

UNIVERSIDADE FEDERAL DO MARANHÃO - UFMA  
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS - CCET  
DEPARTAMENTO DE MATEMÁTICA  
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL -  
PROFMAT

EUGÊNIO MENDES MORAES FILHO

A Teoria da Resposta ao Item na Avaliação Educacional do  
Ensino Básico

São Luís  
2014

EUGÊNIO MENDES MORAES FILHO

A Teoria da Resposta ao Item na Avaliação Educacional do  
Ensino Básico

Dissertação apresentada ao PROFMAT/ Universidade Federal do Maranhão como requisito parcial para a obtenção do grau de Mestre em Matemática.

**Orientador: Prof. Josenildo de Souza Chaves**

**Doutor em Estatística**

São Luís

2014

Filho, Eugênio Mendes Moraes.

A Teoria da Resposta ao Item na Avaliação Educacional do Ensino Básico / Eugênio Mendes Moraes. Filho - 2014

50.p

Dissertação (Mestrado) - PROFMAT, Universidade Federal do Maranhão, 2014.

Orientador: Prof. Josenildo de Souza Chaves

1. Teoria da resposta ao item. 2. Avaliação educacional - Aspecto qualitativo. 3. Curva característica.. I. Título.

CDU 519.2:37091.26

EUGÊNIO MENDES MORAES FILHO

A Teoria da Resposta ao Item na Avaliação Educacional do  
Ensino Básico

Dissertação apresentada ao PROFMAT/ Universidade Federal do Maranhão como requisito parcial para a obtenção do grau de Mestre em Matemática.

Aprovado em

**BANCA EXAMINADORA**

---

Prof. Josenildo de Souza Chaves

Doutor em Estatística

---

Prof. José Antônio Pires Ferreira Marão

Doutor em Física

---

Prof. Mario Tanaka Filho

Doutor em Engenharia Mecânica

*À minha família...*

## **AGRADECIMENTOS**

Primeiramente a Deus por ter me guiado nessa caminhada.

À minha família pelo apoio em todos os momentos, em especial à minha esposa Quézia pelo carinho e compreensão sempre presentes.

Ao PROFMAT pela oportunidade de aperfeiçoamento e crescimento profissional.

Aos professores do curso pelo comprometimento, em especial a meu orientador Prof. Dr. Josenildo Souza pela singular ajuda e dedicação.

*“Conheça todas as teorias, domine todas as técnicas, mas ao tocar uma alma humana seja apenas outra alma humana”.*

*Carl G. Jung*

# Sumário

<b>Lista de Figuras</b> . . . . .	<b>iii</b>
<b>Resumo</b> . . . . .	<b>iv</b>
<b>Abstract</b> . . . . .	<b>v</b>
<b>1 Introdução</b> . . . . .	<b>1</b>
1.0.1 Organização do Trabalho . . . . .	4
<b>2 Curva Característica do Item</b> . . . . .	<b>5</b>
2.1 Modelos Matemáticos em TRI . . . . .	9
2.1.1 Modelo de Rasch ou de um Parâmetro . . . . .	10
2.1.2 Modelo Logístico de Dois Parâmetros . . . . .	12
2.1.3 Modelo Logístico de Três Parâmetros . . . . .	13
2.1.4 Parâmetros da Curva Característica do Item . . . . .	15
<b>3 Estimação dos Parâmetros</b> . . . . .	<b>18</b>
3.0.5 Estimação dos Parâmetros dos Itens pelo Método da Máxima Verossimilhança . . . . .	22
3.1 Estimação dos Parâmetros Considerando o Agrupamento das Habi- dades . . . . .	27
<b>4 Aplicações</b> . . . . .	<b>30</b>

4.1	Utilização do R na Estimação de Parâmetros em Dados do SARESP .	30
4.2	Aplicação do Modelo de TRI com 3 Parâmetros . . . . .	31
4.3	Leitura do Arquivo . . . . .	31
4.3.1	Análise dos Dados Utilizando o Pacote Irtoys . . . . .	37
4.4	Contruindo o Ranking dos Examinados . . . . .	41
<b>5</b>	<b>Considerações Finais . . . . .</b>	<b>45</b>
	<b>Referências . . . . .</b>	<b>47</b>

# Lista de Figuras

2.1	Exemplo de curva característica do item . . . . .	6
2.2	Curvas características com diferentes níveis de dificuldade. . . . .	8
2.3	Curvas características com diferentes níveis de discriminação. . . . .	8
2.4	Curvas características para 4 itens utilizando o ML1. . . . .	12
2.5	Curvas características para 4 itens utilizando o ML2. . . . .	14
2.6	Curvas características para 4 itens utilizando o ML3. . . . .	15
2.7	Curva característica do item e seus parâmetros. . . . .	15
4.1	Função de informação do item para os dados do SARESP 2007 . . . . .	38
4.2	Curva característica do item para os dados do SARESP 2007 . . . . .	39
4.3	Função de informação do teste para os dados do SARESP 2007. . . . .	40

# Resumo

A importância da avaliação educacional tem sido evidenciada desde a última década. Cada vez mais se faz necessária uma abordagem que mostre o aspecto qualitativo de dados educacionais em detrimento de aspectos puramente quantitativos. Nesse contexto surge a Teoria da Resposta ao Item (TRI). No presente trabalho abordamos os fundamentos dessa teoria, com ênfase nos modelos logísticos que relacionam a probabilidade de acerto a determinados itens com a habilidade dos respondentes. Além disso, apresentamos o processo de estimação dos parâmetros dos modelos e das habilidades dos alunos. Uma aplicação com os dados do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) foi realizada. A implementação computacional foi desenvolvida com o uso do software R.

**Palavras-chave:** Avaliação Educacional, Teoria da Resposta ao Item (TRI), Modelo Logístico.

# Abstract

The importance of educational evaluation has been increased since the last decade. Increasingly an approach is necessary to show the qualitative aspect of educational data in detriment to purely quantitative aspects. In this context, the Item Response Theory (IRT) arises. In this academic work we will discuss the fundamentals of this theory, with emphasis on logistic models that correlate probability of success to particular items with the ability of the respondents. In addition, we present the process of estimating of the parameter's models and of student's abilities. An application with data from the Evaluation of Educational Achievement of the São Paulo State (SARESP) system is performed. The computational implementation was developed using the software R.

**Keywords:** Educational Evaluation, Item Response Theory (IRT), Logistic Model .

# Capítulo 1

## Introdução

A avaliação educacional é um dos pilares do processo ensino aprendizagem, através dela é possível comparar o desempenho dos alunos ao longo dos anos, e observar se houve progresso ou não em determinados aspectos de interesse. Além disso uma avaliação educacional bem construída permite-nos atestar a qualidade do ensino, e verificar o desenvolvimento dos alunos em seu contexto de aprendizagem, com vistas a planejar ações futuras que desenvolvam neles essas habilidades básicas (Vendramini, 2005). Além disso, o educador pode aumentar a sua eficiência no trabalho desde que esteja munido de uma ferramenta que lhe permita realizar uma verificação criteriosa dos seus procedimentos de ensino e dos resultados alcançados. O acompanhamento da evolução no desempenho dos alunos indica não somente a realidade dos resultados das metodologias aplicadas, como também determina um conjunto de referências que permitem revisões e intervenções nos procedimentos do educador, bem como planejamento de atividades posteriores.

Ao aplicarmos uma avaliação de aprendizagem, elaboramos uma prova ou teste, e a partir dos resultados obtidos, classificamos os alunos dentro de uma escala que, geralmente, leva em consideração somente a quantidade de questões respondidas corretamente pelos mesmos. No entanto há de se observar que os resultados encontrados irão depender, tanto das questões aplicadas, quanto da amostra de respondentes a qual foi aplicada tal avaliação. Essa abordagem é comumente utilizada na Teoria Clássica dos Testes (TCT), que dessa forma não leva em consideração a

habilidade do indivíduo, apenas verifica seu score final no teste. Entretanto, um estudante poder ser descrito como brilhante ou mediano, em determinados aspectos que dificilmente podem ser medidos por uma avaliação que leva somente em consideração o score obtido, uma vez que certas qualidades de interesse são perceptíveis apenas intuitivamente. Qualidades ou habilidades escolares tais como: facilidade de leitura, saber relacionar várias fontes de informação, habilidades aritméticas, são aspectos dos estudantes que não podem ser medidas diretamente, cada um destes aspectos são denominados pela psicométrica de aspectos não observáveis ou traços latentes. Segundo Francisco (2005), pode-se entender por traço latente as diferentes modalidades estruturais da inteligência que compreendem determinadas operações que o indivíduo utiliza para estabelecer relações entre os objetos físicos, os conceitos, situações, fenômenos e pessoas. As habilidades instrumentais referem-se especificamente ao plano do saber fazer e decorrem, diretamente, do nível estrutural das competências já adquiridas e que se transformam em habilidades. Isto é, a "capacidade de agir eficazmente em um determinado tipo de situação, apoiando-se em conhecimentos, mas sem se limitar a eles" (Perrenoud, 1998). A técnica denominada Teoria da Resposta ao Item (TRI) assume que o desempenho em um teste pode ser explicado em função desses aspectos ou traços latentes. Passaremos a nos referir a essas variáveis com os termos proficiências ou habilidades.

Há uma preocupação crescente com habilidades pouco desenvolvidas de leitura, escrita e aritmética que motiva a busca por testes de competência e habilidades básicas como um meio de certificar a obtenção de uma competência mínima (Anastasi, 2000). De acordo com Andrade (2000), a Teoria da Resposta ao Item (TRI) é uma metodologia que sugere formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item e seus traços latentes. A Teoria da Resposta ao Item foi desenvolvida com propósito de resolver um problema na Teoria Clássica dos Testes que é a dependência da medida de proficiência em relação ao teste aplicado e dos parâmetros dos itens em relação ao conjunto dos respondentes. Segundo a Teoria da Resposta ao Item, a medida de habilidade de um examinado não depende dos itens apresentados a ele, e os parâmetros de discriminação e de dificuldade de cada item não dependem do grupo de respondentes. Na área

educacional é crescente o interesse pela aplicação de técnicas derivadas da TRI, já que esta metodologia visa um tipo de avaliação mais qualitativo, baseada em escalas de habilidades que permite posicionar os estudantes em tal escala de acordo com o nível de proficiência atestado pelo mesmo.

Um outro aspecto que deve ser citado, e que diferencia bastante a Teoria Clássica dos Testes da TRI, é que na Teoria Clássica, pelo fato de a análise psicométrica estar construída sobre o objeto de medição (prova), torna-se quase inviável a comparação entre populações submetidas a provas diferentes, ou pelo menos ao que se denomina formas paralelas de teste. No caso da Teoria da Resposta ao Item, os itens em si, tornam-se os elementos centrais da análise, o que permite comparar populações submetidas a provas distintas, desde que haja apenas alguns elementos, ou itens, em comum. Com isso é também possível avaliar o desenvolvimento de uma determinada série de um ano para outro, ou ainda comparar o desempenho entre escolas públicas ou particulares. Desta forma, várias questões de interesse prático podem ser respondidas de forma mais coerente por meio dessa abordagem.

Segundo Vendramini (2005), há dois pressupostos a serem observados para que possamos aplicar os modelos propostos pela Teoria da Resposta ao Item: o critério de unidimensionalidade (os itens devem medir uma única habilidade); e o critério da independência local (as respostas dos itens não devem ser influenciadas pelas respostas fornecidas a outros itens). Com relação à unidimensionalidade, em geral procura-se reunir itens para os quais supõe-se que uma certa proficiência ou habilidade é dominante. O pressuposto da independência local significa que quando a habilidade latente que está sendo mensurada pelo conjunto de itens é mantida constante, nenhum par de itens pode ser correlacionado. Este pressuposto implica que a probabilidade de acertar um item depende exclusivamente da habilidade do examinado, não da ordem do item dentro da prova, do cansaço, etc. (Lord, 1980).

A TRI fornece modelos que atribuem parâmetros para itens e para indivíduos separadamente de forma a predizer probabilisticamente a resposta de qualquer indivíduo a qualquer item. Requena (1990) ressalta que as funções de resposta ao item estabelecem as relações, matematicamente formalizadas, de como cada resposta depende de certo nível ou grau de habilidade no traço considerado. Geralmente os

itens podem ser avaliados por meio de modelos de um, dois ou três parâmetros. O modelo de um parâmetro envolve apenas a "dificuldade"(parâmetro  $b$ ); o de dois, envolve além da dificuldade a "discriminação"(parâmetro  $a$ ); e o de três, além dos dois já citados,  $a$  e  $b$ , leva em consideração a probabilidade de "acerto ao acaso"(parâmetro  $c$ ). O parâmetro  $\theta$  representa a habilidade dos examinados.

O objetivo principal deste trabalho é apresentar os conceitos fundamentais da Teoria da Resposta ao Item, e avaliar a utilização deste método como alternativa ao que é feito na Teoria Clássica dos Testes. Em particular, utilizar a metodologia apresentada para analisar dados de rendimento escolar de alunos do Ensino Médio.

### 1.0.1 Organização do Trabalho

Este trabalho está organizado em cinco capítulos. No Capítulo 2 são apresentados os fundamentos da Teoria da Resposta ao Item, com ênfase na estrutura da curva característica do item, seus parâmetros e os modelos matemáticos utilizados para adequação dos dados à curva, bem como, os três modelos logísticos de 1, 2 e 3 parâmetros. Por fim, fazemos uma breve discussão de como os parâmetros da curva característica do item modelam seu traçado e a resultante prática dessas variações em um teste.

No Capítulo 3 são abordados os procedimentos de estimação, tanto dos parâmetros dos itens da CCI quanto das habilidades dos respondentes. Abordamos também o processo de estimação por máxima verossimilhança e discutimos algumas dificuldades na aplicação destes procedimentos.

No Capítulo 4 mostramos uma aplicação dos procedimentos utilizados na TRI fazendo uso de dados da Secretaria de Rendimento Escolar do estado de São Paulo com o uso do software R (R Core Team, 2013), tanto para estimação quanto para construção das curvas características dos itens e outras curvas que compõem a teoria.

No Capítulo 5 apresentamos as considerações finais.

## Capítulo 2

# Curva Característica do Item

Neste capítulo serão apresentados os modelos utilizados em Teoria de Resposta ao Item. Desde o modelo de um parâmetro até o modelo de 3 parâmetros que utilizaremos para aplicações no Capítulo 4.

De acordo com Hambleton & Rogers (1991), há dois pressupostos relacionados com os modelos da TRI: a estrutura matemática da Curva Característica do Item e a estrutura dimensional dos dados do teste. Segundo Baker (2001), a Curva Característica do Item é o "bloco de construção" da TRI, todo restante do desenvolvimento dessa teoria depende do comportamento desta curva.

Inicialmente, precisamos entender que para medir a quantidade de traços latentes de uma pessoa é necessário uma escala de medida, ou seja, uma escala com uma certa métrica (Baker, 2001). Há, todavia, uma grande dificuldade técnica em definir tal escala, assim como o posicionamento dos números na escala e a quantidade de traços latentes envolvidos. Assume-se portanto, para abordagem teórica dos conceitos iniciais da TRI uma escala arbitrária, levando-se em consideração apenas que tal escala de habilidade terá centro em zero, e uma métrica que variará entre -2 a 2 a -6 a 6, ainda que o intervalo desta escala possa variar, teoricamente, entre  $(-\infty, +\infty)$ .

O enfoque usual para se medir a quantidade de traços latentes ou habilidades de uma pessoa é desenvolver um teste consistindo de um número de itens (questões). Cada um desses itens irá medir um aspecto particular da habilidade em questão. A

princípio esses testes podem tanto possuir questões de resposta livre, quanto questões de múltipla escolha, desde que os itens de resposta livre possam ser corrigidos de forma dicotômica. Em termos práticos, no entanto, pela dificuldade de se criar uma escala de fácil manejo para avaliações deste tipo, principalmente devido ao fator da subjetividade, que influencia na correção da prova, a grande maioria dos testes desenvolvidos em TRI constituem-se de itens de múltipla escolha, sendo que cada resposta correta recebe o valor 1 e as demais alternativas recebem o valor 0, e por esse motivo, esses itens são geralmente denominados itens binários.

Segundo Baker (1992), um pressuposto lógico é que cada examinado em um teste de itens possua uma certa quantidade de habilidades mínimas, que fará com que o mesmo possa ser situado dentro da escala de habilidades. Esse valor será representado pela letra grega  $\theta$ . Em cada nível de habilidade, haverá uma certa probabilidade de que um examinado com aquela habilidade dê uma resposta correta a um item. Essa probabilidade é denotada por  $P(\theta)$ . Em um teste típico essa probabilidade será baixa para examinados com baixa habilidade e será alta para examinados com alta habilidade. Traçando  $P(\theta)$  como uma função da habilidade teremos como resultado uma curva suave em forma de  $S$ , como mostra a Figura 2.1.

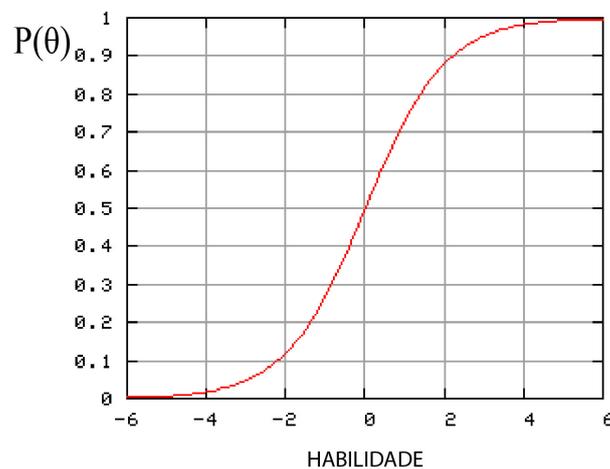


FIGURA 2.1: Exemplo de curva característica do item

Essa curva descreve a relação entre a probabilidade de resposta correta a um item e sua escala de habilidade. Na Teoria da Resposta ao Item tal curva é conhecida como Curva Característica do Item. Cada item em um teste irá possuir

uma curva característica própria.

Há duas propriedades que determinam a forma da curva característica: a dificuldade e a discriminação do item. A dificuldade descreve o quanto a escala de habilidades será alongada, por exemplo, uma função de item fácil, concentra examinados de baixa habilidade enquanto que uma função de item difícil, concentra examinados com alta habilidade, dessa forma, a dificuldade é um indicador local. A segunda propriedade técnica é a discriminação, a qual aponta o quão bem um item pode distinguir entre examinados que possuem habilidades abaixo ou acima da exigida para acertar o item. Essa propriedade essencialmente é notada pelo declive da curva característica do item em sua região central. Quanto mais inclinada a curva, mais poder de discriminação possuirá o item. Quanto mais assentada ou linear for a curva, menos adequado será o item para discriminar os examinados uma vez que a probabilidade de resposta correta em um nível de baixa habilidade é quase a mesma que em um nível de alta habilidade.

A partir da discussão a seguir, ver (Baker, 2001) conseguimos perceber a afirmação de a dificuldade do item ser um índice local, ou indicador local. A Figura 2.2 mostra três curvas características do item, plotadas na mesma escala. Elas possuem o nível de discriminação muito próximos, mas diferem em suas respectivas dificuldades. A curva *I1* representa um item fácil, uma vez que, mesmo indivíduos com baixa habilidade tem grande probabilidade de acertá-lo. A curva *I2*, mostra um item com dificuldade média uma vez que no meio da escala de habilidade a probabilidade de acerto da questão é em torno de cinquenta por cento. Por fim, a curva *I3* representa um item de maior dificuldade, nota-se que mesmo indivíduos com com alta habilidade tem apenas cerca de oitenta por cento de chance e acertá-lo.

Por sua vez, a Figura 2.3 contém três curvas características do item, possuindo nível de dificuldade muito próximo, só que desta vez o nível de discriminação é diferente para cada curva. A curva *I2*, mais linear, representa um item com baixo nível de discriminação, uma vez que ela não distingue bem entre entre examinados que tenham alta ou baixa habilidade, pelo fato da probabilidade de acerto para ambos os grupos serem bem próximas. A curva *I1*, que apresenta uma maior inclinação, representa um item com nível de discriminação médio, e por fim a curva

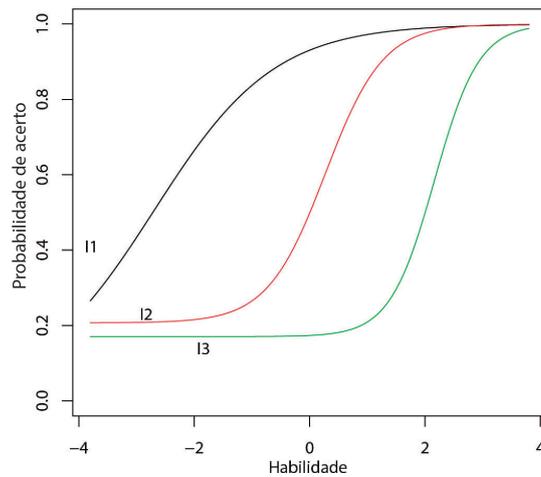


FIGURA 2.2: Curvas características com diferentes níveis de dificuldade.

mais inclinada, a *I3*, representa um item com alto índice de discriminação, percebe-se que para indivíduos com habilidade abaixo da metade da escala, a probabilidade de acerto no item é pequena, abaixo dos cinquenta por cento, e para indivíduos com habilidade acima da metade da escala, a probabilidade de acerto do item aproxima-se mais de cem por cento.

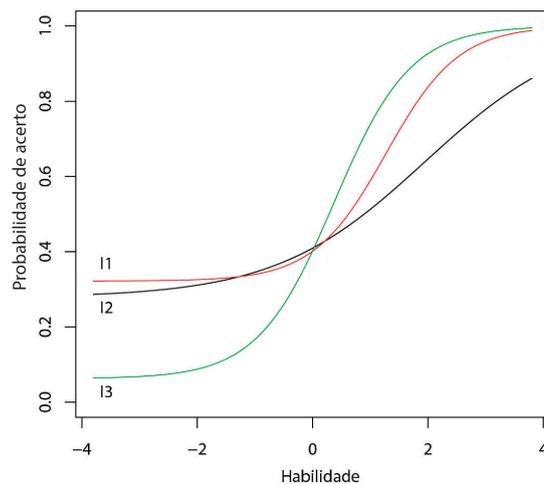


FIGURA 2.3: Curvas características com diferentes níveis de discriminação.

## 2.1 Modelos Matemáticos em TRI

Até agora foi feita uma abordagem geral das propriedades da Curva Característica do Item. Nesta seção iremos introduzir os modelos matemáticos para a Curva Característica do Item. Esses modelos partem de uma equação matemática que representa a relação entre a probabilidade de resposta correta a um item e a habilidade do respondente. Essa relação é sempre expressa de tal forma que quanto maior for a habilidade, maior a probabilidade de acerto no item. Além disso esses modelos juntamente com seus parâmetros, fornecem um veículo para a comunicação de informações sobre as propriedades técnicas dos itens.

Segundo Valle (1999) os modelos propostos dependem fundamentalmente de três fatores:

- da natureza do item - dicotômicos ou não dicotômicos;
- do número de populações envolvidas;
- do número de traços latentes que estão sendo medidos - apenas um ou mais de um.

Neste trabalho são considerados apenas os modelos que avaliam apenas um único traço latente ou habilidade, os chamados modelos unidimensionais. Na Teoria da Resposta ao Item, o modelo matemático padrão para a curva característica do item é obtido a partir da função logística. Tal função tem as características e o formato das funções definidas na seção anterior. Particularmente a função logística é muito utilizada em ciências biológicas, nos modelos de crescimento de plantas e reprodução de animais. Foi primeiramente deduzida em 1844, e dentro do contexto da TRI vem sendo utilizada desde a década de 1950. Como consequência, principalmente, de sua simplicidade, e por ser uma função explícita dos parâmetros, o que permite um melhor manejo e maiores possibilidades de resposta computacional nos cálculos das estimativas. Diz-se isto, principalmente pelo fato de que no início do desenvolvimento da TRI, o modelo utilizado era baseado na distribuição normal acumulada, que é uma função implícita dos parâmetros.

Na verdade os primeiros modelos de resposta ao item surgidos na década de 50, eram modelos em que se considerava que uma única habilidade, de um único grupo estava sendo medida por um teste onde os itens eram corrigidos de maneira dicotômica. Estes modelos foram primeiramente desenvolvidos na forma de uma função ogiva normal, como já citado, mas necessitava-se por questões de conveniência de um modelo mais simples. Birnbaum (1968) foi motivado pelo trabalho de Lord (1952) porém, sugeriu a troca da função ogiva normal pelo modelo logístico de dois parâmetros. Além disso Birnbaum introduziu o parâmetro de acerto ao acaso, também conhecido como "chute". Paralelamente a estes dois, Rasch (1960) por volta da década de 50 também, começou a desenvolver trabalhos utilizando modelos dicotômicos que acabaram levando seu nome e ficaram conhecidos como modelos de Rasch.

### 2.1.1 Modelo de Rasch ou de um Parâmetro

O modelo agora descrito foi primeiramente publicado pelo matemático dinamarquês George Rasch por volta de 1960. Ainda que Rasch já havia começado seu trabalho em medidas educacionais e psicometria por volta de 1940. Neste modelo, o parâmetro de discriminação do item, para um modelo de dois parâmetros da curva característica, tem seu valor fixado como  $a = 1$  para todos os itens; somente o parâmetro de dificuldade pode tomar diferentes valores. Por esse motivo o modelo de Rasch é frequentemente chamado de modelo logístico de um parâmetro (Baker, 2001).

Trata-se um modelo para dados dicotômicos, que expressa a relação entre a habilidade do respondente e a dificuldade do item proposto. Adota-se portanto o valor 1 para uma resposta correta e o valor 0 para uma resposta incorreta. Dado que a habilidade do respondente é representada pela letra grega  $\theta$ , e o parâmetro de dificuldade é representado pela letra  $b$ , temos que o modelo em questão, assume que a probabilidade de uma resposta ser correta para um determinado item, ou é dada pela diferença entre a habilidade do respondente e a dificuldade do item, ou seja,  $\theta - b$ .

Dessa forma a função logística para o modelo de um parâmetro é dada por:

$$P(\theta) = \frac{\exp(\theta - b)}{1 + \exp(\theta - b)}. \quad (2.1)$$

Ou como é mais comumente encontrado:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}, \quad (2.2)$$

em que:

$P_i(\theta)$ : a função resposta do item, que é a probabilidade de um examinado escolhido ao acaso, possuindo habilidade  $\theta$  responder corretamente ao item;

$\theta$ : nível de proficiência do respondente;

$b_i$  : parâmetro que representa a dificuldade do item  $i$ , medido na mesma escala da habilidade;

$D$ : um fator de escala usado para aproximar a função logística da ogiva Gaussiana com valor de 1,7.

A Curva Característica do Item para o modelo de 1 parâmetro (também conhecido como ML1) é mostrada na Figura 2.4.

Segundo Rasch (1960), quando a habilidade é igual à dificuldade, no modelo acima temos  $P(\theta) = \frac{1}{2}$ , ou seja, a probabilidade de acerto em um item quando a habilidade se iguala à dificuldade do item é de 50%. Caso a habilidade seja maior do que a dificuldade do item, ou seja,  $\theta > b$ , temos que  $P(\theta)$  toma valores positivos. No caso onde  $\theta < b$ ,  $P(\theta)$  toma valores negativos na escala de habilidade.

Há duas suposições assumidas para se aplicar os modelos da Teoria de Resposta ao Item. A primeira é a independência local, que significa dizer que a resposta de um item não deve influenciar na resposta de outro (essa suposição é fundamental para o processo de estimação dos parâmetros do modelo). Uma outra suposição para aplicação do modelo é a chamada unidimensionalidade do teste, ou seja, deve haver apenas um habilidade responsável pela realização de todos os itens

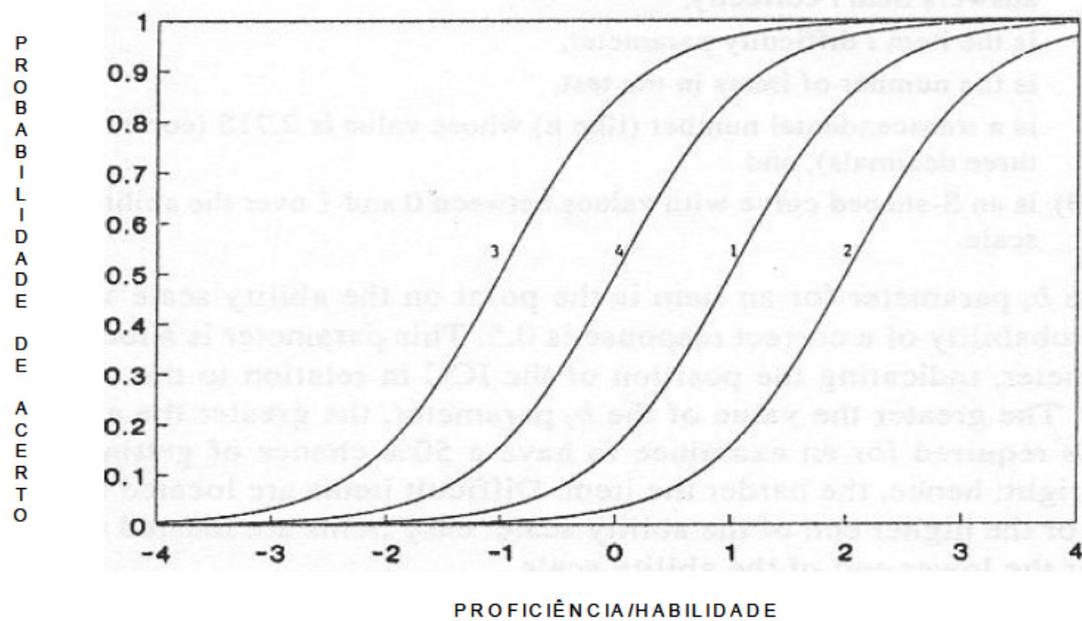


FIGURA 2.4: Curvas características para 4 itens utilizando o ML1.

do teste. Com relação à unidimensionalidade, segundo (Andrade, 2000), qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que mais de um traço latente entra na execução de qualquer tarefa. Dessa forma, é suficiente admitir que haja uma habilidade dominante responsável pelo conjunto de itens. Este fator é o que se supõe estar sendo medido pelo teste.

O modelo de Rasch centra-se na estimação conjunta da dificuldade ( $b_i$ ) dos  $n$  itens  $i$ ,  $i = 1, 2, \dots, n$  e das proficiências dos  $j$  examinados  $\theta_j$  em uma mesma escala.

A vantagem aqui em relação à Teoria Clássica dos Testes é que a TRI permite analisar as interações entre os respondentes e os itens, em termos de poder identificar as situações onde o respondente teria uma melhor probabilidade de responder corretamente aos itens do teste.

### 2.1.2 Modelo Logístico de Dois Parâmetros

Segundo Francisco (2005) Este foi o primeiro modelo criado para a Teoria da Resposta ao Item. Inicialmente esse modelo era baseado na função Gaussiana

padronizada e tinha como expressão:

$$P(U_i|\theta) = \int_{-\infty}^{a_i(\theta-b_1)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz. \quad (2.3)$$

Posteriormente, Birnbaum (1968) substituiu a função Gaussiana padronizada pela função logística e o modelo passou a ser escrito como:

$$P(U_i|\theta_j) = \frac{\exp(Da_i(\theta_j - b_i))}{1 + \exp(Da_i(\theta_j - b_i))}, \quad (2.4)$$

em que:

$P(U_i|\theta_j)$ : Função Característica do Item;

$a_i$ : É o parâmetro que representa o poder de discriminação do item  $i$ , com valor proporcional à inclinação da tangente à curva no ponto de abscissa  $b_i$ ;

$b_i$ : É o grau de dificuldade do item;

$\theta_j$ : é o nível de proficiência do respondente  $j$ ;

$D$ : é um fator de escala usado para aproximar a função logística da ogiva Gaussiana com valor 1,7.

Na Figura 2.5 está representado 4 curvas características para o modelo logístico de 2 parâmetros, também conhecido como ML2.

Tanto o modelo de um ou dois parâmetros não leva em consideração a possibilidade de o aluno acertar por acaso a questão. Pensando-se nisso procurou-se um novo modelo que incorpora-se esta possibilidade.

### 2.1.3 Modelo Logístico de Três Parâmetros

O modelo logístico de três parâmetros é definido por:

$$P(U_{ij} = 1|\theta) = c_i + \frac{(1 - c_i)\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]} \quad (2.5)$$

$$= c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta - b_i)]}, \quad (2.6)$$

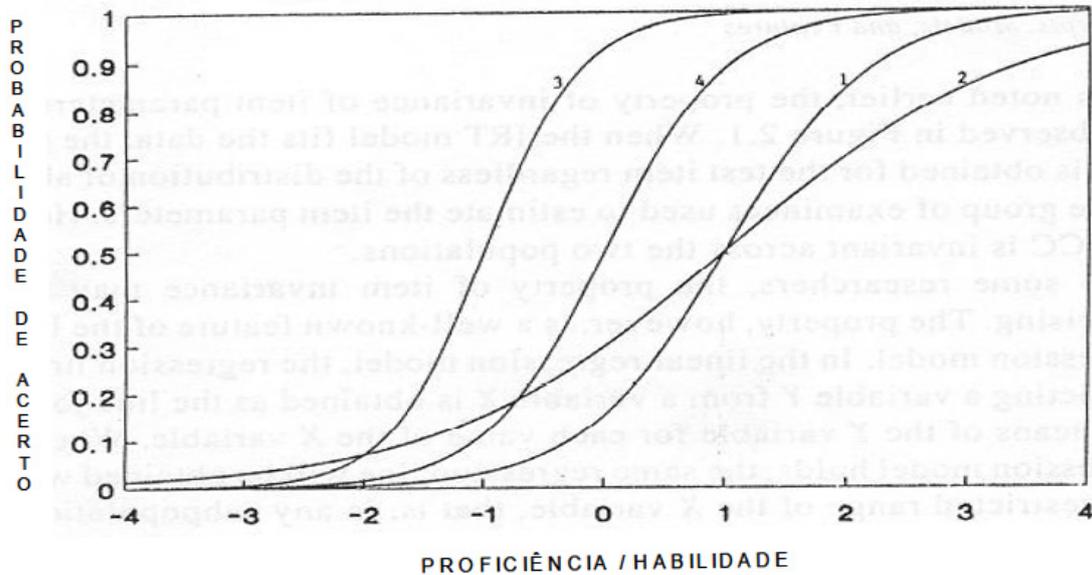


FIGURA 2.5: Curvas características para 4 itens utilizando o ML2.

$j = 1, \dots, J$ , em que,  $P(U_i = 1|\theta_j)$ : é a probabilidade do  $j$ -ésimo indivíduo escolhido ao acaso com grau de proficiência  $\theta_j$  responder corretamente ao  $i$ -ésimo item;

$U_{ij}$ : variável dicotômica que assume o valor 1 quando o  $j$ -ésimo indivíduo responde corretamente ao item  $i$ , e assume 0 quando o  $j$ -ésimo respondente não responde corretamente ao item  $i$ ;

$\theta_j$ : representa o grau de proficiência do  $j$ -ésimo respondente;

$a_i$ : é o parâmetro que corresponde ao índice de discriminação do item;

$b_i$ : é o parâmetro que corresponde ao grau de dificuldade do item;

$c_i$ : é o parâmetro que representa a probabilidade de acerto ao acaso;

D: é um fator de escala usado para aproximar a função logística da ogiva Gaussiana com valor 1,7.

Na Figura 2.6 estão representadas 4 curvas características do item utilizando o modelo logístico de 3 parâmetros, também conhecido como ML3.

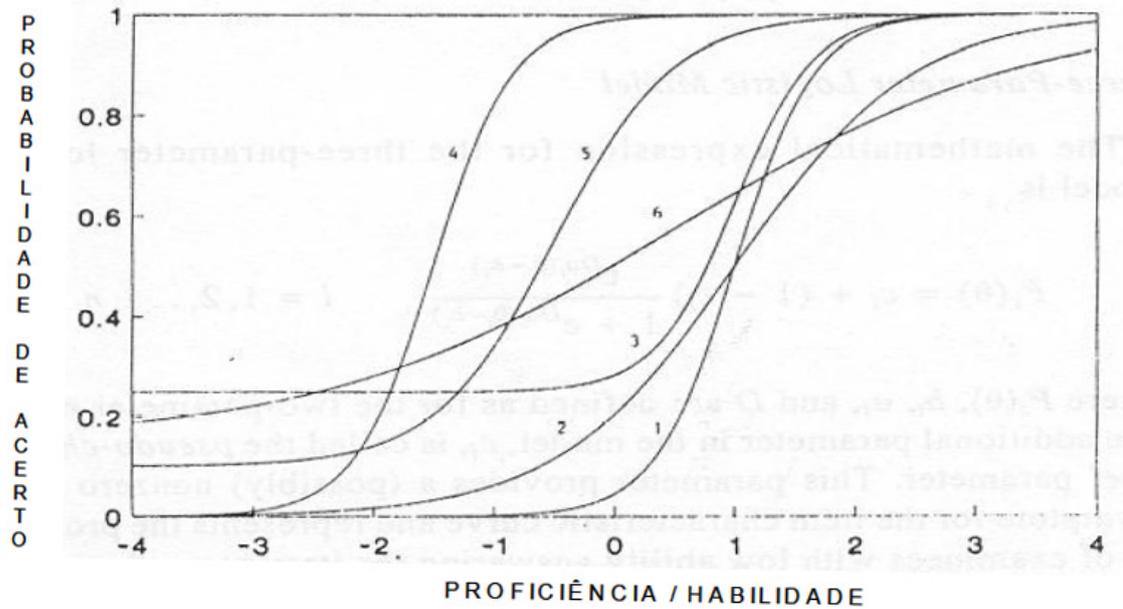


FIGURA 2.6: Curvas características para 4 itens utilizando o ML3.

### 2.1.4 Parâmetros da Curva Característica do Item

Segundo Andrade (2000), a  $P(U_{ij} = 1|1)$  pode ser vista como a proporção de respostas corretas ao item  $i$  dentre todos os indivíduos da população com habilidade  $\theta_j$ . A curva característica do item para  $P(U_{ij} = 1|1)$  é mostrada na Figura 2.7.

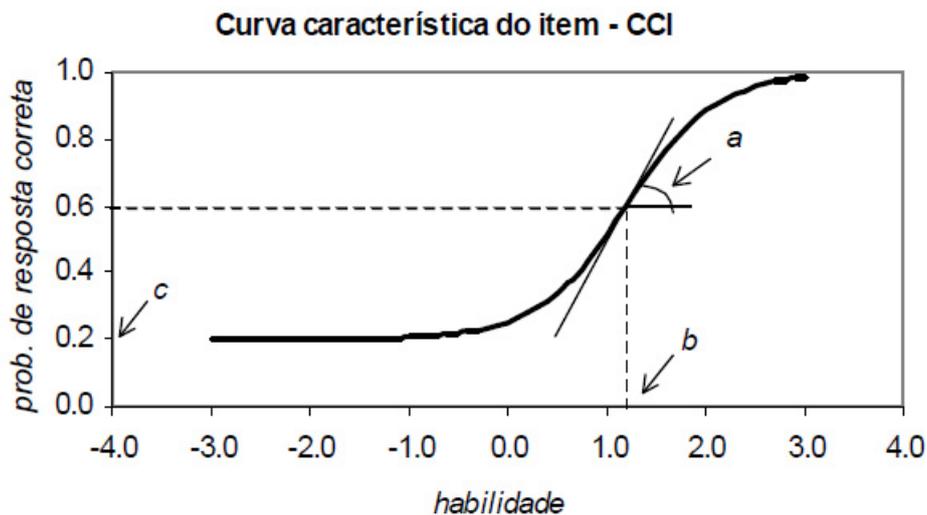


FIGURA 2.7: Curva característica do item e seus parâmetros.

Queremos enfatizar algumas propriedades dos parâmetros da Curva Característica do Item. Como já foi citado, o parâmetro  $a$  diz respeito à discriminação, o que implica dizer que este parâmetro está relacionado ao nível de conhecimento atestado pelos examinados. O valor de  $a$  é proporcional à derivada da tangente à curva no ponto  $b$ , ou seja, onde a curva muda de concavidade. Valores de  $a$  negativos não são esperados uma vez que isso significaria que a probabilidade de responder corretamente ao item diminuiria com o aumento da habilidade do examinado, para mais detalhes veja Francisco (2005), Andrade (2000), e Baker (1992).

Em termos práticos, uma curva mais íngreme sugere um grande poder discriminativo, o que ocorre para valores altos (positivos) do parâmetro  $a$ . O contrário ocorre quando o valor do parâmetro  $a$  tende a zero, o que fará com que a curva se torne mais achatada. Esse parâmetro também nos informa a qualidade que o item possui, e conseqüentemente o quão apropriada é a escala que está se usando.

O parâmetro  $b$  como já informado anteriormente, é medido na mesma escala da habilidade ( $\theta$ ). Em nosso contexto de estudo ele está relacionado com a dificuldade que um aluno tem em responder um determinado item. Sempre que  $b$  toma valores maiores, significa que a dificuldade do item aumenta e conseqüentemente é exigido do aluno um nível maior de habilidade. Segundo Andrade (2000), o parâmetro  $b$  representa a habilidade necessária para uma probabilidade de acerto igual a  $\frac{1+c}{2}$ .

O parâmetro  $c$  representa a probabilidade de um aluno com baixa habilidade responder corretamente ao item, também chamado de parâmetro de acerto ao acaso. Dessa forma, quando não é permitido chutar,  $c$  é igual a 0 e  $b$  representa o ponto da escala da habilidade onde a probabilidade de acertar o item é de 0,5 (Andrade, 2000).

Temos também ainda a considerar, a Função de Informação do Item, que determina a contribuição de cada item para a avaliação da habilidade. De modo geral, itens com alto poder discriminativo contribuem mais para a fidedignidade da medida do que itens com baixo poder de discriminação.

A função de informação do item é dada pela expressão

$$I_i(\theta) = \frac{[\frac{d}{d\theta} P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}, \quad (2.7)$$

em que:

$I_i(\theta)$  é a informação fornecida pelo item  $i$  no nível de habilidade  $\theta$ ;

$P_i(\theta) = P(X_{ij} = 1|\theta)$  e  $Q_i(\theta) = 1 - P_i(\theta)$

A função de informação do teste,  $I(\theta)$ , é dada pela soma de informações de cada item (Hambleton & Rogers, 1991).

$$I(\theta) = \sum_{i=1}^I I_i(\theta). \quad (2.8)$$

Para o modelo logístico de três parâmetros temos a expressão da função de informação do item é:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[ \frac{P_i(\theta) - c_i}{1 - c_i} \right]^2. \quad (2.9)$$

Trazendo para nossa discussão sobre as implicações dos valores de cada parâmetro, temos que a informação é maior quando:

$b_i$  se aproxima de  $\theta$ ;

quanto mais for o valor de  $a_i$

e quanto mais  $c_i$  se aproximar de 0.

## Capítulo 3

# Estimação dos Parâmetros

Uma das primeiras tarefas na TRI é a estimação dos parâmetros dos itens, assim como, das habilidades dos respondentes. Nos modelos da TRI, a probabilidade de uma resposta depende da habilidade do respondente,  $\theta_j$ , e dos parâmetros que caracterizam o item. Há, no entanto, três situações que podem ocorrer: a primeira é quando se conhece os parâmetros dos itens, restando-nos portanto estimar as habilidades dos respondentes; uma outra situação ocorre quando são conhecidas as habilidades dos respondentes e restamos estimar os parâmetros dos itens; por fim há o caso mais comum que ocorre, quando não conhecemos nem os parâmetros dos itens, nem as habilidades dos respondentes. Na verdade o que geralmente são conhecidas são as respostas dos examinados aos itens do teste. Em suma, o problema da estimação é portanto determinar o valor de  $\theta_j$  para cada examinado e os parâmetros que compõe cada item do teste. Fazendo um paralelo com os modelos clássicos de regressão, onde os parâmetros que caracterizam o modelo de regressão (os coeficientes de regressão) devem ser estimados, nota-se duas grandes diferenças entre estes modelos e os modelos da TRI. O primeiro, é que o modelo de regressão geralmente é linear. Segundo, e mais importante, o regressor (variável independente) na análise de regressão é observável. Na TRI a variável  $\theta$  é não observável. Esse aspecto dificulta bastante o problema de estimação dos parâmetros do modelo.

Em geral os métodos utilizados em TRI para estimação dos parâmetros dos

modelos são baseados em máxima verossimilhança, com auxílio de algum procedimento iterativo, como o algoritmo de Newton Raphson ou o "scoring" de Fisher ver Andrade (2000). Alguns procedimentos Bayesianos também são aplicados com alguma frequência. A estimativa do parâmetro do item então fornece as informações das propriedades técnicas dos itens do teste. Na TRI, o processo de estimação dos parâmetros dos itens é conhecido como *calibração*.

Na situação em que desejamos estimar tanto os parâmetros do itens, quanto as habilidades, há duas abordagens usuais: estimação conjunta, parâmetros dos itens e habilidades, ou em duas etapas, primeiro a estimação dos parâmetros dos itens e, posteriormente, das habilidades (Andrade, 2000).

Há alguns pontos a serem considerados, levando-se em conta cada tipo de abordagem. Segundo Andrade (2000), no caso da estimação conjunta, o número de parâmetros a serem estimados simultaneamente pode ser extremamente grande. Por exemplo, para o modelo logístico de três parâmetros, esse número seria de  $3I + n$ , onde  $I$  representa o vetor de parâmetro dos itens. O que leva a uma grande exigência computacional que envolve a inversão de matrizes de ordem  $3I + n$ . Para contornar esse problema, Birnbaum (1968) propôs um processo conhecido como "back-and-fork" ou "vai e volta" o qual é iniciado com estimativas grosseiras das habilidades e envolve a estimação dos parâmetros dos itens considerando as habilidades conhecidas; após a obtenção das estimativas dos parâmetros dos itens, as estimações das habilidades são feitas considerando conhecidos os parâmetros dos itens. Esses passos são repetidos até que algum critério de parada do processo seja alcançado. A vantagem desse procedimento é ele permitir que os itens sejam estimados individualmente, como consequência da independência local, já assumida anteriormente, o que exige o tratamento de matrizes de ordem 3 para o ML3. Da mesma forma, a partir da independência entre as respostas vindas de indivíduos diferentes, as habilidades passam a ser estimadas individualmente, o que diminui drasticamente a exigência computacional. Há entretanto algumas considerações sobre esse procedimento que precisam ser levantadas: sabe-se que, para os parâmetros dos itens conhecidos, os EMV's das habilidades convergem para os seus verdadeiros valores quando o número de itens cresce; com as habilidades conhecidas,

os EMV's dos parâmetros dos itens  $\hat{\zeta}_i$ , convergem para os seus verdadeiros valores quando o número de indivíduos cresce. Na estimação conjunta, as habilidades são denominadas de parâmetros incidentais, pois o número destes parâmetros ( $\theta_j$ ) cresce com o número de indivíduos; os parâmetros dos itens são denominados de parâmetros estruturais, e o número desses parâmetros não se altera quando a amostra cresce. Essas denominações são devidas a Neyman & Scott (1948), que notaram, em um contexto diferente da TRI, que na presença de parâmetros incidentais os EMV's dos parâmetros dos itens podem ser assintoticamente tendenciosos. Esse problema de falta de consistência dos parâmetros dos itens (ou habilidades) na presença de um número muito grande de indivíduos foi notado por Andersen (1973) e demonstrado para o modelo de Rasch. Porém, quando o número de itens e o número de indivíduos crescem, os EMV's dos parâmetros dos itens e das habilidades podem ser não-viciados, como sugerido por Lord (1968) e demonstrado apenas para o modelo de Rasch. Resultados numéricos obtidos por Lord (1975) e Swaminathan & Gifford (1983) reforçam a conjectura de que os estimadores de máxima verossimilhança dos parâmetros dos itens e das habilidades são não-viciados, quando o número de itens e o número de indivíduos crescem.

Devido ao fato de existir tal inconsistência sobre o viesamento ou não dos EMV's quando da estimação conjunta das habilidades dos respondentes e os parâmetros de habilidade dos mesmos, buscou-se ao desenvolvimento da estimação em duas partes. Este método baseia-se na existência de uma distribuição associada à proficiência dos indivíduos da população  $\mathbb{I}$ . Isso possibilita que a estimação dos itens seja feita pelo método da máxima verossimilhança marginal, ou seja, considerando uma determinada distribuição para a proficiência dos indivíduos de  $\mathbb{I}$ , cuja função densidade de probabilidade (fdp) é  $g(\theta|\eta)$ , onde  $\eta$  é o vetor de parâmetros associados à  $\mathbb{I}$  e integrando a função de verossimilhança com relação a  $\theta$ . Segundo Andrade (2000), após a estimação dos parâmetros dos itens, as proficiências são estimadas individualmente por máxima verossimilhança, ou pela moda ou média da distribuição condicional de  $\theta_j$ , sendo  $u_j = (u_{j1}, \dots, u_{jI})$ , o vetor de respostas do indivíduo  $j$ ,  $j = 1, \dots, n$ , com  $\zeta_i = (a_i, b_i, c_i)$ , e o vetor de parâmetros do item  $i$ ,  $i = 1, \dots, I$ , conhecidos.

Ainda, segundo Andrade (2000), embora este método tenha a vantagem de envolver, na primeira etapa, apenas a estimação dos parâmetros dos itens, a estimação é feita através de aplicação de métodos numéricos que dependem das derivadas segundas da log-verossimilhança com relação à  $\zeta_i$  e  $\zeta_k$ , com  $i, k = 1, \dots, I$ , que podem ser não nulas para  $i \neq k$ . O que acarreta termos que calcular a inversa de matrizes de ordem  $3I \times 3I$  para o Modelo Logístico de Três Parâmetros, o que não deixa de ser bastante trabalhoso do ponto de vista computacional. Tendo estas dificuldades em mente, Bock & Aitkin (1991) fizeram uma modificação no modelo de Bock & Lieberman (1991) adicionando a suposição de independência entre os itens, de forma que as derivadas segundas citadas acima para  $i \neq k$  sejam nulas. Dessa forma a matriz  $3I \times 3I$  de derivadas segundas, no Modelo Logístico de 3 parâmetros, torna-se bloco diagonal, o que possibilita que os parâmetros dos itens possam ser estimados individualmente. Há ainda outras formas de estimação dos parâmetros dos itens e das habilidades dos examinados, entretanto aquelas aqui já citadas são as mais importantes e são as que fizemos uso neste trabalho.

Antes de prosseguirmos apresentaremos algumas notações utilizadas para o processo de estimação, além de suposições que servirão de suporte para podermos desenvolver melhor os modelos propostos.

Seja  $\theta_j$  a habilidade e  $U_{ij}$  a variável aleatória que representa a resposta binária (ou dicotomizada) do indivíduo  $j$  ao item  $i$ , com

$$U_{ji} = \begin{cases} 1, \text{ caso a resposta esteja correta, com } P(U_{ji} = 1) = \theta_j \\ 0, \text{ caso a resposta esteja incorreta, com } P(U_{ji} = 0) = 1 - \theta_j \end{cases}$$

Sejam  $U_j = (U_{j1}, U_{j2}, \dots, U_{jI})$  o vetor aleatório de respostas do indivíduo  $j$  e  $U = (U_1, U_2, \dots, U_n)$  o conjunto de todas as respostas dadas. Da mesma forma representamos as observações por  $u_{ji}, u_j$  e  $u_{..}$ . Seja então  $\theta = (\theta_1, \dots, \theta_n)$  o vetor de habilidades dos  $n$  indivíduos e seja  $\zeta = (\zeta_1, \dots, \zeta_I)$  o conjunto dos parâmetros dos itens.

As suposições a serem aceitas serão as seguintes:

- As respostas dadas por examinados diferentes para determinado item são

independentes;

- Examinados com a mesma proficiência ( $\theta$ ) repondem a todos os itens do teste de forma independente.

### 3.0.5 Estimação dos Parâmetros dos Itens pelo Método da Máxima Verossimilhança

Considera-se agora o caso onde é conhecida a habilidade  $\theta$  (constante) do respondente a um determinado teste. Nosso objetivo é estimar os parâmetros dos itens que compõem esse teste. Na prática, essa situação não ocorre, mas o estudo matemático desse caso é importante para o desenvolvimento do caso mais complexo, que é estimar os parâmetros dos itens ( $a_i, b_i, c_i$ ) e das proficiências ( $\theta$ ) simultaneamente.

Estimar os parâmetros de todos os itens de um teste simultaneamente requer recursos computacionais muito dispendiosos. Dessa forma, na maioria dos procedimentos em TRI, os parâmetros dos itens são estimados item por item. Considerando, portanto que nosso objetivo é estimar um único item, digamos o item  $i$ , para obtermos os parâmetros dos  $I$  itens que compõem o teste, deveremos repetir o processo independentemente  $I$  vezes.

A estimação dos parâmetros de qualquer um dos modelos logísticos da TRI é feita utilizando-se o método da máxima verossimilhança, quando as proficiências são conhecidas. Utilizamos o Modelo Logístico de 3 Parâmetros para exemplificação, uma vez que este é o mais utilizado e além disso os modelos de dois e três parâmetros podem ser derivados deste.

Considerando portanto o vetor  $U_j$ , de respostas correspondentes ao examinado  $j$ , para os  $n$  itens do teste, e sendo  $U$  a matriz do conjunto de vetores das respostas de ordem  $m \times n$ . Dentro deste contexto, tem-se que a função de verossimilhança do vetor de parâmetros dos itens

$$\underline{\zeta} = (\zeta_1, \zeta_2, \dots, \zeta_I),$$

é dada por

$$L(\zeta) = P(U_{..} = u_{..} | \theta, \zeta).$$

Como:

$$L(\zeta) = \prod_{i=1}^n P(\underline{U}_j = \underline{u}_j | \theta_j, \underline{\zeta}_i),$$

e

$$L(\zeta) = \prod_{j=1}^n \prod_{i=1}^I P(U_{ji} = u_{ji} | \theta_j, \underline{\zeta}_i).$$

Onde se tem que a distribuição de  $U_{ji}$  só depende de  $\underline{\zeta}$  através de  $\zeta_i$

Usando a notação

$$P_{ji} = P(U_{ji}=1 | \theta_j, \zeta_i) \text{ e } Q_{ji} = 1 - P_{ji},$$

tem-se:

$$P(U_{ji} = u_{ji} | \theta_j, \zeta_i) = P(U_{ji} = 1 | \theta_j, \zeta_i)^{u_{ji}} P(U_{ji} = 0 | \theta_j, \zeta_i)^{1-u_{ji}}$$

.

Logo:

$$P(U_{ji} = u_{ji} | \theta_j, \zeta_i) = P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}},$$

e

$$L(\zeta) = \prod_{j=1}^n \prod_{i=1}^I P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}}.$$

Aplicando-se o logaritmo na função de verossimilhança obtém-se a função log-verossimilhança que é o logaritmo natural da função de verossimilhança, e pode ser escrita como:

$$\log L(\zeta) = \sum_{j=1}^n \sum_{i=1}^I \{u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}\}.$$

que é então derivada em relação à cada um dos parâmetros  $a_i$ ,  $b_i$  e  $c_i$ . Os EMV's de  $\zeta_i$ ,  $i = 1, \dots, I$ , são os valores que maximizam a função de verossimilhança, obtidos pelas soluções da equação:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = 0, \quad \text{com } i = 1, \dots, I.$$

Escrevendo a equação acima de outra forma, temos:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ u_{ji} \frac{\partial(\log P_{ji})}{\partial \zeta_i} + (1 - u_{ji}) \frac{\partial(\log Q_{ji})}{\partial \zeta_i} \right\}.$$

Aplicando-se a derivada do logaritmo, tem-se:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ u_{ji} \frac{1}{P_{ji}} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right) + (1 - u_{ji}) \frac{1}{Q_{ji}} \left( \frac{\partial(1 - P_{ji})}{\partial \zeta_i} \right) \right\} \quad (3.1)$$

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ u_{ji} \frac{1}{P_{ji}} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right) + (1 - u_{ji}) \frac{1}{Q_{ji} \left( -\frac{\partial P_{ji}}{\partial \zeta_i} \right)} \right\} \quad (3.2)$$

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ u_{ji} \frac{1}{P_{ji}} - (1 - u_{ji}) \frac{1}{Q_{ji}} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right) \quad (3.3)$$

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ \frac{u_{ji} Q_{ji} - P_{ji} + u_{ji} P_{ji}}{P_{ji} Q_{ji}} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right) \right\} \quad (3.4)$$

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ \frac{u_{ji}(Q_{ji} + P_{ji}) - P_{ji}}{P_{ji} Q_{ji}} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right). \quad (3.5)$$

Como  $Q_{ji} + P_{ji} = 1$ , temos que:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right).$$

Para as demonstrações a seguir utilizaremos a seguinte convenção  $W_{ji} = \frac{P_{ji}^* Q_{ji}^*}{P_{ji} Q_{ji}}$  e  $P_{ji} Q_{ji} = \frac{P_{ji}^* Q_{ji}^*}{W_{ji}}$ , onde  $P_{ji}^* = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$ , ou, equivalentemente  $P_{ji}^* = \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1}$  (Andrade, 2000) que é a probabilidade de resposta correta.

Além disso temos,  $Q_{ji}^* = 1 - P_{ji}^*$ , desta forma,

$$Q_{ji}^* = 1 - \{1 + e^{-Da_i(\theta_j - b_j)}\}^{-1}$$

$$Q_{ji}^* = \frac{e^{-Da_i(\theta_j - b_i)}}{1 + e^{-Da_i(\theta_j - b_i)}}.$$

Daí temos que

$$Q_{ji}^* = e^{-Da_i(\theta_j - b_i)} (1 + e^{-Da_i(\theta_j - b_i)})^{-1},$$

que corresponde à probabilidade de resposta incorreta.

Aplicando-se estas considerações à equação

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right).$$

Temos:

$$\begin{aligned} \frac{\partial \log L(\zeta)}{\zeta_i} &= \sum_{j=1}^n \left\{ \frac{u_{ji} - P_{ji}}{\frac{P_{ji}^* Q_{ji}^*}{W_{ji}}} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right), \\ \frac{\partial \log L(\zeta)}{\zeta_i} &= \sum_{j=1}^n \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^* Q_{ji}^*} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right). \end{aligned}$$

Para obtermos as equações de estimação dos parâmetros  $a_i$ ,  $b_i$  e  $c_i$ , precisamos das expressões a seguir, que são derivadas do modelo logístico de três parâmetros:

$$P(U_{ji} = 1 | \theta_j) = C_i + (1 - C_i)(1 + e^{-Da_i(\theta_j - b_i)})^{-1}.$$

Derivando a expressão do ML3 em relação ao parâmetro  $a_i$ , temos:

$$\begin{aligned} \frac{\partial P_{ji}}{\partial a_i} &= (1 - c_i)(-1) (1 + e^{-Da_i(\theta_j - b_i)})^{-2} e^{-D(\theta_j - b_i)} (-D(\theta_j - b_i)) + (1 + e^{(-Da_i(\theta_j - b_i))})^{-1} \times 0 \\ \frac{\partial P_{ji}}{\partial a_i} &= D(1 - c_i)(\theta_j - b_i) (1 + e^{-Da_i(\theta_j - b_i)})^{-2} e^{-D(\theta_j - b_i)}. \end{aligned}$$

Sendo  $P_{ji}^* = \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1}$  e  $Q_{ji}^* = 1 - P_{ji}^*$ , temos que,

$$\begin{aligned} P_{ji}^* Q_{ji}^* &= (1 + e^{-Da_i(\theta_j - b_i)})^{-1} \left[ 1 - (1 + e^{-Da_i(\theta_j - b_i)})^{-1} \right] \\ P_{ji}^* Q_{ji}^* &= (1 + e^{-Da_i(\theta_j - b_i)})^{-1} \frac{e^{-Da_i(\theta_j - b_i)}}{1 + e^{-Da_i(\theta_j - b_i)}} \\ P_{ji}^* Q_{ji}^* &= \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \frac{e^{-Da_i(\theta_j - b_i)}}{1 + e^{-Da_i(\theta_j - b_i)}} \\ P_{ji}^* Q_{ji}^* &= \frac{e^{-Da_i(\theta_j - b_i)}}{(1 + e^{-Da_i(\theta_j - b_i)})^2}. \end{aligned}$$

Substituindo este valor na expressão  $(1 + e^{-Da_i(\theta_j - b_i)})^{-2} e^{-Da_i(\theta_j - b_i)} = P_{ji}^* Q_{ji}^*$ , temos para o parâmetro de discriminação  $a_i$

$$\frac{\partial P_{ji}}{\partial a_i} = D(1 - c_i)(\theta_j - b_i) P_{ji}^* Q_{ji}^*.$$

Da mesma forma, aplicamos a derivada do ML3 em relação a  $b_i$ , para obtermos a equação de estimação do parâmetro de dificuldade do item, e obtemos a expressão

$$\frac{\partial P_{ji}}{\partial b_i} = -Da_i(1 - c_i) (1 + e^{-Da_i(\theta_j - b_i)})^{-2} e^{-D(\theta_j - b_i)},$$

e substituindo  $P_{ji}^* Q_{ji}^*$  pelo valor encontrado na expressão da derivada do ML3 em relação ao parâmetro  $b_i$ , obtemos

$$\frac{\partial P_{ji}}{\partial b_i} = -Da_i(1 - c_i) P_{ji}^* Q_{ji}^*.$$

De forma idêntica, prosseguimos para encontrar a equação de estimação do parâmetro de acerto ao acaso do item ( $c_i$ ), derivando o ML3 em relação a  $c_i$ , obtendo a expressão a seguir

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

Ou equivalentemente,

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

Desta forma

$$\frac{\partial P_{ji}}{\partial c_i} = Q_{ji}^*.$$

Estas equações nos permitem determinar as EMV's para os três parâmetros de cada item, para mais detalhes ver (Andrade, 2000).

### 3.1 Estimação dos Parâmetros Considerando o Agrupamento das Habilidades

Uma outra forma de estimar os parâmetros dos itens é considerar as proficiências dos respondentes agrupados. Isso torna-se possível quando os valores de  $\theta$  são conhecidos, permitindo-nos separar as habilidades em  $q$  intervalos, cujos valores centrais (ou alguma medida central dessas habilidades) denotaremos por  $\bar{\theta}_k$ ,  $k = 1, 2, \dots, q$ . Consideremos, para fins práticos, que todos os indivíduos pertencentes à categoria  $k$  possuam proficiências  $\bar{\theta}_k$ , com isto reduzimos significativamente a exigência computacional, o que torna este método de estimação bastante atrativo.

Considere, portanto, os  $q$  grupos de  $f_{ki}$ , com  $k = 1, \dots, q$ , indivíduos com proficiências  $\bar{\theta}_k$  conhecidas. Selecionamos então estes  $q$  grupos de uma população  $\Pi$  de respondentes do item  $i$ . Seja  $r_{ki}$  o número de indivíduos do grupo  $k$  que responderam corretamente ao item  $i$ .

Considerando a independência local entre as repostas dos diferentes examinados, podemos assumir que  $r_{ki}$ ,  $K = 1, \dots, q$ , tem distribuição Binomial com parâmetros  $f_{ki}$  e  $P_{ki}$ , sendo  $P_{ki}$  o modelo logístico de 3 parâmetros.

Substituindo-se  $\theta_j$  por  $\bar{\theta}_k$ , a função de verossimilhança assume a forma

$$L(\zeta) = \prod_{k=1}^q \prod_{i=1}^I \left\{ \binom{f_{ki}}{r_{ki}} P_{ki}^{r_{ki}} Q_{ki}^{f_{ki} - r_{ki}} \right\}.$$

A log-verossimilhança será dada portanto por

$$\log L(\zeta) = C + \sum_{k=1}^q \sum_{i=1}^I \{r_{ki} \log P_{ki} + (f_{ki} - r_{ki}) \log Q_{ki}\}, \quad (3.6)$$

em que:

$$C = \sum_{k=1}^q \sum_{i=1}^I \binom{f_{ki}}{r_{ki}},$$

é constante em relação a  $\zeta$ .

Tomando-se a derivada da equação (3.6) em relação a  $\zeta_i$ , obtemos:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{k=1}^q \frac{1}{P_{ki} Q_{ki}} (r_{ki} - f_{ki} P_{ki}) \left( \frac{\partial P_{ki}}{\partial \zeta_i} \right).$$

Sendo

$$P_{ji} Q_{ji} = \frac{P_{ji}^* Q_{ji}^*}{W_{ji}},$$

obtemos

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{k=1}^q (r_{ki} - f_{ki} P_{ki}) \frac{W_{ki}}{P_{ki}^* Q_{ki}^*} \left( \frac{\partial P_{ki}}{\partial \zeta_i} \right).$$

Agora repetimos o procedimento de derivar O ML3 em relação a cada um dos parâmetros, como já feito anteriormente, ficando com as seguintes expressões:

$$\frac{\partial P_{ji}}{\partial a_i} = D(1 - c_i)(\theta_j - b_i) P_{ki}^* Q_{ki}^*, \quad (3.7)$$

$$\frac{\partial P_{ji}}{\partial b_i} = -D a_i (1 - c_i) P_{ki}^* Q_{ki}^*, \quad (3.8)$$

$$\frac{\partial P_{ji}}{\partial c_i} = Q_{ki}^*. \quad (3.9)$$

Desta forma as equações de estimações para os parâmetros dos itens,  $a_i$ ,  $b_i$  e  $c_i$ , quando as proficiências são agrupadas, são dadas por:

$$a_i : \quad D(1 - c_i) \sum_{k=1}^q (r_{ki} - f_{ki}P_{ki})(\bar{\theta}_k - b_i)W_{ki} = 0, \quad (3.10)$$

$$b_i : \quad -Da_i(1 - c_i) \sum_{k=1}^q (r_{ki} - f_{ki}P_{ki})W_{ki} = 0, \quad (3.11)$$

$$c_i = \sum_{k=1}^q (r_{ki} - f_{ki}P_{ki}) \frac{W_{ki}}{P_{ki}^*} = 0 \quad (3.12)$$

A solução do sistema formado pelas equações (3.10)-(3.12) não pode ser obtida diretamente, por isso, faz-se uso de métodos iterativos como Método de Newton-Raphson ou o "scoring" de Fisher para obtermos as EMV's dos parâmetros dos itens.

Optamos fazer uso do software R Core Team (2013) para otimização dos processos de estimação e construção das curvas relativas à TRI, para mais detalhes teóricos sobre o uso do método de Newton Raphson ou "scoring" de Fisher, sugerimos a leitura de Francisco (2005) e Andrade (2000).

# Capítulo 4

## Aplicações

Neste capítulo utilizamos o modelo TRI com três parâmetros para analisar dados do sistema SARESP

### 4.1 Utilização do R na Estimação de Parâmetros em Dados do SARESP

Utilizaremos o apoio computacional do sistema R (R Core Team, 2013). R é uma linguagem e ambiente para computação estatística e gráficos. Faz parte da filosofia do Projeto GNU e está disponível como Software Livre sob os termos da "Licença Pública Geral do GNU" da Fundação do Software Livre (Free Software Foundation's GNU General Public License) na forma de código fonte. Ele compila e roda sobre uma larga variedade de plataformas UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS (Fontana, 2009).

O R é um software de código aberto que permite fazer o tratamento dos dados utilizando pacotes otimizados para tarefas relacionadas com a psicometria fazendo uso da TRI. Seguimos a sugestão de (Anjos & Andrade, 2012) fazendo uso de um conjunto de pacotes organizados na área de Psicometria chamado *Psychometrics*, que pode ser acessado em <http://cran-r.c3sl.ufpr.br/web/views/Psychometrics.html>. Em particular utilizamos os pacotes *irtoys* (Partchev, I., 2010), *ltm* (Rizopoulos,

2006), Deducer (Fellows, 2012), plink (Weeks, 2010) e CTT (Willse, 2014). A versão do R utilizada foi a "R version 3.1.1 (2014-07-10)".

## 4.2 Aplicação do Modelo de TRI com 3 Parâmetros

Os dados a seguir foram obtidos do SARESP - Sistema de Avaliação de Rendimento Escolar do Estado São Paulo e correspondem a um teste com trinta itens de língua portuguesa aplicados a alunos do Ensino Médio dos três turnos, matutino, vespertino e noturno, e estão disponíveis em <http://www.ufpr.br/~aanjos/TRI/sinape/dados/saresp.dat>, tendo como fonte a Secretaria Estadual da Educação de São Paulo (Anjos & Andrade, 2012).

## 4.3 Leitura do Arquivo

Antes de continuarmos, devemos indicar a forma como os dados estão apresentados para podermos identificar cada característica de interesse neste exemplo.

1		ABDCAABCDCAADBDDABDDDCDCADADBC
2		ADCDAADBCCCABABDBCABADCABBBDBC
3		BDCACABACACABADDABDBADCABDBBBC
1	1 011001138433m07	ADDCAADBDBABDDABABCBCBDCCAADBC
1	1 011002964093m07	DBACAACDABACDBBAABDDBCAACDADBC
1	1 011004154243m07	ABDCAABCDDAADBDDABDADCDDADACBC
1	1 011005367283m07	DCDCACABADCCCCADCBCABCCCCAABCD
1	1 011007519633m07	DBDCABCDBAABDBDCCDBDDBADBBABBA
1	1 011008054863m07	DDBCAACDBAAACBBBCBDDDBADBCACBC
1	1 011008066103m07	DBADBACDBAACCBBCABDDCBADBDABBC
1	1 011008864603m07	DBDDADDCDCDCDDDAABBDBDBABDBBBB
1	1 011010038713m07	ADCCABCDAABCADBABABDACBADBCACDC

O arquivo está organizado da seguinte forma: as três primeiras linhas são o gabarito do teste, cada uma referente a um turno de ensino, a linha 1 corresponde ao gabarito do turno matutino, a 2 ao gabarito do turno vespertino e por fim, a linha 3 corresponde ao gabarito do teste aplicado aos alunos do turno noturno.

Abaixo das três primeiras linhas, a primeira coluna corresponde ao turno de ensino, sendo 1 para matutino, 2 para vespertino e 3 para o noturno. A segunda coluna corresponde ao teste aplicado; a terceira, que é uma sequência com 15 algarismos, corresponde a um identificador referente a cada aluno, e por fim, a última coluna são as respostas dos alunos aos trinta itens do teste em questão.

Para fazermos a leitura destes dados basta inserir-mos as seguintes linhas no software R:

```
> saresp<-read.fwf(  
'http://www.ufpr.br/~aanjos/TRI/sinape/dados/saresp.dat',  
widths=c(1,-1,1,-1,12,3,-1,rep(1,30)),  
header=FALSE,skip=3,na.strings=' ')  
colnames(saresp)<-c('grupo','escola','id','turno',  
paste('i',1:30,sep=' '))
```

Vejamos portanto o significado de cada argumento na linha de comando: o comando `saresp<-read.fwf()`, juntamente com o endereço virtual fornecido, permite ao R buscar e ler o arquivo indicado (no caso o arquivo do SARESP cuja fonte já foi indicada). Já o argumento `widths=c(1,-1,1,-1,12,3,-1,rep(1,30))` primeiramente indica ao programa que ele leia a primeira coluna com 1 dígito (que faz referência ao turno no qual foi aplicado o teste), não leia a segunda coluna, leia a terceira com 3, também com 1 dígito, não leia a coluna 4, leia o código de identificação dos alunos, que são os próximos 12 números como sendo apenas uma coluna, os próximos 3 algarismos também apenas como uma coluna, não leia a próxima coluna e por fim, considere cada uma das letras como uma coluna.

Os argumentos `header=FALSE`, indicam que as colunas do arquivo não possuem nomes, O `skip=3` pedi que o programa ignore as três primeiras linhas, e o `na.strings=' '`) que considere os espaços em branco como *missing*.

Usamos o `head(saresp)` para obter os dados:

```

      grupo  escola  id          turno  i1 i2 i3 i4 i5 i6 i7 i8 i9 i10 i11 i12 i13 i14
1         1      1  11001138433 m07    A D D C A A D B D B A  B  D  D
2         1      1  11002964093 m07    D B A C A A C D A B A  C  D  B
3         1      1  11004154243 m07    A B D C A A B C D D A  A  D  B
4         1      1  11005367283 m07    D C D C A C A B A D C  C  C  C
5         1      1  11007519633 m07    D B D C A B C D B A A  B  D  B
6         1      1  11008054863 m07    D D B C A A C D B A A  A  C  B

      i15 i16 i17 i18 i19 i20 i21 i22 i23 i24 i25 i26 i27 i28 i29 i30
1  A   B   A   B   C   B   C   B   D   C   C   A   A   D   B   C
2  B   A   A   B   D   D   B   C   A   A   C   D   A   D   B   C
3  D   D   A   B   D   A   D   C   D   D   A   D   A   C   B   C
4  A   D   C   B   C   A   B   C   C   C   C   A   A   B   C   D
5  D   C   C   D   B   D   D   B   A   D   B   B   A   B   B   A
6  B   B   C   B   D   D   D   B   A   D   B   C   A   C   B   C

```

O gabarito pode ser obtido com a seguinte sequência de comandos:

```

gabarito<-read.fwf('http://www.ufpr.br/~aanjos/TRI/dados/saresp.dat',
widths=c(-20,rep(1,30)),header=F,nr=3)

colnames(gabarito)<-c(paste('i',1:30,sep=))

gabarito

```

O que nos da o seguinte resultado:

```

      i1 i2 i3 i4 i5 i6 i7 i8 i9 i10 i11 i12 i13 i14 i15 i16 i17 i18 i19 i20 i21
1  A  B  D  C  A  A  B  C  D  C  A  A  D  B  D  D  A  B  D  D  D
2  A  D  C  D  A  A  D  B  C  C  C  A  B  A  B  D  B  C  A  B  A
3  B  D  C  A  C  A  B  A  C  A  C  A  B  A  D  D  A  B  D  B  A

      i22 i23 i24 i25 i26 i27 i28 i29 i30
1  C  D  C  A  D  A  D  B  C
2  D  C  A  B  B  B  D  B  C
3  D  C  A  B  D  B  B  B  C

```

Como queremos apenas exemplificar aqui, indicamos ao R que leia apenas as três primeiras linhas do arquivo utilizando a opção `nr=3`. Há agora um procedimento que é necessário ser feito para que os dados possam ser lidos e interpretados sob o ponto de vista da TRI. Observe que as respostas aos itens estão na forma de letras, a saber, as alternativas *a*, *b*, *c*, e *d*; e como já foi indicado no início deste trabalho,

precisamos que estes dados sejam dicotomizados. Segundo Anjos & Andrade (2012) as principais funções conhecidas para análise de dados da TRI requerem que as respostas estejam em um objeto do tipo `data.frame` ou `matrix`, com números 0 e 1 (0 indica resposta incorreta e 1 uma resposta correta).

Faremos portanto uso do pacote `Deducer` utilizando a função `recode.variable()` deste pacote para podermos converter as respostas de letras para números. Para carregá-lo basta usar o comando `library(Deducer)`

Faremos isto utilizando o seguinte comando:

```
dados<-recode.variables(saresp,"'A'->1; 'B'->2; 'C'->3; 'D'->4")
```

Para visualizarmos os dados utilizamos o comando `head(dados)`, uma vez que, com o comando acima criamos o objeto `dados` e para que os nomes das colunas possa ter a mesma nomenclatura do arquivo `saresp` original, utilizamos o comando `names(dados)<-names(saresp)`.

Para análise dos dados, vamos fazer uso apenas dos dados do turno matutino. Podemos extrair tais dados da seguinte maneira

```
manha<-subset(dados,grupo==1,select=names(dados))
```

```
head(manha)
```

	grupo	escola	id	turno	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14
1	1	1	11001138433	m07	1	4	4	3	1	1	4	2	4	2	1	2	4	4
2	1	1	11002964093	m07	4	2	1	3	1	1	3	4	1	2	1	3	4	2
3	1	1	11004154243	m07	1	2	4	3	1	1	2	3	4	4	1	1	4	2
4	1	1	11005367283	m07	4	3	4	3	1	3	1	2	1	4	3	3	3	3
5	1	1	11007519633	m07	4	2	4	3	1	2	3	4	2	1	1	2	4	2
6	1	1	11008054863	m07	4	4	2	3	1	1	3	4	2	1	1	1	3	2

	i15	i16	i17	i18	i19	i20	i21	i22	i23	i24	i25	i26	i27	i28	i29	i30
1	1	2	1	2	3	2	3	2	4	3	3	1	1	4	2	3
2	2	1	1	2	4	4	2	3	1	1	3	4	1	4	2	3
3	4	4	1	2	4	1	4	3	4	4	1	4	1	3	2	3
4	1	4	3	2	3	1	2	3	3	3	3	1	1	2	3	4
5	4	3	3	4	2	4	4	2	1	4	2	2	1	2	2	1
6	2	2	3	2	4	4	4	2	1	4	2	3	1	3	2	3

Como é do nosso interesse, a princípio, somente os dados do turno matutino, faremos a recodificação do gabarito usando os seguintes comandos:

```
gab<-recode.variables(gabarito,"'A'->1; 'B'->2; 'C'->3; 'D'->4")
```

```
colnames(gab)<-c(paste('i',1:30,sep=))
```

```
gab
```

E tem como saída

```

  i1 i2 i3 i4 i5 i6 i7 i8 i9 i10 i11 i12 i13 i14 i15 i16 i17 i18 i19 i20 i21
1 1 2 4 3 1 1 2 3 4 3 1 1 4 2 4 4 1 2 4 4 4
2 1 4 3 4 1 1 4 2 3 3 3 1 2 1 2 4 2 3 1 2 1
3 2 4 3 1 3 1 2 1 3 1 3 1 2 1 4 4 1 2 4 2 1

  i22 i23 i24 i25 i26 i27 i28 i29 i30
1 3 4 3 1 4 1 4 2 3
2 4 3 1 2 2 2 4 2 3
3 4 3 1 2 4 2 2 2 3

```

No objeto `dados` as respostas estão gravadas como 1, 2, 3 e 4, cada um desses números correspondendo às alternativas a,b,c e d respectivamente. Faremos portanto uso do pacote `ltm` com o intuito de transformar as respostas para a forma 0 ou 1 utilizando a função `multi.choice()` do pacote citado, a partir das informações obtidas do gabarito. Para isso criamos o objeto `manha.NA` com as informações já dicotomizadas.

Antes de mais nada, carregamos o pacote `ltm` utilizando o comando `library(ltm)`.

Feito isto utilizamos o comando:

```
manha.NA<-mult.choice(manha[,5:34],as.numeric(gab[1,]))
```

```
head(manha.NA)
```

Obtendo os seguintes dados:

```

  i1 i2 i3 i4 i5 i6 i7 i8 i9 i10 i11 i12 i13 i14 i15
1 1 1 0 1 1 1 1 0 0 1 0 1 0 0
2 0 1 0 1 1 1 0 0 0 0 1 0 1 1 0
3 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1
4 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0
5 0 1 1 1 1 0 0 0 0 0 1 0 1 1 1
6 0 0 0 1 1 1 0 0 0 0 1 1 0 1 0

```

	i16	i17	i18	i19	i20	i21	i22	i23	i24	i25	i26	i27	i28	i29	i30
1	0	1	1	0	0	0	0	1	1	0	0	1	1	1	1
2	0	1	1	1	1	0	1	0	0	0	1	1	1	1	1
3	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1
4	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0
5	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0
6	0	0	1	1	1	1	0	0	0	0	0	1	0	1	1

Algumas observações podem ainda ser feitas com relação a dados em falta, estes dados aparecem quando o examinado não dá resposta para algum item, e pode ser observado utilizando-se o comando:

```

descript(manha.NA)$missin

```

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
Freq	0	6.0000000	4.0000000	0	0	0	1.0000000	2.0000000	3.0000000	1.0000000
(%)	0	0.5994006	0.3996004	0	0	0	0.0999001	0.1998002	0.2997003	0.0999001

	i11	i12	i13	i14	i15	i16	i17	i18	i19
Freq	0	0	4.0000000	1.0000000	1.0000000	1.0000000	0	0	4.0000000
(%)	0	0	0.3996004	0.0999001	0.0999001	0.0999001	0	0	0.3996004

	i20	i21	i22	i23	i24	i25	i26	i27
Freq	1.0000000	0	2.0000000	4.0000000	4.0000000	2.0000000	3.0000000	1.0000000
(%)	0.0999001	0	0.1998002	0.3996004	0.3996004	0.1998002	0.2997003	0.0999001

	i28	i29	i30
Freq	3.0000000	3.0000000	1.0000000
(%)	0.2997003	0.2997003	0.0999001

Esse tipo de dado é conhecido como 'missings' ou 'NA's', quando um pacote do R não permite a leitura desse tipo de dado, as respostas faltantes devem ser codificadas como incorretas (Anjos & Andrade, 2012). Para fazer isso, usamos o seguinte comando, para o objeto `manha`:

```
manha.f<-ifelse(is.na(manha.NA)==T,0,manha.NA)
```

### 4.3.1 Análise dos Dados Utilizando o Pacote Irtoys

#### Estimação dos Parâmetros

Antes de tudo, devemos carregar o pacote `irtoys` com o comando `library(irtoys)` O processo de estimação dos parâmetros, para o modelos logístico de três parâmetros, também conhecido como calibração em TRI, pode ser obtido da seguinte maneira, fazendo uso da função `est()` no R:

```
manha.f.par<-est(manha.f, model = "3PL", engine = "ltm",
  nqp = 20)
```

Criando portanto um objeto de nome `manha.f.par` que contém os valores estimados dos parâmetros para o ML3. Para visualizar os primeiros parâmetros, fazemos uso do comando `head(manha.f.par)`, obtendo como resultado:

```

$est
      [1]      [2]      [3]
i1  0.6644912  1.99392048  0.253467616
i2  1.9176240  0.26004408  0.194251538
i3  1.7516722  1.11478121  0.158545335
i4  1.0914463  0.25838770  0.144479287
i5  0.9412978 -2.72278521  0.042459114
i6  0.5695324 -0.42061784  0.035576380
i7  0.8229451  0.79664518  0.102547303
i8  0.5208884  1.34752364  0.001570670
i9  1.5328465  1.04475287  0.176233636
i10 1.4767474  1.29191358  0.309627553
i11 0.6023265 -0.64063531  0.010022951
i12 0.7842418  0.98748358  0.327779311
i13 1.1718633 -0.10201865  0.040756908
i14 1.2411512  0.24302102  0.246493438
i15 1.8518783  2.25958982  0.266862520
i16 1.6486812  1.69808984  0.159598509
i17 2.2123951  0.78535678  0.488306475
i18 3.2381238  0.01230251  0.733241528
i19 0.9762965  0.45642929  0.084285320
i20 1.2949585  0.51635292  0.587890671
i21 0.5838297  0.98182525  0.002208187
i22 3.5570522  2.12163173  0.175740316
i23 2.4368194  2.29198787  0.215549013
i24 0.7976424  2.58066634  0.401730984
i25 0.9672422  0.99653800  0.003259543
i26 1.3648268 -0.24152498  0.194079847
i27 2.0420187  0.61792463  0.581292571
i28 1.5260226  0.36114174  0.067286984
i29 1.2263268 -1.08025199  0.002740888
i30 1.3711384 -0.61151383  0.428043019

$se
```

A Função de Informação do Item fazendo o uso do seguinte comando:

```
plot(iif(manha.f.par$est),label=T)
```

e obtemos a Figura 4.1

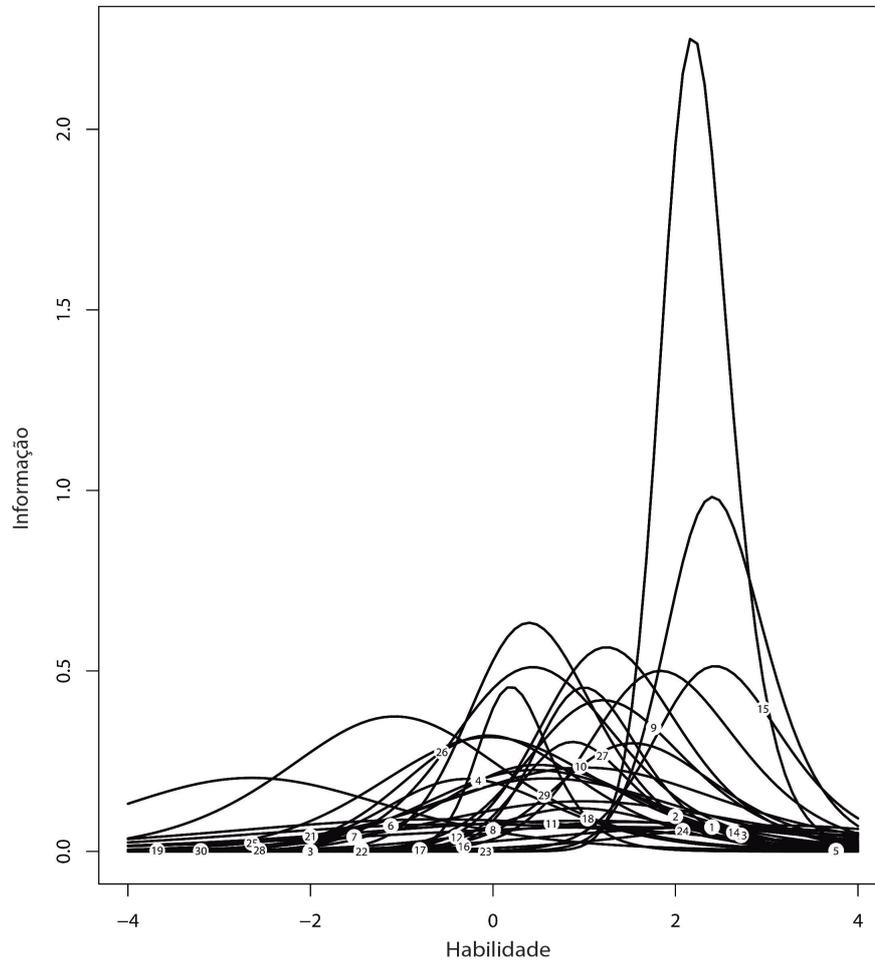


FIGURA 4.1: Função de informação do item para os dados do SARESP 2007

Já a Curva Característica do Item pode ser obtida através do comando:

```
plot(irf(manha.f.par$est),label=T)
```

Tendo com resultado o gráfico mostrado na Figura 4.2.

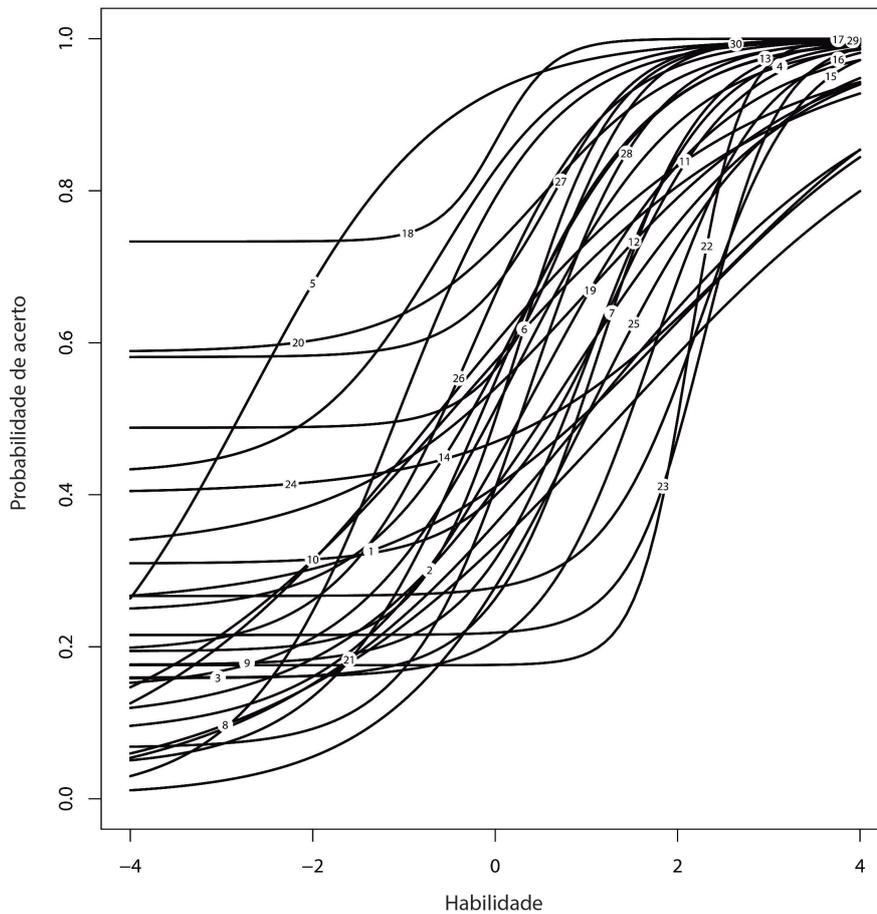


FIGURA 4.2: Curva característica do item para os dados do SARESP 2007

### Estimação das Habilidades dos Respondentes

Tendo sido feitas as devidas estimativas para os valores dos parâmetros dos ML3 acima, vamos agora estimar a habilidade  $\theta$  dos examinados. Para isto faremos uso da função `eap()`, sendo que precisamos indicar o arquivo que contém as respostas e o objeto com os parâmetros estimados que obtemos anteriormente. Aplicamos então o comando:

```
manha.f.sco=eap(manha.f,ip=manha.f.par$est,qu=normal.qu())
```

Obtendo como resultado o objeto `manha.f.sco` que contém o valor de  $\theta$  estimado para cada indivíduo da população em questão. Para visualizarmos parte

destes dados, fazemos uso do `head(manha.f.sco)`, obtendo os valores abaixo:

	est	sem	n
[1,]	0.06110404	0.4771474	30
[2,]	0.58837056	0.4089677	30
[3,]	1.92623806	0.4680314	30
[4,]	-1.71143568	0.6046441	30
[5,]	-0.47840008	0.4540995	30
[6,]	-0.29040561	0.4498074	30

Agora, fazendo uso do comando `plot(tif(manha.f.par$est), label=T)`, plotamos a função de informação do teste que tem como gráfico a Figura 4.3

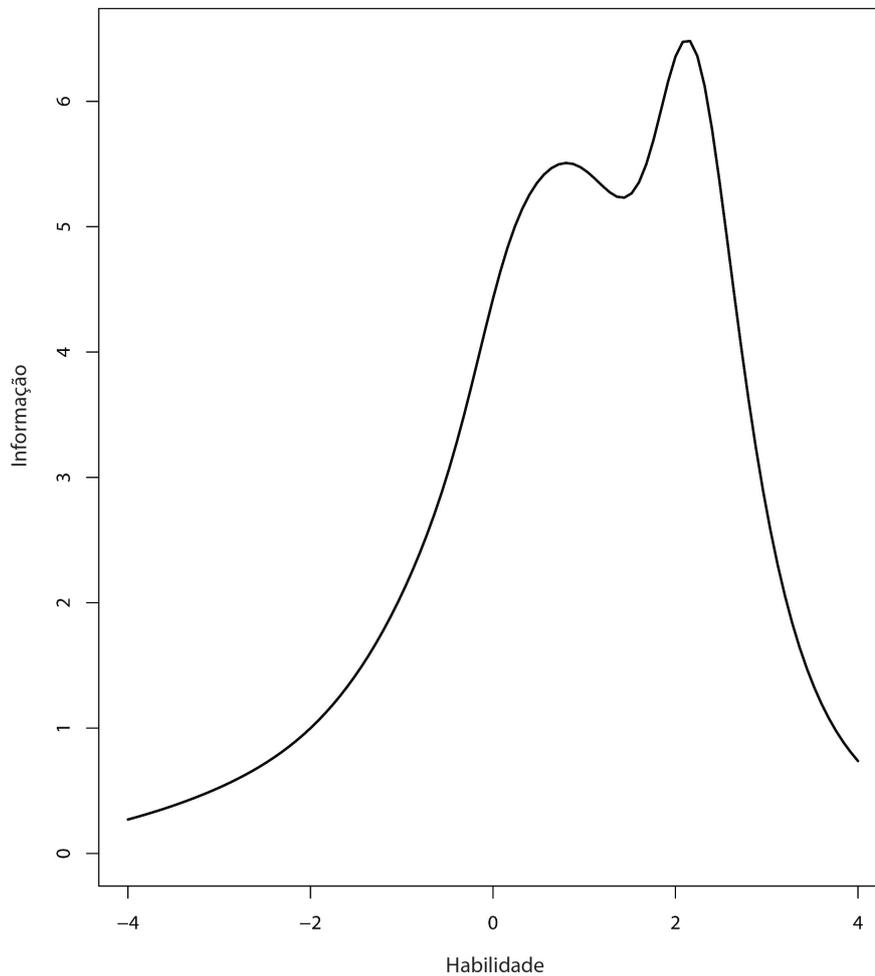


FIGURA 4.3: Função de informação do teste para os dados do SARESP 2007.

Se quisermos comparar a relação entre a habilidade dos alunos e o escore,

fazemos isto com o comando:

```
trf(manha.f.par$est)

$х
 [1] -4.00 -3.92 -3.84 -3.76 -3.68 -3.60 -3.52 -3.44 -3.36 -3.28 -3.20 -3.12
[13] -3.04 -2.96 -2.88 -2.80 -2.72 -2.64 -2.56 -2.48 -2.40 -2.32 -2.24 -2.16
[25] -2.08 -2.00 -1.92 -1.84 -1.76 -1.68 -1.60 -1.52 -1.44 -1.36 -1.28 -1.20
[37] -1.12 -1.04 -0.96 -0.88 -0.80 -0.72 -0.64 -0.56 -0.48 -0.40 -0.32 -0.24
[49] -0.16 -0.08 0.00 0.08 0.16 0.24 0.32 0.40 0.48 0.56 0.64 0.72
[61] 0.80 0.88 0.96 1.04 1.12 1.20 1.28 1.36 1.44 1.52 1.60 1.68
[73] 1.76 1.84 1.92 2.00 2.08 2.16 2.24 2.32 2.40 2.48 2.56 2.64
[85] 2.72 2.80 2.88 2.96 3.04 3.12 3.20 3.28 3.36 3.44 3.52 3.60
[97] 3.68 3.76 3.84 3.92 4.00

$f
 [1] 7.122584 7.160749 7.200909 7.243151 7.287566 7.334249 7.383300
 [8] 7.434820 7.488918 7.545709 7.605312 7.667854 7.733471 7.802306
[15] 7.874511 7.950250 8.029696 8.113035 8.200463 8.292190 8.388439
[22] 8.489444 8.595453 8.706727 8.823539 8.946172 9.074921 9.210090
[29] 9.351993 9.500950 9.657290 9.821346 9.993459 10.173971 10.363235
[36] 10.561603 10.769436 10.987101 11.214970 11.453420 11.702830 11.963574
[43] 12.236012 12.520467 12.817201 13.126378 13.448016 13.781942 14.127754
[50] 14.484793 14.852163 15.228760 15.613342 16.004599 16.401214 16.801899
[57] 17.205413 17.610561 18.016190 18.421193 18.824518 19.225194 19.622366
[64] 20.015323 20.403542 20.786712 21.164758 21.537847 21.906371 22.270914
[71] 22.632169 22.990819 23.347364 23.701919 24.054004 24.402397 24.745122
[78] 25.079605 25.402995 25.712560 26.006043 26.281884 26.539281 26.778120
[85] 26.998827 27.202210 27.389312 27.561295 27.719362 27.864703 27.998454
[92] 28.121684 28.235379 28.340442 28.437691 28.527863 28.611621 28.689555
[99] 28.762195 28.830010 28.893420

$ni
 [1] 30

attr("class")
 [1] "trf"
```

A leitura desta informação é feita da seguinte maneira: o valor de  $x$  representa a habilidade estimada, enquanto que o valor de  $f$  representa o correspondente score esperado.

## 4.4 Construindo o Ranking dos Examinados

O comando a seguir criará um objeto com as seguintes informações: scores, posição e número de acertos de cada examinado.

```
final.rank<-data.frame('escore'=manha.f.sco[,1],
'posição'=rank(manha.f.sco[,1]),
'acertos'=margin2table(manha.f)[-1002,31])
```

Obtendo os resultados a seguir:

	escore	posição	acertos
1	0.06110404	538	16
2	0.58837056	748	17
3	1.92623806	987	26
4	-1.71143568	15	8
5	-0.47840008	326	12
6	-0.29040561	398	13

Se quisermos ordenar os resultados, fazemos uso do comando:

```
final.acertos<- final.rank[order(final.rank$acertos),]  
head(final.acertos)
```

Tendo como resultado os valores abaixo:

	escore	posição	acertos
25	-0.9453425	156	7
88	-1.9417899	2	7
128	-1.6560666	18	7
149	-1.5962607	24	7
183	-1.8380660	7	7
192	-1.6005866	23	7

Ou ainda:

```
tail(final.acertos)

          escore   posição  acertos
422  2.450968   999      27
803  2.348995   997      27
859  2.177786   993      27
229  2.604056  1000      28
628  2.360057   998      28
558  2.766691  1001      29
```

Se o objetivo é ordenar em função da classificação, modificamos o comando acima utilizando o argumento `escore` e obtemos o resultado mostrado a seguir.

```
final.escore<- final.rank[order(final.rank$escore),]

head(final.escore)

          escore   posição  acertos
257  -1.942270    1          7
88   -1.941790    2          7
971  -1.905838    3          7
315  -1.888097    4          8
325  -1.857824    5         10
395  -1.847286    6          7
```

```
tail(final.escore)
```

Caso seja de interesse podemos alterar a escala da habilidade, para obtermos, por exemplo, escores não negativos. Faremos uma alteração exemplificativa da escala  $N(0,1)$  para a  $N(500,10)$  com o comando:

```
novo.score<-score.transform(manha.f.sco[,1],
```

---

	escore	posição	acertos
409	2.335383	996	27
803	2.348995	997	27
628	2.360057	998	28
422	2.450968	999	27
229	2.604056	1000	28
558	2.766691	1001	29

```
mu.new = 500, sd.new = 100, normalize = FALSE)
```

```
round(head(novo.score$new, n=30), 2)
```

Os procedimentos que foram feitos até agora, podem também ser efetuados por intermédio do pacote `ltm`. A seguinte sequência de comandos nos dá, respectivamente: a estimativa dos parâmetros, o erro padrão das estimativas dos parâmetros, e as curvas características dos itens.

```
manha.f.tpm <- tpm(manha.f)
```

```
manha.f.tpm
```

```
summary(manha.f.tpm)
```

```
plot(manha.f.tpm, legend=F)
```

Para mais detalhes, ver Anjos & Andrade (2012).

# Capítulo 5

## Considerações Finais

O processo de avaliação educacional é um dos aspectos mais importantes e por vezes menos aperfeiçoado. Neste processo podemos observar se o conhecimento transmitido está sendo absorvido pelos alunos. Sem dúvida, há um grande interesse que os alunos desenvolvam habilidades mínimas que lhes permitam uma melhor desenvoltura ao longo do Ensino Médio e posteriormente no acadêmico. Além disso, avaliar a qualidade do ensino oportuniza a conscientização sobre os pontos fortes e fracos do processo ensino-aprendizagem, apontando caminhos para mudanças.

Neste trabalho apresentamos os fundamentos da Teoria da Resposta ao Item, fazendo uso principalmente do modelo logístico de três parâmetros nas nossas aplicações, com o apoio computacional do (R Core Team, 2013). A análise dos parâmetros da curva característica do item mostram a importância da reflexão sobre a construção de itens que melhorem a qualidade das avaliações ou testes a qual estão submetidos os estudantes.

Os resultados obtidos no Capítulo 4, permitem uma melhor classificação dos alunos em determinados testes, levando em consideração não somente o escore obtido mas apontando para os traços latentes em desenvolvimento.

Análises de dados do cotidiano do educador e de outros sistemas de rendimento escolar podem ser também realizadas. Além disso, como sugestão para trabalhos futuros temos a análise do processo de equalização, que permite comparar o desenvolvimento seriado de alunos ou mesmo comparar o desempenho de alunos

em turnos diferentes ou em contextos educacionais diferentes como por exemplo estudantes de escolas públicas e estudantes de escolas particulares.

# Referências

- Anastasi, A. & Urbina, S. (2000). *Testagem Psicológica*. Artes Médicas, Porto Alegre-RS.
- Andersen, E. B. (1973). *Conditional inference in multiple choice questionnaires*. British Journal of Mathematical and Statistical Psychology, 26,, 31-44.
- Andrade, D. F.; Tavares, H. V. R. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. 14 SINAPE, Caxambú-MG.
- Anjos, A. d. & Andrade, D. F. (2012). *Teoria da Resposta ao Item com uso do R*. 20 SINAPE, João Pessoa - PB.
- Baker, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. Marcel Dekker, Inc, New York.
- Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC, University of Wisconsin.
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. Statistical Theories of Mental Test Scores, Reding, MA.
- Bock, R. D. & Aitkin, M. (1991). *Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm*. Psychometrika, 46, 433-459.
- Bock, R. D. & Lieberman, M. (1991). *Fitting a response model for n dichotomously scored items*. Psychometrika, 35, 179-197.
- Fellows, I. (2012). Deducer: A data analysis gui for R. *Journal of Statistical Software*, 49(8), 1–15.
- Fontana, B. da Silva, D. J. A. M. B. (2009). *Minicurso de Estatística Básica: Introdução ao Software R*. Universidade Federal de Santa Maria, Santa Maria - RS.
- Francisco, R. (2005). *Aplicação da Teoria da Resposta ao Item (TRI) no Exame Nacional de Cursos (E.N.C.) da UNICENTRO*. Dissertação de mestrado, UFPR – Universidade Federal do Paraná, Paraná, Brasil.
- Hambleton, R. k. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publications, Newbury Park.

- Lord, F. M. (1952). *The Relation of the Reliability of Multiplechoice Tests to the Distribution or Item Difficulties*. *Psychometrika* 17, 181-194.
- Lord, F. M. (1968). *An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model*. *Educational and Psychological Measurement*, 28, 989-1020.
- Lord, F. M. (1975). *Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters*. (Research Bulletin RB-75-33), Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum, Hillsdale - NJ.
- Neyman, J. & Scott, E. L. (1948). *Consistent estimation from partially consistent observations*. *Econometrica* 16, 1-32.
- Partchev, I. (2010). *irtoys: Simple interface to the estimation and plotting of IRT models*. R package version 0.1.3., S.l.
- Perrenoud, P. (1998). *Avaliação. Da Excelência à Regulação das Aprendizagens*. Artmed Editora, Bruxelles: De Boeck.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen.
- Requena, C. S. (1990). *Psicometria Teoria Y Practica en la Construcción de Tests*. Normas, Madrid - Espanha.
- Rizopoulos, D. (2006). ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25.
- Swaminathan, H. & Gifford, J. A. (1983). *Estimation in the three-parameter latent trait model*. In D. Weiss (Ed.), *New horizons in testing*, New York: Academic Press, 13-29.
- Valle, R. C. (1999). *Teoria da Resposta ao Item*. Dissertação de mestrado, IME – Universidade de São Paulo, São Paulo, Brasil.
- Vendramini, C. M. M., D. A. S. (2005). *Teoria de Resposta ao Item na Análise de uma Prova de Estatística em Universitários*. *Psico-USF*, v.10,n.2, São Paulo- SP.
- Weeks, J. P. (2010). plink: An R package for linking mixed-format tests using irt-based methods. *Journal of Statistical Software*, 35(12), 1–33.
- Willse, J. T. (2014). *CTT: Classical Test Theory Functions*. R package version 2.1.