

---

# Modelos alternativos da TRI para dados politômicos

*Marcelo Andrade da Silva*

---



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

MARCELO ANDRADE DA SILVA

**MODELOS ALTERNATIVOS DA TRI PARA DADOS POLITÔMICOS**

Tese apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Jorge Luis Bazán Guzmán

**São Carlos  
Abril de 2019**



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

MARCELO ANDRADE DA SILVA

**ALTERNATIVE POLYTOMOUS IRT MODELS**

Doctoral dissertation submitted to the Department of Statistics – DEs-UFSCar and to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Interinstitutional Graduate Program in Statistics. FINAL VERSION

Advisor: Prof. Dr. Jorge Luis Bazán Guzmán

**São Carlos  
April de 2019**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Marcelo Andrade da Silva, realizada em 22/03/2019:

Prof. Dr. Vicente Garibay Cancho  
USP

Prof. Dra. Mariana Cúri  
USP

Prof. Dr. Caio Lucidius Naberezny Azevedo  
UNICAMP

Prof. Dr. Victor Hugo Lachos Dávila  
UCoñn

Prof. Dra. Márcia D'Elia Branco  
USP

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Mariana Cúri e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ão) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Vicente Garibay Cancho

*Dedico esta pesquisa à minha  
querida esposa e melhor amiga,  
Raquel, pelo companheirismo e  
apoio incondicionais. Depois de  
Jesus Cristo, ela é o maior presente  
que Deus já me deu nesta vida.*





# AGRADECIMENTOS

---

---

Ao bondoso Deus, por se fazer conhecido para aqueles que o buscam.

Ao professor Jorge Bazán, pelo conhecimento transmitido, pelo tempo dedicado neste trabalho, pelos conselhos e por toda a ajuda, que foram preciosos para o desenvolvimento desta pesquisa.

Aos professores Caio Azevedo e Mariana Cúri, pelas valiosas contribuições fornecidas no Exame de Qualificação.

À professora Anne Corinne Huggins-Manley, da Universidade da Flórida, Estados Unidos, pela receptividade e pela parceria nesta pesquisa.

À professora Luzia Trinca, que me apresentou o universo acadêmico e me incentivou a ingressar no doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro oferecido para a realização desta pesquisa.

Ao Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP (PIPGEs), pela confiança depositada em mim quando fui aceito como aluno regular e por todo o apoio oferecido para a realização desta pesquisa. Em especial, a todos os professores e colegas do PIPGEs que contribuíram com minha formação.

Ao nosso grupo de pesquisa de “Modelos de Variáveis Latentes” (MVL), pelos valiosos encontros que proporcionaram muitas ideias para esta pesquisa.

À Universidade de São Paulo (USP) e à Universidade Federal de São Carlos (UFSCar), por todo o suporte oferecido ao longo destes últimos quatro anos.

À minha amada esposa, Raquel, pelo amor, carinho, paciência e apoio emocional.

Aos meus pais, Ronaldo e Sonia, pelo apoio e incentivo aos estudos.

Aos meus amigos da Igreja Aliança Cristã e Missionária de Bauru, por me ensinarem os verdadeiros valores eternos.

Por fim, a todos aqueles que, por quaisquer motivos, não mencionei, por contribuir direta ou indiretamente com a realização desta pesquisa.

Guardo com carinho em meu coração cada pessoa que fez parte desta etapa da minha vida. Muito obrigado.



# RESUMO

DA SILVA, M. A. **Modelos alternativos da TRI para dados politômicos**. 2019. 123 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Departamento de Estatística, Universidade Federal de São Carlos, São Carlos – SP, 2019.

Os modelos da teoria da resposta ao item (TRI) para dados politômicos são frequentemente utilizados na análise de dados provenientes das ciências comportamentais e sociais. Do ponto de vista prático, dados politômicos são mais informativos do que dados dicotômicos, uma vez que considera mais de duas categorias de resposta em cada item do teste, tornando atrativos os modelos designados a esse tipo de dados. A proposta desta pesquisa é explorar modelos alternativos da TRI para dados politômicos e suas extensões multidimensionais, preenchendo algumas lacunas existentes na literatura. Especificamente, os capítulos deste trabalho seguem uma sequência de construção da modelagem da TRI. Primeiramente, realizamos um estudo para auxiliar os leitores na escolha entre dois dos principais modelos da TRI para dados politômicos no contexto unidimensional: o modelo de resposta gradual (RG) e o modelo de crédito parcial generalizado (CPG). Conduzimos uma análise de sensibilidade de prioris para escolher um cenário de prioris adequado para cada modelo e verificamos o desempenho de alguns critérios de comparação de modelos frente a estes modelos através de um estudo de simulação. Em seguida, estendemos o modelo CPG unidimensional para o contexto bifator, propondo o modelo CPG-bifator, em que são considerados um traço latente global e traços latentes específicos através de uma estrutura aditiva em sua formulação. Além disso, flexibilizamos a estrutura do modelo CPG-bifator, tornando possível o seu uso com outras funções de ligação além da usual logito, tais como probito e clog-log. Na sequência, incorporamos a relação entre os itens do teste e as dimensões do traço latente dos indivíduos na formulação dos modelos da teoria da resposta ao item multidimensionais (TRIM) através da matriz  $\mathbf{Q}$ , um componente presente na grande maioria dos modelos de diagnóstico cognitivo (MDC), tornando acessível aos usuários uma maneira simples de expressar a relação item-traço nos modelos da TRIM. Por fim, propomos um método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM. Em particular, utilizamos no estudo o modelo CPG multidimensional com matriz  $\mathbf{Q}$  incorporada em sua formulação. Os diferentes estudos de simulação e as aplicações realizadas nesta pesquisa mostraram que estes modelos são modelos alternativos para a análise de dados politômicos e que podem ser utilizados pelos usuários na prática.

**Palavras-chave:** inferência Bayesiana; modelos da TRI multidimensionais; modelos para dados politômicos; teoria da resposta ao item; variáveis latentes.



# ABSTRACT

DA SILVA, M. A. **Alternative polytomous IRT models**. 2019. 123 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Departamento de Estatística, Universidade Federal de São Carlos, São Carlos – SP, 2019.

The item response theory (IRT) models for polytomous data are frequently used in the analysis of data coming from the behavioral and social sciences. From a practical point of view, polytomous data are more informative than dichotomous data, since it considers more than two response categories in each test item, making the models assigned to this type of data attractive. The purpose of this research is to explore alternative polytomous IRT models and their multidimensional extensions, filling some gaps in the literature. Specifically, the chapters of this work follow a construction sequence of IRT modeling. Firstly, we conducted a study to assist readers in choosing between two of the major polytomous IRT models in the one-dimensional context: the graded response (GR) model and the generalized partial credit (GPC) model. We conducted a sensitivity analysis of priors to choose a suitable priors scenario for each model and we verified the performance of some model comparison criteria against these models through a simulation study. Then, we extend the one-dimensional GPC model to the bifactor context, proposing the GPC-bifactor model, in which a global latent trait and specific latent traits are considered through an additive structure in its formulation. In addition, we flexibilize the structure of the GPC-bifactor model, making possible its use with other link functions beyond the usual logit, such as probit and clog-log. Then, we incorporate the relation between the items and the latent trait dimensions of the individuals in the formulation of the multidimensional item response theory (MIRT) models through the  $\mathbf{Q}$ -matrix, a component present in the vast majority of cognitive diagnostic models (CDM), making it easy for users to express the item-trait relationship in MIRT models. Finally, we propose a validation method using the  $\mathbf{Q}$ -matrix in MIRT models. In particular, we used in the study the multidimensional GPC model with  $\mathbf{Q}$ -matrix embedded in its formulation. The different simulation studies and the applications performed in this research showed that these models are alternative models for the analysis of polytomous data and that can be used by the users in practice.

**Keywords:** Bayesian inference; multidimensional IRT models; models for polytomous data; item response theory; latent variables.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – CCCIs sob o modelo RG supondo um item $j$ com $a_j = 1,5$ , $b_1 = -2$ , $b_2 = -0,7$ , $b_3 = 1$ e $b_4 = 2$ . . . . .	27
Figura 2 – CCCIs sob o modelo CPG supondo um item $j$ com $a_j = 1$ , $b_1 = 0$ , $b_2 = -1,5$ , $b_3 = 0,5$ e $b_4 = 2$ . . . . .	29
Figura 3 – <i>Boxplots</i> da distribuição da medida AVR B para os parâmetros de discriminação, dificuldade e indivíduos relativo aos cenários de prioris do modelo RG (A1, A2 e A3) e do modelo CPG (B1, B2 e B3). . . . .	38
Figura 4 – PEC de cada critério de comparação de modelo apresentado em um <i>heat map</i> (mapa de calor). . . . .	40
Figura 5 – Descrição dos 21 itens que compõem o BDI e a frequência relativa das respostas de cada item apresentada em um <i>heat map</i> (mapa de calor). . . . .	42
Figura 6 – Estimativas dos parâmetros de discriminação e intervalos de credibilidade de 95% dos modelos RG e CPG. . . . .	43
Figura 7 – Estimativas dos parâmetros de dificuldade e intervalos de credibilidade de 95% para o modelo RG. . . . .	43
Figura 8 – Estimativas dos parâmetros de dificuldade e intervalo de credibilidade de 95% para o modelo CPG. . . . .	44
Figura 9 – Gráfico de dispersão para as estimativas de traços latentes (modelos RG e CPG). Devido à diferença na natureza de ambos os modelos, o significado dos traços latentes em cada modelo não é o mesmo. . . . .	45
Figura 10 – $p$ -valor Bayesiano por item (modelos RG e CPG) com base na distância qui-quadrada. . . . .	46
Figura 11 – Os valores de AVR B dos parâmetros dos itens do modelo CPG-bifator para os quatro cenários construídos no estudo de recuperação de parâmetros. . . . .	58
Figura 12 – Estrutura considerada para a aplicação em dados de mobile banking. . . . .	59
Figura 13 – $p$ -valor Bayesiano por item para os dados de mobile banking. . . . .	61
Figura 14 – Estimativas dos parâmetros de discriminação global e específica para os dados de mobile banking. . . . .	62
Figura 15 – Estimativas dos parâmetros de dificuldade de acordo com as categorias para os dados de mobile banking. . . . .	63
Figura 16 – Resumo das estimativas dos parâmetros de indivíduos. . . . .	63

Figura 17 – CCCIs condicionais considerando o valor do traço latente específico como zero. . . . .	65
Figura 18 – CCCIs condicionais considerando o valor do traço latente global como zero. . . . .	66
Figura 19 – $p$ -valor Bayesiano por item para os dados de respostas de estudantes da Áustria. . . . .	78
Figura 20 – Estimativas dos parâmetros de discriminação global e específica para os dados de respostas de estudantes da Áustria. . . . .	78
Figura 21 – Estimativas dos parâmetros de dificuldade de acordo com as categorias de resposta dos itens para os dados de respostas de estudantes da Áustria. . . . .	79
Figura 22 – Resumo das estimativas dos parâmetros de indivíduos. . . . .	80
Figura 23 – Curvas características condicionais do item considerando o valor do traço latente específico como zero. . . . .	81
Figura 24 – Curvas características condicionais do item considerando o valor do traço latente global como zero. . . . .	82
Figura 25 – Representação gráfica do algoritmo de troca por ponto adaptado para busca de matriz $\mathbf{Q}$ . . . . .	91
Figura 26 – Exemplo de código R utilizado para chamar a linguagem de modelagem Stan através do pacote “RStan”. . . . .	119
Figura 27 – Código Stan do modelo de resposta gradual apresentado no Capítulo 2. . . . .	120
Figura 28 – Código Stan do modelo de crédito parcial generalizado apresentado no Capítulo 2. . . . .	121
Figura 29 – Código Stan do modelo de crédito parcial generalizado bifator com função de ligação clog-log apresentado no Capítulo 3. . . . .	122
Figura 30 – Código Stan do modelo de crédito parcial generalizado multidimensional com matriz $\mathbf{Q}$ apresentado no Capítulo 4. . . . .	123



# LISTA DE TABELAS

---

---

Tabela 1	– Uma representação da distribuição Categorizada apresentada em 2.1 . . . . .	25
Tabela 2	– Distribuições a priori para os parâmetros dos itens do modelo RG retiradas da literatura . . . . .	31
Tabela 3	– Distribuições a priori para os parâmetros dos itens do modelo CPG retiradas da literatura . . . . .	31
Tabela 4	– Média da RMSE (viés) para os 3 cenários de prioris de cada modelo . . . . .	37
Tabela 5	– Valores dos critérios de comparação de modelos para os 3 cenários considerados para os modelos RG e CPG . . . . .	39
Tabela 6	– Valores dos cinco critérios de comparação de modelos, $p$ -valor Bayesiano global e tempo, em minutos, do algoritmo NUTS para os modelos RG e CPG ajustados aos dados de BDI. . . . .	45
Tabela 7	– Valores de RMSE para os parâmetros dos itens a partir do modelos CPG-bifator com as três funções de ligação consideradas . . . . .	57
Tabela 8	– Valores dos critérios de comparação de modelos, $p$ -valor Bayesiano global e tempo, em minutos, do algoritmo NUTS referentes ao ajuste do modelo CPG-bifator com funções de ligação logito, probito e clog-log aos dados de mobile banking . . . . .	60
Tabela 9	– Frequência das categorias de resposta em cada item e as matrizes $\mathbf{Q}$ - <i>custom</i> e $\mathbf{Q}$ - <i>bifat</i> . . . . .	76
Tabela 10	– Valores dos índices de comparação de modelos DIC, WAIC e LOO, $p$ -valor Bayesiano global e tempo, em minutos, referentes ao ajuste do modelo CPGM-Q com a matriz $\mathbf{Q}$ - <i>custom</i> e com a matriz $\mathbf{Q}$ - <i>bifat</i> aos dados de respostas de estudantes da Áustria . . . . .	77
Tabela 11	– As oito classes de atributos dispostas em dois grupos . . . . .	87
Tabela 12	– Matriz $\mathbf{Q}_V$ e matrizes $\mathbf{Q}$ encontradas a partir dos cinco critérios propostos considerando o Cenário 1 do estudo de simulação . . . . .	95
Tabela 13	– Matriz $\mathbf{Q}_V$ e matrizes $\mathbf{Q}$ encontradas a partir dos cinco critérios propostos considerando o Cenário 2 do estudo de simulação . . . . .	96
Tabela 14	– Descrição dos itens, matriz $\mathbf{Q}_L$ e frequência das respostas em cada categoria. . . . .	97
Tabela 15	– Matriz $\mathbf{Q}_L$ utilizada por Fragoso e Cúri (2013) e DA SILVA <i>et al.</i> (2018) e matrizes $\mathbf{Q}$ encontradas a partir dos cinco critérios propostos considerando os dados do BDI . . . . .	99



# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	19
1.1	Objetivos e visão geral . . . . .	21
2	ESCOLHA ENTRE DOIS MODELOS POLITÔMICOS DA TRI . . . . .	23
2.1	Introdução . . . . .	23
2.2	Modelos da TRI para itens politômicos . . . . .	25
2.3	Estimação Bayesiana via NUTS . . . . .	29
2.4	Estudos de simulação . . . . .	36
2.5	Aplicação . . . . .	41
2.6	Conclusões e comentários . . . . .	46
3	UM MODELO DE CRÉDITO PARCIAL GENERALIZADO BIFATOR COM DIFERENTES FUNÇÕES DE LIGAÇÃO . . . . .	49
3.1	Introdução . . . . .	49
3.2	O modelo CPG-bifator . . . . .	51
3.3	Estimação Bayesiana para o modelo CPG-bifator . . . . .	53
3.4	Estudo de simulação . . . . .	56
3.5	Aplicação . . . . .	58
3.6	Conclusões e comentários . . . . .	67
4	UM MODELO DE CRÉDITO PARCIAL GENERALIZADO MULTIDIMENSIONAL COM MATRIZ Q . . . . .	69
4.1	Introdução . . . . .	69
4.2	Os modelos da TRIM para dados politômicos . . . . .	71
4.2.1	<i>Incorporando a matriz Q em modelos compensatórios</i> . . . . .	72
4.2.2	<i>O modelo CPGM-Q</i> . . . . .	73
4.3	Estimação Bayesiana para o modelo CPGM-Q . . . . .	73
4.4	Aplicação . . . . .	75
4.5	Conclusões e comentários . . . . .	83
5	VALIDAÇÃO DA MATRIZ Q EM MODELOS DA TRIM . . . . .	85
5.1	Introdução . . . . .	85
5.2	O método de validação da matriz Q para o modelo DINA . . . . .	86
5.3	Validação da matriz Q em modelos da TRIM . . . . .	88

5.4	Estudo de simulação . . . . .	93
5.5	Aplicação . . . . .	97
5.6	Conclusões e comentários . . . . .	99
6	COMENTÁRIOS FINAIS E DESENVOLVIMENTOS FUTUROS . .	101
6.1	Comentários finais . . . . .	101
6.2	Desenvolvimentos futuros . . . . .	102
	REFERÊNCIAS . . . . .	105
APÊNDICE A	O MÉTODO MONTE CARLO HAMILTONIANO E O ALGORITMO NO-U-TURN SAMPLER . . . . .	115
APÊNDICE B	CÓDIGO R E STAN . . . . .	119

---

## INTRODUÇÃO

---

No contexto de avaliação, os indivíduos pertencentes à população de interesse, que podem ser estudantes, consumidores, pacientes, entre outros, são submetidos a um instrumento de medida tal como um teste, questionário ou avaliação clínica. Os instrumentos de medida são compostos de itens propostos por especialistas com a finalidade de avaliar variáveis latentes de interesse dos indivíduos, usualmente chamadas de traços latentes, que são características não observáveis, tais como: proficiência em uma avaliação educacional, satisfação do consumidor sobre um produto ou um serviço e medidas de sintomas depressivos. Os modelos da teoria da resposta ao item (TRI) visam estimar os traços latentes relacionados aos indivíduos e os parâmetros associados aos itens do instrumento de medida.

Um item pode ser classificado de acordo com o número de categorias de resposta que ele possui. Itens que possuem apenas duas categorias de resposta, tais como certo/errado ou verdadeiro/falso, são denominados dicotômicos. Em contraste, itens que possuem mais de duas categorias de resposta são denominados não dicotômicos ou, mais comumente, politômicos. Os itens politômicos são divididos basicamente em duas classes: 1) nominal, em que as categorias de resposta não possuem ordenação entre si; e 2) ordinal, em que as categorias de resposta possuem ordenação entre si. Um dos formatos de itens politômicos ordinais mais conhecidos e utilizados é de escala tipo Likert ([LIKERT, 1932](#)), geralmente constituída de categorias de resposta que variam da “total discordância” até a “total concordância” sobre determinada afirmação. Itens politômicos tem sido frequentemente utilizados por especialistas nas áreas de medição educacional e psicológica por fornecerem uma experiência nova aos indivíduos submetidos e também por reunirem mais informação estatística sobre os traços latentes medidos ([COHEN, 1983](#); [NERING](#); [OSTINI, 2010](#)).

Além dos itens dicotômicos e politômicos, é possível utilizar os modelos da TRI

considerando itens de resposta livre, ou seja, itens sem categorias, desde que as respostas dos indivíduos sejam categorizadas posteriormente, isto é, atribuir para cada resposta livre uma pontuação  $k$ , em que  $k \in \{1, 2, \dots, m\}, m \in \mathbb{Z}^+$ . Assim, se  $m = 2$ , o item pode ser considerado dicotômico e, se  $m > 2$ , o item pode ser considerado politômico.

A maioria dos modelos da TRI para dados politômicos são propostos sob a suposição de categorias de resposta ordenadas. Dentre eles, o modelo de resposta gradual (modelo RG) proposto por Samejima (1969) e o modelo de crédito parcial generalizado (modelo CPG) proposto por Muraki (1992) são amplamente conhecidos e utilizados. Ambos modelos são descritos em detalhes no próximo capítulo. Uma visão geral dos modelos da TRI para dados politômicos pode ser encontrada em Embretson e Reise (2000) e em Ostini e Nering (2006).

Podemos encontrar na literatura vários métodos para estimar os parâmetros dos modelos da TRI para dados politômicos. Entre eles, métodos frequentistas tais como o algoritmo de maximização da esperança ou o método da quadratura adaptativo (BOCK; AITKIN, 1980; MURAKI, 1990; MURAKI, 1992) e também abordagens Bayesianas, utilizando métodos de Monte Carlo via cadeias de Markov (MCMC), através dos conhecidos métodos Metropolis Hastings (METROPOLIS *et al.*, 1953; HASTINGS, 1970) e Gibbs Sampling (GEMAN; GEMAN, 1984; GELFAND; SMITH, 1990). Outra opção encontrada para ajustar os modelos da TRI é utilizando pacotes disponíveis no software R, tais como “ltm” (RIZOPOULOS, 2006) e “mirt” (CHALMERS, 2012). Também é possível utilizar softwares de modelagem estatística, tais como WinBUGS (LUNN *et al.*, 2000a), OpenBUGS (THOMAS, 2007), JAGS (PLUMMER, 2003) e Stan (CARPENTER *et al.*, 2017).

Ao longo do tempo, muitos modelos da TRI foram estendidos para o caso multidimensional, dando origem aos modelos da TRI multidimensionais (TRIM). Situações em que deseja-se medir múltiplas características latentes de indivíduos ou situações em que a suposição de unidimensionalidade dos traços latentes é violada demandam modelos com uma estrutura multidimensional. Por exemplo, em uma avaliação designada a medir a competência matemática dos alunos, a suposição de um traço unidimensional é frequentemente difícil de ser sustentada na prática (ACKERMAN, 1994; RECKASE, 2009). A resposta a itens de matemática pode depender não apenas da habilidade dos alunos em matemática, mas também em leitura e interpretação de texto. Assim, os modelos da TRIM são promissores. Os principais desenvolvimentos da TRIM podem ser encontrados em (RECKASE, 2009).

No contexto de variáveis latentes multidimensionais, uma nova classe de modelos, conhecida como modelos de diagnóstico cognitivo (MDC), ver Rupp, Templin e Henson (2010), tornou-se recentemente um novo campo empolgante de pesquisa em variáveis latentes. Estes modelos tem por objetivo classificar os indivíduos em grupos com base na posse/não posse de certos atributos, fornecendo informações específicas sobre seus pontos

fortes e fracos. Os MDC possuem em sua formulação uma matriz comumente conhecida como matriz  $\mathbf{Q}$  (TATSUOKA, 1983), uma componente bastante simples e intuitiva para o usuário especificar quais itens medem quais dimensões do traço latente.

## 1.1 Objetivos e visão geral

O principal objetivo desta pesquisa é propor modelos da TRI para dados politômicos que sejam mais adequados a dados reais e desenvolver extensões para a medição de traços latentes multidimensionais, incorporando o uso de informações relevantes acerca da relação entre item-traço através da matriz  $\mathbf{Q}$ . Para a estimação dos parâmetros destes modelos, adotamos uma abordagem Bayesiana, que, com os avanços computacionais das últimas décadas, tornou-se útil no contexto de modelo da TRI devido a sua complexidade e ao grande número de parâmetros a serem estimados (FOX, 2010). Para isso, utilizamos o algoritmo No-U-Turn Sampler (NUTS) através do software Stan.

Nossos objetivos específicos são:

1. Estimar de maneira eficiente os parâmetros dos modelos da TRI para dados politômicos (especificamente o modelo RG e o modelo CPG) a partir de um enfoque Bayesiano, desenvolver estudos de sensibilidade de prioris e abordar a escolha entre estes modelos utilizando critérios de comparação de modelos.
2. Estender o modelo CPG para o contexto multidimensional, considerando um modelo bifator com diferentes funções de ligação.
3. Propor um modelo CPG multidimensional considerando informações da relação item-traço através da matriz  $\mathbf{Q}$ , generalizando o seu uso para os modelos da TRIM.
4. Propor um método empírico de validação de matriz  $\mathbf{Q}$  em modelos da TRIM através de um algoritmo de busca de matrizes.

Este trabalho é organizado da seguinte maneira. No Capítulo 2, apresentamos detalhadamente os modelos RG e CPG e realizamos um estudo comparativo visando auxiliar o leitor na escolha do modelo mais adequado aos dados. No Capítulo 3, propomos um modelo CPG bifator com funções de ligação flexíveis como um modelo alternativo para o contexto de dados politômicos. No Capítulo 4, apresentamos uma nova formulação para os modelos da TRIM visando incorporar informações da relação item-traço através da matriz  $\mathbf{Q}$ . No Capítulo 5, propomos um método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM. Finalmente, no Capítulo 6, apresentamos alguns comentários gerais e possíveis extensões deste trabalho atual.





---

# ESCOLHA ENTRE DOIS MODELOS POLITÔMICOS DA TRI

---

---

## 2.1 Introdução

Dentre os modelos da teoria da resposta ao item (TRI) para dados politômicos, o modelo de resposta gradual (modelo RG) proposto por Samejima (1969) e o modelo de crédito parcial generalizado (modelo CPG) proposto por Muraki (1992) são amplamente conhecidos e utilizados. O modelo RG é uma extensão do modelo logístico de dois parâmetros (LORD, 1952; BIRNBAUM, 1968) para itens politômicos e é chamado de *modelo de diferença* ou *modelo indireto* por Thissen e Steinberg (1986) devido às probabilidades de respostas serem modeladas indiretamente através da diferença entre as probabilidades acumuladas adjacentes. Em contraste, o modelo CPG é uma generalização do modelo de crédito parcial (modelo CP), que foi estendido por Masters (1982) a partir do modelo de Rasch (RASCH, 1980) para  $m > 2$  categorias de resposta. Por ter suas probabilidades modeladas diretamente, Thissen e Steinberg (1986) o chama de *modelo direto* ou *modelo de “divisão pelo total”*.

Um ponto que merece atenção no contexto de dados politômicos é a maneira de escolher o modelo mais adequado ao conjunto de dados. Devemos considerar as características do conjunto de dados no momento da escolha do modelo, verificando, por exemplo, se as categorias de resposta dos itens são nominais ou ordinais. Também devemos checar se a natureza do modelo encaixa-se no contexto dos dados, como, por exemplo, o modelo CPG, que é proposto para situações onde um item é pontuado parcialmente e o indivíduo ganha mais créditos na medida em que sua resposta se aproxima da completa. Além disso, podemos usar uma abordagem matemática existente para verificar o ajuste do modelo, como os critérios de comparação de modelos ou outros índices existentes. Por exemplo, Stern e Sinharay (2005), Sinharay, Johnson e Stern (2006) e Fox (2010) utilizaram em mo-

delos da TRI um procedimento chamado *Posterior Predictive Model Checking* (PPMC), que tem por objetivo avaliar o ajuste do modelo utilizando a distribuição preditiva à posteriori (GELMAN *et al.*, 2014). Maydeu-Olivares, Drasgow e Mead (1994) utilizam uma medida chamada de índice de observador ideal (IOI) para comparar os modelos RG e CP. Em geral, não existe consenso sobre qual abordagem utilizar para selecionar o modelo politômico da TRI mais adequado para um específico conjunto de dados e, portanto, recomenda-se utilizar mais de um procedimento para auxiliar na escolha do modelo mais apropriado.

O modelo RG e modelo CPG foram selecionados para este estudo por serem modelos da TRI para dados politômicos ordenados amplamente conhecidos e por gerarem indecisão aos pesquisadores no momento de escolher um deles para ajustar o conjunto de dados de interesse. Um exemplo prático deste problema é referente aos dados do Inventário de Depressão de Beck (BDI), descrito em Beck *et al.* (1961). Neste contexto, o modelo RG é utilizado por Castro, Trentini e Riboldi (2010), porém Fliege *et al.* (2009) utilizam o modelo CPG.

O objetivo deste capítulo é realizar um estudo que auxilie os leitores na escolha entre estes dois modelos da TRI para dados politômicos amplamente conhecidos: o modelo RG e o modelo CPG. Para isto, verificamos o desempenho de alguns critérios Bayesianos de comparação de modelos frente a estes modelos realizando um estudo de simulação. Além disso, investigamos o desempenho da recuperação de parâmetros de ambos os modelos e conduzimos uma análise de sensibilidade de prioris para escolher um cenário de prioris adequado para cada modelo. A estimação dos parâmetros dos modelos é feita através do algoritmo No-U-Turn Sampler (NUTS), descrito em Hoffman e Gelman (2014), via software Stan (CARPENTER *et al.*, 2017). Finalmente, apresentamos uma aplicação ajustando os modelos RG e CPG aos dados do BDI para avaliar características de depressão, afim de verificar o modelo mais adequado através dos critérios Bayesianos de comparação de modelos com melhor desempenho de acordo com o estudo de simulação realizado, demonstrando uma forma de justificar a escolha do modelo apropriado a ser utilizado na pesquisa.

Este capítulo está organizado da seguinte forma: nesta seção apresentamos uma visão geral dos modelos politômicos da TRI e a proposta deste capítulo. Na Seção 2.2, apresentamos formalmente os modelos RG e CPG. Na seção 2.3, apresentamos algumas especificações Bayesianas para ambos os modelos e conduzimos uma revisão dos principais critérios de comparação de modelos. Na Seção 2.4, conduzimos dois estudos de simulação. No primeiro, realizamos uma recuperação de parâmetros para os modelos RG e CPG e conduzimos uma análise de sensibilidade de distribuições a priori. O segundo estudo avalia o desempenho dos cinco critérios de comparação de modelos para escolher entre os dois modelos politômicos da TRI utilizados. Na Seção 2.5, apresentamos uma aplicação para

dados reais de saúde mental. Especificamente, avaliamos qual é o melhor modelo para depressão considerando o conjunto de dados BDI. Na seção 2.6, apresentamos algumas conclusões e comentários.

## 2.2 Modelos da TRI para itens politômicos

Para apresentar os modelos RG e CPG, consideramos a situação em que um instrumento de medida com  $J$  itens é aplicado a  $N$  indivíduos em que, para cada item  $j$ , há  $m_j$  categorias de resposta. Seja a variável aleatória  $Y_{ij}$  a resposta do indivíduo  $i$  ao item  $j$ , de tal modo que

$$Y_{ij}|\theta_i, \xi_j \sim \text{Categorizada}(P_{ij1}, P_{ij2}, \dots, P_{ijm_j}), \quad (2.1)$$

em que  $-\infty < \theta_i < +\infty$  é o traço latente do indivíduo  $i$ ,  $\xi_j$  é o vetor de parâmetros associado ao item  $j$ ,  $P_{ijk}$  é a probabilidade do indivíduo  $i$  selecionar uma específica categoria  $k$  no item  $j$ , com  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, J$  e  $k = 1, 2, \dots, m_j$ . A distribuição categorizada apresentada em 2.1 é uma generalização da distribuição de Bernoulli para variáveis que assumem uma quantidade finita de valores  $m_j$ , tal que  $m_j > 2$ , para  $j = 1, 2, \dots, J$ . Na língua portuguesa, esta distribuição é também conhecida por distribuição categórica. Podemos representar a Expressão 2.1 a partir da Tabela 1.

Tabela 1 – Uma representação da distribuição Categorizada apresentada em 2.1

$k$	1	2	...	$m_j$
$P(Y_{ij} = k \theta_i, \xi_j)$	$P_{ij1}$	$P_{ij2}$	...	$P_{ijm_j}$

A variável aleatória  $Y_{ij}$  assume o valor  $k$ ,  $k \in \{1, 2, \dots, m_j\}$  com probabilidade  $P_{ijk}$ . Obviamente,  $\sum_{k=1}^{m_j} P_{ijk} = 1$ .

### ***O modelo de resposta gradual***

De acordo com o modelo RG de Samejima (1969), a probabilidade  $P_{ijk}$  é modelada a partir das probabilidades acumuladas  $P_{ijk}^+$ , que são as probabilidade de um indivíduo  $i$  selecionar uma categoria específica  $k$  ou outra inferior no item  $j$ , e é definida como

$$P_{ijk}^+ = P(Y_{ij} \leq k|\theta_i, \xi_j) = \begin{cases} L(\eta_{ijk}), & \text{se } k \in \{1, 2, \dots, m_j - 1\} \\ 1, & \text{se } k = m_j \end{cases}, \quad (2.2)$$

em que  $\theta_i$  é o traço latente do indivíduo  $i$ ,  $\xi_j = (a_j, b_{j1}^*, \dots, b_{j[m_j-1]}^*)'$ ,  $a_j > 0$  é o parâmetro de discriminação do item  $j$ ,  $b_{jk}^* = a_j b_{jk}$ ,  $\forall k \in \{1, 2, \dots, m_j - 1\}$ ,  $b_{jk}$  é o parâmetro de dificuldade da categoria  $k$  pertencente ao item  $j$ ,  $L(\cdot)$  é tipicamente conhecido como *função*

de resposta do item e frequentemente é utilizada a distribuição logística acumulada, dada por

$$L(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

No modelo RG, um item  $j$  possui  $m_j$  categorias de resposta e, portanto, para cada item existem  $m_j - 1$  parâmetros de dificuldade. O preditor latente  $\eta_{ijk}$  é então definido como

$$\eta_{ijk} = a_j(b_{jk} - \theta_i) = b_{jk}^* - a_j\theta_i, \quad (2.3)$$

para  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, J$  e  $k = 1, 2, \dots, m_j - 1$ . Curtis (2010) e Johnson e Albert (1999) usam a parametrização apresentada na última igualdade em 2.3 quando utiliza inferência Bayesiana. Esta alternativa é preferida porque reduz o tempo computacional (FOX, 2010). Pelo uso desta parametrização, os parâmetros de dificuldade são facilmente recuperados por  $b_{jk} = \frac{b_{jk}^*}{a_j}$ .

É necessário definir

$$-\infty < b_{j[k-1]}^* \leq b_{jk}^* < +\infty, \quad \forall j = 1, 2, \dots, J \text{ e } \forall k = 2, 3, \dots, m_j - 1. \quad (2.4)$$

Esta é uma condição fundamental para o modelo RG, pois ela garante que as probabilidades acumuladas sejam não decrescentes.

A probabilidade de um indivíduo  $i$  selecionar uma categoria de resposta  $k$  em um item  $j$  é obtida pela diferença das probabilidades acumuladas adjacentes, isto é,

$$P_{ijk} = P(Y_{ij} = k | \theta_i, \xi_j) = \begin{cases} P_{ijk}^+, & \text{se } k = 1 \\ P_{ijk}^+ - P_{ij[k-1]}^+, & \text{se } 2 \leq k < m_j. \\ 1 - P_{ij[k-1]}^+, & \text{se } k = m_j \end{cases} \quad (2.5)$$

Reescrevemos a Equação 2.5 da seguinte forma

$$P_{ijk} = P(Y_{ij} = k | \theta_i, \xi_j) = \begin{cases} L(a_j\theta_i - b_{jk}^*), & \text{se } k = 1 \\ L(a_j\theta_i - b_{jk}^*) - L(a_j\theta_i - b_{j[k-1]}^*), & \text{se } 2 \leq k < m_j. \\ 1 - L(a_j\theta_i - b_{j[k-1]}^*), & \text{se } k = m_j \end{cases} \quad (2.6)$$

Para descrever graficamente a relação entre diferentes níveis do traço latente e a probabilidade de resposta a uma particular categoria em um item definida pelo modelo RG, a curva característica da categoria do item (CCCI) pode ser construída. Como exemplo, apresentamos na Figura 1 as CCCIs sob o modelo RG para um item  $j$  com 5 categorias de resposta, parâmetro de discriminação  $a_j = 1,5$  e parâmetros de dificuldade  $b_1 = -2$ ,  $b_2 = -0,7$ ,  $b_3 = 1$  e  $b_4 = 2$ .

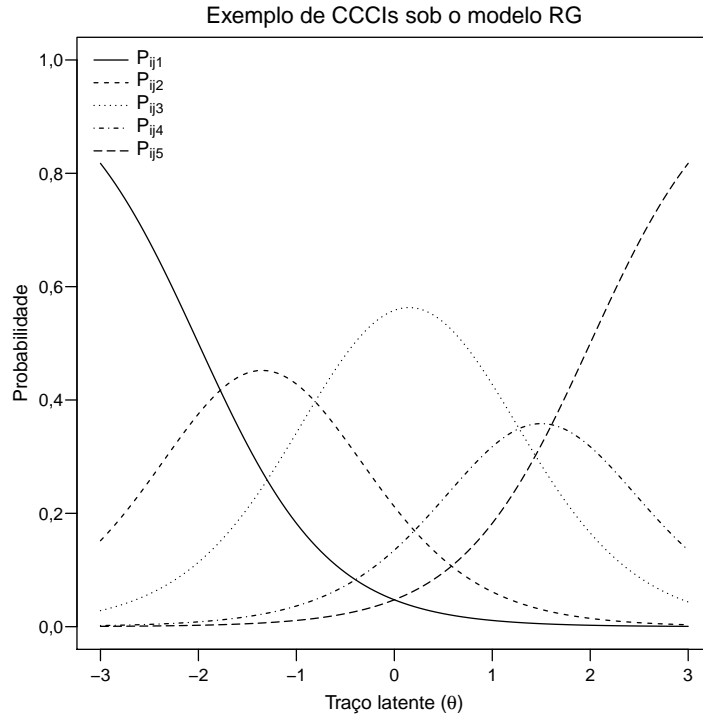


Figura 1 – CCCIs sob o modelo RG supondo um item  $j$  com  $a_j = 1,5$ ,  $b_1 = -2$ ,  $b_2 = -0,7$ ,  $b_3 = 1$  e  $b_4 = 2$ .

### ***O modelo de crédito parcial generalizado***

O modelo CPG, formulado por Muraki (1992), é uma alternativa para itens politômicos com categorias de resposta ordenadas. O modelo foi baseado na suposição de que a probabilidade de um indivíduo selecionar a categoria de resposta  $k$  sobre a categoria  $k - 1$  para o item  $j$  segue uma probabilidade condicional dada pelo modelo logístico de dois parâmetros, isto é,

$$P(Y_{ij} = k | Y_{ij} = k - 1) = P_{ijk|k-1} = \frac{P_{ijk}}{P_{ij[k-1]} + P_{ijk}} = \frac{\exp[a_j(\theta_i - b_{jk})]}{1 + \exp[a_j(\theta_i - b_{jk})]}. \quad (2.7)$$

O modelo CPG assume que duas categorias de resposta adjacentes em um item politômico podem ser consideradas como categorias dicotômicas e, portanto, a probabilidade de uma pessoa com um certo nível de habilidade alcançar a categoria de pontuação  $k$  em vez de  $k - 1$  pode ser descrita por um modelo da TRI para itens dicotômicos.

A partir da Equação 2.7, temos que  $P_{ijk} = \exp[a_j(\theta_i - b_{jk})] \cdot P_{ij[k-1]}$ . Em seguida, definimos

$$\lambda_k = \exp[a_j(\theta_i - b_{jk})], \quad (2.8)$$

o que implica em  $P_{ijk} = \lambda_k \cdot P_{ij[k-1]}$ . Dessa forma,

$$\begin{aligned} P_{ij2} &= \lambda_2 \cdot P_{ij1} \\ P_{ij3} &= \lambda_3 \cdot P_{ij2} = \lambda_3 \cdot \lambda_2 \cdot P_{ij1} \\ P_{ij4} &= \lambda_4 \cdot P_{ij3} = \lambda_4 \cdot \lambda_3 \cdot \lambda_2 \cdot P_{ij1} \\ &\vdots \\ P_{ijm_j} &= \lambda_{m_j} \cdot P_{ij[m_j-1]} = \lambda_{m_j} \cdot \lambda_{m_j-1} \cdot \dots \cdot \lambda_2 \cdot P_{ij1}. \end{aligned}$$

Como  $\sum_{k=1}^{m_j} P_{ijk} = 1$ , temos que  $P_{ij1} \cdot (1 + \sum_{l=2}^{m_j} \prod_{h=2}^l \lambda_h) = 1$ . Isolando  $P_{ij1}$ , encontramos

$$P_{ij1} = \frac{1}{(1 + \sum_{l=2}^{m_j} \prod_{h=2}^l \lambda_h)} = \frac{\lambda_1}{\lambda_1 \cdot (1 + \sum_{l=2}^{m_j} \prod_{h=2}^l \lambda_h)} = \frac{\lambda_1}{\sum_{l=1}^{m_j} \prod_{h=1}^l \lambda_h}.$$

Substituindo  $P_{ij1}$  nas igualdades acima, temos

$$\begin{aligned} P_{ij2} &= \frac{\lambda_1 \cdot \lambda_2}{\sum_{l=1}^{m_j} \prod_{h=1}^l \lambda_h} \\ P_{ij3} &= \frac{\lambda_1 \cdot \lambda_2 \cdot \lambda_3}{\sum_{l=1}^{m_j} \prod_{h=1}^l \lambda_h} \\ &\vdots \\ P_{ijm_j} &= \frac{\lambda_1 \cdot \dots \cdot \lambda_{[m_j-1]} \cdot \lambda_{m_j}}{\sum_{l=1}^{m_j} \prod_{h=1}^l \lambda_h}. \end{aligned}$$

Dessa forma, podemos escrever

$$P_{ijk} = \frac{\prod_{h=1}^k \lambda_h}{\sum_{l=1}^{m_j} \prod_{h=1}^l \lambda_h}.$$

Portanto, considerando a Equação 2.8, a probabilidade de um indivíduo  $i$  selecionar a categoria de resposta  $k$  em um item  $j$  é dada por

$$P_{ijk} = P(Y_{ij} = k | \theta_i, \xi_j) = \frac{\exp [\sum_{h=1}^k a_j (\theta_i - b_{jh})]}{\sum_{l=1}^{m_j} \exp [\sum_{h=1}^l a_j (\theta_i - b_{jh})]}, \quad (2.9)$$

em que  $\xi_j = (a_j, b_{j1}, \dots, b_{jm_j})'$ ,  $a_j > 0$  é o parâmetro de discriminação do item  $j$  e  $b_{j1}$  é arbitrariamente fixado em 0. Então, existem  $m_j - 1$  parâmetros de etapa,  $b_{j2}, \dots, b_{jm_j}$ , que são os parâmetros de dificuldade de mover da categoria de resposta  $k - 1$  para a categoria de resposta  $k$  no item  $j$ . Podemos decompor o parâmetro  $b_{jk}$ , de modo que  $b_{jk} = b_j - d_{jk}$ , em que  $b_j$  é o parâmetro de localização que caracteriza a dificuldade geral do item  $j$  e  $d_{jk}$  é o parâmetro de limite da categoria, interpretado como a dificuldade relativa da etapa  $k$  na comparação de outras etapas dentro de um item. O parâmetro de limite da última categoria para cada item é definido como o oposto da soma dos parâmetros de limite das categorias anteriores, isto é,  $d_{jm_j} = -\sum_{k=1}^{m_j-1} d_{jk}$ . Essa restrição faz com que a soma dos parâmetros de limite das categorias seja igual a 0 para cada item.

Assim como no modelo RG, podemos descrever graficamente a relação entre diferentes níveis do traço latente e a probabilidade de resposta a uma particular categoria em um item definida pelo modelo CPG através da curva característica da categoria do item (CCCI). Apresentamos na Figura 2 as CCCIs sob o modelo CPG para um item  $j$  com 4 categorias de resposta, parâmetro de discriminação  $a_j = 1$  e parâmetros de dificuldade  $b_1 = 0$ ,  $b_2 = -1,5$ ,  $b_3 = 0,5$  e  $b_4 = 2$ .

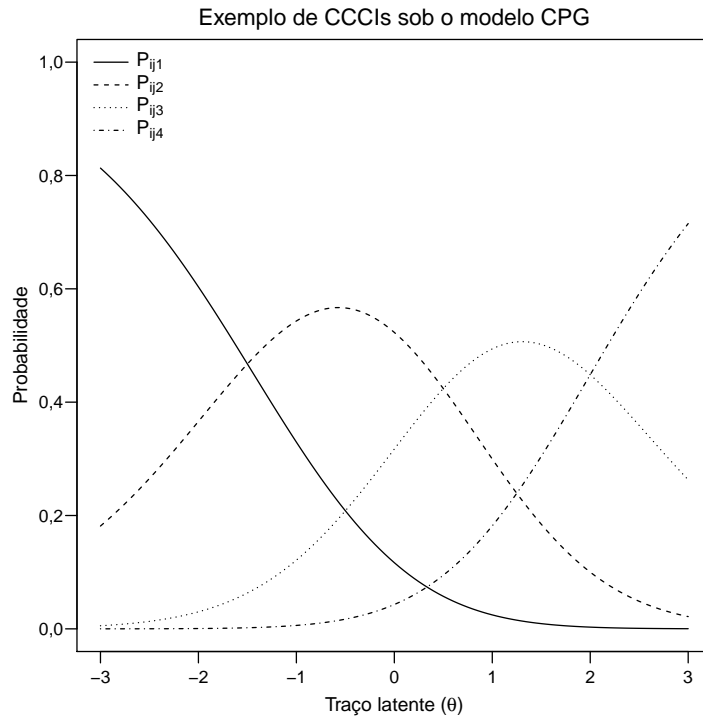


Figura 2 – CCCIs sob o modelo CPG supondo um item  $j$  com  $a_j = 1$ ,  $b_1 = 0$ ,  $b_2 = -1,5$ ,  $b_3 = 0,5$  e  $b_4 = 2$ .

A construção dos modelos RG e CPG para dados aumentados pode ser encontrada em [Ferreira \(2014\)](#).

## 2.3 Estimação Bayesiana via NUTS

Nesta seção, apresentamos uma abordagem Bayesiana para os modelos RG e CPG através do algoritmo MCMC No-U-Turn Sampler (NUTS), uma extensão do método Monte Carlo Hamiltoniano (HMC), proposto por [Neal \(2011\)](#).

### Verossimilhança

Dada uma matriz  $\mathbf{Y} = [y_{ij}]_{N \times J}$  representando as respostas de  $N$  indivíduos submetidos a um instrumento de medida com  $J$  itens, o vetor de parâmetros de indivíduos  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)'$ , e o vetor de parâmetros de itens  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_J)'$ , em que  $\theta_i$  e  $\xi_j$

são definidos na Seção 2.2, a verossimilhança dos modelos RG e CPG pode ser escrita como

$$\ell(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^N \prod_{j=1}^J P(Y_{ij} = y_{ij}|\boldsymbol{\theta}_i, \boldsymbol{\xi}_j) = \prod_{i=1}^N \prod_{j=1}^J \prod_{k=1}^{m_j} [P_{ijk}]^{w_{ijk}}, \quad (2.10)$$

em que  $P_{ijk}$  é dada pela Equação 2.6 e pela Equação 2.9, respectivamente para o modelo RG e para o modelo CPG,  $w_{ijk} = 1$  se a  $k$ -ésima categoria do item  $j$  é selecionada pelo indivíduo  $i$  e  $w_{ijk} = 0$ , caso contrário. A variável  $w_{ijk}$  faz com que a verossimilhança considere apenas as probabilidades das categorias de resposta selecionadas pelos indivíduos, permitindo o uso de banco de dados com respostas faltantes, pois se o indivíduo  $i$  não responde um item  $j$ , então  $w_{ijk} = 0$  para  $k = 1, 2, \dots, m_j$ .

### **Especificação das distribuições a priori e a posteriori**

Em inferência Bayesiana, assume-se que os parâmetros de interesse são variáveis aleatórias e, por isso, faz-se necessário especificar distribuições de probabilidades a priori que reflitam o nosso conhecimento prévio sobre o comportamento destes parâmetros. Este é um passo importante na análise Bayesiana, porque a distribuição de probabilidades a posteriori é influenciada tanto pela amostra quanto pelas distribuições de probabilidades a priori.

Para a especificação de uma distribuição a priori conjunta dos parâmetros dos itens e dos indivíduos para o modelo RG, assume-se que existe uma estrutura de prioris independentes, com exceção dos parâmetros de dificuldade, pois devem respeitar a restrição dada pela Expressão 2.4. Assim, a distribuição a priori conjunta dos parâmetros do modelo RG pode ser escrita como

$$P_{RG}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^N p(\boldsymbol{\theta}_i) \prod_{j=1}^J \left[ p(a_j) p(b_{j1}) \prod_{k=2}^{m_j-1} p(b_{jk}|b_{j[k-1]}) \right]. \quad (2.11)$$

Para a especificação de uma distribuição a priori conjunta dos parâmetros dos itens e dos indivíduos para o modelo CPG, assume-se que existe uma estrutura de prioris independentes da seguinte forma

$$P_{CPG}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^N p(\boldsymbol{\theta}_i) \prod_{j=1}^J \left[ p(a_j) \prod_{k=1}^{m_j} p(b_{jk}) \right]. \quad (2.12)$$

Na maioria dos casos, os pesquisadores adotam a distribuição normal padrão para a distribuição a priori da variável latente  $\boldsymbol{\theta}_i, i = 1, 2, \dots, N$ , mas tanto no modelo RG como no modelo CPG são utilizadas diferentes prioris para os parâmetros de discriminação e dificuldade. Apresentamos separadamente algumas prioris usadas pelos pesquisadores nos modelos RG e CPG. Na Tabela 2, apresentamos três cenários de prioris para os parâmetros de itens do modelo RG.



Tabela 2 – Distribuições a priori para os parâmetros dos itens do modelo RG retiradas da literatura

Parâmetros	Cenários de prioris		
	A1	A2	A3
$a_j$	$\ln N(0;1)$	$\ln N(0; \sqrt{2})$	$N(1;2,5)T(0; +\infty)$
$b_{j1}$	$N(0;2)$	$U(-5; +5)$	–
$b_{j[k+1]}$	$N(0;2)T(b_{jk}; +\infty)$	$U(-5; +5)T(b_{jk}; +\infty)$	–
$b_{j1}^*$	–	–	$N(0;2,5)$
$b_{j[k+1]}^*$	–	–	$N(0;2,5)T(b_{jk}; +\infty)$

A notação  $N(\mu, \sigma)$  indica uma distribuição normal com média  $\mu$  e desvio padrão  $\sigma$ ,  $\ln N(\mu, \sigma)$  denota uma distribuição lognormal com parâmetros de localização  $\mu$  e escala  $\sigma$ ,  $I(b_{jk})$  indica que o parâmetro  $b_{j[k+1]}$  é amostrado de tal forma que  $b_{j[k+1]} > b_{jk}$ , que é uma restrição do modelo RG,  $U(c, d)$  denota uma distribuição uniforme no intervalo  $(c, d)$  e  $T(c, d)$  indica que a distribuição considerada é truncada por valores menores que  $c$  e maiores que  $d$ . O cenário de prioris A1 foi proposto por [Zhu e Stone \(2012\)](#), o cenário A2 foi proposto por [Kieftenbeld e Natesan \(2012\)](#) e o cenário A3 foi proposto por [Curtis \(2010\)](#). Especificamente [Curtis \(2010\)](#) utiliza prioris para os parâmetros  $b_{jk}^*$ . Entretanto, como apresentado na Seção 2.2,  $b_{jk}$  pode ser escrito em função de  $b_{jk}^*$ . Na Tabela 3, nós listamos três cenários de prioris para o modelo CPG.

Tabela 3 – Distribuições a priori para os parâmetros dos itens do modelo CPG retiradas da literatura

Parâmetros	Cenários de prioris		
	B1	B2	B3
$a_j$	$\ln N(0;5)$	$N(0; \sqrt{1000})T(0; +\infty)$	$N(1;2,5)T(0, +\infty)$
$b_{jk}$	–	$N(0, \sqrt{1000})T(-3, 5; +3, 5)$	$N(0;1)$
$b_j$	$N(0;5)$	–	–
$d_{jk}$	$N(0;5)$	–	–

O cenário de prioris B1 foi proposto por [Stone e Zhu \(2015\)](#), o cenário B2 foi proposto por [Zhu, Robinson e Torenvlied \(2014\)](#) e o cenário B3 foi proposto por [Li e Baser \(2012\)](#), baseado em [Curtis \(2010\)](#). Especificamente [Stone e Zhu \(2015\)](#) utiliza prioris para os parâmetros  $b_j$  e  $d_{jk}$ . Entretanto, como apresentado na Seção 2.2,  $b_{jk}$  pode ser escrito em função de  $b_j$  e  $d_{jk}$ . Na Seção 2.4, nós conduzimos uma análise de sensibilidade das distribuições a priori consideradas nos cenários apresentados nas Tabelas 2 e 3.

Considerando a verossimilhança na Equação 2.10 e a distribuição a priori conjunta do modelo RG dada pela Equação 2.11, a distribuição conjunta a posteriori do modelo

RG é dada por

$$\begin{aligned}
 P_{RG}(\boldsymbol{\theta}, \boldsymbol{\xi} | \mathbf{Y}) &\propto P_{RG}(\boldsymbol{\theta}, \boldsymbol{\xi}) \ell(\mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\xi}) \\
 &\propto \prod_{i=1}^N \left[ \phi(\boldsymbol{\theta}_i) \prod_{j=1}^J P(Y_{ij} = y_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\xi}_j) \right] \\
 &\quad \times \prod_{j=1}^J \left[ p(a_j) p(b_{j1}) \prod_{k=2}^{m_j-1} p(b_{jk} | b_{j[k-1]}) \right],
 \end{aligned} \tag{2.13}$$

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

Considerando a verossimilhança na Equação 2.10 e a distribuição a priori conjunta do modelo CPG dada pela Equação 2.12, a distribuição conjunta a posteriori do modelo CPG é dada por

$$\begin{aligned}
 P_{CPG}(\boldsymbol{\theta}, \boldsymbol{\xi} | \mathbf{Y}) &\propto P_{CPG}(\boldsymbol{\theta}, \boldsymbol{\xi}) \ell(\mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\xi}) \\
 &\propto \prod_{i=1}^N \left[ \phi(\boldsymbol{\theta}_i) \prod_{j=1}^J P(Y_{ij} = y_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\xi}_j) \right] \prod_{j=1}^J \left[ P(a_j) \prod_{k=1}^{m_j} P(b_{jk}) \right].
 \end{aligned} \tag{2.14}$$

As distribuições a posteriori dos parâmetros dada pelas Equações 2.13 e 2.14 tem uma forma intratável. Assim, para obter estimativas dos parâmetros dos modelos RG e CPG, utilizamos o algoritmo NUTS via software Stan (CARPENTER *et al.*, 2017). Com sintaxe similar ao BUGS, o software Stan fornece inferência Bayesiana completa para modelos estatísticos através de dois possíveis algoritmos MCMC: HMC e NUTS. Para distribuições a posteriori complexas, ou seja, com grande quantidade de parâmetros, os algoritmos HMC e NUTS são mais eficientes em termos computacionais do que Gibbs Sampling e Metropolis-Hastings (GRANT *et al.*, 2017), utilizando menos iterações para convergência, o que justifica o nosso uso. Uma descrição dos algoritmos HMC e NUTS é apresentada no Apêndice A. Além disso, apresentamos nas Figuras 27 e 28, no Apêndice B, o código em Stan dos modelos RG e CPG, respectivamente.

## **Critérios de comparação de modelos**

Para comparar os modelos RG e CPG, consideramos 5 critérios de comparação de modelos: *Deviance Information Criterion* (DIC), *expected Akaike information criterion* (EAIC), *expected Bayesian (or Schwarz) and Akaike information criterion* (EBIC), *Watanabe-Akaike information criterion* (WAIC) e *leave-one-out cross-validation* (LOO-CV ou simplesmente LOO). O DIC é discutido por Spiegelhalter *et al.* (2002) e amplamente utilizado em comparações de modelos da TRI, como podemos ver em Bazán, Branco e Bolfarine (2006), Li *et al.* (2009) e Fragoso e Cúri (2013). Da mesma forma, o EAIC (BROOKS, 2002) e o EBIC (CARLIN; LOUIS, 2001) também podem ser usados para comparar diferentes modelos da TRI. Por outro lado, as medidas WAIC (WATANABE,

2010) e LOO (VEHTARI; GELMAN; GABRY, 2017) são menos conhecidas por terem sido recentemente propostas.

Os critérios DIC, EAIC e EBIC são baseados em uma medida de complexidade do modelo dada pelo número efetivo de parâmetros  $\rho_D = \mathbb{E}[D(\boldsymbol{\theta}, \boldsymbol{\xi})] - D(\mathbb{E}[\boldsymbol{\theta}], \mathbb{E}[\boldsymbol{\xi}])$ , em que

$$D(\boldsymbol{\theta}, \boldsymbol{\xi}) = -2\ln(\ell(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\xi})) = -2 \sum_{i=1}^N \sum_{j=1}^J \sum_{k=1}^{m_j} w_{ijk} \ln(P_{ijk})$$

é denominada *deviance*. Em  $G$  valores simulados para os parâmetros  $\boldsymbol{\theta}$  e  $\boldsymbol{\xi}$ , podemos estimar  $\mathbb{E}[D(\boldsymbol{\theta}, \boldsymbol{\xi})]$  por

$$\widehat{D}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \frac{1}{G} \sum_{g=1}^G D(\boldsymbol{\theta}^{(g)}, \boldsymbol{\xi}^{(g)}).$$

Para estimar  $D(\mathbb{E}[\boldsymbol{\theta}], \mathbb{E}[\boldsymbol{\xi}])$ , nós podemos utilizar a média da posteriori dos parâmetros, ou seja,

$$\widehat{D}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\xi}}) = D\left(\frac{1}{G} \sum_{g=1}^G \boldsymbol{\theta}^{(g)}, \frac{1}{G} \sum_{g=1}^G \boldsymbol{\xi}^{(g)}\right).$$

Então, as estimativas dos critérios DIC, EAIC e EBIC são, respectivamente,

$$\widehat{\text{DIC}} = \widehat{D}(\boldsymbol{\theta}, \boldsymbol{\xi}) + \widehat{\rho}_D = 2\widehat{D}(\boldsymbol{\theta}, \boldsymbol{\xi}) - \widehat{D}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\xi}}),$$

$$\widehat{\text{EAIC}} = \widehat{D}(\boldsymbol{\theta}, \boldsymbol{\xi}) + 2p,$$

e

$$\widehat{\text{EBIC}} = \widehat{D}(\boldsymbol{\theta}, \boldsymbol{\xi}) + p \log n,$$

em que  $p$  é o número de parâmetros do modelo e  $n = N \times J$  é o número total de observações, ou seja, o número de elementos da matriz  $\mathbf{Y}$ . Para os critérios EAIC e EBIC, respectivamente,  $2p$  e  $p \log n$  são fixados para penalizar a média a posteriori da *deviance*. Para os modelos RG e CPG,  $p = N + \sum_{j=1}^J m_j$ .

O critério WAIC é baseado em uma medida de precisão preditiva denominada *expected log point predictive density* (elpd), estimada por

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \widehat{p}_{\text{waic}},$$

em que  $\widehat{\text{lpd}}$  é denominado *computed log pointwise predictive density* e pode ser estimado por

$$\widehat{\text{lpd}} = \sum_{i=1}^N \sum_{j=1}^J \ln \left( \frac{1}{G} \sum_{g=1}^G P(Y_{ij} = y_{ij} | \boldsymbol{\theta}_i^{(g)}, \boldsymbol{\xi}_j^{(g)}) \right),$$

e  $\widehat{p}_{\text{waic}}$  é o número efetivo de parâmetros estimados definido por

$$\widehat{p}_{\text{waic}} = \sum_{i=1}^N \sum_{j=1}^J V_{g=1}^G \left( \ln \left( P(Y_{ij} = y_{ij} | \boldsymbol{\theta}_i^{(g)}, \boldsymbol{\xi}_j^{(g)}) \right) \right),$$

em que  $V_{g=1}^G a_g = \frac{1}{G-1} \sum_{g=1}^G (a_g - \bar{a})^2$ . Então, utilizando a escala da *deviance* para ser comparável aos critérios já vistos anteriormente, nós temos

$$\widehat{\text{WAIC}} = -2\widehat{\text{elpd}}_{\text{waic}}.$$

*Leave-one-out cross-validation* (LOO) é uma medida que fornece a precisão da predição de pontos fora da amostra para a abordagem Bayesiana utilizando log-verossimilhança. Uma maneira aproximada de calcular o LOO foi proposta por [Vehtari, Gelman e Gabry \(2017\)](#) através do *Pareto smoothed importance sampling* (PSIS), um método que utiliza a distribuição de Pareto e pesos de importância. A estimativa do PSIS da densidade preditiva pontual do log do LOO esperado é dada por

$$\widehat{\text{elpd}}_{\text{psis-loo}} = \sum_{i=1}^N \sum_{j=1}^J \ln \left( \frac{\sum_{g=1}^G \kappa_{ij}^{(g)} P(Y_{ij} = y_{ij} | \theta_i^{(g)}, \xi_j^{(g)})}{\sum_{g=1}^G \kappa_{ij}^{(g)}} \right),$$

em que  $\kappa_{ij}^{(g)}$  é um vetor de pesos construído a partir do PSIS. Assim como o critério WAIC, o critério LOO pode ser calculado por

$$\widehat{\text{LOO}} = -2\widehat{\text{elpd}}_{\text{psis-loo}}.$$

Considerando estes critérios, o modelo que melhor se ajustou ao conjunto de dados é o modelo que obteve o menor valor do critério. De acordo com [Spiegelhalter et al. \(2002\)](#), é recomendado utilizar de mais de um critério para comparar dois modelos.

[Burnham e Anderson \(2002\)](#) propuseram interpretar os valores dos critérios de comparação de modelos utilizando a diferença destes valores em relação ao melhor modelo em cada critério considerado. Esta diferença é dada por

$$\Delta_{\text{CRIT}h} = \text{CRIT}_h - \text{CRIT}_{\min}, \quad (2.15)$$

em que  $\text{CRIT}_h$  é o valor do critério para o modelo  $h$ , considerando  $h = 1, 2, \dots, M$ , tal que  $M$  é o número total de modelos que estão sendo ajustados e  $\text{CRIT}_{\min}$  é o valor do critério do melhor modelo ajustado. [Burnham e Anderson \(2002\)](#) apresentam valores indicativos para auxiliar na interpretação das diferenças dos valores dos critérios. Os modelos com  $0 < \Delta_{\text{CRIT}} < 2$  devem ser considerados adequados para os dados, enquanto que modelos com  $\Delta_{\text{CRIT}} > 10$  podem ser descartados. Esta maneira de interpretar os valores dos critérios de comparação de modelos é um método da estatística clássico aplicado aqui em um contexto Bayesiano.

### ***Chechagem preditiva a posteriori do modelo***

A checagem preditiva a posteriori de um modelo é um procedimento que avalia o ajuste do modelo aos dados utilizando a distribuição preditiva a posteriori. [Stern e Sinharay \(2005\)](#), [Sinharay, Johnson e Stern \(2006\)](#) e [Fox \(2010\)](#) abordam este procedimento

para modelos da TRI. A ideia é simular conjuntos de dados a partir do modelo utilizado e comparar esses dados simulados com os dados observados através de uma medida de discrepância escolhida adequadamente. Os dados simulados devem ser semelhantes aos dados observados. Uma diferença sistemática entre eles, ou seja, uma diferença persistente, indica que o modelo é inadequado para os dados ou para o contexto.

A distribuição preditiva a posteriori dos dados é dada por

$$p(\mathbf{y}^{rep}|\mathbf{y}^{obs}) = \int p(\mathbf{y}^{rep}|\vartheta) p(\vartheta|\mathbf{y}^{obs}) d\vartheta, \quad (2.16)$$

em que  $\mathbf{y}^{obs}$  é a matriz de dados observados,  $\mathbf{y}^{rep}$  é a matriz de dados simulados gerada a partir de sua distribuição preditiva a posteriori e  $\vartheta$  é o conjunto de parâmetros do modelo.

A implementação deste procedimento pode ser feita via algoritmo MCMC de forma simples. Depois que as amostras dos parâmetros do modelo são geradas pelo algoritmo MCMC a partir de suas distribuições a posteriori, as seguintes etapas devem ser realizadas para simular as réplicas preditivas usando os modelos RG e CPG:

- **Passo 1:** Gerar estimativas para os parâmetros dos itens  $a_j$  e  $b_{jk}$ , em que  $j = 1, 2, \dots, J$  e  $k = 1, 2, \dots, M_j$ , a partir da distribuição a posteriori  $p(a_j, b_{j1}, b_{j2}, \dots, b_{jM_j} | \mathbf{y}^{obs})$ .
- **Passo 2:** Gerar estimativas para os parâmetros dos indivíduos  $\theta_i^{rep}$  em que  $i = 1, 2, \dots, N$  a partir da distribuição a posteriori  $p(\theta_i | \mathbf{y}^{obs})$ .
- **Passo 3:** Utilizar os valores gerados nos passos anteriores para calcular as probabilidades preditas para os modelos RG e CPG a partir da Equação 2.6 e Equação 2.9, respectivamente.
- **Passo 4:** Simular a réplica da matriz de resposta  $\mathbf{y}^{rep}$  a partir das probabilidades preditas calculadas no **Passo 3**.

Executando estes passos por  $T$  vezes, obtemos  $T$  réplicas do conjunto de parâmetros do modelo:  $\vartheta_{(1)}, \vartheta_{(2)}, \dots, \vartheta_{(T)}$ , e  $T$  réplicas dos dados:  $\mathbf{y}_{(1)}^{rep}, \mathbf{y}_{(2)}^{rep}, \dots, \mathbf{y}_{(T)}^{rep}$ . Utilizamos estas réplicas para verificar o ajuste do modelo aos dados através do  $p$ -valor Bayesiano, definido como

$$p_D = P(D(\mathbf{y}^{rep}, \vartheta) \geq D(\mathbf{y}^{obs}, \vartheta) | \mathbf{y}^{obs}), \quad (2.17)$$

em que  $D(\cdot, \vartheta)$  é chamada de medida de discrepância. Na prática, podemos obter o  $p$ -valor Bayesiano a partir da proporção de valores nas  $T$  réplicas, tal que  $D(\mathbf{y}_{(t)}^{rep}, \vartheta_{(t)}) \geq D(\mathbf{y}^{obs}, \vartheta_{(t)})$ , para  $t = 1, 2, \dots, T$ . Dizemos que  $p_D$  é a proporção de valores da função  $D(\cdot, \vartheta)$  aplicada nas réplicas que excedem o valor de  $D(\cdot, \vartheta)$  aplicada nos dados observados. A discrepância entre o modelo ajustado e os dados é indicada por valores de  $p_D$  perto de zero ou um, uma vez que, assumindo que o modelo está bem ajustado,  $D(\mathbf{y}^{rep}, \vartheta)$

não deve ser maior ou menor que  $D(\mathbf{y}^{obs}, \vartheta)$ . De acordo com Sahu (2002), um bom ajuste implica em  $p$ -valor Bayesiano entre 0,1 e 0,9.

Sinharay, Johnson e Stern (2006) sugerem utilizar uma medida de discrepância global que considera a distribuição das categorias de resposta observadas de toda a matriz de dados, denominada distância qui-quadrada, definida por

$$D(\mathbf{y}, \vartheta) = \sum_{k=1}^{\max\{m_1, \dots, m_J\}} \frac{[F^k - \mathbb{E}(F^k)]^2}{\mathbb{E}(F^k)}, \quad (2.18)$$

em que  $F^k$  é a frequência da categoria de resposta  $k$  na matriz  $\mathbf{y}$ ,  $\mathbb{E}(F^k)$  é a frequência esperada da categoria de resposta  $k$  na matriz  $\mathbf{y}$ , calculada pela média dos conjuntos de dados replicados. Esta medida de discrepância global é utilizada para o cálculo de um único  $p$ -valor Bayesiano global através da Equação 2.17.

De maneira similar, podemos verificar a qualidade do ajuste em um particular item. Para isso, definimos a distância qui-quadrada para um item  $j$  como

$$D_j(\mathbf{y}, \vartheta) = \sum_{k=1}^{\max\{m_1, \dots, m_J\}} \frac{[F^{k,j} - \mathbb{E}(F^{k,j})]^2}{\mathbb{E}(F^{k,j})}, \quad (2.19)$$

em que  $F^{k,j}$  é a frequência da categoria de resposta  $k$  para o item  $j$ ,  $\mathbb{E}(F^{k,j})$  é a frequência esperada da categoria de resposta  $k$  para o item  $j$ , calculada a partir das  $T$  réplicas. Esta medida de discrepância é utilizada para o cálculo do  $p$ -valor Bayesiano específico para o item  $j$  através da Equação 2.17. Neste caso, podemos calcular um  $p$ -valor Bayesiano para cada item existente.

## 2.4 Estudos de simulação

Nesta seção, conduzimos dois estudos de simulação utilizando os modelos RG e CPG. No primeiro estudo, analisamos a sensibilidade dos cenários de distribuições a priori descritos nas Tabelas 2 e 3 através de medidas apropriadas para avaliar a recuperação de parâmetros. No segundo estudo, comparamos o desempenho dos cinco critérios de comparação de modelos descritos na Seção 2.3, checando o ajuste dos modelos RG e CPG com os cenários de distribuições a priori selecionados no primeiro estudo de simulação.

### ***Análise de sensibilidade dos cenários de distribuições a priori***

Em Kieftenbeld e Natesan (2012), um experimento realizado para verificar a recuperação de parâmetros do modelo RG utilizando Gibbs Sampling (MCMC) indicou que o tamanho mínimo da amostra e o número mínimo de itens para um resultado satisfatório no estudo de recuperação de parâmetros é de, respectivamente, 300 indivíduos e 15 itens. Assim, usamos o tamanho da amostra de 300 indivíduos, a quantidade de 15 itens,

cada item com quatro categorias de resposta, ou seja,  $N = 300$ ,  $J = 15$  e  $m_j = 4$ , para  $j = 1, 2, \dots, J$ .

Inicialmente, simulamos a partir de uma distribuição normal padrão os parâmetros dos indivíduos  $\theta_i$ ,  $1 \leq i \leq N$ , de uma distribuição normal truncada para valores positivos com média 1 e desvio padrão  $\sqrt{2}$  os parâmetros de inclinação  $a_j$ ,  $1 \leq j \leq J$  e de uma distribuição uniforme entre  $-3$  e  $+3$  os parâmetros  $b_{jk}$ ,  $1 \leq k \leq M_j$ . Estes valores simulados foram fixados para gerar 20 matrizes de resposta  $\mathbf{Y}$  a partir do modelo RG e 20 matrizes de resposta  $\mathbf{Y}$  a partir do modelo CPG. Cada conjunto de dados dos dois modelos foi ajustado considerando os três cenários de prioris (A1, A2 e A3 para o modelo RG, de acordo com a Tabela 2 e B1, B2 e B3 para o modelo CPG, de acordo com a Tabela 3). Em cada ajuste foi utilizada uma cadeia com 10.000 iterações, sendo as primeiras 5.000 descartadas. A convergência foi verificada pelo diagnóstico de Gelman e Rubin (GELMAN; RUBIN, 1992), indicando que a cadeia atingiu o equilíbrio satisfatoriamente.

Para comparar o desempenho das estimativas dos modelos utilizando diferentes distribuições a priori e avaliar a recuperação dos parâmetros dos itens e dos indivíduos, consideramos algumas estatísticas apropriadas. A primeira estatística utilizada foi a raiz quadrada do erro quadrático médio (*root mean square error*, RMSE), uma medida de precisão definida como

$$\text{RMSE} = \sqrt{\frac{1}{20} \sum_{r=1}^{20} (\hat{\pi}_{lr} - \pi_l)^2},$$

em que  $\pi_l$  é um parâmetro do item ou do indivíduo com  $l$  sendo um índice conveniente para este parâmetro e  $\hat{\pi}_{lr}$  sua respectiva estimativa obtida na réplica  $r$ ,  $r = 1, 2, \dots, 20$ . A segunda estatística utilizada é o viés (ou *bias*), dado por

$$\text{Bias} = \hat{\pi}_l - \pi_l,$$

em que  $\hat{\pi}_l = \frac{1}{20} \sum_{r=1}^{20} \hat{\pi}_{lr}$  e, finalmente, utilizamos o valor absoluto do viés relativo (*absolute values of relative bias*, AVR B), dado por

$$\text{AVRB} = \frac{|\hat{\pi}_l - \pi_l|}{|\pi_l|}.$$

O valor da RMSE e o valor do viés para os parâmetros dos itens e dos indivíduos são apresentados na Tabela 4 para cada modelo considerando as diferentes prioris.

Tabela 4 – Média da RMSE (viés) para os 3 cenários de prioris de cada modelo

	Modelo RG			Modelo CPG		
	A1	A2	A3	B1	B2	B3
$a_j$	0,141 (-0,046)	0,163 (-0,109)	0,136 (+0,009)	0,168 (-0,096)	0,190 (-0,065)	0,143 (-0,036)
$b_{jk}$	0,189 (+0,013)	0,289 (-0,061)	0,213 (+0,020)	0,261 (-0,039)	0,249 (+0,002)	0,199 (+0,004)
$\theta_i$	0,343 (+0,032)	0,343 (+0,025)	0,339 (+0,028)	0,292 (+0,025)	0,309 (+0,026)	0,285 (+0,028)

Os valores da RMSE foram relativamente pequenos para os parâmetros de inclinação em todas as prioris, com o cenário de prioris B2 sendo o pior caso. Para os parâmetros de dificuldade, os piores valores da RMSE foram relacionados às distribuições a priori A2, B1 e B2, todos acima de 0,24. Os valores da RMSE dos parâmetros dos indivíduos foram semelhantes, com o cenário de prioris A3 melhor para o modelo RG e B3 melhor para o modelo CPG. Ao realizar a análise do viés, os cenários de prioris A2, B1 e B2 mostraram alguns valores indesejáveis (distantes de zero), enquanto os cenários A3 e B3 apresentaram viés próximo de zero para os parâmetros dos itens e dos indivíduos. Em geral, a partir dos valores apresentados na Tabela 4, concluímos que os dois modelos são insensíveis aos cenários de prioris especificados, isto é, a escolha da distribuição a priori em ambos os modelos pode ser feita arbitrariamente. No entanto, notamos que a priori B2 para o modelo CPG apresenta um valor maior da medida RMSE para o parâmetro de inclinação do que outras prioris, indicando uma maior variabilidade em suas estimativas.

Os resultados referentes ao AVR<sub>B</sub> são apresentados a partir de dezoito *boxplots* na Figura 3.

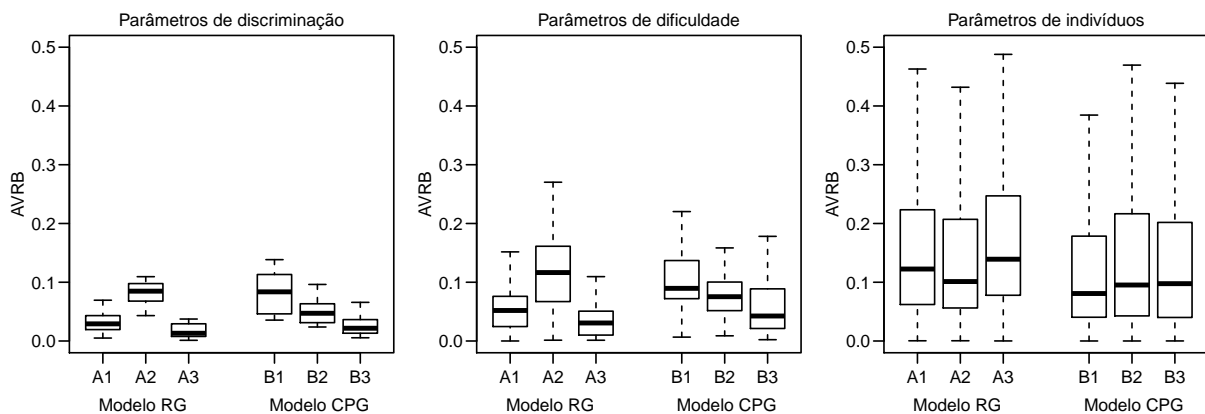


Figura 3 – *Boxplots* da distribuição da medida AVR<sub>B</sub> para os parâmetros de discriminação, dificuldade e indivíduos relativo aos cenários de prioris do modelo RG (A1, A2 e A3) e do modelo CPG (B1, B2 e B3).

Observando os valores desta medida referentes aos parâmetros de discriminação e dificuldade, verificamos que, entre os cenários de prioris do modelo RG, A2 tem o maior valor para AVR<sub>B</sub> e A3 tem o menor valor para AVR<sub>B</sub>. Do mesmo modo, entre os cenários do modelo CPG, B1 tem valores mais elevados do que os outros cenários de prioris. Em contraste, considerando os parâmetros de indivíduos, vemos que os cenários A2 e B1 possuem menores valores de AVR<sub>B</sub> em comparação com os demais cenários pertencentes aos modelos RG e CPG, respectivamente. Como a diferença de AVR<sub>B</sub> em cada cenário é pequena, podemos considerar que estes resultados indicam uma similaridade entre os valores de AVR<sub>B</sub> dos parâmetros para todos os cenários de prioris de ambos modelos.

Além da abordagem apresentada anteriormente para verificar a recuperação de



parâmetros em cada cenário, utilizamos os cinco critérios de comparação de modelos descritos na subseção 2.3 para comparar o ajuste do modelo RG com os cenários de prioris A1, A2 e A3 e do modelo CPG com cenários de prioris B1, B2 e B3. Os valores dos critérios são apresentados na Tabela 5 juntamente com a abordagem de Burnham e Anderson (2002) e Spiegelhalter *et al.* (2002) para nos ajudar na interpretação.

Tabela 5 – Valores dos critérios de comparação de modelos para os 3 cenários considerados para os modelos RG e CPG

	Modelo RG			Modelo CPG		
	A1	A2	A3	B1	B2	B3
DIC	13987,87	13990,87	13985,99	13355,84	14350,77	13354,49
$\Delta_{DIC}$	1,88	4,88	0,00	1,35	996,28	0,00
EAIC	14658,68	14664,62	14652,11	14038,06	14887,19	14032,35
$\Delta_{EAIC}$	6,57	12,51	0,00	5,71	854,84	0,00
EBIC	18535,37	18541,21	18528,80	17914,75	18763,88	17909,04
$\Delta_{EBIC}$	6,57	12,41	0,00	5,71	854,84	0,00
WAIC	14060,56	14063,52	14060,50	13444,27	14455,07	13443,64
$\Delta_{WAIC}$	0,06	3,02	0,00	0,63	1011,43	0,00
LOO	14068,51	14071,18	14068,43	13459,94	14551,16	13459,89
$\Delta_{LOO}$	0,08	2,75	0,00	0,05	1091,27	0,00

Em geral, os valores dos critérios em cada cenário de prioris foram próximos em ambos os modelos, atingindo, em alguns casos, diferenças muito pequenas, como podemos ver, por exemplo, pelo valor de  $\Delta_{WAIC} = 0,06$  para o cenário de prioris A1 no modelo RG e  $\Delta_{LOO} = 0,05$  para o cenário de prioris B1 no modelo CPG. A única exceção foi o cenário de prioris B2 no modelo CPG, que obteve  $\Delta_{EAIC} = \Delta_{EBIC} = 854,84$  como a menor diferença em relação ao menor valor de critério e, portanto, pode ser desconsiderada. Possivelmente isto ocorre porque o cenário B2 possui valores altos para o desvio padrão das distribuições, tornando as prioris menos informativas. O desempenho ligeiramente melhor do cenário A3 para o modelo RG e B3 para o modelo CPG fornece evidências de que são os melhores cenários de prioris e, portanto, são os cenários de prioris escolhidos para o próximo estudo de simulação e para a aplicação deste capítulo.

### ***Desempenho dos critérios de comparação de modelos frente aos modelos RG e CPG***

Realizamos um estudo do desempenho dos critérios de comparação de modelos descritos na seção 2.3 para a escolha do modelo apropriado. Para desenvolver essa análise, optamos por utilizar as prioris A3 e B3 para os modelos RG e CPG, respectivamente, com base na análise de sensibilidade apresentada. Consideramos 15 cenários baseados em Kieftenbeld e Natesan (2012) construídos variando o número de indivíduos  $N \in \{50, 75, 150, 300, 500\}$  e o número de itens  $J \in \{7, 15, 30\}$ . Para cada um dos 15 cenários

rios, 50 matrizes  $\mathbf{Y}$  foram geradas a partir do modelo RG, então tanto o modelo RG como o modelo CPG foram ajustados com estes dados e os cinco critérios de comparação de modelos foram calculados para observar o comportamento dessas medidas nos cenários estabelecidos. O mesmo procedimento foi conduzido para conjuntos de dados simulados a partir do modelo CPG.

A Figura 4 apresenta 10 gráficos do tipo *heat map* (mapa de calor) contendo a porcentagem de vezes que cada critério de comparação de modelos indicou o modelo correto (isto é, o modelo a partir do qual os dados foram simulados) considerando os 15 cenários. Este percentual foi denominado percentual de escolha correta (PEC). Assim, em cada gráfico, fixamos o critério de comparação de modelos (DIC, EAIC, EBIC, WAIC e LOO) e o modelo utilizado para gerar a matriz de respostas  $\mathbf{Y}$  (modelo RG e CPG), confrontando o PEC em cada cenário.

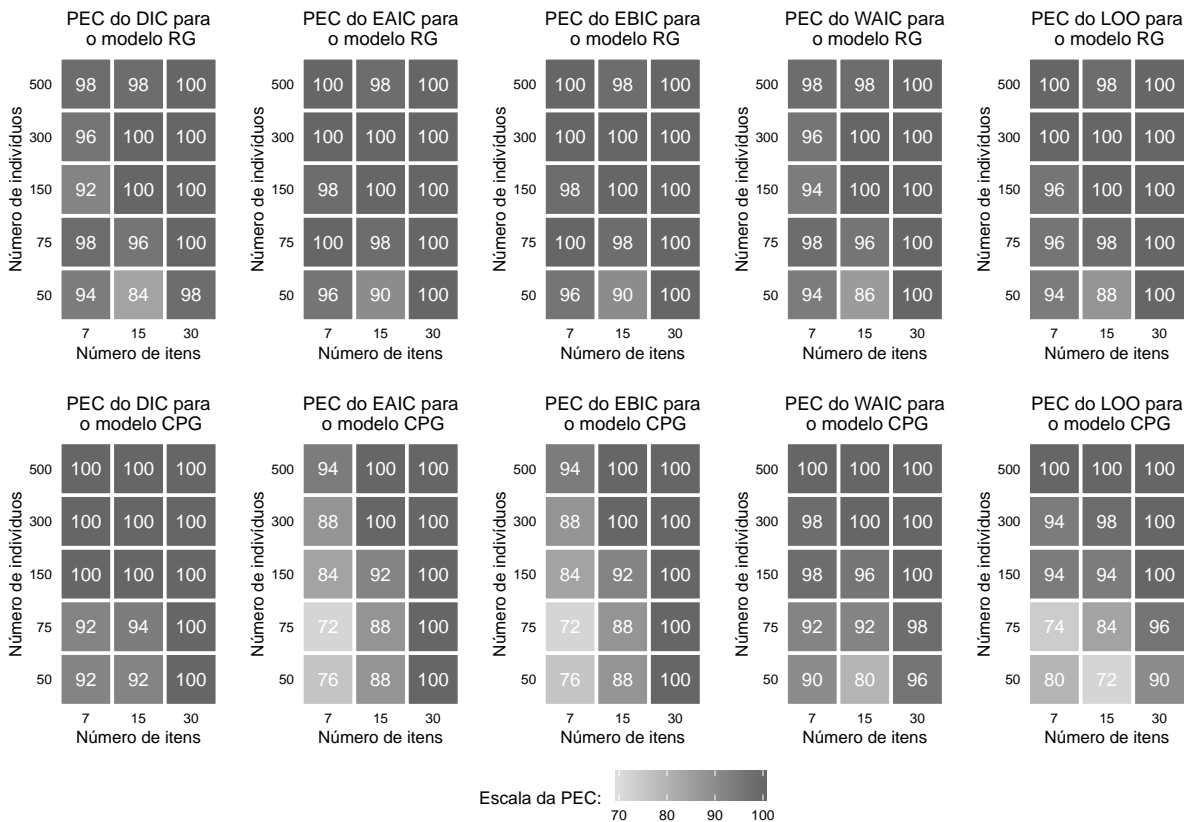


Figura 4 – PEC de cada critério de comparação de modelo apresentado em um *heat map* (mapa de calor).

Em geral, à medida que o número de indivíduos e itens aumentam, o PEC do modelo também aumenta em todos os critérios, de tal forma que, para  $N \geq 150$  e  $J \geq 30$ , os cinco critérios examinados podem ser utilizados indistintamente. No entanto, é evidente que os critérios são mais sensíveis ao número de itens do que ao número de indivíduos, porque quando  $J$  é pequeno, independente do valor de  $N$ , os percentuais são substancialmente menores. Além disso, para  $N$  pequeno e  $J$  pequeno, os critérios de comparação de

modelos são instáveis, pois o PEC chega a oscilar abaixo de 80% em alguns critérios.

De acordo com os valores do PEC, podemos dizer que para o cenário em que  $N = 50$  e  $J = 7$ , o índice DIC indica corretamente o modelo RG em 94% dos casos e o modelo CPG em 92% dos casos. Considerando o menor PEC, podemos dizer que temos uma probabilidade de selecionar o modelo correto de 92% utilizando o índice DIC para o cenário  $N = 50$  e  $J = 7$ . Analogamente, considerando o mesmo cenário, podemos dizer que os critérios EAIC e EBIC apontam para o melhor modelo com uma probabilidade de 24% de falha. Esta ideia aplica-se a outros critérios e cenários.

Para o cenário com  $N = 50$  e  $J = 7$ , o critério DIC apresentou o melhor desempenho, seguido pelo critério WAIC. No cenário com  $N = 50$  e  $J = 15$ , os melhores critérios foram EAIC e EBIC, ambos com a mesma PEC. Finalmente, para o cenário com  $N = 150$  e  $J = 7$ , os critérios WAIC e LOO obtiveram o melhor desempenho, seguido do critério DIC. Além disso, não houve um critério de comparação de modelos que destacou-se para  $\mathbf{Y}$  simulado a partir do modelo RG. Em contraste, para  $\mathbf{Y}$  simulado a partir do modelo CPG, o índice DIC apresentou o PEC maior do que os demais critérios em cenários com  $N$  e  $J$  pequenos. Em resumo, o critério DIC é um bom critério, como tem sido utilizado a prática em modelos da TRI.

## 2.5 Aplicação

Nesta seção, apresentamos uma aplicação para ilustrar a comparação entre os modelos RG e CPG utilizando dados do Inventário da Depressão de Beck (Beck Depression Inventory, BDI), ver [Beck et al. \(1961\)](#). O Dr. Teng Chei-Tung do Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, forneceu um conjunto de dados com 1.111 respostas de estudantes coletadas a partir de uma versão do BDI com 21 itens que medem características de depressão (sintomas e atitudes) com intensidade variando de neutro a um nível máximo de severidade, classificado em 1 (baixa intensidade) a 4 (alta intensidade). Os dados podem ser encontrados em [Fragoso e Cúri \(2013\)](#).

A Figura 5 contém a descrição dos 21 itens que compõem o BDI e a frequência relativa (em porcentagem) das respostas dos 1.111 indivíduos para cada categoria de resposta apresentada em um *heat map* (mapa de calor).

As categorias de resposta 1 e 2, que são referentes à baixa intensidade, são as mais escolhidas. A frequência média destas duas categorias de resposta é de 91%. Em contraste, a categoria de resposta 4 é a menos escolhida, com uma frequência média de 1,2%. Apenas os itens 1 (Tristeza), 8 (Inferioridade) e 17 (Cansaço) possuem maior frequência na categoria de resposta 2.

Ajustamos os modelos RG e CPG a este conjunto de dados do BDI através da

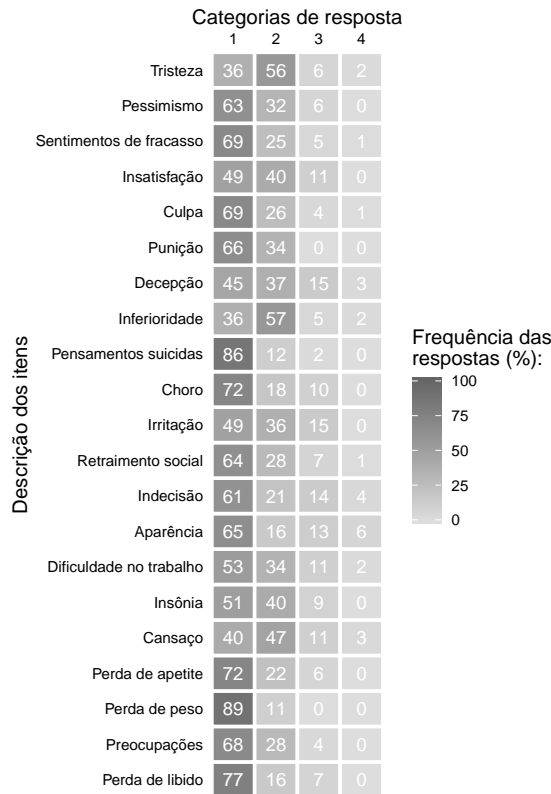


Figura 5 – Descrição dos 21 itens que compõem o BDI e a frequência relativa das respostas de cada item apresentada em um *heat map* (mapa de calor).

abordagem Bayesiana descrita na Seção 2.3. Apesar de sabermos que os parâmetros dos modelos RG e CPG não são diretamente comparáveis (OSTINI; NERING, 2006), apresentamos as estimativas dos parâmetros de discriminação e dos parâmetros de indivíduos de forma a confrontar seus valores. A Figura 6 mostra as estimativas dos parâmetros de discriminação e os intervalos de credibilidade de 95% de ambos modelos.

Podemos observar que as estimativas dos parâmetros de discriminação do modelo RG foram maior do que do modelo CPG em 18 dos 21 itens. Além disso, as estimativas dos parâmetros de discriminação do item 19 do modelo RG e dos itens 14 e 21 do modelo CPG ficaram abaixo do valor de referência 0,6 (SANTOS; AZEVEDO; BOLFARINE, 2013), que, apesar de ser um valor de referência tipicamente utilizado em modelos dicotômicos com função de ligação proibito, está sendo utilizado aqui para indicar os itens que não discriminam adequadamente o traço latente avaliado pelo BDI.

As Figuras 7 e 8 apresentam as estimativas dos parâmetros de dificuldade e os intervalos de credibilidade de 95% dos modelos RG e CPG, respectivamente, para os dados de BDI considerados nesta aplicação. De acordo com as especificações do modelo RG na Seção 2.2, um item  $j$  do BDI possui os parâmetros de dificuldade  $b_{jk}$ , para  $k = 1, 2, 3$ . Estes parâmetros correspondem ao valor do traço latente tal que a probabilidade de um indivíduo selecionar uma categoria de resposta superior à  $k$ -ésima é 0,5. Por exemplo, sob

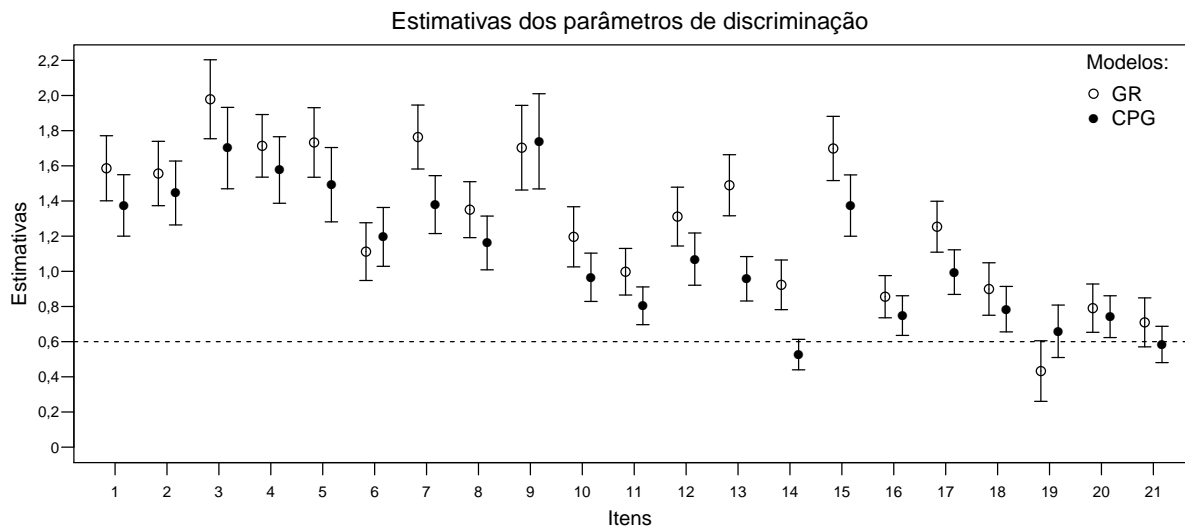


Figura 6 – Estimativas dos parâmetros de discriminação e intervalos de credibilidade de 95% dos modelos RG e CPG.

o modelo RG, o item 1 possui as dificuldades  $-2,25$ ,  $0,35$  e  $1,45$ . Assim, podemos dizer que um indivíduo com traço latente  $\theta = -2,25$  terá probabilidade  $0,5$  de selecionar as categorias 2, 3 ou 4 no item 1. Da mesma forma, um indivíduo com traço latente  $\theta = 0,35$  terá probabilidade  $0,5$  de selecionar as categorias 3 ou 4 no item 1.

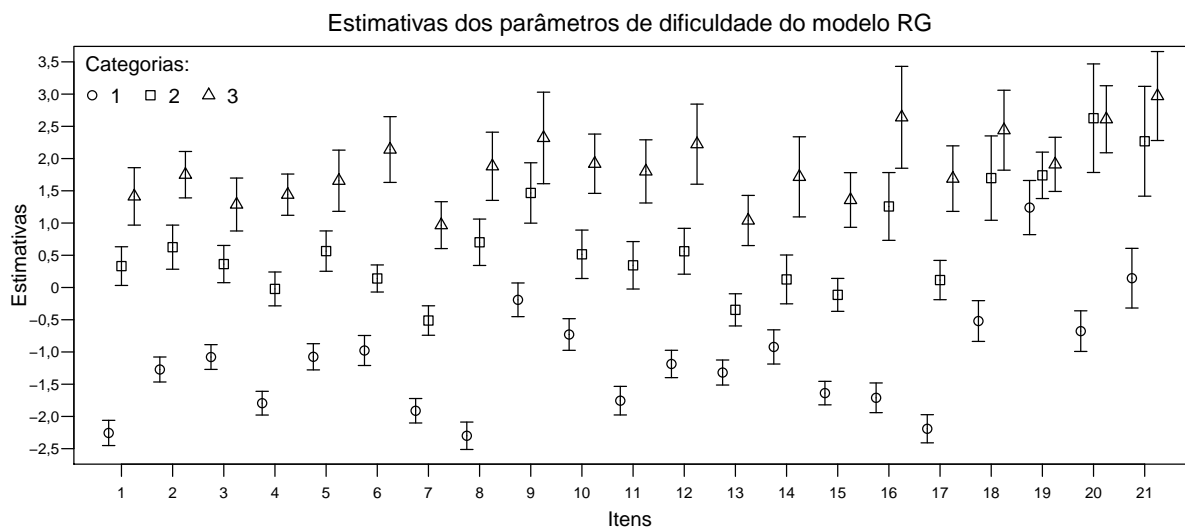


Figura 7 – Estimativas dos parâmetros de dificuldade e intervalos de credibilidade de 95% para o modelo RG.

Em contraste, considerando o modelo CPG, um item  $j$  do BDI possui os parâmetros de dificuldade  $b_{jk}$ , para  $k = 1, 2, 3, 4$ , tal que  $b_{j1} = 0$ . Assim como no modelo RG, que permite interpretar os valores dos parâmetros de dificuldade associando-os aos valores dos traços latentes dos indivíduos, no modelo CPG podemos relacionar parâmetro de dificuldade e traço latente de um indivíduo com as probabilidades deste indivíduo selecionar as

categorias de resposta de um determinado item, tal que

$$P_{ijk} < P_{ij[k-1]}, \text{ se } \theta_i < b_{jk}$$

$$P_{ijk} = P_{ij[k-1]}, \text{ se } \theta_i = b_{jk}$$

$$P_{ijk} > P_{ij[k-1]}, \text{ se } \theta_i > b_{jk}$$

Ou seja, a relação de ordem entre o traço latente do indivíduo e o valor da dificuldade referente à  $k$ -ésima categoria de resposta de um item influencia na relação de ordem das probabilidades adjacentes  $P_{ij[k-1]}$  e  $P_{ijk}$ . Por exemplo, sob o modelo CPG, o item 1 possui as dificuldades  $-2,25$ ,  $0,55$  e  $0,4$ . Então, podemos dizer que um indivíduo com traço latente igual a  $-2,25$  terá a mesma probabilidade de selecionar a categoria 1 e a categoria 2, enquanto que um indivíduo com traço latente inferior a  $-2,25$  terá probabilidade maior em selecionar a categoria 1 do que a categoria 2.

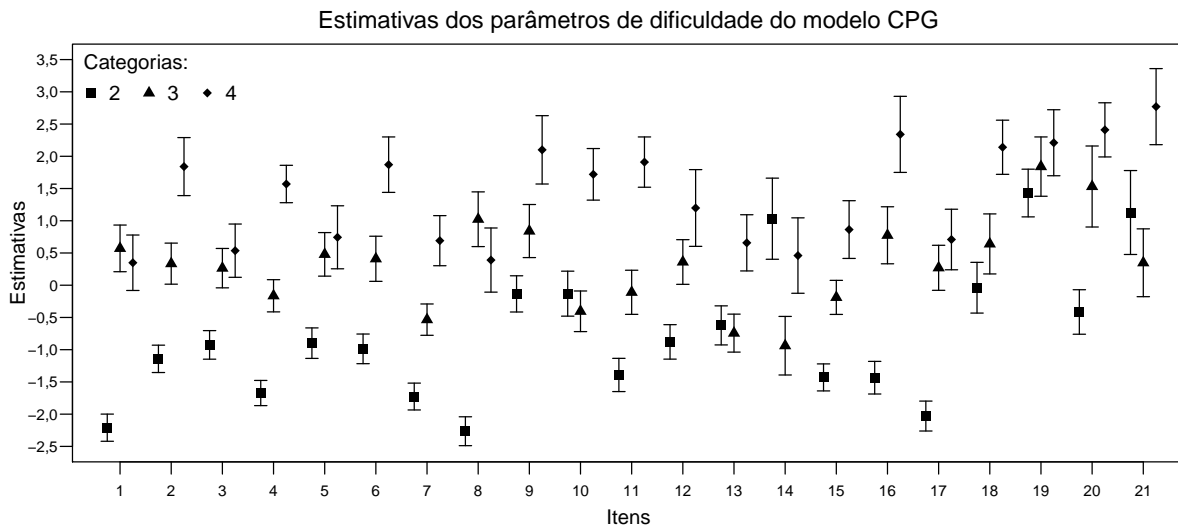


Figura 8 – Estimativas dos parâmetros de dificuldade e intervalo de credibilidade de 95% para o modelo CPG.

Embora as estimativas dos parâmetros dos modelos RG e CPG não tenham relação devido à natureza de ambos os modelos (OSTINI; NERING, 2006), conceitualmente as estimativas dos traços latentes dos indivíduos obtidas a partir dos modelos RG e CPG são comparáveis, pois explicam o padrão de respostas dadas pelos indivíduos. Dessa forma, apresentamos na Figura 9 uma comparação entre as estimativas dos parâmetros de indivíduos a partir de um gráfico de dispersão. A bissetriz dos quadrantes ímpares foi colocada na Figura 9 como referência ao leitor. Pontos que estão próximos à bissetriz indicam que as estimativas em ambos modelos são próximas. Em geral, as estimativas dos traços latentes em ambos modelos são próximas. Entretanto, podemos observar que todos os pontos com valor acima de 1,5 (no eixo horizontal) estão abaixo da bissetriz, indicando que o modelo RG forneceu estimativas maiores do que o modelo CPG para estes traços latentes.

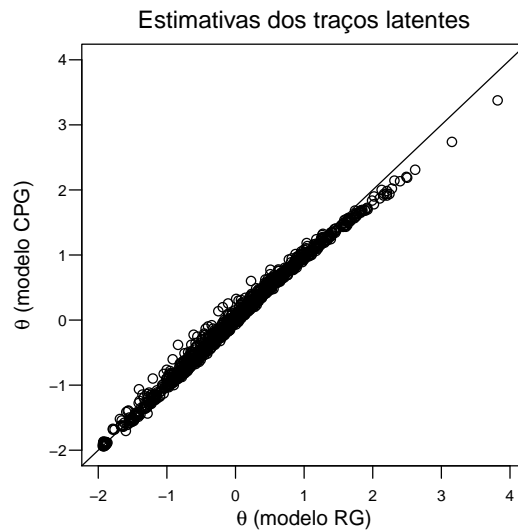


Figura 9 – Gráfico de dispersão para as estimativas de traços latentes (modelos RG e CPG). Devido à diferença na natureza de ambos os modelos, o significado dos traços latentes em cada modelo não é o mesmo.

Para verificar qual modelo é mais adequado para ser utilizado na análise dos dados do BDI, nós ajustamos os modelos RG e CPG e, em seguida, calculamos os valores dos critérios de comparação de modelos discutidos na seção 2.3. A Tabela 6 apresenta estes valores dos critérios e também o  $p$ -valor Bayesiano global para cada modelo. Como podemos observar, os 5 critérios indicaram o modelo RG como sendo mais adequado aos dados do que o modelo CPG. As diferenças entre os valores dos critérios,  $\Delta_{\text{CRIT}}$ , são consideráveis, o que evidencia a adequabilidade do modelo RG para os dados de BDI desta aplicação. Além disso, o  $p$ -valor Bayesiano global do modelo RG, 0,283, pertence ao intervalo  $[0, 1; 0, 9]$ , indicando, de acordo com Sahu (2002), um bom ajuste do modelo.

Tabela 6 – Valores dos cinco critérios de comparação de modelos,  $p$ -valor Bayesiano global e tempo, em minutos, do algoritmo NUTS para os modelos RG e CPG ajustados aos dados de BDI.

	DIC	EAIC	EBIC	WAIC	LOO	$p$ -valor	Tempo
Modelo RG	33881,23	35332,76	44961,52	33994,71	34002,45	0,283	113,69
Modelo CPG	33898,41	35371,39	45000,15	34029,84	34042,19	0,204	169,31
$\Delta_{\text{CRIT}}$	17,17	38,63	38,63	35,13	39,74	–	–

Os resultados da análise de ajuste dos itens para os modelos RG e CPG, baseada na distância qui-quadrada apresentada na Equação 2.19, são apresentados na Figura 10.

Podemos observar que a maioria dos itens possui  $p$ -valor Bayesiano entre 0,1 e 0,9 em ambos modelos, o que indica, segundo (SAHU, 2002), que os modelos são bem ajustados para a maioria dos itens. No modelo RG, os itens 8, 13 e 14 possuem  $p$ -valor acima de 0,9 e no modelo CPG, os itens 14 e 17 possuem  $p$ -valor acima de 0,9, o que pode indicar, por exemplo, má formulação destes itens.

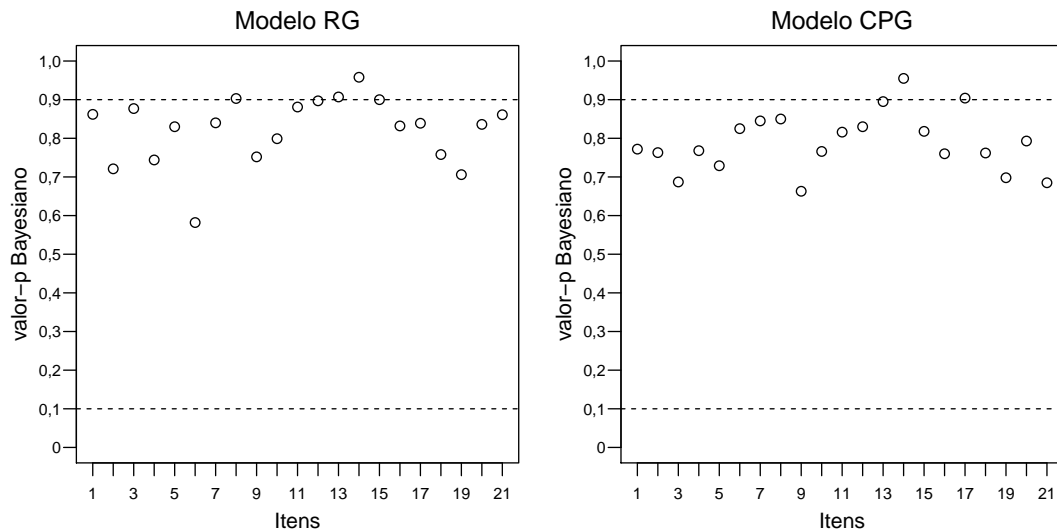


Figura 10 –  $p$ -valor Bayesiano por item (modelos RG e CPG) com base na distância qui-quadrada.

## 2.6 Conclusões e comentários

Neste capítulo, realizamos um estudo comparativo de dois modelos da TRI para itens politômicos: o modelo de resposta gradual (RG) e o modelo de crédito parcial generalizado (CPG). Estes modelos foram escolhidos para este estudo por serem os modelos mais conhecidos e utilizados dentre os modelos da TRI para itens politômicos com categorias de resposta ordenadas. A falta de um estudo que auxilie os usuários na escolha entre estes dois modelos torna esta decisão incerta, como abordamos na Seção 2.5 para o teste BDI.

Conduzimos dois estudos de simulação utilizando o algoritmo NUTS para estimar os parâmetros dos modelos RG e CPG. No primeiro estudo, realizamos uma análise de sensibilidade das distribuições a priori para escolher um cenário de prioris para cada modelo. Em geral, através das estatísticas consideradas para avaliar a recuperação dos parâmetros, observamos que o cenário de distribuições a priori para o modelo RG, proposto por [Curtis \(2010\)](#) resultou em uma precisão ligeiramente maior nas estimativas dos parâmetros em relação aos demais cenários. Para o modelo CPG, o cenário de prioris proposto por [Li e Baser \(2012\)](#) também resultou em uma precisão ligeiramente maior nas estimativas dos parâmetros. Uma possível justificativa para que estes cenários de prioris tenham fornecido estimativas ligeiramente mais precisas é o fato de que, em geral, essas distribuições tem desvio padrão menor do que as outras distribuições consideradas. Além disso, os valores dos cinco critérios de comparação de modelos para essas duas prioris foram ligeiramente inferiores aos demais, o que nos levou a utilizá-los no segundo estudo de simulação e na aplicação deste capítulo. A única exceção é o cenário de prioris proposto por [Zhu, Robinson e Torenvlied \(2014\)](#) para o modelo CPG, pois os critérios de comparação de modelos



indicaram que o modelo CPG com este cenário de prioris não é adequado se comparado com os cenários de prioris de [Li e Baser \(2012\)](#) e [Curtis \(2010\)](#).

No segundo estudo de simulação, realizamos uma análise de desempenho dos cinco critérios de comparação de modelos considerados neste capítulo. Para isso, simulamos 50 amostras de cada modelo, ajustamos estas amostras utilizando ambos os modelos com as prioris selecionadas no primeiro estudo de simulação e calculamos os 5 critérios de comparação de modelos para cada amostra ajustada. O esperado seria que os critérios indicassem o modelo que inicialmente simulou as amostras, embora isso não tenha ocorrido com 100% frequência para todas as amostras. No entanto, as amostras com  $N$  e  $J$  grandes foram associadas ao modelo correto em 100% dos casos pelos cinco critérios de comparação de modelos. Além disso, pode-se ver que quando o número de itens é baixo, os critérios “acertam” o modelo com uma frequência mais baixa.

Este estudo mostra o uso dos cinco critérios de comparação de modelos para selecionar entre dois modelos diferentes da TRI para dados politômicos: o modelo RG e o modelo CPG. Embora esses dois modelos não estejam aninhados, isto é, ambos os modelos não usam parâmetros diretamente comparáveis, concluímos que alguns dos critérios de comparação de modelos investigados são adequados e confiáveis para serem utilizados na escolha do modelo mais apropriado. Assim, quando os pesquisadores e profissionais estão escolhendo entre os modelos, não há necessidade de fazer uma escolha apenas arbitrária ou subjetiva. Em vez disso, ao trabalhar com dados politômicos ordenados, recomendamos ajustar tanto o modelo RG quanto o modelo CPG e comparar os ajustes com os critérios DIC, WAIC e LOO. Com base nos estudos, não recomendamos o uso dos critérios EAIC e EBIC para comparar os modelos, porque eles tiveram um desempenho inferior nos cenários com um número pequeno de indivíduos e itens.



---

# UM MODELO DE CRÉDITO PARCIAL GENERALIZADO BIFATOR COM DIFERENTES FUNÇÕES DE LIGAÇÃO

---

---

## 3.1 Introdução

É comum encontrarmos na prática situações que demandam a suposição de multidimensionalidade para os modelos da teoria da resposta ao item (TRI). Uma das primeiras apresentações formais de um modelo da teoria da resposta ao item multidimensional (TRIM) é feita por Samejima (1974) com a proposta de um modelo para itens de resposta contínua. Algumas propostas iniciais de modelos da TRIM para itens dicotômicos podem ser encontradas em Sympson (1978) e Whitely (1980). As extensões do modelo de resposta gradual (modelo RG) de Samejima (1969) e do modelo de crédito parcial generalizado (modelo CPG) de Muraki (1992) para o caso multidimensional podem ser encontradas em Muraki e Carlson (1995) e Yao e Schwarz (2006a), respectivamente.

Segundo Toland *et al.* (2017), podemos classificar os modelos da TRIM em 4 diferentes categorias de acordo com as diferentes estruturas multidimensionais dos traços latentes. Os modelos pertencentes à primeira categoria são chamados por Sheng e Wikle (2007) de modelos da TRI multi-unidimensionais e são caracterizados por utilizar separadamente uma estrutura unidimensional a cada grupo de itens, denominado de subteste. Os modelos pertencentes à segunda categoria são os modelos da TRIM hierárquicos (SHENG; WIKLE, 2008), pois correlacionam um traço latente global (chamado de traço de segunda ordem) com traços específicos (chamados de traços de primeira ordem), formando uma hierarquia entre eles. A terceira categoria é composta pelos modelos da TRIM aditivos (SHENG; WIKLE, 2009; MARTELLI; MATTEUCCI; MIGNANI, 2016) e são definidos a partir de uma estrutura aditiva entre o traço latente global e os traços latentes específi-

cos. A quarta categoria é composta pelos modelos da TRI bifatores (REISE; MORIZOT; HAYS, 2007; REISE, 2012; TOLAND *et al.*, 2017), que são os modelos da TRIM aditivos com a restrição de ortogonalidade entre todos os traços latentes considerados. Ou seja, os modelos da TRI bifatores possuem uma estrutura aditiva entre o traço latente global e os traços latentes específicos e possuem as correlações entre os traços latentes definidas como zero.

Os modelos da TRI bifatores associam biunivocamente os traços latentes específicos aos subtestes, ou seja, cada traço latente específico é associado com um, e apenas um, subteste. Assim, as respostas dos indivíduos a cada item são diretamente influenciadas tanto pelo traço latente global quanto pelo traço latente específico correspondente ao subteste do item. Ao contrário dos modelos da TRIM hierárquicos, os modelos da TRI bifatores possuem traços latentes global e específicos definidos no mesmo nível conceitual, ou seja, os traços latentes específicos não são definidos a partir do traço latente global. De acordo com Chen, West e Sousa (2006) e Immekus e Imbrie (2008), os modelos da TRI bifatores facilitam a interpretação da influência do traço latente global em cada item e em cada subteste, devido à independência entre os traços latentes.

Os modelos da TRI bifatores para dados dicotômicos são cada vez mais utilizados em pesquisas de medição em educação e psicologia (GIBBONS; RUSH; IMMEKUS, 2009; RIJMEN, 2010; REISE, 2012). Entretanto, os modelos da TRI bifatores para dados politômicos ainda são pouco explorados na literatura. Gibbons *et al.* (2007) propuseram a extensão do modelo RG na estrutura bifator (modelo RG-bifator). Toland *et al.* (2017) apresentaram o modelo RG-bifator em detalhes, fornecendo uma ilustração do seu uso através de uma aplicação à dados reais. Martelli, Matteucci e Mignani (2016) apresentam um modelo da TRIM aditivo de resposta gradual em que, em alguns cenários do estudo de simulação, os autores definem as correlações entre os traços latentes como zero, tornando o modelo proposto um modelo RG-bifator.

É notável a ausência de estudos que propõem a extensão para uma estrutura bifator de outros modelo da TRI para dados politômicos, como, por exemplo, o modelo CPG. Dessa forma, propomos neste capítulo um modelo de crédito parcial generalizado bifator (modelo CPG-bifator) como um modelo alternativo para o caso de dados politômicos. Além disso, como foi apresentado no Capítulo 2, o modelo CPG foi originalmente desenvolvido considerando a função de ligação logito (MURAKI, 1992). Assim, flexibilizamos o modelo CPG-bifator para o seu uso com outras funções de ligação, como proibito e clog-log. Para a estimação dos parâmetros, utilizamos uma abordagem Bayesiana através do algoritmo No-U-Turn Sampler (NUTS), uma extensão do método de Monte Carlo Hamiltoniano (HMC) descrito em Hoffman e Gelman (2014).

Este capítulo está organizado da seguinte forma: nesta seção apresentamos uma visão geral dos modelos da TRI bifatores e a proposta deste capítulo. Na Seção 3.2, apre-

sentamos formalmente o modelo CPG-bifator com diferentes funções de ligação. Na Seção 3.3, apresentamos informações sobre a estimação dos parâmetros do modelo CPG-bifator através de uma abordagem Bayesiana. Na Seção 3.4, realizamos um estudo de simulação para verificar a recuperação dos parâmetros do modelo. Na Seção 3.5, apresentamos uma aplicação do modelo proposto neste capítulo para ilustrar o seu uso na prática. Por fim, na Seção 3.6, apresentamos algumas conclusões e comentários finais.

## 3.2 O modelo CPG-bifator

Para apresentar o modelo CPG-bifator e o desenvolvimento para o seu uso com outras funções de ligação, consideramos a seguinte situação: (i)  $N$  indivíduos que respondem aos itens de um teste composto de  $D$  subtestes; (ii) cada subteste  $d$  contém  $J_d$  itens e é designado a avaliar a  $d$ -ésima dimensão do traço latente dos indivíduos; (iii) o item  $j$  do subteste  $d$  é composto por  $m_{jd}$  categorias de resposta. Assim, definimos a variável aleatória  $Y_{ijd}$  como a resposta do indivíduo  $i$  ao item  $j$  do subteste  $d$ , tal que

$$Y_{ijd} | \theta_{0i}, \theta_{di}, \xi_{jd} \sim \text{Categorizada}(P_{ijd1}, P_{ijd2}, \dots, P_{ijdm_{jd}}), \quad (3.1)$$

em que  $\theta_{0i}$  é o traço latente global do  $i$ -ésimo indivíduo,  $\theta_{di}$  é o traço latente específico da dimensão  $d$  do  $i$ -ésimo indivíduo,  $\xi_{jd} = (a_{0jd}, a_{jd}, b_{jd1}, b_{jd2}, \dots, b_{jdm_{jd}})'$  é o vetor de parâmetros associado ao item  $j$  do subteste  $d$  tal que  $a_{0jd}$  e  $a_{jd}$  são, respectivamente, parâmetros de discriminação global e específico do item  $j$  do subteste  $d$ ,  $b_{jdk}$  é o parâmetro de dificuldade referente à  $k$ -ésima categoria de resposta do item  $j$  do subteste  $d$  e  $P_{ijdk}$  é a probabilidade do indivíduo  $i$  responder a  $k$ -ésima categoria de resposta do item  $j$  do subteste  $d$ , para  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, J_d$ ,  $d = 1, 2, \dots, D$  e  $k = 1, 2, \dots, m_{jd}$ . A distribuição Categorizada apresentada na Expressão 3.1 é definida na Seção 2.2 e exemplificada na Tabela 1.

De acordo com Muraki (1992), o modelo CPG unidimensional é formulado partindo do pressuposto de que para cada uma das categorias de resposta sucessivas, a probabilidade de selecionar a  $k$ -ésima categoria sobre a categoria anterior é definida a partir de um modelo dicotômico. Visando estender o modelo de Muraki (1992) para uma estrutura bifator com funções de ligação flexíveis, definimos esta probabilidade de maneira geral por meio de uma função de distribuição acumulada (FDA), denotada por  $F(\cdot)$ , tal que

$$P(Y_{ijd} = k | Y_{ijd} = k - 1) = \frac{P_{ijdk}}{P_{ijd[k-1]} + P_{ijdk}} = F(\eta_{ijdk}) \quad (3.2)$$

para  $k = 2, 3, \dots, m_{jd}$ , em que  $\eta_{ijdk}$  é chamado de preditor linear latente e é definido como

$$\eta_{ijdk} = a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk}. \quad (3.3)$$

A partir do desenvolvimento da última igualdade da Equação 3.2, temos que

$$P_{ijk} = \frac{F(\eta_{ijk})}{1 - F(\eta_{ijk})} \cdot P_{ijd[k-1]}, \quad (3.4)$$

para  $k = 2, 3, \dots, m_{jd}$ . Assim, definimos

$$\lambda_k = \frac{F(\eta_{ijk})}{1 - F(\eta_{ijk})}.$$

Efetuada a substituição na Equação 3.4, obtemos  $P_{ijk} = \lambda_k \cdot P_{ijd[k-1]}$ . Com isso,

$$\begin{aligned} P_{ijd2} &= \lambda_2 \cdot P_{ijd1} \\ P_{ijd3} &= \lambda_3 \cdot P_{ijd2} = \lambda_3 \cdot \lambda_2 \cdot P_{ijd1} \\ P_{ijd4} &= \lambda_4 \cdot P_{ijd3} = \lambda_4 \cdot \lambda_3 \cdot \lambda_2 \cdot P_{ijd1} \\ &\vdots \\ P_{ijdm_j} &= \lambda_{m_j} \cdot P_{ijd[m_j-1]} = \lambda_{m_j} \cdot \lambda_{m_j-1} \cdot \dots \cdot \lambda_2 \cdot P_{ijd1}. \end{aligned} \quad (3.5)$$

Como  $\sum_{k=1}^{m_{jd}} P_{ijk} = 1$ , temos que  $P_{ijd1} \cdot \left(1 + \sum_{l=2}^{m_{jd}} \prod_{h=2}^l \lambda_h\right) = 1$ . Isolando  $P_{ijd1}$ , encontramos

$$P_{ijd1} = \frac{1}{\left(1 + \sum_{l=2}^{m_{jd}} \prod_{h=2}^l \lambda_h\right)} = \frac{\lambda_1}{\lambda_1 \cdot \left(1 + \sum_{l=2}^{m_{jd}} \prod_{h=2}^l \lambda_h\right)} = \frac{\lambda_1}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l \lambda_h}.$$

Substituindo  $P_{ijd1}$  nas igualdades 3.5, obtemos

$$\begin{aligned} P_{ijd2} &= \frac{\lambda_1 \cdot \lambda_2}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l \lambda_h} \\ P_{ijd3} &= \frac{\lambda_1 \cdot \lambda_2 \cdot \lambda_3}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l \lambda_h} \\ &\vdots \\ P_{ijdm_{jd}} &= \frac{\lambda_1 \cdot \dots \cdot \lambda_{[m_{jd}-1]} \cdot \lambda_{m_{jd}}}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l \lambda_h}. \end{aligned}$$

Portanto, a probabilidade de um indivíduo  $i$  selecionar a  $k$ -ésima categoria de resposta no item  $j$  do subteste  $d$  é dada por

$$P_{ijk} = \frac{\prod_{h=1}^k \lambda_h}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l \lambda_h} = \frac{\prod_{h=1}^k \frac{F(\eta_{ijdh})}{1 - F(\eta_{ijdh})}}{\sum_{l=1}^{m_j} \prod_{h=1}^l \frac{F(\eta_{ijdh})}{1 - F(\eta_{ijdh})}}. \quad (3.6)$$

Uma FDA inversa, denotada por  $F^{-1}(\cdot)$ , é tipicamente conhecida como função de ligação. Duas das funções de ligação mais utilizadas em modelos estatísticos são a função probito, denotada por  $\Phi^{-1}(\cdot)$ , em que  $\Phi(\cdot)$  é a FDA da distribuição normal padrão e a função logito, denotada por  $L^{-1}(\cdot)$ , em que  $L(\cdot)$  é a FDA da distribuição logística padrão.

Além destas duas funções de ligação, estamos considerando neste capítulo a função log-log complementar (clog-log), definida por  $CLL(x) = \log(-\log(1-x))$ , cuja função inversa  $CLL^{-1}(x) = 1 - \exp(\exp(-x))$  é conhecida como a FDA da distribuição log-Weibull.

Assim, considerando  $L(\cdot)$  na Equação 3.6, isto é, definindo  $F(\cdot) = L(\cdot)$ , obtemos o modelo CPG-bifator com função de ligação logito, que, após alguns cálculos pode ser simplificado e, a partir da definição do preditor linear latente na Equação 3.3, pode ser escrito como

$$P_{ijk} = P(Y_{ijk} = k | \theta_{0i}, \theta_{di}, \xi_{jd}) = \frac{\exp(\sum_{h=1}^k (a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk}))}{\sum_{l=1}^{m_{jd}} \exp(\sum_{h=1}^l (a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk}))}. \quad (3.7)$$

Da mesma forma, podemos obter o modelo CPG-bifator com função de ligação probito definindo  $F(\cdot) = \Phi(\cdot)$  na Equação 3.6, ou seja,

$$P_{ijk} = P(Y_{ijk} = k | \theta_{0i}, \theta_{di}, \xi_{jd}) = \frac{\prod_{h=1}^k \frac{\Phi(a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk})}{1 - \Phi(a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk})}}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l \frac{\Phi(a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk})}{1 - \Phi(a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk})}}. \quad (3.8)$$

Analogamente, obtemos o modelo CPG-bifator com função de ligação clog-log definindo  $F(\cdot) = CLL^{-1}(\cdot)$  na Equação 3.6, que, após alguns cálculos, pode ser simplificado e escrito como

$$\begin{aligned} P_{ijk} &= P(Y_{ijk} = k | \theta_{0i}, \theta_{di}, \xi_{jd}) \\ &= \frac{\prod_{h=1}^k (\exp(\exp(-(a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk}))) - 1)}{\sum_{l=1}^{m_{jd}} \prod_{h=1}^l (\exp(\exp(-(a_{0jd}\theta_{0i} + a_{jd}\theta_{di} - b_{jdk}))) - 1)}. \end{aligned} \quad (3.9)$$

A partir da Equação 3.6, outras funções de ligação podem ser utilizadas e, portanto, outras versões do modelo CPG-bifator podem ser formuladas. Por este motivo, dizemos que o modelo CPG-bifator proposto neste capítulo é flexível a funções de ligação. Estamos utilizando no modelo CPG-bifator as funções de ligação logito, probito e clog-log.

### 3.3 Estimação Bayesiana para o modelo CPG-bifator

Nesta seção, apresentamos a estimação para o modelo CPG-bifator através de uma abordagem Bayesiana utilizando o algoritmo NUTS. Para isso, descrevemos a verossimilhança do modelo, as distribuições a priori consideradas e os métodos de avaliação de ajuste utilizados.

#### Verossimilhança

Para definirmos a verossimilhança para o modelo CPG-bifator, consideramos um vetor  $\mathbf{Y}$  contendo as respostas dos  $n$  indivíduos aos itens dos  $D$  subtestes, um vetor  $\theta_0$

contendo os traços latentes globais de todos os indivíduos, um vetor  $\theta$  contendo os traços latentes específicos de todos os indivíduos e  $\xi$  um vetor contendo os parâmetros associados a todos os itens. Assim, a verossimilhança para o modelo CPG-bifator pode ser escrita como

$$\ell(\mathbf{Y}|\theta_0, \theta, \xi) = \prod_{i=1}^N \prod_{d=1}^D \prod_{j=1}^{J_d} \prod_{k=1}^{m_{jd}} [P_{ijk}]^{w_{ijk}}, \quad (3.10)$$

em que  $P_{ijk}$  é dada pelas Equações 3.7, 3.8 e 3.9 considerando as funções de ligação logito, proibito e clog-clog, respectivamente. A variável  $w_{ijk}$  é definida como 1 se a  $k$ -ésima categoria de resposta do item  $j$  do subteste  $d$  é selecionada pelo indivíduo  $i$  e 0, caso contrário. Esta variável faz com que a verossimilhança considere apenas as probabilidades das categorias de resposta selecionadas pelos indivíduos, permitindo o uso de conjunto de dados com respostas faltantes.

### Especificação das distribuições a priori e a posteriori

A distribuição a priori comumente utilizada para os traços latentes em modelos da TRI é a distribuição normal padrão. Assim, para  $i = 1, 2, \dots, N$  e  $d = 1, 2, \dots, D$ ,

$$\begin{aligned} \theta_{0i} &\sim N(\mu_{\theta_0}, \sigma_{\theta_0}), \\ \theta_{di} &\sim N(\mu_{\theta}, \sigma_{\theta}), \end{aligned}$$

em que  $\mu_{\theta_0}$  e  $\mu_{\theta}$  são as médias de  $\theta_{0i}$  e  $\theta_{di}$ , respectivamente, e  $\sigma_{\theta_0}$  e  $\sigma_{\theta}$  são os desvios padrão de  $\theta_{0i}$  e  $\theta_{di}$ , respectivamente.

Para a distribuição a priori dos parâmetros de discriminação, a distribuição normal truncada a valores positivos é uma das opções mais utilizadas em modelos da TRI, pois estes parâmetros são definidos apenas para valores positivos. Assim, para  $d = 1, 2, \dots, D$  e  $j = 1, 2, \dots, J_d$ ,

$$\begin{aligned} a_{0jd} &\sim N(\mu_{a_0}, \sigma_{a_0})T[0, +\infty], \\ a_{jd} &\sim N(\mu_a, \sigma_a)T[0, +\infty], \end{aligned}$$

em que  $\mu_{a_0}$  e  $\mu_a$  são médias de  $a_{0jd}$  e  $a_{jd}$ , respectivamente, e  $\sigma_{a_0}$  e  $\sigma_a$  são os desvios padrão de  $a_{0jd}$  e  $a_{jd}$ , respectivamente.

A escolha da distribuição a priori para os parâmetros de dificuldade também foi baseada em Curtis (2010) e Martelli, Matteucci e Mignani (2016). Para  $d = 1, 2, \dots, D$ ,  $j = 1, 2, \dots, J_d$  e  $k = 1, 2, \dots, m_{jd}$ ,

$$b_{jdk} \sim N(\mu_b, \sigma_b),$$

em que  $\mu_b$  e  $\sigma_b$  são a média e o desvio padrão de  $b_{jdk}$ .

De acordo com Béguin e Glas (2001), Fragoso e Cúri (2013) e Reckase (2009), para garantir a identificabilidade do modelo, basta definir  $\mu_{\theta_0} = \mu_{\theta} = 0$  e  $\sigma_{\theta_0} = \sigma_{\theta} = 1$ . Além



disso, considerando a análise de sensibilidade de prioris do Capítulo 2, definimos  $\mu_{a_0} = \mu_a = 1$ ,  $\mu_b = 0$ ,  $\sigma_{a_0} = \sigma_a = 2,5$  e  $\sigma_b = 1$ . Utilizamos estas configurações para os estudos de simulação e para a aplicação deste capítulo. Com estas definições, assumindo a existência de uma estrutura de prioris independentes, podemos especificar uma distribuição a priori conjunta para os parâmetros do modelo CPG-bifator, tal que

$$p(\theta_0, \theta, \xi) = \prod_{i=1}^N \left[ \phi(\theta_{0i}; 0; 1) \prod_{d=1}^D \phi(\theta_{di}; 0; 1) \right] \times \prod_{d=1}^D \prod_{j=1}^{J_d} \left[ \phi_{T[0,+\infty]}(a_{jd}; 1; 2,5) \phi_{T[0,+\infty]}(a_{0jd}; 1; 2,5) \prod_{k=1}^{m_{jd}} \phi(b_{jdk}; 0; 1) \right], \quad (3.11)$$

em que  $\phi(\cdot; \mu; \sigma)$  corresponde à função densidade de probabilidade da distribuição normal com média igual a  $\mu$  e desvio padrão igual a  $\sigma$  e  $\phi_{T[0,+\infty]}(\cdot; \mu; \sigma)$  corresponde à função densidade de probabilidade da distribuição normal truncada com média igual a  $\mu$  e desvio padrão igual a  $\sigma$ .

A partir da verossimilhança definida na Equação 3.10 e da distribuição a priori conjunta dada pela Equação 3.11, a distribuição a posteriori dos parâmetros do modelo CPG-bifator é dada por

$$p(\theta_0, \theta, \xi | \mathbf{y}) \propto \ell(\mathbf{y} | \theta_0, \theta, \xi) \times p(\theta_0, \theta, \xi). \quad (3.12)$$

A distribuição a posteriori dos parâmetros dada pela Equação 3.12 tem uma forma intratável. Assim, para obter estimativas dos parâmetros do modelo CPG-bifator, utilizamos o algoritmo NUTS.

### **Método MCMC via algoritmo NUTS**

Alguns métodos MCMC amplamente utilizados nas últimas décadas para estimar parâmetros de modelos estatísticos, tais como Metropolis Hastings (METROPOLIS *et al.*, 1953; HASTINGS, 1970) e Gibbs sampling (GEMAN; GEMAN, 1984; GELFAND; SMITH, 1990) estão sendo substituídos por métodos alternativos que mostram-se mais eficientes. É o caso do método Monte Carlo Hamiltoniano (HMC), que tem sido amplamente explorado nos últimos anos e ganhado espaço dentre os pesquisadores. O método HMC utiliza a dinâmica Hamiltoniana ao invés de uma distribuição de probabilidade para construir uma cadeia de Markov, tornando a convergência da simulação mais rápida e com menor autocorrelação (GIROLAMI; CALDERHEAD, 2011).

Com o objetivo de tratar alguns problemas práticos do método HMC, Hoffman e Gelman (2014) propuseram um novo método a partir do HMC, chamado de No-U-Turn Sampler (NUTS). Detalhes do método HMC e do algoritmo NUTS são apresentados no Apêndice A e também podem ser encontrados em Hoffman e Gelman (2014) e DA SILVA *et al.* (2018). Utilizamos o algoritmo NUTS através do software Stan (CARPENTER *et*

al., 2017), um software lançado em 2012, gratuito e de código aberto. Apresentamos na Figura 29, no Apêndice B, o código em Stan do modelo CPG-bifator.

### 3.4 Estudo de simulação

Conduzimos um estudo de simulação visando analisar a recuperação de parâmetros do modelo CPG-bifator através do algoritmo NUTS. Para isto, consideramos a configuração utilizada na aplicação deste capítulo, em que o número de indivíduos é  $N = 333$  e o número de dimensões é  $D = 7$ , com o número de itens em cada dimensão definido por  $\mathbf{J} = (J_1, J_2, \dots, J_7) = (4, 5, 4, 4, 5, 4, 4)$ . Para analisar a influência do número de indivíduos e do número de dimensões na precisão das estimativas, consideramos outros 3 cenários neste estudo de simulação construídos a partir de  $N \in \{333, 666\}$ , ou seja, dobrando o número de indivíduos utilizado na aplicação deste capítulo, e  $D \in \{2, 7\}$ , tal que para  $D = 2$ , definimos  $\mathbf{J} = (J_1, J_2) = (15, 15)$  e para  $D = 7$ , consideramos  $\mathbf{J} = (J_1, J_2, \dots, J_7) = (4, 5, 4, 4, 5, 4, 4)$ . Assim, utilizamos os seguintes cenários no estudo de simulação: (1)  $N = 333$  e  $\mathbf{J} = (4, 5, 4, 4, 5, 4, 4)$ ; (2)  $N = 333$  e  $\mathbf{J} = (15, 15)$ ; (3)  $N = 666$  e  $\mathbf{J} = (4, 5, 4, 4, 5, 4, 4)$ ; e (4)  $N = 666$  e  $\mathbf{J} = (15, 15)$ .

Para cada cenário, geramos 10 réplicas a partir do modelo CPG-bifator com as funções de ligação logito, probito e clog-log. Então, ajustamos as réplicas também a partir do modelo CPG-bifator com as respectivas funções de ligação utilizando o algoritmo NUTS através do pacote R chamado Rstan (STAN DEVELOPMENT TEAM, 2017a), utilizando 2.000 iterações e descartando a primeira metade. A convergência foi verificada pelo diagnóstico de Gelman e Rubin (GELMAN; RUBIN, 1992).

Para verificar a recuperação de parâmetros do modelo CPG-bifator, consideramos algumas estatísticas apropriadas. A primeira estatística utilizada foi a raiz quadrada do erro quadrático médio (*root mean square error*, RMSE), uma medida de precisão definida como

$$\text{RMSE} = \sqrt{\frac{1}{10} \sum_{r=1}^{10} (\hat{\pi}_{lr} - \pi_l)^2},$$

em que  $\pi_l$  é um parâmetro do item ou do indivíduo com  $l$  sendo um índice conveniente para este parâmetro e  $\hat{\pi}_{lr}$  sua respectiva estimativa obtida na réplica  $r$ ,  $r = 1, 2, \dots, 10$ . A medida RMSE nos permite avaliar a precisão da estimativa encontrada a partir do ajuste do modelo aos dados simulados, tal que valores de RMSE menores indicam estimativas mais precisas. A segunda estatística utilizada é o valor absoluto do viés relativo (*absolute values of relative bias*, AVR B), dado por

$$\text{AVRB} = \frac{|\hat{\text{Bias}}|}{|\pi_l|},$$

em que  $\mathbf{B}ias = (\hat{\pi}_l - \pi_l)$ , com  $\hat{\pi}_l = \frac{1}{20} \sum_{r=1}^{10} \hat{\pi}_{lr}$ . A medida AVRB nos permite comparar parâmetros em diferentes escalas, tal que valores de AVRB menores indicam estimativas menos viesadas. Os valores de RMSE para os parâmetros dos itens são apresentados na Tabela 7 considerando as 3 funções de ligação utilizadas no modelo CPG-bifator: logito, probito e clog-log.

Tabela 7 – Valores de RMSE para os parâmetros dos itens a partir do modelos CPG-bifator com as três funções de ligação consideradas

Funções de ligação	$N$	$\mathbf{J}$	$a_0$	$a$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$
Logito	333	(4, 5, 4, 4, 5, 4, 4)	0,150	0,245	0,312	0,231	0,294	0,261	0,297
		(15, 15)	0,119	0,148	0,249	0,217	0,243	0,268	0,255
	666	(4, 5, 4, 4, 5, 4, 4)	0,092	0,107	0,224	0,172	0,215	0,225	0,242
		(15, 15)	0,072	0,072	0,175	0,164	0,173	0,218	0,189
Probita	333	(4, 5, 4, 4, 5, 4, 4)	0,087	0,135	0,191	0,150	0,180	0,177	0,181
		(15, 15)	0,073	0,076	0,159	0,149	0,147	0,172	0,170
	666	(4, 5, 4, 4, 5, 4, 4)	0,077	0,069	0,144	0,107	0,132	0,131	0,129
		(15, 15)	0,050	0,045	0,116	0,099	0,121	0,115	0,116
Clog-log	333	(4, 5, 4, 4, 5, 4, 4)	0,094	0,159	0,195	0,167	0,200	0,211	0,225
		(15, 15)	0,085	0,098	0,163	0,147	0,168	0,207	0,183
	666	(4, 5, 4, 4, 5, 4, 4)	0,077	0,077	0,152	0,122	0,146	0,160	0,164
		(15, 15)	0,051	0,053	0,121	0,115	0,136	0,148	0,123

Como podemos ver na Tabela 7, as funções de ligação logito, probito e clog-log foram consideradas para o ajuste do modelo CPG-bifator aos dados gerados nos 4 cenários de estudo. Em geral, os valores de RMSE indicam que os parâmetros dos itens foram bem recuperados em todos os 4 cenários. Como esperado, aumentando o número de indivíduos, o valor de RMSE diminui em todos os casos, indicando maior precisão. Um exemplo disto pode ser visto na função de ligação logito, no cenário  $N = 333$ ,  $\mathbf{J} = (4, 5, 4, 4, 5, 4, 4)$  para o parâmetro  $a_0$ , que tem valor de RMSE igual a 0,150. Ao aumentar o número de indivíduos para 666, o valor de RMSE passa para 0,092. Também como esperado, diminuindo o número de dimensões de 7 para 2 e, conseqüentemente, aumentando o número de itens em cada dimensão, uma vez que o número total de itens é mantido igual a 30, o valor de RMSE diminui. Vemos isto no mesmo cenário anterior, em que o valor de RMSE passa de 0,150 para 0,119. A única exceção neste caso é para o parâmetro  $b_5$ , com função de ligação logito no cenário  $N = 333$ ,  $\mathbf{J} = (4, 5, 4, 4, 5, 4, 4)$ , em que de RMSE passa de 0,261 para 0,268. Por fim, comparando os valores de RMSE para as funções de ligação, vemos que a função probito obteve os melhores resultados, seguida da função clog-log, com valores de RMSE um pouco maiores, em geral.

Além dos valores de RMSE, apresentamos os valores de AVRB na Figura 11. Cada um dos 4 diagramas representa um cenário.

Nos 4 cenários para todos os parâmetros dos itens, os valores da mediana de AVRB ficaram abaixo de 0,25, indicando um baixo viés relativo. Além disso, em geral, o aumento

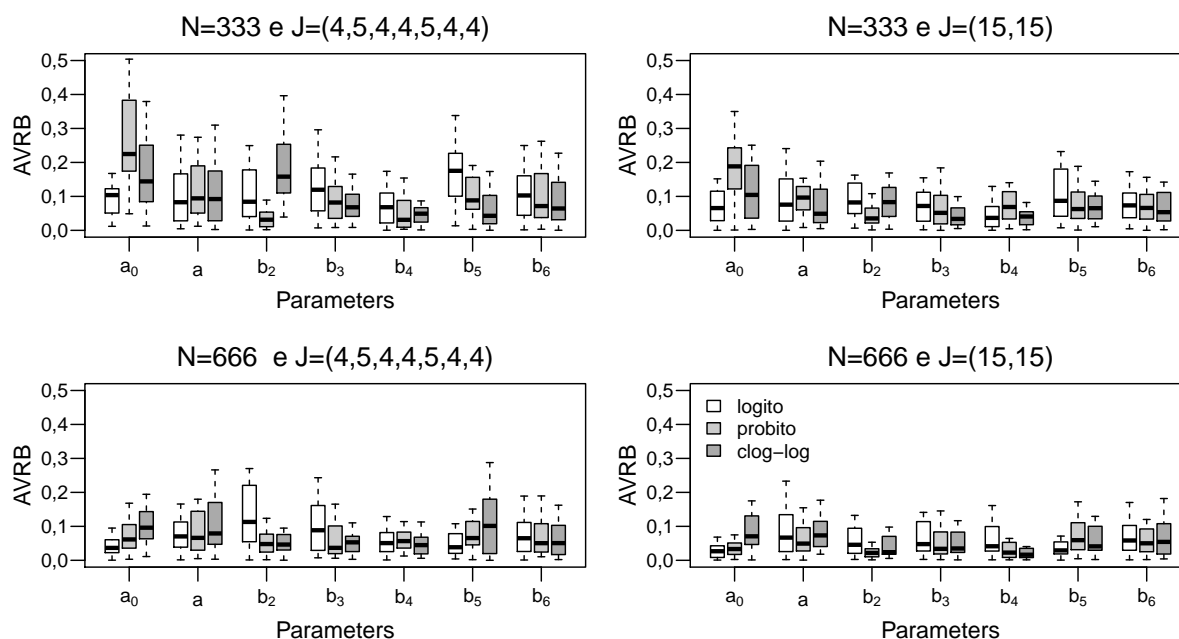


Figura 11 – Os valores de AVR dos parâmetros dos itens do modelo CPG-bifator para os quatro cenários construídos no estudo de recuperação de parâmetros.

no número de indivíduos e a diminuição de dimensões tornaram o valor do AVR mais baixo, como era esperado.

### 3.5 Aplicação

Apresentamos uma aplicação para ilustrar o uso do modelo CPG-bifator utilizando dados reais que mensuram atitudes de não usuários do serviço de mobile banking. O conjunto de dados é referente a uma pesquisa realizada na Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, juntamente com uma amostra de 333 clientes de vários bancos que não eram usuários do serviço de mobile banking, gentilmente fornecido pelo Prof. Dr. José Afonso Mazzon. Uma descrição detalhada deste banco de dados pode ser encontrada em [Püschel, Mazzon e Hernandez \(2010\)](#).

Foram considerados 7 traços latentes específicos para avaliar atitudes de não usuários frente ao mobile banking, compreendendo um total de 30 itens com 6 categorias de resposta em uma escala de discordância-concordância: (i) compatibilidade (4 itens), interpretada como a intensidade em que uma inovação é percebida como consistente com os valores e necessidades do cliente; (ii) vantagem relativa (5 itens), entendida como a intensidade em que o serviço de mobile banking é percebido como a melhor alternativa aos serviços disponíveis; (iii) visibilidade (4 itens), corresponde ao quanto o mobile banking pode ser observado pelos indivíduos antes de sua adoção; (iv) resultados (4 itens), interpretado como a intensidade em que os benefícios do uso do mobile banking são claros para os indivíduos; (v) imagem (5 itens), entendida como a intensidade em que o uso

do mobile banking melhora o status do indivíduo no sistema social; (vi) testabilidade (4 itens), interpretada como a intensidade em que os bancos permitiram que seus clientes experimentassem o serviço de mobile banking; (vii) facilidade de uso percebido (4 itens), entendida como a intensidade de esforço necessário para que um indivíduo utilize este o serviço de mobile banking.

Além dos 7 traços latentes específicos, consideramos um traço latente global correspondente à atitudes gerais mais positivas ou negativas em relação ao serviço móvel bancário. Assim, quanto maiores as pontuações relacionadas a esses traços latentes, maior a probabilidade dos clientes utilizarem serviços bancários móveis. Quanto menor a pontuação, mais esforços de marketing serão exigidos dos bancos para que seus clientes mudem as atitudes em relação à adoção deste serviço. A Figura 12 ilustra a estrutura adotada nesta aplicação. Como podemos ver, a resposta de um indivíduo a um item dentro de um subteste é influenciada tanto pelo traço latente específico deste indivíduo relacionado ao subteste como pelo traço latente global deste indivíduo. Além disso, podemos observar que não há hierarquia entre traço latente global e traços latentes específicos.

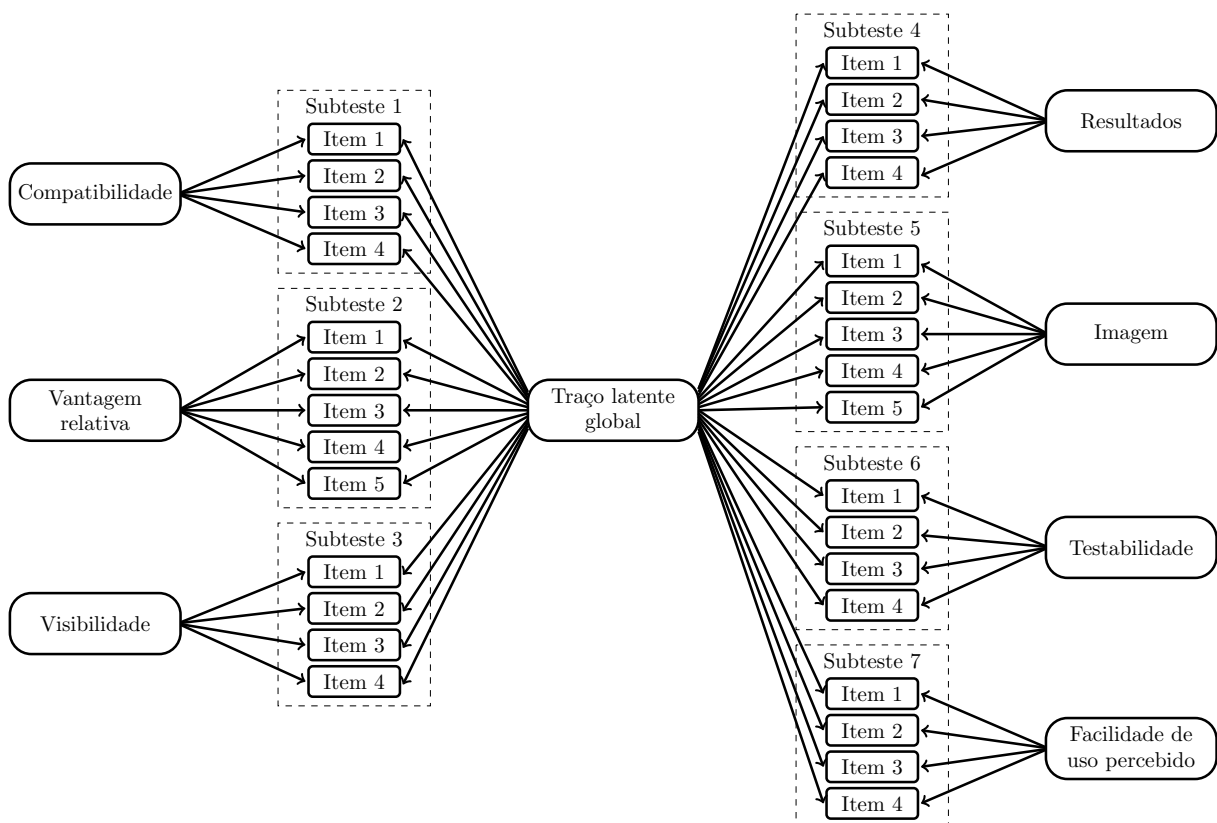


Figura 12 – Estrutura considerada para a aplicação em dados de mobile banking.

Além disso, dois aspectos referentes aos parâmetros dos itens devem ser analisados: o primeiro corresponde à qualidade dos itens em discriminar os indivíduos com diferentes níveis de atitudes em cada traço latente, e o segundo refere-se à dificuldade de um indivíduo mover-se de uma categoria de resposta para a seguinte em um dado item, ou seja, às

dificuldades dos itens.

Ajustamos o modelo CPG-bifator com funções de ligação logito, probito e clog-log a este conjunto de dados de mobile banking utilizando o algoritmo NUTS. Usamos 10.000 iterações, sendo a primeira metade *burn-in* e, portanto, descartada. A convergência das distribuições a posteriori dos parâmetros do modelo CPG-bifator foi verificada através da estimativa do fator de redução de escala potencial (GELMAN; RUBIN, 1992), denotado por  $\hat{R}$ . De acordo com Gelman *et al.* (2014), em problemas de dados reais, valores de  $\hat{R} < 1,1$  indicam convergência. Todos os parâmetros do modelo com as funções de ligação logito, probito e clog-log tiveram o valor de  $\hat{R}$  abaixo de 1,007, 1,007 e 1,005, respectivamente, indicando convergência.

Para verificar a função de ligação mais adequada para ser utilizada no modelo CPG-bifator para os dados de mobile banking, apresentamos na Tabela 8 o  $p$ -valor Bayesiano global calculado a partir da verificação preditiva a posteriori do modelo (PPMC) proposto por Rubin (1984) e os valores de três critérios de comparação de modelos: Deviance Information Criterion (DIC), que pode ser encontrado em Spiegelhalter *et al.* (2002), widely available information criterion (WAIC), proposto por Watanabe (2010) e leave-one-out cross-validation (LOO-CV ou simplesmente LOO), proposto por Vehtari, Gelman e Gabry (2017). Estes 3 critérios mostraram ser adequados e confiáveis para o modelo de resposta gradual unidimensional e para o modelo de crédito parcial generalizado unidimensional no Capítulo 2 e, apesar de não termos garantias de que serão adequados e confiáveis para o modelo CPG-bifator, utilizamos-os para comparar o modelo CPG-bifator com as funções de ligação logito, probito e clog-log. Para uma melhor interpretação dos critérios de comparação de modelos, também apresentamos na Tabela 8 as diferenças dos valores dos critérios em relação ao melhor modelo (modelo cujo valor de critério é o menor), denotado por  $\Delta_{\text{CRIT}}$ . Esta abordagem pode ser encontrada em detalhes em Burnham e Anderson (2002).

Tabela 8 – Valores dos critérios de comparação de modelos,  $p$ -valor Bayesiano global e tempo, em minutos, do algoritmo NUTS referentes ao ajuste do modelo CPG-bifator com funções de ligação logito, probito e clog-log aos dados de mobile banking

	Funções de ligação		
	Logito	Probita	Clog-log
DIC	24665,07	24434,23	24391,90
$\Delta_{\text{DIC}}$	273,17	42,33	0,00
WAIC	25140,53	25003,56	24964,25
$\Delta_{\text{WAIC}}$	176,28	39,31	0,00
LOO	25309,66	25264,32	25259,18
$\Delta_{\text{LOO}}$	50,48	5,14	0,00
$p$ -valor	0,03	0,09	0,24
Tempo	372,51	587,19	554,33

De acordo com Burnham e Anderson (2002) e Spiegelhalter *et al.* (2002), um

modelo deve ser descartado se a diferença de valor de critério  $\Delta_{\text{CRIT}}$  for maior que 10. A Tabela 8 mostra que o modelo CPG-bifator com função de ligação clog-log obteve o menor valor nos três critérios considerados. As diferenças dos valores dos índices DIC, WAIC e LOO entre o modelo CPG-bifator com as funções de ligação clog-log e logito foram 273,17, 176,28 e 50,48, respectivamente. Dessa forma, podemos descartar o modelo CPG-bifator com função logito para os dados de mobile banking utilizados. As diferenças dos valores dos índices DIC, WAIC e LOO entre o modelo CPG-bifator com as funções de ligação clog-log e probito foram 42,33, 39,31 e 5,14, respectivamente. Embora o valor do índice LOO seja menor que 10, podemos descartar o modelo CPG-bifator com função de ligação probito, porque o seu  $p$ -valor Bayesiano global é inferior a 0,10. Portanto, a partir da verificação preditiva a posteriori do modelo e dos critérios de comparação de ajuste de modelos, consideramos o modelo CPG-bifator com função de ligação clog-log o mais adequado para os dados de mobile banking.

A verificação preditiva a posteriori do modelo (PPMC) também possibilita a checagem do ajuste de um modelo da TRI aos dados item a item. Os resultados da PPMC item a item para o modelo CPG-bifator com funções de ligação logito, probito e clog-log são apresentados na Figura 13. Podemos observar que a maioria dos itens possui  $p$ -valor Bayesiano entre 0,1 e 0,9 para o modelo CPG-bifator com as 3 funções de ligação consideradas, o que indica, segundo Sahu (2002), que o modelo é bem ajustado para a maioria dos itens.

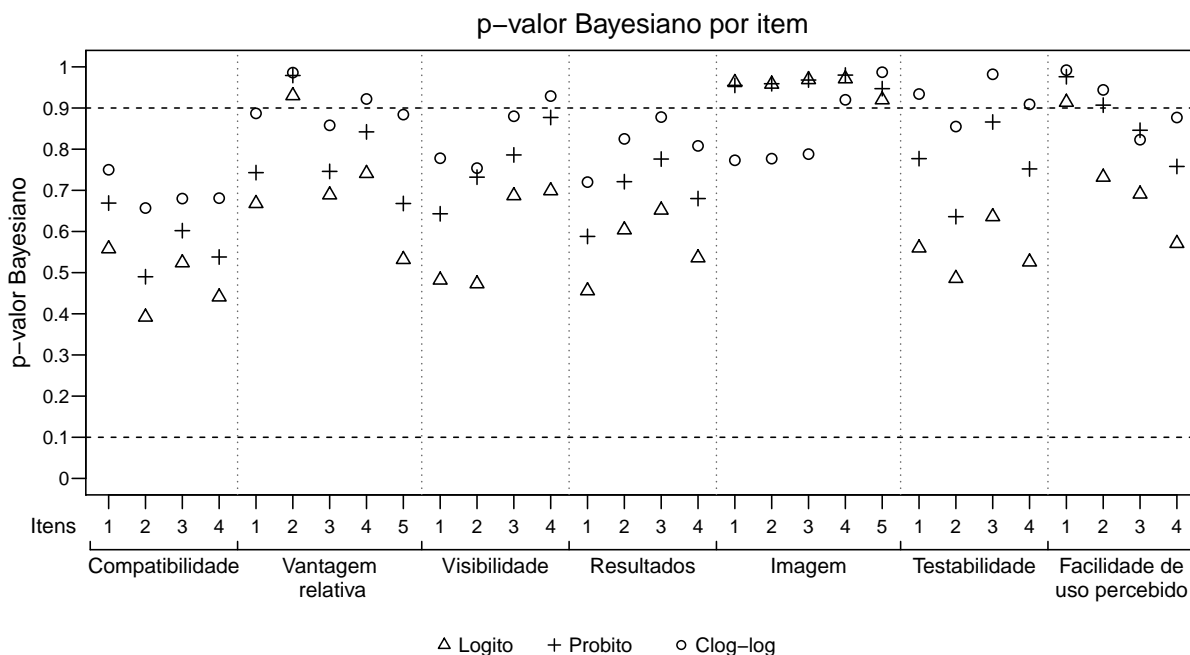


Figura 13 –  $p$ -valor Bayesiano por item para os dados de mobile banking.

Apresentamos na Figura 14 as estimativas dos parâmetros de discriminação juntamente com seus intervalos de credibilidade de 95% encontrados a partir do modelo

CPG-bifator com função de ligação clog-log.

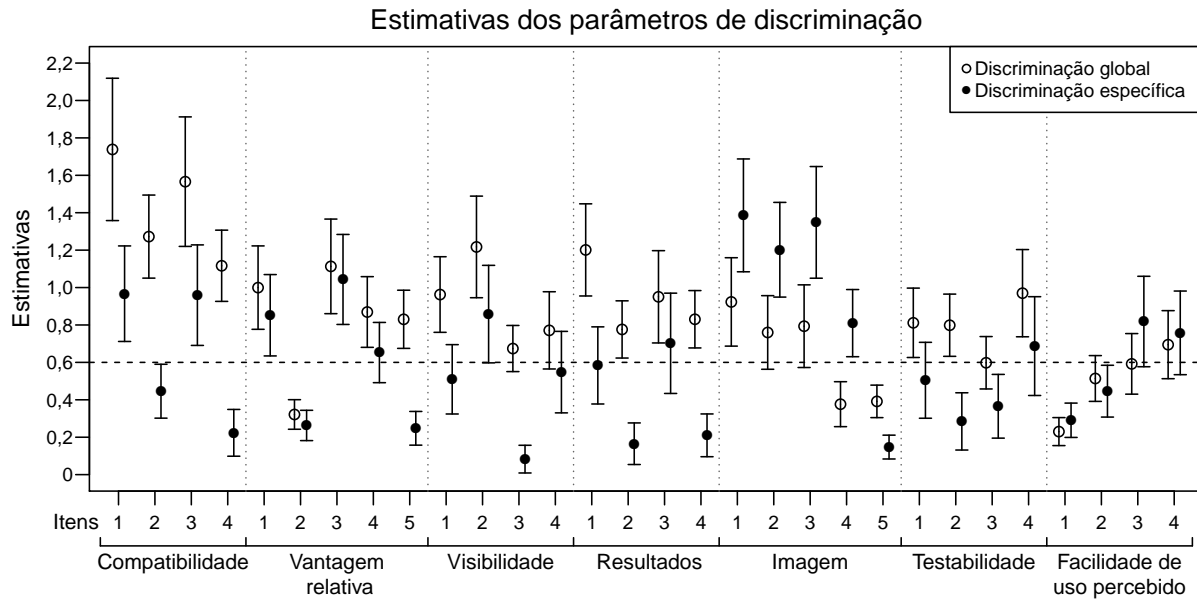


Figura 14 – Estimativas dos parâmetros de discriminação global e específica para os dados de mobile banking.

Com base nas estimativas pontuais, os valores dos parâmetros de discriminação global foram maiores do que os de discriminação específica em 23 dos 30 itens. Isto significa que estes 23 itens distinguem melhor os indivíduos com diferentes níveis do traço latente global do que do traço latente específico. Os 7 itens que possuem as estimativas de parâmetros de discriminação específica maiores do que global pertencem às dimensões “Imagem” e “Facilidade de uso percebido”. Notamos também que o item 2 da dimensão “Vantagem relativa”, item 5 da dimensão “Imagem”, e itens 1 e 2 da dimensão “Facilidade de uso percebido” possuem as estimativas de parâmetros de discriminação (global e específica) menores que 0,6 (SANTOS; AZEVEDO; BOLFARINE, 2013), que, apesar de ser um valor de referência tipicamente utilizado em modelos dicotômicos com função de ligação probito, está sendo utilizado aqui para indicar os itens com baixo poder de discriminação tanto para o traço latente global quanto para o traço latente específico considerado, sendo possível verificar a possibilidade de remoção destes itens.

Apresentamos na Figura 15 as estimativas dos parâmetros de dificuldade juntamente com seus intervalos de credibilidade de 95% encontrados a partir do modelo CPG-bifator com função de ligação clog-log.

Em geral, as estimativas dos parâmetros de dificuldade seguiram a ordem das categorias, embora isto não seja necessário nos modelos de crédito parcial e suas extensões. As estimativas dos itens que avaliam a dimensão “Imagem” são altas e se destacam das demais estimativas. Além disso, em geral, podemos ver que as diferenças entre  $b_{jd2}$  e  $b_{jd3}$ , e  $b_{jd3}$  e  $b_{jd4}$  são menores que as diferenças entre  $b_{jd4}$  e  $b_{jd5}$ , e  $b_{jd5}$  e  $b_{jd6}$ . Este resultado mostra que a dificuldade de um indivíduo evoluir de uma atitude no ponto 4 da escala



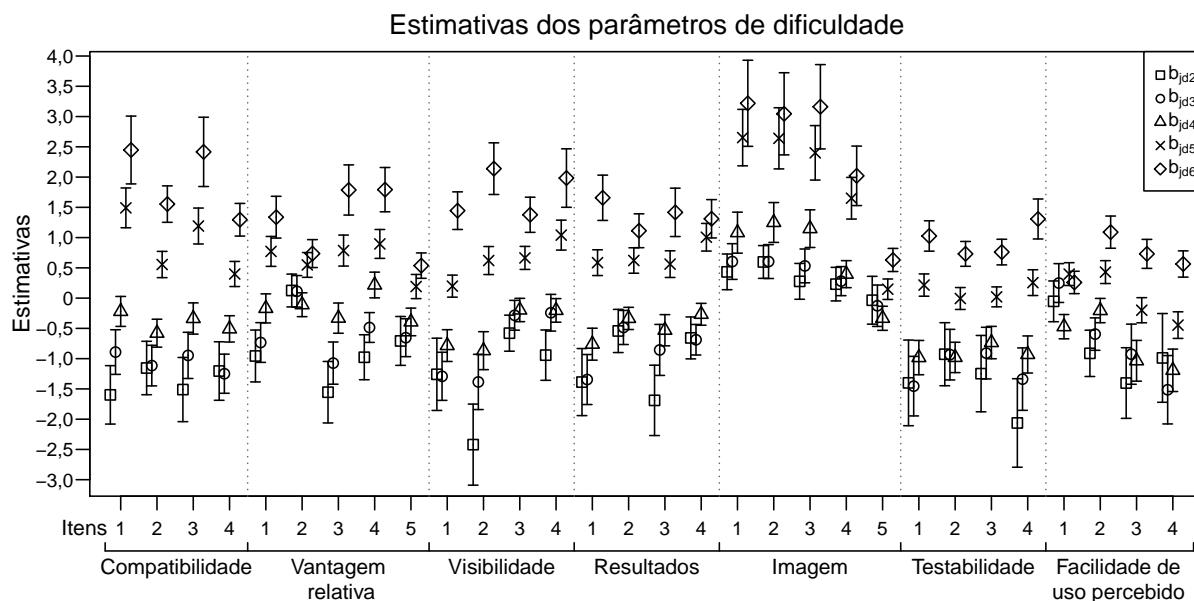


Figura 15 – Estimativas dos parâmetros de dificuldade de acordo com as categorias para os dados de mobile banking.

(concordo pouco) para o ponto 5 (concordo parcialmente) e deste para o ponto 6 (concordo totalmente) é alta, muito maior que a dificuldade para se mover do ponto 1 para o ponto 2 ou 3 (pontos de discordância).

Apresentamos na Figura 16 oito boxplots representando as distribuições das estimativas dos parâmetros de indivíduos separadas por dimensão encontradas a partir do modelo CPG-bifator com função de ligação clog-log.

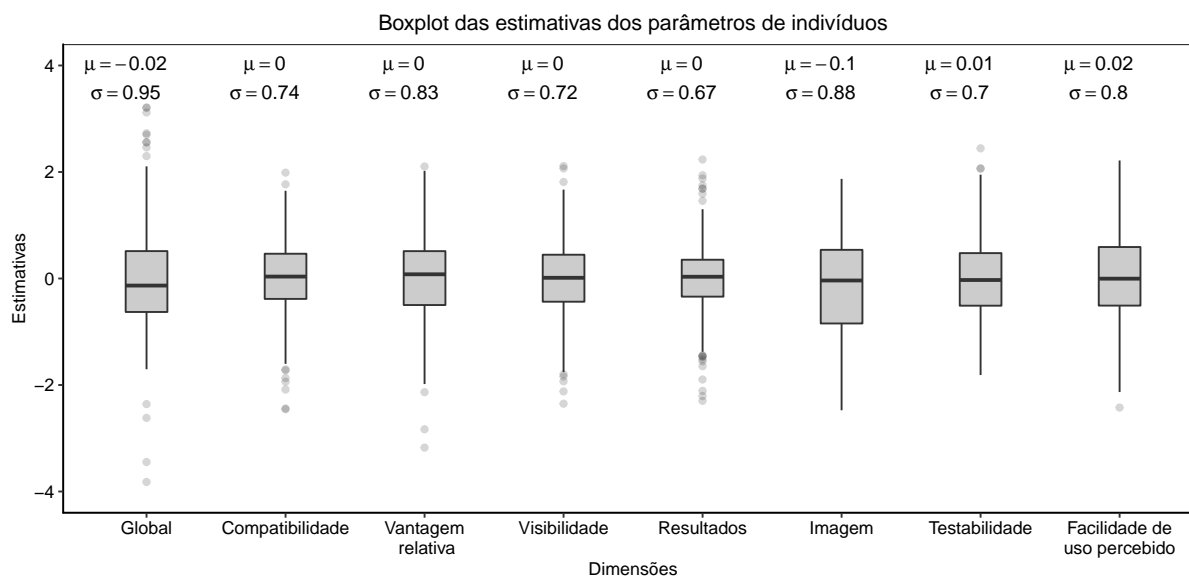


Figura 16 – Resumo das estimativas dos parâmetros de indivíduos.

De acordo com a Figura 16, a média das estimativas dos parâmetros dos indivíduos em cada dimensão foi próxima de zero. O desvio padrão da dimensão global foi o maior de

todos, com um valor igual a 0,95. Em contraste, o desvio padrão da dimensão “Resultados” foi o menor, com um valor igual a 0,67. Em termos de marketing, as estimativas dos parâmetros de indivíduos são úteis para traçar estratégias de abordagem pessoal.

Para descrever graficamente a relação entre diferentes níveis do traço latente e a probabilidade de resposta a uma particular categoria em um item, a curva característica da categoria do item (CCCI) pode ser construída. Como os modelos da TRI bifatores definem as probabilidades tanto pelo traço latente global quanto por um traço latente específico, as suas CCCIs são tridimensionais, o que dificulta a visualização e interpretação dos leitores. Para lidar com este problema, Toland *et al.* (2017) sugerem a construção de CCCIs condicionais, que são determinadas ao fixar o valor de uma das dimensões do traço latente (ou global ou específica) na média, ou seja, igual a zero. Assim, é possível visualizar a variação da probabilidade à medida que o valor de uma das dimensões do traço latente varia, dado que a outra dimensão está fixada em sua média.

Por exemplo, a construção das CCCIs condicionais para um item específico fixando o traço latente global em zero a partir do modelo CPG-bifator com função de ligação clog-log é feita através da Equação 3.9 com os valores estimados dos parâmetros do item, com o traço latente global fixado em zero e com o traço latente específico livre para variar no intervalo  $[-3, +3]$ . Assim, para cada valor do traço latente específico entre  $-3$  e  $+3$ , um diferente  $P_{ijk}$  é calculado para formar as curvas.

As Figuras 17 e 18 mostram as CCCIs condicionais dos 30 itens dos dados de mobile banking encontradas a partir do modelo CPG-bifator com função de ligação clog-log.

Observando o primeiro gráfico da Figura 17 (CCCI condicionais referentes ao item 1 da dimensão “Compatibilidade”), podemos ver que existem intervalos de valores para o traço latente global que tornam cada categoria de resposta mais provável de ser escolhida. Esses intervalos são definidos pela projeção das interseções das curvas relacionadas às categorias adjacentes no eixo horizontal. Por exemplo, indivíduos com valor de  $\theta_0$  inferior a  $-1,15$  possuem maior probabilidade de responder à categoria 1 (discordo totalmente), enquanto que indivíduos com  $\theta_0$  entre  $-1,15$  e  $-0,75$  possuem maior probabilidade de responder à categoria 2 (discordo parcialmente).

Algumas CCCIs mostram a existência de itens em que não há categoria de resposta predominante para qualquer intervalo de valor do traço latente. Um exemplo disto são as CCCIs condicionais para o traço latente específico do item 2 da dimensão “Compatibilidade”, no segundo gráfico da Figura 18. Vemos que as curvas referentes às categorias 1 e 2 não são superiores à curva referente à categoria 3 para nenhum valor de  $\theta_1$ . Como se trata de CCCIs condicionais, uma investigação mais detalhada é necessária antes de decidir, por exemplo, retirar uma destas categorias.

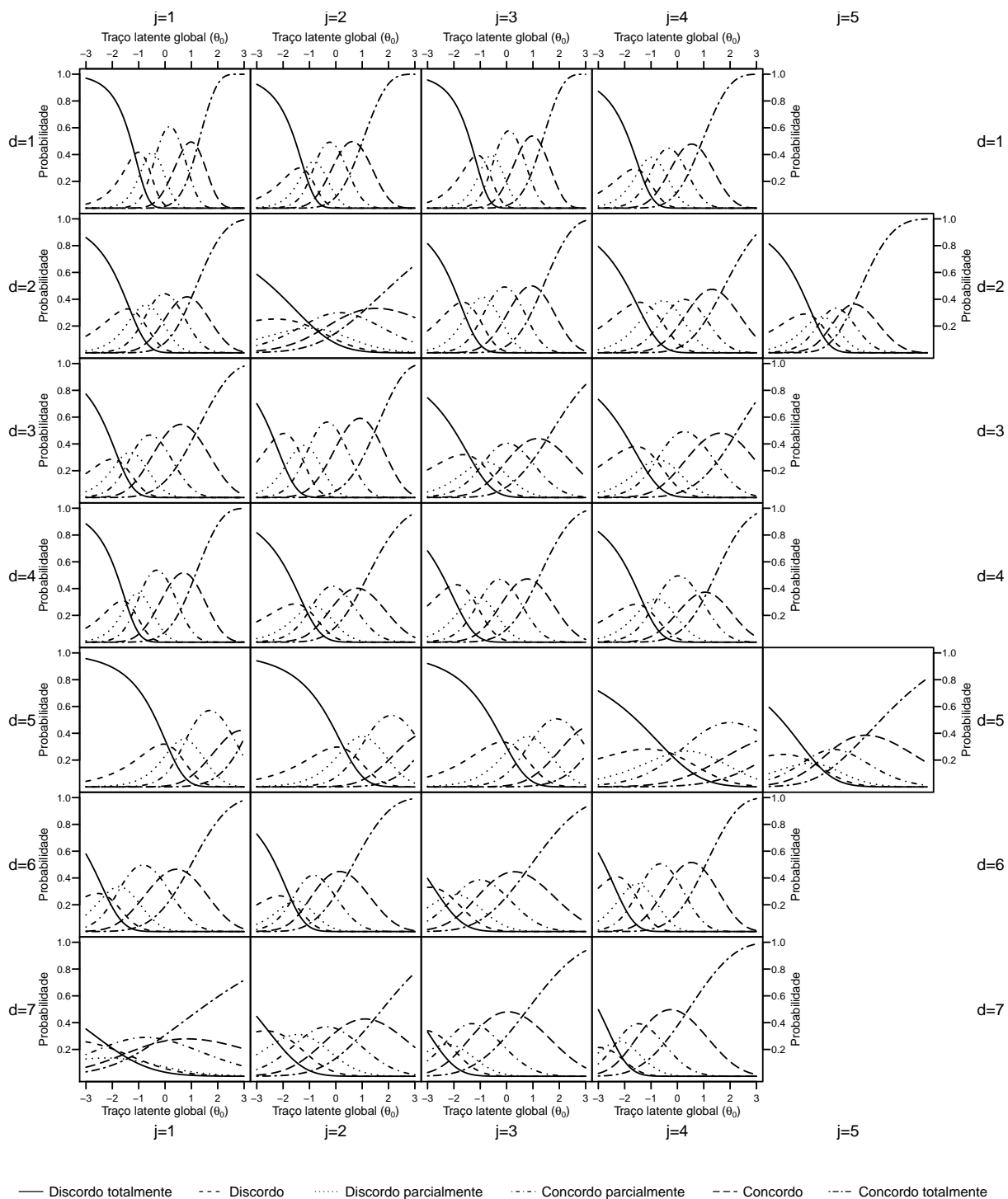


Figura 17 – CCCIs condicionais considerando o valor do traço latente específico como zero.

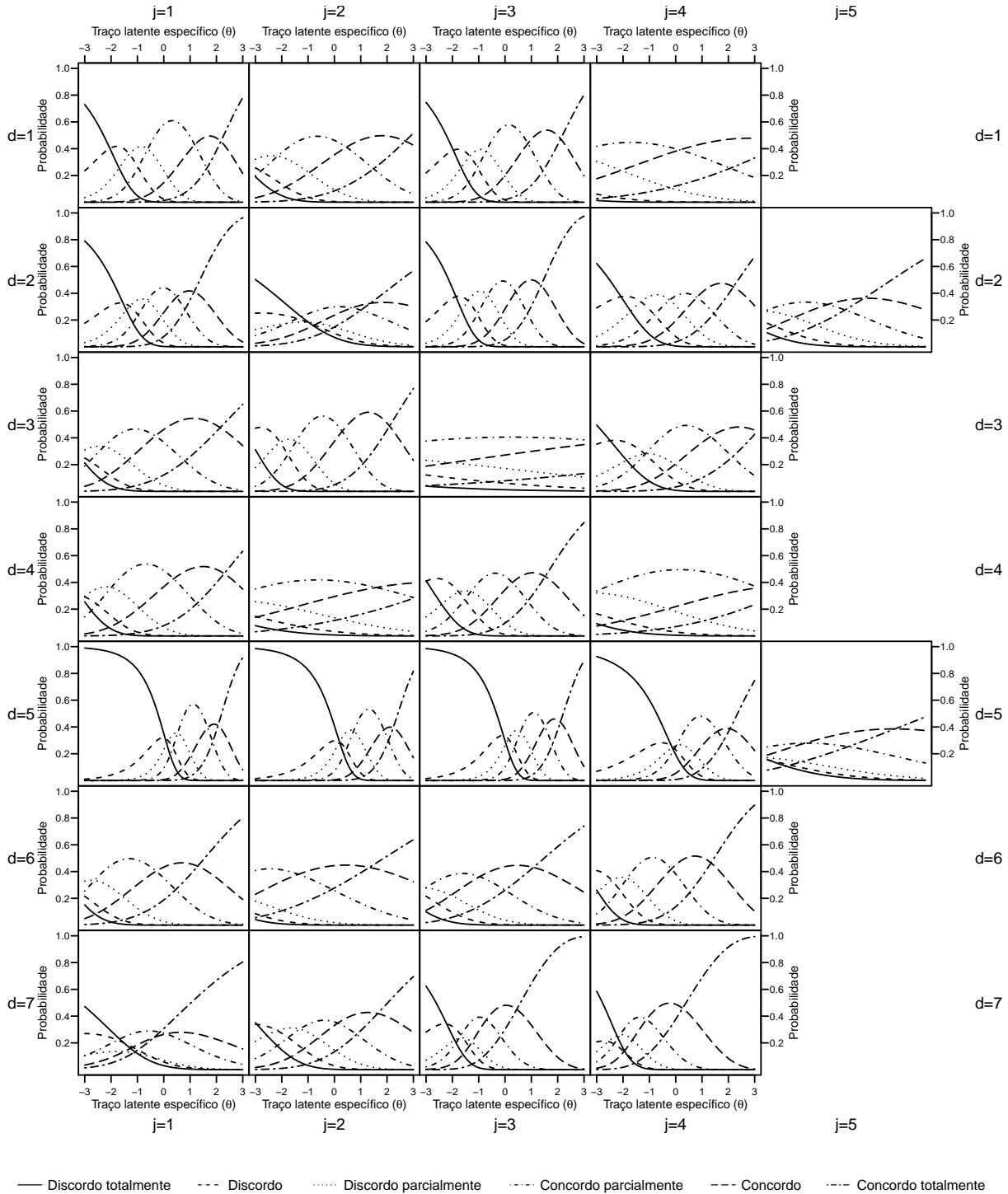


Figura 18 – CCCIs condicionais considerando o valor do traço latente global como zero.

## 3.6 Conclusões e comentários

Neste capítulo, estendemos o modelo CPG unidimensional de [Muraki \(1992\)](#) para o modelo CPG-bifator com funções de ligação flexíveis, ou seja, após algumas manipulações algébricas apresentadas na [Seção 3.2](#), outras funções de ligação podem ser consideradas no modelo CPG-bifator, seguindo as propostas de funções de ligação alternativas em modelos da TRI, como em [Albert e Chib \(1993\)](#), [Samejima \(2000\)](#) e [Bazán, Branco e Bolfarine \(2006\)](#). Além da função de ligação logito, utilizada no modelo de Muraki, também utilizamos neste capítulo as funções probito e clog-log. Esta formulação flexível do modelo bifac-GPC inclui como um caso particular o modelo CPG unidimensional flexível, que pode ser explorado em desenvolvimentos futuros.

O estudo de simulação foi proposto para avaliar a recuperação de parâmetros do modelo CPG-bifator com funções de ligação logito, probito e clog-log. Para isto, consideramos quatro diferentes cenários definidos a partir das variações do número de indivíduos (333 e 666), dimensões (2 e 7) e o número de itens em cada dimensão (15 itens em cada dimensão para os cenários com 2 dimensões e, para os cenários com 7 dimensões, duas dimensões com 5 itens e cinco dimensões com 4 itens). Os resultados deste estudo mostraram que o algoritmo NUTS via software Stan fornece estimativas precisas para todos os casos avaliados. Como esperado, aumentando o número de indivíduos e itens por dimensão, as estatísticas consideradas neste estudo indicam maior precisão nas estimativas dos parâmetros do modelo.

Os resultados apresentados na aplicação deste capítulo exemplificam o uso do modelo CPG-bifator e ajudam o leitor a entender a sua estrutura e a interpretação de seus parâmetros. O modelo CPG-bifator foi ajustado aos dados de mobile banking utilizando as funções de ligação logito, probito e clog-log, sendo esta última indicada pelos critérios de comparação de modelos e pela verificação preditiva a posteriori do modelo como a função de ligação mais adequada aos dados e, portanto, escolhida para a análise realizada. Apresentamos as estimativas dos parâmetros de discriminação e dificuldade dos itens, seus intervalos de credibilidade de 95% e sua interpretação para o contexto dos dados em mobile banking. Além disso, apresentamos as CCCI's condicionais e sua interpretação.

Podemos ver alguns benefícios práticos do modelo CPG-bifator com funções de ligação flexíveis. Primeiro, pesquisadores e profissionais agora passam a ter mais opções de modelos bifatores para dados de respostas politômicas ordenadas, juntamente com o modelo RG-bifator com funções de ligação logito e probito propostos por [Toland \*et al.\* \(2017\)](#) e [Martelli, Matteucci e Mignani \(2016\)](#), respectivamente. Em segundo lugar, a inclusão de outras funções de ligação no modelo CPG-bifator torna possível a comparação e a escolha do modelo com a função de ligação que melhor se ajusta aos dados, como foi feito na aplicação deste capítulo, utilizando os dados de mobile banking, em que o modelo CPG-bifator com função de ligação clog-log foi escolhido como o mais adequado aos dados.

Além disso, o uso do software Stan para estimar os parâmetros do modelo CPG-bifator facilita a inserção de diferentes funções de ligação no modelo pelo usuário. O software Stan também permite que os critérios de comparação de modelos sejam facilmente calculados, como DIC, WAIC e LOO, utilizados na aplicação.

---

# UM MODELO DE CRÉDITO PARCIAL GENERALIZADO MULTIDIMENSIONAL COM MATRIZ Q

---

## 4.1 Introdução

Os modelos da teoria da resposta ao item multidimensionais (TRIM) utilizam dados de respostas de indivíduos a itens para estimar múltiplos traços latentes de interesse, tornando-os úteis, na prática, em várias áreas de estudo, tais como em medições educacionais e psicológicas e outras áreas da ciência. Por exemplo, alguns testes de personalidade tem por objetivo avaliar mais de um fator de personalidade dos indivíduos, como o “Modelo de cinco fatores” (GOLDBERG, 1993), uma metodologia bastante conhecida na psicologia que mapeia cinco supostas dimensões de personalidades para o estudo de características humanas. Além disso, os modelos da TRIM são utilizados em situações em que itens de um teste exigem múltiplas dimensões do traço latente para ser respondido. Por exemplo, em um teste para medir a capacidade matemática dos alunos, os itens também irão requerer conhecimento de leitura e interpretação de texto. A partir deste exemplo, podemos compreender o motivo de alguns autores afirmarem que nenhum teste pode ser verdadeiramente unidimensional (THISSEN, 2016).

Apesar dos modelos da TRIM serem promissores, na prática eles exigem testes com muitos itens e grandes tamanhos de amostra para obter estimativas de parâmetros precisas. Isto ocorre porque o número de parâmetros dos modelos da TRIM é muito maior do que o número de parâmetros dos modelos da TRI unidimensionais (HABERMAN; SINHARAY, 2010; MAYDEU-OLIVARES, 2001). Além disso, outro aspecto que dificulta o uso de modelos da TRIM é a ausência de um procedimento específico para o usuário definir os itens que irão medir as dimensões do traço latente, ou seja, explicitar a relação

entre itens e dimensões do traço latente.

Um procedimento bastante simples e intuitivo para especificar a relação item-traço é utilizado por uma nova classe de modelos de variáveis latentes multidimensionais, conhecida como modelos de diagnóstico cognitivo (MDC). Detalhes destes modelos podem ser encontrados em [Rupp, Templin e Henson \(2010\)](#). A modelagem através dos MDC tornou-se recentemente um novo campo empolgante de pesquisa em variáveis latentes. Estes modelos tem por objetivo classificar os indivíduos em grupos com base na posse/não posse de certos atributos, fornecendo informações específicas sobre seus pontos fortes e fracos. Os MDC dispõem de uma matriz comumente conhecida como matriz  $\mathbf{Q}$  ([TATSUOKA, 1983](#)), um recurso bastante simples e eficiente para o usuário especificar quais itens medem quais dimensões do traço latente. Os elementos da matriz  $\mathbf{Q}$  são definidos como

$$q_{id} = \begin{cases} 1, & \text{se o item } i \text{ está relacionado à dimensão } d \\ 0, & \text{caso contrário} \end{cases}.$$

Por exemplo, em um teste com 10 itens planejado para avaliar 4 dimensões do traço latente, consideramos uma matriz  $\mathbf{Q}$ , tal que

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Considerando esta matriz  $\mathbf{Q}$ , vemos que o item 1 é projetado para medir apenas a dimensão 1 do traço latente, enquanto o item 5 é projetado para medir as dimensões 1 e 2. Assim, se um item não é elaborado para medir uma dimensão específica do traço latente, esta informação é incorporada pelos MDC e o ajuste é realizado considerando com precisão o planejamento do teste.

Neste capítulo, propomos uma nova formulação para os modelos da TRIM visando incorporar informações da relação item-traço através da matriz  $\mathbf{Q}$ . Especificamente, reescrevemos o modelo de crédito parcial generalizado multidimensional (CPGM) de [Yao e Schwarz \(2006b\)](#), considerando a matriz  $\mathbf{Q}$  em sua formulação. Assim, tornamos acessível aos usuários uma maneira de expressar claramente a relação item-traço nos modelos da TRIM. Para a estimação dos parâmetros do modelo CPGM com matriz  $\mathbf{Q}$ , utilizamos uma



abordagem Bayesiana através do algoritmo No-U-Turn Sampler (NUTS), uma extensão do método de Monte Carlo Hamiltoniano (HMC) descrito em [Hoffman e Gelman \(2014\)](#).

Este capítulo está organizado da seguinte forma: nesta seção apresentamos uma visão geral dos modelos da TRIM, do procedimento de especificação da relação entre item-traço nos MDC através da matriz  $\mathbf{Q}$  e a proposta deste capítulo. Na Seção 4.2, apresentamos a formulação dos modelos da TRIM para dados politômicos com a matriz  $\mathbf{Q}$  e, especificamente, a formulação do modelo CPGM com matriz  $\mathbf{Q}$ , que nós chamamos de modelo CPGM-Q. Na Seção 4.3, apresentamos informações sobre a estimação dos parâmetros do modelo CPGM-Q através de uma abordagem Bayesiana. Na Seção 4.4, apresentamos uma aplicação do modelo proposto neste capítulo para ilustrar o seu uso. Por fim, na Seção 4.5, apresentamos algumas conclusões, comentários e sugestões para futuras pesquisas.

## 4.2 Os modelos da TRIM para dados politômicos

Os modelos da TRIM são utilizados no contexto em que um teste composto de  $J$  itens é designado para avaliar  $D$  dimensões de um traço latente. Como estamos considerando modelos para dados politômicos, cada item  $j$  deve possuir  $m_j > 2$  categorias de resposta. Assim, definimos a variável aleatória  $Y_{ij}$  como a resposta do indivíduo  $i$  ao item  $j$ , para  $i = 1, 2, \dots, N$  e  $j = 1, 2, \dots, J$ , tal que

$$Y_{ij} | \theta_i, \xi_j \sim \text{Categorizada}(P_{ij1}, P_{ij2}, \dots, P_{ijm_j}), \quad (4.1)$$

em que  $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iD})'$  é o vetor de parâmetros do indivíduo  $i$  tal que  $\theta_{id}$  é a  $d$ -ésima dimensão do traço latente pertencente ao indivíduo  $i$ ;  $\xi_j = (a_{j1}, a_{j2}, \dots, a_{jD}, b_{j1}, b_{j2}, \dots, b_{jm_j})'$  é o vetor de parâmetros associado com o  $j$ -ésimo item, sendo  $a_{jd}$  o parâmetro de discriminação específico da dimensão  $d$  e  $b_{jk}$  o parâmetros de dificuldade referente à  $k$ -ésima categoria do item  $j$ ; e  $P_{ijk}$  é a probabilidade do indivíduo  $i$  responder a  $k$ -ésima categoria de resposta do item  $j$ , definida de acordo com o modelo da TRIM. A distribuição Categorizada apresentada na Expressão 4.1 é definida na Seção 2.2 e exemplificada na Tabela 1.

Existem diferentes modelos da TRIM para dados politômicos e cada um deles define as probabilidades  $P_{ijk}$  de maneira diferente. Neste sentido, [Reckase \(2009\)](#) apresenta uma classificação para os modelos da TRIM de acordo com a maneira em que as coordenadas do traço latente  $\theta_i$  são utilizadas para modelar as probabilidades de resposta. A primeira classe é composta por modelos da TRIM que são formulados a partir de combinações lineares das coordenadas de  $\theta_i$ . Essas combinações lineares podem resultar na mesma soma com diferentes combinações de valores de  $\theta_i$ . Por exemplo, se o valor de uma coordenada de  $\theta_i$  for alto, a soma será a mesma se o valor de outra coordenada

for suficientemente pequeno, causando um efeito de compensação no preditor linear latente do modelo. Portanto, os modelos da TRIM que utilizam combinações lineares das coordenadas de  $\theta_i$  são chamados de modelos compensatórios.

A segunda classe é composta por modelos da TRIM que modelam a probabilidade de resposta em um item através do produto de probabilidades calculadas a partir de modelos unidimensionais, reduzindo o efeito de compensação. Esses modelos são chamados de modelos não compensatórios. Dentro destas duas classes de modelos da TRIM, existem vários modelos formulados na literatura. Na sequência, apresentamos a proposta de incorporar a matriz  $\mathbf{Q}$  em modelos compensatórios.

### 4.2.1 Incorporando a matriz $\mathbf{Q}$ em modelos compensatórios

O preditor linear latente dos modelos compensatórios é especificado em termos da combinação linear das coordenadas de  $\theta_i$ , que é dada por

$$\mathbf{a}'_j \theta_i = a_{j1} \theta_{i1} + a_{j2} \theta_{i2} + \dots + a_{jD} \theta_{iD} = \sum_{d=1}^D a_{jd} \theta_{id}, \quad (4.2)$$

em que  $\mathbf{a}_j = (a_{j1}, a_{j2}, \dots, a_{jD})'$  é o vetor de parâmetros de discriminação do item  $j$ .

Visando incorporar a matriz  $\mathbf{Q}$  em modelos da TRIM compensatórios, inserimos os seus elementos na combinação linear em 4.2, tal que

$$\mathbf{q}'_j \text{diag}(\mathbf{a}_j) \theta_i = q_{j1} a_{j1} \theta_{i1} + q_{j2} a_{j2} \theta_{i2} + \dots + q_{jD} a_{jD} \theta_{iD} = \sum_{d=1}^D q_{jd} a_{jd} \theta_{id}, \quad (4.3)$$

em que  $\mathbf{q}_j = (q_{j1}, q_{j2}, \dots, q_{jD})'$  representa a  $j$ -ésima linha da matriz  $\mathbf{Q}$ ,  $q_{jd}$  é o elemento da  $j$ -ésima linha e  $d$ -ésima coluna da matriz  $\mathbf{Q}$  tal que  $q_{jd} = 1$  se o item  $j$  avalia a dimensão  $d$  e  $q_{jd} = 0$  caso contrário, e  $\text{diag}(\mathbf{a}_j)$  é a matriz diagonal com os elementos de  $\mathbf{a}_j$ . Assim, se  $q_{jd} = 0$ , o termo de soma correspondente será nula. Em termos práticos, isto significa que a probabilidade do indivíduo  $i$  responder qualquer uma das categorias do item  $j$  não é influenciada pela  $d$ -ésima coordenada de  $\theta_i$ .

Para fornecer um exemplo específico ao leitor, incorporamos a matriz  $\mathbf{Q}$  no modelo CPGM, propondo o modelo CPGM-Q, que é apresentado na sequência e utilizado no exemplo prático deste capítulo.

### 4.2.2 O modelo CPGM-Q

A probabilidade do indivíduo  $i$  responder a  $k$ -ésima categoria do item  $j$  sob o modelo CPGM-Q é

$$P_{ijk} = P(Y_{ij} = k | \theta_i, \xi_j) = \frac{\exp\left(k \sum_{d=1}^D q_{jd} a_{jd} \theta_{id} - \sum_{h=1}^k b_{jh}\right)}{\sum_{l=1}^{m_j} \exp\left(l \sum_{d=1}^D q_{jd} a_{jd} \theta_{id} - \sum_{h=1}^l b_{jh}\right)}. \quad (4.4)$$

Assim como no modelo de crédito parcial generalizado unidimensional proposto por [Muraki \(1992\)](#),  $b_{j1}$  é arbitrariamente definido como 0, o que faz cada item  $j$  possuir  $m_j - 1$  parâmetros de dificuldade.

A partir de uma matriz  $\mathbf{Q}$  conveniente, podemos obter, por exemplo, o modelo CPG unidimensional abordado no Capítulo 2 ou o modelo CPG-bifator, proposto no Capítulo 3. A autonomia que a formulação do modelo CPGM-Q oferece ao usuário é um benefício da incorporação da matriz  $\mathbf{Q}$  nos modelos da TRIM.

[Reckase \(2009\)](#) e [Fragoso e Cúri \(2013\)](#) apresentam os parâmetros de discriminação e dificuldade multidimensionais para os modelos da TRIM, que são comparáveis aos parâmetros de discriminação e dificuldade no contexto unidimensional. Considerando a incorporação da matriz  $\mathbf{Q}$  nos modelos TRIM, o parâmetro de discriminação multidimensional para o item  $j$  é dado por

$$MDISC_j = \sqrt{\sum_{d=1}^D q_{jd} a_{jd}^2}. \quad (4.5)$$

Os elementos da matriz  $q_{jd}$  (para  $d = 1, 2, \dots, D$ ) fazem com que o  $MDISC_j$  considere apenas os parâmetros de discriminação associados às dimensões que cada item mede. Da mesma forma, a dificuldade multidimensional do item  $j$  e categoria  $k$ , denotado por  $MDIFF_{jk}$ , é definido por

$$MDIFF_{jk} = \frac{b_{jk}}{MDISC_j}. \quad (4.6)$$

Assim, podemos calcular a discriminação multidimensional e a dificuldade multidimensional para todos os itens designados a avaliar mais de uma dimensão dos traços latentes dos indivíduos.

## 4.3 Estimação Bayesiana para o modelo CPGM-Q

Nesta seção, apresentamos a abordagem Bayesiana utilizada na estimação dos parâmetros do modelo CPGM-Q. Para isso, especificamos a verossimilhança do modelo, as distribuições a priori consideradas, a distribuição a posteriori e os métodos de avaliação de ajuste utilizados.

## Verossimilhança

Seja  $\mathbf{y}$  o vetor que contém as respostas dos  $N$  indivíduos aos  $J$  itens. A verossimilhança para o modelo CPGM-Q pode ser escrita como

$$\ell(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^N \prod_{j=1}^J \prod_{k=1}^{m_j} [P_{ijk}]^{w_{ijk}}, \quad (4.7)$$

em que  $\boldsymbol{\theta}$  é o vetor dos parâmetros de indivíduos,  $\boldsymbol{\xi}$  é o vetor de parâmetros de todos os itens,  $P_{ijk}$  é definido pela Equação 4.4 e  $w_{ijk} = 1$  se a  $k$ -ésima categoria de resposta do item  $j$  é selecionado pelo indivíduo  $i$  e  $w_{ijk} = 0$ , caso contrário. Esta variável faz com que a verossimilhança considere apenas as probabilidades das categorias de resposta selecionadas pelos indivíduos, permitindo o uso de conjunto de dados com respostas faltantes.

## Especificação das distribuições a priori e a posteriori

Definimos as distribuições a priori para os parâmetros do modelo CPGM-Q de acordo com Curtis (2010) e Martelli, Matteucci e Mignani (2016), tais que

$$\begin{aligned} a_{jd} &\sim N(\boldsymbol{\mu}_a, \boldsymbol{\sigma}_a)T(0, +\infty), \text{ para } (j, d) \in \mathbb{C}_Q, \\ b_{jk} &\sim N(\boldsymbol{\mu}_b, \boldsymbol{\sigma}_b), \text{ para } j = 1, 2, \dots, J, \\ \boldsymbol{\theta}_{id} &\sim N(\boldsymbol{\mu}_\theta, \boldsymbol{\sigma}_\theta), \text{ para } i = 1, 2, \dots, N \text{ e } d = 1, 2, \dots, D, \end{aligned} \quad (4.8)$$

em que  $\boldsymbol{\mu}_a$ ,  $\boldsymbol{\mu}_b$  e  $\boldsymbol{\mu}_\theta$  são as médias de  $a_{jd}$ ,  $b_{jk}$  e  $\boldsymbol{\theta}_{id}$ , respectivamente,  $\boldsymbol{\sigma}_a$ ,  $\boldsymbol{\sigma}_b$  e  $\boldsymbol{\sigma}_\theta$  são o desvio padrão de  $a_{jd}$ ,  $b_{jk}$  e  $\boldsymbol{\theta}_{id}$ , respectivamente, e  $\mathbb{C}_Q = \{(j, d) : q_{jd} = 1, j = 1, 2, \dots, J, d = 1, 2, \dots, D\}$ . De acordo com Béguin e Glas (2001), Fragoso e Cúri (2013) e Reckase (2009), para garantir a identificabilidade do modelo, basta definir  $\boldsymbol{\mu}_\theta = \mathbf{0}$  e  $\boldsymbol{\sigma}_\theta = \mathbf{1}$ . Além disso, considerando a análise de sensibilidade de prioris do Capítulo 2, definimos  $\boldsymbol{\mu}_a = \mathbf{1}$ ,  $\boldsymbol{\mu}_b = \mathbf{0}$ ,  $\boldsymbol{\sigma}_a = 2,5$  e  $\boldsymbol{\sigma}_b = \mathbf{1}$ . Utilizamos estas configurações para os estudos de simulação e para a aplicação deste capítulo.

Para a especificação de uma distribuição a priori conjunta dos parâmetros dos itens e dos indivíduos do modelo CPGM-Q, assume-se que existe uma estrutura de prioris independentes da seguinte forma

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\xi}) &= p(\boldsymbol{\theta}) p(\boldsymbol{\xi}) \\ &= \prod_{i=1}^N \left[ \prod_{d=1}^D p(\boldsymbol{\theta}_{id}) \right] \times \prod_{(j,d) \in \mathbb{C}_Q} p(a_{jd}) \times \prod_{j=1}^J \prod_{k=1}^{m_j} p(b_{jk}) \\ &= \prod_{i=1}^N \left[ \prod_{d=1}^D \phi(\boldsymbol{\theta}_{id}; \mathbf{0}, \mathbf{1}) \right] \times \prod_{(j,d) \in \mathbb{C}_Q} \phi_{T[0,+\infty]}(a_{jd}; \mathbf{0}, \mathbf{1}) \times \prod_{j=1}^J \prod_{k=1}^{m_j} \phi(b_{jk}; \mathbf{0}, \mathbf{1}), \end{aligned} \quad (4.9)$$

em que  $\phi(\cdot; \boldsymbol{\mu}; \boldsymbol{\sigma})$  corresponde à função densidade de probabilidade da distribuição normal com média igual a  $\boldsymbol{\mu}$  e desvio padrão igual a  $\boldsymbol{\sigma}$  e  $\phi_{T[0,+\infty]}(\cdot; \boldsymbol{\mu}; \boldsymbol{\sigma})$  corresponde à função

densidade de probabilidade da distribuição normal truncada com média igual a  $\mu$  e desvio padrão igual a  $\sigma$ .

Considerando a verossimilhança definida na Equação 5.7 e a distribuição a priori conjunta dos parâmetros dos itens e dos indivíduos do modelo CPGM-Q definida na Equação 4.9, a distribuição a posteriori dos parâmetros do modelo CPGM-Q é dada por

$$p(\boldsymbol{\theta}, \boldsymbol{\xi} | \mathbf{y}) \propto \ell(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\xi}) \times p(\boldsymbol{\theta}, \boldsymbol{\xi}). \quad (4.10)$$

A distribuição a posteriori apresentada na Equação 4.10 tem uma forma intratável e, portanto, não é possível obter analiticamente as distribuições marginais a posteriori. Assim, para obter estimativas dos parâmetros dos modelos RG e CPG, utilizamos o método MCMC através do algoritmo NUTS, que será apresentado a seguir.

### **Método MCMC via algoritmo NUTS**

O algoritmo NUTS é um método MCMC proposto por Hoffman e Gelman (2014) a partir do método de Monte Carlo Hamiltoniano (HMC) com o objetivo de construir cadeias de Markov de maneira mais eficiente em relação à outros métodos MCMC amplamente utilizados na inferência Bayesiana, como Metropolis Hastings (METROPOLIS *et al.*, 1953; HASTINGS, 1970) e Gibbs Sampling (GEMAN; GEMAN, 1984; GELFAND; SMITH, 1990). Estas comparações podem ser encontradas em Girolami e Calderhead (2011), Grant *et al.* (2017) e DA SILVA *et al.* (2018). Estes métodos tradicionais utilizam uma distribuição de probabilidade para construir uma cadeia de Markov e exigem um número muito grande de iterações. Em contraste, o HMC e consequentemente o NUTS utilizam a dinâmica hamiltoniana (NEAL, 2011) para construir uma cadeia de Markov, fazendo a simulação convergir mais rapidamente para a distribuição alvo com baixa autocorrelação. Uma explicação do método HMC e do algoritmo NUTS pode ser encontrada no Apêndice A e em Hoffman e Gelman (2014) e DA SILVA *et al.* (2018).

Neste capítulo, utilizamos o algoritmo NUTS via Stan para o ajuste do modelo CPGM-Q. O Stan é um software lançado em 2012 e que atualmente está na versão 2.6. A sintaxe do Stan é de fácil uso, similar a outros softwares, tais como BUGS (LUNN *et al.*, 2000b) e JAGS (PLUMMER, 2003). Além disso, o usuário Stan conta com um guia de usuário completo (STAN DEVELOPMENT TEAM, 2017b) para facilitar a implementação de modelos estatísticos com estimação bayesiana. Apresentamos na Figura 30, no Apêndice B, o código em Stan do modelo CPGM-Q.

## **4.4 Aplicação**

Apresentamos uma aplicação para ilustrar o uso do modelo CPGM-Q utilizando diferentes matrizes  $\mathbf{Q}$ . O conjunto de dados utilizado é referente a uma subamostra de

uma pesquisa com estudantes do ensino básico na Áustria. Este conjunto de dados está disponível no pacote R chamado “CDM” (ROBITZSCH *et al.*, 2018). Ele contém respostas de 2400 indivíduos a 13 itens. Os itens 1 a 5 medem a característica “atividade cultural” e possuem três opções de resposta: 1 denota que o indivíduo nunca participa da atividade, 2 denota uma ou duas vezes por ano e 3 denota mais que duas vezes por ano. Os itens 6 a 9 medem a característica “autoconceito matemático” e possuem quatro opções de resposta: 1, 2, 3 e 4, em que um número maior indica um autoconceito mais alto. Os itens 10 a 13 medem a característica “felicidade frente à matemática” e também possuem quatro opções de resposta: 1, 2, 3 e 4, em que um número maior indica felicidade frente à matemática mais intensa.

Assim, propomos utilizar o modelo CPGM-Q com uma matriz  $\mathbf{Q}$  customizada (chamada de  $\mathbf{Q}$ -*custom*) com 13 linhas, cada uma representando um item e com 3 colunas, uma para cada dimensão do traço latente. Uma segunda proposta é utilizar o modelo CPGM-Q com uma matriz  $\mathbf{Q}$  bifatora (chamada de  $\mathbf{Q}$ -*bifat*), considerando neste cenário um traço latente global, de tal forma que cada item irá medir um traço latente específico e o traço latente global. Ambas as matrizes  $\mathbf{Q}$  são apresentadas na Tabela 9, que também contém a frequência de resposta para cada item.

Tabela 9 – Frequência das categorias de resposta em cada item e as matrizes  $\mathbf{Q}$ -*custom* e  $\mathbf{Q}$ -*bifat*

	Categorias de resposta				$\mathbf{Q}$ - <i>custom</i>			$\mathbf{Q}$ - <i>bifat</i>			
	1	2	3	4	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
Item 1	.50	.35	.15	–	1	0	0	1	1	0	0
Item 2	.64	.25	.11	–	1	0	0	1	1	0	0
Item 3	.73	.20	.07	–	1	0	0	1	1	0	0
Item 4	.18	.29	.53	–	1	0	0	1	1	0	0
Item 5	.64	.21	.15	–	1	0	0	1	1	0	0
Item 6	.31	.46	.20	.03	0	1	0	1	0	1	0
Item 7	.39	.37	.18	.05	0	1	0	1	0	1	0
Item 8	.45	.31	.15	.08	0	1	0	1	0	1	0
Item 9	.20	.41	.28	.10	0	1	0	1	0	1	0
Item 10	.04	.12	.38	.46	0	0	1	1	0	0	1
Item 11	.07	.26	.37	.29	0	0	1	1	0	0	1
Item 12	.19	.34	.26	.20	0	0	1	1	0	0	1
Item 13	.15	.33	.30	.22	0	0	1	1	0	0	1

Podemos notar que as categorias de resposta inferiores (1 e 2) possuem uma frequência maior que as categorias de resposta superiores na maioria dos itens (exceto para os itens 4, 10, 11 e 13). Além disso, a simplicidade da estrutura dos dados no tocante à relação item-traço é destacada através das matrizes  $\mathbf{Q}$ , em que vemos os itens planejados para medir apenas uma característica específica de interesse.

O ajuste do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*custom* e com a matriz  $\mathbf{Q}$ -*bifat* foi realizado através do algoritmo NUTS, utilizando 10.000 iterações, sendo a primeira

metade *burn-in* e, portanto, descartada. A verificação da convergência das distribuições a posteriori dos parâmetros foi feita através da estimativa do fator de redução de escala potencial (GELMAN; RUBIN, 1992), denotado por  $\hat{R}$ . Todos os parâmetros do modelo com as matrizes  $\mathbf{Q}$ -*custom* e  $\mathbf{Q}$ -*bifat* tiveram o valor de  $\hat{R}$  abaixo de 1,02. Como estes valores são inferiores ao valor de referência 1,1, indicado por Gelman *et al.* (2014), concluímos que método convergiu, conforme esperávamos.

Para selecionar a matriz  $\mathbf{Q}$  mais adequada para ser utilizada no modelo CPGM-Q para o conjunto de dados de respostas de estudantes da Áustria, apresentamos na Tabela 10 o  $p$ -valor Bayesiano global calculado a partir da verificação preditiva a posteriori do modelo (PPMC) proposto por Rubin (1984) e os valores de três critérios de comparação de modelos: *Deviance Information Criterion* (DIC), que pode ser encontrado em Spiegelhalter *et al.* (2002), *widely available information criterion* (WAIC), proposto por Watanabe (2010) e *leave-one-out cross-validation* (LOO-CV ou simplesmente LOO), proposto por Vehtari, Gelman e Gabry (2017). Estes 3 critérios foram escolhidos para comparar o modelo CPGM-Q com as matrizes  $\mathbf{Q}$ -*custom* e  $\mathbf{Q}$ -*bifat* por terem tido bom desempenho no estudo de simulação do Capítulo 2.

Tabela 10 – Valores dos índices de comparação de modelos DIC, WAIC e LOO,  $p$ -valor Bayesiano global e tempo, em minutos, referentes ao ajuste do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*custom* e com a matriz  $\mathbf{Q}$ -*bifat* aos dados de respostas de estudantes da Áustria

	DIC	WAIC	LOO	$p$ -valor	Tempo
Modelo CPGM-Q <i>custom</i>	49624,25	51380,53	52107,62	0,850	1335,11
Modelo CPGM-Q <i>bifat</i>	48288,95	50352,30	51338,83	0,836	876,46
$\Delta_{\text{CRIT}}$	1335,30	1028,23	768,79	–	–

Os valores dos três critérios foram menores para o modelo CPGM-Q *bifat* com diferença alta entre os valores de cada critério ( $\Delta_{\text{CRIT}}$ ). Além disso, o  $p$ -valor Bayesiano global do modelo CPGM-Q *bifat*, 0,836, pertence ao intervalo  $[0,1;0,9]$ , indicando, de acordo com Sahu (2002), um bom ajuste do modelo. A PPMC também possibilita a checagem do ajuste de um modelo da TRI aos dados item a item, como mostrado nos capítulos anteriores. Os resultados desta checagem são apresentados na Figura 19.

Obtiveram  $p$ -valor Bayesiano acima de 0,9 apenas os itens 3, 7 e 12 a partir do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*custom* e apenas os itens 3 e 12 a partir do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifat*. De acordo com Sahu (2002), o modelo CPGM-Q com ambas matrizes  $\mathbf{Q}$  é bem ajustado para a maioria dos itens.

Assim, a partir dos índices de ajuste apresentados na Tabela 10 e do  $p$ -valor Bayesiano por item apresentado na Figura 19, consideramos o modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifat* o mais adequado para os dados de respostas de estudantes da Áustria.

Apresentamos na Figura 20 as estimativas dos parâmetros de discriminação global

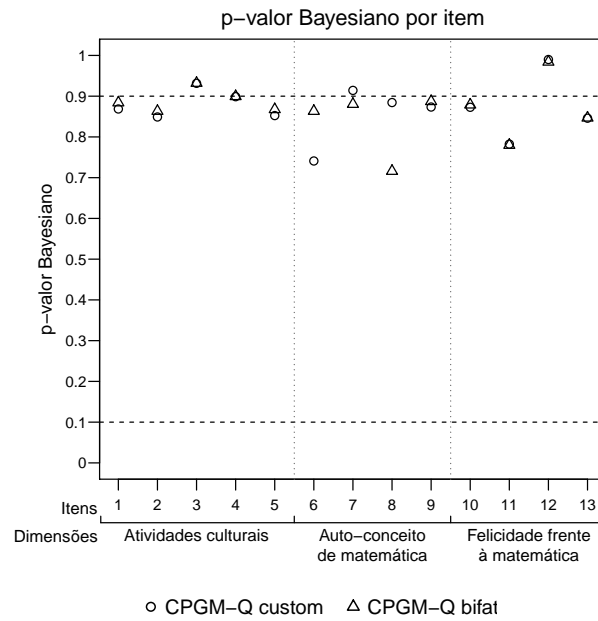


Figura 19 –  $p$ -valor Bayesiano por item para os dados de respostas de estudantes da Áustria.

e específica juntamente com seus intervalos de credibilidade de 95% encontrados a partir do modelo CPGM-Q com a matriz  $Q$ -*bifat*.

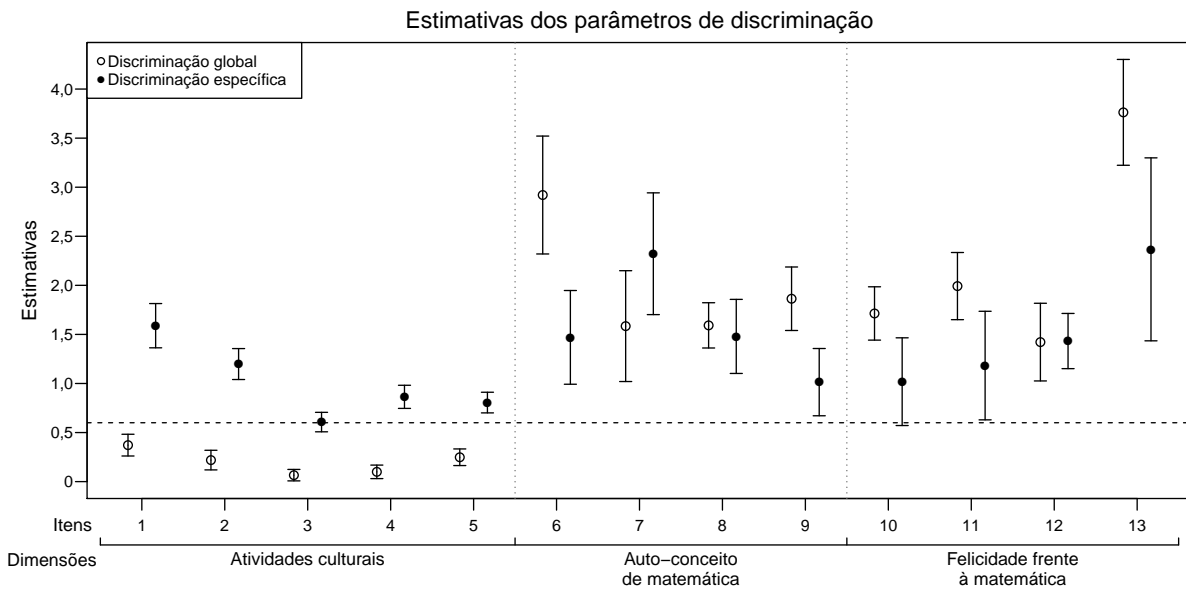


Figura 20 – Estimativas dos parâmetros de discriminação global e específica para os dados de respostas de estudantes da Áustria.

O item 3 (dimensão “Atividades culturais”) possui a menor estimativa para os parâmetros de discriminação global e específico (0,07 e 0,61, respectivamente). Já o item 13 (dimensão “Felicidade frente à Matemática”) possui a maior estimativa para os parâmetros de discriminação global e específica (3,76 e 2,37, respectivamente). Especificamente acerca dos cinco itens da dimensão “Atividades culturais”, vemos que as estimativas dos



parâmetros de discriminação global estão abaixo do valor de referência tipicamente utilizado em modelos dicotômicos com função de ligação probito, que é 0,6 (AZEVEDO; BOLFARINE; ANDRADE, 2012), o que não ocorre com as estimativas de discriminação específica. Assim, podemos afirmar que as probabilidades de responder a esses itens são mais influenciadas pelos traços específicos do que pelo traço global. Este cenário é equilibrado para as demais dimensões. Os itens 7 e 8 (dimensão “Auto-conceito de matemática”) e o item 12 (dimensão “Felicidade frente à matemática”) possuem discriminações específicas maiores que as discriminações globais, enquanto os itens 6 e 9 (dimensão “Auto-conceito de matemática”) e os itens 10, 11 e 13 (dimensão “Felicidade frente à matemática”) possuem discriminações globais maiores que as discriminações específicas.

Apresentamos na Figura 21 as estimativas dos parâmetros de dificuldade juntamente com seus intervalos de credibilidade de 95% encontrados a partir do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifat*.

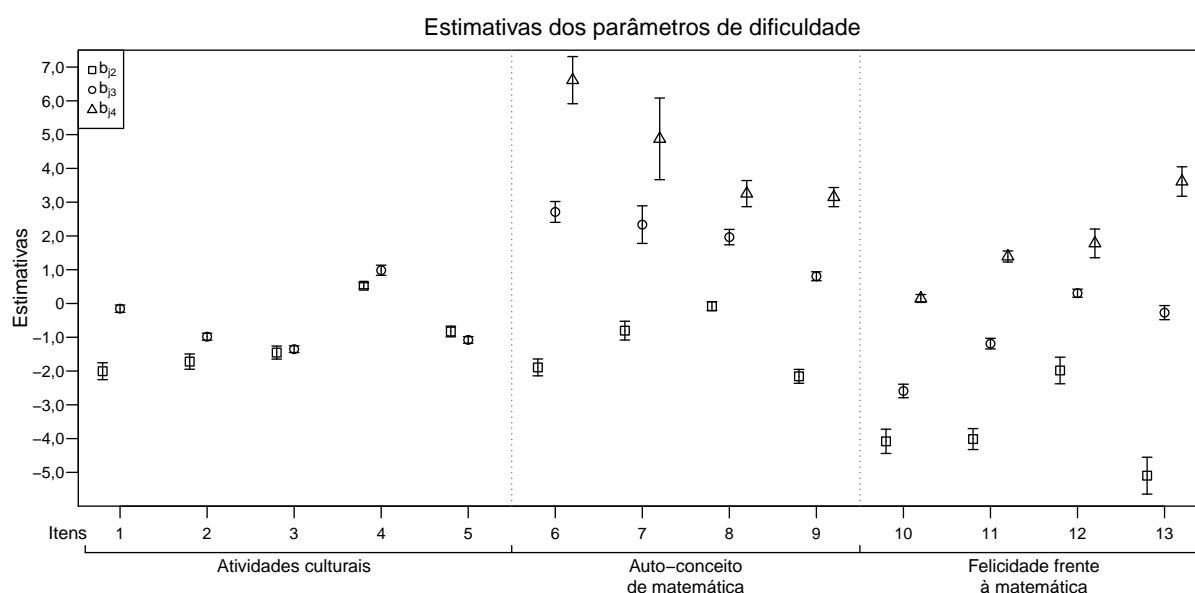


Figura 21 – Estimativas dos parâmetros de dificuldade de acordo com as categorias de resposta dos itens para os dados de respostas de estudantes da Áustria.

Cada uma das três dimensões possui itens com diferentes níveis de dificuldade, ou seja, itens com baixa dificuldade e itens com alta dificuldade. A dimensão “Atividades culturais” possui dificuldades entre  $-2,00$  e  $0,98$ , a dimensão “Auto-conceito de matemática” possui dificuldade entre  $-2,16$  e  $6,61$  e a dimensão “Felicidade frente à matemática” possui dificuldade entre  $-5,10$  e  $3,61$ . Esta variação de dificuldade dentro de cada dimensão torna possível a avaliação de diferentes níveis de traços latentes de indivíduos.

Apresentamos na Figura 22 quatro *boxplots* representando as distribuições das estimativas dos parâmetros de indivíduos separadas por dimensão encontradas a partir do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifat*.

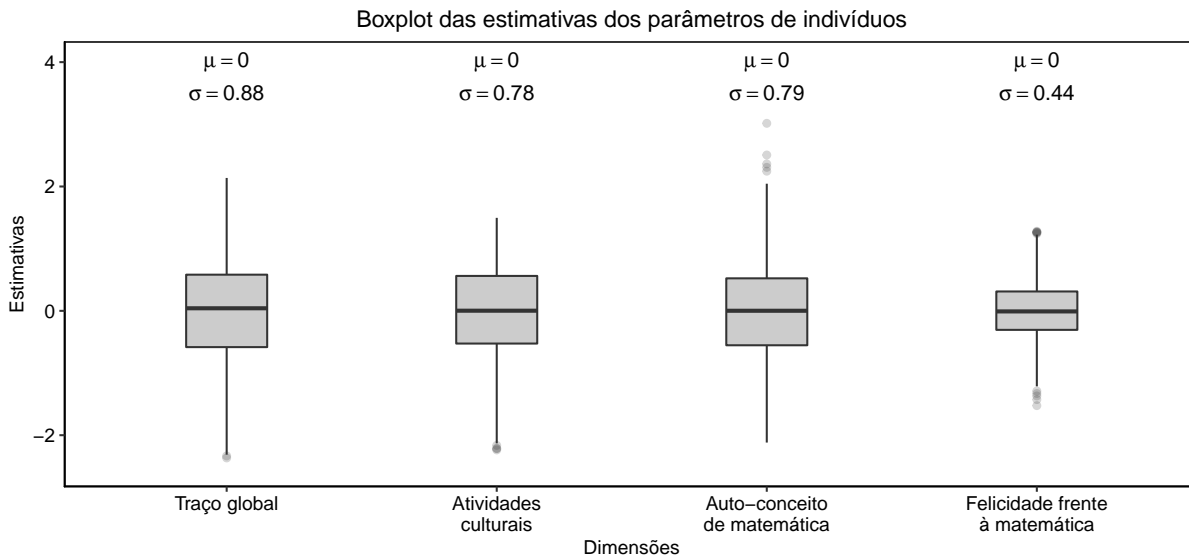


Figura 22 – Resumo das estimativas dos parâmetros de indivíduos.

As quatro dimensões do traço latente apresentam média igual a zero. Entretanto, o traço global possui o maior desvio padrão e a dimensão “Felicidade frente à matemática” possui o menor desvio padrão. Em termos práticos, as estimativas dos parâmetros de indivíduos são úteis para programar estratégias de ação para modificar aspectos culturais e de opinião dos estudantes acerca da matemática.

Para descrever graficamente a relação entre diferentes níveis do traço latente e a probabilidade de resposta a uma particular categoria, construímos as curvas características das categorias dos itens (CCCI) condicionais (TOLAND *et al.*, 2017), determinadas ao fixar o valor de uma das dimensões do traço latente (ou global ou específica) na média, ou seja, igual a zero. Assim, é possível visualizar a variação da probabilidade à medida que o valor de uma das dimensões do traço latente varia, dado que a outra dimensão está fixada em sua média. As Figuras 23 e 24 mostram as CCCIs condicionais dos 13 itens dos dados de respostas de estudantes da Áustria encontradas a partir do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifac*.

Na Figura 23, as CCCIs condicionais referentes aos itens 1 à 5 possuem pouca inclinação, ilustrando os efeitos dos valores pequenos das discriminações globais apresentadas na Figura 20, que são inferiores a 0,6. Além disso, notamos que a CCCI condicional referente à categoria 4 do item 7 não é predominante para nenhum valor do traço latente global entre  $-3$  e  $+3$ , ou seja, para qualquer valor de  $\theta_0 \in [-3, +3]$ , a probabilidade de resposta 4 no item 7 não é superior às demais categorias de resposta, dado o traço latente específico fixado na média.

Na Figura 24, as CCCIs condicionais referentes aos itens 3 (categoria 1), 6 (categoria 4) e 13 (categoria 1) também mostram que, para qualquer valor do traço latente específico entre  $-3$  e  $+3$ , a probabilidade de resposta não é superior às demais catego-

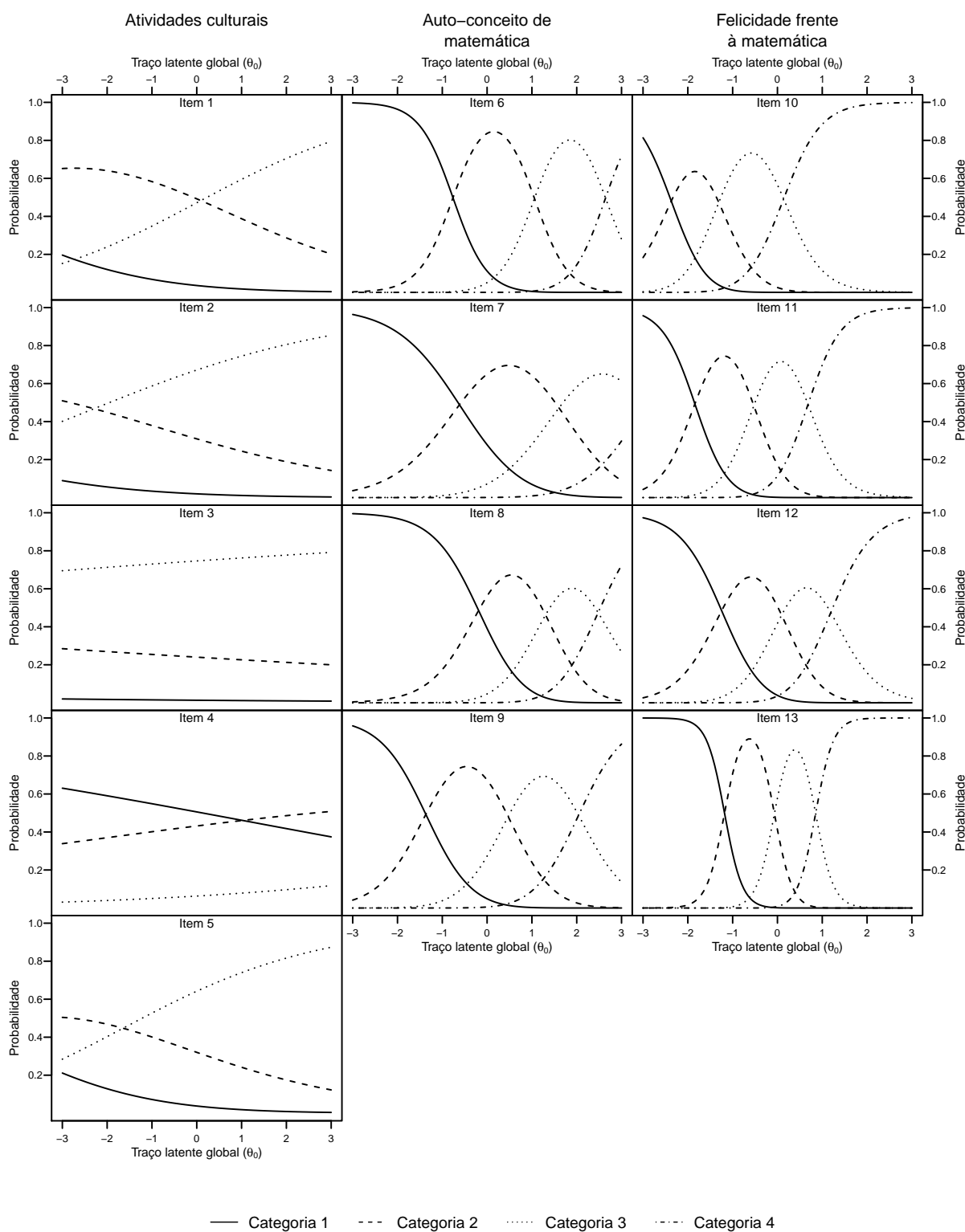


Figura 23 – Curvas características condicionais do item considerando o valor do traço latente específico como zero.

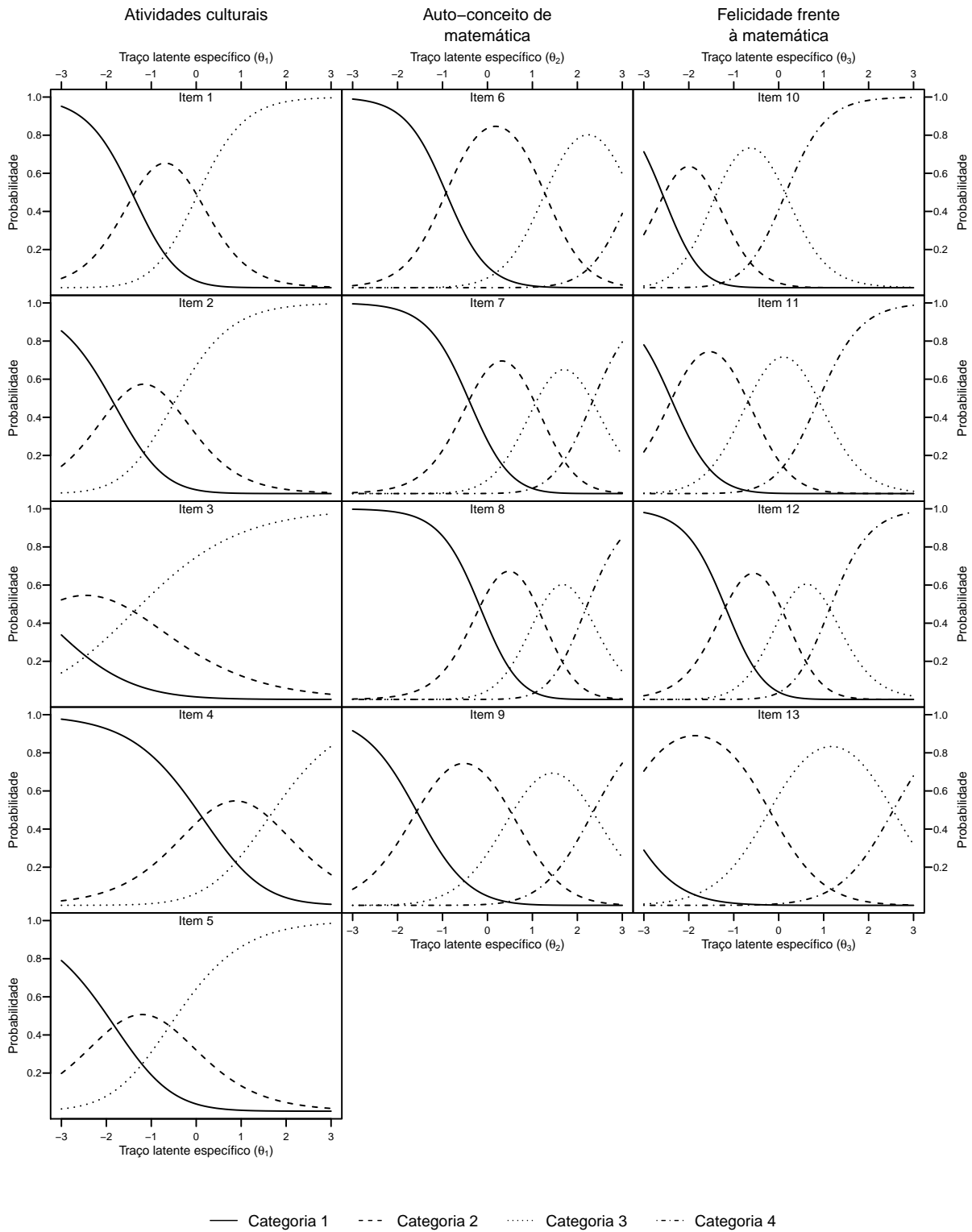


Figura 24 – Curvas características condicionais do item considerando o valor do traço latente global como zero.

rias de resposta, dado o traço latente global fixado na média. Como se trata de CCCIs condicionais, uma investigação mais detalhada é necessária antes de decidir, por exemplo, retirar estas categorias de resposta de seus respectivos itens.

## 4.5 Conclusões e comentários

Neste capítulo, propomos reescrever os modelos da TRIM de tal forma a considerar informações da relação item-traço através de um instrumento amplamente utilizado em MDC: a matriz  $\mathbf{Q}$ . Esta modificação fornece ao usuário não apenas uma maneira fácil e eficiente de especificar a relação item-traço em um problema prático, mas também estabelece uma proximidade entre os modelos da TRIM e os MDC. Além disso, o uso da matriz  $\mathbf{Q}$  em modelos da TRIM diminui o número de parâmetros de discriminação do modelo, simplificando o processo de estimação, aumentando a precisão das estimativas de traços latentes e, na maioria das vezes, encurtando o tempo de estimação.

Apesar do uso da matriz  $\mathbf{Q}$  em modelos da TRIM proporcionar os benefícios mencionados, devemos ter cautela para definir e utilizar a matriz  $\mathbf{Q}$ , pois é recomendado que ela seja definida com ajuda de especialistas na área do teste (CHEN *et al.*, 2015; LIU; HUGGINS-MANLEY; BRADSHAW, 2017). Além disso, informações na literatura e suporte empírico são de grande importância para a construção da matriz  $\mathbf{Q}$ .

Os resultados apresentados na aplicação deste capítulo exemplificam o uso do modelo CPGM-Q e ajudam o leitor a entender a sua estrutura e a interpretação de seus parâmetros. O modelo CPGM-Q foi ajustado aos dados de respostas de estudantes da Áustria utilizando duas matrizes  $\mathbf{Q}$ . A primeira, denominada matriz  $\mathbf{Q}$ -*custom*, é definida para customizar a relação item-traço, tal que cada item passa a avaliar apenas uma dimensão específica do traço latente. A segunda matriz utilizada na aplicação é denominada matriz  $\mathbf{Q}$ -*bifat* e é definida considerando neste cenário um traço latente global de tal forma que cada item passa a medir um traço latente específico e o traço latente global. Os índices de ajuste utilizados para comparar a adequação do modelo CPGM-Q aos dados com ambas matrizes  $\mathbf{Q}$  indicaram o modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifat* como a melhor opção de ajuste. Assim, apresentamos as estimativas dos parâmetros do modelo CPGM-Q com a matriz  $\mathbf{Q}$ -*bifat*, seus intervalos de credibilidade de 95%. Além disso, apresentamos as CCCIs condicionais e sua interpretação.



---

# VALIDAÇÃO DA MATRIZ $\mathbf{Q}$ EM MODELOS DA TRIM

---

---

## 5.1 Introdução

No Capítulo 4, incorporamos a relação entre os itens do teste e as dimensões do traço latente dos indivíduos na formulação dos modelos da teoria da resposta ao item multidimensionais (TRIM) através da matriz  $\mathbf{Q}$ , um componente presente na grande maioria dos modelos de diagnóstico cognitivo (MDC). Para utilizar um modelo com matriz  $\mathbf{Q}$  na prática, seja ele um MDC ou um modelo da TRIM com matriz  $\mathbf{Q}$  incorporada em sua formulação, é necessário primeiro estabelecer a relação entre itens e traço latente construindo uma matriz  $\mathbf{Q}$  apropriada. Embora este processo de construção da matriz  $\mathbf{Q}$  seja tipicamente feita por especialistas no tema dos itens, ele é um processo subjetivo e, portanto, pode implicar em equívocos, resultando em importantes problemas práticos, como, por exemplo, influenciar a estimação dos parâmetros do modelo (TORRE; CHIU, 2016). Por conta disso, a verificação da adequabilidade da matriz  $\mathbf{Q}$  através de métodos empíricos é recomendada (CHEN *et al.*, 2015; LIU; HUGGINS-MANLEY; BRADSHAW, 2017).

Neste capítulo, propomos um método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM. Em particular, utilizamos no estudo o modelo de crédito parcial generalizado multidimensional com matriz  $\mathbf{Q}$  incorporada em sua formulação. O método baseia-se na proposta de validação da matriz  $\mathbf{Q}$  de Torre (2008) para o modelo DINA, em que são considerados os dados disponíveis, o modelo escolhido e um critério projetado para fornecer informação máxima sobre alguma propriedade de interesse. O conhecimento especializado sobre a área do teste não é considerado no método.

Este capítulo está organizado da seguinte forma: nesta seção apresentamos uma visão geral da necessidade de validar uma matriz  $\mathbf{Q}$  em modelos que utilizam este com-

ponente e a proposta deste capítulo. Na Seção 5.2, apresentamos o modelo DINA e o método de validação da matriz  $\mathbf{Q}$  baseado em uma medida de discriminação dos itens. Na Seção 5.3, apresentamos os critérios sugeridos para serem utilizados no método proposto e o algoritmo utilizado para busca de matriz  $\mathbf{Q}$ . Na Seção 5.4, realizamos um estudo de simulação para verificar o desempenho dos critérios propostos. Por fim, na Seção 5.6, apresentamos algumas conclusões e comentários.

## 5.2 O método de validação da matriz $\mathbf{Q}$ para o modelo DINA

Nesta seção, descrevemos o método de validação da matriz  $\mathbf{Q}$  proposto por Torre (2008) para o modelo DINA. Para isto, apresentamos o modelo DINA e definimos o índice de discriminação dos itens utilizado na validação da matriz  $\mathbf{Q}$ .

### Visão geral do modelo DINA

O modelo DINA (*deterministic inputs, noisy “and” gate*) é apropriado para itens que requerem uma conjunção de atributos igualmente importantes para serem corretamente respondidos. Ele é considerado um modelo parcimonioso por ter sua formulação simples e com poucos parâmetros.

Para apresentar o modelo DINA, vamos considerar um instrumento de medida com  $J$  itens aplicados a  $N$  indivíduos, visando classificá-los de acordo com seus domínios em  $K$  habilidades. Como um indivíduo pode ou não possuir uma habilidade, existem  $C = 2^K$  possíveis classes de atributos, que são denotados por vetores  $\alpha_c = (\alpha_{c1}, \alpha_{c2}, \dots, \alpha_{cK})'$ , tal que  $\alpha_{ck} = 1$ , se o perfil  $c$  possui a habilidade  $k$  e  $\alpha_{ck} = 0$ , caso contrário, com  $c = 1, 2, \dots, C$  e  $k = 1, 2, \dots, K$ . Então, consideramos a variável aleatória  $Y_{ij}$  como a resposta do indivíduo  $i$  ao item  $j$ , para  $i = 1, 2, \dots, N$  e  $j = 1, 2, \dots, J$ , tal que

$$Y_{ij} | \alpha_i, \Omega_j \sim \text{Bernoulli}(P_{ij}), \quad (5.1)$$

em que  $\Omega_j = (g_j, s_j)'$  é o vetor de parâmetros do item  $j$  tal que  $g_j$ , denominado *guessing*, é a probabilidade de um indivíduo que não possui as habilidades necessárias requisitadas pelo item  $j$  “adivinhar” a resposta correta deste item e  $s_j$ , denominado *slipping*, é a probabilidade de um indivíduo que possui as habilidades necessárias requisitadas pelo item  $j$  “escorregar”, ou seja, errar a questão e  $P_{ij}$  é a probabilidade do indivíduo  $i$  responder corretamente o item  $j$ , definida por

$$P_{ij} = P(Y_{ij} = 1 | \alpha_i, \Omega_j) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}, \quad (5.2)$$



em que  $\eta_{ij}$  assume o valor 1 se, e somente se, o indivíduo  $i$  possui todas as habilidades requisitadas para responder corretamente o item  $j$ , ou seja,

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} = \mathbb{1}(\alpha'_i \mathbf{q}_j = \mathbf{q}'_j \mathbf{q}_j)$$

em que  $\mathbb{1}(\cdot)$  é a função indicadora. Os parâmetros *guessing* e *slipping* podem ser definidos em termos de  $\eta_{ij}$ , tal que

$$g_j = P(Y_{ij} = 1 | \eta_{ij} = 0) \text{ e } s_j = P(Y_{ij} = 0 | \eta_{ij} = 1).$$

Sob o modelo DINA, fixado um item  $j$ , as  $C$  possíveis classes de atributos são divididas em dois grupos: o grupo em que  $\eta_{j0}$ , formado pelas classes de atributos com pelo menos um dos atributos requisitados pelo item  $j$  ausente e o grupo em que  $\eta_{j1}$ , formado pelas classes de atributos com todos os atributos requisitados para o item  $j$ . Para ilustrar, vamos supor um teste destinado a avaliar 3 habilidades distintas e um item  $j$  específico deste teste que requer que um indivíduo possua domínio sobre a habilidade 1 para que ele responda corretamente o item, ou seja,  $\mathbf{q}_j = (1, 0, 0)'$ . A Tabela 11 apresenta as 8 classes de atributos existentes para este contexto separadas de acordo com os grupos  $\eta_{j0}$  e  $\eta_{j1}$ .

Tabela 11 – As oito classes de atributos dispostas em dois grupos

$\eta_{j0}$	$\eta_{j1}$
(0, 0, 0)	(1, 0, 0)
(0, 1, 0)	(1, 1, 0)
(0, 0, 1)	(1, 0, 1)
(0, 1, 1)	(1, 1, 1)

Note que, a partir da Equação 5.2, indivíduos com classes de atributos em um mesmo grupo tem a mesma probabilidade de responder corretamente o item  $j$ . Especificamente, indivíduos com classes de atributos pertencentes ao grupo  $\eta_{j0}$  tem probabilidade de responder corretamente o item  $j$  igual a  $g_j$ , enquanto indivíduos com classes de atributos pertencentes ao grupo  $\eta_{j1}$  tem probabilidade de responder corretamente o item  $j$  igual a  $1 - s_j$ . Assim, para que o modelo seja coerente, é necessário que a restrição  $1 - s_j > g_j$  seja verificada (JUNKER; SIJTSMA, 2001).

### O índice de discriminação dos itens

Comumente, nos modelos da TRI, um dos parâmetro dos itens a ser estimado é o parâmetro de discriminação, que indica o poder do item em distinguir indivíduos com diferentes valores de traços latentes (ANDRADE; TAVARES; VALLE, 2000). Para o caso

dicotômico, o parâmetro de discriminação de um item exerce influência na curva característica do item (CCI), enquanto que, para o caso politômico, o parâmetro de discriminação de uma categoria de resposta de um item exerce influência na curva característica da categoria do item (CCCI). Especificamente, valores muito altos implicam em curvas bastante inclinadas, gerando diferenças altas entre probabilidades de resposta de indivíduos com valores de traço latente em torno da dificuldade do item.

Em contraste, o modelo DINA não possui em sua definição um parâmetro de item semelhante ao parâmetro de discriminação nos modelos da TRI. Para contornar isto, Torre (2008) propôs um índice de discriminação para o item  $j$ , denotado por  $\delta_j$  e definido como a diferença das probabilidades de sucesso entre o grupo  $\eta_{j1}$  e o grupo  $\eta_{j0}$ , ou seja,

$$\delta_j = 1 - s_j - g_j. \quad (5.3)$$

Assim, itens que diferenciam mais os indivíduos pertencentes aos grupos  $\eta_{j0}$  e  $\eta_{j1}$  possuem  $\delta_j$  maiores, fazendo alusão à interpretação dos parâmetros de discriminação nos modelos da TRI.

### **O método de validação da matriz $\mathbf{Q}$**

Torre (2008) notou que o índice de discriminação  $\delta_j$  não é uma característica inerente do item  $j$ , mas sim uma característica que depende do vetor  $\mathbf{q}_j$  escolhido para o item. Assim, o método de validação da matriz  $\mathbf{Q}$  utilizando o modelo DINA proposto por Torre (2008) consiste em encontrar um vetor  $\mathbf{q}_j$  que forneça o maior valor de  $\delta_j$  para cada item  $j$ . Ou seja,

$$\mathbf{q}_j = \arg \max_{\alpha_c}(\delta_j) = \arg \max_{\alpha_c}(1 - s_j - g_j) = \arg \min_{\alpha_c}(s_j + g_j), \quad (5.4)$$

para  $j = 1, 2, \dots, J$ . A ideia deste método é encontrar uma matriz  $\mathbf{Q}$  que possibilite modelar probabilidades de sucesso mais heterogêneas entre os grupos  $\eta_{j0}$  e  $\eta_{j1}$  para cada item  $j$  através do modelo DINA. Uma matriz  $\mathbf{Q}$  mal especificada pode levar a uma modelagem com baixo poder de distinguir os grupos latentes.

A busca pela matriz  $\mathbf{Q}$  que maximiza os índices de discriminação dos itens sob o modelo DINA pode ser realizada através de um algoritmo de busca exaustiva, ou seja, para cada uma das  $2^K - 1$  possibilidades atribuídas ao vetor  $\mathbf{q}_j$  ( $2^K$  possíveis classes de atributos, exceto a sequência de zeros), o índice de discriminação do item  $j$ ,  $\delta_j$ , é calculado. Então, a classe de atributos que maximiza  $\delta_j$  é escolhida para preencher a  $j$ -ésima linha da matriz  $\mathbf{Q}$ . Isto é feito item a item separadamente.

## **5.3 Validação da matriz $\mathbf{Q}$ em modelos da TRIM**

Baseados no método empírico para validar as especificações da matriz  $\mathbf{Q}$  proposto por Torre (2008) para o modelo DINA descrito na Seção 5.2, propomos um método de

validação da matriz  $\mathbf{Q}$  em modelos da TRIM. Especificamente, utilizamos no estudo o modelo de crédito parcial generalizado multidimensional com matriz  $\mathbf{Q}$  incorporada em sua formulação (modelo CPGM-Q), apresentado no Capítulo 4.

A ideia do método é encontrar uma matriz  $\mathbf{Q}$  de forma que algum critério de interesse seja otimizado. Para isso, a seguir descrevemos o modelo CPGM-Q, definimos alguns critérios a serem otimizados para a busca da matriz  $\mathbf{Q}$  e descrevemos o algoritmo de busca de matrizes utilizado neste capítulo.

### O modelo CPGM-Q

Para apresentar o modelo CPGM-Q, consideramos um teste composto de  $J$  itens designados para avaliar  $D$  dimensões do traço latente. Cada item  $j$  possui  $m_j > 2$  categorias de resposta. Assim, definimos a variável aleatória  $Y_{ij}$  como a resposta do indivíduo  $i$  ao item  $j$ , tal que

$$Y_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\xi}_j \sim \text{Categorizada}(P_{ij1}, P_{ij2}, \dots, P_{ijm_j}), \quad (5.5)$$

em que  $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iD})$  é o vetor de parâmetros do indivíduo  $i$  tal que  $\theta_{id}$  é a  $d$ -ésima dimensão do traço latente pertencente ao indivíduo  $i$ ;  $\boldsymbol{\xi}_j = (a_{j1}, a_{j2}, \dots, a_{jD}, b_{j1}, b_{j2}, \dots, b_{jm_j})$  é o vetor de parâmetros associado com o  $j$ -ésimo item, sendo  $a_{jd}$  o parâmetro de discriminação específico da dimensão  $d$  e  $b_{jk}$  o parâmetros de dificuldade referente à  $k$ -ésima categoria do item  $j$ ; e  $P_{ijk}$  é a probabilidade do indivíduo  $i$  responder a  $k$ -ésima categoria de resposta do item  $j$ , definida como

$$P_{ijk} = P(Y_{ij} = k | \boldsymbol{\theta}_i, \boldsymbol{\xi}_j) = \frac{\exp\left(k \sum_{d=1}^D q_{jd} a_{jd} \theta_{id} - \sum_{h=1}^k b_{jh}\right)}{\sum_{l=1}^{m_j} \exp\left(l \sum_{d=1}^D q_{jd} a_{jd} \theta_{id} - \sum_{h=1}^l b_{jh}\right)}. \quad (5.6)$$

A verossimilhança para o modelo CPGM-Q pode ser escrita como

$$\ell(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i=1}^N \prod_{j=1}^J \prod_{k=1}^{m_j} [P_{ijk}]^{w_{ijk}}, \quad (5.7)$$

em que  $\mathbf{y}$  o vetor que contém as respostas dos  $N$  indivíduos aos  $J$  itens,  $\boldsymbol{\theta}$  é o vetor dos parâmetros de indivíduos,  $\boldsymbol{\xi}$  é o vetor de parâmetros de todos os itens e  $w_{ijk} = 1$ , se a  $k$ -ésima categoria de resposta do item  $j$  é selecionado pelo indivíduo  $i$  e  $w_{ijk} = 0$ , caso contrário. Esta variável faz com que a verossimilhança considere apenas as probabilidades das categorias de resposta selecionadas pelos indivíduos, permitindo o uso de conjunto de dados com respostas faltantes.

Como a metodologia utilizada exige que o modelo seja ajustado aos dados muitas vezes com diferentes matrizes  $\mathbf{Q}$ , métodos Bayesianos tornam-se inviáveis devido ao tempo computacional. Assim, neste capítulo, a estimação dos parâmetros do modelo CPGM-Q é

feita utilizando um algoritmo de maximização de expectativa (algoritmo EM) via pacote do software R chamado “mirt” (CHALMERS, 2012).

### Critérios propostos

Os critérios de interesse foram propostos de acordo com propriedades desejáveis na análise dos dados, assim como Torre (2008) propôs utilizar o índice de discriminação dos itens como critério de busca da matriz  $\mathbf{Q}$  para o modelo DINA. Especificamente, propomos cinco critérios para investigar o comportamento deles na busca da matriz  $\mathbf{Q}$  para modelos da TRIM. O primeiro deles, denominado  $DISC_{\text{média}}$ , consiste em encontrar uma matriz  $\mathbf{Q}$  que maximize a média dos parâmetros de discriminação dos itens, ou seja,

$$\max_{\mathbf{Q} \in \Theta} \frac{\sum_{(j,d) \in \mathbb{C}_Q} a_{jd}}{\sum_{j=1}^J \sum_{d=1}^D q_{jd}}, \quad (5.8)$$

em que  $\mathbb{C}_Q = \{(j,d) : q_{jd} = 1, j = 1, 2, \dots, J, d = 1, 2, \dots, D\}$ . O segundo critério, chamado  $MDISC_{\text{média}}$ , propõe encontrar uma matriz  $\mathbf{Q}$  que maximize a média dos parâmetros de discriminação multidimensional, isto é,

$$\max_{\mathbf{Q} \in \Theta} \frac{\sum_{j=1}^J MDISC_j}{J}, \quad (5.9)$$

em que  $MDISC_j = \sqrt{\sum_{d=1}^D q_{jd} a_{jd}^2}$  é o parâmetro de discriminação multidimensional para o item  $j$ . O terceiro critério, denominado  $DISC_{\text{mín}}$ , visa buscar uma matriz  $\mathbf{Q}$  que maximize a discriminação mínima, ou seja,

$$\max_{\mathbf{Q} \in \Theta} \min_{(j,d) \in \mathbb{C}_Q} \{a_{jd}\}. \quad (5.10)$$

Em contraste, o quarto critério utilizado no estudo, chamado  $AIC$ , não está associado aos parâmetros de discriminação como os três primeiros. Visando um bom ajuste do modelo aos dados, propomos buscar uma matriz  $\mathbf{Q}$  que minimize o valor do *Akaike information criterion* (AKAIKE, 1973), isto é,

$$\min_{\mathbf{Q} \in \Theta} (-2 \log(L) + 2p), \quad (5.11)$$

em que  $L$  é o valor máximo da verossimilhança e  $p$  é o número de parâmetros estimados no modelo. O componente  $2 \times p$  é entendido como uma função de penalidade para evitar superparametrização no modelo. Assim como o quarto critério, o quinto critério utilizado no estudo, chamado  $BIC$ , também visa um bom ajuste do modelo aos dados. Especificamente, propomos buscar uma matriz  $\mathbf{Q}$  que minimize o valor do *Bayesian information criterion* (SCHWARZ, 1978), isto é,

$$\min_{\mathbf{Q} \in \Theta} (-2 \log(L) + p \log(n)), \quad (5.12)$$

em que  $n = N \times J$  é o número total de observações.

### Algoritmo de troca

Em contraste ao método de validação da matriz  $\mathbf{Q}$  para o modelo DINA, que constrói uma matriz  $\mathbf{Q}$  ótima linha a linha separadamente, o método de validação da matriz  $\mathbf{Q}$  para modelos da TRIM deve considerar todos os itens simultaneamente para construir uma matriz  $\mathbf{Q}$  ótima, pois na TRI existe uma interação entre itens e traços latentes de modo a impossibilitar a construção de uma matriz  $\mathbf{Q}$  ótima linha a linha. Assim, uma forma de buscar uma matriz  $\mathbf{Q}$  ótima é através da busca exaustiva, ou seja, após definido o número de dimensões do traço latente, o número de itens e o critério utilizado, devemos construir todas as possíveis matrizes  $\mathbf{Q}$ , calcular o valor do critério para cada uma delas e selecionar a matriz com o valor de critério ótimo. Entretanto, a quantidade de matrizes  $\mathbf{Q}$  distintas aumenta exponencialmente de acordo com a dimensão delas, tornando o método de busca exaustiva inviável. Uma alternativa é utilizar um método eficiente e sistemático para a busca de matrizes ótimas, como, por exemplo, um algoritmo de troca (FEDOROV, 1972), que é um método proposto originalmente para buscar matrizes de delineamentos ótimos (ATKINSON; DONEV; TOBIAS, 2007).

Uma das principais versões do algoritmo de troca é chamada de algoritmo de troca por ponto (*point-exchange algorithm*) e pode ser encontrada em (NGUYEN; MILLER, 1992). A ideia deste método é realizar subsequentes trocas dos pontos da matriz, ou seja, das linhas da matriz, visando melhorar o critério considerado e, assim, construir uma matriz ótima para o critério. A busca é repetida para um número pré-determinado de matrizes iniciais distintas, a fim de que a solução encontrada tenha maior chance de ser um ótimo global e não apenas um ótimo local. Uma representação gráfica do algoritmo de troca por ponto adaptado para busca de matriz  $\mathbf{Q}$  é fornecido na Figura 25.

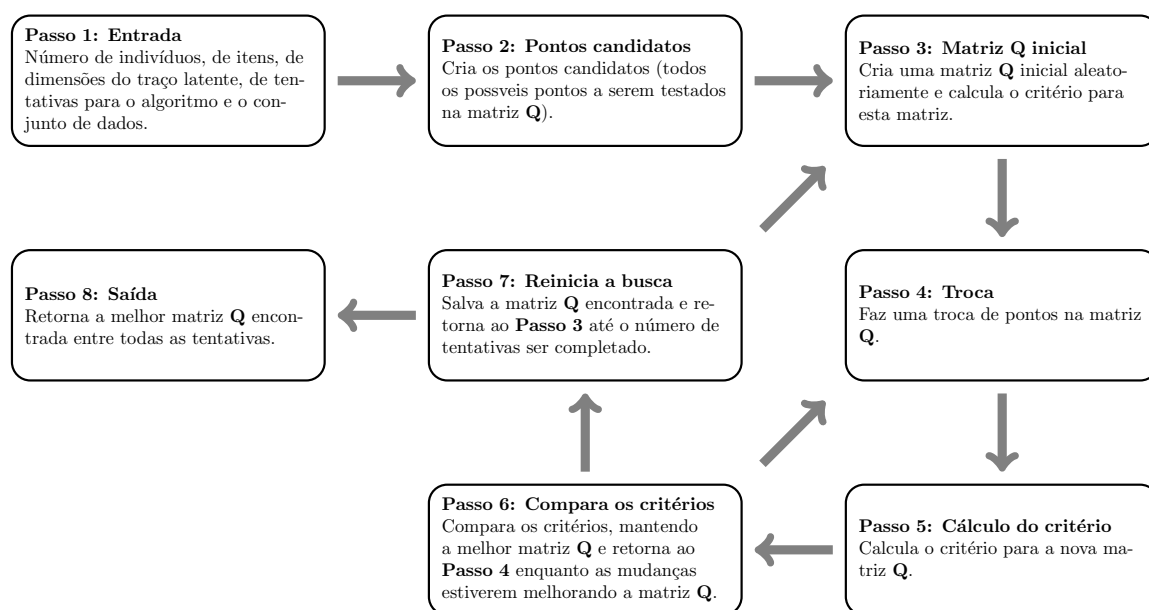


Figura 25 – Representação gráfica do algoritmo de troca por ponto adaptado para busca de matriz  $\mathbf{Q}$ .

Além da representação gráfica, apresentamos no Algoritmo 1 um pseudocódigo do algoritmo de troca por ponto adaptado para busca de matriz  $\mathbf{Q}$ .

---

**Algoritmo 1:** Algoritmo de Troca por Ponto adaptado para busca de matriz  $\mathbf{Q}$

---

**Entrada:**  $N, J, D, nTentativas$ ;  
 $N$ : número de indivíduos;  
 $J$ : número de itens;  
 $D$ : número de dimensões do traço latente;  
 $nTentativas$ : número de tentativas do algoritmo;

- 1 Construir um vetor de pontos candidatos  $cand$  com  $nCand$  pontos;
- 2  $m \leftarrow 1$ ;
- 3 **Para**  $m \leq nTentativas$  **faça**
- 4      $\mathbf{Q} \leftarrow$  matriz binária inicial gerada aleatoriamente;
- 5      $crit \leftarrow$  cálculo do valor do critério;
- 6      $melhorou \leftarrow 1$ ;
- 7     **Enquanto**  $melhorou = 1$  **faça**
- 8          $melhorou \leftarrow 0$ ;
- 9          $j \leftarrow 1$ ;
- 10        **Para**  $j \leq J$  **faça**
- 11             $c \leftarrow 1$ ;
- 12            **Para**  $c \leq nCand$  **faça**
- 13                **Se**  $\mathbf{Q}[j][.] \neq cand[c]$  **então**
- 14                     $q_{aux} \leftarrow \mathbf{Q}[j][.]$ ;
- 15                     $\mathbf{Q}[j][.] \leftarrow cand[c]$ ;
- 16                    **Se**  $\mathbf{Q}$  *não possui* *coluna de zeros* **então**
- 17                         $crit_0 \leftarrow$  cálculo do valor do critério;
- 18                        **Se**  $crit_0 > crit$  **então**
- 19                             $melhorou \leftarrow 1$ ;
- 20                             $crit \leftarrow crit_0$ ;
- 21                        **Fim**
- 22                        **Senão**
- 23                             $\mathbf{Q}[j][.] \leftarrow q_{aux}$ ;
- 24                        **Fim**
- 25                        **Fim**
- 26                        **Senão**
- 27                             $\mathbf{Q}[j][.] \leftarrow q_{aux}$ ;
- 28                        **Fim**
- 29                        **Fim**
- 30                        **Fim**
- 31                        **Fim**
- 32     **Fim**
- 33     **Se**  $m == 1$  **então**
- 34          $\mathbf{Q}_{melhor} \leftarrow \mathbf{Q}$ ;
- 35          $crit_{melhor} \leftarrow crit$ ;
- 36     **Fim**
- 37     **Senão**
- 38         **Se**  $crit > crit_{melhor}$  **então**
- 39             $\mathbf{Q}_{melhor} \leftarrow \mathbf{Q}$ ;
- 40             $crit_{melhor} \leftarrow crit$ ;
- 41         **Fim**
- 42     **Fim**
- 43 **Fim**
- 44 **Retorna**  $\mathbf{Q}_{melhor}$

---

Inicialmente, as variáveis de entrada são definidas pelo usuário: número de indivíduos, número de itens, número de dimensões do traço latente e número de tentativas

do algoritmo. Em seguida, um vetor com todos os pontos candidatos é contruído (linha 1). O procedimento a partir da linha 3 é repetido para cada tentativa pré-definida. Na linha 4, uma matriz  $\mathbf{Q}$  inicial é gerada aleatoriamente e, na linha 5, o valor do critério referente à matriz  $\mathbf{Q}$  gerada é calculado de acordo com o critério escolhido (Equações 5.8, 5.9, 5.10, 5.11 ou 5.12). Os procedimentos entre as linhas 7 e 32 são realizados enquanto o valor do critério estiver melhorando. A linha 10 é responsável por percorrer as linhas da matriz  $\mathbf{Q}$  e, para cada uma destas linhas, percorre-se todos os pontos candidatos (linha 12). Na linha 13 é verificado se a linha  $j$  atual da matriz  $\mathbf{Q}$  é diferente do ponto candidato de índice  $c$ . Em caso afirmativo, o algoritmo armazena a linha  $j$  atual da matriz  $\mathbf{Q}$  na variável auxiliar  $q_{aux}$  (linha 14), realiza a troca na matriz  $\mathbf{Q}$  (linha 15) e verifica se, após a troca, a matriz  $\mathbf{Q}$  não possui colunas de zeros (linha 16). Em caso negativo, ou seja, se a nova matriz  $\mathbf{Q}$  possui alguma coluna de zeros, a troca é desfeita (linha 27). Em caso afirmativo, o valor do critério referente à nova matriz  $\mathbf{Q}$  é calculado (linha 17) e é verificado se o critério melhorou em relação à melhor matriz  $\mathbf{Q}$  encontrada até o momento (linha 18). Em caso afirmativo, a nova matriz  $\mathbf{Q}$  passa a ser a melhor matriz  $\mathbf{Q}$  já encontrada, juntamente com o seu respectivo valor de critério (linha 20). Em caso negativo, a troca é desfeita (linha 23). Quando o algoritmo percorrer todas as linhas da matriz  $\mathbf{Q}$  e todos os pontos candidatos e não obter melhora na matriz  $\mathbf{Q}$ , a melhor matriz  $\mathbf{Q}$  encontrada e seu respectivo valor de critério são salvos (linhas 33 a 42). Em seguida, o algoritmo começa uma nova tentativa até completar o número de tentativas pré-definido.

Embora o algoritmo de troca seja mais eficiente do que a busca exaustiva, ele é um método heurístico, ou seja, não há garantia de que a matriz encontrada seja ótima. Além disso, o número de pontos candidatos aumenta exponencialmente à medida que o número de dimensões do traço latente  $D$  aumenta. Especificamente, são, ao todo,  $2^D - 1$  pontos candidatos (todas as combinações de zeros e uns, exceto o ponto formado apenas de zeros). Como consequência, o tempo de execução do algoritmo aumenta na mesma ordem.

## 5.4 Estudo de simulação

Realizamos um estudo de simulação visando analisar o desempenho dos critérios propostos na Seção 5.3. Para isto, definimos dois cenários, um formado pelo número de indivíduos  $N = 500$ , número de dimensões  $D = 2$  e número de itens  $J = 10$ , cada item com  $m_j = 4$  categorias de resposta. O outro cenário formado pelo número de indivíduos  $N = 500$ , número de dimensões  $D = 3$  e número de itens  $J = 15$ , cada item também com  $m_j = 4$  categorias de resposta. Foram definidas 20 tentativas para o algoritmo de troca.

Para cada cenário, geramos aleatoriamente uma matriz  $\mathbf{Q}$ , denotada por  $\mathbf{Q}_V$ , e simulamos os parâmetros dos indivíduos  $\theta_{id}$ , para  $1 \leq i \leq N$  e  $1 \leq d \leq D$ , a partir de uma

distribuição normal padrão, os parâmetros de discriminação dos itens  $a_{jd}$ , para  $(j, d) \in \mathbb{C}_Q$ , a partir de uma distribuição uniforme entre 0,6 e 1,4 tal que  $0,6 < MDISC_j < 1,4$ , para  $1 \leq j \leq J$ , e os parâmetros de dificuldade dos itens  $b_{jk}$ , para  $1 \leq j \leq J$  e  $1 \leq k \leq m_j$ , a partir de uma distribuição uniforme entre  $-2,4$  e  $2,4$ . Estes valores são utilizados juntamente com a matriz  $\mathbf{Q}_V$  para simular uma matriz de respostas a partir do modelo CPGM-Q. Então, esta matriz de respostas é utilizada no algoritmo de troca para buscar a melhor matriz  $\mathbf{Q}$  de acordo com cada um dos cinco critérios propostos na Seção 5.3.

### **Cenário 1: $D = 2$ e $J = 10$**

A Tabela 12 apresenta a matriz  $\mathbf{Q}_V$  utilizada na simulação dos dados e as matrizes  $\mathbf{Q}$  encontradas pelo algoritmo de troca a partir dos cinco critérios propostos considerando o Cenário 1. Para estabelecer diferenças e comparar as matrizes, a Tabela 12 também apresenta a classificação ordenada da matriz verdadeira de acordo com cada critério utilizado (posição da matriz  $\mathbf{Q}_V$ ), o percentual de elementos de cada matriz  $\mathbf{Q}$  encontrada em comum com os elementos da matriz  $\mathbf{Q}_V$  (similaridade) e a eficiência das matrizes  $\mathbf{Q}$  em relação a cada critério, definida pela razão entre o valor do critério da matriz  $\mathbf{Q}$  encontrada e o valor do critério da melhor matriz  $\mathbf{Q}$ . Em outras palavras, a eficiência de uma matriz  $\mathbf{Q}$  qualquer em relação a um critério específico é um percentual que relaciona o valor absoluto do critério desta matriz  $\mathbf{Q}$  com o valor absoluto do critério da melhor matriz  $\mathbf{Q}$ .

A matriz I é a matriz  $\mathbf{Q}_V$  gerada aleatoriamente e utilizada para simular a matriz de respostas. Ela também é a matriz  $\mathbf{Q}$  encontrada com o menor valor dos critérios  $AIC$  e  $BIC$ . Portanto, os critérios  $AIC$  e  $BIC$  recuperaram 100% da matriz verdadeira  $\mathbf{Q}_V$ , utilizada na simulação da matriz de respostas. Além disso, a única eficiência baixa desta matriz é a eficiência no critério  $DISC_{\min}$ , com 44,22%.

A matriz II é a matriz  $\mathbf{Q}$  encontrada com o maior valor do critério  $DISC_{\text{média}}$ , ou seja, a maior média dos parâmetros de discriminação. Considerando este critério, a matriz  $\mathbf{Q}_V$  se encontra na 67ª posição, ou seja, existem pelo menos outras 66 matrizes  $\mathbf{Q}$  distintas cujos valores do critério  $DISC_{\text{média}}$  são maiores do que o valor do mesmo critério referente à matriz  $\mathbf{Q}_V$ . Além disso, a matriz II possui 60,00% de seus elementos equivalentes à matriz verdadeira  $\mathbf{Q}_V$ . Vemos que a matriz II possui 91,16% de eficiência no critério  $MDISC_{\text{média}}$ , 28,50% de eficiência no critério  $DISC_{\min}$ , 95,93% de eficiência no critério  $AIC$  e 97,41% de eficiência no critério  $BIC$ . Outro detalhe a ser observado nesta matriz é que nenhum item avalia ambas as dimensões do traço latente.

A matriz III é a matriz  $\mathbf{Q}$  encontrada com o maior valor do critério  $MDISC_{\text{média}}$ , ou seja, a maior média dos parâmetros de discriminação multidimensionais. De acordo com o algoritmo, a matriz  $\mathbf{Q}_V$  se encontra na 10ª posição considerando o critério  $MDISC_{\text{média}}$ . A matriz III possui 65,00% de seus elementos equivalentes à matriz verdadeira  $\mathbf{Q}_V$ . Além



Tabela 12 – Matriz  $\mathbf{Q}_V$  e matrizes  $\mathbf{Q}$  encontradas a partir dos cinco critérios propostos considerando o Cenário 1 do estudo de simulação

	$\mathbf{Q}_V; AIC; BIC$		$DISC_{m\acute{e}dia}$		$MDISC_{m\acute{e}dia}$		$DISC_{m\acute{m}n}$	
	I		II		III		IV	
	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
Item 1	1	0	1	0	0	1	1	0
Item 2	0	1	1	0	1	1	1	0
Item 3	1	1	1	0	1	1	0	1
Item 4	1	1	0	1	1	1	0	1
Item 5	1	1	1	0	1	1	1	1
Item 6	1	0	1	0	1	0	1	1
Item 7	0	1	0	1	0	1	1	1
Item 8	1	0	0	1	1	1	1	0
Item 9	1	1	1	0	1	0	0	1
Item 10	0	1	0	1	1	0	1	1
Posição da $\mathbf{Q}_V$	–		67 <sup>a</sup>		10 <sup>a</sup>		252 <sup>a</sup>	
Similaridade	–		60,00%		65,00%		60,00%	
Ef. $DISC_{m\acute{e}dia}$	71,08%		100,00%		79,59%		51,98%	
Ef. $MDISC_{m\acute{e}dia}$	80,34%		91,16%		100,00%		69,06%	
Ef. $DISC_{m\acute{m}n}$	44,22%		28,50%		19,11%		100,00%	
Ef. $AIC$	100,00%		95,93%		97,40%		97,33%	
Ef. $BIC$	100,00%		97,41%		96,16%		97,38%	

disso, vemos que as eficiências da matriz III são parecidas com as eficiências da matriz II, sendo baixa apenas para o critério  $DISC_{m\acute{m}n}$ .

A matriz IV é a matriz  $\mathbf{Q}$  encontrada com o maior valor do critério  $DISC_{m\acute{m}n}$ , ou seja, o maior mínimo entre os parâmetros de discriminação. A matriz  $\mathbf{Q}_V$  se encontra na 252<sup>a</sup> posição considerando o critério  $DISC_{m\acute{m}n}$ . A matriz IV possui 60,00% de seus elementos equivalentes à matriz verdadeira  $\mathbf{Q}_V$ . Além disso, vemos que as eficiências da matriz IV nos critérios  $DISC_{m\acute{e}dia}$  e  $MDISC_{m\acute{e}dia}$  são inferiores à 70%, enquanto a eficiência da matriz IV nos critérios  $AIC$  e  $BIC$  são superiores a 97%.

### Cenário 2: $D = 3$ e $J = 15$

A Tabela 13 apresenta a matriz  $\mathbf{Q}_V$ , as matrizes  $\mathbf{Q}$  obtidas pelo algoritmo de troca em cada critério proposto considerando o Cenário 2 e as mesmas medidas apresentadas na Tabela 12 para efeito de comparação entre as matrizes.

A matriz I é a matriz  $\mathbf{Q}_V$  gerada aleatoriamente e utilizada para simular a matriz de respostas. Ela também é a matriz  $\mathbf{Q}$  encontrada com o menor valor do critério  $BIC$ . Portanto, o critério  $BIC$  recupera 100% da matriz verdadeira  $\mathbf{Q}_V$ , utilizada na simulação da matriz de respostas. Além disso, assim como no Cenário 1, a única eficiência baixa desta matriz é a eficiência no critério  $DISC_{m\acute{m}n}$ , com 29,38%.

Tabela 13 – Matriz  $\mathbf{Q}_V$  e matrizes  $\mathbf{Q}$  encontradas a partir dos cinco critérios propostos considerando o Cenário 2 do estudo de simulação

	$\mathbf{Q}_V; BIC$			$DISC_{m\u00e9dia}$			$MDISC_{m\u00e9dia}$			$DISC_{m\u00edn}$			$AIC$		
	I			II			III			IV			V		
	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1$	$\theta_2$	$\theta_3$
Item 1	1	0	0	0	0	1	1	1	0	0	0	1	1	0	0
Item 2	1	0	1	1	0	0	1	1	1	1	0	1	1	0	1
Item 3	1	1	1	0	1	0	0	1	1	1	0	0	1	1	1
Item 4	1	1	1	1	0	1	1	1	0	1	0	0	1	1	1
Item 5	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
Item 6	0	1	0	0	1	0	1	1	0	0	1	0	0	1	0
Item 7	1	0	1	0	1	0	0	1	0	1	0	0	1	0	1
Item 8	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0
Item 9	0	1	1	0	1	0	1	1	0	0	0	1	1	1	1
Item 10	0	0	1	0	1	0	1	1	1	1	0	1	0	1	1
Item 11	0	1	1	0	0	1	0	1	0	0	1	1	0	1	1
Item 12	0	1	1	0	1	0	0	1	1	0	1	1	0	1	1
Item 13	0	1	0	0	1	0	1	1	1	0	1	1	0	1	0
Item 14	1	1	1	0	1	0	1	0	0	0	1	0	1	1	1
Item 15	0	0	1	0	1	1	1	0	1	1	1	1	0	0	1
Posição da $\mathbf{Q}_V$	–			781 <sup>a</sup>			60 <sup>a</sup>			250 <sup>a</sup>			2 <sup>a</sup>		
Similaridade	–			60,00%			57,78%			64,44%			95,56%		
Ef. $DISC_{m\u00e9dia}$	56,34%			100,00%			32,20%			44,77%			51,58%		
Ef. $MDISC_{m\u00e9dia}$	73,92%			91,11%			100,00%			61,29%			73,77%		
Ef. $DISC_{m\u00edn}$	29,38%			4,35%			0,00%			100,00%			29,51%		
Ef. $AIC$	99,98%			94,32%			96,42%			96,77%			100,00%		
Ef. $BIC$	100,00%			94,67%			96,47%			96,95%			99,97%		

A matriz II é a matriz  $\mathbf{Q}$  encontrada com o maior valor do critério  $DISC_{m\u00e9dia}$ , ou seja, a maior média dos parâmetros de discriminação. Podemos observar uma grande quantidade de zeros nesta matriz. São, ao todo, 28 zeros, contra 18 zeros na matriz  $\mathbf{Q}_V$ . Apenas os itens 4 e 15 avaliam mais de uma dimensão do traço latente. Além disso, vemos que existem pelo menos outras 780 matrizes  $\mathbf{Q}$  distintas com valor do critério  $DISC_{m\u00e9dia}$  superior ao da matriz  $\mathbf{Q}_V$ .

A matriz III é a matriz  $\mathbf{Q}$  encontrada com o maior valor do critério  $MDISC_{m\u00e9dia}$ , ou seja, a maior média dos parâmetros de discriminação multidimensionais. Vemos que a eficiência da matriz III no critério  $DISC_{m\u00edn}$  é próxima de zero.

A matriz IV é a matriz  $\mathbf{Q}$  encontrada com o maior valor do critério  $DISC_{m\u00edn}$ , ou seja, o maior mínimo entre os parâmetros de discriminação. Vemos que existem pelo menos outras 249 matrizes  $\mathbf{Q}$  distintas com valor do critério  $DISC_{m\u00edn}$  superior ao da matriz  $\mathbf{Q}_V$ .

A matriz V é a matriz  $\mathbf{Q}$  encontrada com o menor valor do critério  $AIC$ . Vemos que existe apenas uma matriz  $\mathbf{Q}$  distinta com valor do critério  $AIC$  superior ao da matriz  $\mathbf{Q}_V$ , que é justamente a matriz V. Apenas os elementos  $q_{9,1}$   $q_{10,2}$  diferem da matriz  $\mathbf{Q}_V$ .

Como consequência, a matriz  $V$  é 99,97% eficiente para o critério  $BIC$ .

## 5.5 Aplicação

Nesta seção, aplicamos o método proposto para validar matriz  $\mathbf{Q}$  utilizando modelos da TRIM apresentado neste capítulo em dados do Inventário da Depressão de Beck (Beck Depression Inventory, BDI), ver [Beck et al. \(1961\)](#). É o mesmo conjunto de dados utilizado no Capítulo 2. São 1.111 estudantes que respondem uma versão do BDI com 21 itens destinados a medir duas dimensões de características de depressão, a dimensão cognitiva ( $\theta_1$ ) e a dimensão somático-afetiva ( $\theta_2$ ), com intensidade variando de neutro a um nível máximo de severidade, classificado em 1 (baixa intensidade) a 4 (alta intensidade). Os dados podem ser encontrados em [Fragoso e Cúri \(2013\)](#).

A Tabela 14 contém a descrição dos 21 itens que compõem o BDI utilizado na coleta dos dados considerados nesta aplicação, a matriz  $\mathbf{Q}$  utilizada nesta versão do BDI por [Fragoso e Cúri \(2013\)](#) e [DA SILVA et al. \(2018\)](#), denotada aqui por  $\mathbf{Q}_L$ , e a frequência relativa (em porcentagem) das respostas dos 1.111 indivíduos para cada categoria de resposta.

Tabela 14 – Descrição dos itens, matriz  $\mathbf{Q}_L$  e frequência das respostas em cada categoria.

Itens	Descrição	Matriz $\mathbf{Q}_L$		Categorias de resposta			
		$\theta_1$	$\theta_2$	1	2	3	4
1	Tristeza	1	1	0,36	0,56	0,06	0,02
2	Pessimismo	1	1	0,63	0,32	0,06	0,00
3	Sentimentos de fracasso	1	0	0,69	0,25	0,05	0,01
4	Insatisfação	1	1	0,49	0,40	0,11	0,00
5	Culpa	1	0	0,69	0,26	0,04	0,01
6	Punição	1	0	0,66	0,34	0,00	0,00
7	Decepção	1	0	0,45	0,37	0,15	0,03
8	Inferioridade	1	0	0,36	0,57	0,05	0,02
9	Pensamentos suicidas	1	1	0,86	0,12	0,02	0,00
10	Choro	1	1	0,72	0,18	0,10	0,00
11	Irritação	0	1	0,49	0,36	0,15	0,00
12	Retraimento social	1	1	0,64	0,28	0,07	0,01
13	Indecisão	1	1	0,61	0,21	0,14	0,04
14	Aparência	1	1	0,65	0,16	0,13	0,06
15	Dificuldade no trabalho	1	1	0,53	0,34	0,11	0,02
16	Insônia	0	1	0,51	0,40	0,09	0,00
17	Cansaço	0	1	0,40	0,47	0,11	0,03
18	Perda de apetite	1	1	0,72	0,22	0,06	0,00
19	Perda de peso	0	1	0,89	0,11	0,00	0,00
20	Preocupações	1	1	0,68	0,28	0,04	0,00
21	Perda de libido	0	1	0,77	0,16	0,07	0,00

Como podemos ver, as categorias de resposta 1 e 2, que são referentes à baixa

intensidade, são as mais escolhidas. A frequência média destas duas categorias de resposta é de 91%. Em contraste, a categoria de resposta 4 é a menos escolhida, com uma frequência média de 1,2%. Apenas os itens 1 (Tristeza), 8 (Inferioridade) e 17 (Cansaço) possuem maior frequência na categoria de resposta 2. Os demais itens possuem maior frequência de resposta na categoria 1.

O algoritmo de troca é utilizado para buscar a melhor matriz  $\mathbf{Q}$  para o conjunto de dados do BDI de acordo com cada um dos cinco critérios propostos na Seção 5.3. Assim como no estudo de simulação, o modelo CPGM-Q é utilizado. Foram definidas 20 tentativas para o algoritmo de troca.

A Tabela 15 apresenta a matriz  $\mathbf{Q}_L$ , as matrizes  $\mathbf{Q}$  obtidas pelo algoritmo de troca em cada critério e as mesmas medidas apresentadas na Tabela 12 e na Tabela 13 para efeito de comparação entre as matrizes.

A matriz I é a matriz  $\mathbf{Q}$  utilizada na literatura para estes dados de BDI. Ela possui eficiência de apenas 44,88% para o critério  $DISC_{\min}$ . Por outro lado, a eficiência da matriz I para os critérios  $AIC$  e  $BIC$  é de 99,40%, indicando que o modelo CPGM-Q se ajusta aos dados de BDI com a matriz  $\mathbf{Q}_L$  quase tão bem quanto as matrizes V e VI.

A matriz II, encontrada a partir do critério  $DISC_{\text{média}}$ , apresenta a mesma característica das matrizes  $\mathbf{Q}$  encontradas a partir deste critério nos dois cenários do estudo de simulação: uma grande quantidade de zeros. Especificamente, de acordo com esta matriz, nenhum item avalia ambas dimensões do traço latente. Além disso, existem pelo menos outras 548 matrizes  $\mathbf{Q}$  distintas com valor do critério  $DISC_{\text{média}}$  superior ao da matriz  $\mathbf{Q}_L$ .

A matriz III, encontrada a partir do critério  $MDISC_{\text{média}}$ , tem similaridade com a matriz  $\mathbf{Q}_L$  (matriz I) maior do que a matriz II, 64,29% contra 59,52%. Além disso, a matriz III possui eficiência para os critérios  $AIC$  e  $BIC$  superior à eficiência da matriz II, 99,20% contra 98,68% para o critério  $AIC$  e 99,24% contra 98,83% para o critério  $BIC$ . Vemos também que existem pelo menos outras 217 matrizes  $\mathbf{Q}$  distintas com valor do critério  $MDISC_{\text{média}}$  superior ao da matriz  $\mathbf{Q}_L$ .

A matriz IV, encontrada a partir do critério  $DISC_{\min}$ , é a matriz com menor similaridade com a matriz  $\mathbf{Q}_L$  (matriz I), apenas 52,38%. É também a matriz com menor eficiência para os critérios  $AIC$  e  $BIC$ , 96,79% e 96,95%, respectivamente. Além disso, vemos que existem pelo menos outras 148 matrizes  $\mathbf{Q}$  distintas com valor do critério  $DISC_{\min}$  superior ao da matriz  $\mathbf{Q}_L$ .

A matriz V, encontrada a partir do critério  $AIC$ , é a matriz com maior similaridade com a matriz  $\mathbf{Q}_L$  (matriz I), 69,05%. Ela possui eficiência alta para o critério  $BIC$ , 99,95%, e eficiência baixa para o critério  $DISC_{\min}$ , 22,00%. Além disso, existem pelo menos outras 163 matrizes  $\mathbf{Q}$  distintas com valor do critério  $AIC$  superior ao da matriz  $\mathbf{Q}_L$ .

A matriz VI, encontrada a partir do critério  $BIC$ , possui similaridade de 61,90%

Tabela 15 – Matriz  $\mathbf{Q}_L$  utilizada por Fragoso e Cúri (2013) e DA SILVA *et al.* (2018) e matrizes  $\mathbf{Q}$  encontradas a partir dos cinco critérios propostos considerando os dados do BDI

	$\mathbf{Q}_L$		$DISC_{\text{média}}$		$MDISC_{\text{média}}$		$DISC_{\text{mín}}$		$AIC$		$BIC$	
	I		II		III		IV		V		VI	
	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
Item 1	1	1	1	0	1	1	1	1	1	1	1	0
Item 2	1	1	1	0	1	1	1	1	0	1	1	0
Item 3	1	0	1	0	1	0	1	0	1	1	1	1
Item 4	1	1	1	0	1	1	0	1	0	1	1	0
Item 5	1	0	1	0	1	1	1	0	1	1	1	1
Item 6	1	0	1	0	1	1	0	1	1	1	1	1
Item 7	1	0	1	0	1	1	1	0	1	1	1	1
Item 8	1	0	1	0	1	0	1	0	1	1	1	1
Item 9	1	1	1	0	1	0	0	1	1	1	1	0
Item 10	1	1	1	0	1	0	1	0	0	1	1	0
Item 11	0	1	1	0	1	0	1	0	1	1	1	0
Item 12	1	1	1	0	1	0	1	0	1	1	1	1
Item 13	1	1	1	0	1	1	1	0	0	1	1	0
Item 14	1	1	1	0	1	0	0	1	1	1	1	0
Item 15	1	1	1	0	1	0	0	1	1	1	1	0
Item 16	0	1	0	1	0	1	1	0	0	1	1	0
Item 17	0	1	0	1	1	1	1	0	0	1	1	0
Item 18	1	1	1	0	1	0	1	0	0	1	1	0
Item 19	0	1	1	0	0	1	1	0	1	1	1	0
Item 20	1	1	1	0	1	0	0	1	1	1	1	0
Item 21	0	1	1	0	1	0	0	1	1	1	1	1
Posição da $\mathbf{Q}_L$	–		549 <sup>a</sup>		218 <sup>a</sup>		149 <sup>a</sup>		164 <sup>a</sup>		132 <sup>a</sup>	
Similaridade	–		59,52%		64,29%		52,38%		69,05%		61,90%	
Ef. $DISC_{\text{média}}$	73,16%		100,00%		83,41%		82,35%		68,52%		83,43%	
Ef. $MDISC_{\text{média}}$	85,78%		90,97%		100,00%		79,39%		92,11%		91,15%	
Ef. $DISC_{\text{mín}}$	44,88%		67,52%		11,69%		100,00%		22,00%		78,38%	
Ef. $AIC$	99,40%		98,68%		99,20%		96,79%		100,00%		99,95%	
Ef. $BIC$	99,40%		98,83%		99,24%		96,95%		99,95%		100,00%	

com a matriz  $\mathbf{Q}_L$  (matriz I). Ela possui eficiência alta para o critério  $AIC$ , 99,95%. Além disso, existem pelo menos outras 131 matrizes  $\mathbf{Q}$  distintas com valor do critério  $AIC$  superior ao da matriz  $\mathbf{Q}_L$ .

## 5.6 Conclusões e comentários

Neste capítulo, propomos um método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM baseado em um critério projetado para fornecer informação máxima sobre alguma propriedade de interesse, conforme a proposta de validação da matriz  $\mathbf{Q}$  de Torre (2008) para o modelo DINA. Para isso, utilizamos o algoritmo de troca por ponto (NGUYEN; MILLER, 1992), um método eficiente e sistemático para a busca de matrizes, visando

encontrar uma matriz  $\mathbf{Q}$  que maximize alguma propriedade de interesse. Além disso, o método fornece informações de eficiência que podem ser úteis na reavaliação de uma matriz  $\mathbf{Q}$ . Um estudo de simulação foi realizado para analisar o desempenho do método proposto utilizando dois diferentes cenários construídos a partir da variação do número de itens e do número de dimensões do traço latente. Para ilustrar o uso do método de validação da matriz  $\mathbf{Q}$  proposto em dados reais, utilizamos um conjunto de dados com respostas de 1.111 estudantes a uma versão do BDI com 21 itens.

No estudo de simulação, os critérios formulados a partir de medidas de comparação de modelos (*AIC* e *BIC*) tiveram desempenho melhor do que os critérios formulados a partir dos parâmetros de discriminação. No Cenário 1, que é o cenário definido com menos itens e menos dimensões do traço latente, o método recuperou a matriz  $\mathbf{Q}$  utilizada na simulação dos dados utilizando os critérios *AIC* e *BIC*. Já no Cenário 2, o método recuperou a matriz  $\mathbf{Q}$  utilizando o critério *BIC*. Entretanto, o método também apresentou um bom desempenho a partir do critério *AIC*, encontrando uma matriz  $\mathbf{Q}$  em que mais de 95% de seus elementos são similares à matriz  $\mathbf{Q}$  utilizada na simulação dos dados.

Apesar de terem apresentado um desempenho inferior aos critérios *AIC* e *BIC*, os critérios formulados com base na proposta de Torre (2008), a partir dos parâmetros de discriminação, apresentaram um desempenho razoável. Em particular, as matrizes  $\mathbf{Q}$  encontradas pelo método a partir do critério  $DISC_{\text{média}}$  nos dois cenários do estudo de simulação e na aplicação possuem uma quantidade elevada de zeros. Isto é justificável, pois, ao buscar uma matriz  $\mathbf{Q}$  que maximize a média das discriminações, o método tende a eliminar as discriminações com valores menores em cada item, uma vez que estes valores reduzem a média. Já as matrizes  $\mathbf{Q}$  encontradas pelo método a partir dos critérios  $MDISC_{\text{média}}$  e  $DISC_{\text{mín}}$  apresentaram similaridade razoável com a matriz  $\mathbf{Q}$  utilizada na simulação dos dados. Os critérios formulados a partir dos parâmetros de discriminação podem ser utilizados na medição da eficiência de matrizes  $\mathbf{Q}$  encontradas através de outros critérios.

O método proposto neste capítulo é uma ferramenta que fornece informações estatísticas sobre a matriz  $\mathbf{Q}$ . Assim, este método tem por objetivo complementar o conhecimento de um especialista no tema dos itens. Portanto, para uma implementação bem-sucedida da validação da  $\mathbf{Q}$  na prática, é necessário utilizar informações estatísticas em conjunto com o conhecimento de um especialista no tema dos itens.

---

## COMENTÁRIOS FINAIS E DESENVOLVIMENTOS FUTUROS

---

Este capítulo é fornece uma descrição das contribuições desta pesquisa e de possíveis desenvolvimentos futuros.

### 6.1 Comentários finais

Nesta tese, ao longo dos capítulos, exploramos algumas lacunas existentes na literatura acerca dos modelos da TRI para dados politômicos e suas extensões multidimensionais, procurando estudar alguns aspectos de modelagem fundamentais neste contexto. Construimos, a partir do estudo de comparação do ajuste de dois dos modelos unidimensionais mais conhecidos: o modelo de resposta gradual e o modelo de crédito parcial generalizado, uma sequência até chegarmos na proposta de um método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM, passando pela proposta de um modelo de crédito parcial generalizado bifator flexível à funções de ligação e pela incorporação da matriz  $\mathbf{Q}$  pelo modelo de crédito parcial generalizado multidimensional.

Ao investigar os modelos unidimensionais no Capítulo 2, a partir de um estudo de simulação, concluímos que alguns dos critérios de comparação de modelos analisados são adequados e confiáveis para serem utilizados na escolha do modelo mais apropriado aos dados. Com isso, pesquisadores e profissionais podem selecionar o modelo não apenas por uma escolha teórica, mas também utilizando os critérios que tiveram um bom desempenho no estudo.

Contemplando uma estrutura multidimensional, o modelo de crédito parcial generalizado bifator, proposto no Capítulo 3, passa a ser uma alternativa de modelos bifatores para dados de respostas politômicas frente ao modelo de resposta gradual bifator de [Toland \*et al.\* \(2017\)](#) e [Martelli, Matteucci e Mignani \(2016\)](#). Além disso, a literatura passa a

ter formalmente um modelo de crédito parcial generalizado com funções de ligação diferentes da tradicional logística, implementada na proposta do modelo por Muraki (1992). No contexto multidimensional, uma preocupação recorrente é acerca da não identificabilidade do modelo, que ocorre quando mais de um conjunto de parâmetros produz o mesmo valor de probabilidade. De acordo com Béguin e Glas (2001), Fragoso e Cúri (2013) e Reckase (2009), a identificabilidade do modelo é garantida ao especificar uma escala para os traços latentes, com média igual a 0 e o desvio padrão igual a 1. Apesar disso, entendemos que este assunto está em aberto e estudos mais aprofundados envolvendo a identificabilidade de modelos da TRIM propostos neste trabalho devem ser realizados.

Na sequência, incorporamos a matriz  $\mathbf{Q}$ , amplamente utilizada em MDC, nos modelos da TRIM no Capítulo 4. Especificamente, utilizamos o modelo de CPG para demonstrar a proposta. Dentre os benefícios desta alteração, podemos citar a maneira fácil e eficiente de especificar a relação item-traço em um problema prático, a proximidade entre os modelos da TRIM e os MDC e a possível diminuição do número de parâmetros de discriminação do modelo. Além disso, é importante mencionar que, a partir de uma matriz  $\mathbf{Q}$  conveniente, o modelo proposto pode tornar-se o modelo bifator do Capítulo 3 ou o modelo de crédito parcial unidimensional explorado no Capítulo 2.

Por fim, no Capítulo 5, propomos um método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM para evitar problemas de modelagem devido a uma matriz  $\mathbf{Q}$  mal especificada. Particularmente, utilizamos o modelo CPGM-Q, formulado no Capítulo 4, visando analisar o desempenho do método e ilustrar o seu uso na prática.

Esta tese é baseada em quatro manuscritos desenvolvidos durante o período de doutorado. Três manuscritos já foram publicados (DA SILVA; BAZÁN; HUGGINS-MANLEY, 2019; DA SILVA *et al.*, 2018; DA SILVA *et al.*, 2019) e um está sob revisão (DA SILVA *et al.*, 2019). Como esta tese é composta de capítulos independentes, pode ocorrer variação no número de réplicas nos estudos de simulação e variação no número de iterações utilizadas no algoritmo NUTS para o ajuste dos modelos da TRI.

## 6.2 Desenvolvimentos futuros

Existem muitas possibilidades de continuação desta pesquisa. Sugerimos a ampliação do estudo de desempenho dos critérios de comparação de modelos apresentado no Capítulo 2, utilizando outros modelos e suas extensões multidimensionais e outros índices e métodos de verificação de ajuste de modelos. Isto forneceria uma metodologia sólida e confiável para pesquisadores e profissionais na etapa de escolha do modelo estatístico mais adequado ao conjunto de dados.

Outra sugestão seria ampliar a proposta do modelo de crédito parcial generalizado bifator apresentada no Capítulo 3 para um modelo de crédito parcial generalizado aditivo



multidimensional, que considera uma estrutura de correlação entre os traços latentes específicos e o traço latente global, como foi feito por [Martelli, Matteucci e Mignani \(2016\)](#) com o modelo de resposta gradual ou utilizando uma estrutura hierárquica de variáveis latentes.

Além disso, a proposta de incorporar a matriz  $\mathbf{Q}$  em modelos da TRIM apresentada no Capítulo 4 abre um leque de novas possibilidades para novos trabalhos. Primeiro, seria útil investigar os efeitos da especificação da matriz  $\mathbf{Q}$  na estimação dos modelos da TRIM, assim como é feito em [Kunina-Habenicht, Rupp e Wilhelm \(2012\)](#), [Gao \*et al.\* \(2017\)](#) e [Rupp e Templin \(2008\)](#) para os modelos de diagnóstico cognitivo. Segundo, um estudo de sensibilidade de prioris sobre modelos que incorporaram a matriz  $\mathbf{Q}$ , incluindo casos como a não normalidade de traços latentes, como em [Azevedo, Bolfarine e Andrade \(2011\)](#), e assimetria de funções de ligação, como em [Bazán, Branco e Bolfarine \(2006\)](#). Terceiro, como os modelos da TRIM tornaram-se mais próximos dos MDC por conta da incorporação da matriz  $\mathbf{Q}$ , estudos comparativos entre ambas classes de modelos tornaram-se viáveis. Quarto, uma investigação para verificar se os critérios de comparação de modelos podem ser utilizados para verificar dimensionalidade do teste através do ajuste do modelo aos dados com diferentes matrizes  $\mathbf{Q}$ . Por fim, um estudo para verificar se a incorporação da matriz  $\mathbf{Q}$  em modelos da TRIM evita problemas de identificabilidade no modelo, visto que a matriz  $\mathbf{Q}$  inseri restrições na relação item-traço.

A proposta do método de validação de matriz  $\mathbf{Q}$  em modelos da TRIM no Capítulo 5 é um passo inicial na compreensão de como uma matriz  $\mathbf{Q}$  pode ser validada empiricamente. A partir daí, surgem novas possibilidades de continuação desta pesquisa. Entre elas, uma adaptação do algoritmo de troca pode ser implementada, visando deixar alguns elementos da matriz  $\mathbf{Q}$  fixos e, a partir disso, buscar matrizes  $\mathbf{Q}$  utilizando os critérios abordados nesta pesquisa e/ou novos critérios, podendo englobar critérios de modo a buscar matrizes  $\mathbf{Q}$  que otimizem mais de uma propriedade de interesse simultaneamente.

Também sugerimos estudos acerca da análise de resíduos para os modelos propostos nesta tese e o desenvolvimento de um pacote estatístico incorporando os diferentes códigos destes modelos para que sejam disponibilizados aos usuários. Além disso, para o caso de grandes bases de dados, seria útil desenvolver o algoritmo NUTS diretamente em C++ para os modelos desta pesquisa, visando diminuir o tempo computacional de estimação. Finalmente, como os modelos da TRI podem ser vistos como modelos de regressão multivariados, é possível desenvolver modelos de regressão para dados politômicos considerando os resultados apresentados nesta tese.



## REFERÊNCIAS

---

---

ACKERMAN, T. A. Using multidimensional item response theory to understand what items and tests are measuring. **Applied Measurement in Education**, Taylor & Francis, v. 7, n. 4, p. 255–278, 1994. Citado na página 20.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: **Selected Papers of Hirotugu Akaike**. [S.l.]: Springer, 1973. p. 199–213. Citado na página 90.

ALBERT, J. H.; CHIB, S. Bayesian analysis of binary and polychotomous response data. **Journal of the American statistical Association**, Taylor & Francis Group, v. 88, n. 422, p. 669–679, 1993. Citado na página 67.

ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. **ABE, Sao Paulo**, 2000. Citado na página 87.

ATKINSON, A.; DONEV, A.; TOBIAS, R. **Optimum experimental designs, with SAS**. [S.l.]: Oxford University Press, 2007. v. 34. Citado na página 91.

AZEVEDO, C. L.; BOLFARINE, H.; ANDRADE, D. F. Parameter recovery for a skew-normal irt model under a bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 82, n. 11, p. 1679–1699, 2012. Citado na página 79.

AZEVEDO, C. L. N.; BOLFARINE, H.; ANDRADE, D. F. Bayesian inference for a skew-normal irt model under the centred parameterization. **Computacional Statistics and Data Analysis**, v. 55, p. 353–365, 2011. Citado na página 103.

BAZÁN, J. L.; BRANCO, M. D.; BOLFARINE, H. A skew item response model. **Bayesian Analysis**, v. 1, n. 4, p. 861–892, 2006. Citado nas páginas 32, 67 e 103.

BECK, A. T.; WARD, C. H.; MENDELSON, M.; ERBAUGH, J. An inventory for measuring depression. **Archives of General Psychiatry**, v. 4, p. 53–63, 1961. Citado nas páginas 24, 41 e 97.

BÉGUIN, A.; GLAS, C. Mcmc estimation and some model-fit analysis of multidimensional irt models. **Psychometrika**, v. 66, n. 4, p. 541–561, 2001. Citado nas páginas 54, 74 e 102.

BIRNBAUM, A. Some latent trait models and their use in inferring and examinee's ability. In: LORD, F. I.; NOVICK, M. (Ed.). **Statistical theories of mental test scores**. [S.l.]: Reading, MA: Addison -Wesley, 1968. Citado na página 23.

BOCK, R. D.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: application of an em algorithm. **Psychometrika**, v. 46, n. 4, p. 443–459, 1980. Citado na página 20.

BROOKS, S. P. Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde, A. **Journal Royal Statistical Society, Series B**, v. 64, n. 4, p. 616–618, 2002. Citado na página 32.

BURNHAM, K.; ANDERSON, D. **Model selection and multimodel inference: a practical information-theoretic approach**. 2nd. ed. New York: Springer, 2002. Citado nas páginas 34, 39 e 60.

CARLIN, B. P.; LOUIS, T. A. **Bayes and Empirical Bayes Methods for Data Analysis**. second edition. [S.l.]: Chapman & Hall/CRC, Boca Raton, 2001. 440 p. Citado na página 32.

CARPENTER, B.; GELMAN, A.; HOFFMAN, M.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M.; GUO, J.; LI, P.; RIDDELL, A. Stan: A probabilistic programming language. **Journal of Statistical Software**, v. 76, n. 1, p. 1–32, 2017. Citado nas páginas 20, 24, 32 e 56.

CASTRO, S. M. J.; TRENTINI, C.; RIBOLDI, J. Item response theory applied to the Beck Depression Inventory. **Revista Brasileira de Epidemiologia**, v. 13, n. 3, p. 1–13, 2010. Citado na página 24.

CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1–29, 2012. Citado nas páginas 20 e 90.

CHEN, F. F.; WEST, S. G.; SOUSA, K. H. A comparison of bifactor and second-order models of quality of life. **Multivariate Behavioral Research**, v. 41, n. 2, p. 189–225, 2006. Citado na página 50.

CHEN, Y.; LIU, J.; XU, G.; YING, Z. Statistical analysis of q-matrix based diagnostic classification models. **Journal of the American Statistical Association**, Taylor & Francis, v. 110, n. 510, p. 850–866, 2015. Citado nas páginas 83 e 85.

COHEN, J. The cost of dichotomization. **Applied Psychological Measurement**, v. 7, n. 3, p. 249–253, 1983. Citado na página 19.

CURTIS, S. M. Bugs code for item response theory. **Journal of Statistical Software**, v. 36, Code Snippet 1, p. 1–34, 2010. Citado nas páginas 26, 31, 46, 47, 54 e 74.

DA SILVA, M. A.; BAZÁN, J. L.; HUGGINS-MANLEY, A. C. Sensitivity analysis and choosing between alternative polytomous IRT models using Bayesian model comparison criteria. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 48, n. 2, p. 601–620, 2019. Citado na página 102.

DA SILVA, M. A.; DE OLIVEIRA, E. S. B.; VON DAVIER, A. A.; BAZÁN, J. L. Estimating the DINA model parameters using the No-U-Turn Sampler. **Biometrical Journal**, v. 60, n. 2, p. 352–368, 2018. Citado nas páginas 15, 55, 75, 97 e 99.

DA SILVA, M. A.; HUGGINS-MANLEY, A. C.; MAZZON, J. A.; BAZÁN, J. L. A bifactor generalized partial credit model with flexible link functions: a novel approach in survey data. **Journal of Applied Statistics**, Taylor & Francis, 2019. Citado na página 102.

DA SILVA, M. A.; LIU, R.; HUGGINS-MANLEY, A. C.; BAZÁN, J. L. Incorporating the Q-matrix into multidimensional item response theory models. **Educational and Psychological Measurement**, SAGE Publications Sage CA: Los Angeles, CA, 2018. Citado na página 102.

\_\_\_\_\_. Estimating multidimensional polytomous item response theory models with q-matrices, with applications in stan. **Submetido**, 2019. Citado na página 102.

EMBRETSON, S. E.; REISE, S. P. **Item response theory for psychologists**. Maheah. [S.l.]: New Jersey: Lawrence Erlbaum Associates, Publishers, 2000. Citado na página 20.

FEDOROV, V. V. Theory of optimal experiments. **Academic Press**, 1972. Citado na página 91.

FERREIRA, E. V. **Modelos da Teoria de Resposta ao Item assimétricos de grupos múltiplos para respostas politômicas nominais e ordinais sob um enfoque bayesiano**. Dissertação (Mestrado) — Universidade Estadual de Campinas, The address of the publisher, 2 2014. Citado na página 29.

FLIEGE, H.; BECKER, J.; WALTER, O. B.; ROSE, M.; BJORNER, J. B.; KLAPP, B. F. Evaluation of a computer-adaptive test for the assessment of depression (d-cat) in clinical application. **International Journal of Methods in Psychiatric Research**, v. 1, n. 18, p. 23–36, 2009. Citado na página 24.

FOX, J.-P. **Bayesian item response modeling**. [S.l.]: Springer: New York, 2010. 313 p. Citado nas páginas 21, 23, 26 e 34.

FRAGOSO, T. M.; CÚRI, M. Improving psychometric assessment of the beck depression inventory using multidimensional item response theory. **Biometrical Journal**, v. 55, p. 527–540, 2013. Citado nas páginas 15, 32, 41, 54, 73, 74, 97, 99 e 102.

GAO, M.; MILLER, D.; ; LIU, R. The impact of Q-matrix misspecification and model misuse on classification accuracy in the generalized dina model. **Journal of Measurement and Evaluation in Education and Psychology**, v. 8, n. 4, p. 391–403, 2017. Citado na página 103.

GELFAND, A.; SMITH, A. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, p. 398–409, 1990. Citado nas páginas 20, 55, 75 e 115.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; DUNSON, D. B.; VEHTARI, A.; RUBIN, D. B. **Bayesian Data Analysis**. 3rd. ed. New York: Chapman & Hall, 2014. 639 p. Citado nas páginas 24, 60 e 77.

GELMAN, A.; RUBIN, D. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, n. 4, p. 457–472, 1992. Citado nas páginas 37, 56, 60 e 77.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 6, n. 6, p. 721–741, 1984. Citado nas páginas 20, 55, 75 e 115.

- GIBBONS, R. D.; BOCK, R. D.; HEDEKER, D.; WEISS, D. J.; SEGAWA, E.; BHAI-MIK, D. K.; KUPFER, D. J.; FRANK, E.; GROCHOCINSKI, V. J.; STOVER, A. Full-information item bifactor analysis of graded response data. **Applied Psychological Measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 31, n. 1, p. 4–19, 2007. Citado na página 50.
- GIBBONS, R. D.; RUSH, A. J.; IMMEKUS, J. C. On the psychometric validity of the domains of the pdsq: An illustration of the bi-factor item response theory model. **Journal of psychiatric research**, Elsevier, v. 43, n. 4, p. 401–410, 2009. Citado na página 50.
- GIROLAMI, M.; CALDERHEAD, B. Riemann manifold langevin and hamiltonian monte carlo methods. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 73, n. 2, p. 123–214, 2011. Citado nas páginas 55, 75 e 115.
- GOLDBERG, L. The structure of phenotypic personality traits. **American Psychologist**, v. 48, p. 26–34, 1993. Citado na página 69.
- GRANT, R. L.; FURR, D. C.; CARPENTER, B.; GELMAN, A. Fitting bayesian item response models in stata and stan. **Stata Journal**, Stata Press, v. 17, n. 2, p. 343–357(15), 2017. Citado nas páginas 32 e 75.
- HABERMAN, S.; SINHARAY, S. Reporting of subscores using multidimensional item response theory. **Psychometrika**, v. 75, n. 2, p. 209–227, 2010. Citado na página 69.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, v. 57, p. 97–109, 1970. Citado nas páginas 20, 55, 75 e 115.
- HOFFMAN, M. D.; GELMAN, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. **Journal of Machine Learning Research**, v. 15, p. 1593–1623, 2014. Citado nas páginas 24, 50, 55, 71, 75 e 117.
- IMMEKUS, J. C.; IMBRIE, P. Dimensionality assessment using the full-information item bifactor analysis for graded response data: An illustration with the state metacognitive inventory. **Educational and Psychological Measurement**, Sage Publications Sage CA: Los Angeles, CA, v. 68, n. 4, p. 695–709, 2008. Citado na página 50.
- JOHNSON, V. E.; ALBERT, J. H. **Ordinal data modeling**. New York: Springer-Verlag, 1999. 258 p. Citado na página 26.
- JUNKER, B. W.; SIJTSMA, K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. **Applied Psychological Measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 25, n. 3, p. 258–272, 2001. Citado na página 87.
- KIEFTENBELD, V.; NATESAN, P. Recovery of graded response model parameters: a comparison of marginal maximum likelihood and markov chain monte carlo estimation. **Applied Psychological Measurement**, v. 36, n. 5, p. 399–419, 2012. Citado nas páginas 31, 36 e 39.
- Kunina-Habenicht, O.; RUPP, A.; WILHELM, O. The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. **Journal of Educational Measurement**, v. 49, n. 1, p. 59–81, 2012. Citado na página 103.

LI, F.; COHEN, A.; KIM, S.; CHO, S. Model selection methods for mixture dichotomous irt models. **Applied Psychological Measurement**, v. 33, n. 5, p. 353–373, 2009. Citado na página 32.

LI, Y.; BASER, R. Using r and winbugs to fit a generalized partial credit model for developing and evaluating patient-reported outcomes assessments. **Statistics in Medicine**, v. 31, p. 2010–2026, 2012. Citado nas páginas 31, 46 e 47.

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, v. 140, p. 1–55, 1932. Citado na página 19.

LIU, R.; HUGGINS-MANLEY, A. C.; BRADSHAW, L. The impact of q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. **Educational and psychological measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 77, n. 2, p. 220–240, 2017. Citado nas páginas 83 e 85.

LORD, F. A theory of test scores. **Psychometric Monograph**, v. 7, 1952. Citado na página 23.

LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. Winbugs - a bayesian modelling framework: Concepts, structure, and extensibility. **Statistics and Computing**, v. 10, n. 4, p. 325–337, 2000. Citado na página 20.

\_\_\_\_\_. Winbugs - a bayesian modelling framework: concepts, structure, and extensibility. **Statistics and Computing**, v. 10, n. 4, p. 325–337, 2000. Citado na página 75.

MARTELLI, I.; MATTEUCCI, M.; MIGNANI, S. Bayesian estimation of a multidimensional additive graded response model for correlated traits. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 45, n. 5, p. 1636–1654, 2016. Citado nas páginas 49, 50, 54, 67, 74, 101 e 103.

MASTERS, G. A rasch model for partial credit scoring. **Psychometrika**, v. 47, n. 1, p. 149–174, 1982. Citado na página 23.

MAYDEU-OLIVARES, A. Multidimensional item response theory modeling of binary data: Large sample properties of noharm estimates. **Journal of Educational and Behavioral Statistics**, v. 26, n. 1, p. 51–71, 2001. Citado na página 69.

MAYDEU-OLIVARES, A.; DRASGOW, F.; MEAD, A. Distinguishing among parameter item response models for polychotomous ordered data. **Applied Psychological Measurement**, v. 18, n. 3, p. 245–256, 1994. Citado na página 24.

METROPOLIS, N.; ROSENBLUTH, A.; ROSENBLUTH, M.; TELLER, A.; TELLER, E. Equation of state calculations by fast computing machines. **The journal of chemical physics**, v. 21, p. 1087–1092, 1953. Citado nas páginas 20, 55, 75 e 115.

MURAKI, E. Fitting a polytomous item response model to likert-type data. **Applied Psychological Measurement**, v. 14, n. 1, p. 59–71, 1990. Citado na página 20.

\_\_\_\_\_. A generalized partial credit model: application of an em algorithm. **Applied Psychological Measurement**, v. 16, n. 2, p. 159–176, 1992. Citado nas páginas 20, 23, 27, 49, 50, 51, 67, 73 e 102.

- MURAKI, E.; CARLSON, J. E. Full-information factor analysis for polytomous item responses. **Applied Psychological Measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 19, n. 1, p. 73–90, 1995. Citado na página 49.
- NEAL, R. Mcmc using hamiltonian dynamics. In: BROOKS A. GELMAN, G. L. J. I. S.; MENG, X.-L. (Ed.). **Handbook of Markov chain Monte Carlo**. [S.l.]: Chapman and Hall/CRC, 2011. cap. 5, p. 116–162. Citado nas páginas 29, 75 e 117.
- NERING, M.; OSTINI, R. **Handbook of polytomous item response theory models**. New York: Routledge, 2010. Citado na página 19.
- NGUYEN, N.-K.; MILLER, A. J. A review of some exchange algorithms for constructing discrete d-optimal designs. **Computational Statistics & Data Analysis**, Elsevier, v. 14, n. 4, p. 489–498, 1992. Citado nas páginas 91 e 99.
- OSTINI, R.; NERING, M. L. **Polytomous item response theory models: quantitative applications in the social sciences**. Thousand Oaks, CA: SAGE, 2006. 107 p. Citado nas páginas 20, 42 e 44.
- PLUMMER, M. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: **DSC 2003 Working Papers**. [S.l.: s.n.], 2003. p. 1–8. Citado nas páginas 20 e 75.
- PÜSCHEL, J.; MAZZON, J. A.; HERNANDEZ, J. M. C. Mobile banking: proposition of an integrated adoption intention framework. **International Journal of bank marketing**, Emerald Group Publishing Limited, v. 28, n. 5, p. 389–409, 2010. Citado na página 58.
- RASCH, G. **Probabilistic models for some intelligence and attainment tests**. [S.l.]: The University of Chicago Press: Chicago, 1980. Citado na página 23.
- RECKASE, M. **Multidimensional Item Response Theory**. 1. ed. New York: Springer-Verlag, 2009. Citado nas páginas 20, 54, 71, 73, 74 e 102.
- REISE, S. P. The rediscovery of bifactor measurement models. **Multivariate behavioral research**, Taylor & Francis, v. 47, n. 5, p. 667–696, 2012. Citado na página 50.
- REISE, S. P.; MORIZOT, J.; HAYS, R. D. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. **Quality of Life Research**, Springer, v. 16, n. 1, p. 19–31, 2007. Citado na página 50.
- RIJMEN, F. Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional irt model. **Journal of Educational Measurement**, Wiley Online Library, v. 47, n. 3, p. 361–372, 2010. Citado na página 50.
- RIZOPOULOS, D. ltm: An r package for latent variable modelling and item response theory analyses. **Journal of Statistical Software**, v. 17, n. 5, p. 1–25, 2006. Disponível em: <<http://www.jstatsoft.org/v17/i05/>>. Citado na página 20.
- ROBITZSCH, A.; KIEFER, T.; GEORGE, A. C.; UENLUE, A.; ROBITZSCH, M. A. Package ‘cdm’. 2018. Citado na página 76.



- RUBIN, D. B. Bayesianly justifiable and relevant frequency calculations for the applied statistician. **The Annals of Statistics**, JSTOR, p. 1151–1172, 1984. Citado nas páginas 60 e 77.
- RUPP, A.; TEMPLIN, J. The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. **Educational and Psychological Measurement**, v. 68, n. 1, p. 78–96, 2008. Citado na página 103.
- RUPP, A.; TEMPLIN, J.; HENSON, R. **Diagnostic measurement: Theory, methods, and applications**. [S.l.]: Guilford Press, 2010. Citado nas páginas 20 e 70.
- SAHU, S. Bayesian estimation and model choice in item response models. **Journal of Statistical Computation and Simulation**, v. 72, n. 3, p. 217–232, 2002. Citado nas páginas 36, 45, 61 e 77.
- SAMEJIMA, F. Estimation of latent ability using a response pattern of graded scores. **Psychometrika Monograph**, n. 17, 1969. Citado nas páginas 20, 23, 25 e 49.
- \_\_\_\_\_. Normal ogive model on the continuous response level in the multidimensional latent space. **Psychometrika**, Springer, v. 39, n. 1, p. 111–121, 1974. Citado na página 49.
- \_\_\_\_\_. Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. **Psychometrika**, Springer, v. 65, n. 3, p. 319–335, 2000. Citado na página 67.
- SANTOS, J. R.; AZEVEDO, C. L.; BOLFARINE, H. A multiple group item response theory model with centered skew-normal latent trait distributions under a Bayesian framework. **Journal of Applied Statistics**, Taylor & Francis, v. 40, n. 10, p. 2129–2149, 2013. Citado nas páginas 42 e 62.
- SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citado na página 90.
- SHENG, Y.; WIKLE, C. Comparing multiunidimensional and unidimensional item response theory models. **Educational and Psychological Measurement**, v. 67, n. 6, p. 899–919, 2007. Citado na página 49.
- \_\_\_\_\_. Bayesian multidimensional IRT models with a hierarchical structure. **Educational and Psychological Measurement**, v. 68, n. 3, p. 413–430, 2008. Citado na página 49.
- \_\_\_\_\_. Bayesian IRT models incorporating general and specific abilities. **Behaviormetrica**, v. 36, n. 1, p. 27–48, 2009. Citado na página 49.
- SINHARAY, S.; JOHNSON, M.; STERN, H. Posterior predictive assessment of item response theory models. **Applied Psychological Measurement**, v. 30, n. 4, p. 298–321, 2006. Citado nas páginas 23, 34 e 36.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDEN, A. van der. Bayesian measures of model complexity and fit. **Journal Royal Statistical Society, Series B**, v. 64, n. 4, p. 583–639, 2002. Citado nas páginas 32, 34, 39, 60 e 77.
- STAN DEVELOPMENT TEAM. **RStan: the R interface to Stan**. 2017. <<http://mc-stan.org>>. Citado na página 56.

\_\_\_\_\_. **Stan Modeling Language Users Guide and Reference Manual, Version 2.17.0**. 2017. <<http://mc-stan.org>>. Citado na página 75.

STERN, H.; SINHARAY, S. Bayesian model checking and model diagnostics. In: DEY DIPAK K., R. C. E. B. M. T. I.; STATISTICS, C. I. H. of (Ed.). **Bayesian Modelling, Thinking and Computation**. The Netherlands: Elsevier, 2005. Citado nas páginas 23 e 34.

STONE, C. A.; ZHU, X. **Bayesian analysis of item response theory models using SAS®**. NC: SAS Institute: Cary, 2015. 282 p. Citado na página 31.

SYMPSON, J. B. A model for testing with multidimensional items. In: UNIVERSITY OF MINNESOTA, DEPARTMENT OF PSYCHOLOGY, PSYCHOMETRIC METHODS PROGRAM MINNEAPOLIS. **Proceedings of the 1977 computerized adaptive testing conference**. [S.l.], 1978. Citado na página 49.

TATSUOKA, K. Rule space: An approach for dealing with misconceptions based on item response theory. **Journal of Educational Measurement**, v. 20, n. 4, p. 345–354, 1983. Citado nas páginas 21 e 70.

THISSEN, D. Bad questions: An essay involving item response theory. **Journal of Educational and Behavioral Statistics**, v. 41, n. 1, p. 81–89, 2016. Citado na página 69.

THISSEN, D.; STEINBERG, L. A taxonomy of item response models. **Psychometrika**, v. 51, p. 567–577, 1986. Citado na página 23.

THOMAS, A. **OpenBUGS: Constructing Mcmc Software**. [S.l.]: John Wiley & Sons Incorporated, 2007. (Wiley Series in Computational Statistics Series). Citado na página 20.

TOLAND, M. D.; SULIS, I.; GIAMBONA, F.; PORCU, M.; CAMPBELL, J. M. Introduction to bifactor polytomous item response theory analysis. **Journal of School Psychology**, v. 60, p. 41 – 63, 2017. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022440516300693>>. Citado nas páginas 49, 50, 64, 67, 80 e 101.

TORRE, J. de la. An empirically based method of Q-matrix validation for the dina model: Development and applications. **Journal of educational measurement**, Wiley Online Library, v. 45, n. 4, p. 343–362, 2008. Citado nas páginas 85, 86, 88, 90, 99 e 100.

TORRE, J. de la; CHIU, C.-Y. General method of empirical Q-matrix validation. **Psychometrika**, Springer, v. 81, n. 2, p. 253–273, 2016. Citado na página 85.

VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. **Statistics and Computing**, v. 27, n. 5, p. 1413–1432, 2017. Citado nas páginas 33, 34, 60 e 77.

WATANABE, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. **Journal of Machine Learning Research**, v. 11, p. 3571–3594, 2010. Citado nas páginas 33, 60 e 77.

WHITELEY, S. E. Multicomponent latent trait models for ability tests. **Psychometrika**, Springer, v. 45, n. 4, p. 479–494, 1980. Citado na página 49.

YAO, L.; SCHWARZ, R. D. A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. **Applied psychological measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 30, n. 6, p. 469–492, 2006. Citado na página 49.

\_\_\_\_\_. A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. **Applied Psychological Measurement**, v. 30, n. 6, p. 469–492, 2006. Citado na página 70.

ZHU, L.; ROBINSON, S. E.; TORENVLIED, R. A bayesian approach to measurement bias in networking studies. **American Review of Public Administration**, p. 1–23, 2014. Citado nas páginas 31 e 46.

ZHU, X.; STONE, C. Bayesian comparison of alternative graded response models for performance assessment applications. **Educational and Psychological Measurement**, v. 72, n. 5, p. 774–799, 2012. Citado na página 31.



---

# O MÉTODO MONTE CARLO HAMILTONIANO E O ALGORITMO NO-U-TURN SAMPLER

---

---

Alguns métodos MCMC amplamente utilizados nas últimas décadas para estimar parâmetros de modelos estatísticos, tais como Metropolis Hastings ([METROPOLIS \*et al.\*, 1953](#); [HASTINGS, 1970](#)) e Gibbs sampling ([GEMAN; GEMAN, 1984](#); [GELFAND; SMITH, 1990](#)), tendem a explorar o espaço paramétrico via passos aleatórios ineficientes, aumentando a autocorrelação dos valores gerados e o tempo necessário para esses métodos convergirem para a distribuição alvo. Desta forma, estes métodos estão sendo substituídos por métodos alternativos que mostram-se mais eficientes. É o caso do método Monte Carlo Hamiltoniano (HMC), que tem sido amplamente explorado nos últimos anos e ganhado espaço dentre os pesquisadores. O método HMC utiliza a dinâmica Hamiltoniana ao invés de uma distribuição de probabilidade para construir uma cadeia de Markov, tornando a convergência da simulação mais rápida e com menor autocorrelação ([GIROLAMI; CALDERHEAD, 2011](#)).

A dinâmica Hamiltoniana pode ser utilizada para descrever o movimento de um objeto através de sua localização, dada por um vetor  $\mathbf{x}$  e *momentum* (velocidade de massa), dado por um vetor  $\mathbf{p}$ , ambos em algum instante  $t$ . As dimensões desses vetores correspondem à dimensão considerada do problema, que estamos denotando por  $d$ . Associado a cada localização do objeto existe uma energia potencial denotada por  $U(\mathbf{x})$  e, associada a cada *momentum* do objeto, existe uma energia cinética denotada por  $K(\mathbf{p})$ . A energia total do sistema, denominada Hamiltoniano, é definida como a soma da energia potencial e energia cinética, ou seja,

$$H(\mathbf{x}, \mathbf{p}) = U(\mathbf{x}) + K(\mathbf{p}).$$

Podemos determinar a localização e o *momentum* de um objeto a qualquer instante  $t$  a

partir do cálculo das derivadas parciais do Hamiltoniano

$$\frac{\partial x_i}{\partial t} = \frac{\partial H}{\partial p_i} = \frac{\partial K(\mathbf{p})}{\partial p_i}, \quad (\text{A.1})$$

$$\frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial x_i} = -\frac{\partial U(\mathbf{x})}{\partial x_i}, \quad (\text{A.2})$$

para  $i = 1, 2, \dots, d$ .

A implementação computacional da dinâmica Hamiltoniana é feita pela aproximação das Equações A.1 e A.2 através da discretização do tempo, dividindo-o em pequenas fatias de comprimento  $\delta$ . Essa aproximação melhora à medida em que  $\delta$  diminui. A abordagem utilizada no método Monte Carlo Hamiltoniano para realizar a discretização do tempo é conhecida como o método *Leap Frog*. Em suma, o método *Leap Frog* consiste nos três passos descritos a seguir.

1. Comece atualizando as variáveis de *momentum* durante um pequeno intervalo de tempo  $\delta/2$

$$p_i(t + \delta/2) = p_i(t) - (\delta/2) \frac{\partial U}{\partial x_i(t)}.$$

2. Atualize as variáveis de localização durante o intervalo de tempo  $\delta$  usando os novos valores das variáveis de *momentum*.

$$x_i(t + \delta) = x_i(t) + \delta \frac{\partial K}{\partial p_i(t + \delta/2)}.$$

3. Complete a atualização das variáveis de *momentum* sobre o outro pequeno intervalo de tempo  $\delta/2$

$$p_i(t + \delta) = p_i(t + \delta/2) - (\delta/2) \frac{\partial U}{\partial x_i(t + \delta)}.$$

A ideia do Monte Carlo Hamiltoniano é usar a dinâmica Hamiltoniana para amostrar valores de uma distribuição alvo  $p(\mathbf{x})$  através do conceito de uma distribuição canônica da mecânica estatística. Definimos a distribuição canônica para qualquer função de energia  $E(\boldsymbol{\theta})$  sobre um conjunto de variáveis  $\boldsymbol{\theta}$  como

$$p(\boldsymbol{\theta}) = \frac{1}{Z} e^{-E(\boldsymbol{\theta})},$$

em que  $Z$  é a constante de normalização. Como distribuições de probabilidade não escalonadas podem ser amostradas via métodos MCMC e como sabemos que a função de energia para a dinâmica Hamiltoniana é uma combinação de energias potencial e cinética, isto é,  $E(\boldsymbol{\theta}) = H(\mathbf{x}, \mathbf{p}) = U(\mathbf{x}) + K(\mathbf{p})$ , então podemos escrever a distribuição canônica da função de energia para a dinâmica Hamiltoniana da seguinte forma

$$\begin{aligned} p(\mathbf{x}, \mathbf{p}) &\propto e^{-H(\mathbf{x}, \mathbf{p})} \\ &= e^{-[U(\mathbf{x}) + K(\mathbf{p})]} \\ &= e^{-U(\mathbf{x})} \cdot e^{-K(\mathbf{p})} \\ &\propto p(\mathbf{x}) \cdot p(\mathbf{p}). \end{aligned}$$

Note que  $\mathbf{x}$  e  $\mathbf{p}$  são independentes. A função de energia potencial é definida como

$$U(\mathbf{x}) = -\log[\pi(\mathbf{x}) \cdot \ell(\mathbf{x}|D)],$$

em que  $\pi(\mathbf{x})$  é a densidade a priori e  $\ell(\mathbf{x}|D)$  é a função de verossimilhança dado os dados  $D$ . A função de energia cinética é definida como

$$K(\mathbf{p}) = \sum_{i=1}^d \frac{p_i^2}{2}.$$

Como  $\mathbf{x}$  e  $\mathbf{p}$  são independentes, podemos escolher qualquer distribuição para as variáveis auxiliares  $\mathbf{p}$ . No entanto, a opção mais comum é a distribuição normal padrão.

O HMC é iniciado a partir da simulação de valores  $\mathbf{x}^{(0)}$  para as variáveis de interesse de sua distribuição a priori. Então, já dentro de uma estrutura de repetição, os valores atuais  $\mathbf{p}_0$  são obtidos para as variáveis auxiliares da distribuição normal padrão. Para cada iteração  $t$  são consideradas como variáveis de interesse os valores atuais  $\mathbf{x}_0 = \mathbf{x}^{(t-1)}$ , em que  $\mathbf{x}^{(t-1)}$  é encontrado na iteração  $(t-1)$ . Para a iteração  $t = 1$ , o valor atual considerado é o valor inicialmente simulado  $\mathbf{x}^{(0)}$ , ou seja,  $\mathbf{x}_0 = \mathbf{x}^{(0)}$ . Em seguida, o método *Leap frog* é utilizado com os valores  $\mathbf{x}_0$  e  $\mathbf{p}_0$  para  $L$  passos e tamanho do passo  $\delta$  para obter os valores propostos  $\mathbf{x}^*$  e  $\mathbf{p}^*$ . Com estes valores propostos, calcula-se a probabilidade de aceitação, um procedimento análogo ao índice de aceitação no algoritmo Metropolis-Hastings, dado por

$$\alpha = \min(1, \exp(-U(\mathbf{x}^*) + U(\mathbf{x}_0) - K(\mathbf{p}^*) + K(\mathbf{p}_0))).$$

Um número aleatório  $u$  é simulado a partir da distribuição uniforme com limitantes 0 e 1, ou seja,  $\text{Uniforme}(0, 1)$ . Se  $u \leq \alpha$ , então  $\mathbf{x}^*$  é aceito e define-se o próximo estado da cadeia de Markov  $\mathbf{x}^{(t)} = \mathbf{x}^*$ . Caso contrário,  $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$ . Então, retorna-se para o início da estrutura de repetição  $T$  vezes.

O tamanho do passo  $\delta$  e o número de passos  $L$  são parâmetros do HMC definidos pelo usuário. Definir estes parâmetros é uma tarefa difícil de ser realizada, porque ela exige conhecimento especializado do método e do contexto. A escolha errada destes parâmetros pode causar problemas, tais como: (1) poder computacional desperdiçado (tamanho de passo excessivamente pequeno); (2) simulação imprecisa e alta taxa de rejeição (tamanho de passo excessivamente grande); (3) comportamento de passeio aleatório lento (número de passos excessivamente pequeno); e (4) trajetória refazer seus passos (número de passos excessivamente grande). Maiores detalhes do método HMC podem ser encontrados em Neal (2011).

O algoritmo No-U-turn Sampler (NUTS), descrito por Hoffman e Gelman (2014), é uma extensão do HMC, proposto com o objetivo de simular valores aleatórios de uma distribuição alvo eliminando a necessidade de definir o número de passos  $L$  a serem percorridos em cada iteração e o tamanho  $\delta$  de cada passo. Em resumo, enquanto o método

HMC executa o algoritmo *Leap Frog*  $L$  vezes e seleciona o último estado gerado com um procedimento análogo à razão de aceitação do algoritmo Metropolis-Hastings, o algoritmo NUTS constrói uma árvore de nós e seleciona um destes nós para o próximo estado na cadeia. Para isso, é introduzida uma variável de corte  $u$  com distribuição condicional  $u|\mathbf{x}^{(t-1)}, \mathbf{p}_0 \sim \text{Uniforme}\left(0, \exp\{-U(\mathbf{x}^{(t-1)}) - K(\mathbf{p}_0)\}\right)$ . Cada árvore de nós é construída a partir de subárvores até que o critério de parada

$$(\tilde{\mathbf{x}} - \mathbf{x}) \cdot \tilde{\mathbf{p}} < \mathbf{0} \quad (\text{A.3})$$

seja satisfeito, em que  $\tilde{\mathbf{x}}$  é o vetor de interesse atual,  $\mathbf{x}$  é o vetor de interesse anterior e  $\tilde{\mathbf{p}}$  é o vetor auxiliar atual. Este critério simula os nós até que o vetor atual  $\tilde{\mathbf{x}}$  comece a retroceder em direção ao vetor anterior  $\mathbf{x}$ . Por isso o nome “No-U-Turn”. Para cada subárvore  $j$  ( $j = 0, 1, 2, \dots$ ), simula-se aleatoriamente a direção  $\mathbf{v}_j$  do caminho a ser percorrido para construir  $2^j$  nós, tal que  $\mathbf{v}_j \sim \text{Uniforme}(\{-1, +1\})$ .

O  $k$ -ésimo nó  $(\mathbf{x}_{(k)}, \mathbf{p}_{(k)})$ , para  $k = 1, 2, \dots, 2^j$ , da  $j$ -ésima subárvore é construído a partir do algoritmo *Leap Frog* e é selecionado com probabilidade

$$\frac{\mathbb{1}(u \leq \exp\{-U(\mathbf{x}_{(k)}) - K(\mathbf{p}_{(k)})\})}{n'}$$

em que

$$n' = \sum_{h=1}^k \mathbb{1}(u \leq \exp\{-U(\mathbf{x}_{(h)}) - K(\mathbf{p}_{(h)})\}),$$

ou seja,  $n'$  é o número de nós construídos na  $j$ -ésima subárvore até o momento, tal que

$$\mathbb{1}(u \leq \exp\{-U(\mathbf{x}_{(h)}) - K(\mathbf{p}_{(h)})\}) = 1$$

para  $h = 1, 2, \dots, k$ .

Ao término do procedimento de construção da  $J$ -ésima subárvore, um nó estará escolhido para substituir o candidato a novo estado da cadeia. Esta substituição ocorre com probabilidade  $\frac{n'}{n}$ , em que

$$n' = \sum_{h=1}^{2^j} \mathbb{1}(u \leq \exp\{-U(\mathbf{x}_{(h)}) - K(\mathbf{p}_{(h)})\}) \quad \text{e} \quad n = \sum_{j=0}^J \sum_{h=1}^{2^j} \mathbb{1}(u \leq \exp\{-U(\mathbf{x}_{(h)}) - K(\mathbf{p}_{(h)})\}).$$

Em outras palavras, uma árvore é construída até o critério de parada descrito em (A.3) ser satisfeito, ou seja, até que uma das extremidades da trajetória atinja a outra extremidade da árvore. Assim, o vetor atual  $\tilde{\mathbf{x}}$  é o novo estado da cadeia.



---

## CÓDIGO R E STAN

---

Neste apêndice, apresentamos um exemplo do código R para utilizar o software Stan. Além disso, apresentamos o código Stan dos modelos de resposta gradual e de crédito parcial generalizado apresentados no Capítulo 2, do modelo de crédito parcial generalizado bifator apresentado no Capítulo 3 e do modelo de crédito parcial generalizado multidimensional com matriz  $\mathbf{Q}$  apresentado no Capítulo 4.

---

```

1 library(rstan)                                #chama o pacote Stan previamente instalado
2 dados <- read.csv("dados.csv")                #lê os dados contidos em um arquivo csv
3 Y <- as.matrix(dados)                         #transforma os dados em uma matriz
4 N <- nrow(Y)                                  #número de indivíduos
5 J <- ncol(Y)                                  #número de itens
6 m <- rep(4,J)                                 #vetor com o número de categorias de resposta
7 parametros <- c("a","b","theta")             #define os parâmetros que serão retornados
8
9 ##### A função Stan #####
10 #define o arquivo com o código Stan do modelo (file)
11 #define os parâmetros do modelo que serão retornados (pars)
12 #define o número de cadeias (chains)
13 #define o número de thin
14 #define o número de iterações (iter)
15 #define o número de burn-in (warmup)
16 stanfit <- stan(file = 'modelo.stan',
17                 pars = parametros,
18                 chains = 1,
19                 thin = 1,
20                 iter = 10000,
21                 warmup = 5000)
22 stanfit                                         #apresenta o retorno do Stan

```

---

Figura 26 – Exemplo de código R utilizado para chamar a linguagem de modelagem Stan através do pacote “RStan”.

---

```

1 data {
2   int<lower=1> N;           //número de indivíduos
3   int<lower=1> J;           //número de itens
4   int<lower=1> m[J];       //número de categorias de respostas
5   int<lower=1, upper=max(m)> Y[N, J]; //matriz de respostas
6 }
7
8 parameters {
9   real<lower=0> a[J];       //vetor de parâmetros de discriminação
10  ordered[max(m)-1] b_star[J]; //vetor de parâmetros de dificuldade
11  real theta[N];           //vetor de parâmetros de indivíduos
12 }
13
14 transformed parameters {
15   simplex[max(m)] prob[N, J];
16   real Pmais[N, J, max(m)];
17   for (i in 1:N) {
18     for (j in 1:J) {
19       for (k in 1:(m[j]-1))
20         Pmais[i, j, k] <- inv_logit(b_star[j, k] - a[j]*theta[i]);
21       Pmais[i, j, m[j]] <- 1;
22     }
23     for (j in 1:J){
24       prob[i, j, 1] <- Pmais[i, j, 1];
25       for (k in 2:m[j])
26         prob[i, j, k] <- Pmais[i, j, k] - Pmais[i, j, k-1];
27     } } }
28
29 model {
30   theta ~ normal(0, 1);
31   for (j in 1:J) {
32     a[j] ~ normal(1,2.5) T[0,];
33     b_star[j] ~ normal(0, 2.5);
34   }
35   for (i in 1:N) {
36     for (j in 1:J)
37       Y[i, j] ~ categorical(prob[i, j, 1:m[j]]);
38 } }

```

---

Figura 27 – Código Stan do modelo de resposta gradual apresentado no Capítulo 2.

---

```

1 data {
2   int<lower=1> N;           //número de indivíduos
3   int<lower=1> J;           //número de itens
4   int<lower=1> m[J];       //número de categorias de respostas
5   int<lower=1, upper=max(m)> Y[N, J]; //matriz de respostas
6 }
7
8 parameters {
9   real<lower=0> a[J];       //vetor de parâmetros de discriminação
10  real b_aux[J, max(m)-1]; //vetor de parâmetros de dificuldade
11  real theta[N];           //vetor de parâmetros de indivíduos
12 }
13
14 transformed parameters {
15   simplex[max(m)] prob[N, J];
16   real eta[N, J, max(m)];
17   vector[max(m)] psum[N, J];
18   vector[max(m)] exp_psum[N, J];
19   real b[J, max(m)];
20   for (j in 1:J) {
21     b[j, 1] <- 0;
22     for (k in 2:m[j])
23       b[j, k] <- b_aux[j, k-1];
24   }
25   for (i in 1:N) {
26     for (j in 1:J) {
27       for (k in 1:m[j]) {
28         eta[i, j, k] <- a[j] * (theta[i] - b[j, k]);
29         psum[i, j, k] <- sum(eta[i, j, 1:k]);
30         exp_psum[i, j, k] <- exp(psum[i, j, k]);
31       } } }
32   for (i in 1:N) {
33     for (j in 1:J)
34       prob[i, j, 1:m[j]] <- softmax(psum[i, j, 1:m[j]]);
35   } }
36
37 model {
38   theta ~ normal(0, 1);
39   for (j in 1:J) {
40     a[j] ~ normal(1, 2.5) T[0,];
41     for (k in 1:(m[j]-1))
42       b_aux[j, k] ~ normal(0, 2.5);
43   }
44   for (i in 1:N) {
45     for (j in 1:J)
46       Y[i, j] ~ categorical(prob[i, j, 1:m[j]]);
47   } }

```

---

Figura 28 – Código Stan do modelo de crédito parcial generalizado apresentado no Capítulo 2.

---

```

1 data {
2   int<lower=1> N; //número de indivíduos
3   int<lower=1> D; //número de subtestes/dimensões
4   int<lower=1> J[D]; //vetor com número de itens em cada dimensão
5   int m[maxJ, D]; //número de categorias de respostas
6   int<lower=1> maxm; //maior elemento do vetor m
7   int<lower=1> maxJ; //maior elemento do vetor J
8   int<lower=1> summ; //soma dos elementos do vetor m
9   int<lower=1> sumJ; //soma dos elementos do vetor J
10  int<lower=1> lenY; //número de observações (elementos de Y)
11  int<lower=1, upper=maxm> Y[lenY]; //matriz de respostas
12  int a_idx[maxJ, D]; //índices dos parâmetros de discriminação
13  int b_idx[maxJ, D, maxm-1]; //índices dos parâmetros de dificuldade
14  int theta0_idx[N]; //índices dos parâmetros globais de indiv
15  int theta_idx[D, N]; //índices dos parâmetros específicos de indiv
16  int Y_idx[N, maxJ, D]; //índices da matriz de respostas e de probs
17 }
18 parameters {
19   real<lower=0> a0[sumJ]; //vetor de parâmetros de discriminação global
20   real<lower=0> a[sumJ]; //vetor de parâmetros de discriminação espec
21   real b[summ-sumJ]; //vetor de parâmetros de dificuldade
22   real theta0[N]; //vetor de parâmetros de indivíduos
23   real theta[N*D]; //vetor de parâmetros de indivíduos
24 }
25 transformed parameters {
26   real eta[N, maxJ, D, maxm];
27   real Cloglog_eta[N, maxJ, D, maxm];
28   real odds[N, maxJ, D, maxm];
29   real prod_odds[N, maxJ, D, maxm];
30   real sum_prod_odds[N, maxJ, D];
31   real b_aux;
32   simplex[maxm] prob[lenY];
33   for (i in 1:N) {
34     for (d in 1:D) {
35       for (j in 1:J[d]) {
36         for (k in 1:m[j, d]) {
37           if (k == 1)
38             b_aux <- 0;
39           else
40             b_aux <- b[b_idx[j, d, k-1]];
41           eta[i, j, d, k] <- a0[a_idx[j, d]]*theta0[theta0_idx[i]] + a[a_idx[j, d]]*
              theta[theta_idx[d, i]] - b_aux;
42           if (eta[i, j, d, k] > 2)
43             eta[i, j, d, k] <- 2;
44           Cloglog_eta[i, j, d, k] <- inv_cloglog(eta[i, j, d, k]);
45           odds[i, j, d, k] <- Cloglog_eta[i, j, d, k]/(1-Cloglog_eta[i, j, d, k]);
46           prod_odds[i, j, d, k] <- prod(odds[i, j, d, 1:k]);
47         }
48         sum_prod_odds[i, j, d] <- sum(prod_odds[i, j, d, 1:m[j, d]]);
49         for (k in 1:m[j, d]) {
50           if (Y_idx[i, j, d] != 0)
51             prob[Y_idx[i, j, d], k] <- prod_odds[i, j, d, k]/sum_prod_odds[i, j, d];
52         } } } }
53 model {
54   for (i in 1:N) {
55     theta0[theta0_idx[i]] ~ normal(0, 1);
56     for (d in 1:D)
57       theta[theta_idx[d, i]] ~ normal(0, 1);
58   }
59   for (d in 1:D) {
60     for (j in 1:J[d]) {
61       a0[a_idx[j, d]] ~ normal(0,1) T[0,];
62       a[a_idx[j, d]] ~ normal(0,1) T[0,];
63       for (k in 1:(m[j, d]-1))
64         b[b_idx[j, d, k]] ~ normal(0, 1);
65     } }
66   for (l in 1:lenY)
67     Y[l] ~ categorical(prob[l, 1:maxm]);
68 }

```

---

Figura 29 – Código Stan do modelo de crédito parcial generalizado bifator com função de ligação clog-log apresentado no Capítulo 3.

---

```

1 data {
2   int<lower=1> N;           //number of individuals
3   int<lower=1> J;           //number of items
4   int<lower=1> m[J];        //number of response categories
5   int<lower=1> D;           //number of dimensions
6   int<lower=0, upper=1> Q[J, D]; //Q-matrix
7   int<lower=1> maxm;        //element of highest value of the vector m
8   int<lower=1> sumQ;        //sum of the elements of the vector m
9   int<lower=1> sumQ;        //sum of the elements of the Q-matrix
10  int<lower=1> lenY;        //number of observations
11  int<lower=1> y[lenY];     //matrix of responses
12  int<lower=1> a_idx[J, D]; //index of discriminations parameters
13  int<lower=1> theta_idx[N, D]; //index of individuals parameters
14  int<lower=1> Y_idx[N, J]; //index of response matrix and probs
15 }
16
17 parameters {
18   real<lower=0> a[sumQ];    //vector of discrimination parameters
19   ordered[maxm-1] b[J];    //vector of difficulty parameters
20   real theta[N*D];        //vector of individual parameters
21 }
22
23 transformed parameters {
24   vector[maxm] eta[N, J];
25   vector[maxm] psum[N, J];
26   simplex[maxm] prob[lenY];
27   real combL;
28   real b_aux;
29   for (l in 1:lenY)
30     prob[l, maxm] <- 0;
31   for (i in 1:N) {
32     for (j in 1:J) {
33       combL <- 0;
34       for (d in 1:D) {
35         if (Q[j, d] == 1)
36           combL <- combL + a[a_idx[j, d]]*theta[theta_idx[i, d]];
37       }
38       for (k in 1:m[j]) {
39         if (k == 1)
40           b_aux <- 0;
41         else
42           b_aux <- b[b_idx[j, k-1]];
43         eta[i, j, k] <- combL - b_aux;
44         psum[i, j, k] <- sum(eta[i, j, 1:k]);
45       }
46       if (Y_idx[i, j] != 0)
47         prob[Y_idx[i, j], 1:m[j]] <- softmax(psum[i, j, 1:m[j]]);
48     } } }
49
50 model {
51   for (i in 1:N) {
52     for (d in 1:D)
53       theta[theta_idx[i, d]] ~ normal(0, 1);
54   }
55   for (j in 1:J) {
56     for (d in 1:D) {
57       if (Q[j, d] == 1)
58         a[a_idx[j, d]] ~ normal(1, 2.5) T[0,];
59     }
60     b[j] ~ normal(0, 2.5);
61   }
62   for (l in 1:lenY)
63     Y[l] ~ categorical(prob[l]);
64 }

```

---

Figura 30 – Código Stan do modelo de crédito parcial generalizado multidimensional com matriz **Q** apresentado no Capítulo 4.