

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Contribuições para modelos gerais de diagnóstico cognitivo

Renato da Silva Fernandes

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Renato da Silva Fernandes

Contribuições para modelos gerais de diagnóstico cognitivo

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

Área de Concentração: Estatística

Orientador: Prof. Dr. Jorge Luis Bazán Guzmán

USP – São Carlos
Maio de 2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

F363c Fernandes, Renato da Silva
 Contribuições para modelos gerais de diagnóstico
 cognitivo / Renato da Silva Fernandes; orientador
 Jorge Luis Bazán Guzmán. -- São Carlos, 2024.
 113 p.

 Tese (Doutorado - Programa Interinstitucional de
 Pós-graduação em Estatística) -- Instituto de Ciências
 Matemáticas e de Computação, Universidade de São
 Paulo, 2024.

 1. modelos de variáveis latentes. 2. modelos de
 diagnóstico cognitivo. 3. abordagem Bayesiana. I.
 Guzmán, Jorge Luis Bazán, orient. II. Título.

Renato da Silva Fernandes

Contributions to general cognitive diagnosis models

Thesis submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Jorge Luis Bazán Guzmán

USP – São Carlos
May 2024

Dedico esse trabalho à minha filha, Isabela, à minha esposa, Lidiane e ao meu pai, Renato.

AGRADECIMENTOS

Agradeço ao meu orientador professor Dr. Jorge Luis Bazán Guzmán, pela paciência, atenção, e pelas importantes contribuições na pesquisa desenvolvida.

Agradeço à professora Dra. Mariana Curi, por sua atenção, sugestões, contribuições e por atuar como minha coorientadora, mesmo que de forma extra oficial. Agradeço aos membros da banca de qualificação por suas valiosas observações e apontamentos que contribuíram para a melhoria do trabalho de pesquisa dessa tese.

Agradeço aos membros da banca de qualificação e da banca examinadora, Dr. Heliton Ribeiro Tavares, Dr. Jorge Andrés González Burgos, Dr. Luis Hilmar Valdivieso Serrano, Dr. Marcelo Andrade da Silva, por suas valiosas observações e apontamentos que contribuíram para a melhoria do trabalho de pesquisa dessa tese.

Agradeço ao IFSP (Instituto Federal de Educação, Ciência e Tecnologia de São Paulo) por conceder o afastamento das atividades profissionais para qualificação, permitindo que eu me dedicasse integralmente à realização desse curso de Doutorado.

Agradeço ao CeMEAI -USP (Centro de Pesquisa em Matemática Aplicada à Indústria) pela disponibilização de recursos computacionais do *cluster* Euler, imprescindível para a execução dos estudos de simulações apresentados nesta tese.

RESUMO

FERNANDES, RENATO S. **Contribuições para modelos gerais de diagnóstico cognitivo.** 2024. 113 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Os Modelos de Diagnóstico Cognitivo (MDC) são modelos de variáveis latentes discretas que visam a determinar o padrão de posse de habilidades ou atributos de um indivíduo com base em suas respostas a um teste. Essa classe inclui os modelos de diagnóstico gerais, cuja formulação permite que vários outros MDC sejam obtidos como um caso especial do primeiro. Este trabalho apresenta contribuições para esses modelos gerais em uma abordagem Bayesiana.

Primeiro, apresentamos formulações Bayesianas para dois modelos gerais de diagnóstico cognitivo, o modelo DINA generalizado (G-DINA) - baseado no modelo *Deterministic Input, Noisy "AND" Gate*, cuja sigla em inglês, DINA, será mantida - para respostas dicotômicas; e o modelo de diagnóstico cognitivo generalizado para respostas politômicas (MGDCP). Tais formulações incluem novos conjuntos de restrições para os parâmetros dos itens para melhorar a estimação e a interpretabilidade dos parâmetros do modelo. Ambos foram implementados utilizando o software JAGS, cujas implementações são disponibilizadas nesta tese. Para ambos os modelos, foi elaborado um estudo de simulação a fim de avaliar a acurácia na recuperação dos parâmetros do método de estimação Bayesiano e os resultados comparados com aqueles obtidos com o método de estimação clássico. Ademais, os resultados de ambos os estudos indicam que o método de estimação Bayesiano proposto recupera todos os parâmetros com acurácia igual ou superior ao do método de estimação clássico nos cenários avaliados.

Em um exemplo de aplicação, os dois modelos propostos são utilizados para examinar as respostas de 1111 estudantes universitários ao Inventário de Depressão de Beck (BDI) utilizando MDC. Primeiro, é realizada uma análise comparativa entre os resultados obtidos com o modelo DINA (adotado na formulação original dessa metodologia) e o modelo G-DINA, ambos aplicados aos dados dicotomizados. Depois, comparamos os resultados obtidos com o modelo G-DINA, aplicado aos dados dicotomizados e o MGDCP, aplicado aos dados politômicos originais.

Por fim, em outro exemplo de aplicação, apresentamos um novo sistema de recomendação para filmes que incorpora um MDC em sua formulação. O MDC é utilizado para fazer a previsão da classificação que um usuário daria a cada item e recomendar os itens com base nessas previsões. O método proposto foi aplicado a dois conjuntos de dados de filmes e seu desempenho foi comparado a outros sistemas de recomendação encontrados na literatura, apresentando resultados superiores aos dos métodos concorrentes.

Palavras-chave: modelos de variáveis latentes, modelos de diagnóstico cognitivo, abordagem Bayesiana.

ABSTRACT

FERNANDES, RENATO S. **Contributions to general cognitive diagnosis models**. 2024. 113 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Cognitive Diagnosis Models (CDMs) are discrete latent variable models that aim to determine an individual's pattern of possession of skills or attributes based on their test responses. This class includes the general diagnosis models, whose formulation allows various other CDMs to be obtained as a special case of the former. This work presents contributions to these general models under a Bayesian approach.

First, we present Bayesian formulations for two general diagnosis models, the Generalized DINA (G-DINA) model, designed for dichotomous responses; and the General Polytomous Diagnosis Model (GPDM), designed for polytomous responses. These formulations include new sets of constraints on the item parameters to improve the estimation and interpretability of the model parameters. For both models, an estimation method was implemented using the JAGS software, which is available in this thesis. Moreover, for both models, a simulation study was designed to evaluate the parameters recovery accuracy of this Bayesian estimation method and the results were compared with those obtained with the standard classical estimation method. In both studies, the results indicate that the proposed Bayesian estimation method recovers all parameters with accuracy equal to or better than the classical estimation method under the evaluated scenarios.

In an example of application, we use the proposed models to examine the responses of 1111 college students to the Beck Depression Inventory (BDI) using a CDM in the modeling process. In the first instance, we compare the results obtained with the DINA model (adopted in the original formulation of this methodology) and the G-DINA model, both applied to the dichotomized data. In the second instance, we compare the results obtained with the G-DINA model, applied to dichotomized data, and the GPDM, applied to original polytomous data.

At last, in another example of application, we present a new recommendation system for movies that incorporates a CDM in its formulation. The CDM is used to predict the ratings that a user would give to each item and make recommendations based on these predictions. The proposed method was applied to two movie datasets and its performance compared to other recommendation systems found in the literature, presenting results superior to the competing methods.

Keywords: Latent variable models, Cognitive diagnosis models, Bayesian approach.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração da obtenção do vetor reduzido de atributos para alguns casos em que $K = 4$ e $K_j = 2$	31
Figura 2 – Ilustração do mapeamento entre classes e grupos latentes para alguns casos com $K = 3$	32
Figura 3 – Ilustração da bimodalidade na distribuição à posteriori dos parâmetros dos itens do modelo G-DINA quando nenhuma restrição é imposta.	39
Figura 4 – Ilustração das restrições impostas aos parâmetros dos itens com os conjuntos de restrições RM, RMaP e RMeP, para itens com $K_j = 3$	40
Figura 5 – Resultados da acurácia na recuperação dos parâmetros dos itens no estudo de simulação	69
Figura 6 – Probabilidades de sucesso estimadas com o modelo DINA e G-DINA para os itens do teste BDI.	77
Figura 7 – Gráfico das probabilidades estimadas para o conjunto de dados do BDI	80
Figura 8 – Distribuição dos 10 filmes mais bem avaliados no conjunto de validação de DB1	93
Figura 9 – Distribuição dos 10 filmes mais bem avaliados pelos usuários segundo as notas estimadas pelo MGDCP	94
Figura 10 – Distribuição das notas previstas pelo MGDCP de três usuários para o BD1	95
Figura 11 – MAP(10) do MGDCP melhorado para diferentes pontos de corte	96

LISTA DE QUADROS

Quadro 1 – Exemplos de combinações de parâmetros obtidas em função do vetor de atributos reduzidos com a FRI do G-DINA	34
Quadro 2 – Resumo dos cenários do estudo de simulação do G-DINA	45
Quadro 3 – Correspondência entre o parâmetro reportado e os originais dos modelos DINA e G-DINA.	74

LISTA DE TABELAS

Tabela 1 – Matriz Q do cenário I do estudo de simulação do G-DINA	46
Tabela 2 – Parâmetros estruturais do cenário I do estudo de simulação do G-DINA . . .	46
Tabela 3 – Matriz Q do cenário II do estudo de simulação do G-DINA	47
Tabela 4 – Parâmetros estruturais do cenário II do estudo de simulação do G-DINA . .	47
Tabela 5 – Matriz Q do cenário III e IV do estudo de simulação do G-DINA	48
Tabela 6 – Parâmetros estruturais do cenário III do estudo de simulação do G-DINA . .	48
Tabela 7 – Parâmetros estruturais do cenário IV do estudo de simulação do G-DINA . .	48
Tabela 8 – RMSE médio obtidos pelos diferentes métodos de estimação no estudo de simulação do G-DINA	51
Tabela 9 – ACCR médio obtido pelos diferentes métodos de estimação no estudo de simulação do G-DINA	52
Tabela 10 – PCCR médio obtido pelos diferentes métodos de estimação no estudo de simulação do G-DINA	53
Tabela 11 – Tempo médio, em segundos, para a execução dos algoritmos de estimação do G-DINA	54
Tabela 12 – Matriz Q dos cenários I e II do estudo de simulação do MGDCP	63
Tabela 13 – Matriz Q do cenário III do estudo de simulação do MGDCP	64
Tabela 14 – Parâmetros dos itens dos cenários I e II para o nível de resposta $h = 1$ do estudo de simulação do MGDCP	64
Tabela 15 – Parâmetros dos itens dos cenários I e II para o nível de resposta $h = 1$ do estudo de simulação do MGDCP	65
Tabela 16 – Formulas utilizadas para obtenção dos parâmetros dos itens para níveis de respostas $h > 1$ no estudo de simulação do MGDCP	65
Tabela 17 – Resultados da acurácia na recuperação dos atributos latentes no estudo de simulação	68
Tabela 18 – Matriz Q do utilizada nas metodologias de classificação das respostas do BDI	72
Tabela 19 – Comparação do ajuste dos modelos DINA e G-DINA para os dados do BDI	75
Tabela 20 – Tabela de contingência do perfil latente estimado pelos modelo DINA e G-DINA aos respondentes do teste BDI	76
Tabela 21 – Proporção dos perfis latentes estimada com o modelo DINA e G-DINA . . .	78
Tabela 22 – Probabilidades estimadas com os modelos G-DINA e MGDCP para o con- junto de dados do BDI	81

Tabela 22 – Probabilidades estimadas com os modelos G-DINA e MGDCP para o conjunto de dados do BDI	82
Tabela 23 – Tabela de contingência do perfil latente estimado pelos modelo G-DINA e MGDCP aos respondentes do teste BDI	83
Tabela 24 – Tabela comparativa entre dois indivíduos	84
Tabela 25 – Acurácia na predição das notas para os conjuntos de Dados DB1 e DB2	97
Tabela 26 – MAP(k) para os conjuntos de dados DB1 e DB2, considerando apenas os itens avaliados.	98
Tabela 27 – MAP(k) para os conjuntos de dados DB1 e DB2, considerando os dados ausentes.	98

LISTA DE ABREVIATURAS E SIGLAS

A-CDM	<i>Additive Cognitive Diagnosis Model</i> (ver MDCA)
ACCR	<i>Attributes' Correct Classification Ratio</i> (taxa de classificação correta dos atributos)
BDI	<i>Beck Depression Inventory</i> (Inventário de Depressão de Beck)
CBF	<i>Content-Based Filtering</i>
CDM	<i>Cognitive Diagnosis Model</i> (ver MDC)
CFG	<i>Content Filtering of Gender</i>
DIC	<i>Deviance Information Criterion</i>
DINA	<i>Deterministic inputs, noise "AND" gate</i>
DINO	<i>Deterministic inputs, noise "OR" gate</i>
EM	<i>Expectation - Maximization</i>
ENEM	Exame Nacional do Ensino Médio
FRI	Função de Resposta ao Item
FRVA	Função de Redução de Vetor de Atributos
G-DINA	<i>Generalized DINA</i> (DINA Generalizado)
GDM	<i>General Diagnostic Model</i> (ver MGDC)
GPDM	<i>General Polytomous Diagnosis Model</i> (ver MGDCP)
HLPC	<i>Higher and Lower Probability Constraints</i> (ver RMaMeP)
HPC	<i>Higher Probability Constraints</i> (ver RMaP)
JAGS	<i>Just Another Gibb's Sampler</i>
KNN	<i>K-nearest neighbors</i>
LCDM	<i>Log-linear Cognitive Diagnostic Model</i> (ver MDCLog)
LLM	<i>Linear Logistic Model</i> (ver MLL)
LPC	<i>Lower Probability Constraints</i> (ver RMeP)
MAE	<i>Mean Absolute Error</i> (erro médio absoluto)
MAP(k)	<i>Mean Average Precision at k</i>
MC	<i>Monotonic Constraints</i> (ver RM)
MCMC	<i>Markov Chain Monte Carlo</i> (Monte Carlo via Cadeia de Markov)
MCMC	<i>Monte Carlo Markov Chain</i>
MDC	Modelos de Diagnóstico Cognitivo
MDCA	Modelos de Diagnóstico Cognitivo Aditivo
MDCLog	Modelo de Diagnóstico Cognitivo Log-linear

MGDC	Modelo Geral de Diagnóstico Cognitivo
MGDCP	Modelo Geral de Diagnóstico Cognitivo para respostas Politômicas
MLL	Modelo Logístico Linear
MRR	Modelo Reduzido Reunificado
NeuMF	<i>Neural Network with Matrix Factorization</i>
PCCR	<i>Patterns' Correct Classification Ratio</i> (taxa de classificação correta dos padrões)
R-RUM	<i>Reduced Reunified Model</i> (ver MRR)
RFMaMeP	Restrição Flexível da Maior e da Menor Probabilidade
RM	Restrição Monotônica
RMaMeP	Restrição da Maior e da Menor Probabilidade
RMaP	Restrição da Maior Probabilidade
RMeP	Restrição da Menor Probabilidade
RMSE	<i>Root of Mean Squared Error</i> (raiz do erro quadrático médio)
RMSEA	<i>Root of Mean Squared Error of an Approximation</i> (raiz quadrada média do erro de aproximação)
SAEB	Sistema de Avaliação da Educação Básica
SARESP	Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo
soft HLPC	<i>soft Higher and lower Probability Constraints</i> (ver RFMaMeP)
TRI	Teoria de Resposta ao Item
VAE	<i>Variational Auto Encoder</i>
WAIC	<i>Widely Applicable Information Criterion</i>

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Contexto da pesquisa	23
1.2	Os modelos de diagnóstico cognitivo	23
1.3	Objetivos	25
1.4	Estrutura do trabalho	26
2	UMA ABORDAGEM BAYESIANA PARA O MODELO G-DINA	27
2.1	Introdução	27
2.2	Os modelos de diagnóstico cognitivo	28
2.3	O modelo G-DINA	30
2.4	Propostas de restrição dos parâmetros do item para o G-DINA	38
2.5	Formulação Bayesiana para o modelo G-DINA	41
2.6	Estudo de simulação	45
2.7	Discussão	54
3	UMA ABORDAGEM BAYESIANA PARA O MODELO MGDCP	55
3.1	Introdução	55
3.2	O modelo MGDCP	56
3.3	Formulação Bayesiana para o MGDCP	59
3.4	Estudo de simulação	62
3.5	Discussão	70
4	APLICAÇÃO - DADOS DE DEPRESSÃO	71
4.1	Introdução	71
4.2	Metodologia de classificação da depressão utilizando MDC	71
4.3	A análise comparativa entre DINA e G-DINA	73
4.4	Análise comparativa entre G-DINA e MGDCP	78
4.5	Discussão	85
5	APLICAÇÃO - SISTEMA DE RECOMENDAÇÃO	87
5.1	Introdução	87
5.2	O sistema de recomendação com MGDCP	88
5.3	Outros sistemas de recomendação	89
5.4	Metodologia para avaliação dos métodos	90

5.5	Os conjuntos de dados	92
5.6	Inclusão da popularidade no sistema de recomendação com MGDCP	92
5.7	Resultados	96
5.8	Discussão	99
6	CONCLUSÃO E COMENTÁRIOS FINAIS	101
6.1	Comentários finais	101
6.2	Produção acadêmica durante o Doutorado	104
6.3	Desenvolvimentos futuros	105
	REFERÊNCIAS	107

INTRODUÇÃO

1.1 Contexto da pesquisa

Nesta tese são apresentadas novas contribuições acerca dos modelos gerais de diagnóstico cognitivo com enfoque na inferência Bayesiana. Neste capítulo, apresentamos os primeiros detalhes sobre o objeto e os objetivos desta pesquisa de Doutorado.

Os modelos em estudo são englobados na categoria dos modelos de variáveis latentes, que são construções hipotéticas das quais não existem métodos operacionais para realizar uma medição direta. São exemplos: inteligência, habilidade verbal, ambição. Apesar de não ser possível medir essas variáveis diretamente, seu efeito sobre outras variáveis são possíveis de serem observados e, portanto, podem ser estudadas (EVERETT, 1984).

1.2 Os modelos de diagnóstico cognitivo

Os Modelos de Diagnóstico Cognitivo (MDC, ou CDM, em inglês) são uma classe de modelos de variáveis latentes discretas que têm como objetivo determinar quais atributos ou habilidades um indivíduo possui ou não com base nas respostas desse indivíduo em um teste (DIBELLO; ROUSSOS; STOUT, 2006; RUPP; TEMPLIN; HENSON, 2010; GEORGE; ROBITZSCH, 2015).

Existe uma grande variedade de MDC na literatura, tais como *Deterministic Input Noisy “AND” Gate* (DINA; Haertel, 1989), *Deterministic Input Noisy “OR” Gate* (DINO; Templin e Henson, 2006)¹, modelo reduzido reunificado (MRR, ou R-RUM, do termo em inglês *Reduced Reunified Model*; Hartz, 2002), modelo de diagnóstico cognitivo aditivo (MDCA, ou A-CDM, do termo em inglês *Additive Cognitive Diagnosis Model*; de la Torre, 2011) e o modelo logístico linear (MLL, ou LLM, do termo em inglês *Linear Logistic Model*; Maris, 1999).

¹ Nessa tese, optamos por manter os nomes e siglas originais dos modelos DINA e DINO por não haver tradução conhecida para o nomes desses modelos e por as siglas em inglês serem adotadas usualmente na literatura.

Existem aplicações dos MDC em diversas áreas, das quais podemos destacar as aplicações no campo da avaliação educacional (CHEN; CHEN, 2016a; LI; HUNTER; LEI, 2016; YAMAGUCHI; OKADA, 2018), da avaliação psicométrica (CHEN *et al.*, 2019), da avaliação de desordens psicológicas (TEMPLIN; HENSON, 2006; DE LA TORRE; ARK; ROSSI, 2018; SILVA *et al.*, 2018), entre outros.

Apesar da popularidade dos MDC ter crescido nos últimos anos no mundo, essa classe de modelos ainda é pouco conhecida no Brasil. Em termos de modelos de variáveis latentes, os modelos da Teoria de Resposta ao Item (TRI) são muito mais populares no Brasil, sendo utilizado em exames de larga escala do Brasil, tais como o Sistema de Avaliação da Educação Básica (SAEB), o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) (ANDRADE; TAVARES; VALLE, 2000) e, mais recentemente, o Exame Nacional do Ensino Médio (ENEM) (JUNIOR, 2010).

Com relação às duas classes de modelos citadas, a principal diferença está no tipo de variável latente considerada. Nos modelos da TRI, o traço latente é representado por uma ou mais variáveis latentes contínuas que representam a proficiência ou habilidade do indivíduo avaliado. Já nos MDC, a variável latente é um vetor binário em que cada componente representa a posse ou não de um determinado atributo ou habilidade pelo indivíduo.

Tais diferenças tem um grande impacto na aplicabilidade desses modelos. Considerando o contexto de avaliação educacional, por exemplo, os modelos TRI são mais indicados quando o objetivo é verificar o nível de conhecimento de cada aluno, para efeitos de comparação e classificação. Por sua vez, os MDC são capazes de identificar os pontos fortes e fracos de cada aluno, permitindo uma medição mais eficaz do aprendizado e do progresso de cada aluno (DE LA TORRE, 2009). As informações fornecidas pelos MDC podem ser utilizadas para aprimorar o processo de ensino aprendizagem, por meio de planejamento de novas atividades de acordo com as dificuldades dos alunos e até mesmo a elaboração de atividades didáticas personalizadas, de acordo com as necessidades específicas de cada aluno.

Um elemento fundamental na definição dos MDC é a matriz Q (TATSUOKA, 1983), que especifica quais atributos são avaliados por um item do teste. Usualmente, essa matriz é definida por especialistas sobre os assuntos avaliados pelo questionário. Por envolver critérios subjetivos, pode ocorrer da matriz Q ser mal especificada. Nesse sentido, existem diversos estudos sobre o impacto da especificação incorreta dessa matriz na estimação dos parâmetros (RUPP; TEMPLIN, 2008; MIAO; MILLER; REN, 2017) e diversas propostas de metodologias de validação (DE LA TORRE, 2008; DE LA TORRE; CHIU, 2016; WANG *et al.*, 2018; NÁJERA *et al.*, 2020) e refinamento (CHIU, 2013) dessa matriz. Uma outra abordagem possível consiste em obter a matriz Q por meio de um processo de estimação Chen *et al.* (2018), Wang, Cai e Tu (2020).

Outro ponto importante é a escolha do modelo de diagnóstico cognitivo mais adequado a uma aplicação de interesse. Nesse aspecto, destacamos uma sub-classe de modelos denominados *modelos gerais de diagnóstico cognitivo*, cuja a principal característica é a capacidade de se

obter outros MDC como casos particulares. Tais modelos são bastante interessantes devido à versatilidade que possuem. Dentro dos modelos gerais, existe o DINA Generalizado (G-DINA; de la Torre, 2011)², o modelo de diagnóstico cognitivo log-linear (MDCLog, ou LCDM, do termo em inglês *Log-linear Cognitive Diagnostic Model*; Henson, Templin e Willse, 2009) e o modelo geral de diagnóstico cognitivo (MGDC, ou GDM, do termo em inglês *General Diagnostic Model*; von Davier, 2019), quando consideramos os modelos para respostas dicotômicas; para respostas politômicas, há o modelo geral de diagnóstico cognitivo para respostas politômicas (MGDCP, ou GPDM, do termo em inglês *General Polytomous Diagnosis Model*; Chen e de la Torre, 2018).

Os modelos gerais de diagnóstico cognitivo constituem o objeto de pesquisa de Doutorado apresentado nesta tese, cujos objetivos são apresentados na próxima seção.

1.3 Objetivos

1.3.1 Objetivo geral

Contribuir para o desenvolvimento teórico e prático dos modelos gerais da classe MDC, sob uma perspectiva Bayesiana, incluindo, mas, não se limitando, à proposição de novos modelos e desenvolvimento de novos métodos de estimação.

1.3.2 Objetivos específicos

1. Desenvolver uma abordagem Bayesiana para o modelo G-DINA, incluindo a proposição de dois novos conjuntos de restrições dos parâmetros dos itens, a implementação computacional de um método de estimação por meio de Monte Carlo via Cadeia de Markov (MCMC, do termo em inglês *Markov Chain Monte Carlo*) e a verificação da recuperação dos parâmetros por meio de estudo de simulação.
2. Desenvolver uma abordagem Bayesiana para o MGDCP, incluindo a proposição um novo conjunto de restrições dos parâmetros dos itens, a implementação computacional de um método de estimação via MCMC, e a verificação da recuperação dos parâmetros por meio de estudo de simulação.
3. Apresentar uma aplicação dos métodos de estimação propostos na análise de dados de depressão para as respostas do Inventário de Depressão de Beck.
4. Apresentar uma aplicação dos modelos de diagnóstico cognitivo em sistemas de recomendações de filmes.

² Aqui, adotamos novamente a sigla G-DINA, do termo em inglês *Generalized DINA*, por ser amplamente utilizada na literatura.

1.4 Estrutura do trabalho

No capítulo 2 é apresentada uma abordagem Bayesiana para o modelo G-DINA. Primeiro, são apresentados os detalhes dos modelos de diagnóstico cognitivo e do modelo G-DINA em específico. A seguir, são introduzidos dois novos conjuntos de restrição dos parâmetros dos itens, bem como a justificativa para tais propostas. Em seguida, são apresentados os detalhes da formulação Bayesiana do modelo G-DINA proposta neste trabalho. Por fim, é apresentado o estudo de simulação desenvolvido para avaliar a acurácia da recuperação dos parâmetros do modelo G-DINA por meio do método de estimação Bayesiano proposto.

No capítulo 3, é apresentada uma abordagem Bayesiana para o modelo MGDCP. Primeiro, são apresentados os detalhes do MGDCP. Na sequência, é apresentada uma formulação Bayesiana para o MGDCP, com a inclusão de um novo conjunto de restrições. Por fim, é apresentado o estudo de simulação desenvolvido para avaliar a acurácia da recuperação dos parâmetros do método de estimação Bayesiano para o MGDCP.

No capítulo 4, é apresentado um exemplo de aplicação no qual os modelos G-DINA e MGDCP Bayesianos, propostos nos capítulos anteriores, são utilizados para analisar os dados referente à depressão. Mais especificamente, são analisadas as respostas do Inventário de Depressão de Beck por meio de uma adaptação da metodologia de classificação da depressão proposta por [Silva et al. \(2018\)](#), na qual o modelo MDC utilizado originalmente é substituído pelos modelos G-DINA e MGDCP.

No capítulo 5, é apresentada uma nova aplicação na qual um modelo de diagnóstico cognitivo é utilizado na formulação de um sistema de recomendação de filmes. Esse sistema utiliza o MGDCP para estimar os parâmetros dos itens e o perfil latente dos indivíduos e criar listas de sugestões de filmes personalizadas a partir da predição das notas de avaliação desses indivíduos aos filmes do banco de dados.

Por fim, no capítulo 6 são discutidos os resultados e as descobertas apresentadas nesta tese. Também são elencadas as produções acadêmicas resultantes desta pesquisa de Doutorado e apresentados alguns tópicos a serem investigados no futuro.

UMA ABORDAGEM BAYESIANA PARA O MODELO G-DINA

2.1 Introdução

O G-DINA é um modelo de diagnóstico cognitivo para respostas dicotômicas que tem, como ponto forte, a grande flexibilidade decorrente de sua formulação, que engloba diversos outros MDC como casos particulares (DE LA TORRE, 2011), tais como o DINA, o DINO, o modelo de diagnóstico cognitivo aditivo (MDCA), o modelo reduzido reunificado (MRR), o modelo logístico linear (MLL) e o modelo de diagnóstico cognitivo log-linear (MDCLog).

Esse modelo tem se popularizado nos últimos anos, o que demonstra o número de publicações envolvendo esse modelo (CHEN; CHEN, 2016a; CHEN; CHEN, 2016b; LI; HUNTER; LEI, 2016; RAVAND, 2016; KOEDSRI; NGUDGRATOKE, 2018; RAVAND; ROBITZSCH, 2018; HWANG; CHOI; HAND, 2020; ARYADOUST, 2021; TONEKABONI; RAVAND; REZVANI, 2021; LI *et al.*, 2021; DELAFONTAINE *et al.*, 2022; WU *et al.*, 2022; KREITCHMANN *et al.*, 2023; LIANG *et al.*, 2023). Embora esses exemplos de aplicações sejam quase que exclusivamente dentro do contexto de avaliação educacional, esse modelo pode ser aplicado em diversas outras áreas, dentre as quais podemos destacar a área médica e a psicológica.

Também é comum encontrar estudos que comparam o ajuste do modelo G-DINA com outros MDC (LI; HUNTER; LEI, 2016; RAVAND; ROBITZSCH, 2018; ARYADOUST, 2021); em quase todos os casos, o modelo G-DINA se ajusta melhor aos dados.

Embora exista na literatura métodos de estimação Bayesiano para todos os modelos mencionados acima como casos especiais de G-DINA, nesta tese é apresentada a primeira formulação Bayesiana para esse modelo considerando sua forma saturada e a função de ligação de identidade.

Um dos benefícios da abordagem Bayesiana é fornecer a distribuição a posteriori dos

parâmetros, que contém mais informações do que as estimativas pontuais. Tais distribuições podem ser utilizadas, por exemplo, para construir intervalos de credibilidades dos parâmetros de interesse. Além disso, os métodos de estimação Bayesiano possuem a vantagem de não dependerem de teoria assintótica para garantir a validade dos resultados e são mais robustos ao lidar com modelos complexos (ZHAN *et al.*, 2019).

Neste capítulo, é apresentado um método de estimação Bayesiano para o modelo G-DINA saturado com a função de ligação de identidade. Também são apresentadas duas propostas de restrição que visam a aprimorar a estimativa de parâmetros. Adicionalmente, é apresentado um estudo de simulação conduzido para avaliar a acurácia na recuperação dos parâmetros do método proposto.

2.2 Os modelos de diagnóstico cognitivo

Nesta seção, apresentaremos os elementos fundamentais para a definição de um Modelo de Diagnóstico Cognitivo. Por simplicidade, apresentaremos esses elementos dentro do contexto mais usual: a aplicação de um questionário para avaliar quais atributos o respondente possui ou não possui. Existem aplicações em outros contextos, nos quais o significado e interpretação das variáveis envolvidas podem ser diferentes do que será apresentado aqui. Todavia, a compreensão do funcionamento desses modelos, dentro do contexto que será apresentado, é suficiente para que um pesquisador consiga utilizar os MDC em outros contextos, realizando as adaptações necessárias na definição das variáveis.

Considere um questionário com J itens que avaliam K atributos, respondidos por N indivíduos, onde cada item é avaliado de forma dicotômica, sendo J , K e N números naturais.

Seja Y uma matriz de dimensão $N \times J$, denominada de matriz de resposta, em que o elemento da posição (i, j) , denotado por y_{ij} , é um valor observado definido da seguinte forma:

$$y_{ij} = \begin{cases} 1, & \text{se o indivíduo } i \text{ responde ao item } j \text{ corretamente;} \\ 0, & \text{caso contrário} \end{cases}$$

Ainda, denotaremos por Y_j a variável aleatória que corresponde à resposta do item j ($= 1, \dots, J$) e assume valores 0 ou 1, em que 1 representa uma resposta correta e 0, o oposto.

Seja Q a matriz que indica os atributos requeridos em cada item, em que q_{jk} representa o elemento da posição (j, k) e é definido da seguinte forma:

$$q_{jk} = \begin{cases} 1, & \text{se o atributo } k \text{ é requerido para responder o item } j; \\ 0, & \text{caso contrário} \end{cases}$$

Usualmente, a matriz Q é conhecida e definida anteriormente à etapa de análise dos dados, sendo definida com auxílio de especialistas no assunto do teste ou com base em informações adicionais disponíveis sobre os itens do teste. Existem estudos, nos quais a matriz Q é tida como desconhecida e considerada como um parâmetro do modelo a ser estimado (CHEN *et al.*, 2018). Contudo, nesta tese, será considerado que a matriz Q é previamente conhecida, exceto quando indicado diferente. Ainda sobre a matriz Q , utilizaremos \mathbf{q}_j para denotar o vetor correspondente à j -ésima linha dessa matriz.

Nessa classe de modelos, a variável latente de interesse é um vetor binário de dimensão K denominado vetor de atributos e denotado por $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^T$, em que α_k é o indicador da posse do atributo k ($= 1, \dots, K$). Nesses termos, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{iK})^T$, com $i = 1, \dots, N$ representa o vetor de atributos do indivíduo i , em que α_{ik} , para $k = 1, \dots, K$, é uma variável binária não observável definida como:

$$\alpha_{ik} = \begin{cases} 1, & \text{se o indivíduo } i \text{ possui o atributo } k; \\ 0, & \text{caso contrário.} \end{cases}$$

Adicionalmente, utilizaremos \mathbf{A} para denotar a matriz binária de dimensão $N \times K$ cuja i -ésima linha ($i = 1, \dots, N$) corresponde ao vetor $\boldsymbol{\alpha}_i$. Tal matriz será denominada, doravante, de matriz de atributos.

Os modelos de Diagnóstico Cognitivo são definidos com base nessas três matrizes (Y , Q e \mathbf{A}). O que distingue um MDC do outro é a forma que tais matrizes são combinadas para se obter a *função de resposta ao item* (FRI). A FRI é a função que relaciona a variável latente que representa a habilidade do indivíduo (o vetor de atributos, neste caso) e a probabilidade de sucesso no item em questão.

A seguir, apresentamos a definição de dois MDC muito utilizados na literatura.

2.2.1 Os modelos DINA e DINO

Seja $P_j(\boldsymbol{\alpha}_i) = Pr(Y_j = 1 | \boldsymbol{\alpha} = \boldsymbol{\alpha}_i)$ a probabilidade do indivíduo i responder corretamente o item j , dado que esse indivíduo possui vetor de atributos $\boldsymbol{\alpha}_i$. Por sinteticidade, iremos nos referir à probabilidade de responder corretamente um item j como *probabilidade de sucesso*.

O modelo DINA (HAERTEL, 1989; JUNKER; SIJTSMA, 2001; DE LA TORRE, 2009) é, provavelmente, o mais popular dos MDC. A FRI do modelo DINA é dado por:

$$P_j(\boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad \text{em que } \eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$$

para todo $i = 1, \dots, N$, $k = 1, \dots, K$ e $j = 1, \dots, J$.

Nesse modelo, cada item j possui dois parâmetros: g_j (denominado *guessing*) e s_j (denominado *slipping*). Note que, no modelo DINA, $\eta_{ij} = 1$ se, somente se, o indivíduo possui todos os atributos requeridos pelo item j . Dessa forma, $g_j = Pr(Y_j = 1 | \eta_{ij} = 0)$ representa a probabilidade do indivíduo acertar o item j , dado que ele não possui todos os atributos requeridos por esse item e pode ser interpretado como a probabilidade de acertar o item j no “chute”. De modo análogo, $s_j = Pr(Y_j = 0 | \eta_{ij} = 1)$ representa a probabilidade do indivíduo errar o item j , dado que ele possui todos os atributos requeridos por esse item e pode ser interpretado como a probabilidade de cometer um “deslize” no item j . Conforme salientado por [de la Torre \(2011\)](#), os termos *guessing* e *slipping* são apenas mnemônicos, pois os fenômenos de acertar um item, sem dominar todos os atributos requeridos, ou errar um item, dominando todos os atributos requeridos, podem ocorrer por outras razões que configurem um “chute” ou “deslize”.

O modelo DINO ([TEMPLIN; HENSON, 2006](#)) possui suas semelhanças com o DINA, porém adotando uma lógica um pouco diferente. A FRI do modelo DINO é dado por:

$$P_j(\boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad \text{em que} \quad \eta_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$$

para todo $i = 1, \dots, N$, $k = 1, \dots, K$ e $j = 1, \dots, J$.

Novamente, há dois parâmetros para cada item: g_j (*guessing*) e s_j (*slipping*), porém com interpretações diferentes. Note que, no modelo DINO, $\eta_{ij} = 0$ se, somente se, o indivíduo não possui nenhum dos atributos requeridos pelo item j . No caso do DINO, o parâmetro *guessing* $g_j = Pr(Y_j = 1 | \eta_{ij} = 0)$ representa a probabilidade do indivíduo acertar o item j , dado que ele não possui nenhum dos atributos requeridos por esse item; já o parâmetro *slipping* $s_j = Pr(Y_j = 0 | \eta_{ij} = 1)$ representa a probabilidade do indivíduo errar o item j , dado que ele possui pelo menos um dos atributos requeridos por esse item.

2.3 O modelo G-DINA

O modelo G-DINA foi proposto por ([DE LA TORRE, 2011](#)) como uma generalização do modelo DINA. Nele, a função resposta do item j só é afetada pelos atributos requeridos por esse item. Sendo assim, os atributos irrelevantes podem ser removidos do processo de modelagem. Com esse intuito, definimos o *vetor reduzido de atributos*, como segue:

Seja $K_j = \sum_{k=1}^K q_{jk}$ o número de atributos requeridos pelo item j ; denotamos por $\boldsymbol{\alpha}_i^{*j} = (\alpha_{i1}^*, \dots, \alpha_{iK_j}^*)^T$ o vetor reduzido de atributos do indivíduo i para o item j , que é obtido tomando o vetor (completo) de atributos $\boldsymbol{\alpha}_i$ e mantendo apenas os componentes requeridos para responder o item j . Em outras palavras, $\boldsymbol{\alpha}_i^{*j}$ é o sub-vetor de $\boldsymbol{\alpha}_i$, formado pelos componentes α_{ik} , $k = 1, \dots, K$, tais que $q_{jk} = 1$.

A Figura 1 ilustra como o vetor reduzido $\boldsymbol{\alpha}_i^{*j}$ é obtido em uma situação no qual o item requer dois dos quatro atributos avaliados (isto é, $K = 4$ e $K_j = 2$).

Figura 1 – Ilustração da obtenção do vetor reduzido de atributos para alguns casos em que $K = 4$ e $K_j = 2$

\mathbf{q}_j	vetor completo $\boldsymbol{\alpha}$	vetor reduzido $\boldsymbol{\alpha}^{*j}$
$(1, 0, 0, 1)^T$	$(1, 1, 1, 0)^T$	$(1, 0)^T$
	$(1, 0, 1, 0)^T$	$(1, 0)^T$
	$(0, 1, 1, 0)^T$	$(0, 0)^T$
$(1, 1, 0, 0)^T$	$(1, 1, 1, 0)^T$	$(1, 1)^T$
	$(1, 0, 1, 0)^T$	$(1, 0)^T$
	$(0, 1, 1, 0)^T$	$(0, 1)^T$

Fonte: Elaborada pelo autor.

Mais formalmente, podemos definir a função *FRVA* (sigla para função de redução de vetor de atributos), que recebe o vetor de atributos completos $\boldsymbol{\alpha}_i$ e o vetor \mathbf{q}_j , e retorna o vetor reduzido de atributos $\boldsymbol{\alpha}_i^{*j}$. Seja $\zeta(m, \mathbf{q}_j) = \min\{k_0 : \sum_{k=1}^{k_0} q_{jk} \leq m\}$; a função *FRVA* é definida como apresentado na [Equação 2.1](#).

$$FRVA : \{0, 1\}^K \times \{0, 1\}^K \longrightarrow \bigcup_{k=1}^K \{0, 1\}^k \quad (2.1)$$

$$(\boldsymbol{\alpha}; \mathbf{q}_j) \longmapsto \boldsymbol{\alpha}^{*j} = (\alpha_1^*, \dots, \alpha_{K_j}^*)^T$$

em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^T$; $\mathbf{q}_j = (q_{j1}, \dots, q_{jK})^T$ e $\alpha_m^* := \alpha_{\zeta(m, \mathbf{q}_j)}^*$ para todo $m = 1, \dots, K_j$.

Faremos, agora, algumas importantes observações sobre os vetores de atributos: $\boldsymbol{\alpha}_i$ é um vetor binário de dimensão K e existem $C = 2^K$ possíveis valores para esse vetor. Sendo assim, podemos tratar o vetor de atributos $\boldsymbol{\alpha}_i$ como uma variável categórica com C possíveis valores. Utilizaremos $\boldsymbol{\alpha}_{(c)} = (\alpha_{c1}, \dots, \alpha_{cK})^T$ para denotar o perfil de atributos da classe latente $c = 1, \dots, C$; sendo $\alpha_{ck} = 1$, se os indivíduos da classe c possuem o atributo k ($= 1, \dots, K$) e $\alpha_{ck} = 0$, caso contrário. Note que, para cada $i = 1, \dots, N$, existe um único $c \in \{1, \dots, C\}$ tal que $\boldsymbol{\alpha}_i \equiv \boldsymbol{\alpha}_{(c)}$.

Convém explicar alguns detalhes da notação utilizada para os vetores de atributos neste trabalho: $\boldsymbol{\alpha}_i$ (índice sem parênteses) será utilizado para denotar o vetor de atributos de um determinado indivíduo i ; $\boldsymbol{\alpha}_{(c)}$ (com o índice entre parênteses) para denotar vetor de atributos de uma determinada classe latente e $\boldsymbol{\alpha}$ (sem índice), para representar a variável, de forma geral. Essa distinção será útil na apresentação do modelo e do método de estimação.

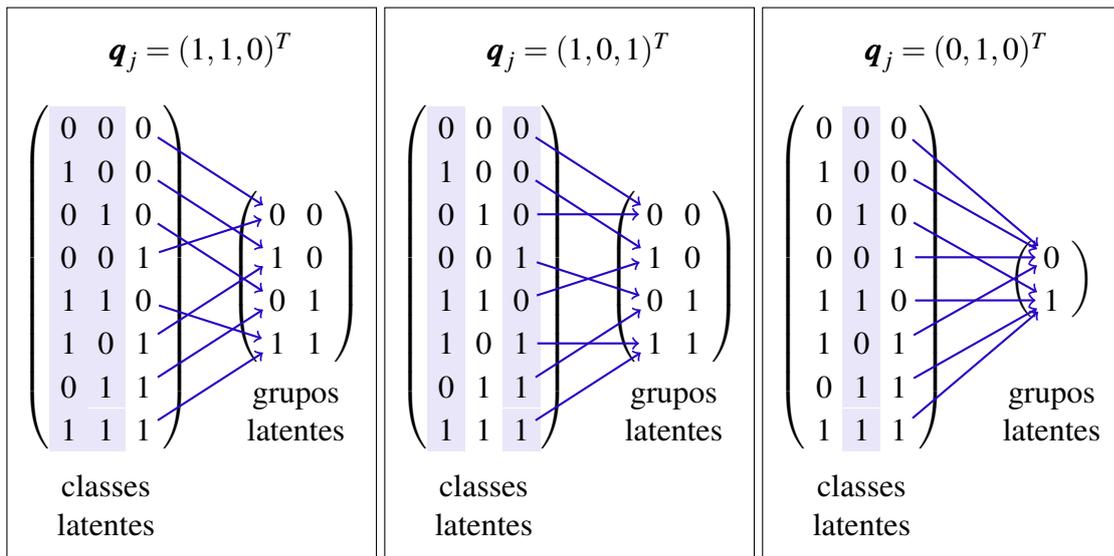
De modo similar, para cada j , os vetores de atributos reduzidos podem ser tratados como variáveis categóricas, sendo $C_j = 2^{K_j}$ o número de possíveis valores que essa variável pode assumir. Analogamente, utilizaremos $\boldsymbol{\alpha}_{(l)}^{*j} = (\alpha_{l1}^*, \dots, \alpha_{lk}^*, \dots, \alpha_{lK_j}^*)^T$ para denotar o grupo latente¹ $l = 1, \dots, C_j$ em relação ao item j ; sendo $\alpha_{lk}^* = 1$, se os indivíduos que pertencem ao grupo latente l em relação ao item j possuem o k -ésimo atributo requerido por esse item e $\alpha_{lk}^* = 0$, caso contrário. Note que, para cada $j = 1, \dots, J$ e $i = 1, \dots, N$, existe um único $l \in \{1, \dots, C_j\}$

tal que $\alpha_i^{*j} \equiv \alpha_{(l)}^{*j}$. Assim como para os vetores de atributos, α_i^{*j} (índice sem parênteses) será utilizado para denotar o vetor reduzido de atributos de um determinado indivíduo e $\alpha_{(l)}^{*j}$ (com o índice entre parênteses) para denotar o vetor correspondente a um determinado grupo latente e α^{*j} , a variável, de forma geral.

Outro ponto importante a se destacar é que, para cada item j , o conjunto das classes latentes $\{\alpha_{(1)}, \dots, \alpha_{(C)}\}$ é mapeado de forma unívoca com o conjunto de grupos latentes $\{\alpha_{(1)}^{*j}, \dots, \alpha_{(C_j)}^{*j}\}$. Ademais, para cada j , o conjunto de classes é particionado em C_j subconjuntos com 2^{K-K_j} elementos, sendo cada um desses subconjuntos associado a um elemento distinto do conjunto de grupos latentes.

A Figura 2 ilustra esse mapeamento para um teste com $K = 3$, considerando três vetores q_j distintos.

Figura 2 – Ilustração do mapeamento entre classes e grupos latentes para alguns casos com $K = 3$



Fonte: Elaborada pelo autor.

2.3.1 Função de resposta ao item do modelo G-DINA

Seja $P_j(\alpha) = Pr(Y_j = 1 | \alpha = \alpha_{(c)})$ a probabilidade de um indivíduo com vetor de atributos $\alpha_{(c)}$ responder o item j corretamente. No modelo G-DINA, a função de probabilidade da resposta Y_j não é afetada pelos atributos não requeridos por esse item. Desse modo, a probabilidade apresentada acima pode ser escrita como $Pr(Y_j = 1 | \alpha^{*j} = \alpha_{(l)}^{*j})$, onde $\alpha_{(l)}^{*j}$ representa o vetor de atributos reduzido correspondente à $\alpha_{(c)}$ para o item j . Desse modo, temos:

¹ Aqui, o termo *grupo latente* é utilizado para fazer distinção entre as classes latentes previamente definidas, embora sejam também *classes de equivalência*. Mais especificamente, o termo “grupos latentes” é utilizado para nos referir às classes de vetores reduzidos de atributos, enquanto classes latentes, às classes de vetores (completos) de atributos.

$$P_j^*(\boldsymbol{\alpha}_i^{*j}) := Pr(Y_j = 1 | \boldsymbol{\alpha}^{*j} = \boldsymbol{\alpha}_i^{*j}) \equiv Pr(Y_j = 1 | \boldsymbol{\alpha} = \boldsymbol{\alpha}_i) \quad (2.2)$$

Aqui, P_j e P_j^* são utilizados para denotar as funções que recebem como argumento o vetor de atributos completo e reduzido, respectivamente. Por sucintez, iremos nos referir à $P_j(\boldsymbol{\alpha}_{(c)})$ e $P_j^*(\boldsymbol{\alpha}_{(l)}^{*j})$ como a probabilidade de sucesso, no item j , da classe latente c e do grupo latente j , respectivamente.

Nesses termos, a função de resposta ao item (FRI) do modelo G-DINA é dado pela expressão:

$$g\left(P_j^*\left(\boldsymbol{\alpha}_{(l)}^{*j}\right)\right) := \delta_{j0} + \sum_{k=1}^{K_j} \delta_{jk} \alpha_{lk}^* + \sum_{k'=k+1}^{K_j} \sum_{k=1}^{K_j-1} \delta_{jkk'} \alpha_{lk}^* \alpha_{lk'}^* + \dots + \delta_{j12\dots K_j} \prod_{k=1}^{K_j} \alpha_{lk}^* \quad (2.3)$$

para todo $\boldsymbol{\alpha}_{(l)}^{*j}$, $l = 1, \dots, C_j$. Nessa equação, $g(\cdot)$ é uma função de ligação, que pode ser a função *identidade*, *log* ou *logit* (DE LA TORRE, 2011) e δ_j são os parâmetros do item j .

Fazendo um paralelo com modelos de regressão, para cada $j = 1, \dots, J$, o conjunto de equações 2.2 e 2.3 representa um *Modelo Linear Generalizado*, com todos os efeitos de interação, no qual os componentes do vetor reduzido de atributos α_{lk}^* , $k = 1, 2, \dots, K_j$ fazem papel de variáveis preditoras (exceto pelo fato de serem latentes) e y_{ij} , o papel de variável resposta. Por esse motivo, podemos interpretar os parâmetros dos itens de modo análogo aos parâmetros de um modelo linear generalizado. Considerando a função de ligação identidade, os parâmetros δ_j podem ser interpretados da seguinte maneira:

- δ_{j0} é o intercepto do item j e representa a probabilidade de *linha de base*, que, nesse contexto, corresponde à probabilidade de um indivíduo responder corretamente o item j , quando esse indivíduo não possui nenhum dos atributos requeridos por esse item.
- δ_{jk} , $k = 1, \dots, K_j$, é o efeito principal do componente α_{lk}^* e representa a diferença entre a probabilidade de sucesso no item j de um indivíduo, que, dentre os atributos requeridos por esse item, possui apenas o k -ésimo atributo e a probabilidade de linha de base, δ_{j0} ;
- $\delta_{jkk'}$, $k = 1, \dots, K_j - 1$ e $k' = k + 1, \dots, K_j$, é o efeito de interação (de primeira ordem) entre os componentes α_{lk}^* e $\alpha_{lk'}^*$ e representa a diferença entre a probabilidade de sucesso no item j de um indivíduo, que, dentre os atributos requeridos por esse item, possui apenas k -ésimo e k' -ésimo atributos e a soma da probabilidade de linha de base δ_{j0} e dos efeitos principais, δ_{jk} e $\delta_{jk'}$;
- $\delta_{j12\dots K_j}$ é o efeito de interação (de ordem $K_j - 1$) entre os componentes α_{l1}^* , α_{l2}^* , \dots , $\alpha_{lK_j}^*$ e representa a diferença entre a probabilidade de sucesso no item j de um indivíduo que possui todos os atributos requeridos por esse item e a soma da probabilidade de linha de

base, de todos os efeitos principais e todos os efeitos de interação de ordem menor que $K_j - 1$.

Para melhor entendimento da função apresentada na [Equação 2.3](#), o [Quadro 1](#) ilustra as diferentes combinações de parâmetros em função dos vetores reduzidos de atributos para um item que requer três atributos.

Quadro 1 – Exemplos de combinações de parâmetros obtidas em função do vetor de atributos reduzidos com a FRI do G-DINA

Vetor reduzido de atributos (argumento da FRI)	Combinação de parâmetros resultante (segundo membro da FRI)
(0, 0, 0)	δ_{j0}
(1, 0, 0)	$\delta_{j0} + \delta_{j1}$
(0, 1, 0)	$\delta_{j0} + \delta_{j2}$
(0, 0, 1)	$\delta_{j0} + \delta_{j3}$
(1, 1, 0)	$\delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j12}$
(1, 0, 1)	$\delta_{j0} + \delta_{j1} + \delta_{j3} + \delta_{j13}$
(0, 1, 1)	$\delta_{j0} + \delta_{j2} + \delta_{j3} + \delta_{j23}$
(1, 1, 1)	$\delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j3} + \delta_{j12} + \delta_{j13} + \delta_{j23} + \delta_{j123}$

Fonte: Elaborada pelo autor.

2.3.2 Casos especiais do G-DINA

Impondo algumas restrições na FRI apresentada na [equação 2.3](#), diversos MDC podem ser obtidos como casos especiais do G-DINA ([DE LA TORRE, 2009](#)). Eis alguns exemplos:

- DINA: é obtido, tomando a função de ligação identidade e fixando todos os parâmetros δ_j como zero, exceto os parâmetros δ_{j0} e $\delta_{j12\dots K_j}$. Desse modo, o modelo DINA pode ser escrito como:

$$P_j^*(\boldsymbol{\alpha}_{(I)}^{*j}) = \delta_{j0} + \delta_{j12\dots K_j} \prod_{k=1}^{K_j} \alpha_{lk}^*$$

- DINO: é obtido, tomando a função de ligação identidade e fixando todos os parâmetros dos itens de modo que $\delta_{jk} = -\delta_{jk'k''} = \dots = (-1)^{K_j} \delta_{j12\dots K_j}$ para $k = 1, \dots, K_j$, $k' = 1, \dots, K_j - 1$ e $k'' > k', \dots, K_j$. Desse modo, o modelo DINO pode ser escrito como:

$$P_j^*(\boldsymbol{\alpha}_{(I)}^{*j}) := \delta_{j0} + \sum_{k=1}^{K_j} \delta_{j1} \alpha_{lk}^* - \sum_{k'=k+1}^{K_j} \sum_{k=1}^{K_j-1} \delta_{j1} \alpha_{lk}^* \alpha_{lk'}^* + \dots + (-1)^{K_j} \delta_{j1} \prod_{k=1}^{K_j} \alpha_{lk}^*$$

- Modelo de diagnóstico cognitivo aditivo (MDCA): é obtido, tomando a função de ligação identidade e fixando todos os efeitos de interação como zero. O modelo MDCA pode ser escrito como:

$$P_j^*(\boldsymbol{\alpha}_{(I)}^{*j}) = \delta_{j0} + \sum_{k=1}^{K_j} \delta_{jk} \alpha_{lk}^*$$

- Modelo logístico linear (MLL): é obtido, tomando a função de ligação *logit* e fixando todos os efeitos de interação como zero. O modelo MLL pode ser escrito como:

$$\text{logit} \left(P_j^* (\boldsymbol{\alpha}_{(l)}^{*j}) \right) = \delta_{j0} + \sum_{k=1}^{K_j} \delta_{jk} \alpha_{lk}^*$$

- modelo reduzido reunificado (MRR): é obtido, tomando a função de ligação *log* e fixando todos os efeitos de interação como zero. O modelo MRR pode ser escrito como:

$$\log \left(P_j^* (\boldsymbol{\alpha}_{(l)}^{*j}) \right) = \delta_{j0} + \sum_{k=1}^{K_j} \delta_{jk} \alpha_{lk}^*$$

- Modelo de diagnóstico cognitivo log-linear (MDCLog): é equivalente à versão saturada do G-DINA com função de ligação *logit*. O modelo MDCLog pode ser escrito como:

$$\text{logit} \left(P_j^* (\boldsymbol{\alpha}_{(l)}^{*j}) \right) = \delta_{j0} + \sum_{k=1}^{K_j} \delta_{jk} \alpha_{lk}^* + \sum_{k'=k+1}^{K_j} \sum_{k=1}^{k'-1} \delta_{jkk'} \alpha_{lk}^* \alpha_{lk'}^* + \dots + \delta_{j12\dots K_j} \prod_{k=1}^{K_j} \alpha_{lk}^*$$

2.3.3 Parâmetros estruturais do modelo

Considere a [Equação 2.3](#). Quando utilizamos a identidade como função de ligação, existem uma série de restrições nos parâmetros do item. Por exemplo, para todo item j , o intercepto δ_{j0} deve ser não negativo e os efeitos principais δ_{jk} deve satisfazer a inequação $0 < \delta_{jk} + \delta_{j0} < 1$, para todo $k = 1, \dots, K_j$. De forma geral, os parâmetros dos itens estão restritos a assumirem valores tais que, para cada vetor de atributo reduzido $\boldsymbol{\alpha}_{(l)}^{*j}$, a soma da combinação de resultante, ao se tomar $\boldsymbol{\alpha}_{(l)}^{*j}$ como argumento da FRI, deve pertencer ao intervalo $(0, 1)$. Dada essas restrições, é mais conveniente considerarmos uma parametrização alternativa baseada nas probabilidades de sucesso para a abordagem Bayesiana que será apresentada.

Seja $\boldsymbol{\delta}_j = (\delta_{j0}, \delta_{j1}, \dots, \delta_{j12\dots K_j})^T$ o vetor de dimensão C_j formado pelos parâmetros do item j ; seja $\theta_{jl} := P_j^* (\boldsymbol{\alpha}_{(l)}^{*j})$ a probabilidade de sucesso no item j dos indivíduos do grupo latente l do referido item; seja $\mathbf{P}_j = (\theta_{j1}, \dots, \theta_{jC_j})^T$ o vetor de probabilidades de sucesso do item j . Considerando a versão saturada do modelo G-DINA (isto é, o modelo sem restrições adicionais aos parâmetros), existe uma bijeção entre $\boldsymbol{\delta}_j$ e \mathbf{P}_j expressa pela [Equação 2.4](#).

$$g(\mathbf{P}_j) = M_j \cdot \boldsymbol{\delta}_j \tag{2.4}$$

em que M_j é uma matriz binária de dimensão $C_j \times C_j$, denominada de *matriz saturada de design* ([DE LA TORRE, 2011](#)).

Seja m_{jlv} o elemento da posição (l, v) da matriz M_j . Os elementos da l -ésima linha de M_j indicam quais componentes de $\boldsymbol{\delta}_j$ contribuem e quais não, na soma que resulta na probabilidade de sucesso θ_{jl} , sendo $m_{jlv} = 1$, se o v -ésimo componente de $\boldsymbol{\delta}_j$ contribui para a soma que compõe $\theta_{jl} := P_j^* (\boldsymbol{\alpha}_{(l)}^{*j})$ e $m_{jlv} = 0$, caso contrário.

Para efeito de ilustração, considere um item j que requer três atributos. Nesse caso, os vetores e matriz que constituem a [Equação 2.4](#) são:

$$\mathbf{P}_j = \begin{pmatrix} \theta_{j1} \\ \theta_{j2} \\ \theta_{j3} \\ \theta_{j4} \\ \theta_{j5} \\ \theta_{j6} \\ \theta_{j7} \\ \theta_{j8} \end{pmatrix} = \begin{pmatrix} P_j^*(0,0,0) \\ P_j^*(1,0,0) \\ P_j^*(0,1,0) \\ P_j^*(0,0,1) \\ P_j^*(1,1,0) \\ P_j^*(1,0,1) \\ P_j^*(0,1,1) \\ P_j^*(1,1,1) \end{pmatrix} M_j = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \text{ and } \boldsymbol{\delta}_j = \begin{pmatrix} \delta_{j0} \\ \delta_{j1} \\ \delta_{j2} \\ \delta_{j3} \\ \delta_{j12} \\ \delta_{j13} \\ \delta_{j23} \\ \delta_{j123} \end{pmatrix}$$

Note que, a matriz M_j depende somente do número de atributos requeridos pelo item j (K_j), não sendo necessário especificar quais são esses atributos. Ainda, para qualquer K_j , M_j é uma matriz triangular inferior cujo determinante é igual a 1. Logo, M_j é invertível e $\boldsymbol{\delta}_j$ pode ser obtido a partir de \mathbf{P}_j por meio da equação $\boldsymbol{\delta}_j = (M_j)^{-1} \cdot g(\mathbf{P}_j)$.

Pelos motivos apontados acima, na versão saturada do modelo G-DINA, tanto \mathbf{P}_j quanto $\boldsymbol{\delta}_j$ podem ser utilizados como parâmetros dos itens, sem prejuízo (MA; JIANG, 2021). No entanto, a parametrização baseada nas probabilidades possui vantagem de que, para cada item $j = 1, \dots, J$, cada componente do vetor $\mathbf{P}_j = (\theta_{j1}, \dots, \theta_{jC_j})^T$ pode assumir qualquer valor no intervalo $(0, 1)$, sem restrições, para qualquer função de ligação. Por esse motivo, \mathbf{P}_j será utilizado como o vetor de parâmetros estruturais do item j na abordagem Bayesiana do modelo G-DINA. Vale ressaltar que o método de estimação clássica usual também utiliza a parametrização baseada nas probabilidades (DE LA TORRE, 2011; MA; DE LA TORRE, 2020; ROBITZSCH *et al.*, 2020).

Nesse caso, diremos que os parâmetros $\theta_{j1}, \dots, \theta_{jC_j}$ são independentes entre si, no sentido de que o valor assumido por θ_{jc} , $c = 1, \dots, C_j$, não interfere no valor que $\theta_{jc'}$ pode assumir, $c' \in \{1, \dots, C_j\}$, $c \neq c'$.

Convém ressaltar também que a independência dos componentes de \mathbf{P}_j . Para ilustrar casos em que não há independência entre os parâmetros, considere os modelos DINA e DINO, que possuem apenas dois parâmetros estruturais por item e que representam casos especiais do modelo G-DINA. Para esses dois modelos, temos que:

$$\mathbf{P}_j^{DINA} = \begin{pmatrix} \theta_{j1} \\ \vdots \\ \theta_{jC_j} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \delta_{j0} \\ \delta_{j12\dots K_j} \end{pmatrix} \quad \text{e} \quad \mathbf{P}_j^{DINO} = \begin{pmatrix} \theta_{j1} \\ \vdots \\ \theta_{jC_j} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \delta_{j0} \\ \delta_{j1} \end{pmatrix}$$

onde \mathbf{P}_j^{DINA} e \mathbf{P}_j^{DINO} representam, respectivamente, os vetores de parâmetros estruturais dos modelos DINA e DINO, na forma de probabilidades.

Note que, nos casos acima, os vetores \mathbf{P}_j^{DINA} e \mathbf{P}_j^{DINO} possuem apenas dois graus de liberdade e, para qualquer item com $K_j > 1$, os componentes desses vetores não são independentes como acontece na versão saturada do modelo G-DINA.

2.3.4 Função de verossimilhança do G-DINA

Nessa subseção, é apresentada a função de verossimilhança do modelo G-DINA. Para defini-la, note que $Y_j | \boldsymbol{\alpha}_i$ tem distribuição *Bernoulli* ($P_j(\boldsymbol{\alpha}_i)$) para todo $i = 1, \dots, N$ e $j = 1, \dots, J$. No modelo G-DINA, assim como em outros MDC, assumimos que diferentes indivíduos respondem o teste de forma independente e que as respostas de cada indivíduo para os diferentes itens são condicionalmente independentes, dado o perfil latente desse indivíduo. Desse modo, os vetores $(y_{i1}, \dots, y_{iJ})^T$, $i = 1, \dots, N$ (que representam as respostas do indivíduo i e correspondem à i -ésima linha da matriz Y) são conjuntamente independentes, dado a matriz latente \mathbf{A} .

Seja $\mathbf{P} = (\mathbf{P}_1^T, \mathbf{P}_2^T, \dots, \mathbf{P}_J^T)^T$ o vetor de parâmetros estruturais do modelo, um vetor de dimensão $\sum_{j=1}^J C_j$ formado pela justaposição dos vetores de parâmetros \mathbf{P}_j para $j = 1, \dots, J$, definidos anteriormente.

A função de verossimilhança dos parâmetros \mathbf{A} e \mathbf{P} , para a matriz de resposta \mathbf{Y} , é dada pela Equação 2.5.

$$L_{\mathbf{Y}}(\mathbf{A}; \mathbf{P}) = \prod_{i=1}^N \prod_{j=1}^J P_j^*(\boldsymbol{\alpha}_i^{*j})^{y_{ij}} (1 - P_j^*(\boldsymbol{\alpha}_i^{*j}))^{1-y_{ij}} \quad (2.5)$$

Seja $\xi_{ij} \in \{1, \dots, C_j\}$ a variável que representa o grupo latente ao qual o indivíduo i pertence com relação ao item j , sendo que $\xi_{ji} = l$ indica que o indivíduo i pertence ao grupo l em relação ao item j ou, equivalentemente, que $AFRV(\boldsymbol{\alpha}_i, \mathbf{q}_j) = \boldsymbol{\alpha}_{(\xi_{ji})}^{*j}$. Invertendo a ordem dos produtórios na Equação 2.5 e agrupando as respostas dos indivíduos que pertencem ao mesmo grupo latente, temos que a função de verossimilhança do G-DINA pode ser reescrita conforme a Equação 2.6.

$$L_{\mathbf{Y}}(\mathbf{A}; \mathbf{P}) = \prod_{j=1}^J \prod_{l=1}^{C_j} \prod_{i: \xi_{ji}=l} \theta_{jl}^{Y_{ij}} (1 - \theta_{jl})^{1-Y_{ij}} \quad (2.6)$$

2.4 Propostas de restrição dos parâmetros do item para o G-DINA

Nessa seção, apresentaremos alguns problemas decorrentes da falta de restrição nos parâmetros dos itens do modelo G-DINA e a solução mais usual para tais problemas.

2.4.1 Problemas decorrentes da falta de restrição nos parâmetros dos itens

O G-DINA é um modelo bastante flexível e, em sua forma saturada, não há nenhuma restrição imposta às probabilidades θ_{jl} , além de estarem limitadas no intervalo $(0, 1)$. Essa falta de restrições pode levar ao aparecimento de alguns resultados contraintuitivo. Exemplificando: sem a inclusão de restrições, existem pontos no espaço paramétrico do modelo G-DINA, que representam a situação na qual os indivíduos menos aptos a responder um item (isto é, aqueles que não possuem nenhum dos atributos requeridos pelo referido item) apresentam uma probabilidade maior de respondê-lo corretamente do que os indivíduos mais aptos (isto é, aqueles que possuem todos os atributos requeridos pelo referido item). Note que a situação descrita acima representa normalmente um quadro inesperado na maioria das aplicações reais.

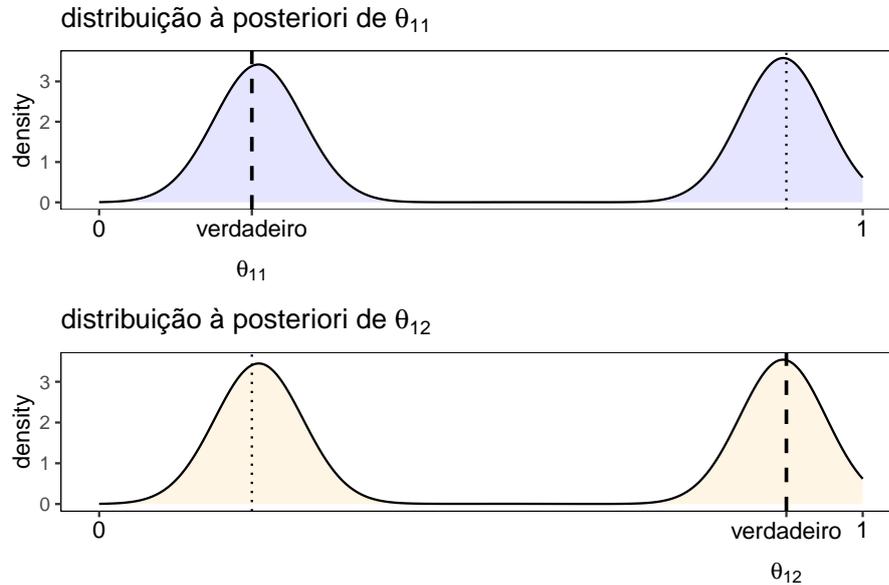
Além disso, um problema adicional surge quando consideramos uma formulação Bayesiana para o G-DINA sem a imposição de restrições.

Foi observado em análises preliminares que, ao ajustar o modelo G-DINA sob a abordagem Bayesiana, sem a inclusão de restrições nos parâmetros dos itens, a distribuição a posteriori de θ_{jl} tende a apresentar bimodalidade (detalhes sobre a formulação Bayesiana considerada serão apresentados na seção 2.5).

O comportamento mencionado anteriormente pode ser facilmente replicado por meio de um estudo de simulação. A Figura 3 apresenta a distribuição a posteriori dos parâmetros θ_{11} e θ_{12} , obtidos em um estudo de simulação preliminar, em que o primeiro item requer apenas 1 atributo (isto é, $K_1 = 1$); θ_{1l} , $l = 1, 2$ representa a probabilidade dos indivíduos do grupo latente l responder o primeiro item corretamente.

Podemos observar na Figura 3 que ambos os parâmetros apresentam distribuição bimodal semelhante, com picos nos verdadeiros valores (isto é, valores simulados) de θ_{12} e θ_{11} . Observe que, nesse caso, não é possível determinar qual pico corresponde a qual parâmetro, não sendo possível obter uma boa estimativa pontual para esses parâmetros, com base nas informações disponíveis. Tal comportamento sugere um problema de identificabilidade relacionado aos parâmetros dos itens.

Figura 3 – Ilustração da bimodalidade na distribuição à posteriori dos parâmetros dos itens do modelo G-DINA quando nenhuma restrição é imposta.



Nota: a linha tracejada representa os verdadeiro valor (isto é, o valor simulado) do parâmetro

Fonte: Elaborada pelo autor.

2.4.2 Propostas de restrições dos parâmetros dos itens para o G-DINA

Uma solução para evitar alguns resultados indesejáveis mencionados anteriormente é incluir restrições sobre os parâmetros estruturais. Nesta subseção, discutiremos três conjuntos de restrições: (a) restrição monotônica (denotado pela sigla RM, ou MC, do termo em inglês, *Monotonic Constraints*) apresentada em [de la Torre \(2011\)](#), (b) restrição da menor probabilidade (denotado pela sigla RMeP, ou LPC, do termo em inglês, *Lower Probability Constraints*) e (c) restrição da maior probabilidade (denotado pela sigla RMaP, ou HPC, do termo em inglês, *Higher Probability Constraints*).

Sejam $\mathbf{0}_{K_j}$ e $\mathbf{1}_{K_j}$ os vetores de dimensão K_j , cujos os elementos são todos iguais a 0 e 1, respectivamente. Seja $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^T$ e $\boldsymbol{\alpha}' = (\alpha'_1, \dots, \alpha'_K)^T$ dois vetores de dimensão K . Considere a relação de ordem parcial entre vetores “ \prec ”, definida de modo que $\boldsymbol{\alpha} \prec \boldsymbol{\alpha}'$ se, somente se, $\alpha_k \leq \alpha'_k$ para todo $k = 1, \dots, K$ e $\alpha_{k_0} < \alpha'_{k_0}$ para algum k_0 .

Dizemos que o item j satisfaz a restrição monotônica se o domínio de um atributo adicional - dentre aqueles requeridos pelo item em questão - implica o aumento ou inalteração da probabilidade de sucesso nesse item. Em termos matemáticos, MC é satisfeita se $\boldsymbol{\alpha}_{(l)}^{*j} \prec \boldsymbol{\alpha}_{(l')}^{*j}$, implica $\theta_{jl} \leq \theta_{j'l'}$ para todo $\theta_{jl}, \theta_{j'l'} \in \mathbf{P}_j$.

Por sua vez, dizemos que o item j satisfaz a restrição de menor probabilidade, se o grupo de indivíduos que não possuem nenhum dos atributos requeridos por esse item tem a menor probabilidade de sucesso nesse item dentre todos os grupos. Em termos matemáticos, sendo

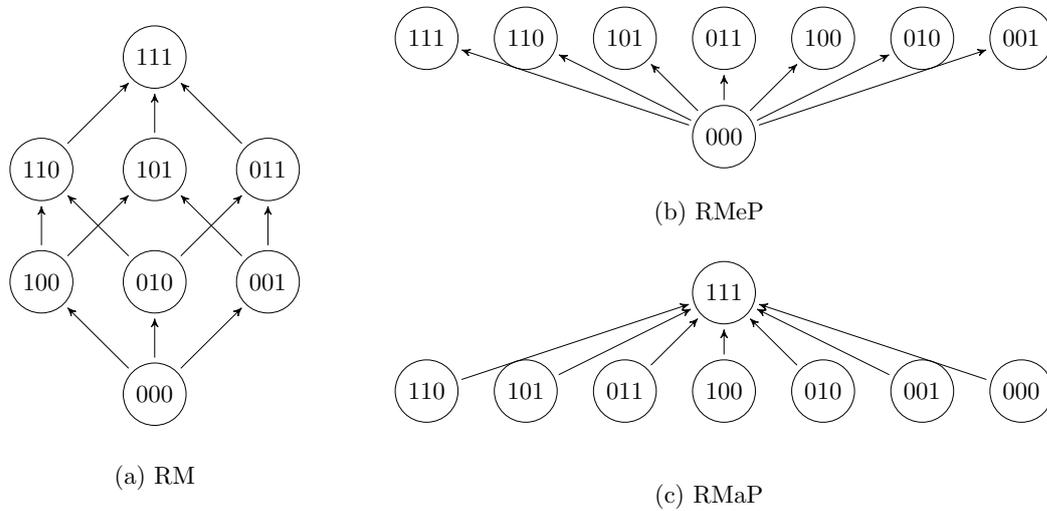
$\theta_{j1} = P(Y_j = 1 | \boldsymbol{\alpha}^{*j} = \mathbf{0}_{K_j})$, RMeP é satisfeita se $\theta_{j1} \leq \theta_{jl}$ para todo $\theta_{jl} \in \mathbf{P}_j$.

De modo similar, dizemos que o item j satisfaz a restrição de maior probabilidade, se o grupo de indivíduos que possui todos os atributos requeridos por esse item tem a maior probabilidade de sucesso nesse item dentre todos os grupos. Em termos matemáticos, sendo $\theta_{jC_j} = P(Y_j = 1 | \boldsymbol{\alpha}^{*j} = \mathbf{1}_{K_j})$, RMaP é satisfeita se $\theta_{jC_j} \geq \theta_{jl}$ para todo $\theta_{jl} \in \mathbf{P}_j$.

Com respeito aos conjuntos de restrições apresentados, dizemos que um modelo satisfaz um determinado conjunto de restrição se, somente se, essas restrições são satisfeitas para todo item $j = 1, \dots, J$. Caso contrário, dizemos que esse modelo viola o referido conjunto de restrições.

A Figura 4 ilustra as relação de ordem que devem ser satisfeitas em cada conjunto de restrição para o caso de um item que requer três atributos. Os grafos orientados (a), (b) e (c) representam, respectivamente, os conjuntos de restrições RM, RMeP e RMaP; os vértices representam os grupos latentes; e a aresta de $\boldsymbol{\alpha}$ para $\boldsymbol{\alpha}'$ indica que a probabilidade associada à $\boldsymbol{\alpha}$ ($P_j^*(\boldsymbol{\alpha})$) é menor que a probabilidade associada à $\boldsymbol{\alpha}'$ ($P_j^*(\boldsymbol{\alpha}')$). Note que o grafo correspondente a RM tem uma complexidade bem maior do que a RMeP e a RMaP, requerendo que uma série de restrições sejam satisfeitas.

Figura 4 – Ilustração das restrições impostas aos parâmetros dos itens com os conjuntos de restrições RM, RMaP e RMeP, para itens com $K_j = 3$



Nota: nos grafos, $\boldsymbol{\alpha} \rightarrow \boldsymbol{\alpha}'$ representa que $P_j^*(\boldsymbol{\alpha}) \leq P_j^*(\boldsymbol{\alpha}')$.

Fonte: Elaborada pelo autor.

Faremos agora algumas considerações sobre os conjuntos de restrições apresentados.

Embora, em muitas aplicações, seja razoável presumir que o domínio de uma nova habilidade ou atributo eleve a probabilidade de um indivíduo responder corretamente aos itens de um teste, conforme instituído na RM, há situações em que essa implicação não é verdadeira. Citando o exemplo dado por [de la Torre \(2011\)](#), considere um teste em que existem itens de múltipla escolha com distratores elaborados para detectar a ausência de um conjunto de

conhecimentos ou habilidades específicas. Tomando um desses itens, pode ocorrer que certos grupos de indivíduos que possuem alguns dos atributos requeridos, mas não todos, selecionem um desses distratores com alta probabilidade, resultando numa probabilidade de sucesso mais baixa do que a do grupo que não possui nenhum desses atributos e escolhe uma alternativa ao acaso.

Com relação à pertinência, destacamos que, apesar da RM evitar resultados contraintuitivos - como aquele citado no início dessa subseção - ela pode ser excessivamente restritiva em algumas situações, além de ser mais complexa de ser implementada. Por outro lado, a RMeP e RMaP são menos restritivas e mais simples de serem implementadas, além de atuarem como um relaxamento da RM, pois o cumprimento da última implica o cumprimento das duas primeiras². Dessa forma, quem considera válido impor a restrição monotônica no modelo G-DINA, em uma determinada aplicação, também deve considerar válido impor as outras duas restrições.

Por fim, é importante ressaltar que em aplicações de dados reais, essas restrições funcionam como hipóteses de modelagem e, como os parâmetros do itens são geralmente desconhecidos, não é possível verificar se tais hipóteses são válidas ou não. Portanto, é mais parcimonioso escolher a hipótese mais relaxada (RMeP ou RMaP) em vez da mais restrita (RM).

Com base nessas considerações, os conjuntos de restrições RMeP e RMaP foram incluídos na formulação Bayesiana do modelo G-DINA apresentado na seção 2.5. Por sua vez, a RM não foi utilizada, devido à sua alta complexidade estrutural, o que demandaria uma implementação customizada de acordo com a matriz Q utilizada.

2.5 Formulação Bayesiana para o modelo G-DINA

Nesta seção apresentaremos uma formulação Bayesiana para a versão saturada do modelo G-DINA com a função de ligação identidade. Nas subseções seguintes serão apresentadas as distribuições a priori adotadas e o modelo hierárquico resultante. Além disso, comentaremos sobre as principais diferenças entre nossa abordagem e outros trabalhos da literatura.

2.5.1 Escolha das distribuições a priori

Consideramos três conjuntos de parâmetros de interesse: os parâmetros estruturais (ou parâmetros dos itens), o perfil latente dos indivíduos, e a distribuição populacional desses perfis.

Os parâmetros estruturais do teste como um todo podem ser representados pelo vetor $\mathbf{P} = (\mathbf{P}_1^T, \mathbf{P}_2^T, \dots, \mathbf{P}_J^T)^T$, conforme definido anteriormente. Considerando a hipótese de independência local assumida na definição do modelo, é razoável assumir que os vetores \mathbf{P}_j , $j = 1, \dots, J$ são independentes. Ademais, na versão saturada do G-DINA, para cada item j , os componentes de

² Note que $\mathbf{0}_{K_j} \prec \boldsymbol{\alpha}_{(I)}^{*j} \prec \mathbf{1}_{K_j}$ para todo $\boldsymbol{\alpha}_{(I)}^{*j}$ diferente de $\mathbf{0}_{K_j}$ e $\mathbf{1}_{K_j}$. Sendo assim, a restrição monotônica implica que $\mathbf{0}_{K_j} = \theta_{j1} \leq \theta_{jl}$ (RMeP) e $\mathbf{1}_{K_j} = \theta_{jCj} \geq \theta_{jl}$ (RMaP) para todo $\theta_{jl} \in \mathbf{P}_j$.

$\mathbf{P}_j = (\theta_{j1}, \dots, \theta_{jl}, \dots, \theta_{jC_j})^T$ são, também, independentes. Sendo assim, podemos atribuir a cada parâmetro θ_{jl} , $j = 1, \dots, J$ e $l = 1, \dots, C_j$ distribuições *Beta*'s independentes. Isto é:

$$\theta_{jl} \sim \text{Beta}(a_{jl}, b_{jl}), \quad j = 1, \dots, J \text{ e } l = 1, \dots, C_j$$

onde $a_{jl} > 0$ e $b_{jl} > 0$ são os hiperparâmetros da variável aleatória θ_{jl} . Neste trabalho, a fim de obter prioris pouco informativas, tomamos $a_{jl} = b_{jl} = 1$ para todo $j = 1, \dots, J$ e $l = 1, \dots, C_j$.

A função densidade de probabilidade do vetor \mathbf{P} pode ser escrita como:

$$P(\mathbf{P}) = \prod_{j=1}^J \prod_{l=1}^{C_j} f_{jl}(\theta_{jl}) = \prod_{j=1}^J \prod_{l=1}^{C_j} \frac{1}{\mathbf{B}(a_{jl}, b_{jl})} \theta_{jl}^{a_{jl}-1} (1 - \theta_{jl})^{b_{jl}-1}$$

em que $f_{jl}(\cdot)$ representa a função de densidade de probabilidade da distribuição *Beta*(a_{jl}, b_{jl}) e $\mathbf{B}(a_{jl}, b_{jl}) = \frac{\Gamma(a_{jl})\Gamma(b_{jl})}{\Gamma(a_{jl} + b_{jl})}$.

Seja \mathbf{A} é uma matriz binária de dimensão $N \times K$, onde a linha $i = 1, \dots, N$ representa o vetor de atributos do indivíduo i , denotado adiante, por $\boldsymbol{\alpha}_i$. Neste trabalho, definimos distribuições para o perfil latente do indivíduo, seguindo a formulação: seja $V = (v_1, \dots, v_N)^T$ um vetor de números inteiros, em que $v_i \in \{1, \dots, C\}$ representa a classe latente a qual o indivíduo i pertence. Isto é, $v_i = c$ indica que o indivíduo i ($= 1, \dots, N$) pertence à c -ésima classe latente ($c = 1, \dots, C$) ou, equivalentemente, que $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{(c)}$. Dessa forma, podemos atribuir uma distribuição categórica para v_i , $j = 1, \dots, N$, sendo essas variáveis conjuntamente independentes.

Seja $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C)^T$ um vetor de dimensão C , em que $\pi_c \geq 0$ e $\sum_{c=1}^C \pi_c = 1$. Nesse contexto, $\boldsymbol{\pi}$ representa a distribuição populacional das categorias latentes, onde π_c , $c = 1, \dots, C$ é a probabilidade a priori de um indivíduo pertencer à classe latente c . Assumindo π_c é desconhecido, podemos considerá-lo um vetor aleatório com distribuição Dirichlet.

Sendo assim, temos:

$$\begin{aligned} v_i | \boldsymbol{\pi} &\sim \text{categórica}(\boldsymbol{\pi}), \text{ for } i = 1, \dots, N \\ \boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \end{aligned}$$

em que $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_C)^T$, com $\lambda_c > 0$ para $c = 1, \dots, C$. $\boldsymbol{\lambda}$ é o (hiper)parâmetro de concentração de $\boldsymbol{\pi}$, cujo valor é definido de acordo com informação a priori disponível sobre $\boldsymbol{\pi}$. Nesse trabalho, a fim de obter uma priori pouco informativa, tomamos $\boldsymbol{\lambda} = \mathbf{1}_C$

Note que, fixado uma indexação para classes latentes, há uma relação biunívoca entre \mathbf{A} e V , sendo $Pr(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{(c)}) \equiv Pr(v_i = c)$ para todo i e todo c . Desse modo, a função de probabilidade de $\mathbf{A} | \boldsymbol{\pi}$ e de densidade de $\boldsymbol{\pi}$ pode ser escrita, respectivamente como:

$$P(\mathbf{A}|\boldsymbol{\pi}) = \prod_{i=1}^N \prod_{c=1}^C \pi_c^{\mathbb{I}(\alpha_i = \alpha_{(c)})}$$

$$P(\boldsymbol{\pi}) = \frac{1}{B(\boldsymbol{\lambda})} \prod_{c=1}^C \pi_c^{\lambda_c - 1}$$

em que $B(\boldsymbol{\lambda}) = \frac{\prod_{c=1}^C \Gamma(\lambda_c)}{\Gamma(\sum_{c=1}^C \lambda_c)}$ é função beta multivariada.

2.5.2 Modelo hierárquico Bayesiano

Com base nas distribuições a priori e função de verosimilhança definidas anteriormente, o modelo G-DINA pode ser representado, sob uma ótica Bayesiana, pela seguinte estrutura hierárquica:

$$Y_{ij} | \alpha_i, \mathbf{P}_j \sim \text{Bernoulli} \left(\theta_{j\xi_{ij}} \right); i = 1, \dots, N; j = 1, \dots, J,$$

em que ξ_{ij} é tal que $\alpha_{\xi_{ij}}^{*j} = \text{FRVA}(\alpha_i; \mathbf{q}_j)$

$$\theta_{jl} \sim \text{Beta}(a_{jl}, b_{jl}) \quad j = 1, \dots, J, \text{ e } l = 1, \dots, C_j \quad (2.7)$$

$$v_i | \boldsymbol{\pi} \sim \text{Categórica}(\boldsymbol{\pi}) \quad i \in \{1, \dots, N\}$$

em que $\alpha_i = \alpha_{(v_i)}$

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\lambda})$$

em que $a_{jl} > 0$, $b_{jl} > 0$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_C)^T$ com $\lambda_c > 0$ para todo $c = 1, \dots, C$ são constantes definidas pelo usuário do modelo. Neste trabalho, adotamos $a_{jl} = b_{jl} = 1$, para $j = 1, \dots, J$ e $l = 1, \dots, C_j$; e $\lambda_c = 1$, para $c = 1, \dots, C$.

Considerando as distribuições definidas na estrutura hierárquica 2.7 e aplicando o teorema de Bayes, temos que:

$$P(\mathbf{A}; \boldsymbol{\pi}; \mathbf{P} | \mathbf{Y}) \propto L(\mathbf{Y} | \mathbf{A}; \mathbf{P}) \times P(\mathbf{A} | \boldsymbol{\pi}) \times P(\boldsymbol{\pi}) \times P(\mathbf{P})$$

dado que \mathbf{P} não depende de \mathbf{A} nem de $\boldsymbol{\pi}$.

Dessa forma, a distribuição a posteriori dos parâmetros \mathbf{A} , \mathbf{P} e $\boldsymbol{\pi}$ no modelo G-DINA é dada pela expressão:

$$p(\mathbf{A}; \mathbf{P}; \boldsymbol{\pi} | \mathbf{Y}) \propto \left[\prod_{j=1}^J \left(\prod_{l=1}^{C_j} \theta_{jl}^{a_{jl}-1} (1 - \theta_{jl})^{b_{jl}-1} \prod_{i:\xi_{ij}=l} \theta_{jl}^{Y_{ij}} (1 - \theta_{jl})^{1-Y_{ij}} \right) \right] \cdot \left[\prod_{c=1}^C \pi_c^{\lambda_c - 1} \prod_{i=1}^N \pi_c^{\mathbb{I}(v_i=c)} \right]$$

Com algumas manipulações algébricas, podemos reescrever a equação anterior como:

$$P(\mathbf{A}; \mathbf{P}; \boldsymbol{\pi} | \mathbf{Y}) \propto \prod_{j=1}^J \prod_{l=1}^{C_j} \theta_{jl}^{a_{jl} + S_{lj} - 1} (1 - \theta_{jl})^{b_{jl} + n_{lj} - S_{lj} - 1} \prod_{c=1}^C \pi_c^{\lambda_c + n_c - 1}$$

em que $n_{lj} = \sum_{i=1}^N \mathbb{I}(\xi_{ij} = l)$, $S_{lj} = \sum_{i=1}^N Y_{ij} \mathbb{I}(\xi_{ij} = l)$, e $n_c = \sum_{i=1}^N \mathbb{I}(v_i = c)$.

Finalmente, o conjunto de restrição RMeP pode ser incluído no modelo G-DINA a partir da equação anterior, resultando na expressão:

$$P(\mathbf{A}; \mathbf{P}; \boldsymbol{\pi} | \mathbf{Y}) \propto \prod_{j=1}^J \prod_{l=1}^{C_j} \theta_{jl}^{a_{jl} + S_{lj} - 1} (1 - \theta_{jl})^{b_{jl} + n_{lj} - S_{lj} - 1} \mathbb{I}(\theta_{jl} \geq \theta_{j1}) \prod_{c=1}^C \pi_c^{\lambda_c + n_c - 1} \quad (2.8)$$

Convém destacar que o fator $\mathbb{I}(\theta_{jl} \geq \theta_{j1})$, na Equação 2.8, representa o conjunto de restrições RMeP. Para incluir o conjunto de restrições RMaP, basta substituir a expressão mencionada por $\mathbb{I}(\theta_{jl} \leq \theta_{jC_j})$.

Não é possível encontrar a constante de proporcionalidade na equação 2.8, seria necessário integrar o lado direito da expressão em relação aos parâmetros \mathbf{A} , \mathbf{P} e $\boldsymbol{\pi}$; porém, não é conhecida uma solução analítica para a integral resultante. Por esse motivo, um algoritmo MCMC é utilizado para obter uma aproximação da distribuição a posteriori aproximada do modelo G-DINA, implementado com o software JAGS (PLUMMER *et al.*, 2016). O código utilizado encontra-se no endereço eletrônico <<https://github.com/renato-fernandes-usp/cdm>>.

2.5.3 Comparação com outras formulações Bayesianas

Um ponto chave do modelo G-DINA é sua capacidade de representar vários MDC por meio do uso de diferentes funções de ligação e restrições nos parâmetros do item. Dessa forma, existem outros MDC que compartilham a mesma FRI do modelo G-DINA. Nesta subseção, discutiremos as principais diferenças entre a abordagem Bayesiana para o modelo G-DINA, apresentada neste capítulo e as existentes na literatura para modelos similares ou equivalentes ao G-DINA.

Essa análise se limita aos modelos saturados, uma vez que os não saturados (por exemplo, DINA e DINO) já diferem do modelo em análise na quantidade de parâmetros. Cumprindo esse requisito, podemos mencionar três trabalhos. Henson, Templin e Willse (2009), Zhan *et al.* (2019) apresentam formulações Bayesianas para o MDCLog, que é equivalente à versão saturada do modelo G-DINA com função de ligação *logit*. Além dessas duas, existe a formulação de Culpepper (2019), que utiliza a função de ligação *probit* e inclui a matriz \mathbf{Q} como um parâmetro a ser estimado, conjuntamente com os demais parâmetros do modelo.

Além da diferença na função de ligação utilizada, esses trabalhos consideram os parâmetros $\boldsymbol{\delta}$ como parâmetro de interesse. Isto é, a formulação utilizada nesses trabalhos atribuem distribuições a priori para $\boldsymbol{\delta}_j = (\delta_{j0}, \delta_{j1}, \dots, \delta_{j12\dots K_j})^T$. Sendo assim, as probabilidades de

sucesso associadas a cada classe latente são obtidas por meio de uma transformação desses parâmetros, de acordo com FRI (Equação 2.3).

Por outro lado, na abordagem apresentada neste capítulo, os parâmetros de interesse são as probabilidades de sucesso e, sendo assim, nenhuma transformação é necessária. Essa última abordagem tem a vantagem de ter uma interpretação mais direta dos parâmetros, uma vez que representam a probabilidade de sucesso nos itens associados a cada classe latente, enquanto a interpretação dos parâmetros nos outros trabalhos depende da função de ligação adotada.

2.6 Estudo de simulação

Nesta seção, apresentaremos um estudo de simulação para avaliar a recuperação dos parâmetros do modelo G-DINA, utilizando a formulação Bayesiana definida na seção anterior. O objetivo deste estudo é verificar se o método de estimativa Bayesiano para o G-DINA tem um bom desempenho na recuperação dos parâmetros ou não. Para isso, a acurácia na recuperação de parâmetros do método de estimativa Bayesiano é comparada com o do método de estimação clássica, que, nesse caso, é utilizado como uma medida base para a acurácia.

2.6.1 Detalhes dos cenários

Quadro 2 – Resumo dos cenários do estudo de simulação do G-DINA

Cenário	J	K	RMaP é satisfeita?	RMeP é satisfeita?	RM é satisfeita?
I	10	3	sim	sim	sim
II	30	5	sim	sim	sim
III	20	4	sim	sim	não
IV	20	4	sim	sim	não

Fonte: Elaborada pelo autor.

Para este estudo, foram considerados quatro cenários de simulação (I a IV), variando a matriz Q (com diferentes números de itens e atributos) e os parâmetros estruturais. As principais características desses cenários estão resumidas no Quadro 2.

Nos cenários I e II, foi utilizado as matrizes Q , apresentadas nas Tabelas 1 e 3, e os parâmetros estruturais, apresentados nas Tabelas 2 e 4, respectivamente. As matrizes Q e parâmetros estruturais apresentados acima foram retirados de dois conjuntos de dados simulados disponíveis no pacote **GDINA** (MA; DE LA TORRE, 2016a) do R, em que as restrições monotônicas são satisfeitas. Os cenários III e IV utilizam, ambos, a matriz Q , apresentada na Tabela 5, e os parâmetros estruturais, apresentados nas Tabelas 6 e 7, respectivamente. Destacamos que, nesses dois últimos cenários, o conjunto de restrições RMeP e RMaP são satisfeitos, mas o RM não.

Tabela 1 – Matriz Q do cenário I do estudo de simulação do G-DINA

item	atributos		
	α_1	α_2	α_3
1	1	0	0
2	0	1	0
3	0	0	1
4	1	0	1
5	0	1	1
6	1	1	0
7	1	0	1
8	1	1	0
9	0	1	1
10	1	1	1

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).

Tabela 2 – Parâmetros estruturais do cenário I do estudo de simulação do G-DINA

item (j)	θ_{j1}	θ_{j2}	θ_{j3}	θ_{j4}	θ_{j5}	θ_{j6}	θ_{j7}	θ_{j8}
	$P_j^*(0)$	$P_j^*(1)$						
	$P_j^*(00)$	$P_j^*(10)$	$P_j^*(01)$	$P_j^*(11)$				
	$P_j^*(000)$	$P_j^*(100)$	$P_j^*(010)$	$P_j^*(001)$	$P_j^*(110)$	$P_j^*(101)$	$P_j^*(011)$	$P_j^*(111)$
1	0.2	0.9	-	-	-	-	-	-
2	0.1	0.8	-	-	-	-	-	-
3	0.1	0.9	-	-	-	-	-	-
4	0.1	0.3	0.5	0.9	-	-	-	-
5	0.1	0.1	0.1	0.8	-	-	-	-
6	0.2	0.9	0.9	0.9	-	-	-	-
7	0.1	0.45	0.45	0.8	-	-	-	-
8	0.1	0.28	0.28	0.8	-	-	-	-
9	0.1	0.4	0.4	0.8	-	-	-	-
10	0.1	0.2	0.3	0.4	0.4	0.5	0.7	0.9

Nota: a primeira linha do cabeçalho representa os parâmetros dos itens na forma θ_{jc} , enquanto a segunda, terceira e quarta linha, representam, respectivamente, as probabilidades correspondente $P_j^*(\alpha^{*j})$ para os itens com K_j igual a 1, 2 e 3.

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).

Tabela 3 – Matriz Q do cenário II do estudo de simulação do G-DINA

item	atributo					item	atributo				
	α_1	α_2	α_3	α_4	α_5		α_1	α_2	α_3	α_4	α_5
1	1	0	0	0	0	16	0	1	0	1	0
2	0	1	0	0	0	17	0	1	0	0	1
3	0	0	1	0	0	18	0	0	1	1	0
4	0	0	0	1	0	19	0	0	1	0	1
5	0	0	0	0	1	20	0	0	0	1	1
6	1	0	0	0	0	21	1	1	1	0	0
7	0	1	0	0	0	22	1	1	0	1	0
8	0	0	1	0	0	23	1	1	0	0	1
9	0	0	0	1	0	24	1	0	1	1	0
10	0	0	0	0	1	25	1	0	1	0	1
11	1	1	0	0	0	26	1	0	0	1	1
12	1	0	1	0	0	27	0	1	1	1	0
13	1	0	0	1	0	28	0	1	1	0	1
14	1	0	0	0	1	29	0	1	0	1	1
15	0	1	1	0	0	30	0	0	1	1	1

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).

Tabela 4 – Parâmetros estruturais do cenário II do estudo de simulação do G-DINA

item (j)	θ_{j1}	θ_{j2}	θ_{j3}	θ_{j4}	θ_{j5}	θ_{j6}	θ_{j7}	θ_{j8}
	$P_j^*(0)$	$P_j^*(1)$						
	$P_j^*(00)$	$P_j^*(10)$	$P_j^*(01)$	$P_j^*(11)$				
	$P_j^*(000)$	$P_j^*(100)$	$P_j^*(010)$	$P_j^*(001)$	$P_j^*(110)$	$P_j^*(101)$	$P_j^*(011)$	$P_j^*(111)$
1 a 5	0.1	0.8	-	-	-	-	-	-
6 a 10	0.2	0.9	-	-	-	-	-	-
11 a 14	0.2	0.5	0.5	0.8	-	-	-	-
16 a 20	0.1	0.1	0.1	0.8	-	-	-	-
21 a 25	0.2	0.8	0.8	0.8	0.8	0.8	0.8	0.8
26 a 27	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8
28 a 30	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.8

Nota: a primeira linha do cabeçalho representa os parâmetros dos itens na forma θ_{jc} , enquanto a segunda, terceira e quarta linha, representam, respectivamente, as probabilidades correspondentes $P_j^*(\alpha^{*j})$ para os itens com K_j igual a 1, 2 e 3.

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).

Tabela 5 – Matriz Q do cenário III e IV do estudo de simulação do G-DINA

item	atributo				item	atributo			
	α_1	α_2	α_3	α_4		α_1	α_2	α_3	α_4
1	1	0	0	0	11	0	0	1	1
2	0	1	0	0	12	1	0	0	1
3	0	0	1	0	13	1	1	1	0
4	0	0	1	1	14	1	1	0	1
5	1	0	0	0	15	1	0	1	1
6	0	1	0	0	16	0	1	1	1
7	0	0	1	0	17	1	1	1	0
8	0	0	1	1	18	1	1	0	1
9	1	1	0	0	19	1	0	1	1
10	0	1	1	0	20	0	1	1	1

Fonte: Elaborada pelo autor.

Tabela 6 – Parâmetros estruturais do cenário III do estudo de simulação do G-DINA

item (j)	θ_{j1}	θ_{j2}	θ_{j3}	θ_{j4}	θ_{j5}	θ_{j6}	θ_{j7}	θ_{j8}
	$P_j^*(0)$	$P_j^*(1)$						
	$P_j^*(00)$	$P_j^*(10)$	$P_j^*(01)$	$P_j^*(11)$				
	$P_j^*(000)$	$P_j^*(100)$	$P_j^*(010)$	$P_j^*(001)$	$P_j^*(110)$	$P_j^*(101)$	$P_j^*(011)$	$P_j^*(111)$
1 a 8	0.2	0.9	-	-	-	-	-	-
9 a 12	0.1	0.5	0.1	0.8	-	-	-	-
13 a 20	0.2	0.6	0.6	0.6	0.3	0.3	0.3	0.9

Nota: a primeira linha do cabeçalho representa os parâmetros dos itens na forma θ_{jc} , enquanto a segunda, terceira e quarta linha, representam, respectivamente, as probabilidades correspondente $P_j^*(\alpha^{*j})$ para os itens com K_j igual a 1, 2 e 3.

Fonte: Elaborada pelo autor.

Tabela 7 – Parâmetros estruturais do cenário IV do estudo de simulação do G-DINA

item (j)	θ_{j1}	θ_{j2}	θ_{j3}	θ_{j4}	θ_{j5}	θ_{j6}	θ_{j7}	θ_{j8}
	$P_j^*(0)$	$P_j^*(1)$						
	$P_j^*(00)$	$P_j^*(10)$	$P_j^*(01)$	$P_j^*(11)$				
	$P_j^*(000)$	$P_j^*(100)$	$P_j^*(010)$	$P_j^*(001)$	$P_j^*(110)$	$P_j^*(101)$	$P_j^*(011)$	$P_j^*(111)$
1 a 8	0.4	0.8	-	-	-	-	-	-
9 a 12	0.3	0.5	0.3	0.8	-	-	-	-
13 a 20	0.2	0.7	0.7	0.7	0.5	0.5	0.5	0.9

Nota: a primeira linha do cabeçalho representa os parâmetros dos itens na forma θ_{jc} , enquanto a segunda, terceira e quarta linha, representam, respectivamente, as probabilidades correspondente $P_j^*(\alpha^{*j})$ para os itens com K_j igual a 1, 2 e 3.

Fonte: Elaborada pelo autor.

Apresentamos agora detalhes sobre a geração dos dados. Para avaliar o impacto do tamanho da amostra na recuperação dos parâmetros, simulamos, para cada cenário, testes com 50, 200 e 1000 respondentes. Para cada cenário e tamanho de amostra, os vetores de atributos foram gerados, associando uma classe latente para cada indivíduo de forma aleatória, com igual probabilidade para cada classe. Feito isso, foram gerados 100 replicações para cada cenário e tamanho de amostra, mantendo-se fixos todos os parâmetros do modelo.

Para cada conjunto de dados gerados, estimamos os parâmetros do modelo G-DINA, utilizando os métodos de estimação Bayesiano e clássico. Para entender melhor a influência das restrições nos parâmetros estruturais, consideramos, para a abordagem Bayesiana, as estimativas obtidas com a imposição do conjunto de restrições RMeP e RMaP.

Para a abordagem clássica, consideramos as estimativas obtidas com e sem a imposição da restrição monotônica. É importante destacar que essas são as duas únicas opções de restrições nos parâmetros estruturais disponíveis nos pacotes para estimação do G-DINA do R, **CDM** (ROBITZSCH *et al.*, 2020) e **GDINA** (MA; DE LA TORRE, 2016a).

2.6.2 Detalhes dos métodos de estimação

Para a estimação Bayesiana, a distribuição posterior aproximada dos parâmetros \mathbf{P} , \mathbf{A} e $\boldsymbol{\pi}$ foram obtidas com um algoritmo MCMC implementado com o software *JAGS* (PLUMMER *et al.*, 2016). No MCMC, foram utilizadas duas cadeias com 10000 iterações cada, sendo primeiras 5000 iterações de cada cadeia descartadas. Não foi utilizado *Thinning*. Sendo assim, foram obtidos 10000 valores de amostragem para cada parâmetro.

A convergência do MCMC foi verificada utilizando a estatística de Gelman e Rubin \hat{R} (GELMAN; RUBIN, 1992) calculado através do algoritmo proposto por Vehtari *et al.* (2021). Segundo Brooks e Gelman (1998), valores de \hat{R} próximo de 1 indicam que a distribuição das cadeias MCMC estão próximas da distribuição-alvo. O valor de corte utilizado usualmente como indicativo de convergência é 1.1 (ZHAN *et al.*, 2019). Neste estudo, todos os valores de \hat{R} são menores que 1,05 (com 99% dos valores menores que 1,008), indicando que todas as simulações convergiram.

O método de estimação clássica é baseado no estimador de máxima verosimilhança marginalizada (MMLE). Por não ser possível determinar o MMLE algebricamente, uma solução numérica aproximada é obtida utilizando um algoritmo *Expectation-Maximization* (EM). Nesse estudo de simulação, as estimativas para abordagem clássica foram obtidas com o pacote **GDINA** (MA; DE LA TORRE, 2020) do R.

Nesse trabalho, a acurácia na recuperação dos parâmetros foi calculada a partir dos estimadores pontuais de cada parâmetros. Procedemos de tal forma com o propósito de viabilizar a utilização das mesmas medidas em todos os métodos de estimação.

No método de estimação Bayesiano, a média amostral foi utilizada como estimador

pontual de ponto dos parâmetros de itens θ_{jl} , $j = 1, \dots, J$, $l = 1, \dots, C_j$; enquanto a moda amostral foi utilizada para estimar a classe latente de cada indivíduo, por se tratar de uma categórica. Já no método clássico, os parâmetros dos itens foram obtidos a partir da aproximação do MMLE mencionado anteriormente, e os vetores de atributos dos indivíduos por meio do estimador *Maximum a Posteriori*. Convém ressaltar que essas são as estimativas reportadas por padrão pelo pacote **GDINA** (MA; DE LA TORRE, 2016a).

Para avaliar a acurácia da recuperação dos parâmetros estruturais, consideramos a raiz do erro quadrático médio (RMSE) de cada parâmetro θ_{jl} . Já para a acurácia dos atributos latentes dos indivíduos, consideramos o ACCR (*Attributes' Correct Classification Ratio*, ou taxa de classificação correta dos atributos, em tradução livre) e o PCCR (*Patterns' Correct Classification Ratio*, ou taxa de classificação correta dos padrões, em tradução livre) Song *et al.* (2012) que medem, respectivamente, a proporção que cada componente e cada vetor α_i foram estimados corretamente. Para facilitar a visualização dos resultados, são reportados apenas os valores médios do RMSE, ACCR e PCCR obtidos para cada cenário e tamanho de amostra, calculados por meio das seguintes fórmulas:

$$\begin{aligned} \text{RMSE}(\theta_{jl}) &= \sqrt{\frac{1}{R} \sum_{r=1}^R (\widehat{\theta}_{jl}^{(r)} - \theta_{jl})^2} \\ \overline{\text{RMSE}} &= \frac{1}{\sum_{j=1}^J C_j} \sum_{j=1}^J \sum_{l=1}^{C_j} \text{RMSE}(\theta_{jl}) \\ \text{PCCR}^{(r)} &= \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\widehat{\alpha}_i^{(r)} = \alpha_i) \\ \overline{\text{PCCR}} &= \frac{1}{R} \sum_{r=1}^R \text{PCCR}^{(r)} \\ \text{ACCR}^{(r)} &= \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(\widehat{\alpha}_{ik}^{(r)} = \alpha_{ik}) \\ \overline{\text{ACCR}} &= \frac{1}{R} \sum_{r=1}^R \text{ACCR}^{(r)} \end{aligned}$$

em que $R = 100$ é o número total de replicações; $r = 1, \dots, R$ é o índice que indica a replicação considerada na avaliação da fórmula; $\mathbb{I}(\cdot)$ representa a função indicadora; $\widehat{\theta}_{jl}^{(r)}$, $\widehat{\alpha}_i^{(r)}$ e $\widehat{\alpha}_{ik}^{(r)}$ são, respectivamente, as estimativas pontuais obtidas na r -ésima replicação do l -ésimo parâmetro do item j , θ_{jl} ; do vetor de atributos do indivíduo i , α_i ; e do k -ésimo componente do vetor mencionado anteriormente, α_{ik} .

2.6.3 Resultados

Nesta subseção, são apresentados os resultados referentes a acurácia na recuperação dos parâmetros do método de estimação Bayesiano, com os conjuntos de restrições RMeP e RMaP; e do método de estimação clássica, sem a imposição de restrições e com a imposição do RM

(detalhes dessa implementação podem ser encontrados em [Hong, Chang e Tsai \(2016\)](#) e [Ma e de la Torre \(2020\)](#)).

É importante destacar que, nos cenários III e IV, são reportados os resultados obtidos com a imposição da restrição monotônica no método clássico, apesar de sabermos que esse conjunto de restrições não são satisfeitas nesses cenários. Procedemos dessa forma para avaliar o impacto de, equivocadamente, incluir a restrição monotônica no processo de estimação, quando essa não é satisfeita pelo modelo que gerou os dados.

A Tabela 8 apresenta os resultados referentes à recuperação dos parâmetros estruturais, em termos do RMSE médio.

Tabela 8 – RMSE médio obtidos pelos diferentes métodos de estimação no estudo de simulação do G-DINA

Cenário	N	Clássico		Clássico com RM		Bayesiano com RMaP		Bayesiano com RMeP	
I	50	0.199	(0.041)	0.152	(0.021)	0.126	(0.019)	0.148	(0.022)
	200	0.097	(0.018)	0.087	(0.014)	0.080	(0.012)	0.079	(0.013)
	1000	0.040	(0.007)	0.038	(0.006)	0.038	(0.007)	0.038	(0.007)
II	50	0.195	(0.018)	0.144	(0.013)	0.136	(0.010)	0.134	(0.010)
	200	0.095	(0.008)	0.076	(0.006)	0.080	(0.006)	0.078	(0.005)
	1000	0.037	(0.002)	0.032	(0.002)	0.036	(0.002)	0.037	(0.002)
III	50	0.209	(0.022)	0.202	(0.021)	0.131	(0.015)	0.142	(0.013)
	200	0.103	(0.012)	0.132	(0.006)	0.086	(0.009)	0.090	(0.031)
	1000	0.042	(0.004)	0.110	(0.001)	0.041	(0.004)	0.039	(0.004)
IV	50	0.303	(0.027)	0.256	(0.024)	0.156	(0.010)	0.151	(0.010)
	200	0.212	(0.029)	0.197	(0.026)	0.121	(0.018)	0.119	(0.014)
	1000	0.074	(0.009)	0.096	(0.007)	0.063	(0.006)	0.064	(0.006)

Nota: os valores entre parênteses denotam o desvio padrão.

Fonte: Dados da pesquisa.

Podemos observar, na Tabela 8, que, de forma geral, o método de estimação Bayesiano (com ambos tipo de restrição) tem acurácia similar ou superior ao método clássico sem a imposição de restrições, sendo a diferença de acurácia em favor do método Bayesiano mais evidente para os menores tamanhos de amostra. Ademais, na comparação com o método clássico com a inclusão da restrição monotônica, o método Bayesiano apresenta acurácia similar ao primeiro nos cenários I e II, e superior nos cenários III e IV.

Ainda sobre essa tabela, comparando o RMSE médio para o mesmo tamanho de amostra entre os diferentes cenários, nota-se que a acurácia do método Bayesiano apresenta pouca variação (independentemente da restrição incluída), enquanto o método clássico apresenta variações consideravelmente maiores. Esse resultado indica que o método de estimação Bayesiano é menos sensível à variações nas especificações do teste, tais como o número de itens e os valores dos parâmetros estruturais.

A Tabela 9 apresenta os resultados da recuperação dos atributos latentes dos indivíduos em termos do ACCR médio.

Tabela 9 – ACCR médio obtido pelos diferentes métodos de estimação no estudo de simulação do G-DINA

Cenário	N	Clássico	Clássico com RM	Bayesiano com RMaP	Bayesiano com RMeP
I	50	86.1 % (3.7 %)	86.6 % (3.5 %)	86.1 % (3.6 %)	85.2 % (3.9 %)
	200	88.5 % (1.8 %)	88.6 % (1.7 %)	88.9 % (1.5 %)	89.0 % (1.4 %)
	1000	90.1 % (0.6 %)	90.1 % (0.6 %)	90.1 % (0.6 %)	90.1 % (0.6 %)
II	50	91.6 % (2.1 %)	91.9 % (2.1 %)	91.9 % (2.0 %)	91.7 % (1.9 %)
	200	93.3 % (1.0 %)	93.8 % (1.0 %)	93.8 % (0.9 %)	93.9 % (1.0 %)
	1000	94.6 % (0.3 %)	94.6 % (0.3 %)	94.6 % (0.3 %)	94.6 % (0.3 %)
III	50	90.3 % (2.8 %)	89.1 % (3.0 %)	91.2 % (2.7 %)	90.8 % (2.5 %)
	200	93.0 % (1.3 %)	92.4 % (1.3 %)	93.4 % (1.2 %)	92.8 % (2.8 %)
	1000	94.4 % (0.4 %)	93.6 % (0.4 %)	94.4 % (0.4 %)	94.4 % (0.5 %)
IV	50	71.7 % (3.7 %)	71.2 % (3.9 %)	72.6 % (3.9 %)	72.7 % (3.9 %)
	200	73.2 % (2.8 %)	73.0 % (2.9 %)	76.3 % (2.7 %)	76.1 % (2.2 %)
	1000	78.3 % (1.0 %)	77.9 % (1.2 %)	78.6 % (0.9 %)	78.5 % (0.9 %)

Nota: os valores entre parênteses denotam o desvio padrão.

Fonte: Dados da pesquisa.

Podemos notar, na Tabela 9, que não há diferença relevante entre os métodos avaliados na recuperação com atributos latentes. Além disso, os resultados sugerem que essa medida é mais afetada pelos valores dos parâmetros dos itens, do que pelo número de itens ou tamanho da amostra.

A Tabela 10 apresenta os resultados da recuperação do perfil latente dos indivíduos em termos do PCCR médio.

Podemos observar, na Tabela 9, que, comparado ao método de estimação clássico sem a inclusão de restrições, o método de estimação Bayesiano apresenta maior acurácia nesse quesito no cenário III, para $N = 50$ e, no cenário IV, para $N = 50$ e $N = 200$, independentemente do conjunto de restrição utilizado; nos demais casos, a acurácia de ambos os métodos é similar. Por sua vez, quando comparado ao método de estimação clássico, com a inclusão da restrição monotônica, o método de estimação Bayesiano apresenta acurácia similar ao primeiro nos cenários I e II, porém, consideravelmente superior nos cenários III e IV.

A seguir, destacamos algumas descobertas advindas da análise em conjunto das três tabelas mencionadas.

Primeiro, com relação ao conjunto de restrições incluídas no método de estimação Bayesiano, notamos que não há diferenças relevantes de acurácia na recuperação dos parâmetros, quando comparados os resultados obtidos com a inclusão do conjunto RMeP ou do RMaP. Em todos os quesitos, a diferença de resultados entre ambos é menor que uma unidade de desvio

Tabela 10 – PCCR médio obtido pelos diferentes métodos de estimação no estudo de simulação do G-DINA

Cenário	N	Clássico	Clássico com RM	Bayesiano com RMaP	Bayesiano com RMeP
I	50	65.1 % (8.6 %)	66.7 % (7.7 %)	64.9 % (8.4 %)	62.1 % (9.6 %)
	200	71.0 % (4.3 %)	71.2 % (4.2 %)	71.7 % (3.6 %)	72.0 % (3.5 %)
	1000	74.9 % (1.4 %)	75.0 % (1.4 %)	74.9 % (1.4 %)	74.8 % (1.4 %)
II	50	66.9 % (7.1 %)	67.7 % (7.6 %)	68.2 % (6.8 %)	67.1 % (7.1 %)
	200	73.2 % (3.8 %)	74.5 % (3.7 %)	75.2 % (3.3 %)	75.1 % (3.7 %)
	1000	77.8 % (1.2 %)	78.0 % (1.2 %)	77.9 % (1.2 %)	77.8 % (1.2 %)
III	50	71.5 % (8.2 %)	65.1 % (8.7 %)	74.0 % (8.2 %)	72.8 % (7.4 %)
	200	79.2 % (3.5 %)	75.5 % (3.8 %)	80.3 % (3.2 %)	78.5 % (9.3 %)
	1000	83.3 % (1.2 %)	79.5 % (1.3 %)	83.4 % (1.2 %)	83.4 % (1.2 %)
IV	50	30.2 % (8.5 %)	27.3 % (7.9 %)	35.9 % (7.4 %)	34.4 % (8.3 %)
	200	37.3 % (6.7 %)	31.5 % (6.5 %)	45.3 % (5.7 %)	43.7 % (5.1 %)
	1000	50.5 % (2.0 %)	45.8 % (2.7 %)	51.3 % (1.8 %)	51.1 % (1.9 %)

Nota: os valores entre parênteses denotam o desvio padrão.

Fonte: Dados da pesquisa.

padrão observado da medida em questão.

O segundo ponto refere-se à inclusão da restrição monotônica no método de estimação clássico. Podemos observar que o impacto positivo (em termos de incremento na acurácia) obtido com a inclusão desse conjunto de restrição, quando esse é compatível, é muito menos significativo do que o impacto negativo obtido com a inclusão inadequada do mesmo, isto é, quando as hipóteses do conjunto de restrições RM, não são condizentes com parâmetros reais dos itens do teste. Sendo assim, é aconselhável empregar as restrições monotônicas com precaução.

Outro ponto de destaque é o custo computacional dos métodos avaliados. A Tabela 11 apresenta o tempo médio (em segundos) necessário para executar o algoritmo de estimação em cada teste simulado com cada método.

Podemos observar na Tabela 11, que, conforme esperado, o método de estimação Bayesiano demanda mais tempo para ser executado do que o método clássico. Isso ocorre devido ao processo de formação das cadeias de Markov, que envolve a geração de amostras aleatórias de todos os parâmetros do modelo. Consequentemente, o tempo de execução do método Bayesiano aumenta conforme o número de itens, atributos e indivíduos aumenta, como pode ser observado na tabela mencionada.

Por fim, baseado nos resultados desse estudo de simulação, podemos concluir que o método de estimação Bayesiano para o G-DINA é confiável e possui acurácia igual ou superior ao método de estimação clássico.

Tabela 11 – Tempo médio, em segundos, para a execução dos algoritmos de estimação do G-DINA

Cenário	N	Clássico	Clássico com RM	Bayesiano com RMaP	Bayesiano com RMeP
I	50	0.08	2.22	42.32	42.63
	200	0.13	3.09	192.29	192.03
	1000	0.09	1.78	1215.83	1211.37
II	50	0.11	3.49	248.90	249.07
	200	0.29	5.34	1112.10	1101.72
	1000	0.24	1.87	7091.60	7094.73
III	50	0.08	2.86	126.25	125.33
	200	0.13	3.51	571.07	567.77
	1000	0.12	1.38	3773.05	3761.72
IV	50	0.11	4.67	138.68	138.00
	200	0.43	13.37	592.44	591.99
	1000	0.59	9.89	3710.62	3736.05

Fonte: Dados da pesquisa.

2.7 Discussão

A principal contribuição desse capítulo é a apresentação de uma abordagem Bayesiana para o modelo G-DINA, com inclusão de dois novos conjuntos de restrições para os parâmetros dos itens, denominados *restrição da menor probabilidade* (RMeP) e *restrição da maior probabilidade* (RMaP). Para realizar a estimação dos parâmetros, foi apresentado a formulação de um modelo hierárquico para o G-DINA e desenvolvido uma implementação computacional do método de estimação utilizando o software *JAGS*.

O estudo de simulação apresentado na seção 2.6 demonstrou que o método de estimação Bayesiano, com ambos os conjuntos de restrições propostos, recuperam adequadamente todos os parâmetros com acurácia comparável ao método de estimação clássico, sendo mais preciso que esse último em algumas situações, como nos casos com tamanho de amostras pequeno. Além disso, os resultados desse estudo de simulação sugerem que é mais seguro impor restrições menos restritivas (isto é, utilizar o RMeP ou o RMaP em vez do RM), devido ao impacto negativo sobre a acurácia que ocorre quando restrições mais rigorosas são utilizadas inadequadamente.

UMA ABORDAGEM BAYESIANA PARA O MODELO MGDPCP

3.1 Introdução

Os modelos de diagnóstico cognitivo mais conhecidos, como o DINA ([HAERTEL, 1989](#)), o DINO ([TEMPLIN; HENSON, 2006](#)), o modelo de diagnóstico cognitivo aditivo (MDCA) ([DE LA TORRE, 2011](#)), o modelo de diagnóstico cognitivo log-linear ([HENSON; TEMPLIN; WILLSE, 2009](#)) e G-DINA ([DE LA TORRE, 2011](#)), são projetados para lidar apenas com itens de respostas dicotômicas.

Quando um teste é formado por itens com outros tipos de respostas; tais como itens avaliados em diferentes níveis, questões de múltipla escolha, entre outros; o procedimento usual para utilizar os MDC, nesses testes, é dicotomizar tais respostas. Embora útil, esse procedimento resulta em uma perda de informação, que afeta a classificação ou diagnóstico do respondente ([MA; DE LA TORRE, 2016b](#)).

Poucos modelos de MDC são projetados para abranger respostas politômicas. Dentre eles, podemos citar: o DINA de crédito parcial (*Partial Credit DINA*; [de la Torre, 2010](#), apud [Ma e de la Torre, 2016b](#)), o MGPC para respostas graduadas (*GDM for graded responses*; [von Davier, 2008](#)), o G-DINA sequencial (*sequential G-DINA*; [Ma e de la Torre, 2016b](#)) e o modelo geral de diagnóstico cognitivo para respostas politômicas (*General Polytomous Diagnosis Model*; [Chen e de la Torre, 2018](#)).

O enfoque desse capítulo é o modelo geral de diagnóstico cognitivo para respostas politômicas (MGDCP). A formulação desse modelo é baseada na combinação do processo de modelagem do modelo G-DINA para respostas dicotômicas com o processo de subdivisão dos itens utilizados nos modelos de respostas graduadas ([CHEN; DE LA TORRE, 2018](#)). Mais precisamente, em cada item, os diferentes níveis de respostas são tratados como subitens

dicotômicos, que são modelados de acordo com a FRI do modelo G-DINA.

Esse modelo foi selecionado entre os demais, devido às características relevantes que possui. Ele pode lidar com respostas ordinais e categóricas; possui uma formulação geral capaz de abranger vários sub-modelos como casos particulares; não exige que, em cada item, seja determinado o rol de atributos requeridos para cada nível de resposta (como ocorre no G-DINA sequencial, por exemplo).

Neste capítulo, apresentaremos uma formulação Bayesiana para o MGDCP com a inclusão de uma nova proposta de restrição dos parâmetros, baseada na escolha de hiperparâmetros, projetada para melhorar a interpretação do modelo e a estimativa de parâmetros. É apresentado, também, um estudo de simulação, desenvolvido para avaliar a recuperação dos parâmetros do método de estimação Bayesiano baseado na formulação proposta.

3.2 O modelo MGDCP

Por ter sua formulação baseada no modelo G-DINA, há muita similaridade entre o MGDCP e o referido modelo. Desse modo, diversas definições são iguais ou muito semelhantes.

Para definir o modelo MGDCP, considere, assim como anteriormente, um teste formado por J itens que avaliam K atributos e respondido por N indivíduos.

Considere que as respostas de cada item $j = 1, \dots, J$ são classificadas em H_j níveis. Aqui, por simplicidade de notação e sem perda de generalidade, os níveis de respostas serão representados por números inteiros de 1 a H_j . Seja Y a matriz de resposta do teste, uma matriz de valores observados de dimensão $N \times J$, onde o elemento y_{ij} , com $i = 1, \dots, N$ e $j = 1, \dots, J$, representa a resposta do indivíduo i para o item J . Ainda, seja Y_j a variável aleatória que representa a resposta do item j , cujo conjunto suporte é $\{1, \dots, H_j\}$.

Seja Q uma matriz de dimensão $J \times K$ com elementos $q_{jk} \in \{0, 1\}$, com $j = 1, \dots, J$ e $k = 1, \dots, K$, onde $q_{jk} = 1$, se o atributo k é requerido pelo item j ; e $q_{jk} = 0$, caso contrário. Nesse contexto, dizemos que o atributo k é requerido pelo item j se a posse desse atributo influencia na resposta do item j ; isto é, se a distribuição de probabilidade de Y_j é afetada pela presença ou não do atributo k .

Ademais, todas as variáveis relacionadas aos atributos latentes seguem a mesma definição utilizada no G-DINA. Isso se aplica aos vetores de atributos do indivíduo i , α_i , às classes latentes $\alpha_{(c)}$, aos vetores reduzidos de atributos α^{*j} , aos grupos latentes $\alpha_{(l)}^{*j}$ (a definição dessas variáveis pode ser encontrada no capítulo 2). Do mesmo modo, temos as constantes $C = 2^K$, $K_j = \sum_{k=1}^K q_{jk}$ e $C_j = 2^{K_j}$ que seguem com a mesma definição.

3.2.1 Função de resposta ao item do MGDCP

No MGDCP, cada item é subdividido em H_j subitens dicotômicos, sendo cada um desses sub-itens modelados, seguindo a formulação do modelo G-DINA. Seja $P_{jh}(\boldsymbol{\alpha}_{(l)}^{*j}) := Pr(Y_j = h | \boldsymbol{\alpha}^{*j} = \boldsymbol{\alpha}_{(l)}^{*j})$, com $j = 1, \dots, J$, $l = 1, \dots, C_j$ e $h = 1, \dots, H_j$, a probabilidade de um indivíduo dar uma resposta de nível h ($= 1, \dots, H_j$) para o item j ($= 1, \dots, J$), dado que esse indivíduo pertence ao grupo latente l ($= 1, \dots, C_j$) do referido item. É importante ressaltar que as probabilidades são restringidas de modo a satisfazer a condição $\sum_{h=1}^{H_j} P_{jh}(\boldsymbol{\alpha}_{(l)}^{*j}) = 1$ para todo j e l .

A função de resposta ao item (FRI) do MGDCP para o item j e nível de resposta h é dada pela Equação 3.1

$$g\left(P_{jh}(\boldsymbol{\alpha}_{(l)}^{*j})\right) := \delta_{jh0} + \sum_{k=1}^{K_j} \delta_{jhk} \alpha_{lk}^* + \sum_{k'=k+1}^{K_j} \sum_{k=1}^{K_j-1} \delta_{jhkk'} \alpha_{lk}^* \alpha_{lk'}^* + \dots + \delta_{jh12\dots K_j} \prod_{k=1}^{K_j} \alpha_{lk}^* \quad (3.1)$$

em que $g(\cdot)$ é uma função de ligação e δ_{jh} são os parâmetros do item j para a resposta do nível h , sendo que δ_{jh0} , δ_{jhk} , $\delta_{jhkk'}$ representam, respectivamente, o intercepto, o efeito principal do k -ésimo atributo exigido e o efeito de interação entre o k -ésimo e o k' -ésimo atributo exigido, referentes à resposta de nível h do item j . A interpretação desses parâmetros é similar ao modelo G-DINA (ver p. 2.3.1), porém atribuída agora ao nível de resposta h do item j , em vez de apenas ao item j . Analogamente, as escolhas usuais para função de ligação são *identidade*, *log* ou *logit* (CHEN; DE LA TORRE, 2018).

Seja $\theta_{jlh} := P_{jh}(\boldsymbol{\alpha}_{(l)}^{*j})$ a probabilidade da resposta h no item j associada aos indivíduos do grupo l do referido item. Para cada item $j = 1, \dots, J$ e cada nível de resposta $h = 1, \dots, H_j$, é possível estabelecer uma bijeção entre os vetores $\mathbf{P}_{jh} = (\theta_{j1h}, \dots, \theta_{jC_jh})^T$ e $\boldsymbol{\delta}_{jh} = (\delta_{jh0}, \delta_{jh1}, \dots, \delta_{jh12\dots K_j})^T$, por meio da equação $g(\mathbf{P}_{jh}) = M_j \boldsymbol{\delta}_{jh}$, em que M_j é a matriz saturada de *design* (veja p.38). Sendo assim, a parametrização, baseada nas probabilidades θ , será utilizada no método de estimação clássico e Bayesiano do MGDCP, apresentados mais adiante.

O método de estimação clássico, proposto por Chen e de la Torre (2018), utiliza um algoritmo *Expectation - Maximization* (EM) para obter uma aproximação do MMLL. Convém ressaltar que o algoritmo de estimação do trabalho original foi desenvolvido para o software **Ox** e não está disponível para uso em nenhum local acessível. Por esse motivo, nas aplicações em que a estimação clássica do MGDCP é utilizada, é empregado uma implementação em R (R Core Team, 2021) do algoritmo citado anteriormente, desenvolvida pelo autor desta tese, seguindo os procedimentos apresentados por Chen e de la Torre (2018).

3.2.2 Função de verossimilhança do MGDCP

Seja $\boldsymbol{\theta}_{jl} = (\theta_{jl1}, \dots, \theta_{jlH_j})^T$ o vetor de dimensão H_j formado pelas probabilidades dos diferentes níveis de resposta do item j associadas ao grupo latente l do referido item. Seja $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{11}^T, \dots, \boldsymbol{\theta}_{1C_1}^T, \dots, \boldsymbol{\theta}_{J1}^T, \dots, \boldsymbol{\theta}_{JC_J}^T \right)^T$ o vetor de dimensão $\sum_{j=1}^J C_j H_j$ formado pela justaposição dos vetores de parâmetros $\boldsymbol{\theta}_{jl}$, para $j = 1, \dots, J$ e $l = 1, \dots, C_j$.

Seja $v_i \in \{1, \dots, C\}$ em que $v_i = c$ indica que o indivíduo i ($= 1, \dots, N$) pertence à c -ésima classe latente ($c = 1, \dots, C$) (ou seja, $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{(c)}$). Seja $\xi_{ij} \in \{1, \dots, C_j\}$ a variável que indica o grupo latente ao qual o indivíduo i pertence, sendo que $\xi_{ij} = l$ indica que $\boldsymbol{\alpha}_i^{*j} = \boldsymbol{\alpha}_{(l)}^{*j}$. Equivalentemente, x_{ij} pode ser definido como sendo o índice tal que $\boldsymbol{\alpha}_{(x_{ij})}^{*j} = \text{FRVA}(\boldsymbol{\alpha}_i, \mathbf{q}_j)$.

Nesses termos, a função de verossimilhança do modelo MGDCP é dada pela Equação 3.2

$$L = P(\mathbf{Y}|\mathbf{A}, \boldsymbol{\theta}) = \prod_{i=1}^N \prod_{j=1}^J \prod_{l=1}^{C_j} \prod_{h=1}^{H_j} \theta_{jhl}^{\mathbb{I}(\xi_{ji}=l)\mathbb{I}(Y_{ij}=h)} \quad (3.2)$$

3.2.3 Conjuntos de restrições dos parâmetros para modelos de respostas politômicas

O modelo MGDCP é uma extensão do modelo G-DINA para respostas politômicas e apresenta os mesmos problemas que ocorrem com o último, devido à falta de restrição nos parâmetros dos itens, conforme descritos na seção 2.4. Sendo assim, a imposição de restrições nos parâmetros dos itens é importante para a estimação e interpretabilidade dos parâmetros.

Na seção 2.4, foram definidos três conjuntos de restrições: monotônica (RM), da menor probabilidade (RMeP) e da maior probabilidade (RMaP). No entanto, tais definições são restritas aos modelos com respostas dicotômicas. Por esse motivo, é necessário apresentar extensões dessas definições para modelos com respostas politômicas.

Para itens de resposta dicotômica, RM é definido de modo que $\boldsymbol{\alpha} \prec \boldsymbol{\alpha}'$ implica $Pr(Y_j = 1 | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}) \leq Pr(Y_j = 1 | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}')$, onde \prec é relação de ordem entre vetores definida de modo que $(\alpha_1, \dots, \alpha_K) \prec (\alpha'_1, \dots, \alpha'_K)$ se, somente se, $\alpha_k \leq \alpha'_k$ para todo $k = 1, \dots, K$, e $\alpha_{k_0} < \alpha'_{k_0}$ para algum $k_0 \in \{1, \dots, K\}$. [Chen e de la Torre \(2018\)](#) expandem o conceito da restrição monotônica para respostas politômicas com a seguinte definição:

A restrição monotônica (RM) é satisfeita para o item j , se e somente se, $\boldsymbol{\alpha} \prec \boldsymbol{\alpha}'$ implica $Pr(Y_j > h | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}) \leq Pr(Y_j > h | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}')$ para toda resposta $h = 1, \dots, H_j$.

Baseado nessa definição, o conjuntos de restrições RMeP e RMaP podem ser estendidos para itens de resposta politômica, como segue:

A restrição de menor probabilidade (RMeP) é satisfeita para o item j , se e somente se, $Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \mathbf{0}) \leq Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \boldsymbol{\alpha}_{(l)}^{*j})$ para todo $h = 1, \dots, H_j$.

De modo similar, a restrição de maior probabilidade (RMaP) é satisfeita para o item j , se e somente se, $Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \mathbf{a}) \geq Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \boldsymbol{\alpha}_{(l)}^{*j})$ para todo $h = 1, \dots, H_j$.

Assim, podemos definir o conjunto de restrições de maior e menor probabilidade (denotado doravante por RMaMeP, ou, equivalentemente, por HLPC, do termo em inglês *Higher and Lower Probability Constraints*) como o conjunto de restrições, no qual RMaP e RMeP devem ser simultaneamente satisfeitas. Isto é, a restrição RMaMeP é satisfeita para o item j , se e somente se, $Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \mathbf{1}) \leq Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \boldsymbol{\alpha}_{(l)}^{*j}) \leq Pr(Y_j > h | \boldsymbol{\alpha}^{*j} = \mathbf{0})$ para todo $h = 1, \dots, H_j$.

Por fim, dizemos que o modelo MGDCP satisfaz alguns desses conjuntos de restrições definidos anteriormente, se as restrições desse conjunto são satisfeitas para os itens $j = 1, \dots, J$.

3.3 Formulação Bayesiana para o MGDCP

Nesta seção apresentamos uma formulação Bayesiana para o modelo MGDCP.

Considerando distribuições a priori para os parâmetros dos itens θ_{jlh} , para $j = 1, \dots, J$, $l = 1, \dots, C_j$ e $h = 1, \dots, H_j$; para as classes latentes dos indivíduos, representadas por v_i , para $i = 1, \dots, N$ e para o vetor de proporções das classes latentes $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C)^T$; o MGDCP pode ser representado na abordagem Bayesiana, por meio da seguinte estrutura hierárquica:

$$\begin{aligned}
 Y_{ij} | \boldsymbol{\alpha}_i, \boldsymbol{\theta} &\sim \text{Categórica}(\boldsymbol{\theta}_{j\xi_{ij}}) \quad i = 1, \dots, N; j = 1, \dots, J, \\
 \text{em que } \xi_{ij} &\text{ é tal que } \quad \boldsymbol{\alpha}_{\xi_{ij}}^{*j} = \text{FRVA}(\boldsymbol{\alpha}_i; \mathbf{q}_j) \\
 \boldsymbol{\theta}_{jl} &\sim \text{Dirichlet}(\boldsymbol{\rho}_{jl}) \quad j = 1, \dots, J, \text{ e } l = 1, \dots, C_j \\
 v_i | \boldsymbol{\pi} &\sim \text{Categórica}(\boldsymbol{\pi}) \quad i \in \{1, \dots, N\} \\
 \text{em que } \boldsymbol{\alpha}_i &= \boldsymbol{\alpha}_{(v_i)} \\
 \boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\lambda})
 \end{aligned} \tag{3.3}$$

em que $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C)^T$, com $\pi_c > 0$ para todo $c = 1, \dots, C$ e $\sum_{c=1}^C \pi_c = 1$, representa a distribuição das classes latentes; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_C)^T$ e $\boldsymbol{\rho}_{jl} = (\rho_{jl1}, \dots, \rho_{jlH_j})^T$ são, respectivamente, vetores de dimensão C e H_j e representam os hiperparâmetros de $\boldsymbol{\pi}$ e $\boldsymbol{\theta}_{jl}$.

Para obter prioris pouco informativas, basta tomar $\boldsymbol{\lambda}$ e $\boldsymbol{\rho}_{jl}$ para $j = 1, \dots, J$ e $l = 1, \dots, H_j$ como vetores com todos os componentes iguais a 1. No entanto, utilizar prioris pouco informativas para $\boldsymbol{\rho}_{jl}$ resulta em problemas na estimação devido a falta de restrição dos parâmetros, conforme descritos nas seções 3.2.3 e 2.4. Por esse motivo, adotamos uma escolha diferente para os hiperparâmetros $\boldsymbol{\rho}_{jl}$, que será apresentada a seguir.

3.3.1 Proposta de restrição baseada na escolha dos hiperparâmetros

Em formulações Bayesianas, as distribuições a priori dos parâmetros representam o conhecimento prévio ou as crenças sobre a distribuição dos parâmetros do modelo. Se houver

informações preliminares sobre os parâmetros, elas deverão ser incorporadas ao modelo por meio da escolha das distribuições a priori. Sendo assim, nos modelos MDC, a distribuição a priori dos parâmetros do item deve estar em concordância com as restrições impostas ao modelo. Portanto, a forma mais natural de introduzir novos conjuntos de restrições é por meio da escolha das prioris.

Neste trabalho, propomos um método alternativo para evitar os resultados indesejáveis, descritos nas seções 3.2.3 e 2.4, baseado estritamente na escolha dos hiperparâmetros para os parâmetros do item. Essa proposta teve como objetivo abranger tanto a RMaP quanto a RMeP, sem ter que aumentar a complexidade da estrutura hierárquica apresentada em 3.3.

O raciocínio dessa proposta é desbalancear as distribuições a priori dos parâmetros estruturais de modo que o conjunto de restrições RMaMeP seja satisfeito com uma alta probabilidade. Isto é, em vez de impor uma série de restrições na distribuição dos parâmetros dos itens de modo que o conjunto RMaMeP seja certamente satisfeito, essa proposta inclui condições mais brandas, nas quais o conjunto de restrições mencionado possui uma alta probabilidade de ser satisfeito, sendo, porém, essa probabilidade menor que 1. Em compensação, obtém-se uma estrutura hierárquica mais simples, na qual as distribuições a priori dos parâmetros dos itens são conjuntamente independentes, fato que contribui para diminuir a complexidade do modelo.

Denominamos essa proposta de *restrição flexível de maior e menor probabilidade*, denotada, doravante, por RFMaMeP. O termo utilizado faz referência ao conceito de restrições flexíveis (*soft constraints*, em inglês), utilizado nas áreas de Otimização e Aprendizado de Máquina, para representar um conjunto de condições que, quando violadas, geram uma penalização do valor-alvo, mas não invalidam a solução proposta (MESEGUER; ROSSI; SCHIEX, 2006).

Para impor a RFMaMeP, definimos, para cada $j = 1, \dots, J$, o hiperparâmetro $\boldsymbol{\rho}_{jl}$ da seguinte forma:

$$\begin{aligned} \boldsymbol{\rho}_{j1} &= \left(C_j, \frac{1}{H_j - 1}, \dots, \frac{1}{H_j - 1} \right)^T \\ \boldsymbol{\rho}_{jl} &= \left(\frac{C_j + 1}{2}, \frac{C_j + 1}{2(H_j - 1)}, \dots, \frac{C_j + 1}{2(H_j - 1)} \right)^T \quad \text{para } l = 2, \dots, C_j - 1 \\ \boldsymbol{\rho}_{jC_j} &= \left(1, \frac{C_j}{H_j - 1}, \dots, \frac{C_j}{H_j - 1} \right)^T \end{aligned} \quad (3.4)$$

Justificativa do método

Apresentamos aqui algumas justificativas para a escolha dos hiperparâmetros apresentados na Equação 3.4.

O primeiro ponto a ser destacado é que essa formulação é baseada na interpretação de que o primeiro nível de resposta ($h = 1$) representa um *feedback* negativo, enquanto os demais

($h = 2, \dots, H_j$) representam diferentes níveis de *feedback* positivo no item avaliado. Em um teste educacional, $h = 1$ pode representar uma resposta errada em uma questão, enquanto os demais valores de h representam as respostas corretas em diferentes níveis (parcialmente correta ou totalmente correta). Em testes psicométricos, nos quais os itens medem a frequência que uma situação ocorre com o respondente, $h = 1$ representa que a situação descrita no item nunca ocorre, enquanto os demais valores de h representam diferentes níveis de ocorrência (raramente, ocasionalmente, frequentemente ou muito frequentemente).

Seja $S_{jl} = \sum_{h=1}^{H_j} \rho_{jlh}$ o parâmetro de concentração do vetor $\boldsymbol{\rho}_{jl}$. Note que, para cada item j , os hiperparâmetros foram definidos de modo que $S_{jl} = C_j + 1$ para todo l , que é subdividido nas categorias das respostas de *feedback* negativo e positivo, de acordo com o grupo latente correspondente.

Essa divisão foi feita de modo a refletir, o mais fielmente possível, o comportamento esperado para os parâmetros dos itens quando os conjuntos de restrições RMaP e RMeP são satisfeitos.

O parâmetro de concentração é dividido entre as duas categorias de *feedback* nas seguintes proporções: C_j para 1, para o primeiro grupo latente ($l = 1$); 1 para C_j , para o último grupo latente ($l = C_j$); e 1 para 1, para os demais grupos latentes ($l = 2, \dots, C_j - 1$, para $C_j > 1$).

Com essa configuração, a distribuição da probabilidade (a priori) é mais concentrada nas respostas de *feedback* negativo para o primeiro grupo latente, enquanto o oposto ocorre com o último grupo latente. Para os demais grupos latentes, a distribuição de probabilidade (a priori) é igualmente dividida entre o conjunto de respostas de *feedback* positivo e negativo. Essa distribuição favorece a satisfação das condições apresentadas para o conjunto RMaMeP.

Definida a divisão do parâmetro de concentração S_{jl} entre as duas categorias de *feedback*, a parcela da concentração destinada ao *feedback* positivo é igualmente dividida entre os $H_j - 1$ níveis de respostas presentes nessa categoria, representando um equilíbrio na distribuição desses níveis de resposta.

Ressaltamos que a escolha dos hiperparâmetros refletem o caso no qual o *feedback* negativo é representado pelo primeiro nível de resposta. Caso o número de níveis, que representam o *feedback* negativo, seja diferente em uma determinada aplicação, a repartição do parâmetro de concentração, dentro de cada vetor $\boldsymbol{\rho}_{jl}$, pode ser ajustada de forma a melhor representar a realidade.

Além disso, a estratégia RFMaMeP possui as seguintes propriedades:

Em primeiro lugar, com a estratégia RFMaMeP, as respostas esperadas dos itens, dado a classe latente do indivíduo, preservam a relação de ordem induzida pelo conjunto de restrições RMaMeP. Isto é, para todo $j = 1, \dots, J$, temos que $E[Y_j | \boldsymbol{\alpha}^{*j} = \mathbf{0}] \leq E[Y_j | \boldsymbol{\alpha}^{*j} = \boldsymbol{\alpha}_{(l)}^{*j}] \leq E[Y_j | \boldsymbol{\alpha}^{*j} = \mathbf{1}]$ para todo $\boldsymbol{\alpha}_{(l)}^{*j}$.

Em segundo lugar, essa estratégia produz a mesma distribuição a priori marginal do conjunto de restrições de maior e menor probabilidade para os parâmetros θ_{jl1} , $j = 1, \dots, J$ e $l = 1, \dots, C_j$, que representam os parâmetros dos itens associados ao primeiro nível de resposta.

Essa última afirmação decorre dos seguintes fatos. Para cada item j , temos que, independente do número de níveis de respostas H_j desse item, para satisfazer os conjuntos de restrições RMaMeP, θ_{j11} e θ_{jC1} devem ser, respectivamente, o menor e maior elemento do conjunto de $\{\theta_{jl1} | l = 1, \dots, C_j\}$. Sabe-se que, se X_1, \dots, X_d são variáveis aleatórias independentes e identicamente distribuídas com distribuição $Beta(1, 1)$, então as variáveis aleatórias $X_{(1)} = \min\{X_1, \dots, X_d\}$ e $X_{(d)} = \max\{X_1, \dots, X_d\}$ possuem distribuição $Beta(d, 1)$ e $Beta(1, d)$, respectivamente. Portanto, ao adotar a estratégia RFMaMeP, as distribuições marginais a priori de θ_{j11} e θ_{jC1} são, respectivamente, $Beta(C_j, 1)$ e $Beta(1, C_j)$; que corresponde exatamente às distribuições que esses parâmetros teriam, caso o conjunto de restrições RMaMeP fosse imposto.

Outra vantagem da RFMaMeP é a facilidade de sua implementação. Para impor o conjunto de restrições RM, uma série de sistemas de inequações precisam ser satisfeitas para cada item j . Assim, a incorporação dessas restrições em um modelo Bayesiano exigiria uma estrutura hierárquica sofisticada. Mesmo para itens dicotômicos, não há uma implementação geral conhecida da RM que funcione para todas as matrizes Q válidas, sendo necessário uma implementação personalizada para cada matriz Q .

Em relação às RMaP e RMeP, apesar de existir implementações desses conjuntos de restrições para itens de resposta dicotômica (conforme apresentado no capítulo 2), uma solução geral para estender essas restrições para respostas politômicas não é conhecida. Em contrapartida, a implementação da RFMaMeP é simples de ser implementada e modificada, se necessário. O código para a estimativa do modelo proposto com a inclusão da estratégia RFMaMeP utilizando o software JAGS (PLUMMER *et al.*, 2016) em R encontra-se no endereço eletrônico <<https://github.com/renato-fernandes-usp/cdm>>.

3.4 Estudo de simulação

Nesta seção, apresentamos um estudo de simulação desenvolvido para avaliar a recuperação dos parâmetros da abordagem Bayesiana para o GDPM com a estratégia RFMaMeP, apresentada na seção anterior.

O objetivo desse estudo é avaliar se a recuperação do método estimação Bayesiano é satisfatória ou não. Por esse motivo, os resultados obtidos com o método de estimação Bayesiano são comparados aos obtidos com o método de estimação clássico, sendo esse último utilizado como base de comparação da acurácia na recuperação dos parâmetros.

3.4.1 Detalhes dos cenários

O estudo de simulação foi delineado em três cenários, que foram inspirados em dois conjuntos de dados simulados para respostas dicotômicas, disponíveis no pacote GDINA (MA; DE LA TORRE, 2016a), denominados **sim10gdina** e **sim30gdina**.

Os cenários I e II são baseados na estrutura do **sim10gdina**. Eles possuem $J = 10$ itens que avaliam $K = 3$ atributos; ambos usam a mesma matriz Q e têm os mesmos valores para os parâmetros do item associados ao primeiro nível de respostas; isto é, os parâmetros θ_{jl1} são idênticos nos dois cenários para todo $j = 1, \dots, J$ e $l = 1, \dots, C_j$.

Os cenários mencionados diferem no número de níveis de respostas de cada item, definidos como $H_j = 3$ e $H_j = 4$ para todos os itens, respectivamente; e pela regra utilizada para definir os demais parâmetros dos itens, conforme detalharemos mais adiante.

Da mesma forma, o cenário III é baseado na estrutura de **sim30gdina**, contando com $J = 30$ itens que avaliam $K = 5$ atributos. Além disso, o número de níveis de respostas dos itens foi fixado em $H_j = 3$ para todos os itens.

As Tabelas 12 e 13¹ representam as matrizes Q utilizadas nos cenários I e II e cenário III, respectivamente. Os parâmetros dos itens para o primeiro nível de resposta dos cenários I e II são apresentados na Tabela 14 e do cenário III, na Tabela 15. Por fim, na Tabela 16 são apresentadas as fórmulas utilizadas para obter os parâmetros dos itens para os níveis de resposta $h > 1$.

Tabela 12 – Matriz Q dos cenários I e II do estudo de simulação do MGDCP

item	atributos		
	α_1	α_2	α_3
1	1	0	0
2	0	1	0
3	0	0	1
4	1	0	1
5	0	1	1
6	1	1	0
7	1	0	1
8	1	1	0
9	0	1	1
10	1	1	1

Fonte: Adaptada de Ma e de la Torre (2016a).

¹ Note que essas matrizes Q são a mesmas que foram utilizadas nos cenários I e II do estudo de simulação apresentado no capítulo 2.

Tabela 13 – Matriz Q do cenário III do estudo de simulação do MGDCP

item	atributo					item	atributo				
	α_1	α_2	α_3	α_4	α_5		α_1	α_2	α_3	α_4	α_5
1	1	0	0	0	0	16	0	1	0	1	0
2	0	1	0	0	0	17	0	1	0	0	1
3	0	0	1	0	0	18	0	0	1	1	0
4	0	0	0	1	0	19	0	0	1	0	1
5	0	0	0	0	1	20	0	0	0	1	1
6	1	0	0	0	0	21	1	1	1	0	0
7	0	1	0	0	0	22	1	1	0	1	0
8	0	0	1	0	0	23	1	1	0	0	1
9	0	0	0	1	0	24	1	0	1	1	0
10	0	0	0	0	1	25	1	0	1	0	1
11	1	1	0	0	0	26	1	0	0	1	1
12	1	0	1	0	0	27	0	1	1	1	0
13	1	0	0	1	0	28	0	1	1	0	1
14	1	0	0	0	1	29	0	1	0	1	1
15	0	1	1	0	0	30	0	0	1	1	1

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).Tabela 14 – Parâmetros dos itens dos cenários I e II para o nível de resposta $h = 1$ do estudo de simulação do MGDCP

item (j)	θ_{j11}	θ_{j21}	θ_{j31}	θ_{j41}	θ_{j51}	θ_{j61}	θ_{j71}	θ_{j81}
	$P_{j1}(0)$	$P_{j1}(1)$						
	$P_{j1}(00)$	$P_{j1}(10)$	$P_{j1}(01)$	$P_{j1}(11)$				
	$P_{j1}(000)$	$P_{j1}(100)$	$P_{j1}(010)$	$P_{j1}(001)$	$P_{j1}(110)$	$P_{j1}(101)$	$P_{j1}(011)$	$P_{j1}(111)$
1	0.8	0.1	-	-	-	-	-	-
2	0.9	0.2	-	-	-	-	-	-
3	0.9	0.1	-	-	-	-	-	-
4	0.9	0.7	0.5	0.1	-	-	-	-
5	0.9	0.9	0.9	0.2	-	-	-	-
6	0.8	0.1	0.1	0.1	-	-	-	-
7	0.9	0.55	0.55	0.2	-	-	-	-
8	0.9	0.72	0.72	0.2	-	-	-	-
9	0.9	0.6	0.6	0.2	-	-	-	-
10	0.9	0.8	0.7	0.6	0.6	0.5	0.3	0.1

Nota: a primeira linha do cabeçalho representa os parâmetros dos itens na forma θ_{jc} , enquanto a segunda, terceira e quarta linha, representam, respectivamente, as probabilidades correspondente $P_{j1}(\alpha^{*j})$ para os itens com K_j igual a 1, 2 e 3.

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).

Tabela 15 – Parâmetros dos itens dos cenários I e II para o nível de resposta $h = 1$ do estudo de simulação do MGDCP

item (j)	θ_{j11}	θ_{j21}	θ_{j31}	θ_{j41}	θ_{j51}	θ_{j61}	θ_{j71}	θ_{j81}
	$P_{j1}(0)$	$P_{j1}(1)$						
	$P_{j1}(00)$	$P_{j1}(10)$	$P_{j1}(01)$	$P_{j1}(11)$				
	$P_{j1}(000)$	$P_{j1}(100)$	$P_{j1}(010)$	$P_{j1}(001)$	$P_{j1}(110)$	$P_{j1}(101)$	$P_{j1}(011)$	$P_{j1}(111)$
1 a 5	0.9	0.2	-	-	-	-	-	-
6 a 10	0.8	0.1	-	-	-	-	-	-
11 a 14	0.8	0.5	0.5	0.2	-	-	-	-
16 a 20	0.9	0.9	0.9	0.2	-	-	-	-
21 a 25	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2
26 a 27	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.2
28 a 30	0.8	0.6	0.6	0.6	0.4	0.4	0.4	0.2

Nota: a primeira linha do cabeçalho representa os parâmetros dos itens na forma θ_{jc} , enquanto a segunda, terceira e quarta linha, representam, respectivamente, as probabilidades correspondente $P_{j1}(\alpha^{*j})$ para os itens com K_j igual a 1, 2 e 3.

Fonte: Adaptada de [Ma e de la Torre \(2016a\)](#).

Tabela 16 – Formulas utilizadas para obtenção dos parâmetros dos itens para níveis de respostas $h > 1$ no estudo de simulação do MGDCP

	$l \backslash h$	parâmetro do item θ_{jlh}		
		2	3	4
Cenários I e III	1	$0.8(1 - \theta_{j11})$	$0.2(1 - \theta_{j11})$	-
	$2, \dots, C_j - 1$	$0.5(1 - \theta_{j11})$	$0.5(1 - \theta_{j11})$	-
	C_j	$0.4(1 - \theta_{j11})$	$0.6(1 - \theta_{j11})$	-
Cenário II	1	$0.6(1 - \theta_{j11})$	$0.3(1 - \theta_{j11})$	$0.1(1 - \theta_{j11})$
	$2, \dots, C_j - 1$	$0.4(1 - \theta_{j11})$	$0.4(1 - \theta_{j11})$	$0.2(1 - \theta_{j11})$
	C_j	$0.3(1 - \theta_{j11})$	$0.4(1 - \theta_{j11})$	$0.3(1 - \theta_{j11})$

Fonte: Elaborada pelo autor.

Com relação ao processo de geração de dados, simulamos testes com $N = 100, 200, 500, 1000$ respondentes (que representam o tamanhos de amostras) para cada cenário. Ademais, os vetores latentes dos indivíduos foram gerados uma única vez para cada cenário e tamanho de amostra, sendo o perfil latente de cada indivíduo escolhido aleatoriamente dentre todas as possibilidades, com igual probabilidade para todos. Por fim, especificados os parâmetros do modelo, cada teste foi replicado $R = 200$ vezes.

3.4.2 Detalhes dos métodos de estimação

Para o método de estimação clássico, foi utilizado o algoritmo EM para obter uma aproximação MMLE, utilizando o algoritmo implementado em R pelo autor desta tese. Tal algoritmo permite a inclusão dos conjuntos de restrições de menor probabilidade, maior probabilidade, maior e menor probabilidade, e monotônica.

Para o método de estimação Bayesiano, adotamos a formulação com a estratégia RF-MaMeP, conforme apresentado na seção 3.3.1. Uma aproximação da distribuição a posteriori foi obtida, utilizando um algoritmo MCMC, implementado no software JAGS (PLUMMER *et al.*, 2016). Para o MCMC, foram utilizadas duas cadeias com 27000 iterações, sendo, as 7000 primeiras, descartadas. Não foi utilizado *thining*. Desse modo, foram geradas 40000 amostras de cada parâmetro.

A acurácia na recuperação dos parâmetros foi medida a partir das estimativas pontuais dos parâmetros. Para a abordagem frequentista, a última interação do algoritmo EM foi adotada como estimativa pontual para os parâmetros do item e o vetor de atributos dos indivíduos foi obtido a partir do método *Maximum a Posteriori*, calculado a partir das estimativas pontuais dos parâmetros do modelo. Para o método Bayesiano, a média da amostra MCMC e a moda² foram adotadas como estimativa pontual para os parâmetros do item e a classe latente dos indivíduos, respectivamente.

Para os parâmetros do item, consideramos a raiz do erro quadrático médio (RMSE) de cada parâmetro ($RMSE(\theta_{jlh})$) para avaliar a acurácia da recuperação. Além disso, a fim de comparar os métodos de estimativa, obtemos $\Delta RMSE(\theta_{jlh})$, definido como a diferença entre o RMSE de θ_{jlh} obtido com o método de estimativa Frequentista e Bayesiano. Tais medidas foram obtidas através das seguintes formulas:

$$RMSE(\theta_{jlh}) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_{jlh}^{(r)} - \theta_{jlh})^2}$$

$$\Delta RMSE(\theta_{jlh}) = RMSE_C(\theta_{jlh}) - RMSE_B(\theta_{jlh})$$

² Aqui, a moda amostral foi utilizada pois o parâmetro estimado é uma variável categórica, sendo a média amostral inapropriada para esse tipo de variável.

em que R ($=200$) é o total de replicações; o sobrescrito (r) , $r = 1, \dots, R$ indica a replicação considerada para a avaliação; e $\widehat{\theta}_{j|lh}^{(r)}$ representa a estimativa pontual do parâmetro $\theta_{j|lh}$, obtida na r -ésima replicação. Ainda, $\text{RMSE}_C(\theta_{j|lh})$ e $\text{RMSE}_B(\theta_{j|lh})$ correspondem ao RMSE do parâmetro $\theta_{j|lh}$, referente aos métodos de estimação Clássico e Bayesiano, respectivamente.

Para o vetor de atributos latentes dos indivíduos, consideramos o *Attribute Correct Classification Ratio* ($\text{ACCR}^{(r)}$) e o *Pattern Correct Classification Ratio* ($\text{PCCR}^{(r)}$), que medem, respectivamente, a taxa de recuperação por componente e por vetor dos atributos latentes dos indivíduos, calculados para cada replicação. Além disso, essas medidas são sintetizadas pelo valor médio entre as replicações, e representadas por $\overline{\text{ACCR}}$ e $\overline{\text{PCCR}}$, respectivamente. Tais medidas foram obtidas por meio das seguintes formulas:

$$\begin{aligned} \text{PCCR}^{(r)} &= \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\widehat{\alpha}_i^{(r)} = \alpha_i) \\ \text{ACCR}^{(r)} &= \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(\widehat{\alpha}_{ik}^{(r)} = \alpha_{ik}) \\ \overline{\text{PCCR}} &= \frac{1}{R} \sum_{i=1}^R \text{PCCR}^{(r)} \\ \overline{\text{ACCR}} &= \frac{1}{R} \sum_{i=1}^R \text{ACCR}^{(r)} \end{aligned}$$

em que $\widehat{\alpha}_i^{(r)}$ e $\widehat{\alpha}_{ik}^{(r)}$ representam, respectivamente, as estimativa pontuais do vetor α_i e do k -ésimo componente desse vetor, α_{ik} , obtidas na r -ésima replicação. As demais quantidades apresentadas já foram previamente definidas.

3.4.3 Resultados do estudo de simulação

Inicialmente, foi avaliada a convergência do método MCMC, utilizando a estatística de Gelman Rubin (GELMAN; RUBIN, 1992) - denotada, doravante, por \widehat{R} - utilizando o algoritmo proposto por Vehtari *et al.* (2021). Valores de \widehat{R} próximos de 1 indicam que a cadeia de Markov convergiu para a distribuição-alvo, sendo 1.1 um valor amplamente utilizado como ponto de corte como critério de convergência (ZHAN *et al.*, 2019; DU *et al.*, 2022). Nesse estudo de simulação, o \widehat{R} de todos os parâmetros foi menor que 1.1 para todas as replicações e nenhum problema de convergência foi detectado.

Os resultados referentes à acurácia na recuperação dos atributos latentes dos indivíduos são sumarizados na tabela 17, no qual são apresentados os valores médios do ACCR e do PCCR obtido com os dois métodos de estimação.

Podemos observar na Tabela 17, que ambos os métodos apresentam resultados muito semelhantes em todos os cenários e tamanhos de amostra em termos da recuperação dos atributos

Tabela 17 – Resultados da acurácia na recuperação dos atributos latentes no estudo de simulação

Cenário	N	método de estimação	ACCR	PCCR	
I	100	Clássico	87.5 % (2.5 %)	68.9 % (5.7 %)	
		Bayesiano	88.3 % (2.4 %)	70.5 % (5.5 %)	
	200	Clássico	89.4 % (1.6 %)	73.2 % (4.0 %)	
		Bayesiano	89.6 % (1.6 %)	73.5 % (3.9 %)	
	500	Clássico	90.7 % (0.9 %)	76.4 % (2.0 %)	
		Bayesiano	90.5 % (0.9 %)	76.0 % (2.2 %)	
	1000	Clássico	91.0 % (0.6 %)	77.1 % (1.3 %)	
		Bayesiano	90.9 % (0.6 %)	76.9 % (1.4 %)	
	II	100	Clássico	86.0 % (3.0 %)	65.0 % (6.9 %)
			Bayesiano	86.8 % (2.6 %)	66.8 % (6.0 %)
200		Clássico	88.2 % (1.6 %)	70.5 % (3.9 %)	
		Bayesiano	88.4 % (1.9 %)	70.7 % (4.4 %)	
500		Clássico	90.0 % (1.0 %)	74.7 % (2.2 %)	
		Bayesiano	89.7 % (1.1 %)	74.1 % (2.5 %)	
1000		Clássico	90.6 % (0.6 %)	76.2 % (1.4 %)	
		Bayesiano	90.3 % (0.7 %)	75.6 % (1.6 %)	
III		100	Clássico	92.4 % (1.6 %)	69.0 % (5.7 %)
			Bayesiano	93.0 % (1.4 %)	71.7 % (5.2 %)
	200	Clássico	93.9 % (1.0 %)	75.1 % (3.6 %)	
		Bayesiano	94.1 % (0.9 %)	75.9 % (3.4 %)	
	500	Clássico	95.1 % (0.4 %)	79.5 % (1.7 %)	
		Bayesiano	95.1 % (0.5 %)	79.5 % (1.9 %)	
	1000	Clássico	95.4 % (0.3 %)	80.8 % (1.3 %)	
		Bayesiano	95.4 % (0.3 %)	80.7 % (1.5 %)	

Nota: os valores entre parênteses denotam o desvio padrão.

Fonte: Dados da pesquisa.

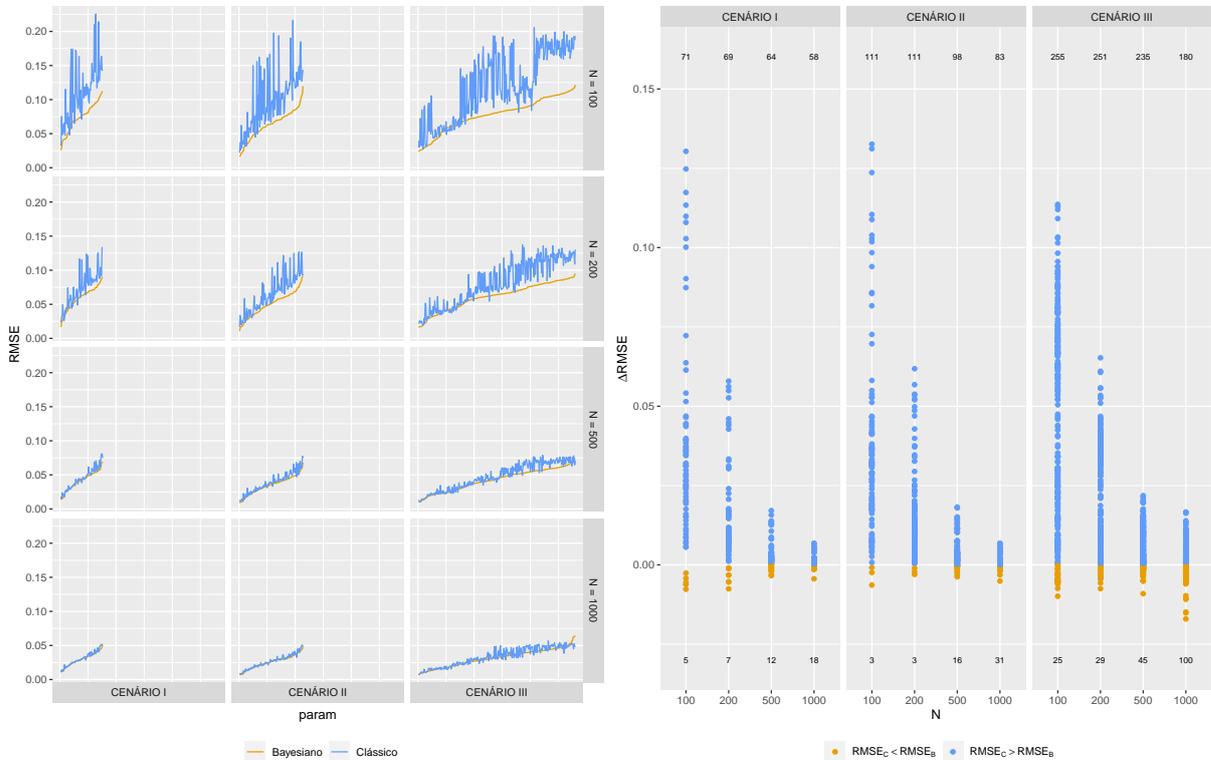
(medido pelo ACCR) e do perfil latente (medido pelo PCCR), não havendo diferença significativa entre os métodos nesse quesito. Não houve também variação relevante entre os resultados do cenário I e II, indicando que o número de níveis de respostas não afeta de forma significativa a acurácia da recuperação dos parâmetros.

A Figura 5 apresenta os resultados relativos à acurácia na recuperação dos parâmetros do item. Em (a) é exibido os valores de RMSE, obtidos com cada método de estimação para cada parâmetro estrutural para os diferentes cenários e tamanho de amostra, enquanto em (b), a diferença no valor do RMSE entre os dois métodos, em termos do Δ RMSE, em que um valor positivo de Δ RMSE indica que o RMSE, obtido pelo método de estimação clássico, é maior do que o Bayesiano para um parâmetro específico.

Destacamos que, na Figura 5 (a), os parâmetros dos itens foram ordenados, para cada

cenário e tamanho de amostra, em ordem crescente do RMSE obtido pelo método de estimação Bayesiano, a fim de facilitar a visualização da tendência observada.

Figura 5 – Resultados da acurácia na recuperação dos parâmetros dos itens no estudo de simulação



(a) RMSE dos parâmetros dos itens

(b) Diferença de RMSE dos parâmetros dos itens

Nota: (a) cada ponto representado no eixo x representa um parâmetro do item. Para melhor visualização, os parâmetros foram ordenados segundo o valor do RMSE obtido com o método de estimação Bayesiano, do menor para o maior. (b) cada ponto representa um parâmetro do item, havendo, respectivamente, 76, 114 e 280 pontos nos cenários I, II e III. Os valores destacados, acima e abaixo de cada gráfico, representam, respectivamente, o número de pontos com valores ΔRMSE positivos e negativos.

Fonte: Dados da pesquisa.

Com relação aos cenários I e II, podemos observar, na Figura 5, que o método de estimação Bayesiano apresenta RMSE menor do que o clássico para quase todos os parâmetros das amostras de tamanho 100 e 200. Além disso, nos poucos pontos onde o método clássico tem RMSE menor (representados pelos pontos laranja em (b)), a diferença é muito pequena (menor que 0.01). Em contrapartida, existem parâmetros onde o RMSE da estimação com o método clássico tem valor até 0.10 superior do que o da estimação Bayesiana. Ademais, podemos observar que a diferença de RMSE entre os dois métodos diminui à medida que o tamanho da amostra aumenta, sendo que, para o maior tamanho de amostra, essa diferença torna-se desprezível.

Para o cenário III, o método de estimação Bayesiano continua apresentando melhores resultados do que o método clássico para a maioria dos parâmetros, embora essa superioridade não seja tão expressiva quanto a observada para os outros cenários. Podemos notar também

que, nesse cenário, o RMSE diminui mais lentamente com o aumento da amostra. Desse modo, amostras maiores são necessárias para que a recuperação dos parâmetros desse cenário atinja um nível de acurácia próximo àqueles obtidos para as amostras de tamanho $N=1000$ dos cenários I e II.

Por fim, com base nos resultados apresentados, podemos concluir que o método de estimação Bayesiano do MGDPC recupera os parâmetros de forma satisfatória, apresentando acurácia superior ao método clássico na recuperação dos parâmetros dos itens para as amostras pequenas.

3.5 Discussão

Neste capítulo foi apresentado uma formulação Bayesiana para o MGDPC, que inclui um novo conjunto de restrições para os parâmetros, que foi denominado de restrições flexíveis de maior e menor probabilidade, denotado por RFMaMeP. Esse conjunto de restrições tem, como vantagem, a facilidade de implementação, que depende apenas da definição dos hiperparâmetros ρ_{jl} .

A recuperação dos parâmetros pelo método de estimação Bayesiano proposto foi avaliado por meio de um estudo de simulação, no qual, a acurácia da recuperação dos parâmetros do método Bayesiano foi comparada a do método de estimação clássico. Esse estudo de simulação demonstrou que o método de estimação Bayesiano do MGDPC apresenta melhor acurácia do que o método de estimação clássico, especialmente para os tamanhos de amostra menores ($N = 100$ e $N = 200$). Para a maior amostra avaliada ($N = 1000$), o método Bayesiano é, ainda, ligeiramente superior. Porém, nesse último caso, a diferença de acurácia entre os dois métodos é negligenciável em termos práticos de aplicação.

APLICAÇÃO - DADOS DE DEPRESSÃO

4.1 Introdução

Neste capítulo, é apresentado um exemplo de aplicação no qual os métodos de estimação Bayesiano do modelo G-DINA e do MGDCP, apresentados, respectivamente, nos capítulos 2 e 3, são utilizados para classificar indivíduos segundo o perfil latente de depressão que possuem a partir de suas respostas no Inventário de Depressão de Beck (denotado, doravante, pela sigla BDI, oriunda do termo em inglês, *Beck Depression Inventory*, adotado usualmente na literatura). Nessa aplicação são analisados os dados referentes às respostas de 1111 estudantes universitários brasileiros ao BDI.

Este capítulo está organizado da seguinte forma: primeiro são apresentados os detalhes do teste e a metodologia de classificação utilizada; em seguida, é realizada uma análise comparativa dos resultados obtidos utilizando o modelo DINA e o modelo G-DINA para os dados dicotomizados; depois, é feita uma comparação entre os resultados obtidos, utilizando o G-DINA com os dados dicotomizados e o MGDCP com os dados na sua forma politômica original.

4.2 Metodologia de classificação da depressão utilizando MDC

O BDI é um teste psicométrico autorreportado composto por 21 itens de múltipla escolha, que medem sintomas da depressão. Cada item está relacionado a um sintoma diferente e possui quatro opções de respostas pontuadas de 0 a 3; em que 0 indica ausência do sintoma e 3, a maior severidade do sintoma. O método tradicional de diagnóstico é baseado na soma da pontuação dos itens, sendo que indivíduos com pontuação total entre, 0 e 15, são classificados como não depressivos, indivíduos com pontuação total, entre 16 e 20, são classificados como disfóricos e indivíduos, com pontuação total acima de 20, são classificados como depressivos (SILVA *et al.*,

2018).

Primeiro, [Fragoso e Cúri \(2013\)](#) analisaram a estrutura dos fatores latentes do Inventário de Depressão de Beck, utilizando modelos TRI multidimensionais. Nesse estudo, foi constatado que as respostas dos itens podem ser relacionadas a duas dimensões latentes: a primeira, associada aos sintomas de ordem cognitiva e a segunda, aos sintomas de ordem somático-afetiva. Além disso, foi identificado quais itens estavam associados a cada dimensão.

Baseado nas descobertas de [Fragoso e Cúri \(2013\)](#), [Silva et al. \(2018\)](#) propôs uma metodologia de classificação da depressão utilizando o modelo DINA, considerando as duas dimensões da depressão mencionadas anteriormente e a relação entre os itens e as dimensões da depressão. Tal relação foi utilizada para a construção da matriz Q , apresentada na Tabela 18, utilizada no modelo.

Tabela 18 – Matriz Q do utilizada nas metodologias de classificação das respostas do BDI

item	descrição do sintoma	Dimensão cognitiva (α_1)	Dimensão somática-afetiva (α_2)
1	Tristeza	1	1
2	Pessimismo	1	1
3	Senso de fracasso	1	0
4	Falta de satisfação	1	1
5	Sentimento de culpa	1	0
6	Senso de punição	1	0
7	Aversão a si mesmo	1	0
8	Auto acusação	1	0
9	Desejos suicidas	1	1
10	Crises de choro	1	1
11	Irritabilidade	0	1
12	Retraimento social	1	1
13	Indecisão	1	1
14	Distorção da imagem do corpo	1	1
15	Inibição do trabalho	1	1
16	Distúrbio do sono	0	1
17	Fadiga	0	1
18	Perda de apetite	1	1
19	Perda de peso	0	1
20	Preocupação somática	1	1
21	Perda de libido	0	1

Fonte: Adaptada de [Oliveira, Wang e Bazán \(2023\)](#).

A metodologia de [Silva et al. \(2018\)](#) considera um vetor de atributos bidimensional $\alpha = (\alpha_1, \alpha_2)^T$ em que $\alpha_k = 1$, com $k = 1, 2$, indica, respectivamente, a presença de sintomas na dimensão cognitiva ($k = 1$) e somático-afetiva ($k = 2$); e $\alpha_k = 0$, a ausência de sintomas nas dimensões mencionadas. Dessa forma, cada indivíduo é classificado de acordo com o perfil latente que possui, a saber:

- indivíduos com perfil latente $\boldsymbol{\alpha}_{(1)} = (0, 0)^T$ são classificados como não sintomáticos;
- indivíduos com perfil latente $\boldsymbol{\alpha}_{(2)} = (1, 0)^T$ são classificados como sintomáticos na dimensão cognitiva;
- indivíduos com perfil latente $\boldsymbol{\alpha}_{(3)} = (0, 1)^T$ são classificados como sintomáticos na dimensão somático-afetiva;
- indivíduos com perfil latente $\boldsymbol{\alpha}_{(4)} = (1, 1)^T$ são classificados como sintomáticos em ambas as dimensões.

Para compatibilizar os dados com a aplicação de modelo DINA e G-DINA, as respostas do teste BDI foram dicotomizadas de modo que:

$$Y_{ij} = \begin{cases} 0, & \text{se o indivíduo } i \text{ escolheu a alternativa com pontuação 0 no item } j \\ 1, & \text{se o indivíduo } i \text{ escolheu a alternativa com pontuação 1, 2 ou 3 no item } j \end{cases}$$

para todo $i = 1, \dots, 1111$ e $j = 1, \dots, 21$.

4.3 A análise comparativa entre DINA e G-DINA

Nesta seção apresentamos uma adaptação do método proposto por (SILVA *et al.*, 2018), na qual o modelo G-DINA é utilizado no lugar do modelo DINA no método de classificação da depressão para dados do BDI, descrita na seção anterior. Ressalta-se que as demais características da metodologia de classificação, como a matriz Q adotada e a forma de dicotomização das respostas, não foram alteradas.

Apresentamos uma análise comparativa entre a metodologia de classificação, utilizando o modelo DINA e o modelo G-DINA. Nessa análise, os dados dos 1111 respondentes foram ajustados, utilizando o modelo DINA e o modelo G-DINA. Em ambos os modelos, os parâmetros foram estimados sobre uma abordagem Bayesiana, utilizando algoritmos MCMC, implementado com o software JAGS. O código do modelo JAGS utilizado para estimação do modelo DINA pode ser encontrado no material suplementar de Silva *et al.* (2018); o código referente ao modelo G-DINA pode ser encontrado no endereço eletrônico <<https://github.com/renato-fernandes-usp/cdm>>.

Com relação ao MCMC, foi adotada a mesma configuração para os dois modelos, sendo geradas duas cadeias de Markov com 10000 iterações cada, nas quais as 5000 primeiras iterações de cada cadeia foram descartadas. Não foi utilizado *thinning*. Desse modo, foram obtidas amostras de 10000 elementos para cada parâmetro.

A convergência das cadeias de Markov foram verificadas utilizando a estatística de Gelman-Rubin \hat{R} . Segundo Zhan *et al.* (2019), para problemas com dados reais, valores \hat{R}

menores que 1,1 indicam convergência. Nessa aplicação, os valores \widehat{R} de todos os parâmetros foram menores que 1,003 para os modelos DINA e para o modelo G-DINA, indicando a convergência de ambos.

A respeito dos parâmetros dos itens, é importante destacar que as formulações do DINA e do G-DINA utilizam diferentes parâmetros dos itens. Na parametrização do DINA, utilizada nessa aplicação, os parâmetros do item j são representados pelo vetor $(g_j, s_j)^T$, onde g_j (parâmetro *guessing*) representa a probabilidade de sucesso no item j do indivíduo que não possui todos os atributos requeridos por esse item e s_j (parâmetro *slipping*) representa a probabilidade de fracasso no item j do indivíduo que possui todos os atributos requeridos por esse item.

Já para o modelo G-DINA, os parâmetros do item j são representados pelo vetor $(\theta_{j1}, \dots, \theta_{jC_j})^T$, onde θ_{jl} representa a probabilidade de um indivíduo do grupo latente l ($= 1, \dots, C_j$) responder a esse item corretamente.

Para facilitar a interpretação dos resultados e possibilitar uma comparação direta das estimativas, optamos por apresentar, no lugar dos parâmetros originais dos modelos, as probabilidades $P_j(\boldsymbol{\alpha}_{(c)})$, $c = 1, \dots, 4$, em que $P_j(\boldsymbol{\alpha}_{(c)}) = Pr(Y_j = 1 | \boldsymbol{\alpha} = \boldsymbol{\alpha}_{(c)})$ representa a probabilidade de sucesso dos indivíduos da classe latente c no item j .

O [Quadro 3](#) resume a correspondência entre os parâmetros originais dos modelos DINA e G-DINA e o parâmetro reportado nos resultados.

Quadro 3 – Correspondência entre o parâmetro reportado e os originais dos modelos DINA e G-DINA.

\mathbf{q}_j	$P_j(\boldsymbol{\alpha}_{(c)})$	parâmetro correspondente	
		DINA	G-DINA
$(1,0)^T$	$P_j(00)$	g_j	θ_{j1}
	$P_j(10)$	$1 - s_j$	θ_{j2}
	$P_j(01)$	g_j	θ_{j1}
	$P_j(11)$	$1 - s_j$	θ_{j2}
$(0,1)^T$	$P_j(00)$	g_j	θ_{j1}
	$P_j(10)$	g_j	θ_{j1}
	$P_j(01)$	$1 - s_j$	θ_{j2}
	$P_j(11)$	$1 - s_j$	θ_{j2}
$(1,1)^T$	$P_j(00)$	g_j	θ_{j1}
	$P_j(10)$	g_j	θ_{j2}
	$P_j(01)$	g_j	θ_{j3}
	$P_j(11)$	$1 - s_j$	θ_{j4}

Fonte: Elaborada pelo autor.

Em suma, nessa análise, foram comparadas as estimativas dos seguintes parâmetros:

1. vetores de atributos $\widehat{\boldsymbol{\alpha}}$, para $i = 1, \dots, 1111$;
2. probabilidades de sucesso $pr(\boldsymbol{\alpha}_{(c)})$, para $j = 1, \dots, J$ e $c = 1, \dots, 4$;
3. proporções das classes latentes $\widehat{\pi}_c$, $c = 1, \dots, 4$

4.3.1 Resultados

O primeiro ponto analisado nesse comparativo é o ajuste dos modelos. Foram comparados o modelo DINA e o G-DINA com as duas restrições propostas (RMaP e RMeP), utilizando as seguintes medidas: o *Deviance Information Criterion* (DIC) (SPIEGELHALTER *et al.*, 2002), o *Widely Applicable Information Criterion* (WAIC) (VEHTARI; GELMAN; GABRY, 2017) e a raiz quadrada média do erro de aproximação do item j (RMSEA $_j$) (LEI; LI, 2016).

Para os MDC, o RMSEA $_j$ do item j é dado pela expressão:

$$\text{RMSEA}_j = \sqrt{2 \sum_{c=0}^{c=2^K} \pi_c \left(\hat{P}_j(\boldsymbol{\alpha}_{(c)}) - \frac{R_{jc}}{I_c} \right)^2}$$

em que I_{jc} é o número esperado de indivíduos na c -ésima classe latente e R_{jc} , o número esperado de indivíduos na c -ésima classe latente que dá uma resposta com pontuação positiva ao item j . Ademais, por sucintez, reportamos nessa análise apenas o RMSEA médio, $\overline{\text{RMSEA}} = \frac{1}{J} \sum_{j=1}^J \text{RMSEA}_j$.

Apesar dos diferentes significados dessas três medidas, na comparação de modelos, valores menores indicam melhor ajuste do modelo.

Ainda, com relação aos conjuntos de restrições utilizados, é importante destacar que adotamos uma formulação do modelo DINA que inclui a RM, que, no caso particular desse modelo, é equivalente à RMaP e à RMeP.

A tabela 19 apresenta os valores de DIC, WAIC e RMSEA médio, obtidos com o DINA, G-DINA com os conjuntos de restrições RMaP e RMeP.

Tabela 19 – Comparação do ajuste dos modelos DINA e G-DINA para os dados do BDI

modelo	DIC	WAIC	$\overline{\text{RMSEA}}$
DINA	25538.6	34116.6	0.0365
G-DINA - RMeP	24684.4	33280.3	0.0257
G-DINA - RMaP	24668.0	33284.6	0.0258

Fonte: Dados da pesquisa.

Podemos observar, na Tabela 19, que o modelo G-DINA com a RMaP e com a RMeP apresentaram valores muito próximos para as três medidas, sendo esses valores, significativamente menores do que os apresentados pelo modelo DINA. Portanto, de acordo com esses critérios, o modelo G-DINA é mais adequado do que o modelo DINA para essa aplicação, independente do conjunto de restrições considerado para o primeiro.

Na segunda parte de nossa análise, comparamos os valores estimados dos parâmetros com ambos os modelos. Nosso objetivo é elucidar as diferenças e semelhanças nas estimativas decorrentes da escolha do modelo.

Antes de continuar, salientamos que as estimativas dos parâmetros do G-DINA obtidas com ambos os conjuntos de restrições foram muito próximos: a diferença média absoluta entre os parâmetros estruturais é de 0.001, e mais de 99% dos indivíduos foram igualmente classificados. Por esse motivo, por concisão, no decorrer dessa seção, os resultados referente ao G-DINA com a RMaP serão omitidos, sendo apresentados apenas os resultados referente ao modelo DINA e ao G-DINA com a RMeP.

A Tabela 20 apresenta o cruzamento da classificação dos indivíduos segundo o perfil latente estimado pelos modelos DINA e G-DINA.

Tabela 20 – Tabela de contingência do perfil latente estimado pelos modelo DINA e G-DINA aos respondentes do teste BDI

		G-DINA				Total
		$\hat{\alpha} = (0, 0)$	$\hat{\alpha} = (1, 0)$	$\hat{\alpha} = (0, 1)$	$\hat{\alpha} = (1, 1)$	
DINA	$\hat{\alpha}_1$	378	11	46	0	435
	$\hat{\alpha}_2$	16	109	1	0	126
	$\hat{\alpha}_3$	37	4	65	0	106
	$\hat{\alpha}_4$	0	63	92	289	444
Total		431	187	204	289	1111

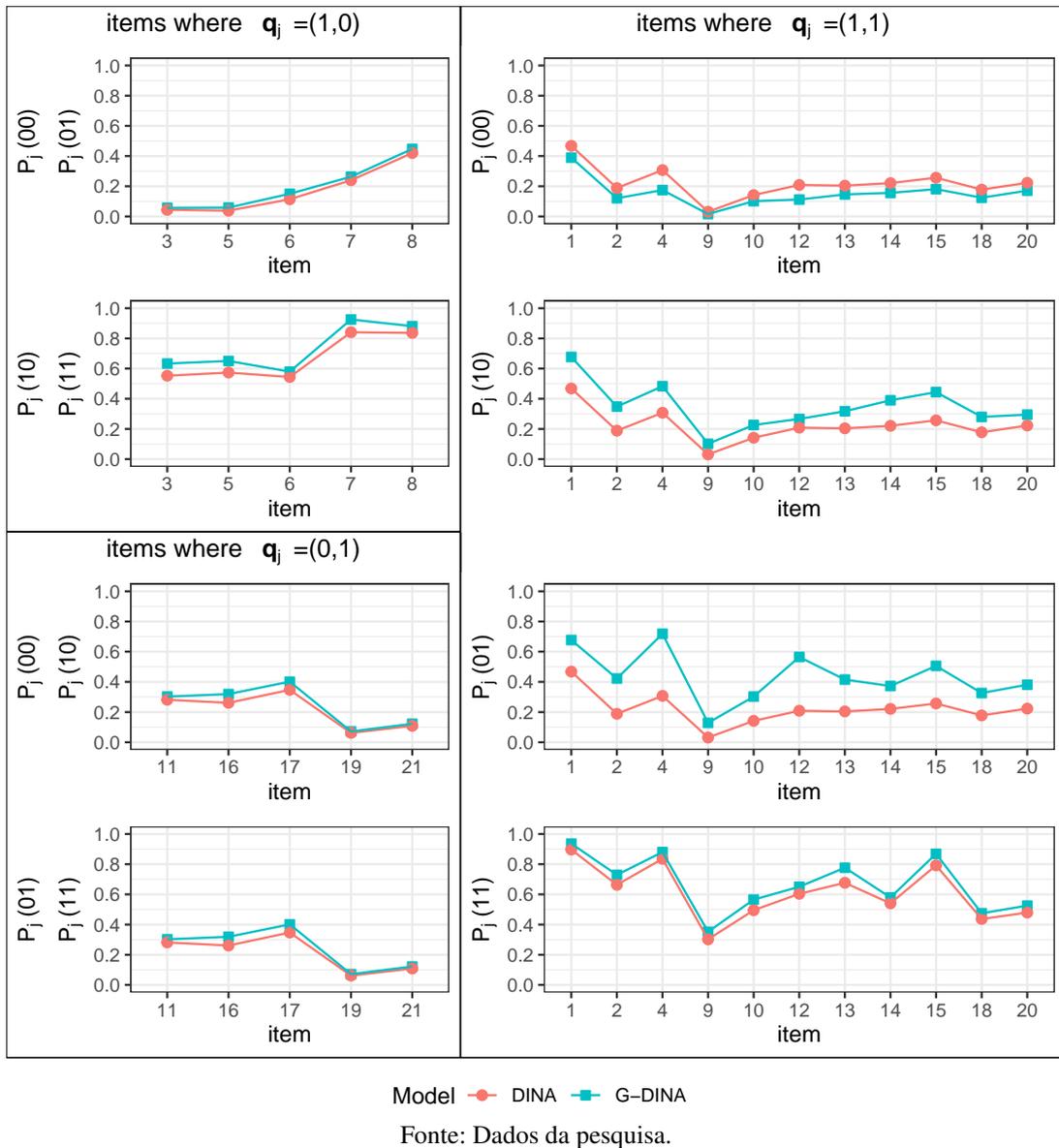
Fonte: Dados da pesquisa.

Podemos observar, na Tabela 20, que todos os indivíduos classificados como sintomáticos em ambas as dimensões ($\hat{\alpha} = (1, 1)^T$) com o G-DINA, também foram classificados dessa forma com o DINA. Por outro lado, entre os indivíduos classificados dessa forma com o modelo DINA, apenas 65% receberam a mesma classificação com o G-DINA, sendo que os 35% restantes foram diagnosticados como sintomáticos em apenas uma dimensão ($\hat{\alpha}$ igual a $(1, 0)^T$ ou $(0, 1)^T$). Esses resultados indicam que a classificação dos indivíduos pode ser bastante afetada pela escolha do modelo e que uma grande parcela dos indivíduos, classificados como sintomático nas duas dimensões com o modelo DINA, receberiam uma classificação mais branda com o modelo G-DINA.

A Figura 6 apresenta as estimativas dos modelos DINA e G-DINA para as probabilidades $P_j(\alpha_{(c)})$ para os diferentes itens j e classes latentes c .

No contexto dessa aplicação, $P_j(\alpha_{(c)})$ representa a probabilidade de um indivíduo com perfil latente $\alpha_{(c)}$ escolher uma alternativa com pontuação positiva para o j -ésimo item do BDI. Podemos observar, na Figura 6, que nos itens relacionados as duas dimensões da depressão, a probabilidade estimada de um indivíduo sintomático, em uma única dimensão, escolher uma resposta com pontuação positiva (representadas pelos gráficos de $P_j(10)$ e $P_j(01)$ na segunda coluna da figura) é consideravelmente menor para o modelo DINA. Essa diferença sistemática nas probabilidades associadas aos indivíduos sintomáticos, em apenas uma dimensão, é, possivelmente, a causa da variação na classificação dos indivíduos observada, na Tabela 20 e discutida anteriormente.

Figura 6 – Probabilidades de sucesso estimadas com o modelo DINA e G-DINA para os itens do teste BDI.



A Tabela 21 traz a proporção estimada de indivíduos pertencentes a cada classe latente com o modelo DINA e o modelo G-DINA, bem como o intervalo de credibilidade desses parâmetros.

Podemos observar, na Tabela 21, que a proporção estimada de indivíduos não depressivos ($\hat{\alpha} = (0,0)^T$) é similar em ambos os modelos, enquanto a proporção estimada de sintomáticos nas duas dimensões ($\hat{\alpha} = (1,1)^T$) é consideravelmente maior no modelo DINA, fato condizente com os resultados apresentados anteriormente.

Note que todos os resultados apresentados indicam que o modelo DINA tende a classificar os indivíduos como sintomáticos nas duas dimensões com uma frequência muito maior do que o modelo G-DINA. Silva *et al.* (2018) observou que o modelo DINA tem a tendência de superestimar os diagnósticos de depressão, sendo tal comportamento associado à dicotomização

Tabela 21 – Proporção dos perfis latentes estimada com o modelo DINA e G-DINA

c	(descrição)	$\hat{\pi}_c$ (DINA)		$\hat{\pi}_c$ (G-DINA)	
		média	HPD (95%)	média	HPD (95%)
1	não depressivo	0.360	(0.315; 0.406)	0.377	(0.336; 0.417)
2	sintomático na dimensão cognitiva	0.124	(0.094; 0.154)	0.173	(0.137; 0.212)
3	sintomático na dimensão somático-afetiva	0.124	(0.085; 0.165)	0.190	(0.150; 0.232)
4	sintomático em ambas dimensões	0.392	(0.354; 0.430)	0.260	(0.225; 0.296)

*HPD (95%): intervalos de credibilidade HPD de 95%.

Fonte: Dados da pesquisa.

das respostas do BDI. Sem descartar o efeito da dicotomização na classificação dos indivíduos, ao comparar os resultados obtidos pelos dois modelos, podemos concluir que a estrutura do modelo DINA tem uma forte influência nessa tendência de superestimação, possivelmente por atribuir, nos itens relacionados as duas dimensões, a mesma probabilidade de resposta com pontuação positiva para indivíduos não sintomáticos e para os sintomáticos em uma única dimensão.

Por fim, devido à tendência do modelo DINA superestimar o diagnóstico de depressão e ter pior ajuste aos dados em comparação ao G-DINA, podemos concluir que, na metodologia de classificação da depressão para respostas do BDI, a utilização do modelo G-DINA é mais apropriada do que o modelo DINA, originalmente proposto em (SILVA *et al.*, 2018).

4.4 Análise comparativa entre G-DINA e MGDCP

Nos estudos mencionados anteriormente, as respostas do BDI foram dicotomizadas para se adequarem aos MDC utilizados. Uma alternativa interessante seria usar um modelo MDC para respostas politômicas, como o MGDCP. Por esse motivo, realizamos uma análise comparativa do uso do G-DINA com as respostas dicotomizadas e do MGDCP com as respostas politômicas, utilizando a metodologia de classificação de depressão apresentada na seção 4.2.

Por clareza das notações que serão utilizadas, convém apresentar com precisão as variáveis respostas consideradas em cada modelo. Seja $Z_{ij} \in \{0, 1, 2, 3\}$, $i = 1, \dots, N$, $j = 1, \dots, J$ as respostas originais do teste BDI. Para o modelo G-DINA, consideramos as respostas dicotomizadas X_{ij} , em que $X_{ij} = 0$ se $Z_{ij} = 0$, representando a ausência do sintoma descrito no item j ; e $X_{ij} = 1$, se $Z_{ij} \geq 1$, representando a manifestação do sintoma mencionado. Além disso, por compatibilidade à definição apresentada na seção 3.2, consideramos no MGDCP $Y_{ij} = Z_{ij} + 1$ como variável resposta. Desse modo, temos que o número de níveis de resposta H_j é igual a 4, para todos os itens do MGDCP e igual a 2, para o G-DINA.

Com relação ao método de estimação, os modelos G-DINA e MGDCP foram ajustados utilizando as respectivas abordagens Bayesianas apresentadas nos capítulos 2 e 3 desta tese. Com

relação ao conjunto de restrições adotados, foi utilizado a RMeP, no G-DINA, e a RFMaMeP, no MGDCP. Ademais, os parâmetros de ambos os modelos foram estimados, utilizando um algoritmo MCMC implementado com o software *JAGS*.

4.4.1 Resultados

Os primeiros comentários referem-se às configurações do MCMC para o modelo MGDCP e sua convergência. Nessa aplicação, foram geradas, para o MGDCP, duas cadeias com 60000 iterações em cada, sendo as primeiras 15000 descartadas. Não foi utilizado *thinning*. Sob essa configuração, o algoritmo MCMC convergiu com sucesso, sendo o \hat{R} de todos os parâmetros menor que 1,06.

Note que o número de iterações utilizado aqui foi muito maior do que o empregado no estudo de simulação, apresentado na seção 3.4. O maior número de iterações foi necessário pois, nas cadeias de Markov, alguns parâmetros dos itens se misturavam mais lentamente. Esse comportamento foi notado principalmente para os parâmetros dos itens, cuja distribuição se concentrava muito próximo do zero.

Para o modelo G-DINA, adotamos a mesma configuração da seção anterior, com duas cadeias com 10000 iterações cada, sendo as primeiras 5000 descartadas, sem *thinning*. O maior valor de \hat{R} foi 1.003, indicando a convergência do MCMC.

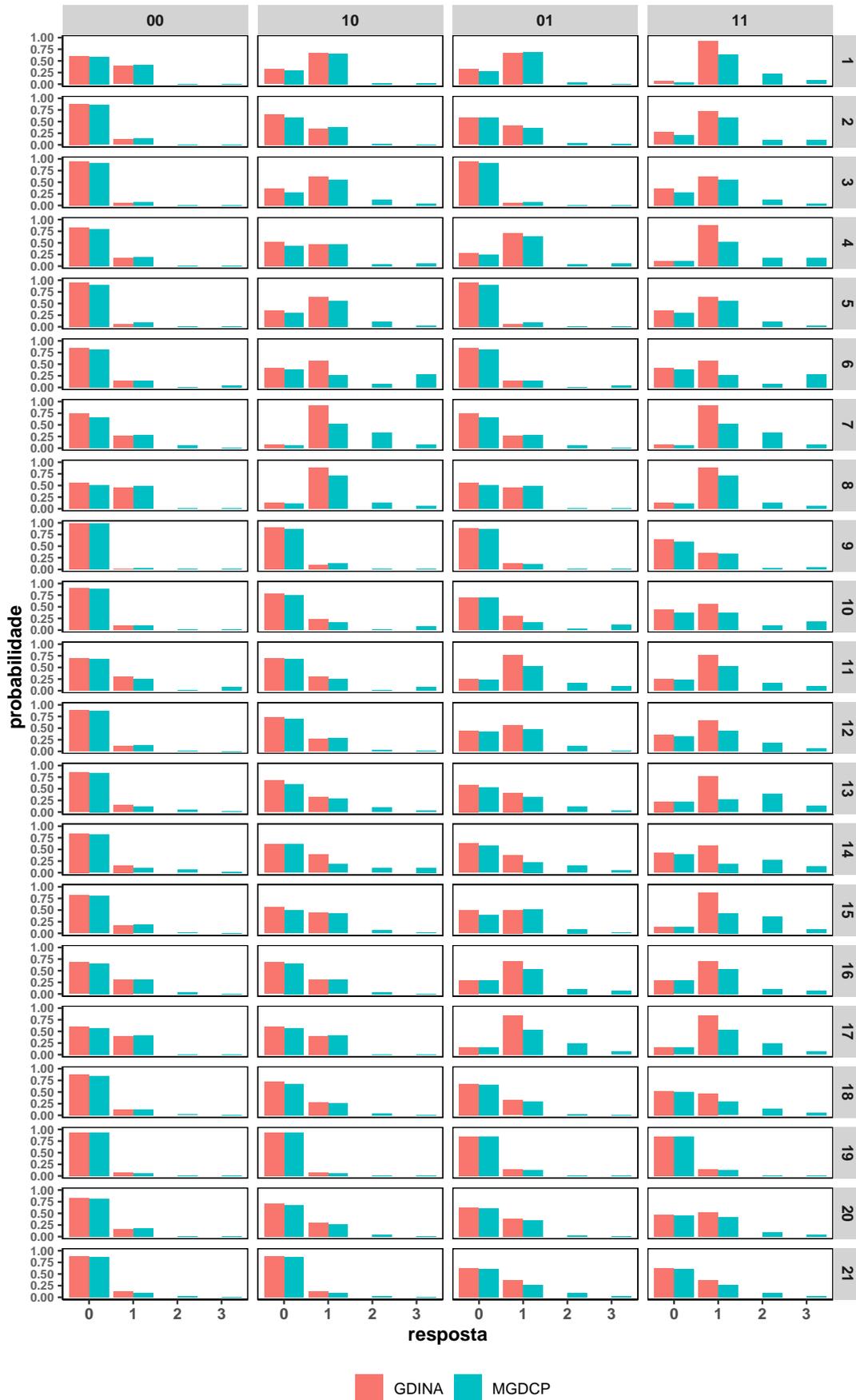
Com relação ao conjunto de restrições impostas ao modelo MGDCP, observamos que a estratégia RFMaMeP surtiu o efeito desejado. Analisando as estimativas dos parâmetros dos itens obtidas com o MGDCP, verificamos que as restrições de maior e de menor probabilidade foram satisfeitas para todos os itens.

Ressaltamos que, esse conjunto de restrições é apropriado para essa aplicação, pois é esperado que os indivíduos com depressão tenham probabilidades maiores de reportar os sintomas mais graves e que os indivíduos não depressivos tenham probabilidades maiores de reportar os sintomas mais leves, ou a ausência deles.

Com respeito aos parâmetros dos itens, temos na Figura 7 e Tabela 22, os valores estimados de $P_{jh}(\boldsymbol{\alpha}_{(c)}) = Pr(Y_j = h | \boldsymbol{\alpha} = \boldsymbol{\alpha}_{(c)})$ para cada modelo, em que $P_{jh}(\boldsymbol{\alpha}_{(c)})$ representa a probabilidade do indivíduo com perfil latente $\boldsymbol{\alpha}_{(c)}$ ($c = 1, \dots, 4$) escolher a resposta h ($= 0, \dots, 4$) no item j ($= 1, \dots, 21$) do teste BDI, na forma de gráfico e tabela, respectivamente.

Podemos observar, na Figura 7 e Tabela 22, que, para alguns itens, tais como os que remetem aos desejos suicidas (9) e à perda de peso (19), a probabilidade de um indivíduo escolher respostas de nível 2 ou 3 é muito baixa, independente do perfil latente que possui. Esse resultado pode ser explicado pelo fato que esses níveis de respostas indicam maior gravidade no sintomas de depressão, fato que não é frequente na população estudada. Consideramos que esses resultados podem ser interessantes a pesquisadores do tema, sendo que análises clínicas adicionais podem auxiliar novas descobertas relacionadas a esse aspecto.

Figura 7 – Gráfico das probabilidades estimadas para o conjunto de dados do BDI



Fonte: Dados da pesquisa.

Tabela 22 – Probabilidades estimadas com os modelos G-DINA e MGDCP para o conjunto de dados do BDI

(continua)

item	h	MGDCP				G-DINA			
		θ_{j1h}	θ_{j2h}	θ_{j3h}	θ_{j4h}	θ'_{j1h}	θ'_{j2h}	θ'_{j3h}	θ'_{j4h}
1	0	0.584	0.292	0.275	0.044	0.610	0.323	0.323	0.064
	1	0.412	0.662	0.684	0.639	0.390	0.677	0.677	0.936
	2	0.003	0.018	0.035	0.233				
	3	0.001	0.028	0.005	0.084				
2	0	0.853	0.588	0.589	0.203	0.879	0.652	0.578	0.271
	1	0.137	0.381	0.359	0.586	0.121	0.348	0.422	0.729
	2	0.005	0.022	0.035	0.105				
	3	0.005	0.009	0.017	0.105				
3	0	0.918	0.289	0.918	0.289	0.943	0.367	0.943	0.367
	1	0.074	0.555	0.074	0.555	0.057	0.633	0.057	0.633
	2	0.005	0.119	0.005	0.119				
	3	0.003	0.037	0.003	0.037				
4	0	0.788	0.431	0.256	0.107	0.825	0.517	0.282	0.120
	1	0.205	0.474	0.645	0.521	0.175	0.483	0.718	0.880
	2	0.002	0.037	0.042	0.188				
	3	0.005	0.057	0.056	0.185				
5	0	0.899	0.300	0.899	0.300	0.941	0.350	0.941	0.350
	1	0.094	0.561	0.094	0.561	0.059	0.650	0.059	0.650
	2	0.005	0.109	0.005	0.109				
	3	0.002	0.029	0.002	0.029				
6	0	0.818	0.385	0.818	0.385	0.850	0.420	0.850	0.420
	1	0.143	0.262	0.143	0.262	0.150	0.580	0.150	0.580
	2	0.004	0.072	0.004	0.072				
	3	0.035	0.281	0.035	0.281				
7	0	0.665	0.065	0.665	0.065	0.737	0.075	0.737	0.075
	1	0.282	0.514	0.282	0.514	0.263	0.925	0.263	0.925
	2	0.053	0.339	0.053	0.339				
	3	0.001	0.082	0.001	0.082				
8	0	0.509	0.108	0.509	0.108	0.553	0.119	0.553	0.119
	1	0.484	0.713	0.484	0.713	0.447	0.881	0.447	0.881
	2	0.007	0.128	0.007	0.128				
	3	0.001	0.052	0.001	0.052				
9	0	0.979	0.857	0.871	0.591	0.984	0.899	0.872	0.648
	1	0.019	0.123	0.119	0.337	0.016	0.101	0.128	0.352
	2	0.001	0.015	0.006	0.029				
	3	0.001	0.005	0.004	0.042				
10	0	0.890	0.747	0.688	0.362	0.899	0.774	0.698	0.435
	1	0.094	0.162	0.170	0.364	0.101	0.226	0.302	0.565
	2	0.002	0.009	0.026	0.098				
	3	0.013	0.081	0.116	0.176				
11	0	0.677	0.677	0.228	0.228	0.698	0.698	0.242	0.242
	1	0.244	0.244	0.520	0.520	0.302	0.302	0.758	0.758
	2	0.007	0.007	0.156	0.156				
	3	0.073	0.073	0.096	0.096				

Tabela 22 – Probabilidades estimadas com os modelos G-DINA e MGDCP para o conjunto de dados do BDI

(conclusão)									
item	h	MGDCP				G-DINA			
		θ_{j1h}	θ_{j2h}	θ_{j3h}	θ_{j4h}	θ'_{j1h}	θ'_{j2h}	θ'_{j3h}	θ'_{j4h}
12	0	0.859	0.698	0.424	0.319	0.888	0.734	0.435	0.349
	1	0.126	0.277	0.467	0.430	0.112	0.266	0.565	0.651
	2	0.014	0.019	0.105	0.188				
	3	0.001	0.006	0.004	0.062				
13	0	0.840	0.593	0.529	0.214	0.854	0.683	0.585	0.223
	1	0.107	0.287	0.320	0.264	0.146	0.317	0.415	0.777
	2	0.047	0.091	0.118	0.389				
	3	0.006	0.029	0.033	0.133				
14	0	0.822	0.606	0.579	0.393	0.844	0.610	0.628	0.420
	1	0.102	0.187	0.224	0.193	0.156	0.390	0.372	0.580
	2	0.059	0.101	0.146	0.279				
	3	0.018	0.105	0.051	0.135				
15	0	0.803	0.493	0.389	0.131	0.819	0.556	0.494	0.131
	1	0.183	0.427	0.511	0.437	0.181	0.444	0.506	0.869
	2	0.011	0.067	0.092	0.354				
	3	0.003	0.012	0.009	0.079				
16	0	0.658	0.658	0.290	0.290	0.681	0.681	0.294	0.294
	1	0.303	0.303	0.533	0.533	0.319	0.319	0.706	0.706
	2	0.035	0.035	0.105	0.105				
	3	0.004	0.004	0.072	0.072				
17	0	0.573	0.573	0.153	0.153	0.599	0.599	0.157	0.157
	1	0.413	0.413	0.535	0.535	0.401	0.401	0.843	0.843
	2	0.012	0.012	0.245	0.245				
	3	0.002	0.002	0.067	0.067				
18	0	0.853	0.681	0.664	0.508	0.876	0.720	0.674	0.525
	1	0.130	0.271	0.294	0.291	0.124	0.280	0.326	0.475
	2	0.016	0.035	0.032	0.141				
	3	0.001	0.013	0.011	0.060				
19	0	0.928	0.928	0.842	0.842	0.929	0.929	0.850	0.850
	1	0.060	0.060	0.128	0.128	0.071	0.071	0.150	0.150
	2	0.008	0.008	0.015	0.015				
	3	0.004	0.004	0.016	0.016				
20	0	0.809	0.675	0.606	0.456	0.829	0.705	0.619	0.475
	1	0.179	0.273	0.355	0.414	0.171	0.295	0.381	0.525
	2	0.011	0.043	0.033	0.091				
	3	0.001	0.009	0.005	0.038				
21	0	0.869	0.869	0.614	0.614	0.878	0.878	0.626	0.626
	1	0.097	0.097	0.258	0.258	0.122	0.122	0.374	0.374
	2	0.026	0.026	0.101	0.101				
	3	0.009	0.009	0.027	0.027				

Fonte: Dados da pesquisa.

Outro aspecto importante de se destacar é a relação existente entre o modelo G-DINA e o MGDCP e seus respectivos parâmetros dos itens. Baseado na definição dos modelos, sem a inclusão de restrições, haveria uma perfeita concordância entre o G-DINA e o MGDCP dentro da formulação adotada nessa aplicação, no sentido de que a medida de probabilidade associadas às respostas de *feedback* negativo (ausência de sintoma) e positivo (presença de sintomas) são exatamente as mesmas nos dois modelos para todos os itens e todas classes latentes. Em termos matemáticos, temos, nesse caso, que $P_{j0}^{MGDCP}(\boldsymbol{\alpha}_{(c)}) = Pr(Y_j = 0 | \boldsymbol{\alpha} = \boldsymbol{\alpha}_{(c)}) = Pr(Z_j = 0 | \boldsymbol{\alpha} = \boldsymbol{\alpha}_{(c)}) = P_{j0}^{GDINA}(\boldsymbol{\alpha}_{(c)})$ para todo $c = 1, \dots, 4$ e todo $j = 1, \dots, J$, onde Y_j representa a resposta do item j no formato original e X_j , a resposta dicotomizada.

Porém, como são utilizados conjuntos de restrições diferentes, a concordância entre o modelo G-DINA e o MGDCP mencionada anteriormente não é garantida. Sendo assim, podemos avaliar a concordância entre esses modelos, comparando os valores estimados de $P_{j0}(\boldsymbol{\alpha}_{(c)})$ com ambos os modelos, para todo c e j . Quanto mais próximos são esses valores, maior é a concordância entre os modelos.

Podemos observar, na Figura 7, que, na maioria dos itens, os valores estimados da probabilidade do nível de resposta 0 são bem próximos entre os dois modelos. A diferença média absoluta entre essas probabilidades é de aproximadamente 0.03, indicando uma grande concordância entre os modelos, mesmo tendo diferentes conjuntos de restrições aplicadas em cada modelo.

O próximo ponto a ser analisado é a classificação dos indivíduos de acordo com os modelos. A Tabela 23 apresenta o cruzamento da classificação dos indivíduos segundo o perfil latente estimado pelos modelos G-DINA e MGDCP.

Tabela 23 – Tabela de contingência do perfil latente estimado pelos modelo G-DINA e MGDCP aos respondentes do teste BDI

		MGDCP				Total
		$\hat{\boldsymbol{\alpha}} = (0, 0)$	$\hat{\boldsymbol{\alpha}} = (1, 0)$	$\hat{\boldsymbol{\alpha}} = (0, 1)$	$\hat{\boldsymbol{\alpha}} = (1, 1)$	
G-DINA	$\hat{\boldsymbol{\alpha}} = (0, 0)$	426	0	5	0	431
	$\hat{\boldsymbol{\alpha}} = (1, 0)$	48	118	9	12	187
	$\hat{\boldsymbol{\alpha}} = (0, 1)$	41	3	156	4	204
	$\hat{\boldsymbol{\alpha}} = (1, 1)$	0	28	37	224	289
Total		515	149	207	240	1111

Fonte: Dados da pesquisa.

Podemos observar, na Tabela 23, que a maioria dos indivíduos é classificada igualmente para os dois modelos (diagonal principal da tabela), mas há diferenças consideráveis. Há 154 indivíduos, que são diagnosticados com mais sintomas com o modelo G-DINA do que com o MGDCP (isto é, mudam de não sintomático para sintomático; ou de sintomático, em uma dimensão, para sintomático em ambas as dimensões), enquanto o oposto ocorre apenas com 21 indivíduos. Essas diferenças podem ser consequência do processo de dicotomização e podem ser discutidas mais detalhadamente pelos especialistas.

É importante observar que a dicotomização das respostas desconsidera a gravidade dos sintomas declarados pelos indivíduos, o que pode afetar consideravelmente o seu diagnóstico, conforme ilustrado na Tabela 24.

Nessa tabela são apresentadas as respostas de dois indivíduos e a probabilidade a posteriori de pertencer a cada classe latente, segundo os modelos MGDCP e G-DINA. Para fins de concisão, os itens em que ambos os indivíduos responderam igualmente foram omitidos.

Tabela 24 – Tabela comparativa entre dois indivíduos

	j	Indivíduo A		Indivíduo B	
		X_{ij}	Y_{ij}	X_{ij}	Y_{ij}
Resposta do item	1	1	2	0	1
	2	1	2	1	3
	6	1	2	1	4
	7	1	2	1	3
	15	1	2	0	1
	16	1	2	1	3
	α	G-DINA	MGDCP	G-DINA	MGDCP
Probabilidade a posteriori	(0,0)	0.03	0.40	0.34	0.35
das classes latentes	(1,0)	0.46	0.29	0.56	0.55
$Pr(\alpha_i = \alpha \mathbf{Y})$	(0,1)	0.10	0.27	0.09	0.07
classe latente estimada	(1,1)	0.42	0.04	0.01	0.02
		(1,0)	(0,0)	(1,0)	(1,0)

Nota: (a) o indivíduo A e B correspondem, respectivamente, aos indivíduos $i = 1046$ e $i = 589$ no conjunto de dados. (b) os itens cujo a resposta original foi a mesma para ambos indivíduos foram omitidas.

Fonte: Dados da pesquisa.

Analisando a tabela 24 e considerando apenas os dados dicotomizados (X_{ij}), podemos notar que o indivíduo A relatou mais sintomas do que o indivíduo B, já que todas as respostas do primeiro (dicotomizadas) são maiores ou iguais as do segundo. Dessa forma, o primeiro tem maior probabilidade de ser sintomático em, pelo menos, uma dimensão, em comparação com o segundo, segundo o modelo G-DINA. Nesse caso, o indivíduo A tem uma probabilidade de apenas 3% de não ser sintomático, contra 34% do indivíduo B.

Porém, considerando os dados politômicos (Y_{ij}), observamos uma tendência diferente dos dados. O indivíduo A aparenta ter sintomas menos intensos de depressão do que o indivíduo B, já que as respostas do primeiro são menores ou iguais as do segundo em todos os itens, exceto nos itens 1 (tristeza) e 15 (inibição do trabalho). Consequentemente, a probabilidade estimada do indivíduo A não ser depressivo (40%) é maior do que a do indivíduo B (35%), segundo o MGDCP.

Além disso, se consideramos apenas a estimativa pontual do perfil latente, ao trocar o modelo G-DINA com respostas dicotomizadas para o MGDCP com as respostas originais, a classificação do indivíduo A muda de sintomático na dimensão cognitiva para não sintomático,

enquanto a classificação do indivíduo B permanece a mesma.

Concluindo, dado o alto nível de concordância na distribuição da probabilidade de resposta entre os dois modelos, podemos inferir que as diferenças observadas se devem, principalmente, ao tipo de dados considerados. Ademais, como a dicotomização representa uma perda de informações, é razoável supor que a classificação obtida pelo MGDCP com dados politômicos seja mais precisa, porque se baseia em todas as informações disponíveis no teste. Portanto, a utilização do modelo MGDCP com os dados originais deve ser privilegiada, apesar de sua maior complexidade.

4.5 Discussão

Neste capítulo foram analisados dados referentes ao Inventário de Depressão de Beck, aplicando-se uma metodologia para a classificação da depressão, que utiliza modelos de diagnóstico cognitivo no processo de modelagem. Tal metodologia foi proposta por [Silva et al. \(2018\)](#), utilizando o modelo DINA. Neste estudo, apresentamos duas adaptações dessa metodologia, utilizando os modelos G-DINA e MGDCP no lugar do modelo originalmente empregado.

Na primeira análise, foram comparados os resultados obtidos com o modelo DINA e G-DINA. Para esses modelos, as respostas são dicotomizadas de acordo com a presença ou ausência de sintomas do item. Os resultados dessa análise indicaram que o modelo G-DINA é mais apropriado para essa aplicação e que o modelo DINA tende a superestimar os diagnósticos de depressão.

Na segunda análise, foram comparados os resultados obtidos com o modelo G-DINA, para os dados dicotomizados e com o MGDCP, para os dados politômicos originais. Os dois modelos apresentaram bons resultados, respeitadas as diferenças inerentes a cada modelo, como a interpretação dos parâmetros. Contudo, podemos observar que, embora a dicotomização das respostas seja uma opção interessante, ela afeta consideravelmente o diagnóstico de alguns indivíduos. Além disso, caso houvessem apenas duas categorias de respostas para cada item, não é possível assegurar que as respostas dadas pelos indivíduos com essa nesse contexto coincidiriam com as respostas dicotomizadas à partir do teste original, com quatro categorias de respostas por item. Por esse motivo, defendemos que, nas aplicações de MDC para fins de diagnóstico, a dicotomização deve ser evitada ou utilizada com cautela, especialmente, se os dados politômicos originais estiverem disponíveis. Em vez disso, incentivamos, nesse tipo de situação, o uso do MGDCP.

Por fim, sobre a metodologia de classificação de depressão apresentada, é importante destacar que o enfoque deste trabalho foi no processo de modelagem e a análise estatística. Para a utilização dessa metodologia para o diagnóstico médico, faz-se necessário, conforme ressaltado por [Silva et al. \(2018\)](#), uma investigação aprofundada, envolvendo especialistas em psiquiatria, além da realização de estudos de validação dos resultados.

APLICAÇÃO - SISTEMA DE RECOMENDAÇÃO

5.1 Introdução

O avanço da tecnologia e o grande aumento na oferta e na utilização de serviços online, dos mais diversos setores da economia proporcionam ao usuário acesso a uma infinidade de produtos, serviços, conteúdos digitais, etc. Essa sobrecarga de dados dificulta o processo de encontrar conteúdo relevante e útil na Internet (ROY; DUTTA, 2022), tornando os métodos de filtragem de informações indispensáveis.

Os sistemas de recomendação atuam como ferramentas de filtragem capazes de oferecer aos usuários, conteúdos ou informações personalizadas, reduzindo o esforço do usuário e o tempo necessário para resultados que sejam relevantes (ROY; DUTTA, 2022). Esses sistemas buscam processar um vasto conjunto de informações coletadas (avaliações online, páginas visitadas, tempo de permanência em uma página, produtos vistos em uma loja online, etc) para apresentar a cada usuário uma lista personalizada de itens que possam lhe interessar. Devido ao seu grande potencial, esses sistemas de recomendação estão sendo cada vez mais usados em um grande número de aplicações, como Web, livros, cursos online, filmes, músicas, comércio eletrônico, etc (ROY; DUTTA, 2022).

Neste capítulo, apresentamos uma aplicação na qual um modelo de diagnóstico cognitivo é utilizado como peça central. Trata-se de um sistema de recomendação para filmes, que utiliza o MGDCP para criar uma lista de recomendação de filmes personalizada de acordo com perfil latente desse indivíduo.

Ressaltamos que a maior parte do conteúdo desse capítulo foi desenvolvido em colaboração com outros pesquisadores e foi publicado em Marana *et al.* (2024). Contudo, nesta tese são fornecidos maiores detalhes do modelo matemático e do tratamento estatístico utilizado, em

comparação ao artigo.

5.2 O sistema de recomendação com MGDCP

Nessa seção é apresentado o sistema de recomendação de filmes que utiliza o MGDCP, conforme proposto em [Marana et al. \(2024\)](#).

Primeiro, apresentaremos os detalhes da modelagem. No contexto dessa aplicação, cada filme representa um item. As classificações dadas aos filmes representam as respostas; os gêneros cinematográficos representam os atributos; e o perfil latente do indivíduo é representado por um vetor binário, indicando quais gêneros cinematográficos esse indivíduo aprecia e quais não.

Considere uma base de dados com J itens e N usuários. Seja Y a matriz resposta de dimensão $N \times J$, onde $y_{ij} = h$ se o indivíduo i ($= 1, \dots, N$) avaliou o filme j ($= 1, \dots, J$) com a nota h ($= 1, \dots, H$); considere adicionalmente, que $y_{ij} = 0$ indica que o indivíduo i não avaliou o item j (isto é, 0 é utilizado para representar dados faltantes - *missing data*).

Adicionalmente, considere a matriz Q de dimensão $J \times K$ com elementos q_{jk} , em que $q_{jk} = 1$ indica que o j -ésimo filme ($j = 1, \dots, J$) se enquadra dentro do k -ésimo gênero. Além disso, para cada $i = 1, \dots, N$, $\alpha_i = (\alpha_{i1}, \dots, \alpha_{iK})^T$ é um vetor de dimensão K , onde $\alpha_{ik} = 1$ indica que o indivíduo i aprecia o k -ésimo ($k = 1, \dots, K$) gênero cinematográfico, e $\alpha_{ik} = 0$, o oposto.

A primeira etapa para a utilização do sistema de recomendação consiste em ajustar o modelo MGDCP, utilizando os dados relativos às avaliações dos usuários e classificação por gêneros dos filmes. Obtém-se assim as estimativas dos parâmetros dos itens $\hat{\theta}_{j|h}$ e o vetor de proporções π .

Com base nos parâmetros estimados do modelo, obtém-se, para cada indivíduo, uma estimativa da distribuição a posteriori de α_i , por meio da fórmula:

$$\hat{\gamma}_i(\alpha_{(c)}) = Pr(\alpha_i = \alpha_{(c)}) = \frac{\prod_{j \in O_i} Pr(Y_j = y_{ij} | \alpha_i = \alpha_{(c)}) \cdot \pi_c}{\sum_{c=1}^C \prod_{j=1}^J Pr(Y_j = y_{ij} | \alpha_i = \alpha_{(c)}) \cdot \pi_c}$$

em que O_i representa o conjunto de itens com respostas observadas do indivíduo i (isto é, o conjunto de filmes que foram avaliados por esse usuário) e $c = 1, \dots, C$ representa a classe latente.

Utilizando as estimativas da distribuição a posteriori de α_i e dos parâmetros dos itens, pode-se obter a predição da nota para os itens não observados, denotado, doravante, por \hat{y}_{ij} . Para isso, consideramos a probabilidade de resposta ao item j , dado as respostas observadas do indivíduo i , dado função apresentada na Equação 5.1:

$$r_{ij}(h) = Pr(Y_j = h | \mathbf{y}_i^{obs}) = \sum_{c=1}^C Pr(Y_j = h | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{(c)}) \cdot \hat{\gamma}_i(\boldsymbol{\alpha}_{(c)}) \quad (5.1)$$

em que \mathbf{y}_i^{obs} representa o vetor de todas as respostas observadas do indivíduo i ; e $Pr(Y_j = h | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{(c)})$ é calculada com base nas estimativas dos parâmetros itens obtidas anteriormente.

Com base nessa função de probabilidade, obtemos \hat{y}_{ij} como a nota esperada do item j para o indivíduo i , calculada pela fórmula:

$$\hat{y}_{ij} = \sum_{h=1}^{H_j} h \cdot r(h)$$

Por fim, o sistema de recomendação, baseado no MGDPCP, atua da seguinte forma: para cada indivíduo de interesse i , obtém-se as estimativas de notas \hat{y}_{ij} para todo item j ainda não avaliado e retorna um vetor formado pelos índices j dos B , itens com maiores notas estimadas da lista mencionada.

Convém ressaltar que, para essa aplicação, foi utilizado o método de estimação clássico, utilizando o algoritmo EM-MMLE. Embora, em teoria, seja possível empregar o método de estimação Bayesiana nessa aplicação, a alta dimensionalidade do conjunto de dados analisados e a alta demanda computacional tornariam a execução do algoritmo impraticável.

5.3 Outros sistemas de recomendação

Nesta seção, descrevemos os sistemas de recomendação, considerados para comparação com o método proposto.

O primeiro é o *Content Filtering of Gender* (CFG) (BOBADILLA *et al.*, 2011), um método tradicional que será descrito em detalhes mais adiante.

O segundo e o terceiro são o *Neural Network with Matrix Factorization* (NeuMF) e o *Variational Autoencoder* (VAE). Esses dois métodos modernos utilizam técnicas de aprendizado de máquina com redes neurais. Detalhes sobre esses dois métodos podem ser encontrados em Microsoft (2023), He *et al.* (2017), Liu *et al.* (2022).

O CFG é um sistema de recomendação da classe *Content-Based Filtering* (CBF), que busca recomendar a cada usuário itens similares àqueles que demonstraram interesse. Para isso, é calculado a similaridade entre todos os pares de itens, que é utilizada para atribuir diferentes pesos às notas. Isto é, dados dois itens, j_1 e j_2 , quanto maior a similaridade entre eles, maior é o a influência da nota de j_1 na nota estimada de j_2 .

Embora outras metodologias de cálculo possam ser utilizadas, reproduzimos aqui a metodologia presente em Marana *et al.* (2024), na qual o algoritmo *K-nearest neighbors* (KNN) é utilizado com a métrica de *Jaccard* para estimar as notas das avaliações dos usuários.

Considere a matriz Q , definida como na seção anterior. Para quaisquer $j_1, j_2 \in \{1, \dots, J\}$, $j_1 \neq j_2$, definimos o coeficiente de similaridade de *Jaccard* entre j_1 e j_2 pela seguinte fórmula:

$$u(j_1, j_2) = \frac{\sum_{k=1}^K q_{j_1 k} \wedge q_{j_2 k}}{\sum_{k=1}^K q_{j_1 k} \vee q_{j_2 k}}$$

em que \wedge e \vee representam, respectivamente, os operadores lógicos “e” e “ou”, que retorna o valor 1 para valores verdadeiros e 0, caso contrário.

Seja $\kappa < J$ o número de vizinhos considerados no algoritmo KNN. Seja, para cada $j \in \{1, \dots, J\}$, R_j o conjunto dos κ itens mais similares ao item j . Isto é, $R_j = \{j_1, \dots, j_\kappa\}$ onde $u(j, j_1), \dots, u(j, j_\kappa)$ são os κ maiores valores do conjunto $\{u(j, j') : j' \in \{1, \dots, J\}, j' \neq j\}$. A predição de y_{ij} é dada pela fórmula:

$$\hat{y}_{ij} = \frac{\sum_{j' \in R_j} u(j, j') \cdot y_{ij'}}{\sum_{j' \in R_j} u(j, j')} \quad (5.2)$$

Alternativamente, caso não seja possível obter R_j , por não haver uma quantidade suficiente de itens similares (isto é, caso j seja tal que $\sum_{j' \neq j} u(j, j') < \kappa$), a predição da nota do indivíduo i para o item j é calculada pela seguinte fórmula:

$$\hat{y}_{ij} = \frac{1}{T} \sum_{i'=1}^N \sum_{j'=1}^J y_{i'j'} - \frac{1}{T \cdot j} \sum_{i'=1}^N y_{i'j} + \frac{1}{T_i} \sum_{j'=1}^J y_{ij'} \quad (5.3)$$

em que $T = \sum_{i=1}^N \sum_{j=1}^J \mathbb{I}(y_{ij} \neq 0)$, $T_{\cdot j} = \sum_{i=1}^N \mathbb{I}(y_{ij} \neq 0)$ e $T_i = \sum_{j=1}^J \mathbb{I}(y_{ij} \neq 0)$ representam, respectivamente, o número de notas observadas para todos itens e indivíduos, para o item j , e para o indivíduo i .

5.4 Metodologia para avaliação dos métodos

O conjunto de dados foi dividido em três subconjuntos: treino, validação e teste.

O primeiro é utilizado para ajustar o modelo e obter estimativas para os parâmetros dos itens e para o vetor de proporções. O segundo é requerido apenas para os métodos NeuMF e VAE, não sendo utilizado nos métodos CFG e no sistema de recomendação com o MGDPCP. O terceiro é utilizado para avaliar a acurácia das predições e a precisão na recomendação dos itens.

Para realizar as avaliações citadas no final do parágrafo anterior, foi realizado um procedimento de mascaramento aleatorizado dos dados do conjunto de treino, no qual, para cada indivíduo i , metade de suas avaliações são sorteadas ao acaso e omitidas. Os dados não omitidos são utilizados para realizar a predição dos demais itens, isto é, aqueles que não receberam uma avaliação ou que a avaliação foi omitida.

5.4.1 Medidas utilizadas na avaliação dos métodos

Nesta seção, descrevemos as medidas utilizadas para avaliar os diferentes sistemas de recomendação.

RMSE e MAE

O *Root of Mean Squared Error* ($RMSE_y$) e o *Mean Absolute Error* (MAE_y) foram calculados com base nas notas estimadas para cada indivíduo para os itens que tiveram as notas omitidas, utilizando as seguintes fórmulas:

$$RMSE_y = \sqrt{\frac{\sum_{i=1}^N \sum_{j \in O_i} (\hat{y}_{ij} - y_{ij})^2}{T}} \quad (5.4)$$

$$MAE_y = \frac{\sum_{i=1}^N \sum_{j \in O_i} |\hat{y}_{ij} - y_{ij}|}{T} \quad (5.5)$$

em que \hat{y}_{ij} e y_{ij} representam, respectivamente, a nota estimada e a nota observada do indivíduo i para o item j ; O_i é o conjunto de itens com respostas observadas do indivíduo i e T , o número total de notas observadas.

Note que somente é possível avaliar tais medidas para os métodos que fornecem, em alguma de suas etapas, estimativas das notas dos usuários, que, no caso, restringem-se ao método de classificação com o MGDCP e o CFG.

MAP

O *Mean Average Precision at k* (denotado por $MAP(k)$ ¹) representa a precisão média da relevância das k -ésimas primeiras recomendações. No contexto dessa aplicação, tal valor é calculado por meio da seguinte fórmula:

$$MAP(k) = \frac{1}{N} \sum_{i=1}^N AP_i(k) \quad (5.6)$$

$$AP_i(k) = \sum_{t=1}^k \frac{1}{k} \mathbb{I}(y_{ij_{it}} > y_0)$$

em que j_{it} e $y_{ij_{it}}$ representam, respectivamente, o t -ésimo item recomendado para o indivíduo i e a avaliação dada (isto é, o valor observado de fato) por esse usuário para esse item; e y_0 é uma constante que representa a menor nota considerada como uma avaliação positiva. Isto é, y_0 representa a menor nota de avaliação que um item deve receber para ser considerado como relevante ao usuário, para a avaliação da medida $MAP(k)$. Nessa aplicação, y_0 foi fixado com o valor 3.5.

¹ A notação original e comumente utilizada é $MAP@k$, onde k representa o número de recomendações avaliadas. Optamos, porém, por utilizar uma notação alternativa $MAP(k)$, por se assemelhar mais às notações utilizadas na Estatística.

5.5 Os conjuntos de dados

Para avaliar os métodos de recomendação propostos, dois conjuntos de dados foram avaliados, *MovieLens 20Me* e a *Anime Recommendation Database* que, doravante, serão denominados, DB1 e DB2. Cada conjunto de dados é formado por duas matrizes, sendo que a primeira contém as informações sobre os gêneros cinematográficos de cada item presente na base de dados (utilizada para definir a matriz Q utilizada na formulação do modelo MDC) e a segunda contém as avaliações dadas pelos usuários a cada um desses itens, sendo os que não foram avaliados, por um determinado usuário, denotados por *NA*.

Nesse estudo, foi realizado uma filtragem preliminar dos dados, a fim de tornar a obtenção e análise dos dados factíveis em um contexto de recursos computacionais moderado. De fato, a aplicação foi inteiramente executada em um computador pessoal. Nessa filtragem, foram selecionados os gêneros e itens mais populares (isto é, aqueles com maiores números de avaliações), tendo sido selecionados quatro gêneros e 1680 itens do BD1 e quatro gêneros e 1457 itens do BD2.

Além disso, para a análise do BD1, foram selecionados, ao acaso, 6978 indivíduos para treino; 1000, para validação; e 1000, para teste. Para o BD2, as quantidades de indivíduos para os conjuntos treino, validação e teste foram, respectivamente, 6540, 1000 e 1000.

5.6 Inclusão da popularidade no sistema de recomendação com MGDCP

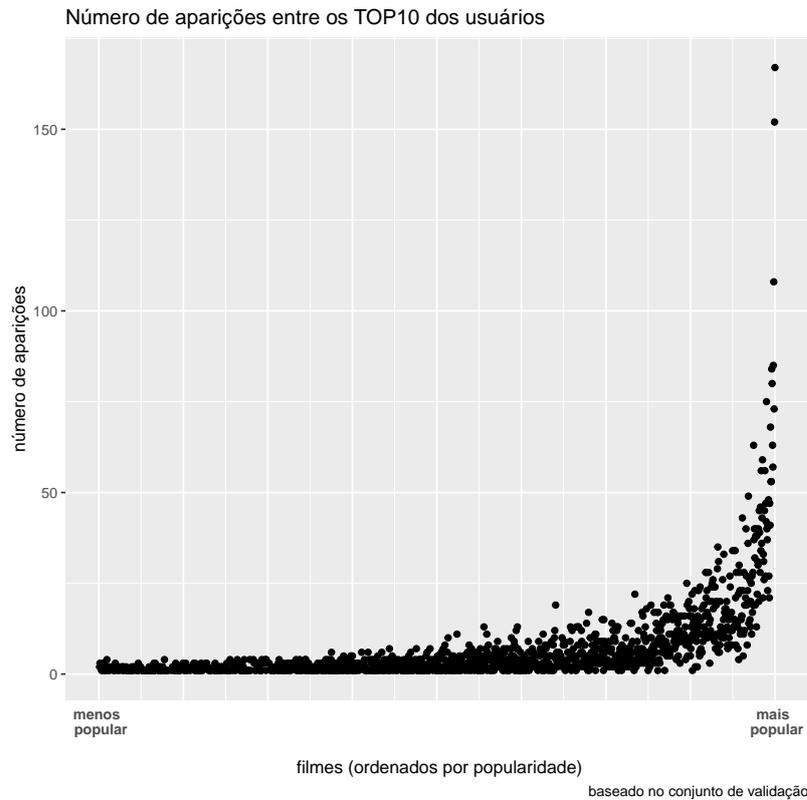
Nesta seção, analisaremos o impacto da popularidade na taxa de aprovação de um item e apresentaremos uma estratégia que visa a aprimorar o sistema de recomendação com o MGDCP, mediante a inclusão do fator popularidade dos itens.

Neste trabalho, adotamos o número de avaliações recebidas como medida de popularidade de um item. A figura 8 apresenta o número de aparições de um item na lista dos 10 mais bem avaliados em função do número total de avaliações recebidas. Esses gráfico foi baseado no conjunto de validação do BD1. Além disso, para facilitar a visualização, os itens (eixo horizontal) foram ordenados em ordem crescente de número de avaliações.

Já a Figura 9 apresenta dados semelhantes, porém, baseados nas predições do MGDCP. Isto é, o gráfico traz o número de aparições na lista dos 10 itens mais bem avaliados, segundo a predição do MGDCP, em função da popularidade dos itens.

Interpretando a Figura 8, podemos observar que os filmes mais populares figuram entre os 10 melhores avaliados pelo usuário mais frequentemente, sendo que a relação entre popularidade de um item e o seu número de aparições entre os 10 mais bem avaliados é aproximadamente exponencial.

Figura 8 – Distribuição dos 10 filmes mais bem avaliados no conjunto de validação de DB1



Fonte: Adaptada de [Marana et al. \(2024\)](#).

Uma possível justificativa para esse comportamento é que um filme torna-se popular, quando possui uma grande aceitação pelo público em geral, recebendo boas críticas de usuários e dos meios de comunicação, o que estimula novos usuários a assistirem tal filme. Dessa forma, os filmes mais populares apresentam, normalmente, alta taxa de aprovação dentre os diversos grupos de usuários com diferentes perfis.

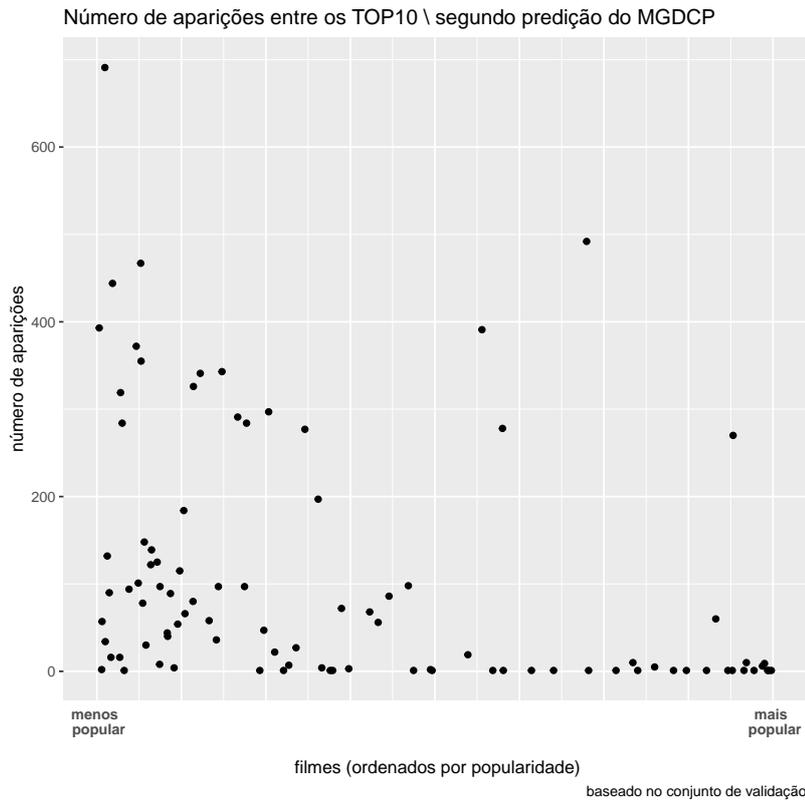
Por outro lado observa-se, na Figura 9, que não há uma influência direta da popularidade na nota estimada pelo MGDCP e, conseqüentemente, nos itens recomendados pelo sistema baseado nesse modelo. Ademais, alguns filmes, que são pouco populares e que raramente aparecem entre os 10 mais bem avaliados entre os usuários, aparecem frequentemente na lista dos 10 mais bem avaliados, segundo a classificação do MGDCP.

Buscando compreender melhor o motivo dessa disparidade, analisamos as notas estimadas pelo MGDCP para cada indivíduo.

A figura 10 apresenta as notas estimadas pelo MGDCP para todos os filmes para três usuários, com os filmes ordenados por popularidade. Nesses gráficos são destacados os n filmes ($n = 10, 50, 100$) com maior nota predita pelo MGDCP, que, chamaremos de TOP10, TOP50 e TOP100.

Note que a diferença entre a nota mais baixa do TOP10 e do TOP50 desses três usuários

Figura 9 – Distribuição dos 10 filmes mais bem avaliados pelos usuários segundo as notas estimadas pelo MGDCP



Fonte: Adaptada de [Marana et al. \(2024\)](#).

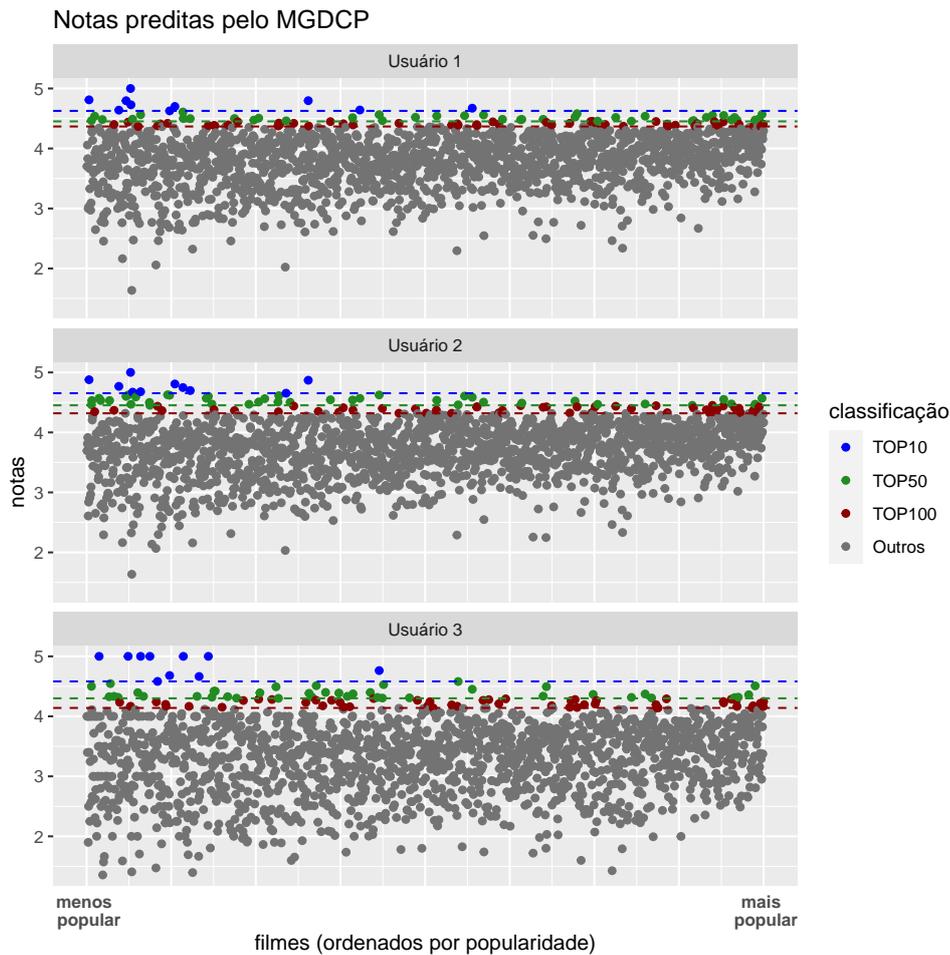
é muito pequena, sendo, ainda menor, a diferença entre as menores notas do TOP50 e o TOP100. Desse modo, pequenas diferenças na nota preditas pelo MGDCP (que podem ser, inclusive, estatisticamente insignificante) podem representar uma grande variação na classificação dos itens, influenciando fortemente a lista de recomendações resultantes.

Buscando superar essa característica indesejada, foi proposto um novo sistema de recomendação que combina as classificações obtidas pelo MGDCP e a popularidade dos itens, sendo denominado *MGDCP melhorado* (*Enhanced GPDM*, em inglês).

Nesse novo sistema de recomendação, as notas dos itens são estimadas com o MGDCP, como anteriormente. No entanto, para a formação da lista de recomendações, tomamos, para cada indivíduo i , uma lista preliminar de recomendações com BP itens e essa lista é, então, ordenada de acordo com a popularidade dos itens e os BF itens mais populares são recomendados para o indivíduo i .

Para definir o número de itens da primeira lista, consideramos um ponto de corte $p \in (0, 1)$, definido de modo que $BP = p \cdot J$, onde J é o número total de itens na base de dados. Note que o valor de BP reflete a influência da popularidade na lista de recomendações, sendo que $p = 1$ corresponde a uma recomendação baseada puramente no critério popularidade, enquanto $p = 1/N$ corresponde a recomendar os itens, baseados apenas nas estimativas do MGDCP.

Figura 10 – Distribuição das notas previstas pelo MGDCP de três usuários para o BD1



Fonte: Adaptada de [Marana et al. \(2024\)](#).

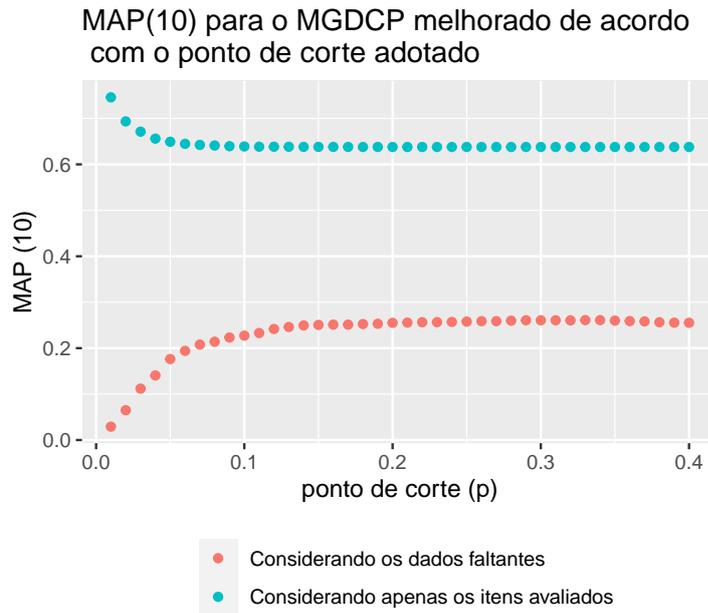
Para essa aplicação, a escolha de p foi baseada em resultados empíricos obtidos a partir do conjunto de validação. Foi realizado uma varredura, tomando p variando de 1% a 40%, de 1% em 1%, avaliando o impacto causado no desempenho do MGDCP melhorado segundo o critério MAP(10).

A Figura 11 apresenta o valor de MAP(10) obtido com o MGDCP melhorado, utilizando diferentes pontos de corte p para conjunto de dados de validação do DB1.

Podemos observar, na Figura 11, que, à medida que aumentamos o ponto de corte, a precisão, considerando-se os valores ausentes, tende a aumentar, enquanto, a precisão, considerando-se apenas os itens classificados, tende a diminuir. Além disso, observamos que os valores do MAP(10) tendem a se estabilizar, depois de algum ponto p , e o ganho efetivo do aumento do valor do ponto de corte tende a ser insignificante em algum ponto.

Por essa razão, escolhemos como ponto de corte o menor valor de p (13%), de modo a que o incremento na precisão (em termos de MAP(10)) seja inferior a um valor fixo δ . Nesse trabalho, foi adotado $\delta = 0.005$. Note que a escolha desses valores depende de escolhas subjetivas

Figura 11 – MAP(10) do MGDCP melhorado para diferentes pontos de corte



Fonte: Adaptada de [Marana et al. \(2024\)](#).

e outros critérios, para escolha de p , poderiam ter sido adotados.

5.7 Resultados

Na primeira análise, avaliamos a acurácia na predição das notas obtidas com o MGDCP para os conjuntos de dados DB1 e DB2. Para efeito de comparação, os resultados da acurácia desse modelo são comparadas com o método de predição CFG, adotando valores de k igual a 50, 100 e 200.

Ressaltamos que os métodos NeuMF e VAE não figuram nesse comparativo, pois não fazem predições das notas (de maneira explícita, ao menos). Além disso, as predições do método MGDCP melhorado são as mesmas do método MGDCP, não havendo assim a necessidade de reportar novamente tais valores.

Na Tabela 25 são apresentados os resultados do RMSE e MAE das predições das notas preditas pelo MGDCP e CFG, conforme descrito no parágrafo anterior.

Podemos observar, na Tabela 25, que as predições do MGDCP são mais precisas do que as obtidas com o CFG em todos os casos analisados, apresentando valores menores de MAE e RMSE nos dois conjuntos de dados.

Aqui se faz pertinente um comentário adicional sobre os resultados apresentados pelo método CFG para os diferentes valores de k utilizados. Quando não há itens semelhantes suficientes ao item em avaliação, o método CFG calcula a nota baseado-se nas médias geral das

Tabela 25 – Acurácia na predição das notas para os conjuntos de Dados DB1 e DB2

método	DB1		DB2	
	MAE	RMSE	MAE	RMSE
MGDCP	0.662	0.961	0.467	0.748
CFG (k = 50)	0.806	1.034	0.665	0.850
CFG (k = 100)	0.873	1.111	0.712	0.901
CFG (k = 200)	0.927	1.173	0.735	0.926

Fonte: Adaptada de [Marana et al. \(2024\)](#).

notas, do usuário e do item em questão. Pode-se observar, na Tabela 25, que quanto maior o valor de k , pior é o desempenho do método CFG. Tal resultado pode ser explicado pela esparsidade dos conjuntos de dados e ao valor de k adotado, o que acarreta em pequeno número de itens que satisfazem o critério para do algoritmo KNN (fórmula da Equação 5.2), sendo utilizado, nesses casos, a fórmula alternativa da Equação 5.3 ([MARANA et al., 2024](#)).

Nesse estudo, utilizamos a medida $MAP(k)$ para avaliar a relevância das recomendações de cada sistema avaliado. Para isso, consideramos duas metodologias distintas.

Na primeira, os itens não avaliados por um determinado usuário foram desconsiderados, isto é, os dados ausentes foram filtrados e as listas de itens avaliadas no $MAP(k)$ foram formadas exclusivamente por itens cujas avaliações foram omitidas. Na segunda, as listas de recomendações incluem os itens com avaliações omitidas e os que não foram avaliados (*missing data*).

Veja que a primeira medida é uma aproximação otimista do $MAP(k)$ com dados completos, pois é calculada apenas dentre os itens avaliados, que, no conjunto de dados analisados, possuem mais avaliações positivas (isto é, acima de 3.5) do que negativas. Por outro lado, o cálculo realizado com a inclusão dos dados ausentes é uma aproximação pessimista do $MAP(k)$ com dados completos, pois, nesse caso, os itens que são recomendados, mas não foram avaliados, são considerados como um item com avaliação negativa, o que não reflete necessariamente a realidade. De qualquer forma, todos os modelos são comparados, seguindo o mesmo critério, portanto, as inferências sobre os resultados do $MAP(k)$, apresentadas a seguir, são válidas, a despeito das considerações apresentadas nesse parágrafo.

A Tabela 26 apresenta os valores do $MAP(k)$ para $k = 10, 30, 50$, considerando na avaliação apenas os itens observados para o conjunto de dados DB1 e DB2.

Podemos observar, na Tabela 26, que o sistema que utiliza o MGDCP fornece melhores recomendações, comparado aos demais métodos competidores em ambos os conjuntos de dados, representado por maiores valores de $MAP(10)$, $MAP(30)$ e $MAP(50)$. Os métodos que utilizam técnicas de aprendizado de máquina, por sua vez, apresentam resultados ligeiramente piores aos do MGDCP melhorado. Ainda, o CFG possui os piores resultados dentre os métodos analisados. Esse resultado inferior pode ter sido influenciado pelo número reduzido de gêneros considerados. Ademais, esses resultados sugerem que ambos os métodos propostos demonstram

Tabela 26 – MAP(k) para os conjuntos de dados DB1 e DB2, considerando apenas os itens avaliados.

método	DB1			DB2		
	MAP(10)	MAP(30)	MAP(50)	MAP(10)	MAP(30)	MAP(50)
MGDCP	0.749	0.754	0.779	0.925	0.907	0.923
CFG ($k = 50$)	0.574	0.621	0.665	0.785	0.812	0.846
CFG ($k = 100$)	0.555	0.603	0.651	0.784	0.812	0.846
CFG ($k = 200$)	0.554	0.599	0.646	0.785	0.812	0.846
VAE	0.649	0.642	0.668	0.857	0.855	0.880
NeuMF	0.613	0.634	0.676	0.868	0.836	0.838
MGDCP melhorado	0.638	0.652	0.692	0.883	0.876	0.896

Nota: MGDCP melhorado obtido tomando o ponto de corte $p = 13$.

Fonte: Adaptada de [Marana et al. \(2024\)](#).

um desempenho superior na recomendação dos itens, quando são considerados apenas os itens já classificados na avaliação do MAP(k) (isto é, desconsiderando os dados ausentes).

A Tabela 27 apresenta os valores do MAP(k) para $k = 10, 30, 50$ para o conjunto de dados DB1 e DB2, considerando os dados ausentes como uma avaliação negativa do item.

Tabela 27 – MAP(k) para os conjuntos de dados DB1 e DB2, considerando os dados ausentes.

método	DB1			DB2		
	MAP(10)	MAP(30)	MAP(50)	MAP(10)	MAP(30)	MAP(50)
MGDCP	0.003	0.004	0.006	0.014	0.014	0.017
CFG ($k = 50$)	0.008	0.005	0.005	0.007	0.004	0.004
CFG ($k = 100$)	0.007	0.005	0.004	0.008	0.005	0.005
VAE	0.181	0.114	0.109	0.162	0.099	0.097
NEUMF	0.132	0.081	0.077	0.208	0.130	0.128
MGDCP melhorado	0.247	0.141	0.123	0.221	0.132	0.124

Nota: MGDCP melhorado obtido tomando o ponto de corte $p = 13$.

Fonte: Adaptada de [Marana et al. \(2024\)](#).

Com relação ao formato adotado para imputar os dados ausentes, a abordagem adotada envolve classificar as avaliações não observadas como indicadores do desinteresse ou da aversão do usuário a esses itens. Por esse motivo, parte da redução da precisão observada, da Tabela 26 para a Tabela 27, é explicada pela esparsidade do conjunto de dados, que aumenta a probabilidade dos itens recomendados para um indivíduo estarem na lista dos que não foram avaliados.

Como consequência desse tratamento, nota-se que todos os métodos avaliados apresentam uma queda considerável de desempenho. Nota-se, na Tabela 27, que o declínio mais expressivo foi do sistema de recomendação com o MGDCP, cujo desempenho passa a ser comparável com o pior método, o CFG. Em contrapartida, o MGDCP melhorado demonstra resultados mais favoráveis, devido à incorporação do critério de popularidade, o que ajuda na seleção dos itens a serem recomendados.

5.8 Discussão

A principal contribuição desse capítulo é a proposição de um novo sistema de recomendação para filmes que aplica um modelo de diagnóstico cognitivo para obter a preferência latente dos usuários e recomendar novos filmes, com base nessas preferências. Uma proposta adicional, denominada MGDCP melhorado, que inclui o fator popularidade para a recomendação dos itens, também foi considerada.

O sistema de recomendação utilizando o MGDCP demonstrou um desempenho superior ao método clássico CFG na previsão das notas dadas aos itens pelos indivíduos para os dois conjuntos de dados analisados.

Com relação à relevância dos itens recomendados, o método de recomendação MGDCP melhorado apresenta bons resultados para a medida $MAP(k)$ para ambas as formas de cálculo considerada (isto é, com e sem a inclusão dos dados faltantes), superando o método clássico CFG e os métodos e os que utilizam aprendizado de máquina com redes neurais, NeuMF e VAE. Por outro lado, o método baseado apenas no MGDCP apresenta bons resultados apenas quando os dados faltantes não são considerados. Desse modo, consideramos o método MGDCP melhorado mais apropriado, devido a consistência de seus resultados.

CONCLUSÃO E COMENTÁRIOS FINAIS

6.1 Comentários finais

No capítulo 2, foi apresentada uma versão Bayesiana para o modelo G-DINA. Para realizar a estimação dos parâmetros, foi proposta uma formulação de um modelo hierárquico e desenvolvido uma implementação desse modelo, utilizando o software *JAGS*. Uma das novidades dessa proposta foi a inclusão de dois novos conjuntos de restrições para os parâmetros dos itens, denominados restrição da menor probabilidade (RMeP) e restrição da maior probabilidade (RMaP). Esses conjuntos de restrições têm a vantagem de fazer menos suposições sobre os itens, além de serem mais simples de implementar em comparação com as restrições monotônicas (utilizada normalmente na literatura), sem prejuízo para a interpretabilidade dos resultados.

O estudo de simulação apresentado na seção 2.6 demonstrou que o método de estimação Bayesiano para o G-DINA recupera todos os parâmetros do modelo com acurácia similar ou superior ao método de estimação clássico, em todos os cenários considerados, independente do conjunto de restrições adotado.

Sobre a escolha do conjunto de restrição, dentre RMaP e RMeP, recomendamos que o profissional escolha aquele cujas hipóteses sejam mais adequadas à aplicação. Se, no contexto da aplicação, for razoável presumir que os indivíduos menos capazes têm menor probabilidade de responder corretamente a um item, então, a RMeP deve ser escolhida. Se, por outro lado, for razoável presumir que os indivíduos mais capazes têm maior probabilidade de responder corretamente a um item, então, a RMaP deve ser escolhida. Em muitos casos, ambas as condições são igualmente razoáveis e as duas podem ser escolhidas de forma intercambiável. Além disso, os resultados do estudo de simulação mostram que, nos casos em que ambos os conjuntos de restrições são satisfeitos, não há diferenças significativas na acurácia da estimação, quando comparados os resultados obtidos com a inclusão da RMaP ou da RMeP.

No capítulo 3 foi apresentado uma formulação Bayesiana para o MGDPCP, que inclui um

novo conjunto de restrições para os parâmetros dos itens, que fora denominado restrição flexível de maior e menor probabilidade (RFMaMeP).

O MGDCP é um Modelo de Diagnóstico Cognitivo geral para respostas politômicas. Apesar de um método de estimação clássico ter sido proposto por [Chen e de la Torre \(2018\)](#), ele foi desenvolvido para o software **Ox** e não foi disponibilizado em nenhum local acessível. Por esse motivo, desenvolvemos também um código de estimação clássico em R, que fora utilizado no estudo de simulação do capítulo 3 e na aplicação apresentada no capítulo 5.

Assim, como acontece no modelo G-DINA, a ausência de restrições nos parâmetros dos itens pode levar a resultados indesejáveis e prejudicar a interpretabilidade dos parâmetros do modelo. Contudo, os conjuntos de restrições dos parâmetros dos itens são mais complexos de serem implementados para MDC para respostas politômicas. Não há nenhuma solução conhecida para a implementação dos conjuntos de restrição RMaP, RMeP ou RM para o MGDCP ou qualquer outro MDC, para respostas politômicas, similar.

Tendo em vista essa dificuldade, foi proposto a inclusão de um conjunto de restrição flexível (que são formadas por um conjunto de condições desejáveis, porém não obrigatórias), em vez restrições rígidas (que são formadas por condições que devem ser, obrigatoriamente satisfeitas), como as utilizadas no modelo G-DINA, por exemplo. Esse conjunto de restrições flexíveis busca, por meio da escolha dos hiperparâmetros da distribuição a priori, produzir o mesmo efeito que seriam obtidos com a inclusão da restrição de maior e menor probabilidade na distribuição a posteriori dos parâmetros dos itens.

Novamente, a recuperação dos parâmetros do modelo pelo método de estimação proposto foi avaliada por meio de um estudo de simulação. Esse estudo demonstrou que o método de estimação Bayesiano do MGDCP apresenta acurácia similar ou superior que o método de estimação clássico, em todos os cenários considerados.

Um aspecto negativo da abordagem Bayesiana é o alto custo computacional, quando comparada à abordagem frequentista. Observamos que o custo computacional do método de estimação Bayesiano do modelo G-DINA e do MGDCP são consideravelmente maiores do que dos métodos clássicos, principalmente quando o teste considerado envolve um número maior de atributos e itens. Com respeito a esse quesito, vale ressaltar que, embora o método de estimação frequentista seja mais rápido, a estimação Bayesiana fornece uma aproximação da distribuição a posteriori de todos os parâmetros do modelo, permitindo conclusões mais precisas.

Considere, por exemplo, a metodologia de classificação da depressão, apresentada no capítulo 4. Enquanto o método de estimação clássico fornece apenas um diagnóstico fechado para cada indivíduo, a estimação Bayesiana é capaz de informar a probabilidade do indivíduo ser não-depressivo, sintomático apenas na dimensão cognitiva, apenas na dimensão somático-afetiva, ou em ambas as dimensões.

Além disso, com o rápido crescimento do poder computacional que vivenciamos, a

execução do método de estimação Bayesiano será consideravelmente mais rápida em poucos anos, fazendo que as vantagens da sua utilização superem as desvantagens. Desse modo, recomendamos a utilização dos métodos de estimação Bayesiano para os modelos G-DINA e MGDCP, em detrimento do método clássico, exceto para as aplicações cuja alta dimensionalidade do conjunto de dados e da matriz Q inviabilizam sua utilização.

No capítulo 4 foram analisados os dados de 1111 estudantes universitários ao Inventário de Depressão de Beck (BDI), aplicando a metodologia para a classificação da depressão proposta por [Silva et al. \(2018\)](#), porém com a utilização dos modelos G-DINA e MGDCP no lugar do modelo DINA.

Uma diferença importante entre os dois modelos é o tipo de resposta considerado na análise. O modelo G-DINA utiliza as respostas dicotomizadas, enquanto o MGDCP é capaz de lidar com as respostas politômicas originais. Ambos os modelos propostos apresentaram resultados interessantes e se mostraram adequados para essa aplicação, de acordo com o tipo de dado considerado na análise.

A escolha entre os dois modelos deve ser pautada, principalmente, nas características do conjunto de dados em análise. A utilização do modelo G-DINA é adequada para os casos em que apenas as respostas dicotomizadas estão disponíveis, ou quando o número de respondentes é muito grande e o tempo de processamento do modelo para resposta politômicas é impraticável. Nos demais casos, a utilização do MGDCP é aconselhável, por utilizar toda informação disponível para a classificação dos indivíduos.

Ainda, sobre a metodologia de classificação de depressão apresentada, é importante destacar que o enfoque deste trabalho foi no processo de modelagem e a análise estatística. Para a utilização dessa metodologia para o diagnóstico médico, faz-se necessário, conforme ressaltado por [Silva et al. \(2018\)](#), uma investigação aprofundada, envolvendo especialistas em psiquiatria, além da realização de estudos de validação dos resultados.

No capítulo 5 é apresentado um sistema de recomendação de filmes que utiliza o modelo MGDCP para fazer a predição das notas que os usuários dariam aos filmes, de acordo com as avaliações disponíveis.

Na análise dos dados da aplicação apresentada no capítulo mencionado, foi observado que há uma forte correlação entre a popularidade de um item e a probabilidade desse item estar dentre os mais bem avaliados pelo usuário. Contudo, no sistema proposto inicialmente, que faz suas recomendações puramente baseado nas predições de notas feitas pelo MGDCP, não há uma conexão entre os itens recomendados e a popularidade dos filmes. Por isso, foi proposto um aprimoramento desse sistema com a inclusão do fator popularidade na criação das listas de itens a serem recomendados. Esse novo sistema foi denominado de MGDCP melhorado.

Convém ressaltar que, para a aplicação apresentada no capítulo 5, foi utilizado método de estimação clássico para o MGDCP. Procedemos dessa forma porque o processo de estimação

Bayesiano do MGDCP via MCMC, conforme implementado no capítulo 3, envolve o armazenamento dos valores amostrados de todos os parâmetros do modelo de cada iteração do MCMC. Desse modo, devido à alta dimensionalidade da matriz Q e do conjunto de respostas avaliados, seria necessário uma quantidade massiva de memória para executar o método de estimação Bayesiano do MGDCP para o conjunto de dados considerados nessa aplicação.

Na análise comparativa realizada, o MGDCP melhorado apresentou resultados superiores à métodos modernos baseados em aprendizado de máquina com redes neurais, o *Neural Network with Matrix Factorization* (NeuMF) e *Variational Auto Encoder* (VAE), para a medida avaliada.

Uma limitação desse resultado é o fato dos métodos terem sido avaliados apenas com base na métrica $MAP(k)$. Essa avaliação pode ser ampliada em futuros estudos para incluir outras métricas e analisar outros conjuntos de dados.

Apesar disso, os resultados obtidos demonstram que o MGDCP melhorado é um método promissor a ser explorado como uma ferramenta poderosa para a aquisição de informações adicionais sobre as preferências latentes dos usuários de forma não invasiva, isto é, são baseadas em informações fornecidas de forma espontânea pelos usuários.

6.2 Produção acadêmica durante o Doutorado

Nesta seção listamos as produções acadêmicas decorrentes desta pesquisa de Doutorado.

Produções de artigos

- Artigo publicado:

MARANA, F. T.; FERNANDES, R. d.; GUZMÁN, J. L. B.; de Carvalho, A. C. P. d. F.; CÚRI, M. What rating they will probably give: A cognitive diagnosis approach for recommending items based on polytomous responses and latent attributes. **Expert Systems with Applications**, v. 245, p. 122981, 2024. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417423034838>>.

- Artigo submetido e em fase de revisão:

FERNANDES, R. d. S.; BAZÁN, J. L.; CÚRI, M. A bayesian approach for the G-DINA model. submetido ao *Brazilian Journal of Probability and Statistics* (em fase de revisão), 2024.

Seminários e apresentação em eventos

- Apresentação de trabalho na modalidade comunicação oral:

FERNANDES, R. S.; BAZÁN, J.L.; CÚRI, M. A Bayesian approach for the G-DINA model. In: ESCOLA DE MODELOS DE REGRESSÃO. 17. 2021. Rio de Janeiro. *Resumos das sessões de comunicação oral*. Rio de Janeiro. Escola Nacional de Ciências Estatísticas do IBGE. Rio de Janeiro. 2021. p. 16. Disponível em: <<https://eventos.ibge.gov.br/downloads/emr2021/arquivos/RESUMOS%20COMUNICACAO%20ORAL.pdf>>. Acesso em: 20 de fevereiro de 2024.

- Apresentação de seminário:

FERNANDES, R. S.; BAZÁN, J.L.; CÚRI, M. Uma abordagem Bayesiana para o modelo G-DINA. In: SEMINÁRIOS DO GRUPO DE MODELOS DE VARIÁVEIS LATENTES. 2022. São Carlos. Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo. 2022. Disponível em: <<https://www.icmc.usp.br/eventos/5832-uma-abordagem-bayesiana-para-o-modelo-g-dina>>. Acesso em: 20 de fevereiro de 2024.

- Apresentação de trabalho em evento internacional na modalidade *Contributed Abstracts*:

FERNANDES, R. S.; BAZÁN, J.L.; CÚRI, M. Analyzing different Constraints on item parameters in the Bayesian estimation of G-DINA model. In: INTERNATIONAL CONFERENCE ON DATA SCIENCE. 2023. Santiago. *Book of Abstracts...* Santiago. Universidad Diego Portales. 2023. p. 116. Disponível em: <https://icds2023.cl/wp-content/uploads/2023/11/BoA_ICDS2023.pdf>. Acesso em: 20 de fevereiro de 2024.

6.3 Desenvolvimentos futuros

Nessa seção, apresentamos algumas possibilidades de continuação de pesquisa.

Com relação aos métodos de estimação Bayesiano dos modelos G-DINA e MGDCP, foi utilizado, para ambos os modelos, o software JAGS, que utiliza, principalmente, o método MCMC denominado *Gibbs Sampling* (ou, amostrador de Gibbs, em português). Contudo, existem métodos MCMC mais eficientes, tais como os métodos *Hamiltonian Monte Carlo* (HMC) (SILVA *et al.*, 2018). Sendo assim, uma possibilidade de pesquisa é a implementação de métodos de estimação Bayesiano dos modelos G-DINA e MGDCP utilizando o software *Stan* (CARPENTER *et al.*, 2017; Stan Development Team, 2018). Ademais, as implementações mencionadas foram feitas em R (R Core Team, 2021); novas implementações, utilizando outras linguagens de programação, tais como Python (ROSSUM; DRAKE, 2009), facilitaria a utilização desses modelos por pesquisadores de outras áreas.

Outro ponto importante é relacionado ao custo computacional desses métodos. As implementações computacionais apresentadas nesta tese não são otimizadas para aplicação em testes de larga escala e possivelmente não são viáveis esse tipo de aplicação. Sendo assim, um

tópico de pesquisa interessante é o desenvolvimento de novos métodos de estimação Bayesiano, para os modelos G-DINA e MGDCP, que sejam eficientes em testes de larga escala.

Por fim, existem questões relacionadas à identificabilidade dos modelos gerais de diagnóstico cognitivo a serem exploradas. Podemos citar, por exemplo, o comportamento bimodal da distribuição dos parâmetros dos itens do modelo G-DINA, quando não há a imposição de restrições nesse modelo (ver seção 2.4, p. 40). Esse comportamento sugere um problema de identificabilidade do modelo G-DINA, que também se estende ao MGDCP. Embora já existam diversas pesquisas sobre a identificabilidade dos modelos de diagnóstico cognitivo (XU; ZHANG, 2016; XU *et al.*, 2017; CHEN *et al.*, 2019; FANG; LIU; YING, 2019; GU; XU, 2019; GU; XU, 2020; GU; XU, 2021), há muito a ser explorado sobre esse tema. O estudo das condições necessárias e suficientes, relacionadas aos parâmetros dos itens, para a identificabilidade dos modelos gerais de diagnóstico cognitivo, é um tópico a ser abordado em pesquisas futuras.

REFERÊNCIAS

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 24.

ARYADOUST, V. A cognitive diagnostic assessment study of the listening test of the singapore–cambridge general certificate of education o level: Application of DINA, DINO, G-DINA, HO-DINA, and RRUM. **International Journal of Listening**, Routledge, v. 35, n. 1, p. 29–52, 2021. Citado na página 27.

BOBADILLA, J.; HERNANDO, A.; ORTEGA, F.; BERNAL, J. A framework for collaborative filtering recommender systems. **Expert Systems with Applications**, v. 38, n. 12, p. 14609–14623, 2011. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417411008049>>. Citado na página 89.

BROOKS, S. P.; GELMAN, A. General methods for monitoring convergence of iterative simulations. **Journal of computational and graphical statistics**, Taylor & Francis, v. 7, n. 4, p. 434–455, 1998. Citado na página 49.

CARPENTER, B.; GELMAN, A.; HOFFMAN, M. D.; LEE, D.; GOODRICH, B.; BETANCOURT, M.; BRUBAKER, M. A.; GUO, J.; LI, P.; RIDDELL, A. Stan: A probabilistic programming language. **Journal of statistical software**, NIH Public Access, v. 76, 2017. Citado na página 105.

CHEN, H.; CHEN, J. Exploring reading comprehension skill relationships through the G-DINA model. **Educational Psychology**, Taylor & Francis, v. 36, n. 6, p. 1049–1064, 2016. Citado nas páginas 24 e 27.

_____. Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. **Language Assessment Quarterly**, Taylor & Francis, v. 13, n. 3, p. 218–230, 2016. Citado na página 27.

CHEN, J.; DE LA TORRE, J. Introducing the general polytomous diagnosis modeling framework. **Frontiers in Psychology**, Frontiers Media SA, v. 9, p. 1474, 2018. ISSN 1664-1078. Disponível em: <<https://www.frontiersin.org/article/10.3389/fpsyg.2018.01474>>. Citado nas páginas 25, 55, 57, 58 e 102.

CHEN, Y.; CULPEPPER, S. A.; CHEN, Y.; DOUGLAS, J. Bayesian estimation of the DINA Q matrix. **Psychometrika**, Springer, v. 83, n. 1, p. 89–108, 2018. Citado nas páginas 24 e 29.

CHEN, Y.-H.; SENK, S. L.; THOMPSON, D. R.; VOOGT, K. Examining psychometric properties and level classification of the van hiele geometry test using ctt and cdm frameworks. **Journal of Educational Measurement**, Wiley Online Library, v. 56, n. 4, p. 733–756, 2019. Citado nas páginas 24 e 106.

CHIU, C.-Y. Statistical refinement of the Q-matrix in cognitive diagnosis. **Applied Psychological Measurement**, Sage Publications Sage CA: Los Angeles, CA, v. 37, n. 8, p. 598–618, 2013. Citado na página 24.

- CULPEPPER, S. A. Estimating the cognitive diagnosis Q matrix with expert knowledge: Application to the fraction-subtraction dataset. **Psychometrika**, Springer, v. 84, n. 2, p. 333–357, 2019. Citado na página 44.
- DE LA TORRE, J. 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 na página 24.
- _____. DINA model and parameter estimation: A didactic. **Journal of educational and behavioral statistics**, Sage Publications Sage CA: Los Angeles, CA, v. 34, n. 1, p. 115–130, 2009. Citado nas páginas 24, 29 e 34.
- _____. The partial-credit DINA model. In: **International Meeting of the Psychometric Society**. [S.l.: s.n.], 2010. Citado na página 55.
- _____. The generalized DINA model framework. **Psychometrika**, v. 76, n. 2, p. 179–199, April 2011. Citado nas páginas 23, 25, 27, 30, 33, 35, 36, 39, 40 e 55.
- DE LA TORRE, J.; ARK, L. A. van der; ROSSI, G. Analysis of clinical data from a cognitive diagnosis modeling framework. **Measurement and Evaluation in Counseling and Development**, Taylor & Francis, v. 51, n. 4, p. 281–296, 2018. Citado na página 24.
- DE LA TORRE, J.; CHIU, C.-Y. A general method of empirical Q-matrix validation. **Psychometrika**, Springer, v. 81, n. 2, p. 253–273, 2016. Citado na página 24.
- DELAFONTAINE, J.; CHEN, C.; PARK, J. Y.; NOORTGATE, W. Van den. Using country-specific q-matrices for cognitive diagnostic assessments with international large-scale data. **Large-scale Assessments in Education**, Springer, v. 10, n. 1, p. 19, 2022. Citado na página 27.
- DIBELLO, L. V.; ROUSSOS, L. A.; STOUT, W. 31a review of cognitively diagnostic assessment and a summary of psychometric models. **Handbook of statistics**, Elsevier, v. 26, p. 979–1030, 2006. Citado na página 23.
- DU, H.; KE, Z.; JIANG, G.; HUANG, S. The performances of gelman-rubin and geweke's convergence diagnostics of monte carlo markov chains in bayesian analysis. **Journal of Behavioral Data Science**, v. 2, n. 2, p. 47–72, 2022. Citado na página 67.
- EVERETT, B. **An introduction to latent variable models**. [S.l.]: Chapman and Hall, 1984. Citado na página 23.
- FANG, G.; LIU, J.; YING, Z. On the identifiability of diagnostic classification models. **Psychometrika**, Springer, v. 84, p. 19–40, 2019. Citado na página 106.
- FERNANDES, R. d. S.; BAZÁN, J. L.; CÚRI, M. A bayesian approach for the G-DINA model. submetido ao Brazilian Journal of Probability and Statistics (em fase de revisão), 2024. Citado na página 104.
- FRAGOSO, T. M.; CÚRI, M. Improving psychometric assessment of the beck depression inventory using multidimensional item response theory. **Biometrical Journal**, Wiley Online Library, v. 55, n. 4, p. 527–540, 2013. Citado na página 72.
- GELMAN, A.; RUBIN, D. B. Inference from Iterative Simulation Using Multiple Sequences. **Statistical Science**, Institute of Mathematical Statistics, v. 7, n. 4, p. 457 – 472, 1992. Disponível em: <<https://doi.org/10.1214/ss/1177011136>>. Citado nas páginas 49 e 67.

GEORGE, A. C.; ROBITZSCH, A. Cognitive diagnosis models in r: A didactic. **The Quantitative Methods for Psychology**, v. 11, n. 3, p. 189–205, 2015. Citado na página 23.

GU, Y.; XU, G. The sufficient and necessary condition for the identifiability and estimability of the dina model. **Psychometrika**, Springer, v. 84, p. 468–483, 2019. Citado na página 106.

_____. Partial identifiability of restricted latent class models. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 48, n. 4, p. 2082 – 2107, 2020. Disponível em: <<https://doi.org/10.1214/19-AOS1878>>. Citado na página 106.

_____. Sufficient and necessary conditions for the identifiability of the q-matrix. **Statistica Sinica**, JSTOR, v. 31, n. 1, p. 449–472, 2021. Citado na página 106.

HAERTEL, E. H. Using restricted latent class models to map the skill structure of achievement items. **Journal of Educational Measurement**, Wiley Online Library, v. 26, n. 4, p. 301–321, 1989. Citado nas páginas 23, 29 e 55.

HARTZ, S. A bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality (doctoral dissertation). **University of Illinois at Urbana-Champaign**, 2002. Citado na página 23.

HE, X.; LIAO, L.; ZHANG, H.; NIE, L.; HU, X.; CHUA, T.-S. Neural collaborative filtering. In: **Proceedings of the 26th International Conference on World Wide Web**. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017. (WWW '17), p. 173–182. ISBN 9781450349130. Disponível em: <<https://doi.org/10.1145/3038912.3052569>>. Citado na página 89.

HENSON, R. A.; TEMPLIN, J. L.; WILLSE, J. T. Defining a family of cognitive diagnosis models using log-linear models with latent variables. **Psychometrika**, Springer, v. 74, p. 191–210, 2009. Citado nas páginas 25, 44 e 55.

HONG, C.-Y.; CHANG, Y.-W.; TSAI, R.-C. Estimation of generalized DINA model with order restrictions. **Journal of Classification**, Springer, v. 33, n. 3, p. 460–484, 2016. Citado na página 51.

HWANG, J.; CHOI, K. M.; HAND, B. Examining domain-general use of reasoning across science and mathematics through performance on standardized assessments. **Canadian Journal of Science, Mathematics and Technology Education**, Springer, v. 20, n. 3, p. 521–537, 2020. Citado na página 27.

JUNIOR, F. d. J. M. Aplicações da teoria da resposta ao item (tri) no brasil. **Revista Brasileira de Biometria**, v. 28, n. 4, p. 137–70, 2010. Citado na página 24.

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 29.

KOEDSRI, A.; NGUDGRATOKE, S. Diagnostic assessment of scientific literacy of lower secondary school students using G-DINA model. **Yearbook of Teacher Education**, p. 268, 2018. Citado na página 27.

KREITCHMANN, R. S.; TORRE, J. de la; SORREL, M. A.; NÁJERA, P.; ABAD, F. J. Improving reliability estimation in cognitive diagnosis modeling. **Behavior Research Methods**, Springer, v. 55, n. 7, p. 3446–3460, 2023. Citado na página 27.

- LEI, P.-W.; LI, H. Performance of fit indices in choosing correct cognitive diagnostic models and q-matrices. **Applied Psychological Measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 40, n. 6, p. 405–417, 2016. Citado na página 75.
- LI, H.; HUNTER, C. V.; LEI, P.-W. The selection of cognitive diagnostic models for a reading comprehension test. **Language Testing**, Sage Publications Sage UK: London, England, v. 33, n. 3, p. 391–409, 2016. Citado nas páginas 24 e 27.
- LI, L.; AN, Y.; REN, J.; WEI, X. Research on the cognitive diagnosis of chinese listening comprehension ability based on the g-dina model. **Frontiers in Psychology**, v. 12, 2021. Disponível em: <<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.714568>>. Citado na página 27.
- LIANG, M. Z.; CHEN, P.; KNOBF, M.; MOLASSIOTIS, A.; TANG, Y.; HU, G. Y.; SUN, Z.; YU, Y. L.; YE, Z. J. Measuring resilience by cognitive diagnosis models and its prediction of 6-month quality of life in be resilient to breast cancer (brbc). **Frontiers in Psychiatry**, Frontiers, v. 14, p. 1102258, 2023. Citado na página 27.
- LIU, J.; XIAO, Y.; ZHU, K.; ZHENG, W.; HSU, C.-H. Hybrid variational autoencoder for collaborative filtering. In: IEEE. **2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)**. [S.l.], 2022. p. 251–256. Citado na página 89.
- MA, W.; DE LA TORRE, J. **GDINA: The generalized DINA model framework, R package version 0.13.0**. 2016. Citado nas páginas 45, 46, 47, 49, 50, 63, 64 e 65.
- _____. A sequential cognitive diagnosis model for polytomous responses. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 69, n. 3, p. 253–275, 2016. Citado na página 55.
- _____. GDINA: An R package for cognitive diagnosis modeling. **Journal of Statistical Software**, v. 93, n. 14, p. 1–26, 2020. Citado nas páginas 36, 49 e 51.
- MA, W.; JIANG, Z. Estimating cognitive diagnosis models in small samples: Bayes modal estimation and monotonic constraints. **Applied Psychological Measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 45, n. 2, p. 95–111, 2021. Disponível em: <<https://doi.org/10.1177/0146621620977681>>. Citado na página 36.
- MARANA, F. T.; FERNANDES, R. d.; GUZMÁN, J. L. B.; de Carvalho, A. C. P. d. F.; CÚRI, M. What rating they will probably give: A cognitive diagnosis approach for recommending items based on polytomous responses and latent attributes. **Expert Systems with Applications**, v. 245, p. 122981, 2024. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417423034838>>. Citado nas páginas 87, 88, 89, 93, 94, 95, 96, 97, 98 e 104.
- MARIS, E. Estimating multiple classification latent class models. **Psychometrika**, Springer, v. 64, p. 187–212, 1999. Citado na página 23.
- MESEGUER, P.; ROSSI, F.; SCHIEX, T. Chapter 9 - soft constraints. In: ROSSI, F.; van Beek, P.; WALSH, T. (Ed.). **Handbook of Constraint Programming**. Elsevier, 2006, (Foundations of Artificial Intelligence, v. 2). p. 281–328. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1574652606800131>>. Citado na página 60.

- MIAO, G.; MILLER, M. D.; REN, L. 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 24.
- MICROSOFT. **Recommender**. [S.l.]: GitHub, 2023. <<https://github.com/microsoft/recommenders>>. Citado na página 89.
- NÁJERA, P.; SORREL, M. A.; DE LA TORRE, J.; ABAD, F. J. Improving robustness in Q-matrix validation using an iterative and dynamic procedure. **Applied Psychological Measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 44, n. 6, p. 431–446, 2020. Citado na página 24.
- OLIVEIRA, E. S. de; WANG, X.; BAZÁN, J. L. A classification model for continuous responses: Identifying risk perception groups on health-related activities. **Biometrical Journal**, Wiley Online Library, v. 65, n. 4, p. 2100222, 2023. Citado na página 72.
- PLUMMER, M. *et al.* rjags: Bayesian graphical models using mcmc. **R package version**, v. 4, n. 6, 2016. Citado nas páginas 44, 49, 62 e 66.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 57 e 105.
- RAVAND, H. Application of a cognitive diagnostic model to a high-stakes reading comprehension test. **Journal of Psychoeducational Assessment**, Sage Publications Sage CA: Los Angeles, CA, v. 34, n. 8, p. 782–799, 2016. Citado na página 27.
- RAVAND, H.; ROBITZSCH, A. Cognitive diagnostic model of best choice: a study of reading comprehension. **Educational Psychology**, Taylor & Francis, v. 38, n. 10, p. 1255–1277, 2018. Citado na página 27.
- ROBITZSCH, A.; KIEFER, T.; GEORGE, A. C.; ÜNLÜ, A. **CDM: Cognitive Diagnosis Modeling**. [S.l.], 2020. R package version 7.5-15. Disponível em: <<https://CRAN.R-project.org/package=CDM>>. Citado nas páginas 36 e 49.
- ROSSUM, G. V.; DRAKE, F. L. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009. ISBN 1441412697. Citado na página 105.
- ROY, D.; DUTTA, M. A systematic review and research perspective on recommender systems. **Journal of Big Data**, Springer, v. 9, n. 1, p. 59, 2022. Citado na página 87.
- RUPP, A. A.; TEMPLIN, J. The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the dina model. **Educational and Psychological Measurement**, Sage Publications Sage CA: Los Angeles, CA, v. 68, n. 1, p. 78–96, 2008. Citado na página 24.
- RUPP, A. A.; TEMPLIN, J.; HENSON, R. A. **Diagnostic measurement: Theory, methods, and applications**. [S.l.]: Guilford press, 2010. Citado na página 23.
- SILVA, M. A. d.; OLIVEIRA, E. S. B. d.; DAVIER, A. A. v.; GUZMÁN, J. L. B. Estimating the DINA model parameters using the no-u-turn sampler. **Biometrical Journal**, WILEY-VCH Verlag GmbH and Co. KGaA, 2018. Citado nas páginas 24, 26, 72, 73, 77, 78, 85, 103 e 105.

SONG, L.-h.; WANG, W.-y.; DAI, H.-q.; DING, S.-l. Using reversible jump mcmc for cognitive diagnostic model selection. In: IEEE. **2012 9th International Conference on Fuzzy Systems and Knowledge Discovery**. [S.l.], 2012. p. 1572–1576. Citado na página 50.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. van der. Bayesian measures of model complexity and fit. **Journal of the royal statistical society: Series b (statistical methodology)**, v. 64, n. 4, p. 583–639, 2002. Citado na página 75.

Stan Development Team. **RStan: the R interface to Stan**. 2018. R package version 2.17.3. Disponível em: <<http://mc-stan.org/17>>. Citado na página 105.

TATSUOKA, K. K. Rule space: An approach for dealing with misconceptions based on item response theory. **Journal of educational measurement**, JSTOR, p. 345–354, 1983. Citado na página 24.

TEMPLIN, J. L.; HENSON, R. A. Measurement of psychological disorders using cognitive diagnosis models. **Psychological methods**, American Psychological Association, v. 11, n. 3, p. 287, 2006. Citado nas páginas 23, 24, 30 e 55.

TONEKABONI, F. R.; RAVAND, H.; REZVANI, R. The construction and validation of a Q-matrix for a high-stakes reading comprehension test: A G-DINA study. **International Journal of Language Testing**, Tabaran Institute of Higher Education, v. 11, n. 1, p. 58–87, March 2021. Citado na página 27.

VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. **Statistics and computing**, Springer, v. 27, p. 1413–1432, 2017. Citado na página 75.

VEHTARI, A.; GELMAN, A.; SIMPSON, D.; CARPENTER, B.; BÜRKNER, P.-C. Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC (with Discussion). **Bayesian Analysis**, International Society for Bayesian Analysis, v. 16, n. 2, p. 667 – 718, 2021. Disponível em: <<https://doi.org/10.1214/20-BA1221>>. Citado nas páginas 49 e 67.

VON DAVIER, M. A general diagnostic model applied to language testing data. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 61, n. 2, p. 287–307, 2008. Citado na página 55.

_____. The general diagnostic model. In: _____. **Handbook of Diagnostic Classification Models: Models and Model Extensions, Applications, Software Packages**. Cham: Springer International Publishing, 2019. p. 133–153. ISBN 978-3-030-05584-4. Disponível em: <https://doi.org/10.1007/978-3-030-05584-4_6>. Citado na página 25.

WANG, D.; CAI, Y.; TU, D. Q-matrix estimation methods for cognitive diagnosis models: Based on partial known q-matrix. **Multivariate Behavioral Research**, Taylor & Francis, p. 1–13, 2020. Citado na página 24.

WANG, W.; SONG, L.; DING, S.; MENG, Y.; CAO, C.; JIE, Y. An em-based method for Q-matrix validation. **Applied psychological measurement**, SAGE Publications Sage CA: Los Angeles, CA, v. 42, n. 6, p. 446–459, 2018. Citado na página 24.

WU, X.; ZHANG, Y.; WU, R.; CHANG, H.-H. A comparative study on cognitive diagnostic assessment of mathematical key competencies and learning trajectories: —pisa data analysis based on 19, 454 students from 8 countries. **Current Psychology**, Springer, p. 1–13, 2022. Citado na página [27](#).

XU, G. *et al.* Identifiability of restricted latent class models with binary responses. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 45, n. 2, p. 675–707, 2017. Citado na página [106](#).

XU, G.; ZHANG, S. Identifiability of diagnostic classification models. **Psychometrika**, Springer, v. 81, n. 3, p. 625–649, 2016. Citado na página [106](#).

YAMAGUCHI, K.; OKADA, K. Comparison among cognitive diagnostic models for the timss 2007 fourth grade mathematics assessment. **PloS one**, Public Library of Science San Francisco, CA USA, v. 13, n. 2, p. e0188691, 2018. Citado na página [24](#).

ZHAN, P.; JIAO, H.; MAN, K.; WANG, L. Using jags for bayesian cognitive diagnosis modeling: A tutorial. **Journal of Educational and Behavioral Statistics**, SAGE Publications Sage CA: Los Angeles, CA, v. 44, n. 4, p. 473–503, 2019. Citado nas páginas [28](#), [44](#), [49](#), [67](#) e [73](#).

