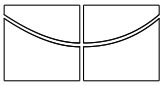


Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Dissertação de Mestrado

Modelo de resposta ao item com controle da heterogeneidade atribuída a fatores conhecidos

por

Rômulo Andrade da Silva

Orientador: Prof. Afrânio Márcio Corrêa Vieira

Rômulo Andrade da Silva

Modelo de resposta ao item com controle da heterogeneidade atribuída a fatores conhecidos

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Universidade de Brasília

Brasília, 2013

Dedicatória

Este trabalho é dedicado a Vicente de Paula da Silva e Lúcia Andrade da Silva.

Agradecimentos

A priori agradeço a Deus por ter me proporcionado saúde, sabedoria e força para perseverar com a minha pesquisa.

Aos meus familiares e amigos que me ajudaram direto e indiretamente, ao meu orientador Prof. Dr. Afrânio Márcio Corrêa Vieira que foi bastante atencioso e prestativo, pela amizade e por ter acreditado no meu potencial.

Em especial, aos meus pais Vicente de Paula da Silva e Lucia Andrade da Silva, aos meus irmãos Vitor Paulo e Karine Maria, a minha noiva Ligia Cipriano Beviláqua e ao Sr. Arnaldo Bevilaqua Vieira e a Sra. Regina Celi Cipriano Bevilaqua, que tanto me motivaram, apoiaram e contribuíram com a realização da dissertação.

Sumário

Resumo

Uma das pressuposições, no processo de estimação dos parâmetros dos modelos tradicionais de resposta ao item, é a independência condicional entre as respostas de diferentes indivíduos. Porém, muitas vezes essa pressuposição é relaxada, por exemplo, quando aplicada em larga escala nas avaliações de sistemas educacionais, o que pode ocasionar variabilidade extra não considerada pelos modelos usuais. A proposta é usar potenciais fontes de heterogeneidade como variáveis explicativas de um efeito aleatório multiplicativo no modelo de Rasch. Esse efeito, consequentemente, acomodará a superdispersão presente nos dados e controlará a pressuposição de independência condicional entre clusters de indivíduos. O modelo foi ajustado aos dados da Prova Brasil 2007, trazendo novas interpretações de grupos. Contudo, a nova abordagem probabilística de considerar informações extras dos indivíduos no momento do ajuste se mostra útil na fase de calibração dos itens.

Palavras Chave: *1. Teoria da resposta ao item 2. Modelo de Rasch 3.Heterogeneidade 4. Superdispersão 5. Prova Brasil.*

Abstract

One of the assumptions in the estimation process of the parameters of the traditional models of item response is conditional independence between the responses of different individuals. Nonetheless, this assumption is often relaxed, for example, when applied to large-scale evaluations of educational systems, which can cause extra variability not considered by usual models. The proposal is to use potential sources of heterogeneity as explanatory variables in a random effect multiplicative Rasch model. This effect, therefore, will accommodate the overdispersion in the data and control the conditional independence assumption between clusters of individuals. The model was adjusted to data from ProvaBrasil 2007 (Brazil Test 2007), bringing new interpretations of groups. However, the new probabilistic approach on considering extra information of individuals at the time of adjustment proves to be useful in the calibration phase of the items.

key words: *1. Item response theory 2. Rasch Model 3. heterogeneity 4. overdispersion 5. ProvaBrasil.*

Introdução

Os estudos de Lord (1952) e Rasch (1960, apud Andrade, Tavares e Valle, 2000) deram o passo inicial para a Teoria da Resposta ao Item - TRI. Esses estudos foram e continuam sendo bastante empregados em Psicometria e em avaliação educacional. Essa teoria propõe modelos paramétricos para traços latentes, isto é, medidas que não são observadas diretamente. Na TRI, basicamente, é apresentada ao sujeito uma série de estímulos, como itens de um teste, e ele responde aos mesmos. Com base na análise das respostas dadas a cada item em específico, infere-se sobre o traço lantente do sujeito (conhecimento, aptidão, sentimento), hipotetizando a existência de uma relação entre as respostas dadas e o nível do traço latente (PASQUALI, 2009). Devido às grandes aplicações e possibilidades de análises comparativas, a TRI despertou o interesse em orgãos de avaliação educacional. Podemos citar o Sistema Nacional de Avaliação da Educação Básica (SAEB), o Projeto PDF/Fundescola e o Exame Nacional de Ensino Médio (ENEM) coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), do MEC. Atualmente, a teoria vem se aperfeiçoando cada vez mais, motivada pela necessidade de melhor representar os fenômenos educacionais.

Sabe-se que uma das pressuposições no processo de estimação dos parâmetros dos modelos tradicionais de resposta ao item é a *independência condicional* entre as respostas de diferentes indivíduos, ou seja, os indivíduos respondem independentemente o teste. Porém, muitas vezes essa suposição é relaxada, por exemplo, quando aplicada em larga escala nas avaliações educacionais (grupos de alunos pertencentes a uma mesma sala sofrem influência das características de uma mesma escola, cidade ou região). Usualmente, nas grandes avaliações, a correlação entre as respostas aos itens, ignorando o conhecimento do aluno, é considerável. Até o momento, há poucas

alternativas para contornar esse problema. Motivado por essa deficiência, a proposta deste trabalho é usar potenciais fatores geradores de heterogeneidade, como, por exemplo: se o aluno possui reprovação ou não, ou se é de origem de escola pública ou privada, que são geralmente desprezadas no momento do ajuste dos modelos da TRI, como variáveis explicativas de um efeito multiplicativo no modelo de Rasch (RASCH, 1960). Entretanto, será utilizado a idéia da família de Modelos Lineares Generalizados Mistos Conjugados de Molenberghs, Verbeke, Demétrio e Vieira (2010). Há algumas vantagens com essa abordagem. Pode-se dar destaque ao fato de lidar com a estrutura naturalmente hierárquica das fontes de variação e de se ter resultados conhecidos dessa família, Silva (2012).

Contudo, este trabalho contribuirá com uma nova proposta de modelos de resposta ao item que estime de forma mais realística a precisão dos parâmetros, levando em conta a heterogeneidade atribuída a fatores conhecidos. Será apresentada, também, a implementação de um algoritmo computacional de estimação dos parâmetros do modelo adaptado por meio da abordagem Bayesiana, bem como a compararação da eficácia do modelo proposto para dados heterogêneos frente ao tradicional, utilizando as bases de dados da Prova Brasil 2007 (SAEB), disponibilizadas gratuitamente pelo Instituto Nacional de Estudos e Pesquisas Educacionais - INEP/MEC (www.inep.gov.br).

Com esse intuito, aborda-se uma revisão metodológica no Capítulo 1 tratando especificamente do modelo de Rasch utilizado na área de avaliação educacional. Define-se as famílias de modelos que servirão de estrutura para o modelo proposto: os Modelos Lineares Generalizados, os Modelos Lineares Generalizados Mistos e os Modelos Lineares Generalizados Mistos Conjugados (MLGMC). Apresenta-se também uma seção no Capítulo 1 sobre noções de inferência bayesiana. Discute-se os aspectos importantes da estimação como o problema das integrais e suas soluções em métodos aproximados e a seleção de modelos. Discute-se também o modelo adaptado de Silva (2012), que trata da heterogeneidade atribuida a fatores desconhecidos.

No Capítulo 2, é tratada a proposta de modelagem probabilística de TRI com controle de heterogenidade atribuído a fatores conhecidos, usando o modelo de Rasch de 1 parâmetro. A idéia é usar fatores conhecidos que causam heterogeneidade por meio de um preditor linear, que se relaciona a um efeito multiplicativo no modelo

de Rasch através da ligação logística. Será focada a utilização da metodologia de inferência bayesiana via Monte Carlo por Cadeias de Markov (MCMC), já que essa metodologia contempla bem a estrutura natural hierárquica dos dados. No Capítulo 3, é feita uma comparação da eficácia do modelo proposto para dados heterogêneos simulados frente ao tradicional e aplica-se o mesmo à base de dados da Prova Brasil de 2007 do Sistema de Avaliação do Ensino Básico (SAEB).

Capítulo 1

Revisão Metodológica

1.1 Teoria da Resposta ao Item - TRI

Esta teoria propõe modelos paramétricos para cálculo de probabilidades assumindo a existência de traços latentes, isto é, medidas que não são observadas diretamente, de grande interesse na psicometria e avaliação educacional. Na TRI, basicamente, é apresentada ao sujeito uma série de estímulos, como itens de um teste ou um questionário, e ele responde aos mesmos. Com base na análise das respostas dadas a cada item específico, infere-se sobre o traço latente atribuído ao sujeito (conhecimento, aptidão, sentimento), hipotetizando uma relação entre as respostas dadas e o nível do traço latente. Existem diversos modelos de TRI que buscam a relação entre uma resposta coerente do indivíduo a um item e seus traços latentes. Na literatura estatística e psicométrica predominam três modelos para respostas dicotômicas, que se diferenciam pelos parâmetros abordados, a saber: modelo que considera apenas a dificuldade do item, que considera a dificuldade e a discriminação do item (capacidade do item de separar sujeitos com magnitudes próximas do mesmo traço) e por fim o modelo que considera a dificuldade, discriminação e a resposta dada ao acaso (ou seja, o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente).

Pode-se dizer que a TRI estende ou até mesmo substitui parte dos conceitos da Teoria Clássica dos Testes - TCT, também chamada de Teoria Clássica da Medida. Essas duas metodologias se destacam em avaliações educacionais e em análises psi-

cométricas (testes cognitivos). Porém, a TRI é uma metodologia mais sofisticada que supera certas limitações da TCT, a saber (PASQUALI, 2009):

- os parâmetros dos itens não dependem da amostra de respondentes;
- os escores dos respondentes independem do teste utilizado;
- o modelo quantifica características do item e não do teste como um todo (TCT), ou seja, a análise não fica associada ao particular conjunto de itens, e sim ao grau de dificuldade do item.
- o modelo oferece medida de precisão para cada nível do traço latente;

Basicamente, a TCT tem por objetivo explicar o resultado final total, considerando o teste como um todo, apresentado o escore total ou o percentual de acerto do indivíduo. Por outro lado, como já comentado, a TRI é mais refinada e se preocupa em modelar a probabilidade de acerto ou de aceitação (em testes de preferência) por trás de cada item. E, assim, é possível realizar estudos em diversas áreas, como por exemplo, na Psicometria, Cúri (2006) estuda as propriedades psicométricas e a validade transcultural do BDI (instrumento para medir a gravidade dos sintomas depressivos). Castro (2008) aplica a teoria na avaliação da intensidade de sintomas depressivos. Na área da Saúde, Mesbah et al. (2002) utiliza a TRI no estudo da qualidade de vida. Na área da Educação, por exemplo, pode-se medir o conhecimento de um aluno em relação a outros, avaliar o desenvolvimento da aprendizagem de alunos de uma determinada série de um ano para outro (mesmo com provas distintas) ou comparar o conhecimento de alunos de regiões distintas ou de escolas diferentes (com pelo menos alguns itens em comum), ver Andrade e Klein (1999). Devido a estas aplicações, a TRI despertou o interesse de grandes orgãos que realizam avaliação educacional: o Sistema de Avaliação da Educação Básica (Saeb), o Projeto PDF/Fundescola, a Prova Brasil e o Exame Nacional de Ensino Médio (Enem) coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), do MEC. Neste trabalho, a TRI será usada na avaliação da Educação usando como base o modelo de Rasch (1960). Andrade et al. (2000) revisaram muito bem a teoria nesse campo como um conjunto de modelos matemáticos que procuram representar a probabilidade de

um indivíduo responder corretamente um item como função dos parâmetros do item e do traço latente do respondente.

1.1.1 Definição do modelo de Rasch para respostas dicotômicas

Considere que n indivíduos respondam I itens (verdadeiro ou falso) de um teste. Seja U_{ij} uma variável aleatória dicotômica tomando valor 1 quando o respondente j acerta o item i e 0 quando o respondente j erra o item i, $j = 1, 2, \ldots, n$ e $i = 1, 2, \ldots, I$. A probabilidade do respondente j acertar o item i ($U_{ij} = 1$) é dada por:

$$I\!P(U_{ij} = 1|\theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}}, \tag{1.1}$$

em que θ_j corresponde ao traço latente (conhecimento ou habilidade) do respondente j e b_i ao parâmetro de dificuldade do item i, tomado na mesma escala da habilidade. Note que a probabilidade é dada pela função logística de $\theta_j - b_i$. Tradicionalmente, a curva formada por essa probabilidade em função de θ_j é chamada de Curva Característica do Item - CCI. Existe uma versão desse modelo que introduz uma constante multiplicativa $D = 1,7$ em $(\theta_j - b_i)$ para tornar a curva logística aproximada à curva da distribuição normal acumulada. A probabilidade de acerto do item cresce não linearmente, de acordo com o traço latente θ (conhecimento ou habilidade), na forma de uma sigmóide (S). Note que, se $\theta_j = b_i$, a probabilidade de acerto é de 50%. Visto de outra forma pela Curva Característica do Item, o parâmetro de dificuldade b_i do item corresponde ao ponto na escala do traço latente θ onde a probabilidade de acerto é 0,5, de forma que se $\theta_j > b_i$ a probabilidade de acerto é maior que 50% e vice-versa.

Convém destacar que o modelo de Rasch é um modelo unidimensional da TRI . Ou seja, o processo de estimação do modelo pressupõe que os itens estão medindo um único traço latente. Assim, a habilidade θ do indivíduo que se busca estimar deve ser o único fator responsável pelas respostas dadas. Usualmente, essa pressuposição é automaticamente assumida, reforçada pela idéia de que exista uma habilidade dominante. Porém, modelos desse tipo sofrem uma limitação considerável que, dependendo do teste, não trata da influência, por exemplo, da língua portuguesa na probabilidade de acerto de um item em um teste de matemática, uma vez que o respondente pode errar por não ter suficiente domínio de estrutura da língua e não compreender corretamente

a questão envolvida no item. Apesar de existirem modelos da TRI multidimensionais que tratam desta abordagem, estes não serão tratados aqui, pois está fora do escopo desta dissertação. Outros modelos de TRI podem ser encontrados em Bock e Zimowski (1997) e em Reckase (2009).

1.1.2 Pressuposições do modelo

Pode-se dizer que na TRI três pressuposições são necessárias para o processo de estimação. A primeira é a independência condicional entre as respostas fornecidas por respondentes distintos. A segunda é que as respostas fornecidas aos itens pelo mesmo respondente são independentes, dada sua habilidade (independência local). Ou seja, o indivíduo não adquire conhecimento no decorrer da prova, pois as respostas são independentes entre si. A terceira pressuposição, como já citada anteriormente, diz respeito aos testes unidimensionais (geralmente assumida como verdadeira). São os testes em que apenas uma aptidão ou traço latente é responsável pela realização de seus itens. Em outras palavras, é esse fator que se supõe estar sendo medido pelo teste. Assim, para os testes unidimensionais (o que será tratado aqui) supõe-se que apenas uma habilidade está envolvida na realização do teste, isto é, há uma habilidade dominante.

1.1.3 Estimação por Máxima Verossimilhança Marginal

O processo de estimação por Máxima Verossimilhança Marginal proposto inicialmente por Bock e Lieberman (1970) apud Andrade et al. (2000) é mais utilizado na abordagem frequentista. Isso se deve ao fato de contornar possíveis problemas de identificabilidade na estimação dos parâmetros dos itens, no caso da estimação conjunta de θ_j e b_i, encontrados por Andersen (1973). Um aperfeiçoamento desse método, dado por Bock e Aitkin (1981), exigiu menos esforço computacional. Eles reformularam as equações dos parâmetros dos itens e adaptaram o algoritmo EM (proposto por

Dempster, Laird e Rubim, 1977), preservando ainda as propriedades assintóticas. A idéia desse método é dividir o processo de estimação em duas etapas: na primeira, os parâmetros dos itens são estimados através da marginalização da função de verossimilhança conjunta, de modo que sejam independentes do efeito aleatório θ; na segunda etapa, estima-se as habilidades, considerando os parâmetros dos itens conhecidos e iguais aos obtidos na primeira etapa. Como os parâmetros das habilidades são desconhecidos, Andersen (1980) sugeriu considerar a existência de uma distribuição de probabilidade latente (denotada na literatura por Π) associada às habilidades, e que os dados obtidos com a realização dos testes correspondem a uma amostra dessa população. Dessa forma, a estimação parte do princípio de marginalizar a verossimilhança integrando-a com relação à distribuição das habilidades, eliminando, assim, esses parâmetros. Considere $g(\theta|\boldsymbol{\eta})$ uma função densidade de probabilidade associada ao traço latente (conhecimento), que seja duplamente diferenciável. As componentes de $\boldsymbol{\eta}$ são os parâmetros associados a Π, conhecidos e finitos. Seja $I\!P(U_{.j} = u_{.j}|b_i, \theta_j, \boldsymbol{\eta})$ a função de probabilidade condicional da variável aleatória dicotômica representando o acerto do indivíduo j dado o parâmetro de dificuldade do item, traço latente do indivíduo j e o conjunto de parâmetros associados à distribuição de probabilidade do traço latente. Com isso, a probabilidade marginal de $U_{.j}$ é dada por:

$$\begin{aligned} I\!P(U_{.j} = u_{.j}|b_i, \boldsymbol{\eta}) &= \int_{\mathbf{R}} I\!P(U_{.j} = u_{.j}|b_i, \theta_j, \boldsymbol{\eta}) g(\theta|\boldsymbol{\eta}) d\theta \\ &= \int_{\mathbf{R}} I\!P(U_{.j} = u_{.j}|b_i, \theta_j) g(\theta|\boldsymbol{\eta}) d\theta, \end{aligned} \tag{1.2}$$

pois a distribuição de $U_{.j}$ não é função de $\boldsymbol{\eta}$. Pela suposição de que as respostas são independentes entre diferentes respondentes, tem-se a verossimilhança:

$$L(b_i, \boldsymbol{\eta}) = \prod_{j=1}^{n} I\!P(U_{.j}|b_i, \boldsymbol{\eta}),$$

e a log-verossimilhança:

$$l(b_i, \boldsymbol{\eta}) \;=\; \sum_{j=1}^{n} ln I\!P(U_{.j}|b_i, \boldsymbol{\eta}) \tag{1.3}$$

Derivando a log-verossimilhança em relação a b_i, obtem-se a seguinte equação de estimação para b_i:

$$\sum_{l=1}^{q}[(\overline{r}_{il} - \overline{f}_{il}P_{il})W_{il}]g_j^*(\theta) = 0, \tag{1.4}$$

em que: $\overline{r}_{il} = \sum_{j=1}^{n} u_{ij}g^*(\overline{\theta}_l)$, $\overline{f}_{il} = \sum_{j}^{n} g^*(\overline{\theta}_l), W_i = \frac{P_i^* Q_i^*}{P_i Q_i}, P_i = \{1+e^{(-\theta-b_i)}\}^{-1}, g_j^*(\theta) = \frac{I\!P(U_{.j}|b_i,\theta)g(\theta,\boldsymbol{\eta})}{\int I\!P(U_{.j}|b_i,\theta)g(\theta,\boldsymbol{\eta})d\theta}$.

Como a equação não possui solução analítica explícita, deve-se usar métodos iterativos, em particular o algoritmo EM, ver Azevedo (2003) ou Andrade et al. (2000).

1.2 Modelos Lineares Generalizados - MLG

Os modelos lineares generalizados (MLG) são uma extensão dos modelos lineares clássicos, proposta por Nelder e Wedderburn (1972). Nesses modelos, assume-se que a variável resposta possa ter qualquer distribuição da família exponencial na forma canônica com a introdução de um parâmetro de perturbação. Uma das vantagens desses modelos é escolher uma relação funcional mais apropriada entre a média da variável resposta e o preditor linear, além de não se restringir ao erro seguindo uma distribuição normal, no caso do modelo clássico de regressão. Felizmente, essa metodologia já está bastante difundida na literatura, como em Cordeiro (1986), McCullagh e Nelder (1989), Collet (1991), Demétrio (2002), Dobson (2001) e Paula (2004), e serviu de base para o desenvolvimento dos Modelos Lineares Generalizados Mistos.

1.2.1 Especificação do MLG

Sejam $Y_1,\ Y_2, \ldots, Y_n$ variáveis aleatórias independentes com média $\mu_1,\ \mu_2, \ldots,\ \mu_n$ de uma distribuição da **família exponencial** na forma canônica com a introdução de um parâmetro de pertubação $a_i(\phi) > 0$, ou seja:

$$f(y_i, \theta_i, \phi) = \exp\{a_i(\phi)^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\}, \tag{1.5}$$

em que $b(.)$ e $c(.)$ são funções conhecidas. Têm-se também que $E(Y_i) = b^{'}(\theta_i) = \mu_i$, $\text{Var}(Y_i) = a(\phi)b^{''}(\theta_i) = a_i(\phi)V_i$, com $V = \partial\mu/\partial\theta$ denominado de *função de variância* dependendo unicamente de μ. Esse conjunto de variáveis aleatórias é definido como **componente aleatória**, que corresponderão à amostra da variável resposta. Seja a **componente sistemática**:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}. \tag{1.6}$$

Tem-se que $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$ é o preditor linear, X representa a matriz que contém os fatores e covariáveis do modelo e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $p < n$, o vetor de parâmetros desconhecidos a serem estimados.

Então os MLG's são definidos pela especificação da componente aleatória (1.5) e pela componente sistemática (1.6) sendo *relacionados* por uma **função de ligação**, $\eta_i = g(\mu_i)$, sendo $g(.)$ uma função monótona e diferenciável. Note que na definição não existe um erro aleatório ε adicionado à componente sistemática do modelo, como no caso de modelos de regressão linear, uma vez que o componente aleatório é implicitamente definido pela especificação da distribuição pertencente à família exponencial.

A escolha dessa distribuição vai depender da natureza dos dados, quase sempre levado em conta o suporte da função. Por exemplo: para dados contínuos, as mais usadas são normal, normal inversa e gama (dados com assimetria à direita); para dados de contagens, têm-se a binomial negativa, a Poisson e outras pertencentes à classe de distribuição em série de potências modificada (DSPG) (BRITO; CORDEIRO; DEMÉTRIO, 2010); e para proporções, utiliza-se a binomial ou beta-binomial (ensaios do tipo dose-resposta). A matriz do modelo vai definir matematicamente o desenho experimental ou estudo observacional.

1.2.2 Estimação do vetor de parâmetros $\boldsymbol{\beta}$

Existem vários métodos de se estimar o vetor de parâmetros $\boldsymbol{\beta}$. Aqui utiliza-se o método da máxima verossimilhança. Assim, para um MLG, o valor estimado do

vetor $\boldsymbol{\beta}$ é o que maximiza o logaritmo da função de verossimilhança:

$$l = l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l(\theta_i) = \sum_{i=1}^{n} \Big\{ a(\phi)^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi) \Big\}. \tag{1.7}$$

Precisa-se obter a função escore, pois os valores dos β's que maximizam a função $l(\boldsymbol{\theta})$ são dados pela solução do sistema $\boldsymbol{U_\beta} = \frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{0}$. Dessa forma, dado que $\mu_i = b^{'}(\theta)$ e $\frac{\partial \mu_i}{\partial \theta_i} = V_i$, tem-se que a função escore é:

$$U(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = a(\phi)^{-1}\boldsymbol{X}^T\boldsymbol{W}^{1/2}\boldsymbol{V}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}),$$

em que a matriz W (matriz de pesos) e V são definidas por:

$$\boldsymbol{W} = \text{diag}[w_1, \ldots, w_n],\ w_i = \Big(\frac{\partial \mu_i}{\partial \eta_i}\Big)^2 \frac{1}{V_i}$$

$$\boldsymbol{V} = \text{diag}[v_1, \ldots, v_n],\ v_i = \frac{\partial \mu_i}{\partial \theta_i}.$$

Para um certo elemento típico, U_r, tem-se:

$$U_r = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_r} = \sum_{i=1}^{n} \frac{dl_i}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \beta_r}, \tag{1.8}$$

pois pela regra da cadeia, tem-se que l é função de θ_i ($l(\theta) = f(\theta_1, \ldots, \theta_i, \ldots, \theta_n)$), que é função de μ_i ($\theta_i = \int V_i^{-1} d\mu_i = q(\mu_i)$), que por sua vez é função de η_i ($\mu_i = g^{-1}(\eta_i) = h(\eta_i)$), e finalmente η_i é função de β ($\eta_i = \sum_{r=1}^{p} x_{ir}\beta_r$). Assim, a expressão (18) fica dada por:

$$U_r = \frac{1}{a(\phi)} \sum_{i=1}^{n} [y_i - b^{'}(\theta_i)] \frac{1}{\frac{d\mu_i}{d\theta_i}} \frac{d\mu_i}{d\eta_i} x_{ir}. \tag{1.9}$$

Fazendo a substituição de $b^{'}(\theta)$ por $\mu_i =$ e $\frac{d\mu_i}{d\theta_i}$ por V_i em (1.9), tem-se:

$$U_r = \frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ir}.$$

Aplica-se o processo iterativo de Newton-Raphson para a obtenção da solução do sistema $\boldsymbol{U(\beta)} = \frac{dl}{d\boldsymbol{\beta}} = \mathbf{0}$, já que geralmente o sistema não apresenta solução analítica, por se tratar de um sistema de equações não lineares. Este processo pode

ser reescrito como um processo iterativo de mínimos quadrados dado por:

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{W}^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(m)} \boldsymbol{z}^{(m)},\ m = 0, 1, \dots\ , \tag{1.10}$$

em que, $\boldsymbol{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \boldsymbol{W}^{-1/2(m)} \boldsymbol{V}^{-1/2(m)} (\boldsymbol{y} - \boldsymbol{\mu})^{(m)}$. O processo acima pode ser utilizado para qualquer MLG. Para mais detalhes sobre a estimação, veja Demétrio (2002) e Paula (2004).

1.3 Modelos Lineares Generalizados Mistos - MLGM

Breslow e Clayton (1993) propuseram a classe dos Modelos Lineares Generalizados Mistos - MLGM (*Generalized Linear Mixed Models*). Os autores tomaram como base os Modelos Linerares Generalizados e adicionaram ao preditor linar ($\mathbf{X}\boldsymbol{\beta}$) um vetor de efeitos aleatórios $\boldsymbol{u}$ (também chamado de efeito latente), com o objetivo de captar variações não consideradas no modelo e que podem influenciar nos resultados. No MLG, modela-se uma função do vetor de médias ($\boldsymbol{\eta}$), já nessa classe, modela-se uma função do vetor de médias condicionais $\boldsymbol{\mu} = I\!E(\mathbf{Y}|\boldsymbol{u})$. Com isso, consegue-se analisar dados com diferentes estruturas de dependência, como acomodar a informação serial contida nas medidas longitudinais não balanceadas ou controlar a superdispersão em dados não longitudinais. Piepho (1999), por exemplo, pôde controlar a superdispersão existente em dados acrescentando efeitos aleatórios ao modelo.

1.3.1 Especificação do MLGM

Seja uma distribuição condicional de $\mathbf{Y}|\boldsymbol{u}$ seguindo alguma distribuição da família exponencial. O vetor $\mathbf{Y}$ representa a variável resposta do modelo e os seus elementos são condicionalmente independentes dado $\boldsymbol{u}$, isto é:

$$\begin{aligned} Y_i|\boldsymbol{u} &\overset{\text{i.d.}}{\sim} f_{Y_i|\boldsymbol{u}}(y_i|\boldsymbol{u}) \\ f_{Y_i|\boldsymbol{u}}(y_i|\boldsymbol{u}) &= \exp\Big\{\frac{w_i}{\phi}[y_i\theta_i - b(\theta_i)] + c(y_i;\phi)\Big\}. \end{aligned} \tag{1.11}$$

Tem-se que a média condicional de y_i dado $\boldsymbol{u}$ é μ_i ($I\!E[Y_i|\boldsymbol{u}] = \mu_i = \frac{db(\theta_i)}{d\theta_i}$). Esta média relaciona-se com o preditor linear através de uma função de ligação, monótona, diferenciável e pré-definida (como nos MLG's):

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + z_i^\top \boldsymbol{u}, \tag{1.12}$$

com $\mathbf{x}_i^\top$ representado a i-ésima linha da matriz do modelo para efeitos fixos, $\boldsymbol{\beta}$ é o vetor de parâmetros dos efeitos fixos, $z_i^\top$ é a i-ésima linha da matriz do modelo para os efeitos aleatórios e $\boldsymbol{u}$ o vetor de efeitos aleatórios. Define-se, ainda, uma distribuição para os efeitos aleatórios:

$$\boldsymbol{u} \;\sim\; f_{\boldsymbol{u}}(\boldsymbol{U}). \tag{1.13}$$

Geralmente, adota-se $\boldsymbol{u} \sim I\!N_q(\mathbf{0}, \mathbf{D})$. Lee e Nelder (1996) apresentam estudos sobre efeitos aleatórios não-normais usando distribuições conjugadas.

Em síntese, alguns resultados importantes dos MLGM's:

$$\begin{aligned} E(Y_i) &= E[E(Y_i|\boldsymbol{u})] = E[\mu_i] \\ &= E[g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i^\top \boldsymbol{u})] \end{aligned}$$

$$\begin{aligned} \mathrm{Var}[Y_i] &= \mathrm{Var}[E(Y_i|\boldsymbol{u})] + E[\mathrm{Var}(Y_i|\boldsymbol{u})] \\ &= \mathrm{Var}[g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i^\top \boldsymbol{u})] + E\Big[\frac{w_i}{\phi}\mathrm{Var}\{g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i^\top \boldsymbol{u})\}\Big] \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}[Y_i, Y_j] &= \mathrm{Cov}[E(Y_i|\boldsymbol{u}), E(Y_j|\boldsymbol{u})] + E[\mathrm{Cov}(Y_i, Y_j|\boldsymbol{u})] \\ &= \mathrm{Cov}[\mu_i, \mu_j] + E(0) \\ &= \mathrm{Cov}[g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i^\top \boldsymbol{u}), g^{-1}(\mathbf{x}_j^\top \boldsymbol{\beta} + z_j^\top \boldsymbol{u})] \end{aligned}$$

1.3.2 Estimação

A seguir, aborda-se o método de Máxima Verossimilhança, sugerido por Schall (1991), para a estimação dos parâmetros do modelo (11). Geralmente será necessário aplicar algum método iterativo, como o algoritmo EM (ver Dempster, Laird e Rubin (1977)). Outros métodos de estimação podem ser encontrados em Breslow e Clayton (1993), McGilchrist (1994) ou em Lee e Nelder (1996). O processo de estimação aqui descrito servirá apenas a título de revisão, pois neste trabalho utilizar-se-á a abordagem bayesiana para a estimação dos parâmetros. Denotando por γ os parâmetros dos efeitos aleatórios da distribuição $f_{\boldsymbol{U}}(\boldsymbol{u})$, a função de log-verossimilhança é dada por:

$$l(\boldsymbol{\beta}, \gamma) = \log \int f_{\mathbf{Y}|\boldsymbol{u}}(\mathbf{y}|\boldsymbol{u}) f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u} = \log f_{\mathbf{Y}}(y),$$

a qual, para os parâmetros fixos $\boldsymbol{\beta}$, é maximizada de acordo com a relação abaixo:

$$\mathbf{X}^{\top} I\!E[\boldsymbol{W}^{*}|\mathbf{y}] \;=\; \mathbf{X}^{\top} I\!E[\boldsymbol{W}^{*}\boldsymbol{\mu}|\mathbf{y}], \tag{1.14}$$

em que $W^{*}\text{diag}\Big[a(\phi)V(\mu_i)\frac{\partial \eta}{\partial \mu}\Big]^{-1}$. Com respeito ao parâmetro γ da função $f_{\boldsymbol{U}}(\boldsymbol{u})$, tem-se a seguinte equação de máxima verossimilhança:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \gamma)}{\partial \gamma} &= \int \frac{\partial \log f_{\boldsymbol{U}}(\boldsymbol{u})}{\partial \gamma} f_{\boldsymbol{U}|\mathbf{y}}(\boldsymbol{u}|\mathbf{y}) d\boldsymbol{u} \\ &= I\!E\Big[\frac{\partial \log f_{\boldsymbol{U}}(\boldsymbol{u})}{\partial \gamma}\Big|\mathbf{y}\Big]. \end{aligned}$$

McCulloch e Searle (2001) apresentam diversas técnicas computacionais para encontrar os estimadores de máxima verossimilhança dos MGLM's como, por exemplo, quadratura numérica, algoritmo EM, Markov Chain Monte Carlo (MCMC) e algoritmo de aproximação estocástica. A seguir serão mostrados os passos do algoritmo EM para calcular as estimativas.

Etapas do algoritmo:

1. Passo inicial (m=0): adotam-se valores iniciais para $\boldsymbol{\beta}^{(0)}$, $\theta^{(0)}$ e $\mathbf{D}^{(0)}$;

2. Calculam-se:

   - $\boldsymbol{\beta}^{(m+1)}$ e $\theta^{(m+1)}$ para maximizar $I\!E(\log f_{\mathbf{y}|\boldsymbol{U}}(\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}))$;
   - $\mathbf{D}^{(m+1)}$ para maximizar $I\!E(f_{\boldsymbol{U}}(\boldsymbol{u}|\mathbf{D})|\mathbf{y})$;
   - Faça $m = m + 1$.

3. Se a convergência ocorre, os valores obtidos são as estimativas de máxima verrossimilhança, caso contrário, retorna-se ao passo 2.

1.4 Modelos Lineares Generalizados Mistos Conjugados - MLGMC

Os Modelos Lineares Generalizados Mistos Conjugados, proposto por Molenberghs, Verbeke, Demétrio e Vieira (2010), são uma das mais recentes e flexíveis famílias de modelos combinados. O modelo tem como base a estrutura dos MLGM's, e como já discutido, são modelos que se adaptam bem às medidas repetidas não-normais. Sabe-se que dados binários, de proporção ou de contagens com medidas repetidas podem apresentar superdispersão (conhecido também como "sobredispersão" ou variação "extra-Bernoulli/binomial/Poisson"). O fenômeno da superdispersão ocorre quando a variabilidade dos dados não é adequadamente descrita pelo modelo. Isto é causado, dentre outros motivos, pela má especificação do preditor linear, pela variabilidade do material experimental, gerando um componente aleatório que não é considerado no modelo, por variáveis não-observáveis omitidas, *outliers* ou pela função de ligação inadequada. Contudo, a não consideração da superdispersão presente nos dados pelo modelo gera subestimação dos erros padrões, levando a uma inflação do erro tipo I. Os MLGM são utilizados também para modelagem da superdispersão quando os dados não são medidos longitudinalmente, dando a esta classe uma grande flexibilidade (BRESLOW e CLAYTON, 1993). Porém, quando ocorrem medidas repetidas e superdispersão (ou também efeitos de fragilidade, no caso de tempos de sobrevivência) simultaneamente, os MLGMC demonstam ser uma alternativa de interesse (MOLENBERGHS et al., 2010). Desse modo, a família de modelos proposto pelos autores combina tanto a estrutura dos modelos para superdispersão

(incluindo uma variável de efeito aleatório com distribuição conjugada à variável resposta) quanto o efeito aleatório normal, separadamente, na classe dos MLG's, para modelar a estrutura de dependência temporal das medidas repetidas. A seguir, define-se modelos dessa família para dados binários.

1.4.1 Especificação do Modelo Combinado Logit-Bernoulli-Normal-Beta

Seja Y_{ij} uma variável aleatória representando a ocorrência ($Y_{ij} = 1$) ou não ($Y_{ij} = 0$) de um evento do i-ésimo elemento no j-ésimo período, sendo $i = 1, \ldots, N$ e $j = 1, \ldots, n_i$. Denotando por γ_{ij} o parâmetro aleatório para a superdispersão (variação extra-Bernoulli) na $(ij)-$ésima observação, $\boldsymbol{u}_{i_{q\times 1}}$ o vetor aleatório q dimensional para a correlação serial, associado ao vetor $\boldsymbol{z}^{\top}_{ij_{1\times q}}$ com q fatores e/ou covariáveis; $\boldsymbol{\beta}_{p\times 1}$ o vetor de parâmetros fixos de dimensão p associado ao vetor $\mathbf{x}^{\top}_{ij_{1\times p}}$ com p fatores e/ou covariáveis; $\boldsymbol{D}$ a matriz de variâncias e covariâncias, com parâmetros fixos e desconhecidos e $I\!P_{ij}$ a probabilidade de sucesso para a (ij)-ésima observação.

$$
\begin{aligned}
Y_{ij}|\boldsymbol{u}_i &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \gamma_{ij}\frac{\exp\{x^{\top}_{ij}\boldsymbol{\beta} + z^{\top}_{ij}\boldsymbol{u}_i\}}{1+\exp\{x^{\top}_{ij}\boldsymbol{\beta} + z^{\top}_{ij}\boldsymbol{u}_i\}} \\
\gamma_{ij} &\sim \text{Beta}(\alpha, \kappa) \\
\boldsymbol{u}_i &\sim I\!N_q(\mathbf{0}, \boldsymbol{D})
\end{aligned}
\tag{1.15}
$$

Note que o parâmetro de superdispersão entra na forma de um efeito multiplicativo no preditor linear e tem distribuição Beta com parâmetros fixos e desconhecidos. O modelo assume ainda que os parâmetros aleatórios γ_{ij} e $\boldsymbol{u}_i$ são independentes. Assim, pode-se escrever o seguinte resultado:

$$
f(y_{ij}, \gamma_{ij}, \boldsymbol{u}_i) = f(y_{ij}|\gamma_{ij}, \boldsymbol{u}_i) f(\boldsymbol{u}_i) f(\gamma_{ij}), \tag{1.16}
$$

ou ainda,

$$f(\mathbf{y}_i, \boldsymbol{\gamma}_i, \boldsymbol{u}_i) = \prod_{j=1}^{n_i} f(y_{ij}|\gamma_{ij}, \boldsymbol{u}_i) f(\boldsymbol{u}_i) f(\gamma_{ij}), \tag{1.17}$$

com a função de probabilidade condicionada de $Y_{ij}|\gamma_{ij}, \boldsymbol{u}_i$ dada por:

$$f(y_{ij}|\gamma_{ij}, \boldsymbol{u}_i) = \left[\gamma_{ij} \frac{\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}{1 + \exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}\right]^{y_{ij}} \left[1 - \gamma_{ij} \frac{\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}{1 + \exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}\right]^{1-y_{ij}}.$$

Uma outra versão desse modelo para dados binários é utilizar a função de ligação probito em (1.15). Assim, a mudança ocorre na probabilidade de sucesso que passa a ser definida por $\pi_{ij} = \gamma_{ij}\Phi(x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i)$. Para mais detalhes sobre esse modelo e o processo de estimação, ver Vieira (2008) e Molenberghs et al. (2010).

1.4.2 Estimação por Máxima Verossimilhança

Para obter os estimadores de máxima verossimilhança de β e $\boldsymbol{D}$, será necessário encontrar a distribuição marginal de $\mathbf{Y}_i$. Desse modo, tomando a expressão (1.17) e integrando com respeito a $\boldsymbol{u}_i$, obtém-se

$$f(\mathbf{y}_i, \gamma_i) = \int \prod_{j=1}^{n_i} f(y_{ij}|\gamma_{ij}, \boldsymbol{u}_i) f(\boldsymbol{u}_i) f(\gamma_{ij}) d\boldsymbol{u}_i. \tag{1.18}$$

Pela especificação do modelo em (1.15), assume-se que as funções de densidade de $\boldsymbol{u}_i$ e γ_{ij} sejam dadas por:

$$f(\boldsymbol{u}_i) = \frac{1}{\sqrt{(2\pi)^{n_i}|\boldsymbol{D}|}} \exp\left[-\frac{1}{2}\boldsymbol{u}_i^\top \boldsymbol{D}^{-1} \boldsymbol{u}_i\right], \quad \boldsymbol{u}_i \in \mathbf{R}^q, \tag{1.19}$$

$$f(\gamma_{ij}) = \frac{\gamma_{ij}^{\alpha-1}(1-\gamma_{ij})^{\kappa-1}}{B(\alpha, \kappa)}, \quad \gamma_{ij} \in [0, 1]. \tag{1.20}$$

Pelas expressões (1.18), (1.19) e (1.20), tem-se a função de verossimilhança ainda

condicionada a $\boldsymbol{\gamma}$:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{D}, \alpha, \kappa|\boldsymbol{\gamma}) &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \left[\gamma_{ij} \frac{\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}{1+\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}} \right]^{y_{ij}} \\
&\times \left[1 - \gamma_{ij} \frac{\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}{1+\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}} \right]^{1-y_{ij}} \\
&\times \frac{1}{\sqrt{(2\pi)^{n_i}|\boldsymbol{D}|}} \exp\left[-\frac{1}{2} \boldsymbol{u}_i^\top \boldsymbol{D}^{-1} \boldsymbol{u}_i \right] \frac{\gamma_{ij}^{\alpha-1}(1-\gamma_{ij})^{\kappa-1}}{B(\alpha,\kappa)} d\boldsymbol{u}_i. \qquad (1.21)
\end{aligned}
$$

No entanto, uma alternativa para retirar a influência desse efeito aleatório é integrar $f(y_{ij}|\gamma_{ij}, \boldsymbol{u}_i)$ analiticamente sobre o domínio de γ_{ij}. Utilizando-se do artifício de que $f(y_{ij}|\gamma_{ij}, \boldsymbol{u}_i) = f(y_{ij}=0|\gamma_{ij}, \boldsymbol{u}_i)^{y_{ij}} f(y_{ij}=1|\gamma_{ij}, \boldsymbol{u}_i)^{1-y_{ij}}$ e reescrevendo $k_{ij} = \frac{\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}{1+\exp\{x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \boldsymbol{u}_i\}}$, tem-se que:

$$
\begin{aligned}
f(y_{ij}=1|\boldsymbol{u}_i) &= \int_0^1 f(y_{ij}=1|\gamma_{ij}, \boldsymbol{u}_i) f(\gamma_{ij}) d\gamma_{ij} \\
&= \int_0^1 \gamma_{ij} k_{ij} \gamma_{ij}^{\alpha-1} (1-\gamma_{ij})^{\kappa-1} [b(\alpha,\beta)]^{-1} d\gamma_{ij} \\
&= k_{ij} \frac{\Gamma(\alpha+1)\Gamma(\kappa)}{\Gamma(\alpha+\kappa+1)} \frac{\Gamma(\alpha+\kappa)}{\Gamma(\alpha)\Gamma(\kappa)} \\
&= \alpha k_{ij} (\alpha+\kappa)^{-1}. \qquad (1.22)
\end{aligned}
$$

Similarmente, obtém-se que:

$$
f(y_{ij}=0|\boldsymbol{u}_i) = \frac{(1-k_{ij})\alpha + \kappa}{\alpha+\kappa}. \qquad (1.23)
$$

Com isso, substituindo as expressões (1.22) e (1.23) em (1.21), chega-se à função de verossimilhança marginal, dependendo das soluções das integrais sobre $\boldsymbol{u}_i$:

$$
L(\boldsymbol{\beta}, \boldsymbol{D}, \alpha, \kappa) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \frac{1}{\alpha+\beta} (\alpha k_{ij})^{y_{ij}} [(1-k_{ij})\alpha + \kappa]^{1-y_{ij}} f(\boldsymbol{u}_i|\boldsymbol{D}) d\boldsymbol{u}_i. \qquad (1.24)
$$

Porém, a função de verossimilhança marginal fica dependendo de integrais sem

solução analítica, necessitando usar técnicas de otimização e métodos iterativos. Por outro lado, expressões bem mais simplificadas foram encontradas ao se utilizar a função de ligação probito, tornando o método menos intensivo computacionalmente, uma vez que as integrais foram solucionadas analíticamente (Vieira, 2008). Assim, pode-se utilizar as estimativas do modelo com ligação probito para obter aproximações para as estimativas do modelo caso se utilize a ligação logito. Bastando, para isso, multiplicar o preditor linear pela constante $c = 16\sqrt{3}/15\pi$.

1.4.3 Adaptação do Modelo de Rasch aos MLGMC para Heterogeneidade Atribuída à Fonte Desconhecida

Viu-se na seção 1.4 que os MLGMC tratam de uma variação extra-Bernoulli, captando toda a heterogeneidade provinda de fonte desconhecida. Com base nisso, Silva (2012) propôs adaptar o modelo de resposta ao item com 1 parâmetro (modelo de Rasch) a essa família. A mudança diz respeito ao índice i do modelo, que a princípio era designado para representar a resposta do indivíduo j no tempo, passando a indicar a qual item esta resposta está relacionada. Existem algumas vantagens ao se fazer essa abordagem. A primeira diz respeito à estrutura de covariância: o modelo adaptado permite admitir outros tipos de estruturas da matriz de variâncias e covariâncias (entre as respostas de um mesmo indivíduo, antes designado para as medidas longitudinais), relaxando a suposição de que as respostas dos itens sejam independentes. A segunda vantagem leva em conta a abordagem na classe dos MLG's, aproveitando, desta forma, os resultados e propriedades específicas conhecidas dessa família. Uma outra vantagem trata da possibilidade de se acomodar a heterogeneidade atribuída a uma fonte desconhecida. Foi visto que os modelos Bernoulli podem apresentar superdispersão e que esta variação extra pode ser oriunda de fontes desconhecidas ou variáveis como: faixa etária, níveis socioeconômicos, escola pública ou privada etc, mas que não são consideradas no modelo. Deste modo, pode-se controlar a violação do pressuposto de independência entre indivíduos. O modelo pode ser descrito como:

$$
\begin{aligned}
Y_{ij}|\theta_i &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \gamma_i \Phi(\theta_i - b_j) \\
\gamma_i &\sim \text{Beta}(\alpha, \kappa) \\
\theta_i &\sim I\!N(0,1)
\end{aligned}
\tag{1.25}
$$

em que θ_i é o traço latente (conhecimento) do respondente; b_i é parâmetro de dificuldade do item; γ_i é traço latente que irá acomodar a heterogeneidade atribuída à fonte desconhecida ou superdisperção com distribuição Beta, e π_{ij} é probabilidade do respondente i acertar o item j do teste.

Com essa adaptação, toda a fonte de heterogeneidade fica alocada em γ, porém desconsidera níveis de fatores como renda, origem de escola pública ou privada de cada indivíduo. Como γ varia entre 0 e 1, a probabilidade de sucesso de todos os respondentes da amostra é reduzida na mesma intensidade, levando a uma subestimação de uns ou superestimação para outros.

1.5 Estatística Bayesiana

A estatística Bayesiana na presença de incerteza atribui distribuições probabilísticas aos parâmetros de um modelo, e utilizando resultados da teoria da probabilidade estima-se os parâmetros, juntamente com os dados. Esta é uma teoria de contraponto à escola clássica da Estatística, que assume o parâmetro do modelo como um valor fixo e desconhecido e utiliza somente os dados para estimá-lo. Agregado a isso, tem-se um fundamento bayesiano: permitir incorporar informações *a priori* a respeito dos parâmetros de interesse. E essa informação pode e deve ser aumentada com a informação dos dados observáveis.

Contudo, na metodologia bayesiana a informação sobre o vetor de parâmetros $\boldsymbol{\theta}$ deve ser expressa (resumida) por meio de uma distribuição de probabilidade, denominada de distribuição ***a priori*** $\pi(\boldsymbol{\theta})$. A amostra $(\mathbf{y}|\boldsymbol{\theta})$, por sua vez, fornecerá a verossimilhança (plausibilidade) de cada valor possível de $\boldsymbol{\theta}$, ou seja, com os dados fixos variam-se os parâmetros com o intuito de achar uma combinação mais verossímil

para os dados observados. Pelo teorema de Bayes, essas duas informações são combinadas, resultando na distribuição ***a posteriori*** $\pi(\boldsymbol{\theta}|\mathbf{y})$. A distribuição *a posteriori* é utilizada para fazer inferência sobre os parâmetros. Seja $\boldsymbol{\theta}$ um vetor de parâmetros de interesse e $\pi(\boldsymbol{\theta})$ sua distribuição *a priori*. Definindo também a função de verossimilhança $L(\boldsymbol{\theta}|\mathbf{y})$, tem-se que, pelo teorema de Bayes, a distribuição *a posteriori* é dada por:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{y})}{\int \pi(\boldsymbol{\theta},\mathbf{y})d\boldsymbol{\theta}} \propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{y}). \tag{1.26}$$

Note que $\int \pi(\boldsymbol{\theta},\mathbf{y})d\boldsymbol{\theta} = f(\mathbf{y})$ não depende de $\boldsymbol{\theta}$. Portanto, trata-se apenas de uma constante no que diz respeito à distribuição *a posteriori*. O enfoque da inferência bayesiana baseia-se na distribuição *a posteriori* de $\boldsymbol{\theta}$, por conter toda a informação probabilística a respeito do parâmetro de interesse. Geralmente, resume-se essa informação através de estimação pontual (média, mediana, moda ou quantis) ou por intervalo (intervalo de credibilidade). Diversos livros tratam de detalhes sobre esses conceitos inferênciais, em particular, Migon e Gamerman (1999), Hoff (2009) e Gelman et al. (2004).

1.5.1 Distribuição a priori

A informação *a priori* sobre um vetor de parâmetros deve ser incorporada na forma de uma distribuição de probabilidade. A distribuição a priori representa o estado atual do conhecimento ou ignorância sobre o vetor de parâmetros (antes de obter os dados). Essa informação em grande parte das vezes é subjetiva e se origina de um especialista da área ou então parte de estudos estatísticos anteriores. Porém, em modelagens em que se trabalham com vários parâmetros, torna-se difícil definir uma distribuição *a priori*. Nesse caso, faz-se necessário definir uma *priori* que não influencie na análise dos dados, conhecida por *flat*, vaga, difusa ou não informativa. Portanto, há dois tipos de *priori*: as informativas e as não informativas.

Uma idéia para adotar *priori* não informativa é utilizar a distribuição uniforme ($\pi(\boldsymbol{\theta}) \propto$ `const.`). A literatura Bayesiana classifica uma distribuição de duas formas:

própria ou imprópria. Uma distribuição será **própria** se a integral sobre o espaço paramétrico resulta em valor finito, e imprópria se resulta em ∞ ($\int \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty$). *Prioris* não informativas podem ser próprias ou impróprias. Sabe-se que ao se utilizar *prioris* próprias, as *posterioris* marginais resultantes serão também próprias. Porém, *prioris* impróprias podem resultar tanto em *posterioris* marginais próprias quanto impróprias. O problema de se ter *posteriori* imprópria é que a inferência sobre os parâmetros fica inviabilizada. Desse modo, faz-se necessário verificar se a *posteriori* é própria ao utilizar uma *priori* imprópria. Têm-se também as densidades **não normalizadas**, que resultam em uma constante positiva, ou seja, são densidades a menos de uma constante que não depende dos parâmetros de interesse. Essas geralmente são utilizadas, pois simplificam os cálculos das integrais envolvidas na obtenção das marginais *a posteriori*.

É importante que se tenha uma boa interação entre o pesquisador e o estatístico para que a informação sobre o fenômeno seja bem compilada na forma de uma distribuição. A distribuição a *priori*, como qualquer outra, possui parâmetros, que são denominados de **hiperparâmetros** para diferenciá-los de fato dos parâmetros de interesse $\boldsymbol{\theta}$. Existe uma classe de prioris próprias que é bastante utilizada, denominadas prioris conjugadas. Isso quer dizer que as distribuições *a priori* e *a posteriori* devem pertencer à mesma classe de distribuições. Assim, a atualização do conhecimento que se tem de $\boldsymbol{\theta}$ envolve apenas uma mudança nos hiperparâmetros. Esse artifício simplifica os cálculos da *posteriori* que geralmente são complicados e necessitam de auxílio computacional. Comumente, a conjugação é feita por membros da família exponencial.

Em modelos mais complexos (com efeitos fixos e aleatórios), a distribuição *a priori* é dividida em estágios, pois além de facilitar a especificação do mesmo, a abordagem se torna natural em determinadas situações experimentais e observacionais. Modelos deste tipo são denominados de modelos hierárquicos. Assim, os parâmetros da distribuição *a priori* também seguem alguma distribuição, configurando, no caso, uma hierarquia em 2 estágios. Os níveis de hierarquia condizem com as quantidades de distribuições especificadas para os hiperparâmetros. Contudo, a distribuição *a posteriori*

no caso de 2 estágios é dada por:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) \propto L(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y})\pi(\boldsymbol{\theta}|\boldsymbol{\phi})\pi(\boldsymbol{\phi}) \propto L(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta}|\boldsymbol{\phi})\pi(\boldsymbol{\phi})$$

Mais detalhes sobre os modelos bayesianos podem ser encontrados em Gelman et al. (2004).

1.5.2 Aproximações *a posteriori*

Uma das desvantagens da metodologia bayesiana é a complexidade das integrais envolvidas nos cálculos que geralmente aparecem em modelos com grandes dimensões do espaço paramétrico. E, na maioria das vezes, isso faz com que a obtenção da *posteriori* se resuma em cálculo de integral, que geralmente não são analiticamente tratáveis. Ou seja, não é possível encontrar uma solução exata das integrais. Pode-se dizer que, devido a esse problema, houve um atraso dos modelos bayesianos, já que a implementação destes dependiam de computadores capazes de executar algoritmos intensivos. Porém, com o avanço computacional na década de 90, juntamente com a publicação do trabalho de Gelfand e Smith (1990), surgiram vários métodos de simulação nesta área.

Existem diversos métodos de aproximação, como os determinísticos e o de simulação. Neste trabalho serão explorados os principais métodos de simulação estocástica, os métodos de Monte Carlo via cadeias de Markov. Dentre os mais utilizados para a construção da cadeia de Markov está o Amostrador de Gibbs, proposto por Geman e Geman (1984) apud Resende (2011) e o método de Metropolis-Hastings.

Esses métodos são bem explorados e descritos na literatura bayesiana, e a idéia básica é gerar uma amostra da distribuição *a posteriori* através da construção de uma cadeia de Markov de modo que a distribuição estacionária desta seja a distribuição de interesse.

1.5.2.1 Metropolis-Hastings

O algoritmo de Metropolis-Hastings foi proposto por Metropolis et al. (1953) e posteriormente estendido por Hastings (1970) apud Resende (2011). A idéia do

método é gerar amostras de uma distribuição (a posteriori) através de valores gerados por uma distribuição auxiliar (proposta) e aceitos com dada probabilidade. A probabilidade de aceitação de cada valor faz com que seja atingida a convergência da cadeia para a distribuição.

Seja $p(.)$ uma função de probabilidade de interesse e $q(.)$ uma função de probabilidade proposta. O algoritmo segue os seguintes passos:

I Determina-se um valor inicial arbitrário $\psi^{(0)}$ para ψ e inicializa-se o contador j=1;

II Gera-se um valor ξ da distribuição proposta $q(\xi|\psi^{(j-1)})$.

III Aceita-se o valor gerado em II com probabilidade $\texttt{min}\left\{1, \frac{p(\xi)q(\psi^{(j-1)}|\xi)}{p(\psi^{(j-1)})q(\xi|\psi^{(j-1)})}\right\}$. Se for aceito, $\psi^{(j)} = \xi$. Caso contrário, a cadeia não se move e $\psi^{(j)} = \psi^{(j-1)}$.

IV Atualiza-se o contador de j para $j + 1$ e retorna-se ao passo II até que a convergência seja obtida.

Considerando que, a partir da k-ésima iteração, a cadeia atinja a convergência (k suficientemente grande), então os valores simulados $(\psi^{(k)}, \psi^{(k+1)}, \ldots, \psi^{(k+n)})$ podem ser usados como uma amostra da distribuição $p(.)$, *a posteriori*. Não se sabe ao certo quantas iterações do processo são necessárias para que se tenha uma cadeia estacionária, porém, pode-se verificar a convergência da mesma através de análise gráfica e de técnicas de diagnóstico de convergência e, assim, escolher um valor apropriado de k. Esse assunto será visto na seção 1.5.2.3. Uma vantagem desse método é que ele pode ser implementado apenas conhecendo parcialmente a distribuição de interesse, isto é, a função de probabilidade a menos de uma constante que não dependa do parâmetro de interesse. Esse fato é importante pois facilita os cálculos bayesianos para achar a distribuição *a posteriori*.

1.5.2.2 Gibbs Sampling

É um caso particular do algoritmo de Metropolis-Hastings. Aqui as distribuições propostas são as distribuições condicionais completas de cada parâmetro e a probabilidade de aceitação é igual a 1. A distribuição condicional completa de um

parâmetro é a distribuição deste parâmetro condicional à informação de todos os outros parâmetros, dada da forma $\pi(\theta_i|\boldsymbol{\theta}_{-i})$, onde $\boldsymbol{\theta}_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d)^\top$. Nesse caso não existe mecanismo de aceitação e a cadeia irá sempre se mover para um novo valor. A seguir as etapas do algoritmo:

I Determinam-se valores iniciais $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})^\top$ e inicializa-se o contador j=1;

II Obtém-se um novo valor de $\boldsymbol{\theta}^{(t)}$ a partir de $\boldsymbol{\theta}^{(t-1)}$ através da geração sucessiva dos valores:

$$\begin{aligned}
\theta_1^{(t)} &\sim \pi(\theta_1^{(t-1)}|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
\theta_2^{(t)} &\sim \pi(\theta_2^{(t-1)}|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
&\vdots \\
\theta_d^{(t)} &\sim \pi(\theta_d^{(t-1)}|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{d-1}^{(t)})
\end{aligned}$$

III Atualiza-se o contador de j para $j + 1$ e retorna-se ao passo II até que a convergência seja obtida.

É importante destacar que mesmo para um problema de grandes dimensões envolvendo distribuição *a priori* hierárquica, o algoritmo trata as simulações de modo univariado, o que vem a ser uma vantagem computacional. Porém, é necessário conhecer as distribuições condicionais completas para que seja possível gerar amostras a partir delas.

1.5.2.3 Avaliação da Convergência da Cadeia de Markov

Alguns cuidados devem ser tomados para se ter uma boa amostra da distribuição *a posteriori*, através das simulações de MCMC. É necessário verificar se a cadeia atingiu a estacionariedade, de forma a representar uma amostra independente da distribuição de interesse. Sabe-se que os vetores de parâmetros iniciais simulados pelos amostradores citados anteriormente, em geral, são auto-correlacionados, tornando-se um problema para as inferências. Gilks et al. (1996) e Cowles e Carlin (1996)

apresentam estudos sobre o diagnóstico da cadeia. O número de iterações a ser definido vai depender da estrutura de correlação, que, consequentemente, implicará na rapidez da convergência.

Usualmente, uma maneira direta de avaliar a convergência é considerar alguns valores iniciais da cadeia e monitorar graficamente a trajetória. Espera-se que a partir de uma certa iteração as cadeias se estabilizem em torno de uma média (comum) e variância constante. É importante avaliar a convergência de todos os parâmetros e não apenas aqueles de interesse. Cowles e Carlin (1996) compararam diversos testes estatísticos de diagnósticos como o de Gelman e Rubin (1992), Raftery e Lewis (1992), Heidelbelger e Welch (1983) e Geweke (1992). Porém, o autor não pôde dizer qual era o mais eficiente. Esses testes tratam de condições necessárias, mas não suficientes, de convergência da cadeia. Não há testes conclusivos que indiquem de fato que a cadeia convergiu, há apenas indícios.

Alguns procedimentos são utilizados para contornar o problema da auto-correlação e assim obter uma "boa" amostra. A primeira é considerar o *Burn-in* (aquecimento) que consiste em descartar as primeiras (500, 1.000 ou 2.000, dependendo do modelo) iterações, buscando eliminar o efeito dos valores iniciais (fase transiente). Um outro procedimento, denominado *thin*, trata de definir um valor k, gerar (se for possível) uma cadeia mais longa e tomar um valor amostrado a cada k valores gerados, de forma a burlar o efeito da correlação entre as amostras simuladas. No entanto, essa técnica pode reduzir consideravelmente o tamanho amostral simulado. Recomenda-se não utilizá-la para fazer inferências sobre medidas de posição. A seguir será apresentado um breve relato sobre os principais testes de avaliação de convergência.

Critério de Gelman e Rubin

Proposto por Gelman e Rubin (1992), o critério utiliza cadeias paralelas, ou seja, simula várias cadeias com diferentes valores iniciais do espaço paramétrico. Os últimos 50% dos valores de cada cadeia são comparados no que diz respeito aos valores inferênciais. Se forem bem similares, isto indica que a cadeia alcançou ou esta próxima da convergência. O teste se baseia na razão de variâncias e indica aceitação de convergência se o fator de redução de escala $\hat{R}_c$ estiver entre 1 e 1,1.

Critério de Raftery e Lewis

Critério proposto por Raftery e Lewis (1992) que busca definir o número de ite-

rações a serem descartadas (*burn-in*), o espaçamento entre as iterações (*Thin*) e o número total de iterações de forma a considerar uma subamostra independente. O interesse desse teste é avaliar a precisão dos quantis estimados, e não diz respeito à convergência em si da cadeia. Mais detalhes em Brooks e Roberts (1999).

Critério de Geweke

Geweke (1992) propôs esse teste para detectar a falta de convergência. Baseia-se no teste de igualdade de médias (Teste Z bilateral) do começo e do fim dos valores da iteração da cadeia, usualmente utiliza-se os 10% primeiros e os últimos 50%.

Critério de Heidelberger e Welch

O teste, proposto por Heidelberger e Welch (1981), é dividido em duas partes. Primeiro testa a hipótese de estacionariedade da cadeia através da estatística de *Cramér-von Mises*. Esse teste pode ser feito sucessivamente eliminando as primeiras iterações até identificar a t-ésima em que o mesmo passa a ser aceito. A segunda parte do teste consiste em verificar se a cadeia, a partir da t-ésima iteração, possui dados suficientes para estimar precisamente a média *a posteriori* com uma certa precisão, usando o teste *half-width*.

1.5.3 Seleção de Modelos

É bem comum em experimentos e pesquisas existir o questionamento filosófico sobre como algumas variáveis de medidas estão relacionadas (interligadas) e quais afetam o resultado investigado. Diante disso, as escolas estatísticas durante anos se posicionaram sobre o assunto criando e aprimorando critérios de seleção de modelos. A seguir serão apresentados os principais métodos de seleção e avaliação de qualidade de ajuste. Kadane e Lazar (2004) apresentam métodos e critérios de seleção de modelos, e essa publicação servirá de referêncial para a revisão que se segue.

Critério de Informação de Akaike

Proposto por Akaike (1974), o Critério de Informação de Akaike (AIC) é usado para comparar modelos, no sentido de escolher o mais ajustado aos dados. Não é um teste e nem indica se o modelo selecionado pelo critério está ou não bem ajustado, apenas compara com os demais. Contudo, o melhor modelo é aquele que possui menor

medida de AIC, dada por:

$$\texttt{AIC}(\mathcal{M}_i) = 2d_i - 2L(\boldsymbol{\theta}) \tag{1.27}$$

em que $\mathcal{M}_i$, com $i = 1, \ldots, q$, é a classe de modelos alternativos e d_i o número de parâmetros. Uma outra versão desse critério, o AICc (corrigido para amostras finitas) é apresentado em Burnham e Anderson (2002), recomendado para utilização em pequenas amostras ou d (número de parâmetros) grande:

$$\texttt{AICc}(\mathcal{M}_i) = \texttt{AIC}(\mathcal{M}_i) + \frac{2d_i(d_i+1)}{n-k-1} \tag{1.28}$$

sendo n o tamanho amostral. Burnham e Anderson (2002) sugerem o uso do AICc quando a razão n/p é pequena (< 40). Note que ambos os critérios penalizam modelos com muitos parâmetros, desencorajando superajustes.

Critério de Informação Bayesiano

O Critério de Informação Bayesiano (Bayesian Information Criterion), também conhecido por BIC, foi proposto por Schwarz (1978) e modificado por Carlin e Louis (1996). Deve-se preferir o modelo que apresenta a maior medida de BIC, dado pela expressão a seguir:

$$\texttt{BIC}(\mathcal{M}_i) = I\!E(\ln[L(\boldsymbol{\theta})]) - \frac{1}{2}d_i \ln(n). \tag{1.29}$$

sendo $I\!E(ln[L(\boldsymbol{\theta})])$ o valor esperado em relação a densidade *a posteriori*, da função log-verossimilhança.

Critério Desvio-Informação

Proposto por Spiegelhalter et al. (2002), o Critério Desvio-Informação (DIC) é uma generalização do BIC. Esse critério se destaca dos outros por poder ser incorporado em modelos para os quais *a posteriori* foi obtida por meio de simulações de Monte Carlo via cadeias de Markov.

Considerando $\{\boldsymbol{\theta}_i^{(1)}, \ldots, \boldsymbol{\theta}_i^{(r)}\}$ a amostra *a posteriori* dos parâmetros do modelo $\mathcal{M}_i$, a implementação computacional do DIC é dada por:

$$\texttt{DIC}(\mathcal{M}_i) \approx -\frac{4}{M}\sum_{j=1}^{M} \log L(\boldsymbol{\theta}_i^{(j)}) + 2\log L(\hat{\boldsymbol{\theta}}_i) \tag{1.30}$$

em que $\hat{\boldsymbol{\theta}}_i = \frac{1}{M}\sum_{j=1}^{r} \boldsymbol{\theta}_i^{(j)}$ é uma estimativa de $\widetilde{\boldsymbol{\theta}}_i$. Brooks (2002) traz mais detalhes sobre o critério e a implementação.

Fator de Bayes

Esse método de seleção é usado para comparação entre dois modelos, fazendo uso da razão das verosimilhanças marginais. Jeffreys (1961) propôs o critério, e define-o da seguinte forma:

$$B_{ij} = \frac{f(\mathbf{y}|\mathcal{M}_i)}{f(\mathbf{y}|\mathcal{M}_j)} \tag{1.31}$$

em que $f(\mathbf{y}|\mathcal{M})$ é a verossimilhança marginal definida por:

$$f(\mathbf{y}|\mathcal{M}) = \int L(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Desse modo, se B_{ij} é maior que 1, é preferível escolher o modelo i. Por outro lado, se B_{ij} for menor que 1, é preferível escolher o modelo j. O Fator de Bayes pode ser interpretado a favor de um modelo de acordo, segundo Jeffreys (1961), com as seguintes categorias:

Tabela 1.1: Interpretação do Fator de Bayes.

B_{ij}	$Log(B_{ij})$	Evidência a favor do modelo $\mathcal{M}_i$
1 a 3,2	$< 0,5$	Muito Fraca
3,2 a 10	0,5 a 1	Fraca
10 a 100	1 a 2	Forte
> 100	> 2	Muito Forte

Fonte: Meyes, 2009

Uma proposta computacional para a implementação desse método (ver Raftery, 1996) usando as amostras geradas por simulação MCMC é dada pela estimação das verossimilhanças marginais:

$$\hat{f}(\mathbf{y}|\mathcal{M}) = \frac{1}{r}\sum_{j=1}^{r} L(\boldsymbol{\theta}^{(j)}|\mathbf{y}). \tag{1.32}$$

Consequentemente, tem-se:

$$\hat{B}_{ij} = \frac{\hat{f}(\mathbf{y}|\mathcal{M}_i)}{\hat{f}(\mathbf{y}|\mathcal{M}_j)}, \tag{1.33}$$

devendo preferir o modelo $\mathcal{M}_i$ se $\hat{B}_{ij} < 1$.

Capítulo 2

Modelo de Resposta ao Item Adaptado ao Controle da Heterogeneidade

A seguir será apresentada a definição do modelo de resposta ao item proposto neste trabalho. O modelo propõe buscar ajuste em dados que estão sujeitos a fatores externos (conhecidos), já que essa presença, entre outros efeitos, causa um **aumento da heterogeneidade**.

A heterogeneidade nos modelos da TRI, usada em avaliação educacional, pode ser devida aos agrupamentos (clusters) de alunos gerados por fatores conhecidos ou desconhecidos (por não se ter dados disponíveis), por exemplo, se não fosse medido quem estudou em escola pública ou privada. Sabe-se empiricamente que esse fator gera probabilidades de acerto diferentes para os dois grupos e, certamente, vai comprometer o pressuposto de independência condicional entre os respondentes, gerando assim heterogeneidade. Por outro lado, se esse fator é medido, mas não é levado em conta pelo modelo, ele continua gerando heterogeneidade. A Figura 2.1 mostra bem esse exemplo pelas curvas característica dos itens de 1 a 9 da prova de matemática da 8.ª série do ano de 2005 do Sistema nacional de avaliação básica - SAEB. O ajuste foi feito pelo modelo de Rasch com 1 parâmetro para cada grupo de rede de ensino. Observa-se que para alunos da rede privada as probabilidades de acerto são maiores para todos os itens aqui analisados, dado uma mesma habilidade. Realizou-se também

o ajuste nos mesmos dados para cada grupo de cor de pele considerado pelo estudante, utilizando modelo de Rasch com 1 parâmetro, ver Figura 2.2. Nota-se que a cor da pele também influência nas respostas dadas. Alunos considerados com cor de pele branca e amarela levam ligeira vantagem sobre os de cor negra e indígena. Esse fato é mais acentuado nos itens 4 e 7.

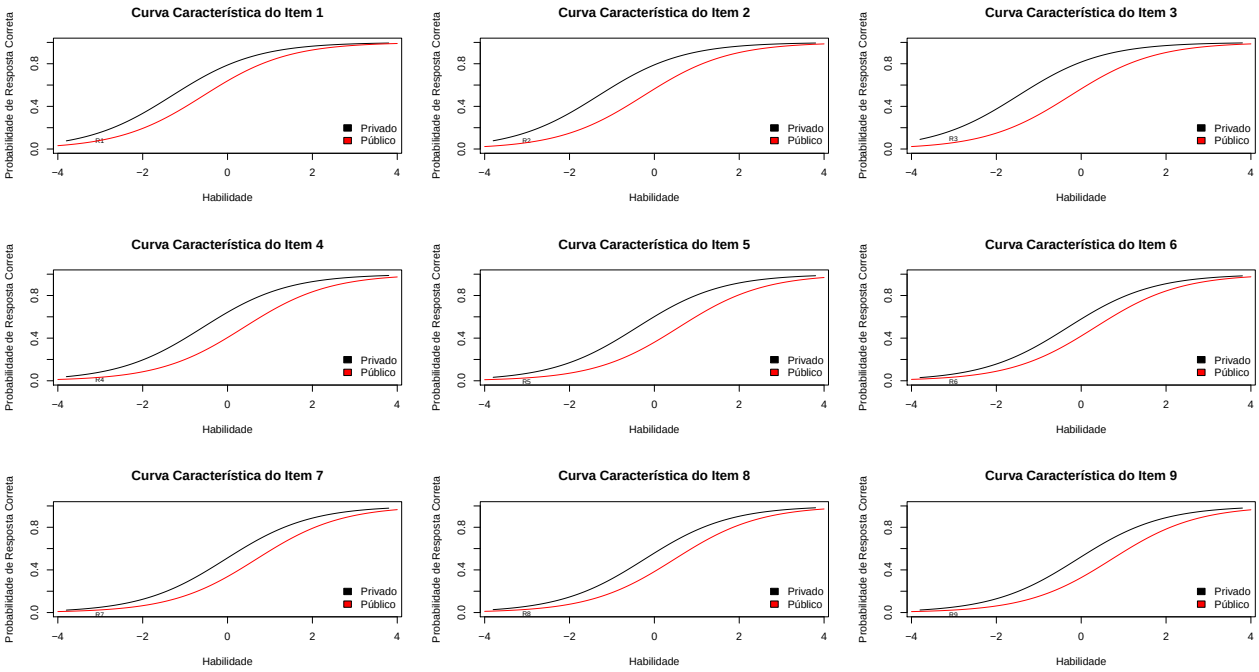


Figura 2.1: Curvas características dos itens 1 a 9 da prova de matemática 8.ª série do ano de 2005 - SAEB, por rede de ensino.

O modelo aqui tratado será capaz de usar essas e outras informações no ajuste dos parâmetros de interesse, contribuindo em estimativas mais precisas, pois levará em conta a heterogeneidade atribuída a fatores conhecidos. O mesmo terá abordagem bayesiana e a metodologia MCMC será aplicada para o seu ajuste, através dos pacotes *MCMCpack* e *R2WinBUGS* do ambiente computacional **R**. Apesar do modelo ser essencialmente bayesiano, o mesmo pode ser visto de modo frequentista, bastanto para isso desconsiderar as distribuições *a priori* para θ_j, b_i e β_k das expressões (2.3), (2.4) e (2.5), respectivamentes. O modelo de TRI com controle da heterogeneidade atribuída a fatores conhecidos é definido por:

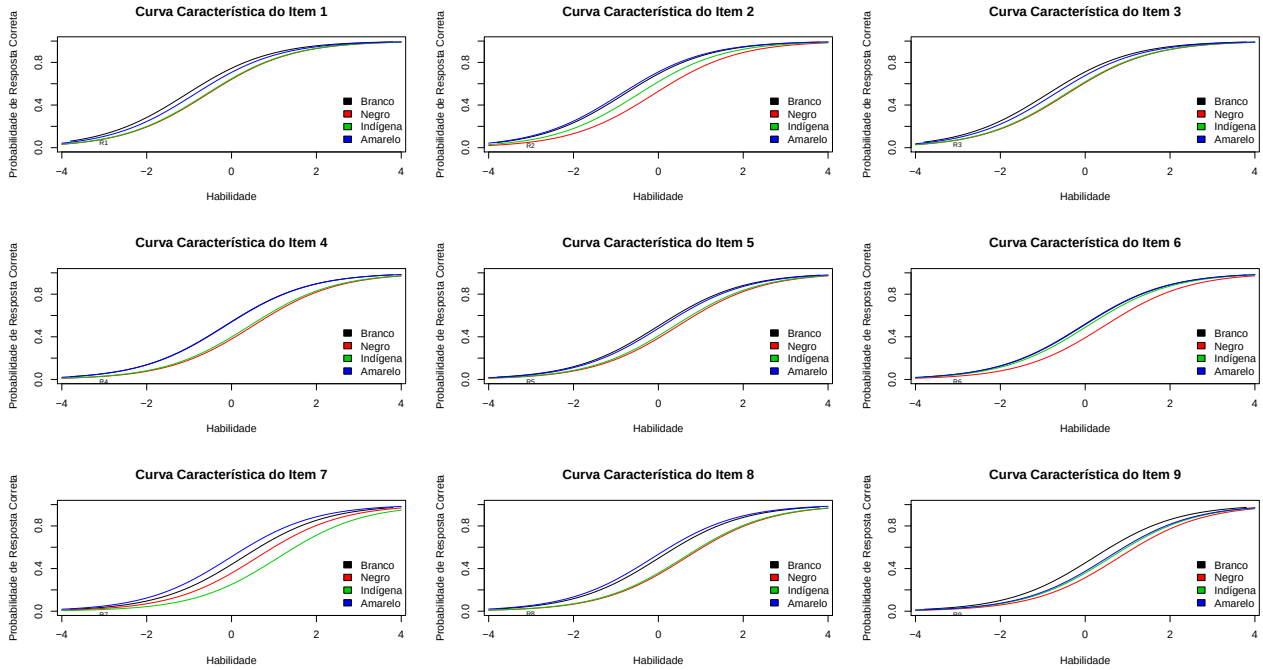


Figura 2.2: Curvas características dos itens 1 a 9 da prova de matemática 8.ª série do ano de 2005 - SAEB, por grupo de cor.

$$
\begin{aligned}
f(Y_{ij}=1|\theta_j) &= \frac{e^{(\theta_j - b_i)}}{1+e^{(\theta_j-b_i)}}\psi_j, \quad i=1,\ldots,I\ ;\ j=1,\ldots,n; && (2.1)\\
\text{logit}(\psi_j) &= \mathbf{x}_j^\top \boldsymbol{\beta}; && (2.2)\\
\theta_j &\sim N(0,1); && (2.3)\\
b_i &\sim N(0,\sigma_b^2); && (2.4)\\
\beta_k &\sim N(0,\sigma_\beta^2), \quad k=1,\ldots,p; && (2.5)
\end{aligned}
$$

em que Y_{ij} representa a variável aleatória que assume valor 1 caso o indivíduo j acerte o item i, e assume valor 0 caso contrário; θ_j corresponde ao traço latente (conhecimento ou habilidade) do respondente j, $j=1,\ldots,n$, assumindo que θ_j possui distribuição *a priori* normal padrão; b_i é parâmetro de dificuldade do item i, medido na mesma escala do traço latente. Assume-se que b_i possui distribuição *a priori* normal com hiperparâmetros 0 e σ_b^2. ψ_j é o efeito multiplicativo com informações do respondente j, ligado por meio da função logit ao preditor linear $\mathrm{x}_j\boldsymbol{\beta}$, e $\boldsymbol{\beta}$ é o vetor p dimensional de parâmetros desconhecidos cuja distribuição *a priori* é normal. Para um específico β_k, essa distribuição possui hiperparâmetros 0 e σ_β^2. O vetor $\mathbf{x}_j^\top =$

$(x_{1j}, x_{2j}, \ldots, x_{pj})$ representa as características do indivíduo j com as potenciais fontes de heterogeneidade do indivíduo. Define-se, ainda, que $\mathbf{X} = (\mathbf{x}_1{}^\top, \mathbf{x}_2{}^\top, \ldots, \mathbf{x}_n{}^\top)^\top$ é a matriz com as informações de todos os indivíduos da amostra. Denota-se por $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^\top$ e $\boldsymbol{b} = (b_1, b_2, \ldots, b_I)^\top$ os vetores contemplando os parâmetros de traços latentes e de dificuldades dos itens, respectivamente.

Note que é necessário ter disponível os possíveis fatores causadores da heterogeneidade, a fim de serem alocados na matriz $\mathbf{X}$. De acordo com as definições tomadas, tem-se as seguintes funções de densidade *a priori*:

$$f(\theta_j) = \sqrt{\frac{1}{2\pi}}\exp\Big\{\frac{-\theta_j^2}{2}\Big\} \propto \exp\Big\{\frac{-\theta_j^2}{2}\Big\}; \tag{2.6}$$

$$f(b_i) = \sqrt{\frac{\sigma_b^{-2}}{2\pi}}\exp\Big\{\frac{-\sigma_b^{-2}}{2}b_i^2\Big\} \propto \exp\Big\{\frac{-\sigma_b^{-2}}{2}b_i^2\Big\}; \tag{2.7}$$

$$f(\beta_k) = \sqrt{\frac{\sigma_\beta^{-2}}{2\pi}}\exp\Big\{\frac{-\sigma_\beta^{-2}}{2}\beta_k^2\Big\} \propto \exp\Big\{\frac{-\sigma_\beta^{-2}}{2}\beta^2\Big\}. \tag{2.8}$$

Pelo Teorema de Bayes, tem-se a função de densidade de probabilidade conjunta *a posteriori*:

$$f(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{b}|\mathbf{X}, \mathbf{Y}) \propto L(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\beta}|\mathbf{X}, \mathbf{Y}) f(\boldsymbol{\theta}) f(\boldsymbol{b}) f(\boldsymbol{\beta}). \tag{2.9}$$

Considerando que $Y_{ij} \sim \text{Bin}(p_{ij}\psi_j)$, em que $\text{logit}(p)$=$\theta_j - b_i$ e $\text{logit}(\psi_j) = \mathbf{x}_j^\top\boldsymbol{\beta}$, a função de verossimilhança é dada por:

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{b}|\mathbf{X}, \mathbf{Y}) &= \prod_{i=1}^{I}\prod_{j=1}^{n} f(U_{ij}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{b}, \mathbf{X}) \\
&= \prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{\exp\{\theta_j - b_i\}}{1+\exp\{\theta_j - b_i\}}\frac{\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}}{1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}}\right)^{y_{ij}}\left(1 - \frac{\exp\{\theta_j - b_i\}}{1+\exp\{\theta_j - b_i\}}\frac{\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}}{1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}}\right)^{1-y_{ij}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{\exp\{\theta_j - b_i + \mathbf{x}_j^\top\boldsymbol{\beta}\}}{[1+\exp\{\theta_j - b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right)^{y_{ij}} \times \\
&\quad \times\left(\frac{[1+\exp\{\theta_j - b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}] - \exp\{\theta_j - b_i + \mathbf{x}_j^\top\boldsymbol{\beta}\}}{[1+\exp\{\theta_j - b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right)^{1-y_{ij}}
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{\exp\{\theta_j - b_i + \mathbf{x}_j^\top\boldsymbol{\beta}\}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right)^{y_{ij}}\left(\frac{1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right)\times \\
&\times\left(\frac{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}{1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}}\right)^{y_{ij}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{n}(\exp\{\theta_j - b_i + \mathbf{x}_j^\top\boldsymbol{\beta}\})^{y_{ij}}\left(\frac{(1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\})^{1-y_{ij}}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right) \\
&= \exp\Big[\sum_{i=1}^{I}\sum_{j=1}^{n}\{y_{ij}\theta_j - y_{ij}b_i + y_{ij}\mathbf{x}_j^\top\boldsymbol{\beta}\}\Big]\prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{(1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\})^{1-y_{ij}}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right) \\
&= \exp\{\mathbf{1}_{1\times I}^\top\mathbf{Y}\boldsymbol{\theta} - \boldsymbol{b}^\top\mathbf{Y}\mathbf{1}_{n\times 1} + \mathbf{1}_{1\times I}^\top\mathbf{Y}\mathbf{X}\boldsymbol{\beta}\}\prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{[1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]^{1-y_{ij}}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right)
\end{aligned}
\tag{2.10}
$$

Aplicando (2.6), (2.7), (2.8), (2.10) em (2.9) tem-se a seguinte função de densidade conjunta *a posteriori*:

$$
\begin{aligned}
P(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{b}|\mathbf{X},\mathbf{Y}) &\propto L(\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{\beta}|\mathbf{X},\mathbf{Y})P(\boldsymbol{\theta})P(\boldsymbol{b})P(\boldsymbol{\beta}) \\
&\propto \exp\{\mathbf{1}_{1\times I}^\top\mathbf{Y}\boldsymbol{\theta} - \boldsymbol{b}^\top\mathbf{Y}\mathbf{1}_{n\times 1} + \mathbf{1}_{1\times I}^\top\mathbf{Y}\mathbf{X}\boldsymbol{\beta}\}\prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{[1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]^{1-y_{ij}}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right)\times \\
&\quad\times\prod_{j=1}^{n}\Big[\exp\Big\{\frac{-\theta_j^2}{2}\Big\}\Big]\prod_{i=1}^{n}\Big[\exp\Big\{\frac{-\sigma_b^{-2}}{2}b_i^2\Big\}\Big]\prod_{k=1}^{p}\Big[\exp\Big\{\frac{-\sigma_\beta^{-2}}{2}\beta_k^2\Big\}\Big],
\end{aligned}
$$

e portanto,

$$
\begin{aligned}
P(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{b}|\mathbf{X},\mathbf{Y}) &\propto \exp\{\mathbf{1}_{1\times I}^\top\mathbf{Y}\boldsymbol{\theta} - \boldsymbol{b}^\top\mathbf{Y}\mathbf{1}_{n\times 1} + \mathbf{1}_{1\times I}^\top\mathbf{Y}\mathbf{X}\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\theta}^\top\boldsymbol{\theta} - \frac{1}{2\sigma_b}\boldsymbol{b}^\top\boldsymbol{b} - \frac{1}{2\sigma_\beta}\boldsymbol{\beta}^\top\boldsymbol{\beta}\}\times \\
&\quad\times\prod_{i=1}^{I}\prod_{j=1}^{n}\left(\frac{[1+\exp\{\theta_j-b_i\}+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]^{1-y_{ij}}}{[1+\exp\{\theta_j-b_i\}][1+\exp\{\mathbf{x}_j^\top\boldsymbol{\beta}\}]}\right).
\end{aligned}
\tag{2.11}
$$

2.1 Aplicação em Dados Simulados

Aqui apresentamos resultados numéricos da aplicação do modelo em dados simulados computacionalmente. Nesse ambiente controlado, pôde-se analisar a viabilidade do modelo e o pacote computacional mais adequado para seu ajuste. Adotou-se aqui a sigla MACH para se referir ao modelo adaptado ao controle da heterogeneidade.

Pré-Teste

Foi realizada a implementação computacional do modelo proposto através do pacote *MCMCpack* utilizado a função *MCMCmetrop1R* e, em seguida, através da função *gibbs* do pacote *LearnBayes*, ambos do ambiente computacional **R**. No pré-teste do modelo, os dados simulados tiveram a seguinte configuração: 500 respondentes, 5 itens e 2 covariáveis para heterogeneidade. De modo que os traços latentes dos respondentes foram amostrados de uma distribuição Normal padrão, o vetor de parâmetros de dificuldade dos itens dado por $\boldsymbol{b} = (-2, -1, 0, 1, 2)$ e os parâmetros das covariáveis para a superdispersão $\boldsymbol{\beta} = (1, 1.5)$. A matriz de características $\mathbf{X}_{500\times2}$ foi construída de forma a possuir 4 grupos de respondentes (4 níveis de ψ). Assim, para cada coluna da matrix $\mathbf{X}_{500\times2}$, foram sorteados 250 respondentes para assumir característica de valor 1, e os demais passaram a assumir característica de valor 2. De forma que possua 4 grupos de 125 respondentes, em média. Portanto, os níveis de ψ assumem os valores: 0.993, 0.982, 0.971 e 0.924. Para a geração da cadeia foram consideradas 1.000 iterações, *thin* 1 e *Burn in* de 100 iterações. A utilização do algoritmo de Metropolis, pela função *MCMCmetrop1R*, foi inviabilizada devido ao grande número de parâmetros, fazendo com que a taxa de aceitação tendesse a zero. Por outro lado, o amostrador de gibbs, executado pela função *gibbs*, apresentou estimativas plausíveis em relação aos dados simulado, ver Tabela 2.1. A simulação levou em torno de duas horas e meia para ser concluída em um computador Dual Core E2140 1.60 GHz, 2 GB DDR3 com sistema operacional Windows XP 32 bit.

Tabela 2.1: Resultados do ajuste do modelo com 500 respondentes, 5 itens e 2 fatores de heterogeneidade. Pré-teste com 1.000 iterações do amostrador de Gibbs.

Parâmetros	Valor Real	Valor estimado					
		Estatísticas					
		Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
b_1	-2	-2,662	-2,059	-1,946	-1,947	-1,826	-1,507
b_2	-1	-1,473	-1,066	-0,979	-0,981	-0,889	-0,548
b_3	0	-0,358	-0,107	-0,019	-0,023	0,058	0,398
b_4	1	0,579	0,876	0,958	0,958	1,040	1,335
b_5	2	1,465	1,801	1,905	1,906	2,000	2,389
β_1	1	-0,560	0,937	0,346	1,405	1,833	4,203
β_2	1,5	-0,687	1,004	1,400	1,442	1,874	4,420

Os resultados preliminares do modelo mostraram-se satisfatórios. Os valores reais dos parâmetros pertenceram ao intervalo entre o 1° e o 3° quartil das amostras gera-

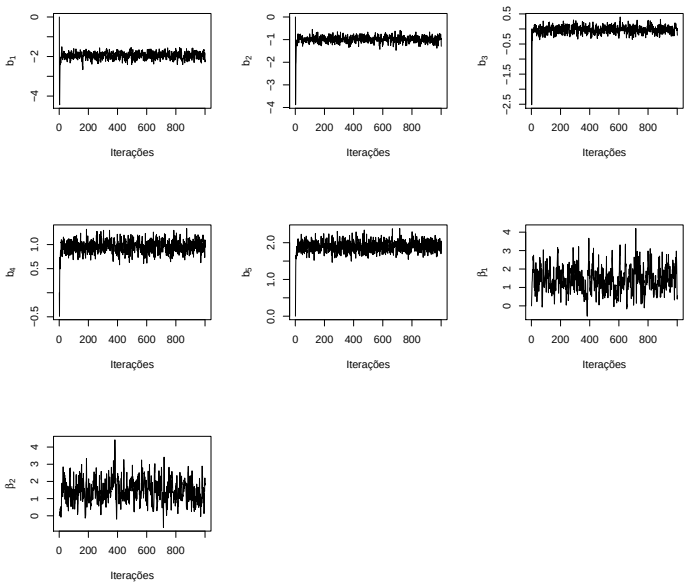


Figura 2.3: Gráfico de iterações do amostrador de Gibbs para os parâmetros de dificuldade e de superdispersão.

das por Gibbs. Porém, nota-se claramente que as cadeias de β_1 e β_2 não atingiram a convergência. Ver Figura 2.3. Ressalta-se ainda o fato de se ter utilizado uma amostra simulada relativamente pequena de 500 respondentes, 5 itens e 2 covariáveis para heterogeneidade, com apenas 1.000 iterações de Gibbs. Apesar do pacote *LearnBayes* ter se mostrado mais viável que o *MCMCmetrop1R* para esse problema, não foi possível realizar simulação com grandes amostras devido ao esforço computacional se intensificar exponencialmente com o aumento das iterações e do tamanho amostral. Parecendo, assim, ser inviável em uma aplicação real, passando mais de sete dias para se obter estimativas a partir de dados com 2.000 alunos e 5.000 iterações.

Desse modo, foi necessário explorar em outros softwares estatísticos meios de se ajustar o modelo em tempo hábil. O software Winbugs 14 manipulado pelo **R** através do pacote *R2WinBugs* mostrou-se bem superior que o algoritmo de gibbs executado pelo *LearnBayes*, pelo fato de monitorar apenas as cadeias de interesse, $\boldsymbol{b}$ e $\boldsymbol{\beta}$. Finalizando o processo de estimação com 1.500 respondentes e 100.000 iterações, com parâmetros idênticos ao do pré-teste, em torno de 4 horas (computador Atom N330 Dual Core 1,6 GHz, 2 GB de RAM com sistema operacional Microsoft Windows XP).

Simulação

Pelos resultados do pré-teste, viu-se que a execução da simulação pelo Winbugs 14 através do pacote *R2WinBugs* do **R** foi mais viável. Assim, optou-se por utilizar esse pacote para uma simulação com amostra grande.

A simulação foi realizada em um computador i7 - 2.2ghz, 8gb ddr3, placa de vídeo GEFORCE 630M com sistema operacional Microsoft Windows 7 - 64 bits. Os dados tiveram a seguinte configuração: 5.000 respondentes, 23 itens e 2 covariáveis para a heterogeneidade. A matriz de características $\mathbf{X}_{5000\times2}$ foi construída de forma a possuir 4 grupos de respondentes (4 níveis de ψ). Desse modo, para cada coluna da matriz $\mathbf{X}_{5000\times2}$, foram sorteados 2.500 respondentes para assumir característica de valor 1, e os demais passaram a assumir característica de valor 2. De forma que possua 4 grupos de 1.250 respondentes, em média. Os traços latentes dos respondentes foram amostrados de uma distribuição Normal padrão. O vetor de parâmetros de dificuldade dos itens foi composto por 5 repetições de cada um dos seguintes valores: -2,-1,0,1, respectivamente, três repetições de 2 e uma de 3, totalizando 23 parâmetros. Os parâmetros das covariáveis para a superdispersão foram $\boldsymbol{\beta} = (1, 1.5)$. De acordo como $\mathbf{X}_{5000\times2}$ e $\boldsymbol{\beta}$ foram definidos, os valores de ψ assumem os valores: 0.993, 0.982, 0.971 e 0.924. Monitorou-se apenas os vetores de parâmetros $\boldsymbol{b}$ e $\boldsymbol{\beta}$. Foram geradas duas cadeias para cada parâmetro com a seguinte configuração: 1.000.000 iterações, *thin* 10 e *Burn in* de 500.000 iterações. Definiu-se esses dados gerados, por conveniência, de simulação 1. O tempo gasto na simulação foi de 19 dias. Os resultados se encontram na Tabela 2.2.

Observa-se pelos gráficos de diagnósticos, em anexo, um indício de convergência das cadeias. Observa-se que as 2 cadeias geradas para cada parâmetro oscilam em torno de uma média comum, sem apresentar tendências, com variância constante. Os testes de Geweke, mostrado na Figura 2.14 em anexo, também indicam que as cadeias convergiram.

Nota-se que as estimativas obtidas para os parâmetros de dificuldade são plausíveis, dado que o intervalo entre os quantis de ordem 97,5% e 2,5% contém o verdadeiro valor do parâmetro. Além disso, as medidas de posição se encontram bem próximas destes reais valores. Por outro lado, as estimativas dos parâmetros para a superdispersão não foram tão satisfatórias, apesar de o intervalo conter o valor real.

Tabela 2.2: Resultados do ajuste com 5.000 respondentes, 24 itens e 2 fatores de heterogeneidade. Simulação MCMC com 1.000.000 iterações do amostrador de Gibbs.

Parâmetros	Valor Real	Valor estimado						
		Estatísticas						
		Média	Desvio Padrão	2,5%	25%	50%	75%	97,5%
beta[1]	1	1.271	0.153	0.984	1.166	1.266	1.371	1.586
beta[2]	1.5	1.313	0.154	1.025	1.208	1.309	1.414	1.630
b[1]	-2	-2.033	0.053	-2.137	-2.068	-2.033	-1.997	-1.930
b[2]	-2	-1.986	0.052	-2.088	-2.020	-1.985	-1.950	-1.886
b[3]	-2	-1.942	0.051	-2.043	-1.976	-1.942	-1.908	-1.844
b[4]	-2	-1.976	0.052	-2.079	-2.010	-1.975	-1.941	-1.876
b[5]	-2	-1.937	0.051	-2.038	-1.971	-1.937	-1.902	-1.838
b[6]	-1	-0.960	0.040	-1.040	-0.987	-0.960	-0.933	-0.882
b[7]	-1	-1.003	0.040	-1.082	-1.030	-1.003	-0.976	-0.924
b[8]	-1	-1.002	0.040	-1.081	-1.029	-1.002	-0.974	-0.923
b[9]	-1	-0.972	0.040	-1.051	-0.998	-0.972	-0.945	-0.894
b[10]	-1	-0.991	0.040	-1.069	-1.017	-0.990	-0.964	-0.912
b[11]	0	0.035	0.036	-0.035	0.011	0.035	0.059	0.105
b[12]	0	0.013	0.036	-0.057	-0.011	0.013	0.037	0.083
b[13]	0	-0.020	0.036	-0.091	-0.044	-0.020	0.004	0.050
b[14]	0	0.028	0.036	-0.042	0.004	0.028	0.052	0.098
b[15]	0	0.074	0.036	0.004	0.050	0.074	0.098	0.144
b[16]	1	0.964	0.037	0.891	0.939	0.964	0.989	1.037
b[17]	1	1.002	0.038	0.929	0.977	1.002	1.028	1.076
b[18]	1	1.029	0.038	0.955	1.003	1.029	1.054	1.103
b[19]	1	1.024	0.038	0.951	0.999	1.024	1.050	1.098
b[20]	1	1.027	0.038	0.954	1.002	1.027	1.053	1.101
b[21]	2	2.070	0.046	1.981	2.039	2.070	2.100	2.159
b[22]	2	1.983	0.045	1.896	1.953	1.982	2.012	2.071
b[23]	2	2.037	0.045	1.949	2.007	2.037	2.068	2.127
b[24]	3	2.939	0.060	2.823	2.899	2.939	2.979	3.057

2.2 Comparação entre o modelo adaptado e o tradicional de Rasch

Comparou-se nesta seção os ajustes feitos pelo modelo adaptado para heterogeneidade com o modelo tradicional de Rasch. Foram utilizados os mesmos dados heterogêneos da simulação 1, seção anterior, para se estimar os parâmetros do modelo tradicional de Rasch. As curvas características dos 23 itens desse ajuste se encontram na Figura 2.4. A Tabela 2.3 mostra os valores estimados para os parâmetros de dificuldade dos dois modelos.

Com o modelo adaptado para heterogeneidade, um item qualquer passa a ter mais de uma curva característica, ou seja, tantas curvas quantos forem os níveis de ψ. Isto é, os grupos (clusters) que causam heterogeneidade nos dados passam a ser controlados. Com isso, apesar de o item possuir o mesmo nível de dificuldade para todos respondentes, as covariáveis dos mesmos fazem com que a probabilidade de acerto seja diferente entre os grupos. Note que, para dados heterogêneos, o modelo tradicional superestima a dificuldade em todos os itens, ver Tabela 2.3. Esse fato fica mais

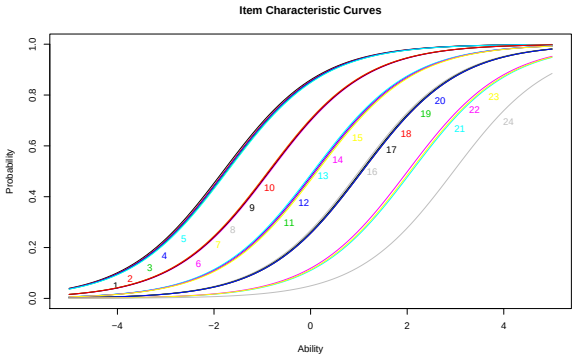


Figura 2.4: Gráfico das Curvas Características do Itens ajustadas pelo modelo tradicional de Rasch, dados da simulação 1.

evidente no próximo ajuste feito a partir dos dados simulados, com heterogeneidade acentuada. Pela Figura 2.5, observa-se as curvas características do item 1 do ajuste de ambos os modelos. É importante destacar o fato que ocorre no ajuste feito pelo modelo tradicional frente ao adaptado. Nota-se que, pelo ajuste do MACH, cada grupo possui um limite superior de probabilidade diferente na CCI. O Modelo tradicional superestima a dificuldade do item a fim de contemplar um ajuste que satisfaça todos os grupos existentes. Por outro lado, o MACH estima corretamente a dificuldade do item para todos os alunos, porém, penaliza na probabilidade de acerto de acordo com as características, uns mais e outros menos. Esse fato não foi tão evidente pois a heterogeneidade nos dados foi relativamente baixa.

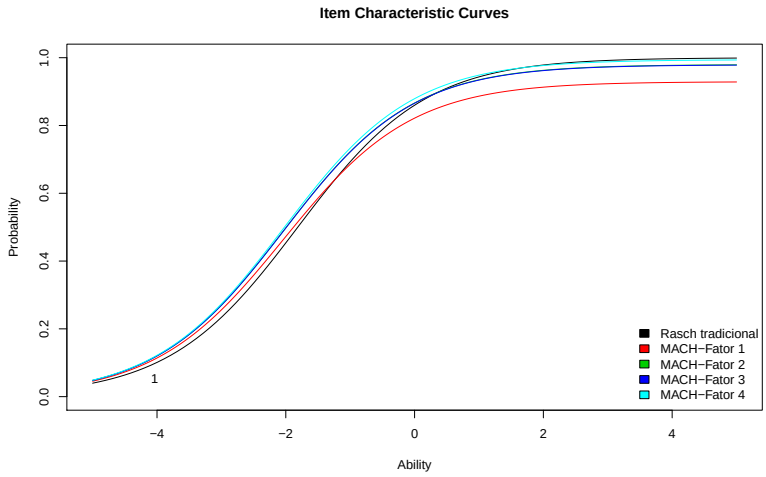


Figura 2.5: CCI 1 ajustada pelo modelo tradicional de Rasch e pelo MACH, dados da simulação 1.

Outro cenário

Tabela 2.3: Comparação entre o ajuste do modelo adaptado e o ajuste do modelo tradicional de Rasch. Dados da simulação 1. MCMC com 2 cadeias de 1.000.000 iterações do amostrador de Gibbs.

Parâmetros	Valor Real	Estimativas					Diferença (a-b)
		Modelo Adaptado (MCMC)			Modelo Tradicional		
		Média (a)	Mediana	DP	Estimativa (b)	EP	
beta[1]	1	1,271	1,266	0,153	-	-	
beta[2]	1,5	1,313	1,309	0,154	-	-	
b[1]	-2	-2,033	-2,033	0,053	-1,813	0,042	-0,220
b[2]	-2	-1,986	-1,985	0,052	-1,779	0,042	-0,207
b[3]	-2	-1,942	-1,942	0,051	-1,735	0,041	-0,207
b[4]	-2	-1,976	-1,975	0,052	-1,761	0,042	-0,215
b[5]	-2	-1,937	-1,937	0,051	-1,722	0,041	-0,215
b[6]	-1	-0,960	-0,960	0,040	-0,836	0,036	-0,124
b[7]	-1	-1,003	-1,003	0,040	-0,878	0,036	-0,125
b[8]	-1	-1,002	-1,002	0,040	-0,871	0,036	-0,131
b[9]	-1	-0,972	-0,972	0,040	-0,853	0,036	-0,119
b[10]	-1	-0,991	-0,990	0,040	-0,868	0,036	-0,123
b[11]	0	0,035	0,035	0,036	0,106	0,034	-0,071
b[12]	0	0,013	0,013	0,036	0,081	0,034	-0,068
b[13]	0	-0,020	-0,020	0,036	0,051	0,034	-0,071
b[14]	0	0,028	0,028	0,036	0,097	0,034	-0,069
b[15]	0	0,074	0,074	0,036	0,140	0,034	-0,066
b[16]	1	0,964	0,964	0,037	1,004	0,036	-0,040
b[17]	1	1,002	1,002	0,038	1,044	0,037	-0,042
b[18]	1	1,029	1,029	0,038	1,069	0,037	-0,040
b[19]	1	1,024	1,024	0,038	1,066	0,037	-0,042
b[20]	1	1,027	1,027	0,038	1,069	0,037	-0,042
b[21]	2	2,070	2,070	0,046	2,093	0,045	-0,023
b[22]	2	1,983	1,982	0,045	2,007	0,044	-0,024
b[23]	2	2,037	2,037	0,045	2,062	0,045	-0,025
b[24]	3	2,939	2,939	0,060	2,960	0,059	-0,021

Nota: DP e EP representam desvio padrão e erro padrão, respectivamente.

No cenário a seguir foi acentuada a heterogeneidade nos dados para se perceber com mais clareza o que acontece com os dois modelos. Foram consideradas nessa simulação as mesmas configurações dos dados do pré-teste, com exceção do número de respondentes, que passaram para 2.000, e dos parâmetros para superdispersão, assumindo $\boldsymbol{\beta} = (0.5, 1)$. De acordo como $\boldsymbol{\beta}$ e $\mathbf{X}_{2.000\times 2}$ foram definidos, os níveis de ψ passam a ser mais acentuados: 0.952, 0.924, 0.881 e 0.817. Denominou-se esses dados gerados por simulação 2. A Tabela 2.4 apresenta os resultados dos ajustes. A Figura 2.6 apresenta as cadeias dos parâmetros geradas pelo ajuste (via MCMC) do modelo adaptado.

Observa-se a partir das duas comparações feitas que o erro padrão das estimativas do modelo tradicional é menor que o desvio padrão das estimativas do modelo adaptado. Esse fato é mais perceptível na simulação 2, onde os dados são mais heterogêneos. Uma justificativa para esse fenômeno é que o modelo tradicional subestima a variabilidade existente nos dados heterogêneos (com superdisperção). Consequentemente, isso acarreta na inflação do erro do tipo I, obtendo, assim, valores de P arti-

Tabela 2.4: Comparação entre o ajuste do modelo adaptado e o ajuste do modelo tradicional de Rasch. Dados da simulação 2. MCMC com 100.000 iterações.

Parâmetros	Valor Real	Estimativas					Diferença (a-b)
		Modelo Adaptado			Modelo Tradicional		
		Média (a)	Mediana	DP	Estimativa (b)	EP	
beta[1]	0,5	0,437	0,437	0.153	-	-	-
beta[2]	1	1,048	1,045	0.154	-	-	-
b[1]	-2	-1,940	-1,997	0.053	-1,269	0,038	-0,666
b[2]	-1	-0,997	-0,997	0.052	-0,585	0,035	-0,412
b[3]	0	0,002	0,003	0.051	0,275	0,043	-0,273
b[4]	1	1,019	1,020	0.052	1,209	0,037	-0,190
b[5]	2	1,947	1,947	0.051	2,097	0,045	-0,150

Nota: DP e EP representam desvio padrão e erro padrão, respectivamente.

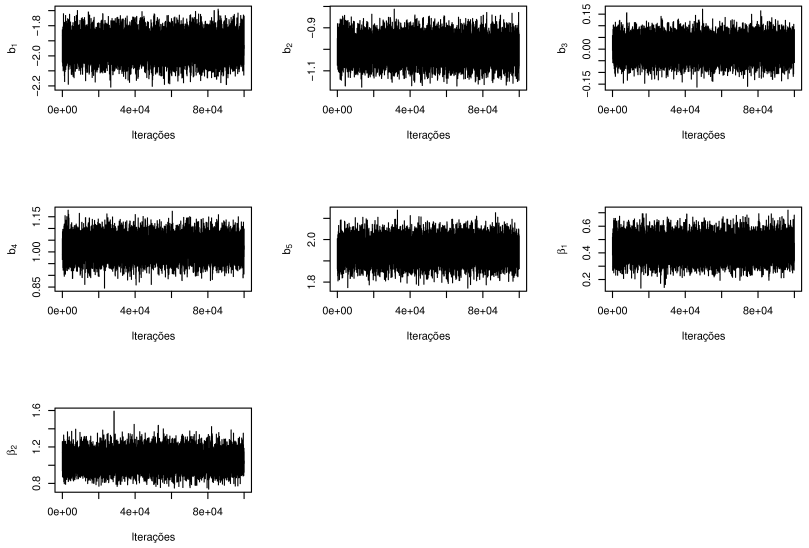


Figura 2.6: Gráfico de iterações do amostrador de Gibbs para os parâmetros de dificuldade e de superdispersão. Simulação 2.

ficialmente significativos (VIEIRA, 2008). Outro ponto que merece destaque é o fato do modelo tradicional apresentar maiores superestimativas para itens fáceis quando comparados com os mais difíceis.

Note que os fatores externos conhecidos que interferem na probabilidade de acerto do item entram no modelo apenas penalizando os grupos, uns mais e outros menos, conservando a dificuldade dos itens. Essa penalização é dada pelo efeito multiplicativo ψ, e serve de base na comparação entre grupos. Por outro lado, esse efeito também está controlando a heterogeneidade existente.

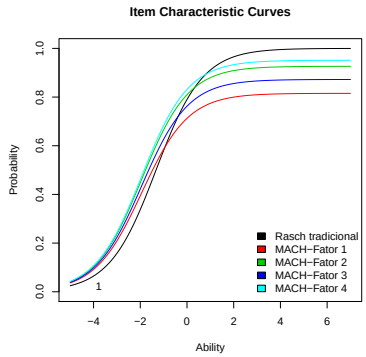


Figura 2.7: CCI 1 ajustada pelo modelo tradicional de Rasch e pelo MACH, dados da simulação 2.

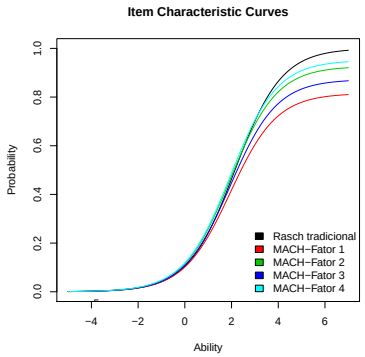


Figura 2.8: CCI 5 ajustada pelo modelo tradicional de Rasch e pelo MACH, dados da simulação 2.

Capítulo 3

Aplicação do modelo adaptado aos dados da Prova Brasil 2007

Neste capítulo foi feita a aplicação do modelo tradicional de Rasch e do adaptado para heterogeneidade aqui proposto, bem como a comparação entre os mesmos. A seguir, explanou-se sobre o banco de dados da Prova Brasil, e como foi extraída uma parte para realizar este estudo.

Sobre a Prova Brasil

A Prova Brasil é um exame nacional de larga escala aplicado a cada dois anos a alunos da Educação Básica, 4ª e 8ª série (quinto e nono anos) do Ensino Fundamental, das escolas públicas urbanas e rurais. É um dos instrumentos de avaliação do sistema educacional brasileiro do Inep/MEC (INEP, 2012). Segundo o MEC, a Prova Brasil tem objetivo de auxiliar nas tomadas de decisões do Governo Federal, Estadual e Municipal, na direção de escola, professores e comunidade escolar, em relação à busca da melhoria da qualidade de ensino. A avaliação dos alunos é feita sob a ótica de duas habilidades: língua portuguesa - associada a competências de leitura e interpretação de texto - e matemática, associada à competência em resolução de problemas. A partir dessas avaliações pode-se acompanhar ao longo do tempo a evolução das escolas, das redes e do sistema como um todo, além de subsidiar os cálculos do Índice de Desenvolvimento da Educação Básica (Ideb). Assim, é possível realizar ações voltadas para a correção de distorções e direcionar recursos técnicos e financeiros para as áreas

prioritárias, visando a redução das desigualdades nelas existentes, (INEP, 2012).

A Prova Brasil surgiu no ano de 2005, quando o Sistema de Avaliação da Educação Básica foi reestruturado. O SAEB dividiu-se em duas avaliações complementares: Avaliação Nacional da Educação Básica (Aneb) e a Avaliação Nacional do Rendimento Escolar (Anresc). A Anresc, conhecida como Prova Brasil, abrange de maneira censitária os estudantes das redes públicas e oferece resultados de cada escola participante, das redes no âmbito dos municípios, dos estados, das regiões e do Brasil. A Aneb, conhecida também como SAEB , abrange de maneira amostral os estudantes das redes públicas e privadas do país. Ela oferece resultados de desempenho apenas para o Brasil, regiões e unidades da Federação, ou seja, não há resultado por escola e por município.

Quando realizado a prova, os envolvidos na avaliação, como os alunos, professores e o diretor da escola, respondem a um questionário onde há perguntas sobre eles, sobre o ensino e sobre a escola. Há também o questionário socioeconômico em que os estudantes fornecem informações sobre fatores que podem estar associados ao desempenho. O propósito do questionário é construir um perfil das características desses envolvidos, e também possibilitar a investigação de fatores que podem interferir positivamente ou negativamente sobre o aprendizado dos alunos (INEP, 2012). O questionário do aluno, ver Figuras 3.23 e 3.24 em anexo, foi usado neste trabalho para identificar os fatores que causam heterogeneidade nos dados, e, assim, incorporá-los ao modelo adaptado, como segue a proposta.

Quanto à organização da prova, são feitos 21 tipos de cadernos de provas e cada aluno responde apenas um. Cada caderno possui quatro blocos de questões, dois de português e dois de matemática. As questões são de múltipla escolha, com quatro ou cinco alternativas de resposta. Para os alunos do 5° ano, o caderno de provas possui 22 itens de português e 22 itens de matemática. Já para os alunos do 9° ano, o caderno possui 26 itens para cada disciplina. O tempo para a realização da prova é 2 horas e 30 minutos. Os bancos de dados das avaliações passadas estão disponíveis gratuitamente no site do INEP (www.inep.gov.br).

Seleção de uma parte do banco de dados

A avaliação da Prova Brasil é quase que censitária. Cerca de 5,5 milhões de alunos fizeram a prova no ano de 2007, em mais de 50 mil escolas em todo o país. Para este estudo foi extraída uma parte do banco de dados da Prova Brasil 2007. Selecionou-se os alunos da 4ª série (quinto ano) da prova de Matemática que realizaram o caderno de provas número "1". Essa porção corresponde a 109.939 alunos para os 22 itens da prova de matemática, mesmo banco utilizado por Silva (2011) para controlar a heterogeneidade oriunda de fatores desconhecidos. Porém, essa quantidade de alunos (parâmetros de habilidade) ainda é problema de esforço real para se ajustar o modelo proposto aqui via MCMC. Desse modo, sorteou-se 5.000 alunos dessa parcela para compor os dados que serão utilizados nos ajustes a seguir.

3.1 Ajuste do modelo tradicional de Rasch

Realizou-se o ajuste usando o modelo tradicional de Rasch através do pacote *ltm* do ambiente computacional **R**. Primeiramente, foi feito o ajuste geral com todos os estudantes, sem estratificá-los por alguma característica. A Tabela 3.2, em anexo, mostra as estimativas obtidas. A Figura 3.1 apresenta as curvas características ajustadas pelo modelo de Rasch para os 22 itens da prova.

Note que esse ajuste foi realizado sem considerar possível existência de heterogeneidade nos dados. Observe que, se forem levadas em consideração as informações extras disponíveis no questionário do aluno, é possível investigar quais fatores interferem no seu desempenho. Se esse fenômeno ocorrer, então é possível que alunos (ou grupos de alunos) não respondem independentemente a prova, ou seja, os fatores alteram a probabilidade de acerto entre grupos, prejudicando, assim, o pressuposto da independência condicional.

Silva (2011) mostra que o banco de dados da Prova Brasil 2007 (parcela com 109.939 alunos da 4ª série submetidos ao caderno 1 da prova de matemática) apresenta heterogeneidade. A fim de comprovar esse fenômeno, considerou-se os itens "2", "35" e "38" do questionário do aluno para se ajustar modelos separados para cada nível de resposta dos respectivos itens. A seguir são apresentadas algumas estatísticas

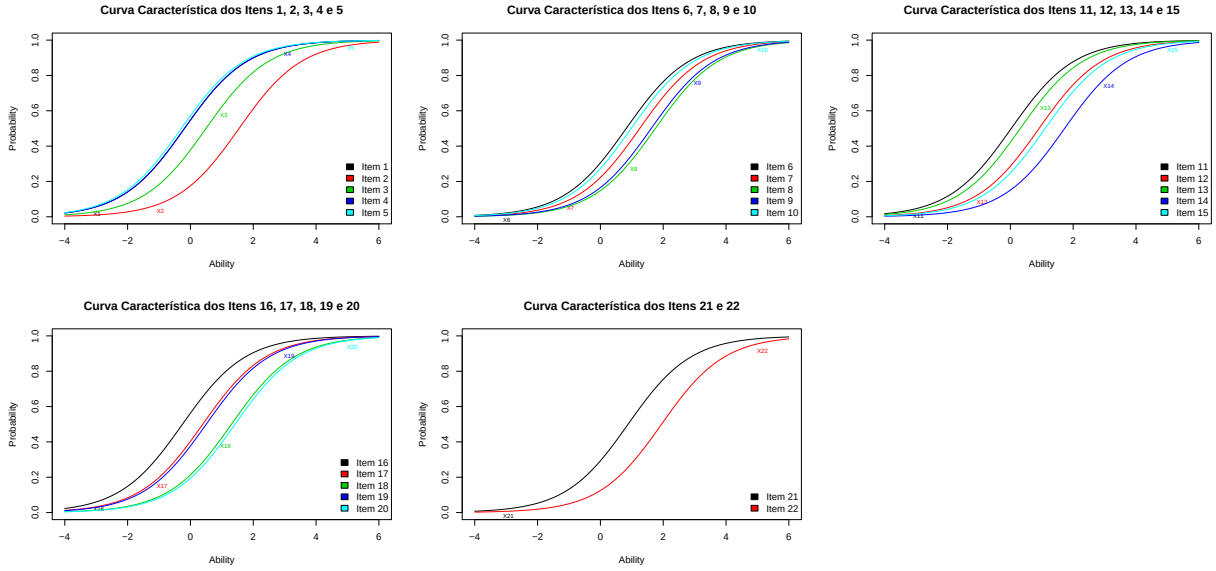


Figura 3.1: CCI ajustada pelo modelo tradicional de Rasch, dados da Prova Brasil 2007.

descritivas de cada nível de fator.

Tabela 3.1: Estatísticas descritivas dos fatores avaliados.

FATOR		ESTATÍSTICAS			
		Número de alunos	% geral de acerto	% geral de erro	Não responderam
Trab. fora de casa	Trabalha fora	768	0,304	0,696	474
	Não trabalha fora	3758	0,352	0,648	
Reprovação	Nenhuma	2969	0,365	0,636	495
	Pelo menos uma	1536	0,307	0,693	
Étnico-Racial	Branco e amarelo	1357	0,345	0,655	434
	Pardo	2535	0,353	0,647	
	Preto e índio	674	0,320	0,680	

Fator 1: Trabalhar fora de casa - Questão 35

Esse fator possui dois níveis: se o aluno trabalha fora de casa (x=1) ou não trabalha (x=0). Ajustou-se um modelo para ambos os extratos. Observa-se nas Tabelas 3.7 e 3.8, em anexo, que as estimativas de quase todos os parâmetros de dificuldade $\boldsymbol{b}$ são superiores para os alunos que trabalham fora casa, ou seja, quem não trabalha fora tem mais chance de acerto nos itens da prova. Nos gráficos das curvas características de alguns itens, esse fato é claramente notado, Figura 3.2. No entanto, o fator "Trabalhar fora de casa" prejudica o pressuposto de independência condicional e causa heterogeneidade. Devendo, assim, considerá-lo no modelo proposto.

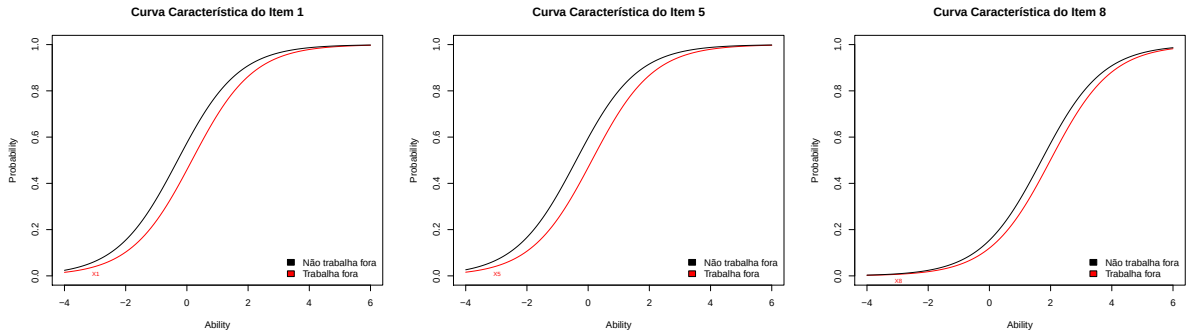


Figura 3.2: CCI ajustada pelo modelo tradicional de Rasch, comparação sobre trabalho fora de casa. Dados da Prova Brasil 2007.

Fator 2: Reprovação - Questão 38

O fator "Reprovação" possui originalmente 3 níveis: "o aluno não possui reprovação", "possui uma reprovação" e "mais de uma reprovação". As duas últimas categorias foram agregadas por possuírem ajustes parecidos, passando a compor a categoria "pelo menos uma reprovação". Assim, definiu-se: x=0 se o aluno não possui reprovação; e x=1 se o aluno possui pelo menos uma reprovação. É evidente também a interferência do fator na probabilidade de acerto dos itens. Nota-se nas Tabelas 3.5 e 3.6, em anexo, que as estimativas dos parâmetros de dificuldade $\boldsymbol{b}$ são superiores para os alunos que reprovaram pelo menos uma vez, com exceção do itens 15 e 17. Isto é, o fator causa heterogeneidade. Na Figura 3.3 é mostrado algumas curvas características que explicitam o fenômeno.

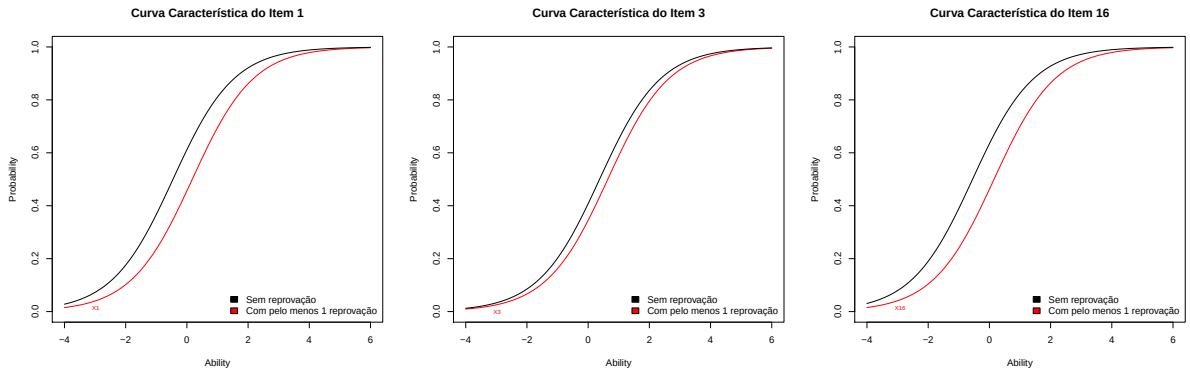


Figura 3.3: CCI ajustada pelo modelo tradicional de Rasch, comparação sobre reprovação. Dados da Prova Brasil 2007.

Fator 3: Étnico-Racial - Questão 2

Este fator possui 5 categorias: branco, pardo, preto, amarelo e índio. Dois pares dessas categorias foram agregados por dois motivos: primeiro, por possuírem ajustes similares, e, segundo, para reduzir esforço computacional do ajuste via MCMC do modelo proposto. Assim, este banco de dados passou a ter as três categorias: "branco e amarelo" (x=0), "pardo" (x=1) e "índio e preto" (x=2). Mais uma vez é evidente a perturbação que esses fatores geram nas probabilidades de acerto dos itens ao criarem certos grupos. As Tabelas 3.7, 3.8 e 3.9, em anexo, apresentam as estimativas dos parâmetros dos modelos. Pela Figura 3.4, nota-se que as estimativas dos parâmetros de dificuldade dos pardos são ligeiramente superiores que as dos considerados pretos e índios. Indicando que os itens são relativamente mais fáceis para aquele grupo.

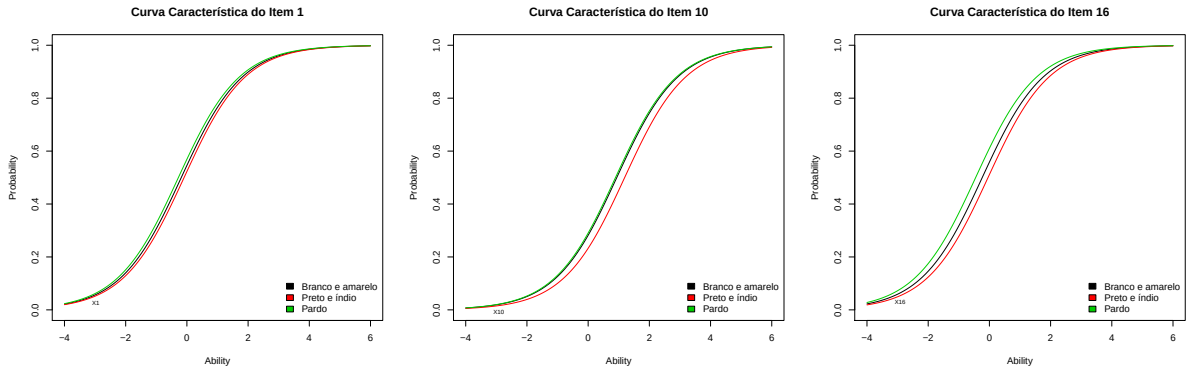


Figura 3.4: CCI ajustada pelo modelo tradicional de Rasch, comparação étnico-racial. Dados da Prova Brasil 2007.

3.2 Ajuste do modelo com controle da heterogeneidade atribuída a fatores conhecidos

O modelo adaptado para heterogeneidade, proposto neste trabalho, foi submetido ao ajuste para os dados da Prova Brasil 2007. Viu-se na seção anterior que os fatores: "trabalhar fora de casa", "reprovação" e "étnico-racial" são causadores potenciais de heterogeneidade. No entanto, esses fatores foram incorporados ao modelo como covariáveis do efeito ψ. Utilizou-se a idéia do procedimento de seleção Forward (DRAPER e SMITH, 1981) para obter melhor ajuste com base nessas covariáveis. A

idéia é ajustar modelos adicionando cada covariável por vez, e incorporá-las quando forem significativas. A seguir é apresentado ajustes para cada uma das covariáveis. Os mesmos foram realizados através do pacote computacional *R2WinBugs* do *software* **R** 2.15. As simulações MCMC tiveram 500.000 iterações, *burn in* de 250.000 e *thin* igual a 10 iterações, com 1 cadeia para cada parâmetro. Monitorou-se os parâmetros $\boldsymbol{b}$ e $\boldsymbol{\beta}$. O tempo gasto no processamento dos ajustes foi de 5 dias em um computador Intel Xeon 2.40 GHz - 2 processadores, 16 GB de memória, com sistema operacional Windows Server Standart 64 bit.

Fator 1: Trabalhar fora de casa

A Tabela 3.10, em anexo, mostra as estimativas obtidas do modelo. Os gráficos de diagnósticos indicam a convergência das cadeias de $\boldsymbol{\beta}$, ver Figura 3.5. Os parâmetros da superdispersão foram significativos de acordo com o intervalo de credibilidade, ver Figura 3.6. Comparando as curvas características dos itens, observa-se similar comportamento ao estudo de simulação, Figura 3.7. De acordo com o preditor linear do efeito estimado, $\text{logit}(\hat{\psi}) = 0,909 + 1,158x$, o grupo que "não trabalha fora" (x=1) possui acréscimo de 1,158 unidades, ou seja, o efeito ψ desse grupo é de 0,887 contra 0,713 do grupo que "trabalha fora", Figura 2.15. Sabe-se que esse efeito incorpora-se ao modelo tradicional de Rasch multiplicando-o. Com base nisso, pode-se dizer que o grupo que "trabalha fora de casa" ($\hat{\psi} = 0,713$) tem uma redução na probabilidade de acerto maior que no grupo que "não trabalha fora", pois o $\hat{\psi}$ desse foi de 0,887, que reduz pouco a probabilidade em relação ao outro efeito. Assim, considerando estes dois valores (0,713; 0,887) para efeito de comparação entre grupos, é possível dizer que o desempenho do grupo que não trabalha fora de casa é 24% $(0,887/0,713 = 1,24)$ maior que o desempenho do grupo que trabalha fora. Note que, além de controlar a heterogeneidade causada pelo fator em estudo, é possível utilizar ψ_j para comparar os níveis desse mesmo. Outra observação que merece destaque é o fato das estimativas das dificuldades dos itens serem menores e com desvios padrões maiores quando comparadas com as estimativas do modelo tradicional. Esse fenômeno também ocorreu no estudo de simulação, indicando possivelmente a captação de maior variabilidade nas estimativas.

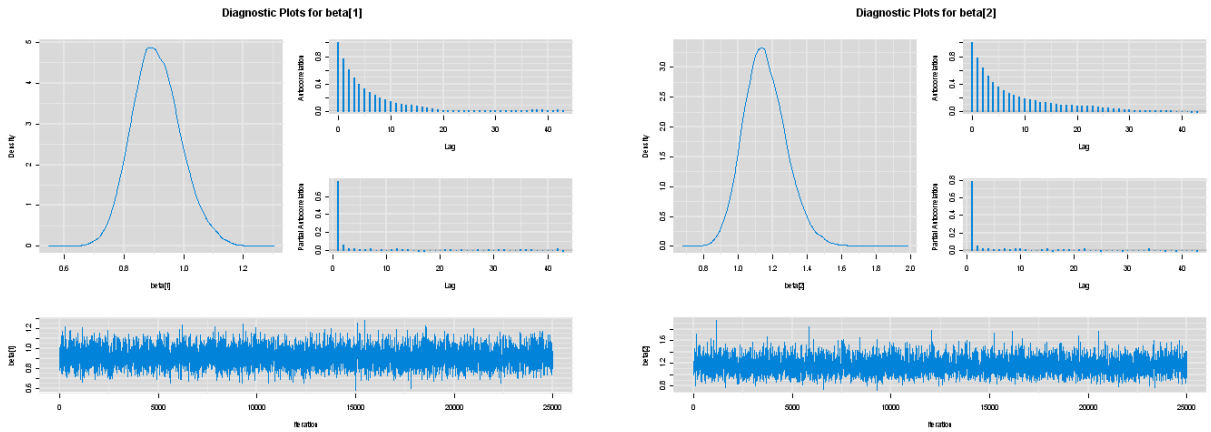


Figura 3.5: Gráficos de diagnóstico das cadeias dos parâmetros β_1 e β_2 do ajuste ponderado pelo fator "Trabalha fora de casa", simulação MCMC.

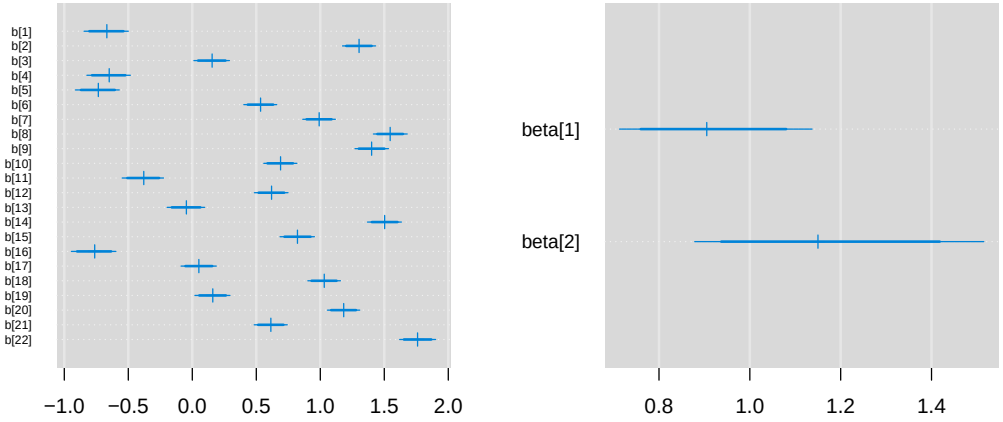


Figura 3.6: Intervalos de credibilidade dos parâmetros ajustados, ao nível de 95% e 99%. Modelo ponderado pelo fator "Trabalha fora de casa", simulação MCMC.

Fator 2: Reprovação

Os gráficos de diagnósticos não indicam satisfatória convergência das cadeias de $\boldsymbol{\beta}$, o que levou a utilizar um *thin* adicional de 10 iterações, melhorando a independência entre os valores amostrados, sendo possível, assim, fazer inferências sobre os mesmos, ver Figura 3.8. Pela Figura 3.10, observa-se que ajuste feito somente com o fator reprovação também é significativo em relação aos parâmetros da superdispersão. A Tabela 3.11 em anexo mostra as estimativas obtidas. Neste caso, os parâmetros estimados do preditor linear foram: $\text{logit}(\hat{\psi}) = 2,099 - 1,280x$, lembrando que adotou-

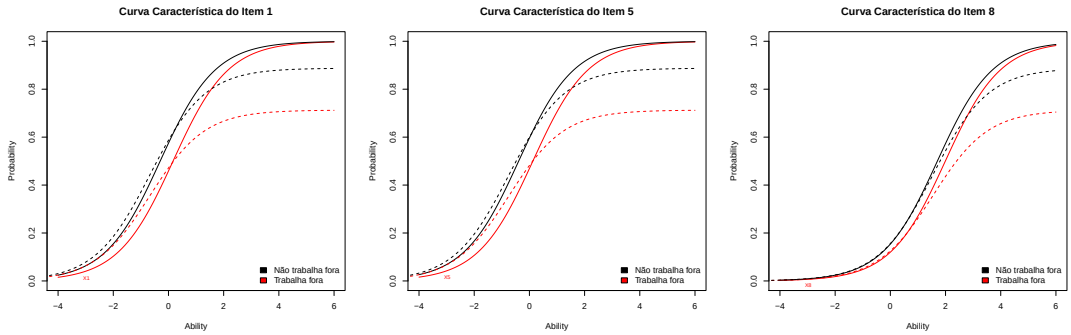


Figura 3.7: Curvas características dos itens 1, 5 e 8 ajustada pelo modelo tradicional e adaptado, pelas linhas contínuas e tracejadas, respectivamente - controlando o fator "Trabalha fora de casa".

se no ajuste a notação x=0 se o aluno não foi reprovado e x=1 se o aluno que foi reprovado pelo menos 1 vez. Assim, quem possui pelo menos uma reprovação obteve $\hat{\psi} = 0,694$, e para quem nunca foi reprovado o $\hat{\psi}$ foi de 0,891. Ou seja, como explicado anteriormente, o desempenho de quem nunca foi reprovado é 28% (0,891/0,694=1,28) maior em relação aqueles que reprovaram pelo menos uma vez. A Figura 3.10 mostra as curvas características de alguns itens deste ajuste, comparando com o modelo tradicional.

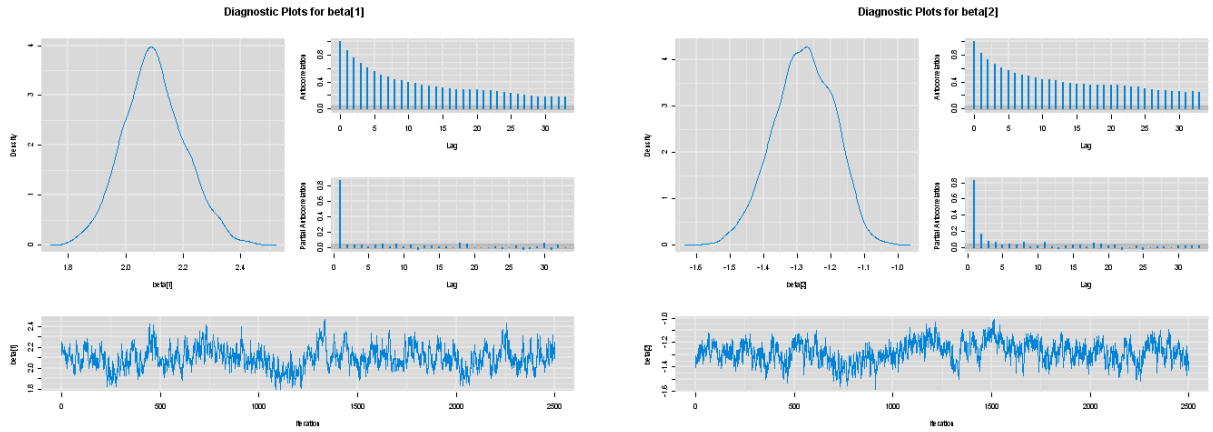


Figura 3.8: Gráficos de diagnóstico das cadeias dos parâmetros β_1 e β_2 do ajuste sob o fator "Reprovação", simulação MCMC.

Fator 3: Étnico-Racial

Neste ajuste as cadeias dos parâmetros simulados, via MCMC, atingiram satisfatória convergência. Os gráficos de diagnósticos encontram-se na Figura 3.11. Os

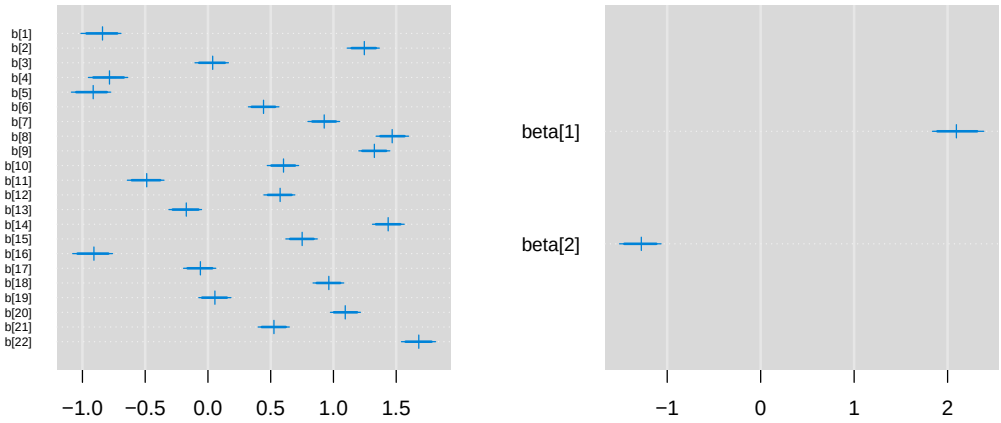


Figura 3.9: Intervalos de credibilidade dos parâmetros ajustados, ao nível de 95% e 99%. Modelo ponderado pelo fator "reprovação", simulação MCMC.

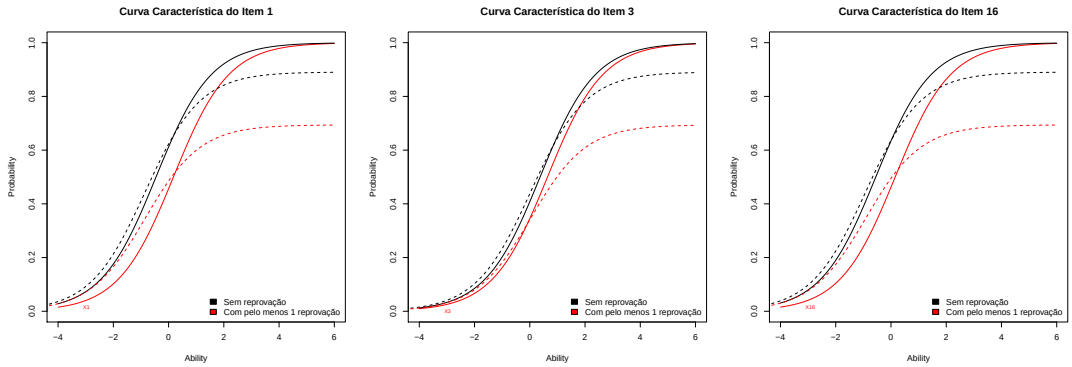


Figura 3.10: Curvas características dos itens 1, 3 e 16 ajustada pelo modelo tradicional e adaptado, pelas linhas contínuas e tracejadas, respectivamente - controlando o fator "reprovação".

parâmetros do preditor linear do efeito ψ foram significativos, de acordo com os intervalos de credibilidades apresentados na Figura 3.12. Aqui, como o fator possui 3 níveis, a matriz X de covariáveis foi formulada da seguinte forma: a primeira coluna com todos os valores iguais a 1; a segunda coluna, x_1, assumindo valor 1 se o aluno é classificado com "pardo" e 0 caso contrário; a terceira coluna, x_2, assumindo valor 1 se o aluno é considerado "preto ou índio" e 0 caso contrário. Desse modo, o grupo de referência para efeito de comparação é o "branco ou amarelo'. O efeito multiplicativo estimado foi de $\text{logit}(\hat{\psi}) = 1,732 + 0,276x_1 - 0,456x_2$. Ou seja, o grupo "branco ou amarelo" possui $\hat{\psi} = 0,850$, o grupo "pardo" em relação a esse

possui $\hat{\psi} = 0,881$ e o grupo "preto ou índio", também em relação àquele, possui $\hat{\psi} = 0,782$. Contudo, pode-se dizer que o grupo "pardo" tem aumento de 3,7% $(0,881/0,849 = 1,037)$ no desempenho em relação ao grupo "branco e amarelo". Por outro lado, o grupo "branco e amarelo" apresenta aumento de 8,7% $(0,850/0,782 = 1,087)$ no desempenho quando comparado com o grupo "preto e índio". Pelas curvas características dos itens, Figura 3.13, observa-se as diferenças entre os grupos. É notório o fato de existir diferentes desempenhos entre os grupos nos diferentes itens, como é visto no ajuste pelo modelo tradicional para cada grupo. Porém, o modelo adaptado estima essa diferença entre grupos com base em todos os itens. Isto é, o efeito existente em cada grupo é igual para todos os itens. É importante frisar que o modelo supõe a existência de heterogeneidade, e não o funcionamento diferencial do item - DIF, ver Adriola (2001).

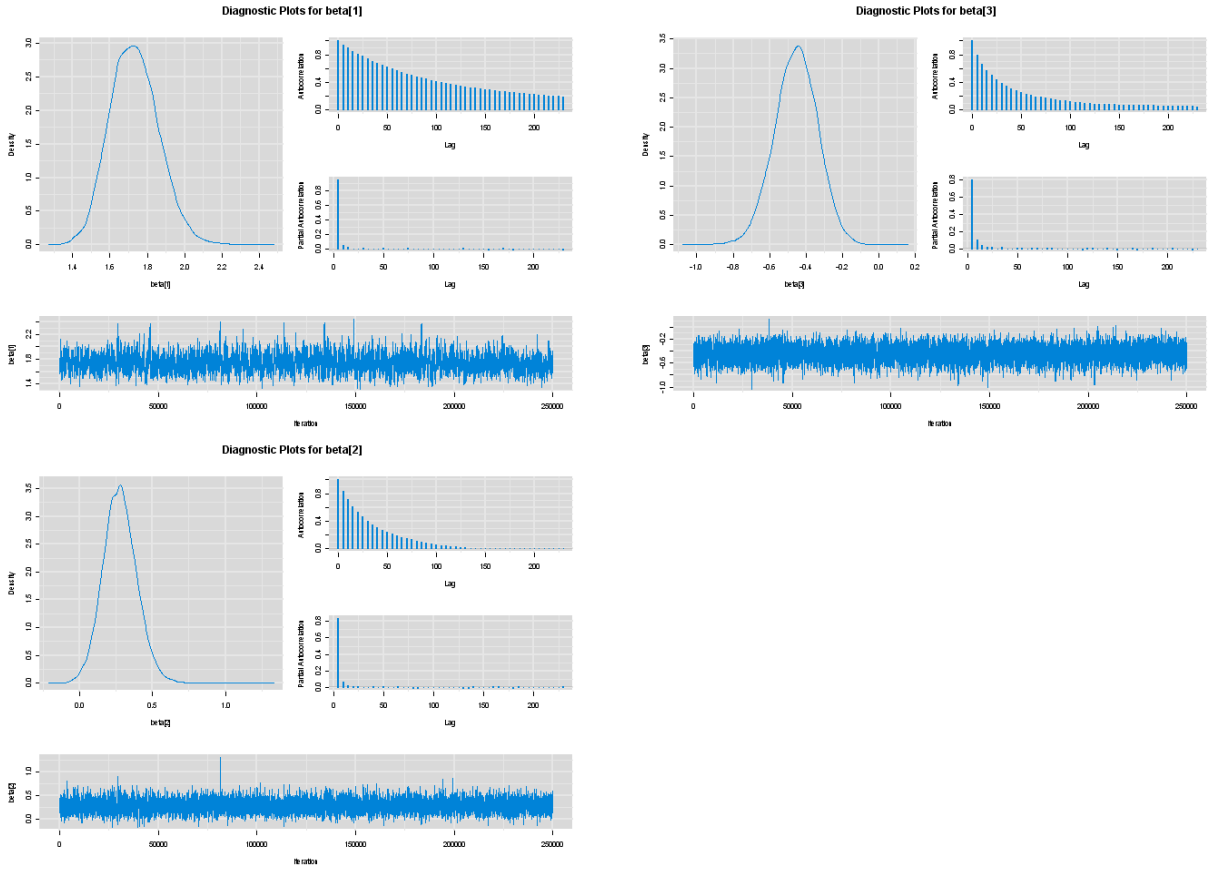


Figura 3.11: Gráficos de diagnóstico das cadeias dos parâmetros β_1, β_2 e β_3 do ajuste sob o fator "étnico-racial", simulação MCMC.

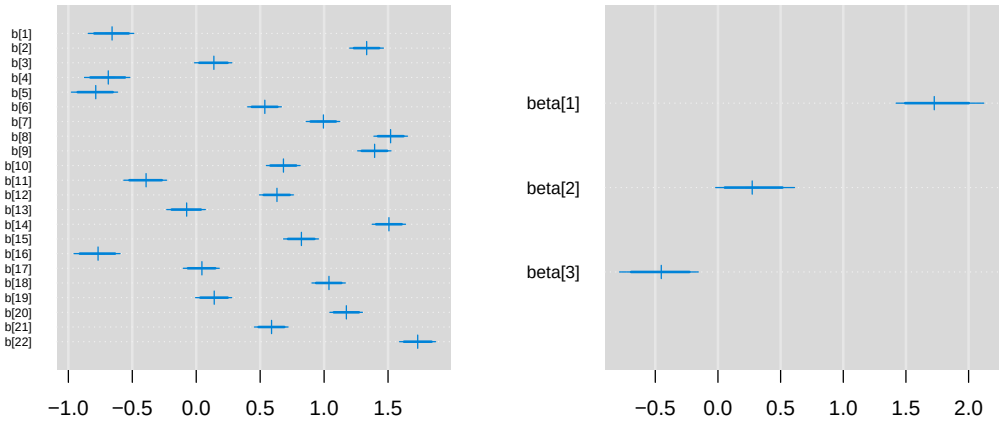


Figura 3.12: Intervalos de credibilidade dos parâmetros ajustados, ao nível de 95% e 99%. Modelo ponderado pelo fator "étnico-racial", simulação MCMC.

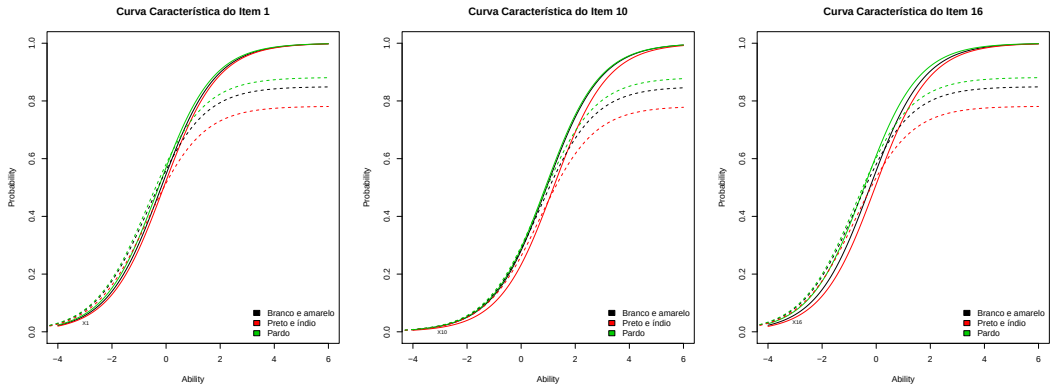


Figura 3.13: Curvas características dos itens 1, 10 e 16 ajustada pelo modelo tradicional e adaptado, pelas linhas contínuas e tracejadas, respectivamente - controlando o fator "étinco-racial".

Fatores: trabalha fora de casa × reprovação

Ajustou-se aqui o modelo controlando os fatores "trabalha fora de casa" e "reprovação", bem como a interação entre os mesmos, pois ambos foram significativos individualmente. As cadeias foram melhoradas quanto à independência dos valores amostrados adicionado *thin* de 10 iterações, porém não houve grandes mudanças nos valores estimados, Figura 3.14. Nota-se que as circunstâncias que prejudicam o desempenho do aluno como reprovação ou trabalhar fora de casa são potencializadas quando ambas ocorrem simultaneamente, ver Figura 3.15. O efeito ψ estimado foi de

$\text{logit}(\hat{\psi}) = 1,455 + 0,497x_1 - 1,055x_2$, em que x_1 é a variável *dummy* representando o aluno que não trabalha fora por valor 1 e 0 caso contrário; e x_2 a variável *dummy* para o fator reprovação, com valor 1 para quem tem pelo menos uma reprovação e 0 caso contrário. Assim, o grupo de referência é o que trabalha fora e não tem reprovação. Com isso, tem-se que para o grupo sem reprovação, não trabalhar aumenta o desempenho em 18,6%. Para o grupo que trabalha fora, ter reprovação reduz o desempenho em 18,8%. Por fim, para o grupo que trabalha fora e não tem reprovação, não trabalhar fora e ter pelo menos 1 reprovação reduz o desempenho em 0,2%.

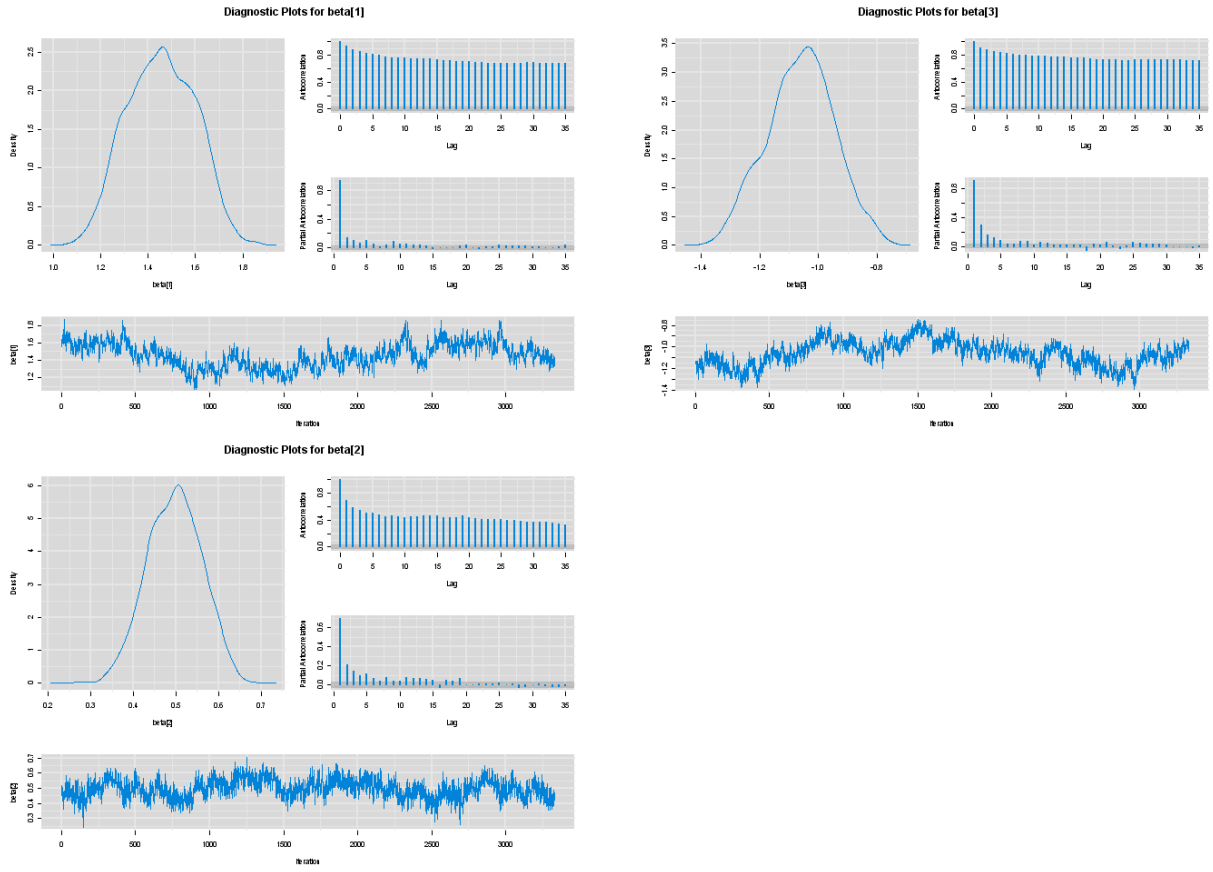


Figura 3.14: Gráficos de diagnóstico das cadeias dos parâmetros β_1, β_2 e β_3 do ajuste sob os fatores "reprovação e trabalho fora de casa", simulação MCMC.

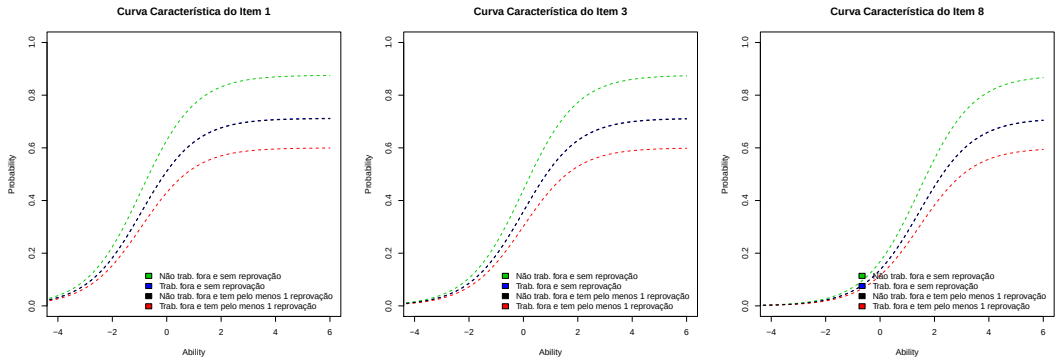


Figura 3.15: Curvas características dos itens 1, 3 e 8 ajustada pelo modelo adaptado - controlando os fatores “reprovação e trabalho fora de casa”.

Conclusão

O objetivo deste trabalho foi apresentar uma proposta de modelo de resposta ao item com controle da heterogeneidade. O modelo proposto incorpora ao tradicional de Rasch um efeito multiplicativo cujo preditor linear contempla características de indivíduos que são potenciais geradoras de grupos (*clusters*). Esta proposta segue uma linha paralela a que foi apresentada em Silva (2012), que propõe modelos de TRI controlando a heterogeneidade oriunda de fatores desconhecidos.

O processo de estimação adotado para o modelo foi baseado no método bayesiano. Primeiramente, pensou-se em utilizar a abordagem do algoritmo de Metropolis usando o ambiente computacional R, porém, foi inviável trabalhar com tal método, pois o modelo incorpora uma grande quantidade de parâmetros e causa intenso esforço computacional. Devido a esse problema, trabalhou-se com o software WinBugs através do R pelo pacote R2WinBUGS, no qual pode-se monitorar apenas as cadeias de interesse, $\boldsymbol{b}$ e $\boldsymbol{\beta}$. Com isso, vários ajustes foram realizados a partir de dados simulados a fim de estudar o comportamento do modelo. Verificou-se que o mesmo estimou satisfatoriamente os parâmetros definidos no gerador de amostras heterogêneas. Isto é, os intervalos de credibilidades dos ajustes contemplam os verdadeiros parâmetros da amostra. Verificou-se também que o modelo tradicional apresentou certas deficiências quando se comparou com o adaptado para dados heterogêneos. Os parâmetros de dificuldades foram superestimados por aquele modelo, principalmente para itens mais fáceis. Além disso, observou-se a partir das comparações feitas que os erros padrões das estimativas do modelo tradicional foram menores que os desvios padrões das estimativas do modelo adaptado. Esse fato foi mais perceptível na Simulação 2 do capítulo 2, onde os dados são mais heterogêneos. Uma justificativa para esse fenômeno é que o modelo tradicional subestima a variabilidade existente nos dados heterogêneos

(com superdispersão). Consequentemente, isso acarreta na inflação do erro do tipo I, obtendo, assim, valores de P artificialmente significativos (VIEIRA, 2008).

A aplicação do modelo proposto nos dados do SAEB - Prova Brasil 2007 foi uma das motivações do trabalho, visto que esse banco de dados apresenta questionários com características de alunos que podem ser utilizadas como fatores no modelo. Ajustou-se individualmente um modelo para cada fator analisado, a saber: se trabalha ou não fora de casa, se possui ou não reprovação e fatores étnico-raciais, "branco ou amarelo", "pardo" ou "preto ou índio". Em seguida um ajuste foi realizado com os fatores "trabalho fora de casa" e "reprovação", simultaneamente. A comparação entre os ajustes do modelo proposto e do tradicional de Rasch para esses dados também apresentou comportamento similar à comparação feita no estudo de simulação. No modelo tradicional, os parâmetros de dificuldades dos itens foram superestimados e os desvios padrões foram menores em relação aos obtidos pelo adaptado.

O modelo proposto conseguiu atingir seu objetivo de controlar a heterogeneidade, retornando, consequentemente, estimativas bem melhores que o tradicional. Além disso, foi possível também estudar descritivamente a comparação entre os níveis dos fatores através do efeito para cada grupo. Esse estudo indica aplicabilidade na fase de calibração dos itens, pois se pode entender ou até mesmo tentar criar itens que sofram menos efeitos da heterogeneidade intrínsecos a população.

Apesar das vantagens do modelo proposto, o mesmo apresenta certa limitação no que diz respeito ao esforço computacional empregado no ajuste. Com o aumento da amostra em mais de 5.000 indivíduos e no número de covariáveis, acima de 3, o tempo no processo cresce exponencialmente passando de mais de cinco dias, com média de 500.000 iterações. Assim, uma possibilidade de expansão desse trabalho seria adotar a metodologia da inferência bayesiana aproximada para modelos latentes Gaussianos usando integrais aproximadas de Laplace - INLA, proposta por Rue et al. (2009), metodologia que pode tornar o processo de estimação mais viável em tempo e eficácia.

Referências Bibliográficas

[1] AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6,p. 716-723, 1974.

[2] ANDERSEN, E.B. Conditional inference in multiple choice questionnaires. **British Journal of Mathematical and Statistical Psychology**, v. 26, p. 31-44, 1973.

[3] ANDERSEN, E.B. Discrete Statistical Models with Social Science Applications. **North-Holand Publishing Company**, New York, 1980.

[4] ANDRADE, D.F.; KLEIN, R. Métodos estatísticos para avaliação educacional: teoria da resposta ao item. Boletim da ABE, v. 43, p. 21-28, 1999.

[5] ANDRADE, D.F.; TAVARES, H. R.; VALLE, R. D. C. **Teoria de resposta ao item: conceitos e aplicações**. São Paulo: Associação Brasileira de Estatística, 2000.

[6] ADRIOLA, W. B. Descrição dos principais métodos para detectar o funcionamento diferencial dos itens (DIF). **Psicologia: Reflexão e Crítica**, v. 14, p. 643-652, 2001.

[7] AZEVEDO, C.L.N. **Métodos de Estimação na Teoria de Resposta ao Item**. Dissertação(Mestrado em Estatística). Universidade de São Paulo - IME/USP, São Paulo, 2003.

[8] BOCK, R.D.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm. **Psychometrika**, v. 45, p. 433-459, 1981.

[9] BOCK, R.D.; LIEBERMAN, M. Fitting a response model for n dichotomously scored items. **Psychometrika**, v. 35, p. 179-197, 1970.

[10] BOCK, R.D.; ZIMOWSKI, M.F. Multiple Grup IRT. In **Handbook of Modern Item Response Theory**. W.J. van der Linden and. R. K. Hambleton Eds. New York: Spring-Verlag, 1997.

[11] BRESLOW, N.E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, v. 88, n. 421, p. 9-25, 1993.

[12] BRITO, E.; CORDEIRO, G.M.; DEMÉTRIO, C.G. Distribuições em séries de potências modificadas **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 4, p. 1-20, 2010.

[13] BROOKS, S. P.; ROBERTS, G. O. On Quantile Estimation and Markov Chain Monte Carlo Convergence. **Biometrika**, v. 86, n. 3, p. 710-717, 1999.

[14] BURNHAM, K. P.; ANDERSON, D.R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2 ed. **Springer-Verlag**, 2002.

[15] CARLIN, B. P.; LOUIS, T. A. **Bayes and Empirical Bayes Methods for Data Analysis.** 2 ed. London: Chapman & Hall, 2000.

[16] CASTRO. S. M. **Teoria da Resposta ao Item: aplicação na avaliação da intensidade de sintomas depressivos.** Tese de doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Epidemiologia. Porto Alegre. 2008.

[17] COLLETT, D. **Modelling Binary Data.** London: Chapman & Hall. 1991, 369p.

[18] CORDEIRO, G.M. **Modelos Lineares Generalizados**. Universidade Estadual de Campinas - IMECC, Campinas, p. 286, 1986.

[19] COWLES, M.K.; CARLIN, B.P.Convergence Diagnostics: A Comparative Review. **Journal of the American Statistical Association,** v.91, n. 434, p. 883-904, 1996.

[20] CÚRI, M. **Análise de questionários com itens constrangedores**. Tese de doutorado (Doutorado em Estatística). Universidade de São Paulo - IME/USP, São Paulo, 2006.

[21] DEMÉTRIO, C.G. **Modelos Lineares Generalizados em Experimentação Agronômica**. Universidade de FSão Paulo - ESALQ/USP, São Paulo, 2002.

[22] DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihoodfrom incomplete data via the EM algorithm (with discussion). **Journal of the Royal Statistical Society Series B**, v. 39, p. 1-38, 1977.

[23] DOBSON, A.J. **An introduction to generalized linear models.** 2.ed. London: Chapman & Hall. 2001, 225p.

[54] DRAPER, N.; SMITH, H. Applied regression analysis. 2 ed. **John Wiley & Sons**, New York, 709 p, 1981.

[25] GELFAND, A.E.; SMITH, A.F. Sampling-based Approaches to Calculation Marginal Densities, **Journal of the American Statistical Association**, v. 85, p. 398-409, 1990.

[26] GELMAN, A.; STERN, J.; RUBIN, H. **Bayesian Data Analysis.** 2.ed. London: Chapman & Hall. 2004.

[27] GELMAM, A.; RUBIN, D.B.Inference from iterative simulation using multiple sequences. **Statistical Science,** n. 7, p. 457-511, 1992.

[28] GEMAN, S.; GEMAN D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 6 n. 6, p. 721-741, 1984.

[29] GEWEKE, J. Evaluating the accuracy of sampling-basead aproaches to the calculation of posterior moments. **Bayesian Statistics,** Oxford, UK: Oxford University Press, p. 169-193, 1992.

[30] GILKS, W.R.; RICHARDSON, S.; SPIEGELHALTER, D. **Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics.** Chapman & Hall/CRC Interdisciplinary Statistics, 1996.

[31] HEIDELBERGER, P.; WELCH, P.D. Simulation run length control in the presence of an initial trasient. **Operations Research,** n. 31, p. 97-109, 1983.

[32] HOFF, P. **A First Course in Bayesian Statistical Methods.** New York: Springer. 2009.

[33] HASTINGS, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. **Biometrika**, v. 57, n. 1, p. 97-109, 1970.

[34] JEFFREYS, H. (1961). **Theory of Probability.** 3 ed. Oxford Univ, 1961.

[35] KADANE, J.B.; LAZAR, N. Methods and Criteria for Model Selection. **JASA**, v. 99, n. 465, p. 279-290, 2004.

[36] LEE, Y.; NELDER, J.A. Hierarchical generalized linear models. With discussion. **Journal of Statistical Society, Series B**, v. 58, p. 619-678, 1996.

[37] LORD, F.M. A theory of test scores. **Iowa City**: Psychometric Monograph n. 7, 1952. Psychometric Society.

[38] McCULLGH, P.; NELDER, J.A. **Generalized Linear Models.** 2.ed. London: Chapman & Hall. 1989, 511p.

[39] McCULLOCH, C. E.; SEARLE, S. R. **Generalized, Linear and Mixed Models**. New York: Wiley-Interscience, 2001.

[40] McGILCHRIST, C.A. Estimation in generalized mixed models, **Journal of the Royal Statistical Society B**, v.56, p. 61-69, 1994.

[41] MESBAH, M., COLE, B.F.; LEE, M. L. T. **Statistical methods for quality of life studies: design, measurements and analysis.** Boston: Kluwer Academic Publishers (2002).

[42] METROPOLIS, N; ROSENBLUTH, A.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. **Equation of State Calculations by Fast Computing Machines.** Journal of Chemical Physics, v.21, p.1087-1092, 1953.

[43] MIGON, H.S.; GAMERMAN, D. **Statistical Inference: An Integrated Approach.** Londres: Ed. Edward Arnold, 1999.

[44] MOLENBERGHS, G.; VERBEKE, G.; DEMÉTRIO, C.G.B.; VIEIRA, A.M.C. A family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects. **Statistical Science**. v. 25, n. 3, p. 325-347, 2010.

[45] NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models, **Journal of the Royal Statistical Society A**, v. 74, p. 221-232, 1972.

[46] PAULA, G.A. **Modelos de Regressão com Apoio Computacional**. Universidade de São Paulo - IME/USP, São Paulo, 2004.

[47] PASQUALI, L. **Psicometria - Teoria dos testes na psicologia e na educação.** 3.ed. Patrópolis, RJ: Vozes; 2009.

[48] PIEPHO; H.P. Analysing disease incidence data from designed experiments by generalized linear mixed models. **Plant Pathology**, v.48 (5), p. 668-674, 1999.

[49] RAFTERY, A.E.; LEWIS, S. How many iterations in the Gibbs sampler?. **Bayesian Statistics,** Oxford, U.K.: Osxford University, p. 763-773, 1992.

[50] RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**, Copenhagen: Danish, University of Chicago Press, 1980.

[51] R DEVELOPMENT CORE TEAM. **R:** a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2008. Disponível em: <http://www.R-projet.org.>. Acesso em: 27 mar. 2012.

[52] RECKASE, M. **Muldimensional Item Response Theory.** Springer: New York. Institute for Educational Research, 2009.

[53] RESENDE, C.M.C. **Inferência Bayesiana Aproximada em Modelos de Espaço de Estados**. Dissertação de mestrado (Mestrado em Estatística). Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

[54] RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. **J. R. Statistical Society B** v. 71, p. 319-392, 2009)

[55] SCHWARZ, G. E. Estimating the dimension of a model. **Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

[56] SHALL, R. Estimation in generalized linear models with random effects. **Biometrika**, v. 78, p. 719-727, 1991.

[57] SILVA, F.P. **Uma adaptação do modelo de resposta ao item para mensuração de heterogeneidade atribuída à fonte desconhecida**. Dissertação de mestrado (Mestrado em Estatística). Universidade de Brasília, Brasília, 2012.

[58] SPIEGELHALTER, D.; BEST, N.; CARLIN, B.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society, Series B**, v. 64, p. 583-639, 2002.

[59] VIEIRA, A. M. C. **Modelagem simultânea de média e dispersão e aplicações na pesquisa agronômina.** Piracicaba: ESALQ, Universidade de São Paulo. Tese de Doutoramento. 2008, 117 p.

Anexos A

```
## Carregando pacotes
library(R2WinBUGS)
library(mcmcplots)
library(MCMCpack)
library(LearnBayes)
library(gibbs.met)
library(coda)

#################### GERA AMOSTRA

n=5000;mu=0       # n: Número de respondentes; mu: Média do traço latente.
b=c(rep(-2,5),rep(-1,5),rep(0,5),rep(1,5),rep(2,3),3)  # b: Vetor de parâmetros de dificuldade dos itens.
Beta=c(1,1.5)     # Beta: Vetor de parâmetros de superdispersão.
theta=rnorm(n,mu) # Gerando amostra das habilidades/proficiências.
X=matrix(c(sample(c(1,2),n,"T"),sample(c(1,2),n,"T")),ncol=2) # X: Matriz de covariáveis.
logit=function(x){   # Definindo a função logit.
(l=1/(1+exp(-x)))}
psi=logit(X%*%Beta)  # Definido psi.
y1=matrix(,ncol=24,nrow=n)  # Definindo a matriz de dados (respostas).
for(i in 1:24){             # Para cada item gera uma amotra para os n respondentes.
y1[,i]=rbinom(n,1,p=logit(theta-b[i])*psi)}
plot(psi)             # Ver níveis de psi.

#####  Estimação pelo pacote R2WinBUGS via WinBUgs

model.file2 <- system.file(package="R2WinBUGS", "model", "modelo.txt") # lendo o modelo
file.show(model.file2) #ver modelo

####  Definição Modelo
#model{for(j in 1:n){
#logit(psi[j])<- (inprod(beta[],X[j,]))
#for(i in 1:It){
#        Y[j,i]~dbern(par[j,i])
#                logit(p[j,i])<- (theta[j]-b[i])
#                par[j,i]<-psi[j]*p[j,i]
#        }}
#for(i in 1:It){b[i]~dnorm(0,1)}
#for(j in 1:n){theta[j]~dnorm(0,1)}
#for(k in 1:c){beta[k]~dnorm(1,1)}
#}

data2<-list(n=n,It=length(b),Y=y1,X=X,c=2) # entrando com os dados

parameters2  <- c("beta","b") # parametros a serem monitorados

sim <- bugs(data2, inits=NULL, parameters2, model.file2,
n.chains=2, n.iter=1000000, n.thin=10,
bugs.directory="c:/Arquivos de programas/WinBUGS14/",debug=T)

#save(sim,file="resu.RData")

load("resu.RData")#Precisa mudar o diretório de trabalho para o local do arquivo resu.RData

summary(sim)
print(sim)
plot(sim)
```

```
caterplot(sim, "b")
caterplot(sim, "beta")
traplot(sim, "b")
traplot(sim, "beta")
denplot(sim, "b")
denplot(sim, "beta")
mcmcplot(sim)
mc=as.mcmc(sim)
plot(mc)
raftery.diag(mc)
attributes(mc)
geweke.diag(mc)
geweke.plot(mc)

##############################################################################################

#### Estimação pelo pacote library(MCMCpack)

################ Log da função a posteriori Geral - com b, Beta e Theta desconhecidos.
################

funpost=function(THETA,x,y,N){
b=THETA[1:5];Beta=THETA[6:7];theta=THETA[8:(7+N)]
 sigb=1;sigB=1;l=numeric();        ###
 for(i in 1:5){
 l[i]=sum((1-y[,i])*log(1+exp(theta-b[i])+exp(X%*%Beta))-log((1+exp(theta-b[i]))*(1+exp(X%*%Beta))))}
 R=sum(theta%*%y)-sum(y%*%b)+sum(t(X%*%Beta)%*%y)-.5*theta%*%theta-.5*b%*%b/sigb-.5*Beta%*%Beta/sigB + sum(l)
}

### Valor Inicial
B=rep(0,n+7) # Definindo os valores iniciais para n+7 parâmetros.

## Metropolis
post.samp1 <- MCMCmetrop1R(funpost, theta.init=B,x=X,y=y1,N=n,
thin=1, mcmc=10000, burnin=100,logfun=TRUE,verbose=1000,force.samp = TRUE)
summary(post.samp1)
#Problema: Taxa de aceitação 0.

## Gibbs
B=rep(0,n+7) # Definindo os valores iniciais para n+7 parâmetros.

s=gibbs(funpost,start=B,m=500,scale=rep(1,(n+7)),x=X,y=y1,N=n)

bB=matrix(,nrow=6,ncol=7)   # Calculando estatísticas da amostra gerada
for(i in 1:7){bB[,i]=summary(s$par[,i])} # para os 7 parâmetros.
bB
plot(s$par[,8],ty='l',main="t1") # Ver cadeia de theta1
plot(s$par[,9],ty='l',main="t2") # Ver cadeia de theta2
# Estima!

## Outro comando do Pacote LearnBayes para metropolis
B=rep(0,n+7) # Definindo os valores iniciais para n+7 parâmetros.
varcov=diag(rep(1,n+7))
proposal=list(var=varcov,scale=1)
s=rwmetrop(funpost,proposal,start=B,m=1000,x=X,y=y1,N=n)
#Problema: Taxa de aceitação 0.

bB=matrix(,nrow=6,ncol=n)   # Calculando estatísticas da amostra gerada
```

```
for(i in 1:7){bB[,i]=summary(s$par[,i])} # para os 7 parâmetros.
bB
plot(s$par[,500],ty='l')

## Gibbs_met
# Muito esforço. Algoritmo de Gibbs com atualizações internas de metropolis
B=rep(0,n+7) # Definindo os valores iniciais para n+7 parâmetros.
mc_mvn <- gibbs_met(log_f=funpost,no_var=(n+7),
ini_value=B,iters=3,iters_met=(n+7),stepsizes_met=rep(1,(n+7)), x=X, y = y1,N=n)
```

Tabela 3.2: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos).

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	-0,192	0,034	b[12]	0,914	0,036
b[2]	1,548	0,040	b[13]	0,317	0,034
b[3]	0,496	0,034	b[14]	1,728	0,042
b[4]	-0,189	0,034	b[15]	1,110	0,037
b[5]	-0,289	0,034	b[16]	-0,253	0,034
b[6]	0,823	0,035	b[17]	0,393	0,034
b[7]	1,252	0,038	b[18]	1,307	0,038
b[8]	1,750	0,042	b[19]	0,503	0,034
b[9]	1,620	0,041	b[20]	1,422	0,039
b[10]	0,979	0,036	b[21]	0,882	0,036
b[11]	0,030	0,035	b[22]	1,948	0,046

Tabela 3.3: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos que trabalham fora de casa.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	0,166	0,085	b[12]	0,762	0,089
b[2]	1,564	0,102	b[13]	0,527	0,087
b[3]	0,558	0,087	b[14]	1,695	0,105
b[4]	0,190	0,085	b[15]	0,950	0,091
b[5]	0,125	0,085	b[16]	0,273	0,085
b[6]	0,895	0,090	b[17]	0,659	0,088
b[7]	1,430	0,099	b[18]	1,372	0,098
b[8]	1,993	0,114	b[19]	0,758	0,089
b[9]	1,607	0,104	b[20]	1,555	0,103
b[10]	1,296	0,098	b[21]	1,103	0,095
b[11]	0,377	0,088	b[22]	2,032	0,121

Tabela 3.4: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos que não trabalham fora de casa.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	-0,301	0,039	b[12]	0,919	0,041
b[2]	1,516	0,046	b[13]	0,252	0,039
b[3]	0,459	0,039	b[14]	1,715	0,048
b[4]	-0,303	0,039	b[15]	1,126	0,042
b[5]	-0,389	0,039	b[16]	-0,432	0,039
b[6]	0,783	0,040	b[17]	0,308	0,039
b[7]	1,196	0,043	b[18]	1,248	0,043
b[8]	1,704	0,048	b[19]	0,413	0,039
b[9]	1,613	0,047	b[20]	1,369	0,045
b[10]	0,894	0,041	b[21]	0,836	0,041
b[11]	-0,062	0,040	b[22]	1,937	0,053

Tabela 3.5: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos com pelo menos 1 reprovação.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	0,176	0,060	b[12]	0,865	0,063
b[2]	1,621	0,073	b[13]	0,582	0,061
b[3]	0,642	0,062	b[14]	1,766	0,075
b[4]	0,070	0,060	b[15]	0,939	0,064
b[5]	0,058	0,060	b[16]	0,152	0,060
b[6]	0,991	0,064	b[17]	0,734	0,062
b[7]	1,323	0,068	b[18]	1,362	0,0690
b[8]	1,958	0,080	b[19]	0,794	0,063
b[9]	1,670	0,074	b[20]	1,578	0,073
b[10]	1,059	0,066	b[21]	1,033	0,066
b[11]	0,223	0,062	b[22]	1,876	0,081

Tabela 3.6: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos sem reprovação.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	-0,455	0,044	b[12]	0,942	0,046
b[2]	1,485	0,051	b[13]	0,138	0,044
b[3]	0,373	0,044	b[14]	1,677	0,053
b[4]	-0,381	0,044	b[15]	1,178	0,048
b[5]	-0,539	0,044	b[16]	-0,550	0,044
b[6]	0,704	0,045	b[17]	0,182	0,044
b[7]	1,192	0,048	b[18]	1,226	0,049
b[8]	1,632	0,053	b[19]	0,304	0,044
b[9]	1,567	0,052	b[20]	1,292	0,049
b[10]	0,887	0,047	b[21]	0,792	0,046
b[11]	-0,094	0,045	b[22]	1,960	0,060

[h]

[h]

Tabela 3.7: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos considerados brancos e amarelos.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	-0,184	0,064	b[12]	0,763	0,067
b[2]	1,689	0,079	b[13]	0,299	0,064
b[3]	0,469	0,065	b[14]	1,64	0,078
b[4]	-0,269	0,064	b[15]	1,076	0,070
b[5]	-0,302	0,064	b[16]	-0,234	0,064
b[6]	0,779	0,067	b[17]	0,35	0,065
b[7]	1,344	0,073	b[18]	1,214	0,072
b[8]	1,685	0,079	b[19]	0,525	0,066
b[9]	1,641	0,078	b[20]	1,366	0,074
b[10]	0,936	0,069	b[21]	0,840	0,069
b[11]	-0,022	0,066	b[22]	1,910	0,087

Tabela 3.8: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos considerados pardos.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	-0,278	0,047	b[12]	0,966	0,050
b[2]	1,499	0,055	b[13]	0,220	0,047
b[3]	0,429	0,047	b[14]	1,783	0,059
b[4]	-0,313	0,047	b[15]	1,105	0,051
b[5]	-0,431	0,047	b[16]	-0,449	0,047
b[6]	0,787	0,049	b[17]	0,297	0,047
b[7]	1,173	0,052	b[18]	1,291	0,053
b[8]	1,739	0,058	b[19]	0,352	0,047
b[9]	1,542	0,056	b[20]	1,365	0,054
b[10]	0,894	0,050	b[21]	0,807	0,050
b[11]	-0,045	0,048	b[22]	1,916	0,064

Tabela 3.9: Estimativas do Modelo de Rasch tradicional para os dados da Prova Brasil 2007 (Resumido - 5.000 alunos). Somente alunos considerados pretos e índios.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Valor Estimado	Erro Padrão		Valor Estimado	Erro Padrão
b[1]	-0,092	0,091	b[12]	0,952	0,097
b[2]	1,474	0,106	b[13]	0,431	0,092
b[3]	0,593	0,093	b[14]	1,615	0,110
b[4]	0,009	0,091	b[15]	1,131	0,099
b[5]	-0,132	0,091	b[16]	-0,044	0,091
b[6]	0,936	0,097	b[17]	0,615	0,093
b[7]	1,241	0,101	b[18]	1,338	0,103
b[8]	1,725	0,1139	b[19]	0,701	0,094
b[9]	1,800	0,115	b[20]	1,530	0,108
b[10]	1,193	0,102	b[21]	1,073	0,100
b[11]	0,203	0,094	b[22]	1,981	0,126

Tabela 3.10: Estimativas do Modelo adaptado para heterogeneidade, controlando o fator “trabalha fora de casa”.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Média	Desvio Padrão		Média	Desvio Padrão
beta[1]	0,909	0,082	beta[2]	1,158	0,123
b[1]	0,166	0,085	b[12]	0,762	0,089
b[2]	1,564	0,102	b[13]	0,527	0,087
b[3]	0,558	0,087	b[14]	1,695	0,105
b[4]	0,190	0,085	b[15]	0,950	0,091
b[5]	0,125	0,085	b[16]	0,273	0,085
b[6]	0,895	0,090	b[17]	0,659	0,088
b[7]	1,430	0,099	b[18]	1,372	0,098
b[8]	1,993	0,114	b[19]	0,758	0,089
b[9]	1,607	0,104	b[20]	1,555	0,103
b[10]	1,296	0,098	b[21]	1,103	0,095
b[11]	0,377	0,088	b[22]	2,032	0,121

Tabela 3.11: Estimativas do Modelo adaptado para heterogeneidade, controlando o fator “reprovação’.

PARÂMETROS	Modelo Tradicional		PARÂMETROS	Modelo Tradicional	
	Estimativas			Estimativas	
	Média	Desvio Padrão		Média	Desvio Padrão
beta[1]	2,099	0,107	beta[2]	-1,28	0,089
b[1]	-0,842	0,063	b[12]	0,574	0,048
b[2]	1,244	0,049	b[13]	-0,174	0,053
b[3]	0,036	0,052	b[14]	1,436	0,05
b[4]	-0,786	0,061	b[15]	0,75	0,048
b[5]	-0,918	0,062	b[16]	-0,911	0,062
b[6]	0,443	0,048	b[17]	-0,062	0,05
b[7]	0,926	0,048	b[18]	0,963	0,048
b[8]	1,468	0,049	b[19]	0,054	0,051
b[9]	1,326	0,048	b[20]	1,095	0,047
b[10]	0,601	0,048	b[21]	0,526	0,049
b[11]	-0,489	0,057	b[22]	1,68	0,053

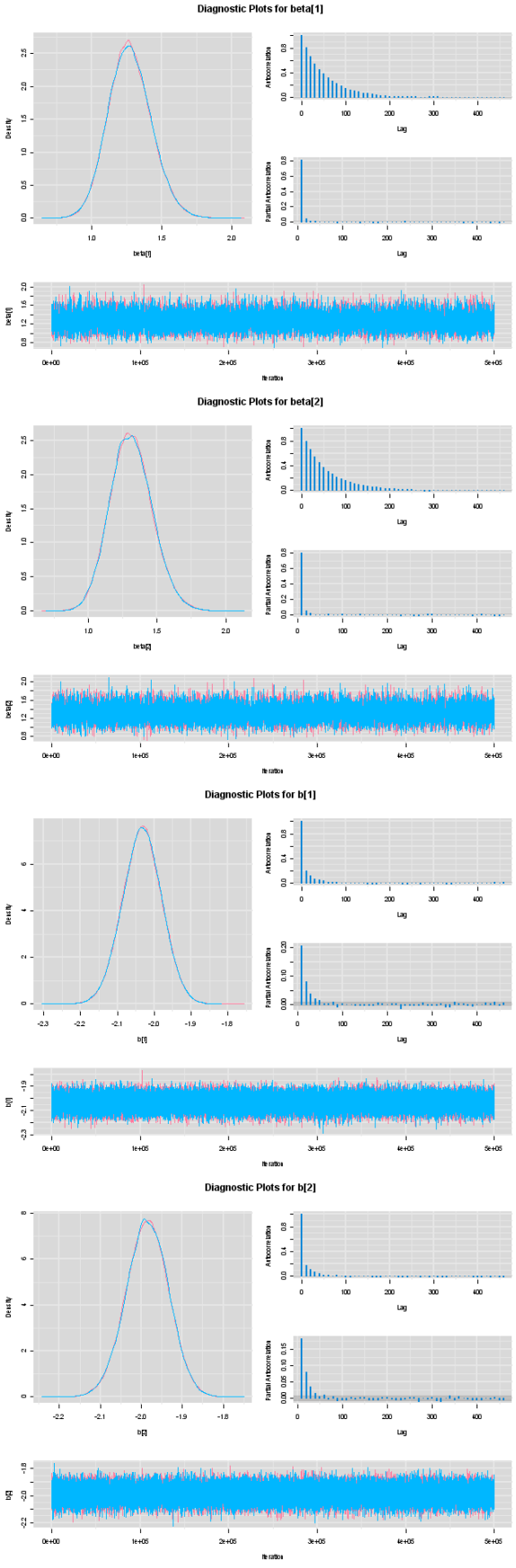


Figura 3.16: Gráficos de diagnóstico das cadeias dos parâmetros β_1, β_2, b_1 e b_2 da simulação MCMC.

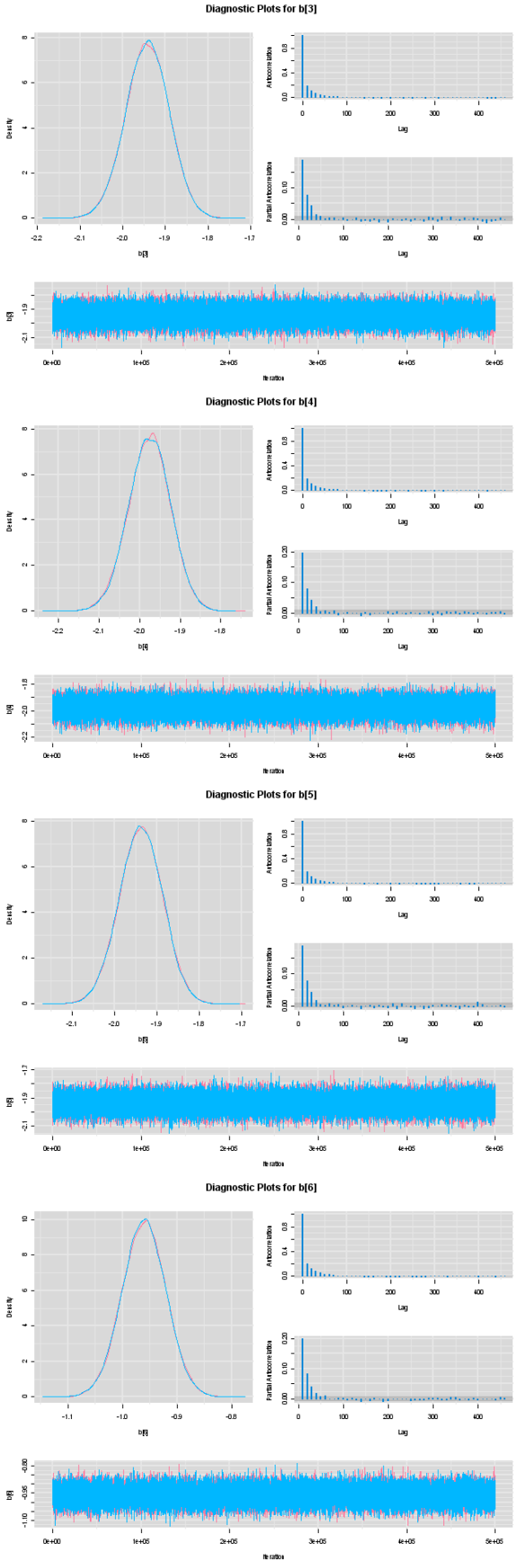
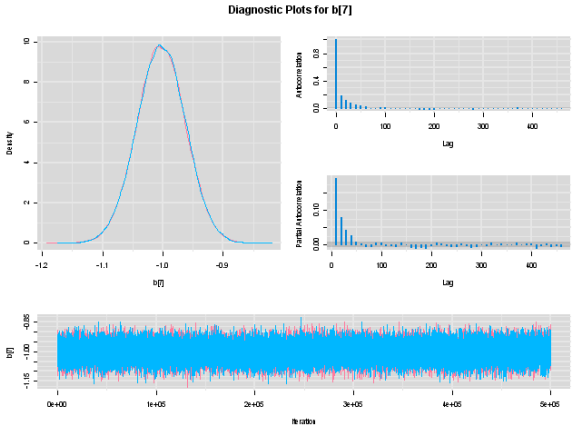
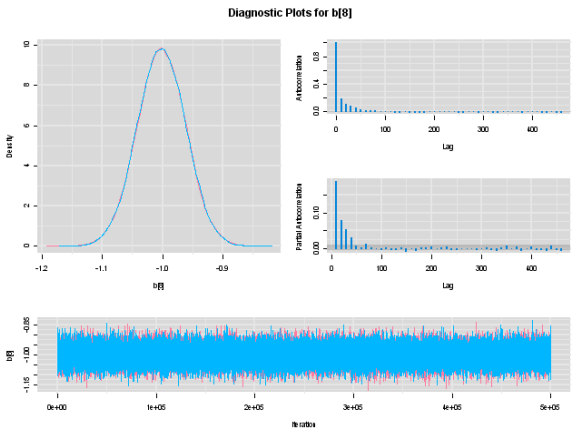
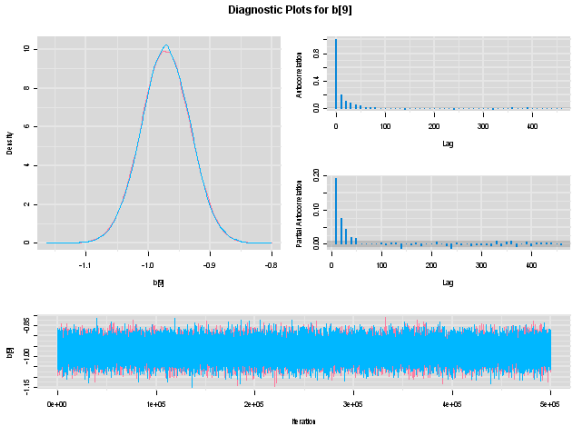


Figura 3.17: Gráficos de diagnóstico das cadeias dos parâmetros b_4, b_3, b_5 e b_6 da simulação MCMC.







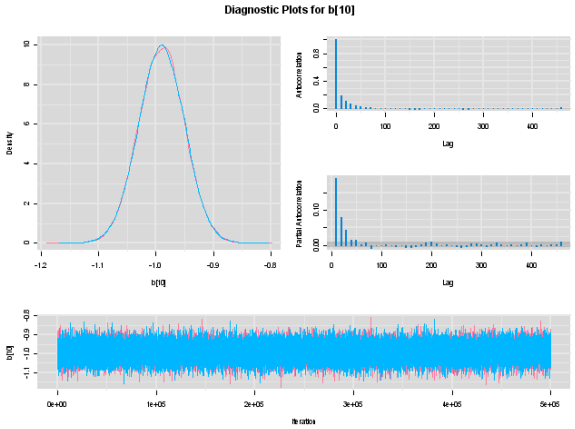
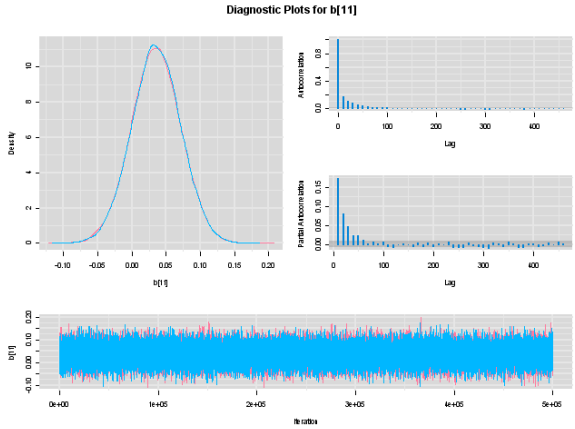
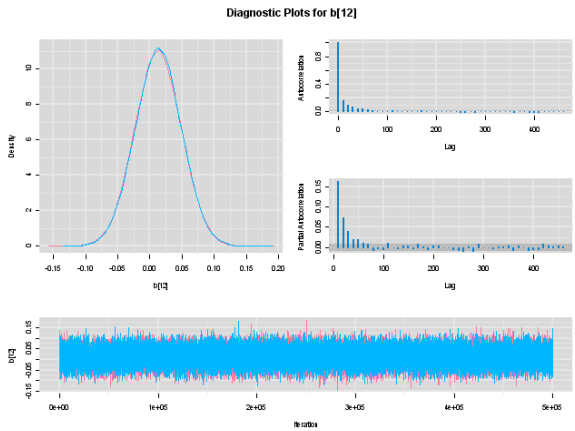
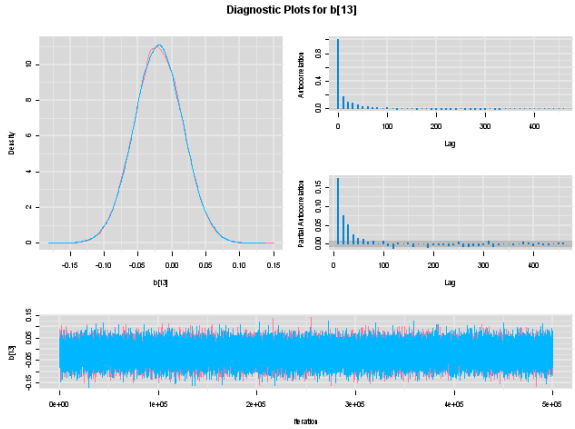


Figura 3.18: Gráficos de diagnóstico das cadeias dos parâmetros b_7, b_8, b_9 e b_{10} da simulação MCMC.







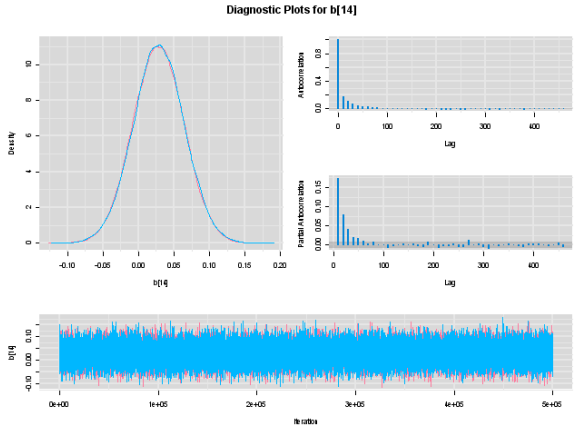
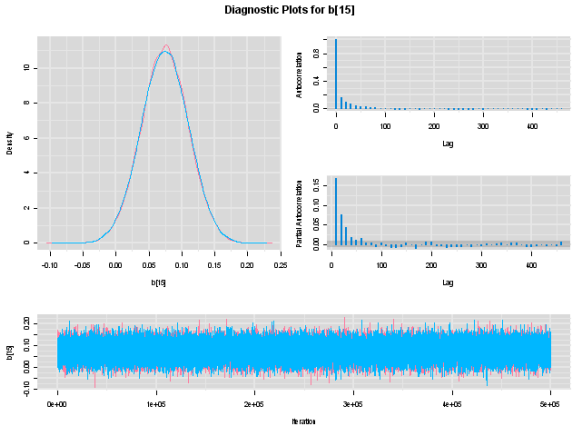
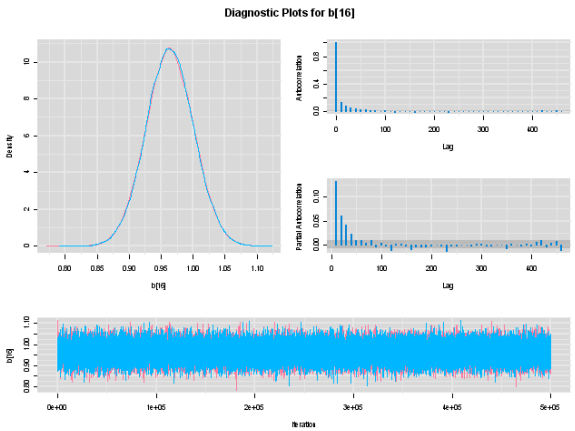
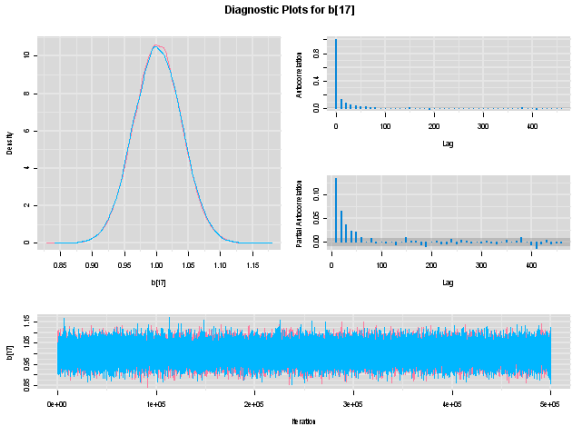


Figura 3.19: Gráficos de diagnóstico das cadeias dos parâmetros b_{11}, b_{12}, b_{13} e b_{14} da simulação MCMC.







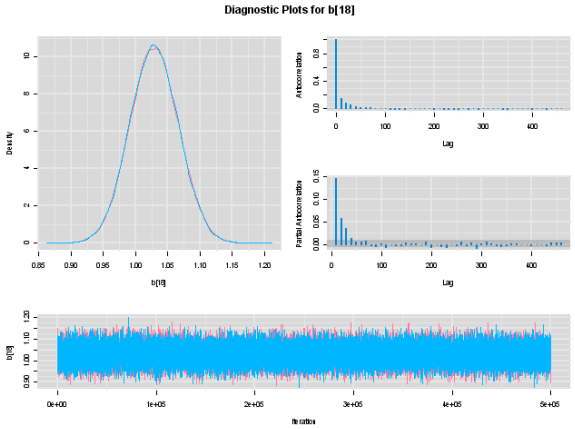
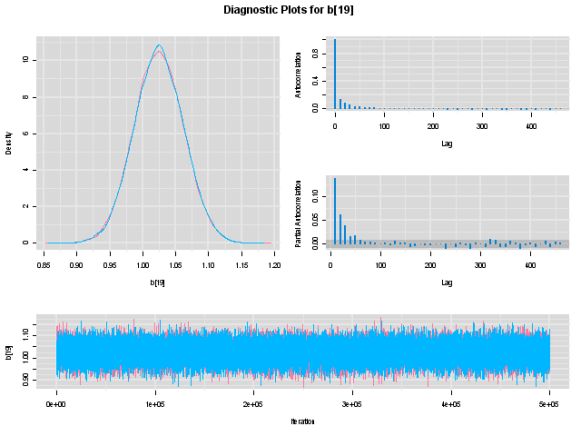
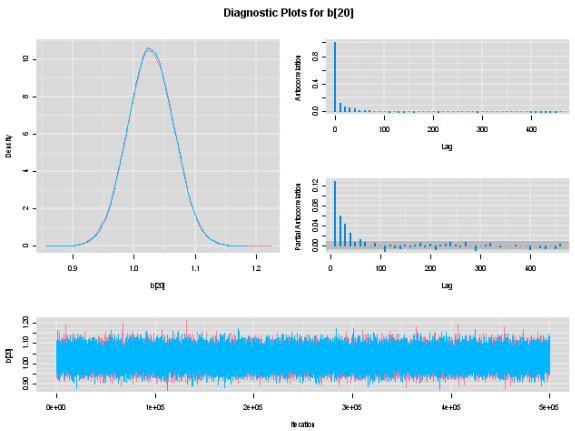
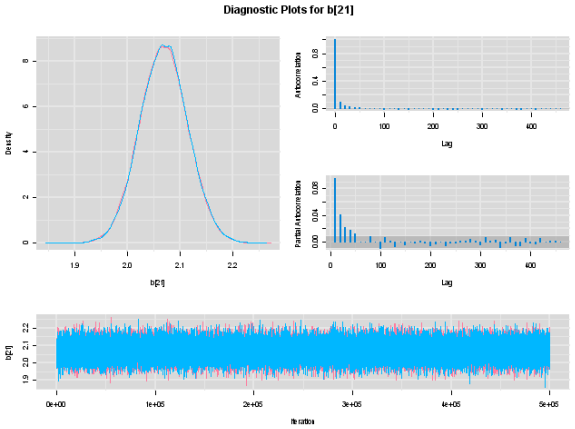


Figura 3.20: Gráficos de diagnóstico das cadeias dos parâmetros b_{15}, b_{16}, b_{17} e b_{18} da simulação MCMC.







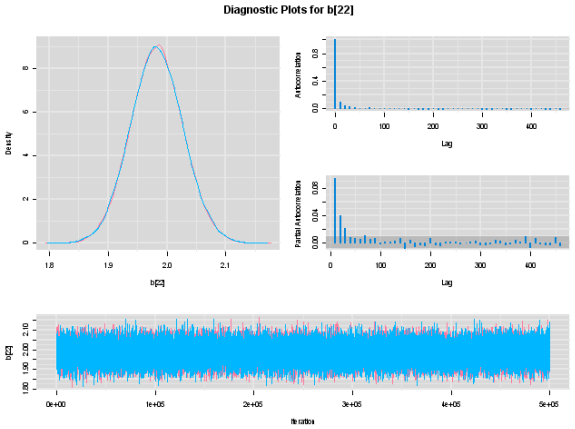


Figura 3.21: Gráficos de diagnóstico das cadeias dos parâmetros b_{19}, b_{20}, b_{21} e b_{22} da simulação MCMC.

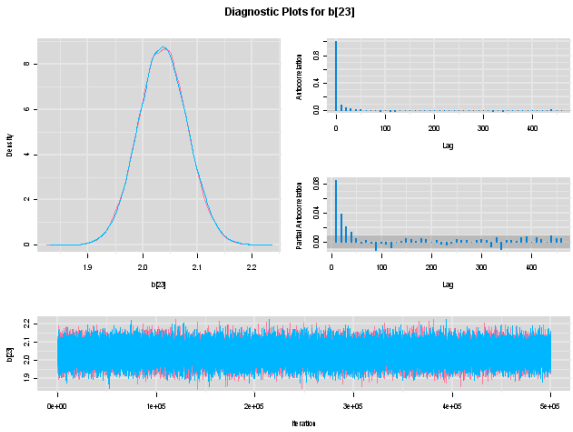


Figura 3.22: Gráficos de diagnóstico da cadeias do parâmetro b_{23} da simulação MCMC.

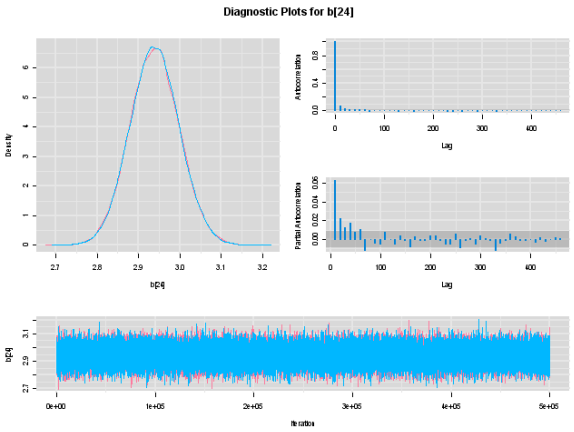


Figura 3.23: Gráficos de diagnóstico da cadeia do parâmetro b_{24} da simulação MCMC.

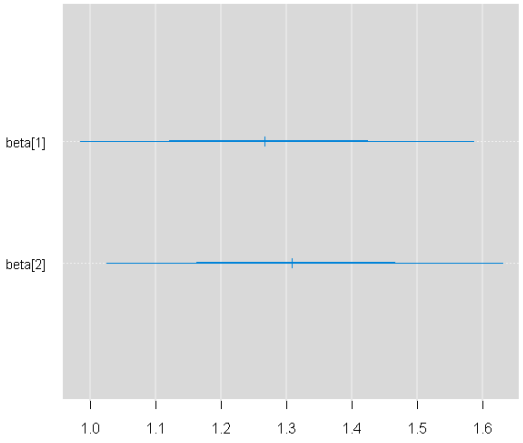


Figura 3.24: Intervalo de 95% e 68% de credibilidade para os parâmetros $\boldsymbol{\beta}$'s da simulação MCMC.

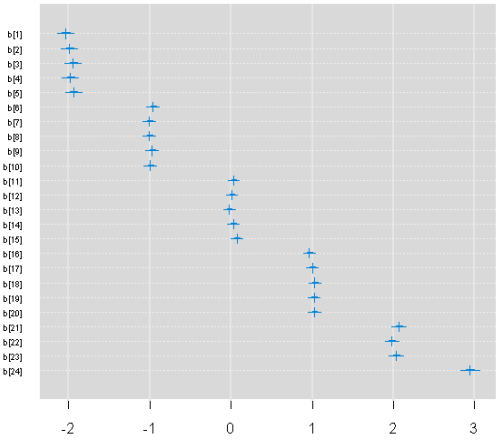


Figura 3.25: Intervalo de 95% e 68% de credibilidade para os parâmetros $\boldsymbol{b}$'s da simulação MCMC.

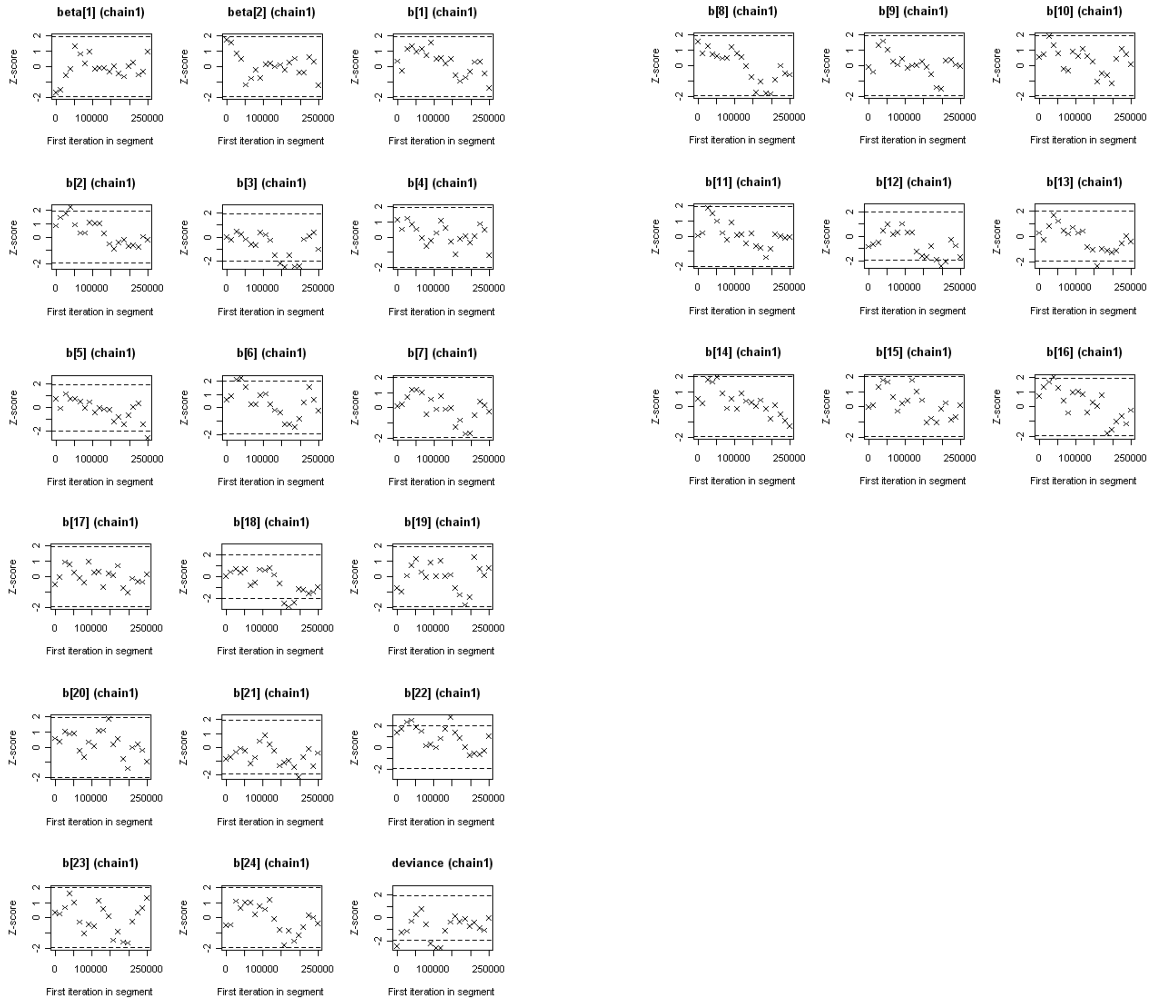


Figura 3.26: Gráficos de diagnóstico Geweke para os parâmetro da simulação MCMC.

INEP Ministério da Educação — BRASIL UM PAÍS DE TODOS GOVERNO FEDERAL

SAEB E PROVA BRASIL – 2007

CADERNO

PARA USO DO APLICADOR

← Aluno ausente

← Aluno presente e NÃO respondeu a prova

← Aluno respondeu no Caderno de Prova, mas NÃO preencheu a Folha de Respostas

← Aluno transcreveu corretamente as marcações do Caderno de Prova para a Folha de Respostas

← ANE

6552494005

ESCREVA O SEU NOME COMPLETO

CÓDIGO DA TURMA

Folha de Respostas

4.ª Série (5.º Ano) do Ensino Fundamental

BLOCO 1		BLOCO 2		BLOCO 3		BLOCO 4	
1	A B C D	12	A B C D	1	A B C D	12	A B C D
2	A B C D	13	A B C D	2	A B C D	13	A B C D
3	A B C D	14	A B C D	3	A B C D	14	A B C D
4	A B C D	15	A B C D	4	A B C D	15	A B C D
5	A B C D	16	A B C D	5	A B C D	16	A B C D
6	A B C D	17	A B C D	6	A B C D	17	A B C D
7	A B C D	18	A B C D	7	A B C D	18	A B C D
8	A B C D	19	A B C D	8	A B C D	19	A B C D
9	A B C D	20	A B C D	9	A B C D	20	A B C D
10	A B C D	21	A B C D	10	A B C D	21	A B C D
11	A B C D	22	A B C D	11	A B C D	22	A B C D

QUESTIONÁRIO DO ALUNO (a ser respondido após a prova)

1. Sexo:
A Masculino. B Feminino.

2. Como você se considera?
A Branco(a). B Pardo(a). C Preto(a). D Amarelo(a). E Indígena.

3. Qual é o mês do seu aniversário?
A Janeiro. B Fevereiro. C Março. D Abril. E Maio. F Junho. G Julho. H Agosto. I Setembro. J Outubro. K Novembro. L Dezembro.

4. Qual é a sua idade?
A 8 anos ou menos. B 9 anos. C 10 anos. D 11 anos. E 12 anos. F 13 anos. G 14 anos. H 15 anos ou mais.

5. Na sua casa tem televisão em cores?
A Sim, uma. B Sim, duas. C Sim, três ou mais. D Não tem.

6. Na sua casa tem rádio?
A Sim, um. B Sim, dois. C Sim, três ou mais. D Não tem.

7. Na sua casa tem videocassete ou DVD?
A Sim. B Não.

8. Na sua casa tem geladeira?
A Sim, uma. B Duas ou mais. C Não tem.

9. Na sua casa tem freezer separado da geladeira?
A Sim. B Não. C Não sei.

10. Na sua casa tem uma máquina de lavar roupa? (Não é tanquinho)
A Sim. B Não.

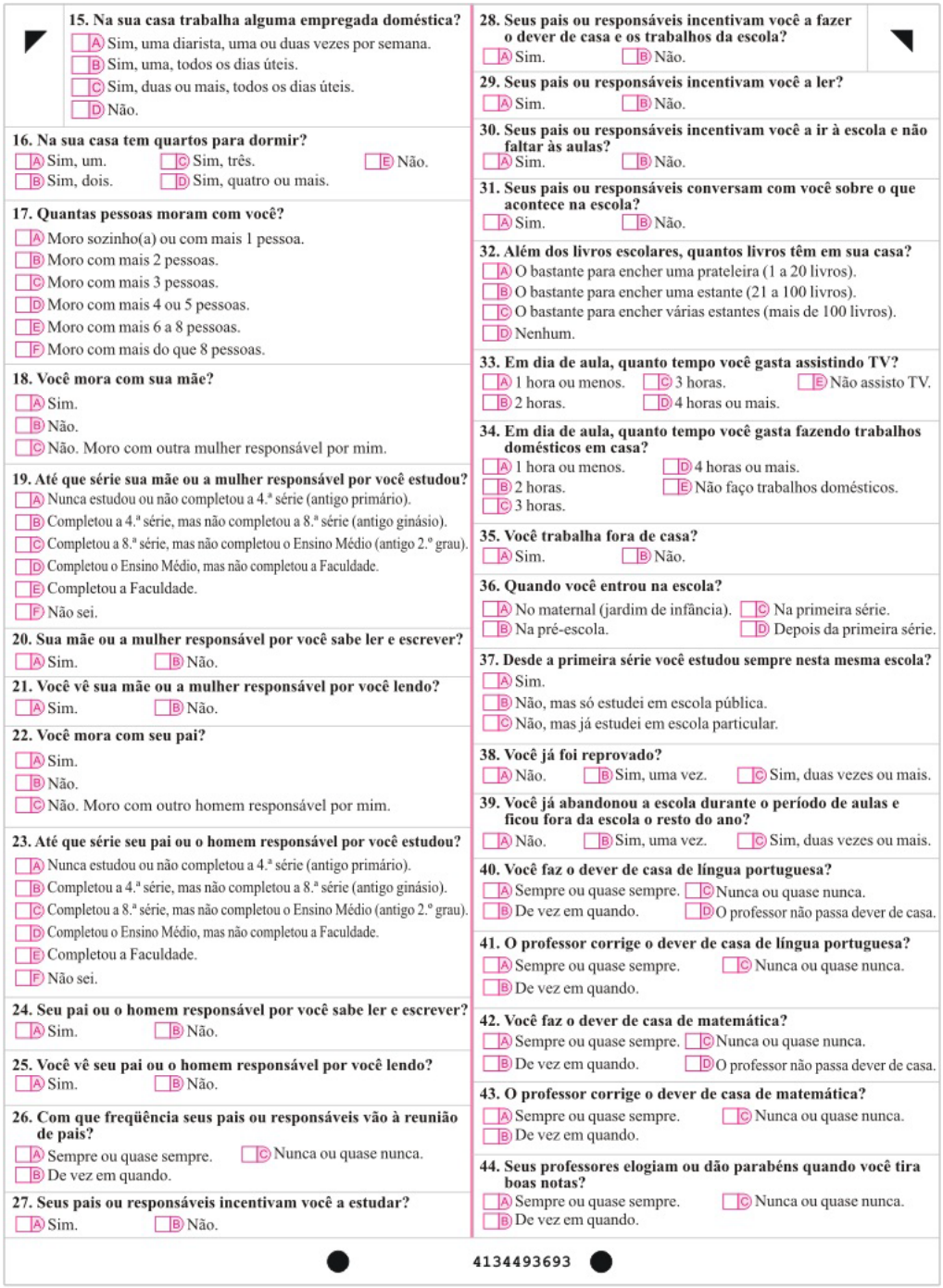
11. Na sua casa tem aspirador de pó?
A Sim. B Não.

12. Na sua casa tem carro?
A Sim, um. B Sim, dois. C Sim, três ou mais. D Não.

13. Na sua casa tem um computador?
A Sim, com internet. B Sim, sem internet. C Não.

14. Na sua casa tem banheiro?
A Sim, um. B Sim, dois. C Sim, três. D Sim, mais de três. E Não.

Figura 3.27: Questionário de aluno da Prova Brasil 2007, parte 1.



15. Na sua casa trabalha alguma empregada doméstica?
- A Sim, uma diarista, uma ou duas vezes por semana.
- B Sim, uma, todos os dias úteis.
- C Sim, duas ou mais, todos os dias úteis.
- D Não.

16. Na sua casa tem quartos para dormir?
- A Sim, um.
- B Sim, dois.
- C Sim, três.
- D Sim, quatro ou mais.
- E Não.

17. Quantas pessoas moram com você?
- A Moro sozinho(a) ou com mais 1 pessoa.
- B Moro com mais 2 pessoas.
- C Moro com mais 3 pessoas.
- D Moro com mais 4 ou 5 pessoas.
- E Moro com mais 6 a 8 pessoas.
- F Moro com mais do que 8 pessoas.

18. Você mora com sua mãe?
- A Sim.
- B Não.
- C Não. Moro com outra mulher responsável por mim.

19. Até que série sua mãe ou a mulher responsável por você estudou?
- A Nunca estudou ou não completou a 4.ª série (antigo primário).
- B Completou a 4.ª série, mas não completou a 8.ª série (antigo ginásio).
- C Completou a 8.ª série, mas não completou o Ensino Médio (antigo 2.º grau).
- D Completou o Ensino Médio, mas não completou a Faculdade.
- E Completou a Faculdade.
- F Não sei.

20. Sua mãe ou a mulher responsável por você sabe ler e escrever?
- A Sim.
- B Não.

21. Você vê sua mãe ou a mulher responsável por você lendo?
- A Sim.
- B Não.

22. Você mora com seu pai?
- A Sim.
- B Não.
- C Não. Moro com outro homem responsável por mim.

23. Até que série seu pai ou o homem responsável por você estudou?
- A Nunca estudou ou não completou a 4.ª série (antigo primário).
- B Completou a 4.ª série, mas não completou a 8.ª série (antigo ginásio).
- C Completou a 8.ª série, mas não completou o Ensino Médio (antigo 2.º grau).
- D Completou o Ensino Médio, mas não completou a Faculdade.
- E Completou a Faculdade.
- F Não sei.

24. Seu pai ou o homem responsável por você sabe ler e escrever?
- A Sim.
- B Não.

25. Você vê seu pai ou o homem responsável por você lendo?
- A Sim.
- B Não.

26. Com que freqüência seus pais ou responsáveis vão à reunião de pais?
- A Sempre ou quase sempre.
- B De vez em quando.
- C Nunca ou quase nunca.

27. Seus pais ou responsáveis incentivam você a estudar?
- A Sim.
- B Não.

28. Seus pais ou responsáveis incentivam você a fazer o dever de casa e os trabalhos da escola?
- A Sim.
- B Não.

29. Seus pais ou responsáveis incentivam você a ler?
- A Sim.
- B Não.

30. Seus pais ou responsáveis incentivam você a ir à escola e não faltar às aulas?
- A Sim.
- B Não.

31. Seus pais ou responsáveis conversam com você sobre o que acontece na escola?
- A Sim.
- B Não.

32. Além dos livros escolares, quantos livros têm em sua casa?
- A O bastante para encher uma prateleira (1 a 20 livros).
- B O bastante para encher uma estante (21 a 100 livros).
- C O bastante para encher várias estantes (mais de 100 livros).
- D Nenhum.

33. Em dia de aula, quanto tempo você gasta assistindo TV?
- A 1 hora ou menos.
- B 2 horas.
- C 3 horas.
- D 4 horas ou mais.
- E Não assisto TV.

34. Em dia de aula, quanto tempo você gasta fazendo trabalhos domésticos em casa?
- A 1 hora ou menos.
- B 2 horas.
- C 3 horas.
- D 4 horas ou mais.
- E Não faço trabalhos domésticos.

35. Você trabalha fora de casa?
- A Sim.
- B Não.

36. Quando você entrou na escola?
- A No maternal (jardim de infância).
- B Na pré-escola.
- C Na primeira série.
- D Depois da primeira série.

37. Desde a primeira série você estudou sempre nesta mesma escola?
- A Sim.
- B Não, mas só estudei em escola pública.
- C Não, mas já estudei em escola particular.

38. Você já foi reprovado?
- A Não.
- B Sim, uma vez.
- C Sim, duas vezes ou mais.

39. Você já abandonou a escola durante o período de aulas e ficou fora da escola o resto do ano?
- A Não.
- B Sim, uma vez.
- C Sim, duas vezes ou mais.

40. Você faz o dever de casa de língua portuguesa?
- A Sempre ou quase sempre.
- B De vez em quando.
- C Nunca ou quase nunca.
- D O professor não passa dever de casa.

41. O professor corrige o dever de casa de língua portuguesa?
- A Sempre ou quase sempre.
- B De vez em quando.
- C Nunca ou quase nunca.

42. Você faz o dever de casa de matemática?
- A Sempre ou quase sempre.
- B De vez em quando.
- C Nunca ou quase nunca.
- D O professor não passa dever de casa.

43. O professor corrige o dever de casa de matemática?
- A Sempre ou quase sempre.
- B De vez em quando.
- C Nunca ou quase nunca.

44. Seus professores elogiam ou dão parabéns quando você tira boas notas?
- A Sempre ou quase sempre.
- B De vez em quando.
- C Nunca ou quase nunca.

4134493693

Figura 3.28: Questionário de aluno da Prova Brasil 2007, parte 2.