



PUC

RIO

Dissertação de Mestrado

Da resposta ao traço latente a TRI como um problema de otimização

Gabriel Minchio Ayres

Pontifícia Universidade Católica do Rio de Janeiro
Centro Técnico Científico
Departamento de Matemática

Rio de Janeiro, 26 de setembro de 2025



Pontifícia
Universidade
Católica do
Rio de Janeiro

Dissertação de Mestrado

Da resposta ao traço latente

a TRI como um problema de otimização

Gabriel Minchio Ayres

Orientação: Professor Dr. Sinésio Pesco

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Matemática pelo Programa de Pós-graduação em Matemática do Departamento de Matemática da PUC-Rio.

Rio de Janeiro, 26 de setembro de 2025



Pontifícia
Universidade
Católica do
Rio de Janeiro

Da resposta ao traço latente

a TRI como um problema de otimização

Gabriel Minchio Ayres

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Matemática pelo Programa de Pós-graduação em Matemática do Departamento de Matemática da PUC-Rio. Aprovada pela Comissão examinadora abaixo:

Professor Dr. Sinésio Pesco
Orientador
Departamento de Matemática – PUC-Rio

Professor Dr. Marcos Craizer
Pontifícia Universidade Católica – PUC-Rio

Professor Dr. Luiz Manoel Silva Fiqueiredo
Universidade Federal Fluminense

Rio de Janeiro, 26 de setembro de 2025



Todos os direitos reservados. A reprodução, total ou parcial, do trabalho é proibida sem autorização da universidade, do autor e do orientador.

Gabriel Minchio Ayres

Egresso do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo (Ifes), antigo Cefetes, é Licenciado em Matemática pela Universidade Federal do Espírito Santo (Ufes). Atua como professor de Matemática no Ensino Fundamental II e no Ensino Médio da rede privada e pública, incluindo a Educação de Jovens e Adultos (EJA), na Grande Vitória/ES. É técnico pedagógico na Secretaria da Educação do Espírito Santo (SEDU), com experiência em avaliações em larga escala como Enem, SAEB e PAEBES.

Ficha Catalográfica

Ayres, Gabriel Minchio

Da resposta ao traço latente : a TRI como um problema de otimização / Gabriel Minchio Ayres ; orientação: Sinésio Pesco. – 2025.

209 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2025.

Inclui bibliografia

1. Matemática – Teses. 2. Teoria da Resposta ao Item. 3. TRI. 4. Modelos logísticos. 5. Psicometria. 6. Otimização. I. Pesco, Sinésio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 510

À professora Olga Tovar, por acender a fagulha que tornou a matemática um caminho possível para mim.

Agradecimentos

Agradeço à minha mãe, Maria Helena, e à minha irmã, Gabriela. Foram meu porto seguro, minha base firme. Sem o apoio de vocês, nada disso teria sido possível. Obrigado por acreditarem em mim e me darem tudo o que precisei para seguir em frente.

Aos colegas de trabalho, que foram muito mais do que isso. O incentivo, as conversas e a torcida de vocês fizeram toda a diferença, especialmente nos momentos mais difíceis. Sou grato por cada palavra de apoio.

À turma do PROFMAT da PUC-Rio de 2023, obrigado pela parceria, pelas risadas e pela amizade. Ter vocês por perto tornou o caminho mais leve e a jornada muito mais divertida.

Aos professores do PROFMAT, meu sincero agradecimento por todo o conhecimento compartilhado e pela disposição em ajudar sempre. A atenção e o cuidado de vocês foram fundamentais para que eu me sentisse motivado.

À comunidade da PUC-Rio, meu agradecimento pelo acolhimento ao longo de toda a jornada. Em especial, à professora Christine Sertã, pelas aulas sempre marcantes e pela atenção constante como Coordenadora do PROFMAT. À professora Dania Gonzalez e ao professor Luís Eduardo, pela atenção, escuta e orientações que me ajudaram a seguir com mais segurança. À Mariana Soares, do Departamento de Matemática, minha gratidão pela disponibilidade e pela ajuda nos momentos em que tudo parecia mais difícil. Vocês tornaram esse percurso mais leve e possível.

Ao professor Sinésio, meu orientador, muito obrigado por topar esse desafio comigo. Sua paciência, clareza e precisão foram essenciais para que este trabalho tomasse forma. Agradeço por cada conversa, cada sugestão e por me guiar com tanta generosidade.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Código de Financiamento 001.

Resumo

Ayres, Gabriel Minchio; Pesco, Sinésio. **Da resposta ao traço latente: a TRI como um problema de otimização.** Rio de Janeiro, 2025. 209 p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Avaliar a aprendizagem dos estudantes de forma consistente, precisa e significativa exige mais do que simplesmente contar acertos em testes. A Teoria da Resposta ao Item (TRI) propõe uma abordagem robusta para interpretar o desempenho dos indivíduos com base em modelos probabilísticos que consideram tanto as características dos itens quanto os traços latentes dos respondentes. Neste contexto, esta dissertação analisa a TRI com foco na formulação matemática dos modelos psicométricos logísticos, especialmente o modelo logístico de três parâmetros (ML3P). O trabalho modela a função de verossimilhança conjunta e investiga a estimativa dos parâmetros como um problema de otimização não linear multidimensional. Para resolvê-lo, a pesquisa adota o *Joint Maximum Likelihood Estimation* (JMLE) como método de estimativa dos parâmetros e aplica algoritmos, como o de Newton-Raphson, para encontrar os valores que maximizam a função objetivo nesse processo. Como aplicação, o estudo propõe estratégias que utilizam itens de edições anteriores do Enem para monitorar a aprendizagem matemática de estudantes do Ensino Médio, promovendo uma leitura mais precisa e pedagógica de seus desempenhos. Com isso, a dissertação busca contribuir para o desenvolvimento de instrumentos avaliativos capazes de mensurar o conhecimento com maior equidade e profundidade, aproximando a TRI da prática docente e da sala de aula.

Palavras-chave

Teoria da Resposta ao Item; TRI; modelos logísticos; psicometria; otimização; Newton-Raphson; Enem; verossimilhança; avaliação.

Abstract

Ayres, Gabriel Minchio; Pesco, Sinésio. **From Response to Latent Trait: IRT as an Optimization Problem.** Rio de Janeiro, 2025. 209 p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Assessing students' learning in a consistent, accurate, and meaningful way requires more than simply counting correct answers on tests. Item Response Theory (IRT) offers a robust framework for interpreting individual performance based on probabilistic models that take into account both the characteristics of test items and the latent traits of respondents. Within this context, this dissertation analyzes IRT with a focus on the mathematical formulation of logistic psychometric models, particularly the three-parameter logistic model (3PL). The study models the joint likelihood function and examines parameter estimation as a multidimensional nonlinear optimization problem. To address this, the research employs the Joint Maximum Likelihood Estimation (JMLE) method and applies algorithms such as Newton-Raphson to determine the parameter values that maximize the objective function. As a practical application, the study proposes strategies that use items from previous editions of the Brazilian National High School Exam (Enem) to monitor students' mathematical learning, promoting a more precise and pedagogically informed interpretation of their performance. In doing so, the dissertation aims to contribute to the development of assessment tools capable of measuring knowledge with greater fairness and depth, bringing IRT closer to teaching practice and the classroom context.

Keywords

Item Response Theory; IRT; logistic models; psychometrics; optimization; Newton-Raphson; Enem; likelihood; educational assessment.

Sumário

1	Introdução	17
2	Um pouco de psicometria	22
2.1	A medida em ciências psicossociais	22
2.2	A Psicometria	23
2.2.1	Os testes psicométricos	25
2.2.2	Psicometria Clássica	26
2.2.3	Psicometria Moderna	28
2.3	Fundamentos da TRI	30
3	Conceitos estatísticos preliminares	32
3.1	Modelos matemáticos e experimentos aleatórios	32
3.2	Variáveis aleatórias	33
3.3	Distribuição de probabilidade	35
3.4	Modelos probabilísticos	36
3.4.1	Distribuição de Bernoulli	37
3.4.2	Distribuição Binomial	38
3.4.3	Distribuição Normal	39
3.5	Função de distribuição acumulada	40
3.5.1	A ogiva normal padrão	41
3.6	Função logística	42
4	Modelos logísticos unidimensionais	46
4.1	Matriz de padrão de resposta	46
4.2	Modelos Logísticos	48
4.2.1	Modelo Logístico de 3 parâmetros da TRI (Modelo de Birnbaum) ..	50
4.2.2	Modelo Logístico de 2 parâmetros da TRI (Modelo de Lord)	51
4.2.3	Modelo Logístico de 1 parâmetro da TRI (Modelo de Rasch)	51
4.3	O fator de escala D	53
4.4	Cálculo da probabilidade de acerto	54
4.5	Cálculo da probabilidade de erro	57
4.6	Comparação entre os modelos logísticos	58

5	<i>Joint Maximum Likelihood Estimation</i> no ML3P	60
5.1	Estimadores e estimativas	60
5.2	Estimador de Máxima Verossimilhança	61
5.3	Função de verossimilhança	63
5.4	Função de log-verossimilhança	64
5.5	Estimação conjunta dos parâmetros e aptidões	66
6	Calibração dos itens no ML3P usando JMEL	68
6.1	A função de verossimilhança dos parâmetros no ML3P	68
6.2	A função de log-verossimilhança dos parâmetros	72
6.3	Separabilidade por itens da log-verossimilhança $\ell(\Lambda)$	74
6.4	Estimador dos parâmetros dos itens.....	76
6.5	Fórmula de recorrência dos parâmetros por Newton-Raphson	83
6.6	A Hessiana da função log-verossimilhança dos parâmetros	88
7	<i>Scoring</i> das aptidões usando JMEL.....	100
7.1	A verossimilhança e a log-verossimilhança das aptidões	100
7.2	Separabilidade por indivíduos da log-verossimilhança $\ell(\theta)$	101
7.3	Estimador da aptidão	102
7.4	Fórmula de recorrência das habilidades por Newton-Raphson	105
8	Escala de proficiência	109
8.1	Curva Característica do Item (CCI).....	109
8.1.1	A CCI e o parâmetro de dificuldade b	114
8.1.2	A CCI e o parâmetro de discriminação a	116
8.1.3	A CCI e o parâmetro de acerto ao acaso c	119
8.2	A curva de probabilidade de erro	120
8.3	Escala de proficiência	121
8.4	Interpretação pedagógica	122
8.5	Comparação de escores	125
8.6	Coerência pedagógica	125
9	As Crônicas do Enem: O Teorema, a Distribuição e a Escala	130
9.1	O Teorema de Bayes e as distribuições <i>a priori</i> e <i>a posteriori</i>	130
9.2	Valor esperado de uma variável aleatória contínua	132

9.3 JMLE e MMLE	133
9.4 Estimação dos parâmetros e das proficiências no Enem	135
9.4.1 Estimação dos parâmetros dos itens via MMLE	135
10 Conclusão	139
11 Referências bibliográficas	140
Apêndice A – Recurso Educacional	146
Apêndice B – Estatísticas descritivas do Enem de Matemática 2023	142
Apêndice C – Estatísticas descritivas do Enem de Matemática 2024	192
Apêndice D – CCI dos itens de Matemática do Enem 2024.....	202
Anexo A – Parâmetros dos itens de Matemática do Enem 2024	205
Anexo B – Matriz de habilidades de Matemática para o Enem	206
Anexo C – Objetos do conhecimento de Matemática para o Enem	208

Lista de ilustrações

Figura 3.1 – Espaço amostral.....	34
Figura 3.2 – Variável aleatória.....	34
Figura 3.3 – Função massa de probabilidade	36
Figura 8.1 – Escala de Proficiência e Interpretação Pedagógica	123
Figura 8.2 – Questão 180, prova amarela, Enem 2021.....	124

Gráfico 3.1 – Função massa de prob. da distribuição de Bernoulli	38
Gráfico 3.2 – Função massa de probabilidade da distribuição Binomial ..	39
Gráfico 3.3 – Função densidade de probabilidade	40
Gráfico 3.4 – Ogiva da distribuição de Bernoulli.....	41
Gráfico 3.5 – Ogiva normal padrão	42
Gráfico 3.6 – Função logística padrão.....	43
Gráfico 3.7 – Gráfico da derivada da função logística padrão.....	43
Gráfico 3.8 – Ogiva normal padrão e sigmoide da função logística	44
Gráfico 3.9 – Ajuste da curva logística padrão à ogiva padrão normal	45
Gráfico 4.1 – Distribuição normal padrão das aptidões na população	51
Gráfico 4.2 – Probabilidade de acerto e nível de aptidão	56
Gráfico 4.3 – Probabilidade de erro e nível de aptidão	58
Gráfico 8.1 – Análise Gráfica do Item 1.....	110
Gráfico 8.2 – Dispersão dos resultados (erros e acertos) do item 1.....	110
Gráfico 8.3 – Regressão logística dos resultados (erros e acertos)	111
Gráfico 8.4 – Dispersão de erros e acertos tipo <i>bubble plot</i>	111
Gráfico 8.5 – Curva Característica do Item 1 na métrica (0, 1)	112
Gráfico 8.6 – Assíntotas horizontais da CCI.....	113
Gráfico 8.7 – Curva Característica do Item 1	113
Gráfico 8.8 – Curva Característica do Item dos itens 2 a 10	114
Gráfico 8.9 – Curvas com diferentes discriminações	115
Gráfico 8.10 – Parâmetro de discriminação	116
Gráfico 8.11 –Efeito do parâmetro de discriminação.....	117
Gráfico 8.12 – Parâmetro de discriminação próximo de zero.....	118

Gráfico 8.13 – Discriminação perfeita	118
Gráfico 8.14 – Discriminação negativa	119
Gráfico 8.15 – Parâmetro de acerto ao acaso	120
Gráfico 8.16 – Probabilidade de erro	121
Gráfico 8.17 – Escala de Proficiência	124

Lista de tabelas e quadros

Tabela 4.1 – Matriz de padrão de resposta $U_{7 \times 9}$	54
Tabela 4.2 – Valor dos parâmetros dos itens do teste	55
Tabela 4.3 – Probabilidade de acerto em cada item do teste,	55
Tabela 4.4 – Probabilidade de erro em cada item do teste,	57
Tabela 4.5 – Comparaçao entre os modelos logísticos.....	59
Tabela 8.1 – Parâmetros do teste hipotético	112
Quadro 8.1 – Classificação dos itens quando à discriminação	115
Quadro 8.2 – Classificação da discriminação dos itens na TRI.....	119
Quadro 8.3 – Estudo de caso: parâmetros dos itens	126
Quadro 8.4 – Estudo de caso: aptidões do testandos.....	126
Quadro 8.5 – Estudo de caso: probabilidades de acerto.....	126
Quadro 8.6 – Estudo de caso 1: verossimilhança	127
Quadro 8.7 – Estudo de caso 2: verossimilhança	127
Quadro 8.8 – Estudo de caso 3: verossimilhança	127
Quadro 8.9 – Comparaçao dos escores.....	128

Lista de abreviaturas e siglas

AGI	Análise Gráfica do Item
CCI	Curva Característica do Item
EAP	Expected a Posteriori
EB	Empirical Bayes
EMV	Estimador de Máxima Verossimilhança
EMVC	Estimador de Máxima Verossimilhança Conjunta
Enem	Exame Nacional do Ensino Médio
FDA	Função de Distribuição Acumulada
FDP	Função Densidade de Probabilidade
IRT	Item Response Theory
JMLE	Joint Maximum Likelihood Estimation
MAP	Maximum a Posteriori
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
ML1P	Modelo Logístico de 1 parâmetro da TRI (Modelo de Rasch)
ML2P	Modelo Logístico de 2 parâmetros da TRI (Modelo de Lord)
ML3P	Modelo Logístico de 3 parâmetros (Modelo de Birnbaum)
MMLE	Marginal Maximum Likelihood Estimation
TCT	Teoria Clássica dos Testes
TRI	Teoria da Resposta ao Item

Lista de símbolos

a_i	Parâmetro de discriminação do item
b_i	Parâmetro de dificuldade do item
c_i	Parâmetro de pseudochute
n	Número de itens
m	Números de indivíduos testandos
D	Fator de escala logística-normal padrão
X_i	Variável aleatória discreta resultado do item i
X_j	Vetor aleatório do resultado de cada item do indivíduo j
X_{ji}	Variável aleatória disc. do resultado do item i respondido pelo indivíduo j
u_{ji}	Resultado que de fato o indivíduo j obteve ao responder 0 item i
U_j	Vetor com todos os resultados u_{ji} que de fato o indivíduo j obteve
U	Matriz $m \times n$ com as respostas de fato obtidas pelos m indivíduos nos n itens
λ_i	Vetor dos parâmetros (a_i, b_i, c_i) do item i
Λ	Vetor dos vetores dos parâmetros $(\lambda_1, \dots, \lambda_n)$ dos n itens
θ_j	Aptidão do indivíduo j
Θ	Vetor de aptidões $(\theta_1, \dots, \theta_m)$
L	Função de verossimilhança
ℓ	Função de log-verossimilhança
Ψ	Função logística
$\nabla_{\lambda_i} P$	Jacobiana da função de probabilidade do ML3P
$\nabla_{\lambda_i}^2 P$	Hessiana da função de probabilidade do ML3P
$H\ell$	Hessiana global (Λ) da log-verossimilhança
$h_i\ell$	Hessiana local (λ) da log-verossimilhança com relação ao item i

*Entre a resposta e o acerto,
há um traço invisível que a matemática pode revelar.*
Gabriel Minchio Ayres

1

Introdução

Dois estudantes que participaram do Exame Nacional do Ensino Médio (Enem) analisam juntos o boletim de desempenho divulgado pelo Inep, comparando suas respectivas notas nas quatro provas objetivas do exame. Como o boletim informa apenas a pontuação de cada área do conhecimento, sem detalhar o número de acertos, eles utilizam o gabarito oficial para corrigir manualmente as provas e descobrir quantas questões acertaram. Ao compararem seus respectivos resultados na prova de Matemática e suas Tecnologias, percebem algo curioso: enquanto um acertou 30 questões e obteve 696,0 pontos, o outro, com apenas 20 acertos, recebeu 701,6 pontos. Essa aparente incoerência os leva a questionar se marcaram o gabarito corretamente ou se houve algum erro no processamento das notas.

Embora fictícia, essa situação foi inspirada em dados reais da edição de 2023 do Enem e reflete um cenário bastante comum, também presente em outras avaliações externas, como o Saeb, o Paebe (no Espírito Santo) e exames internacionais de proficiência em língua inglesa, como o TOEFL iBT. Nessas situações, é comum questionar a forma como as notas são calculadas, mas raramente se discute o verdadeiro significado desses resultados em termos de aprendizagem. No contexto da rotina escolar, o processo é um pouco diferente: as notas, quando baseadas nos acertos de testes, são frequentemente aceitas sem grandes questionamentos, apesar de ainda nesse caso não haver muita reflexão sobre o que elas realmente revelam sobre o aprendizado do estudante.

Para ilustrar, vamos considerar outra situação hipotética: dois estudantes realizam a mesma avaliação na escola. Um deles, Maria, é dedicada, participa das aulas e sempre cumpre as atividades, mas costuma ser insegura em situações de avaliação. O outro estudante, João, ao contrário, não demonstra interesse pelas atividades escolares e apresenta dificuldades de concentração, o que impacta significativamente em seu aprendizado. Ambos realizam um teste com dez questões, sendo quatro consideradas difíceis e seis fáceis, segundo o próprio professor. Após a correção, verifica-se que ambos acertaram cinco questões, porém questões distintas um do outro. No entanto, do total de acertos, Maria acertou três das difíceis, enquanto João acertou apenas uma.

Essa nova situação abre espaço para várias reflexões, que vão além do possível estado emocional dos estudantes durante a prova ou da gestão do tempo disponível para responder às questões. Por exemplo, de imediato podemos nos questionar se o fato de ambos os estudantes terem obtido o mesmo número de acertos implica em terem desenvolvido, no mesmo nível, as habilidades avaliadas. Outras seriam: se Maria acertou a maioria das questões mais complexas, por que teve dificuldade com as consideradas fáceis? Será que essas questões, embora simples, exigiam habilidades que ela ainda precisa desenvolver? Se sim, quais habilidades? E João, mesmo sendo agitado e desmotivado, conseguiu acertar metade do teste. O que pode ter contribuído para esse desempenho apesar das suas dificuldades? Além disso, chama atenção o fato de João ter acertado uma das questões difíceis e, mesmo assim, errar outras que supostamente seriam mais acessíveis. E o instrumento utilizado, a prova: será que foi, de fato, adequado para medir o conhecimento dos estudantes submetidos a ele, trazendo um retrato próximo do real aprendizado de cada um?

Perceba que, sem critérios claros de análise, essas reflexões permanecem subjetivas, o que dificulta uma leitura mais precisa do que cada estudante aprendeu, do que precisa melhorar, ou a partir de onde pode avançar. Assim, a nota de um teste obtida da forma tradicional, com base apenas no número de acertos, acaba funcionando apenas como um número, usado para definir se o estudante será promovido (ou não), sem necessariamente indicar em que nível do processo de aprendizagem ele realmente se encontra. Nesse contexto, embora a prática avaliativa nas escolas ainda esteja fortemente associada à classificação e promoção dos estudantes, há autores que propõem uma perspectiva mais formativa e reflexiva desse processo. Por exemplo, para Luckesi,

[...] a aferição da aprendizagem manifesta-se como um processo de compreensão dos avanços, limites e dificuldades que os educandos estão encontrando para atingir os objetivos do curso, disciplina ou atividade da qual estão participando. A avaliação é, neste contexto, um excelente mecanismo subsidiário da condução da ação (Luckesi, 2013, p. 77).

Ainda sob esse prisma, muitos sistemas educacionais estabelecem metas que os estudantes devem alcançar para serem promovidos, indicando, assim, o progresso esperado em sua aprendizagem. Essas metas geralmente estão bem definidas numericamente, como ocorre no sistema de ensino público estadual do Espírito Santo, do Rio de Janeiro e de vários outros estados, onde os estudantes são aprovados ao atingir 60 dos 100 pontos distribuídos ao longo do ano letivo em cada um dos componentes curriculares que não são avaliados por conceito. No entanto, o que essa meta realmente representa em termos de aprendizado não é tão claro. Essa falta de definição faz com que as escolas, por diversos motivos, priorizem e fomentem o cumprimento numérico da meta, que não necessariamente reflete uma aquisição de mínimo aprendizado.

Por exemplo, dois estudantes de sistemas educacionais diferentes que atingem 60% de aproveitamento final em Matemática não necessariamente consolidaram as mesmas habilidades, muito menos no mesmo nível. Na verdade, tal fato é bem presente ainda que com estudantes de uma mesma turma e não ocorre apenas pela imprecisão das metas, mas também pela ausência de padronização nos instrumentos de medida das proficiências parciais dos estudantes a cada período e nas métricas estabelecidas. Afinal, o que significa, objetivamente, obter 60% de aproveitamento em relação a todo o currículo prescrito ou desenvolvido no ano letivo? E como medir isso sem contradições ou ambiguidades nos instrumentos e métricas utilizados? Quais são as fragilidades e as potencialidades do estudante que atinge esse aproveitamento mínimo e quais habilidades ele precisa fortalecer para continuar seu desenvolvimento, e em que medida?

Assim, há dois pontos a se considerar no processo avaliativo diante das problemáticas anteriormente apresentadas. O primeiro seria o estabelecimento de algum instrumento ou metodologia que possa ser utilizado nos diferentes sistemas de ensino, a nível escolar, a fim de se padronizar o monitoramento da aprendizagem dos estudantes e que permita a comparação de proficiências de forma atemporal. O segundo envolve uma discussão mais profunda em termos de currículo – que não será feita aqui –, com a definição clara do mínimo que o estudante precisa aprender em cada etapa de ensino, isto é, qual a mínima proficiência em cada habilidade que o educando precisa atingir para ser considerado apto a ser promovido, que traduziria a meta de 60% na situação anterior. O desafio dessa empreitada é justamente vincular esses dois pontos: estabelecer um instrumento de medida padronizado que

possa verificar, com boa assertividade, se o estudante atingiu a proficiência mínima estabelecida.

A elaboração de instrumentos que possam monitorar a aprendizagem do estudante, como descrito, exige tempo e dedicação, recursos já limitados no dia a dia dos professores, além de conhecimentos específicos. Tendo isso em vista, com este trabalho pretende-se discutir o primeiro dos pontos descritos anteriormente – justamente a elaboração de instrumentos de avaliação – a partir da análise dos testes e resultados do Enem, um instrumento amplamente utilizado no Brasil para avaliar a proficiência dos estudantes em diferentes áreas do conhecimento.

A partir de 2009, foi implementada a Teoria da Resposta ao Item (TRI) como metodologia para a elaboração das provas do Enem, bem como para se obter a proficiência dos seus participantes. A TRI se desenvolveu a partir de uma área da psicologia, chamada Psicometria, que, de forma geral, usa métodos estatísticos e matemáticos na otimização de funções de verossimilhança para fazer medições de variáveis inobserváveis psicológicas que influenciam o comportamento humano, como o desempenho em provas e testes, de forma que é possível estabelecer uma medida à qualidade desse comportamento. No caso do Enem, essa medida gera a proficiência do participante em cada área do conhecimento, ou seja, a pontuação dele nas provas objetivas, como ocorreu com os dois estudantes na situação inicialmente apresentada.

Ao utilizar a TRI para estimar a proficiência de um indivíduo em uma determinada habilidade de Matemática, por exemplo, é como se esse sujeito despejasse sobre uma mesa todo o seu conhecimento em relação a essa habilidade e fosse possível medi-lo com um instrumento de medida. Nesse caso, o instrumento seria um teste construído com essa metodologia e a unidade de medida seria estabelecida a partir de uma escala de proficiência obtida pela própria metodologia aplicada.

Para determinar a proficiência, a TRI considera as características das questões da prova e a coerência das respostas. Isso permite comparar as notas dos participantes de diferentes edições do Enem, garantindo uma avaliação mais precisa e justa. Dessa forma, os resultados do exame podem ser utilizados em programas como o Sistema de Seleção Unificada (Sisu) e o Programa Universidade para Todos (ProUni) como forma de acesso ao ensino superior.

A análise e discussão que serão feitas aqui têm como objetivo, além de apresentar os conceitos que fundamentam a TRI e descrever a matemática utilizada,

propor, ao final, o uso de questões de edições anteriores do Enem no monitoramento da aprendizagem dos estudantes em Matemática. Dessa forma, será possível, por meio de testes, uma compreensão sistemática das habilidades de Matemática (e dos demais componentes) que foram (ou não) consolidadas pelos estudantes, tentando evitar discussões subjetivas, um vez que entre a resposta e o acerto, há um traço invisível que a matemática pode revelar. É essa busca por uma leitura mais precisa e profunda do conhecimento que direciona esta dissertação à análise da Teoria da Resposta ao Item.

Para isso, primeiramente, serão apresentados os princípios que fundamentam a metodologia aplicada na elaboração das provas do Enem, seguido de uma descrição geral dos procedimentos matemáticos utilizados para determinar a proficiência dos estudantes nas áreas do conhecimento. Por fim, antes de propor uma forma de utilizar as questões do Enem para monitorar o desenvolvimento em Matemática dos estudantes do Ensino Médio, será realizada uma breve análise estatística descritiva dos resultados das edições de 2023 e 2024 do Enem, especificamente na área de Matemática e suas Tecnologias, considerando, prioritariamente, quando possível, o desempenho dos participantes matriculados na 3^a série do Ensino Médio no ano de cada edição, em âmbito nacional.

Assim, pretende-se contribuir para um entendimento mais aprofundado dos docentes sobre a aplicação da TRI nas avaliações de proficiência, destacando como esse método pode ser utilizado para monitorar o aprendizado dos estudantes de forma mais precisa e eficaz, principalmente em Matemática. Além disso, busca-se promover reflexões sobre a importância de instrumentos padronizados na mensuração do conhecimento e dar subsídios para o desenvolvimento de avaliações desse gênero na rotina escolar.

2

Um pouco de psicometria

Embora esta exposição tenha como foco a matemática aplicada à Teoria da Resposta ao Item (TRI) e suas aplicações no ensino e na aprendizagem da matemática, é essencial considerar os conceitos da Psicologia que fundamentam essa teoria. Afinal, a TRI não pertence propriamente à Matemática ou à Estatística, mas sim à Psicometria, um ramo da Psicologia que utiliza matemática e análises estatísticas como ferramentas. Assim, a primeira parte deste trabalho tem como objetivo, ainda que de forma breve, apresentar conceitos da psicologia, especialmente da Psicometria, que servirão como base para os temas abordados ao longo do texto.

2.1

A medida em ciências psicossociais

Medir é comparar o que se quer mensurar com uma unidade de medida, atribuindo um número ou gradação ao atributo medido. Ao que se está medindo, chamamos grandeza, cuja medida pode ser, de acordo com Pasquali (2003), de quatro tipos: medida fundamental, medida derivada, medida por lei ou medida por teoria.

Uma medida é considerada fundamental quando pode ser determinada diretamente, ou seja, quando é possível utilizar um instrumento que forneça o valor da grandeza de imediato, utilizando uma unidade de medida primitiva. A medição ocorre pela comparação dessa unidade com o atributo mensurado. Nesse caso, o instrumento utilizado compartilha a mesma característica essencial do que está sendo medido e, ou corresponde à própria unidade primitiva, ou é formado por várias dessas unidades. Exemplos disso são as medidas de grandezas como massa, comprimento e tempo.

Outras grandezas não possuem uma unidade de medida primitiva. Para serem medidas, elas dependem, a princípio, da combinação de duas ou mais unidades primitivas, assim como de seus respectivos instrumentos e técnicas de medição. Essas grandezas, na verdade, são compostas por outras que se relacionam de alguma forma, sem que seja necessária uma lei ou teoria para estabelecer essa relação. Nesse contexto, podemos dizer que tais grandezas podem ser medidas de forma parcialmente direta, isto é, para medi-las, a princípio, basta medir cada grandeza

envolvida utilizando suas respectivas unidades e instrumentos de medida e, em seguida, estabelecer a relação entre elas. O resultado obtido é uma medida derivada. Exemplos disso são as grandezas velocidade e densidade demográfica.

Diferentemente das medidas fundamentais e derivadas, há casos em que uma grandeza não pode ser determinada diretamente e sua medição depende de uma lei ou teoria. A principal diferença entre lei e teoria científica está na finalidade de cada uma, não havendo uma hierarquia entre elas. Enquanto uma lei científica descreve um fenômeno natural, sem explicar o motivo, uma teoria científica busca explicar o porquê e o mecanismo por trás desse fenômeno, integrando múltiplas evidências, hipóteses e leis dentro de um quadro coerente.

Assim, podemos entender que algumas grandezas não podem ser medidas de forma direta, como ocorre com quase todos os atributos em ciências psicossociais. Essas grandezas não possuem uma unidade primitiva, mas podem ser mensuradas quando se estabelece uma lei que as relate, obtendo, nesse caso, uma medida por lei. Um exemplo clássico é a força com a qual dois corpos mutuamente se atraem, estabelecida pela Lei da Gravitação Universal proposta por Isaac Newton.

Por outro lado, na ausência de uma lei, uma grandeza desse tipo pode ser medida por meio de uma teoria que proponha uma relação hipotética entre ela e outras grandezas dentro de um fenômeno específico. Nessa última situação, obtemos uma medida por teoria, fundamentada em axiomas ou postulados e não em fatos empíricos, como ocorre nas leis. Um exemplo são as medidas relacionadas a comportamentos (τ) e a traços latentes ou aptidões (θ), abordadas pela Teoria Psicométrica ou Teoria dos Testes Psicológicos, desenvolvidas pela Psicometria.

2.2

A Psicometria

A Psicometria é um ramo da Psicologia que se preocupa em medir características psicológicas. Ela discute e procura explicar as respostas dadas por um sujeito a um teste. Conforme Luiz Pasquali,

Etimologicamente, psicometria representa a teoria e a técnica de medida dos processos mentais, especialmente aplicada na área da Psicologia e da Educação. Ela se fundamenta na teoria da medida

em ciências em geral, ou seja, do método quantitativo que tem, como principal característica e vantagem, o fato de representar o conhecimento da natureza com maior precisão do que a utilização da linguagem comum para descrever a observação dos fenômenos naturais" (Pasquali, 2009, p. 993).

Ainda segundo Pasquali (2003), a Psicometria estuda os processos psicológicos a partir da relação entre duas variáveis. A primeira delas é o estímulo comportamental (variável observável), dado ao indivíduo por meio de um teste. Já a segunda variável, dependente da primeira, decorre do próprio indivíduo e é escolhida segundo a abordagem adotada. Essa abordagem, por sua vez, é fundamentada em duas concepções centrais que orientam o estudo da mente e do comportamento: o monismo materialista e o dualismo interacionista.

O monismo materialista considera que o próprio pensamento é físico, no sentido de que cada comportamento é um processo psicológico indissociável, sendo os fenômenos mentais produtos exclusivos da atividade cerebral. Com essa abordagem, a segunda variável estaria no nível do físico (variável observável), sendo ela o **comportamento** ou **tau (τ)**, manifestado pelo indivíduo após ter contato com o estímulo comportamental a que ele foi exposto inicialmente por meio do teste. É dessa concepção que surge a Psicometria Clássica, desenvolvendo a Teoria Clássica dos Testes (TCT), que, no processo de medida de características psicológicas, relaciona apenas variáveis comportamentais (teste e tau), ou seja, variáveis observáveis.

Já o dualismo interacionista defende a interação entre mente e corpo como elementos distintos, mas interligados, no sentido de que um processo psicológico influencia o comportamento do indivíduo e vice-versa. Nessa visão, a segunda variável estaria no nível do psíquico (variável não observável), sendo uma variável hipotética, chamada de **traço latente, aptidão, proficiência, habilidade ou theta (θ)**, que influencia o comportamento do indivíduo a partir do estímulo inicial dado a ele por meio do teste. É nessa abordagem que surge a Psicometria Moderna com a Teoria da Resposta ao Item (TRI), que relaciona uma variável observável (teste) com uma outra variável não observável hipotética (theta) no processo de medida de características psicológicas.

2.2.1

Os testes psicométricos

Comumente, os testes psicométricos são compostos por questões objetivas, chamadas **itens**. Esses itens possuem características específicas, conhecidas como **parâmetros**, que influenciam diretamente a medição do desempenho do indivíduo no teste. Esses parâmetros são estimados numericamente, e seus valores variam conforme a **teoria** e o **modelo** estatístico adotado. Além disso, tanto o número de parâmetros quanto quais parâmetros são utilizados dependem da abordagem teórica e do modelo estatístico escolhido.

Ao realizar o teste, o sujeito recebe um ponto para cada item respondido corretamente e nenhum ponto para os demais. Ao final, é obtido o **escore total**, ou simplesmente escore, que é a soma das pontuações obtidas. Como em toda medida há um **erro**, que pode surgir de diferentes fontes (o próprio instrumento de medida, o procedimento utilizado, fatores aleatórios ou erros de amostragem), esse escore total é chamado também de **escore bruto** e, retirando esse erro do escore bruto, obtemos o **escore verdadeiro**, que é a real medida do que o teste se propõe a medir.

Nesse sentido, podemos falar em **formas paralelas** de um teste. Dois testes com itens diferentes um do outro são paralelos quando medem a mesma coisa da mesma forma, ou seja, quando a aplicação dos dois testes a um mesmo grupo gera escores verdadeiros iguais para cada indivíduo, com a mesma variância dos erros. Perceba que, com relação à condição relativa aos erros de medida, a exigência para dois testes serem formas paralelas um do outro é que a variância dos erros precisa ser igual nos dois testes e não que a distribuição dos erros seja a mesma, ou seja, não estamos dizendo que, para todo grupo testado, o escore total de um indivíduo nos dois testes deve ter o mesmo erro, mas sim que a variância dos erros em um dos testes deve ser igual à variância dos erros no outro teste. Testes que obedecem totalmente aos dois critérios para formas paralelas são chamados de formas estritamente paralelas uma à outra.

Dentre os testes psicológicos apontados por Urbina (2007), destacamos para os fins desse trabalho os testes de aptidão e os testes de preferência. Enquanto os **testes de aptidão** avaliam habilidades e capacidades cognitivas, como raciocínio lógico, memória e velocidade de processamento, os **testes de preferência** investigam personalidade, interesses, tendências e atitudes individuais. A TCT e a TRI

podem ser utilizadas tanto em testes de aptidão quanto em testes de preferência, auxiliando na precisão da interpretação dos resultados. Nesse sentido, antes de aprofundar as especificidades da TRI, foco desta exposição, serão apresentadas, de maneira sucinta, as principais características de cada uma dessas duas teorias, conforme Pasquali (2003), traçando paralelos entre elas no contexto dos testes de aptidão, que são amplamente utilizados na determinação da proficiência de estudantes nas diversas habilidades das áreas do conhecimento.

2.2.2 Psicometria Clássica

A Teoria Clássica dos Testes se concentra na interpretação do resultado geral da avaliação, buscando compreender o significado do escore total. Para exemplificar, lembre-se da segunda situação mencionada no capítulo introdutório, em que dois estudantes, Maria e João, realizaram um teste com 10 questões, acertando, cada um, 5 questões distintas um do outro. Tomando essa situação como exemplo, na TCT, a preocupação estaria em compreender o significado desse 5, independentemente de quais questões foram acertadas. Nesse caso, os dois estudantes seriam colocados no mesmo patamar, uma vez que obtiveram o mesmo escore. Assim, o foco da Psicometria Clássica está concentrado no comportamento (τ) do sujeito, refletido no escore total do teste, sendo, então, de grande interesse por essa teoria a produção de testes de qualidade.

A ideia central do modelo teórico da TCT é determinar o escore verdadeiro (V) do indivíduo a partir do escore total (T) obtido no teste, considerando o possível erro de medida (ε), sendo o escore total a soma do escore verdadeiro com o erro da medida.

$$T = V + \varepsilon$$

O que ocorre é que, nessa equação, o escore verdadeiro e o erro são incógnitas e, a cada indivíduo que realiza o teste, duas novas incógnitas são adicionadas por meio de uma nova equação, formando, assim, um sistema que, com alta probabilidade, se torna um sistema cuja solução é inviável para essa situação caso não sejam feitas algumas considerações.

$$\begin{cases} T_1 = V_1 + \varepsilon_1 \\ T_2 = V_2 + \varepsilon_2 \\ T_3 = V_3 + \varepsilon_3 \\ \vdots \\ T_n = V_n + \varepsilon_n \end{cases}$$

Assim, para conseguir driblar tal situação e conseguir estimar o escore verdadeiro, algumas suposições foram feitas e, a partir delas, a TCT se ergueu, já estando bem axiomatizada no meio do século passado (Pasquali, 2018). Não adentraremos em tais considerações, uma vez que fogem ao objetivo aqui proposto. No entanto, tais suposições geram consequências para este modelo, das quais destacamos:

- i.* Os parâmetros dos itens considerados na TCT (dificuldade e discriminação) dependem da amostra de indivíduos submetidos à testagem. Porém, como apontou Thurstone, “um instrumento de medição não deve ser seriamente afetado em sua função de medição pelo objeto que está sendo medido” (Thurstone, 1928, p. 547, tradução nossa). Assim, se o teste for aplicado a amostras distintas, a TCT trabalharia de forma análoga a uma balança que mede a massa de duas pessoas de maneiras distintas simplesmente pelo fato de serem pessoas diferentes, mesmo que tenham a mesma massa, como se a massa do indivíduo fosse definida pela balança e não pelo atributo medido (Rabelo, 2013). Para se chegar a análises de proficiência a nível do item, é extremamente importante que não ocorram variações significativas nos resultados quando o teste é aplicado a amostras diferentes. Além disso, no contexto em que estamos, é desejável ter um teste que possa ser aplicado a grupos distintos, em épocas diferentes, mas cujos resultados possam ser comparáveis, e isso só é possível se o teste medir a proficiência dos estudantes da mesma forma, independentemente do grupo testado ou da época em que o teste foi aplicado;
- ii.* Na TCT, o resultado da medida de certo comportamento depende do teste utilizado. Isso significa que dois testes distintos que objetivam medir o mesmo comportamento de um mesmo indivíduo tendem a ter resultados diferentes. Para que tais testes produzam o mesmo resultado, é necessário que

sejam formas estritamente paralelas uma à outra, o que, na prática, é muito difícil de se obter;

- iii.* Não é possível avaliar de forma mais profunda, para além do acerto e do erro, o comportamento do indivíduo frente a cada item isoladamente. Isso ocorre porque a análise do desempenho do sujeito em cada um dos itens que compõem o teste é feita em função de todos os outros itens, por meio do escore total;
- iv.* Na TCT, “os testes são elaborados para avaliar maximamente os sujeitos de habilidades medianas, sendo, por isso, bem menos apropriados e válidos para avaliar sujeitos com habilidades superiores ou de pouca habilidade” (Pasquali, 2018, p. 13).

Buscando superar esses e outros impasses da Psicometria Clássica, diversos pesquisadores desenvolveram metodologias alternativas à TCT ao longo do século passado. Essas novas abordagens procuraram minimizar as limitações da Teoria Clássica dos Testes, propondo modelos que consideram características individuais dos sujeitos (traços latentes) e propriedades dos itens (parâmetros) de forma mais independente da amostra utilizada. Dessa maneira, novas perspectivas foram introduzidas na Psicometria, permitindo avanços significativos na precisão e na interpretação dos resultados dos testes psicológicos e educacionais.

2.2.3 **Psicometria Moderna**

Para superar as limitações da TCT, outras estratégias para aferir o desempenho em testes de aptidão foram desenvolvidas, culminando com o que mais tarde viria a se chamar de Teoria da Resposta ao Item ou Psicometria Moderna. Algumas dessas estratégias, segundo Pasquali (2003), procuravam remediar os quatro pontos de fragilidade da TCT destacados anteriormente, de forma que, nessa nova teoria:

- i.* os parâmetros dos itens não dependessem da amostra de testandos;
- ii.* a comparação dos escores obtidos em testes distintos não dependesse de formas estritamente paralelas;

- iii.* fosse estabelecido um modelo a nível do item e não a nível do escore total, o que permitiria a análise de um item sem depender dos demais;
- iv.* os testes avaliassem de igual modo os sujeitos de todos os níveis de habilidade, não só os de habilidade mediana.

Ainda de acordo com Pasquali (2003), a TRI não se interessa pelo desempenho relativo ao escore total de um indivíduo em um teste, mas sim pelo desempenho em cada um dos itens do teste. Mais do que isso, tal teoria objetiva determinar a probabilidade de um indivíduo com certa aptidão acertar dado item, bem como quais fatores afetam essa probabilidade. Por exemplo, retomando a segunda situação apresentada na introdução, a TRI não olharia para as cinco questões que Maria e João acertaram, mas sim para as dez questões do teste individualmente, relacionando o erro ou acerto de cada estudante em cada uma delas com a probabilidade de acerto, com as características do item e com o nível de habilidade de cada um deles.

Embora a base teórica da TRI já estivesse bem estruturada na década de 1960, seu avanço foi lento devido à falta de tecnologia capaz de executar eficientemente a matemática relacionada. Foi apenas a partir da década de 1980, com a relativa facilitação de acesso a computadores, com o desenvolvimento das linguagens de programação, com a criação de softwares dedicados a executar os algoritmos dos modelos matemáticos da TRI e com recursos computacionais específicos, que essa metodologia pôde ser, de fato, aplicada na prática para medir aptidões, além de ser aprimorada.

A TRI relaciona a qualidade do comportamento do sujeito (variável observável) com o seu respectivo traço latente (variável não observável) diante de um estímulo (variável também observável). De acordo com Pasquali (2018), o **traço latente** é uma característica psicológica ou habilidade que não pode ser observada diretamente, mas que influencia o comportamento ou as respostas de um indivíduo em determinadas situações, como em testes ou questionários, sendo o traço latente a causa e o desempenho no teste o efeito.

É razoável considerar que a distribuição da densidade de dado traço latente na população é aproximadamente normal, ou seja, poucos indivíduos têm nível baixo ou alto e a maior parte tem nível mediano de dado traço latente. Com essa

consideração, para medir um traço latente específico, o indivíduo testando recebe um estímulo observável – um item ou conjunto de itens relacionados a esse traço. Em resposta, o indivíduo manifesta um comportamento observável – a resposta aos itens – cuja qualidade indica o nível do traço latente, uma característica psicológica hipotética e não observável. Assim, a partir dos resultados de testes, infere-se a presença e a magnitude do traço latente em estudo.

No Brasil, a TRI teve uma contribuição significativa na análise dos resultados do Sistema de Avaliação da Educação Básica (Saeb), a partir de 1993, e do Exame Nacional do Ensino Médio (Enem), a partir de 2009, com Luiz Pasquali sendo pioneiro nesse tipo de análise no país, especialmente na estimativa da proficiência dos estudantes nessas avaliações. No estado do Espírito Santo, a TRI é utilizada nas avaliações do Programa de Avaliação da Educação Básica do Espírito Santo (PAE-BES) e, no Rio de Janeiro, nas avaliações do Sistema de Avaliação da Educação do Estado do Rio de Janeiro (SAERJ), ambas aplicadas a estudantes do 5º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio para medir a proficiência dos estudantes em Língua Portuguesa e Matemática.

2.3 Fundamentos da TRI

A TRI utiliza métodos matemático-estatísticos para estimar valores para os parâmetros dos itens. Uma vez conhecidos tais parâmetros, é possível relacioná-los com o nível do traço latente necessário para acertar certo item. Além disso, conhecendo o valor dos parâmetros dos itens que compõem um teste, é possível criar testes paralelos com itens de parâmetros semelhantes. Isso torna viável a comparação de resultados obtidos em diferentes amostras de indivíduos, em diversos períodos, mesmo utilizando testes distintos. Para que isso funcione, a TRI assume como axiomas a unidimensionalidade do desempenho e a independência local dos itens.

A **unidimensionalidade** sugere que, apesar do desempenho de um indivíduo em um item estar relacionado a vários traços latentes, existe um traço dominante na execução dessa tarefa, e é esse traço que será medido. Já a **independência local** dos itens postula que a resposta dada por um indivíduo a um item não é influenciada pelas respostas dadas por ele a outros itens do mesmo teste, o mesmo ocorrendo para as respostas de indivíduos diferentes (**independência individual**).

Assumindo esses dois axiomas, é possível expressar o desempenho provável de um indivíduo acertar um item em função do seu traço latente e em função dos parâmetros de cada item do teste. Conforme destaca Luiz Pasquali,

“A partir das respostas dadas pelo sujeito, isto é, analisando as suas respostas aos itens especificados, pode-se inferir sobre o traço latente do sujeito, hipotetizando relações entre as respostas observadas deste sujeito com o nível do seu traço latente. Estas relações podem ser expressas através de uma equação matemática que descreve a forma de função que estas relações assumem” (Pasquali, 2003, p. 83).

Para isso, ao longo dos anos, vários modelos foram (e ainda têm sido) desenvolvidos. No próximo capítulo, será apresentado o principal modelo da TRI utilizado em avaliações de proficiências cujos testes são constituídos por itens objetivos.

3

Conceitos estatísticos preliminares

Modelos estatísticos, como os aplicados na Teoria da Resposta ao Item, são sustentados por uma série de conceitos e ferramentas da matemática e da estatística. Dessa forma, para fixar tais conceitos que serão aqui utilizados e para tornar essa exposição mais fluida, neste capítulo serão apresentados, de forma concisa, algumas definições e resultados que serão utilizados nos próximos capítulos.

3.1

Modelos matemáticos e experimentos aleatórios

Modelos matemáticos são construídos com o objetivo de representar fenômenos reais por meio de relações formais entre variáveis. De forma geral, os modelos matemáticos podem ser determinísticos, em que as condições sob as quais um experimento é executado determinam o resultado do experimento, ou não-determinísticos, em que as condições da experimentação determinam somente o comportamento provável dos possíveis resultados. Mais que isso: mantendo as mesmas condições, aplica-se um modelo determinístico ao experimento em que é possível repeti-lo e esperar obter o mesmo resultado. São os experimentos determinísticos. Não sendo possível repetir o experimento e obter o mesmo resultado, mesmo que se conheçam os possíveis resultados, estamos diante de um experimento aleatório.

Em relação à aplicação de um teste, estamos diante de um experimento que é difícil de ser repetido sob as mesmas condições, tendo em vista que um indivíduo que é submetido a um mesmo teste não é “o mesmo” daquele anterior ao contato com a prova.

[...] os, sujeitos, ao tomarem uma forma paralela, nunca serão exatamente os mesmos, pois há os problemas de maturação [...], tais como a aprendizagem, o cansaço, a motivação e outros, que mudam da aplicação de uma forma para outra, inclusive diferencialmente para diferentes sujeitos, tornando a comparação entre as duas formas não mais paralelas (Pasquali, 2018, p. 13).

Porém, há componentes aleatórios na aplicação de um teste. Ao responder um teste, cada indivíduo tem seu comportamento, manifestado pelas respostas dadas, influenciado tanto pelo seu nível de proficiência (traço latente), quanto por fatores internos intrínsecos (como os citados acima) ou externos (ambientais) a ele, que podem ter origem aleatória. Além disso, como os testes que serão tratados aqui (testes de proficiência) geralmente são objetivos, mesmo um indivíduo com alta habilidade pode errar um item fácil, ao passo que um indivíduo com baixa proficiência pode acertar um item difícil por sorte. Assim, a aplicação de um teste envolve componentes aleatórios provenientes tanto dos indivíduos (ocorrência ou existência de fatores internos e externos), quanto do ato de responder ao teste (resultado da resposta e acerto ao acaso), tornando essencial o seu tratamento como um experimento aleatório, tomando o cuidado de fazer considerações ou usar algum procedimento que permita hipotetizar condições para isso.

3.2 Variáveis aleatórias

Seja S o conjunto de todos os resultados de um experimento aleatório ε . Considere também uma função $P:S \rightarrow \mathbb{R}$ que associa a cada resultado $s \in S$ desse experimento um número real $P(s) \geq 0$ tal que $\sum_{s \in S} P(s) = 1$. Conforme Scheinerman (2011), chamamos a função P de probabilidade do resultado do experimento aleatório ε e o par (S, P) de espaço amostral. Além disso, um subconjunto $A \subseteq S$ é chamado de evento de ε , sendo $P(A) = \sum_{a \in A} P(a)$ a probabilidade do evento A .

Quando S é infinito enumerável ou não-enumerável, o espaço amostral (S, P) é dito infinito. Já quando S é finito, (S, P) é chamado espaço amostral finito e, caso os k resultados possíveis $s \in S$ sejam igualmente verossímeis, ou seja, sejam igualmente prováveis, temos $P(s) = \frac{1}{k}$. Nesse caso, dizemos que (S, P) é um espaço amostral equiprovável.

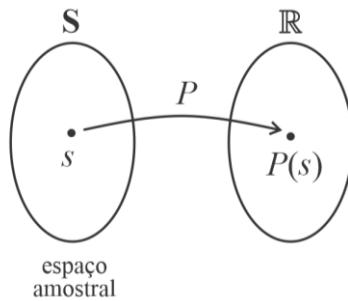


Figura 3.1 – Espaço amostral
Fonte: autoria própria.

Embora naturalmente haja interesse nos resultados do espaço amostral, muitas vezes estamos interessados em resultados que derivam do espaço amostral, mas que não são necessariamente eventos do espaço amostral. Considere o espaço amostral (S, P) . Uma função $X : S \rightarrow V$ que associa os possíveis resultados de um evento aleatório aos elementos de algum conjunto V é chamada de variável aleatória. Perceba que os valores de uma variável aleatória não estão restritos a valores reais. Assim, podemos ter $V = \mathbb{R}$, $V = \mathbb{R}^3$, $V = \mathbb{R}^n$, por exemplo, ou qualquer outro conjunto que descreva os possíveis resultados de interesse derivados do espaço amostral (S, P) .

A expressão variável aleatória é, talvez, uma das mais inadequadas em toda a matemática. Uma variável aleatória não é aleatória nem uma variável! É uma função definida em um espaço amostral. As variáveis aleatórias são utilizadas para modelar grandezas cujo valor é aleatório (Scheinerman, 2011, p. 311).

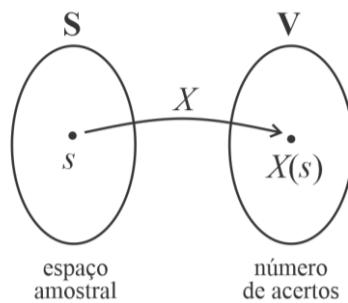


Figura 3.2 – Variável aleatória
Fonte: autoria própria.

Quando V é finito ou infinito enumerável, ou seja, quando a variável aleatória assume um número de valores que é finito ou infinito enumerável, ela é chamada de variável aleatória discreta. Agora, quando a variável aleatória assume valores em um conjunto infinito não enumerável, ela é chamada de variável aleatória contínua. No caso da aplicação de um teste de proficiência, temos, a princípio, uma variável aleatória discreta a considerar: a variável aleatória resultado da resposta a um item, denotada aqui por X , em que $S = \{c, e\}$ e $V = \{1, 0\}$, com $X(c) = 1$ e $X(e) = 0$, uma vez que as respostas (acertos ou erros) são tratadas como variáveis aleatórias condicionadas aos parâmetros do item e à proficiência latente do sujeito.

3.3 Distribuição de probabilidade

A descrição das probabilidades associadas aos possíveis valores de uma variável aleatória é chamada de distribuição de probabilidade. Essa descrição pode ser feita por meio de um diagrama, de uma tabela ou, mais comumente, por uma função, cuja caracterização depende se a variável aleatória é discreta ou contínua.

Chamamos de função massa de probabilidade da variável aleatória discreta $X: S \rightarrow V$ à função $p: V \rightarrow \mathbb{R}$ que associa a probabilidade de ocorrência $p(x_i)$ de cada valor $X(s) = x_i \in V$, com $s \in S$, que satisfaça as condições:

- i. $p(x_i) \geq 0$ para todo i ,
- ii. $\sum_{i=1}^n p(x_i) = 1$, com $i \in I_n = \{0, 1, 2, \dots, n\}$, se V for finito,
- iii. $\sum_{i=1}^{\infty} p(x_i) = 1$, com $i \in \mathbb{N}$, se V for infinito enumerável.

A função massa de probabilidade associa probabilidade a valores isolados da variável aleatória.

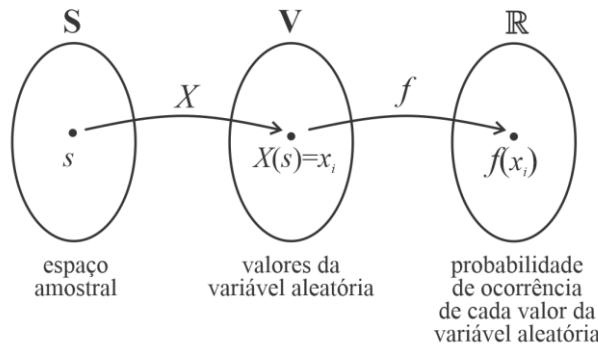


Figura 3.3 – Função massa de probabilidade
Fonte: autoria própria.

Chamamos de função densidade de probabilidade (FDP) da variável aleatória contínua $X : S \rightarrow V$ à função $f : V \rightarrow \mathbb{R}$ que satisfaça:

- i. $f(x) \geq 0$ para todo $x \in V$,
- ii. $\int_{-\infty}^{\infty} f(x) dx = 1$, com $f(x) = 0$ para todo $x \notin V$,
- iii. $P(a \leq X \leq b) = \int_a^b f(x) dx$ para quaisquer a, b , com $-\infty < a < b < \infty$.

A função densidade de probabilidade não associa probabilidade a valores isolados da variável aleatória, uma vez que, para certo $x_0 \in V$, teríamos

$$P(X = x_0) = \int_{x_0}^{x_0} f(x) dx = 0 \quad ,$$

ocorrendo o mesmo para todo $x \in V$, apesar da soma dessas probabilidades ser 1. Da mesma forma, $f(x)$ não representa a probabilidade alguma, produzindo probabilidades apenas quando for integrada entre dois limites. Além disso, pelo menos teoricamente, qualquer função definida em V e que satisfaça as três condições definirá uma variável aleatória contínua.

3.4 Modelos probabilísticos

O estudo de uma variável aleatória tem como objetivo compreender seu comportamento por meio das ocorrências possíveis. Por exemplo, as frequências relati-

vas servem como estimativas de probabilidade. Com base em suposições apropriadas, modelos teóricos, chamados modelos probabilísticos ou modelos de distribuição e frequência, podem reproduzir padrões de distribuição de frequências observados, mesmo sem acesso direto ao fenômeno. Algumas variáveis aleatórias se ajustam bem a contextos práticos, sendo fundamentais para construir modelos aplicáveis e estimar parâmetros com precisão (Bussab; Morettin, 2010).

A seguir, destacaremos brevemente alguns desses modelos que serão utilizados ao longo dessa exposição.

3.4.1 Distribuição de Bernoulli

Um **experimento de Bernoulli** é um experimento aleatório que possui apenas dois resultados possíveis, geralmente chamados de sucesso e fracasso, com probabilidades fixas. Assim, a variável aleatória X , resultado do experimento de Bernoulli, chamada variável de Bernoulli, é uma variável aleatória discreta dicotômica que assume apenas dois valores, geralmente 0 (fracasso) e 1 (sucesso), com função massa de probabilidade $p : V = \{0, 1\} \rightarrow \mathbb{R}$ tal que

$$\begin{cases} p(1) = P(X = 1) = p \\ p(0) = P(X = 0) = 1 - p \end{cases}$$

em que $P(X = 1) = p$ é a probabilidade de sucesso, com $0 < p < 1$.

No contexto dos testes de proficiência com itens objetivos compostos por cinco opções de respostas, sendo apenas uma a correta, a distribuição de Bernoulli pode modelar a probabilidade de acerto da resposta dada a um item por um indivíduo. Nesse caso, considerando que a resposta pelo indivíduo venha apenas da escolha ao acaso de uma das alternativas do item, a probabilidade de acerto é $p(1) = P(X = 1) = 0,2$. A figura a seguir representa essa situação, para a aplicação de um item objetivo a um indivíduo.

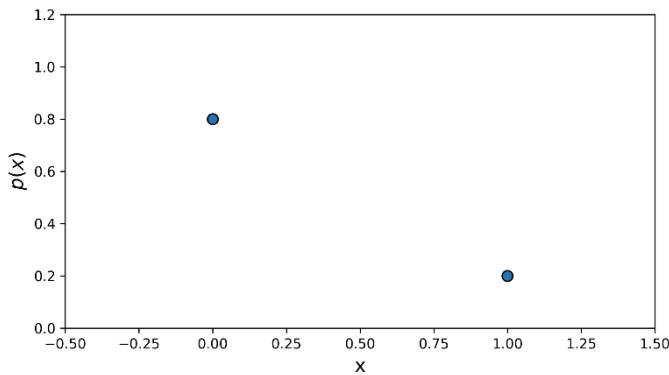


Gráfico 3.1 – Função massa de probabilidade da distribuição de Bernoulli
Fonte: autoria própria

3.4.2 Distribuição Binomial

Ao repetir n vezes um experimento de Bernoulli, supondo que as repetições sejam independentes, ou seja, o resultado de um experimento não influencia o resultado de outro, e sendo sempre p , com $0 < p < 1$, a probabilidade de sucesso em cada repetição, teremos uma sequência de k sucessos e $n - k$ fracassos ou, alternativamente, k “uns” e $n - k$ zeros. Tal experimento é chamado de **experimento binomial** em que a variável aleatória X , número de sucessos em n ensaios de Bernoulli ou número de sucessos em um experimento binomial, é uma variável aleatória discreta, chamada de variável binomial, e a probabilidade de se obter k sucessos é expressa pela função massa de probabilidade $p: V = \{1, 2, \dots, n\} \rightarrow \mathbb{R}$ definida por

$$P(X = k | n, p) = \binom{n}{k} p^k q^{n-k},$$

ou $P(X = k) = \binom{n}{k} p^k q^{n-k},$

ou, simplesmente, $P(k) = \binom{n}{k} p^k q^{n-k}$, com $k = 0, 1, \dots, n$.

Perceba que em um experimento binomial não estamos interessados na ordem em que os sucessos ocorrem, mas sim no número total de sucessos (Bussab; Morettin, 2010).

No contexto dos testes de proficiência com itens objetivos compostos de cinco opções de respostas, sendo apenas uma a correta, a distribuição binomial pode

modelar a probabilidade de ocorrência de um certo número de acertos na aplicação de um teste com um item objetivo a um grupo de n indivíduos. Nesse caso, considerando que a resposta de cada indivíduo venha apenas da escolha ao acaso de uma das alternativas do item, a probabilidade de acerto de cada ensaio de Bernoulli é $p(1) = 0,2$. A figura a seguir, representa a distribuição de probabilidade obtida na aplicação de um teste composto por um item objetivo a um grupo de 10 indivíduos usando o modelo binomial.

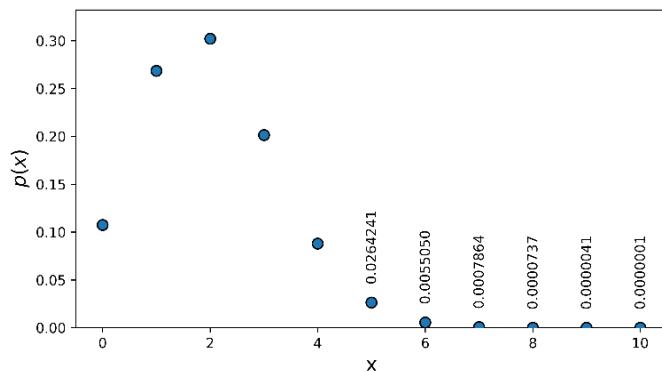


Gráfico 3.2 – Função massa de probabilidade da distribuição Binomial
Fonte: autoria própria

Podemos observar, por exemplo, que a probabilidade de haver dois acertos (dois indivíduos que acertem o item) é de 30%, enquanto a probabilidade de haver 5 acertos está abaixo de 10%.

3.4.3 Distribuição Normal

Uma variável aleatória contínua X tem distribuição normal de probabilidade se sua Função Densidade de Probabilidade (FDP) é dada por

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

em que os parâmetros são a média μ , com $-\infty < \mu < \infty$, e a variância σ^2 , com $\sigma^2 > 0$. Quando $\mu = 0$ e $\sigma^2 = 1$, obtemos

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}$$

que é a distribuição normal padrão. O gráfico da distribuição normal padrão de uma variável aleatória contínua, chamada curva gaussiana, está representado na figura a seguir.

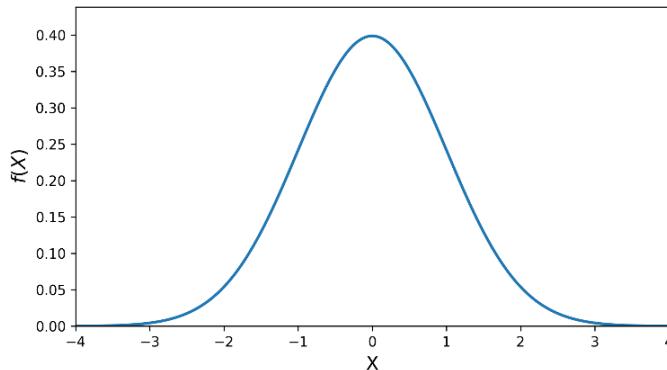


Gráfico 3.3 – Função densidade de probabilidade

Fonte: autoria própria

Uma distribuição normal qualquer pode ser ajustada para a distribuição normal padrão fazendo a mudança de variável $Z = \frac{X - \mu}{\sigma}$.

A suposição de que uma aptidão qualquer se distribua normalmente em uma população, no sentido de que poucos indivíduos possuem um nível de aptidão muito baixo ou muito alto e a maioria tem nível mediano, é largamente adotada em modelagem estatística em psicometria, especialmente na TRI.

3.5 Função de distribuição acumulada

Uma função de distribuição acumulada (FDA) $F: \mathbb{R} \rightarrow \mathbb{R}$ da variável aleatória contínua X é definida como $F(x) = P(X \leq x)$, sendo

i. $F(x) = \sum_i p(x_i)$ para todo $i = 1, 2, \dots$ tal que $x_i \leq x$, com p sendo uma

função massa de probabilidade, se X for uma variável aleatória discreta;

ii. $F(x) = \int_{-\infty}^x f(t) dt$, com f sendo uma função densidade de probabilidade, se X for uma variável aleatória contínua;

iii. $0 \leq F(x) \leq 1$;

iv. se $a < b$, então $F(a) < F(b)$, ou seja, F é não-decrescente,

v. $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$;

vi. $P(a \leq x \leq b) = F(b) - F(a)$.

A representação gráfica de uma função de distribuição acumulada é chamada de ogiva, às vezes estendendo-se essa nomenclatura à própria FDA de uma variável aleatória. A ogiva mostra de forma mais clara como os dados se acumulam ao longo de um intervalo. Se X for uma variável aleatória discreta, então sua ogiva possui saltos, sendo uma função em degraus, como, por exemplo, a ogiva da distribuição de Bernoulli que modela probabilidade acerto da resposta dada a um item de um teste dicotômico, representada na figura a seguir.

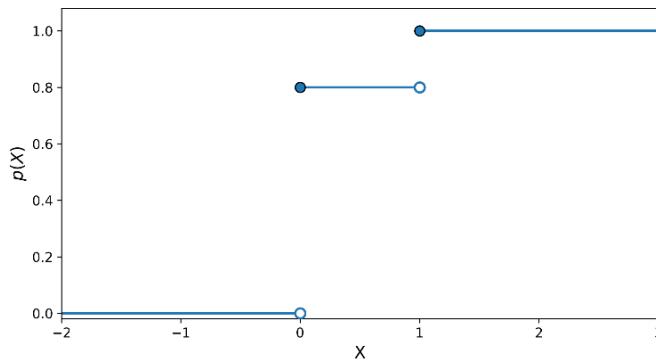


Gráfico 3.4 – Ogiva da distribuição de Bernoulli

Fonte: autoria própria

Caso a variável aleatória X seja contínua, então sua FDA será uma função contínua para todo $x \in \mathbb{R}$ e sua ogiva não possui saltos, como ocorre com a função de distribuição acumulada de uma variável contínua com distribuição normal, cujo gráfico (ou a própria FDA) é chamado de ogiva normal.

3.5.1

A ogiva normal padrão

A função de distribuição acumulada $F(x)$ de uma variável contínua X com distribuição normal padrão ($\mu = 0$ e $\sigma^2 = 1$), denotada por $\Phi(x)$, é dada por

$$\begin{aligned}
 \Phi(x) &= \int_{-\infty}^x f(z) dz \\
 &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz
 \end{aligned}$$

cujo gráfico, chamado de ogiva normal padrão, está representado na figura a seguir.

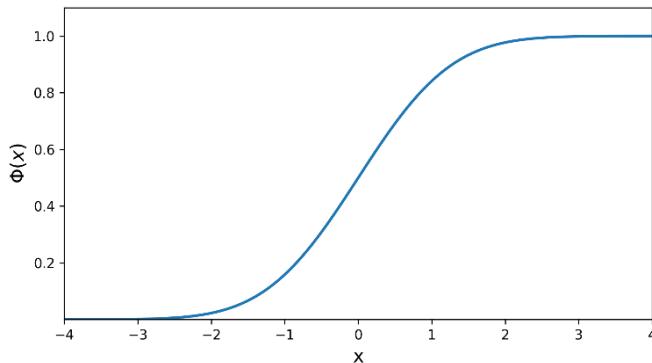


Gráfico 3.5 – Ogiva normal padrão

Fonte: autoria própria

Os modelos pioneiros da TRI faziam uso de Φ para relacionar a probabilidade de um indivíduo com aptidão θ acerta um item. No entanto, a manipulação de Φ exigia um gasto computacional muito grande num período com recursos tecnológicos escassos para esse processamento. Assim, a função Φ , que gera a ogiva normal, foi preferida por outra mais fácil de trabalhar computacionalmente falando: a função logística.

3.6 Função logística

Uma função logística é uma função sigmoide (função cujo gráfico é uma curva em forma de “S”) dada genericamente por

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

em que L é o valor máximo de f , x_0 é o valor da abscissa do ponto de simetria da curva sigmoide e k é um parâmetro relacionado com a inclinação da

curva chamado taxa de crescimento logístico. Quando $k=1$, $x_0=0$ e $L=1$, obtemos

$$f(x) = \frac{1}{1+e^{-x}} \text{ ou } f(x) = \frac{1}{1+e^{-x}} \cdot \frac{e^x}{e^x} = \frac{e^x}{1+e^x}$$

que é a função logística padrão, cujo gráfico está representado a seguir.

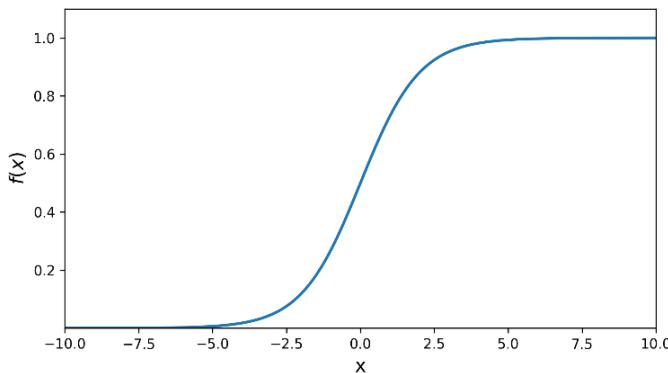


Gráfico 3.6 – Função logística padrão
Fonte: autoria própria

A função logística padrão é uma curva sigmoidal monótona crescente, limitada entre 0 e 1, com simetria em torno do ponto $(0; 0,5)$. Sua derivada

$$f'(x) = \frac{e^{-x}}{1+2e^{-x}+e^{-2x}}$$

sempre positiva e simétrica, pode ser usada para reconstruir a função original por integração, o que destaca seu caráter estrutural autogerador.

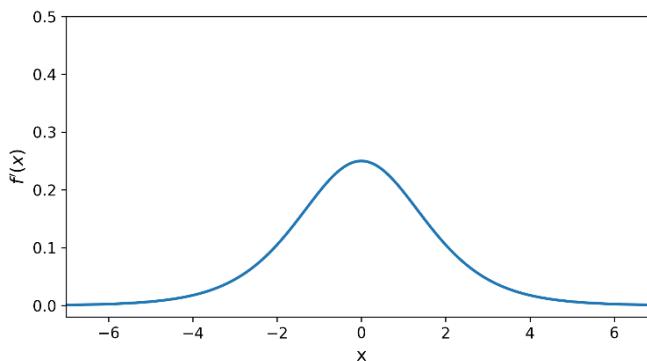


Gráfico 3.7 – Gráfico da derivada da função logística padrão
Fonte: autoria própria

Além disso, essa derivada está relacionada a distribuições de probabilidade simétricas, conferindo à função uma interpretação estatística. A função também é invariante sob transformações afins, o que permite ajustar sua forma e posição sem perder suas propriedades sigmoides (John A. Drakopoulos, 1995).

Perceba que o gráfico da função logística padrão $f(x)$ se parece muito com o da ogiva normal $\Phi(x)$. No entanto, quando plotamos ambas as curvas num mesmo gráfico, vemos diferença. Veja:

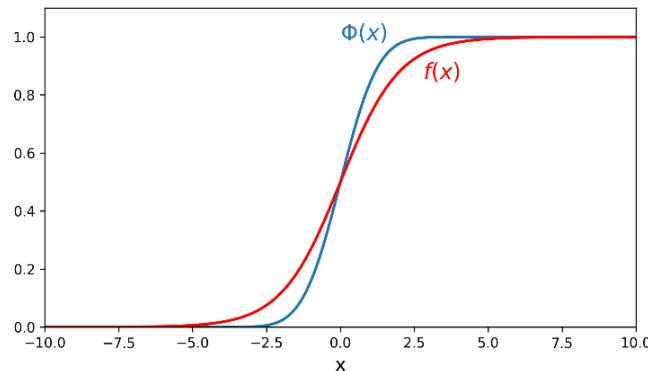


Gráfico 3.8 – Ogiva normal padrão e sigmoide da função logística
Fonte: autoria própria.

Por ser mais fácil de ser processada, computacionalmente falando, e por se ajustar melhor aos dados, os modelos probabilísticos modernos da TRI, inicialmente desenvolvidos utilizando a função ogiva normal, fazem uso da função logística (de Andrade; Tavares; da Cunha Valle, 2000). Essa classe de modelos é chamada de modelos logísticos. No entanto, para ajustar a curva logística à curva da ogiva normal, foi adicionado o parâmetro D à expressão da curva logística, passando a ser descrita por

$$f(x) = \frac{L}{1 + e^{-Dk(x-x_0)}}$$

Esse ajuste, historicamente, foi inicialmente feito de forma empírica, chegando ao valor $D = 1,7$. A figura a seguir mostra esse ajuste para diferentes valores de D .

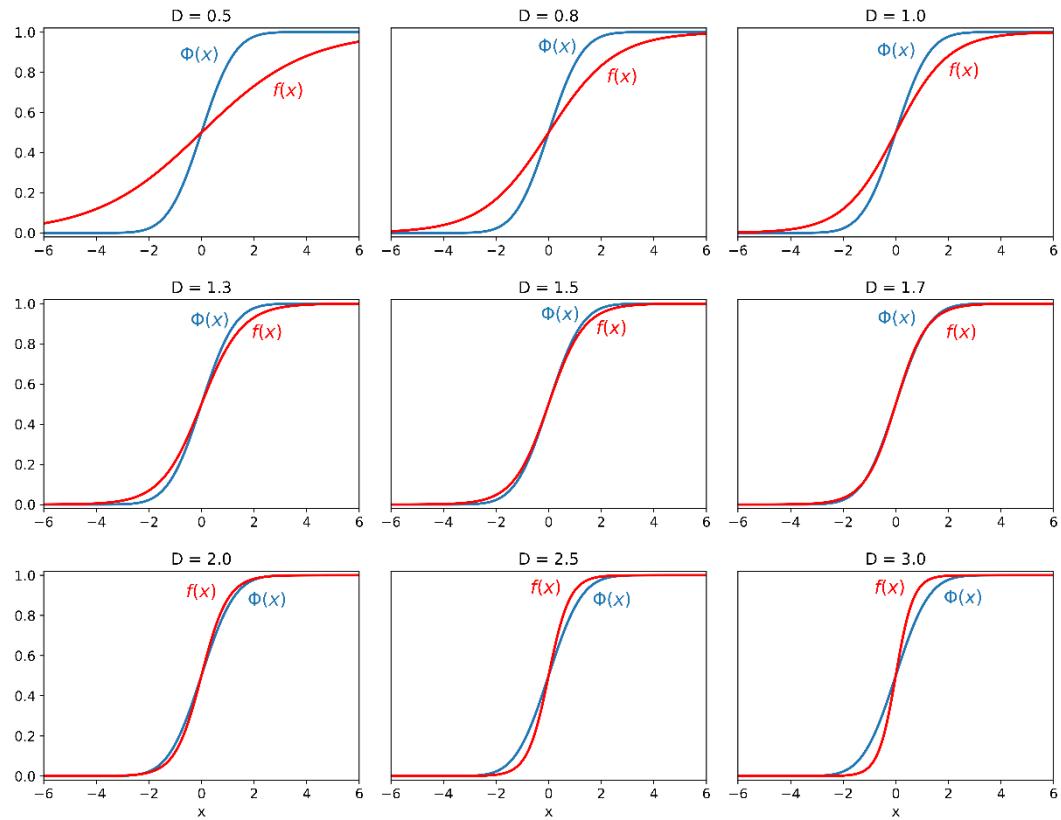


Gráfico 3.9 – Ajuste da curva logística padrão à ogiva padrão normal

Fonte: autoria própria.

4

Modelos logísticos unidimensionais

A TRI utiliza modelos matemáticos para estimar a probabilidade de acerto em um item em função dos parâmetros do item e da aptidão (traço latente) dos respondentes. Assim, “deseja-se descobrir qual o valor do traço latente (da habilidade) do indivíduo que melhor explica o acerto ou o erro em cada item individualmente” (Rabelo, 2013, p. 129). Existem uma série de modelos da TRI. Dentre eles, para os testes de proficiência, destacam-se os modelos logísticos, cuja estrutura vem das funções logísticas. A função logística garante que a probabilidade de acerto varie suavemente entre 0 e 1 conforme a habilidade do sujeito. Por sua vez, os modelos logísticos unidimensionais assumem que essa habilidade é representada por um único traço latente, e que cada item avalia predominantemente essa dimensão. Nesse capítulo, serão apresentados alguns desses modelos da família logística unidimensional da TRI para estimar a proficiência dos testandos em testes de proficiência compostos por itens dicotômicos.

4.1

Matriz de padrão de resposta

Um item dicotômico apresenta apenas dois valores mutuamente exclusivos para o resultado da resposta dada por um indivíduo a esse item. No nosso caso, o resultado dessas respostas pode assumir apenas os valores “certo” ou “errado”. Assim, seja X_i a variável aleatória discreta que representa os possíveis resultados para as respostas dadas ao item i , sendo

$$X_i = \begin{cases} 0, & \text{se a resposta estiver errada} \\ 1, & \text{se a resposta estiver correta} \end{cases}.$$

Dessa forma, após a aplicação de um teste com n itens objetivos a um grupo de m indivíduos, os registros de acertos e erros são feitos para cada participante como uma sequência de zeros e uns. Assim, obtemos o vetor aleatório $X_j = (X_{j1}, X_{j2}, \dots, X_{ji}, \dots, X_{jn})$, também representado por $X_{j1}X_{j2} \dots X_{ji} \dots X_{jn}$ de

forma mais simples, em que cada elemento X_{ji} é a variável aleatória resultado da resposta do indivíduo j ao item i , com $i, j, m, n \in \mathbb{N}^*$, $1 \leq i \leq n$, e $1 \leq j \leq m$.

Agora, considere u_{ji} o resultado (certo o errado) que o indivíduo j de fato obteve com sua resposta (alternativa marcada) dada ao item i do teste, com

$$u_{ji} = \begin{cases} 0, & \text{se a resposta dada pelo indivíduo } j \text{ ao item } i \text{ estiver errada} \\ 1, & \text{se a resposta dada pelo indivíduo } j \text{ ao item } i \text{ estiver correta,} \end{cases}$$

e considere também $U_j = (u_{j1}, u_{j2}, \dots, u_{ji}, \dots, u_{jn})$ o vetor cujos elementos são os resultados que, de fato, o indivíduo j obteve com suas respostas dadas aos itens $1, 2, \dots, i, \dots, n$, chamado de **vetor padrão de resposta** do indivíduo j , de forma que

$$X_j = U_j \Leftrightarrow X_{j1} = u_{j1}, \dots, X_{ji} = u_{ji}, \dots, X_{jn} = u_{jn} .$$

Além disso, a matriz

$$U_{m \times n} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_j \\ \vdots \\ U_m \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1i} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2i} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{j1} & u_{j2} & \cdots & u_{ji} & \cdots & u_{jn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mj} & \cdots & u_{mn} \end{bmatrix} ,$$

cujas linhas são os vetores U_j , é chamada de **matriz de padrão de resposta**, que representa o conjunto de respostas obtidas a partir da aplicação do teste, considerada um amostra de respostas da população que pode ser submetida a esse teste. Nessa matriz, a soma dos elementos da linha j (ou dos elementos do vetor U_j) representa o escore total obtido pelo indivíduo j . Por exemplo, uma saída típica dessa matriz U para um teste aplicado a cinco estudantes ($m=5$) com dez itens ($n=10$) seria

$$U_{5 \times 10} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

em que, nesse caso, todos os cinco indivíduos tiveram um escore de 6 acertos, uma vez que, para cada indivíduo j , $\sum_{i=1}^{10} u_{ji} = 6$. Nessa matriz, o indivíduo $j=3$ errou o item $i=4$ ($u_{34} = 0$), o indivíduo $j=4$ acertou o item $i=7$ ($u_{47} = 1$) e o padrão de resposta do indivíduo $j=5$ é $X_5 = (1, 0, 1, 1, 0, 0, 0, 1, 1, 1)$ ou $X_5 = 1011000111$.

Por fim, sendo $X = (X_1, X_2, \dots, X_j, \dots, X_m)$ a matriz aleatória em que cada entrada X_j é o vetor aleatório $X_j = (X_{j1}, X_{j2}, \dots, X_{ji}, \dots, X_{jn})^T$, se $X = U$, temos

$$\begin{aligned}
 X_{m \times n} = U_{m \times n} &\Leftrightarrow \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_j \\ \vdots \\ U_m \end{bmatrix} \\
 &\Leftrightarrow \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1i} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2i} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{ji} & \cdots & X_{jn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mj} & \cdots & X_{mn} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1i} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2i} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{j1} & u_{j2} & \cdots & u_{ji} & \cdots & u_{jn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mj} & \cdots & u_{mn} \end{bmatrix} \\
 &\Leftrightarrow X_{ji} = u_{ji}
 \end{aligned}$$

4.2 Modelos Logísticos

Diferente da TCT, a TRI avalia não só a quantidade de itens acertados, mas também as razões por trás dos acertos ou erros. Em essência, a TRI determina o nível do traço latente necessário para que uma pessoa acerte um item específico. Isso é feito por meio de uma função que relaciona o nível do traço latente θ (variável não observável) e os parâmetros a, b, c, d, \dots considerados do item (variável observável) com a probabilidade de acerto do item, de forma que, conhecendo os parâmetros dos itens, é possível estimar o nível do traço latente do indivíduo, e vice-versa. Isso é possível graças ao axioma da independência local, que permite que tais

parâmetros sejam obtidos individualmente para cada item, bem como permite que a proficiência de cada indivíduo seja também estimada individualmente.

Nessas condições, dado um item que avalia um certo traço latente, quanto maior o nível desse traço em um indivíduo, maior a probabilidade de ele acertar o item e vice-versa. Assim, parece razoável supor que uma função que relaciona a probabilidade de um indivíduo acertar um item com o seu respectivo nível de aptidão é uma função monótona não decrescente. Então, a probabilidade P que um indivíduo tem de acertar um item ($X = 1$) depende do nível de aptidão θ do indivíduo e dos parâmetros a, b, c, d, \dots considerados do item, ou seja,

$$P(X = 1) = f(\theta, a, b, c, d, \dots)$$

Vários modelos de funções probabilísticas envolvendo tais variáveis foram propostos a partir da segunda metade do século XX com a finalidade de mensurar o desempenho de indivíduos em testes de proficiência. O modelo atualmente utilizado nos principais testes de proficiência constituídos por itens objetivos dicotômicos (resposta certa ou errada) é o modelo logístico de 3 parâmetros, desenvolvido pelo estadunidense Frederic Lord (1912-2000) e posteriormente aperfeiçoado pelo também americano Allan Birnbaum (1923-1976).

Lord (1952) foi o primeiro a desenvolver o modelo unidimensional de 2 parâmetros, baseado na distribuição normal acumulada (ogiva normal). Após algumas aplicações desse modelo, o próprio Lord sentiu a necessidade da incorporação de um parâmetro que tratasse do problema do acerto casual. Assim, surgiu o modelo de 3 parâmetros. Anos mais tarde, Birnbaum (1968) substituiu, em ambos os modelos, a função ogiva normal pela função logística, matematicamente mais conveniente, pois é uma função explícita dos parâmetros do item e de habilidade e não envolve integração. Independentemente do trabalho de Lord, Rasch (1960) propôs o modelo unidimensional de 1 parâmetro, expresso também como modelo de ogiva normal e, também mais tarde, descrito por um modelo logístico por Wright (1968) (de Andrade; Tavares; da Cunha Valle, 2000, p. 4).

Assim, os modelos logísticos de um parâmetro (ML1P), de dois parâmetros (ML2P) e de três parâmetros (ML3P) nasceram de modelos que originalmente usavam a função ogiva normal, sendo mais tarde atualizados para uma versão que usa

como base a função logística. A seguir, vamos apresentar cada um deles, não em ordem cronológica de desenvolvimento, mas sim partindo do mais geral deles.

4.2.1

Modelo Logístico de 3 parâmetros da TRI (Modelo de Birnbaum)

O modelo logístico de três parâmetros considera a probabilidade P que um indivíduo tem de acertar um item ($X = 1$) como dependente do seu nível de aptidão θ (traço latente) e de três parâmetros do item, a saber: discriminação (a), dificuldade (b) e acerto ao acaso (c). Assim, o ML3P define $P(X = 1) = f(\theta, a, b, c)$ como

$$P_{\text{ML3P}} : P(X = 1 | \theta, a, b, c) = c + (1 - c) \frac{1}{1 + e^{-Da(\theta-b)}} \quad (4.1)$$

em que $P(X = 1 | \theta, a, b, c)$ é a probabilidade de que a resposta dada pelo indivíduo de aptidão θ ao item de parâmetro de discriminação a , de dificuldade b e de acerto ao acaso c , esteja correta.

O parâmetro de discriminação a indica a capacidade que um item tem de diferenciar indivíduos com proficiências diferentes, enquanto o parâmetro de dificuldade b indica o mínimo de proficiência que um sujeito precisa ter para uma chance de 50% de acertar um item. Já o parâmetro de acerto ao acaso c indica o quanto provável uma resposta certa pode ser dada ao item por acaso ou, de outra forma, ele representa o quanto provável um indivíduo com baixa habilidade pode responder corretamente ao item (Rabelo, 2013).

Considerando a distribuição de θ na população próxima da normal padrão, temos, teoricamente, $-\infty < \theta, a, b < \infty$. Porém, sabendo que, nesse caso, 99,74% dos indivíduos terão θ dentro do intervalo entre -3 e $+3$, nas aplicações, tipicamente temos $-3 < \theta < 3$, $0 < a < 3$ e $-3 < b < 3$. Já para o parâmetro de acerto ao acaso, temos $0 < c < 1$ (Pasquali, 2018).

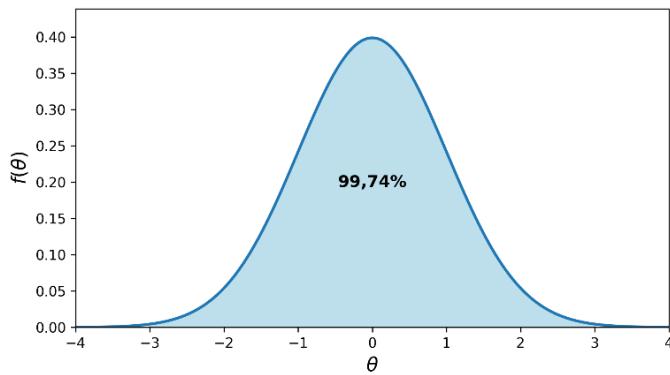


Gráfico 4.1 – Distribuição normal padrão das aptidões na população

Fonte: autoria própria

Além disso, “ D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes aos da função ogiva normal” (de Andrade; Tavares; da Cunha Valle, 2000, p. 10).

4.2.2

Modelo Logístico de 2 parâmetros da TRI (Modelo de Lord)

O modelo logístico de dois parâmetros considera a probabilidade P que um indivíduo tem de acertar um item ($X = 1$) como dependente do seu nível de aptidão θ (traço latente) e dos parâmetros discriminação (a) e dificuldade (b) do item. Dessa forma, o ML2P define $P(X = 1) = f(\theta, a, b)$ como

$$P_{\text{ML2P}} : P(X = 1 | \theta, a, b) = \frac{1}{1 + e^{-Da(\theta-b)}} \quad (4.2)$$

Perceba que o ML2P é um caso particular do ML3P quando $c = 0$ e que

$$P_{\text{ML3P}} = c + (1 - c) \cdot P_{\text{ML2P}}$$

4.2.3

Modelo Logístico de 1 parâmetro da TRI (Modelo de Rasch)

O modelo logístico de um parâmetro (ML1P) considera a probabilidade P que um indivíduo tem de acertar um item ($X = 1$) como dependente do seu nível de aptidão θ (traço latente) e do parâmetro de dificuldade (b) do item. A ideia por trás desse modelo é a suposição de que a chance (*odds*) de um sujeito acertar um item é expressa pela razão entre a probabilidade de sucesso e a de fracasso

$1 - P(X = 1)$, sendo o logaritmo dessa razão a diferença entre a habilidade θ do indivíduo e a dificuldade b do item (Pasquali, 2018). Assim, temos

$$\begin{aligned} \ln \frac{P(X = 1)}{1 - P(X = 1)} &= \theta - b \Rightarrow e^{\theta - b} = \frac{P(X = 1)}{1 - P(X = 1)} \\ &\Rightarrow (1 + e^{\theta - b}) \cdot P(X = 1) = e^{\theta - b} \\ &\Rightarrow P(X = 1) = \frac{e^{\theta - b}}{1 + e^{\theta - b}} \\ &\Rightarrow P(X = 1) = \frac{1}{1 + e^{-(\theta - b)}} \end{aligned}$$

Dessa forma, aplicando o ajuste de escala para a ogiva normal, o ML1P define $P(X = 1) = f(\theta, b)$ como

$$P_{\text{ML1P}} : P(X = 1 | \theta, b) = \frac{1}{1 + e^{-D(\theta - b)}} \quad (4.3)$$

Há alguns impasses em relação à classificação do ML1P. Alguns autores, como Baker, sugerem que esse modelo seria a particularização do modelo ML2P com $a = 1$ para todos os itens, conforme idealizado inicialmente por Rasch, o primeiro a publicar um modelo logístico de 1 parâmetro.

Sob este modelo, o parâmetro de discriminação do modelo logístico de dois parâmetros é fixado em um valor de $a = 1,0$ para todos os itens; apenas o parâmetro de dificuldade pode assumir diferentes valores. Por causa disso, o modelo de Rasch é frequentemente referido como o modelo logístico de um parâmetro (Baker, 2001, p. 25, tradução nossa).

Para outros, como Pasquali, essa classificação é mais flexível, sendo uma particularização do ML2P com $a = 1$ ou outro valor diferente de zero, mas constante para todos os itens, com $c = 0$ sendo o acerto ao acaso (Pasquali, 2018, p. 33). Nesse sentido, o ML1P seria um modelo distinto do modelo de Rasch. Aqui seguiremos as ideias de Baker, considerando o ML1P como uma particularização do modelo ML2P com $a = 1$ e $c = 0$, assim como idealizado originalmente por Rasch.

4.3

O fator de escala D

Conforme anteriormente discutido, o valor 1,7 para o fator de escala D foi historicamente obtido de forma empírica. Porém, Haley (1952), no contexto da modelagem estatística da relação dose-resposta, que é central em estudos de toxicologia, farmacologia e bioensaios, ao realizar uma comparação analítica entre a função logística e a função probit (baseada na distribuição normal acumulada), encontra formalmente 1,702 como o valor de D que minimiza a diferença entre essas duas curvas. Nota-se, então, que esse fator é uma aproximação específica que alinha a função logística à função normal em suas formas padronizadas, mas não necessariamente a qualquer outra forma variada dessas funções, no sentido de que se os valores de μ ou σ^2 forem diferentes de $\mu = 0$ e $\sigma^2 = 1$ ou houver modificações estruturais nessas funções, então pode ser que a diferença entre as duas funções não será reduzida pelo fator $D = 1,702$.

Tendo isso em vista, o valor $D = 1,702$ é amplamente adotado nos modelos logísticos da Teoria de Resposta ao Item (TRI) como uma constante de escala que aproxima a curva logística da função de distribuição acumulada da normal padrão. Essa prática de aproximar as duas curvas é mais comum nos modelos de dois parâmetros, uma vez que a curva resultante é simétrica e varia entre zero e um, permitindo uma correspondência bastante próxima com o modelo de dois parâmetros que usa a função da ogival proposto por Lord. O mesmo ocorre para o modelo logístico de um parâmetro, que também é simétrico e varia de 0 a 1.

Entretanto, quando se trata do modelo logístico de três parâmetros (ML3P), essa correspondência direta entre esse modelo e o modelo de Lord deixa de ser válida. Ao acrescentar o parâmetro c para mapear o acerto ao acaso,

[...] infelizmente, [...] algumas das boas propriedades matemáticas da função logística foram perdidas. No entanto, o modelo resultante passou a ser conhecido como modelo logístico de três parâmetros, mesmo que tecnicamente já não seja mais um modelo logístico (Baker, 2001, p. 28, tradução nossa).

De fato, o acréscimo do parâmetro altera a simetria da curva e afasta seu comportamento da função normal padrão. Inclusive, é por isso que o ML3P, apesar de ser chamado de logístico, não é considerado como tal por Baker e outros.

Nesse sentido, como o valor $D = 1,702$ foi originalmente calculado em uma situação em que $c = 0$ (ogiva normal e função logística padrão), sua aplicação ao ML3P não resulta na aproximação desejada à função normal padrão. Apesar disso, seu uso nos modelos de três parâmetros permanece por conveniência, padronização e história por facilitar comparações com modelos da normal acumulada, ainda que sem equivalência exata comprovada, como discutido por Haley.

4.4 Cálculo da probabilidade de acerto

Considere a matriz de padrão de resposta a seguir. Como discutido na introdução deste trabalho, tais dados isoladamente não conseguem responder a algumas perguntas que são feitas, tanto em relação aos resultados obtidos por cada um dos indivíduos quanto em relação aos itens que compõem o teste. Um questionamento frequente relacionado às questões do teste diz respeito ao motivo pelo qual alguns itens apresentam maior taxa de acerto do que outros. No caso da tabela a seguir, por exemplo, podemos nos perguntar: por que os itens 4 e 9 tiveram mais acertos que o item 7? Quais características tais itens possuem que fazem refletir esse resultado?

Indivíduo	Item $i = 1$	Item $i = 2$	Item $i = 3$	Item $i = 4$	Item $i = 5$	Item $i = 6$	Item $i = 7$	Item $i = 8$	Item $i = 9$
$j = 1$	0	0	0	0	0	0	0	0	0
$j = 2$	0	0	0	1	0	0	0	0	1
$j = 3$	1	0	0	1	0	0	0	0	1
$j = 4$	1	0	0	1	0	1	0	1	1
$j = 5$	1	1	1	1	1	1	0	1	1
$j = 6$	1	1	1	1	1	1	1	1	1
$j = 7$	1	1	1	1	1	1	1	1	1

Tabela 4.1 – Matriz de padrão de resposta $U_{7 \times 9}$

Fonte: autoria própria

A TRI, por meio de seus modelos, nos dá respostas a essas perguntas. No caso do ML3P, como ponto de partida temos os dados da matriz padrão de resposta, e como incógnitas temos, de um lado, os parâmetros a , b e c de cada item e, do

outro, a proficiência θ do indivíduo. Dessa forma, para determinar uma dessas incógnitas, precisamos supor que a outra é conhecida, ou seja, se queremos estimar o conjunto de parâmetros de cada item, precisamos supor que as proficiências dos indivíduos são conhecidas, e vice-versa.

Por exemplo, vamos supor conhecidos os parâmetros do ML3P dos nove itens do teste da tabela anterior, conforme a tabela a seguir, mas desconhecidos os resultados do teste (erros e acertos), ou seja, imagine que ainda não tenhamos em mãos a matriz de padrão de resposta da tabela 2 acima. Além disso, suponha também que o nível de aptidão θ de um certo indivíduo também seja desconhecido.

Item	<i>a</i>	<i>b</i>	<i>c</i>
1	1,20	-1,0	0,20
2	0,80	0,00	0,25
3	1,50	1,00	0,15
4	0,90	-2	0,20
5	1,70	0,50	0,25
6	1,30	-0,50	0,20
7	2,00	1,50	0,10
8	0,70	0,00	0,30
9	1,10	-1,50	0,20

Tabela 4.2 – Valor dos parâmetros dos itens do teste

Fonte: autoria própria

Agora vamos calcular a probabilidade desse indivíduo acertar cada um desses itens, antes de termos os dados do resultado do teste. Como não sabemos qual a aptidão θ desse indivíduo, vamos fazer suposições. Qual seria essa probabilidade caso esse indivíduo tivesse $\theta = -3$? E qual seria ela, se $\theta = -2$? E assim por diante. É como se estivéssemos calculando essa probabilidade para vários indivíduos diferentes cujos valores de θ são conhecidos e diferentes uns dos outros. A tabela a seguir mostra esses valores.

Item	$\theta = -3$	$\theta = -2$	$\theta = -1$	$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 3$
1	0,2665	0,3852	0,6000	0,8148	0,9335	0,9787	0,9935
2	0,3124	0,3760	0,4825	0,6250	0,7675	0,8740	0,9376
3	0,1521	0,1593	0,1903	0,3051	0,5750	0,8449	0,9597
4	0,4312	0,6000	0,7688	0,8865	0,9496	0,9787	0,9912
5	0,2519	0,2605	0,3043	0,4746	0,7754	0,9457	0,9895
6	0,2299	0,2996	0,4744	0,7256	0,9004	0,9701	0,9916
7	0,1001	0,1008	0,1060	0,1427	0,3420	0,7580	0,9573
8	0,3764	0,4385	0,5323	0,6500	0,7677	0,8615	0,9236
9	0,3289	0,4927	0,7073	0,8711	0,9519	0,9833	0,9944

Tabela 4.3 – Probabilidade de acerto em cada item do teste,

Fonte: autoria própria

Para analisar esses resultados, podemos plotar esses pontos, para cada item, em um gráfico que relaciona a probabilidade de acerto $P_i(\theta)$ com os supostos valores de θ , considerando os respectivos valores dos parâmetros dos itens. Veja:

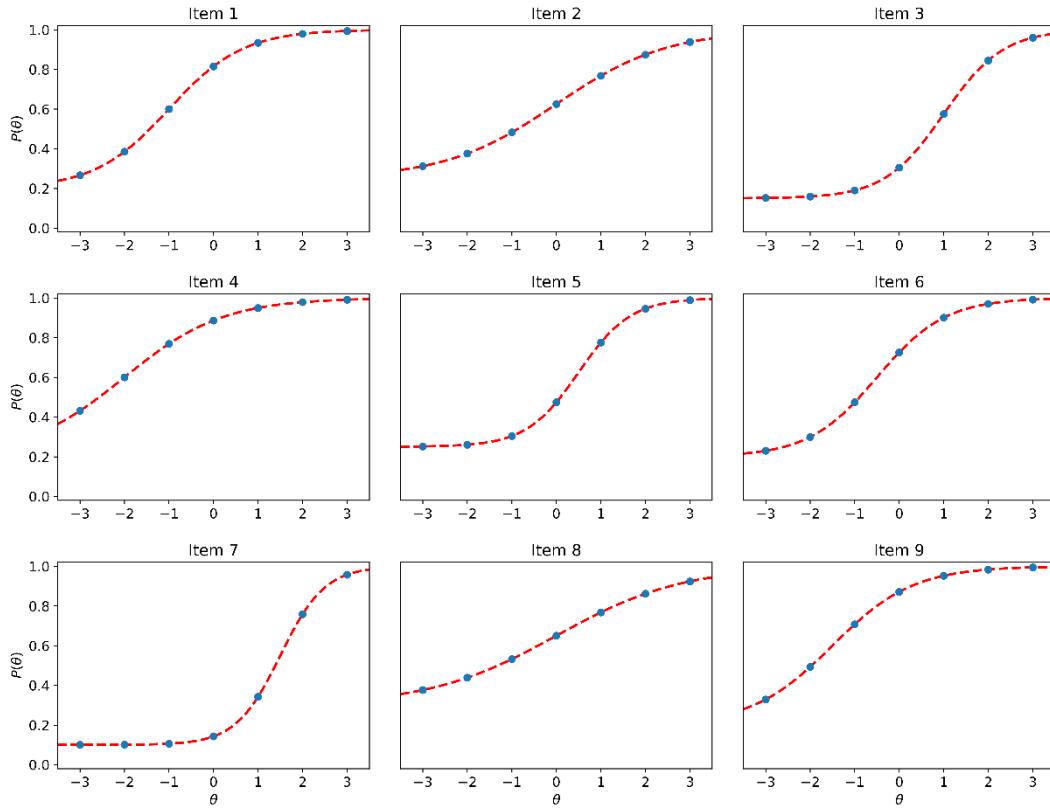


Gráfico 4.2 – Probabilidade de acerto e nível de aptidão

Fonte: autoria própria

Os sete pontos plotados em cada gráfico seguem a tendência indicada pela linha pontilhada, que os interpola. Como é de se esperar, essa curva segue uma tendência logística, ou seja, de uma curva sigmoide. Mais que isso: essa tendência é crescente, o que faz sentido, uma vez que essa é uma curva de tendência da probabilidade de acerto, de forma que, quanto maior é o valor de aptidão do indivíduo, maior é a chance de ele acertar o item.

Então, nesse exemplo, note que o procedimento adotado foi

[...] estimar as probabilidades de acerto antes de termos os dados empíricos; isto se chama cálculo da *probabilidade*. Se você fizer os cálculos em cima de dados já coletados, isto é, *a posteriori*, então se fala de *verossimilhança* (Pasquali, 2018, p. 92).

A partir do momento que passamos a conhecer o resultado do teste, temos uma atualização em relação à nossa crença inicial sobre a distribuição de probabilidade envolvida, que nos permitirá, para cada item, localizar com maior precisão a aptidão θ do indivíduo nos gráficos anteriores. Nesse sentido, conforme Pasquali (2018, p. 95), queremos descobrir qual “[...] é o nível de θ que melhor explica os resultados obtidos pelo sujeito”, o que é possível de ser feito por meio de estimadores. O caso em que conhecemos as aptidões θ dos indivíduos e desconhecemos os parâmetros pode ser pensado de maneira análoga.

4.5 Cálculo da probabilidade de erro

A partir da probabilidade P de acerto a um item, podemos calcular a probabilidade de erro por meio do seu complementar probabilístico $Q = 1 - P$. Assim, do exemplo anterior, obtemos as seguintes probabilidades de erro $Q_i(\theta)$:

Item	$\theta = -3$	$\theta = -2$	$\theta = -1$	$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 3$
1	0,7335	0,6148	0,4000	0,1852	0,0665	0,0213	0,0065
2	0,6876	0,6240	0,5175	0,3750	0,2325	0,1260	0,0624
3	0,8479	0,8407	0,8097	0,6949	0,4250	0,1551	0,0403
4	0,5688	0,4000	0,2312	0,1135	0,0504	0,0213	0,0088
5	0,7481	0,7395	0,6957	0,5254	0,2246	0,0543	0,0105
6	0,7701	0,7004	0,5256	0,2744	0,0996	0,0299	0,0084
7	0,8999	0,8992	0,8940	0,8573	0,6580	0,2420	0,0427
8	0,6236	0,5615	0,4677	0,3500	0,2323	0,1385	0,0764
9	0,6711	0,5073	0,2927	0,1289	0,0481	0,0167	0,0056

Tabela 4.4 – Probabilidade de erro em cada item do teste,
Fonte: autoria própria

Da mesma forma, podemos plotar esses pontos em um gráfico, conforme figura a seguir. Repare que cada curva de tendência da probabilidade de erro $Q(\theta)$ tem crescimento contrário em relação às curvas de tendência de $P(\theta)$.

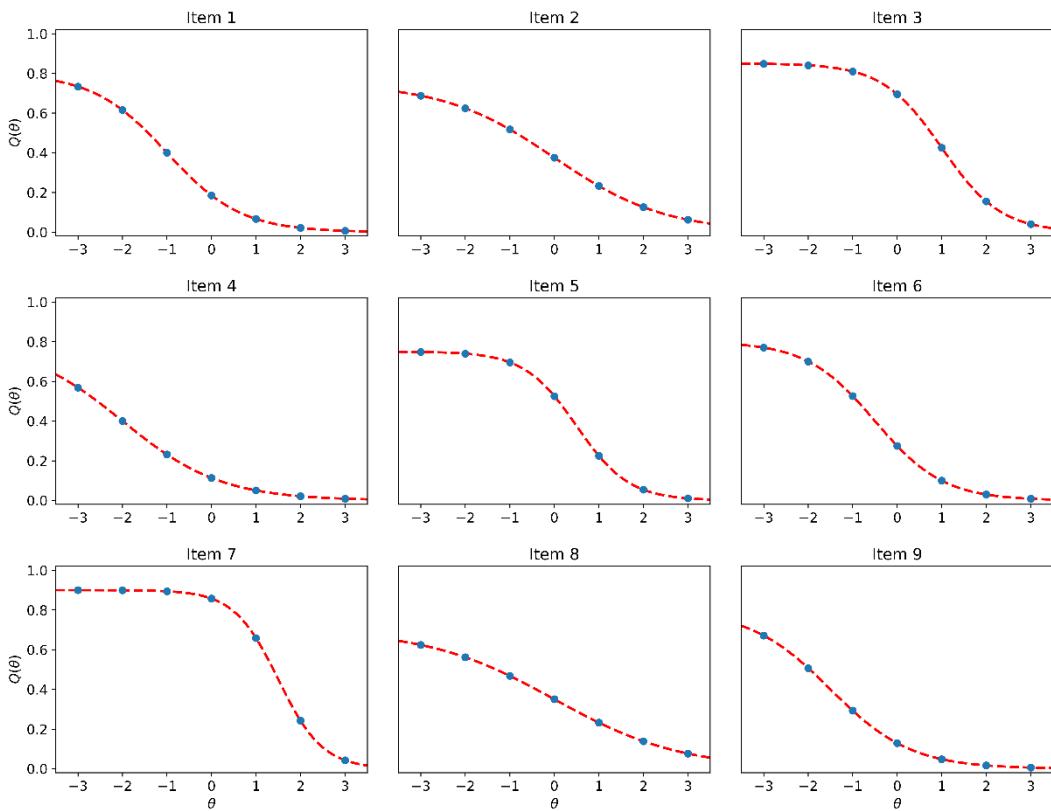


Gráfico 4.3 – Probabilidade de erro e nível de aptidão

Fonte: autoria própria

4.6 Comparação entre os modelos logísticos

Embora seja comum apresentar os modelos logísticos de um, dois e três parâmetros de forma progressiva, como se o modelo de um parâmetro fosse a base para os demais, historicamente, esse desenvolvimento não ocorreu dessa forma, como visto anteriormente. A abordagem aqui proposta, que também não segue a ordem cronológica do desenvolvimento dos modelos, tem como objetivo evidenciar as diferenças e semelhanças entre eles. Assim, os modelos de um e dois parâmetros podem ser vistos como casos particulares do modelo de três parâmetros, com restrições específicas sobre os parâmetros de discriminação e acerto ao acaso.

A seguir, apresenta-se uma comparação entre esses modelos por meio de uma tabela que mostra valores aproximados da probabilidade de acerto de vários itens, em cada um dos três modelos logísticos, para um indivíduo de aptidão $\theta = 0,5$, usando $D = 1$ e $e = 2,72$.

Item	<i>a</i>	<i>b</i>	<i>c</i>	P_{ML1P}	P_{ML2P}	P_{ML3P}
1	0,8	0,6	0,15	0,4750	0,4800	0,5580
2	0,8	0,6	0,25	0,4750	0,4800	0,6100
3	0,8	1,8	0,15	0,2140	0,2610	0,3719
4	0,8	1,8	0,25	0,2140	0,2610	0,4458
5	1,2	0,6	0,15	0,4750	0,4700	0,5495
6	1,2	0,6	0,25	0,4750	0,4700	0,6025
7	1,2	1,8	0,15	0,2140	0,1735	0,2975
8	1,2	1,8	0,25	0,2140	0,1735	0,3801

Tabela 4.5 – Comparação entre os modelos logísticos

Fonte: autoria própria

5

Joint Maximum Likelihood Estimation no ML3P

Nosso objetivo é, a partir das respostas dadas a um teste, determinar a proficiência do testando usando o ML3P. Para isso, precisamos, antes, obter o nível de aptidão do testando relacionado a cada um dos itens, uma vez que, de acordo com o princípio da unidimensionalidade, cada item avalia apenas um traço latente, que é aquele dominante na realização da ação necessária para resolver a questão. Por sua vez, para chegarmos a uma estimativa do nível de aptidão de um indivíduo avaliado em um item, precisamos, a partir dos dados contidos na matriz de padrão de respostas e por meio dos modelos probabilísticos, determinar o quanto provável é esse indivíduo acerto o item.

Esse processo, em certo sentido, se contrapõe àquele em que desejamos calcular a probabilidade de um evento ocorrer, partindo de um modelo probabilístico com parâmetros conhecidos, mas cujos dados de observação estão ausentes porque o experimento não foi realizado. No nosso caso em estudo, fazemos o caminho inverso: assumimos a validade do modelo ML3P e, com base nos dados coletados, buscamos estimar os parâmetros que melhor explicam os resultados. Uma vez estimados esses parâmetros, podemos então atualizar o ML3P e calcular a probabilidade *a posteriori* de um indivíduo ter respondido de determinada forma, condicionada à sua habilidade latente. Neste capítulo, será apresentado o conceito de verossimilhança e de função de verossimilhança, elementos que nos permitirão obter tanto estimativas dos parâmetros dos itens quanto das aptidões dos indivíduos por meio da *Joint Maximum Likelihood Estimation* (JMLE).

5.1 Estimadores e estimativas

Conforme Bussab e Morettin (2010, p. 296), “[...] a Inferência Estatística tem por objetivo fazer generalizações sobre uma população, com base nos dados de uma amostra”. O que queremos saber sobre uma população é chamado de parâmetro populacional. Como não temos acesso a toda a população, usamos uma amostra para tentar descobrir esse valor. Esse processo é chamado de estimação, isto é, usamos características da amostra para estimar características da população.

Parâmetros são medidas que descrevem características da população, enquanto estatísticas são medidas calculadas a partir da amostra. Quando usamos uma estatística com o objetivo de estimar um parâmetro populacional, ela recebe o nome de estimador. Formalmente, dada uma amostra aleatória (X_1, \dots, X_n) de uma variável aleatória X , um estimador T do parâmetro λ é qualquer função dos dados amostrais $T = g(X_1, \dots, X_n)$. A estimativa, por fim, é o valor numérico que o estimador assume em uma amostra específica, ou seja, o estimador é a regra usada para obter uma aproximação do parâmetro populacional, enquanto a estimativa é o resultado numérico obtido ao aplicar essa regra aos dados amostrais (Bussab; Morettin, 2010).

5.2

Estimador de Máxima Verossimilhança

Dentre os diversos métodos de estimação, destaca-se o da máxima verossimilhança, cujo resultado é o **estimador de máxima verossimilhança (EMV)**, também conhecido por **MLE (Maximum Likelihood Estimator)**. Esse método baseia-se na chamada função de verossimilhança, que expressa a probabilidade de se observar a amostra obtida em função dos parâmetros desconhecidos do modelo. A ideia por trás desse método é a seguinte: fixando-se um dentre os vários valores possíveis para o parâmetro (ou para os parâmetros) e supondo esse valor o verdadeiro, quão provável seria obter os dados observados? Essa pergunta é feita e respondida para todos os valores possíveis do parâmetro, de forma que a função de verossimilhança avalia quão bem diferentes valores dele explicam os dados observados no sentido de permitir mapear quais valores do parâmetro levam o modelo a gerar observações mais próximas daquelas que foram efetivamente obtidas.

Note que a função de verossimilhança, por si só, não é um estimador, mas sim uma ferramenta utilizada para construí-lo. De fato, os dados são fixos, pois já foram observados, enquanto a verossimilhança é uma função dos parâmetros, variando conforme o valor atribuído a eles. O estimador por máxima verossimilhança (EMV) é, portanto, uma função dos dados que indica qual valor do parâmetro que maximiza a função de verossimilhança. No nosso caso, os estimadores dos parâmetros são as funções de estimação que serão obtidas por meio da função de verossimilhança e

cujas estimativas serão obtidas por meio de um processo iterativo, como veremos a seguir, neste e no próximo capítulo.

Estatisticamente, encontrar o EMV de um parâmetro significa obter o valor que melhor representa esse parâmetro dentro da distribuição de probabilidade que gerou os dados observados. Nesse contexto, o princípio da verossimilhança afirma que “devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a ‘mais provável’ ” (Bussab & Morettin, 2010, p. 308).

A inferência estatística possui três abordagens principais: a **frequentista**, a **bayesiana tradicional** e a **bayesiana empírica**. A primeira trata os parâmetros como fixos e desconhecidos, baseando-se apenas nos dados observados. A segunda considera os parâmetros como variáveis aleatórias, incorporando um conhecimento prévio por meio de uma **distribuição de probabilidade *a priori*** desses parâmetros, que é aquela que hipotetizamos antes de obtermos os dados. Já a abordagem bayesiana empírica adota uma perspectiva híbrida, estimando a distribuição *a priori* dos parâmetros diretamente a partir dos dados, aplicando o Teorema de Bayes para combiná-la com a verossimilhança. Dentro dessas abordagens, destacam-se duas formas de lidar com a função de verossimilhança: a marginal e a conjunta. Podemos dizer que, tanto a verossimilhança conjunta quanto a marginal são tipos de função de verossimilhança que medem o quanto bem um conjunto de parâmetros explica os dados observados, mas com enfoques distintos uma da outra.

A verossimilhança marginal, ou **Maximum Marginal Likelihood Estimation (MMLE)**, é característica da abordagem bayesiana empírica e resulta da integração sobre variáveis latentes ou auxiliares, refletindo a incerteza associada a essas quantidades não observadas. Para grandes amostras, tais variáveis latentes ou auxiliares geralmente causam incômodo por sobrecarregar o processo, uma vez que o tornam computacionalmente custoso, como também ocorre na verossimilhança conjunta. Para contornar esse problema, na verossimilhança marginal, tais variáveis são tratadas como variáveis aleatórias com distribuição *a priori* conhecida (ou hipotetizada) e fixada (não ajustada com base nos dados observados e não dependente dos parâmetros que são estimados), mas os parâmetros de interesse continuam sendo estimados via máxima verossimilhança. Assim, nessa abordagem para o caso da TRI e do ML3P, os parâmetros a , b , e c dos itens seriam tratados como fixos, mas

desconhecidos, enquanto a aptidão θ é tratada como uma variável aleatória com distribuição *a priori* conhecida.

Neste capítulo, vamos utilizar estimadores de máxima verossimilhança conjunta, ou ***Joint Maximum Likelihood Estimation (JMLE)***, para determinar tanto os parâmetros dos itens de um teste no ML3P da TRI quanto as aptidões do indivíduos submetidos a esse teste em um mesmo procedimento. A verossimilhança conjunta, típica da abordagem frequentista, considera todos os parâmetros simultaneamente. Tal abordagem, no caso da TRI e do ML3P, além dos parâmetros a , b , e c dos itens, que são tratados como fixos, mas desconhecidos, a aptidão θ também é considerada um parâmetro fixo e desconhecido.

5.3 Função de verossimilhança

Matematicamente, considere (X_1, \dots, X_n) uma amostra aleatória de uma variável aleatória X qualquer de uma distribuição de probabilidade com $f(X|\lambda)$ sendo a função massa de probabilidade, se X for discreta, ou sendo a função densidade de probabilidade, se X for contínua. Seja também λ um parâmetro (ou vetor de parâmetros $\lambda = (a, b, c, \dots)$) dessa distribuição e (x_1, \dots, x_n) , com, $X_1 = x_1, \dots, X_n = x_n$, uma observação. Como, por definição de amostra aleatória, as n variáveis aleatórias (X_1, \dots, X_n) são independentes e têm a mesma distribuição de probabilidade de X (Meyer, 1983), a **função de probabilidade conjunta** de (x_1, \dots, x_n) , condicionada ao parâmetro λ (ou ao vetor $\lambda = (a, b, c, \dots)$), representa a distribuição simultânea das probabilidades associadas a combinações de valores de (X_1, \dots, X_n) , sendo dada por

$$f(X_1 = x_1, \dots, X_n = x_n | \lambda) = \prod_{i=1}^n f(x_i | \lambda) .$$

Essa função de probabilidade conjunta é a base para a construção da função de verossimilhança, usada para obter o estimador de máxima verossimilhança e, por fim, estimar o parâmetro λ (ou ao vetor $\lambda = (a, b, c, \dots)$) com base nos dados

observados. Essa expressão diz que a probabilidade (ou densidade de probabilidade) de se observar os valores (x_1, \dots, x_n) , dado o parâmetro λ , é igual ao produto das probabilidades (ou densidades) individuais de cada observação, assumindo que elas são independentes e seguem a mesma distribuição de probabilidade.

Para variáveis discretas, essa função, chamada de função de massa de probabilidade conjunta, atribui diretamente uma probabilidade para a ocorrência simultânea de combinações específicas de valores das variáveis envolvidas. Já para variáveis contínuas, ela é chamada de densidade de probabilidade conjunta e, nesse caso, a probabilidade de ocorrência simultânea está associada à “área sob a superfície da densidade” correspondente a uma “região” definida do espaço amostral.

Perceba que, quando conhecemos os parâmetros da distribuição, mas não conhecemos os dados, $f(X_1 = x_1, \dots, X_n = x_n | \lambda)$ é uma massa de probabilidade (ou densidade de probabilidade), que é função dos dados e não dos parâmetros. Agora, se conhecemos os dados (x_1, \dots, x_n) , mas não os parâmetros, chamamos a função de probabilidade conjunta f de função de verossimilhança conjunta $L(\lambda | x_1, \dots, x_n)$, de **função de verossimilhança**, ou, simplesmente, verossimilhança do parâmetro λ (ou vetor de parâmetros $\lambda = (a, b, c, \dots)$), dada a observação (x_1, \dots, x_n) , sendo essa uma função dos parâmetros e não dos dados. Assim, fazendo $\mathbf{x} = (x_1, \dots, x_n)$, temos

$$\begin{aligned} L(\lambda | x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i | \lambda), \\ \text{ou } L(\lambda | \mathbf{x}) &= \prod_{i=1}^n f(x_i | \lambda), \\ \text{ou, simplesmente, } L(\lambda) &= \prod_{i=1}^n f(x_i | \lambda) . \end{aligned}$$

5.4 Função de log-verossimilhança

Para estimar o parâmetro λ (ou vetor de parâmetros $\lambda = (a, b, c, \dots)$) em estudo de uma distribuição $f(X | \lambda)$, basta obter a função de verossimilhança

$L(\lambda | x_1, \dots, x_n)$ da observação (x_1, \dots, x_n) , que depende do parâmetro desconhecido λ (ou do vetor de parâmetros $\lambda = (a, b, c, \dots)$) e, depois, obter o valor do EMV $\hat{\lambda}$ (ou do vetor de parâmetros $\hat{\lambda} = (\hat{a}, \hat{b}, \hat{c}, \dots)$), que maximiza essa função. Para isso, em muitos casos, pode ser conveniente maximizar o logaritmo da função de verossimilhança $\ell(\lambda)$, chamada de **função de log-verossimilhança**, em vez da função de verossimilhança original $L(\lambda)$. Dessa forma, fazendo $\mathbf{x} = (x_1, \dots, x_n)$, representamos a log-verossimilhança por

$$\begin{aligned}\ell(\lambda | x_1, \dots, x_n) &= \log_N L(\lambda | x_1, \dots, x_n), \\ \text{ou } \ell(\lambda | \mathbf{x}) &= \log_N L(\lambda | \mathbf{x}), \\ \text{ou, simplesmente, } \ell(\lambda) &= \log_N L(\lambda | \mathbf{x}),\end{aligned}$$

com $N \in \mathbb{R}$ escolhido adequadamente para cada situação, com $N > 0$ e $N \neq 1$ (Bussab; Morettin, 2010).

De ordem prática, uma das principais razões para se preferir o uso do logaritmo da verossimilhança à própria é que o logaritmo transforma as multiplicações de probabilidades em somas, e os expoentes em fatores de produtos, o que é muito fácil de se lidar. Além disso, multiplicar muitas probabilidades, cujos valores se encontram quase que em sua totalidades entre zero e 1, pode levar a números muito próximos de zero, o que, geralmente, resulta em imprecisão. Outro motivo, é que é inevitável o uso de derivadas ao maximizar a verossimilhança, e derivar somas é muito mais simples que derivar produtos.

Com essas considerações, o **estimador de máxima verossimilhança conjunta (EMVC)** é uma função dos dados, obtida por meio da função de verossimilhança (ou função de log-verossimilhança), usada para encontrar uma estimativa $\hat{\lambda}$ para o parâmetro λ (ou $\lambda = (a, b, c, \dots)$). Tal estimativa é aquela que maximiza a verossimilhança ou (log-verossimilhança). Nesse processo, como geralmente não há uma solução fechada para o estimador assim obtido, a maximização de $\ell(\lambda)$, nesse caso, pode ser obtida por meio de métodos numéricos iterativos, como o algoritmo de Newton-Raphson.

5.5 Estimação conjunta dos parâmetros e aptidões

Com a matriz de padrão de respostas U em mãos, o primeiro objetivo é estimar, a partir desses dados observados, os parâmetros do modelo logístico de três parâmetros (ML3P) para cada item. Esse processo, em certo sentido, se contrapõe ao cálculo de probabilidades usando um modelo probabilístico, no qual partimos de parâmetros conhecidos para calcular a probabilidade de ocorrência de certos eventos ainda não observados, ou seja, ainda não temos os dados do experimento. No nosso caso em estudo, fazemos o caminho inverso: assumimos a validade do modelo ML3P e, com base nos dados coletados, buscamos estimar os parâmetros que melhor explicam os resultados. Uma vez estimados esses parâmetros, podemos então atualizar o ML3P e calcular a probabilidade *a posteriori* de um indivíduo ter respondido de determinada forma, condicionada à sua habilidade latente.

Na aplicação de um teste, alguns cenários são possíveis em relação ao conhecimento dos parâmetros a_i , b_i e c_i dos itens e da aptidão θ de cada indivíduo. Pode ser que antes de se obter os resultados do teste já se conheçam os parâmetros, como no caso de itens pré-testados ou itens que já foram utilizados em testes anteriores, mas se desconheça a aptidão de cada indivíduo, como ocorreu no exemplo dado no início da seção 4.4. O contrário também é possível, isto é, pode ser que se conheçam as aptidões do testando antes da aplicação do teste, mas se desconheçam os parâmetros dos itens. Porém, o que mais comumente ocorre é, após a aplicação e correção de um teste, temos um conjunto de dados referentes aos erros e acertos dos indivíduos testados em cada item do teste – matriz padrão de resposta –, mas não temos nem o valor dos parâmetros dos itens, nem o das aptidões dos indivíduos. Trataremos desse caso aqui, chamado de **estimação conjunta** dos parâmetros dos itens e das aptidões.

Na estimação conjunta tradicional dos parâmetros dos itens e das aptidões (habilidades ou proficiência) de cada testando, teríamos $3m + n$ parâmetros a serem estimados simultaneamente, o que pode exigir enorme gasto computacional. Um procedimento alternativo *back-and-forth* mas ainda considerado como uma estimação conjunta, foi proposto por Birnbaum, cujo passo inicial é obter uma primeira estimativa dos parâmetros dos itens, partindo de uma atribuição grosseira para o valor das aptidões, por exemplo, por meio de escores padronizados (de Andrade;

Tavares; da Cunha Valle, 2000) ou fazendo $\theta_j = 0$ (Pasquali, 2018), baseando-se na suposição de que, a nível populacional, a distribuição das habilidades tende a seguir uma normal padrão (de Andrade; Tavares; da Cunha Valle, 2000).

Partindo dessa primeira estimativa para os parâmetros dos itens, procede-se com algum procedimento iterativo para obter estimativas melhores para eles, até algum critério de parada ocorrer. Com estimativas melhoradas para os parâmetros dos itens, um procedimento análogo é feito para refinar a estimativa inicial das aptidões. Em seguida, retorna-se à primeira etapa com os valores das aptidões atualizados e procede-se com novas estimativas para os parâmetros do itens, cujos valores são utilizados para obter novas atualizações para os valores das aptidões, e assim por diante. Esse processo de “vai e volta” se repete até que algum critério global de parada seja atingido (de Andrade; Tavares; da Cunha Valle, 2000). Esse será o procedimento adotado aqui, uma vez que reduz a exigência computacional, além de permitir tanto a estimação individual dos parâmetros dos itens quanto das aptidões dos testandos. No entanto,

“[...] esse procedimento tem um problema sério: sabe-se que, para os parâmetros dos itens conhecidos, os Estimadores de Máxima Verossimilhança (EMV) das habilidades convergem [...] para os seus verdadeiros valores quando o número de itens cresce; com as habilidades conhecidas, os EMV dos parâmetros dos itens [...] convergem para os seus verdadeiros valores quando o número de indivíduos i cresce” (Rabelo, 2013, p. 28).

6

Calibração dos itens no ML3P usando JMLE

A seguir, descreveremos como os parâmetros a_i , b_i e c_i dos itens podem ser estimados no ML3P. A esse processo de estimação dos parâmetros chamamos de **calibração ou parametrização dos itens**. Existem diferentes formas de se fazer isso, partindo da escolha da ferramenta para se obter o estimador (verossimilhança conjunta ou marginal, por exemplo), passando pela abordagem (estimação conjunta dos parâmetros dos itens e da aptidão dos indivíduos ou em duas etapa), chegando ao processo iterativo a ser utilizado para se obter a estimativa (como os métodos de Newton-Raphson, “*Scoring*” de Fisher ou procedimentos bayesianos, sendo estes últimos baseados em técnicas numéricas para inferência a partir da distribuição *a posteriori*). Considerando os objetivos aqui propostos, seguiremos o mesmo caminho de Rabelo (2013), apresentando as principais etapas desse processo, sem a intenção de esgotá-lo. Ademais, cabe ressaltar que tal estimação é feita por meio de métodos numéricos iterativos com auxílio computacional, o que explica a aplicação da TRI de forma mais acentuada somente após a década de 70, com os avanços tecnológicos.

6.1 A função de verossimilhança dos parâmetros no ML3P

Para estimar tanto os valores dos parâmetros quanto o das aptidões, usaremos seus estimadores de máxima verossimilhança conjunta (EMVC) obtidas por meio de uma função dos parâmetros dos itens e das aptidões dos testandos: a chamada função de verossimilhança conjunta, relacionada à matriz de padrão de resposta U . Dessa forma, em qualquer caso, precisamos antes de tudo obter a função de verossimilhança para o ML3P, modelo utilizado nesse estudo. Inicialmente, vamos considerar o caso em que queiramos obter a verossimilhança para estimar os parâmetros dos itens. Assim, vamos supor conhecidos o nível de aptidão dos testandos, por exemplo, fazendo estimativas grosseiras iniciais para cada θ_j .

Representando, por simplicidade, os três parâmetros do item i no ML3P por meio do vetor $\lambda_i = (a_i, b_i, c_i)$, podemos escrever a probabilidade do indivíduo j ,

que possui nível de habilidade θ_j , acertar o item i , de parâmetros $\lambda_i = (a_i, b_i, c_i)$, por

$$P_{ji}(X_{ji} = 1 | \theta_j, \lambda_i), \text{ ou} \\ P_{ji}(\theta_j, \lambda_i), \text{ ou, simplesmente, } P_{ji}$$

e, dessa forma, podemos representar a probabilidade do indivíduo j , que possui nível de habilidade θ_j , errar o item i , de parâmetros $\lambda_i = (a_i, b_i, c_i)$, por

$$P_{ji}(X_{ji} = 0 | \theta_j, \lambda_i), \text{ ou} \\ 1 - P_{ji}(X_{ji} = 1 | \theta_j, \lambda_i) = 1 - P_{ji}(\theta_j, \lambda_i) = 1 - P_{ji}, \text{ ou} \\ Q_{ji}(\theta_j, \lambda_i), \text{ ou, simplesmente, } Q_{ji} . \quad (6.1)$$

Note que a probabilidade do indivíduo j , que possui nível de habilidade θ_j , dar uma resposta ao item i , de parâmetros λ_i , cujo resultado é u_{ji} , pode ser escrita como

$$P_{ji}(X_{ji} = u_{ji} | \theta_j, \lambda_i) = P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} .$$

De fato, se o indivíduo j acertar o item i , então $u_{ji} = 1$. Assim,

$$P_{ji}(X_{ji} = 1 | \theta_j, \lambda_i) = P_{ji}^1 \cdot Q_{ji}^{1-1} = P_{ji} ,$$

mas, caso erre, $u_{ji} = 0$ e, então,

$$P_{ji}(X_{ji} = 0 | \theta_j, \lambda_i) = P_{ji}^0 \cdot Q_{ji}^{1-0} = Q_{ji} .$$

Em outras palavras, X_{ji} tem distribuição de Bernoulli, com

$$\begin{cases} p_{ji}(1) = (X_{ji} = 1 | \theta_j, \lambda_i) = P_{ji} \\ p_{ji}(0) = (X_{ji} = 0 | \theta_j, \lambda_i) = Q_{ji} \end{cases} .$$

Pela independência local dos itens, o acerto de um item pelo indivíduo j não afeta a probabilidade de acerto de outro item do teste pelo mesmo indivíduo. Dessa forma, a realização dos valores $X_{j1} = u_{j1}$, $X_{j2} = u_{j2}$, ..., $X_{ji} = u_{ji}$ e $X_{jn} = u_{jn}$ corresponde à ocorrência de eventos independentes. Portanto, usando $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ para representar, ordenadamente, as trincas de parâmetros $\lambda_i = (a_i, b_i, c_i)$

do M3PL dos n itens do teste, temos que a probabilidade de determinado indivíduo j de nível de aptidão θ_j dar o padrão de resposta $U_j = (u_{j1}, u_{j2}, \dots, u_{ji}, \dots, u_{jn})$ aos itens $1 \leq i \leq n$ de parâmetros λ_i é a mesma que a probabilidade de esse indivíduo dar, simultaneamente, cada uma de suas respostas individuais u_{ji} , e é dada por

$$\begin{aligned}
 P_j(X_j = U_j | \theta_j, \Lambda) &= P_j((X_{j1} = u_{j1} | \theta_j, \lambda_1), (X_{j2} = u_{j2} | \theta_j, \lambda_2), \\
 &\quad \dots, (X_{ji} = u_{ji} | \theta_j, \lambda_i), \dots, (X_{jn} = u_{jn} | \theta_j, \lambda_n)) \\
 &= P_{ji} \left(\bigcap_{i=1}^n X_{ji} = u_{ji} | \theta_j, \lambda_i \right) \\
 &= \prod_{i=1}^n P_{ji}(X_{ji} = u_{ji} | \theta_j, \lambda_i) \\
 &= \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} ,
 \end{aligned}$$

que é a probabilidade conjunta da ocorrência simultânea de $X_{j1} = u_{j1}$, $X_{j2} = u_{j2}$, \dots , $X_{ji} = u_{ji}$, \dots e $X_{jn} = u_{jn}$ para cada indivíduo j , em todos os itens.

Por sua vez, pela independência individual, as respostas oriundas de indivíduos diferentes são independentes, isto é, a probabilidade de um indivíduo j dar o padrão de respostas U_j não afeta a probabilidade do indivíduo k dar o padrão de respostas U_k , com $1 \leq j \leq m$, $1 \leq k \leq m$ e $k \neq j$. Assim, a realização dos valores $X_1 = U_1$, $X_2 = U_2$, \dots , $X_j = U_j$, \dots e $X_m = U_m$ corresponde à ocorrência de eventos independentes e a probabilidade de os indivíduos $1 \leq j \leq m$ de nível de aptidão θ_j darem respostas u_{ji} aos itens $1 \leq i \leq n$ de parâmetros λ_i tais que $X = U$ é a mesma que a probabilidade de esses mesmos indivíduos darem o respectivo padrão de resposta U_j a esses itens.

Nesse sentido, usando $\Theta = (\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_m)$ para representar, ordenadamente, as aptidões de cada indivíduo submetido ao teste, a probabilidade conjunta da ocorrência simultânea dos resultados $X_{j1} = u_{j1}$, $X_{j2} = u_{j2}$ \dots , $X_{ji} = u_{ji}$, \dots e $X_{jn} = u_{jn}$, com $1 \leq j \leq m$ e $1 \leq i \leq n$ de todos os m indivíduos testados em todos os n itens do teste, ou seja, a probabilidade de ocorrência de U é dada por

$$\begin{aligned}
P(X = U | \Theta, \Lambda) &= P((X_1 = U_1 | \theta_1, \Lambda), (X_2 = U_2 | \theta_2, \Lambda), \\
&\quad \dots, (X_j = U_j | \theta_j, \Lambda), \dots, (X_m = U_m | \theta_m, \Lambda)) \\
&= P_j \left(\bigcap_{j=1}^m X_j = U_j | \theta_j, \Lambda \right) \\
&= \prod_{j=1}^m P_j(X_j = U_j | \theta_j, \Lambda) \\
&= \prod_{j=1}^m \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \quad ,
\end{aligned}$$

que é a função de verossimilhança conjunta $L(\Lambda | U)$, ou simplesmente $L(\Lambda)$, associada aos valores amostrais observados de U no ML3P, dependendo apenas dos parâmetros desconhecidos $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_n)$, uma vez que temos os dados de repostas armazenados em U , D é fator de escala conhecido e supomos, inicialmente, que os valores das aptidões θ_j dos testandos eram conhecidos. Assim,

$$\begin{aligned}
L(\Lambda) &= L(\Lambda | U) \\
&= P(X = U | \Theta, \Lambda) \\
&= \prod_{j=1}^m \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \quad (6.2)
\end{aligned}$$

para $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ e com

$$P_{ji} = P_{ji}(X_{ji} = 1 | \theta_j, \lambda_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-D\lambda_i(\theta_j - b_i)}}$$

sendo a probabilidade do indivíduo j , de aptidão θ_j , acertar $(X_{ji} = 1)$ o item i , de parâmetros λ_i , e

$$Q_{ji} = P_{ji}(X_{ji} = 0 | \theta_j, \lambda_i) = 1 - P_{ji}$$

a probabilidade do indivíduo j , de aptidão θ_j , errar $(X_{ji} = 0)$ o item i , de parâmetros λ_i .

6.2 A função de log-verossimilhança dos parâmetros

Para obter as estimativas $\hat{\lambda} = (\hat{a}_i, \hat{b}_i, \hat{c}_i)$ por meio dos estimadores de máxima verossimilhança conjunta (EMVC) dos parâmetros $\lambda_i = (a_i, b_i, c_i)$ de cada item, basta maximizar a função de verossimilhança dos parâmetros dos itens $L(\Lambda)$ com relação aos valores amostrais armazenados em U , obtidos por meio dos resultados de erros e acertos das respostas dadas pelos m indivíduos submetidos ao teste com n itens dicotômicos. Porém, como foi dito anteriormente, na maioria dos casos, é mais conveniente maximizar a log-verossimilhança $\ell(\Lambda) = \log L(\Lambda)$, afinal, ao fazer isso, evitaremos trabalhar com tantos expoentes, o que por si só já facilita bastante. Além disso, o logaritmo aplicado preservará o crescimento da função original. Assim, a partir da equação (6.2), obtemos a log-verossimilhança dos parâmetros.

$$L(\Lambda) = \prod_{j=1}^m \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \Leftrightarrow \ell(\Lambda) = \log L(\Lambda) = \log \left(\prod_{j=1}^m \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \right).$$

Perceba, então, que maximizar a log-verossimilhança $\ell(\Lambda)$ é equivalente a maximizar a verossimilhança $L(\Lambda)$. Dessa forma, para obter os EMVC $\hat{\Lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_i, \dots, \hat{\lambda}_n)$ dos parâmetros $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_n)$ dos itens no ML3P, por praticidade, vamos preferir maximizar $\ell(\Lambda)$, que é dada por

$$\begin{aligned} \ell(\Lambda) &= \log L(\Lambda) \\ &= \log \left(\prod_{j=1}^m \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \right) \\ &= \sum_{j=1}^m \sum_{i=1}^n \log \left(P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \right) \\ &= \sum_{j=1}^m \sum_{i=1}^n \left(\log P_{ji}^{u_{ji}} + \log Q_{ji}^{1-u_{ji}} \right) \\ &= \sum_{j=1}^m \sum_{i=1}^n \left(u_{ji} \log P_{ji} + (1-u_{ji}) \log Q_{ji} \right) \end{aligned} \tag{6.3}$$

sendo $u_{ji} \in \{0, 1\}$ a resposta do indivíduo j , de aptidão θ_j , ao item i , de parâmetros λ_i , note que $\ell_{ji} = \ell_{ji}(\theta_j, \lambda_i) = u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}$ é a contribuição da resposta u_{ji} para a função de log-verossimilhança, de forma que se a resposta for correta, $u_{ji} = 1$ e, então, $\ell_{ji} = u_{ji} \log P_{ji}$, mas se for errada, $u_{ji} = 0$ e, dessa forma, $\ell_{ji} = \log Q_{ji}$.

Como a adição, subtração, multiplicação e composição de funções escalares contínuas a várias variáveis, nos devidos domínios, são também contínuas (Bortolossi, 2002), a função $\ell(\Lambda)$, que é uma função escalar de \mathbb{R}^{3n} em \mathbb{R} , é contínua. De fato, perceba que $\ell(\Lambda)$ é a soma, diferença, produto e composição das funções f_k , com $f_1(\Lambda) = \log X$, $f_2(\Lambda) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-D\theta_i(\theta_j - b_i)}}$, $f_3(\Lambda) = 0$ e $f_4(\Lambda) = 1$, escalares e contínuas em seus respectivos domínios $D_k \subset \mathbb{R}^{3n}$, com $k = 1, 2, 3$ e 4 . Note ainda que $f_2(\Lambda) \in D_1$, $\forall \Lambda \in \mathbb{R}^{3n}$, pois $0 < f_2(\Lambda) < 1$.

Outro fato importante é que toda função escalar contínua a várias variáveis definida num conjunto compacto (conjunto limitado e fechado) não vazio possui pelo menos um máximo global e pelo menos um mínimo global, que é o Teorema de Weierstrass no caso de funções de várias variáveis (Bortolossi, 2002). Note que, a princípio, esse teorema não pode ser aplicado a $\ell(\Lambda)$, pois, apesar de essa função ser contínua e o seu domínio $D_\ell = \mathbb{R}^{3n}$ ser fechado em \mathbb{R}^{3n} (o conjunto A dos pontos de fronteira de \mathbb{R}^{3n} é vazio, de forma que $A = \emptyset \subset \mathbb{R}^{3n}$), ele não é limitado (não existe uma bola fechada $\overline{B_r(0)}$ de centro em $\vec{0}_{3n} = (0, 0, \dots, 0)$ e raio r tal que $D_\ell = \mathbb{R}^{3n} \subset \overline{B_r(0)}$).

Porém, na prática, o valor dos parâmetros dos itens $\lambda_i = (a_i, b_i, c_i)$ tipicamente varia em intervalos reais bem comportados, a saber: $0 < a_i < 3$, $-3 < b_i < 3$, $0 < c_i < 1$ e os valores das aptidões, considerando-as com distribuição próxima da normal padrão na população, se situam no intervalo $-3 < \theta < 3$. Assim, tomando uma tolerância segura para os intervalos dos parâmetros, digamos $-5 < a_i < 5$, $-5 < b_i < 5$ e $5 < c_i < 5$, obtemos uma bola fechada $\overline{B_5(0)}$ de centro em

$\vec{0}_{3n} = (0, 0, \dots, 0)$ e raio $r = 5$ tal que $\hat{D}_\ell = \{\Lambda \in \mathbb{R}^{3n} \mid \lambda_i \in \mathbb{R}^3, -5 \leq a_i, b_i, c_i \leq 5,$ com $\forall i = 1, 2, \dots, n \} \subset \overline{B_5(0)}$.

Nesse caso, \hat{D}_ℓ é compacto, pois é fechado e limitado e, assim, como $\ell(\Lambda)$ é contínua, o Teorema de Weierstrass passa a ser válido e, então, ℓ possui pelo menos um máximo global e pelo menos um mínimo global em \hat{D}_ℓ . Como todo extremo global é também um extremo local, pela Regra de Fermat, e como $\ell(\Lambda)$ é uma função escalar de $3n$ variáveis, se $\hat{\Lambda}$ é um ponto crítico de $\ell(\Lambda)$ e é também um ponto interior de $D \subset \hat{D}_\ell \subset \mathbb{R}^{3n}$, então a matriz jacobiana de ℓ , matriz de derivadas parciais de primeira ordem, se anula em $\hat{\Lambda}$ (Bortolossi, 2002), ou, de outra forma, $\nabla \ell_{\lambda_i}(\Lambda) = \vec{0}$. Logo, é possível encontrar os EMVC $\hat{\Lambda}$ dos parâmetros Λ por meio da maximização da função de log-verossimilhança, uma vez que, no conjunto $\hat{D}_\ell = \mathbb{R}^{3n}$ do domínio de $\ell(\Lambda)$ considerado, há, pelo menos, um ponto de máximo local.

6.3 Separabilidade por itens da log-verossimilhança $\ell(\Lambda)$

Partindo da equação (6.3),

$$\ell(\Lambda) = \sum_{j=1}^m \sum_{i=1}^n \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right)$$

é a log-verossimilhança dos parâmetros dos itens e

$$\nabla \ell(\Lambda) = (\nabla_{\lambda_1} \ell(\Lambda), \nabla_{\lambda_2} \ell(\Lambda), \dots, \nabla_{\lambda_n} \ell(\Lambda), \dots, \nabla_{\lambda_m} \ell(\Lambda))$$

é o gradiente de ℓ no ponto $\Lambda = (\lambda_1, \dots, \lambda_n)$, temos

$$\begin{aligned} \nabla \ell(\Lambda) &= \nabla \left(\sum_{j=1}^m \sum_{i=1}^n \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) \right) \\ &= \sum_{j=1}^m \nabla \left(\sum_{i=1}^n \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) \right) \end{aligned}$$

em que, ao fixar $i = k$, temos

$$\nabla \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) = 0, \forall i \neq k \quad (6.4)$$

uma vez que, nesse caso, apenas os termos em k variam e os demais termos em $i \neq k$ são constantes em relação a k , de forma que

$$\begin{aligned} \nabla_{\lambda_k} \ell(\Lambda) &= \nabla_{\lambda_k} \left(\sum_{i=1}^n \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) \right) \\ &\stackrel{[6.2]}{=} \nabla_{\lambda_k} \left(u_{jk} \log P_{jk}(\lambda_k) + (1-u_{jk}) \log Q_{jk}(\lambda_k) \right) = \nabla_{\lambda_k} \ell(\lambda_k) \end{aligned} \quad (6.5)$$

Note que essa última igualdade reflete a independência local dos itens, um dos princípios fundamentais da TRI. Ele se manifesta aqui indicando que ℓ é separável por itens, no sentido de que podemos estimar os parâmetros por blocos, o que significa que é possível estimar os parâmetros de cada item separadamente, com menor custo computacional.

De fato, podemos fazer dois tipos de abordagem para o gradiente da função de verossimilhança ℓ . Uma delas, com o gradiente global $\nabla_{\lambda_i} \ell(\Lambda)$, em que ℓ depende do vetor completo de parâmetros dos itens Λ e estamos derivando em relação a certo vetor λ_i de parâmetros de um item i fixado, isto é,

$$\nabla_{\lambda_i} \ell(\Lambda) = (\nabla_{\lambda_1} \ell(\Lambda), \dots, \nabla_{\lambda_i} \ell(\Lambda), \dots, \nabla_{\lambda_n} \ell(\Lambda))$$

como feito anteriormente, e que, na forma matricial, fica

$$\nabla_{\lambda_i} \ell(\Lambda) = \begin{bmatrix} \nabla_{\lambda_1} \ell(\Lambda) \\ \nabla_{\lambda_2} \ell(\Lambda) \\ \vdots \\ \nabla_{\lambda_i} \ell(\Lambda) \\ \vdots \\ \nabla_{\lambda_n} \ell(\Lambda) \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial \ell}{\partial a_1}(\Lambda), \frac{\partial \ell}{\partial b_1}(\Lambda), \frac{\partial \ell}{\partial c_1}(\Lambda) \right) \\ \left(\frac{\partial \ell}{\partial a_2}(\Lambda), \frac{\partial \ell}{\partial b_2}(\Lambda), \frac{\partial \ell}{\partial c_2}(\Lambda) \right) \\ \vdots \\ \left(\frac{\partial \ell}{\partial a_i}(\Lambda), \frac{\partial \ell}{\partial b_i}(\Lambda), \frac{\partial \ell}{\partial c_i}(\Lambda) \right) \\ \vdots \\ \left(\frac{\partial \ell}{\partial a_n}(\Lambda), \frac{\partial \ell}{\partial b_n}(\Lambda), \frac{\partial \ell}{\partial c_n}(\Lambda) \right) \end{bmatrix} = \begin{bmatrix} \vec{0}_3 \\ \vec{0}_3 \\ \vdots \\ \vec{0}_3 \end{bmatrix}$$

com $\vec{0}_3 = (0, 0, 0)$. Então, $\nabla_{\lambda_i} \ell(\Lambda)$ é um vetor de dimensão $n \times 1$ e apenas o bloco correspondente ao item i carrega uma derivada não nula. Os demais são nulos porque os λ_k , $k \neq i$, são tratados como constantes na derivação com respeito a λ_i .

Ainda, perceba que $\nabla_{\lambda_i} \ell(\Lambda)$ é função apenas de λ_i , não dependendo de λ_k , para $k \neq i$. Assim, podemos escrever, de forma simplificada, $\nabla_{\lambda_i} \ell(\Lambda)$ como $\nabla_{\lambda_i} \ell(\lambda_i)$ (de Andrade; Tavares; da Cunha Valle, 2000). E esta seria a outra abordagem: usar diretamente o gradiente local $\nabla_{\lambda_i} \ell(\lambda_i)$, em que a função ℓ depende apenas dos parâmetros λ_i do item i , e a derivada aplicada também é em relação a esse vetor, ou seja,

$$\nabla_{\lambda_i} \ell(\lambda_i) = \left(\frac{\partial \ell}{\partial a_i}(\lambda_i), \frac{\partial \ell}{\partial b_i}(\lambda_i), \frac{\partial \ell}{\partial c_i}(\lambda_i) \right)$$

e será essa a abordagem que faremos aqui.

6.4 Estimador dos parâmetros dos itens

Antes de avançarmos na determinação dos estimadores dos parâmetros dos itens, vamos apresentar algumas considerações e resultados que serão utilizados ao longo deste capítulo. Para começar, apenas por conveniência, vamos denotar a “parte logística” do ML3P

$$f(x) = \frac{1}{1 + e^{-D a_i (\theta_j - b_i)}}$$

por

$$\Psi_{ji}(z_{ji}) = \frac{1}{1 + e^{-z_{ji}}} , \text{ com } z_{ji} = D a_i (\theta_j - b_i) . \quad (6.6)$$

Assim, temos

$$\Psi'_{ji}(z) = \frac{\partial \Psi_{ji}}{\partial z}(z) = \frac{e^{-z_{ji}}}{(1 + e^{-z_{ji}})^2} . \quad (6.7)$$

Além disso,

$$\begin{aligned}\frac{1}{1+e^{-z_{ji}}} &= \Psi_{ji}(z_{ji}) \Rightarrow 1 - \frac{1}{1+e^{-z_{ji}}} = 1 - \Psi_{ji}(z_{ji}) \\ &\Rightarrow \frac{e^{-z_{ji}}}{1+e^{-z_{ji}}} = 1 - \Psi_{ji}(z_{ji})\end{aligned}$$

e, multiplicando ambos os lados da última igualdade por $\Psi_{ji}(z_{ji})$, obtemos

$$\begin{aligned}\frac{e^{-z_{ji}}}{1+e^{-z_{ji}}} &= 1 - \Psi_{ji}(z_{ji}) \Rightarrow \Psi_{ji}(z_{ji}) \cdot \frac{e^{-z_{ji}}}{1+e^{-z_{ji}}} = \Psi_{ji}(z_{ji}) \cdot (1 - \Psi_{ji}(z_{ji})) \\ &\Rightarrow \frac{1}{1+e^{-z_{ji}}} \cdot \frac{e^{-z_{ji}}}{1+e^{-z_{ji}}} = \Psi_{ji}(z_{ji}) \cdot (1 - \Psi_{ji}(z_{ji})) \\ &\Rightarrow \frac{e^{-z_{ji}}}{(1+e^{-z_{ji}})^2} = \Psi_{ji}(z_{ji}) \cdot (1 - \Psi_{ji}(z_{ji})) \\ &\stackrel{[6.7]}{\Rightarrow} \Psi'_{ji}(z_{ji}) = \Psi_{ji}(z_{ji}) \cdot (1 - \Psi_{ji}(z_{ji})) \quad . \quad (6.8)\end{aligned}$$

Agora, fazendo $P_{ji} = P_{ji}(\lambda_{ji})$, $Q_{ji} = Q_{ji}(\lambda_{ji})$, $\Psi_{ji} = \Psi(z_{ji})$, e $\Psi'_{ji} = \Psi'_{ji}$, para simplificar a escrita, e denotando $\Psi'_{ji} \cdot \frac{1}{P_{ji} \cdot Q_{ji}}$ por W_{ji} , temos

- i.* $\Psi'_{ji} \cdot \frac{1}{P_{ji} \cdot Q_{ji}} = W_{ji}$
- ii.* $\Psi_{ji} \cdot (1 - \Psi_{ji}) \cdot \frac{1}{P_{ji} \cdot Q_{ji}} = W_{ji} \quad (6.9)$
- iii.* $(1 - \Psi_{ji}) \cdot \frac{1}{P_{ji} \cdot Q_{ji}} = \frac{W_{ji}}{\Psi_{ji}}$
- iv.* $\frac{1}{P_{ji} \cdot Q_{ji}} = \frac{W_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})}$

Agora, note que, como no nosso problema θ_j e u_{ji} são conhecidos (um grosseiramente estimado e o outro observado, respectivamente) e $D=1,7$ para a escala normal (ou $D=1$ para se manter na escala logística), veja que P_{ji} e $Q_{ji}=1-P_{ji}$ são funções de $\lambda_i=(a_i, b_i, c_i)$. Dessa forma,

$$P_{ji} = P_{ji}(\lambda_i) = c_i + (1-c_i) \cdot \frac{1}{1+e^{-Da_i(\theta_j-b_i)}}$$

$$\text{e } Q_{ji} = Q_{ji}(\lambda_i) = 1 - P_{ji}(\lambda_i) .$$

Assim, vamos determinar as expressões dos componentes do gradiente

$$\nabla_{\lambda_i} P_{ji}(\lambda_i) = \left(\frac{\partial P_{ji}}{\partial a_i}(\lambda_i), \frac{\partial P_{ji}}{\partial b_i}(\lambda_i), \frac{\partial P_{ji}}{\partial c_i}(\lambda_i) \right) ,$$

que são as derivadas parciais de primeira ordem de P_{ji} em relação a $\lambda_i=(a_i, b_i, c_i)$.

Para simplificar um pouco mais a escrita, apenas por praticidade, denotaremos $\lambda_i=(a_i, b_i, c_i)$ por $\lambda=(a, b, c)$, θ_j por θ , z_{ji} por z . Assim, partindo da equação (4.1), temos

$$P = c + (1-c) \cdot \frac{1}{1+e^{-Da(\theta-b)}} \stackrel{(6.6)}{\Rightarrow} P = c + (1-c) \cdot \Psi . \quad (6.10)$$

Dessa forma, P também se torna função de z , que por sua vez é função de θ e b . Assim, ao calcular os componentes do gradiente $\nabla_{\lambda_i} P_{ji}(\lambda_i)$, vamos precisar das derivadas parciais de primeira ordem de $z = Da(\theta-b)$, que são dadas por

$$i. \quad \frac{\partial z}{\partial a}(\theta, a, b) = D(\theta-b) \quad (6.11)$$

$$ii. \quad \frac{\partial z}{\partial b}(\theta, a, b) = -Da \quad (6.12)$$

$$iii. \quad \frac{\partial z}{\partial c}(\theta, a, b) = 0 \quad (6.13)$$

Agora, calculando as primeiras derivadas parciais de P , obtemos

$$\begin{aligned}
i. \quad \frac{\partial P}{\partial a}(\lambda) &= \frac{\partial}{\partial a} \cdot [c + (1-c) \cdot \Psi] \\
&= (1-c) \cdot \left[\frac{\partial \Psi}{\partial z} \cdot \frac{\partial z}{\partial a} \right] \\
&\stackrel{[6.8]}{=} (1-c) \cdot \Psi(1-\Psi) \cdot \frac{\partial z}{\partial a} \\
&\stackrel{[6.11]}{=} D(\theta-b)(1-c) \cdot \Psi(1-\Psi)
\end{aligned} \tag{6.14}$$

$$\begin{aligned}
ii. \quad \frac{\partial P}{\partial b}(\lambda) &= \frac{\partial}{\partial b} \cdot [c + (1-c) \cdot \Psi] \\
&= (1-c) \cdot \left[\frac{\partial \Psi}{\partial z} \cdot \frac{\partial z}{\partial b} \right] \\
&\stackrel{[6.8]}{=} (1-c) \cdot \Psi(1-\Psi) \cdot \frac{\partial z}{\partial b} \\
&\stackrel{[6.12]}{=} -Da(1-c) \cdot \Psi(1-\Psi)
\end{aligned} \tag{6.15}$$

$$\begin{aligned}
iii. \quad \frac{\partial P}{\partial c}(\lambda) &= \frac{\partial}{\partial c} \cdot [c + (1-c) \cdot \Psi] \\
&= 1 + \left(-\Psi - \frac{\partial \Psi}{\partial z} \cdot \frac{\partial z}{\partial c} \right) \\
&\stackrel{[6.13]}{=} 1 + \left(-\Psi - \frac{\partial \Psi}{\partial z} \cdot 0 \right) \\
&= 1 - \Psi
\end{aligned} \tag{6.16}$$

Por último, antes de encontrar uma expressão para os estimadores dos parâmetros dos itens, vamos incluir nos nossos resultados alguns gradientes que surgiram durante esse processo.

$$\begin{aligned}
i. \quad \nabla \log P_{ji}(\lambda_i) &= \frac{\nabla P_{ji}(\lambda_i)}{P_{ji}(\lambda_i)} \\
ii. \quad \nabla \log Q_{ji}(\lambda_i) &= \frac{\nabla Q_{ji}(\lambda_i)}{Q_{ji}(\lambda_i)} = \frac{\nabla(1 - P_{ij}(\lambda_i))}{Q_{ji}(\lambda_i)} = -\frac{\nabla P_{ij}(\lambda_i)}{Q_{ji}(\lambda_i)} \\
iii. \quad \nabla Q_{ji}(\lambda_i) &= \nabla(1 - P_{ji}(\lambda_i)) = -\nabla P_{ji}(\lambda_i)
\end{aligned} \tag{6.17}$$

Finalmente, com esses resultados, podemos encontrar os estimador de máxima verossimilhança, que, para cada item, é o valor $\hat{\lambda}_i$ que maximiza $\ell(\lambda_i)$. Assim, suponde que tal ponto exista, ele é um ponto de máximo de ℓ , assim precisamos pesquisas os pontos críticos dessa superfície, fazendo $\nabla_{\lambda_i} \ell(\lambda_i) = \vec{0}_3$. Assim, calculando o gradiente de $\ell(\lambda_i)$ aplicado ao ponto λ_i , temos

$$\begin{aligned}
 \nabla_{\lambda_i} \ell(\lambda_i) &\stackrel{(6.3)}{=} \nabla_{\lambda_i} \sum_{j=1}^m \sum_{i=1}^n \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) \\
 &= \sum_{j=1}^m \nabla_{\lambda_i} \left(\sum_{i=1}^n \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) \right) \\
 &\stackrel{(6.4)}{=} \sum_{j=1}^m \nabla_{\lambda_i} \left(u_{ji} \log P_{ji}(\lambda_i) + (1-u_{ji}) \log Q_{ji}(\lambda_i) \right) \\
 &\stackrel{(6.17)}{=} \sum_{j=1}^m \left(u_{ji} \frac{\nabla_{\lambda_i} P_{ji}(\lambda_i)}{P_{ji}(\lambda_i)} - (1-u_{ji}) \frac{\nabla_{\lambda_i} Q_{ji}(\lambda_i)}{Q_{ji}(\lambda_i)} \right) \\
 &= \sum_{j=1}^m (u_{ji} - P_{ji}(\lambda_i)) \cdot \nabla_{\lambda_i} P_{ji}(\lambda_i) \cdot \frac{1}{P_{ji}(\lambda_i) \cdot Q_{ji}(\lambda_i)} \tag{6.18}
 \end{aligned}$$

em que:

- $u_{ji} - P_{ji}(\lambda_i)$ é o erro de predição ou resíduo probabilístico;
- $\frac{1}{P_{ji}(\lambda_i) \cdot Q_{ji}(\lambda_i)}$ é o fator de escalonamento da curvatura da função de verossimilhança, que amplifica o impacto do resíduo probabilístico quando a probabilidade está próxima de 0 ou de 1;
- $\nabla_{\lambda_i} P_{ji}(\lambda_i)$ é sensibilidade da probabilidade;

Ainda, podemos determinar cada componente do gradiente $\nabla_{\lambda_i} \ell(\lambda_i)$, a partir da equação (6.18) e usando os resultados obtidas nas equações (6.14), (6.15) e (6.16). Assim, obtemos

$$\begin{aligned}
\frac{\partial \ell}{\partial a_i}(\lambda_i) &= \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \left(\frac{\partial P_{ji}}{\partial a_i}(\lambda_i) \right) \cdot \frac{1}{P_{ji} \cdot Q_{ji}} \\
&\stackrel{[6.14]}{=} \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \left(D(\theta_j - b_i)(1 - c_i) \cdot \Psi_{ji}(1 - \Psi_{ji}) \right) \cdot \frac{1}{P_{ji} Q_{ji}} \\
&\stackrel{[6.9]}{=} D(1 - c_i) \cdot \sum_{j=1}^m (u_{ji} - P_{ji})(\theta_j - b_i) \cdot W_{ji}
\end{aligned} \tag{6.19}$$

$$\begin{aligned}
\frac{\partial \ell}{\partial b_i}(\lambda_i) &= \frac{\partial \log L}{\partial b_i}(\lambda_i) \\
&= \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \left(\frac{\partial P_{ji}}{\partial b_i} \right) \cdot \frac{1}{P_{ji} Q_{ji}} \\
&\stackrel{[6.15]}{=} \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot (-D a_i (1 - c_i) \cdot \Psi(1 - \Psi)) \cdot \frac{1}{P_{ji} Q_{ji}} \\
&\stackrel{[6.9]}{=} -D a_i (1 - c_i) \cdot \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot W_{ji}
\end{aligned} \tag{6.20}$$

$$\begin{aligned}
\frac{\partial \ell}{\partial c_i}(\lambda_i) &= \frac{\partial \log L(\lambda_i)}{\partial c_i} \\
&= \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \left(\frac{\partial P_{ji}}{\partial c_i} \right) \cdot \frac{1}{P_{ji} Q_{ji}} \\
&\stackrel{[6.16]}{=} \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot (1 - \Psi) \cdot \frac{1}{P_{ji} Q_{ji}} \\
&\stackrel{[6.9]}{=} \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji}}
\end{aligned} \tag{6.21}$$

que são chamadas de equações de estimação dos parâmetros.

Assim, como dito anteriormente, a estimativa de máxima verossimilhança conjunta $\hat{\lambda}_i = (\hat{a}_i, \hat{b}_i, \hat{c}_i)$ de $\lambda_i = (a_i, b_i, c_i)$ para cada i é dada pelo ponto crítico de $\ell(\lambda_i)$ que a maximiza, obtido fazendo $\nabla_{\lambda_i} \ell(\lambda_i) = \vec{0}_{3 \times 1}$. A existência desse ponto crítico já foi discutida anteriormente. No entanto, para que esse ponto crítico seja, de fato, um máximo local, é necessário que a matriz Hessiana $H\ell(\lambda_i)$ seja definida negativa, o que significa que todos os seus autovalores precisam ser estritamente

negativos. Existindo ponto de máximo, para determiná-lo, precisamos, então, resolver o sistema

$$\left\{ \begin{array}{l} D(1-c_i) \cdot \sum_{j=1}^m (u_{ji} - P_{ji})(\theta_j - b_i) \cdot W_{ji} = 0 \\ -Da_i(1-c_i) \cdot \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot W_{ji} = 0 \\ \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji}} = 0 \end{array} \right.$$

cujas equações podem ser vistas como funções implícitas que envolvem os dados observados como umas das variáveis (pois variam de item para item), sendo as demais variáveis os parâmetros $\lambda_i = (a_i, b_i, c_i)$ dos itens, uma vez que, nesse caso, os θ_j possuem valores arbitrário e D é uma constante. Logo, cada uma dessas funções compõe o estimador de máxima verossimilhança dos parâmetros dos itens.

No entanto, a resolução desse sistema pode não ser uma tarefa simples de ser feita usando métodos de resolução convencionais.

Na maioria das vezes, as condições obtidas das derivadas de primeira ordem para o problema de maximização da função de verossimilhança não permitem a obtenção de uma solução explícita para os estimadores em questão. O sistema de equações gerado é quase sempre não linear, obrigando que a maximização seja feita por algum processo numérico. Os procedimentos de otimização numéricos funcionam de forma recursiva, sendo os valores dos parâmetros na etapa $k+1$ uma função de seus valores na etapa k . O algoritmo numérico consiste em tentar um valor para cada parâmetro e, depois, corrigi-lo sucessivamente até que algum critério de convergência seja atendido, quando, então, chega-se a um máximo para a função de verossimilhança. Em alguns casos, quando não ocorre convergência, o processo de iteração tem de ser interrompido depois de um número de iterações preestabelecido (Rabelo, 2013, p. 163).

A seguir, será descrito como o método de Newton-Raphson pode ser aplicado para estimar o valor dos parâmetros de cada item no ML3P da TRI na resolução desse sistema.

6.5 Fórmula de recorrência dos parâmetros por Newton-Raphson

Para resolver $\nabla_{\lambda_i} \ell(\lambda_i) = \vec{0}_{3 \times 1}$ e encontrar $\hat{\lambda}_i = (\hat{a}_i, \hat{b}_i, \hat{c}_i)$, estimativa do valor exato $\lambda_i^* = (a_i^*, b_i^*, c_i^*)$ dos parâmetros $\lambda_i = (a_i, b_i, c_i)$ de cada item que maximiza $\ell(\Lambda)$, sendo $\Lambda = (\lambda_1, \dots, \lambda_n)$, pode-se usar o método de Newton-Raphson em uma aproximação da função ℓ obtida por expansão de Taylor.

Conforme Arenales e Darezzo (2015), a ideia por trás desse método é aproximar $\ell(\Lambda)$ por uma função $q(\Lambda)$ em torno de um ponto inicial $\hat{\Lambda}^{(0)} = (\hat{\lambda}_1^{(0)}, \dots, \hat{\lambda}_n^{(0)}) \in D \subset \mathbb{R}^{3n}$, com $\hat{\lambda}_i^{(0)} = (\hat{a}_i^{(0)}, \hat{b}_i^{(0)}, \hat{c}_i^{(0)})$, sendo $\hat{\Lambda}^{(0)}$ uma estimativa grosseira do ponto de máximo exato $\Lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$ de ℓ , com $\lambda_i^* = (a_i^*, b_i^*, c_i^*)$ e D uma restrição do domínio de ℓ na qual há grande probabilidade de se encontrar esse ponto de máximo. O ponto de máximo $\hat{\Lambda}^{(1)} = (\hat{\lambda}_1^{(1)}, \dots, \hat{\lambda}_n^{(1)})$ de $q(\Lambda)$ nessa restrição, com $\hat{\lambda}_i^{(1)} = (a_i^{(1)}, b_i^{(1)}, c_i^{(1)})$, fornece uma atualização para a aproximação da estimativa inicial do ponto de máximo exato Λ^* de $\ell(\Lambda)$ com a vantagem de ser mais conveniente de calcular a aproximação $\hat{\Lambda}^{(1)}$ do que calcular Λ^* diretamente por $\nabla_{\lambda_i} \ell(\lambda_i) = \vec{0}_{3 \times 1}$, computacionalmente falando.

Então, esse novo ponto $\hat{\Lambda}^{(1)}$ é usado para obter uma melhor aproximação de ℓ por q e, dessa forma, obter uma melhor estimativa $\hat{\Lambda}^{(2)}$ para o ponto de máximo Λ^* de ℓ . Esse procedimento se repete até que algum critério de parada seja atingido, como, por exemplo, atingir certo número de iterações ou a diferença entre a estimativa para a outra ficar dentro de uma certa tolerância.

No nosso caso, para aplicar esse método, vamos expandir a função $\ell(\Lambda)$ $\ell(\Lambda)$ por série de Taylor até a derivada segunda na vizinhança de um ponto inicial $\hat{\Lambda}^{(k)} \in \mathbb{R}^{3n}$, obtendo o polinômio de Taylor de segundo grau

$$q(\Lambda) = \ell(\hat{\Lambda}^{(k)}) + \nabla \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)}) + \frac{1}{2} (\Lambda - \hat{\Lambda}^{(k)})^T \cdot H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)}) ,$$

que se aproxima de $\ell(\Lambda)$ na vizinhança de $\hat{\Lambda}^{(k)}$, ou seja, $\ell(\Lambda) \approx q(\Lambda)$ para $\Lambda \approx \hat{\Lambda}^{(k)}$, em que

$$\nabla \ell(\hat{\Lambda}^{(k)}) = \begin{bmatrix} \nabla_{\lambda_1} \ell(\hat{\Lambda}^{(k)}) \\ \nabla_{\lambda_2} \ell(\hat{\Lambda}^{(k)}) \\ \vdots \\ \nabla_{\lambda_i} \ell(\hat{\Lambda}^{(k)}) \\ \vdots \\ \nabla_{\lambda_n} \ell(\hat{\Lambda}^{(k)}) \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial \ell}{\partial a_1}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial b_1}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial c_1}(\hat{\Lambda}^{(k)}) \right) \\ \left(\frac{\partial \ell}{\partial a_2}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial b_2}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial c_2}(\hat{\Lambda}^{(k)}) \right) \\ \vdots \\ \left(\frac{\partial \ell}{\partial a_i}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial b_i}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial c_i}(\hat{\Lambda}^{(k)}) \right) \\ \vdots \\ \left(\frac{\partial \ell}{\partial a_n}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial b_n}(\hat{\Lambda}^{(k)}), \frac{\partial \ell}{\partial c_n}(\hat{\Lambda}^{(k)}) \right) \end{bmatrix}_{3n \times 1}$$

é o gradiente global de ℓ (ou a matriz Jacobiana de ℓ) avaliado no ponto $\hat{\Lambda}^{(k)}$ e a matriz

$$H\ell(\hat{\Lambda}^{(k)}) = \nabla^2 \ell(\hat{\Lambda}^{(k)}) = \nabla_{\lambda_i} (\nabla_{\lambda_{i'}} \ell(\hat{\Lambda}^{(k)}))^T = \left[\frac{\partial^2}{\partial \lambda_i \partial \lambda_{i'}^T} (\hat{\Lambda}^{(k)}) \right]_{n \times n},$$

com $i, i' = 1, 2, \dots, n$, é a matriz Hessiana global de ℓ avaliada no ponto $\hat{\Lambda}^{(k)}$, dada por

$$H\ell(\hat{\Lambda}^{(k)}) = \begin{bmatrix} h_{11} \ell(\hat{\Lambda}^{(k)}) & h_{12} \ell(\hat{\Lambda}^{(k)}) & \dots & h_{1n} \ell(\hat{\Lambda}^{(k)}) \\ h_{21} \ell(\hat{\Lambda}^{(k)}) & h_{22} \ell(\hat{\Lambda}^{(k)}) & \dots & h_{2n} \ell(\hat{\Lambda}^{(k)}) \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} \ell(\hat{\Lambda}^{(k)}) & h_{n2} \ell(\hat{\Lambda}^{(k)}) & \dots & h_{nn} \ell(\hat{\Lambda}^{(k)}) \end{bmatrix}$$

em que cada bloco $h_{ii'} \ell(\hat{\Lambda}^{(k)}) \in \mathbb{R}^{3 \times 3}$ é definido por

$$h_{ii'} \ell(\hat{\Lambda}^{(k)}) = \frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_{i'}^T} (\hat{\Lambda}^{(k)}) = \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_i \partial a_{i'}}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial b_{i'}}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial c_{i'}}(\hat{\Lambda}^{(k)}) \\ \frac{\partial^2 \ell}{\partial b_i \partial a_{i'}}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial b_{i'}}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial c_{i'}}(\hat{\Lambda}^{(k)}) \\ \frac{\partial^2 \ell}{\partial c_i \partial a_{i'}}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial b_{i'}}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial c_{i'}}(\hat{\Lambda}^{(k)}) \end{bmatrix},$$

sendo $\frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_{i'}^T}(\Lambda)$ a derivada de ordem 2 de ℓ calculada primeiro em relação ao vetor $\lambda_{i'}^T$, que é o vetor transposto de $\lambda_{i'}$, e, depois, em relação ao vetor λ_i .

Note que, $h_{ii'}\ell(\hat{\Lambda}^{(k)}) \neq 0_{3 \times 3}$ apenas se $i = i'$, caso contrário, $h_{ii'} = 0_{3 \times 3}$. De fato, considere dois parâmetros quaisquer r_i e $s_{i'}$ de ℓ , com $r_i \in (a_i, b_i, c_i) = \lambda_i$ e $s_{i'} \in (a_{i'}, b_{i'}, c_{i'}) = \lambda_{i'}$ tais que $i \neq i'$. Lembrando que $\ell(\Lambda)$ é função de $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ e de $\Theta = (\theta_1, \theta_2, \dots, \theta_m)$, mas que, nesse procedimento, os componentes θ_j de Θ foram grosseiramente estimados inicialmente, então $\frac{\partial \ell}{\partial s_{i'}}(\Lambda^{(k)})$ independe de r_i , de forma que $\frac{\partial \ell}{\partial s_{i'}}(\Lambda^{(k)})$ é tratada como constante na derivação com relação a r_i . Dessa forma, temos

$$\frac{\partial^2 \ell}{\partial r_i \partial s_{i'}}(\Lambda^{(k)}) = \frac{\partial}{\partial r_i} \left(\frac{\partial \ell}{\partial s_{i'}}(\Lambda^{(k)}) \right) = 0$$

e, portanto, cada bloco $h_{ii'}\ell(\hat{\Lambda}^{(k)})$ da Hessiana $H\ell(\hat{\Lambda}^{(k)})$ em que $i \neq i'$ é dado por

$$h_{ii'}\ell(\hat{\Lambda}^{(k)}) = \frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_{i'}^T}(\hat{\Lambda}^{(k)}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0_{3 \times 3}$$

e, assim,

$$H\ell(\hat{\Lambda}^{(k)}) = \begin{bmatrix} h_{11}\ell(\hat{\Lambda}^{(k)}) & 0 & \dots & 0 \\ 0 & h_{22}\ell(\hat{\Lambda}^{(k)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_{nn}\ell(\hat{\Lambda}^{(k)}) \end{bmatrix}$$

o que mostra que, além do gradiente $\nabla_{\lambda_i} \ell(\Lambda)$ ser separável por itens, a Hessiana global $H\ell(\hat{\Lambda}^{(k)})$ também é, sendo $h_{ii}\ell(\hat{\Lambda}^{(k)})$ a matriz Hessiana local do item i , que denotaremos a partir de agora por $h_i\ell(\hat{\Lambda}^{(k)})$.

Nesse sentido, as Hessianas locais $h_i\ell(\hat{\Lambda}^{(k)})$ de cada item são funções apenas de λ_i , não dependendo de $\lambda_{i'}$, para $i' \neq i$, uma vez que os parâmetros de $\lambda_{i'}$ são tratados como constantes na derivação com respeito aos parâmetros de λ_i . Assim, podemos escrever, de forma simplificada, a Hessiana local do item i $h_i\ell(\hat{\Lambda}^{(k)})$ como $h_i\ell(\hat{\lambda}_i^{(k)})$, isto é,

$$h_i\ell(\hat{\Lambda}^{(k)}) = \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_i \partial a_i}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial b_i}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial c_i}(\hat{\Lambda}^{(k)}) \\ \frac{\partial^2 \ell}{\partial b_i \partial a_i}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial b_i}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial c_i}(\hat{\Lambda}^{(k)}) \\ \frac{\partial^2 \ell}{\partial c_i \partial a_i}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial b_i}(\hat{\Lambda}^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial c_i}(\hat{\Lambda}^{(k)}) \end{bmatrix} =$$

$$\begin{bmatrix} \frac{\partial^2 \ell}{\partial a_i \partial a_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial b_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial c_i}(\hat{\lambda}_i^{(k)}) \\ \frac{\partial^2 \ell}{\partial b_i \partial a_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial b_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial c_i}(\hat{\lambda}_i^{(k)}) \\ \frac{\partial^2 \ell}{\partial c_i \partial a_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial b_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial c_i}(\hat{\lambda}_i^{(k)}) \end{bmatrix} = h_i\ell(\hat{\lambda}_i^{(k)})$$

Isso ocorre porque a log-verossimilhança dos itens é independente localmente, fruto do princípio da independência local dos itens da TRI. Assim, os parâmetros de um item não afetam os demais, não havendo interação cruzada entre derivadas de diferentes itens. Como consequência, ao focarmos diretamente na derivada $\nabla_{\lambda_i} \ell(\lambda_i^{(k)})$, evitamos construir e manipular a matriz Hessiana global $H\ell(\hat{\Lambda}^{(k)})$ de dimensão $3n \times 3n$. Isso será especialmente útil na aplicação do algoritmo de Newton-Raphson, que poderá operar no nível dos itens com base nas Hessianas locais dos itens $h_i\ell(\hat{\lambda}_i^{(k)})$ de cada item, simplificando tanto o cálculo quanto a interpretação dos resultados.

A partir daí, para determinar o ponto de máximo de ℓ , inicialmente derivamos com relação a Λ , a função

$$q(\Lambda) = \ell(\hat{\Lambda}^{(k)}) + \nabla \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)}) + \frac{1}{2} (\Lambda - \hat{\Lambda}^{(k)})^T \cdot H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)}) ,$$

que foi obtida por expansão de Taylor de segunda ordem da função ℓ em torno de $\hat{\Lambda}^{(k)}$, e obtemos

$$\begin{aligned} \nabla q(\Lambda) &= \vec{0}_{3n \times 1} + \nabla \ell(\hat{\Lambda}^{(k)}) + H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)}) \\ &= \nabla \ell(\hat{\Lambda}^{(k)}) + H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)}) . \end{aligned}$$

Igualando a zero para calcular o ponto crítico em questão, temos

$$\begin{aligned} \nabla \ell(\hat{\Lambda}^{(k)}) + H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)})^T &= \nabla q(\Lambda) \\ \nabla \ell(\hat{\Lambda}^{(k)}) + H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)})^T &= \vec{0}_{3n} \\ H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)})^T &= -\nabla \ell(\hat{\Lambda}^{(k)}) \\ [H \ell(\hat{\Lambda}^{(k)})]^{-1} \cdot H \ell(\hat{\Lambda}^{(k)}) \cdot (\Lambda - \hat{\Lambda}^{(k)})^T &= -[H \ell(\hat{\Lambda}^{(k)})]^{-1} \cdot \nabla \ell(\hat{\Lambda}^{(k)}) \\ \Lambda &= \hat{\Lambda}^{(k)} - [H \ell(\hat{\Lambda}^{(k)})]^{-1} \cdot \nabla \ell(\hat{\Lambda}^{(k)}) \end{aligned}$$

de onde encontramos

$$\Lambda_{(3n \times 1)}^{(k+1)} = \hat{\Lambda}_{(3n \times 1)}^{(k)} - [H \ell(\hat{\Lambda}^{(k)})]_{(3n \times 3n)}^{-1} \cdot [\nabla \ell(\hat{\Lambda}^{(k)})]_{(3n \times 1)}$$

que é a fórmula recursiva global do método de Newton-Raphson aplicado à log-verossimilhança ℓ . O termo $d_k = -[H \ell(\hat{\Lambda}^{(k)})]^{-1} \cdot \nabla \ell(\hat{\Lambda}^{(k)})$ indica a direção de busca, ou seja, a direção em que a função mais cresce.

Agora, usando a separabilidade por itens de ℓ , podemos aplicar localmente o método de Newton-Raphson, isto é, de forma separada para cada vetor λ_i de parâmetros de um item em específico, obtendo

$$\hat{\lambda}_i^{(k+1)} = \hat{\lambda}_i^{(k)} - \left[h_i \ell(\hat{\lambda}_i^{(k)}) \right]^{-1} \cdot \nabla_{\lambda_i} \ell(\hat{\lambda}_i^{(k)})$$

Então, nessa abordagem a nível de item, temos

$$\begin{aligned} \ell: D \subset \mathbb{R}^3 &\rightarrow \mathbb{R} \\ \lambda_i &\mapsto \ell(\lambda_i) \end{aligned}$$

sendo a log-verossimilhança com $\lambda_i = (a_i, b_i, c_i)$, além de

$$\begin{aligned} \nabla_{\lambda_i} \ell: D \subset \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ \lambda_i &\mapsto \nabla_{\lambda_i} \ell(\lambda_i) \end{aligned}$$

sendo o gradiente de ℓ dado por

$$\nabla_{\lambda_i} \ell(\lambda_i) = \left(\frac{\partial \ell}{\partial a_i}(\lambda_i), \frac{\partial \ell}{\partial b_i}(\lambda_i), \frac{\partial \ell}{\partial c_i}(\lambda_i) \right)$$

e

$$h_i \ell(\lambda_i) = \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_i \partial a_i}(\lambda_i) & \frac{\partial^2 \ell}{\partial a_i \partial b_i}(\lambda_i) & \frac{\partial^2 \ell}{\partial a_i \partial c_i}(\lambda_i) \\ \frac{\partial^2 \ell}{\partial b_i \partial a_i}(\lambda_i) & \frac{\partial^2 \ell}{\partial b_i \partial b_i}(\lambda_i) & \frac{\partial^2 \ell}{\partial b_i \partial c_i}(\lambda_i) \\ \frac{\partial^2 \ell}{\partial c_i \partial a_i}(\lambda_i) & \frac{\partial^2 \ell}{\partial c_i \partial b_i}(\lambda_i) & \frac{\partial^2 \ell}{\partial c_i \partial c_i}(\lambda_i) \end{bmatrix}$$

sendo a Hessiana local de ℓ do item i aplicada ao ponto λ_i e $[h_i \ell(\lambda_i)]^{-1}$ a sua inversa.

6.6 A Hessiana da função log-verossimilhança dos parâmetros

Usando o método de Newton-Raphson na log-verossimilhança local dos parâmetros dos itens $\ell(\lambda_i)$, encontramos a fórmula de recorrência

$$\hat{\lambda}_i^{(k+1)} = \hat{\lambda}_i^{(k)} - \left[h_i \ell(\hat{\lambda}_i^{(k)}) \right]^{-1} \cdot \nabla_{\lambda_i} \ell(\hat{\lambda}_i^{(k)}) \Rightarrow$$

$$\begin{bmatrix} a_i^{(k+1)} \\ b_i^{(k+1)} \\ c_i^{(k+1)} \end{bmatrix} = \begin{bmatrix} a_i^{(k)} \\ b_i^{(k)} \\ c_i^{(k)} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_i \partial a_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial b_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial a_i \partial c_i}(\hat{\lambda}_i^{(k)}) \\ \frac{\partial^2 \ell}{\partial b_i \partial a_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial b_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial b_i \partial c_i}(\hat{\lambda}_i^{(k)}) \\ \frac{\partial^2 \ell}{\partial c_i \partial a_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial b_i}(\hat{\lambda}_i^{(k)}) & \frac{\partial^2 \ell}{\partial c_i \partial c_i}(\hat{\lambda}_i^{(k)}) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{\partial \ell}{\partial a_i}(\hat{\lambda}_i^{(k)}) \\ \frac{\partial \ell}{\partial b_i}(\hat{\lambda}_i^{(k)}) \\ \frac{\partial \ell}{\partial c_i}(\hat{\lambda}_i^{(k)}) \end{bmatrix}$$

para determinar uma estiva para os parâmetros $\lambda_i = (a_i, b_i, c_i)$ do item i . Como $\hat{\lambda}_i^{(k)}$ é um valor conhecido de λ_i na iteração k , para a iteração ser realizada e o valor de $\hat{\lambda}_i^{(k+1)}$ ser estimado, é necessário saber como determinar a Hessiana local $h_i \ell(\hat{\lambda}_i^{(k)})$ e sua inversa $\left[h_i \ell(\hat{\lambda}_i^{(k)}) \right]^{-1}$, bem como o gradiente $\nabla_{\lambda_i} \ell(\hat{\lambda}_i^{(k)})$.

Lembrando que, no nosso problema, apenas as respostas u_{ji} são conhecidas e que θ_{ji} é inicialmente arbitrado, $\nabla_{\lambda_i} \ell$ depende apenas de λ_i para ser valorado, cujo valor $\lambda_i = \hat{\lambda}_i^{(k)}$ será atribuído a cada iteração para encontrar $\hat{\lambda}_i^{(k+1)}$. De fato, a expressão geral do gradiente $\nabla_{\lambda_i} \ell(\lambda_i) = \left(\frac{\partial \ell}{\partial a_i}(\lambda_i), \frac{\partial \ell}{\partial b_i}(\lambda_i), \frac{\partial \ell}{\partial c_i}(\lambda_i) \right)$, que em sua forma matricial, é a matriz Jacobiana local da função ℓ aplicada ao ponto λ_i , foi obtida na equação (6.18) e é dada por

$$\nabla_{\lambda_i} \ell(\lambda_i) = \sum_{j=1}^m (u_{ji} - P_{ji}(\lambda_i)) \cdot \frac{1}{P_{ji}(\lambda_i) \cdot Q_{ji}(\lambda_i)} \cdot \nabla_{\lambda_i} P_{ji}(\lambda_i) \quad ,$$

cujos componentes, obtidos nas equações (6.19), (6.20) e (6.21), são

- i. $\frac{\partial \ell}{\partial a_i}(\lambda_i) = D(1 - c_i) \cdot \sum_{j=1}^m (u_{ji} - P_{ji})(\theta_j - b_i) \cdot W_{ji} \quad ,$
- ii. $\frac{\partial \ell}{\partial b_i}(\lambda_i) = -D a_i (1 - c_i) \cdot \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot W_{ji} \quad \text{e}$
- iii. $\frac{\partial \ell}{\partial c_i}(\lambda_i) = \sum_{j=1}^m (u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji}} \quad ,$

sendo chamadas de equações de estimação dos parâmetros, todas dependendo apenas de $\lambda_i = (a_i, b_i, c_i)$.

Além disso,

$$\nabla_{\lambda_i} P_{ji}(\lambda_i) = \left(\frac{\partial P_{ji}}{\partial a_i}(\lambda_i), \frac{\partial P_{ji}}{\partial b_i}(\lambda_i), \frac{\partial P_{ji}}{\partial c_i}(\lambda_i) \right)$$

é o gradiente de P_{ji} no ponto λ_{ji} , que em sua forma matricial, é a matriz Jacobiana local da função de probabilidade do ML3P do item i aplicada ao ponto λ_i , cujos componentes, obtidos pelas equações (6.14), (6.15) e (6.16), são dados por

$$i. \quad \frac{\partial P_{ji}}{\partial a_i}(\lambda_i) = D(\theta_j - b_i)(1 - c_i) \cdot \Psi_{ji} \cdot (1 - \Psi_{ji}) ,$$

$$ii. \quad \frac{\partial P_{ji}}{\partial b_i}(\lambda_i) = -D a_i (1 - c_i) \cdot \Psi_{ji} \cdot (1 - \Psi_{ji}) \quad \text{e}$$

$$iii. \quad \frac{\partial P_{ji}}{\partial c_i}(\lambda_i) = 1 - \Psi_{ji} ,$$

e, assim, todas as expressões de seus componentes também dependem apenas de $\lambda_i = (a_i, b_i, c_i)$. Dessa forma, $\nabla_{\lambda_i} \ell(\lambda_i)$ está completamente determinado, pois é, realmente, depende apenas de $\lambda_i = (a_i, b_i, c_i)$ na situação estudada.

Podemos, ainda, fazer um ajuste na expressão do gradiente $\nabla_{\lambda_i} \ell(\lambda_i)$ da equação (6.18), que pode ser reescrita, de forma mais simples, como

$$\begin{aligned} \nabla_{\lambda_i} \ell(\lambda_i) &= \sum_{j=1}^m (u_{ji} - P_{ji}(\lambda_i)) \cdot \frac{1}{P_{ji}(\lambda_i) \cdot Q_{ji}(\lambda_i)} \cdot \nabla_{\lambda_i} P_{ji}(\lambda_i) \\ &\stackrel{[6.9]}{=} \sum_{j=1}^m (u_{ji} - P_{ji}(\lambda_i)) \cdot \frac{W_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \cdot \nabla_{\lambda_i} P_{ji}(\lambda_i) \\ &= \sum_{j=1}^m (u_{ji} - P_{ji}(\lambda_i)) \cdot W_{ji} \cdot \frac{\nabla_{\lambda_i} P_{ji}(\lambda_i)}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \end{aligned}$$

e, fazendo $N_{ji} = \frac{\nabla_{\lambda_i} P_{ji}(\lambda_i)}{\Psi_{ji} \cdot (1 - \Psi_{ji})}$, obtemos

$$\nabla_{\lambda_i} \ell(\lambda_i) = \sum_{j=1}^m (u_{ji} - P_{ji}(\lambda_i)) \cdot W_{ji} \cdot h_{ij} \quad , \quad (6.22)$$

de forma que, ainda, $\nabla_{\lambda_i} \ell(\lambda_i)$ depende apenas de $\lambda_i = (a_i, b_i, c_i)$.

No entanto, para usarmos a fórmula de recorrência dos parâmetros e determinar estimativas para $\lambda_i = (a_i, b_i, c_i)$, ainda precisamos obter uma expressão para a Hessiana local $h\ell(\lambda_i)$ para, posteriormente, calcularmos $[h\ell(\lambda_i)]^{-1}$. E é isso o que faremos agora. Para isso, apresentar alguns resultados que serão utilizados nesse processo.

Inicialmente, para simplificar a escrita, faremos $P_{ji} = P_{ji}(\lambda_i)$, $Q_{ji} = Q_{ji}(\lambda_i)$, $\Psi_{ji} = \Psi_{ji}(z_{ji})$, $v_{ji} = v_{ji}(\lambda_i)$ e $\ell = \ell(\lambda_i)$. Assim, veja que

$$\begin{aligned} \nabla_{\lambda_i} (P_{ji} \cdot Q_{ji}) &= (\nabla_{\lambda_i} P_{ji}) \cdot Q_{ji} + P_{ji} \cdot (\nabla_{\lambda_i} Q_{ji}) \\ &\stackrel{[6.1]}{=} (\nabla_{\lambda_i} P_{ji}) \cdot (1 - P_{ji}) + P_{ji} \cdot (\nabla_{\lambda_i} Q_{ji}) \\ &\stackrel{[6.17]}{=} \nabla_{\lambda_i} P_{ji} - (\nabla_{\lambda_i} P_{ji}) \cdot P_{ji} - P_{ji} \cdot (\nabla_{\lambda_i} P_{ji}) \\ &= (1 - 2P_{ji}) \cdot (\nabla_{\lambda_i} P_{ji}) \quad . \end{aligned} \quad (6.23)$$

Agora, fazendo $v_{ji} = \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}}$ e observando que $u_{ji} = u_{ji}^2$ e que $P_{ji}^2 = 2P_{ji}^2 - P_{ji}^2 + P_{ji} - P_{ji}$, temos

$$\begin{aligned}
v_{ji}^2 &= \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \cdot (u_{ji}^2 - 2P_{ji} \cdot u_{ji} + P_{ji}^2) \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \cdot (u_{ji} - 2P_{ji} \cdot u_{ji} + P_{ji}^2) \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \cdot (u_{ji} - 2P_{ji} \cdot u_{ji} + 2P_{ji}^2 - P_{ji}^2 + P_{ji} - P_{ji}) \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \cdot (P_{ji} - P_{ji}^2 + u_{ji} - 2P_{ji} \cdot u_{ji} - P_{ji} + 2P_{ji}^2) \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \cdot [P_{ji} \cdot (1 - P_{ji}) + (u_{ji} - P_{ji}) \cdot (1 - 2P_{ji})] \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \cdot [P_{ji}Q_{ji} + (u_{ji} - P_{ji}) \cdot (1 - 2P_{ji})]
\end{aligned} \tag{6.24}$$

Além disso,

$$\begin{aligned}
\nabla_{\lambda_{ji}} v_{ji} &= \nabla_{\lambda_{ji}} \left(\frac{u - P_{ji}}{P_{ji}Q_{ji}} \right) \\
&= \frac{-\nabla_{\lambda_{ji}} P_{ji} \cdot P_{ji}Q_{ji} - (u - P_{ji}) \cdot \nabla_{\lambda_{ji}} (P_{ji}Q_{ji})}{(P_{ji}Q_{ji})^2} \\
&= -\left\{ \frac{1}{(P_{ji}Q_{ji})^2} \cdot [\nabla_{\lambda_{ji}} P_{ji} \cdot P_{ji}Q_{ji} + (u - P_{ji}) \cdot \nabla_{\lambda_{ji}} (P_{ji}Q_{ji})] \right\} \\
&\stackrel{[6.23]}{=} -\left\{ \frac{1}{(P_{ji}Q_{ji})^2} \cdot [\nabla_{\lambda_{ji}} P_{ji} \cdot P_{ji}Q_{ji} + (u - P_{ji}) \cdot (1 - 2P_{ji}) \cdot \nabla P_{ji}] \right\} \\
&= -\left\{ \frac{1}{(P_{ji}Q_{ji})^2} \cdot [P_{ji}Q_{ji} + (u - P_{ji}) \cdot (1 - 2P_{ji})] \right\} \cdot \nabla_{\lambda_{ji}} P_{ji} \\
&\stackrel{[6.22]}{=} -v_{ji}^2 \cdot \nabla_{\lambda_{ji}} P_{ji}
\end{aligned} \tag{6.25}$$

Com esses resultados preliminares, podemos determinar uma expressão para a Hessiana local $h_i^\ell(\lambda_i)$, obtendo

$$\begin{aligned}
h_i \ell(\lambda_i) &= \nabla_{\lambda_i}^2 \ell \\
&= \nabla_{\lambda_i} \left(\nabla_{\lambda_i} \ell \right)^T \\
&\stackrel{(6.18)}{=} \nabla_{\lambda_i} \left(\sum_{j=1}^m \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \cdot \nabla_{\lambda_i} P_{ji} \right)^T \\
&= \sum_{j=1}^m \nabla_{\lambda_i} \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \cdot \nabla_{\lambda_i} P_{ji} \right)^T \\
&= \sum_{j=1}^m \nabla_{\lambda_i} \left(v_{ji} \cdot \nabla_{\lambda_i} P_{ji} \right)^T \\
&= \sum_{j=1}^m \nabla_{\lambda_i} \left(v_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T \right) \\
&= \sum_{j=1}^m \left[\nabla v_{ji} \cdot \nabla_{\lambda_i} P_{ji}^T + v_{ji} \cdot \nabla_{\lambda_i} \left(\nabla_{\lambda_i} P_{ji} \right)^T \right] \\
&= \sum_{j=1}^m \left[\nabla v_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T + v_{ji} \cdot \nabla_{\lambda_i}^2 P_{ji} \right] \\
&\stackrel{(6.25)}{=} \sum_{j=1}^m \left[-v_{ji}^2 \cdot \nabla_{\lambda_i} P_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T + v_{ji} \cdot \nabla_{\lambda_i}^2 P_{ji} \right] \\
&= \sum_{j=1}^m \left[v_{ji} \cdot \nabla_{\lambda_i}^2 P_{ji} - v_{ji}^2 \cdot \nabla_{\lambda_i} P_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T \right] \\
&= \sum_{j=1}^m \left[\frac{u_{ji} - P_{ji}}{P_{ji} \cdot Q_{ji}} \cdot \nabla_{\lambda_i}^2 P_{ji} - \left(\frac{u_{ji} - P_{ji}}{P_{ji} \cdot Q_{ji}} \right)^2 \cdot \nabla_{\lambda_i} P_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T \right] \\
&= \sum_{j=1}^m \left[\left(u_{ji} - P_{ji} \right) \frac{1}{P_{ji} \cdot Q_{ji}} \cdot \nabla_{\lambda_i}^2 P_{ji} - \right. \\
&\quad \left. \left(u_{ji} - P_{ji} \right)^2 \left(\frac{1}{P_{ji} \cdot Q_{ji}} \right)^2 \cdot \nabla_{\lambda_i} P_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T \right] \\
&\stackrel{(6.9)}{=} \sum_{j=1}^m \left[\left(u_{ji} - P_{ji} \right) \frac{W_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \cdot \nabla_{\lambda_i}^2 P_{ji} - \right. \\
&\quad \left. \left(u_{ji} - P_{ji} \right)^2 \left(\frac{W_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \right)^2 \cdot \nabla_{\lambda_i} P_{ji} \cdot \left(\nabla_{\lambda_i} P_{ji} \right)^T \right] \\
&= \sum_{j=1}^m \left[W_{ji} \left(u_{ji} - P_{ji} \right) \frac{\nabla_{\lambda_i}^2 P_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} - \right. \\
&\quad \left. W_{ji}^2 \left(u_{ji} - P_{ji} \right)^2 \cdot \frac{\nabla_{\lambda_i} P_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \cdot \left(\frac{\nabla_{\lambda_i} P_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \right)^T \right]
\end{aligned}$$

que, fazendo $M_{ji} = \frac{\nabla_{\lambda_i}^2 P_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})}$ e $N_{ji} = \frac{\nabla_{\lambda_i} P_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})}$, fica

$$\begin{aligned} h_i \ell(\lambda_i) &= \sum_{j=1}^m \left[W_{ji} (u_{ji} - P_{ji}) \cdot M_{ji} - W_{ji}^2 (u_{ji} - P_{ji})^2 \cdot N_{ji} \cdot N_{ji}^T \right] \\ &= \sum_{j=1}^m \left\{ W_{ji} (u_{ji} - P_{ji}) \cdot \left[M_{ji} - (u_{ji} - P_{ji}) \cdot N_{ji} \cdot N_{ji}^T \right] \right\} \end{aligned}$$

Como W_{ji} e P_{ji} estão definidas respectivamente, pelas equações (6.9) e (4.1), todas dependentes de $\lambda_i = (a_i, b_i, c_i)$, para que $h_i \ell(\lambda_i)$ fique completamente determinada em função das variáveis conhecidas ou das variáveis objetivo do nosso problema de estimativação dos parâmetros dos itens, só precisamos calcular os termos N_{ji} e M_{ji} . Note que, como Ψ_{ji} foi definida pela equação (6.6), basta determinar $\nabla_{\lambda_i} P_{ji}(\lambda_i)$ e $\nabla_{\lambda_i}^2 P_{ji}(\lambda_i)$, que são, respectivamente, a matriz Jacobiana e a matriz Hessiana da função de probabilidade de acerto P_{ji} do ML3P aplicadas no ponto $\lambda_i = (a_i, b_i, c_i)$, e são dadas por

$$\nabla_{\lambda_i} P_{ji}(\lambda_i) = \begin{bmatrix} \frac{\partial P}{\partial a}(\lambda) & \frac{\partial P}{\partial b}(\lambda) & \frac{\partial P}{\partial c}(\lambda) \end{bmatrix}$$

e

$$\nabla_{\lambda_i}^2 P_{ji}(\lambda_i) = \begin{bmatrix} \frac{\partial^2 P_{ji}}{\partial a_i \partial a_i}(\lambda_i) & \frac{\partial^2 P_{ji}}{\partial a_i \partial b_i}(\lambda_i) & \frac{\partial^2 P_{ji}}{\partial a_i \partial c_i}(\lambda_i) \\ \frac{\partial^2 P_{ji}}{\partial b_i \partial a_i}(\lambda_i) & \frac{\partial^2 P_{ji}}{\partial b_i \partial b_i}(\lambda_i) & \frac{\partial^2 P_{ji}}{\partial b_i \partial c_i}(\lambda_i) \\ \frac{\partial^2 P_{ji}}{\partial c_i \partial a_i}(\lambda_i) & \frac{\partial^2 P_{ji}}{\partial c_i \partial b_i}(\lambda_i) & \frac{\partial^2 P_{ji}}{\partial c_i \partial c_i}(\lambda_i) \end{bmatrix}.$$

Para isso, vamos determinar uma expressão para cada elemento de N_{ji} e de M_{ji} . Vamos iniciar simplificando ainda mais a escrita, apenas por praticidade, denotando $\lambda_i = (a_i, b_i, c_i)$ por ν , θ_j por θ , u_{ji} por u , z_{ji} por z , $P_{ji}(\lambda_{ji})$ por P , $Q_{ji}(\lambda_{ji})$ por Q , $\Psi(z)$ por Ψ , $\Psi'(z) = \frac{\partial \Psi}{\partial z}(z)$ por Ψ' , $\Psi''(z) = \frac{\partial^2 \Psi}{\partial z \partial z}(z)$ por

Ψ'' , $v_{ji}(\lambda_i) = \frac{u-P}{PQ}$ por v e, finalmente, W_{ji} por W . Além disso, como de costume aqui, vamos apresentar alguns resultados preliminares que serão utilizados nesse para isso.

Vamos começar pela segunda derivada em relação à z , dada por

$$\begin{aligned}
 \Psi'' &= \frac{\partial^2 \Psi}{\partial z} \\
 &= \frac{\partial}{\partial z} \left(\frac{\partial \Psi}{\partial z} \right) \\
 &\stackrel{[6.8]}{=} \frac{\partial}{\partial z} [\Psi \cdot (1 - \Psi)] \\
 &= \Psi \cdot (1 - \Psi) \cdot (1 - \Psi) - \Psi \cdot [\Psi \cdot (1 - \Psi)] \\
 &= \Psi \cdot (1 - \Psi) \cdot (1 - 2\Psi) \quad . \tag{6.26}
 \end{aligned}$$

Já as derivadas parciais de primeira ordem de $z = Da(\theta - b)$ são dadas por

- i. $\frac{\partial z}{\partial a}(\theta, a, b) = D(\theta - b)$,
- ii. $\frac{\partial z}{\partial b}(\theta, a, b) = -Da$, (6.27)
- iii. $\frac{\partial z}{\partial c}(\theta, a, b) = 0$.

Escrevendo $\Psi(1 - \Psi)$ como $\frac{\partial \Psi}{\partial z}$ nas equações (6.14), (6.15) e (6.16), para as segundas derivadas encontramos:

$$\begin{aligned}
 \frac{\partial^2 P}{\partial a \partial a}(\lambda) &= \frac{\partial}{\partial a} \cdot \left(\frac{\partial P}{\partial a} \right) \\
 &\stackrel{[6.14]}{=} D(\theta - b)(1 - c) \cdot \left[\frac{\partial}{\partial a} \left(\frac{\partial \Psi}{\partial z} \right) \right] \\
 &= D(\theta - b)(1 - c) \cdot \frac{\partial^2 \Psi}{\partial z} \cdot \frac{\partial z}{\partial a} \\
 &\stackrel{[6.26]}{=} D(\theta - b)(1 - c) \cdot \Psi(1 - \Psi)(1 - 2\Psi) \cdot \frac{\partial z}{\partial a} \\
 &\stackrel{[6.27]}{=} D(\theta - b)(1 - c) \cdot \Psi(1 - \Psi)(1 - 2\Psi) \cdot D(\theta - b) \\
 &= D^2(\theta - b)^2(1 - c) \cdot \Psi(1 - \Psi)(1 - 2\Psi)
 \end{aligned}$$

como segunda derivada parcial pura de P em relação a a ,

$$\begin{aligned}
 \frac{\partial^2 P}{\partial b \partial b}(\lambda) &= \frac{\partial}{\partial b} \cdot \left(\frac{\partial P}{\partial b} \right) \\
 &\stackrel{[6.15]}{=} -Da(1-c) \cdot \left[\frac{\partial}{\partial b} \left(\frac{\partial \Psi}{\partial z} \right) \right] \\
 &= -Da(1-c) \cdot \frac{\partial^2 \Psi}{\partial z \partial b} \\
 &\stackrel{[6.26]}{=} -Da(1-c) \cdot \Psi(1-\Psi)(1-2\Psi) \cdot \frac{\partial z}{\partial b} \\
 &\stackrel{[6.27]}{=} -Da(1-c) \cdot \Psi(1-\Psi)(1-2\Psi) \cdot (-Da) \\
 &= D^2 a^2 (1-c) \cdot \Psi(1-\Psi)(1-2\Psi)
 \end{aligned}$$

como segunda derivada parcial pura de P em relação a b e

$$\begin{aligned}
 \frac{\partial^2 P}{\partial c \partial c}(\lambda) &= \frac{\partial}{\partial c} \cdot \left(\frac{\partial P}{\partial c} \right) \\
 &\stackrel{[6.16]}{=} \frac{\partial}{\partial c} (1-\Psi) \\
 &= -\frac{\partial \Psi}{\partial c} \\
 &= -\frac{\partial \Psi}{\partial c} \cdot \frac{\partial z}{\partial c} \\
 &\stackrel{[6.27]}{=} -\frac{\partial \Psi}{\partial c} \cdot 0 \\
 &= 0
 \end{aligned}$$

como segunda derivada parcial pura de P em relação a c .

Finalmente, vamos calcular as segundas derivadas mistas de P . Note que a matriz Hessiana $h_i \ell(\lambda_i)$ é simétrica, ou seja, dados quaisquer dois parâmetros r e

s de ℓ , com $r, s \in (a_i, b_i, c_i) = \lambda_i$ e $r \neq s$, temos $\frac{\partial^2 \ell}{\partial r \partial s}(\lambda_i) = \frac{\partial^2 \ell}{\partial s \partial r}(\lambda_i)$, pois ℓ é

de classe C^2 . Assim, $\frac{\partial^2 P_{ji}}{\partial a_i \partial b_i}(\lambda_i) = \frac{\partial^2 P_{ji}}{\partial b_i \partial a_i}(\lambda_i)$, $\frac{\partial^2 P_{ji}}{\partial a_i \partial c_i}(\lambda_i) = \frac{\partial^2 P_{ji}}{\partial c_i \partial a_i}(\lambda_i)$ e

$\frac{\partial^2 P_{ji}}{\partial b_i \partial c_i}(\lambda_i) = \frac{\partial^2 P_{ji}}{\partial c_i \partial b_i}(\lambda_i)$. Dessa forma, temos

$$\begin{aligned}
\frac{\partial^2 P}{\partial a \partial b}(\lambda) &= \frac{\partial^2 P}{\partial b \partial a}(\lambda) \\
&= \frac{\partial}{\partial b} \left(\frac{\partial P}{\partial a} \right) \\
&\stackrel{[6.14]}{=} \frac{\partial}{\partial b} \left[(1-c) \cdot \frac{\partial \Psi}{\partial z} \cdot D(\theta-b) \right] \\
&= D(1-c) \cdot \frac{\partial}{\partial b} \left[\frac{\partial \Psi}{\partial z} \cdot (\theta-b) \right] \\
&= D(1-c) \cdot \left[\frac{\partial}{\partial b} \left(\frac{\partial \Psi}{\partial z} \right) \cdot (\theta-b) + \frac{\partial \Psi}{\partial z} \cdot (-1) \right] \\
&= D(1-c) \cdot \left[\frac{\partial^2 \Psi}{\partial z \partial b} \cdot \frac{\partial z}{\partial b} \cdot (\theta-b) - \frac{\partial \Psi}{\partial z} \right] \\
&\stackrel{[6.26]}{=} D(1-c) \cdot \left[\Psi(1-\Psi)(1-2\Psi) \cdot \frac{\partial z}{\partial b} \cdot (\theta-b) - \frac{\partial \Psi}{\partial z} \right] \\
&\stackrel{[6.27]}{=} D(1-c) \cdot \left[\Psi(1-\Psi)(1-2\Psi) \cdot (-Da) \cdot (\theta-b) - \frac{\partial \Psi}{\partial z} \right] \\
&\stackrel{[6.8]}{=} D(1-c) \cdot \left[\Psi(1-\Psi)(1-2\Psi) \cdot (-Da) \cdot (\theta-b) - \Psi(1-\Psi) \right] \\
&= -D(1-c) \cdot \Psi(1-\Psi) \left[(1-2\Psi) \cdot Da \cdot (\theta-b) + 1 \right]
\end{aligned}$$

como segunda derivada parcial mista de P em relação a a e b ,

$$\begin{aligned}
\frac{\partial^2 P}{\partial a \partial c}(\lambda) &= \frac{\partial^2 P}{\partial c \partial a}(\lambda) \\
&= \frac{\partial}{\partial c} \left(\frac{\partial P}{\partial a} \right) \\
&\stackrel{[6.14]}{=} \frac{\partial}{\partial c} \left[(1-c) \cdot \frac{\partial \Psi}{\partial z} \cdot D(\theta-b) \right] \\
&= D(\theta-b) \cdot \frac{\partial}{\partial c} \left[(1-c) \cdot \frac{\partial \Psi}{\partial z} \right] \\
&= D(\theta-b) \cdot \left[(-1) \cdot \frac{\partial \Psi}{\partial z} + (1-c) \cdot \frac{\partial}{\partial c} \left(\frac{\partial \Psi}{\partial z} \right) \right] \\
&= D(\theta-b) \cdot \left[-\frac{\partial \Psi}{\partial z} + (1-c) \cdot \frac{\partial^2 \Psi}{\partial z \partial c} \cdot \frac{\partial z}{\partial c} \right] \\
&\stackrel{[6.8]}{=} D(\theta-b) \cdot \left[-\Psi(1-\Psi) + (1-c) \cdot \frac{\partial^2 \Psi}{\partial z \partial c} \cdot \frac{\partial z}{\partial c} \right] \\
&\stackrel{[6.27]}{=} D(\theta-b) \cdot \left[-\Psi(1-\Psi) + (1-c) \cdot \frac{\partial^2 \Psi}{\partial z \partial c} \cdot 0 \right] \\
&= -D(\theta-b) \cdot \Psi(1-\Psi)
\end{aligned}$$

como segunda derivada parcial mista de P em relação a a e c ,

$$\begin{aligned}
 \frac{\partial^2 P}{\partial b \partial c}(\lambda) &= \frac{\partial^2 P}{\partial c \partial b}(\lambda) \\
 &= \frac{\partial}{\partial c} \left(\frac{\partial P}{\partial b} \right) \\
 &\stackrel{[6.15]}{=} \frac{\partial}{\partial c} \left[(1-c) \cdot \frac{\partial \Psi}{\partial z} \cdot (-Da) \right] \\
 &= -Da \cdot \frac{\partial}{\partial c} \left[(1-c) \cdot \frac{\partial \Psi}{\partial z} \right] \\
 &= -Da \cdot \left[(-1) \cdot \frac{\partial \Psi}{\partial z} + (1-c) \cdot \frac{\partial}{\partial c} \left(\frac{\partial \Psi}{\partial z} \right) \right] \\
 &= -Da \cdot \left[-\frac{\partial \Psi}{\partial z} + (1-c) \cdot \frac{\partial^2 \Psi}{\partial z^2} \cdot \frac{\partial z}{\partial c} \right] \\
 &\stackrel{[6.8]}{=} -Da \cdot \left[-\Psi(1-\Psi) + (1-c) \cdot \frac{\partial^2 \Psi}{\partial z^2} \cdot \frac{\partial z}{\partial c} \right] \\
 &\stackrel{[6.27]}{=} -Da \cdot \left[-\Psi(1-\Psi) + (1-c) \cdot \frac{\partial^2 \Psi}{\partial z^2} \cdot 0 \right] \\
 &= Da \cdot \Psi(1-\Psi)
 \end{aligned}$$

como segunda derivada parcial mista de P em relação a b e c .

Assim, a partir das primeiras e segundas derivadas de P , vemos que a Jacobiana $\nabla_{\lambda_i} P_{ji}(\lambda_i)$ e a Hessiana $\nabla_{\lambda_i}^2 P_{ji}(\lambda_i)$ são dadas por

$$\begin{aligned}
 \nabla_{\lambda_i} P_{ji}(\lambda_i) &= \\
 &= \begin{bmatrix} \frac{\partial P}{\partial a}(\lambda) & \frac{\partial P}{\partial b}(\lambda) & \frac{\partial P}{\partial c}(\lambda) \end{bmatrix} \\
 &= \begin{bmatrix} D(\theta-b)(1-c) \cdot \Psi \cdot (1-\Psi) & -Da(1-c) \cdot \Psi \cdot (1-\Psi) & 1-\Psi \end{bmatrix}
 \end{aligned}$$

e

$$\begin{aligned}
 \nabla_{\lambda_i}^2 P_{ji}(\lambda_i) &= \\
 &= \begin{bmatrix} D^2(\theta-b)^2(1-c) \cdot & -D(1-c) \cdot \Psi \cdot (1-\Psi) \cdot & -D(\theta-b) \cdot \Psi \cdot (1-\Psi) \cdot \\ \Psi(1-\Psi)(1-2\Psi) & [(1-2\Psi) \cdot Da \cdot (\theta-b) + 1] & \\ -D(1-c) \cdot \Psi \cdot (1-\Psi) \cdot & D^2 a^2 (1-c) \cdot \Psi \cdot (1-\Psi) \cdot (1-2\Psi) & Da \cdot \Psi \cdot (1-\Psi) \\ [(1-2\Psi) \cdot Da \cdot (\theta-b) + 1] & & \\ -D(\theta-b) \cdot \Psi \cdot (1-\Psi) & Da \cdot \Psi \cdot (1-\Psi) & 0 \end{bmatrix}
 \end{aligned}$$

de forma que N_{ji} e de M_{ji} podem ser obtidas por

$$N_{ji} = \frac{1}{\Psi(1-\Psi)} \cdot \nabla_{\lambda_i} P_{ji}(\lambda_i)$$

$$= \begin{bmatrix} D(\theta-b)(1-c) & -Da(1-c) & \frac{1}{\Psi} \end{bmatrix}$$

e

$$M_{ji} = \frac{1}{\Psi(1-\Psi)} \cdot \nabla_{\lambda_i}^2 P_{ji}(\lambda_i)$$

$$= \begin{bmatrix} D^2(\theta-b)^2(1-c)(1-2\Psi) & -D(1-c)[(1-2\Psi) \cdot Da \cdot (\theta-b) + 1] & -D(\theta-b) \\ -D(1-c)[(1-2\Psi) \cdot Da \cdot (\theta-b) + 1] & D^2a^2(1-c)(1-2\Psi) & Da \\ -D(\theta-b) & Da & 0 \end{bmatrix}$$

e dependem apenas de λ_i . Assim, a Hessiana local

$$h_i \ell(\lambda_i) = \sum_{j=1}^m \left[W_{ji} \cdot (u_{ji} - P_{ji}) \cdot (M_{ji} - W_{ji} \cdot (u_{ji} - P_{ji}) \cdot N_{ji} \cdot N_{ji}^T) \right]$$

está bem definida e também depende apenas de λ_i .

Logo, escolhendo um valor para θ_j , a Hessiana local $h_i \ell(\lambda_i)$ depende apenas dos parâmetros $\lambda_i = (a_i, b_i, c_i)$ dos itens, e sua inversa pode ser obtida por meio de algoritmo e procedimento computacional, bem como as interações de

$$\hat{\lambda}_i^{(k+1)} = \hat{\lambda}_i^{(k)} - \left[h_i \ell(\hat{\lambda}_i^{(k)}) \right]^{-1} \cdot \nabla_{\lambda_i} \ell(\hat{\lambda}_i^{(k)})$$

do algoritmo de Newton-Raphson aplicado na obtenção desses parâmetros.

Scoring das aptidões usando JMLE

Com uma primeira estimativa do valor dos parâmetros dos itens encontrada na primeira etapa da estimativa conjunta no processo *back-and-forth*, o próximo passo é fazer as primeiras estimativas das aptidões dos indivíduos. Para cada item respondido pelo testando, deve-se identificar o valor de θ que melhor explica seus acertos e erros ao longo do teste. Como cada item avalia um traço latente específico, sendo θ justamente a medida desse traço, ao estimar os valores de θ relacionados à todos os itens, torna-se possível, posteriormente, calcular a proficiência geral do indivíduo com base em seu desempenho na avaliação.

O termo *scoring* (ou *escoragem*, em português) refere-se ao processo de estimar o nível de aptidão ou proficiência de um indivíduo com base em suas respostas aos itens do teste, utilizando modelos estatísticos como a Teoria de Resposta ao Item. Assim, de forma análoga ao capítulo anterior, passaremos aqui a descrever o *scoring* das aptidões θ , dados os parâmetros a , b e c , por meio do seu Estimador de Máxima Verossimilhança (EMV). Nesse procedimento, θ passa a ter “papel de parâmetro”.

7.1 A verossimilhança e a log-verossimilhança das aptidões

Agora, vamos considerar o caso em que queremos obter a verossimilhança para estimar as aptidões dos testandos. Assim, vamos supor conhecidos os parâmetros dos itens, por exemplo, aproveitando itens já utilizados em outros testes ou a partir da primeira estimativa do processo de estimação conjunta dos parâmetros dos itens, descrito anteriormente.

Na seção 6.1, encontramos a função de verossimilhança dos parâmetros dos itens Λ supondo arbitrados valores de Θ em um procedimento de estimação conjunta, que englobava todos os parâmetros dos itens bem como os nível de aptidão de todos os testandos. Assim, é fácil ver que a expressão da função de verossimilhança para as habilidades é a mesma da função de verossimilhança para os parâmetros dos itens, porém em função das aptidões Θ , e não de Λ , com

$\Theta = (\theta_1, \dots, \theta_j, \dots, \theta_m)$ e $j = 1, 2, \dots, m$, uma vez que, agora, são conhecidos os parâmetros dos itens. Porém, a verossimilhança $L(\Theta)$ ainda está associada aos valores amostrais observados U , no ML3P, no sentido de ser utilizada para obter o valor dos parâmetros desconhecidos, no caso Θ , que maximizam a probabilidade de se obter a amostra particular U observada, ou seja, o valor que torna aquela amostra a “mais provável”. Dessa forma,

$$\begin{aligned} L(\Theta) &= L(\Theta | U) \\ &= P(X = U | \Lambda, \Theta) \\ &= \prod_{j=1}^m \prod_{i=1}^n P_{ji}^{u_{ji}} \cdot Q_{ji}^{1-u_{ji}} \end{aligned} \tag{7.1}$$

é a função de verossimilhança conjunta $L(\Theta | U)$, ou simplesmente $L(\Theta)$, das habilidades, dado o padrão de respostas U dos testados.

Da mesma maneira, a expressão da função de verossimilhança das habilidades será a mesma que a da verossimilhança das aptidões obtida na equação (6.3), porém, em função de Θ , ou seja,

$$\ell(\Theta) = \sum_{j=1}^m \sum_{i=1}^n (u_{ji} \log P_{ji} + (1-u_{ji}) \log Q_{ji}) \tag{7.2}$$

7.2 Separabilidade por indivíduos da log-verossimilhança $\ell(\theta)$

Note que, ao fixar $j = t$, temos

$$\frac{\partial}{\partial \theta_j} (u_{ji} \log P_{ji} + (1-u_{ji}) \log Q_{ji}) = 0, \quad \forall j \neq t, \tag{7.3}$$

pois, nessa situação, somente os termos em t variam e os demais termos em $j \neq t$ são constantes em relação a t , de forma que

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_t}(\Theta) &= \frac{\partial \ell}{\partial \theta_t} \sum_{j=1}^m (u_{ji} \log P_{ji} + (1-u_{ji}) \log Q_{ji}) = \\ &= \frac{\partial \ell}{\partial \theta_t} (u_{ti} \log P_{ti} + (1-u_{ti}) \log Q_{ti}) = \frac{\partial \ell_t}{\partial \theta_t}(\Theta) \end{aligned}$$

com $\frac{\partial \ell_t}{\partial \theta_t}(\Theta)$ sendo a derivada da log-verossimilhança individual associada ao indivíduo t , que é um dos componentes da derivada da log-verossimilhança global $\frac{\partial \ell}{\partial \theta_t}(\Theta)$. O resultado acima reflete o princípio da independência individual da TRI, que postula que a resposta dada por um indivíduo não influencia nem é influenciada pela resposta de outro indivíduo, uma vez que cada P_{ji} depende apenas do seu respectivo θ_j e, assim, podemos estimar cada habilidade independentemente.

De fato, para $\nabla_{\theta_j} \ell(\Theta)$, temos

$$\nabla_{\theta_j} \ell(\Theta) = \begin{bmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_j} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial \ell_1}{\partial \theta_1} & 0 & \dots & 0 & \dots & 0 \\ 0 & \frac{\partial \ell_2}{\partial \theta_2} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & \frac{\partial \ell_j}{\partial \theta_j} & \dots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \frac{\partial \ell_m}{\partial \theta_m} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{m \times 1} = \begin{bmatrix} \frac{\partial \ell_1}{\partial \theta_1} \\ \frac{\partial \ell_2}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell_j}{\partial \theta_j} \\ \vdots \\ \frac{\partial \ell_m}{\partial \theta_m} \end{bmatrix}$$

de onde vemos que, para certo indivíduo j fixado, o único componente não nulo

em $\frac{\partial \ell}{\partial \theta_j}(\Theta)$ é $\frac{\partial \ell_j}{\partial \theta_j}(\Theta)$. Nesse sentido, podemos denotar $\frac{\partial \ell}{\partial \theta_j}(\Theta) = \frac{d \ell}{d \theta_j}(\theta_j)$.

7.3 Estimador da aptidão

Note agora que P_{ji} e $Q_{ji} = 1 - P_{ji}$ são funções de θ_j , ou seja,

$$P_{ji} = P_{ji}(\theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-D a(\theta_j - b)}} \quad \text{e} \quad Q_{ji} = Q_{ji}(\theta_j) = 1 - P_{ji} \quad . \quad (7.4)$$

Com isso em vista, para determinar o estimador da aptidão do indivíduo j , precisamos antes ser capazes de calcular $\frac{\partial \ell}{\partial \theta_j}(\Theta)$. Nessa direção, vamos encontrar alguns resultados prévios que serão utilizados. Então, da equação (6.6), temos

$$\begin{aligned}\frac{\partial z_{ji}}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j} (D a_i (\theta_j - b_i)) \\ &= 0 \cdot (\theta_j - b_i) + D a_i \cdot 1 \\ &= D a_i \quad .\end{aligned}\tag{7.5}$$

Além disso, temos

$$\begin{aligned}\frac{\partial P_{ji}}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j} (c_i + (1 - c_i) \cdot \Psi_{ji}) \\ &= (1 - c_i) \cdot \frac{\partial \Psi_{ji}}{\partial \theta_j} \\ &= (1 - c_i) \cdot \frac{\partial \Psi_{ji}}{\partial z} \cdot \frac{\partial z}{\partial \theta_j} \\ &= D a_i \cdot (1 - c_i) \cdot \Psi_{ji} (1 - \Psi_{ji}) \quad .\end{aligned}\tag{7.6}$$

E, por último,

$$\begin{aligned}i. \quad \frac{\partial}{\partial \theta_j} \log P_{ji} &= \frac{1}{P_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \\ ii. \quad \frac{\partial}{\partial \theta_j} \log Q_{ji} &= \frac{1}{Q_{ji}} \cdot \frac{\partial Q_{ji}}{\partial \theta_j} = \frac{1}{Q_{ji}} \cdot \frac{\partial}{\partial \theta_j} (1 - P_{ji}) = -\frac{1}{Q_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \quad .\end{aligned}\tag{7.7}$$

Com esses resultados, já somos capazes de determinar $\frac{\partial \ell}{\partial \theta_j}(\Theta)$. Veja:

$$\begin{aligned}\frac{\partial \ell}{\partial \theta_j}(\Theta) &= \frac{\partial}{\partial \theta_j} \sum_{j=1}^m \sum_{i=1}^n (u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}) \\ &= \frac{\partial}{\partial \theta_j} \sum_{i=1}^m \sum_{j=1}^n (u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \theta_j} \sum_{j=1}^m (u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}) \\ &\stackrel{[7.3]}{=} \sum_{i=1}^n \frac{\partial}{\partial \theta_j} (u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji})\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left(u_{ji} \cdot \frac{\partial}{\partial \theta_j} \log P_{ji} + (1-u_{ji}) \cdot \frac{\partial}{\partial \theta_j} \log Q_{ji} \right) \\
&\stackrel{[7.7]}{=} \sum_{i=1}^n \left(u_{ji} \cdot \frac{1}{P_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} - (1-u_{ji}) \cdot \frac{1}{Q_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \right) \\
&= \sum_{i=1}^n \frac{u_{ji} - P_{ji}}{P_{ji} \cdot Q_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \tag{7.8}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n (u_{ji} - P_{ji}) \cdot \frac{1}{P_{ji} \cdot Q_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \\
&\stackrel{[6.7]}{=} \sum_{i=1}^n (u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji} \cdot (1-\Psi_{ji})} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \tag{7.9}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{[7.6]}{=} \sum_{i=1}^n (u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji} \cdot (1-\Psi_{ji})} \cdot D a_i (1-c_i) \cdot \Psi_{ji} (1-\Psi_{ji}) \\
&= D \cdot \sum_{i=1}^n a_i \cdot (u_{ji} - P_{ji}) \cdot (1-c_i) \cdot W_{ji} \tag{7.10}
\end{aligned}$$

em que, ao fazer $\frac{\partial \ell}{\partial \theta_j}(\Theta) = 0$ para procurar pontos críticos, em especial os valores

de θ que maximizam $\ell(\theta)$, obtemos o estimador da aptidão $\hat{\theta}_j$.

$$D \cdot \sum_{i=1}^n a_i \cdot (u_{ji} - P_{ji}) \cdot (1-c_i) \cdot W_{ji} = 0 \tag{7.11}$$

Como todo estimador, o estimador da aptidão obtido é uma função dos dados observados. De fato, os parâmetros dos itens $\lambda_i = (a_i, b_i, c_i)$ já foram estimados, neste caso, e D é uma constante, de forma que, para cada indivíduo j , teremos um vetor padrão de resposta U_j e, para cada item i , teremos um valor de u_{ji} . Com todos esses valores em mãos, procuramos, dentre todos os valores possíveis de θ_j , o maior que zera o estimador de aptidão e, pela característica da função de log-verosimilhança, isso ocorre justamente para $\hat{\theta}_j$, que é a estimativa de máxima verossimilhança da aptidão do indivíduo j .

Novamente, o estimador assim obtido não apresenta uma solução explícita para obter as estimativas que queremos, sendo novamente necessário algum método numérico para fazer a estimativa.

7.4 Fórmula de recorrência das habilidades por Newton-Raphson

Com procedimento análogo ao feito na seção 6.5 para os parâmetros dos itens e considerando a separabilidade de $\ell(\Theta)$ por indivíduos, podemos aplicar o método de Newton-Raphson localmente, encontrando

$$\hat{\theta}_j^{(k+1)} = \hat{\theta}_j^{(k)} - \left[H\left(\hat{\theta}_j^{(k)}\right) \right]^{-1} \cdot h\left(\hat{\theta}_j^{(k)}\right) , \quad (7.12)$$

que é uma recorrência que permite obter uma estimativa $\hat{\theta}_j$ do valor real θ_j^* da habilidade θ_j do indivíduo j . Para essa recorrência ficar completamente definida, precisamos determinar $h(\theta_j)$, que é a primeira derivada parcial de $\ell(\Theta)$ (Jacobiana local) em relação a θ_j , e $H(\theta_j)$, segunda derivada parcial de $\ell(\Theta)$ em relação a θ_j (Hessiana local).

Uma expressão para $h(\theta_j) = \frac{\partial \ell}{\partial \theta_j}(\Theta)$ já foi obtida na equação (7.9). No entanto, vamos reescrevê-la para facilitar o uso do algoritmo usado no método de Newton-Raphson. Assim, partindo da equação (7.9), temos

$$\begin{aligned} h(\theta_j) &= \frac{\partial \ell}{\partial \theta_j}(\Theta) \\ &= \sum_{i=1}^n (u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \\ &= \sum_{i=1}^n (u_{ji} - P_{ji}) \cdot W_{ji} \cdot h_{ji} \end{aligned} \quad (7.13)$$

com

$$N'_{ji} = \frac{1}{\Psi_{ji} \cdot (1 - \Psi_{ji})} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \quad (7.14)$$

Já para, para determinados uma expressão para $H(\theta_j)$, precisamos de alguns resultados anteriores. Inicialmente, note que:

$$\frac{\partial Q_{ji}}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} (1 - P_{ji}) = -\frac{\partial P_{ji}}{\partial \theta_j} \quad (7.15)$$

Assim,

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} (P_{ji} Q_{ji}) &= \frac{\partial P_{ji}}{\partial \theta_j} \cdot Q_{ji} + P_{ji} \cdot \frac{\partial Q_{ji}}{\partial \theta_j} \\
&\stackrel{[7.4]}{=} \frac{\partial P_{ji}}{\partial \theta_j} \cdot (1 - P_{ji}) + P_{ji} \cdot \frac{\partial Q_{ji}}{\partial \theta_j} \\
&\stackrel{[7.15]}{=} \frac{\partial P_{ji}}{\partial \theta_j} \cdot (1 - P_{ji}) - P_{ji} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \\
&= (1 - 2P_{ji}) \cdot \frac{\partial P_{ji}}{\partial \theta_j}
\end{aligned} \tag{7.16}$$

Além disso, como $u_{ji}^2 = u_{ji}$ e $P_{ji}^2 = 2P_{ji}^2 - P_{ji}^2 + P_{ji} - P_{ji}$, temos

$$\begin{aligned}
\left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right)^2 &= \frac{1}{(P_{ji} Q_{ji})^2} \cdot (u_{ji}^2 - 2P_{ji}u + P_{ji}^2) \\
&= \frac{1}{(P_{ji} Q_{ji})^2} \cdot (u_{ji} - 2P_{ji}u + 2P_{ji}^2 - P_{ji}^2 + P_{ji} - P_{ji}) \\
&= \frac{1}{(P_{ji} Q_{ji})^2} \cdot [P_{ji}(1 - P_{ji}) + (u_{ji} - P_{ji})(1 - 2P_{ji})] \\
&= \frac{1}{(P_{ji} Q_{ji})^2} \cdot [P_{ji} Q_{ji} + (u_{ji} - P_{ji})(1 - 2P_{ji})]
\end{aligned} \tag{7.17}$$

e, dessa forma,

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) &= \frac{1}{(P_{ji} Q_{ji})^2} \cdot \left[\left(\frac{\partial}{\partial \theta_j} (u_{ji} - P_{ji}) \right) \cdot P_{ji} Q_{ji} - (u_{ji} - P_{ji}) \cdot \frac{\partial}{\partial \theta_j} (P_{ji} Q_{ji}) \right] \\
&= \frac{1}{(P_{ji} Q_{ji})^2} \cdot \left[\left(-\frac{\partial P_{ji}}{\partial \theta_j} \right) \cdot P_{ji} Q_{ji} - (u_{ji} - P_{ji}) \cdot \frac{\partial}{\partial \theta_j} (P_{ji} Q_{ji}) \right] \\
&= -\frac{1}{(P_{ji} Q_{ji})^2} \cdot \left[\frac{\partial P_{ji}}{\partial \theta_j} \cdot P_{ji} Q_{ji} + (u_{ji} - P_{ji}) \cdot \frac{\partial}{\partial \theta_j} (P_{ji} Q_{ji}) \right] \\
&\stackrel{[7.16]}{=} -\frac{1}{(P_{ji} Q_{ji})^2} \cdot \left[\frac{\partial P_{ji}}{\partial \theta_j} \cdot P_{ji} Q_{ji} + (u_{ji} - P_{ji}) \cdot (1 - 2P_{ji}) \cdot \frac{\partial P_{ji}}{\partial \theta_j} \right]
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{(P_{ji}Q_{ji})^2} \cdot [P_{ji}Q_{ji} + (u_{ji} - P_{ji})(1 - 2P_{ji})] \cdot \frac{\partial P_{ji}}{\partial \theta_j} \\
&= -\left\{ \frac{1}{(P_{ji}Q_{ji})^2} \cdot [P_{ji}Q_{ji} + (u_{ji} - P_{ji})(1 - 2P_{ji})] \right\} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \\
&= -\left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 \cdot \frac{\partial P_{ji}}{\partial \theta_j}
\end{aligned} \tag{7.18}$$

Agora, calculando a Hessiana, temos

$$\begin{aligned}
H(\theta_j) &= \frac{\partial^2 \ell}{\partial \theta_j \partial \theta_j}(\theta_j) \\
&= \frac{\partial}{\partial \theta_j} \left[\frac{\partial \ell}{\partial \theta_j}(\theta_j) \right] \\
&\stackrel{[7.8]}{=} \frac{\partial}{\partial \theta_j} \left(\sum_{i=1}^n \frac{u_{ji} - P_{ji}}{P_{ji} \cdot Q_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \right) \\
&= \sum_{i=1}^n \frac{\partial}{\partial \theta_j} \left(\frac{u_{ji} - P_{ji}}{P_{ji} \cdot Q_{ji}} \cdot \frac{\partial P_{ji}}{\partial \theta_j} \right) \\
&\stackrel{[7.18]}{=} \sum_{i=1}^n \left[-\left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 \cdot \frac{\partial P_{ji}}{\partial \theta_j} \cdot \frac{\partial P_{ji}}{\partial \theta_j} + \frac{u_{ji} - P_{ji}}{P_{ji} \cdot Q_{ji}} \cdot \frac{\partial}{\partial \theta_j} \left(\frac{\partial P_{ji}}{\partial \theta_j} \right) \right] \\
&= \sum_{i=1}^n \left[(u_{ji} - P_{ji}) \cdot \frac{1}{P_{ji} \cdot Q_{ji}} \cdot \left(\frac{\partial^2 P_{ji}}{\partial \theta_j \partial \theta_j} \right) - \right. \\
&\quad \left. (u_{ji} - P_{ji})^2 \cdot \left(\frac{1}{P_{ji}Q_{ji}} \right)^2 \cdot \left(\frac{\partial P_{ji}}{\partial \theta_j} \right)^2 \right] \\
&\stackrel{[6.9]}{=} \sum_{i=1}^n \left[(u_{ji} - P_{ji}) \cdot \frac{W_{ji}}{\Psi_{ji}(1 - \Psi_{ji})} \cdot \left(\frac{\partial^2 P_{ji}}{\partial \theta_j \partial \theta_j} \right) - \right. \\
&\quad \left. (u_{ji} - P_{ji})^2 \cdot \left(\frac{W_{ji}}{\Psi_{ji}(1 - \Psi_{ji})} \right)^2 \cdot \left(\frac{\partial P_{ji}}{\partial \theta_j} \right)^2 \right] \\
&= \sum_{i=1}^n \left[(u_{ji} - P_{ji}) \cdot W_{ji} \cdot M'_{ji} - (u_{ji} - P_{ji})^2 \cdot W_{ji}^2 \cdot N'_{ji}^2 \right] \\
&= \sum_{i=1}^n W_{ji} \cdot (u_{ji} - P_{ji}) \cdot \left[M'_{ji} - (u_{ji} - P_{ji}) \cdot W_{ji} \cdot N'_{ji}^2 \right]
\end{aligned} \tag{7.19}$$

com N'_{ji} dada pela equação (7.14) e

$$M'_{ji} = \frac{1}{\Psi_{ji}(1-\Psi_{ji})} \cdot \left(\frac{\partial^2 P_{ji}}{\partial \theta_j \partial \theta_j} \right) \quad .$$

Com isso, a recorrência da equação (7.12)

$$\hat{\theta}_j^{(k+1)} = \hat{\theta}_j^{(k)} - \left[H(\hat{\theta}_j^{(k)}) \right]^{-1} \cdot h(\hat{\theta}_j^{(k)})$$

fica completa e as iterações levam a uma estimativa da aptidão de cada indivíduo.

Cabe ressaltar que o θ assim estimado para cada participante já representa sua proficiência latente final, ou seja, é o resultado da modelagem considerando todas as respostas aos itens do teste dados pelo testando e os parâmetros dos itens.

8

Escala de proficiência

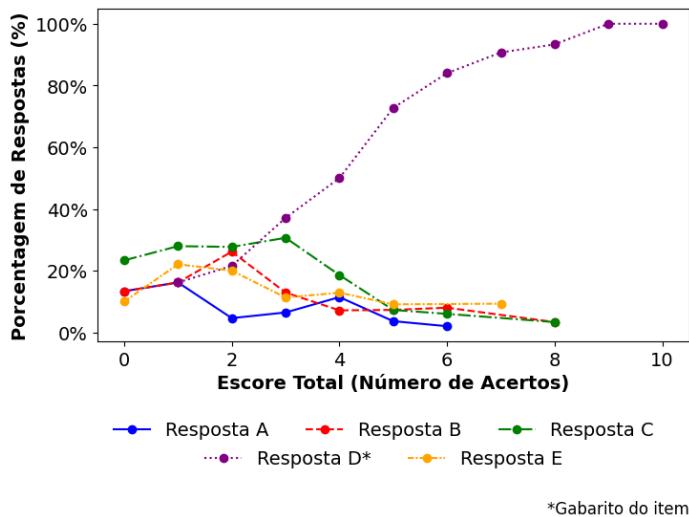
Com os parâmetros dos itens já definidos, podemos agora analisar melhor a relação existente entre a proficiência e a probabilidade de acerto em cada item. Compreender essa relação é fundamental, tanto para validar os itens que irão compor os testes quanto para assegurar a padronização das avaliações. Neste capítulo, vamos explorar uma representação gráfica que traz diversas informações sobre o item e seu potencial para medir, com precisão, a proficiência do testando.

Além disso, apresentaremos a escala de proficiência, que serve como referência para interpretar, pedagogicamente, os resultados obtidos. Por fim, discutiremos os mecanismos utilizados na TRI que permitem a comparação de proficiências, mesmo em aplicações distintas de uma mesma avaliação.

8.1 Curva Característica do Item (CCI)

Vamos retomar o teste aplicado a Maria e a João, apresentado como segundo exemplo hipotético na introdução deste trabalho. Ambos acertaram cinco dos dez itens do teste, porém itens diferentes. Agora, ainda usando dados hipotéticos, analisaremos os resultados de todas as turmas cujos estudantes foram submetidos a esse teste. Vamos supor que, ao todo, 500 estudantes foram avaliados. Aqui, é importante lembrar que o teste é composto por dez itens objetivos, cada um com cinco alternativas de resposta, das quais apenas uma está correta.

O gráfico a seguir mostra, para o item 1 desse teste, a proporção de escolha de cada uma das cinco alternativas de resposta, por escore total, dos 500 estudantes avaliados. Esse gráfico, chamado de **Análise Gráfica do Item (AGI)**, é tipicamente usado na TCT para análise de itens. Perceba que, apesar de ser possível obter várias conclusões interessantes sobre o item a partir da AGI, todas elas não individualizam cada item, uma vez que a AGI relaciona as respostas do item com o escore total. Além disso, tais conclusões também não individualizam o indivíduo testando.



Com o intuito de encontrar a melhor explicação para o escore total do indivíduo, a TRI inova ao tirar o foco da possível influência existente entre o escore total e o acerto de um item em específico para analisar a relação existente entre a probabilidade de acerto e o nível do traço latente exigido na resolução do item. Note que, se fôssemos plotar em um gráfico de dispersão o escore total dos testandos em relação ao respectivo resultado u_{ji} de erro ou de acerto nesse mesmo item 1, teríamos um depósito de pontos ao longo das retas $y = 0$ e $y = 1$.

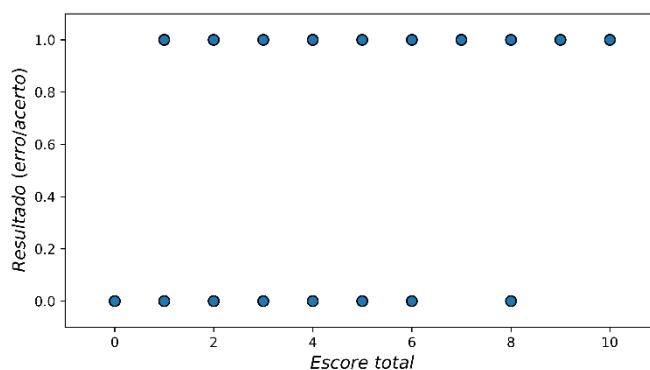


Gráfico 8.2 – Dispersão dos resultados (erros e acertos) do item 1

Fonte: autoria própria

Essa dispersão dá origem a uma curva no formato de “S”, chamada sigmoide, com duas assíntotas horizontais, que pode ser obtida por meio de uma regressão logística. Note que essa curva mostra uma tendência de acerto nesse item, indicada pela maior ou menor densidade de pontos em algumas regiões ao longo das retas $y = 0$ e $y = 1$. Veja:

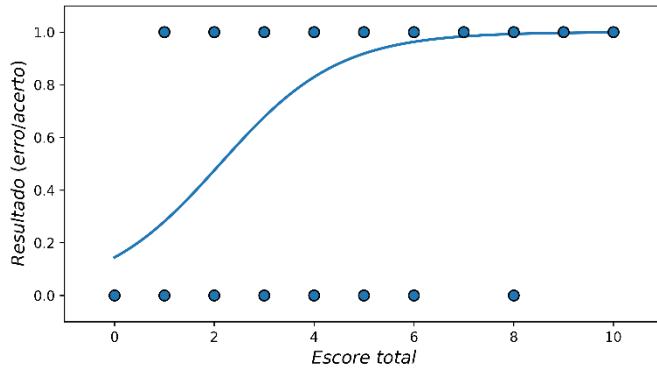


Gráfico 8.3 – Regressão logística dos resultados (erros e acertos)

Fonte: autoria própria

Como os dados de escore, nesse caso, tem natureza discreta, a dispersão não nos traz muita informação. No entanto, ao usar um *bubble plot*, em que as bolhas são proporcionais à frequência em cada escore, torna-se possível evidenciar, de forma progressiva, a curva sigmoidal e a transição de densidade entre erro e acerto à medida que o escore aumenta. Observe ainda que, sobre a assíntota de acertos $y = 1$, a densidade da dispersão nos extremos é significativamente menor em comparação com a região intermediária. Esse padrão mostra a tendência de uma distribuição normal dos traços latentes na população.

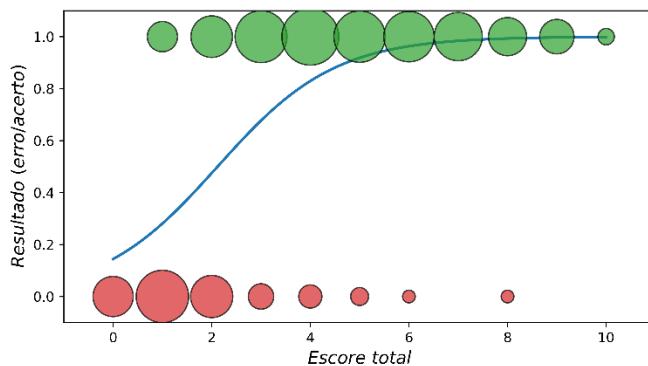


Gráfico 8.4 – Dispersão de erros e acertos tipo *bubble plot*

Fonte: autoria própria

O ML3P descreve justamente essa relação por meio de uma saída gráfica típica, chamada de Curva Característica do Item: a CCI. A expressão “característica do item” significa que cada item tem a sua curva, que traduz o efeito dos seus respectivos parâmetros para cada nível de $\theta \in \mathbb{R}$, de $-\infty$ a $+\infty$. Assim, para obter a CCI de um item, temos antes que obter os seus parâmetros e atualizar o modelo com tais valores. Dessa forma, o modelo passa a ser uma função de θ , $P(\theta)$.

A seguir, estão os parâmetros dos itens desse teste hipotético obtidos usando a metodologia de estimativa de parâmetros e de aptidões utilizada pelo Enem (MMLE e EAP), segundo o Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2021b), que comentaremos no Capítulo 9. Como base de dados para esse exemplo, foi utilizado uma amostra dos microdados da aplicação de 2023 do Enem, obtidos no site do Inep.

Item	<i>a</i>	<i>b</i>	<i>c</i>
1	1,865	-0,726	0,000
2	1,963	-0,282	0,000
3	2,269	0,240	0,158
4	0,952	0,316	0,000
5	1,776	0,528	0,025
6	1,918	1,067	0,202
7	1,118	0,610	0,000
8	1,178	1,473	0,084
9	1,654	0,901	0,021
10	2,342	1,381	0,015

Tabela 8.1 – Parâmetros do teste hipotético

Fonte: autoria própria

A partir dessas informações, plotamos a CCI do Item 1, confirmando uma tendência próxima da apresentada no gráfico de dispersão anterior.

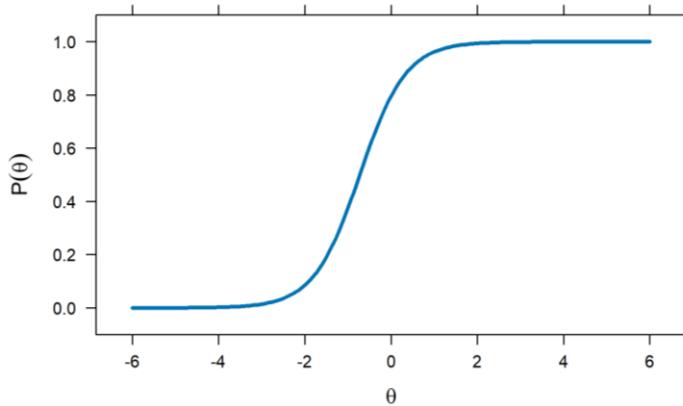


Gráfico 8.5 – Curva Característica do Item 1 na métrica (0, 1)

Fonte: autoria própria

Adicionalmente, veja que os valores assumidos pela CCI estão limitados pelas assíntotas horizontais $y=0$ e $y=1$, consequência da origem probabilística do modelo utilizado (ML3P).

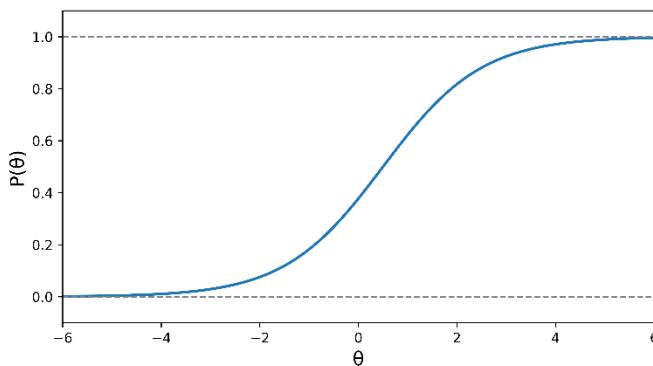


Gráfico 8.6 – Assíntotas horizontais da CCI

Fonte: autoria própria

Cabe ressaltar que essa probabilidade se refere à probabilidade de acerto, e não a de erro. Além disso, perceba que a CCI da imagem anterior está na métrica de média 0 e desvio padrão 1, chamada de métrica ou escala (0, 1), uma vez que os resultados são gerados por um modelo que, por padrão, opera nessa métrica. Porém, como valores negativos de proficiência podem causar certa estranheza, embora indiquem apenas que o nível de aptidão do indivíduo está alguns desvios abaixo da média zero, é possível transformar essa escala por meio de uma simples conversão linear. Um exemplo comum é a métrica adotada pelo Enem, com média 500 e desvio padrão 100, conforme definido pelo Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2021b), responsável pela aplicação e organização do exame. Dessa forma, com essa métrica, a CCI anterior, fica conforme a imagem a seguir, mantendo a forma e as dimensões, modificando apenas a escala numérica.

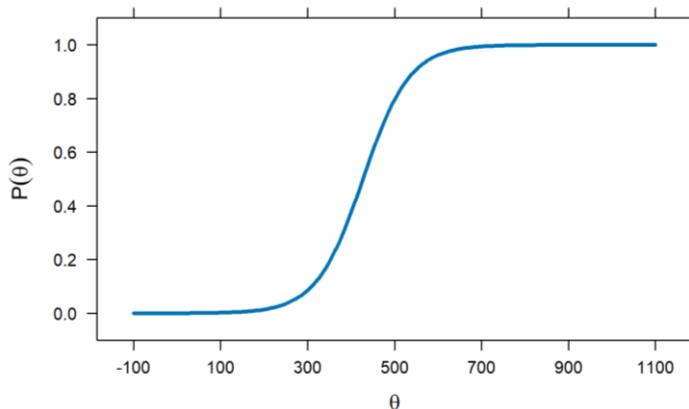


Gráfico 8.7 – Curva Característica do Item 1

Fonte: autoria própria

A CCI reflete diretamente os efeitos dos parâmetros associados a cada item. Compare os valores apresentados na Tabela 8.1 com os gráficos dos demais itens do teste na métrica de média 500 e desvio padrão 100, ilustrados a seguir, e observe como cada parâmetro influencia o formato e o comportamento da curva.

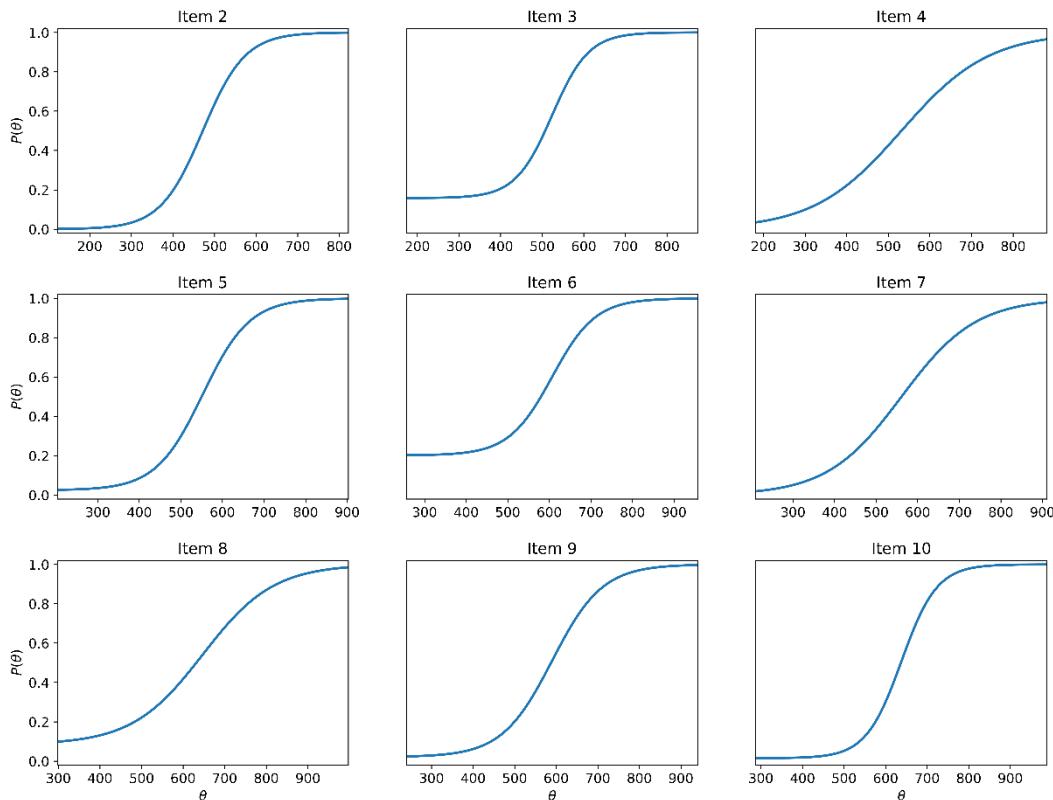


Gráfico 8.8 – Curva Característica do Item dos itens 2 a 10
Fonte: autoria própria

8.1.1 A CCI e o parâmetro de dificuldade b

O parâmetro de dificuldade b indica o valor mínimo da aptidão θ que o indivíduo precisa possuir para ter uma probabilidade de $\frac{1+c}{2}$ de acertar o item (Rabelo, 2013). Em outras palavras, trata-se do ponto da escala de aptidão em que a probabilidade de resposta correta é igual a $\frac{1+c}{2}$. Caso $c=0$, então b indica a

proficiência mínima necessária para que o indivíduo tenha uma chance de 50% de acertar o item.

O valor desse parâmetro, teoricamente, varia de $-\infty$ a $+\infty$, porém, na prática, ele se situa entre -3 e $+3$, na escala $(0, 1)$, que é a faixa de dificuldade onde normalmente se encontram os itens válidos. Quanto maior o valor desse parâmetro, mais difícil é o item no sentido de que maior deve ser o nível de aptidão do indivíduo para que ele tenha uma chance de $\frac{1+c}{2}$ de acertar o item. Assim, itens com valores de dificuldade fora dessa faixa geralmente não são utilizados em testes, pois, na maioria das vezes, possuem problemas de elaboração (Rabelo, 2013).

Vamos ver mais adiante que a abscissa do ponto de inflexão da CCI corresponde ao parâmetro de dificuldade do item. A figura a seguir apresenta as curvas de três itens: os itens II e III possuem a mesma dificuldade, superior à do item I.

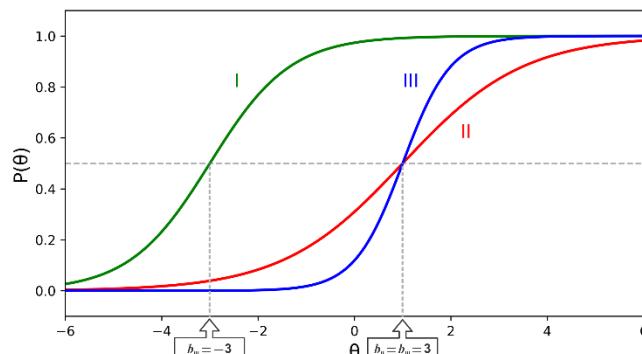


Gráfico 8.9 – Curvas com diferentes discriminações
Fonte: autoria própria

De acordo com Rabelo (2013), é esperado uma distribuição normal para os níveis de dificuldade dos itens de um teste, como a maioria dos autores da área de avaliação e psicometria adotam. A classificação dos itens, na escala $(0, 1)$, por nível de dificuldade, bem como sua respectiva distribuição no teste, está representada no quadro a seguir.

Classificação	b	Distribuição
Muito fácil	$b \leq -1,28$	10%
Fácil	$-1,28 < b < -0,52$	20%
Mediano	$-0,52 \leq b \leq 0,52$	40%
Difícil	$0,52 < b < 1,28$	20%
Muito difícil	$b \geq 1,28$	10%

Quadro 8.1 – Classificação dos itens quando à discriminação
Fonte: Rabelo, 2013, p. 134, adaptado

8.1.2 A CCI e o parâmetro de discriminação α

A discriminação de um item é a sua capacidade em diferenciar indivíduos com níveis de aptidão próximas ao nível de dificuldade do item. De acordo com Rabelo (2013), “[...] a discriminação é identificada como um valor proporcional à inclinação da curva no ponto onde ela muda de concavidade (ponto de inflexão)”.

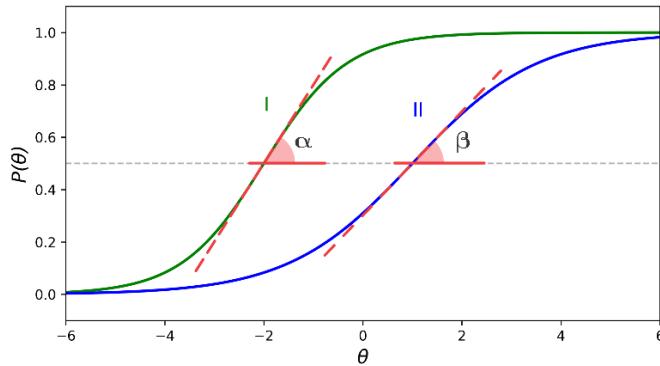


Gráfico 8.10 – Parâmetro de discriminação

Fonte: autoria própria

O ponto de inflexão de uma curva é aquele em que o gráfico da função muda de concava para convexa ou vice-versa. Podemos encontrá-lo por meio da derivada segunda da função. No caso da CCI de um item i , como P é função de θ , partimos da primeira derivada de $P(\theta)$ em relação a θ , conforme a equação (7.6). Lembrando que Ψ depende de z e que, por sua vez, z é função de θ , temos que a segunda derivada de $P(\theta)$ em relação a θ é dada por

$$\begin{aligned}
 \frac{d^2P}{d\theta^2} &= \frac{d}{d\theta} \left(\frac{dP}{d\theta} \right) \\
 &\stackrel{[7.6]}{=} \frac{d}{d\theta} \left[Da(1-c) \cdot \Psi(1-\Psi) \right] \\
 &\stackrel{[6.8]}{=} \frac{d}{d\theta} \left[Da(1-c) \cdot \Psi' \right] \\
 &= Da(1-c) \cdot \frac{d\Psi'}{d\theta} \\
 &\cdot \quad = Da(1-c) \cdot \Psi'' \frac{dz}{d\theta} \\
 &\stackrel{[7.5]}{=} Da(1-c) \cdot \Psi'' \cdot Da \\
 &= D^2a^2(1-c) \cdot \Psi'' \\
 &\stackrel{[6.26]}{=} D^2a^2(1-c) \cdot \Psi(1-\Psi)(1-2\Psi)
 \end{aligned}$$

e, para o ponto de inflexão,

$$\frac{d^2P}{d\theta^2} = 0 \Rightarrow D^2a^2(1-c) \cdot \Psi(1-\Psi)(1-2\Psi) = 0 \quad .$$

Como $\Psi = \frac{1}{1+e^{-z}}$ é a função logística em que $\lim_{x \rightarrow \infty} \Psi = 1$, então $\Psi = 1$ ocorre apenas assintoticamente. Dessa forma, como $a > 0$, $D > 0$ e $0 \leq c < 1$, então o único fator capaz de anular a última expressão acima é $1-2\Psi$, de modo que

$$1-2\Psi = 0 \Rightarrow \Psi = \frac{1}{2}$$

e, assim,

$$\begin{aligned} \Psi = \frac{1}{1+e^{-z}} = \frac{1}{2} &\Rightarrow \frac{1}{1+e^{-z}} = \frac{1}{2} \\ &\Rightarrow 1+e^{-z} = 2 \\ &\Rightarrow e^{-z} = 1 \\ &\Rightarrow z = 0 \quad . \end{aligned}$$

Por fim, como $z = Da(\theta - b)$,

$$\begin{aligned} z = 0 &\Rightarrow Da(\theta - b) = 0 \\ &\Rightarrow \theta - b = 0 \\ &\Rightarrow \theta = b \quad . \end{aligned}$$

Logo, o ponto de inflexão na CCI do ML3P ocorre quando θ é a dificuldade do item. Dessa forma, quanto maior a inclinação da CCI no ponto de inflexão, mais discriminativo é o item, ou seja, mais separa indivíduos com nível de aptidões diferentes.

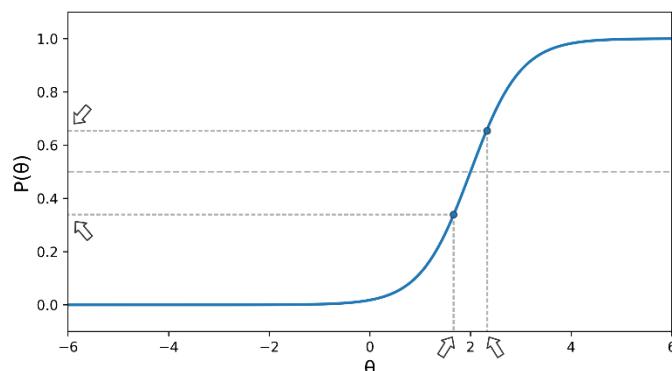


Gráfico 8.11 –Efeito do parâmetro de discriminação
Fonte: autoria própria

Por outro lado,

Quanto mais achatada for a curva, menos discriminativo será o item. Valores do parâmetro a próximo de zero indicam que o item tem pouco poder de discriminação, ou seja, indivíduos com habilidades muito diferentes têm, aproximadamente, a mesma probabilidade de responder corretamente o item (Rabelo, 2013, p. 138).

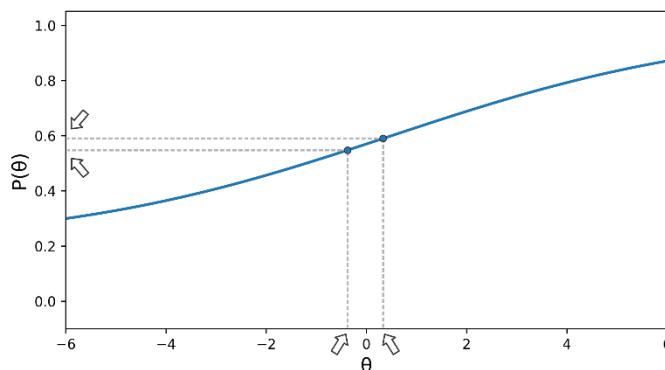


Gráfico 8.12 – Parâmetro de discriminação próximo de zero

Fonte: autoria própria

Além disso, conforme Pasquali, “[...] se um item apresenta uma discriminação perfeita, então o ângulo, de incidência da curva seria de 90 graus, ou seja, uma perpendicular. Nesse caso, o item é capaz de discriminar diferenças infinitesimalmente mínimas de níveis de teta” (Pasquali, 2018, p. 29).

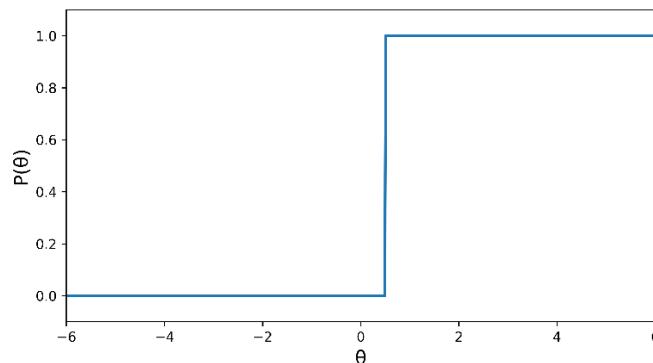


Gráfico 8.13 – Discriminação perfeita

Fonte: Pasquali, 2018, p. 29, adaptado

Dessa forma, note que, pelo menos teoricamente, os valores do parâmetro a podem variar de $-\infty$ a $+\infty$. No entanto, valores negativos ficam um tanto quanto estranhos, pois indicariam que indivíduos de maior aptidão têm mais chance de errar o item do que aqueles com menor aptidão, uma vez que a CCI, agora, estaria com um formato refletido.

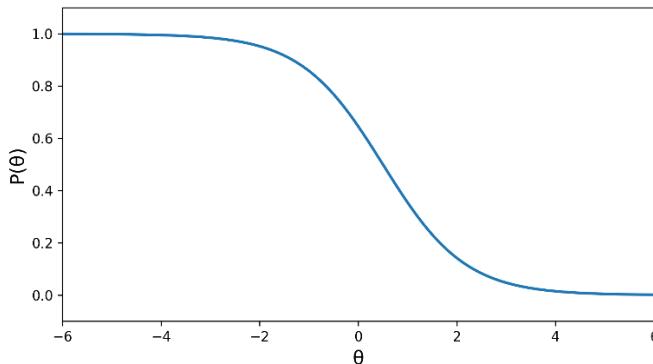


Gráfico 8.14 – Discriminação negativa

Fonte: autoria própria

Na prática, porém, os valores do parâmetro a se concentram entre -3 e $+3$ na escala $(0, 1)$. Conforme Rabelo (2013), um item é considerado discriminativo se o parâmetro a é maior que $0,7$, sendo que alguns autores categorizam a discriminação dos itens conforme o quadro seguir.

a	Discriminação
$a = 0$	Nenhuma
$0 < a \leq 0,35$	Muito baixa
$0,35 < a \leq 0,65$	Baixa
$0,65 < a \leq 1,35$	Moderada
$1,35 < a \leq 1,7$	Alta
$a > 1,7$	Muito alta

Quadro 8.2 – Classificação da discriminação dos itens na TRI

Fonte: Rabelo, 2013, p. 138

8.1.3 A CCI e o parâmetro de acerto ao acaso c

O parâmetro c mede a probabilidade de acerto ao acaso de uma resposta dada a um item por um indivíduo de baixa proficiência. Sabemos que a chance de acerto de um item ao escolher, ao acaso, uma de suas cinco alternativas, é de 20%. Porém,

essa é uma abordagem que desconsidera fatores de atualização, como marcação de alternativas atrativas ou eliminação de alternativas repelentes, fazendo com que a probabilidade inicial de acerto aumente ou reduza, conforme a natureza do item. Apesar de ser designado de parâmetro de chute ao acaso, muitos preferem chamar o parâmetro c de pseudochute, justamente porque a escolha ao acaso de uma opção de resposta pode sofrer algum tipo de influência, não sendo essa escolha tão aleatória assim.

O parâmetro c corresponde ao ponto de intersecção entre a assíntota inferior da CCI e o eixo y. Dessa forma, quanto mais para cima estiver essa assíntota horizontal, maior é a chance de um indivíduo com baixa proficiência acertar o item. Como medida de probabilidade, $0 \leq c \leq 1$ e, para um item com cinco opções de resposta, valores de c menores que 0,2 são aceitáveis (Rabelo, 2013). Na figura a seguir, um indivíduo com baixa proficiência tem mais chance de acertar o item I do que o item II.

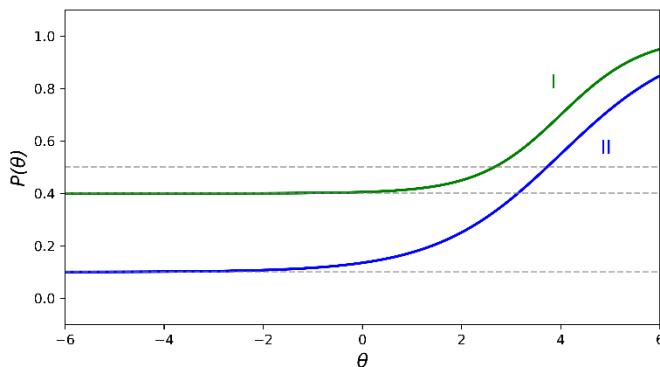


Gráfico 8.15 – Parâmetro de acerto ao acaso
Fonte: autoria própria

8.2 A curva de probabilidade de erro

Naturalmente, a CCI é a representação gráfica da probabilidade de acerto por excelência da TRI, muito embora, de forma indireta, seja possível obter informações da probabilidade de erro do item para cada um dos $\theta \in \mathbb{R}$. No entanto, a função $Q(\theta) = 1 - P(\theta)$ é a mais adequada para representar a probabilidade de erro. A figura a seguir mostra a curva da probabilidade de erro.

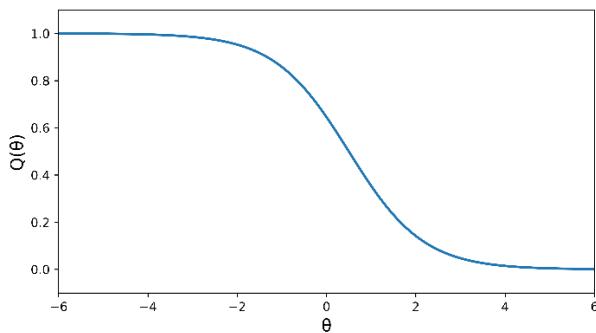


Gráfico 8.16 – Probabilidade de erro

Fonte: autoria própria

Note que, conforme o esperado, a curva de erro tem crescimento contrário ao da CCI, uma vez que, para θ pequenos, há maior probabilidade de erro do item.

8.3 Escala de proficiência

Ao usar o ML3P para determinar os parâmetros dos itens e, a partir deles, as proficiências por meio do JMLE, tais valores são obtidos na métrica (0, 1). Um próximo passo é ajustar essa métrica, caso necessário, tanto para os parâmetros (dificuldade) quanto para as proficiências. Com todos os parâmetros (dificuldade) e proficiências na mesma escala, o que podemos dizer sobre a aprendizagem do estudante? Por exemplo, o que significa ter determinada proficiência? É possível inferir que o estudante se encontra dentro do previsto em relação ao que devia saber? Quais são os conhecimento e aptidões que o estudante demonstrou ter domínio? E quais deveria aperfeiçoar? Nesse sentido, há dois aspectos a se falar: o primeiro está relacionado ao procedimento que dá sentido e significado às aptidões obtidas por meio de um teste de proficiência; o segundo, em como comparar proficiências vindas de versões diferentes de um mesmo teste e em como produzir testes paralelos.

Tais reflexões possibilitaram a criação da escala de proficiência, que coloca tanto os itens quanto o testando em uma mesma régua, organizados, respectivamente, pelo parâmetro de dificuldade e pelo nível de aptidão (ou proficiência). A ideia por trás de posicionar itens e testandos em uma mesma régua é comparar o respectivo desempenho com o nível de aptidão necessário para acertar cada item. Dessa forma, um indivíduo que estiver posicionado numa posição acima do nível de dificuldade de certo item certamente tem alta probabilidade de acertar esse e

todos os itens abaixo do nível em que está. Assim, questões mais fáceis estão posicionadas na região inferior (à esquerda) da régua e as mais difíceis, acima (à direita). Na verdade,

Geralmente, o posicionamento do item está um pouco acima do parâmetro de dificuldade. A questão é sempre posicionada no valor da régua em que a probabilidade de acerto está próxima de 0,65, o qual representa que participantes nesse nível possuem alta probabilidade de dominar o conteúdo da questão (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2021a).

8.4 Interpretação pedagógica

Após localizar os itens na escala de proficiência, seguimos com a próxima etapa: a interpretação pedagógica das escalas de proficiência. Essa etapa é planejada antes mesmo da elaboração dos itens, com a confecção da Matriz de Competências e Habilidades que serão alvo da avaliação. No Enem, cada área do conhecimento possui uma matriz. Especificamente, a matriz da área de Matemática e suas Tecnologias possui trinta habilidades distribuídas por sete competências, de forma que, cada habilidade pertence a uma única competência e cada item avalia uma única habilidade. É evidente que o estudante precisa mover vários conhecimentos e ações para resolver um item, mas existe uma habilidade que se destaca e é essa que, tecnicamente, será avaliada.

Junto com a Matriz de Habilidades e Competências do Enem, está um conjunto de objetos do conhecimento, que justamente são os conhecimentos que o estudante precisa dispor, para, a partir deles, realizar as ações solicitadas em cada item. O Enem organiza essas ações, a nível macro, em cinco Eixos Cognitivos comuns a todas as áreas do conhecimento, que influenciam a ação que o estudante terá diante do item. É justamente isso que será medido no teste: o quanto bem o estudante organiza e relaciona os objetos do conhecimento em relação ao eixo cognitivo solicitado. Essa convergência caracteriza uma habilidade de forma que, habilidades próximas nesse sentido foram agrupadas em competências na Matriz de Habilidades e Competências.

A interpretação pedagógica é a análise do desempenho do estudante no teste a partir da escala de proficiência, a partir da qual é possível inferir o nível de domínio do testando em relação às habilidades avaliadas. Nesse sentido, é altamente recomendável que para cada habilidade exista uma escala de proficiência de maneira a facilitar a interpretação pedagógica. A matriz completa da área de Matemática e suas Tecnologia pode ser encontrada no site oficial do Inep, mas também pode ser encontrada ao final deste trabalho, nos Anexos B e C.

Como a prova de Matemática do Enem possui 45 itens e a matriz de Matemática tem 30 habilidades, algumas dessas habilidades são medidas mais de uma vez. A seguir, está um exemplo retirado do Guia do Participante do Enem de 2025, em que é mostrado um exemplo de escala de proficiência e também uma interpretação pedagógica, considerando quais conhecimento e habilidades é provável que o estudante domine a depender da posição dele nessa régua.

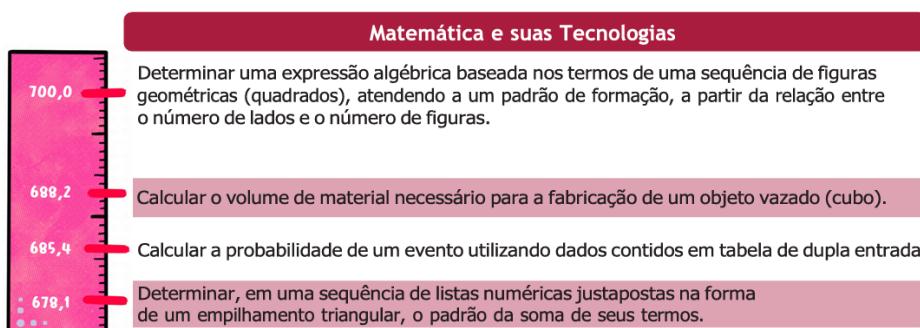


Figura 8.1 – Escala de Proficiência e Interpretação Pedagógica

Fonte: Guia do Estudante/Inep, 2021, pág. 12. Acesso: 28 ago. 2025

A figura a seguir contém os últimos itens da habilidade “H6: Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.”, do caderno amarelo, usadas para avaliar os participantes do Enem nos últimos 10 anos na aplicação regular. Nela, cada bolha identifica a questão, por posição na prova e ano de aplicação da prova de forma que, quanto mais para cima o ponto está, mais recente é o item; e quanto mais para a direta, mais difícil é o item. Além disso, a métrica da escala está conforme a métrica usada no Enem.

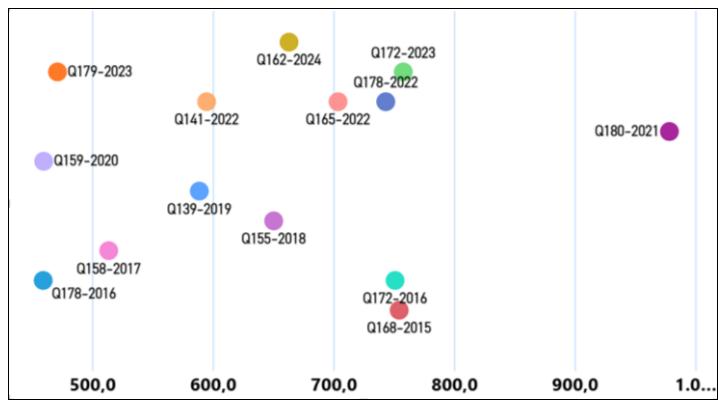


Gráfico 8.17 – Escala de Proficiência

Fonte: autoria própria

Por exemplo, o item 180 do caderno amarelo da prova regular de 2021 avaliou a Habilidade 16 e, como esse item está mais à direita que os outros, o seu parâmetro de dificuldade é maior que os demais, indicando que esse foi um item difícil.

O Atomium, representado na imagem, é um dos principais pontos turísticos de Bruxelas. Ele foi construído em 1958 para a primeira grande exposição mundial depois da Segunda Guerra Mundial, a Feira Mundial de Bruxelas.

Trata-se de uma estrutura metálica construída no formato de um cubo. Essa estrutura está apoiada por um dos vértices sobre uma base paralela ao plano do solo, e a diagonal do cubo, contendo esse vértice, é ortogonal ao plano da base. Centradas nos vértices desse cubo, foram construídas oito esferas metálicas, e uma outra esfera foi construída centrada no ponto de interseção das diagonais do cubo. As oito esferas sobre os vértices são interligadas segundo suas arestas, e a esfera central se conecta a elas pelas diagonais do cubo.

Todas essas interligações são feitas por tubos cilíndricos que possuem escadas em seu interior, permitindo o deslocamento de pessoas pela parte interna da estrutura. Na diagonal ortogonal à base, o deslocamento é feito por um elevador, que permite o deslocamento entre as esferas da base e a esfera do ponto mais alto, passando pela esfera central.

Considere um visitante que se deslocou pelo interior do Atomium sempre em linha reta e seguindo o menor trajeto entre dois vértices, passando por todas as arestas e todas as diagonais do cubo.

A projeção ortogonal sobre o plano do solo do trajeto percorrido por esse visitante é representada por



Disponível em: <http://tripedatrip.com>.
Acesso em: 25 out. 2019.

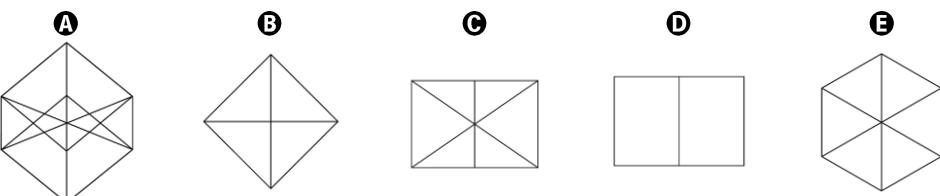


Figura 8.2 – Questão 180, prova amarela, Enem 2021

Fonte: Prova Regular Amarela/Inep, 2021, pág. 31

8.5 Comparação de escores

Um grande desafio das avaliações em larga escala reside na comparabilidade, visto que, uma vez aplicado o teste, deseja-se mensurar e distinguir resultados, inclusive de públicos, locais e períodos diferentes. A TRI permite que itens de diferentes provas ou aplicações sejam comparáveis entre si, desde que estejam calibrados na mesma escala de proficiência. Isso é possível porque os parâmetros dos itens são estimados em uma métrica comum, o que garante que a posição de cada item na escala represente o nível de proficiência necessário para acertá-lo.

Além disso, mesmo que dois conjuntos de itens sejam calibrados em escalas distintas, é possível realizar um processo de equalização, que consiste em alinhar as duas escalas por meio de transformações lineares. Esse procedimento garante que os parâmetros dos itens e as proficiências dos participantes possam ser comparados diretamente, mesmo que tenham sido obtidos em momentos ou populações diferentes.

A equalização pode ser feita por diversos métodos, como o uso de itens comuns ou itens âncora entre provas ou por técnicas estatísticas que ajustam os parâmetros para que fiquem na mesma régua.

Equalizar, no caso da TRI, significa colocar parâmetros de itens vindos de provas distintas em uma escala comum, tornando-os comparáveis. No âmbito do Enem, a equalização se dá via itens comuns, que no caso são os itens advindos dos pré-testes. Esses itens funcionam como itens de ligação e garantem que os parâmetros dos itens procedentes de diferentes aplicações estejam na mesma escala e, portanto, sejam comparáveis (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2021b).

8.6 Coerência pedagógica

Para finalizar este capítulo, faremos três estudos de caso. Em todos eles, usaremos o modelo ML3P com a JMLE, nas mesmas condições, quais sejam:

- Teste com 3 itens, todos com os parâmetros de discriminação e pseudochute iguais, variando apenas o nível de dificuldade, sendo um fácil, um médio e

outro difícil, na métrica (0, 1), organizados na tabela a seguir, de cima para baixo, do menor para o maior valor de b .

Item	a	b	c
Item 1 (fácil)	1,0	-2,0	0,2
Item 2 (médio)	1,0	0,0	0,2
Item 3 (difícil)	1,0	2,0	0,2

Quadro 8.3 – Estudo de caso: parâmetros dos itens

Fonte: autoria própria

- Cinco participantes, sendo a proficiência de cada um deles conhecida de um teste anterior, também na métrica (0, 1). Os estudantes estão organizados na tabela por ordem de aptidão, de cima para baixo, da menor para a maior.

Estudante	θ
João	-2,0
Maria	-1,0
Pedro	0,0
José	1,0
Ana	2,0

Quadro 8.4 – Estudo de caso: aptidões do testados

Fonte: autoria própria

Assim, as probabilidades de acerto também ficarão constantes. Na tabela a seguir, P_1 , P_2 e P_3 representam as probabilidades de acerto respectivamente dos itens 1, 2 e 3.

Estudante	P_1 (fácil)	P_2 (médio)	P_3 (difícil)
João	0,6000	0,2953	0,2144
Maria	0,7849	0,4151	0,2379
Pedro	0,9047	0,6000	0,2953
José	0,9621	0,7849	0,4151
Ana	0,9856	0,9047	0,6000

Quadro 8.5 – Estudo de caso: probabilidades de acerto

Fonte: autoria própria

1) Vamos à primeira situação: se todos os estudantes submetidos ao teste tivessem o padrão de resposta 011, ou seja, errassem a primeira, que é uma questão fácil, e acertassem as demais, que são mais difíceis. Qual deles teria o melhor resultado?

Usando a equação (6.2) da verossimilhança conjunta, vamos calcular a verossimilhança para cada um deles, fazendo $(1 - P_{j1}) \cdot P_{j2} \cdot P_{j3}$.

Estudante	θ	Resposta	Verossimilhança
João	-2,0	011	0,0253
Maria	-1,0	011	0,0213
Pedro	0,0	011	0,0169
José	1,0	011	0,0124
Ana	2,0	011	0,0078

Quadro 8.6 – Estudo de caso 1: verossimilhança

Fonte: autoria própria

2) Vamos agora ao segundo caso: se todos os estudantes submetidos ao teste tivessem o padrão de resposta 001, ou seja, errassem o primeiro e o segundo item, mas acertassem o terceiro, que é um item mais difícil. Qual deles teria o melhor resultado?

Usando novamente a equação (6.2), podemos calcular a verossimilhança conjunta, agora com padrão $(1 - P_{j1}) \cdot (1 - P_{j2}) \cdot P_{j3}$. Assim, obtemos:

Estudante	θ	Resposta	Verossimilhança
João	-2,0	001	0,0604
Maria	-1,0	001	0,0299
Pedro	0,0	001	0,0113
José	1,0	001	0,0034
Ana	2,0	001	0,0008

Quadro 8.7 – Estudo de caso 2: verossimilhança

Fonte: autoria própria

3) Vamos agora ao último caso: se todos os estudantes submetidos ao teste tivessem o padrão de resposta 100, ou seja, acertassem a primeira, que é uma questão fácil, e errassem as demais, que são mais difíceis. Qual deles teria o melhor resultado?

Usando novamente a equação (6.2), podemos calcular a verossimilhança conjunta, agora com padrão $P_{j1} \cdot (1 - P_{j2}) \cdot (1 - P_{j3})$. Assim, obtemos:

Estudante	θ	Resposta	Verossimilhança
João	-2,0	100	0,3321
Maria	-1,0	100	0,3498
Pedro	0,0	100	0,2550
José	1,0	100	0,1211
Ana	2,0	100	0,0376

Quadro 8.8 – Estudo de caso 3: verossimilhança

Fonte: autoria própria

Note que, apesar de esses estudos de caso não condizerem com situações realistas é possível, mesmo assim, tirar algumas conclusões interessantes. No primeiro caso, a partir da tabela de valores da verossimilhança podemos inferir que,

para o padrão 011, a verossimilhança é maior para valores mais baixos de θ . Mesmo que um desses participantes tenha acertado os itens mais difíceis, o modelo considera mais plausível que ele tenha proficiência baixa, porque o erro no item fácil é estatisticamente “inaceitável” para alguém com alta proficiência, porém, mais aceitável para alguém com baixa proficiência.

Acertos em itens difíceis não compensam erros em itens fáceis. Isso acontece porque o modelo 3PL já admite que pessoas com proficiência média ou baixa podem acertar itens difíceis por sorte ou esforço. Mas errar um item fácil, cuja probabilidade de acerto é altíssima para qualquer θ acima de $-1,0$, derruba a verossimilhança dos valores altos de θ .

Já no segundo caso, podemos perceber que o padrão de resposta 001 é mais coerente com uma habilidade baixa, pois o modelo considera plausível que alguém com baixa proficiência erre os itens mais simples e acerte o difícil por sorte ou esforço. À medida que a proficiência aumenta, esse padrão se torna estatisticamente menos aceitável, já que erros em itens fáceis são penalizados fortemente.

Na terceira situação, a verossimilhança conjunta, calculada com o padrão de resposta 100, indica que esse padrão é mais coerente com proficiências baixas, pois o modelo considera plausível que estudantes com menor habilidade acertem o item fácil e falhem nos demais. À medida que a proficiência aumenta, espera-se maior consistência nos acertos, especialmente nos itens mais difíceis, e por isso os valores de verossimilhança diminuem para os estudantes com θ mais elevado. O modelo, portanto, favorece uma coerência entre o padrão de resposta e o nível de proficiência estimado, penalizando incoerências em proficiências altas.

Uma outra análise, a partir da comparação das proficiências obtidas por um mesmo estudante, também pode ser bem interessante. Veja:

Estudante	θ	Resposta 011	Resposta 001	Resposta 100
João	-2,0	0,0253	0,0604	0,3321
Maria	-1,0	0,0213	0,0299	0,3498
Pedro	0,0	0,0169	0,0113	0,2550
José	1,0	0,0124	0,0034	0,1211
Ana	2,0	0,0078	0,0008	0,0376

Quadro 8.9 – Comparação dos escores

Fonte: autoria própria

Perceba que todos os estudantes conseguiram melhores resultados quando acertam uma única questão: a mais fácil. Então, em avaliações que usam a TRI

como metodologia, não necessariamente recebem as maiores notas aqueles que acertam os itens mais difíceis ou os que acertam mais itens: é necessário apresentar uma coerência pedagógica, acertando primeiro os itens mais fáceis e progredindo até os mais difíceis.

Os Apêndices B e C contém uma pequena análise estatística descritiva das provas de Matemática e suas Tecnologias do Enem 2023 e 2024, respectivamente. Dentre as análises feitas, estão aquelas relativas às melhores e às piores coerências pedagógicas nessas provas, conforme amostra utilizada. No Apêndice A, é possível encontrar um Recurso Educacional que utiliza as informações contidas nos microdados do Enem para construir uma escala de proficiência simplificada com a finalidade de monitorar a aprendizagem em Matemática dos estudantes.

9

As Crônicas do Enem: O Teorema, a Distribuição e a Escala

O Exame Nacional do Ensino Médio (Enem) é uma avaliação de larga escala aplicada anualmente no Brasil, voltada principalmente para estudantes que estão concludo ou já concluíram o ensino médio. Criado em 1998, o exame inicialmente tinha como objetivo avaliar a qualidade da educação básica no país, mas ao longo dos anos passou a desempenhar um papel central no acesso ao ensino superior. Atualmente, os resultados do exame são utilizados em processos seletivos de universidades públicas (via SISU), programas de financiamento estudantil (como o FIES), bolsas de estudo (via ProUni), além de servirem como critério de ingresso em instituições estrangeiras.

A partir de 2009, o Enem passou por uma reformulação metodológica significativa, adotando a Teoria da Resposta ao Item (TRI) como base para a construção e correção das provas. Essa abordagem permite uma avaliação mais precisa da proficiência dos participantes, levando em consideração não apenas o número de acertos, mas também as características dos itens e a coerência pedagógica das respostas. Com a TRI, o exame passou a oferecer uma escala de proficiência mais robusta e comparável entre diferentes edições, fortalecendo sua credibilidade como instrumento de seleção e diagnóstico educacional.

No Enem, para estimar os parâmetros e as proficiências dos estudantes, é utilizado uma abordagem estatística híbrida frequentista/bayesiana, envolvendo MMLE (*Marginal Maximum Likelihood Estimation*), para a determinação dos parâmetros dos itens, EAP (*Expected a Posteriori*), para a estimação das proficiências, além da quadratura de Gauss-Hermite, para o cálculo numérico aproximado de integrais exponenciais envolvidas nesse procedimento. Neste capítulo, apresentaremos, de forma sucinta, a metodologia utilizada para o cálculo da proficiência dos estudantes no Enem, passando pelo Teorema de Bayes, pelas distribuições *a priori* e *a posteriori* e pelo valor esperado para uma variável aleatória contínua.

9.1 O Teorema de Bayes e as distribuições *a priori* e *a posteriori*

O Teorema de Bayes permite atualizar a probabilidade de um evento com base em novas informações, sendo essencial para a probabilidade condicional, ou

seja, quando queremos saber a chance de algo acontecer dado que outra coisa já aconteceu. Assim, o teorema mostra como revisar uma probabilidade inicial quando recebemos novas evidências, gerando uma probabilidade atualizada.

Segundo Meyer (1983), considere A e B dois eventos associados a um experimento aleatório. Denotando por $P(A|B)$ a probabilidade da ocorrência do evento A, sabendo que o evento B ocorreu, com $P(B) > 0$, temos

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B) \cdot P(B) , \quad (9.1)$$

que, em outras palavras, é a probabilidade condicionada da ocorrência do evento B à ocorrência do evento A. Por outro lado, com $P(A) > 0$, temos,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(B|A) \cdot P(A) . \quad (9.2)$$

Das equações (9.1) e (9.2), encontramos

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \xrightarrow{+ P(B)} P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} , \quad (9.3)$$

que é a versão mais simples do Teorema de Bayes, nos ensinando como calcular a probabilidade de algo ser verdadeiro depois de observada uma informação relevante. Por exemplo, sendo $P(A)$ a probabilidade de um estudante ter alta proficiência em Matemática e $P(B)$ a probabilidade de o estudante ter acertado um item difícil, então a probabilidade desse aluno ter alta proficiência, dado que ele acertou um item difícil, é dada pela equação (9.3).

A distribuição *a priori* é a probabilidade assumida para um parâmetro antes da observação dos dados. Em contextos bayesianos, ou seja, qualquer situação em que usamos o Teorema de Bayes para atualizar nossas crenças ou estimativas com base em novas evidências, a distribuição *a priori* representa o que se sabe (ou se

assume saber) sobre uma variável aleatória com base em conhecimento anterior ou suposições teóricas. Já a distribuição *a posteriori* é a distribuição atualizada dessa variável obtida após a observação dos dados. Ela combina a distribuição *a priori* com a verossimilhança das observações usando o Teorema de Bayes. Assim, a distribuição *a posteriori* representa o que se passa a acreditar sobre essa variável após considerar os dados observados.

9.2 Valor esperado de uma variável aleatória contínua

De acordo com Meyer (1983), se X é uma variável contínua com função densidade de probabilidade $g(x)$, então

$$E[X] = \int_{-\infty}^{\infty} x \cdot g(x) dx \quad (9.4)$$

é o seu valor esperado.

Então, para se obter o valor esperado de uma variável aleatória contínua, é necessário calcular uma integral que, nada mais é, que a soma de infinitas parcelas compostas pelo produto de x com $g(x)$. Perceba que tal integral está sendo dividida pela soma dos valores da densidade de probabilidade $g(x)$, uma vez que $g(x)$ “integra 1”. Logo, podemos dizer que o valor esperado de uma variável aleatória contínua é uma média ponderada, num caso contínuo, de todos os valores que essa variável pode assumir, onde os pesos são dados pela função densidade de probabilidade.

O valor esperado de uma variável aleatória contínua, portanto, é calculado por meio da integral do produto entre cada valor possível e sua respectiva densidade de probabilidade, equivalendo a uma média ponderada. No método EAP, utilizado para estimar a proficiência dos participantes no Enem, esse princípio é aplicado diretamente: a proficiência de cada participante é obtida por meio da média dos possíveis valores do traço latente θ , segundo a distribuição *a posteriori* – obtida ao ponderar a distribuição *a priori* pela verossimilhança das respostas observadas. Esse mesmo raciocínio fundamenta a construção de escalas em avaliações baseadas na Teoria da Resposta ao Item.

9.3 JMLE e MMLE

Da seção 5.3, a verossimilhança é uma função dos parâmetros do modelo. Mais especificamente, ela representa a probabilidade de se observar os dados obtidos, dado um conjunto de parâmetros. Por essa razão, embora se origine de uma função de probabilidade, a verossimilhança não é uma distribuição de probabilidade sobre os parâmetros e, portanto, não precisa “integrar 1”. Ainda assim, ela constitui uma ferramenta fundamental para a estimação desses parâmetros

Dito isso, retomando a discussão que foi feita na seção 5.2 o “JMLE (*Joint Maximum Likelihood Estimation*) e o MMLE (*Marginal Maximum Likelihood Estimation*) são métodos distintos de estimação de parâmetros que usam a função de verossimilhança. O JMLE estima simultaneamente os parâmetros dos itens λ e os traços latentes θ dos indivíduos, como foi descrito ao longo deste trabalho. Já o MMLE marginaliza os traços latentes, tratando-os como variáveis aleatórias e integrando-os fora da função de verossimilhança. Dizemos, nesse caso, que estamos calculando a verossimilhança marginalizando sobre θ .

Então, no caso do MMLE, assume-se uma distribuição (função densidade de probabilidade) $g(x)$ *a priori* para θ , por exemplo, a normal padrão, e realiza-se a integração dessa variável na função de verossimilhança. Com isso, θ deixa de ser tratado como parâmetro fixo e passa a ser considerado uma variável aleatória com distribuição conhecida. O objetivo é eliminar θ da função de verossimilhança, permitindo a estimação dos parâmetros dos itens de forma independente dos traços latentes.

Diferentemente do JMLE, em que a estimação dos parâmetros dos itens e dos traços latentes ocorre simultaneamente em uma única etapa, o MMLE realiza esse processo em duas etapas distintas: primeiro, estima-se os parâmetros dos itens com base na verossimilhança marginal, de forma independente de θ , e só depois os traços individuais são estimados. Os traços individuais θ podem ser estimados por meio de abordagens frequentistas, como o MLE (*Maximum Likelihood Estimation*), ou bayesianas, como MAP (*Maximum a Posteriori*), EAP (*Expected a Posteriori*), MCMC (*Markov Chain Monte Carlo*) ou EB (*Empirical Bayes*).

Na abordagem MMLE, a distribuição de θ é previamente definida e permanece fixa ao longo da estimação dos itens. Diferentemente de algumas abordagens bayesianas (como o MCMC) ou o EB, o MMLE assume uma distribuição *a priori*

para θ e não a ajusta com base nos dados que são observados durante ou após o processo de estimação. No entanto, quando o MMLE é integrado a métodos bayesianos de escoragem, como o EAP, dentro de uma abordagem híbrida, a distribuição inicial de θ pode ser interpretada como uma distribuição *a priori*. Isso ocorre porque ela é reutilizada na etapa de estimação da proficiência, sendo combinada com a verossimilhança das respostas individuais para formar a distribuição *a posteriori*.

Assim, embora o MMLE, isoladamente, não revise a distribuição de θ , sua combinação com técnicas bayesianas confere à distribuição original o papel de uma *priori* no sentido clássico. A distribuição posterior resultante incorpora tanto a informação prévia quanto os dados empíricos observados. Esse processo é especialmente relevante no contexto da Teoria de Resposta ao Item (TRI), onde a escoragem precisa refletir essa integração entre conhecimento prévio e evidência observada.

Por fim, no MMLE, a função de verossimilhança marginal é construída a partir da função de verossimilhança conjunta. Assim, a função de verossimilhança marginal contém o mesmo produtório da verossimilhança conjunta, mas ela não é a verossimilhança conjunta completa, porque θ não é tratado mais como parâmetro a ser estimado, mas sim como variável aleatória com distribuição definida. Ao integrar sobre todos os possíveis valores de θ , passamos a não tratar todos os valores como igualmente prováveis. Em vez disso, damos mais peso aos valores de θ que são mais prováveis segundo a distribuição assumida, que geralmente é uma normal padrão. Isso torna a verossimilhança marginal uma espécie de “média ponderada” das verossimilhanças condicionais.

Além disso, ao integrar θ , evitamos ter que estimar um parâmetro para cada indivíduo, o que seria impraticável em grandes amostras, contornando alguns problemas que aparecem na verossimilhança conjunta devido a isso. A única questão é que a integral obtida dessa forma não é tão simples de ser calculada de forma analítica, mas pode ser resolvida numericamente, por exemplo, por meio da quadratura de Gauss-Hermite. O importante nesse procedimento é que θ foi “eliminado da função de verossimilhança”, mas sem ignorar sua influência, não sendo necessário, portanto, estimar um θ para cada indivíduo ou estimá-lo conjuntamente com os parâmetros dos itens para determinar os parâmetros dos itens.

9.4 Estimação dos parâmetros e das proficiências no Enem

Este trabalho tem como foco o processo de otimização na estimação das aptidões e proficiências em testes de proficiência com a abordagem frequentista da JMML aplicada à TRI. No entanto, uma notável aplicação da Teoria da Resposta ao Item se encontra na determinação de proficiências de estudantes em avaliações em larga escala como o Programa de Avaliação da Educação Básica do Espírito Santo (Paebes), Sistema de Avaliação da Educação Básica (Saeb), o Exame Nacional de Desempenho dos Estudantes (ENADE), Programa Internacional de Avaliação de Estudantes (PISA), *Test of English as a Foreign Language Internet Based Test* (TOEFL iBT), dentre outras, cuja abordagem principal é bayesiana (tradicional ou bayesiana empírica).

Entre as avaliações em larga escala que se beneficiam da aplicação da Teoria de Resposta ao Item com abordagem não frequentista, destaca-se o Exame Nacional do Ensino Médio (Enem). Como consta na publicação Procedimentos de Análise do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2021b), o Enem adota uma abordagem distinta, baseada na estimação marginal por MMLE, um método frequentista, seguida da escoragem dos participantes com o EAP, um método bayesiano. Essa estrutura em duas etapas permite lidar com grandes volumes de dados, otimizado melhor o processo computacional ,quando compactado com o JMML.

A seguir, serão apresentadas, de maneira concisa, as etapas da metodologia usada no Enem, desde a estimação dos parâmetros até o *scoring* dos participantes, de acordo com a publicação do Inep citada anteriormente e a publicação Entendendo sua Nota no Enem, também do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2021a).

9.4.1 Estimação dos parâmetros dos itens via MMLE

Com a matriz de padrões de desempenho consolidada em mãos, inicialmente, assume-se uma distribuição *a priori* para as proficiências θ . Segundo o Inep, esta distribuição é a normal padrão (média 0 e desvio padrão 1), porém é deixado margem para a escolha de outra *a priori*. Com isso, pela independência dos itens, a probabilidade de determinado indivíduo j de nível de aptidão θ_j dar o padrão de

resposta $U_j = (u_{j1}, \dots, u_{jn})$ é a mesma que a probabilidade de esse indivíduo dar, simultaneamente, cada uma de suas respostas individuais u_{ji} , ou seja,

$$\begin{aligned}
 P_j(X_j = U_j | \theta_j, \Lambda) &= P_j((X_{j1} = u_{j1} | \theta_j, \lambda_1), (X_{j2} = u_{j2} | \theta_j, \lambda_2), \\
 &\quad \dots, (X_{ji} = u_{ji} | \theta_j, \lambda_i), \dots, (X_{jn} = u_{jn} | \theta_j, \lambda_n)) \\
 &= P_{ji} \left(\bigcap_{i=1}^n X_{ji} = u_{ji} | \theta_j, \lambda_i \right) \\
 &= \prod_{i=1}^n P_{ji}(X_{ji} = u_{ji} | \theta_j, \lambda_i) , \tag{9.5}
 \end{aligned}$$

chamada de verossimilhança condicional das respostas do participante j , assumindo que conhecemos (ou supomos) a distribuição *a priori* de θ_j por meio da densidade de probabilidade $g(\theta_j)$.

A partir dessa expressão, vamos marginalizar sobre θ_j , integrando, para isso.

Dessa forma,

$$P_j(X_j = U_j | \Lambda) = \int_{-\infty}^{\infty} P_j(X_j = U_j | \theta_j, \Lambda) \cdot g(\theta_j) d\theta_j , \tag{9.6}$$

$$\stackrel{[9.5]}{=} \int_{-\infty}^{\infty} \prod_{i=1}^n P_{ji}(X_{ji} = u_{ji} | \theta_j, \lambda_i) \cdot g(\theta_j) d\theta_j \tag{9.7}$$

que é a verossimilhança marginal para o indivíduo j .

Agora, pela independência individual, as respostas de indivíduos diferentes são independentes, isto é, a probabilidade de um indivíduo j dar o padrão de respostas U_j não afeta a probabilidade do indivíduo k dar o padrão de respostas U_k .

Assim, a realização dos valores $X_1 = U_1$, $X_2 = U_2$, \dots , $X_j = U_j$, \dots e $X_m = U_m$ corresponde à ocorrência de eventos independentes e a probabilidade de os indivíduos de nível de aptidão θ_j darem respostas u_{ji} aos itens de parâmetros λ_i tais que $X = U$, é a mesma que a probabilidade de esses mesmos indivíduos darem o respectivo padrão de resposta U_j a esses itens. Daí

$$\begin{aligned}
 L(\Lambda) &= \prod_{j=1}^m P_j(X_j = U_j | \Lambda) \\
 &\stackrel{[9.7]}{=} \prod_{j=1}^m \int_{-\infty}^{\infty} \prod_{i=1}^n P_{ji}(X_{ji} = u_{ji} | \theta_j, \lambda_i) \cdot g(\theta_j) d\theta_j
 \end{aligned} \tag{9.8}$$

é a verossimilhança marginal com distribuição *a priori* de θ conhecida e igual a densidade de probabilidade $g(\theta_j)$.

Devido à dificuldade de resolver convencionalmente a integral da função de verossimilhança marginal da equação (9.8), usa-se, em geral, a quadratura Gaussiana, mais especificamente a quadratura de Gauss-Hermite, que é um método de integração numérica voltado para funções associadas à distribuição normal.

A integral definida acima não possui uma forma fechada para um modelo geral de resposta ao item. Ela deve ser avaliada numericamente: para $g(\theta)$ normal, o método de escolha é a fórmula da quadratura de Gauss-Hermite,

$$\prod_{j=1}^m P_j(X_j = U_j | \Lambda) \approx \sum_{q=1}^Q P_j(x_q) A(x_q)$$

onde x_q é o ponto de quadratura, ou "nó", e $A(x_q)$ é o peso de quadratura correspondente. Os valores dos pontos e pesos são tabelados ou podem ser gerados a partir de rotinas computacionais amplamente disponíveis. Limites de erro da aproximação ao valor verdadeiro da integral estão disponíveis (Bock; Gibbons, 2021, p. 158, tradução nossa).

O número Q se refere a quantidade de pontos de quadratura que, de acordo com o Inep, no caso do Enem, são 40.

Após a aproximação da verossimilhança marginal via quadratura de Gauss-Hermite, o algoritmo EM é utilizado para encontrar os valores dos parâmetros dos itens que maximizam essa função. No Anexo A deste trabalho está uma tabela com os parâmetros dos itens de Matemática do caderno amarelo da aplicação regular do Enem de 2024 e, no Apêndice D, estão as respectivas CCI.

Com os itens calibrados, o próximo passo é a estimativa da aptidão de θ_j de cada testando. De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2021a), o método EAP (*Expected A Posteriori*), bayesiano, é utilizado para isso, e consiste em calcular o valor esperado da distribuição *a posteriori* de θ , dado o padrão de respostas do participante.

10 Conclusão

Nesta dissertação, a Teoria da Resposta ao Item (TRI) foi explorada sob a perspectiva de um problema de otimização não linear multidimensional, com ênfase na formulação matemática dos modelos logísticos unidimensionais, em particular o modelo logístico de três parâmetros (ML3P). Iniciamos com os fundamentos da psicometria e conceitos estatísticos preliminares, passando pela modelagem da função de verossimilhança conjunta e pela adoção do método *Joint Maximum Likelihood Estimation* (JMLE) para a estimação simultânea dos parâmetros dos itens e das aptidões dos respondentes. Utilizamos o algoritmo de Newton-Raphson para maximizar a função objetivo, obtendo fórmulas de recorrência e a matriz Hessiana para itens e aptidões. Além disso, discutimos a curva característica do item (CCI), a escala de proficiência e estratégias para o monitoramento da aprendizagem matemática de estudantes do Ensino Médio, usando itens de edições anteriores do Enem como ferramenta para uma interpretação pedagógica mais precisa e equitativa.

Os resultados obtidos incluem uma apresentação detalhada das derivadas e expressões analíticas necessárias para a implementação do JMLE no ML3P, demonstrando como o processo de otimização permite estimar parâmetros que capturam a discriminação, dificuldade e acerto ao acaso dos itens, bem como as proficiências latentes. Aplicações práticas, como estudos de caso sobre coerência pedagógica, revelaram que o modelo prioriza padrões de resposta consistentes, penalizando incoerências e valorizando acertos em itens alinhados ao nível de aptidão esperado. Essas análises reforçam o potencial da TRI para avaliações educacionais mais robustas, aproximando a teoria da prática docente e favorecendo intervenções personalizadas na sala de aula.

Por fim, o capítulo dedicado às "Crônicas do Enem" introduz elementos da abordagem bayesiana, como o Teorema de Bayes, distribuições *a priori* e *a posteriori*, e métodos híbridos como o *Marginal Maximum Likelihood Estimation* (MMLE) combinado com *Expected a Posteriori* (EAP), utilizados no Enem para lidar com grandes volumes de dados e melhorar a precisão das estimativas. Essa integração frequentista-bayesiana representa um caminho promissor para prosseguir com a pesquisa, permitindo explorar avanços na estimação de proficiências em avaliações educacionais em larga escala e seu uso nas unidades escolares.

11

Referências bibliográficas

ARENALES, S.; DAREZZO, A. **Cálculo Numérico: aprendizagem com apoio de software**. 2. ed. São Paulo: Cengage Learning, 2015.

BAKER, F. B. **The Basics of Item Response Theory**. 2. ed. [S. l.]: ERIC, 2001. Disponível em: <https://files.eric.ed.gov/fulltext/ED458219.pdf>. Acesso em: 29 jun. 2025.

BOCK, R. D.; GIBBONS, R. D. **Item Response Theory**. Hoboken: John Wiley & Sons, 2021.

BORTOLOSSI, H. J. **Cálculo Diferencial a Várias variáveis: uma introdução à teoria da otimização**. Rio de Janeiro, São Paulo: Editora PUC-Rio, Edições Loyola, 2002.

BUSSAB, W. de O.; MORETTIN, P. A. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010.

DE ANDRADE, D. F.; TAVARES, H. R.; DA CUNHA VALLE, R. **Teria da Resposta ao Item: conceitos e aplicações**. São Paulo: ABE, 2000.

HALEY, K. D. C. **Estimation of the Dosage Mortality Relationship When the Dose Is Subject to Error**. 1952. Tese (Doutorado em Estatística) – Stanford University, Stanford, 1952.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Entenda Sua nota no enem: guia do participante**. Brasília, DF: Inep, 2021a. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_examens_da_educacao_basica/entenda_a_sua_nota_no_enem_guia_do_participante.pdf.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Exame Nacional do Ensino Médio - Enem: procedimentos de análise**. Brasília, DF: Inep, 2021b. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_examens_da_educacao_basica/enem_procedimentos_de_analise.pdf.

JOHN A. DRAKOPOULOS. Sigmoidal theory. **Fuzzy Sets and Systems**, [s. l.], v. 76, n. 3, p. 349–363, 1995. [https://doi.org/10.1016/0165-0114\(95\)00028-5](https://doi.org/10.1016/0165-0114(95)00028-5).

LUCKESI, C. C. **Avaliação da aprendizagem escolar: estudos e proposições**. 22. ed. São Paulo: Cortez, 2013.

MEYER, P. L. **Probabilidade: aplicações à estatística**. 2. ed. Rio de Janeiro: LTC, 1983.

PASQUALI, L. **Psicometria: Teoria dos Testes na Psicologia e na Educação**. 5. ed. Petrópolis: Vozes, 2003.

PASQUALI, L. Revista da Escola de Enfermagem da USP. **Psicometria**, [s. l.], v. 43, p. 992–999, 2009. <https://doi.org/10.1590/S0080-62342009000500002>.

PASQUALI, L. **TRI - Teoria da Resposta ao item: teoria, procedimentos e aplicações**. 1. ed. Curitiba: Appris, 2018.

RABELO, M. **Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro**. Rio de Janeiro: SBM, 2013.

SCHEINERMAN, E. R. **Matemática Discreta: uma introdução**. 2. ed. São Paulo: Cengage Learning, 2011.

THURSTONE, L. L. Attitudes Can Be Measured. **American Journal of Sociology**, [s. l.], v. 33, n. 4, p. 529–554, 1 jan. 1928. <https://doi.org/10.1086/214483>.

URBINA, S. **Fundamentos da Testagem Psicológica**. 1. ed. Porto Alegre: Artmed, 2007.

APÊNDICE A – Recurso Educacional

PONTIFÍCIA U
NIVERSIDADE C
ATÓLICA
DO RIO DE JANEIRO



Gabriel Minchio Ayres

MONITORAMENTO DA APRENDIZAGEM MATEMÁTICA COM A TRI escalas de proficiência do Enem simplificadas

Recurso Educacional

Recurso Educacional apresentado como requisito parcial para obtenção do grau de Mestre, no Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

Orientador: Prof. Sinésio Pesco

Rio de Janeiro,
Setembro de 2025

Sumário

1	Introdução.....	147
2	Sobre o recurso educacional.....	148
2.1	Público-alvo e pré-requisitos	149
2.2	Materiais e recursos	149
2.3	Inovação, replicabilidade e impacto	149
3	Escala de Proficiência Simplificada.....	151
3.1	Obtenção das informações dos itens	151
3.2	Construção das escalas de proficiência simplificadas	154
4	Propostas de atividades.....	157
4.1	Proposta 1: caminho autônomo com apoio entre pares	157
4.2	Proposta 2: intervenção guiada pelo Professor.....	157
4.3	Proposta 3: desafio gamificado entre estudantes	158
4.4	Plano de Aula Integrado (para qualquer das três propostas).....	158
5	Para finalizar	160
	Tabelas e Escalas de Proficiência Simplificadas.....	161
	Referências bibliográficas.....	191

Carta ao professor

Este recurso educacional foi desenvolvido com o objetivo de apoiar professores do Ensino Médio na utilização da Teoria da Resposta ao Item (TRI) como ferramenta de monitoramento da aprendizagem em matemática. A proposta está centrada na construção de escalas de proficiência simplificadas por habilidade, a partir da análise dos microdados do Exame Nacional do Ensino Médio (Enem). Por meio da identificação dos parâmetros de dificuldade dos itens e da organização desses itens em uma escala ordenada, semelhante à apresentada no capítulo 8 da dissertação *“Da resposta ao traço latente: a TRI como um problema de otimização”*, o professor poderá localizar o estudante em uma posição relativa à habilidade avaliada.

Essa localização é especialmente útil para avaliações diagnósticas e estratégias de recomposição da aprendizagem, pois permite a proposição de trilhas de estudo personalizadas, favorecendo uma abordagem formativa da avaliação. Ao considerar o conhecimento prévio do estudante, o professor pode planejar intervenções mais precisas, promovendo o avanço progressivo no domínio dos objetos do conhecimento e habilidades da Matemática.

O recurso foi pensado para ser acessível e aplicável em diferentes contextos escolares, respeitando os limites técnicos e pedagógicos da TRI. Sua estrutura permite que seja utilizado mesmo por professores com pouca familiaridade com tecnologia da informação, ao mesmo tempo em que oferece possibilidades de aprofundamento para aqueles que desejam explorar mais. Recomenda-se que o professor explore os materiais complementares disponibilizados, como tutoriais, exemplos de aplicação e referências teóricas, a fim de potencializar o uso do recurso em sua prática docente e ampliar sua capacidade de análise e intervenção sobre os processos de aprendizagem dos estudantes.

Carta ao estudante

Você já se perguntou por que algumas questões do ENEM parecem mais difíceis que outras, mesmo tratando do mesmo conteúdo? Ou por que às vezes acerta uma pergunta complexa e erra uma mais simples? Isso não acontece por acaso. Existe uma forma de entender melhor seu desempenho e usar isso a seu favor: a Teoria da Resposta ao Item (TRI).

Este recurso educacional foi criado para te ajudar a enxergar o aprendizado de matemática como uma jornada, com pontos de partida, caminhos possíveis e metas claras. A ideia é que você possa se localizar nessa trajetória e, com o apoio do professor, traçar estratégias para avançar, respeitando o seu ritmo e o que você já sabe.

Aqui, você vai conhecer uma ferramenta que pode te ajudar bastante: a escala de proficiência por habilidade simplificada. Com ela, é possível entender como seu desempenho em questões anteriores do Enem revela seu nível de domínio em cada habilidade matemática. A partir disso, dá para montar um plano de estudos mais inteligente, focado no que realmente precisa ser reforçado. Nada de estudar tudo ao mesmo tempo, o foco é evoluir com estratégia. Esse recurso pode ser usado inclusive nas aulas de Projeto de Vida ou Estudo Orientado, se sua escola oferecer esses componentes. Vale a pena mostrar ao seu professor.

O mais importante: este material não foi feito para te julgar, mas para te ajudar. Ele foi pensado para te dar autonomia, clareza e motivação. Você vai entender melhor seus pontos fortes, saber onde precisa se dedicar mais e, principalmente, perceber que aprender matemática pode ser mais estratégico e menos sofrido. E isso inclui também a preparação para o ENEM.

Então, bora começar? Seu caminho na matemática está só começando e, agora, você tem uma ferramenta para caminhar com mais confiança.

O autor

Gabriel Minchio, autor deste recurso educacional, é egresso do ensino médio do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo – Ifes, instituição que, à época de sua formação, ainda se denominava Centro Federal de Educação Tecnológica do Espírito Santo (Cefetes). Motivado pelo ambiente acadêmico e pela atuação inspiradora de seus professores, cursou Licenciatura em Matemática na Universidade Federal do Espírito Santo, a Ufes. Desde então, tem atuado como professor de Matemática tanto no Ensino Fundamental II quanto no Ensino Médio da rede privada e, atualmente, exerce a docência no Ensino Médio e na Educação de Jovens e Adultos - Ensino Médio, na Grande Vitória, além de atuar como técnico pedagógico na Secretaria da Educação do Espírito Santo – SEDU.

Ao longo de sua trajetória como docente, Gabriel Minchio esteve constantemente envolvido na preparação de estudantes para o Exame Nacional do Ensino Médio (Enem), avaliação que é fundamentada na Teoria da Resposta ao Item (TRI). Atuando na SEDU, foi diversas vezes convidado a analisar o desempenho da rede estadual em avaliações como o Sistema de Avaliação da Educação Básica (SAEB) e o Programa de Avaliação da Educação Básica do Espírito Santo (PAEBES), contribuindo com estudos e análises voltados ao Índice de Desenvolvimento da Educação Básica (Ideb) da Rede Estadual. Essas vivências despertaram seu interesse por uma compreensão mais aprofundada da TRI, o que o motivou a ingressar no Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, na Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Sob orientação do professor Sinesio Pesco, desenvolveu a dissertação intitulada *“Da resposta ao traço latente: a TRI como um problema de otimização”*, da qual este recurso educacional é parte integrante.

1

Introdução

A Teoria da Resposta ao Item (TRI) é uma abordagem estatística que permite estimar a proficiência dos estudantes com base em seus padrões de resposta, levando em consideração características dos itens como dificuldade, discriminação e acerto casual. Embora amplamente utilizada em avaliações de larga escala, como o Enem e o SAEB, sua aplicação em contextos escolares ainda enfrenta desafios, especialmente pela complexidade técnica envolvida e pela limitação de recursos disponíveis. No entanto, seus benefícios são expressivos: a TRI oferece uma leitura mais precisa do desempenho dos alunos, permite comparações entre avaliações realizadas em diferentes momentos e favorece práticas mais justas de avaliação.

Este recurso educacional foi concebido a partir da sistematização e adaptação de diversas práticas docentes desenvolvidas pelo autor ao longo dos anos, aliadas a análises recorrentes dos microdados do Enem. Trata-se de um texto didático com foco na avaliação diagnóstica e na divulgação tecnológica, alinhado às linhas de pesquisa do PROFMAT, especialmente nas áreas de avaliação educacional. O trabalho tem como principais referências os livros “*TRI: teoria, procedimentos e aplicações*”, de Luiz Pasquali (Editora Appris), e “*Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro*”, da coleção PROFMAT, escrito por Mauro Rabelo e publicado pela Sociedade Brasileira de Matemática (SBM).

Este material apresenta como proposta central a construção de escalas de proficiência por habilidade, utilizando os microdados do Enem como base. A partir da extração dos parâmetros de dificuldade dos itens, o professor pode organizar uma reta de proficiência simplificada, posicionar os estudantes nela e, com base nessa localização, propor caminhos de aprendizagem personalizados, respeitando o ponto de partida de cada aluno. Essa abordagem permite ao professor atuar de forma mais estratégica, promovendo intervenções pedagógicas mais precisas e alinhadas às reais necessidades dos estudantes.

Deseja-se ao leitor uma leitura proveitosa e uma experiência enriquecedora com este material. Para aqueles que desejam se aprofundar no tema, os livros citados são excelentes pontos de partida.

2

Sobre o recurso educacional

Este recurso educacional tem como objetivo principal oferecer aos professores do Ensino Médio uma abordagem prática e acessível para o monitoramento da aprendizagem em matemática, fundamentada na Teoria da Resposta ao Item (TRI). A proposta busca promover o uso pedagógico dos microdados do Exame Nacional do Ensino Médio (Enem), permitindo a construção de escalas de proficiência simplificadas por habilidade. Ao integrar dados reais, ferramentas computacionais e uma perspectiva diagnóstica e formativa da avaliação, o material visa ampliar o repertório avaliativo do professor e fortalecer o vínculo entre avaliação e aprendizagem.

A finalidade deste recurso é apoiar o trabalho docente com instrumentos que possibilitem uma leitura mais precisa do desempenho dos estudantes, contribuindo para o planejamento de intervenções pedagógicas mais eficazes. Ao permitir a localização do aluno em uma escala de proficiência, o professor pode construir trilhas de aprendizagem personalizadas, respeitando o conhecimento prévio e o ritmo de cada estudante. Essa abordagem também contribui para a preparação dos alunos para avaliações externas, como o próprio Enem, ao oferecer estimativas de proficiência comparáveis às dos participantes oficiais.

Embora destinado principalmente a professores de matemática do Ensino Médio, o recurso também pode ser explorado por estudantes interessados em compreender melhor seu processo de aprendizagem e por pesquisadores da área de avaliação educacional. Sua estrutura foi pensada para ser acessível e aplicável em diferentes contextos escolares, permitindo que profissionais com distintos níveis de familiaridade com tecnologia possam utilizá-lo de forma efetiva, especialmente por meio da construção e interpretação de escalas de proficiência simplificadas.

Sabemos dos desafios enfrentados pelos professores, especialmente no ensino da Matemática, diante da escassez de tempo, dos recursos escolares limitados e do baixo engajamento dos estudantes, uma realidade ainda comum em muitas escolas. Diante desse cenário, este material foi pensado como uma alternativa viável e motivadora, que convida o professor a experimentar uma abordagem avaliativa mais significativa e conectada à realidade dos alunos. Em vez de recorrer a estratégias genéricas ou pouco direcionadas, o recurso propõe ferramentas que permitem identificar com mais precisão as lacunas de aprendizagem e os caminhos possíveis para superá-las. Com isso, o tempo em sala de aula pode ser melhor aproveitado, com ações pedagógicas mais focadas e alinhadas às reais necessidades dos estudantes, tornando o processo de ensino mais eficiente e menos baseado em tentativas.

Espera-se que, com a utilização deste material, os professores possam desenvolver práticas avaliativas mais robustas e estratégicas, promovendo uma cultura de avaliação voltada para o desenvolvimento contínuo dos estudantes. Ao tornar visível o percurso de aprendizagem e oferecer instrumentos para sua análise, o recurso contribui para uma educação matemática mais justa, personalizada e alinhada às demandas contemporâneas da escola e da sociedade.

2.1 Público-alvo e pré-requisitos

Este recurso educacional é voltado principalmente para professores e estudantes do Ensino Médio, com atenção especial aos docentes de Matemática. No entanto, também pode ser adaptado por professores de outras áreas e trabalhado de forma conjunta com docentes de componentes integradores, como Projeto de Vida e Estudo Orientado, quando presentes na grade curricular.

A proposta envolve o uso de recursos tecnológicos acessíveis, como planilhas eletrônicas, que facilitam a organização e análise dos dados utilizados na construção das escalas de proficiência. Embora o domínio dessas tecnologias possa contribuir para uma implementação mais fluida, o mais importante é a disposição para aprender e aplicar as práticas sugeridas. Pensando nisso, o material foi elaborado com orientações passo a passo, buscando atender inclusive aqueles que têm pouca familiaridade com informática.

Vale destacar que as ferramentas digitais utilizadas neste recurso estão atualizadas até 30 de setembro de 2025. Caso haja mudanças futuras, recomenda-se consultar a documentação oficial de cada ferramenta para verificar os procedimentos vigentes. Mais do que seguir instruções técnicas, o objetivo é oferecer ideias que possam ser adaptadas conforme os recursos disponíveis no momento da aplicação, garantindo flexibilidade e aplicabilidade em diferentes contextos escolares.

2.2 Materiais e recursos

Para a implementação deste recurso educacional, será necessário o uso de alguns materiais básicos e recursos tecnológicos que viabilizam sua aplicação em ambientes físicos e digitais. A lista de itens inclui:

- Material impresso, caso os testes ou atividades sejam aplicados em papel;
- Computador com acesso à internet;
- Aplicativo de leitura e edição de planilhas eletrônicas, como o Microsoft Excel.

Embora o Excel seja amplamente utilizado, existem outras opções disponíveis no mercado que também atendem às necessidades deste recurso, como o Google Planilhas, LibreOffice Calc e WPS Office.

2.3 Inovação, replicabilidade e impacto

Este recurso educacional apresenta alto teor inovador ao disponibilizar, de forma acessível, ferramentas inéditas para aplicação da Teoria da Resposta ao Item (TRI) em sala de aula. O conhecimento técnico necessário para utilizar a TRI, bem como o domínio das metodologias associadas, costuma ser restrito a contextos muito específicos, o que torna sua presença no cotidiano escolar ainda limitada. Pensando nisso, este material foi concebido com foco na divulgação e na

democratização do acesso a essas abordagens, permitindo que professores possam incorporá-las à prática pedagógica de maneira simples, porém eficaz.

A proposta integra dados reais do Enem com procedimentos acessíveis de organização e análise, promovendo uma leitura mais precisa da aprendizagem dos estudantes por meio da construção de escalas de proficiência simplificadas. Trata-se da sistematização de uma série de atividades desenvolvidas pelo autor ao longo de sua trajetória docente, reunidas e adaptadas para formar um conjunto coerente e aplicável em diferentes contextos escolares.

Com alto potencial de impacto, este recurso educacional foi idealizado para atender tanto escolas públicas quanto privadas, abrangendo públicos com diferentes níveis de familiaridade tecnológica e distintas condições de infraestrutura. Ao oferecer alternativas práticas e escaláveis, contribui para ampliar o repertório avaliativo dos professores e fortalecer o vínculo entre avaliação e aprendizagem.

3

Escala de Proficiência Simplificada

Neste primeiro momento, iremos preparar o terreno para a execução nas aulas. Utilizaremos os microdados do Enem para extrair informações sobre os itens da prova, especificamente sobre o parâmetro de dificuldade dos itens. A partir desses dados, será possível organizar os itens em uma escala ordenada, do mais fácil ao mais difícil, construindo o que chamaremos de escala de proficiência simplificada. Essa escala funciona como uma régua que mede o nível de dificuldade dos itens, sobre a qual posicionaremos também o estudante em termos de quais questões ele consegue ou não responder.

Chamaremos essa escala de “simplificada” porque ela utiliza um procedimento menos técnico do que os adotados em avaliações de larga escala, mas ainda assim oferece uma estimativa bastante útil da proficiência dos alunos. Após a aplicação dos itens dessa escala aos estudantes, por meio de um teste, por exemplo, os resultados obtidos permitirão ao professor posicionar cada estudante em um ponto da reta de proficiência, identificando quais habilidades já foram consolidadas e quais ainda precisam ser desenvolvidas.

Essa estimativa inicial serve como base para traçar caminhos de aprendizagem mais precisos e personalizados, orientando os próximos passos de cada estudante de acordo com seu nível atual. Com isso, o professor deixa de atuar de forma genérica e passa a construir intervenções mais estratégicas, respeitando o ritmo e as necessidades reais de cada aluno.

3.1

Obtenção das informações dos itens

Os microdados são conjuntos de informações detalhadas e anonimizadas coletadas em grandes avaliações, como o Enem. Eles incluem dados individuais de cada participante, como respostas às questões, notas, dados socioeconômicos e parâmetros dos itens (questões das provas), organizados em arquivos que permitem análises estatísticas aprofundadas. Diferente dos resultados agregados, os microdados oferecem uma visão granular, ideal para estudos educacionais, construção de escalas de proficiência e desenvolvimento de ferramentas como as propostas neste recurso educacional.

Vamos mostrar como acessar os microdados do Enem da aplicação de 2024, mas passo semelhante pode ser feito para as aplicações dos demais anos. Lembrando que esse passo a passo pode sofrer alterações com o passar do tempo;

1. Acesse a página do portal oficial do Inep que contém os microdados do Enem no endereço a seguir:

<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>

2. Clique no link de download do ano que você quer obter os microdados (no nosso caso, 2024);

3. Você encontrará um arquivo compactado (geralmente em formato .zip) contendo os microdados do ano que você escolheu;
4. Após o download, extraia os arquivos. Eles geralmente incluem:
 - Pasta DADOS: arquivos com dados brutos dos participantes e dos itens;
 - Pasta DICIONÁRIO de variáveis (em .pdf ou .txt);
 - Pasta INPUTS dos scripts utilizados pelo Inep;
 - Pasta DOCUMENTOS TÉCNICOS;
 - Pasta PROVAS E GABARITOS;
5. Na pasta DADOS, abra o arquivo ITENS_PROVA_2024 em um aplicativo de edição de planilhas eletrônicas. Nesse arquivo, cada linha contém informações de um item utilizado no Enem do ano em questão.

As provas do Enem são organizadas em cadernos com diferentes cores, em que as questões são as mesmas, numa mesma aplicação (se não forem provas adaptadas, pois, nesse caso, alguns itens podem sofrer alterações ou serem substituídos), porém apresentadas em outra ordem. Vamos aqui obter os parâmetros da prova de Matemática do caderno amarelo. Para isso, usaremos o Excel. Assim, se necessário, pesquise sobre os comandos utilizados aqui na página oficial do Excel em

<https://support.microsoft.com/pt-br/excel>

- Dito isso, com o arquivo ITENS_PROVA_2024 aberto,
1. Clique em qualquer parte da planilha que contenha algum dado;
 2. Vá em DADOS/FILTRO e ative os filtros;
 3. Na coluna SG_AREA, filtre apenas por MT;
 4. Clique em qualquer parte da planilha que contenha algum dado;
 5. Pressione CONTROL+T, para selecionar toda a planilha de dados filtrada;
 6. Pressione CONTROL+C, para copiar todos os dados filtrados;
 7. Crie uma nova planilha, no mesmo arquivo que está trabalhando;
 8. Clique na célula A1;
 9. Pressione CONTROL+V, para colar os dados copiados;
 10. Clique em qualquer parte dessa nova planilha que contenha algum dado;
 11. Vá em DADOS/CLASSIFICAR e classifique:
 - os dados da coluna TX_COR, de A a Z;
 - os dados da coluna TX_PROVA, do menor para o maior; e
 - os dados da coluna CO_POSICAO do menor para o maior.
 12. Clique em qualquer parte da planilha que contenha algum dado;
 13. Vá em DADOS/REMOVER DUPLICATAS;
 14. Na janela que se abrir, clique em ANULAR TODAS AS SELEÇÕES;
 15. Na mesma janela, selecione:
 - CO_HABILIDADE,
 - NU_PARAM_A,
 - NU_PARAM_B; e
 - NU_PARAM_C.
 16. Ainda nessa mesma janela, clique em OK;
 17. Clique em qualquer parte da planilha que contenha algum dado;
 18. Vá em DADOS/FILTRO e ative os filtros;
 19. Na coluna TX_COR, filtre apenas por AMARELA.

Se tudo tiver dado certo, você terá as 45 questões da prova de Matemática das aplicações regular e as 45 questões PPL/reaplicação, totalizando 90 questões. Porém, estamos interessados, aqui, apenas na aplicação regular. Geralmente, as questões da aplicação regular já estão classificadas com relação às habilidades da Matriz de Habilidades de Competências do Enem. Já os itens das outras aplicações podem ou não ter essa classificação. Assim, se nessa última planilha você verificar que a coluna CO_HABILIDADE contém alguns valores preenchidos e outros não, basta fazer um filtro nessa coluna, desprezando os valores vazios, e você obterá as informações dos itens da primeira aplicação (regular) do Enem de 2024 relativos à área de Matemática suas tecnologias.

No entanto, quando este material foi produzido, não havia valores vazios nessa última planilha. Então, para filtrar apenas as questões da aplicação regular,

1. Na coluna CO_PROVA, filtre um código de cada vez, até o filtrado conter 45 linhas. Quando esse roteiro foi feito, o primeiro código que gerou um filtrado de 45 linhas era o 1324, que foi selecionado para o filtro;
2. Acesse o link a seguir. Ele te levará para a página do Inep de provas e gabarito do Enem;

<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem/provas-e-gabaritos>

3. Clique em 2024, desça a página e encontre “2º Dia – Caderno 5 – Amarelo – Aplicação Regular” ou algo parecido;
4. Clique em Gabarito. Será exibido no seu navegador ou baixado para o seu computador um arquivo com os gabaritos da prova de Matemática e da prova de Ciências da Natureza;
5. Agora, compare esse gabarito das questões 136 a 180, que são referentes aos itens de Matemática, com o gabarito dos respectivos itens da planilha, que está na coluna TX_GABARITO;
 - Se as informações dos dois gabaritos corresponderem, então o filtrado da planilha se refere às questões da aplicação regular;
 - Caso contrário, o filtrado se refere à aplicação PPL/reaplicação, e você deverá filtrar o outro código para obter os itens da aplicação regular. Filtrando pelo outro código, você pode acessar o link anterior e encontrar “2º Dia – Caderno 5 – Amarelo – Reaplicação/PPL”, clicar em Gabarito, e proceder novamente com a comparação dos gabaritos para se certificar de que as questões que foram filtradas se referem à aplicação PPL do Enem de 2024. Quando este roteiro foi feito, o código da prova regular de 2024 foi 1408.
6. Com as linhas que correspondem aos itens de Matemática da aplicação regular do Enem de 2024, você pode selecionar e copiar essas informações para uma outra planilha e salvar esse arquivo.

Pronto, já temos as informações relativas aos itens de Matemática da aplicação regular do Enem de 2024. Note que nessa base há uma coluna chamada NU_PARAM_B. É ela que informa o nível de dificuldade do item, em uma métrica de média 500 e desvio padrão 100, que é a métrica usada no Enem. Além disso, nesse mesmo banco de dados, há uma coluna chamada CO_HABILIDADE, que identifica qual das 30 habilidades da matriz de referência do Enem para a área de Matemática e suas habilidades foi avaliada em cada item.

Você deverá repetir o procedimento que foi feito para obter as informações relativas aos itens de Matemática da aplicação regular do Enem de 2024 para alguns outros anos, até que tenha uma quantidade de itens suficiente para construir uma escala de proficiência simplificada. Aqui, foi feito esse procedimento para as aplicações de 2015 até 2024, obtendo os registros de dificuldade dos itens da área de Matemática e suas Tecnologias das aplicações regulares dos últimos 10 anos, totalizando 441 itens, já descontados os itens anulados ou que não convergiram na aplicação da TRI.

3.2 Construção das escalas de proficiência simplificadas

Com os dados dos itens em mãos, você deverá:

1. empilhar todos os registros, copiando e colando cada base obtida de anos diferentes em uma única planilha, observando se as colunas se mantêm e estão todas na mesma ordem ou se precisa de algum ajuste;
2. Feito isso, você deverá filtrar a coluna CO_HABILIDADE por habilidade, da primeira habilidade até a trigésima, uma por vez, e registrar a posição do item na prova amarela, dada pela coluna CO_POSICAO, e o respectivo nível de dificuldade, dado pela coluna NU_PARAM_B. A ideia é, com a informação dessas três colunas, obtermos uma tabela para cada habilidade.

Por exemplo, a tabela abaixo, obtida seguindo os procedimentos descritos anteriormente, contém todos os itens da Habilidade 1 de Matemática do Enem, no período de 2015 a 2024. Você deverá obter uma tabela que, para o que iremos fazer, só importam essas colunas. As demais podem ser ignoradas por hora. Observe que, além de ser mantida a coluna TX_GABARITO, foi acrescentada a coluna ANO, que não havia originalmente na planilha de dados dos itens.

ANO	CO_POSICAO	TX_GABARITO	CO_HABILIDADE	NU_PARAM_B
2020	149	A	1	-0,08552
2022	144	E	1	-0,02398
2019	148	A	1	0,37213
2019	144	D	1	0,42930
2023	168	C	1	0,63295
2015	177	E	1	0,96427
2024	178	B	1	1,01771
2021	136	D	1	1,21252
2024	176	D	1	1,22835
2015	169	C	1	1,26471
2021	137	D	1	1,40250
2016	158	D	1	1,52149
2022	159	C	1	1,92189
2017	174	E	1	1,94320
2020	147	C	1	2,31084
2018	164	A	1	2,41940
2017	164	C	1	2,65817

Tabela 3.1 – Tabela de itens da Habilidade 1

Fonte: Inep

Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/micrdados>

Acesso: 30 ago. 2025

Obtendo essa base de dados para todas as 30 habilidades, podemos construir nossa escala de proficiência simplificada, uma para cada habilidade, utilizando um gráfico de dispersão, em que o eixo x corresponde ao nível de dificuldade do item e o eixo y, ao ano do item. Assim, para a Habilidade 6,

H6: Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.

obtivemos a escala de proficiência simplificada a seguir.

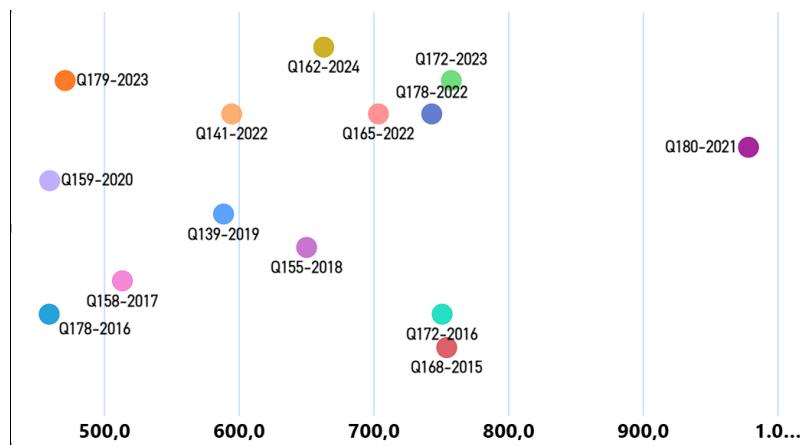
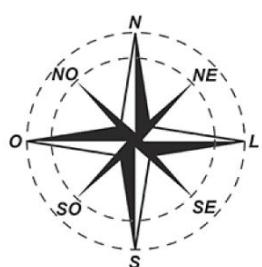


Figura 3.1 – Escala de Proficiência Simplificada para a Habilidade 16 do Enem
Fonte: autoria própria

Assim, se o estudante resolver o item 165 da prova de 2022, mas não conseguir fazer o item 178 de 2022, podemos inferir que a sua proficiência nessa habilidade é próxima de 700. Além disso, supõe-se que esse estudante tem alta probabilidade de acertar os itens à esquerda do item Q165-2022, mas começa a ter baixa probabilidade de acertar as questões à direita da questão desse item. Por exemplo, vamos ver o caso em que o estudante consegue fazer a questão 155 de 2018 (Q155-2018), mas não consegue fazer a questão 162 de 2024 (Q162-2024).

A rosa dos ventos é uma figura que representa oito sentidos, que dividem o círculo em partes iguais.



Uma câmera de vigilância está fixada no teto de um shopping e sua lente pode ser direcionada remotamente, através de um controlador, para qualquer sentido. A lente da câmera está apontada inicialmente no

sentido Oeste e o seu controlador efetua três mudanças consecutivas, a saber:

- 1^a mudança: 135° no sentido anti-horário;
- 2^a mudança: 60° no sentido horário;
- 3^a mudança: 45° no sentido anti-horário.

Após a 3^a mudança, ele é orientado a reposicionar a câmera, com a menor amplitude possível, no sentido Noroeste (NO) devido a um movimento suspeito de um cliente.

Qual mudança de sentido o controlador deve efetuar para reposicionar a câmera?

- A** 75° no sentido horário
- B** 105° no sentido anti-horário
- C** 120° no sentido anti-horário
- D** 135° no sentido anti-horário
- E** 165° no sentido horário

Figura 3.2 – Item 155, aplicação regular de 2018, prova amarela
Fonte: Inep.

Esse item, assim como o seguinte, pertence à Habilidade 6 da Matriz de Habilidade e Competências do Enem, conforme enunciada anteriormente, solicita que o estudante mova conhecimentos elementares sobre geometria para localizar e descrever movimentos no espaço. Porém, os dois itens exigem abordagens distintas: enquanto um item explora a variação angular, o outro explora a projeção ortogonal. Veja, agora, o item 162 de 2024:

Em um jogo virtual para celular, um personagem pode percorrer trajetórias retilíneas voando ou se deslocando ao longo de paredes. Considere que o personagem descreve a trajetória ABCDEF, em que os pontos A, D e E estão em um plano paralelo ao que contém os pontos B e C, sendo esses dois planos ortogonais ao plano da base que contém o ponto F, conforme a figura.

A projeção ortogonal, sobre o plano da base, da trajetória ABCDEF descrita pelo personagem é

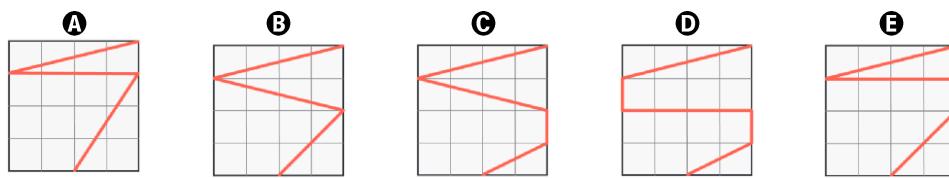
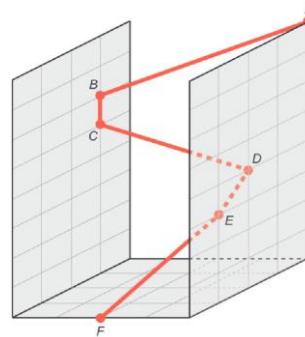


Figura 3.3 – Item 162, aplicação regular de 2024, prova amarela
Fonte: Inep.

Assim, a partir das particularidades observadas na transição entre o erro e o acerto de um item, o professor pode orientar o estudante em seu percurso formativo, direcionando-o para avançar em itens que estão em posições mais à direita ou mais à esquerda daquela em que ele se encontra, dependendo se conseguiu ou não responder corretamente a um item inicial. No caso dos dois itens acima, por exemplo, o estudante demonstrou ter bom domínio de localização em relação à posição angular, mas não em relação à projeção ortogonal. Esse é um indicativo de onde o estudante está nessa habilidade e por onde pode continuar a sua jornada de aprendizado.

Embora esse processo de interrupção na escala nem sempre ocorra de forma linear, ele serve como um bom indicativo das intervenções pedagógicas necessárias. Além disso, permite identificar com objetividade quais ações o estudante precisa desenvolver para aprimorar seu domínio tanto das habilidades envolvidas no item em que houve interrupção quanto dos objetos de conhecimento relacionados a essas habilidades.

Para facilitar a aplicação dessas escalas, disponibilizamos, como anexo deste recurso educacional, as 30 tabelas com suas respectivas escalas de proficiência simplificadas, correspondentes às habilidades de Matemática avaliadas nos itens válidos das edições regulares do Enem (caderno amarelo) de 2015 a 2024.

4

Propostas de atividades

Com as escalas de interesse já definidas, podemos agora pensar em formas de integrá-las às atividades previstas nos planejamentos. Uma proposta é utilizá-las em dinâmicas diversificadas, que além de fornecerem insights valiosos para aplicações futuras, também favorecem o engajamento e a proatividade dos estudantes.

Vejamos algumas possibilidades a seguir.

4.1 Proposta 1: caminho autônomo com apoio entre pares

Na perspectiva de oferecer ao estudante momentos para exercer sua proatividade e promover engajamento, pode-se construir uma atividade como a que será sugerida aqui.

Descrição:

- Cada estudante recebe uma escala de proficiência simplificada e um caderno com questões organizadas conforme essa escala;
- A proposta é que o aluno tente resolver os itens começando por um ponto intermediário;
- Se não conseguir resolver, ele recua para um item à esquerda (mais fácil). Se conseguir, avança para um item à direita (mais difícil);
- Esse processo pode ser feito individualmente ou em duplas, promovendo apoio mútuo e troca de estratégias.

Objetivo: Desenvolver autonomia, autorregulação e consciência sobre o próprio nível de domínio, além de estimular a colaboração entre pares.

4.2 Proposta 2: intervenção guiada pelo Professor.

Em casos em que é necessário um acompanhamento mais próximo, pode-se seguir pela sugestão a seguir.

Descrição:

- O professor seleciona uma série de atividades distribuídas ao longo da escala de proficiência;
- Cada estudante recebe um conjunto inicial de questões e tenta resolvê-las;
- Com base no desempenho, o professor indica novos itens alinhados ao nível de domínio demonstrado pelo estudante, podendo até ajustar a escala, usando uma mais adaptada ao domínio que o estudante apresentou ter;
- A ideia é que o professor atue como orientador, oferecendo desafios progressivos e personalizados.

Objetivo: Promover intervenções pedagógicas mais precisas, respeitando o ritmo de aprendizagem e favorecendo a recomposição de aprendizado.

4.3 Proposta 3: desafio gamificado entre estudantes

Caso seja necessário estimular os estudantes com atividades que os coloquem no centro do processo de aprendizagem, a sugestão a seguir é uma boa opção.

Descrição:

- Cada estudante recebe uma escala de proficiência e escolhe um “caminho” dentro dela, partindo de algum ponto que o professor tiver marcado, ligando itens próximos;
- Esse caminho é desenhado na própria escala e depois proposto como desafio a outro colega, que tentará resolvê-lo, passando por todos os pontos, na ordem escolhida pelo colega que o desafiou;
- A atividade pode ser realizada ao longo de várias aulas, com pontuação acumulativa, criando um ambiente de gamificação. Os caminhos podem ser avaliados pela criatividade, coerência e nível de desafio, por exemplo;
- Nessa perspectiva, é possível jogar em duplas, grupos e até promover uma competição entre turmas!

Objetivo: Estimular o engajamento, a criatividade e o pensamento estratégico, além de promover a revisão de conteúdos de forma lúdica.

4.4 Plano de Aula Integrado (para qualquer das três propostas)

Tema da aula: Exploração de habilidades matemáticas por meio da escala de proficiência simplificada.

Objetivo da aula: Utilizar a escala de proficiência como ferramenta para personalizar o percurso de aprendizagem dos estudantes, promovendo autonomia, colaboração e intervenção pedagógica estratégica.

Tempo estimado: 50 minutos

Etapa	Tempo	Descrição
Organização da sala	5 min	Distribuição das escalas, cadernos de questões e agrupamento dos alunos.
Trabalho individual ou em pares	20 min	Estudantes resolvem os itens conforme a proposta escolhida.
Discussão entre colegas	10 min	Compartilham estratégias, dificuldades e descobertas.
Intervenção do professor	10 min	Circula pela sala, faz perguntas orientadoras, incentiva e motiva.
Fechamento coletivo	5 min	Professor retoma os objetivos, destaca avanços e propõe continuidade.

Avaliação: O mecanismo avaliativo deixaremos em aberto, uma vez que depende da estratégia que o professor adotará.

Observações para o professor:

- Antes da aula, construa a régua de proficiência com base nas habilidades que deseja trabalhar;
- Teste a escala, verificando se as questões realmente pertencem às habilidades descritas pelo Inep;
- Prepare os materiais com antecedência: escalas impressas, cadernos com questões organizadas por nível;
- Durante a aula, evite dar respostas diretas. Prefira perguntas que ajudem o estudante a pensar;
- Incentive os alunos a registrarem seus avanços e dificuldades, criando um histórico de aprendizagem. Inclusive, a escala com esses registros pode ser um mecanismo de avaliação.

5

Para finalizar

A construção e o uso de escalas de proficiência simplificadas representam uma oportunidade valiosa para transformar a avaliação em um processo mais estratégico, formativo e centrado no estudante. Ao organizar os itens por nível de dificuldade e associá-los às habilidades específicas que se deseja desenvolver, o professor ganha uma ferramenta poderosa para planejar intervenções pedagógicas mais precisas, promover trilhas de aprendizagem personalizadas e estimular a autonomia dos alunos.

Além das propostas já apresentadas, como o uso da escala em atividades entre pares, a orientação guiada pelo professor e os desafios gamificados, outras possibilidades podem ser exploradas. Por exemplo, a escala pode ser utilizada como instrumento de autoavaliação, em que o próprio estudante marca os itens que já domina e identifica aqueles que ainda precisa estudar. Também pode servir como base para rodas de conversa sobre estratégias de resolução de problemas, promovendo o desenvolvimento metacognitivo. Em contextos de recuperação paralela, a escala pode orientar o planejamento de atividades específicas para grupos com dificuldades semelhantes, otimizando o tempo e os recursos disponíveis.

Independentemente da estratégia adotada, o mais importante é que o professor esteja atento ao potencial da escala como ferramenta de escuta pedagógica. Ela permite enxergar o estudante em movimento, reconhecer seus avanços e oferecer apoio onde ele mais precisa. Ao tornar visível o percurso de aprendizagem, a escala contribui para uma educação matemática mais justa, mais humana e mais eficaz.

Em tempo, cabe ressaltar que, apesar de a fonte do material usado nas práticas aqui sugeridas ser o Enem, esse fato não inviabiliza a sua utilização no Ensino Fundamental 2. De fato, já há uma gama muito grande de itens alinhados curricularmente a essa etapa de ensino.

Por fim, a seguir, disponibilizo 30 escalas, cada uma referente a uma habilidade do Enem para a área de Matemática e suas Tecnologias para você já usar. Essas escalas foram construídas com as aplicações regulares do Enem dos últimos 10 anos e retiradas da prova amarela. Desejo a você, caro professor, um excelente trabalho com seus alunos. Que este material possa inspirar práticas potentes, fortalecer vínculos e transformar a sala de aula em um espaço de descoberta, crescimento e protagonismo.

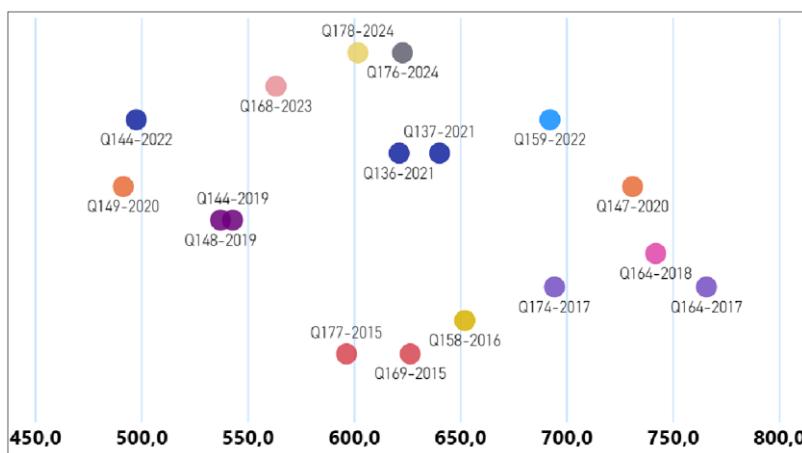
Tabelas e Escalas de Proficiência Simplificadas

A seguir, estão, por habilidade, as tabelas e as escalas de proficiência simplificadas dos itens da Área de Matemática e suas Tecnologias, referentes aos cadernos amarelos da aplicação regular do Enem, dos últimos 10 anos.

As tabelas foram organizadas em ordem crescente de dificuldade (parâmetros b) de forma que, quanto maior é esse parâmetro, mais difícil é o item. Assim, nas tabelas, os itens aumentam de dificuldade de cima para baixo. Já na escala de proficiência simplificada, a dificuldade dos itens aumenta de baixo para cima.

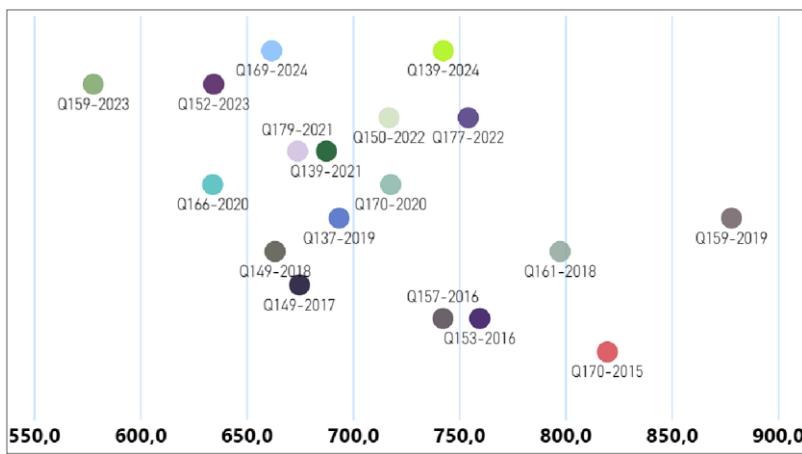
H1: Reconhecer, no contexto social, diferentes significados e representações dos números e operações – naturais, inteiros, racionais ou reais.

Ano	Item	Gabarito	Dificuldade
2020	149	A	491,4
2022	144	E	497,6
2019	148	A	537,2
2019	144	D	542,9
2023	168	C	563,3
2015	177	E	596,4
2024	178	B	601,8
2021	136	D	621,3
2024	176	D	622,8
2015	169	C	626,5
2021	137	D	640,3
2016	158	D	652,1
2022	159	C	692,2
2017	174	E	694,3
2020	147	C	731,1
2018	164	A	741,9
2017	164	C	765,8



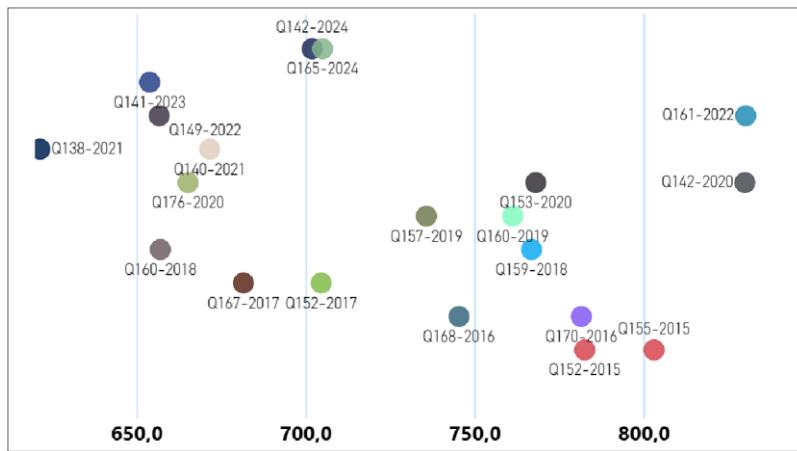
H2: Identificar padrões numéricos ou princípios de contagem.

Ano	Item	Gabarito	Dificuldade
2023	159	D	577,8
2020	166	B	634,0
2023	152	E	634,5
2024	169	A	661,7
2018	149	E	663,3
2021	179	E	673,9
2017	149	B	674,8
2021	139	A	687,5
2019	137	E	693,4
2022	150	B	716,9
2020	170	E	717,7
2016	157	A	742,3
2024	139	E	742,5
2022	177	B	754,2
2016	153	E	759,6
2018	161	C	797,5
2015	170	A	819,6
2019	159	E	878,1



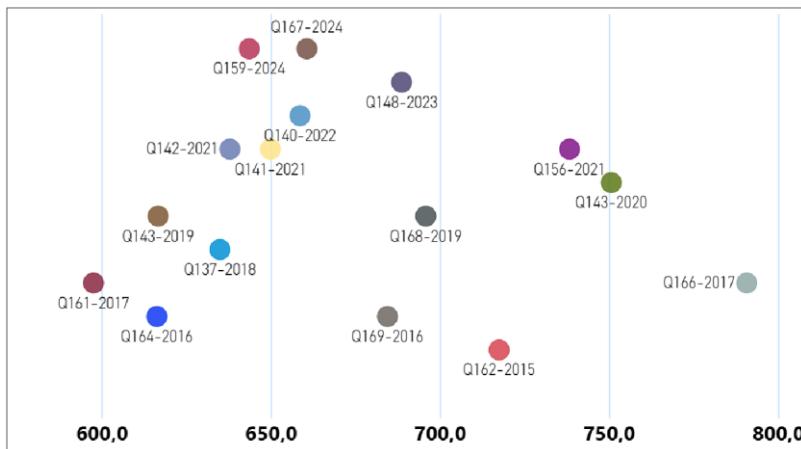
H3: Resolver situação-problema envolvendo conhecimentos numéricos.

Ano	Item	Gabarito	Dificuldade
2021	138	C	621,5
2023	141	C	653,9
2022	149	C	656,7
2018	160	C	657,0
2020	176	C	665,2
2021	140	B	671,6
2017	167	E	681,6
2024	142	D	701,9
2017	152	B	704,6
2024	165	B	705,0
2019	157	E	735,7
2016	168	A	745,3
2019	160	C	761,2
2018	159	C	766,8
2020	153	A	768,0
2016	170	D	781,5
2015	152	D	782,5
2015	155	E	803,0
2020	142	C	829,9
2022	161	B	830,1



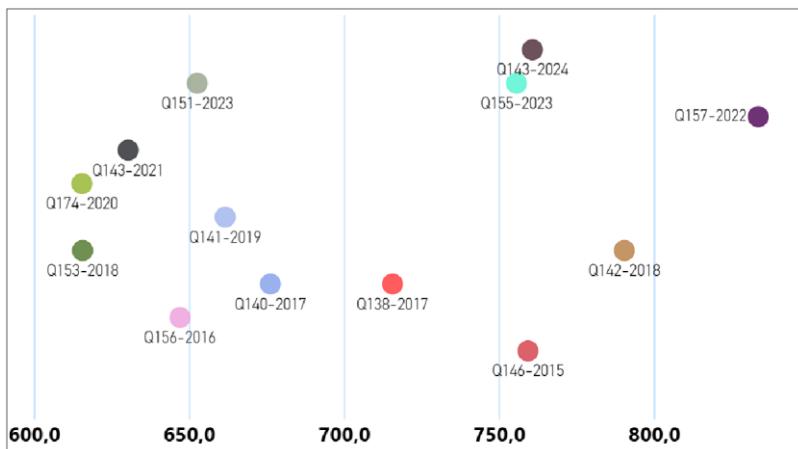
H4: Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.

Ano	Item	Gabarito	Dificuldade
2017	161	C	597,6
2016	164	A	616,3
2019	143	A	616,7
2018	137	D	634,9
2021	142	C	637,9
2024	159	C	643,6
2021	141	D	649,9
2022	140	A	658,6
2024	167	B	660,6
2016	169	B	684,5
2023	148	E	688,7
2019	168	C	695,8
2015	162	C	717,5
2021	156	E	738,3
2020	143	C	750,6
2017	166	D	790,7



H5: Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.

Ano	Item	Gabarito	Dificuldade
2020	174	B	615,4
2018	153	B	615,6
2021	143	C	630,3
2016	156	C	647,1
2023	151	B	652,6
2019	141	E	661,6
2017	140	E	676,2
2017	138	E	715,6
2023	155	D	755,6
2015	146	E	759,3
2024	143	A	760,8
2018	142	B	790,4
2022	157	B	833,6



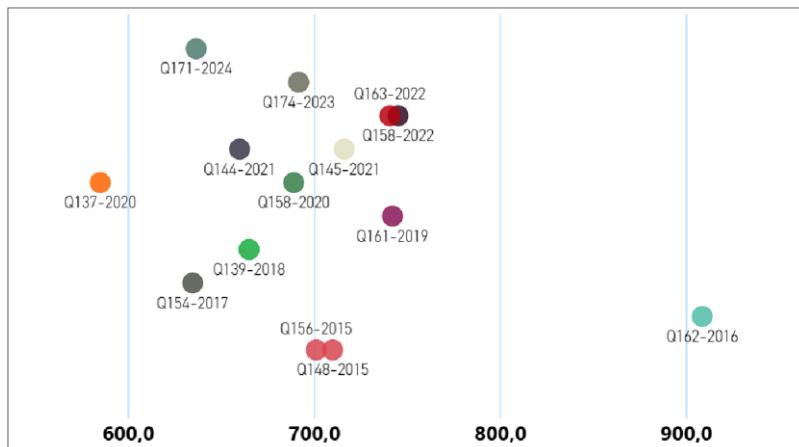
H6: Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.

Ano	Item	Gabarito	Dificuldade
2016	178	C	459,2
2020	159	B	459,6
2023	179	E	471,1
2017	158	A	513,6
2019	139	E	588,6
2022	141	C	594,7
2018	155	E	650,3
2024	162	C	663,0
2022	165	E	703,6
2022	178	A	743,2
2016	172	E	750,9
2015	168	E	754,3
2023	172	A	757,7
2021	180	E	978,3



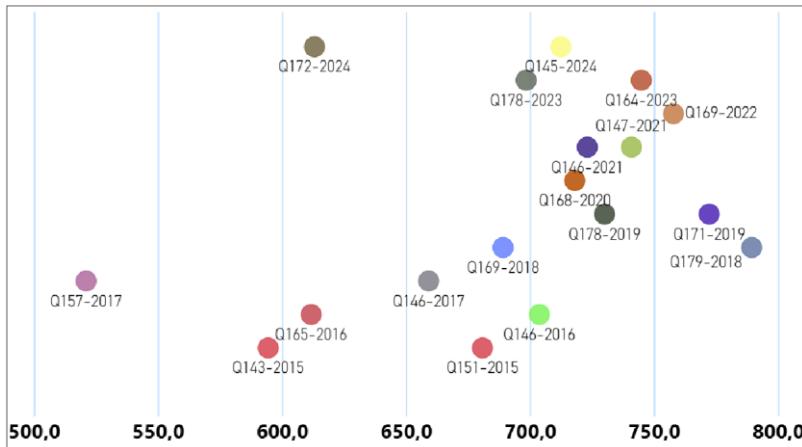
H7: Identificar características de figuras planas ou espaciais.

Ano	Item	Gabarito	Dificuldade
2020	137	E	585,0
2017	154	E	634,7
2024	171	D	636,5
2021	144	A	659,9
2018	139	E	665,0
2020	158	C	689,0
2023	174	B	691,6
2015	156	C	701,1
2015	148	E	710,0
2021	145	E	716,2
2022	163	C	740,7
2019	161	A	742,1
2022	158	D	745,1
2016	162	E	908,7



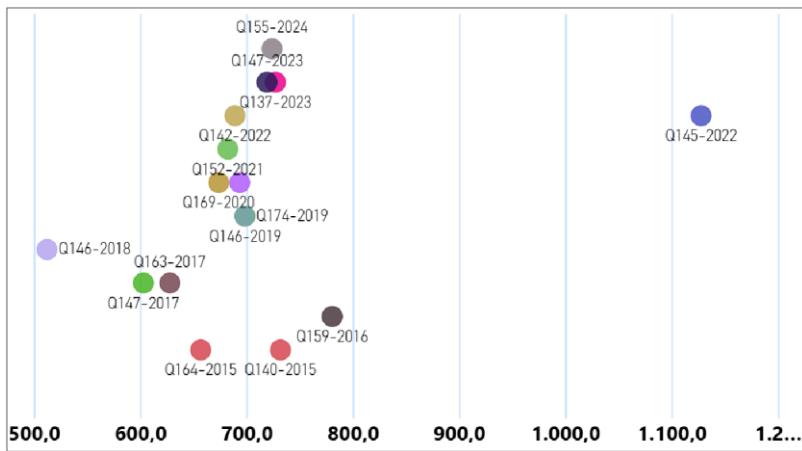
H8: Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma.

Ano	Item	Gabarito	Dificuldade
2017	157	D	521,0
2015	143	A	594,3
2016	165	C	611,7
2024	172	D	613,0
2017	146	B	659,1
2015	151	A	680,8
2018	169	A	689,1
2023	178	E	698,4
2016	146	D	703,8
2024	145	C	712,4
2020	168	B	718,0
2021	146	D	723,1
2019	178	D	730,0
2021	147	C	740,9
2023	164	B	744,8
2022	169	D	757,8
2019	171	D	772,2
2018	179	D	789,3



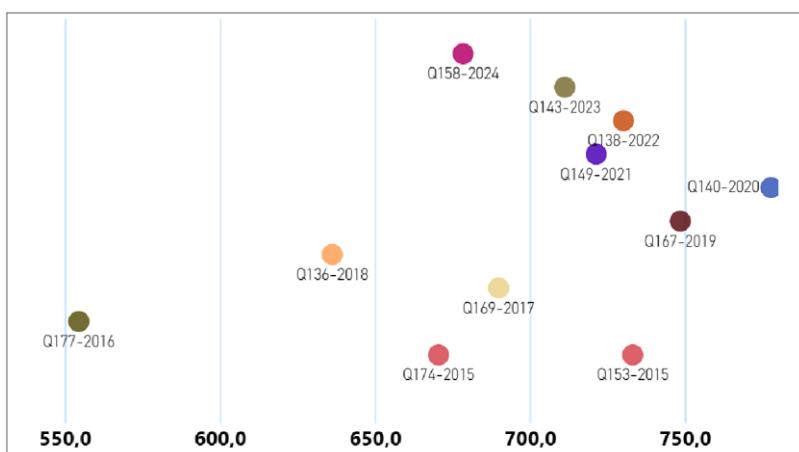
H9: Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos como solução de problemas do cotidiano.

Ano	Item	Gabarito	Dificuldade
2018	146	A	512,2
2017	147	B	602,9
2017	163	D	627,7
2015	164	C	656,7
2020	169	E	673,6
2021	152	B	682,2
2022	142	E	689,0
2020	178	D	693,5
2019	174	A	698,1
2019	146	E	698,7
2023	147	B	719,1
2024	155	E	723,7
2023	137	C	727,0
2015	140	A	731,7
2016	159	B	780,4
2022	145	D	1127,5



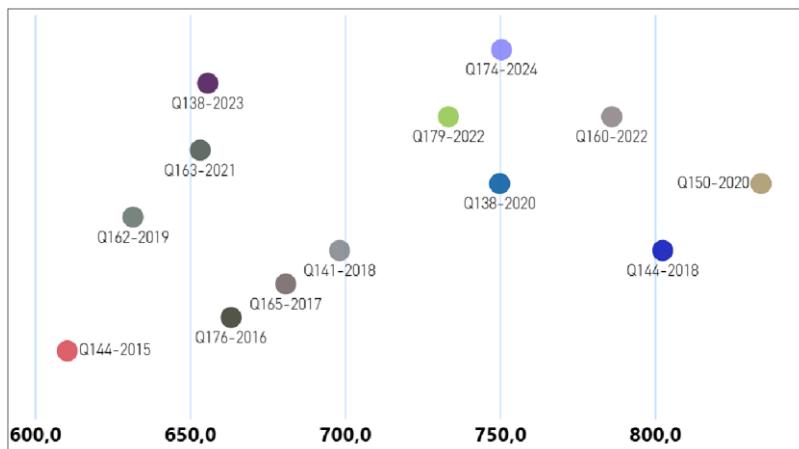
H10: Identificar relações entre grandezas e unidades de medida.

Ano	Item	Gabarito	Dificuldade
2016	177	D	554,4
2018	136	A	636,1
2015	174	E	670,5
2024	158	C	678,4
2017	169	B	689,8
2023	143	B	711,2
2021	149	B	721,3
2022	138	A	730,1
2015	153	C	733,1
2019	167	D	748,4
2020	140	C	777,7



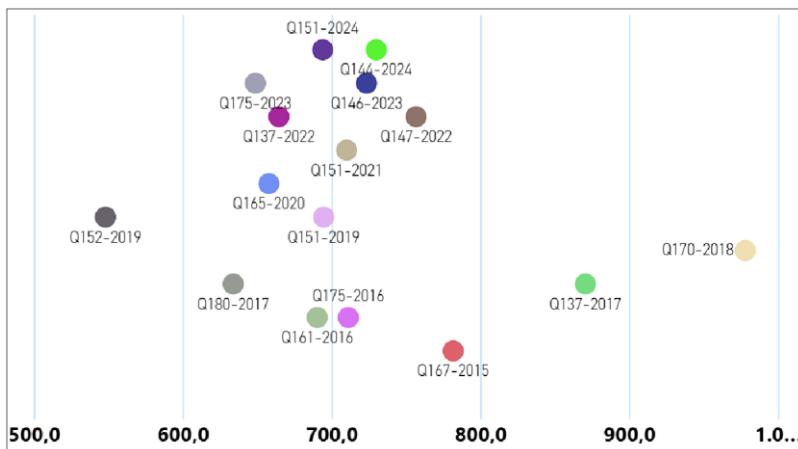
H11: Utilizar a noção de escalas na leitura de representação de situação do cotidiano.

Ano	Item	Gabarito	Dificuldade
2015	144	D	610,4
2019	162	C	631,6
2021	163	C	653,2
2023	138	E	655,7
2016	176	A	663,2
2017	165	C	680,8
2018	141	A	698,2
2022	179	A	733,3
2020	138	B	749,9
2024	174	E	750,4
2022	160	A	786,0
2018	144	C	802,4
2020	150	B	834,2



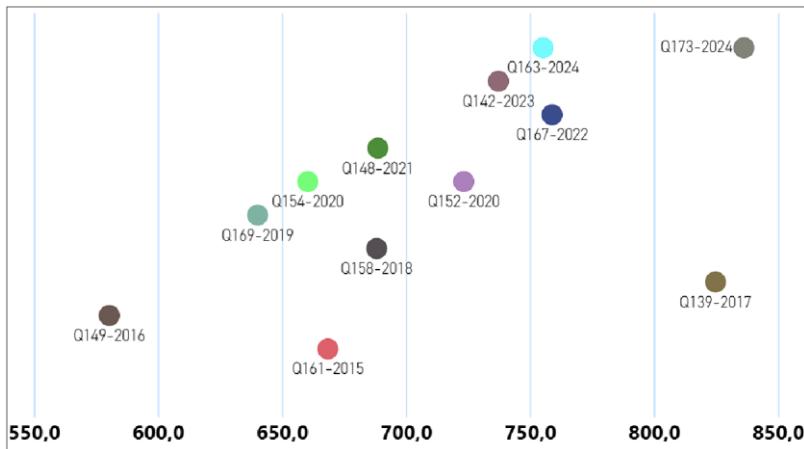
H12: Resolver situação-problema que envolva medidas de grandezas.

Ano	Item	Gabarito	Dificuldade
2019	152	C	547,9
2017	180	C	633,8
2023	175	D	648,7
2020	165	B	657,7
2022	137	B	664,5
2016	161	C	690,2
2024	151	D	693,9
2019	151	B	694,5
2021	151	D	710,0
2016	175	D	711,1
2023	146	C	723,4
2024	144	D	729,8
2022	147	E	756,7
2015	167	D	781,6
2017	137	C	870,5
2018	170	B	977,9



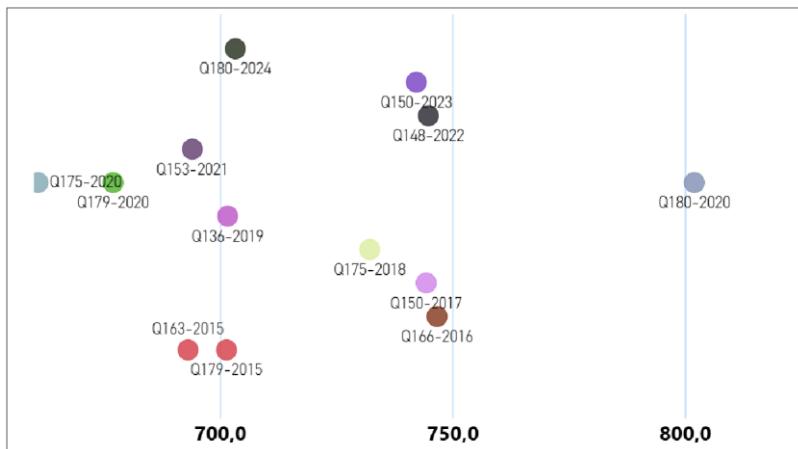
H13: Avaliar o resultado de uma medição na construção de um argumento consistente.

Ano	Item	Gabarito	Dificuldade
2016	149	B	580,3
2019	169	C	640,1
2020	154	B	660,3
2015	161	A	668,4
2018	158	D	688,1
2021	148	E	688,6
2020	152	C	723,3
2023	142	C	737,2
2024	163	A	755,3
2022	167	B	758,9
2017	139	C	824,7
2024	173	C	836,2



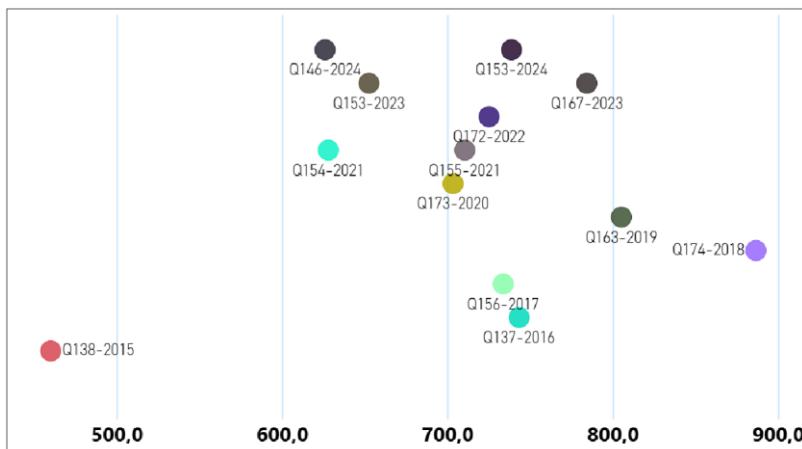
H14: Avaliar proposta de intervenção na realidade utilizando conhecimentos geométricos relacionados a grandezas e medidas.

Ano	Item	Gabarito	Dificuldade
2020	175	A	660,8
2020	179	A	677,0
2015	163	C	693,1
2021	153	D	694,0
2015	179	C	701,4
2019	136	B	701,6
2024	180	E	703,3
2018	175	A	732,2
2023	150	A	742,2
2017	150	B	744,3
2022	148	C	744,8
2016	166	B	746,7
2020	180	C	802,0



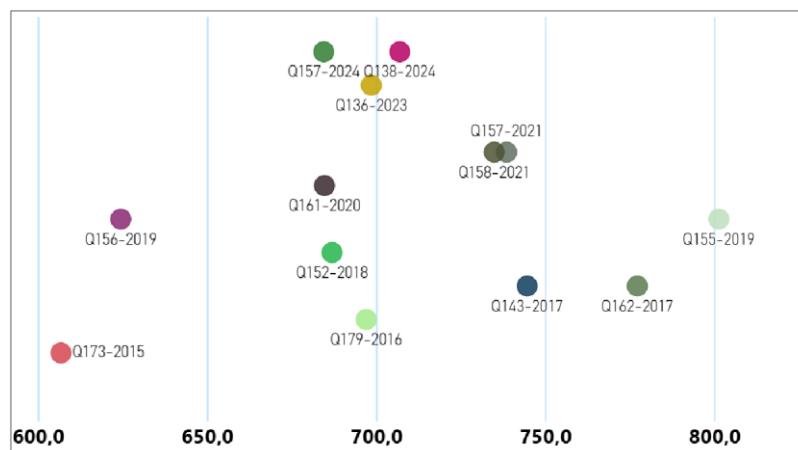
15: Identificar a relação de dependência entre grandezas

Ano	Item	Gabarito	Dificuldade
2015	138	B	460,0
2024	146	A	625,9
2021	154	A	627,9
2023	153	A	652,4
2020	173	B	703,4
2021	155	E	710,5
2022	172	D	725,2
2017	156	D	733,7
2024	153	B	738,7
2016	137	B	743,3
2023	167	B	784,3
2019	163	A	805,2
2018	174	A	886,5



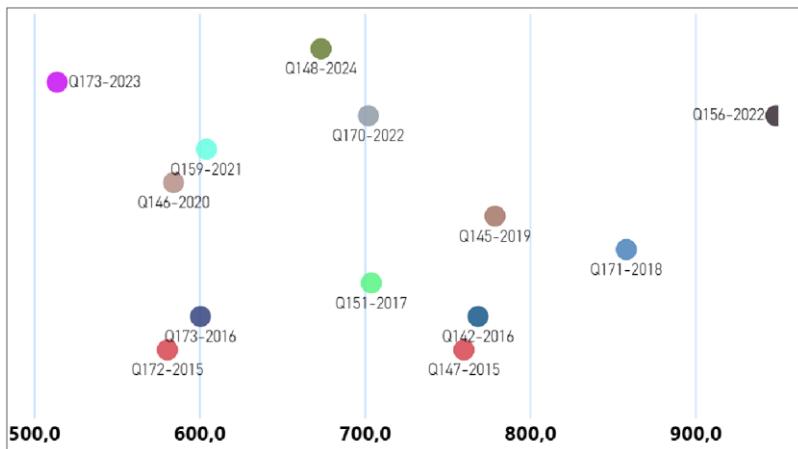
H16: Resolver situação-problema envolvendo a variação de grandezas, direta ou inversamente proporcionais.

Ano	Item	Gabarito	Dificuldade
2015	173	B	606,7
2019	156	B	624,4
2024	157	D	684,5
2020	161	E	684,7
2018	152	B	686,9
2016	179	E	697,0
2023	136	D	698,5
2024	138	C	706,9
2021	158	A	734,9
2021	157	B	738,5
2017	143	D	744,6
2017	162	B	777,2
2019	155	A	801,3



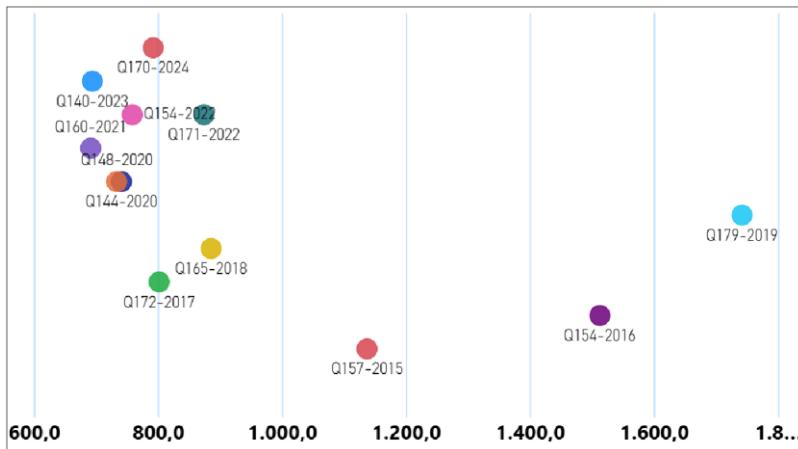
H17: Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação.

Ano	Item	Gabarito	Dificuldade
2023	173	E	513,9
2015	172	D	580,6
2020	146	C	584,3
2016	173	B	600,6
2021	159	D	604,2
2024	148	D	673,5
2022	170	C	702,1
2017	151	D	703,9
2015	147	A	760,0
2016	142	E	768,5
2019	145	B	778,8
2018	171	C	858,2
2022	156	C	948,7



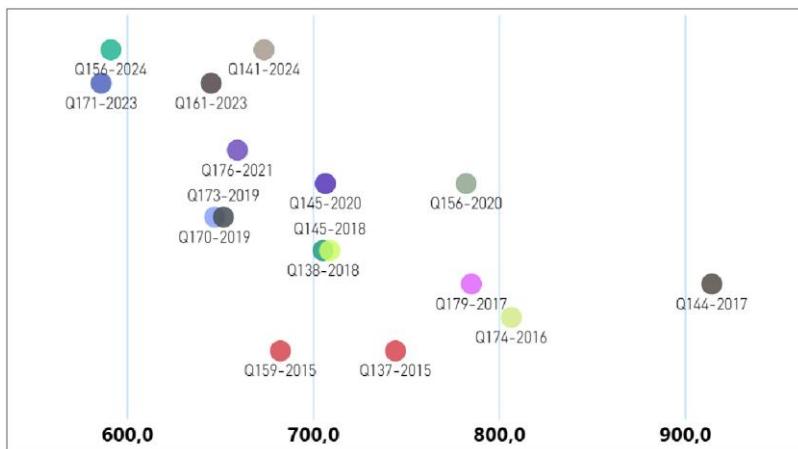
H18: Avaliar propostas de intervenção na realidade envolvendo variação de grandezas.

Ano	Item	Gabarito	Dificuldade
2021	160	C	691,3
2023	140	A	693,9
2020	148	D	733,1
2020	144	C	740,8
2022	154	E	758,4
2024	170	D	792,0
2017	172	C	801,6
2022	171	C	873,8
2018	165	C	885,4
2015	157	A	1136,8
2016	154	A	1512,4
2019	179	D	1741,7



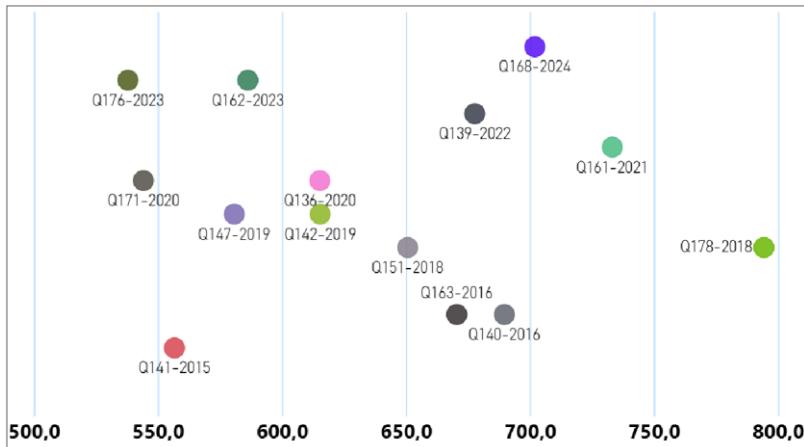
H19: Identificar representações algébricas que expressem a relação entre grandezas.

Ano	Item	Gabarito	Dificuldade
2023	171	D	586,0
2024	156	B	591,2
2023	161	A	645,1
2019	170	E	647,0
2019	173	D	651,8
2021	176	A	659,3
2024	141	E	673,6
2015	159	E	682,4
2018	138	A	705,3
2020	145	A	706,6
2018	145	A	708,9
2015	137	B	744,2
2020	156	A	782,3
2017	179	A	785,1
2016	174	C	806,6
2017	144	A	914,3



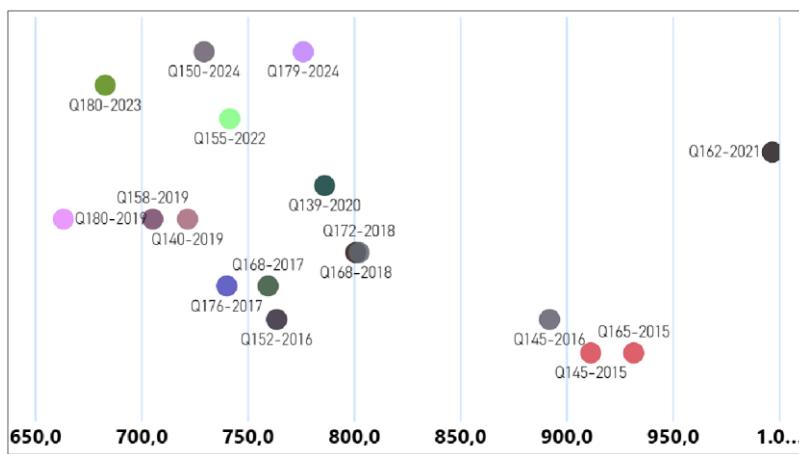
H20: Interpretar gráfico cartesiano que represente relações entre grandezas.

Ano	Item	Gabarito	Dificuldade
2023	176	A	537,8
2020	171	E	544,1
2015	141	D	556,6
2019	147	B	580,7
2023	162	B	586,2
2020	136	A	615,2
2019	142	A	615,3
2018	151	E	650,6
2016	163	B	670,4
2022	139	A	677,7
2016	140	C	689,6
2024	168	B	701,9
2021	161	A	733,2
2018	178	B	794,2



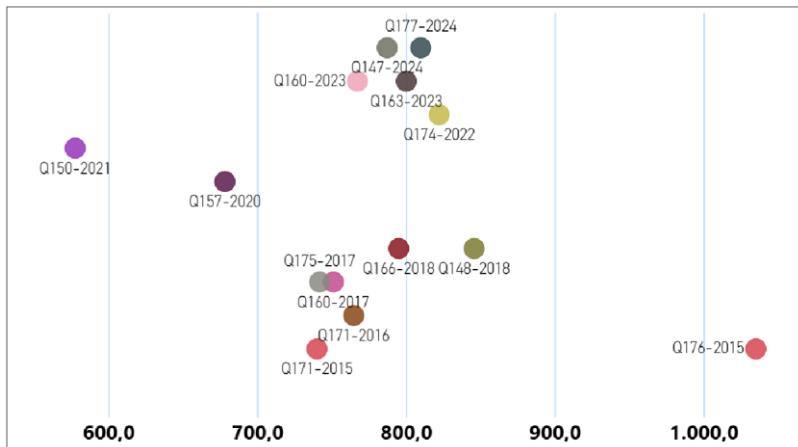
H21: Resolver situação-problema cuja modelagem envolva conhecimentos algébricos.

Ano	Item	Gabarito	Dificuldade
2019	180	D	663,3
2023	180	B	683,0
2019	158	C	705,3
2019	140	E	721,7
2024	150	C	729,4
2017	176	D	740,1
2022	155	E	741,5
2017	168	D	759,6
2016	152	C	763,6
2024	179	C	776,1
2020	139	C	786,2
2018	168	D	800,8
2018	172	B	802,4
2016	145	D	892,0
2015	145	A	911,4
2015	165	E	931,6
2021	162	E	996,8



H22: Utilizar conhecimentos algébricos/geométricos como recurso para a construção de argumentação.

Ano	Item	Gabarito	Dificuldade
2021	150	B	577,6
2020	157	B	678,2
2015	171	B	740,1
2017	175	A	741,8
2017	160	E	751,1
2016	171	B	764,7
2023	160	D	767,2
2024	147	A	787,2
2018	166	E	795,0
2023	163	B	800,1
2024	177	B	809,9
2022	174	D	822,2
2018	148	B	845,7
2015	176	D	1035,0



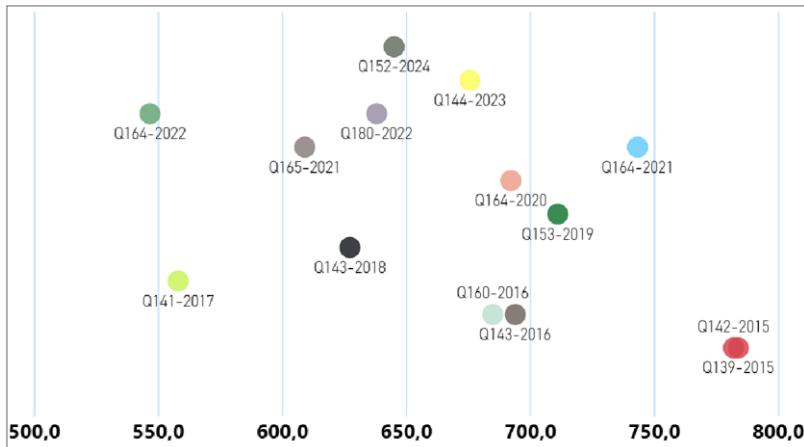
23: Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.

Ano	Item	Gabarito	Dificuldade
2015	154	B	606,3
2024	160	C	611,7
2023	156	D	627,7
2019	154	B	653,8
2020	177	C	710,9
2016	155	E	711,3
2022	176	C	737,3
2018	177	B	831,9
2015	136	D	851,9
2017	145	D	965,6



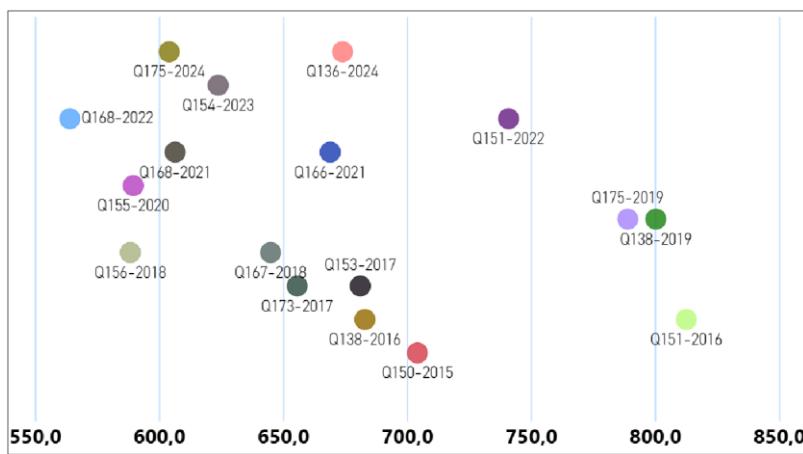
H24: Utilizar informações expressas em gráficos ou tabelas para fazer inferências.

Ano	Item	Gabarito	Dificuldade
2022	164	B	546,7
2017	141	E	558,1
2021	165	B	609,1
2018	143	B	627,3
2022	180	C	638,1
2024	152	B	645,1
2023	144	D	675,7
2016	160	E	684,9
2020	164	D	692,3
2016	143	A	694,0
2019	153	D	711,1
2021	164	D	743,3
2015	142	C	782,0
2015	139	B	783,9



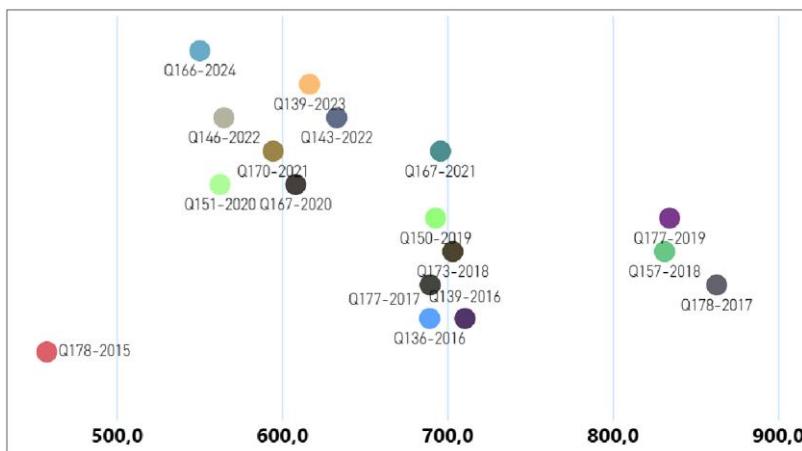
H25: Resolver problema com dados apresentados em tabelas ou gráficos.

Ano	Item	Gabarito	Dificuldade
2022	168	D	564,0
2018	156	E	588,4
2020	155	C	589,6
2024	175	A	604,0
2021	168	C	606,5
2023	154	B	623,8
2018	167	D	644,9
2017	173	A	655,6
2021	166	C	669,0
2024	136	C	673,9
2017	153	A	681,1
2016	138	A	682,9
2015	150	C	704,2
2022	151	E	740,9
2019	175	B	788,9
2019	138	D	800,3
2016	151	D	812,6



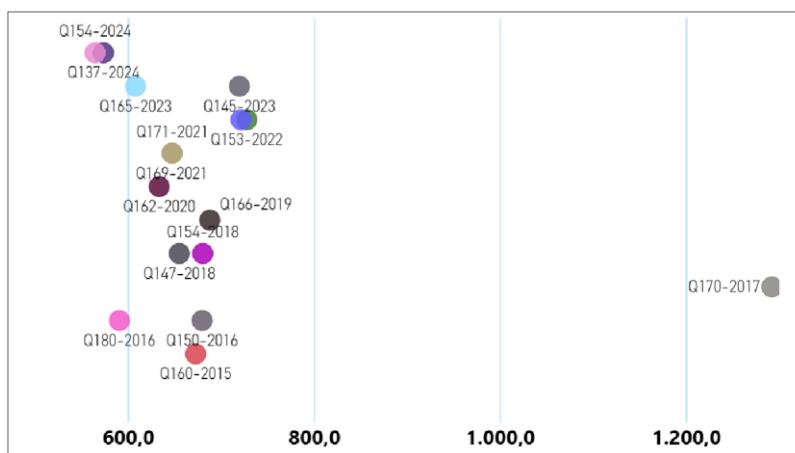
H26: Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.

Ano	Item	Gabarito	Dificuldade
2015	178	B	457,7
2024	166	A	550,2
2020	151	C	562,4
2022	146	A	564,8
2021	170	D	594,5
2020	167	D	608,2
2023	139	C	616,5
2022	143	B	633,0
2016	136	C	689,3
2017	177	A	689,6
2019	150	A	692,9
2021	167	C	695,8
2018	173	C	703,3
2016	139	A	710,6
2018	157	A	831,3
2019	177	C	834,3
2017	178	E	862,8



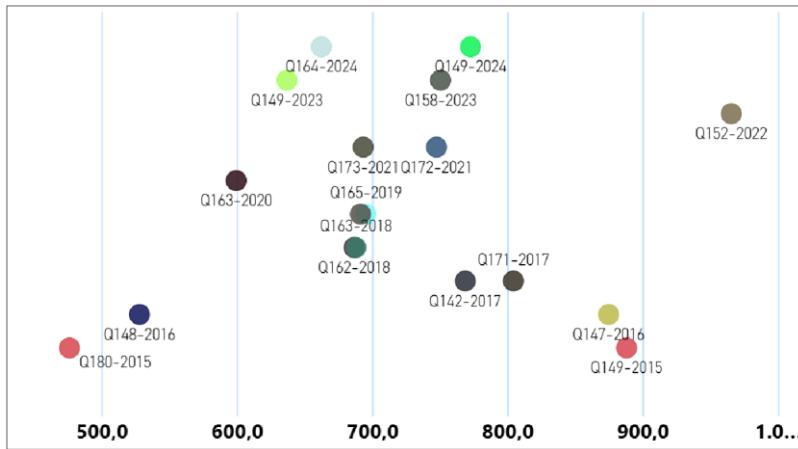
H27: Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos.

Ano	Item	Gabarito	Dificuldade
2024	154	C	564,5
2024	137	E	573,5
2016	180	E	590,6
2023	165	C	608,0
2020	162	C	633,4
2021	169	C	646,8
2021	171	D	647,5
2018	147	D	655,0
2015	160	D	672,6
2016	150	D	679,6
2018	154	D	680,2
2019	166	B	687,9
2023	145	E	719,6
2022	173	B	721,4
2022	153	D	727,6
2017	170	B	1292,1



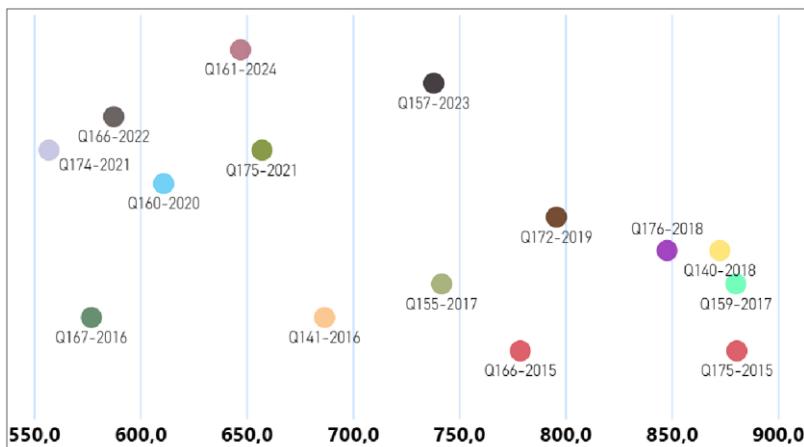
H28: Resolver situação-problema que envolva conhecimentos de estatística e probabilidade.

Ano	Item	Gabarito	Dificuldade
2015	180	C	476,1
2016	148	D	527,9
2020	163	E	599,4
2023	149	E	636,8
2024	164	B	662,3
2018	162	C	686,5
2018	163	D	687,8
2019	176	E	691,2
2021	173	E	693,2
2019	165	C	695,3
2021	172	C	747,5
2023	158	A	750,5
2017	142	C	768,7
2024	149	E	772,7
2017	171	A	804,3
2016	147	C	874,7
2015	149	D	887,9
2022	152	C	965,3



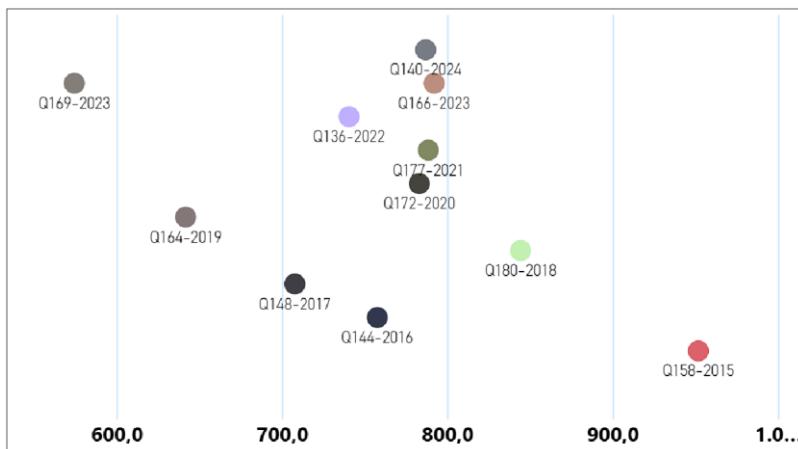
H29: Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação.

Ano	Item	Gabarito	Dificuldade
2021	174	E	557,0
2016	167	D	576,9
2022	166	A	587,5
2020	160	A	610,9
2024	161	A	647,1
2021	175	B	657,2
2016	141	C	686,6
2023	157	A	738,0
2017	155	B	741,7
2015	166	B	778,7
2019	172	C	795,7
2018	176	E	847,8
2018	140	D	872,6
2017	159	B	880,1
2015	175	E	880,6



H30: Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade.

Ano	Item	Gabarito	Dificuldade
2023	169	C	574,3
2019	164	D	641,7
2017	148	D	707,6
2022	136	D	740,4
2016	144	B	757,5
2020	172	D	783,0
2024	140	B	786,9
2021	177	B	788,5
2023	166	C	791,9
2018	180	D	844,3
2015	158	A	951,7



Referências bibliográficas

PASQUALI, L. **Psicometria: Teoria dos Testes na Psicologia e na Educação.** 5. ed. Petrópolis: Vozes, 2003.

RABELO, M. **Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro.** Rio de Janeiro: SBM, 2013.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Entenda a sua nota no Enem: guia do participante.** Brasília, DF: INEP, 2021a.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Exame Nacional do Ensino Médio – Enem: procedimentos de análise.** Brasília, DF: INEP, 2021b.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Matriz de Referência do ENEM.** Brasília: INEP, [s.d.]. Disponível em: https://download.inep.gov.br/download/enem/matriz_referencia.pdf. Acesso em: 30 ago. 2025.

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Microdados do Enem. Brasília: Inep, [2020?]. Disponível em:

<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 30 ago. 2025.

APÊNDICE B – Estatísticas descritivas do Enem 2023

Os relatórios estatísticos apresentados a seguir referem-se à edição do Enem 2023, em todas as suas aplicações, considerando exclusivamente os participantes que:

- estavam cursando o ensino médio regular no ano da realização do exame, com previsão de conclusão ainda em 2023;
- estiveram presentes no dia da prova de Matemática e suas Tecnologias;
- possuíam o campo “Sigla da Unidade da Federação da escola” devidamente preenchido;
- não tiveram suas provas anuladas ou zeradas.

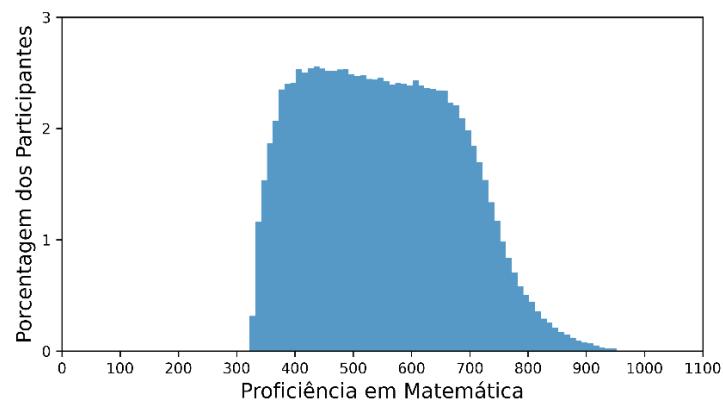
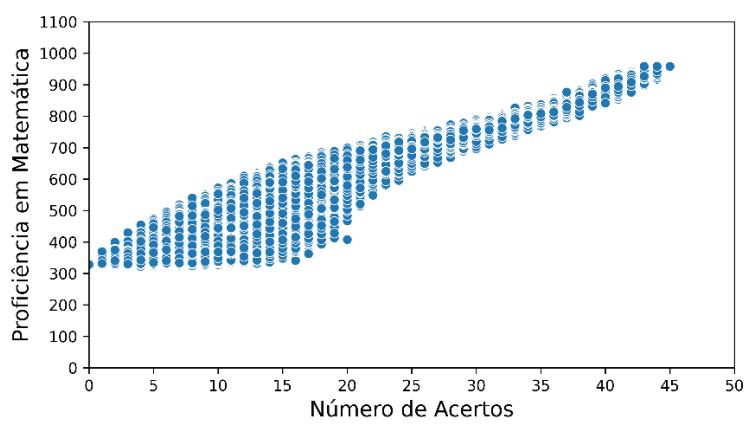
Com esse recorte, foi obtida uma amostra de 705.646 participantes distribuídos nas seguintes redes de ensino:

- Municipal: 5492 estudantes;
- Estadual: 443465 estudantes;
- Federal: 43458 estudantes;
- Privada: 213231 estudantes.

É importante destacar que, devido à metodologia da Teoria de Resposta ao Item (TRI) utilizada pelo Enem, a proficiência em cada área do conhecimento nunca será igual a zero. Isso só ocorre em casos específicos, como quando o estudante realiza marcações duplas no cartão-resposta, deixa questões em branco ou preenche o cartão de forma ilegível. Mesmo que o participante erre todos os itens de uma determinada área, ainda assim não será atribuída proficiência zero, justamente por conta das características estatísticas da TRI.

Além disso, o Enem utiliza uma escala padronizada, em que a média das proficiências é 500 e o desvio padrão é 100. Esse modelo evita confusões, como o aparecimento de notas negativas que podem ocorrer em escalas com média zero e desvio padrão igual a 1. A padronização adotada em todas as edições do exame permite comparar o desempenho dos participantes de forma consistente, independentemente do ano de aplicação.

Os dados utilizados foram extraídos dos microdados do Enem, disponibilizados pelo Inep no portal eletrônico <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados>.

Distribuição de frequência relativa da amostra em relação à proficiência.**Dispersão das proficiências em relação ao número de acertos**

**Distribuição de frequência absoluta
da amostra em relação ao número de acertos**

Acertos	Mínima	Máxima	Média	Desvio Padrão	Frequência
0	328,20	328,20	328,20	0,00	32
1	328,20	369,60	335,71	11,36	92
2	328,20	399,10	341,31	14,84	378
3	328,20	429,80	348,92	19,34	1422
4	322,10	455,00	355,22	22,57	3804
5	324,90	473,70	365,90	26,97	8283
6	326,90	497,10	376,45	31,40	15029
7	326,30	518,90	388,27	34,99	23526
8	322,70	540,30	402,30	38,61	32593
9	327,60	549,40	417,81	42,92	40259
10	328,00	572,30	436,08	46,23	45749
11	332,70	588,90	456,36	48,82	48291
12	332,60	609,80	478,59	50,22	48432
13	332,20	620,10	501,74	49,78	46291
14	335,80	637,90	526,11	48,24	43152
15	347,50	651,50	549,66	44,95	39181
16	341,20	663,80	572,25	40,96	35572
17	362,50	672,70	593,38	36,11	31387
18	392,80	684,60	612,35	31,32	28116
19	413,40	690,60	629,25	27,20	25263
20	408,00	701,00	645,09	23,29	22939
21	514,40	710,20	659,15	20,68	20584
22	548,80	719,10	672,01	18,24	18406
23	578,10	734,90	683,79	16,70	16673
24	595,90	737,10	695,01	15,28	14883
25	624,20	745,20	705,61	14,03	13392
26	641,20	755,50	716,10	12,94	11829
27	653,30	758,90	726,09	12,24	10375
28	668,50	772,70	735,79	11,67	9214
29	689,40	777,90	745,73	11,10	7912
30	696,00	790,30	755,57	10,78	6864
31	711,10	797,90	765,57	10,36	5962
32	726,60	807,70	775,78	10,06	5088
33	737,30	826,80	786,22	9,90	4498
34	757,00	831,90	797,72	9,82	3836
35	767,20	844,50	808,66	9,80	3317
36	782,30	857,10	820,04	10,24	2821
37	794,70	876,40	832,57	10,19	2347
38	801,90	880,00	845,92	10,65	2098
39	828,60	902,50	859,84	11,09	1680
40	840,90	918,80	874,92	12,24	1384
41	854,10	935,80	891,87	13,03	1088
42	872,50	939,80	908,69	13,20	830
43	891,90	958,60	929,45	13,75	456
44	917,80	958,60	947,03	11,28	239
45	958,60	958,60	958,60	0,00	79

Coerência pedagógica

Nos relatórios a seguir, cada linha representa o desempenho de um participante na prova de Matemática, com foco na coerência pedagógica das respostas. As colunas indicam:

- **ACERTOS:** total de questões respondidas corretamente;
 - **NOTA:** proficiência obtida pelo participante;
 - **PADRÃO DE RESPOSTA:** sequência binária que representa erros (0) e acertos (1) em cada item da prova;
 - **COERÊNCIA:** grau de alinhamento pedagógico das respostas (quanto mais acertos nos itens mais fáceis, maior a coerência);
 - **INICIAIS:** número de acertos nas primeiras 20 questões (mais fáceis);
 - **FINAIS:** número de acertos nas últimas 25 questões (mais difíceis).

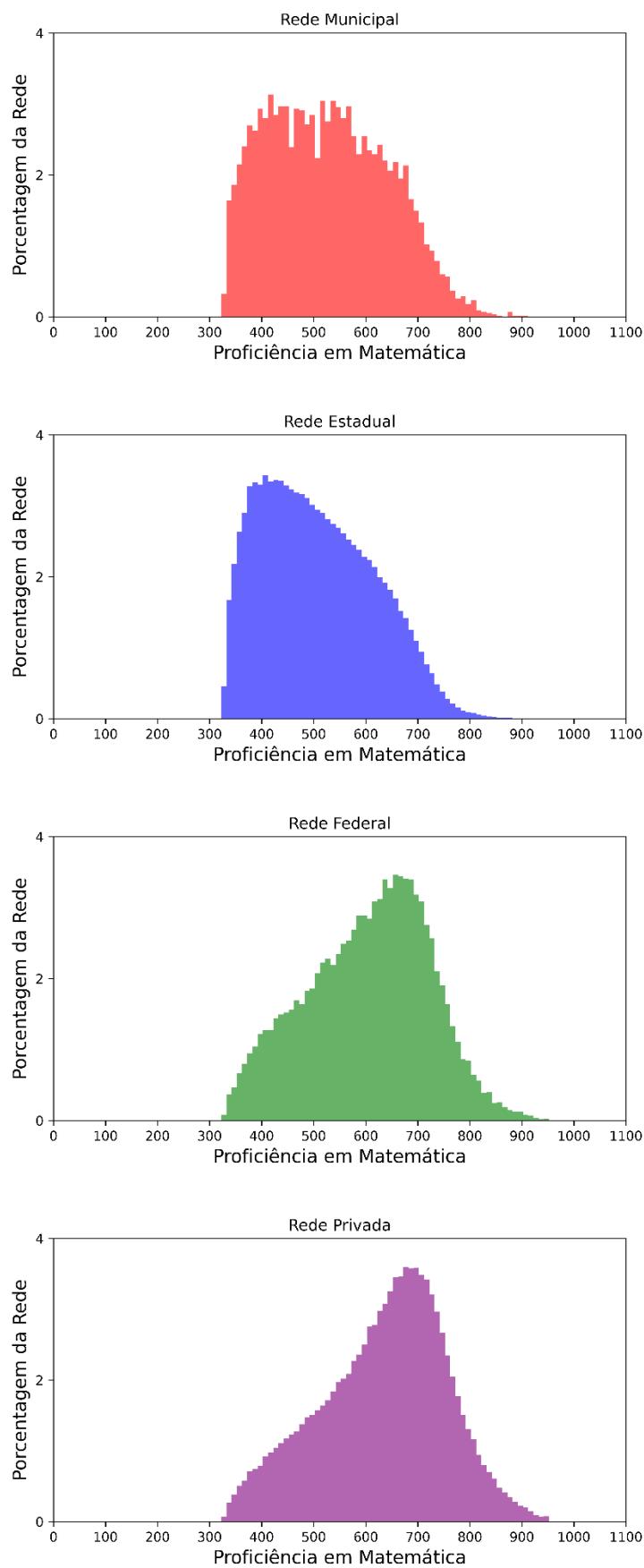
Além disso, para gerar esse relatório, os itens das provas foram reorganizados em ordem crescente de dificuldade, ou seja. Assim, na coluna PADRÃO DE RESPOSTA, os primeiros itens são fáceis e os últimos mais dificeis.

- Melhor e pior coerência pedagógica do maior desvio-padrão em relação à proficiência

- Melhor e pior coerência pedagógica por 20 acertos

ACERTOS	NOTA	PADRÃO DE RESPOSTA	COERENCIA	INICIAIS	FINAIS
20	701	0111111111110110100010111100000000000000000000000000	Melhor	15	5
20	408	00010011011100000000100100011011010011111101	Pior	6	14

- Melhor e pior coerência pedagógica para a proficiência 500 (média)

Distribuição de FR da amostra em relação às Redes de Ensino

APÊNDICE C – Estatísticas descritivas do Enem 2024

Na edição do Enem de 2024, o Inep divulgou os microdados do Enem separando o banco de dados de participantes do banco de dados de resultados, de forma a não ser possível relacioná-los. Dessa forma, os relatórios estatísticos apresentados a seguir consideram os participantes que:

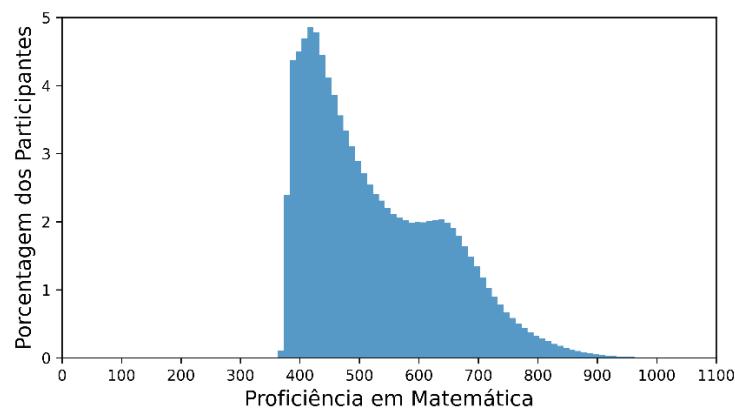
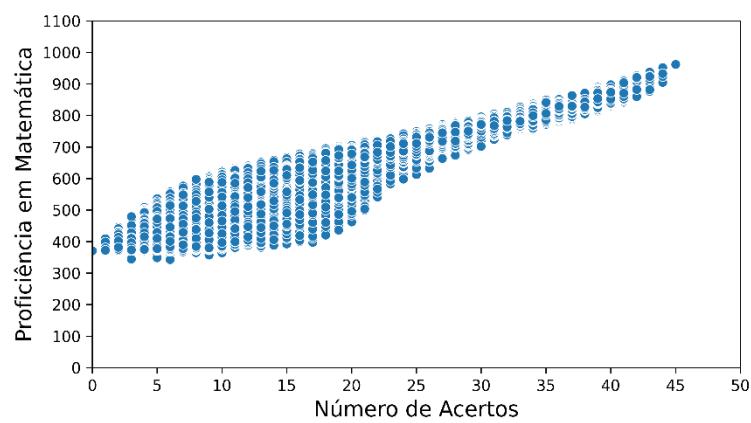
- possuíam o campo “Sigla da Unidade da Federação da escola” devidamente preenchido;
- não tiveram suas provas anuladas ou zeradas.

Com esse recorte, foi obtida uma amostra de 3.004.169 participantes.

É importante destacar que, devido à metodologia da Teoria de Resposta ao Item (TRI) utilizada pelo Enem, a proficiência em cada área do conhecimento nunca será igual a zero. Isso só ocorre em casos específicos, como quando o estudante realiza marcações duplas no cartão-resposta, deixa questões em branco ou preenche o cartão de forma ilegível. Mesmo que o participante erre todos os itens de uma determinada área, ainda assim não será atribuída proficiência zero, justamente por conta das características estatísticas da TRI.

Além disso, o Enem utiliza uma escala padronizada, em que a média das proficiências é 500 e o desvio padrão é 100. Esse modelo evita confusões, como o aparecimento de notas negativas que podem ocorrer em escalas com média zero e desvio padrão igual a 1. A padronização adotada em todas as edições do exame permite comparar o desempenho dos participantes de forma consistente, independentemente do ano de aplicação.

Os dados utilizados foram extraídos dos microdados do Enem, disponibilizados pelo Inep no portal eletrônico <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados>.

Distribuição de frequência relativa da amostra em relação à proficiência.**Dispersão das proficiências em relação ao número de acertos**

**Distribuição de frequência absoluta
da amostra em relação ao número de acertos.**

Acertos	Mínima	Máxima	Média	Desvio Padrão	Frequência
0	369,40	371,00	370,99	0,13	324
1	371,00	409,30	376,93	9,11	1244
2	369,60	445,00	382,41	12,85	5193
3	344,60	479,00	389,34	16,98	16607
4	371,30	509,10	396,50	20,81	42383
5	348,90	535,90	404,78	24,89	85913
6	342,80	557,40	414,25	28,97	143621
7	367,90	576,10	425,04	33,37	205284
8	365,20	597,50	437,49	38,09	255521
9	357,50	609,30	451,86	42,94	283543
10	364,60	623,10	468,90	47,69	281051
11	374,00	633,30	488,15	51,84	258159
12	379,10	643,60	510,34	54,60	223446
13	381,30	653,10	534,51	55,40	184936
14	386,00	662,30	559,81	53,35	151308
15	391,30	670,30	583,97	48,40	122336
16	397,90	678,10	606,23	41,66	101072
17	397,70	687,40	625,30	34,38	85294
18	419,30	694,80	641,41	27,71	72416
19	436,40	701,90	655,30	22,55	62139
20	460,50	708,90	667,15	19,04	53653
21	499,70	719,60	678,28	16,26	46836
22	541,10	725,40	688,41	14,73	41238
23	583,40	735,90	698,28	13,45	36380
24	598,10	743,20	707,87	12,47	31289
25	612,60	752,40	717,21	11,77	27825
26	632,40	757,80	726,44	11,30	24234
27	663,90	771,20	735,46	10,84	21103
28	673,80	778,50	744,54	10,62	18580
29	691,40	788,10	753,54	10,54	16556
30	702,10	795,90	762,84	10,21	14672
31	722,70	804,70	771,65	10,22	12699
32	737,50	814,40	780,77	10,30	11565
33	752,90	827,40	790,45	10,31	10295
34	756,60	836,50	799,94	10,25	9130
35	772,30	849,40	810,04	10,41	8239
36	781,10	858,30	820,27	10,67	7407
37	787,60	867,50	831,38	11,00	6382
38	805,00	877,00	842,65	11,43	5646
39	814,60	890,50	855,06	12,29	4929
40	830,70	901,20	868,37	12,89	4069
41	841,60	915,10	882,39	13,32	3458
42	857,20	927,70	899,06	13,52	2693
43	876,80	940,60	918,16	13,14	1901
44	904,90	951,80	938,64	11,13	1094
45	961,90	961,90	961,90	0,00	506

Coerência Pedagógica

Nos relatórios a seguir, cada linha representa o desempenho de um participante na prova de Matemática, com foco na coerência pedagógica das respostas. As colunas indicam:

- **ACERTOS:** total de questões respondidas corretamente;
- **NOTA:** proficiência obtida pelo participante;
- **PADRÃO DE RESPOSTA:** sequência binária que representa erros (0) e acertos (1) em cada item da prova;
- **COERÊNCIA:** grau de alinhamento pedagógico das respostas (quanto mais acertos nos itens mais fáceis, maior a coerência);
- **INICIAIS:** número de acertos nas primeiras 20 questões (mais fáceis);
- **FINAIS:** número de acertos nas últimas 25 questões (mais dificeis).

Além disso, para gerar esse relatório, os itens das provas foram reorganizados em ordem crescente de dificuldade, ou seja. Assim, na coluna PADRÃO DE RESPOSTA, os primeiros itens são fáceis e os últimos mais dificeis.

- Melhor e pior coerência pedagógica do maior desvio-padrão em relação à proficiência

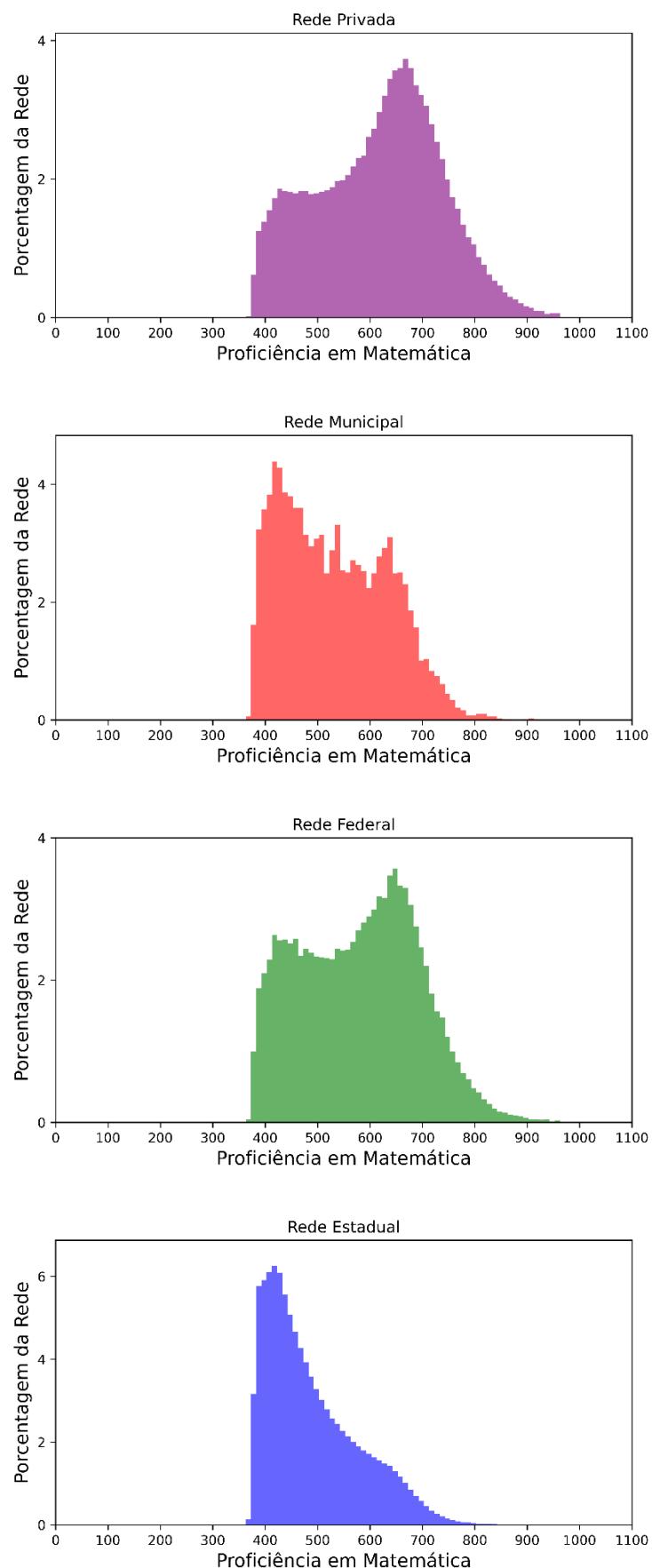
ACERTOS	NOTA	PADRÃO DE RESPOSTA	COERENCIA	INICIAIS	FINAIS
13	653,1	111111111100100000100000000000000000000000000000000	Melhor	13	0
13	381,3	000000000010010000001001001110110101000001100	Pior	2	11

- Melhor e pior coerência pedagógica por 20 acertos

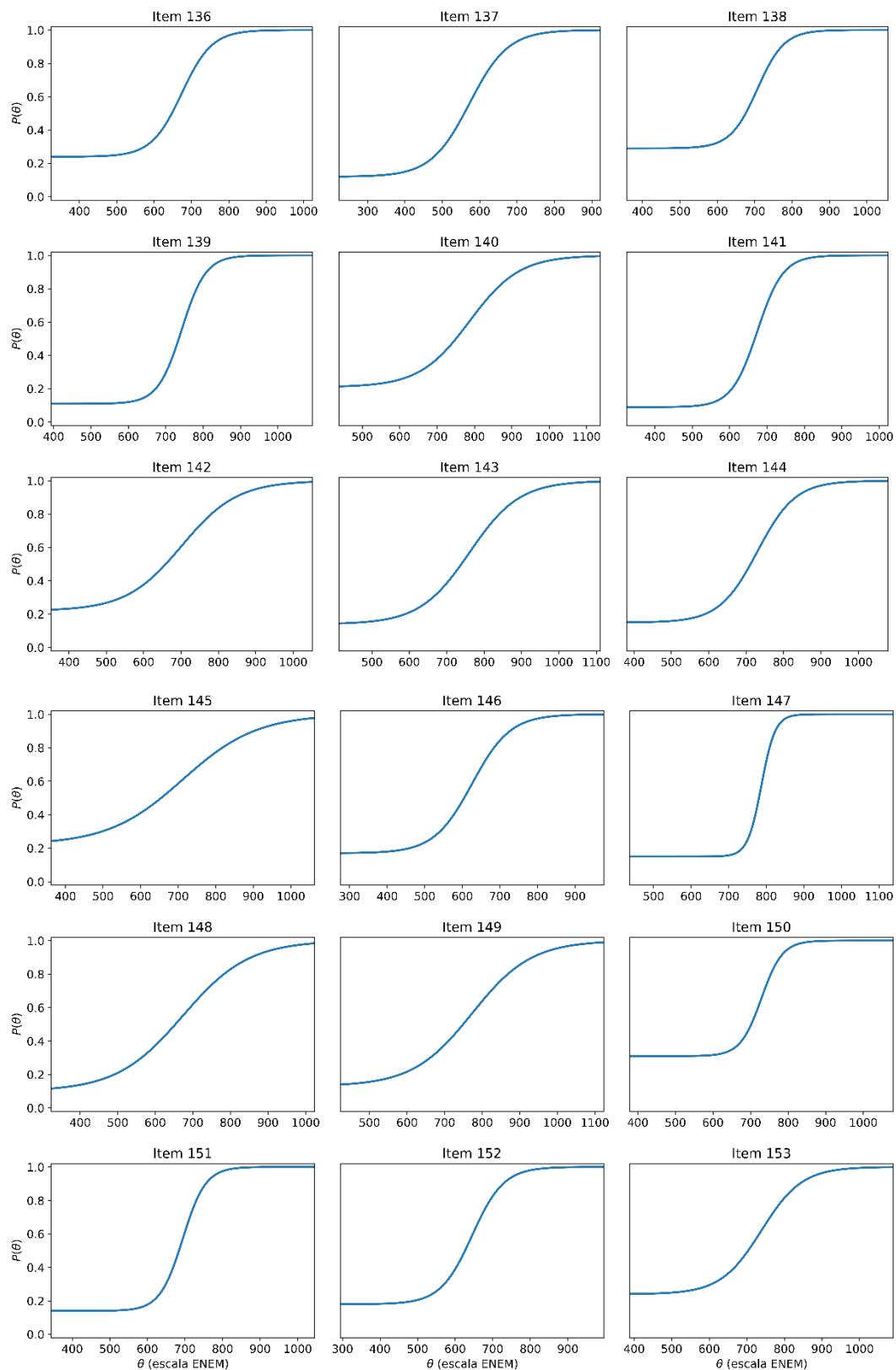
ACERTOS	NOTA	PADRÃO DE RESPOSTA	COERENCIA	INICIAIS	FINAIS
20	708,9	11110111011011101101100110000000000000000000000	Melhor	15	5
20	460,5	000000010000011100100110011001111010001111011	Pior	5	15

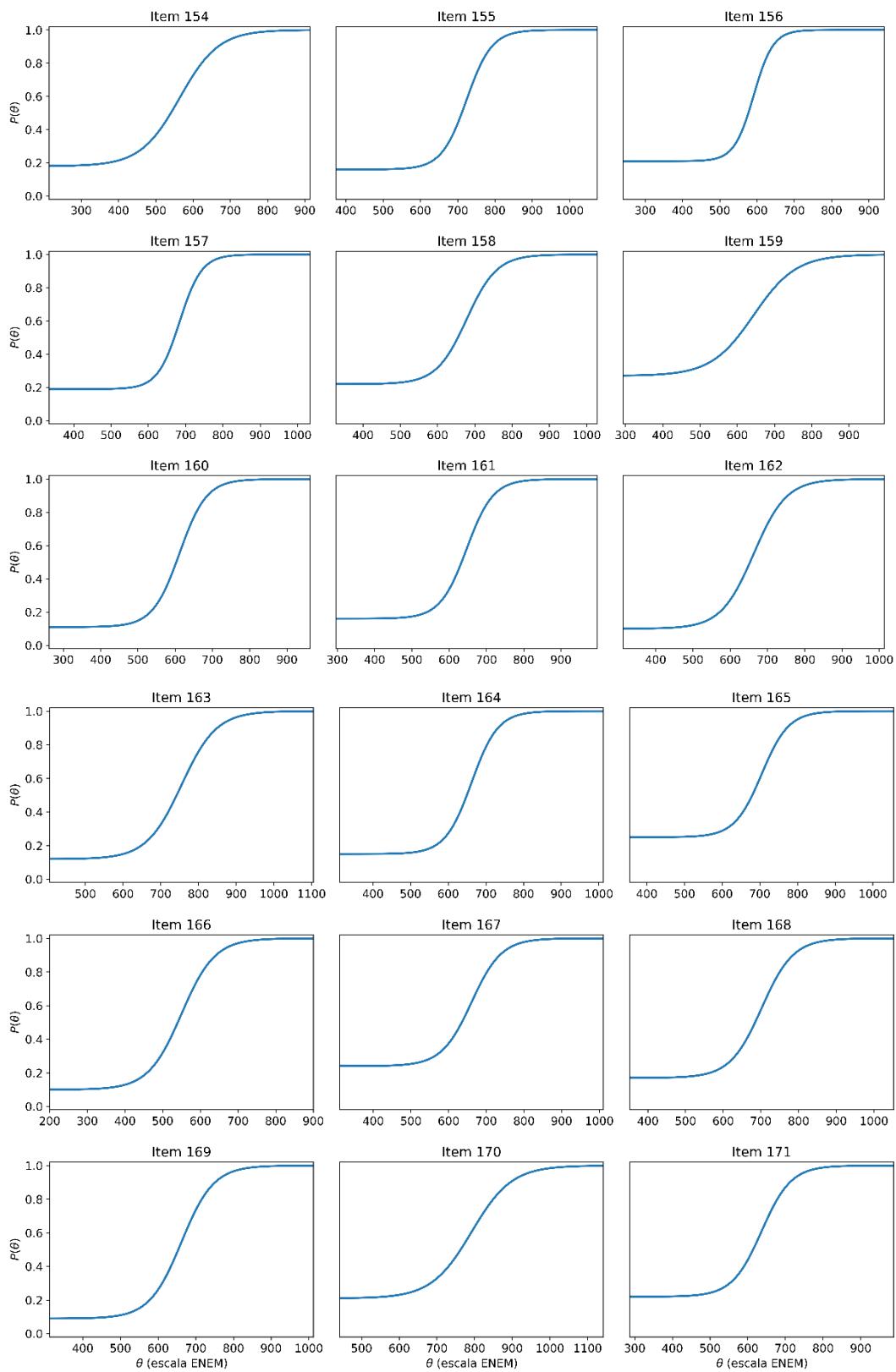
- Melhor e pior coerência pedagógica para a proficiência 500 (média)

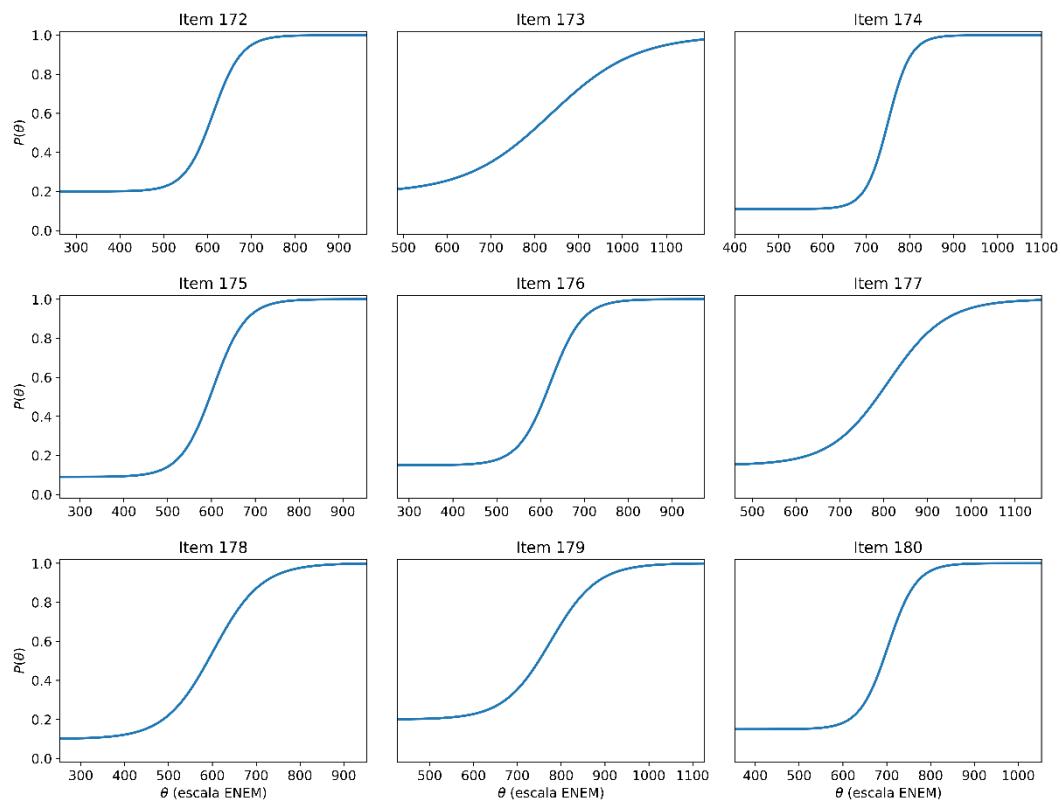
ACERTOS	NOTA	PADRÃO DE RESPOSTA	COERENCIA	INICIAIS	FINAIS
5	500	1100011000	Melhor	4	1
18	500	000010100010101000001110110001010100110100011	Pior	5	13

Distribuição de FR da amostra em relação às Redes de Ensino

APÊNDICE D – CCI dos itens da prova da área de Matemática e suas Tecnologias da aplicação regular de 2024 do Enem







Fonte: Inep.

Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 30 ago. 2025

Observação: Plots gerados pelo autor, a partir dos microdados do Enem/Inep

ANEXO A – Parâmetros dos itens da prova da área de Matemática e suas Tecnologias da aplicação regular de 2024 do Enem

Item	Gabarito	Habilidade	Parâmetro a	Parâmetro b	Parâmetro c
136	C	25	2,46865	673,917	523,8
137	E	27	1,93962	573,478	512,3
138	C	16	2,79451	706,903	529,4
139	E	2	3,12311	742,519	511,3
140	B	30	1,49785	786,851	521,2
141	E	19	2,90615	673,585	509,5
142	D	3	1,34998	701,873	521,9
143	A	5	1,49745	760,757	514,2
144	D	12	1,97145	729,841	514,6
145	C	8	1,0066	712,389	521,7
146	A	15	1,96986	625,923	516,8
147	A	22	5,44912	787,200	515,4
148	D	17	1,13982	673,496	510,1
149	E	28	1,26643	772,688	513,4
150	C	21	3,58253	729,438	531,2
151	D	12	3,29248	693,864	513,6
152	B	24	2,38298	645,126	518,4
153	B	15	1,84675	738,735	524,3
154	C	27	1,91272	564,470	518,1
155	E	9	2,96006	723,749	516,5
156	B	19	3,84247	591,232	520,6
157	D	16	3,37555	684,456	519,3
158	C	10	2,46593	678,408	521,5
159	C	4	1,76236	643,583	527,4
160	C	23	2,78314	611,704	511,3
161	A	29	2,81885	647,104	515,6
162	C	6	2,27814	663,004	509,6
163	A	13	2,16007	755,251	512,2
164	B	28	2,86713	662,276	515,3
165	B	3	2,79303	704,988	525,0
166	A	26	2,26960	550,185	510,0
167	B	4	2,52245	660,614	523,6
168	B	20	2,39398	701,893	517,2
169	A	2	2,35211	661,749	508,7
170	D	18	1,87853	791,962	520,9
171	D	7	2,56039	636,529	522,2
172	D	8	3,09197	613,019	520,0
173	C	13	1,02692	836,222	518,7
174	E	11	3,90485	750,442	511,2
175	A	25	2,70251	604,041	509,0
176	D	1	2,73571	622,835	514,6
177	B	22	1,52360	809,917	515,1
178	B	1	1,82455	601,771	510,0
179	C	21	1,89157	776,100	519,9
180	E	14	3,11951	703,277	514,9

Fonte: Inep

Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 30 ago. 2025

Observação: escala do parâmetro *b* alterada para média 500 e desvio padrão 100

ANEXO B – Matriz de Habilidades e Competências para a prova da Área de Matemática e suas Tecnologias do Enem

COMPETÊNCIAS	HABILIDADES
Competência de área 1 Construir significados para os números naturais, inteiros, racionais e reais.	H1: Reconhecer, no contexto social, diferentes significados e representações dos números e operações – naturais, inteiros, racionais ou reais. H2: Identificar padrões numéricos ou princípios de contagem. H3: Resolver situação-problema envolvendo conhecimentos numéricos. H4: Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas. H5: Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.
Competência de área 2 Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela.	H6: Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional. H7: Identificar características de figuras planas ou espaciais. H8: Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma. H9: Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos como solução de problemas do cotidiano.
Competência de área 3 Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano.	H10: Identificar relações entre grandezas e unidades de medida. H11: Utilizar a noção de escalas na leitura de representação de situação do cotidiano. H12: Resolver situação-problema que envolva medidas de grandezas. H13: Avaliar o resultado de uma medição na construção de um argumento consistente. H14: Avaliar proposta de intervenção na realidade utilizando conhecimentos geométricos relacionados a grandezas e medidas.

COMPETÊNCIAS	HABILIDADES
Competência de área 4 Construir noções de variação de grandezas para a compreensão da realidade e a solução de problemas do cotidiano.	H15: Identificar a relação de dependência entre grandezas. H16: Resolver situação-problema envolvendo a variação de grandezas, direta ou inversamente proporcionais. H17: Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação. H18: Avaliar propostas de intervenção na realidade envolvendo variação de grandezas.
Competência de área 5 Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas.	H19: Identificar representações algébricas que expressem a relação entre grandezas. H20: Interpretar gráfico cartesiano que represente relações entre grandezas. H21: Resolver situação-problema cuja modelagem envolva conhecimentos algébricos. H22: Utilizar conhecimentos algébricos/geométricos como recurso para a construção de argumentação. H23: Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.
Competência de área 6 Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extração, interpolação e interpretação.	H24: Utilizar informações expressas em gráficos ou tabelas para fazer inferências. H25: Resolver problema com dados apresentados em tabelas ou gráficos. H26: Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.
Competência de área 7 Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística.	H27: Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos. H28: Resolver situação-problema que envolva conhecimentos de estatística e probabilidade. H29: Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação. H30: Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade.

Fonte: Inep.

Disponível em: https://download.inep.gov.br/download/enem/matriz_referencia.pdf. Acesso em: 30 ago. 2025.

ANEXO C – Objetos do conhecimento para a prova da Área de Matemática e suas Tecnologias do Enem

CONHECIMENTOS	OBJETOS DO CONHECIMENTO
Conhecimentos numéricos	<ul style="list-style-type: none"> • Operações em conjuntos numéricos (naturais, inteiros, racionais e reais); • Desigualdades; • Divisibilidade e fatoração; • Razões e proporções; • Relações de dependência entre grandezas; • Porcentagem e juros; • Sequências e progressões; • Princípios de contagem.
Conhecimentos geométricos	<ul style="list-style-type: none"> • Características das figuras geométricas planas e espaciais; • Grandezas, unidades de medida e escalas; • Comprimentos, áreas e volumes; • Ângulos; • Posições de retas; • Simetrias de figuras planas ou espaciais; • Congruência e semelhança de triângulos; • Teorema de Tales; • Relações métricas nos triângulos; • Circunferências; • Trigonometria do ângulo agudo.
Conhecimentos de estatística e probabilidade	<ul style="list-style-type: none"> • Representação e análise de dados; • Medidas de tendência central (médias, moda e mediana); • Desvios e variância; • Noções de probabilidade.
Conhecimentos algébricos	<ul style="list-style-type: none"> • Gráficos e funções; • Funções algébricas do 1º e dos 2º graus, polinomiais, racionais, exponenciais e logarítmicas; • Equações e inequações; • Relações no ciclo trigonométrico e funções trigonométricas.
Conhecimentos algébricos/geométricos	<ul style="list-style-type: none"> • Plano cartesiano; • Retas; • Circunferências; • Paralelismo e perpendicularidade; • Sistemas de equações.

Fonte: Inep.

Disponível em: https://download.inep.gov.br/download/enem/matriz_referencia.pdf. Acesso em: 30 ago. 2025.