



UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

# ICR: ÍNDICE DE COERÊNCIA DE RESPOSTA E SUAS PROPRIEDADES

Lizandra Farias de Oliveira

Orientação: **Prof. Dr. Héilton Ribeiro Tavares**  
Coorientação: **Profa. Dra. Maria Regina Madruga Tavares**

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

Belém  
2021

Lizandra Farias de Oliveira

# ICR: ÍNDICE DE COERÊNCIA DE RESPOSTA E SUAS PROPRIEDADES

Dissertação apresentada ao Curso de Mestrado em Matemática e Estatística da Universidade Federal do Pará, como pré-requisito para a obtenção do título de Mestre em Estatística.

Orientação: **Prof. Dr. Héilton Ribeiro Tavares**

Coorientação: **Profa. Dra. Maria Regina Madruga Tavares**

Belém

2021

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD  
Sistema de Bibliotecas da Universidade Federal do Pará  
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

---

O48i Oliveira, Lizandra Farias de.  
ICR: Índice de Coerência de Resposta e suas Propriedades /  
Lizandra Farias de Oliveira. — 2021.  
xii, 80 f. : il. color.

Orientador(a): Prof. Dr. Héilton Ribeiro Tavares  
Coorientação: Prof<sup>a</sup>. Dra. Maria Regina Madruga Tavares  
Dissertação (Mestrado) - Universidade Federal do Pará,  
Instituto de Ciências Exatas e Naturais, Programa de Pós-  
Graduação em Matemática e Estatística, Belém, 2021.

1. Avaliação em larga escala. 2. Teoria da Resposta ao  
Item. 3. Coerência. I. Título.

CDD 310

---

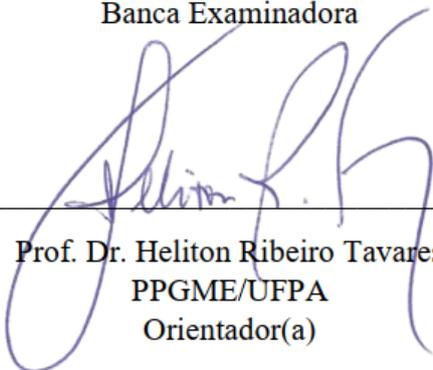
Lizandra Farias de Oliveira

## **ICR: ÍNDICE DE COERÊNCIA DE RESPOSTA E SUAS PROPRIEDADES**

Esta Dissertação foi julgada e aprovada para a obtenção do grau de Mestre em Estatística, no Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará.

Belém, 30 de julho de 2021.

Banca Examinadora



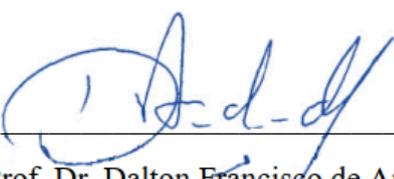
---

Prof. Dr. Heliton Ribeiro Tavares  
PPGME/UFPA  
Orientador(a)



---

Profa. Dra. Maria Regina Madruga Tavares  
PPGME/UFPA  
Coorientador(a)



---

Prof. Dr. Dalton Francisco de Andrade  
PPGMGA/UFSC  
Examinador(a)

---

# Agradecimentos

---

Quero agradecer imensamente ao Nosso Bom e Misericordioso Deus por tanto me ajudar a chegar até aqui, mesmo em meio a tantas dificuldades. Sou grata também aos anjos e santos que junto à Virgem do Carmo intercederam por mim para que eu pudesse concluir este trabalho.

Quero agradecer ao meu esposo, Marcos, que é um grande apoio para mim, segurou minha mão nas horas mais difíceis, me ajudando no que fosse possível.

Sou grata à minha tia Mara e aos meus primos, Mayk e Samayra, que me receberam em sua casa, me apoiaram quando mais precisei. Espero um dia poder retribuir o quanto fizeram por mim.

Sou grata pelas pessoas que conheci. Robinson, obrigada pelas conversas, pela troca de conhecimento, pelo apoio moral, pela amizade. Agradeço a tantos outros colegas de turma que sem saber me ajudaram a enfrentar a carga de estar longe de casa.

Agradeço ao meu orientador, professor Héilton, que me passou tanto conhecimento da qual nem pude dar conta. A essência desse trabalho é mais dele do que minha. Agradeço também a banca avaliadora por suas palavras que me auxiliaram a fazer um trabalho melhor.

Finalmente, gostaria de agradecer à UFPA pelo ensino gratuito de qualidade, a todos que compõem o PPGME e à CAPES, que sem os quais essa dissertação dificilmente poderia ter sido realizada. Agradeço, em particular, à Ana Carolina (Carol), secretária do PPGME, que sempre foi profissional e atenciosa ao responder minhas dúvidas.

*“Ninguém é suficientemente perfeito, que não possa aprender com o outro e, ninguém é totalmente estruído de valores que não possa ensinar algo ao seu irmão”*

*São Francisco de Assis*

---

# Resumo

---

Este trabalho apresenta uma proposta de um estimador de **coerência**, denominado de **Índice de Coerência de Resposta (ICR)**, baseado na distribuição *Poisson-Binomial*, que visa identificar indivíduos que responderam de forma aleatória a todos os itens de um teste ou de um questionário, envolvendo itens dicotômicos. Foi estudada a distribuição de probabilidade da estatística do ICR e várias de suas propriedades, através de resultados analíticos e/ou simulações computacionais. Foi proposta uma regra de decisão de forma a categorizar se todas as respostas de cada indivíduo têm indícios de aleatoriedade. Esta metodologia foi aplicada a dados reais, estimando a proporção de indivíduos que responderam de forma totalmente casual aos itens, bem como na re-estimação dos parâmetros dos itens sem tal conjunto de indivíduos. Tal aplicação teve resultados positivos, na qual as probabilidades de acerto para cada item foi reajustada fazendo com que alunos que marcaram aleatoriamente tivessem uma pequena queda nas probabilidades de acerto de cada item e, à medida que o valor do ICR do indivíduo aumenta, a probabilidade de acertar o item também cresceu.

**PALAVRAS-CHAVE:** Avaliação em larga escala, Teoria da Resposta ao Item, Coerência.

---

# Abstract

---

This work presents a proposal for a **coherence** estimator, called the **Response Coherence Index (ICR)**, based on the *Poisson-Binomial* distribution, which aims to identify individuals who responded randomly to all items of a test or a questionnaire, involving dichotomous items. The probability distribution of ICR statistics and several of its properties were studied through analytical results and/or computational simulations. A decision rule was proposed in order to categorize whether all responses of each individual have signs random clues. This methodology was applied to real data, estimating the proportion of individuals who answered completely randomly to the items, as well as for the re-estimate of the parameters of the items without such a set of individuals. Positive results have been obtained, in which the probabilities of success for each item were readjusted so that students who randomly marked response had a small drop in the probabilities of success of each item and, as the value of the individual's ICR increases, the probability of success of the item also grown.

**KEYWORDS:** Large-scale evaluation, Item Response Theory, Coherence.

---

# Sumário

---

<b>Agradecimentos</b>	<b>v</b>
<b>Resumo</b>	<b>vii</b>
<b>Abstract</b>	<b>viii</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Figuras</b>	<b>xii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Aspectos gerais . . . . .	1
1.2 Justificativa e Importância da Dissertação . . . . .	3
1.3 Objetivos . . . . .	4
1.3.1 Objetivo Geral . . . . .	4
1.3.2 Objetivos Específicos . . . . .	4
1.4 Sumário da Dissertação . . . . .	4
<b>2 As Distribuições Poisson-Binomial e Binomial Composta</b>	<b>6</b>
2.1 A Distribuição Binomial . . . . .	6
2.2 A Poisson-Binomial e suas propriedades . . . . .	7
2.3 A Binomial Composta e suas propriedades . . . . .	10
<b>3 A Teoria da Resposta ao item</b>	<b>13</b>
3.1 Conceitos Iniciais . . . . .	13
3.2 O Modelo Logístico de Quatro Parâmetros e seus Submodelos . . . . .	14
3.3 Processo de Estimação pelo Método da Máxima Verossimilhança Marginal .	16
3.4 Estimação de Parâmetros com o uso do Pacote mirt do R . . . . .	18
<b>4 ICR: Índice de Coerência de Resposta</b>	<b>22</b>
4.1 A Variável Escore (S) e suas Propriedades . . . . .	22
4.2 Índice de Coerência de Resposta - ICR . . . . .	25
4.3 Estudo do ICR com parâmetros dos itens e habilidades conhecidos . . . . .	28
4.4 Estudo do ICR com parâmetros dos itens conhecidos e habilidades desco- nhecidas . . . . .	29
4.5 Estudo do ICR com parâmetros dos itens e habilidades desconhecidos . . . . .	31
4.6 Robustez . . . . .	33

4.7 Critério de Decisão . . . . .	39
<b>5 Aplicação a dados reais: ENEM-2019</b>	<b>47</b>
5.1 Sobre o Exame Nacional do Ensino Médio . . . . .	47
5.2 Estimação dos parâmetros dos itens e habilidades . . . . .	48
5.3 Estatísticas do ICR . . . . .	50
<b>6 Conclusões e Considerações Gerais</b>	<b>54</b>
6.1 Trabalhos Futuros . . . . .	54
<b>Referências</b>	<b>56</b>
<b>A Algoritmo para organização da base de dados do ENEM-2019, Fortaleza-CE</b>	<b>58</b>
<b>B Algoritmo para gerar estatísticas dos escores de cada área do ENEM-2019, Fortaleza-CE</b>	<b>64</b>
<b>C Algoritmo de construção do ICR</b>	<b>66</b>
<b>D Algoritmo para aplicação do ICR aos dados de Ciências Humanas e suas Tecnologias do ENEM-2019, Fortaleza-CE</b>	<b>75</b>

---

## Lista de Tabelas

---

4.1	Algumas Fórmulas Estatísticas para Medidas Amostras. . . . .	24
4.2	Algumas Estatísticas do ENEM-2019, Fortaleza-CE. . . . .	24
4.3	Tabela com estatísticas do ICR usando EAP / VP com habilidades desco- nhecidas . . . . .	31
4.4	Tabela com estatísticas do ICR usando EAP /VP. . . . .	33
4.5	Algumas estatísticas para avaliar a robustez usando VP. . . . .	38
4.6	Algumas estatísticas para avaliar a robustez usando EAP. . . . .	38

---

# Lista de Figuras

---

3.1	Exemplo de uma Curva Característica do Item. . . . .	14
3.2	Exemplo de gráfico usando a função <code>plot()</code> . . . . .	21
3.3	Exemplo de gráfico usando a função <code>itemplot()</code> . . . . .	21
4.1	Histograma dos Escores no ENEM-2019, Fortaleza-CE. . . . .	23
4.2	Gráfico do ICR com parâmetros conhecidos . . . . .	29
4.3	Gráficos do ICR com parâmetros dos itens conhecidos e habilidades desconhecidas . . . . .	30
4.4	Recuperação dos parâmetros dos itens . . . . .	32
4.5	Gráficos do ICR com parâmetros desconhecidos . . . . .	32
4.6	Robustez em relação ao Modelo de Resposta ao Item . . . . .	33
4.7	Robustez em relação a distribuição inicial de $\theta$ . . . . .	35
4.8	Robustez em relação ao número de indivíduos . . . . .	36
4.9	Robustez em relação ao número de itens . . . . .	37
4.10	Gráfico com as densidades do ICR e de $Z$ . . . . .	40
4.11	Gráfico com as densidades de ICR, $Z$ e $Z'$ considerando 22 itens e 4 alternativas. . . . .	42
4.12	Gráfico no $\mathbb{R}^2$ dos valores de $z$ . . . . .	43
4.13	Gráfico no $\mathbb{R}^3$ dos valores de $z$ . . . . .	43
4.14	Gráfico do modelo 1 estimado . . . . .	44
4.15	Gráfico do modelo 2 estimado . . . . .	45
4.16	Gráfico do modelo 2 estimado ( $\mathbb{R}^2$ ) . . . . .	46
5.1	Estimação dos parâmetros dos itens de Ciências Humanas . . . . .	49
5.2	Proporção de acertos no item 3 de Ciências Humanas . . . . .	49
5.3	Gráfico das habilidades estimadas em Ciências Humanas . . . . .	50
5.4	Curva do ICR usando dados de Ciências Humanas do ENEM - 2019, Fortaleza-CE . . . . .	51
5.5	Gráficos para comparação das estimativas considerando dados com e sem aleatoriedade . . . . .	51
5.6	Gráficos para comparação das estimativas por item, considerando dados com e sem aleatoriedade . . . . .	52
5.7	Comparação da soma das probabilidades de acerto obtidas antes e depois da exclusão de dados aleatórios para diferentes valores do ICR. . . . .	53

---

# Capítulo 1

## Introdução

---

### 1.1 Aspectos gerais

Na área educacional é comum a utilização de provas como um instrumento de avaliação em diversas áreas de conhecimento e em diversas situações. As provas tem por objetivo verificar o conhecimento do aluno, serve para avaliar o desempenho do mesmo de forma a acompanhar seu processo de aprendizagem, verificando em que nível se encontra.

Há provas que além de servirem como instrumento avaliativo servem também como processo seletivo para recrutar o indivíduo a uma finalidade específica, seja a um cargo em uma empresa ou uma vaga na universidade. As provas podem ser apresentadas de forma dissertativa ou objetiva. A primeira é aquela em que o aluno tem liberdade ao elaborar sua resposta, utilizando suas próprias palavras. A segunda é a qual o aluno pode, por exemplo, completar lacunas, assinalar verdadeiro ou falso, ou simplesmente escolher uma alternativa entre várias apresentadas (RAMPAZZO, 2011).

A depender da finalidade da prova é vantajoso utilizar de provas objetivas, pois com elas é possível varrer uma maior extensão de conhecimentos e habilidades, desde conhecimentos mais simples a mais complexos e não depender muito tempo na correção.

Porém, a grande desvantagem das provas objetivas é que há a possibilidade de acertar as questões sem ter conhecimento do assunto. Isso pode ocorrer: por eliminação, onde se sabe quais alternativas são incorretas; por “chute”, onde o respondente acerta ao acaso numa espécie de adivinhação; ou também por trapaça (fraude), onde o aluno adquire a resposta correta por meios ilícitos.

Sobre a prática de fraudes, tem surgido nos últimos anos muitos trabalhos focados em buscar detectar se houve fraude ou não ao responder uma prova. Em avaliações de larga escala não é possível verificar de maneira simples se houve fraude nas respostas, é necessário evidências substanciais e há casos em que essas evidências podem ser medidas através de índices estatísticos (CIZEK e WOLLACK, 2017).

Os editores Cizek e Wollack (2017) mostram em seu livro a contribuição de vários autores, que focam em métodos quantitativos para detectar tipos de ações suspeitas de trapaça em testes, como, por exemplo, semelhanças improváveis nas respostas, cópia de respostas, características de *background* semelhantes entre examinandos com padrões de resposta anômalos, pré-conhecimento do conteúdo do teste e entre outros.

No trabalho de Souza (2019) são apresentados métodos para detecção de potenciais fraudes em avaliações de larga escala, expondo diferentes índices de detecção de cópia e similaridade de respostas, trazendo uma proposta para otimizar esses métodos, que delongam tempo computacional. Na mesma linha, temos também o trabalho de Meza (2020) cuja dissertação contribui com a proposta de uma estatística que minimiza significativamente a taxa de falso positivo, e evita que alguém seja acusado injustamente de ter fraudado o teste.

O presente trabalho traz uma nova proposta, que tem como foco detectar os indivíduos que marcam a prova no “chute”, fazendo uma análise da coerência do padrão de resposta dos candidatos, do qual podemos tirar conclusões se há ou não indícios de aleatoriedade. É coerente que indivíduos com baixa proficiência acertem menos questões e aqueles com alta proficiência, mais questões, e é necessário olhar não apenas a quantidade de respostas certas, mas também a coerência. Por exemplo, considere um determinado teste de múltipla escolha e um padrão de respostas em que o indivíduo marca corretamente apenas as questões difíceis. O número de questões certas pode não corresponder ao nível de conhecimento da pessoa. Neste caso, suspeita-se que o respondente tenha “chutado” tais questões, pois não é coerente que alguém seja capaz de acertar os itens difíceis e não acertar os fáceis, dado que os conhecimentos são acumulativos, em que primeiro se aprende os conteúdos simples e depois os mais complexos. Então, o número de questões certas por si só não necessariamente condiz com a proficiência do indivíduo.

A Teoria da Resposta ao Item (TRI), é um instrumento de medida estatística que procura verificar a coerência nos padrões de respostas, buscando realmente avaliar o conhecimento do aluno sobre determinado assunto, tentando dar um parecer não apenas quantitativo, mas também qualitativo, buscando evitar que pessoas obtenham nota altas por simplesmente acertarem uma grande quantidade de questões de maneira aleatória, e algumas vezes desonesta.

Há no Brasil o Exame Nacional do Ensino Médio (ENEM), que é um conjunto de provas objetivas e uma redação, o qual avalia o domínio do candidato sobre conhecimentos,

adquiridos no ensino médio escolar regular, em quatro grandes áreas, são elas: Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias; Matemática e suas Tecnologias; e Linguagens e Códigos e suas Tecnologias. Este exame usa a TRI para correções de suas avaliações e a nota obtida no ENEM é utilizada por grandes universidades como forma de seleção de ingresso, de maneira parcial ou integral. E eis que dada a importância da finalidade de tal prova é necessário que haja um grande cuidado ao analisar seus resultados. Deve-se garantir que as notas estimadas sejam precisas e coerentes.

Assim, este trabalho visa trazer uma contribuição ao estimar os resultados de provas de larga escala, buscando verificar se as respostas dada pelo candidato está conforme sua proficiência, ou seja, se tem coerência. Propõe-se um estimador que ajude, a partir de um critério bem estabelecido, a identificar indivíduos que possam ter respondido a prova de maneira totalmente aleatória, e, assim, trazer uma melhoria na qualidade das estimativas.

## **1.2 Justificativa e Importância da Dissertação**

Em todo e qualquer teste há a necessidade de agir com honestidade, tanto da parte do aplicador como do respondente para que o impacto ao aplicar o teste tenha resultados mais justos.

Existem avaliações cujos resultados são usados para analisar os conhecimentos e as habilidades dos estudantes de uma escola, de uma cidade ou até de um país, para que a partir disto sejam formuladas políticas e programas educacionais com o objetivo de melhorar a qualidade do ensino. Há também provas cuja finalidade é selecionar os alunos para uma vaga na universidade, por exemplo, o ENEM. Em ambos os casos é necessário que as notas finais obtidas sejam as mais verdadeiras possíveis, dada a importância do resultado das provas.

Porém, é uma realidade que nessas provas de múltipla escolha muitos alunos marcam as alternativas dos itens de maneira aleatória, porque não sabem o conteúdo, porque a prova é longa e não houve tempo para ler e pensar nas questões, porque se distraiu ou porque simplesmente não há o interesse de fazer a prova, o que acaba dificultando a calibração dos itens. Por isso, é preciso identificar onde houve essa aleatoriedade, pois: se o candidato "chutou" tudo ele merece uma menor nota, seja porque não usou seu conhecimento, seja porque realmente sua proficiência é baixa; e se a finalidade é, por exemplo, obter um

índice do desempenho de uma escola, os alunos que "chutam" podem diminuir o valor desse índice e acabar prejudicando a escola. Seja qual for a situação, tais indivíduos não devem atrapalhar o processo de calibração dos parâmetros dos itens, se caso forem desconhecidos.

Dessa forma, este trabalho apresenta um estimador de coerência, denominado Índice de Coerência de Resposta (ICR), que visa verificar a conformidade nas respostas daqueles submetidos a um teste, buscando identificar quais indivíduos responderam a prova de forma totalmente aleatória. Além de buscar uma melhor qualidade ao fazer as estimativas dos parâmetros.

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

O objetivo principal desta dissertação é propor um estimador de coerência, denominado ICR - Índice de Coerência de Resposta, baseado na distribuição Poisson-Binomial, que visa identificar indivíduos que responderam de forma aleatória a todos os itens de um teste.

### 1.3.2 Objetivos Específicos

- i) Estudar a distribuição estatística do ICR e suas propriedades;
- ii) Propor uma regra de decisão de forma a categorizar as respostas de cada indivíduo, indicando se têm ou não indícios de aleatoriedade em todos os itens;
- iii) Aplicar a estatística ICR a dados reais, provenientes da edição do ENEM do ano de 2019.

## 1.4 Sumário da Dissertação

O presente trabalho está dividido em seis capítulos, são eles:

- No primeiro capítulo se aborda os aspectos gerais, justificativa e importância deste trabalho, os objetivos geral e específicos e o sumário da dissertação.
- No segundo capítulo é apresentada as distribuições Poisson-Binomial e Binomial Composta e suas propriedades.

- 
- No terceiro capítulo é citado brevemente a Teoria da Resposta ao Item, seus principais modelos e o uso do pacote mirt do R para estimação dos parâmetros.
  - No quarto capítulo trata-se do Índice de Coerência de Resposta (ICR) no qual se faz um estudo de sua distribuição, sua robustez e é proposto um critério de decisão.
  - No quinto capítulo a distribuição do ICR é aplicada aos dados do ENEM 2019.
  - No sexto capítulo se finaliza este trabalho com a conclusão e algumas considerações finais.

---

## Capítulo 2

# As Distribuições Poisson-Binomial e Binomial Composta

---

### 2.1 A Distribuição Binomial

Considere um teste que contenha  $n$  questões independentes, onde cada uma é um ensaio de Bernoulli, ou seja, admite apenas dois resultados: acerto, com probabilidade  $p$ , ( $0 \leq p \leq 1$ ), e erro, com probabilidade  $q = 1 - p$ . Considere, em particular a probabilidade de acerto nas  $x$  primeiras questões do teste e erro nas  $n - x$  restantes, então temos

$$\underbrace{p \cdot p \cdots p}_x \cdot \underbrace{q \cdot q \cdots q}_{n-x} = p^x q^{n-x}$$

Agora, caso se queira a probabilidade de obter os mesmos  $x$  acertos e  $n - x$  erros, porém em uma sequência diferente será obtido o mesmo resultado acima, pois os acertos e erros serão somente rearranjados.

Para obter o número de sequências possíveis para este mesmo caso, basta permutar os  $p$ 's e  $q$ 's e logo tem-se

$$\frac{n!}{x!(n-x)!} = \binom{n}{x}$$

possibilidades distintas.

Então, neste caso, a probabilidade de obter  $x$  acertos em um teste de  $n$  questões independentes é dada por

$$\binom{n}{x} p^x q^{n-x},$$

que caracteriza o modelo denominado na literatura de distribuição binomial.

**Definição:** Uma variável aleatória discreta  $X$  é dita ter distribuição Binomial com parâmetros  $(n, p)$ , denotada como  $X \sim Bin(n, p)$ , onde  $n \in \mathbb{N}$ ,  $0 \leq p \leq 1$  e  $q = 1 - p$ , se

sua função de probabilidade é dada por

$$f(x|n, p) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = \{0, 1, 2, \dots, n\} \quad (2.1)$$

Perceba que  $X \sim Bin(n, p)$  pode ser escrita como  $X = \sum_{i=1}^n X_i$ , onde cada  $X_i \sim Bernoulli(p)$  independentes.

É importante lembrar que ao fazer uso da distribuição binomial é necessário ter  $n$  repetições de um ensaio de Bernoulli, que sejam independentes e com probabilidade de sucesso constante em cada ensaio.

## 2.2 A Poisson-Binomial e suas propriedades

Considere o mesmo caso da seção anterior, porém agora as questões têm probabilidades de acerto não necessariamente iguais, ou seja, cada questão  $i$  ( $i = 1, 2, \dots, n$ ) tem distribuição  $Bernoulli(p_i)$ . Tomando, por exemplo, os valores  $n = 3$  e  $x = 2$  se tem que

$$\begin{aligned} P(X = 2) &= p_1 p_2 q_3 + p_1 p_3 q_2 + p_2 p_3 q_1 \\ &= p_1 p_2 (1 - p_3) + p_1 p_3 (1 - p_2) + p_2 p_3 (1 - p_1) \\ &= \prod_{i \in \{1,2\}} p_i \prod_{i \in \{3\}} (1 - p_i) + \prod_{i \in \{1,3\}} p_i \prod_{i \in \{2\}} (1 - p_i) + \prod_{i \in \{2,3\}} p_i \prod_{i \in \{1\}} (1 - p_i) \end{aligned}$$

Utilizou-se destes valores para que o leitor note que quanto maior seja  $n$  mais parcelas terão que ser somadas, o que torna inviável o cálculo direto das probabilidades.

Isto gera um modelo mais geral para soma de variáveis aleatórias Bernoulli independentes que não são identicamente distribuídas, denominado na literatura por distribuição Poisson-Binomial (PB).

**Definição:** Uma variável aleatória discreta  $X$  é dita ter distribuição Poisson-Binomial com parâmetros  $\mathbf{p} = \{p_1, \dots, p_n\}$ , denotada como  $X \sim PB(\mathbf{p})$ , onde  $0 \leq p_i \leq 1$ , se sua função de probabilidade é dada por

$$f(x|\mathbf{p}) = P(X = x) = \sum_{A \in B_x} \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i) \quad (2.2)$$

onde  $B_x = \{A | A \subseteq \{1, 2, \dots, n\}, \#A = x\}$ , no qual  $\#A$  denota a cardinalidade de  $A$ . Para melhor visualizar  $B_x$  é considerado novamente  $n = 3$ , assim

$$B_0 = \{\};$$

$$B_1 = \{\{1\}, \{2\}, \{3\}\};$$

$$B_2 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}; \text{ e}$$

$$B_3 = \{\{1, 2, 3\}\}$$

são os conjuntos utilizados no somatório para obter as correspondentes probabilidades de ocorrer  $x$  sucessos.

Adiante é apresentada algumas propriedades da distribuição Poisson-Binomial. Dado que esta é uma soma de Bernoulli's independentes, então se tem os seguintes resultados:

- Média

$$\mu = E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p_i \quad (2.3)$$

- Variância

$$\sigma^2 = Var(X) = Var\left(\sum_{i=1}^n X_i\right) \stackrel{indep.}{=} \sum_{i=1}^n Var(X_i) = \sum_{i=1}^n p_i q_i \quad (2.4)$$

- Função Geradora de Probabilidade

$$G_X(t) = E(t^X) = E(t^{\sum_{i=1}^n X_i}) \stackrel{indep.}{=} \prod_{i=1}^n E(t^{X_i}) = \prod_{i=1}^n (p_i t + q_i) \quad (2.5)$$

- Função Geradora de Momentos

$$M_X(t) = E(e^{tX}) = E(e^{t\sum_{i=1}^n X_i}) \stackrel{indep.}{=} \prod_{i=1}^n E(e^{tX_i}) = \prod_{i=1}^n (p_i e^t + q_i) \quad (2.6)$$

- Função Característica (FC)

$$\varphi_X(t) = E(e^{itX}) = E(e^{it\sum_{j=1}^n X_j}) \stackrel{indep.}{=} \prod_{j=1}^n E(e^{itX_j}) = \prod_{j=1}^n (p_j e^{it} + q_j), \quad \mathbf{i} = \sqrt{-1} \quad (2.7)$$

- Função de Distribuição Acumulada (FDA)

$$F_X(x|\mathbf{p}) = P(X \leq x) = \sum_{k=0}^x f(k|\mathbf{p}) = \sum_{k=0}^x \sum_{A \in B_x} \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i) \quad (2.8)$$

Os cálculos oriundos da distribuição PB são trabalhosos de realizar partindo diretamente de sua definição, por isso alguns autores desenvolveram métodos, exatos ou aproximados, como alternativas mais eficientes de se manipular. Este trabalho traz alguns desses métodos.

O primeiro método que a ser tratado, desenvolvido por Hong (2013), é baseado na aplicação de Transformadas de Fourier Discreta (TFD) na função característica de PB, chamado comumente de *DFT-CF method*.

Tomando a FC em (2.7) com valores de  $t = \omega l$ , onde  $l = 0, 1, \dots, n$  e  $\omega = 2\pi/(n+1)$  tem-se

$$\varphi_X(\omega l) = E(e^{i\omega l X}) = \sum_{x=0}^n e^{i\omega l x} f(x|\mathbf{p}) \quad (2.9)$$

e convenientemente multiplicando por  $1/(n+1)$

$$\frac{1}{n+1} \varphi_X(\omega l) = \frac{1}{n+1} \sum_{x=0}^n e^{i\omega l x} f(x|\mathbf{p}) \quad (2.10)$$

Agora aplicando a TFD em (2.10) tem-se

$$\begin{aligned} f(x|\mathbf{p}) &= \sum_{l=0}^n \frac{1}{n+1} \varphi_X(\omega l) e^{-i\omega l x} \\ &= \frac{1}{n+1} \sum_{l=0}^n \varphi_X(\omega l) e^{-i\omega l x}, \quad x = 0, 1, \dots, n \end{aligned} \quad (2.11)$$

então teremos que a FDA da Poisson-Binomial é dada por

$$F_X(x) = \sum_{k=0}^x \frac{1}{n+1} \sum_{l=0}^n \varphi_X(\omega l) e^{-i\omega l k} \quad (2.12)$$

$$= \frac{1}{n+1} \sum_{k=0}^x \sum_{l=0}^n \varphi_X(\omega l) e^{-i\omega l k} \quad (2.13)$$

Um segundo método para o cálculo das probabilidades da função de distribuição de PB é pela aproximação normal, com base no conhecido Teorema do Limite Central. Esta aproximação é dada por

$$F_X(x) \approx \Phi\left(\frac{x + 0,5 - \mu}{\sqrt{\sigma^2}}\right), \quad x = 0, 1, \dots, n \quad (2.14)$$

onde  $\Phi$  é a função de distribuição acumulada da normal padrão, 0,5 é a chamada correção de continuidade e  $\mu$  e  $\sigma^2$  são dados em (2.3) e (2.4), respectivamente.

Outra alternativa é a fórmula recursiva elaborada por Lord e Wingersky (1984), denominada de algoritmo LW, tal que com  $n$  ensaios independentes tem-se

$$f_n(x|\mathbf{p}) = f_{n-1}(x-1|\mathbf{p}) \cdot p_n + f_{n-1}(x|\mathbf{p}) \cdot q_n, \quad x \in (1, 2, \dots, n-1) \quad (2.15)$$

no qual

$$f_1(0|\mathbf{p}) = q_1;$$

$$f_1(1|\mathbf{p}) = p_1;$$

$$f_n(0|\mathbf{p}) = f_{n-1}(0|\mathbf{p}) \cdot q_n; \text{ e}$$

$$f_n(n|\mathbf{p}) = f_{n-1}(n-1|\mathbf{p}) \cdot p_n.$$

Tomando como exemplo  $n = 3$  tem-se

$$\begin{aligned} f_3(0|\mathbf{p}) &= f_2(0|\mathbf{p}) \cdot q_3 = f_1(0|\mathbf{p}) \cdot q_2 q_3 = q_1 q_2 q_3 \\ f_3(1|\mathbf{p}) &= f_2(0|\mathbf{p}) \cdot p_3 + f_2(1|\mathbf{p}) \cdot q_3 = f_1(0|\mathbf{p}) \cdot q_2 p_3 + [f_1(0|\mathbf{p}) \cdot p_2 + f_1(1|\mathbf{p}) \cdot q_2] \cdot q_3 \\ &= q_1 q_2 p_3 + [q_1 p_2 + p_1 q_2] q_3 = q_1 q_2 p_3 + q_1 p_2 q_3 + p_1 q_2 q_3 \\ f_3(2|\mathbf{p}) &= f_2(1|\mathbf{p}) \cdot p_3 + f_2(2|\mathbf{p}) \cdot q_3 = [f_1(0|\mathbf{p}) \cdot p_2 + f_1(1|\mathbf{p}) \cdot q_2] \cdot p_3 + [f_1(1|\mathbf{p}) \cdot p_2] \cdot q_3 \\ &= [q_1 p_2 + p_1 q_2] \cdot p_3 + p_1 p_2 q_3 = q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3 \\ f_3(3|\mathbf{p}) &= f_2(2|\mathbf{p}) \cdot p_3 = f_1(1|\mathbf{p}) \cdot p_2 p_3 = p_1 p_2 p_3 \end{aligned}$$

Estes resultados são exatamente os mesmos encontrados se calculados pela definição da distribuição de Poisson-Binomial em (2.2).

Existem outros métodos utilizados para obter as probabilidades através do modelo PB que podem ser consultados em González, Wiberg e von Davier (2016).

## 2.3 A Binomial Composta e suas propriedades

Considere agora  $X$  não somente como um número total de acertos, mas também como um padrão de resposta de determinado indivíduo. Suponha que um teste contenha, em equilíbrio, questões fáceis e difíceis, então um candidato pode acertar metade das questões, porém sua nota, que varia de 0 a 10, não será necessariamente 5, mas será de acordo com suas questões acertadas, variando de mais alta, se acertar as difíceis, à mais baixa, se acertar as fáceis. Esse padrão de respostas nos indicará o nível do sujeito, ou sua habilidade.

Em Teoria da Resposta ao Item Lord e Novick (1968) desenvolvem o modelo de distribuição Binomial Composto (BC) baseado nesse padrão de respostas.

**Definição:** Seja  $X$  a soma do número de respostas certas em um teste de  $n$  questões (independentes entre si) de um indivíduo com habilidade  $\theta$ , tal que

$$X = \sum_{i=1}^n X_i$$

onde  $X_i$  assume valor 1 com probabilidade  $p_i$ , se o indivíduo respondeu corretamente

a questão  $i$ , e valor 0 com probabilidade  $q_i = 1 - p_i$ , se ao responder errou. Então, a distribuição condicional de  $X$  dado  $\theta$ , chamada distribuição binomial composta, é

$$f(x|\theta) = \sum_{T_x} \left( \prod_{i=1}^n p_i^{x_i} q_i^{1-x_i} \right) \quad (2.16)$$

em que  $T_x$  é o conjunto que contém todos os padrões de respostas com exatamente  $x$  respostas corretas.

Ao assumir, em particular,  $n = 3$  temos o seguinte:

$$\begin{aligned} T_0 &= \{(0, 0, 0)\} & \Rightarrow & f(0|\theta) = q_1 q_2 q_3; \\ T_1 &= \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} & \Rightarrow & f(1|\theta) = p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3; \\ T_2 &= \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\} & \Rightarrow & f(2|\theta) = p_1 p_2 q_3 + p_1 q_2 p_3 + q_1 p_2 p_3; \\ T_3 &= \{(1, 1, 1)\} & \Rightarrow & f(3|\theta) = p_1 p_2 p_3. \end{aligned}$$

Note que são exatamente as mesmas probabilidades encontradas se tomada a distribuição Poisson-Binomial para o mesmo valor de  $n = 3$ . Segundo González, Wiberg e von Davier (2016), as distribuições PB e BC geram as mesmas probabilidades, o que as torna equivalentes. Por serem equivalentes as propriedades encontradas de 2.3 a 2.8 serão válidas para a Binomial Composta, assim como também o algoritmo LW.

A distribuição BC pode ser aproximada também utilizando a seguinte equação, composta por Lord e Novich (1968):

$$\begin{aligned} f(x|\theta) = \delta_n(x) &+ \frac{n}{2} V_2 C_2(x) + \frac{n}{3} V_3 C_3(x) + \left( \frac{n}{4} V_4 - \frac{n^2}{8} V_2^2 \right) C_4(x) \\ &+ \left( \frac{n}{5} V_5 - \frac{5n^2}{6} V_2 V_3 \right) C_5(x) + \dots, \quad x = 0, 1, \dots, n \end{aligned} \quad (2.17)$$

onde

$$\begin{aligned} \delta_n(x) &= \binom{n}{x} (\bar{p})^x (1 - \bar{p})^{n-x}, \quad x = 0, 1, \dots, n \\ C_r(x) &= \sum_{v=0}^r (-1)^{v+1} \binom{r}{v} \delta_{n-r}(x-v), \quad r = 2, 3, \dots, n \\ V_r &= \frac{1}{n} \sum_{i=1}^n (p_i - \bar{p})^r, \quad r = 2, 3, \dots, n \\ \bar{p} &= \frac{1}{n} \sum_{i=1}^n p_i \end{aligned}$$

A Equação 2.17 é uma série finita de ordem  $n$ , na qual sendo usados todos os  $n$  termos teremos uma identidade exata. Do ponto de vista prático, é possível truncar a série, de tal maneira que quanto mais próximo de  $n$  seja a ordem dos termos mais aproximada é a série. Se tomado apenas o primeiro termo da série então se terá uma distribuição binomial de parâmetros  $(n, \bar{p})$ .

---

## Capítulo 3

# A Teoria da Resposta ao item

---

### 3.1 Conceitos Iniciais

A Teoria da Resposta ao Item (TRI) propõe modelos matemáticos para medir traços latentes, ou seja, características não observáveis que não podem ser medidas diretamente, por exemplo, a satisfação profissional de uma pessoa, a qualidade de vida em um país ou a proficiência de um indivíduo em determinada área de conhecimento. Essa metodologia se baseia não somente no número total de acertos de um teste, mas visa sobretudo as propriedades dos itens (questões) do mesmo e a habilidade (ou habilidades) do respondente. Ela sugere que quanto maior a habilidade do indivíduo, maior a probabilidade de acerto no item (ANDRADE; TAVARES; VALLE, 2000).

A TRI reúne modelos que têm como pressupostos a unidimensionalidade do teste e deriva que existe independência local entre os itens. O primeiro diz que a probabilidade de acertar um item depende unicamente de um fator ou habilidade, a saber,  $\theta$ . O segundo, também chamado de independência condicional, afirma que para uma pessoa com determinado valor  $\theta$  sua resposta a um item não é influenciada por suas respostas aos outros itens, ou seja, as respostas são independentes (MUÑIZ, 2018). Matematicamente, dado um vetor de respostas  $U_1, U_2, \dots, U_n$  de determinado indivíduo com habilidade  $\theta$ , tem-se

$$P(U_1, U_2, \dots, U_n | \theta) = P(U_1 | \theta) P(U_2 | \theta) \cdots P(U_n | \theta) = \prod_{i=1}^n P(U_i | \theta) \quad (3.1)$$

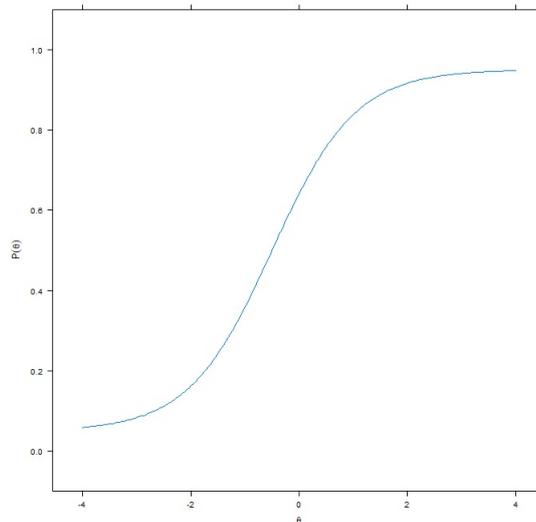
onde a variável  $U_i$  ( $i = 1, 2, \dots, n$ ) admite apenas dois valores:  $u_i = 1$ , quando se acerta, ou  $u_i = 0$ , quando se erra o item.

Existem também em TRI modelos multidimensionais (MIRT - *Multidimensional Item Response Theory*), onde o item dependerá de mais de uma habilidade. Por exemplo, um item que exige tanto habilidade verbal (interpretação de texto) quanto habilidade matemática, neste caso deverá ser ajustado um modelo de duas dimensões. Neste trabalho será

abordado apenas sobre os modelos unidimensionais, mas para leituras sobre modelos de resposta ao item multidimensionais pode ser consultado Reckase, 2009.

Os modelos da TRI assumem que existe uma relação funcional entre  $\theta$  e a probabilidade de acertar o item e esta dita função é denominada Curva Característica do Item (CCI). As CCI utilizadas em Teoria da Resposta ao Item serão do "tipo S", conforme o exemplo na Figura 3.1, e para delinear-las é necessário conhecer seus parâmetros, os quais serão descritos na seção seguinte.

Figura 3.1 *Exemplo de uma Curva Característica do Item.*



Conforme o número de parâmetros considerados e o tipo de função matemática adotada teremos diferentes modelos de CCI. No caso dos modelos bidimensionais, a CCI recebe a denominação de "superfície característica do item" e para modelos com mais de duas dimensões se usa o termo "hiper superfície de resposta ao item", pois nesses casos não se tem mais uma curva no plano, mas sim uma função em espaço multidimensional (KOSE e DEMIRTASLI, 2012).

## 3.2 O Modelo Logístico de Quatro Parâmetros e seus Submodelos

Em Teoria da Resposta ao Item tem-se utilizado frequentemente duas funções matemáticas para a CCI, a curva normal acumulada (ou ogiva normal) e a função logística, da qual trataremos neste trabalho.

Para se conheça a forma da CCI é necessário que se saiba os valores dos parâmetros dos itens e quais parâmetros são considerados. Em seguida, apresenta-se quatro parâmetros, denominados  $a, b, c$  e  $\gamma$ , usados nos modelos de resposta ao item.

O parâmetro  $a$ , denominado parâmetro de discriminação ou parâmetro de inclinação, mede a capacidade do item em discriminar os valores altos e baixos da habilidade e seu valor é proporcional à inclinação da reta tangente à CCI no ponto de máxima inclinação da curva.

O parâmetro  $b$ , chamado parâmetro de dificuldade, indica quão difícil o item é e seu valor corresponde a  $\theta$  no ponto de máxima inclinação da CCI.

O parâmetro  $c$ , representa a probabilidade de resposta correta por um indivíduo com baixo valor de  $\theta$ .

E o parâmetro  $\gamma$ , corresponde à probabilidade de um sujeito com alta habilidade não responder o item corretamente. Seu valor é ligeiramente inferior a 1.

A seguir, os modelos que serão apresentados assumem que os itens do teste são dicotômicos, corrigidos como certo ou errado, ou dicotomizados, quando o item é de caráter aberto.

O primeiro modelo aqui citado, apresentado em Barton e Lord (1981), é o modelo logístico de quatro parâmetros (ML4) e é dado por

$$P_{ji}(\theta_j) \equiv P(U_{ji} = 1|\theta_j) = c_i + (\gamma_i - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}, \quad (3.2)$$

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

$P(U_{ji} = 1|\theta_j)$  é a probabilidade de um indivíduo  $j$  com habilidade  $\theta_j$  acertar o item  $i$ ;

$\theta_j$  representa a habilidade (traço latente ou proficiência) do indivíduo  $j$ ;

$a_i$  é o parâmetro de discriminação do item  $i$ ;

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

$c_i$  é o valor de  $P_{ji}(\theta_j)$  quando  $\theta_j = -\infty$ ;

$\gamma_i$  é o valor de  $P_{ji}(\theta_j)$  quando  $\theta_j = +\infty$ ;

$e$  é a base do logaritmo neperiano (aproximadamente 2,718); e

$D$  é constante, que ao tomar o valor 1,7 a função logística se aproxima a ogiva normal.

Três outros modelos são oriundos do ML4, como se fossem versões restritas, onde se fixam os valores de um ou mais parâmetros. Ao tomar  $\gamma = 1$  tem-se o modelo logístico de três parâmetros (ML3). Para obter o modelo logístico de dois parâmetros (ML2) é adotado

também  $c = 0$ . E finalmente, o modelo logístico de um parâmetro (ML1), ou modelo de Rasch, é obtido fixando ainda  $a = 1$ . Portanto, os modelos tornam-se:

$$ML3 : P(U_{ji} = 1|\theta_j) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}, j = 1, \dots, n \text{ e } i = 1, \dots, I \quad (3.3)$$

$$ML2 : P(U_{ji} = 1|\theta_j) = \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}, j = 1, \dots, n \text{ e } i = 1, \dots, I \quad (3.4)$$

$$ML1 : P(U_{ji} = 1|\theta_j) = \frac{e^{D(\theta_j - b_i)}}{1 + e^{D(\theta_j - b_i)}}, j = 1, \dots, n \text{ e } i = 1, \dots, I \quad (3.5)$$

### 3.3 Processo de Estimação pelo Método da Máxima Verossimilhança Marginal

Após a seleção de um dos modelos, o passo seguinte é estimar (ou calibrar) os parâmetros de cada item e o valor de  $\theta$  para cada indivíduo a partir das respostas empíricas das pessoas aos itens.

Para isto, um dos meios utilizados é o método de Máxima Verossimilhança Marginal (MVM), que consiste em estimar primeiramente os parâmetros dos itens, considerando uma função densidade de probabilidade (fdp) para  $\theta$  (não necessariamente no sentido bayesiano), e depois, com estes já conhecidos, estimar as habilidades dos indivíduos.

Considera-se, primeiramente, que as habilidades  $\theta_j (j = 1, 2, \dots, n)$  são realizações de uma variável aleatória contínua  $\theta$ , com fdp  $g(\theta|\eta)$ , onde o vetor  $\eta$  é conhecido. Para a estimação dos parâmetros dos itens deve-se marginalizar a função de verossimilhança, integrando-a com respeito a  $\theta$ .

Tomando o vetor aleatório  $\mathbf{U}_j. = (\mathbf{U}_{j1}, \mathbf{U}_{j2}, \dots, \mathbf{U}_{jI})$  de respostas do indivíduo  $j$ , a probabilidade marginal de  $\mathbf{U}_j.$  é dada por

$$P(\mathbf{U}_j. = \mathbf{u}_j. | \zeta, \eta) \equiv P(\mathbf{u}_j. | \zeta, \eta) = \int_{\mathbb{R}} P(\mathbf{u}_j. | \theta, \zeta, \eta) g(\theta | \eta) d\theta$$

e utilizando que a distribuição de  $\mathbf{U}_j.$  não depende de  $\eta$ , então

$$P(\mathbf{u}_j. | \zeta, \eta) = \int_{\mathbb{R}} P(\mathbf{u}_j. | \theta, \zeta) g(\theta | \eta) d\theta \quad (3.6)$$

Usando a independência entre as respostas de diferentes indivíduos, pode-se escrever a verossimilhança associada ao vetor  $\mathbf{U}_.. = (\mathbf{U}_{1.}, \mathbf{U}_{2.}, \dots, \mathbf{U}_{n.})$ , como

$$L(\zeta, \eta) = P(\mathbf{u}_.. | \zeta, \eta) = \prod_{j=1}^n P(\mathbf{u}_j. | \zeta, \eta), \quad (3.7)$$

porém, por vantagens computacionais, tem se utilizado a abordagem de Padrão de Respostas, em que os dados seguem uma distribuição multinomial, na qual

$$L(\zeta, \eta) = \frac{n!}{r_1! \cdots r_s!} [P(\mathbf{u}_1 | \zeta, \eta)]^{r_1} \cdots [P(\mathbf{u}_s | \zeta, \eta)]^{r_s} \quad (3.8)$$

Essa abordagem consiste em tomar a frequência dos padrões de respostas. Se um teste contém  $I$  itens e as respostas registradas como 1 - certo ou 0 - errada, então são possíveis  $2^I$  padrões de respostas diferentes. Seja  $r_l$  o total de sujeitos que respondem um padrão  $l$ , onde  $l = 1, 2, \dots, s$ , tal que  $s \leq \min(n, 2^I)$  é o número de padrão de respostas com  $r_l > 0$  de amostra com  $n$  indivíduos, portanto

$$\sum_{l=1}^s r_l = n$$

Assim, a frequência de padrões de respostas terá distribuição multinomial com parâmetros  $(n, P(\mathbf{u}_1 | \zeta, \eta), \dots, P(\mathbf{u}_s | \zeta, \eta))$ , que é exatamente a função de verossimilhança dada na Equação 3.8.

Agora, deve-se encontrar os valores dos parâmetros que conjuntamente maximizam a função  $L(\zeta, \eta)$ , ou seja, é preciso resolver

$$\frac{\partial \log L(\zeta, \eta)}{\partial \zeta_i} = 0, \quad i = 1, 2, \dots, I \quad (3.9)$$

O logaritmo de 3.8 é

$$\begin{aligned} \log L(\zeta, \eta) &= C + r_1 \log P(\mathbf{u}_1 | \zeta, \eta) + \cdots + r_s \log P(\mathbf{u}_s | \zeta, \eta) \\ &= C + \sum_{l=1}^s r_l \log P(\mathbf{u}_l | \zeta, \eta) \end{aligned}$$

onde  $C$  não depende dos parâmetros dos itens.

Então,

$$\begin{aligned} \frac{\partial \log L(\zeta, \eta)}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left\{ \sum_{l=1}^s r_l \log P(\mathbf{u}_l | \zeta, \eta) \right\} \\ &= \sum_{l=1}^s r_l \frac{1}{P(\mathbf{u}_l | \zeta, \eta)} \frac{\partial P(\mathbf{u}_l | \zeta, \eta)}{\partial \zeta_i} \end{aligned} \quad (3.10)$$

Segundo Andrade, Tavares e Valle (2000), a derivada da Equação 3.6, contida em 3.9 para o padrão de resposta  $l$ , pode ser escrita como

$$\frac{\partial P(\mathbf{u}_l | \zeta, \eta)}{\partial \zeta_i} = \int_{\mathbb{R}} \left[ \frac{(u_{li} - P_i)}{P_i Q_i} \left( \frac{\partial P_i}{\partial \zeta_i} \right) \right] P(u_l | \theta, \zeta) g(\theta | \eta) d\theta \quad (3.11)$$

onde  $P_i = P(U_i = 1|\theta, \zeta_i)$  e  $Q_i = P(U_i = 0|\theta, \zeta_i)$ .

Considerando agora

$$W_i - \frac{P_i^* Q_i^*}{P_i Q_i}, \quad \text{onde } P_i^* = \{1 + e^{-Da_i(\theta - b_i)}\}^{-1} \text{ e } Q_i^* = 1 - P_i^*$$

$$\text{e } g_i^*(\theta) \equiv g(\theta|\mathbf{u}_{l.}, \zeta, \eta) = \frac{P(u_{li}|\theta, \zeta)g(\theta|\eta)}{P(u_{li}|\zeta, \eta)}$$

então, após alguns passos, da Equação 3.9 são encontradas as seguintes equações para estimação dos parâmetros do item  $i$ ,  $i = 1, 2, \dots, I$

$$a_i : D(\gamma_i - c_i) \sum_{l=1}^s r_l \int_{\mathbb{R}} [(u_{li} - P_i)(\theta - b_i)W_i] g_i^*(\theta) d\theta = 0 \quad (3.12)$$

$$b_i : -Da_i(\gamma_i - c_i) \sum_{l=1}^s r_l \int_{\mathbb{R}} [(u_{li} - P_i)W_i] g_i^*(\theta) d\theta = 0 \quad (3.13)$$

$$c_i : \sum_{l=1}^s r_l \int_{\mathbb{R}} \left[ (u_{li} - P_i) \frac{W_i}{P_i^*} \right] g_i^*(\theta) d\theta = 0 \quad (3.14)$$

$$\gamma_i : \sum_{l=1}^s r_l \int_{\mathbb{R}} \left[ (u_{li} - P_i) \frac{W_i}{Q_i^*} \right] g_i^*(\theta) d\theta = 0 \quad (3.15)$$

as quais não possuem solução explícita, tendo, assim, que recorrer a métodos numéricos iterativos.

Após os parâmetros dos itens calibrados é feita a segunda etapa do processo de estimação, que é estimar os valores de  $\theta$ . Estes podem ser estimados individualmente por máxima verossimilhança ou pela moda ou média da distribuição condicional de  $(\theta_j|\mathbf{u}_j)$ .

### 3.4 Estimação de Parâmetros com o uso do Pacote mirt do R

Através do software livre R é possível fazer análise estatística de dados. Um dos pacotes utilizados no R para a estimação de parâmetros, via máxima verossimilhança, em modelos multidimensionais da TRI é o MIRT (Multidimensional Item Response Theory), porém podendo também ser usado para dados unidimensionais. Este pacote foi apresentado em Chalmers (2012), onde o autor expõe brevemente a implementação do mesmo. No site <https://github.com/philchalmers/mirt/wiki> também é possível ter mais informações do uso do pacote mirt.

Para ilustrar o emprego do mirt faz-se uso do conjunto de dados LSAT7, disponível do

R, no qual contém 5 variáveis (itens) dicotômicas e 1000 observações. Esses dados estão em forma de tabela de frequência levando em conta os padrões de respostas idênticos, porém para estimar os modelos os dados devem estar em formato expandido, para isso é usado a seguinte função:

```
> library("mirt")
> lsat <- expand.table(LSAT7)
```

A função `mirt(data = lsat, model = 1)` tem como padrão estimar o modelo TRI de dois parâmetros, onde `data` representa o banco de dados dicotomizados e `model` é a dimensão, ou número de fatores, do modelo. Para estimar outros tipos de modelos pode ser usado o argumento `itemtype = ' '` dentro da função e definido qual modelo a ser utilizado, como `'Rasch'`, `'2PL'`, `'3PL'`, `'4PL'` para modelos Rasch, de 2, 3 e 4 parâmetros, respectivamente, ou para definir um modelo diferente para cada item em particular use-se `'c(2PL, 2PL, 3PL, 3PL, 4PL)'`. Portanto, por exemplo, tem-se

```
> doisPL <- mirt(lsat, 1)
> tresPL <- mirt(lsat, 1, itemtype = '3PL')
> mixPL <- mirt(lsat, 1, itemtype = 'c(2PL, 2PL, 3PL, 3PL, 4PL)')
```

O comando para extração dos parâmetros é dado por

```
> coef(doisPL, simplify = TRUE)$items
      a1      d  g  u
Item.1 0.9879254 1.8560605 0 1
Item.2 1.0808847 0.8079786 0 1
Item.3 1.7058006 1.8042187 0 1
Item.4 0.7651853 0.4859966 0 1
Item.5 0.7357980 1.8545127 0 1

> coef(doisPL, IRTpars = TRUE, simplify = TRUE)$items
      a      b  g  u
Item.1 0.9879254 -1.8787456 0 1
Item.2 1.0808847 -0.7475160 0 1
Item.3 1.7058006 -1.0576962 0 1
Item.4 0.7651853 -0.6351358 0 1
Item.5 0.7357980 -2.5204102 0 1
```

onde na primeira saída tem-se as colunas `a1` que representa o parâmetro de discriminação, `d` o intercepto, onde  $d = -ab$ , `g` derivado do inglês *guessing* que corresponde ao parâmetro  $c$  e `u` também do inglês *upper* que equivale ao parâmetros  $\gamma$ . E na segunda saída temos os coeficientes na forma tradicional, onde `a` é o parâmetro de discriminação, `b` é o parâmetro de dificuldade e `g` e `u` são os mesmos da primeira saída.

Com a função abaixo pode-se obter as estimativas das habilidades  $\theta$ .

```
> theta <- fscores(doisPL)
```

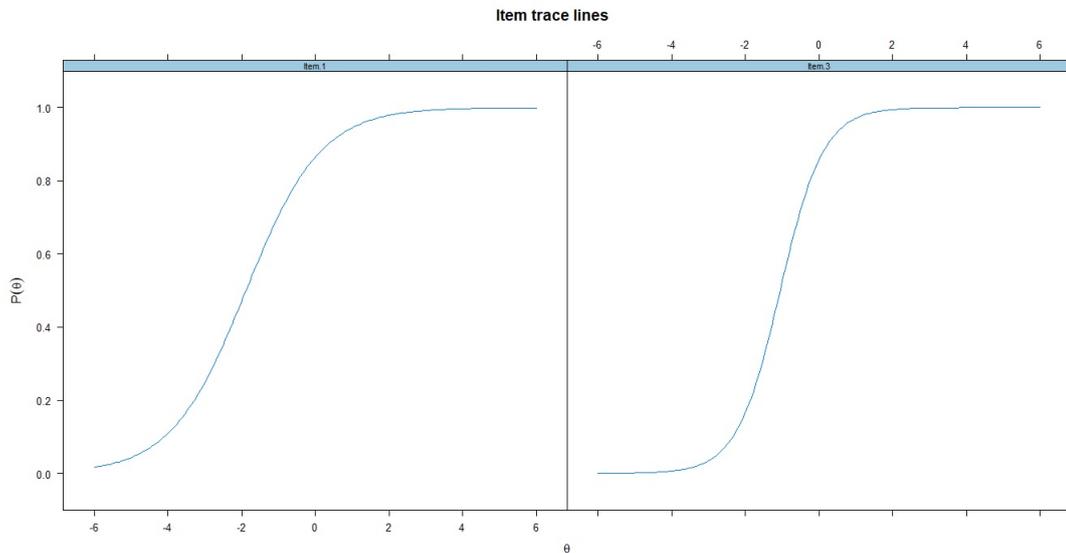
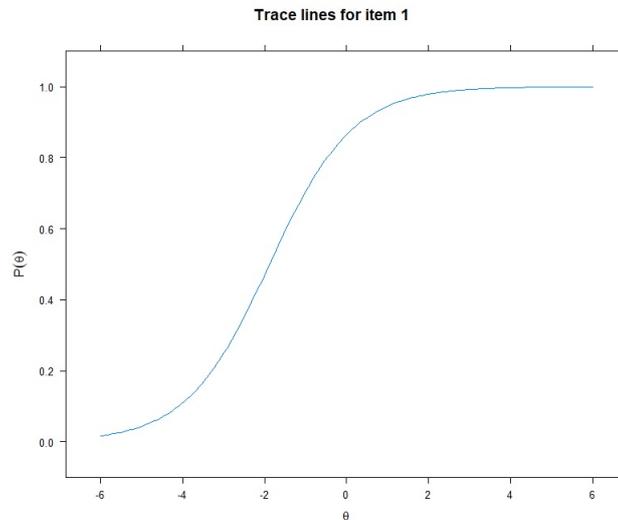
```
> head(theta)
```

```
          F1
[1,] -1.869854
[2,] -1.869854
[3,] -1.869854
[4,] -1.869854
[5,] -1.869854
[6,] -1.869854
```

E para gerar gráficos pode-se utilizar as funções `plot()` ou `itemplot()`, onde na primeira é possível escolher mais de um item para visualizar graficamente através do argumento `which.items`. Os gráficos gerados pelos dois comandos abaixo são mostrados nas Figuras 3.2 e 3.3.

```
> plot(doisPL, type = 'trace', which.items = c(1,3))
```

```
> itemplot(doisPL, item = 1)
```

Figura 3.2 *Exemplo de gráfico usando a função plot()*Figura 3.3 *Exemplo de gráfico usando a função itemplot()*

---

## Capítulo 4

# ICR: Índice de Coerência de Resposta

---

Neste capítulo é abordado o tema central deste trabalho, o estimador ICR, que será usado para identificar indivíduos que responderam de forma totalmente aleatória aos itens (dicotômicos) de um teste ou de um questionário.

Primeiro é apresentado a variável escore, que será usada na definição do ICR. Depois é feito um estudo do índice em questão, analisando como se comporta em diferentes situações. E, por fim, a demonstração do critério de decisão, onde efetivamente é delimitado, a partir do ICR, quando há ou não indícios de aleatoriedade no padrão de resposta.

### 4.1 A Variável Escore ( $S$ ) e suas Propriedades

A variável escore ( $S$ ) corresponde ao número de acertos de um indivíduo no teste, ou seja

$$S = \sum_{i=1}^n X_i$$

onde  $n$  é o total de itens do teste e  $X_i$  é a resposta do indivíduo ao item  $i$ , que toma os valores zero, para respostas incorretas, e um para as corretas. Assim, como o teste é composto de  $n$  itens, então  $S$  varia de 0 a  $n$ .

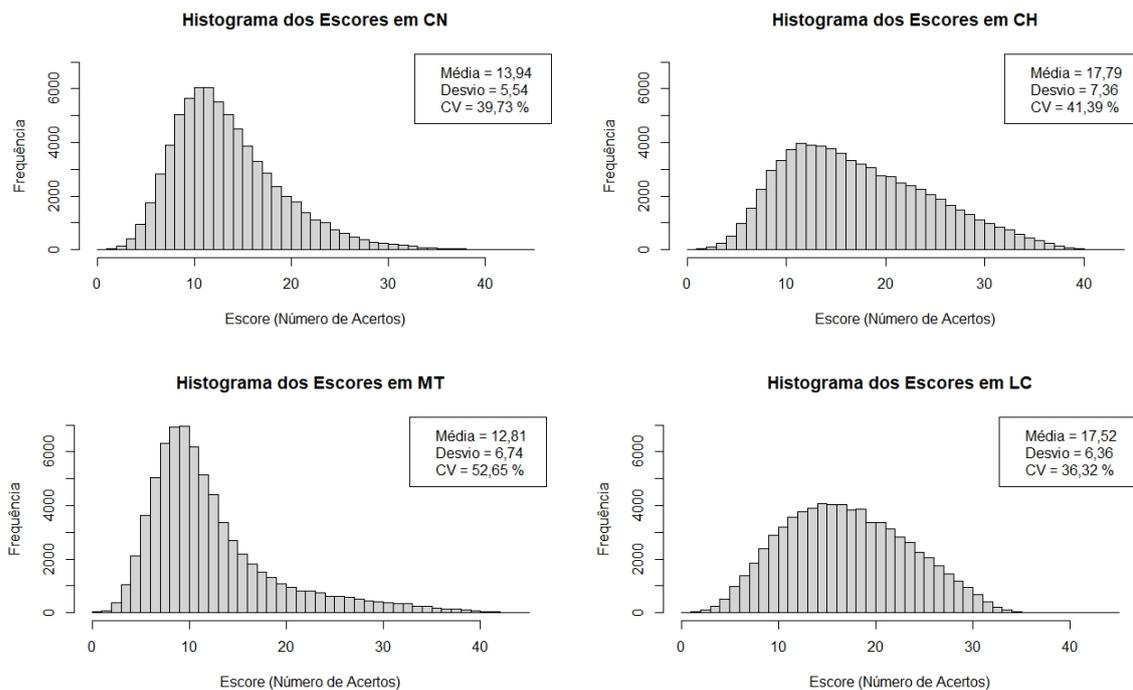
Considere, por exemplo, as respostas dos candidatos que fizeram o ENEM (Exame Nacional do Ensino Médio) no ano de 2019 na cidade de Fortaleza - CE. O ENEM é composto por 180 questões, que abrange quatro áreas de conhecimento, são elas: CN - Ciências da Natureza e suas Tecnologias; CH - Ciências Humanas e suas Tecnologias; MT - Matemática e suas Tecnologias; e LC - Linguagens, Códigos e suas Tecnologias. Cada área contém 45 questões, porém em LC desconsidera-se os itens referentes à língua estrangeira,

ou seja, são computadas apenas 40 questões, e em CH uma questão foi anulada (questão 90 do caderno azul), assim esta também não será considerada.

A partir dos citados dados foram calculadas algumas estatísticas e dispostas na Tabela 4.2. Para os cálculos das medidas amostrais foram utilizadas as fórmulas que constam na Tabela 4.1.

A Figura 4.1 apresenta o histograma dos escores em cada área do ENEM-2019, nela observa-se que a distribuição do escore é distinta para cada área, nos quatro casos os dados se concentram à esquerda dos gráficos, o que significa que a maioria dos respondentes acertaram menos da metade da prova por área. Graficamente nota-se que há assimetria nas distribuições.

Figura 4.1 *Histograma dos Escores no ENEM-2019, Fortaleza-CE.*



O coeficiente de assimetria é aqui denominado  $AS$ . Se  $AS < 0$ , a distribuição tem caudas mais pesadas à esquerda e diz-se que há assimetria negativa (ou à esquerda). Se  $AS > 0$ , as caudas da distribuição são mais pesadas a direita e tem-se assimetria positiva (ou à direita). E se a distribuição é simétrica,  $AS = 0$ . Então, calculando o coeficiente de assimetria de cada área de conhecimento do ENEM tem-se que todos são positivos (ver Tabela 4.2), ou seja, possuem assimetria positiva.

Tabela 4.1 *Algumas Fórmulas Estatísticas para Medidas Amostrais.*

Estatística	Fórmula	Função no R	Pacote
Média	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	<code>mean</code>	<code>base</code>
Desvio Padrão	$s = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}}$	<code>sd</code>	<code>stats</code>
Coef. Variação	$CV = \frac{s}{\bar{X}} 100$		
Coef. Assimetria	$AS = \frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s} \right)^3$	<code>skewness</code>	<code>e1071</code>
Coef. Curtose	$K = \frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s} \right)^4 - 3$	<code>kurtosis</code>	<code>e1071</code>

Tabela 4.2 *Algumas Estatísticas do ENEM-2019, Fortaleza-CE.*

Área	Média	Desvio Padrão	Coef. Variação (%)	Coef. Assimetria	Coef. Curtose
CN	13,94370	5,53982	39,72989	0,93392	1,09137
CH	17,78737	7,36207	41,38930	0,49832	-0,41285
MT	12,80573	6,74235	52,65106	1,53882	2,47246
LC	17,52173	6,36375	36,31922	0,19792	-0,61556

Graficamente, a Figura 4.1 também mostra que as distribuições tem "formas de achatamento" diferentes, em CN e MT as distribuições são mais finas e altas, e em CH e LC elas são mais grossas e baixas. O coeficiente de curtose é que traz informação sobre esse tipo de "formato" das distribuições. Se  $K < 0$  a distribuição é chamada de platicúrtica, se  $K = 0$ , de mesocúrtica e se  $K > 0$  tem-se uma distribuição leptocúrtica. Assim, nas áreas de CN e MT as distribuições são leptocúrticas, pois têm coeficientes de curtose negativos, e em CH e LC seus coeficientes de curtose são positivos, logo possuem distribuição platicúrtica.

Portanto, em média os alunos acertaram mais questões em Ciências Humanas e Linguagens e Códigos, áreas que envolvem mais a leitura, cerca de 17 itens corretos, porém em LC há uma menor dispersão dos dados em relação à média do que em CH. Já nas áreas de cálculo, MT e CN, os alunos acertaram em média de 12 a 14 itens, com  $CV$  próximos de 39% e 52%, respectivamente, o que nos diz que há uma grande variabilidade nos dados em torno da média.

## 4.2 Índice de Coerência de Resposta - ICR

A partir das respostas a um teste de múltipla escolha, pode-se construir uma matriz de probabilidades. Cada item tem seus parâmetros e cada indivíduo sua proficiência, daí tem-se uma matriz, onde para cada item  $i$  e indivíduo  $j$  há uma probabilidade  $P_{i,j}$  distinta.

É sabido que ao se somar  $n$  Bernoulli's, que são variáveis aleatórias de 0's e 1's com probabilidade de um igual a  $p$ , se tem, se independentes, uma Binomial de parâmetros  $n$  e  $p$ . Mas, como descrito acima, as probabilidades não são iguais, porém possuem independência local, ou seja, dada a proficiência existe independência, então somando as variáveis tem-se uma Poisson-Binomial, que tem como parâmetros as probabilidades de cada item.

Como aqui se está trabalhando com somas de variáveis aleatórias independentes, é proposto uma padronização nos escores dos respondentes, ou seja, subtraindo pela média e dividindo pelo desvio padrão, o que, como se verá mais adiante, torna a distribuição próxima de uma distribuição Normal.

**Definição:** O Índice de Coerência de Resposta (ICR) será definido como

$$ICR = \frac{S - E(S)}{DP(S)}, \quad (4.1)$$

onde

$S$  é a variável score,  $S = \sum_{i=1}^n X_i$ , tal que  $X_i$  é a resposta ao item (0 ou 1);

$E(S)$  é a esperança matemática de  $S$ ,  $E(S) = \sum_{i=1}^n p_i$ , tal que  $p_i$  é a probabilidade de acertar o item  $i$ ; e

$DP(S)$  é o desvio-padrão de  $S$ ,  $DP(S) = \sum_{i=1}^n p_i(1 - p_i)$ .

A variável  $S$  é soma de Bernoulli's independentes com parâmetros diferentes, então  $S$  tem distribuição Poisson-Binomial, disso tem-se que o valor esperado do score ou score verdadeiro,  $E(S)$ , é a soma das probabilidades de acerto de cada item (veja Equação 2.3) e cada uma dessas probabilidades são oriundas da TRI.

Partindo diretamente da definição da distribuição Poisson-Binomial é laborioso calcular as probabilidades a medida que o número de amostras aumenta, conseqüentemente, é trabalhoso também calcular os valores do ICR, então faz-se necessário uma aproximação para que se possa fazer os cálculos de maneira mais simplificada.

Como a variável  $S$  é soma de Bernoulli's independentes com parâmetros distintos, não

é possível utilizar o Teorema Central do Limite (TCL), que exige que se tenha parâmetros iguais (as variáveis devem ser independentes e identicamente distribuídas), para calcular uma convergência. Então, é utilizado o Teorema Central do Limite de Liapunov (BLAI-OTTA e DELIEUTRAZ, 2004) que trata deste caso e, então, será provado que o ICR tende em distribuição para a Normal(0,1).

### Teorema Central do Limite de Liapunov

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes cujos três primeiros momentos existem, e sejam  $m_i = E(X_i)$ ,  $\sigma_i^2 = Var(X_i) < \infty$ , e  $a_i = E(|X_i - m_i|^3)$ .

Considere, ainda,

$$s_n = \sqrt{\sum_{i=1}^n \sigma_i^2} \quad A_n = \sqrt[3]{\sum_{i=1}^n a_i} \quad S_n = \frac{\sum_{i=1}^n (X_i - m_i)}{s_n}$$

Se a condição de Liapunov acontece

$$\lim_{n \rightarrow \infty} \frac{A_n}{s_n} = 0$$

, então  $S_n$  convergirá em distribuição para a Normal(0,1), ou seja,

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x)$$

onde  $F_n$  e  $\Phi(x)$  são as funções de distribuição de  $S_n$  e da Normal(0,1), respectivamente.

O ICR é basicamente a variável  $S_n$  citada no enunciado do Teorema, lembrando que a variável  $S$  tem distribuição Poisson-Binomial, então

$$\begin{aligned} ICR &= \frac{S - E(S)}{DP(S)} = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)}{\sqrt{\sum_{i=1}^n Var(X_i)}} \\ &= \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n m_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} = \frac{\sum_{i=1}^n (X_i - m_i)}{s_n} = S_n \end{aligned}$$

Sendo assim, veja a demonstração do Teorema.

### Demonstração:

Para verificar a condição de Liapunov para o ICR, antes tem-se que

$$\begin{aligned} E(|X_i - m_i|^3) &= \sum_{x_i=0}^1 |x_i - p_i|^3 p_i^{x_i} (1 - p_i)^{1-x_i} = |-p_i|^3 (1 - p_i) + |1 - p_i|^3 p_i \\ &= p_i^3 (1 - p_i) + (1 - p_i)^3 p_i, \quad 0 \leq p_i \leq 1, \quad 1 \leq i \leq n \end{aligned} \quad (4.2)$$

Logo,

$$\frac{A_n}{s_n} = \frac{\sqrt[3]{\sum_{i=1}^n [p_i^3(1-p_i) + (1-p_i)^3 p_i]}}{\sqrt{\sum_{i=1}^n p_i(1-p_i)}}. \quad (4.3)$$

Perceba que

$$\frac{p_i(1-p_i)}{2} \leq p_i^3(1-p_i) + (1-p_i)^3 p_i, \quad (4.4)$$

pois

$$\begin{aligned} \frac{p_i(1-p_i)}{2} &\leq p_i^3(1-p_i) + (1-p_i)^3 p_i \\ \frac{p_i(1-p_i)}{2} &\leq p_i(1-p_i)(p_i^2 + (1-p_i)^2) \\ \frac{1}{2} &\leq 2p_i^2 - 2p_i + 1, \quad 0 < p_i < 1 \end{aligned}$$

e  $1/2$  é ponto de mínimo de  $2p_i^2 - 2p_i + 1$ .

Além disso,

$$p_i^3(1-p_i) + (1-p_i)^3 p_i \leq p_i(1-p_i) \quad (4.5)$$

pois,

$$\begin{aligned} p_i^3(1-p_i) + (1-p_i)^3 p_i &\leq p_i(1-p_i) \\ p_i(1-p_i)(p_i^2 + (1-p_i)^2) &\leq p_i(1-p_i) \\ 2p_i^2 - 2p_i + 1 &\leq 1, \quad 0 < p_i < 1 \end{aligned}$$

Assim, pelas desigualdades 4.4 e 4.5 tem-se que

$$\begin{aligned} \frac{p_i(1-p_i)}{2} &\leq p_i^3(1-p_i) + (1-p_i)^3 p_i \leq p_i(1-p_i) \\ \sum_{i=1}^n \frac{p_i(1-p_i)}{2} &\leq \sum_{i=1}^n [p_i^3(1-p_i) + (1-p_i)^3 p_i] \leq \sum_{i=1}^n p_i(1-p_i) \\ \sqrt[3]{\frac{1}{2} \sum_{i=1}^n p_i(1-p_i)} &\leq \sqrt[3]{\sum_{i=1}^n [p_i^3(1-p_i) + (1-p_i)^3 p_i]} \leq \sqrt[3]{\sum_{i=1}^n p_i(1-p_i)} \\ \frac{\sqrt[3]{\sum_{i=1}^n p_i(1-p_i)}}{\sqrt[3]{2} \sqrt{\sum_{i=1}^n p_i(1-p_i)}} &\leq \frac{\sqrt[3]{\sum_{i=1}^n [p_i^3(1-p_i) + (1-p_i)^3 p_i]}}{\sqrt{\sum_{i=1}^n p_i(1-p_i)}} \leq \frac{\sqrt[3]{\sum_{i=1}^n p_i(1-p_i)}}{\sqrt{\sum_{i=1}^n p_i(1-p_i)}} \end{aligned}$$

$$\frac{1}{\sqrt[3]{2}\sqrt[6]{\sum_{i=1}^n p_i(1-p_i)}} \leq \frac{A_n}{s_n} \leq \frac{1}{\sqrt[6]{\sum_{i=1}^n p_i(1-p_i)}}$$

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt[3]{2}\sqrt[6]{\sum_{k=1}^n p_i(1-p_i)}} \leq \lim_{n \rightarrow \infty} \frac{A_n}{s_n} \leq \lim_{n \rightarrow \infty} \frac{1}{\sqrt[6]{\sum_{i=1}^n p_i(1-p_i)}}$$

Em ambos os denominadores extremos tem-se dentro da raiz soma de  $n$  parcelas positivas. Como  $n$  tende ao infinito então essa soma também tenderá ao infinito, embora de forma lenta pois tais parcelas estão contidas no intervalo  $(0,1)$ . Assim, os dois limites extremos tendem a zero. Então, pelo teorema do confronto

$$\lim_{n \rightarrow \infty} \frac{A_n}{s_n} = 0$$

e, portanto, é válida a condição de Liapunov. Assim, a variável  $ICR$  convergirá em distribuição para a  $Normal(0, 1)$ .

### 4.3 Estudo do ICR com parâmetros dos itens e habilidades conhecidos

Na seção anterior foi visto matematicamente que o ICR tende em distribuição a uma Normal Padrão. Agora será visto como a função se comporta empiricamente utilizando dados simulados.

Para a simulação dos parâmetros dos itens, de discriminação, dificuldade e acerto casual, foram utilizadas funções baseadas na distribuição uniforme e para as habilidades os dados foram extraídos de uma  $Normal(0,1)$ . É tomado o número de itens igual a 22, considerando 50000 indivíduos e o modelo logístico de três parâmetros para a simulação dos dados. O algoritmo da simulação completo pode ser visto no Apêndice C.

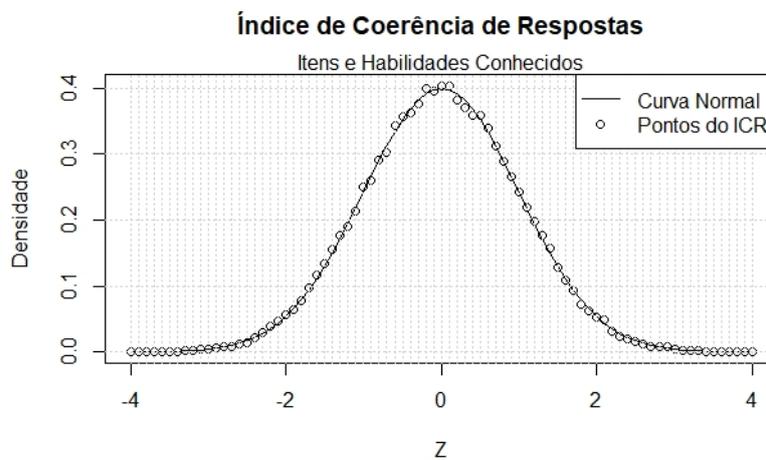
Foram simulados todos os parâmetros dos itens e das habilidades e assim gerados os valores do ICR considerando os parâmetros conhecidos. A Figura 4.2 mostra o gráfico dos pontos referentes ao ICR, e a curva Normal Padrão. Mesmo com um número pequeno de itens, os dados se ajustam bem à curva normal, exceto quando se aproxima do topo que os dados fogem um pouco, tendo uma média igual a -0,00561 e desvio padrão 0,99815,

com percentis 5% e 95% iguais a -1,6457 e 1,6244, respectivamente e um  $EQM(ICR) = 0,00002$ . Este último é o chamado Erro Quadrático Médio e é obtido a partir da seguinte fórmula:

$$EQM(ICR) = \frac{1}{n} \sum_{i=1}^n (f(x_i) - \varphi(x_i))^2,$$

onde  $f(x_i)$  é a função densidade do ICR e  $\varphi(x_i)$  é a densidade da distribuição Normal(0,1).

Figura 4.2 Gráfico do ICR com parâmetros conhecidos



## 4.4 Estudo do ICR com parâmetros dos itens conhecidos e habilidades desconhecidas

Nesta seção, a base de dados foi gerada através de simulação conforme dito na seção anterior, porém são considerados conhecidos apenas os parâmetros referentes aos itens e as habilidades são tomadas como desconhecidas. Então, é considerado como informação conhecida as respostas dos indivíduos e os parâmetros a, b e c.

Para estimar as habilidades foi utilizada a função `fscores` do pacote `mirt` do R e usados dois métodos bayesianos, a Esperança da Posteriori (EAP) e os Valores Plausíveis (VP).

O método EAP é bastante conhecido e é obtido de forma direta, calculando a esperança da distribuição a posteriori de  $\theta$ .

A estimação das habilidades por valores plausíveis consiste em obter, de maneira iterativa, uma amostra da distribuição a posteriori  $f(\theta|\mathbf{U} = \mathbf{u})$  utilizando um algoritmo de rejeição.

No algoritmo de rejeição é gerado um valor  $\theta^*$  inicial da priori  $f(\theta)$  e em seguida é

gerado um vetor de padrão de resposta  $\mathbf{u}^*$  de  $P(\mathbf{U} = \mathbf{u}|\theta^*)$ , onde  $\mathbf{u}$  é um vetor de 0's, quando a resposta é incorreta, e 1's, quando corretas. Esse processo se repete até que  $\theta^*$  produza  $\mathbf{u}^*$  tal que  $\mathbf{u}^* = \mathbf{u}$ . Ao final,  $\theta^*$  será um valor plausível para o indivíduo com padrão de resposta  $\mathbf{u}$ . Para melhor compreensão, em síntese, veja o seguinte esquema.

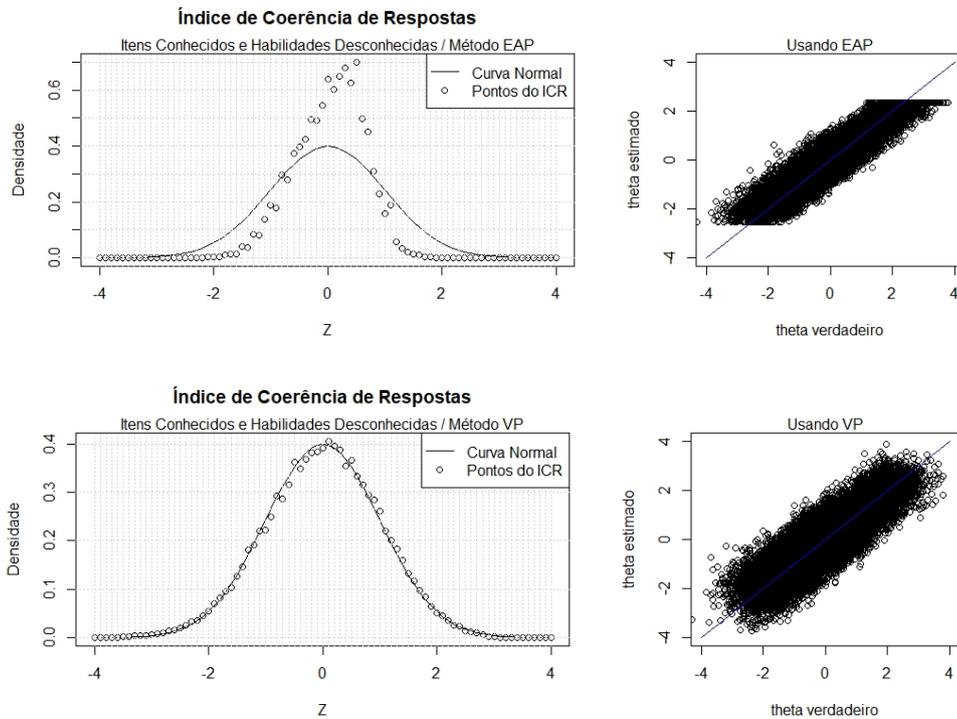
Algoritmo de rejeição para  $f(\theta|\mathbf{U} = \mathbf{u})$

---

- 1: Repita
- 2:     gere  $\theta \sim f(\theta)$
- 3:     gere  $\mathbf{u}^* \sim P(\mathbf{U} = \mathbf{u}|\theta)$
- 4: até  $\mathbf{u}^* = \mathbf{u}$
- 5: Defina  $\theta = \theta^*$

Na Figura 4.3 tem-se os gráficos resultantes ao estimar as habilidades e gerar a distribuição do ICR, e na Tabela 4.3 foram colocados algumas estatísticas obtidas ao usar cada método. Quando é usado o método de valores plausíveis a curva do ICR se aproxima bastante à normal padrão, porém com o método EAP tem-se uma menor dispersão.

Figura 4.3 Gráficos do ICR com parâmetros dos itens conhecidos e habilidades desconhecidas



Para estimação das habilidades o erro é calculado por  $EQM(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n (\hat{\theta} - \theta)^2$ , onde  $\theta$  é a habilidade verdadeira e  $\hat{\theta}$  é a habilidade estimada. Usando EAP se obteve um EQM

## 4.5 Estudo do ICR com parâmetros dos itens e habilidades desconhecidos 31

de  $\theta$  menor do que quando usado VP. Porém, ambos os métodos produziram uma média e desvio padrão próximos de zero e um, respectivamente, para  $\theta$ .

Tabela 4.3 *Tabela com estatísticas do ICR usando EAP / VP com habilidades desconhecidas*

Método de Estimação	Média(ICR)	DP(ICR)	Percentis (5%/95%)	EQM(ICR)	EQM( $\hat{\theta}$ )
Esperança da Posteriori	0,03940	0,60274	-1,01895/0,95145	0,00919	0,15483
Valores Plausíveis	-0,00168	1,01129	-1,69078/1,63377	0,00006	0,31128

Em resumo, ao calcular a esperança da posteriori tem-se um menor erro ao estimar as habilidades, porém a distribuição do ICR não teve um bom ajuste. E ao obter os valores plausíveis para as habilidades foi tido um bom ajuste do ICR com a normal, mas o erro das proficiências foi maior.

## 4.5 Estudo do ICR com parâmetros dos itens e habilidades desconhecidos

Agora, como informação conhecida considera-se apenas a base de dados com as respostas dos indivíduos, é admitido que não se tem conhecimento nem dos parâmetros dos itens nem das habilidades. Embora, esses dados sejam conhecidos, pois foram simulados para gerar a matriz de respostas, eles são assumidos desconhecidos para se fazer o estudo do ICR nessa circunstância específica.

Quando estimados os parâmetros  $a$ ,  $b$  e  $c$ , que são referentes aos itens, obteve-se uma boa estimação (veja Figura 4.4). Para o parâmetro de discriminação tivemos  $EQM(\hat{a}) = 0,00171$ , para o de dificuldade,  $EQM(\hat{b}) = 0,00204$ , e para o de acerto casual,  $EQM(\hat{c}) = 0,00091$ .

No estudo da distribuição do ICR verificou-se que ao "perturbar" os parâmetros dos itens, fazendo pequenas mudanças nos seus valores, a curva do ICR não se modificou tanto, mas há uma grande influência no produto final quando usados métodos diferentes para estimação dos parâmetros das habilidades, semelhante ao que foi visto na seção anterior.

Na Figura 4.5 são apresentados os gráficos obtidos considerando os parâmetros desconhecidos, tanto dos itens como das habilidades. À esquerda o ajuste não ficou como desejado, quando se usa a esperança da posteriori para  $\theta$ . A média ficou próxima de zero,

mas o desvio padrão se encontrou por volta de 0,6. Outras estatísticas foram colocadas resumidas na Tabela 4.4.

Figura 4.4 *Recuperação dos parâmetros dos itens*

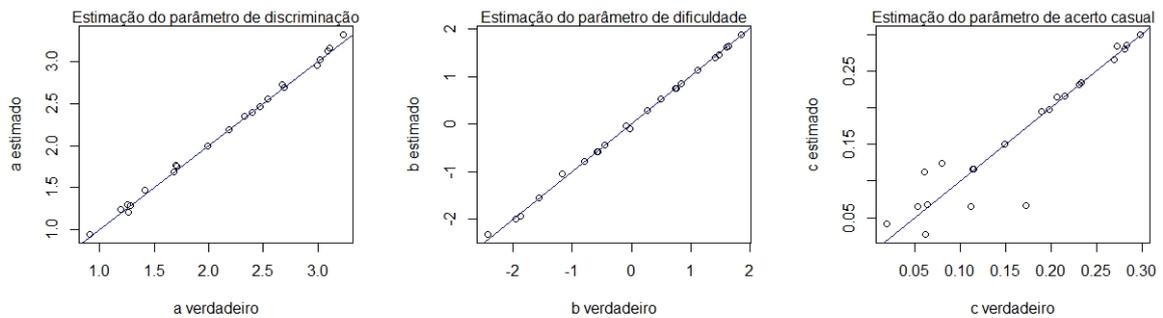
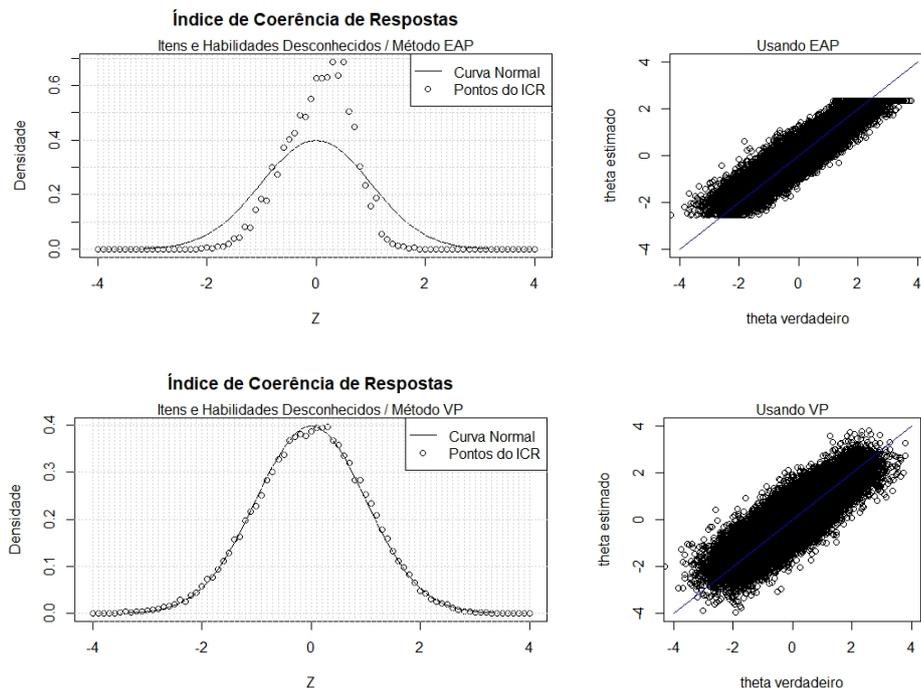


Figura 4.5 *Gráficos do ICR com parâmetros desconhecidos*



Conforme a Tabela 4.4, quando são usados valores plausíveis, tem-se uma média e desvio padrão bem próximos de 0 e 1, respectivamente.

Ao estimar as habilidades  $\theta$ , o erro de estimação, usando o método EAP, é menor do que quando utilizado valores plausíveis.

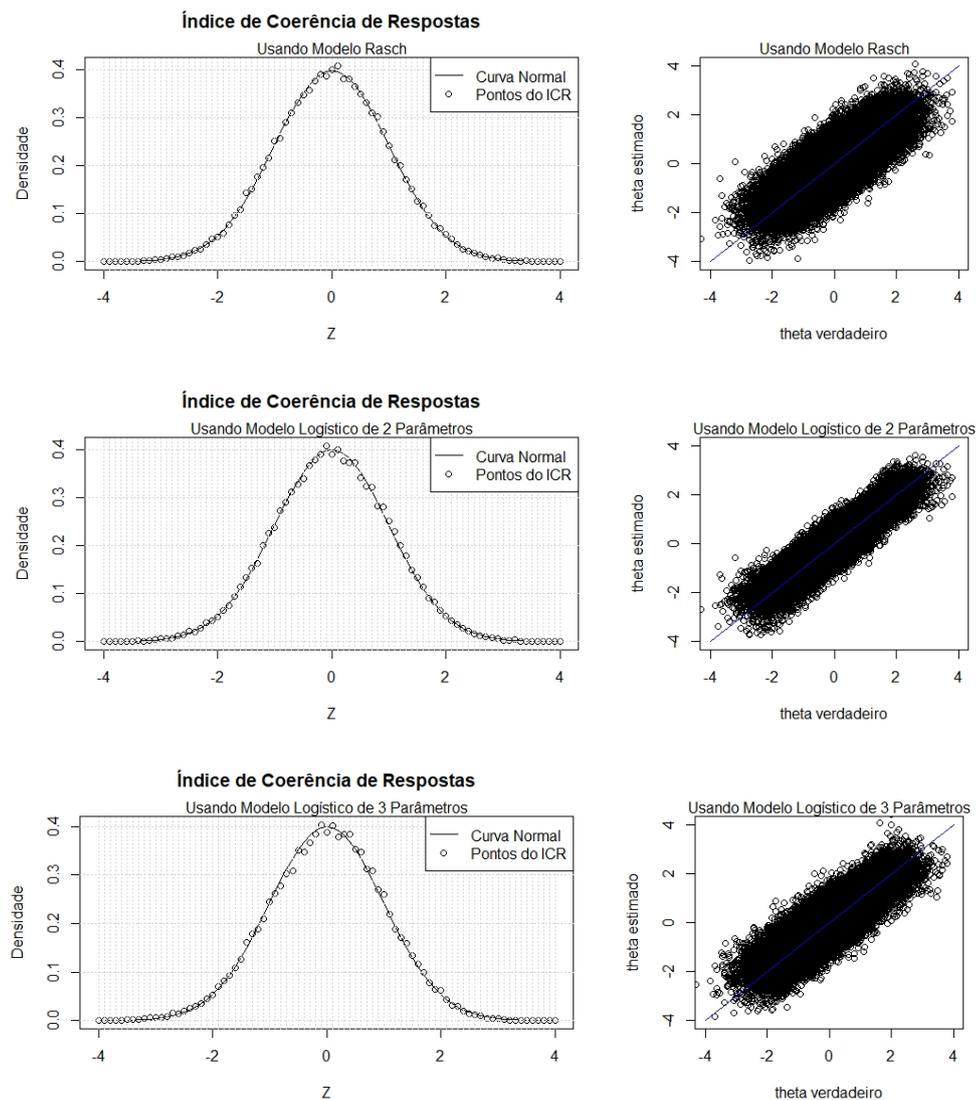
Tabela 4.4 Tabela com estatísticas do ICR usando EAP /VP.

Método de Estimação	Média(ICR)	DP(ICR)	Percentis (5%/95%)	EQM(ICR)	EQM( $\hat{\theta}$ )
Esperança da Posteriori	0,03897	0,60460	-1,02716/0,95250	0,00910	0,15494
Valores Plausíveis	-0,00037	1,00799	-1,68085/1,62474	0,00004	0,30741

## 4.6 Robustez

Inicialmente, são admitidas suposições preliminares específicas para a simulação dos dados, a saber, número de itens e de indivíduos, distribuição das proficiências (ou habili-

Figura 4.6 Robustez em relação ao Modelo de Resposta ao Item



dades) e o modelo matemático de resposta ao item. Se for feita alguma alteração nesses elementos o produto final, que é a distribuição do ICR, continuará estável ou haverá mudanças consideráveis? Executar esse questionamento é verificar a robustez do ICR.

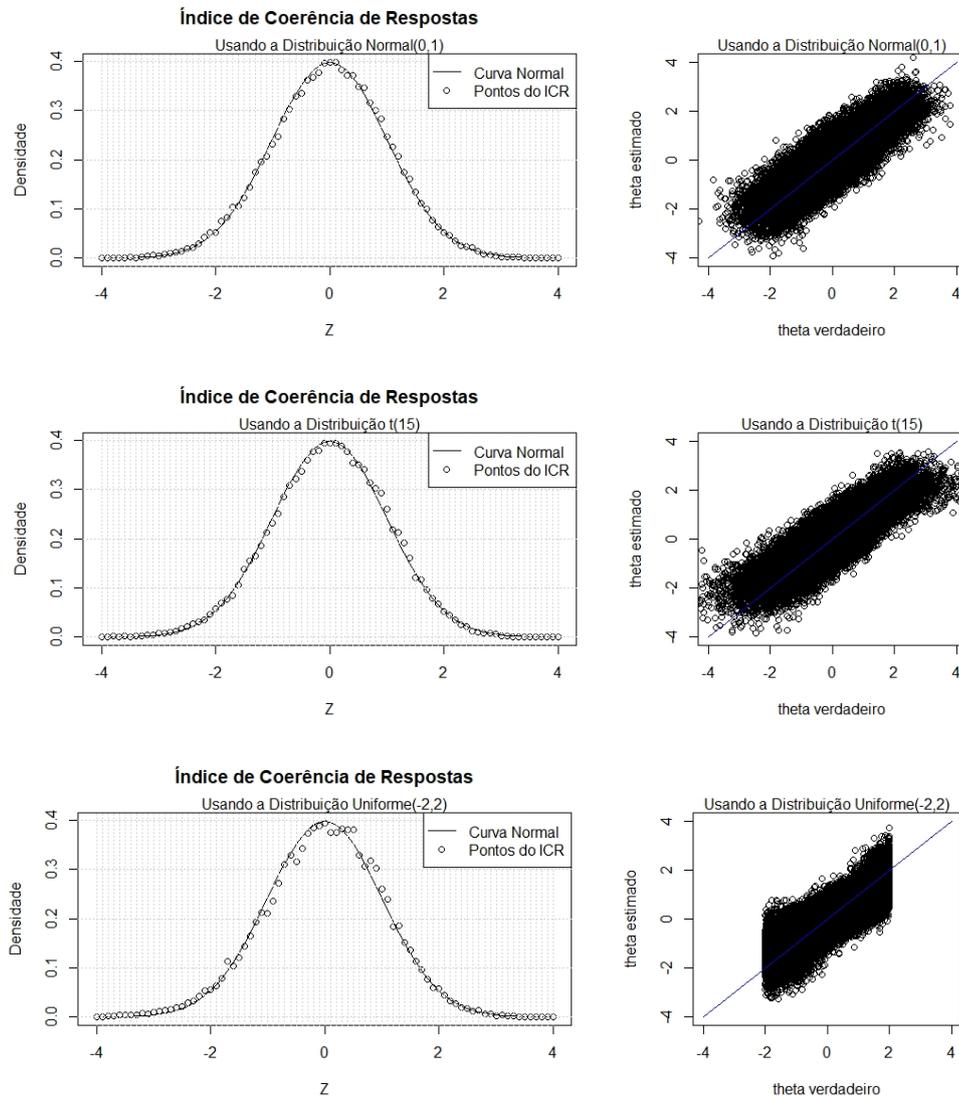
Foram organizadas as Tabelas 4.5 e 4.6, que contêm as estatísticas encontradas ao avaliar a robustez usando valores plausíveis e a esperança a posteriori, respectivamente, para estimar as habilidades. Contendo também em resumo quais os dados utilizados para cada caso de robustez a ser verificada, indicando em negrito, para melhor visualizar, qual elemento foi alterado.

Em TRI, os modelos matemáticos unidimensionais podem conter um ou mais parâmetros referentes ao item. Os mais conhecidos e utilizados são os modelos logísticos de 1, 2 e 3 parâmetros. Primeiramente, é verificado como se comporta a distribuição do ICR com relação a esses três modelos.

Na Figura 4.6 tem-se os gráficos resultantes ao utilizar os distintos modelos de resposta ao item. Todos se comportam de maneira similar na distribuição do ICR, porém ao estimar os parâmetros das proficiências os mesmos produzem resultados distintos. Para os casos em que foram usados os modelos de Rasch (de um parâmetro) e o de três parâmetros se obteve os maiores EQM's de  $\hat{\theta}$ , conforme mostra a Tabela 4.5, o que produz gráficos mais largos comparados ao que usa o modelo de 2 parâmetros.

O parâmetro  $\theta$ , que é o parâmetro das habilidades dos indivíduos, pode assumir teoricamente qualquer valor entre  $-\infty$  e  $\infty$ , comumente assumindo uma escala (0,1), onde 0 representa a média e 1 o desvio padrão das habilidades. Para conferir a robustez do ICR em relação a distribuição atribuída a  $\theta$  tomam-se as distribuições Normal(0,1), t(15) e Uniforme(-2,2), as quais têm escala igual ou próxima a (0,1).

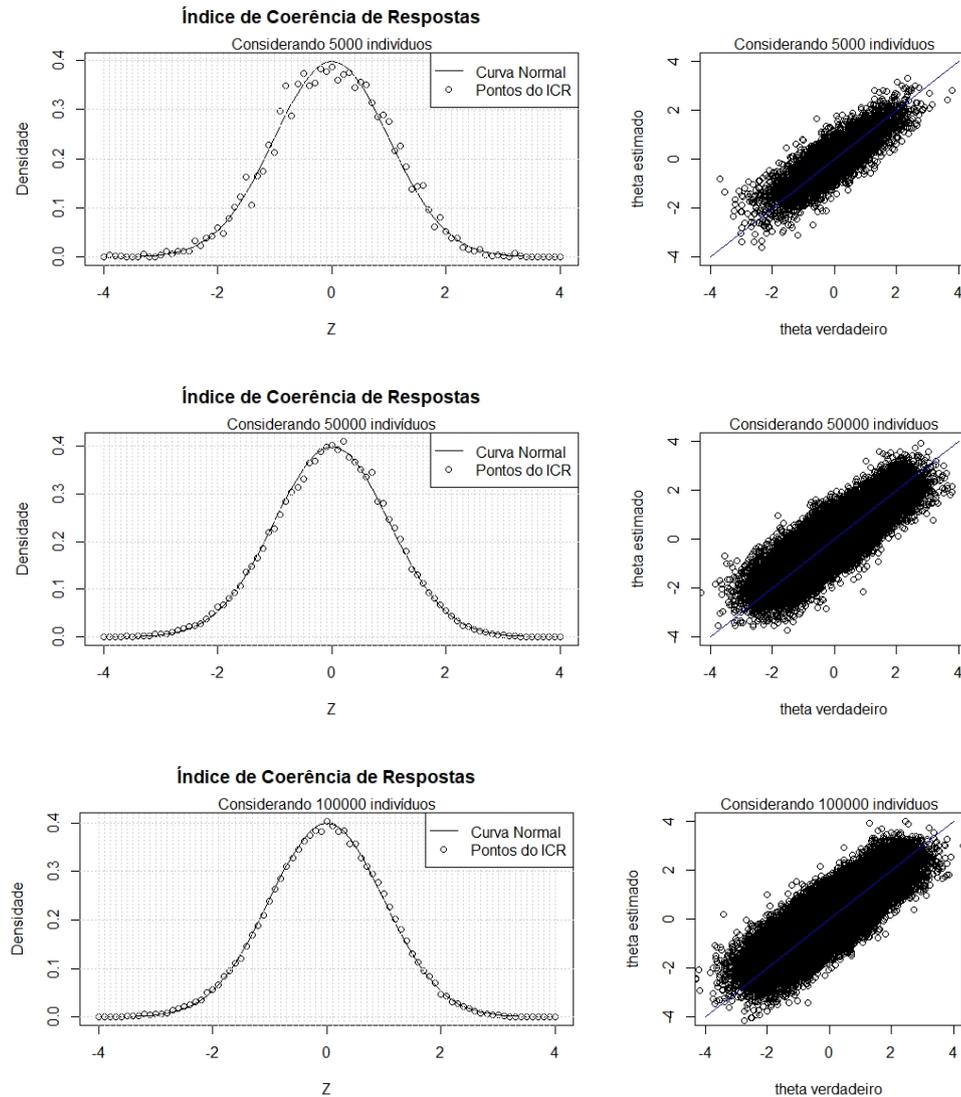
Na Figura 4.7 são apresentados os gráficos de estimação das proficiências e nota-se que ao usar a distribuição Uniforme(-2,2), aparentemente, esta terá um maior erro associado as habilidades decorrente de esta ser a mais distinta dentre as três distribuições usadas. Porém, nos três casos se obteve um  $EQM(\hat{\theta})$  próximo a 0,38. E, nas três situações a distribuição do ICR também permaneceu estável, com um maior EQM no caso da Uniforme.

Figura 4.7 Robustez em relação a distribuição inicial de  $\theta$ 

É analisado agora a robustez do ICR em relação ao número de indivíduos. Escolheu-se as quantidades 5000, 50000 e 100000 respondentes e apresentados graficamente os resultados na Figura 4.8. Para o número menor de indivíduos os pontos do ICR ficaram mais dispersos e com os demais valores os dados se encontraram melhor ajustados. Ao olhar para as estimativas das habilidades não há grandes diferenças nos erros calculados, todos estão próximos de 0,38.

No Brasil, existem o SAEB (Avaliação da Educação Básica) e o ENEM (Exame Nacional do Ensino Médio) que são avaliações de larga escala que utilizam a TRI. As provas do

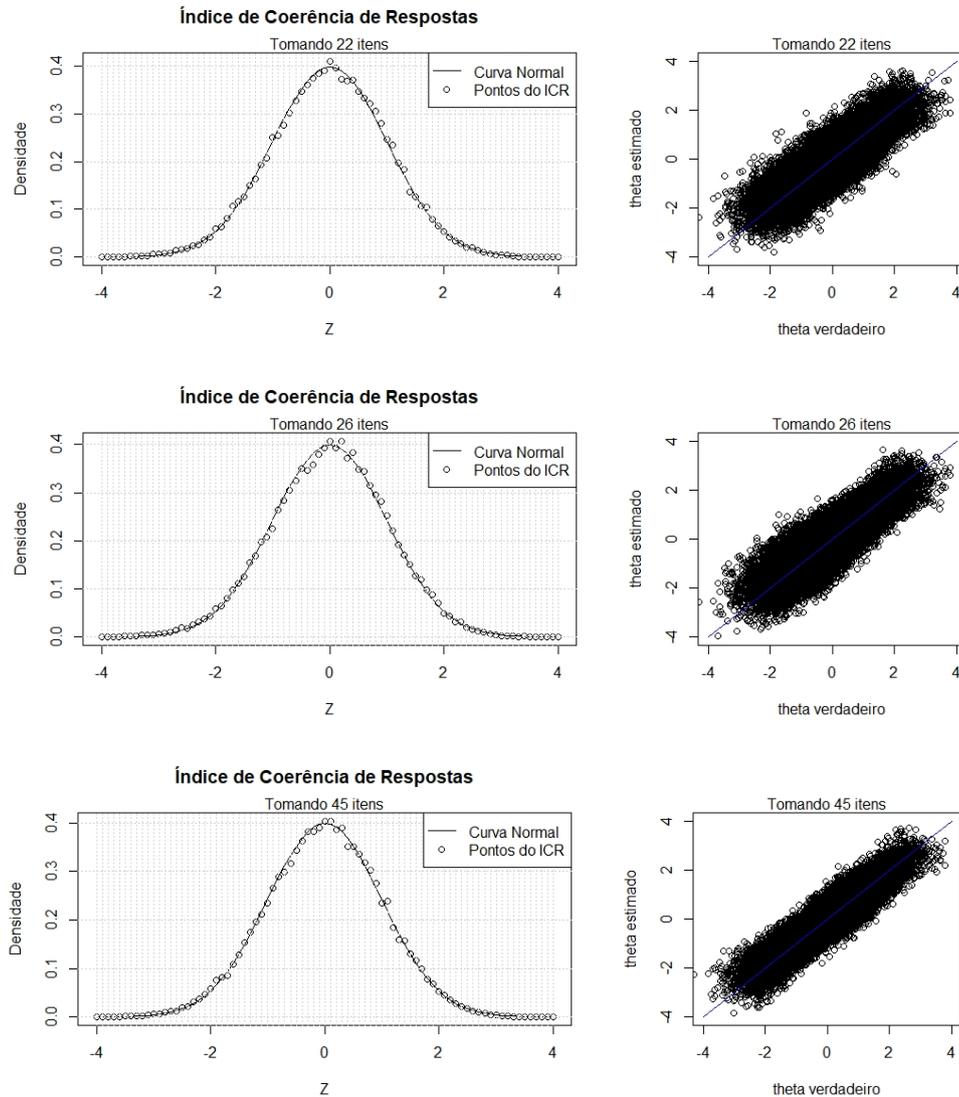
Figura 4.8 Robustez em relação ao número de indivíduos



SAEB aplicadas aos alunos do 5<sup>o</sup> ano do ensino fundamental contêm 22 itens de cada disciplina e as do 9<sup>o</sup> ano, 26. Já as provas no ENEM são compostas por 45 itens em cada área de conhecimento. Baseados nessas quantidades é que se averigua a robustez do ICR em relação ao número de itens.

Ao observar, na Figura 4.9, os gráficos à direita referentes a estimação de  $\theta$  vê-se que quanto mais itens mais fina a espessura do gráfico, ou seja, o erro em relação as habilidades diminui conforme o aumento do número de itens. Isso se confirma ao calcular o *EQM* de cada  $\theta$  estimado mostrado na Tabela 4.5.

Figura 4.9 Robustez em relação ao número de itens



Na mesma Figura 4.9 tem-se também os gráficos referentes à distribuição do ICR e é possível notar que o ajuste visualmente melhorou conforme o aumento de itens, mas estando praticamente o mesmo nas três ocasiões, o que mostra a robustez do ICR em relação ao número de itens.

É verificado também a robustez utilizando a esperança da posteriori para estimar  $\theta$ . Estão apresentadas na Tabela 4.6 as estatísticas obtidas para este caso. Foram omitidos da tabela, mas na maioria das situações a distribuição do ICR teve, aproximadamente, média zero e desvio padrão 0,6. Os EQM(ICR) foram maiores que os obtidos quando

foram usados valores plausíveis, já os  $EQM(\hat{\theta})$  foram menores. E os percentis de 5% e 95% estão em parte cerca de -1 e 1, respectivamente.

Embora a distribuição do ICR não tenha se ajustado a curva normal quando é usado o

Tabela 4.5 *Algumas estatísticas para avaliar a robustez usando VP.*

Itens	Indivíduos	Dist. de $\theta$	Mod. Resp. Item	EQM(ICR)	Percentis (5%/95%)	EQM( $\hat{\theta}$ )
22	50000	Normal(0,1)	<b>Rasch</b>	0,00002	-1,63942/1,64974	0,43163
22	50000	Normal(0,1)	<b>2 parâmetros</b>	0,00002	-1,64282/1,63878	0,19445
22	50000	Normal(0,1)	<b>3 parâmetros</b>	0,00005	-1,67857/1,63579	0,30751
22	50000	<b>Normal(0,1)</b>	3 parâmetros	0,00004	-1,69647/1,63634	0,30885
22	50000	<b>Uniforme(-2,2)</b>	3 parâmetros	0,00013	-1,74302/1,63291	0,33541
22	50000	<b>t(15)</b>	3 parâmetros	0,00005	-1,67878/1,62558	0,31667
22	<b>5000</b>	Normal(0,1)	3 parâmetros	0,00026	-1,64474/1,61552	0,30137
22	<b>50000</b>	Normal(0,1)	3 parâmetros	0,00005	-1,68801/1,63198	0,31034
22	<b>100000</b>	Normal(0,1)	3 parâmetros	0,00002	-1,69311/1,63468	0,30919
<b>22</b>	50000	Normal(0,1)	3 parâmetros	0,00004	-1,68019/1,64567	0,31013
<b>26</b>	50000	Normal(0,1)	3 parâmetros	0,00004	-1,68700/1,64421	0,28430
<b>45</b>	50000	Normal(0,1)	3 parâmetros	0,00003	-1,68219/1,64563	0,16911

Tabela 4.6 *Algumas estatísticas para avaliar a robustez usando EAP.*

Itens	Indivíduos	Dist. de $\theta$	Mod. Resp. Item	EQM(ICR)	Percentis (5%/95%)	EQM( $\hat{\theta}$ )
22	50000	Normal(0,1)	<b>Rasch</b>	0,06675	-0,75529/0,76143	0,21582
22	50000	Normal(0,1)	<b>2 parâmetros</b>	0,05651	-0,74778/0,70431	0,09765
22	50000	Normal(0,1)	<b>3 parâmetros</b>	0,00910	-1,02716/0,95250	0,15494
22	50000	<b>Normal(0,1)</b>	3 parâmetros	0,00910	-1,02716/0,95250	0,15494
22	50000	<b>Uniforme(-2,2)</b>	3 parâmetros	0,01379	-0,88120/0,94094	0,18771
22	50000	<b>t(15)</b>	3 parâmetros	0,01011	-1,02099/0,95491	0,17553
22	<b>5000</b>	Normal(0,1)	3 parâmetros	0,00964	-1,014623/0,95482	0,16440
22	<b>50000</b>	Normal(0,1)	3 parâmetros	0,00910	-1,02716/0,95250	0,15494
22	<b>100000</b>	Normal(0,1)	3 parâmetros	0,00929	-1,02416/0,94200	0,15482
<b>22</b>	50000	Normal(0,1)	3 parâmetros	0,00910	-1,02716/0,95250	0,15494
<b>26</b>	50000	Normal(0,1)	3 parâmetros	0,01163	-1,01187/0,91311	0,14273
<b>45</b>	50000	Normal(0,1)	3 parâmetros	0,01072	-0,94324/0,89568	0,08504

método EAP, ela se comportou de maneira semelhante ao caso com valores plausíveis. O ICR se manteve estável, sem grandes mudanças, em todas as situações realizadas. Tendo uma maior influência quando há um aumento no número de itens que resulta em um menor erro na estimativa de  $\theta$ .

## 4.7 Critério de Decisão

Para ser possível classificar os padrões de respostas dos indivíduos, como aleatório ou não, foi simulado, primeiramente, um conjunto de dados em que todas as pessoas submetidas ao teste o responderam de maneira totalmente aleatória.

Para estes dados simulados, tem-se que para cada questão do teste existe probabilidade  $1/NALT$  de acerto, onde  $NALT$  é o número de alternativas que contém cada questão do teste. Logo, se há um total de  $I$  questões, então a variável escore, que neste caso é chamada de  $S_2$ , terá distribuição binomial de parâmetros  $(I, 1/NALT)$ . E, pelo Teorema Central do Limite, tem-se que a distribuição de  $S_2$  pode ser aproximada pela normal, ou seja, assintoticamente tem-se que

$$S_2 \sim Normal\left(\frac{I}{NALT}, \frac{I(NALT - 1)}{NALT^2}\right) \quad (4.6)$$

Ao tomar o ICR nas duas situações, considerando o padrão de resposta inicial e supondo respostas totalmente aleatórias, tem-se

$$ICR = \frac{S - E(S)}{DP(S)}, \text{ e} \quad (4.7)$$

$$Z = \frac{S_2 - E(S)}{DP(S)} \quad (4.8)$$

Em 4.8 é denominado  $Z$  para diferenciar as duas distribuições.

Como  $S_2$  tende a uma normal, a distribuição de  $Z$  também tenderá a uma normal, pois é função de  $S_2$ , com parâmetros

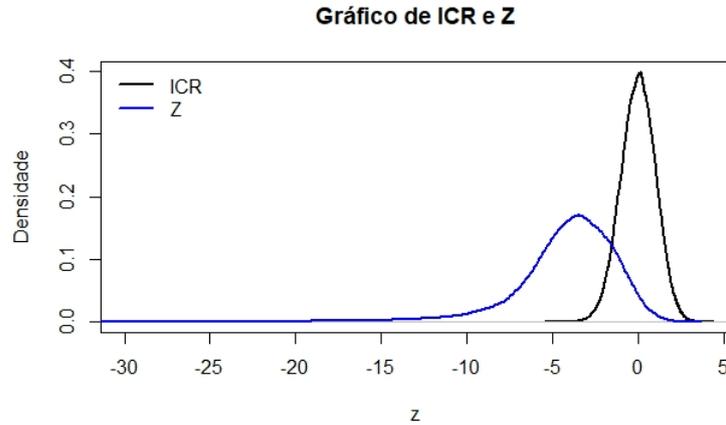
$$\mu = \frac{I}{NALT \cdot DP(S)} - \frac{E(S)}{DP(S)}, \text{ e}$$

$$\sigma^2 = \frac{I(NALT - 1)}{[NALT \cdot DP(S)]^2}$$

e, como já foi visto no início deste capítulo, o ICR tem distribuição assintótica  $Normal(0,1)$ .

A partir disso, é possível montar, por exemplo, o seguinte gráfico

Figura 4.10 Gráfico com as densidades do ICR e de  $Z$



A Figura 4.10 apresenta em um gráfico as densidades do ICR e de  $Z$ , onde as duas distribuições têm um ponto  $z$  em comum e será justamente esse ponto o critério de decisão adotado neste trabalho. Considera-se que indivíduos que obtiverem  $ICR \geq z$  têm padrão de resposta coerente e aqueles com  $ICR < z$ , padrão de resposta totalmente aleatório.

Portanto, é preciso encontrar o valor de  $z$ , tal que  $Z = ICR$ . Logo,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

$$\exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right) = \sigma \exp\left(-\frac{1}{2}z^2\right)$$

$$-\frac{1}{2} \frac{(z^2 - 2z\mu + \mu^2)}{\sigma^2} = \log \sigma - \frac{z^2}{2}$$

$$-z^2 + 2z\mu - \mu^2 = 2\sigma^2 \log \sigma - \sigma^2 z^2$$

$$(\sigma^2 - 1)z^2 + 2\mu z - (\mu^2 + 2\sigma^2 \log \sigma) = 0.$$

Assim,

$$\begin{aligned}
 \Delta &= (2\mu)^2 + 4(\sigma^2 - 1)(\mu^2 + 2\sigma^2 \log \sigma) \\
 &= 4\mu^2 + 4(\sigma^2\mu^2 + 2\sigma^4 \log \sigma - \mu^2 - 2\sigma^2 \log \sigma) \\
 &= 4\mu^2 - 4\mu^2 + 4\sigma^2(\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma) \\
 &= 4\sigma^2(\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma).
 \end{aligned}$$

Então

$$\begin{aligned}
 z &= \frac{-2\mu \pm \sqrt{4\sigma^2(\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma)}}{2(\sigma^2 - 1)} \\
 &= \frac{-2\mu \pm 2\sigma\sqrt{\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma}}{2(\sigma^2 - 1)} \\
 &= \frac{-\mu \pm \sigma\sqrt{\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma}}{\sigma^2 - 1}.
 \end{aligned}$$

Como a curva  $Z$  está mais a esquerda da curva normal padrão e mais dispersa, então  $\mu < 0$  e  $\sigma > 1$ . Além disso, o valor de  $z$  em que  $Z = ICR$  é negativo. Assim,

$$z = \frac{-\mu - \sigma\sqrt{\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma}}{\sigma^2 - 1}. \quad (4.9)$$

Como a distribuição de  $Z$  tem caudas muito pesadas à esquerda, esta foi limitada inferiormente tal que

$$\min(Z) = \text{média}(Z) - 2 * dp(Z)$$

para que assim ela se torne um pouco mais simétrica.

Assim, tem-se uma nova distribuição, a  $Z$  limitada, chamada  $Z'$ . E, a partir de  $Z'$ , é possível estimar os valores de  $\mu$  e  $\sigma^2$ . Logo,

$$\hat{\mu} = \bar{Z}' = \frac{1}{N} \sum_{i=1}^N Z'_i \text{ e}$$

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (Z'_i - \bar{Z}')^2,$$

onde  $N$  é o número de amostras que contém  $Z'$ .

Tomando, como exemplo, o caso em que se tem 22 itens e 4 alternativas, então

$$\hat{\mu} = -3,801917$$

$$\hat{\sigma}^2 = 2,271706^2 = 5,160646$$

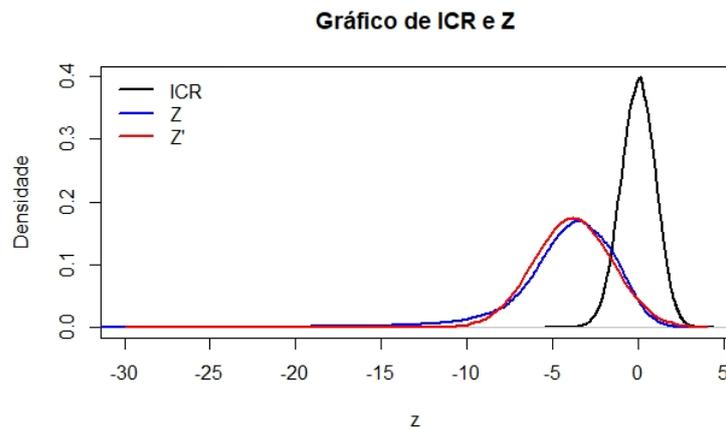
Logo,  $Z' \sim Normal(-3, 80; 5, 16)$ . Assim,

$$z = \frac{-\mu - \sigma\sqrt{\mu^2 + 2\sigma^2 \log \sigma - 2 \log \sigma}}{\sigma^2 - 1}$$

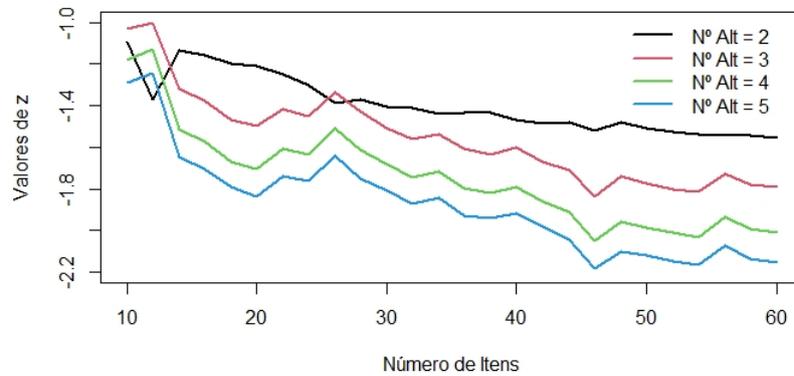
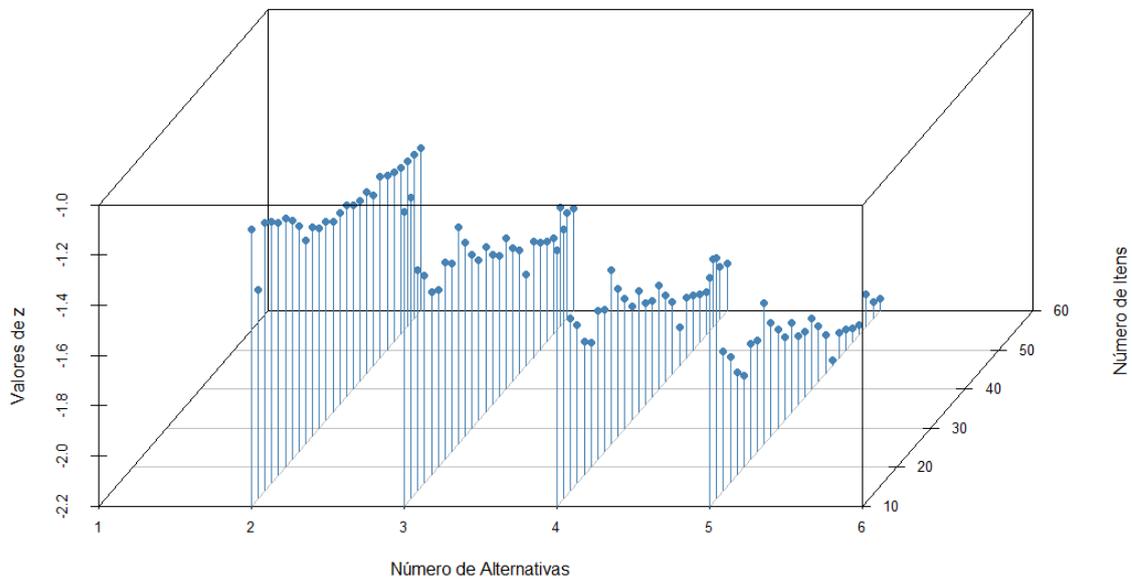
$$z = -1.605068$$

Portanto, para indivíduos com  $ICR \geq -1,605068$  é dito que estes tiveram padrões de resposta coerente e aqueles com  $ICR < -1,605068$  chutaram todas as respostas. Na Figura 4.11 tem-se o gráfico gerado para este exemplo.

Figura 4.11 Gráfico com as densidades de  $ICR$ ,  $Z$  e  $Z'$  considerando 22 itens e 4 alternativas.



A partir da fórmula encontrada para  $z$  na Equação 4.9 são gerados vários outros valores de  $z$ , variando a quantidade de itens, de 10 a 60, e alternativas, de 2 a 5. Assim, são construídos os seguintes gráficos.

Figura 4.12 Gráfico no  $\mathbb{R}^2$  dos valores de  $z$ Figura 4.13 Gráfico no  $\mathbb{R}^3$  dos valores de  $z$ 

Nas Figuras 4.12 e 4.13 a medida que o número de itens aumenta o valor de  $z$  diminui, e assim também acontece quando o número de alternativas aumenta.

Assim, com os dados encontrados são estimados dois modelos de ajuste em que  $z$  depende do número de alternativas e do número de itens. O primeiro, modelo 1, foi um modelo linear, tal que

$$z = \beta_0 + \beta_1 NALT + \beta_2 I + \epsilon,$$

e o segundo, modelo 2, foi um modelo não linear, tal que

$$z = \beta_0 + \beta_1 NALT + \beta_2 \log(I) + \epsilon,$$

onde:

$\beta_0$ ,  $\beta_1$  e  $\beta_2$  são os coeficientes desconhecidos;

$NALT$  é o número de alternativas que contém os itens;

$I$  é o número de itens que contém o teste; e

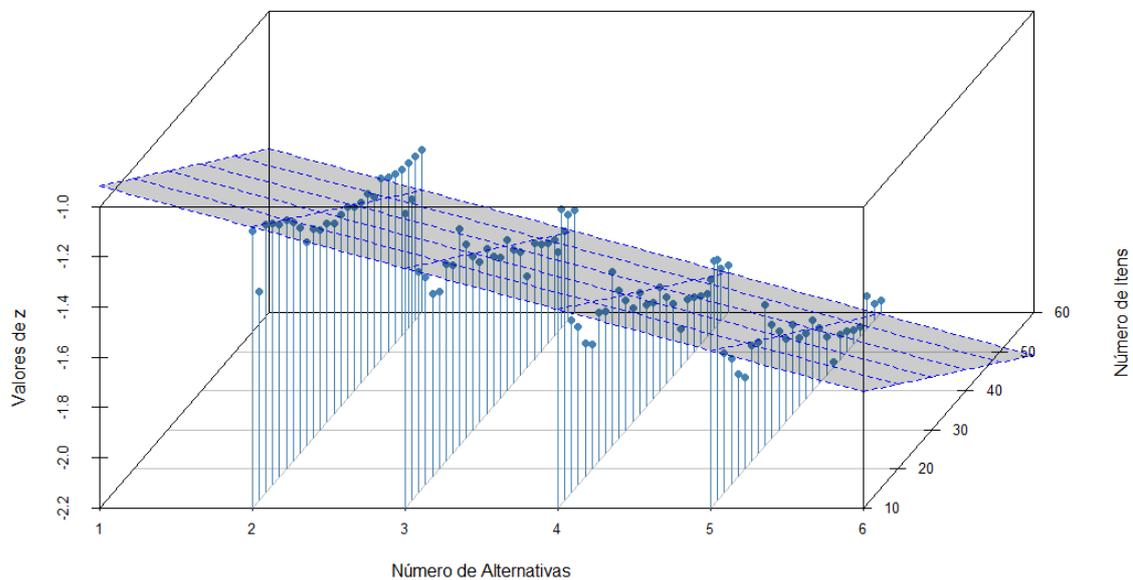
$\epsilon$  é o erro do modelo.

Os modelos estimados foram:

$$\text{modelo 1 : } z = -0,62679 - 0,16424 * NALT - 0,01261 * I \quad (4.10)$$

$$\text{modelo 2 : } z = 0,25780 - 0,16424 * NALT - 0,38522 * \log(I) \quad (4.11)$$

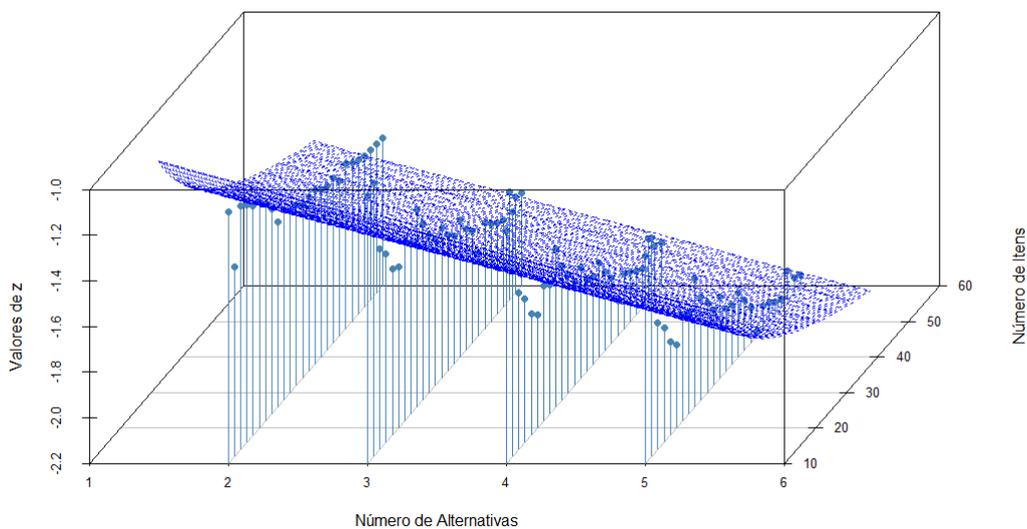
Figura 4.14 *Gráfico do modelo 1 estimado*



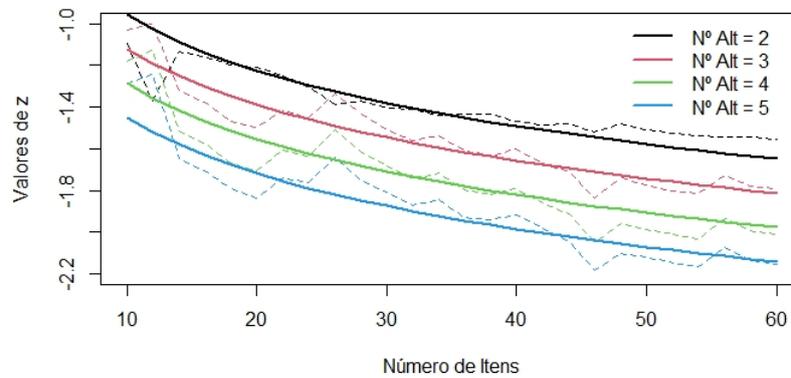
A Figura 4.14 apresenta o gráfico gerado com o ajuste do modelo linear estimado. Em relação ao número de alternativas os dados parecem de fato seguir linearmente conforme o modelo estimado, mas em relação o número de itens o modelo estimado decresce de forma mais rápida dos que os dados.

A Figura 4.15 apresenta o gráfico sob o modelo não-linear estimado. Novamente, em relação ao número de alternativas os dados se comportam linearmente, mas em relação ao número de itens os dados decrescem conforme uma função logarítmica, onde a medida que aumenta a quantidade de itens os valores de  $z$  decresce com menor variação, o que torna visualmente um melhor ajuste em relação ao outro modelo.

Figura 4.15 *Gráfico do modelo 2 estimado*



Para a escolha do melhor modelo foi utilizado o Critério de Informação de Akaike (AIC), que considera o modelo de melhor ajuste aquele com o menor valor do AIC. Para o modelo 1 foi obtido um AIC de  $-177.477$  e para o modelo2,  $-202.705$ . Logo, o modelo que melhor se ajustou aos dados foi o segundo. Na Figura 4.16 é apresentado um outro gráfico, em  $\mathbb{R}^2$ , com as curvas de ajuste do modelo 2, mostrando que o número de itens decresce conforme a função logarítmica estimada.

Figura 4.16 *Gráfico do modelo 2 estimado ( $\mathbb{R}^2$ )*

Portanto, para o critério de decisão foi encontrado que o valor de  $z$  estimado é

$$z = -0,25780 - 0,16424 * NALT - 0,38522 * \log(I), \quad (4.12)$$

em que  $z$  depende apenas do número de itens e do número de alternativas.

---

## Capítulo 5

# Aplicação a dados reais: ENEM-2019

---

### 5.1 Sobre o Exame Nacional do Ensino Médio

O Exame Nacional do Ensino Médio (ENEM) é um exame individual aplicado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), autarquia federal vinculada ao Ministério da Educação (MEC). Criado em 1998, era composto por 63 questões e uma redação, com o objetivo fundamental de avaliar o desempenho dos alunos concluintes e egressos do ensino médio (BRASIL, 2002).

Pouco tempo depois, a nota do ENEM passou a ser utilizada como critério de concessão de bolsas de estudos integrais ou parciais em universidades particulares pelo Programa Universidade para Todos (ProUni). Com isso, cresceu o interesse dos alunos em participar da prova com a finalidade de tentar conseguir uma das bolsas e conseqüentemente uma vaga na universidade.

Desde 2009, o ENEM passou a ser utilizado como meio de acesso ao ensino público superior, através do Sistema de Seleção Unificada (SISU). No mesmo período, a prova teve um aumento do número de questões, passando de 63 para 180, divididas em quatro grandes áreas (Linguagens e Códigos, Ciências Humanas, Ciências da Natureza e Matemática), contendo 45 questões cada, e continuando com uma redação.

Com o passar dos anos, o ENEM acabou se tornando o maior exame de vestibular do Brasil (BRASIL, 2019). Assim, se tornou necessário um aperfeiçoamento tanto da prova quanto da aplicação da mesma. As medidas de segurança estão ficando mais rígidas e propostas de tornar as notas mais verdadeiras e confiáveis têm surgido, como os trabalhos de Rodrigues (2016), Moraes (2019), Souza (2019) e Meza (2020) que são focados na detecção de fraudes.

Assim, este trabalho veio também com uma proposta para ajudar na qualidade das estimativas. Como já citado, foi proposto um estimador de coerência, o ICR, que visa identificar indivíduos com padrão de resposta que possa ter indícios de aleatoriedade em

todas as respostas. Para aplicação desse estimador escolheu-se os dados do ENEM 2019, em particular, os dados dos alunos da cidade de Fortaleza, no estado do Ceará. Esses dados, assim também como de outras edições do ENEM, podem ser consultados no site do INEP, na página de microdados (BRASIL, 2021).

No ENEM de 2019 se inscreveram para a prova 5.095.388 candidatos, tendo a maior taxa de participação da história, com 3.935.237 pessoas presentes nos dois dias de aplicação da mesma (BRASIL, 2020). No município de Fortaleza, cerca de 95 mil candidatos se inscreveram para o ENEM, porém apenas 71094 compareceram efetivamente.

Para aplicação do ICR foram desconsideradas as provas adaptadas e de reaplicação, restando assim 71016 candidatos.

Nas seções seguintes são apresentadas as estimativas dos parâmetros dos itens encontradas para a prova de Ciências Humanas e suas Tecnologias e também os resultados da aplicação do *ICR*.

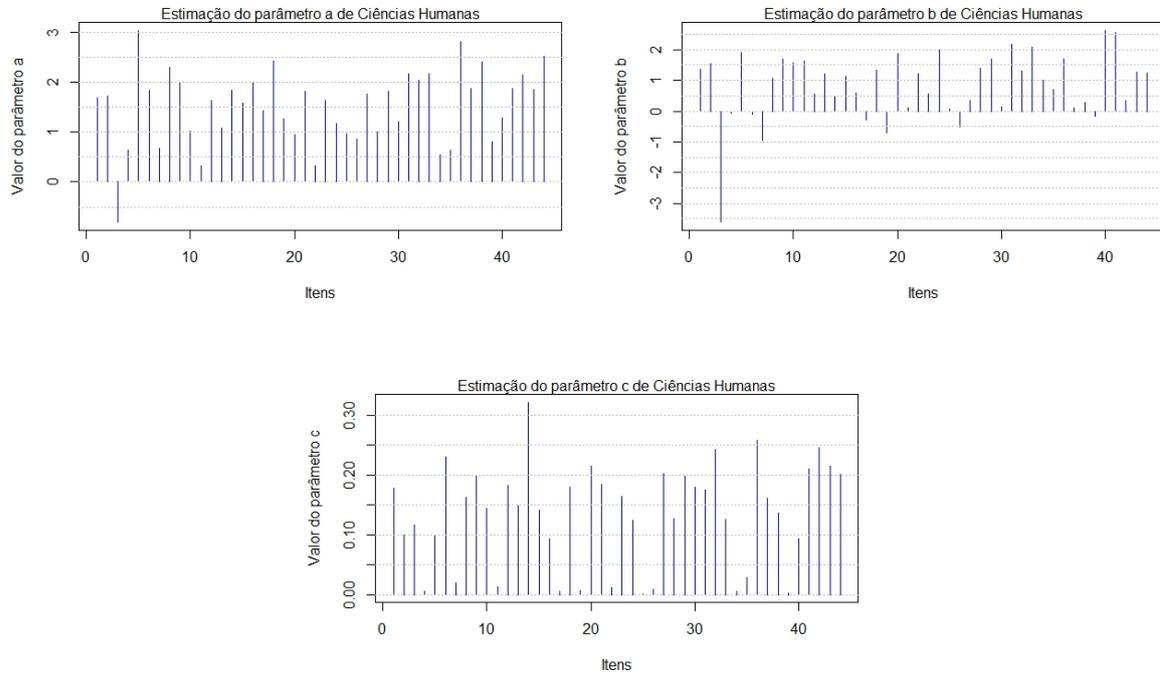
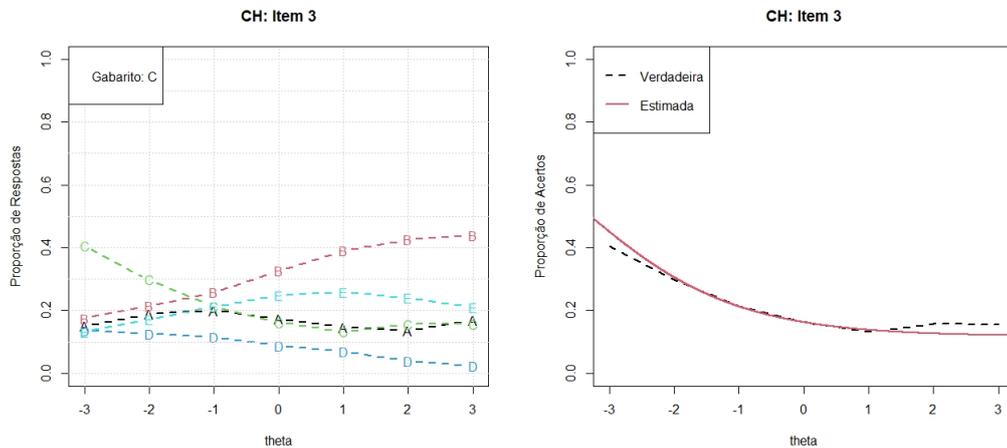
## 5.2 Estimação dos parâmetros dos itens e habilidades

Nesta seção são apresentadas as estimativas obtidas na área de Ciências Humanas e suas Tecnologias do ENEM 2019 dos candidatos de Fortaleza - CE.

De início, temos na Figura 5.1 as estimativas dos parâmetros dos itens. O item 45 foi anulado, então esse item foi retirado da base, restando, assim, apenas 44 itens.

No gráfico do parâmetro de discriminação (a) o item 3 foi estimado com valor negativo, o que diz que a CCI (Curva Característica do Item) desse item é decrescente, ou seja, candidatos com alta proficiência têm baixa probabilidade de acerto, o que não deveria acontecer. Este item talvez teve algum problema quando foi formulado, seja no enunciado, sejam nas alternativas que podem ter sido mal elaboradas, por exemplo.

Na Figura 5.2 são apresentados dois gráficos contendo a proporção de acertos no item 3 conforme a proficiência dos candidatos, que é dada pelo valor do parâmetro  $\theta$ . A proporção dos candidatos com proficiência maior foi menor, o que concorda com a estimativa encontrada. O gabarito do item 3 é a alternativa C, mas a alternativa B foi a mais assinalada pelos candidatos à medida que  $\theta$  aumenta, mostrando, assim, que por algum motivo a alternativa B pareceu ser a correta. No geral, as pessoas descartaram a alternativa D, mas não conseguiram escolher entre as outras alternativas. As curvas se mantêm parecidas, com proporções semelhantes e pouca acentuação, não ultrapassando a metade.

Figura 5.1 *Estimação dos parâmetros dos itens de Ciências Humanas*Figura 5.2 *Proporção de acertos no item 3 de Ciências Humanas*

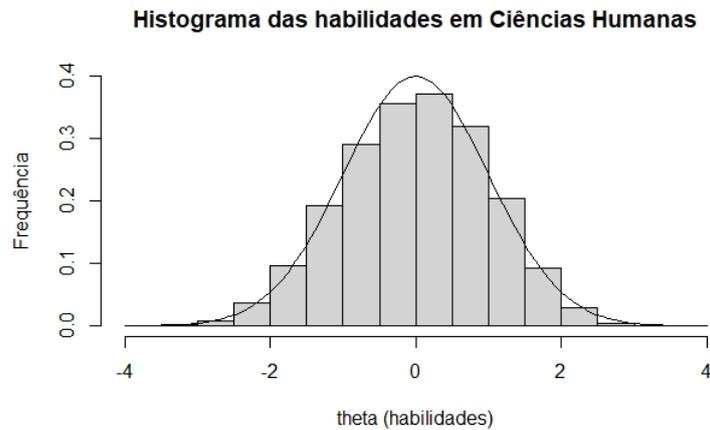
Apesar de o valor do parâmetro de discriminação ser negativo no item 3 ele concorda com a realidade, conforme é apresentado no gráfico à direita na Figura 5.2. A curva estimada ficou bem ajustada à curva verdadeira.

Em todos os itens, o parâmetro de dificuldade ( $b$ ) se manteve dentro do comum esperado, com estimativas dentro do intervalo  $(-5,5)$ . Assim também ocorre com o parâmetro de acerto casual ( $c$ ), com valores entre  $(0, 0.35)$ .

Como a CCI do item 3 resultou incomum, esse item é desconsiderado e a matriz de dados ficou apenas com os 43 itens restantes.

Na Figura 5.3 é apresentado o histograma das habilidades dos candidatos, que tem média igual a  $-0.00081$  e desvio padrão igual a  $0.99887$ . A distribuição das habilidades é praticamente simétrica, com um coeficiente de assimetria igual a  $-0.1112445$ , e platicúrtica com um coeficiente de curtose igual a  $-0.3067976$ . Lembrando que os valores de  $\theta$  foram estimados pelo método de valores plausíveis.

Figura 5.3 *Gráfico das habilidades estimadas em Ciências Humanas*



## 5.3 Estatísticas do ICR

Nesta seção o ICR é aplicado aos dados dos candidatos que fizeram a prova de Ciências Humanas do ENEM-2019 no município de Fortaleza - CE.

A Figura 5.4 mostra a curva do ICR que resultou bem próxima da curva normal com  $EQM(ICR) = 0,00004$ , média  $-0,00095$ , desvio padrão  $1.02460$  e percentis de 5% e 95% iguais a  $-1,70361$  e  $1,64739$ , respectivamente.

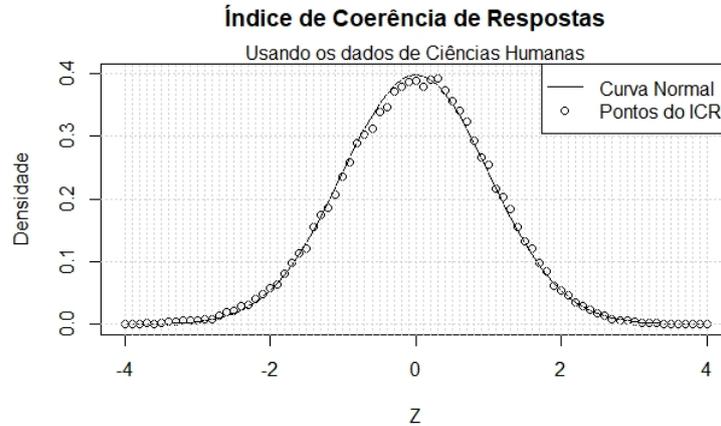
Cada item das provas do ENEM contém 5 alternativas e como restaram apenas 43 itens, tem-se, para o critério de decisão, um valor de  $z$  igual a

$$z = -0,25780 - 0,16424.5 - 0,38522. \log(43)$$

$$z = -2,52789$$

onde  $z$  foi calculado a partir da fórmula dada em 4.12.

Figura 5.4 *Curva do ICR usando dados de Ciências Humanas do ENEM - 2019, Fortaleza-CE*



Houve 648 indivíduos dos 71016 cujo  $ICR \leq -2$ , 52789. Assim, seus padrões de resposta são ditos ter aleatoriedade em todos os itens e, conseqüentemente, foram desconsiderados da análise. Portanto, o conjunto de dados restantes possui, estatisticamente, apenas padrões de respostas coerentes.

Foram estimados novamente os parâmetros dos itens apenas com as respostas coerentes. Nas Figuras 5.5 e 5.6 são apresentados os gráficos de comparação entre as estimativas encontradas com os dados completos, que contém padrões de respostas aleatórios, e os dados sem aleatoriedade. No geral, as estimativas tiveram poucas diferenças, sendo o item 22 com a maior diferença nas estimativas dos parâmetros b e c.

Figura 5.5 *Gráficos para comparação das estimativas considerando dados com e sem aleatoriedade*

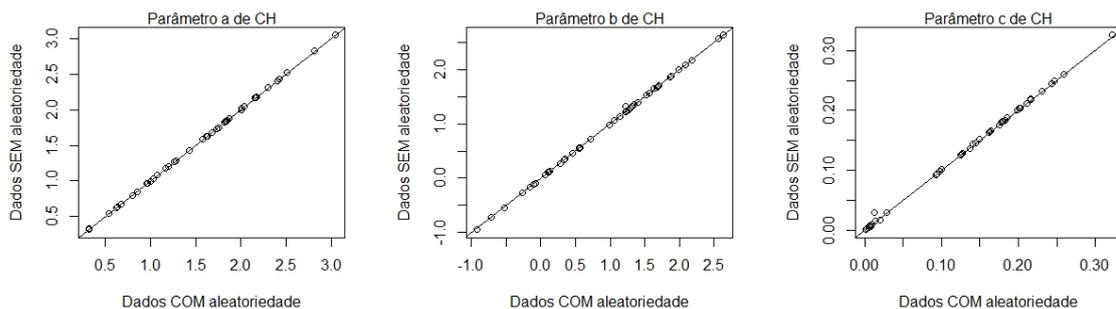
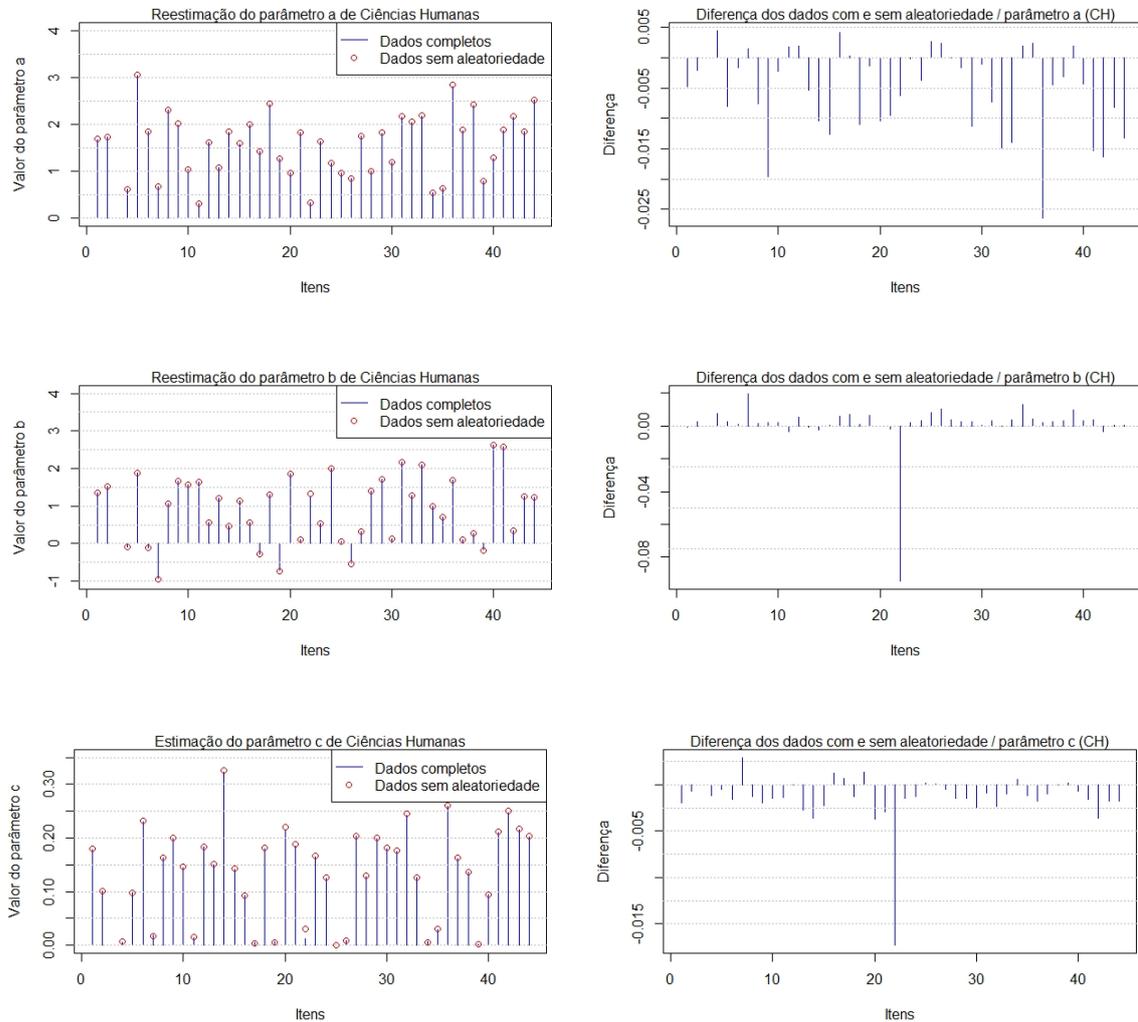


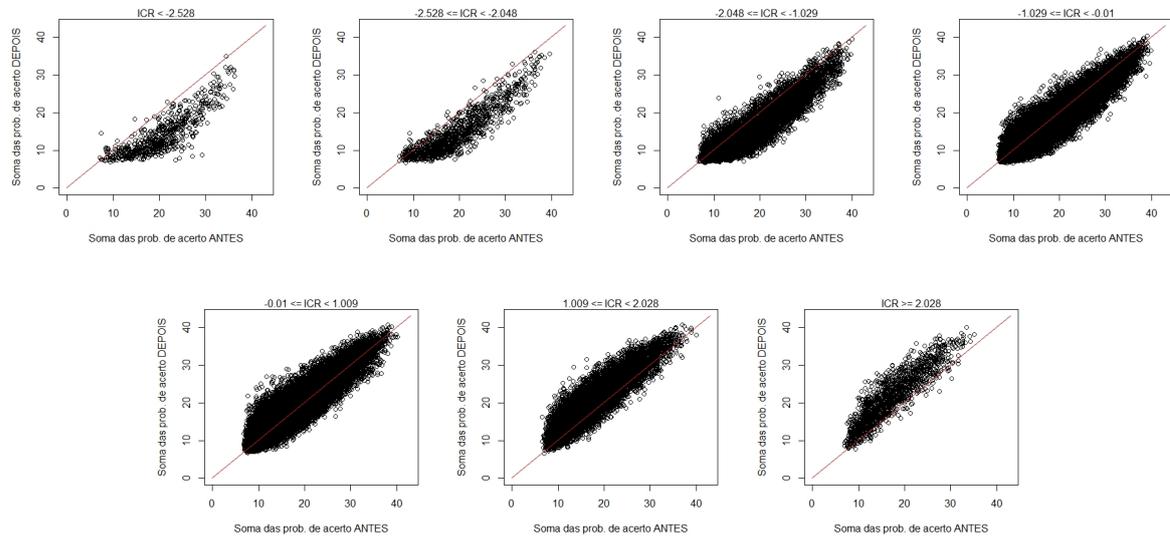
Figura 5.6 Gráficos para comparação das estimativas por item, considerando dados com e sem aleatoriedade



Embora as estimativas encontradas, após serem desconsiderados os padrões totalmente aleatórios, não tenham grandes diferenças comparadas aos dados completos, elas influenciaram no resultado da soma das probabilidades de acerto de cada candidato. Na Figura 5.7 os candidatos com  $ICR < -2,528$ , que são os que chutaram todas as respostas, tiveram soma de probabilidades de acerto menor após a re-estimação, ou seja, no geral, em cada item esses candidatos tiveram uma menor probabilidade de acerto. E, a medida que o valor do  $ICR$  aumenta as somas também são maiores.

Assim, todos os candidatos sofreram algum ajuste nas suas estimativas, principalmente aqueles com  $ICR$  mais distante da média, tornando-as mais cabíveis.

Figura 5.7 *Comparação da soma das probabilidades de acerto obtidas antes e depois da exclusão de dados aleatórios para diferentes valores do ICR.*



---

## Capítulo 6

# Conclusões e Considerações Gerais

---

Neste trabalho foram apresentadas, primeiramente, as distribuições Binomial, Poisson-Binomial e Binomial Composta e suas propriedades e, ainda, ligeiramente, a Teoria da Resposta ao Item, que são de grande importância para a construção do ICR.

O ICR - Índice de Coerência de Resposta - vem como uma proposta de identificar indivíduos que marcaram no "chute" todos os itens de um teste ou de um questionário, que envolvem itens dicotômicos, e, a partir disso, tentar melhorar a qualidade das estimativas buscando um resultado mais coerente e verdadeiro.

Foi proposto também um critério de decisão, onde, dados o número de itens e de alternativas, é possível obter um corte no qual são categorizados em indivíduos que provavelmente marcaram a prova coerente e aqueles que responderam ao acaso.

Por fim, o ICR é aplicado em dados reais, mais precisamente na prova de Ciências Humanas do ENEM-2019 dos alunos de Fortaleza-CE, onde são estimados os parâmetros dos itens e os mesmos são re-estimados sem a presença das respostas com aleatoriedade. Tal aplicação teve resultados positivos, na qual as probabilidades de acerto para cada item foi reajustada fazendo com que alunos que marcaram todos os itens aleatoriamente tivessem uma pequena queda nas probabilidades de acerto de cada item e, a medida que o valor do ICR do indivíduo aumenta, a probabilidade de acertar o item também cresceu.

## 6.1 Trabalhos Futuros

Recomenda-se para trabalhos futuros:

- Aplicar o ICR às demais áreas do ENEM-2019;
- Verificar a robustez do ICR utilizando distribuições assimétricas para as habilidades, já que neste trabalho foi testada supondo apenas distribuições simétricas;

- Gerar mais valores plausíveis e construir, por exemplo, a média para estimar os valores das proficiências de cada indivíduo;
- Construir um ICR que pudesse ser usado em itens não dicotômicos;
- Construir um ICR que identificasse indivíduos que tiveram padrão de respostas parcialmente aleatório;
- Trabalhar com dados que tiveram respostas deixadas em branco;
- Avaliar o porquê de ocorrerem os "chutes". Por exemplo, pode ser que indivíduo esteja passando por problemas socioemocionais ou a ordem em que os itens estão dispostos influencia nas respostas, caso a prova seja muito extensa; e etc.

---

# Referências

---

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C . **Teoria da Resposta ao Item: Conceitos e Aplicações**. São Paulo: ABE, 2000. 154 p.

BARTON, Mark A.; LORD, Frederic M. **An Upper Asymptote for the Three-Parameter Logistic Item-Response Model**. New Jersey: Educational Testing Service, jul. 1981.

BLAIOTTA, Jimena; DELIEUTRAZ, Pablo. **Teorema central del limite**. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, 2004.

BRASIL. **Exame nacional do ensino médio: Documento básico**. Brasília, 2002. Disponível em: <<http://www.dominiopublico.gov.br/download/texto/me000115.pdf>>. Acesso em junho de 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Enem**. Brasília: Inep, 2021. Disponível em:<<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>>. Acesso em junho de 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Histórico**. Brasília: Inep, 2019. Disponível em:<<http://portal.inep.gov.br/enem/historico>>. Acesso em junho de 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Resultados Enem 2019**. Brasília: Inep, 2020. Disponível em:< [http://portal.mec.gov.br/images/17.01.2020\\_Apresentao\\_Enem\\_resultados.pdf](http://portal.mec.gov.br/images/17.01.2020_Apresentao_Enem_resultados.pdf)>. Acesso em junho de 2021.

CHALMERS, R. Philip. mirt: A Multidimensional Item Response Theory Package for the R Environment. **Journal of Statistical Software**, v. 48, mai. 2012.

CIZEK, Gregory J.; WOLLACK, James A. **Handbook of quantitative methods for detecting cheating on tests**. New York: Routledge, 2017.

---

GONZÁLEZ, J.; WIBERG, M.; VON DAVIER, A. A. (2016). A Note on the Poisson's Binomial Distribution in Item Response Theory. **Applied Psychological Measurement**, v. 40(4), p. 302–310, 2016.

HONG, Yili; TECH, Virginia. On computing the distribution function for the Poisson binomial distribution. **Computational Statistics & Data Analysis**, v. 59, p. 41-51, mar. 2013.

LORD, Frederic M.; NOVICK, Melvin R. **Statistical Theories of Mental Test Scores**. Massachusetts: Addison-Wesley, 1968. 568 p.

LORD, Frederic M.; WINGERSKY, Marilyn S. Comparison of IRT true-score and equi-percentile observed-score "equatings". **Applied Psychological Measurement**, v. 8, n.4, p. 453-461, 1984.

KOSE, Ibrahim A.; DEMIRTASLI, Nukhet C. Demirtasli. Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. **Procedia - Social and Behavioral Sciences**, v. 46, p. 135-140, 2012.

MEZA, Robinson O. **Seleção quantílica no pacote testfraud para detecção de fraudes em testes**. Dissertação (Mestrado em Estatística) - Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2020.

MORAES, Alice N. **O estado da arte dos métodos estatísticos para detecção de fraudes em testes e aplicações**. Dissertação (Mestrado em Estatística) - Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2019.

MUÑIZ, José. **Introducción a la Psicometría: Teoría Clásica y TRI**. Madrid: Pirámide, 2018. 376 p.

RAMPAZZO, Sandra R. **Instrumentos de avaliação: reflexões e possibilidades de uso no processo de ensino e aprendizagem**. Londrina, 2011.

RECKASE, Mark D. **Statistics for Social and Behavioral Sciences: Multidimensional item response theory**. New York: Springer, 2009. 355p.

RODRIGUES, José I. L. **Métodos estatísticos para detecção de fraudes em testes**. Dissertação (Mestrado em Estatística) - Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2016.

SOUZA, Miguel M. **Implementação e otimização do pacote testfraud para detecção de fraudes em testes**. Dissertação (Mestrado em Estatística) - Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2019.

---

## Apêndice A

# Algoritmo para organização da base de dados do ENEM-2019, Fortaleza-CE

---

```
#####  
# Carregando pacotes  
if(!require(data.table)) install.packages("data.table"); library(data.table)  
if(!require(stringr)) install.packages("stringr"); library(stringr)  
  
#####  
# Informando o local do arquivo  
setwd("_")  
  
#####  
# Carregando os dados. Disponível em:  
# https://www.gov.br/inep/pt-br/acao-a-informacao/dados-abertos/microdados/enem  
  
ENEM_2019 <- data.table::fread(input='MICRODADOS_ENEM_2019.csv',  
integer64='character',  
skip=0, #Ler do inicio  
nrow=-1, #Ler todos os registros  
na.strings = "",  
showProgress = TRUE)  
  
#####  
# Selecionando os candidatos de Fortaleza-Cear  
ENEM19_FORT <- ENEM_2019[NO_MUNICIPIO_PROVA == "Fortaleza"]  
  
#####  
# Selecionando apenas os candidatos presentes em todas as provas  
ENEM19_FORT_P <- subset(ENEM19_FORT, TP_PRESENCA_MT == 1 &  
TP_PRESENCA_CN == 1 & TP_PRESENCA_CH == 1 &  
TP_PRESENCA_LC == 1 &  
NU_NOTA_MT > 0 & NU_NOTA_CN > 0 &  
NU_NOTA_CH > 0 & NU_NOTA_LC > 0)  
  
#####  
# Retirando os candidatos em que as provas sao adaptadas e de reaplicacao  
ENEM19_FORT_Q <- subset(ENEM19_FORT_P, CO_PROVA_CN >= 503 & CO_PROVA_CN <= 506  
& CO_PROVA_CH >= 507 & CO_PROVA_CH <= 510  
& CO_PROVA_LC >= 511 & CO_PROVA_LC <= 514
```

---

```
& CO_PROVA_MT >= 515 & CO_PROVA_MT <= 518)
```

```
#####  
# Banco Ciencias da Natureza (CN)
```

```
matriz_resp = matrix(0, nrow(ENEM19_FORT_Q), 45)  
matriz_gab = matrix(0, nrow(ENEM19_FORT_Q), 45)  
lin = nrow(matriz_resp)  
col = ncol(matriz_resp)  
matriz_resp_categ = matrix(0, lin, col)
```

```
for (i in 1:nrow(ENEM19_FORT_Q)) {  
  matriz_resp[i,] = str_split(ENEM19_FORT_Q$TX_RESPOSTAS_CN[i], "", simplify = TRUE)  
  matriz_gab[i,] = str_split(ENEM19_FORT_Q$TX_GABARITO_CN[i], "", simplify = TRUE)  
}
```

```
for (i in 1:lin) {  
  for (j in 1:col) {  
    if (matriz_resp[i,j] == matriz_gab[i,j]) {  
      matriz_resp_categ[i,j] = 1  
    }  
  }  
}
```

```
CN_NOM = matriz_resp ; CN_NOM = as.data.frame(CN_NOM)  
CN_DIC = matriz_resp_categ; CN_DIC = as.data.frame(CN_DIC)
```

```
#####  
# Banco Ciencias Humanas (CH)
```

```
matriz_resp = matrix(0, nrow(ENEM19_FORT_Q), 45)  
matriz_gab = matrix(0, nrow(ENEM19_FORT_Q), 45)  
lin = nrow(matriz_resp)  
col = ncol(matriz_resp)  
matriz_resp_categ = matrix(0, lin, col)
```

```
for (i in 1:nrow(ENEM19_FORT_Q)) {  
  matriz_resp[i,] = str_split(ENEM19_FORT_Q$TX_RESPOSTAS_CH[i], "", simplify = TRUE)  
  matriz_gab[i,] = str_split(ENEM19_FORT_Q$TX_GABARITO_CH[i], "", simplify = TRUE)  
}
```

```
for (i in 1:lin) {  
  for (j in 1:col) {  
    if (matriz_resp[i,j] == matriz_gab[i,j]) {  
      matriz_resp_categ[i,j] = 1  
    }  
  }  
}
```

```
CH_NOM = matriz_resp ; CH_NOM = as.data.frame(CH_NOM)  
CH_DIC = matriz_resp_categ; CH_DIC = as.data.frame(CH_DIC)
```

```
#####  
# Banco Matematica (MT)
```

```
matriz_resp = matrix(0, nrow(ENEM19_FORT_Q), 45)
```

---

```

matriz_gab = matrix(0, nrow(ENEM19_FORT_Q), 45)
lin = nrow(matriz_resp)
col = ncol(matriz_resp)
matriz_resp_categ = matrix(0, lin, col)

for (i in 1:nrow(ENEM19_FORT_Q)){
  matriz_resp[i,] = str_split(ENEM19_FORT_Q$TX_RESPOSTAS_MT[i], " ", simplify = TRUE)
  matriz_gab[i,] = str_split(ENEM19_FORT_Q$TX_GABARITO_MT[i], " ", simplify = TRUE)
}

for (i in 1:lin) {
  for (j in 1:col) {
    if (matriz_resp[i,j] == matriz_gab[i,j]) {
      matriz_resp_categ[i,j] = 1
    }
  }
}

MT_NOM = matriz_resp ; MT_NOM = as.data.frame(MT_NOM)
MT_DIC = matriz_resp_categ; MT_DIC = as.data.frame(MT_DIC)

#####
# Banco Linguagens e C digos (LC)

matriz_resp = matrix(0, nrow(ENEM19_FORT_Q), 50)
matriz_gab = matrix(0, nrow(ENEM19_FORT_Q), 50)

for (i in 1:nrow(ENEM19_FORT_Q)){
  matriz_resp[i,] = str_split(ENEM19_FORT_Q$TX_RESPOSTAS_LC[i], " ", simplify = TRUE)
  matriz_gab[i,] = str_split(ENEM19_FORT_Q$TX_GABARITO_LC[i], " ", simplify = TRUE)
}

matriz_resp = matriz_resp[,11:50]
matriz_gab = matriz_gab[,11:50]
lin = nrow(matriz_resp)
col = ncol(matriz_resp)
matriz_resp_categ = matrix(0, lin, col)

for (i in 1:lin) {
  for (j in 1:col) {
    if (matriz_resp[i,j] == matriz_gab[i,j]) {
      matriz_resp_categ[i,j] = 1
    }
  }
}

LC_NOM = matriz_resp ; LC_NOM = as.data.frame(LC_NOM)
LC_DIC = matriz_resp_categ; LC_DIC = as.data.frame(LC_DIC)

#####
# Organizando respostas de acordo com o caderno (cores)
# Os dados foram ordenados (ainda no excel) conforme CO_PROVA
posi_itens = read.csv2("ITENS_PROVA_2019.csv", header = T)[1:740,]

branca = which(posi_itens$TX_COR == "Branca")

```

---

```

posi_itens$TX_COR[branca] = "Cinza"

cad_azul = subset(posi_itens, TX_COR == "Azul") #referencia
cad_amarela = subset(posi_itens, TX_COR == "Amarela")
cad_rosa = subset(posi_itens, TX_COR == "Rosa")
cad_cinza = subset(posi_itens, TX_COR == "Cinza")

ord_azul = match(cad_azul$CO_ITEM, cad_azul$CO_ITEM)
ord_amarela = match(cad_amarela$CO_ITEM, cad_amarela$CO_ITEM)
ord_rosa = match(cad_rosa$CO_ITEM, cad_rosa$CO_ITEM)
ord_cinza = match(cad_cinza$CO_ITEM, cad_cinza$CO_ITEM)

azul_CN = ord_azul[1:45]
azul_CH = ord_azul[46:90]
azul_LC = ord_azul[101:140]
azul_MT = ord_azul[141:185]

amarela_CN = ord_amarela[1:45]
amarela_CH = ord_amarela[46:90]
amarela_LC = ord_amarela[101:140] #nao consideramos questoes de lingua estrangeira
amarela_MT = ord_amarela[141:185]

rosa_CN = ord_rosa[1:45]
rosa_CH = ord_rosa[46:90]
rosa_LC = ord_rosa[101:140]
rosa_MT = ord_rosa[141:185]

cinza_CN = ord_cinza[1:45]
cinza_CH = ord_cinza[46:90]
cinza_LC = ord_cinza[101:140]
cinza_MT = ord_cinza[141:185]

# posicoes na base de dados
resp_azul_CN = which(ENEM19_FORT_Q$CO_PROVA_CN == 503)
resp_azul_CH = which(ENEM19_FORT_Q$CO_PROVA_CH == 507)
resp_azul_LC = which(ENEM19_FORT_Q$CO_PROVA_LC == 511)
resp_azul_MT = which(ENEM19_FORT_Q$CO_PROVA_MT == 515)

resp_amarela_CN = which(ENEM19_FORT_Q$CO_PROVA_CN == 504)
resp_amarela_CH = which(ENEM19_FORT_Q$CO_PROVA_CH == 508)
resp_amarela_LC = which(ENEM19_FORT_Q$CO_PROVA_LC == 512)
resp_amarela_MT = which(ENEM19_FORT_Q$CO_PROVA_MT == 516)

resp_cinza_CN = which(ENEM19_FORT_Q$CO_PROVA_CN == 505)
resp_cinza_CH = which(ENEM19_FORT_Q$CO_PROVA_CH == 509)
resp_cinza_LC = which(ENEM19_FORT_Q$CO_PROVA_LC == 514)
resp_cinza_MT = which(ENEM19_FORT_Q$CO_PROVA_MT == 518)

resp_rosa_CN = which(ENEM19_FORT_Q$CO_PROVA_CN == 506)
resp_rosa_CH = which(ENEM19_FORT_Q$CO_PROVA_CH == 510)
resp_rosa_LC = which(ENEM19_FORT_Q$CO_PROVA_LC == 513)
resp_rosa_MT = which(ENEM19_FORT_Q$CO_PROVA_MT == 517)

# ordenando as respostas

```

---

```

CN_NOM_ord = CN_NOM ; CN_DIC_ord = CN_DIC
CN_NOM_ord[resp_amarela_CN,] = CN_NOM[resp_amarela_CN, amarela_CN]
CN_DIC_ord[resp_amarela_CN,] = CN_DIC[resp_amarela_CN, amarela_CN]
CN_NOM_ord[resp_rosa_CN,] = CN_NOM[resp_rosa_CN, rosa_CN]
CN_DIC_ord[resp_rosa_CN,] = CN_DIC[resp_rosa_CN, rosa_CN]
CN_NOM_ord[resp_cinza_CN,] = CN_NOM[resp_cinza_CN, cinza_CN]
CN_DIC_ord[resp_cinza_CN,] = CN_DIC[resp_cinza_CN, cinza_CN]

CH_NOM_ord = CH_NOM ; CH_DIC_ord = CH_DIC
CH_NOM_ord[resp_amarela_CH,] = CH_NOM[resp_amarela_CH, amarela_CH-45]
CH_DIC_ord[resp_amarela_CH,] = CH_DIC[resp_amarela_CH, amarela_CH-45]
CH_NOM_ord[resp_rosa_CH,] = CH_NOM[resp_rosa_CH, rosa_CH-45]
CH_DIC_ord[resp_rosa_CH,] = CH_DIC[resp_rosa_CH, rosa_CH-45]
CH_NOM_ord[resp_cinza_CH,] = CH_NOM[resp_cinza_CH, cinza_CH-45]
CH_DIC_ord[resp_cinza_CH,] = CH_DIC[resp_cinza_CH, cinza_CH-45]

LC_NOM_ord = LC_NOM ; LC_DIC_ord = LC_DIC
LC_NOM_ord[resp_amarela_LC,] = LC_NOM[resp_amarela_LC, amarela_LC-100]
LC_DIC_ord[resp_amarela_LC,] = LC_DIC[resp_amarela_LC, amarela_LC-100]
LC_NOM_ord[resp_rosa_LC,] = LC_NOM[resp_rosa_LC, rosa_LC-100]
LC_DIC_ord[resp_rosa_LC,] = LC_DIC[resp_rosa_LC, rosa_LC-100]
LC_NOM_ord[resp_cinza_LC,] = LC_NOM[resp_cinza_LC, cinza_LC-100]
LC_DIC_ord[resp_cinza_LC,] = LC_DIC[resp_cinza_LC, cinza_LC-100]

MT_NOM_ord = MT_NOM ; MT_DIC_ord = MT_DIC
MT_NOM_ord[resp_amarela_MT,] = MT_NOM[resp_amarela_MT, amarela_MT-140]
MT_DIC_ord[resp_amarela_MT,] = MT_DIC[resp_amarela_MT, amarela_MT-140]
MT_NOM_ord[resp_rosa_MT,] = MT_NOM[resp_rosa_MT, rosa_MT-140]
MT_DIC_ord[resp_rosa_MT,] = MT_DIC[resp_rosa_MT, rosa_MT-140]
MT_NOM_ord[resp_cinza_MT,] = MT_NOM[resp_cinza_MT, cinza_MT-140]
MT_DIC_ord[resp_cinza_MT,] = MT_DIC[resp_cinza_MT, cinza_MT-140]

# resposta diferentes das alternativas s o consideradas NA
LC_NOM_ord[which(LC_NOM_ord == ".", arr.ind = T)] = NA
LC_NOM_ord[which(LC_NOM_ord == "*", arr.ind = T)] = NA
LC_NOM_ord[which(is.na(LC_NOM_ord), arr.ind = T)] = "NA"

MT_NOM_ord[which(MT_NOM_ord == ".", arr.ind = T)] = NA
MT_NOM_ord[which(MT_NOM_ord == "*", arr.ind = T)] = NA
MT_NOM_ord[which(is.na(MT_NOM_ord), arr.ind = T)] = "NA"

CN_NOM_ord[which(CN_NOM_ord == ".", arr.ind = T)] = NA
CN_NOM_ord[which(CN_NOM_ord == "*", arr.ind = T)] = NA
CN_NOM_ord[which(is.na(CN_NOM_ord), arr.ind = T)] = "NA"

CH_NOM_ord[which(CH_NOM_ord == ".", arr.ind = T)] = NA
CH_NOM_ord[which(CH_NOM_ord == "*", arr.ind = T)] = NA
CH_NOM_ord[which(is.na(CH_NOM_ord), arr.ind = T)] = "NA"

# gabarito
gab_CN = cad_azul$TX_GABARITO[1:45]
gab_CH = cad_azul$TX_GABARITO[46:90]
gab_LC = cad_azul$TX_GABARITO[101:140]
gab_MT = cad_azul$TX_GABARITO[141:185]

```

```
# item 45 do caderno azul de CH foi anulado  
CH_NOM_ord = CH_NOM_ord[,-45]  
CH_DIC_ord = CH_DIC_ord[,-45]
```

---

## Apêndice B

# Algoritmo para gerar estatísticas dos escores de cada área do ENEM-2019, Fortaleza-CE

---

```
#####  
# Carregando pacote  
if(!require (e1071)) install.packages("e1071"); library(e1071)  
  
#####  
# Escore e algumas estatísticas de cada área de Conhecimento  
# As bases de dados  
# CH_DIC_ord  
# CN_DIC_ord  
# MT_DIC_ord  
# LC_DIC_ord  
# foram geradas no Algoritmo A  
  
### CN ###  
SCORE_CN = rowSums(CN_DIC_ord)  
mean(SCORE_CN) # Media  
sd(SCORE_CN) # Desvio Padrao  
100*sd(SCORE_CN)/mean(SCORE_CN) # Coeficiente de Variacao  
skewness(SCORE_CN) # Coeficiente de Assimetria  
kurtosis(SCORE_CN) # Coeficiente de Curtose  
  
x=seq(0,45)  
hist(SCORE_CN, breaks=x, main="Histograma dos Escores em CN",  
 xlab = "Escore (Número de Acertos)", ylab = "Frequência",  
 ylim=c(0,7000))  
legend("topright",  
 legend=c("Média = 13,94", "Desvio = 5,54", "CV = 39,73%"))  
  
### CH ###  
SCORE_CH = rowSums(CH_DIC_ord)  
mean(SCORE_CH) # Media  
sd(SCORE_CH) # Desvio Padrao  
100*sd(SCORE_CH)/mean(SCORE_CH) # Coeficiente de Variacao  
skewness(SCORE_CH) # Coeficiente de Assimetria  
kurtosis(SCORE_CH) # Coeficiente de Curtose
```

---

```
x=seq(0,44)
hist(SCORE_CH, breaks=x, main="Histograma dos Escores em CH",
      xlab = "Escore (N mero de Acertos)", ylab = "Frequencia",
      ylim=c(0,7000))
legend("topright",
       legend = c("Media = 17,79", "Desvio = 7,36", "CV = 41,39%" ))
```

```
### MT ###
SCORE_MT = rowSums(MT_DIC_ord)
mean(SCORE_MT)           # Media
sd(SCORE_MT)            # Desvio Padrao
100*sd(SCORE_MT)/mean(SCORE_MT) # Coeficiente de Variacao
skewness(SCORE_MT)     # Coeficiente de Assimetria
kurtosis(SCORE_MT)     # Coeficiente de Curtose
```

```
x=seq(0,45)
hist(SCORE_MT, breaks=x, main="Histograma dos Escores em MT",
      xlab = "Escore (N mero de Acertos)", ylab = "Frequencia",
      ylim=c(0,7000))
legend("topright",
       legend = c("Media = 12,81", "Desvio = 6,74", "CV = 52,65%" ))
```

```
### LC ###
SCORE_LC = rowSums(LC_DIC_ord)
mean(SCORE_LC)           # Media
sd(SCORE_LC)            # Desvio Padrao
100*sd(SCORE_LC)/mean(SCORE_LC) # Coeficiente de Variacao
skewness(SCORE_LC)     # Coeficiente de Assimetria
kurtosis(SCORE_LC)     # Coeficiente de Curtose
```

```
x=seq(0,45)
hist(SCORE_LC, breaks=x, main="Histograma dos Escores em LC",
      xlab = "Escore (N mero de Acertos)", ylab = "Frequencia",
      ylim=c(0,7000))
legend("topright",
       legend = c("Media = 17,52", "Desvio = 6,36", "CV = 36,32%" ))
```

---

## Apêndice C

# Algoritmo de construção do ICR

---

```
#####  
##%  
##%      DEFINICOES INICIAIS E SINTAXES INTERMEDIARIAS      %%  
##%  
#####  
  
#####  
Carrega_Pacotes=function(){  
  if(!require(mirt)){install.packages("mirt"); library(mirt) }  
  if(!require(mirtCAT)){install.packages("mirtCAT"); library(mirtCAT) }  
  if(!require(irtoys)){install.packages("irtoys"); library(irtoys) }  
  if(!require(maxLik)){install.packages("maxLik"); library(maxLik) }  
  if(!require(readxl)){install.packages("readxl"); library(readxl) }  
  if(!require(beepr)) install.packages("beepr"); library(beepr)  
}  
#####  
  
#####  
Gera_Itens=function(I,kPL,bfixo=T,ordenab=T){  
  # I = numero de itens;  
  # kPL = modelo log stico (1, 2 ou 3 parametros);  
  # bfixo = valores do parametro b, fixos ou aleatorios (default TRUE);  
  # ordenab = valores de b ser o ordenados, do mais facil ao mais dificil  
  (default TRUE);  
  
  #—— parametro a ——#  
  set.seed(1)  
  a=runif(I)*2.5+.75; #vetor linha (1 x I)  
  a=a/D  
  
  #—— parametro b ——#  
  bmin=-2.5  
  if (bfixo) b=bmin+seq(0,I-1)*abs(2*bmin)/(I-1) # de 0 a I-1 tem I elementos  
  else set.seed(1); b=runif(I)*abs(2*bmin)+bmin;  
  
  if (ordenab) b=sort(b) #vetor linha (1 x I)  
  
  #—— parametro c ——#  
  set.seed(1)  
  c=runif(I)*.3;
```

```

#— modelo logístico —#
if (kPL=="1PL" | kPL==1) {a=u=rep(1,I); c=rep(0,I)}
if (kPL=="2PL" | kPL==2) {c=rep(0,I); u=rep(1,I)}
if (kPL=="3PL" | kPL==3) {u=rep(1,I)}

#— matriz de parametros dos itens —#
zeta=cbind(a,b,c,u)
colnames(zeta)=c("a","b","c","u")
cat("\nParametros de itens: \n"); print(head(zeta))

return(zeta)
}
#####
#####
Gera_Habilidades = function(n, dist=1, par1=0, par2=1, par3=15){
  # n = numero de repeticoes;
  # dist = distribuicao a ser usada (1-Normal, 2-Uniforme, 3-T de Student)
  (default 1);
  # par1 = primeiro parametro da distribuicao;
  # par2 = segundo parametro da distribuicao;

  #— escolha da distribuicao —#
  set.seed(1)
  if (dist==1) {theta=rnorm(n,par1,par2); cat("\nGerando Normal com media ",
  par1," e variancia ",par2,".\n")}
  if (dist==2) {theta=runif(n,par1,par2); cat("\nGerando Uniforme com parametros ",
  par1," e ",par2,".\n")}
  if (dist==3) {theta=rt(n,par3); cat("\nGerando t-Student com parametro ",
  par3,".\n")}

  return(theta)
}
#####
#####
Gera_Dados=function(zeta,theta){
  # zeta = matriz de parametros dos itens (saida de Gera_Items);
  # theta = vetor de valores das habilidades (saida de Gera_Habilidades);

  #— organizando os vetores de parametros em linhas —#
  I=nrow(zeta); n=NROW(theta)
  a=zeta[,1]; a=t(a) #vetor linha (1 x I)
  b=zeta[,2]; b=t(b) #vetor linha (1 x I)
  c=zeta[,3]; c=t(c) #vetor linha (1 x I)

  #— repeticao dos parametros em n linhas de individuos —#
  ones=rep(1,n) # vetor coluna de uns (n x 1)
  ax=kronecker(ones,a) ; #ax tem dimensao n x I : a (1 x I) repetido em n linhas
  bx=kronecker(ones,b) ; #bx tem dimensao n x I : b (1 x I) repetido em n linhas
  cx=kronecker(ones,c) ; #cx tem dimensao n x I : c (1 x I) repetido em n linhas
  #dim(ax);dim(bx);dim(cx)

  #— repeticao das habilidades em I colunas de Itens —#

```

---

```

ones=matrix(rep(1,I),1,I) # vetor linha de uns (1 x I)
thetak=kronecker(theta,ones) ;
#dim(thetak);

#--- matriz dicotomica U ---#
P=cx+(1-cx)/(1+exp(-D*ax*(thetak-bx))) #P: matriz de probabilidades

set.seed(1)
X=runif(n*I); dim(X)=c(n,I);           #X: Matrix obtida de uma Unif(0,1)
U=1*(X<P);                             #U: matriz de dados (0,1)
colnames(U)=paste0("Item.",1:I)

return(U)
}
#####
#####
Grafico=function(X, dx=.1){
  # X = Soma de Bernoullis;
  # dx = distancia de um ponto a outro

  cat("\nMinimo: ", min(X), ", ", "Maximo: ", max(X), "\n")
  x=c(-1000, seq(-4,4,dx), 1000)
  H = hist(X, breaks=x+dx/2, plot = FALSE)
  Freq=H$density; Freq=Freq[-length(Freq)]
  x=x[-c(1, length(x))]
  y.max=max(Freq)
  y=sum(dnorm(x))
  plot(x,Freq, main=" indice de Coerencia de Respostas", xlab="Z",
        ylab="Densidade", ylim=c(0,max(dnorm(x),y.max)))

  normal = curve(dnorm(x), n = 81, add=TRUE)
  abline(h=seq(0,1,.1), col="lightgray", lty="dotted")
  abline(v=(seq(-4,4,dx)), col="lightgray", lty="dotted")

  return(list(Freq = Freq, normal = normal$y, erro = Freq - normal$y))
}
#####
#####
Estima_Fixos=function(Ux, zeta){
  # Ux = dados dicotomicos;
  # zeta = matrix de parametros.

  # Construindo o INTERCEPTO para o MIRT, d=-a*b
  zeta[,2] = -zeta[,1]*zeta[,2]

  pars <- mirt(Ux, 1, itemtype = '3PL', pars = 'values')

  #trocando os parametros do template do mirt pelos valores fixados, e
  FALSE para nao estimar
  k=1
  It = NROW(zeta)
  for (i in 1:It){

```

```

ip=4*i-3
FIXED = 1:It
if (i %in% FIXED) {
  lin.a=4*i-3; pars$value[lin.a]=zeta[k,1];      pars$est[lin.a]=FALSE
  lin.b=4*i-2; pars$value[lin.b]=zeta[k,2];      pars$est[lin.b]=FALSE
  lin.c=4*i-1; pars$value[lin.c]=zeta[k,3];      pars$est[lin.c]=FALSE
  k=k+1; #cat("\nO item",k,"foi atualizado!\n")
}
else{ # (priori, media, desvio-padrao)
  pars[ip,10] = "lnorm" # Mesmo do BILOG-MG
  pars[ip,11:12] = c(1,exp(0.5)) # Mesmo do BILOG-MG

  pars[ip+1,10] = 'norm' # Mesmo do BILOG-MG
  pars[ip+1,11:12] = c(0,2) # Mesmo do BILOG-MG

  pars[ip+2,10] = 'norm' # No BILOG-MG e 'Beta(5,17)'.
  pars[ip+2,11:12] = c(-1,.33) # No BILOG-MG e Beta(5,17) pela parametrizacao.
}
}

pars.f = mirt(Ux,1, itemtype = '3PL', pars = pars)

return(pars.f)
}
#####
#####
Coerencia=function(Ux, ZETA=NULL, THETA=NULL, ZETA_FIXO=FALSE, METODO){
  # Ux = Base de dados dicotomicos;
  # ZETA = Matriz de parametros dos itens (default NULL);
  # THETA = Vetor de valores das habilidades (default THETA);
  # ZETA_FIXO = Matriz de parametros fixadas (default FALSE);
  # METODO = M todo de estima o das habilidades (EAP ou VP).

  #---verificando se a entrada de dados contem apenas os valores dicotomicos---#
  Verifica_Itens=function(k=20){
    for (i in 1:ncol(Ux)) if (length(table(Ux[1:1000,i]))>k)
      stop("\nFavor fornecer somente os dados de itens. A execucao sera finalizada!\n")
  }
  Verifica_Itens()

  if (METODO=="EAP") metodo = "EAP"
  if (METODO=="VP") metodo = "plausible"

  #--- Valores dos parametros zeta e theta ---#
  # ZETA e THETA verdadeiros
  if (!is.null(ZETA) & !is.null(THETA) & ZETA_FIXO==FALSE) {
    zetax=ZETA; thetax=THETA
  }
  # ZETA verdadeiro e THETA estimado
  if (!is.null(ZETA) & is.null(THETA) & ZETA_FIXO==TRUE) {
    mod=ZETA
    zetax=coef(mod, IRTpars=TRUE, simplify=TRUE)[[1]]
  }
}

```

```

thetax=fcores(mod, method = metodo)
cat("\n\nProficiencias_estimadas_com_media: ",round(mean(thetax),3),
    "\ne_desvio-padrao: ", round(sd(thetax),3),'\n')
}
# ZETA e THETA estimados
if (is.null(ZETA) & is.null(THETA) & ZETA_FIXO==FALSE) {
  PL=paste0(kPL,"PL")
  if (kPL == 1) PL = "Rasch";
  mod = mirt(Ux,1, itemtype = PL, technical = list(NCYCLES = 2000))
  zetax=coef(mod, IRTpars=TRUE, simplify=TRUE)[[1]]
  thetax=fcores(mod, method = metodo)
  cat("\n\nProficiencias_estimadas_com_media: ",round(mean(thetax),3),
      "\ne_desvio-padrao: ", round(sd(thetax),3),'\n')
}

#--- Probabilidades pelo modelo 4PL Multidimensional---#
a = zetax[,1]; b = zetax[,2]; c = zetax[,3]; gm= zetax[,4];
d=-a*b #Vetor (nx1)
N=nrow(Ux)
I=ncol(Ux)

eta = thetax%*% t(a) + matrix(d,N,I,byrow=TRUE);
eta = t(eta) # n x I (a'theta+d)
P = c + (gm-c)/(1+exp(-eta)); P=t(P) # n x I

#--- Valores do ICR ---#
S=rowSums(Ux, na.rm = T)
P=P*(!is.na(Ux))
ES=rowSums(P, na.rm = T)
DPS=sqrt(rowSums(P*(1-P), na.rm = T))
ICR=(S-ES)/DPS
cat("\nMedia_do_ICR: ",round(mean(ICR),5), "\ne_Desvio-Padrao_do_ICR: ",
    round(sd(ICR),5),'\n')

Gf = Grafico(ICR)
EQM = round(mean(Gf$erro^2),5)
return(list(ZETA=zetax, THETA=thetax, P=P, S=S, ES=ES, DPS=DPS,
            ICR=ICR, EQM=EQM, perc=quantile(ICR,c(0.05,0.95))) )
}
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
#%
#%
#%
#%
#%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
#***** DEFINICOES PARA O PROCESSO *****
I = 22 # Numero de Itens
N = 50000 # Numero de Individuos
NALT = 4 # N mero de Alternativas
kPL = 3 # Modelo Logistico: 1(PL), 2(PL) or 3(PL)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

---

```

Carrega_Pacotes()

zeta = Gera_Itens(I, kPL)          # Gerando parametros dos itens

theta = Gera_Habilidades(N, dist=1) # Gerando habilidades

U = Gera_Dados(zeta, theta)      # Gerando conjunto de dados

#####
# Usando ZETA & THETA verdadeiros

ICR11 = Coerencia(U, ZETA=zeta, THETA=theta, METODO="VP")
mtext("Itens□e□Habilidades□Conhecidos")
legend("topright", legend = c("Curva□Normal", "Pontos□do□ICR"),
       lty = c(1,NA), pch = c(NA,1))

#####
# Usando ZETA verdadeiro e THETA estimado

zeta2 = Estima_Fixos(Ux=U, zeta = zeta)

ICR10 = Coerencia(U, ZETA=zeta2, ZETA_FIXO = TRUE, METODO = "VP")
mtext("Itens□Conhecidos□e□Habilidades□Desconhecidas□/□M todo□VP")
legend("topright", legend = c("Curva□Normal", "Pontos□do□ICR"),
       lty = c(1,NA), pch = c(NA,1))
mean((theta-ICR10$THETA)^2)
plot(theta, ICR10$THETA, xlab="theta□verdadeiro", ylab="theta□estimado",
     xlim=c(-4,4), ylim=c(-4,4))
mtext("Usando□VP")
lines(c(-4,4),c(-4,4), col = "blue")

#####
# Usando ZETA & THETA estimados

ICR00 = Coerencia(U, METODO = "VP");
mtext("Itens□e□Habilidades□Desconhecidos□/□M todo□VP")
legend("topright", legend = c("Curva□Normal", "Pontos□do□ICR"),
       lty = c(1,NA), pch = c(NA,1))
mean((theta-ICR00$THETA)^2)
plot(theta, ICR00$THETA, xlab="theta□verdadeiro", ylab="theta□estimado",
     xlim=c(-4,4), ylim=c(-4,4))
mtext("Usando□VP")
lines(c(-4,4),c(-4,4), col = "blue")

#####
# CRITERIO DE DECISAO

#####
# Considerando dados totalmente aleatorios
Dados_aleat = function(ICR, ES, DPS, NALT){
  # ICR = Vetor de valores do ICR

```

```

# ES = Vetor de esperanças do Escores;
# DPS = vetor de desvios padrao do Escores;
# NALT = numero de alternativas nos itens.

#— Conjunto de dados Aleat rios —#
P = 1/NALT
set.seed(1)
X=runif(N*I); dim(X)=c(N,I);           #X: Matrix obtida de uma Unif(0,1)
U=1*(X<P);                             #U: matriz de dados (0,1)
colnames(U)=paste0("Item.",1:I)

#— Valores do Z —#
S=rowSums(U, na.rm = T)
Z=(S-ES)/DPS
cat("\nMedia do Z: ",round(mean(Z),5), " e Desvio Padrao do Z: ", round(sd(Z),5),
    '\n')

cat("\nMinimo de Z: ", min(Z), " e Maximo: ",max(Z), "\n")

#— Estimando os valores de mu e sigma —#
min_Z2 = mean(Z) - 2*sd(Z)
Z2 = subset(Z, Z>min_Z2)
prop = NROW(Z2)/NROW(Z)
mu = mean(Z2)
sig = sd(Z2)
sig2 = sig^2
cat("\nZ' e Normal com media ", mu, " e vari ncia ",sig2, "\n")
cat("\nZ' estimado com ", prop, "% dos dados\n")

z = (-mu-sig*sqrt(mu^2+2*sig2*log(sig)-2*log(sig)))/(sig2-1)
cat("\nSe o indiv duo tem ICR >=", z,
    ", ent o ele tem padr o de resposta coerente.\n")
cat("\nSe o indiv duo tem ICR <=", z,
    ", ent o ele n o tem padr o de resposta coerente.\n")

#— Grafico —#
plot(density(ICR), xlim = c(-30, 4), main="Gr fico de ICR e Z", xlab="z",
     ylab="Densidade", lwd = 2)
lines(density(Z), col="blue", lwd = 2)
curve(dnorm(x, mu, sig), add=TRUE, col="red", lwd = 2)
legend("topleft", legend=c("ICR", "Z", "Z'"), lty=c(1,1,1),
      col=c("black", "blue", "red"), lwd=c(2,2,2), bty=c("n", "n", "n"))

x1 = z ; x2 = z
y1 = 0 ; y2 = dnorm(z, mu, sig)

plot(density(ICR), xlim = c(-5, 0), main="Gr fico de ICR e Z", xlab="z",
     ylab="Densidade", lwd = 2)
lines(density(Z), col="blue", lwd=2)
curve(dnorm(x, mu, sig), add=TRUE, col="red", lwd=2)
lines(c(x1, x2), c(y1, y2), lty = 2)
legend("topleft", legend=c("ICR", "Z", "Z'"), lty=c(1,1,1),
      col=c("black", "blue", "red"), lwd=c(2,2,2), bty=c("n", "n", "n"))

```

```

    return(list(S_linha = S, Z = round(Z,3), Z_linha = round(Z2,3)))
}
#####

Dados_aleat(ICR00$ICR, ICR00$ES, ICR00$DPS, NALT)

#####
# Encontrando uma funcao geral para o criterio de decisao

# Foram gerados os varios pontos de corte do criterio de decisao
# Fixamos as alternativas e variamos o numero de itens (20, 22, ..., 60)
# Apresentamos aqui apenas os resultados que foram obtidos gerando um por um
  a partir das funcoes anteriores
# Com duas alternativas
z_2 = c(-1.097226, -1.370016, -1.133898, -1.160509, -1.195897, -1.211221,
        -1.251258, -1.301745, -1.389349, -1.370644, -1.404584, -1.409438,
        -1.440798, -1.435346, -1.435346, -1.466299, -1.482857, -1.479546,
        -1.519633, -1.478436, -1.507093, -1.524149, -1.536190, -1.541337,
        -1.544184, -1.553153)
# Com tres alternativas
z_3 = c(-1.028853, -1.003402, -1.320887, -1.377150, -1.469799, -1.495891,
        -1.416187, -1.449341, -1.338089, -1.428643, -1.509676, -1.560666,
        -1.539357, -1.604984, -1.637171, -1.597797, -1.669256, -1.710789,
        -1.837316, -1.736888, -1.771231, -1.799539, -1.815446, -1.725840,
        -1.776827, -1.792139)
# Com quatro alternativas
z_4 = c(-1.179052, -1.129694, -1.515464, -1.572889, -1.667564, -1.704978,
        -1.609026, -1.633998, -1.507925, -1.614638, -1.683839, -1.745252,
        -1.718131, -1.797334, -1.817962, -1.789292, -1.857038, -1.913778,
        -2.049050, -1.958472, -1.984268, -2.011698, -2.029781, -1.932594,
        -1.994034, -2.010803)
# Com cinco alternativas
z_5 = c(-1.289659, -1.241522, -1.645303, -1.701288, -1.792579, -1.836835,
        -1.738794, -1.759624, -1.638071, -1.749941, -1.806706, -1.871432,
        -1.844495, -1.927609, -1.939870, -1.919654, -1.981957, -2.045076,
        -2.179669, -2.101012, -2.119719, -2.145984, -2.165386, -2.071170,
        -2.135609, -2.153346)
dados_z = c(z_2, z_3, z_4, z_5)

n_alt = c(rep(2,26), rep(3,26), rep(4, 26), rep(5,26))
n_itens = rep(seq(10,60, 2), 4)
dados = cbind(n_alt, n_itens, dados_z); dados = as.data.frame(dados)
library(scatterplot3d)
teste = scatterplot3d(dados, xlab = "N mero de Alternativas",
                     ylab = "N mero de Itens", zlab = "z", pch=16,
                     xlim=c(2.5, 5.5), angle=60, color="steelblue", type="h")

# gerando dois modelos distintos para ajuste
modelo1 = lm(data = dados, formula = dados_z ~ n_alt + n_itens)
modelo2 = lm(data = dados, formula = dados_z ~ n_alt + log(n_itens))
summary(modelo)
teste = scatterplot3d(dados, pch=16, xlim=c(1.5, 5.5), angle=70,
                     color="steelblue", type="h", xlab = "N mero de Alternativas",
                     ylab = "N mero de Itens", zlab = "Valores de z")

```

---

```

teste = scatterplot3d(dados, pch=16, xlim=c(1.5, 5.5), angle=170,
                      color="steelblue", type="h", xlab = "N mero_de_Alternativas",
                      ylab = "N mero_de_Itens", zlab = "Valores_de_z")
teste$plane3d(modelo1, draw_polygon = TRUE, col="blue")
teste$contour3d(modelo2, lty="24", x.count = 200, col="blue")
dados_z2 = -0.25780 - 0.16424*n_alt - 0.38522*log(n_itens)
cbind(dados, dados_z2)

plot(seq(10,60,2), z_2, type="l", ylim=c(-2.2, -1), lwd=1, lty=2,
      xlab="N mero_de_Itens", ylab="Valores_de_z")
lines(seq(10,60,2), z_3, col=2, lwd=1, lty=2)
lines(seq(10,60,2), z_4, col=3, lwd=1, lty=2)
lines(seq(10,60,2), z_5, col=4, lwd=1, lty=2)
legend("topright",
       legend = c("N Alt=2", "N Alt=3", "N Alt=4", "N Alt=5"),
       col=c(1,2,3,4), lwd=c(2,2,2,2), bty=c("n", "n", "n", "n"))
lines(n_itens[1:26], 0.25780 - 0.16424*2 - 0.38522*log(n_itens[1:26]), lwd=2, col=1)
lines(n_itens[1:26], 0.25780 - 0.16424*3 - 0.38522*log(n_itens[1:26]), lwd=2, col=2)
lines(n_itens[1:26], 0.25780 - 0.16424*4 - 0.38522*log(n_itens[1:26]), lwd=2, col=3)
lines(n_itens[1:26], 0.25780 - 0.16424*5 - 0.38522*log(n_itens[1:26]), lwd=2, col=4)

# escolhendo o melhor modelo
modelo1 = lm(formula = z_4 ~ n_itens)
modelo2 = lm(formula = z_4 ~ log(n_itens[1:26]))
AIC(modelo1)
AIC(modelo2) # aquela com menor valor de AIC eh o modelo de melhor ajuste

```

---

## Apêndice D

# Algoritmo para aplicação do ICR aos dados de Ciências Humanas e suas Tecnologias do ENEM-2019, Fortaleza-CE

---

```
#####  
# Tomamos os dados CH_DIC_ord  
# que foram obtidos no Algoritmo A  
# e a fun o Coerencia()  
# que foi obtida no Algoritmo B  
  
if(!require(scales)) install.packages("scales"); library(scales)  
  
# ----- #  
# Estatísticas do ICR para área de Ciências Humanas #  
# ----- #  
CH_ICR = Coerencia(CH_DIC_ord, METODO = "VP")  
mtext("Usando os dados de Ciências Humanas")  
legend("topright", legend = c("Curva Normal", "Pontos do ICR"), lty = c(1,NA),  
      pch = c(NA,1))  
  
# Graficos dos parametros dos itens  
#a  
plot(seq(1,44), CH_ICR$ZETA[,1], type="h", col="blue", xlab="Itens",  
      ylab="Valor do parametro a")  
mtext("Estimativa do parametro a de Ciências Humanas")  
abline(h=seq(-1,3,0.5), col="gray", lty="dotted")  
  
#b  
plot(seq(1,44), CH_ICR$ZETA[,2], type="h", col="blue", xlab="Itens",  
      ylab="Valor do parametro b")  
mtext("Estimativa do parametro b de Ciências Humanas")  
abline(h=seq(-4,3,0.5), col="gray", lty="dotted")  
  
#c  
plot(seq(1,44), CH_ICR$ZETA[,3], type="h", col="blue", xlab="Itens",  
      ylab="Valor do parametro c")  
mtext("Estimativa do parametro c de Ciências Humanas")  
abline(h=seq(0,0.35,0.05), col="gray", lty="dotted")
```

```

# Grafico dos parametros dos individuos
hist(CH_ICR$THETA, freq=FALSE,
      main="Histograma das habilidades em Ci ncias Humanas",
      xlab="theta(habilidades)", ylab="Frequ ncia", xlim=c(-4,4), ylim=c(0,0.4))
curve(dnorm(x), add=TRUE)#mean(CH_ICR$THETA)
#sd(CH_ICR$THETA)
#100*sd(CH_ICR$THETA)/mean(CH_ICR$THETA)
#library(e1071)
#skewness(CH_ICR$THETA)
#kurtosis(CH_ICR$THETA)

#####
# analisando a CCI do item 3
alt = c("A", "B", "C", "D", "E")

CHt = cbind(CH_NOM_ord, CH_ICR$THETA)
CHt = CHt[order(CHt[,45]),]

table(round(CHt[,45],0)) #usamos para encontrar os "tamanhos" de theta
(numero de individuo em cada escore)

ini = c(7, 379, 5085, 22237, 48024, 66560, 70851)
fin = c(378, 5084, 22236, 48023, 66559, 70850, 71016)

CH_freq = matrix(0, nrow=7, ncol=5)
colnames(CH_freq) = alt

for(i in 1:7) CH_freq[i,] = t(table(CHt[ini[i]:fin[i],5]))[1:5])

for (i in 1:7) CH_freq[i,] = CH_freq[i,]/sum(CH_freq[i,])

# Grafico1
plot(seq(-3,3), CH_freq[,1], type = "b", lty = 2, pch="A", col=1, ylim=c(0,1),
      xlab="theta", ylab="Propor o de Respostas", main= "CH: Item 1", lwd=2)
lines(seq(-3,3), CH_freq[,2], type = "b", lty = 2, pch="B", col=2, lwd=2)
lines(seq(-3,3), CH_freq[,3], type = "b", lty = 2, pch="C", col=3, lwd=2)
lines(seq(-3,3), CH_freq[,4], type = "b", lty = 2, pch="D", col=4, lwd=2)
lines(seq(-3,3), CH_freq[,5], type = "b", lty = 2, pch="E", col=5, lwd=2)
abline(v=seq(-3,3,1), col="lightgray", lty="dotted")
abline(h=seq(0,1,0.1), col="lightgray", lty="dotted")
legend("topleft", legend = "Gabarito: C")

# Grafico2
CHtt = cbind(CH_ICR$P, CH_ICR$THETA)
CHtt = CHtt[order(CHtt[,45]),]
plot(seq(-3,3), CH_freq[,3], type = "l", lty = 2, xlab="theta",
      ylab="Propor o de Acertos", main= "CH: Item 3", lwd=2, ylim=c(0,1))
lines(CHtt[,45], CHtt[,3], col = 2, lwd=2)
legend("topleft", legend=c("Verdadeira", "Estimada"), lty=c(2,1), lwd=c(2,2),
      col=c(1,2))

#####
# Retirando os dados do item 3 de CH

```

---

```

CH_DIC_sem3 = CH_DIC_ord[, -3]
S=rowSums(CH_DIC_sem3, na.rm = T)
P = CH_ICR$P[, -3]
ES = rowSums(P, na.rm = T)
DPS = sqrt(rowSums(P*(1-P), na.rm = T))
ICR = (S-ES)/DPS
cat("\nMedia do ICR: ", round(mean(ICR), 5), " e Desvio-Padrao do ICR: ",
    round(sd(ICR), 5), "\n")

Gf = Grafico(ICR)
mtext("Usando os dados de Ci ncias Humanas")
legend("topright", legend = c("Curva Normal", "Pontos do ICR"),
      lty = c(1, NA), pch = c(NA, 1))

mean(ICR)
sd(ICR)
round(mean(Gf$erro^2), 5)
quantile(ICR, c(0.05, 0.95))

# ----- #
# Retirando os indiv duos que colaram e reestimando os parametros #
# ----- #

# ----- CH ----- #
I = 43 ; NALT = 5
z = - 0.25780 - 0.16424*NALT - 0.38522*log(I)
# z = -2.52789

# Enumeramos os individuos e adicionamos aos dados junto com o respectivo ICR
CH2 = cbind(CH_DIC_sem3, seq(1:71016), ICR)
colnames(CH2)[44:45] = c("Ind", "ICR")
CH2 = subset(CH2, ICR >= z)

#reestimamos os par metros considerando apenas respostas coerentes
mod_CH2 = mirt(CH2[1:43], 1, itemtype = "3PL", technical = list(NCYCLES = 2000))
zeta_CH2 = coef(mod_CH2, IRTpars=TRUE, simplify=TRUE)[[1]]

# Consideramos novamente o conjunto de dados completo para estimar as proficiencias
zeta_CH2 = Estima_Fixos(CH_DIC_sem3, zeta_CH2)
CH2_ICR = Coerencia(CH_DIC_sem3, zeta_CH2, ZETA_FIXO = TRUE, METODO = "VP")
mtext("Dados de Ci ncias Humanas sem aleatoriedade")
legend("topright", legend = c("Curva Normal", "Pontos do ICR"), lty = c(1, NA),
      pch = c(NA, 1))

# mean(CH2_ICR$ICR)
# sd(CH2_ICR$ICR)
# kurtosis(CH2_ICR$ICR)
# skewness(CH2_ICR$ICR)
# CH2_ICR$EQM

# Graficos dos parametros dos itens
#a
plot(c(1, 2, seq(4, 44)), CH_ICR$ZETA[-3, 1], type="h", col="blue", xlab="Itens",
     ylab="Valor do parametro", ylim=c(0, 4))

```

```

mtext(" Reestima o do par metro a de Ci ncias Humanas")
abline(h=seq(0,4,0.5), col="gray", lty="dotted")
lines(c(1,2,seq(4,44)), CH2_ICR$ZETA[,1], col="red", type="p")
legend("topright", legend=c("Dados completos", "Dados sem aleatoriedade"),
        lty = c(1,NA), pch = c(NA,1), col=c("blue", "red"))

plot(c(1,2,seq(4,44)),CH_ICR$ZETA[-3,1]-CH2_ICR$ZETA[,1], type="h", col="blue",
      xlab="Itens", ylab="Diferen a")
mtext(" Diferen a dos dados com e sem aleatoriedade / par metro a (CH)")
abline(h=seq(-0.03,0.01,0.005), col="gray", lty="dotted")

plot(CH_ICR$ZETA[-3,1], CH2_ICR$ZETA[,1], xlab="Dados_COM aleatoriedade",
      ylab="Dados_SEM aleatoriedade")
lines(c(0,4), c(0,4))
mtext(" Par metro a de CH")

#b
plot(c(1,2,seq(4,44)), CH_ICR$ZETA[-3,2], type="h", col="blue", xlab="Itens",
      ylab="Valor do par metro b", ylim=c(-1,4))
mtext(" Reestima o do par metro b de Ci ncias Humanas")
abline(h=seq(-1,4,0.5), col="gray", lty="dotted")
lines(c(1,2,seq(4,44)), CH2_ICR$ZETA[,2], col="red", type="p")
legend("topright", legend=c("Dados completos", "Dados sem aleatoriedade"),
        lty = c(1,NA), pch = c(NA,1), col=c("blue", "red"))

plot(c(1,2,seq(4,44)),CH_ICR$ZETA[-3,2]-CH2_ICR$ZETA[,2], type="h", col="blue",
      xlab="Itens", ylab="Diferen a")
mtext(" Diferen a dos dados com e sem aleatoriedade / par metro b (CH)")
abline(h=seq(-0.1,0.1,0.025), col="gray", lty="dotted")

plot(CH_ICR$ZETA[-3,2], CH2_ICR$ZETA[,2], xlab="Dados_COM aleatoriedade",
      ylab="Dados_SEM aleatoriedade")
lines(c(-2,3), c(-2,3))
mtext(" Par metro b de CH")

#c
plot(c(1,2,seq(4,44)), CH_ICR$ZETA[-3,3], type="h", col="blue", xlab="Itens",
      ylab="Valor do par metro c", ylim=c(0,0.35))
mtext(" Estima o do par metro c de Ci ncias Humanas")
abline(h=seq(0,0.35,0.05), col="gray", lty="dotted")
lines(c(1,2,seq(4,44)), CH2_ICR$ZETA[,3], col="red", type="p")
legend("topright", legend=c("Dados completos", "Dados sem aleatoriedade"),
        lty = c(1,NA), pch = c(NA,1), col=c("blue", "red"))

plot(c(1,2,seq(4,44)),CH_ICR$ZETA[-3,3]-CH2_ICR$ZETA[,3], type="h", col="blue",
      xlab="Itens", ylab="Diferen a")
mtext(" Diferen a dos dados com e sem aleatoriedade / par metro c (CH)")
abline(h=seq(-0.015,0.01,0.0025), col="gray", lty="dotted")

plot(CH_ICR$ZETA[-3,3], CH2_ICR$ZETA[,3], xlab="Dados_COM aleatoriedade",
      ylab="Dados_SEM aleatoriedade")
lines(c(-1,1), c(-1,1))
mtext(" Par metro c de CH")

```

---

```

# Grafico dos parametros dos individuos
# library(scales)
# hist(CH_ICR$THETA, main="Histograma das habilidades em Ci ncias Humanas",
      xlab="theta (habilidades)", ylab="Frequ ncia", col=alpha("red", 0.6),
      xlim=c(-4,4))
# hist(CH2_ICR$THETA, main=NULL, col=alpha("blue", 0.4), add=T)
# hist(CH_ICR$THETA-CH2_ICR$THETA,
      main="Diferen as das habilidades em CH com e sem aleatoriedade",
      xlab="theta (habilidades)", ylab="Frequ ncia", xlim=c(-4,4))

#####
# Comparando as Somas das probabilidades de acerto antes e depois

Somap_antes = ES
Somap_depois = CH2_ICR$ES

CH_chute = subset(cbind(ICR, Somap_antes, Somap_depois), ICR<z)
plot(CH_chute[,2], CH_chute[,3], xlab="Soma das prob. de acerto ANTES",
     ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")
mtext("ICR< -2.528")

#CH_coerente = subset(cbind(ICR, Somap_antes, Somap_depois), ICR>=z)
#plot(CH_coerente[,2], CH_coerente[,3], xlab="Soma das prob. de acerto ANTES",
     ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
#lines(c(0,43),c(0,43), col = "red")
#mtext("ICR >= z")

CH_coerente2 = subset(cbind(ICR, Somap_antes, Somap_depois), ICR>=z & ICR<(-2.048))
plot(CH_coerente2[,2], CH_coerente2[,3], xlab="Soma das prob. de acerto ANTES",
     ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")
mtext("-2.528<=ICR<-2.048")

CH_coerente3 = subset(cbind(ICR, Somap_antes, Somap_depois),
                      ICR>=(-2.048) & ICR<(-1.029))
plot(CH_coerente3[,2], CH_coerente3[,3], xlab="Soma das prob. de acerto ANTES",
     ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")
mtext("-2.048<=ICR<-1.029")

CH_coerente4 = subset(cbind(ICR, Somap_antes, Somap_depois),
                      ICR>=(-1.029) & ICR<(-0.01))
plot(CH_coerente4[,2], CH_coerente4[,3], xlab="Soma das prob. de acerto ANTES",
     ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")
mtext("-1.029<=ICR<-0.01")

CH_coerente5 = subset(cbind(ICR, Somap_antes, Somap_depois),
                      ICR>=(-0.01) & ICR<1.009)
plot(CH_coerente5[,2], CH_coerente5[,3], xlab="Soma das prob. de acerto ANTES",
     ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")

```

---

```
mtext(" -0.01<=ICR<1.009 ")

CH_coerente6 = subset(cbind(ICR, Somap_antes, Somap_depois),
                      ICR>=1.009 & ICR<2.028)
plot(CH_coerente6[,2], CH_coerente6[,3], xlab="Soma das prob. de acerto ANTES",
      ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")
mtext(" 1.009<=ICR<2.028 ")

CH_coerente7 = subset(cbind(ICR, Somap_antes, Somap_depois), ICR>=2.028)
plot(CH_coerente7[,2], CH_coerente7[,3], xlab="Soma das prob. de acerto ANTES",
      ylab="Soma das prob. de acerto DEPOIS", xlim=c(0,43), ylim=c(0,43))
lines(c(0,43),c(0,43), col = "red")
mtext(" ICR>=2.028 ")
```