

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Programa de Pós-Graduação em Estatística

Samuel de Souza Barbosa

**Modelo Bayesiano de Teoria da Resposta ao Item com estrutura
hierárquica e multinível para os traços latentes aplicado aos
dados do SAEB**

Belo Horizonte

2023

Samuel de Souza Barbosa

Modelo Bayesiano de Teoria da Resposta ao Item com estrutura hierárquica e multinível para os traços latentes aplicado aos dados do SAEB

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Doutor Flávio Bamberra Gonçalves
Coorientador: Profa. Doutora Juliane Venturelli Silva Lima

Belo Horizonte

2023

Barbosa, Samuel de Souza.

B238m Modelo Bayesiano de Teoria da Resposta ao Item com
estrutura hierárquica e multinível para os traços latentes
aplicado aos dados do SAEB / Samuel de Souza Barbosa.–
2023.
83 f. il.

Orientador: Flávio Bambirra Gonçalves
Coorientadora: Juliane Venturelli Silva Lima
Dissertação (mestrado) - Universidade Federal de Minas
Gerais, Instituto de Ciências Exatas, Departamento de Estatística
Referências: f.70-71.

1. Estatística – Teses. 2. Teoria da resposta do item – Teses.
3. Avaliação Educacional – Teses. I. Gonçalves, Flávio Bambirra
II. Lima, Juliane Venturelli Silva. III. Universidade Federal de
Minas Gerais, Instituto de Ciências Exatas, Departamento de
Estatística. IIII. Título.

CDU 519.2(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

FOLHA DE APROVAÇÃO

Modelo Bayesiano de Teoria da Resposta ao Item com estrutura hierárquica e multinível para os traços latentes aplicado aos dados do SAEB.

SAMUEL DE SOUZA BARBOSA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Mestre em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 16 de novembro de 2023, pela banca constituída pelos membros:

Prof. Flávio Bambirra Gonçalves - Orientador
DEST/UFMG

Prof. Juliane Venturéli Silva Lima
FAB - Força Aérea Brasileira

Prof. Vinícius Diniz Mayrink
DEST/UFMG



Documento assinado digitalmente

Dalton Francisco de Andrade

Data: 22/11/2023 15:11:45-0300

CPF: ***.559.248-**

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Dalton Francisco de Andrade
UFSC

Belo Horizonte, 16 de novembro de 2023.



ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO DO ALUNO SAMUEL DE SOUZA BARBOSA, MATRICULADO, SOB O Nº 2021656394, NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 16 DE NOVEMBRO DE 2023.

Aos 16 de novembro de 2023, às 15h, em reunião pública presencial de número 273 (conforme orientações para a atividade de defesa de dissertação durante a vigência da Portaria PRPG nº 1819), na sala 2030, do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora, homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de dissertação do aluno SAMUEL DE SOUZA BARBOSA, nº matrícula 2021656394, intitulada: "*Modelo Bayesiano de Teoria da Resposta ao Item com estrutura hierárquica e multinível para os traços latentes aplicado aos dados do SAEB*", requisito final para obtenção do Grau de mestre em Estatística. Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Flávio Bambirra Gonçalves, passou a palavra ao aluno para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do aluno. Após a defesa, os membros da banca examinadora reuniram-se reservadamente, sem a presença do aluno e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

- (X) Aprovado.
- () Reprovado com resubmissão do texto em ____ dias.
- () Reprovado com resubmissão do texto e nova defesa em ____ dias.
- () Reprovado.

Flávio Bambirra Gonçalves (DEST/UFMG)
Orientador

Juliane Venturolli Silva Lima (FAB)
Coorientadora

Vinícius Diniz Mayrink (DEST/UFMG)

Dalton Francisco de Andrade (INE-CTC/UFSC)

O resultado final foi comunicado publicamente ao aluno pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 16 de novembro de 2023.

Resumo

Este trabalho propõe um modelo hierárquico multinível para inferência de habilidades e sub-habilidades a partir da Teoria da Resposta ao Item. O modelo permite a inclusão de covariáveis no nível da escola e o agrupamento de alunos, fatores importantes no estudo das avaliações educacionais. O modelo é aplicado aos dados do Sistema de Avaliação da Educação Básica (SAEB) 2019 e os resultados analisados frente ao modelo logístico unidimensional de três parâmetros utilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Palavras-chave: Estatística; Teoria da Resposta ao Item; Avaliação Educacional; SAEB

Abstract

This work proposes a hierarchical multilevel model for inferring proficiency in a discipline and its areas based on Item Response Theory. The model allows for the inclusion of school-level covariates and the clustering of students, which are important factors in the study of educational assessments. The model is applied to the data from the Basic Education Assessment System (Saeb) 2019, and the results are analyzed in comparison to the unidimensional three-parameter logistic model used by the National Institute for Educational Studies and Research Anísio Teixeira (Inep).

Keywords: Statistics; Item Response Theory; Educational Assessments; SAEB

Lista de ilustrações

Figura 1 – Número de itens distintos por tema	17
Figura 2 – Composição dos cadernos de teste - 5º ano do Ensino Fundamental	18
Figura 3 – Curva Característica do Item - Parâmetro de Dificuldade	20
Figura 4 – Curva Característica do Item - Parâmetro de Discriminação	21
Figura 5 – Ilustração da hierarquia entre traços latentes	26
Figura 6 – Distribuição do número de alunos avaliados por escola no SAEB - 3ª série do Ensino Médio	35
Figura 7 – Cenário 1 - Nulo	41
Figura 8 – Cenário 2 - Hierárquico	43
Figura 9 – Cenário 3 - Multinível	45
Figura 10 – Cenário 4 - Completo	47
Figura 11 – Variação 1 - Modelo Nulo	51
Figura 12 – Variação 2 - Modelo Hierárquico	52
Figura 13 – Variação 3 - Modelo Multinível	54
Figura 14 – Variação 4 - Modelo Completo	55
Figura 15 – Hierarquia proposta entre traços latentes - Matemática	57
Figura 16 – Hierarquia proposta entre traços latentes - Língua Portuguesa	59
Figura 17 – Resultados do modelo com dados do SAEB - Matemática	64
Figura 18 – Decomposição dos efeitos da escola - Prova de Matemática do SAEB	65
Figura 19 – Resultados do modelo com dados do SAEB - Língua Portuguesa	68
Figura 20 – Cenário 1 - Nulo	74
Figura 21 – Cenário 2 - Hierárquico	75
Figura 22 – Cenário 3 - Multinível	76
Figura 23 – Cenário 4 - Completo	77
Figura 24 – Variação 1 - Modelo Nulo	78
Figura 25 – Variação 2 - Modelo Hierárquico	79
Figura 26 – Variação 3 - Modelo Multinível	80
Figura 27 – Variação 4 - Modelo Completo	81

Figura 28 – Traceplot do modelo com dados do SAEB - Matemática	82
Figura 29 – Traceplot do modelo com dados do SAEB - Língua Portuguesa	83

Lista de tabelas

Tabela 1 – Detalhamento da População Alvo do SAEB 2019	15
Tabela 2 – Quantitativo de temas, habilidades e itens nas matrizes de referência do SAEB 2019	16
Tabela 3 – Exemplo de uma matriz Q	23
Tabela 4 – Fatores do processo gerador de dados	34
Tabela 5 – Cenários para análise de variáveis do modelo	38
Tabela 6 – Média do Erro Quadrático Médio, Viés e Variância da estimação no modelo proposto em cada conjunto simulado de dados	40
Tabela 7 – Resultados do Cenário 1 - Nulo	42
Tabela 8 – Resultados do Cenário 2 - Hierárquico	44
Tabela 9 – Resultados do Cenário 3 - Multinível	46
Tabela 10 – Resultados do Cenário 4 - Completo	48
Tabela 11 – Variações do modelo para simulação	49
Tabela 12 – Média do Erro Quadrático Médio, Viés e Variância da estimação em cada modelo	50
Tabela 13 – Resultados da Variação 1 - Modelo Nulo	51
Tabela 14 – Resultados da Variação 2 - Modelo Hierárquico	53
Tabela 15 – Resultados da Variação 3 - Modelo Multinível	54
Tabela 16 – Resultados da Variação 4 - Modelo Completo	56
Tabela 17 – Traços latentes, temas e número de itens distintos por tema na prova de Matemática da 3 ^a série do ensino médio - SAEB 2019	57
Tabela 18 – Quantitativo de alunos por número de itens respondidos em cada tema de Matemática	58
Tabela 19 – Traços latentes, temas e número de itens distintos por tema na prova de Língua Portuguesa da 3 ^a série do ensino médio - SAEB 2019	59
Tabela 20 – Quantitativo de alunos por número de itens respondidos em cada tema de Língua Portuguesa	60
Tabela 21 – Inferência com dados do SAEB - Matemática	62

Tabela 22 – Inferência com dados do SAEB - Língua Portuguesa	66
Tabela 23 – Recorte de temas e habilidades da Matriz de Referência do SAEB 2019:	
Matemática 3ª série do ensino médio	84

Sumário

1	Introdução	12
2	Sistema de Avaliação da Educação Básica (SAEB)	14
2.1	Matrizes de Referência	16
2.2	Composição das provas	18
3	Modelos da Teoria da Resposta ao Item	19
3.1	Modelo logístico unidimensional de três parâmetros	19
3.1.1	Parâmetros dos itens	19
3.1.2	Identificação de modelos da TRI	21
3.2	Modelos probito unidimensional, multidimensional e multiunidimensional de três parâmetros	22
3.3	Modelo unidimensional multinível de três parâmetros com agrupamento de alunos em escolas	24
3.4	Modelo multidimensional de três parâmetros com hierarquia nos traços latentes	26
4	Modelo multidimensional com hierarquia nos traços latentes e agrupamento de indivíduos	28
4.1	Inclusão de covariáveis e agrupamento de alunos em escolas	28
4.2	Multiunidimensionalidade	29
4.3	Elicitação Bayesiana do modelo proposto	29
5	Estimação do modelo proposto e variações	31
5.1	Hamiltonian Monte Carlo	31
5.2	No U-Turn Sampler	32
5.3	Estatísticas de diagnóstico	32
6	Simulações	34
6.1	Desempenho do modelo proposto em diferentes conjuntos de dados	34
6.1.1	Resultados do modelo proposto em cenários simulados de dados	39
6.2	Variações do modelo	49

6.2.1	Resultados das variações do modelo	49
7	Aplicação Empírica	57
7.1	Resultados do modelo proposto aplicado aos dados do SAEB	61
7.1.1	Resultados do modelo para Matemática	61
7.1.2	Resultados do modelo para Língua Portuguesa	66
8	Considerações Finais	69
	Referências	71
	Apêndices	73

1 Introdução

A Teoria da Resposta ao Item (TRI) é um conjunto de modelos matemáticos que procura representar a probabilidade de um indivíduo dar uma certa resposta a um item como função de parâmetros do item e da habilidade (ou habilidades) dos respondentes (Andrade; Tavares; Valle, 2000). Nos modelos da TRI, as habilidades são usualmente consideradas traços latentes, isto é, constructos hipotéticos, não observáveis, que determinam as probabilidades de resposta aos itens do teste (Reckase, 2009).

Os modelos propostos na literatura variam, dentre outros aspectos, pela natureza do item — dicotômico ou não dicotômico — e pela quantidade de habilidades que estão sendo medidas — apenas uma ou mais de uma. Os itens dicotômicos (ou dicotomizados) são aqueles cuja resposta é corrigida como certa ou errada. Já a resposta a itens não dicotômicos ou não dicotomizados podem assumir valores arbitrários.

Nesta dissertação serão considerados somente os modelos para itens dicotômicos ou dicotomizados, uma vez que na maior parte das avaliações padronizadas, tais como as provas de múltipla escolha do Sistema de Avaliação da Educação Básica (SAEB) e do Exame Nacional do Ensino Médio (Enem), são aplicados itens dicotomizáveis.

Em relação à quantidade de habilidades, os modelos unidimensionais assumem que existe uma proficiência determina as probabilidades de resposta a todos os itens do teste. Se denotarmos por Y_{ij} a resposta do aluno j ao item i , em um teste composto por I itens aplicado a J estudantes, no modelo unidimensional temos que $P(\mathbf{Y}_j) = f(\theta_j)$, em que $\theta_j \in \mathbb{R}$ é a habilidade do aluno j (ou a proficiência que resume suas habilidades), sendo \mathbf{Y}_j o vetor de respostas do estudante aos itens $i = 1, \dots, I$ e f uma função de probabilidade.

Os modelos multidimensionais, por sua vez, assumem que múltiplos fatores determinam as probabilidades de resposta do conjunto de itens, isto é, $P(\mathbf{Y}_j) = f(\boldsymbol{\theta}_j)$, em que $\boldsymbol{\theta}_j \in \mathbb{R}^D$ é o conjunto de habilidades do aluno j (ou um vetor multidimensional que resume suas habilidades), sendo D um inteiro positivo maior do que um.

Além disso, nos modelos conhecidos como multiunidimensionais, múltiplas dimensões determinam as respostas à totalidade de itens testados, porém cada item está associado

somente a uma dimensão, isto é, temos ainda que $P(\mathbf{Y}_j) = f(\boldsymbol{\theta}_j)$ com $\boldsymbol{\theta}_j \in \mathbb{R}^D$, porém $P(Y_{ij}) = f(\theta_{jd})$, $d \in \{1, \dots, D\}$, com $\theta_{jd} \in \mathbb{R}$ para todo j e d .

Um dos modelos mais comuns na Teoria da Resposta ao Item, o modelo logístico de três parâmetros, foi proposto por [Birnbaum \(1968\)](#). Em sua forma multidimensional, múltiplos traços latentes determinam as probabilidades de resposta correta a itens dicotômicos.

Nas avaliações educacionais do ENEM e do SAEB, os testes são construídos a partir das chamadas matrizes de referência, que definem múltiplas habilidades avaliadas, agrupadas em temas ou áreas do conhecimento. Dessa forma, a construção desses testes induzem uma modelagem multidimensional, em que cada dimensão resume as habilidades de um tema, e hierárquica, pois a cada tema ou área estão associadas as habilidades que o compõem. Como, nessas avaliações, cada item do teste é destinado a medir uma única habilidade, a estrutura das matrizes de referência pode ser incorporada aos modelos, por exemplo, através da matriz Q , uma matriz binária que indica a relação entre itens e habilidades (ou traços latentes). Cada entrada na matriz Q é igual a um se o item se destina a medir a habilidade em questão, e igual a zero caso contrário ([Da Silva et al., 2019](#)).

Diversas extensões do modelo multidimensional foram propostas na literatura para incorporar hierarquia entre traços latentes. Neste trabalho, utilizaremos como referência o modelo proposto por [Venturelli S. L., Gonçalves e Andrade \(2020\)](#). Outra referência importante será o modelo de [Fox \(2010\)](#), que incorpora a hierarquia de alunos em escolas e permite a inclusão de covariáveis explicativas.

O objetivo desta dissertação é propor um modelo multiunidimensional que incorpore tanto a hierarquia entre traços latentes quanto a hierarquia entre alunos e escolas, possibilitando, ainda, a inclusão de covariáveis para alunos ou escolas.

Na próxima Seção (2), descrevemos o SAEB, as matrizes de referência e a composição das provas. Na Seção 3, faremos a revisão dos modelos de TRI mencionados. Em seguida, na Seção 4, apresentaremos o modelo proposto. Descreveremos brevemente como serão estimados os resultados do modelo proposto e suas variações na Seção 5. A Seção 6 traz o resultado do uso do modelo em algumas simulações e a Seção 7 a aplicação empírica aos dados do SAEB.

2 Sistema de Avaliação da Educação Básica (SAEB)

O SAEB é um conjunto de avaliações externas em larga escala que permite que as escolas e as redes municipais e estaduais de ensino avaliem a qualidade da educação oferecida aos estudantes.

Os testes do SAEB são elaborados a partir de matrizes de referência. As matrizes são compostas de temas, e em cada tema, os conteúdos associados a competências e habilidades desejáveis para cada série e para cada disciplina são detalhados, cada uma especificando o que os itens das provas devem medir. A Tabela 23, no Apêndice, relaciona as habilidades e temas da Matriz de Referência de Matemática da 3ª série do Ensino Médio.

Em cada ciclo da avaliação, o conjunto de itens aplicados nos testes é posicionado na escala de proficiência a partir dos parâmetros calculados com base na Teoria de Resposta ao Item (TRI). São calculadas também medidas de proficiência para os alunos nas disciplinas avaliadas (Língua Portuguesa e Matemática), através de modelos unidimensionais. Essas medidas correspondem ao valor estimado para a habilidade do aluno na respectiva disciplina, em uma escala arbitrada com média é 250 e desvio padrão 50.

Os resultados são obtidos a partir do modelo unidimensional logístico de três parâmetros, descrito na Seção 3, Subseção 3.1. A escolha desse modelo teve como base a concepção pedagógica das matrizes de referência da avaliação, que previa a existência de um fator predominante responsável pelas respostas às questões de cada prova. Além disso, a opção pela TRI tem como intuito garantir a comparabilidade entre anos e possibilitar a construção de escalas pedagógicas (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2014; Klein, 2003).

Além da proficiência dos alunos, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), autarquia do Ministério da Educação responsável pelo SAEB, publica também os dados das respostas dos estudantes aos itens e as matrizes de referência, permitindo que sejam estimados modelos multidimensionais e hierárquicos.

O SAEB avalia bianualmente escolas e alunos das redes privadas e públicas do país, em áreas urbanas e rurais, matriculados no 5º ano e 9º ano do Ensino Fundamental (EF) e

nas 3^a/4^a séries do Ensino Médio (EM) regulares. A avaliação é feita de forma censitária na rede pública, e de forma amostral na rede privada, conforme detalhamento na Tabela 1.

Tabela 1 – Detalhamento da População Alvo do SAEB 2019

Rede de Ensino	Mínimo de Matrículas	Disciplinas	Séries	Tipo de Aplicação
Escolas públicas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	5 ^o ano e 9 ^o ano do Ensino Fundamental e 3 ^a e 4 ^a séries do Ensino Médio (tradicional e integrado)	Censitária
Escolas privadas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	5 ^o ano e 9 ^o ano do Ensino Fundamental e 3 ^a e 4 ^a séries do Ensino Médio (tradicional e integrado)	Amostral
Escolas públicas e privadas localizadas em zonas urbanas e rurais	10	Ciências Humanas e Ciências da Natureza	9 ^o ano do Ensino Fundamental	Amostral
Escolas públicas e privadas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	2 ^o ano do Ensino Fundamental	Amostral
Instituições públicas ou conveniadas com o setor público, localizadas em zonas urbanas e rurais	-	-	Creche ou pré-escola da etapa da educação infantil	Amostral (Estudo piloto)

Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019)

O universo referente ao 5^o ano do Ensino Fundamental, por exemplo, compreendeu 2.882.901 alunos em 2019, a maioria da rede pública municipal (1.942.895). Conforme veremos a seguir, cada aluno participante do 5^o ano do Ensino Fundamental responde a 44 itens de teste, resultando em um grande volume de dados disponíveis para inferência. O quantitativo de respostas por tema depende do desenho das provas, feito a partir das matrizes de referência, que discutimos a seguir.

2.1 Matrizes de Referência

Os testes do SAEB são elaborados a partir de matrizes de referência. As matrizes são compostas de temas, e em cada tema, as habilidades desejáveis para cada série e para cada disciplina, especificando o que os itens das provas devem medir. A Tabela 23, no Apêndice, exemplifica dois temas da matriz de Matemática da 3ª série do Ensino Médio.

Cada item do teste é destinado a avaliar somente uma habilidade da respectiva matriz. A Matriz de Referência de Língua Portuguesa do 5º ano do Ensino Fundamental, por exemplo, é composta de cinco temas, que agrupam 14 competências, às quais foram destinados 77 itens de teste na prova aplicada em 2019. A Tabela 2 resume o quantitativo de temas, habilidades e itens no SAEB 2019.

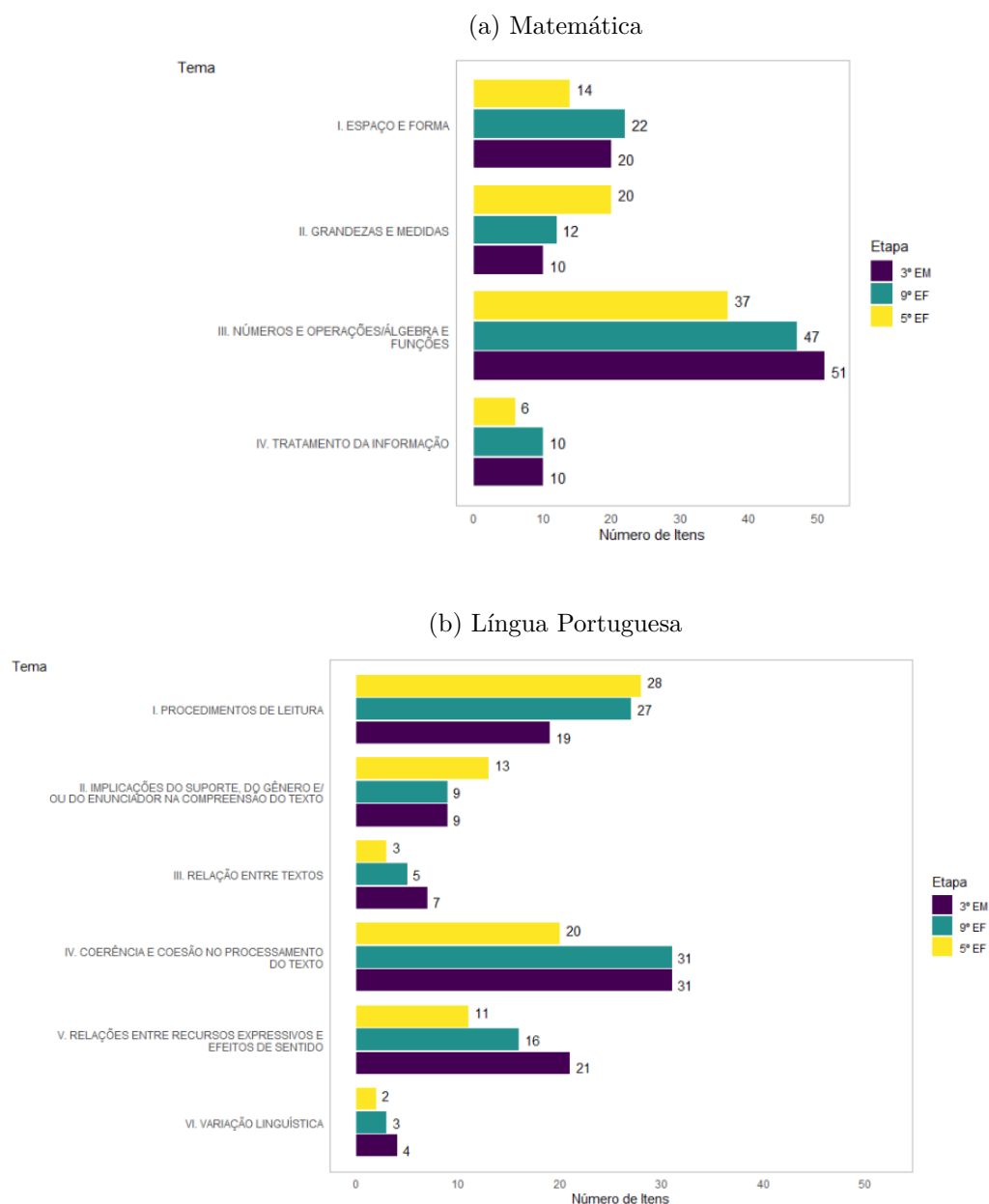
Tabela 2 – Quantitativo de temas, habilidades e itens nas matrizes de referência do SAEB 2019

Etapa	Disciplina	Temas	Habilidades	Itens
5º EF	Língua Portuguesa	5	14	77
	Matemática	4	28	77
9º EF	Língua Portuguesa	6	21	91
	Matemática	4	37	91
3º EM	Língua Portuguesa	6	21	91
	Matemática	4	35	91

Fonte: Elaboração própria a partir de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019)

A Figura 1 exibe a quantidade de itens distintos por tema das matrizes de referência de Língua Portuguesa e Matemática. Observa-se que não há homogeneidade do número de itens por tema, o que pode resultar em amostras efetivas distintas para a estimação das subhabilidades, e, conseqüentemente, diferentes graus de precisão nas estimativas geradas.

Figura 1 – Número de itens distintos por tema



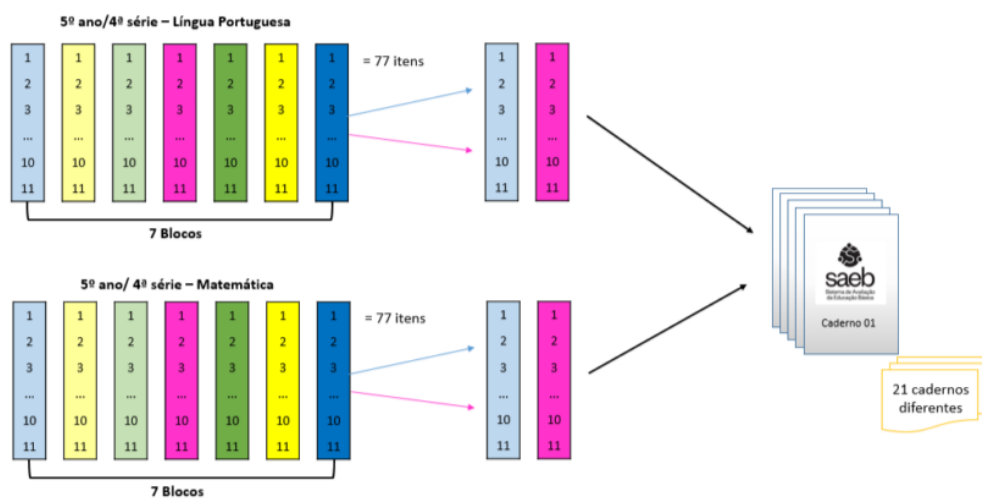
Fonte: Elaboração própria a partir de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019)

2.2 Composição das provas

Para montar os cadernos de prova do SAEB, o INEP utiliza a metodologia denominada Blocos Incompletos Balanceados (BIB), cujo objetivo é permitir que um grande número de itens (questões) seja aplicado ao conjunto de alunos avaliados, sem que cada aluno precise responder a todas as questões que fazem parte do teste naquele ciclo.

Na 3ª série do Ensino Fundamental, são 7 blocos de 11 itens, totalizando 77 itens em cada disciplina. Se numerarmos os blocos em cada disciplina avaliada (Língua Portuguesa e Matemática) de um a sete, cada caderno de teste é montado agrupando-se dois blocos de Língua Portuguesa e os dois blocos correspondentes (de mesma numeração) de Matemática. Dessa forma, são construídos $\binom{7}{2} = 21$ cadernos. Cada aluno responde a somente um caderno, que contém 22 itens de cada disciplina, ou seja, 44 itens respondidos no total. A Figura 2 ilustra essa composição:

Figura 2 – Composição dos cadernos de teste - 5º ano do Ensino Fundamental



Fonte: (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2019)

No 9º ano do Ensino Fundamental e na 3ª série do Ensino médio, são 7 blocos de 13 itens em cada disciplina e série/ano. Cada caderno de teste é montado agrupando-se dois blocos de Língua Portuguesa e os dois blocos correspondentes de Matemática. O aluno responde, portanto, a um caderno contendo 26 itens de Língua Portuguesa e 26 itens de Matemática, ou seja, 52 itens respondidos no total.

3 Modelos da Teoria da Resposta ao Item

Nesta seção descreveremos o modelo logístico unidimensional de três parâmetros, utilizado pelo INEP para os resultados do SAEB e o modelo probito unidimensional de três parâmetros, considerando respostas dicotômicas. Além disso, apresentaremos também as extensões do modelo probito para a forma multidimensional (Reckase, 2009), para o uso da matriz Q (Da Silva et al., 2019), para a abordagem multinível, considerando a hierarquia entre traços latentes (Venturelli S. L.; Gonçalves; Andrade, 2020), para o agrupamento de alunos em escolas e para a inclusão de covariáveis (Fox, 2010).

Os modelos mencionados serão expostos separadamente. Na Seção 4 apresentamos uma proposta que reúne elementos de todos esses modelos.

3.1 Modelo logístico unidimensional de três parâmetros

O modelo logístico unidimensional de três parâmetros é definido por

$$P(Y_{ij} = 1) = c_i + \frac{(1 - c_i)}{1 + \exp[-ka_i(\theta_j - b_i)]}, \quad (1)$$

em que Y_{ij} corresponde a uma resposta certa ($Y_{ij} = 1$) ou errada ($Y_{ij} = 0$) do aluno j ao item i , com $j = 1, \dots, J$ e $i = 1, \dots, I$. Os parâmetros dos itens, $a_i \in \mathbb{R}^+$, $b_i \in \mathbb{R}$ e $c_i \in [0, 1]$ correspondem, respectivamente, ao grau de discriminação do item, à sua dificuldade, e à probabilidade de um acerto ao acaso, sendo a constante k um fator de escala. A equação (1) é denominada função de resposta ao item.

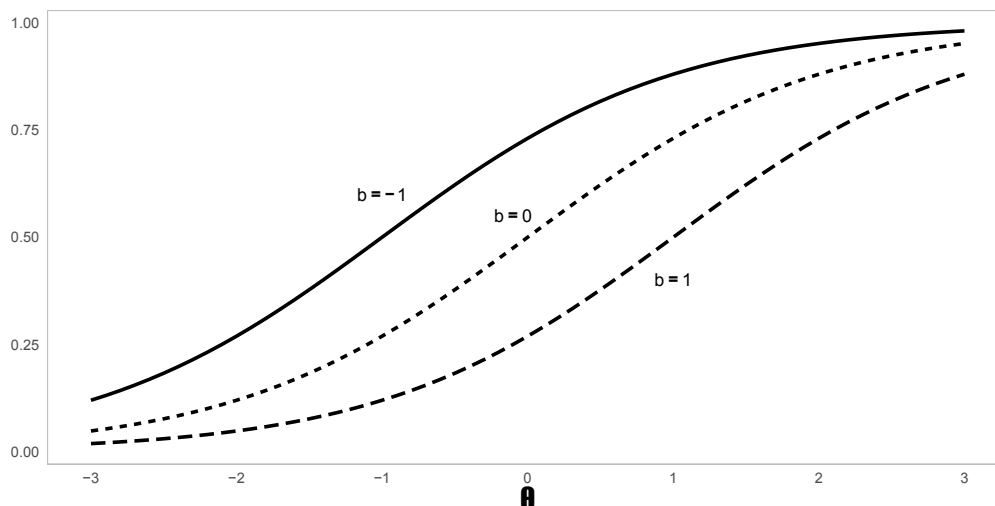
Nesse modelo, fixados os parâmetros do item, somente um parâmetro de habilidade, θ_j , importa para a probabilidade de acerto de cada item.

3.1.1 Parâmetros dos itens

Para simplificar a notação, vamos denotar a probabilidade de acerto por P . Dessa forma, segue de (1) que $P = c_i + (1 - c_i)/2$ quando $\theta_j = b_i$. Além disso, para cada valor de θ_j , quanto maior o valor de b_i menor a probabilidade de acerto, o que permite interpretar b_i como parâmetro de dificuldade do item i .

A Figura 3 exibe a forma de (1) para valores fixos de a , b , c , θ e k , conhecida como a Curva Característica do Item. Observe que, fixados os demais parâmetros, conforme aumenta o grau de dificuldade do item (b), menor a probabilidade de acerto dos indivíduos em qualquer nível de habilidade (θ). Importante observar que o parâmetro b está na mesma escala de θ .

Figura 3 – Curva Característica do Item - Parâmetro de Dificuldade



Relação entre a probabilidade de acerto a itens com diferentes graus de dificuldade ($b = -1$, $b = 0$ e $b = 1$) e o parâmetro de habilidade θ , considerando $a = 1$, $c = 0$ e $k = 1$.

Fonte: Elaboração própria

Outro parâmetro importante do modelo é o grau de discriminação do item, a_i . Observe que

$$\frac{\partial P}{\partial \theta_j} = k a_i (1 - c_i) P (1 - P) \quad (2)$$

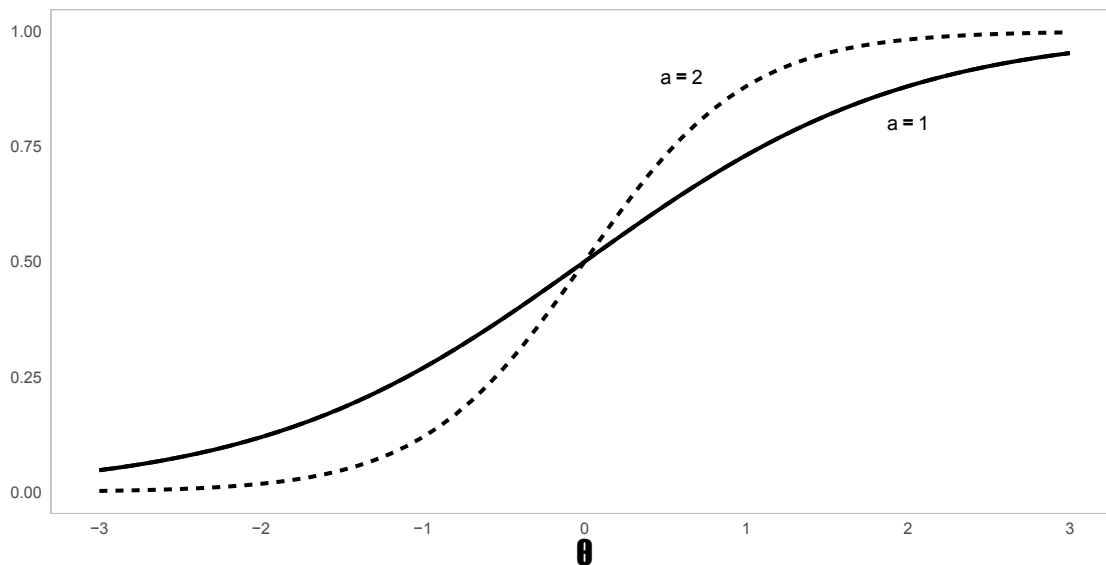
isto é, que o efeito de θ_j na função de resposta ao item depende linearmente de a_i . Quanto maior a_i , mais a probabilidade de acerto varia com θ_j . Assim, a_i é interpretado como grau de discriminação do item.

Vale notar ainda que $\frac{\partial P}{\partial \theta_j}$ assume valor máximo quando $P = 1 - P$, ou seja, quando $P = 1/2$, o que acontece quando $\theta_j = b_i$, como já vimos. Em palavras, a inclinação da função de resposta ao item é máxima quando $\theta_j = b_i$.

A Figura 4 exibe dois itens com o mesmo grau de dificuldade ($b = 0$), mas com valores distintos de discriminação a . Observe que pequenas alterações em θ na vizinhança de $\theta = b$

causam maiores alterações nas probabilidades de acerto para o item com $a = 2$. Dizemos que esse item tem, portanto, maior poder de discriminação nessa vizinhança de habilidades.

Figura 4 – Curva Característica do Item - Parâmetro de Discriminação



Relação entre a probabilidade de acerto a itens com diferentes graus de discriminação ($a = 1$ e $a = 2$) e o parâmetro de habilidade θ , considerando $b = 0$, $c = 0$ e $k = 1$.

Fonte: Elaboração própria

Por fim, observamos de (1) que c_i corresponde à parcela de probabilidade de acerto que não depende diretamente dos parâmetros de discriminação e dificuldade do item nem da habilidade do indivíduo. Assim, o parâmetro c_i pode ser interpretado como a probabilidade de acerto ao acaso.

3.1.2 Identificação de modelos da TRI

Um modelo estatístico pode ser definido como um conjunto de distribuições de probabilidade \mathcal{P} sobre um espaço amostral Ω (McCullagh, 2002). Um modelo estatístico paramétrico é um espaço paramétrico Θ com uma função $P : \Theta \rightarrow \mathcal{P}(\Omega)$, que atribui a cada $\theta \in \Theta$ uma distribuição de probabilidade P_θ em Ω . Um modelo Bayesiano requer um componente adicional na forma de uma distribuição a priori em Θ .

Ainda segundo McCullagh (2002), uma parametrização é dita identificável se parâmetros distintos dão origem a distribuições distintas, isto é, se $\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$. Assim, os parâmetros de um modelo são identificáveis se, e somente se, $P : \Theta \rightarrow \mathcal{P}(\Omega)$ é injetiva.

Considerando o modelo logístico unidimensional de três parâmetros em (1), temos, tomando $\theta'_j = \theta_j + \delta$ e $b'_i = b_i + \delta$, que $P(Y_{ij} = 1; \theta_j, b_i) = P(Y_{ij} = 1; \theta'_j, b'_i)$ para todo $\delta \in \mathbb{R}$, ou seja, os parâmetros não são identificáveis. Da mesma forma, $a' = a\delta$ e $\tilde{\theta}_j = \theta_j/\delta$ resultam na mesma probabilidade de resposta correta.

Assim, a métrica (locação e escala) dos parâmetro de habilidade é conhecida a menos de uma transformação afim. Vários tipos de restrições para obter identificação foram propostos na literatura. A locação pode ser identificada fixando-se, por exemplo, o nível de habilidade de um indivíduo específico, denominado indivíduo padrão (Fox, 2010), ou a média da habilidade da população.

Modelos Bayesianos de resposta ao item, porém, não são identificados fixando-se um único parâmetro em virtude de complicações que esse procedimento causa na especificação de prioris e condicionais completas em algoritmos MCMC. Uma forma de se obter identificação é impor restrições em hiperparâmetros do modelo Bayesiano. Assume-se, usualmente, $\boldsymbol{\theta} \sim \mathcal{N}(0, 1)$. Martín (2021) mostra que, com algumas condições adicionais, essa hipótese é suficiente para identificação dos parâmetros de habilidade e dos itens.

3.2 Modelos probito unidimensional, multidimensional e multiunidimensional de três parâmetros

O modelo probito unidimensional de três parâmetros é similar ao modelo logístico, utilizando-se, porém, como função de ligação a distribuição acumulada da Normal padrão:

$$P(Y_{ij} = 1|\theta_j) = c_i + (1 - c_i)\Phi(a_i(\theta_j - b_i)), \quad (3)$$

sendo Φ a função distribuição acumulada da Normal padrão. Se tomarmos $k = 1/1.7$ na equação (1), os modelos probito e logístico apresentam a mesma escala dos parâmetros e mesma distribuição de probabilidades.

A extensão do modelo anterior para sua versão multidimensional é dada por

$$P(Y_{ij} = 1|\boldsymbol{\theta}_j) = c_i + (1 - c_i)\Phi(\mathbf{a}_i^\top(\boldsymbol{\theta}_j - \mathbf{b}_i)), \quad (4)$$

com $\boldsymbol{\theta}_j \in \mathbb{R}^D$ vetor das D habilidades consideradas no modelo, $\mathbf{a}_i \in \mathbb{R}^D$ representando os parâmetros de discriminação do item referentes a cada uma dessas habilidades e $\mathbf{b}_i \in \mathbb{R}^D$ o vetor obtido pela repetição D vezes de b_i .

Já a versão multiunidimensional do modelo de três parâmetros pode ser obtida por meio do uso adequado da Matriz Q (Da Silva et al., 2019), uma matriz binária que indica a relação entre itens e habilidades (Tabela 3). Cada entrada na matriz Q é igual a um se o item se destina a medir a habilidade em questão, e igual a zero caso contrário. No caso multiunidimensional, cada linha da matriz Q soma um, pois cada item avalia somente uma habilidade.

Tabela 3 – Exemplo de uma matriz Q

	Habilidade 1	Habilidade 2	Habilidade 3	Habilidade 4	Habilidade 5
Item 1	1	0	0	0	0
Item 2	1	0	0	0	0
Item 3	0	1	0	0	0
Item 4	0	1	0	0	0
Item 5	0	0	1	0	0
Item 6	0	0	1	0	0
Item 7	0	0	0	1	0
Item 8	0	0	0	1	0
Item 9	0	0	0	1	0
Item 10	0	0	0	0	1

Fonte: Adaptado de (Da Silva et al., 2019)

No caso multiunidimensional, cada linha da matriz Q tem apenas uma entrada igual a um, sendo todas as demais iguais a zero. Considerando-se o modelo multidimensional ($\boldsymbol{\theta}_j \in \mathbb{R}^D$), podemos obter a versão multiunidimensional por

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i)\Phi(a_i(Q_i^\top \boldsymbol{\theta}_j - b_i)), \quad (5)$$

em que Q_i é a i -ésima linha da matriz Q . Dessa forma, a habilidade d do aluno j (θ_{jd}) explica a probabilidade da resposta ao item i somente se $q_{id} = 1$, isto é, se o item i mensura a habilidade d . Observe que, no caso multiunidimensional, temos, novamente, $a \in \mathbb{R}^I$, em oposição a $a \in \mathbb{R}^{I \times D}$ como no modelo multidimensional.

3.3 Modelo unidimensional multinível de três parâmetros com agrupamento de alunos em escolas

Outra extensão do modelo unidimensional de três parâmetros corresponde ao modelo multinível de [Fox \(2010\)](#):

$$\theta_{js} = \beta_s + e_{js} \quad (6)$$

em que

$$\beta_s = \gamma + u_s. \quad (7)$$

Nesse modelo, S escolas reúnem seus alunos J_s . A habilidade θ_{js} do aluno j na escola s é modelada por um intercepto da escola β_s e um termo de erro e_{js} . O intercepto da escola, por sua vez, é composto de uma média geral γ e um desvio da escola em relação à média.

O intercepto da escola β_s é tratado como efeito aleatório, para o qual é comum se assumir uma distribuição Normal, isto é, assume-se que $u_s \stackrel{\text{i.i.d.}}{\sim} N(0, \tau^2)$. Para o termo de erro assume-se $e_{js} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

Observe que essa especificação fixa uma dispersão comum para a habilidade dos alunos em escolas distintas (σ^2). Uma possível flexibilização desse modelo consiste em permitir que a variância de e_{js} seja específica para cada escola, isto é, $e_{js} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_s^2)$.

Devido ao efeito aleatório, as habilidades de alunos da mesma escola não são independentes:

$$\begin{aligned}
\text{Cov}(\theta_{js}, \theta_{j's'}) &= \text{Cov}(\beta_s + e_{js}, \beta'_s + e_{j's'}) \\
&= \text{Cov}(u_s + e_{js}, u'_s + e_{j's'}) \\
&= \text{Cov}(u_s, u'_s) + \text{Cov}(e_{js}, e_{j's'}) \\
&= \begin{cases} 0 & \text{se } s \neq s' \\ \tau^2 & \text{se } s = s', j \neq j' \\ \tau^2 + \sigma^2 & \text{se } s = s', j = j' \end{cases}
\end{aligned}$$

Segue que a correlação entre as habilidades de alunos distintos, na mesma escola, é dada por

$$\begin{aligned}
\rho_{j_s, j'_s} &= \frac{\text{Cov}(\theta_{j_s}, \theta_{j'_s})}{\sqrt{\text{Var } \theta_{j_s}} \sqrt{\text{Var } \theta_{j'_s}}} \\
&= \frac{\tau^2}{\sqrt{\text{Cov}(\theta_{j_s}, \theta_{j_s})} \sqrt{\text{Cov}(\theta_{j'_s}, \theta_{j'_s})}} \\
&= \frac{\tau^2}{\sqrt{\tau^2 + \sigma^2} \sqrt{\tau^2 + \sigma^2}} \\
&= \frac{\tau^2}{\tau^2 + \sigma^2}
\end{aligned}$$

e corresponde à proporção da variância das habilidades individuais atribuída à escola s .

Fox (2010) propõe, ainda, a possibilidade da inclusão de covariáveis para alunos e escolas. Seja $\mathbf{w}_s = (w_{1s}, \dots, w_{Vs})^\top$ o vetor V -dimensional de covariáveis para a escola s . O modelo passa, então, a ser dado por:

$$\theta_{js} = \mathbf{w}_s^\top \boldsymbol{\gamma} + u_s + e_{js} \quad (8)$$

com $\gamma \in \mathbb{R}^V$.

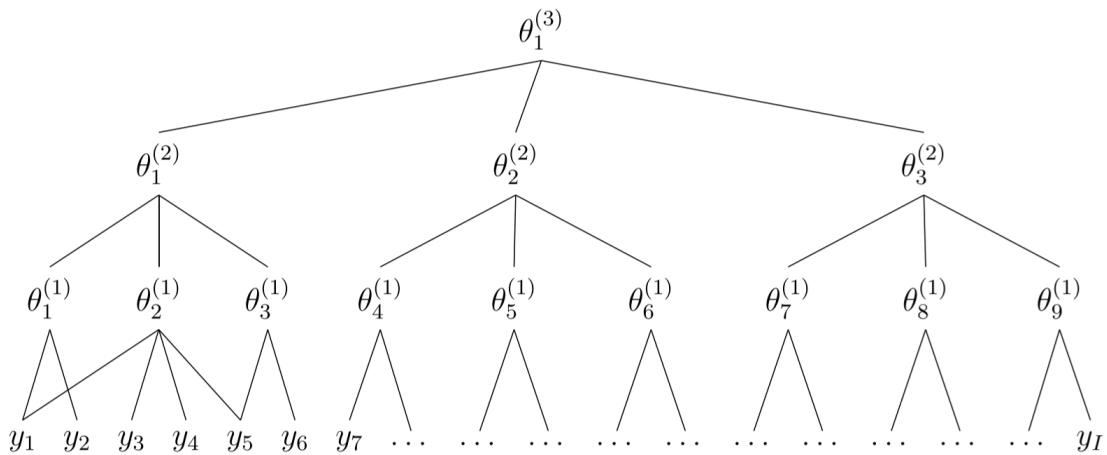
3.4 Modelo multidimensional de três parâmetros com hierarquia nos traços latentes

Venturelli S. L., Gonçalves e Andrade (2020) propõem um modelo multidimensional com hierarquia nos traços latentes. Nesse modelo, $\theta_j^{(k)} = (\theta_{1j}^{(k)}, \dots, \theta_{D_k j}^{(k)})$ corresponde ao vetor de traços latentes do indivíduo j no nível hierárquico k , $k = 1, \dots, K$. A seguinte estrutura hierárquica é considerada:

$$\theta_j^{(k)} = \lambda^{(k+1)} \theta_j^{(k+1)} + \epsilon_j^{(k)}, \quad k = 1, \dots, K - 1 \quad (9)$$

com $\lambda^{(k+1)} = [\lambda_{dd'}]$ sendo a matriz $D_k \times D_{k+1}$ de coeficientes relacionando o d -ésimo traço latente do nível k com o d' -ésimo traço latente no nível $k + 1$ para todo $j = 1, \dots, J$.

Figura 5 – Ilustração da hierarquia entre traços latentes



Fonte: Venturelli S. L., Gonçalves e Andrade (2020)

Os autores impõem duas restrições sobre as matrizes $\lambda^{(k+1)}$. A primeira é que um traço latente no nível k esteja relacionado somente a um traço latente no nível $k + 1$. Isto implica que cada linha da matriz $\lambda^{(k+1)}$ tem somente um elemento não nulo. A segunda é uma restrição usual em modelos de análise de fatores, na qual um traço latente no nível $k + 1$ está relacionado com pelo menos três traços do nível k , necessária para haver identificação do modelo.

Na estrutura hierárquica em (9), o nível mais detalhado de traços latentes $\theta_j^{(1)}$ é o que determina diretamente as probabilidades de respostas. Dessa forma, a estrutura de análise

de fatores combina esses traços para definir os demais traços nos níveis superiores (indicados por índices maiores). Assim, os traços nos níveis superiores (índices maiores) podem ser interpretados como aqueles que melhor caracterizam uma habilidade geral definida como uma composição de habilidades do nível inferior.

Como $\boldsymbol{\theta}_j^{(1)}$ determina diretamente as probabilidades de respostas, temos que

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i) \Phi(\mathbf{a}_i^\top (\boldsymbol{\theta}_j^{(1)} - \mathbf{b}_i)) \quad (10)$$

para todo i e j , sendo Φ a função distribuição acumulada da Normal padrão.

Na abordagem Bayesiana, a especificação completa das distribuições a priori para os traços latentes é feita assumindo-se ainda:

$$\theta_{dj}^{(K)} \sim N(0, 1), \quad \forall d, \forall j. \quad (11)$$

$$\epsilon_{dj}^{(k)} | \lambda_d^{(k+1)} \sim N \left(0, 1 - \sum_{d'=1}^{D_{k+1}} (\lambda_{dd'}^{(k+1)})^2 \right). \quad (12)$$

$$\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \left[\prod_{j=1}^J \prod_{d=1}^{D_K} \pi(\theta_{dj}^{(K)}) \right] \left[\prod_{j=1}^J \prod_{k=1}^{K-1} \prod_{d=1}^{D_k} \pi(\theta_{dj}^{(k)} | \lambda_{dd'}^{(k+1)}, \theta_{d'j}^{(k+1)}) \right] \left[\prod_{k=2}^K \prod_{d=1}^{D_k} \prod_{d'=1}^{D_{k+1}} \pi(\lambda_{dd'}^{(k)}) \right]. \quad (13)$$

Os autores assume, (12) para que a escala de θ_j^k mantenha distribuição $N(0, 1)$ para todo j e k . Vale notar que somente um elemento da soma $\sum_{d'=1}^{D_{k+1}} (\lambda_{dd'}^{(k+1)})^2$ é diferente de zero. Além disso, a adoção da priori $\lambda_{dd'}^{(k)} \sim \text{Uniforme}(-1, 1)$ para todos os coeficientes das matrizes $\boldsymbol{\lambda}^{(k+1)}$ faz com que cada $\lambda_{dd'}^{(k+1)}$ seja a correlação entre $\theta_{jd}^{(k)}$ e $\theta_{jd'}^{(k+1)}$.

Por fim, completando a especificação Bayesiana do modelo, [Venturelli S. L., Gonçalves e Andrade \(2020\)](#) assumem ainda que $(\mathbf{a}_i, \mathbf{b}_i) \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, com $\boldsymbol{\mu}$ e $\boldsymbol{\Lambda}$ ambos conhecidos e $\boldsymbol{\Lambda} = \text{diag} \{ \sigma_{a_1}^2, \dots, \sigma_{a_{D_1}}^2, \sigma_{b_i}^2 \}$. Com as hipóteses mencionadas, os autores mostram que todo $\theta_{dj}^{(k)}$ tem distribuição a priori normal com média zero e variância igual a um, de modo a fixar a escala dos traços latentes em todos os níveis hierárquicos, permitindo a identificação do modelo em relação a $\boldsymbol{\theta}$.

4 Modelo multidimensional com hierarquia nos traços latentes e agrupamento de indivíduos

Nesta seção vamos especificar um modelo multiunidimensional de três parâmetros com hierarquia nos traços latentes, covariáveis para alunos e escolas, e agrupamento de alunos em escolas.

Considerando o modelo proposto por [Venturelli S. L., Gonçalves e Andrade \(2020\)](#), vamos assumir a seguinte estrutura hierárquica:

$$\boldsymbol{\theta}_j^{(k)} = \boldsymbol{\lambda}^{(k+1)} \boldsymbol{\theta}_j^{(k+1)} + \boldsymbol{\epsilon}_j^{(k)}, \quad k = 1, \dots, K - 1. \quad (14)$$

para $\boldsymbol{\theta}_j^{(k)} = (\theta_{1j}^{(k)}, \dots, \theta_{D_k j}^{(k)})$ vetor de traços latentes do indivíduo j no nível hierárquico k , para $k = 1, \dots, K$. Em (14), $\boldsymbol{\lambda}^{(k+1)} = [\lambda_{dd'}]$ é a matriz $D_k \times D_{k+1}$ de coeficientes relacionando o d -ésimo traço latente do nível k com o d' -ésimo traço latente no nível $k + 1$ para todo $j = 1, \dots, J$.

4.1 Inclusão de covariáveis e agrupamento de alunos em escolas

Este trabalho propõe a inclusão das covariáveis e o agrupamento de alunos em escolas no nível superior do modelo ($k = K$), assumindo $\boldsymbol{\theta}_j^{(K)}$ unidimensional. Assim, utilizando uma modificação do modelo de [Fox \(2010\)](#), considerando escolas ($s = 1, \dots, S$) agrupando os alunos ($j = 1, \dots, J_s$), temos

$$\theta_{js}^{(K)} = \mathbf{w}_s^\top \boldsymbol{\gamma} + u_s + e_{js}. \quad (15)$$

sendo $\mathbf{w}_s \in \mathbb{R}^V$ o vetor de covariáveis da escola s , u_s o desvio em relação à parcela da habilidade atribuída à escola e e_{js} desvio idiossincrático do aluno.

Além disso, assumimos que $u_s \sim N(0, \tau^2)$ e $e_{js} \sim N(0, \sigma^2)$, isto é, o termo de erro no nível do aluno tem variância comum entre todos os alunos em todas as escolas. Outra possível especificação seria $e_{js} \sim N(0, \sigma_s^2)$. Neste caso, os desvios referentes aos alunos têm variância específica em cada escola, e poderíamos adotar, por exemplo, $\sigma_s^2 \sim \text{Gamma}(\nu/2, \nu/2)$, com

ν conhecido. Nesta dissertação, vamos considerar o caso com variância comum entre as escolas.

4.2 Multiunidimensionalidade

A última modificação necessária para o modelo proposto é torná-lo multiunidimensional. O modelo de [Venturelli S. L., Gonçalves e Andrade \(2020\)](#) é multidimensional, de forma que cada item pode estar associado a múltiplas habilidades. No modelo aqui proposto, vamos considerar o caso particular em que cada item testa uma única habilidade, que se alcança pelo uso da matriz Q :

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i) \Phi(a_i(\mathbf{Q}_i^\top \boldsymbol{\theta}_j - b_i)). \quad (16)$$

Vale lembrar, ainda, que a estrutura hierárquica em (14) será determinada, nas aplicações empíricas, pelas matrizes de referência que reúnem habilidades em áreas ou temas.

4.3 Elicitação Bayesiana do modelo proposto

Conforme exposto, descrevemos o modelo proposto na abordagem Bayesiana pelo conjunto de equações a seguir. O modelo para os dados é dado por (16), repetida a seguir:

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i) \Phi(a_i(\mathbf{Q}_i^\top \boldsymbol{\theta}_j - b_i)). \quad (16)$$

A equação (14) estabelece estrutura hierárquica entre os parâmetros de habilidade

$$\boldsymbol{\theta}_j^{(k)} = \boldsymbol{\lambda}^{(k+1)} \boldsymbol{\theta}_j^{(k+1)} + \boldsymbol{\epsilon}_j^{(k)}, \quad k = 1, \dots, K - 1. \quad (14)$$

com o parâmetro mais geral de habilidade dado pela equação (15):

$$\boldsymbol{\theta}_{js}^{(K)} = \mathbf{w}_s^\top \boldsymbol{\gamma} + u_s + e_{js}. \quad (15)$$

Além disso, assumimos que u_s e e_{js} são independentes, com $u_s \stackrel{\text{i.i.d.}}{\sim} N(0, \tau^2)$ e $e_{js} \stackrel{\text{i.i.d.}}{\sim} N(0, 1 - \tau^2)$, o que implica em

$$\theta_j^{(K)} \sim N(\mathbf{w}_s^\top \boldsymbol{\gamma}, 1). \quad (17)$$

De maneira similar a [Venturelli S. L., Gonçalves e Andrade \(2020\)](#), assumimos ainda que

$$\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{s=1}^S \left[\prod_{j=1}^{J_s} \pi(u_s) \pi(\theta_{js}^K | u_s) \right] \left[\prod_{j=1}^{J_s} \prod_{k=1}^{K-1} \prod_{d=1}^{D_k} \pi(\theta_{jkd}^k | \lambda_{dd'}^{(k+1)}, \theta_{jkd'}^{(k+1)}) \right] \left[\prod_{k=2}^K \prod_{d=1}^{D_k} \prod_{d'=1}^{D_{k+1}} \pi(\lambda_{dd'}^{(k)}) \right], \quad (18)$$

$$\epsilon_{jkd}^{(k)} | \lambda_{dd'}^{(k+1)} \sim N\left(0, 1 - \sum_{d'=1}^{D_{k+1}} (\lambda_{dd'}^{(k+1)})^2\right), \quad (19)$$

$$\lambda_{dd'}^{(k)} \sim \text{Uniforme}(0, 1), \quad (20)$$

o que representa independência entre efeitos aleatórios das escolas (u_s), independência das respostas entre alunos condicional a u_s , independência das subhabilidades condicional a λ e à habilidade do nível imediatamente superior, e distribuições de λ independentes. Observe, ainda, de [14](#) e [19](#) que $\pi(\theta_{jkd}^k | \lambda_{dd'}^{(k+1)}, \theta_{jkd'}^{(k+1)}) \sim N(\lambda_{dd'}^{k+1} \theta_{jkd'}^{(k+1)}, 1 - (\lambda_{dd'}^{k+1})^2)$.

Por fim, nos estudos simulados e na inferência com os dados do SAEB, utilizamos as seguintes distribuições a priori:

$$a_i \sim U[0, 10]$$

$$b_i \sim N(0, 100)$$

$$c_i \sim U[0, 0.25]$$

$$\gamma \sim N(0, 100)$$

$$\tau^2 \sim U(0, 1).$$

Observe que apenas a priori de c_i é informativa, que assim adotamos para facilitar a estimação do parâmetro c , que tipicamente apresenta maior viés e variância na estimação.

5 Estimação do modelo proposto e variações

Nesta seção, vamos descrever brevemente como serão estimados os parâmetros do modelo proposto e de suas variações, tanto nas simulações quanto na aplicação aos dados do SAEB.

Será utilizado em todas as estimações o software Stan ([Stan Development Team, 2023](#)) através do pacote cmdstanR ([Gabry; Češnovar; Johnson, 2023](#)), desenvolvido na linguagem estatística R ([R Core Team, 2023](#)).

A estimação no Stan utiliza um algoritmo de Monte Carlo de Cadeia de Markov (MCMC). Em particular, adota-se o *No U-Turn Sampler* (NUTS), uma variante do método de *Monte Carlo Hamiltoniano* ([Hoffman; Gelman, 2011](#)).

5.1 Hamiltonian Monte Carlo

O Método de Monte Carlo Hamiltoniano (HMC) é uma técnica de amostragem que foi desenvolvida para explorar eficientemente espaços de parâmetros de alta dimensão. Ele combina conceitos da física hamiltoniana com métodos de Monte Carlo para produzir amostras representativas da distribuição alvo.

O HMC evita o comportamento ineficiente de *random walks* e supera problemas de correlação entre variáveis que afetam muitos métodos MCMC, ao tomar passos informados por gradiente de primeira ordem. Essas características permitem que ele convirja para distribuições-alvo de alta dimensão mais rapidamente do que métodos mais simples, como o *Metropolis Random Walk* ou a amostragem de Gibbs.

Na física bidimensional, a dinâmica hamiltoniana pode ser visualizada como o movimento de uma partícula fictícia (representando o estado atual do sistema), em uma superfície de altura variada.

A partícula possui uma energia potencial $U(\theta)$, proporcional à altura da superfície, e uma energia cinética, $K(\rho)$, em que θ representa o estado atual e ρ é uma variável auxiliar correspondente ao *momentum* da partícula.

Na aplicação ao MCMC, a energia potencial correspondente ao inverso aditivo do log da densidade ($U(\theta) = -\log p(y, \theta)$). A partir da energia potencial e da energia cinética,

resolve-se a trajetória da partícula através das equações da dinâmica hamiltoniana. As posições da partícula na trajetória encontrada correspondem às propostas para amostragem, que são aceitas ou não de acordo com a consistência dessas posições com a energia do sistema.

Para uma descrição precisa e completa do HMC sugerimos a leitura de (NEAL, 2011).

5.2 No U-Turn Sampler

Apesar dos benefícios em relação aos métodos mais simples de amostragem por MCMC, o desempenho do HMC é altamente sensível a dois parâmetros: um tamanho de passo ε e um número desejado de passos L . Em particular, se L for muito pequeno, o algoritmo exibirá um comportamento de *random walk*, indesejável, enquanto se L for muito grande, o algoritmo desperdiçará computação.

O *No U-Turn Sampler* (NUTS) é uma extensão do HMC que elimina a necessidade de definir um número de passos L . O NUTS utiliza um algoritmo recursivo para construir um conjunto de pontos candidatos prováveis que abrange uma ampla gama da distribuição-alvo, parando automaticamente quando começa a retroceder e refazer seus passos.

O NUTS também conta com um método para adaptar o parâmetro de tamanho de passo ε dinamicamente. Dessa forma, o algoritmo pode ser usado sem ajustes manuais.

5.3 Estatísticas de diagnóstico

Uma das formas de verificar a convergência de uma cadeia para a distribuição de equilíbrio é compará-la a outras cadeias inicializadas aleatoriamente. Essa é a motivação da estatística \hat{R} de Gelman e Rubin (1992).

O \hat{R} mede a razão entre a variância média das amostras dentro de cada cadeia em relação à variância das amostras agrupadas entre as cadeias; se todas as cadeias estiverem em equilíbrio, esses valores serão iguais e o \hat{R} será igual a um. Se as cadeias não convergirem para uma distribuição comum, o \hat{R} será maior do que um.

Outro problema comum nos métodos MCMC é a existência de autocorrelação, que aumenta

a incerteza da estimação de quantidades de interesse a posteriori. Uma das formas de diagnosticá-lo é o cálculo do tamanho efetivo da amostra (*Effective Sample Size - ESS*).

De forma geral, o ESS de uma quantidade de interesse captura *quantas amostras independentes contêm o mesmo volume de informação que a amostra dependente obtida pelo algoritmo MCMC*. Assim, para amostras independentes, o ESS equivale ao número de iterações. Para amostras correlacionadas, o ESS será inferior a esse número.

Duas estatísticas de amostra efetiva são calculadas no Stan: *ESS Bulk* e *ESS Tail*. O ESS Bulk estima a eficiência amostral para medidas resumo de locação tais como a média e mediana. Para resultados finais, [Stan Development Team \(2023\)](#) recomenda que o ESS Bulk seja maior do que 100 vezes o número de cadeias. Por exemplo, ao rodar quadro cadeias, recomenda-se um ESS Bulk de pelo menos 400. Já o ESS Tail calcula o mínimo entre as amostras efetivas dos quantis 5% e 95%, sendo útil para diagnosticar problemas de escala e mistura lenta na cauda da distribuição de interesse.

6 Simulações

Nesta seção faremos dois conjuntos de análises simuladas do modelo. No primeiro, vamos gerar dados de quatro modelos distintos em termos de hierarquia entre habilidades e presença ou ausência de fatores multinível. Em seguida estimaremos o modelo proposto em cada conjunto de dados gerado e avaliaremos os resultados. No segundo, vamos gerar dados do modelo proposto completo e testar quatro variações desse modelo, que desconsiderem hierarquia nos traços latentes e/ou fatores multinível.

6.1 Desempenho do modelo proposto em diferentes conjuntos de dados

A Tabela 4 elenca as variáveis do modelo proposto, sua descrição e se vamos considerar diferentes valores na análise desta seção.

Tabela 4 – Fatores do processo gerador de dados

Fator	Descrição	Variação na simulação
S	Total de escolas	Não
J_s	Número de alunos na escola s	Não
K	Níveis hierárquicos no modelo	Não
D	Total de habilidades no segundo nível hierárquico	Não
W	Covariáveis das escolas	Sim
λ	Relação hierárquica entre as habilidades	Sim
γ	Coefficientes das covariáveis das escolas	Sim
τ^2	Variância do “efeito escola”	Sim

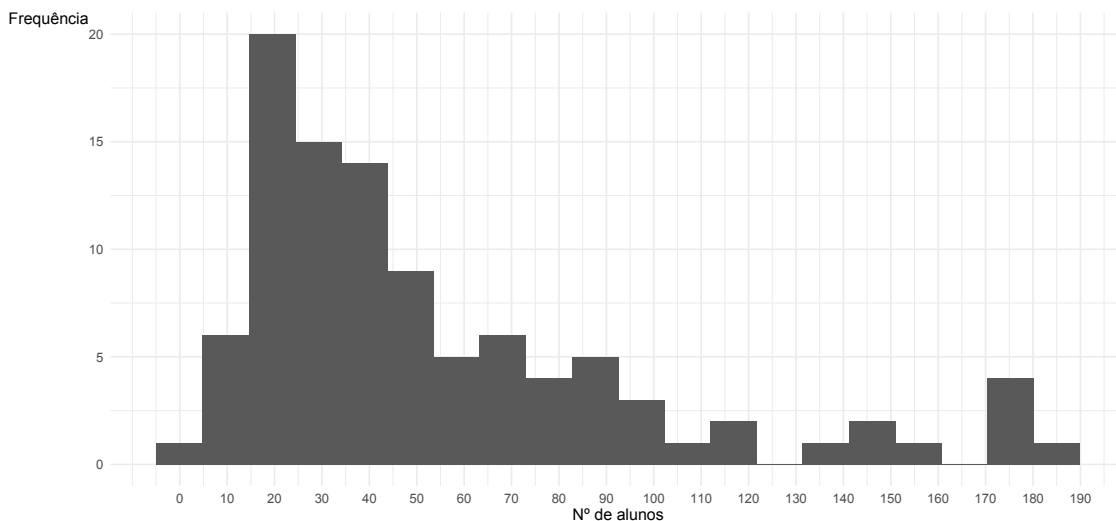
Fonte: Elaboração própria

Na aplicação empírica, o total de escolas S corresponde ao número de escolas que participaram da avaliação ou um subconjunto de interesse. Para a simulação, adotaremos o valor arbitrário $S = 100$.

Na simulação, o número de alunos J_s de cada escola foi amostrado do SAEB, resultando no total de 5.438 alunos distribuídos em 100 escolas com a distribuição de tamanho conforme a Figura 6.

Na aplicação, estimaremos uma habilidade geral $\theta^{(2)}$ e uma habilidade $\theta_d^{(1)}$ para cada tema d da Matriz de Referência do SAEB, o que resulta em dois níveis hierárquicos, e, portanto,

Figura 6 – Distribuição do número de alunos avaliados por escola no SAEB - 3^a série do Ensino Médio



Distribuição obtida a partir de uma amostra aleatória de 100 escolas.

Fonte: Elaboração própria

$K = 2$. Além disso, como a Matriz de Referência do SAEB agrupa os itens em quatro temas, adotamos $D = 4$.

O impacto de λ no modelo vem das equações (14) e (?). Observe que, quando $\lambda_d = 1$, temos $\theta_{jd}^{(1)} = \theta_j^{(2)}$, de modo que $\lambda = [1, 1, 1, 1]$ corresponde, efetivamente, ao caso unidimensional.

Por outro lado, quando $\lambda_d = 0$, $\theta_{jd}^{(1)}$ é independente de $\theta_j^{(2)}$, sendo completamente determinado por ϵ_{jd} . Nesse caso, os efeitos da escola não influenciam a sub-habilidade d , que passa a depender somente do aluno j . O caso em que $\lambda = [0, 0, 0, 0]$ corresponde, portanto, a uma situação multidimensional com habilidades independentes.

Para a simulação, adotaremos dois cenários:

- $\lambda = [1, 1, 1, 1]$: habilidade geral determina majoritariamente as probabilidades de resposta, correspondendo a um teste unidimensional.
- $\lambda = [0.9, 0.8, 0.7, 0.6]$: algumas sub-habilidades tem maior relação com a habilidade geral do que outras, similar ao que esperamos observar na aplicação empírica.

Tal como na aplicação do SAEB, serão consideradas duas covariáveis: Nível Socioeconômico

médios dos alunos da escola (NSE) e o Índice de Complexidade da Gestão Escolar (ICG). O NSE é um traço latente calculado pelo INEP a partir de dados do questionário contextual dos alunos, e o ICG um traço latente calculado a partir de dados da escola como número de alunos, turnos e modalidades de ensino que oferta.

Como γ determina o efeito das covariáveis sobre a habilidade geral e sub-habilidades, também adotaremos dois cenários:

- $\gamma = [0, 0]$: covariáveis irrelevantes; e
- $\gamma = [0.5, 0.2]$: covariáveis com efeitos na mesma escala da soma dos fatores desconhecidos da escola (u_s) e da variação individual da habilidade (e_{js}).

Por fim, as habilidades gerais de cada aluno são obtidas de $u_s \sim N(0, \tau^2)$ e $e_{js} \sim N(0, 1 - \tau^2)$. É possível demonstrar que τ^2 corresponde à proporção da variância da habilidade geral $\theta_{js}^{(2)}$ que se deve a variações entre as médias das escolas:

De (15) temos que

$$\begin{aligned} \text{Var}(\theta_{js}^{(2)}) &= \text{Var}(\mathbf{w}_s^\top \boldsymbol{\gamma} + u_s + e_{js}) \\ &= \text{Var}(u_s + e_{js}) \\ &= \tau^2 + \sigma^2, \end{aligned}$$

pois $\mathbf{w}_s^\top \boldsymbol{\gamma}$ é constante e u_s e e_{js} são independentes com $u_s \sim N(0, \tau^2)$ e $e_{js} \sim N(0, \sigma^2)$. Como $\tau^2 + \sigma^2 = 1$, temos que $\frac{\tau^2}{\tau^2 + \sigma^2} = \tau^2$, logo τ^2 é a proporção da variância em $\theta_{js}^{(2)}$ que se deve a perturbações nas médias das escolas.

Podemos obter $\hat{\tau}^2$ empiricamente como

$$\hat{\tau}^2 = \frac{\text{Var}([\hat{\mu}_1, \dots, \hat{\mu}_S])}{\text{Var}([\hat{\mu}_1, \dots, \hat{\mu}_S]) + \text{Var}([\hat{\theta}_{11}^{(2)}, \dots, \hat{\theta}_{JS}^{(2)})},$$

com $\hat{\mu}_s$ sendo a proficiência média da escola s e $\hat{\theta}_{js}^{(2)}$ a proficiência do aluno j na escola s . Utilizando os dados do SAEB 2021, com resultados calculados pelo INEP através do

modelo logístico unidimensional de três parâmetros, obtemos $\hat{\tau}^2 \approx 0.2$.

Observe, da equação (15), que τ^2 tem efeito direto em $\theta_{js}^{(2)}$. Se considerarmos o caso extremo em que não há efeito de covariáveis ($\gamma = [0, 0]$), então quando $\tau^2 = 1$, temos $u_s \sim N(0, 1)$ e $e = 0$, e daí que a habilidade geral corresponde ao efeito da escola u_s , com $\theta_{js}^{(2)} = u_s \sim N(0, 1)$. Neste caso, formam-se grupos de alunos com a mesma habilidade geral, correspondentes ao efeito da escola a que pertencem.

Por outro lado, quando $\tau^2 = 0$, temos $e_{js} \sim N(0, 1)$, e, se $\gamma = 0$, segue que $\theta_{js}^{(2)} = e_{js} \sim N(0, 1)$. Nessa situação não há efeito escola, toda a variação na habilidade geral deve-se à variação nas habilidades individuais.

Quando $\gamma \neq [0, 0]$, o efeito escola passa a ser composto de um efeito fixo, devido às covariáveis, e o efeito aleatório u_s . Neste caso, soma-se à distribuição de $\theta_{js}^{(2)}$ o valor de $\gamma^\top \mathbf{w}_s$. Como w_{1s} e w_{2s} assumem valores em distribuições aproximadamente normais, cada, a média de $\theta_{js}^{(2)}$ passa depende dos coeficientes de $\gamma = [\gamma_1, \gamma_2]$, de modo que, quanto maiores as magnitudes desses coeficientes, mais a distribuição de $\theta^{(2)}$ passa a ser determinada pelo efeito fixo.

Assim, considerando as variações propostas, serão analisados os cenários descritos na Tabela 5.

Tabela 5 – Cenários para análise de variáveis do modelo

Cenário	Descrição	Valor das variáveis
1. Nulo	Unidimensional, sem covariáveis e sem “efeito escola”	$\lambda = [1, 1, 1, 1]$ $\gamma = [0, 0]$ $\tau^2 = 0$
2. Hierárquico	Hierárquico, multidimensional, sem efeito de covariáveis e sem “efeito escola”	$\lambda = [0.9, 0.8, 0.7, 0.6]$ $\gamma = [0, 0]$ $\tau^2 = 0$
3. Multinível	Unidimensional, com covariáveis e “efeito escola”	$\lambda = [1, 1, 1, 1]$ $\gamma = [0.5, 0.2]$ $\tau^2 = 0.2$
4. Completo	Hierárquico, multidimensional, com covariáveis e “efeito escola”	$\lambda = [0.9, 0.8, 0.7, 0.6]$ $\gamma = [0.5, 0.2]$ $\tau^2 = 0.2$

Fonte: Elaboração própria.

6.1.1 Resultados do modelo proposto em cenários simulados de dados

A Tabela 6 apresenta a média de Erro Quadrático Médio (EQM), Viés e Variância do modelo proposto nos quatro conjuntos de dados simulados. Em cada cenário, foram obtidas 1.000 amostras coletadas após *warm-up* de 1.000 iterações em 4 cadeias de Markov com inicializações distintas, totalizando 4.000 amostras.

Considere um estimador $\hat{\theta}_n = f(Y)$ de θ , por exemplo, para uma amostra de tamanho n . Então o EQM desse estimador é dado por $\text{EQM}(\hat{\theta}_n) = E_{\theta|Y}[(\hat{\theta}_k - \theta)^2]$, com $k = 1, \dots, n$ sendo o número de estimativas obtidas para θ . O Viés desse estimador é dado por $\text{Viés}(\hat{\theta}_n) = E_{\theta|Y}(\hat{\theta}_n) - \theta$, e a variância por $\text{Var}(\hat{\theta}_n) = E_{\theta|Y}[(\hat{\theta}_n - E(\hat{\theta}_n))^2]$. Nas Tabelas 6 e 12 são reportadas as médias de cada uma dessas medidas para cada conjunto de parâmetros estimados, isto é, reporta-se a média, por exemplo, $\text{EQM}(\hat{\theta}^{(2)}) = \frac{\sum_j \text{EQM}(\theta_j^{(2)})}{J}$ para $j = 1, \dots, J$.

Os resultados indicam que o EQM de $\theta^{(2)}$ e, portanto, de $\theta^{(1)}$, acompanham a complexidade do modelo gerador dos dados, sendo mais baixos para os Cenários 1 (Nulo) e 3 (Multinível), e mais altos para os Cenários 2 (Hierárquico) e 4 (Completo). O maior EQM se deve à maior variância dos estimadores nos cenários 2 e 4, sendo o viés de estimação baixo em todos os cenários. Vale notar ainda que não há diferenças significativas de viés ou variância de estimação entre cenários nos parâmetros dos itens.

As Figuras 7 a 10 e as Tabelas 7 a 10 apresentam os resultados para cada um dos cenários. Nelas, os gráficos de dispersão contrastam os valores verdadeiros e os valores estimados, que correspondem à média à posteriori de cada parâmetro. Já os gráficos de densidade correspondem às densidades a posteriori marginais empíricas, com uma curva para cada cadeia da amostragem.

As Tabelas 7 a 10 apresentam o valor verdadeiro, o valor estimado (média a posteriori e intervalo de credibilidade a posteriori de 95%), desvio padrão da distribuição marginal a posteriori e as estatísticas de diagnóstico \hat{R} e tamanhos efetivos da amostra no núcleo e nas caudas da distribuição (ESS Bulk e ESS Tail, respectivamente). No caso dos parâmetros $\theta^{(2)}$, $\theta^{(1)}$, a , b e c , são exibidos somente um exemplar de cada, visto que somam, ao todo, milhares de parâmetros, um para cada aluno no caso dos θ , e um para cada item do teste

Tabela 6 – Média do Erro Quadrático Médio, Viés e Variância da estimação no modelo proposto em cada conjunto simulado de dados

Medida	Cenário	$\theta^{(2)}$	a	b	c	$\theta^{(1)}$
Erro Quadrático Médio	1 - Nulo	0.028	0.022	0.003	0.0007	[0.028, 0.028, 0.028, 0.028]
	2 - Hierárquico	0.165	0.033	0.003	0.0006	[0.098, 0.095, 0.119, 0.101]
	3 - Multinível	0.038	0.099	0.003	0.0005	[0.038, 0.038, 0.038, 0.038]
	4 - Completo	0.161	0.012	0.003	0.0004	[0.099, 0.111, 0.101, 0.109]
Viés	1 - Nulo	0.006	0.046	0.016	0.0034	[0.006, 0.006, 0.006, 0.006]
	2 - Hierárquico	-0.017	0.040	-0.021	0.0003	[-0.016, -0.020, -0.023, 0.004]
	3 - Multinível	-0.041	0.032	-0.036	0.0063	[-0.041, -0.041, -0.041, -0.041]
	4 - Completo	0.025	0.052	0.015	0.0007	[0.021, 0.017, 0.014, 0.033]
Variância	1 - Nulo	0.030	0.044	0.003	0.0008	[0.031, 0.031, 0.032, 0.031]
	2 - Hierárquico	0.164	0.080	0.004	0.0007	[0.092, 0.091, 0.117, 0.105]
	3 - Multinível	0.040	0.099	0.006	0.0005	[0.041, 0.041, 0.041, 0.043]
	4 - Completo	0.150	0.010	0.004	0.0007	[0.092, 0.104, 0.101, 0.104]

Fonte: Elaboração própria.

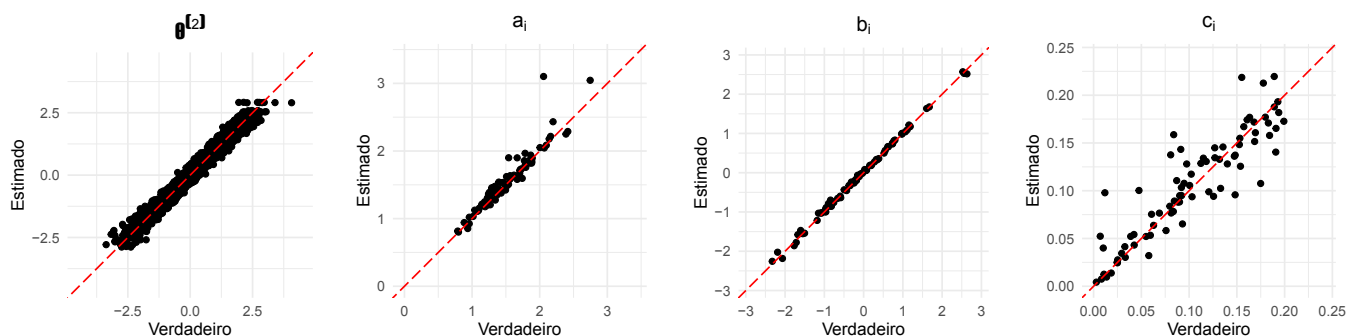
Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

no caso de a, b e c .

De modo geral, observamos nos resultados que as estimativas pontuais se aproximam dos valores verdadeiros em todos os cenários. Vale notar que as estatísticas de diagnóstico da inferência de λ , parâmetro da hierarquia, indicam baixo tamanho da amostra efetiva no núcleo (*ESS bulk*) e nas caudas (*ESS tail*) da distribuição, além de baixa concordância entre cadeias (\hat{R} elevados) nos cenários em que não há de fato hierarquia no processo gerador dos dados verdadeiros (Cenários 1 e 3). Apesar disso, as estimativas pontuais e intervalares desses parâmetros foram precisas, e não prejudicaram a estimação dos demais parâmetros. Os valores elevados de \hat{R} e reduzidos de amostra efetiva se devem à escala dos valores estimados, muito próximos da unidade, variando na terceira casa decimal, para um parâmetro de correlação, que assume valores entre 0 e 1, como se observa nos traceplots (ver Apêndice).

A Figura 7 exibe os resultados para o cenário em que os dados foram gerados de um modelo unidimensional, sem o agrupamento de alunos em escolas e sem o uso de covariáveis. Os gráficos de dispersão indicam estimativa dos parâmetros de habilidade e dos itens sem viés evidente e com baixa variância.

Figura 7 – Cenário 1 - Nulo



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Na Tabela 7 podemos ver que os valores estimados estão próximos dos valores verdadeiros para todos os parâmetros, com baixo desvio padrão. Os valores para as estatísticas de diagnóstico estão de acordo com o indicado conforme a discussão realizada na Subseção 5.3 da Seção 5, com exceção dos parâmetros de hierarquia, que, apesar do \hat{R} elevado e baixos tamanhos de amostra efetiva, apresentaram estimativa pontual e intervalar precisa, e não prejudicaram a estimação dos demais parâmetros.

Tabela 7 – Resultados do Cenário 1 - Nulo

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	-0.26	-0.15 [-0.38, 0.08]	0.14	1.00	4957.58	3026.90
$\theta_{1,9}^{(1)}$	0.73	0.73 [0.48, 0.99]	0.15	1.00	4772.41	3117.83
a_{11}	1.40	1.42 [1.33, 1.51]	0.05	1.00	1216.55	2810.85
b_{32}	0.39	0.36 [0.30, 0.42]	0.03	1.01	549.20	1781.49
c_{57}	0.19	0.19 [0.18, 0.21]	0.01	1.00	4131.38	2976.43
γ_1	0.00	-0.01 [-0.03, 0.01]	0.01	1.03	126.25	375.29
γ_2	0.00	-0.00 [-0.03, 0.02]	0.01	1.03	189.46	331.40
λ_1	1.00	1.00 [1.00, 1.00]	0.00	1.51	7.32	34.29
λ_2	1.00	1.00 [1.00, 1.00]	0.00	1.97	5.57	11.54
λ_3	1.00	1.00 [1.00, 1.00]	0.00	1.55	7.26	12.11
λ_4	1.00	1.00 [1.00, 1.00]	0.00	1.60	6.85	11.36
τ^2	0.00	0.00 [0.00, 0.00]	0.00	1.07	52.19	95.70

Fonte: Elaboração própria.

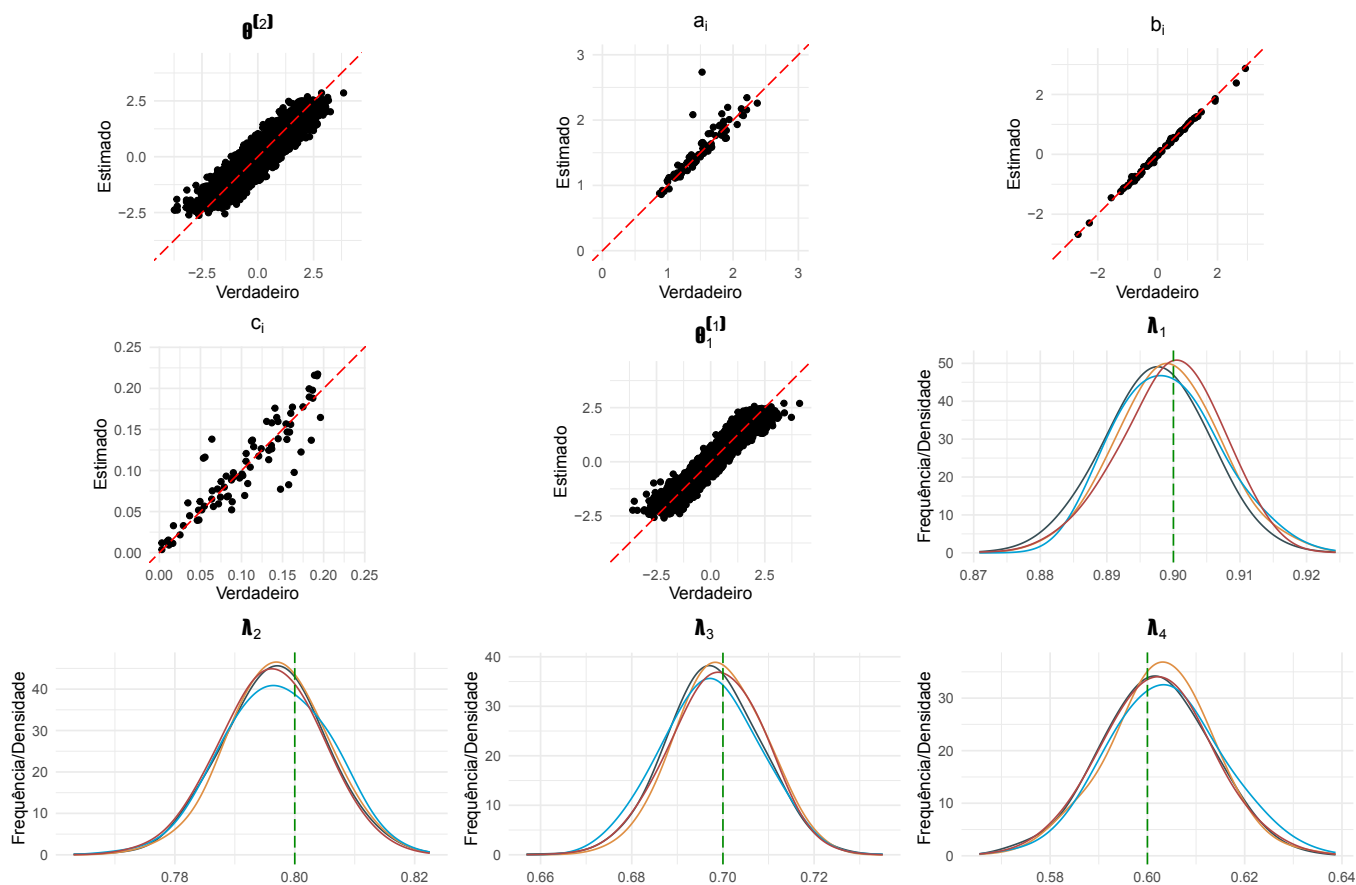
Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

A Figura 8 exibe os resultados para o cenário em que os dados foram gerados de um modelo com hierarquia entre a habilidade geral e as sub-habilidades, porém sem o agrupamento de alunos em escolas e sem o uso de covariáveis. Devido à presença de hierarquia, os valores estimados para $\theta^{(2)}$ e para cada $\theta^{(1)}$ são diferentes. Observe que há maior variância na estimação de $\theta^{(2)}$ se comparada à do cenário 1 (Figura 7).

As distribuições marginais a posteriori de λ indicam estimação adequada dos parâmetros de hierarquia, o que se confirma na Tabela 8. Além disso, a Tabela 8 indica também estimação precisa dos parâmetros multinível (τ^2 e γ).

Figura 8 – Cenário 2 - Hierárquico



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Tabela 8 – Resultados do Cenário 2 - Hierárquico

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	-1.66	-0.68 [-1.33, -0.03]	0.39	1.00	2572.30	2775.39
$\theta_{1,9}^{(1)}$	-0.35	-0.80 [-1.24, -0.37]	0.27	1.00	4168.08	2994.52
a_{11}	0.87	0.88 [0.77, 0.99]	0.07	1.00	3230.56	2973.94
b_{32}	0.59	0.53 [0.48, 0.57]	0.03	1.00	1874.49	2923.82
c_{57}	0.16	0.10 [0.03, 0.16]	0.04	1.00	987.13	1939.52
γ_1	0.00	0.01 [-0.01, 0.04]	0.02	1.00	1406.86	2045.79
γ_2	0.00	0.02 [-0.01, 0.05]	0.02	1.00	780.15	1488.54
λ_1	0.90	0.90 [0.89, 0.91]	0.01	1.01	220.48	344.79
λ_2	0.80	0.80 [0.78, 0.81]	0.01	1.01	627.03	1376.76
λ_3	0.70	0.70 [0.68, 0.71]	0.01	1.00	1000.06	1847.04
λ_4	0.60	0.60 [0.58, 0.62]	0.01	1.00	962.18	1755.61
τ^2	0.00	0.00 [0.00, 0.00]	0.00	1.07	77.18	145.19

Fonte: Elaboração própria.

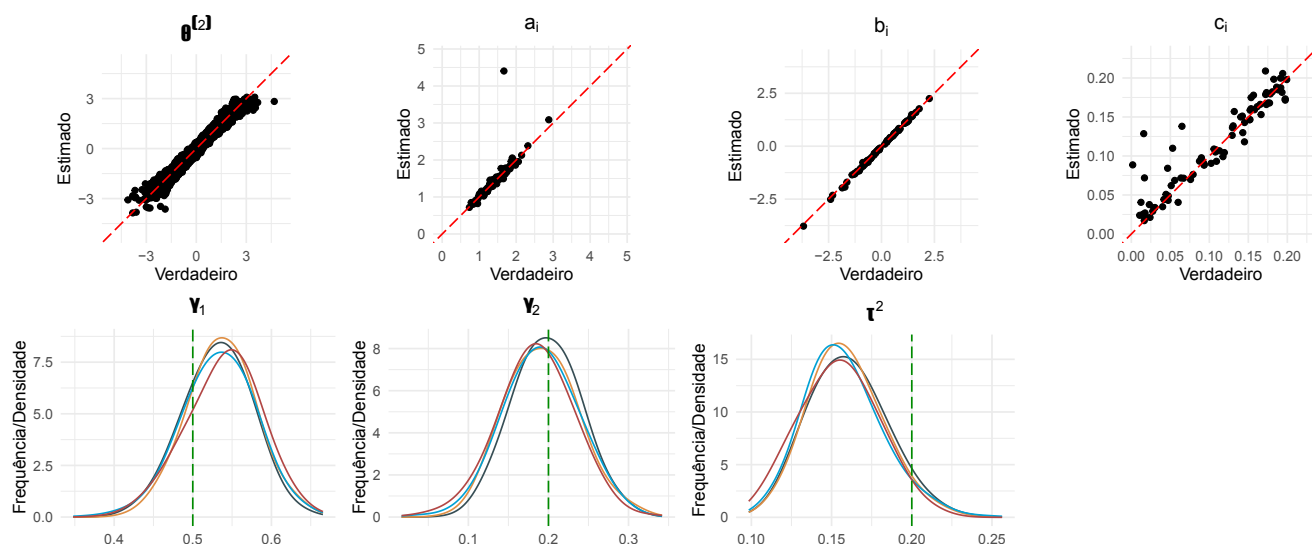
Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

No cenário 3, os dados também foram gerados de um modelo unidimensional, porém com o agrupamento de alunos em escolas e o uso de covariáveis. A Figura 9 exhibe graficamente os resultados, indicando ausência de viés e baixa variância na estimação dos parâmetros de habilidade e dos itens, além da estimação precisa dos coeficientes da regressão (γ). Observe que, tal como no cenário 1, a variância da estimação de $\theta^{(2)}$ é baixa se comparada à dos cenários com fator hierárquico (Figuras 8 e 10).

A Tabela 9 indica estimação precisa dos parâmetros de hierarquia, apesar das estatísticas de diagnóstico pouco favoráveis. Isso acontece devido à precisão da escala em que estão sendo estimados os valores. Os parâmetros da hierarquia assumem valores entre 0 e 1, correspondendo, efetivamente à correlação entre a habilidade geral e cada subhabilidade. Como pode se observar nos traceplots do Apêndice A, os valores amostrados de λ variam na terceira casa decimal, entre 0.996 e 1.000. Dessa forma, não há baixa concordância entre cadeias, apesar dos \hat{R} altos, nem baixo tamanho efetivo das amostras, visto que todos os valores amostrados estão próximos do valor verdadeiro.

Figura 9 – Cenário 3 - Multinível



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Tabela 9 – Resultados do Cenário 3 - Multinível

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	0.21	0.17 [-0.09, 0.43]	0.15	1.00	925.62	2658.19
$\theta_{1,9}^{(1)}$	-0.41	-0.33 [-0.59, -0.07]	0.16	1.00	1338.81	2296.33
a_{11}	1.43	1.37 [1.27, 1.48]	0.06	1.00	842.63	2052.10
b_{32}	0.52	0.46 [0.39, 0.54]	0.05	1.04	65.62	81.41
c_{57}	0.19	0.18 [0.16, 0.21]	0.02	1.00	2538.44	2576.15
γ_1	0.50	0.54 [0.46, 0.61]	0.04	1.01	295.06	415.75
γ_2	0.20	0.19 [0.12, 0.26]	0.04	1.01	497.63	691.83
λ_1	1.00	1.00 [1.00, 1.00]	0.00	1.99	5.46	11.41
λ_2	1.00	1.00 [1.00, 1.00]	0.00	2.98	4.56	11.40
λ_3	1.00	1.00 [1.00, 1.00]	0.00	2.45	4.89	11.20
λ_4	1.00	1.00 [0.99, 1.00]	0.00	1.59	6.81	27.60
τ^2	0.20	0.16 [0.12, 0.20]	0.02	1.01	162.51	351.37

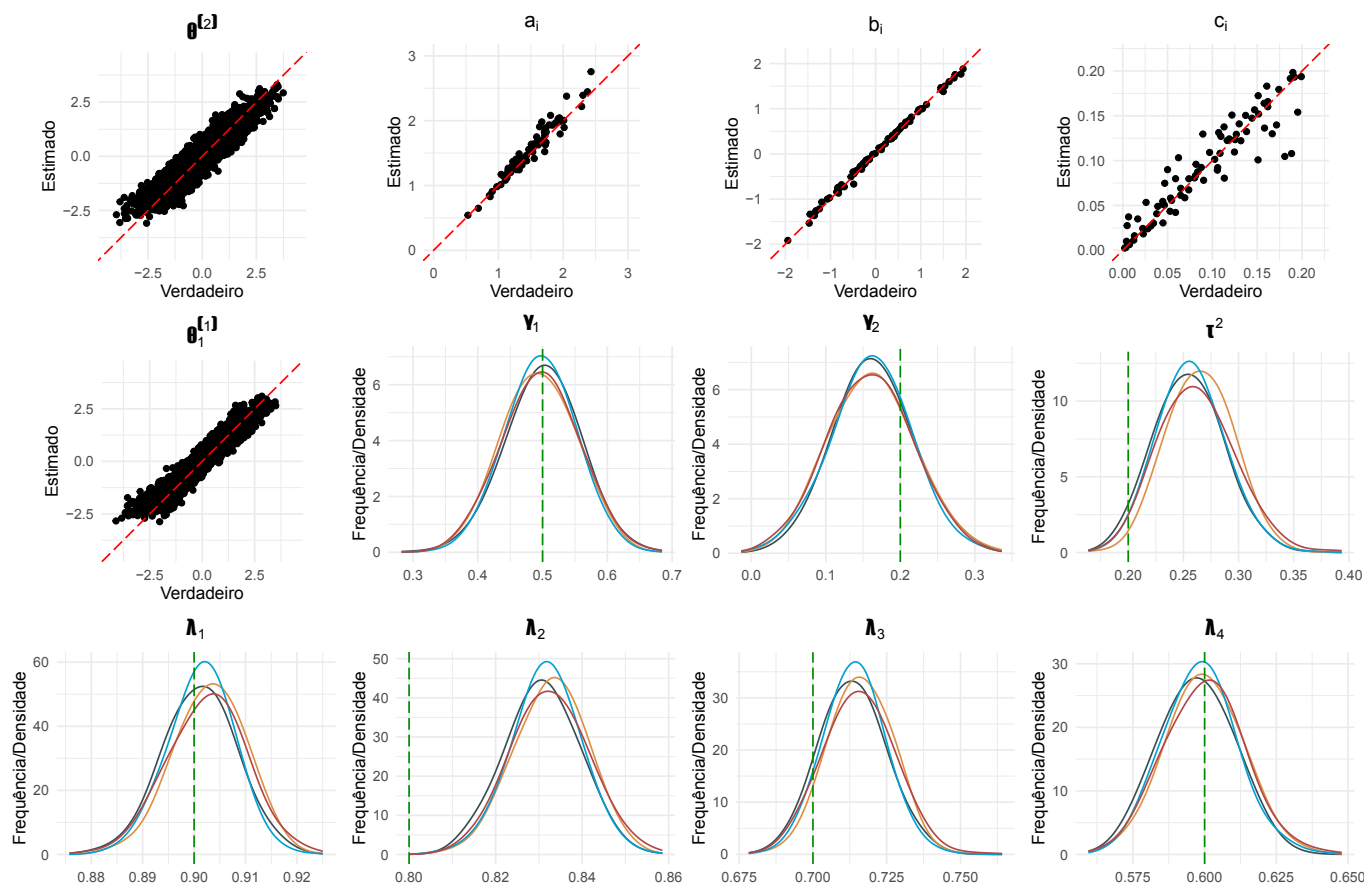
Fonte: Elaboração própria.

Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

No cenário 4, os dados foram gerados do modelo completo, com hierarquia entre os traços latentes e o com o fator multinível. Observe que, em virtude da presença de hierarquia, os valores estimados para $\theta^{(2)}$ e para cada $\theta^{(1)}$ são diferentes. Destaca-se a estimação pontual e intervalar apropriada dos parâmetros de hierarquia (λ), cujos valores estão reportados na Tabela 10.

Figura 10 – Cenário 4 - Completo



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Os gráficos com os traceplots em cada cenário estão disponíveis no Apêndice A.

Tabela 10 – Resultados do Cenário 4 - Completo

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	1.01	0.77 [0.15, 1.41]	0.38	1.00	4183.16	2593.43
$\theta_{1,9}^{(1)}$	2.84	3.10 [2.23, 4.14]	0.59	1.00	3370.67	2593.23
a_{11}	1.19	1.20 [1.04, 1.37]	0.10	1.00	2183.00	3177.69
b_{32}	-1.07	-1.00 [-1.15, -0.86]	0.09	1.00	1594.93	2639.90
c_{57}	0.13	0.12 [0.11, 0.14]	0.01	1.00	3756.04	3246.80
γ_1	0.50	0.50 [0.41, 0.59]	0.05	1.00	2797.66	2573.09
γ_2	0.20	0.16 [0.07, 0.25]	0.05	1.00	3562.46	2495.44
λ_1	0.90	0.90 [0.89, 0.91]	0.01	1.01	294.45	881.90
λ_2	0.80	0.83 [0.82, 0.85]	0.01	1.01	343.54	1247.79
λ_3	0.70	0.71 [0.70, 0.73]	0.01	1.01	334.85	978.39
λ_4	0.60	0.60 [0.58, 0.62]	0.01	1.01	327.28	1182.00
τ^2	0.20	0.26 [0.21, 0.31]	0.03	1.01	205.55	642.16

Fonte: Elaboração própria.

Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

6.2 Variações do modelo

Nesta seção avaliaremos variações do modelo, com a presença ou ausência de hierarquia, multidimensionalidade, covariáveis e agrupamento de alunos em escolas no *modelo*. Consideraremos os cenários na Tabela 11.

Nos dados verdadeiros desta análise, todos esses fatores estão presentes. Os dados correspondem ao Cenário 4 (completo) da análise do processo gerador de dados (Tabela 5).

Tabela 11 – Variações do modelo para simulação

Variação do Modelo	Hierarquia	Multinível	Covariáveis
1. Nulo	Não	Não	Não
2. Hierárquico	Sim	Não	Não
3. Multinível	Não	Sim	Sim
4. Completo	Sim	Sim	Sim

Fonte: Elaboração própria

6.2.1 Resultados das variações do modelo

A Tabela 12 apresenta a média de Erro Quadrático Médio (EQM), Viés e Variância na estimação das variações dos modelos, aplicados ao conjunto de dados gerado de um processo com hierarquia nos traços latentes fatores multinível (agrupamento de alunos em escolas e presença de covariáveis). Em cada variação do modelo, foram obtidas 1.000 amostras coletadas após *warm-up* de 1.000 iterações em 4 cadeias de Markov inicializadas aleatoriamente com valores distintos.

Os resultados na Tabela 12 mostram que a Variação 4 (modelo completo) apresenta os menores valores de Erro Quadrático Médio para todos os parâmetros, o que indica que a inclusão do fatores hierárquico e multinível no modelo estimado, quando os dados verdadeiros foram de fato gerados de um processo com a presença de tais fatores, aprimora a estimação dos parâmetros.

As seguintes observações podem ser feitas a partir da Tabela 12:

- O modelo completo apresenta o menor EQM para $\theta^{(2)}$, a , b , c e $\theta^{(1)}$;

- A inclusão do fator hierárquico reduz o viés de estimação de a e de c , reduz a variância dos estimadores de b e c , e aumenta a variância dos estimadores de $\theta^{(2)}$.

Tabela 12 – Média do Erro Quadrático Médio, Viés e Variância da estimação em cada modelo

Medida	Variação do modelo	$\theta^{(2)}$	a	b	c	$\theta^{(1)}$
Erro Quadrático Médio	1 - Nulo	0.202	0.280	0.032	0.0015	-
	2 - Hierárquico	0.185	0.030	0.007	0.0005	[0.109, 0.118, 0.104, 0.109]
	3 - Multinível	0.191	0.346	0.052	0.0014	-
	4 - Completo	0.161	0.012	0.003	0.0004	[0.099, 0.110, 0.101, 0.109]
Viés	1 - Nulo	0.004	-0.437	-0.001	0.0023	-
	2 - Hierárquico	0.006	0.127	-0.005	0.0001	[0.004, 0.001, 0.001, 0.022]
	3 - Multinível	0.023	-0.508	0.022	0.0026	-
	4 - Completo	0.022	0.051	0.013	0.0007	[0.019, 0.014, 0.012, 0.032]
Variância	1 - Nulo	0.045	0.003	0.006	0.0010	-
	2 - Hierárquico	0.134	0.010	0.002	0.0007	[0.080, 0.092, 0.092, 0.097]
	3 - Multinível	0.053	0.003	0.009	0.0010	-
	4 - Completo	0.150	0.010	0.004	0.0007	[0.093, 0.104, 0.102, 0.104]

Fonte: Elaboração própria.

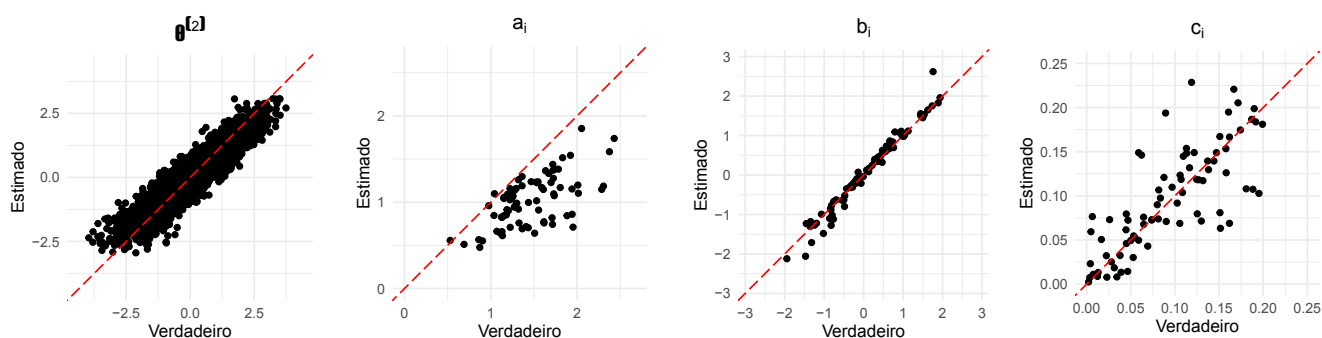
Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

As Figuras 11 a 14 e as Tabelas 13 a 16 apresentam os resultados para cada uma das variações do modelo.

As Tabelas 13 a 16 apresentam o valor verdadeiro, o valor estimado (média a posteriori e intervalo de credibilidade a posteriori de 95%), desvio padrão da distribuição marginal a posteriori e as estatísticas de diagnóstico \hat{R} e tamanhos efetivos da amostra no núcleo e nas caudas da distribuição (ESS Bulk e ESS Tail, respectivamente). No caso dos parâmetros $\theta^{(2)}$, $\theta^{(1)}$, a , b e c , são exibidos somente um exemplar de cada, visto que somam, ao todo, milhares de parâmetros, um para cada aluno no caso dos θ , e um para cada item do teste no caso de a , b e c .

A Figura 11 apresenta, graficamente, os resultados da variação correspondente ao modelo unidimensional, sem o agrupamento de alunos em escolas e sem a inclusão de covariáveis. Observa-se vício sobretudo na estimação dos parâmetros de discriminação dos itens (a).

Figura 11 – Variação 1 - Modelo Nulo



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Tabela 13 – Resultados da Variação 1 - Modelo Nulo

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	1.01	0.61 [0.27, 0.95]	0.18	1.00	5450.71	2979.75
a_{11}	1.19	0.91 [0.77, 1.07]	0.08	1.00	2244.48	2364.88
b_{32}	-1.07	-0.99 [-1.15, -0.88]	0.07	1.00	982.17	2025.32
c_{57}	0.13	0.12 [0.09, 0.14]	0.01	1.00	2401.54	2676.54

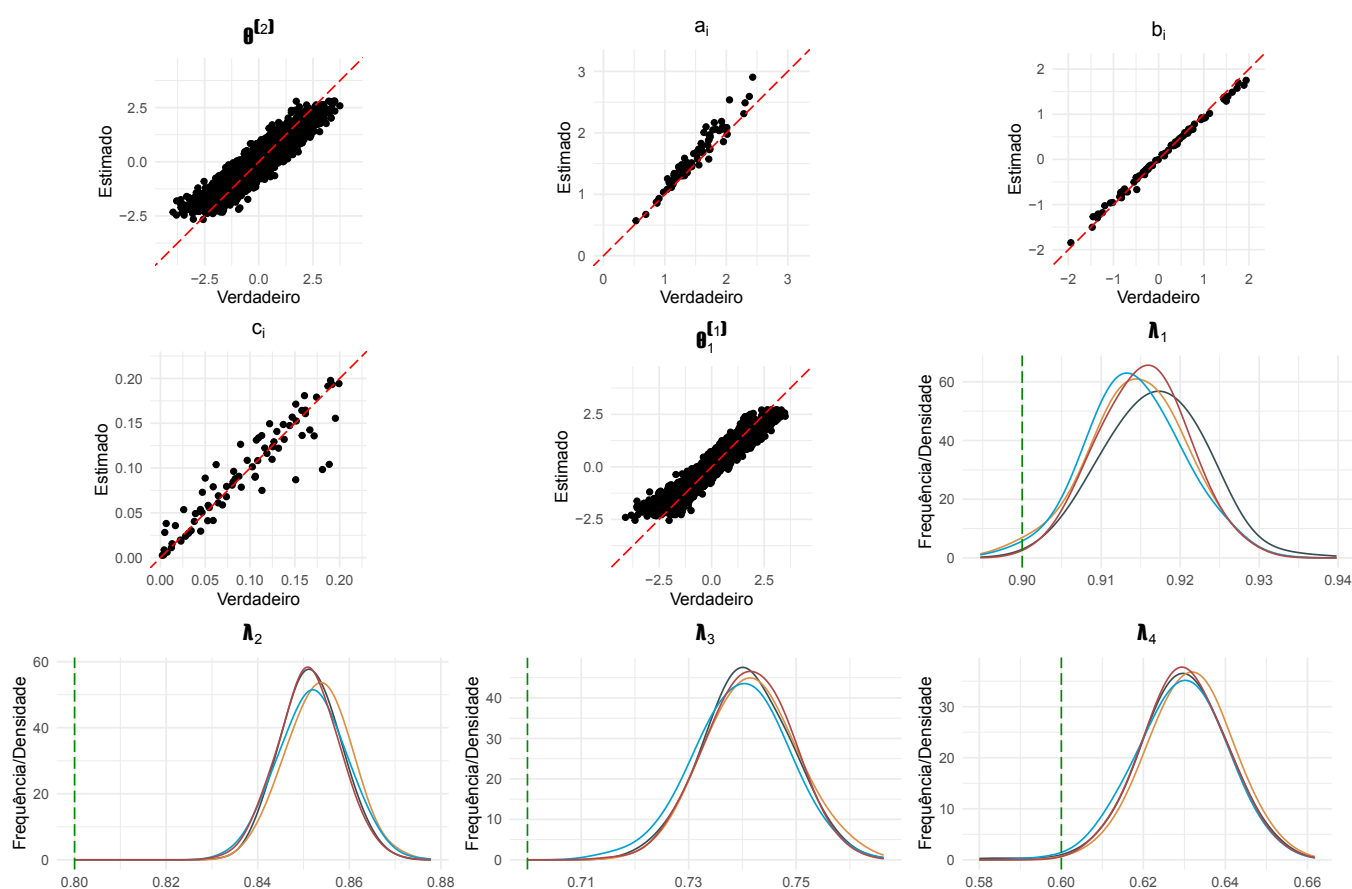
Fonte: Elaboração própria.

Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

Na segunda variação do modelo temos hierarquia entre a habilidade geral e as sub-habilidades, mas ausência de agrupamento de alunos em escolas e dos parâmetros multinível. A Figura 12 mostra que a presença do fator hierárquico reduz o vício na estimação do parâmetro a e a variância nos valores estimados para c . Porém, a ausência dos parâmetros multinível induzem vício, apesar de baixo, na estimação de λ , o que pode ser visto também na Tabela 14.

Figura 12 – Variação 2 - Modelo Hierárquico



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Tabela 14 – Resultados da Variação 2 - Modelo Hierárquico

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	1.01	0.64 [-0.06, 1.32]	0.35	1.00	1689.62	2204.33
$\theta_{1,109}^{(1)}$	2.84	2.72 [1.82, 3.86]	0.52	1.00	6305.89	2846.84
a_{11}	1.19	1.27 [1.08, 1.47]	0.10	1.00	2488.97	2823.46
b_{32}	-1.07	-0.96 [-1.10, -0.84]	0.07	1.01	831.10	1440.69
c_{57}	0.13	0.12 [0.11, 0.14]	0.01	1.00	3315.51	3164.17
λ_1	0.90	0.92 [0.90, 0.93]	0.01	1.02	157.30	294.59
λ_2	0.80	0.85 [0.84, 0.87]	0.01	1.01	346.92	788.15
λ_3	0.70	0.74 [0.72, 0.76]	0.01	1.01	568.11	1228.27
λ_4	0.60	0.63 [0.61, 0.65]	0.01	1.01	510.93	1102.34

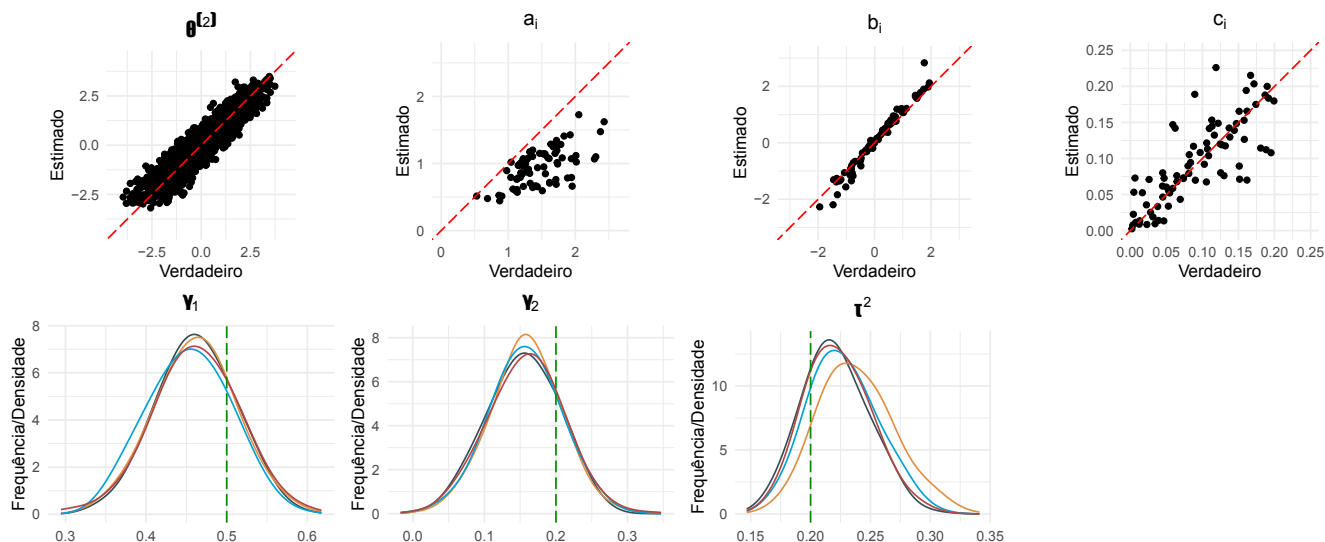
Fonte: Elaboração própria.

Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

Na Figura 13 observamos que mesmo no modelo unidimensional com agrupamento de alunos em escolas e presença de covariáveis, a ausência do fator hierárquico causa vício na estimação do parâmetro a , tal como ocorre no modelo unidimensional (Variação 1).

Figura 13 – Variação 3 - Modelo Multinível



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Além disso, a Tabela 15 indica estimação pontual e intervalar razoável para os parâmetros multinível (τ^2 e γ).

Tabela 15 – Resultados da Variação 3 - Modelo Multinível

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	1.01	0.69 [0.30, 1.07]	0.20	1.00	1424.82	2402.65
a_{11}	1.19	0.86 [0.73, 1.01]	0.07	1.00	1113.49	1707.12
b_{32}	-1.07	-1.06 [-1.26, -0.88]	0.10	1.02	187.95	483.37
c_{57}	0.13	0.12 [0.09, 0.14]	0.01	1.00	2500.16	2283.60
γ_1	0.50	0.46 [0.36, 0.55]	0.05	1.00	375.61	686.56
γ_2	0.20	0.16 [0.06, 0.25]	0.05	1.01	386.75	701.89
τ^2	0.20	0.23 [0.18, 0.29]	0.03	1.04	96.19	180.52

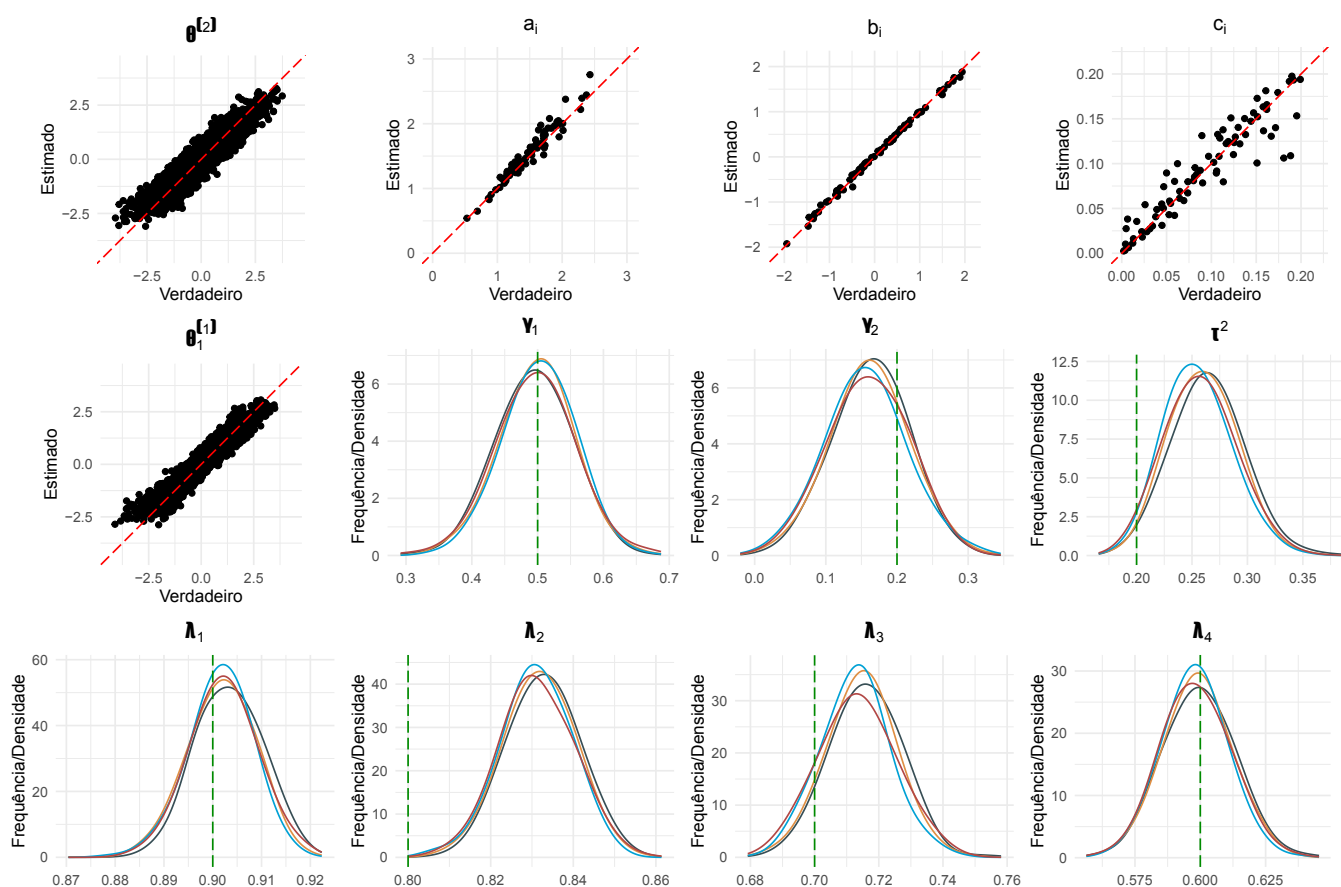
Fonte: Elaboração própria.

Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

A Variação 4 corresponde ao modelo proposto completo, com hierarquia nos traços latentes, agrupamento de alunos em escolas e presença de covariáveis. A Figura 14 mostra baixo viés e variância na estimação dos parâmetros de habilidade e dos itens (mesmo no caso de λ_2 , se considerarmos a escala desse parâmetro de correlação, que assume valores entre 0 e 1), estimação apropriada dos coeficientes da regressão (γ) e baixo vício na estimação de λ . A estimação de τ^2 é a que apresenta maior vício e variância. As estimações pontuais e intervalares constam na Tabela 16.

Figura 14 – Variação 4 - Modelo Completo



Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Os gráficos com os traceplots em cada variação estão disponíveis no Apêndice B.

Tabela 16 – Resultados da Variação 4 - Modelo Completo

Parâmetro	Valor verdadeiro	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	1.01	0.77 [0.04, 1.48]	0.37	1.00	2892.75	2512.43
$\theta_{1,9}^{(1)}$	2.84	3.07 [2.06, 4.28]	0.57	1.00	4863.97	3174.42
a_{11}	1.19	1.20 [1.02, 1.41]	0.10	1.00	2122.76	2884.57
b_{32}	-1.07	-1.00 [-1.18, -0.83]	0.09	1.01	374.07	1045.21
c_{57}	0.13	0.12 [0.10, 0.14]	0.01	1.00	4422.72	3421.59
γ_1	0.50	0.50 [0.39, 0.61]	0.06	1.00	1059.96	1829.03
γ_2	0.20	0.16 [0.05, 0.27]	0.05	1.00	1220.98	1572.86
λ_1	0.90	0.90 [0.89, 0.92]	0.01	1.01	226.28	772.81
λ_2	0.80	0.83 [0.81, 0.85]	0.01	1.01	190.38	586.89
λ_3	0.70	0.71 [0.69, 0.74]	0.01	1.02	194.79	556.80
λ_4	0.60	0.60 [0.57, 0.62]	0.01	1.01	232.76	702.41
τ^2	0.20	0.26 [0.20, 0.32]	0.03	1.02	147.38	487.51

Fonte: Elaboração própria.

Dados gerados de $J = 5.438$ alunos distribuídos em $s = 100$ escolas, estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

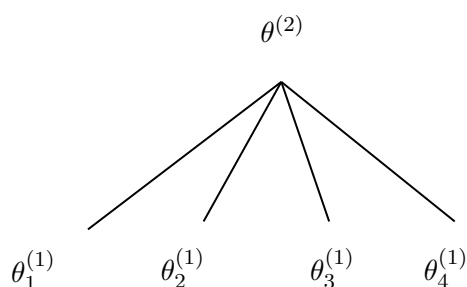
A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

7 Aplicação Empírica

Aplicamos o modelo proposto ao conjunto de dados de todas as escolas da rede estadual de Belo Horizonte (MG) cujos alunos realizaram a prova da 3ª série do ensino médio no SAEB 2019. A inferência será realizada considerando a estrutura hierárquica induzida pela Matriz de Referência.

Na prova de Matemática do SAEB, temos 91 itens organizados em quatro temas. Vale notar que, devido à utilização dos Blocos Balanceados Incompletos no desenho das provas, nenhum aluno responde a todos os 91 itens. De fato, conforme exposto na seção (2), cada aluno responde somente a 26 itens na prova da 3ª série do Ensino Médio em Matemática.

Figura 15 – Hierarquia proposta entre traços latentes - Matemática



Na hierarquia da Matriz de Referência, $\theta^{(2)}$ resume a proficiência do aluno em matemática, relacionada às habilidades em quatro temas: Espaço e Forma, Grandezas e Medidas, Números e Operações / Álgebra e Funções e Tratamento da Informação

Fonte: Elaboração própria

Tabela 17 – Traços latentes, temas e número de itens distintos por tema na prova de Matemática da 3ª série do ensino médio - SAEB 2019

Traço latente	Tema	Número distinto de itens
$\theta^{(2)}$	Matemática	91
$\theta_1^{(1)}$	Espaço e Forma	20
$\theta_2^{(1)}$	Grandezas e Medidas	9
$\theta_3^{(1)}$	Números e Operações / Álgebra e Funções	51
$\theta_4^{(1)}$	Tratamento da Informação	11

Fonte: Elaboração própria

A Tabela 18 indica o número de alunos que respondeu a cada quantidade de itens em cada tema da Matriz de Referência. Observa-se, em conformidade com a Tabela 17, que

a maioria dos alunos responde a uma quantidade baixa de itens nos temas *Grandezas e Medidas e Tratamento da Informação*. Por outro lado, todos os alunos responderam 12 itens ou mais no tema *Números e Operações / Álgebra e Funções*, que apresenta a maior quantidade de itens distintos no total.

Tabela 18 – Quantitativo de alunos por número de itens respondidos em cada tema de Matemática

Número de itens respondidos	Tema			
	Espaço e Forma	Grandezas e Medidas	Números e Operações / Álgebra e Funções	Tratamento da Informação
1	0	1208	0	0
2	0	2373	0	1197
3	408	3589	0	4794
4	1213	1208	0	2387
5	1985	0	0	0
6	1972	0	0	0
7	2397	0	0	0
8	403	0	0	0
12	0	0	1208	0
13	0	0	2391	0
14	0	0	400	0
15	0	0	1160	0
16	0	0	2003	0
17	0	0	808	0
19	0	0	408	0

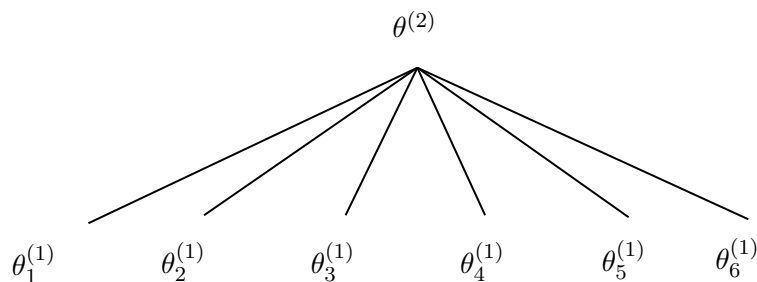
Quantitativo dentre o total de 8.378 alunos que realizaram a prova de matemática da 3ª série do ensino médio nas escolas estaduais de Belo Horizonte no SAEB 2019.

Fonte: Elaboração própria.

Já em Língua Portuguesa temos também 91 itens, porém agrupados em seis temas distintos. Nessa disciplina também há a utilização dos Blocos Balanceados Incompletos no desenho das provas. Assim, nem todos os alunos respondem todos os 91 itens. Tal como em Matemática, cada aluno responde somente a 26 itens na prova da 3ª série do Ensino Médio em Língua Portuguesa.

A Tabela 20 indica o número de alunos que respondeu a cada quantidade de itens em

Figura 16 – Hierarquia proposta entre traços latentes - Língua Portuguesa



Na hierarquia da Matriz de Referência, $\theta^{(2)}$ resume a habilidade geral do aluno em Língua Portuguesa, relacionada às habilidades em seis temas: Procedimentos de leitura, Implicações do suporte, compreensão do texto do gênero e/ou do enunciador na compreensão do texto, Relação entre textos, Coerência e coesão no processamento do texto, Relações entre recursos expressivos e efeitos de sentido e Variação linguística

Fonte: Elaboração própria

Tabela 19 – Traços latentes, temas e número de itens distintos por tema na prova de Língua Portuguesa da 3ª série do ensino médio - SAEB 2019

Traço latente	Tema	Número distinto de itens
$\theta^{(2)}$	Língua Portuguesa	91
$\theta_1^{(1)}$	Procedimentos de leitura	19
$\theta_2^{(1)}$	Implicações do suporte, compreensão do texto do gênero e/ou do enunciador na compreensão do texto	9
$\theta_3^{(1)}$	Relação entre textos	7
$\theta_4^{(1)}$	Coerência e coesão no processamento do texto	31
$\theta_5^{(1)}$	Relações entre recursos expressivos e efeitos de sentido	21
$\theta_6^{(1)}$	Variação linguística	4

Fonte: Elaboração própria

cada tema da Matriz de Referência de Língua Portuguesa. Observa-se que muitos alunos não responderam itens no tema *Variação linguística*. Por outro lado, todos os alunos responderam cinco itens ou mais no tema *Coerência e coesão no processamento do texto*, que apresenta a maior quantidade de itens distintos no total.

Tabela 20 – Quantitativo de alunos por número de itens respondidos em cada tema de Língua Portuguesa

Número de itens respondidos	Tema					
	Procedimentos de leitura	Implicações do suporte, compreensão do texto do gênero e/ou do enunciador na compreensão do texto	Relação entre textos	Coerência e coesão no processamento do texto	Relações entre recursos expressivos e efeitos de sentido	Variação linguística
0	0	377	0	0	0	3985
1	405	1597	1964	0	0	0
2	395	1989	4407	0	0	3999
3	1179	2405	2007	0	802	0
4	1209	1202	0	0	402	394
5	1209	808	0	410	1589	0
6	1207	0	0	389	3167	0
7	1203	0	0	1203	0	0
8	402	0	0	1606	2418	0
9	780	0	0	1620	0	0
10	389	0	0	1173	0	0
11	0	0	0	1577	0	0
12	0	0	0	400	0	0

Quantitativo dentre o total de 8.378 alunos que realizaram a prova de Língua Portuguesa da 3ª série do ensino médio nas escolas estaduais de Belo Horizonte no SAEB 2019.

Fonte: Elaboração própria.

7.1 Resultados do modelo proposto aplicado aos dados do SAEB

7.1.1 Resultados do modelo para Matemática

Em Matemática, a estimação do modelo proposto para os dados do SAEB produz evidência contrária à hipótese de unidimensionalidade, adotada na aplicação usual do modelo logístico de três parâmetros. A estimação pontual (média a posteriori) dos parâmetros $\hat{\lambda}_2 = 0.88$ e $\hat{\lambda}_4 = 0.91$ (temas *Grandezas e Medidas* e *Tratamento da Informação* da Matriz de Referência, respectivamente) estão distantes da unidade e os respectivos intervalos de credibilidade de 95% não incluem esse valor. Os demais temas apresentam estimação pontual e intervalar mais próximas da unidade, indicando maior correlação dentre essas habilidades e a habilidade geral $\theta^{(2)}$ (Tabela 21).

Um dos potenciais usos dos resultados multidimensionais é a identificação de temas ou habilidades que possam se beneficiar de reforço do aprendizado por aluno ou escola, aplicação que apresentamos como sugestão para trabalhos futuros.

Além disso, a estimação pontual e intervalar de τ^2 indica haver correlação entre respostas de alunos da mesma escola, em contrário à hipótese de independência das respostas usualmente assumida na estimação pelo modelo logístico de três parâmetros.

Outro resultado a destacar é a presença de efeito estimado do Nível Sócioeconômico Médio da escola (γ_1) sobre a habilidade geral $\theta^{(2)}$, em acordo com a literatura sobre desempenho educacional e contexto socioeconômico (ver, por exemplo, [Sirin, 2005](#)). Além disso, observamos efeito negativo, também esperado, do Índice de Complexidade da Gestão Escolar (γ_2), embora o intervalo de credibilidade de 95% inclua o zero no caso desse indicador.

Tabela 21 – Inferência com dados do SAEB - Matemática

Parâmetro	Valor INEP ⁴	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	-0.91	-0.89 [-1.97, 0.06]	0.51	1.00	6979.60	6477.95
a_{11}	-	0.69 [0.61, 0.78]	0.04	1.00	5911.46	6637.91
b_{32}	-	0.67 [0.44, 0.94]	0.13	1.00	908.30	2315.94
c_{57}	-	0.12 [0.11, 0.14]	0.01	1.00	6639.75	6648.37
$\theta_{1,109}^{(1)}$	-	1.52 [0.55, 2.43]	0.48	1.00	7664.56	6555.15
γ_1	-	0.31 [0.24, 0.37]	0.03	1.00	1800.73	2698.18
γ_2	-	-0.09 [-0.19, 0.02]	0.05	1.01	433.84	699.26
λ_1	-	0.97 [0.95, 0.99]	0.01	1.45	8.05	36.87
λ_2	-	0.88 [0.83, 0.94]	0.03	1.24	14.06	59.85
λ_3	-	0.99 [0.98, 1.00]	0.01	1.97	5.53	47.73
λ_4	-	0.91 [0.88, 0.93]	0.01	1.07	66.24	133.59
τ^2	-	0.07 [0.04, 0.09]	0.01	1.00	1917.72	3987.36

Fonte: Elaboração própria.

Valores estimados de $J = 8.378$ alunos da 3ª série do Ensino Médio das 75 escolas da rede estadual participantes do SAEB 2019 em Belo Horizonte.

Dados estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

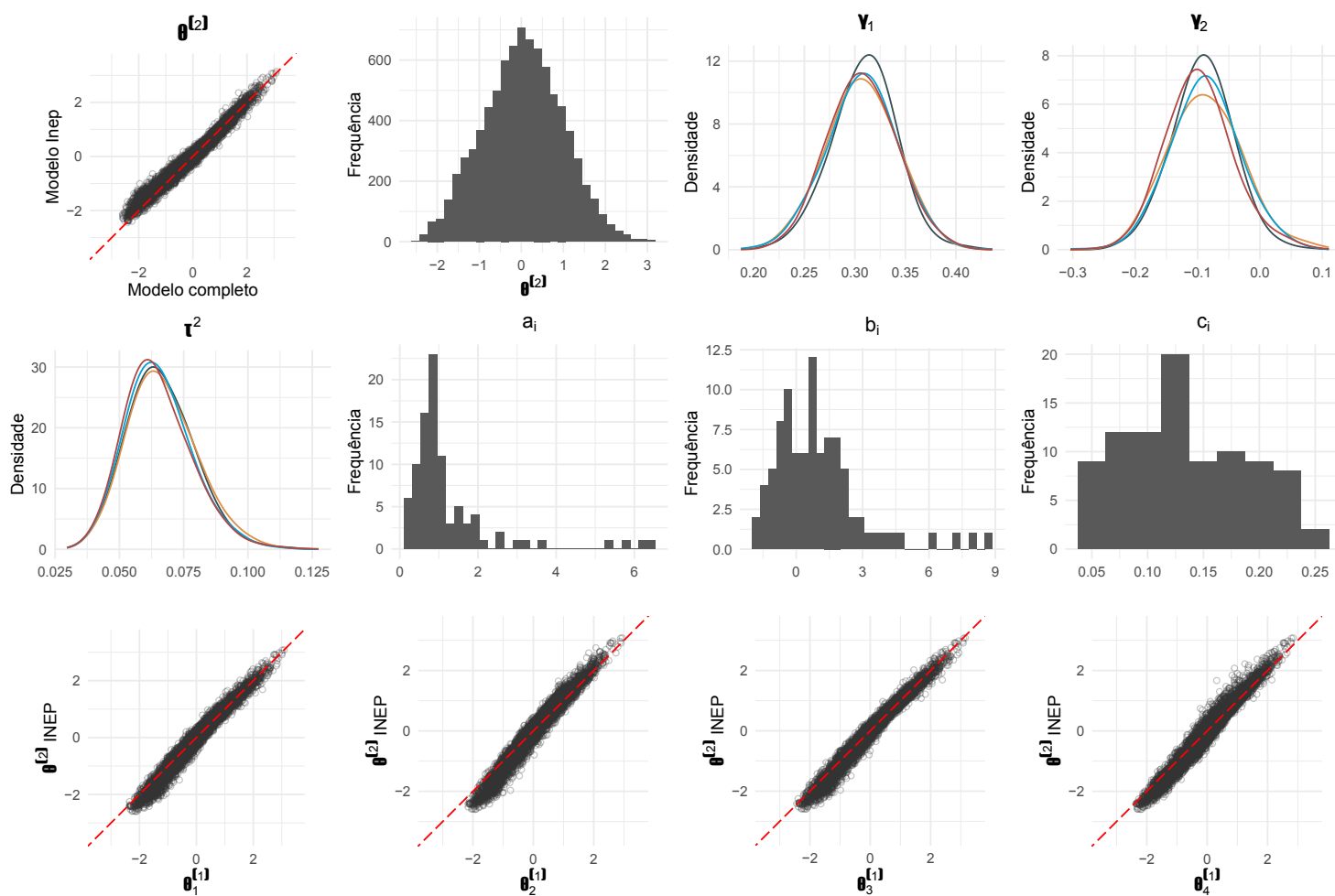
A Figura 17 exibe a correlação entre a proficiência estimada pelo INEP e a habilidade geral e temáticas estimadas no modelo proposto, além das distribuições marginais a posteriori de $\theta^{(2)}$, γ , τ^2 , a , b e c . Destaca-se que, para o tema *Grandezas e Medidas*, as maiores diferenças entre a proficiência estimada pelo INEP e os valores estimados no modelo proposto ocorrem para valores mais baixos de habilidade, isto é, os valores estimados de $\theta_1^{(1)}$, no gráfico de dispersão, tendem a se distanciar da reta de 45° nos valores mais baixos. Isso significa que alunos com habilidade baixa em *Grandezas e Medidas* tendem a obter desempenho um pouco melhor nos itens dos outros temas.

Já no caso do tema *Tratamento da Informação*, o outro tema cuja correlação com $\theta^{(2)}$ está mais distante da unidade, há também diferenças nos valores mais altos de habilidade, de modo que muitos alunos com bom desempenho em itens de outros temas apresentam habilidade um pouco mais baixa em *Tratamento da Informação*.

Vale notar ainda, que o modelo permite a decomposição do efeito da escola no desempenho dos alunos entre fatores conhecidos e desconhecidos. Na Figura 18, os alunos da escola 11, por exemplo, apresentam uma habilidade média superior aos efeitos conhecidos da escola (Nível Socioeconômico Médio e Complexidade da Gestão Escolar) conforme o subgráfico (b), pois o efeito de fatores não conhecidos é alto (subgráfico (c)). O inverso ocorre para os alunos da escola 66.

O traceplot das cadeias estimadas com os dados da prova de Matemática do SAEB constam no Apêndice C.

Figura 17 – Resultados do modelo com dados do SAEB - Matemática

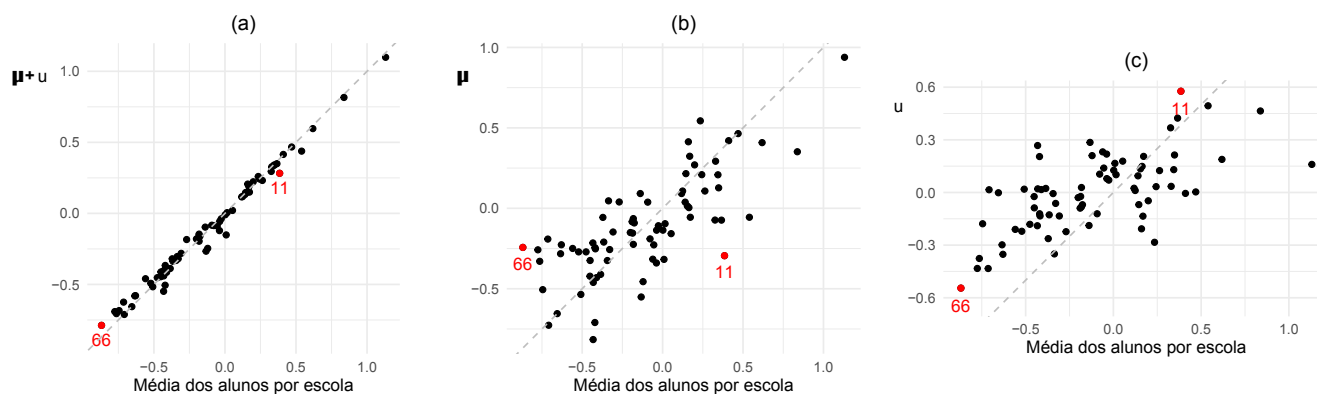


Valores estimados de $J = 8.378$ alunos da 3ª série do Ensino Médio das 75 escolas da rede estadual participantes do SAEB 2019 em Belo Horizonte.

Dados estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Figura 18 – Decomposição dos efeitos da escola - Prova de Matemática do SAEB



O modelo proposto permite decompor efeitos de fatores conhecidos e desconhecidos no desempenho de alunos de uma escola. Os alunos da escola 11, por exemplo, apresentam uma habilidade média superior aos efeitos conhecidos da escola (Nível Socioeconômico Médio e Complexidade da Gestão Escolar), conforme o subgráfico (b), pois o efeito de fatores não conhecidos é alto (subgráfico (c)). O inverso ocorre para os alunos da escola 66.

Fonte: Elaboração própria

7.1.2 Resultados do modelo para Língua Portuguesa

Em Língua Portuguesa, a estimação dos parâmetros de hierarquia está mais próxima da unidimensionalidade. Somente um dos seis parâmetros tem estimativa pontual um pouco mais distante da unidade ($\hat{\lambda}_6 = 0.93$).

Tal como em Matemática, a estimação pontual e intervalar de τ^2 indica haver correlação entre respostas de alunos da mesma escola, em contrário à hipótese de independência das respostas usualmente assumida na estimação pelo modelo logístico de três parâmetros.

Em relação às covariáveis, também há presença de efeito estimado do Nível Sócioeconômico Médio da escola (γ_1) sobre a habilidade geral $\theta^{(2)}$ em Língua Portuguesa. Além disso, a estimativa pontual do coeficiente do Índice de Complexidade da Gestão Escolar (γ_2) também é negativa, embora o intervalo de credibilidade de 95% inclua o zero, tal como em Matemática.

Tabela 22 – Inferência com dados do SAEB - Língua Portuguesa

Parâmetro	Valor INEP	Valor estimado	Desvio padrão	\hat{R}	ESS Bulk	ESS Tail
$\theta_{67}^{(2)}$	1.44	1.34 [0.50, 2.24]	0.44	1.00	4088.32	2673.46
a_{11}	-	0.56 [0.47, 0.69]	0.06	1.00	2679.50	2739.57
b_{32}	-	0.08 [-0.27, 0.34]	0.15	1.02	415.70	726.24
c_{57}	-	0.11 [0.01, 0.23]	0.06	1.00	1515.30	2282.94
$\theta_{1,109}^{(1)}$	-	-0.70 [-1.45, -0.02]	0.37	1.00	3546.47	2899.46
γ_1	-	0.23 [0.16, 0.29]	0.04	1.01	1037.77	1262.59
γ_2	-	-0.06 [-0.17, 0.05]	0.06	1.02	188.23	341.75
λ_1	-	0.99 [0.98, 0.99]	0.00	1.54	7.16	13.81
λ_2	-	0.98 [0.97, 0.99]	0.00	1.85	5.82	26.21
λ_3	-	0.97 [0.95, 0.99]	0.01	1.54	7.21	24.08
λ_4	-	0.99 [0.98, 1.00]	0.00	2.05	5.38	10.86
λ_5	-	0.99 [0.98, 0.99]	0.00	1.30	11.23	55.47
λ_6	-	0.93 [0.90, 0.97]	0.02	1.42	8.31	25.97
τ^2	-	0.08 [0.05, 0.11]	0.01	1.00	895.08	1713.36

Fonte: Elaboração própria.

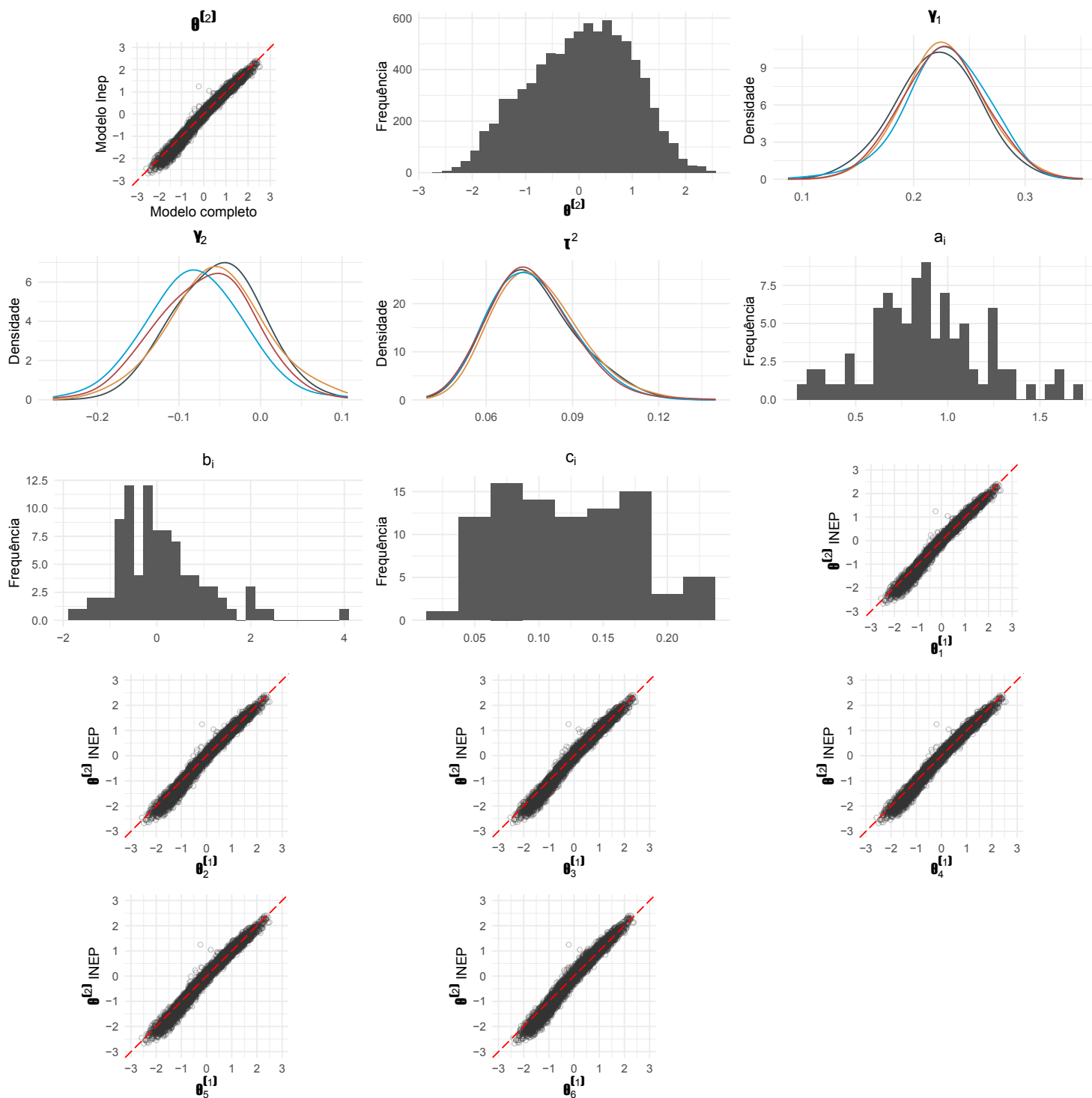
Valores estimados de $J = 8.378$ alunos da 3ª série do Ensino Médio das 75 escolas da rede estadual participantes do SAEB 2019 em Belo Horizonte.

Dados estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

A coluna “Valor estimado” indica a média e o intervalo de credibilidade de 95% (quantis 0.025 e 0.975, respectivamente).

A Figura 19 exibe a correlação entre a proficiência estimada pelo INEP e a habilidade geral e temáticas estimadas no modelo proposto, além das distribuições marginais a posteriori de $\theta^{(2)}$, γ , τ^2 , a , b e c . De modo geral, os gráficos de dispersão tendem a indicar unidimensionalidade no caso da prova de Língua Portuguesa.

Figura 19 – Resultados do modelo com dados do SAEB - Língua Portuguesa



Valores estimados de $J = 8.378$ alunos da 3ª série do Ensino Médio das 75 escolas da rede estadual participantes do SAEB 2019 em Belo Horizonte.

Dados estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

8 Considerações Finais

O cálculo da proficiência em Língua Portuguesa e em Matemática a partir de dados do SAEB, realizado pelo INEP, utiliza o modelo logístico de três parâmetros, que pressupõe unidimensionalidade e independência das respostas dos alunos entre itens e também das respostas dos outros alunos. A Matriz de Referência do SAEB, contudo, tem uma estrutura multidimensional, e, além disso, a literatura educacional evidencia a importância do contexto escolar no aprendizado, de modo que alunos da mesma escola ou do mesmo grupo tendem a responder itens de prova de maneira similar ou correlacionada.

Nesta dissertação, propomos um modelo multiunidimensional que reúne os itens nos temas a que pertencem na Matriz de Referência, e que agrupa alunos em escolas, introduzindo correlação das habilidades de alunos do mesmo grupo, além de permitir a inclusão de covariáveis da escola ou de alunos na estimação.

Realizamos uma série de simulações tanto para verificar a capacidade do modelo de estimar corretamente parâmetros gerados a partir de diferentes processos geradores, quanto para comparar seu desempenho a modelos sem presença de hierarquia nos traços latentes ou agrupamento de alunos em escolas e uso de covariáveis.

Aplicamos, ainda, o modelo proposto aos dados da prova de Matemática e de Língua Portuguesa realizada pelos alunos do 3ª série do Ensino Médio das escolas estaduais em Belo Horizonte (MG) no SAEB 2019. Obtemos evidência contrária à hipótese de unidimensionalidade no caso da prova de Matemática, e a favor dessa hipótese na prova de Língua Portuguesa.

Além disso, a estimação pontual e intervalar de τ^2 indicou haver correlação entre respostas de alunos da mesma escola em ambas as provas, em contrário à hipótese de independência das respostas usualmente assumida na estimação pelo modelo logístico de três parâmetros.

Também nas duas provas, identificamos, ainda, a presença de efeito de características observadas e não observadas da escola sobre o desempenho dos alunos. Os resultados do modelo multiunidimensional permitem, ainda, a identificação de temas ou habilidades específicas que possam se beneficiar de reforço do aprendizado, por aluno ou escola.

Como sugestão para trabalhos futuros, indicamos a extensão do modelo para três níveis hierárquicos, que não apenas reúna os itens em temas da Matriz de Referência mas em temas e habilidades. Dessa forma, seriam obtidos três níveis de proficiência: a proficiência geral do aluno em Matemática, por exemplo, sua proficiência nos grupos de itens relacionados (por exemplo, *Grandezas e Medidas* ou *Álgebra e Funções*), e sua proficiência em cada habilidade da Matriz (exemplo: *Resolver problema envolvendo o cálculo de perímetro de figuras planas*).

Outra forma de aplicação sugerida seria, também, reunir os alunos pela turma em oposição ao agrupamento somente por escola, ou, ainda, agrupá-los pela escola e depois pela turma, em dois níveis, tornando a aplicação mais próxima da realidade escolar.

Por fim, outra possível extensão do modelo consiste em flexibilizar a especificação da variância da habilidade dos alunos nas escolas. Nesta dissertação assumimos que $u_s \sim N(0, \tau^2)$ e $e_{js} \sim N(0, \sigma^2)$, isto é, o termo de erro no nível do aluno tem variância comum para os alunos em todas as escolas. Outra possível especificação seria $e_{js} \sim N(0, \sigma_s^2)$. Neste caso, os desvios referentes aos alunos têm variância específica em cada escola, e uma modelagem mais flexível poderia adotar, por exemplo, $\sigma_s^2 \sim \text{Gamma}(\nu/2, \nu/2)$, com ν conhecido.

Referências

- Andrade, D. F. d.; Tavares, H. R.; Valle, R. d. C. *Teoria da Resposta ao Item: Conceitos e Aplicações*. [S.l.: s.n.], 2000. Citado na página 12.
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In: LORD, F.; NOVICK, M. (Ed.). *Statistical Theories of Mental Test Scores*. [S.l.]: Addison-Wesley, 1968. p. 397–479. Citado na página 13.
- Da Silva, M. A. et al. Incorporating the Q-Matrix Into Multidimensional Item Response Theory Models. *Educational and Psychological Measurement*, v. 79, n. 4, p. 665–687, 2019. Citado 3 vezes nas páginas 13, 19 e 23.
- Fox, J.-P. *Bayesian Item Response Modeling*. [S.l.]: Springer New York, 2010. Citado 6 vezes nas páginas 13, 19, 22, 24, 25 e 28.
- Gabry, J.; Češnovar, R.; Johnson, A. *cmdstanr: R Interface to 'CmdStan'*. [S.l.], 2023. <https://mc-stan.org/cmdstanr/>, <https://discourse.mc-stan.org>. Citado na página 31.
- 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 na página 32.
- Hoffman, M. D.; Gelman, A. *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*. 2011. Citado na página 31.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Nota Explicativa Resultados Prova Brasil 2013*. 2014. Recuperado em 28 de fevereiro de 2021. Disponível em: <https://download.inep.gov.br/mailling/2014/nota_explicativa_prova_brasil_2013.pdf>. Citado na página 14.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Microdados do SAEB 2019*. 2019. Recuperado em 28 de fevereiro de 2021. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-ainformacao/dados-abertos/microdados/saeb>>. Citado 5 vezes nas páginas 15, 16, 17, 18 e 84.
- Klein, R. Utilização da teoria de resposta ao item no sistema nacional de avaliação da educação básica (saeb). *Revista Meta: Avaliação*, v. 1, n. 2, p. 125–140, 2003. Disponível em: <<https://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38>>. Citado na página 14.
- Martín, E. S. Identification of item response theory models. In: LINDEN, W. J. van der (Ed.). *Handbook of Item Response Theory, Statistical Tools, Volume 2*. [S.l.]: Taylor & Francis, 2021. v. 63, n. 3, cap. 8, p. 431–433. Citado na página 22.
- McCullagh, P. What is a statistical model? *The Annals of Statistics*, Institute of Mathematical Statistics, v. 30, n. 5, p. 1225–1267, 2002. ISSN 00905364. Disponível em: <<http://www.jstor.org/stable/1558705>>. Citado na página 21.
- NEAL, R. Mcmc using hamiltonian dynamics. In: _____. Chapman and Hall/CRC, 2011. Disponível em: <<http://dx.doi.org/10.1201/b10905-6>>. Citado na página 32.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2023. Disponível em: <<https://www.R-project.org/>>. Citado na página 31.

Reckase, M. *Multidimensional Item Response Theory*. [S.l.]: Springer New York, 2009. Citado 2 vezes nas páginas 12 e 19.

Sirin, S. R. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 75, n. 3, p. 417–453, 2005. Citado na página 61.

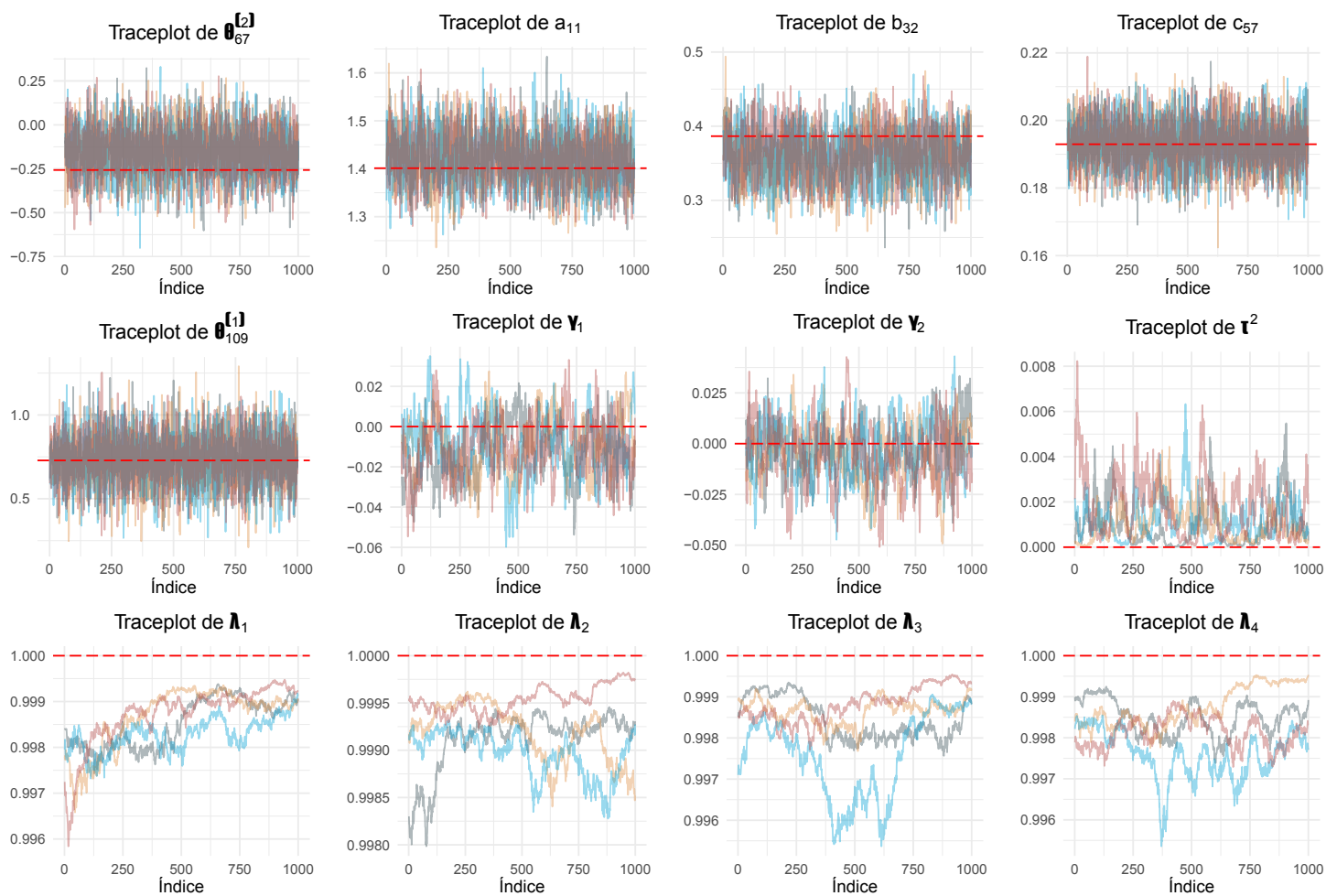
Stan Development Team. *Stan: A C++ Library for Probability and Sampling*. 2023. Version 2.31.0. Computer software. Disponível em: <<https://mc-stan.org/>>. Citado 2 vezes nas páginas 31 e 33.

Venturelli S. L., J.; Gonçalves, F. B.; Andrade, D. F. *Multidimensional Bayesian IRT Model for Hierarchical Latent Structures*. arXiv, 2020. Disponível em: <<https://arxiv.org/abs/2006.09966>>. Citado 7 vezes nas páginas 13, 19, 26, 27, 28, 29 e 30.

Apêndices

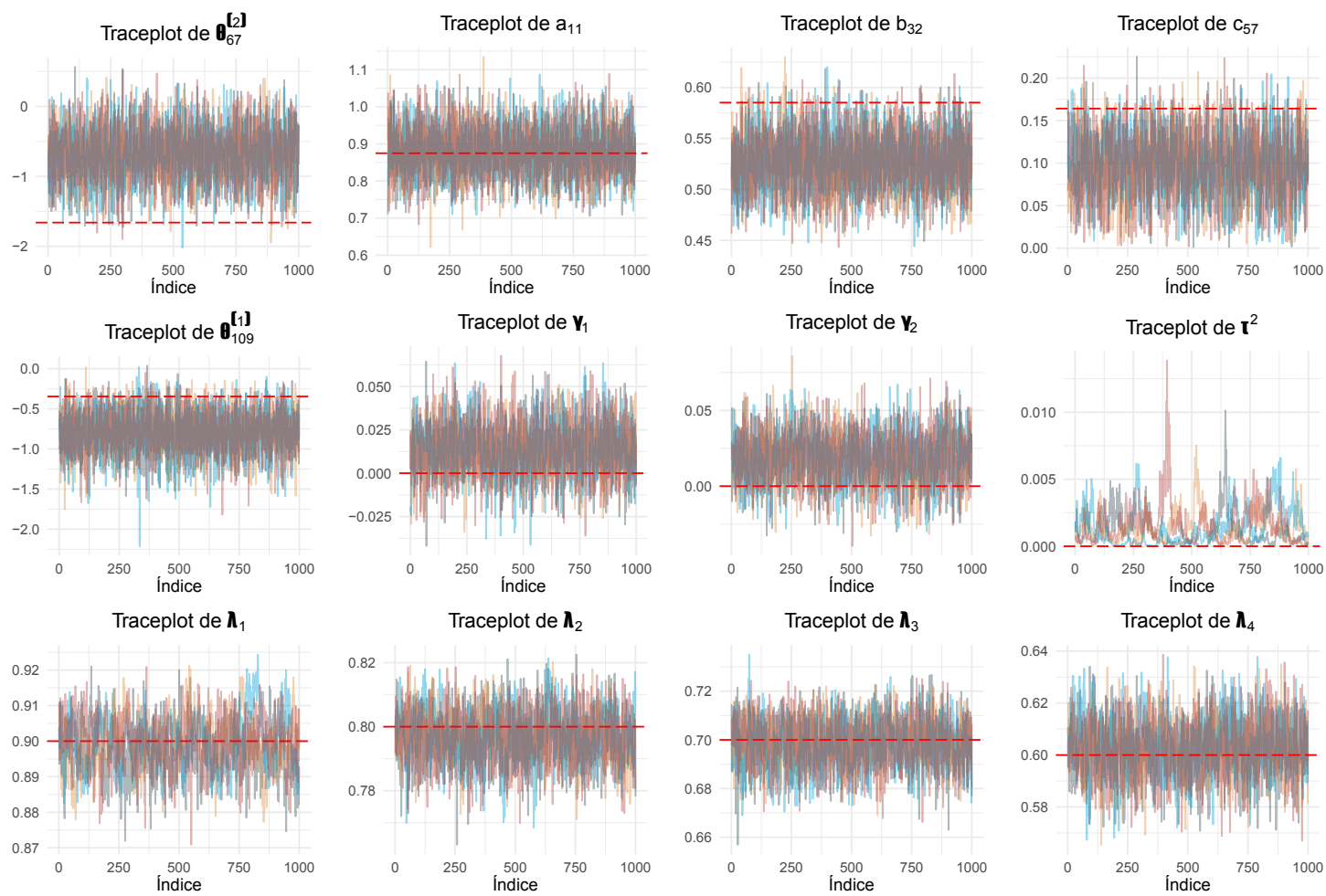
Apêndice A - Traceplots dos cenários para o processo gerador de dados

Figura 20 – Cenário 1 - Nulo



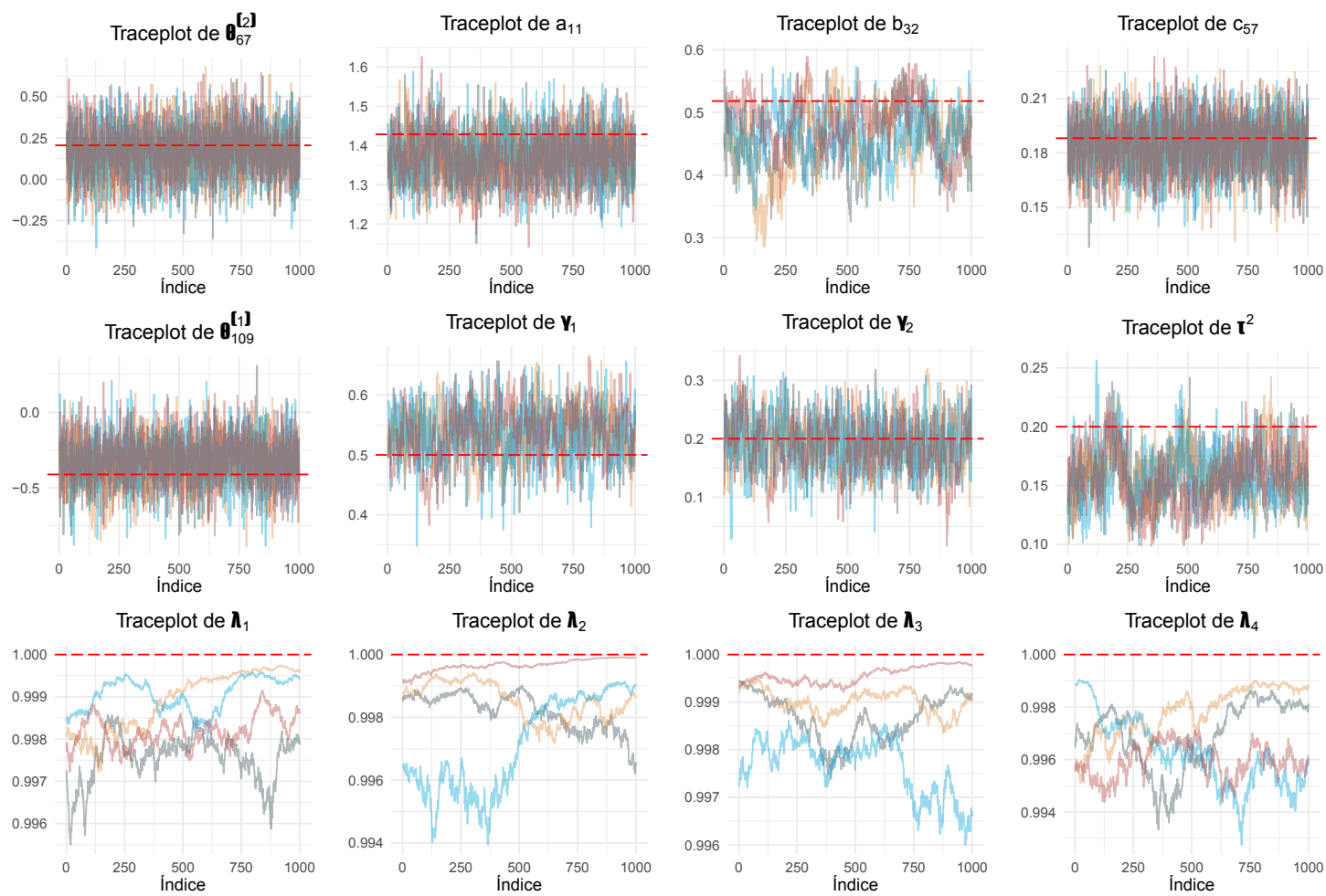
Fonte: Elaboração própria

Figura 21 – Cenário 2 - Hierárquico



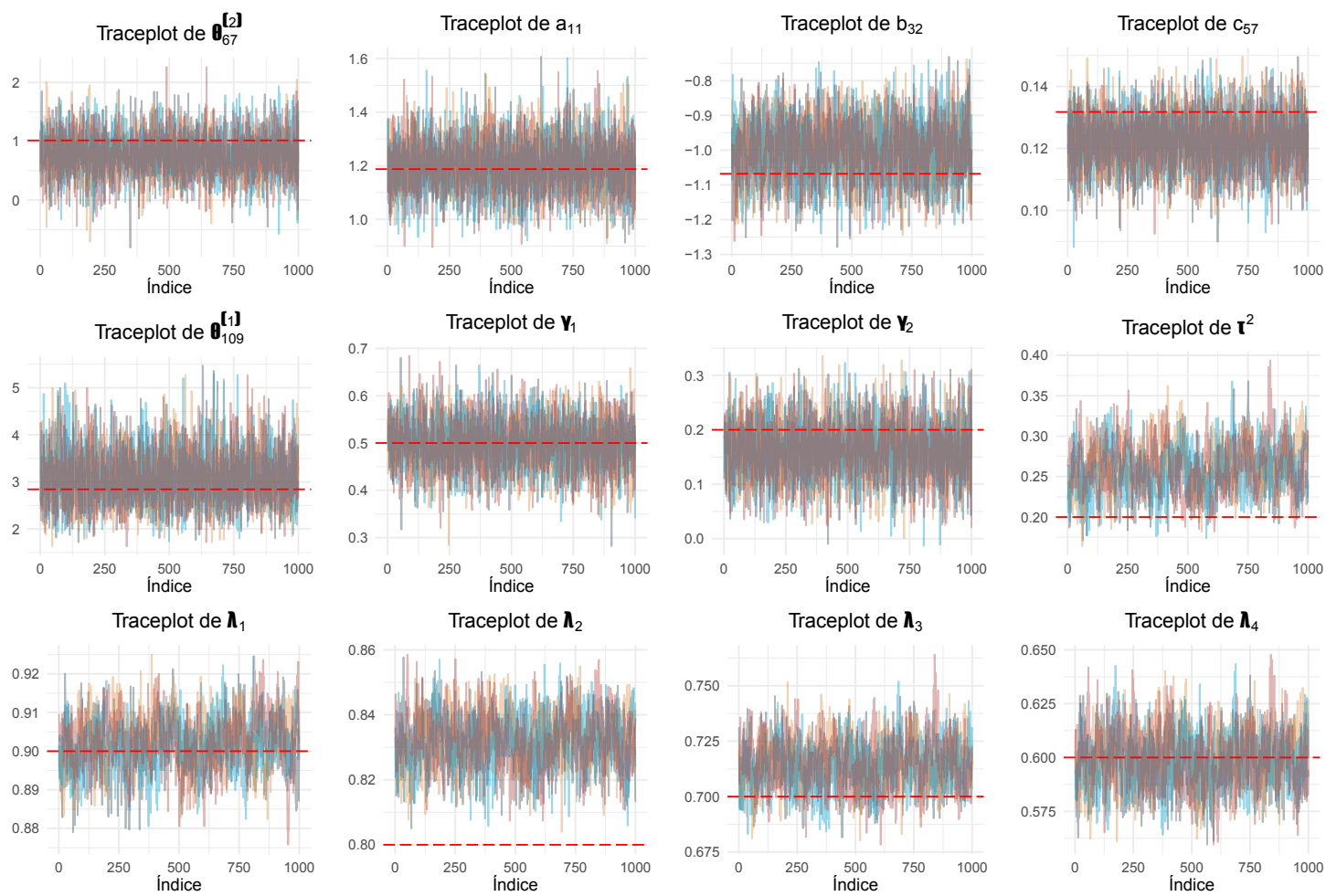
Fonte: Elaboração própria

Figura 22 – Cenário 3 - Multinível



Fonte: Elaboração própria

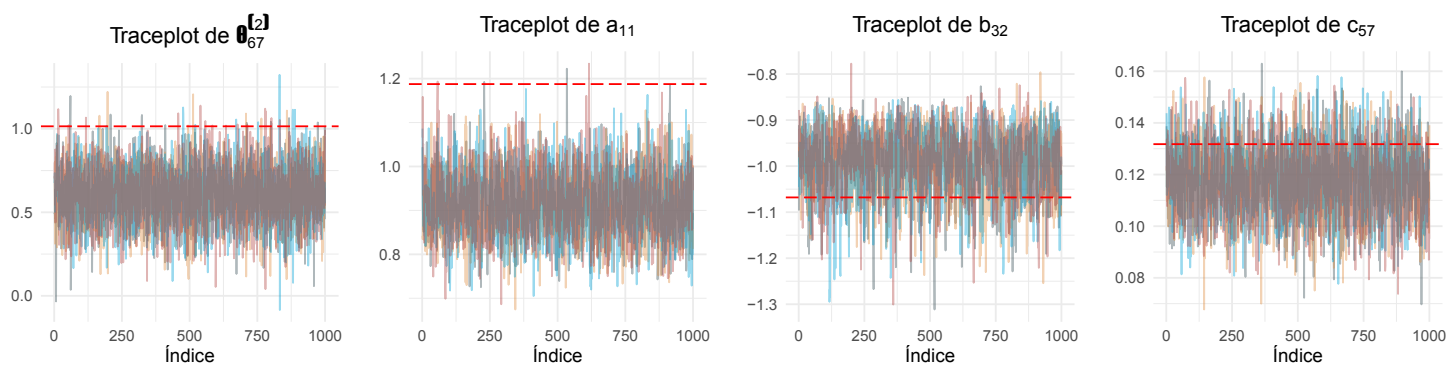
Figura 23 – Cenário 4 - Completo



Fonte: Elaboração própria

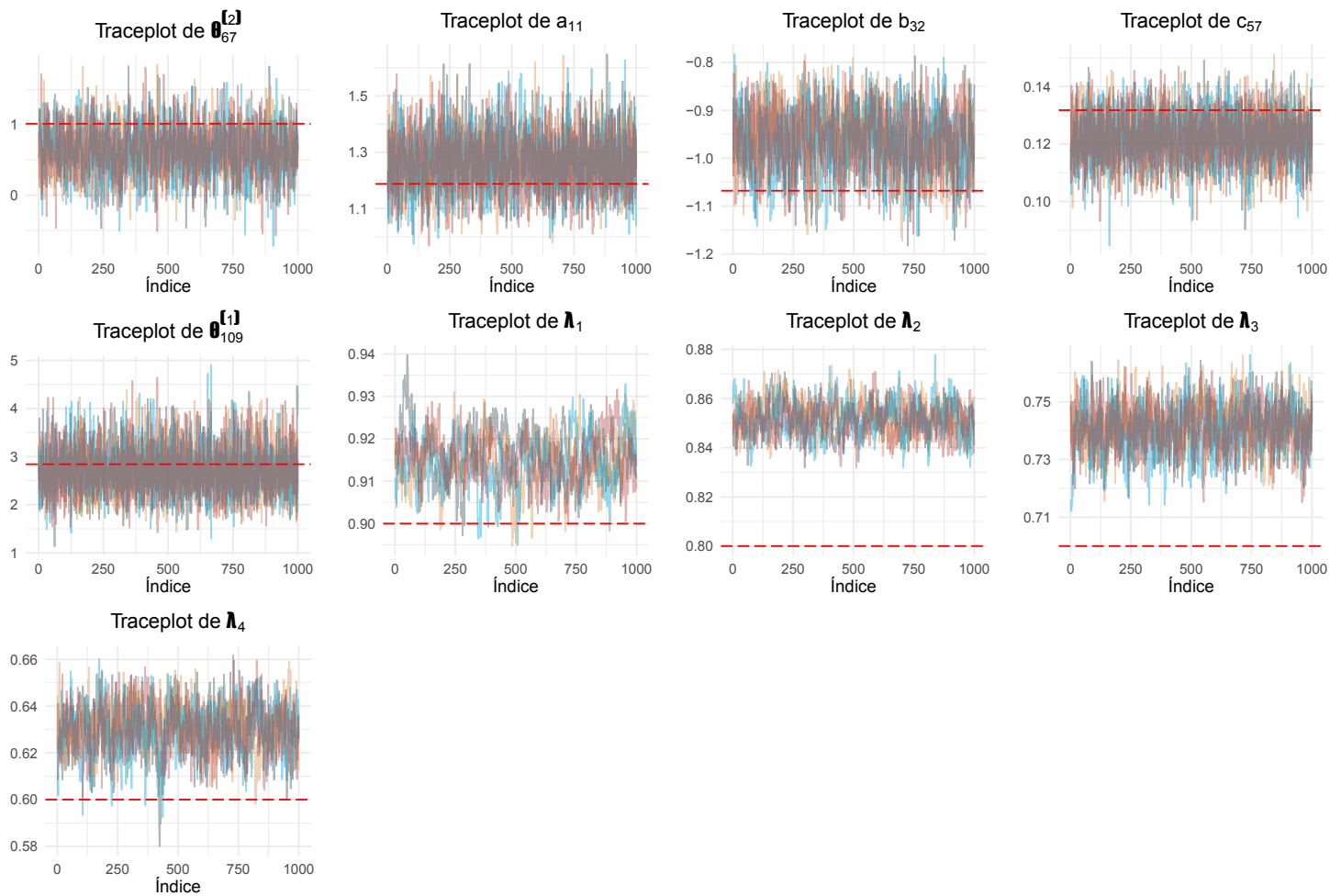
Apêndice B - Traceplots das variações do modelo

Figura 24 – Variação 1 - Modelo Nulo



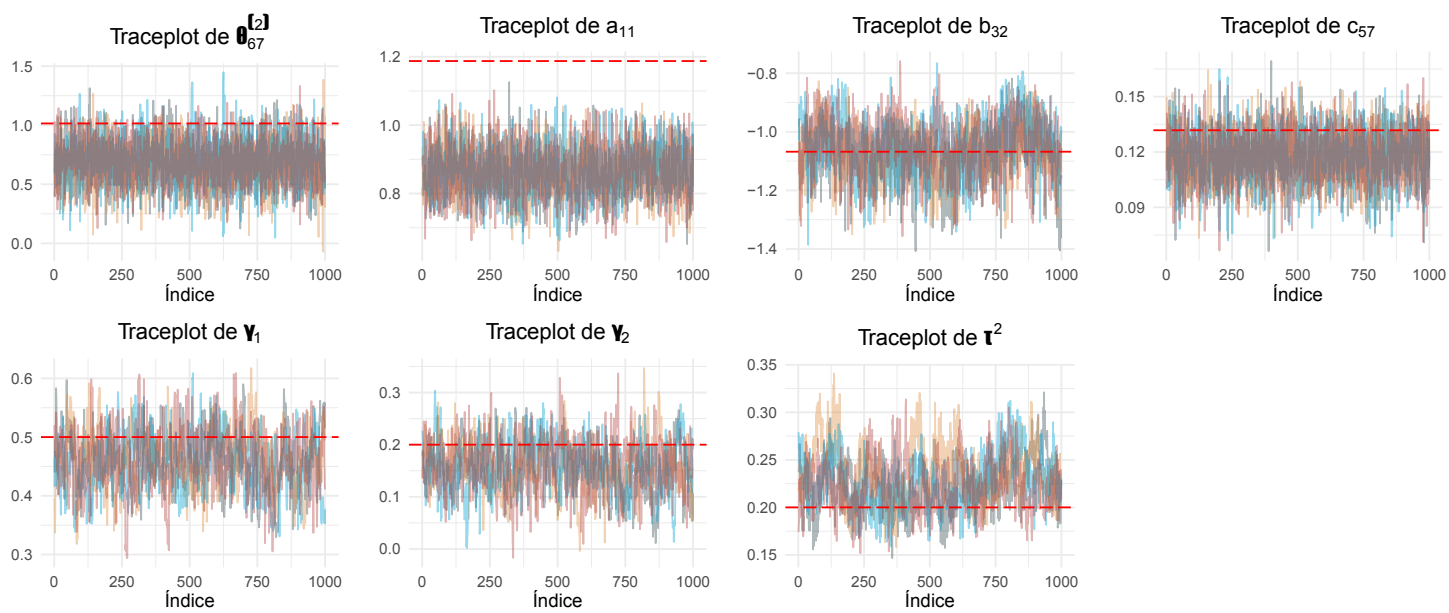
Fonte: Elaboração própria

Figura 25 – Variação 2 - Modelo Hierárquico



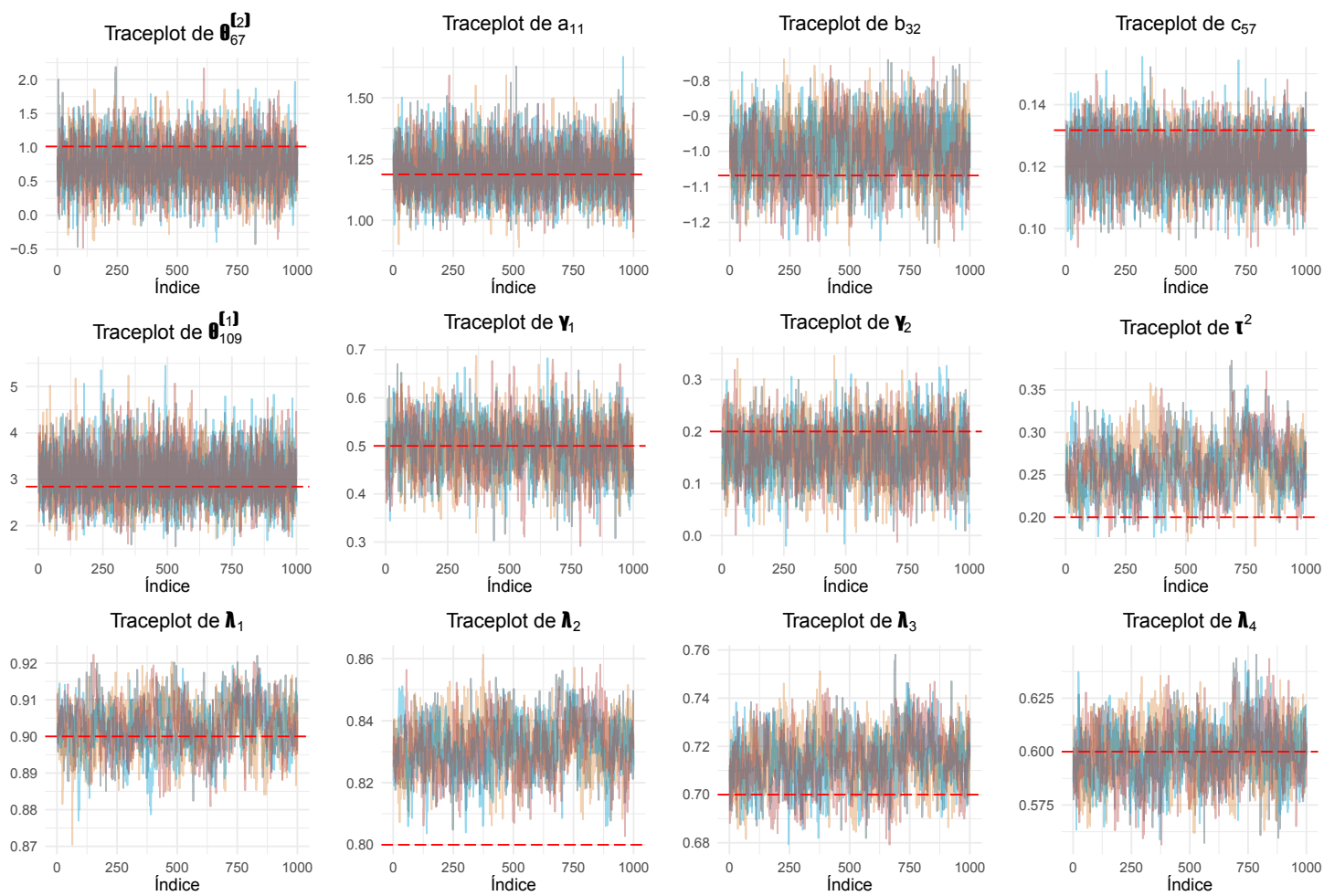
Fonte: Elaboração própria

Figura 26 – Variação 3 - Modelo Multinível



Fonte: Elaboração própria

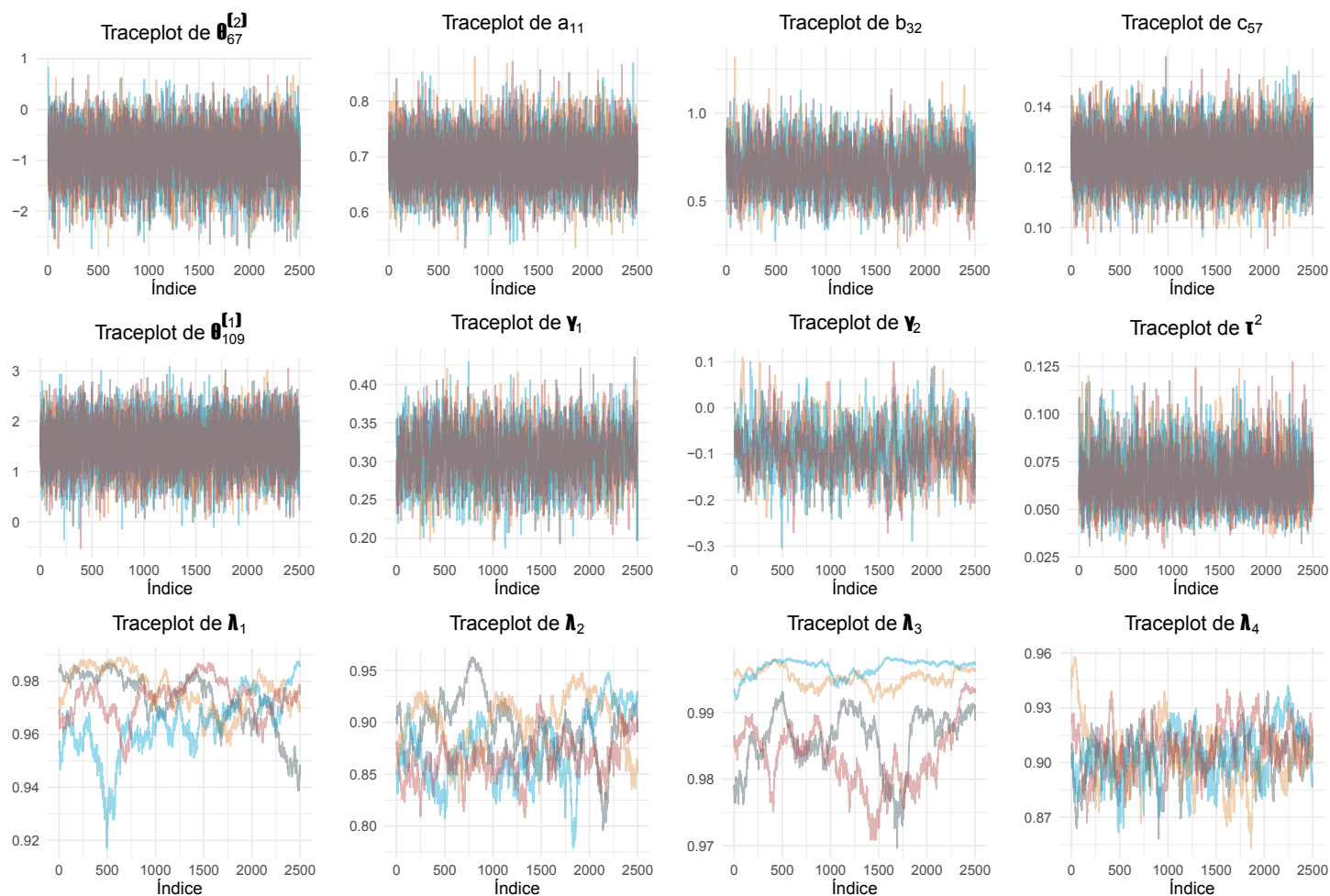
Figura 27 – Variação 4 - Modelo Completo



Fonte: Elaboração própria

Apêndice C - Traceplots da aplicação do modelo proposto aos dados do SAEB

Figura 28 – Traceplot do modelo com dados do SAEB - Matemática

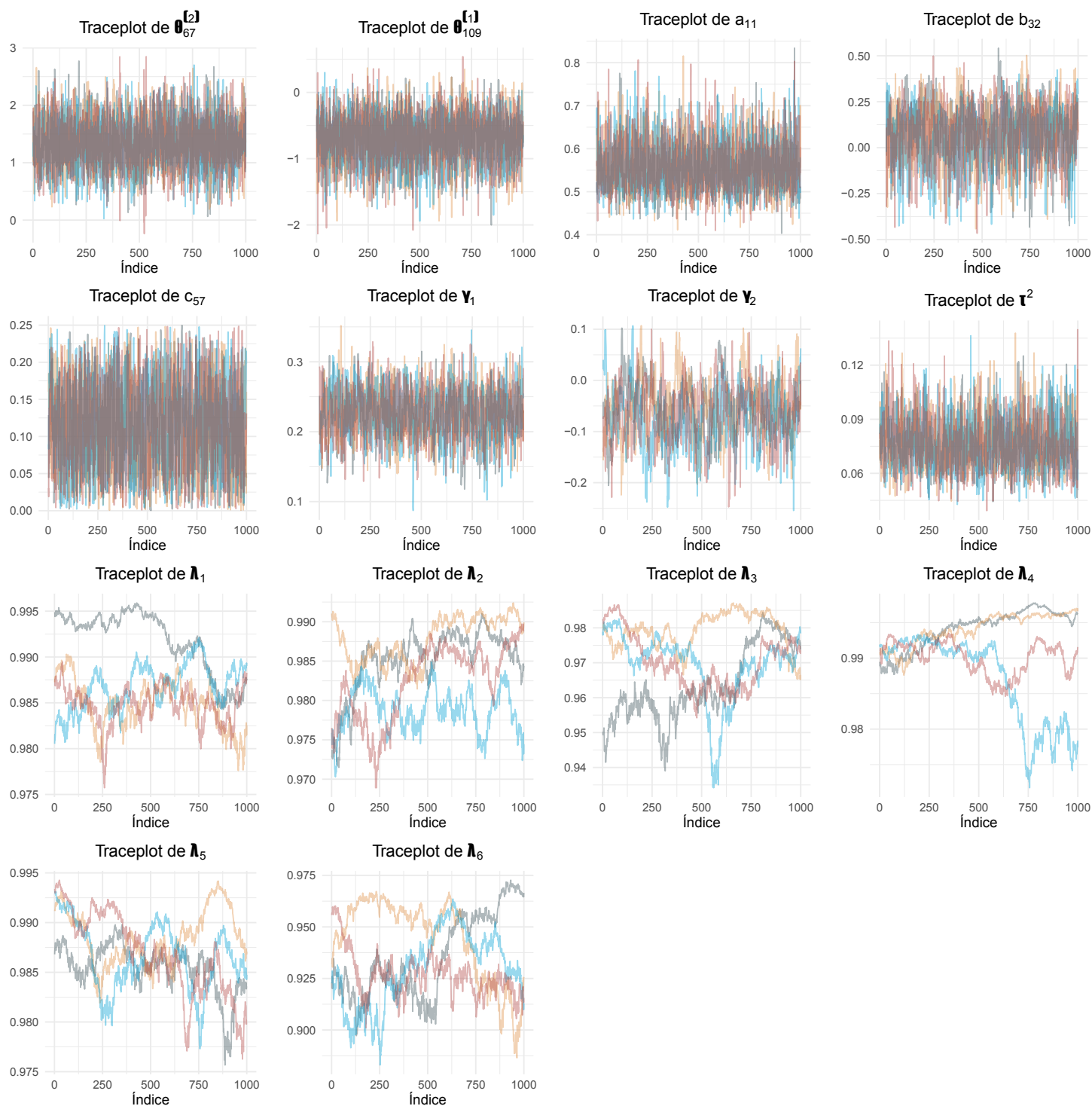


Valores estimados de $J = 8.378$ alunos da 3^a série do Ensino Médio das 75 escolas da rede estadual participantes do SAEB 2019 em Belo Horizonte.

Dados estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Figura 29 – Traceplot do modelo com dados do SAEB - Língua Portuguesa



Valores estimados de $J = 8.378$ alunos da 3ª série do Ensino Médio das 75 escolas da rede estadual participantes do SAEB 2019 em Belo Horizonte.

Dados estimados a partir de 4 cadeias de 1.000 amostras coletadas após *warm-up* de 1.000 iterações.

Fonte: Elaboração própria

Apêndice D - Matriz de Referência de Matemática da 3ª série do Ensino Médio

Tabela 23 – Recorte de temas e habilidades da Matriz de Referência do SAEB 2019: Matemática 3ª série do ensino médio

Tema	Habilidade
Espaço e Forma	D1. Identificar figuras semelhantes mediante o reconhecimento de relações de proporcionalidade.
	D2. Reconhecer aplicações das relações métricas do triângulo retângulo em um problema que envolva figuras planas ou espaciais.
	D3. Relacionar diferentes poliedros ou corpos redondos com suas planificações ou vistas.
	D4. Identificar a relação entre o número de vértices, faces e/ou arestas de poliedros expressa em um problema.
	D5. Resolver problema que envolva razões trigonométricas no triângulo retângulo (seno, cosseno, tangente).
	D6. Identificar a localização de pontos no plano cartesiano.
	D7. Interpretar geometricamente os coeficientes da equação de uma reta.
	D8. Identificar a equação de uma reta apresentada a partir de dois pontos dados ou de um ponto e sua inclinação.
	D9. Relacionar a determinação do ponto de interseção de duas ou mais retas com a resolução de um sistema de equações com duas incógnitas.
	D10. Reconhecer, dentre as equações do 2º grau com duas incógnitas, as que representam circunferências.
Grandezas e Medidas	D11. Resolver problema envolvendo o cálculo de perímetro de figuras planas.
	D12. Resolver problema envolvendo o cálculo de área de figuras planas.
	D13. Resolver problema envolvendo a área total e/ou volume de um sólido (prisma, pirâmide, cilindro, cone, esfera).

Fonte: Adaptado de [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira \(2019\)](#)