



Universidade Federal de Goiás  
Instituto de Matemática e Estatística  
Programa de Mestrado Profissional em  
Matemática em Rede Nacional



# Teoria de Resposta ao Item aplicada no ENEM

Sidney Tadeu Santiago Costa

Goiânia

2017

## TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR AS TESES E DISSERTAÇÕES ELETRÔNICAS NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

**1**            **1. Identificação do material bibliográfico:**     **Dissertação**             **Tese**

**1**            **2. Identificação da Tese ou Dissertação**

**2**

Nome completo do autor: Sidney Tadeu Santiago Costa

Título do trabalho: Teoria de Resposta ao Item Aplicada no ENEM

### **3. Informações de acesso ao documento:**

Concorda com a liberação total do documento  SIM             NÃO<sup>1</sup>

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.



Data: 03 / 03 / 2017

---

Assinatura do (a) autor (a) <sup>2</sup>

---

<sup>1</sup> Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

<sup>2</sup>A assinatura deve ser escaneada.

Sidney Tadeu Santiago Costa

# Teoria de Resposta ao Item aplicada no ENEM

Trabalho de Conclusão de Curso apresentado ao Instituto de Matemática e Estatística da Universidade Federal de Goiás, como parte dos requisitos para obtenção do grau de Mestre em Matemática.

Área de Concentração: Estatística

Goiânia

2017

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

COSTA, SIDNEY TADEU SANTIAGO

Teoria de Resposta ao Item aplicada no ENEM [manuscrito] /  
SIDNEY TADEU SANTIAGO COSTA. - 2017.

72 f.: il.

Orientador: Prof. Fabiano Fortunato Teixeira dos Santos; co orientador José Waldemar da Silva.

Trabalho de Conclusão de Curso Stricto Sensu (Stricto Sensu) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2017.

Bibliografia. Apêndice.

Inclui símbolos, gráfico, tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Teoria de Resposta ao Item. 2. Método da Máxima Verossimilhança. I. dos Santos, Fabiano Fortunato Teixeira, orient. II. Título.

CDU 519.2



Universidade Federal de Goiás-UFG  
Instituto de Matemática e Estatística-IME  
Mestrado profissional em Matemática em Rede  
Nacional - PROFMAT/UFG



Campus Samambaia – Caixa Postal 131 – CEP: 74.001-970 – Goiânia-GO.  
Fones: (62) 3521-1208 e 3521-1137 [www.ime.ufg.br](http://www.ime.ufg.br)

**Ata da reunião da Banca Examinadora da Defesa de Trabalho de Conclusão de Curso do aluno Sidney Tadeu Santiago Costa** – Aos três dias do mês de março do ano de dois mil e dezessete (03/03/2017), às 09:00 horas, reuniram-se os componentes da Banca Examinadora: Prof. Dr. Fabiano Fortunato Teixeira dos Santos – Orientador; Prof.<sup>a</sup>. Dr.<sup>a</sup>. Rosane Gomes Pereira e Prof. Dr. José Eder Salvador de Vasconcelos, para, sob a presidência do primeiro, e em sessão pública realizada no LEMAT, procederem a avaliação da defesa intitulada: “**Teoria de Resposta ao Item Aplicada no ENEM**”, em nível de Mestrado, área de concentração Matemática do Ensino Básico, de autoria de Sidney Tadeu Santiago Costa discente do Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT da Universidade Federal de Goiás. A sessão foi aberta pelo Presidente da banca, Prof. Dr. Fabiano Fortunato Teixeira dos Santos, que fez a apresentação formal dos membros da banca. A seguir, a palavra foi concedida ao autor do TCC que, em 30 minutos procedeu a apresentação de seu trabalho. Terminada a apresentação, cada membro da banca arguiu o examinando, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1075/2012 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta os Programas de Pós-Graduação da UFG e procedidas as correções recomendadas, o Trabalho foi **APROVADO** por unanimidade, considerando-se integralmente cumprido este requisito para fins de obtenção do título de **MESTRE EM MATEMÁTICA**, na área de concentração Matemática do Ensino Básico pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega na secretaria do IME da versão definitiva do trabalho, com as devidas correções supervisionadas e aprovadas pelo orientador. Cumpridas as formalidades de pauta, às 10:00 horas a presidência da mesa encerrou a sessão e para constar, eu, Sonia Maria de Oliveira, secretária do PROFMAT/UFG, lavrei a presente Ata que, depois de lida e aprovada, é assinada pelos membros da Banca Examinadora em quatro vias de igual teor.

Prof. Dr. Fabiano Fortunato Teixeira dos Santos  
Presidente – IME/UFG

Prof.ª. Dr.ª. Rosane Gomes Pereira  
Membro – IME/UFG

Prof. Dr. José Eder Salvador de Vasconcelos  
Membro – IFG/GOIÂNIA

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

**Sidney Tadeu Santiago Costa** é bacharel em Ciência da Computação e licenciado em Matemática, concluiu a pós graduação *latu senso* em matemática aplicada e atua desde 1995 como professor de matemática no ensino médio e pré vestibulares. Ministrou aulas no ensino superior nos cursos de engenharia e administração por 5 anos.

*Dedico esse trabalho a minha esposa Jaqueline Ferreira Fernandes que me apoiou em todos os momentos e aos meus filhos Ana Júlia e Heitor que dão sentido em tudo na minha vida.*

# Lista de Figuras

1.1	Curva característica do item . . . . .	18
2.1	Gráfico da probabilidade de sucesso em função da habilidade . . . . .	25
2.2	Alteração da constante de proporção . . . . .	26
2.3	Variações do parâmetro de dificuldade . . . . .	27
2.4	Variações do parâmetro de discriminação . . . . .	28
2.5	Escala de transformação do acerto casual . . . . .	29
2.6	Curva característica ideal para o parâmetro de discriminação . . . . .	31
3.1	Algoritmo de Newton-Raphson . . . . .	41
3.2	Estimação dos parâmetros e das habilidades . . . . .	42
3.3	Quantidade de iterações realizadas na estimação . . . . .	43
3.4	Planilha com 10 itens e 16 respondentes . . . . .	44
3.5	Planilha com os valores estimados . . . . .	46
4.1	Resultados para ciências humanas - software R . . . . .	49
4.2	Resultados para ciências humanas - planilha eletrônica . . . . .	50
4.3	Comparação entre resultados para ciências humanas - faixa de 36 até 41 acertos . . . . .	52
4.4	Comparação entre resultados para ciências humanas - faixa de 14 até 21 acertos . . . . .	53
4.5	Comparação entre resultados para ciências humanas - faixa de 3 até 16 acertos . . . . .	54



# Sumário

<b>1</b>	<b>Introdução</b>	<b>13</b>
1.1	Teoria clássica . . . . .	14
1.2	Visão geral da TRI . . . . .	15
1.3	A curva característica do item - CCI . . . . .	18
<b>2</b>	<b>Referencial Teórico</b>	<b>20</b>
2.1	Evolução histórica dos modelos . . . . .	20
2.1.1	Suposições da TRI . . . . .	20
2.1.2	Modelagem matemática . . . . .	22
2.2	Características gerais dos parâmetros . . . . .	30
<b>3</b>	<b>Estimação dos Parâmetros</b>	<b>33</b>
3.1	Método da máxima verossimilhança . . . . .	34
3.2	Aplicação das fórmulas . . . . .	39
3.3	Implementação do processo de estimação . . . . .	43
<b>4</b>	<b>Resultados e Discussões</b>	<b>48</b>
4.1	Tratamento dos dados no software R . . . . .	48
4.2	Tratamento dos dados nos algoritmos implementados . . . . .	50
<b>5</b>	<b>Conclusão</b>	<b>56</b>

# Agradecimentos

Agradeço aos amigos que me apoiaram no desenvolvimento da pesquisa, ao orientador professor Dr. Fabiano que, apesar da distância, possibilitou a realização do trabalho.

Um agradecimento especial ao professor Dr. José Waldemar da Universidade Federal de Uberlândia, que se propôs, de forma atenciosa e prestativa, ajudar nessa pesquisa.

Agradeço aos alunos que contribuíram com dúvidas durante as oficinas sobre a Teoria de Resposta ao Item no ENEM.

# Resumo

Com a nota obtida no Exame Nacional do Ensino Médio - ENEM os estudantes podem se candidatar a vagas em diversas instituições públicas de ensino superior e programas do governo, por exemplo, o programa Universidade para Todos (Prouni) e o Fundo de Financiamento Estudantil (Fies). O ENEM utiliza uma metodologia de correção das questões objetivas denominada Teoria de Resposta ao Item - TRI, que possui vários aspectos que são diferentes da Teoria Clássica dos Testes - TCT. O principal fator que determina o resultado de um sujeito em um processo avaliativo onde se utiliza a TCT, é o número de respostas corretas, enquanto na TRI, além da quantidade de acertos é fundamental se analisar quais respostas estão corretas. O objetivo deste trabalho é explicar o que é a TRI e como se aplica essa metodologia em avaliações de larga escala.

Será feita uma abordagem histórica dos modelos logísticos utilizados pela TRI e a justificativa da existência de cada parâmetro que compõe a equação principal da modelagem. Para determinar cada parâmetro que compõe o modelo da TRI e calcular a nota final de cada candidato, será utilizado um procedimento de otimização denominado Método da Máxima Verossimilhança - MMV.

As ferramentas computacionais no trabalho foram o software R, com pacotes desenvolvidos para aplicação da TRI e a linguagem de programação Visual Basic para programar funções, denominadas como macros, em planilhas eletrônicas.

**Palavras-Chaves:** Teoria de Resposta ao Item, Método da Máxima Verossimilhança.

# Abstract

With the note gotten in the Exame Nacional do Ensino Médio - ENEM the students can apply the vacant in diverse public institutions of superior education and programs of the government, for example, the program Universidade para Todos (Prouni) and the Fundo de Financiamento Estudantil (Fies). The ENEM uses a methodology of correction of the objective questions called Theory of Reply to the Item - TRI, that has some aspects that are different of the Classic Theory of the Tests - TCT. The main factor that determines the result of a citizen in a avaliativo process where if uses the TCT, is the number of correct answers, while in the TRI, beyond the amount of rightness is basic if to analyze which answers they are correct. The objective of this work is to explain what it is the TRI and as if it applies this methodology in evaluations of wide scale.

A historical boarding of the logistic models used by the TRI and the justification of the existence of each parameter will be made that composes the main equation of the modeling. To determine each parameter that composes the model of the TRI and to calculate the final note of each candidate, a procedure of called optimization will be used Method of Maximum Probability - MMV.

The computational tools in the work had been software R, with packages developed for application of the TRI and the Visual programming language beginner's all-purpose symbolic instruction code to program functions, called as macros, in electronic spread sheets.

**Key Works:** Item Response Theory - IRT, Maximum Likelihood Method - MMV.

# Capítulo 1

## Introdução

Este trabalho sobre a Teoria de Resposta ao Item (TRI) tem como foco a avaliação no sistema educacional para processos em que são envolvidos vários indivíduos, conhecidos como avaliações em larga escala. O Exame Nacional do Ensino Médio será o referencial prático de toda a pesquisa e a aplicação das modelagens será com simulados aplicados aos alunos do ensino básico em várias unidades de uma escola de ensino médio.

A TRI surgiu em meados dos anos 50, segundo Thurstone (1959), para resolver o problema clássico de psicometria relacionado a testes de inteligência; como propor um teste de inteligência cujos resultados não variem de acordo com os instrumentos utilizados? Imagine um problema análogo: o que aconteceria se o “metro” variasse na medição de uma largura e altura de um prédio? Pois bem, para criar um instrumento independente do objeto a ser avaliado, a TRI utiliza modelos e algoritmos matemáticos de complexa operacionalidade, razão esta que justifica a utilização dessa modelagem a partir da evolução da computação.

O ENEM é uma prova aplicada uma vez ao ano e propõe uma avaliação por competências e habilidades ao final do ensino básico e também confere a cada participante um parâmetro para a auto-avaliação e orientação de seu processo de formação continuada. Segundo Rabelo (2011), a partir de 2009, agregou-se novas funcionalidades ao exame, ampliando-se o caráter de processo seletivo para o acesso às instituições de educação superior, além de incorporar funções anteriormente atribuídas ao Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA) do ensino médio. Para implantar essas mudanças, foi necessário recorrer às técnicas oriundas da TRI. A utilização dessa teoria no ENEM abriu a possibilidade de se construir uma série histórica

do desempenho dos estudantes e dos egressos do ensino médio brasileiro, como é feito com os resultados do SAEB e da Prova Brasil, segundo Silva Carvalho (2001).

Este trabalho será dividido em cinco partes. O primeiro capítulo é a introdução, que possui a motivação e justificativa do tema escolhido bem como fundamentações teóricas para a utilização da Teoria de Resposta ao Item. No segundo capítulo, apresentaremos toda a evolução teórica dos modelos matemáticos da TRI. No terceiro, uma abordagem das metodologias de estimação dos parâmetros e da habilidade de cada respondente serão expostas. No quarto, será feita uma aplicação de todas as modelagens através do software R e com algoritmos implementados do Visual Basic.

## 1.1 Teoria clássica

Quando os processos de avaliação e/ou processos de seleção utilizam escores brutos ou padronizados obtidos com a análise direta dos acertos, quanto maior sua nota (número de acertos) maior será sua classificação. Estatisticamente, esse procedimento caracteriza análises e interpretações sempre associadas à prova como um todo e não se aprofunda nas características de cada item. Essa é a vertente da Teoria Clássica dos Testes - TCT, conforme descrito por Vianna (2014). Considere, por exemplo, um teste com 45 itens aplicado para 1000 indivíduos, onde vários acertaram 30 itens, mas ninguém acertou os mesmos exercícios. Será que todos que acertam o mesmo número de itens, têm o mesmo conhecimento sobre tais assuntos? Provavelmente não, mas a TCT não difere sujeitos que têm o mesmo número de acertos.

Para diferir cada sujeito dentro do teste, é necessário que se analise cada item que o compõe. Essa é a principal diretriz da Teoria de Resposta ao Item: considerar características psicométricas de cada item, capazes de diferenciar os indivíduos que realizaram um teste. Em muitas situações de medidas sociológicas, psicológicas ou educacionais, a variável de interesse é de entendimento intuitivo para todos, no caso de uma avaliação educacional como o ENEM, a variável é o conhecimento necessário para resolver um problema. Porém, na maioria das vezes, ela não é observável diretamente. É isto que a psicometria define como variáveis não observáveis ou habilidades ou traços latentes. Embora essas variáveis possam ser facilmente descritas e listadas, por exemplo, a inteligência, a habilidade em executar uma tarefa, ansiedade, o nível de entendimento de texto etc, elas não podem ser medidas diretamente como o peso ou altura de uma pessoa. Apesar de todas serem características implícitas a cada ser humano. A meta das

medidas educacionais e psicológicas é determinar como os traços latentes se processam na pessoa.

A TRI, de acordo com Lord (1968), se divide em duas etapas: a primeira é mensurar parâmetros psicométricos como dificuldade, discriminação e probabilidade de acerto casual, a partir das respostas de todos os sujeitos; em seguida, utiliza-se o conjunto de respostas de cada respondente e um modelo matemático que depende de tais parâmetros para calcular o traço latente de cada participante do teste. Em linhas gerais, a metodologia consiste em utilizar as respostas de todos os sujeitos para caracterizar cada item do teste através de uma função e, com todas as funções dos itens, calcula-se a habilidade de cada respondente para o padrão de respostas dicotômicas realizado no teste, ou seja, quais itens acertou e quais errou.

No caso de processos de avaliação, cria-se uma escala de proficiência para atribuir uma nota para cada habilidade. Para medir altura e comprimento que são observadas diretamente, pode-se utilizar o metro. A escala de proficiência é uma unidade análoga ao metro, destinada ao processo de medir habilidades. No caso do ENEM, faz-se uso de uma métrica com referência central igual a 500 (nota mínima média obtida para conclusão do ensino médio em 2009) e desvio padrão 100, conforme proposto por Pasquali (2007).

## 1.2 Visão geral da TRI

As avaliações fazem parte do processo de ensino e aprendizagem, elas envolvem troca entre pessoas e devem sempre ser construtivas, como ressalta Perrenoud (1999), “antes de regular as aprendizagens, a avaliação regula o trabalho, as atividades, as relações de autoridade e a cooperação em aula (...)”

Quando se cria um teste, o objetivo principal é medir um traço latente: o nível de conhecimento que o sujeito tem sobre determinado assunto, ou seja, qual conhecimento que ele possui para resolver as situações problemas. Os testes podem ser compostos por itens discursivos e itens objetivos. A prova discursiva é extremamente frágil à presença de erros e difícil confiabilidade (fidedignidade), pois dois avaliadores podem fazer leituras antagônicas de um mesmo contexto ocasionando resultados discrepantes em uma mesma avaliação. Entre pesquisadores da área de avaliação, o teste de resposta livre (discursivo) e o teste objetivo devem ser utilizados para qualquer processo educacional; entretanto, quando o número de examinados é muito grande, existe a necessidade de

resultados precisos e altamente fidedignos, imparciais, justos e válidos. O teste objetivo representa uma opção interessante, pois a correção fica livre de erros de julgamentos por parte dos examinadores, elimina-se a possibilidade de que um examinador valorize testes com boa apresentação e com características que não são relevantes no processo pretendido, e principalmente, tem-se rapidez na correção, conforme ressalta Vianna (2014).

Para que se tenha uma visão geral de como funciona a TRI, considere um teste composto por 10 itens com 5 alternativas e apenas uma correta, que foi aplicado para 16 sujeitos. As respostas foram organizadas em uma tabela conforme a Tabela 1.1. As respostas são dicotômicas, ou seja, 1 indica acerto e 0 erro. Após uma soma dos acertos, surgem dúvidas fundamentais para o processo de avaliação. Como diferir os respondentes que acertaram a mesma quantidade de itens? Será possível que um sujeito que acertou menos itens, tenha um conhecimento maior que outro que acertou mais? Por que muitas pessoas acertaram o item 01 enquanto poucas o item 10? Todos os itens têm o mesmo peso, no propósito de medir o nível de conhecimento do sujeito? Para que seja possível determinar o grau de conhecimento (habilidade) dos indivíduos que realizam o teste, a TRI propõe uma análise individual de cada item, mensurando características como discriminação, dificuldade e acerto casual e em seguida determinando a probabilidade de sucesso de um sujeito  $j$ , com habilidade  $\theta_j$ , em um item  $i$ , denota-se por  $P_i(\theta_j)$ . A habilidade de um sujeito é um valor real que varia no espectro de  $]-\infty, \infty]$ . Todo o processo é estruturado por um modelo matemático que será apresentado no Capítulo 2.



Tabela 1.1: Respostas após a aplicação do teste

Nº	Sujeitos	I-01	I-02	I-03	I-04	I-05	I-06	I-07	I-08	I-09	I-10	Total
1	Ana	1	1	1	1	1	1	1	1	1	0	9
2	Bia	1	1	1	1	1	1	1	1	0	0	8
3	Carlos	1	1	1	1	1	1	1	0	0	0	7
4	Débora	1	1	1	1	1	1	0	0	0	0	6
5	Érica	1	1	1	1	1	0	0	0	0	0	5
6	Fábio	1	1	1	1	0	0	0	0	0	0	4
7	Gil	1	1	1	0	0	0	0	0	0	0	3
8	Heitor	1	1	0	0	0	0	0	0	0	0	2
9	Igor	1	0	0	0	0	0	0	0	0	0	1
10	Júlia	0	0	0	0	0	0	0	0	0	1	1
11	Kelly	1	1	1	1	1	0	0	0	0	0	5
12	Leo	0	0	0	0	0	1	1	1	1	1	5
13	Mário	1	1	1	0	0	0	0	0	1	1	5
14	Nubia	1	1	0	0	0	0	0	1	1	1	5
15	Otávio	0	0	0	1	1	1	1	1	0	0	5
16	Paulo	0	0	1	1	1	1	1	0	0	0	5
	Total	12	11	10	9	8	7	6	5	4	4	

A representação gráfica do modelo matemático utilizado na TRI é denominada Curva Característica do Item - CCI. É possível compreender o que significa probabilidade de sucesso com este gráfico.

### 1.3 A curva característica do item - CCI

Competência é a capacidade de mobilização de recursos cognitivos, socioafetivos ou psicomotores, estruturados em rede, com vistas a estabelecer relações com e entre objetos, situações, fenômenos e pessoas para resolver, encaminhar e enfrentar situações complexas. Segundo Perrenoud (1999), uma das características importantes da noção de competência é desafiar o sujeito a mobilizar os recursos no contexto de situação-problema para tomar decisões favoráveis a seu objetivo ou a suas metas. As habilidades decorrem das competências adquiridas e referem-se ao plano imediato do saber fazer. Por maior que seja a habilidade de uma pessoa na resolução de um item, ela pode não ter sucesso e, por menor que seja sua habilidade, no caso do ENEM, ela pode fazer uma escolha aleatória da resposta e acertá-la. É claro que o conceito de probabilidade de acerto torna-se intrínseco nesse processo. O gráfico da probabilidade de acerto em função da habilidade é uma curva chamada de sigmóide, conforme a figura abaixo.

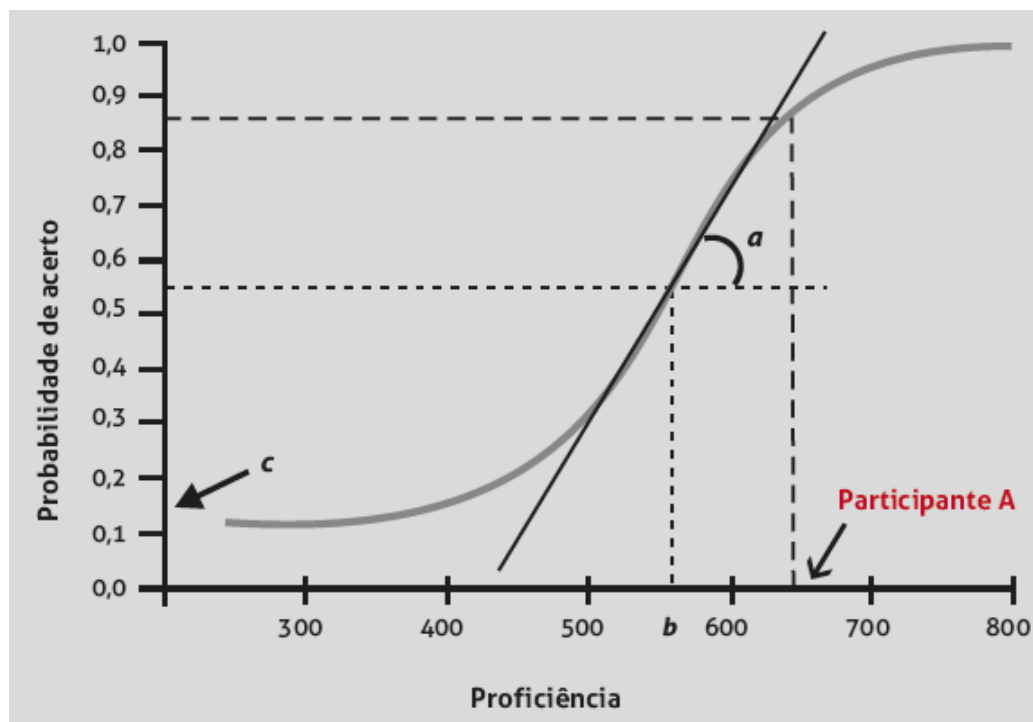


Figura 1.1: Curva característica do item

A proficiência do sujeito  $j$  é a habilidade  $\theta_j$  que ele tem para resolver o item  $i$ , representado no eixo horizontal. A métrica teórica desta variável vai de  $-\infty$  a  $+\infty$  mas, na prática, os valores estão aproximadamente de 0 a 1000, segundo Pasquali

(2007). A probabilidade de sucesso  $P_i(\theta_j)$  é uma função que depende da habilidade do respondente ( $\theta_j$ ) cujo gráfico é denominado de curva característica do item - CCI, apresentado na figura 1.1. Nesta função, quanto maior a habilidade respondente, maior será a probabilidade de sucesso no item, assim o gráfico é sempre crescente. A curva possui uma assíntota superior igual a 1 e uma inferior igual ao valor do parâmetro de acerto casual. Na prática, interpretamos que, quando uma pessoa sabe resolver o item, a probabilidade de acertar tende a 100% e, caso contrário, ela deve escolher a resposta aleatoriamente, ou seja, a probabilidade de sucesso é igual ao valor do parâmetro de acerto casual, denotado pela letra  $c$ . A CCI possui um ponto de inflexão onde a inclinação da reta tangente representa o parâmetro de discriminação do item. A derivada da função em cada ponto mostra que antes do ponto de inflexão, o crescimento da probabilidade de acerto é maior e depois menor.

Para justificar e aplicar toda a modelagem matemática da TRI, existe uma evolução histórica de cada modelo que foi apoiado em fatos e teoremas que serão explicados e aplicados com exemplos no próximo capítulo deste trabalho.

# Capítulo 2

## Referencial Teórico

A Teoria de Resposta ao Item é uma metodologia que mensura o nível de conhecimento sobre determinado conteúdo que uma pessoa tem ao realizar um teste. Para compreender todo o processo matemático existente nessa teoria, este capítulo apresenta os postulados da TRI e contextualiza a evolução histórica dos modelos matemáticos compostos por equações diferenciais ordinárias - EDO's.

### 2.1 Evolução histórica dos modelos

A proposta da TRI é expressar, por funções, a relação existente entre variáveis observadas e variáveis hipotéticas, chamadas de habilidade, traço latente ou proficiência. Assim, se conhecemos as características das variáveis observadas (parâmetros dos itens de um teste), estas se tornam constantes na equação, passam a ser denotadas como parâmetros e a equação torna-se solucionável, permitindo que se estime então o nível de aptidão do sujeito e vice-versa; isto é, se for conhecido o nível do traço latente do sujeito, é possível estimar as características dos itens respondidos por este sujeito. A seguir tem-se os dois postulados da Teoria de Resposta ao Item.

#### 2.1.1 Suposições da TRI

A teoria do traço latente possui dois princípios especiais ou básicos: **a unidimensionalidade** e **a independência local**. São suposições fundamentais que são ditas

hipóteses e não podem ser demonstradas e sim testadas, logo são aceitas ou não. Note que um modelo matemático para funcionar e ser útil precisa de algumas suposições entre o modelo e os dados, estabelecendo relações entre variáveis observáveis e empíricas. Mesmo não sendo demonstradas previamente, as suposições de forma indireta através de resultados podem levar a modelos práticos satisfatórios.

## A unidimensionalidade

As teorias dos traços latentes mostram que por de trás de um desempenho comportamental qualquer, existe um conjunto de variáveis que podem ser observadas e determinadas. Assim, um sujeito que se situa em um espaço de  $n$  dimensões, possui uma proficiência que pode ser expressa como função de um vetor de habilidades  $(\theta_1, \theta_2, \theta_3, \dots, \theta_n)$ , ou seja:

$$\text{Proficiência} = f(\theta_1, \theta_2, \theta_3, \dots, \theta_n)$$

É razoável aceitar que o pensamento humano seja multideterminado, ou seja, um grupo de traços latentes são executados na resolução de um problema. Fazendo uma análise superficial, esse fato compromete o princípio em questão, todavia, é suficiente assumir que na execução de uma tarefa, há uma aptidão dominante responsável pelo desempenho do respondente. Muitas pesquisas buscam soluções para modelos multidimensionais, entretanto, a TRI não possui modelos satisfatórios para esse caso, segundo Pasquali (2007). Resumindo, embora um sujeito possua um conjunto de habilidades  $(\theta_1, \theta_2, \theta_3, \dots, \theta_n)$  necessárias para resolver um item, este postulado assume que existe um  $\theta = \theta_j$ , para algum  $j$ , que será dominante, ou seja,

$$\text{Proficiência} = f(\theta).$$

## A independência local

Considerando que exista uma aptidão dominante responsável pela resposta de um indivíduo, assume-se o fato de que as respostas fornecidas a dois itens diferentes são eventos independentes. Esse é o axioma da Independência Local. Assim, este princípio decorre que a sequência de respostas do sujeito a uma série de itens é igual ao produto das probabilidades de sucesso ou fracasso de cada item separadamente.

Seja a probabilidade do sujeito  $j$ , com habilidade  $\theta_j$  que deu a resposta dicotômica  $u_{ji}$  no item  $i$ , denotada por  $P(u_{ji}|\theta_j)$ . A proposta matemática do princípio da

independência local, considerando que a probabilidade de sucesso em cada item do teste é um evento independente, é representada da seguinte forma:

$$P(u_{j1}, u_{j2}, u_{j3}, \dots, u_{jI} | \theta_j) = P_1(u_{j1} | \theta_j) \cdot P_2(u_{j2} | \theta_j) \cdot \dots \cdot P_I(u_{jI} | \theta_j) ,$$

onde o número total de sujeitos é  $N$ , o número total de itens é  $I$ ,  $i \in \{1, 2, \dots, I\}$  e  $j \in \{1, 2, \dots, N\}$ . A habilidade do sujeito  $j$  é considerada aptidão dominante para a resolução do item  $i$  e é denotada por  $\theta_j$ . A resposta do sujeito  $j$  dada ao item  $i$  é dicotômica, ou seja,  $u_{ji} = 1$  resposta correta e caso contrário,  $u_{ji} = 0$ . Por fim, a probabilidade de sucesso é  $P_i(u_i = 1 | \theta_j)$  e a probabilidade de fracasso é  $P_i(u_i = 0 | \theta_j)$ .

De acordo com a independência local, se houver alguma relação entre as respostas de um sujeito em vários itens, a justificativa não será por causa da habilidade dominante; logo, mantendo constantes tais fatores, é possível concluir que a única fonte de variação de resposta será em função da aptidão, assim as respostas tornam-se independentes e influenciadas apenas pelo teta dominante. Em resumo, a independência local implica na unidimensionalidade; então, se a última for comprovada a primeira também será. Outra informação relevante é que a independência local propõe a independência das respostas do sujeito, isto não significa que os itens não possam ser relacionados. Quando existe uma relação entre os itens, ela será definida pelo grupo de respondentes que possuem traços latentes próximos, logo respondem de forma semelhante.

Quando um indivíduo responde a uma série de itens, ele produz um padrão de respostas, composto pelos produtos das probabilidades ( $P_i$  para acerto e  $Q_i = 1 - P_i$  para erro). Como exemplo, para três itens e uma probabilidade individual de acerto igual a  $\frac{1}{5}$ , um padrão de resposta com acerto nas duas primeiras e erro na última é dado por:  $P_1 \cdot P_2 \cdot Q_3 = 0,2 \cdot 0,2 \cdot 0,8 = 0,032$ .

Na próxima seção, tem-se a dedução da equação que determina a probabilidade de sucesso em função de habilidade do sujeito.

### 2.1.2 Modelagem matemática

Vários fenômenos envolvem a variação de uma grandeza em função de outra, levando naturalmente a modelos baseados em equações diferenciais. Pode-se ter variações temporais da posição de um objeto, temperatura de um material, concentração de um agente químico, nutriente ou poluente em um meio, umidade do ar, quantidade de habitantes de uma cidade, densidade de bactérias, massa de um gás, valor de uma mercadoria, câmbio entre moedas, produto interno bruto e outras. Existem também as

variações em relação a outras grandezas como temperatura e posição, densidade e temperatura, a probabilidade de sucesso em um item e a habilidade de quem o respondeu, entre outras, segundo Kent (2012). As equações diferenciais representam diversas leis fundamentais. Podem ser empíricas, como em reações químicas, ou heurísticas, como em dinâmica populacional, modelo que foi aplicado na TRI, segundo Baker (2001), em meados de 1950. O ajuste da função utilizada na TRI é análogo ao processo utilizado no crescimento populacional e passou por duas propostas. Inicialmente, a proposta de Malthus, onde a taxa de variação do número de indivíduos no instante  $t$  é proporcional à quantidade de indivíduos presentes no ambiente neste instante. Seja  $N$  o número de indivíduos em um tempo  $t$  e  $k$  uma constante de proporção; assim a EDO é

$$\frac{dN}{dt} = kN.$$

Supondo  $N(0) = N_0$ , a solução é dada por:

$$N(t) = N_0 e^{kt}.$$

Para  $k < 0$  a população decresce e, caso contrário, a população cresce sem limites. Como existem limitações do ambiente como a falta de alimentos para toda a população, doenças, predadores, alterações climáticas e outras, este modelo torna-se pouco satisfatório para um longo prazo.

A segunda modelagem, denominada Modelo de Verhulst-Pearl, considera que a taxa de variação do número de indivíduos é proporcional à quantidade de indivíduos presentes no ambiente ( $N$ ) e quantidade de indivíduos que completaria a capacidade do ambiente ( $L$ ). Neste caso a EDO é escrita assim

$$\frac{dN}{dt} = kN \left(1 - \frac{L}{N}\right),$$

onde  $k$  é uma constante de proporção, com  $N(0) = N_0$  tem solução particular dada por:

$$N(t) = \frac{L}{1 + \left(\frac{L}{N_0} - 1\right)e^{-kt}}.$$

Neste caso, o número de indivíduos ficará sempre entre a quantidade existente no tempo zero e a capacidade do ambiente, podendo ser determinada mesmo quando o tempo é grande. Para a TRI, tem-se que a variável que representa o número de indivíduos na dinâmica populacional passa ser a probabilidade de acerto em um item, de acordo com Fletcher (1994). A equação diferencial será apresentada adiante. Considere

uma pessoa com habilidade  $\theta$ ; ao resolver um item, sua probabilidade de sucesso será  $P$  e fracasso,  $1 - P$ . Para uma habilidade nula ( $\theta = 0$ ) a probabilidade de sucesso e fracasso são iguais, ou seja,  $\theta = 0$  e  $P = \frac{1}{2}$ . Assim a taxa de variação da probabilidade é proporcional à probabilidade de sucesso e fracasso. Ressaltando que a probabilidade de sucesso  $P$  é uma função que depende da habilidade  $\theta$ , ou seja,  $P = f(\theta)$ , a EDO é

$$\frac{dP}{d\theta} = kP(1 - P). \quad (2.1)$$

Considere  $k$  uma constante de proporção. Utilizando o método da separação de variáveis proposto no estudo de equações diferenciais, tem-se:

$$\int \frac{dP}{P(1 - P)} = \int kd\theta.$$

Como  $\frac{1}{P(1 - P)} = \frac{1}{P} + \frac{1}{1 - P}$ , segue que

$$\int \left( \frac{1}{P} + \frac{1}{1 - P} \right) dP = \int kd\theta.$$

Logo,

$$\ln(P) - \ln(1 - P) = k\theta + C \Rightarrow \ln\left(\frac{P}{1 - P}\right) = k\theta + C.$$

Segue que

$$e^{k\theta} e^C = \frac{P}{1 - P}.$$

Como  $\theta = 0$  e  $P = \frac{1}{2}$ , fazendo as substituições devidas chega-se em  $C = 1$ . Finalizando a resolução da EDO, chega-se na equação que determina a probabilidade de sucesso em função da habilidade de um sujeito

$$e^{k\theta} = \frac{P}{1 - P};$$

ou seja,

$$P = P(\theta) = \frac{1}{1 + e^{-k\theta}}. \quad (2.2)$$



Adotando no eixo das abscissas a escala de proficiência denominada habilidade  $\theta$  e no eixo das ordenadas a probabilidade de acerto  $P(\theta)$ , o gráfico da função (2.2), é representado pela figura adiante.

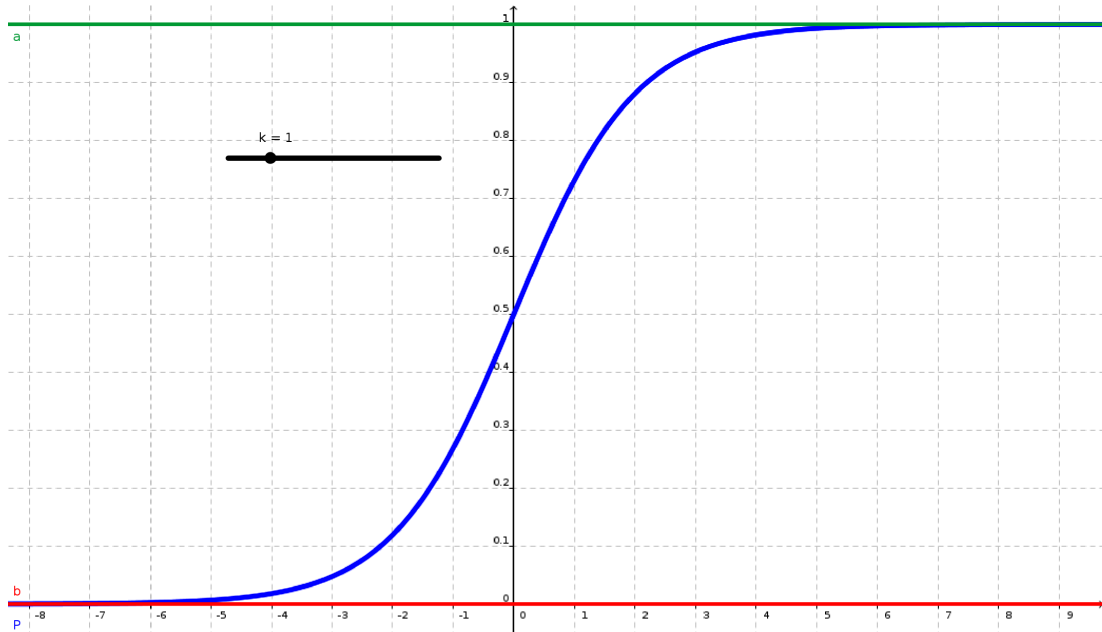


Figura 2.1: Gráfico da probabilidade de sucesso em função da habilidade

Nota-se que o gráfico é uma curva crescente. Para habilidades muito pequenas, a probabilidade de sucesso tende a zero e, em caso oposto, a probabilidade tende a um. Existe um ponto de inflexão em  $\theta = 0$  e para valores do domínio  $\theta < -3$  ou  $\theta > 3$ , a variação na probabilidade torna-se muito pequena, fato que pode ser verificado calculando-se a derivada em tais pontos. Com relação à constante de proporção  $k$ , segundo Rabelo (2011), para  $k = 1,7$  a função (2.2), possui o gráfico semelhante ao gráfico da função de distribuição de uma Gaussiana. A figura 2.2 mostra a variação da constante, de  $k = 1$  para  $k = 1,7$ .

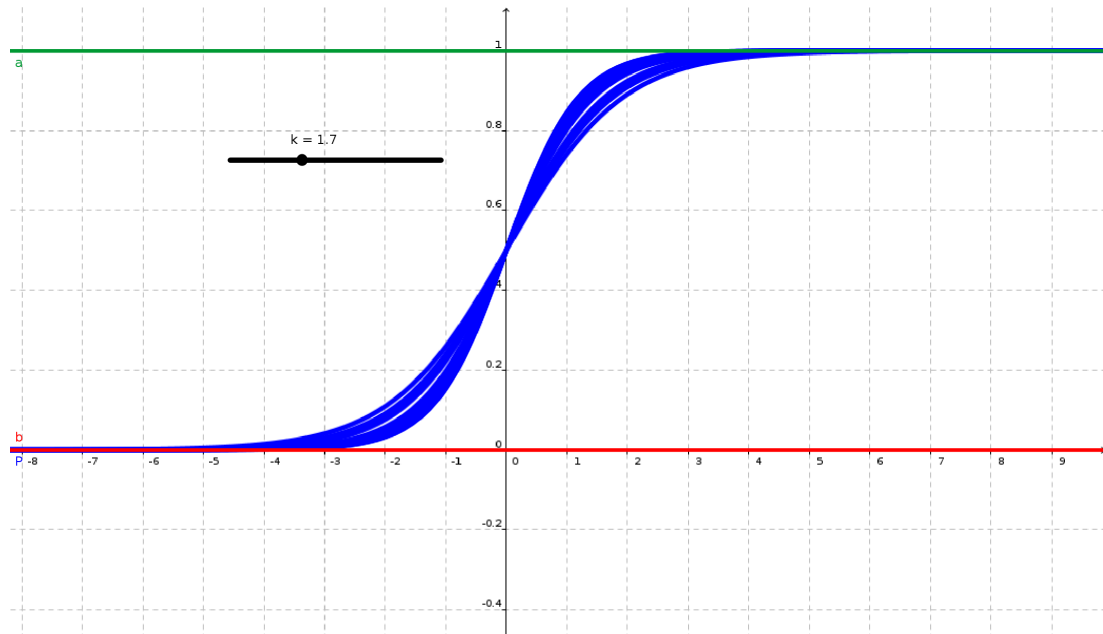


Figura 2.2: Alteração da constante de proporção

Para  $k < 0$ , a função não tem representatividade, pois é decrescente (a habilidade aumenta e a probabilidade diminui) e para  $k = 0$  não existe variação de probabilidade, tem-se uma função constante que não se aplica na TRI. Por fim, para simplificar as demonstrações dos modelos que introduziram os parâmetros psicométricos de discriminação, dificuldade e acerto casual, adota-se  $k = 1$ , como estabelecido em Pasquali (2007).

O modelo logístico com um parâmetro surge pela translação horizontal do gráfico da função (2.2) denotado por  $b$ . Se  $b < 0$  gráfico desloca para a direita e caso contrário para a esquerda. Para  $b = 0$  o ponto de inflexão será  $\theta = 0$ . Este parâmetro está na mesma escala da habilidade. O modelo apresentado na função (2.2) sofre uma modificação determinando o modelo logístico - ML1, representado pela lei:

$$P(\theta) = \frac{1}{1 + e^{-(\theta-b)}}. \quad (2.3)$$

A figura 2.3 representa as translações definidas pelas modificações do parâmetro de dificuldade aplicado na função (2.3).

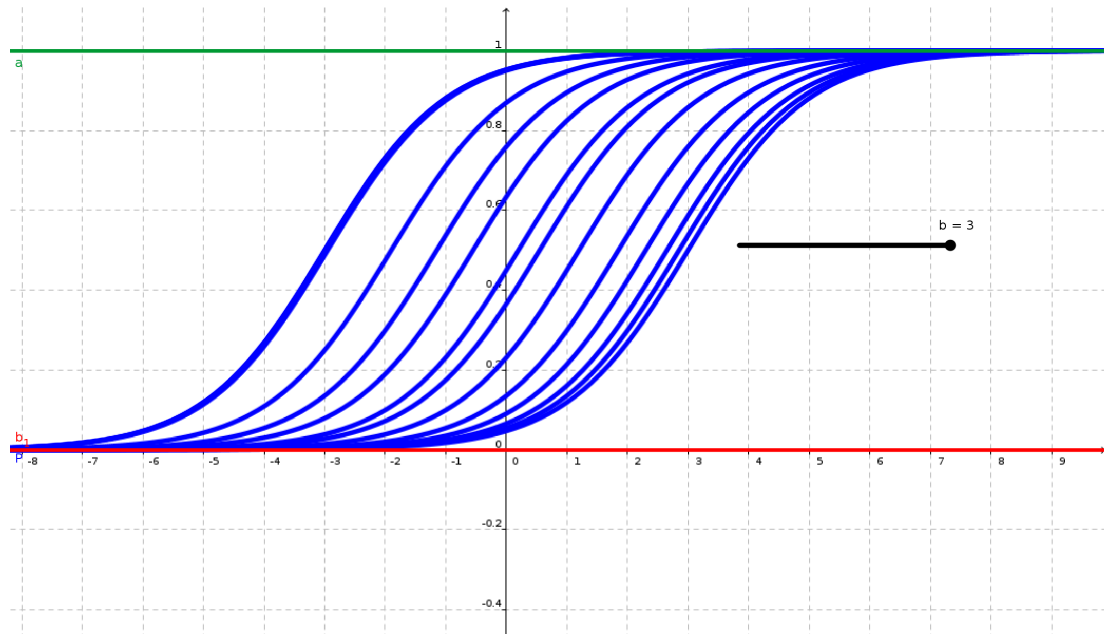


Figura 2.3: Variações do parâmetro de dificuldade

Considere um valor fixo para a probabilidade de sucesso em um item, o deslocamento para a direita do gráfico (2.3) significa que a habilidade para acertar o item deve aumentar, conclui-se que a dificuldade do item aumenta. Quando o deslocamento ocorre para a esquerda, a dificuldade diminui.

Para o modelo logístico de dois parâmetros - ML2, surge um parâmetro capaz de diferenciar sujeitos com habilidades muito próximas, chamado discriminação ( $a$ ). Retomando a equação diferencial apresentada na equação 2.1 e considerando que a taxa de probabilidade de sucesso seja proporcional à probabilidade de sucesso, fracasso e ao valor do parâmetro de discriminação do item, chega-se na seguinte EDO

$$\frac{dP}{d\theta} = kP(1 - P)a. \quad (2.4)$$

Note que o parâmetro discriminação ( $a$ ) no processo de resolução das integrais é uma constante, pois os diferenciais são  $dP$  e  $d\theta$ . Assim, utilizando a mesma técnica para resolver equações diferenciais separáveis apresentado anteriormente, o modelo logístico de dois parâmetros é modelado por

$$P(\theta) = \frac{1}{1 + e^{-a(\theta-b)}}. \quad (2.5)$$

Aumentando o parâmetro de discriminação, a curva sigmóide torna-se bem abaulada, determinando valores diferentes para habilidades bem próximas, conforme apresentado na figura adiante.

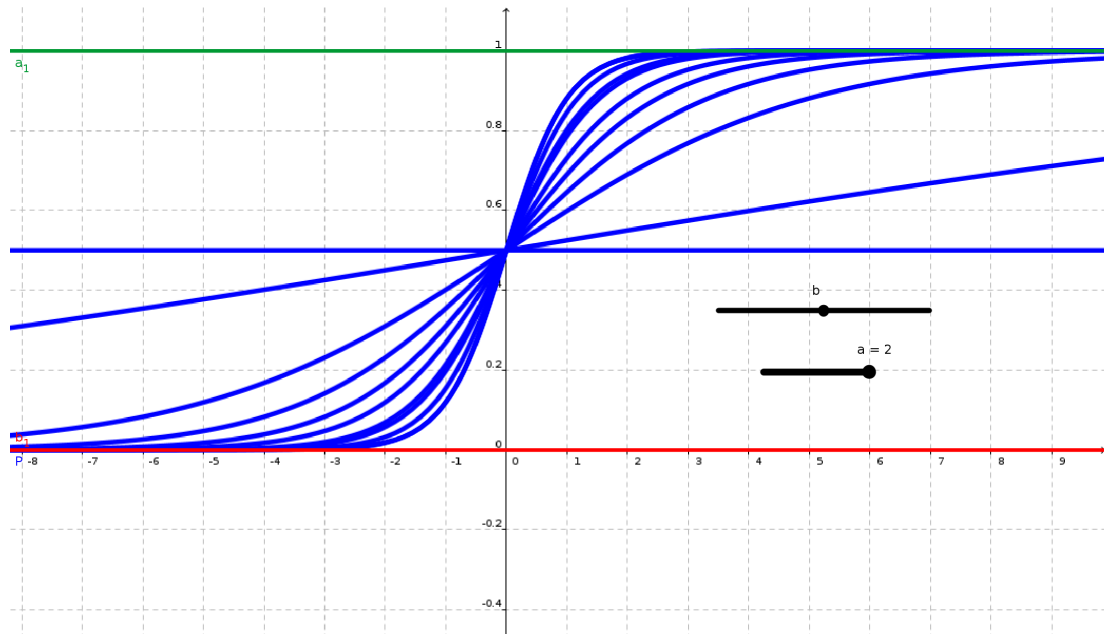


Figura 2.4: Variações do parâmetro de discriminação

Por fim, tem-se o modelo logístico com três parâmetros - ML3, com o parâmetro de acerto casual denotado por  $c$ . No Enem, assim como qualquer avaliação objetiva, a escolha de uma resposta para um item pode ser feita de forma aleatória e, com cinco alternativas (apenas uma correta) a probabilidade mínima de sucesso, passa de 0 para 0.2. Para introduzir esta propriedade na função 2.5 é necessário realizar uma transformação linear. A figura 2.5 mostra os valores de referência utilizados para a transformação dos valores de probabilidade.

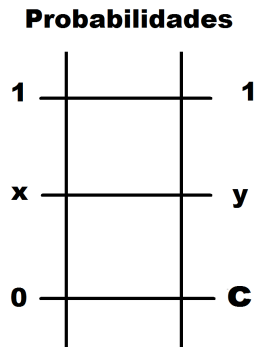


Figura 2.5: Escala de transformação do acerto casual

Para finalizar a função a ser utilizada na TRI, falta apenas considerar a escolha aleatória da resposta do item, ou seja, o chute. Por se tratar de itens objetivos, a menor probabilidade de acerto no item deixa de ser zero (assíntota inferior) e passa para 20% (cada item tem cinco alternativas sendo apenas uma correta). Esta característica desloca a parte inferior da CCI para cima. O deslocamento do limite inferior é feito por uma mudança de escala que utiliza os valores extremos como referência. Note que a probabilidade mínima na escala do modelo de dois parâmetros é zero, e no novo modelo é igual ao valor parâmetro de acerto casual ( $c = 20\%$ ), já os valores máximos continuam sendo um, assim, a ideia do modelo de três parâmetros é criar uma correspondência de cada probabilidade do ML2, no intervalo de 0 a 1, com um único valor de probabilidade no intervalo de  $c$  até 1. Essa correspondência é feita de forma proporcional, cada variação de probabilidades em um modelo é proporcional à variação de probabilidades do outro. Seja  $x$  o valor da probabilidade na função (2.5) e  $y$  a nova probabilidade no ML3, a proporção entre as variações de probabilidades será

$$\frac{x - 0}{y - c} = \frac{1 - 0}{1 - c},$$

ou seja,

$$y = c + (1 - c)x.$$

Assim o modelo com três parâmetros fica:

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}} \quad (2.6)$$

Esta função representa a modelagem fundamental da TRI, o modelo matemático que, uma vez determinados os parâmetros de dificuldade ( $b$ ), a discriminação ( $a$ ) e o acerto casual ( $c$ ); é possível calcular a probabilidade de um sujeito acertar o item de um teste.

Para reescrever a igualdade 2.6 , considera-se que o teste é composto por  $I$  itens que serão resolvidos por  $N$  de indivíduos com  $i \in \{1, 2, \dots, I\}$  e  $j \in \{1, 2, \dots, N\}$ .

Segue que:

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-a_i \cdot (\theta_j - b_i)}}, \quad (2.7)$$

onde  $\theta_j$  é a habilidade do indivíduo  $j$ ,  $P_i(\theta_j)$  probabilidade do indivíduo  $j$  acertar o item  $i$ ,  $a_i$  é a discriminação do item  $i$ ,  $b_i$  é a dificuldade do item  $i$  e  $c_i$  é o acerto casual do item  $i$ .

A próxima seção aborda as principais características dos parâmetros e os valores esperados para que a função 2.7 possa representar a probabilidade de acerto em um item.

## 2.2 Características gerais dos parâmetros

No ML3 existe uma habilidade latente que justifica o acerto ou erro de um sujeito em cada item isoladamente. O modelo matemático proposto na função 2.7 define que, para cada item, existe um nível de discriminação ( $a$ ), a dificuldade do item ( $b$ ) e acerto casual ( $c$ ). Os valores de cada parâmetro estão em uma escala contínua e possuem interpretações específicas. No próximo capítulo será feita uma abordagem de como os valores de cada um dos parâmetros são calculados.

A discriminação deve ser concebida como a capacidade do item em diferenciar indivíduos com habilidades diferentes. Na TRI, a discriminação é o parâmetro capaz de diferenciar os sujeitos com traços latentes próximos na competência a ser aferida. Na curva característica de cada item, a discriminação é dada pelo coeficiente angular da reta tangente no ponto de inflexão (ponto onde o gráfico muda a concavidade). Quanto maior o valor do parâmetro de discriminação ( $a$ ), mais íngreme será a curva e mais discriminativo será o item. Valores próximos de zero indicam que o item tem baixo poder de diferenciar os respondentes que sabem dos que não têm o conhecimento necessário para responder a questão, assim não cumpre os propósitos de uma avaliação educacional. Valores negativos para o parâmetro de discriminação indicam que o

item se comporta de maneira incoerente, pois o aumento da habilidade implicaria na diminuição da probabilidade de acerto, assim considera-se apenas os valores positivos ( $a > 0$ ).

Para o parâmetro de discriminação, espera-se valores entre 0 e 2, sendo que os valores mais apropriados são aqueles maiores que um, de acordo com Rabelo (2011). Diante disso, o gráfico ideal é dado pela figura abaixo.

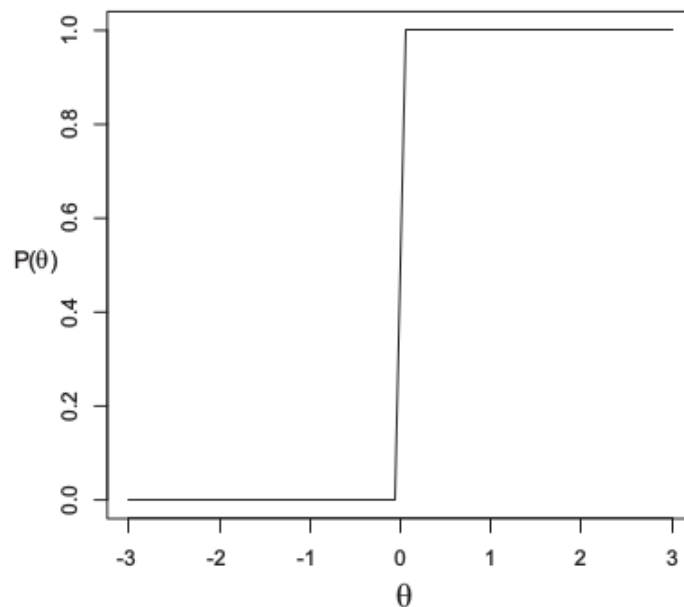


Figura 2.6: Curva característica ideal para o parâmetro de discriminação

A dificuldade é o valor da aptidão  $\theta$  necessário para se obter uma probabilidade de acerto igual a  $\frac{1+c}{2}$ , ou seja, quando a habilidade é igual a dificuldade ( $\theta - b = 0$ ) na função 2.7. Quando não é permitido a escolha aleatória, a dificuldade é o valor da habilidade que produz a probabilidade de 0.5 de sucesso para o item. Quanto maior for o nível de aptidão necessário para que o sujeito acerte o item, maior será a dificuldade do mesmo. A dificuldade é representada em uma escala contínua que varia no intervalo  $(-\infty, \infty)$ . Itens com valores de  $b$  menores que  $-3$  são considerados muito fáceis, entre  $-3$  e  $-1$  fáceis, entre  $-1$  e  $1$  intermediários, entre  $1$  e  $3$  difíceis e os valores maiores que  $3$  muito difíceis. O nível de dificuldade ideal para os itens de um teste depende da sua finalidade. Em avaliações educacionais, recomenda-se uma distribuição dentro de uma curva normal: 10% dos itens em cada uma das faixas extremas (muito fácil

e muito difícil), 20% em cada uma das faixas seguintes (fácil e difícil) e 40% na faixa média (intermediária), segundo Pasquali (2007). A seguir tem-se a tabela dos níveis de dificuldade esperados para um item.

Tabela 2.1: Níveis para o parâmetro dificuldade

Classificação	Valores para o parâmetro
Muito Fácil	$b \leq -3$
Fácil	$-3 < b < -1$
Intermediário	$-1 \leq b < 1$
Difícil	$1 \leq b < 3$
Muito Difícil	$b \geq 3$

O Parâmetro de acerto casual aumenta a probabilidade de acerto no item, deslocando para cima a assíntota inferior; no caso do Enem, esse valor é 20%. Nota-se que esse parâmetro aumenta a probabilidade de sucesso, veja o exemplo a seguir. Para  $\theta = 2$ ,  $a = b = 1$  e  $c = 0$ , a probabilidade de sucesso, aplicando a equação 2.5 resulta em

$$P(\theta = 2) = \frac{1}{1 + 2 \cdot 7^{-1 \cdot (2-1)}} = 0,73$$

e no ML3 da equação 2.6 o resultado fica

$$P(\theta = 2) = 0,2 + 0,8 \cdot \frac{1}{1 + 2 \cdot 7^{-1 \cdot (2-1)}} = 0,78.$$

Isto é, com o acerto casual, o item torna-se hipoteticamente mais fácil, subindo sua probabilidade de acerto de 73% para 78%.

Cada item possui seus parâmetros com valores específicos que são calculados por estimação. Este procedimento matemático será explicado no próximo capítulo. A habilidade de cada respondente também será estimada da mesma forma. A estimação supõe um valor inicial para cada parâmetro e para a habilidade, em seguida, aplica-se várias regras e os valores são atualizados verificando se houve uma convergência dos valores calculados nas iterações anteriores, tornando tais valores constantes, denominados os parâmetros dos itens. Os parâmetros calculados definem uma função para cada item e com as funções de todos os itens é possível calcular a habilidade que cada sujeito possui. Com a habilidade ou proficiência, tem-se a nota final obtida no teste.



## Capítulo 3

# Estimação dos Parâmetros

A estimação de parâmetros trabalha com probabilidades de eventos independentes. De fato, a independência das respostas dos respondentes é um postulado na teoria (independência local). Dois ou mais eventos consideram-se independentes entre si se a ocorrência ou a não-ocorrência de um não influencia a ocorrência do(s) outro(s), segundo Stevenco (1981). Se dois ou mais eventos são independentes, então a probabilidade de ocorrência simultânea de todos é igual ao produto de suas probabilidades individuais. Para  $n$  eventos  $A_1, A_2, \dots, A_n$ , temos:

$$P(A_1 \text{ e } A_2 \text{ e } \dots \text{ e } A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n)$$

Quando os eventos são mutuamente excludentes, a probabilidade de ocorrência de qualquer deles (por definição, não podem ocorrer dois ou mais conjuntamente) é a soma das probabilidades individuais. Generalizando, temos:

$$P(A_1 \text{ ou } A_2 \text{ ou } \dots \text{ ou } A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

Para estimar os parâmetros e a habilidade, utiliza-se um padrão de respostas conforme apresentado na seção 2.1.1.

Uma das etapas mais importantes da TRI é a estimação dos parâmetros dos itens e das habilidades dos respondentes. A probabilidade de uma resposta correta a um determinado item depende da habilidade do indivíduo e dos parâmetros que caracterizam o item. Mas, em geral, ambos são desconhecidos. Apenas as respostas dos indivíduos aos itens do teste são conhecidas.

Tem-se um problema bastante complicado cujo objetivo é calcular os parâmetros (discriminação, dificuldade e acerto casual) e em seguida determinar a proficiência

(habilidade) de um sujeito. Para resolver o problema, a TRI utiliza o método da máxima verossimilhança. Este método é um processo matemático que otimiza uma função (chamada função de verossimilhança) através de algoritmos iterativos. Neste trabalho utilizaremos o algoritmo de Newton-Raphson.

Na próxima seção serão apresentadas todos os modelos matemáticos utilizados na aplicação do método da máxima verossimilhança para estimar os parâmetros de todos os itens e as habilidades de todos os respondentes que realizaram um teste.

### 3.1 Método da máxima verossimilhança

Considere um teste composto por  $I$  itens respondidos por  $N$  sujeitos ( $i \in \{1, 2, \dots, I\}$  e  $j \in \{1, 2, \dots, N\}$ ) de forma que cada resposta, dada pela pessoa  $j$  ao item  $i$ , seja determinada pela função  $u_{ji}$  de forma que para  $u_{ji} = 1$  a resposta do sujeito está correta e para  $u_{ji} = 0$  a resposta está errada. A probabilidade de sucesso do sujeito  $j$  no item  $i$  é calculada pela função 2.7 e será expressa por  $P_i(\theta_j)$  ou simplesmente  $P_i$ . Por fim, tem-se o evento complementar definido pelo fracasso no item e denotado por  $Q_i = 1 - P_i$ .

Pelo princípio da independência local, define-se uma função  $l$ , como produto das probabilidades para um padrão de respostas dadas pelo sujeito  $j$  a todos os  $I$  itens. A função  $l$  é definida por:

$$l = \prod_{i=1}^I P_i(\theta_j)^{u_{ji}} \cdot Q_i(\theta_j)^{1-u_{ji}}. \quad (3.1)$$

O objetivo é calcular o valor de cada parâmetro  $a_i$ ,  $b_i$  e  $c_i$  de todos os itens e a habilidade  $\theta_j$  de cada respondente, que determinam o maior valor para a função  $l$ . Para isso, a TRI utiliza o método da máxima verossimilhança. A função (3.1) será chamada de função de verossimilhança. Em linhas gerais, o processo consiste em calcular uma raiz da derivada da função  $l$ . Sabe-se que a derivada de um produto de funções é bem mais complicada que a derivada da soma, dessa forma uma técnica útil, consiste em tomar o logaritmo em ambos os membros de (3.1). Assim, chamando de  $L = \ln(l)$ , obtemos:

$$\begin{aligned}
L &= \ln \left( \prod_{i=1}^I P_i(\theta_j)^{u_{ji}} \cdot Q_i(\theta_j)^{1-u_{ji}} \right) \\
&= \sum_{i=1}^I \ln(P_i(\theta_j)^{u_{ji}} \cdot Q_i(\theta_j)^{1-u_{ji}}) \\
&= \sum_{i=1}^I u_{ji} \cdot \ln(P_i(\theta_j)) + (1 - u_{ji}) \cdot \ln(Q_i(\theta_j)). \tag{3.2}
\end{aligned}$$

Para verificar se existe, e em caso positivo determinar a raiz da derivada implícita da função (3.2) em relação a cada parâmetro e em seguida em relação à habilidade, utilizaremos o método iterativo de Newton-Raphson. Este método foi inicialmente introduzido por Isaac Newton para determinar as raízes de uma função por meio de sucessivas iterações. Neste método, começa-se com um valor arbitrário para  $x$  e, em seguida, em repetidas iterações, vão se fazendo pequenas mudanças para melhorar a aproximação à solução. Mais tarde, Joseph Raphson melhorou o método de Newton, introduzindo as derivadas nas repetidas iterações para resolver equações lineares e não-lineares.

Para compreender melhor o algoritmo, considere uma função  $f(x)$  real, de variável real que possui uma raiz. Para encontrar o valor desta raiz, devemos encontrar um valor de  $x$  tal que  $f(x) = 0$ . Utilizando o método de Newton-Raphson, fazemos uma atribuição inicial  $x_n = x_0$ , com  $n \in \mathbb{N}$ , calculamos a derivada da função no ponto com abscissa  $x_0$  e aplicamos, para  $n = 0$ , a fórmula :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{3.3}$$

onde  $x_n$  é valor para  $x$ ,  $x_{n+1}$  é um valor obtido com a aplicação da fórmula (3.5),  $f(x_n)$  é o valor da função  $f$  em  $x_n$  e  $f'(x_n)$  é o valor da primeira derivada da função  $f$  em  $x_n$ . Com o valor obtido com a aplicação da fórmula (3.5), calcula-se a diferença entre  $x_1$  e  $x_0$ , caso a diferença seja tão pequena quanto se queira, a raiz foi encontrada, caso contrário, atualizamos o valor de  $x_n$  e realizamos uma nova iteração.

Para aplicar Newton-Raphson na função (3.2) a função  $f(x)$  será a primeira derivada de  $L$  porque o objetivo é encontrar o valor máximo da função  $L$ , ou seja,  $L' = 0$  e  $f'(x)$  será a segunda derivada da função  $L$ . Sem perda de generalidade, adota-se

$P_{ji}(u_{ji} = 1|\theta_j) = P$ ,  $P_{ji}(u_{ji} = 0|\theta_j) = Q$ ,  $\theta_j = \theta$ ,  $a_i = a$ ,  $b_i = b$ ,  $c_i = c$ ,  $u_{ji} = u$  e as representações de derivadas implícitas como  $\frac{\partial L}{\partial \theta} = L'_\theta$ ,  $\frac{\partial^2 L}{\partial \theta^2} = L''_\theta$ ,  $\frac{\partial L}{\partial a} = L'_a$ ,  $\frac{\partial^2 L}{\partial a^2} = L''_a$ ,  $\frac{\partial L}{\partial b} = L'_b$ ,  $\frac{\partial^2 L}{\partial b^2} = L''_b$ ,  $\frac{\partial L}{\partial c} = L'_c$  e  $\frac{\partial^2 L}{\partial c^2} = L''_c$ .

A seguir tem-se o algoritmo de Newton-Raphson para calcular cada um dos parâmetros e a habilidade de cada sujeito.

$$\left\{ \begin{array}{l} \theta_{n+1} = \theta_n - \frac{L'_\theta(\theta_n)}{L''_\theta(\theta_n)} \\ a_{n+1} = a_n - \frac{L'_a(a_n)}{L''_a(a_n)} \\ b_{n+1} = b_n - \frac{L'_b(b_n)}{L''_b(b_n)} \\ c_{n+1} = c_n - \frac{L'_c(c_n)}{L''_c(c_n)} \end{array} \right. \quad (3.4)$$

Para que o método da máxima verossimilhança seja executado faltam as derivadas da função (3.2) que serão determinadas adiante.

Para determinar os parâmetros, o somatório refere-se ao conjunto de sujeitos  $j \in \{1, 2, \dots, N\}$  e para a habilidade, ao conjunto de itens  $i \in \{1, 2, \dots, I\}$ . Cada aplicação de Newton-Raphson é feita separadamente, assim adotamos a notação sem os contadores  $i$  e  $j$  para o conjunto de itens. Neste momento as derivadas serão realizadas para a equação (2.7) do ML3, assim, sem perdas de generalidades, deixaremos as variações do somatório, considerando que a derivada da soma é a soma das derivadas de uma função. Reescrevendo a igualdade (3.2), tem-se:

$$L = u \ln P + (1 - u) \ln Q.$$

A seguir utilizaremos a regra da cadeia para derivada de funções compostas. Vamos derivar a função (3.2) em relação à função 2.7, em seguida faremos as derivadas da função (2.7) em relação aos parâmetros ( $a$ ,  $b$  e  $c$ ) e em relação a habilidade ( $\theta$ ).

Agora vamos determinar a primeira derivada da função de verossimilhança  $L$ :

$$\begin{aligned} L' &= u \frac{1}{P} P' + (1 - u) \frac{1}{Q} Q' \\ &= \frac{Q P' u + Q' P - Q' P u}{PQ} \\ &= \frac{Q P' u + Q' P - Q' P u}{PQ}. \end{aligned} \quad (3.5)$$

Lembrando que  $Q' = -P'$ , obtemos

$$\begin{aligned} L' &= \frac{QP'u - P'P + P'Pu}{PQ} \\ &= \frac{P'u(Q + P) - P'P}{PQ}. \end{aligned}$$

Como  $Q + P = 1$ , segue que

$$L' = \frac{P'u - P'P}{PQ} = \frac{(u - P)}{PQ}P'.$$

Fazendo a substituição  $T = \frac{u - P}{PQ}$ , chegamos a primeira derivada da função  $L$

$$L' = TP'.$$

A última etapa é calcular a derivada de segunda ordem da função de máxima verossimilhança. Primeiramente vamos calcular a derivada de  $T$

$$\begin{aligned} T' &= \frac{(u - P)'PQ - (u - P)(PQ)'}{(PQ)^2} \\ &= \frac{-P'PQ - (u - P)(P'Q + PQ')}{(PQ)^2} \\ &= \frac{-P'PQ - (uP'Q + uPQ' - PP'Q - P^2Q')}{(PQ)^2} \\ &= \frac{-P'PQ - uP'Q - uPQ' + PP'Q + P^2Q'}{(PQ)^2} \\ &= \frac{-uP'(1 - P) - uP(-P') + P^2(-P')}{(PQ)^2} \\ &= \frac{-uP' + uP'P + uPP' - P^2P'}{(PQ)^2} \\ &= \frac{-uP' + 2uP'P - P^2P'}{(PQ)^2} \\ &= \frac{-P'(u - 2uP + P^2)}{(PQ)^2}. \end{aligned}$$

Sabe-se que  $u = 0$  ou  $u = 1$ , o que permite utilizar  $u^2 = u$ . Segue então, que

$$\begin{aligned} T' &= \frac{-P'(u - P)^2}{(PQ)^2} \\ &= -P'T^2. \end{aligned} \tag{3.6}$$

Agora a segunda derivada de  $L$  será:

$$\begin{aligned} L'' &= T'P' + TP'' \\ &= -P'T^2P' + TP'' \\ &= -(P'T)^2 + TP''. \end{aligned} \tag{3.7}$$

Lembrando que:  $L' = TP'$ , finalizamos a segunda derivada

$$L'' = TP'' - (L')^2.$$

Retomando as notações para um conjunto de itens (e quando necessário, conjunto de sujeitos), tem-se:

$$L'_j = \sum_{i=1}^I \frac{u_{ji} - P_i}{P_i Q_i} \cdot P'_i \tag{3.8}$$

$$L''_j = \sum_{i=1}^I \frac{u_{ji} - P_i}{P_i Q_i} \cdot P''_i - (L'_j)^2. \tag{3.9}$$

Para finalizar o método da máxima verossimilhança temos que calcular as derivadas da função (2.7) em relação a cada parâmetro e a habilidade. Como

$$P = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta_j - b_i)}},$$

a primeira derivada em relação a habilidade é

$$P'_\theta = \frac{a_i(1 - c_i)e^{-a_i(\theta - b_i)}}{(1 + e^{-a_i(\theta - b_i)})^2}$$

e a segunda derivada é

$$P''_{\theta} = \frac{2a_i^2(1-c_i)e^{-2a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^3} - \frac{a_i^2(1-c_i)e^{-a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^2}.$$

Agora, derivando em relação ao parâmetro de discriminação obtemos

$$P'_a = -\frac{(1-c_i)(1-b_i)e^{-a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^2}$$

e a segunda derivada é

$$P''_a = \frac{2(1-c_i)(b_i-\theta)^2e^{-2a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^3} - \frac{(1-c_i)(b_i-\theta)^2e^{-a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^2}.$$

Em relação ao parâmetro dificuldade, a primeira derivada é

$$P'_b = -\frac{a_i(1-c_i)(1-b_i)e^{-a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^2}$$

e a segunda derivada é

$$P''_b = \frac{2a_i^2(1-c_i)e^{-2a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^3} - \frac{a_i^2(1-c_i)e^{-a_i(\theta-b_i)}}{(1+e^{-a_i(\theta-b_i)})^2}.$$

Finalmente, a primeira derivada em relação ao parâmetro de acerto casual é

$$P'_c = 1 - \frac{1}{1+e^{-a_i(\theta-b_i)}}$$

e a segunda derivada é

$$P''_c = 0.$$

Todas as derivadas foram calculadas para a aplicação do algoritmo de Newton-Raphson na função de máxima verossimilhança para estimar os parâmetros de cada item  $i$  ( $a_i$ ,  $b_i$  e  $c_i$ ) e a habilidade  $\theta_j$  de cada sujeito  $j$  que realizou o teste.

Na próxima seção, vamos organizar a utilização de todas as fórmulas que foram deduzidas para a aplicação do método da máxima verossimilhança.

## 3.2 Aplicação das fórmulas

Para a estimação dos parâmetros e das habilidades o algoritmo de Newton-Raphson (N-R) será aplicado várias vezes. Na primeira etapa, o algoritmo é utilizado para calcular os parâmetros e depois para determinar as habilidades.

Cada iteração do algoritmo consiste em fazer uma atribuição inicial para a incógnita que se deseja estimar (parâmetro ou habilidade), em seguida, as funções (3.5) e (3.6) são utilizadas para o grupo de sujeitos, quando se pretende calcular os parâmetros e, para o grupo de itens, quando o objetivo é calcular o traço latente. Para finalizar a iteração, atualiza-se o valor procurado e, caso a variação entre o valor anterior e o atual, seja menor que um valor fixado, por exemplo  $10^{-6}$ , finaliza-se o processo determinando o valor para a incógnita. Se o valor ainda for maior que o esperado, reinicia-se os procedimentos com uma nova iteração.

Em resumo, o algoritmo de N-R será mostrado na figura a seguir.



Atribuições iniciais:  $\alpha_1, b_1, c_1$  ou  $\theta_1$

**Para**  $i = 1$  até  $I$  (quando for habilidade) ou **Para**  $j = 1$  até  $N$  (quando for parâmetros)

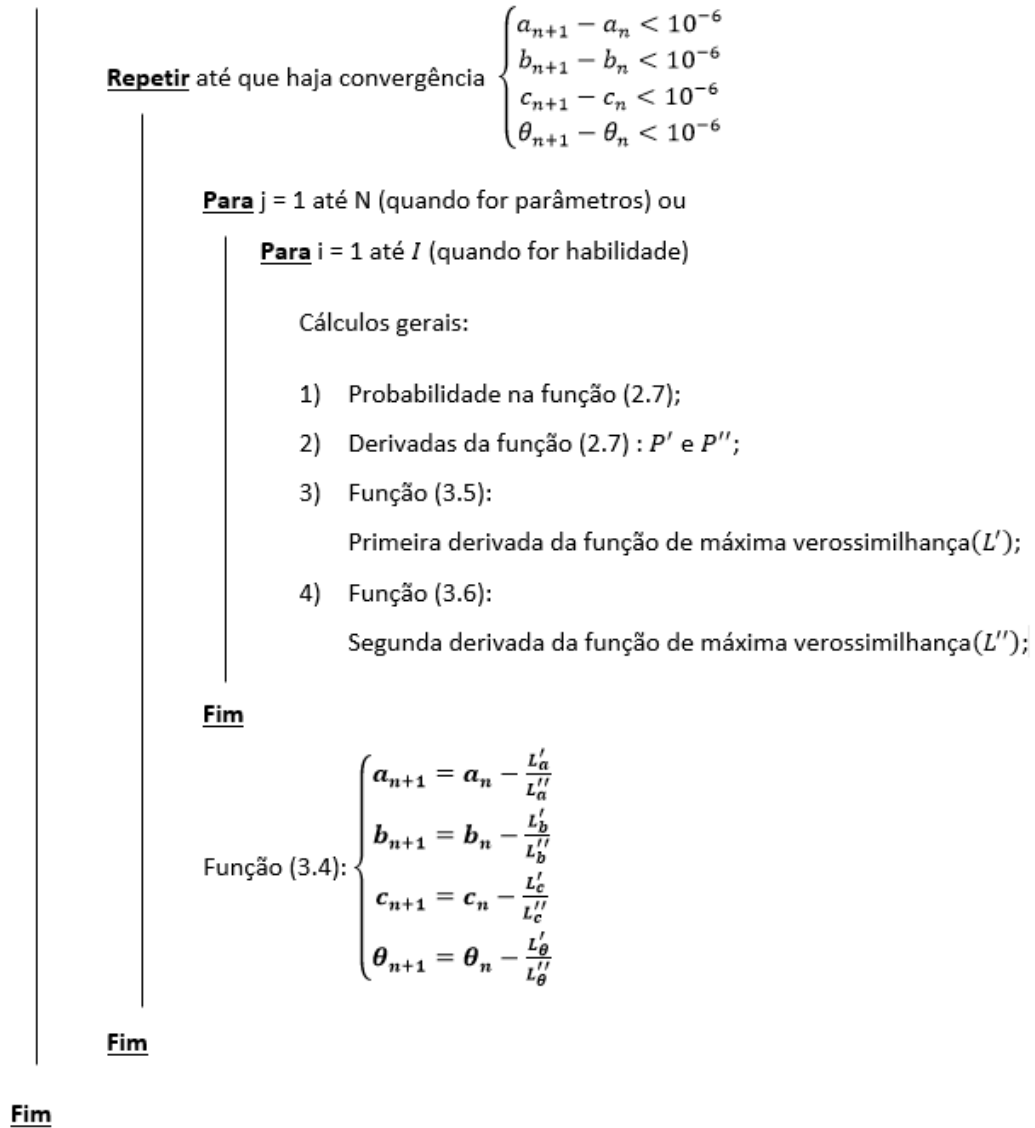


Figura 3.1: Algoritmo de Newton-Raphson

A rotina apresentada na figura (3.1) será aplicada para cada item determinando os parâmetros e para cada sujeito calculando a traço latente. Considerando um teste com  $I$  itens e  $N$  sujeitos, o algoritmo N-R será executado  $3I$  vezes para calcular os

parâmetros e depois mais  $N$  vezes para calcular as habilidades, ou seja,  $3I + N$  vezes para cada etapa da estimação. No exemplo da tabela (1.1), o algoritmo será executado 30 vezes para determinar os parâmetros e mais 16 vezes para determinar as habilidades dos sujeitos, logo N-R será executado 46 vezes.

Para completar o processo de estimação devemos repetir a proposta da figura (3.1) por várias vezes, até que os valores dos parâmetros e das habilidades não tenham variações consideráveis. Para esse procedimento, não foi estabelecido um critério de parada, mas sim um numero fixo de repetições (chamaremos de  $T$ ). Esse fato se justifica pela complexidade de programação necessária para comparar  $3I$  parâmetros e  $N$  habilidades simultaneamente.

Na próxima figura será apresentada a estrutura de repetição para o processo de estimação.

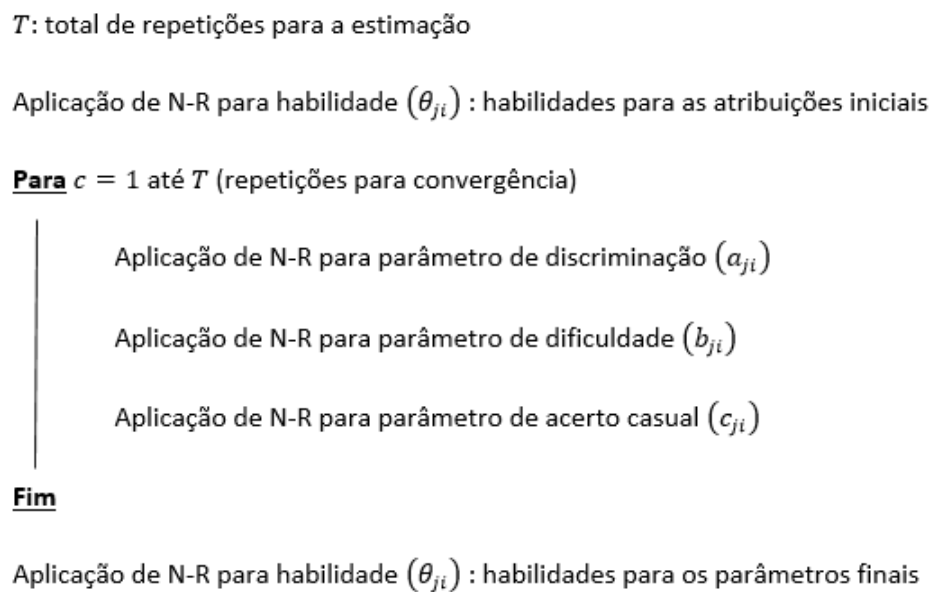


Figura 3.2: Estimação dos parâmetros e das habilidades

No processo de estimação dos parâmetros e das habilidades cada rotina deve ser executada várias vezes. Sendo assim, sem programação, seria extremamente exaustivo os procedimentos matemáticos estabelecidos na TRI. Como exemplo, o cálculo da probabilidade de sucesso, proposto na função (2.7) no exemplo da tabela (1.1), para uma

repetição é realizado 3.084 vezes e para 100 repetições, onde ocorre a convergência dos valores, a fórmula foi utilizada 56.284 vezes. A figura a seguir mostra essa simulação.

<b>Antes da convergência</b>				<b>Depois da convergência</b>			
Iterações (N-R):		3.084		Iterações (N-R):		56.284	
Repetições		100	100	Repetições		100	100
Itens:		10		Itens:		10	
Sujeitos:		16		Sujeitos:		16	
A p a g a	Ordena	Discriminação		Ordena	Discriminação		
	Início	Dificuldade		Início	Dificuldade		
	Habilidade	Acaso		Habilidade	Acaso		
	S	Acertos	Nota	S	Acertos	Nota	
1	9	790,927	1	9	815,863		

Figura 3.3: Quantidade de iterações realizadas na estimação

Na próxima seção, serão implementados todos os algoritmos apresentados para a estimação dos parâmetros e das habilidades.

### 3.3 Implementação do processo de estimação

Para implementar os algoritmos apresentados nas seções anteriores, foi criada uma planilha eletrônica e as todas as funções do método da máxima verossimilhança foram desenvolvidos em uma linguagem de programação chamada de Visual Basic, que permite interações entre as células da planilha e rotinas de programação. A planilha permite uma visualização didática dos resultados obtidos. As respostas da tabela 1.1 com 10 itens e 16 respondentes foram inseridas nesta planilha, conforme apresenta a figura (3.4).

Itens (N-R):		ITEM - 01	ITEM - 02	ITEM - 03	ITEM - 04	ITEM - 05	ITEM - 06	ITEM - 07	ITEM - 08	ITEM - 09	ITEM - 10
0	100	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
10	100	0,25	0,31	0,38	0,44	0,50	0,56	0,63	0,69	0,75	0,75
16	100	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20
Registres											
Itens:											
Sujeitos:											
Ordem											
Início											
Habilidade											
Acessos											
5	Nota	12	11	10	9	8	7	6	5	4	4
1		1	1	1	1	1	1	1	1	1	1
2		1	1	1	1	1	1	1	1	1	0
3		1	1	1	1	1	1	1	1	0	0
4		1	1	1	1	1	1	1	0	0	0
5		1	1	1	1	1	0	0	0	0	0
6		1	1	1	1	0	0	0	0	0	0
7		1	1	1	1	0	0	0	0	0	0
8		1	1	1	0	0	0	0	0	0	0
9		0	0	0	0	0	0	0	0	0	0
10		0	0	0	0	0	0	0	0	0	1
11		1	1	1	1	1	0	0	0	0	0
12		0	0	0	0	0	1	1	1	1	1
13		1	1	1	1	0	0	0	0	0	1
14		1	1	1	0	0	0	0	1	1	1
15		0	0	0	1	1	1	1	1	0	0
16		0	0	0	1	1	1	1	0	0	0

Figura 3.4: Planilha com 10 itens e 16 respondentes

A parte superior esquerda apresenta os botões com os principais comandos, seguidos das curvas características dos 45 itens que definem o traço latente. Logo abaixo, a primeira coluna apresenta a identificação dos sujeitos ( $1, 2, 3, \dots, N$ ), a nota e as demais colunas tem as respostas binárias dos  $N$  sujeitos aos 45 itens. Esta planilha possui as estimativas iniciais para a discriminação de cada item igual a um ( $a = 1$ ), recomendado por Pasquali (2007), a dificuldade de cada item determinada pela razão entre erros no item e total de itens (por exemplo, o item 01 tem 12 acertos logo,  $b = \frac{4}{16} = 0,25$ ) e o acerto casual igual a 20% para todos os itens, pois cada item possui cinco alternativas e apenas uma correta.

Foram implementados os algoritmos para a estimação de cada parâmetro e a habilidade, indicados pelos botões rotulados por Discriminação, Dificuldade, Acaso e Habilidade. Foi desenvolvida uma rotina (indicada pelo rótulo Iterações) que realiza chamadas consecutivas das rotinas de estimação para discriminação, dificuldade, acaso e habilidade, respectivamente, para que se verifique a convergência dos valores.

Segundo o guia do estudante criado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP, o cálculo da nota é realizado em uma escala que depende de dois valores: valor de posição ou de referência, para o qual foi atribuído o valor 500, que representa o desempenho médio dos concluintes do ensino médio da rede pública de 2009 (primeiro ano de aplicação da TRI) que realizaram o exame naquele ano e, valor de dispersão, para o qual foi atribuído o valor 100, que representa uma medida de variabilidade média das notas desses concluintes em relação ao desempenho médio 500. Esse valor é conhecido como desvio padrão. Resumindo, com a estimação das habilidades, o cálculo da nota é feito pela equação  $N = 100\theta + 500$ .

Após a realização das estimativas os resultados foram apresentados na tabela a seguir.

Itens		ITEM - 01	ITEM - 02	ITEM - 03	ITEM - 04	ITEM - 05	ITEM - 06	ITEM - 07	ITEM - 08	ITEM - 09	ITEM - 10
Repetições		100	100	100	100	100	100	100	100	100	100
Itens:		10	10	10	10	10	10	10	10	10	10
Sujeitos:		16	16	16	16	16	16	16	16	16	16
Discriminação		0,61	0,69	1,32	1,62	1,54	1,58	1,99	1,09	1,70	1,00
Ordem		1	1	1	1	1	1	1	1	1	1
Início		0,63	-1,62	-0,74	-0,15	0,10	0,42	0,42	1,11	2,93	2,46
Dificuldade		0,63	0,00	0,00	0,20	0,20	0,20	0,05	0,07	0,20	0,21
Aceite		12	11	10	9	8	7	6	5	4	4
Nota		815,863	700,147	636,872	561,555	550,660	508,970	461,770	458,657	415,637	407,059
Aceitos		1	1	1	1	1	1	1	1	1	1
1		9	8	7	6	5	4	3	2	1	1
2		8	7	6	5	4	3	2	1	1	1
3		7	6	5	4	3	2	1	1	1	1
4		6	5	4	3	2	1	1	1	1	1
5		5	4	3	2	1	1	1	1	1	1
6		4	3	2	1	1	1	1	1	1	1
7		3	2	1	1	1	1	1	1	1	1
8		2	1	1	1	1	1	1	1	1	1
9		1	1	1	1	1	1	1	1	1	1
10		1	1	1	1	1	1	1	1	1	1
11		5	4	3	2	1	1	1	1	1	1
12		5	4	3	2	1	1	1	1	1	1
13		5	4	3	2	1	1	1	1	1	1
14		5	4	3	2	1	1	1	1	1	1
15		5	4	3	2	1	1	1	1	1	1
16		5	4	3	2	1	1	1	1	1	1
17		5	4	3	2	1	1	1	1	1	1
18		5	4	3	2	1	1	1	1	1	1
19		5	4	3	2	1	1	1	1	1	1
20		5	4	3	2	1	1	1	1	1	1
21		5	4	3	2	1	1	1	1	1	1
22		5	4	3	2	1	1	1	1	1	1
23		5	4	3	2	1	1	1	1	1	1
24		5	4	3	2	1	1	1	1	1	1
25		5	4	3	2	1	1	1	1	1	1
26		5	4	3	2	1	1	1	1	1	1
27		5	4	3	2	1	1	1	1	1	1
28		5	4	3	2	1	1	1	1	1	1
29		5	4	3	2	1	1	1	1	1	1
30		5	4	3	2	1	1	1	1	1	1
31		5	4	3	2	1	1	1	1	1	1
32		5	4	3	2	1	1	1	1	1	1
33		5	4	3	2	1	1	1	1	1	1
34		5	4	3	2	1	1	1	1	1	1
35		5	4	3	2	1	1	1	1	1	1
36		5	4	3	2	1	1	1	1	1	1
37		5	4	3	2	1	1	1	1	1	1
38		5	4	3	2	1	1	1	1	1	1
39		5	4	3	2	1	1	1	1	1	1
40		5	4	3	2	1	1	1	1	1	1
41		5	4	3	2	1	1	1	1	1	1
42		5	4	3	2	1	1	1	1	1	1
43		5	4	3	2	1	1	1	1	1	1
44		5	4	3	2	1	1	1	1	1	1
45		5	4	3	2	1	1	1	1	1	1
46		5	4	3	2	1	1	1	1	1	1
47		5	4	3	2	1	1	1	1	1	1
48		5	4	3	2	1	1	1	1	1	1
49		5	4	3	2	1	1	1	1	1	1
50		5	4	3	2	1	1	1	1	1	1
51		5	4	3	2	1	1	1	1	1	1
52		5	4	3	2	1	1	1	1	1	1
53		5	4	3	2	1	1	1	1	1	1
54		5	4	3	2	1	1	1	1	1	1
55		5	4	3	2	1	1	1	1	1	1
56		5	4	3	2	1	1	1	1	1	1
57		5	4	3	2	1	1	1	1	1	1
58		5	4	3	2	1	1	1	1	1	1
59		5	4	3	2	1	1	1	1	1	1
60		5	4	3	2	1	1	1	1	1	1
61		5	4	3	2	1	1	1	1	1	1
62		5	4	3	2	1	1	1	1	1	1
63		5	4	3	2	1	1	1	1	1	1
64		5	4	3	2	1	1	1	1	1	1
65		5	4	3	2	1	1	1	1	1	1
66		5	4	3	2	1	1	1	1	1	1
67		5	4	3	2	1	1	1	1	1	1
68		5	4	3	2	1	1	1	1	1	1
69		5	4	3	2	1	1	1	1	1	1
70		5	4	3	2	1	1	1	1	1	1
71		5	4	3	2	1	1	1	1	1	1
72		5	4	3	2	1	1	1	1	1	1
73		5	4	3	2	1	1	1	1	1	1
74		5	4	3	2	1	1	1	1	1	1
75		5	4	3	2	1	1	1	1	1	1
76		5	4	3	2	1	1	1	1	1	1
77		5	4	3	2	1	1	1	1	1	1
78		5	4	3	2	1	1	1	1	1	1
79		5	4	3	2	1	1	1	1	1	1
80		5	4	3	2	1	1	1	1	1	1
81		5	4	3	2	1	1	1	1	1	1
82		5	4	3	2	1	1	1	1	1	1
83		5	4	3	2	1	1	1	1	1	1
84		5	4	3	2	1	1	1	1	1	1
85		5	4	3	2	1	1	1	1	1	1
86		5	4	3	2	1	1	1	1	1	1
87		5	4	3	2	1	1	1	1	1	1
88		5	4	3	2	1	1	1	1	1	1
89		5	4	3	2	1	1	1	1	1	1
90		5	4	3	2	1	1	1	1	1	1
91		5	4	3	2	1	1	1	1	1	1
92		5	4	3	2	1	1	1	1	1	1
93		5	4	3	2	1	1	1	1	1	1
94		5	4	3	2	1	1	1	1	1	1
95		5	4	3	2	1	1	1	1	1	1
96		5	4	3	2	1	1	1	1	1	1
97		5	4	3	2	1	1	1	1	1	1
98		5	4	3	2	1	1	1	1	1	1
99		5	4	3	2	1	1	1	1	1	1
100		5	4	3	2	1	1	1	1	1	1

Figura 3.5: Planilha com os valores estimados

Na primeira coluna da planilha apresentada na figura (3.5) está a classificação final após uma ordenação. Por exemplo, o quinto colocado foi o sujeito de identificação igual a 16, na tabela 1.1, o Paulo. Nota-se que todas as estimações atendem ao padrão esperado, o item mais difícil (Item 10) teve a menor quantidade de acertos, vários sujeitos acertaram 5 itens: o que teve a melhor nota (561.555) acertou os itens mais fáceis e errou os mais difíceis enquanto o respondente que teve a menor nota (362.963), acertou os itens mais difíceis e errou os mais fáceis, isso é uma incoerência pedagógica. O único empate ocorreu porque os dois sujeitos tiveram as mesmas respostas, logo tem o mesmo traço latente. Resumindo, a TRI é uma modelagem matemática que permite gerar uma proficiência (nota) onde além do número de acertos, é fundamental que exista uma coerência pedagógica nos acertos. Todos os cálculos são feitos com procedimentos matemáticos aplicados das respostas dos sujeitos e não por uma avaliação de quem elaborou o item. Note que em um processo de avaliação educacional, é comum que dois professores façam avaliações diferentes para as características pedagógicas de um item (o que é muito fácil para um, pode não ser para outro), gerando resultados distintos e subjetivos em uma correção. Isso não ocorre em nenhum momento da aplicação da TRI, todos os cálculos foram realizados com base em equações, conforme foi demonstrado neste trabalho.

No Apêndice 1 estão todos os códigos implementados na linguagem de programação Visual Basic e foram aplicados nas planilhas apresentadas neste capítulo. O resultados obtidos com as simulações de 10 itens e 16 sujeitos, apresentados na tabela da figura 3.2, convergiram gerando valores satisfatórios. Entretanto, a aplicação da TRI no ENEM é realizada para 45 itens por área do conhecimento, produzindo uma proficiência, por área, para cada estudante que realiza o exame. Para uma avaliação em larga escala, várias técnicas são utilizadas para tratamento da informação, caso não haja convergência dos valores na aplicação dos algoritmos. Atualmente, existem vários softwares estatísticos, com diversos pacotes desenvolvidos para realizar o tratamento de dados, na aplicação da TRI; alguns com um custo elevado, como exemplo o *BILOG (Analysis of binary response data)* e outros denominados softwares livres, como o Software R, que será utilizado na próxima etapa deste trabalho para aplicar a TRI em um teste simulado ao ENEM, com quatro áreas do conhecimento e 45 itens cada uma, realizado com alunos no ensino médio e pré-vestibulares de Uberlândia.

No próximo capítulo, será realizada uma comparação entre as notas obtidas com as rotinas programadas para aplicação da TRI no software R e as notas determinadas pelos algoritmos desenvolvidos neste capítulo.

# Capítulo 4

## Resultados e Discussões

Para aplicar a metodologia proposta nos capítulos anteriores, foi realizado um teste com 180 itens, sendo 45 itens para cada traço latente: Matemática e suas Tecnologias; Linguagens, Códigos e suas Tecnologias; Ciências da Natureza e suas Tecnologias e Ciências Humanas e suas Tecnologias. Este teste foi aplicado em dois dias; no primeiro, para um total de 811 alunos e no segundo, 668. Foi uma simulação ao que ocorre no ENEM: mesmos horários e itens seguindo recomendações pedagógicas direcionadas pelas diretrizes no INEP. As respostas foram inseridas em um planilha eletrônica e em seguida foram realizados dois procedimentos: o primeiro foi criar a rotina para a utilização da TRI através do software R e o segundo, aplicar os algoritmos desenvolvidos para estimar os parâmetros e a habilidade apresentados no capítulo anterior. Para concluir, foi feita uma comparação entre os resultados obtidos em ambas simulações. Na próxima seção será apresentado os resultados analisados no software R.

### 4.1 Tratamento dos dados no software R

Os comandos utilizados nos aplicativos desenvolvidos estão no Apêndice 2. Adiante tem-se uma explicação das funções utilizadas. Para aplicação a TRI no simulado, o primeiro passo é corrigir as respostas com o gabarito, ou seja, criar a tabela com acerto (1) e erro (0). O segundo passo, é realizar a estimação dos parâmetros através da função *tpm* (three parameter model) do pacote *ltm* (Latent Trait Model). O terceiro passo é estimar as habilidades pelo comando *factore.scores*. O quarto passo é a transformação



das habilidades calculadas para a métrica estabelecida pelo INEP ( $N = 100\theta + 500$ ). Com as notas, criou-se uma classificação para os participantes. Por fim, foi feita a construção da curva característica de cada item, com a função *plot*, mostrada na figura a seguir.

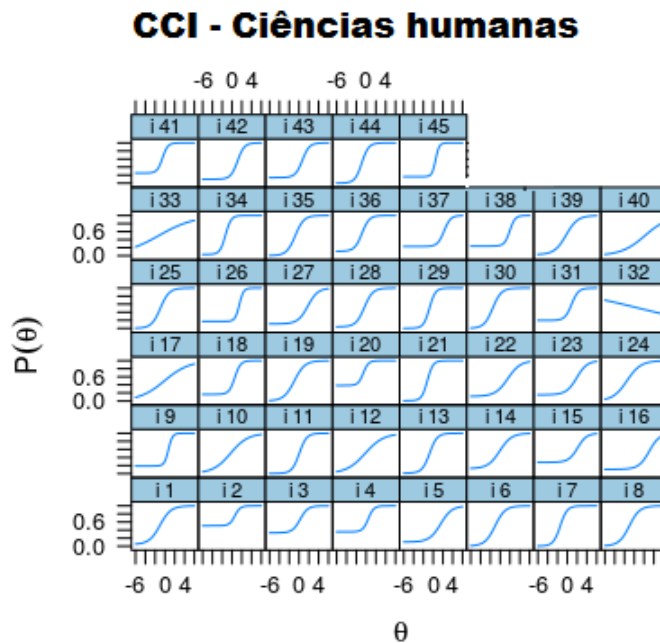


Figura 4.1: Resultados para ciências humanas - software R

Os procedimentos foram criados para determinar a proficiência das quatro áreas, mas por uma questão de simplificação, a análise relatada neste trabalho foi apenas para os 45 primeiros itens do banco de respostas, a saber, Ciências Humanas e suas Tecnologias. Os resultados de todas as áreas foram processados e divulgados para todos os participantes do teste.

Na próxima seção estão os resultados obtidos com o processamento das respostas através dos algoritmos implementados na planilha eletrônica apresentada no capítulo anterior.

## 4.2 Tratamento dos dados nos algoritmos implementados

Os resultados obtidos com a estimação dos parâmetros pelo método da máxima verossimilhança apresentaram as curvas características dos itens apresentados na figura a seguir.

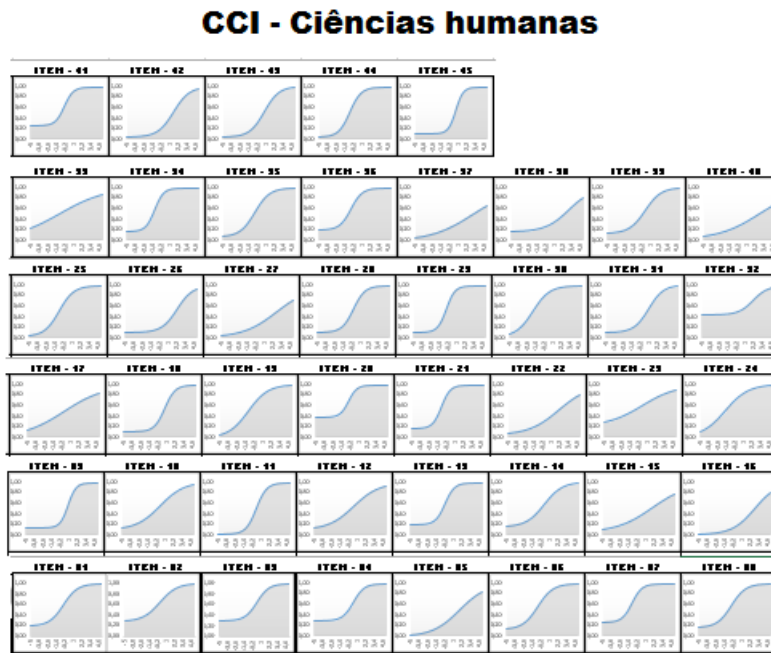


Figura 4.2: Resultados para ciências humanas - planilha eletrônica

Nota-se que diversos itens tem gráficos semelhantes em ambas simulações (Figura 4.1 e 4.2). O método da máxima verossimilhança, contudo, tem alguns problemas, em especial o que Lord (1968) chama de efeito teto e efeito piso. Quando um sujeito acerta quase todos os itens, a estimação de sua aptidão por meio da máxima verossimilhança vai produzir um teta infinito positivo, ou seja,  $\theta = +\infty$ . Este é o efeito teto, assim não há convergência para os valores procurados e para que o algoritmo encerre a rotina foi definido um critério de parada baseado nos valores esperados, conforme exposto na seção 2.2. Por outro lado, quando o sujeito erra muitos itens, novamente a estimação da sua aptidão vai produzir teta infinito, desta vez, negativo, isto é,  $\theta = -\infty$ . Este é

o efeito piso e o procedimento de finalização do algoritmo foi o mesmo.

Para superar estas dificuldades na estimação dos parâmetros por meio da máxima verossimilhança, são utilizados outros métodos de estimação, tais como, o método Bayesiano, o método de Máximo a Posteriori - MAP, o A Posteriori Esperado - EAP, a Máxima Verossimilhança Marginal - MML. Entretanto, estes serão estudos para projetos futuros e não serão abordados neste trabalho de conclusão.

É claro que para o intervalo mediano, onde os padrões de resposta geram a convergência do método, os resultados são muito próximos. Este fato fica exposto na taxa de aproximação (diferença percentual entre a nota obtida no software R e os algoritmos implementados) calculada na última coluna da figura 4.3. Para os exemplos, na faixa entre 24 e 30 acertos, este percentual não chega a 1%. É importante relatar que, a quantidade de sujeitos, com essa propriedade, no banco de respostas completo, é bem maior do que o exposto.

Ao final da aplicação da TRI no simulado, foi enviado para os alunos participantes, um relatório com a nota estimada em cada uma das quatro áreas do conhecimento, junto com a quantidade de acertos obtidos. Foi realizada uma explicação sobre a metodologia para viabilizar a interpretação dos resultados.

As figuras abaixo mostram os resultados das notas finais no software R e nos algoritmos desenvolvidos no Visual Basic. Temos duas tabelas, a primeira obtida com o algoritmos desenvolvidos para o Visual Basic e a segunda com os comandos executados no software R. Em ambas, a primeira coluna indica uma identificação do sujeito, a segunda coluna, o número de acertos e a terceira coluna, a nota final com a TRI. A segunda tabela possui uma coluna a mais que indica a variação percentual entre as duas notas calculadas.

Apresentaremos três faixas de acertos escolhidas dentre os 811 candidatos que fizeram o simulado.

A p a g a	<b>Itens: 45</b>					
	<b>Sujeitos: 811</b>					
	Ordена	Discriminação				
	Início	Dificuldade				
Habilidade		Acase	Ordена	Início		
<b>LTM - SOFTWARE R</b>						
S	Acertos	Nota	S	Acertos.CH	Nota.Final.CH	Dif. Perc.
550	40	858,838	396	41	917,607	6,40%
396	41	855,354	550	40	913,787	6,39%
520	40	845,021	420	40	899,817	6,09%
316	41	844,512	657	41	897,457	5,90%
420	40	843,444	639	41	884,484	4,64%
639	41	826,804	520	40	882,659	6,33%
713	39	826,258	606	41	873,132	5,37%
657	41	825,662	527	39	868,017	4,88%
398	40	824,467	316	41	861,290	4,28%
606	41	821,164	456	39	860,914	4,62%
547	39	818,304	751	39	853,818	4,16%
456	39	818,292	288	36	850,645	3,80%
408	39	817,081	398	40	848,203	3,67%
527	39	816,471	408	39	846,732	3,57%
751	39	812,776	291	39	839,503	3,18%

Figura 4.3: Comparação entre resultados para ciências humanas - faixa de 36 até 41 acertos

No Visual Basic, o indivíduo de identificação 550 ficou em primeiro lugar enquanto o número 396 em segundo. No software R, ocorreu a inversão destas posições. As variações percentuais são pequenas, entretanto podemos notar que várias pessoas com o mesmo número de acertos tiveram notas diferentes, fato justificado pelo padrão de respostas de cada sujeito, conforme propõe a TRI.

A p a g a	Itens: 45					
	Sujeitos: 811					
	Ordena	Discriminação				
	Início	Dificuldade				
Habilidade		Acaso				
			Ordena	Início		
<b>LTM - SOFTWARE R</b>						
S	Acertos	Nota	S	Acertos.CH	Nota.Final.CH	Dif. Perc.
185	25	499,470	21	25	500,679	0,24%
797	25	499,439	387	29	500,404	0,19%
197	23	498,753	273	25	499,976	0,24%
34	24	498,302	596	23	499,928	0,33%
51	24	497,997	75	24	499,901	0,38%
776	23	497,745	720	22	499,814	0,41%
602	24	497,530	170	23	499,341	0,36%
217	22	497,239	16	25	499,114	0,38%
254	22	497,200	776	23	499,062	0,37%
170	23	497,055	626	22	498,349	0,26%
209	22	496,826	243	23	498,300	0,30%
693	23	496,572	133	23	498,218	0,33%
389	23	496,464	97	22	496,800	0,07%
97	22	496,042	797	25	496,175	0,03%
196	22	495,910	46	24	495,926	0,00%
725	22	495,795	163	26	495,428	-0,07%
96	24	495,674	436	25	495,247	-0,09%
204	24	495,662	226	24	495,172	-0,10%
213	21	495,486	804	22	494,990	-0,10%
232	23	495,426	576	21	494,885	-0,11%
709	22	495,299	166	27	494,832	-0,09%
620	23	494,549	23	17	493,948	-0,12%

Figura 4.4: Comparação entre resultados para ciências humanas - faixa de 14 até 21 acertos

Para a faixa de acertos apresentada na figura 4.4, podemos perceber que a diferença entre as notas obtidas pela rotina implementada para o software R e os algoritmos desenvolvidos no Visual Basic, diminui para um valor absoluto menor que 1%. Na faixa anterior, mostrada na figura 4.3, a diferença não atingiu o valor, em módulo, de

7%. Quanto mais próximos dos extremos forem os padrões de respostas, maior será a diferenças entre resultados obtidos no algoritmo e o software R.

Um fato considerável é que essa diferença vai interferir na colocação do candidato, por exemplo, o candidato com identificação 97, acertou 22 itens e teve a nota 496,042 no Visual Basic e 496,800 no software R, gerando uma diferença de posições na classificação geral do traço latente de Ciências Humanas e suas Tecnologias. A justificativa será abordada no final deste capítulo.

A p a g a		Itens: 45		Sujeitos: 811		Ordena		Início	
		Ordena	Discriminação						
		Início	Dificuldade						
		Habilidade	Acaso						
				LTM - SOFTWARE R					
S	Acertos	Nota		S	Acertos.CH	Nota.Final.CH	Dif. Perc.		
484	11	278,589		585	13	236,450	-17,82%		
195	10	276,366		670	8	222,416	-24,26%		
346	8	274,357		419	10	220,345	-24,51%		
198	12	271,352		29	8	199,737	-35,85%		
627	10	265,838		702	8	193,460	-37,41%		
182	9	263,876		485	3	188,499	-39,99%		
702	8	259,359		591	11	185,333	-39,94%		
670	8	258,270		5	10	180,087	-43,41%		
5	10	249,291		724	12	173,284	-43,86%		
375	9	230,235		246	9	169,138	-36,12%		
307	12	224,668		179	16	164,196	-36,83%		
649	10	210,585		307	12	164,160	-28,28%		
557	10	201,504		653	12	161,051	-25,12%		
64	10	198,343		64	10	160,359	-23,69%		
246	9	182,801		557	10	159,084	-14,91%		
383	6	164,088		546	14	157,371	-4,27%		
716	10	163,342		182	9	152,080	-7,41%		
144	12	158,574		55	11	149,918	-5,77%		
81	11	152,433		605	11	140,327	-8,63%		
594	7	150,686		198	12	135,497	-11,21%		
653	12	143,722		102	10	134,755	-6,65%		
643	3	143,474		195	10	119,615	-19,95%		

Figura 4.5: Comparação entre resultados para ciências humanas - faixa de 3 até 16 acertos

Para estes sujeitos a diferença percentual foi maior devido ao padrão de respostas ser composto por poucos itens corretos, dificultando a convergência dos valores calculados nas iterações executadas nos algoritmos. Apesar de não haver convergência para padrões de respostas extremos (acertar ou errar quase todos), as aproximações foram consideradas satisfatórias, ressaltando que o objetivo do trabalho é a compreensão da metodologia aplicada na TRI.

Para superar estas dificuldades da estimação de parâmetros por meio da máxima verossimilhança, são utilizados outros métodos de estimação, tais como, o método Bayesiano, o método de Máximo a Posteriori (MAP, maximum a posteriori), o A Posteriori Esperado (EAP, expected a posteriori) e a Máxima Verossimilhança Marginal (MML). O EAP é utilizado pelas bibliotecas do software R enquanto o método utilizado do Visual Basic foi o MMV, justificando as diferenças encontradas.

No próximo capítulo, teremos a conclusão do trabalho.

# Capítulo 5

## Conclusão

O processo de ensino e aprendizagem é dinâmico e precisa de constantes mudanças para se adaptar às diversas realidades nos quais ele se insere. A avaliação torna-se uma etapa fundamental para a melhoria na qualidade da educação. A Teoria de Resposta ao Item mostra que um teste composto por vários itens, pode ser analisado sob uma perspectiva muito mais eficiente e fidedigna do que a proposta na Teoria Clássica dos Testes. Com o uso de ferramentas computacionais, a metodologia fundamentada na determinação de um traço latente se mostra uma excelente ferramenta pedagógica para avaliar o processo educacional.

Compreender a teoria que estrutura a TRI é uma tarefa difícil e complexa. Difundir tal proposta revela-se ainda mais desafiador. A tarefa pode ser árdua mas deve ser feita, pois ela propõe um nota fundamentada em uma coerência pedagógica e não apenas no número de acertos, algo muito sensato no processo de ensino aprendizagem. Apresentar, com dados concretos para os estudantes do ensino médio, o resultado da execução dos procedimentos da TRI, causou sentimentos variados: indagações de como esse processo pode ser tão preciso ao ponto de detectar um chute; despertou a necessidade de uma estratégia de resolução de provas buscando um padrão de respostas mais eficiente; e o mais importante, a confiança no processo de avaliação em que eles estão inseridos; com a TRI eles perceberam que é possível calcular o nível de competência que um participante possui para resolver situações problema propostas nos itens.

Implementar cada um dos algoritmos foi muito importante para entender como funciona a TRI, e principalmente, para pensar em novos projetos para a divulgação da metodologia nas avaliações educacionais.





Next j

%% ' ALGORITMO DE

'NEWTON-RAPHSON PARA TODOS SUJEITOS

For j = 1 To totalIndividuos

h(j) = Cells(8 + j, 5)

dr1 = 0

dr2 = 0

E = 1

While (Abs(E) > 0.00000001)

dr1 = 0

dr2 = 0

For i = 1 To totalItens

P = pr(3, i) + (1 - pr(3, i)) /

(1 + Exp(-D \* pr(1, i) \* (h(j) - pr(2, i))))

T = (r(j, i) - P) / (P \* (1 - P))

D1P = pr(1, i) \* (1 - pr(3, i)) \* D \* Exp(-D \* pr(1, i) \*

(h(j) - pr(2, i)))/(1 + Exp(-D \* pr(1, i) \*

(h(j) - pr(2, i)))) ^ 2

D2P = 2 \* pr(1, i) ^ 2 \* (1 - pr(3, i)) \* (D ^ 2) \*

Exp(-2 \* D \* pr(1, i) \* (h(j) - pr(2, i)))

/ (1 + Exp(-D \* pr(1, i) \* (h(j) - pr(2, i)))) ^ 3

- pr(1, i) ^ 2 \* (1 - pr(3, i)) \* (D ^ 2) \*

Exp(-D \* pr(1, i) \* (h(j) - pr(2, i)))

/ (1 + Exp(-D \* pr(1, i) \* (h(j) - pr(2, i)))) ^ 2

dr1 = dr1 + T \* D1P

```

dr2 = dr2 + T * D2P - (T * D1P) ^ 2
Next i
E = dr1 / dr2
h(j) = h(j) - E
hAnterior = h(j)
If h(j) > 3 Then
E = 0.0000000001          ' PARADA PARA CASO DE DIVERGENCIA
h(j) = hAnterior
End If
If h(j) < -3 Then
E = 0.0000000001
h(j) = hAnterior
End If
Wend
Cells(8 + j, 5) = h(j)
Cells(8 + j, 6) = 100 * h(j) + 500
Next j
End Sub

```

Rotina para estimar a discriminação.

```

Sub estimaDiscriminacao()
Dim pr(3, 180)
Dim h(10000)
Dim r(10000, 180)
D = Cells(7, 5)
%
```

```

totalItens = Cells(3, 6)      ' LEITURAS INICIAIS:
totalIndividuos = Cells(4, 6)' - TOTAL DE ITENS
For i = 1 To totalItens      ' - TOTAL DE SUJEITOS
pr(1, i) = Cells(5, 7 + i)   ' - PARÂMETROS INICIAIS
pr(2, i) = Cells(6, 7 + i)
pr(3, i) = Cells(7, 7 + i)
Next i

For j = 1 To totalIndividuos
For i = 1 To totalItens
r(j, i) = Cells(8 + j, 7 + i)
Next i
Next j

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
For i = 1 To totalItens      'NEWTON-RAPHSON PARA CADA ITEM
dr1 = 0
dr2 = 0
E = 1
While (Abs(E) > 0.00000001)
dr1 = 0
dr2 = 0
For j = 1 To totalIndividuos
h(j) = Cells(8 + j, 5)
P = pr(3, i) + (1 - pr(3, i)) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i))))
T = (r(j, i) - P) / (P * (1 - P))

```

```

D1P = -(1 - pr(3, i)) * (pr(2, i) - h(j)) * D *
Exp(-D * pr(1, i) * (h(j) - pr(2, i))) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i)))) ^ 2
D2P = 2 * (pr(2, i) - h(j)) ^ 2 * (1 - pr(3, i)) * (D ^ 2) *
Exp(-2 * D * pr(1, i) * (h(j) - pr(2, i))) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i)))) ^ 3 -
(pr(2, i) - h(j)) ^ 2 * (1 - pr(3, i)) * (D ^ 2) *
Exp(-D * pr(1, i) * (h(j) - pr(2, i))) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i)))) ^ 2
dr1 = dr1 + T * D1P
dr2 = dr2 + T * D2P - (T * D1P) ^ 2
Next j
E = dr1 / dr2
aAnterior = pr(1, i)
pr(1, i) = pr(1, i) - E
If pr(1, i) > 2 Then
E = 0.0000000001 ' PARADA PARA CASO DE DIVERGENCIA
pr(1, i) = aAnterior
End If
If pr(1, i) < 0 Then
E = 0.0000000001
pr(1, i) = aAnterior
End If
Wend
With Sheets("Parametros")

```

```

.Cells(1 + i, 4) = pr(1, i)
End With
Next i
End Sub

```

Rotina para estimar a dificuldade de cada item.

```

Sub estimaDificuldade()
Dim pr(3, 180)
Dim h(10000)
Dim r(10000, 180)
D = Cells(7, 5)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
totalItens = Cells(3, 6)      ' LEITURAS INICIAIS:
totalIndividuos = Cells(4, 6)' - TOTAL DE ITENS
For i = 1 To totalItens      ' - TOTAL DE SUJEITOS
pr(1, i) = Cells(5, 7 + i)  ' - PARÂMETROS INICIAIS
pr(2, i) = Cells(6, 7 + i)
pr(3, i) = Cells(7, 7 + i)
Next i
For j = 1 To totalIndividuos
For i = 1 To totalItens
r(j, i) = Cells(8 + j, 7 + i)
Next i
Next j
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
For i = 1 To totalItens      'NEWTON-RAPHSON PARA CADA ITEM

```

```

dr1 = 0
dr2 = 0
E = 1
While (Abs(E) > 0.00000001)
dr1 = 0
dr2 = 0
For j = 1 To totalIndividuos
h(j) = Cells(8 + j, 5)
P = pr(3, i) + (1 - pr(3, i)) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i))))
T = (r(j, i) - P) / (P * (1 - P))
D1P = -pr(1, i) * (1 - pr(3, i)) * D *
Exp(-D * pr(1, i) * (h(j) - pr(2, i))) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i)))) ^ 2
D2P = 2 * pr(1, i) ^ 2 * (1 - pr(3, i)) * (D ^ 2) *
Exp(-2 * D * pr(1, i) * (h(j) - pr(2, i))) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i)))) ^ 3 - pr(1, i) ^ 2 *
(1 - pr(3, i)) * (D ^ 2) * Exp(-D * pr(1, i) * (h(j) - pr(2, i)))
/ (1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i)))) ^ 2
dr1 = dr1 + T * D1P
dr2 = dr2 + T * D2P - (T * D1P) ^ 2
Next j
E = dr1 / dr2
bAnterior = pr(2, i)
pr(2, i) = pr(2, i) - E

```





```

pr(2, i) = Cells(6, 7 + i)
pr(3, i) = Cells(7, 7 + i)
Next i
For j = 1 To totalIndividuos
For i = 1 To totalItens
r(j, i) = Cells(8 + j, 7 + i)
Next i
Next j
For i = 1 To totalItens   'NEWTON-RAPHSON PARA CADA ITEM
dr1 = 0
dr2 = 0
E = 1
While (Abs(E) > 0.00000001)
dr1 = 0
dr2 = 0
For j = 1 To totalIndividuos
h(j) = Cells(8 + j, 5)
P = pr(3, i) + (1 - pr(3, i)) /
(1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i))))
T = (r(j, i) - P) / (P * (1 - P))
D1P = 1 - 1 / (1 + Exp(-D * pr(1, i) * (h(j) - pr(2, i))))
D2P = 0
dr1 = dr1 + T * D1P
dr2 = dr2 + T * D2P - (T * D1P) ^ 2
Next j

```

```

E = dr1 / dr2
cAnterior = pr(3, i)
pr(3, i) = pr(3, i) - E
If pr(3, i) > 1 Then
E = 0.0000000001 ' PARADA PARA CASO DE DIVERGENCIA
pr(3, i) = cAnterior
End If
If pr(3, i) < 0 Then
E = 0.0000000001
pr(3, i) = cAnterior
End If
Wend
With Sheets("Parametros")
. Cells(1 + i, 2) = pr(3, i)
End With
Next i
End Sub

```

## Apêndice 2

Os códigos apresentados foram desenvolvidos para o software R. As etapas do processo de aplicação da TRI estão descritas adiante.

Inicialmente foi feita uma correção das respostas dos estudantes criando uma planilha eletrônica do tipo "xlsx" com as respostas binárias ("1" acerto e "0" erro). Em seguida foram desenvolvidas as rotinas para:

Importar as respostas para o software R:

```
## ENTRADA COM AS RESPOSTAS DOS ESTUDANTES
##      QUE PARTICIPARAM DO SIMULADO
primeiroDia = read.xlsx("PrimeiroDia.xlsx", sheetName = "todos")
segundoDia = read.xlsx("SegundoDia.xlsx", sheetName = "todos")
CH = primeiroDia[,c(2:46)]
CN = primeiroDia[,c(47:91)]
LC = segundoDia[,c(2:46)]
MA = segundoDia[,c(47:91)]
colnames(CH) = c(paste("i", 1:45))
colnames(CN) = c(paste("i", 1:45))
colnames(LC) = c(paste("i", 1:45))
colnames(MA) = c(paste("i", 1:45))
```

Estimar os parâmetros de todos os itens:

```
# ESTIMAÇÃO DOS PARAMETROS
parametros.CH = tpm(CH)
parametros.CN = tpm(CN)
parametros.LC = tpm(LC)
parametros.MA = tpm(MA)
```

```

write.xlsx(coef(parametros.CH), file = "/parametrosCH.xlsx")
write.xlsx(coef(parametros.CN), file = "/parametrosCN.xlsx")
write.xlsx(coef(parametros.LC), file = "/parametrosCL.xlsx")
write.xlsx(coef(parametros.MA), file = "/parametrosMA.xlsx")

```

Estimar cada traço latente: Ciências Humanas, Ciências da Natureza, Linguagens e Códigos e Matemática.

```

# DETERMINAÇÃO DO ESCORE DE CADA PADRÃO DE RESPOSTAS
theta.p.CH = factor.scores(parametros.CH,
                           method = "EAP", resp.patterns = CH)
theta.p.CN = factor.scores(parametros.CN,
                           method = "EAP", resp.patterns = CN)
theta.p.LC = factor.scores(parametros.LC,
                           method = "EAP", resp.patterns = LC)
theta.p.MA = factor.scores(parametros.MA,
                           method = "EAP", resp.patterns = MA)

```

Determinação da escala de proficiência e classificação geral.

```

notas.CH = NA
notas.CN = NA
notas.LC = NA
notas.MA = NA
for (i in 1:NROW(theta.p.CH$score.dat$z1)){
notas.CH[i] = 100*theta.p.CH$score.dat$z1[i]+500
notas.CN[i] = 100*theta.p.CN$score.dat$z1[i]+500
}

```

```

for (i in 1:NROW(theta.p.LC$score.dat$z1)){
notas.LC[i] = 100*theta.p.LC$score.dat$z1[i]+500
notas.MA[i] = 100*theta.p.MA$score.dat$z1[i]+500
}
# CLASSIFICAO GERAL
classificacao.CH = sort.list(theta.p.CH$score.dat$z1,
                             decreasing = T)
classificacao.CN = sort.list(theta.p.CN$score.dat$z1,
                             decreasing = T)
classificacao.LC = sort.list(theta.p.LC$score.dat$z1,
                             decreasing = T)
classificacao.MA = sort.list(theta.p.MA$score.dat$z1,
                             decreasing = T)

posicao.CH =NA
posicao.CN =NA
posicao.LC =NA
posicao.MA =NA
for (i in 1:NROW(classificacao.CH)){
posicao.CH[classificacao.CH[i]] = i
posicao.CN[classificacao.CN[i]] = i
}
for (i in 1:NROW(classificacao.LC)){
posicao.LC[classificacao.LC[i]] = i
posicao.MA[classificacao.MA[i]] = i
}

```

```

}
geral.CH = data.frame("Acertos CH" = acertos.CH,
"Posição CH" = posicao.CH , "Nota Final CH" = notas.CH)
geral.CN = data.frame("Acertos CN" = acertos.CN,
"Posição CN" = posicao.CN , "Nota Final CN" = notas.CN)
geral.LC = data.frame("Acertos LC" = acertos.LC,
"Posição LC" = posicao.LC , "Nota Final LC" = notas.LC)
geral.MA = data.frame("Acertos MA" = acertos.MA,
"Posição MA" = posicao.MA , "Nota Final MA" = notas.MA)

write.xlsx (geral.CH , file = "/ResultadoGeral.CH.xlsx")
write.xlsx (geral.CN , file = "/ResultadoGeral.CN.xlsx")
write.xlsx (geral.LC , file = "/ResultadoGeral.LC.xlsx")
write.xlsx (geral.MA , file = "/ResultadoGeral.MA.xlsx")

```

# Referências Bibliográficas

- Baker, F. B. (2001). *The basics of item response theory*.
- Fletcher, P. R. (1994). *A Teoria da Resposta ao Item: medidas invariantes do desempenho escolar*. revista da Fundação Cesgranrio Rio de Janeiro.
- Kent, N. (2012). *Equações Diferenciais*.
- Lord, F. M. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Pasquali, L. (2007). *Teoria de resposta ao Item Procedimentos e Aplicações*.
- Perrenoud, P. (1999). *Avaliação: da excelência à regulação das aprendizagens-entre duas lógicas*.
- Rabelo, M. (2011). *Análise comparativa dos processos de avaliação educacional em larga escala*.
- Silva Carvalho, E. (2001). A concepção de educação profissional no exame nacional de cursos.
- Stevenson, J. (1981). *Estatística*.
- Thurstone, L. (1959). *The measurement of values*. Chicago University Press.
- Vianna, H. M. (2014). *Estudos em avaliação educacional*.