



JUSCELIA DIAS MENDONÇA

**Modelos Multidimensionais de Item na Avaliação
de Ingresso da UFLA**

**LAVRAS - MG
2016**

JUSCELIA DIAS MENDONÇA

**Modelos Multidimensionais de Item na Avaliação
de Ingresso da UFLA**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Estatística e Experimentação Agrope-
cuária, área de concentração em Estatística e Experi-
mentação Agropecuária, para a obtenção do título de
Doutor.

Dr. Júlio Sílvio de Sousa Bueno Filho

Orientador

LAVRAS - MG

2016

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pela própria autora.

Mendonça, Juscelia Dias.

Modelos Multidimensionais de Item na Avaliação de Ingresso da
UFLA / Juscelia Dias Mendonça. – Lavras : UFLA, 2016.

126 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2016.

Orientador: Júlio Sílvio de Souza Bueno Filho.

Bibliografia.

1. Teoria de resposta ao item multidimensional. 2. Seleção de
modelos. 3. Avaliação educacional. 4. Dimensão do modelo de item.
5. Forward. I. Universidade Federal de Lavras. II. Título.

JUSCELIA DIAS MENDONÇA

**Modelos Multidimensionais de Item na Avaliação
de Ingresso da UFLA**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Estatística e Experimentação Agrope-
cuária, área de concentração em Estatística e Experi-
mentação Agropecuária, para a obtenção do título de
Doutor.

APROVADO em 25 de agosto de 2016.

Dr. Júlio Sílvio de Sousa Bueno Filho
Orientador

LAVRAS - MG

2016

AGRADECIMENTOS

Houve um tempo em minha vida que ter o título de doutor era algo "inatingível", um verdadeiro sonho, porém oportunidades foram surgindo e hoje estou tornando este sonho realidade. Contudo, para chegar até aqui foi necessário muito esforço da minha parte e também, a colaboração e apoio de várias pessoas.

Em primeiro lugar, agradeço a Deus por ter fornecido as condições necessárias para a conclusão deste trabalho, por orientar meus caminhos e decisões.

Agradeço à minha mãe, por todo apoio durante minha vida, em especial nestes anos de aluna da pós-graduação. Obrigada por me apoiar, incentivar, acalmar e por cuidar tão bem de meu filho nos momentos que estive ausente.

Agradeço ao meu filho Mateus, por ser um bom filho, proporcionando assim tranquilidade para concentrar em meus estudos. Agradeço também pelo apoio e compreensão nos momentos de ausência. Saiba que sua existência foi um grande incentivo para completar essa jornada.

Agradeço ao meu orientador, Professor Júlio Sílvio de Sousa Bueno Filho por todo tempo dedicado a este trabalho, pela disposição em ajudar, pelas conversas sobre a tese e pela paciência. Sem sua ajuda este trabalho não seria concluído.

Agradeço a todos os professores que tive durante toda minha vida, em especial ao Professor Ednaldo Carvalho Guimarães, pessoa que me falou da pós graduação da UFLA, forneceu materiais de estudo e motivou a ingressar no mestrado.

Agradeço a todos os professores e funcionários do Departamento de

Ciências Exatas sempre dispostos a ajudar no que foi necessário. Em especial, agradeço ao Professor Ulisses Azevedo Leitão por toda ajuda na conclusão do mestrado e também, pela disposição em ajudar na elaboração do segundo artigo deste trabalho.

Agradeço a todos os meus colegas do doutorado, em especial à Larissa, Danielle e Tábata. Vocês são muito especiais, aprendi muito com vocês. Larissa, obrigada por sempre me receber tão bem em sua casa sempre que fui à Lavras.

Agradeço a todos os meus colegas de trabalho na UFMT, em especial ao Anderson, Juliano e Marcelino, pela paciência e disposição em ajudar.

Agradeço ao meu irmão Juliano, às minhas amigas e familiares de Uberlândia, obrigada pelo incentivo!

A todos, muito obrigada!

LISTA DE FIGURAS

Figura 1	A: teste multivariado sem correlação entre os traços latentes; B: teste multivariado com correlação entre os traços latentes; C: teste com itens multidimensionais.	22
Figura 2	Retas com todos os pontos onde a probabilidade de acerto do item é $\frac{1}{2}$	32
Figura 3	Gráfico com contornos do item com: $a_1 = 0,75$ $a_2 = 1,5$ e $d = -0,75$	34
Figura 4	Item A: $a_1 = 0,7$, $a_2 = 0$, $c = 0,6$ $d = -2$; Item B: $a_1 = 0,9$, $a_2 = 0,7$, $c = 0$, $d = 1,4$; Item C: $a_1 = 0,8$, $a_2 = 0,8$, $c = 0,2$, $d = -9,7$; Item D: $a_1 = 1,3$, $a_2 = 3,2$, $c = 0,2$, $d = -0,6$;	35
Figura 5	Gráfico com Três Curvas Características do Item com discriminação $a_j = 1,7$, acerto casual $c_j = 0$ e diferentes dificuldades	37
Figura 6	Gráfico de contornos e setas para superfície do item B da Figura 4 Parâmetros do Item B: $a_1 = 0,9$, $a_2 = 0,7$, $c = 0$, $d = 1,4$, $A = 1,14$, $B = -1,23$	38
Figura 7	Gráfico de setas para dois itens	41
Figura 8	SRI e gráficos de contorno com o vetor dos itens 1 e 2 da Tabela 1	42
Figura 1	A: teste multivariado sem correlação entre os traços latentes; B: teste multivariado com correlação entre os traços latentes; C: teste com itens multidimensionais.	99
Figura 2	Gráfico de vetores dos itens 5, 39, 42 e 57, representantes típicos de cada grupo formado ao analisar as habilidades como tendo 2 dimensões.	106
Figura 3	Gráfico de vetores de todos os itens da prova	107
Figura 4	Posicionamento dos itens por grupo, com relação à dificuldade e discriminação.	110
Figura 5	Superfície de Resposta do Item 9.	112
Figura 6	Superfície de Resposta do Item 18.	113
Figura 7	Gráficos de contornos com setas dos itens 9, 18, 50 e 36.	113
Figura 8	Superfície de Resposta do Item 50.	115
Figura 9	Superfície de Resposta do Item 36.	116
Figura 10	Dispersão das habilidades.	119
Figura 11	Gráfico de SIT e Contorno da SIT com posicionamento dos itens.	120

LISTA DE TABELAS

Tabela 1	Parâmetros de dois itens multidimensionais.	40
Tabela 2	Provas do Exame Vestibular do segundo semestre de 2006 da Universidade Federal de Lavras (UFLA)	64
Tabela 3	Provas do Exame Vestibular do segundo semestre de 2006 da Universidade Federal de Lavras (UFLA)	65
Tabela 4	Comparação dos Modelos Unidimensionais, Bidimensionais e Tridimensionais conforme procedimento descrito em Béguin e Glas (2001)	66
Tabela 1	Provas do Exame Vestibular do segundo semestre de 2006 da Universidade Federal de Lavras (UFLA)	82
Tabela 2	Resultados dos testes para a inclusão da segunda dimensão a cada item do teste	83
Tabela 3	Critério DIC para comparação de modelos	84
Tabela 4	Estimativas dos parâmetros (moda <i>a posteriori</i>) do modelo selecionado pelo método <i>forward</i>	86
Tabela 5	Coefficientes de correlação (%) de Pearson entre as estimativas (moda <i>a posteriori</i>) dos parâmetros de discriminação multidimensional estimados para o modelo segundo a estratégia <i>forward</i> de seleção e para os modelos unidimensional, bidimensional e tridimensional	87
Tabela 6	Coefficientes de correlação (%) de Pearson entre as estimativas (moda <i>a posteriori</i>) dos parâmetros de dificuldade multidimensional estimados para o modelo segundo a estratégia <i>forward</i> de seleção e para os modelos unidimensional, bidimensional e tridimensional	88
Tabela 7	Coefficientes de correlação (%) de Pearson entre as estimativas (moda <i>a posteriori</i>) dos parâmetros de acerto casual estimados para o modelo segundo a estratégia de seleção <i>forward</i> e para os modelos unidimensional, bidimensional e tridimensional	88
Tabela 8	Coefficientes de correlação (%) de Pearson entre as estimativas (moda <i>a posteriori</i>) dos parâmetros da habilidade sintética unidimensional para o modelo segundo a estratégia de seleção <i>forward</i> e para os modelos unidimensional, bidimensional e tridimensional	89

Tabela 9	Coeficientes de correlação de Spearman (%) entre as estimativas (moda <i>a posteriori</i>) dos parâmetros da habilidade sintética unidimensional para o modelo segundo a estratégia de seleção <i>forward</i> e para os modelos unidimensional, bidimensional e tridimensional	89
Tabela 10	Estatísticas descritivas do desempenho dos examinados	90
Tabela 11	Classificação de alguns examinados considerando a nota da prova e a habilidade estimada pelo método <i>forward</i> (HF) e a habilidade sintética estimada pelo modelo bidimensional (HB)	90
Tabela 1	Ângulos separadores de itens	103
Tabela 2	Parâmetros estimados pelo modelo bidimensional.	105
Tabela 3	Distribuição dos itens dentro de cada grupo.	108
Tabela 4	Análise didático-pedagógica dos 10 itens mais difíceis e de 3 itens com grande discriminação.	109
Tabela 5	Distribuição dos itens da prova por grupos.	117
Tabela 6	Estatísticas descritivas das notas dos examinados	118

SUMÁRIO

	PRIMEIRA PARTE	12
1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	17
2.1	Modelos da Teoria de Resposta ao Item	23
2.2	Modelos Unidimensionais da TRI	25
2.3	Modelos Multidimensionais da TRI	27
2.3.1	Modelos Multidimensionais Compensatórios	28
2.3.2	Modelos Multidimensionais Parcialmente Compensatórios	28
2.4	Modelos Multidimensionais Compensatórios de Resposta ao Item	28
2.5	Parâmetro de Dificuldade e Discriminação	36
2.6	Suposições dos Modelos de TRIM	43
2.7	Função de Verossimilhança	44
2.8	Dimensionalidade de um conjunto de dados	45
3	MÉTODOS DE ESTIMAÇÃO	47
3.1	Versão Bayesiana da TRIM: Béguin e Glas (2001)	47
4	MATERIAL E MÉTODOS	64
5	RESULTADOS	65
6	REFERÊNCIAS	67
	SEGUNDA PARTE - ARTIGOS	71
	ARTIGO 1- Determinação da Dimensão em Modelos de Item Para Dados dos Candidatos ao Curso de Agronomia na Avaliação de Ingresso à UFLA 2006-2	72
1	INTRODUÇÃO	73
2	MATERIAL E MÉTODOS	75
3	RESULTADOS	83
4	CONCLUSÃO	91
	REFERÊNCIAS	91
	ARTIGO 2 - Análise Pedagógica de uma Prova Multidisciplinar Usando Modelos Bidimensionais de Item	95
1	INTRODUÇÃO	95

2	METODOLOGIA	100
3	RESULTADOS	104
4	CONCLUSÃO	122
	REFERÊNCIAS	123
	CONSIDERAÇÕES FINAIS	126

RESUMO

A Teoria de Resposta ao Item (TRI) tem-se mostrado uma metodologia adequada na análise de dados educacionais, todavia a escolha de um, dentre os diversos modelos disponíveis, não é uma tarefa simples, em especial quando se pode variar a escolha da dimensão dos itens. Neste trabalho procuramos elucidar aspectos estatísticos e potencialidades de interpretação de modelos multidimensionais da TRI (TRIM). Na primeira parte deste trabalho a literatura da área é revista, com enfoque para modelos multidimensionais compensatórios de item. Esta revisão incluiu, para fins didáticos, a descrição completa de um modelo de análise bayesiana de modelos TRIM, que é o estado da arte na literatura. Esta análise foi exemplificada com dados obtidos na avaliação de ingresso do segundo semestre de 2006, pelos candidatos ao curso de Agronomia da Universidade Federal de Lavras (UFLA). O modelo mais bem ajustado nesta etapa foi o bidimensional. A segunda parte é composta de dois artigos, sendo o primeiro uma proposta de seleção de modelos do tipo *forward* para identificar a dimensão adequada para modelar cada item do mesmo conjunto de dados. No segundo artigo encontra-se a análise pedagógica baseada nos mesmos dados, usando o modelo bidimensional. Um capítulo final de considerações gerais resume o que se aprendeu nesta pesquisa.

Palavras Chave: Teoria de resposta ao item multidimensional. Seleção de modelos. Avaliação educacional. Aplicação. Forward.

ABSTRACT

Item Response Theory (IRT) has been shown as a worthy tool to analyse educational data. However, choosing one out of many models is not a simple task. This is specially true when you can vary the dimension of models for each item. In this thesis we try to enlighten some statistical properties and the analytical potential to interpret multidimensional models of IRT (MIRT). In the first part a review of the literature is done focussing on compensatory multidimensional models for items. This part bring a didactic effort to present a state of art description of the fully Bayesian MIRT analysis. This analysis is illustrated with data from entrance examination of candidates for Agronomy undergraduate program in the second semester of 2006 at the Universidade Federal de Lavras (UFLA). The best model in this section is the one that has two dimensions for all items. Second part has two papers. The first discusses a forward type of model selection procedure to identify different dimensions for each model. The second bring a pedagogical analysis based on the same data using two dimensions model. As a final chapter we draw the last remarks on what was learned in our investigation.

Key-words: Educational evaluation. Forward. Model selection. Multidimensional item response theory.

PRIMEIRA PARTE

1 INTRODUÇÃO

A Teoria de Resposta ao Item (TRI) é um conjunto de modelos matemáticos para análise de variáveis latentes. Atualmente é utilizada em diversas áreas como: educação, qualidade de vida, psiquiatria, psicologia, satisfação do consumidor e várias outras. Na área educacional, a TRI tem sido usada em diversas avaliações no Brasil e no mundo. Através de seus resultados o governo, as escolas e professores tem implementado medidas para melhoria da educação.

Os modelos da TRI requerem que o pressuposto da independência local, ou condicional, seja atendido. A validade deste pressuposto implica que as respostas de diferentes examinados, a diferentes itens sejam independentes. A independência citada é utilizada no cálculo da função de verossimilhança do modelo adotado (LORD, 1980). A propriedade de invariância dos parâmetros, que permite estimar a habilidade dos examinados, independentemente dos parâmetros dos itens, também é consequência do pressuposto da independência local.

Dessa forma, observa-se que o atendimento do pressuposto de independência local é extremamente importante nos modelos da TRI. Este pressuposto é atendido, somente se, o modelo utilizado possui o número correto de habilidades necessárias para que todos examinados possam responder corretamente aos itens de um teste (LORD, 1980). Logo, para utilizar qualquer modelo da TRI é essencial a determinação da correta dimensão do conjunto de dados a ser analisado (CONDE, 2007).

Os modelos da TRI podem ser classificados em compensatórios e

não compensatórios. Nos modelos compensatórios, admite-se que, a mesma probabilidade de acerto obtida com valores moderados de habilidades, possa ser obtida compensando-se baixos valores de quaisquer habilidade do modelo, com valores altos em outras habilidades. Nos modelos parcialmente compensatórios essa propriedade não ocorre.

Um modelo da TRI que utiliza apenas uma habilidade é chamado de unidimensional. Já os modelos que apresentam mais de uma habilidade são chamados de multidimensionais. Os modelos de TRI unidimensionais admitem que, cada item de um teste mede apenas um único traço latente e todos os itens de um teste, medem o mesmo traço latente. No entanto, essa suposição não é natural pois, o desempenho humano em qualquer atividade está ligado a diversos fatores. Dessa forma, é mais natural pensar que os níveis de mais de uma habilidade, presentes no indivíduo, estão determinando se o mesmo irá acertar ou errar um item de um teste.

Comumente nos modelos da TRI unidimensional, admite-se a ideia de que, dentre todas as várias variáveis que caracterizam o desempenho humano, existe um único fator dominante, sendo esse fator, o principal responsável pela resposta de indivíduo a um item. Contudo, a existência desse fator deve ser investigada. Nojosa (2002) constatou que a prova do ENEM, que utiliza a TRI unidimensional em sua análise, mede no mínimo 5 habilidades diferentes e não apenas uma, conforme exigido pelo modelo utilizado para análise.

Por todos os motivos citados, a determinação do número de habilidades, ou eixos necessários para modelar a interação entre itens e examinados, presente na matriz de dados a ser analisada por modelos da TRI é extremamente importante. Quando é necessário mais que um eixo para realizar

tal modelagem, será usado um modelo da Teoria de Resposta ao Item Multidimensional (TRIM).

Na literatura são apresentadas diversas formas de determinação da dimensão de conjunto de dados como: o DIMTEST, o DETECT, a Análise Paralela, diversas versões do Teste Qui-quadrado, versões estas que contam com modificações da estatística qui-quadrado (RECKASE, 2009). Estas técnicas não são interessantes para a seleção de modelos com diferentes dimensões, pois têm fundamentos não-paramétricos e não compreendem uma clara estratégia de testes sequenciais, como utilizado em modelos encaixados de regressão. Pouco se encontra na literatura de TRIM sobre a seleção de modelos com dimensionalidades diferentes para cada item. No entanto, isto poderia ser feito com medidas de adequação como os critérios de informação com base na verossimilhança (direta ou marginal). Dentre estes destacam-se o Deviance Information Criterion, DIC; o Akaike Information Criterion, AIC e o Bayesian Information Criterion, BIC (RAFTERY et al., 2007).

Pelo que foi exposto, reiteramos que a motivação deste trabalho vem da importância da correta determinação das dimensões dos modelos de itens que se ajustam a um conjunto de dados em TRI. Isto pode ter implicações na avaliação da qualidade dos itens e na estimação de habilidades globais dos examinados.

OBJETIVOS

Os objetivos deste trabalho são:

- apresentar modelos multidimensionais compensatórios de item e uma implementação numérica destes em R;
- desenvolver um protocolo para a determinação da dimensão de um conjunto de dados para posterior modelagem através da TRI;
- aplicar o protocolo e identificar a dimensão de um exame real;
- verificar as possibilidades de discussão da qualidade dos itens ao interpretar os parâmetros de modelos multidimensionais em um exame real.

2 REFERENCIAL TEÓRICO

Variáveis que não podem diretamente ser mensuráveis são chamadas de variáveis latentes. Dessa forma, o conhecimento é uma variável latente. Em diversas situações têm-se como objetivo medir o conhecimento dos indivíduos e uma maneira de medir o conhecimento é pela observação da resposta que um examinado dá a um item, pois são variáveis relacionadas.

A Teoria de Resposta ao Item (TRI) propõe modelos para variáveis latentes. A TRI, também denominada por Teoria da Curva Característica ou Teoria do Traço Latente é um conjunto de modelos matemáticos que considera o item como unidade básica de análise e procura representar a probabilidade de um indivíduo dar certa resposta a um item em função dos parâmetros do item e do(s) traço(s) latente(s) do indivíduo (ANDRADE; TAVARES; VALLE, 2000).

Na educação, a TRI tem sido usada no Brasil e no mundo em diversas avaliações como: o PISA (Programme for International Student Assessment), realizado pela Organização para a Cooperação e o Desenvolvimento Econômico (OECD); o teste TOEFL (Test of English as a Foreign Language) exame de proficiência em língua inglesa, o SAEB (Sistema de Avaliação da Educação Básica), a Prova Brasil e o ENEM (Exame Nacional do Ensino Médio), os três últimos realizados pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira).

Os resultados obtidos em avaliações são muito importantes, pois são reflexo de uma determinada realidade e geradores de tomadas de decisões. Através desses resultados, o governo planeja e implementa as políticas públicas para melhoria da Educação, gerencia a distribuição de recursos e mobiliza as redes de ensino. As escolas e professores, por sua vez, tomam

conhecimento do nível de desenvolvimento de seus alunos e buscam caminhos pedagógicos para possíveis intervenções.

Cada vez mais, os resultados obtidos pelas avaliações educacionais são utilizados como: índice da qualidade da educação oferecida pelas escolas. A divulgação de resultados como o indicador nacional de desenvolvimento da Educação, o IDEB, em jornais e em outros meios de comunicação, tem provocado discussões sobre a qualidade da educação brasileira. Existe uma grande variação de resultados nas diferentes localidades brasileiras e, também, entre os resultados brasileiros com os de outros países. A comparação com outros países é feita através dos resultados obtidos pela avaliação PISA, na qual o Brasil ocupa atualmente a 60^a posição em um total de 76 países avaliados.

O IDEB foi criado pelo INEP em 2007, sendo medido em uma escala de zero a dez. Sintetiza dois conceitos igualmente importantes para a qualidade da educação: aprovação e média de desempenho dos estudantes em língua portuguesa e matemática. O indicador é calculado a partir dos dados sobre aprovação escolar, obtidos no Censo Escolar, e médias de desempenho nas avaliações do INEP, o SAEB e a Prova Brasil. O IDEB é um indicador de qualidade e foi desenvolvido para avaliar a educação brasileira em cada unidade da federação, município e escola pública. Ele é também usado para acompanhar a melhoria da qualidade da educação de acordo com metas preestabelecidas. Seu uso todavia, tem sido criticado por focar apenas nos resultados finais, não considerando as diversas variáveis explicativas do desempenho dos estudantes como o perfil dos alunos e as características do estabelecimento de ensino.

A avaliação PISA avalia conhecimentos de leitura, matemática e ci-

ências de estudantes na faixa dos 15 anos, idade em que se pressupõe o término da escolaridade básica obrigatória na maioria dos países. O programa é desenvolvido e coordenado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) e em cada país participante há uma coordenação nacional. No Brasil, o PISA é coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (<http://portal.inep.gov.br/web/portal-ideb>)

Portanto, é fácil perceber a importância da análise correta dos dados obtidos em avaliações educacionais, pois estas são responsáveis por diversas ações do governo e sociedade. O uso da TRI como metodologia da análise desses dados trouxe diversos avanços comparada com sua precursora, a Teoria Clássica dos Testes (TCT). Entre as vantagens do uso da TRI, pode-se citar:

1. Diferentes pessoas, ou a mesma pessoa, em diferentes ocasiões podem ter suas habilidades comparadas a partir de itens comuns nos testes (técnica de equalização).
2. Os parâmetros dos itens obtidos por meio da TRI são medidas estatisticamente independentes da amostra de respondentes, ou seja, são constantes independentemente da amostra na qual são aplicadas (propriedade da invariância).
3. O grau de dificuldade das questões é considerado para a composição da nota do examinado. Dessa forma, pessoas com o mesmo número de acertos poderão ter notas diferentes, pois, dá-se mais valor às questões mais difíceis.
4. A invariância dos parâmetros de itens possibilita a utilização da testa-

gem adaptativa computadorizada (CAT), em que o computador escolhe as próximas questões de acordo com o desempenho do examinado em questões anteriores.

A TRI foi utilizada no Brasil pela primeira vez em 1995, no terceiro ciclo de avaliação do SAEB. Anteriormente era usada a Teoria Clássica dos Testes, em que a nota de cada examinado é obtida pela proporção de questões acertadas. No caso de questões com pesos diferentes, os pesos são pré-determinados arbitrariamente pelos elaboradores da prova. Assim é obtida a nota de cada item e a nota do teste é novamente obtida pela soma da nota de todos os itens.

Pela TCT, o índice de dificuldade das questões é calculado apenas pelo percentual de erros, variando de acordo com o do nível de habilidade dos respondentes, ou seja, um mesmo teste submetido a dois grupos de respondentes, em que o primeiro possui elevado nível de habilidade, o segundo baixo nível de habilidade, poderá apresentar, dois valores distintos de índice de dificuldade, isto é, um valor pequeno produzido para o primeiro grupo e um valor grande para o segundo grupo.

O índice de discriminação das questões, ou seja, a capacidade do item diferenciar os examinados de alta habilidade dos de baixa habilidade, é calculado pela TCT através das correlações entre o item e a nota total (correlações bisserial e ponto-bisserial). Já o percentual de acertos ao acaso é fixo na TCT e estimado pela probabilidade de acerto da questão por mero acaso ($1/n^o$ de alternativas do item).

A Teoria de Resposta ao Item Multidimensional (TRIM) é definida por Reckase (2009) como um conjunto de modelos estatísticos que busca modelar a interação entre as características dos examinados e dos itens.

Os modelos da TRIM, também conhecidos na literatura como Modelos da Teoria de Resposta ao Item (TRI) Multidimensionais Dentro-do-Item, são modelos adequados para itens que medem dois ou mais de um traços latentes.

A dimensão de um item é o número mínimo de habilidades necessárias para que determinado grupo de indivíduos possa respondê-lo corretamente. Na TRIM, há uma interação entre respondentes e itens, por isso, não existem itens unidimensionais ou multidimensionais e sim, respostas que precisam ser analisadas com determinado modelo, ou seja, um mesmo item pode ser multidimensional para um determinado grupo de examinados e unidimensional para outro grupo. (RECKASE, 2009).

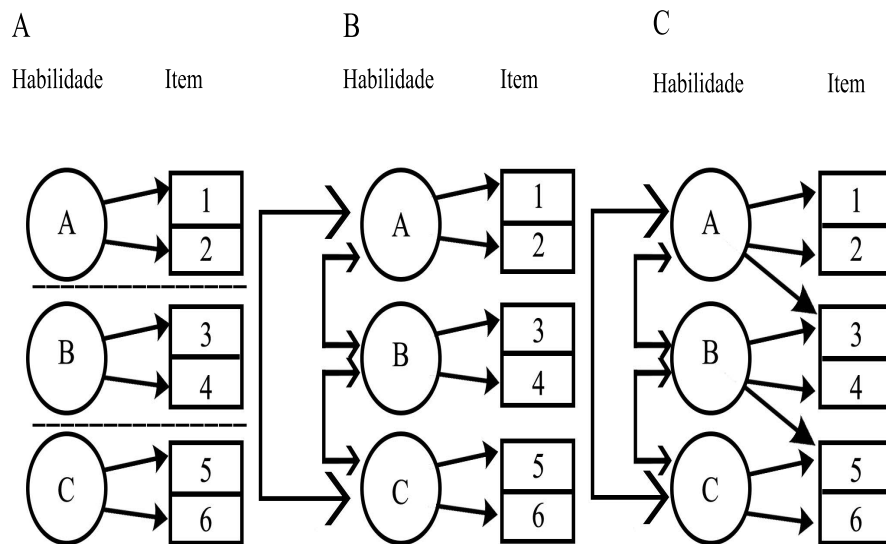
Os modelos, Unidimensional e Multivariado são casos particulares do modelo Multidimensional. O modelo Multidimensional é também chamado de Multidimensional Dentro do Item e, o modelo Multivariado é também chamado de Multidimensional Entre Itens (VELOSO, 2014;MATOS, 2008). O modelo multivariado é adequado para analisar testes, ou seja, conjuntos de itens que podem ser divididos em sub-testes que possuem apenas itens unidimensionais. Na Figura 1 observa-se um esquema com um modelos Multivariados (A e B) e modelos Multidimensionais (C).

- Em A, observa-se um teste multivariado que pode ser dividido em três sub-testes unidimensionais, para cada item 1,2,...,6 é necessária apenas uma habilidade para que o item seja respondido corretamente. As habilidades A, B e C não são correlacionadas e, cada um dos sub-testes usa uma única habilidade, A, B e C respectivamente para ser respondido (três sub-testes unidimensionais).
- Assim como em A, tem-se em B um teste multivariado que também

pode ser dividido em 3 sub-testes unidimensionais. Para responder a prova toda são necessárias as habilidades correlacionadas A, B e C, cada item do teste necessita apenas de uma habilidade para ser respondido corretamente.

- Em C, observa-se um teste composto por itens que necessitam de mais de uma habilidade para serem respondidos corretamente (itens 3 e 5), por isso, trata-se de um teste multidimensional que não pode ser dividido em sub-testes unidimensionais.

Figura 1 –**A**: teste multivariado sem correlação entre os traços latentes;**B**: teste multivariado com correlação entre os traços latentes; **C**: teste com itens multidimensionais.



Fonte: Adaptado de Wang (2004).

As respostas de um indivíduo a um teste podem ser classificadas de

forma:

- **Dicotômica:** Nesse caso as respostas são consideradas apenas como certas ou erradas.
- **Politômica:** Além das respostas certo ou errado são atribuídas probabilidades para cada categoria.
- **Contínua:** São usados valores dentro de intervalos reais para classificar as respostas, muito usado nas questões abertas.

É importante ressaltar que as questões abertas também podem ser corrigidas como certas ou erradas.

2.1 Modelos da Teoria de Resposta ao Item

A resposta de um examinado a um item é uma variável aleatória representada por Y_{ij} , $i=1,\dots,N$ (N é o número total de examinados), $j=1,\dots,J$ (J é o número total de itens do teste). As respostas de todos os examinados, a todos os itens de um teste, são representadas pela matriz de respostas \mathbf{Y} .

Em itens dicotômicos, ou seja, corrigidos como certo ou errado, a variável aleatória Y_{ij} assume os valores 0 ou 1; o valor 0 representa uma resposta incorreta ao item e o valor 1, por sua vez, representa a resposta correta. Dessa forma, a variável resposta Y_{ij} possui distribuição de probabilidade de Bernoulli(P_{ij}) (VELOSO, 2014):

$$f_{Y_{ij}}(y_{ij}; \pi_{ij}) = \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}}. \quad (1)$$

Os modelos de TRI para itens dicotômicos possuem a seguinte representação genérica exposta no modelo probabilístico:

$$\pi_{ij} = P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, \mathbf{b}_j, c_j) = c_j + (1 - c_j)F(\eta_{ij}) \quad (2)$$

A probabilidade da Equação 2 é chamada de Função de Resposta ao Item (FRI) na qual:

- $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}, \dots, \theta_{iK})$ é um vetor K -dimensional com as habilidades do indivíduo i , $\boldsymbol{\theta}_i \in \mathbb{R}^K$;
- $\mathbf{a}_j = (a_{j1}, a_{j2}, \dots, a_{jk})$ é um vetor com dimensão K que contém os parâmetros relacionados à discriminação do item j , $\mathbf{a}_j \in \mathbb{R}^K$;
- $\mathbf{b}_j = (b_{j1}, b_{j2}, \dots, b_{jk})$ é um vetor com dimensão K que contém os parâmetros de dificuldade do item, $\mathbf{b}_j \in \mathbb{R}^K$;
- c_j é o parâmetro que representa a probabilidade de acerto de indivíduos de baixa habilidade, $c_j \in \mathbb{R}$;
- η_{ij} é o preditor linear, uma função adequadamente escolhida dos parâmetros $\boldsymbol{\theta}_i$, \mathbf{a}_j e \mathbf{b}_j , $\eta_{ij} \in \mathbb{R}$,

$$\eta_{ij} = \mathbf{a}_j(\boldsymbol{\theta}'_i + \mathbf{b}'_j);$$
- F é uma função de distribuição, probito ou logito por exemplo;
- K é a dimensão do item, ou seja, o número de habilidades necessárias para que o item seja respondido corretamente.

Portanto, através da Equação 2, verifica-se que a probabilidade de acerto de um indivíduo a um item depende da habilidade do indivíduo e das características do item.

Os vetores K-dimensionais $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}, \dots, \theta_{iK})$, são chamados por Lord (1968) de espaço latente completo se nele estiverem presentes todas as variáveis latentes θ_{ik} que discriminam os examinados. A proficiência de cada examinado é determinada pela posição do examinado em cada traço latente.

2.2 Modelos Unidimensionais da TRI

Quando $K=1$, tem-se os Modelos Unidimensionais da Teoria de Resposta ao Item, ou Teoria de Resposta ao Item unidimensional. Nestes modelos, se a função de ligação F , mostrada na Equação 2, for a função distribuição normal tem-se os modelos denominados probitos, já se a função F usada for a distribuição logística, tem-se os modelos logitos.

Modelo da Ogiva Normal de 1 Parâmetro ou Modelo Probit de 1 Parâmetro

Este modelo possui apenas o parâmetro associado à dificuldade do item, ou seja, o parâmetro b_j . Este parâmetro deve estar na mesma escala da habilidade e pode assumir qualquer valor real. O preditor linear do modelo é: $\theta_i - b_j$; Assim a FRI é dada por:

$$P(Y_{ij}|\theta_i, b_j) = \Phi(\theta_i - b_j), \text{ com } \theta_i \in \mathbb{R} \text{ e } b_j \in \mathbb{R}. \quad (3)$$

Modelo da Ogiva Normal de 2 Parâmetros ou Modelo Probit de 2 Parâmetros

Neste modelo o preditor linear é: $a_j(\theta_i - b_j)$. Veja que se $a_j = 1$, o modelo de 2 parâmetros é igual ao de 1 parâmetro, o parâmetro a_j é proporcional à inclinação da reta tangente à curva no ponto de inflexão (BRAGION, 2011). Assim a FRI é dada por:

$$P(Y_{ij} = 1|\theta_i, a_j, b_j) = \Phi[a_j(\theta_i - b_j)] \text{ com } \theta_i \in \mathbb{R}, b_j \in \mathbb{R} \text{ e } a_j \in \mathbb{R}. \quad (4)$$

Modelo da Ogiva Normal de 3 Parâmetros ou Modelo Probit de 3 Parâmetro

Neste modelo, além dos parâmetros descritos anteriormente, também é utilizado um parâmetro para representar o acerto casual do item pelos examinados de baixa habilidade, denotado por c_j . Birnbaum (1968) foi quem introduziu este parâmetro ao modelo considerando que examinados com baixa habilidade podem obter respostas corretas para os itens. Observe que para obter os modelos anteriores basta fazer $c_j = 0$.

Dessa forma, a FRI é dada por:

$$P(Y_{ij} = 1|\theta_i, a_j, b_j, c_j) = c_j + (1 - c_j)\Phi[a_j(\theta_i - b_j)].$$

Modelos Logísticos de 1, 2 e 3 parâmetros

Birnbaum (1968) foi o primeiro a usar o modelo logístico para substituir o modelo com a ogiva normal. Tal substituição foi realizada, por dois fatores:

1. A simplicidade do modelo logístico (ou Função Distribuição Logística) no qual não é envolvida nenhuma integral.
2. A proximidade das duas funções de distribuição quando é usado um parâmetro de escala $D=1,702$.

Dessa forma, a FRI de uma forma genérica para os modelos logísticos de 1,2 e 3 parâmetros pode ser obtida pelo modelo probabilístico (MENDONÇA, 2010):

$$\begin{aligned}\pi_{ij} &= P(Y_{ij} = 1|\theta_i, a_j, b_j, c_j) = c_j + (1 - c_j)\Psi(\eta_{ij}) \\ &= c_j + (1 - c_j)\frac{1}{1 + e^{-D[a_j(\theta_i - b_j)]}}.\end{aligned}\tag{5}$$

(Modelo Logístico de 3 Parâmetros)

O Modelo Logístico de 1 parâmetro é obtido substituindo $a_j = 1$ e $c_j = 0$ na Equação 5. Já o Modelo Logístico de 2 parâmetros é obtido substituindo apenas $c_j = 0$.

2.3 Modelos Multidimensionais da TRI

Os Modelos Multidimensionais da TRI podem ser classificados como: Modelos Compensatórios e Parcialmente Compensatórios.

2.3.1 Modelos Multidimensionais Compensatórios

Nos modelos compensatórios, um ou mais valores baixos das habilidades presentes no modelo, podem ser compensados, com altos valores em outras habilidades. Dessa forma, a mesma probabilidade é obtida quando são utilizados valores moderados das habilidades. Nesses modelos, as habilidades são combinadas linearmente e essa combinação é o preditor linear, geralmente probito ou logito.

2.3.2 Modelos Multidimensionais Parcialmente Compensatórios

Em modelos parcialmente compensatórios, não é admissível que se possa obter, a mesma probabilidade obtida com habilidades moderadas usando, combinação de altas e baixas habilidades. Esse tipo de modelo separa as habilidade de um item em partes, usando um modelo unidimensional para cada habilidade, e a probabilidade total é obtida através do produto das probabilidades de modelo (RECKASE, 2009).

Neste trabalho serão apresentados modelos compensatórios multidimensionais para itens dicotômicos .

2.4 Modelos Multidimensionais Compensatórios de Resposta ao Item

Os modelos multidimensionais de resposta ao item, para itens dicotômicos são expressos da seguinte forma:

$$\pi_{ij} = P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) = c_j + (1 - c_j)F(\eta_{ij}). \quad (6)$$

O preditor linear é dado por:

$$\begin{aligned}\eta_{ij} &= \mathbf{a}_j(\boldsymbol{\theta}'_i - \mathbf{b}'_j) = \mathbf{a}_j\boldsymbol{\theta}'_i - \mathbf{a}_j\mathbf{b}'_j = a_{j1}\theta_{i1} \\ &+ a_{j2}\theta_{i2} + \dots + a_{jK}\theta_{iK} - a_{j1}b_{j1} - a_{j2}b_{j2} - \dots \\ &- a_{jk}b_{jk}.\end{aligned}$$

Fazendo : $-a_{j1}b_{j1} - a_{j2}b_{j2} - \dots - a_{jK}b_{jK} = -\sum_{k=1}^K a_{jk}b_{jk} = d_j$
, obtem-se a seguinte expressão:

$$\eta_{ij} = \sum_{k=1}^K a_{jk}\theta_{jk} + d_j$$

,

ou na forma vetorial:

$$\eta_{ij} = \mathbf{a}_j(\boldsymbol{\theta}'_i - \mathbf{b}'_j) = \mathbf{a}_j\boldsymbol{\theta}'_i + d_j.$$

Logo:

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) = c_j + (1 - c_j)F(\mathbf{a}_j\boldsymbol{\theta}'_i + d_j).$$

- $\boldsymbol{\theta} = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}, \dots, \theta_{iK})$ é o vetor de habilidades do i-ésimo indivíduo e K é a dimensão do item, $\boldsymbol{\theta}_i \in \mathbb{R}^K$;
- $\mathbf{a}_j = (a_{j1}, a_{j2}, \dots, a_{kj})$ é o vetor relacionado à discriminação do item, $\mathbf{a}_j \in \mathbb{R}^K$;
- $\mathbf{b}_j = (b_{j1}, b_{j2}, \dots, b_{jk})$ é um vetor com dimensão K que contém os parâmetros de dificuldade do item, $b_j \in \mathbb{R}^K$;
- d_j é o parâmetro relacionado à dificuldade do item, $d_j \in \mathbb{R}$;
- c_j é o parâmetro relacionado ao acerto de indivíduos de baixa habilidade, $c_j \in \mathbb{R}$;

- $P(Y_{ij} = 1|\boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j)$ é a probabilidade do indivíduo i com habilidade $\boldsymbol{\theta}_i$ acertar o item j com parâmetros de \mathbf{a}_j, d_j e c_j .

Se na Equação 6, a função F é a Função de Distribuição Normal, o modelo é então chamado de Modelo Multidimensional Probit de 3 Parâmetros. O Modelo Multidimensional Probit de 2 Parâmetros pode ser obtido fazendo $c_j = 0$ e o Modelo Multidimensional Probit de 1 Parâmetro pode ser obtido substituindo $c_j = 0$ e \mathbf{a}_j por um vetor de dimensão $1 \times K$ com todos os elementos iguais a 1.

Por sua vez, se na Equação 6 a função F é a Função de Distribuição Logística, o modelo é chamado de Modelo Multidimensional Logito de 1,2 ou 3 parâmetros como a seguir:

Modelo Multidimensional Logito de 1 Parâmetro

$$P(Y_{ij} = 1|\boldsymbol{\theta}_i, d_j) = c_j + (1 - c_j)\Psi(-D\boldsymbol{\theta}_i^t) = \frac{1}{1 + e^{-D(\boldsymbol{\theta}_i^t + d_j)}}. \quad (7)$$

Modelo Multidimensional Logito de 2 Parâmetros

$$P(Y_{ij} = 1|\boldsymbol{\theta}_i, \mathbf{a}_j, d_j) = c_j + (1 - c_j)\Psi[-D(\mathbf{a}_j\boldsymbol{\theta}_i^t)] = \frac{1}{1 + e^{-D(\mathbf{a}_j\boldsymbol{\theta}_i^t + d_j)}}. \quad (8)$$

Modelo Multidimensional Logito de 3 Parâmetros

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) = c_j + (1 - c_j)\Psi[-D(\mathbf{a}_j\boldsymbol{\theta}'_i + d_j)]. \quad (9)$$

Características de um Modelo Compensatório Multidimensional com Dimensão 2 ($K = 2$)

O preditor linear $\eta_{ij} = \sum_{k=1}^2 a_{jk}\theta_{ik} + d_j$ é uma função linear dos elementos de $\boldsymbol{\theta}_i$. Dessa forma, o gráfico de η_{ij} é uma reta, na qual, d_j é o intercepto e \mathbf{a}_j é o vetor de inclinação da reta. Portanto, quando η_{ij} é igual a uma constante qualquer, ou seja, $\eta_{ij} = q$, tem-se uma reta com todos os valores de $\boldsymbol{\theta}_i$ que satisfazem a equação: $\sum_{k=1}^2 a_{jk}\theta_{ik} + d_j = q$.

Uma situação interessante é quando $\eta_{ij} = \sum_{k=1}^2 a_{jk}\theta_{ik} + d_j = 0$ assim, com o Modelo Multidimensional Logito de 2 Parâmetros:

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j) = \frac{1}{1 + e^{-D(\sum_{k=1}^2 a_{jk}\theta_{ik} + d_j)}} = \frac{1}{1 + e^{(0)}} = \frac{1}{2}.$$

No Modelo Unidimensional Logístico de 2 parâmetros, a probabilidade de um indivíduo acertar a um determinado item é igual à $\frac{1}{2}$, apenas, quando sua habilidade for igual à dificuldade do item, ou seja, $\theta_i = b_j$. No entanto, nos modelos multidimensionais existem infinitas combinações de $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}, \dots, \theta_{iK})$ que resultam na probabilidade $\frac{1}{2}$. Um exemplo pode ser obtido com o seguinte modelo bidimensional:

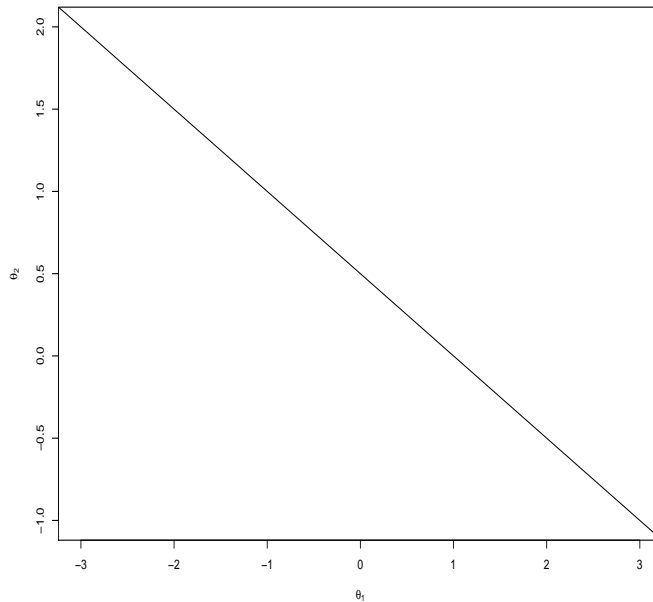
$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j) = \frac{1}{1 + e^{(0,75\theta_{i1} + 1,5\theta_{i2} - 0,75)}}. \quad (10)$$

ndo

A probabilidade $\frac{1}{2}$ é obtida qua: $(0,75\theta_{i1} + 1,5\theta_{i2} - 0,75) = 0$ ou seja, $\theta_{i2} = -0,5\theta_{i1} + 0,5$. Na Figura 2 observa-se os eixos (θ_1, θ_2) e, também, areta cujos pontos são os valores que satisfazem a seguinte probabilidade:

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j) = \frac{1}{1 + e^{(0,75\theta_{i1} + 1,5\theta_{i2} - 0,75)}} = \frac{1}{2}.$$

Figura 2 –Reta com todos os pontos onde a probabilidade de acerto do item é $\frac{1}{2}$.

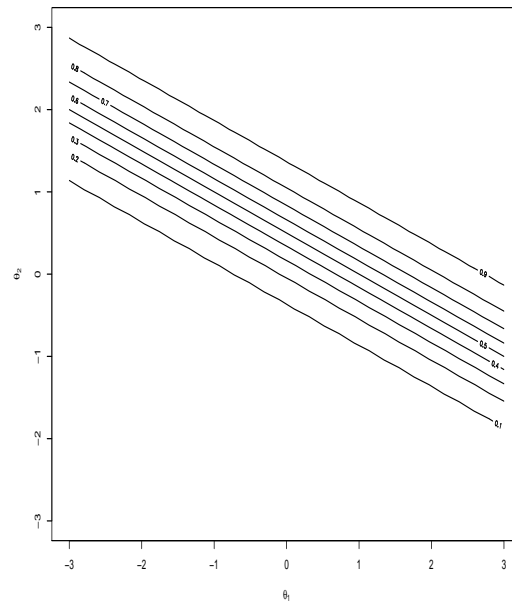


A análise dos pontos da reta permite uma compreensão melhor dos modelos compensatórios em itens bidimensionais. Observa-se que baixos valores em uma das coordenadas, θ_1 ou θ_2 , podem ser compensados com altos valores da outra coordenada de forma que a probabilidade de acerto do indivíduo seja constante.

Na Figura 3 está presente o gráfico de contornos do item descrito na Equação 10. Observa-se neste gráfico diferentes retas, os pontos pertencentes a cada reta retornam sempre a mesma probabilidade de acerto ao item, para os diferentes pares ordenados θ_1, θ_2 . Na reta com probabilidade 0,1, por exemplo, encontra-se todos os pares ordenados (θ_1, θ_2) com os quais o examinado terá probabilidade 0,1 de acertar o item, o mesmo acontece com as demais retas.

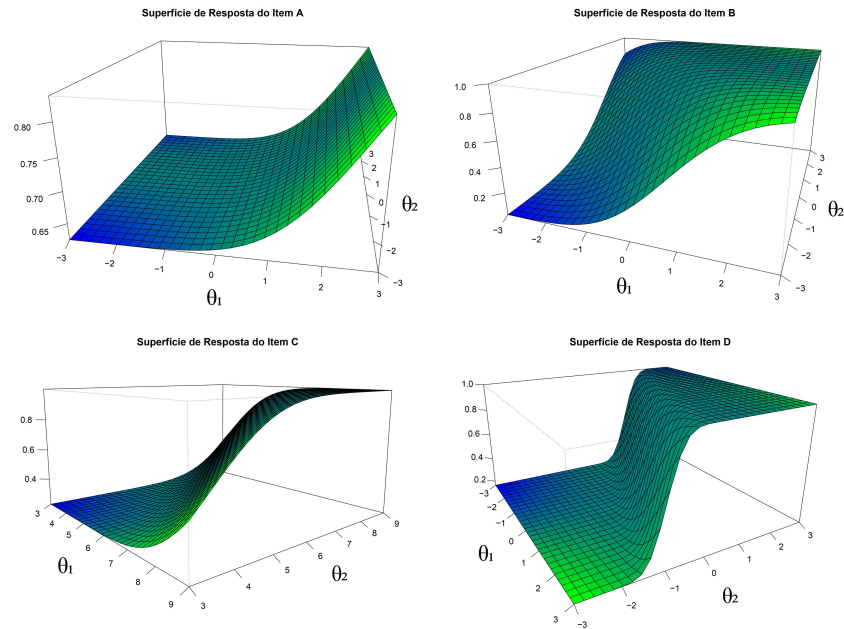
As retas presentes no gráfico de contorno de um item são sempre paralelas e a distância entre elas indica se acréscimos de habilidades implicam em um grande, ou pequeno, aumento na probabilidade de acerto. Retas próximas indicam que pequenos acréscimos na habilidade geram grandes acréscimos na probabilidade de acerto do indivíduo (item com grande discriminação) e o contrário, retas muito distantes, indicam que para ter um pequeno acréscimo na probabilidade de acerto é necessário um grande acréscimo nas habilidades (item com pequena discriminação).

Figura 3 –Gráfico com contornos do item com: $a_1 = 0,75$ $a_2 = 1,5$ e $d = -0,75$



O gráfico de um item multidimensional é chamado de Superfície de Resposta ao Item (SRI). A Figura 4 apresenta 4 diferentes SRI.

Figura 4 – **Item A:** $a_1 = 0,7$, $a_2 = 0$, $c = 0,6$, $d = -2$; **Item B:** $a_1 = 0,9$, $a_2 = 0,7$, $c = 0$, $d = 1,4$; **Item C:** $a_1 = 0,8$, $a_2 = 0,8$, $c = 0,2$, $d = -9,7$; **Item D:** $a_1 = 1,3$, $a_2 = 3,2$, $c = 0,2$, $d = -0,6$; .



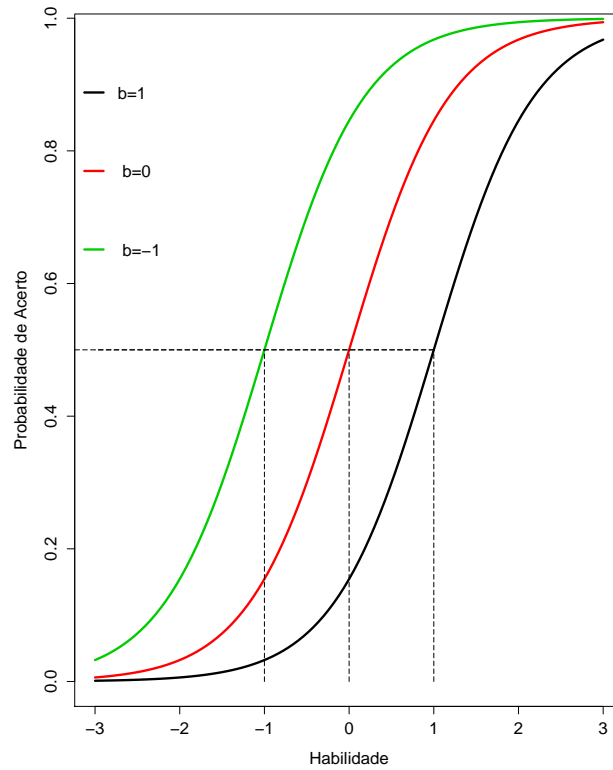
Analisando os 4 gráficos da figura, observa-se que o parâmetro c_j fornece uma assíntota para a Superfície de Resposta ao Item. Além disso, analisando em especial a SRI do item D, que possui um valor muito maior em a_2 do que em a_1 , observa-se que tornando fixo qualquer valor de θ_1 e aumentando os valores de θ_2 , a inclinação da superfície, ou seja, a probabilidade de acerto aumenta muito mais do que quando fixar-se θ_2 e aumenta-se os valores de θ_1 . Logo, o parâmetro de discriminação a_j está relacionado com a inclinação da SRI, ou seja, com a taxa de mudança da probabilidade de acerto em cada eixo.

2.5 Parâmetro de Dificuldade e Discriminação

Na Teoria de Resposta ao Item Unidimensional, os parâmetros de discriminação (a_j) e dificuldade (b_j) estão relacionados com a forma que a Curva Característica do Item (CCI) assume. O parâmetro de dificuldade indica o valor de θ_i no qual a CCI possui maior inclinação (ponto de inflexão), neste ponto a probabilidade de acerto é $0,5(1 + c_j)$. Já o parâmetro de discriminação é proporcional à inclinação da reta tangente no ponto de inflexão.

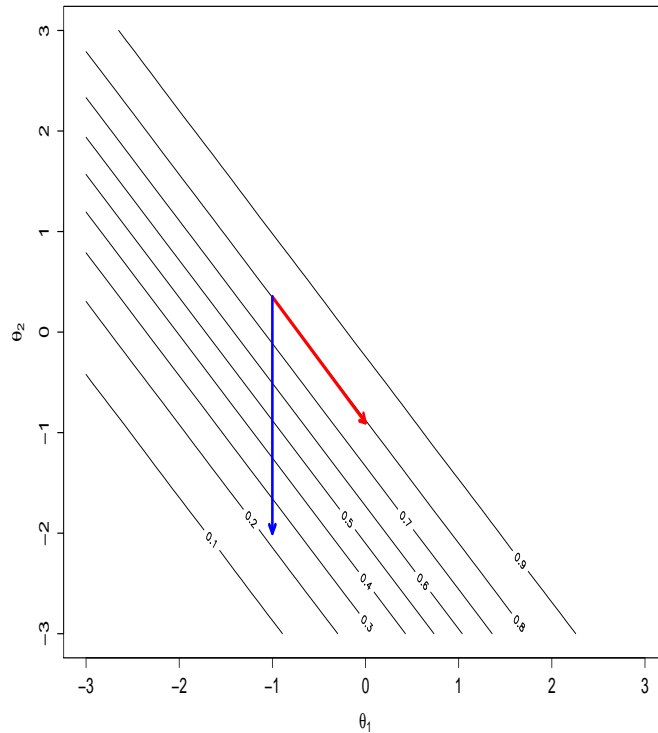
Além disso, o parâmetro b_j indica a distância entre o ponto inicial da escala de θ e o ponto onde ocorre a maior inclinação da CCI. Dessa forma, um item que possui parâmetro b_j positivo é um item no qual a maior inclinação da CCI acontece a direita do valor inicial da escala das habilidades, e, por sua vez, um item que possui parâmetro b_j negativo é um item onde a maior inclinação da CCI acontece a esquerda do valor inicial da escala de θ conforme mostra a Figura 5:

Figura 5 –Gráfico com Três Curvas Características do Item com discriminação $a_j = 1,7$, acerto casual $c_j = 0$ e diferentes dificuldades



Na TRIM estas interpretações para os parâmetros de discriminação e dificuldade podem ser generalizadas, mas há alguns fatores de complexidade a serem considerados. A inclinação de uma superfície é dependente da direção do movimento ao longo da superfície de modo que o ponto de inclinação mais acentuado depende do sentido que está sendo considerado. Tal fato é visualizado claramente na Figura 6.

Figura 6 –Gráfico de contornos e setas para superfície do item B da Figura 4 **Parâmetros do Item B:** $a_1 = 0,9$, $a_2 = 0,7$, $c = 0$, $d = 1,4$, $A = 1,14$, $B = -1,23$



Na Figura 6, observa-se que deslocamentos na SRI na direção do conjunto de pontos do espaço apontados pela seta vermelha produzirão a mesma probabilidade de resposta correta para este item. Já deslocamentos na SRI, na direção do conjunto de pontos da seta azul, produzem uma mudança de aproximadamente 0,6 na probabilidade de acerto do item.

A discriminação de um item indica o quanto a mudança no vetor de habilidades implica em mudanças nas probabilidades de acerto. Em itens

com alta discriminação a SRI pode ser dividida em 3 partes, uma região com alta probabilidade de acerto, uma região com baixa probabilidade de acerto e uma estreita região de transição entre as duas regiões citadas anteriormente. A SRI D da Figura 4, possui estas características, por isso o item D possui maior discriminação que os demais.

O poder de discriminação para itens multidimensionais \mathbf{A}_j é dado por:

$$\mathbf{A}_j = \sqrt{\sum_{k=1}^K a_{jk}^2}. \quad (11)$$

O parâmetro d_j obtido através do produto $-\mathbf{a}_{jk}b_j$ está relacionado à dificuldade do item, entretanto não pode ser interpretado diretamente como a dificuldade do item. O parâmetro de dificuldade multidimensional que possui tal interpretação é o \mathbf{B}_j , que é obtido pela seguinte expressão:

$$\mathbf{B}_j = \frac{-d_j}{\mathbf{A}_j}. \quad (12)$$

O parâmetro \mathbf{B}_j indica a distância da origem dos eixos θ_i até o ponto de maior inclinação, na direção de maior inclinação (RECKASE, 2009).

Reckase, M. D. e McKinley, R. L.(1991) apresentaram um gráfico muito utilizado para representar itens bidimensionais, denominado Gráfico de Vetor, que resume várias informações do item.

O comprimento do vetor é dado por \mathbf{A}_j , conforme a Equação 11. A base do vetor é o ponto de maior inclinação na direção de maior inclinação e a distância da base da seta até a origem é igual ao parâmetro \mathbf{B}_j conforme Equação 12. A direção do vetor é também a direção de maior inclinação da SRI e o ângulo que o vetor forma com cada um dos eixos θ_{jk} é dado por

α_{jk} (ACKERMAN, 2009):

$$\alpha_{jk} = \text{arco-cosseno} \left(\frac{\mathbf{a}_{jk}}{\mathbf{A}_j} \right). \quad (13)$$

Para itens bidimensionais, os ângulos α_{j1} e α_{j2} também podem ser obtidos por:

$$\alpha_{j1} = \text{arco-tangente} \left(\frac{a_{j2}}{a_{j1}} \right)$$

e

$$\alpha_{j2} = \text{arco-tangente} \left(\frac{a_{j1}}{a_{j2}} \right).$$

Um exemplo com os parâmetros para dois itens é dado na Tabela 1:

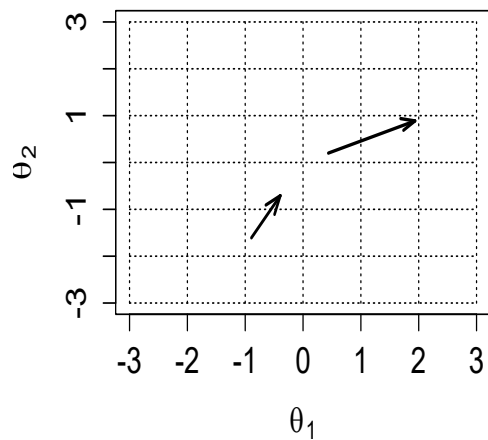
Tabela 1 –Parâmetros de dois itens multidimensionais.

Item	a_{j1}	a_{j2}	d_j	A_j	B_j	α_1	α_2
1	1,5	0,7	-0,8	1,65	0,48	25,01°	64,98°
2	0,5	0,9	1,9	1,03	-1,84	60,95°	29,05°

O gráfico de vetores dos dois itens é dada pela Figura 7. Observa-se que a distância entre a origem dos eixos e a base do vetor é dada pelo parâmetro \mathbf{B}_j . O tamanho da seta é dado pelo parâmetro \mathbf{A}_j e a direção da seta, ou seja, o ângulo entre a seta e os eixo θ_1 é dado pelo arco tangente entre (a_1/a_2) . Já o ângulo entre a seta e o eixo θ_2 é dado pelo arco tangente de (a_2/a_1) . As setas pertencentes ao primeiro quadrante identificam itens difíceis, as setas pertencentes ao terceiro quadrante, indicam itens fáceis e somente encontraremos setas no segundo e quarto quadrantes em itens com

um dos parâmetros a_{jk} negativos (RECKASE, 2009).

Figura 7 –Gráfico de setas para dois itens

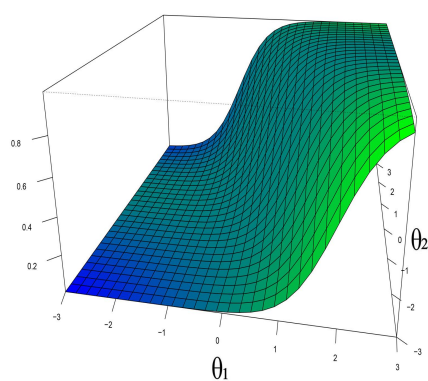


O ponto de maior inclinação, na direção de maior inclinação é obtido quando o vetor que representa o item cruza com o contorno de probabilidade $0,5(1+c_j)$ (BRYANT D.; DAVIS L, 2011). Quanto mais à direita da origem for a posição deste ponto, mais difícil o item será e quanto mais à esquerda, mais fácil. Mais detalhes sobre os parâmetros \mathbf{A}_j e \mathbf{B}_j e sua dedução formal podem ser encontrados em Reckase, M. D (1985) e Reckase, M. D.; McKinley, R. L (1991).

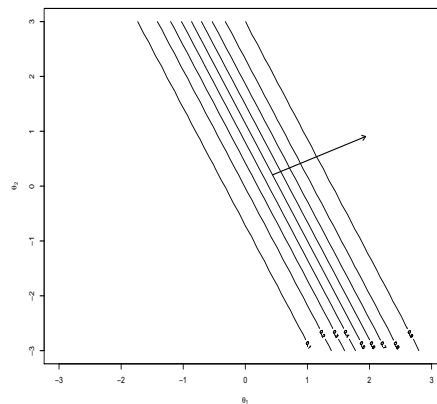
A seguir serão apresentadas as SRI e os gráficos de contorno com os vetores dos dois itens do exemplo na Tabela 1. Observa-se que a direção do vetor é a direção de maior inclinação da SRI, e o ponto de encontro do vetor, com o contorno de probabilidade 0,5 (em ambos exemplos $c_j = 0$) é

o ponto de maior inclinação, na direção de maior inclinação da SRI.

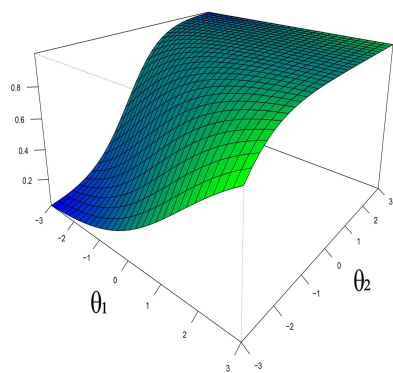
Figura 8 –SRI e gráficos de contorno com o vetor dos itens 1 e 2 da Tabela 1



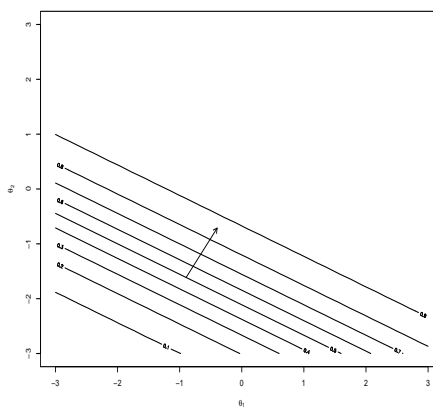
(a) SRI item 1



(b) Contornos com seta item 1



(c) SRI item 2



(d) Contornos com seta item 2

2.6 Suposições dos Modelos de TRIM

Lord(1980) e Rackase (2009), falam da suposição chamada Independência Local (ou Condicional). Esta suposição diz que as respostas de um examinado a diferentes itens em um teste são estatisticamente independentes. Em outras palavras, as respostas de um examinado a um item não afetam, de forma positiva ou negativa, a sua resposta a qualquer outro item do teste. O examinado não aprende fazendo a prova, o vetor de traços latentes inicial do examinado permanece inalterado, até a resposta do último item do teste. Quando a Independência Local existe, a probabilidade de resposta do teste, para qualquer examinado, pode ser escrita como o produto da probabilidade de resposta deste examinado em cada item.

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \dots, Y_{iJ} = y_{iJ} | \boldsymbol{\theta}) = P(Y_{i1} = y_{i1} | \boldsymbol{\theta}) \cdot P(Y_{i2} = y_{i2} | \boldsymbol{\theta}) \cdot \dots \cdot P(Y_{iJ} = y_{iJ} | \boldsymbol{\theta})$$

Lord(1980) cita como consequência da Independência Local, a propriedade em que, a ordem da apresentação dos itens em um teste também não afeta a probabilidade de resposta do examinado. O que é óbvio pois, se a Independência Local é verdadeira, a ordem dos fatores não alterará o produto. Em termos práticos, a ordem em que os itens aparecem ao longo do teste não altera as respostas do indivíduo. Outra consequência citada pelo autor é que respostas de examinados diferentes, mas com o mesmo nível de habilidade, são independentes.

Em seu trabalho, Lord (1980) afirma que a suposição de Independência Local é equivalente à correta dimensionalidade do teste, pois para dois indivíduos com mesmos vetores $\boldsymbol{\theta}_i$ é esperado que a probabilidade de acerto de ambos indivíduos seja a mesma. Porém, tal fato só ocorre com a correta determinação da dimensionalidade do instrumento de medida. Caso

haja alguma habilidade que não esteja contida no vetor de traço latentes e um dos indivíduos tenha um nível maior desta habilidade que o outro, a probabilidade de ambos responder ao item não será a mesma e, com isso, a Independência Local não é verdadeira.

Portanto, os modelos da TRI devem conter a dimensão verdadeira do teste. Montenegro (2010) define a dimensionalidade como: “dimensionalidade é a dimensão mínima no espaço de habilidades requerida para se obter independência condicional”. Logo, é de fundamental importância a verificação da dimensionalidade de um conjunto de dados analisados por modelos da TRIM.

2.7 Função de Verossimilhança

A suposição de independência condicional informa que são independentes, as respostas de um examinado a diferentes itens de um teste e, também, as respostas de examinados diferentes a um item. Dessa forma pode-se escrever a função de verossimilhança:

$$\begin{aligned}
 L(\boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j | Y) &= P(\mathbf{Y}_{ij} = \mathbf{y}_{ij} | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) \\
 &= \prod_{j=1}^J \prod_{i=1}^N P(Y_{ij} = y_{ij} | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) \\
 &= \prod_{j=1}^J \prod_{i=1}^N (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{(1-y_{ij})}, \text{ sendo:} \quad (14)
 \end{aligned}$$

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) = \pi_{ij}.$$

2.8 Dimensionalidade de um conjunto de dados

Rackase (2009) cita 4 procedimentos para determinar a dimensão de um conjunto de dados, quais sejam:

1. DIMTEST (STOUT,1999; STOUT,2001) (do Inglês Dimensionality Test): É um procedimento estatístico não paramétrico, implementado em um programa também chamado DIMTEST. Este procedimento usa os seguintes pressupostos: Independência local, condicional e monotocidade. (Monotocidade é um pressuposto onde espera-se que indivíduos que possuam maior habilidade, em algum elemento do vetor parâmetros θ_i , possuam maior probabilidade de acerto do item).
2. DETECT (ZHANG; STOUT,1999)(do Inglês *Dimensionality Evaluation To Enumerate Contributing Traits*): É um outro procedimento, também não paramétrico, com a mesma base teórica do DIMTEST e também é implementado em um programa com o seu nome. O DETECT usa, além dos pressupostos de independência local, condicional e monotocidade, o pressuposto de independência entre os pares $\mathbf{Y}_{ij} \neq \mathbf{Y}_{iJ}$, para todo θ_i .
3. Análise Paralela: Apesar da abordagem apresentada no trabalho ter sido sugerida por Ledesma e Valero-Mora (2007), ela tem uma longa história com a análise de fatores provavelmente sugerida pela primeira vez por Horn (1965).
4. Diferença Qui-Quadrado: Este procedimento é um Teste de Razão de Verossimilhanças, onde uma estatística que possui distribuição assintoticamente Qui-quadrado é testada. Neste procedimento é usado o

teste Qui-Quadrado (χ^2) para testar, o ajuste entre modelos com K e $K + 1$ dimensões. Dessa forma é determinado o número de eixos necessários para modelar a interação entre itens e examinados presente na matriz de dados \mathbf{Y}_{ij} .

A estatística usada para o teste é:

$$\chi^2 = 2[(y_{i,j} * \log(\hat{\pi}_{i,j}) + (1 - y_{i,j}) * \log(1 - \hat{\pi}_{i,j})) - 2[(y_{i,j} * \log(\hat{\pi}'_{i,j}) + (1 - y_{i,j}) * \log(1 - \hat{\pi}'_{i,j}))].$$

Nesta expressão:

- $y_{i,j}$ é a resposta dada pelo examinado i ao item j ;
- $\hat{\pi}_{i,j}$ é a probabilidade do indivíduo i acertar o item j com os parâmetros do modelo com K dimensões testado, essa probabilidade é um resultado obtido pelo estimador de máxima verossimilhança usado;
- $\hat{\pi}'$ é a probabilidade do indivíduo i acertar o item j com o parâmetros do modelo com $K + 1$ dimensões testado. Essa probabilidade é um resultado obtido pelo estimador de máxima verossimilhança usado;
- $\hat{\pi}_{i,j}$ e $\hat{\pi}'$ são definidos conforme a função de resposta ao item FRI considerada.

A hipótese nula testada neste teste é de que dimensões adicionais não fornecem melhora no ajuste do modelo. Dessa forma, devem ser testados modelos com dimensões crescentes. Inicialmente, o ajuste do modelo de 1 dimensão versus 2 dimensões é testado, se a hipótese nula for rejeitada significa que, dimensões adicionais fornecem melhora no ajuste, então, deve

ser testado o ajuste do modelo de 2 dimensões versus 3 e por conseguinte, sucessivamente até que a hipótese nula não seja mais rejeitada e, então, o modelo de menor dimensão é aceito.

Se a verossimilhança do modelo testado for do tipo:

$L(\mathbf{Y}_{ij}|\boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j)$ o grau de liberdade usado para encontrar o quantil da distribuição Qui-quadrado é dado pelo número de examinados, mais o número de itens menos o número de dimensões $N + J - K$.

Por sua vez se a verossimilhança do modelo testado for do tipo:

$L(\mathbf{Y}_{ij}, \boldsymbol{\theta}|\mathbf{a}_j, d_j, c_j)$ o grau de liberdade usado para encontrar o quantil da distribuição Qui-quadrado será o número de itens menos o número de dimensões $J - K$

3 MÉTODOS DE ESTIMAÇÃO

A estimação dos parâmetros de itens e examinados dos modelos da TRIM pode ser feita pela inferência clássica ou bayesiana. Na inferência clássica a estimação é feita por vários métodos como: verossimilhança conjunta, verossimilhança marginal e verossimilhança condicional. Azevedo (2003) discutiu os principais métodos de estimação. A seguir será apresentada uma proposta bayesiana para a estimação dos parâmetros dos modelos de TRIM.

3.1 Versão Bayesiana da TRIM: Béguin e Glas (2001)

Béguin e Glas (2001) propuseram um procedimento Bayesiano para estimação do modelo multidimensional probito de 3 parâmetros. Este procedimento usa o esquema de dados aumentados que consiste em tratar dados

latentes, ou faltantes, como parâmetros a serem estimados. Apesar de neste método se introduzir mais parâmetros a serem estimados, as densidades condicionais ficam, em geral, mais simples de serem amostradas. O esquema de dados aumentados facilita a utilização de algoritmos Monte Carlo via Cadeias de Markov MCMC. O procedimento completo será apresentado a seguir.

A probabilidade de resposta correta a um item com o modelo multidimensional probito de 3 parâmetros é:

$$\pi_{ij} = P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) = c_j + (1 - c_j)\Phi(\eta_{ij}) \quad (15)$$

A Equação 15 informa que a probabilidade de um indivíduo acertar a um item é a probabilidade do indivíduo chutar e acertar mais a probabilidade dele não chutar e acertar. Logo, ou o indivíduo sabe e não chuta, ou não sabe e chuta. Dessa forma é introduzida uma variável latente chamada de W_{ij} que assumirá os seguintes valores:

$$W_{ij} = \begin{cases} 1, & \text{se o indivíduo } i \text{ conhece a resposta correta do item } j \\ 0, & \text{se o indivíduo } i \text{ não conhece a resposta correta do item } j \end{cases}$$

A variável $W_{ij} = w_{ij} | Y_{ij} = 1$ tem distribuição Bernoulli com probabilidade de sucesso dada por $p_{ij} = \frac{\Phi(\eta_{ij})}{\Phi(\eta_{ij}) + c_j(1 - \Phi(\eta_{ij}))}$, ou seja:

$$P(W_{ij} = w_{ij} | Y_{ij} = 1) = p_{ij}^{W_{ij}}(1 - p_{ij})^{(1 - W_{ij})}.$$

Assim, quando $W_{ij} = 1$ tem-se que:

$$P(W_{ij} = 1|Y_{ij} = 1) = \left[\frac{\Phi(\eta_{ij})}{\Phi(\eta_{ij}) + c_j(1 - \Phi(\eta_{ij}))} \right],$$

ou seja $P(W_{ij} = 1|Y_{ij} = 1) \propto \Phi(\eta_{ij})$.

Por outro lado, quando $W_{ij} = 0$, tem-se que:

$$P(W_{ij} = 0|Y_{ij} = 1) = \left[\frac{c_j(1 - \Phi(\eta_{ij}))}{\Phi(\eta_{ij}) + c_j(1 - \Phi(\eta_{ij}))} \right],$$

ou seja: $P(W_{ij} = 0|Y_{ij} = 1) \propto c_j(1 - \Phi(\eta))$.

Dessa forma, a probabilidade $P(W_{ij} = w_{ij}|Y_{ij} = y_{ij})$ pode ser reescrita da seguinte forma:

$$\begin{aligned} P(W_{ij} = 1|Y_{ij} = 1, \eta_{ij}, c_j) &\propto \Phi(\eta) \\ P(W_{ij} = 0|Y_{ij} = 1, \eta_{ij}, c_j) &\propto c_j(1 - \Phi(\eta)) \\ P(W_{ij} = 1|Y_{ij} = 0, \eta_{ij}, c_j) &= 0 \\ P(W_{ij} = 0|Y_{ij} = 0, \eta_{ij}, c_j) &= 1 \end{aligned}$$

No procedimento de dados aumentados introduz-se a variável Z_{ij} definida da seguinte forma:

$$P(z_{ij}|w_{ij}, \eta_{ij}) \propto \phi(z_{ij}; \eta_{ij}, 1)[I(z_{ij} > 0).I(w_{ij} = 1) + I(z_{ij} < 0).I(w_{ij} = 0)].$$

Onde $\phi(z_{ij}; \eta_{ij}, 1)$ representa a densidade da distribuição normal com média η_{ij} e desvio padrão 1 e $I(\cdot)$ representa uma função indicadora que assume o valor zero, se o argumento é falso e um, se é verdadeiro.

Para completar o modelo são assumidas as seguintes distribuições a

priori:

$$\begin{aligned}
 \boldsymbol{\theta}_i &\sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \\
 \boldsymbol{\mu} &\sim N_n(\boldsymbol{\mu}_n, \frac{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{K_n}) \\
 \boldsymbol{\Sigma}_{\boldsymbol{\theta}} &\sim W^{-1}(\boldsymbol{\Lambda}, v) \\
 \mathbf{a}_j &\sim N_n(\boldsymbol{\mu}_{\mathbf{a}}, \boldsymbol{\Sigma}_{\mathbf{a}}) \\
 d_j &\sim N(\mu_d, \sigma_d^2) \\
 c_j &\sim \text{Beta}(\alpha_1, \alpha_2)
 \end{aligned}$$

Assim, a posteriori conjunta de $\boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j, \boldsymbol{\mu}, \boldsymbol{\theta}_i, \mathbf{W}_{ij}, \mathbf{Z}_{ij}$ é dada por:

$$\begin{aligned}
 &P(\boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}, \mathbf{W}_{ij}, \boldsymbol{\mu}, \mathbf{Z}_{ij} | \mathbf{y}_{ij}) = P(\mathbf{Z}_{ij}, \mathbf{W}_{ij} | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) \\
 &\times P(\boldsymbol{\theta}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \times P(\boldsymbol{\mu} | \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \times P(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \times P(\mathbf{a}_j) \times P(d_j) \times P(c_j) \\
 &= \prod_{i=1}^n \prod_{j=1}^J P(Z_{ij} | \eta_{ij}, W_{ij}) \times P(W_{ij} | y_{ij}, \eta_{ij}, c_j) \times P(\boldsymbol{\theta}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \\
 &\times P(\boldsymbol{\mu} | \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \times P(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \times P(\mathbf{a}_j) \times P(d_j) \times P(c_j).
 \end{aligned}$$

Para $t = 1, 2, B, \dots, M$, onde B é o *burn in* e M é o tamanho de amostra gerada. O algoritmo proposto em Béguin e Glas (2001) simula iterativamente todas as quantidades desconhecidas da seguinte forma:

- O algoritmo é inicializado com valores iniciais convenientes;
- Amostrar $\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ através de sua condicional completa;
- Amostrar $\mathbf{Z}_{ij}, \mathbf{W}_{ij}$ através de sua condicional completa;

- Amostrar $\boldsymbol{\theta}$ através de sua condicional completa;
- Amostrar \mathbf{a}_j e d_j através de sua condicional completa;
- Amostrar \mathbf{c}_j através de sua condicional completa.

As condicionais completas necessárias serão apresentadas a seguir. Existe uma observação importante a ser feita sobre esse procedimento; são usadas restrições para obter identificabilidade da escala de medida, a restrição é feita em K (k é a dimensão do modelo) itens. Nestes itens o parâmetro $d_j = 0$ e o parâmetro $a_{jk} = 1$, se $j = k$ e $a_{jk} = 0$, se $j \neq k$, $j = 1, \dots, K$ e $k = j + 1, \dots, K$.

Obtenção das Distribuições Condicionais Completas

Propriedades de formas quadráticas (BOX; TIAO, 2011)

Sejam \mathbf{x} , \mathbf{a} e \mathbf{b} vetores de dimensão $p \times 1$ e \mathbf{A} e \mathbf{B} matrizes simétricas positivas definidas de forma que $\mathbf{A} + \mathbf{B}$ seja não singular, então é válida a igualdade:

$$\begin{aligned} (\mathbf{x} - \mathbf{a})' \mathbf{A} (\mathbf{x} - \mathbf{a}) + (\mathbf{x} - \mathbf{b})' \mathbf{B} (\mathbf{x} - \mathbf{b}) &= (\mathbf{x} - \mathbf{c})' (\mathbf{A} + \mathbf{B}) (\mathbf{x} - \mathbf{c}) \\ &\quad (\mathbf{a} - \mathbf{b})' \mathbf{C} (\mathbf{a} - \mathbf{b}), \end{aligned} \quad (16)$$

em que:

$$\begin{aligned} \mathbf{c} &= (\mathbf{A} + \mathbf{B})^{-1} (\mathbf{A}\mathbf{a} + \mathbf{B}\mathbf{b}) \\ \mathbf{C} &= \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{B} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}. \end{aligned}$$

Posterior do Coeficiente β de um Modelo de Regressão Linear Normal/Normal (BOX; TIAO, 2011)

Em um modelo linear de regressão normal tem-se que:

$$y_i = \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2).$$

O mesmo modelo escrito pode ser escrito matricialmente como:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \\ \boldsymbol{\epsilon} &\sim N_N(\mathbf{0}, \sigma^2 \mathbf{I}_N), \end{aligned}$$

em que σ^2 é conhecido e \mathbf{X} sendo uma matriz $N \times p$ cuja i -ésima linha é \mathbf{X}'_i .

Se:

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (17)$$

tem-se que a posteriori de $\boldsymbol{\beta}$ é dada por:

$$\boldsymbol{\beta} | \mathbf{y} \stackrel{dist}{=} \boldsymbol{\beta} | \hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta). \quad (18)$$

Sendo:

$$\boldsymbol{\mu}_\beta = \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + (\mathbf{X}' \frac{1}{\sigma^2} \mathbf{X})^{-1} \right)^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_0), \quad (19)$$

$$\boldsymbol{\Sigma}_\beta = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 (\mathbf{X}' \frac{1}{\sigma^2} \mathbf{X})^{-1} \right)^{-1} \boldsymbol{\Sigma}_0, \quad (20)$$

em que $\hat{\boldsymbol{\beta}}$ é o estimador de máxima verossimilhança (EMV) que é dado por:

$$\hat{\beta} = \left(\mathbf{X}' \frac{1}{\sigma^2} \mathbf{X} \right) \mathbf{X}' \frac{1}{\sigma^2} \mathbf{y}.$$

Outra forma de escrever a distribuição a posteriori (19) é:

$$\beta | \mathbf{y} = \beta | \hat{\beta} \sim N_P \left(\Psi \left[\Sigma_0^{-1} \mu_0 + \left(\mathbf{X}' \frac{1}{\sigma^2} \mathbf{X} \right) \hat{\beta} \right], \Psi \right) \quad (21)$$

ou

$$\beta | \mathbf{y} = \beta | \hat{\beta} \sim N_P \left(\Psi \left[\Sigma_0^{-1} \mu_0 + \mathbf{X}' \frac{1}{\sigma^2} \mathbf{y} \right], \Psi \right),$$

com $\Psi = \left(\Sigma_0^{-1} + \mathbf{X}' \frac{1}{\sigma^2} \mathbf{X} \right)^{-1}$.

Condicional completa de Σ_θ

$$\begin{aligned}
P(\Sigma_\theta | \dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n \phi(\boldsymbol{\theta}_i; \boldsymbol{\mu}, \Sigma_\theta) \right] \times W^{-1}(\boldsymbol{\Lambda}, v) \times \phi\left(\boldsymbol{\mu}; \boldsymbol{\mu}_n, \left(\frac{\Sigma_\theta}{K_n}\right)\right) \\
P(\Sigma_\theta | \dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n (2\pi)^{-\frac{p}{2}} |\Sigma_\theta|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\mu})'(\Sigma_\theta)^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu})} \right] \\
&\times |\boldsymbol{\Lambda}|^{\frac{v}{2}} |\Sigma_\theta|^{-\frac{(v+p+1)}{2}} e^{-\frac{1}{2}\text{tr}(\boldsymbol{\Lambda}^{-1}\Sigma_\theta^{-1})} 2^{\frac{vp}{2}} \Gamma_p\left(\frac{v}{2}\right) \\
&\times (2\pi)^{-\frac{p}{2}} \left| \frac{\Sigma_\theta}{K_n} \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)' \left(\frac{\Sigma_\theta}{K_n}\right)^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)} \\
P(\Sigma_\theta | \dots, \mathbf{y}) &\propto |\Sigma_\theta|^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\mu})'(\Sigma_\theta)^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu})} \\
&\times |\Sigma_\theta|^{-\frac{(v+p+1)}{2}} e^{-\frac{1}{2}\text{tr}(\boldsymbol{\Lambda}^{-1}\Sigma_\theta^{-1})} \\
&\times \left| \frac{\Sigma_\theta}{K_n} \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)' \left(\frac{\Sigma_\theta}{K_n}\right)^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)} \\
P(\Sigma_\theta | \dots, \mathbf{y}) &\propto |\Sigma_\theta|^{-\frac{n}{2}} \times |\Sigma_\theta|^{-\frac{1}{2}} \times |\Sigma_\theta|^{-\frac{(v+p+1)}{2}} \\
&\times e^{\left[-\frac{1}{2}\text{tr}\left[\sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\mu})(\boldsymbol{\theta}_i - \boldsymbol{\mu})' + K_n(\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)'\right](\Sigma_\theta)^{-1}\right]} \\
&\times e^{-\frac{1}{2}\text{tr}(\boldsymbol{\Lambda}^{-1}\Sigma_\theta^{-1})} \\
P(\Sigma_\theta | \dots, \mathbf{y}) &\propto |\Sigma_\theta|^{-\frac{(v+p+1+n+1)}{2}} \\
&\times e^{\left[-\frac{1}{2}\text{tr}\left[\sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\mu})(\boldsymbol{\theta}_i - \boldsymbol{\mu})' + K_n(\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)' + \boldsymbol{\Lambda}^{-1}\right](\Sigma_\theta)^{-1}\right]}.
\end{aligned}$$

Assim:

$$\Sigma_\theta | \dots, \mathbf{y} \sim W^{-1}(\mathbf{S}, n + v + 1),$$

$$\text{em que } \mathbf{S} = \left[\sum_{i=1}^n ((\boldsymbol{\theta}_i - \boldsymbol{\mu})(\boldsymbol{\theta}_i - \boldsymbol{\mu})' + K_n(\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)' + \boldsymbol{\Lambda}^{-1}) \right].$$

Condicional completa de μ

$$\begin{aligned}
 P(\mu|\dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n \phi(\boldsymbol{\theta}_i; \mu, \boldsymbol{\Sigma}_\theta) \right] \times \phi\left(\mu; \mu_n, \frac{\boldsymbol{\Sigma}_\theta}{K_n}\right) \\
 P(\mu|\dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\theta}_i - \mu)'(\boldsymbol{\Sigma}_\theta)^{-1}(\boldsymbol{\theta}_i - \mu)} \right] \\
 &\quad \times (2\pi)^{-\frac{p}{2}} \left| \frac{\boldsymbol{\Sigma}_\theta}{K_n} \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu - \mu_n)' \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} (\mu - \mu_n)} \\
 P(\mu|\dots, \mathbf{y}) &\propto e^{-\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \mu)'(\boldsymbol{\Sigma}_\theta)^{-1}(\boldsymbol{\theta}_i - \mu)} \times e^{-\frac{1}{2}(\mu - \mu_n)' \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} (\mu - \mu_n)}.
 \end{aligned}$$

Sabe-se que:

$$\begin{aligned}
 \sum_{i=1}^n (\boldsymbol{\theta}_i - \mu)'(\boldsymbol{\Sigma}_\theta)^{-1}(\boldsymbol{\theta}_i - \mu) &= \sum_{i=1}^n \boldsymbol{\theta}_i'(\boldsymbol{\Sigma}_\theta)^{-1}\boldsymbol{\theta}_i - \sum_{i=1}^n \boldsymbol{\theta}_i'(\boldsymbol{\Sigma}_\theta)^{-1}\mu + \quad (22) \\
 &\quad - \mu'(\boldsymbol{\Sigma}_\theta)^{-1} \sum_{i=1}^n \boldsymbol{\theta}_i + \sum_{i=1}^n \mu'(\boldsymbol{\Sigma}_\theta)^{-1}\mu.
 \end{aligned}$$

$$\sum_{i=1}^n \boldsymbol{\theta}_i'(\boldsymbol{\Sigma}_\theta)^{-1}\boldsymbol{\theta}_i \quad \text{é uma constante.}$$

$$\sum_{i=1}^n \boldsymbol{\theta}_i'(\boldsymbol{\Sigma}_\theta)^{-1}\mu = (\mu'(\boldsymbol{\Sigma}_\theta)^{-1} \sum_{i=1}^n \boldsymbol{\theta}_i)'$$

e

$$\sum_{i=1}^n \mu'(\boldsymbol{\Sigma}_\theta)^{-1}\mu = \mu'(n\boldsymbol{\Sigma}_\theta)^{-1}\mu.$$

Pode-ser reescrever a Equação 22 da seguinte forma:

$$\begin{aligned}
\sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\mu})' (\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\mu}) &= \text{Constante} - \sum_{i=1}^n \frac{\boldsymbol{\theta}_i'}{n} (n\boldsymbol{\Sigma}_\theta)^{-1} \boldsymbol{\mu} \\
&- \boldsymbol{\mu}' (n\boldsymbol{\Sigma}_\theta)^{-1} \sum_{i=1}^n \frac{\boldsymbol{\theta}_i}{n} + \boldsymbol{\mu}' (n\boldsymbol{\Sigma}_\theta)^{-1} \boldsymbol{\mu} \\
&= \text{Constante} - \bar{\boldsymbol{\theta}}' (n\boldsymbol{\Sigma}_\theta)^{-1} \boldsymbol{\mu} \\
&- \boldsymbol{\mu}' (n\boldsymbol{\Sigma}_\theta)^{-1} \bar{\boldsymbol{\theta}} + \boldsymbol{\mu}' (n\boldsymbol{\Sigma}_\theta)^{-1} \boldsymbol{\mu} \\
&= (\boldsymbol{\mu} - \bar{\boldsymbol{\theta}})' (n\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\theta}}).
\end{aligned}$$

Assim:

$$\begin{aligned}
P(\boldsymbol{\mu} | \dots, \mathbf{y}) &\propto e^{-\frac{1}{2} \sum_{i=1}^n (\boldsymbol{\theta}_i - \boldsymbol{\mu})' (\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\mu})} \times e^{-\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)' \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)} \\
&\propto e^{-\frac{1}{2} (\boldsymbol{\mu} - \bar{\boldsymbol{\theta}})' (n\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\theta}})} \times e^{-\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)' \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)} \\
&\propto e^{-\frac{1}{2} (\boldsymbol{\mu} - \bar{\boldsymbol{\theta}})' (n\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\theta}}) + \frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)' \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)}.
\end{aligned}$$

Dessa forma, usando o resultado (16), obtem-se:

$$\boldsymbol{\mu} | \dots, \mathbf{y} \sim N_k(\mathbf{F}\mathbf{E}, \mathbf{F}^{-1})$$

em que:

$$\mathbf{F} = \left((n\boldsymbol{\Sigma}_\theta)^{-1} + \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} \right)^{-1} \mathbf{E} = \left((n\boldsymbol{\Sigma}_\theta)^{-1} (\bar{\boldsymbol{\theta}}) + \left(\frac{\boldsymbol{\Sigma}_\theta}{K_n} \right)^{-1} (\boldsymbol{\mu}_n) \right).$$

Condicional completa de Z_{ij}, W_{ij}

$$\begin{aligned}
P(\mathbf{Z}_{ij}, \mathbf{W}_{ij} | \dots, \mathbf{y}) &\propto \prod_{j=1}^m \prod_{i=1}^n \phi(\mathbf{Z}_{ij}, \mathbf{W}_{ij}; \eta_{i,j}, 1) I(Z_{ij} > 0, W_{ij} = 1, y_{i,j} = 1) \Phi(\eta_{i,j}) + \\
&+ \phi(\mathbf{Z}_{ij}; \eta_{i,j}, 1) I(Z_{ij} \leq 0, W_{ij} = 0, y_{i,j} = 1) c_j (1 - \phi(\mathbf{Z}_{ij})) + \\
&+ \phi(\mathbf{Z}_{ij}; \eta_{i,j}, 1) I(Z_{ij} > 0, W_{ij} = 1, y_{i,j} = 0) 0 + \\
&+ \phi(\mathbf{Z}_{ij}; \eta_{i,j}, 1) I(Z_{ij} \leq 0, W_{ij} = 0, y_{i,j} = 0) 1.
\end{aligned}$$

Condicional completa de θ

Distribuição condicional completa associada a um determinado traço latente θ_i .

$$\begin{aligned}
P(\theta_i | \dots, \mathbf{y}) &\propto \left[\prod_{j=1}^m \phi(\mathbf{Z}_{ij}; \boldsymbol{\eta}_{ij}, 1) \right] \times \phi(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta) \\
P(\theta_i | \dots, \mathbf{y}) &\propto \left[\prod_{j=1}^m e^{\frac{-1}{2}(\mathbf{Z}_{ij} - \boldsymbol{\eta}_{ij})^2} \right] \times e^{\frac{-1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\mu})' (\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\mu})} \\
P(\theta_i | \dots, \mathbf{y}) &\propto \left[\prod_{j=1}^m e^{\frac{-1}{2}(\mathbf{Z}_{ij} - (\mathbf{a}_j \boldsymbol{\theta}_i - d_j))^2} \right] \times e^{\frac{-1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\mu})' (\boldsymbol{\Sigma}_\theta)^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\mu})}.
\end{aligned}$$

Aplicando a decomposição de Cholesky na matriz $\boldsymbol{\Sigma}_\theta$ tem-se que:
 $\boldsymbol{\Sigma}_\theta = \mathbf{L}\mathbf{L}^t$.

Padronizando a variável θ_i que possui média $\boldsymbol{\mu}$, variância $\boldsymbol{\Sigma}_\theta$ e desvio padrão \mathbf{L} , obtém-se:

$$(\boldsymbol{\theta}_i - \boldsymbol{\mu})(\mathbf{L}^{-1}) = \boldsymbol{\theta}_i^0 \quad (23)$$

$$\boldsymbol{\theta}_i^0 = (\boldsymbol{\theta}_i - \boldsymbol{\mu})\mathbf{L}^{-1} \quad (24)$$

$$\boldsymbol{\theta}_i = \mathbf{L}\boldsymbol{\theta}_i^0 + \boldsymbol{\mu}. \quad (25)$$

Assim, a condicional completa fica da seguinte forma:

$$P(\boldsymbol{\theta}_i | \dots, \mathbf{y}) \propto \left[\prod_{j=1}^m e^{-\frac{1}{2}(\mathbf{Z}_{ij} - \mathbf{a}_j \boldsymbol{\theta}_i - d_j)^2} \right] \times e^{-\frac{1}{2}(\boldsymbol{\theta}_i^0 - \mathbf{0})'(\mathbf{I})^{-1}(\boldsymbol{\theta}_i^0 - \mathbf{0})}.$$

A variável \mathbf{Z}_{ij} possui distribuição normal com média η e variância 1 $\mathbf{Z}_{ij} \sim N(\boldsymbol{\eta}, 1)$. Pode-se reescrever a variável \mathbf{Z}_{ij} utilizando o seguinte modelo hierárquico:

$$\mathbf{Z}_{ij} = \mathbf{a}_j \boldsymbol{\theta}_i - d_j + \boldsymbol{\epsilon}_j$$

$$\mathbf{Z}_{ij} + d_j = \mathbf{a}_j \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_j$$

$$\mathbf{Z}_{ij} + d_j = \mathbf{a}_j(\mathbf{L}\boldsymbol{\theta}_i^0 + \boldsymbol{\mu}) + \boldsymbol{\epsilon}_j$$

$$\mathbf{Z}_{ij} - \mathbf{a}_j \mathbf{L}\boldsymbol{\theta}_i^0 - \mathbf{a}_j \boldsymbol{\mu} + d_j = \boldsymbol{\epsilon}_j.$$

A condicional completa é então dada por:

$$P(\boldsymbol{\theta}_i | \dots, \mathbf{y}) \propto \left[\prod_{j=1}^m e^{-\frac{1}{2}(\mathbf{a}_j \boldsymbol{\theta}_i - d_j + \boldsymbol{\epsilon}_j - \mathbf{a}_j \boldsymbol{\theta}_i - d_j)^2} \right] \times e^{-\frac{1}{2}(\boldsymbol{\theta}_i^0 - \mathbf{0})'(\mathbf{I})^{-1}(\boldsymbol{\theta}_i^0 - \mathbf{0})}$$

$$P(\boldsymbol{\theta}_i | \dots, \mathbf{y}) \propto \left[\prod_{j=1}^m e^{-\frac{1}{2}\boldsymbol{\epsilon}_j^2} \right] \times e^{-\frac{1}{2}(\boldsymbol{\theta}_i^0 - \mathbf{0})'(\mathbf{I})^{-1}(\boldsymbol{\theta}_i^0 - \mathbf{0})}.$$

Fazendo:

$$\mathbf{H} = \mathbf{Z}_{ij} - \mathbf{a}\boldsymbol{\mu} + d_j,$$

o erro pode ser escrito da seguinte forma:

$$\begin{aligned}\epsilon'_j \epsilon_j &= (\mathbf{Z}_{ij} - \mathbf{aL}\boldsymbol{\theta}_i^0 - \mathbf{a}\boldsymbol{\mu} + d_j)' (\mathbf{Z}_{ij} - \mathbf{aL}\boldsymbol{\theta}_i^0 - \mathbf{a}\boldsymbol{\mu} + d_j) \\ \epsilon'_j \epsilon_j &= (\mathbf{H} - \mathbf{aL}\boldsymbol{\theta}_i^0)' (\mathbf{H} - \mathbf{aL}\boldsymbol{\theta}_i^0).\end{aligned}$$

Assim, fazendo $\epsilon'_j \epsilon_j =$ Soma de Quadrados do Erro (SQE), obtem-se:

$$\begin{aligned}SQE &= (\mathbf{H}'\mathbf{H} - \mathbf{H}'\mathbf{aL}\boldsymbol{\theta}_i^0 - (\mathbf{aL}\boldsymbol{\theta}_i^0)'\mathbf{H} + (\mathbf{aL}\boldsymbol{\theta}_i^0)'(\mathbf{aL}\boldsymbol{\theta}_i^0)) \\ \epsilon'_j \epsilon_j &= (\mathbf{H} - \mathbf{aL}\boldsymbol{\theta}_i^0)' (\mathbf{H} - \mathbf{aL}\boldsymbol{\theta}_i^0).\end{aligned}$$

O estimador de $\hat{\boldsymbol{\theta}}_i^0$ será obtido pelo método dos Mínimos Quadrados da seguinte forma:

$$\begin{aligned}\frac{\partial SQE}{\partial \boldsymbol{\theta}_i^0} &= -\mathbf{H}'\mathbf{aL} - (\mathbf{aL})'\mathbf{H} + (\mathbf{aL})'(\mathbf{aL}\boldsymbol{\theta}_i^0) + (\mathbf{aL}\boldsymbol{\theta}_i^0)'(\mathbf{aL}) \\ 0 &= -\mathbf{H}'\mathbf{aL} - (\mathbf{aL})'\mathbf{H} + (\mathbf{aL})'(\mathbf{aL}\hat{\boldsymbol{\theta}}_i^0) + (\mathbf{aL}\hat{\boldsymbol{\theta}}_i^0)'(\mathbf{aL}) \\ 0 &= -2\mathbf{H}'\mathbf{aL} + 2(\mathbf{aL})'(\mathbf{aL}\hat{\boldsymbol{\theta}}_i^0) \\ (\mathbf{aL})'(\mathbf{aL}\hat{\boldsymbol{\theta}}_i^0) &= \mathbf{H}'\mathbf{aL} \\ [(\mathbf{aL})'(\mathbf{aL})]^{-1}(\mathbf{aL})'(\mathbf{aL}\hat{\boldsymbol{\theta}}_i^0) &= [(\mathbf{aL})'(\mathbf{aL})]^{-1}\mathbf{H}'\mathbf{aL} \\ \hat{\boldsymbol{\theta}}_i^0 &= [(\mathbf{aL})'(\mathbf{aL})]^{-1}\mathbf{H}'\mathbf{aL}.\end{aligned}$$

Lembrando que $\boldsymbol{\theta}_i^0 \sim N(\mathbf{0}, \mathbf{I})$, pela Equação 21:

$$\boldsymbol{\theta}_i^0 | \hat{\boldsymbol{\theta}}_i^0 \sim N((\mathbf{I} + (\mathbf{aL})'(\mathbf{aL}))^{-1}(\mathbf{aL})'(\mathbf{aL})\hat{\boldsymbol{\theta}}_i^0, (\mathbf{I} + (\mathbf{aL})'(\mathbf{aL}))^{-1}).$$

Condicional completa de (b_j, \mathbf{a}_j)

Distribuição condicional completa associada ao parâmetro de discriminação e dificuldade de cada item. O procedimento para obtenção desta condicional completa é análogo ao descrito na obtenção de θ_i .

Nesta condicional será usada a seguinte notação:

$$\begin{pmatrix} b_j \\ \mathbf{a}_j \end{pmatrix} = \boldsymbol{\beta}_j.$$

Dessa forma, $\eta_{ij} = -b_j + \boldsymbol{\theta}'_i \mathbf{a}_j = (-1, \boldsymbol{\theta}'_i) \begin{pmatrix} b_j \\ \mathbf{a}_j \end{pmatrix} = \boldsymbol{\Theta}' \boldsymbol{\beta}_j$.

$$P(\boldsymbol{\beta}_j | \dots, \mathbf{y}) \propto \left[\prod_{i=1}^n \phi(\mathbf{Z}_{ij}; \boldsymbol{\eta}_{ij}, 1) \right] \times \phi(\boldsymbol{\beta}_j; \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$$

$$P(\boldsymbol{\beta}_j | \dots, \mathbf{y}) \propto \left[\prod_{i=1}^n e^{-\frac{1}{2}(\mathbf{Z}_{ij} - \boldsymbol{\eta}_{ij})^2} \right] \times e^{-\frac{1}{2}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_\beta)'(\boldsymbol{\Sigma}_\beta)^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_\beta)}.$$

Aplicando a decomposição de Cholesky na matriz $\boldsymbol{\Sigma}_\beta$, tem-se que $\boldsymbol{\Sigma}_\beta = \mathbf{L}\mathbf{L}'$ e pode-se reescrever a variável \mathbf{Z}_{ij} utilizando o seguinte modelo hierárquico:

$$\begin{aligned} Z_{ij} &= \boldsymbol{\Theta}' \boldsymbol{\beta}_j + \epsilon_i \\ \mathbf{Z}_{ij} - \boldsymbol{\Theta}' \boldsymbol{\beta}_j &= \epsilon_i \\ \epsilon_i &= \mathbf{Z}_{ij} - \boldsymbol{\Theta}' \boldsymbol{\beta}_j. \end{aligned}$$

A condicional completa é então dada por:

$$\begin{aligned}
 P(\beta_j | \dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n e^{-\frac{1}{2}(\mathbf{z}_{ij} - (\Theta' \beta_j))^2} \right] \times e^{-\frac{1}{2}(\beta_j - \mu_\beta)'(\Sigma_\beta)^{-1}(\beta_j - \mu_\beta)} \\
 P(\beta_j | \dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n e^{-\frac{1}{2}(\Theta' \beta_j + \epsilon_i - (\Theta' \beta_j))^2} \right] \times e^{-\frac{1}{2}(\beta_j - \mu_\beta)'(\Sigma_\beta)^{-1}(\beta_j - \mu_\beta)} \\
 P(\beta_j | \dots, \mathbf{y}) &\propto \left[\prod_{i=1}^n e^{-\frac{1}{2}(\epsilon_i)^2} \right] \times e^{-\frac{1}{2}(\beta_j - \mu_\beta)'(\Sigma_\beta)^{-1}(\beta_j - \mu_\beta)}.
 \end{aligned}$$

O erro pode ser escrito da seguinte forma:

$$\begin{aligned}
 \epsilon'_j \epsilon_j &= (\mathbf{Z}_{ij} - \Theta' \beta_j)' (\mathbf{Z}_{ij} - \Theta' \beta_j) \\
 \epsilon'_j \epsilon_j &= (\mathbf{Z}'_{ij} \mathbf{Z}_{ij} - \mathbf{Z}'_{ij} \Theta' \beta_j - (\Theta' \beta_j)' \mathbf{Z}_{ij} + (\Theta' \beta_j)' \Theta' \beta_j).
 \end{aligned}$$

Assim, fazendo $\epsilon'_j \epsilon_j =$ Soma de Quadrados do Erro (SQE), obtem-se:

$$SQE = (\mathbf{Z}'_{ij} \mathbf{Z}_{ij} - \mathbf{Z}'_{ij} \Theta' \beta_j - \beta'_j \Theta' \mathbf{Z}_{ij} + (\Theta' \beta_j)' \Theta' \beta_j).$$

O estimador $\hat{\beta}_j$ será obtido pelo método dos Mínimos Quadrados da seguinte forma:

$$\frac{\partial SQE}{\partial \beta_j} = -\mathbf{Z}'_{ij} \Theta' - \Theta \mathbf{Z}_{ij} + \Theta \Theta' \beta_j + \beta'_j \Theta \Theta'.$$

Fazendo :

$$\frac{\partial SQE}{\partial \beta_j} = 0,$$

obtem-se:

$$\begin{aligned} -2\Theta\mathbf{Z}_{ij} + 2\Theta\Theta'\hat{\beta}_j &= 0 \\ \Theta\Theta'\hat{\beta}_j &= \Theta\mathbf{Z}_{ij} \\ (\Theta\Theta')^{-1}(\Theta\Theta'\hat{\beta}_j) &= (\Theta\Theta')^{-1}\Theta\mathbf{Z}_{ij} \\ \hat{\beta}_j &= (\Theta\Theta')^{-1}\Theta\mathbf{Z}_{ij}. \end{aligned}$$

Lembrando que $\beta_j \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, por (21):

$$\beta_j | \hat{\beta}_j \sim N\left((\boldsymbol{\Sigma}_\beta^{-1} + \Theta'\Theta)^{-1}(\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta + \Theta'\hat{\beta}_j), (\boldsymbol{\Sigma}_\beta^{-1} + \Theta'\Theta)^{-1}\right).$$

Condicional completa de c_j

Seja $t_j = \sum_{i=1}^n \mathbf{I}(W_{ij} = 0)$ o número de indivíduos que não sabem a resposta correta do item e chutam a resposta. Dentre os indivíduos que respondem ao item j e chutam, alguns irão acertar, outros não. A probabilidade de um indivíduo não saber a resposta de um item, chutar e acertar, ou seja, a probabilidade de acerto casual é dada por: $P(Y_{ij} = 1 | W_{ij} = 0) = c_j$.

Seja o número total de indivíduos que não sabem a resposta correta do item j , chutam e acertam, dado por:

$$S_j = \sum_{i|W_{ij}}^n y_{ij}.$$

Assim, S_j tem distribuição Binomial, ou seja: $S_j \sim Bin(t_j, c_j)$.

Usando para c_i uma priori conjugada não informativa $Beta(\alpha_1, \alpha_2)$, a posteriori pode ser obtida da seguinte forma:

$$\begin{aligned}
 P(c_j | \dots, \mathbf{y}) &\propto Bin(t_j, c_j) \times Beta(\alpha_1, \alpha_2) \\
 P(c_j | \dots, \mathbf{y}) &\propto \binom{t_j}{S_j} c_j^{S_j} (1 - c_j)^{t_j - S_j} \times \frac{c_j^{\alpha_1 - 1} (1 - c_j)^{\alpha_2 - 1}}{Beta(\alpha_1, \alpha_2)} \\
 P(c_j | \dots, \mathbf{y}) &\propto c_j^{S_j} (1 - c_j)^{t_j - S_j} \times c_j^{\alpha_1 - 1} (1 - c_j)^{\alpha_2 - 1} \\
 P(c_j | \dots, \mathbf{y}) &\propto c_j^{S_j + \alpha_1 - 1} (1 - c_j)^{t_j - S_j + \alpha_2 - 1} \\
 P(c_j | \dots, \mathbf{y}) &\propto \frac{c_j^{S_j + \alpha_1 - 1} (1 - c_j)^{t_j - S_j + \alpha_2 - 1}}{Beta(S_j + \alpha_1 - 1, t_j - S_j + \alpha_2 - 1)}.
 \end{aligned}$$

Assim: $c_j | \dots, \mathbf{y} \sim Beta(S_j + \alpha_1 - 1, t_j - S_j + \alpha_2 - 1)$.

4 MATERIAL E MÉTODOS

Dados Reais

Os dados utilizados nesse trabalho referem-se à primeira fase do Exame Vestibular da Universidade Federal de Lavras (UFLA), realizado no segundo semestre do ano de 2006. O exame foi dividido em 9 provas, contendo um total de 66 itens, com 4 alternativas cada. Na prova de Língua Estrangeira, o candidato escolheu entre Inglês ou Espanhol.

Por razões operacionais, as provas de Língua Estrangeira e Filosofia não foram consideradas na análise. A prova de Língua Estrangeira foi excluída, por não ter sido respondida por todos os examinados. Além disso, alguns candidatos optaram pelas questões de Inglês e outros as pelas questões de Espanhol. Já a prova de filosofia foi também descartada da análise, por conter apenas dois itens. Assim, foram analisados 56 itens divididos em 7 provas, conforme mostra a Tabela 2:

Tabela 2 – Provas do Exame Vestibular do segundo semestre de 2006 da Universidade Federal de Lavras (UFLA)

Prova	Número de Itens	Itens
Português	10	1 a 10
Geografia	8	11 a 18
História	6	19 a 24
Biologia	8	35 a 42
Física	8	43 a 50
Matemática	8	51 a 58
Química	8	59 a 66

Este exame foi escolhido por ter sido anteriormente analisado com o modelo unidimensional por Bragion (2007, 2011) e multivariado por Veloso (2014). O vestibular foi aplicado a dez cursos, conforme a Tabela 3:

Tabela 3 – Provas do Exame Vestibular do segundo semestre de 2006 da Universidade Federal de Lavras (UFLA)

Sigla	Curso	Candidatos
AD	Administração	284
AL	Engenharia de Alimentos	229
AG	Agronomia	651
CB	Ciências Biológicas	326
CC	Ciência da Computação	235
EA	Engenharia Agrícola	154
EF	Engenharia Florestal	238
MV	Medicina Veterinária	604
QI	Química	104
ZO	Zootectina	228

Os resultados deste trabalho referem-se apenas à análise dos alunos do curso de Agronomia. Esse grupo foi escolhido por ser o grupo com maior número de examinados.

5 RESULTADOS

Escolha da dimensão do modelo

Analisou-se as respostas dos examinados ao vestibular UFLA 2006-2 conforme procedimento descrito por Béguin e Glas (2001). Foram utilizados modelos unidimensionais, bidimensionais e tridimensionais na análise. O Deviance Information Criterion (DIC) e o Akaike Information Criterion (AIC) (RAFTERY et al., 2007) de cada um dos modelos estão apresentados na Tabela 4. O modelo bidimensional apresentou menores valores de DIC e AIC por isso é o modelo que melhor se ajustou aos dados.

Tabela 4 – Comparação dos Modelos Unidimensionais, Bidimensionais e Tridimensionais conforme procedimento descrito em Béguin e Glas (2001)

Modelo	DIC	AIC
Unidimensional	113468,2	107186,9
Bidimensional	104534,3	96264,8
Tridimensional	129334,3	114405,2

Um dos objetivos deste trabalho é apresentar uma proposta de seleção de modelos (grau de multidimensionalidade) da TRI através de um procedimento análogo aos métodos do tipo *forward* usados em regressão linear múltipla (HOCKING, 1976). Isto porque suspeita-se que nem todos os itens precisam ser analisados como bidimensionais para que o modelo bidimensional seja melhor ajustado que o unidimensional. Com a seleção de modelos pretende-se identificar um modelo parcimonioso em que os itens possam ter dimensões diferentes.

Nos modelos testados utilizando o procedimento de Béguin e Glas (2001) todos os itens eram exclusivamente ou unidimensionais, ou bidimensionais, ou tridimensionais. Inicialmente pensou-se em utilizar este mesmo procedimento para encontrar o melhor modelo através do procedimento *forward*, todavia devido ao tempo necessário para sua execução, isso é inviável. Uma alternativa factível foi implementada usando o pacote *mirt* (Chalmes 2012) do **R** (R CORE TEAM, 2014).

A parte 2 deste trabalho é composta de 2 artigos, no primeiro está a descrição do procedimento *forward* para seleção de modelos multidimensionais da TRI e no segundo artigo está uma aplicação da TRI multidimensional (usando o modelo bidimensional para todos os itens) aos dados da avaliação de ingresso da UFLA no ano de 2006-2, anteriormente descrita.

6 REFERÊNCIAS

ACKERMAN, T. A.; GIERL, M. J.; WALKER, C. M. Using multidimensional item response theory to evaluate educational and psychological tests. **Educational Measurement: Issues and Practice**, v. 22, n. 3, p. 37-51, 2003.

ANDRADE D. F.; TAVARES H. R.; VALLE R.C.; **Teoria da resposta ao item: Conceitos e Aplicações**. São Paulo: Associação Brasileira de Estatística - SINAPE, 2000.

AZEVEDO, C. L. N. **Métodos de estimação na teoria de resposta ao item**. Dissertação (Mestrado em Estatística)-Universidade de São Paulo, São Paulo, 121p., 2003.

BÉGUIN, A. A.; GLAS, C. A.W. MCMC estimation and some model-fit analysis of multidimensional IRT models. **Psychometrika**, v. 66, n. 4, p. 541-561, 2001.

BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. **Statistical theories of mental test scores**, 1968.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: J. Wiley, 2011.

BRAGION, M. L. L. **Um modelo de Teoria de Resposta ao Item para os dados do Vestibular 2006-2 da UFLA** 88p. Dissertação (Mestrado)

- Universidade Federal de Lavras, Lavras, 2007.

BRAGION, M. L. L. **Análise combinada de exames vestibulares da Universidade Federal de Lavras usando a teoria de resposta ao item** 187 p. Tese (Doutorado) - Universidade Federal de Lavras, Lavras, 2011.

BRYANT, D.; DAVIS, L. Item vector plots for the multidimensional three-parameter logistic model. **Applied Psychological Measurement**, v. 35, n. 5, p. 393, 2011.

CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1-29, 2012.

CONDE, F. N.; LAROS, J. A. Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. **Avaliação Psicológica**, v. 6, n. 2, p. 205-215, 2007.

LORD, F. M. **Applications of item response theory to practical testing problems**. Routledge, 1980.

LORD, F. M.; NOVICK, M. R. **Statistical Theories of Mental Test Score**, Reading: Addison-Wesley, 1968.

MATOS, G. S. **Modelos multidimensionais da TRI com distribuições assimétricas para traços latentes** Tese (Doutorado) - Universidade de

São Paulo, São Paulo-SP, 2008.

MENDONÇA, J.D. **Análise da eficiência da estimação de parâmetros da TRI pelo software ICL** 128p. Dissertação (Mestrado) - Universidade Federal de Lavras, Lavras, 2010.

MONTENEGRO, A. M. **Multidimensional item response models**. Tese De Doutorado-Universidad Nacional de Colombia, 2010.

NOJOSA, R. T. Teoria da Resposta ao Item (TRI): modelos multidimensionais. **Estudos em Avaliação Educacional**, n. 25, p. 123-166, 2002.
RAFTERY, A. E.; NEWTON, M. A.; SATAGOPAN, J. M.; KRIVITSIY,

P. (2007) Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion). In **Bayesian Statistics 8** (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), p. 1-45. Oxford: Oxford University Press.

RECKASE, M. D. The difficulty of test items that measure more than one ability. **Applied Psychological Measurement**, v. 9, n. 4, p. 401-412, 1985.

RECKASE, M. D. **Multidimensional Item Response Theory** New York, NY: Springer, 2009.

RECKASE, M. D.; MCKINLEY, R. L. The discriminating power of items

that measure more than one dimension. **Applied Psychological Measurement**, v. 15, n. 4, p. 361-373, 1991.

VELOSO, M. V. S. **Contribuições ao estudo de modelos multivariados na teoria de resposta ao item**. Tese de Doutorado - Universidade Federal de Lavras, 2014.

WANG, W.C.; CHEN, P.H.; CHENG, Y. Y. Improving measurement precision of test batteries using multidimensional item response models. **Psychological methods**, v. 9, n. 1, p. 116, 2004.

STOUT, W. et al. Dimtest. **Computer software, The William Stout Institute for Measurement, Champaign, IL**, 1999.

STOUT, W.; FROELICH, A. G.; GAO, F. Using resampling methods to produce an improved DIMTEST procedure. In: **Essays on item response theory**. Springer New York, p. 357-375, 2001.

ZHANG, J. M., STOUT, W. The theoretical DETECT index of dimensionality and its application to approximate simple structure. **Psychometrika** 64:213-249, 1999.

HOCKING, R. R. A Biometrics invited paper. The analysis and selection of variables in linear regression. **Biometrics**, v. 32, n. 1, p. 1-49, 1976.

SEGUNDA PARTE

Nesta seção são apresentados dois artigos, o primeiro sobre um método de seleção de modelos do tipo *forward* e o segundo sobre as possibilidades de discussão pedagógica de uma prova sob a luz do modelo bidimensional de item.

ARTIGO 1- Determinação da Dimensão em Modelos de Item Para Dados dos Candidatos ao Curso de Agronomia na Avaliação de Ingresso à UFLA 2006-2

RESUMO

As atividades humanas dependem de diversos fatores, o mesmo ocorre com o conhecimento e capacidade de execução de tarefas. Para se averiguar a quantidade de habilidades necessárias para a realização de um teste de conhecimento pode-se utilizar a Teoria de Resposta ao Item (TRI). Existem diferentes modelos de TRI e a análise das dimensões interfere na escolha do modelo. Na literatura são encontradas diversas formas para determinar a dimensão do melhor modelo explicativo de dados respostas a itens, no entanto tais técnicas frequentemente consideram que todos os itens da prova tem a mesma dimensão, o que suspeita-se não ser sempre verdadeiro. Este trabalho teve como objetivo utilizar procedimentos do tipo *stepwise-forward*, que são comumente utilizados na seleção de modelos em regressão linear múltipla, para a seleção de dimensionalidade em modelos da TRI. Nos resultados deste trabalho este método foi eficiente em selecionar modelos parcimoniosos, mas levou a dificuldades de interpretação. Os modelos bidimensionais de item foram também parcimoniosos e têm maior potencial explicativo. Suspeita-se que a seleção de modelos realizadas pelo método Backward traga melhores resultados na escolha das habilidades presentes em cada item, por preservar itens unidimensionais diferentes.

Palavras chave: Dimensão do modelo de item. Forward. Seleção de Modelos. Teoria de Resposta ao Item Multidimensional.

1 INTRODUÇÃO

Existem diversos fatores causais do desempenho humano em qualquer atividade realizada. Portanto, não é natural pensar que apenas uma habilidade presente no indivíduo está determinando se o mesmo irá acertar, ou errar, um item de um teste (RECKASE, 2009).

Dessa forma, parece contraditório utilizar modelos unidimensionais da Teoria de Resposta ao Item (TRI unidimensional) para analisar dados educacionais, pois os mesmos admitem, que cada item de um teste mede apenas um único traço latente e todos os itens de um teste, medem o mesmo traço latente. Embora a unidimensionalidade seja desejável, pois torna o processo de estimação dos parâmetros menos complexo.

Frequentemente admite-se a ideia da existência de uma variável dominante, ou seja, dentre todas as várias variáveis que explicam o desempenho humano em determinada atividade, existe uma dominante, justificando assim o uso de um modelo unidimensional da TRI. No entanto, não é correto admitir a existência desse fator, sem uma prévia investigação. Nojosa (2002) estimou que em uma prova do ENEM, que utiliza a TRI unidimensional em sua análise, era mais adequado medir ao menos cinco habilidades diferentes, fugindo às pressuposições do modelo de análise.

A independência local é um pressuposto dos modelos da TRI. Este pressuposto informa que as respostas de todos os examinados a todos os itens são independentes, assim como, as respostas de um mesmo examinado a diferentes itens. Dessa forma é obtida a função verossimilhança dos modelos da TRI (LORD, 1980).

Pelo exposto anteriormente, a determinação da correta da dimensão do modelo a ser utilizado é primordial, pois apenas quando essa dimensão é

estabelecida corretamente, pode-se utilizar do pressuposto da independência local (CONDE, 2007).

Diversas formas de determinação da dimensão dos modelos de item estão presentes na literatura, como: o DIMTEST, o DETECT, a Análise Paralela, diversas versões de Testes Qui-quadrado (Reckase, 2009), além de medidas de adequação do modelo como: Deviance Information Criterion (DIC), Akaike Information Criterion (AIC) e Bayesian Information Criterion (BIC) (RAFTERY et al., 2007).

Em geral, para modelos de regressão, utiliza-se com sucesso procedimentos simples de seleção de modelos do tipo *stepwise-forward* (HOCKING, 1967). Neste trabalho procuramos nos inspirar em procedimentos deste tipo para determinar a quantidade de habilidades que, incluídas ao modelo trarão melhoria no ajuste e desta forma, estimar a dimensão do conjunto de dados. Tal procedimento tem a vantagem de poder considerar a quantidade de habilidades necessárias em um item por vez, e não da prova como um todo.

Desta forma, este trabalho tem por objetivo apresentar e testar a utilização do procedimento de seleção de modelos do tipo *forward* para estimar o número de habilidades necessárias para que cada item seja respondido corretamente. Serão também apresentadas aplicações dos métodos propostos utilizando os dados obtidos na primeira fase da Avaliação de Ingresso da Universidade Federal de Lavras (UFLA), realizado no segundo semestre do ano de 2006. Os resultados obtidos pelo método apresentado serão comparados a modelos contendo ora apenas itens unidimensionais, ora bidimensionais e ora tridimensionais.

2 MATERIAL E MÉTODOS

Em itens dicotômicos, ou seja, corrigidos como certo ou errado, a variável aleatória Y_{ij} é dicotomizada assumindo os valores 0 ou 1; o valor 0 representa uma resposta incorreta ao item e o valor 1, por sua vez, representa a resposta correta. Um modelo para itens dicotômicos muito utilizado na TRIM é o Modelo Multidimensional Probit de 3 Parâmetros no qual, a probabilidade de um examinado acertar a um item em função de sua habilidade e também, das características do item é dada por:

$$\pi_{ij} = P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j) = c_j + (1 - c_j) \Phi(\mathbf{a}_j \boldsymbol{\theta}'_i + d_j) = c_j + (1 - c_j) \Phi(\eta_{ij}), \quad (1)$$

em que:

- Y_{ij} resposta dada pelo examinado i ao item j ;
- $i = 1, \dots, N$ número de examinados respondentes ao teste;
- $j = 1, \dots, J$ número de itens do teste;
- $\boldsymbol{\theta} = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}, \dots, \theta_{iK})$ é o vetor de habilidades do i -ésimo indivíduo e K é a dimensão do item, $\boldsymbol{\theta}' \in \mathbb{R}^K$;
- $\mathbf{a}_j = (a_{j1}, a_{j2}, \dots, a_{jK})$ é o vetor relacionado à discriminação do item, $\mathbf{a}_j \in \mathbb{R}^K$;
- d_j é o parâmetro relacionado à dificuldade do item, $d_j \in \mathbb{R}$;
- c_j é o parâmetro relacionado ao acerto de indivíduos de baixa habilidade, $c_j \in \mathbb{R}$;

- $\eta_{ij} = \mathbf{a}_j(\boldsymbol{\theta}'_i + d_j)$, $\eta_{ij} \in \mathbb{R}$ e é o preditor linear, uma função adequadamente escolhida dos parâmetros $\boldsymbol{\theta}'_i$, \mathbf{a}_j e d_j ;
- Φ é uma função de distribuição normal acumulada, probito ou logito por exemplo;
- $P(Y_{ij} = 1|\boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j)$ é a probabilidade do indivíduo i com habilidade $\boldsymbol{\theta}_i$ acertar o item j , dados os parâmetros de item \mathbf{a}_j , d_j e c_j .

O poder de discriminação \mathbf{A}_j e dificuldade \mathbf{B}_j para itens multidimensionais são dados por (RECKASE, 2009):

$$\mathbf{A}_j = \sqrt{\sum_{k=1}^K a_{ik}^2} \quad (2)$$

e

$$\mathbf{B}_j = \frac{-d_j}{A_j}. \quad (3)$$

Seja $\boldsymbol{\tau}_j = (\tau_{j1}, \tau_{j2}, \dots, \tau_{jk}, \dots, \tau_{jK})^{(t)}$ um vetor contendo as seguintes razões entre os parâmetros de discriminação $\tau_{jk} = \frac{a_{jk}}{A_j}$. Seja $\boldsymbol{\Sigma}$ a matriz de covariância das habilidades estimadas para o modelo multidimensional. O índice sintético unidimensional θ_u é dado por (DÍAZ, 2010):

$$\theta_u = \frac{\boldsymbol{\tau}_j^t \boldsymbol{\theta}_i}{\sqrt{\boldsymbol{\tau}_j^t \boldsymbol{\Sigma} \boldsymbol{\tau}_j}}. \quad (4)$$

Para implementar o procedimento *forward* utilizado nesse trabalho foi utilizado o pacote *mirt* (Chalmes, 2012) do **R** (R CORE TEAM, 2014). A função *mirt.model()* do pacote *mirt* permite escolher a dimensão de cada item separadamente, facilitando a implementação dos métodos. Para evitar problemas de convergência em modelos com variadas dimensões foi utilizado o método de estimação quasi-Monte Carlo EM (Bock et al. 1988; Bock

and Aitkin 1981). Note que esta opção foi apenas de natureza prática, pois a inferência bayesiana utilizando cadeias de Markov completas seria muito demorada para fins de comparação entre os métodos de seleção de modelos. Para a identificabilidade da escala de medidas do modelo, utilizou-se a rotação varimax (BERNAARDS C.A.; JENNRICH R.I., 2005).

Método da Máxima Verossimilhança Marginal (MVM)

A função de verossimilhança condicional associada ao i -ésimo respondente, com probabilidades de sucesso π_{ij} conforme descrita em (1) é dada por:

$$L_j(\mathbf{a}_j, d_j, c_j, \boldsymbol{\theta} | \mathbf{Y}_i) = P(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{a}_j, d_j, c_j, \boldsymbol{\theta}) = \prod_{j=1}^J \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(1-y_{ij})}. \quad (5)$$

Assumindo-se uma distribuição a priori $g(\boldsymbol{\theta})$ multivariada normal, a distribuição marginal da qual se obteve as estimativas é dada por:

$$\begin{aligned} P(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{a}_j, d_j, c_j) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^J p_{ij}^{y_{ij}} (1 - p_{ij})^{(1-y_{ij})} g(\boldsymbol{\theta}) d(\boldsymbol{\theta}) \\ &= \int_{\boldsymbol{\theta}} L_j(\mathbf{a}_j, d_j, c_j, | \mathbf{Y}_i) g(\boldsymbol{\theta}) d(\boldsymbol{\theta}). \end{aligned} \quad (6)$$

A integral descrita em (6) pode ser calculada através do método da quadratura gaussiana (ABRAMOWITZ; STEGUN, 1964).

Moda a Posteriori

Para a inferência Bayesiana foi assumida uma distribuição de probabilidade *a priori* para cada parâmetro do modelo. A densidade *a posteriori* conjunta, é dada por:

$$P(c_j, \mathbf{a}_j, d_j, \boldsymbol{\theta} | \mathbf{Y}_{ij}) \propto L(\mathbf{Y}_{ij} | \mathbf{a}_j, d_j, c_j, \boldsymbol{\theta}) \times P(\mathbf{a}_j, d_j, c_j, \boldsymbol{\theta} | \boldsymbol{\zeta}, \boldsymbol{\vartheta}). \quad (7)$$

A verossimilhança $L(\cdot)$ é definida como em (5), $\boldsymbol{\zeta}$ e $\boldsymbol{\vartheta}$ representam hiperparâmetros das distribuições *a priori* dos itens e da habilidades dos examinados reespectivamente. A distribuição *a posteriori* (7) é marginalizada sobre o vetor de habilidades $\boldsymbol{\theta}$ da mesma forma que no método MVM e também sobre os hiperparâmetros das distribuições de itens. Dessa forma a *a posteriori* é reescrita da seguinte forma:

$$P(c_j, \mathbf{a}_j, d_j | \mathbf{Y}_{ij}) \propto \int \left(\left[\int_{\boldsymbol{\theta}} L(\mathbf{Y}_{ij} | \mathbf{a}_j, d_j, c_j, \boldsymbol{\theta}) P(\boldsymbol{\theta} | \boldsymbol{\vartheta}) d(\boldsymbol{\theta}) \right] P(\mathbf{a} | \boldsymbol{\zeta}_a) P(d | \boldsymbol{\zeta}_d) P(c | \boldsymbol{\zeta}_c) \right) \times g(\boldsymbol{\zeta}_a) g(\boldsymbol{\zeta}_d) g(\boldsymbol{\zeta}_c) d(\boldsymbol{\zeta}_a) d(\boldsymbol{\zeta}_b) d(\boldsymbol{\zeta}_c) \quad (8)$$

Os passos do algoritmo EM seguem de forma análoga aos realizados na estimação pelo método da MVM. O procedimento para estimação das habilidades pode ser encontrado em Embretson e Reise (2013); Thissen et al. (1995) e Warm (1989).

Neste trabalho assumiu-se as seguintes distribuições *a priori* para os

parâmetros:

$$\begin{aligned}\mathbf{a}_j &\sim \ln_k \mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\ d_j &\sim N(\mu_d, \sigma_d^2) \\ c_j &\sim N(\mu_c, \sigma_c^2) \\ \boldsymbol{\theta} &\sim Nk(\mu_\theta, \boldsymbol{\Sigma}_\theta).\end{aligned}$$

Assumiu-se que os hiperparâmetros das distribuições a priori de \mathbf{a}_j , d_j e c_j eram: $\mu_{\mathbf{a}} = \mathbf{0}$, $\boldsymbol{\Sigma}_a = 0,5 \times \mathbf{I}_k$, $\mu_d = 0$, $\sigma_d = 3$, $\mu_c = 0$, $\sigma_c = 100$. Embora a priori para o parâmetro c_j admita massa para valores negativos ou maiores que 1, isto apenas indica que o procedimento de estimação aprenderá com os dados mais rapidamente. Os hiperparâmetros das distribuição *a priori* para $\boldsymbol{\theta}$ foram $\mu_\theta = \mathbf{0}$ e $\boldsymbol{\Sigma}_\theta = \mathbf{I}_k$.

A priori log-normal para o parâmetro \mathbf{a}_j foi escolhida para garantir que apenas valores positivos deste parâmetro sejam estimados garantindo assim, que os modelos apresentados neste trabalho preservem a monotonicidade das superfícies de item.

Métodos de seleção de modelos

Por analogia a modelos de regressão (HOCKING, 1976), o método *forward* pode ser utilizado para determinar a dimensão K de cada item do teste, ou seja, a dimensão dos vetores $\boldsymbol{\theta}'$ e \mathbf{a}_j . Este método consiste em uma sequência de testes, que possuem como objetivo verificar se o acréscimo de uma dimensão, a algum item do teste, fornece melhoria significativa ao ajuste do modelo.

As dimensões são acrescentadas em apenas um item do teste por vez.

A cada acréscimo calcula-se algum critério de qualidade do ajuste do modelo. Neste trabalho, foi utilizado o Deviance Information Criterion (DIC) para a seleção de modelos. Por este critério, o modelo que apresenta menor valor de DIC é escolhido como que melhor se ajusta aos dados (RAFTERY et al., 2007).

Proposta para determinar a dimensão de um conjunto de dados usando o método *forward*

A inclusão de cada dimensão sucessiva ($K = 2, 3$, etc) a todos os itens do teste é realizada em $1, 2, \dots, j, j \leq J$ passos. Em cada passo existirá um modelo tomado como inicial, chamado de Modelo Base (\mathbf{M}_0), ao qual será acrescentada uma dimensão, em apenas um dos seus $J - (j - 1)$ itens, que ainda possuem dimensão $k - 1$, dessa forma serão testados $J - (j - 1)$ modelos. Posteriormente, será testado se o acréscimo de uma dimensão, a somente um dos $J - (j - 1)$ itens do Modelo Base, trouxe melhoria ao ajuste do modelo aos dados.

O processo se inicia com o passo 1, com a inclusão da dimensão 2, neste passo são realizados J testes. O preditor linear do modelo base e dos modelos testados são da seguinte forma:

Preditor linear do Modelo Base:

$$\eta_{ij} = a_{1j}\theta_{1i} + d_j$$

Preditor linear dos demais modelos a serem comparados com o Mo-

delo Base no passo 1:

$$\begin{aligned}\eta_{ij} &= a_{1j}\theta_{1i} + a_{2j}\theta_{1i} + d_j, \\ a_{2j} &\neq 0, \text{ para apenas um item } j, j = 1, \dots, J.\end{aligned}$$

No passo 2 são realizados $J - 1$ testes, o Modelo Base a ser utilizado é o modelo escolhido no passo 1. Os modelos do passo 2, cujos ajustes serão testados, incluem outro item bidimensional além do item j^* . Assim, os preditores lineares do Modelo Base e dos demais modelos, cujo ajuste serão testados são da seguinte forma:

Preditor Linear do Modelo Base do passo 2:

$$\begin{aligned}\eta_{ij} &= a_{1j}\theta_{1i} + a_{2j}\theta_{1i} + d_j, \\ a_{2j} &\neq 0, \text{ apenas para o item } j^* \text{ escolhido no passo 1.}\end{aligned}$$

Preditor Linear dos demais modelos testados no passo 2:

$$\begin{aligned}\eta_{ij} &= a_{1j}\theta_{1i} + a_{2j}\theta_{1i} + d_j, \\ a_{2j} &\neq 0, \text{ para o item } j^* \text{ e para apenas um item} \\ & j \neq j^*, j = 1, \dots, J.\end{aligned}$$

No passo 3, o Modelo Base passa a ser o modelo obtido no passo 2. O processo é então realizado novamente até que a inclusão de uma dimensão, a qualquer item do Modelo Base do passo j não traga melhorias ao seu ajuste.

A inclusão da dimensão 2 tem até J passos. Caso o processo não encerre ainda na dimensão 2, será então testado se a inclusão da dimensão 3, a algum item do teste, fornece melhoria ao ajuste dos dados e assim

sucessivamente.

Dados utilizados

Os dados utilizados para ilustrar as duas propostas de seleção de modelos referem-se à primeira fase da avaliação de Ingresso da Universidade Federal de Lavras (UFLA), para o processo seletivo do segundo semestre de 2006. As provas analisadas estão disponíveis no seguinte endereço:

<www.dips.ufla.br/copese/provasAnteriores.asp?tipo=P1>. O exame foi dividido em 9 provas, contendo um total de 66 itens, com 4 alternativas cada. Na prova de Língua Estrangeira, o candidato escolheu entre Inglês ou Espanhol.

As provas de Língua Estrangeira e Filosofia não foram consideradas nesta análise, sendo analisados 56 itens divididos em 7 provas, conforme mostra a Tabela 1:

Tabela 1 – Provas do Exame Vestibular do segundo semestre de 2006 da Universidade Federal de Lavras (UFLA)

Prova	Número de Itens	Itens
Português	10	1 a 10
Geografia	8	11 a 18
História	6	19 a 24
Biologia	8	35 a 42
Física	8	43 a 50
Matemática	8	51 a 58
Química	8	59 a 66

Para este conjunto de dados procedeu-se à seleção de modelos com a proposta *forward* anteriormente descrita. O procedimento *forward* encerrou-se ainda na dimensão dois, sendo escolhido um modelo com 20 itens bidi-

mensionais e os demais unidimensionais. A seguir serão apresentados os índices DIC do modelo escolhido e também os índices dos modelos com todos os itens bidimensionais e tridimensionais.

3 RESULTADOS

O procedimento *forward* foi realizado até o passo 21, ainda na inclusão da segunda dimensão. Neste passo, a adição de nenhum item trouxe melhoria ao ajuste quando comparado ao seu Modelo Base. A Tabela 2 informa o item que foi incluído em cada passo e também, a dimensão de cada item no modelo final escolhido.

Tabela 2 – Resultados dos testes para a inclusão da segunda dimensão a cada item do teste

Passo	Itens Bidimensionais		Escolhido
		Prévio(s)	
1	Nenhum		50
2	50		20
3	50 e 20		38
4	50, 20 e 38		18
5	50, 20, 38 e 18		46
6	50, 20, 38, 18 e 46		37
7	50, 20, 38, 18, 46 e 37		54
8	50, 20, 38, 18, 46, 37 e 54		55
9	50, 20, 38, 18, 46, 37, 54 e 55		16
10	50, 20, 38, 18, 46, 37, 54, 55 e 16		1
11	50, 20, 38, 18, 46, 37, 54, 55, 16 e 1		40
12	50, 20, 38, 18, 46, 37, 54, 55, 16, 1 e 40		22
13	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40 e 22		12
14	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22 e 12		43
15	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12 e 43		39
16	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12, 43 e 39		17
17	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12, 43, 39 e 17		24
18	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12, 43, 39, 17 e 24		36
19	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12, 43, 39, 17, 24 e 36		11
20	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12, 43, 39, 17, 24, 36 e 11		49
21	50, 20, 38, 18, 46, 27, 54, 55, 16, 1, 40, 22, 12, 43, 39, 17, 24, 36, 11 e 49	Nenhum	

O valor encontrado do índice DIC do modelo obtido pelo método Forward e pelos modelos com todos os itens unidimensionais, bidimensionais e tridimensionais encontra-se na Tabela 3:

Tabela 3 –Critério DIC para comparação de modelos

Modelo	DIC
Proposta Forward	46.847,29
Modelo Unidimensional	50.638,49
Modelo Bidimensional	47.226,47
Modelo Tridimensional	47.717,99

O menor valor de DIC corresponde ao modelo obtido com o método Forward, portanto, este é considerado o modelo que melhor se ajusta aos dados. O maior valor de DIC encontrado foi obtido com o modelo unidimensional, sendo este portanto, o modelo com pior ajuste aos dados. O modelo bidimensional tem DIC menor que o tridimensional e deve ser preferido a este por ser mais parcimonioso.

A seguir serão apresentados os parâmetros estimados (moda *a posteriori*) do modelo selecionado pelo método Forward.

Observa-se que a prova de Português, Geografia, História, Biologia, Física, Matemática e Química contaram com 1, 5, 3, 5, 4, 2 e 0 itens bidimensionais respectivamente. Os itens bidimensionais tiveram, na maioria dos itens, valores parecidos entre a_{1j} e a_{2j} .

Através destes resultados poderia se concluir que as provas de Química e Português necessitam da mesma habilidade para serem respondidas corretamente. Todavia este resultado contradiz os resultados encontrados em Veloso (2014), que analisou o mesmo conjunto de dados usando um modelo unidimensional multivariado e concluiu que a prova media duas habilidades. Além disso, Veloso (2014) verifica a existência de dois grupos de provas relacionados a cada uma das habilidades, sendo um grupo formado pelas provas de português, geografia e história e o outro grupo constituído pelas provas de biologia, matemática, física e química. O item 58, que

apresentou discriminação baixíssima foi analisado por especialistas e não foi encontrado nenhum problema em sua formulação.

Dessa forma, apesar do modelo obtido pelo método *forward* ter tido melhor ajuste aos dados que os demais modelos, sua interpretação contradiz outros estudos. Logo, o modelo com todos os itens bidimensionais pode ter resultados mais interpretáveis pedagogicamente.

Tabela 4 – Estimativas dos parâmetros (moda *a posteriori*) do modelo selecionado pelo método *forward*

Item	a_1	a_2	d	c	A	B
Prova de Português						
1	1,00	0,88	0,28	0,69	1,33	-0,21
2	0,77	-	-0,80	0,44	0,77	1,04
3	0,80	-	-1,07	0,32	0,80	1,33
4	0,73	-	-0,35	0,40	0,73	0,47
5	0,68	-	-0,48	0,41	0,68	0,72
6	0,72	-	-2,32	0,25	0,72	3,23
7	0,63	-	-0,94	0,26	0,63	1,50
8	0,73	-	-2,87	0,13	0,73	3,92
9	1,06	-	0,65	0,51	1,06	-0,61
10	0,69	-	-0,05	0,27	0,69	0,07
Prova de Geografia						
11	0,86	0,93	-0,08	0,54	1,27	0,06
12	0,73	0,70	0,18	0,48	1,01	-0,18
13	0,55	-	0,01	0,37	0,55	-0,02
14	0,58	-	-0,55	0,34	0,58	0,94
15	0,71	-	-1,38	0,65	0,71	1,94
16	0,65	0,72	-0,97	0,25	0,97	1,00
17	0,75	0,67	0,38	0,40	1,01	-0,38
18	0,66	0,60	0,66	0,60	0,89	-0,74
Prova de História						
19	0,64	-	0,48	0,50	0,64	-0,75
20	0,66	0,80	-1,87	0,47	1,04	1,80
21	0,79	-	-2,84	0,22	0,79	3,58
22	0,78	0,78	-3,48	0,22	1,11	3,15
23	0,17	-	-0,98	0,19	0,17	5,81
24	0,61	0,58	0,38	0,23	0,84	-0,45
Prova de Biologia						
35	0,73	-	-1,10	0,27	0,73	1,50
36	1,68	0,81	0,62	0,29	1,86	-0,34
37	1,42	0,73	0,62	0,27	1,59	-0,39
38	1,53	0,69	0,74	0,10	1,68	-0,44
39	0,98	0,68	0,01	0,22	1,19	-0,01
40	1,44	0,61	0,88	0,20	1,57	-0,56
41	0,20	-	-0,74	0,25	0,20	3,69
42	2,23	-	1,00	0,27	2,23	-0,45
Prova de Física						
43	0,81	0,88	-0,67	0,16	1,20	0,56
44	0,91	-	-2,36	0,15	0,91	2,60
45	0,85	-	-0,76	0,13	0,85	0,89
46	0,74	0,63	-0,09	0,24	0,98	0,10
47	0,70	-	-2,37	0,23	0,70	3,37
48	0,94	-	0,45	0,10	0,94	-0,48
49	0,82	0,51	-0,96	0,15	0,82	1,16
50	0,75	0,74	-3,38	0,27	1,05	3,21
Prova de Matemática						
51	1,13	-	-0,53	0,10	1,13	0,47
52	0,81	-	-2,12	0,17	0,81	2,61
53	0,84	-	-1,94	0,16	0,84	2,30
54	0,71	0,73	-0,49	0,16	1,02	0,48
55	0,69	0,88	-1,84	0,30	1,12	1,64
56	1,23	-	-0,68	0,08	1,23	0,55
57	0,17	-	-0,47	0,28	0,17	2,81
58	<0,0006	-	-2,21	0,20	0,00	3456,34
Prova de Química						
59	1,83	-	0,63	0,13	1,83	-0,34
60	1,51	-	-0,05	0,23	1,51	0,03
61	0,30	-	0,20	0,29	0,30	-0,67
62	0,86	-	-0,94	0,11	0,86	1,09
63	0,65	-	-2,89	0,17	0,65	4,42
64	1,31	-	-0,06	0,28	1,31	0,05
65	1,85	-	0,85	0,19	1,85	-0,46
66	1,26	-	-0,58	0,07	1,26	0,46

A seguir serão apresentadas as correlações de Pearson entre as estimativas (moda *a posteriori*) dos parâmetros estimados para cada um dos 4 modelos, para as habilidades estimadas será também apresentada a correlação de Spearman. Os parâmetros obtidos pelo modelo *forward* serão denotados por: A_F, B_F, c_F e θ_F ; os parâmetros obtidos pelo modelo unidimensional serão denotados por: A_1, B_1, c_1 e θ_1 ; os parâmetros obtidos pelo modelo bidimensional serão denotados por: A_2, B_2, c_2 e θ_2 ; por fim, os parâmetros obtidos pelo modelo tridimensional serão denotados por: A_3, B_3, c_3 e θ_3 .

Estas correlações serão apresentadas para melhor compreensão de diferenças e semelhanças entre os quatro modelos apresentados.

Tabela 5 – Coeficientes de correlação (%) de Pearson entre as estimativas (moda *a posteriori*) dos parâmetros de discriminação multidimensional estimados para o modelo segundo a estratégia *forward* de seleção e para os modelos unidimensional, bidimensional e tridimensional

	A_1	A_2	A_3
A_F	-5,94	51,18	54,89
A_1		-8,77	-7,77
A_2			83,42

A discriminação multidimensional do modelo escolhido é mais relacionada com a dos modelos bidimensional e tridimensional. A discriminação do modelo unidimensional possui baixa correlação negativa com a discriminação de qualquer outro modelo.

Tabela 6 – Coeficientes de correlação (%) de Pearson entre as estimativas (moda *a posteriori*) dos parâmetros de dificuldade multidimensional estimados para o modelo segundo a estratégia *forward* de seleção e para os modelos unidimensional, bidimensional e tridimensional

	B_1	B_2	B_3
B_F	13,23	24,52	-1,31
B_1		2,68	-36,04
B_2			37,59

A dificuldade multidimensional do modelo escolhido apresentou baixas correlações com todas as dificuldades dos demais modelos.

Tabela 7 – Coeficientes de correlação (%) de Pearson entre as estimativas (moda *a posteriori*) dos parâmetros de acerto casual estimados para o modelo segundo a estratégia de seleção *forward* e para os modelos unidimensional, bidimensional e tridimensional

	c_1	c_2	c_3
c_F	58,94	88,84	88,93
c_1		53,59	51,41
c_2			98,71

Observa-se altas correlações entre todos os parâmetros de acerto casual c_j estimados pelos modelos. Tal fato era esperado pois este parâmetro é independente do modelo adotado, por isso as correlações para este parâmetro foram bem maiores que as demais.

Tabela 8 – Coeficientes de correlação (%) de Pearson entre as estimativas (moda *a posteriori*) dos parâmetros da habilidade sintética unidimensional para o modelo segundo a estratégia de seleção *forward* e para os modelos unidimensional, bidimensional e tridimensional

	θ_1	θ_2	θ_3	Notas
θ_F	91,28	93,68	95,91	87,00
θ_1		88,92	88,74	91,36
θ_2			78,55	89,60
θ_3				88,71

Observou-se altas correlações entre as habilidades sintéticas estimadas pelos diferentes modelos e também com relação às notas obtidas. A menor correlação foi observada entre as habilidades dos modelos bidimensional e tridimensional.

A menor correlação entre a habilidade estimada e a nota foi observada com o modelo selecionado pelo método *forward*, no entanto esta correlação ainda é alta.

Tabela 9 – Coeficientes de correlação de Spearman (%) entre as estimativas (moda *a posteriori*) dos parâmetros da habilidade sintética unidimensional para o modelo segundo a estratégia de seleção *forward* e para os modelos unidimensional, bidimensional e tridimensional

	θ_1	θ_2	θ_3	Notas
θ_F	88,14	92,16	94,73	90,85
θ_1		86,87	85,88	85,59
θ_2			97,73	90,85
θ_3				90,44

A correlação de Spearman assim como a de Pearson apresentou altos valores de correlações positivas, o que significa que a ordem das habilidades sintéticas dos modelos, assim como a ordem das notas são altamente correlacionadas.

A seguir será apresentada uma análise descritiva das notas apresentadas pelos examinados e também, uma comparação da posição de alguns examinados caso o critério utilizado fosse, ao invés da nota, a habilidade estimada pela proposta Forward e pelo modelo bidimensional. Estes modelos foram escolhidos pois apresentaram os melhores ajustes aos dados.

Tabela 10 –Estatísticas descritivas do desempenho dos examinados

	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Nota	7,00	18,00	22,00	21,96	25,00	38,00

A prova continha 56 questões, logo a média foi inferior a 50% da nota e a maior e menor nota observada foram 38 e 7, respectivamente. Observa-se que 75% dos estudantes tiveram nota inferior a 25 pontos e 25% dos estudantes tiveram nota inferior a 18.

Na Tabela 11 são apresentados os resultados de alguns examinados que receberam a nota mínima, a nota máxima, notas próximas ao 1º e 3º quartil e também, nota próxima à média da prova.

Tabela 11 –Classificação de alguns examinados considerando a nota da prova e a habilidade estimada pelo método *forward* (HF) e a habilidade sintética estimada pelo modelo bidimensional (HB)

Examinado	Nota		HF		HB	
	Valor	Ordem	Valor	Ordem	Valor	Ordem
A	7	651	-2,07	651	-2,94	651
B	18	516	-1,40	436	-2,61	599
C	22	288	-0,91	212	-2,31	385
D	25	161	-0,28	77	-1,84	143
E	38	1	0,62	2	-0,03	1

A principal tendência de interpretação que conseguimos identificar é que, para a seleção de candidatos, o erro na dimensionalidade dos modelos pode não ter maior significado, mas seguramente terá na interpretação da

qualidade dos itens.

Embora tenha se ajustado bem aos dados, o procedimento *forward* é muito ingênuo ao atribuir modelos unidimensionais, na mesma dimensão, a itens que potencialmente são unidimensionais em dimensões diferentes.

Isto levaria a dificuldades de interpretação e interpretações contraditórias, mesmo que o modelo seja o melhor ajustado. Potencialmente, procedimentos do tipo *backward* deveriam ser utilizados para corrigir este problema. Assim, permitiríamos que dimensões diferentes se manifestem como unidimensionais em modelos mais simples.

4 CONCLUSÃO

O procedimento de seleção de modelos *forward* apresentou melhor ajuste aos dados. Isto pode significar necessidade de reinterpretar modelos unidimensionais com modelos com maior número de dimensões. Para fins de seleção de candidatos, a habilidade mostrou-se bastante correlacionada à nota final, por qualquer dos modelos. No entanto, para fins de avaliação da qualidade dos itens os modelos tendem a ser muito diferentes. Pretende-se estender este tipo de avaliação para a seleção *backward*.

REFERÊNCIAS

ABRAMOWITZ, M.; STEGUN, I. A. **Handbook of mathematical functions: with formulas, graphs, and mathematical tables.** Courier Corporation, 1964.

BERNAARDS C.A.; JENNRICH R.I. Gradient Projection Algorithms and

Software for Arbitrary Rotation Criteria in Factor Analysis." **Educational and Psychological Measurement**, 65, p.676-696, 2005.

BOCK, R. D.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. **Psychometrika**, v. 46, n. 4, p. 443-459, 1981.

BOCK, R. D.; GIBBONS, R.; MURAKI, E.. Full-information item factor analysis. **Applied Psychological Measurement**, v. 12, n. 3, p. 261-280, 1988.

CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1-29, 2012.

CONDE, F. N.; LAROS, J. A. Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. **Avaliação Psicológica**, v. 6, n. 2, p. 205-215, 2007.

DÍAZ, A. M. M. **Multidimensional Item Response Theory Models where the ability has a latent linear structure** Tese (Doutorado) - Universidad Nacional de Colombia, Colômbia, 2010.

EMBRETSON, S. E.; REISE, S. P. **Item response theory**. Psychology Press, 2013.

HOCKING, R. R. A Biometrics invited paper. The analysis and selection

of variables in linear regression. **Biometrics**, v. 32, n. 1, p. 1-49, 1976.

LORD, F. M. **Applications of item response theory to practical testing problems**. Routledge, 1980.

NOJOSA, R. T. Teoria da Resposta ao Item (TRI): modelos multidimensionais. **Estudos em Avaliação Educacional**, n. 25, p. 123-166, 2002.

R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2012. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>. Acesso em 14 de ago. 2016.

RAFTERY, A. E.; NEWTON, M. A.; SATAGOPAN, J. M.; KRIVITSIY, P. (2007) Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion). In **Bayesian Statistics 8** (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), p. 1-45. Oxford: Oxford University Press.

RECKASE, M. D. **Multidimensional Item Response Theory** New York, NY: Springer, 2009.

THISSEN, D. et al. Item response theory for scores on tests including polytomous items with ordered responses. **Applied Psychological Measurement**, v. 19, n. 1, p. 39-49, 1995.

WARM, T. A. Weighted likelihood estimation of ability in item response theory. **Psychometrika**, v. 54, n. 3, p. 427-450, 1989.

ARTIGO 2- Análise Pedagógica de uma Prova Multidisciplinar Usando Modelos Bidimensionais de Item

RESUMO

Este trabalho mostra a aplicação de modelos da teoria da resposta ao item multidimensional em um processo seletivo multidisciplinar. A aplicação da metodologia permitiu identificar duas habilidades predominantes e também, classificar as provas. Esta classificação foi realizada usando o ângulo que cada vetor de item forma com o eixo da habilidade que mais explica a variabilidade do grau de dificuldade do item. Foram formados quatro grupos segundo a importância relativa de cada habilidade na resposta dos examinados para cada item. A discussão das estimativas obtidas com especialistas foi capaz de corroborar a identificação de itens com problemas de formulação, seja por problemas estruturais ou por imprecisões nas alternativas apresentadas. O artigo compara os resultados das avaliações utilizando modelos unidimensionais, univariados e multivariados.

PALAVRAS-CHAVE: TEORIA DE RESPOSTA AO ITEM MULTIDIMENSIONAL, AVALIAÇÃO EDUCACIONAL, APLICAÇÃO, VESTIBULAR MULTIDISCIPLINAR

1 INTRODUÇÃO

A proficiência sobre determinado assunto, assim como o grau de ansiedade, o grau de satisfação, bem como outros aspectos de interesse psicológico e comportamental, não podem ser medidos diretamente, sendo chamados de variáveis latentes. Para quantificar tais variáveis são toma-

das medidas indiretas em variáveis secundárias supostamente relacionadas. Assim, mede-se o conhecimento de um indivíduo indiretamente pela sua resposta a um item.

A Teoria de Resposta ao Item (TRI) propõe modelos com variáveis latentes que explicam a expectativa de acerto em itens específicos de um teste. Uma avaliação educacional reflete aspectos da realidade do ensino que podem ser elucidados com uma teoria que trate os parâmetros de itens de forma independente do grau de habilidade dos examinados. Desta forma, a TRI pode auxiliar no entendimento dos resultados observados e orientar na tomada de decisões pedagógicas. A fundamentação teórica para estes estudos é descrita na literatura como Pesquisa em Eficácia Educacional, ou EER, do inglês Educational Effectiveness Research. Uma revisão dos avanços metodológicos da área pode ser encontrada em Creemers, Kyriakides e Sammons (2010). Segundo Gustafsson (2010), a EER está atualmente em uma fase na qual o desenvolvimento de modelos teóricos objetivando explicar as relações intrínsecas entre os diferentes níveis e componentes de um sistema educacional tem um alto nível de prioridade. Este objetivo não pode ser atingido a menos que a dinâmica e a natureza variável do fenômeno educacional sejam adequadamente capturados, e isto requer ainda elucidar os mecanismos causais fundamentais dos sistemas educacionais. (GUSTAFSSON, 2010, p. 99.).

O desenvolvimento de modelos teóricos na busca por relações em diferentes níveis e componentes, tem levado a duas estratégias e abordagens distintas. Por um lado, crescem o número de estudos de modelagem multiníveis (LUYFTEN e SAMMONS, 2010), em que os dados são analisados em estruturas hierarquizadas. O exemplo mais conhecido desta abordagem

é a análise de dados de estudantes agrupados em turmas dentro de escolas de um sistema educacional. No entanto, a análise hierarquizada pode ser estendida para estruturas maiores, como regiões, cidades e países. Estes estudos frequentemente procuram evidenciar (BONAMINO e BERNADO, 2012) a contribuição da turma, do professor e da escola para o rendimento escolar (LEITÃO, 2015).

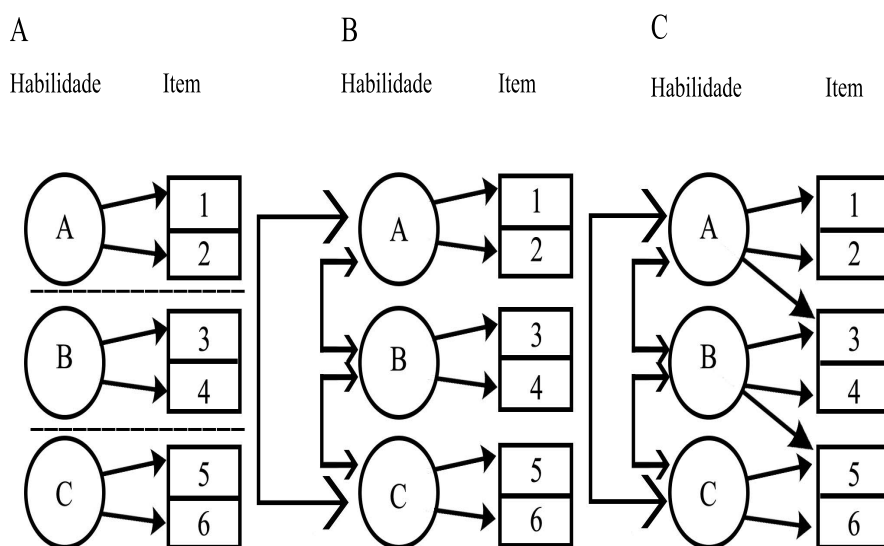
A segunda estratégia diz respeito à modelagem multidimensional da TRI (TRIM), que é relativamente recente e pouco conhecida no país, fora dos círculos teóricos de Estatística. Neste tipo de estudo procuram-se evidências da presença de mais de uma habilidade explicativa da expectativa de acerto em um item. Sabe-se que a dinâmica e a natureza variável do fenômeno educacional se devem, em parte, à complicada inter-relação entre o conjunto de habilidades necessárias ao seu desenvolvimento. Isto sugere um papel importante a ser explorado por modelos da TRIM.

Desta forma, o presente estudo foca na investigação da possibilidade de utilizar a TRIM como análise norteadora do processo da análise didático-pedagógica do item. Buscou-se avançar na interpretação de dois elementos fundamentais para a avaliação do rendimento educacional: o item de avaliação e a heterogeneidade de habilidades dos estudantes. Os fundamentos conceituais da distinção entre seus modelos multidimensionais e multivariados são que, nestes últimos, apenas uma habilidade é necessária para responder corretamente cada item, ou seja, são modelos unidimensionais para cada item. No entanto, para responder o teste admite-se o uso de mais de uma habilidade, presente em diferentes itens. Assim, pode-se dividir um teste multivariado em subtestes unidimensionais, (VELOSO, 2014; MATOS, 2008). Já nos modelos multidimensionais, existem itens que precisam

de mais de uma habilidade para ser respondidos corretamente. A maneira como tais habilidades interagem na resposta pode também ser modelada de diferentes formas. A Figura 1, adaptada de Wang (2004), mostra a diferença entre modelos multivariados (A e B) e modelos Multidimensionais (C).

A dimensão de um item é o número mínimo de habilidades que interferem de forma relevante na expectativa de acerto de um item. Na literatura são apresentadas diversas formas de determinação da dimensão de conjunto de dados como: o DIMTEST, o DETECT, a Análise Paralela (RECKASE, 2009), diversas versões de testes com base na distribuição Qui-quadrado, além de medidas de adequação do modelo como: *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) e *Deviance Information Criterion* (DIC) (RAFTERY et al., 2006).

Figura 1 –**A**: teste multivariado sem correlação entre os traços latentes; **B**: teste multivariado com correlação entre os traços latentes; **C**: teste com itens multidimensionais.



Fonte: Adaptado de Wang (2004).

No presente artigo é apresentada uma análise multidimensional para a prova da 1^a fase Vestibular da UFLA 2006-2. Este conjunto de dados foi escolhido por anteriormente ter sido analisado com um modelo unidimensional por Bragion (2011) e para outro conjunto de estudantes por um modelo multivariado por Veloso, (2014). As provas analisadas estão disponíveis no seguinte endereço:

<www.dips.ufla.br/copese/provasAnteriores.asp?tipo=P1>. Comparações entre os resultados aqui obtidos e os dos demais autores também fazem parte da discussão.

2 METODOLOGIA

Os resultados apresentados neste trabalho foram obtidos por meio do Modelo Multidimensional Probit de 3 parâmetros para itens dicotômicos mostrado a seguir.

Tomando $i = 1, \dots, N$, como o número de examinados, $j = 1, \dots, J$, como o número de itens do teste, a variável aleatória Y_{ij} representa a resposta do examinado i ao item j . Itens dicotômicos são corrigidos como certo ou errado, assim, a variável aleatória Y_{ij} assume apenas valores 0 ou 1. Portanto a probabilidade de um examinado acertar um item, condicionada à sua habilidade e às características do item (\mathbf{a}_j, d_j, c_j) é dada pelo seguinte modelo, chamado Modelo Multidimensional Probit de 3 Parâmetros:

$$P_{ij} = P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, b_j, c_j) = c_j + (1 - c_j)\Phi(\eta_{ij}),$$

$$\eta_{ij} = \sum_{k=1}^K a_{jk}\theta_{ik} + d_j,$$

- $\boldsymbol{\theta} = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}, \dots, \theta_{iK})$ é o vetor K -dimensional de habilidades do i -ésimo indivíduo, $\boldsymbol{\theta} \in \mathbb{R}^K$;
- $\mathbf{a}_j = (a_{j1}, a_{j2}, \dots, a_{jK})$ é o vetor K -dimensional relacionado à discriminação do item, $\mathbf{a}_j \in \mathbb{R}^K$;
- d_j é o parâmetro relacionado à dificuldade do item, $d_j \in \mathbb{R}$;
- c_j é o parâmetro relacionado ao acerto de indivíduos de baixa habilidade, $c_j \in \mathbb{R}$;
- $P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j)$ é a probabilidade do indivíduo i com habilidade $\boldsymbol{\theta}_i$ acertar o item j ;

- η_{ij} é o preditor linear, uma função adequadamente escolhida dos parâmetros $\mathbf{a}_j, \boldsymbol{\theta}$ e d_j , $\eta_{ij} \in \mathbb{R}$;
- Φ é a função de distribuição acumulada normal.

Para itens multidimensionais, o poder de discriminação do item (\mathbf{A}_j), a dificuldade (\mathbf{B}_j) e o ângulo γ_{jk} que mostra a posição do item j em relação ao eixo da habilidade k , são dados, respectivamente, por:

$$\mathbf{A}_j = \sqrt{\sum_{k=1}^K a_{jk}^2}, \quad (1)$$

$$\mathbf{B}_j = \frac{-d_j}{A_j} \quad (2)$$

e

$$\gamma_{jk} = \arccos\left(\frac{\mathbf{a}_{jk}}{\mathbf{A}_j}\right). \quad (3)$$

No software **R** (R CORE TEAM, 2014), pode-se encontrar diversos pacotes que fazem a estimação dos parâmetros da TRI, dentre eles, destaca-se o pacote *mirt* (CHALMERS, 2012), que foi criado para estimar os parâmetros dos modelos multidimensionais dicotômicos e politômicos. Estão implementados dois algoritmos que realizam as estimativas dos parâmetros da TRIM baseados nos princípios desenvolvidos por Bock e Aitkin (1981) e Bock, Gibbons e Muraki (1988). As estimativas de parâmetros apresentadas neste trabalho foram produzidas usando o pacote *mirt* e os resultados referem-se a modelos que preservam a monotonicidade das superfícies de item. Esta é uma propriedade segundo a qual a expectativa de se obter uma resposta correta aumenta, conforme o aumento das habilidades presentes nos indivíduos, ou seja, habilidades maiores sempre implicam em

maior expectativa de sucesso. Isto não é uma imposição matemática usual dos modelos TRIM e, para preservar tal propriedade, foram especificadas distribuições a priori próprias para os parâmetros de item, como descrito a seguir.

Para os parâmetros de discriminação \mathbf{a}_j foram utilizadas distribuições a priori log-normais com média 0 e desvio padrão 0,5; para o parâmetro relacionado à dificuldade d_j foi utilizada a distribuição a priori normal com média 0 e desvio padrão 3, para o parâmetro c_j utilizou-se como priori uma distribuição normal com média 0 e desvio padrão 100. Note que embora esta priori admita massa para parâmetro de acerto casual negativo ou maior que 1, isto apenas indica que o procedimento de estimação "aprenderá" com os dados mais rapidamente.

Os dados utilizados foram as respostas obtidas pelos 651 candidatos ao curso de Agronomia na primeira fase do Exame Vestibular da Universidade Federal de Lavras (UFLA), realizado no segundo semestre do ano de 2006. A parte utilizada do exame consiste em provas de sete (7) disciplinas, com um total de 56 itens, com quatro (4) alternativas cada, sendo que as provas de Geografia, Biologia, Física, Matemática e Química continham oito (8) itens cada, a prova de Português dez (10) itens e a prova de História seis (6) itens.

As respostas foram analisadas, utilizando um modelo bidimensional. Este modelo foi escolhido após a análise da dimensionalidade do conjunto de dados. A metodologia desta escolha distancia dos objetivos deste artigo.

O ângulo γ_{j1} , definido na Equação 3, indica a relevância da habilidade θ_1 na resposta ao item. Os itens com posicionamento semelhante em relação ao eixo θ_1 foram agrupados em quatro (4) grupos, conforme a

Tabela 1 a seguir:

Tabela 1 – Ângulos separadores de itens

Grupo	Ângulo	Descrição dos itens do grupo
1	$0^\circ \leq \gamma_{j1} < 10^\circ$	A habilidade θ_1 é extremamente importante para que a resposta do indivíduo seja correta e a habilidade θ_2 é praticamente desnecessária.
2	$10^\circ < \gamma_{j1} \leq 45^\circ$	Ambas as habilidades θ_1 e θ_2 são importantes para que a resposta do indivíduo seja correta, porém a habilidade θ_1 tem maior influência.
3	$45^\circ < \gamma_{j1} \leq 60^\circ$	Ambas as habilidades são importantes, porém a habilidade θ_2 tem maior influência.
4	$60^\circ < \gamma_{j1} \leq 90^\circ$	A habilidade θ_1 é praticamente desnecessária para a resposta correta do indivíduo e a habilidade θ_2 é extremamente importante.

Os valores 10° e 60° foram escolhidos por meio de um procedimento que procurou minimizar a variação dentro de cada setor angular e maximizar a variação entre os setores angulares. O ângulo de 45° foi utilizado para preservar a simetria dos setores (ACKERMAN et.al., 2003).

Outro instrumento de discussão utilizado foi o gráfico de vetores dos itens (RECKASE; MCKINLEY, 1991). Neste gráfico todos os itens foram apresentados na forma de vetores cujo comprimento é dado por \mathbf{A}_j , a direção do vetor é a direção de maior inclinação da Superfície de Resposta ao Item (SRI). O ponto que é a base do vetor é o ponto de maior inclinação, na direção de maior inclinação da SRI e a distância desse ponto, que é a base da seta, até a origem dos planos é igual ao parâmetro \mathbf{B}_j .

De posse dos resultados numéricos, procedeu-se à análise didático-pedagógica dos itens que mais se destacaram em termos dos parâmetros de dificuldade e de discriminação multidimensionais dos itens. Para tanto, os itens foram discutidos com especialistas das diversas áreas, procurando elucidar os porquês dos seus aspectos distintivos.

O teste Exato de Fisher foi usado para testar a hipótese de que

os agrupamentos das habilidades são independentes das disciplinas, considerando um nível de significância de 5% . Adicionalmente produziu-se a Superfície de Informação do Teste (SIT), superfície de resposta em que a quantidade de informação dos itens é calculada como função das habilidades (RECKASE, 2009).

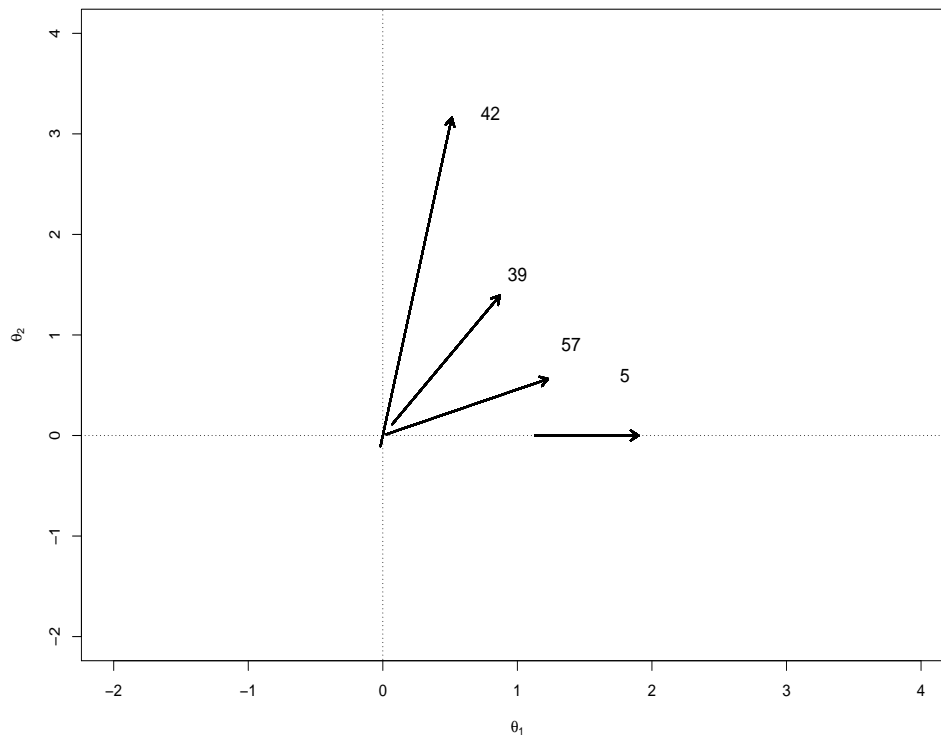
3 RESULTADOS

Na Tabela 2 são apresentadas as estimativas dos parâmetros do modelo bidimensional de item para cada um dos itens, separados por provas. Nesta tabela encontram-se também os ângulos que foram usados para o agrupamento e podem ser identificadas numericamente propriedades de itens. Um resumo sinótico destes resultados, no entanto, depende dos gráficos e tabelas auxiliares que se seguem. A Figura 2 contém o gráfico de vetores de 4 itens sendo cada item, representante de cada um dos grupos. Os itens 5, 57, 39 e 42 pertencem aos grupos 1, 2, 3 e 4 respectivamente. A partir desta figura e seus ângulos é possível estabelecer os agrupamentos da Tabela 3 e estudar suas propriedades.

Tabela 2 – Parâmetros estimados pelo modelo bidimensional.

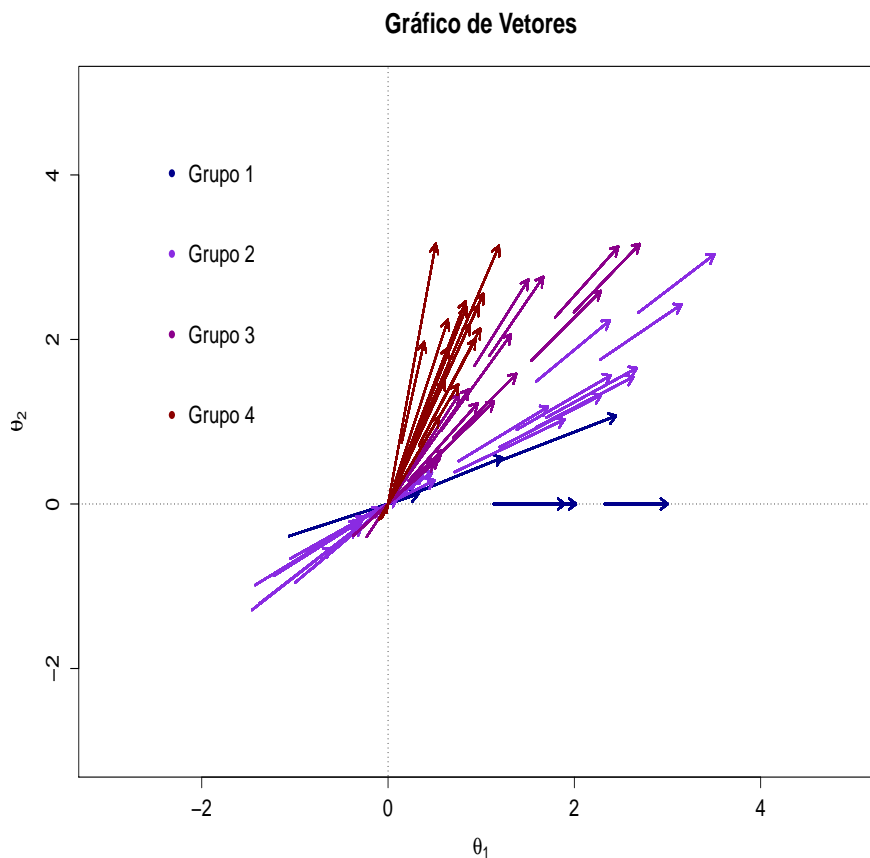
Item	a_{j1}	a_{j2}	d_j	c_j	A_j	B_j	γ_{j1}
Prova de Português							
1	1.05	0.92	2.57	0.00	1.40	-1.84	41.17
2	0.78	0.52	0.97	0.06	0.93	-1.04	33.50
3	0.68	0.78	-1.10	0.31	1.04	1.06	48.67
4	0.79	0.50	1.16	0.00	0.93	-1.24	32.56
5	0.77	0.00	-0.87	0.43	0.77	1.13	0.00
6	0.71	0.83	-3.35	0.27	1.09	3.06	49.42
7	0.90	0.52	0.48	0.04	1.04	-0.46	29.98
8	0.68	0.86	-3.17	0.14	1.10	2.89	51.44
9	0.68	0.00	-1.58	0.63	0.68	2.31	0.00
10	0.71	0.72	0.24	0.20	1.01	-0.23	45.40
Prova de Geografia							
11	0.94	0.67	1.74	0.01	1.16	-1.51	35.52
12	0.72	0.69	1.38	0.02	1.00	-1.38	44.03
13	0.97	0.65	1.22	0.17	1.17	-1.04	33.77
14	0.61	0.68	0.04	0.23	0.91	-0.04	47.80
15	0.81	0.00	-0.98	0.63	0.81	1.20	0.00
16	0.75	0.82	-0.64	0.20	1.11	0.58	47.51
17	1.12	1.08	1.77	0.00	1.56	-1.14	43.88
18	0.85	0.75	2.21	0.04	1.14	-1.94	41.32
Prova de História							
19	1.10	0.76	2.32	0.00	1.34	-1.73	34.67
20	1.09	0.63	-1.74	0.47	1.26	1.38	30.08
21	1.04	0.61	-2.23	0.21	1.21	1.84	30.48
22	0.82	0.71	-3.85	0.22	1.08	3.55	40.59
23	1.18	0.64	-1.09	0.26	1.34	0.82	28.32
Prova de Biologia							
24	0.99	0.96	1.18	0.00	1.38	-0.86	44.03
35	0.67	0.55	0.22	0.01	0.87	-0.25	39.06
36	1.25	3.29	0.57	0.16	3.52	-0.16	69.17
37	0.99	1.73	0.92	0.04	1.99	-0.46	60.23
38	0.79	2.34	-0.32	0.00	2.47	0.13	71.31
39	0.80	1.28	-0.20	0.16	1.51	0.13	58.00
40	0.68	1.68	0.34	0.01	1.82	-0.19	68.08
41	0.79	0.74	-2.36	0.39	1.08	2.19	43.05
42	0.53	3.27	0.36	0.00	3.31	-0.11	80.71
Prova de Física							
43	0.59	1.26	-1.14	0.13	1.39	0.82	64.97
44	0.75	0.85	-2.63	0.15	1.13	2.33	48.56
45	0.74	0.94	-0.43	0.05	1.20	0.36	51.78
46	0.87	0.90	0.67	0.00	1.25	-0.54	46.13
47	1.01	0.66	-1.99	0.22	1.21	1.65	33.16
48	0.64	1.24	0.29	0.00	1.39	-0.21	62.88
49	0.24	1.22	-0.95	0.00	1.24	0.77	78.65
50	0.87	0.67	-3.16	0.26	1.10	2.87	37.52
Prova de Matemática							
51	0.68	1.69	-1.42	0.11	1.82	0.78	68.09
52	0.58	0.96	-2.36	0.16	1.12	2.10	58.84
53	0.58	1.05	-2.30	0.16	1.20	1.91	61.05
54	0.55	1.06	-0.53	0.07	1.19	0.45	62.62
55	0.96	0.66	-1.07	0.25	1.16	0.92	34.39
56	0.66	1.42	-1.23	0.05	1.57	0.79	65.13
57	1.20	0.55	-0.04	0.33	1.32	0.03	24.85
58	1.28	0.56	-1.78	0.23	1.40	1.27	23.51
Prova de Química							
59	0.58	2.02	-0.48	0.00	2.11	0.23	74.04
60	0.50	1.47	-0.71	0.13	1.55	0.46	71.05
61	1.39	0.51	1.68	0.17	1.48	-1.13	20.27
62	0.75	1.17	-1.47	0.14	1.39	1.06	57.25
63	0.97	0.60	-2.28	0.16	1.14	2.00	31.65
64	0.69	1.72	-0.92	0.27	1.85	0.50	67.94
65	0.97	2.42	-0.38	0.16	2.61	0.14	68.08
66	0.56	1.59	-1.43	0.04	1.69	0.85	70.75

Figura 2 –Gráfico de vetores dos itens 5, 39, 42 e 57, representantes típicos de cada grupo formado ao analisar as habilidades como tendo 2 dimensões.



Na Figura 2, que apresenta quatro itens típicos de cada grupo, observa-se que os vetores referentes aos itens 5 e 57 são muito mais próximos do eixo θ_1 que os demais vetores. Essa proximidade indica que são itens que necessitam muito mais da habilidade θ_1 para serem respondidos corretamente que da habilidade θ_2 .

Figura 3 –Gráfico de vetores de todos os itens da prova



Na Figura 3, que contém os vetores de todos os itens da prova, observa-se que a maioria dos itens pertence aos grupos 2 e 3, ou seja, questões que estão relacionadas às duas habilidades. A maioria dos itens é considerada difícil, pois estão posicionados à direita da origem. Observa-se um grupo de itens muito difíceis, com base dos vetores muito distantes da origem e também, um grupo de itens com discriminação muito grande, ou seja, com vetores bem maiores que as demais. Estes itens foram analisados

por especialistas e os resultados são resumidamente apresentados na Tabela 4.

A distribuição de todos os itens da prova, dentro de cada grupo está apresentada na Tabela 3, a seguir.

Tabela 3 –Distribuição dos itens dentro de cada grupo.

Grupo	Itens
1	5, 9, 15, 58 e 61.
2	1, 2, 4, 7, 11, 12, 13, 17, 18, 19, 20, 21, 22, 23, 24, 35, 41, 47, 50, 55, 57 e 63.
3	3, 6, 8, 10, 14, 16, 39, 44, 45, 46, 52, 53 e 62.
4	36, 37, 38, 40, 42, 43, 48, 49, 51, 54, 56, 59, 60, 64, 65 e 66.

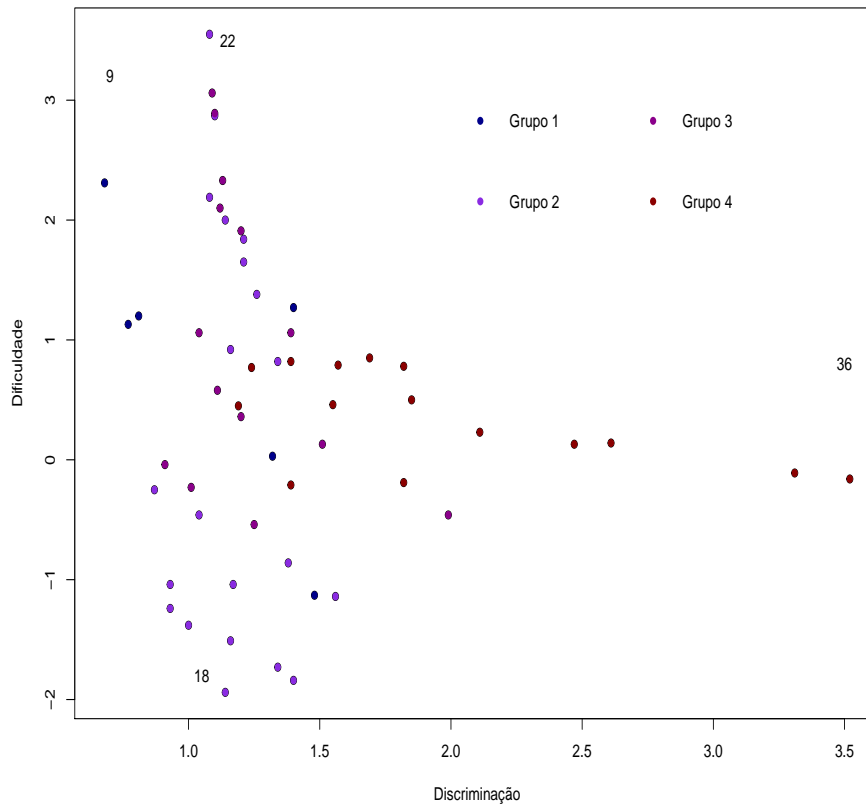
Na Tabela 4 encontra-se a análise dos 10 itens mais difíceis da prova e também de 3 itens com grande discriminação. O estudo destes itens indicou que muitos deles eram difíceis por apresentarem enunciados confusos, problemas na elaboração, além da presença de mais de uma alternativa correta, ou nenhuma alternativa correta. As características dos itens com grande discriminação foram muito diferentes, isto é não foram observados problemas na elaboração. Estes itens conseguiam diferenciar examinados com alta habilidade, de examinados com baixa habilidade .

Tabela 4 –Análise didático-pedagógica dos 10 itens mais difíceis e de 3 itens com grande discriminação.

Item	Análise dos 10 itens mais difíceis da prova.
6 e 8	Não apresentam nenhuma resposta correta dentre as alternativas, deveriam ter sido anuladas.
9	Questão relacionada à leitura de uma obra de Guimarães Rosa. Este item discrimina quem leu o livro de quem não leu, mas não quem estudou mais ou está preparado para o resto da prova
22	Apresenta duas alternativas corretas como resposta, deveria ter sido anulada
41	A estrutura da questão é difícil e complexa. Os conceitos abordados são difíceis. As alternativas são plausíveis mais difíceis.
44	Não é uma questão difícil desde que o aluno entenda que o comprimento é proporcional à massa.
50	Item bem elaborado, contudo aborda um conteúdo pouco trabalhado no ensino médio o que faz dela uma questão difícil para os examinados que não viram este conteúdo.
52	Item com enunciado complexo pede-se para identificar a alternativa incorreta. Não é um item difícil desde que o aluno saiba que a notação da letra c trata de matriz inversa.
53	Item muito difícil, em seu enunciado é introduzido um conceito novo que o aluno tem que entender para então assinalar a alternativa correta que é a mais difícil dentre todas.
63	Item com enunciado complexo pede-se para identificar a alternativa incorreta. É um item que necessita de muita atenção para ser respondido, dessa forma não deveria estar localizado no fim da prova. Não possui grande complexidade se o aluno souber estequiometria e termodinâmica, mas ainda assim exige atenção e cuidado com detalhes
Análise de 3 itens com grande discriminação da prova.	
36	É um item no qual os alunos que sabem sistemática evolutiva conseguem responder e os que não viram não conseguem.
38	É um item trivial para examinados que sabem genética de sistema A,B,0 os que não sabem não conseguem responder.
42	É um item no qual os alunos que sabem sobre a AIDS conseguem responder já os que não sabem, não conseguem responder

A Figura 4 apresenta o posicionamento dos itens considerando os parâmetros de dificuldade e discriminação multidimensionais, \mathbf{A}_j e \mathbf{B}_j .

Figura 4 –Posicionamento dos itens por grupo, com relação à dificuldade e discriminação.



Observa-se que os itens com maior dificuldade da prova discriminam pouco, já os itens com grande discriminação são fáceis. Os itens pertencentes aos grupos 4 possuem maior discriminação e os itens do grupo 1 menor discriminação. Os itens do grupo 1 podem ser considerados difíceis,

pois possuem dificuldade maior que 0, já os itens do grupo 2 mostraram-se bastante heterogêneos com relação à dificuldade. Neste grupo encontra-se o item mais fácil e também, o mais difícil da prova. Os itens do grupo 3, também variaram muito com relação à dificuldade, no entanto menos quando comparados ao grupo 2. Já os itens do grupo 4 apresentaram a menor variação com relação à dificuldade.

Verificou-se também, que o item mais fácil da prova foi o 18, o mais difícil que deveria ter sido anulado foi o 22, aquele com maior discriminação o 36 e com menor discriminação o 9. A seguir serão apresentados os gráficos e discussões relacionados aos itens 9, 18, 50 e 36, o item 50 será discutido, pois foi o item mais difícil da prova que não apresentou problemas em sua formulação.

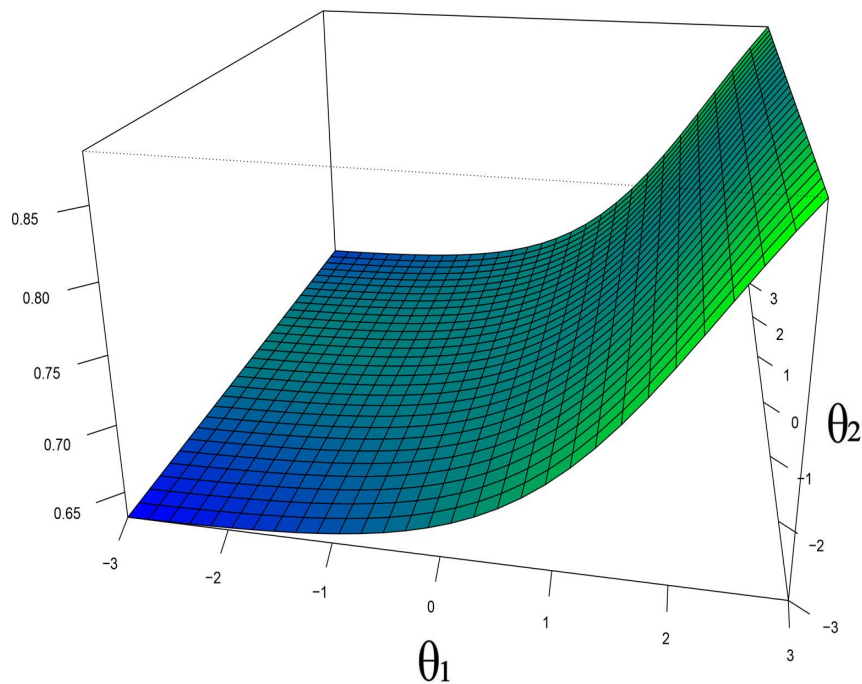
Discussão do Item 9

O item 9 apresentou menor discriminação dentre todos os itens da prova, com parâmetro $A_9 = 0,69$. Em sua SRI apresentada na Figura 5, observa-se que a mudança da região com baixa probabilidade de acerto, para a região de habilidades com alta probabilidade de acerto, ocorre suavemente. Este comportamento é muito diferente ao observado na SRI do item 36 na Figura 9, item com maior discriminação da prova.

O pequeno valor de discriminação é também observado, no tamanho do vetor correspondente a este item, no gráfico de contornos da Figura 7. O vetor correspondente ao item 9 é o menor de todos. Outro fato interessante deste item é o alto valor de acerto casual 0,63. Observou-se, na resolução da prova por especialistas, que mesmo indivíduos que não leram o texto conseguem eliminar duas alternativas, apenas observando as opções a serem

assinaladas.

Figura 5 –Superfície de Resposta do Item 9.



Discussão do Item 18

O item 18 mostrou-se o item mais fácil da prova, pois apresentou o menor parâmetro de dificuldade multidimensional, $B_{18} = -1,94$. Salienta-se que o parâmetro de dificuldade multidimensional apresentado foi negativo, significando que o vetor deste item está a esquerda da origem dos espaços, no 3º quadrante.

Figura 6 –Superfície de Resposta do Item 18.

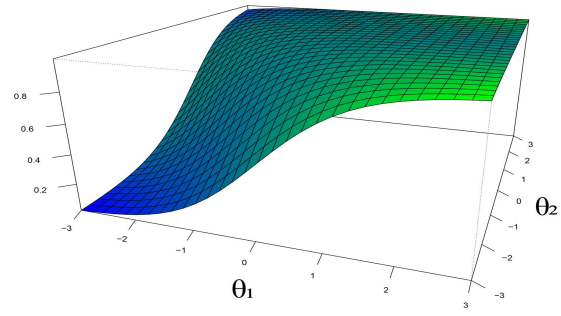
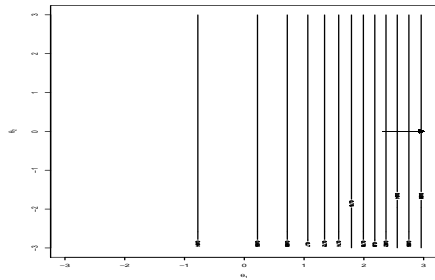
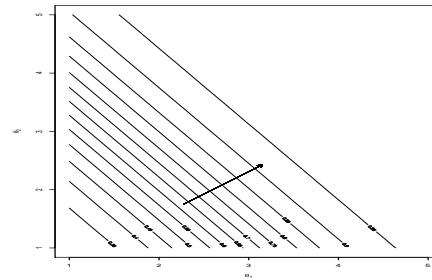


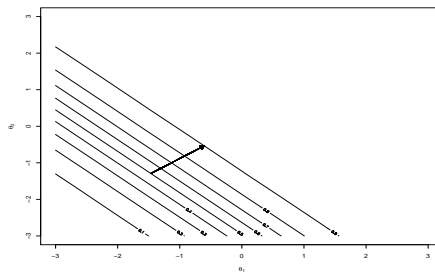
Figura 7 –Gráficos de contornos com setas dos itens 9, 18, 50 e 36.



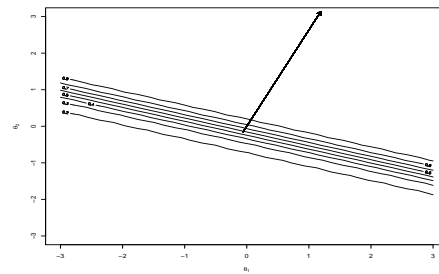
(a) Contornos com seta item 9



(b) Contornos com seta item 50



(c) Contornos com seta item 18



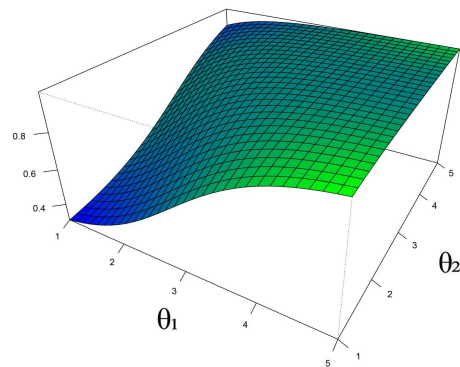
(d) Contornos com seta item 36

A dificuldade de um item multidimensional é a sua distância da origem até o ponto de maior inclinação, na direção de maior inclinação da SRI (RECKASE, 2009). Tal ponto ocorre no cruzamento do vetor direcional com o contorno equiprovável $0,5 + \frac{\epsilon}{2}$ (BRYANT D.; DAVIS L., 2011). Analisando a Figura 7 (a, b, c, e d), observa-se que o ponto onde ocorre o cruzamento é exatamente a base do vetor, e o ponto de encontro, ou seja, a base da seta que está mais à esquerda da origem é a do item 18, mostrando que este é o item que necessita de menores valores de habilidade θ_1 e θ_2 para atingir o ponto de maior inclinação, na direção de maior inclinação das SRI, por isso ele é considerado o item mais fácil da prova. Tal fato é também observado na SRI deste item (Figura 6), a região de transição de baixa probabilidade de acerto, para alta probabilidade de acerto ocorre para pequenos valores de ambas as habilidades.

Discussão do Item 50

O item 50 mostrou-se o mais difícil da prova dentre os que não deveriam ter sido anulados, com parâmetros $B_{50} = 2,87$. Observando a base dos vetores da Figura 7, verifica-se que o item que contém a base mais à direita é o item 50, por este motivo este é o item mais difícil da prova, ou seja, é o item que necessita de maiores valores de habilidade θ_1 e θ_2 para atingir o ponto de maior inclinação, na direção de maior inclinação da SRI. Na Figura 8, observa-se que a região de transição de baixa probabilidade de acerto, para alta probabilidade de acerto ocorre para altos valores de ambas as habilidades.

Figura 8 –Superfície de Resposta do Item 50.

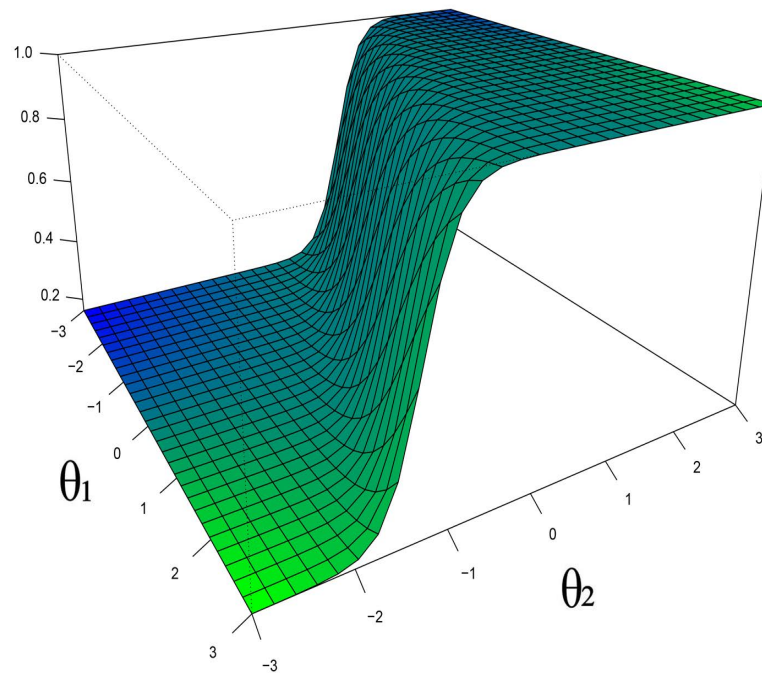


Discussão do Item 36

O item 36 foi o item que apresentou maior valor de discriminação na prova, com parâmetro $A_{(36)} = 3,52$. Na Figura 9 observa-se que a SRI do item 36 pode ser dividida em 3 partes, uma região com alta probabilidade de acerto, uma região com baixa probabilidade de acerto e uma estreita região de transição entre as duas regiões citadas anteriormente.

O tamanho do vetor correspondente a este item é também o maior de todos (Figura 7-d). A proximidade dos contornos equiprováveis e o tamanho do vetor indicam que, neste item, pequenas mudanças nos valores das habilidades θ_1 e θ_2 resultam em grandes mudanças na probabilidade de acerto.

Figura 9 –Superfície de Resposta do Item 36.



Na Tabela 5 encontra-se a distribuição dos grupos dentro de cada prova. Os itens das provas de Português, Geografia e História possuem 75% de seus itens pertencentes aos grupos 1 e 2. Já as provas de Biologia, Física, Matemática e Química possuem aproximadamente 72% de seus itens pertencentes aos grupos 3 e 4. Tal fato indica que a habilidade 1 é uma habilidade relacionada à leitura, interpretação de textos e a habilidade 2 é uma habilidade mais relacionada à conhecimentos específicos. A prova em estudo apresentou diversas questões, que para serem respondidas ne-

cessitavam da prévia leitura de um texto. Isso pode explicar a quantidade de itens pertencentes aos grupos 2 e 3, ou seja, questões que necessitam conjuntamente das habilidades θ_1 e θ_2 .

Para melhor interpretação dos resultados obtidos na Tabela 5 foi realizado o Teste Exato de Fisher. Os itens dos grupos 1 e 2, que estão mais relacionados com a habilidade θ_1 , foram agrupados em um só grupo e os itens dos grupos 3 e 4, mais relacionados com a habilidade θ_2 , foram agrupados em outro grupo. As provas de Português, Geografia e História, que apresentaram itens com semelhante distribuição com relação aos 4 grupos, foram agrupadas em um tipo denominado Ciências Humanas (CH), as provas de Biologia, Física, Matemática e Química foram agrupadas em um outro tipo, denominado Ciências Exatas e Biologia (CE).

Tabela 5 –Distribuição dos itens da prova por grupos.

Grupos	Ciências Humanas(CH)			Ciências Exatas e Biologia(CE)			
	Português	Geografia	História	Biologia	Física	Matemática	Química
1	2	1	0	0	0	0	0
2	4	5	6	2	2	3	2
3	4	2	0	1	3	1	1
4	0	0	0	5	3	4	5

As hipóteses testadas pelo Teste Exato de Fisher são de independência entre tipos e grupos. O teste determina se a proporção dos Grupos (1 e 2) e (3 e 4) diferem significativamente nos tipos CH ou CE. Rejeitar tal hipótese é de extrema importância metodológica, pois permite associar significados aos traços latentes medidos pelas habilidades.

Como o valor-p da amostra observada foi de aproximadamente 0,001, rejeita-se a hipótese de independência, ou seja, os Grupos (1 e 2) e (3 e 4) interagem significativamente com os tipos CH ou CE. As provas CH estão mais ligadas à habilidade θ_1 e as provas CE estão mais ligadas à habilidade

θ_2 .

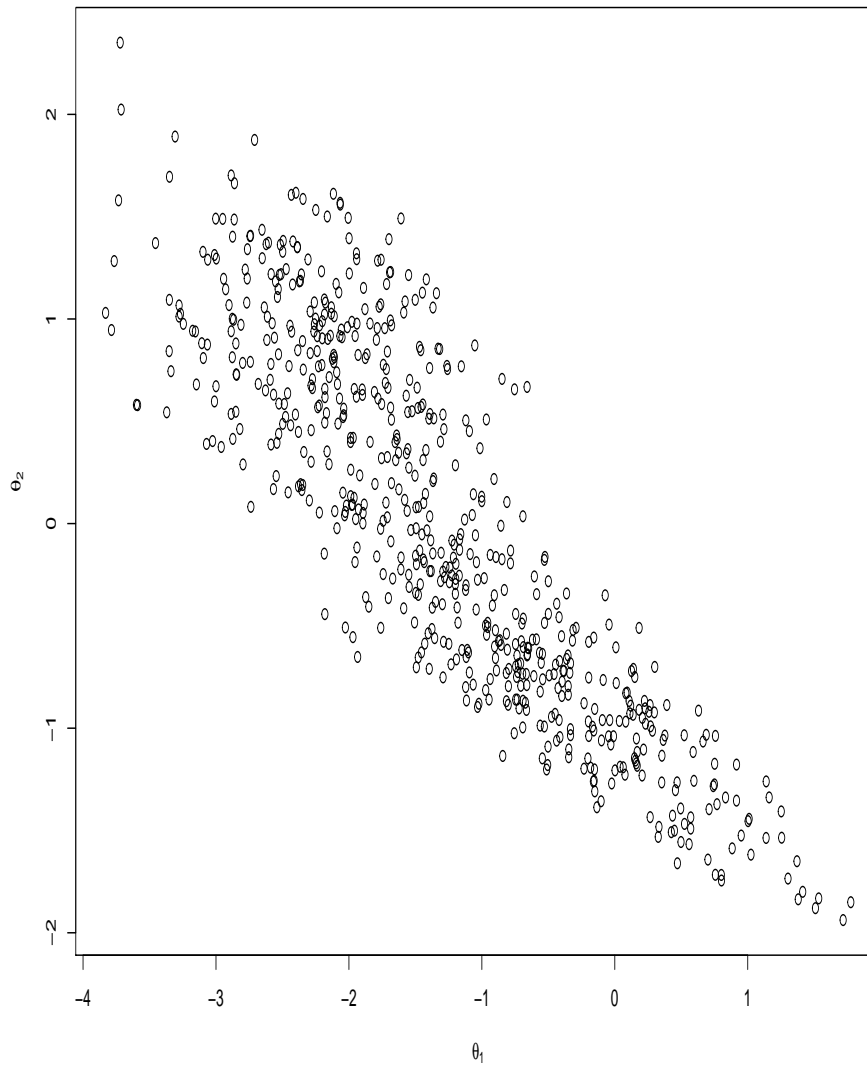
A Tabela 6 a seguir, apresenta o desempenho dos alunos na prova. Nenhum examinado errou ou acertou todas as questões. A média de acerto foi inferior à 50% da nota e 75% dos examinados obtiveram nota igual ou inferior a 25 pontos.

Tabela 6 –Estatísticas descritivas das notas dos examinados

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
7,00	18,00	22,00	21,96	25,0	38,0

Na Figura 10 é apresentado o gráfico de dispersão das habilidades θ_1 e θ_2 estimadas para todos os 651 candidatos. A correlação encontrada entre as duas habilidades foi aproximadamente $\rho = -0,86$. A correlação é alta e negativa. Tal fato justifica o modelo com todos os itens bidimensionais ter sido escolhido inicialmente na seleção do modelo.

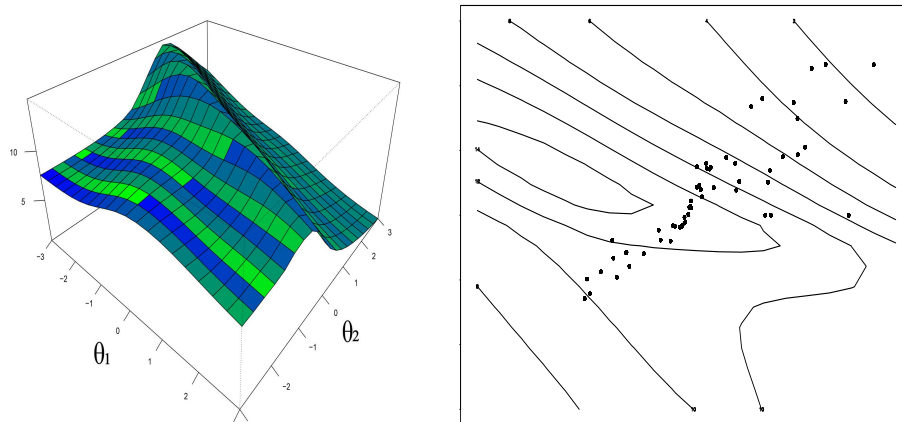
Figura 10 –Dispersão das habilidades..



Os itens foram em sua grande maioria difíceis. Isso pode ser vi-

sualizado no gráfico de vetores dos itens apresentados na Figura 3, que pertenciam em maioria, ao 1º quadrante e também, no gráfico de contornos da Superfície de Informação do Teste (SIT) apresentado na Figura 11. Observa-se que todos os itens ficaram fora da região de maior informação (região 14), o que indica que o vestibular foi, no geral, difícil para o grupo de examinados.

Figura 11 –Gráfico de SIT e Contorno da SIT com posicionamento dos itens.



(a) Superfície de Informação do Teste

(b) Contornos da Superfície de Informação do Teste

Comparação com Resultados Anteriores

Comparações entre os resultados obtidos com os modelos multidimensional e unidimensional.

Bragion (2011) analisou o mesmo conjunto de dados usando o Modelo Logístico Unidimensional de 3 parâmetros. Em suas análises os itens mais difíceis apresentaram maior discriminação e os mais fáceis menor dis-

criminação. Nos resultados obtidos neste trabalho, observou-se na Figura 4, que os itens mais difíceis possuem discriminação mais baixa e os com maior discriminação são fáceis. Observou-se um considerável número de itens com acerto casual acima de 0,25. Neste trabalho observou-se o mesmo fato, uma vez que 11 itens apresentaram estimativa pontual maior que 0,25.

Em seu trabalho, Bragion (2011) analisou separadamente os itens 8, 11 e 74 (corresponde ao item 66 deste trabalho). No item 8, o modelo bidimensional indicou um item muito difícil, que levou à análise de especialistas que concluíram não haver uma alternativa correta para o item. O item 11 foi considerado um item ruim por possuir baixa discriminação e dificuldade. Na análise usando o modelo multidimensional, tal item foi considerado fácil, mas com razoável discriminação e não se identificou problemas com o item. Já o item 74 (corresponde ao item 66 deste trabalho), última questão da prova, foi considerado um item bom por possuir bom poder de discriminação e dificuldade em ambos trabalhos.

Os resultados com modelos unidimensionais mostraram que as provas mais difíceis e com maior discriminação foram as provas de Inglês, Biologia, Física, Química e Matemática. A análise com o modelo multidimensional mostrou que as provas de Biologia, Física, Química e Matemática foram mais relacionadas com a habilidade θ_2 e esta habilidade está relacionada com a discriminação do item. A prova de Inglês não foi analisada neste trabalho.

As provas de Português, Geografia e História foram mais relacionadas com a habilidade θ_1 . Contudo, a maioria dos itens dessa prova pertencem ao grupo 2, ou seja, há uma mistura das habilidades. Isso faz com que ocorra uma grande variação na dificuldade dessas questões. Dessa forma

não se pode afirmar que essas provas foram mais fáceis ou mais difíceis que as provas de Biologia, Física, Matemática e Química

Comparações entre os resultados obtidos com os modelos multidimensional e multivariado

Veloso (2014) percebeu que as habilidades dos examinados estimadas são fortemente associadas positivamente. A associação acontece entre as provas de humanas (Português, Geografia, História, Filosofia) e também entre as provas nas áreas de exatas e biológicas (Biologia, Física, Matemática e Química). A associação observada foi mais forte entre as habilidades das áreas de exatas e biológicas do que entre as habilidades em humanas e sociais. Entre as áreas, habilidades tiveram associações próximas de zero. O modelo bidimensional de itens permite chegar à mesma divisão de provas. Entretanto, a correlação entre as habilidades estimadas foi altamente negativa.

4 CONCLUSÃO

A TRIM utilizando modelos bidimensionais de itens permitiu explorar melhor as características particulares presentes nos itens e identificar potenciais problemas didático-pedagógicos decorrentes de sua elaboração e, ou, estrutura. Neste estudo, a análise foi corroborada pela experiência didática de especialistas, o que deu significado aos modelos ajustados. Quanto às distribuições das habilidades entre os candidatos, o modelo permitiu identificar heterogeneidade entre as diferentes habilidades, bem como, indicar com maior precisão deficiências na formação dos estudantes. Assim sendo,

recomenda-se como atividade rotineira a análise exploratória de itens com modelos semelhantes ao aqui utilizado, pois permite interpretar com maior detalhe exames multidisciplinares do tipo vestibular e propor explicações didático-pedagógicas razoáveis junto a especialistas da educação.

REFERÊNCIAS

ACKERMAN, T. A.; GIERL, M. J.; WALKER, C. M. Using multidimensional item response theory to evaluate educational and psychological tests. **Educational Measurement: Issues and Practice**, v. 22, n. 3, p. 37-51, 2003.

BOCK, R. D.; AITKIN, M.. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. **Psychometrika**, v. 46, n. 4, p. 443-459, 1981.

BOCK, R. D.; GIBBONS, R.; MURAKI, E. Full-information item factor analysis. **Applied Psychological Measurement**, v. 12, n. 3, p. 261-280, 1988.

BONAMINO, A. E ; BERNADO, E. S. **A organização de turmas de alunos nos anos iniciais do Ensino Fundamental e desempenho em leitura.** In.: Avaliação em Larga Escala: Questões Polêmicas. WERLE, F.O.C. (Org.) Brasília. Liber Livro Ed. pp. 75 - 96, 2012.

BRAGION, M. L. L. **Análise combinada de exames vestibulares da Universidade Federal de Lavras usando a teoria de resposta ao**

item. Tese de Doutorado - Universidade Federal de Lavras, Lavras, 2011.

BRYANT, D.; DAVIS, L. Item vector plots for the multidimensional three-parameter logistic model. **Applied Psychological Measurement**, v. 35, n. 5, p. 393, 2011.

CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1-29, 2012.

CREEMERS, B. P. M; KYRIAKIDES, L.; SAMMONS, P. **Methodological advances in educational effectiveness research**. Routledge, 2010.

Raftery, A. E., Newton, M. A., Satagopan, J. M. and Krivitsiy, P. (2007) Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion). In **Bayesian Statistics 8** (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), pp. 1-45. Oxford: Oxford University Press.

GUSTAFSSON, J.E. Longitudinal designs. In B CREEMERS, B.; KYRIAKIDES, L.; SAMMONS, P. (Eds.), **Methodological advances in educational effectiveness research**. London, UK: Routledge Taylor Francis. pp. 77-101, 2010.

LEITÃO, U. A., Social index of educational effectiveness: A new approach from the perspective of promoting equity. **Education Policy Analysis**

Archives, 23(50), 2015. Disponível em:

<<http://dx.doi.org/10.14507/epaa.v23.1915>> Acesso em 30 de jun. 2016.

LUYFTEN, H.; SAMMONS, P. **Multilevel Modelling**. In B CREE-MERS, B.; KYRIAKIDES, L.; SAMMONS, P. (Eds.), *Methodological advances in educational effectiveness research*. London, UK: Routledge Taylor Francis. pp. 246-276, 2010

MATOS, G. M. **Modelos multidimensionais da TRI com distribuições assimétricas para os traços latentes**. Tese de Doutorado - Universidade de São Paulo, São Paulo, 2008.

R DEVELOPMENT CORE TEAM **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2012. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.

RECKASE, M. D. **Multidimensional Item Response Theory** New York, NY: Springer, 2009.

VELOSO, M. V. S. **Contribuições ao estudo de modelos multivariados na teoria de resposta ao item**. Tese de Doutorado - Universidade Federal de Lavras, Lavras, 2014.

WANG, W.C.; CHEN, P. H.; CHENG, Y. Y. Improving measurement precision of test batteries using multidimensional item response models. **Psychological methods**, v. 9, n. 1, p. 116, 2004.

CONSIDERAÇÕES FINAIS

Esta tese procurou discutir aspectos relevantes do ponto de vista estatístico de aplicações de modelos multidimensionais de item a uma avaliação de admissão na Universidade Federal de Lavras.

Os resultados obtidos pelo procedimento de Béguin e Glas (2001) mostraram-se consistentes na estimação dos parâmetros dos modelos da TRI, porém nossa implementação em R deste processo é extremamente demorada.

As estimativas produzidas pelo pacote *mirt* são rápidas, no entanto observou-se que os modelos de itens são muito sensíveis a escolhas das distribuições *a priori*, resultando em grandes diferenças marcantes nos parâmetros estimados. Com o procedimento *forward* pode-se encontrar um modelo que melhor se ajusta aos dados entretanto, este resultado nem sempre é interpretável pedagogicamente, uma proposta de seleção *backward* talvez produza modelos mais interpretáveis, ficando como sugestão para trabalhos futuros.

O modelo bidimensional mostrou ser muito útil, pois foi capaz de apontar itens com problemas na prova, identificando em quais provas cada uma das habilidades era mais utilizada. Pretende-se estender esse tipo de análise para avaliações nacionais do tipo ENEM.

