



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

Testes de adequabilidade de ajuste  
em Teoria de Resposta ao Item

por

Bernardo Nogueira Schlemper

Orientador: Prof. Dr. Antônio Eduardo Gomes

Junho de 2014

Bernardo Nogueira Schlemper

**Testes de adequabilidade de ajuste  
em Teoria de Resposta ao Item**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

**Universidade de Brasília**

**Brasília, Junho de 2014**

TERMO DE APROVAÇÃO

Nome do aluno

TESTES DE ADEQUABILIDADE DE AJUSTE  
EM TEORIA DE RESPOSTA AO ITEM

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Data da defesa: 25 de junho de 2014

Orientador:

---

Prof. Dr. Antônio Eduardo Gomes  
Departamento de Estatística, UnB

Comissão Examinadora:

---

Prof. Dr. André Luiz Fernandes Cançado  
Departamento de Estatística, UnB

---

Prof. Dr. Dalton Francisco de Andrade  
Departamento de Informática e Estatística, Universidade Federal de Santa Catarina

Brasília, Junho de 2014

## Ficha Catalográfica

**SCHLEMPER, BERNARDO NOGUEIRA**

Teses de adequabilidade de ajuste em Teoria de Resposta  
ao Item, (UnB - IE, Mestre em Estatística, 2010).

Dissertação de Mestrado - Universidade de Brasília. Departamento de Estatística  
- Instituto de Ciências Exatas.

1. Teste de adequabilidade
2. Estimador Não-paramétrico de Máxima Verossimilhança
3. Bootstrap
4. Regressão Isotônica
5. Kernel

É concedida à Universidade de Brasília a permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta monografia de Projeto Final pode ser reproduzida sem a autorização por escrito do autor.

Bernardo Nogueira Schlemper

*Aos meus pais, amigos,  
e todos que de alguma forma estiveram  
presentes durante a elaboração desse trabalho*

# Agradecimentos

À minha família, que sempre esteve ao meu lado, dentre eles minha mãe sempre com palavras de incentivo e sem a qual nada disso seria possível, seja por todos os ensinamentos ao longo da vida ou por todos os esforços que fez para me ver feliz, meu irmão e eterno parceiro pro que der e vier, Luiz por todos os conselhos e por todas as vezes que me ajudou com toda sua experiência e inteligência para tomar as melhores decisões, meu pai por todos os momentos de alegria.

Aos meus amigos e segunda família Leonardo (Zéo), André, Tomaz, Alexis, Sérgio e Sarah, por estarem sempre comigo dando força, motivação pra continuar, e estando presentes nos momentos mais engraçados e também os mais importantes desde que os conheço. Vocês são os melhores amigos que alguém pode ter!

Aos meus amigos que moram longe, Riva, Greg, Bene, Abib, Ian, Wiki, Jambo, Dimenor, Saidh, Sabino, Palandré, Tafaro, Thawan, Paiva, Yuri, Marcuxo, Dandan, que me proporcionam 24 horas de risos por dia e com quem posso sempre desabafar quando a coisa fica séria.

À Janaina, a quem eu devo muito do que sou hoje, por ter me ajudado em todos os aspectos da minha vida, a ser um estudante e uma pessoa melhor, com todos os toques que me fizeram e fazem sempre pensar que posso melhorar, que sempre acreditou em mim até quando nem eu acreditava. Você é especial.

Ao professor e meu orientador Antônio Eduardo, pela orientação e cuja ajuda foi fundamental em todas as etapas do trabalho.

Aos membros da banca examinadora, pelas críticas e sugestões que contribuíram para a melhora desse e de futuros trabalhos.

# Sumário

<b>Lista de Figuras</b>	<b>2</b>
<b>Resumo</b>	<b>3</b>
<b>1 Conceitos Básicos em Teoria de Resposta ao Item</b>	<b>5</b>
1.1 Principais modelos em TRI . . . . .	5
1.1.1 Modelos para itens dicotômicos . . . . .	6
1.1.2 Modelos para itens não dicotômicos . . . . .	12
<b>2 Cronologia da TRI e dos testes não paramétricos</b>	<b>19</b>
<b>3 Testes de adequabilidade de ajuste para modelos de Teoria de Resposta ao Item</b>	<b>27</b>
3.1 Teste baseado em suavização por kernel (Ramsay, 1991) . . . . .	27
3.2 Teste baseado na distribuição Qui-quadrado (Orlando & Thissen, 2000)	29
3.3 Teste baseado em probabilidades a posteriori (Stone, 2000) . . . . .	36
3.4 Teste Baseado em Regressão Isotônica . . . . .	37
<b>4 Simulação e Aplicação</b>	<b>42</b>
4.1 Elaboração da simulação . . . . .	42
4.2 Aplicação . . . . .	47
<b>5 Conclusões e Trabalhos futuros</b>	<b>66</b>
5.1 Conclusões . . . . .	66
5.2 Trabalhos futuros . . . . .	66

# Lista de Figuras

1.1	Curva característica do item . . . . .	8
1.2	Curvas características do item (linha contínua) e curvas de informação (linha pontilhada) para diferentes valores dos parâmetros . . . . .	11
3.1	Regressão isotônica e minorante convexo máximo . . . . .	40
4.1	Gráfico com CCIs para o item 19 do ML3 utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	61
4.2	Gráfico com CCIs para o item 35 do ML3 utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	61
4.3	Gráfico com CCIs para o item 4 do ML3 com 6 itens fora do modelo utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	62
4.4	Gráfico com CCIs para o item 39 do ML3 com 6 itens fora do modelo utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	62
4.5	Gráfico com CCIs para o item 13 do ML3 com 18 itens fora do modelo utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	63
4.6	Gráfico com CCIs para o item 14 do ML3 com 18 itens fora do modelo utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	63
4.7	Gráfico com CCIs para o item 14 do ML2 utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	64
4.8	Gráfico com CCIs para o item 37 do ML2 utilizando $\theta_{estimado}$ via modelo paramétrico no ENPMV . . . . .	64
4.9	Gráfico com ENMPV e CCIs estimadas para itens da prova do SA-RESP utilizando score padrão no ENPMV . . . . .	65



# Resumo

Neste trabalho, são apresentados testes de adequabilidade de ajuste para modelos de teoria de resposta ao item (TRI). Um novo teste baseado na estimação não paramétrica da curva característica do item (CCI) obtida a partir da aplicação de técnicas de regressão isotônica é apresentado. O teste foi implementado no software R, e foram feitas simulações para observar seu desempenho na avaliação da adequabilidade de modelos paramétricos para a curva característica do item, juntamente com o teste utilizando a estimação via suavização por kernel (Ramsay,1991).

**Palavras Chave:** *Teste de Adequabilidade, Regressão Isotônica, Kernel, Estimador Não-paramétrico de Máxima Verossimilhança, Bootstrap.*

# INTRODUÇÃO

A Teoria de Resposta ao Item (TRI) abrange um conjunto de modelos estatísticos que vem ganhando destaque no últimos tempos principalmente graças a sua aplicação na avaliação educacional. Algumas avaliações brasileiras importantes já utilizam modelos de TRI, como o Sistema de Avaliação da Educação Básica (Saeb) e o Exame Nacional do Ensino Médio (Enem).

A TRI é muito utilizada em medidas psicométricas, com destaque para sua utilização na avaliação de conhecimento e habilidades. Tal utilização se dá pelo fato de que a probabilidade de acerto a um determinado item de um dado teste é condicionada à habilidade e conhecimento do respondente, ou seja, a modelagem é feita de forma que espera-se que indivíduos com maior conhecimento em um determinado assunto acertem mais questões em um teste que outro indivíduo com menos conhecimento em modelos cumulativos, e também que em itens considerados mais difíceis haja mais acertos de indivíduos com conhecimento maior.

Com isso, a TRI apresenta diferenças consideráveis em relação a Teoria Clássica, pois essa se baseia no escore bruto ou padronizado do indivíduo em um teste. Porém tal abordagem apresenta limitações, como a dependência do conjunto de itens que compõem o instrumento de medida, e dessa forma, limita sua aplicabilidade. Enquanto a Teoria Clássica tem foco na prova, a TRI tem foco nos itens.

# Capítulo 1

## Conceitos Básicos em Teoria de Resposta ao Item

### 1.1 Principais modelos em TRI

Dentro da TRI existem diversos modelos matemáticos, cujo objetivo é descrever de maneira mais precisa possível a probabilidade de um indivíduo responder corretamente a um determinado item como função dos parâmetros do item e de habilidades do respondente. Essa relação se apresenta de forma que quanto maior a habilidade do respondente, maior será a probabilidade de que ele acerte o item.

Tais modelos dependem basicamente de três fatores:

- natureza do item,
- número de populações envolvidas (populações de respondentes),
- quantidade de traços latentes sendo medida.

A natureza do item corresponde ao item ser dicotômico ou não dicotômico. O número de populações envolvidas e a quantidade de traços latentes dizem respeito a apenas uma população ou um traço latente, ou mais de um desses fatores, respectivamente.

Vale lembrar que há uma distinção entre grupo e população aqui. O que será chamado de grupo é uma amostra de elementos (indivíduos) de uma população. Esse

conceito se assemelha ao processo de amostragem aleatória simples, dessa forma, quando for utilizado um grupo de respondentes, refere-se a uma amostra de indivíduos de uma mesma população.

### 1.1.1 Modelos para itens dicotômicos

Os modelos para itens dicotômicos são utilizados não somente para itens de resposta livre avaliados de maneira dicotomizada mas também para itens de múltipla escolha dicotomizados, que são aqueles corrigidos como certo ou errado.

Os modelos mais utilizados para itens dicotômicos são os logísticos, sendo eles em geral divididos em três tipos, de acordo com a quantidade de parâmetros que são utilizados para descrever o item. Esses três tipos de modelos consideram os seguintes parâmetros:

- dificuldade do item,
- dificuldade do item e discriminação,
- dificuldade do item, discriminação e probabilidade de resposta correta do indivíduo com baixa habilidade.

O parâmetro de discriminação deve assumir um valor mínimo, que garanta probabilidades diferentes de acerto a respondentes com proficiências diferentes. O grau de informação do item está diretamente relacionado a esse parâmetro.

O modelo logístico com três parâmetros (ML3) é o mais completo e os outros dois mais simples (com 1 e 2 parâmetros) podem ser obtidos a partir dele, portanto.

### Modelo logístico de 3 parâmetros (ML3)

O modelo logístico unidimensional de 3 parâmetros é dado por

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (1.1)$$

com  $i=1,2,\dots,I$  e  $j=1,2,\dots,n$ , onde:

$U_{ij}$  é a variável dicotômica que assume o valor 0 quando o indivíduo  $j$  responde de maneira incorreta o item  $i$ , e assume valor 1 quando o indivíduo responde corretamente o item  $i$ .

$\theta_j$  é a habilidade (ou proficiência) do  $j$ -ésimo indivíduo (traço latente).

$P(U_{ij} = 1|\theta_j)$  é a probabilidade do indivíduo  $j$  com habilidade  $\theta_j$  responder o item  $i$  de maneira correta. Essa probabilidade, vista como função da habilidade, também é chamada de Função de Resposta do Item (FRI) ou Curva Característica do Item (CCI), e seus parâmetros são:

$b_i$  é o parâmetro de dificuldade do item  $i$ ;

$a_i$  é o parâmetro de discriminação na região de  $b_i$ ;

$c_i$  é o parâmetro do item que descreve a probabilidade de resposta correta ao item  $i$  do indivíduo com baixa habilidade.

O parâmetro  $b_i$  é medido na mesma escala da habilidade, e  $a_i$  tem valor igual à inclinação da Curva Característica do Item (CCI) no ponto  $b_i$ .  $P(U_{ij} = 1|\theta_j)$  é explicada como a proporção de respostas corretas dadas ao item  $i$  de todos os respondentes que possuem habilidade  $\theta_j$ .

A Figura 1.1 demonstra a relação entre os parâmetros do modelo e  $P(U_{ij} = 1|\theta_j)$ .

A escala de habilidade é arbitrária, ou seja, as relações de ordem entre seus pontos são mais importantes que sua magnitude. O parâmetro  $b$ , como dito anteriormente, é medido na mesma unidade da habilidade, e o parâmetro  $c$  varia entre 0 e 1 por se tratar de uma probabilidade.

O parâmetro de discriminação  $a$  é proporcional à derivada da tangente da curva no ponto de inflexão. Dessa forma, não se esperam valores negativos para esse parâmetro, uma vez que isso implicaria em dizer que a probabilidade de responder um item de maneira correta diminui a medida que a habilidade do indivíduo aumenta. Valores altos para esse parâmetro implicariam em uma curva com grande inclinação, o que significa que esse item possui grande poder de discriminação separando os indivíduos que respondem corretamente ou não a ele, e valores baixos indicam baixo poder de

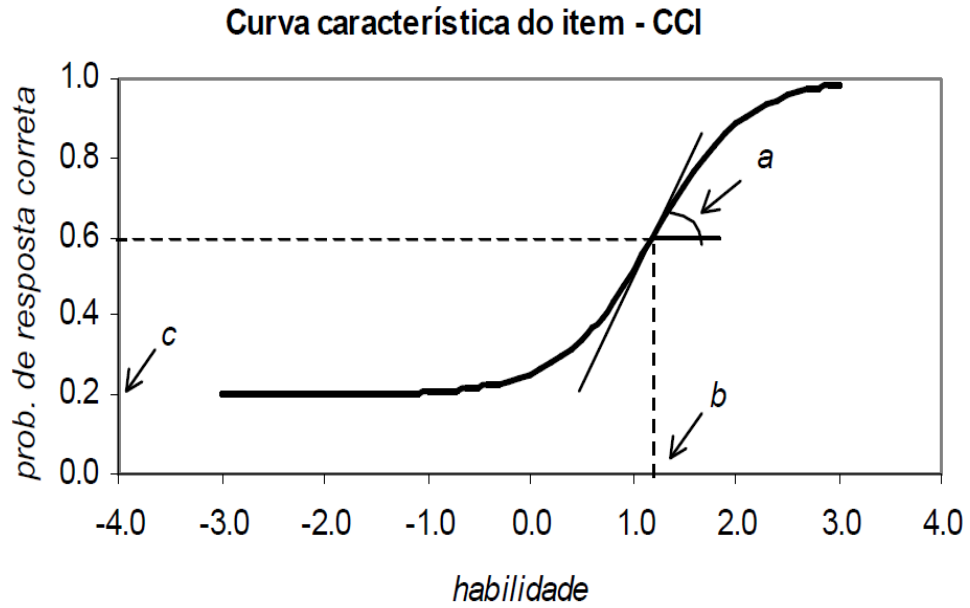


Figura 1.1: *Exemplo de Curva Característica do Item com parâmetros evidenciados* discriminação do item.

Conforme dito anteriormente, os modelos mais simples ML1 (modelo logístico unidimensional de 1 parâmetro, também conhecido como modelo de Rasch) e ML2 (modelo logístico unidimensional de 2 parâmetros), que possuem, respectivamente, 1 e 2 parâmetros, são obtidos de ML3 sem dificuldade. O modelo ML2 por exemplo é obtido de ML3 somente retirando a probabilidade de acerto ao acaso ( $c = 0$ ), sendo dado por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (1.2)$$

com  $i=1,2,\dots,I$  e  $j=1,2,\dots,n$ .

Caso, além de desconsiderar a resposta ao acaso, todos os itens possuam o mesmo poder de discriminação, obtém-se o modelo logístico unidimensional de 1 parâmetro (ML1), que assume a seguinte forma:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}} \quad (1.3)$$

com  $i=1,2,\dots,I$  e  $j=1,2,\dots,n$ .

## Função de Informação do Item

Juntamente com a Curva Característica do Item (CCI), a função de informação do item é uma medida muito utilizada em TRI. Ela serve como análise da quantidade de informação que um item ou teste contém para medir a habilidade do respondente. Essa função possui a forma

$$I_i(\theta) = \frac{\left[ \frac{d}{d\theta} P_i(\theta) \right]^2}{P_i(\theta) Q_i(\theta)}, \quad (1.4)$$

onde  $I_i(\theta)$  é a informação no nível de habilidade  $\theta$  fornecida pelo item  $i$ ,  $P_i(\theta) = P(U_{ij} = 1|\theta)$  e  $Q_i(\theta) = 1 - P_i(\theta)$ .

Para o ML3, a equação acima pode ser escrita da seguinte forma:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[ \frac{P_i(\theta) - c_i}{1 - c_i} \right]^2. \quad (1.5)$$

Tal equação mostra a influência dos três parâmetros no resultado da informação do item, pois observa-se que essa informação aumenta ao passo que

- $b_i$  se aproxima do valor de  $\theta$ ;
- o valor de  $a_i$  aumenta;
- $c_i$  se aproxima de 0.

## Função de Informação do Teste

A soma das informações de cada item compondo o teste resulta na informação fornecida pelo teste, conforme a equação

$$I(\theta) = \sum_{i=1}^I I_i(\theta) \quad (1.6)$$

A função de informação do teste também pode ser representada pelo erro-padrão da medida, conhecida em TRI como erro-padrão de estimação, dado por

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (1.7)$$

Nota-se que todas as medidas de informação citadas até agora dependem do valor de  $\theta$ . Dessa forma, verifica-se que a amplitude do intervalo de confiança para  $\theta$  também dependerá desse valor.

Na figura 1.2 a seguir, estão exemplificadas algumas curvas características de itens e curvas de informação de itens com valores diferentes para os parâmetros.

Percebe-se que quanto maior o valor do parâmetro  $a$  (parâmetro de discriminação ou inclinação) mais acentuada a curva característica se torna. Tal fato resulta em uma diferença de probabilidade de resposta certa maior entre indivíduos com habilidades 1 e 2, sendo tais diferenças iguais a 0,25 e 0,37 nos itens 2 e 4, respectivamente. Ou seja, pode-se afirmar que o item 4 discrimina de maneira mais evidente esses indivíduos que o item 2.

Quanto ao parâmetro  $b$  (parâmetro de dificuldade ou posição), pode-se observar que quanto maior for o seu valor, maior será a habilidade necessária do indivíduo para responder corretamente ao item. Ou seja, um item com maior valor de  $b$  é mais difícil que outro item com um valor menor de  $b$ .

Para cada item, está associado um intervalo na escala de habilidade no qual o item tem maior poder de discriminação. Esse intervalo está demonstrado nos gráficos pelas curvas pontilhadas (curvas de informação), sendo esses intervalos definidos em torno do parâmetro  $b$ . Dessa forma, define-se bons alunos tendo como base os itens considerados difíceis.



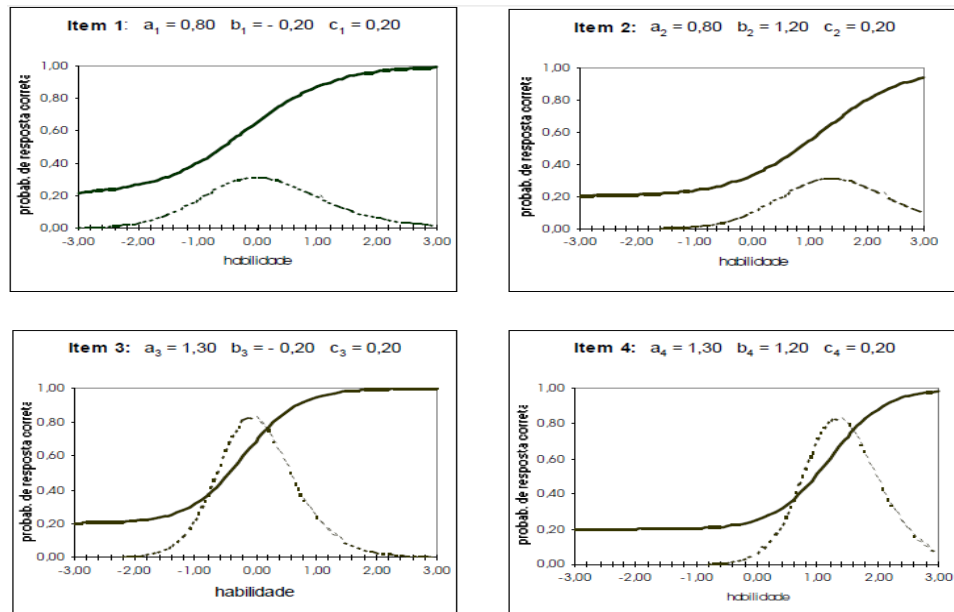


Figura 1.2: *Curvas Características do Item com alterações dos parâmetros*

## Escala de Habilidade

Na TRI, a habilidade pode assumir qualquer valor entre  $-\infty$  e  $\infty$ , o que não acontece na medida score em um teste com  $N$  questões dicotômicas (certo ou errado). Dessa forma, torna-se necessária a criação de uma origem para a escala de habilidade, assim como uma unidade de medida para a definição da escala. Tais valores são escolhidos de modo que representem o valor médio e o desvio-padrão das habilidades dos respondentes.

É comum que se utilize os valores 0 para a média e 1 para o desvio-padrão, e nesse caso o parâmetro  $b$  normalmente varia entre  $-2$  e  $2$ . Para o parâmetro  $a$  são esperados valores entre 0 e 2, sendo valores maiores que 1 os mais adequados.

Apesar da utilização desses valores ser mais comum, não há diferença caso outros valores fossem escolhidos, já que as relações de ordem existentes entre seus pontos são o que realmente importa. Um indivíduo com habilidade 1,2, por exemplo, está 1,2 desvios-padrão acima da habilidade média, que seria o mesmo caso de um indivíduo com habilidade 260 caso a escala escolhida fosse a de média 200 e desvio-padrão igual a 50. A transformação da escala se dá da seguinte forma:

$$a(\theta - b) = (a/50)[(50 \times \theta + 200) - (50 \times b + 200)] = a^*(\theta^* - b^*), \quad (1.8)$$

sendo  $a(\theta - b)$  a parte do modelo probabilístico proposto envolvida na transformação. Para tal, temos:

1.  $\theta^* = 50 \times \theta + 200$
2.  $b^* = 50 \times b + 200$
3.  $a^* = a/50$
4.  $P(U_i = 1|\theta) = P(U_i = 1|\theta^*)$

No caso do item 1 mostrado anteriormente, na escala  $(0, 1)$ , os valores de  $a$  e  $b$  são  $0,8$  e  $-0,2$  respectivamente, e seus correspondentes na escala  $(200, 50)$  são  $0,016$  e  $190$ . Da mesma forma, um indivíduo com habilidade  $1$  na escala  $(0, 1)$  tem na nova escala sua habilidade igual a  $250$ , e

$$\begin{aligned} P(U_1 = 1|\theta = 1) &= 0,2 + (1 - 0,2) \frac{1}{1 + e^{-1,7 \times 0,8 \times (1 - (-0,2))}} \\ &= 0,2 + (1 - 0,2) \frac{1}{1 + e^{-1,7 \times 0,016 \times (250 - 190)}} \\ P(U_1 = 1|\theta^* = 250) &= 0,869 \end{aligned} \quad (1.9)$$

o que significa que a probabilidade de um certo indivíduo responder a um item de maneira correta é a mesma independente da escala usada para medir sua habilidade. Dessa forma, só podemos analisar itens baseando-nos em seus parâmetros se conhecermos a escala na qual eles foram determinados.

### 1.1.2 Modelos para itens não dicotômicos

Nessa seção, serão abordados alguns dos principais modelos tanto para análises de itens abertos, quanto para itens avaliados de forma graduada, que são aqueles

cuja elaboração inclui categorias intermediárias organizadas de forma gradual entre o acerto e o erro. Nesse último tipo, além de avaliar se o respondente acertou ou não a resposta, avalia-se qual foi a resposta dada por ele.

## Modelo de Resposta Nominal

O Modelo de Resposta Nominal foi desenvolvido por Bock (1972) tendo como base o modelo logístico de dois parâmetros e pode ser aplicado a todas as categorias de resposta escolhidas em um teste com itens de múltipla escolha.

Esse modelo tem como objetivo maximizar a precisão da habilidade estimada usando toda a informação presente nas respostas dos indivíduos, não apenas a escolha correta ou não para um determinado item. A probabilidade de um indivíduo  $j$  selecionar uma opção  $k$  como resposta para o item  $i$  é dada por:

$$P_{i,k}(\theta_j) = \frac{e^{a_{i,k}^+(\theta_j - b_{i,k}^+)}}{\sum_{h=1}^{m_i} e^{a_{i,h}^+(\theta_j - b_{i,h}^+)}} \quad (1.10)$$

com  $i = 1, 2, \dots, I, j = 1, 2, \dots, n$ , e  $k = 1, 2, \dots, m_i$ .

Em cada  $\theta_j$ , a soma das probabilidades sobre as  $m_i$  opções de resposta  $\sum_{k=1}^{m_i} P_{i,k}(\theta_j)$  é 1. As quantidades  $(a_{i,k}^+; b_{i,k}^+)$  são parâmetros do item  $i$  referentes à  $k$ -ésima opção. Tal modelo não assume qualquer forma de ordenação a priori das opções de resposta do item.

## Modelo de Resposta Gradual

No modelo de Resposta Gradual, elaborado por Samejima (1969), assume-se uma ordenação para as categorias de resposta dos itens. Assim como o Modelo de Resposta Nominal, esse modelo visa obter mais informação das respostas do que apenas o acerto ou não do indivíduo a um determinado item.

Nesse modelo supõe-se que os escores das categorias de um item  $i$  são ordenados do menor ao maior e denotados por  $k = 0, 1, \dots, m_i$  onde  $(m_i + 1)$  é o número de

categorias do  $i$ -ésimo item. A equação que expressa a probabilidade do indivíduo  $j$  escolher uma dada categoria ou outra mais alta do item  $i$  pode ser expressa por uma extensão do modelo logístico de 2 parâmetros e é dada por

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} \quad (1.11)$$

com  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, n$ , e  $k = 0, 1, \dots, m_i$ , em que  $b_{i,k}$  é o parâmetro de dificuldade da  $k$ -ésima categoria de respostas do item  $i$ .

Os outros parâmetros são semelhantes aos definidos anteriormente para os outros modelos.

Para os casos de itens dicotômicos, o parâmetro de inclinação era chamado de discriminação do item, o que não é válido para itens não dicotômicos, uma vez que nesses casos a discriminação de uma categoria se dá não somente pelo parâmetro de inclinação mas também pela distância das categorias de dificuldades adjacentes.

Pela definição temos

$$b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}, \quad (1.12)$$

o que significa que obrigatoriamente deve haver uma ordenação entre o nível de dificuldade das categorias de dado item  $i$ , de acordo com a classificação de seus escores.

A probabilidade de um indivíduo  $j$  obter um escore  $k$  no item  $i$  é dada por

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j). \quad (1.13)$$

Samejima (1969) também definiu  $P_{i,0}^+(\theta_j)$  e  $P_{i,m_i+1}^+(\theta_j)$  de tal forma que

$$P_{i,0}^+(\theta_j) = 1 \quad \text{e} \quad P_{i,m_i+1}^+(\theta_j) = 0.$$

Dessa forma, obtém-se

$$P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j) = 1 - P_{i,1}^+(\theta_j) \quad (1.14)$$

e

$$P_{i,m}(\theta_j) = P_{i,m}^+(\theta_j) - P_{i,m_i+1}^+(\theta_j) = P_{i,m}^+(\theta_j), \quad (1.15)$$

o que resulta em

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}}. \quad (1.16)$$

Percebe-se que  $m_i$  valores de dificuldade e o parâmetro de inclinação do item precisam ser estimados para um item com  $m_i + 1$  categorias, ou seja, o número de parâmetros a ser estimado é igual ao número de categorias de resposta.

## Modelo de Escala Gradual

O Modelo de Escala Gradual, elaborado por Andrich (1978), é um caso específico do modelo de resposta gradual, sendo da mesma forma aplicado aos itens com categorias de respostas ordenadas. No entanto a diferença é que agora assume-se que os escores das categorias são igualmente espaçados. Tal modelo é dado pela equação:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_k)}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_{k+1})}} \quad (1.17)$$

com  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, n$ , e  $k = 0, 1, \dots, m$ , em que  $b_i$  é o parâmetro de locação do item  $i$ , e  $d_k$  o parâmetro de categoria.

Sendo  $P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j) \geq 0$ , temos que  $d_k - d_{k+1} \geq 0$ , o que significa que  $d_1 \geq d_2 \geq \dots \geq d_m$ .

Como dito anteriormente, a diferença entre esse modelo e o modelo de escala gradual se dá pelas categorias de resposta serem equidistantes. Dessa forma, o parâmetro  $b_{i,k}$  agora é decomposto em um parâmetro  $b_i$  de locação do item, e  $d_k$  um parâmetro de categoria, de forma que  $b_{i,k} = b_i - d_k$ .

Como os parâmetros  $d_k$  não dependem do item, caso os itens pertencentes a um teste tiverem suas próprias categorias de resposta que diferem no número, esse modelo não é apropriado.

## Modelo de Crédito Parcial

O Modelo de Crédito Parcial, criado por Masters (1982), também é utilizado para o caso de respostas múltiplas de categorias ordenadas. A grande diferença desse modelo para o modelo gradual está no fato desse modelo pertencer à família dos modelos de Rasch, sendo ele uma extensão do modelo de Rasch para itens dicotômicos. Sendo assim, todos os parâmetros do modelo são de locação e assume-se poder de discriminação igual para os itens.

Tomando um determinado item  $i$  com  $m_i + 1$  categorias de resposta ordenáveis ( $k = 0, 1, \dots, m_i$ ), esse modelo é determinado pela equação

$$P_{i,k}(\theta_j) = \frac{\exp \left[ \sum_{u=0}^k (\theta_j - b_{i,u}) \right]}{\sum_{u=0}^{m_i} \exp \left[ \sum_{v=0}^u (\theta_j - b_{i,v}) \right]} \quad (1.18)$$

com  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, n$ ,  $k = 0, 1, \dots, m_i$  e  $b_{i,0} \equiv 0$ , em que  $P_{i,k}(\theta_j)$  é a probabilidade de um indivíduo com habilidade  $\theta_j$  escolher a categoria  $k$  dentre as  $m_i + 1$  categorias possíveis do item  $i$ .

O parâmetro  $b_{i,k}$  é o parâmetro do item que regula a probabilidade do indivíduo escolher a categoria  $k$  no item  $i$  ao invés da categoria adjacente ( $k - 1$ ). Cada  $b_{i,k}$  corresponde ao valor de habilidade em que o respondente apresenta a mesma probabilidade de responder às categorias  $k$  e ( $k - 1$ ), ou seja,  $P_{i,k}(\theta_j) = P_{i,k-1}(\theta_j)$ .

No modelo de crédito parcial, para itens com  $m_i + 1$  categorias de resposta, é necessária a estimação de  $m_i$  parâmetros, o que, para itens com apenas duas categorias de resposta, torna o modelo igual ao modelo de Rasch para itens dicotômicos.

Esse modelo indica que a probabilidade de um indivíduo responder à categoria  $k$  é dada como a probabilidade condicional sobre as categorias  $(k - 1)$  e  $k$ . Em outras palavras, o modelo de crédito parcial mostra os pontos na escala de proficiência em que uma categoria de resposta se torna mais provável que a resposta anterior, dado que o indivíduo completou os passos anteriores.

## Modelo de Crédito Parcial Generalizado

O Modelo de Crédito Parcial Generalizado foi elaborado por Muraki (1992), e foi baseado no modelo de crédito parcial criado por Masters (1982), relaxando a hipótese de igualdade do poder de discriminação para todos os itens. A equação que descreve tal modelo é:

$$P_{i,k}(\theta_j) = \frac{\exp\left[\sum_{u=0}^k Da_i(\theta_j - b_{i,u})\right]}{\sum_{u=0}^{m_i} \exp\left[\sum_{v=0}^u Da_i(\theta_j - b_{i,v})\right]} \quad (1.19)$$

com  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, n$ , e  $k = 0, 1, \dots, m_i$ .

Para  $m_i + 1$  categorias de resposta,  $m_i$  parâmetros podem ser identificados. Qualquer valor pode ser atribuído a qualquer um dos  $m_i + 1$  parâmetros de dificuldade, uma vez que o termo incluso no parâmetro é cancelado no numerador e no denominador do modelo. Em geral, define-se  $b_{i,0} \equiv 0$ .

Os parâmetros  $b_{i,k}$  de categoria são os pontos de interceptação na escala  $\theta_j$  das curvas de  $P_{i,k-1}(\theta_j)$  e  $P_{i,k}(\theta_j)$ . Essas curvas se interceptam apenas uma vez e tal intercepto pode ser verificado em qualquer ponto da escala  $\theta_j$ . Assim, assumindo a hipótese de que  $a_i \geq 0$ , observa-se que:

- se  $\theta_j = b_{i,k}$  então  $P_{i,k}(\theta_j) = P_{i,k-1}(\theta_j)$ ;
- se  $\theta_j > b_{i,k}$  então  $P_{i,k}(\theta_j) > P_{i,k-1}(\theta_j)$ ;
- se  $\theta_j < b_{i,k}$  então  $P_{i,k}(\theta_j) < P_{i,k-1}(\theta_j)$ ;

podendo o parâmetro  $b_{i,k}$  ser decomposto como  $b_{i,k} = b_i - d_k$ , de forma análoga ao modelo de escala gradual. A diferença, no entanto, se dá pelo fato de que nesse modelo os valores  $d_k$  não obrigatoriamente são ordenados sequencialmente em um item. Esse parâmetro é a dificuldade relativa da categoria  $k$  em relação às outras categorias do item.

Para esse modelo, itens com  $m_i + 1$  categorias de resposta necessitam da estimação de um parâmetro de inclinação e um parâmetro de locação para cada item e  $m_i$  parâmetros de locação. Assim, um teste com  $N$  itens terá  $\sum_{i=1}^N m_i + 2N$  parâmetros de item a serem estimados.



## Capítulo 2

# Cronologia da TRI e dos testes não paramétricos

A Teoria de Resposta ao Item (TRI) utiliza a modelagem matemática da relação entre o nível de conhecimento do respondente  $\theta$  e as probabilidades de resposta aos itens de um teste  $P(\theta)$ . Nos últimos anos, a família dos modelos de TRI se expandiu de maneira significativa, com a criação de novas técnicas para avaliação de testes, itens, respondentes e para estimação de parâmetros.

Em diversas áreas de aplicação, modelos de estimação paramétrica (PIRT) são usados. Tais modelos assumem uma forma funcional específica para a curva característica do item (CCI), normalmente logística, e proporcionam uma redução substancial dos dados, sendo apenas dois ou três parâmetros necessários para a construção da CCI. No entanto, a utilidade dos modelos paramétricos dependem do grau de correspondência entre o modelo para estimação da CCI e evidência empírica.

Por definição a Curva Característica do Item é não observável, e sua estimação exige que se suavize o ruído sem viesar o sinal. Uma das primeiras aplicações de métodos de avaliação alternativos aos clássicos foi utilizado por Lord & Novick (1968). Nesse estudo, gráficos de regressões para itens de um teste foram usados, baseados em respostas dadas a um teste composto por 89 itens de vocabulário e 59 itens de matemática com 103.275 respondentes. Esse grande número de examinados permitiu a obtenção de estimadores confiáveis para a proporção de respostas corretas aos itens para cada grupo de escore. No entanto, ficou claro que para esse método não pa-

ramétrico de estimação da FRI seriam necessárias adaptações antes que ele pudesse ser aplicado a dados de atitude (traço latente), que frequentemente consistem em ter entre 100 a 500 pessoas e 4 a 14 itens.

Na estimação de CCIs através da proporção de respostas corretas em grupos de escore restante (excluindo-se o item sendo analisado), há uma delicada troca entre o poder contra os vários desvios, o que requer uma estrutura de grupos refinada, e estabilidade, o que requer um tamanho grande para os grupos. Molenaar & Sijtsma (2000) recomendaram utilizar um número mínimo para os grupos diferente do que se utilizava e confiar nos resultados caso tal mudança não tivesse tanto efeito. Caso esse efeito fosse muito forte, deveria ser feita uma inspeção gráfica e das estatísticas no software MSP5 (software utilizado por eles no estudo da TRI) para verificar a causa desse efeito, mas uma conclusão definitiva ainda seria difícil de ser obtida.

A avaliação do ajuste do modelo exige dois processos complementares: verificar alguns pressupostos como unidimensionalidade ou independência local e avaliar a concordância entre as observações e as predições do modelo. O ajuste do modelo deve ser avaliado tanto ao nível de teste quanto ao nível de item. Em geral, a avaliação do modelo ao nível de item tende a ser mais rigorosa que ao nível de teste, já que espera-se que nem todas as condições e requisitos sejam satisfeitos.

Existem diversos procedimentos propostos para a avaliação da adequabilidade do ajuste das CCIs. No entanto, muitos deles apresentam desvantagens. Em alguns métodos mais tradicionais, como no proposto por Yen (1981), no qual os respondentes são agrupados tendo como base seus valores estimados de  $\theta$ , é possível encontrar erros. Como consequência disso, muitos itens são erroneamente identificados como não ajustados. Embora alguns outros métodos tenham conseguido superar esse problema, como os propostos por Orlando & Thissen (2000) e Stone (2000), muitos desses métodos utilizam estatísticas  $\chi^2$  de Pearson ou estatísticas baseadas no  $\chi^2$ , que são muito sensíveis ao tamanho da amostra, o que faz com que caso o modelo não seja muito confiável, por exemplo, o  $\chi^2$  normalmente será significativo para grandes amostras.

Outro motivo para a agregação de dois grupos, realizada frequentemente, além dos valores estimados de  $\theta$ , é o pequeno número de respondentes dentro desses grupos. Esse agrupamento foi inspirado pelo algoritmo de “pool adjacent violators”, conhecido

como algoritmo PAVA na regressão isotônica (Barlow, Bartholomew, Bremner, & Brunk, 1972) e outros procedimentos semelhantes utilizando adequabilidade do ajuste através do  $\chi^2$  dos dados amostrais.

Uma possibilidade de avaliação do ajuste de modelos proposta por Swaminathan et al. (2007) sugere que se utilize a análise gráfica em conjunto com procedimentos estatísticos. Outra estratégia, sugerida por Douglas & Cohen (2001), é comparar a CCI paramétrica com a CCI estimada através de um modelo de TRI não paramétrico (NIRT). Os modelos não paramétricos não assumem uma forma funcional particular para a CCI. Dessa forma, modelos não paramétricos podem ser usados para estimar as CCIs sem restrições de forma e podem gerar uma estimativa mais confiável para a verdadeira CCI em testes mais longos e tamanhos de amostra não muito grandes sob suposições fracas (Douglas, 1997).

Durante a construção de um teste, alguns aspectos como acurácia, facilidade na interpretação, aplicabilidade e validade, nem sempre apontam para a mesma direção quando se trata da decisão a ser tomada. Dessa forma, a construção de uma escala satisfatória na Teoria de Resposta ao Item Não Paramétrica depende da capacidade de encontrar um balanço entre tais requisitos. A TRI oferece uma possibilidade para esse processo, e os modelos de TRI não paramétricos podem ser bastante úteis, já que cresceram de maneira significativa e abrangem uma grande quantidade de métodos baseados em suposições brandas para os modelos.

Junker & Sijtsma (2001) propuseram três motivos para a utilização dos modelos não paramétricos:

1. Disponibilizar um conhecimento mais profundo sobre como os modelos paramétricos funcionam;
2. Oferecer uma possibilidade mais abrangente de aplicações nas quais uma família de modelos paramétricos se ajusta de maneira não satisfatória;
3. Criar procedimentos de fácil utilização para dados com números muito menores de itens e respondentes que os usados em testes de larga escala.

Algumas aplicações do NIRT para medição de traços como atitude e personalidade foram utilizadas por Mokken (1971) e Molenaar (1991). Tendo como inspiração a

razão 3 listada anteriormente, eles preferiram trabalhar com medidas globais simples, como popularidades de itens ou a estatística  $H$  de Loevinger (1948), que é tanto o número de erros de Guttman redimensionado como a correlação ou covariância redimensionada. Já outros, inspirados pelas razões 1 e 2, como Bolt (2001), Douglas & Cohen (2001), Ramsay (1991), Stout (1990), utilizaram o NIRT para medição da habilidade e desempenho dos respondentes. Tais autores em geral possuem informação suficiente para estimar as funções de resposta ao item ou as matrizes de covariância condicionadas aos valores do traço latente.

Independente da motivação, ambas as abordagens aparentemente possuem muito em comum. As duas propostas apresentam o interesse comum de determinar quão útil a teoria do traço latente pode ser quando assume-se apenas monotonicidade das funções de resposta ao item, sem forçá-las a possuírem formatos logísticos ou ogivas da distribuição normal. Esse é um desafio para matrizes com grande quantidade de dados, mas ainda mais para pequenas matrizes nas quais muito refinamento na análise causa instabilidade.

Uma das vantagens da estimação não paramétrica é que ela não necessita que os respondentes sejam agrupados pelo seus valores estimados de  $\theta$ , e fornece uma exibição gráfica conveniente de possíveis desajustes. Além disso, a raiz da diferença quadrática média entre a CCI não paramétrica e a CCI paramétrica pode ser computada através dos valores de  $\theta$ . Esse índice é conhecido pela sigla RISE (do inglês, “root integrated squared error”) ou RISD (“root integrated squared difference”), e sua significância pode ser aproximada através de procedimentos bootstrap paramétricos, conforme sugerido por Douglas & Cohen (2001).

Mokken (1971) sugeriu a utilização do coeficiente  $H$  de Loevinger (1948), que assume seu máximo valor de 1 quando a tabela  $2 \times 2$  de respostas para um par de itens dicotômicos não possui erros de Guttman (erros nos quais itens fáceis são respondidos de maneira errada, e itens difíceis são respondidos corretamente). Isso serve também para os casos em que dificuldades diferentes dos itens mantêm a correlação máxima em um valor baixo. O coeficiente  $H$ , como já foi dito, é tanto o número de erros de Guttman redimensionado como a correlação ou covariância redimensionada, e ambas as interpretações podem ser estendidas para um coeficiente ao nível de item e de escala, em que  $H$  é a chave para a construção tanto para a versão exploratória quanto para

a confirmatória da análise de escala de Mokken.

Muitas das propriedades de  $H$  foram provadas por Mokken (1971), e outros autores desenvolveram refinamentos e extensões para o coeficiente. Molenaar (1982, 1991) obteve uma extensão de  $H$  de itens dicotômicos para politômicos. Snijders (2001) trabalhou com a estrutura multi-nível entre pessoas (por exemplo, quando as pessoas classificam seus vizinhos ou amigos, o coeficiente  $H$  é calculado tanto dentro dos grupos de vizinhos e amigos como entre os grupos). Sijtsma (1998) destaca alguns dos principais trabalhos desenvolvidos a partir do coeficiente  $H$  de Loevinger.

Stout (1987, 1990) substituiu a covariância de um item por uma covariância condicional dado o valor do traço latente. Isso foi estimado através do condicionamento de um subteste adequado selecionado (DIMTEST; presente no trabalho de Stout, Goodwin-Froelich, & Gao, 2001), ou do escore total do teste, ou do escore restante excluindo-se os dois itens em que a associação foi considerada (DETECT, utilizado por Zhang & Stout, 1999a, 1999b).

Para grupos de baixo escore ou baixo escore restante, essas estimações da covariância local são mais instáveis que as estimações locais de dificuldade do item. Agregar grupos adjacentes causaria problemas, pois a heterogeneidade em valores verdadeiros de traços latentes cresceria em excesso. O subteste DETECT omite grupos pequenos, mas são necessárias pesquisas em outras soluções.

Associações em pares de itens condicionais e não condicionais proporcionaram medidas de associação mais efetivas ao nível de população, assim como estimadores mais confiáveis ao nível de amostra. Habing (2001) explorou mudanças sistemáticas na covariância condicional através da amplitude do traço latente. Bolt (2001) apresentou representações gráficas baseadas em todas as covariâncias condicionais dos pares. Croon (1991), Hoijsink & Molenaar (1997), e Vermut (2001) começaram a utilizar uma metodologia de classes latentes em NIRT, facilitando o estudo de propriedades de amostragem das estimações e ajustes das estatísticas, que podem ser úteis em pesquisas saindo das associações entre pares e indo para distribuições de probabilidade para trios de itens ou todos os itens simultaneamente.

Métodos e algoritmos mais modernos, como suavização via kernel, bootstrap paramétrico, algoritmo EM ou MCMC (Markov Chain Monte Carlo), possibilitam novos estudos em NIRT, reduzindo a dependência em distribuições normais para o traço la-

tente, ou ogivas normais e logísticas. Junker (2001) e Junker & Sijtsma (2001b) mostraram como até modelos detalhados de processamento mental levando a um certo padrão de resposta necessitam não ser baseados em famílias paramétricas de FRIs. Huisman (1999), Huisman & Molenaar (2001) e Laros & Tellegen (1991) deixaram claro que o modelo não paramétrico de Mokken (1971) pode ser utilizado para testes adaptativos e imputação de dados faltantes. Com o objetivo de encontrar um teste computacional adaptável não paramétrico, são necessários métodos para armazenar FRIs não paramétricas em bancos de itens para definir funções de informação no contexto não paramétrico.

Van der Ark (2001), Sijtsma & van der Ark (2001), e Hemker, Sijtsma, Molenaar, & Junker (1996, 1997) demonstraram que o domínio de itens politômicos é mais complexo que o de itens dicotômicos. Embora ganhar mais pontos em um mesmo item seja numericamente equivalente a ganhar mais pontos em mais itens, isso pode ser diferente psicometricamente. Esses autores mostraram que o modelo de resposta gradual não paramétrico, usado previamente para expandir o modelo de Mokken para itens politômicos (Molenaar, 1982, 1991), é o mais abrangente dentre todos os modelos politômicos de TRI.

Usar mais de duas categorias de respostas ordenadas pode aumentar a confiança do teste e diminuir a frustração do respondente, particularmente em medições de personalidade e atitude. Além disso, pode também aumentar o perigo de mal ajuste do modelo e de pares de respondentes incorretamente ordenados.

As metodologias não paramétricas para adequabilidade de ajuste ao nível de pessoas e para estimação da confiança também existem, como em Meijer (1994), Meijer & van Krimpen-Stoop (2001) para o primeiro caso, e Meijer, Sijtsma, & Molenaar (1995), e Sijtsma & Molenaar (1987) para o segundo caso. O equacionamento do teste requer a predição do escore do respondente em um teste ainda não feito, e isso é feito utilizando o NIRT já que a imputação é possível para o par item-respondente que ainda não foi observado, conforme Huisman & Molenaar (2001). Outra área em que há ainda poucas pesquisas é na verificação de modelos para os três pressupostos do NIRT, unidimensionalidade, independência local e monotonicidade.

O método de suavização via Kernel é um dos mais usados para estimação não paramétrica de CCIs com a finalidade de verificar a adequabilidade do ajuste. Wells

& Bolt (2008) compararam o RISE da suavização via Kernel com duas estatísticas de ajuste muito utilizadas, o  $S - X^2$  proposto por Orlando & Thissen (2000) e o  $G^2$  BILOG de Mislevy & Bock (1990), em várias condições diferentes como a extensão do teste, tamanho da amostra e porcentagem de itens mal ajustados. O ajuste do modelo logístico de 2 parâmetros (ML2) foi avaliado, e os itens mal ajustados seguiam o ML3, o modelo de resposta nominal misto, e o modelo cosseno hiperbólico, significando desajuste em diferentes pontos da escala de habilidade. Os índices RISE e o  $S - X^2$  controlaram o erro do Tipo 1, mas o RISE se mostrou um teste mais poderoso dadas as condições propostas.

Ramsay (1991) utilizou métodos de suavização via kernel para dois testes de 100 itens feitos por aproximadamente 300 respondentes. Foi observado que algumas das CCIs eram muito diferentes do modelo logístico, e posteriormente, aplicações do software TESTGRAF (Ramsay, 1993), revelaram conclusões semelhantes. Douglas & Cohen (2001) também mostraram a utilidade dos métodos de suavização via Kernel. Tais métodos necessitavam da escolha de um parâmetro de suavização (janela) e de uma função Kernel. Assim, realizar uma análise sensitiva através da variação dessas escolhas é algo a se levar em consideração, principalmente nos raros casos em que as decisões importantes de qual teste será usado forem baseadas apenas na estimação da FRI. Com o desenvolvimento dos métodos computacionais, é possível obter detalhes acerca de CCIs a serem estudadas. No entanto, isso não deve causar um excesso de observações sobre pequenos erros, pois nesses casos todos os modelos estão de alguma forma incorretos. A questão principal deve ser o quanto a discrepância do modelo ajustado influencia de maneira significativa as conclusões a serem tomadas.

Embora em geral a suavização via Kernel tenha apresentado melhores resultados nos estudos existentes, é preciso ressaltar que a CCI empírica pode ser obtida também através da estimação da probabilidade a posteriori de uma resposta correta como a proporção  $A_q/N_q$  onde  $A_q$  é a “pseudocontagem” da quantidade de respondentes com habilidade  $\theta_q$  que respondem corretamente o item e  $N_q$  é a “pseudocontagem” do número de respondentes com habilidade  $\theta_q$ . Essas pseudocontagens são computadas no *Expectationstep* (E step) e utilizadas no passo de maximização (M step) do algoritmo EM para a estimação dos parâmetros dos itens (Mislevy & Bock, 1990).

Outro estudo que merece destaque foi realizado por Molenaar & Sijtsma (2000),

que implementaram no software MSP5 um método simples de suavização no qual agregavam grupos de escore vizinhos caso eles fossem muito pequenos. De acordo com o estudo, quando se trabalha com um número entre 4 e 14 itens, é importante que se reduza o viés através da substituição do escore total do teste pelo escore descon siderando o item a ser avaliado. Para verificar o modelo através do MSP5, é possível fazer a inspeção visual das curvas estimadas, observar o número, tamanho e significância das violações ou conferir um resumo das estatísticas obtidas combinando as informações que se possui em um valor geral por item. Embora os resultados possam ser instáveis para testes pequenos e pequenos tamanhos de amostras, o procedimento normalmente é satisfatório em distinguir CCIs verdadeiramente não monótonas de variações da amostra.

Sueiro & Abad (2011) propuseram usar a distribuição a posteriori do traço latente como um modelo não paramétrico e comparar o desempenho de um índice baseado nesse método com outro modelo baseado na suavização via kernel. Nesse estudo as taxas de erro e poder foram avaliadas utilizando o ML2 e três tipos de maus ajustes aos itens. Os resultados apresentados mostraram que para os itens ajustados as distâncias entre as CCIs paramétricas e não paramétricas diminuíram a medida que o tamanho da amostra foi aumentado para ambos os casos. O RISE da suavização via Kernel também decresceu com o aumento do tamanho do teste. Métodos bootstrap foram usados para obter um teste de significância, e os dois procedimentos apresentaram desempenhos satisfatórios quanto às taxas do erro tipo 1. Quanto ao poder do teste o método das probabilidades a posteriori foi superior, principalmente para pequenas amostras, embora em testes curtos ambos os procedimentos tenham apresentado desempenhos parecidos.

Pode-se observar que diversos testes já foram utilizados para não só estimar a curva característica do item, mas para vários outros objetivos na TRI, com diversas abordagens, diferentes esquemas de amostragem e utilização de métodos computacionais. Neste trabalho, um novo teste baseado na estimação não paramétrica da CCI sob a suposição de monotonicidade é apresentado e testado. Esse teste utiliza a teoria da regressão isotônica para o cálculo da estrutura da CCI, verifica-se seu ajuste à curva e compara-se seu desempenho com os outros testes a serem implementados.



# Capítulo 3

## Testes de adequabilidade de ajuste para modelos de Teoria de Resposta ao Item

### 3.1 Teste baseado em suavização por kernel (Ramsay, 1991)

Ramsay(1991) explorou o uso da suavização por kernel para estudar a relação entre o ranking de habilidade dos indivíduos e a variável binária indicando a resposta correta ou não de um determinado item. Para tanto, a análise foi dividida em quatro etapas:

1. Ranqueamento: Estimar o ranking  $r_j$  do  $j$ -ésimo indivíduo ranqueando os valores  $T_j$  de uma estatística  $T$ .
2. Substituir os rankings  $r_j$  pelos quantis de uma distribuição que parece apropriada para os dados em questão. A distribuição Normal padrão é a mais utilizada, mas autores que pensam na habilidade como porcentagem preferem distribuições como a Beta.
3. Ordenação: Ordenar os padrões de resposta dos indivíduos  $(X_{i1}, \dots, X_{in})$  pelo ranking estimado do escore, ou seja, o  $i$ -ésimo padrão de resposta no teste

ordenado  $(X_{(i)1}, \dots, X_{(i)n})$  correspondente ao escore de habilidade do  $i$ -ésimo examinado pelo tamanho  $T_i$ .

4. Suavização: Estimar a probabilidade de resposta positiva do  $m$ -ésimo indivíduo para o  $i$ -ésimo item suavizando a relação entre o vetor de tamanho  $I$  de indicadores de opção do item escolhido e o vetor das habilidades nos pontos de quadratura  $i$ .

O escore total do indivíduo é a estatística mais óbvia para ser utilizada como ranqueamento. Como Mokken (1971) apontou, a monotonicidade da curva característica do item para todos os itens usados para calcular o escore final é uma condição suficiente de consistência do escore final como base para o ranking conforme cresce o número de itens.

O problema em utilizar o escore total está no fato de que para testes menores e com grande número de respondentes, podem haver muitos empates. Para minimizar possíveis vieses devido à ordem em que os resultados desses testes são gravados, pode ser melhor aleatorizar a ordenação dos respondentes dentro do grupo do escore final.

No entanto, o escore total não é um bom indicador da habilidade do indivíduo ou uma boa base para o ranqueamento uma vez que não considera a dificuldade do item ou a informação disponível nas opções de resposta. Uma estatística utilizada para o ranqueamento é a seguinte:

$$W_{im} = \text{logit}(P_{im}^{(75)}) - \text{logit}(P_{im}^{(25)}), \quad (3.1)$$

onde  $P_{im}^{(75)}$  e  $P_{im}^{(25)}$  são as proporções de indivíduos com os 25% maiores e menores escores finais, respectivamente, que escolheram a opção  $m$  para o item  $i$  em um determinado teste de múltipla escolha.

Assim, a estatística utilizada é

$$T_j = \sum_i \sum_m Y_{im}^j W_{im}, \quad (3.2)$$

sendo  $T_j$  então a soma dos pesos  $W_{im}$  para as opções escolhidas.

Para a suavização,  $\hat{P}(\theta)$  é a média ponderada dada por

$$\hat{P}_{im}(\theta) = \sum_{j=1}^J w_j(\theta) Y_{im}^{(j)}, \quad (3.3)$$

onde os pesos  $w_j(\theta)$  são definidos de forma que sejam máximos quando  $q_j = \theta$ .

Os pesos adequados são definidos em termos de uma função kernel  $K(u)$  que é máxima em  $u = 0$  e que convirja para zero em ambos os lados. Também é fundamental um parâmetro de discriminação  $h$  que controle quão rápido o peso atinge o valor zero, definido por

$$w_j(\theta) \propto K\left[\frac{q_j - \theta}{h}\right] \quad (3.4)$$

A necessidade de manter  $\hat{P}$  entre  $[0, 1]$  impõe as restrições adicionais  $w_j(\theta) \geq 0$  e  $\sum_j w_j(\theta) = 1$ , e como consequência, é preferível utilizar funções kernel  $K$  não negativas e os pesos de Nadaraya-Watson definidos como

$$w_j(\theta) = \frac{K[(q_j - \theta)/h]}{\sum_k K[(q_k - \theta)/h]}. \quad (3.5)$$

Dessa forma, a estimação via suavização se torna

$$\hat{P}_{im}(\theta) = \frac{\sum_j K\left(\frac{q_j - \theta}{h}\right) Y_{im}^{(j)}}{\sum_j K\left(\frac{q_j - \theta}{h}\right)} \quad (3.6)$$

## 3.2 Teste baseado na distribuição Qui-quadrado (Orlando & Thissen, 2000)

Orlando and Thissen (2000) apresentaram novos índices de adequabilidade de ajuste para modelos da TRI dicotômicos. Foram testados o comportamento do  $\chi^2$  de Pearson ( $S - \chi^2$ ) e a razão de verossimilhança  $G^2(S - G^2)$  através de simulações e comparados com dois índices similares aos já utilizados ( $Q_1 - \chi^2$  e  $Q_1 - G^2$ ).

Tendo definido que TRI é uma coleção de técnicas de modelagem para análise de itens, testes e respondentes, o modelo para respostas a itens dicotômicos geralmente especifica que a probabilidade de um padrão de resposta  $x$  é dado por

$$P(x) = \int \prod_i \{T_i(\theta)^{x_i} [1 - T_i(\theta)]^{1-x_i}\} \phi(\theta) \partial\theta \quad (3.7)$$

onde  $x$  é o vetor resposta,  $T_i(\theta)$  é a probabilidade de acerto no item  $i$  como função da habilidade  $\theta$ , e  $\phi(\theta)$  é a distribuição da população para  $\theta$ .

Estatísticas de adequabilidade de ajuste para os modelos ML2 e ML3 já haviam sido construídas. Para computar tais estatísticas, os dados observados devem ser organizados de maneira a serem comparados com as previsões do modelo. O procedimento para a construção dessas medidas geralmente se dá pelos seguintes passos:

1. Estimar  $\theta$  e os parâmetros dos itens
2. Ordenar os respondentes de acordo com seus valores estimados de  $\theta$
3. Formar subgrupos dos respondentes ordenados
4. Calcular a proporção de respondentes em cada subgrupo que acertou e errou cada item
5. Comparar essas proporções observadas com as preditas pelo modelo utilizando uma estatística  $\chi^2$  ou representação gráfica

Uma dessas medidas,  $Q_1$ , proposta por Yen(1981), é dada por

$$Q_{1i} = \sum_{k=1}^{10} \frac{N_k(O_{ik} - E_{ik})^2}{E_{ik}} + \sum_{k=1}^{10} \frac{N_k[(1 - O_{ik}) - (1 - E_{ik})]^2}{1 - E_{ik}} = \sum_{k=1}^{10} \frac{N_k(O_{ik} - E_{ik})^2}{E_{ik}(1 - E_{ik})} \quad (3.8)$$

Para o item  $i$ , as proporções observadas ( $O_{ik}$ ) são obtidas através da partição da escala de  $\theta$  em 10 intervalos de tal forma que o número de respondentes em cada intervalo seja aproximadamente igual. Após isso, as respostas observadas para cada intervalo são inseridas em uma tabela de contingência  $10 \times 2$ . As proporções esperadas são computadas do modelo como a probabilidade média predita de resposta correta em

cada intervalo. Sob  $H_0$ , a estatística  $Q_1$  é distribuída como uma  $\chi^2$  com  $10 - m$  graus de liberdade (onde  $m$  é o número de parâmetros do item) sob certas circunstâncias, conforme mostrado por Yen (1981). Bock (1972) também utilizou uma medida de ajuste ao item baseado na distribuição  $\chi^2$ , mas essa difere de  $Q_1$  em alguns aspectos, como o número de intervalos que é variável, e na estimação das proporções esperadas, que é computada através da mediana ao invés da média.

A razão de verossimilhança  $G^2$  desenvolvida por Mckinley & Mills (1985) é baseada em passos semelhantes aos realizados para a obtenção de  $Q_1$ . Os respondentes são ordenados e particionados em 10 intervalos de tamanhos iguais de acordo com seus  $\theta$  estimados. As respostas corretas e erradas de cada grupo são computadas e obtém-se  $G^2$  através de

$$G_i^2 = 2 \sum_{k=1}^{10} N_k \left[ O_{ik} \ln \left( \frac{O_{ik}}{E_{ik}} \right) + (1 - O_{ik}) \ln \left( \frac{1 - O_{ik}}{1 - E_{ik}} \right) \right], \quad (3.9)$$

com  $10 - m$  graus de liberdade.

Embora estatísticas de ajuste ao item sejam úteis na detecção de desajustes de modelos, taxas aceitáveis do erro Tipo  $I$  para essas estatísticas ainda não foram estabelecidas de maneira clara. O problema é que estas estatísticas para ajuste do item não são construídas da mesma maneira que estatísticas de adequabilidade  $\chi^2$  tradicionais. Com exceção de algumas estatísticas para o modelo de Rasch, as medidas de ajuste dependem em algum grau dos estimadores de  $\theta$  para obter as proporções observadas, que devem estar disponíveis antes que o modelo esteja ajustado. Proporções observadas de dependentes modelos fazem com que seja difícil determinar a distribuição da estatística de ajuste. Isso se deve ao fato da incerteza quanto aos graus de liberdade para aproximações  $\chi^2$ . A perda de graus de liberdade devido a estimativa de  $\theta$  é insignificante para qualquer item individualmente (Yen, 1981).

Outro problema é o método pelo qual respondentes são agrupados em intervalos. Tais agrupamentos em grupos de tamanhos iguais é altamente dependente da amostra, o que faz com que os pontos de corte e também o número de intervalos afetem a estatística de ajuste a ser obtida.

Poderia ser obtido um índice melhor caso os respondentes fossem agrupados de acordo com alguma característica dos dados observados ao invés do  $\theta$  estimado, como as proporções observadas de respostas corretas e incorretas ou o escore bruto. As proporções esperadas mais acessíveis de ML2 e ML3 estão associados com os padrões de resposta, não com o escore bruto. Dessa forma, uma comparação direta dessas proporções esperadas com os dados observados não é viável. No entanto, uma tabela de respostas esperadas pode ser obtida das previsões dos modelos ML2 e ML3 para cada item e intervalo de  $\theta$ . Essa tabela pode então ser comparada com os dados observados.

Devido a dependências entre tabelas de contagens observadas para cada item no mesmo teste, o número de graus de liberdade associados a uma estatística calculada a partir da classificação cruzada de grupos de pontuação e respostas dos itens permanece incerta. No entanto, este método apresenta uma vantagem sobre estatísticas em que as frequências observadas são apenas uma função dos dados e os respondentes não precisam ser agrupados de uma forma arbitrária e dependentes do modelo. Este método também pode ser estendido para modelos politômicos. O desafio é obter as frequências esperadas de respostas corretas e incorretas para cada escore bruto de cada item. O método desenvolvido para calcular tais frequências é realizado de forma a:

1. Introduzir duas novas estatísticas de ajuste baseadas nesse novo método computacional,
2. Examinar a performance desses novos índices com dados ajustados e desajustados,
3. Comparar o comportamento desses índices com os índices de  $\chi^2$  de Pearson e a razão de verossimilhança  $G^2$ .

Um método de predição de distribuições de verossimilhança conjuntas para cada escore bruto foi brevemente descrito por Lord e Wingersky (1984), e posteriormente desenvolvido por Thissen, Pommerich, Billeaud e Williams (1995). Esse método utiliza um algoritmo recursivo que constrói a verossimilhança conjunta para cada grupo dividido por escore, item a item.

Para a construção da estatística ajustada ao item, o objetivo é determinar uma verossimilhança para cada escore possível desconsiderando um item por vez, para cada item (caso em que o número de verossimilhanças é o produto do número de itens pelo número de escores possíveis). Em seguida o item retirado é acrescentado novamente para se obter a proporção de examinados com escore  $k$  que acertaram o item  $i$ . Primeiramente se obtém as distribuições de verossimilhança para cada escore, e então esse resultado é utilizado para se obter a distribuição de verossimilhança para cada escore sem determinado item.

Para modelos dicotômicos as distribuições de verossimilhança para os escores são obtidas da maneira descrita a seguir. As verossimilhanças para os escores 0 e 1 são pré-determinadas de modo a serem iguais às funções de resposta do item para respostas certas e erradas no primeiro item, de tal forma:

$$S_0^* = 1 - T_1 \quad \text{e} \quad S_1^* = T_1, \quad (3.10)$$

onde  $S_k^*$  é o valor interino da verossimilhança do escore  $k$  e  $T_1$  é a função de resposta ao item para a resposta correta ao item 1. Vale ressaltar que tais notações suprimem o parâmetro  $\theta$ , ou seja,  $S_k$  e  $T_i$  são funções de  $\theta$ ,  $S_k(\theta)$  e  $T_i(\theta)$ , e não valores escalares.

Em seguida, adiciona-se cada item  $i$  ao teste

$$S_0 = (1 - T_i)S_0^* \quad \text{e} \quad S_k = T_i S_{k-1}^* + (1 - T_i)S_k^* \quad (3.11)$$

para cada  $k = 1, 2, \dots, i - 1$ ; e

$$S_i = T_i S_{i-1}^* \quad (3.12)$$

Após a adição de cada item, o novo  $S_k$  substitui  $S_k^*$  para todos os escores computados para o item anterior.

Na última iteração do algoritmo, as verossimilhanças conjuntas foram acumuladas para cada escore exceto o último, tornando a verossimilhança conjunta para o escore  $k$  para todos os itens

$$S_k = T_{\text{último}} S_{k-1}^* + (1 - T_{\text{último}}) S_k^* \quad (3.13)$$

onde

$S_k$  é a distribuição posterior do escore para o grupo  $k$ ;

$T_{\text{último}}$  é a Função de Resposta ao Item do último item;

$S_{k-1}^*$  é a distribuição posterior do escore para o grupo  $k - 1$  sem o último item, e;

$S_k^*$  é a distribuição posterior do escore para o grupo  $k$  sem o último item.

Uma variação desse algoritmo recursivo fornece as proporções preditas do modelo de acertos e erros para cada item para cada escore final. Após a equação anterior ser computada, o algoritmo recursivo é repetido para todo  $i$ , onde um item é omitido em cada iteração, gerando as verossimilhanças conjuntas para grupo de escore sem o item  $i$  ( $S_k^{*i}$ ).

Uma vez que as verossimilhanças dos escores com todos os itens e com a exclusão de cada item foram obtidas, elas podem ser combinadas com cada item omitido anteriormente para verificar a proporção desejada de respondentes com escore  $k$  que responderam corretamente o item  $i$ .

$$E_{ik} = \frac{\int T_i S_{k-1}^{*1} \phi(\theta) \partial \theta}{\int S_k \phi(\theta) \partial \theta} \quad (3.14)$$

As integrais da equação acima são aproximadas usando quadratura retangular sobre incrementos igualmente espaçados de  $\theta$  de  $-4,5$  a  $4,5$ . Outros métodos para avaliação numérica de tais integrais podem ser encontrados em Stroud(1974).

Quanto aos novo índices propostos por Orlando e Thissen (2000), sua utilização se deve ao fato de nunca terem sido encontrados resultados claros acerca da utilidade da comparação entre os índices  $\chi^2$  e  $G^2$ , por isso em seu artigo, Orlando e Thissen apresentaram ambos os índices. O índice  $\chi^2$  é descrito por

$$S - \chi_i^2 = \sum_{k=1}^{n-1} N_k \frac{O_{ik} - E_{ik}}{E_{ik}(1 - E_{ik})}, \quad (3.15)$$

onde as proporções observadas ( $O_{ik}$ ) para o item  $i$  e escore  $k$  são computadas dos dados, e as proporções esperadas ( $E_{ik}$ ) são computadas através das equações 3.13 e 3.14. A soma é sobre o número possível de escores menos dois, já que a proporção



de respondentes que acertam e tem escore 0 é 0, e a proporção de respondentes que erram e tem escore 1 é 1. O índice  $G^2$  proposto é

$$S - G_i^2 = 2 \sum_{k=1}^{n-1} N_k \left[ O_{ik} \ln \frac{O_{ik}}{E_{ik}} + (1 - O_{ik}) \ln \frac{1 - O_{ik}}{1 - E_{ik}} \right], \quad (3.16)$$

em que as proporções observadas e esperadas são definidas da mesma forma que em 3.15.

## Frequência mínima das células

Um problema que pode ocorrer ao se trabalhar com tal método é o caso em que as frequências esperadas de respostas certas e erradas são muito pequenas para um dado intervalo de escore. Frequências pequenas nas linhas das tabelas usadas para construir as estatísticas ajustadas podem causar um decréscimo na acurácia da aproximação de  $\chi^2$  para sua distribuição. A fim de evitar tal problema, um algoritmo foi criado para avaliar a quantidade de valores esperados. Esse algoritmo começa seu processo agregando os valores das células a partir das extremidades (pelo maior e menor escores ou grupos de escore) até o centro da tabela até que todas as caselas tenham uma contagem suficientemente grande. Realiza-se o mesmo processo com a tabela dos valores observados para que possa ser feita a comparação.

Vários livros na literatura como o de Cochran (1952) sugeriram um mínimo de 5 para frequência das células. Larntz (1978) escreveu que o  $\chi^2$  aparentava alcançar a frequência desejada de rejeição sob a condições nulas, em que todas as frequências esperadas das células eram maiores que 1. Já a estatística  $G^2$  é mais sensível a pequenos valores esperados nas células. Larntz (1978) observou que, com muitas células contendo frequências esperadas entre 1,5 e 4, a estatística  $G^2$  rejeitava a hipótese nula muito frequentemente.

Com base em tais fatos, Orlando e Thissen (2000) realizaram análises prévias nas quais compararam resultados obtidos com frequência mínima das células de 1 e 4. Os resultados mostraram que o  $\chi^2$  apresentou desempenho satisfatório com uma frequência mínima de 1, o que não ocorreu para  $G^2$ . Com a frequência mínima igual a 4,  $G^2$  apresentou melhoras mas ainda assim resultados insatisfatórios. Mesmo com a

implementação do algoritmo, foi observado que após a agregação sobravam poucas linhas para o cálculo dos graus de liberdade e havia também a omissão de muitos itens em cada análise. Com isso, Orlando e Thissen(2000) realizaram o estudo considerando a frequência mínima de 1 para as células, mas ressaltando que  $G^2$  apresentaria resultados pobres.

### 3.3 Teste baseado em probabilidades a posteriori (Stone, 2000)

Outra abordagem para a estimação da curva característica não paramétrica do item é baseada em probabilidades a posteriori (Stone, 2000). Tais probabilidades a posteriori são advindas de processos bayesianos. Nesse caso, a distribuição condicional da habilidade  $\theta$  de um indivíduo dado um padrão de resposta  $x$  é dada por

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \quad (3.17)$$

onde  $P(x|\theta)$  é a probabilidade condicional de resposta no padrão  $x$  (função de verossimilhança),  $P(\theta)$  é a distribuição a priori de  $\theta$  (geralmente utiliza-se a distribuição normal padronizada) e  $P(x)$  é a probabilidade marginal de resposta no padrão  $x$ . Utilizando-se  $q$  pontos de quadratura, a aproximação numérica correspondente para  $P(\theta|x)$  é

$$P(\theta_q|x_j) \approx \frac{P(x_j|\theta_q)w(\theta_q)}{\sum_{k=1}^Q P(x_j|\theta_k)w(\theta_k)} \quad (3.18)$$

onde o numerador é a probabilidade condicional do padrão de resposta do  $j$ -ésimo respondente ao conjunto de itens dada a habilidade  $\theta_q$  multiplicada pelo peso  $w(\theta_q)$ , que corresponde à função densidade da normal no ponto de quadratura  $\theta_q$ .

Para itens dicotômicos, probabilidades não paramétricas baseadas em probabilidades a posteriori são dadas por

$$P_{posterior,i}(\theta_q) = \frac{A_{iq}}{N_{iq}} = \frac{\sum_{j=1}^N P(\theta_q|x_j)Y_{ij}}{\sum_{j=1}^N P(\theta_q|x_j)} \quad (3.19)$$

onde  $Y_{ij}$  é 1 se o  $j$ -ésimo respondente acerta o item  $i$  e 0 caso contrário.  $A_{iq}$  é a contagem de respondentes no grupo de habilidade  $q$  que acerta corretamente o item  $i$ , e é obtido como a soma das probabilidades a posteriori dentre os indivíduos que responderam corretamente, e  $N_q$  é a contagem de respondentes no nível de habilidade  $q$ , e é obtido através da soma das probabilidades a posteriori de todos os respondentes.

A curva característica não paramétrica do item é determinada com base nas probabilidades a posteriori similares às propostas por Levine e Williams na estimação empírica da CCI (Levine & Williams, 1991,1993), sendo esse um estudo melhorado da soma simples não paramétrica proposta por Samejima(1983).

Existem várias semelhanças entre a estimação via suavização por kernel e a baseada em probabilidades a posteriori. No entanto, também há diferenças fundamentais entre essas abordagens.

Enquanto na suavização via kernel o escore final é usado para ranquear os indivíduos de acordo com sua habilidade, na utilização das probabilidades a posteriori espera-se que o ranking seja melhorado já que  $P(\theta_q|x)$  é baseado no padrão de resposta.

Outra diferença se dá no fato de que na suavização por kernel, a mesma função kernel é utilizada para todos os  $\theta_j$ s, o parâmetro de discriminação  $h$  comum para todos os itens controla o desvio-padrão da função kernel, e deve ser escolhido pelo usuário (Samejima, 1998). Em contrapartida, utilizando-se as probabilidades a posteriori, não assume-se forma matemática para  $P(\theta_q|x)$  e a contribuição do respondente para a estimação da curva característica do item é distribuída em diversos níveis de acordo com a variância da distribuição a posteriori. Quão maior é a incerteza da estimação da habilidade, maior é a função suavizadora.

### 3.4 Teste Baseado em Regressão Isotônica

Tendo como base o conhecimento dos vários testes já realizados para o ajuste de curvas características do item, pensou-se em utilizar um teste baseado na regressão

isotônica para tal objetivo. O termo isotônica ou monotônica refere-se ao fato de a variável resposta aumentar com o aumento da variável independente.

A regressão isotônica teve sua teoria sumarizada por Barlow et al. (1972), e consiste em encontrar uma função não decrescente que minimize a soma de quadrados dos erros. Ela difere da regressão usual pois se utiliza informação a priori sobre uma possível relação de ordem na variável resposta. Na teoria da regressão isotônica, tal informação é usada de maneira que se espere um aumento na eficiência do ajuste ao utilizá-la.

Tal método é útil na aplicação à Teoria de Resposta ao Item para a estimação não paramétrica da Curva Característica do Item, uma vez que, na construção dessas curvas, elas apresentam, em geral, forma não decrescente, com o aumento da probabilidade de resposta correta conforme cresce a habilidade do examinado. Para a construção do teste baseado na regressão isotônica, primeiro será apresentada uma breve teoria de como utilizar tal método.

**Definição 3.4.1** Seja  $(X_1, T_1), (X_2, T_2), \dots, (X_n, T_n)$  uma amostra aleatória onde  $X_i$  (tempo de sobrevivência) e  $T_i$  (instante de observação) são variáveis aleatórias independentes, não negativas com funções de distribuição  $F$  e  $G$ , respectivamente. Suponha também que sejam observadas apenas as variáveis  $T_i$  e  $\delta_i = 1_A$ , sendo

$$1_A = \begin{cases} 1, & \text{se } X_i \leq T_i \\ 0, & \text{se } c.c. \end{cases}$$

Essa estrutura de dados é chamada de **caso 1 de censura intervalar**.

Assim, quando há o caso 1 de censura intervalar (status corrente), o estimador não-paramétrico de máxima verossimilhança de  $F$  é o valor  $\hat{F}(T_i), i = 1, \dots, n$ , que maximiza a expressão:

$$\log L(F) = \sum_{i=1}^n \{ \delta_i \log F(T_i) + (1 - \delta_i) \log(1 - F(T_i)) \}, \quad (3.20)$$

sob a restrição de que  $F(T_i) \leq F(T_{i+1})$ , para  $T_i \leq T_{i+1}$

Assume-se sem perda de generalidade que  $0 \leq T_1 \leq T_2 \leq \dots \leq T_n$

O procedimento utilizado para encontrar essas estimativas baseia-se em um algoritmo para cálculo de regressões isotônicas e foi proposto por Groeneboom e Wellner (1992).

**Definição 3.4.2** Uma função real  $h$  em  $D$  é isotônica se  $x \leq y$  implica  $h(x) \leq h(y)$ ,  $\forall x, y \in D$ .

**Definição 3.4.3** Seja  $s$  uma função qualquer em  $D$  e  $w$  uma função positiva em  $D$ . Uma função isotônica  $s^*$  em  $D$  é uma regressão isotônica de  $s$  com pesos  $w$  se, e somente se, minimiza, na classe de funções isotônicas  $h$  em  $D$ , a soma

$$\sum_{x \in D} [s(x) - h(x)]^2 w(x) \quad (3.21)$$

**Teorema 3.4.1** Se  $h$  é isotônica em  $D$ , a imagem de  $h$  está em  $I \subset \mathfrak{R}$  e se  $\phi$  é uma função estritamente convexa, então  $s^*$  é a única função que maximiza

$$\sum_x \{\phi(h(x)) + s(x) - h(x)\} \varphi(h(x)) w(x) \quad \text{em que} \quad \varphi(x) = \frac{d\phi}{dx}. \quad (3.22)$$

A prova se encontra em Barlow et al. (1972), pág. 42.

Observa-se que a expressão para o  $\log L(F)$  também pode ser escrita na forma acima com  $g(T_i) = \delta_i$ ,  $w(T_i) = 1$ ,  $\phi(x) = x \log x + (1 - x) \log(1 - x)$ ,  $\varphi = \phi'(x) = \log(x) - \log(1 - x)$ .

Dessa forma, pelo Teorema 1.1.1, o estimador não-paramétrico de máxima verossimilhança de  $F$  é dado pela regressão isotônica  $g^*$ , calculado por:

$$y_m = g^*(T_m) = \max_{i \leq m} \min_{k \geq m} \frac{\sum_{i \leq j \leq k} \delta_j}{k - i + 1} \quad (3.23)$$

onde  $g^*(T_m) = \hat{F}(T_m)$ ,  $m = 1, \dots, n$

Tal solução também pode ser encontrada no plano cartesiano, ao se plotar os pontos  $(0, 0)$  e  $\left(i, \sum_{j \leq i} \delta_j\right)$  e em seguida, calculando-se o minorante convexo máximo dos pontos em  $[0, n]$ . A definição de minorante convexo máximo encontra-se a seguir.

**Definição 1.1.4.** O minorante convexo máximo é definido como sendo a função

$H^* : [0, n] \rightarrow \mathfrak{R}$  tal que

$$H^*(t) = \sup\{H(t); H(i) \leq \sum_{j \leq i} \delta_j, 0 \leq i \leq n, H(0) = 0, H \text{ convexa}\}$$

Vale lembrar que caso  $\delta_i = 0, 1 \leq i \leq k_1$  e  $\delta_i = 1, k_2 \leq i \leq n$  para  $0 < k_1 < k_2 < n$ , então  $y_i = 0, 1 \leq i \leq k_1$  e  $y_i = 1, k_2 \leq i \leq n$ , com o valor de  $y_i$  sendo a derivada à esquerda de  $H^*$  em  $i$  para  $i = 1, \dots, n$ .

A figura a seguir exemplifica a regressão isotônica e o minorante convexo máximo.

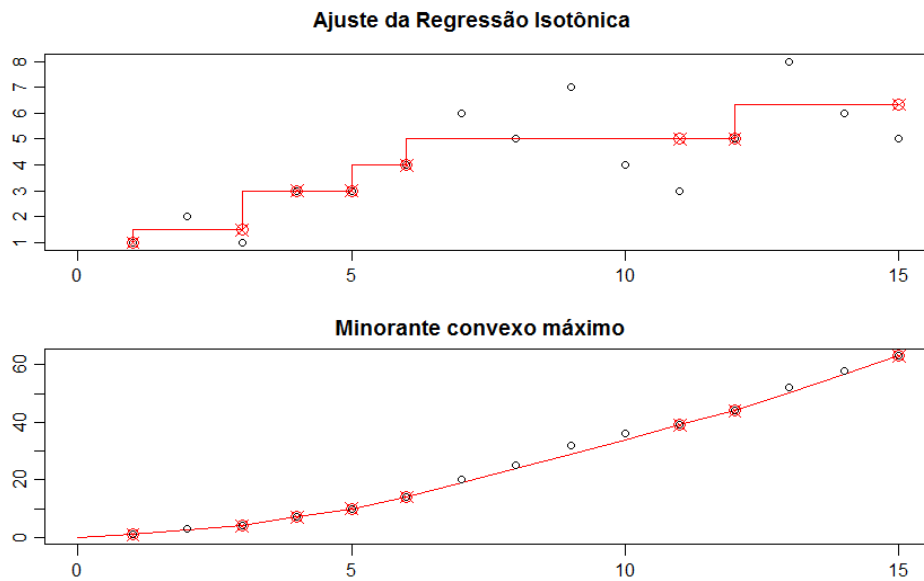


Figura 3.1: *Exemplo de regressão isotônica e minorante Convexo máximo*

O teste de adequabilidade baseado no estimador não paramétrico para  $p_i(\theta_j)$  no ML3 e ML2 será realizado utilizando a estimativa não paramétrica da curva característica do item (CCI) baseada em técnicas de regressão isotônica.

Para a realização desse teste para o  $i$ -ésimo item, tendo como base a teoria de regressão isotônica, o estimador não paramétrico de máxima verossimilhança (ENPMV)  $\hat{F}_i(\hat{\theta}_j)$  sob a suposição de que  $F$  é não decrescente em relação a  $\hat{\theta}_1, \dots, \hat{\theta}_n$  é dado pela regressão isotônica da função  $g_i(\hat{\theta}_j) = u_{ij}$  com pesos  $w_i(\hat{\theta}_j) = 1$ , onde  $u_{ij}$  é indicadora de resposta correta ao  $i$ -ésimo item pelo respondente  $j$ .

A maneira mais prática para obtenção de  $\hat{F}_i(\hat{\theta}_j)$  consiste em tomá-lo como sendo

a derivada à esquerda em  $\sum_{l=1}^j w_l(\hat{\theta}_l) = j$  da função minorante convexa máxima do diagrama de soma acumulada formado pelos pontos  $P_0 = (0, 0)$  e  $P_j = (j, \sum_{l=1}^j u_{il})$ .

Para o ML3, fazemos  $\hat{F}_i^*(\hat{\theta}_j) = \max\{\hat{c}_i, \hat{F}_i(\hat{\theta}_j)\}$  que maximiza (3.20) sob a restrição adicional de que  $F_i(\theta_j) \geq \hat{c}_i$ , e consequentemente  $\hat{F}_i(\hat{\theta}_j)^*$  é o ENPMV para  $p_i(\theta_j)$  em (1.1).

Testa-se então a hipótese nula  $H_0$  de que os dados são advindos de um ML2 ou ML3, utilizando-se para isso uma distância  $d$  entre  $\hat{p}_i$  e  $\hat{F}_i$  no ML2, ou entre  $\hat{p}_i$  e  $\hat{F}_i^*$  no ML3 como estatística do teste. Por exemplo:

$$d_{0i} = \sup_{\theta \in \mathfrak{R}} |\hat{F}_i(\theta) - \hat{p}_i(\theta)|, \quad (3.24)$$

$$d_{1i} = \int_{\mathfrak{R}} |\hat{F}_i(\theta) - \hat{p}_i(\theta)| \phi(\theta) d\theta, \quad (3.25)$$

ou

$$d_{2i} = \sqrt{\int_{\mathfrak{R}} |\hat{F}_i(\theta) - \hat{p}_i(\theta)|^2 \phi(\theta) d\theta}. \quad (3.26)$$

onde  $\phi$  é a função densidade de probabilidade da distribuição  $N(0,1)$ , que pode ser substituída por outra distribuição para as habilidades.

O  $p$ -valor é obtido através do método de reamostragem bootstrap paramétrico, por meio da geração de  $L$  conjuntos aleatórios de respostas  $\{u_{ij}^{(l)}, j = 1, \dots, n; l = 1, \dots, L\}$  para cada indivíduo a partir da CCI paramétrica  $\hat{p}_i$  estimada para o  $i$ -ésimo item com probabilidade de resposta positiva dada por  $P(u_{ij} = 1 | \theta_j) = \hat{p}_{ij} = \hat{p}_i(\theta_j)$  no ML2 ou no ML3. Para o  $l$ -ésimo conjunto de respostas gerado, calcula-se  $\hat{F}_i^{(l)}(\hat{\theta}_j)$ ,  $l = 1, \dots, L$  e a distância entre  $\hat{F}_i^{(l)}$  e  $\hat{p}_i$ , e o  $p$ -valor utilizado é a proporção dos  $L$  conjuntos para os quais  $d(\hat{F}_i^{(l)}, \hat{p}_i) > d(\hat{F}_i, \hat{p}_i)$  ou  $d(\hat{F}_i^{*(l)}, \hat{p}_i) > d(\hat{F}_i^*, \hat{p}_i)$ , ou seja:

$$p - \text{valor} = \frac{\#\{m : d(\hat{F}_i^{(l)}, \hat{p}_i) > d(\hat{F}_i, \hat{p}_i)\}}{L} \quad (3.27)$$

# Capítulo 4

## Simulação e Aplicação

### 4.1 Elaboração da simulação

Para a realização dos testes, utilizou-se o software R versão 3.0.3. Foram implementados no R algoritmos que criassem as funções de distribuição do tempo de falha, para os quais se calcula o estimador não paramétrico de máxima verossimilhança (ENPMV) para dados com status corrente, além de funções que calculassem as diversas integrais referentes às distâncias entre as curvas características dos itens (valor absoluto da diferença e quadrado da diferença) e os estimadores não paramétrico, não paramétrico suavizado via núcleo estimador de kernel e o estimador proposto por Ramsay (1991).

Em seguida foi implementado o processo de simulação com auxílio do pacote *irtoys*. Esse pacote é de extrema utilidade visto que possui várias ferramentas que facilitam a utilização de modelos de TRI, tanto para criação de estimadores quanto para representação gráfica de modelos de TRI.

Para a simulação, define-se os valores dos parâmetros para a geração dos dados, as proficiências reais e com base nessas informações, são gerados os dados. Após isso, inicia-se o processo de obtenção dos estimadores, primeiramente através da estimação não paramétrica de  $\theta$ .

Os cálculos das estatísticas de teste são feitos a partir da geração de dados completos, os itens são calibrados e estima-se as proficiências. Ordenam-se os dados dos respondentes em ordem crescente de score. São criadas várias distâncias entre as



curvas estimadas e a curva real, sendo as utilizadas nessa dissertação:

- Distância quadrática entre o ENPMV e CCI paramétrica ajustada;
- Distância quadrática entre o ENPMV suavizado via kernel e CCI paramétrica ajustada;
- Distância quadrática entre a CCI criada via estimadores de Nadaraya-Watson e CCI paramétrica ajustada.

Para a criação dessas curvas características estimadas, é necessário que haja um processo de suavização dos valores gerados nas estimações. Segundo Ramsay e Silverman (2005), é muito importante ter cuidado na elaboração do processo de suavização, pois caso contrário, a curva estimada pode não representar de maneira correta a curva real e suas características.

O parâmetro  $h$  apresentado anteriormente é conhecido como parâmetro de suavização, e recebe esse nome pois ele controla a suavização. Quanto maior for seu valor, mais suave será a curva, e valores mais baixos de  $h$  resultam em estimativas mais oscilantes, já que apenas algumas observações mais próximas de  $\theta$  são efetivamente consideradas nos pesos.

A escolha do parâmetro  $h$ , conhecido também como *bandwidth*, influencia na suavidade das curvas pois à medida que o aumentamos, diminui-se a quantidade  $\frac{(q_j - \theta)}{h}$ , e dessa forma os pesos dos valores que estão longe do ponto a ser estimado serão maiores. Vários métodos podem ser encontrados sobre como definir o tamanho de  $h$ , sejam eles ligados à variação dos pontos ou até ao tamanho do intervalo de suavização, critérios computacionais minimizando o erro quadrático médio ou métodos de validação cruzada.

Vários experimentos sugerem que  $h = I^{-1/5}$  é um valor com resultados satisfatórios, de forma que seja  $h = 0,2$  para  $I$  de tamanho 3000 e  $h = 0,3$  para tamanho 500. A preferência por um valor de  $h$  ideal varia de acordo com o objetivo do usuário no que diz respeito a vício e variância. Quando se aumenta o valor de  $h$ , a variância do estimador diminui e o vício do estimador aumenta. Dessa forma deve-se escolher o que mais se adequar às pretensões do usuário.

Nessa dissertação, utilizou-se um valor de  $h = \frac{1,1}{(n^{1/5})}$ , conforme a literatura propõe, já que ao serem realizados testes com valores maiores de  $h$ , havia uma tendência à hipersuavização das curvas, principalmente para o caso da regressão via kernel de Nadaraya-Watson.

Para o cálculo do  $p$ -valor, utilizou-se a técnica *bootstrap*, pela qual foram geradas 500 amostras bootstrap, com os parâmetros pré definidos anteriormente a partir dos modelos ajustados. Para cada uma dessas amostras são feitos ajustes dos modelos paramétricos, estimação da proficiência via modelos paramétricos, estimação não paramétrica de  $\theta$  (via escore padrão e  $\theta_{estimado}$ ) e ordenação dos dados referentes aos respondentes em ordem crescente de escore e criadas matrizes com as distâncias para cada amostra bootstrap.

Calcula-se as estatísticas de teste para a amostra bootstrap da mesma forma realizada anteriormente, tanto para o ENPMV quanto para o estimador proposto por Ramsay(1991). Por fim são calculados os  $p$ -valores para as distâncias criadas entre as CCIs estimadas e a CCI ajustada para averiguação da adequabilidade dos testes.

Para todos os casos estudados o modelo considerado foi o de 3 parâmetros (ML3) definido previamente. Foram criados nas simulações provas diferentes para cada caso, considerando a estimação não paramétrica de  $\theta$  de duas maneiras: via escore padrão e via modelos paramétricos ( $\theta_{estimado}$ ). Além de provas com todos os itens seguindo o ML3, foram realizadas simulações com testes em que alguns itens não seguiam tal modelo, com a finalidade de verificar se os estimadores criados detectavam a fuga do modelo proposto.

Para os casos em que todos os itens seguiam o ML3, foram criadas provas com 48 itens. Quanto às provas contendo itens que não seguiam o ML3, foram determinadas as seguintes formas para esses itens:

- Itens possuindo CCI normal assimétrica;
- Itens possuindo CCI de 4 parâmetros (limite superior menor que 1);
- Itens possuindo CCI com plateau intermediário.

O primeiro tipo de item acontece geralmente quando alguns respondentes de habilidade moderada acertam com maior probabilidade quando não sabem e “chutam” o

resultado. Essa CCI foi descrita por Wainer e Thissen (1987) como tendo a forma:

$$P_i(\theta) = \frac{c_i}{1 + \exp[1,702a_i(\theta_j - (b_i + d_i))]} + \frac{1}{1 + \exp[-1,702a_i(\theta_j - b_i)]} \quad (4.1)$$

onde  $a_i$ ,  $b_i$  e  $c_i$  são os parâmetros do ML3 e  $d$  é um número positivo.

O segundo tipo de item segue o ML4 proposto por Barton e Lord (1981), usando um parâmetro a mais que define um limite superior para a proficiência. Tal modelo é definido por

$$P_i(\theta) = \frac{(1 - c_i)d_i}{1 + \exp[-1,702a_i(\theta_j - b_i)]}, \quad (4.2)$$

onde  $d_i$  é o parâmetro extra que limita a curva.

O terceiro tipo apresenta itens com CCI que possuem um plateau, conforme descrito por Bock et al.(1973) e é dado por

$$P_i(\theta) = c_i + (1 - c_i) \left[ \frac{x_i}{1 + \exp[-1,702a_i(\theta_j - b_i)]} + \frac{1 - x_i}{1 + \exp[-1,702a_i(\theta_j - b_i - d_i)]} \right] \quad (4.3)$$

onde  $d_i$  é um número positivo que indica o tamanho do plateau, e  $x_i$  é a altura desse plateau.

Foram criadas provas contendo 48 ou 50 itens no total para esses casos, dividindo-as em provas que continham 6 itens não seguindo o ML3 (2 itens com cada uma das CCIs descritas acima) e provas que continham 18 itens não seguindo o ML3 (6 itens com cada uma das CCIs acima).

Além disso também criou-se provas em que os itens seguiam o ML2, sendo alguns deles também possuindo itens com CCIs conforme as exceções descritas anteriormente, também com a intenção de verificar o comportamento dos testes propostos frente a esses dados. Dessa forma, foram criadas no total 10 provas diferentes, sendo elas:

1. Prova com 48 itens seguindo ML3 e estimação não paramétrica de  $\theta$  via score padrão;
2. Prova com 48 itens seguindo ML3 e estimação não paramétrica de  $\theta$  via modelos paramétricos;
3. Prova com 48 itens, sendo 42 seguindo ML3 e outros 6 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via score padrão;
4. Prova com 48 itens, sendo 42 seguindo ML3 e outros 6 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via modelos paramétricos;
5. Prova com 50 itens, sendo 32 seguindo ML3 e outros 18 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via score padrão;
6. Prova com 50 itens, sendo 32 seguindo ML3 e outros 18 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via modelos paramétricos;
7. Prova com 48 itens seguindo ML2 e estimação não paramétrica de  $\theta$  via score padrão;
8. Prova com 48 itens seguindo ML2 e estimação não paramétrica de  $\theta$  via modelos paramétricos;
9. Prova com 50 itens, sendo 32 seguindo ML2 e outros 18 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via score padrão;
10. Prova com 50 itens, sendo 32 seguindo ML2 e outros 18 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via modelos paramétricos;

Para todos os exames, foi utilizado o número de mil respondentes.

## 4.2 Aplicação

Além das provas criadas através de simulação, o algoritmo foi implementado também para dados reais, utilizando os dados do SARESP, em que dois testes foram aplicados em alunos do 3º ano do ensino médio de turnos diferentes (manhã, tarde e noite) em 2007. Cada teste possui 30 itens, sendo 5 itens comuns aos dois testes. Para esse trabalho, como não há intenção de trabalhar com a equalização dos testes para os diferentes grupos, selecionou-se os alunos de apenas um turno (manhã). O turno da manhã continha mil e um alunos respondendo ao teste, valor semelhante ao utilizado nas simulações implementadas.

A seguir, são apresentadas as tabelas contendo os resultados dos testes utilizados, apresentando os parâmetros criados nas simulações, os parâmetros estimados, assim como os p-valores dos três testes.

Tabela 4.1: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 48 itens seguindo ML3 e estimação não paramétrica de  $\theta$  via escore padrão

Item	a	$\hat{a}$	b	$\hat{b}$	c	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,671	-1,4	0,054	0,2	0,458	0,954	0,936	0,854
Item 2	0,5	0,783	-1	-0,333	0,2	0,284	0,870	0,818	0,846
Item 3	0,5	0,452	-0,6	-1,658	0,2	0,007	0,250	0,280	0,204
Item 4	0,5	0,872	-0,2	0,696	0,2	0,375	0,948	0,952	0,966
Item 5	0,5	0,604	0,2	0,334	0,2	0,216	0,900	0,878	0,792
Item 6	0,5	0,882	0,6	1,708	0,2	0,385	0,738	0,742	0,522
Item 7	0,5	0,743	1	2,053	0,2	0,365	0,826	0,840	0,880
Item 8	0,5	0,465	1,4	1,000	0,2	0,155	0,976	0,958	0,812
Item 9	0,8	1,183	-1,4	-0,464	0,2	0,459	0,926	0,892	0,680
Item 10	0,8	0,563	-1	-2,200	0,2	0,002	0,204	0,216	0,130
Item 11	0,8	1,064	-0,6	0,558	0,2	0,474	0,930	0,946	0,160
Item 12	0,8	0,705	-0,2	-0,068	0,2	0,211	0,902	0,858	0,490
Item 13	0,8	0,717	0,2	0,013	0,2	0,165	0,866	0,820	0,740
Item 14	0,8	0,563	0,6	0,122	0,2	0,106	0,810	0,792	0,692
Item 15	0,8	0,703	1	1,362	0,2	0,247	0,862	0,840	0,528
Item 16	0,8	0,488	1,4	0,579	0,2	0,001	0,790	0,766	0,922
Item 17	1,1	1,449	-1,4	-0,672	0,2	0,494	0,924	0,888	0,308
Item 18	1,1	1,348	-1	-0,361	0,2	0,403	0,944	0,914	0,344
Item 19	1,1	0,871	-0,6	-1,365	0,2	0,004	0,484	0,488	0,434
Item 20	1,1	1,032	-0,2	-0,311	0,2	0,105	0,528	0,520	0,552
Item 21	1,1	1,125	0,2	0,557	0,2	0,291	0,886	0,916	0,608
Item 22	1,1	1,063	0,6	0,468	0,2	0,135	0,988	0,980	0,046
Item 23	1,1	1,451	1	0,994	0,2	0,206	0,718	0,712	0,418
Item 24	1,1	1,271	1,4	1,565	0,2	0,242	0,622	0,618	0,506
Item 25	1,4	1,421	-1,4	-1,077	0,2	0,341	0,756	0,698	0,506
Item 26	1,4	1,575	-1	-0,804	0,2	0,265	0,726	0,714	0,128
Item 27	1,4	1,572	-0,6	-0,446	0,2	0,288	0,852	0,804	0,522
Item 28	1,4	1,949	-0,2	0,030	0,2	0,326	0,918	0,912	0,190
Item 29	1,4	2,610	0,2	0,575	0,2	0,339	0,822	0,808	0,320
Item 30	1,4	1,458	0,6	0,548	0,2	0,214	0,824	0,872	0,502
Item 31	1,4	1,389	1	1,192	0,2	0,230	0,784	0,770	0,348
Item 32	1,4	1,352	1,4	1,565	0,2	0,195	0,706	0,748	0,540
Item 33	1,7	1,506	-1,4	-1,688	0,2	0,000	0,486	0,482	0,396
Item 34	1,7	1,539	-1	-1,080	0,2	0,110	0,592	0,574	0,592
Item 35	1,7	1,472	-0,6	-0,469	0,2	0,295	0,780	0,734	0,844
Item 36	1,7	1,708	-0,2	-0,145	0,2	0,190	0,584	0,576	0,486
Item 37	1,7	1,552	0,2	0,347	0,2	0,224	0,806	0,868	0,788
Item 38	1,7	1,748	0,6	0,789	0,2	0,220	0,784	0,780	0,956
Item 39	1,7	2,380	1	1,023	0,2	0,202	0,914	0,894	0,696
Item 40	1,7	1,118	1,4	1,467	0,2	0,149	0,554	0,542	0,552
Item 41	2	1,902	-1,4	-1,619	0,2	0,000	0,344	0,348	0,194
Item 42	2	1,997	-1	-1,095	0,2	0,001	0,496	0,490	0,352
Item 43	2	1,943	-0,6	-0,451	0,2	0,205	0,648	0,648	0,474
Item 44	2	1,998	-0,2	-0,159	0,2	0,164	0,484	0,524	0,530
Item 45	2	1,849	0,2	0,192	0,2	0,157	0,696	0,680	0,320
Item 46	2	2,235	0,6	0,575	0,2	0,170	0,786	0,758	0,504
Item 47	2	2,487	1	1,176	0,2	0,230	0,886	0,884	0,392
Item 48	2	2,783	1,4	1,440	0,2	0,215	0,732	0,790	0,004

Tabela 4.2: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 48 itens seguindo ML3 e estimação não paramétrica da CCI utilizando  $\theta_{estimado}$  via modelos paramétricos

Item	a	$\hat{a}$	b	$\hat{b}$	c	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,671	-1,4	0,054	0,2	0,458	0,982	0,994	0,586
Item 2	0,5	0,783	-1	-0,333	0,2	0,284	0,846	0,786	0,590
Item 3	0,5	0,452	-0,6	-1,658	0,2	0,007	0,216	0,234	0,110
Item 4	0,5	0,872	-0,2	0,696	0,2	0,375	0,786	0,670	0,756
Item 5	0,5	0,604	0,2	0,334	0,2	0,216	0,878	0,870	0,528
Item 6	0,5	0,882	0,6	1,708	0,2	0,385	0,472	0,438	0,510
Item 7	0,5	0,743	1	2,053	0,2	0,365	0,338	0,290	0,354
Item 8	0,5	0,465	1,4	1,000	0,2	0,155	0,456	0,490	0,282
Item 9	0,8	1,183	-1,4	-0,464	0,2	0,459	0,826	0,810	0,500
Item 10	0,8	0,563	-1	-2,200	0,2	0,002	0,378	0,370	0,286
Item 11	0,8	1,064	-0,6	0,558	0,2	0,474	0,748	0,726	0,212
Item 12	0,8	0,705	-0,2	-0,068	0,2	0,211	0,896	0,896	0,716
Item 13	0,8	0,717	0,2	0,013	0,2	0,165	0,430	0,626	0,448
Item 14	0,8	0,563	0,6	0,122	0,2	0,106	0,778	0,810	0,640
Item 15	0,8	0,703	1	1,362	0,2	0,247	0,204	0,352	0,266
Item 16	0,8	0,488	1,4	0,579	0,2	0,001	0,770	0,720	0,610
Item 17	1,1	1,449	-1,4	-0,672	0,2	0,494	0,852	0,816	0,470
Item 18	1,1	1,348	-1	-0,361	0,2	0,403	0,400	0,526	0,486
Item 19	1,1	0,871	-0,6	-1,365	0,2	0,004	0,150	0,170	0,050
Item 20	1,1	1,032	-0,2	-0,311	0,2	0,105	0,804	0,748	0,732
Item 21	1,1	1,125	0,2	0,557	0,2	0,291	0,706	0,744	0,554
Item 22	1,1	1,063	0,6	0,468	0,2	0,135	0,466	0,428	0,324
Item 23	1,1	1,451	1	0,994	0,2	0,206	0,734	0,784	0,702
Item 24	1,1	1,271	1,4	1,565	0,2	0,242	0,512	0,684	0,630
Item 25	1,4	1,421	-1,4	-1,077	0,2	0,341	0,578	0,502	0,360
Item 26	1,4	1,575	-1	-0,804	0,2	0,265	0,592	0,734	0,600
Item 27	1,4	1,572	-0,6	-0,446	0,2	0,288	0,452	0,704	0,548
Item 28	1,4	1,949	-0,2	0,030	0,2	0,326	0,534	0,402	0,372
Item 29	1,4	2,610	0,2	0,575	0,2	0,339	0,756	0,862	0,718
Item 30	1,4	1,458	0,6	0,548	0,2	0,214	0,706	0,504	0,570
Item 31	1,4	1,389	1	1,192	0,2	0,230	0,768	0,696	0,682
Item 32	1,4	1,352	1,4	1,565	0,2	0,195	0,344	0,584	0,656
Item 33	1,7	1,506	-1,4	-1,688	0,2	0,000	0,470	0,430	0,414
Item 34	1,7	1,539	-1	-1,080	0,2	0,110	0,472	0,462	0,352
Item 35	1,7	1,472	-0,6	-0,469	0,2	0,295	0,876	0,948	0,860
Item 36	1,7	1,708	-0,2	-0,145	0,2	0,190	0,726	0,716	0,522
Item 37	1,7	1,552	0,2	0,347	0,2	0,224	0,682	0,936	0,748
Item 38	1,7	1,748	0,6	0,789	0,2	0,220	0,534	0,464	0,532
Item 39	1,7	2,380	1	1,023	0,2	0,202	0,100	0,602	0,546
Item 40	1,7	1,118	1,4	1,467	0,2	0,149	0,482	0,532	0,626
Item 41	2	1,902	-1,4	-1,619	0,2	0,000	0,220	0,374	0,454
Item 42	2	1,997	-1	-1,095	0,2	0,001	0,576	0,464	0,310
Item 43	2	1,943	-0,6	-0,451	0,2	0,205	0,512	0,362	0,354
Item 44	2	1,998	-0,2	-0,159	0,2	0,164	0,866	0,866	0,722
Item 45	2	1,849	0,2	0,192	0,2	0,157	0,514	0,226	0,252
Item 46	2	2,235	0,6	0,575	0,2	0,170	0,210	0,196	0,092
Item 47	2	2,487	1	1,176	0,2	0,230	0,918	0,854	0,796
Item 48	2	2,783	1,4	1,440	0,2	0,215	0,382	0,448	0,302

Tabela 4.3: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 48 itens, sendo 42 seguindo ML3 e outros 6 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via escore padrão

Item	a	$\hat{a}$	b	$\hat{b}$	c	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,771	-1,25	-1,721	0,2	0,001	0,352	0,360	0,294
Item 2	0,5	0,970	-0,75	-0,492	0,2	0,213	0,868	0,822	0,444
Item 3	0,5	0,793	-0,25	-0,478	0,2	0,161	0,678	0,616	0,644
Item 4	0,5	0,780	0,25	0,120	0,2	0,140	0,964	0,958	0,874
Item 5	0,5	0,921	0,75	1,031	0,2	0,243	0,556	0,504	0,572
Item 6	0,5	1,196	1,25	1,489	0,2	0,258	0,392	0,360	0,642
Item 7	0,75	1,651	-1,25	-0,410	0,2	0,535	0,950	0,952	0,908
Item 8	0,75	1,702	-0,75	-0,179	0,2	0,397	0,974	0,972	0,920
Item 9	0,75	1,521	-0,25	-0,065	0,2	0,266	0,978	0,982	0,988
Item 10	0,75	1,011	0,25	0,208	0,2	0,191	0,970	0,976	0,936
Item 11	0,75	0,979	0,75	0,923	0,2	0,192	0,436	0,478	0,266
Item 12	0,75	1,799	1,25	1,518	0,2	0,254	0,112	0,060	0,088
Item 13	1	1,602	-1,25	-1,522	0,2	0,000	0,402	0,408	0,456
Item 14	1	1,393	-0,75	-0,924	0,2	0,199	0,660	0,650	0,728
Item 15	1	1,362	-0,25	-0,393	0,2	0,156	0,852	0,838	0,858
Item 16	1	1,410	0,25	0,146	0,2	0,152	0,958	0,974	1,000
Item 17	1	1,759	0,75	0,815	0,2	0,218	0,244	0,216	0,132
Item 18	1	1,496	1,25	1,260	0,2	0,159	0,056	0,060	0,122
Item 19	1,25	1,997	-1,25	-1,644	0,2	0,000	0,414	0,424	0,472
Item 20	1,25	2,043	-0,75	-0,740	0,2	0,153	0,836	0,832	0,820
Item 21	1,25	2,060	-0,25	-0,071	0,2	0,260	0,984	0,988	0,970
Item 22	1,25	2,161	0,25	0,291	0,2	0,176	0,956	0,966	0,992
Item 23	1,25	2,134	0,75	0,869	0,2	0,178	0,038	0,034	0,142
Item 24	1,25	2,176	1,25	1,286	0,2	0,183	0,012	0,012	0,088
Item 25	1,5	2,522	-1,25	-1,071	0,2	0,337	0,842	0,808	0,550
Item 26	1,5	3,000	-0,75	-0,582	0,2	0,269	0,978	0,964	0,986
Item 27	1,5	2,723	-0,25	-0,177	0,2	0,246	0,992	0,990	0,978
Item 28	1,5	3,113	0,25	0,348	0,2	0,242	0,884	0,830	0,686
Item 29	1,5	3,779	0,75	0,847	0,2	0,236	0,100	0,114	0,604
Item 30	1,5	2,708	1,25	1,325	0,2	0,222	0,102	0,092	0,140
Item 31	1,75	2,924	-1,25	-1,164	0,2	0,312	0,722	0,732	0,524
Item 32	1,75	3,067	-0,75	-0,698	0,2	0,221	0,968	0,968	0,898
Item 33	1,75	2,642	-0,25	-0,257	0,2	0,183	0,988	0,992	0,970
Item 34	1,75	2,784	0,25	0,259	0,2	0,199	0,970	0,976	1,000
Item 35	1,75	2,341	0,75	0,886	0,2	0,181	0,014	0,054	0,018
Item 36	1,75	2,783	1,25	1,374	0,2	0,193	0,114	0,090	0,150
Item 37	2	2,638	-1,25	-1,551	0,2	0,169	0,616	0,616	0,432
Item 38	2	3,701	-0,75	-0,644	0,2	0,240	0,986	0,992	0,710
Item 39	2	3,543	-0,25	-0,163	0,2	0,199	0,994	0,990	0,980
Item 40	2	2,626	0,25	0,232	0,2	0,156	0,976	0,984	0,956
Item 41	2	4,022	0,75	0,820	0,2	0,210	0,000	0,000	0,000
Item 42	2	2,517	1,25	1,307	0,2	0,178	0,060	0,036	0,106
Item 43	-5	3,181	1,5	0,929	0,2	0,232	0,140	0,088	0,118
Item 44	5	2,363	-1	-0,254	0,2	0,117	0,978	0,978	0,982
Item 45	1,5	0,861	-1	-1,254	0,2	0,000	0,182	0,218	0,096
Item 46	2	3,406	1	1,142	0,2	0,179	0,014	0,014	0,102
Item 47	3	0,481	-1	-0,190	0,2	0,000	0,132	0,200	0,250
Item 48	3,5	1,198	0	0,907	0,2	0,145	0,094	0,156	0,000



Tabela 4.4: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 48 itens, sendo 42 seguindo ML3 e outros 6 com CCIs alternativas, e estimação não paramétrica da CCI utilizando  $\theta_{estimado}$  via modelos paramétricos

Item	a	$\hat{a}$	b	$\hat{b}$	c	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,771	-1,25	-1,721	0,2	0,001	0,172	0,224	0,138
Item 2	0,5	0,970	-0,75	-0,492	0,2	0,213	0,444	0,582	0,530
Item 3	0,5	0,793	-0,25	-0,478	0,2	0,161	0,506	0,490	0,426
Item 4	0,5	0,780	0,25	0,120	0,2	0,140	0,948	0,924	0,716
Item 5	0,5	0,921	0,75	1,031	0,2	0,243	0,622	0,362	0,594
Item 6	0,5	1,196	1,25	1,489	0,2	0,258	0,680	0,658	0,658
Item 7	0,75	1,651	-1,25	-0,410	0,2	0,535	0,702	0,682	0,388
Item 8	0,75	1,702	-0,75	-0,179	0,2	0,397	0,334	0,326	0,380
Item 9	0,75	1,521	-0,25	-0,065	0,2	0,266	0,902	0,798	0,538
Item 10	0,75	1,011	0,25	0,208	0,2	0,191	0,476	0,426	0,384
Item 11	0,75	0,979	0,75	0,923	0,2	0,192	0,162	0,258	0,308
Item 12	0,75	1,799	1,25	1,518	0,2	0,254	0,762	0,766	0,606
Item 13	1	1,602	-1,25	-1,522	0,2	0,000	0,434	0,514	0,486
Item 14	1	1,393	-0,75	-0,924	0,2	0,199	0,498	0,458	0,456
Item 15	1	1,362	-0,25	-0,393	0,2	0,156	0,664	0,688	0,676
Item 16	1	1,410	0,25	0,146	0,2	0,152	0,808	0,804	0,670
Item 17	1	1,759	0,75	0,815	0,2	0,218	0,746	0,772	0,788
Item 18	1	1,496	1,25	1,260	0,2	0,159	0,482	0,494	0,422
Item 19	1,25	1,997	-1,25	-1,644	0,2	0,000	0,088	0,192	0,288
Item 20	1,25	2,043	-0,75	-0,740	0,2	0,153	0,652	0,566	0,818
Item 21	1,25	2,060	-0,25	-0,071	0,2	0,260	0,504	0,510	0,478
Item 22	1,25	2,161	0,25	0,291	0,2	0,176	0,882	0,882	0,658
Item 23	1,25	2,134	0,75	0,869	0,2	0,178	0,262	0,704	0,210
Item 24	1,25	2,176	1,25	1,286	0,2	0,183	0,324	0,292	0,490
Item 25	1,5	2,522	-1,25	-1,071	0,2	0,337	0,324	0,466	0,370
Item 26	1,5	3,000	-0,75	-0,582	0,2	0,269	0,592	0,482	0,712
Item 27	1,5	2,723	-0,25	-0,177	0,2	0,246	0,598	0,670	0,534
Item 28	1,5	3,113	0,25	0,348	0,2	0,242	0,446	0,536	0,664
Item 29	1,5	3,779	0,75	0,847	0,2	0,236	0,606	0,536	0,018
Item 30	1,5	2,708	1,25	1,325	0,2	0,222	0,396	0,820	0,874
Item 31	1,75	2,924	-1,25	-1,164	0,2	0,312	0,966	0,960	0,706
Item 32	1,75	3,067	-0,75	-0,698	0,2	0,221	0,678	0,448	0,528
Item 33	1,75	2,642	-0,25	-0,257	0,2	0,183	0,540	0,270	0,388
Item 34	1,75	2,784	0,25	0,259	0,2	0,199	0,896	0,952	0,850
Item 35	1,75	2,341	0,75	0,886	0,2	0,181	0,382	0,766	0,552
Item 36	1,75	2,783	1,25	1,374	0,2	0,193	0,618	0,620	0,680
Item 37	2	2,638	-1,25	-1,551	0,2	0,169	0,820	0,676	0,484
Item 38	2	3,701	-0,75	-0,644	0,2	0,240	0,444	0,764	0,826
Item 39	2	3,543	-0,25	-0,163	0,2	0,199	0,012	0,044	0,074
Item 40	2	2,626	0,25	0,232	0,2	0,156	0,318	0,260	0,192
Item 41	2	4,022	0,75	0,820	0,2	0,210	0,602	0,332	0,762
Item 42	2	2,517	1,25	1,307	0,2	0,178	0,794	0,724	0,664
Item 43	-5	3,181	1,5	0,929	0,2	0,232	0,802	0,856	0,284
Item 44	5	2,363	-1	-0,254	0,2	0,117	0,486	0,496	0,408
Item 45	1,5	0,861	-1	-1,254	0,2	0,000	0,322	0,252	0,494
Item 46	2	3,406	1	1,142	0,2	0,179	0,852	0,780	0,830
Item 47	3	0,481	-1	-0,190	0,2	0,000	0,004	0,008	0,016
Item 48	3,5	1,198	0	0,907	0,2	0,145	0,016	0,066	0,076

Tabela 4.5: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 50 itens, sendo 32 seguindo ML3 e outros 18 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via escore padrão

Item	a	$\hat{a}$	b	$\hat{b}$	c	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,821	-1,4	-1,289	0,2	0,255	0,148	0,178	0,094
Item 2	0,5	1,085	-1	-0,345	0,2	0,343	0,660	0,634	0,778
Item 3	0,5	0,761	-0,6	-1,237	0,2	0,031	0,262	0,252	0,210
Item 4	0,5	0,919	-0,2	0,109	0,2	0,268	0,814	0,792	0,942
Item 5	0,5	1,027	0,2	0,390	0,2	0,259	0,740	0,736	0,770
Item 6	0,5	1,260	0,6	1,094	0,2	0,303	0,708	0,706	0,622
Item 7	0,5	0,913	1	1,383	0,2	0,252	0,560	0,562	0,762
Item 8	0,5	0,982	1,4	1,504	0,2	0,257	0,694	0,700	0,808
Item 9	1	1,736	-1,4	-1,225	0,2	0,336	0,476	0,452	0,374
Item 10	1	1,331	-1	-1,311	0,2	0,170	0,460	0,454	0,448
Item 11	1	1,352	-0,6	-0,984	0,2	0,006	0,306	0,286	0,212
Item 12	1	1,442	-0,2	-0,163	0,2	0,190	0,476	0,438	0,360
Item 13	1	1,364	0,2	0,134	0,2	0,148	0,320	0,452	0,462
Item 14	1	1,424	0,6	0,644	0,2	0,219	0,660	0,634	0,510
Item 15	1	1,583	1	1,258	0,2	0,220	0,834	0,846	0,866
Item 16	1	1,087	1,4	1,412	0,2	0,147	0,642	0,650	0,840
Item 17	1,5	2,994	-1,4	-1,114	0,2	0,445	0,488	0,468	0,258
Item 18	1,5	2,454	-1	-0,841	0,2	0,295	0,440	0,416	0,374
Item 19	1,5	1,860	-0,6	-1,000	0,2	0,041	0,196	0,158	0,124
Item 20	1,5	2,134	-0,2	-0,300	0,2	0,134	0,248	0,250	0,276
Item 21	1,5	2,349	0,2	0,375	0,2	0,238	0,650	0,578	0,480
Item 22	1,5	2,366	0,6	0,573	0,2	0,158	0,808	0,780	0,048
Item 23	1,5	2,377	1	1,042	0,2	0,167	0,938	0,932	0,860
Item 24	1,5	2,901	1,4	1,369	0,2	0,179	0,916	0,918	0,928
Item 25	2	2,533	-1,4	-1,718	0,2	0,010	0,376	0,370	0,272
Item 26	2	4,732	-1	-0,781	0,2	0,360	0,618	0,576	0,344
Item 27	2	3,830	-0,6	-0,671	0,2	0,186	0,406	0,368	0,308
Item 28	2	3,205	-0,2	-0,200	0,2	0,216	0,174	0,166	0,232
Item 29	2	6,096	0,2	0,439	0,2	0,270	0,706	0,682	0,400
Item 30	2	3,508	0,6	0,709	0,2	0,232	0,892	0,874	0,716
Item 31	2	5,249	1	1,080	0,2	0,206	0,604	0,436	0,026
Item 32	2	6,136	1,4	1,411	0,2	0,196	0,988	0,966	0,470
Item 33	-5	2,610	1,5	0,707	0,2	0,178	0,874	0,874	0,612
Item 34	-4	3,618	1	0,372	0,2	0,251	0,658	0,588	0,360
Item 35	-3	2,890	0,5	-0,090	0,2	0,261	0,246	0,198	0,556
Item 36	3	2,333	-0,5	0,175	0,2	0,164	0,366	0,282	0,208
Item 37	4	2,104	-1	-0,430	0,2	0,128	0,322	0,310	0,296
Item 38	5	2,118	-1	-0,384	0,2	0,105	0,254	0,272	0,224
Item 39	0,5	0,531	-1	-0,940	0,2	0,004	0,218	0,198	0,212
Item 40	1,1	0,983	-1	-1,283	0,2	0,001	0,160	0,144	0,048
Item 41	1,7	1,138	0	0,152	0,2	0,066	0,294	0,292	0,064
Item 42	2	1,701	0	0,090	0,2	0,069	0,160	0,194	0,258
Item 43	1,4	1,965	1	1,257	0,2	0,194	0,910	0,906	0,740
Item 44	0,8	1,240	1	1,017	0,2	0,147	0,732	0,726	0,846
Item 45	3,5	0,560	-1	-0,422	0,2	0,001	0,050	0,060	0,094
Item 46	3,5	1,129	-1	-0,174	0,2	0,075	0,230	0,352	0,236
Item 47	3	0,470	-1	-0,205	0,2	0,003	0,208	0,250	0,256
Item 48	3	1,131	0	0,898	0,2	0,140	0,730	0,754	0,000
Item 49	3,5	0,719	0	0,638	0,2	0,023	0,944	0,982	0,120
Item 50	3,5	0,909	0	0,606	0,2	0,053	0,822	0,958	0,620

Tabela 4.6: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 50 itens, sendo 32 seguindo ML3 e outros 18 com CCIs alternativas, e estimação não paramétrica da CCI utilizando  $\theta_{estimado}$  via modelos paramétricos

Item	a	$\hat{a}$	b	$\hat{b}$	c	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,821	-1,4	-1,289	0,2	0,255	0,528	0,502	0,468
Item 2	0,5	1,085	-1	-0,345	0,2	0,343	0,818	0,846	0,654
Item 3	0,5	0,761	-0,6	-1,237	0,2	0,031	0,024	0,042	0,018
Item 4	0,5	0,919	-0,2	0,109	0,2	0,268	0,228	0,264	0,350
Item 5	0,5	1,027	0,2	0,390	0,2	0,259	0,514	0,446	0,462
Item 6	0,5	1,260	0,6	1,094	0,2	0,303	0,410	0,256	0,552
Item 7	0,5	0,913	1	1,383	0,2	0,252	0,650	0,562	0,726
Item 8	0,5	0,982	1,4	1,504	0,2	0,257	0,528	0,420	0,622
Item 9	1	1,736	-1,4	-1,225	0,2	0,336	0,772	0,686	0,506
Item 10	1	1,331	-1	-1,311	0,2	0,170	0,722	0,528	0,428
Item 11	1	1,352	-0,6	-0,984	0,2	0,006	0,302	0,312	0,356
Item 12	1	1,442	-0,2	-0,163	0,2	0,190	0,556	0,526	0,372
Item 13	1	1,364	0,2	0,134	0,2	0,148	0,016	0,024	0,104
Item 14	1	1,424	0,6	0,644	0,2	0,219	0,916	0,762	0,850
Item 15	1	1,583	1	1,258	0,2	0,220	0,830	0,814	0,676
Item 16	1	1,087	1,4	1,412	0,2	0,147	0,838	0,772	0,752
Item 17	1,5	2,994	-1,4	-1,114	0,2	0,445	0,508	0,618	0,076
Item 18	1,5	2,454	-1	-0,841	0,2	0,295	0,844	0,830	0,754
Item 19	1,5	1,860	-0,6	-1,000	0,2	0,041	0,042	0,014	0,044
Item 20	1,5	2,134	-0,2	-0,300	0,2	0,134	0,758	0,852	0,882
Item 21	1,5	2,349	0,2	0,375	0,2	0,238	0,552	0,742	0,966
Item 22	1,5	2,366	0,6	0,573	0,2	0,158	0,220	0,136	0,128
Item 23	1,5	2,377	1	1,042	0,2	0,167	0,630	0,370	0,726
Item 24	1,5	2,901	1,4	1,369	0,2	0,179	0,674	0,496	0,428
Item 25	2	2,533	-1,4	-1,718	0,2	0,010	0,358	0,260	0,266
Item 26	2	4,732	-1	-0,781	0,2	0,360	0,712	0,752	0,792
Item 27	2	3,830	-0,6	-0,671	0,2	0,186	0,008	0,110	0,124
Item 28	2	3,205	-0,2	-0,200	0,2	0,216	0,340	0,180	0,414
Item 29	2	6,096	0,2	0,439	0,2	0,270	0,494	0,598	0,632
Item 30	2	3,508	0,6	0,709	0,2	0,232	0,802	0,378	0,550
Item 31	2	5,249	1	1,080	0,2	0,206	0,964	0,664	0,832
Item 32	2	6,136	1,4	1,411	0,2	0,196	0,456	0,396	0,320
Item 33	-5	2,610	1,5	0,707	0,2	0,178	0,084	0,220	0,138
Item 34	-4	3,618	1	0,372	0,2	0,251	0,830	0,410	0,594
Item 35	-3	2,890	0,5	-0,090	0,2	0,261	0,846	0,706	0,664
Item 36	3	2,333	-0,5	0,175	0,2	0,164	0,734	0,812	0,842
Item 37	4	2,104	-1	-0,430	0,2	0,128	0,824	0,832	0,908
Item 38	5	2,118	-1	-0,384	0,2	0,105	0,896	0,978	0,958
Item 39	0,5	0,531	-1	-0,940	0,2	0,004	0,450	0,416	0,274
Item 40	1,1	0,983	-1	-1,283	0,2	0,001	0,430	0,368	0,504
Item 41	1,7	1,138	0	0,152	0,2	0,066	0,968	0,966	0,922
Item 42	2	1,701	0	0,090	0,2	0,069	0,840	0,768	0,934
Item 43	1,4	1,965	1	1,257	0,2	0,194	0,072	0,232	0,480
Item 44	0,8	1,240	1	1,017	0,2	0,147	0,356	0,522	0,650
Item 45	3,5	0,560	-1	-0,422	0,2	0,001	0,016	0,016	0,012
Item 46	3,5	1,129	-1	-0,174	0,2	0,075	0,000	0,000	0,000
Item 47	3	0,470	-1	-0,205	0,2	0,003	0,012	0,036	0,006
Item 48	3	1,131	0	0,898	0,2	0,140	0,478	0,568	0,314
Item 49	3,5	0,719	0	0,638	0,2	0,023	0,360	0,502	0,130
Item 50	3,5	0,909	0	0,606	0,2	0,053	0,000	0,082	0,216

Tabela 4.7: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 48 itens seguindo ML2 e estimação não paramétrica de  $\theta$  via escore padrão

Item	a	$\hat{a}$	b	$\hat{b}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,408	-1,4	-1,874	0,542	0,580	0,530
Item 2	0,5	0,448	-1	-1,103	0,396	0,470	0,454
Item 3	0,5	0,425	-0,6	-0,521	0,804	0,738	0,006
Item 4	0,5	0,493	-0,2	-0,124	0,774	0,760	0,488
Item 5	0,5	0,525	0,2	0,152	0,512	0,432	0,570
Item 6	0,5	0,527	0,6	0,542	0,766	0,756	0,212
Item 7	0,5	0,431	1	1,292	0,508	0,616	0,700
Item 8	0,5	0,352	1,4	1,820	0,348	0,346	0,264
Item 9	0,8	0,801	-1,4	-1,351	0,976	0,980	0,218
Item 10	0,8	0,891	-1	-1,020	0,798	0,780	0,754
Item 11	0,8	0,773	-0,6	-0,762	0,494	0,508	0,594
Item 12	0,8	0,744	-0,2	-0,153	0,612	0,544	0,724
Item 13	0,8	0,907	0,2	0,123	0,556	0,652	0,450
Item 14	0,8	0,799	0,6	0,756	0,304	0,404	0,380
Item 15	0,8	0,682	1	1,248	0,784	0,804	0,378
Item 16	0,8	0,886	1,4	1,333	0,558	0,576	0,786
Item 17	1,1	1,096	-1,4	-1,324	0,962	0,950	0,530
Item 18	1,1	1,028	-1	-1,048	0,856	0,852	0,786
Item 19	1,1	1,127	-0,6	-0,505	0,804	0,776	0,782
Item 20	1,1	1,241	-0,2	-0,231	0,908	0,876	0,760
Item 21	1,1	1,135	0,2	0,249	0,474	0,398	0,540
Item 22	1,1	1,190	0,6	0,596	0,548	0,516	0,450
Item 23	1,1	1,215	1	1,016	0,402	0,458	0,302
Item 24	1,1	0,918	1,4	1,432	0,536	0,496	0,298
Item 25	1,4	1,014	-1,4	-1,763	0,566	0,584	0,710
Item 26	1,4	1,251	-1	-1,022	0,766	0,828	0,750
Item 27	1,4	1,119	-0,6	-0,741	0,896	0,844	0,702
Item 28	1,4	1,308	-0,2	-0,008	0,290	0,848	0,704
Item 29	1,4	1,347	0,2	0,190	0,388	0,422	0,324
Item 30	1,4	1,167	0,6	0,689	0,804	0,784	0,406
Item 31	1,4	1,685	1	0,990	0,748	0,742	0,348
Item 32	1,4	1,319	1,4	1,477	0,594	0,614	0,462
Item 33	1,7	1,864	-1,4	-1,309	0,774	0,764	0,472
Item 34	1,7	1,839	-1	-1,054	0,832	0,830	0,642
Item 35	1,7	1,725	-0,6	-0,587	0,784	0,720	0,698
Item 36	1,7	1,461	-0,2	-0,255	0,654	0,610	0,824
Item 37	1,7	1,802	0,2	0,270	0,332	0,278	0,002
Item 38	1,7	1,589	0,6	0,590	0,576	0,504	0,420
Item 39	1,7	1,656	1	1,003	0,648	0,700	0,456
Item 40	1,7	1,684	1,4	1,548	0,800	0,856	0,086
Item 41	2	2,067	-1,4	-1,470	0,868	0,856	0,454
Item 42	2	1,902	-1	-1,011	0,982	0,970	0,868
Item 43	2	1,879	-0,6	-0,613	0,730	0,736	0,842
Item 44	2	1,872	-0,2	-0,206	0,860	0,792	0,682
Item 45	2	1,890	0,2	0,276	0,456	0,496	0,372
Item 46	2	2,239	0,6	0,619	0,654	0,624	0,310
Item 47	2	2,031	1	1,039	0,922	0,888	0,390
Item 48	2	2,055	1,4	1,407	0,760	0,762	0,370

Tabela 4.8: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 48 itens seguindo ML2 e estimação não paramétrica da CCI utilizando  $\theta_{estimado}$  via modelos paramétricos

Item	a	$\hat{a}$	b	$\hat{b}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,408	-1,4	-1,874	0,438	0,516	0,680
Item 2	0,5	0,448	-1	-1,103	0,260	0,404	0,270
Item 3	0,5	0,425	-0,6	-0,521	0,528	0,532	0,134
Item 4	0,5	0,493	-0,2	-0,124	0,598	0,396	0,540
Item 5	0,5	0,525	0,2	0,152	0,524	0,366	0,510
Item 6	0,5	0,527	0,6	0,542	0,524	0,534	0,338
Item 7	0,5	0,431	1	1,292	0,238	0,334	0,182
Item 8	0,5	0,352	1,4	1,820	0,808	0,792	0,556
Item 9	0,8	0,801	-1,4	-1,351	0,150	0,126	0,038
Item 10	0,8	0,891	-1	-1,020	0,910	0,820	0,838
Item 11	0,8	0,773	-0,6	-0,762	0,772	0,728	0,946
Item 12	0,8	0,744	-0,2	-0,153	0,142	0,092	0,248
Item 13	0,8	0,907	0,2	0,123	0,132	0,162	0,166
Item 14	0,8	0,799	0,6	0,756	0,044	0,074	0,030
Item 15	0,8	0,682	1	1,248	0,704	0,834	0,902
Item 16	0,8	0,886	1,4	1,333	0,562	0,564	0,330
Item 17	1,1	1,096	-1,4	-1,324	0,516	0,502	0,386
Item 18	1,1	1,028	-1	-1,048	0,430	0,402	0,520
Item 19	1,1	1,127	-0,6	-0,505	0,334	0,376	0,434
Item 20	1,1	1,241	-0,2	-0,231	0,270	0,160	0,178
Item 21	1,1	1,135	0,2	0,249	0,196	0,162	0,162
Item 22	1,1	1,190	0,6	0,596	0,210	0,458	0,400
Item 23	1,1	1,215	1	1,016	0,148	0,406	0,128
Item 24	1,1	0,918	1,4	1,432	0,728	0,742	0,526
Item 25	1,4	1,014	-1,4	-1,763	0,686	0,694	0,724
Item 26	1,4	1,251	-1	-1,022	0,120	0,152	0,288
Item 27	1,4	1,119	-0,6	-0,741	0,698	0,686	0,940
Item 28	1,4	1,308	-0,2	-0,008	0,234	0,264	0,430
Item 29	1,4	1,347	0,2	0,190	0,106	0,100	0,090
Item 30	1,4	1,167	0,6	0,689	0,388	0,278	0,146
Item 31	1,4	1,685	1	0,990	0,340	0,550	0,230
Item 32	1,4	1,319	1,4	1,477	0,608	0,758	0,272
Item 33	1,7	1,864	-1,4	-1,309	0,658	0,654	0,756
Item 34	1,7	1,839	-1	-1,054	0,940	0,920	0,958
Item 35	1,7	1,725	-0,6	-0,587	0,206	0,214	0,240
Item 36	1,7	1,461	-0,2	-0,255	0,696	0,406	0,722
Item 37	1,7	1,802	0,2	0,270	0,946	0,932	0,844
Item 38	1,7	1,589	0,6	0,590	0,254	0,390	0,478
Item 39	1,7	1,656	1	1,003	0,662	0,742	0,754
Item 40	1,7	1,684	1,4	1,548	0,650	0,590	0,432
Item 41	2	2,067	-1,4	-1,470	0,574	0,860	0,816
Item 42	2	1,902	-1	-1,011	0,424	0,692	0,508
Item 43	2	1,879	-0,6	-0,613	0,492	0,648	0,874
Item 44	2	1,872	-0,2	-0,206	0,626	0,542	0,578
Item 45	2	1,890	0,2	0,276	0,636	0,758	0,694
Item 46	2	2,239	0,6	0,619	0,562	0,472	0,654
Item 47	2	2,031	1	1,039	0,196	0,308	0,296
Item 48	2	2,055	1,4	1,407	0,748	0,896	0,512

Tabela 4.9: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 50 itens, sendo 32 seguindo ML2 e outros 18 com CCIs alternativas, e estimação não paramétrica de  $\theta$  via escore padrão

Item	a	$\hat{a}$	b	$\hat{b}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,739	-1,4	-1,887	0,450	0,516	0,148
Item 2	0,5	0,739	-1	-1,220	0,210	0,258	0,158
Item 3	0,5	0,709	-0,6	-0,768	0,210	0,256	0,076
Item 4	0,5	0,697	-0,2	-0,252	0,772	0,736	0,986
Item 5	0,5	0,714	0,2	0,214	0,922	0,864	0,946
Item 6	0,5	0,714	0,6	0,545	0,800	0,810	0,860
Item 7	0,5	0,638	1	1,319	0,802	0,780	0,854
Item 8	0,5	0,702	1,4	1,565	0,882	0,878	0,842
Item 9	1	1,413	-1,4	-1,645	0,414	0,414	0,072
Item 10	1	1,791	-1	-1,105	0,182	0,168	0,086
Item 11	1	1,580	-0,6	-0,806	0,128	0,142	0,172
Item 12	1	1,385	-0,2	-0,277	0,972	0,656	0,850
Item 13	1	1,701	0,2	0,116	0,848	0,838	0,764
Item 14	1	1,570	0,6	0,667	0,804	0,850	0,842
Item 15	1	1,516	1	1,075	0,716	0,712	0,850
Item 16	1	1,570	1,4	1,397	0,876	0,888	0,828
Item 17	1,5	2,366	-1,4	-1,587	0,590	0,544	0,004
Item 18	1,5	2,033	-1	-1,214	0,120	0,100	0,084
Item 19	1,5	2,188	-0,6	-0,716	0,066	0,072	0,140
Item 20	1,5	2,292	-0,2	-0,268	0,794	0,624	0,866
Item 21	1,5	2,258	0,2	0,191	0,828	0,842	0,740
Item 22	1,5	2,256	0,6	0,619	0,900	0,904	0,876
Item 23	1,5	2,389	1	1,055	0,804	0,788	0,854
Item 24	1,5	2,241	1,4	1,467	0,656	0,678	0,682
Item 25	2	2,802	-1,4	-1,717	0,502	0,384	0,012
Item 26	2	2,743	-1	-1,213	0,296	0,214	0,062
Item 27	2	2,807	-0,6	-0,760	0,144	0,154	0,106
Item 28	2	2,551	-0,2	-0,195	0,866	0,868	0,858
Item 29	2	2,712	0,2	0,210	0,890	0,858	0,814
Item 30	2	2,855	0,6	0,651	0,892	0,886	0,758
Item 31	2	3,435	1	1,055	0,696	0,806	0,664
Item 32	2	3,117	1,4	1,476	0,524	0,538	0,764
Item 33	-5	2,907	1,5	0,812	0,864	0,860	0,888
Item 34	-4	2,197	1	0,252	0,920	0,950	0,948
Item 35	-3	2,432	0,5	-0,299	0,340	0,186	0,404
Item 36	3	2,248	-0,5	0,219	0,730	0,774	0,608
Item 37	4	2,097	-1	-0,287	0,804	0,512	0,774
Item 38	5	2,380	-1	-0,336	0,278	0,264	0,750
Item 39	0,5	0,451	-1	-0,417	0,396	0,690	0,432
Item 40	1,1	0,937	-1	-1,029	0,176	0,184	0,092
Item 41	1,7	1,658	0	0,335	0,444	0,678	0,256
Item 42	2	2,080	0	0,152	0,860	0,856	0,826
Item 43	1,4	2,027	1	1,404	0,444	0,492	0,764
Item 44	0,8	1,095	1	1,261	0,95	0,938	0,892
Item 45	3,5	0,452	-1	0,660	0,000	0,030	0,000
Item 46	3,5	1,057	-1	-0,083	0,002	0,124	0,076
Item 47	3	0,679	-1	0,521	0,072	0,272	0,164
Item 48	3	1,418	0	0,979	0,902	0,918	0,560
Item 49	3,5	1,183	0	1,144	0,468	0,828	0,032
Item 50	3,5	1,563	0	0,860	0,624	0,788	0,432

Tabela 4.10: Tabela com parâmetros, parâmetros estimados e p-valores das distâncias quadráticas da prova com 50 itens, sendo 32 seguindo ML2 e outros 18 com CCIs alternativas, e estimação não paramétrica da CCI utilizando  $\theta_{estimado}$  via modelos paramétricos

Item	a	$\hat{a}$	b	$\hat{b}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,5	0,739	-1,4	-1,887	0,182	0,242	0,162
Item 2	0,5	0,739	-1	-1,220	0,076	0,086	0,046
Item 3	0,5	0,709	-0,6	-0,768	0,534	0,434	0,390
Item 4	0,5	0,697	-0,2	-0,252	0,532	0,426	0,840
Item 5	0,5	0,714	0,2	0,214	0,906	0,828	0,978
Item 6	0,5	0,714	0,6	0,545	0,882	0,770	0,894
Item 7	0,5	0,638	1	1,319	0,082	0,106	0,704
Item 8	0,5	0,702	1,4	1,565	0,778	0,768	0,894
Item 9	1	1,413	-1,4	-1,645	0,496	0,608	0,184
Item 10	1	1,791	-1	-1,105	0,618	0,596	0,552
Item 11	1	1,580	-0,6	-0,806	0,814	0,838	0,844
Item 12	1	1,385	-0,2	-0,277	0,374	0,300	0,496
Item 13	1	1,701	0,2	0,116	0,092	0,124	0,324
Item 14	1	1,570	0,6	0,667	0,082	0,072	0,472
Item 15	1	1,516	1	1,075	0,900	0,976	0,922
Item 16	1	1,570	1,4	1,397	0,686	0,704	0,742
Item 17	1,5	2,366	-1,4	-1,587	0,306	0,142	0,474
Item 18	1,5	2,033	-1	-1,214	0,496	0,572	0,912
Item 19	1,5	2,188	-0,6	-0,716	0,364	0,548	0,646
Item 20	1,5	2,292	-0,2	-0,268	0,656	0,512	0,852
Item 21	1,5	2,258	0,2	0,191	0,482	0,668	0,414
Item 22	1,5	2,256	0,6	0,619	0,772	0,712	0,794
Item 23	1,5	2,389	1	1,055	0,186	0,256	0,840
Item 24	1,5	2,241	1,4	1,467	0,216	0,550	0,834
Item 25	2	2,802	-1,4	-1,717	0,050	0,170	0,498
Item 26	2	2,743	-1	-1,213	0,146	0,106	0,392
Item 27	2	2,807	-0,6	-0,760	0,018	0,038	0,310
Item 28	2	2,551	-0,2	-0,195	0,050	0,082	0,724
Item 29	2	2,712	0,2	0,210	0,990	0,980	0,962
Item 30	2	2,855	0,6	0,651	0,586	0,514	0,670
Item 31	2	3,435	1	1,055	0,504	0,244	0,850
Item 32	2	3,117	1,4	1,476	0,254	0,174	0,676
Item 33	-5	2,907	1,5	0,812	0,672	0,756	0,894
Item 34	-4	2,197	1	0,252	0,052	0,204	0,822
Item 35	-3	2,432	0,5	-0,299	0,004	0,002	0,272
Item 36	3	2,248	-0,5	0,219	0,238	0,174	0,368
Item 37	4	2,097	-1	-0,287	0,542	0,360	0,414
Item 38	5	2,380	-1	-0,336	0,206	0,116	0,624
Item 39	0,5	0,451	-1	-0,417	0,272	0,400	0,646
Item 40	1,1	0,937	-1	-1,029	0,504	0,518	0,474
Item 41	1,7	1,658	0	0,335	0,300	0,248	0,408
Item 42	2	2,080	0	0,152	0,974	0,942	0,974
Item 43	1,4	2,027	1	1,404	0,018	0,014	0,454
Item 44	0,8	1,095	1	1,261	0,340	0,356	0,728
Item 45	3,5	0,452	-1	0,660	0,000	0,000	0,000
Item 46	3,5	1,057	-1	-0,083	0,000	0,000	0,000
Item 47	3	0,679	-1	0,521	0,000	0,000	0,000
Item 48	3	1,418	0	0,979	0,008	0,004	0,002
Item 49	3,5	1,183	0	1,144	0,000	0,000	0,000
Item 50	3,5	1,563	0	0,860	0,000	0,000	0,000

Tabela 4.11: Tabela com parâmetros estimados e p-valores das distâncias quadráticas da prova do SARESP, turno manhã, com 30 itens e estimação não paramétrica de  $\theta$  via escore padrão

Item	$\hat{a}$	$\hat{b}$	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,243	1,977	0,632	0,392	0,424	0,404
Item 2	0,191	0,253	1,895	0,402	0,410	0,618
Item 3	0,156	1,114	1,718	0,862	0,858	0,602
Item 4	0,142	0,253	1,082	0,520	0,596	0,894
Item 5	0,045	-2,729	0,938	0,000	0,000	0,106
Item 6	0,041	-0,402	0,570	0,496	0,526	0,830
Item 7	0,097	0,784	0,811	0,646	0,642	0,600
Item 8	0,002	1,351	0,520	0,994	0,998	0,944
Item 9	0,175	1,043	1,522	0,668	0,708	0,600
Item 10	0,309	1,292	1,468	0,352	0,378	0,276
Item 11	0,011	-0,638	0,603	0,298	0,316	0,456
Item 12	0,323	0,971	0,771	0,344	0,390	0,684
Item 13	0,039	-0,106	1,168	0,734	0,744	0,764
Item 14	0,243	0,233	1,228	0,398	0,444	0,762
Item 15	0,265	2,296	1,747	0,392	0,400	0,428
Item 16	0,158	1,708	1,607	0,814	0,802	0,372
Item 17	0,493	0,797	2,315	0,032	0,036	0,786
Item 18	0,731	0,001	3,140	0,006	0,018	0,988
Item 19	0,075	0,433	0,955	0,634	0,646	0,628
Item 20	0,591	0,533	1,324	0,080	0,102	0,890
Item 21	0,003	0,984	0,584	0,712	0,694	0,738
Item 22	0,173	2,196	2,833	0,484	0,524	0,056
Item 23	0,215	2,351	2,250	0,728	0,726	0,770
Item 24	0,402	2,587	0,796	0,276	0,332	0,174
Item 25	0,003	0,999	0,964	0,914	0,900	0,772
Item 26	0,192	-0,247	1,360	0,398	0,418	0,586
Item 27	0,578	0,606	1,991	0,018	0,020	0,230
Item 28	0,066	0,359	1,515	0,706	0,678	0,650
Item 29	0,003	-1,085	1,220	0,234	0,258	0,346
Item 30	0,426	-0,620	1,363	0,248	0,256	0,514



Tabela 4.12: Tabela com parâmetros estimados e p-valores das distâncias quadráticas da prova do SARESP, turno manhã, com 30 itens e estimação não paramétrica da CCI utilizando  $\theta_{estimado}$  via modelos paramétricos

Item	$\hat{a}$	$\hat{b}$	$\hat{c}$	p-valor distância quadrática	p-valor dist. quadrática suavizada	p-valor Nadaraya- Watson
Item 1	0,243	1,977	0,632	0,268	0,200	0,202
Item 2	0,191	0,253	1,895	0,388	0,378	0,166
Item 3	0,156	1,114	1,718	0,884	0,894	0,708
Item 4	0,142	0,253	1,082	0,584	0,580	0,388
Item 5	0,045	-2,729	0,938	0,440	0,406	0,270
Item 6	0,041	-0,402	0,570	0,924	0,884	0,702
Item 7	0,097	0,784	0,811	0,652	0,596	0,274
Item 8	0,002	1,351	0,520	0,174	0,392	0,026
Item 9	0,175	1,043	1,522	0,812	0,896	0,494
Item 10	0,309	1,292	1,468	0,764	0,700	0,730
Item 11	0,011	-0,638	0,603	0,510	0,584	0,124
Item 12	0,323	0,971	0,771	0,474	0,524	0,354
Item 13	0,039	-0,106	1,168	0,870	0,722	0,556
Item 14	0,243	0,233	1,228	0,496	0,522	0,194
Item 15	0,265	2,296	1,747	0,576	0,556	0,142
Item 16	0,158	1,708	1,607	0,784	0,632	0,616
Item 17	0,493	0,797	2,315	0,386	0,406	0,556
Item 18	0,731	0,001	3,140	0,370	0,356	0,984
Item 19	0,075	0,433	0,955	0,634	0,554	0,344
Item 20	0,591	0,533	1,324	0,458	0,474	0,574
Item 21	0,003	0,984	0,584	0,768	0,850	0,492
Item 22	0,173	2,196	2,833	0,212	0,258	0,332
Item 23	0,215	2,351	2,250	0,610	0,588	0,398
Item 24	0,402	2,587	0,796	0,448	0,396	0,290
Item 25	0,003	0,999	0,964	0,920	0,914	0,662
Item 26	0,192	-0,247	1,360	0,950	0,910	0,780
Item 27	0,578	0,606	1,991	0,324	0,308	0,356
Item 28	0,066	0,359	1,515	0,950	0,828	0,712
Item 29	0,003	-1,085	1,220	0,478	0,482	0,346
Item 30	0,426	-0,620	1,363	0,864	0,830	0,900

Os três testes utilizados apresentaram resultados satisfatórios. Em geral, os p-valores dos testes eram semelhantes para os mesmos modelos quando se utilizou o escore padrão e o  $\theta_{estimado}$  via modelos paramétricos para a estimação não paramétrica das CCIs.

Quando as provas não possuíam itens não seguindo o ML3, foram raros os casos em que o p-valor indicou uma má adequação no ajuste dos testes implementados. Na maioria dos casos, os p-valores das distâncias quadráticas da regressão via kernel de Nadaraya-Watson apresentaram valores menores que os p-valores das distâncias quadráticas e distâncias quadráticas suavizadas.

Um p-valor menor indica um ajuste pior do teste, ou seja, uma distância maior entre a curva estimada e a CCI real, o que graficamente pode ser verificado de maneira mais intuitiva. Dessa forma, foram criados gráficos para alguns itens das diferentes provas simuladas e um gráfico com itens da prova do SARESP, com a finalidade de mostrar o que essas diferenças entre os p-valores dos itens indicam quando as curvas são analisadas.

Na comparação entre os gráficos dos itens n°19 e n°35 das provas com todos os 48 itens seguindo os ML3, Figuras 4.1 e 4.2 respectivamente, é possível observar como as curvas criadas estão distantes da CCI real no item n°19, em especial a gerada pelo estimador de Nadaraya-Watson, enquanto para o item n°35 as curvas possuem menos oscilações em torno da CCI real.

Para a prova com itens seguindo o ML2, é clara também a diferença entre os itens representados nas Figuras 4.7 e 4.8. O item n°14 dessa prova apresentou o p-valor da distância quadrática suavizada como o maior, sendo esse 0,074, enquanto o item n°37 teve como menor p-valor o referente ao teste da regressão via kernel de Nadaraya-Watson ( $p\text{-valor} = 0,844$ ). Para esses dois itens é evidente a proximidade das curvas estimadas com a CCI real no item n°37, e as distâncias entre elas no item n°14, em especial no caso do estimador de Nadarya-Watson.

Para a prova do SARESP utilizando o escore padrão na estimação não paramétrica de  $\theta$ , a Figura 4.9 representando as curvas de quatro itens mostra uma evidente diferença entre as estimações para cada um desses itens. As curvas para o item n°13 estão alinhadas, enquanto para o item n°5 é cada uma delas apresenta uma forma distinta. A seguir estão os gráficos para esse itens citados e outros.

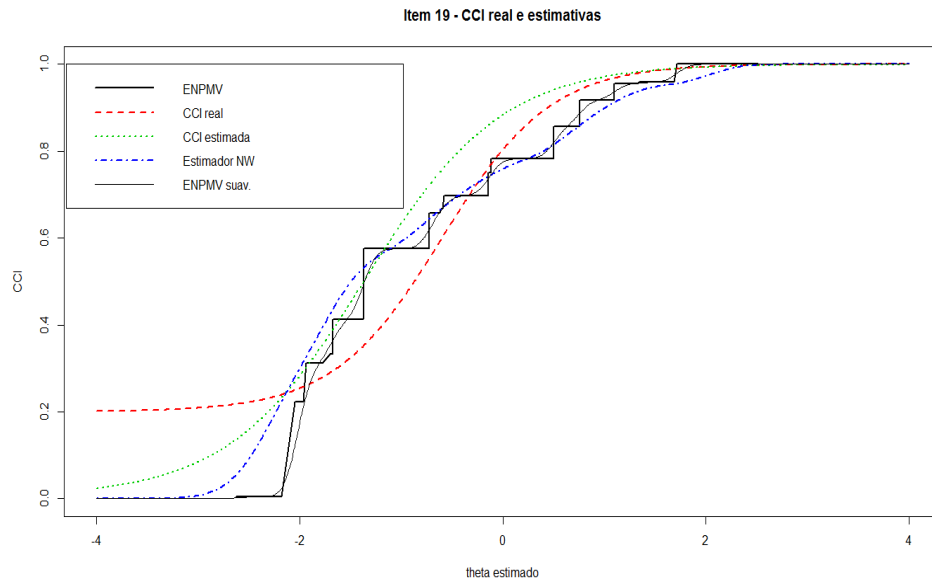


Figura 4.1: *CCIs para o item nº 19 do ML3 utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

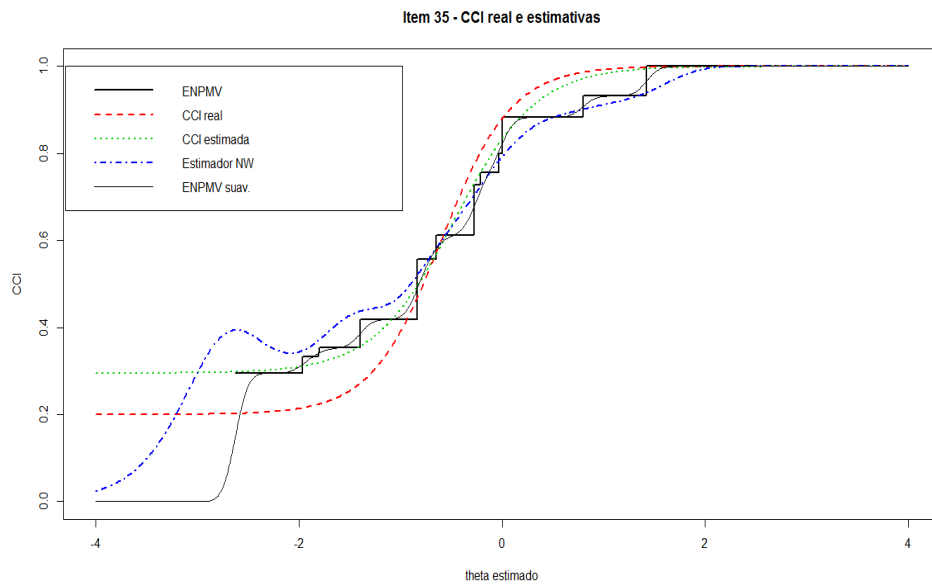


Figura 4.2: *CCIs para o item nº 35 do ML3 utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

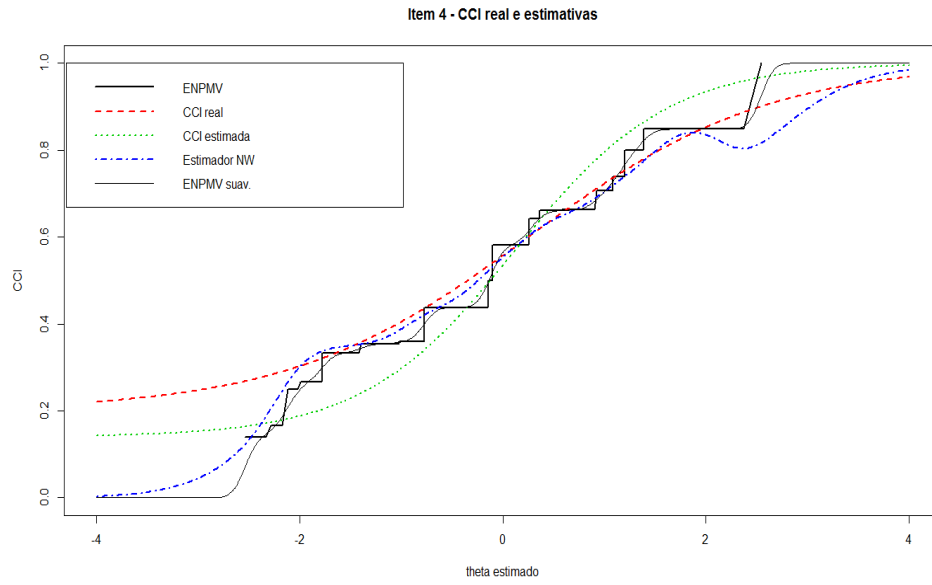


Figura 4.3: *CCIs para o item nº4 do ML3 com 6 itens fora do modelo utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

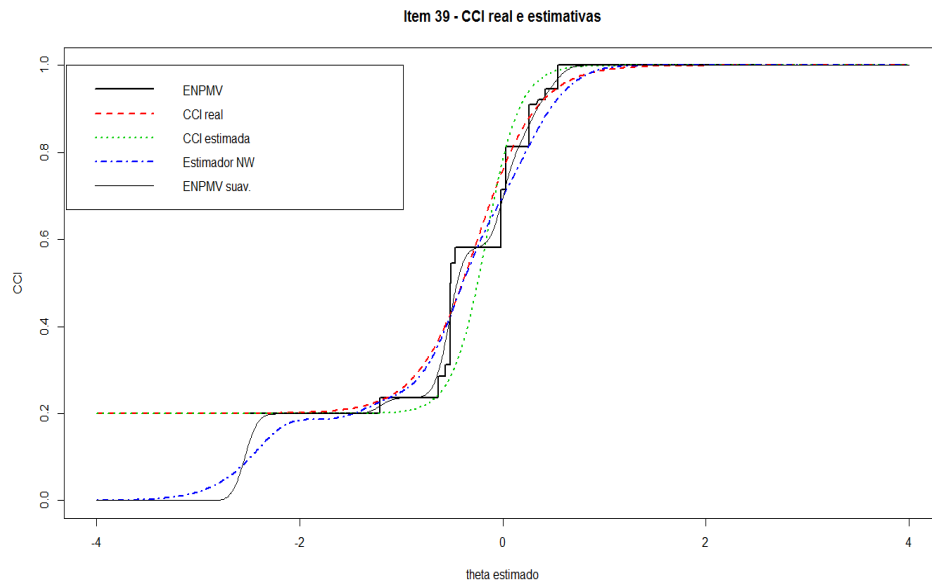


Figura 4.4: *CCIs para o item nº39 do ML3 com 6 itens fora do modelo utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

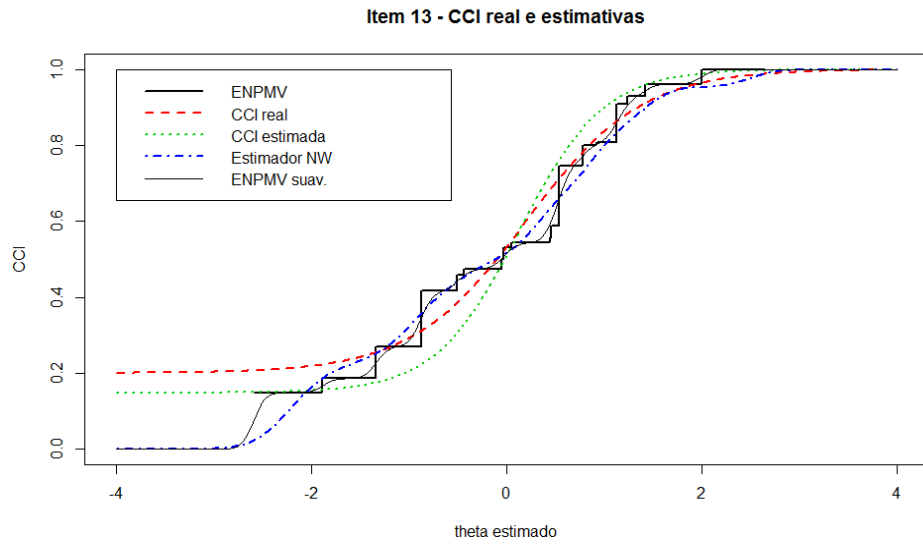


Figura 4.5: *CCIs para o item nº13 do ML3 com 18 itens fora do modelo utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

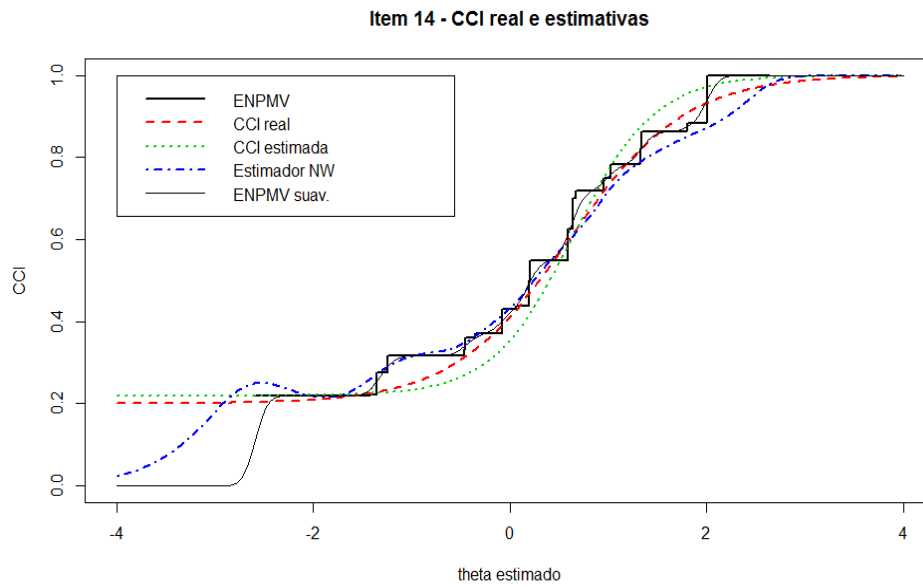


Figura 4.6: *CCIs para o item nº14 do ML3 com 18 itens fora do modelo utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

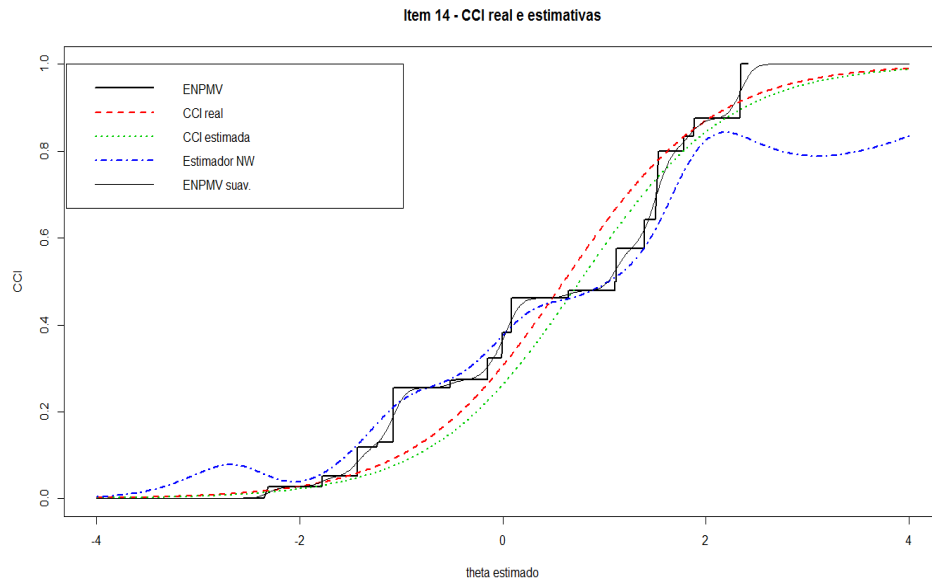


Figura 4.7: *CCIs para o item n° 14 do ML2 utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

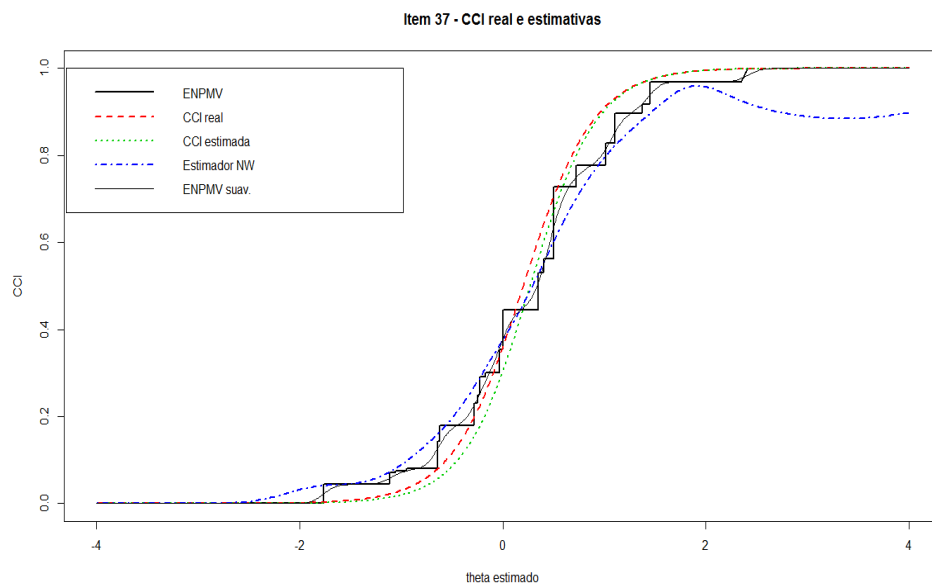


Figura 4.8: *CCIs para o item n° 37 do ML2 utilizando  $\theta_{estimado}$  via modelo paramétrico no ENPMV*

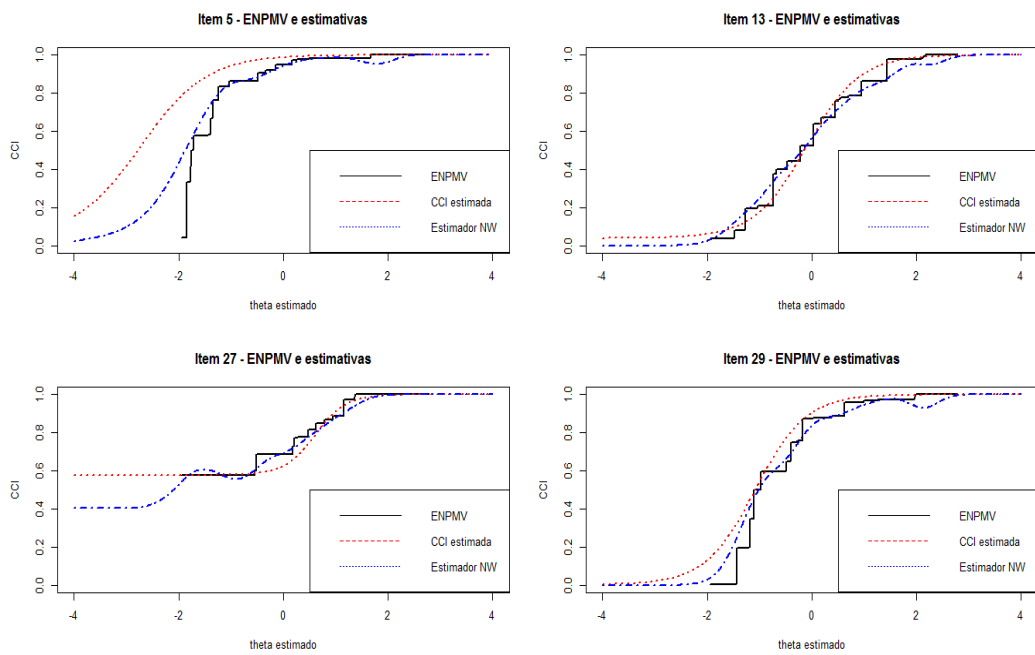


Figura 4.9: *ENPMV e CCI*s estimadas para itens da prova do SARESP utilizando escore padrão no ENPMV

# Capítulo 5

## Conclusões e Trabalhos futuros

### 5.1 Conclusões

De maneira geral, os três testes utilizados neste trabalho não rejeitaram a hipótese de que a CCI segue o modelo logístico para os itens cuja CCI de fato segue este modelo.

Os três teste detectaram as CCIs que não seguem o modelo logístico em alguns casos, especialmente quando a CCI possui um patamar intermediário (plateau). No entanto, para alguns itens, há discrepância entre os p-valores dos três testes. São necessários estudos mais exaustivos para avaliar melhor o desempenho dos testes.

### 5.2 Trabalhos futuros

Existem diversos outros testes de adequabilidade, e pretende-se implementar alguns desses testes no futuro para comparação de resultados com os elaborados nessa dissertação. Dois que merecem destaque foram os abordados nesse trabalho no Capítulo 3, o teste baseado na distribuição Qui-quadrado, elaborado por Orlando & Thissen (2000), e o teste baseado em probabilidades a posteriori, elaborado por Stone (2000).

Ambos os testes já foram citados em publicações, mas julga-se interessante no futuro fazer um trabalho envolvendo a comparação dos testes presentes nessa dissertação com esses dois, além de outros que possam ser interessantes e apresentem



bons resultados.

Outra proposta possível para ser desenvolvida em trabalhos posteriores é o estudo de técnicas para desenvolvimento de testes visando verificar a adequabilidade do ajuste em modelos politômicos como o modelo de resposta gradual.

# Referências Bibliográficas

Andrade, D. F., Tavares, H. R., Valle, R. C.(2000). "Teoria da Resposta ao Item. Conceitos e Aplicações". Associação Brasileira de Estatística: São Paulo.

Andrich, D. (1978). "A rating formulation for ordered response categories". *Psychometrika*, 43, 561-573.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. e Brunk, H. D. 1972. "Statistical Inference Under Order Restrictions". John Wiley & Sons, New York.

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item response model (Research Report No. 81-20). Princeton, NJ: Educational Testing Service.

Bock, R. D., Wainer, H., Petersen, A., Thissen, D., Murray, J., & Roche, A. (1973). A parametrization for individual human growth curves. *Human Biology*, 45, 63-80.

Bolt, D. M. (2001). Conditional covariance-based representation of multidimensional test structure. *Applied Psychological Measurement*, 25, 244-257.

Cochran, William G. The Chi-Square Test of Goodness of Fit. *The Annals of Mathematical Statistics* 23 (1952), no. 3, 315–345.

Croon, M. A. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology*, 44, 315-331.

Douglas, J. (1997). Joint consistency of nonparametric item characteristic curve and ability estimation. *Psychometrika*, 62, 7-28.

Douglas, J., & Cohen, A. (2001). "Nonparametric item response function estimation for assessing parametric model fit". *Applied Psychological Measurement*, 25(3), 234-243.

Groeneboom, P., Wellner, J. A. (1992). "Information Bounds and Nonparametric Maximum Likelihood Estimation". Birkhauser Verlag, Berlin.

Habing, B. (2001). Nonparametric regression and the parametric bootstrap for local dependence assessment. *Applied Psychological Measurement*, 25, 221-233.

Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991). "Fundamentals of Item Response Theory". Newbury Park: Sage Publications.

Hojtink, H., & Molenaar, I. W. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika*, 62, 171-189.

Huisman, J. M. E. (1999). Item nonresponse: Occurrence, causes and imputation of missing answers to test items (M & T Series No. 32). Leiden, The Netherlands: DSWO Press.

Huisman, J. M. E., & Molenaar, I. W. (2001). Imputation of missing scale data with item response models. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 222-244). New York: Springer-Verlag.

Junker, B. W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 247-276). New York: Springer-Verlag.

Junker, B., & Sijtsma, K. (2001b). Cognitive assessment models with few assumptions and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.

Larntz, K. (1978). Small sample comparisons of exact levels for chi-square goodness of fit statistics. *Journal of the American Statistical Association*, 73, 253-263.

Laros, J. A., & Tellegen, P. J. (1991). Construction and validation of the SON-R 5.5-17, the Snijders- Oomen non-verbal intelligence test. Groningen, The Netherlands: Wolters-Noordhoff.

Lee, Y.-S. (2004). "A comparison of nonparametric ICC estimation for small samples and short tests". Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Lee, Y.-S. (2007). "A comparison of methods for nonparametric estimation of item characteristic curves for binary items". *Applied Psychological Measurement*, 31, 121-134

Levine, M. V. & Williams, B. A. (1991, May). An overview and evaluation of nonparametric IRF estimation strategies. Paper presented at the Office of Naval Research Contractors Meeting on Model-Based Measurement, Princeton, NJ.

Levine, M. V, & Williams, B. (1993). "Nonparametric models for polychotomously scored item responses: Analysis and integration". Unpublished manuscript.

Loevinger, J. (1948). "The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis". *Psychological Bulletin*, 45, 507-529.

Lord, F.M. (1980). "Applications of Item Response Theory to Practical Testing Problems". Hillsdale: Lawrence Erlbaum Associates

Lord, F.M., Norvick, M.R. (1968). "Statistical Theories of Mental Test Score". Reading: Addison-Wesley

Masters, G. N. (1982). "A Rasch model for partial credit scoring". *Psychometrika*, 47, 149-174.

McKinley, R. L., & Mills, C. N. (1985). A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement*, 9, 49-57.

Meijer, R. R. (1994). The number of Guttman errors as a simple and powerful person-fit statistic. *Applied Psychological Measurement*, 18, 311-314.

Meijer, R. R., Sijtsma, K., & Molenaar, I. W. (1995). Reliability estimation for single dichotomous items based on Mokken's IRT model. *Applied Psychological Measurement*, 19, 323-335.

Meijer, R. R., & van Krimpen-Stoop, E. M. L. A. (2001). Person fit across subgroups: An achievement testing example. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 377-390). New York: Springer-Verlag.

Mokken, R. J. (1971). *A theory and procedure of scale analysis, with applications in political research*. New York: Walter de Gruyter-Mouton.

Molenaar, I. W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden*, 8, 145-164.

Molenaar, I. W. (1991). A weighted Loevinger Hcoefficient extending Mokken scaling to multcategory items. *Kwantitatieve Methoden*, 37, 97-117.

Molenaar, I. W. (2001) "Thirty Years of Nonparametric Item Response Theory". *Applied Psychological Measurement*, 25, 295-299.

Muraki, E. (1992). "A generalized partial credit model : Application of an EM algorithm". *Applied Psychological Measurement*, 16, 159-176.

Orlando, M., & Thissen, D. (2000). "Likelihood-based item-fit indices for dichotomous item response theory models". *Applied Psychological Measurement*, 24, 50-65.

Orlando, M., & Thissen, D. (2003). "Further investigation of the performance of  $S - X^2$ : An item fit index to use with dichotomous item response theory models". *Applied Psychological Measurement*, 24, 289-298.

Ramsay, J. (1991). "Kernel smoothing approaches to nonparametric item charac-

teristic curve estimation". *Psychometrika*, 56, 611-630.

Ramsay, J.O. & Silverman, B.W. (2005) *Functional data analysis*. New York: Springer.

Samejima, F. (1983). "Some methods and approaches of estimating the operation characteristics of discrete item responses". In H. Wainer and S. Messick [Eds.] *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord*. Hillsdale, NJ: Lawrence Erlbaum

Samejima, F. (1998). "Efficient nonparametric approaches for estimating the operating characteristics of discrete item responses". *Psychometrika*, 63, 111-131.

Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika*, 52, 79-97.

Snijders, T. A. B. (2001). Two-level nonparametric scaling for dichotomous data. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 319-338). New York: Springer-Verlag

Stone, C. A. (2000). "Monte Carlo based null distribution for an alternative goodness-of-fit test statistic in IRT models". *Journal of Educational Measurement*, 37, 58-75.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55, 293-325.

Stout, W. F., Goodwin-Froelich, A., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 357-375). New York: Springer-Verlag.

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement*, 25, 283-294.

Wainer, H., & Thissen, D. (1987). Estimating ability with the wrong model. *Journal of Educational Statistics*, 12, 339-368.

Yen, W. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5, 245-262.

Zhang, J., & Stout, W. F. (1999a). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64, 129-152.

Zhang, J., & Stout, W. F. (1999b). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64, 213-249.