

UM MODELO DE TEORIA DE RESPOSTA AO
ITEM PARA OS DADOS DO VESTIBULAR
2006-2 DA UFLA

MARIA DE LOURDES LIMA BRAGION

2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

MARIA DE LOURDES LIMA BRAGION

UM MODELO DE TEORIA DE RESPOSTA AO ITEM PARA
OS DADOS DO VESTIBULAR 2006-2 DA UFLA

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação e Agronomia, área de concentração e Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

Orientador

Prof. Dr. Júlio Sílvio de Sousa
Bueno Filho

LAVRAS
MINAS GERAIS-BRASIL
2007

**Ficha Catalográfica Preparada pela Divisão de Processos
Técnicos da Biblioteca Central da UFLA**

Bragion, Maria de Lourdes Lima.

Um modelo de Teoria de Resposta ao Item para os dados do Vestibular 2006-2 da UFLA / Maria de Lourdes Lima Bragion. - Lavras: UFLA, 2007.

88 p. : il.

Orientador: Júlio Sílvio de Sousa Bueno Filho.

Dissertação (Mestrado) - UFLA.

Bibliografia.

1. Teoria de Resposta ao Item. 2. Inferência Bayesiana.

I. Universidade Federal de Lavras. II. Título.

CDD - 519.54

MARIA DE LOURDES LIMA BRAGION

UM MODELO DE TEORIA DE RESPOSTA AO ITEM PARA
OS DADOS DO VESTIBULAR 2006-2 DA UFLA

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

APROVADA em 15 de fevereiro de 2007.

Prof^a. Dr^a. Rosângela Helena Loschi UFMG

Prof. Dr. Lucas Monteiro Chaves UFLA

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

UFLA

(Orientador)

LAVRAS

MINAS GERAIS-BRASIL

Em primeiro lugar, **louvo e agradeco** a Deus, pela oportunidade que me deu de realizar e concluir mais esta importante etapa de minha vida.

Dedico esta vitória:

Ao meu grande amor, Nivaldo, pelo apoio, compreensão e compreensão, nos momentos de maior necessidade.

Aos meus filhos, Daniel e Mariana, pela confiança que depositaram em mim.

Ao meu pai, que sei que teria muito orgulho se pudesse estar aqui.

A minha mãe, pela educação que me deu, a qual muito contribuiu para que eu chegasse onde cheguei.

AGRADECIMENTOS

A Deus, por ter me concedido saúde e força, tornando possível concluir este trabalho.

Ao meu pai, Nivaldo, por ter tido paciência, suportando comigo as agústias e espinhos que passei nesta caminhada. Aos meus filhos, Daniel e Mariana, por compreenderem as muitas vezes em que estive ausente, devido aos compromissos de meus estudos. Daniel, meu filho, te agradeço muito por sua grande ajuda na parte computacional.

Agradeço aos meus pais, porque foi por meio deles que pude começar a percorrer a jornada da vida e pude chegar onde cheguei.

Ao meu orientador, Júlio Sílvio de Sousa Bueno Filho, pela sua honestidade, dedicação e esforço em me ajudar no desenvolvimento deste trabalho.

A todos os professores e funcionários do Departamento de Ciências Exatas.

A todos os colegas do mestrado e do doutorado em Estatística, especialmente à Patrícia e Elisa, pela convivência agradável e por terem sido minhas grandes companheiras de estudo.

A FAPEMIG, por me conceder uma bolsa durante todo esse período, o que tornou possível minha dedicação a este trabalho.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	iii
RESUMO	vi
ABSTRACT	vii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 A Curva Característica do Item (CCI)	3
2.2 Modelos para Teoria de Resposta ao Item	9
2.2.1 Modelo logístico de 1 parâmetro	10
2.2.2 Modelo logístico de 2 parâmetros	11
2.2.3 Modelo logístico de 3 parâmetros	11
2.3 Propriedade da TRI	13
2.3.1 Ilustrando a propriedade de invariância do item	13
2.4 Métodos de estimação	15
2.5 Inferência bayesiana	16
2.5.1 Monte Carlo via cadeias de Markov (MCMC)	19
2.5.2 Algoritmo Metropolis & Hastings	20
3 MATERIAIS E MÉTODOS	22
3.1 Materiais	22
3.2 Metodologia	23
3.3 Distribuição a Posteriori	26
3.3.1 Modelo logístico de 3 parâmetros	27
3.3.2 Implementação do processo amostral	28
3.3.3 Exemplo ilustrativo da implementação do algoritmo	29
4 RESULTADOS E DISCUSSÃO	34

4.1	Intervalos de máxima densidade de probabilidade (H.P.D.) para os parâmetros dos itens	34
4.2	Condições gerais sobre os parâmetros dos itens e ajuste do modelo	48
4.3	Condições específicas sobre os parâmetros dos itens	55
4.3.1	Considerações para a disciplina de Biologia e Matemática	55
4.3.2	Análise do curso de Medicina Veterinária.	68
5	CONCLUSÕES	78
6	REFERÊNCIAS BIBLIOGRÁFICAS	79
7	ANEXOS	82

LISTA DE TABELAS

1	Intervalos com as respectivas classificações e valores esperados na priori para o parâmetro a	27
2	Intervalos com as respectivas classificações e valores esperados na priori para o parâmetro b	28
3	Valores estimados dos parâmetros a (discriminação do item) e b (dificuldade do item) para o modelo de 2 parâmetros e de b para o modelo de 1 parâmetro (Baker, 2001).	33
4	Escores brutos com parâmetro com os valores estimados das habilidades (θ) para o modelo de 2 parâmetros e 1 parâmetro (Baker, 2001).	33
5	Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>) para o parâmetro c e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.	35
6	Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>) para o parâmetro c e respectivo erro de Monte Carlo, para o curso de Agronomia.	37
7	Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>) para o parâmetro c e respectivo erro de Monte Carlo, para o curso de Engenharia Agrícola.	39
8	Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>) para o parâmetro a e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.	42

9	Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>) para o parâmetro b e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.	45
10	Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>) para o parâmetro θ dos 30 primeiros alunos e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.	47
11	Correlação entre as estimativas médias dos parâmetros a , b e c para os cursos do Vestibular 2006-2 da UFLA.	48
12	Correlação da nota do aluno \times habilidade média estimada	50
13	Média dos parâmetros discriminação (a), dificuldade (b) e acerto casual (c), por disciplina.	50
14	Estimativas a <i>posteriori</i> por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - <i>highest probability density</i>), para os parâmetros a e b de alguns itens do Curso de Agronomia e respectivo erro de Monte Carlo.	72
15	Estimativas a <i>posteriori</i> por ponto e por intervalo (I.C.: intervalo de credibilidade, HPD: <i>highest probability density</i>), para o parâmetro θ de alguns candidatos do curso de Medicina Veterinária e erro de Monte Carlo.	76
16	Vetor resposta de um teste de 10 itens e 16 examinados.	83

LISTA DE FIGURAS

1	Curva característica do item - CCI	4
2	CCI com diferentes valores para b (parâmetro dificuldade do item) e esse o valor de a (parâmetro discriminação do item).	5
3	CCI com diferentes valores para a (parâmetro discriminação do item) e esse o valor de b (parâmetro dificuldade do item).	6
4	CCI Ideal.	8
5	CCI com diferentes valores para c (parâmetro da probabilidade de acerto casual) e esses os valores de a (parâmetro discriminação) e b (parâmetro dificuldade do item).	9
6	CCI para as proporções de respostas corretas para um grupo de examinados com habilidade entre -3 a -1.	14
7	CCI para as proporções de respostas corretas para um grupo de examinados com habilidade entre 1 a 3.	14
8	CCI para as proporções de respostas corretas para um grupo de examinados com habilidade entre -3 a 3.	15
9	Poder de discriminação a para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	52
10	Grau de dificuldade b para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	53
11	Probabilidade de acerto casual c para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	54

12	Questão 38 do Vestibular 2006-2 da UFLA, correspondente ao item 30 e estudo.	56
13	Poder de discriminação (a) da prova de Biologia para os cursos de 1 (Admissão), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	58
14	Grau de dificuldade (b) da prova de Biologia para os cursos de 1 (Admissão), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	59
15	Probabilidade de acerto casual (c) da prova de Biologia para os cursos de 1 (Admissão), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	60
16	Questão 52 do Vestibular 2006-2 da UFLA, correspondente ao item 44 e estudo.	61
17	Questão 57 do Vestibular 2006-2 da UFLA, correspondente ao item 49 e estudo.	63
18	Poder de discriminação (a) da prova de Matemática para os cursos de 1 (Admissão), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	65
19	Grau de dificuldade (b) da prova de Matemática para os cursos de 1 (Admissão), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	66

20	Probabilidade de acerto casual (c) da prova de Matemática para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).	67
21	Poder de discriminação a para o curso de Medicina Veterinária.	69
22	Grau de dificuldade b para o curso de Medicina Veterinária. .	70
23	Probabilidade de acerto casual c para o curso de Medicina Veterinária.	71
24	Representação gráfica das cadeias geradas pelo algoritmo Metropolis-Hastings e da densidade a <i>posteriori</i> dos parâmetros a , b e c , para o item 58.	73
25	CCI dos itens 30, 49 e 58 para o Curso de Medicina Veterinária.	74
26	Representação gráfica das cadeias geradas pelo Metropolis-Hastings e da densidade a <i>posteriori</i> do parâmetro θ dos candidatos 19, 25 e 26.	75
27	Gráfico da correlação entre a nota do vestibulando e sua respectiva habilidade estimada.	76

RESUMO

BRAGION, Maria de Lourdes Lima. **Um modelo de teoria de resposta ao item para os dados do vestibular 2006-2 da UFPA.** 2007. 88 p. Dissertação (Mestrado em Agronomia / Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.

Neste trabalho ajustou-se um modelo de Teoria de Resposta ao Item (TRI) aos dados do Vestibular 2006-2 da universidade Federal de Lavras (UFLA). O objetivo principal deste estudo foi investigar a qualidade das questões quanto ao seu grau de dificuldade e poder de discriminação. Adotou-se o modelo logístico de três parâmetros: poder de discriminação (**a**), grau de dificuldade (**b**) e probabilidade de acerto casual (**c**), além da habilidade individual do candidato (θ). Os ajustes deste modelo aos dados de cada curso (considerados populações distintas de candidatos) foram feitos separadamente, via inferência bayesiana, usando o algoritmo Metropolis & Hastings. As habilidades individuais apresentaram alta correlação com as notas observadas. Houve baixo índice de acerto casual. As questões mais difíceis, e mais variadas, foram de Matemática, seguidas de Física, Química, História, Biologia e Português. Geografia e Filosofia apresentaram um baixo nível de dificuldade. Quanto ao poder de discriminação, as questões de Biologia apresentaram o melhor resultado, com ótima discriminação. Boa discriminação foi obtida pelas questões de Química, Física e Matemática. As demais não trouxeram uma contribuição relevante. A título de ilustração, uma breve discussão das provas para candidatos de um curso é apresentada. O modelo proposto de TRI mostrou-se importante ferramenta para entender e intervir na qualidade do vestibular da UFLA.

Palavras-chaves: algoritmo Metropolis-Hastings, Monte Carlo via Cadeia de Markov, inferência bayesiana, teoria de resposta ao item.

***Comitê Orientador:** Júlio Sílvio de Sousa Bueno Filho - UFLA. (Orientador)

ABSTRACT

BRAGION, Maria de Lourdes Lima. **An Item Response Theory Model for UFLA's 2006-2 admission examn.** 2007. 88 p. Dissertation (Master in Agronomy /Statistics and

1 INTRODUÇÃO

Selecionar indivíduos que possuam maiores habilidades é o principal objetivo dos exames vestibulares. Porém, não se pode medir tais habilidades das pessoas a partir do seu estatura ou seu peso. O instrumento de medida que tem sido tradicionalmente usado para esse fim é o resultado de uma prova, no qual são avaliados seus escores brutos. Todas as análises e interpretações estão baseadas nessa prova como um todo. Isso faz com que não seja possível comparar indivíduos que não foram submetidos à mesma prova.

A teoria de resposta ao item (TRI) está baseada na avaliação de cada um dos itens, e não dos escores brutos da prova. Esta teoria é definida como um conjunto de modelos para a probabilidade de uma pessoa dar uma resposta correta a um determinado item, e é função de sua habilidade e de características do item. Essa habilidade é inerente a cada indivíduo e não depende da prova a ele aplicada, assim como os parâmetros do item também possuem características próprias e não dependentes de quem responde. Essa é uma propriedade muito importante da TRI: a invariância de seus parâmetros em uma determinada população (Baker, 2001). Sendo assim, a TRI possibilita usar ferramentas de estimação de habilidades pessoais e características dos itens, tendo um enorme impacto potencial no planejamento e avaliação de processos de ensino-aprendizagem e exames de seleção dos mais variados tipos.

Várias questões de interesse prático na área da Educação também têm sido respondidas com o auxílio da TRI, como por exemplo, comparar o desempenho de alunos de escolas públicas e privadas ou avaliar o desenvolvimento de uma série de um ano para outro. Algumas dessas aplicações da TRI e avaliações educacionais brasileiras podem ser encontradas em Valle (1999).

Vários modelos de TRI já foram propostos na literatura. Os trabalhos pioneiros foram os de Lord (1952) e de Rasch (1960), os quais foram os primeiros autores a propor modelos para ítemos para o cálculo de tais

probabilidades no contexto da psicometria. Inicialmente, o modelo estatístico utilizado foi o da distribuição normal acumulada, porém, Birnbau (1968), propôs modelar a probabilidade acumulada com a função logística, pelo fato de ela possuir os parâmetros do item e da habilidade de forma explícita, tornando mais fácil sua visualização e, também, por não envolver integração. Desde então, essa teoria tem tido notável avanço.

O objetivo deste trabalho é aplicar os conceitos da TRI para os dados do Vestibular 2006-2 da UFLA, estimar as habilidades dos vestibulandos e comparar com seus escores brutos e estimar os parâmetros de seus itens, avaliando quais questões cumpriram melhor o objetivo a que se propõe, qual seja, discriminar adequadamente os melhores candidatos. Com a identificação de tais itens, espera-se estar contribuindo para a melhoria do seu processo seletivo. Para tanto, estudaremos os cursos e as provas, sob o ponto de vista da probabilidade de acerto casual, do grau de dificuldade e do poder de discriminação do item.

No próximo capítulo, será apresentado um breve referencial, e que os principais conceitos e modelos sobre esta teoria são enunciados, fornecendo um exemplo para ilustrar a propriedade de invariância do item. As ideias básicas sobre inferência bayesiana e o método MCMC também são apresentadas. No capítulo 3, são descritos os dados do vestibular 2006-2 da UFLA e a metodologia adotada. O capítulo 4 é dedicado a apresentação dos resultados alcançados e a discussões dos mesmos e, no capítulo 5, relacionam-se as conclusões finais. Por último, são citadas as referências bibliográficas que constam no texto. Dados do exemplo adicional e os programas utilizados são apresentados nos apêndices.

2 REFERENCIAL TEÓRICO

Os primeiros trabalhos com TRI foram realizados a partir de meados do século passado. Porém, com as ferramentas estatísticas por ela utilizadas são muito mais complexas que as utilizadas pela forma usual de medir, o seu crescimento e divulgação somente se tornaram possíveis com o avanço dos recursos computacionais. Atualmente, esta técnica tem se tornado predominante nas avaliações em diversos países. No Brasil, a primeira aplicação dessa teoria ocorreu em 1995 na análise do Sistema Nacional de Avaliação da Educação Básica (SAEB). A partir de então, seu uso nas avaliações educacionais brasileiras tem sido valorizado e incentivado pelos órgãos governamentais, com o Ministério da Educação.

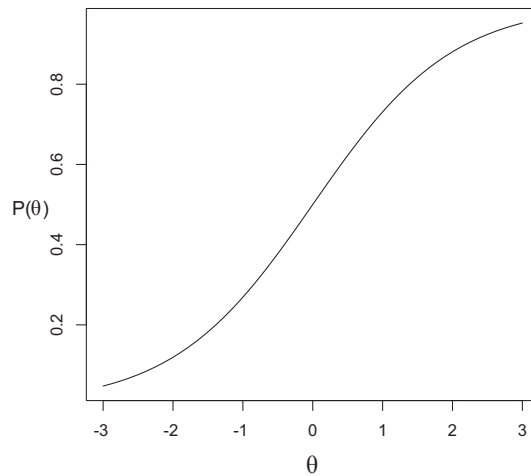


FIGURA 1: Curva característica do item - CCI

A CCI possui três propriedades que a caracteriza . São elas:

i) dificuldade do item - representada pela letra **b**, por sua vez, representa o ponto, na escala de habilidade, e que o indivíduo, com aquela habilidade, possui 50% de probabilidade de responder corretamente a esse item . Por isso, é também chamado, índice de locação;

ii) discriminação do item - representada pela letra **a**, representa a inclinação da curva no ponto **b** e faz com que se torne possível diferenciar entre examinados que estão abaixo ou acima do índice de locação. Por exemplo, se a curva é chata, então o item é capaz de discriminar, pois a probabilidade de resposta correta àquele item para níveis de habilidade menores é, aproximadamente, a mesma que para níveis de habilidade maiores que **b**. Portanto, quanto maior for o valor do parâmetro **a**, mais sensível torna-se o modelo às variações na habilidade e torno de seu ponto de dificuldade. Por isso, ele é chamado de parâmetro de discriminação do item ;

iii) probabilidade de acerto casual - representada pela letra **c**, representa a probabilidade de que o examinado, com baixa habilidade, venha

a acertar a questão, se θ que ele saiba a resposta, isto é,

$$\lim_{\theta \rightarrow -\infty} P(\theta) = c$$

Exemplos de CCI:

a) Comparação de CCI com a mesma discriminação e diferentes dificuldades

Na Figura 2 estão representadas três CCI com o mesmo valor para o parâmetro discriminação e diferentes valores para o parâmetro dificuldade.

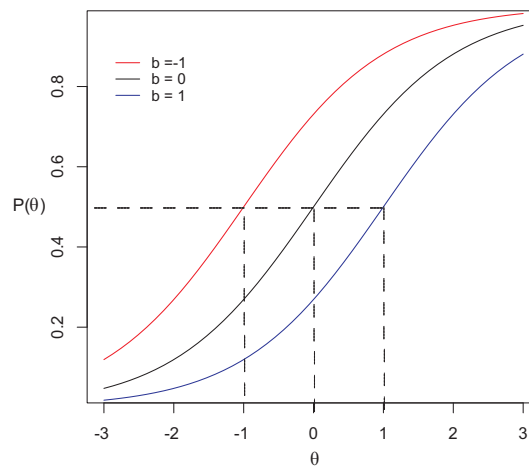


FIGURA 2: CCI com diferentes valores para \mathbf{b} (parâmetro dificuldade do item) e o mesmo valor de \mathbf{a} (parâmetro discriminação do item).

A curva vermelha representa o item fácil, pois as probabilidades de se obter 50% de probabilidade de resposta correta a esse item são dadas por um indivíduo com habilidade -1. A curva preta representa o item de média dificuldade, pois indivíduos com habilidade média (0) possuem 50% de probabilidade de responder corretamente a esse item. A curva azul representa o item mais difícil, pois, a probabilidade de 50% para que se acerte esse item é alcançada por indivíduos com habilidade mais alta (1).

Portanto, a classificação de um item como sendo fácil, médio ou difícil vai depender de onde ele se encontra na escala da habilidade. Se um indivíduo com pouca habilidade possui uma probabilidade alta de responder corretamente a esse item, ele é considerado fácil. Se apenas aquele que possui habilidade alta possui 50% de probabilidade de responder corretamente a esse item, ele é considerado difícil. Porém, mesmo o item difícil pode possuir uma baixa discriminação, o que vai depender do parâmetro **a**. Portanto, ele serve para adicionar significado ao parâmetro **b**.

Um item fácil é útil para comparar indivíduos com baixa habilidade, enquanto que um item difícil é útil para a comparação entre indivíduos com alta habilidade.

b) Comparação de CCI com a mesma dificuldade e diferente discriminação.

Um exemplo de três CCI com o mesmo valor para o parâmetro dificuldade e diferentes valores para o parâmetro discriminação é representado na Figura 3.

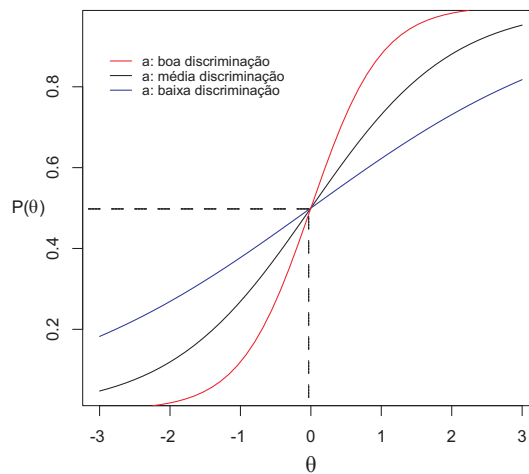


FIGURA 3: CCI com diferentes valores para **a** (parâmetro discriminação do item) e o mesmo valor de **b** (parâmetro dificuldade do item).

A curva vermelha representa um item que discrimina bem, pois a inclinação da curva no ponto onde $b=0$ é bastante íngreme. Isso faz com que a probabilidade de um indivíduo responder corretamente a esse item e que possua habilidade pouco acima do índice de locação, seja bem maior que a probabilidade de um indivíduo que possua habilidade pouco abaixo desse índice. A curva preta possui um nível de discriminação médio, enquanto a curva azul possui um nível de discriminação muito baixo, pois a probabilidade de resposta correta modifica-se muito pouco abaixo e acima do ponto b . Observa-se que mesmo o indivíduo com baixa habilidade possui alta probabilidade de responder corretamente a esse item.

Poderia-se considerar a CCI, que se encontra na Figura 4, como sendo aquela que apresenta poder de discriminação ideal. Porém, ela tem o inconveniente de somente servir para distinguir entre indivíduos que possuíssem habilidades bem próximas de $\theta = 0$. Um pouco além deste valor já não teria nenhum poder de discriminação.

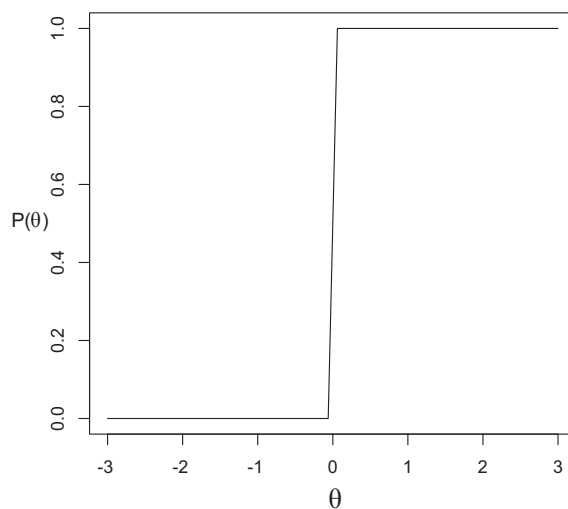


FIGURA 4: CCI Ideal.

c) Comparação de CCI com diferentes probabilidades de acerto casual

Na Figura 5 estão representadas três CCI com diferentes valores para a probabilidade de acerto casual.

Pode-se notar que a diferença na probabilidade de acerto entre indivíduos com habilidades diferentes diminui quando o valor do parâmetro c aumenta. Por exemplo: para a curva azul, a probabilidade de que um indivíduo com habilidade -1 acerte essa questão é de, aproximadamente, 0,55 e outro, com habilidade 1, de, aproximadamente, 0,85. A diferença aproximada na probabilidade entre eles é, pois, de 0,3. Já para a curva preta, a probabilidade de acerto para o indivíduo com habilidade -1 é de, aproximadamente, 0,25 e, para o indivíduo com habilidade 1, de, aproximadamente, 0,75, ou seja, a diferença aproximada de 0,5. Portanto, quando o parâmetro c passou do valor 0,0 (curva preta) para o valor 0,4 (curva azul), isto é, teve seu valor aumentado, a diferença na probabilidade de acerto entre estes dois indivíduos diminuiu: passou de 0,5 para 0,3. Isso significa que o poder de discriminação diminuiu.

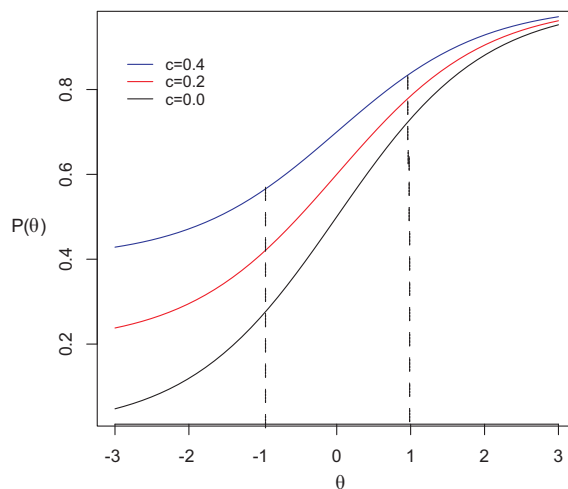


FIGURA 5: CCI com diferentes valores para c (parâmetro da probabilidade de acerto casual) e os valores de a (parâmetro de discriminação) e b (parâmetro de dificuldade do item).

2.2 Modelos para Teoria de Resposta ao Item

Vários modelos têm sido propostos na literatura para TRI. Essa diversidade deve-se a três fatores: 1) o item ser de caráter discreto (e.g., binário, associado a acerto e questões fechadas) ou contínuo (associado à nota e questões abertas); 2) envolver uma ou mais populações e 3) modelagem da habilidade que está sendo avaliada. Modelos e que mais de uma habilidade está sendo medida (modelos multidimensionais) podem ser encontrados em Hableton & Cook (1977); modelo de resposta gradual, e que não são consideradas simplesmente as respostas como corretas ou incorretas, e Saehji (1969) e modelos para duas ou mais populações, e Bock & Zimowski (1997).

Os primeiros modelos foram desenvolvidos por Lord (1952) para itens binários, uma população e uma única habilidade, na forma de uma função normal acumulada. Mais tarde, eles foram descritos por meio de uma função

logística por Birnbau (1968). Dependendo do número de parâmetros relacionados ao item, essa função pode ser classificada em três tipos de modelos, os quais estão descritos abaixo e são os que mais são utilizados pela TRI atualmente, para itens binários.

A variável e os parâmetros destes modelos são representados como:

- Y_{ij} : variável dicotômica que pode assumir os valores 1 (quando o examinado i responder corretamente ao item j) ou 0 (quando ele responder incorretamente);
- θ_i : parâmetro da habilidade do examinado i ;
- a_j : parâmetro do poder de discriminação do item j .
- b_j : parâmetro do grau de dificuldade do item j .
- c_j : parâmetro da resposta casual, isto é, a probabilidade do examinado obter uma resposta correta a um item puramente ao acaso.

2.2.1 Modelo logístico de 1 parâmetro

Neste modelo, somente o grau de dificuldade é considerado para descrever o item. Inicialmente, foi proposto por Rasch (1960) e expresso como um modelo logístico por Wright (1968). É dado pela equação:

$$P(Y_{ij} = 1|\theta_i) = \frac{1}{1 + e^{-(\theta_i - b_j)}}$$

Diferentemente da avaliação usual, em que a habilidade é diretamente proporcional à nota e que varia de $\mathbf{0}$ a \mathbf{n} , sendo \mathbf{n} o número de itens de um teste, a escala de habilidade pode assumir, teoricamente, qualquer valor real entre $-\infty$ a $+\infty$.

O parâmetro \mathbf{b} é definido como o ponto na escala de habilidade no qual o examinado, com essa dada habilidade, tem a probabilidade 0,5 de responder corretamente a esse item, ou seja,

$$\frac{1}{1 + e^{-(\theta_i - b_j)}} = 0,5$$

$$1 + e^{-(\theta_i - b_j)} = 2 \quad \Rightarrow \quad e^{-(\theta_i - b_j)} = 1 \quad \Rightarrow \quad e^{-(\theta_i - b_j)} = e^0$$

$$-(\theta_i - b_j) = 0 \quad \Rightarrow \quad \theta_i = b_j$$

2.2.2 Modelo logístico de 2 parâmetros

Neste modelo, além do grau de dificuldade, o item também é descrito pelo seu poder de discriminação. Foi Lord (1952) que desenvolveu este modelo, baseado na distribuição normal acumulada e, em 1968, Birnbaum o descreveu como o modelo logístico. É dado pela equação:

$$P(Y_{ij} = 1|\theta_i) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$$

O valor

É dado pela equação:

$$P(Y_{ij} = 1|\theta_i) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$$

Esse modelo representa a probabilidade condicional: probabilidade de que um indivíduo acerte o item j , dada sua habilidade. A primeira parcela refere-se à probabilidade de que ele o acerte, dado que isso ocorreu de forma casual e dada a sua habilidade, isto é, $c_j = P(Y_{ij} = 1|C, \theta_i)$, sendo C a escolha aleatória da alternativa. A segunda parcela corresponde à probabilidade do acerto, dado que ele sabia a resposta e dada a sua habilidade, ou seja, $(1 - c_j)p = P(Y_{ij} = 1|NC, \theta_i)$, e que NC representa a escolha da alternativa de forma não aleatória.

Como se trata de uma probabilidade, o domínio do parâmetro c está entre 0 e 1.

Este modelo é o mais completo, pois a probabilidade de que um indivíduo venha a acertar um item por meio do acaso é realmente possível, e, portanto, deve ser levado em consideração.

É interessante notar que, nesse modelo, o valor do parâmetro b não é mais o ponto da escala onde o examinado, com essa habilidade, tem a probabilidade de 0,5 de responder corretamente a esse item. Agora, quando $b = \theta$, a probabilidade de resposta correta fica

$$\begin{aligned} P(Y_{ij} = 1|\theta_i) &= c + (1 - c)(0,5) = \\ &= c + \frac{1 - c}{2} = \frac{2c + 1 - c}{2} = \frac{c + 1}{2} \end{aligned}$$

Isso significa que o parâmetro b , agora, define o ponto na escala de habilidade e que a probabilidade de resposta correta ao item corresponde ao valor médio entre c e 1.

2.3 Propriedade da TRI

A TRI possui uma propriedade muito importante: a invariância de seus parâmetros, isto é:

i) os valores dos parâmetros do item não dependem do nível de habilidade do examinado;

ii) o nível de habilidade do examinado é invariante com relação aos valores dos parâmetros dos itens usados para determiná-lo.

2.3.1 Ilustrando a propriedade de invariância do item

Para elucidar esta propriedade, será repetida, a seguir, a apresentação de Baker (2001), para um conjunto ligeiramente distinto de dados simulados.

Suponha-se que apenas dois grupos de examinados fossem testados de uma mesma população: o primeiro grupo com habilidade entre -3 a -1 e o segundo grupo com habilidade entre 1 a 3. Plotando-se as proporções de resposta correta para ambos os grupos e ajustando-se uma CCI, por meio da estimação de seus parâmetros, a essas proporções, obtêm-se os Gráficos que estão representados de forma ilustrativa nas Figuras 6 e 7.

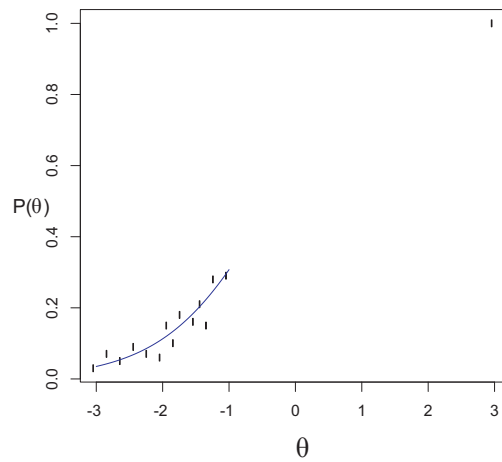


FIGURA 6: CCI para as proporções de respostas corretas para um grupo de examinados com habilidade entre -3 a -1.

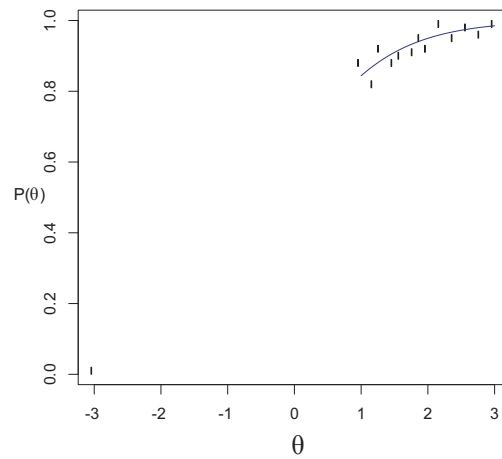


FIGURA 7: CCI para as proporções de respostas corretas para um grupo de examinados com habilidade entre 1 a 3.

As estimativas obtidas para os parâmetros desse item, para os dois grupos separadamente, teria os mesmos valores, isto é, $a_1 = a_2$ e $b_1 = b_2$. Se for plotada, no mesmo gráfico, a CCI desse item para ambos os grupos, verifica-se que ela será coincidente, como se pode verificar pelo gráfico da

Figura 8.

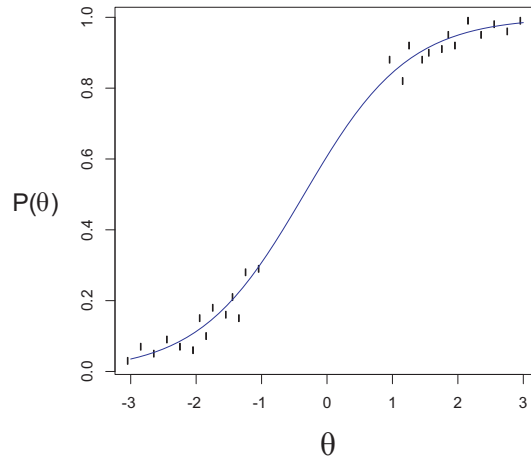


FIGURA 8: CCI para as proporções de respostas corretas para um grupo de examinados com habilidade entre -3 a 3.

Poderia ser que o grupo 1 considerasse esse item difícil e o grupo 2, fácil. Mas, o grau de dificuldade desse item, neste caso, é $b = 0$, isto é, de média dificuldade. Isto deixa claro que a característica fácil, média ou difícil é uma propriedade do item e não dependente do grupo que responde a ele.

2.4 Métodos de estimação

Vários são os métodos utilizados para a estimação dos parâmetros na TRI. Azevedo (2003) desenvolveu um trabalho em que apresenta e discute os principais deles.

Inicialmente, a estimação dos parâmetros foi feita pelo método da máxima verossimilhança. Um exemplo dessa metodologia pode ser visto em Baker (2001). Porém, por meio deste método, se um item é respondido corretamente ou incorretamente por todos os examinados, não é possível estimar seus parâmetros, assim como, se um examinado responder corretamente ou incorretamente

ente a todos os itens, talvez não é possível estimar sua habilidade. Mais recentemente, métodos bayesianos de estimação foram propostos e, com isso, tais problemas foram resolvidos (Andrade, 2000).

Neste trabalho, é considerada apenas uma população de cada vez, sendo os itens de caráter binário e a característica da habilidade, unidimensional. O processo de estimação dos parâmetros foi feito via inferência bayesiana, utilizando-se o algoritmo Metropolis & Hastings (Hastings, 1970; Metrópolis et al., 1953;).

2.5 Inferência bayesiana

A inferência bayesiana é caracterizada por incorporar informação à priori aos dados (Paulino et al., 2003).

A definição de inferência é a de que ela é uma rama da Estatística que visa construir técnicas que permitam fazer afirmações sobre os parâmetros populacionais, com base em dados amostrais. A distribuição dessa amostra é dada por uma função de verossimilhança. Ela define uma relação entre os dados e um parâmetro desconhecido β (Mood et al., 1974). Na inferência bayesiana, toda informação que se tenha sobre esse parâmetro, antes de ser observada a amostra, deve ser considerada e incorporada aos dados, por meio de uma distribuição a *priori*. A distribuição resultante é dada por uma distribuição a *posteriori* (O'Hagan, 1994).

Um procedimento para construir-se a distribuição a *posteriori* partindo da distribuição a *priori* é o teorema de Bayes.

Seja $A = \{\cup A_i, i \in N^*\}$, uma partição do espaço amostral Ω e B um evento qualquer de Ω . Utilizando-se o teorema de Bayes, tem-se:

$$P(A_i|B) = \frac{P(A_i).P(B|A_i)}{P(B)}$$

onde cada elemento desta expressão, no contexto de inferência, pode ser interpretado como:

$P(A_i|B)$: probabilidade a *posteriori*;

$P(A_i)$: probabilidade a *priori*;
 $P(B|A_i)$: probabilidade de B dado A_i ;
 $P(B)$: marginal de B, igual a:

$$P(B) = \sum P(A_i) \cdot P[B|A_i]$$

que não depende de A_i . Sendo assim, a distribuição condicional de interesse pode ser escrita como:

$$P(A_i|B) \propto P(A_i) \cdot P(B|A_i)$$

Para o caso contínuo, denotando-se por β o parâmetro de interesse e por y_1, \dots, y_n os dados observados, temos:

$$f(\beta|y_1, \dots, y_n) = \frac{\prod_{i=1}^n f(y_i|\beta) \cdot f(\beta)}{\int_{\beta} \prod_{i=1}^n f(y_i|\beta) \cdot f(\beta) d\beta}$$

e

$$f(\beta|y_1, \dots, y_n) \propto \prod_{i=1}^n f(y_i|\beta) \cdot f(\beta)$$

A distribuição $f(\beta)$ é para o caso de modelagem de inferência, denominada distribuição a *priori* e resume a informação que é conhecida sobre o parâmetro, prévia ao conhecimento dos dados. A distribuição $f(\beta|y_i)$ é chamada distribuição a *posteriori* e reflete o que é conhecido sobre as distribuições a priori após a obtenção dos dados. Assim, a função de verossimilhança modifica ou atualiza o conhecimento que se tinha sobre o parâmetro. O teorema de Bayes é, pois, uma ferramenta que atualiza a informação feita sobre os parâmetros, com base na informação de uma amostra que já ocorreu (Box & Tiao, 1992).

As distribuições a *priori* podem ser informativas ou não informativas. Ela é dita não informativa quando não se tem nenhuma informação sobre o parâmetro. Quando houver pouca informação sobre os parâmetros, as distribuições a *priori* podem ser chamadas de "vagas" ("*flat*"). Ao ser

considerados todos os possíveis valores de β como igualmente prováveis, pode-se atribuir uma distribuição a priori uniforme, ficando assim $p(\beta) \propto k^{-1}$. Outra proposta de distribuição a priori não informativa foi feita por Jeffreys (1961). Tal proposta visa minimizar o conteúdo de informação, com o objetivo de que o máximo desta seja extraído da amostra. Tais distribuições a priori são muito usadas, talvez, para a apresentação de resultados "pouco perturbados" por suposições anteriores. Ao final de um processo de inferência bayesiana, pode-se avaliar o quanto o conhecimento da distribuição a priori afeta o conhecimento da distribuição a posteriori e o quanto houve de influência da distribuição a priori sobre os resultados.

Toda inferência bayesiana é feita com base na distribuição a posteriori. Porém, esta distribuição é, muitas vezes, desconhecida ou muito complexa, tornando difícil operações de integração, fundamental no caso contínuo. Por exemplo: muitas vezes, deseja-se conhecer a distribuição conjunta de um determinado vetor de parâmetro β ($\beta = \beta_1, \beta_2, \dots, \beta_p$). Para um determinado parâmetro β_i , essa distribuição é chamada distribuição marginal e a forma de obtê-la é integrar a distribuição a posteriori conjunta em relação aos demais parâmetros, isto é,

$$f(\beta_i) = \int \dots \int f(\beta_{i, \beta_{-i}}) d\beta_{-i},$$

onde β_{-i} corresponde ao conjunto complementar de β_i .

Outras vezes, deseja-se a média a posteriori ou qualquer esperança de funções do parâmetro β , isto é,

$$E[g(\beta)|y] = \int g(\beta)p(\beta|y)d\beta$$

sendo $g(\beta)$ a função de interesse.

A inferência exata somente será feita calculando-se estas integrais analiticamente, o que, com raras exceções, é impossível ou muito trabalhoso. A alternativa, então, é fazer uso de aproximações numéricas. Paulino et al. (2003) apresenta várias estratégias sugeridas para resolver o problema da integração: aproximação da distribuição a posteriori por uma dis-

tribuição normal multivariada, método de Laplace, métodos de quadratura numérica, métodos diretos de Monte Carlo e métodos de Monte Carlo via cadeias de Markov (MCMC). Estes métodos estão em ordem crescente de pressuposições e complexidades computacionais.

Será feita, agora, uma breve descrição do método MCMC, por ter sido este o método utilizado neste trabalho.

2.5.1 Monte Carlo via cadeias de Markov (MCMC)

O método de Monte Carlo via cadeias de Markov (MCMC) consiste num dos métodos numéricos para obter uma amostra da função de in

distribuições completas não são conhecidas, como é o caso do nosso trabalho, utiliza-se o algoritmo MH. Maiores detalhes sobre o algoritmo AG podem ser encontrados em Gaieran (1997).

2.5.2 Algoritmo Metropolis & Hastings

O algoritmo Metropolis & Hastings consiste em um poderoso algoritmo do tipo MCMC para gerar valores para a distribuição de interesse utilizando uma distribuição conhecida como auxiliar. Foi desenvolvido através dos trabalhos de Hastings (1970) e Metropolis et al. (1953). Esta distribuição será chamada de **distribuição geradora de candidatas** e será representada por $q(\beta)$.

O procedimento usado para que os valores amostrados pertençam à distribuição de interesse será por meio dos seguintes passos:

1 - atribui-se um valor ("chute") inicial β como primeiro valor da amostra;

2 - calcula-se $P(\beta)$ e $q(\beta)$;

3 - amostra-se β^* da distribuição candidata e calcula-se $P(\beta^*)$ e $q(\beta^*)$;

4 - obtém-se a razão:

$$r = \frac{\frac{P(\beta^*)}{q(\beta^*)}}{\frac{P(\beta)}{q(\beta)}} = \frac{P(\beta^*) \cdot q(\beta)}{P(\beta) \cdot q(\beta^*)}$$

5 - em uma probabilidade de aceitação:

$$\alpha(\beta, \beta^*) = \min\left\{1, \frac{P(\beta^*) \cdot q(\beta)}{P(\beta) \cdot q(\beta^*)}\right\}$$

6 - gera-se $u \sim U(0, 1)$;

7 - aceita-se β^* com o valor da densidade de interesse, se $u \leq \alpha$. Caso contrário, rejeita-se β^* , repete-se β na amostra e retorna-se ao passo 3.

Esse processo é repetido até que o número de iterações desejada seja atingido.

Segundo Chib & Greenberg (1995), várias são as famílias que podem ser escolhidas para a densidade candidata. No entanto, explorando-se a forma de $P(\beta)$, essa escolha pode ser feita de modo a simplificar a implementação do algoritmo.

Para anular a influência do "chute" inicial, desprezam-se os valores das primeiras iterações. Isto é chamado de "*burn-in*" e, para que não haja correlação entre os dados, os valores das iterações são tomados de certa e certa distância, o que é chamado de *thinning*.

3 MATERIAIS E MÉTODOS

Os dados utilizados para o desenvolvimento deste trabalho referem-se ao Vestibular 2006-2 da Universidade Federal de Lavras (UFLA).

A estimação dos parâmetros foi feita por meio da inferência bayesiana, utilizando-se o método Monte Carlo via cadeia de Markov (MCMC). O modelo adotado foi o modelo logístico de 3 parâmetros.

Para melhor visualização e compreensão dos conceitos básicos da TRI e dos resultados alcançados por meio desse método, foi ajustado um exemplo de Baker (2001).

3.1 Materiais

A prova do Vestibular 2006-2 da UFLA constou de 66 itens, distribuídos da seguinte forma:

- Português - questões de 1-10;
- Geografia - questões de 11-18;
- História - questões de 19-24;
- Filosofia - questões de 25-26;
- Língua Estrangeira - questões 27-34;
- Biologia - questões 35-42;
- Física - questões 43-50;
- Matemática - questões 51-58;
- Química - questões 59-66.

Como havia, na Língua Estrangeira, a opção entre o Inglês e o Espanhol, as questões desta prova não foram incluídas na análise, com o objetivo de simplificar, passando a ser considerados 58 itens.

Fora oferecidos 10 cursos: Ad inistração (AD), Engenharia de Ali entos (AL), Agrono ia (AG), Ciências Biológicas (CB), Ciências da Co putação (CC), Engenharia Agrícola (EA), Engenharia Florestal (EF), Medicina Veterinária (MV), Quí mica (QI) e Zootecnia (ZO).

Para fins de si plificação do trabalho co putacional, considerou-se que cada curso for a u a população diferente, sendo a etodologia aplicada separada ente a cada u deles. O nú ero de candidatos por curso foi o seguinte: AD = 284; AL = 229; AG = 651; CB = 326; CC = 235; EA = 154; EF = 238; MV = 604; QI = 104; ZO = 228.

As análises fora feitas e u co putador equipado co procesador AMD-64 Dual, co 1 Gb de e ória RAM. Estas análises consu ira 14 dias de processa ento, utilizando-se o R-2.4.0.

Os progra as utilizados para i ple entar a etodologia bayesiana fora elaborados na language R (R Develop ent Core Tea , 2005). Para a verificação das cadeias geradas do MCMC, fez-se uso da biblioteca CODA (Plu er et al., 2006). Esses progra as estão apresentados no Apêndice B.

3.2 Metodologia

Cada ite de u teste possui duas possibilidades de resposta: certa ou errada. Considerando o valor 1 para a resposta correta e 0 para a resposta incorreta, te -se a seguinte distribuição de Bernoulli:

$$f_{Y_{ij}}(y_{ij}; p) = p^{y_{ij}}(1 - p)^{1 - y_{ij}}$$

e que,

$y_{ij} = \{0, 1\}$: variável aleatória das possíveis respostas do indivíduo i para o ite j ;

p : probabilidade de y_{ij} ser igual a 1.

Como \mathbf{p} representa o valor da probabilidade do examinado \mathbf{i} responder corretamente ao item \mathbf{j} e, essa probabilidade está em função de sua habilidade e dos parâmetros associados a esse item, \mathbf{p} também é uma variável aleatória e, portanto, possui uma distribuição, que é dada pelo modelo 3 parâmetros, o qual se encontra especificado nas páginas 12, do capítulo 2.

Sendo aplicado um teste com \mathbf{J} itens para \mathbf{I} indivíduos e considerando para \mathbf{p} o modelo de 3 parâmetros, tem-se a seguinte função de verossimilhança:

$$f_{Y_{11}, \dots, Y_{IJ}}(y_{11}, \dots, y_{ij} | a_j, b_j, c_j, \theta_i) = \prod_{i,j}^{I,J} [c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}]^{y_{ij}} \cdot \{1 - [c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}]\}^{(1 - y_{ij})}$$

O procedimento que foi usado para a estimação dos parâmetros neste trabalho foi o da teoria bayesiana.

Para o parâmetro \mathbf{a} , a distribuição a *priori* escolhida foi a Lognormal, pois a CCI é não decrescente e, portanto, valores negativos não fazem sentido para ele. Se considerasse os valores negativos para este parâmetro, isso significaria que quanto maior a habilidade de um indivíduo, menor seria sua probabilidade de acertar a questão, o que seria incoerente. Sua distribuição é, pois:

$$f_{A_j}(a_j) = \frac{1}{a_j \sqrt{2\pi}} e^{-\frac{1}{2}[\ln(a_j)]^2}$$

Para o parâmetro \mathbf{b} , foi escolhida a distribuição Unifórm e (-3,3), pois, para este parâmetro, não se tem muita informação a *priori*. A escolha do intervalo [-3,3] foi para relacionar o domínio de \mathbf{b} ao de θ . Não foi escolhida uma distribuição a *priori* normal para este parâmetro, devido ao fato de que este se encontra dependente de quem elabora as questões, não sendo

razoável supor que tenha distribuição normal. Sua distribuição a *priori* escolhida foi, pois:

$$f_{B_j}(b_j) = \frac{1}{6}$$

Note que o domínio de θ é $(-\infty, \infty)$ e o domínio de \mathbf{b} é $[-3, 3]$. Estas duas variáveis devem ser associadas, porém, desejou-se possibilitar que $\theta < -3$ e $\theta > 3$ refletisse candidatos com habilidade aquém ou além da exigida pela prova. O uso da distribuição Normal $(0,1)$ para θ , no entanto, minimiza a expectativa de que tais valores de habilidade ocorram.

A distribuição a *priori* escolhida para o parâmetro \mathbf{c} foi uma distribuição Beta(1,3), pois, tratando-se de uma probabilidade, a informação de que se dispôs foi a de que ele se encontra no intervalo $(0,1)$ e, como as questões do vestibular possuíam 4 alternativas, o valor esperado para um acerto casual é de 0,25. Outros valores também poderia ser sido escolhidos para essa distribuição, mas desejara-se aqueles que minimizasse $\alpha + \beta$, a fim de que fosse pouco informativa (α e β são os parâmetros da Beta, sendo o valor obtido pela soma correspondente ao espaço amostral de um binomial).

Outra alternativa poderia ser a de considerar a moda e a média, como sendo igual a 0,25. Neste caso, uma distribuição Beta(2,4), por exemplo, resultaria numa moda igual a 0,25, porém, seu espaço amostral não seria finito e o valor esperado seria de, aproximadamente, 0,33, que pode ser considerado alto.

Sua distribuição, então, é:

$$f_{C_j}(c_j) = \frac{1}{B(1, 3)}(1 - c_j)^2$$

Utilizou-se como distribuição a *priori* para o parâmetro θ uma distribuição Normal $(0,1)$, pois, para as habilidades dos candidatos, é coerente admitir que as características desta distribuição estão de acordo com a popu-

lação e estudo. Sua distribuição é:

$$f_{\Theta_i}(\theta_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta_i^2}$$

3.3 Distribuição a Posteriori

Multiplicando-se todas as distribuições *a priori* pela função de verossimilhança, tem-se a distribuição conjunta *a posteriori*, dada por:

$$f_{(A_j, B_j, C_j, \theta_i | Y_{ij})}(a_j, b_j, c_j, \theta_i | y_{ij}) \propto \prod_{i,j}^{I,J} f_{A_j}(a_j) \cdot f_{B_j}(b_j) \cdot f_{C_j}(c_j) \cdot f_{\Theta_i}(\theta_i).$$

$$\cdot \prod_{i,j}^{I,J} f_{(Y_{ij} | A_j, B_j, C_j, \theta_i)}(y_{ij} | a_j, b_j, c_j, \theta_i)$$

$$f_{(A_j, B_j, C_j, \theta_i | Y_{ij})}(a_j, b_j, c_j, \theta_i | y_{ij}) \propto \prod_{i,j}^{I,J} \frac{1}{a_j \sqrt{2\pi}} e^{-\frac{1}{2}[\ln(a_j)]^2} \cdot \frac{1}{6} \cdot \frac{1}{B(1, 3)} (1-c_j)^2 \cdot \frac{1}{\sqrt{2\pi}}$$

$$e^{-\frac{1}{2}\theta_i^2} \cdot \prod_{i,j}^{I,J} \left[c_j + (1-c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \right]^{y_{ij}} \cdot \left\{ 1 - \left[c_j + (1-c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \right] \right\}^{(1-y_{ij})}$$

$$f_{(A_j, B_j, C_j, \theta_i | Y_{ij})}(a_j, b_j, c_j, \theta_i | y_{ij}) \propto \prod_{i,j}^{I,J} \frac{(1-c_j)^2}{12\pi a_j} \cdot \frac{e^{-\frac{1}{2}[\ln^2(a_j) + \theta^2]}}{B(1, 3)}.$$

$$\cdot \prod_{i,j}^{I,J} \left[c_j + (1-c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \right]^{y_{ij}} \cdot \left\{ 1 - \left[c_j + (1-c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \right] \right\}^{(1-y_{ij})}$$

3.3.1 Modelo logístico de 3 parâmetros

O modelo utilizado neste trabalho foi o modelo logístico de 3 parâmetros, pois a probabilidade de que um indivíduo venha a acertar a questão de forma casual, uma vez que ele não sabe a resposta, é a probabilidade que realmente existe e, portanto, deve ser levada em consideração. Além disso, este modelo é o mais completo e o mais utilizado atualmente, principalmente nos sistemas de avaliação educacional (como, por exemplo, pelo SAEB).

Utilizou-se o intervalo de máxima densidade de probabilidade (ou HPD, do inglês *highest posterior densit*) do parâmetro " c ", com o objetivo de confirmar a relevância deste parâmetro. Considera-se que, se houver um número considerável de intervalos com valores menores ou maiores do que 0,25 para a probabilidade de acerto casual, o uso de um modelo que leve em consideração o acréscimo deste parâmetro, como é o caso do modelo logístico de 3 parâmetros, será o mais indicado.

Para a classificação dos parâmetros do item serão usados os critérios que se encontram nas Tabelas 1 e 2:

a) Para o parâmetro de discriminação (a):

TABELA 1: Intervalos com as respectivas classificações e valores esperados na priori para o parâmetro a .

Intervalo para a	Classificação	% de ocorrência na priori
$a > 3$	Ótima (O)	13,34%
$2 < a \leq 3$	Alta (A)	25,58%
$1 < a \leq 2$	Média (M)	11,02%
$a \leq 1$	Baixa (B)	50,06%

Este critério foi estabelecido com base na distribuição a priori escolhida (lognormal), cuja média é de 1,65. Com estes valores, espera-se que grande parte dos itens obtenha baixo poder de discriminação (50,06%); 38,92% esteja acima da média e apenas 13,34% deles tenha um poder de discriminação ótimo.

b) Para o parâmetro dificuldade (**b**):

TABELA 2: Intervalos com as respectivas classificações e valores esperados na priori para o parâmetro **b**.

Intervalo para b	Classificação	% de ocorrência na priori
$b > 2$	Difícil (D)	1/6
$0,5 < b \leq 2$	Média (M)	1/4
$-2 < b \leq 0,5$	Fácil (F)	5/12
$b \leq -2$	Muito Fácil (MF)	1/6

Com este critério, espera-se que sejam poucas as questões que apresentem graus de dificuldade extremos, isto é, difícil ou muito fácil. Acredita-se que a maior parte delas seja fáceis, sendo que apenas 25% possuam grau médio de dificuldade. Segundo Andrade et al. (2000), os valores do parâmetro **b** variam (tipicamente) entre -2 e 2. Para o parâmetro **a**, esperam-se valores entre 0 e 2, sendo mais apropriados valores maiores que 1.

3.3.2 Implementação do processo amostral

Foi gerada uma cadeia inicial com 100.000 iterações para todos os parâmetros. Descartaram-se as 500 primeiras observações ("*burn-in*") e adotada de pontos amostrais e uma a cada 25 iterações ("*thinning*"). Isso foi feito para garantir a segurança do processo amostral. A amostra válida constou de 4.000 observações.

A rotina implementada no CODA para garantir a convergência das cadeias geradas pelo critério de Raftery & Lewis estabelece o tamanho inicial do *burn-in* e do *thinning*. Foram fornecidos valores inferiores a 100 para o *burn-in* e inferiores a 20 para o *thinning*. Como foram usados valores superiores a estes, aceitou-se a amostra com o sentido da distribuição a *posteriori*.

Para implementar o algoritmo MH, por simplicidade, foi escolhida a distribuição geradora de candidatas com o sentido a respeito da distribuição a *priori* dos parâmetros de interesse.

Sendo a probabilidade de aceitação dada por:

$$\alpha(\beta, \beta^*) = \min\left\{1, \frac{P(\beta^*) \cdot q(\beta)}{P(\beta) \cdot q(\beta^*)}\right\}$$

que pode ser escrita como:

$$\alpha(\beta, \beta^*) = \min\left\{1, \frac{\psi(\beta^*) \cdot \prod f(\beta^*) \cdot q(\beta)}{\psi(\beta) \cdot \prod f(\beta) \cdot q(\beta^*)}\right\}$$

onde $\psi(\cdot)$ é a distribuição a *priori* e $\prod f(\cdot)$, a função de verossimilhança da distribuição de interesse, se for escolhida $q(\cdot) = \psi(\cdot)$, tem-se:

$$\beta(\beta, \beta^*) = \min\left\{1, \frac{\psi(\beta^*) \cdot \prod f(\beta^*) \cdot \psi(\beta)}{\psi(\beta) \cdot \prod f(\beta) \cdot \psi(\beta^*)}\right\} = \min\left\{1, \frac{\prod f(\beta^*)}{\prod f(\beta)}\right\}$$

o que torna mais simples a fórmula algébrica da probabilidade de aceitação (que passa ser a razão de verossimilhança).

Dessa forma, as distribuições candidatas escolhidas para gerar valores aos elementos dos vetores de parâmetros \mathbf{a} , \mathbf{b} , \mathbf{c} , e θ foram, respectivamente, a lognormal, a uniforme, a beta e a normal.

3.3.3 Exemplo ilustrativo da implementação do algoritmo

Utilizando-se o exemplo ilustrativo de Baker (2001), foi aplicado o método MCMC, considerando o modelo de 2 parâmetros. Os dados desse exemplo se encontram no Anexo A e refere-se a 10 itens respondidos por 16 indivíduos. Será dada a seguir, a sequência dos passos realizados:

1) Valores iniciais (escolha arbitrária):

$\mathbf{a} = [0,1066 \ 0,3209 \ 0,6202 \ 1,3260 \ 0,9207 \ 0,4816 \ 0,5783 \ 0,0368 \ 0,8314 \ 0,6252]'$

$$b = [0,2140 \ -0,2920 \ -0,2572 \ -0,9786 \ -0,3320 \ -1,5781 \ -1,2077 \ 0,8649 \\ -1,0024 \ -0,7350]'$$

$$\theta = [1,4767 \ 0,5067 \ -1,3396 \ -0,4372 \ 0,9408 \ -0,4571 \ 0,2523 \ -0,0069 \\ 0,1047 \ -1,9899 \ -0,1924 \ -1,4474 \ 0,4313 \ -0,5443 \ 0,6309 \ 0,4281]'$$

e que o que o apóstrofo (') representa o vetor transposto.

A probabilidade de acerto é dada por:

$$P(Y_{ij} = 1|\theta_i) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$$

A verossimilhança:

$$L(y_{1,1}, \dots, y_{16,10}|\beta) = 0,4664 \times 0,3618 \times \dots \times 0,7666 \times 0,6742$$

$$\text{sendo } \beta = [a_j, b_j, \theta_i]'$$

$$\ln[L(y_{1,1}, \dots, y_{16,10}|\beta)] = -132,0127$$

2) Gera-se um valor de u a lognormal (candidata) para $a_1 = 0,2937$.

3) Calcula-se a verossimilhança com este valor de a_1 e obtém-se:

$$\ln[L(y_{1,1}, \dots, y_{16,10}|\beta)] = -130,9249$$

4) A probabilidade de aceitação α é dada por:

$$\alpha(0,1066; 0,2937) = \min\left\{1, \frac{\exp(-130,9249)}{\exp(-132,0127)}\right\} = \min\{1; 2,9677\}$$

$$\alpha = 1$$

5) Gera-se u de $U[-3,3]$: $u = 0,1869$.

Como $0,1869 < 1$, aceita-se $0,2937$ como o valor da a ostra.
 Repete-se o processo para todos a_j . Faz-se o mesmo com b_j e θ_i .
 O vetor de a_1 , neste passo, é dado por: $(0,1066; 0,2937)$.

6) Gera-se, novamente, outro valor para a_1 da lognormal (candidata)
 $= 0,6363$.

A log-verossimilhança com esse novo valor é: $-134,5665$.

A probabilidade de aceitação, agora, fica:

$$\alpha(0,2937; 0,6363) = \min\left\{1; \frac{\exp(-134,5665)}{\exp(-130,9249)}\right\} = \min\{1; 0,0262\}$$

$$\alpha = 0,0262$$

7) Gera-se outro valor u de $U[-3,3]$: $u = 0,0932$.

Como $0,0932 > 0,0262$, rejeita-se $0,6363$. Repete-se, então,
 $0,2937$ como o valor da a ostra.

8) Novamente repete-se o processo com todos os a_j, b_j e θ_i .

O vetor de a_1 até esse passo é: $(0,1066; 0,2937; 0,2937)$.

9) Repete-se esse processo para 100.000 iterações, queiando-se as
 primeiras 500 ("burn in") e tomando-se uma amostra a cada 25 iterações
 ("thinning").

As estimativas obtidas para os valores dos itens e das habilidades,
 neste exemplo, se encontram nas Tabelas 3 e 4. A coluna "1 Parâmetro" se
 refere às estimativas obtidas por Baker (2001) pelo método de máxima
 verossimilhança, fazendo uso do modelo logístico de 1 parâmetro.

Nota-se que os parâmetros \mathbf{b} e θ apresentaram estimativas um pouco
 diferentes às do método de máxima verossimilhança obtidas por Baker (2001).
 Isto se deve ao fato do acréscimo do parâmetro \mathbf{a} no modelo. No entanto,
 as correlações de Pearson e de Spearman para o parâmetro \mathbf{b} estimado com
 o modelo de 2 parâmetros e o método utilizado por Baker foram de 0,90 e

0,98, respectivamente. Para θ , a correlação da nota com a habilidade estimada pelo modelo de 2 parâmetros foi de 0,92 e 0,94, respectivamente. Por meio dos valores obtidos por estas correlações, pode-se verificar que elas são altamente correlacionadas.

A Tabela do vetor de respostas para este exemplo se encontra no Anexo A.

TABELA 3: Valores estimados dos parâmetros **a** (discriminação do item) e **b** (dificuldade do item) para o modelo de 2 parâmetros e de **b** para o modelo de 1 parâmetro (Baker, 2001).

Ite	2 Parâmetros		1 Parâmetro
	a	b	b
1	4,00	-2,68	-2,22
2	0,81	-0,80	-0,30
3	0,16	-0,56	-0,17
4	1,75	0,67	0,59
5	0,48	-2,15	-1,20
6	4,00	-0,59	-0,02
7	2,48	-0,28	0,02
8	0,81	0,30	0,22
9	4,00	-0,19	0,15
10	0,45	1,66	0,26

TABELA 4: Escores brutos comparado com os valores estimados das habilidades (θ) para o modelo de 2 parâmetros e 1 parâmetro (Baker, 2001).

Examinado	Nota Bruta	θ (1 par.)	θ (2 par.)
1	-1,50	2	-3,23
2	-1,50	2	-1,95
3	0,02	5	-1,22
4	-0,42	4	-1,56
5	-2,37	1	-3,80
6	-0,91	3	-1,32
7	-0,42	4	-0,38
8	-0,42	4	0,13
9	-0,42	4	0,10
10	-0,91	3	-1,69
11	2,33	9	2,69
12	2,33	9	2,55
13	0,46	6	-1,11
14	2,33	9	2,54
15	2,33	9	2,74
16	***	10	2,92

4 RESULTADOS E DISCUSSÃO

4.1 Intervalos de máxima densidade de probabilidade (H.P.D.) para os parâmetros dos itens

As Tabelas 5, 6 e 7 referem-se aos intervalos HPD para o parâmetro "c" dos cursos de Medicina Veterinária, Agronomia e Engenharia de Alimentos, os quais foram colocados como exemplo. Para os demais cursos, os resultados obtidos são semelhantes.

Pode-se notar que, nas três situações, encontra-se elevado número de HPD que não abrange $c = 0,25$. Se tivesse os 5% dos itens e que o HPD não abrangesse $c = 0,25$, poderia-se considerar este parâmetro como constante e igual a 0,25. Cinco por cento de 58 itens corresponde a 3 itens e, no entanto, observa-se uma quantidade muito maior de intervalos que ultrapassaram este valor. O desejo, então, é poder estimar qual é o valor deste parâmetro, o que é possível por meio do modelo logístico de 3 parâmetros. Este, pois, pode talvez ser considerado um forte argumento a favor de adotá-lo como o mais indicado para o este trabalho.

Outra observação digna de nota diz respeito à disciplina de Geografia. Observa-se que, para ela, não houve nenhum HPD com $c < 0,25$, sendo, então, possível que todas as suas questões possam ter obtido uma probabilidade de acerto casual acima da média esperada, enquanto que, para as disciplinas de Física, Química e Matemática, principalmente para o curso de Agronomia, muitas de suas questões não favoreceram o acerto de forma casual.

TABELA 5: Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de $\hat{\mu}$ à densidade de probabilidade - *highest probability density*) para o parâmetro \mathbf{c} e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.

Disciplina	Ítem	Média	I.C. - HPD		Erro de M.C.
			inferior	superior	
Português	1	0,151	0,000	0,390	0,004
	2	0,163	0,000	0,402	0,006
	3	0,178	0,000	0,409	0,016
	4	0,160	0,000	0,396	0,006
	5	0,123	0,000	0,309	0,004
	6	0,258	0,221	0,297	0,001
	7	0,093	0,000	0,242	0,003
	8	0,144	0,115	0,172	0,001
	9	0,278	0,000	0,611	0,013
	10	0,079	0,000	0,199	0,002
Geografia	11	0,112	0,000	0,288	0,003
	12	0,171	0,000	0,417	0,006
	13	0,135	0,000	0,332	0,005
	14	0,155	0,000	0,348	0,005
	15	0,211	0,000	0,480	0,007
	16	0,140	0,000	0,342	0,009
	17	0,121	0,000	0,316	0,004
	18	0,111	0,000	0,306	0,004
História	19	0,103	0,000	0,285	0,003
	20	0,190	0,000	0,457	0,008
	21	0,154	0,000	0,275	0,006
	22	0,224	0,189	0,260	0,001
	23	0,145	0,000	0,391	0,015
	24	0,110	0,000	0,264	0,003
Filosofia	25	0,121	0,000	0,323	0,003
	26	0,067	0,000	0,191	0,002

cont...

cont...					
	27	0,136	0,000	0,396	0,017
	28	0,248	0,188	0,308	0,002
	29	0,271	0,157	0,360	0,002
Biologia	30	0,108	0,074	0,143	0,001
	31	0,266	0,198	0,329	0,001
	32	0,249	0,201	0,294	0,001
	33	0,214	0,000	0,422	0,018
	34	0,253	0,208	0,298	0,001
	35	0,210	0,172	0,249	0,001
	36	0,132	0,099	0,162	0,001
	37	0,237	0,201	0,275	0,001
Física	38	0,203	0,003	0,375	0,015
	39	0,205	0,141	0,256	0,001
	40	0,298	0,235	0,359	0,001
	41	0,132	0,091	0,179	0,001
	42	0,264	0,227	0,307	0,002
	43	0,167	0,116	0,219	0,001
	44	0,165	0,132	0,198	0,001
	45	0,149	0,079	0,210	0,002
Mateática	46	0,186	0,077	0,263	0,002
	47	0,242	0,003	0,360	0,014
	48	0,096	0,060	0,134	0,001
	49	0,178	0,000	0,379	0,008
	50	0,305	0,162	0,364	0,007
	51	0,099	0,067	0,130	0,001
	52	0,202	0,162	0,243	0,001
	53	0,327	0,000	0,606	0,011
Química	54	0,138	0,084	0,185	0,001
	55	0,255	0,222	0,290	0,001
	56	0,264	0,219	0,308	0,001
	57	0,208	0,167	0,249	0,001
	58	0,086	0,059	0,115	0,001

Obs.:Dados em negrito refere-se a HPD que não contém $c=0,25$.

TABELA 6: Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - *highest probability density*) para o parâmetro \mathbf{c} e respectivo erro de Monte Carlo, para o curso de Agronomia.

Disciplina	Ítem	Média inferior	I.C. - HPD		Erro de M.C.
			superior		
Português	1	0,252	0,000	0,584	0,016
	2	0,201	0,001	0,521	0,020
	3	0,254	0,008	0,410	0,016
	4	0,268	0,006	0,543	0,023
	5	0,185	0,000	0,508	0,014
	6	0,269	0,228	0,310	0,001
	7	0,284	0,011	0,403	0,016
	8	0,148	0,119	0,180	0,001
	9	0,242	0,000	0,571	0,015
	10	0,243	0,004	0,463	0,014
Geografia	11	0,339	0,024	0,634	0,021
	12	0,219	0,000	0,515	0,012
	13	0,201	0,000	0,521	0,012
	14	0,160	0,000	0,395	0,009
	15	0,336	0,044	0,600	0,009
	16	0,194	0,000	0,345	0,012
	17	0,171	0,000	0,436	0,010
	18	0,288	0,000	0,574	0,011
História	19	0,270	0,000	0,577	0,014
	20	0,164	0,000	0,503	0,020
	21	0,235	0,198	0,269	0,001
	22	0,224	0,191	0,255	0,001
	23	0,321	0,283	0,387	0,007
Filosofia	24	0,137	0,000	0,365	0,007
	25	0,237	0,000	0,579	0,013
	26	0,266	0,000	0,616	0,020

cont...

cont...					
	27	0,210	0,004	0,370	0,015
	28	0,216	0,115	0,315	0,004
	29	0,187	0,019	0,318	0,006
Biologia	30	0,065	0,000	0,125	0,003
	31	0,181	0,007	0,306	0,006
	32	0,151	0,005	0,254	0,006
	33	0,376	0,074	0,464	0,012
	34	0,200	0,099	0,293	0,005
	35	0,130	0,002	0,223	0,003
	36	0,167	0,134	0,200	0,001
	37	0,142	0,006	0,237	0,005
Física	38	0,187	0,002	0,370	0,010
	39	0,253	0,220	0,287	0,001
	40	0,096	0,000	0,210	0,002
	41	0,096	0,000	0,185	0,005
	42	0,269	0,234	0,302	0,001
	43	0,124	0,039	0,191	0,002
	44	0,183	0,123	0,232	0,001
	45	0,183	0,132	0,229	0,001
Mateática	46	0,091	0,000	0,198	0,002
	47	0,328	0,287	0,367	0,001
	48	0,078	0,002	0,136	0,002
	49	0,218	0,000	0,489	0,001
	50	0,272	0,236	0,305	0,001
	51	0,096	0,002	0,164	0,004
	52	0,225	0,138	0,300	0,002
	53	0,365	0,030	0,591	0,026
Química	54	0,169	0,062	0,236	0,002
	55	0,192	0,162	0,223	0,001
	56	0,289	0,158	0,379	0,003
	57	0,210	0,118	0,282	0,003
	58	0,096	0,034	0,153	0,002

Obs.:Dados em negrito refere-se a HPD que não contém $c=0,25$.

TABELA 7: Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de $\hat{\mu}$ à densidade de probabilidade - *highest probability density*) para o parâmetro \mathbf{c} e respectivo erro de Monte Carlo, para o curso de Engenharia Agrícola.

Disciplina	Ítem	média	I.C. - HPD		Erro de M.C.
			inferior	superior	
Português	1	0,300	0,000	0,617	0,006
	2	0,225	0,000	0,525	0,004
	3	0,228	0,000	0,392	0,004
	4	0,247	0,000	0,548	0,007
	5	0,251	0,003	0,513	0,006
	6	0,203	0,005	0,323	0,002
	7	0,144	0,000	0,338	0,003
	8	0,097	0,025	0,163	0,001
	9	0,289	0,000	0,599	0,006
	10	0,203	0,000	0,439	0,004
Geografia	11	0,232	0,000	0,533	0,004
	12	0,245	0,000	0,523	0,004
	13	0,212	0,000	0,453	0,004
	14	0,234	0,000	0,486	0,004
	15	0,388	0,037	0,707	0,006
	16	0,167	0,000	0,364	0,003
	17	0,243	0,000	0,521	0,005
	18	0,310	0,000	0,641	0,006
História	19	0,256	0,000	0,567	0,005
	20	0,226	0,000	0,457	0,005
	21	0,146	0,021	0,241	0,001
	22	0,137	0,051	0,217	0,001
	23	0,218	0,000	0,403	0,004
Filosofia	24	0,181	0,000	0,430	0,003
	25	0,185	0,000	0,464	0,004
	26	0,191	0,000	0,453	0,004

cont...

cont...					
	27	0,228	0,000	0,387	0,005
	28	0,091	0,000	0,222	0,002
	29	0,175	0,000	0,399	0,003
Biologia	30	0,088	0,019	0,160	0,001
	31	0,154	0,000	0,306	0,002
	32	0,159	0,000	0,293	0,002
	33	0,201	0,000	0,451	0,005
	34	0,320	0,186	0,454	0,002
	35	0,111	0,000	0,229	0,002
	36	0,158	0,026	0,257	0,001
	37	0,187	0,039	0,284	0,001
Física	38	0,149	0,000	0,342	0,002
	39	0,129	0,000	0,248	0,002
	40	0,248	0,003	0,428	0,004
	41	0,149	0,000	0,273	0,002
	42	0,262	0,007	0,382	0,004
	43	0,145	0,000	0,264	0,002
	44	0,173	0,047	0,267	0,001
	45	0,126	0,042	0,195	0,001
Mateática	46	0,205	0,041	0,327	0,002
	47	0,155	0,000	0,335	0,003
	48	0,078	0,000	0,169	0,001
	49	0,251	0,004	0,527	0,007
	50	0,151	0,000	0,260	0,003
	51	0,131	0,020	0,230	0,001
	52	0,216	0,008	0,374	0,002
	53	0,249	0,000	0,510	0,009
Química	54	0,124	0,000	0,223	0,001
	55	0,161	0,088	0,240	0,001
	56	0,136	0,000	0,283	0,002
	57	0,234	0,045	0,354	0,002
	58	0,110	0,023	0,177	0,001

Obs.:Dados em negrito refere-se a HPD que não contém $c=0,25$.

Pelos resultados da Tabela 8, observando-se os intervalos HPDs do parâmetro \mathbf{a} , pode-se observar que, para o curso de Medicina Veterinária, as questões da disciplina de Geografia, exceto a de número 16, foram consideradas de baixo poder de discriminação. O mesmo pode-se dizer com relação à disciplina de Filosofia. Português também teve a maioria de suas questões com essa característica. O oposto ocorreu com as disciplinas de Biologia e Física, e que todas as questões apresentaram poder de discriminação elevado. Com exceção da questão 49, de Matemática e 53, de Química, estas duas disciplinas também atingiram o ótimo poder de discriminação.

TABELA 8: Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de $\hat{\alpha}$ à densidade de probabilidade - *highest probability density*) para o parâmetro \mathbf{a} e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.

Disciplina	Ítem	Média	I.C. - HPD		Erro de M.C.
			inferior	superior	
Português	1	0,624B	0,376	0,882	0,005
	2	0,411B	0,165	0,649	0,003
	3	1,035	0,085	4,283	0,147
	4	0,387B	0,149	0,622	0,004
	5	0,485B	0,255	0,742	0,003
	6	4,757	0,944	11,231	0,060
	7	0,667B	0,394	0,979	0,004
	8	5,044	1,290	11,684	0,059
	9	0,549B	0,273	0,919	0,010
	10	0,711	0,418	1,007	0,004
Geografia	11	0,491B	0,253	0,764	0,003
	12	0,437B	0,169	0,722	0,005
	13	0,412B	0,151	0,676	0,006
	14	0,366B	0,099	0,699	0,011
	15	0,402B	0,189	0,636	0,003
	16	0,859	0,156	3,067	0,089
	17	0,411B	0,166	0,645	0,004
	18	0,903	0,577	1,229	0,007
História	19	0,874	0,564	1,226	0,008
	20	0,358B	0,022	0,835	0,027
	21	1,197	0,272	3,426	0,067
	22	4,772	1,156	11,040	0,072
	23	0,963	0,099	4,313	0,177
Filosofia	24	0,490B	0,219	0,809	0,005
	25	0,762	0,483	1,055	0,006
	26	0,995	0,715	1,320	0,005

cont...

cont...					
	27	0,910	0,090	4,185	0,191
	28	5,894	3,040	9,322	0,096
	29	2,172	0,776	3,685	0,026
Biologia	30	4,964	3,248	6,888	0,031
	31	2,452	0,931	4,216	0,028
	32	5,281	2,940	7,979	0,046
	33	1,079	0,078	4,000	0,121
	34	20,943	9,073	42,628	1,401
	35	5,122	0,987	9,967	0,107
	36	4,524	0,884	10,946	0,076
	37	7,962	0,954	18,403	0,260
Física	38	2,815	0,169	12,603	0,435
	39	3,063	0,387	7,326	0,053
	40	2,913	1,081	5,220	0,034
	41	1,974	0,986	3,069	0,018
	42	4,717	0,346	11,007	0,078
	43	3,228	0,768	5,990	0,047
	44	4,748	1,036	11,826	0,076
	45	2,284	0,522	5,295	0,047
Mateática	46	1,814	0,457	3,411	0,033
	47	2,420	0,170	8,076	0,190
	48	2,414	1,053	4,046	0,028
	49	0,239B	0,018	0,361	0,081
	50	5,259	0,210	12,895	0,137
	51	4,502	2,961	6,165	0,034
	52	7,779	4,149	13,056	0,141
	53	0,384B	0,034	0,625	0,081
Química	54	2,277	0,657	4,286	0,043
	55	6,223	1,599	13,681	0,083
	56	5,104	2,095	8,641	0,071
	57	4,984	3,209	7,484	0,046
	58	2,918	1,517	4,445	0,027

Dados em **negrito** referem-se a HPD que contém $a > 3$: Ótimo poder de discriminação.

B: refere-se a HPD que somente contém $a < 0,5$: Baixo poder de discriminação.

Fazendo-se a análise do grau de dificuldade para este curso, por meio da observação dos resultados da Tabela 10, nota-se que a disciplina que contém as questões mais difíceis é Matemática, seguida pela Física, repetindo o que já foi constatado de forma geral anteriormente. Houve muita variabilidade e relação à questão 49 (Matemática). Isto quer dizer que ela foi considerada muito fácil, para um grupo de candidatos e difícil, para outros. Isso também ocorreu com as questões 53 (Química), 20 (História) e 13 (Geografia).

A disciplina de Biologia, que apresentou maior poder de discriminação, teve apenas duas questões consideradas difíceis, que foram as de número 27 e 33, porém, caracteriza-se por apresentar um grau de dificuldade médio para todas as demais questões.

Não houve nenhuma questão que pudesse ser considerada muito fácil por todos os candidatos (HPD com valores de $b < -2$). Houve questões fáceis nas disciplinas de Português, Geografia, História e Filosofia, mas nenhuma em Biologia, Física, Matemática e Química. No entanto, estas últimas quatro foram as que apresentaram maior poder de discriminação, de acordo com a Tabela 8. Observa-se, com isto, que questões fáceis não implicam, necessariamente, em melhor poder de discriminação.

TABELA 9: Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de $\hat{\mu}$ à densidade de probabilidade - *highest probability density*) para o parâmetro \mathbf{b} e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.

Disciplina	Item	Média	I.C. - HPD		Erro de M.C.
			inferior	superior	
Português	1	-1,871 F	-2,993	-0,956	0,024
	2	-0,496	-2,178	1,538	0,044
	3	2,421	1,555	3,000	0,018
	4	-0,388	-2,153	1,693	0,050
	5	-0,334	-1,317	1,029	0,023
	6	2,650	2,184	3,000	0,005
	7	0,410	-0,182	1,119	0,015
	8	2,686	2,255	3,000	0,006
	9	-1,481	-3,000	0,541	0,079
	10	0,517	0,011	1,091	0,010
Geografia	11	-0,064	-0,958	1,138	0,019
	12	-0,445	-2,124	1,482	0,037
	13	0,424	-0,807	2,046	0,035
	14	1,576	0,335	2,997	0,026
	15	-1,531 F	-2,993	0,400	0,051
	16	2,502	1,843	2,999	0,010
	17	-0,085	-1,327	1,614	0,029
	18	-1,422 F	-2,122	-0,762	0,017
História	19	-1,358 F	-2,047	-0,661	0,018
	20	1,079	-0,832	3,000	0,051
	21	2,553	1,980	3,000	0,009
	22	2,589	2,068	3,000	0,007
	23	2,459	1,652	3,000	0,014
	24	1,190	0,254	2,296	0,017
Filosofia	25	-1,836 F	-2,709	-1,049	0,018
	26	-1,086 F	-1,511	-0,679	0,008

cont...

cont...					
	27	2,391	1,521	3,000	0,017
	28	0,790	0,621	0,989	0,008
	29	1,086	0,762	1,408	0,008
Biologia	30	1,032	0,877	1,179	0,008
	31	1,395	1,074	1,755	0,008
	32	1,002	0,834	1,164	0,008
	33	2,394	1,454	3,000	0,021
	34	0,813	0,663	0,948	0,010
	35	1,700	1,341	2,158	0,012
	36	2,485	1,969	3,000	0,008
	37	1,859	1,481	2,534	0,012
Física	38	2,068	1,443	2,991	0,022
	39	2,574	2,040	3,000	0,007
	40	1,417	1,121	1,739	0,007
	41	1,885	1,411	2,473	0,013
	42	2,544	1,997	3,000	0,007
	43	1,709	1,329	2,186	0,012
	44	2,701	2,299	3,000	0,004
	45	2,525	1,970	3,000	0,008
Mateática	46	1,800	1,313	2,497	0,014
	47	2,510	1,922	2,999	0,009
	48	1,908	1,475	2,490	0,012
	49	0,298	-2,356	3,000	0,056
	50	2,622	2,115	3,000	0,005
	51	1,153	0,985	1,306	0,008
	52	1,041	0,903	1,190	0,007
	53	-0,283	-2,950	2,803	0,110
Química	54	2,155	1,641	2,900	0,015
	55	2,723	2,341	3,000	0,005
	56	1,288	1,109	1,490	0,007
	57	1,129	0,969	1,296	0,008
	58	1,815	1,460	2,251	0,010

Dados em negrito refere-se a HPD que contém $b > 2$: Grau de dificuldade considerado difícil.

F: refere-se a HPD que somente contém $b < -2$: Grau de dificuldade considerado fácil.

TABELA 10: Estimativas a posteriori por ponto e por intervalo (IC-HPD: intervalo de credibilidade de $\hat{\theta}$ a densidade de probabilidade - *highest probability density*) para o parâmetro θ dos 30 primeiros alunos e respectivo erro de Monte Carlo, para o curso de Medicina Veterinária.

Candidato	média	I.C. - HPD		Nota normalizada	Erro de M.C.
		inferior	superior		
1	-1,310	-2,637	-0,012	-1,448	0,015
2	-0,132	-1,205	0,853	-0,828	0,010
3	0,020	-0,821	0,762	-0,621	0,010
4	0,006	-1,035	0,952	-0,931	0,009
5	-0,125	-1,149	0,743	-0,828	0,008
6	0,806	-0,797	1,452	-0,517	0,014
7	1,164	0,823	1,528	0,103	0,007
8	-0,006	-0,906	0,777	-0,517	0,008
9	-1,019	-2,311	0,805	-1,138	0,015
10	-0,390	-1,426	0,550	-1,345	0,010
11	-0,040	-0,923	0,755	-0,414	0,009
12	1,368	1,025	1,738	-0,207	0,008
13	-0,818	-2,087	0,473	-0,724	0,013
14	0,933	0,293	1,399	-0,207	0,007
15	0,371	-1,195	1,282	-0,103	0,012
16	-2,168	-3,445	-0,881	-2,069	0,026
17	0,043	-0,903	0,947	-0,517	0,010
18	1,008	0,539	1,436	0,000	0,007
19	1,468	1,077	1,909	0,517	0,008
20	1,130	0,754	1,541	0,310	0,008
21	-0,460	-1,504	0,544	-1,138	0,011
22	-1,263	-2,634	0,036	-1,241	0,014
23	0,255	-0,651	0,895	-0,517	0,009
24	0,413	-0,589	1,108	-0,310	0,010
25	-0,853	-2,371	0,516	-1,138	0,013
26	-1,207	-2,572	0,102	-1,552	0,011
27	-2,347	-3,842	-1,128	-1,655	0,037
28	-1,494	-2,855	-0,087	-1,448	0,012
29	-0,056	-1,332	1,108	-0,310	0,016
30	0,198	-0,561	0,800	-0,207	0,009

Dados em negrito referem-se às notas normalizadas que se encontram fora do HPD da habilidade estimada.

De acordo com os resultados apresentados na Tabela 10, constata-se que houve alguns candidatos que não tiveram suas notas normalizadas dentro do intervalo HPD. Para estes alunos, a estimativa de suas habilidades não refletiu a nota obtida no vestibular.

4.2 Considerações gerais sobre os parâmetros dos itens e ajuste do modelo

Com base nos dados da Tabela 11, constata-se as correlações entre as estimativas médias dos parâmetros.

TABELA 11: Correlação entre as estimativas médias dos parâmetros **a**, **b** e **c** para os cursos do Vestibular 2006-2 da UFLA.

Ite	a,b	a,c	b,c
Ad inistração	0,38**	0,49**	0,22*
Eng.Ali entos	0,37**	0,37**	-0,13
Agrono ia	0,59**	0,23*	-0,10
C.Biológicas	0,35**	0,03	-0,03
C.Co putação	0,36**	-0,07	-0,11
Eng.Agrícola	0,36**	-0,25*	-0,52**
Eng.Florestal	0,46**	-0,28*	-0,53**
Med.Veterinária	0,30**	0,37**	0,21
Química	0,31**	-0,39**	-0,29*
Zoologia	0,40**	0,00	-0,19

* Significativo, a 5%, pelo teste **t**.

** Significativo, a 1%, pelo teste **t**.

Fazendo-se uma análise das correlações entre as estimativas médias dos parâmetros **a** e **b**, observa-se que a correlação foi significativa e altamente positiva. Isso significa que quanto mais difícil a questão, maior seu poder de discriminação.

Para as estimativas dos parâmetros **a** e **c**, dos cursos de Ciências Biológicas, Ciências da Computação e Zootecnia, elas não estão significativamente correlacionadas. Para os cursos de Ad inistração, Engenharia de Ali entos, Agrono ia e Medicina Veterinária, esta correlação é significativamente positiva e, para os cursos de Engenharia Agrícola, Engenharia

Florestal e Química, significativamente negativa. Em relação aos cursos para as quais esta correlação se mostrou positiva, não significa que quanto maior a probabilidade de acerto casual, maior a discriminação, mas sim que quanto maior o **valor** estimado do parâmetro **c**, maior o **valor** estimado do parâmetro **a**. O que está correlacionado são seus valores estimados e não seus significados. Neste caso, o que acontece é que, quando o valor estimado do parâmetro **c** aumenta, o espaço no eixo de $P(\theta)$, que corresponde a todos os valores de θ da CCI, diminui. Isto faz com que a CCI se torne mais íngreme e para que as diferenças entre as probabilidades de acerto de indivíduos com habilidades diferentes possa ser sentidas, isto é, tenha mais a discriminação. Logo, o seu valor estimado (do parâmetro **a**) aumenta, mas não necessariamente seu poder de discriminação.

A correlação entre as estimativas dos parâmetros **b** e **c** foi positiva para alguns cursos; para outros, foi negativa e, para outros ainda, não correlacionadas. O único curso em que ela se mostrou positivamente significativa, foi para Administração.

Os resultados da Tabela 12 mostram a correlação entre a nota do candidato e sua habilidade média estimada por Pearson e por Spearman. Verifica-se que ela foi muito boa em ambos os casos, pois alcançou valores bem elevados, o que contribuiu para maior confiabilidade no processo de estimação. Com base nesta tabela, pode-se considerar que o modelo foi bem ajustado.

TABELA 12: Correlação da nota do aluno \times habilidade média estimada

Curso	Correlação de Pearson	Correlação de Spearman
Administração	0,83	0,83
Eng.Aliados	0,93	0,92
Agronomia	0,90	0,89
C.Biológicas	0,87	0,92
C.Computação	0,90	0,87
Eng.Agrícola	0,91	0,95
Eng.Florestal	0,91	0,91
Med.Veterinária	0,93	0,92
Química	0,79	0,80
Zoologia	0,87	0,87

Analisando, agora, os gráficos que se encontram nas Figuras 9 a 11 e a Tabela 13.

TABELA 13: Média dos parâmetros discriminação (**a**), dificuldade (**b**) e acerto casual (**c**), por disciplina.

Disciplina	a	b	c
Português	1,45(M)	0,89(M)	0,216
Geografia	0,80(B)	0,40(F)	0,227
História	1,86(M)	1,48(M)	0,206
Filosofia	0,79(B)	-1,00(F)	0,217
Biologia	3,57(O)	1,22(M)	0,202
Física	2,66(A)	1,97(M)	0,179
Matemática	2,44(A)	2,13(D)	0,178
Química	2,93(A)	1,47(M)	0,188
$a > 3$: O (Ótima)		$b > 2$: D (difícil)	
$2 < a \leq 3$: A (Alta)		$0,5 < b \leq 2$: M (Média)	
$1 < a \leq 2$: M (Média)		$-2 < b \leq 0,5$: F (Fácil)	
$a \leq 1$: B (Baixa)		$b \leq -2$: MF (Muito fácil)	

Pode-se observar, por meio das Figuras 9 a 11 e resultados na Tabela 13, que as questões que apresentaram maior grau de dificuldade se encontram nas disciplinas de Matemática e Física, tendo estas a menor probabilidade de acerto ao acaso. Porém, não foram as que apresentaram maior discriminação. Maior poder de discriminação foi apresentado pelas questões de

Biologia e Química, as quais apresentaram uma probabilidade de acerto ao acaso razoável, assim como o grau de dificuldade também é razoável.

Já as questões de Filosofia e Geografia foram as disciplinas para as quais houve probabilidade de acerto ao acaso, consideradas as mais fáceis e de menor discriminação.

De forma geral, pode-se fazer as seguintes considerações sobre cada parâmetro:

1) acerto casual (c): em média, nenhuma disciplina obteve um valor maior que 25% para este parâmetro. Isto indica que o Vestibular 2006-2 da UFLA obteve um baixo índice de acerto casual, favorecendo a obtenção de notas baixas.

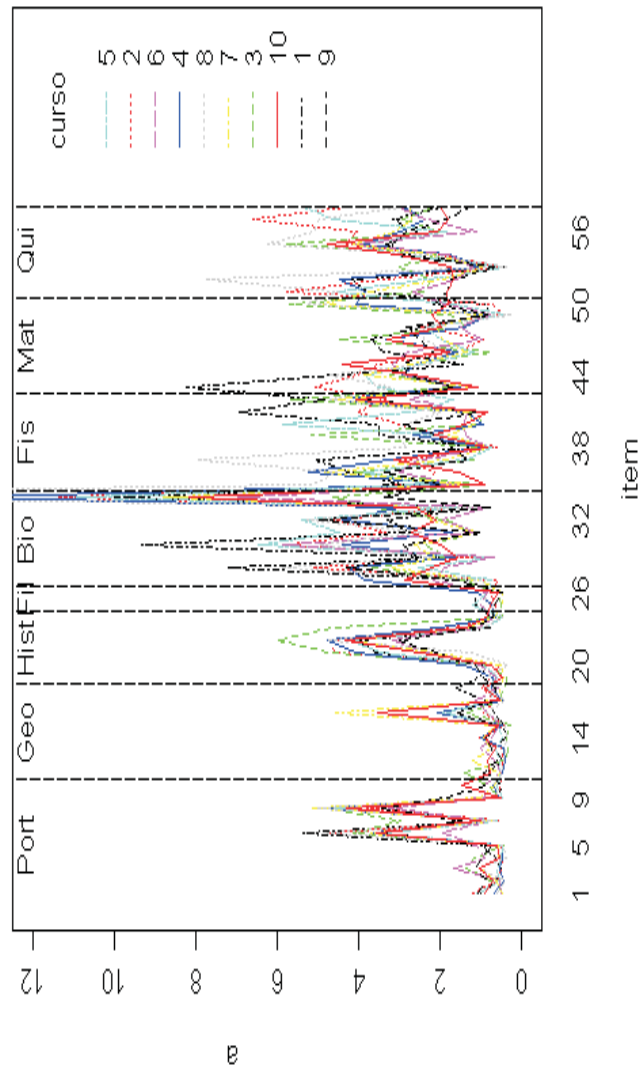


FIGURA 9: Poder de discriminação a para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

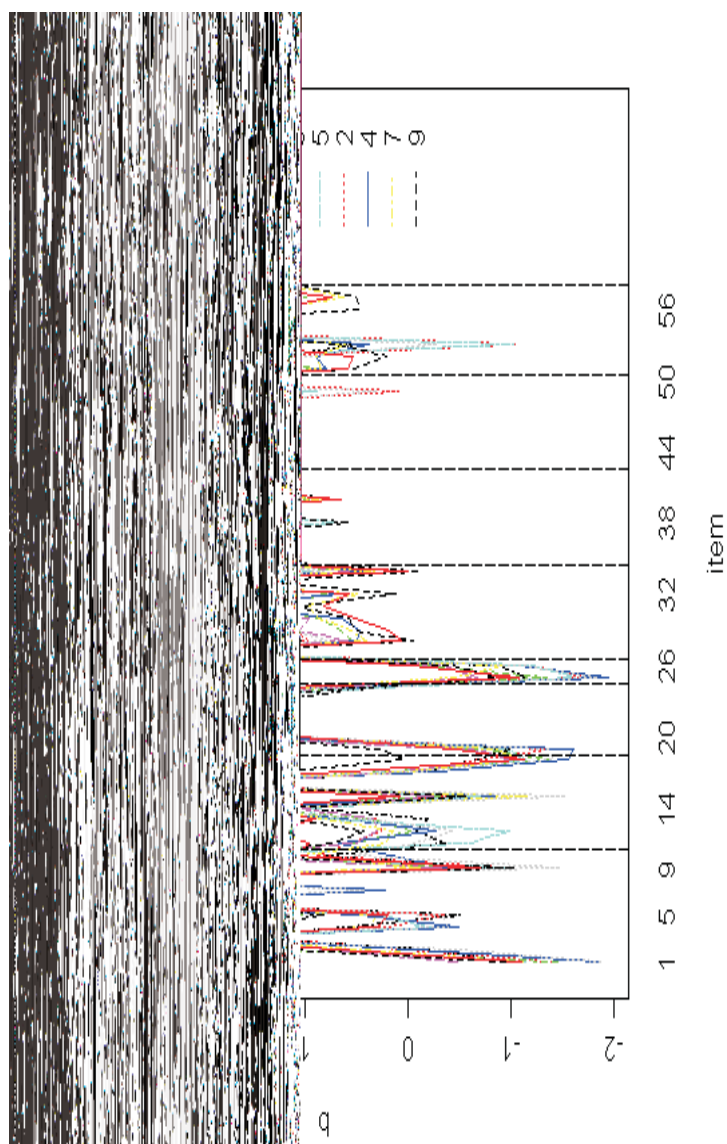


FIGURA 10: Grau de dificuldade \mathbf{b} para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

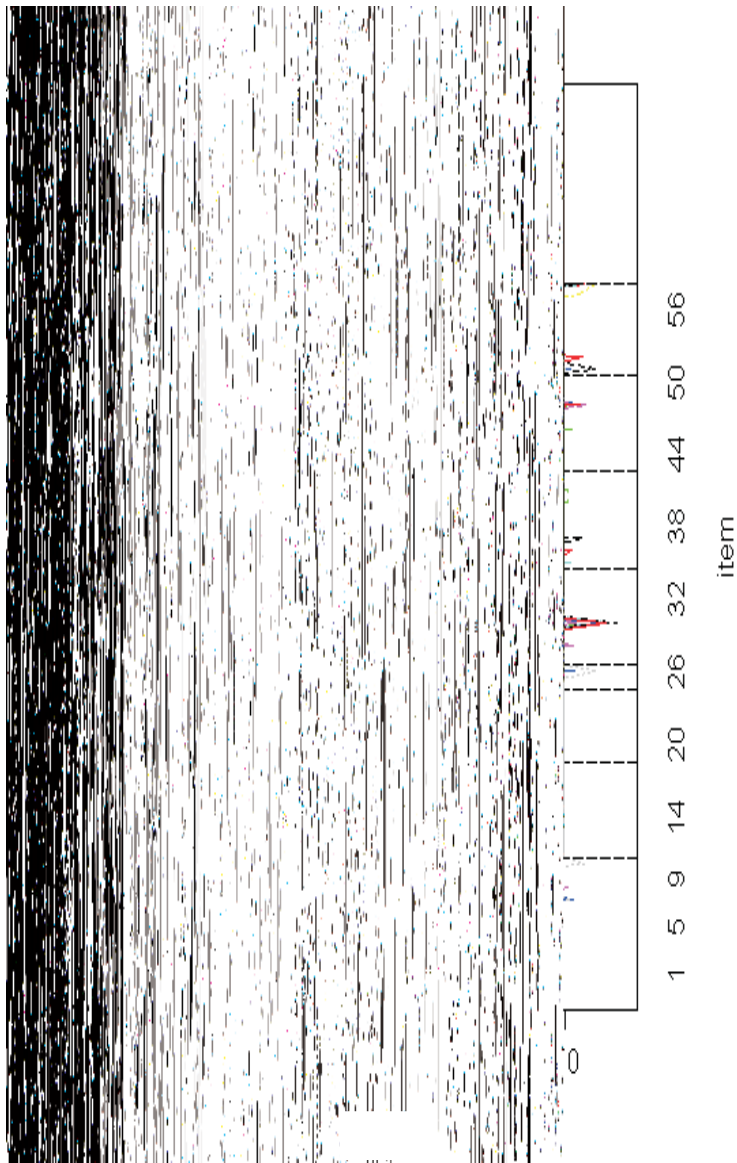


FIGURA 11: Probabilidade de acerto casual c para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agricultura), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

4.3 Considerações específicas sobre os parâmetros dos itens

4.3.1 Considerações para a disciplina de Biologia e Matemática

A escolha dessas duas disciplinas se deve ao fato de terem sido as que se caracterizaram por apresentar maior poder de discriminação e maior dificuldade, respectivamente.

Analisando-se os parâmetros dos itens da disciplina de Biologia, cujos gráficos se encontram nas Figuras de 13 a 15, as questões que obtiveram maior valor para o parâmetro discriminação foram as questões 34, 30 e 28, respectivamente, tendo todas elas um grau médio de dificuldade. A questão 34 foi considerada a mais fácil dentre todas.

Quanto à probabilidade de que se acertasse ao acaso, a questão 34 foi a que apresentou maior probabilidade e a questão 30, a menor.

A questão considerada a mais difícil foi a questão 27, que apresentou uma boa discriminação.

Na Figura 12 está reproduzida a questão de número 30, que se destacou por ter um alto poder de discriminação e pouca probabilidade de acerto casual.

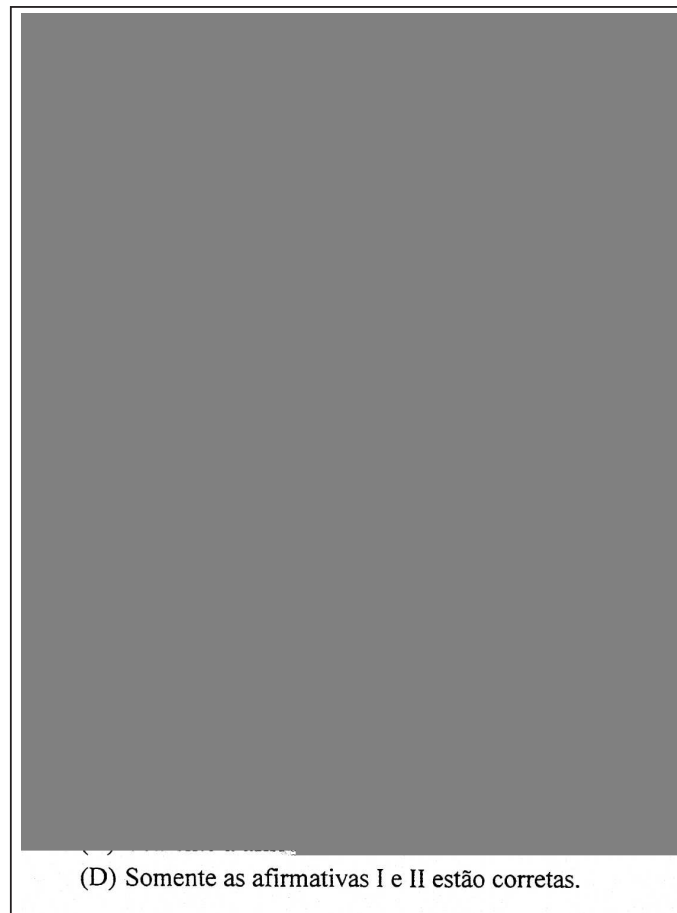


FIGURA 12: Questão 38 do Vestibular 2006-2 da UFLA, correspondente ao item 30 e estudo.

Os valores para os parâmetros desse item, de forma geral e para o curso que mais e menos discriminou, foram:

Geal: $a_{30} = 4,67$; $b_{30} = 0,87$; $c_{30} = 0,09$;

Administração: $a_{30} = 9,31$; $b_{30} = 1,09$; $c_{30} = 0,10$;

Zootecnia: $a_{30} = 2,01$; $b_{30} = 0,51$; $c_{30} = 0,06$;

Esta questão se destacou, dentro desta disciplina, pelo fato de ter obtido valor considerado o poder de discriminação, médio grau de difi-

dade e baixa probabilidade de acerto ao acaso. Para os candidatos do curso de Zootecnia, pode-se concluir que eles conhecem o assunto (**b** baixo). Os alunos de Ciências Humanas tiveram um pouco mais de dificuldade, sendo que, para este curso, esta questão teve um alto poder de discriminação. A questão é pouco trabalhosa, mas rica em conceitos e bem formulada.

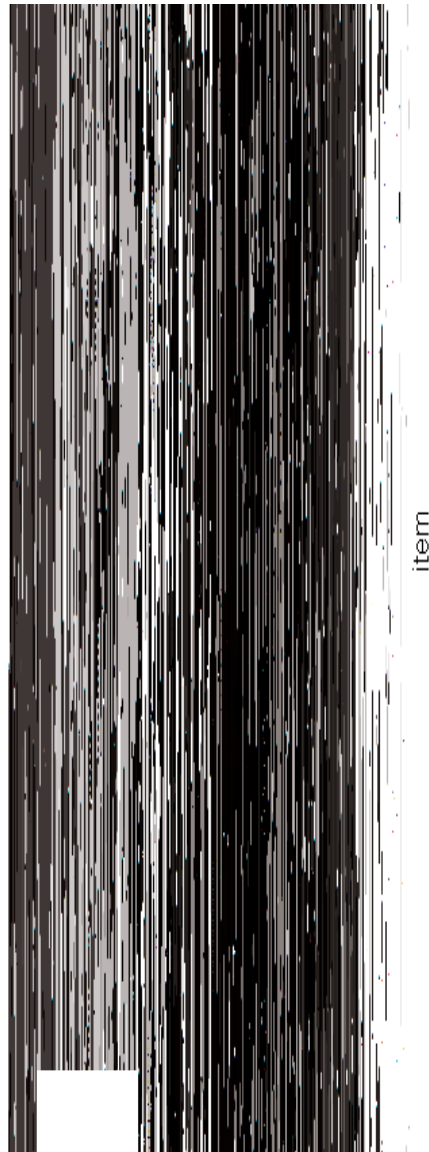


FIGURA 13: Poder de discriminação (a) da prova de Biologia para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

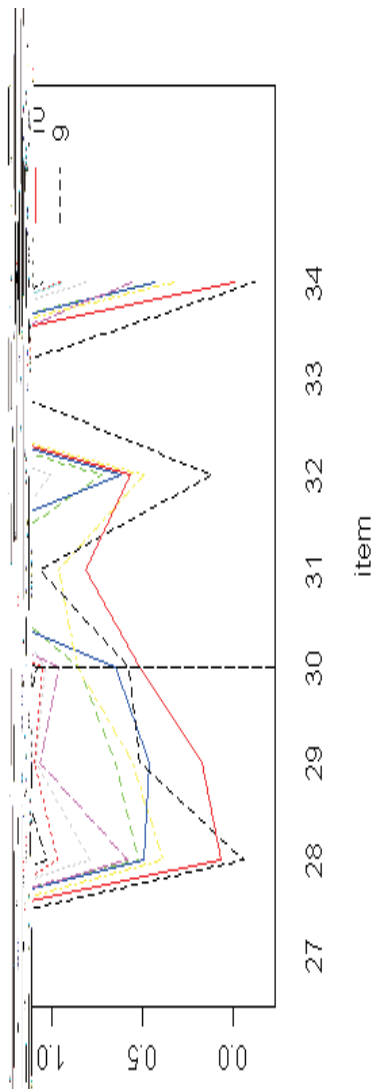


FIGURA 14: Grau de dificuldade (b) da prova de Biologia para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

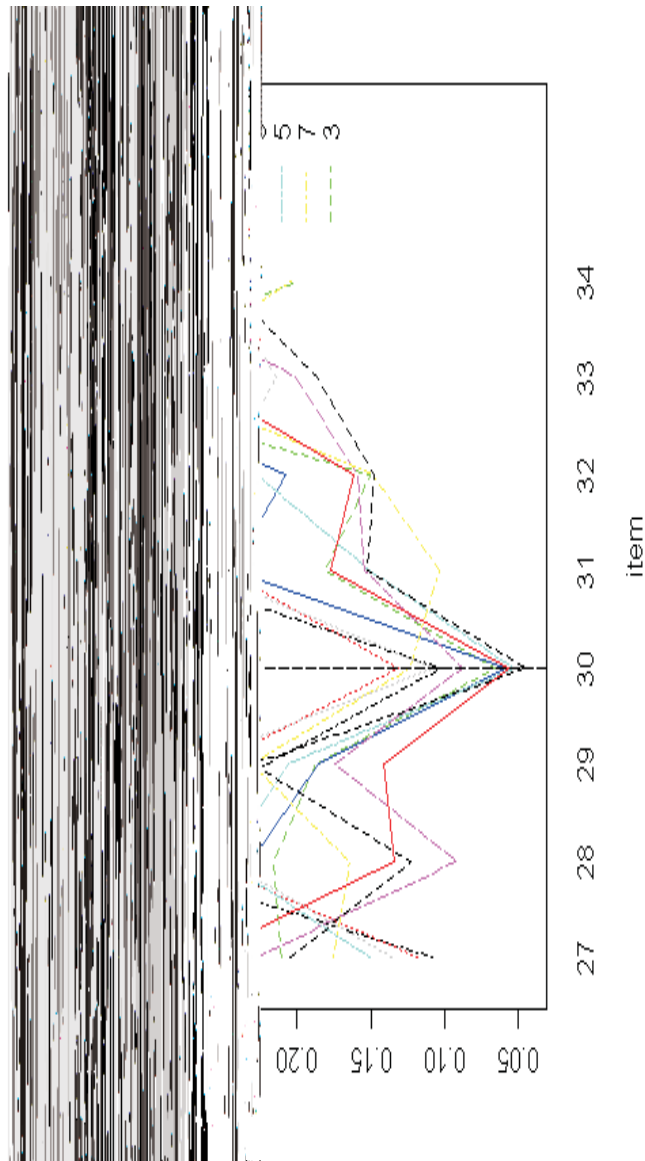


FIGURA 15: Probabilidade de acerto casual (c) da prova de Biologia para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

Fazendo-se a análise para a disciplina de Matemática, cujos gráficos se encontram nas Figuras de 18 a 20, a questão mais difícil foi a de número 44, apresentando um bom grau de discriminação e baixa probabilidade de acerto ao acaso. Houve muita probabilidade de acerto ao acaso para a questão 47 e pouco para a questão 48. A questão 49 foi a considerada mais fácil e com baixo poder de discriminação.

As questões 44 (mais difícil) e 49 (mais fácil), estão reproduzidas nas Figuras 16 e 17.



FIGURA 16: Questão 52 do Vestibular 2006-2 da UFLA, correspondente ao item 44 e estudo.

Valores para os parâmetros desse item, de forma geral e para os cursos nos quais teve maior e menor grau de dificuldade:

Geal: $a_{44} = 3,55; b_{44} = 2,55; c_{44} = 0,16;$

Med.Veterinária: $a_{44} = 4,75; b_{44} = 2,70; c_{44} = 0,16;$

Zootecnia: $a_{44} = 2,68; b_{44} = 2,42; c_{44} = 0,14;$

Esta questão envolve vários conceitos simultâneos: determinante e suas propriedades, inversa de matrizes e resolução de sistemas. Isso faz com

que ela seja uma questão difícil (para todos os cursos, o valor desse parâmetro foi superior a 2, sendo o menor deles, 2,42). O fato de ter sido pedido para se assinalar a alternativa incorreta, obriga o candidato a resolver todas as alternativas, pois ele precisa descobrir qual alternativa não está correta e não apenas resolver a questão e buscar qual delas corresponde à solução que ele encontrou. Portanto, ou ele sabe ou não sabe, isto é, tem um bom poder de discriminação. Este fator, talvez, advinha do fato de que o conteúdo de matrizes é muito extenso e, muitas vezes, o aluno não o viu por completo no ensino médio. Quanto ao seu enunciado, está bem claro e objetivo.

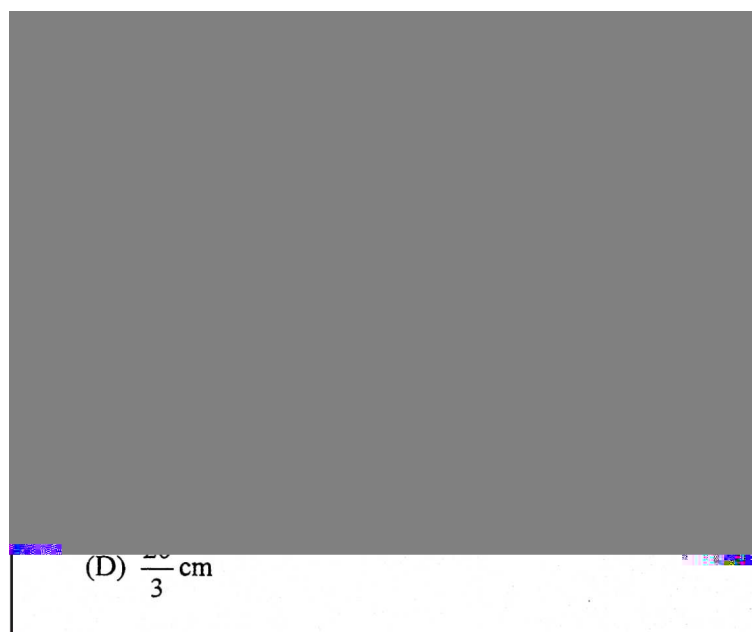


FIGURA 17: Questão 57 do Vestibular 2006-2 da UFLA, correspondente ao item 49 e estudado.

Valores para os parâmetros desse item, de forma geral e para os cursos nos quais teve maior e menor grau de dificuldade:

Geal: $a_{49} = 0,92$; $b_{49} = 1,33$; $c_{49} = 0,21$;

Zootecnia: $a_{49} = 2,19$; $b_{49} = 2,33$; $c_{49} = 0,29$;

Eng. de Alimentos: $a_{49} = 0,45$; $b_{49} = 0,09$; $c_{49} = 0,17$;

Esta questão é considerada uma questão fácil, já que poucos conceitos são cobrados por ela: ângulos, perímetro de circunferência e regra de três, conteúdo de 6^a série do ensino fundamental. Porém, seu diagrama aparenta ser complicado, o que, talvez, tenha sido a razão de se elevar seu grau de dificuldade de baixo para mediano (sendo, no entanto, mais fácil dentre as questões de Matemática), resultando em um alto índice de acerto casual e baixo poder de discriminação. Cabe ressaltar que o único curso para

o qual ela apresentou uma boa discriminação foi para o de Zootecnia. Também somente para o curso de Agronomia é que foi considerado como médio poder de discriminação (1,73). Para todos os demais cursos, os valores desse parâmetro foram inferiores a 1.

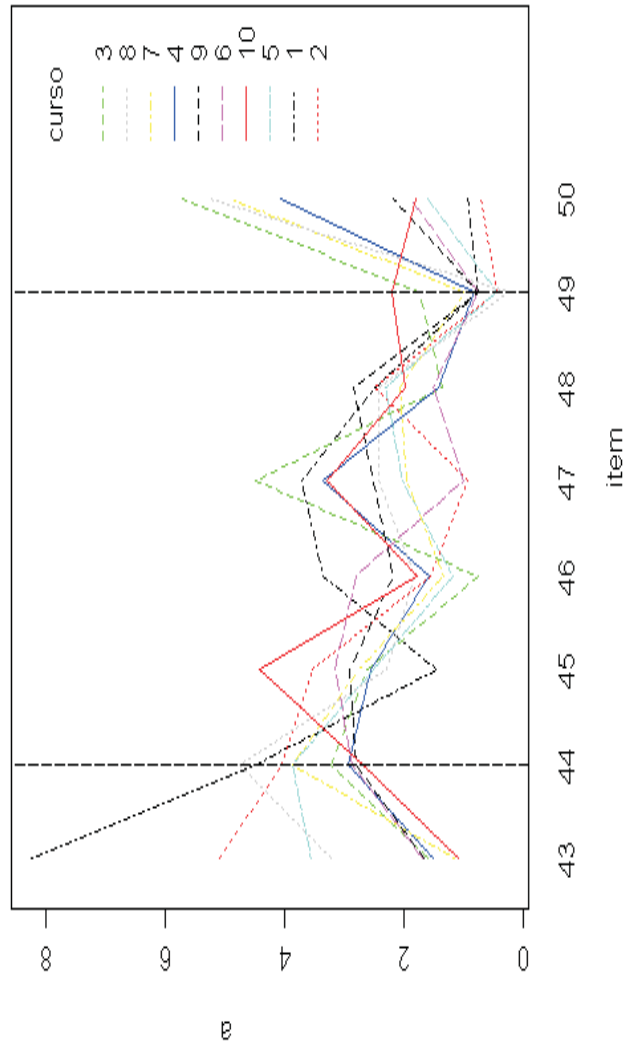


FIGURA 18: Poder de discriminação (a) da prova de Matemática para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

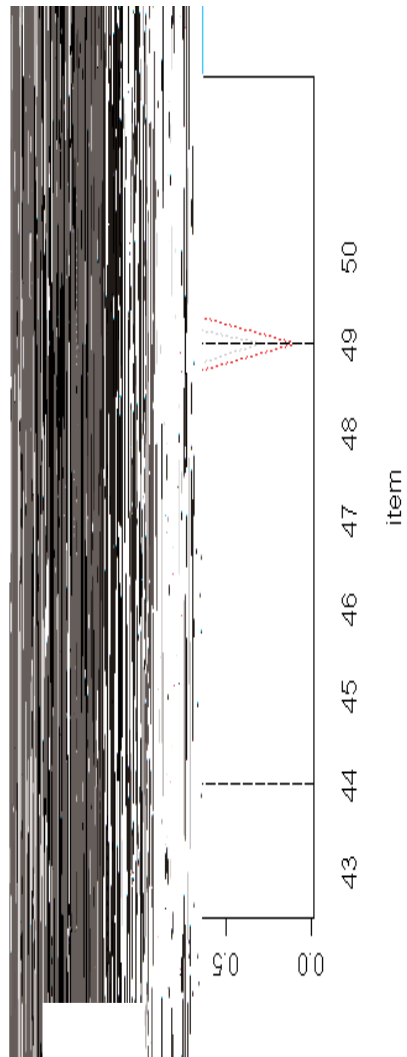


FIGURA 19: Grau de dificuldade (b) da prova de Matemática para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agricultura), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

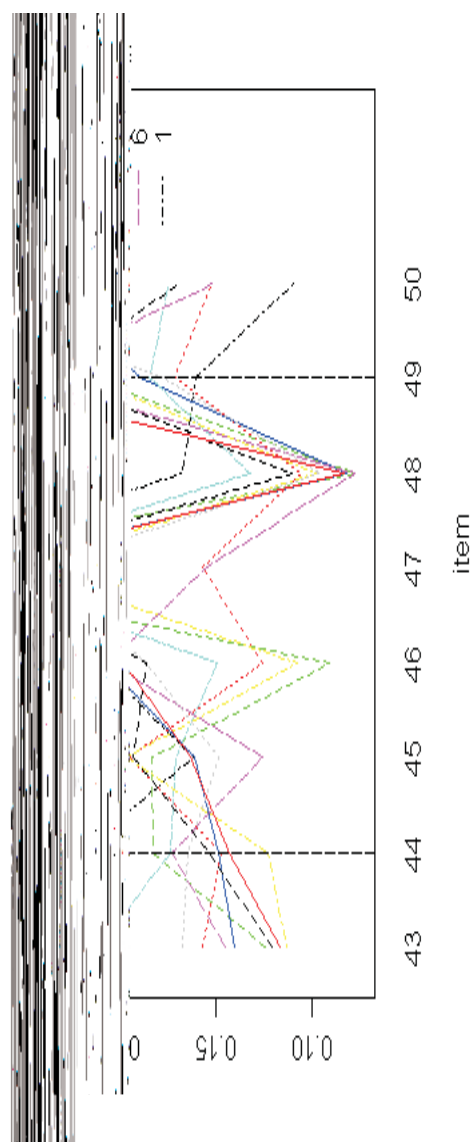


FIGURA 20: Probabilidade de acerto casual (c) da prova de Matemática para os cursos de 1 (Administração), 2 (Engenharia de Alimentos), 3 (Agronomia), 4 (Ciências Biológicas), 5 (Ciências da Computação), 6 (Engenharia Agrícola), 7 (Engenharia Florestal), 8 (Medicina Veterinária), 9 (Química) e 10 (Zootecnia).

4.3.2 Análise do curso de Medicina Veterinária.

Nas Figuras de 21 a 23, encontram-se os gráficos de acerto casual, grau de dificuldade e poder de discriminação para cada questão.

Observa-se que houve muitos itens com valores de discriminação acima de 3,00, o que é considerado um valor de ótimo poder de discriminação, tendo o item 49 (Matemática) sido o que menos discriminou, neste curso.

Um fato que se observa é quanto à disciplina de Geografia, que se caracterizou pela baixa discriminação apresentada pelas suas questões (de 11 a 18).

Quanto à probabilidade de acerto casual, a questão 26 (Filosofia) foi a que apresentou menor valor, sendo, no entanto, uma questão fácil e de baixa discriminação. A questão 58 também obteve um valor muito baixo para esse parâmetro. A de maior probabilidade foi a questão 53 (Química), de dificuldade média e também de baixo poder de discriminação.

Houve muitas questões com bom grau de dificuldade, sendo as mais fáceis, a de número 1 (Português), com baixa discriminação e baixa probabilidade de acerto casual.

Para este curso, a questão 30 (Biologia) se apresentou com o uma questão com alta discriminação, grau de dificuldade médio e baixa probabilidade de acerto casual.

Os valores para os parâmetros destes itens encontram-se na Tabela 14.

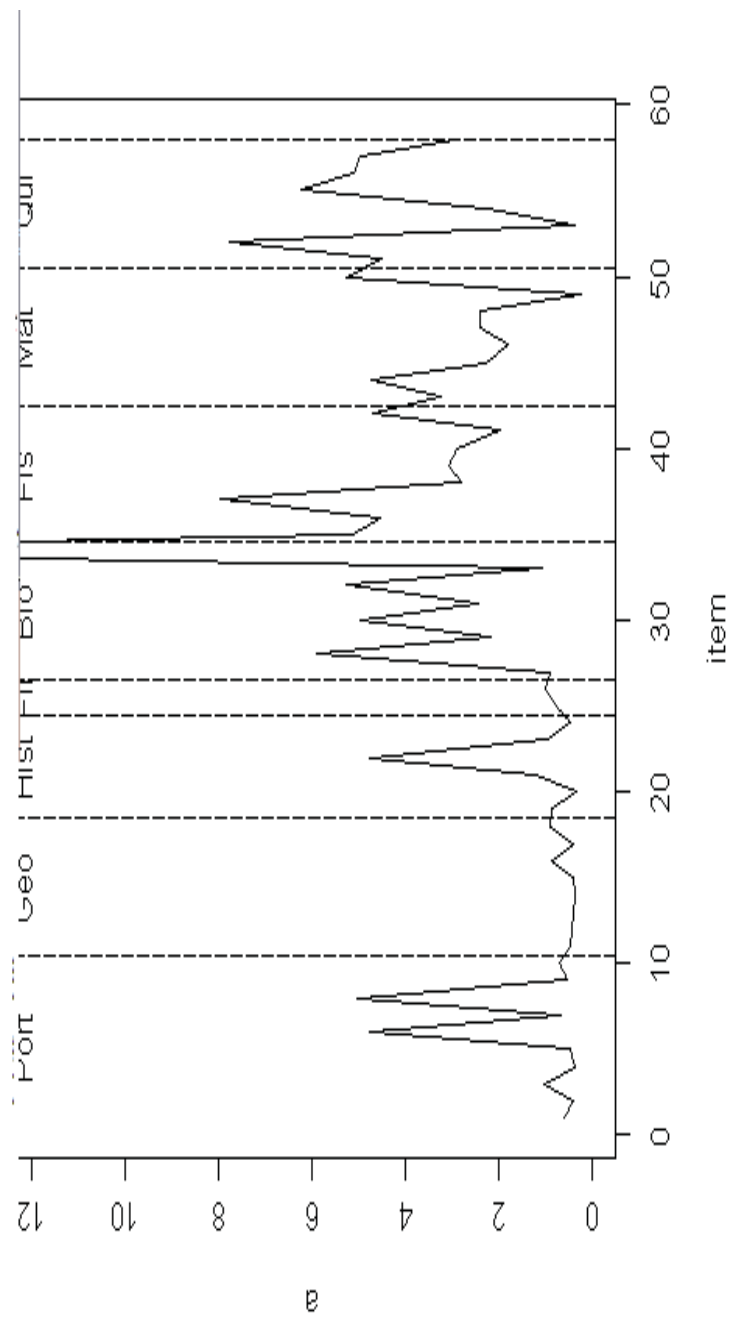


FIGURA 21: Poder de discriminação a para o curso de Medicina Veterinária.

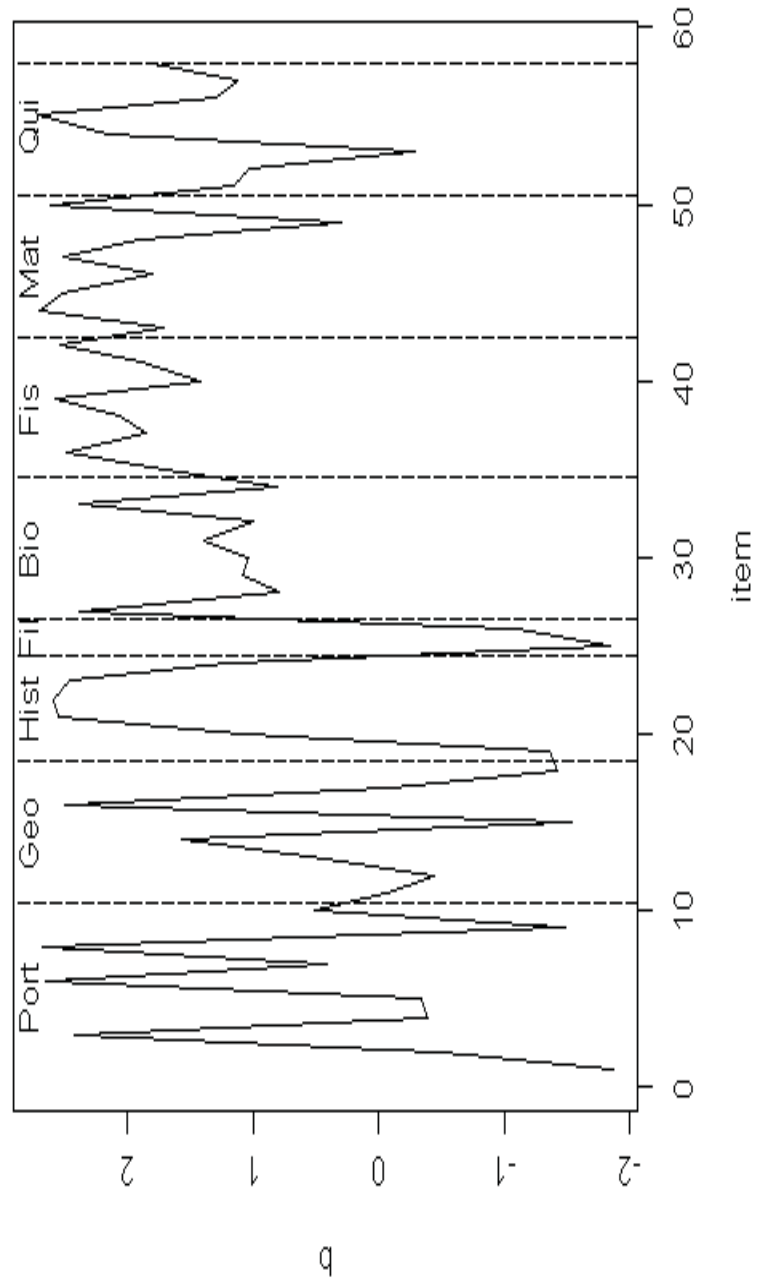


FIGURA 22: Grau de dificuldade **b** para o curso de Medicina Veterinária.

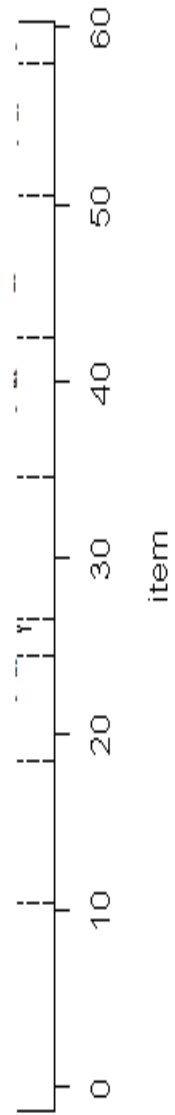


FIGURA 23: Probabilidade de acerto casual c para o curso de Medicina Veterinária.

TABELA 14: Estimativas a *posteriori* por ponto e por intervalo (IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade - *highest probability density*), para os parâmetros **a** e **b** de alguns itens do Curso de Agronomia e respectivo erro de Monte Carlo.

Itm	Parâmetro	média	I.C. - HPD		Erro de M.C.
			inferior	superior	
1	a	0,624	0,376	0,882	0,005
	b	-1,871	-2,993	-0,956	0,024
	c	0,151	0,000	0,390	0,004
26	a	0,995	0,715	1,319	0,005
	b	-1,086	-1,511	-0,679	0,008
	c	0,067	0,000	0,191	0,002
30	a	4,964	3,248	6,888	0,031
	b	1,032	0,877	1,179	0,008
	c	0,108	0,074	0,143	0,001
49	a	0,239	0,018	0,361	0,081
	b	0,298	-2,356	3,000	0,056
	c	0,178	0,000	0,379	0,008
53	a	0,383	0,034	0,625	0,081
	b	-0,283	-2,945	2,803	0,110
	c	0,327	0,000	0,606	0,011
58	a	2,918	1,517	4,444	0,027
	b	1,815	1,460	2,251	0,010
	c	0,086	0,059	0,115	0,001

As cadeias geradas pelo método MCMC para os parâmetros do item 58 encontram-se representadas na Figura 24.

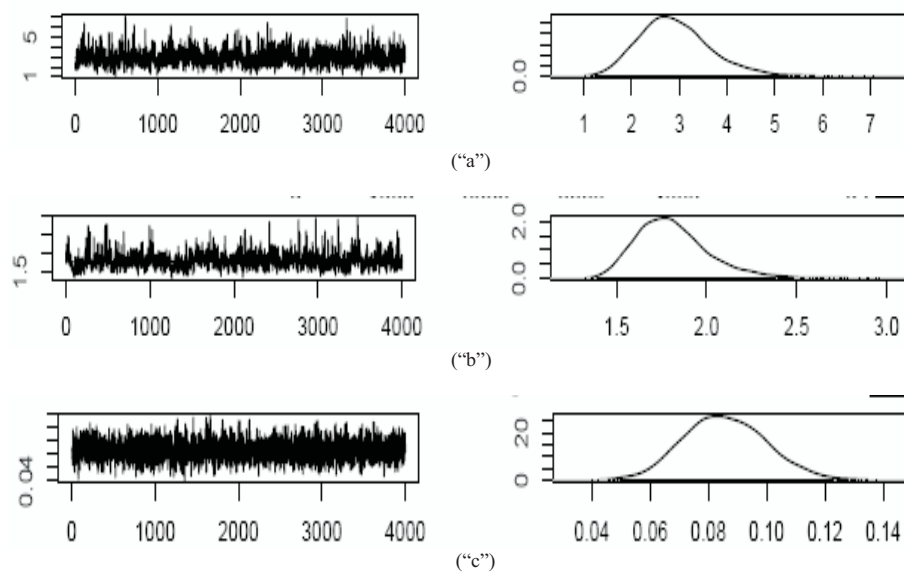
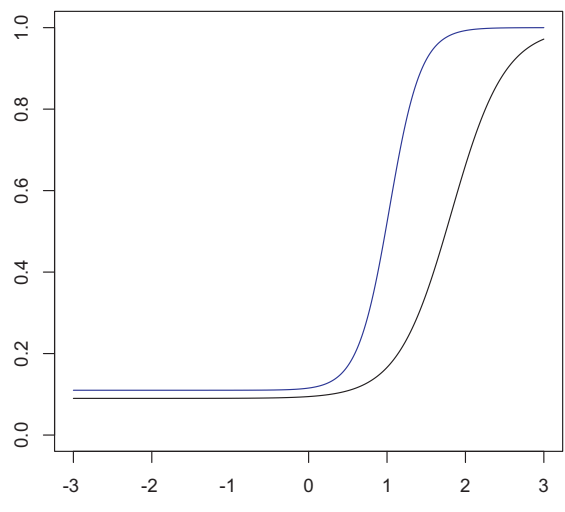


FIGURA 24: Representação gráfica das cadeias geradas pelo algoritmo Metropolis-Hastings e da densidade *a posteriori* dos parâmetros **a**, **b** e **c**, para o item 58.

A partir da observação destes gráficos não se nota irregularidades no processo de geração das amostras. A distribuição *a posteriori* para cada parâmetro, (**a**, **b** e **c**), mostrou um comportamento aproximadamente simétrico. Este item apresentou alto poder de discriminação (2,92), um grau de dificuldade médio (1,82) e baixíssima probabilidade de acerto casual (0,09).

A CCI deste item, juntamente com a CCI dos itens 30 e 49, está representada na Figura 25.



consegue discriminar entre candidatos, mesmo que estes tenham habilidades bem discrepantes.

Alguns exemplos de cadeias geradas pelo MCMC estão representados na Figura 26.

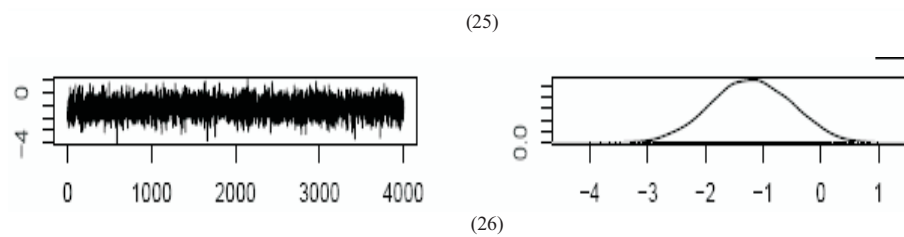
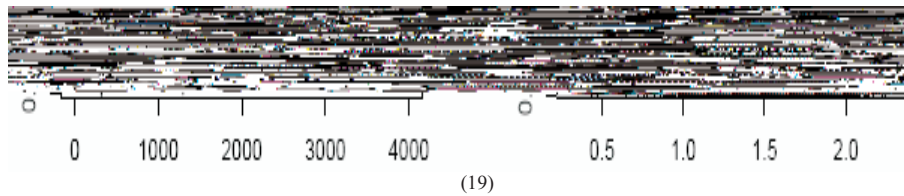


FIGURA 26: Representação gráfica das cadeias geradas pelo Metropolis-Hastings e da densidade *a posteriori* do parâmetro θ dos candidatos 19, 25 e 26.

Por meio destes gráficos, pode-se observar que o processo amostral não apresentou flagrantes irregularidades, sendo a distribuição *a posteriori* do parâmetro θ , aproximadamente, simétrica. Observa-se que o candidato 19 teve uma habilidade estimada acima da média, enquanto que os candidatos 25 e 26, abaixo. Os valores para estas estimativas, por ponto e por intervalo, destes três candidatos, estão na Tabela 15.

TABELA 15: Estimativas *a posteriori* por ponto e por intervalo (I.C.: intervalo de credibilidade, HPD: *highest probability density*), para o parâmetro θ de alguns candidatos do curso de Medicina Veterinária e erro de Monte Carlo.

Candidato	média	I.C. HPD		Erro de M.C.
		inferior	superior	
19	1,468	1,077	1,909	0,008
25	-0,853	-2,371	0,516	0,013
26	-1,207	-2,572	0,102	0,011

O gráfico da correlação entre a nota dos vestibulandos e sua habilidade estimada estão representados na Figura 27.

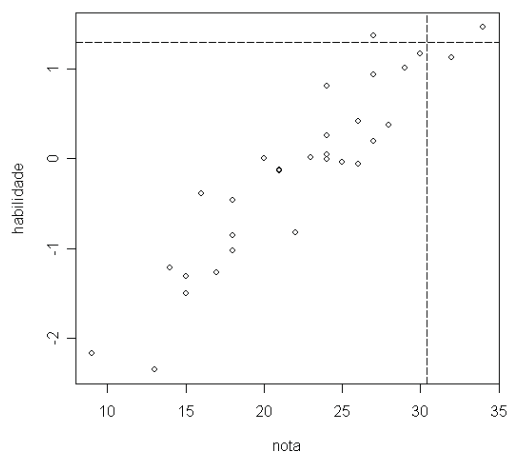


FIGURA 27: Gráfico da correlação entre a nota do vestibulando e sua respectiva habilidade estimada.

Por meio deste gráfico, pode-se observar que a correlação foi muito boa. Seu valor foi de 0,92. Isso faz com que haja mais segurança quanto aos resultados do processo avaliativo realizado.

As linhas pontilhadas têm por objetivo destacar os dois melhores candidatos quanto à nota e quanto à habilidade. Nota-se que há algumas discrepâncias. Por exemplo, dentre os dois candidatos de maior habilidade,

apenas u esteve entre os que tivera melhores notas. Isso alerta para o fato de que u a classificação feita levando em consideração apenas a nota, pode levar a más escolhas.

5 CONCLUSÕES

Para o Vestibular 2006-2 da UFLA, pode-se resumir os principais resultados como sendo os seguintes: os modelos de TRI permitiram analisar em detalhes aspectos de discriminação, dificuldade e acerto casual por provas e por cursos. Suas questões apresentaram correlação positiva entre o grau de dificuldade e o poder de discriminação, indicando que questões mais fáceis não levam a aumento no poder de discriminação. Em geral, a probabilidade de acerto casual não está correlacionada ao grau de dificuldade, nem ao poder de discriminação.

Quanto às provas, a disciplina de Biologia apresentou ótimo poder de discriminação; as disciplinas de Química, Física e Matemática, alto; História e Português, médio e Geografia e Filosofia, baixo. Em relação ao grau de dificuldade, a disciplina de Matemática foi considerada difícil; Física, Química, História, Biologia e Português, consideradas com média dificuldade; Geografia e Filosofia, fáceis. Houve baixa probabilidade de acerto casual. Quanto à habilidade estimada, esta apresentou alta correlação com as notas.

Com base nestes resultados, foi possível elaborar as seguintes conclusões:

- a TRI mostrou-se útil para discutir a qualidade das provas e questões no vestibular 2006-2 da UFLA, o que pode servir de referência para desenvolvimentos futuros e planejamento do processo seletivo nesta instituição;
- de forma geral, dissipou-se um conceito previamente estabelecido nesta instituição que associava o grau de dificuldade diretamente ao poder de discriminação dos itens;
- em todas as provas e para todos os cursos, é possível estabelecer diferenças entre os poderes de discriminação das questões, no entanto, apenas um monitoramento com os responsáveis das áreas e especialistas nas provas pode elucidar os motivos e possíveis regularidades que levem a melhorias no processo de elaboração de novos itens.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item**: conceitos e aplicações. SINAPE, 2000.

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

BAKER, F. B. **The basics of item response theory**. 2. ed. Wisconsin: University of Wisconsin, 2001. 172 p.

BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. In: LORD, F. M.; NOVICK, M. R. (Ed.) **Statistical Theories of Mental Test Scores**. Reading, M.A.: Addison-Wesley, 1968.

BOCK, R. D.; ZIMOWSKI, M. F. Multiple Group IRT. In: van der Linden, W. J. ; HAMBLETON, R. K. (Ed.). **Handbook of Modern Item Response Theory**. New York: Springer-Verlag, 1997.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: John Wiley & Sons, 1992. 588 p.

CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hasting Algorithm. **The American Statistician**, Alexandria, v. 49, n. 4, p. 327-335, Nov. 1995.

GAMERMAN, D. **Markov chain Monte Carlo**: Stochastic Simulation for Bayesian Inference. London: Chapman & Hall, 1997.

GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Alexandria, v. 85, n. 410, p. 398-409, June 1990.

- JEFREYS, H.; **Theory of probability**. Oxford: Clarendon, 1961.
- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. **Fundamentals of Item Response Theory**. Newbury Park: Sage Publications, 1991.
- HAMBLETON, R. K.; COOK, L. L. Latent trait models and their use in the analysis of educational test data. **Journal of Educational Measurement**, Washington, v. 14, n. 2, p. 75-96, 1997.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, Mar. 1970.
- LORD, F. M. **A theory of test scores**. Psychometric Monograph, 7, 1952.
- METROPOLIS, N.; ROSEMBLUT, A. W.; ROSEMBLUT, M. N.; TELLER, A. H.; THELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, New York, v. 21, p. 1087-1092, 1953.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. Tokio: McGraw-Hill Kogakusha, 1974. 564 p.
- O'HAGAN, A. **Kendall's advanced theory of statistics**. London: Arnold, 1994. 330 p. (Bayesian Inference, vol 2B)
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Portugal, 2003. 447 p.
- PLUMMER, M.; BEST, N.; COWLES, K; VINES, K. **The coda Package**. Output analysis and diagnostics for MCMC, GPL Version 2 or later, 2006.

RASCH, G. **Probabilistic models for some intelligence and attainment tests**. Copenhagen: Danish Institute for Educational Research, 1960.

R Development Core Team. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing. Disponível em : <<http://www.R-project.org>>. Acesso em : 10 dez. 2005.

SAMEJIMA, F. A. **Estimation of latent ability using a response pattern of graded scores**. Psychometric Monograph, 17, 1969.

TIERNEY, L. Markov chains for exploring posterior. **The Annals of Statistics** Hayward, v. 22, n. 4, p. 1701-1762, Dec. 1994.

VALLE, R.C. **Teoria da Resposta ao Item**. 1999. 99 p. Dissertação (Mestrado) - Universidade de São Paulo, São Paulo.

WRIGHT, B. D. Sample-free test calibration and person measurement. In: INVITATIONAL CONFERENCE ON TESTING PROBLEMS, 1967, Princeton. **Proceedings ...** Princeton, N. J.: Educational Testing Service, 1968.

7 ANEXOS

ANEXOS	PÁGINA
Anexo A.....	73
Anexo B.....	74

ANEXO A: Tabela com o vetor resposta para a um exemplo de Baker (2001).

TABELA 16: Vetor resposta de um teste de 10 itens e 16 examinados.

		1	2	3	4	5	6	7	8	9	10	RS
	01	0	0	1	0	0	0	0	1	0	0	2
	02	1	0	1	0	0	0	0	0	0	0	2
	03	1	1	1	0	1	0	1	0	0	0	5
E	04	1	1	1	0	1	0	0	0	0	0	4
X	05	0	0	0	0	1	0	0	0	0	0	1
A	06	1	1	0	1	0	0	0	0	0	0	3
M	07	1	0	0	0	0	1	1	1	0	0	4
I	08	1	0	0	0	1	1	0	0	1	0	4
N	09	1	0	1	0	0	1	0	0	1	0	4
A	10	1	0	0	0	1	0	0	0	0	1	3
D	11	1	1	0	1	1	1	1	1	1	1	9
O	12	1	1	1	1	1	1	1	1	1	0	9
	13	1	1	1	0	1	0	1	0	0	1	6
	14	1	1	1	1	1	1	1	1	1	0	9
	15	1	1	0	1	1	1	1	1	1	1	9
	16	1	1	1	1	1	1	1	1	1	1	10

FONTE: Baker (2001), pg. 132-2-133.

ANEXO B: Programa utilizado no procedimento de obtenção de amostras da densidade posterior pelo método MCMC.

Programa para o Curso de Agronomia-AG 3 parâmetros (para os demais cursos, o programa só altera na leitura dos dados).

```
dados <- read.table("VAG20062.csv")

n <- length(dados[,1]) q <- 58 escolha <- 4

<-as.matrix(cbind(dados[1:n,4:(26+3)],dados[1:n,(35+3):(66+3)]))

# Chutes iniciais a <- rlnorm(q) b <- runif(q,-3,3) c
<-rbeta(q,1,(escolha-1)) theta <- rnorm(n)

pi <- 0* vb <- pi

for(i in 1:n)
{
  for(j in 1:q)
  {
    pi[i,j] <- c[j]+(1-c[j])/(1+exp(-a[j]*(theta[i]-b[j])))
    vb[i,j] <- (pi[i,j]^( [i,j]))*((1-pi[i,j])^(1- [i,j]))
  }
}

# Definindo tamanhos da cadeia

N <- 100000 B <- 500 J <- 25

alpha.a <- rep(1,q) alpha.b <- rep(1,q) alpha.c
<-rep(1,q) alpha.theta <- rep(1,n)

Chute <- cbind(t(a),t(b),t(c),t(theta))

# Atualizando a, # Usando priori e geradora de candidatos log-normal

for(count in 1:N)
{
```

```

pic <- pi
vbc <- vb

for(j in 1:q)
{
ac <- rlnorm(1)
pic[,j] <- c[j] + (1-c[j])/(1+exp(-ac*(theta-b[j])))
vbc[,j] <- (pic[,j]^( [ ,j]))*((1-pic[,j])^(1- [ ,j]))
alpha.a[j] <- exp(sum(log(vbc[,j])-log(vb[,j])))
if(alpha.a[j] >= 1){alpha.a[j] <- 1}
if(runif(1) < alpha.a[j])
{
a[j] <- ac
pi[,j] <- pic[,j]
vb[,j] <- vbc[,j]
}
pic <- pi
vbc <- vb
}

# Atualizando b # Usando priori e geradora de candidatos normal
padronizada

for(j in 1:q)
{
bc <- runif(1,-3,3)
pic[,j] <- c[j] + (1-c[j])/(1+exp(-a[j]*(theta-bc)))
vbc[,j] <- (pic[,j]^( [ ,j]))*((1-pic[,j])^(1- [ ,j]))
alpha.b[j] <- exp(sum(log(vbc[,j])-log(vb[,j])))
if(alpha.b[j] >= 1){alpha.b[j] <- 1}
if(runif(1) < alpha.b[j])
{
b[j] <- bc
pi[,j] <- pic[,j]
vb[,j] <- vbc[,j]
}
pic <- pi
vbc <- vb
}

```

```

}

# Atualizando c # Usando priori e geradora de candidatos normal
padronizada

for(j in 1:q)
{
  cc <- rbeta(1,1,(escolha-1))
  pic[,j] <- cc + (1-cc)/(1+exp(-a[j]*(theta-b[j])))
  vbc[,j] <- (pic[,j]^( [ ,j]))*((1-pic[,j])^(1- [ ,j]))
  alpha.c[j] <- exp(sum(log(vbc[,j])-log(vb[,j])))
  if(alpha.c[j] >= 1){alpha.c[j] <- 1}
  if(runif(1) < alpha.c[j])
  {
    c[j] <- cc
    pi[,j] <- pic[,j]
    vb[,j] <- vbc[,j]
  }
  pic <- pi
  vbc <- vb
}

# Atualizando theta # Usando priori e geradora de candidatos normal
padronizada

for(i in 1:n)
{
  thetac <- rnorm(1)
  pic[i,] <- c + (1-c)/(1+exp(-a*(thetac-b)))
  vbc[i,] <- (pic[i,]^( [i,]))*((1-pic[i,])^(1- [i,]))
  alpha.theta[i] <- exp(sum(log(vbc[i,])-log(vb[i,])))
  if(alpha.theta[i] >= 1){alpha.theta[i] <- 1}
  if(runif(1) < alpha.theta[i])
  {
    theta[i] <- thetac
    pi[i,] <- pic[i,]
    vb[i,] <- vbc[i,]
  }
}

```

```

    }
    pic <- pi
    vbc <- vb
  }

  if ( (count-B)%%J == 0 )
  {
    write(t(a), file="CAGA3.lst", q, append=TRUE)
    write(t(b), file="CAGB3.lst", q, append=TRUE)
    write(t(c), file="CAGC3.lst", q, append=TRUE)
    write(t(theta[1:30]), file="CAGT3.lst", n, append=TRUE)
  }
}

```

Programa para gerar estatísticas e gráficos de resultados
 3 parâmetros

```

CadeiaA <- read.table("CAGA3.txt",h=F) CadeiaB
<-read.table("CAGB3.txt",h=F) CadeiaC
<-read.table("CAGC3.txt",h=F) CadeiaTheta
<-read.table("CAGT3.txt",h=F)

pdf(file="Graf3Par.pdf")

librar (coda)

CAm <- mcmc(CadeiaA) CBm <- mcmc(CadeiaB) CCm <- mcmc(CadeiaC) CTm
<- mcmc(CadeiaTheta)

rafter .diag(CAm) rafter .diag(CBm) rafter .diag(CCm)
rafter .diag(CTm)

HPDinterval(CAm) HPDinterval(CBm) HPDinterval(CCm) HPDinterval(CTm)

```

```

summar (CAm) summar (CBm) summar (CCm) summar (CTm)

plot(CAm) plot(CBm) plot(CCm) plot(CTm)

mcpair(CAm) mcpair(CBm) mcpair(CCm) mcpair(CTm)

mediaA <- 1:58 sdA      <- 1:58 mediaB <- 1:58 sdB      <- 1:58
mediaC <- 1:58 sdC      <- 1:58 mediaT <- 1:30 sdT      <- 1:30

for(i in 1:58)
{
  mediaA[i] <- mean(CAm[,i])
  sdA[i]    <- sd(CAm[,i])
  mediaB[i] <- mean(CBm[,i])
  sdB[i]    <- sd(CBm[,i])
  mediaC[i] <- mean(CCm[,i])
  sdC[i]    <- sd(CCm[,i])
}

for(j in 1:30)
{
  mediaT[j] <- mean(CTm[,j])
  sdT[j]    <- sd(CTm[,j])
}

mediaA mediaB mediaC mediaT sdA sdB sdC sdT

```


Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)