Desempenho Escolar em Pernambuco: Análise dos Itens e das Habilidades usando Teoria Clássica e TRI

LILIAN MARIA SANTOS

Orientador: Maria Cristina Falcão Raposo

Co-orientador: Manoel Raimundo de Sena Junior

Área de Concentração: Estatística Aplicada

Dissertação submetida como requerimento parcial para obtenção do grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2008

Santos, Lilian Maria
Desempenho escolar em Pernambuco: análise dos itens e das habilidades usando teoria clássica e TRI / Lilian Maria Santos. – Recife : O Autor, 2008.
xii, 89 folhas : il., fig., tab.

Dissertação (mestrado) - Universidade Federal de Pernambuco. CCEN. Departamento de Estatística, 2008.

Inclui bibliografia e apêndices.

1. Estatística aplicada. 2. TRI. 3. Avaliação educacional. 4. Desempenho escolar. I. Título.

310 CDD (22.ed.) MEI2008-15

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

29 de fevereiro de 2008
(data)

Nós recomendamos que a dissertação de mestrado de autoria de

Lílian Maria Santos

intitulada

"Desempenho escolar em Pernambuco: análise dos itens e das habilidades usando Teoria Clássica e TRI"

seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Klaus Leite Pinto Vasconcellos
Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Maria Cristina Falcão Raposo
Maria Cristina Falcão Raposo orientador

Marcel D. Toledo Vieira
Marcel Toledo Vieira **(UFJF)**

Claudia Regina Lima
Claudia Regina Oliveira de Paiva Lima

Este documento será anexado à versão final da dissertação.

Este trabalho é carinhosamente dedicado

aos meus pais.

Agradecimentos

Aos meus pais, Domingos e Lindaura, pelo amor incondicional, carinho, dedicação, confiança, paciência, amizade e por terem me ensinado os valores e princípios que até hoje norteiam a formação do meu caráter.

Aos meus irmãos, Nilton, Nilza, Paulo, Lindinalva, Elma, Geisa e Janete, pelo amor incondicional, carinho, confiança e amizade.

Ao meu namorado, Liba, por estar sempre caminhando junto comigo nessa jornada.

À professora Maria Cristina Falcão Raposo, pela orientação e pelo exemplo de profissional, mãe, amiga e professora.

Ao professor Manoel Raimundo de Sena Junior, pela co-orientação.

À professora Maria Luiza Santos, pelo apoio a esse trabalho.

Aos meus amigos do Curso de Doutorado, Tatiene, Tarciana e Daniel, pela amizade sincera e companheirismo.

À minha amiga Jane pela amizade e convivência.

Aos meus amigos: Marcelo, Hemílio, Raphael e Valmir pelo apoio, atenção e amizade.

Às minhas amigas: Lidia e Juliana, pela amizade e parceria nos momentos de ansiedade, estudo e de alegria compartilhados.
Aos meus amigos do mestrado, Larissa, Abraão, Fábio, Silvinha e Edwin pela amizade e companheirismo.

Aos meus amigos do "ano da tese", Uilton, Manoel, Isabel, Andréa, Alice, Wagner, Olga e Cícero.

Aos meus amigos de turma que não seguiram comigo no mestrado, Bruno, Cecílio, Marcele, Alan, Vera, Andréa e Cácio.

Aos meus amigos que já concluiram, Katya, Carlos, Geraldo, Ênio, Lusmarina, Themis, Arthur.

À Valéria Bittencourt, pelo enorme carinho, paciência e amizade com que sempre tratou a mim e aos demais alunos do mestrado.

Aos professores do Departamento de Estatística, em especial aos professores Francisco Cribari–Neto, Klaus Leite Pinto Vasconcellos, Cristiano Ferraz e Claúdia Regina Lima.

Aos funcionários do Departamento de Estatística, em especial a Maurício, Neto, Cândido, Cícero e Tonho por toda ajuda a mim dispensada.

À banca examinadora, pelas valiosas críticas e sugestões.

À SEDUC - Secretaria de Educação do Estado de Pernambuco por disponibilizar os dados para esse trabalho.

Ao CNPQ, pelo apoio financeiro.

Se cada dia cai

Se cada dia cai, dentro de cada noite,
há um poço
onde a claridade está presa.

Há que sentar-se na beira
do poço da sombra
e pescar luz caída
com paciência.

Pablo Neruda.

Resumo

A partir da década de 80 a TRI (Teoria da Resposta ao Item) passou a ser tópico de pesquisa dominante entre os especialistas em medidas e sua aplicação se dá em diversas áreas do conhecimento, entre elas a educacional. Neste trabalho, o objetivo, além de apresentar um resumo da teoria existente de análise de itens de uma prova, bem como das estimativas das habilidades usando a TRI, apresenta o resultado da análise de dados do SAEPE (Sistema de Avaliação Educacional de Pernambuco) de 2005, das provas de Português e Matemática da 3^a série do Ensino Médio de Pernambuco. Os resultados da análise dos dados revelam que em média, as habilidades em português e matemática são diferentes por sexo e por região de residência, e ainda que na prova realizada, foram identificados 38% dos itens de português e 80% dos itens de matemática como itens de elevado grau de dificuldade.

Palavras-chave:

TRI, Avaliação Educacional, Desempenho Escolar.

Abstract

Since 1980's, the Item Response theory (IRT) has been very important subject among measurement experts due to applications in several fields including education. In this work, an introduction of that theory is presented and applied to item analysis in high school exams. Using IRT, hability estimates were obtained from 2005 SAEPE data analysis of Portuguese and Mathematics exams applied to third grade high school students in the state of Pernambuco, Brazil. Data analysis results show on average, that habilities in Portuguese and Mathematics disciplines are different according to student sex and residence location. It was also revealed that 38% of Portuguese exam items and 80% of Mathematics exam items were quite difficult.

Keywords:

IRT (Item Response Theory), Educational Evaluation, School Performance

Índice

Lista de Figuras **xi**

Lista de Tabelas **xii**

1 Introdução **1**

1.1 Avaliação Educacional 1

1.2 Teoria da Resposta ao Item 3

1.3 Teoria Clássica dos Testes-TCT 5

1.4 Vantagens da TRI sobre a TCT 8

1.5 Objetivo do Trabalho 9

1.6 Organização do Trabalho 9

2 Modelos da resposta ao item **11**

2.1 Modelos para itens dicotômicos ou dicotomizados 11

2.2 Curva característica do item 13

2.3 Suposições 15

3 Métodos de estimação **17**

3.1 Estimação dos Parâmetros dos Itens 18

3.2 Estimação das Habilidades 25

3.3 Estimação conjunta dos parâmetros dos itens e habilidades 30
3.4 Máxima verossimilhança marginal 32
3.5 Métodos iterativos 37

4 Equalização **41**

4.1 Delineamento de grupos não-equivalentes com itens comuns 42
4.2 Diferentes tipos de equalização 42
4.2.1 Um único grupo fazendo uma única prova 42
4.2.2 Um único grupo fazendo duas provas totalmente distintas 43
4.2.3 Um único grupo fazendo duas provas parcialmente distintas 43
4.2.4 Dois grupos fazendo uma única prova 43
4.2.5 Dois grupos fazendo duas provas totalmente distintas 43
4.2.6 Dois grupos fazendo duas provas parcialmente distintas 43
4.3 Diferentes problemas de estimação 44
4.3.1 Quando todos os itens são novos 44
4.3.2 Quando todos os itens já estão calibrados 44
4.3.3 Quando alguns itens são novos e outros já estão calibrados 44
4.4 A escala de Habilidade 45
4.4.1 Construção e interpretação de escalas de habilidade 46
4.5 Equalização a posteriori 47

5 Aplicação **51**

5.1 Método Utilizado 52
5.1.1 Métodos para a Calibração dos Itens 54
5.1.2 Métodos Implementados para a Estimação das Habilidades 54
5.1.3 Métodos Utilizados para Comparar Habilidades 56
5.2 Resultados e Discussões 56
5.2.1 Análise das Estimativas dos Indicadores - SAEPE 2005 56
5.2.2 Análise dos Itens 57

5.2.3 Análise das Habilidades dos indivíduos 66

6 Conclusões e Sugestões de Trabalhos Futuros **74**

6.1 Sugestões de Trabalhos Futuros 75

Apêndice A **76**

Apêndice B **82**

Referências Bibliográficas **85**

Lista de Figuras

2.1 Curva Característica do Item 14

5.1 Curva Característica do item 56 de português (o mais difícil) 61
5.2 Curva Característica do item 13 de português (item bom) 62
5.3 Curva Característica do item 02 de português (o menos difícil) 63
5.4 Curva Característica do item 79 de matemática (o mais difícil) 64
5.5 Curva Característica do item 06 (o mais discriminante) 65
5.6 Curva Característica do item 61 de matemática (item bom) 66
5.7 Histograma das proficiências/habilidades de Português 67
5.8 Histograma das proficiências/habilidades de Matemática 67

Lista de Tabelas

5.1 Média e Desvio-Padrão das Estimativas dos Indicadores- SAEPE 2005 (Português) 57

5.2 Média e Desvio-Padrão das Estimativas dos Indicadores- SAEPE 2005 (Matemática) 57

5.3 Número de itens com problemas pela TCT 58

5.4 Número de itens que apresentam problemas com os parâmetros da TRI 59

5.5 Número de itens âncora para os níveis âncora da prova de Português e Matemática 60

5.6 Parâmetros de alguns dos itens de Português 60

5.7 Parâmetros de alguns itens de Matemática 63

5.8 Estatísticas descritivas das proficiências/habilidades de Português e Matemática . 66

5.9 Média e mediana das habilidades dos alunos por sexo de Português 68

5.10 Média e mediana das habilidades dos alunos por sexo de Matemática 68

5.11 Média e mediana das habilidades dos alunos da Região Metropolitana e Não-Metropolitana de Português 69

5.12 Média e mediana das habilidades dos alunos da Região Metropolitana e Não-metropolitana de Matemática 69

5.13 Número de escolas e médias das variáveis nos 5 grupos 71

5.14 Proporção das variáveis binárias nos 5 grupos 72

5.15 Médias e mediana das habilidades de português por grupo de escolas 72

5.16 Médias e medianas das habilidades de matemática por grupo de escolas 72

CAPÍTULO 1

Introdução

1.1 Avaliação Educacional

Até o início dos anos 90 o processo de avaliação educacional tinha como objetivo apenas obter resultados classificatórios.

Atualmente, a discussão sobre esse tema é fundamentada em medidas de conhecimento e habilidades cognitivas adquiridas pelos indivíduos e, no Brasil, usadas a partir do Sistema Nacional de Avaliação do Ensino Básico (SAEB) criado em 1990, do Exame Nacional do Ensino Médio (ENEM), implantado pelo Ministério da Educação em 1998, do Exame Nacional de Desempenho dos Estudantes (ENADE), criado em 2004 pelo MEC para substituir o antigo "provão", e ainda de outros exames de avaliação estaduais como o Sistema de Avaliação Educacional de Pernambuco (SAEPE), criado em 2000 com a proposta de ser bianual, onde o objetivo principal é a melhoria da qualidade da educação e dos cursos oferecidos aos estudantes.

O SAEB avalia a qualidade, a eqüidade e a eficiência do ensino e da aprendizagem, no âmbito do Ensino Fundamental e do Ensino Médio. Aplicado a cada dois anos, o SAEB também avalia o que os alunos sabem e são capazes de fazer em diversas situações de seu período escolar, levando em conta as condições existentes nas escolas brasileiras. Para isso, são utilizados instrumentos específicos como provas aplicadas a alunos de escolas selecionadas por amostragem em todas as unidades da Federação, nas quais é medido o desempenho acadêmico dos estudantes, e

questionários pelos quais são investigados os fatores intra e extra-escolares associados ao desempenho dos alunos. Assim, o SAEB é um instrumento essencial de apoio a todos que lidam com a educação em nosso país.

As avaliações criadas nos últimos anos pelo MEC, como o ENADE, conhecido até o ano de 2003 como "provão", foram colocadas para a sociedade com a finalidade de melhorar a qualidade dos cursos superiores oferecidos aos estudantes e também verificar as proficiências básicas dos concluintes dos cursos de graduações.

O Sistema de Avaliação Educacional de Pernambuco (SAEPE) tem como objetivo principal desenvolver um trabalho permanente de monitoria e de incentivos para a melhoria da qualidade do ensino e o consequente desempenho das escolas e foi aplicado nos anos 2002 e 2005.

Através da implementação do SAEPE tem-se procurado conhecer o que os alunos sabem e são capazes de fazer, em diversos momentos de seu percurso escolar visando a melhoria da qualidade, da eficiência e da eqüidade da educação básica. Com o SAEPE, está sendo possível disponibilizar para as escolas, os órgãos municipais, regionais e estaduais informações ou indicadores, como:

1. Qualidade do ensino ministrado e sua melhoria, indicada pelas médias de proficiência demonstrada pelos alunos nas provas e pela evolução deste desempenho;

2. eficiência da escola e sua melhoria, indicadas pelas taxas de promoção da escola e sua evolução;

3. capacidade da escola de diversificar as oportunidades de aprendizagem oferecidas aos alunos, indicada pela desistência de atividades sistemáticas de recuperação e de aceleração da aprendizagem e da extensão da jornada escolar e extra-escolar;

4. padrões de oferta educacional que a escola apresenta, tomando como referência equipamentos, ambientes escolares (sua situação e manutenção), nível de preparo e qualificação de seus recursos humanos, condições de acompanhar os avanços tecnológicos, etc.;

5. padrões de gestão escolar, que remetem a existência, à atuação de organismos colegiados na escola e à sua capacidade de gestão autônoma;

6. Integração no meio social, no que se refere à participação da comunidade extra-escolar na vida escolar.

1.2 Teoria da Resposta ao Item

A Teoria da Resposta ao Item (TRI) é uma teoria do traço latente (habilidade ou aptidão) aplicada a teste de habilidades ou de desempenho. Essa teoria se refere a uma família de modelos matemáticos cujo objetivo é relacionar variáveis observáveis (itens de um teste por exemplo) e traços hipotéticos não observáveis ou aptidões que são responsáveis pelo surgimento das variáveis observáveis, ou seja, das respostas ou comportamentos emitidos pelo sujeito que são as variáveis observáveis. A resposta que o indivíduo dá ao item depende do nível de habilidade que o indivíduo possui. Assim, a habilidade é a causa e a resposta do indivíduo é o efeito.

Quando estas relações são expressas numa equação matemática, constando de variáveis e constantes, tem-se um modelo ou teoria do traço latente. Então, se algumas das características das variáveis observadas (como os itens de um teste) são conhecidas, estas se tornam constantes na equação e, a partir desta nova equação é possível estimar o nível de desempenho do sujeito e vice-versa, isto é, se for conhecido o nível de habilidade, é possível estimar os parâmetros (características) dos itens respondidos por este indivíduo.

A TRI se baseia em dois postulados básicos: [Nojosa (2001)]

1. O desempenho do sujeito numa tarefa (item de um teste), que pode ser predito a partir de um conjunto de fatores ou variáveis hipotéticas, ditos aptidões ou traços latentes (identificados na TRI com a letra grega θ); o θ sendo a causa e, o desempenho o efeito. Ou seja, comportamento=função(traço latente);

2. a relação entre o desempenho e a habilidade que pode ser descrita por uma equação matemática monotônica crescente, chamada de *Curva Característica do Item*- CCI, definida adiante no capítulo 2.

A TRI originou-se entre os anos de 1935 e 1940. Entre os precursores da TRI encontram-se os trabalhos de Richardson (1936), comparando os parâmetros dos itens obtido pela Teoria Clássica

da Psicometria com os modelos que hoje usam a TRI; os trabalhos de Lawley & Richardson (1943), Lawley (1944), indicando alguns métodos para estimar os parâmetros dos itens, os quais se afastavam da Teoria Clássica; o trabalho de Tucker (1946), que parece ter sido o primeiro a utilizar a expressão *Curva Característica do Item*, que constitui um conceito chave na TRI e o trabalho de Lazersfeld (1950) que introduziu o conceito de traço latente, conceito que se constitui um parâmetro chave da nova TRI.

Porém, o responsável mais direto da TRI moderna é [Lord (1952)] por ter elaborado não apenas um modelo teórico como também métodos para estimar os parâmetros dos itens dentro da nova teoria, utilizando o modelo da ogiva normal. Os modelos elaborados por Lord se aplicam a testes onde as respostas são dicotômicas, isto é, certo ou errado. Depois, Samejima (1972) elaborou modelos para tratar respostas politômicas e mesmo para dados contínuos, como é o caso por exemplo de alguns testes de personalidade. Outra contribuição importante na história da TRI foi dada por Birnbaum (1957) que substituiu as curvas de ogiva por curvas logísticas, isto é, baseadas nos logaritmos, tornando o tratamento matemático e a interpretação dos dados mais fácil.

A partir das décadas de 1970 e 1980, a TRI passou a ser tópico de pesquisa dominante entre os especialistas em medidas. Como a complexidade matemática no campo da TRI é enorme, o progresso vertiginoso nas máquinas de processamento (microcomputadores) possibilitou a viabilização dos cálculos que o modelo TRI exige. Com este progresso foi possível, nos anos 80, o desenvolvimento de softwares apropriados para tais cálculos.

Hoje, a TRI tem sido incorporada em muitos softwares de análise de itens, como: BICAL [Wright et al. (1979.)], BILOG [Mislevy & Bock (1984.)], BILOG-MG [Zimowski et al. (1996)], MULTILOG [Thissen (1991)].

A TRI é uma técnica utilizada em várias áreas de conhecimento, como por exemplo, na área educacional [Andrade (1999)]; medicinal [DeRoos & Allen-Meares (1998)]; na área psicossocial [Granger & Deutsch (1998)], etc.

Mendoza et al. (2005), usaram a TRI na análise psicométrica dos itens que compõem o desenho da figura humana que é um dos instrumentos mais divulgados e utilizado na prática de

avaliação psicológica de pessoas.

Uma das grandes vantagens dessa técnica é que ela permite fazer comparações entre habilidades de indivíduos de populações diferentes quando são submetidos a testes que tenham alguns itens comuns, ou ainda, a comparação de indivíduos de mesma população submetidos a testes totalmente diferentes, uma vez que a TRI tem como elementos centrais os itens e não a prova como um todo [Valle (1999)], ou seja, ela surgiu como uma forma de considerar cada item individualmente, sem relevar os escores totais, daí as conclusões independem propriamente do teste, mas de cada item que o compõe [Andrade et al. (2000)].

Inúmeras aplicações da TRI têm sido exploradas nas últimas décadas, tais como: criação de banco de itens, avaliação adaptativa computadorizada, equalização de provas, avaliação de mudança cognitiva. Para maiores detalhes das principais aplicações, ver Lord (1980) e Wainer (1989).

1.3 Teoria Clássica dos Testes-TCT

O modelo clássico da psicometria tradicional [Pasquali (1997)] está fundamentado na teoria clássica dos testes, que considera os testes como um conjunto de estímulos comportamentais (itens) cuja qualidade é definida em termos de um critério, que por sua vez, é representado por comportamentos presentes ou futuros.

A Teoria Clássica dos Testes, estava bastante bem axiomatizada já nos anos 50, sobretudo com os trabalhos de Guilford (1936,1954) e Gulliksen (1950). Porém, continha o grave problema que Thurstone (1928,1959) pontuava antes dos anos 30, "Um instrumento de medida, na sua função de medir, não pode ser seriamente afetado pelo objeto de medida".

Embora Thurstone tenha percebido este problema, ele não conseguiu encontrar uma solução para o mesmo. Foi após os anos 50 que os psicometristas começaram a descobrir a solução para o problema, baseados na teoria do traço latente de Lazersfeld (1959), nos trabalhos de Lord (1952) e do dinamarquês Rasch (1960), os quais se tornaram as bases da moderna Teoria da Resposta ao Item.

No modelo clássico, dois construtos são introduzidos: o escore verdadeiro e o erro de medida.

O escore verdadeiro para um indivíduo pode ser definido com um valor esperado dos seus escores obtidos em vários testes, e o erro de medida pode ser definido como a diferença entre o escore verdadeiro e o observado.

Matematicamente tem-se o modelo

$$x = t + \epsilon,$$

onde x, t e ϵ, são, respectivamente, o escore observado, o escore verdadeiro e o erro de medida. As suposições para esse modelo são:

1. $E(\epsilon) = 0$;
2. $\rho(t, \epsilon) = 0$;
3. $\rho(\epsilon_1, \epsilon_2) = 0$,

onde ϵ_1 e ϵ_2 são os erros de medida em duas aplicações de um teste e ρ é o coeficiente de correlação.

Os principais índices calculados na Teoria Clássica para cada item são: índice de dificuldade, o bisserial e o bisserial para cada uma das alternativas (ponto bisserial). A seguir são apresentadas caracterizações gerais sobre esses índices.

Índice de dificuldade

O índice de dificuldade (I) representa a proporção de alunos que acertou o item. Quanto mais alunos acertam a um determinado item, mais fácil é esse item. Esse índice também é conhecido como índice de facilidade, pois quanto maior esse valor, mais fácil é o item. Ele varia de 0 (ninguém acertou o item) até 1 (todos os alunos acertaram o item). Em geral, testes que alcançam um índice médio de dificuldade em torno de 0,5 produzem distribuições de escores no teste com maior variação [Bloom et al. (1971); Vianna (1982); Pasquali (1997) e Mcintire & Miller (2000)]. Pode-se utilizar a seguinte sugestão para interpretação [Condé (2001)]:

- Item fácil: $I > 0,70$;

- item de dificuldade média: $0,30 < I \leq 0,70$;
- item difícil: $I \leq 0,30$.

Bisserial

O Bisserial é um índice de discriminação que indica a correlação entre o desempenho no item e o desempenho no teste como um todo. Espera-se de uma resposta a um item discriminativo que os estudantes que vão bem na prova como um todo, acertem-no, e por sua vez, aqueles que não vão bem, errem-no.

Quanto maior o coeficiente bisserial, maior a capacidade do item de discriminar grupos de estudantes que construíram determinada competência e habilidade, daqueles que não as construíram. Os itens com coeficiente baixo não diferenciam o indivíduo que construiu, daquele que não construiu determinada competência e habilidade. A correlação bisserial é menos influenciada pela dificuldade do item e tende a apresentar menos variação de uma situação de testagem para outra [Wilson et al. (1991)]. Sua fórmula é: [Rodrigues (2006)]

$$r_{bis} = \frac{M_i - M}{S} \times \frac{p_i}{h(p_i)}, \qquad \text{onde}$$

M_i= média no teste dos indivíduos que acertam o item (i);

M= média total do teste;

S= desvio padrão do teste;

p_i=proporção de indivíduos que acertam o item i;

$h(p_i)$= é a ordenada na curva normal no ponto de divisão dos segmentos que contém as proporções p dos casos.

Bisserial por cada uma das alternativas

Quando o cálculo do coeficiente Bisserial é efetuado para cada uma das alternativas, tem-se a correlação da opção de resposta do indivíduo ao item com o seu desempenho na prova como um todo. Assim, espera-se que os alunos que se desempenham bem na prova, tenham feito a opção pela alternativa correta de um determinado item. Caso esses alunos tenham sido atraídos

a responder qualquer uma das outras alternativas que não a certa, o item não é discriminativo e não consegue diferenciar os alunos que construíram, daqueles que não construíram determinada competência e habilidade. Esse índice é uma medida estatística capaz de identificar itens com formulação inadequada ou com erro de gabarito.

1.4 Vantagens da TRI sobre a TCT

A TRI se desenvolveu tendo como um dos objetivos suprir deficiências da Teoria Clássica. Embora a TRI não entre em contradição com os princípios da Teoria Clássica, ela traz uma nova proposta de análise centrada nos itens que supera as principais limitações da Teoria Clássica [Muniz (1994); Hambleton et al. (1978)], além de apresentar novos recursos tecnológicos para a avaliação [Nunes & Primi (2005)] .

Vale ressaltar que a TRI não veio para substituir toda a Teoria Clássica, mas apenas parte dela, particularmente na análise dos itens e no tema da fidedignidade da medida.

Hambleton et al. (1991) apresentam cinco grandes avanços que a TRI trouxe sobre a Teoria Clássica:

1. O cálculo do nível de habilidade do sujeito: Na Teoria Clássica, o escore do sujeito dependia e variava segundo o teste aplicado fosse mais fácil ou mais difícil, ou produzisse maiores ou menores erros. Assim, tais escores não eram comparáveis. Já na TRI, esse cálculo independe da amostra de itens utilizados, ou seja, a habilidade do sujeito é independente do teste.

2. O cálculo dos parâmetros dos itens (dificuldade e discriminação): Na Teoria Clássica, os parâmetros dos itens dependiam muito dos sujeitos amostrados possuirem maior ou menor habilidade. Já na TRI, esse cálculo independe da amostra de sujeitos utilizada, ou seja, os parâmetros dos itens são independentes dos sujeitos.

3. A TRI permite emparelhar itens com a habilidade do sujeito, ou seja, avalia a habilidade de um indivíduo, utilizando itens com dificuldade tal que se situam em torno do tamanho da habilidade do sujeito, sendo assim possível utilizar itens mais fáceis para sujeitos com

habilidades inferiores e itens mais difíceis para indivíduos mais aptos, produzindo escores comparáveis em ambos os casos. Já na Teoria Clássica, sempre é aplicado o mesmo teste para todos os sujeitos, de maneira que, se o teste fosse fácil, avaliaria bem sujeitos de habilidade menor e mal, indivíduos de habilidade superior e, se o teste fosse difícil, faria o contrário.

4. A TRI constitui um modelo que não precisa fazer suposições que aparentam serem improváveis, tais como os erros de medida serem iguais para todos os testandos, como faz a Teoria Clássica.

5. A TRI não precisa trabalhar com testes estritamente paralelos, que é um teste funcional para determinar se o processamento e os resultados de uma nova versão da aplicação são consistentes com o processamento e resultados da antiga versão da aplicação, como exige a Teoria clássica.

1.5 Objetivo do Trabalho

Este trabalho tem como objetivo, além de apresentar uma revisão da teoria existente de análise de itens de uma prova, bem como das estimativas das habilidades usando a TRI, apresentar o resultado da análise de dados do SAEPE 2005 das provas de Português e Matemática da 3^a série do Ensino Médio de Pernambuco.

1.6 Organização do Trabalho

Esse trabalho está estruturado em 6 capítulos, onde na introdução é abordado o tema de estudo, o objetivo e a organização do mesmo. O segundo capítulo trata dos modelos logísticos da TRI de 1, 2 e 3 parâmetros, o terceiro apresenta os métodos de estimação dos parâmetros dos itens e das habilidades para uma única população, o quarto capítulo, mostra os diferentes tipos de equalização, o quinto, apresenta uma breve descrição do método utilizado que é a aplicação do software BILOG, bem como os resultados das análises dos itens e das habilidades dos estudantes da 3^a série do ensino médio de Pernambuco. Já no sexto e último capítulo é apresentada a

conclusão final do trabalho.

A presente dissertação de mestrado foi digitada utilizando o sistema de tipografia LATEX, desenvolvido por Leslie Lamport em 1985, que consiste em uma série de macros ou rotinas do sistema TEX (criado por Donald Knuth na Universidade de Stanford) que facilitam o desenvolvimento da edição do texto. Todos os resultados numéricos e todos gráficos apresentados nesta dissertação de mestrado foram obtidos utilizando o BILOG e o SPSS 11.5.

CAPÍTULO 2

Modelos da resposta ao item

Os modelos existentes na TRI se distinguem na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos podem conter um ou mais parâmetros relacionados ao indivíduo. Para detalhes dos diversos modelos existentes ver [Van Der Linden & Hambleton (1997)] e [Andrade et al. (2000)]. Segundo Andrade et al. (2000), os modelos propostos na literatura dependem fundamentalmente de três fatores:

1. da natureza do item: dicotômicos ou não dicotômicos;
2. do número de populações envolvidas: apenas uma ou mais de uma;
3. e da quantidade de traços latentes que está sendo medida: apenas um (modelo unidimensional) ou mais de um (modelo multidimensional).

2.1 Modelos para itens dicotômicos ou dicotomizados

Existem basicamente três tipos de modelos logísticos para itens dicotômicos ou dicotomizados (itens com mais de duas categorias de resposta ou de resposta aberta, porém corrigidos como certo ou errado) os quais se diferem pelo número de parâmetros utilizados para descrever o item. São os modelos logísticos de 1, 2 e 3 parâmetros, que consideram, respectivamente:

- somente a dificuldade do item;

- a dificuldade e a discriminação;
- a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos de baixa habilidade.

O modelo logístico de 3 parâmetros é o mais completo sendo que os outros dois podem ser facilmente obtidos como casos particulares dele.

Dos modelos propostos pela TRI, o modelo logístico de três parâmetros (ML3), em geral é o mais utilizado e é dado por:

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

com $i = 1, 2, \ldots, m$ e $j = 1, 2, \ldots, n$, onde:

- m é o número de itens;
- n é o número de indivíduos;
- U_{ji} é uma variável dicotômica que assume o valor 1, quando o indivíduo j responde corretamente o item i e, o valor zero caso contrário.
- θ_j representa a habilidade (traço latente) do j-ésimo indivíduo.
- $P(U_{ji} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item - FRI. Pode também ser interpretada como a proporção de respostas corretas do item i dentre os indivíduos da população com habilidade θ_j.
- b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.
- a_i é o parâmetro de discriminação (ou de inclinação) do item i, com valor proporcional à inclinação (*slope*) da Curva Característica do Item - CCI no ponto b_i.

- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual).

- D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultados semelhantes ao da função ogiva normal, o qual foi o primeiro modelo da TRI onde se utilizava a função de distribuição acumulada da normal como a função de resposta ao item ou *curva característica do item* [Van Der Linden & Hambleton (1997)].

O Modelo Logístico de dois Parâmetros (ML2) é dado por:

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

com $i = 1, 2, \ldots, m$ e $j = 1, 2, \ldots, n$. Esse modelo é utilizado quando não existe possibilidade de resposta correta ao acaso, ou seja $c_i = 0$.

O Modelo Logístico de um parâmetro (ML1) é utilizado quando além de não existir possibilidade de resposta casual ($c_i = 0$), todos os itens tiverem o mesmo poder de discriminação, ou seja $a_i = 1$. Ele é dado por:

$$P(U_{ji} = 1|\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}},$$

com $i = 1, 2, \ldots, m$ e $j = 1, 2, \ldots, n$.

2.2 Curva característica do item

A curva característica do item (CCI), ilustrada na figura 2.1 adiante inserida, estabelece a relação entre a probabilidade de um indivíduo acertar um item com os valores da variável ou característica latente que está sendo medida pelo teste, de tal forma que quanto maior a habilidade do indivíduo maior a probabilidade dele acertar o item. Essa relação pode tomar diferentes formas dependendo dos parâmetros de discriminação (a), de dificuldade b e a probabilidade de

acerto casual c. Essas informações podem estar presentes nas equações, possibilitando uma maior caracterização do item.

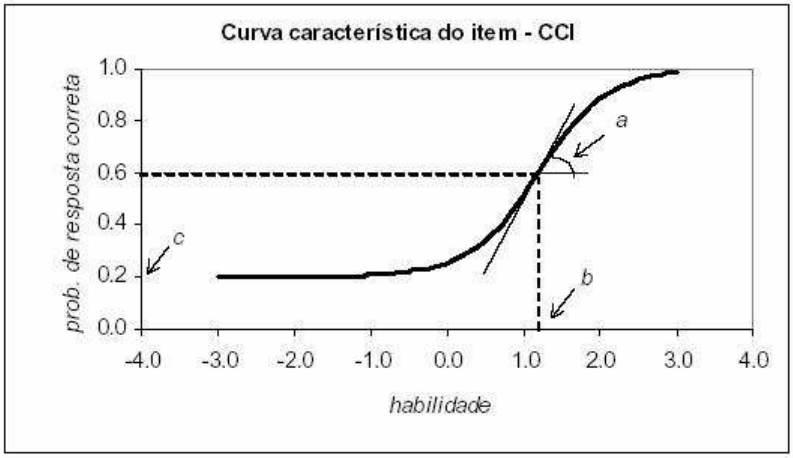


Figura 2.1: Curva Característica do Item

Os valores dos parâmetros a e b dependem da escala utilizada. Por exemplo, se as habilidades de um determinado grupo de respondentes têm média 0 e desvio padrão 1, os valores plausíveis para o parâmetro b variam entre -3,0 e 3,0. Valores próximos de 3,0 correspondem a itens que são muito difíceis e valores próximos a -3,0 correspondem a ítens muito fáceis para esse grupo. Para o parâmetro a, espera-se valores no intervalo $(0; 3)$. Valores próximos de zero indicam que o item tem pouco poder de discriminação (alunos com habilidades bastante diferentes tem aproximadamente a mesma probabilidade de responder corretamente ao item) e valores próximos de 3,0 indicam itens com curvas características muito íngremes que discriminam os indivíduos basicamente em dois grupos: os que possuem habilidade abaixo do valor do parâmetro b e os que possuem habilidade acima do valor do parâmetro b.

Na prática, os estudos vêm mostrando que utilizando a escala logística, os valores mais apropriados para o parâmetro de discriminação estão entre 0,9 e 2,7, já para a escala normal esse intervalo é $(0, 6; 1, 6)$ [Nojosa (2001)]. O parâmetro de dificuldade apresenta valores no intervalo $(-2, 0; 2, 0)$, independente da escala utilizada, logística ou normal. O parâmetro c depende, a princípio, do número de alternativas do item. Por exemplo, para um item com 5 alternativas,

espera-se valores entre 0,1 e 0,3.

O índice de discriminação a, refere-se a inclinação da CCI no ponto de inflexão, isto é, quando a curva corta a linha que corresponde a probabilidade de $(1 + c)/2$ de resposta correta. Quanto maior for a inclinação da curva, maior será o seu valor. Ele é proporcional ao coeficiente angular da reta tangente ao ponto de inclinação máxima, ou seja, onde a probabilidade de acerto for igual a $(1 + c)/2$. Dessa forma, itens com a negativo não são esperados para esse modelo, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da habilidade. Vale ressaltar que a capacidade de disciminação dos itens varia de acordo com o nível de habilidade avaliado (θ) [Nunes & Primi (2005)].

O parâmetro b é um parâmetro de locação que determina a posição na escala da habilidade onde ocorre o ponto de inflexão, ou seja, corresponde ao valor da habilidade em que a probabilidade de acerto for igual a $(1 + c)/2$, daí será $0,5$ se $c = 0$. A métrica teórica desse parâmetro varia de -∞ a $+\infty$, mas na prática varia de -3 a $+3$. Se $b = -3$, o item é extremamente fácil; zero, de dificuldade mediana ; e 3, extremamente difícil. Quanto maior for o valor de b, maior será o nível de habilidade exigida para que o indivíduo tenha a chance de 0,5 de acertar o item. [Pasquali & Primi (2003)]

O parâmetro c representa a probabilidade de um aluno com baixa habilidade responder corretamente o item e é muitas vezes referido como a probabilidade de acerto casual, ou seja, quando a probabilidade de acerto não depende da habilidade. Então quando não é permitido o acerto casual, c é igual a 0 e b representa o ponto na escala da habilidade onde a probabilidade de acertar o item é 0,5.

2.3 Suposições

Antes de referir as suposições necessárias para a modelagem, é importante garantir as seguintes suposições quanto a aplicação de um teste:

- o tempo para a resolução do teste é suficiente para que todos os itens possam ser respondidos por todos os indivíduos e,

- a ordem em que os itens são apresentados aos indivíduos não interfere no desempenho dos mesmos.

A TRI inclui um conjunto de pressupostos acerca dos dados para os quais o modelo será aplicado. Os modelos matemáticos empregados na TRI pressupõem que a probabilidade de um indivíduo responder a um determinado item corretamente depende de sua habilidade e das características do item. Os dois principais pressupostos são o da *unidimensionalidade* e o da *independência local*. A unidimensionalidade supõe que somente uma habilidade esteja sendo medida pelos itens que compõem um teste. A independência local está relacionada ao conceito de unidimensionalidade e pressupõe que as respostas dadas aos itens dependem somente da habilidade que está sendo medida e não de outras habilidades. Dessa forma, as respostas dos indivíduos para qualquer par de itens deverão ser estatisticamente independentes.

CAPÍTULO 3

Métodos de estimação

Os modelos da TRI ficarão completamente especificados a partir da estimação dos parâmetros dos itens e das habilidades dos respondentes. O processo de estimação dos parâmetros dos itens é chamado de calibração [Baker (1992)].

No processo de estimação podem acontecer os seguintes casos:

1. os parâmetros dos itens são conhecidos e precisa-se estimar as habilidades;
2. os parâmetros das habilidades dos respondentes são conhecidos e precisa-se estimar os parâmetros dos itens;
3. os parâmetros das habilidades e dos itens são estimados simultaneamente.

No primeiro caso a solução é dada empregando o método da máxima verossimilhança ou métodos bayesianos, ambos através da aplicação de procedimentos iterativos, como, por exemplo, o método de Newton-Raphson ou " Scoring" de Fisher. O segundo caso tem apenas caráter teórico e é solucionado usando o método da máxima verossimilhança. O terceiro caso, o mais encontrado na prática, pode ser resolvido de duas formas também usando o método da máxima verossimilhança: a estimação conjunta dos parâmetros dos itens e das habilidades dos indivíduos; ou em duas etapas, primeiro a estimação dos parâmetros dos itens e, em seguida, a estimação das habilidades.

A TRI nos possibilita analisar individualmente os itens de um teste fornecendo parâmetros referentes aos itens e parâmetros referentes aos indivíduos. Bem interpretados esses parâmetros fornecem uma grande quantidade de informação necessária para a análise dos itens, identificando itens por sua dificuldade e por seu poder de discriminação entre os indivíduos de maior ou menor nível de habilidade.

Informações mais detalhadas sobre a estimação dos parâmetros podem ser encontradas em [Baker (1987), Baker (1992), Lord (1980), Birnbaum (1968), Swaminathan & Gifford (1983), Hambleton & Swaminathan (1985), Hambleton et al. (1991), Hambleton & Van Der Linden (1996)].

3.1 Estimação dos Parâmetros dos Itens

Seja $U_{j.} = (U_{j1}, U_{j2}, \ldots, U_{jm})$ o vetor aleatório de respostas do indivíduo j, seja também $U_{..} = (U_{1.}, \ldots, U_{n.})$ o conjunto total de respostas. De maneira análoga, as observações serão representadas por $u_{ji}, u_{j.}$ e $u_{..}$. Segue ainda que $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ é o vetor de habilidades dos n indivíduos e $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_m)$ é o vetor cujos elementos $\zeta_i = (a_i, b_i, c_i)$ são os vetores de parâmetros do item i. Como o modelo ML3 é o mais completo, a escolha dele leva a resultados que servirão para os outros dois modelos (ML1 e ML2). Então, usando as suposições de Unidimensionalidade e de Independência Local, podemos escrever a função de verossimilhança da seguinte forma:

$$\begin{aligned} L(\zeta) &= \prod_{j=1}^{n} P(U_{j.} = u_{j.}|\theta_j, \zeta) \\ &= \prod_{j=1}^{n}\prod_{i=1}^{m} P(U_{ji} = u_{ji}|\theta_j, \zeta_i), \end{aligned} \tag{3.1}$$

onde em (3.1) usa-se que a distribuição de U_{ji} só depende de ζ através de ζ_i. Mas,

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

onde: $P_{ji} = P(U_{ji} = 1|\theta_j, \zeta_i)$ e $Q_{ji} = 1 - P_{ji} = P(U_{ji} = 0|\theta_j, \zeta_i)$.

Assim,

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

e daí segue que a log-verossimilhança é escrita da seguinte maneira:

$$\texttt{log}L(\zeta) = \sum_{j=1}^{n}\sum_{i=1}^{m}[u_{ji}\texttt{log}P_{ji} + (1-u_{ji})\texttt{log}Q_{ji}]$$

Os estimadores de Máxima Verossimilhança (EMV) para os parâmetros dos m itens são determinados quando ζ_i satisfaz a equação

$$\frac{\partial \texttt{log}L(\zeta)}{\partial \zeta_i} = 0, \qquad i = 1, 2, \ldots, m.$$

mas,

$$\begin{aligned}
\frac{\partial \text{log}L(\zeta)}{\partial \zeta_i} &= \sum_{j=1}^{n}\left\{u_{ji}\frac{\partial(\text{log}P_{ji})}{\partial\zeta_i} + (1-u_{ji})\frac{\partial(\text{log}Q_{ji})}{\partial\zeta_i}\right\} \\
&= \sum_{j=1}^{n}\left\{u_{ji}\frac{1}{P_{ji}}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right) - (1-u_{ji})\frac{1}{Q_{ji}}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right)\right\} \\
&= \sum_{j=1}^{n}\left\{u_{ji}\frac{1}{P_{ji}} - (1-u_{ji})\frac{1}{Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right) \\
&= \sum_{j=1}^{n}\left\{\frac{u_{ji}Q_{ji} - P_{ji} + u_{ji}P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right) \\
&= \sum_{j=1}^{n}\left\{\frac{u_{ji}(Q_{ji}+P_{ji})}{P_{ji}Q_{ji}} - \frac{P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right) \\
&= \sum_{j=1}^{n}\left\{\frac{u_{ji}(1-P_{ji}+P_{ji}) - P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right) \\
&= \sum_{j=1}^{n}\left\{\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right) \qquad (3.2)
\end{aligned}$$

Definindo: $W_{ji} = \frac{P^*_{ji}Q^*_{ji}}{P_{ji}Q_{ji}}$, onde $P^*_{ji} = \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1}$ e $Q^*_{ji} = 1 - P^*_{ji}$.

tem-se,

$$\frac{\partial \text{log}L(\zeta)}{\partial\zeta_i} = \sum_{j=1}^{n}\left\{(u_{ji}-P_{ji})\frac{W_{ji}}{P^*_{ji}Q^*_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right).$$

Para obter as equações de estimação será preciso as seguintes expressões:

$$\frac{\partial P_{ji}}{\partial a_i} = D(1-c_i)(\theta_j - b_i)P^*_{ji}Q^*_{ji}, \qquad (3.3)$$

$$\frac{\partial P_{ji}}{\partial b_i} = -Da_i(1-c_i)P^*_{ji}Q^*_{ji}, \qquad (3.4)$$

$$\frac{\partial P_{ji}}{\partial c_i} = Q^*_{ji}. \tag{3.5}$$

Para o parâmetro de discriminação, tem-se que:

$$\begin{aligned}
\frac{\partial \log L(\zeta)}{\partial a_i} &= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji}}{\partial a_i} \right) \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) D(1 - c_i)(\theta_j - b_i) P^*_{ji} Q^*_{ji} \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= D(1 - c_i) \sum_{j=1}^{n} (u_{ji} - P_{ji})(\theta_j - b_i) W_{ji}.
\end{aligned}$$

Para o parâmetro de dificuldade, tem-se que:

$$\begin{aligned}
\frac{\partial \log L(\zeta)}{\partial b_i} &= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji}}{\partial b_i} \right) \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= \sum_{j=1}^{n} \left\{ (u_{ji} - P_{ji})(-1) D a_i (1 - c_i) P^*_{ji} Q^*_{ji} \frac{W_{ji}}{P^*_{ji} Q^*_{ji}} \right\} \\
&= -D a_i (1 - c_i) \sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji}.
\end{aligned}$$

Para o parâmetro de acerto casual, tem-se que:

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

Assim sendo, as equações de estimação para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i : \quad D(1 - c_i) \textstyle\sum_{j=1}^{n} (u_{ji} - P_{ji})(\theta_j - b_i) W_{ji} = 0, \tag{3.6}$$

$$b_i : \quad -D a_i (1 - c_i) \textstyle\sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji} = 0, \tag{3.7}$$

$$c_i : \quad \textstyle\sum_{j=1}^{n} (u_{ji} - P_{ji}) \frac{W_{ji}}{P^*_{ji}} = 0. \tag{3.8}$$

Estas equações não possuem solução explícita, então é necessário aplicar algum método iterativo (Newton Raphson ou Escore de Fisher), para a obtenção das estimativas de máxima verossimilhança.

Aplicação do algoritmo de Newton-Raphson para a estimação dos parâmetros dos itens

Segundo Andrade et al. (2000), seja $l(\zeta) = \log L(\zeta)$ a log-verossimilhança, onde $\zeta = (\zeta_1, \ldots, \zeta_m)$, com $\zeta_i = (a_i, b_i, c_i)'$. Se os valores iniciais $\widehat{\zeta}_i^{(0)} = (a_i^{(0)}, b_i^{(0)}, c_i^{(0)})'$ podem ser encontrados para ζ_i, então uma estimativa atualizada será $\widehat{\zeta}_i^{(1)} = \widehat{\zeta}_i^{(0)} + \Delta\widehat{\zeta}_i^{(0)}$, ou seja,

$$\hat{a}_i^{(1)} = \hat{a}_i^{(0)} + \Delta\hat{a}_i^{(0)},$$

$$\hat{b}_i^{(1)} = \hat{b}_i^{(0)} + \Delta\hat{b}_i^{(0)},$$

$$\hat{c}_i^{(1)} = \hat{c}_i^{(0)} + \Delta\hat{c}_i^{(0)},$$

onde $\Delta\hat{a}_i^{(0)} = a_i - \hat{a}_i^{(0)}$, $\Delta\hat{b}_i^{(0)} = b_i - \hat{b}_i^{(0)}$ e $\Delta\hat{c}_i^{(0)} = c_i - \hat{c}_i^{(0)}$ são erros de aproximação. Usando a expansão em série de Taylor de $\partial l(\zeta)/\partial\zeta_i$ em torno de $\widehat{\zeta}_i^{(0)}$, teremos:

$$\frac{\partial l(\zeta)}{\partial a_i} = \frac{\partial l(\widehat{\zeta}_i^{(0)})}{\partial a_i} + \Delta\hat{a}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial a_i^2} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial a_i \partial b_i} + \Delta\hat{c}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial a_i \partial c_i} + R_{a_i}(\widehat{\zeta}_i^{(0)}),$$

$$\frac{\partial l(\zeta)}{\partial b_i} = \frac{\partial l(\widehat{\zeta}_i^{(0)})}{\partial b_i} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial b_i^2} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial b_i \partial a_i} + \Delta\hat{c}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial b_i \partial c_i} + R_{b_i}(\widehat{\zeta}_i^{(0)}),$$

$$\frac{\partial l(\zeta)}{\partial c_i} = \frac{\partial l(\widehat{\zeta}_i^{(0)})}{\partial c_i} + \Delta\hat{a}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial c_i^2} + \Delta\hat{b}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial c_i \partial a_i} + \Delta\hat{c}_i^{(0)}\frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial c_i \partial b_i} + R_{c_i}(\widehat{\zeta}_i^{(0)}),$$

onde $\partial l(\widehat{\zeta}_i)/\partial\alpha_i$ representa a função $\partial l(\zeta_i)/\alpha_i$ avaliada no ponto $\zeta_i = \widehat{\zeta}_i$. Nessas expressões sabe-se que $\partial l(\zeta)/\partial\zeta_i$ é função apenas de ζ_i, não dependendo de ζ_l para $l \neq i$. Por isso, pode-se representá-la de forma simplificada por $\partial l(\zeta_i)/\partial\zeta_i$. Fazendo

$$\frac{\partial l(\zeta_i)}{\partial a_i} = \frac{\partial l(\zeta_i)}{\partial b_i} = \frac{\partial l(\zeta_i)}{\partial c_i} = 0,$$

usando a notação

$$L_1 = \frac{\partial l(\widehat{\zeta}_i^{(0)})}{\partial a_i} \qquad L_{11} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial a_i^2} \qquad L_{12} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial a_i \partial b_i} \qquad L_{13} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial a_i \partial c_i},$$

$$L_2 = \frac{\partial l(\widehat{\zeta}_i^{(0)})}{\partial b_i} \qquad L_{21} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial b_i \partial a_i} \qquad L_{22} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial b_i^2} \qquad L_{23} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial b_i \partial c_i},$$

$$L_3 = \frac{\partial l(\widehat{\zeta}_i^{(0)})}{\partial c_i} \qquad L_{31} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial c_i \partial a_i} \qquad L_{32} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial c_i \partial b_i} \qquad L_{33} = \frac{\partial^2 l(\widehat{\zeta}_i^{(0)})}{\partial c_i^2},$$

e desprezando os restos $R_{a_i}(\widehat{\zeta}_i^{(0)}), R_{b_i}(\widehat{\zeta}_i^{(0)})$ e $R_{c_i}(\widehat{\zeta}_i^{(0)})$, teremos

$$0 = L_1 + L_{11}\Delta\hat{a}_i^{(0)} + L_{12}\Delta\hat{b}_i^{(0)} + L_{13}\Delta\hat{c}_i^{(0)},$$

$$0 = L_2 + L_{12}\Delta\hat{a}_i^{(0)} + L_{22}\Delta\hat{b}_i^{(0)} + L_{23}\Delta\hat{c}_i^{(0)},$$

$$0 = L_3 + L_{13}\Delta\hat{a}_i^{(0)} + L_{23}\Delta\hat{b}_i^{(0)} + L_{33}\Delta\hat{c}_i^{(0)}.$$

colocando o resultado em forma matricial, tem-se

$$-\begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} \Delta\hat{a}_i^{(0)} \\ \Delta\hat{b}_i^{(0)} \\ \Delta\hat{c}_i^{(0)} \end{pmatrix}.$$

Resolvendo o sistema para $\Delta\widehat{\zeta}_i^{(0)}$, tem-se

$$\begin{pmatrix} \Delta\hat{a}_i^{(0)} \\ \Delta\hat{b}_i^{(0)} \\ \Delta\hat{c}_i^{(0)} \end{pmatrix} = -\begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix}^{-1} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix},$$

e finalmente,

$$\begin{pmatrix} \hat{a}_i^{(1)} \\ \hat{b}_i^{(1)} \\ \hat{c}_i^{(1)} \end{pmatrix} = \begin{pmatrix} \hat{a}_i^{(0)} \\ \hat{b}_i^{(0)} \\ \hat{c}_i^{(0)} \end{pmatrix} - \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix}^{-1} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix}.$$

Após obtido $\widehat{\zeta}_i^{(1)}$, este é considerado um novo ponto inicial para a obtenção de $\widehat{\zeta}_i^{(2)}$, e assim por diante. Este processo é repetido até que algum critério de parada seja alcançado. Por exemplo, até que $\Delta\widehat{\zeta}_i^{(t)} = \widehat{\zeta}_i^{(t)} - \widehat{\zeta}_i^{(t-1)}$ seja suficientemente pequeno ou que um número pré-definido, t_{max}, de iterações seja cumprido.

As expressões L_k, para $k = 1, 2, 3$ são dadas por:

$$\begin{aligned}
\frac{\partial \log L(\zeta)}{\partial a_i} &= D(1-c_i)\sum_{j=1}^{n}(u_{ji} - P_{ji})(\theta_j - b_i)W_{ji} \\
\frac{\partial \log L(\zeta)}{\partial b_i} &= -Da_i(1-c_i)\sum_{j=1}^{n}(u_{ji} - P_{ji})W_{ji}. \\
\frac{\partial \log L(\zeta)}{\partial c_i} &= \sum_{j=1}^{n}\left\{(u_{ji} - P_{ji})\frac{W_{ji}}{P_{ji}^*}\right\}.
\end{aligned}$$

e as expressões L_{kl}, $\quad k, l = 1, 2, 3$, são obtidas de

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i \partial \zeta_i'} = \sum_{j=1}^{n}\left\{\left[\frac{\partial}{\partial \zeta_i}\left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}}\right)\right]\left(\frac{\partial P_{ji}}{\partial \zeta_i}\right)' + \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}}\right)\left(\frac{\partial^2 P_{ji}}{\partial \zeta_i \partial \zeta_i'}\right)\right\}$$

$$= \sum_{j=1}^{n} \left\{ \left[\frac{\partial \upsilon ji}{\partial \zeta_i} \right] \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right)' + \upsilon_{ji} \left(\frac{\partial^2 P_{ji}}{\partial \zeta_i \partial \zeta_i'} \right) \right\}$$

onde

$$\upsilon_{ji} = \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}}$$

e

$$\begin{aligned}
\frac{\partial \upsilon_{ji}}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) = \\
&= \frac{1}{(P_{ji}Q_{ji})^2} \left\{ -P_{ji}Q_{ji} \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) - (u_{ji} - P_{ji}) \left(\frac{\partial P_{ji} Q_{ji}}{\partial \zeta_i} \right) \right\} \\
&= \frac{-1}{(P_{ji}Q_{ji})^2} \left\{ P_{ji}Q_{ji} \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) + (u_{ji} - P_{ji}) \left[\left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) - 2P_{ji} \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) \right] \right\} \\
&= \frac{-1}{(P_{ji}Q_{ji})^2} \{ P_{ji}Q_{ji} + (u_{ji} - P_{ji})(1 - 2P_{ji}) \} \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) \\
&= \frac{-1}{(P_{ji}Q_{ji})^2} (u_{ji} - P_{ji})^2 \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) \\
&= -\upsilon_{ji}^2 \left(\frac{\partial P_{ji}}{\partial \zeta_i} \right) \qquad (3.9)
\end{aligned}$$

A igualdade (3.9) segue do fato que $u_{ji} = u_{ji}^2$. Considerando $\widehat{\zeta}_i^{(t)}$ a estimativa de ζ_i na iteração t, então na iteração $t+1$ do algoritmo Newton-Raphson tem-se que

$$\widehat{\zeta}_i^{(t+1)} = \widehat{\zeta}_i^{(t)} - \left[\boldsymbol{H}(\widehat{\zeta}_i^{(t)}) \right]^{-1} \boldsymbol{h}(\widehat{\zeta}_i^{(t)}).$$

onde,

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

e

$$\begin{aligned}
\boldsymbol{H}(\zeta_i) &\equiv \frac{\partial \log L(\zeta)}{\partial \zeta_i \partial \zeta_i'} \\
&= \sum_{j=1}^{n} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right) (P_{ji}^* Q_{ji}^*) \boldsymbol{H}_{ji} - \left(\frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right)^2 (P_{ji}^* Q_{ji}^*)^2 \boldsymbol{h}_{ji} \boldsymbol{h}_{ji}' \right\} \\
&= \sum_{j=1}^{n} (u_{ji} - P_{ji}) W_{ji} \left\{ \boldsymbol{H}_{ji} - (u_{ji} - P_{ji}) W_{ji} \boldsymbol{h}_{ji} \boldsymbol{h}_{ji}' \right\}.
\end{aligned}$$

com

$$\boldsymbol{h}_{ji} = (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial P_{ji}}{\partial \zeta_i}\right) = \begin{pmatrix} D(1-c_i)(\theta_j - b_i) \\ -Da_i(1-c_i) \\ \frac{1}{P^*_{ji}} \end{pmatrix},$$

e

$$\begin{aligned} \boldsymbol{H}_{ji} &= (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial^2 P_{ji}}{\partial\zeta_i\partial\zeta_i'}\right) \\ &= \begin{pmatrix} D^2(1-c_i)(\theta_j-b_i)^2(1-2P^*_{ji}) & . & . \\ -D(1-c_i)\left\{1+Da_i(\theta_j-b_i)(1-2P^*_{ji})\right\} & D^2a_i^2(1-c_i)(1-2P^*_{ji}) & . \\ -D(\theta_j-b_i) & Da_i & 0 \end{pmatrix} \end{aligned}$$

Aplicação do método "Scoring" de Fisher

Para aplicar esse método deve-se substituir os componentes da matriz de derivadas segundas usadas no processo iterativo de Newton-Raphson pelos seus valores esperados. Note que a variável U_{ji} só pode assumir os valores: 1, com probabilidade P_{ji} e 0 com probabilidade Q_{ji}, então U_{ji} tem distribuição *Bernoulli* (P_{ji}). Assim, $E(U_{ji}) = P_{ji}$ e $E(U_{ji} - P_{ji})^2 = Var(U_{ji}) = P_{ji}Q_{ji}$. Logo de $\boldsymbol{H}(\zeta_i) = \sum_{j=1}^{n}(u_{ji} - P_{ji})W_{ji}\left\{\boldsymbol{H}_{ji} - (u_{ji} - P_{ji})W_{ji}\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\right\}$, tem-se que

$$\begin{aligned} \Delta(\zeta_i) &\equiv E(\boldsymbol{H}(\zeta_i)) \\ &= \sum_{j=1}^{n}\{E(U_{ji}-P_{ji})W_{ji}\boldsymbol{H}_{ji} - E(U_{ji}-P_{ji})^2W_{ji}^2\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\} \\ &= \sum_{j=1}^{n}\{-P_{ji}Q_{ji}W_{ji}^2\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\} \\ &= -\sum_{j=1}^{n}\{P^*_{ji}Q^*_{ji}W_{ji}\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\} \end{aligned}$$

Erro-padrão

Os estimadores de máxima verossimilhança gozam de propriedades assintóticas, como vício nulo e eficiência. Sob algumas condições de regularidade, a distribuição assintótica do estimador de máxima verossimilhança, $\widehat{\zeta_i}$, é normal com vetor de média ζ_i e matriz de covariâncias dada pela inversa da matriz de informação

$$I(\zeta_i) = -E\left(\frac{\partial^2 \log L(\zeta)}{\partial\zeta_i\partial\zeta_i'}\right) = -\Delta(\zeta_i),$$

onde $\Delta(\zeta_i) = -\sum_{j=1}^{n}\{P_{ji}^{*}Q_{ji}^{*}W_{ji}\boldsymbol{h}_{ji}\boldsymbol{h}_{ji}'\}$. As raízes quadradas dos elementos diagonais de $[I(\zeta_i)]^{-1}$ fornecem os erros-padrão dos estimadores $\widehat{a_i}, \widehat{b_i}$ e $\widehat{c_i}$.

Escore nulo ou perfeito

Alguns problemas ocorrem na estimação por máxima verossimilhança. Se o item i é respondido incorretamente por todos os indivíduos, ou seja, $u_{ji} = 0$, $j = 1, \cdots, n$, então $L(\zeta) = \prod_{j=1}^{n}\prod_{i=1}^{m} P_{ji}^{u_{ji}}Q_{ji}^{1-u_{ji}}$, resume-se a $L(\zeta) = \prod_{j=1}^{n} Q_{ji}$. Considerando os valores $_i, c_i$, e θ_j fixos, tem-se que mudanças no valor de b_i apenas transladam Q_{ji}, sem alterar seus valores máximo e mínimo. Assim, fixando a_i, c_i, $i = 1, \cdots, m$, $\quad b_l, l \neq i$ e θ_j, o valor que maximiza a verossimilhança será $b_i = -\infty$. Por outro lado, se o item i é respondido corretamente por todos os indivíduos, isto é, $u_{ji} = 1$, então $L(\zeta) = \prod_{j=1}^{n}\prod_{i=1}^{m} P_{ji}^{u_{ji}}Q_{ji}^{1-u_{ji}}$, resume-se a $L(\zeta) = \prod_{j=1}^{n} P_{ji}$. Com o mesmo argumento anterior, tem-se que o estimador de máxima verossimilhança será $b_i = +\infty$. Problemas similares a esse ocorrem com os parâmetros a_i e c_i [Andrade et al. (2000)].

3.2 Estimação das Habilidades

A estimação das habilidades de um número pequeno de indivíduos é mais confiável se forem utilizados itens já calibrados, sendo que esta calibração dos itens deve ser feita com um número grande de indivíduos. Pela independência entre as respostas de diferentes indivíduos e a independência local, pode-se escrever a log-verossimilhança como:

$$\mathtt{log}L(\zeta) = \sum_{j=1}^{n}\sum_{i=1}^{m}[u_{ji}\mathtt{log}P_{ji} + (1-u_{ji})\mathtt{log}Q_{ji}],$$

agora como função de θ e não de ζ, ou seja:

$$\mathtt{log}L(\theta) = \sum_{j=1}^{n}\sum_{i=1}^{m}[u_{ji}\mathtt{log}P_{ji} + (1-u_{ji})\mathtt{log}Q_{ji}]$$

O EMV de θ_j é o valor que maximiza a verossimilhança, ou seja, que soluciona a equação:

$$\frac{\partial \mathrm{log}L(\theta)}{\partial \theta_j} = 0, \quad j = 1, \ldots, n.$$

segue que

$$\frac{\partial \mathrm{log}L(\theta)}{\partial \theta_j} = \sum_{i=1}^{m}\left\{u_{ji}\frac{\partial(\mathrm{log}P_{ji})}{\partial \theta_j} + (1-u_{ji})\frac{\partial(\mathrm{log}Q_{ji})}{\partial \theta_j}\right\}$$

$$
\begin{aligned}
&= \sum_{i=1}^{m}\left\{u_{ji}\frac{1}{P_{ji}}\left(\frac{\partial P_{ji}}{\partial\theta_j}\right)-(1-u_{ji})\frac{1}{Q_{ji}}\left(\frac{\partial P_{ji}}{\partial\theta_j}\right)\right\}\\
&= \sum_{i=1}^{m}\left\{u_{ji}\frac{1}{P_{ji}}-(1-u_{ji})\frac{1}{Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\theta_j}\right)\\
&= \sum_{i=1}^{m}\left\{\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\theta_j}\right)
\end{aligned}
$$

e fazendo a ponderação

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

temos que:

$$\frac{\partial \log L(\theta)}{\partial\theta_j}=\sum_{i=1}^{m}\left\{(u_{ji}-P_{ji})\frac{W_{ji}}{P^*_{ji}Q^*_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\theta_j}\right).$$

mas,

$$\frac{\partial P_{ji}}{\partial\theta_j}=Da_i(1-c_i)P^*_{ji}Q^*_{ji}.$$

então,

$$
\begin{aligned}
\frac{\partial \log L(\theta)}{\partial\theta_j} &= \sum_{i=1}^{m}\left\{(u_{ji}-P_{ji})Da_i(1-c_i)P^*_{ji}Q^*_{ji}\frac{W_{ji}}{P^*_{ji}Q^*_{ji}}\right\}\\
&= D\sum_{i=1}^{m}a_i(1-c_i)(u_{ji}-P_{ji})W_{ji}.
\end{aligned}
$$

Segue então que a equação de estimação $\frac{\partial \log L(\theta)}{\partial\theta_j}=0$, para θ_j, $j=1,\ldots,n$, é:

$$\theta_j : D\sum_{i=1}^{m}a_i(1-c_i)(u_{ji}-P_{ji})W_{ji} = 0. \tag{3.10}$$

Como esta equação não apresenta solução explícita para θ_j, é preciso utilizar algum método iterativo para obter as estimativas desejadas. (Newton-Raphson ou "Scoring" de Fisher)

Aplicação do algoritmo de Newton-Raphson para a estimação das habilidades

De maneira similar ao que foi feito na Seção 3.1, e considerando $\widehat{\theta_j}^{(t)}$ a estimativa de θ_j na iteração t, então na iteração $t+1$ do algoritmo de Newton-Raphson tem-se que:

$$\widehat{\theta_j}^{(t+1)}=\widehat{\theta_j}^{(t)}-\left[H(\widehat{\theta_j}^{(t)})\right]^{-1}h(\widehat{\theta_j}^{(t)})$$

onde,

$$h(\theta_j)\equiv\frac{\partial \log L(\theta)}{\partial\theta_j}=\sum_{i=1}^{m}\left\{\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right\}\left(\frac{\partial P_{ji}}{\partial\theta_j}\right)$$

e

$$
\begin{aligned}
H(\theta_j) &\equiv \frac{\partial^2 \log L(\theta)}{\partial \theta_j^2} \\
&= \frac{\partial}{\partial \theta_j}\left\{\sum_{i=1}^{m}\left[\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\frac{\partial P_{ji}}{\partial \theta_j}\right]\right\} \\
&= \sum_{i=1}^{m}\left\{\frac{\partial}{\partial \theta_j}\left[\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\frac{\partial P_{ji}}{\partial \theta_j}\right]\right\} \\
&= \sum_{i=1}^{m}\left\{\left[\frac{\partial}{\partial \theta_j}\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\right]\left(\frac{\partial P_{ji}}{\partial \theta_j}\right)+\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2}\right)\right\}
\end{aligned}
$$

mas,

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

substituindo em $H(\theta_j)$, temos:

$$
\begin{aligned}
H(\theta_j) &= \sum_{i=1}^{m}\left\{\frac{\partial}{\partial \theta_j}\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\frac{\partial P_{ji}}{\partial \theta_j}+\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\frac{\partial^2 P_{ji}}{\partial \theta_j^2}\right\} \\
&= \sum_{i=1}^{m}\left\{-\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)^2\frac{\partial P_{ji}}{\partial \theta_j}\frac{\partial P_{ji}}{\partial \theta_j}+\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)\frac{\partial^2 P_{ji}}{\partial \theta_j^2}\right\}
\end{aligned}
$$

$$= \sum_{i=1}^{m} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right) \left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2} \right) - \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 \left(\frac{\partial P_{ji}}{\partial \theta_j} \right)^2 \right\} \tag{3.11}$$

como

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

então,

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

porém,

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

então,

$$\begin{aligned}
\frac{\partial^2 P_{ji}}{\partial \theta_j^2} &= Da_i(1 - c_i) \left[(Q_{ji}^* - P_{ji})^* \frac{\partial P_{ji}^*}{\partial \theta_j} \right] \\
&= Da_i(1 - c_i) \left[(1 - 2P_{ji}^*) Da_i P_{ji}^* Q_{ji}^* \right] \\
&= D^2 a_i^2 (1 - c_i)(1 - 2P_{ji}^*) P_{ji}^* Q_{ji}^*
\end{aligned}$$

Agora, voltando para (3.11), temos:

$$\begin{aligned}
H(\theta_j) &= \sum_{i=1}^{m} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right) \left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2} \right) - \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 \left(\frac{\partial P_{ji}}{\partial \theta_j} \right)^2 \right\} \\
&= \sum_{i=1}^{m} \left\{ \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right) D^2 a_i^2 (1 - c_i)(1 - 2P_{ji}^*) P_{ji}^* Q_{ji}^* - \left(\frac{u_{ji} - P_{ji}}{P_{ji}Q_{ji}} \right)^2 (Da_i(1 - c_i)P_{ji}^* Q_{ji}^*)^2 \right\}
\end{aligned}$$

fazendo,

$$h_{ji} = (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial P_{ji}}{\partial \theta_j}\right) = Da_i(1-c_i),$$

$$H_{ji} = (P^*_{ji}Q^*_{ji})^{-1}\left(\frac{\partial^2 P_{ji}}{\partial \theta_j^2}\right) = D^2a_i^2(1-c_i)(1-2P^*_{ji})$$

tem-se que,

$$h(\theta_j) \equiv \frac{\partial \log L(\theta)}{\partial \theta_j} = \sum_{i=1}^{m}\left\{(u_{ji}-P_{ji})\frac{W_{ji}}{P^*_{ji}Q^*_{ji}}\right\}(P^*_{ji}Q^*_{ji})h_{ji} = \sum_{i=1}^{m}(u_{ji}-P_{ji})W_{ji}h_{ji}.$$

e

$$\begin{aligned} H(\theta_j) \equiv \frac{\partial^2 \log L(\theta)}{\partial \theta_j^2} &= \sum_{i=1}^{m}\left\{\left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)(P^*_{ji}Q^*_{ji})H_{ji} - \left(\frac{u_{ji}-P_{ji}}{P_{ji}Q_{ji}}\right)^2(P^*_{ji}Q^*_{ji})^2h_{ji}^2\right\} \\ &= \sum_{i=1}^{m}(u_{ji}-P_{ji})W_{ji}\{H_{ji}-(u_{ji}-P_{ji})W_{ji}h_{ji}^2\} \end{aligned}$$

Aplicação do método "Scoring" de Fisher

Para a aplicação desse método, deve-se substituir os componentes da matriz de derivadas segundas usadas no processo iterativo de Newton-Raphson pelos seus valores esperados. Assim,

$$\begin{aligned} \Delta(\theta_j) &\equiv E(H(\theta_j)) \\ &= \sum_{i=1}^{m}\{E(U_{ji}-P_{ji})W_{ji}H_{ji} - E(U_{ji}-P_{ji})^2W_{ji}^2h_{ji}^2\} \\ &= -\sum_{i=1}^{m}P^*_{ji}Q^*_{ji}W_{ji}h_{ji}^2. \end{aligned}$$

Então, neste caso, a expressão para estimativa de θ_j, $j = 1, \cdots, n$, na iteração $t+1$ será

$$\widehat{\theta}_j^{\,t+1} = \widehat{\theta}_j^{\,t} - [\Delta(\widehat{\theta}_j^{\,t})]^{-1}h(\widehat{\theta}_j^{\,t}).$$

Erro-padrão

Sob algumas condições de regularidaade, a distribuição assintótica do estimador de máxima verossimilhança, $\widehat{\theta_j}$, é normal com média θ_j e variância dada pela inversa da matriz de informação

$$I(\theta_j) = -E\left(\frac{\partial \log L(\theta)}{\partial \theta_j^2}\right) = -\Delta(\theta_j),$$

onde $\Delta(\theta_j)$ é obtida de $\Delta(\theta_j) = -\sum_{i=1}^{m}P^*_{ji}Q^*_{ji}W_{ji}h_{ji}^2$. A raiz quadrada de $I(\theta_j)$ fornece o erro-padrão de $\widehat{\theta_j}$.

Escore nulo ou perfeito

Assim como na estimação dos parâmetros dos itens, existe um problema a ser contornado na estimação por máxima verossimilhança. Se o indivíduo j obtém escore nulo, isto é, $u_{ji} = 0$, $i = 1, \cdots, m$, então a verossimilhança resume-se a $L(\theta) = \prod_{i=1}^{m} Q_{ji}$. Como Q_{ji}, $i = 1, \cdots, m$, é decrescente com θ_j, então $L(\theta)$ também é decrescente com θ_j e assim o estimador de máxima verossimilhança será $\theta_j = -\infty$. Por outro lado, se o indivíduo j obtiver o escore total, ou seja, $u_{ji} = 1$, $i = 1, \cdots, m$, então a verossimilhança resume-se a $L(\theta) = \prod_{i=1}^{m} P_{ji}$. Como P_{ji}, $i = 1, \cdots, m$, é crescente com θ_j, então $L(\theta)$ também é crescente com θ_j e assim o estimador de máxima verossimilhança será $\theta_j = +\infty$.

3.3 Estimação conjunta dos parâmetros dos itens e habilidades

Esta Seção tratará do caso mais comum em que os parâmetros dos itens e das habilidades são desconhecidos. As equações a serem utilizadas foram apresentadas nas equações: (3.6), (3.7), (3.8) e (3.10) e são as seguintes:

$$\begin{aligned}
a_i: &\quad D(1-c_i)\sum_{j=1}^{n}(u_{ji}-P_{ji})(\theta_j - b_i)W_{ji} = 0,\\
b_i: &\quad -Da_i(1-c_i)\sum_{j=1}^{n}(u_{ji}-P_{ji})W_{ji} = 0,\\
c_i: &\quad \sum_{j=1}^{n}(u_{ji}-P_{ji})\frac{W_{ji}}{P^*_{ji}} = 0,\\
\theta_j: &\quad D\sum_{i=1}^{m}a_i(1-c_i)(u_{ji}-P_{ji})W_{ji} = 0
\end{aligned}$$

Porém, estas equações não possuem soluções explícitas para os respectivos EMV. Daí, algum processo iterativo deve ser aplicado no processo de maximização e consequentemente, algumas suposições devem ser adicionadas ao modelo.

Ao contrário das Seções anteriores, onde a métrica (unidade de medida) foi estabelecida pelo conhecimento de um dos conjuntos de parâmetros (habilidades ou itens), na estimação conjunta nenhum desses parâmetros é conhecido e, portanto não há métrica definida. É preciso estabelecer essa métrica devido a um problema denominado *falta de identificabilidade do modelo*. Isso ocorre

porque mais de um conjunto de parâmetros produz o mesmo valor no ML3, e consequentemente, na verossimilhança. Como exemplo, sejam: $\theta_j^* = \alpha\theta_j + \omega$, $b_i^* = \alpha b_i + \omega$, $a_i^* = a_i/\alpha$ e $c_i^* = ci$, onde α e ω são constantes reais com $\alpha > 0$, daí

$$\begin{aligned} P(U_{ji} = 1|\theta_j^*, \zeta_i^*) &= c_i^* + (1 - c_i^*)\left\{1 + \exp\left[-Da_i^*(\theta_j^* - b_i^*)\right]\right\}^{-1} \\ &= c_i + (1 - c_i)\left\{1 + \exp\left[-D\frac{a_i}{\alpha}(\alpha\theta_j + \omega - (\alpha b_i + \omega))\right]\right\}^{-1} \\ &= c_i + (1 - c_i)\left\{1 + \exp\left[-Da_i(\theta_j - b_i)\right]\right\}^{-1} \\ &= P(U_{ji} = 1|\theta_j, \zeta_i). \end{aligned}$$

Essa não-identificabilidade está relacionada com as características da população em estudo. Este problema pode ser solucionado pela especificação de uma medida de posição (média) e outra de dispersão (desvio padrão) para as habilidades. Dessa maneira obtém-se uma métrica para as habilidades e, consequentemente, para os parâmetros do itens. Ou seja, as habilidades e os parâmetros dos itens são estimados na métrica (μ, σ). Em várias situações [Andrade et al. (2000)] adota-se $\mu = 0$ e $\sigma = 1$.

Para aplicar o algoritmo de Newton-Raphson são necessárias as derivadas segundas da log-verossimilhança, com relação a ζ_i e $\theta_j, i = 1, \ldots, m$ e $j = 1, \ldots, n$. Estas derivadas compõem uma matriz $\boldsymbol{H}$, quadrada, de ordem (3m+n) sendo que essa dimensão pode ser suficientemente grande de forma a causar uma enorme exigência computacional. Daí, é preciso explorar um pouco mais a estrutura de $\boldsymbol{H}$.

Algumas considerações na modelagem contribuem para simplificar a estrutura dessa matriz. São elas:

- independência local, implicando

$$\frac{\partial^2 \log L(\zeta, \theta)}{\partial\zeta_i \partial\zeta_l'} = 0, \quad i \neq l.$$

- independência entre as respostas de indivíduos diferentes, implicando

$$\frac{\partial^2 \log L(\zeta, \theta)}{\partial\theta_j \partial\theta_l'} = 0, \quad j \neq l.$$

- independência entre habilidades e itens, implicando

$$\frac{\partial^2 \log L(\zeta, \theta)}{\partial \zeta_i \partial \theta_j'} = 0.$$

Dessa forma, a matriz $\boldsymbol{H}$ torna-se uma matriz bloco-diagonal, na qual os m primeiros blocos são matrizes de ordem 3×3 referentes aos três parâmetros dos itens e os n blocos seguintes são escalares referentes às habilidades dos n indivíduos. Ainda que a matriz $\boldsymbol{H}$ usada no processo iterativo de Newton-Raphson seja simplificada pelas considerações acima, sua dimensão não é alterada. Todavia, com base nessa estrutura bloco-diagonal, Birnbaum (1968) propôs um algoritmo em que os itens e as habilidades são estimados individualmente, utilizando o algoritmo Newton-Raphson ou o método Score de Fisher, no qual cada iteração é composta de dois estágios:

Estágio 1: Começando com estimativas iniciais para as habilidades θ_j (escores padronizados, por exemplo) e tratando estas habilidades como conhecidas, estima-se $\zeta_i, \quad i = 1, \ldots, m$.

Estágio 2: Começando com estimativas iniciais (obtidas no estágio 1) para ζ e tratando estes parâmetros como conhecidos, estima-se as habilidades $\theta_j, \quad j = 1, \ldots, n$

No estágio 1, os itens são estimados empregando o desenvolvimento da Seção 3.1. No estágio 2 as habilidades são estimadas com a teoria desenvolvida na Seção 3.2. Este processo é repetido até a convergência das habilidades e dos parâmetros dos itens.

3.4 Máxima verossimilhança marginal

Esse método proposto por Bock & Lieberman (1970) apresenta algumas vantagens em relação ao método da Máxima Verossimilhança Conjunta. Nesse método a estimação é feita em duas etapas: primeiro os parâmetros dos itens e, depois as habilidades. Como as habilidades não são conhecidas, é preciso usar algum artifício de forma que a verossimilhança não seja mais função das habilidades. Segundo Andersen (1980) ao considerar uma população Π composta por n indivíduos com habilidades θ_j, $j = 1, \ldots, n$, e ao construir a distribuição de frequência acumulada $G(\theta) = (\text{número de indivíduo} : \theta_j \leq \theta)/n$, então, se n for suficientemente grande os θ_j estarão bastante próximos de forma que $G(\theta)$ pode ser aproximada por uma distribuição contínua. A densidade $g(\theta)$, relativa à $G(\theta)$, pode realmente ser considerada a função densidade

para θ no experimento de retirar um indivíduo ao acaso da população Π e observar seu parâmetro θ. Vale ressaltar que, quando atribuímos uma distribuição de probabilidade para θ, não estamos aplicando nenhum argumento bayesiano. A distribuição de θ realmente existe, nesse sentido, como a densidade relativa à distribuição $G(\theta)$.

Assim um artifício para eliminar as habilidades na verossimilhança consiste em marginalizar a verossimilhança integrando-a com relação à distribuição de habilidade. De forma geral, considera-se que as habilidades, θ_j, $j = 1, \ldots, n$, são realizações de uma variável aleatória θ com distribuição contínua e função densidade de probabilidade $g(\theta|\eta)$ duplamente diferenciável, com as componentes de η conhecidas e finitas. Para o caso em que θ tem distribuição normal, tem-se $\eta = (\mu, \sigma^2)$, onde μ é a média e σ^2 é a variância das habilidades dos indivíduos de Π. Portanto, para que os itens sejam estimados na métrica (0,1), deve-se adotar $\mu = 0$ e $\sigma = 1$.

Segundo Bock & Lieberman (1970), a probabilidade marginal de $U_{j.}$ é dada por

$$\begin{aligned} P(u_{j.}|\zeta, \eta) &= \int_{\mathbb{R}} P(u_{j.}|\theta, \zeta, \eta) g(\theta|\eta) d\theta \\ &= \int_{\mathbb{R}} P(u_{j.}|\theta, \zeta) g(\theta|\eta) d\theta, \end{aligned} \tag{3.12}$$

onde em (3.12) usou-se que a distribuição de $U_{j.}$ não é função de η e R representa o conjunto dos números reais. Como as respostas de diferentes indivíduos são independentes, pode-se escrever a probabilidade associada ao vetor de respostas $U..$ como

$$P(u_{..}|\zeta, \eta) = \prod_{j=1}^{n} P(u_{j.}|\zeta, \eta).$$

Ainda que a verossimilhança possa ser escrita dessa forma, é muito comum utilizar a abordagem de *Padrões de resposta*. Como tem-se m itens no total, com 2 possíveis respostas para cada item (0 ou 1), há $S = 2^m$ possíveis respostas (padrões de resposta). Quando o número de indivíduos é grande com relação ao número de itens, pode haver vantagens computacionais em trabalhar com o número de ocorrências dos diferentes padrões de resposta. Daí o índice j não mais representará um indivíduo, mas um padrão de resposta. Seja r_j o número de ocorrências distintas do padrão de resposta j, e ainda $S \leq min(n, S)$ o número de padrões de resposta com

$r_j > 0$. Daí, segue que

$$\sum_{j=1}^{s} r_j = n.$$

Pela independência entre as respostas dos diferentes indivíduos, tem-se que os dados seguem uma distribuição *multinomial*, isto é,

$$L(\zeta, \eta) = \frac{n!}{\prod_{j=1}^{s} r_j!} \prod_{j=1}^{s} [P(u_{j.}|\zeta, \eta)]^{r_j},$$

e portanto, a log-verossimilhança é

$$\log L(\zeta, \eta) = \log \left\{ \frac{n!}{\prod_{j=1}^{s} r_j!} \right\} + \sum_{j=1}^{s} r_j \log P(u_{j.}|\zeta, \eta).$$

As equações de estimação para os itens são dadas por

$$\frac{\partial \log L(\zeta, \eta)}{\partial \zeta_i} = 0, \quad i = 1, \ldots, m,$$

onde

$$\begin{aligned} \frac{\partial \log L(\zeta, \eta)}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \left\{ \sum_{j=1}^{s} r_j \log P(u_{j.}|\zeta, \eta) \right\} \\ &= \sum_{j=1}^{s} r_j \frac{1}{P(u_{j.}|\zeta, \eta)} \frac{\partial P(u_{j.}|\zeta, \eta)}{\partial \zeta_i}. \end{aligned}$$

Mas,

$$\begin{aligned} \frac{\partial P(u_{j.}|\zeta, \eta)}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i} \int_{I\!R} P(u_{j.}|\theta, \zeta) g(\theta|\eta) d\theta \\ &= \int_{I\!R} \left(\frac{\partial}{\partial \zeta_i} P(u_{j.}|\theta, \zeta) \right) g(\theta|\eta) d\theta \qquad (*) \\ &= \int_{I\!R} \left(\frac{\partial}{\partial \zeta_i} \prod_{j=1}^{m} P(u_{jl}|\theta, \zeta_l) \right) g(\theta|\eta) d\theta \end{aligned}$$

$$\begin{aligned} \frac{\partial P(u_{j.}|\zeta, \eta)}{\partial \zeta_i} &= \int_{I\!R} \left(\prod_{l \neq i}^{m} P(u_{jl}|\theta, \zeta_l) \right) \left(\frac{\partial}{\partial \zeta_i} P(u_{ji}|\theta, \zeta_i) \right) g(\theta|\eta) d\theta \\ &= \int_{I\!R} \left(\frac{\partial P(u_{ji}|\theta, \zeta_i)/\partial \zeta_i}{P(u_{ji}|\zeta_i)} \right) P(u_{j.}|\theta, \zeta) g(\theta|\eta) d\theta, \end{aligned}$$

onde a ordem da derivada e da integral em $(*)$ pôde ser permutada com base no Teorema da Convergência Dominada de Lebesgue [Chow & Teicher (1978.)]. Sabe-se que

$$\begin{aligned}\frac{\partial}{\partial\zeta_i}P(u_{ji}|\theta,\zeta_i) &= \frac{\partial}{\partial\zeta_i}\left(P_i^{u_{ji}}Q_i^{1-u_{ji}}\right)\\ &= u_{ji}P_i^{u_{ji}-1}\left(\frac{\partial P_i}{\partial\zeta_i}\right)Q_i^{1-u_{ji}} + (1-u_{ji})Q_i^{-u_{ji}}\left(-\frac{\partial}{\partial\zeta_i}P_i\right)P_i^{u_{ji}}\\ &= \left(u_{ji}P_i^{u_{ji}-1}Q_i^{1-u_{ji}} - (1-u_{ji})Q_i^{-u_{ji}}P_i^{u_{ji}}\right)\frac{\partial P_i}{\partial\zeta_i}. \qquad (3.13)\end{aligned}$$

Observe que o termo entre parênteses da equação (3.13) vale 1 quando $u_{ji}=1$ e vale -1 quando $u_{ji}=0$, então pode-se reescrevê-lo como $(-1)^{u_{ji}+1}$. Daí,

$$\frac{\partial}{\partial\zeta_i}P(u_{ji}|\theta,\zeta_i) = (-1)^{u_{ji}+1}\left(\frac{\partial P_i}{\partial\zeta_i}\right).$$

mas,

$$\begin{aligned}\frac{\partial P(u_{j.}|\zeta,\eta)}{\partial\zeta_i} &= \int_{I\!R}\left(\frac{\partial P(u_{ji}|\theta,\zeta_i)/\partial\zeta_i}{P(u_{ji}|\zeta_i)}\right)P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta\\ &= \int_{I\!R}\frac{(-1)^{u_{ji}+1}}{P(u_{ji}|\zeta_i)}\left(\frac{\partial P_i}{\partial\zeta_i}\right)P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta\\ &= \int_{I\!R}\frac{(-1)^{u_{ji}+1}}{P_i^{u_{ji}}Q_i^{1-u_{ji}}}\left(\frac{\partial P_i}{\partial\zeta_i}\right)P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta\end{aligned}$$

Observe agora que

$$\frac{(-1)^{u_{ji}+1}P_iQ_i}{P_i^{u_{ji}}Q_i^{1-u_{ji}}} = \left\{\begin{array}{cc} Q_i & \text{se } u_{ji}=1;\\ -P_i & \text{se } u_{ji}=0\end{array}\right\}$$

então, este termo pode ser escrito como $u_{ji}-P_i$. E, consequentemente,

$$\frac{(-1)^{u_{ji}+1}}{P_i^{u_{ji}}Q_i^{1-u_{ji}}} = \frac{u_{ji}-P_i}{P_iQ_i}$$

Então,

$$\frac{\partial P(u_{j.}|\zeta,\eta)}{\partial\zeta_i} = \int_{I\!R}\left[\frac{(u_{ji}-P_i)}{P_iQ_i}\left(\frac{\partial P_i}{\partial\zeta_i}\right)\right]P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta$$

Seja agora $W_i = \frac{P_i^*Q_i^*}{P_iQ_i}$, onde

$$P_i^* = \{1+e^{-Da_i(\theta-b_i)}\}^{-1} \qquad e \qquad Q_i^* = 1-P_i^*.$$

Assim,

$$\frac{\partial P(u_{j.}|\zeta,\eta)}{\partial\zeta_i} = \int_{I\!R}\left[(u_{ji}-P_i)\left(\frac{\partial P_i}{\partial\zeta_i}\right)\frac{W_i}{P_i^*Q_i^*}\right]P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta$$

Usando a notação

$$g_j^*(\theta) \equiv g(\theta|u_{j.},\zeta,\eta) = \frac{P(u_{j.}|\theta,\zeta)g(\theta|\eta)}{P(u_{j.}|\zeta,\eta)},$$

a função de verossimilhança pode ser escrita como

$$\frac{\partial \log L(\zeta,\eta)}{\partial \zeta_i} = \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta.$$

As expressões para as derivadas de P_i são dadas nas equações (3.3), (3.4) e (3.5) com P_{ji}, Q_{ji}, P_{ji}^* e Q_{ji}^* substituídas por P_i, Q_i, P_i^* e Q_i^*, respectivamente. Assim, para o parâmetro de discriminação (a_i), tem-se que

$$\begin{aligned}
\frac{\partial \log L(\zeta,\eta)}{\partial\, a_i} &= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial\, a_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) D(1-c_i)(\theta - b_i) P_i^* Q_i^* \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\
&= D(1-c_i) \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i)(\theta - b_i) W_i \right] g_j^*(\theta) d\theta.
\end{aligned}$$

Para o parâmetro de dificuldade (b_i), tem-se que

$$\begin{aligned}
\frac{\partial \log L(\zeta,\eta)}{\partial\, b_i} &= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial\, b_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i)(-1) D a_i (1-c_i) P_i^* Q_i^* \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\
&= -D a_i (1-c_i) \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) W_i \right] g_j^*(\theta) d\theta.
\end{aligned}$$

Para o parâmetro de acerto ao acaso (c_i), tem-se que

$$\begin{aligned}
\frac{\partial \log L(\zeta,\eta)}{\partial\, c_i} &= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial\, c_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) Q_i^* \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \\
&= \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \frac{W_i}{P_i^*} \right] g_j^*(\theta) d\theta
\end{aligned}$$

Resumindo, as equações de estimação para os parâmetros a_i, b_i e c_i são, respectivamente,

$$a_i: \quad D(1-c_i) \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i)(\theta - b_i) W_i \right] g_j^*(\theta) d\theta = 0, \tag{3.14}$$

$$b_i: \quad -D a_i (1-c_i) \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) W_i \right] g_j^*(\theta) d\theta = 0, \tag{3.15}$$

$$c_i: \quad \sum_{j=1}^{s} r_j \int_{I\!R} \left[(u_{ji} - P_i) \frac{W_i}{P_i^*} \right] g_j^*(\theta) d\theta \quad = \quad 0 \tag{3.16}$$

as quais não possuem solução explícita.

3.5 Métodos iterativos

No desenvolvimentos das expressões para a estimação de ζ_i na estimação dos parâmetros dos itens, a propriedade de independência local foi suficiente para garantir que os parâmetros dos itens pudessem ser estimados individualmente, pois a derivada segunda de $\log L(\zeta)$ com relação a ζ_i e ζ_l, para $l \neq i$, era nula. Porém, na estimação por Máxima Verossimilhança Marginal isso não acontece, levando à necessidade da estimação dos m itens conjuntamente. As expressões para as derivadas segundas são obtidas a partir de:

$\frac{\partial^2 \log L(\zeta,\eta)}{\partial \zeta_l \partial \zeta_i'} = \frac{\partial}{\partial \zeta_l} \left[\frac{\partial \log L(\zeta,\eta)}{\partial \zeta_i} \right]'$

$$= \quad \frac{\partial}{\partial \zeta_l} \left[\sum_{j=1}^{s} r_j \frac{1}{P(u_{j.}|\zeta,\eta)} \frac{\partial P(u_{j.}|\zeta,\eta)}{\partial \zeta_i} \right]'$$

$$= \quad \sum_{j=1}^{s} r_j \left\{ \frac{\partial^2 P(u_{j.}|\zeta,\eta)/(\partial \zeta_l \partial \zeta_i')}{P(u_{j.}|\zeta,\eta)} - \left(\frac{\partial P(u_{j.}|\zeta,\eta)/(\partial \zeta_l}{P(u_{j.}|\zeta,\eta)} \right) \left(\frac{\partial P(u_{j.}|\zeta,\eta)/(\partial \zeta_i)}{P(u_{j.}|\zeta,\eta)} \right)' \right\}$$

para $i, l = 1, \ldots, m$. Considerando $\widehat{\zeta}^{(t)}$ a estimativa de ζ na iteração t, então na iteração $t+1$ tem-se que

$$\widehat{\zeta}^{(t+1)} = \widehat{\zeta}^{(t)} - [H_{PI}(\widehat{\zeta}^{(t)})]^{-1} h_{PI}(\widehat{\zeta}^{(t)})$$

Para chegar às expressões de $H_{PI}(\widehat{\zeta})$ e $h_{PI}(\zeta_i)$, adota-se a seguinte notação:

$$\mathrm{v}_{ji} = (u_{ji} - P_i) \frac{W_i}{P_i^* Q_i^*} = \frac{(u_{ji} - P_i)}{P_i Q_i}$$

Sabe-se que,

$$\frac{\partial P(u_{j.}|\zeta,\eta)}{\partial \zeta_i} = \int_{I\!R} \left[(u_{ji} - P_i) \left(\frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] P(u_{j.}|\theta,\zeta) g(\theta|\eta) d\theta$$

fazendo,

$$h_{i(j)} \equiv \frac{\partial P(u_{j.}|\zeta,\eta)/\partial \zeta_i}{P(u_{j.}|\zeta,\eta)} = \int_{I\!R} \left[\mathrm{v}_{ji} \left(\frac{\partial P_i}{\partial \zeta_i} \right) \right] g_j^*(\theta) d\theta$$

e

$$\frac{\partial^2 P(u_{j.}|\zeta,\eta)}{\partial\zeta_i\partial\zeta_i'} = \frac{\partial}{\partial\zeta_i}\left\{\int_{I\!R}\left[\mathrm{v}_{ji}\left(\frac{\partial P_i}{\partial\zeta_i}\right)'\right]P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta\right\}$$

Utilizando

$\frac{\partial P(u_{j.}|\theta,\zeta)}{\partial\zeta_i} = \left[\mathrm{v}_{ji}\left(\frac{\partial P_i}{\partial\zeta_i}\right)\right]P(u_{j.}|\theta,\zeta)$ e o desenvolvimento na aplicação do Algoritmo de Newton Raphson, que resultou em:

$$\frac{\partial \mathrm{v}_{ji}}{\partial\zeta_i} = -\mathrm{v}_{ji}^2\left(\frac{\partial P_{ji}}{\partial\zeta_i}\right), \quad \text{onde} \quad \mathrm{v}_{ji} = \frac{(u_{ji}-P_i)}{P_iQ_i}$$

tem-se que:

$$\begin{aligned}
\frac{\partial^2 P(u_{j.}|\theta,\zeta)}{\partial\zeta_i\partial\zeta_i'} &= \frac{\partial}{\partial\zeta_i}\left\{\left[\mathrm{v}_{ji}\left(\frac{\partial P_i}{\partial\zeta_i}\right)'\right]P(u_{j.}|\theta,\zeta)\right\} \\
&= \frac{\partial}{\partial\zeta_i}\left[\mathrm{v}_{ji}\left(\frac{\partial P_i}{\partial\zeta_i}\right)'\right]P(u_{j.}|\theta,\zeta) + \mathrm{v}_{ji}\left(\frac{\partial P(u_{j.}|\theta,\zeta)}{\partial\zeta_i}\right)\left(\frac{\partial P_i}{\partial\zeta_i}\right)' \\
&= \left[-\mathrm{v}_{ji}^2\left(\frac{\partial P_i}{\partial\zeta_i}\right)\left(\frac{\partial P_i}{\partial P_i}\right)' + \mathrm{v}_{ji}\left(\frac{\partial^2 P_i}{\partial\zeta_i\partial\zeta_i'}\right)\right]P(u_{j.}|\theta,\zeta) + \mathrm{v}_{ji}^2 P(u_{j.}|\theta,\zeta)\left(\frac{\partial P_i}{\partial\zeta_i}\right)\left(\frac{\partial P_i}{\partial\zeta_i}\right)' \\
&= \mathrm{v}_{ji}\left(\frac{\partial^2 P_i}{\partial\zeta_i\partial\zeta_i'}\right)P(u_{j.}|\theta,\zeta)
\end{aligned}$$

Para $l = i$, tem-se que

$$\frac{\partial^2 P(u_{j.}|\zeta,\eta)}{\partial\zeta_i\partial\zeta_i'} = \int_{I\!R}\mathrm{v}_{ji}\left(\frac{\partial^2 P_i}{\partial\zeta_i\partial\zeta_i'}\right)P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta$$

Assim, a primeira parcela em

$$\frac{\partial^2 \log L(\zeta,\eta)}{\partial\zeta_l\partial\zeta_i'} = \sum_{j=1}^{s} r_j\left\{\frac{\partial^2 P(u_{j.}|\zeta,\eta)/(\partial\zeta_l\partial\zeta_i')}{P(u_{j.}|\zeta,\eta)} - \left(\frac{\partial P(u_{j.}|\zeta,\eta)/\partial\zeta_l}{P(u_{j.}|\zeta,\eta)}\right)\left(\frac{\partial P(u_{j.}|\zeta,\eta)/(\partial\zeta_i)}{P(u_{j.}|\zeta,\eta)}\right)'\right\}$$

pode ser escrita para $i = l$, como

$$H_{ii(j)} \equiv \frac{\partial^2 P(u_{j.}|\zeta,\eta)/(\partial\zeta_i\partial\zeta_i')}{P(u_{j.}|\zeta,\eta)} = \int_{I\!R}\mathrm{v}_{ji}\left(\frac{\partial^2 P_i}{\partial\zeta_i\partial\zeta_i'}\right)g_j^*(\theta)d\theta$$

Para $l \neq i$, tem-se que

$$\begin{aligned}
\frac{\partial^2 P(u_{j.}|\zeta,\eta)}{\partial\zeta_l\partial\zeta_i'} &= \frac{\partial}{\partial\zeta_l}\left\{\int_{I\!R}\left[\mathrm{v}_{ji}\left(\frac{\partial P_i}{\partial\zeta_i}\right)'\right]P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta\right\} \\
&= \int_{I\!R}\mathrm{v}_{ji}\left(\frac{\partial P(u_{j.}|\theta,\zeta)}{\partial\zeta_l}\right)\left(\frac{\partial P_i}{\partial\zeta_i}\right)'g(\theta|\eta)d\theta \\
&= \int_{I\!R}\mathrm{v}_{ji}\mathrm{v}_{jl}\left(\frac{\partial P_l}{\partial\zeta_l}\right)\left(\frac{\partial P_i}{\partial\zeta_i}\right)'P(u_{j.}|\theta,\zeta)g(\theta|\eta)d\theta
\end{aligned}$$

Logo, para $l \neq i$,

$$H_{il(j)} \equiv \frac{\partial^2 P(u_{j.}|\zeta,\eta)/(\partial\zeta_l \partial\zeta_i')}{P(u_{j.}|\zeta,\eta)} = \int_{I\!R} \mathrm{v}_{ji}\mathrm{v}_{jl} \left(\frac{\partial P_l}{\partial \zeta_l}\right)\left(\frac{\partial P_i}{\partial \zeta_i}\right)' g_j^*(\theta) d\theta$$

Pode-se agora obter as equações de estimação para ζ. Com as expressões de Newton Raphson:

$$\frac{\partial P_{ji}^* Q_{ji}^*}{\partial \alpha_i} = (1 - 2P_{ji}^*)\frac{\partial P_{ji}^*}{\partial \alpha_i}, \quad \alpha_i \in \{a_i, b_i, c_i\},$$

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

De $\frac{\partial^2 P_i}{(\partial\zeta_i \partial\zeta_i')}, \quad i = 1, \ldots, m$

$$h_{PI}(\zeta) = \begin{pmatrix} h(\zeta_1) \\ \vdots \\ h(\zeta_m) \end{pmatrix} \quad e \quad H_{PI}(\zeta) = \begin{pmatrix} H(\zeta_1,\zeta_1) & \ldots & H(\zeta_1,\zeta_m) \\ \vdots & \ddots & \vdots \\ H(\zeta_m,\zeta_1) & \ldots & H(\zeta_m,\zeta_m) \end{pmatrix}$$

Sejam

$$h_i = (P_i^* Q_i^*)^{-1}\left(\frac{\partial P_i}{\partial \zeta_i}\right) = \begin{pmatrix} D(1-c_i)(\theta - b_i) \\ -Da_i(1-c_i) \\ \frac{1}{P_i^*} \end{pmatrix},$$

$$H_{ii} = (P_i^* Q_i^*)^{-1}\left(\frac{\partial^2 P_i}{\partial\zeta_i \partial\zeta_i'}\right) = \begin{pmatrix} D^2(1-c_i)(\theta-b_i)^2(1-2P_i^*) & . & . \\ -D(1-c_i)\{1 + Da_i(\theta-b_i)(1-2P_i^*)\} & D^2 a_i^2(1-c_i)(1-2P_i^*) & . \\ -D(\theta - b_i) & Da_i & 0 \end{pmatrix}$$

e para $i \neq l$,

$$H_{il} = h_i h_l' = (P_i^* Q_i^*)^{-1}(P_l^* Q_l^*)^{-1}\left(\frac{\partial P_i}{\partial \zeta_i}\right)\left(\frac{\partial P_l}{\partial \zeta_l}\right)'$$

$$H_{il} = \begin{pmatrix} D^2(1-c_i)(\theta-c_l)(\theta-b_i)(\theta-b_l) & -D^2 a_l(1-c_i)(1-c_l)(\theta-b_i) & D(1-c_i)(\theta-b_i)/P_l^* \\ -D^2 a_l(1-c_i)(1-c_l)(\theta-b_l) & D^2 a_i a_l(1-c_i)(1-c_l) & -Da_i(1-c_i)/P_l^* \\ -D(1-c_l)(\theta-b_l)/P_i^* & -Da_l(1-c_l)/P_i^* & [P_i^* P_l^*]^{-1} \end{pmatrix}$$

Retornando a $H_{ii(j)} = \int_{I\!R} \mathrm{v}_{ji} \left(\frac{\partial^2 P_i}{\partial \zeta_i \partial \zeta_i'} \right) g_j^*(\theta) d\theta$, tem-se a primeira parcela de

$$\frac{\partial^2 \log L(\zeta, \eta)}{\partial \zeta_l \partial \zeta_i'} = \sum_{j=1}^{s} r_j \left\{ \frac{\partial^2 P(u_{j.}|\zeta, \eta)/(\partial \zeta_l \partial \zeta_i')}{P(u_{j.}|\zeta, \eta)} - \left(\frac{\partial P(u_{j.}|\zeta, \eta)/(\partial \zeta_l}{P(u_{j.}|\zeta, \eta)} \right) \left(\frac{\partial P(u_{j.}|\zeta, \eta)/(\partial \zeta_i)}{P(u_{j.}|\zeta, \eta)} \right)' \right\}$$

pode ser reescrita como:

$$H_{ii(j)} = \int_{I\!R} (u_{ji} - P_i) W_i H_{ii} g_j^*(\theta) d\theta$$

e, para $i \neq l$,

$$H_{il(j)} = \int_{I\!R} (u_{ji} - P_i)(u_{jl} - P_l) W_i H_{il} g_j^*(\theta) d\theta$$

Daí,

$$H(\zeta_i, \zeta_l) = \frac{\partial^2 \log L(\zeta, \eta)}{\partial \zeta_l \partial \zeta_i'} = \sum_{i=1}^{s} r_j \left\{ H_{il(j)} - h_{i(j)} h'_{l(j)} \right\}$$

Para aplicar o algoritmo "Scoring"de Fisher, nota-se que

$$E[H_{il(j)}] = 0, \quad i, l = 1, 2, \ldots, m \quad e \quad j = 1, 2, \ldots, n.$$

Segue então que:

$$\Delta(\zeta_i, \zeta_l) = E[H(\zeta_i, \zeta_l)] = -\sum_{j=1}^{s} r_j [h_{i(j)} h_{i(j)}].$$

CAPÍTULO 4

Equalização

Equalizar significa equiparar, tornar comparável, na TRI significa colocar parâmetros de itens vindos de provas distintas ou habilidades de respondentes de diferentes grupos, na mesma métrica, ou seja, numa escala comum, tornando os itens e/ou habilidades comparáveis. [Andrade et al. (2000)] Sem a equalização é impossível afirmar, por exemplo, que a média de desempenho de alunos em um Sistema Nacional de Avaliação sofreu mudanças de um ano para o outro. É necessário separar as diferenças de habilidade dos indivíduos avaliados, das diferenças de dificuldade dos instrumentos aplicados [Rabello (2001)].

A importância dessa técnica torna-se evidente quando se percebe que a decisão baseada nos dados de qualquer indivíduo deve ser a mesma independente da forma que ele respondeu e, ainda em situações onde não se quer tomar decisões a nível individual, as análises devem ser feitas levando-se em consideração os escores transformados pela equalização, que são comparáveis.

Existem dois tipos de equalização: a equalização *via população* e a equalização *via itens comuns*. Ou seja, há duas maneiras de colocar parâmetros, tanto de itens quanto de habilidades, numa mesma métrica [Andrade et al. (2000)]. Na primeira usa-se o fato de que se um único grupo de respondentes é submetido a provas distintas, basta que todos os itens sejam calibrados conjuntamente para se ter a garantia de que todos estarão na mesma métrica. E, na equalização via itens comuns, a garantia de que as populações envolvidas terão seus parâmetros em uma

mesma escala será dada pelos itens comuns entre as populações que servirão de ligação entre elas.

4.1 Delineamento de grupos não-equivalentes com itens comuns

Para realizar uma equalização é preciso planejar um delineamento. Como exemplo deste delineamento, pode-se valer de uma aplicação contendo duas formas, X e Y, contendo um conjunto de itens comuns. Para Hambleton & Swaminathan (1995) este delineamento é chamado de delineamento do teste âncora. Para este método, existem duas variações possíveis:

1. os itens comuns contribuem para o cálculo do escore do indivíduo no teste, sendo chamado de itens internos.

2. os itens comuns não contribuem para o cálculo do escore do indivíduo, sendo denominado de itens externos.

Os grupos de itens internos e externos são geralmente aplicados com tempo marcado separadamente.

Para Yang & Houang (1996) independentemente do método de equalização usado em um delineamento que usa itens comuns, o método será mais preciso à medida que aumentar o número de itens comuns. Segundo Andrade (1999) e Yang & Houang (1996) uma referência para o número mínimo de itens comuns é 20% do tamanho total de cada uma das provas a ser equalizada. Além disso,Yang & Houang afirmam que independentemente do delineamento utilizado os testes a serem equalizados devem possuir 35 itens no mínimo.

4.2 Diferentes tipos de equalização

4.2.1 Um único grupo fazendo uma única prova

Este é o caso trivial, em que se aplicam diretamente os modelos matemáticos e os métodos de estimação descritos anteriormente. Ou seja, não é necessário nenhum tipo de equalização.

4.2.2 Um único grupo fazendo duas provas totalmente distintas

Este é o caso clássico que é chamado de equalização via população. Para resolvê-lo, basta que todos os itens de ambas as provas sejam calibrados simultaneamente. O fato de todos os indivíduos representarem uma amostra aleatória de uma mesma população é que garante que todos os parâmetros envolvidos estarão numa mesma escala.

4.2.3 Um único grupo fazendo duas provas parcialmente distintas

Este caso é bastante semelhante ao caso anterior podendo fazer também a equalização via população.

4.2.4 Dois grupos fazendo uma única prova

Este é um exemplo de equalização via itens comuns (só que no caso, todos). Como as duas populações fazem exatamente a mesma prova, basta que os itens sejam calibrados utilizando-se as respostas dos respondentes de ambos os grupos simultaneamente.

4.2.5 Dois grupos fazendo duas provas totalmente distintas

Este é o único caso que não pode ser resolvido pela TRI. É possível calibrar separadamente os itens das duas provas, porém não é possível fazer nenhum tipo de comparação entre os resultados obtidos, uma vez que eles estarão em métricas diferentes. Então não faz sentido comparar os resultados destes dois grupos.

4.2.6 Dois grupos fazendo duas provas parcialmente distintas

Esse também é um caso de equalização via itens comuns. Este caso ilustra o maior avanço da TRI sobre a Teoria Clássica. O uso de itens comuns entre provas distintas aplicadas a populações distintas permite que todos os parâmetros estejam na mesma escala ao final dos processos de estimação, possibilitando comparações e a construção de "escalas de conhecimento" interpretáveis, que são de grande importância na área educacional. A resolução deste caso é bastante semelhante ao que foi descrito na Seção (4.2.4), com a diferença que aqui apenas alguns dos itens (e não a prova toda) fazem a ligação entre as duas populações envolvidas.

4.3 Diferentes problemas de estimação

4.3.1 Quando todos os itens são novos

Neste caso, deseja-se calibrar o conjunto completo de itens. Este é o caso trivial e para resolver este problema, basta utilizar uma das técnicas de estimação descritas no capítulo anterior.

4.3.2 Quando todos os itens já estão calibrados

Este é o caso em que todos o itens já foram calibrados anteriormente, ou seja, deseja-se apenas estimar as habilidades dos respondentes. Este também é um caso bastante frequente na TRI, devido ao impulso que esta teoria deu na criação de *bancos de itens*. Esses bancos são formados por conjuntos de itens que já foram testados e calibrados a partir de um número significativo de indivíduos de uma dada população. Dessa forma, assume-se que os parâmetros desses itens já são "conhecidos", ou seja, assume-se que são conhecidos os verdadeiros valores dos parâmetros desses itens e assim, sempre que se desejar, pode-se aplicar novamente alguns desses itens do banco a outros indivíduos (ou até mesmo a um único indivíduo)e assim estimar apenas suas habilidades que estarão sempre na mesma métrica dos parâmetros dos itens.

Quando se "constrói" um banco de itens, uma informação fundamental é a escala em que aqueles itens foram calibrados, visto que as habilidades de indivíduos que serão estimadas futuramente a partir daqueles itens estarão nesta mesma métrica e portanto, quaisquer comparações diretas só poderão ser feitas com outros indivíduos que também tenham suas habilidades nesta escala.

Dessa forma, para resolver este problema, basta utilizar um dos processos de estimação das habilidades dos indivíduos quando os parâmetros dos itens já são conhecidos.

4.3.3 Quando alguns itens são novos e outros já estão calibrados

Neste caso, deseja-se calibrar alguns itens e manter os parâmetros de outros, que já foram calibrados anteriormente. Este caso também está ligado à criação de bancos de itens, uma vez que um banco de itens está continuamente em formação, ou seja, pode-se estar interessado em acrescentar novos itens ou retirar itens ao conjunto que já se encontra no banco.

Para resolver este problema, deve-se definir uma das populações como sendo a referência, e então, as demais populações serão posicionadas em relação a ela, a fim de evitar problemas de indeterminação de escala.

4.4 A escala de Habilidade

A escala é uma ferramenta utilizada para sistematizar informações. Nela é utilizado o conceito de medida que atribui números aos fenômenos naturais. As habilidades dos estudantes são fenômenos naturais observáveis para os quais podem ser atribuídas quantidades. Uma escala é capaz de representar em que e quanto mais grupos de estudantes construiram competências e habilidades que outros grupos de estudantes ainda não construiram.

Quando se fala em métrica, significa se referir ao tipo de escala utilizada para medir um dado fenômeno, ou seja, se um indivíduo que obtém nota 9 é considerado excelente numa prova de desempenho, então supõe-se que a métrica utilizada é uma escala que vai de 0 a 10, pois se a escala fosse de 0 a 100, então a nota representaria péssimo desempenho. Assim, é de fundamental importância saber a métrica utilizada para poder entender o significado do valor atribuído.

A escala da habilidade é uma escala arbitrária onde o importante são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. O parâmetro c não depende da escala, pois trata-se de uma probabilidade, e como tal, assume sempre valores entre 0 e 1. O parâmetro b representa a habilidade necessária para uma probabilidade de acerto casual igual a $(1 + c)/2$. Assim, quanto maior o valor de b, mais difícil é o item, e vice -versa.

Na TRI a habilidade pode assumir teoricamente qualquer valor real entre $-\infty$ e $+\infty$. Daí, é necessário estabelecer uma origem e uma unidade de medida para a definição da escala. Esses valores são escolhidos de modo a representar, respectivamente, o valor médio e o desvio-padrão das habilidades dos indivíduos da população em estudo. Em muitos casos, utiliza-se a escala com média igual a 0 e desvio-padrão igual a 1, que é representada por escala (0,1). Essa escala é bastante utilizada pela TRI, e nesse caso, os valores do parâmetro b variam (tipicamente) entre -2 e +2. Com relação ao parâmetro a, espera-se valores entres 0 e +2, sendo que os valores mais apropriados de a seriam aqueles maiores que 1 [Andrade et al. (2000)].

4.4.1 Construção e interpretação de escalas de habilidade

Uma escala é constituída por uma sequência numérica. O número em uma escala está sempre associado a uma interpretação. Na escala utilizada pelo SAEB, por exemplo, cada intervalo numérico representa um nível de desempenho de grupos de estudantes e vem acompanhado de uma interpretação das competências e habilidades que eles já construíram no seu processo de desenvolvimento.

A escala é definida por *níveis âncora*, que por sua vez são caracterizados por um conjunto de itens denominados *itens âncora*. Níveis âncora são pontos selecionados pelo analista na escala da habilidade para serem interpretados pedagogicamente. Já os itens âncora são itens selecionados para cada um dos níveis âncora segundo um critério de definição, ou seja, sejam dois níveis âncora consecutivos Y e Z com $Y < Z$. Um determinado item é âncora para o nível Z e somente se as 3 condições abaixo forem satisfeitas simultaneamente: [Andrade et al. (2000)]

1. $P(U = 1|\theta = Z) \geq 0,65$;

2. $P(U = 1|\theta = Y) < 0,50$;

3. $P(U = 1|\theta = Z) - P(U = 1|\theta = Y) \geq 0,30$

Dessa forma, para um item ser âncora em um determinado nível âncora da escala, ele precisa ser respondido corretamente por uma grande proporção de indivíduos (pelo menos 65%) com este nível de habilidade e por uma proporção menor de indivíduos (no máximo 50%) com o nível de habilidade imediatamente anterior. Além disso, a diferença entre a proporção de indivíduos com esses níveis de habilidade que acertam a esse item deve ser de pelo menos 30%. Assim, itens-âncora são itens que caracterizam um ponto ou nível das escalas para o qual a grande maioria dos alunos situados naquele nível acerta o item, ao passo que um percentual considerável de alunos situados ao nível abaixo da escala erra o item. Esses itens têm o objetivo de discriminar pontos na escala que separam alunos que construiram daqueles que não construiram determinadas competências ou habilidades.

Para auxiliar na interpretação das escalas, é preciso identificar um número razoável de itens-âncora para cada um dos níveis da escala. É importante que o conjunto de itens-âncora cubra a maior extensão possível da matriz de referência para que a análise das competências e habilidades seja rica.

A interpretação dos itens âncora é realizada por meio de painéis de especialistas com o objetivo de buscar o significado das respostas dadas pelos alunos em cada um dos níveis da escala, de forma a desenvolver uma descrição de suas habilidades. Esses painéis contam com professores e especialistas nas disciplinas e nas áreas de Educação e Avaliação, com experiência em interpretação de escala em termos de competências e habilidades e nas matrizes de referência.

Através desse trabalho de interpretação, são então definidas quais são as competências e habilidades dos estudantes que se situam em cada um dos níveis da escala.

4.5 Equalização a posteriori

A equalização a posteriori é uma forma de equalizar as populações envolvidas depois de terminado o processo de calibração dos itens. Para isso, procede-se da seguinte maneira: calibra-se separadamente os dois conjuntos de itens, que foram submetidos às duas populações de interesse. A condição necessária é que hajam itens comuns entre os dois conjuntos. Assim, para os itens comuns, tem-se dois conjuntos de estimativas, cada uma na métrica de suas respectivas populações. Assim, através dessas duas estimativas para os itens comuns estabelece-se algum tipo de relação que permita colocar os parâmetros de um dos conjuntos de itens na escala do outro. Com todos os itens na mesma métrica, pode-se então estimar as habilidades de todos os respondentes, que estarão também na mesma escala.

A equalização através da TRI se vale de uma característica desses modelos (delineamento de grupos não-equivalentes com itens comuns), que diz que se o modelo se ajusta bem aos dados, qualquer transformação linear da escala θ também se ajustará. Assim, estimativas diferentes obtidas a partir de diferentes programas estão relacionadas linearmente através das escalas θ. Desse modo, uma equação linear pode ser usada para converter os parâmetros da TRI de uma escala para uma outra escala. As estimativas dos parâmetros resultantes (chamadas de calibradas),

podem então ser usadas para estabelecer escores equivalentes, baseados no número de respostas corretas nas duas formas, e qualquer transformação linear das estimativas dos parâmetros de habilidades obtidos em uma determinada escala, essa não altera a probabilidade de um indivíduo acertar o item, caso os parâmetros de itens também sejam transformados apropriadamente.

A função de probabilidade de θ é dada por:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \qquad i = 1, 2, \ldots, m$$

Primeiramente é efetuada uma transformação linear de θ, gerando θ^* e em seguida têm-se a nova função de probabilidade após a transformação, ou seja,

$$\theta^* = A\theta + B \Leftrightarrow \theta = \frac{\theta^* - B}{A}$$

Aplicando essa transformação em $P_i(\theta)$, tem-se que,

$$P_i(\theta^*) = c_i + (1 - c_i)\frac{e^{Da_i(\frac{\theta^* - B}{A} - b_i)}}{1 + e^{Da_i(\frac{\theta^* - B}{A} - b_i)}} \qquad i = 1, 2, \ldots, m$$

mas,

$a_i\left(\frac{\theta^* - B}{A} - b_i\right) = a_i\left(\frac{\theta^* - B - Ab_i}{A}\right) = \frac{a_i}{A}[\theta^* - (Ab_i + B)] = a^*(\theta^* - b^*)$

onde,

$a^* = \frac{a_i}{A}, \quad b^* = Ab_i + B \quad e \quad \theta^* = A\theta + B$

Então, quando se transforma linearmente a escala de θ, a probabilidade de acerto ao item não se altera, desde que sejam também transformados os parâmetros dos itens a e b, pois o parâmetro c é independente da transformação da escala.

Apesar da frequente utilização da escala (0,1), em termos práticos, não faz a menor diferença estabelecer-se estes valores ou outros quaisquer. O importante são as relações de ordem existentes entre seus pontos. Por exemplo, na escala (0,1) um indivíduo com habilidade 1,20 está 1,20 desvios-padrão acima da habilidade média. Este mesmo indivíduo teria a habilidade 92 e, consequentemente estaria também 1,20 desvios-padrão acima da habilidade média, se a escala utilizada para esta população fosse a escala (80;10). Isto pode ser visto a partir da transformação de escala: [Andrade (2000)]

$$a(\theta - b) = (a/10)[(10 \times \theta + 80) - (10 \times b + 80)] = a^*(\theta^* - b^*),$$

onde $a(\theta - b)$ é a parte do modelo probabilístico proposto envolvida na transformação. Assim tem-se que:

- $\theta^* = 10 \times \theta + 80$,
- $b^* = 10 \times b + 80$,
- $a^* = a/10$,
- $P(U_i = 1|\theta) = P(U_i = 1|\theta^*)$.

Assim sendo, a probabilidade de um indivíduo responder corretamente a um certo item é sempre a mesma, independente da escala utilizada para medir a sua habilidade, ou ainda, a habilidade de um indivíduo é invariante à escala de medida. Dessa forma, não faz qualquer sentido querermos analisar itens a partir dos valores de seus parâmetros a e b sem conhecer a escala na qual eles foram determinados.

Vários métodos, que se baseiam nessas relações lineares existentes entre os parâmetros de um mesmo item medidos em escalas diferentes, poderiam ser então utilizados para determinar os coeficientes A (constante de inclinação) e B (constante de intercepto). A solução mais natural, pelo próprio tipo de relação existente entre os parâmetros, seria determinar esses coeficientes através de uma regressão linear simples. Porém, a crítica feita à utilização desse método é que ele não é simétrico, ou seja, uma regressão de x por y é diferente de uma regressão de y por x.

Um dos métodos de equalização a posteriori existentes que não apresenta esse problema por ser invariante em relação as variáveis utilizadas, é chamado *Média-Desvio (Mean-Sigma)*. Nesse método é utilizado:

$$A = \frac{\sigma_{G1}}{\sigma_{G2}} \quad e \quad B = \mu(b_{G1}) - A\mu(b_{G2})$$

onde,

σ_{G1} e σ_{G2} são os desvios-padrão e $\mu(b_{G1})$ e $\mu(b_{G2})$ as médias amostrais das estimativas dos parâmetros de dificuldade dos itens comuns nos grupos 1 e 2, respectivamente. Do mesmo modo, as habilidades dos respondentes do grupo 2 podem ser colocadas na mesma escala das habilidades

dos respondentes do grupo 1 da seguinte forma:

$$\theta^1_{G2} = A\theta_{G2} + B$$

onde θ^1_{G2} é o valor da habilidade θ_{G2} na escala do grupo 1. Maiores detalhes sobre este e outros métodos de equalização, podem ser encontrados em Kolen & Brennan (1995).

Quando o delineamento de itens comuns é aplicado, a equalização pode ser executada em uma ou duas etapas. Quando realizada em duas etapas, calibram-se as bases separadamente e, depois, efetua-se uma transformação na escala, chamada de *scale linking*, transformando as novas estimativas dos parâmetros dos itens comuns para a escala da primeira avaliação. Da mesma forma, esta transformação é usada para colocar as estimativas de habilidades dos alunos na escala da primeira avaliação. A outra forma consiste na análise simultânea das bases, de forma que, ao final da calibração, as estimativas para itens e indivíduos estejam todas em uma mesma escala.

A equalização é aplicada no SAEB para comparar os escores dos alunos em diversas situações, tais como:

1. comparação de escores dos alunos que frequentam uma mesma série;
2. comparação de escores dos alunos que frequentam séries diferentes;
3. comparação dos escores dos alunos, obtidos em anos diferentes.

CAPÍTULO 5

Aplicação

Neste capítulo será apresentada uma aplicação de análise de itens e de desempenho escolar em Pernambuco através dos dados do SAEPE-2005.

O Sistema de Avaliação Educacional de Pernambuco (SAEPE) cujo objetivo principal é desenvolver um trabalho permanente de monitoria e incentivos para a melhoria da qualidade e do desempenho das escolas, foi criado em 2000 com a proposta de ser bi-anual, porém o primeiro exame só foi realizado em 2002 e o segundo em 2005. Ele é aplicado aos alunos de $2^a, 4^a$ e 8^a séries do Ensino Fundamental e 3^a série do ensino Médio da rede Estadual e Municipal, que abrange 184 municípios.

A Secretaria de Educação do Estado de Pernambuco, em regime de colaboração com as Secretarias Municipais de Educação do Estado, representadas pela União Nacional de Dirigentes Municipais de Educação (UNDIME) e em convênio com o Ministério da Educação, através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP, coletou dados para o SAEPE.

Para atingir os objetivos, vários instrumentos são empregados além das provas de avaliação do desempenho escolar, ou seja, são utilizados questionários (inseridos no apêndice B) que permitem:

1. Obter informações sobre as características da realidade socioeconômica e cultural e hábitos de estudo dos alunos;

2. avaliar o perfil e a prática pedagógica dos professores;

3. avaliar o perfil e as práticas de gestão escolar dos diretores;

4. realizar o levantamento dos equipamentos disponíveis, das características físicas e de conservação das escolas;

Os indicadores resultantes dessas avaliações permitem que se façam associações, correlações, análises hierárquicas e estudos relevantes sobre a realidade educacional do estado de Pernambuco.

Nesse estudo, foi feita a análise dos itens das provas de Português e Matemática da 3^a série do Ensino Médio do SAEPE 2005, bem como das habilidades dos indivíduos referentes a tais disciplinas. Para isso, 48789 alunos distribuídos em 7142 escolas (Estaduais e Municipais), participaram dessa avaliação. As provas de Português e Matemática foram constituídas de 84 itens cada uma.

As provas aplicadas pelo SAEPE, adotaram o delineamento de *Blocos Balanceados Incompletos* (BIB). Nem sempre é viável ou desejável que todos os itens do teste sejam respondidos por todos os indivíduos, porém, muitas vezes faz-se necessário assegurar uma ampla e representativa cobertura do conteúdo da avaliação, dessa forma, tal representação é realizada através do BIB. Ou seja, um conjunto de itens é dividido em um número menor de blocos, estes são então designados para os cadernos, de modo que cada bloco seja emparelhado com outro bloco para formar um caderno.

Por meio desse delinemento, os itens foram divididos em 7 blocos com 12 itens cada, e esses blocos foram divididos em 21 cadernos diferentes com 24 itens cada um. Cada aluno responde a um caderno de português e outro de matemática totalizando 48 itens (24 de português e 24 de matemática).

Os ítens que compõem o teste foram pré-testados.

5.1 Método Utilizado

A análise dos dados foi feita baseada nas duas teorias: a TCT (Teoria Clássica dos Testes) e a TRI (Teoria da Resposta ao Item). Utilizando a TCT, foram calculados os percentuais de

acerto de cada item e o coeficiente de correlação bisserial.

Para a estimação dos parâmetros do modelo da TRI, foi utilizado um programa computacional visto que existe uma grande quantidade de dados que exigem compilação e também pela complexidade das operações. Existem vários softwares que executam os procedimentos da TRI. Aqui no Brasil, os mais utilizados são o BILOG e o BILOG-MG. Segundo Andrade et al. (2000), estes dois programas são específicos para análises via TRI de itens dicotômicos ou dicotomizados, e ainda, ambos têm implementados os modelos unidimensionais logísticos de 1, 2 e 3 parâmetros.

O BILOG permite apenas analisar respondentes provenientes de uma única população, enquanto que o BILOG-MG permite a análise de mais de um grupo de respondentes.

Nesse estudo foi utilizado o programa BILOG, visto que a análise se restringia a uma única população com itens dicotômicos (certo ou errado). Esse programa tem como entrada um arquivo em linguagem própria, (inserido no apêndice) com extensão.blm, onde informa ao programa principal BILOG as especificações de entrada de dados formatados em arquivo texto e salva os resultados de saída, onde SAVE são as especificações de salvamento desses arquivos.

Nos arquivos de saída, tem-se:

- um arquivo com as estimativa dos parâmetros dos modelos da TRI com extensão.par;
- um arquivo com os escores (número de acertos e percentual) e habilidades de todos os indivíduos que compõem o arquivo de dados, com extensão.sco;
- um arquivo com os gráficos da CCI, com extensão.plt;
- um arquivo mostrando os acertos e erros dos examinandos por item, com extensão.sor.

O BILOG executa a análise em 3 etapas, chamadas de fase 1, 2 e 3 que se caracterizam pelo tipo de tarefas realizadas em cada uma delas.

A fase 1, fase de entrada e leitura dos dados, contém dois tipos de informação: a identificação de cada indivíduo com suas respectivas respostas ao teste e o gabarito (que é uma sequência contendo as alternativas corretas dos itens que compõem o teste). Nessa fase, além da verificação de que a leitura dos dados foi feita corretamente, são calculadas algumas estatísticas descritivas,

tais como número de indivíduos submetidos a cada item e algumas correlações de interesse como as correlações bisserial e ponto bisserial, usadas na TCT. Essas estatísticas são utilizadas posteriormente como valores iniciais para os processos de estimação realizados nas fases seguintes. A correlação bisserial fornece um diagnóstico preliminar dos itens, servindo por exemplo, na identificação de itens com problemas no gabarito.

A fase 2 é a fase de calibração dos itens. Aqui são estimados os parâmetros dos itens, com seus respectivos erros-padrão. O BILOG fornece também gráficos contendo algumas informações de interesse, tais como as curvas características dos itens. Juntamente com a curva característica de cada item é fornecido também um teste de ajuste do modelo utilizado.

A fase 3 é a fase da estimação das habilidades dos respondentes. São estimadas as habilidades de cada um dos indivíduos, a partir dos resultados obtidos na fase anterior. Essas habilidades são inicialmente estimadas na escala dos parâmetros dos itens, porém pode-se especificar alguns tipos de mudança na escala, que poderão ser feitas tanto nas habilidades quanto nos parâmetros estimados na fase anterior.

5.1.1 Métodos para a Calibração dos Itens

Inicialmente o programa realiza a calibração dos itens e depois a estimação das habilidades dos respondentes. Dois métodos de estimação para os parâmetros dos itens estão implementados: a máxima verossimilhança marginal e um método bayesiano de estimação por maximização da distribuição marginal a posteriori. Para isso, é necessário a utilização de distribuições de probabilidade para as habilidades dos respondentes. Esses programas assumem que os respondentes representam uma amostra aleatória de uma população de proficiências que pode ser assumida como tendo uma distribuição normal padrão, ou ainda uma distribuição empírica, a ser estimada conjuntamente com os parâmetros dos itens.

5.1.2 Métodos Implementados para a Estimação das Habilidades

Depois de terminada a fase de calibração dos parâmetros dos itens, é feita a estimação das habilidades dos respondentes. O BILOG e o BILOG-MG têm implementado o método

de estimação por máxima verossimilhança, por esperança a posteriori (EAP) e por máximo a posteriori (MAP). No método da máxima verossimilhança, as estimativas das habilidades dos respondentes são calculadas pelo método de Newton-Raphson, utilizando-se uma transformação linear do logito do percentual de acertos dos indivíduos como valores iniciais. Os problemas descritos anteriormente com as estimativas dos respondentes que tiveram erro total ou acerto total, são contornados através de um artifício: os alunos que erraram todos os itens ganham um meio certo no item mais fácil, e os alunos que acertaram todos os itens, perdem um meio certo no item mais difícil. Este método nem sempre fornece boas estimativas.

No método EAP, as estimativas para as habilidades são calculadas utilizando-se pontos de quadratura a fim de aproximar a distribuição a priori das habilidades de cada respondente. O número de pontos de quadratura é definido pelo usuário, que pode também escolher entre uma priori que seja normal (e cujos parâmetros podem ser especificados pelo usuário), ou uma distribuição discreta arbitrária (fornecida pelo usuário), ou ainda uma distribuição discreta empírica, através do uso dos pontos de quadratura e de seus respectivos pesos gerados na fase 2.

As estimativas EAP para as habilidades dos respondentes estão sempre definidas, qualquer que seja o padrão de respostas. Quando se utiliza a estimação por EAP, é fornecida uma estimativa da distribuição de habilidade da população de respondentes, na forma de uma distribuição discreta, dada pelos pontos de quadratura. Esta distribuição é obtida acumulando-se as densidades da posteriori de todos dos sujeitos em cada ponto de quadratura. As somas são então normalizadas para se obter as probabilidades estimadas em cada ponto. Além disso, são fornecidos a média e o desvio-padrão para essa distribuição estimada.

No método MAP, as estimativas das habilidades são calculadas pelo método Newton-Gauss. Este procedimento sempre converge e fornece estimativas para todos os padrões de respostas possíveis. É assumida uma distribuição a priori normal, cujos parâmetros podem ser especificados pelo usuário, sendo que o padrão definido nesses programas é a normal padrão, [Andrade, (2000), pgs.: 123 a 127].

5.1.3 Métodos Utilizados para Comparar Habilidades

A partir do interesse de analisar se a habilidade do aluno é função da escola que ele estuda ou da região em que ele mora, ou ainda do gênero ao qual pertence, foram aplicados alguns procedimentos estatísticos adequados, quais sejam:

1. as escolas foram agrupadas em função de algumas características usando uma técnica estatística de agrupamento;

2. na comparação das habilidades médias por grupos de escola foi utilizado o teste de Kruskal Wallis (teste não paramétrico de comparação de médias de múltiplas populações não normais);

3. na comparação das habilidades médias entre regiões (metropolitana e não-metropolitana) e por sexo (masculino e feminino) foi utilizado o teste de Mann Whitney (teste não paramétrico para comparação de médias de duas populações não normais).

5.2 Resultados e Discussões

As análises estatísticas aplicadas aos dados do SAEPE 2005 (3^a série do ensino médio) que os alunos da rede Estadual e Municipal foram submetidos, resultou na estimação das habilidades/proficiência, do índice de dificuldade e do poder de discriminação de cada item. Para isso, usou-se a TRI.

Na análise da TCT, obteve-se a estimativa da discriminação dos itens através do coeficiente de correlação bisserial, que é o adequado quando se trabalha com uma variável discreta (pontuação total no teste) e outra variável dicotômica (indicação de acerto ou erro do item) [Francisco (2005)].

5.2.1 Análise das Estimativas dos Indicadores - SAEPE 2005

A partir dos dados da tabela 5.1, constata-se que em Português os estudantes apresentaram uma habilidade média baixa de -0,118. A análise dos itens revelou um valor médio de dificuldade de 1,418 que aponta para um teste razoavelmente difícil e um índice médio de discriminação de

1,3 que demonstra que na média, a discriminação é boa, e ainda, uma média do indicador de acerto casual menor que 0,2.

Tabela 5.1: Média e Desvio-Padrão das Estimativas dos Indicadores- SAEPE 2005 (Português)

Indicador	**Média**	**Desvio Padrão**
Habilidade	-0,118	1,364
Discriminação (a)	1,300	0,530
Dificuldade (b)	1,418	1,036
Acerto Casual (c)	0,125	0,045

Quanto à avaliação de Matemática, os dados apresentados na tabela 5.2, revelaram uma habilidade média baixa de -0,162, a análise dos itens revelou um valor médio de dificuldade de 3,612 que aponta para um teste extremamente difícil. O índice médio de discriminação é 1,389 o que demonstra que na média a discriminação é boa. Observa-se também que o índice médio de dificuldade em Matemática é bem maior que Português, porém o desvio padrão também é muito maior, ou seja, em média, Matemática foi mais difícil, mas as questões de Português tinham todas mais ou menos o mesmo nível de dificuldade, enquanto que Matemática tinha questões ou muito fáceis e/ou muito difíceis.

Tabela 5.2: Média e Desvio-Padrão das Estimativas dos Indicadores- SAEPE 2005 (Matemática)

Indicador	**Média**	**Desvio Padrão**
Habilidade	-0,162	1,425
Discriminação (a)	1,389	0,820
Dificuldade (b)	3,612	5,077
Acerto Casual (c)	0,136	0,053

5.2.2 Análise dos Itens

Nesta Seção serão apresentados os principais resultados da análise dos itens usando a TCT e a TRI.

Análise dos Itens pela TCT

O cálculo do índice de dificuldade dos itens, com base na TCT, é calculado baseado no percentual de examinandos que respondem corretamente a um dado item, enquanto a discriminação dos itens foi realizada considerando-se os coeficientes de correlação bisserial. A tabela 5.3 a seguir, mostra que 32 itens (38%) da prova de português e 67 (80%) da prova de matemática foram considerados difíceis ($I < 30$) e 11 itens (13%) da prova de português e 38 (45%) de matemática apresentaram correlação bisserial menor que 0,20, indicando que ou existe algum problema com a construção do item, ou que o conhecimento exigido para resolver a questão não é de domínio de quem supostamente o sabe.

Como os 84 itens usados na elaboração das provas já haviam sido pré-testados, é possível que o elevado percentual (80%) de itens difíceis e de (45%) itens com baixa correlação bisserial seja a consequência do baixo nível de conhecimento dos estudantes em matemática.

Tabela 5.3: Número de itens com problemas pela TCT

Disciplina	**Índice de Dificuldade (<30)**	**Correlação Bisserial (<0,2)**
Português	32 (38%)	11 (13%)
Matemática	67 (80%)	38 (45%)

Análise dos Itens pela TRI

A análise dos resultados da aplicação da TRI é de fundamental importância visto que a unidade de análise é o item e não o teste, como na TCT.

Seguindo Rodrigues (2006) os critérios adotados para o julgamento dos parâmetros da TRI, foram:

- Discriminação ruim: a < 0,60;
- Itens difíceis: b > 2;
- Itens fáceis: b < -2;

- Probabilidade de acerto casual: c > 0,3

A tabela 5.4 a seguir mostra que os maiores problemas surgiram em torno do parâmetro b. Os resultados apresentados, referem-se aos parâmetros $b > 2$, uma vez que não foram encontrados itens com índices menores que -2. Conclui-se então que um número significativo de itens de português (23,8%) e mais da metade dos itens de matemática (61,9%) exigia um elevado grau de habilidade para a sua resolução.

Quanto ao parâmetro de discriminação a, na prova de português, 6 itens (7,1%) não apresentaram um alto poder de discriminação, enquanto que em matemática esse número foi 11 (13,1%). Percebe-se também que dos 20 itens que apresentaram problemas em relação ao parâmetro b na disciplina de português, 5 deles também apresentaram problemas em relação ao parâmetro a. Já em matemática, dos 52 itens que apresentaram problemas com o parâmetro b, 11 deles também tiveram problemas quanto ao parâmetro a.

Não houve problema quanto a probabilidade de acerto casual, pois c foi sempre menor que 0,22.

Tabela 5.4: Número de itens que apresentam problemas com os parâmetros da TRI

Disciplina	**a<0,6**	**b>2,0**	**a<0,6 e b>2,0**
Português	6 (7,1%)	20 (23,8%)	5 (5,95%)
Matemática	11 (13,1%)	52 (61,9%)	11 (13,1%)

Conforme referido na Seção 4.4.1, um item para ser considerado âncora em um determinado nível âncora da escala, deve ser bastante acertado por indivíduos com determinado nível de habilidade e pouco acertado por indivíduos com um nível de habilidade imediatamente inferior.

A tabela 5.5 adiante, mostra que de um total de 84 itens de português e 84 de matemática, apenas 32 itens (38%) são âncora para português e 31 (36,9%) para matemática. Verifica-se também que para a disciplina de português existe apenas 1 item âncora para o nível âncora 0, 13 para os níveis 1 e 2 e 5 itens âncora para o nível âncora 3. Na disciplina de Matemática, não existe item âncora para o nível âncora 0, existem 5 itens âncora para o nível 1, 14 para o nível 2 e 12 para o nível âncora 3. Observa-se também que não existem itens âncora para os níveis

$-1, -2$ e -3 nas provas de português e matemática. A prova de português tem itens âncora mais bem distribuídos pelos níveis âncora que matemática.

Tabela 5.5: Número de itens âncora para os níveis âncora da prova de Português e Matemática

Níveis Âncora	**Português**	**Matemática**
0	1	0
1	13	5
2	13	14
3	5	12

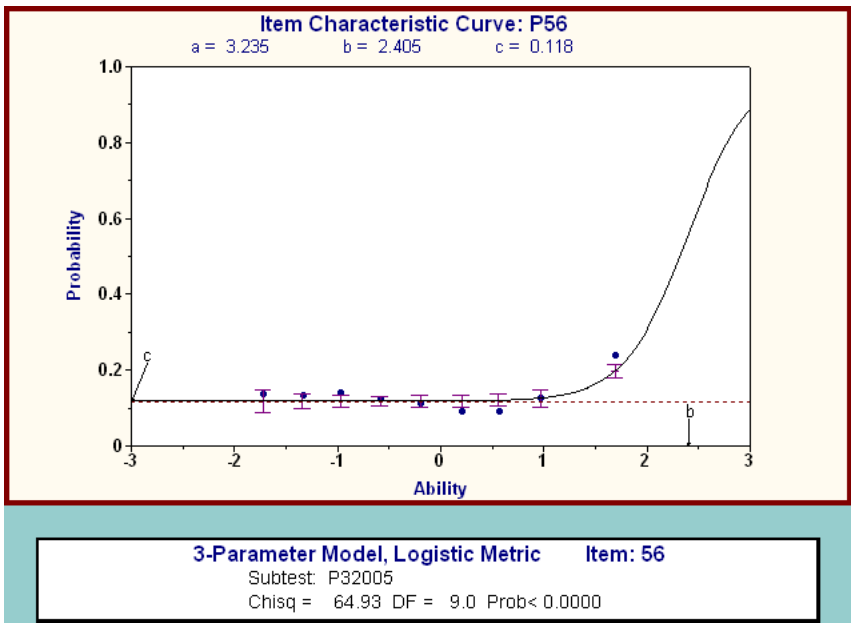
Para explicar melhor a forma analítica dos itens de uma prova foram destacados alguns itens de cada uma das provas. Da prova de português foram destacados 4 itens. A tabela 5.6 mostra os valores dos parâmetros dos dois itens mais difíceis e dos dois menos difíceis dos 84 existentes da prova de português, visto que não existem itens fáceis nessa prova, conforme já referido.

Tabela 5.6: Parâmetros de alguns dos itens de Português

Item	**% de acerto**	**bisserial**	**a**	**b**	**c**
29	13,3	0,22	2,65	2,26	0,11
56	13,1	0,13	3,24	2,40	0,12
2	64,7	0,45	1,48	-0,20	0,21
13	58,2	0,47	1,33	-0,35	0,06

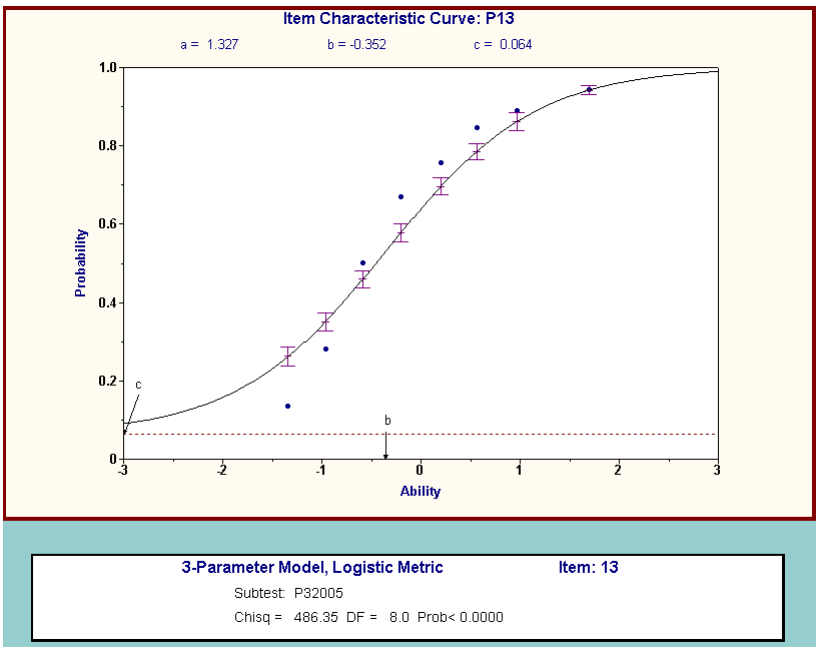
Dos itens de português que apresentaram problemas, o item 56, cuja curva característica é apresentada na Figura 5.1, é o mais difícil tanto pela análise da TCT, onde o percentual de acerto é de 13,1% e o coeficiente bisserial é 0,13, quanto pela TRI, onde $a = 3,24$; $b = 2,4$ e $c = 0,12$. Observa-se que o poder de discriminação deste item é grande, pois é muito difícil, $b = 2,405$, para uma média de dificuldade de 1,418 e, além disso, a probabilidade de acerto aumenta significativamente quando se faz um deslocamento na escala das habilidades. Percebe-se também que a probabilidade de acerto casual é pequena, ou seja $c = 11,8\%$. A curva característica do item 29 é semelhante a do item 56.

Figura 5.1: Curva Característica do item 56 de português (o mais difícil)



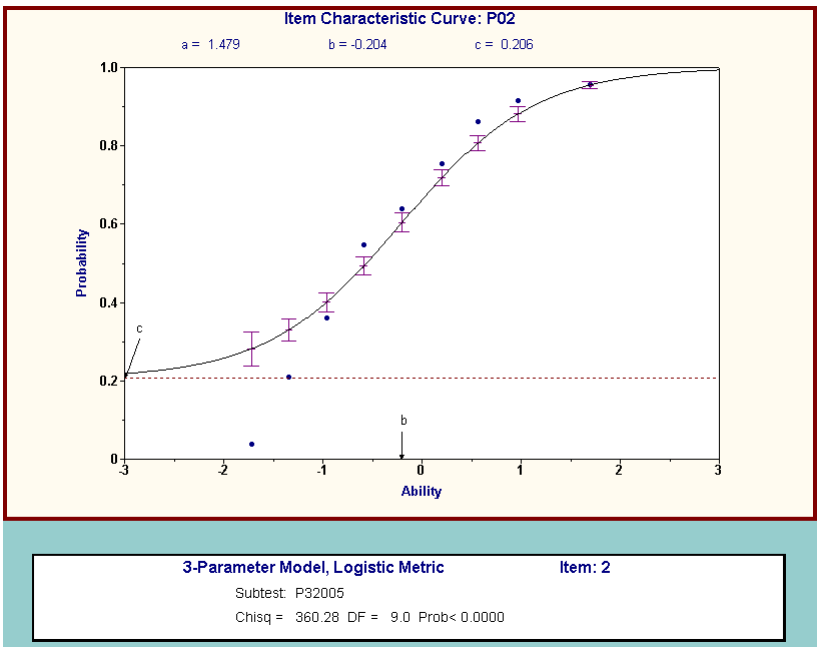
O item 13 ilustrado na figura 5.2, é um item de dificuldade média pois seu percentual de acerto foi de 58,2% e seu índice de dificuldade foi de -0,352. O índice de discriminação desse item também é muito bom (a=1,327) e a probabilidade de acerto casual é baixa (c=0,064). A correlação bisserial também é boa (0,473). Logo ele possui características para ser classificado como bom.

Figura 5.2: Curva Característica do item 13 de português (item bom)



A figura 5.3 a seguir mostra a curva característica do item 2 que conforme os dados já apresentados na tabela 5.6 foi o que apresentou maior percentual de acerto (64.7%). O índice de dificuldade foi de -0.20 e a discriminação foi de 1.48. Este é o item menos difícil dos 84 que foram apresentados na prova de português.

Figura 5.3: Curva Característica do item 02 de português (o menos difícil)



Em se tratando da prova de matemática, a tabela 5.7 a seguir, mostra os valores dos parâmetros dos três itens mais difíceis e dos três menos difíceis de matemática tanto pelo enfoque da TCT quanto pela TRI, dos 84 itens apresentados na prova de matemática. O item 79 é o mais difícil, apesar de não ser o item que tem o menor percentual de acerto (11,9%), entretanto, seu coeficiente bisserial é menor em relação aos outros dois (-0,003) e pela TRI, seu índice de dificuldade $b_{79} = 27,93$ é o maior dentre todos os itens.

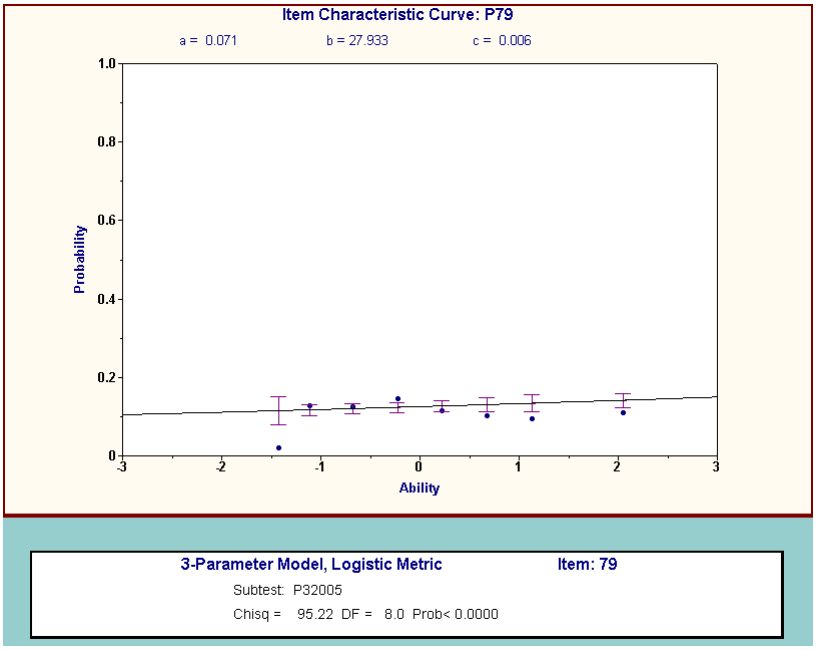
Tabela 5.7: Parâmetros de alguns itens de Matemática

Itens	**% de acerto**	**bisserial**	**a**	**b**	**c**
6	10,2	0,159	3,948	2,586	0,094
62	10,4	0,006	0,104	20,921	0,010
79	11,9	-0,003	0,071	27,933	0,006
7	44,4	0,308	1,257	0,565	0,115
61	53,6	0,367	1,764	0,139	0,075
78	43,5	0,252	0,879	0,595	0,060

A análise da curva característica do item 79 ilustrada na figura 5.4 a seguir, confirma que

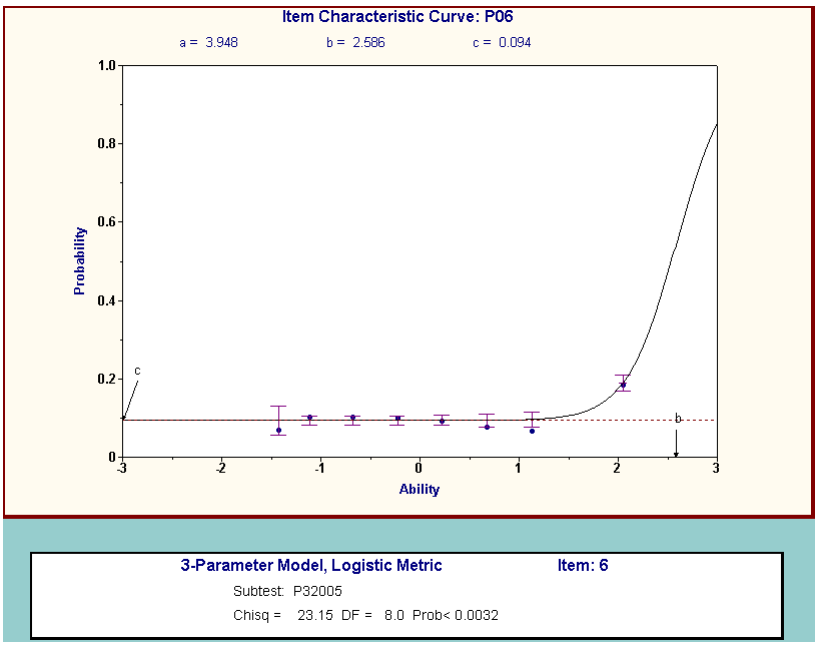
a discriminação deste item (79) é muito pequena, ou seja, qualquer deslocamento na escala da habilidade não causa alteração na probabilidade de acertar o item. Além disso, este item apresentou um grau de extrema dificuldade $b_{79} = 27,93$ para uma média de dificuldade de 3,612 conforme tabela 5.2 apresentada anteriormente.

Figura 5.4: Curva Característica do item 79 de matemática (o mais difícil)



O item 6, conforme figura 5.5 adiante, também é um item muito difícil onde seu percentual de acerto foi de 10,2%, e seu índice de dificuldade 2,586. Além disso, o índice de discriminação desse item é muito alto $a = 3,95$. Observa-se que a curva é muito íngreme, onde um certo deslocamento no eixo da habilidade, causa uma alteração bastante significativa na probabilidade de acertar o item.

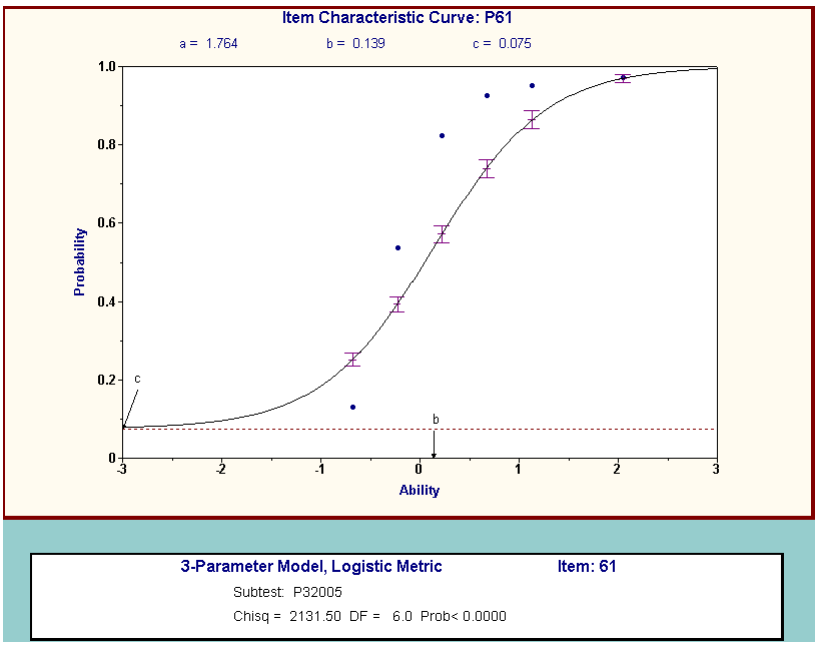
Figura 5.5: Curva Característica do item 06 (o mais discriminante)



Assim, percebe-se que tanto o item 79, como o item 6 são muito difíceis, porém, como o índice de discriminação do item 79, $a_{79} = 0,071$ é muito baixo, ele não consegue classificar alunos de baixo e alto desempenho. Já o item 6, apesar de difícil, tem discriminação $a_6 = 3,95$, ou seja, ele discrimina muito bem os alunos que tem habilidade maior que 2,6 daqueles que não tem.

Por outro lado, a curva característica do item 61, apresentada na figura 5.6 revelou que este é o item mais fácil da prova de matemática apesar de não ser considerado de fato como tal, uma vez que ele apresenta um percentual de acerto de 53,6% relativamente baixo para um item fácil e uma dificuldade mediana ($b = 0,139$). Assim, este é um item de dificuldade média. O índice de discriminação desse item também é muito bom (a=1,764) e a probabilidade de acerto casual é baixa (c=0,075). Assim, esse item pode ser classificado como bom.

Figura 5.6: Curva Característica do item 61 de matemática (item bom)



5.2.3 Análise das Habilidades dos indivíduos

Os dados da Tabela 5.8 apresentam algumas medidas descritivas das habilidades dos alunos nas disciplinas de português e matemática. Vale ressaltar que a escala da habilidade varia de $-4,0$ a $+4,0$. Nas Figuras 5.7 e 5.8, encontram-se os histogramas dos valores das habilidades/proficiências que ilustram claramente a assimetria das duas distribuições analisadas, assimetria esta bem mais acentuada em matemática.

Tabela 5.8: Estatísticas descritivas das proficiências/habilidades de Português e Matemática

Medida	Português	Matemática
Média	-0,118	-0,162
Desvio Padrão	1,364	1,425
Mínimo	-4,0	-4,0
Máximo	4,0	4,0

Os dados mostram que a habilidade média em português (-0,118) foi maior que em matemática (-0,162) e o pico no menor intervalo na Figura 5.7, mostra que mais de 500 alunos obtiveram

desempenho mínimo (habilidade -4,0) em português.

Figura 5.7: Histograma das proficiências/habilidades de Português

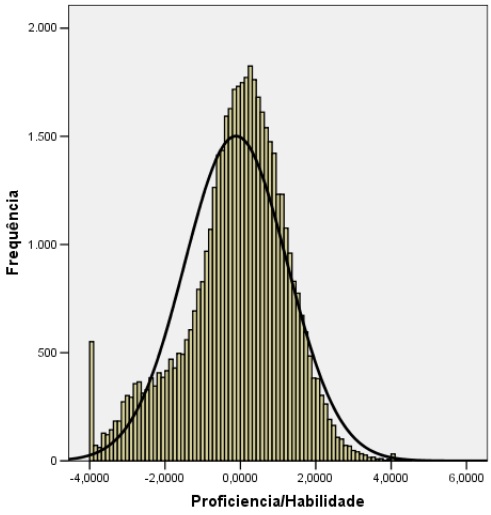
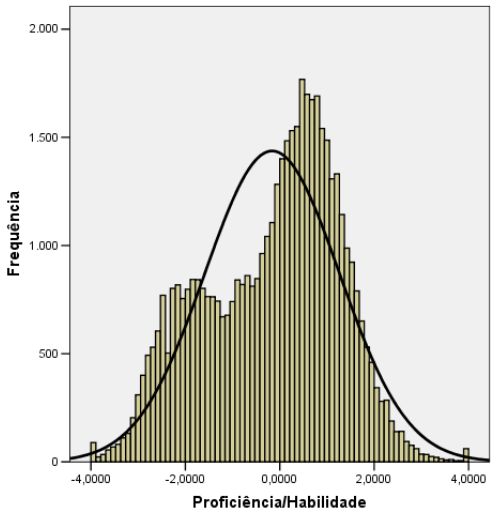


Figura 5.8: Histograma das proficiências/habilidades de Matemática



Análise por Sexo

Separando os estudantes por gênero e analisando suas habilidades, pode-se perceber através da Tabela 5.9, que o desempenho médio dos alunos do sexo masculino em português foi inferior

aos alunos do sexo feminino, com uma habilidade média de -0.205 para os homens e -0.016 para as mulheres e a mediana mostra que 50% das mulheres tiveram habilidade superior a 0,129. O teste de Mann Whitney aplicado em uma subamostra de 5% do total dos alunos comprova que essa diferença é estatisticamente significativa. O teste foi realizado numa subamostra uma vez que amostras muito grandes, em geral, levam sempre a rejeição da hipótese de igualdade.

Tabela 5.9: Média e mediana das habilidades dos alunos por sexo de Português

Sexo	**Número de alunos**	**Média da habilidade**	**Mediana**
Masculino	17814	-0,205	-0,064
Feminino	29260	-0,016	0,129
p-valor=0,000 do teste de Mann Whitney			

No entanto, em matemática ocorreu o inverso, ou seja, a Tabela 5.10 mostra que o desempenho dos alunos do sexo masculino (-0.093) foi melhor que os do sexo feminino (-0.187) e a mediana mostra que 50% dos homens tiveram habilidade superior a 0.171. A diferença não é estatisticamente significativa ao nível de 5% pelo teste de Mann Whitney aplicado em uma subamostra de 5% do total dos alunos. Como as medianas foram positivas e as médias foram negativas, isto indica que a maior parte dos alunos tiveram habilidade superior a 0, porém os alunos que tiveram habilidade inferior a 0, tiveram notas muito baixas, o que confirma a assimetria da distribuição.

Tabela 5.10: Média e mediana das habilidades dos alunos por sexo de Matemática

Sexo	**Número de alunos**	**Média da habilidade**	**Mediana**
Masculino	17814	-0,093	0,171
Feminino	29260	-0,187	0,082
p-valor=0,090 do teste de Mann Whitney			

Análise por Região

Os estudantes pesquisados foram separados em função da localização da escola segundo Região Metropolitana do Recife e Não-Metropolitana.

Os dados apresentados na Tabela 5.11, a seguir, mostram que em média, os alunos das es-

colas da Região Metropolitana obtiveram melhor desempenho na prova de português em relação aos alunos que pertencem a região Não-Metropolitana, sendo que esta diferença não é estatisticamente significativa ao nível de 5% a partir do teste de Mann Whitney aplicado também em uma subamostra de 5% do total dos alunos. Também, a nota mediana dos estudantes das escolas da região metropolitana é de 0,082 bem maior que a nota mediana das escolas da região não-metropolitana (0,01).

Tabela 5.11: Média e mediana das habilidades dos alunos da Região Metropolitana e Não-Metropolitana de Português

Região	**Número de alunos**	**Média da habilidade**	**Mediana**
Metropolitana	18924	-0,066	0,082
Não-Metropolitana	28134	-0,145	0,010
$p - valor = 0,641$ do teste de Mann Whitney			

Quanto a comparação das habilidades em matemática, os dados da Tabela 5.12, a seguir, revelam que em média os alunos que pertencem a Região Metropolitana tiveram melhor desempenho em relação aos alunos da Região Não-Metropolitana. Porém, esta diferença não é estatisticamente significativa ao nível de 5% de significância.

Tabela 5.12: Média e mediana das habilidades dos alunos da Região Metropolitana e Não-metropolitana de Matemática

Região	**Número de alunos**	**Média da habilidade**	**Mediana**
Metropolitana	18930	-0,143	0,106
Não-Metropolitana	28139	-0,170	0,106
p=0,312 do teste de Mann Whitney			

Análise por Grupo de Escola

Para tentar avaliar se a habilidade dos alunos está associada as condições das escolas, foi então aplicada a técnica estatística de análise de agrupamento a fim de agrupar as 7142 escolas

pesquisadas no SAEPE 2005.

Dentro do SAEPE 2005, conforme já referido, além das provas de português e matemática aplicadas aos alunos, foram também obtidas informações das escolas através de um questionário estruturado contendo questões referentes a situação e estado de conservação ou adequação quanto ao prédio da escola, às salas de aula e aos equipamentos existentes.

As variáveis relativas aos aspectos que qualificam o estado de conservação do prédio, são variáveis ordinais numa escala desde: **A** [bom(valor=4)], **B** [regular(valor=3)], **C** [ruim (valor=2)] até **D** [não se aplica/não existe (valor=1)].

Para efeito do agrupamento das escolas foi então definida a variável *conservação* a qual corresponde a soma das respostas referentes a conservação do: telhado, alvenaria (paredes), piso, esquadrias (portas e janelas), instalações hidráulicas e elétrica, pintura, muros (fechamentos) e vidros.

Para caracterizar a situação das condições de funcionamento das *salas* de aula foram somadas as respostas das perguntas em relação aos aspectos: iluminação natural e artificial, ventilação, quadro de giz ou branco, bancas de alunos e mesa do professor, na mesma escala do estado de conservação do prédio.

Também foi considerada a variável *sanitários*, variável ordinal, que corresponde à existência e condições de funcionamento de sanitários para os alunos, assumindo as categorias: não existe, ruim, regular e bom.

Além dessas três variáveis foram também utilizadas as seguintes variáveis binárias (sim:1 e não:0):

- **biblioteca**: existência ou não de biblioteca na escola;
- **computador**: existência ou não de computadores na escola;
- **segurança**: existência ou não de muros, grades ou cercas em condições de garantir segurança aos alunos;
- **proteção**: indica se a escola tem algum sistema de proteção contra incêndio (alarmes de

fumaça ou temperatura, extintores, mangueiras, etc.);

- **depredação**: indica se há ou não sinais de depredação;
- **limpeza**: indica se a escola apresenta-se limpa e bem ordenada;
- **PPP**: a escola possui ou não Projeto Político Pedagógico;
- **APM**: a escola tem ou não Associação de Pais e Mestres;
- **CE**: a escola possui ou não Conselho Escolar.

As escolas foram agrupadas em 5 grupos, de acordo com as características apresentadas pelas variáveis em estudo, através do método não-hierárquico k-médias. Tal método consiste, basicamente, em alocar cada elemento amostral àquele agrupamento cujo centróide (vetor de médias amostral) é o mais próximo do vetor de valores observados para o respectivo elemento.

Os resultados apresentados nas tabelas 5.13 e 5.14 mostram que dos grupos formados, o grupo 5 é constituído de 1473 (23,28%) escolas e o grupo 1 que agrupa 1831 (28,94%) apresentaram melhores condições referentes aos aspectos abordados, ou seja, apresentaram maiores valores para quase todas as variáveis. Já o grupo 3 formado por 879 (13,89%) escolas e o grupo 2 formado por 1128 (17,83%) reúne as escolas em piores condições, pois apresentam os menores valores para quase todas as variáveis.

A partir dos grupos constituídos, foi então analisadas as habilidades dos alunos em português e em matemática.

Tabela 5.13: Número de escolas e médias das variáveis nos 5 grupos

Grupos	**N**	**Conservação**	**Salas**	**Sanitário**
1 (melhor)	1831	18,83	15,22	2,12
2 (pior)	1128	12,71	13,95	1,55
3 (pior)	879	9,10	9,14	1,00
4	1016	16,44	10,32	1,63
5 (melhor)	1473	24,42	15,87	2,51

Tabela 5.14: Proporção das variáveis binárias nos 5 grupos

Grupos	1	2	3	4	5
Biblioteca	0,22	0,09	0,08	0,24	0,41
Computador	0,21	0,08	0,08	0,28	0,46
Segurança	0,41	0,15	0,10	0,38	0,79
Proteção	0,06	0,02	0,01	0,06	0,17
Depredação	0,22	0,35	0,48	0,32	0,16
Limpeza	0,9	0,79	0,6	0,81	0,96
PPP	0,6	0,45	0,39	0,6	0,75
APM	0,23	0,18	0,16	0,21	0,25
CE	0,47	0,29	0,26	0,52	0,71

Os dados ilustrados nas tabelas 5.15 e 5.16 revelam que os alunos pertencentes as escolas do grupo 2, apesar de ser considerado o grupo que apresenta as piores condições em relação às variáveis estudadas, em média, obtiveram melhor desempenho nas provas de português e matemática. Na prova de português, os alunos das escolas do grupo 4 apresentaram em média o pior desempenho, já na prova de matemática o grupo 3 apresentou o pior desempenho em média.

Tabela 5.15: Médias e mediana das habilidades de português por grupo de escolas

Grupo	**Número de alunos**	**Média**	**Mediana**
1 (melhor)	10925	-0,1049	0,0358
2 (pior)	2390	-0,0679	0,0896
3 (pior)	2301	-0,1000	0,0452
4	9637	-0,1663	-0,0156
5 (melhor)	19179	-0,0980	0,0492
$p - valor = 0,090$ do teste de Kruskal Wallis			

Tabela 5.16: Médias e medianas das habilidades de matemática por grupo de escolas

Grupo	**Número de alunos**	**Média**	**Mediana**
1	10925	-0,1167	0,1333
2	2390	-0,0704	0,1948
3	2303	-0,2329	0,0103
4	9638	-0,1972	0,0741
5	19187	-0,1685	0,1071
$p - valor = 0,626$ do teste de Kruskal Wallis			

A análise destas diferenças foi feita a partir da aplicação do teste de Kruskal Wallis, aplicado numa subamostra de 5% da amostra de alunos e, os resultados revelaram que essas diferenças não são estatísticamente significativas.

CAPÍTULO 6

Conclusões e Sugestões de Trabalhos Futuros

O resultado da análise das habilidades dos estudantes da 3^a série do ensino médio de Pernambuco avaliados pelo SAEPE-2005 nas disciplinas de português e matemática, revelaram que:

- a habilidade média dos estudantes de $-0,118$ em Português e $-0,162$ em Matemática (numa escala de -4,0 a 4,0);
- as mulheres apresentaram em média, habilidades em Português estatisticamente maiores que os homens, ocorrendo o contrário em matemática;
- as habilidades dos estudantes residentes na Região Metropolitana do Recife são em média superiores a dos residentes no interior, tanto em Português quanto em Matemática mas, essas diferenças não são estatisticamente significativas;
- as habilidades, em média, não variam segundo as condições físicas das escolas.

Na análise dos resultados de cada um dos itens das provas de português e matemática feita usando a TCT e a TRI, pode-se destacar:

- foram identificados muitos itens com índice de dificuldade elevado (38% da prova de português e 80% da prova de matemática), o que sugere o baixo nível do conhecimento exigido para resolver principalmente os itens de matemática;

- que uma pequena quantidade de itens (7,1% em português e 13,1% em matemática) apresentaram índice de discriminação inadequada.

6.1 Sugestões de Trabalhos Futuros

- Aplicar a equalização para comparar o desempenho dos alunos nos anos de 2002 e 2005 em Pernambuco;
- Fazer essas análises com as séries do Ensino Fundamental.
- Implementar rotinas no R que possibilite aplicar a TRI.

Apêndice A

PROGRAMA BILOG UTILIZADO PARA ANÁLISE DOS DADOS DO SAEPE 2005 - MATEMÁTICA

```
>COMENT
         saepe 2005 3 MATEMATICA.
>GLOBAL  DFNAME='SAEPE2005_3M_novo.DAT', NTEst =1, NPARM=3, LOGistic,
         SAVE;
>SAVE    SCORE='SAEPE2005_3M_novo.SCO',
         PARM='SAEPE2005_3M_novo.PAR';
>LENGTH  NITENS=(84);
>INPUT   NTOT=84, KFNAME='SAEPE2005_3M_novo.DAT',
         NIDCH=13,NALT=5, NFORM=21, NGRoup=1 ;
>ITENS   INUM=(1(1)84), INAME=(P01(1)P84);
>TEST    TNAME=P32005;
>FORM1   LENGTH=24,
         INUM=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
         18,19,20,21,22,23,24);
>FORM2   LENGTH=24,
         INUM=(13,14,15,16,17,18,19,20,21,22,23,24,
         25,26,27,28,29,30,31,32,33,34,35,36);
>FORM3   LENGTH=24,
         INUM=(25,26,27,28,29,30,31,32,33,34,35,36,
         37,38,39,40,41,42,43,44,45,46,47,48);
>FORM4   LENGTH=24,
         INUM=(37,38,39,40,41,42,43,44,45,46,47,48,
         49,50,51,52,53,54,55,56,57,58,59,60);
```

```
>FORM5   LENGTH=24,
          INUM=(49,50,51,52,53,54,55,56,57,58,59,60,
         61,62,63,64,65,66,67,68,69,70,71,72);
>FORM6   LENGTH=24,
         INUM=(61,62,63,64,65,66,67,68,69,70,71,72,
         73,74,75,76,77,78,79,80,81,82,83,84);
>FORM7   LENGTH=24,
         INUM=(73,74,75,76,77,78,79,80,81,82,83,84,
         1,2,3,4,5,6,7,8,9,10,11,12);
>FORM8   LENGTH=24,
         INUM=(1,2,3,4,5,6,7,8,9,10,11,12,
         25,26,27,28,29,30,31,32,33,34,35,36);
>FORM9   LENGTH=24,
         INUM=(13,14,15,16,17,18,19,20,21,22,23,24,
         37,38,39,40,41,42,43,44,45,46,47,48);
>FORM10   LENGTH=24,
         INUM=(25,26,27,28,29,30,31,32,33,34,35,36,
         49,50,51,52,53,54,55,56,57,58,59,60);
>FORM11   LENGTH=24,
         INUM=(37,38,39,40,41,42,43,44,45,46,47,48,
        61,62,63,64,65,66,67,68,69,70,71,72);
>FORM12  LENGTH=24,
         INUM=(49,50,51,52,53,54,55,56,57,58,59,60,
         73,74,75,76,77,78,79,80,81,82,83,84);
>FORM13   LENGTH=24,
         INUM=(61,62,63,64,65,66,67,68,69,70,71,72,
          1,2,3,4,5,6,7,8,9,10,11,12);
>FORM14   LENGTH=24,
```

```
          INUM=(73,74,75,76,77,78,79,80,81,82,83,84,
          13,14,15,16,17,18,19,20,21,22,23,24);
>FORM15   LENGTH=24,
          INUM=(1,2,3,4,5,6,7,8,9,10,11,12,
           37,38,39,40,41,42,43,44,45,46,47,48);
>FORM16   LENGTH=24,
          INUM=(13,14,15,16,17,18,19,20,21,22,23,24,
          49,50,51,52,53,54,55,56,57,58,59,60);
>FORM17   LENGTH=24,
          INUM=(25,26,27,28,29,30,31,32,33,34,35,36,
          61,62,63,64,65,66,67,68,69,70,71,72);
>FORM18   LENGTH=24,
          INUM=(37,38,39,40,41,42,43,44,45,46,47,48,
          73,74,75,76,77,78,79,80,81,82,83,84);
>FORM19   LENGTH=24,
          INUM=(49,50,51,52,53,54,55,56,57,58,59,60,
          1,2,3,4,5,6,7,8,9,10,11,12);
>FORM20   LENGTH=24,
          INUM=(61,62,63,64,65,66,67,68,69,70,71,72,
          13,14,15,16,17,18,19,20,21,22,23,24);
>FORM21   LENGTH=24,
          INUM=(73,74,75,76,77,78,79,80,81,82,83,84,
          25,26,27,28,29,30,31,32,33,34,35,36);
(13A1,I2,24A1)
>CALIB    NQPT=10, IDIst=0, TPRior,
          CYCLE=50, REFERENCE=1, NEWTON=10,
          CRIT=0.001, PLOT=1.0;
>SCORE    IDIST=0, METHOD=1, NOPRINT;
```

PROGRAMA BILOG UTILIZADO PARA ANÁLISE DOS DADOS DO SAEPE 2005 - PORTUGUÊS

```
>COMENT
         saepe 2005 3 PORTUGUES novo.
>GLOBAL  DFNAME='SAEPE2005_3P_novo.DAT', NTEst =1, NPARM=3, LOGistic,
         SAVE;
>SAVE    SCORE='SAEPE2005_3P_novo.SCO',
         PARM='SAEPE2005_3P_novo.PAR';
>LENGTH  NITENS=(84);
>INPUT   NTOT=84, KFNAME='SAEPE2005_3P_novo.DAT',
         NIDCH=13,NALT=5, NFORM=21, NGRoup=1 ;
>ITENS   INUM=(1(1)84), INAME=(P01(1)P84);
>TEST    TNAME=P32005;
>FORM1   LENGTH=24,
         INUM=(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
         18,19,20,21,22,23,24);
>FORM2   LENGTH=24,
         INUM=(13,14,15,16,17,18,19,20,21,22,23,24,
         25,26,27,28,29,30,31,32,33,34,35,36);
>FORM3   LENGTH=24,
         INUM=(25,26,27,28,29,30,31,32,33,34,35,36,
         37,38,39,40,41,42,43,44,45,46,47,48);
>FORM4   LENGTH=24,
         INUM=(37,38,39,40,41,42,43,44,45,46,47,48,
         49,50,51,52,53,54,55,56,57,58,59,60);
>FORM5   LENGTH=24,
```

```
          INUM=(49,50,51,52,53,54,55,56,57,58,59,60,
         61,62,63,64,65,66,67,68,69,70,71,72);
>FORM6   LENGTH=24,
         INUM=(61,62,63,64,65,66,67,68,69,70,71,72,
         73,74,75,76,77,78,79,80,81,82,83,84);
>FORM7   LENGTH=24,
         INUM=(73,74,75,76,77,78,79,80,81,82,83,84,
         1,2,3,4,5,6,7,8,9,10,11,12);
>FORM8   LENGTH=24,
         INUM=(1,2,3,4,5,6,7,8,9,10,11,12,
         25,26,27,28,29,30,31,32,33,34,35,36);
>FORM9   LENGTH=24,
         INUM=(13,14,15,16,17,18,19,20,21,22,23,24,
         37,38,39,40,41,42,43,44,45,46,47,48);
>FORM10   LENGTH=24,
         INUM=(25,26,27,28,29,30,31,32,33,34,35,36,
         49,50,51,52,53,54,55,56,57,58,59,60);
>FORM11   LENGTH=24,
         INUM=(37,38,39,40,41,42,43,44,45,46,47,48,
        61,62,63,64,65,66,67,68,69,70,71,72);
>FORM12  LENGTH=24,
         INUM=(49,50,51,52,53,54,55,56,57,58,59,60,
         73,74,75,76,77,78,79,80,81,82,83,84);
>FORM13   LENGTH=24,
         INUM=(61,62,63,64,65,66,67,68,69,70,71,72,
          1,2,3,4,5,6,7,8,9,10,11,12);
>FORM14   LENGTH=24,
         INUM=(73,74,75,76,77,78,79,80,81,82,83,84,
```

```
          13,14,15,16,17,18,19,20,21,22,23,24);
>FORM15   LENGTH=24,
          INUM=(1,2,3,4,5,6,7,8,9,10,11,12,
           37,38,39,40,41,42,43,44,45,46,47,48);
>FORM16   LENGTH=24,
          INUM=(13,14,15,16,17,18,19,20,21,22,23,24,
          49,50,51,52,53,54,55,56,57,58,59,60);
>FORM17   LENGTH=24,
          INUM=(25,26,27,28,29,30,31,32,33,34,35,36,
          61,62,63,64,65,66,67,68,69,70,71,72);
>FORM18   LENGTH=24,
          INUM=(37,38,39,40,41,42,43,44,45,46,47,48,
          73,74,75,76,77,78,79,80,81,82,83,84);
>FORM19   LENGTH=24,
          INUM=(49,50,51,52,53,54,55,56,57,58,59,60,
          1,2,3,4,5,6,7,8,9,10,11,12);
>FORM20   LENGTH=24,
          INUM=(61,62,63,64,65,66,67,68,69,70,71,72,
          13,14,15,16,17,18,19,20,21,22,23,24);
>FORM21   LENGTH=24,
          INUM=(73,74,75,76,77,78,79,80,81,82,83,84,
          25,26,27,28,29,30,31,32,33,34,35,36);
(13A1,I2,24A1)
>CALIB    NQPT=10, IDIst=0, TPRior,
          CYCLE=50, REFERENCE=1, NEWTON=10,
          CRIT=0.001, PLOT=1.0;
>SCORE    IDIST=0, METHOD=1, NOPRINT;
```

Apêndice B

QUESTIONÁRIO APLICADO PARA OBTENÇÃO DOS DADOS DO SAEPE

SAEPE 2005
SISTEMA DE AVALIAÇÃO EDUCACIONAL DE PERNAMBUCO

MINISTÉRIO DA EDUCAÇÃO - MEC

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA - INEP
AVALIAÇÃO NACIONAL DO RENDIMENTO ESCOLAR - ANRESC

GOVERNO DO ESTADO DE PERNAMBUCO
SECRETARIA DE EDUCAÇÃO E CULTURA
SECRETARIA EXECUTIVA DE DESENVOLVIMENTO DA EDUCAÇÃO
GERÊNCIA DE MONITORAMENTO DA QUALIDADE DO ENSINO
GERÊNCIA DE AVALIAÇÃO

QUESTIONÁRIO DA ESCOLA

Prezado(a) Coordenador(a) da Escola:

A Secretaria de Educação e Cultura do Estado de Pernambuco, em regime de colaboração com as Secretarias Municipais de Educação do Estado, representadas pela União Nacional dos Dirigentes Municipais de Educação (UNDIME) e em convênio com o Ministério da Educação, através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP, está coletando dados para o Sistema de Avaliação Educacional de Pernambuco (SAEPE). A finalidade é apresentar às escolas informações e elementos para melhorar os serviços educacionais do Estado.

É importante ressaltar que não é nosso objetivo avaliar individualmente diretores, professores ou alunos. As informações só serão divulgadas por Escola, Município ou Região. Por isso, damos garantia absoluta de sigilo sobre as informações fornecidas neste questionário.

Você deverá observar e indagar sobre diversos aspectos referentes à existência, situação e estado de conservação ou adequação quanto ao prédio da escola, às salas de aula e aos equipamentos existentes. Os critérios de julgamento deverão ser os seguintes:

A – BOM: o aspecto julgado está em condições de ser utilizado, em bom estado.

B – REGULAR: o aspecto julgado necessita de pequenos reparos, parcialmente suficiente/ satisfatório.

C – RUIM: o aspecto julgado necessita de grande reforma, ou totalmente insuficiente.

D – NÃO SE APLICA/NÃO EXISTE: o aspecto não existe ou o critério não é aplicável.

Agradecemos de antemão sua participação no processo e esperamos, juntos, melhorar a qualidade da educação em nosso Estado.

GOVERNO DO ESTADO
SECRETARIA DE EDUCAÇÃO E CULTURA
INEP
Ministério da Educação
BRASIL UM PAÍS DE TODOS GOVERNO FEDERAL
UNDIME PE
União Nacional dos Dirigentes Municipais de Educação

Indique o estado de conservação dos seguintes aspectos referentes ao prédio da escola:

	BOM	REGULAR	RUIM	NÃO EXISTE
01. Telhado	(A)	(B)	(C)	(D)
02. Alvenaria/ paredes	(A)	(B)	(C)	(D)
03. Piso	(A)	(B)	(C)	(D)
04. Esquadrias: portas e janelas	(A)	(B)	(C)	(D)
05. Instalações hidráulicas	(A)	(B)	(C)	(D)
06. Instalações elétricas	(A)	(B)	(C)	(D)
07. Pintura	(A)	(B)	(C)	(D)
08. Muros / fechamentos	(A)	(B)	(C)	(D)
09. Vidros	(A)	(B)	(C)	(D)

Observe as condições de funcionamento das salas de aula e indique a situação:

	BOM	REGULAR	RUIM	NÃO EXISTE
10. Iluminação natural	(A)	(B)	(C)	(D)
11. Iluminação artificial	(A)	(B)	(C)	(D)
12. Ventilação	(A)	(B)	(C)	(D)
13. Quadro de giz ou branco	(A)	(B)	(C)	(D)
14. Bancas de alunos	(A)	(B)	(C)	(D)
15. Mesa do professor	(A)	(B)	(C)	(D)
16. Armários	(A)	(B)	(C)	(D)

Indique a existência e condições de funcionamento das seguintes instalações:

	BOM	REGULAR	RUIM	NÃO EXISTE
17. Biblioteca	(A)	(B)	(C)	(D)
18. Laboratório de ciências	(A)	(B)	(C)	(D)
19. Laboratório de informática	(A)	(B)	(C)	(D)
20. Oficinas (artes, marcenaria)	(A)	(B)	(C)	(D)
21. Auditório	(A)	(B)	(C)	(D)
22. Quadras de esportes/ginásio	(A)	(B)	(C)	(D)
23. Área de recreio	(A)	(B)	(C)	(D)
24. Área de recreio coberta	(A)	(B)	(C)	(D)
25. Vestiários	(A)	(B)	(C)	(D)
26. Sala do professor	(A)	(B)	(C)	(D)
27. Sala da direção	(A)	(B)	(C)	(D)
28. Secretaria / administração	(A)	(B)	(C)	(D)
29. Cozinha	(A)	(B)	(C)	(D)
30. Despensa	(A)	(B)	(C)	(D)
31. Sala de leitura	(A)	(B)	(C)	(D)
32. Sanitários de alunos	(A)	(B)	(C)	(D)
33. Sanitários de funcionários	(A)	(B)	(C)	(D)
34. Refeitório	(A)	(B)	(C)	(D)

Indique a existência e estado de conservação dos seguintes equipamentos da escola:

	BOM	REGULAR	RUIM	NÃO EXISTE
35. Televisão	(A)	(B)	(C)	(D)
36. Vídeo-cassete	(A)	(B)	(C)	(D)
37. Antena parabólica	(A)	(B)	(C)	(D)
38. Mimeógrafo	(A)	(B)	(C)	(D)
39. Máquina fotocopiadora	(A)	(B)	(C)	(D)
40. Máquina fotográfica	(A)	(B)	(C)	(D)
41. Filmadora	(A)	(B)	(C)	(D)
42. Projetor de *slides*	(A)	(B)	(C)	(D)
43. Retroprojetor	(A)	(B)	(C)	(D)
44. Máquina de datilografia	(A)	(B)	(C)	(D)
45. Aparelho de som	(A)	(B)	(C)	(D)
46. Telefones	(A)	(B)	(C)	(D)
47. Computadores na administração	(A)	(B)	(C)	(D)
48. Computadores para alunos	(A)	(B)	(C)	(D)
49. DVD	(A)	(B)	(C)	(D)
50. *Data-show*	(A)	(B)	(C)	(D)

51. Se tem computadores para uso dos alunos, quantos computadores a escola tem em bom estado?

_____

52. Quantos títulos de livros, aproximadamente, a escola tem para consulta dos alunos (na biblioteca, sala de leitura, salas de aula, prateleiras, etc.)?

(A) Didáticos: _____

(B) Não didáticos: _____

53. Existem muros, grades ou cercas em condições de garantir a segurança dos alunos? *(Caso existam buracos ou aberturas que permitam o acesso de pessoas estranhas, a resposta é NÃO)*
(A) SIM.
(B) NÃO.

54. A escola tem algum sistema de proteção contra incêndio (alarmes de fumaça ou temperatura, extintores, mangueiras, etc.)?
(A) SIM.
(B) NÃO.

55. Que tipo de identificação é utilizado para o acesso do aluno à escola?
(A) Farda.
(B) Crachá.
(C) Nenhum.
(D) Outros.

56. As salas onde são guardados os equipamentos mais caros (computadores, projetores, vídeo, etc.) têm dispositivos para trancá-las (cadeados, grades, travas, etc.)?
(A) SIM.
(B) NÃO.

57. A escola apresenta sinais de depredação (vidros, portas, janelas ou lâmpadas quebradas)?
(A) SIM.
(B) NÃO.

58. A escola apresenta pichação externa?
(A) SIM.
(B) NÃO.

59. A escola apresenta pichação interna?
(A) SIM.
(B) NÃO.

60. A grafitagem é utilizada, na escola, como forma de expressão artístico-cultural?
(A) SIM.
(B) NÃO.

61. A escola apresenta-se limpa e bem ordenada?
(A) SIM.
(B) NÃO.

62. A escola tem Regimento Escolar?
(A) SIM.
(B) NÃO.

63. A Escola possui Projeto Político Pedagógico?
(A) SIM.
(B) NÃO.

64. A escola tem Associação de Pais e mestres (APM)?
(A) SIM.
(B) NÃO.

65. A Escola possui Grêmio Estudantil?
(A) SIM.
(B) NÃO.

66. A Escola possui Conselho Escolar?
(A) SIM.
(B) NÃO.

As questões a seguir só deverão ser respondidas se a escola tiver turmas do Projeto Alfabetizar com Sucesso.

67. Quantas turmas da 1ª etapa do I Ciclo DO PROJETO ALFABETIZAR COM SUCESSO funcionam atualmente nesta escola/estabelecimento de ensino?

_____ turmas.

68. Quantas turmas da 2ª etapa do I Ciclo DO PROJETO ALFABETIZAR COM SUCESSO funcionam atualmente nesta escola/estabelecimento de ensino?

_____ turmas.

69. Quantas turmas da 1ª etapa do II Ciclo do PROJETO ALFABETIZAR COM SUCESSO funcionam atualmente nesta escola/estabelecimento de ensino?

_____ **turmas.**

70. As turmas do PROJETO ALFABETIZAR COM SUCESSO funcionam
(A) apenas no turno da manhã.
(B) apenas no turno da tarde.
(C) apenas no turno da noite.
(D) apenas nos turnos da manhã e tarde.
(E) apenas nos turnos da manhã e noite.
(F) apenas nos turnos da tarde e noite.
(G) Nos três turnos: manhã, tarde e noite.

Referências Bibliográficas

Andersen, E. B. (1980), *Discrete Statistical Models with Social Science Applications*, New York: North-Holand Publishing Company.

Andrade, D. F. (1999), Comparando o Desempenho de Grupos (populações) de Respondentes através da Teoria da Resposta ao Item, PhD thesis, Tese apresentada ao Departamento de Estatística e Matemática Aplicada da UFC para o concurso de professor titular.

Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000), *Teoria da Resposta ao Item: Conceitos e Aplicações.*, 14 SINAPE.

Baker, F. B. (1992), *Item Response Theory - Parameter Estimation Techniques.*, New York: Marcel Dekker, Inc.

Baker, R. (1987), *Classical Test Theory nd Item Response Theory in Test Analysis.*, Special Report n.2: Language Testing Update. University of Edinburgh.

Birnbaum, A. (1957), *Efficient design and use of test of a mental ability for various decision-making problems. (Series Report No. 58-16).*, Washinghton, DC: USAF School of Aviation Medicine.

Birnbaum, A. (1968), Some latent trait models and their use in infering an examinee's ability. in f.m. lord & m. r. novick., *in* 'Statistical Theories of Mental Test Scores', Reding, MA: Addison-Wesley.

Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971), *Handbook on Formative and Summative Evaluation of student Learning*, New York: McGraw-Hill.

Bock, R. D. & Lieberman, M. (1970), 'Fitting a response model for n dichotomously scored items.', *Psychometrika* **35**, 179–197.

Chow, Y. S. & Teicher, H. (1978.), *Probability Theory: Independence, Interchangeability, Martingales*, New York: Springer-Verlag.

Condé, F. N. (2001), Análise empírica de itens, Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais - DAEB/INEP/MEC.

DeRoos, Y. & Allen-Meares, P. (1998), 'Applications of rash analysis: exploring differences in depression between african-american and white children.', *Journal of Social Service Research* **23**, 93–107.

Francisco, R. (2005), Aplicação da teoria da resposta ao item (tri), no exame nacional de cursos (enc) da unicentro., Master's thesis, Universidade Federal do Paraná.

Granger, C. V. & Deutsch, A. (1998), 'Rash analysis of the functional independence measure (fimtm) mastery tes.', *Arch. Phys. Med. Rehabil.* **79**, 52–57.

Guilford, J. P. (1936,1954), *Psychometric Methods*, New York: McGraw-Hill.

Gulliksen, H. (1950), *Theory of Mental Tests*, New York: John Wiley and Sons.

Hambleton, H. K., SWAMINATHAN, H., COOK, L. L., EIGNOR, D. R. & GIFFORD, J. A. (1978), 'Developments in latent trait theory: models, technical issues, and applications.', *Review of Educational Research* **48(4)**, 467–510.

Hambleton, R. K. & Swaminathan, H. (1985), *Item Response Theory: Principles and Applications.*, Boston: Kluwer Academic Publishers.

Hambleton, R. K. & Swaminathan, H. (1995), *Item Response Theory: Principles and Applications*, Boston: Kluwer. Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991), *Fundamentals of item response theory*, Newbury Park, CA: SAGE Publications.

Hambleton, R. K. & Van Der Linden, W. J. (1996), *Modern Item Response Theory.*, Springer.

Kolen, M. J. & Brennan, R. L. (1995), *Test Equating - Methods and Pratices*, New York: Springer.

Lawley, D. N. (1944), 'The factorial analysis of multiple item tests.', *Proceedings of the Royal Society of Edinburgh* **62-A**, 74–82.

Lawley, D. N. & Richardson, M. W. (1943), 'On problems connected with item selection and test construction', *Proceedings of the Royal Society of Edinburgh* **61-A**, 273–287.

Lord, F. M. (1952), 'The relation of the reliability of multiplechoice tests to the distribution or item difficulties.', *Psychometrika* **17**, 181–194.

Lord, F. M. (1980), *Aplications of item response theory to practical testing problems*, New Jersey:Lawrence Erlbaum Associates.

Mcintire, S. A. & Miller, L. A. (2000), *Foundations of Psychological Testing*, New York: McGraw-Hill.

Mendoza, C. E. F., Abad, F. J. & Lelé, A. J. (2005), 'Análise de itens do desenho da figura humana: aplicação de tri', *Psicologia: Teoria e Pesquisa* **vol.21**, nº 2.

Mislevy, R. J. & Bock, R. D. (1984.), *BILOG: Maximum likelihood item analysis and test scoring logistic models.*, Mooresville, In: Scientific Software.

Muniz, J. (1994), *Teoria Clásica de los tests.*, Madrid: Ediciones Pirámide, S. A.

Nojosa, R. T. (2001), Modelos multidimensionais para a teoria da resposta ao item, Master's thesis, Universidade Federal de Pernambuco.

Nunes, C. H. S. S. & Primi, R. (2005), 'Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item', *Avaliação Psicológica* **4(2)**, pp. 141–153.

Pasquali, L. (1997), *Psicometria: Teoria e Aplicações.*, Brasília: Editora da Universidade de Brasília.

Pasquali, L. & Primi, R. (2003), 'Fundamentos da teoria da resposta ao item - tri', *Avaliação Psicológica* **(2)2**, pp. 99–110.

Rabello, G. C. (2001), A metodologia de equalização e o sistema nacional de avaliação da educação básica (saeb), Technical report, Texto apresentado na série de Seminários "SAEB 2001 - Estratégias para a Ação".

Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests.*, Copenhagem: Danish Institute for Educational Research.

Richardson, M. W. (1936), 'The relation between the difficulty and the differential validity of a test.', *Psichometrika* **1**, 33–49.

Rodrigues, M. M. M. (2006), 'Proposta de análise de itens das provas do saeb sob a perspectiva pedagógica e a psicométrica', *Estudos em Avaliação Educacional* **v.17**, n.34.

Samejima, F. (1972), *A general model for tree-response data (Psychometric Monograph, Nº 18)*, Psychometric Society.

Swaminathan, H. & Gifford, J. A. (1983), *Estimation of Parameters in the Three-Prameter Latent Trait Model. In D. Weiss (Ed.)*, New York: Academic Press.

Thissen, D. (1991), *MULTILOG Version 6 User's Guide*, Mooresville, IN: Scientific Software, Inc.

Tucker, L. R. (1946), 'Maximum validity of a test with equivalent items.', *Psychometrika* **11**, 1–13.

Valle, R. C. (1999), Teoria da resposta ao item., Master's thesis, São Paulo: IME/USP.

Van Der Linden, W. J. & Hambleton, R. K. (1997), *Handbook of Modern Item Response Theory.*, New York: Springer-Verlag.

Vianna, H. M. (1982), *Testes em Educação*, São Paulo: Ibrasa.

Wainer, H. (1989), 'The future of item analysis.', *Journal of educational Measurement* **26(2)**, 191–208.

Wilson, D. T., Wood, R. & Gibbons, R. (1991), *TESTFACT: test scoring, item statistics, and item factor analysis.*, Chicago: Scientific Software.

Wright, B. D., Mead, R. J. & Bell, S. R. (1979.), *BICAL: a Rasch Program for the Analysis of Dichotomus Data*, Chicago: MESA.

Yang, W. & Houang, R. T. (1996), The effect of anchor lenght and equating method on the accuracy of test equating: comparisons of linear and irt-based equating using anchor-item design. paper presented at the aera annual conference, division d., new york. zimowski, m. f., muraki, e., mislevy, r. j. & bock, r. d. (1996)., *in* 'BILOG-MG. Multiple-Group IRT analysis and test maintenance for binary items. Chicago: Scientific Software International.'.

Zimowski, M. F., Muraki, E., Mislevy, R. J. & BockK, R. D. (1996), *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items.*, Chicago: Scientific Software, Inc.