

ã 15.225 ã 35.326 ã 25.427 âmetro a8.143 âmetro b8.244
âmetro c8.345 âmetro a9.456 âmetro b9.557 âmetro c9.658
âmetro δ_2 9.759 âmetro λ_2 9.860 ário
contextual completo PROEB 2009B.174



Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Estatística

**NOVAS ESTRUTURAS MULTIDIMENSIONAIS
DA TRI PARA CONSIDERAÇÃO DO EFEITO
DA ANSIEDADE NA PROFICIÊNCIA DO
ALUNO EM AVALIAÇÕES DE LARGA
ESCALA**

SARAH MARTINS SALOMÃO BRODBECK

200855021

Juiz de Fora

2011

SARAH MARTINS SALOMÃO BRODBECK
200855021

**NOVAS ESTRUTURAS MULTIDIMENSIONAIS
DA TRI PARA CONSIDERAÇÃO DO EFEITO
DA ANSIEDADE NA PROFICIÊNCIA DO
ALUNO EM AVALIAÇÕES DE LARGA
ESCALA**

Relatório apresentado à disciplina Monografia do curso de graduação em Estatística, Departamento de Estatística, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Tufi Machado Soares

JUIZ DE FORA

2011

Dedicatória

A todos que, como eu, admiram a matemática da vida.

“O mapa não é o território”

Umberto Eco

“Models should be as simple as possible, but not more so”

Albert Einstein

Agradecimentos

A Deus e à Nossa Senhora, por toda a benção, proteção durante todo esse caminho e durante minha vida inteira.

À minha avó, Célia, cujo orgulho sempre almejo conseguir, cuja presença nesse momento seria minha maior alegria, cuja falta eu sinto a cada momento da minha vida.

À minha mãe, por ser meu exemplo de profissional e pessoa, por ser meu público favorito para todas as minhas apresentações, idéias e teoremas, por entender que, às vezes, só um colo de mãe ajuda.

Ao meu pai, por sempre estar ao meu lado e torcer por mim em todos os momentos, pela paciência nos momentos mais difíceis, pelos conselhos e por nunca me deixar levar pelo meu cansaço.

À minha irmã, Lívia, por, desde o ensino fundamental, me ajudar nos problemas da matemática e da vida, pelo apoio e confiança irrestrita.

À minha irmã, Ana Catarina, pelos conselhos valiosos, pelo apoio nos momentos mais importantes e pelas melhores gargalhadas.

Ao Igor, por estar ao meu lado na estatística e na vida, por falar a mesma língua que eu, pela alegria e ajuda nos momentos mais importantes.

Aos meus cunhados, Guilherme e Nicholas, pelas risadas nos melhores momentos.

À minha avó, Lia, e minha tia, Siegred, por sempre torcerem pela minha vitória.

À minha tia Zezé, que, desde os livros de cálculos até os conselhos sobre as matérias, esteve ao meu lado durante todo o percurso.

Às minhas tias Zitinha e Margarida, pela companhia desde a minha infância e pelo apoio.

À minha tia Cacau, por fazer questão de estar presente e pelo apoio em todos os momentos.

Aos meus amigos: Abby, Tássia, Denise, Raquel Morani, Vanessa, Mariana Salomão, Lucas e Cadu, pela amizade inconfundível, em todos os momentos.

À Raquel Borges, Leiliane e Lu, companheiros guerreiros de todas as dificuldades, loucuras, desesperos, e incontáveis alegrias, do ICE.

Ao Manoel, por sempre depositar fé em mim e por estar pronto para me ajudar em todas as horas.

Ao meu orientador, professor Tufi, pela confiança que depositou em mim à primeira vista, pelo incansável envolvimento e supervisão desse trabalho, por ter despertado em mim, com seu entusiasmo, a paixão pela modelagem.

Ao Neimar e ao Alexandre, pela ajuda nesse trabalho.

Ao professor Ronaldo Bastos, por ter me incentivado a fazer estatística quando eu não sabia o que esperar da vida, e por ter me acompanhado durante todo meu curso.

A todos que me deram oportunidades de ver que estatística é, de fato, o que eu amo e com o que eu quero mexer para o resto de minha vida.

Resumo

Os modelos da Teoria de Resposta ao Item construíram diversos avanços no campo educacional . A partir da TRI, foi possível compreender melhor a relação entre o aluno e o item durante o teste.

Apesar disso, tais modelos determinavam apenas a relação entre a proficiência do aluno e sua probabilidade de acertar determinado item, ignorando fatores emocionais que afetam sua performance cognitiva.

Esse trabalho propõe um modelo que incorpora o efeito da ansiedade do aluno no modelo tradicional da TRI, na tentativa de se obter sua “proficiência purificada” , obtendo, assim, resultados mais “verossímeis”.

Abstract

The Item Response Theory Models has allowed many advances on the educational field. With IRT, it was possible to have a better understanding of the relation between the student and the item of a test.

However, those models only determine the relation between the student's ability and his/her probability of getting the correct response, ignoring the emotional factors that affect his/her cognitive performance.

This paper proposes a model that incorporates the effect of the student's anxiety in IRT traditional model, in the attempt to get the student's "purified ability", and, therefore, obtain results that are closer to reality.

Lista de Tabelas

7.1	Assertivas do constructo “ansiedade”	38
8.1	Correlação entre os valores simulados e estimados	49
9.1	Correlação entre os valores estimados pelo modelo tradicional da TRI e pelo modelo proposto	55
9.2	Comparação entre as proficiências estimadas pelo modelo tradicional da TRI e o modelo proposto	62
9.3	Comparação entre as proficiências estimadas pelo modelo tradicional da TRI e o modelo proposto	62
10.1	Comparação entre as proficiências estimadas pelo modelo tradicional da TRI e o modelo proposto	64

Lista de Figuras

2.1	Curvas seguindo o modelo de Rasch	8
2.2	Curvas seguindo o modelo de 2 parâmetros	9
2.3	Curvas seguindo o modelo de 3 parâmetros	10
4.1	Item mal ajustado pelo modelo da TRI	20
4.2	Item bem ajustado pelo modelo da TRI	20
5.1	Depreciador bicaudal	23
5.2	Comparação das CCI's para valores baixos de ansiedade	25
5.3	Comparação das CCI's para valores intermediários de ansiedade	26
5.4	Comparação das CCI's para valores altos de ansiedade	27
8.1	Valores simulados, estimados e IC's para o parâmetro a	43
8.2	Valores simulados, estimados e IC's para o parâmetro b	44
8.3	Valores simulados, estimados e IC's para o parâmetro c	45
8.4	Valores simulados, estimados e IC's para o parâmetro δ_1	46
8.5	Valores simulados, estimados e IC's para o parâmetro δ_2	47
8.6	Valores simulados, estimados e IC's para o parâmetro λ_1	48
8.7	Valores simulados, estimados e IC's para o parâmetro λ_2	49

9.1	Valores estimados e IC's para o parâmetro δ_1	51
9.2	Valores estimados e IC's para o parâmetro δ_2	52
9.3	Depreciador unicaudal	53
9.4	Valores estimados para o parâmetro a	56
9.5	Valores estimados para o parâmetro b	57
9.6	Valores estimados para o parâmetro c	58
9.7	Valores estimados para o parâmetro δ_2	59
9.8	Valores estimados para o parâmetro λ_2	60
10.1	Ajuste do item 15 pelo modelo proposto	65
10.2	Ajuste do item 15 pelo modelo tradicional da TRI	66
10.3	Ajuste do item 26 pelo modelo proposto	66
10.4	Ajuste do item 26 pelo modelo tradicional da TRI	67
B.1	Questionário contextual	74

Sumário

Resumo	v
Abstract	vi
1 Introdução	1
I Fundamentação teórica	3
2 Modelagem da relação aluno-item: TRI unidimensional	4
2.1 Introdução	4
2.2 Pressupostos da TRI	5
2.3 Modelos da TRI	6
2.4 Estimação	11
3 TRI Multidimensional	13
3.1 Introdução	13
3.2 Modelos compensatórios	14
3.3 Não compensatórios (parcialmente compensatórios)	15
4 Ajustes dos modelos da TRI	17

4.1	Introdução	17
4.2	Não-ajuste	18
5	Modelo da TRI incorporando a ansiedade do aluno	21
6	Simulação bayesiana e estimação dos parâmetros do modelo	28
6.1	MCMC	28
6.1.1	Amostrador de Gibbs	31
6.1.2	Metropolis and Hasting	31
6.2	WinBUGS	32
II	Metodologia	34
7	Bases de dados utilizadas	35
7.1	Introdução	35
7.2	Bases de dados simulados	35
7.3	Base de dados reais: PROEB	37
III	Apresentação dos Resultados	40
8	Resultados da base simulada	41
8.1	Parâmetro a	43
8.2	Parâmetro b	43
8.3	Parâmetro c	44
8.4	Parâmetro δ_1	45

8.5	Parâmetro δ_2	46
8.6	Parâmetro λ_1	47
8.7	Parâmetro λ_2	48
8.8	Correlação entre os valores estimados e simulados	49
9	Resultados da base de dados reais (PROEB)	50
9.1	Introdução	50
9.2	Modelo com depreciador unicaudal	53
9.3	Resultados do modelo	54
9.3.1	Modelo tradicional da TRI	54
9.3.2	Comparação dos parâmetros	55
9.3.3	Comparação das proficiências	60
10	Ajuste do modelo proposto	64
IV	Conclusão e considerações finais	68
	Referências	71
	Apêndice	73
A	Sintaxe para geração dos dados simulados no MATLAB	73
B	Questionário contextual completo PROEB 2009	74
C	Sintaxe montada para o <i>software</i> WinBUGS para a base simulada	75

D	Sintaxe montada para o <i>software</i> WinBUGS com o modelo bicaudal para o PROEB	76
E	Sintaxe montada para o <i>software</i> WinBUGS com o modelo unicaudal para o PROEB	77
F	Sintaxe montada para o <i>software</i> WinBUGS com o modelo tradicional para o PROEB	78

Capítulo 1

Introdução

Até o final da década de 80, a Teoria da Resposta ao Item, ao tratar do desempenho de alunos em testes cognitivos de múltipla escolha, por exemplo em linguagem e matemática, considerava, especificamente, as habilidades e competências aferidas pelos itens e as performances dos alunos somente em relação a elas. Até então, fatores como ansiedade, competitividade e auto-eficácia não eram consideradas como elementos relevantes na resolução dos itens.

Atualmente, na psicometria, para a melhor aferição da performance dos alunos, é cada vez mais freqüente, na literatura, estudos que objetivam associar fatores emocionais ao desempenho dos alunos no teste.

A chamada ansiedade face ao teste é conceituada como a disposição individual para reagir com estados de ansiedade de forma mais intensa e freqüente, ativando cognições de preocupação, pensamentos irrelevantes que interferem com atenção, concentração e realização na situação de avaliações (Spielberger & Vagg *apud* Rosário et al. (2004)).

A ansiedade, como dita no parágrafo anterior, afeta diretamente o desempenho do aluno. As duas variáveis, portanto, possuem correlação negativa, isso é, o aumento

da ansiedade acarreta um desempenho menor.

Essas informações evidenciam mais fortemente a necessidade de um estudo aprofundado acerca da relação entre ansiedade e desempenho escolar, tornando possível uma maior compreensão acerca de como fatores emocionais afetam o aluno na hora da realização de um teste.

O objetivo desse trabalho é, portanto, propor um modelo de Teoria de Resposta ao Item (TRI) que incorpore o impacto da ansiedade na probabilidade de acerto do aluno ao item.

Esse trabalho será dividido em três partes:

- Na primeira parte, será feita uma abordagem teórica, a fim de se fundamentarem todos os conceitos necessários para a compreensão desse trabalho. Os primeiros capítulos abordarão a Teoria de Resposta ao Item Uni e Multidimensional e o ajustes dos modelos de TRI. Em seguida, será apresentado o modelo proposto, juntamente com o método de estimação e uma breve apresentação do *software* utilizado, o WinBUGS.
- A segunda parte conterà a metodologia, abordando as bases de dados utilizadas nesse trabalho (uma base com dados simulados e outra com dados reais).
- A terceira parte apresentará os resultados encontrados e o ajuste do modelo proposto;
- Por fim, serão elencadas as considerações finais acerca do modelo e as principais conclusões do estudo.

Parte I
FUNDAMENTAÇÃO TEÓRICA

Capítulo 2

Modelagem da relação aluno-item: Teoria de Resposta ao Item Unidimensional(TRI)

2.1 Introdução

O primeiro modelo de Teoria de Resposta ao Item (TRI) foi desenvolvido na década de 50, mas foi principalmente durante a década de 90 que a TRI ganhou bastante enfoque no Brasil. Com o uso crescente de avaliações em larga escala e com a crescente necessidade de se encontrar uma teoria que permitisse a comparação entre provas feitas por indivíduos diferentes ou provas feitas pelos mesmos indivíduos em momentos diferentes, a TRI começou a ser mais estudada e utilizada.

Na Teoria Clássica do Teste (TCT), o resultado é obtido através de escores brutos, isto é, a soma dos acertos do examinado em uma avaliação. Nessa teoria, o teste é analisado como um todo, ou seja, os itens são analisados somente em conjunto. Isso dificulta a comparabilidade entre testes e entre alunos, fazendo com que, nessa teoria, seja complicado comparar alunos que fizeram provas diferentes e comparar as características das provas feitas por grupos distintos de alunos. Por conta dessa

característica, diz-se que a TCT é “student dependent” e “test dependent”.

Nesse contexto de aferição de resultados em avaliações, a TRI vem sendo usada como uma teoria mais adequada, uma vez que permite, além da estimação do conhecimento do aluno sobre determinado tema, a análise do item proposto para aferir esse conhecimento, o que torna os problemas encontrados na TCT mais amenos. A TRI permite que a medida de proficiência do aluno e a medida de dificuldade do item sejam mensurados segundo uma mesma escala, e essa propriedade, como se verá, torna possível a comparabilidade de resultados de alunos submetidos a diferentes conjuntos de itens, e por outro lado, torna possível comparar as dificuldades dos itens mesmo quando submetidos a grupos de alunos diferentes.

2.2 Pressupostos da TRI

Geralmente, a TRI trabalha com modelos que utilizam parâmetros que caracterizam o item e um fator latente, que é característica do aluno. Esse fator é denominado habilidade ou, especificamente para o caso de avaliações educacionais, pode ser visto como a proficiência do aluno em determinado assunto. Dados os parâmetros de um determinado item e a habilidade θ de um aluno, é possível calcular a sua probabilidade de acertar o item.

A TRI unidimensional, objeto de estudo dessa seção, pressupõe que apenas uma habilidade é medida em cada item. Na prática, sabe-se que, quando um aluno responde a um item, existem diversas habilidades correlacionadas. Para a aplicação dos métodos da TRI, porém, é necessário que exista o foco para uma habilidade dominante sendo medida.

Essa teoria também pressupõe independência local, isto é, desconsiderando a habilidade do aluno, não existe relação entre as respostas de um determinado aluno para os diversos itens que compõem um teste. Isso significa que a probabilidade de o aluno acertar um determinado item não interfere em sua probabilidade de acertar outro item da prova.

Sejam $U_1, U_2, U_3, \dots, U_j$ as respostas de um determinado aluno para j itens de um teste. Considerando os itens como dicotômicos (certo ou errado), seja $P(U_j = 1|\theta)$ a probabilidade de o aluno acertar o item. A independência local pode ser escrita, em termos estatísticos, como:

$$P(U_1, U_2, \dots, U_j|\theta) = P(U_1|\theta)P(U_2|\theta) \cdots P(U_j|\theta) = \prod_j P(U_j|\theta) \quad (2.1)$$

Se o pressuposto da **invariância dos parâmetros** dos modelos da TRI for observado, então a TRI se torna uma ferramenta bastante útil para comparabilidade dos modelos. Isso significa que, independentemente do grupo de alunos que realiza o teste, os parâmetros do item não variam, e, independentemente de qual teste um aluno realiza, sua habilidade θ também não varia. Este é o caso da maior parte das aplicações e situações práticas.

2.3 Modelos da TRI

Existem diversos modelos da TRI. O que os diferencia é a quantidade de parâmetros que compõem cada um deles, o tipo de resposta admitido, o número de habilidades medidas, entre outros.

A curva que modela o item é conhecida como Curva Característica do Item, e

será referida, nesse trabalho, como CCI.

O modelo de um parâmetro, também chamado de modelo de Rasch (Rasch,1960), considera apenas o parâmetro b , que mede a dificuldade do item. O parâmetro b é um parâmetro de locação e, na escala de habilidade, corresponde ao ponto na qual a probabilidade de acerto é 0,5. Quanto maior o valor de b , maior é a dificuldade do item, e, portanto, maior é a habilidade requerida para que um aluno tenha 50% de chance de acerto em um item. Segundo esse modelo, a probabilidade de um aluno de habilidade θ acertar o item j é dada por:

$$P_j(\theta) = \frac{\exp(\theta - b_j)}{1 + \exp(\theta - b_j)} \quad (2.2)$$

O gráfico abaixo, construído no *software* livre R, revela, claramente, que as curvas são semelhantes e só se diferem pelo valor de b (na curva 1, $b=0.8$ e, na curva 2, $b=0.7$). Além disso, ambas as curvas encontram o eixo y no valor 0, isto é, pessoas com habilidade muito baixa, tendendo ao menos infinito, possuem probabilidade de acerto igual a 0. Dessa forma, a curva 2 é a curva 1 deslocada para a direita.

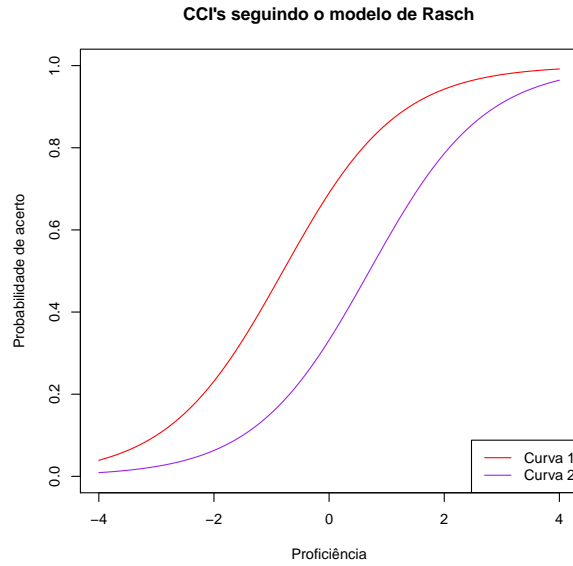


Figura 2.1: Curvas seguindo o modelo de Rasch

O modelo de dois parâmetros (primeiramente proposto por Birbaun e, depois, aprimorado por Lord, acrescentando o fator de escala D) considera o parâmetro b de dificuldade e o parâmetro a , que mede a discriminação do item. Ele mede o quanto o item consegue fazer distinção entre alunos, em relação a suas habilidades. O parâmetro a assume sempre valores positivos, de forma a garantir que o aumento da habilidade implica o aumento da probabilidade de acerto em um dado item. Segundo esse modelo, a probabilidade de um aluno de habilidade θ acertar o item j é dada por:

$$P_i(\theta) = \frac{\exp(Da_i(\theta - b_i))}{1 + \exp(Da_i(\theta - b_i))} \quad (2.3)$$

O gráfico abaixo, construído no *software* R, já revela maior diferença entre as curvas. Ambas encontram o eixo y no valor 0. Apesar disso, agora a curva 2 não pode ser obtida através do deslocamento da 1, uma vez que elas possuem valores

diferentes de discriminação (a curva 1 possui parâmetros $a=0.8$ e $b=-0.8$ e a curva 2 possui parâmetros $a=1.5$ e $b=0.7$).

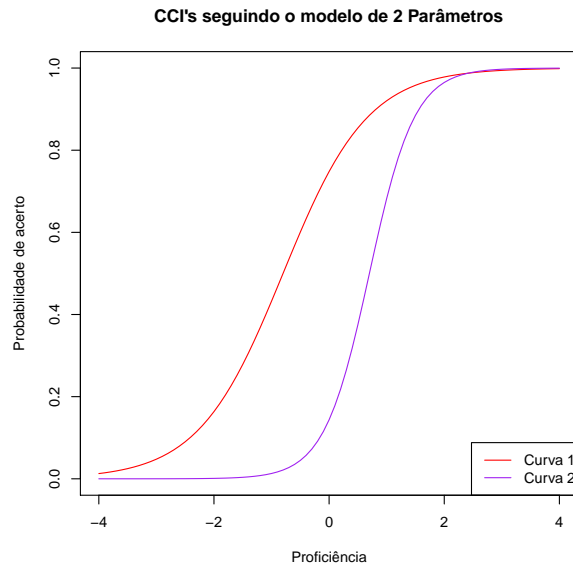


Figura 2.2: Curvas seguindo o modelo de 2 parâmetros

O modelo de três parâmetros (Birbaun, 1968) considera, além dos parâmetros a e b , o parâmetro c , que corresponde à probabilidade de acerto do item por pessoas com habilidade muito baixa. Sob determinadas situações, alguns pesquisadores o interpretam como a “probabilidade de acerto ao acaso”. A probabilidade, por esse modelo, é dada por:

$$P_i(\theta) = c_i + (1 - c_i) * \frac{\exp(Da_i(\theta - b_i))}{1 + \exp(Da_i(\theta - b_i))} \quad (2.4)$$

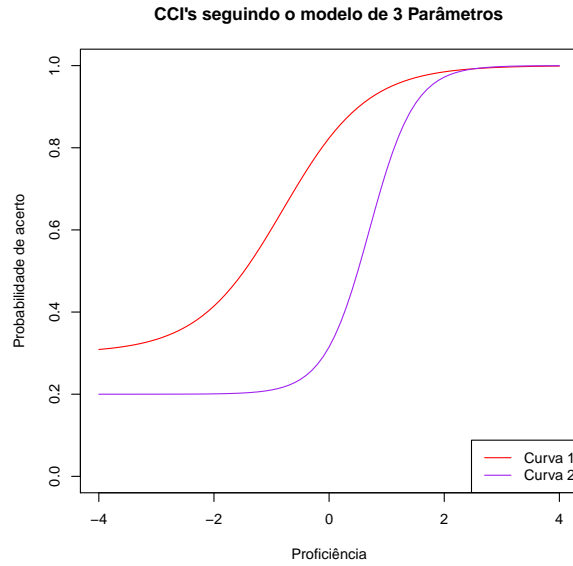


Figura 2.3: Curvas seguindo o modelo de 3 parâmetros

O gráfico acima, construído no *software* R, revela que as curvas não encontram o eixo y no valor igual a 0. Isso significa, como dito anteriormente, que pessoas com pouca proficiência possuem probabilidade de acerto diferente de 0. Como os parâmetros c das duas curvas são diferentes, seus interceptos com o eixo de probabilidade são diferentes (os parâmetros da curva 1 são $a=0.8, b=-1.8, c=0.3$ e os parâmetros da curva 2 são $a=1.5, b=0.7$ e $c=0.2$).

Pode-se trabalhar com o modelo logístico ou com o modelo probito, que considera a distribuição normal acumulada, e que será usada para a os algoritmos propostos por esse trabalho. A aproximação entre os dois se dá quando um parâmetro de escala $D=1,7$ é fixado para o modelo logístico. A função ogiva normal, equivalente ao modelo de três parâmetros acima citado é, em que $z=a_i(\theta-b_i)$:

$$P_j(\theta) = c_j + (1 + c_j) * \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) dz \quad (2.5)$$

2.4 Estimação

Para estimar os parâmetros da TRI, é necessário considerar as três situações possíveis.

- Já existe conhecimento prévio acerca dos parâmetros dos itens (no caso de questões retiradas de bancos de itens), sendo desejado estimar a habilidade do grupo de alunos que realizou o teste;
- Já existe conhecimento prévio acerca da habilidade de certo grupo de alunos (caso esse grupo seja usado para testar novos itens), sendo desejado estimar os parâmetros dos itens;
- Os parâmetros dos itens e as habilidades dos alunos são desconhecidos. Essa é a situação mais recorrente na prática, e requer, portanto, a estimação conjunta dos parâmetros dos itens e das habilidades.

Existem diversas formas de realização dessa última estimação citada. A forma mais utilizada é a máxima verossimilhança marginalizada. Existem, também, muitos métodos bayesianos, como as médias e moda *a posteriori*, que serão utilizados nesse trabalho.

É de interesse deste trabalho lidar, principalmente, com situações nas quais ambos os parâmetros dos itens e as habilidades são desconhecidos; por isso, o foco será na terceira situação exposta acima.

Seja U_{ij} a resposta do i -ésimo aluno ao j -ésimo item do teste e $P_{ij} = P(U_{ij}=1 \mid \theta_i, a_j, b_j, c_j)$ a probabilidade do i -ésimo aluno acertar ao j -ésimo item, e, analogamente,

$Q_{ij} = P(U_{ij}=0 \mid \theta_i, a_j, b_j, c_j) = 1 - P(U_{ij}=1 \mid \theta_i, a_j, b_j, c_j)$ a probabilidade de erro.

Considerando os pressupostos da TRI abordados anteriormente, e considerando um grupo de N alunos realizando um teste com n itens, tem-se:

$$f(U_{ij} \mid \theta, a, b, c) = \prod_{i=1}^N \prod_{j=1}^n P_{ij}^{U_{ij}} Q_{ij}^{1-U_{ij}} \quad (2.6)$$

Esse método de estimação não consegue, porém, estimar a habilidade de alunos que acertaram ou erraram todos os itens, não sendo possível, também, estimar com precisão os parâmetros desse item. Além disso, não existe uma solução única de maximização dessa função, sendo necessárias hipóteses adicionais acerca da distribuição da proficiência (como, por exemplo, a padronização), ou, sob enfoque Bayesiano, *priors* sobre a proficiência e outros parâmetros do modelo.

Uma solução para essas situações é usar a estimação bayesiana, que será aquela que terá maior enfoque nesse trabalho, e que será descrita em detalhes nos capítulos seguintes.

Capítulo 3

TRI multidimensional

3.1 Introdução

A complexidade dos processos psicológicos e educacionais Reckase(2009) evidenciou a necessidade de construção de modelos de Teoria de Resposta ao Item que conseguissem levar em consideração as diversas habilidades requeridas de um aluno para a realização de um item. Como exposto na seção acima, qualquer item requer do aluno um conjunto de habilidades, ora psicológicas e ora cognitivas, para a compreensão, realização, e conseqüentemente, acerto.

Como a aplicação da TRI unidimensional pressupõe unidimensionalidade, ou seja, a existência de uma habilidade dominante, não é possível aplicá-la para modelos mais complexos.

Surgiu, então, a Teoria da Resposta ao Item Multidimensional (MIRT). Os modelos da MIRT consideram que não é apenas um parâmetro que caracteriza o aluno (sua habilidade cognitiva), mas, sim, um vetor de parâmetros, que representa, além de sua proficiência, outras variáveis, como, por exemplo, as do âmbito emocional. Esse vetor, juntamente com os parâmetros característicos do item, irá definir a pro-

babilidade de acerto.

Os modelos da MIRT trabalham com o conceito de **espaço latente completo**, que é um conjunto de todas as variáveis latentes que discriminam os diferentes grupos de alunos (Lord and Novick,1968) segundo o desempenho no item. A indexação k varia entre 1 e m , sendo m o número de dimensões da proficiência, isso é, o número de habilidades requeridas para a realização de determinado item.

O vetor espaço latente completo é, portanto, definido por: $\theta=(\theta_1,\theta_2,\dots,\theta_m)'$. (Reckase,2009)

Os modelos da MIRT também podem abranger um, dois ou três parâmetros, assim como definidos para os modelos da TRI na seção anterior. Também podem ter função de ligação logística ou a ogiva normal.

Além disso, os modelos da MIRT são divididos entre modelos compensatórios e não compensatórios.

3.2 Modelos compensatórios

Os modelos compensatórios são aqueles que consideram que a falta de habilidade em um determinado traço latente pode ser compensada pela alta habilidade em outro. Isso significa que a probabilidade de um aluno acertar um item pode não ser seriamente afetada pelo baixo valor da habilidade θ_1 , por exemplo, se esse aluno possui um alto valor da habilidade θ_2 . Dessa forma, se estabelece uma relação de linearidade.

O modelo compensatório logístico de três parâmetros (Lord,1980) é dado por:

$$P(U_{ij}|\theta_i, a_j, c_j, d_j) = c_j + (1 - c_j) \frac{e^{a_j\theta_i + d_j}}{1 + e^{a_j\theta_i + d_j}}, \quad (3.1)$$

Na equação (3.1), d é um escalar que corresponde à interação entre os parâmetros a e b ($d = a \cdot b$).

Chamando $z_i(\theta_j) = a_i \theta_j + d_i$, o modelo compensatório considerando a ogiva normal (Bock and Schilling, 2003) é:

$$P_{ij}(\theta) = c_j + (1 - c_j) * \int_{-z_j(\theta_i)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt \quad (3.2)$$

3.3 Não compensatórios (parcialmente compensatórios)

Como a natureza compensatória do modelo define que, mesmo que $m - 1$ dimensões do espaço latente completo θ sejam muito baixas, o aluno ainda tem alta probabilidade de acerto de um determinado item caso a m -ésima dimensão seja suficientemente alta (Reckase, 2009), muitas vezes, esse modelo não consegue refletir a realidade. Sympson (1978) evidencia que um aluno, fazendo um teste de matemática, com baixa habilidade de compreensão de linguagem, terá também baixa probabilidade de acertar o item, por ter baixa probabilidade de compreender o que está sendo requerido no enunciado. Dessa forma, a alta habilidade do aluno em matemática não conseguirá compensar sua baixa habilidade em compreensão de linguagem.

Os modelos não-compensatórios consideram que o aluno precisa de uma habilidade suficientemente alta em cada dimensão do item para poder ter alta probabilidade de acerto. Sendo assim, a probabilidade de acerto do aluno é determinada pelo

produto das probabilidades de acerto em cada dimensão, estabelecendo uma relação de não-linearidade, não admitindo, assim, a natureza compensatória.

Reckase enfatiza que a melhor denominação para esse tipo de modelo é “modelo parcialmente compensatório”, pois um alto valor de θ em uma dimensão acarreta maior probabilidade de acerto do que um valor baixo, fazendo com que ocorra, de certa forma, compensação.

Seja m o número de dimensões medidas pelo item. O modelo logístico parcialmente compensatório de três parâmetros é dado por:

$$P(U_{ij}|\theta_i, a_j, c_j, d_j) = c_j + (1 - c_j) \prod_{l=1}^m \frac{e^{1.7a_{jl}(\theta_{il}-b_{jl})}}{1 + e^{1.7a_{jl}(\theta_{il}-b_{jl})}}, \quad (3.3)$$

O presente trabalho fará uso de modelos de natureza multidimensionais parcialmente compensatórios, embora seja uma modelagem diferente da apresentada acima.

Capítulo 4

Ajustes dos modelos da TRI

4.1 Introdução

Como todos os modelos estatísticos, é necessário analisar se o modelo da TRI para cada item possui um bom ajuste aos dados. É possível fazer a verificação desse ajuste graficamente ou através de estatísticas de ajuste.

A verificação gráfica do ajuste deve ser feita com cautela, uma vez que é necessário considerar que é sempre passível de subjetividade e parcialidade, isso é, a análise gráfica está sempre relacionada com a visão do pesquisador. Além disso, a determinação da qualidade do ajuste é extremamente complicada e instável em situações extremas, como, por exemplo, quando um item possui o parâmetro b próximo aos percentis superiores/inferiores ou quando o item segue uma ogiva, porém se distancia de determinado conjunto de dados.

O uso de estatísticas de ajuste ameniza os problemas decorrentes da análise gráfica, uma vez que trabalha com resultados de teste de hipótese e resultados matemáticos. Apesar disso, essa técnica não está isenta de erros. O teste de hipótese não-paramétrico realizado para testar qualidade de ajuste, o teste qui-quadrado,

apresenta limitações, uma vez que não deve ser aplicado para amostras muito pequenas (isto é, quando as frequências esperadas são menores do que cinco), e, no caso de amostras grandes, o teste se torna muito sensível, tendendo a sempre rejeitar a hipótese nula, uma vez que tende a revelar um ajuste não- significativo do modelo.

Dessa forma, a melhor maneira de concluir acerca da qualidade do ajuste do modelo é a utilização conjunta das duas técnicas acima citadas, uma vez que elas apresentam, sob determinados aspectos, complementaridade.

4.2 Não-ajuste

Existem diversos motivos pelos quais os modelos da TRI podem apresentar mau-ajuste, como, por exemplo, a inserção de itens problemáticos no teste (itens que possuem parâmetros muito altos ou muito baixos, referendados às distribuições consagradas na literatura, ou não compatíveis entre si, isto é, alta discriminação para itens muito fáceis e com alta chance de acerto ao acaso, acarretando dificuldade de estimabilidade dos mesmos).

Porém, mesmo quando os itens apresentam bons parâmetros, é possível verificar a existência de problemas de ajuste dos mesmos. Isso pode ocorrer quando ocorre a violação do pressuposto de invariância dos parâmetros ou a violação do pressuposto de simetria da CCI (Samejima,2000)

A violação do pressuposto de invariância dos parâmetros ocorre quando o item se comporta de forma diferente para dois diferentes subgrupos da população. Quando isso ocorre, diz-se que esse item possui DIF (Differential Item Functioning) (Soares et al., 2009). Como os parâmetros do item são diferentes para cada subgrupo,

tal violação acaba por causar o mau ajuste do modelo da TRI que desconsidera a existência do DIF.

Em geral, a TRI para respostas dicotômicas trabalha com ogivas simétricas (Samejima, 2000). A existência da assimetria pode ser resultado da existência de uma variável exógena ao problema. Isso faz com que o efeito latente siga uma curva assimétrica, acarretando, portanto, um mau-ajuste do modelo tradicional da TRI.

No caso do mau ajuste do modelo causado pela existência de uma variável exógena, o melhor ajuste pode ser obtido através da incorporação de tal variável ao modelo. Para que isso seja possível, é necessário, primeiramente, identificar qual é a variável exógena que deve ser considerada no problema e como ela pode ser inserida no modelo, a fim de se garantir uma modelagem mais adequada.

Abaixo, encontram-se dois gráficos elaborados pelo Sisani (Sistema de Análise de Itens), um *software* estatístico desenvolvido pela equipe do CAEd (Centro de Políticas Públicas e Avaliação da Educação). Esse software, além de calcular DIF, estatísticas clássicas e a CCI do item, testa o ajuste dos mesmos. Na primeira figura, a curva se mostra distante dos valores observados, revelando mau ajuste. Na segunda, observa-se um bom ajuste do modelo aos itens.

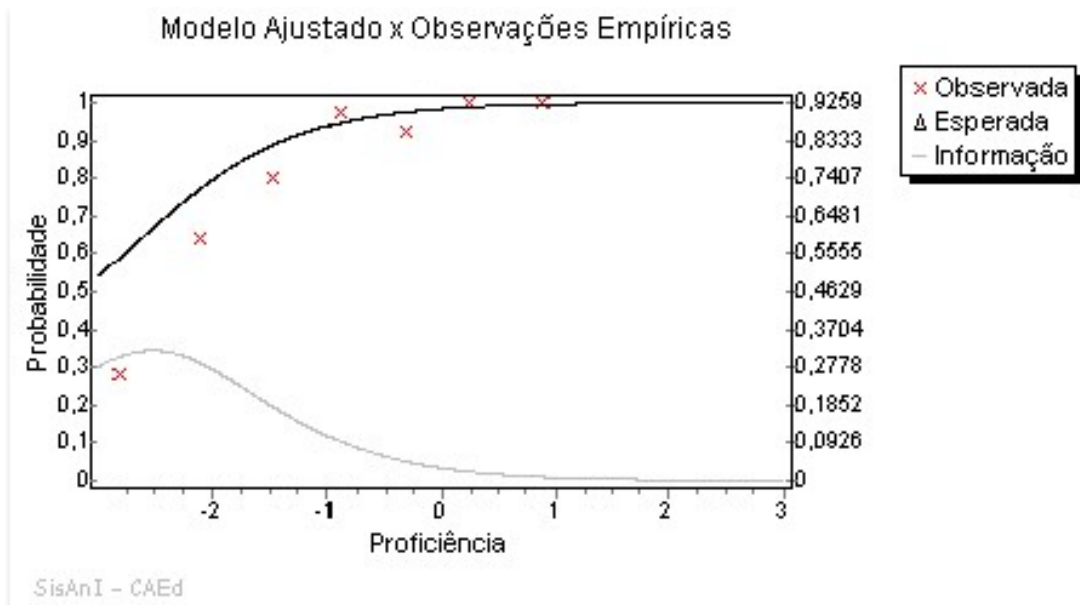


Figura 4.1: Item mal ajustado pelo modelo da TRI

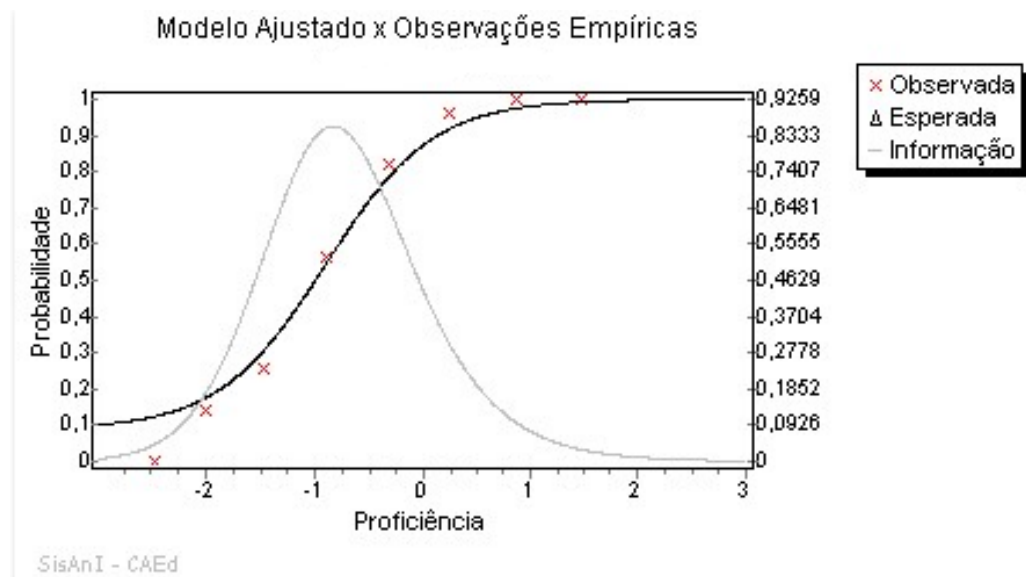


Figura 4.2: Item bem ajustado pelo modelo da TRI

Capítulo 5

Modelo da TRI incorporando a ansiedade do aluno

Como visto na primeira seção desse trabalho, é cada vez mais reconhecido o efeito das variáveis emocionais de um aluno em seu desempenho no teste, isto é, da existência de correlação entre a variável emocional e a proficiência do aluno. Dessa forma, é concebível pensar que uma variável emocional possa ser uma variável exógena e, se essa variável for desconsiderada, pode causar mau ajuste das CCI's.

Com o objetivo, então, de melhorar o ajuste e levando em consideração a discussão da primeira seção, propõe-se a incorporação da informação trazida pela variável emocional ao modelo da TRI.

Tal incorporação tornará possível, portanto, o estudo do impacto que tal variável possui na proficiência, além de possibilitar, simultaneamente, a análise da “proficiência purificada”, isto é, a proficiência que revela a habilidade do aluno testada no item, sem a interferência de outros fatores. Uma desvantagem trazida por tal incorporação é a necessidade de uma reavaliação acerca das interpretações dos parâmetros tradicionais da TRI, que não é, entretanto, objeto pertinente do escopo desse trabalho.

A fim de o modelo continuar sendo estimável e interpretável, é recomendável que tal incorporação possua as seguintes características:

- Unicidade de solução dado a unicidade da distribuição de proficiência. Uma vez definida *a priori* para a proficiência, é fundamental que o modelo apresente solução única para seus parâmetros;
- Robustez às condições iniciais: quando o modelo for implementado computacionalmente, é importante ter uma razoável independência às condições iniciais;
- Boa estimabilidade, pressupondo, portanto, parcimônia e *prioris* adequadas;
- Inatividade quando o item não sofre influência do fator. Caso o item não seja afetado pela variável exógena em questão, é importante que o modelo consiga captar tal característica;
- Adequação aos dados, levando a um bom ajuste.

Como discutido na primeira seção desse trabalho, está posto que a ansiedade afeta o desempenho do aluno na prova, em escalas diferentes.

Na literatura, é possível encontrar diversos estudos que comprovam esse impacto, demonstrando que a ansiedade apresenta efeito em forma de U invertido (Woodman and Hardy, 2001 *apud* Mimoso, 2002). Esse formato indica que existe um nível de ansiedade ótimo, que não prejudica o ser humano na realização de suas tarefas. Tanto níveis de ansiedade mais baixos quanto mais altos podem afetar o desempenho do ser humano, sendo que tal relação não é necessariamente simétrica.

Considerando um modelo de TRI multidimensional parcialmente compensatório, podemos incorporar uma expressão, aqui denominada de depreciador, que estabeleça a relação da ansiedade com o desempenho do aluno. Assim, tal depreciador possui o formato de um U invertido, como é possível observar na figura abaixo, feita no *software* R.

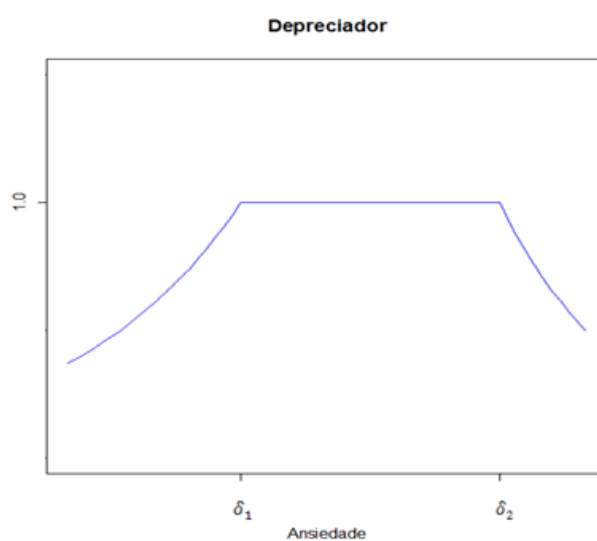


Figura 5.1: Depreciador bicaudal

Esse depreciador, $\eta(\delta_1, \delta_2, \lambda_1, \lambda_2, \varphi)$, pode ser escrito da seguinte forma:

$$\eta(\delta_1, \delta_2, \lambda_1, \lambda_2, \varphi) = \begin{cases} e^{\lambda_1 * (\varphi - \delta_1)} & \text{se } \varphi < \delta_1 \\ 1 & \text{se } \delta_1 \leq \varphi \leq \delta_2 \\ e^{-\lambda_2 * (\varphi - \delta_2)} & \text{se } \varphi > \delta_2 \end{cases} \quad (5.1)$$

Em que:

- φ é a medida do nível de ansiedade do aluno;
- δ_1 e δ_2 são parâmetros do item, medidos na escala de ansiedade. Como é possível ver pela figura, o parâmetro δ_1 representa o valor limite, dentre os

níveis mais baixos, nos quais a ansiedade causa impacto; e o parâmetro δ_2 representa o valor a partir do qual, nos níveis mais altos, a ansiedade também causa impacto;

- λ_1 e λ_2 são parâmetros do item, e determinam a taxa de crescimento/decaimento das curvas exponenciais.

A incorporação desse depreciador na CCI se dá através da multiplicação do mesmo à π_{ij} , sendo que $P(U_{ij} | \theta, a, b, c) = c + (1 - c) * \pi_{ij}$. Seja $P(U_{ij} | \theta, a, b, c, \eta)$ a probabilidade de que um aluno i selecionado ao acaso acerte o item j . Teremos:

$$P(U_{ij} | \theta_i, a, b, c) = \begin{cases} c_j + e^{\lambda_{1j} * (\varphi_i - \delta_{1j})} * \frac{1 - c_j}{1 + \exp(-1.7a_j(\theta_i - b_j))} & \text{se } \varphi_i < \delta_{1j} \\ c_j + \frac{1 - c_j}{1 + \exp(-1.7a_j(\theta_i - b_j))} & \text{se } \delta_{1j} \leq \varphi_i \leq \delta_{2j} \\ c_j + e^{-\lambda_{2j} * (\varphi_i - \delta_{2j})} * \frac{1 - c_j}{1 + \exp(-1.7a_j(\theta_i - b_j))} & \text{se } \varphi_i > \delta_{2j} \end{cases} \quad (5.2)$$

Em que:

- a, b, c são os parâmetros de discriminação, dificuldade e acerto ao acaso, respectivamente, assim como no modelo tradicional da TRI, apresentado no Capítulo 2.
- Os demais parâmetros são os do depreciador, apresentados acima.

A inserção desse depreciador faz com que, para dado nível de ansiedade abaixo de δ_1 ou acima de δ_2 , a probabilidade de acerto, para todos os valores de θ seja menor, como podemos observar pelos gráficos abaixo.

Considerando $\delta_1 = -1$, $\delta_2 = 1,9$, $\lambda_1 = 0,5$, $\lambda_2 = 0,7$, $a = 1,2$, $b = 1$, $c = 0,3$, podemos, com o auxílio do *software* R, comparar as curvas em 3 cenários possíveis:

- Para valores de $\varphi < \delta_1$, representado, no exemplo, por $\varphi=-1.1$:

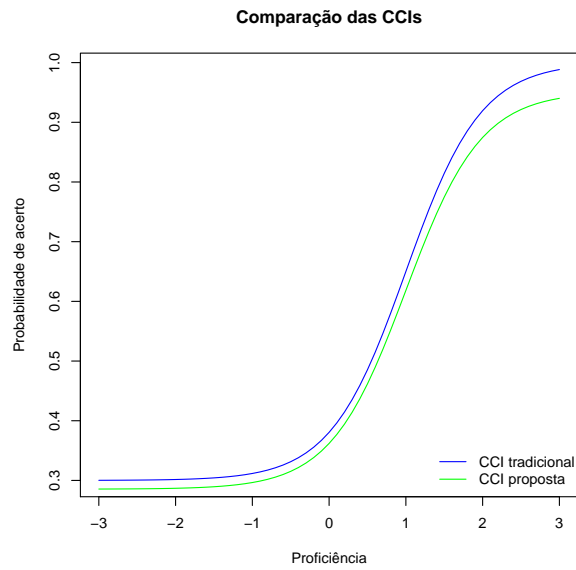


Figura 5.2: Comparação das CCI's para valores baixos de ansiedade

- Para $\delta_1 < \varphi < \delta_2$, representado, no exemplo, por $\varphi=0.5$:

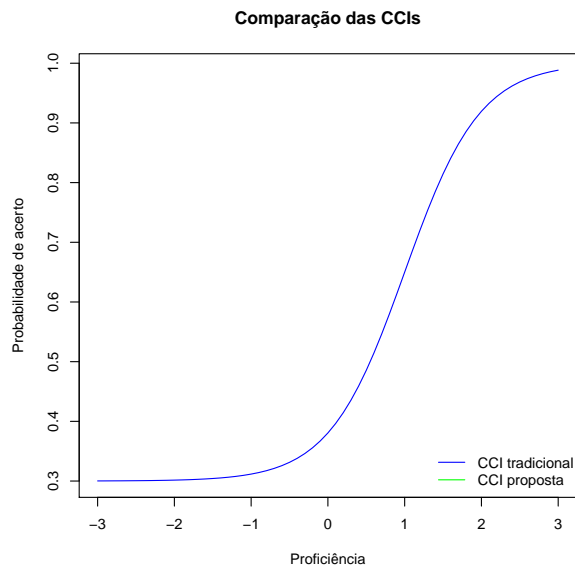
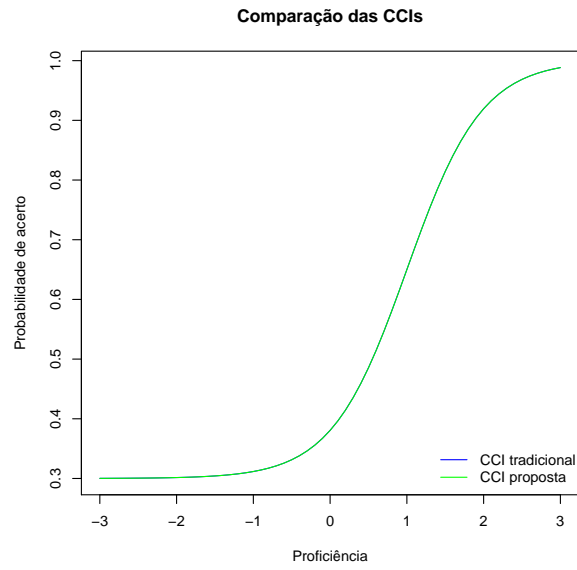


Figura 5.3: Comparação das CCI's para valores intermediários de ansiedade

- Para valores de $\varphi > \delta_2$, representado, no exemplo, por $\varphi=2$:

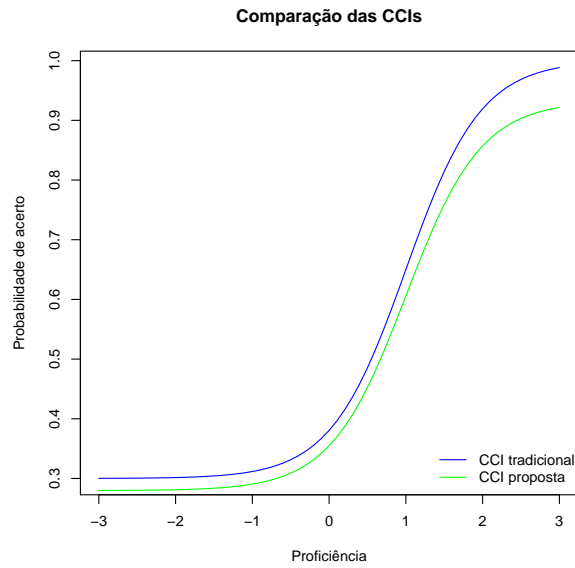


Figura 5.4: Comparação das CCI's para valores altos de ansiedade

Na prática, a redução da probabilidade de acerto significa que, caso o aluno tenha um nível de ansiedade fora do intervalo (δ_1, δ_2) , sua probabilidade de acerto será menor do que aquela predita pelo modelo tradicional da TRI, dados sua proficiência e os parâmetros do item. Dessa forma, a incorporação do depreciador possibilita a obtenção da “proficiência purificada”, como discutido nas seções anteriores.

Capítulo 6

Simulação bayesiana e estimação dos parâmetros do modelo

6.1 MCMC

Como citado nas seções iniciais desse trabalho, uma das formas de estimação dos parâmetros é a abordagem bayesiana. Nessa abordagem, incorpora-se o conhecimento existente *a priori* com as informações trazidas pelos dados.

A informação trazida pelos dados é a Equação de Verossimilhança, que consiste na função de densidade conjunta dos alunos e dos itens. Como a TRI possui como pressuposto a independência local e, também, a independência entre alunos (isto é, a probabilidade de um *i*-ésimo aluno acertar um determinado item não é afetado pela probabilidade de acerto de qualquer outro aluno), a Equação de Verossimilhança para a TRI é dada pelo produto das probabilidades de acerto. Com pensamento análogo àquele apresentado no Capítulo 1 desse trabalho, considerando o modelo proposto acima, isto é, a equação 5.1, temos, portanto, que a probabilidade de erro, denotada por $Q_{ij}(U_{ij}|a,b,c,\eta)$, é dada: $Q_{ij}(U_{ij}|a,b,c,\eta) = 1 - P_{ij}(U_{ij}|a,b,c,\eta)$. Assim, a Equação de Verossimilhança será dada por:

$$f(U_{ij}|\theta, a, b, c, \eta) = \prod_{j=1}^N \prod_{i=1}^n P_{ij}^{U_{ij}} Q_{ij}^{1-U_{ij}} \quad (6.1)$$

O conhecimento existente *a priori* é expresso através das distribuições *a priori* dos parâmetros, nas quais são definidos seus valores esperados e suas variâncias. O conhecimento *a priori* expressa o estado de conhecimento ou ignorância existente acerca dos parâmetros (Box and Tiao, 1992). A *priori* do conjunto de parâmetros dos itens, aqui chamado de ζ , é denotada por $p(\zeta; \theta)$ e, como os parâmetros dos itens são independentes das habilidades dos alunos, temos que: $p(\zeta; \theta) = p(\zeta) * p(\theta)$.

Dada a distribuição *a priori*, a Equação de Verossimilhança e os dados, é possível calcular a distribuição de probabilidade $p(\zeta, \theta | U_{ij})$, que é chamada distribuição *a posteriori*. A partir dessa distribuição, podemos fazer inferências sobre os parâmetros (Box and Tiao, 1992).

Em termos matemáticos, temos:

$$p(\zeta, \theta | U_{ij}) \propto f(U_{ij} | \theta, \zeta) * p(\zeta) * p(\theta) \quad (6.2)$$

Os métodos da “moda *a posteriori*” e “média da *a posteriori*” possuem a vantagem de estimarem os parâmetros para qualquer padrão de resposta (mesmo para casos extremos, como quando o aluno acerta todas as questões da prova, por exemplo), mas possuem a desvantagem de estarem ligadas à *priori*. Dessa forma, é necessário garantir que as distribuições *a priori* para os parâmetros são boas, isto é, que elas possuem média próxima dos valores verdadeiros e variâncias razoavelmente altas. Isso pode evitar viés nos resultados.

Um método de simulação bayesiano que é usado para gerar as amostras da dis-

tribuição conjunta *a posteriori* dos parâmetros dos modelos de resposta ao item é o *Monte Carlo Markov Chain* (MCMC), abordado por Gamerman, 1997.

A estimação dos parâmetros via MCMC se dá através de um algoritmo, que é repetido para todos os parâmetros. Esse algoritmo, em termos gerais, consiste nos seguintes passos:

- Considerando as distribuições *a priori* definidas para os parâmetros e os valores iniciais (que podem ser definidos ou gerados), sorteio de números aleatórios;
- Substituição do valor do parâmetro dentro da Equação de Verossimilhança, dado os demais parâmetros da iteração anterior (ou da condição inicial, no caso da primeira iteração);
- Caso esse novo valor maximize a verossimilhança, ocorre a incorporação do mesmo. Caso o contrário, o valor é descartado;
- Repetição desses passos até atingir o critério de parada (convergência).

Dessa forma, o MCMC “testa” diversos valores, dadas as *prioris* definidas previamente, incorporando aquele que maximiza a verossimilhança. Quando o algoritmo alcança convergência, ele descarta os demais valores do parâmetro, incorporando sempre o mesmo valor (ou valores muito próximos, considerando o erro padrão).

Existem dois métodos de simulação Bayesiana MCMC que merecem destaque: *Metropolis and Hasting* e *Amostrador de Gibbs* (Gibbs Sampler). Esses dois métodos serão, aqui, descritos de forma geral. A combinação de ambos resultará no método utilizado nesse trabalho.

6.1.1 Amostrador de Gibbs

Seja ρ o conjunto de parâmetros que desejamos estimar e Z os dados. Usaremos o símbolo (\cdot) para representar os demais parâmetros que não tenham sido explicitados em uma expressão. Teremos:

- Passo 1: Inicialize um contador de iterações $j = 1$ e defina um vetor de valores iniciais: $\rho^{(0)} = (\rho_1^1, \rho_2^1, \dots, \rho_k - 1^1, \rho_k^1, \dots, \rho_n^1)$;
- Passo 2: Obtenha um novo vetor $\rho^{(j)}$, simulando:

$$\rho_1^j \sim P(\rho_1 | \rho_2^{j-1}, \rho_3^{j-1}, \dots, \rho_n^{j-1}; Z) \propto P(Z | \rho_1, \rho_2^{j-1}, \rho_3^{j-1}, \dots, \rho_n^{j-1}) P(\rho_1 | (\cdot))$$

$$\rho_2^j \sim P(\rho_2 | \rho_1^j, \rho_3^{j-1}, \dots, \rho_n^{j-1}; Z) \propto P(Z | \rho_1, \rho_2, \rho_3^{j-1}, \dots, \rho_n^{j-1}) P(\rho_2 | (\cdot))$$

⋮

$$\rho_n^j \sim P(\rho_n | \rho_1^j, \rho_2^j, \dots, \rho_{n-1}^j; Z) \propto P(Z | \rho_1, \rho_2, \rho_3, \dots, \rho_{n-1}) P(\rho_n | (\cdot)) \quad (6.3)$$

- Passo 3: Pare quando atingir o critério de parada, isto é, quando os vetores aleatórios convergirem para a distribuição estacionária.

6.1.2 Metropolis and Hasting

Usando a mesma notação da seção anterior, suponha que se deseja gerar uma amostra de $P(\rho_k | \rho_1^j, \rho_2^j, \dots, \rho_{k-1}^j; \rho_k^{j-1}, \dots, \rho_n^j)$:

- Passo 1: Inicialize um contador de iterações $j=1$ e defina um vetor de valores iniciais: $\rho^{(0)}=(\rho_1^1, \rho_2^1, \dots, \rho_{k-1}^1, \rho_k^1, \dots, \rho_n^1)$;

- Passo 2: Obtenha um novo vetor ρ^* , simulando uma densidade proposta $q(\rho_k^* | \rho_k^{j-1})$, $k=1, 2, \dots, n$;

- Passo 3: Calcule a razão:

$$r(\rho_k^{j-1}, \rho_k^*) = \frac{P(\rho_k^* | (\cdot)) q(\rho_k^{j-1} | \rho_k^*)}{P(\rho_k^{j-1} | (\cdot)) q(\rho_k^* | \rho_k^{j-1})} \quad (6.4)$$

, tal que:

$$r(\rho_k^{j-1}, \rho_k^*) = \frac{P(Z | \rho_1^j, \rho_2^j, \dots, \rho_{k-1}^j, \rho_k^*, \rho_{k+1}^{j-1}, \dots, \rho_n^{j-1})}{P(Z | \rho_1^j, \rho_2^j, \dots, \rho_{k-1}^j, \rho_k^{j-1}, \rho_{k+1}^{j-1}, \dots, \rho_n^{j-1})} \frac{P(\rho_k^*) q(\rho_k^{j-1} | \rho_k^*)}{P(\rho_k^{j-1}) q(\rho_k^* | \rho_k^{j-1})} \quad (6.5)$$

- Passo 4: Encontre a probabilidade de aceitação do novo valor fazendo: $P_{ac} = \min(1, r(\rho_k^{j-1}, \rho_k^*))$;
- Passo 5: Gere $Z_{ac} \sim \text{Ber}(P_{ac})$;
- Passo 6: Faça $\rho_k^j = \rho_k^*$ se $Z_{ac}=1$ e $\rho_k^j = \rho_k^{j-1}$ se $Z_{ac}=0$;
- Passo 7: Prossiga até a estacionaridade da distribuição $(\rho_k^j | (\cdot))$

6.2 WinBUGS

O *software* WinBUGS é uma versão livre para o Windows do antigo programa BUGS (Bayesian Using Gibbs Samples), que permite a estimação de parâmetros de modelos via Monte Carlo Markov Chain (MCMC).

Esse *software* possui uma linguagem própria, bastante parecida, porém, com a do software R, existindo, nesse programa, um pacote que faz uma ligação direta

com o WinBUGS (R2winBUGS). É possível, também, formular modelos usando o “doodle”, que é uma forma de esquematização própria do *software*. Porém, nesse trabalho, foi usada a sintaxe textual do WinBUGS.

Uma vez montada a sintaxe do modelo, é possível escolher os parâmetros que são de interesse e esses serão monitorados pelas iterações, cuja quantidade fica a critério do pesquisador. Quando as iterações chegam ao fim, é possível obter as estatísticas descritivas (média, mediana, intervalo de confiança, desvio padrão, entre outros), o histórico dos valores dos parâmetros incorporados (importantes para se observar se houve convergência), os quartis, o traço (traço dinâmico de cada parâmetro), a densidade e a auto-correlação nas cadeias dos parâmetros. Além disso, é possível obter os últimos valores incorporados dos parâmetros, o que possibilita ao pesquisador começar uma nova cadeia de iteração.

O uso desse *software* vem crescendo bastante com o advento da estatística bayesiana no Brasil. Além disso, como dito anteriormente, possui uma linguagem similar ao R e a utilização do algoritmo de Gibbs vem se mostrando bastante eficaz na realização de diversos problemas práticos e teóricos.

Parte II
METODOLOGIA

Capítulo 7

Bases de dados utilizadas

7.1 Introdução

A fim de se testar o modelo proposto no capítulo 5, serão usadas duas bases de dados. Uma das bases é composta por dados simulados, tendo como objetivo testar a capacidade de convergência do modelo, assim como a recuperação dos parâmetros pelo programa utilizado, o WinBUGS. A segunda base é composta por dados reais do Programa de Avaliação da Rede Pública de Educação Básica (PROEB), do ano de 2009.

7.2 Bases de dados simulados

A base de dados simulados foi composta através da geração dos parâmetros de um teste de quarenta itens, realizado por dois mil alunos.

Para isso, utilizou-se o programa MATLAB (versão 2007a), e a sintaxe utilizada encontra-se no apêndice A .

As distribuições escolhidas para a geração dos parâmetros foram truncadas para que os valores possíveis estivessem dentro de limites bem conhecidos na prática.

- $a \sim LN(1.7860, 1.3771)|_{(0.5, 3.5)}$;
- $b \sim N(0, 2)|_{(-4, 4)}$;
- $c \sim Beta(5, 17)|_{(0, 0.5)}$;
- $\delta_1 \sim unif(-4, 4)$;
- $\delta_2 \sim unif(-4, 4)$;
- $\lambda_1 \sim gamma(1, 1)$;
- $\lambda_2 \sim gamma(1, 1)$;
- $\varphi \sim N(0, 1)$;
- $\theta \sim N(0, 1)$.

Considerando, assim, os parâmetros dos itens e os dos alunos, calculou-se a probabilidade de acerto do aluno para cada item. A partir dessa probabilidade calculada, foi possível, através da distribuição binomial, criar um padrão de resposta para cada aluno (sendo que 1 corresponde ao acerto e 0 corresponde ao erro).

Uma grande deficiência da simulação de dados é sua incapacidade de manter correlações que existem na realidade. Sabe-se que os parâmetros de um item não mantêm independência entre si, e é intuitivo pensar que a ansiedade também possui correlação com a proficiência. Como os dados são gerados de forma aleatória, os parâmetros são independentes entre si, podendo ocasionar itens com parâmetros não compatíveis, acarretando dificuldade na estimação dos mesmos.

O uso de dados simulados nesse trabalho, como observado anteriormente, terá, portanto, o objetivo de testar o estimador, isto é, testar o algoritmo produzido para o programa WinBUGS.

7.3 Base de dados reais: PROEB

A fim de se testar o modelo numa base de dados reais, foi utilizada a base de dados do Programa de Avaliação da Educação Básica do Estado de Minas Gerais (PROEB) do ano de 2009.

O PROEB é uma avaliação anual que faz parte do Sistema Mineiro de Avaliação da Escola Pública (SIMAVE), instituído pela Secretaria de Educação do Estado em 2000.

O PROEB é composto por testes cognitivos de Língua Portuguesa e Matemática que são aplicados em alunos do 5º, 9º e 12º ano (antigas 4ª e 8ª séries do Ensino Fundamental e 3ª Série do Ensino Médio) e questionários contextuais direcionados a alunos, professores e diretores.

Os alunos, no primeiro dia de prova, fazem o teste cognitivo de Língua Portuguesa e respondem a um questionário que tem como objetivo conhecer o padrão de vida do aluno, avaliando suas características culturais e sócio-econômicas. No segundo dia, os alunos fazem o teste cognitivo de Matemática e respondem a um questionário que possui perguntas de escopo psicológico. Para esse trabalho, foi usado o banco de dados correspondente ao segundo dia de prova dos alunos do 5º ano.

Para a prova do PROEB, são elaborados 169 itens, divididos em 13 blocos compostos por 13 questões. Cada forma de teste consiste em três blocos, de forma

que cada aluno responda a 39 questões. Assim, formas de teste diferentes possuem questões em comum, podendo ter a ordem alterada.

Optou-se, por simplicidade, trabalhar com apenas a primeira forma de teste, formada pelos três primeiros blocos. Foram, portanto, analisados 39 itens de Matemática.

O questionário contextual aplicado juntamente com a prova de matemática é composto por 34 assertivas, às quais o aluno deveria responder por uma escala de Likert de 4 pontos (de “Discordo totalmente” até “Concordo totalmente”). Dentro desse questionário, foram inseridas 9 assertivas pertencentes à escala de ansiedade, que foram utilizadas para a criação do indicador de ansiedade utilizado neste trabalho e que se encontram na tabela abaixo. O questionário contextual completo encontra-se no apêndice B.

Tabela 7.1: Assertivas do constructo “ansiedade”

Item	Assertivas do constructo “ansiedade”
Q02	Eu queria ser feliz como meus amigos.
Q03	Eu fico chateado quando o professor me chama à atenção.
Q04	Eu não tenho confiança nas coisas que eu faço.
Q05	Eu me canso com qualquer coisa.
Q06	Eu fico preocupado com coisas que não são importantes
Q07	Eu fico nervoso porque tenho problemas.
Q08	Eu tenho dificuldade de decidir o que vou fazer.
Q12	Eu tenho dificuldades para escrever as palavras que são ditadas.
Q15	Eu esqueço rápido o que aprendi.

Os alunos que não responderam à nenhuma pergunta do questionário contextual, cujo nível de ansiedade não poderia ser medido, foram retirados da base e o indicador de ansiedade foi criado a partir do *software* estatístico PARSCALE (versão 4.1.2328.4, Janeiro de 2003), que trabalha com itens politômicos.

A base de dados utilizada para o trabalho foi, após tal retirada, composta por 9401 alunos, respondendo a 39 itens de Matemática.

Parte III

**APRESENTAÇÃO DOS
RESULTADOS**

Capítulo 8

Resultados da base simulada

Para a estimação dos parâmetros dos itens simulados, construiu-se uma sintaxe para o *software* WinBUGS, que pode ser encontrada no apêndice C.

As *prioris* designadas para todos os parâmetros, como pode ser observado na sintaxe, foram:

- $a \sim \text{unif}(0, 3.5)$
- $b \sim \text{unif}(-4, 4)$
- $c \sim \text{unif}(0, 1)$
- $\delta_1 \sim \text{unif}(-4, 4)$
- $\delta_2 \sim \text{unif}(0, 8) + \delta_1$
- $\lambda_1 \sim \text{gamma}(2, 3)$
- $\lambda_2 \sim \text{gamma}(2, 3)$
- $\theta \sim N(0, 1)|_{(-4,4)}$

É importante notar, na sintaxe utilizada no WinBUGS, que o modelo de TRI utilizado foi o modelo de probito normal, que, no caso de MCMC, é mais utilizado do que o modelo logístico, como aponta Gúzman em sua tese de doutorado (Gúzman, 2005)

As *prioris* foram escolhidas com o intuito de não causar fortes restrições aos números aleatórios gerados no MCMC, dando, portanto, maior “liberdade” ao estimador.

Foram realizadas 15000 iterações, tendo sido monitoradas as últimas 5000, com *thinning* de 5, a fim de se descorrelatar a cadeia. Dessa forma, ao final, tem-se uma amostra de 1000 valores para cada parâmetro.

É importante considerar que, no algoritmo do WinBUGS, existe a restrição de que $\delta_1 < \delta_2$, intrínseca ao modelo (basta observar a forma do depreciador, na figura 4.1). Apesar disso, como dito anteriormente, a simulação dos parâmetros foi feita de forma independente, através das distribuições acima descritas, o que pode ocasionar $\delta_1 > \delta_2$. Portanto, é possível que existe dificuldade na estimação desses parâmetros.

A seguir, será feita uma análise dos resultados para cada parâmetro do item. Foram consideradas, como estimativas, a mediana da distribuição, uma vez que esse valor é mais robusto do que a média e, considerando o Teorema Central do Limite (que determina que, para tamanhos de amostras muito grandes, qualquer distribuição tende a se assemelhar à distribuição normal), sabe-se que o valor esperado da mediana é o valor esperado da distribuição, uma vez que a distribuição normal é simétrica. As tabelas dos parâmetros, por item, se encontram no CD anexo a este trabalho.

8.1 Parâmetro a

Pelo gráfico abaixo, podemos perceber que, com exceção dos itens 12 e 21, o intervalo de 95% de credibilidade do *software* contém parâmetro a para todos os itens. É importante notar que o valor do parâmetro a simulado para esses itens é de 3.5, valor muito alto e difícil de ser encontrado na prática. Como havia sido observado anteriormente, a simulação de dado possui essa deficiência, o que acaba afetando a estimabilidade dos parâmetros.

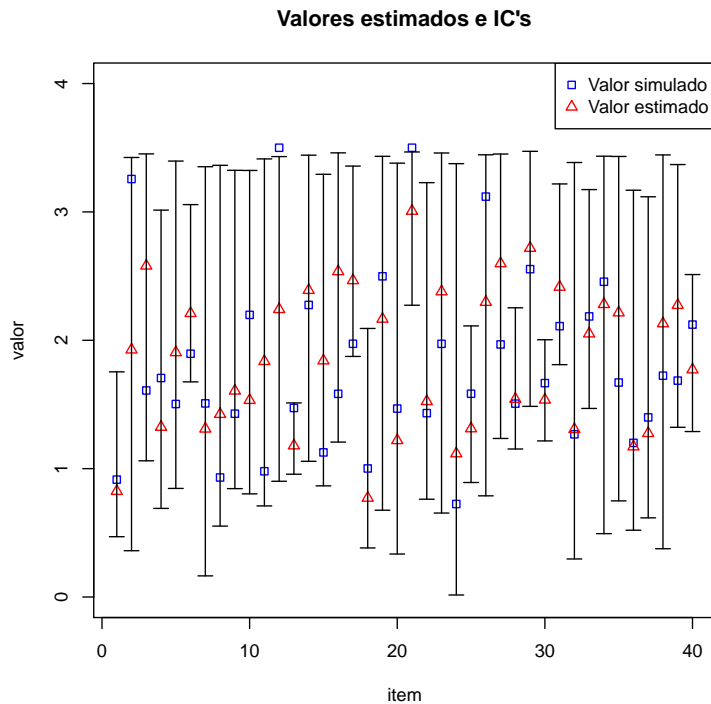


Figura 8.1: Valores simulados, estimados e IC's para o parâmetro a

8.2 Parâmetro b

O gráfico abaixo revela que apenas para cinco itens o intervalo de credibilidade estimado pelo WinBUGS não contém o verdadeiro valor do parâmetro. Para três

desses, os valores simulado para o parâmetro b tiveram módulo igual a 4, ocasionando o mesmo problema mencionado na seção acima. Um desses itens foi o item 12, que possui valor alto para o parâmetro a , podendo afetar a estimabilidade do parâmetro b .

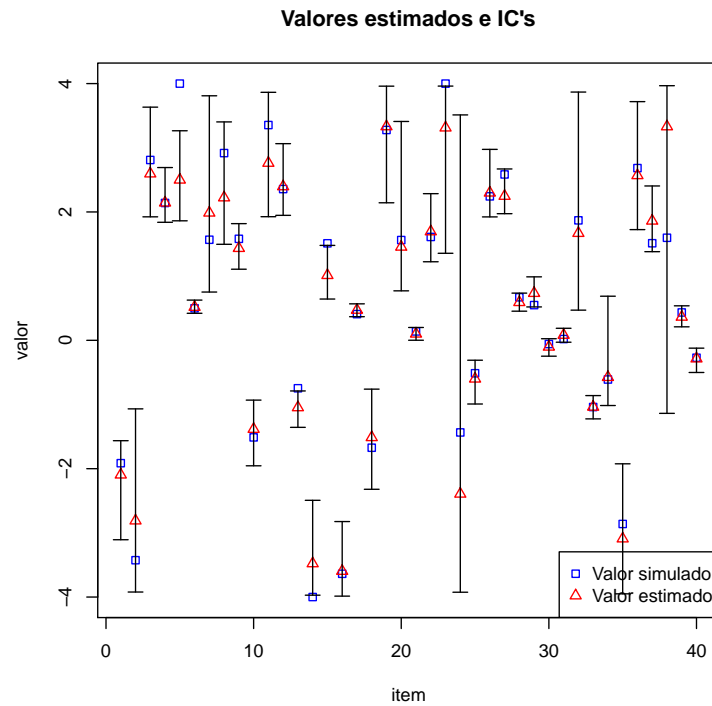


Figura 8.2: Valores simulados, estimados e IC's para o parâmetro b

8.3 Parâmetro c

O gráfico abaixo mostra que, para apenas dois itens (item 5 e item 13), o intervalo de credibilidade estimado não contém o verdadeiro valor do parâmetro. Ambos os itens também não conseguiram ter o parâmetro b estimado corretamente.

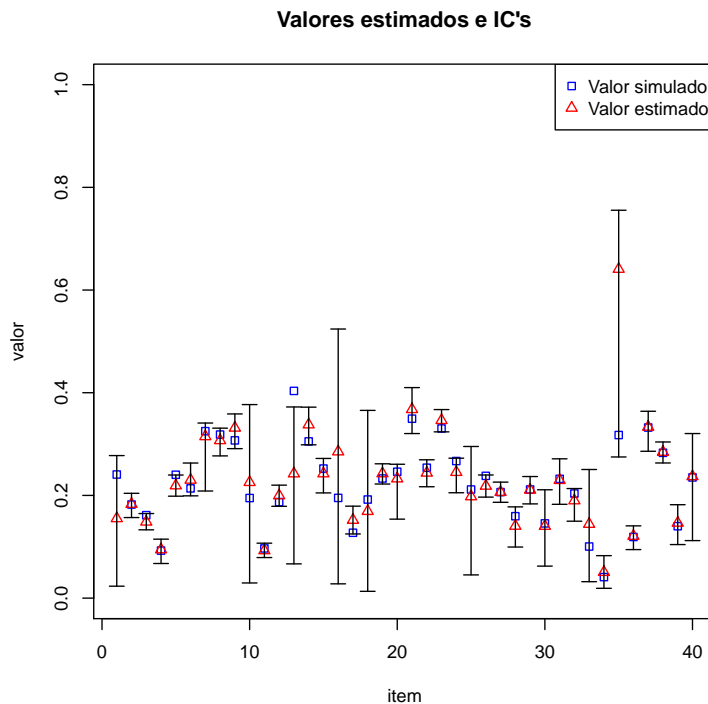


Figura 8.3: Valores simulados, estimados e IC's para o parâmetro c

8.4 Parâmetro δ_1

O gráfico abaixo revela que, para 3 itens, a estimativa do intervalo de credibilidade não contém o verdadeiro valor do parâmetro δ_1 . É importante notar que, nesses itens, o valor simulado para o parâmetro estava abaixo de -2.6 . Os valores gerados para a ansiedade seguem uma distribuição normal, com média 0 e desvio padrão igual a 1. Dessa forma, sabe-se que a probabilidade de terem sido gerado valores abaixo de -2.6 é muito pequena. Dessa forma, pode-se afirmar que esse item não afeta negativamente pessoas com baixa ansiedade. Em termos práticos, nesse caso, a CCI para pessoas com baixos níveis de ansiedade não sofreria depreciação, se assemelhando à CCI tradicional da TRI.

Valores muito baixos do parâmetro δ_1 podem ter pouca estimabilidade, uma vez que não existe informação suficiente nos dados para uma estimação mais precisa.

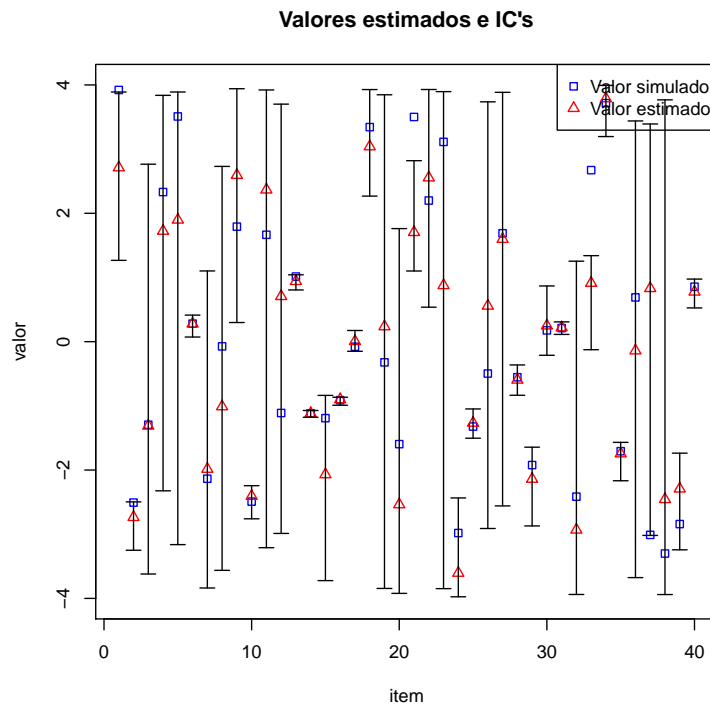


Figura 8.4: Valores simulados, estimados e IC's para o parâmetro δ_1

8.5 Parâmetro δ_2

O gráfico abaixo revela que, apenas para dois itens, o intervalo de credibilidade estimado não contém o verdadeiro valor do parâmetro. Em um desses itens, ocorre o que foi discutido na seção anterior. Nesse caso, o valor simulado para o parâmetro δ_2 é muito alto (acima de 3), resultando em baixa informação nos dados, e, portanto, estimação pouco precisa. O item 5, como observado anteriormente, também não teve os valores dos parâmetros b e c estimado corretamente. Sabe-se que os itens não são independentes e, portanto, a má estimação de um parâmetro pode afetar os demais.

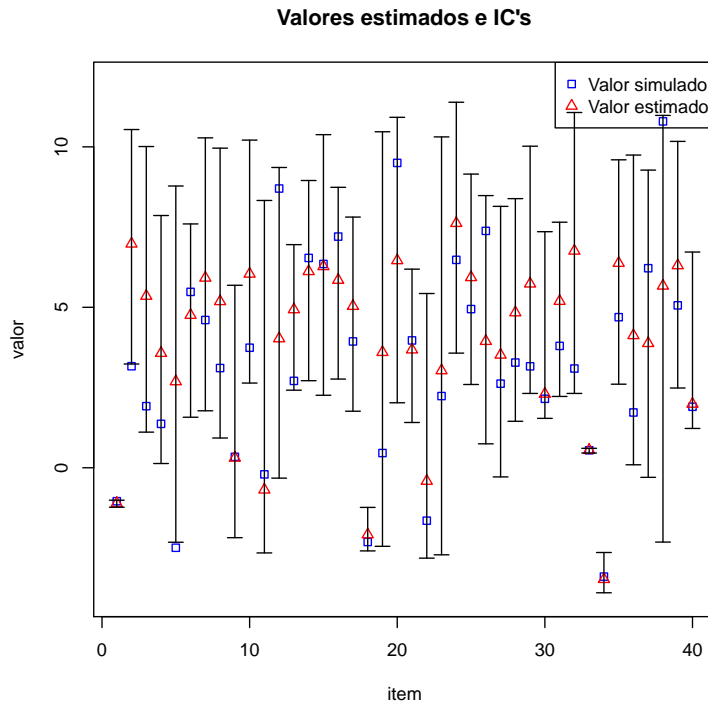


Figura 8.5: Valores simulados, estimados e IC's para o parâmetro δ_2

8.6 Parâmetro λ_1

O gráfico abaixo revela que para 7 itens, o intervalo de credibilidade estimado não contém o verdadeiro valor do parâmetro. Como pode ser observado nas tabelas do CD, para a grande maioria dos casos, isso ocorre quando o parâmetro δ_1 é muito baixo, acarretando o problema discutido anteriormente. Como o item não afeta pessoas com baixo/alto nível de ansiedade, não existe informação suficiente nos dados para a estimação do decaimento.

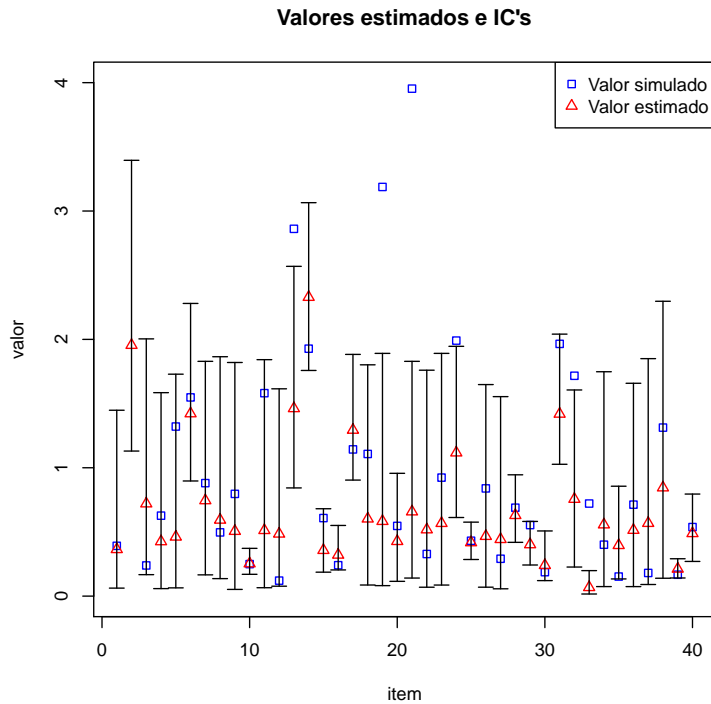


Figura 8.6: Valores simulados, estimados e IC's para o parâmetro λ_1

8.7 Parâmetro λ_2

O gráfico abaixo revela que, para 6 itens, o intervalo de credibilidade estimado não contém o verdadeiro valor do parâmetro. Novamente, como pode ser observado nas tabelas do CD, para a grande maioria dos casos, isso ocorre quando o valor gerado para δ_2 é muito alto, havendo dificuldade de estimação.

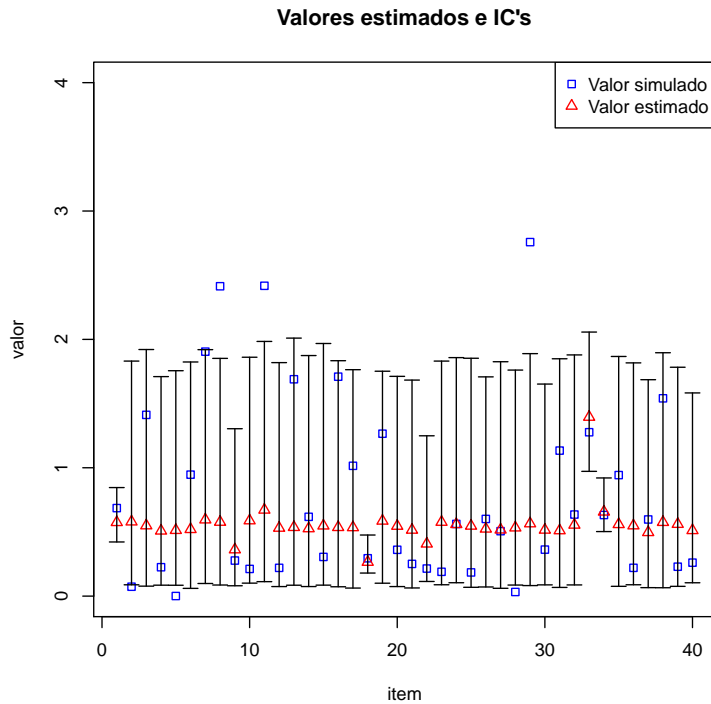


Figura 8.7: Valores simulados, estimados e IC's para o parâmetro λ_2

8.8 Correlação entre os valores estimados e simulados

A tabela abaixo revela as correlações entre os valores estimados e simulados. Alguns valores parecem ser menores do que esperados, devidos às observações feitas nas seções anteriores.

Tabela 8.1: Correlação entre os valores simulados e estimados

Parâmetro	Correlação	Significância
a	0,652	0,000
b	0,973	0,000
c	0,780	0,000
δ_1	0,879	0,000
δ_2	0,743	0,000
λ_1	0,633	0,000
λ_2	0,264	0,099

Capítulo 9

Resultados da base de dados reais (PROEB)

9.1 Introdução

Para a aplicação do modelo proposto na base do PROEB, construiu-se uma sintaxe para o *software* WinBUGS, que pode ser encontrada no apêndice *D*. Foram utilizadas as mesmas *prioris* descritas na seção anterior e, também, o modelo probito normal.

Foram rodadas 30000 iterações, tendo sido monitoradas as últimas 5000, com *thinning* de 5, obtendo-se, portanto, uma amostra composta por mil elementos. Novamente, a mediana da *posteriori* foi considerada como estimativa.

Uma análise aprofundada das estimativas de δ_1 revela que a grande maioria se encontra abaixo do valor -2.5 . Conforme foi discutido na seção 8.4, valores muito baixos de δ_1 revelam que pessoas com baixa ansiedade não são afetadas negativamente, uma vez que o indicador de ansiedade segue uma distribuição Normal, com média 0 e desvio padrão igual a 1, fazendo com que a probabilidade de que existam pessoas com nível de ansiedade abaixo de -2.5 seja muito pequena. Isso pode ser

observado no gráfico abaixo, construído no software R:

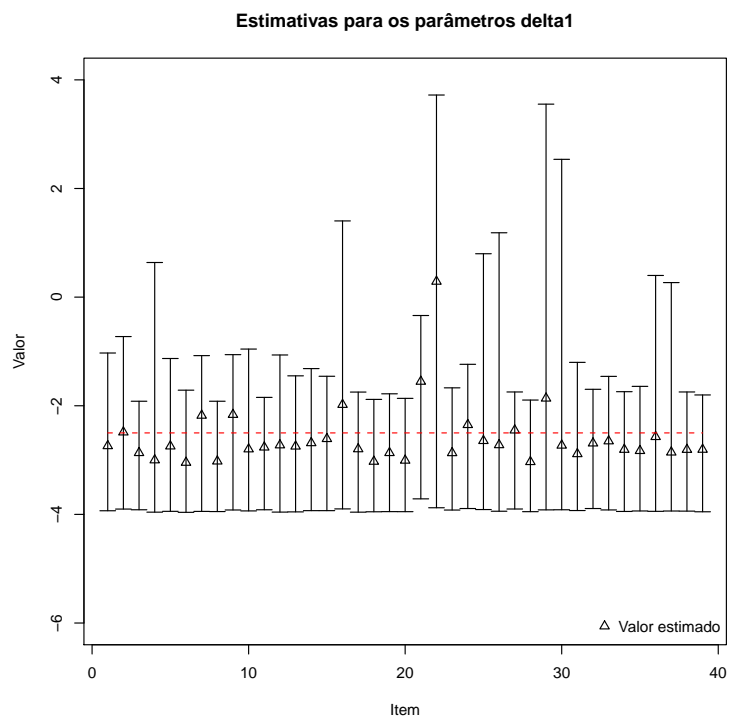


Figura 9.1: Valores estimados e IC's para o parâmetro δ_1

A análise das estimativas de δ_2 revela que a grande maioria se concentra abaixo de 2.5, fazendo com que a cauda direita do depreciador tenha início em valores baixos ou intermediários de ansiedade, revelando que, de fato, há redução da probabilidade de acerto. Isso pode ser observado no gráfico abaixo, construído no *software* R:

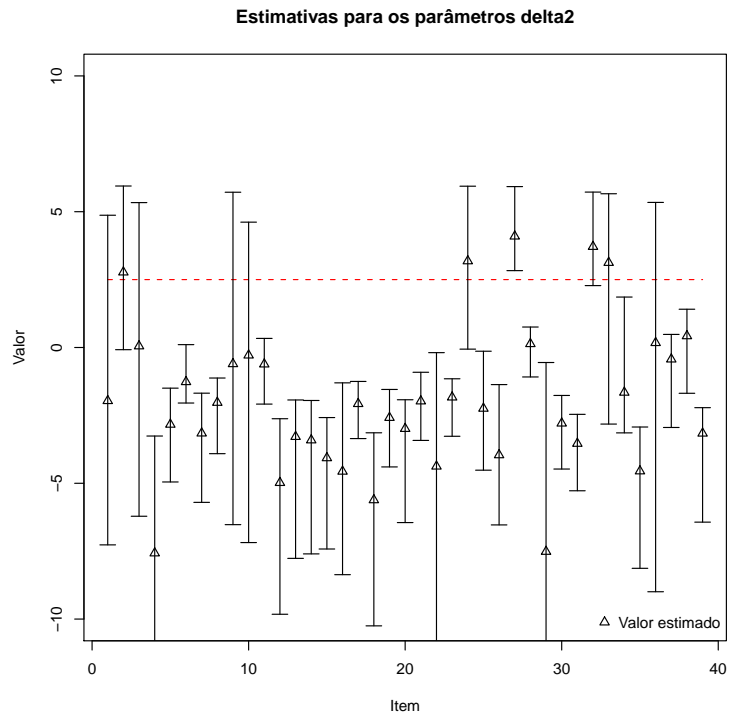


Figura 9.2: Valores estimados e IC's para o parâmetro δ_2

Em outras palavras, pode-se dizer que a cauda direita causa impacto no modelo, enquanto a esquerda, não. É importante notar que a avaliação do PROEB é uma avaliação em larga escala, que tem como objetivo avaliar as escolas públicas do estado de Minas Gerais. Como essa avaliação não tem impacto direto sobre o aluno, é concebível pensar que tal teste pode não causar muita ansiedade no aluno, podendo, inclusive, não penalizar pessoas com baixo nível de ansiedade.

Dessa forma, para esse caso específico, a fim de se manter a parcimônia, será considerado um depreciador unicaudal, que afeta apenas alunos com alto nível de ansiedade.

9.2 Modelo com depreciador unicaudal

O depreciador unicaudal possui uma estrutura mais simples do que o bicaudal, por diminuir a probabilidade de acerto apenas para alunos com alto nível de ansiedade. Dessa forma, a proficiência de alunos com baixo nível de ansiedade pode ser predita através do modelo tradicional da TRI. Sua forma gráfica assemelha-se à figura abaixo:

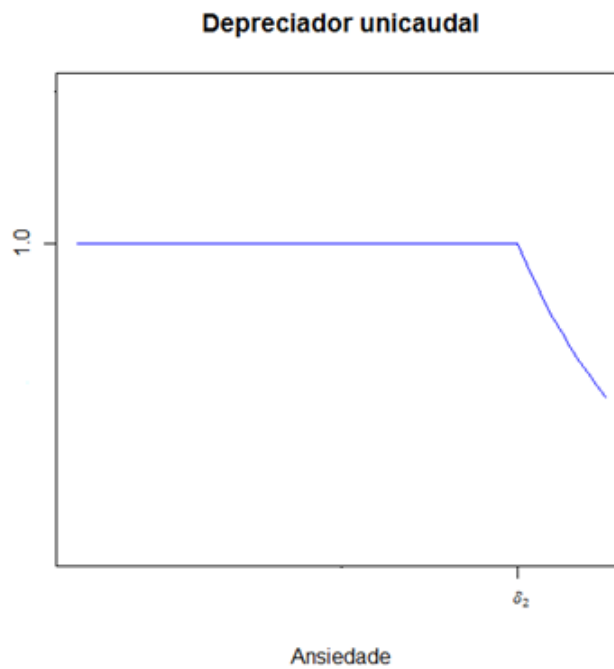


Figura 9.3: Depreciador unicaudal

Matematicamente, o depreciador unicaudal $\eta(\delta_2, \lambda_2, \varphi)$ pode ser escrito da seguinte forma:

$$\eta(\delta_2, \lambda_2, \varphi) = \begin{cases} 1 & \text{se } \varphi \leq \delta_2 \\ e^{-\lambda_2 * (\varphi - \delta_2)} & \text{se } \varphi > \delta_2 \end{cases} \quad (9.1)$$

As variáveis do depreciador unicaudal possuem a mesma interpretação das do depreciador bicaudal.

Dessa forma, o modelo de TRI incorporando tal depreciador pode ser escrito da seguinte forma:

$$P(U_{ij}|\theta_i, a, b, c) = \begin{cases} c_j + \frac{1-c_j}{1+\exp(-1.7a_j(\theta_i-b_j))} & \text{se } \varphi_i \leq \delta_{2j} \\ c_j + e^{-\lambda_{2j}*(\varphi_i-\delta_{2j})} * \frac{1-c_j}{1+\exp(-1.7a_j(\theta_i-b_j))} & \text{se } \varphi_i > \delta_{2j} \end{cases} \quad (9.2)$$

Foi, portanto, construída uma sintaxe para o *software* WinBUGS, encontrada no apêndice E. Foram rodadas 30.000 iterações, sendo monitoradas as últimas 5000 com *thinning* de 5, obtendo-se uma amostra de 1.000 elementos. As *prioris* foram as mesmas utilizadas para o modelo bicaudal, a mediana também foi a estatística de tendência central utilizada, assim como o modelo probito normal. As condições iniciais foram geradas aleatoriamente.

9.3 Resultados do modelo

9.3.1 Modelo tradicional da TRI

Para ser possível fazer uma comparação do modelo proposto com o modelo tradicional da TRI, foi construída uma sintaxe para o WinBUGS usando o modelo tradicional da TRI. As *prioris* para os parâmetros a, b e c foram as mesmas citadas para a sintaxe do modelo proposto, também foi considerado o modelo probito normal e as estimativas consideradas também foram as medianas a *posteriori*. O número de iterações também foi o mesmo: 30000 iterações, tendo sido monitoradas as últimas 5000, com *thinning* de 5, obtendo-se, portanto, uma amostra de 1000 iterações. As

condições iniciais foram geradas aleatoriamente. A sintaxe se encontra no apêndice *F*.

9.3.2 Comparação dos parâmetros

A inclusão do depreciador no modelo tradicional da TRI acaba afetando as estimativas dos parâmetros, uma vez que alteram, também, sua interpretação. Uma estimativa maior do parâmetro b de um item pelo modelo proposto, por exemplo, não implica um aumento de dificuldade desse item, já que a interpretação dos parâmetros também é modificada.

Abaixo, encontra-se uma tabela que estabelece a correlação entre os parâmetros a, b e c do modelo tradicional da TRI e do modelo proposto. Apesar de haver alteração nas estimativas, é possível perceber que ainda existe forte relação entre os mesmos.

Tabela 9.1: Correlação entre os valores estimados pelo modelo tradicional da TRI e pelo modelo proposto

Parâmetro	Correlação	Significância
a	0,816	0,000
b	0,992	0,000
c	0,890	0,000

É importante observar que o parâmetro c deveria ter se mantido constante, e, portanto, era esperado que sua correlação fosse igual 1. Apesar disso, esse é o parâmetro com pior estimabilidade, explicando a correlação de 0,890 encontrada.

Nas subseções abaixo, encontram-se representados os valores estimados para os parâmetros dos itens do PROEB 2009.

Parâmetro a

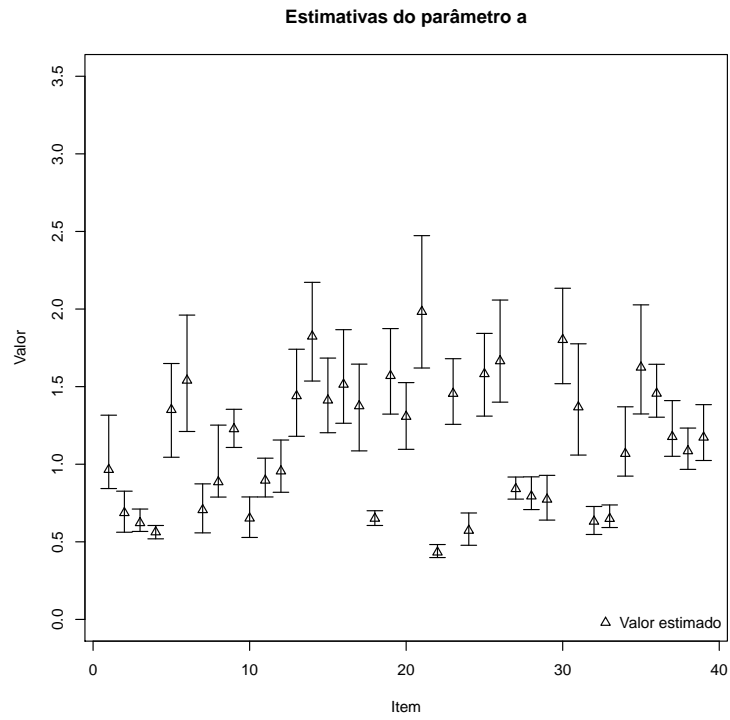


Figura 9.4: Valores estimados para o parâmetro a

Parâmetro b

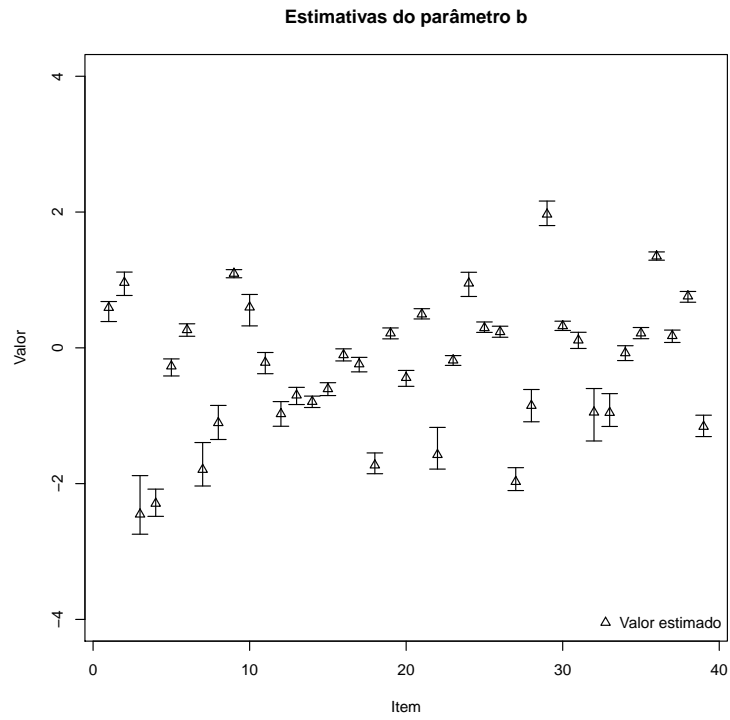


Figura 9.5: Valores estimados para o parâmetro b

Parâmetro c

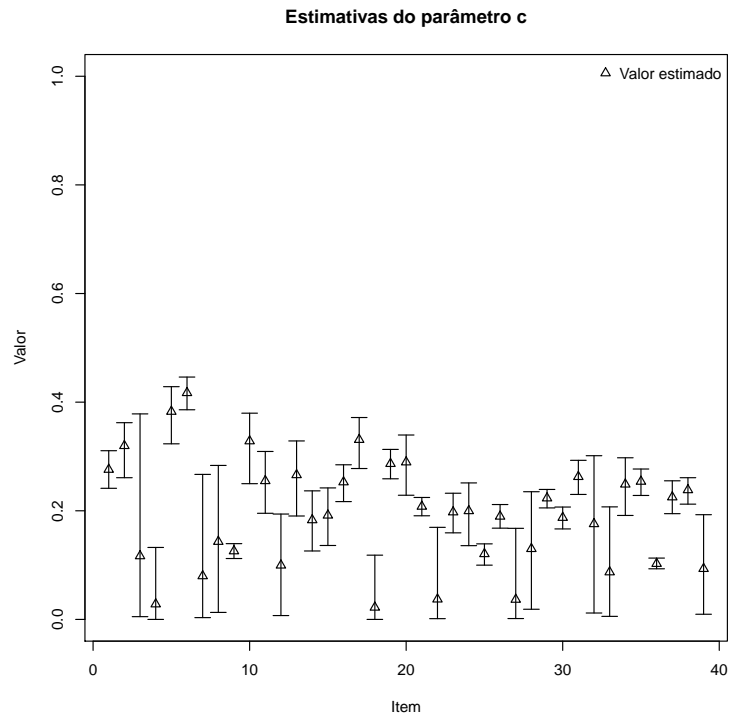


Figura 9.6: Valores estimados para o parâmetro c

Parâmetro δ_2

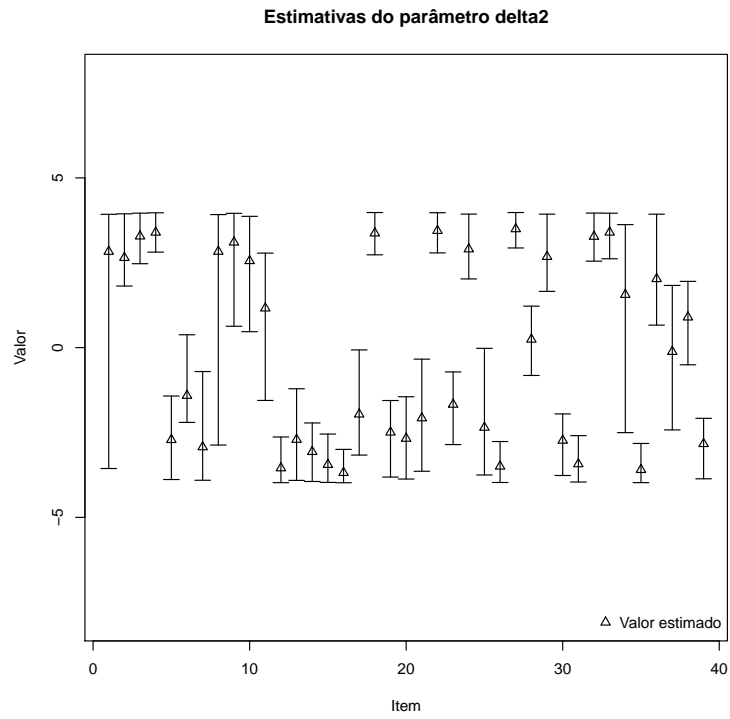


Figura 9.7: Valores estimados para o parâmetro δ_2

Parâmetro λ_2

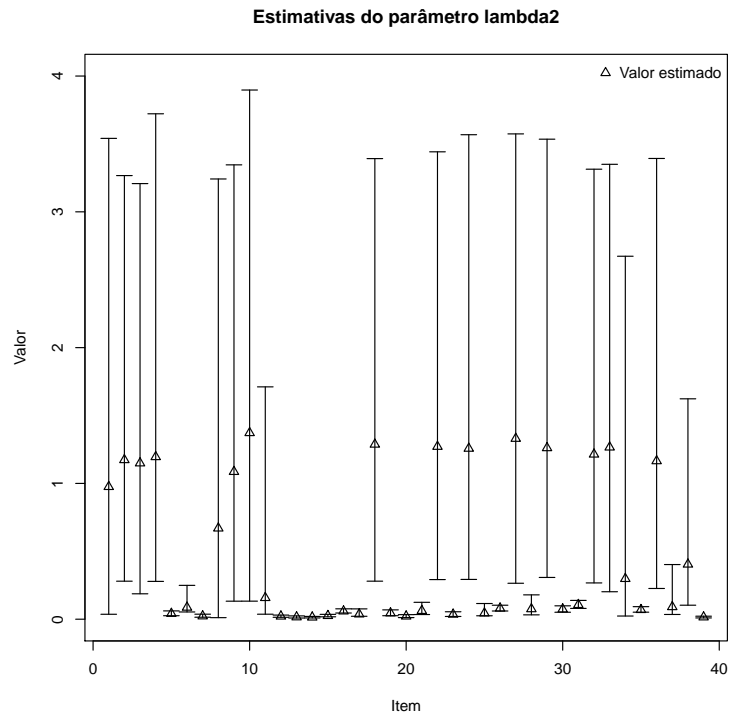


Figura 9.8: Valores estimados para o parâmetro λ_2

9.3.3 Comparação das proficiências

Como observado no capítulo 5, a inclusão do depreciador no modelo tem como objetivo a inclusão do efeito da ansiedade no modelo, permitindo com que fosse calculada a proficiência do aluno, descontando tal impacto.

Dessa forma, como, para a base do PROEB, foi utilizado o modelo com depreciador unicaudal, é esperado que a proficiência de pessoas com alto nível de ansiedade seja aumentada, e a proficiência das demais seja igual àquela predita pelo modelo tradicional da TRI.

Nas sintaxes construídas para o *software* WinBUGS para a estimação dos parâmetros, a proficiência tinha como distribuição *a priori* a normal padrão, fazendo

com que os valores estimados para a proficiência sejam “forçados” a apresentarem média 0 e desvio padrão igual a 1. Isso implica que as escalas de medidas foram diferentes para o modelo tradicional da TRI e para o modelo proposto nesse trabalho. Dessa forma, a fim de se obter comparabilidade entre as proficiências, é necessário equalizá-las, proporcionando-lhes a mesma escala.

Como o modelo utilizado para a base do PROEB é unicaudal, sabe-se que a proficiência dos alunos com nível baixo de ansiedade não deve ser afetada pelo depreciador, fazendo com que ela seja igual àquela predita pelo modelo tradicional de TRI.

Fazendo a hipótese de que os 50 % dos alunos cujos níveis de ansiedade são menores não possuem variação na sua estimativa de proficiência, podemos encontrar os coeficientes β_0 e β_1 tais que $\theta_D^T = \beta_0 + \beta_1 \theta_S$, em que θ_D corresponde ao valor da proficiência estimado pelo modelo proposto, θ_D^T corresponde ao valor da proficiência estimado pelo modelo proposto na escala da proficiência estimada pelo modelo tradicional e θ_S corresponde ao valor da proficiência estimada pelo modelo tradicional.

Aplicando uma regressão linear aos alunos cuja proficiência foram invariantes, encontra-se $\beta_0 = 0.034$ e $\beta_1 = 0.975$, de forma que:

$$\theta_D^T = \frac{0.034 + \theta_D}{0.975} \quad (9.3)$$

Aplicando tal transformação aos alunos que apresentaram os 50 % de níveis de ansiedade mais altos, teremos ambas as proficiências na mesma escala. A amostra é grande (n=9401 alunos), fazendo com que, pelo Teoria do Limite Central, possamos

assumir normalidade e aplicar o teste t-pareado para a comparação das médias. O teste foi aplicado utilizando o *software* SPSS, versão 15..

Tabela 9.2: Comparação entre as proficiências estimadas pelo modelo tradicional da TRI e o modelo proposto

Parâmetros	Média	Desvio padrão	t	Significância
$\theta_D^T - \theta_S$	0,0612463	0,1096443	38,295	0,000

Os resultados acima revelam que o valor de proficiência estimado pelo modelo da TRI que incorpora a ansiedade aumenta a proficiência dos alunos com alto nível de ansiedade. Comparando os 10 % de alunos com maior nível de ansiedade, a diferença é ainda maior, como pode-se observar na tabela abaixo, resultante do output do *software* SPSS, versão 15..

Tabela 9.3: Comparação entre as proficiências estimadas pelo modelo tradicional da TRI e o modelo proposto

Parâmetros	Média	Desvio padrão	t	Significância
$\theta_D^T - \theta_S$	0,1200175	0,1195002	21,773	0,000

Percebe-se, então, que o modelo proposto consegue, de fato, “descontar” o impacto negativo que altos níveis de ansiedade causam no aluno. A diferença, para o caso específico do PROEB, é relativamente baixa. Como já dito antes, isso provavelmente ocorre pelo fato do PROEB ser uma avaliação em larga escala que não possui impacto direto sobre o aluno. Além disso, o modelo proposto provavelmente causará maior diferença na proficiência para medidas de ansiedade mais precisas, que podem ser obtidos através de questionários com perguntas mais específicas do constructo “ansiedade face ao teste”.

Além disso, é interessante notar que esse modelo consegue detectar itens que

causam estresse. No CD anexo, encontram-se todos os parâmetros estimados para os itens do PROEB. Os itens que tiveram parâmetro δ_2 entre -3 e 1 indicam que níveis intermediários de ansiedade já causam depreciação da probabilidade do aluno acertar aquele item, enquanto itens que possuem parâmetro δ_2 acima de 2.5 indicam que o modelo apropriado para esse item é o modelo tradicional da TRI, por motivos já discutidos nas seções anteriores.

Capítulo 10

Ajuste do modelo proposto

O *software* WinBUGS, juntamente com as estimativas dos parâmetros, calcula o deviance do modelo. A estatística *deviance* é utilizada para verificação do ajuste do modelo. Em geral, o valor dessa estatística de um modelo proposto é comparada com o valor do *deviance* do modelo nulo, a fim de determinar qual modelo tem melhor ajuste. É calculada fazendo-se: $2 \cdot \log(\text{verossimilhança})$.

Nesse trabalho, essa estatística será usada para comparar o modelo proposto e o modelo tradicional da TRI. Abaixo, na tabela, se encontra o valor estimado para a estatística *deviance* e os limites inferiores e superiores do intervalo de credibilidade de 95 %.

Tabela 10.1: Comparação entre as proficiências estimadas pelo modelo tradicional da TRI e o modelo proposto

Modelo	Limite Inferior	Deviance estimada	Limite superior
<i>Tradicional</i>	368100	368400	368700
<i>Proposto</i>	367600	367900	368200

Assim, a diferença entre ambos os *deviances* segue uma distribuição χ^2 , com 2 graus de liberdade (diferença de quantidade de parâmetros entre os modelos). Sabe-se que, para 2 graus de liberdade e considerando $\alpha=0.05$, a diferença entre os dois

(500) é significativa. Se a diferença dos deviances for significativa, o modelo com melhor deviance é considerado melhor (Kreft De Leeuw *apud* Laros et al. (2010)).

Como o modelo proposto apresenta menor *deviance*, o modelo de TRI incorporando a ansiedade possui melhor ajuste.

Abaixo, se encontram dois pares de gráficos, comparando o ajuste do modelo proposto e o do modelo tradicional da TRI. É possível perceber, principalmente para esses dois itens (que causam ansiedade nos alunos) que há um melhor ajuste pelo modelo proposto.

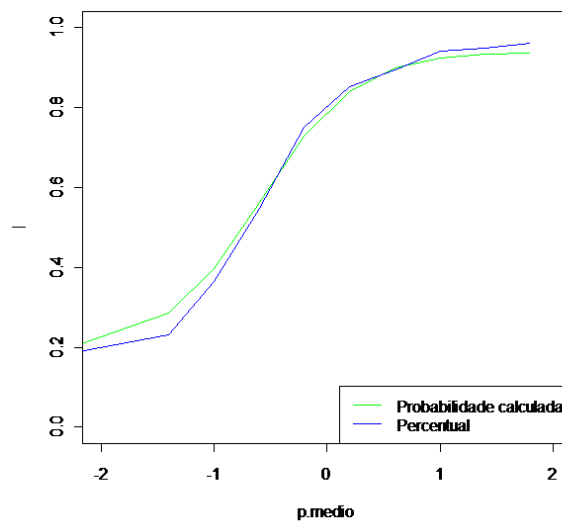


Figura 10.1: Ajuste do item 15 pelo modelo proposto

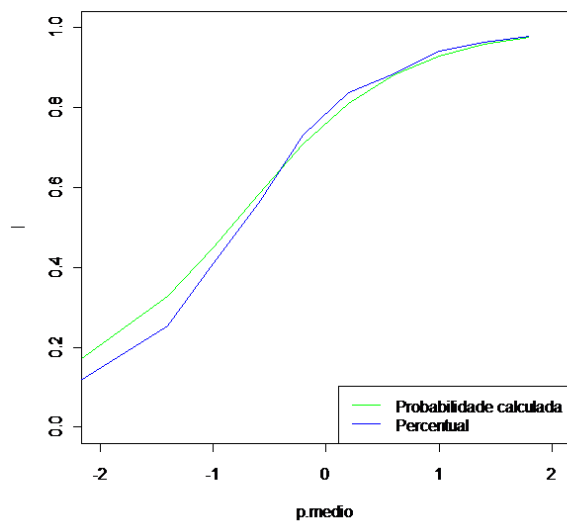


Figura 10.2: Ajuste do item 15 pelo modelo tradicional da TRI

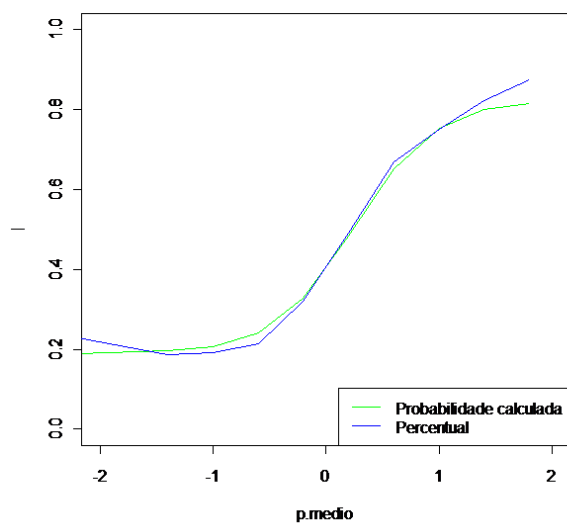


Figura 10.3: Ajuste do item 26 pelo modelo proposto

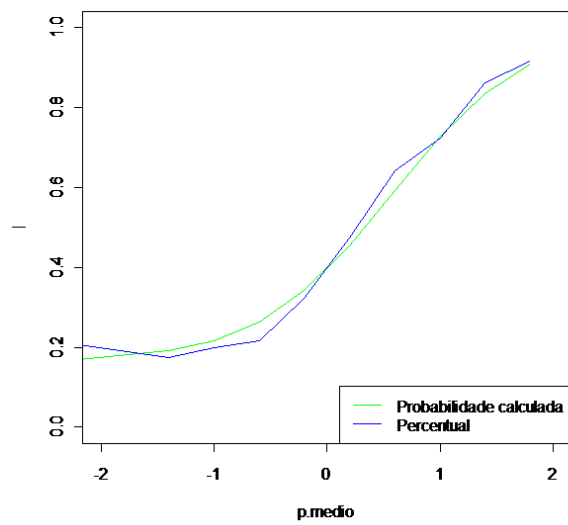


Figura 10.4: Ajuste do item 26 pelo modelo tradicional da TRI

Os gráficos que revelam o ajuste do modelo proposto e do modelo tradicional da TRI aos 39 itens do PROEB encontram-se no CD anexo.

Parte IV

CONCLUSÃO E CONSIDERAÇÕES FINAIS

Esse trabalho tornou possível a incorporação do impacto de um fator emocional na predição da probabilidade de acerto de um item por um aluno.

O modelo de TRI que incorpora a ansiedade possui um melhor ajuste, uma vez que considera mais informações acerca do aluno. No futuro, com a possível criação de novos modelos que incorporem diversos fatores que afetam o aluno no momento da realização do teste, será possível obter um modelo mais próximo da realidade.

O modelo proposto nesse trabalho permitiu a obtenção da proficiência “purificada”, retirando o efeito da ansiedade da probabilidade de acerto do aluno, além de permitir a identificação de itens que causam maior ansiedade. Para se descobrir o motivo pelo qual alguns itens causam mais ansiedade, é necessário uma análise mais profunda, feita por especialistas da área, acerca do item.

A fim de tornar o modelo e a análise mais precisas, é interessante a elaboração de um questionário que possui perguntas mais específicas do constructo “ansiedade face ao teste”, obtendo, portanto, medidas mais precisas. Além disso, é interessante aplicar tal modelo em testes que causam maior ansiedade nos alunos. Como dito anteriormente, o PROEB é uma avaliação em larga escala, e, portanto, não tem como objetivo avaliar o desempenho individual do aluno.

Em termos de modelagem estatística, o modelo pode se tornar mais preciso com a inclusão de uma variável que represente a atuação do depreciador. Em outras palavras, ao invés de se identificar itens que causam ansiedade através dos valores estimados para δ_1 e δ_2 , seria possível fazê-lo através de uma variável dicotômica (que assume o valor 1 se o item causa ansiedade e 0 caso o contrário). Além disso, em

termos de equalização das proficiências, seria interessante, de início, determinar um grupo cuja proficiência seja invariante ao modelo.

O modelo proposto nesse trabalho conseguiu cumprir seu propósito e, certamente, será capaz de abrir caminho para diversos outros modelos que incorporem fatores que afetam alunos no momento de realização de um teste, tornando, cada vez mais possível, a obtenção da verdadeira proficiência do aluno.

Referências Bibliográficas

- Birbaum, A. (1968). Some latent traits models and their use in inferring an examinee's ability. *Statistical Theories of Mental Test Scores*.
- Box, G. & Tiao, G. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.
- Costa, P. D. M. (2005). *Modelos Multidimensionais de Resposta ao Item*.
- de Andrade, D. F., Tavares, H. R., & Valle, R. C. (2010). Tri: Conceitos e aplicações - sinape.
- de Boeck, P. & et al (2006). A local-influence-based diagnostic approach to a speeded item response theory model. *Journal of the Royal Statistics Society*.
- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Sage Publications.
- Laros, J., Marciano, J. L., & de Andrade, J. (2010). Fatores que afetam o desempenho na prova de matemática do seab: Um estudo multinível. *Avaliação Psicológica*.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*.
- Neto, J. H. V. (2010). Inferência bayesiana.
- Rasch, G. (1960). *Probabilistic Modelos for some intelligence and attainment tests*. Institute for Educational Research.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. Springer.

- Rosário, P., Soares, S., & et al (2004). Ansiedade face aos testes e auto-regulação da aprendizagem : variáveis emocionais no aprender. *Psicologia e Educação*.
- Samejima, F. (2000). Logistic positive exponent family of models: virtue of asymmetric item characteristic curves. *Psychometrika*.
- Soares, T. M. (2005). Utilização da teoria de resposta ao item na produção de indicadores sócio econômicos. *Pesquisa Operacional*.
- Soares, T. M., Gonçalves, F., & Gamerman, D. (2009). An integrated bayesian model for dif analysis. *Journal of Educational and Behavioral Statistics*.

Apêndice A

Sintaxe para geração dos dados simulados no MATLAB

```
\% geração de dados para o programa de ansiedade
a=lognrnd(0.5,0.4,40,1);
a=diag(a*(a>0.5)')+ (a<0.5)*0.5;
a=diag(a*(a<3.5)')+ (a>3.5)*3.5;
b=normrnd(0,2,40,1);
b=diag(b*(b>-4)')+ (b<-4)*(-4);
b=diag(b*(b<4)')+ (b>4)*(4);
c=betarnd(5,17,40,1);
c=diag(c*(c<0.5)')+ (c>0.5)*0.5;
ansii=normrnd(0,1,2000,1);
prof=normrnd(0,1,2000,1);
p1=ones(2000,40)./(ones(2000,40)+exp(-1.7*repmat(a',2000,1).*(repmat(prof,1,40)-repmat(b',2000,1))));
ini=random('unif',-4,4,40,1);
lambda=random('gam',1,1,40,1);
mat=ones(2000,1);
ini=ini*mat;
lambda=lambda*mat;
p2= (repmat(ansi,1,40) < ini) + ( exp(-lambda.*(repmat(ansi,1,40)-ini) ) .* (repmat(ansi,1,40) >= ini) );
p= repmat(c',2000,1) + (ones(2000,40)-repmat(c',2000,1)).*p2.*p1;
dadosrest=binornd(1,p);
save resultado
save dados dados -ascii
```


Apêndice B

Questionário contextual completo PROEB 2009

**QUESTIONÁRIO DO ALUNO — 5º ano — ENSINO FUNDAMENTAL
MATEMÁTICA**

Caro(a) aluno(a),
Este questionário vai ajudar a melhorar sua vida escolar. Para que isso aconteça, precisamos de sua colaboração, respondendo, com cuidado e sinceridade, as perguntas abaixo. Por favor, marque só uma resposta para cada questão.

Marque o quanto você concorda com as afirmações abaixo.

1. Eu me sinto chateado quando alguém faz alguma coisa que não gosto.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
2. Eu queria ser feliz como os meus amigos.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
3. Eu fico chateado quando o professor me chama à atenção.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
4. Eu não tenho confiança nas coisas que faço.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
5. Eu me canso com qualquer coisa.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
6. Eu fico preocupado com coisas que não são importantes.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
7. Eu fico nervoso porque tenho problemas.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
8. Eu tenho dificuldade de decidir o que vou fazer.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
9. Eu consigo copiar a matéria com facilidade.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
10. Eu consigo ler com facilidade.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
11. Eu serei um bom aluno na escola.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
12. Eu tenho dificuldades para escrever as palavras que são ditadas.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
13. Eu posso sempre me sair bem na escola.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
14. Eu sei que estudando me sairei bem na vida.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
15. Eu esqueço rápido o que aprendi.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
16. Evito ficar pensando muito para realizar uma atividade.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
17. Só faço tarefas escolares quando é pedido pelo professor.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
18. Eu gosto de ir à escola.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
19. Quero ser o melhor aluno da sala.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
20. Só estudo quando sei que vai haver prova.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
21. Gosto de tarefas difíceis.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
22. Leio apenas o que o professor pede em sala de aula.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
23. Não consigo curtir nenhuma atividade escolar.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
24. Quanto mais difícil a tarefa, mais eu tento.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
25. Não me importa com o que pensam de mim.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
26. Os pais devem exigir o melhor desempenho do filho na escola.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
27. Os pais devem exigir somente aquilo que o filho é capaz de fazer.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
28. Os pais não devem se preocupar com a vida escolar do filho.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
29. Não faz mal os alunos conversarem um pouco durante as aulas.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
30. A conversa em sala de aula pode atrapalhar o professor a dar a matéria.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
31. O professor é exigente quando pede silêncio em sala.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
32. A minha escola oferece o melhor ensino que ela pode.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
33. A escola deve melhorar o ensino oferecido.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo
34. O ensino oferecido na escola me dá boas chances na vida.	1-Discordo	2-Discordo mais que concordo	3-Concordo mais que discordo	4-Concordo

Figura B.1: Questionário contextual

Apêndice F

Sintaxe montada para o *software* WinBUGS com o modelo tradicional para o PROEB

```
model
{
  for( i in 1 : N ) {
    theta[i] ~ dnorm(0,1)I(-4,4)
  }
  for( j in 1:K) {
    b[j] ~ dunif(-4,4)
    a[j] ~ dunif(0,3.5)
    c[j] ~ dunif(0,1)
  }
  for( i in 1 : N) {
    for( j in 1:K) {
      Y[i,j] ~ dbern(p[i,j])
      p[i,j] <- c[j]*(1-c[j])*phi(a[j]*(theta[i]-b[j]))
    }
  }
}
list(N=9401,K=39,
      Y = structure(
        .Data = c(1,0,1,1,1,0,1,1,0,1,1,1,0,0,0,0,0,0,0,1,0,1,0,1,0,0,0,1,1,1,0,0,0,0,1,
        1,1,1,1,0,1,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,0,1,1,0,0,1,1,
        1,0,1,1,0,0,0,1,1,0,0,0,0,0,1,0,0,1,1,0,0,1,0,1,0,0,1,0,0,1,1,0,1,1,1,0,0,1,0,1,
        1,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,
        .
        .
        0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,1,0,0,1,1,1,1,0,1,1,1,0,0,1,1,1,
        1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,0,1,1,1,1,1,1,1),
      .Dim = c(9401, 39))
```