



MANOEL VITOR DE SOUZA VELOSO

**CONTRIBUIÇÕES AO ESTUDO DE MODELOS
MULTIVARIADOS NA TEORIA DE RESPOSTA AO
ITEM**

LAVRAS - MG

2014

MANOEL VITOR DE SOUZA VELOSO

**CONTRIBUIÇÕES AO ESTUDO DE MODELOS MULTIVARIADOS NA
TEORIA DE RESPOSTA AO ITEM**

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

Orientador

Dr. Júlio Sílvio de Sousa Bueno Filho

Coorientador

Dr. Caio Lucidius Naberezny Azevedo

**LAVRAS - MG
2014**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Veloso, Manoel Vitor de Souza.

Contribuições ao estudo de modelos multivariados na teoria de resposta ao item / Manoel Vitor de Souza Veloso. – Lavras : UFLA, 2014.

90 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2014.

Orientador: Júlio Sílvio Souza Bueno Filho.

Bibliografia.

1. TRI. 2. Habilidades multivariadas. 3. Análise bayesiana. 4. Dados aumentados. 5. Gibbs sampling. I. Universidade Federal de Lavras. II. Título.

CDD – 519.542

MANOEL VITOR DE SOUZA VELOSO

**CONTRIBUIÇÕES AO ESTUDO DE MODELOS MULTIVARIADOS NA
TEORIA DE RESPOSTA AO ITEM**

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

APROVADA em 15 de agosto de 2014.

Dr. Marcelo Silva de Oliveira

UFLA

Dr. Eric Batista Ferreira

UNIFAL

Dr. João Domingos Scalon

UFLA

Dr. Marcelo Tavares

UFU

Dr. Júlio Sílvio de Sousa Bueno Filho
Orientador

Dr. Caio Lucidius Naberezny Azevedo
Coorientador

LAVRAS - MG

2014

*A Deus,
à minha família,
aos meus irmãos,
aos meus pais,
aos meus amigos,
ofereço.*

*À minha esposa Luciana e à minha sogra Abigail, com carinho
dedico.*

AGRADECIMENTOS

Uma vitória é constituída por muitas etapas e, em cada uma delas, é possível contar com a colaboração, apoio, incentivo, compreensão, torcida e expectativa de várias pessoas que ajudam a fazer a diferença.

Assim, agradeço a muitas pessoas que ajudaram a fazer a diferença: aos funcionários do Departamento de Ciências Exatas onde desenvolvi as minhas atividades, pelo auxílio e ajuda; aos mestres, sem dúvida, dedico grande parte de meu sucesso. Sem os seus ensinamentos e instruções, esse projeto não poderia ser concluído.

Aos Professores Marcelo Tavares, Rogério e Ednaldo, que apontaram a grande possibilidade de cursar uma pós-graduação na UFLA.

Ao meu coorientador Caio Lucidius Naberezny Azevedo e aos mesmos da banca pelas contribuições, paciência e sugestões.

Ao meu orientador Júlio Sílvio de Sousa Bueno Filho um especial muito obrigado: pelas orientações, pelas tantas vezes que se reuniu comigo para discutir a tese, pelas vindas a Varginha quando estudamos em minha casa, por me receber junto dos seus, pela amizade construída durante todos esses anos, minha família, você, sua esposa Giovana e filhos. Sua dedicação é que possibilitou defender essa tese.

Aos amigos da pós-graduação em Estatística e Experimentação Agropecuária que, de alguma forma, colaboraram para que se pudesse desenvolver e concluir esse projeto. As programações, as atitudes, as palavras de incentivo e apoio, as conversas, risadas, viagens e festas que ajudaram a descontrair foram fundamentais para que eu pudesse chegar aqui: Fábio, Walmes, Ivan, Danilo, Crystian, Jair

Prado, Mariele, Tiago, Diana, Marcelo, Enio, Gislene, Letícia e mais um tanto...
Como é bom ter amigos durante essa caminhada. Espero sempre conviver com
vocês!

Aos amigos da UNIFAL, Campus Avançado de Varginha, por me aguentar
desde junho de 2012, angustiado, falando de tese sem parar. Fui chato, eu sei!
Obrigado por aguentarem!

Ao meu pai, meus irmãos e demais familiares por entenderem que para
fazer o Doutorado tive que abrir mão de muita coisa, inclusive de estar junto da-
queles que tanto amo, durante grandes intervalos de tempo. Foi necessário!

À família que tive a felicidade de constituir e que amo demais, minha es-
posa Luciana e minha sogra Abigail, que estão ao meu lado diariamente, conse-
lhando, rindo, orientando, às vezes até adoecendo junto comigo nas horas de tanta
preocupação e pressão e, o mais importante, cuidando de mim, pacientemente.
Desculpe-me por ocupar tanto vocês!

Agradeço à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ní-
vel Superior) pela bolsa concedida para o desenvolvimento desse projeto.

À Universidade Federal de Lavras pela estrutura e oportunidade de par-
ticipar do Programa de Pós-Graduação em Estatística e Experimentação Agrope-
cuária.

“...Nem lembra se olhou pra trás

Ao primeiro passo, aço, aço....

Porque se chamava homem

Também se chamavam sonhos

E sonhos não envelhecem

Em meio a tantos gases lacrimogênicos

Ficam calmos, calmos, calmos

E lá se vai mais um dia... ”

Clube da Esquina II (Flávio Venturini)

RESUMO

Esse estudo teve como principal motivação descobrir a estrutura de dependência entre habilidades nas provas de um exame de seleção, por meio do ajuste de um Modelo Multivariado de Resposta ao Item. Foram comparadas diferentes estruturas de covariâncias em simulações com o modelo proposto. O ajuste do modelo multivariado foi também comparado ao de modelos univariados da TRI em uma aplicação com dados do Vestibular 2006 – 2 da UFLA. A análise bayesiana foi implementada no programa R por meio de algoritmos do tipo Monte Carlo via Cadeias de Markov (MCMC), usando-se o amostrador de Gibbs com dados aumentados. As habilidades em cada prova foram bem estimadas, tanto para o ajuste com o modelo proposto quanto para o ajuste do modelo univariado, porém, no modelo multivariado foi melhor estimada a associação, resultando também em estimativas de maior correlação entre provas de mesma área (a exemplo, Física e Matemática). O programa implementado se mostrou consistente nas simulações e permitiu identificar as vantagens da aplicação do modelo proposto na análise dos dados reais, em substituição aos modelos univariados. As vantagens do modelo multivariado incluem a capacidade de discriminar grupos de habilidades altamente associadas e gerar estimativas mais precisas.

Palavras-chave: Amostrador de Gibbs. Análise bayesiana. Dados aumentados. Habilidades multivariadas. TRI.

ABSTRACT

The aim of this thesis is to discover the dependence structure between the abilities in an admittance test by fitting an Item Response Multivariate Model. Different covariance structures in the proposed model were compared using simulations. Multivariate IRT model was also compared to the univariate models in an application using data from "Vestibular"2006/2 - UFLA. A fully Bayesian analysis was implemented in the R program by means of Markov Chains Monte Carlo (MCMC) type algorithms, using Gibbs Sampling via augmented data. The abilities in each test were well estimated, by both univariate models and proposed multivariate model. However, the multivariate model was more precise and unveil a strong association between the exams with similar subjects, (e.g. Physics and Mathematics). The implemented program was consistent in simulated examples and allowed us to identify the advantages of the application of the proposed model to real data, instead of using univariate models. Advantages of Multivariate model include the capacity to discriminate groups of abilities highly associated between each other and to yield more precise estimates.

Keywords: Augmented data. Bayesian analysis. Gibbs Sampling. IRT. Multivariate abilities.

LISTA DE FIGURAS

Figura 1	Curva Característica do Item - CCI	23
Figura 2	Relação dos parâmetros na Curva Característica do Item - CCI.	24
Figura 3	CCI's com diferentes valores para b e $a = 1$	25
Figura 4	CCI's com diferentes valores para a e $b = 0$	25
Figura 5	CCI's com diferentes valores para c	26
Figura 6	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 1000$ indivíduos	45
Figura 7	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 1000$ indivíduos	46
Figura 8	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 1000$ indivíduos	47
Figura 9	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 1000$ indivíduos	47
Figura 10	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 1000$ indivíduos	48

Figura 11	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 1000$ indivíduos	49
Figura 12	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para Habilidade média em função do número de itens por prova, com $n = 1000$ indivíduos	50
Figura 13	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para a Habilidade média em função do número de itens por prova, com $n = 1000$ indivíduos	50
Figura 14	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 2000$ indivíduos	54
Figura 15	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 2000$ indivíduos	54
Figura 16	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 2000$ indivíduos	55
Figura 17	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 2000$ indivíduos	56

Figura 18	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 2000$ indivíduos	57
Figura 19	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 2000$ indivíduos	57
Figura 20	Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para Habilidade média em função do número de itens por prova, com $n = 2000$ indivíduos	58
Figura 21	Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para a Habilidade média em função do número de itens por prova, com $n = 2000$ indivíduos	59
Figura 22	Traços das cadeias para o parâmetro a associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC, para os itens 27 e 28	62
Figura 23	Traços das cadeias para o parâmetro b associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC, para os itens 27 e 28	62
Figura 24	Traços das cadeias para o parâmetro c associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC, para os itens 27 e 28	63
Figura 25	Traços das cadeias e densidade a posteriori para as correlações entre as habilidades em Português e Geografia (V2) e entre Português e História (V3). Valores da matriz de correlação entre as habilidades. Cadeias associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC	64

Figura 26	Traços das cadeias e densidade a posteriori para as estimativas das correlações entre as habilidades em Português e Geografia (V2) e entre Português e História (V3), dada pela estimativa da matriz de correlação entre as habilidades, no diagnóstico de Gelman & Rubin, obtidas via MCMC	65
Figura 27	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 estimadas com o modelo univariado na prova de Português.	73

APÊNDICES

Figura 28	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 2.	83
Figura 29	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 3.	84
Figura 30	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 4.	85
Figura 31	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 5.	86
Figura 32	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 6.	87
Figura 33	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 7.	88
Figura 34	Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 8.	89

LISTA DE TABELAS

Tabela 1	Configurações dos exames simulados usando diferentes números de indivíduos, provas por exame, itens por prova e estruturas de covariâncias.	37
Tabela 2	Provas do Exame Vestibular 2006 – 2 da Universidade Federal de Lavras (UFLA)	40
Tabela 3	Cursos oferecidos e quantidade de candidatos por curso na Universidade Federal de Lavras (UFLA), no segundo semestre de 2006	41
Tabela 4	Correlação entre valores paramétricos e valores estimados para todas as configurações simuladas ($n = 1000$)	43
Tabela 5	Erro Quadrático Médio (EQM) para os parâmetros estimados em todas as configurações utilizadas ($n = 1000$)	44
Tabela 6	Correlação entre valores paramétricos e valores estimados para todas as configurações simuladas ($n = 2000$)	52
Tabela 7	Erro Quadrático Médio (EQM) para os parâmetros estimados em todas as configurações utilizadas ($n = 2000$)	53
Tabela 8	Correlações entre as habilidades nas provas estimadas pelo modelo multivariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química	63
Tabela 9	Correlações entre as notas obtidas no teste e as habilidades estimadas nas provas pelo modelo multivariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química	66

Tabela 10	Estimativas <i>a posteriori</i> por ponto (média) e por intervalo (IC- <i>HPD</i> com 95%) para os parâmetros alguns itens de todas as provas do Vestibular 2006 – 2 da UFLA, considerados bons (itens destacados em negrito) e ruins (itens sem destaques) . . .	67
Tabela 11	Estimativas <i>a posteriori</i> por ponto (média) e por intervalo (IC- <i>HPD</i> com 95%) para os parâmetros Habilidades dos indivíduos 29,30,31,32 calculadas com o modelo multivariado na prova de Português do Vestibular 2006 – 2 da UFLA	68
Tabela 12	Correlações entre as habilidades nas provas, estimadas pelo modelo univariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química.	69
Tabela 13	Correlações entre as notas obtidas e as habilidades estimadas nas provas pelo modelo univariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química.	70
Tabela 14	Estimativas <i>a posteriori</i> por ponto (média) e por intervalo (IC- <i>HPD</i> com 95%), na análise univariada, para os parâmetros de alguns itens de todas as provas do Vestibular 2006 – 2 da UFLA.	71
Tabela 15	Estimativas <i>a posteriori</i> por ponto (média) e por intervalo (IC- <i>HPD</i> com 95%) para os parâmetros Habilidades dos indivíduos 29,30,31,32 calculadas com o modelo univariado na prova de Português do Vestibular 2006 – 2 da UFLA	72

APÊNDICES

Tabela 16	Estimativas a <i>posteriori</i> por ponto (média) e por intervalo (IC- <i>HPD</i>) para os parâmetros de itens do Vestibular 2006 – 2 da UFLA.	81
Tabela 17	Estimativas a <i>posteriori</i> por ponto (média) e por intervalo (IC- <i>HPD</i>) para os parâmetros de todos os itens de todas as provas do Vestibular 2006 – 2 da UFLA.	82

SUMÁRIO

1	INTRODUÇÃO	18
2	REFERENCIAL TEÓRICO	20
2.1	A Teoria de Resposta ao Item (TRI)	20
2.1.1	Modelos para a Teoria de Resposta ao Item	21
2.1.2	A Curva Característica do Item	23
2.2	Modelos Multidimensionais da Teoria de Resposta ao Item	25
2.2.1	O Modelo Multivariado da TRI para uma única população	28
3	MATERIAL E MÉTODOS	31
3.1	Implementação do Modelo Probit de 3 Parâmetros Multivariado (MP3PM) usando dados aumentados	31
3.2	Diferentes configurações na simulação do modelo multivariado	36
3.3	Dados para aplicação do modelo multivariado: Vestibular 2006-2 da UFLA	40
4	RESULTADOS E DISCUSSÃO	42
4.1	Estudo de simulação do modelo multivariado de resposta ao item	42
4.2	Aplicação do modelo multivariado nos dados do vestibular 2006-2 da UFLA	61
4.3	Comparação do modelo multivariado proposto com a aplicação do modelo univariado nos dados do vestibular 2006 – 2 da UFLA	68
4.4	Considerações gerais sobre os resultados	74
5	CONCLUSÃO	76
	REFERÊNCIAS	77
	APÊNDICE	81

1 INTRODUÇÃO

A Teoria de Resposta ao Item (TRI) trata da aplicação e compreensão de um conjunto de modelos denominados Modelos de Resposta ao Item (MRI). Tais modelos associam características não observáveis de indivíduos (traços latentes), características dos instrumentos de medida (questionários, testes, provas) e as respostas dadas a esses instrumentos. Mais ainda, os MRI modelam a probabilidade de um indivíduo atingir certo escore, dado os traços latentes (habilidades) desse indivíduo e as características desse item.

A TRI vem sendo aplicada nas principais avaliações educacionais, como SAEB, ENADE, ENEM no Brasil e TOEFL, GRE, PISA no mundo. Devido ao aumento de sua aplicação em diferentes campos de pesquisa, tais como educação, marketing, psiquiatria, genética, e com o surgimento de planejamentos mais complexos, é cada vez mais necessário o uso de modelos de resposta ao item mais flexíveis que possam levar em conta as diversas características em estudo. Por isso, na escolha do melhor modelo, dentre os MRI em uma aplicação específica, é necessário usar métodos de estimação adequados e medidas de diagnósticos que permitam avaliar a qualidade de ajuste desses modelos.

Alguns MRI para respostas dicotômicas são usados na TRI, como o Modelo de Ogiva Normal (modelo probito) e os Modelos Logísticos com 1, 2 ou 3 parâmetros. Esses podem ser caracterizados de acordo com o número de parâmetros de itens que pertençam ao modelo. O modelo mais completo é o de três parâmetros de item. Usando restrições nos parâmetros desses modelos, pode-se obter os modelos de dois ou de um parâmetro. O modelo de três parâmetros descreve a probabilidade de um indivíduo responder corretamente a um item, dada sua habilidade em responder aquele item e a três características do próprio item: o poder de discriminação, o grau de dificuldade e o acerto casual.

A qualidade de predição da distribuição das habilidades dos elementos de uma população pode estar associada à possibilidade de se obter melhores estimativas da variabilidade, usando informações dos indivíduos respondentes que possam estar associadas a suas habilidades, como sua condição socioeconômica e o grau de escolaridade, dentre outras. Para isso existem os Modelos Multidimensionais

de Resposta ao Item, os quais assumem que é necessário mais de um traço latente para modelar apropriadamente as probabilidades de resposta. Os Modelos Multivariados de Resposta ao Item, caso particular dos Modelos Multidimensionais consideram a possível estrutura de covariância, ou grau de correlação, existente entre as habilidades de um indivíduo associadas a várias disciplinas. Os dois são modelos que se referem a dois ou mais traços latentes. Existe uma diferença, o modelo multivariado refere-se a traços latentes distintos, avaliados por diferentes testes, sendo que cada item é associada uma única habilidade. Já o modelo multidimensional considera vários traços latentes associados a um item.

Pode ser interessante também estudar indivíduos de diferentes grupos, com diferentes características, próprias de cada um desses grupos. Essas diferentes características podem induzir comportamentos distintos na distribuição dos traços latentes. As distribuições de probabilidades dos traços latentes poderiam ser específicas para cada grupo, fazendo-se necessário, então, considerar a heterogeneidade dos grupos envolvidos.

O presente trabalho diz respeito a um ajuste de um modelo multivariado de resposta ao item usando dados aumentados. Foram feitas simulações para verificar a consistência da rotina R para análise bayesiana desse modelo. Foram comparadas diferentes estruturas de covariâncias sob o modelo multivariado da TRI e foi comparado o modelo univariado com o modelo multivariado da TRI em uma aplicação com dados reais do Vestibular 2006 – 2 da UFLA. Esse estudo teve como motivação melhorar as estimativas para as habilidades com o modelo proposto, descobrir relações entre habilidades nas provas/disciplinas envolvidas, bem como verificar se o modelo multivariado tem alguma vantagem sobre o univariado.

Uma análise bayesiana plena foi implementada por meio de algoritmos do tipo Monte Carlo via Cadeias de Markov (MCMC), usando-se o amostrador de Gibbs com dados aumentados. Toda análise foi programada e executada no software **R** (R CORE TEAM, 2014), com o uso do pacote CODA - *Convergence Diagnosis and Output Analysis* (PLUMMER et al., 2006).

2 REFERENCIAL TEÓRICO

2.1 A Teoria de Resposta ao Item (TRI)

A Teoria de Resposta ao Item (TRI) engloba um conjunto de modelos com a finalidade de representar a probabilidade de uma pessoa obter um certo escore a um determinado item, em função de sua habilidade e das características do item. A TRI surgiu com a necessidade de superar as limitações da Teoria Clássica dos Testes (TCT) que era uma forma de medir as habilidades de indivíduos submetidos a algum teste.

A TCT só permitia a comparação entre indivíduos ou grupos de indivíduos que tivessem sido submetidos a provas equivalentes ou àquelas que produzissem mesma média e desvio-padrão (LORD, 1980). Outras limitações da TCT, segundo Andrade, Tavares e Valle (2000) são:

- a) seus resultados dependem das características dos indivíduos que realizam o teste;
- b) a habilidade de um examinado é estimada por percentuais de acertos ou escores dos testes;
- c) apresenta a dificuldade de comparar resultados de diferentes testes em diferentes situações.

Essas limitações foram resolvidas pela TRI, a qual trata o examinado de forma diferente:

- a) o enfoque das análises desvincula-se das provas e concentra-se nos itens;
- b) a resposta dada ao item depende do nível de habilidade de cada indivíduo;
- c) o item é um estímulo que leva a uma resposta, dependendo do nível da habilidade;
- d) a habilidade de um indivíduo não depende da prova;

- e) as características dos itens não dependem de quem realiza a prova;
- f) a invariância dos parâmetros permite a comparação entre indivíduos, mesmo em provas com itens diferentes;
- g) as respostas aos itens de um teste são medidas por uma relação probabilística, cuja relação gráfica dessa probabilidade é a curva característica do item (CCI).

Lord (1952) é considerado um dos grandes responsáveis por dar as bases da moderna TRI, seguido dos trabalhos de Birnbaum (1968) e Rasch (1960), devido aos modelos por eles elaborados.

O crescimento e divulgação da TRI só começou a partir da década de 80 com o aumento da capacidade de processamento computacional, já que os algoritmos de estimação usados são muito complexos.

No Brasil, a primeira aplicação da TRI foi em 1995 na análise do Sistema Nacional de Avaliação da Educação Básica (SAEB). A partir daí, seu uso está sendo incentivado, principalmente, pelos órgãos de governo como o Ministério da Educação.

2.1.1 Modelos para a Teoria de Resposta ao Item

Uma apresentação genérica dos modelos para respostas dicotômicas da TRI considera uma variável aleatória Y_{ij} associada ao acerto ou erro na resposta do indivíduo i ao item j , $i = 1, \dots, n$ e $j = 1, \dots, m$. Essa variável pode assumir valor 1 (acerto) ou 0 (erro).

Um modelo amostral para Y_{ij} seria a função de probabilidade (fp) de Bernoulli

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

em que p_{ij} está em função dos parâmetros e é a probabilidade condicional de que o indivíduo i responda correto ao item j . Dessa forma, uma classe geral de modelos para respostas dicotômicas é dado por:

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j)F(m_{ij}), \quad (2)$$

em que:

θ_i é o parâmetro de habilidade do indivíduo i ;

a_j é o parâmetro de discriminação do item j ;

b_j é o parâmetro de dificuldade do item j (índice de locação);

c_j é o parâmetro para a resposta de acerto casual do item j ;

$m_{ij} = a_j(\theta_i - b_j)$ é preditor linear;

$F(\cdot)$ é uma função não decrescente, dada pela função de distribuição acumulada (fda) de uma dada distribuição.

Um dos primeiros modelos dessa classe de TRI foi introduzido por Lord (1952), o qual considerou $F(\cdot) = \Phi(\cdot)$, sendo $\Phi(\cdot)$ a função de distribuição acumulada (fda) da distribuição normal padrão. Ao ser determinada a densidade relacionada a essa distribuição, $f(\cdot)$, nota-se que o parâmetro b_j está relacionado à média e o parâmetro a_j à mudança de concavidade. Quando b_j é igual ao parâmetro θ_i , $p_{ij} = F_\theta(m_{ij}) = F(0) = 0,5$.

Birnbaum (1968) introduziu o modelo logístico de dois parâmetros, em que considerou a função de distribuição acumulada da distribuição logística para $F(\cdot)$, isto é, $F(\cdot) = L(m_{ij})$, sendo

$$\begin{aligned} L(m_{ij}) &= \frac{e^{m_{ij}}}{1 + e^{m_{ij}}} \\ &= \frac{1}{1 + e^{-m_{ij}}}. \end{aligned}$$

Se, nesse modelo, for assumido $a_j = 1$, ele fica conhecido como Modelo de Rasch (RASCH, 1960). Em qualquer um dos casos, percebe-se que o parâmetro c_j de acerto casual não está presente no modelo, porém, há a possibilidade disso ocorrer. É o caso das questões de múltipla escolha, as quais permitem que sejam acertadas mesmo por quem não domine o conhecimento. Para tanto, Birnbaum (1968) propôs a introdução do parâmetro c_j no modelo, chamado de modelo logístico de três parâmetros (ML3P). Considerando o ML3P, o cálculo de p_{ij} será dado por:

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \quad (3)$$

2.1.2 A Curva Característica do Item

Um Modelo de 3 Parâmetros da TRI (M3P), seja ele logístico ou probito, pode ser expresso graficamente através da Curva Característica do Item (CCI). Cada item terá sua própria CCI. Essa curva é obtida ao ser plotado p_{ij} em função de θ , para um determinado item j . Na prática, é a proporção de respostas corretas de todos os indivíduos com habilidade θ_i que está relacionada a esse item, como se vê na Fig. 1, a seguir.

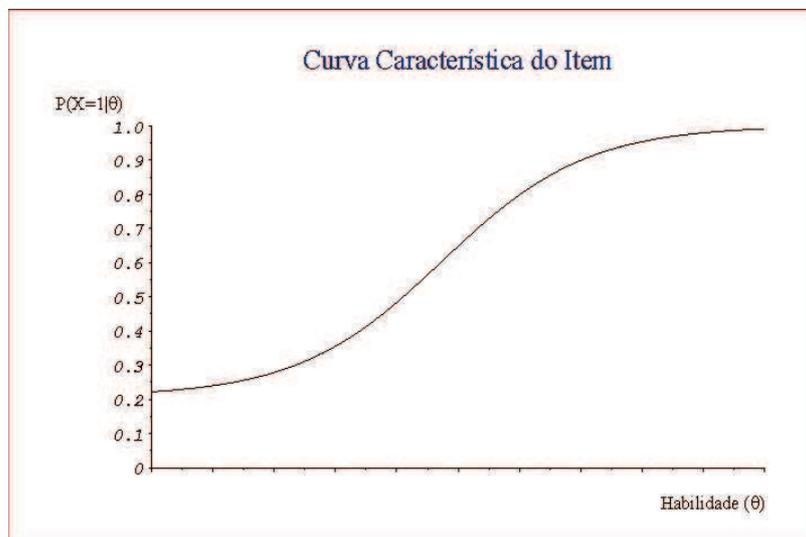


Figura 1 Curva Característica do Item - CCI

Na Figura 1, a CCI parte do valor 0,20 a probabilidade de acerto casual em um item com 5 alternativas, com o uso do M3P. Observa-se que, quanto maior for a habilidade do indivíduo, maior será a probabilidade de que ele responda corretamente ao item, obedecendo uma relação não linear (BRAGION, 2007).

Conforme pode ser visto na Figura 2, uma CCI pode ser descrita da seguinte forma:

- a) θ possui uma escala arbitrária onde o importante são as relações de ordem e não sua magnitude;

- b) a é proporcional à derivada da tangente da curva no ponto de inflexão;
- c) b (medido na mesma escala de θ) representa a habilidade necessária para uma probabilidade de acerto igual a $(1 + c)/2$;
- d) c representa a probabilidade de um indivíduo, com baixa habilidade, responder corretamente o item.

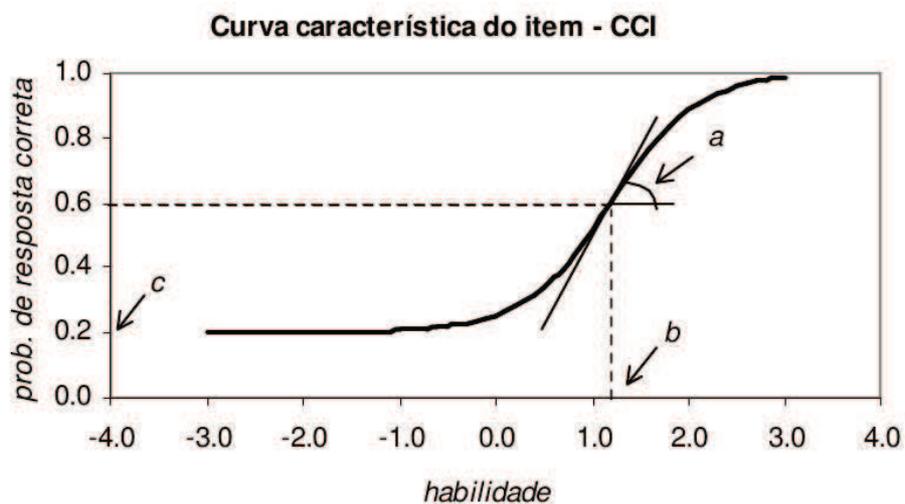


Figura 2 Relação dos parâmetros na Curva Característica do Item - CCI.

A seguir, nas Figuras 3, 4 e 5, foram feitas comparações entre CCI's variando os valores dos parâmetros de itens:

- a) Comparações entre as CCI's com a mesma discriminação e diferentes dificuldades:
- b) Comparações entre as CCI's com a mesma dificuldade e diferentes discriminações:
- c) Comparações entre as CCI's com diferentes valores para o parâmetro de acerto casual:

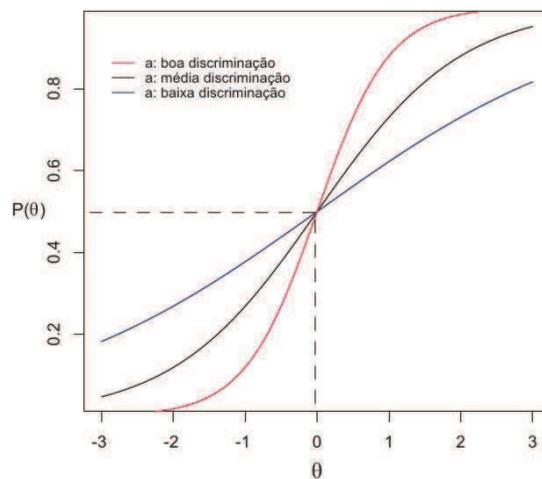


Figura 3 CCI's com diferentes valores para b e $a = 1$.

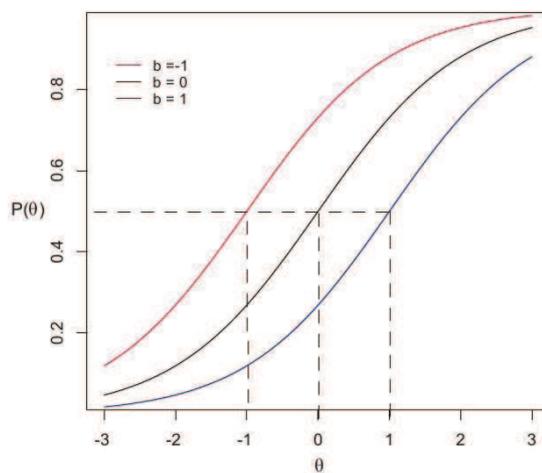


Figura 4 CCI's com diferentes valores para a e $b = 0$.

2.2 Modelos Multidimensionais da Teoria de Resposta ao Item

Alguns pontos acerca da TRI são levantados, como por exemplo, a qualidade de estimação da distribuição das habilidades dos elementos de uma popu-

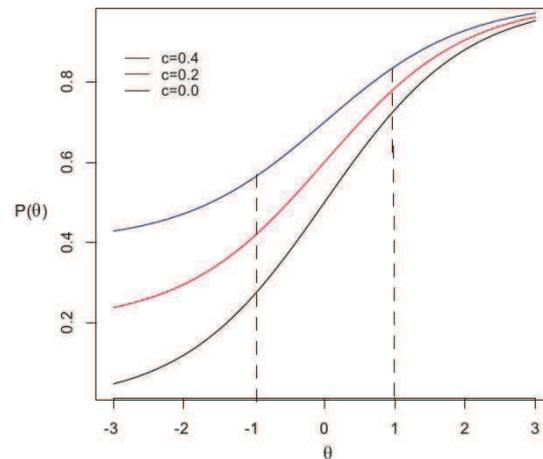


Figura 5 CCI's com diferentes valores para c .

lação (MISLEVY, 1991). Esse autor discute a possibilidade de se obter melhores estimativas da variabilidade usando informações dos indivíduos respondentes que possam estar associadas às suas habilidades, como sua condição socioeconômica e o grau de escolaridade, dentre outras.

Alguns modelos, como os Multidimensionais, (HARTIG; HÖHLER, 2009; NOJOSA, 2001; RECKASE, 2009), assumem que é necessário mais de um traço latente para modelar apropriadamente as probabilidades de resposta. Outros modelos como o Multivariado (MATOS, 2001) e o Longitudinal (TAVARES, 2001), modelam, respectivamente, a estrutura de dependência dos vários traços latentes medidos em diferentes testes e as dependências longitudinais dos traços latentes.

Pode ser interessante também estudar indivíduos de diferentes grupos, com diferentes características próprias de cada um desses grupos. Essas diferentes características podem induzir comportamentos distintos na distribuição dos traços latentes. As distribuições de probabilidades dos traços latentes poderiam ser específicas para cada grupo, fazendo-se necessário, então, considerar a heterogeneidade dos grupos envolvidos. Propostas para modelar as diferenças entre grupos foram apresentadas por Mislevy (1983) e a estimação das densidades dos traços latentes de diversos grupos por Mislevy (1984). Estudos recentes abordaram o

tema, como Azevedo (2008) que considera uma estrutura bayesiana, por meio de um esquema de dados aumentados, para analisar modelos longitudinais com grupos múltiplos, e também por Azevedo, Andrade e Fox (2012) que desenvolveram um algoritmo que usa Gibbs sampling (GS) e Metrópolis-Hastings (M-H) com o objetivo de explorar o potencial dos procedimentos de estimação via MCMC, no ajuste de um modelo bayesiano para o modelo de grupos múltiplos.

Os modelos multivariados de resposta ao item consideram a possível estrutura de covariância existente entre as habilidades de um indivíduo associadas a várias disciplinas. Alguns autores já usaram esse tipo de modelo, como Thomas e Gan (1997) que descrevem um modelo utilizado pelo *National Assessment of Educational Progress (NAEP)*. Já o modelo de Tavares (2001) faz uma generalização do modelo de várias populações independentes de Bock e Zimowski (1997), usando avaliações de uma mesma população ao longo do tempo.

Mais ainda, o modelo multivariado, bem como o modelo de múltiplos grupos, podem ser vistos como casos particulares do modelo longitudinal: o multivariado quando se assume as médias em cada tempo iguais a zero e uma matriz de covariâncias não estruturada (com elementos da diagonal iguais a 1 (MATOS, 2001); o de múltiplos grupos quando se considera uma matriz de covariâncias diagonal (TAVARES, 2001).

No Brasil, algumas avaliações em larga escala, como os exames seletivos para o ensino superior, não levam em conta a possível estrutura de covariância/correlação existente entre os rendimentos nas disciplinas avaliadas. Mesmo submetendo os candidatos a provas com várias disciplinas é feito uso de modelos univariados de resposta ao item. Alguns exames seletivos para o ensino superior, como o ENEM, submetem os candidatos a provas de várias disciplinas, mas realizam as avaliações por meio de modelos univariados, não levando em conta a possível estrutura de covariância/correlação entre as disciplinas avaliadas.

Existe uma diferença entre modelos multivariados e modelos multidimensionais. Os dois são modelos que se referem a dois ou mais traços latentes. Entretanto, o modelo multivariado refere-se a traços latentes distintos, avaliados por diferentes testes, sendo que cada item é associada uma única habilidade. Já o modelo multidimensional considera vários traços latentes associados a um item.

Tanto no modelo univariado quanto no multivariado, há um elevado número de parâmetros a serem estimados, diferentes distribuições de traços latentes e diferentes estruturas de dependências, o que torna complicado o ajuste desses modelos. O método de estimação por Máxima Verossimilhança Marginal (MVM) (BOCK; AITKIN, 1981), conforme Azevedo (2003), tem algumas limitações em relação aos procedimentos numéricos necessários para sua implementação. Mislevy (1986) propôs uma versão bayesiana para o MVM, mas ainda restrito ao mesmo problema de implementação.

A análise bayesiana plena é necessária sendo implementada por meio de algoritmos do tipo Monte Carlo via Cadeias de Markov (MCMC), como Gibbs sampling, (BÉGUIN; GLAS, 2001; SHENG; HEADRICK, 2013), Gibbs sampling via dados aumentados, Metropolis-Hastings, Random Walk no algoritmo Metropolis-Hastings como pode ser visto em Patz e Junker (1999a, 1999b).

Os autores de la Torre e Patz (2005) propuseram um método eficiente para estimar as habilidades levando em conta diferentes estruturas de correlação entre essas habilidades, num modelo Multidimensional. Matos (2001) propôs uma metodologia para modelos multivariados de resposta ao item, que teve como origem os modelos longitudinais de Tavares (2001). Este último havia proposto uma generalização do modelo de Bock e Zimowski (1997) ao considerar a situação em que os dados possuem uma estrutura longitudinal, que permite analisar o desempenho educacional de indivíduos em certas avaliações ao longo do tempo. Sendo assim, Matos construiu um modelo em que uma única população é avaliada em relação a l habilidades ou traços latentes, levando-se em consideração uma possível estrutura de correlação entre as habilidades nas l disciplinas.

2.2.1 O Modelo Multivariado da TRI para uma única população

Nesta seção iremos explicitar a construção de um modelo multivariado para uma única população. Deve-se considerar l áreas distintas do conhecimento (l testes em uma prova) obtendo, dessa forma, um vetor com l traços latentes (habilidades), $\theta_i = (\theta_{i1}, \dots, \theta_{il})'$. Neste modelo não há itens comuns entre os testes. Dessa maneira, cada teste terá m_l itens, totalizando $m = m_1 + \dots + m_l$

itens.

Matos (2001) usou o modelo logístico de 3 parâmetros (ML3P) da TRI por ser um modelo de grande aceitação entre os profissionais das áreas de avaliação, como segue:

$$P(Y_{ijl} = 1 | \theta_{il}, a_{jl}, b_{jl}, c_{jl}) = c_{jl} + (1 - c_{jl}) \frac{1}{1 + \exp[-a_{jl}(\theta_{il} - b_{jl})]} \quad (4)$$

em que θ_{il} é a habilidade do indivíduo i no teste l e a_{jl}, b_{jl}, c_{jl} são os parâmetros associados ao item j no teste l .

Algumas suposições devem ser atendidas nesse modelo:

- a) A probabilidade de acerto a qualquer item, em qualquer teste, pode ser modelada pela mesma função resposta (ML3P);
- b) Suposições de independência:
 - independência local ou condicional, em que uma vez conhecida a habilidade θ_{il} do indivíduo, as respostas aos itens que compõem cada teste são independentes;
 - as respostas aos itens de um teste são fornecidas de forma independente dos outros testes, uma vez conhecidas e fixadas as l habilidades do indivíduo, $\theta_i = (\theta_{i1}, \dots, \theta_{il})'$;
 - respostas ao teste l dependem apenas da habilidade θ_{il} , ou seja, independem da habilidade θ_{ik} , uma vez conhecida θ_{il} , $l \neq k$;
 - indivíduos diferentes respondem aos testes de forma independente.
- c) O tempo destinado ao teste é suficiente para que todos possam responder a todos os itens;
- d) A ordem em que os itens são apresentados aos indivíduos não interfere no desempenho dos mesmos;
- e) A probabilidade de acerto a cada item é função de um único traço latente.

No caso do presente estudo optou-se pelo Modelo de Ogiva Normal Multivariado de 3 Parâmetros, ou seja, um modelo probito, usando dados aumentados, porque esse modelo facilita a implementação do algoritmo utilizado.

3 MATERIAL E MÉTODOS

3.1 Implementação do Modelo Probit de 3 Parâmetros Multivariado (MP3PM) usando dados aumentados

Seja um exame vestibular, composto por p provas (disciplinas) distintas em que Y_{ijl} é a resposta (0 para resposta incorreta e 1 para resposta correta) do i -ésimo indivíduo ao j -ésimo item da l -ésima prova. Será considerado um único grupo de indivíduos respondendo a todos os itens de uma ou mais provas (disciplinas) de um exame/avaliação.

Neste trabalho será usado o Modelo Probit de 3 Parâmetros Multivariado da TRI (MP3PM), no qual a probabilidade de um indivíduo acertar um item, dada sua habilidade e parâmetros do item, é

$$\pi_{ijl} = P(Y_{ijl} = 1 | \theta_{il}, a_{jl}, b_{jl}, c_j) = c_j + (1 - c_j)\Phi(\eta_{ijl}) \quad (5)$$

em que

Φ é a função de distribuição acumulada da normal padrão;

$$\eta_{ijl} = a_{jl}(\theta_{il} - b_{jl});$$

θ_{il} é a habilidade do indivíduo i na prova l ;

a_{jl} é o parâmetro de discriminação do item j na prova l ;

b_{jl} é o parâmetro de dificuldade do item j na prova l ;

c_j é o parâmetro associado à probabilidade de que um indivíduo com habilidade baixa responda corretamente ao item j ;

$i = 1, \dots, n$, $j = 1, \dots, m$ e $l = 1, \dots, p$;

$Y_{ijl} \sim \text{Bernoulli}(\pi_{ijl})$.

Sendo assim, a verossimilhança para Y será:

$$L(y_{ijl}; \theta_{il}, a_{jl}, b_{jl}, c_j) = \prod_{l=1}^p \prod_{i=1}^n \prod_{j=1}^m \pi_{ijl}^{y_{ijl}} (1 - \pi_{ijl})^{1-y_{ijl}} \quad (6)$$

Os dados Y , que são as respostas de n indivíduos para m itens nas l provas,

serão aumentados com variáveis latentes W e Z definidas, conforme Béguin e Glas (2001), com o objetivo de facilitar a implementação do algoritmo MCMC.

A variável W_{ijl} é construída da seguinte forma:

$$W_{ijl} = \begin{cases} 1 & \text{se o indivíduo } i \text{ sabe a resposta correta do item } j \text{ da prova } l \\ 0 & \text{se o indivíduo não sabe a resposta.} \end{cases}$$

Daí, a probabilidade condicional de $W_{ijl} = w_{ijl}$ dado $Y_{ijl} = y_{ijl}$ será:

$$\begin{aligned} P(W_{ijl} = 1 | Y_{ijl} = 1, \eta_{ijl}, c_j) &\propto \Phi(\eta_{ijl}) & (7) \\ P(W_{ijl} = 0 | Y_{ijl} = 1, \eta_{ijl}, c_j) &\propto c_j(1 - \Phi(\eta_{ijl})) \\ P(W_{ijl} = 1 | Y_{ijl} = 0, \eta_{ijl}, c_j) &= 0 \\ P(W_{ijl} = 0 | Y_{ijl} = 0, \eta_{ijl}, c_j) &= 1 \end{aligned}$$

As variáveis Z_{ijl} são independentes e normalmente distribuídas com média η_{ijl} e desvio padrão 1.

Essas variáveis se relacionam com w por: $z_{ijl} > 0$ se $w_{ijl} = 1$ ou $z_{ijl} \leq 0$ se $w_{ijl} = 0$. Então,

$$p(z_{ijl} | w_{ijl}, \eta_{ijl}) \propto \phi(z_{ijl}; \eta_{ijl}, 1) \cdot [I(z_{ijl} \geq 0)I(w_{ijl} = 1) + I(z_{ijl} < 0)I(w_{ijl} = 0)] \quad (8)$$

em que $\phi(\cdot; \eta_{ijl}, 1)$ é a densidade normal com média η_{ijl} e desvio padrão 1 e, $I(\cdot)$ é uma função indicadora que assume 1 se o argumento é verdadeiro e 0 caso contrário.

Dessa forma, tem-se agora uma verossimilhança para os dados aumenta-

dos:

$$\begin{aligned}
p(z_{ijl}, w_{ijl} | y_{ijl}; \theta_{il}, \boldsymbol{\Sigma}, \boldsymbol{\zeta}, c_{jl}) &= \\
&= p(z_{ijl} | w_{ijl}, \eta_{ijl}) p(w_{ijl} | y_{ijl}, \eta_{ijl}, c_{jl}) \propto \\
&\propto \prod_{l=1}^p \prod_{i=1}^n \prod_{j=1}^m [\phi(z_{ijl}; \eta_{ijl}, 1) I(z_{ijl} > 0, w_{ijl} = 1, y_{ijl} = 1)] \Phi(\eta_{ijl}) + \\
&+ [\phi(z_{ijl}; \eta_{ijl}, 1) I(z_{ijl} \leq 0, w_{ijl} = 0, y_{ijl} = 1)] c_{jl} (1 - \Phi(\eta_{ijl})) + \\
&+ [\phi(z_{ijl}; \eta_{ijl}, 1) I(z_{ijl} > 0, w_{ijl} = 1, y_{ijl} = 0)] \cdot 0 + \\
&+ [\phi(z_{ijl}; \eta_{ijl}, 1) I(z_{ijl} \leq 0, w_{ijl} = 0, y_{ijl} = 0)] \cdot 1 \tag{9}
\end{aligned}$$

Essa estrutura de dados aumentados permite obter distribuições condicionais completas com forma funcional mais simples (ALBERT, 1992).

As distribuições *a priori* para os parâmetros são:

- a) a_j e b_j são representados pelo vetor $\boldsymbol{\zeta}$ e terá como *priori* uma distribuição normal bivariada com média e variância definidas da seguinte forma:

$$\boldsymbol{\mu}_{\boldsymbol{\zeta}_0} = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_0} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \tag{10}$$

em que $\mu_a = 1$, $\mu_b = 0$, $\sigma_a^2 = 0,5$, $\sigma_b^2 = 9$ e $p(\boldsymbol{\zeta}) = \prod_{j=1}^m I(a_j > 0)$, assegurando que o parâmetro de discriminação seja sempre positivo.

- b) $\boldsymbol{\theta}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$, em que $\mathbf{0}$ é um vetor de zeros e $\boldsymbol{\Sigma}_\theta$ é uma matriz de correlações de ordem p .
- c) c_j terá *priori* conjugada $Beta(5, 17)$, pelo fato de a média dessa distribuição, com esses parâmetros, aproximar de 0,25, valor que é esperado para questões com 4 alternativas de resposta.
- d) $\boldsymbol{\Sigma}_\theta$ terá *priori* proporcional a uma constante, ou seja, $p(\boldsymbol{\Sigma}_\theta) \propto 1$.

O motivo de adotar esse procedimento é o de simular amostras de uma

distribuição *a posteriori* conjunta de ξ , \mathbf{c} , $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, \mathbf{z} e \mathbf{w} :

$$\begin{aligned}
 p(\boldsymbol{\zeta}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{w}, \boldsymbol{\Sigma} | \mathbf{y}) &= p(\mathbf{z}, \mathbf{w} | \mathbf{y}; \boldsymbol{\zeta}, \mathbf{c}, \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta} | \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \cdot p(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \cdot \\
 &\cdot p(\boldsymbol{\zeta}) \cdot p(\mathbf{c}) = \\
 &= \prod_{l=1}^p \prod_{i=1}^n \prod_{j=1}^m p(z_{ij} | w_{ij}; \eta_{ij}) \cdot p(w_{ij} | y_{ij}; \eta_{ij}, c_j) \cdot \\
 &\cdot p(\boldsymbol{\theta}_i | \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \cdot p(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \cdot p(\boldsymbol{\zeta}) \cdot p(\mathbf{c})
 \end{aligned} \tag{11}$$

O procedimento no modelo multivariado envolve construir as distribuições condicionais completas *a posteriori* de todas as componentes de acordo com os 5 passos a seguir:

1º) Obter $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} | \boldsymbol{\theta}$ de uma Distribuição Inversa Wishart:

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} | \boldsymbol{\theta} \sim IW(\nu, S^{-1})$$

em que $\nu = n$ são os graus de liberdade e $S = \sum_{i=1}^n (\theta_{il} - \bar{\theta}_l)(\theta_{il} - \bar{\theta}_l)'$ com $\bar{\theta}_l$ sendo a média das habilidades na prova l .

2º) Obter $\mathbf{w}, \mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\zeta}, \mathbf{c}, \mathbf{y}$ da seguinte forma:

a) Obter w_{ijl} da distribuição de $W_{ijl} | \mathbf{y}; \boldsymbol{\zeta}, \mathbf{c}, \boldsymbol{\theta}$ como na Eq. 7:

$W_{ijl} \sim \text{Bernoulli}(\pi_w)$ em que

$$\pi_w = \frac{\phi(z|\eta)\Phi(\eta)}{\phi(z|\eta)\Phi(\eta) + \phi(z|\eta)c_j(1 - \Phi(\eta))} = \frac{\Phi(\eta)}{\Phi(\eta) + c_j(1 - \Phi(\eta))}$$

b) Obter z_{ijl} em $Z_{ijl} | \mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \mathbf{y}$, sendo:

$$Z_{ijl} | \mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \mathbf{y} \sim \begin{cases} N(\eta_{ijl}, 1) : & \text{truncada à esquerda do 0, se } w_{ijl} = 1 \\ N(\eta_{ijl}, 1) : & \text{truncada à direita do 0, se } w_{ijl} = 0. \end{cases}$$

3º) Obter $\boldsymbol{\theta} | \mathbf{z}, \boldsymbol{\zeta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ de uma distribuição normal da seguinte forma:

Das Equações 7 e 8 segue que

$$Z_{ijl} + b_{jl} = a_{jl}\theta_{il} + \varepsilon_{ijl}$$

em que ε_{ijl} é um erro normalmente distribuído.

A distribuição condicional de θ implica um modelo normal para a regressão de $Z_{ijl} + b_{jl}$ em a_{jl} , sendo θ_i os coeficientes da regressão, no qual tem uma distribuição *a priori* normal com $\mu = \mathbf{0}$ e Σ_θ . Assim, a distribuição condicional completa de θ_i é normal:

$$\theta_i \sim N_p(\mu^*, \Sigma^*) \quad (12)$$

em que $\hat{\theta}_i = (\mathbf{a}'\mathbf{a})^{-1}\mathbf{a}'(\mathbf{z} + \mathbf{b})$ e \mathbf{a} é uma matriz de dimensão $m \times p$. Sendo assim,

$$\begin{aligned} \Sigma^* &= ((\mathbf{a}'\mathbf{a})^{-1} + \Sigma_\theta^{-1})^{-1} = \mathbf{a}'\mathbf{a} \\ &\text{e} \\ \mu^* &= ((\mathbf{a}'\mathbf{a})^{-1} + \Sigma_\theta^{-1})^{-1}((\mathbf{a}'\mathbf{a})^{-1}\hat{\theta}_i + \Sigma_\theta^{-1}\mu) = \\ &= (\mathbf{a}'\mathbf{a})(\mathbf{a}'\mathbf{a})^{-1}\hat{\theta}_i = \\ &= (\mathbf{a}'\mathbf{a})^{-1}\mathbf{a}'(\mathbf{z} + \mathbf{b}) \end{aligned}$$

4º) Obter $\zeta | \mathbf{z}, \theta$

Combinam-se a distribuição *a priori* normal bivariada para ζ , com média $\mu_{\zeta_0} = (\mu_a, \mu_b)'$ e variância $\Sigma_{\zeta_0} = \text{diag}(\sigma_a^2, \sigma_b^2)$.

Condicionais à \mathbf{z} e θ , as *posteriors* de ζ_1, \dots, ζ_m serão independentes com densidade de ζ_j dada por:

$$\begin{aligned} p(\zeta_j | \mathbf{w}, \theta, \mathbf{z}, \mathbf{y}) &\propto \prod_{i=1}^n \prod_{j=1}^m \phi(z_{ijl}; \eta_{ijl}, 1) \phi(a_{il}; \mu_a, \sigma_a) \cdot \\ &\cdot \phi(b_{jl}; \mu_b, \sigma_b) \end{aligned}$$

em que $\phi(\cdot)$ é uma função densidade de probabilidade da distribuição normal.

Sabendo que $z_{ij} = a_{jl}\theta_{il} - 1b_{jl}$, define-se \mathbf{X} uma matriz $n \times (p + 1)$ com linhas $(\theta_{i1}, \dots, \theta_{ip}, -1)$. Condicional a θ , $\mathbf{z}_j = (z_{1j}, \dots, z_{nj})'$

satisfaz um modelo linear

$$z_j = \mathbf{X}\zeta_j + \varepsilon_j$$

em que $\varepsilon_j = (\varepsilon_{1j}, \dots, \varepsilon_{ij}, \dots, \varepsilon_{nj})'$ iid $N(0,1)$.

A verossimilhança de ζ é uma Normal com média $\hat{\zeta}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'z_j$ e variância $v = (\mathbf{X}'\mathbf{X})^{-1}$.

Combinando a verossimilhança com as *prioris* normais:

$$\zeta_j | \theta, z, \mathbf{y} \sim N\left(\mu_{\zeta_j}, (\Sigma_{\zeta_0}^{-1} + \mathbf{X}'\mathbf{X})^{-1}\right)$$

em que $\mu_{\zeta_j} = (\Sigma_{\zeta_0}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\Sigma_{\zeta_0}^{-1}\mu_{\zeta_0} + \mathbf{X}'z_j)$.

5º) Obter $c | \mathbf{w}, \mathbf{y}$

Tome $t_j = \sum_{i=1}^n \mathbf{I}(w_{ij} = 0)$ sendo o número de indivíduos que não sabem a resposta e que dá a resposta correta. A probabilidade de que esse indivíduo acerte a resposta, dado que ele não sabe a resposta correta é $P(Y_{ij} = 1 | W_{ij} = 0) = c_j$.

O número de respostas corretas obtidas por mero acaso, $s_j = \sum_{i|w_{ij}=0}^n y_{ij}$, tem distribuição Binomial, $B(t_j, c_j)$.

Com uma *priori* Beta conjugada não informativa, a distribuição condicional completa a *posteriori* de c_j é:

$$c_j | \mathbf{w}, \mathbf{y} \sim \text{Beta}(s_j + 1, t_j - s_j + 1) \quad (13)$$

3.2 Diferentes configurações na simulação do modelo multivariado

Nessa seção será avaliada a magnitude das correlações entre os valores paramétricos simulados e estimados e o Erro Quadrático Médio (EQM) para os valores estimados via MCMC (método de Monte Carlo via Cadeias de Markov). As análises serão feitas com base em um único grupo de indivíduos respondendo a um exame com 3 e com 5 provas distintas, cada uma com diferentes números

de itens, sendo 10, 30 e 50 itens e com 3 diferentes estruturas de covariâncias não estruturadas (com elementos da diagonal iguais a 1) para os exames com 3 e com 5 provas. Ao todo foram 36 diferentes configurações, sendo 18 com um grupo de 1000 indivíduos e 18 com um grupo de 2000 indivíduos, sendo 9 configurações para a prova com 3 testes e 9 para a prova com 5 testes. As configurações simuladas estão descritas na Tabela 1.

Tabela 1 Configurações dos exames simulados usando diferentes números de indivíduos, provas por exame, itens por prova e estruturas de covariâncias.

Configurações para os exames simulados					
Indivíduos	Provas por exame	Itens por prova	Estruturas de Covariâncias		
1000	3	10	$I_{(3)}$	$R1_{(3)}$	$R2_{(3)}$
		30			
		50			
	5	10	$I_{(5)}$	$R1_{(5)}$	$R2_{(5)}$
		30			
		50			
2000	3	10	$I_{(3)}$	$R1_{(3)}$	$R2_{(3)}$
		30			
		50			
	5	10	$I_{(5)}$	$R1_{(5)}$	$R2_{(5)}$
		30			
		50			

Nessa Tabela 1 estão as 36 configurações para a simulação do modelo multivariado.

As matrizes de covariâncias serão assumidas como possíveis estruturas de correlações entre as habilidades nas provas. Para o exame com 3 provas medindo uma habilidade por prova são:

$$R1_{(3)} = \begin{bmatrix} 1 & 0,5 & 0,25 \\ & 1 & 0,5 \\ & & 1 \end{bmatrix} \quad R2_{(3)} = \begin{bmatrix} 1 & 0,9 & 0,8 \\ & 1 & 0,9 \\ & & 1 \end{bmatrix}$$

Essas estruturas em um exame com três provas foram pensadas da seguinte forma:

$I_{(3)}$ Estrutura Identidade evidenciando uma situação em que não há relação entre as habilidades do indivíduo nas provas;

$R1_{(3)}$ Estrutura com três provas (Matemática, Português e História, por exemplo), evidenciando que a habilidade de um indivíduo em Matemática razoavelmente relacionada com sua habilidade em Português e com sua habilidade em História, e que a habilidade em Português se relaciona razoavelmente com a habilidade em História.

$R2_{(3)}$ Estrutura com três provas (Matemática, Português e Física, por exemplo), evidenciando que a habilidade de um indivíduo em Matemática está supostamente relacionada tanto com sua habilidade em Português como com sua habilidade em Física, e que a habilidade em Português se relaciona fortemente com a habilidade em Física.

Para a prova com 5 testes/habilidades são:

$$R1_{(5)} = \begin{bmatrix} 1 & 0,5 & 0,0 & 0,0 & 0,0 \\ & 1 & 0,5 & 0,0 & 0,0 \\ & & 1 & 0,5 & 0,0 \\ & & & 1 & 0,5 \\ & & & & 1 \end{bmatrix} \quad R2_{(5)} = \begin{bmatrix} 1 & 0,3 & 0,3 & 0,3 & 0,3 \\ & 1 & 0,3 & 0,3 & 0,3 \\ & & 1 & 0,3 & 0,3 \\ & & & 1 & 0,3 \\ & & & & 1 \end{bmatrix}$$

Já essas estruturas em um exame com 5 provas foram pensadas da seguinte forma:

$I_{(5)}$ Estrutura Identidade evidenciando uma situação em que não há relação entre as habilidades do indivíduo nas provas;

$R1_{(5)}$ Estrutura com 5 provas (Matemática, Português, Geografia, História e Filosofia, por exemplo) evidenciando que a habilidade de um indivíduo em Matemática está razoavelmente relacionada com sua habilidade em Português e não se relaciona com as demais; a habilidade de um indivíduo em

Português está razoavelmente relacionada com sua habilidade em Geografia e não se relaciona com as demais; a habilidade de um indivíduo em Geografia se relaciona razoavelmente com sua habilidade em História e não se relaciona com as demais; a habilidade de um indivíduo em História está relacionada razoavelmente com sua habilidade em Filosofia.

$R2_{(5)}$ Estrutura com 5 provas (Matemática, Português, Física, Química e Biologia, por exemplo) evidenciando que a habilidade de um indivíduo em Matemática está relacionada com sua habilidade em Português e um pouco das demais provas/disciplinas; a habilidade de um indivíduo em Português está relacionada com sua habilidade em Física e nas demais provas/disciplinas; a habilidade de um indivíduo em Física está relacionada com sua habilidade em Química e nas demais provas/disciplinas; a habilidade de um indivíduo em Química está relacionada com sua habilidade em Biologia.

A Estrutura Identidade de ordens 3 e 5 serviu de referência nas comparações do modelo com ou sem correlações entre as habilidades das l provas.

Os valores paramétricos para parâmetros de itens e habilidades foram os seguintes: $a = [0,5; 2,4]$, $b = [-3; 3]$, $c = [0,15; 0,30]$ e $\theta \sim NM(\mu, \Sigma)$ ($\mu = \mathbf{0}_{(\ell)}$ e $\Sigma = R_{(\ell)}$, com $\ell = \{3,5\}$).

Foi usado o método de Monte Carlo via Cadeias de Markov (MCMC) para gerar cadeias para obter uma amostra da distribuição conjunta a posteriori a partir das distribuições completas de cada parâmetro (Gibbs Sampling). Foram geradas cadeias iniciais de tamanho 4000 valores para cada um dos parâmetros dos itens e das habilidades, considerando um burn-in de 100 observações e um thinning de 5. Foi realizado o diagnóstico de Raftery e Lewis e atualizado burn-in e thinning. Para análise da convergência foram adotados o critério de Gelman e Rubin, além de uma análise visual do traço das cadeias.

Após a obtenção das amostras, foram feitas estimativas pontuais e por intervalo de credibilidade (IC) de máxima densidade a posteriori, HPD (*Highest Probability Density Interval*) para cada um dos parâmetros de interesse. Definiu-se um nível de credibilidade 5% para tal intervalo.

3.3 Dados para aplicação do modelo multivariado: Vestibular 2006-2 da UFLA

Os dados utilizados nesse trabalho referem-se ao Exame Vestibular 2006 – 2 da Universidade Federal de Lavras (UFLA). As provas desse exame fazem parte da primeira fase do vestibular daquele ano. Foram utilizados 58 itens, com quatro (4) alternativas, divididos em nove (9) provas. Por razões operacionais a prova de Língua Estrangeira não foi considerada na análise. Isso permite simplificar a comparação entre os modelos considerados restringindo-a às provas em que todos os analisados submeteram-se a todos os itens, conforme a Tabela 2.

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

Provas	Quantidade de itens	Sequência
Português	10	1 a 10
Geografia	8	11 a 18
História	6	19 a 24
Filosofia	2	25 a 26
Biologia	8	27 a 34
Física	8	35 a 42
Matemática	8	43 a 50
Química	8	51 a 58

Esse vestibular foi aplicado a 10 cursos, conforme Tabela 3, mas no presente estudo considera-se apenas 3 cursos (Medicina Veterinária, Química e Zootecnia), totalizando 936 indivíduos respondentes, representando um único grupo de indivíduos, respeitando a metodologia proposta. Esses três cursos escolhidos têm a característica comum de se exigir dos candidatos um conhecimento prévio homogêneo, além do fato de a soma desses candidatos se aproximar de 1000 indivíduos, quantidade também usada nos estudos de simulação.

Tabela 3 Cursos oferecidos e quantidade de candidatos por curso na Universidade Federal de Lavras (UFLA), no segundo semestre de 2006

Sigla	Curso	Nº de candidatos
AD	Administração	284
AL	Engenharia de Alimentos	229
AG	Agronomia	651
CB	Ciências Biológicas	326
CC	Ciência da Computação	235
EA	Engenharia Agrícola	154
EF	Engenharia Florestal	238
MV	Medicina Veterinária	604
QI	Química	104
ZO	Zootecnia	228

Esse exame foi selecionado em decorrência da realização de uma análise univariada com seus dados por Bragion (2007, 2010), o que permite comparar, no presente estudo, o modelo univariado global ao modelo multivariado e aos modelos univariados por provas.

4 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados referentes ao estudo de simulação do modelo multivariado, como proposto na metodologia desta tese, e também referentes à aplicação desse modelo nos dados do Vestibular 2006 – 2 da UFLA.

4.1 Estudo de simulação do modelo multivariado de resposta ao item

Calculou-se o coeficiente de correlação entre os valores paramétricos e estimados para os parâmetros de item, habilidade e matriz de variâncias e covariâncias entre as habilidades, e também o Erro Quadrático Médio (EQM) dos valores estimados via MCMC, com o objetivo de verificar o desempenho do programa usando diferentes configurações na simulação.

A Tab. 4 informa as correlações entre os valores paramétricos e estimados e a Tab. 5 informa os EQM's via MCMC para um exame simulado, aplicado a 1000 indivíduos, com 3 e com 5 provas, com 10, 30 e 50 itens por prova e com 3 estruturas/matrizes de correlações para cada tipo de exame (I , $R1$ e $R2$ de ordem 3 e 5).

Na Tab. 4, para os parâmetros de item, as correlações foram altas, menos a correlação para o parâmetro c . O aumento do número de itens por prova e o aumento do número de provas por exame provocou um pequeno aumento, em geral, das correlações entre os valores paramétricos e estimados para o poder de discriminação (a e \hat{a}) e o grau de dificuldade (b e \hat{b}), com exceção do acerto casual (c e \hat{c}) em que a correlação foi maior quando se tem um exame composto por 3 provas e 30 itens por prova e também no exame composto por 5 provas e 10 itens por prova.

Em relação à estrutura de correlação usada, a Tab. 4 indica que as estruturas $R1$ e $R2$ provocaram um aumento maior do que a estrutura I nas correlações para os parâmetros de item quando se trata de um exame com 3 provas do que um exame com 5.

Para as habilidades, fica evidente na Tab. 4 que as correlações entre os valores paramétricos e estimados (θ e $\hat{\theta}$) são maiores quando usadas as estruturas $R1$

e $R2$ em vez da estrutura I , independente do número de provas por exame. Mais ainda, conforme o número de itens por prova aumenta, as correlações também aumentam para qualquer EC.

Tabela 4 Correlação entre valores paramétricos e valores estimados para todas as configurações simuladas ($n = 1000$)

Configurações			Correlação entre valores paramétricos e estimados												
Provas	Itens	Estrutura	a	b	c	θ_1	θ_2	θ_3	θ_4	θ_5	θ	Σ			
3	10	I	0,63	0,93	0,27	0,79	0,82	0,81	—	—	0,81	0,99			
		R1	0,33	0,91	0,15	0,80	0,85	0,81	—	—	0,82	0,96			
		R2	0,69	0,93	0,15	0,86	0,91	0,90	—	—	0,89	0,75			
	30	I	I	0,63	0,93	0,38	0,93	0,93	0,92	—	—	0,93	0,99		
			R1	0,69	0,93	0,27	0,93	0,93	0,92	—	—	0,93	0,99		
			R2	0,58	0,94	0,49	0,94	0,96	0,94	—	—	0,95	1,00		
		50	I	I	0,64	0,88	0,17	0,96	0,94	0,95	—	—	0,95	0,99	
				R1	0,67	0,92	0,26	0,96	0,95	0,95	—	—	0,95	0,96	
				R2	0,70	0,91	0,28	0,96	0,97	0,96	—	—	0,96	0,99	
		50	10	I	I	0,48	0,94	0,42	0,80	0,85	0,82	0,74	0,84	0,81	0,99
					R1	0,48	0,93	0,30	0,79	0,86	0,85	0,83	0,86	0,84	0,99
					R2	0,40	0,92	0,27	0,81	0,84	0,83	0,77	0,86	0,82	0,98
30	I		I	0,71	0,91	0,15	0,94	0,93	0,87	0,92	0,95	0,92	0,99		
			R1	0,69	0,92	0,20	0,93	0,93	0,91	0,93	0,95	0,93	0,99		
			R2	0,67	0,93	0,30	0,93	0,93	0,90	0,92	0,94	0,92	0,99		
50	I	I	0,72	0,92	0,27	0,96	0,94	0,95	0,95	0,96	0,95	0,99			
		R1	0,65	0,91	0,25	0,96	0,95	0,95	0,96	0,96	0,96	0,99			
		R2	0,65	0,92	0,23	0,96	0,95	0,95	0,96	0,95	0,95	0,99			

Dessa forma, a Tab. 4 mostra que o uso das estruturas I , $R1$ e $R2$ não aumenta, de forma geral, as correlações para o parâmetros de item. O aumento, mesmo que pequeno, se dá pelo efeito do aumento do número de provas e de itens por prova. Já para as habilidades, percebe-se que o aumento das correlações se dá pelo efeito do aumento do número de itens por prova e pelo efeito do uso das diferentes estruturas de correlação, com a vantagem de usar $R1$ e $R2$ em vez de I , tanto para exames com 3 quanto com 5 provas.

Na Tab. 5, em relação aos parâmetros de item, os EQM's foram menores para c , a e b , nessa ordem, sendo quase nulo para c . Em geral, à medida que o número de itens por prova aumenta o EQM para a diminui em um exame com 3 provas; já para o exame com 5 provas, o EQM para a diminui quando o número de itens por prova aumenta de 10 para 30 e aumenta quando passa de 30 para 50 itens. Agora, para o parâmetro b , o EQM diminui quando aumenta de 10 para 30 o número de itens por prova e aumenta quando passa de 30 para 50.

Em relação às habilidades, a Tab. 5 mostra EQM's altos, comparados com os EQM's dos parâmetros de itens. Na maioria das vezes, quando o número de itens por prova aumenta, o EQM das habilidades com estrutura *I* aumentam para qualquer quantidade de prova por exame; quando se tem *R1* e *R2* os EQM's ora aumentam ora diminuem, sem um padrão determinado, independente da quantidade de provas e itens por provas.

Assim, as estruturas de correlações da forma *R1* e *R2* não têm efeito sobre os valores dos EQM's das habilidades. Em relação aos parâmetros de itens, a estrutura *R1* foi a que fez com que o EQM diminuísse à medida que o número de itens por prova aumentasse.

Tabela 5 Erro Quadrático Médio (EQM) para os parâmetros estimados em todas as configurações utilizadas ($n = 1000$)

Configurações			Erro Quadrático Médio (EQM)								
Provas	Itens	Estrutura	a	b	c	θ_1	θ_2	θ_3	θ_4	θ	Σ
3	10	I	0,39	0,88	0,04	1,97	1,93	1,96	—	—	<0,004
		R1	0,53	0,81	0,04	2,00	1,90	1,94	—	—	<0,004
		R2	0,30	0,67	0,03	2,13	2,12	2,09	—	—	<0,002
	30	I	0,31	0,58	0,03	1,81	1,94	1,97	—	—	<0,002
		R1	0,28	0,66	0,03	2,04	1,97	1,99	—	—	<0,001
		R2	0,35	0,60	0,03	1,94	1,96	1,95	—	—	<0,001
		I	0,28	0,86	0,03	2,04	1,98	2,00	—	—	<0,002
		R1	0,27	0,67	0,03	1,95	1,98	2,03	—	—	<0,002
		R2	0,25	0,78	0,03	2,00	1,99	2,02	—	—	<0,001
	50	I	0,56	0,69	0,04	2,02	2,09	2,14	1,96	1,98	<0,004
		R1	0,44	0,69	0,04	1,90	1,93	2,09	2,00	1,92	<0,004
		R2	0,52	0,89	0,04	1,88	2,01	1,96	1,96	1,90	<0,004
I		0,28	0,69	0,03	1,99	1,95	1,93	2,03	2,03	<0,003	
R1		0,28	0,68	0,03	2,09	2,07	1,99	1,99	1,97	<0,002	
R2		0,28	0,62	0,03	2,05	1,90	1,96	2,03	2,03	<0,002	
5	30	I	0,49	2,62	0,03	1,89	2,01	2,04	1,89	1,82	<0,003
		R1	0,41	2,95	0,03	1,96	2,04	1,97	2,05	2,10	<0,002
		R2	0,50	2,50	0,03	2,00	1,96	2,00	2,08	2,07	<0,002
	50	I	0,49	2,62	0,03	1,89	2,01	2,04	1,89	1,82	<0,003
		R1	0,41	2,95	0,03	1,96	2,04	1,97	2,05	2,10	<0,002
		R2	0,50	2,50	0,03	2,00	1,96	2,00	2,08	2,07	<0,002

A seguir, serão exibidos os gráficos referentes à interação entre número de provas e estruturas de correlações (Fig. 6, 8, 10 e 12) e os gráficos de interação entre exames com 3 e 5 provas (Fig. 7, 9, 11 e 13) para verificar a magnitude das correlações entre os valores paramétricos e os estimados em função do número de itens por prova, com $n = 1000$ indivíduos.

As Fig. 6 e 7 dizem respeito ao parâmetro Poder de Discriminação (a). Na Fig. 6 o exame com 5 provas e estrutura I ($5 : I$) provocou um maior aumento da correlação entre a e \hat{a} , dentre todas as interações entre provas e estruturas de correlação, à medida que o número de itens por prova aumenta. Para o exame com 3 provas as estruturas $R2$ e $R1$ fizeram com que as correlações entre a e \hat{a} aumentassem à medida em que aumentava a quantidade de itens por prova. Na Fig. 7 as correlações entre a e \hat{a} aumentam quando o número de itens por prova aumenta para exames com 3 e 5 provas. Quando se tem 10 itens por prova as correlações são maiores no exame com 3 do que com 5 provas. Quando se tem 30 itens a situação inverte e quando passa para 50 a situação se equipara.

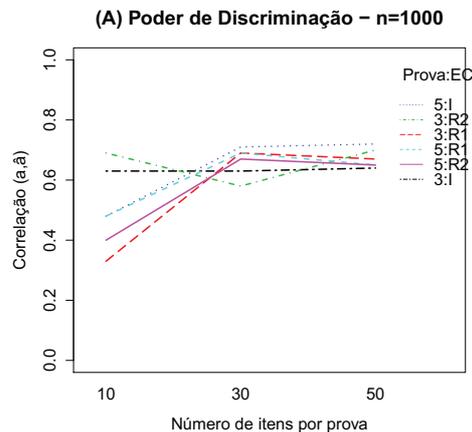


Figura 6 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 1000$ indivíduos

Sendo assim, em se tratando do Poder de Discriminação do item, se o exame tiver 3 provas, as estruturas $R1$ e $R2$ produzem melhor efeito no aumento das correlações entre a e \hat{a} do que quando se usa I com qualquer quantidade de itens por prova. Para o exame com 5 provas, não há razões para usar as estruturas $R1$ e $R2$, já que a estrutura I apresentou um aumento maior das correlações independente do número de itens por prova.

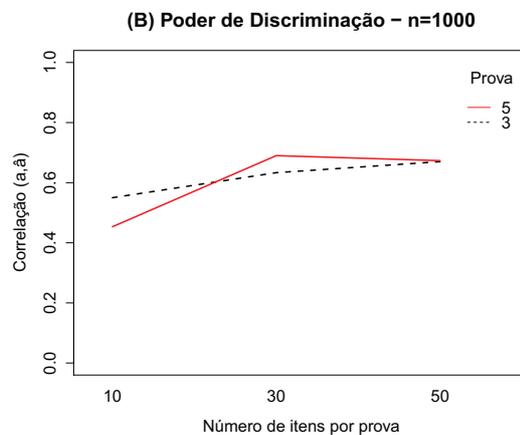


Figura 7 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 1000$ indivíduos

As Fig. 8 e 9 dizem respeito ao parâmetro Grau de Dificuldade (b). A Fig.8 mostra que as correlações entre b e \hat{b} são altas para qualquer quantidade de itens por prova e para qualquer estrutura de correlação associada ao número de provas em um exame. Mesmo assim, quando se tem 50 itens por prova, a estrutura I no exame com 3 provas apresentou as menores correlações e a estrutura $R1$ nesse mesmo exame apresentou as maiores correlações entre b e \hat{b} . Para o exame com 5 provas, a estrutura I se equipara com 3 : $R1$.

A Fig. 9 mostra que os exames com 3 e 5 provas têm desempenho parecido em relação às correlações entre b e \hat{b} . Para qualquer quantidade de itens por prova, as correlações são muito altas (acima de 0.9). O exame com 5 provas leva uma pequena vantagem em relação ao com 3 provas quando a quantidade de itens é igual a 10 e 50.

Logo, para o Grau de Dificuldade, o efeito das interações entre provas e estrutura de correlação na magnitude das correlações entre b e \hat{b} em função do número de itens por prova são parecidos.

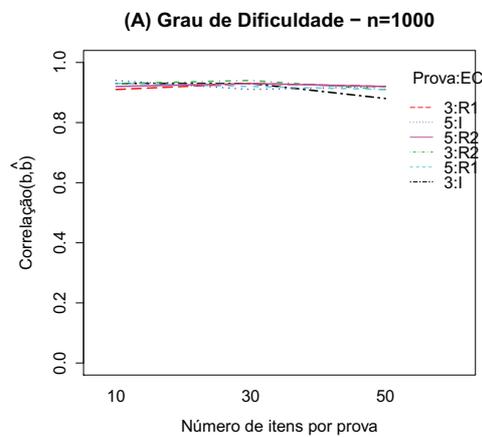


Figura 8 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 1000$ indivíduos

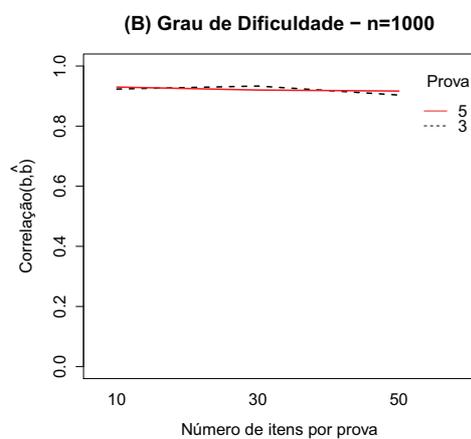


Figura 9 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 1000$ indivíduos

As Fig. 10 e 11 dizem respeito ao parâmetro Acerto Casual (c). A Fig. 10 mostra que as correlações entre c e \hat{c} são relativamente baixas, comparadas como outros parâmetros de itens. Com poucos itens por prova (10), o exame com 5 provas e estrutura I retorna uma correlação acima de 0.4, a mais alta dentre

as interações entre provas e estrutura de correlação. Com 30 itens a interação da estrutura $R1$ e o exame com 5 provas resulta em correlação próxima de 0.5. Já para 50 itens, as correlações caem consideravelmente e, dentre as mais baixas, a interação do exame com 3 provas e estrutura $R2$ foi a que apresentou maior magnitude de correlação.

Na Fig. 11 a interação entre provas por exame mostra que 30 itens por prova retorna correlação máxima quando tal exame tem 3 provas. O exame com 5 provas resultou em maiores correlações apenas quando se tem 10 e 50 itens por prova.

Baseado nisso, o exame com 5 provas tem um melhor efeito no aumento da correlação entre c e \hat{c} para quantidades baixas ou altas de itens por prova. Para quantidade mediana de itens, o efeito de se usar um exame com 3 provas fica evidente. Já para a interação entre número de provas e estrutura de correlação, o caso em que se tem 3 provas com estrutura $R2$ apresentou o melhor efeito no aumento das correlações entre c e \hat{c} à medida que o número de itens por prova também aumenta, do que quando se usa I .

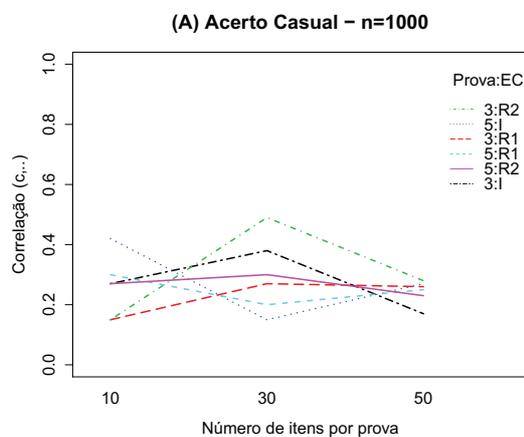


Figura 10 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 1000$ indivíduos

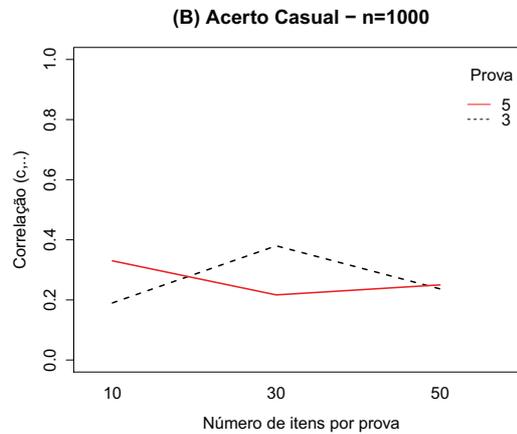


Figura 11 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 1000$ indivíduos

As Fig. 12 e 13 dizem respeito à Habilidade (θ). A Fig. 12 mostra que as correlações entre θ e $\hat{\theta}$ são altas e aumentam ainda mais quando o número de itens por prova também aumenta. Dentre as interações entre quantidades de provas e estruturas de correlação, as que proporcionaram um aumento maior na magnitude das correlações foram a estrutura $R2$ no exame com 3 provas e a estrutura $R1$ no exame com 5 provas.

Na Fig. 13, tanto o exame com 3 quanto o com 5 provas retornaram correlações altas entre θ e $\hat{\theta}$. Mesmo assim, percebe-se que as correlações vão aumentando para ambos os exames à medida que o número de itens por prova também aumenta. Apesar de o desempenho dos exames serem bem parecidos quando cada prova tem 10 ou 30 itens, o exame com 3 provas provocou uma correlação maior que o com 5 provas.

Apesar de as interações nos gráficos anteriores serem muito parecidas, há uma pequena vantagem em se escolher a Estrutura $R2$ para exame com 3 provas e a estrutura $R1$ no exame com 5 provas, pois os efeitos delas provocaram um aumento nas correlações entre θ e $\hat{\theta}$ maior que o efeito de I nos dois tipos de exames.

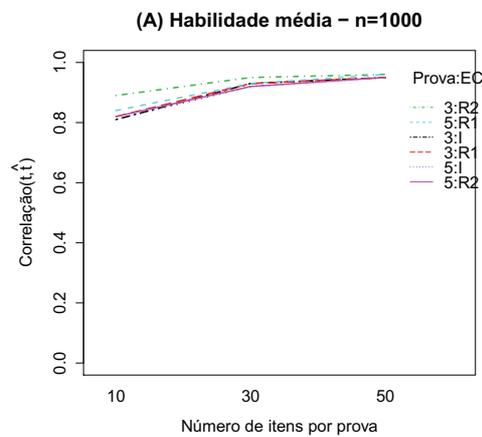


Figura 12 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para Habilidade média em função do número de itens por prova, com $n = 1000$ indivíduos

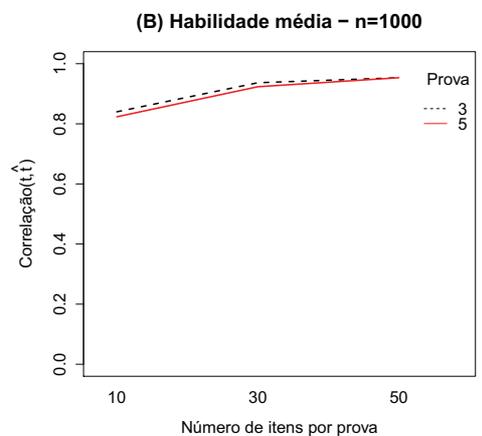


Figura 13 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para a Habilidade média em função do número de itens por prova, com $n = 1000$ indivíduos

Para a simulação com 2000 indivíduos fazendo os exames, a Tab. 6 informa as correlações entre os valores paramétricos e estimados e a Tab. 7 informa

os EQM's via MCMC para exames com 3 e com 5 provas, com 10, 30 e 50 itens por prova e com 3 estruturas/matrices de correlações para cada tipo de exame (I , $R1$ e $R2$ de ordem 3 e 5).

Na Tab. 6, para o parâmetro Poder de Discriminação (a) de itens, se fixar a estrutura I , fica claro que as correlações entre a e \hat{a} aumentam à medida que o número de itens aumenta e à medida que o número de provas por exame também aumenta; o mesmo acontece com a estrutura $R1$ fixada. Já para a estrutura $R2$, quando fixada, pode-se perceber que as correlações aumentam com o aumento do número de itens por prova em um exame com 3 provas, o que não acontece no exame com 5 provas. Para o Grau de Dificuldade (b), se for considerado I , as correlações entre b e \hat{b} aumentam com o aumento de itens por prova e de número de provas por exames. Porém, quando se consideram as estruturas $R1$ e $R2$, as correlações aumentam à medida que o número de itens por prova aumenta no exame com 3 provas e diminuem quando o número de itens por prova aumenta em um exame com 5 provas. Para o Acerto Casual (c), considerando a estrutura I , as correlações entre c e \hat{c} só aumentam, quando o número de itens por prova aumenta, no exame com 3 provas; com a estrutura $R1$, as correlações são altas, e maiores que na estrutura I , quando se tem 30 itens por prova no exame com 3 provas; para a estrutura $R2$, as correlações voltam a aumentar com o aumento do número de itens por prova e para o aumento do número de provas por exame.

Para as habilidades (θ), ainda na Tab. 6, se fixar a estrutura I , as correlações entre θ e $\hat{\theta}$ aumentam à medida que o número de itens por prova aumenta, tanto no exame com 3 provas quanto no exame com 5 provas, mas não com o aumento do número de provas por exame; com o uso da estrutura $R1$ as correlações aumentam com o aumento do número de itens por prova e com o aumento de número de provas por exame. Já com $R2$ fixada, as correlações aumentam com o aumento do número de itens por prova em cada exame, mas entre número de provas por exame, as correlações diminuem de 3 para 5 provas.

Na Tab. 7, no exame com 3 provas, com o uso das estruturas $R1$ e $R2$ os EQM's foram menores do que com o uso da estrutura I para os parâmetros Poder

Tabela 6 Correlação entre valores paramétricos e valores estimados para todas as configurações simuladas ($n = 2000$)

Configurações			Correlação entre valores paramétricos e estimados									
Provas	Itens	R	a	b	c	θ_1	θ_2	θ_3	θ_4	θ_5	θ	Σ
3	10	I	0,44	0,90	0,16	0,79	0,84	0,81	—	—	0,81	0,99
		R1	0,53	0,89	0,25	0,81	0,86	0,83	—	—	0,83	0,99
		R2	0,49	0,91	-0,06	0,88	0,9	0,88	—	—	0,89	0,96
	30	I	0,73	0,93	0,37	0,93	0,91	0,92	—	—	0,92	0,99
		R1	0,67	0,89	0,34	0,93	0,93	0,93	—	—	0,93	0,99
		R2	0,65	0,92	0,17	0,95	0,96	0,94	—	—	0,95	0,99
5	50	I	0,71	0,91	0,33	0,96	0,95	0,96	—	—	0,96	0,99
		R1	0,66	0,93	0,32	0,96	0,95	0,95	—	—	0,95	0,99
		R2	0,74	0,93	0,34	0,96	0,97	0,96	—	—	0,96	0,99
	10	I	0,62	0,92	0,33	0,78	0,84	0,81	0,77	0,84	0,81	0,99
		R1	0,58	0,94	0,24	0,80	0,86	0,84	0,81	0,87	0,84	0,99
		R2	0,46	0,94	0,16	0,79	0,85	0,84	0,77	0,86	0,82	0,99
30	I	0,73	0,92	0,23	0,93	0,93	0,87	0,93	0,95	0,92	0,99	
	R1	0,73	0,92	0,22	0,93	0,93	0,91	0,93	0,94	0,93	0,99	
	R2	0,76	0,92	0,31	0,93	0,93	0,88	0,92	0,95	0,92	0,99	
50	I	0,70	0,93	0,29	0,96	0,95	0,96	0,95	0,95	0,95	0,99	
	R1	0,66	0,92	0,25	0,96	0,96	0,96	0,96	0,95	0,96	0,99	
	R2	0,68	0,93	0,32	0,96	0,95	0,95	0,95	0,95	0,95	0,99	

de Discriminação quando se tem 10 itens por prova e para o Grau de Dificuldade quando se tem 50 itens por prova. Em relação ao Acerto Casual, os EQM's foram baixos (próximos de zero) e com valores equiparados entre as 3 estruturas de correlação usadas. Para o exame com 5 provas, o caso que se destaca é quando se tem 50 itens em cada prova, em que as estruturas $R1$ e $R2$ resultaram EQM's menores do que a estrutura I para o a e para b , com essa quantidade de itens, os EQM's foram muito altos.

Agora, para o parâmetro Habilidade, os EQM's são menores com o uso da estrutura $R1$ e $R2$ do que com I quando se tem 50 itens por prova, tanto para o exame com 3 quanto com 5 provas. Nos outros casos, apenas a estrutura $R1$ fez com que os EQM's fossem menores do que se usar as demais estruturas.

A seguir serão exibidos os gráficos referentes à interação entre número de provas e estruturas de correlações (Fig. 14, 16, 18 e 20) e os gráficos de interação entre exames com 3 e 5 provas (Fig. 15, 17, 19 e 21) para verificar a magnitude das correlações entre os valores paramétricos e os estimados em função do número de itens por prova, com $n = 2000$ indivíduos.

Tabela 7 Erro Quadrático Médio (EQM) para os parâmetros estimados em todas as configurações utilizadas ($n = 2000$)

Configurações			Erro Quadrático Médio (EQM)								
Provas	Itens	R	a	b	c	θ_1	θ_2	θ_3	θ_4	θ_5	Σ
3	10	I	0,45	0,77	0,04	2,01	1,99	1,98	—	—	<0,002
		R1	0,39	0,88	0,04	1,98	1,95	2,02	—	—	<0,002
		R2	0,38	0,82	0,04	2,08	2,03	2,02	—	—	<0,001
	30	I	0,24	0,60	0,02	2,02	1,95	1,98	—	—	<0,001
		R1	0,26	0,74	0,03	1,94	1,95	1,88	—	—	<0,001
		R2	0,26	0,67	0,03	2,09	2,09	2,08	—	—	<0,001
5	50	I	0,23	0,65	0,02	2,03	2,01	2,09	—	—	<0,001
		R1	0,25	0,60	0,03	1,96	1,89	2,00	—	—	<0,001
		R2	0,22	0,64	0,03	1,97	1,93	1,89	—	—	<0,001
	10	I	0,36	0,59	0,03	1,99	2,06	1,97	2,02	2,09	<0,003
		R1	0,39	0,67	0,04	2,00	2,02	2,03	2,08	2,05	<0,003
		R2	0,42	0,53	0,04	2,06	2,05	2,04	2,02	2,03	<0,003
30	I	0,22	0,60	0,03	2,02	1,95	2,00	2,02	2,03	<0,002	
	R1	0,25	0,60	0,03	1,99	1,98	1,97	2,04	1,93	<0,001	
	R2	0,22	0,61	0,03	2,01	2,05	2,01	2,08	2,02	<0,001	
50	I	0,47	2,87	0,03	2,03	2,00	1,98	1,98	2,02	<0,001	
	R1	0,43	2,59	0,03	1,95	2,03	1,92	1,91	1,98	<0,001	
	R2	0,43	2,83	0,03	1,92	1,96	1,95	1,94	2,08	<0,002	

Fig. 14 e 15 dizem respeito ao parâmetro Poder de Discriminação (a).

Na Fig. 14, as correlações entre a e \hat{a} em função do número de itens por prova foram maiores no exame com 3 provas e estrutura $R2$ quando se tem 50 itens em cada prova, sendo que para essa interação entre número de provas e estrutura, as correlações foram aumentando à medida que o número de itens por prova também aumentou; com 30 itens a interação que provocou um maior aumento nas correlações foi entre o exame com 5 provas e estrutura $R2$; quando se tem poucos itens (10), as correlações foram maiores na interação com a estrutura I no exame com 5 provas e menores no exame com 3. Na Fig. 15, as correlações aumentam à medida que o número de itens aumenta, para qualquer quantidade de provas por exame, sendo que com 10 e 30 itens o exame com 5 provas resultou em correlações maiores do que o com 3 provas.

Isso mostrou o efeito positivo de se usar as estruturas $R1$ e $R2$ comparado com o uso de I , no sentido de aumentar as correlações entre a e \hat{a} , em provas com quantidade maior de itens (30 e 50) e que o efeito de se ter 50 itens por prova favoreceu exames com menos provas.

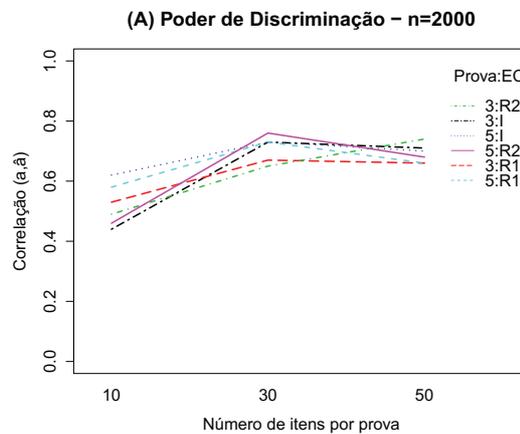


Figura 14 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 2000$ indivíduos

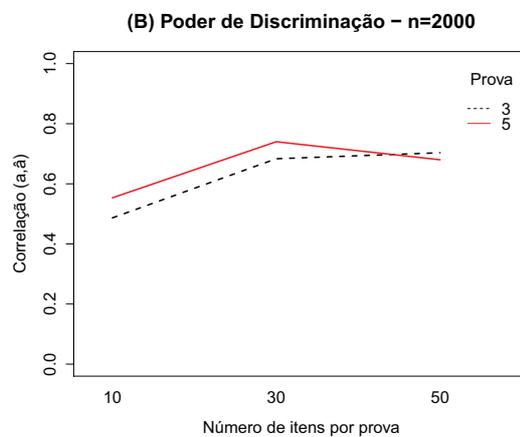


Figura 15 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Poder de Discriminação em função do número de itens por prova, com $n = 2000$ indivíduos

As Fig. 16 e 17 dizem respeito ao parâmetro Grau de Dificuldade (b). Na Fig. 16 percebe-se que as correlações entre b e \hat{b} são altas para qualquer interação

entre número de provas e estruturas de correlações. Mesmo altas, à medida que o número de itens por prova aumenta, as interações que retornaram em maiores correlações foram entre 3 provas por exame e $R1$ e $R2$, respectivamente. Para o exame com 5 provas, a interação com a estrutura I foi a que apresentou uma maior correlação entre b e \hat{b} .

Já na Fig. 17 pode-se observar que tanto o exame com 3 quanto o com 5 provas fazem com que as correlações entre b e \hat{b} sejam altas à medida que o número de itens por prova aumenta. Para provas com 30 e 50 itens os exames se equiparam; com 10 itens o exame com 5 provas resulta em uma correlação maior que no exame com 3.

Dessa forma, dá para ver que o número de provas por exame tem praticamente o mesmo efeito na magnitude das correlações entre b e \hat{b} . Mais ainda, mesmo com efeitos parecidos, as interações entre 3 provas e estruturas $R1$ e $R2$ têm melhor desempenho no aumento das correlações do que com a estrutura I . Para o exame com 5 provas, apenas a interação com a estrutura $R2$ teve melhor desempenho que com as demais, para poucos itens por prova.

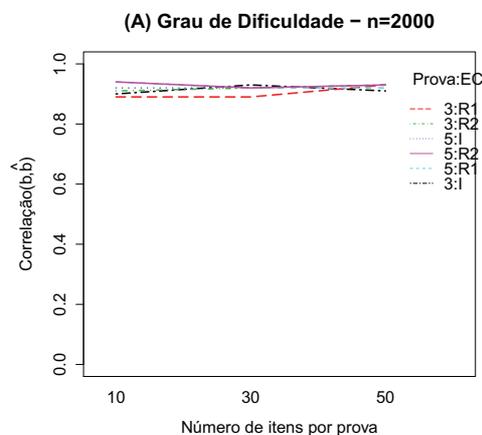


Figura 16 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 2000$ indivíduos

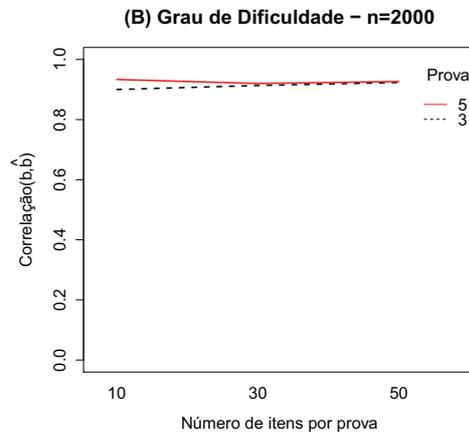


Figura 17 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Grau de Dificuldade em função do número de itens por prova, com $n = 2000$ indivíduos

As Fig. 18 e 19 dizem respeito ao parâmetro Acerto Casual (c). A Fig. 18 ilustra que as correlações entre c e \hat{c} são relativamente baixas comparadas a outros parâmetros de item, para qualquer interação entre número de provas por exame e estrutura de correlação e para qualquer quantidade de itens por prova. Ainda assim dá para perceber que dentre todas as interações, a que se dá entre 3 provas e estrutura $R2$ foi a que provocou um maior aumento nas correlações entre c e \hat{c} na medida em que o número de itens foi aumentado. Para o exame com 5 provas, apenas a interação com a estrutura $R2$ provocou um aumento maior do que com a estrutura I , mesmo assim, menores do que com 3 : $R2$. Na Fig. 19, tanto o exame com 3 como com 5 provas provocou um aumento na magnitude das correlações à medida que o número de itens por prova aumentou. O exame com mais provas só provocou um aumento maior na correlação entre c e \hat{c} quando o número de itens por prova foi pequeno (10). Para quantidades maiores de itens por prova (30,50) o exame com 3 provas aumentou a magnitude das correlações entre c e \hat{c} .

Com isso, exames com menos provas têm efeito superior ao de mais provas, no aumento das correlações entre c e \hat{c} , apenas para 30 e 50 itens por provas

e que, por mais que essas correlações tenham sido baixas, a interação entre o exame com 3 provas e estrutura R2 teve o melhor efeito dentro todas as outras interações no aumento das correlações.

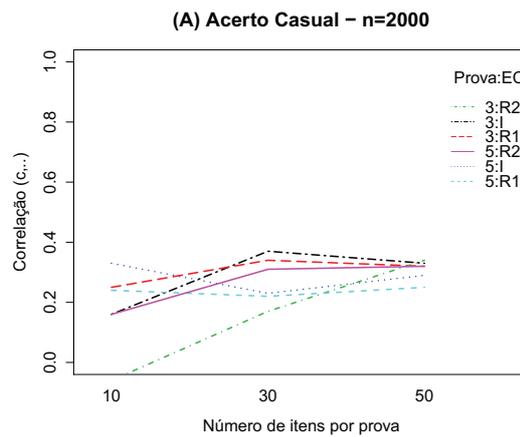


Figura 18 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 2000$ indivíduos

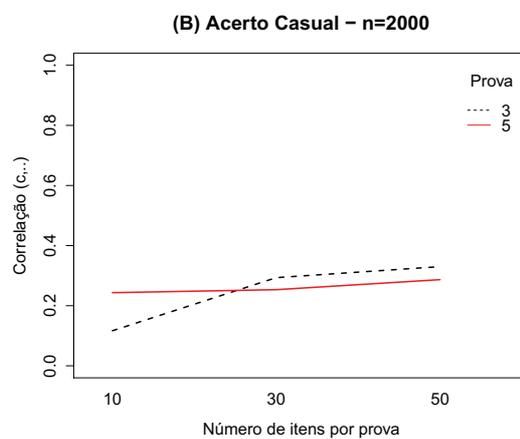


Figura 19 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para o Acerto Casual em função do número de itens por prova, com $n = 2000$ indivíduos

As Fig. 20 e 21 dizem respeito ao parâmetro Habilidade (θ). A Fig. 20 mostra que as correlações entre θ e $\hat{\theta}$ já são altas e aumentam à medida que o número de itens por prova aumenta, para todas as interações entre número de provas por exame e estruturas de correlações. Mesmo com desempenho parecido, as interações entre 3 provas e estrutura *R2* e 5 provas e estrutura *R1* acarretaram uma correlação entre θ e $\hat{\theta}$, para 10 e 30 itens por prova, maior que com as demais interações.

Já a Fig. 21 também indica que os exames com 3 e 5 provas aumentam a correlação entre θ e $\hat{\theta}$ à medida que o número de itens por prova aumenta. Mesmo parecidos o exame com 3 provas provoca um aumento um pouco maior que o de 5 provas nas correlações.

Com isso, dá para ver que, em relação às correlações entre θ e $\hat{\theta}$, exames com 3 provas conseguem obter correlações maiores e que as interações entre 3 provas e *R2* e 5 e *R1* só não conseguem um efeito melhor do que com *I* quando se tem provas com muitos itens (50, no caso).

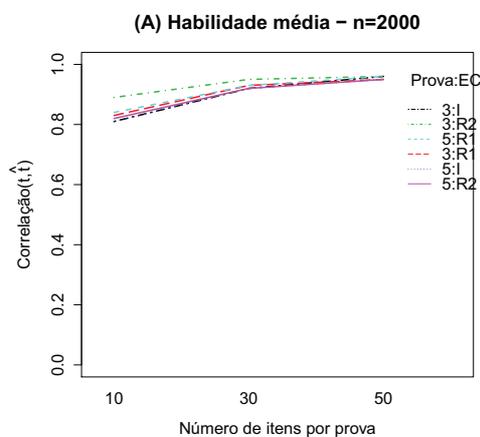


Figura 20 Gráfico de interação entre Provas e Estruturas de Correlações para verificar as correlações entre valores paramétricos e estimados para Habilidade média em função do número de itens por prova, com $n = 2000$ indivíduos

Dados os resultados e as discussões preliminares sobre os efeitos das interações entre número de provas por exame e estrutura de correlação (Prova:EC),

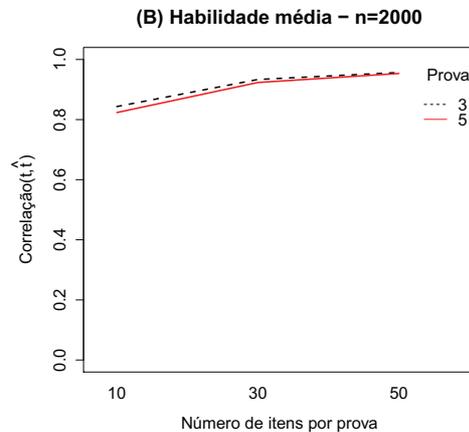


Figura 21 Gráfico de interação entre Provas para verificar as correlações entre valores paramétricos e estimados para a Habilidade média em função do número de itens por prova, com $n = 2000$ indivíduos

nas magnitudes das correlações entre os valores paramétricos e os estimados em função do aumento de número de itens por prova, deve-se comparar esses resultados em relação ao número de indivíduos que respondem aos itens dos exames nas simulações realizadas (1000 e 2000 indivíduos). Comparações parecidas foram feitas por de la Torre e Patz (2005) para verificar a consistência dos estimadores.

Em relação ao parâmetro Poder de Discriminação, com 10 itens por prova, o aumento do número de indivíduos de 1000 para 2000 favorece o uso da matriz identidade I como estrutura de correlação na interação com o exame que possui 5 provas (5 : I). Com 30 itens, o aumento no número de indivíduos favoreceu a interação 5 : R^2 e com 50 itens favoreceu a interação 3 : R^2 . Em relação ao número de provas em um exame, sem se preocupar com a estrutura de correlação usada e direcionando a atenção aos efeitos do número de itens por prova, quando se tem 1000 indivíduos, e a correlação entre a e \hat{a} é maior com o exame com 3, 5 e 5 provas quando se tem 10, 30 e 50 itens, respectivamente. Já quando se tem 2000 indivíduos a correlação entre a e \hat{a} é maior para o exame com 5, 5 e 3 provas quando se tem 10, 30 e 50 itens, respectivamente.

Em relação ao parâmetro Grau de Dificuldade, as correlações entre b e \hat{b}

se mostraram altas independente da quantidade de indivíduos respondentes. Com o aumento de 1000 para 2000 indivíduos o efeito das interações entre número de provas e estruturas de correlações foram em geral os mesmos, não mostrando mudanças significativas. Sem se preocupar com as estruturas de correlações usadas na simulação, os efeitos do número de provas por exame também não foram diferentes com esse aumento no número de indivíduos.

Para o parâmetro Acerto Casual, o aumento no número de indivíduos não alterou o padrão das magnitudes das correlações entre c e \hat{c} , que foram relativamente baixas, comparadas com os outros parâmetros de item, em geral variando entre 0.2 e 0.4. O efeito das interações entre número de provas por exame e estrutura de correlação nas magnitudes das correlações entre c e \hat{c} não mudaram, permanecendo a interação 3 : R2 como a que provocou maior aumento nessas correlações à medida em que o número de itens por prova aumenta. No que diz respeito ao número de provas em um exame, destacam-se as situações em que se tem 50 itens por prova, sendo que, quando se tem 1000 indivíduos, o exame com mais provas provoca um aumento maior nas correlações do que quando se tem menos provas e, quando se tem 2000 indivíduos acontece o contrário, favorecendo o exame com menos provas.

Com o parâmetro associado à habilidade, as correlações entre θ e $\hat{\theta}$ foram altas para as duas quantidades de indivíduos simuladas. Tanto para 1000 quanto para 2000 indivíduos, as diferentes interações entre número de provas e estrutura de correlação tiveram comportamentos semelhantes, provocando um aumento das correlações entre θ e $\hat{\theta}$ à medida que o número de itens por prova aumenta. A diferença em relação à quantidade de indivíduos respondentes, pequena que seja, se deu quando a quantidade de itens por prova foi igual a 50, em que a interação 3 : R2 retornou a maior correlação com 1000 indivíduos e, com 2000 indivíduos foi favorecida a interação 3 : I. Sem se preocupar com as estruturas de correlações, o efeito que o número de provas em um exame teve no aumento das correlações entre com θ e $\hat{\theta}$ foi o mesmo, para qualquer quantidade de itens por prova e para qualquer quantidade de indivíduos simulada.

4.2 Aplicação do modelo multivariado nos dados do vestibular 2006 – 2 da UFLA

Recorrendo-se ao método apresentado na seção 3.1, a metodologia proposta foi ilustrada com uma aplicação em dados reais. Os dados utilizados neste trabalho se referem ao Vestibular 2006 – 2 da UFLA. Esse vestibular foi composto por provas de Português, Geografia, História, Filosofia, Língua Estrangeira (Inglês ou Espanhol), Biologia, Física, Matemática e Química. Para a análise, por questões operacionais, uma amostra de 936 indivíduos (total de candidatos dos cursos de Medicina Veterinária, Zootecnia e Química) foi tomada da população de 3053 indivíduos candidatos a 10 diferentes cursos, por acreditar que essa mostra é homogênea em relação aos indivíduos, e não foi considerada a prova de Língua Estrangeira. Dessa forma, os dados usados na análise se referem aos 936 indivíduos respondendo a 8 provas em um exame Vestibular.

O diagnóstico de Gelman & Rubin indicou que para a maioria das cadeias houve convergência. Os gráficos associados a alguns exemplos com os parâmetros referentes ao poder de discriminação (a), grau de dificuldade (b) e acerto casual (c), respectivamente, pode ser visto nas Fig. 22,23,24. Os traços aqui plotados referem-se aos itens 27 e 28.

Verificou-se convergência das cadeias para os parâmetros referentes às habilidades. Na Fig. 25, estão os traços das respectivas cadeias para os indivíduos 29, 30, 31 e 32, na prova de Português. O mesmo pode ser dito para a convergência da estrutura de correlação R estimada entre as habilidades nas 8 provas desse exame, como pode ser visto na Fig. 26. Isto foi ilustrado para as correlações entre Português e Geografia (V2) e entre Português e História (V3).

O conjunto dos gráficos dos traços e densidades a posteriori dos parâmetros associados a itens, habilidades e correlações não foram mostrados aqui por questão de espaço. Apenas os itens precisaram de maiores tamanhos amostrais para a convergência. A seguir são apresentados os resultados obtidos para os parâmetros em questão.

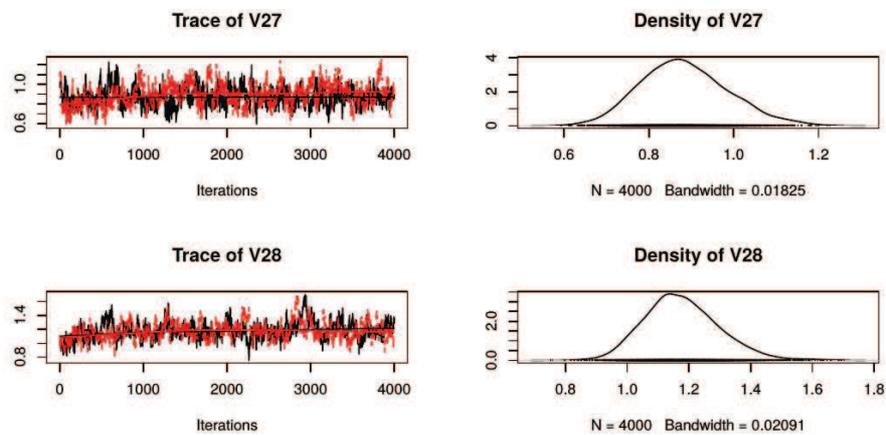


Figura 22 Traços das cadeias para o parâmetro a associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC, para os itens 27 e 28

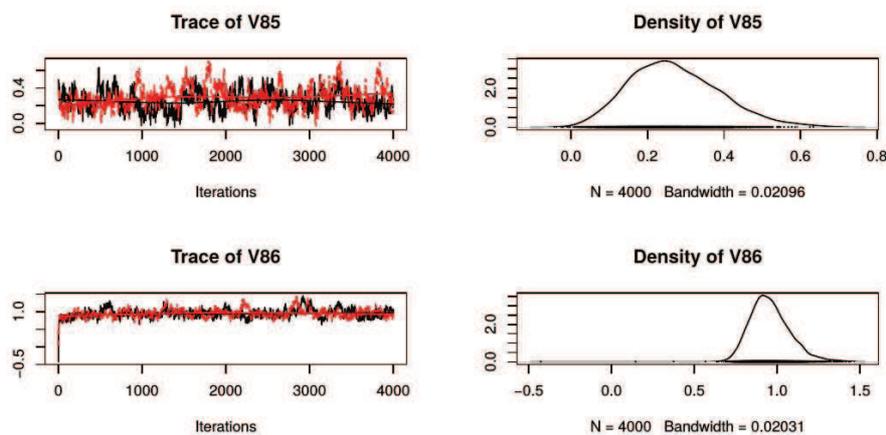


Figura 23 Traços das cadeias para o parâmetro b associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC, para os itens 27 e 28

Percebeu-se que os indivíduos desse exame possuem habilidades fortemente associadas, positivamente, umas às outras nas áreas, consideradas aqui, de humanas ou sociais (nesse caso, Português, Geografia, História, Filosofia) e também nas áreas de exatas e biológicas (Biologia, Física, Matemática, Química). Observou-se uma associação mais forte entre as habilidades das áreas de exatas e biológicas do que entre as habilidades em humanas e sociais. Entre as áreas, as

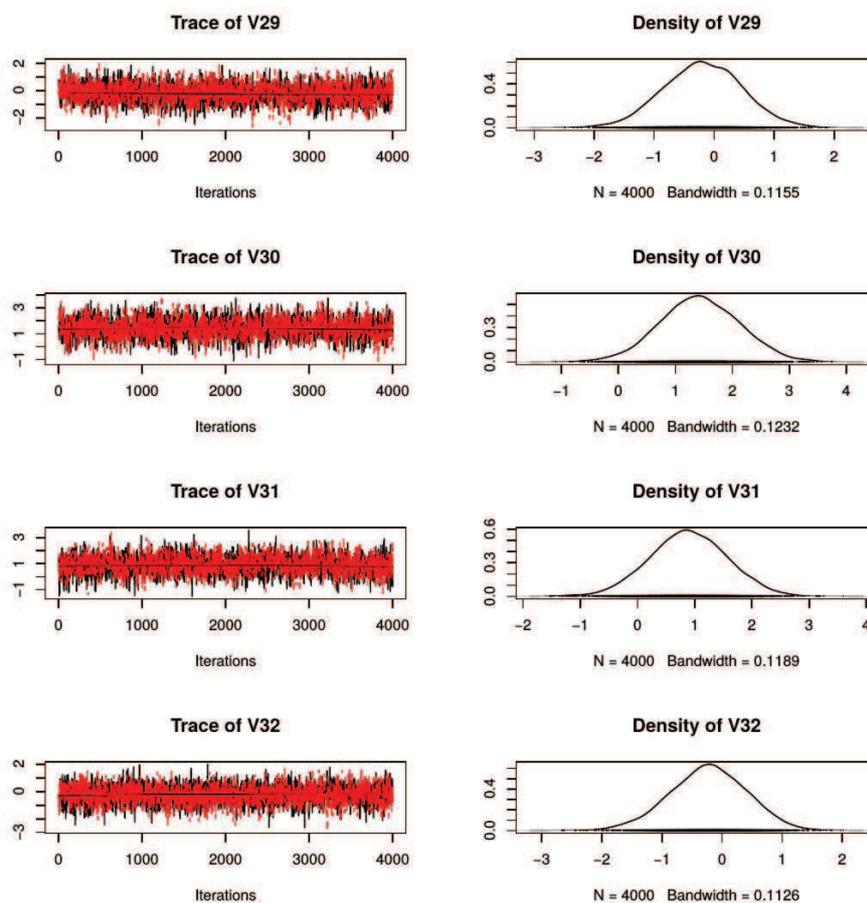


Figura 25 Traços das cadeias e densidade a posteriori para as correlações entre as habilidades em Português e Geografia (V2) e entre Português e História (V3). Valores da matriz de correlação entre as habilidades. Cadeias associadas ao diagnóstico de Gelman & Rubin, obtidas via MCMC

As correlações foram destacadas na Tab. 8 para ressaltar que esse resultado evidencia que o MP3PM discrimina grupos de habilidades por meio da relação de dependência entre as habilidades nas provas de um exame. Os dados desse exame vestibular permitem trabalhar com a ideia de que, se 2 grupos de habilidades estimadas fossem independentes, a partir daí poderiam ser discriminadas 2 grandes áreas do conhecimento.

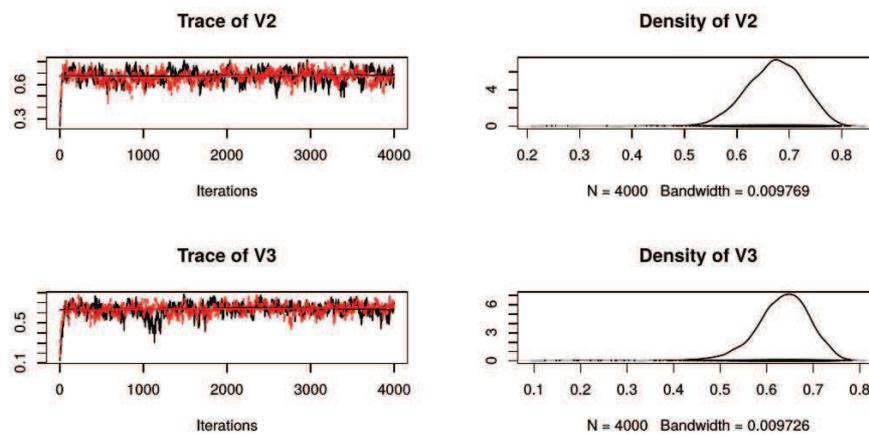


Figura 26 Traços das cadeias e densidade a posteriori para as estimativas das correlações entre as habilidades em Português e Geografia (V2) e entre Português e História (V3), dada pela estimativa da matriz de correlação entre as habilidades, no diagnóstico de Gelman & Rubin, obtidas via MCMC

As correlações entre as notas obtidas pelos indivíduos e as habilidades estimadas nas provas estão descritas na Tab. 9. As notas foram centralizadas e se referem ao número de acertos obtidos por cada indivíduo em cada prova, e a habilidade em cada prova é a habilidade média do indivíduo.

A Tab. 9 indica que a habilidade estimada tem relação direta com a nota obtida numa mesma prova. Vale ressaltar que a nota em Matemática tem pouca associação com a habilidade do indivíduo nessa mesma prova, se comparada com as correlações entre nota e habilidade das demais provas. As correlações com (*) são significativas ao nível de 5% de probabilidade para o teste de t , dizendo que são diferentes de zero. Isso poderá significar, por exemplo, o seguinte: a nota do indivíduo na prova de História pode ser prejudicada se ele tiver habilidade nas provas de Física, Matemática e Química.

A seguir, na Tab. 10, encontram-se descritas as estimativas *a posteriori* por ponto (média) e por intervalos de credibilidade (IC-HPD) para alguns parâmetros de itens estimados. No Apêndice 1 encontra-se a Tab. 16 com essas estimativas

Tabela 9 Correlações entre as notas obtidas no teste e as habilidades estimadas nas provas pelo modelo multivariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química

nota x θ	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
nota.Por	0,79	0,54	0,51	0,51	0,02	0,00	0,01	-0,01
nota.Geo		0,81	0,58	0,51	-0,06	-0,09*	-0,07*	-0,10*
nota.His			0,59	0,38	-0,05	-0,07*	-0,07*	-0,09*
nota.Fil				0,89	-0,05	-0,08*	-0,07*	-0,09*
nota.Bio					0,87	0,83	0,83	0,80
nota.Fis						0,38	0,35	0,34
nota.Mat							0,15	0,12
nota.Qui								0,72

para todos os itens do exame.

Para efeito de discussão, consideramos que: um item discrimina bem o candidato quando o $\hat{a} \geq 1$, um item tem grau de dificuldade média quando $\hat{b} = 0$ e um item tem probabilidade baixa de ser respondido corretamente por mero acaso (acerto casual) quando $\hat{c} \leq 0,25$. E também será considerado que um bom item é aquele que apresenta bom poder de discriminação, grau baixo ou médio de dificuldade e acerto casual menor ou igual a 0,25.

Sendo assim, na Tab. 10 cabe ressaltar alguns casos: ela mostra que os itens da prova de Português têm discriminação baixa, já que o item que apresentou o maior poder de discriminação foi o item 1 ($\hat{a} = 0,66$) e, dentre os outros, só foi considerado bom item porque o grau de dificuldade foi baixo e o acerto casual foi praticamente nulo. Outro caso que vale destacar diz respeito às provas de Física, Matemática e Química, as quais tiveram itens com bom poder de discriminação em questões difíceis, mas com alta probabilidade de acerto casual.

Na Tab. 11 estão ilustradas as habilidades estimadas para os indivíduos números 29; 30; 31; 32 que responderam aos itens da prova de Português. Percebe-se que tanto para indivíduos com habilidade alta (acima de zero) quanto para indivíduos com habilidade baixa (abaixo do zero) as estimativas apresentaram uma variabilidade baixa. Em geral isso aconteceu, inclusive para as

habilidades estimadas nas outras provas.

Tabela 10 Estimativas *a posteriori* por ponto (média) e por intervalo (IC-HPD com 95%) para os parâmetros alguns itens de todas as provas do Vestibular 2006 – 2 da UFLA, considerados bons (itens destacados em negrito) e ruins (itens sem destaques)

Provas	Itens	a			b			c		
		média	IC-HPD		média	IC-HPD		média	IC-HPD	
			inf	sup		inf	sup		inf	sup
Português	1	0,66	0,44	0,88	-0,64	-0,87	-0,36	0,19	0,05	0,36
	9	0,32	0,13	0,50	-0,48	-0,81	-0,11	0,25	0,04	0,48
Geografia	11	0,81	0,43	1,42	0,37	-0,09	1,01	0,25	0,08	0,40
	12	0,53	0,19	1,09	0,38	-0,19	1,37	0,30	0,09	0,53
História	18	1,09	0,69	1,57	-0,69	-0,93	-0,46	0,20	0,06	0,35
	19	1,61	0,99	2,30	-0,66	-1,04	-0,34	0,21	0,08	0,33
Filosofia	20	1,34	0,13	2,27	4,13	2,60	5,80	0,52	0,48	0,55
	24	0,80	0,32	1,59	0,77	0,17	1,63	0,23	0,09	0,36
Biologia	25	2,03	1,16	2,64	-1,10	-1,77	-0,58	0,23	0,06	0,41
	26	1,41	0,75	2,36	-0,45	-0,76	-0,12	0,22	0,06	0,37
Física	28	1,14	0,93	1,36	0,93	0,72	1,14	0,04	0,01	0,07
	29	0,03	0,00	0,08	0,07	-0,24	0,46	0,23	0,06	0,41
Matemática	30	1,30	1,03	1,62	1,17	0,89	1,48	0,04	0,01	0,07
	34	1,74	1,39	2,11	0,28	0,10	0,47	0,07	0,02	0,11
Química	35	0,78	0,31	1,20	1,87	1,18	2,53	0,18	0,14	0,23
	38	1,44	0,89	1,93	2,82	2,16	3,57	0,34	0,30	0,38
Química	40	1,17	0,69	1,60	1,35	0,80	1,96	0,26	0,20	0,32
	43	1,78	0,87	2,65	3,09	1,75	4,44	0,18	0,14	0,21
Química	50	0,41	0,00	1,07	3,29	1,15	4,68	0,29	0,24	0,34
	51	1,31	1,05	1,62	1,13	0,92	1,37	0,03	0,01	0,05
Química	53	0,02	0,00	0,04	0,09	-0,28	0,64	0,26	0,06	0,48
	56	2,33	0,77	3,51	2,87	0,84	4,32	0,26	0,17	0,31
Química	57	1,25	0,84	1,75	1,03	0,61	1,58	0,10	0,04	0,16

Nessa Tab. 11, pontualmente, mesmo existindo médias abaixo, mas próximas, de zero; e outras acima de zero (na maioria dos casos), indica que o indivíduo possui habilidade alta no conteúdo de Português. Porém, o intervalo de credibilidade (IC-HPD), em geral, contém o zero, indicando que as habilidades dos indivíduos têm valor médio.

As habilidades nas demais provas também foram bem estimadas, como pode ser visto em alguns traços das cadeias que estão no Apêndice 2. Foram escolhidos apenas 4 indivíduos por questão operacional, com a intenção de representar todos os outros, mas vale ressaltar que as habilidades de todos os 936 indivíduos que fizeram as provas desse exame vestibular tiveram suas habilidades bem estimadas.

Tabela 11 Estimativas a *posteriori* por ponto (média) e por intervalo (IC-HPD com 95%) para os parâmetros Habilidades dos indivíduos 29,30,31,32 calculadas com o modelo multivariado na prova de Português do Vestibular 2006 – 2 da UFLA

Indivíduos	θ_1 :Português - modelo multivariado			
	média	IC-HPD		desvio padrão
		inferior	superior	
29	-0,20	-1,61	1,07	0,68
30	1,39	-0,07	2,72	0,72
31	0,88	-0,42	2,33	0,70
32	-0,30	-1,60	1,01	0,68

No Apêndice C encontram-se os traços das cadeias para as demais provas para os mesmos indivíduos (Fig. 28, 29, 30, 31, 32, 33 e 34). Esses gráficos foram inseridos com o objetivo de mostrar que as habilidades foram bem estimadas pelo método proposto. Foram escolhidos os mesmos indivíduos citados na Tab. 11, para que os traços ilustrem a qualidade das estimativas das habilidades desses indivíduos.

4.3 Comparação do modelo multivariado proposto com a aplicação do modelo univariado nos dados do vestibular 2006 – 2 da UFLA

A seguir, são apresentados os resultados da aplicação de um modelo para análise univariada de habilidades em cada uma das 8 provas do vestibular, descritas anteriormente. Para cada resultado univariado é feita uma comparação com o multivariado.

A Tab. 12 se refere às correlações entre as habilidades em cada uma das 8 provas e mostra que as habilidades estimadas para cada prova, pelo modelo univariado, foram em geral baixas. Foram identificados os mesmos grupos de habilidades discriminadas com a análise do modelo multivariado (correlações em negrito), porém, evidenciando pouca associação entre as habilidades nas provas pertencentes ao grupo de ciências humanas e sociais e também ao grupo de ciências exatas e, ou, biológicas. A correlação entre as habilidades nas provas do grupo de humanas

com as habilidades no grupo de exatas e biológicas foi praticamente nula. Entretanto, para ter uma confiança maior em se afirmar que são nulas, foi feito um teste t para as correlações entre as habilidades desse grupo (correlações sem destaque em negrito), com 5% de significância (BEST; ROBERTS, 1975).

Tabela 12 Correlações entre as habilidades nas provas, estimadas pelo modelo univariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química

θ_i	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
θ_1	1,00	0,29	0,25	0,31	0,01	0,11*	0,08*	0,01
θ_2		1,00	0,29	0,28	-0,03	0,09*	0,04	-0,05
θ_3			1,00	0,25	-0,04	0,10*	-0,02	-0,07*
θ_4				1,00	-0,02	0,06	0,05	-0,05
θ_5					1,00	0,17	0,20	0,63
θ_6						1,00	0,15	0,22
θ_7							1,00	0,22
θ_8								1,00

Podemos observar que as correlações, marcadas com asterisco (*), foram significativas ao nível de 5%, concluindo que essas correlações são diferentes de zero. Logo, além dos grupos discriminados aqui, a habilidade na prova de Português também está associada de forma positiva às habilidades nas provas de Física e Matemática, a habilidade em Geografia está associada de forma positiva com a habilidade em Física e a habilidade em História está associada de forma negativa à habilidade em Química.

Isso mostra que quando um exame com várias provas é analisado de forma univariada, não se consegue um alto grau de associação entre as habilidades nas provas dos grupos, discriminados como no caso multivariado e, mais ainda, existe associação entre habilidades do grupo de humanas com o de exatas e biológicas, mesmo que baixa.

A Tab. 13 indica se as notas obtidas e as habilidades estimadas nas provas desse exame podem ser interpretadas com o mesmo sentido e também se a nota obtida teve relação com mais de uma habilidade. Na diagonal estão os valores das correlações entre a nota obtida e a habilidade estimada na mesma prova, indicando

que existe forte relação. Os valores fora da diagonal indicam que a nota em uma determinada prova está associada a mais de uma habilidade, inclusive entre provas do grupo de humanas e de exatas. Isso pode ser visto nas correlações com a nota de Português e História, as quais ficaram relacionadas às habilidades nas provas de humanas e também à habilidade em Física, pois as correlações entre a nota de Português e História com a habilidade em Física foram significativas ao nível de 5% de probabilidade para o teste de t (valores com *).

Tabela 13 Correlações entre as notas obtidas e as habilidades estimadas nas provas pelo modelo univariado, sendo θ_1 : Português, θ_2 : Geografia, θ_3 : História, θ_4 : Filosofia, θ_5 : Biologia, θ_6 : Física, θ_7 : Matemática e θ_8 : Química

nota x θ	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
nota.Por	0,90	0,28	0,23	0,31	0,02	0,12*	0,05	0,02
nota.Geo		0,92	0,27	0,27	-0,04	0,05	0,05	-0,06
nota.His			0,78	0,20	-0,04	0,12*	-0,04	-0,05
nota.Fil				1,00	-0,03	0,06	0,05	-0,06
nota.Bio					0,92	0,17	0,20	0,57
nota.Fis						0,85	0,12	0,27
nota.Mat							0,80	0,09
nota.Qui								0,89

Na Tab. 14, encontram-se descritas as estimativas a *posteriori* por ponto (média) e por intervalo de credibilidade (IC-HPD) de 95% para alguns parâmetros de itens estimados na análise univariada. No Apêndice 2 encontra-se a Tab. 17 com essas estimativas para todos os itens do exame.

Para efeito de discussão, assim como foi dito nos resultados da análise multivariada, será considerado que: um item discrimina bem o candidato quando $\hat{a} \geq 1$, um item tem grau de dificuldade médio quando $\hat{b} = 0$ e um item tem probabilidade baixa de ser respondido corretamente por mero acaso (acerto casual) quando $\hat{c} \leq 0,25$. E também será considerado que um bom item é aquele que apresenta bom poder de discriminação, grau de dificuldade baixo ou médio e acerto casual menor ou igual a 0,25.

Os resultados da Tab. 14, comparados com os resultados da Tab. 10, obti-

dos com a análise multivariada, mostram que os itens 1 e 9 da prova de Português não foram considerados bons itens no caso univariado, diferindo assim do resultado na análise multivariada. Mesmo sendo itens com maior poder de discriminação do que no caso multivariado, esses itens apresentaram uma alta probabilidade de serem respondidos corretamente por mero acaso ($\hat{c} > 0,25$). Outra diferença para os resultados da análise multivariada foi em relação aos itens da prova de Física. No contexto multivariado, o item 38 tinha sido identificado como ruim. Já no contexto univariado, o item considerado ruim foi o 40 por ter $\hat{c} > 0,25$, mesmo com bom poder discriminação.

Tabela 14 Estimativas *a posteriori* por ponto (média) e por intervalo (IC-HPD com 95%), na análise univariada, para os parâmetros de alguns itens de todas as provas do Vestibular 2006 – 2 da UFLA

Provas	Itens	a			b			c		
		média	IC-HPD		média	IC-HPD		média	IC-HPD	
			inf	sup		inf	sup		inf	sup
Português	1	0,74	0,48	1,05	-0,66	-1,45	0,34	0,31	0,00	0,58
	9	0,79	0,42	1,22	1,00	-0,66	2,96	0,65	0,39	0,78
Geografia	11	0,87	0,50	1,31	0,84	0,13	1,48	0,36	0,19	0,49
	12	0,91	0,51	1,38	1,49	0,71	2,41	0,50	0,41	0,59
História	18	1,07	0,67	1,51	-0,69	-1,10	-0,19	0,18	0,00	0,40
	19	0,97	0,57	1,49	-0,36	-1,03	0,55	0,26	0,00	0,55
	20	1,04	0,57	1,58	2,56	1,92	3,00	0,50	0,45	0,55
	24	0,94	0,52	1,41	0,74	0,11	1,35	0,21	0,02	0,36
Filosofia	25	1,15	0,73	1,64	-0,79	-1,18	-0,40	0,16	0,00	0,37
	26	1,13	0,70	1,64	-0,56	-0,87	-0,21	0,12	0,00	0,29
Biologia	28	1,25	0,96	1,57	0,74	0,60	0,89	0,03	0,00	0,06
	29	1,16	0,68	1,71	2,63	2,13	3,00	0,58	0,55	0,62
	30	1,28	0,96	1,63	0,83	0,70	0,98	0,02	0,00	0,05
	34	1,86	1,44	2,39	0,13	0,00	0,27	0,04	0,00	0,10
Física	35	0,94	0,54	1,40	1,85	1,05	2,80	0,14	0,01	0,23
	38	0,96	0,52	1,43	1,34	0,47	2,39	0,21	0,03	0,36
	40	1,04	0,56	1,57	2,25	1,24	3,00	0,36	0,25	0,44
	43	1,00	0,55	1,52	1,97	1,18	2,92	0,15	0,03	0,23
Matemática	50	1,19	0,64	1,76	2,78	2,40	3,00	0,29	0,25	0,32
	51	1,08	0,78	1,40	0,91	0,72	1,09	0,03	0,00	0,08
Química	53	1,25	0,75	1,81	2,65	2,18	3,00	0,61	0,58	0,64
	56	0,91	0,54	1,32	1,08	0,69	1,44	0,17	0,05	0,27
	57	1,03	0,74	1,39	0,71	0,50	0,95	0,06	0,00	0,13

Na Tab. 15 e na Fig. 27 estão ilustradas as habilidades estimadas por meio de uma análise univariada de habilidades para os indivíduos números 29,30,31,32 que responderam aos itens da prova de Português. Essa prova foi escolhida como

exemplo para mostrar como ficaram as habilidades estimadas e se elas foram bem estimadas.

As habilidades nas demais provas também foram bem estimadas, mas por questão de organização do texto não foram mostradas aqui. Foram escolhidos apenas 4 indivíduos por questão operacional, com a intenção de representar todos os outros, mas vale ressaltar que todos os 936 indivíduos que fizeram as provas desse exame vestibular foram, assim como na análise multivariada, bem estimados.

A Tab. 15 teve a intenção de mostrar que as habilidades estimadas para os indivíduos tiveram valores médios no conteúdo de Português ($\theta = 0$). Pontualmente, existem médias abaixo e acima de zero, mas não tão altas como a análise multivariada estimou, indicando que o indivíduo possui habilidade em torno do valor zero no conteúdo de Português. Porém, o intervalo de credibilidade (IC-HPD), em geral, contém o zero, indicando que as habilidades dos indivíduos têm valor médio.

Tabela 15 Estimativas a *posteriori* por ponto (média) e por intervalo (IC-HPD com 95%) para os parâmetros Habilidades dos indivíduos 29,30,31,32 calculadas com o modelo univariado na prova de Português do Vestibular 2006 – 2 da UFLA

Indivíduos	θ_1 :Português - modelo univariado			
	média	IC-HPD		desvio padrão
		inferior	superior	
29	-0,09	-1,85	1,62	0,91
30	0,38	-1,61	2,38	1,03
31	0,24	-1,68	2,06	0,97
32	-0,17	-2,00	1,60	0,95

A Fig. 27 ilustra os traços das cadeias e densidades a *posteriori*, para o caso univariado, de alguns indivíduos na prova de Português. Os gráficos de traços das habilidades univariadas estimadas nas outras provas, para esses mesmos indivíduos, também mostram que foram bem estimadas, porém não foram apresentadas aqui. Esses gráficos foram inseridos com o objetivo de indicar que as habilidades foram bem estimadas pelo método univariado, assim como no multivariado.

Foram escolhidos os mesmos indivíduos citados na Tab. 15, para que os

gráficos da Fig. 27 ilustrem a qualidade das estimativas das habilidades desses indivíduos.

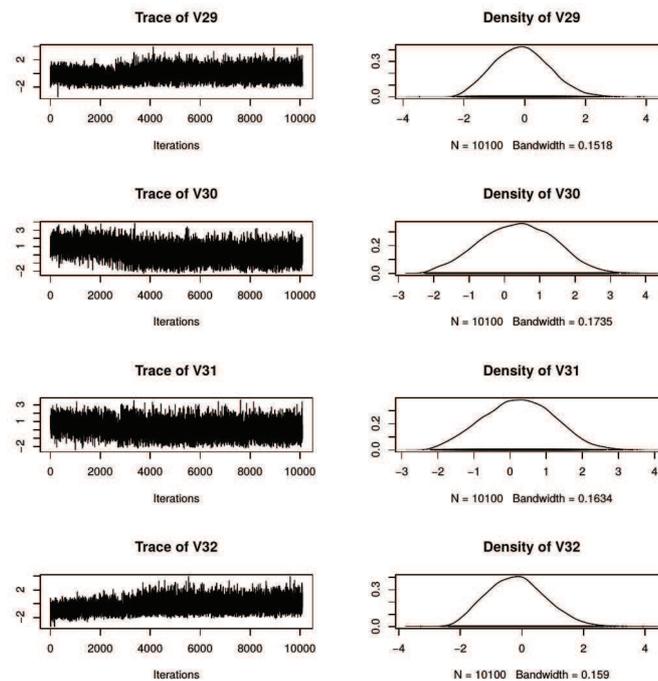


Figura 27 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 estimadas com o modelo univariado na prova de Português

Visto isso, pode-se afirmar que as habilidades estimadas por meio da análise multivariada resultaram em valores mais distantes do zero e também em intervalos de credibilidade com amplitudes menores do que os estimados pela análise univariada. Assim, pode-se dizer que as estimativas para as habilidades no caso multivariado foram mais consistentes do que no univariado.

Em ambas as análises, as habilidades foram bem estimadas, porém com maior precisão no modelo multivariado.

Comparar análises univariadas e multivariadas de habilidades significa co-tejar análises mais simples e fáceis de estimar com análises mais complexas que,

por serem mais flexíveis, permitem gerar hipóteses que revelam melhor as interações entre as habilidades em estudo.

A flexibilidade da análise multivariada não foca apenas uma habilidade individual e específica, desvinculada, por si só. Ela revela o potencial do avaliado (vestibulando) por permitir conhecer este potencial de maneira ampliada: é daí que tal análise revela a associação que existe entre as diversas habilidades. Assim, o vestibulando é conhecido não apenas por ser detentor de uma dada habilidade em uma dada disciplina, mas em diferentes disciplinas que dialogam entre si.

Pode-se dizer que os resultados de uma análise multivariada de habilidades ampliam a informação sobre o conhecimento do examinado. Seu potencial de discriminação é maior porque a associação que se faz por esse tipo de análise revela muito mais em termos de conhecimentos de cada indivíduo. Correlacionam-se objetos de estudo diferentes, mas que exigem habilidades cognitivas afins.

4.4 Considerações gerais sobre os resultados

As estimativas dos parâmetros de itens e habilidades, as matrizes de variâncias e covariâncias e estruturas de correlações estimadas no estudo de simulação, foram consistentes.

Na aplicação do modelo proposto ao Vestibular 2006 – 2 da UFLA, composto por 8 provas de diferentes conteúdos percebeu-se que:

- a) houve uma discriminação forte de de habilidades em duas áreas. Uma primeira é o grupo de habilidades em ciências humanas e sociais e uma segunda é o grupo de habilidades em ciências exatas e biológicas. Esta separação se deu com alta associação positiva entre as habilidades nas provas dentro de cada área;
- b) A nota obtida em uma prova de uma determinada área foi influenciada positivamente pelas habilidades na mesma área e ligeiramente influenciada de forma negativa pelas habilidades da outra área.

Não houve grande mudança qualitativa em relação ao ajuste dos modelos univariados, mas houve mudança em termos de precisão e inerpretabilidade.

Os resultados permitem concluir que o modelo multivariado é mais poderoso em identificar associações e estreitar os intervalos de credibilidade das estimativas.

A implementação com dados aumentados do modelo multivariado é consistente na estimação dos parâmetros envolvidos em provas do tipo vestibular.

No Vestibular 2006 – 2, formaram-se áreas de habilidades entre as provas, que correspondem vagamente a agrupamentos de interesse dos ingressantes.

O modelo multivariado é mais preciso que o univariado na interpretação deste tipo de prova e deve ser preferido.

Pretende-se expandir o modelo proposto aqui para um modelo multivariado com múltiplos grupos e, também, comparar o modelo multivariado com o multidimensional.

5 CONCLUSÃO

A implementação com dados aumentados do modelo multivariado é consistente na estimação dos parâmetros envolvidos em provas do tipo vestibular.

Não houve mudança qualitativa em relação ao ajuste dos modelos univariados, mas houve mudança em termos de precisão e interpretabilidade.

O modelo multivariado é mais poderoso em identificar associações e estreitar os intervalos de credibilidade das estimativas de itens e habilidades.

No Vestibular 2006 – 2, formaram-se áreas de habilidades entre as provas, que correspondem vagamente a agrupamentos de interesse dos ingressantes.

O modelo multivariado é mais preciso que o univariado na interpretação de vestibulares com várias provas e deve ser preferido.

REFERÊNCIAS

- ALBERT, J. H. Bayesian estimation of normal ogive item response functions using Gibbs sampling. **Journal of Educational Statistics**, Washington, v. 17, n. 2, p. 251-269, June 1992.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item: conceitos e aplicações**. Rio de Janeiro: SINAPE, 2000. 154 p.
- AZEVEDO, C. L. N. **Métodos de estimação na teoria da resposta ao item**. 2003. 121 p. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo, 2003.
- AZEVEDO, C. L. N. **Modelos longitudinais de grupos múltiplos multiníveis na teoria da resposta ao item: métodos de estimação e seleção estrutural sob uma perspectiva bayesiana**. 2008. 231 p. Tese (Doutorado em Estatística) - Universidade de São Paulo, São Paulo, 2008.
- AZEVEDO, C. L. N.; ANDRADE, D. F.; FOX, J. A Bayesian generalized multiple group IRT model with model-fit assessment tools. **Computational Statistics and Data Analysis**, Amsterdam v. 56, p. 4399–4412, Dec. 2012.
- BÉGUIN, A. A.; GLAS, C. A. W. MCMC estimation and some model-fit analysis of multidimensional IRT models. **Psychometrika**, Williamsburg, v. 66, n. 4, p. 541-562, Dec. 2001.
- BEST, D. J.; ROBERTS, D. E. Algorithm AS 89: the upper tail probabilities of spearman's rho. **Applied Statistics**, Abington, v. 24, p. 377-379, 1975.
- BIRNBAUM, A. **Statistical theories of mental test scores**. Reading: A. Wesley, 1968. 568 p.
- BOCK, D. R.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: an application of an EM algorithm. **Psychometrika**, Williamsburg, v. 46, n. 4, p. 433-459, 1981.

BOCK, D. R.; ZIMOWSKI, M. F. The multiple groups IRT. In: _____. **Handbook of modern item response theory**. 5th ed. New York: Springer-Verlag, 1997. p. 433–448.

BRAGION, M. L. L. **Análise combinada de exames vestibulares da Universidade Federal de Lavras usando a teoria de resposta ao item**. 2010. 187 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2010.

BRAGION, M. L. L. **Um modelo de teoria de resposta ao item para os dados do vestibular 2006-2 da UFLA**. 2007. 88 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2007.

DE LA TORRE, J.; PATZ, J. R. Making the most of what we have: a practical application of multidimensional item response theory in test scoring. **Journal of Educational and Behavioral Statistics**, Washington, v. 30, n. 3, p. 295-311, 2005.

HARTIG, J.; HÖHLER, J. Multidimensional IRT models for the assessment of competencies. **Studies in Educational Evaluation**, Elmsford, v. 35, n. 2/3, p. 57-63, June/Sept. 2009.

LORD, F. M. **Applications of item response theory to practical testing problems**. Erlbaum: Hillsdale, 1980. 274 p.

LORD, F. M. **A theory os test scores**. 7th ed. Richmond: Psychometric Monograph, 1952. 84 p.

MATOS, G. S. **Teoria da resposta ao item: uma proposta de modelo multivariado**. 2001. 134 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco, Recife, 2001.

MISLEVY, R. J. Bayes modal estimation in item response models. **Psychometrika**, Williamsburg, v. 51, n. 2, p. 177–195, 1986.

MISLEVY, R. J. Estimating latent distribution. **Psychometrika**, Williamsburg, v. 49, n. 3, p. 359–381, 1984.

MISLEVY, R. J. Item response models for grouped data. **Journal of Educational Statistics**, Washington, v. 8, n. 4, p. 271–288, 1983.

MISLEVY, R. J. Randomization-based inference about latent variables from complex samples. **Psychometrika**, Williamsburg, v. 56, n. 2, p. 177–196, 1991.

NOJOSA, R. T. **Modelos multidimensionais na teoria da resposta ao item**. 2001. 111 p. Dissertação (Mestrado em Ciências Matemáticas e de Computação) - Universidade Federal de Pernambuco, Recife, 2001.

PATZ, J. R.; JUNKER, B. W. Applications and extensions of mcmc in irt: multiple item types, missing data, and rated responses. **Journal of Educational and Behavioral Statistics**, Washington, v. 24, n. 4, p. 342-366, 1999a.

PATZ, J. R.; JUNKER, B. W. A straightforward approach to markov chain monte carlo methods for item response models. **Journal of Educational and Behavioral Statistics**, Washington, v. 24, n. 2, p. 146–178, 1999b.

PLUMMER, M. et al. CODA: convergence diagnosis and output analysis for MCMC. **R News**, New York, v. 6, n. 17, p. 7–11, 2006.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna, 2014. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 ago. 2014.

RASCH, G. **Probabilistic models for some intelligence and attainment tests**. Copenhagen: Danish Institute for Educational Research, 1960. 199 p.

RECKASE, M. D. **Multidimensional item response theory**. New York: Springer, 2009. 355 p.

SHENG, Y.; HEADRICK, T. C. Bayesian hierarchical modeling with 3PNO item response models. **American Journal of Mathematics and Statistics**, Rosemead, v. 3, n. 5, p. 281–287, 2013.

TAVARES, H. R. **Teoria da resposta ao item para dados longitudinais**. 2001. 131 p. Tese (Doutorado em Estatística) - Universidade de São Paulo, São Paulo, 2001.

THOMAS, N.; GAN, N. Generating multiple imputations for matrix sampling data analyzed with item response models. **Journal of Educational and Behavioral Statistics**, Washington, v. 22, n. 4, p. 425–445, 1997.

APÊNDICES

APÊNDICE A - Estimativas a posteriori por ponto e por intervalo dos parâmetros de itens na análise multivariada do vestibular 2006-2

Tabela 16 Estimativas a posteriori por ponto (média) e por intervalo (IC-HPD) para os parâmetros de itens do Vestibular 2006 – 2 da UFLA

Provas	Itens	a			b			c		
		média	IC-HPD		média	IC-HPD		média	IC-HPD	
			inf	sup		inf	sup		inf	sup
Português	1	0,66	0,44	0,88	-0,64	-0,87	-0,36	0,19	0,05	0,36
	2	0,54	0,35	0,74	0,03	-0,23	0,33	0,2	0,06	0,34
	3	0,32	0,00	0,69	1,68	0,43	3,71	0,31	0,16	0,43
	4	0,45	0,25	0,65	0,08	-0,22	0,45	0,21	0,07	0,38
	5	0,50	0,29	0,73	0,08	-0,26	0,45	0,18	0,04	0,37
	6	0,17	0,00	0,48	3,11	0,99	4,69	0,25	0,14	0,31
	7	0,59	0,34	0,88	0,54	0,20	0,90	0,16	0,05	0,26
	8	0,17	0,00	0,45	2,63	1,35	4,03	0,13	0,08	0,17
	9	0,32	0,13	0,50	-0,48	-0,81	-0,11	0,25	0,04	0,48
	10	0,53	0,31	0,80	0,47	0,14	0,90	0,17	0,05	0,30
Geografia	11	0,81	0,43	1,42	0,37	-0,09	1,01	0,25	0,08	0,40
	12	0,53	0,19	1,09	0,38	-0,19	1,37	0,3	0,09	0,53
	13	0,40	0,21	0,59	0,28	-0,04	0,65	0,21	0,06	0,34
	14	0,43	0,20	0,68	0,62	0,18	1,08	0,25	0,11	0,38
	15	0,44	0,25	0,64	-0,26	-0,56	0,18	0,24	0,06	0,47
	16	0,38	0,03	0,83	1,33	0,53	2,23	0,23	0,10	0,33
	17	0,67	0,44	1,00	0,20	-0,08	0,57	0,22	0,09	0,35
	18	1,09	0,69	1,57	-0,69	-0,93	-0,46	0,2	0,06	0,35
História	19	1,61	0,99	2,30	-0,66	-1,04	-0,34	0,21	0,08	0,33
	20	1,34	0,13	2,27	4,13	2,60	5,80	0,52	0,48	0,55
	21	0,48	0,00	0,97	1,87	1,01	2,74	0,22	0,13	0,28
	22	0,39	0,00	1,03	3,38	1,67	5,41	0,22	0,18	0,25
	23	0,40	0,08	0,96	1,07	0,40	2,02	0,24	0,10	0,35
	24	0,80	0,32	1,59	0,77	0,17	1,63	0,23	0,09	0,36
Filosofia	25	2,03	1,16	2,64	-1,10	-1,77	-0,58	0,23	0,06	0,41
	26	1,41	0,75	2,36	-0,45	-0,76	-0,12	0,22	0,06	0,37
Biologia	27	0,86	0,67	1,06	0,25	0,03	0,48	0,12	0,04	0,20
	28	1,14	0,93	1,36	0,93	0,72	1,14	0,04	0,01	0,07
	29	0,03	0,00	0,08	0,07	-0,24	0,46	0,23	0,06	0,41
	30	1,30	1,03	1,62	1,17	0,89	1,48	0,04	0,01	0,07
	31	0,72	0,44	1,01	0,97	0,53	1,38	0,18	0,09	0,27
	32	0,97	0,76	1,20	0,64	0,40	0,91	0,1	0,03	0,16
	33	0,20	0,00	0,56	2,17	0,40	3,51	0,35	0,15	0,43
	34	1,74	1,39	2,11	0,28	0,10	0,47	0,07	0,02	0,11
Física	35	0,78	0,31	1,20	1,87	1,18	2,53	0,18	0,14	0,23
	36	0,43	0,00	1,03	2,89	1,24	3,95	0,14	0,08	0,17
	37	1,18	0,65	1,67	2,9	2,19	3,85	0,23	0,19	0,26
	38	1,44	0,89	1,93	2,82	2,16	3,57	0,34	0,30	0,38
	39	0,40	0,00	0,84	2,3	1,43	3,44	0,21	0,17	0,26
	40	1,17	0,69	1,60	1,35	0,80	1,96	0,26	0,20	0,32
	41	0,70	0,51	0,91	1,11	0,87	1,40	0,07	0,03	0,12
	42	0,27	0,00	0,76	2,66	0,91	3,78	0,25	0,12	0,31
Matemática	43	1,78	0,87	2,65	3,09	1,75	4,44	0,18	0,14	0,21
	44	0,35	0,00	0,84	2,38	1,41	3,50	0,15	0,10	0,19
	45	0,53	0,00	1,47	3,49	1,42	5,67	0,17	0,12	0,21
	46	0,91	0,31	1,70	1,94	0,76	3,20	0,2	0,09	0,27
	47	0,92	0,00	1,86	3,29	0,54	5,10	0,3	0,15	0,37
	48	0,90	0,43	1,78	2,02	1,36	3,44	0,08	0,04	0,12
	49	0,02	0,00	0,05	0,4	0,00	0,94	0,26	0,08	0,44
	50	0,41	0,00	1,07	3,29	1,15	4,68	0,29	0,24	0,34
Química	51	1,31	1,05	1,62	1,13	0,92	1,37	0,03	0,01	0,05
	52	1,83	1,17	2,65	1,67	1,00	2,60	0,16	0,11	0,21
	53	0,02	0,00	0,04	0,09	-0,28	0,64	0,26	0,06	0,48
	54	0,81	0,31	1,35	2,37	1,54	3,38	0,17	0,13	0,21
	55	0,25	0,00	0,76	2,81	1,38	3,78	0,25	0,20	0,29
	56	2,33	0,77	3,51	2,87	0,84	4,32	0,26	0,17	0,31
	57	1,25	0,84	1,75	1,03	0,61	1,58	0,1	0,04	0,16
	58	0,86	0,62	1,16	1,44	1,13	1,76	0,05	0,01	0,08

APÊNDICE B - Estimativas a posteriori por ponto e por intervalo dos parâmetros de itens na análise univariada do vestibular 2006-2 da UFLA

Tabela 17 Estimativas a posteriori por ponto (média) e por intervalo (IC-HPD) para os parâmetros de todos os itens de todas as provas do Vestibular 2006 – 2 da UFLA

Provas	Itens	a			b			c		
		média	IC-HPD		média	IC-HPD		média	IC-HPD	
			lower	upper		lower	upper		lower	upper
Português	1	0,74	0,48	1,05	-0,66	-1,45	0,34	0,31	0,00	0,58
	2	0,71	0,44	1,04	0,44	-0,46	1,28	0,31	0,06	0,52
	3	0,98	0,55	1,46	2,24	1,64	2,95	0,35	0,30	0,41
	4	0,83	0,46	1,26	0,81	-0,08	1,56	0,40	0,20	0,54
	5	0,76	0,51	1,07	0,03	-0,46	0,63	0,17	0,00	0,36
	6	1,45	0,94	2,04	2,87	2,64	3,00	0,26	0,23	0,29
	7	0,84	0,44	1,25	1,29	0,62	1,90	0,28	0,14	0,39
	8	1,35	0,83	1,90	2,83	2,54	3,00	0,14	0,11	0,16
	9	0,79	0,42	1,22	1,00	-0,66	2,96	0,65	0,39	0,78
	10	0,87	0,48	1,28	0,89	0,31	1,40	0,25	0,08	0,38
Geografia	11	0,87	0,50	1,31	0,84	0,13	1,48	0,36	0,19	0,49
	12	0,91	0,51	1,38	1,49	0,71	2,41	0,50	0,41	0,59
	13	0,75	0,46	1,10	0,68	-0,03	1,35	0,27	0,07	0,44
	14	0,85	0,50	1,29	1,47	0,82	2,26	0,35	0,24	0,46
	15	0,78	0,44	1,22	0,12	-0,94	1,08	0,42	0,13	0,64
	16	0,88	0,48	1,36	2,52	1,95	3,00	0,29	0,22	0,35
	17	0,83	0,48	1,27	0,64	-0,10	1,27	0,33	0,13	0,49
	18	1,07	0,67	1,51	-0,69	-1,10	-0,19	0,18	0,00	0,40
História	19	0,97	0,57	1,49	-0,36	-1,03	0,55	0,26	0,00	0,55
	20	1,04	0,57	1,58	2,56	1,92	3,00	0,50	0,45	0,55
	21	1,12	0,65	1,66	2,70	2,26	3,00	0,24	0,21	0,28
	22	1,25	0,71	1,79	2,80	2,48	3,00	0,21	0,18	0,24
	23	0,92	0,50	1,39	2,15	1,49	3,00	0,31	0,23	0,38
	24	0,94	0,52	1,41	0,74	0,11	1,35	0,21	0,02	0,36
Filosofia	25	1,15	0,73	1,64	-0,79	-1,18	-0,40	0,16	0,00	0,37
	26	1,13	0,70	1,64	-0,56	-0,87	-0,21	0,12	0,00	0,29
Biologia	27	0,95	0,68	1,27	0,25	-0,03	0,54	0,11	0,00	0,22
	28	1,25	0,96	1,57	0,74	0,60	0,89	0,03	0,00	0,06
	29	1,16	0,68	1,71	2,63	2,13	3,00	0,58	0,55	0,62
	30	1,28	0,96	1,63	0,83	0,70	0,98	0,02	0,00	0,05
	31	0,81	0,45	1,13	1,28	0,92	1,66	0,19	0,09	0,28
	32	0,89	0,64	1,15	0,58	0,36	0,82	0,07	0,00	0,14
	33	1,23	0,73	1,76	2,78	2,44	3,00	0,38	0,35	0,42
	34	1,86	1,44	2,39	0,13	0,00	0,27	0,04	0,00	0,10
Física	35	0,94	0,54	1,40	1,85	1,05	2,80	0,14	0,01	0,23
	36	1,15	0,65	1,67	2,76	2,38	3,00	0,13	0,09	0,16
	37	1,10	0,62	1,64	2,61	2,12	3,00	0,22	0,18	0,26
	38	0,96	0,52	1,43	1,34	0,47	2,39	0,21	0,03	0,36
	39	1,13	0,64	1,69	2,72	2,30	3,00	0,21	0,17	0,25
	40	1,04	0,56	1,57	2,25	1,24	3,00	0,36	0,25	0,44
	41	1,18	0,64	1,75	2,66	2,09	3,00	0,22	0,16	0,26
	42	0,97	0,53	1,50	2,39	1,67	3,00	0,22	0,13	0,29
Matemática	43	1,00	0,55	1,52	1,97	1,18	2,92	0,15	0,03	0,23
	44	1,06	0,61	1,62	2,73	2,28	3,00	0,14	0,10	0,19
	45	1,14	0,66	1,70	2,71	2,28	3,00	0,16	0,12	0,19
	46	1,07	0,58	1,58	2,55	1,97	3,00	0,25	0,19	0,29
	47	0,12	0,12	0,12	2,86	2,61	3,00	0,01	0,00	0,03
	48	1,02	0,57	1,55	2,44	1,82	3,00	0,10	0,03	0,15
	49	1,12	0,59	1,68	2,60	1,94	3,00	0,51	0,46	0,55
	50	1,19	0,64	1,76	2,78	2,40	3,00	0,29	0,25	0,32
Química	51	1,08	0,78	1,40	0,91	0,72	1,09	0,03	0,00	0,08
	52	1,05	0,69	1,48	0,80	0,53	1,06	0,10	0,00	0,18
	53	1,25	0,75	1,81	2,65	2,18	3,00	0,61	0,58	0,64
	54	1,03	0,59	1,51	2,75	2,39	3,00	0,17	0,13	0,21
	55	0,12	0,12	0,12	2,95	2,85	3,00	0,00	0,00	0,01
	56	0,91	0,54	1,32	1,08	0,69	1,44	0,17	0,05	0,27
	57	1,03	0,74	1,39	0,71	0,50	0,95	0,06	0,00	0,13
	58	0,86	0,62	1,14	1,57	1,29	1,88	0,03	0,00	0,06

APÊNDICE C - Gráficos de traços para as habilidades estimadas pelo modelo multivariado

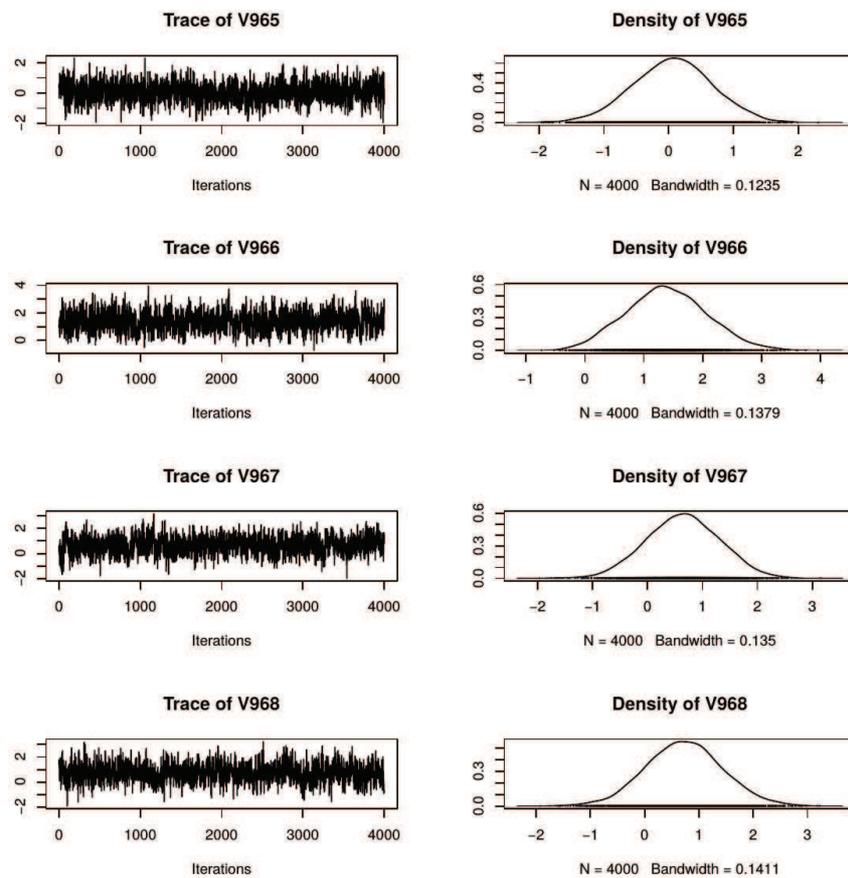


Figura 28 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 2.

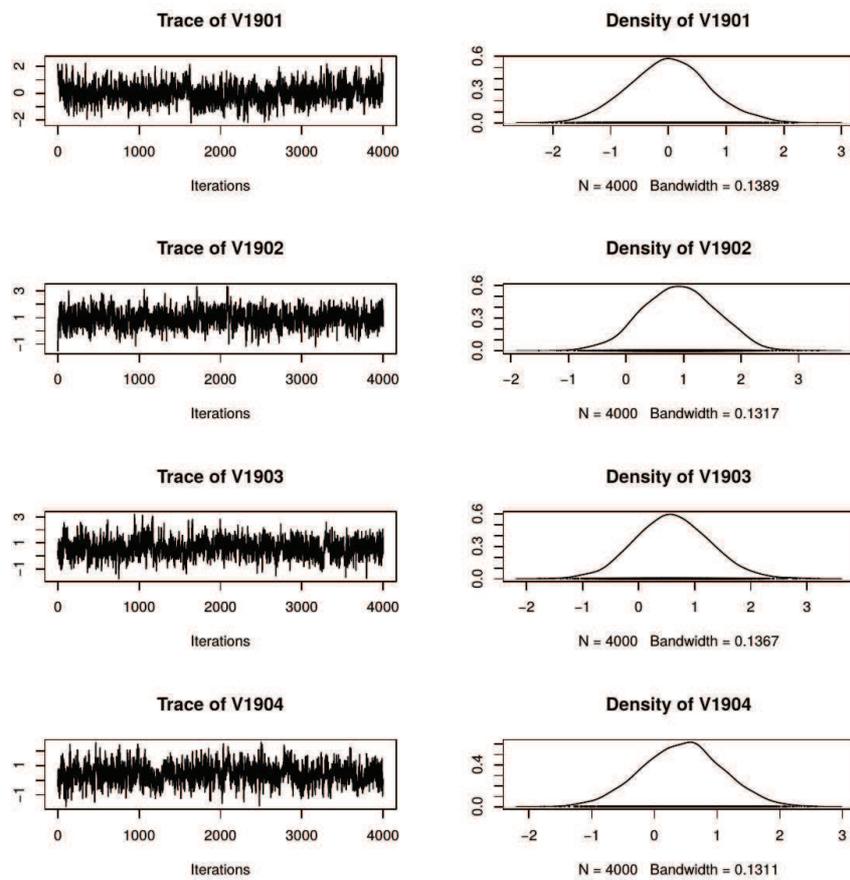


Figura 29 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 3.

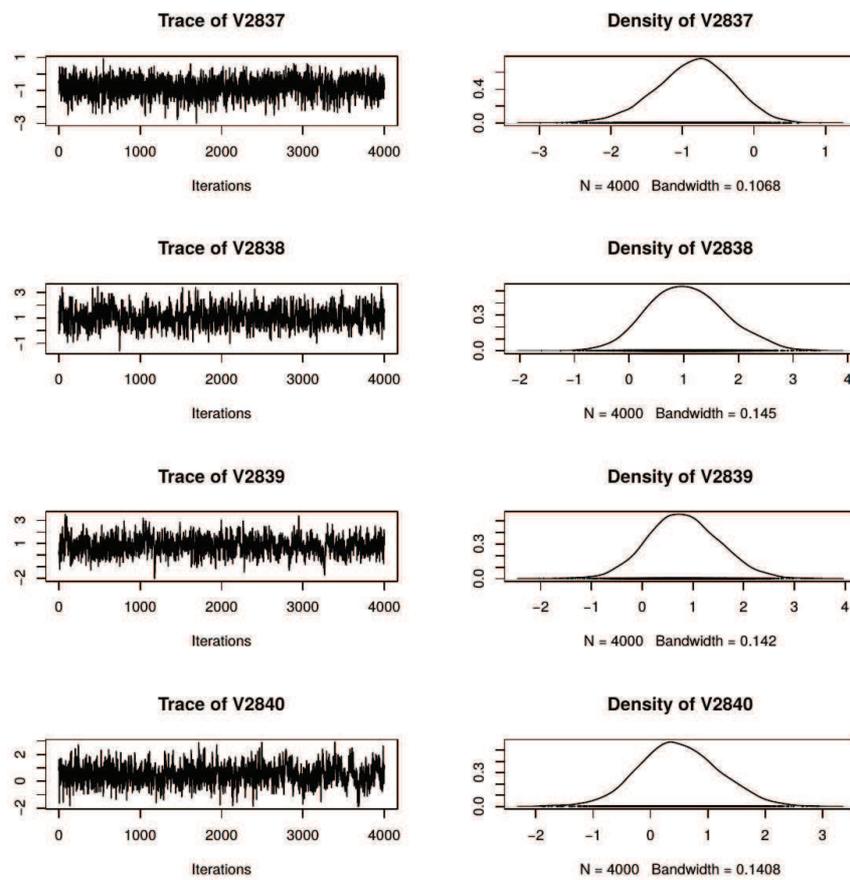


Figura 30 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 4.

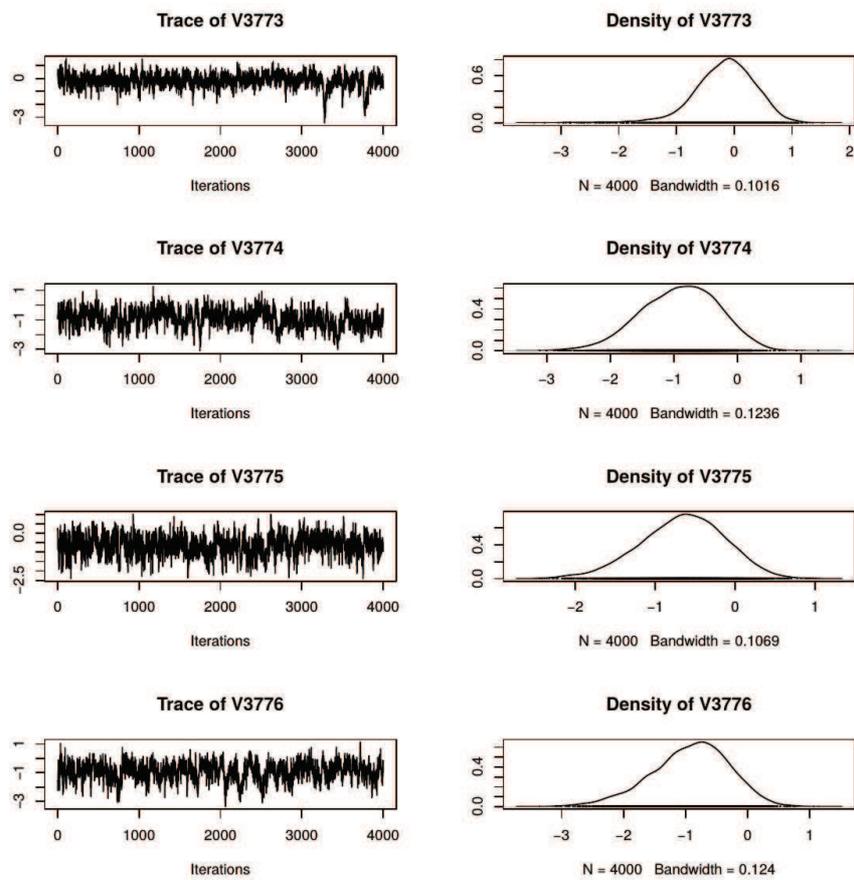


Figura 31 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 5.

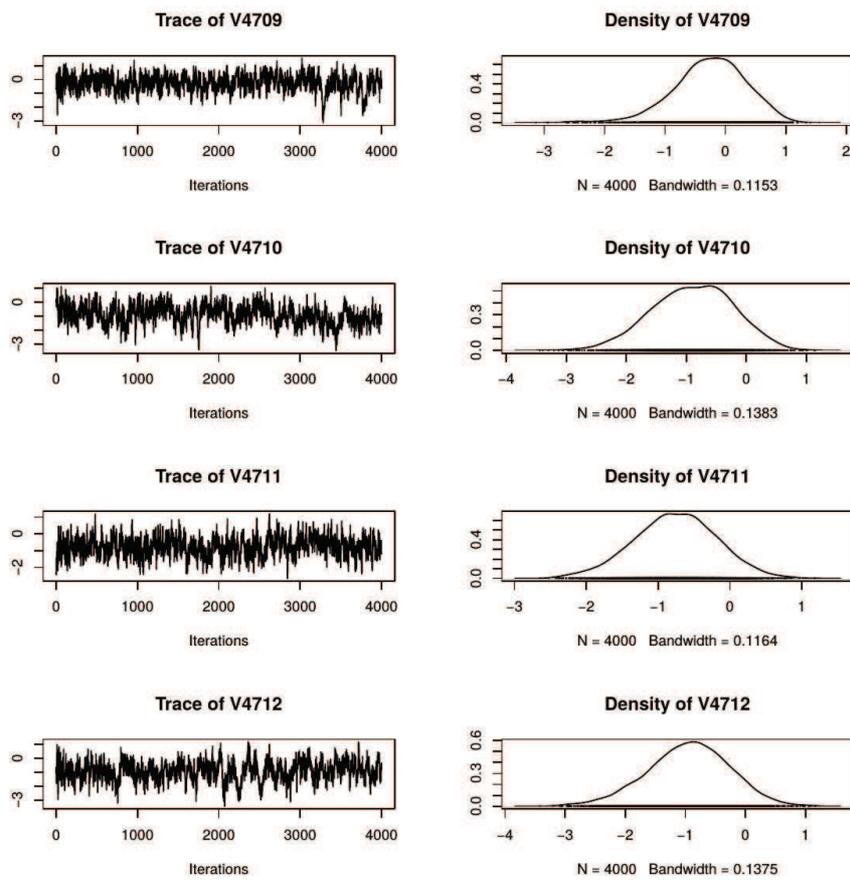


Figura 32 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 6.

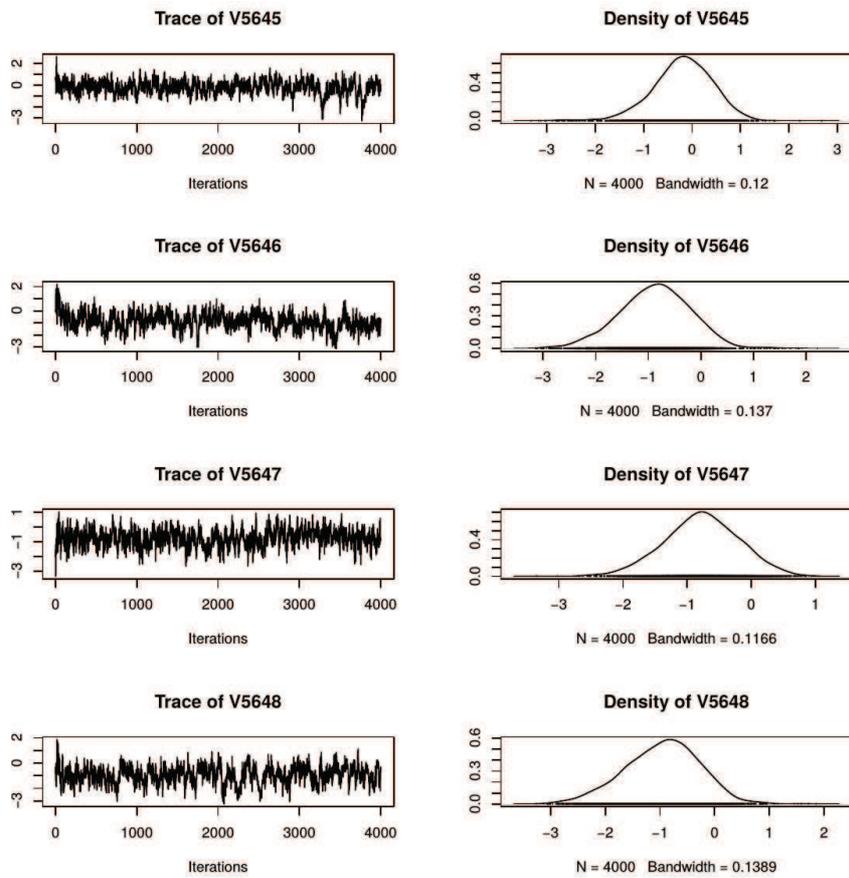


Figura 33 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 7.

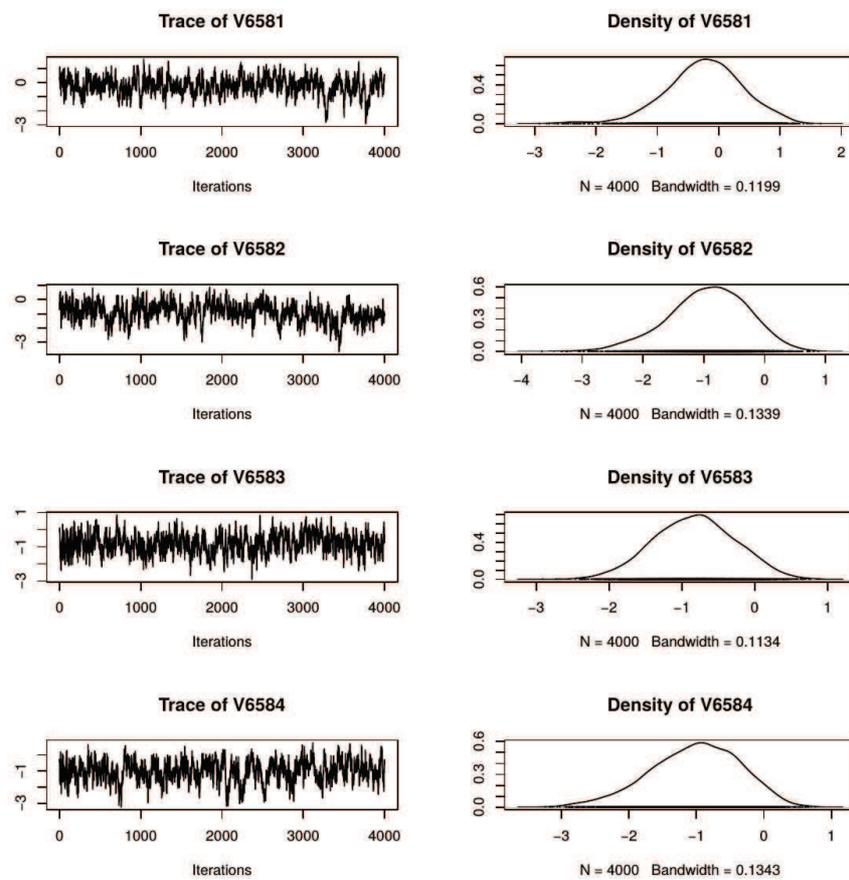


Figura 34 Traços das cadeias e densidades a posteriori obtidas via MCMC para as habilidades dos indivíduos 29,30,31,32 na prova 8.