
**MODELOS MULTIDIMENSIONAIS PARA A
TEORIA DA RESPOSTA AO ITEM**

RONALD TARGINO NOJOSA

Bacharel em Estatística, Universidade Federal do Ceará, 1997

Orientadores

Prof. Dr. Dalton Francisco de Andrade

Prof. Dr. Francisco Cribari Neto

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 2001

Aos meus pais, José Alves e Maria das Dores,
e a minha noiva, Francilene Caetano, minha homenagem.

Dedico este trabalho a
Maurício Mota.

Agradecimentos

Ao Professor Dalton Francisco de Andrade pela orientação, amizade e confiança transmitida durante a elaboração desta dissertação de mestrado.

Ao Professor Francisco Cribari Neto pela orientação e apoio a esta dissertação de mestrado.

Ao Professor João Maurício Araújo Mota pela sua amizade, confiança, incentivo e grandiosa contribuição à minha formação como Estatístico.

Aos professores do Programa de Mestrado em Estatística da UFPE pelo conhecimento transmitido durante meu curso de mestrado, em especial ao Prof. Klaus Leite Pinto Vasconcellos.

Aos colegas do curso de pós-graduação com os quais dividi bons e difíceis momentos, em especial a Gilberto da Silva Matos.

A Francilene Caetano e a Solange Targino pelo apoio fundamental durante o desenvolvimento deste trabalho.

A Valéria e a Raquel pela atenção dispensada durante o curso de pós-graduação.

Aos demais colegas dos Departamentos de Estatística da UFPE e da UFC que contribuíram direta ou indiretamente para realização deste trabalho.

Ao INEP/MEC, em particular a professora Maria Inês Fini, coordenadora do ENEM, pela concessão do banco de dados utilizado neste trabalho.

Aos participantes da banca examinadora pelas sugestões.

A CAPES pelo apoio financeiro.

Resumo

A Teoria da Resposta ao Item (TRI) originou-se na metade da década de 30, mas só recentemente está sendo utilizada em situações práticas, devido às dificuldades das análises estatísticas envolvidas e à falta de programas computacionais específicos. Vários são os modelos de resposta ao item, que se diferenciam na forma da função característica do item e/ou no número de parâmetros atribuídos ao modelo. Atualmente, os modelos utilizados em situações práticas limitam-se a avaliações que medem um traço latente (habilidade) que pode ser representado por uma única dimensão (avaliações unidimensionais). Algumas avaliações, seja pela construção dos itens ou pela própria finalidade da aplicação, não podem, a princípio, serem consideradas unidimensionais. Este é o caso do Exame Nacional do Ensino Médio - ENEM, que foi estruturado segundo uma matriz de 5 competências. Este trabalho apresenta, primeiramente, um breve relato da Teoria Clássica de Medidas apontando suas deficiências e introduzindo a Teoria da Resposta ao Item unidimensional. Em seguida são apresentadas técnicas destinadas à determinação da dimensionalidade de um conjunto de dados. A Análise Fatorial de Informação Plena está entre essas técnicas e é proposta como ferramenta para a verificação da dimensionalidade do ENEM. Uma grande vantagem dessa análise é que ela permite testar a significância estatística dos fatores adicionados sucessivamente ao modelo, em outras palavras, testar a significância do aumento na dimensão de um conjunto de dados. Como critérios para a determinação da dimensionalidade foram utilizados: (i) o teste de qui-quadrado, (ii) a magnitude das cargas fatoriais e (iii) um índice de dimensionalidade. Como índice complementar utilizou-se a porcentagem da variância explicada pelos fatores. Os resultados desse estudo mostraram que o ENEM pode ser considerado um exame pentadimensional, ou seja, a habilidade medida pelo exame pode ser representada por pelo menos 5 dimensões. Mesmo em avaliações propostas unidimensionais, essa análise é extremamente importante dada a dificuldade na elaboração de itens que preservem a unidimensionalidade. O trabalho apresenta, ainda, modelos que tornam possível a modelagem de dados em avaliações multidimensionais e considera o modelo compensatório MC3 para modelar os dados do Exame Nacional do Ensino Médio do ano de 1999.

Abstract

The Item Response Theory was first introduced in the mid 1930s, but only recently it has seen applications. There are several different models in the Item Response Theory, and they differ in the assumptions about the structure of the item and the number of parameters in the model. The models currently being used seek to measure a latent trait (ability) which can be represented by only one dimension (unidimensionality). Some evaluations, however, do not fit in a one-dimensional framework. This is the case of the ‘Exame Nacional do Ensino Médio’ (ENEM), which has been designed according to a matrix of five different skills. At the outset, this dissertation presents a brief overview of the Classical Measurement Theory and of the unidimensional Item Response Theory. We also discuss some techniques that are helpful in determining the dimensionality of a given data set. One of these techniques is the Full Information Factorial Analysis, which we propose as a way to check the dimensionality of ENEM. The results suggest that of this exam is at least fivefold dimensional. We also present models designed to model multidimensional evaluations, and use the compensatory MC3 model to model data from the 1999 ENEM.

Índice

1. Introdução	1
1.1. Teoria Clássica	1
1.2. Teoria Clássica - Limitações	1
1.3. Teoria da Resposta ao Item (TRI) - Modelos Unidimensionais	2
1.3.1. Modelos Unidimensionais	4
1.3.2. Suposições	6
1.3.3. Estimaco	7
1.3.4. Escala de Medidas	7
1.3.5. Comentrios Adicionais	8
1.4. Organizao da Dissertao	8
2. Anlise Fatorial	9
2.1. Introduo	9
2.2. Modelo Fatorial Convencional	9
2.3. Estimaco dos Parmetros	11
2.4. Modelo Fatorial para Dados Dicotmicos ou Dicotomizados	12
2.4.1. O Coeficiente de Correlao Phi	12
2.4.2. A Correlao Tetracrica	14
2.4.3. Casos Heywood	18
2.4.4. Anlise Fatorial de Informao Plena	19
3. Modelos Multidimensionais	23
3.1. Introduo	23
3.2. Motivao	23
3.3. Modelos Matemticos	24
3.3.1. Modelo Compensatrio com Acerto Casual - MC3	24
3.3.2. Modelo Compensatrio sem Acerto Casual - MC2	28
3.3.3. Modelos No-Compensatrios	28
3.3.4. Modelos Ogiva-Normal Multidimensionais	28
4. Parmetro de Dificuldade Multidimensional - DIFICM	30
4.1. Introduo	30
4.2. Motivao	30
4.3. O Parmetro de Dificuldade Multidimensional	31

5. Parâmetro de Discriminação Multidimensional - DISCM.....	34
5.1. Introdução	34
5.2. Conceitualização	34
5.3. O Parâmetro de Discriminação Multidimensional	35
5.4. Função de Informação Multidimensional - INFM.....	37
6. Estimação dos Parâmetros Multidimensionais	39
6.1. Introdução	39
6.2. Estimação dos Parâmetros dos Itens	40
6.3. Estimação das Habilidades	43
6.4. Estimação Conjunta dos Parâmetros dos Itens e das Habilidades.....	45
7. Aplicação em Dados Reais	51
7.1. Introdução	51
7.2. Características do ENEM.....	51
7.3. Dados do ENEM.....	52
7.4. Dimensionalidade do ENEM.....	52
7.5. Análise dos itens do ENEM	56
8. Comentários e Sugestões Finais	58
Apêndice	60
Referências	64

Capítulo 1

Introdução

A Teoria da Resposta ao item (TRI) originou-se entre os anos de 1935 e 1940. Desenvolveu-se tendo como um dos objetivos suprir deficiências da Teoria Clássica de Medidas. A partir dos anos 70 e 80 a TRI passou a ser tópico de pesquisa dominante entre os especialistas em medidas. Mais recentemente, com o avanço computacional, as dificuldades das análises estatísticas foram minimizadas e essa teoria pode mostrar, na prática, sua contribuição em diversas áreas, das quais destaca-se a educacional.

Neste capítulo será apresentado um breve relato da Teoria Clássica de Medidas, ressaltando deficiências que incentivaram o surgimento e o desenvolvimento da TRI. Em seguida um resumo histórico da Teoria da Resposta ao Item, suas características, suposições, alguns dos modelos unidimensionais mais utilizados e, por último, a motivação para os modelos multidimensionais. Na parte final do capítulo será apresentada a organização dessa dissertação de mestrado.

1.1. Teoria Clássica

Qualquer modelo matemático inclui um conjunto de suposições sobre os dados para os quais o modelo é aplicado e especifica o relacionamento entre construtos observados e não observados descritos no modelo. No modelo clássico, dois construtos são introduzidos: o escore verdadeiro e o erro de medida. O escore verdadeiro para um indivíduo pode ser definido como um valor esperado dos seus escores obtidos em vários testes. O erro de medida pode ser definido como a diferença entre o escore verdadeiro e o observado. O modelo clássico supõe que: (1) os erros de medida são aleatórios com média zero e não correlacionados entre si e com os escores verdadeiros e (2) os escores verdadeiros, os observados e os erros de medida são linearmente relacionados. 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 seriam:

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.

1.2. Teoria Clássica - Limitações

A Teoria Clássica apresenta inúmeros resultados úteis, tais como a fórmula de Spearman-Brown e a fórmula-20 de Kuder-Richardson, ambas utilizadas para calcular a fidedig-

nidade¹ de um teste (Lord, 1980; Vianna, 1987). Entretanto, há várias deficiências no modelo clássico. Por exemplo, a dificuldade do item, que é a proporção de indivíduos que acertam ao item, e a discriminação do item, que é a correlação bisseral ou ponto-bisseral, dependem do grupo de indivíduos do qual elas são obtidas. Para uma amostra de indivíduos com habilidade acima da habilidade média, o índice de dificuldade do item será mais alto do que para uma outra amostra de indivíduos com habilidade abaixo da habilidade média. Também, o índice de discriminação tenderia a ser mais alto em amostras com indivíduos mais heterogêneos (em termo de habilidade) do que em amostras mais homogêneas. Isto ocorre por causa do efeito da amostra no coeficiente de correlação (Lord & Novick, 1968). Outro problema é que os escores observado e verdadeiro aumentam e diminuem dependendo da dificuldade do teste. Esse problema fica claro quando são comparados indivíduos submetidos a formas diferentes de um teste, por exemplo, teste com algumas questões diferentes.

Outras duas deficiências do modelo clássico dizem respeito à suposição de erros de medida iguais para todos os indivíduos e à definição de testes paralelos descritos, por exemplo, por Lord (1980) e Hambleton & Swaminathan (1985). Em relação à primeira delas, é suficiente notar que, em testes difíceis, os erros de medida são, provavelmente, maiores para indivíduos com baixa habilidade do que para indivíduos com habilidade alta. Com relação à segunda deficiência citada, satisfazer a definição de testes estritamente paralelos é difícil na prática. A suposição de paralelismo para testes não paralelos resulta em estimativas imprecisas da fidedignidade do teste, do erro-padrão de medida e do tamanho do teste necessário para alcançar a confiança ou fidedignidade necessária. Por último, a Teoria Clássica tem como base, para a avaliação e a seleção de indivíduos, o teste como um todo, ou seja, as análises e interpretações dependem do conjunto de itens que compõem o teste, o que impossibilita a comparação entre indivíduos não submetidos ao mesmo teste. É essa uma das principais críticas à Teoria Clássica.

Após esse breve relato, fica claro que o desejável seria ter (a) estatísticas de itens não dependentes do grupo, (b) escores que não dependessem da dificuldade do teste para descrever a habilidade dos indivíduos e (c) modelos que não requeiram testes estritamente paralelos para avaliar a confiança ou fidedignidade. Há agora evidências substanciais para sugerir que essas três propriedades desejáveis, além de outras, possam ser obtidas em uma outra estrutura de teoria de medidas, conhecida como Teoria da Resposta ao Item (Lord, 1980; Hambleton & Swaminathan, 1985).

1.3. Teoria da Resposta ao Item (TRI) - Modelos Unidimensionais

A origem da TRI pode ser datada da metade da década de 30 e início da década de 40. Em 1936, Richardson derivou o relacionamento entre os parâmetros do modelo da TRI e os parâmetros dos itens do modelo clássico, o que forneceu um caminho inicial para a obtenção das estimativas dos parâmetros para modelos na TRI. O trabalho de Lawley (1943) marcou o início da teoria baseada nos itens dos testes. Lawley mostrou

¹A fidedignidade de um teste refere-se à estabilidade dos seus resultados: se um teste é aplicado inúmeras vezes ao mesmo grupo de indivíduos espera-se que os resultados sejam os mesmos.

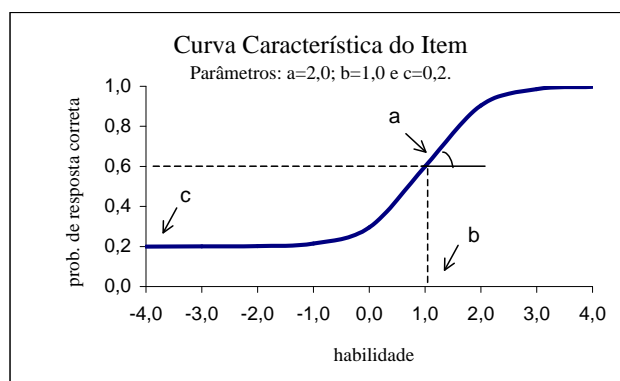
como obter as estimativas de máxima verossimilhança para os parâmetros do modelo de curva característica do item, definiu o escore verdadeiro em termos dos itens e do teste e mostrou que o coeficiente de confiança/fidedignidade pode ser expresso como função dos parâmetros dos itens. A principal extensão desse trabalho se deve a Lord (1952), que mostrou que as várias medidas da teoria clássica poderiam ser expressas como funções dos parâmetros da função característica do item. O trabalho destes dois pesquisadores estabeleceu os conceitos básicos da teoria psicométrica baseada em itens, agora conhecida como Teoria da Resposta ao Item (TRI).

A TRI sustenta-se em dois pilares:

- a performance de um indivíduo em um teste pode ser explicada por um conjunto de fatores chamados traços latentes ou habilidades;
- o relacionamento entre as respostas dos indivíduos a cada item e a habilidade medida pelo teste pode ser descrita por uma função monótona crescente, chamada Função Característica do Item (FCI). Essa função fornece a probabilidade de indivíduos de vários níveis de habilidade responderem corretamente ao item.

A Figura 1.1 mostra uma função característica de item quando o traço latente ou habilidade tem uma única dimensão (unidimensionalidade). Observe que indivíduos com altos valores para a habilidade têm maior probabilidade de responder corretamente ao item do que aqueles com baixa habilidade.

Figura 1.1 - Representação gráfica de uma FCI.



Existem vários modelos de respostas ao item, que se diferenciam na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos contêm um ou mais parâmetros relacionados ao item e um ou mais parâmetros relacionados ao indivíduo. De forma mais geral, os modelos da TRI podem ser organizados na seguinte estrutura:

- modelos unidimensionais ou multidimensionais em relação ao traço latente;
- modelos lineares ou não lineares;
- modelos de resposta dicotômica ou não dicotômica;
- modelos para uma ou mais populações.

Quando as suposições da Teoria da Resposta ao Item estão presentes (veja Subseção 1.3.2), ou pelo menos de forma aproximada a um conjunto de dados, as estimativas obtidas apresentam as seguintes propriedades: (a) assumindo a existência de um grande conjunto de itens, todos medindo o mesmo traço latente, as estimativas das habilidades dos indivíduos são independentes da particular amostra de itens aplicada aos indivíduos, (b) assumindo a existência de uma grande população de indivíduos, as medidas de caracterização dos itens (por exemplo, a dificuldade e a discriminação) são independentes da particular amostra de indivíduos empregada no processo de estimação dos parâmetros dos itens (calibração) e (c) uma estatística indicando a precisão com que cada habilidade foi estimada é fornecida.

Por fim, o conceito de formas paralelas de testes, comum à Teoria Clássica, é substituído por um método estatístico que permite a estimação de diferentes erros-padrão de medidas para indivíduos com diferentes níveis de habilidade.

Comparada à Teoria Clássica uma importante vantagem dos modelos de resposta ao item é que, dado um conjunto de itens que tenham sido ajustados a um modelo de resposta ao item (isto é, os parâmetros dos itens são conhecidos), é possível estimar as habilidades dos indivíduos na mesma escala (de habilidade) de qualquer subconjunto de itens no domínio dos itens que tenham sido ajustados ao modelo. O domínio dos itens diz respeito à homogeneidade no sentido de medir uma única habilidade.

1.3.1. Modelos Unidimensionais

Esta seção apresenta dois modelos unidimensionais para resposta dicotômica. Uma abordagem aos diversos modelos citados na seção anterior pode ser encontrado em Baker (1992) e Andrade, Tavares & Valle (2000). As principais diferenças entre os modelos está na forma matemática da Curva Característica do Item e no número de parâmetros do modelo. Esses modelos são expressões matemáticas que fornecem a probabilidade de resposta correta a um item como função da habilidade do indivíduo e dos parâmetros dos itens.

Modelo Logístico de 3 parâmetros

Dentre os modelos propostos na literatura, um dos mais utilizados é o modelo logístico de 3 parâmetros proposto por Birnbaum (1957). Esse modelo é empregado a itens de múltipla escolha do tipo certo/errado e é dado por

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

onde

X_{ij} é uma variável dicotômica que assume os valores 1, quando o j -ésimo indivíduo responde corretamente ao item i , ou 0 quando o j -ésimo indivíduo não responde corretamente ao item i , com $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, N$;

$P(X_{ij} = 1 \mid \theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente ao item i ;

θ_j representa a habilidade (traço latente, traço ou proficiência) do j -ésimo indivíduo;

a_i é o parâmetro de discriminação do item i ;

b_i é o parâmetro de dificuldade do item i , medido na mesma escala da habilidade;

c_i é o parâmetro que representa a probabilidade de um indivíduo com baixa habilidade responder corretamente ao item i (muitas vezes referido como a probabilidade de resposta correta dada ao acaso);

D é um fator de escala, constante igual a 1. Fazendo $D = 1,702$, aproxima-se a função logística da função ogiva normal (veja pág. 6).

A função $P(X_{ij} = 1 \mid \theta_j)$ assume valores entre 0 e 1 e tem forma S na escala da habilidade. 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 a_i tem valor proporcional à inclinação da Curva Característica do Item (CCI) no ponto b_i ; assim, valores negativos não são esperados, pois indicariam que a probabilidade de responder corretamente ao item diminui com o aumento da habilidade. O parâmetro b_i é o ponto na escala da habilidade onde a probabilidade de responder ao item corretamente é igual a $(1 + c_i)/2$. Assim, quando não é possível o acerto casual (“chute”), b_i representa o ponto na escala onde a probabilidade de acertar ao item é 0,5. O parâmetro c_i não depende da escala, e assume valores entre 0 e 1.

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 itens muito fáceis para esse grupo. Com relação ao parâmetro a , espera-se valores no intervalo $(0; 3,0)$. Valores próximos de 0 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, e para a escala normal esse intervalo é $(0,6; 1,6)$. Para os parâmetros de dificuldade são 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.

Em muitas situações, o acerto casual não é possível, e assim, trabalha-se com o parâmetro c igual a 0. Isto conduz ao modelo logístico de 2 parâmetros, também entre os mais aplicados.

Modelo Ogiva Normal de 3 parâmetros

Esse modelo foi proposto por Lord (1952) e é dado por

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta_j - b_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Com o trabalho de Haley (1952), a equivalência entre este modelo e o logístico de 3 parâmetros se dá para $D = 1,702$. Os resultados obtidos com os dois modelos diferem em menos de 0,01, o que pode ser considerado não-significante. Outros modelos unidimensionais e maiores detalhes podem ser encontrados em Hambleton & Swaminathan (1985).

1.3.2. Suposições

Os modelos unidimensionais têm como suposição fundamental a *unidimensionalidade* do teste, ou seja, supõem que o teste mede apenas um único traço latente (habilidade dominante no conjunto de itens). A *independência local ou condicional* fora citada em alguns trabalhos como suposição. Entretanto, com a publicação de Lord (1980), a independência não mais é vista como suposição, mas sim como conseqüência da correta determinação da dimensionalidade dos dados. A independência condicional significa que dada a habilidade do indivíduo, suas respostas aos diferentes itens do teste são independentes. Essa conseqüência é fundamental no processo de estimação dos parâmetros do modelo. A independência condicional pode ser obtida mesmo quando os dados não são unidimensionais, desde que todas as dimensões da habilidade que influenciam a performance no teste sejam consideradas. Por exemplo, considere um teste de matemática no qual os itens exigem níveis altos da habilidade verbal (interpretação de texto). Indivíduos com níveis baixos dessa habilidade podem não responder corretamente aos itens mesmo possuindo níveis altos da habilidade matemática. Dessa forma, a performance no teste é influenciada por ambas as habilidades: matemática e verbal. Se um modelo unidimensional é ajustado aos dados, considerando apenas a habilidade matemática, a independência condicional não pode ser considerada, pois as respostas aos diferentes itens dependerá da habilidade verbal. Por outro lado, se os indivíduos apresentam níveis próximos de habilidade verbal, considera-se que somente a habilidade matemática influencia na performance sobre os itens; assim, a independência condicional pode ser considerada quando um modelo unidimensional é ajustado. Em testes contendo itens que exigem o conhecimento de procedimentos específicos ou estratégicos para se obter a resposta correta ou que fornecem informações que ajudam a responder outros itens, a independência condicional também não se verifica. A habilidade para detectar procedimentos específicos pode ser vista como outra dimensão além da habilidade testada.

Outras duas suposições são: (a) o tempo para a resolução do teste é suficiente para que todos os itens possam ser respondidos por todos os indivíduos e (b) a ordem em que os itens são apresentados aos indivíduos não interfere no desempenho dos mesmos.

1.3.3. Estimação

Uma vez determinado o modelo da TRI a ser utilizado, é necessário determinar os valores dos parâmetros dos itens e das habilidades dos indivíduos. Nos modelos unidimensionais, cada indivíduo é caracterizado apenas por um parâmetro, θ , e para a caracterização dos itens utilizam-se de 1 a 4 parâmetros, dependendo do modelo utilizado. De modo geral, obtêm-se estimativas para os parâmetros do modelo com erros-padrão pequenos quando o número de itens é de pelo menos 30 e o número de respondentes para cada item é de pelo menos 300. No processo de estimação, nos defrontamos com 3 casos. O primeiro é quando os parâmetros dos itens são conhecidos e deseja-se apenas estimar as habilidades dos indivíduos. No segundo caso são conhecidas as habilidades e deseja-se apenas estimar os parâmetros dos itens. No terceiro, nem os parâmetros dos itens e nem as habilidades dos indivíduos são conhecidos; deseja-se estimar ambos. O primeiro caso começa a ser freqüente na prática, e a solução é dada empregando o método da máxima verossimilhança ou métodos bayesianos, ambos através da aplicação de procedimentos interativos, como, por exemplo, o método de Newton-Raphson. O segundo caso tem apenas caráter teórico e é solucionado usando o método da máxima verossimilhança. O terceiro caso, provavelmente o mais encontrado na prática, é abordado de duas formas: a estimação conjunta dos parâmetros de itens e das habilidades dos indivíduos; ou em duas etapas, primeiro a estimação dos itens e, em seguida, a das habilidades (Baker, 1992).

1.3.4. Escala de Medidas

Nos dois primeiros casos citados na subseção anterior, o conhecimento dos parâmetros (dos itens ou dos indivíduos) implica o conhecimento da escala em que eles foram medidos. No terceiro, não há nenhuma escala definida. Como a habilidade pode assumir qualquer valor real, os valores dos parâmetros que maximizam a função de verossimilhança não podem ser determinados de modo único. É necessário, então, estabelecer uma origem, que representará a média das habilidades dos indivíduos na população, e uma unidade de medida, que representará o desvio-padrão das habilidades, antes do processo de estimação. As habilidades dos indivíduos não são observadas, mas, baseado nas respostas dos indivíduos a um conjunto de itens, é possível estimá-las. A propriedade mais importante da habilidade é que ela não depende do particular conjunto de itens ao qual o indivíduo foi submetido. Essa propriedade possibilita uma comparação direta de itens, testes ou performances de diferentes grupos de indivíduos. A habilidade pode assumir teoricamente qualquer valor real, e muitas são as métricas na qual ela pode ser definida. Usualmente, utiliza-se uma escala com média 0 e desvio-padrão 1. Entretanto, poder-se-ia assumir quaisquer outros valores, pois o importante são as relações de ordem existentes entre os pontos da escala. A especificação da escala elimina a não-identificabilidade do modelo que ocorre devido à produção de um mesmo valor para a função característica do item a partir de diferentes conjuntos de parâmetros. Por exemplo, considerando o modelo logístico de 3 parâmetros, se $\theta'_j = \alpha\theta_j + \beta$, $b'_i = \alpha b_i + \beta$ e $a'_i = a_i/\alpha$, onde α e β são constantes reais com $\alpha > 0$, obtemos $P(X_{ij} = 1 | \theta'_j, b'_i, a'_i) = P(X_{ij} = 1 | \theta_j, b_i, a_i)$. Uma discussão e desenvolvimento completos podem ser encontrados em Baker (1992).

1.3.5. Comentários Adicionais

Os modelos unidimensionais da TRI têm como suposição fundamental a unidimensionalidade do teste. Esta suposição exige que os itens que irão compor o teste sejam elaborados de modo a atendê-la. Entretanto, dependendo do objetivo do teste, nem sempre isso é possível.

A violação da suposição de unidimensionalidade conduz, entre outras, às seguintes conseqüências negativas: (i) a própria validade do item passa a ser questionada e (ii) assumindo um modelo unidimensional com itens multidimensionais não se pode garantir a independência condicional.

Hattie (1985) identificou mais de 30 índices sugeridos para determinar a dimensionalidade de uma medida e os categorizou nos seguintes grupos:

- Grupo 1: índices baseados em padrões de respostas;
- Grupo 2: índices baseados na fidedignidade;
- Grupo 3: índices baseados na análise de componentes principais;
- Grupo 4: índices baseados na análise fatorial;
- Grupo 5: índices baseados na TRI.

Pela categorização de Hattie (1985), a análise fatorial através das matrizes dos coeficientes ϕ e das correlações tetracóricas se enquadram no grupo 4 e a análise fatorial de informação plena no grupo 5. No Capítulo 2 desta dissertação serão abordados esses dois grupos, com referência específica a essas análises.

1.4. Organização da Dissertação

O objetivo principal desta dissertação é inicializar o estudo da Teoria da Resposta ao Item Multidimensional no Brasil. Este estudo exigiu uma revisão dos métodos de Análise Fatorial visando a determinação da dimensionalidade de instrumentos de avaliação e conseqüentemente da dimensão dos modelos a serem ajustados. O Capítulo 1 foi dedicado à revisão de alguns conceitos da Teoria Clássica e da TRI unidimensional. O Capítulo 2 apresenta a análise fatorial convencional - que trata com variáveis contínuas - e a análise fatorial para dados dicotômicos ou dicotomizados. Nesse capítulo é desenvolvida a teoria para a determinação da dimensionalidade de um conjunto de dados. No Capítulo 3 são apresentados alguns modelos multidimensionais da TRI. O Capítulo 4 é dedicado ao desenvolvimento do parâmetro de dificuldade para modelos multidimensionais. O Capítulo 5 aborda o parâmetro de discriminação multidimensional e introduz o conceito de informação multidimensional. O Capítulo 6 trata da estimação dos parâmetros para modelos multidimensionais: parâmetros dos itens e habilidades dos indivíduos. O Capítulo 7 é dedicado a aplicação da teoria desenvolvida nos Capítulos 2, 3, 4, 5 e 6. Esse capítulo apresenta os resultados da avaliação da dimensionalidade do Exame Nacional do Ensino Médio (ENEM) do ano de 1999 e da modelagem dos dados deste Exame através do modelo compensatório MC3. O Capítulo 8 é dedicado às considerações gerais sobre a presente dissertação e algumas sugestões para a continuidade deste estudo. Por fim, alguns resultados citados no Capítulo 7 compõem o Apêndice.

Capítulo 2

Análise Fatorial

2.1. Introdução

A Análise Fatorial trata do relacionamento interno de um conjunto de variáveis. Suas idéias básicas se devem, principalmente, a psicólogos como Charles Spearman, Thomson, Thurstone e Burt, que buscavam obter uma melhor compreensão para a “inteligência”.

Os testes de inteligência eram - e ainda são - montados com uma grande variedade de itens que variam em graus de memorização, habilidade verbal, habilidade matemática, entre outras. A Análise Fatorial foi desenvolvida para analisar esses testes e averiguar se a “inteligência” era determinada por um único fator geral ou por vários fatores.

Na Análise Fatorial convencional, p variáveis observadas são modeladas como funções lineares de um número menor, m , de outras variáveis contínuas, denominadas variáveis latentes ou fatores. Essas variáveis são empregadas na estimação da correlação entre as variáveis observadas.

Os objetivos principais da Análise Fatorial são:

- determinar o número de fatores que fornecem um ajuste satisfatório à matriz de correlação observada;
- estimar os coeficientes de regressão das variáveis observadas nos fatores.

Com isso, almeja-se uma explicação parcimoniosa do relacionamento entre as variáveis observadas.

Basicamente, o modelo fatorial é motivado pelo seguinte argumento: Suponha variáveis que possam ser agrupadas por suas correlações, isto é, suponha que todas as variáveis de um grupo particular de variáveis são fortemente correlacionadas entre si, mas têm, relativamente, baixa correlação com variáveis de outros grupos diferentes. Então, é conceitual que cada grupo de variáveis representa uma única construção básica, ou fator, que é responsável pelas correlações observadas. Por exemplo, correlações de testes de Inglês, Francês e Português sugerem um fator básico: “domínio verbal ou em línguas”. Um segundo grupo de variáveis, representando escores em ciências exatas: Matemática e Física, por exemplo, se avaliado, corresponderia a outro fator. É esse tipo de estrutura que a Análise Fatorial procura determinar.

2.2. Modelo Fatorial Convencional

A Análise Fatorial é, em essência, um método para explicação da variabilidade de dados através de um modelo ajustado. A matriz de covariância ou a de correlação de um número razoável de variáveis $\mathbf{Y} = (Y_1, \dots, Y_p)$ é o objeto da análise. Por hipótese, o interrelacionamento entre as variáveis pode ser explicado por um modelo de regressão linear múltiplo, com os Y 's como variáveis dependentes. A característica que diferencia

a Análise Fatorial do modelo de regressão é que nessa os preditores, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$, não são observados, mas sim, inferidos a partir dos dados. No modelo convencional de Análise Fatorial, assume-se um vetor m -dimensional de variáveis latentes, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$, na população de indivíduos. As observações de uma amostra aleatória de N indivíduos não consistem, entretanto, de valores de $\boldsymbol{\theta}$, mas de valores de p variáveis observadas $\mathbf{Y} = (Y_1, \dots, Y_p)$, onde $p > m$. É assumido, também, que \mathbf{Y} depende estocasticamente de $\boldsymbol{\theta}$ através do seguinte sistema de equações lineares:

$$\begin{aligned} Y_1 &= \lambda_{11}\theta_1 + \lambda_{12}\theta_2 + \dots + \lambda_{1m}\theta_m + e_1 \\ Y_2 &= \lambda_{21}\theta_1 + \lambda_{22}\theta_2 + \dots + \lambda_{2m}\theta_m + e_2 \\ &\quad \vdots \\ Y_p &= \lambda_{p1}\theta_1 + \lambda_{p2}\theta_2 + \dots + \lambda_{pm}\theta_m + e_p \end{aligned}$$

ou, em forma matricial,

$$\mathbf{Y} = \Lambda\boldsymbol{\theta} + \mathbf{e}. \quad (2.1)$$

O coeficiente λ_{ik} , $i = 1, 2, \dots, p$ e $k = 1, 2, \dots, m$, é definido como a carga da i -ésima variável no k -ésimo fator, e a matriz Λ é denominada matriz de cargas fatoriais. $\boldsymbol{\theta}$ é o vetor de fatores comuns, ou, simplesmente, vetor de fatores e \mathbf{e} é o vetor de fatores específicos ou resíduos. Os resíduos são supostos independentes entre si, o que equivale a assumir que dado m fatores comuns, as variáveis observadas são independentes uma das outras. Os fatores, geralmente, são escalonados de modo a tornar suas variâncias iguais a 1. Por conveniência, e sem perda de generalidade, assume-se que $\boldsymbol{\theta}$ e \mathbf{Y} têm média $\mathbf{0}$. Assim, pela Equação 2.1 temos que a matriz de covariância de \mathbf{Y} é dada por

$$\Sigma = \Lambda\Theta\Lambda' + \Psi, \quad (2.2)$$

onde Θ denota a matriz de covariância de $\boldsymbol{\theta}$, e Ψ representa a matriz de covariância do resíduo \mathbf{e} .

Incorporando ao modelo a suposição de que $\Theta = \mathbf{I}_m$, matriz identidade de ordem m , e, também, a de que $\Lambda'\Psi^{-1}\Lambda$ é uma matriz diagonal, reúnem-se as condições necessárias para torná-lo identificável, ou seja, elimina-se o problema de indeterminação do modelo. A indeterminação está relacionada à produção da mesma matriz Σ através de diferentes matrizes Λ e Θ , combinadas segundo a Equação 2.2. Para exemplificar, considere \mathbf{A} uma matriz quadrada de posto completo m com colunas normalizadas, $\Lambda^* = \Lambda\mathbf{A}$ e $\Theta^* = \mathbf{A}^{-1}\Theta\mathbf{A}'^{-1}$. Note que, utilizando a Equação 2.2, se $\Sigma = \Lambda\Theta\Lambda' + \Psi$, então, por simples substituição $\Sigma = \Lambda^*\Theta^*\Lambda'^* + \Psi$, o que caracteriza a indeterminação do modelo. Várias escolhas de \mathbf{A} podem produzir cargas fatoriais que são fáceis de serem explorados visualmente ou interpretados, sem afetar o ajuste do modelo. O processo de obtenção dos valores de Λ^* e Θ^* é chamado rotação de fatores (Harman, 1976).

Ainda com base nas suposições, as variâncias das variáveis observadas são dadas por

$$\sigma_i^2 = \sum_{k=1}^m \lambda_{ik}^2 + \psi_i^2.$$

O primeiro termo, $\sum_{k=1}^m \lambda_{ik}^2$, é conhecido como *comunalidade*, e representa a parte da variância da variável Y_i explicada pelos fatores comuns. O segundo termo, ψ_i^2 , é a variância do i -ésimo resíduo, e_i , e é denominado variância específica ou *unicidade*. Para facilitar a transição para o caso de variáveis dicotomizadas, serão introduzidas algumas suposições sobre as distribuições de \mathbf{Y} , $\boldsymbol{\theta}$ e \mathbf{e} e, também, algumas considerações quando do procedimento de estimação.

Suponha que os resíduos na Equação 2.1 seguem uma distribuição normal multivariada: $\mathbf{e} \sim N_p(\mathbf{0}, \Psi)$. A distribuição de \mathbf{Y} condicional a $\boldsymbol{\theta}$ será dada por

$$(\mathbf{Y} \mid \boldsymbol{\theta}, \Lambda, \Psi) \sim N_p(\Lambda\boldsymbol{\theta}, \Psi). \quad (2.3)$$

Assumindo que $\boldsymbol{\theta} \sim N_m(\mathbf{0}, \mathbf{I}_m)$, determina-se a distribuição marginal de \mathbf{Y} , ou a distribuição de \mathbf{Y} a partir de uma amostra aleatória de indivíduos, pela integração da Equação 2.3 sobre todo o domínio de $\boldsymbol{\theta}$. Ou seja,

$$f(\mathbf{y} \mid \Lambda, \Psi) = \int f(\mathbf{y} \mid \boldsymbol{\theta}, \Lambda, \Psi) f(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Como ambas as densidades sob a integral são normais, a integração pode ser desenvolvida explicitamente. Encontra-se que

$$\mathbf{Y} \sim N_p(\mathbf{0}, \Sigma), \quad (2.4)$$

onde,

$$\Sigma = \Lambda\Lambda' + \Psi.$$

2.3. Estimação dos Parâmetros

A estimação dos parâmetros dos itens é feita através do método de máxima verossimilhança. A função de verossimilhança para as respostas de N indivíduos de amostra aleatória, sob a expressão (2.4), é dada por

$$L(\mathbf{y}_1, \dots, \mathbf{y}_N \mid \Lambda, \Psi) = \prod_{j=1}^N \frac{\exp[(-\mathbf{y}'_j \Sigma^{-1} \mathbf{y}_j)/2]}{(2\pi)^{p/2} |\Sigma|^{1/2}}. \quad (2.5)$$

Para determinar os valores dos parâmetros que satisfazem as equações de verossimilhança, maximiza-se o logaritmo da Equação 2.5 com relação às matrizes de parâmetros Λ e Ψ , diferenciando-o com relação a cada parâmetro e igualando o resultado a zero. Com as suposições introduzidas na seção anterior, garante-se que as estimativas dos parâmetros serão únicas.

A maximização da Equação 2.5 é equivalente à minimização de

$$F = -\frac{N}{2} [tr(\Sigma^{-1} \mathbf{S}) - \log |\Sigma^{-1} \mathbf{S}|],$$

onde \mathbf{S} é a matriz de correlação observada dos Y 's e F é uma função proporcional à log-verossimilhança da Equação 2.5. Note que o produtório envolvendo todos os N indivíduos presente na Equação 2.5 dá lugar a uma outra equação que envolve apenas uma função do vetor de respostas, em termos da matriz de covariância observada. Em outras palavras,

estimativas eficientes de Λ e Ψ podem ser obtidas usando somente os $p(p+1)/2$ elementos da matriz \mathbf{S} e sem perda de informação no que se refere às respostas dos N indivíduos, independentemente de quão grande seja N comparado a $p(p+1)/2$.

Mislevy (1986) cita outros dois métodos para estimação de Λ e Ψ : o método dos mínimos quadrados não ponderados (MQNP) e o método dos mínimos quadrados generalizados (MQG). O primeiro minimiza $F = (S^o - \hat{\Sigma}^o)'(S^o - \hat{\Sigma}^o)$ e o segundo, $F = (S^o - \hat{\Sigma}^o)'W^{-1}(S^o - \hat{\Sigma}^o)$, onde o operador ‘ o ’ indica que uma matriz $M = (\mathbf{m}_1, \dots, \mathbf{m}_n)$ é reescrita como um vetor de colunas da forma $(\mathbf{m}'_1, \dots, \mathbf{m}'_n)'$ e $W = \Sigma \otimes \Sigma$, com \otimes indicando o produto direto de matrizes.

Todos os três métodos citados nesta seção fornecem estimadores consistentes para Λ e Ψ sob as suposições apresentadas inicialmente. Os métodos de máxima verossimilhança e mínimos quadrados generalizados exigem que a matriz \mathbf{S} seja positiva definida e fornecem testes qui-quadrado para ajuste do modelo, permitindo assim avaliar o número de fatores.

2.4. Modelo Fatorial para Dados Dicotômicos ou Dicotomizados

Nesta seção será apresentada a extensão do modelo fatorial tradicional para dados dicotômicos ou dicotomizados. A primeira idéia foi estabelecer uma matriz que substituísse a matriz de covariância para variáveis contínuas e operar de acordo com o que foi apresentado na seção anterior. As soluções iniciais foram dadas através do uso das matrizes de coeficientes *phi* e de correlações tetracóricas. Um outro método, denominado Análise Fatorial de Informação Plena, foi introduzido a partir do artigo de Bock & Aitkin (1981) e será abordado na parte final da seção.

2.4.1. O Coeficiente de Correlação Phi

Não há impedimento para o cálculo da correlação produto momento de Pearson ($\rho(X, Y)$) entre variáveis dicotômicas - “coeficiente *phi*”, como ela é chamada neste caso especial.

$$\rho(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Dessa forma, seria natural aplicar o método da seção anterior para ajustar um modelo fatorial a uma matriz de correlação assim obtida. Entretanto, vários pesquisadores têm demonstrado os problemas inerentes a essa prática.

O principal problema é que os valores do coeficiente *phi* dependem não só do grau de relacionamento entre as variáveis, mas, também, da média individual de cada variável.

As Tabelas 2.1, 2.2 e 2.3 ilustram os resultados obtidos numa simulação com diferentes tamanhos de amostras para o coeficiente *phi*. Na Tabela 2.1, 8 itens foram submetidos a 5000 respondentes. Cada item é uma variável que assume os valores 1 e 0: 1 para respostas corretas e 0 para respostas incorretas. O primeiro item foi respondido incorretamente por apenas um dos indivíduos; o segundo item foi respondido incorretamente por dois indivíduos: o que errou o primeiro item e mais um. Note que somente um indivíduo respondeu diferentemente os dois itens. O resultado para o coeficiente de correlação *phi*

foi 0.707. O terceiro item foi respondido incorretamente pelos dois indivíduos que erraram o item anterior e mais um; o quarto item, pelos três indivíduos anteriores e mais um. Para os itens 1 e 4, apenas 3 dos 5000 indivíduos apresentam respostas diferentes para estes dois itens, isso fez com que a correlação ϕ decrescesse para 0,50. Concluindo a interpretação dos itens, o quinto foi respondido incorretamente pelos 4 indivíduos anteriores anteriores e mais 496, o sexto item, pelos 500 indivíduos anteriores e mais 500, o sétimo item, pelos 1000 indivíduos anteriores e mais 500 e o oitavo item, pelos 1500 indivíduos anteriores e mais 500.

Tabela 2.1 - Correlação ϕ para 8 itens submetidos a 5000 respondentes.

Itens	1	2	3	4	5	6	7	8
2	0,707							
3	0,577	0,816						
4	0,500	0,707	0,866					
5	0,042	0,060	0,074	0,085				
6	0,028	0,040	0,049	0,057	0,667			
7	0,022	0,031	0,037	0,043	0,509	0,763		
8	0,017	0,024	0,030	0,035	0,408	0,612	0,802	1,000

Nas Tabelas 2.2 e 2.3, a mesma análise é feita para outros 8 itens submetidos a 500 e 50 indivíduos, respectivamente. Os itens 1, 2, 3 e 4 apresentam a mesma seqüência de respostas incorretas verificadas na Tabela 2.1, e os itens 5, 6, 7 e 8 apresentam a mesma seqüência e proporção de respostas incorretas. Nota-se facilmente a dependência indesejável do coeficiente de correlação ϕ da média das variáveis envolvidas.

Tabela 2.2 - Correlação ϕ para 8 itens submetidos a 500 respondentes.

Itens	1	2	3	4	5	6	7	8
2	0,706							
3	0,576	0,816						
4	0,498	0,706	0,865					
5	0,134	0,190	0,233	0,269				
6	0,090	0,127	0,155	0,180	0,667			
7	0,068	0,097	0,119	0,137	0,509	0,764		
8	0,055	0,078	0,095	0,110	0,408	0,612	0,802	1,000

Tabela 2.3 - Correlação ϕ para 8 itens submetidos a 50 respondentes.

Itens	1	2	3	4	5	6	7	8
2	0,700							
3	0,565	0,808						
4	0,484	0,692	0,857					
5	0,429	0,612	0,758	0,885				
6	0,286	0,408	0,505	0,590	0,667			
7	0,218	0,312	0,386	0,450	0,509	0,764		
8	0,175	0,250	0,309	0,361	0,408	0,612	0,802	1,000

Quando variáveis binárias são produzidas pela dicotomização de variáveis contínuas, a escolha do ponto de corte afeta os valores do coeficiente ϕ . A análise fatorial feita a partir

da matriz dos coeficientes ϕ de variáveis binárias e produzida pela mesma estrutura de correlação, mas dicotomizadas em diferentes pontos, pode equivaler-se a modelos fatoriais com diferentes estruturas e possibilitar diferentes números de fatores.

Outro problema é que o valor de uma variável dicotômica é limitado, implicando que sua regressão com qualquer variável latente contínua com intervalo infinito não pode ser linear (McDonald & Ahlwat, 1974). Se a correlação for aplicada diretamente em variáveis dicotômicas, o modelo de análise fatorial linear dado pela Equação 2.1 fica mal especificado.

Uma alternativa ao uso da matriz de coeficientes ϕ é apresentada em seguida.

2.4.2. A Correlação Tetracórica

Diferentemente do cálculo das correlações ϕ , onde não houve nenhuma suposição sobre as variáveis observadas envolvidas, consideram-se, agora, as variáveis contínuas Y_i e Y_j que têm distribuição conjunta normal bivariada e que as variáveis observadas, X_i e X_j , que assumem os valores 0 ou 1, são resultantes dessas variáveis através de um processo de dicotomização. Nesse processo, é assumido um ponto de corte para cada uma das variáveis contínuas, denotados por γ_i e γ_j . Os valores das variáveis Y_i e Y_j não são observados diretamente, mas, somente, se assumem valores maiores ou menores do que seus respectivos pontos de corte:

$$\begin{aligned} Y_i &\geq \gamma_i \quad \text{ou} \quad Y_i < \gamma_i \\ Y_j &\geq \gamma_j \quad \text{ou} \quad Y_j < \gamma_j. \end{aligned}$$

Os valores das variáveis observadas X_i e X_j são supostos resultantes das variáveis contínuas da seguinte forma:

$$X_i = \begin{cases} 1, & \text{se } Y_i \geq \gamma_i \\ 0, & \text{se } Y_i < \gamma_i. \end{cases}$$

A variável X_j recebe tratamento análogo.

A idéia agora é inferir a correlação produto momento entre Y_i e Y_j , $\rho(Y_i, Y_j)$, a partir da observação de X_i e X_j . Como exemplo, poder-se-ia considerar a variável Y denotando a habilidade exigida por um item e a variável X , a respectiva resposta ao item: 1 para resposta correta e 0 para resposta incorreta.

Seja π_{ij} a proporção de indivíduos na população com $Y_i \geq \gamma_i$ e $Y_j \geq \gamma_j$. A proporção de indivíduos com $Y_i \geq \gamma_i$, independentemente de Y_j , e a proporção de indivíduos com $Y_j \geq \gamma_j$, independentemente de Y_i , são dadas por π_i e π_j , respectivamente.

Com as suposições sobre a distribuição de (Y_i, Y_j) , temos

$$\pi_{ij} \equiv \int_{\gamma_i}^{\infty} \int_{\gamma_j}^{\infty} f(y_i, y_j, \rho) dy_i dy_j \quad (2.6)$$

$$\pi_i \equiv \int_{\gamma_i}^{\infty} \int_{-\infty}^{\infty} f(y_i, y_j, \rho) dy_i dy_j \quad (2.7)$$

$$\pi_j \equiv \int_{-\infty}^{\infty} \int_{\gamma_j}^{\infty} f(y_i, y_j, \rho) dy_i dy_j, \quad (2.8)$$

onde $f(y_i, y_j, \rho)$ representa a função densidade da distribuição normal padrão bivariada, com ρ representando a correlação entre as variáveis Y_i e Y_j :

$$f(y_i, y_j, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{y_i^2 + y_j^2 - 2\rho y_i y_j}{2(1-\rho^2)}\right].$$

Considerando $\pi_{i\bar{j}}$ a proporção de indivíduos que responderam corretamente o item i , mas responderam incorretamente o item j , e definindo $\pi_{\bar{i}j}$ e $\pi_{\bar{i}\bar{j}}$ analogamente, todas as proporções de respondentes numa tabela de contingência 2×2 estariam determinadas.

Quando γ_i , γ_j e ρ são conhecidos, π_{ij} , π_i e π_j são determinados de forma única a partir das Equações 2.6, 2.7 e 2.8. Entretanto, o caso realista é ter conhecido π_{ij} , π_i e π_j , e daí determinar γ_i , γ_j e ρ , também de forma única. Quando ρ é calculado a partir das proporções da tabela de contingência 2×2 , satisfazendo as expressões (2.6), (2.7) e (2.8), ele é chamado coeficiente de correlação tetracórico, que será denotado por ρ_{ij} (Pearson, 1900). Se a suposição de que (Y_i, Y_j) tem distribuição normal bivariada é verdadeira, então $\rho_{ij} = \rho(Y_i, Y_j)$.

Quando se observam proporções amostrais, ρ_{ij} é estimado por r_{ij} , o coeficiente de correlação tetracório amostral. Para o cálculo de ρ_{ij} ou r_{ij} não há fórmulas fechadas; aproximações computacionais eficientes são dadas por Divgi (1979). Calculando ρ_{ij} para todos os pares de variáveis, no nosso caso, itens, forma-se a matriz de correlações tetracóricas.

A matriz de correlações tetracóricas amostral, S^* , é uma estimativa da matriz de correlação das variáveis Y 's, e assim, procedimentos padrões de análise fatorial podem ser empregados para estimar Λ e Ψ .

Da mesma forma do modelo convencional, será assumida a existência de m variáveis latentes θ . Para $p > m$ variáveis observadas (por exemplo, p itens de um teste), assume-se, também, que a estrutura linear para as p variáveis Y_i , $i = 1, 2, \dots, p$, é dada por

$$Y_i = \lambda_{i1}\theta_1 + \lambda_{i2}\theta_2 + \dots + \lambda_{im}\theta_m + e_i, \quad (2.9)$$

onde e_i são os resíduos. O contraste com a análise fatorial de variáveis observadas está no fato de que os Y 's não são observados diretamente. Observa-se, sim, um vetor de variáveis dicotômicas $\mathbf{X} = (X_1, \dots, X_p)$ com valores determinados de acordo com processo de dicotomização apresentado no início dessa seção.

O passo seguinte é a fatoração de S^* , ou seja, a estimação de Λ e Ψ , e para isso será suposto que os resíduos e_i são distribuídos segundo uma distribuição $N(0, \psi_i^2)$ e são independentes entre si e dos fatores θ . Com as suposições e o processo de dicotomização definido, a probabilidade condicional de um indivíduo j responder corretamente a um item i será dada por

$$\begin{aligned} P(X_{ij} = 1 \mid \boldsymbol{\theta}_j) &= \frac{1}{\sqrt{2\pi}\psi_i} \int_{\gamma_i}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{t - \sum_{k=1}^m \lambda_{ik}\theta_{kj}}{\psi_i}\right)^2\right] dt \\ &= F\left(\frac{\gamma_i - \sum_{k=1}^m \lambda_{ik}\theta_{kj}}{\psi_i}\right) \\ &= F_i(\boldsymbol{\theta}_j) \end{aligned} \quad (2.10)$$

Suponha, ainda, que $\boldsymbol{\theta}$ distribui-se segundo uma $N_m(\mathbf{0}, \Theta)$ na população de interesse. Como no modelo convencional, a distribuição marginal de \mathbf{Y} é $N_p(\mathbf{0}, \Sigma)$, onde, novamente,

$\Sigma = \Lambda\Theta\Lambda' + \Psi$. O fato de que nenhum dos θ 's e Y 's são observados gera uma indeterminância de escala e orientação no modelo. A solução se dá particularizando $\Theta = \mathbf{I}_m$ e $\Sigma_{ii} = 1$ para cada i . Isso implica que

$$\Sigma = \Lambda\Lambda' + \Psi$$

e

$$\psi_i^2 = 1 - \sum_{k=1}^m \lambda_{ik}^2,$$

ou, em notação matricial,

$$\Psi = \mathbf{I} - \text{diag}(\Lambda\Lambda').$$

Sendo $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})$ o vetor de respostas do j -ésimo indivíduo, a função de verossimilhança marginal dos dados para uma amostra aleatória de tamanho N será dada por

$$\begin{aligned} L[(\mathbf{x}_1, \dots, \mathbf{x}_N) | \Lambda, \gamma] &= \prod_{j=1}^N \int f(\mathbf{x}_j | \boldsymbol{\theta}, \Lambda, \gamma) f(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \prod_{j=1}^N \int \prod_i F_i(\boldsymbol{\theta})^{x_{ij}} [1 - F_i(\boldsymbol{\theta})]^{1-x_{ij}} f(\boldsymbol{\theta}) d\boldsymbol{\theta}, \end{aligned} \quad (2.11)$$

onde $f(\boldsymbol{\theta})$ é a função densidade normal padrão multivariada de $\boldsymbol{\theta}$. A Equação 2.11 também pode ser escrita como um produto de todos os s distintos padrões de resposta \mathbf{x}_q , observados com frequência r_q , $q = 1, 2, \dots, s$. A expressão resultante é

$$L = \prod_{q=1}^s \left\{ \int \prod_i F_i(\boldsymbol{\theta})^{x_{qi}} [1 - F_i(\boldsymbol{\theta})]^{1-x_{qi}} f(\boldsymbol{\theta}) d\boldsymbol{\theta} \right\}^{r_q},$$

onde $s \leq \min(N, 2^p)$. O trabalho com padrões de resposta pode ser útil quando o número de observações (indivíduos) é extremamente superior ao número de variáveis (itens). Esse procedimento é empregado em programas computacionais.

A Equação 2.11 é equivalente à Equação 2.5 e, a princípio, o procedimento de estimação para Λ e Ψ pode ser o mesmo daquela seção, ou seja, o método de máxima verossimilhança.

No desenvolvimento dessa análise um ponto merece atenção. Não se pode garantir que S^* seja uma matriz positiva definida. A ausência desta condição impossibilita a aplicação do método de máxima verossimilhança. Nesse caso, a estimação de Λ é feita através do método dos mínimos quadrados não-ponderados, minimizando a quantidade

$$\sum_i \sum_{i < j} (S_{ij}^* - \Sigma_{ij})^2.$$

As vantagens da análise fatorial através da matriz de correlações tetracóricas são: (a) a sua superioridade (melhores estimativas para as correlações) sobre a análise fatorial através do coeficiente *phi* e (b) a sua economia relativa. A economia é no sentido de menos processamento computacional quando comparado a métodos específicos para dados categorizados.

As desvantagens dessa solução podem ser classificadas em três categorias. A primeira começa na tentativa de calcular S^* , onde valores extremos são insatisfatoriamente determinados. A segunda categoria resulta do uso dos mínimos quadrados não-ponderados com a matriz de dados com dois níveis de estimação de erros - nem os erros-padrão nem o ajuste do teste são facilmente avaliados. A terceira categoria resulta do fato de que, diferentemente do caso de variáveis observadas com distribuição normal, a sumarização das variáveis dicotômicas em termos da matriz de covariância não retém, conjuntamente, toda informação sobre o relacionamento dessas variáveis; somente as informações individuais (percentual de acertos) e a informações de pares de variáveis são usadas. Essas limitações da análise fatorial através da matriz tetracórica são superadas com o uso de soluções através do método de mínimos quadrados generalizados.

Considerando Acerto Casual

Na seção anterior considerou-se o caso no qual as respostas são determinadas somente através de θ , não se considerando a possibilidade de acerto casual. Para este caso, a mesma solução pode ser empregada, com uma razão pré-especificada (probabilidade de acerto casual) c_i para cada item, se a proporção observada e a proporção conjunta são apropriadamente ajustadas. Carrol (1945), entre outros citado em Mislevy (1986), propôs correções nas proporções da tabela de contigência 2×2 nas quais são baseadas as correlações tetracóricas. As correções para as proporções marginais e conjuntas são dadas abaixo.

$$\begin{aligned}\pi_i &= (\pi_i^* - c_i)/\bar{c}_i \\ \pi_{it} &= \pi_{it}^* - (c_t/\bar{c}_i)\pi_{it}^* + (c_i c_t/\bar{c}_i \bar{c}_t)\pi_{i\bar{t}}^* \\ \pi_{i\bar{t}} &= (\bar{c}_t)^{-1}\pi_{i\bar{t}}^* - (c_i/\bar{c}_i \bar{c}_t)\pi_{i\bar{t}}^* \\ \pi_{\bar{i}t} &= (\bar{c}_i)^{-1}\pi_{\bar{i}t}^* - (c_t/\bar{c}_i \bar{c}_t)\pi_{\bar{i}t}^* \\ \pi_{\bar{i}\bar{t}} &= (\bar{c}_i \bar{c}_t)^{-1}\pi_{\bar{i}\bar{t}}^*,\end{aligned}$$

onde $\bar{c}_i = 1 - c_i$, $\bar{c}_t = 1 - c_t$ e π^* são valores amostrais correspondentes aos valores populacionais π presentes no início da Subseção 2.4.2. Os ajustes acima podem produzir proporções acima de 1 ou abaixo de 0. A solução é a imposição de valores mínimos e máximos para as proporções ou para os valores de c_i , antes do início da estimação do modelo fatorial. Na Equação 2.10 o acerto casual ao item pode ser introduzido substituindo $F_i(\theta_j)$ por $F_i^*(\theta_j) = c_i + (1 - c_i)F_i(\theta_j)$, onde c_i é a probabilidade de acerto casual ao item.

Solução por Mínimos Quadrados Generalizados(MQG)

O processo de estimação dos parâmetros de Λ e Ψ através do método de MQNP minimiza a soma das diferenças dos quadrados, elemento por elemento, com cada elemento ponderado igualmente. O método dos MQG faz uso mais eficiente dos dados, no sentido de considerar tanto a magnitude da variação quanto o interrelacionamento do erro

amostral entre os elementos. Essa abordagem com as soluções de Christoffersson (1975) e Muthén (1978) será apresentada a seguir.

Na solução de Christoffersson (1975), considera-se $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_p, \pi_{11}, \pi_{21}, \dots, \pi_{it}, \dots)$, com $1 \leq t < i \leq p$, o vetor de valores esperados de π_i e π_{it} , modelados como funções de Λ , $\boldsymbol{\gamma}$ e $\boldsymbol{\pi}^*$, o vetor correspondente de valores observados. Quando o modelo está adequado, a quantidade $\mathbf{e} = \boldsymbol{\pi}^* - \boldsymbol{\pi}$ tem distribuição normal multivariada para grandes amostras com média $\mathbf{0}$ e matriz de covariância Σ_l . Christoffersson fornece uma expressão para um estimador consistente de Σ_l , S_l , e implementa uma solução através do método de mínimos quadrados generalizados para os parâmetros do modelo fatorial. Estimativas consistentes para os parâmetros são dadas pela minimização da função

$$\mathbf{F} = (\boldsymbol{\pi}^* - \boldsymbol{\pi})' S_l^{-1} (\boldsymbol{\pi}^* - \boldsymbol{\pi}).$$

Na solução de Christoffersson (1975), os elementos da matriz S_l incluem termos até quarta ordem marginal, ou seja, proporções de respostas conjuntas de 4 itens. Isso significa que esse método usa mais informações que o dos mínimos quadrados não-ponderados, que utiliza informações de, no máximo, segunda ordem marginal. Essa solução fornece os erros-padrão das estimativas e, ainda, permite avaliar o ajuste do modelo através de teste estatístico. O interesse maior desse teste está na comparação de modelos com diferentes números de fatores, indicando qual deles se ajusta melhor aos dados. Por fim, o modelo exige solução por métodos iterativos, aplicando métodos numéricos às Equações 2.6, 2.7 e 2.8 em cada ciclo.

A solução por MQG de Muthén (1978) minimiza $F = (\mathbf{s} - \boldsymbol{\xi})' S_\delta^{-1} (\mathbf{s} - \boldsymbol{\xi})$, onde $\boldsymbol{\xi}' = (\xi_1, \xi_2)$, com $\xi_1 = \boldsymbol{\gamma}$ e $\boldsymbol{\xi}'_2 = (\sigma_{12}, \dots, \sigma_{it}, \dots)$, e \mathbf{s} é a estimativa amostral dessas quantidades, isto é, os parâmetros “limite” e as correlações tetracóricas amostrais. S_δ é um estimador consistente da matriz de covariância de $\boldsymbol{\delta} = \boldsymbol{\xi} - \mathbf{s}$, dado por uma aproximação da expressão de S_l na solução de Christoffersson. Esse procedimento também utiliza informação de até quarta ordem marginal, fornece estimadores consistente, erros-padrão e teste de ajuste de modelo.

Essas soluções exigem a construção e a imersão da matriz tetracórica assintótica, o que, dependendo do número de itens, pode elevar consideravelmente o processamento computacional.

2.4.3. Casos Heywood

É possível que na construção das matrizes de correlação uma ou mais variâncias específicas (unicidades) sejam iguais a zero (Heywood, 1931). Unicidade igual a zero correspondem a variáveis observadas completamente inseridas no espaço dos fatores, ou modeladas perfeitamente pelas variáveis latentes, sem erros de medida. No caso de unicidades não-positivas, Jöreskog & Sörbom (1980) advertem para um mal ajuste do modelo, que pode ocorrer quando um ou mais fatores são pobremente identificados pelo conjunto de variáveis observadas, ou por flutuações desfavoráveis na amostra quando esta é pequena. Os chamados casos *Heywoods* também ocorrem quando a frequência observada numa tabela de contigência de um par de itens é zero, e como consequência, o valor absoluto da correlação torna-se um.

Na análise fatorial de informação plena, apresentada na seção seguinte, os casos *Heywood* caracterizam-se quando um ou mais parâmetros a aumentam continuamente com o aumento do número de ciclos do algoritmo EM. Dependendo do caso, sugere-se a exclusão dos itens envolvidos ou a aplicação de procedimentos bayesianos.

2.4.4. Análise Fatorial de Informação Plena

Nas seções anteriores consideramos as estimações dos parâmetros através dos métodos de MQNP e MQG no modelo fatorial para variáveis dicotômicas. Essas soluções são denominadas de soluções de “informações limitadas”, no sentido de que somente são utilizadas informações de marginais de ordens inferiores da tabela de contigência que sumaria as respostas dos indivíduos e, portanto, toda informação avaliada para a estimação. Nesta seção será apresentada a análise fatorial de informação plena introduzida por Bock and Aitkin (1981). O método não requer o cálculo de coeficientes de correlação inter-itens e supera vários dos problemas presentes na análise fatorial da matriz tetracórica.

Será assumido novamente o modelo fatorial para variáveis dicotômicas apresentado na Subseção 2.4.2, através da expressão (2.9). As suposições são as seguintes:

- i. os resíduos e_i seguem uma distribuição normal de média 0 e variância σ_i^2 ;
- ii. os resíduos são independentes entre si e dos θ 's;
- iii. $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_m)$, vetor dos fatores, segue uma distribuição normal multivariada com vetor de média $\mathbf{0}$ e matriz de covariâncias \mathbf{I}_m ;
- iv. $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$, vetor de variáveis pseudo-observadas, seguem uma distribuição normal multivariada com vetor de médias $\mathbf{0}$ e matriz de covariância Σ , onde $\Sigma = \Lambda\Lambda' + \Psi$.

Com o intuito de aproximar a teoria da resposta ao item e a análise fatorial, o leitor pode interpretar a variável Y como habilidade geral do indivíduo e $\boldsymbol{\theta}$ como o vetor de habilidades específicas (veja modelo na Seção 2.4.2). Contudo, neste capítulo, $\boldsymbol{\theta}$ será referido simplesmente como vetor de fatores.

A probabilidade de um indivíduo j responder corretamente ao item i , condicionado ao vetor de fatores $\boldsymbol{\theta}_j$, será dada por

$$\begin{aligned} \Phi_i(\boldsymbol{\theta}_j) &= P(X_{ij} = 1 \mid \boldsymbol{\theta}_j) = \frac{1}{\sqrt{2\pi\sigma_i}} \int_{-\infty}^{\gamma_i} \exp \left[-\frac{1}{2} \left(\frac{t - \sum_{k=1}^m \lambda_{ik}\theta_{kj}}{\sigma_i} \right)^2 \right] dt \\ &= F \left(\frac{\gamma_i - \sum_{k=1}^m \lambda_{ik}\theta_{kj}}{\sigma_i} \right), \end{aligned} \quad (2.12)$$

ou seja, $\Phi_i(\boldsymbol{\theta}_j)$ é a distribuição acumulada normal padrão. Assim, a probabilidade do j -ésimo indivíduo apresentar o padrão de respostas $\mathbf{x}_l = [x_{1l}, x_{2l}, \dots, x_{pl}]$, condicionado a $\boldsymbol{\theta}_j$, será dada por

$$P(\mathbf{X} = \mathbf{x}_l \mid \boldsymbol{\theta}_j) = \prod_{i=1}^p [\Phi_i(\boldsymbol{\theta}_j)]^{x_{il}} [1 - \Phi_i(\boldsymbol{\theta}_j)]^{1-x_{il}},$$

com $\Phi_i(\boldsymbol{\theta}_j) = P(X_{il} = 1 \mid \boldsymbol{\theta}_j)$, e a probabilidade marginal será dada por

$$P(\mathbf{X} = \mathbf{x}_l) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} P(\mathbf{X} = \mathbf{x}_l \mid \boldsymbol{\theta}_j) f(\boldsymbol{\theta}) d\theta_1 d\theta_2 \dots d\theta_m = \int L_l(\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

onde f é a função densidade de uma distribuição $N_m(\mathbf{0}, I_m)$. Essa integral pode ser aproximada através do uso de pontos de quadratura de Gauss-Hermite (Stroud & Sechrest, 1966). A expressão resultante é:

$$P_l = P(\mathbf{X} = \mathbf{x}_l) \approx \sum_{\kappa_m}^q \dots \sum_{\kappa_2}^q \sum_{\kappa_1}^q L_l(\mathbf{X}_{\kappa}) A(X_{\kappa_1}) A(X_{\kappa_2}) \dots A(X_{\kappa_m}),$$

onde $L_l(\mathbf{X}_{\kappa}) = P(\mathbf{X} = \mathbf{x}_l \mid \boldsymbol{\theta}_j = \mathbf{X}_{\kappa})$. Assim, a integral no espaço m -dimensional está sendo substituída pelo somatório em um “grid” de q^m pontos de quadratura $\mathbf{X}_{\kappa} = (X_{\kappa_1}, X_{\kappa_2}, \dots, X_{\kappa_m})$, $\kappa = 1, 2, \dots, q$. Como as dimensões do vetor $\boldsymbol{\theta}$ foram assumidas ortogonais, os pesos $A(X_k)$ avaliados em cada ponto são dados pelo produto dos pesos associados a cada coordenada.

Estimação

No processo de estimação da matriz Σ , a matriz das cargas fatoriais Λ é responsável pela determinação da dimensão do vetor de fatores $\boldsymbol{\theta}$. Em outras palavras, e de um modo geral, as magnitudes das cargas fatoriais indicam quantos fatores devem ser contemplados pelo modelo e quanto da variância amostral de cada variável é devida a cada fator.

Trabalhando com padrões de respostas, a probabilidade de ocorrência dos s distintos padrões de respostas na amostra de N indivíduos é dada por

$$P(\mathbf{r} \mid \Lambda, \boldsymbol{\gamma}) = \frac{N!}{r_1! r_2! \dots r_s!} P_1^{r_1} P_2^{r_2} \dots P_s^{r_s}. \quad (2.13)$$

O número de indivíduos que responderam o teste com padrão l de respostas, $l = 1, 2, \dots, s$, é denotado por r_l , e como cada um dos padrões observados “caem” em apenas uma das 2^p categorias de padrões, o vetor de freqüências $\mathbf{r} = (r_1, r_2, \dots, r_s)$, com $s \leq \min(N, 2^p)$, apresenta distribuição multinomial com parâmetros N e P_l . Note que, na Equação 2.13 está sendo utilizado conjuntamente a informação de todos os itens para cada um dos indivíduos. Caracteriza-se aqui, a informação plena.

A solução de informação plena, também denominada máxima verossimilhança marginal, maximiza a Equação 2.13 em relação aos elementos de Λ e $\boldsymbol{\gamma}$, onde $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)$ é o vetor dos parâmetros de “limite”. Essa solução foi introduzida por Bock & Aitkin (1981).

Por conveniência, o argumento da função normal na Equação 2.12 é reescrito da seguinte forma

$$-(\gamma_i - \sum_{k=1}^m \lambda_{ik} \theta_{kj}) / \sigma_i = w_i + \sum_{k=1}^m a_{ik} \theta_{kj}.$$

Agora, o processo de maximização é realizado derivando a log-verossimilhança da Equação 2.13 em relação aos parâmetros w e a , igualando as equações resultantes a zero, e resolvendo em relação a esses parâmetros.

Sendo $\zeta_i = (w_1, w_2, \dots, w_p, a_{11}, a_{12}, \dots, a_{pm})$, a derivada da log-verossimilhança da Equação 2.13 em relação a ζ_i é dada por

$$\frac{d \log P(\mathbf{r} \mid \Lambda, \gamma)}{d \zeta_i} = \frac{d}{d \zeta_i} [C + \sum_{l=1}^s r_l \log P_l],$$

onde C é constante em relação a ζ_i . Bock & Aitkin (1981) apresentam o resultado do desenvolvimento da equação acima, empregando pontos de quadratura. A expressão final é

$$\sum_{\kappa_m=1}^q \cdots \sum_{\kappa_1=1}^q \frac{\bar{r}_{i\kappa} - \bar{N}_\kappa \Phi_i(\mathbf{X}_\kappa)}{\Phi_i(\mathbf{X}_\kappa)(1 - \Phi_i(\mathbf{X}_\kappa))} \frac{d \Phi_i(\mathbf{X}_\kappa)}{d \zeta_i} = 0,$$

onde

$$\bar{N}_\kappa = \sum_{l=1}^s \frac{r_l}{P_l} L_l(\mathbf{X}_\kappa) A(X_{\kappa_1}) A(X_{\kappa_2}) \cdots A(X_{\kappa_m}) = \sum_{l=1}^s r_l P(\mathbf{X}_\kappa \mid \mathbf{x}_l, \Lambda, \gamma) \quad (2.14)$$

e

$$\bar{r}_{i\kappa} = \sum_{l=1}^s x_{il} \frac{r_l}{P_l} L_l(\mathbf{X}_\kappa) A(X_{\kappa_1}) A(X_{\kappa_2}) \cdots A(X_{\kappa_m}) = \sum_{l=1}^s x_{il} r_l P(\mathbf{X}_\kappa \mid \mathbf{x}_l, \Lambda, \gamma). \quad (2.15)$$

A solução das $p + mp$ equações exige métodos iterativos, pois $\bar{r}_{i\kappa}$ e \bar{N}_κ dependem dos parâmetros a e w através de L_l .

Bock & Aitkin (1981) utilizaram uma variação do algoritmo EM (Dempster, Laird & Rubin, 1977) no processo de estimação. De modo geral, o algoritmo EM é um procedimento iterativo para a obtenção de estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade em situações onde há dados faltantes. Os dois passos do algoritmo são dados abaixo.

Passo E: Inicia-se com valores para os parâmetros a e w , calcula-se $L_l(\mathbf{X}_\kappa)$, P_l , $l=1, 2, \dots, s$, e as Equações 2.14 e 2.15, que são respectivamente, o tamanho esperado da amostra no nível κ e a frequência esperada do número de respostas corretas ao item i no nível κ .

Passo M: Com os valores de $\bar{r}_{i\kappa}$ e \bar{N}_κ resultantes do passo E obtêm-se novas estimativas para os parâmetros a e w através de um procedimento iterativo - por exemplo, Newton-Raphson. Essas estimativas são introduzidas novamente no passo E, continuando o ciclo até algum critério de convergência.

A partir das estimativas de máxima verossimilhança dos parâmetros, a_{ik} e w_i , $i = 1, 2, \dots, p$ e $k = 1, 2, \dots, m$, obtêm-se as estimativas dos γ 's e λ 's pelas relações $\hat{\gamma}_i = -\hat{w}_i / \hat{d}_i$ e $\hat{\lambda}_{ik} = \hat{a}_{ik} / \hat{d}_i$, onde $\hat{d}_i = \sqrt{1 + \sum_{k=1}^m \hat{a}_{ik}^2}$.

Teste de Ajuste do Modelo

Se o tamanho da amostra é suficientemente grande, uma estatística para o teste de adequação do modelo (teste da razão de verossimilhança), assumindo o modelo relacionado

a alternativa multinomial, é dado por

$$G^2 = -2 \sum_{l=1}^s r_l \log(NP_l/r_l).$$

G^2 tem, aproximadamente, distribuição qui-quadrado com $s - p(m + 1) + m(m - 1)/2$ graus de liberdade, supondo um modelo com m fatores. Os graus de liberdade refletem o número de padrões de respostas observados, menos o número de parâmetros estimados, mais o número de restrições impostas para efeito de identificação. Entretanto, quando o número de parâmetros é maior que o tamanho da amostra, as frequências esperadas podem ser próximas de zero, tornando o teste pouco confiável.

Quando na formação da matriz de correlação há dados faltantes (“missing data”), a estatística do qui-quadrado torna-se extremamente sensível, no sentido de superestimação. Esse fenômeno acontece, por exemplo, quando da utilização de delineamentos por Blocos Balanceados Incompletos. Outro fato é que em pesquisas da larga escala abrangendo diferentes localidades, os efeitos grupais podem aumentar os valores da diferença do qui-quadrado entre dois modelos. No manual do programa TESTFACT (Wilson, Wood & Gibbons, 1998), os autores recomendam, nos casos citados acima, dividir a estatística teste (diferença de qui-quadrados) por 2 ou 3 antes de avaliar a significância do teste. Neste trabalho, especificamente no Capítulo 7, será utilizado o número 3 como divisor e a estatística teste será denominada *diferença corrigida de qui-quadrados*.

A estatística teste, G^2 , é um índice de falta de ajuste entre o modelo e os dados. Entretanto, seu valor não deve ser interpretado diretamente, somente a diferença dos valores da estatística entre dois modelos é interpretável. De um modo geral, na comparação entre dois modelos: M_1 com m fatores e M_2 com $m + 1$ fatores, se $(G_1^2 - G_2^2)$ for positiva, o segundo modelo se adequa melhor aos dados, caso contrário, o primeiro modelo é preferível.

A dimensionalidade na Teoria da Resposta ao Item é vista como o espaço referente ao vetor θ . No primeiro capítulo, todos os modelos assumem a unidimensionalidade desse vetor, ou seja, $m = 1$. A análise fatorial de informação plena permite testar a significância estatística dos fatores adicionados sucessivamente ao modelo (teste da razão de verossimilhança), e será utilizada nesse trabalho como método para a verificação da dimensionalidade do espaço do vetor θ .

No programa TESTFACT encontram-se implementadas a análise fatorial de informação plena e a análise fatorial através da matriz de correlações tetracóricas.

Capítulo 3

Modelos Multidimensionais

3.1. Introdução

Modelos unidimensionais da Teoria da Resposta ao Item vêm sendo aplicados nos últimos anos em importantes avaliações educacionais. Esses modelos têm como suposição fundamental a unidimensionalidade do teste, ou seja, a suposição de que o teste está medindo uma única habilidade (traço latente). Neste capítulo serão introduzidos os modelos multidimensionais da TRI que contemplam os casos em que mais de uma habilidade é exigida ou necessária para que o indivíduo responda corretamente a um item.

3.2. Motivação

Os modelos unidimensionais de resposta ao item assumem que somente um traço latente - chamaremos sempre habilidade - é necessário para explicar a performance de um indivíduo em um teste, ou, em outras palavras, que o teste exige de forma preponderante apenas uma habilidade. Nessa medida, habilidade, incorpora-se a personalidade, o nível de motivação, o tipo de educação, a ansiedade, a facilidade de “trabalhar sob pressão”, os conhecimentos específicos e adjacentes relacionados ao teste e todos os outros fatores que possam vir a influenciar a performance do indivíduo no teste. Entre tantos fatores a existência de um dominante, referido como habilidade medida pelo teste, sustenta a suposição de unidimensionalidade.

Entretanto, em algumas aplicações a suposição de unidimensionalidade não é atendida, seja pela construção dos itens que compõem o teste ou pela própria finalidade da aplicação. Um exemplo seria um teste de matemática aplicado a um grupo de indivíduos, cujos itens, ou pelo menos alguns deles, exigem a interpretação de um texto antes de propriamente exigir o desenvolvimento matemático. Neste caso, poder-se-ia levantar a hipótese de se estar tratando com um teste bidimensional, no qual se exige tanto habilidade matemática quanto habilidade de interpretação ou verbal. Dessa forma, o desempenho dos indivíduos nesse teste pode ser sensível aos diferentes níveis de suas habilidades matemática e verbal. Poder-se-ia ainda, questionar a exigência de outras habilidades pelo teste, tais como visão geométrica e associação. Mas, se para níveis altos ou baixos dessas habilidades não há mudança significativa na probabilidade de resposta correta ao item, não há razão para que elas sejam consideradas.

Um ponto importante é que um teste pode ser unidimensional para uma população e não ser para outra. Supondo a aplicação de um teste, medindo as habilidades matemática e verbal, a duas populações distintas de indivíduos, se em uma delas todos os indivíduos apresentam níveis altos de habilidade verbal, somente a habilidade matemática afetará a performance no teste. Na segunda população, se a habilidade verbal ocorre de maneira

heterogênea, ambas as habilidades influenciam significativamente a performance dos indivíduos no teste, e assim, o teste seria bidimensional.

3.3. Modelos Matemáticos

A Teoria da Resposta ao Item Multidimensional (TRIM) é uma metodologia relativamente nova na modelagem do relacionamento entre as habilidades de indivíduos e a respectiva matriz de respostas a um conjunto de itens.

A TRIM caminha buscando o mesmo desenvolvimento da Teoria da Resposta ao Item unidimensional, mas muitas são as lacunas a serem preenchidas. Vários modelos foram propostos e alguns deles vêm sendo testados e aplicados.

Nesse trabalho será dada maior ênfase aos modelos compensatórios. Esses modelos são generalizações diretas dos modelos logísticos unidimensionais que, atualmente, estão entre os mais aplicados.

Os dados a serem tratados pelos modelos estarão representados por uma matriz de zeros e uns, correspondendo, respectivamente, a respostas incorretas e corretas aos itens. A matriz de dados é geralmente composta por N linhas, referentes aos indivíduos, e p colunas, referentes aos itens. Assim, a interseção de uma linha com uma coluna representa a resposta do j -ésimo indivíduo, $j = 1, 2, \dots, N$, ao i -ésimo item, $i = 1, 2, \dots, p$. As suposições sobre o mecanismo que cria essa matriz de dados são as seguintes: (1) com um acréscimo em pelo menos uma das dimensões (da habilidade) medidas, a probabilidade de obter uma resposta correta a um item é crescente. Isto é frequentemente chamado de suposição de monotonicidade; (2) a função relativa à probabilidade de resposta correta é suave, no sentido de que as derivadas da função são definidas; (3) a probabilidade da combinação de respostas pode ser determinada a partir do produto das probabilidades de resposta individual, quando as probabilidades são calculadas condicionalmente a um ponto do espaço θ . Essa é uma conseqüência da correta determinação da dimensionalidade de θ - chamada de independência condicional.

Essas suposições são consistentes para vários modelos que relacionam características de indivíduos e de itens.

3.3.1. Modelo Compensatório com Acerto Casual - MC3

A forma básica para o modelo apresentado nesta subseção é uma generalização direta do modelo logístico de 3 parâmetros (Lord, 1980), para o caso onde os indivíduos são representados por um vetor de parâmetros, ao invés de um único escalar. A forma matemática do modelo é dada por

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + \exp[-(D\mathbf{a}'_i\theta_j + d_i)]}, \quad (3.1)$$

onde

X_{ij} é uma variável dicotômica que assume os valores 1, quando o j -ésimo indivíduo responde corretamente ao item i , ou 0 quando o j -ésimo indivíduo não responde corretamente ao item i , com $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, N$;

$\boldsymbol{\theta}'_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{mj})$ representa o vetor habilidade do j -ésimo indivíduo, sendo θ_{kj} , $k = 1, 2, \dots, m$, suas habilidades específicas;

$P(X_{ij} = 1 \mid \boldsymbol{\theta}_j)$ é a probabilidade de um indivíduo j com vetor habilidade $\boldsymbol{\theta}_j$ responder corretamente o item i ;

$\mathbf{a}'_i = (a_{i1}, a_{i2}, \dots, a_{im})$ é o vetor de parâmetros relacionados à discriminação do item i ;

d_i é o parâmetro relacionado à dificuldade do item i , medido na mesma escala da habilidade;

c_i é o parâmetro que representa a probabilidade de um indivíduo com baixa habilidade responder corretamente ao item i (muitas vezes referido como a probabilidade de resposta correta dada ao acaso);

D é um fator de escala, constante e igual a 1.

A denominação compensatório se deve ao fato de que um valor baixo da habilidade em uma das dimensões, pode ser compensado por um valor alto em outra dimensão.

Interpretação do Modelo

A equação para o modelo define uma superfície que fornece a probabilidade de resposta correta para o item como função da posição dos indivíduos no espaço de habilidade especificado pelo vetor $\boldsymbol{\theta}$. Quando há somente duas dimensões, a forma da superfície de probabilidade pode ser representada graficamente. As Figuras 3.1 e 3.2 mostram a superfície de probabilidade para o mesmo item i ($a_{i1} = 0,9$; $a_{i2} = 1,5$; $d_i = -2,0$; $c_i = 0,2$) usando dois diferentes métodos de representação. A Figura 3.1 usa uma superfície tridimensional que enfatiza a monotonicidade do incremento natural da superfície. A Figura 3.2 mostra a superfície como um gráfico de linhas de contorno de igual probabilidade de resposta correta. Essa representação enfatiza que as linhas equiprováveis são linhas retas, e, além disso, são paralelas entre si. Essa característica do modelo é um resultado da forma linear do expoente do e na equação do modelo.

Os parâmetros do modelo podem ser interpretados da seguinte forma:

(1) *Parâmetros dos indivíduos.* Os parâmetros dos indivíduos, habilidades, são os elementos do vetor $\boldsymbol{\theta}_j$. O número de elementos requeridos para modelar adequadamente a matriz de dados é fornecido pela análise fatorial de informação plena introduzida no capítulo anterior. As dimensões podem não ter interpretação clara, mas são importantes para a especificação do modelo. O procedimento de rotação do espaço do vetor $\boldsymbol{\theta}$ possibilita muitas vezes uma melhoria na interpretação das dimensões.

(2) *Parâmetro de discriminação.* O parâmetro de discriminação de um item i é dado por uma função dos elementos a_{ik} , $k = 1, 2, \dots, m$ e $i = 1, 2, \dots, p$, do vetor \mathbf{a}_i . Esses elementos podem ser interpretados, muitas vezes, como k parâmetros de modelos unidimensionais (Lord, 1980). Os elementos do vetor \mathbf{a}_i estão relacionados com a inclinação da superfície de resposta do item na direção dos correspondentes eixos no espaço $\boldsymbol{\theta}$. Os elementos, portanto, indicam a sensibilidade do item para diferenciar habilidades ao longo

dos eixos de θ . No entanto, o poder de discriminação dos diferentes itens dependem da direção estabelecida no espaço de θ .

Figura 3.1 - Superfície de Resposta de um item de parâmetros $a_1 = 0,9$; $a_2 = 1,5$; $d = -2,0$ e $c = 0,2$.

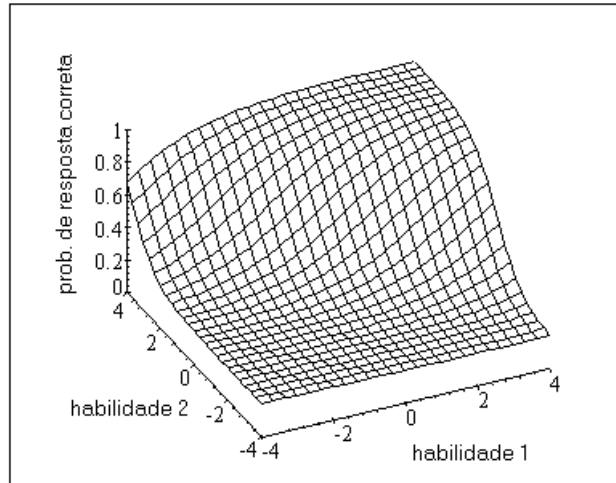
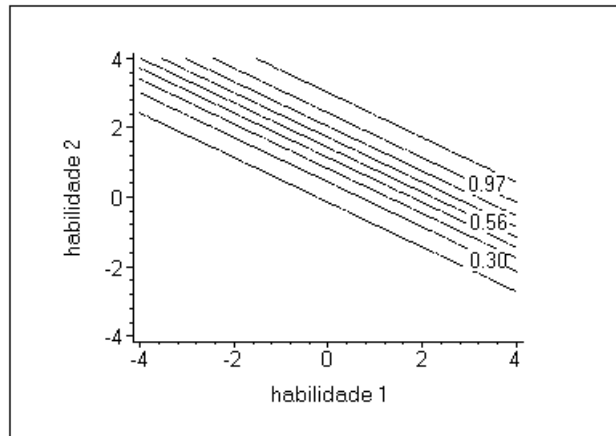


Figura 3.2 - Gráfico de Contorno da SRI dada na Figura 3.1.



O poder de discriminação de um item i multidimensional é dado por

$$DISCM_i = \sqrt{\sum_{k=1}^m a_{ik}^2},$$

onde k é o número de dimensões do vetor θ e a_{ik} é o k -ésimo elemento do vetor \mathbf{a}_i (ver Capítulo 5 para maiores detalhes).

(3) *Parâmetro de dificuldade.* O parâmetro d_i está relacionado à dificuldade do item. No entanto, o valor desse parâmetro não pode ser interpretado da mesma forma como no modelo unidimensional correspondente (modelo logístico de 3 parâmetros). O parâmetro d_i corresponde ao termo $-a_i b_i$ no modelo unidimensional (veja Equação 1.1). Um valor

que é equivalente em interpretação ao parâmetro de dificuldade unidimensional, b_i , é dado por

$$DIFICM_i = \frac{-d_i}{DISCM_i}.$$

$DIFICM_i$ indica a distância da origem do espaço θ ao ponto de maior inclinação na direção de máxima inclinação. Esse é um significado análogo ao parâmetro b_i do modelo unidimensional (ver Capítulo 4 para maiores detalhes).

(4) *Acerto Casual*. O parâmetro c_i tem o mesmo significado dado no modelo logístico unidimensional de 3 parâmetros. O valor desse parâmetro indica a probabilidade de resposta correta ao item por indivíduos com baixa habilidade em todas as dimensões.

Escala de Habilidades

De forma análoga ao caso unidimensional, faz-se necessário estabelecer uma métrica na qual serão interpretados os parâmetros do modelo. Neste trabalho, considera-se que as dimensões do vetor habilidade são ortogonais e que para cada uma delas o valor médio das respectivas habilidades é igual a 0 e o desvio-padrão é igual a 1. No caso multidimensional também não faz diferença alguma quais valores são estabelecidos para as medidas de posição e dispersão da habilidade; o importante são as relações de ordem existentes entre seus pontos. Os parâmetros multidimensionais de dificuldade ($DIFICM$) e de discriminação ($DISCM$) também estão na métrica (0;1), ou seja, possuem média 0 e desvio-padrão 1.

Os valores de $DIFICM$ e $DISCM$ são interpretados da mesma forma que os respectivos parâmetros no caso unidimensional. Dessa forma, espera-se valores entre 0 e 3,0 para $DISCM$ e valores entre $-3,0$ e $3,0$ para $DIFICM$. O parâmetro de acerto casual depende sempre do número de alternativas de resposta para o item. Por exemplo, para um item com 5 alternativas, valores plausíveis estariam no intervalo (0,1 ; 0,3), independente da escala utilizada: logística ou normal. Os modelos multidimensionais Compensatório e Ogiva Normal (Subseção 3.3.4) apresentam a mesma equivalência encontrada nos respectivos modelos unidimensionais, ou seja, para $D = 1,702$ e o mesmo vetor de parâmetros, os dois modelos fornecem resultados bem próximos.

Suposições do Modelo

O modelo descrito acima pressupõe que a dimensionalidade do espaço θ está adequadamente especificada. Isso nos garante a independência condicional. Como introduzido no Capítulo 1, a independência condicional é fundamental no processo de estimação dos parâmetros do modelo (Capítulo 6). Outras suposições são: (a) o tempo para a resolução do teste é suficiente para que todos os itens possam ser respondidos por todos os indivíduos e (b) a ordem em que os itens são apresentados aos indivíduos não interfere no desempenho dos mesmos.

3.3.2. Modelo Compensatório sem Acerto Casual - MC2

Quando não é possível, ou coerente, considerar a possibilidade de acerto casual ao item, o parâmetro c torna-se igual a zero e tem-se o chamado Modelo Compensatório sem Acerto Casual - MC2, dado por

$$P(X_{ij} = 1 | \boldsymbol{\theta}_j) = \frac{\exp(D\mathbf{a}'_i\boldsymbol{\theta}_j + d_i)}{1 + \exp(D\mathbf{a}'_i\boldsymbol{\theta}_j + d_i)}.$$

Um exemplo seria o caso de testes com itens abertos corrigidos como certo ou errado. Os parâmetros deste modelo foram definidos na subseção anterior.

3.3.3. Modelos Não-Compensatórios

O modelo não-compensatório (com acerto casual) foi desenvolvido por Sympson (1978) e é citado em Ackerman (1996). Neste modelo há um parâmetro relacionado à discriminação e outro à dificuldade para cada uma das dimensões. Como os termos no modelo são multiplicativos, a probabilidade condicional de resposta correta é limitada pela menor das probabilidades obtidas nas m dimensões do espaço $\boldsymbol{\theta}$. O modelo é dado por

$$P(X_{ij} = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i) \prod_{k=1}^m \frac{\exp[Da_{ik}(\theta_{kj} + b_{ik})]}{1 + \exp[Da_{ik}(\theta_{kj} + b_{ik})]},$$

onde b_{ik} , $i = 1, 2, \dots, p$ e $k = 1, 2, \dots, m$, é o parâmetro relacionado à dificuldade do item i na dimensão k .

A denominação não-compensatório se deve ao fato de que um nível de habilidade alto em uma dimensão não compensa um nível baixo em outra dimensão. Ackerman (1996) propõe comparações entre este modelo e o modelo compensatório (Equação 3.1) e cita Ackerman (1989), Lim (1993) e Hsu (1995) como referências para a estimação dos parâmetros dos modelos não-compensatórios e compensatórios. Entretanto, programas computacionais ainda não estão disponíveis para este modelo, o que torna sua utilização restrita. Neste trabalho, esses modelos não serão vistos com maiores detalhes. O modelo não-compensatório sem acerto casual é obtido fazendo $c_i = 0$.

3.3.4. Modelos Ogiva-Normal Multidimensionais

O modelo ogiva-normal multidimensional com acerto casual - MN3 é dado por

$$P(X_{ij} = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i) \int_{-\infty}^{Z_i(\boldsymbol{\theta}_j)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad (3.2)$$

onde $Z_i(\boldsymbol{\theta}_j) = a_{i1}\theta_{1j} + a_{i2}\theta_{2j} + \dots + a_{im}\theta_{mj} + d_i$. As definições dos parâmetros desse modelo foram dadas na Subseção 3.3.1. A relação entre este modelo e o modelo compensatório com acerto casual é a mesma dos respectivos modelos unidimensionais, ou seja, fazendo $D = 1,702$, com o mesmo conjunto de parâmetros, os dois modelos fornecem resultados bem próximos.

Note que a Equação 3.2 possui a mesma forma da Equação 2.12. A relação entre análise fatorial e teoria da resposta ao item é tema já discutido e apresentado por alguns pesquisadores. O modelo ogiva-normal está implementado nos dois programas mais utilizados na modelagem de itens multidimensionais: NOHARM (Fraser & McDonald, 1988) e TESTFACT (Wilson, Wood & Gibbons, 1998). Estes programas são comumente empregados para estimação de parâmetros de modelos multidimensionais.

Sem a possibilidade de acerto casual ao item, faz-se o parâmetro c igual a zero e tem-se o chamado modelo ogiva-normal multidimensional sem acerto casual - MN2, dado por

$$P(X_{ij} = 1 | \boldsymbol{\theta}_j) = \int_{-\infty}^{Z_i(\boldsymbol{\theta}_j)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Nos capítulos seguintes serão apresentados de forma detalhada os parâmetros multidimensionais e a definição de informação multidimensional.

Capítulo 4

Parâmetro de Dificuldade Multidimensional - DIFICM

4.1. Introdução

Na Teoria da Resposta ao Item unidimensional o parâmetro de dificuldade b representa a habilidade necessária para uma probabilidade de acerto ao item igual a $(1 + c)/2$, no modelo de 3 parâmetros e, obviamente, $1/2$ no modelo de 2 parâmetros. O parâmetro b é indicado pelo ponto na escala θ onde a inclinação da curva característica do item é máxima. Esse ponto coincide com o ponto de inflexão da curva. Nesta seção será apresentado um parâmetro de dificuldade para itens multidimensionais. A proposta dada por Reckase (1985) é uma generalização do caso unidimensional e descreve a dificuldade multidimensional como a distância, a partir da origem, ao ponto onde a superfície de resposta ao item é mais íngreme.

4.2. Motivação

Estatísticas que descrevem características de itens são comumente empregadas no processo de construção de testes. Estas estatísticas são freqüentemente usadas para produzir formas equivalentes de testes ou para produzir testes com características específicas. De modo geral, essas estatísticas assumem que o item está medindo uma única habilidade ou dimensão. Entretanto, os itens são, geralmente, multidimensionais em algum sentido e, dependendo da intensidade das dimensões, as estatísticas unidimensionais não são apropriadas. Alguns itens medem ou exigem de forma mais dominante uma só habilidade. Para estes itens as estatísticas unidimensionais são razoáveis. Por outro lado, itens que requerem claramente mais de uma habilidade necessitam de um tratamento diferenciado, ou seja, necessitam de medidas que levem em consideração as diferentes dimensões da habilidade.

Os problemas de matemática que envolvem um texto a ser interpretado são exemplos comuns desses tipos de itens. Ambas habilidade verbal (interpretação de texto) e habilidade matemática são necessárias ou exigidas para se obter a resposta correta. Medidas unidimensionais são inapropriadas para esse tipo de item. Como exemplo, poder-se-ia considerar o interesse em ordenar um conjunto de itens desse tipo segundo o parâmetro de dificuldade unidimensional. A ordenação alternar-se-ia dependendo de quão intensas fossem as habilidades verbal ou matemática da amostra de indivíduos.

Na seção seguinte será definida a dificuldade para itens multidimensionais. Essa medida de dificuldade é baseada na generalização multidimensional dos conceitos da Teoria da Resposta ao Item.

4.3. O Parâmetro de Dificuldade Multidimensional

A definição proposta para a medida de dificuldade de itens multidimensionais (*DIFICM*) é baseada em 3 suposições. Primeiro, é assumindo que a probabilidade de resposta correta a um item aumenta monotonicamente com um aumento em pelo menos uma das dimensões medidas. Segundo, um item está locado em um único ponto do espaço multidimensional. A terceira suposição é que o ponto mais razoável para ser utilizado na definição da *DIFICM* é o ponto onde o item apresenta maior discriminação. Este é o ponto onde o item fornece maior informação sobre os indivíduos.

No procedimento para determinar o parâmetro de dificuldade será utilizado, por conveniência, o modelo compensatório sem acerto casual - MC2, que é dado por

$$P(X_{ij} = 1 \mid \boldsymbol{\theta}_j, \mathbf{a}_i, d_i) = \frac{1}{1 + \exp[-(\mathbf{a}'_i \boldsymbol{\theta}_j + d_i)]}. \quad (4.1)$$

Esse modelo pode ser reescrito como

$$P(X_{ij} = 1 \mid \boldsymbol{\theta}_j, \mathbf{a}_i, d_i) = \frac{1}{1 + \exp[-\sum_{k=1}^m a_{ik}(\theta_{kj} - b_{ik})]}, \quad (4.2)$$

onde m é o número de dimensões do modelo, a_{ik} é o k -ésimo elemento do vetor \mathbf{a}_i , θ_{kj} é o k -ésimo elemento de $\boldsymbol{\theta}_j$ e $\sum_{k=1}^m a_{ik}b_{ik} = -d_i$.

Na determinação do ponto no espaço $\boldsymbol{\theta}$ em que a Superfície de Resposta do Item (SRI) tem inclinação máxima, vale ressaltar que a inclinação em um ponto qualquer da superfície é diferente dependendo da direção na qual a inclinação é obtida. Por este motivo, para o desenvolvimento da *DIFICM*, a inclinação em um ponto no espaço $\boldsymbol{\theta}$ será sempre determinada usando a direção estabelecida pela origem do espaço multidimensional e o ponto. A direção e a distância não são únicas para um item. Elas apenas são definidas de forma única quando a origem e a unidade de medida são especificadas no espaço $\boldsymbol{\theta}$.

O procedimento usado para encontrar o ponto de maior inclinação a partir da origem envolve dois passos. Primeiro, o ponto de máxima inclinação em uma particular direção é determinado. Segundo, as inclinações em cada direção são analisadas para determinar a direção que fornece o máximo absoluto.

O modelo dado pela Equação 4.2 é reescrito usando coordenadas polares, de modo a permitir o desenvolvimento com mais uma componente no modelo: a direção tomada no espaço. Isto é feito substituindo cada θ_{kj} por $\theta_j \cos(\alpha_{kj})$, onde θ_j é a distância de $\boldsymbol{\theta}_j$ à origem e α_{kj} é o ângulo entre o k -ésimo eixo e o ponto. A Equação 4.2 reescrita é dada por

$$P(X_{ij} = 1 \mid \theta_j, \mathbf{a}_i, d_i, \boldsymbol{\alpha}_j) = \frac{1}{1 + \exp[-(\sum_{k=1}^m a_{ik}\theta_j \cos(\alpha_{kj}) + d_i)]}. \quad (4.3)$$

As derivadas primeira e segunda da Equação 4.3 em relação a θ_j serão necessárias no desenvolvimento deste capítulo. São elas, respectivamente,

$$\frac{\partial P(X_{ij} = 1 \mid \theta_j, \mathbf{a}_i, d_i, \boldsymbol{\alpha}_j)}{\partial \theta_j} = P_{ij}(1 - P_{ij}) \sum_{k=1}^m a_{ik} \cos(\alpha_{kj}) \quad (4.4)$$

e

$$\frac{\partial^2 P(X_{ij} = 1 \mid \theta_j, \mathbf{a}_i, d_i, \boldsymbol{\alpha}_j)}{\partial \theta_j^2} = \left(\sum_{k=1}^m a_{ik} \cos(\alpha_{kj}) \right)^2 P_{ij} (1 - 3P_{ij} + 2P_{ij}^2), \quad (4.5)$$

onde $P_{ij} = P(X_{ij} = 1 \mid \theta_j, \mathbf{a}_i, d_i, \boldsymbol{\alpha}_j)$.

Igualando a Equação 4.5 a zero, obtém-se a inclinação máxima em uma particular direção $\boldsymbol{\alpha}_j$. Isto é obtido para $P_{ij} = 0,5$. Assim, a inclinação na direção $\boldsymbol{\alpha}_j$ está no seu máximo quando a SRI cruzar o plano 0,5.

A inclinação da superfície de resposta ao item na direção $\boldsymbol{\alpha}_j$ é dada pela primeira derivada do modelo em relação a θ_j (Equação 4.4). Quando $P_{ij} = 0,5$, a inclinação é igual a

$$\frac{1}{4} \sum_{k=1}^m a_{ik} \cos(\alpha_{kj}). \quad (4.6)$$

Quando $\alpha_{kj} = 0$ e $\alpha_{lj} = 90$, $k \neq l$, tem-se que a inclinação da superfície para $P_{ij} = 0,5$ é determinada por uma paralela ao eixo k . A inclinação paralela ao eixo k é dada por $a_{ik}/4$. Assim a_{ik} está relacionado à inclinação da SRI no ponto de máxima inclinação.

Para determinar a direção de máxima inclinação, a expressão (4.6) é diferenciada em relação a $\cos(\alpha_{kj})$. Entretanto, antes de efetuar a diferenciação, a restrição $\sum_{k=1}^m \cos^2(\alpha_{kj}) = 1$ é imposta. Esta restrição é equivalente a assumir que os eixos de $\boldsymbol{\theta}$ são ortogonais. Um sistema com $m - 1$ equações é, então, resolvido para $\cos(\alpha_{ik})$ em termos do parâmetro a . O resultado é

$$\cos(\alpha_{ik}) = \frac{a_{ik}}{(\sum_{k=1}^m a_{ik}^2)^{1/2}}. \quad (4.7)$$

Até este ponto, o vetor $\boldsymbol{\alpha}_j$ foi considerado como um parâmetro do indivíduo, usado para converter coordenadas retangulares para polares. Quando a direção que fornece a máxima inclinação a partir do sistema de equações é determinada, $\boldsymbol{\alpha}$ muda para um parâmetro de item. Por isso, é denotado por $\boldsymbol{\alpha}_i$ nas equações seguintes.

Para determinar a distância ao ponto de maior inclinação, substituímos $\cos(\alpha_{kj})$ na Equação 4.3 pela Equação 4.7, e a equação resultante pode ser resolvida em relação a θ_j com $P_{ij} = 0,5$ - o valor da probabilidade quando a inclinação é máxima. O resultado é

$$DIFICM_i = \frac{-d_i}{(\sum_{k=1}^m a_{ik}^2)^{1/2}}, \quad (4.8)$$

onde $DIFICM$ é o parâmetro de dificuldade para itens multidimensionais. Assim, a dificuldade multidimensional é descrita pela distância da origem ao ponto de maior inclinação da superfície multidimensional e pelos ângulos ou cosenos necessários para descrever a direção correspondente a esse ponto. Esse parâmetro pode ser interpretado da mesma forma que o parâmetro b nos modelos unidimensionais da TRI.

Um resultado de extrema importância diz respeito à comparação dos diversos níveis de dificuldade dos itens. Diferentemente do caso unidimensional, aqui a comparação não é totalmente direta. No caso multidimensional, os itens devem estar medindo a mesma combinação de habilidades, isto é, eles devem estar na mesma direção. Para exemplificar, considere os dados da Tabela 4.1, referentes a um conjunto de 3 itens bidimensionais,

onde a primeira das dimensões retrata a habilidade matemática e a segunda descreve a habilidade verbal (interpretação de texto).

Tabela 4.1 - Parâmetros fictícios de 3 itens

Item	a_{i1}	a_{i2}	d_i	$DIFICM_i$
1	1,69	0,21	0,58	-0,34
2	1,30	0,14	-0,41	0,31
3	0,10	1,85	1,50	-0,81

Verificando os dados da tabela, é razoável comparar as dificuldades dos itens 1 e 2, que estão na mesma direção, ou seja, medem de forma mais enfática a primeira dimensão - a habilidade matemática. Contudo, não seria razoável comparar as dificuldades dos itens 2 e 3, porque estes itens estão realizando medições em direções totalmente diferentes: o item 2 mede predominantemente a habilidade matemática e o item 3, a habilidade verbal.

Itens que meçam melhor uma determinada direção podem ser combinados para formar testes que operam como se estivessem medindo uma única dimensão.

Capítulo 5

Parâmetro de Discriminação Multidimensional - DISCM

5.1. Introdução

As medidas clássicas para a discriminação de itens usualmente utilizadas são as correlações ponto-bisserial e bisserial. Estas medidas são usadas de um modo geral como indicadoras da qualidade do item ou para a seleção de itens na composição de um teste. Na maioria das vezes, as medidas de discriminação da TRI e os conceitos relacionados à informação do item são usados com o mesmo intuito, mas estas medidas também são usadas para especificar a precisão da medida fornecida por um item em diferentes níveis da habilidade ao longo da escala θ .

Tanto as medidas de discriminação do item da Teoria Clássica quanto a da TRI unidimensional são baseadas na suposição de que o teste está medindo um único traço latente - definido pelo escore total ou pela habilidade θ .

Nesta seção será definida a discriminação para itens multidimensionais, ou seja, para o caso em que mais de uma habilidade é requerida para determinar a resposta correta a um item.

5.2. Conceitualização

Nos modelos da TRI unidimensional é assumido que a probabilidade de responder corretamente a um item aumenta com o aumento do nível da habilidade que está sendo medida. Da mesma forma, nos modelos multidimensionais é assumido que a probabilidade de resposta correta a um item aumenta quando há um aumento em cada uma das habilidades requeridas pelo teste.

De modo geral, o poder de discriminação de um item indica quão rápida é a transição de baixa para alta probabilidade de resposta correta a um item.

Um item com alta discriminação divide claramente a região espacial em duas partes, tendo apenas uma estreita região onde as probabilidades são de magnitude intermediária.

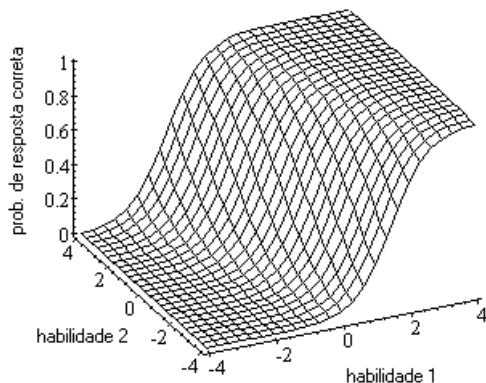
A Figura 5.1 mostra a SRI para dois itens, o primeiro com discriminação moderada e o outro com baixa discriminação. Note que os dois itens não discriminam na mesma direção no espaço θ .

A discriminação de um item está relacionada com a inclinação da superfície de resposta ao item. A inclinação difere dependendo da localização no espaço θ e da direção relativa à superfície naquela localização. Por exemplo, para o item 1 na Figura 5.1 a inclinação é quase nula no ponto $(-2, -2)$ em qualquer direção; no ponto $(2, -3)$ a inclinação é íngreme na direção aproximadamente paralela ao eixo θ_1 (habilidade 1), e é muito baixa, ou quase nula, na direção que passa através da superfície. O nível máximo de discriminação do item está na localização e na direção onde a inclinação é mais íngreme. O poder de discriminação de um item também pode ser descrito relativo a uma particular

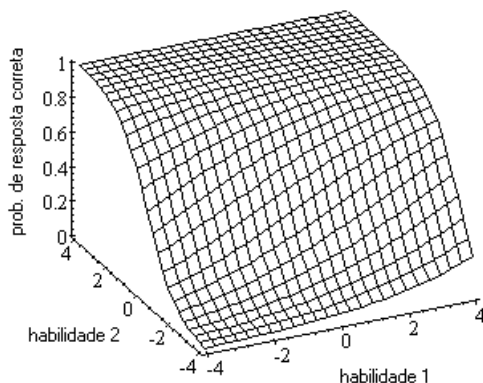
direção no espaço θ . Usando esta medida, dois itens podem ser comparados diretamente para determinar qual apresenta melhor medida para uma particular habilidade.

Figura 5.1 - Superfícies de Resposta de dois itens que variam na discriminação e na dimensão avaliada.

Item 1: $\alpha_{11}=1,9, \alpha_{12}=0,7, d_1=c_1=0$ e $DISCM_1=2,02$.



Item 2: $\alpha_{21}=0,4, \alpha_{22}=1,2, d_2=2, c_2=0$ e $DISCM_2=1,26$.



Na Figura 5.1, por exemplo, o item 2 é melhor que o item 1 para diferenciar dois indivíduos que possuem habilidades nos pontos $(\theta_{11} = 1, \theta_{21} = 0)$ e $(\theta_{12} = 1, \theta_{22} = 1)$ no espaço θ . Note que estes dois indivíduos somente diferem na habilidade 2 ($\theta_2: \theta_{21} \neq \theta_{22}$). Assim, o item 2 apresenta mais informação para medir θ_2 na região especificada no espaço θ . De fato, o item 2 apresenta uma mudança maior nas probabilidades para os dois indivíduos do que o item 1, considerando a direção que é paralela a θ_2 .

5.3. O Parâmetro de Discriminação Multidimensional - DISCM

O parâmetro de discriminação de um item no caso multidimensional fornece o mesmo tipo de informação que é fornecido pelo parâmetro unidimensional correspondente. *DISCM* permite comparar itens numa medida geral (multidimensional) de “qualidade”. Através do *DISCM* tem-se, também, um indicador da magnitude da discriminação do item em cada dimensão do espaço θ .

No caso da TRI unidimensional, o parâmetro de discriminação está relacionado à inclinação da curva característica do item no ponto onde a inclinação é máxima, o ponto de inflexão. O ponto de inflexão é também usado para definir o parâmetro de dificuldade para um item. O parâmetro de discriminação multidimensional é definido pela generalização direta desse conceito, ou seja, *DISCM* está relacionado à inclinação da superfície definida por um modelo multidimensional no ponto de máxima inclinação na direção indicada por *DIFICM* (veja capítulo anterior). A especificação da direção indicada por *DIFICM* elimina problemas de identificabilidade da definição (Reckase & McKinley, 1991). Além disso, dentre todas as direções possíveis, a especificada por *DIFICM* é a de máxima inclinação.

O parâmetro *DISCM* é uma medida de capacidade do item para distinguir entre indivíduos que estão em diferentes localizações no espaço θ . Uma estatística similar também pode ser estabelecida condicionalmente a qualquer direção particular no espaço. Um caso interessante é determinar a discriminação nas direções dos eixos coordenados. Essas estatísticas condicionais dão informações sobre quão bem o item mede uma particular dimensão.

O procedimento matemático para determinar o *DISCM* pode ser resumido em 4 passos:

- i. converte-se a expressão matemática para a SRI para coordenadas polares;
- ii. usa-se a derivada segunda dessa expressão em relação a θ_j para determinar o ponto de máxima inclinação nessa direção;
- iii. determina-se a expressão para a inclinação no ponto de máxima inclinação usando a derivada primeira (em relação a θ_j);
- iv. toma-se a derivada primeira em relação a $\cos(\alpha)$ para determinar a direção da máxima inclinação. Uma função da inclinação nessa direção é proposta como o parâmetro de discriminação.

Esse procedimento fornece um valor para *DISCM* que se relaciona com *DIFICM* da mesma maneira que o parâmetro a se relaciona com o b na TRI unidimensional.

Os passos citados acima foram introduzidos no capítulo anterior, empregando, por conveniência, o modelo compensatório sem acerto casual - MC2 dado pela Equação 4.1. Na definição de *DISCM* será empregado o mesmo modelo, também pelo mesmo motivo. Claro deve ficar que o conceito pode ser estendido a outros modelos.

No capítulo anterior, mostrou-se que a inclinação da superfície na direção α_j é dada pela primeira derivada do modelo em relação a θ_j , Equação 4.5, e que a maximização desse resultado satisfazendo a condição de que $\sum_{k=1}^m \cos^2(\alpha_{kj}) = 1$ é obtida derivando-o em relação a $\cos(\alpha_{kj})$. O resultado fornece a direção de máxima inclinação e é dado por:

$$\cos(\alpha_{ik}) = \frac{a_{ik}}{(\sum_{k=1}^m a_{ik}^2)^{1/2}}, k = 1, 2, \dots, m. \quad (5.1)$$

A inclinação da SRI no ponto de inflexão na direção α_j (veja Seção 4.3) é dado por

$$\frac{1}{4} \sum_{k=1}^m a_{ik} \cos(\alpha_{kj}). \quad (5.2)$$

Substituindo a Equação 5.1 na 5.2, temos a inclinação na direção estabelecida para *DIFICM*:

$$\frac{1}{4} \left(\sum_{k=1}^m a_{ik}^2 \right)^{1/2}.$$

Fazendo um paralelo, para o modelo logístico unidimensional de 2 parâmetros a inclinação no ponto de inflexão é igual a $a_i/4$. Assim, $(\sum_{k=1}^m a_{ik}^2)^{1/2}$ é análogo ao parâmetro a do modelo unidimensional. Portanto, o parâmetro de discriminação para itens multidimensionais pode ser definido como

$$DISCM_i = \left(\sum_{k=1}^m a_{ik}^2 \right)^{1/2}. \quad (5.3)$$

Uma propriedade importante dessa definição é que se o item mede apenas uma dimensão, isto é, quando $a_{ip} > 0$ e $a_{iq} = 0$, $p \neq q$, $DISCM = a_{ip}$. Assim, neste caso especial, $DISCM$ é igual ao parâmetro de discriminação unidimensional.

No caso multidimensional, só é possível comparar diretamente as discriminações de itens se eles estiverem sendo medidos na mesma direção. O leitor deve interpretar as direções como combinações de habilidades, ou seja, se os itens estão na mesma direção, eles estão medindo as habilidades com intensidades semelhantes. Se os itens estão em direções diferentes, primeiramente deve-se selecionar uma direção comum, em seguida recalcular a discriminação para um dos itens, ou eventualmente para os dois, e só então proceder à comparação. Neste caso é conveniente utilizar a fórmula

$$DISCM_{i_D} = \sum_{k=1}^m a_{ik} \cos(\alpha_{ik}).$$

Na comparação de itens segundo as discriminações, a ordenação pode mudar sensivelmente dependendo da direção estabelecida.

5.4. Função de Informação Multidimensional - INFM

A definição de informação multidimensional é uma generalização direta do conceito de informação da TRI unidimensional.

A informação do item multidimensional está relacionada ao parâmetro *DISCM* no sentido de que se um item tem valores altos de discriminação, ele fornece uma grande quantidade de informação. Entretanto, a informação multidimensional difere do parâmetro *DISCM* porque *INFM* está relacionada com a capacidade do item discriminar em cada ponto do espaço θ , e não só no ponto mais íngreme da SRI.

No caso unidimensional, a função de informação do item i para um determinado valor de θ é dada por

$$I_i(\theta) = \frac{[\partial P_i(\theta)/\partial \theta]^2}{P_i(\theta)[1 - P_i(\theta)]}, \quad (5.4)$$

onde $P_i(\theta) = P(X_{ij} = 1 \mid \theta) = P(X_{ij} = 1 \mid \theta, \zeta_i)$, com ζ_i significando o vetor dos parâmetros do item i .

A informação fornecida pelo teste é simplesmente a soma das informações fornecidas por cada item que compõe o mesmo, ou seja,

$$I_T(\theta) = \sum_{i=1}^p I_i(\theta), \quad (5.5)$$

onde p é o número de itens.

Para o caso multidimensional, pode-se usar as Equações 5.4 e 5.5, sendo $P_i(\theta_j)$ definido, por exemplo, pela Equação 4.1. A inclinação no numerador da Equação 5.4 pode ser determinada em várias direções. Na superfície de resposta ao item, cada ponto do espaço θ apresenta várias medidas de inclinação, e não só uma como no caso unidimensional. A inclinação varia dependendo da direção tomada.

Para determinar a inclinação numa determinada direção são utilizadas derivadas direcionais. A derivada direcional é definida como

$$\nabla_{\alpha} P(\theta) = \frac{\partial P(\theta)}{\partial \theta_1} \cos(\alpha_1) + \frac{\partial P(\theta)}{\partial \theta_2} \cos(\alpha_2) + \dots + \frac{\partial P(\theta)}{\partial \theta_m} \cos(\alpha_m), \quad (5.6)$$

onde

α é o vetor dos ângulos com os eixos coordenados no espaço θ ;

α_k , $k = 1, 2, \dots, m$, é o k -ésimo elemento do vetor α ;

θ é o vetor habilidade que define um ponto no espaço θ ;

θ_k é o k -ésimo elemento do vetor θ .

A Equação 5.6 fornece a inclinação na direção α para todo ponto θ no espaço θ . Para o cálculo de *INFM*, a derivada direcional substitui a derivada no numerador da Equação 5.4 (o índice i simplesmente especifica um item). Assim, a função de informação do item pode ser determinada em qualquer direção do espaço θ . Para descrever totalmente a estrutura de informação de um item, muitos gráficos de informação são necessários. Em princípio, a função de informação pode ser determinada por um número infinito de direções a partir da origem. Na prática, a determinação da função de informação para ângulos em intervalos de 10° entre 0° e 90° a partir do eixos é suficiente para determinar onde e para qual combinação de habilidades, o item fornece maior informação.

A função de informação multidimensional considerando o MC2 (Subseção 3.3.2) é dada por $I_i(\theta) = P_i(\theta)[1 - P_i(\theta)][\sum_{k=1}^m a_{ik} \cos(\alpha_k)]^2$. Esta equação fornece a informação no ponto indicado por θ na direção α . Para a função de informação do teste, que é dada pela soma da informação de todos os itens, a direção considerada deve ser a mesma para todos os itens.

Capítulo 6

Estimação dos Parâmetros Multidimensionais

6.1. Introdução

A estimação dos parâmetros do modelo é uma das etapas mais importantes na utilização da Teoria da Resposta ao Item. É através do modelo que se procura explicar a interação existente entre indivíduos e itens. Nos capítulos anteriores foram introduzidos alguns dos modelos multidimensionais e definidos seus parâmetros. Neste capítulo será abordado o processo de estimação dos parâmetros dos itens e das habilidades dos indivíduos (parâmetros do modelo).

O objetivo na estimação é encontrar o conjunto de valores para os parâmetros que maximize a verossimilhança das respostas aos itens. Neste ponto, vale ressaltar que, de um modo geral, apenas uma matriz $N \times p$ de respostas binárias é conhecida. A forma básica da equação de verossimilhança é dada por

$$L = \prod_{j=1}^N \prod_{i=1}^p P(X_{ij} = x_{ij} \mid \boldsymbol{\theta}_j, \boldsymbol{\alpha}_i, d_i, c_i),$$

onde x_{ij} é a resposta dicotômica (0 ou 1) do j -ésimo indivíduo ao i -ésimo item.

No processo de estimação são três as situações possíveis:

- i. as habilidades dos indivíduos são conhecidas e deseja-se apenas estimar os parâmetros dos itens;
- ii. os parâmetros dos itens são conhecidos e deseja-se estimar as habilidades dos indivíduos;
- iii. não se conhece os parâmetros dos itens e nem as habilidades dos indivíduos e deseja-se estimar ambas.

Nestas três situações a solução pode ser dada pelo Método de Máxima Verossimilhança através de algum processo iterativo, como, por exemplo, os algoritmos de Newton-Raphson (Judge et al., 1982) ou Escore de Fisher (Judge et al., 1982), ou, ainda, por procedimentos bayesianos.

Em Baker (1992) e Andrade, Tavares & Valle (2000) o leitor encontra uma excelente descrição do processo de estimação dos parâmetros do modelo para o caso unidimensional. Para o caso multidimensional o desenvolvimento é análogo, podendo seguir, de forma geral, a mesma seqüência de procedimentos do caso univariado apresentado na segunda referência citada acima. Para o desenvolvimento deste capítulo será utilizado o modelo compensatório representado na Equação 3.1.

6.2. Estimação dos Parâmetros dos Itens

Nesta seção será apresentada a estimação dos parâmetros dos itens através do Método da Máxima Verossimilhança, considerando conhecidas as habilidades dos indivíduos. Esta situação não é esperada na prática, mas seu desenvolvimento é importante para o caso mais complexo, que é a última das três situações do proceso de estimação citadas anteriormente.

Considerando a independência entre as respostas dos indivíduos e a independência condicional, a equação de verossimilhança das respostas aos itens é dada por

$$L(\zeta) = \prod_{j=1}^N P(\mathbf{X}_j = \mathbf{x}_j \mid \boldsymbol{\theta}_j, \zeta) = \prod_{j=1}^N \prod_{i=1}^p P(X_{ij} = x_{ij} \mid \boldsymbol{\theta}_j, \zeta_i), \quad (6.1)$$

onde $\zeta = (\zeta'_1, \zeta'_2, \dots, \zeta'_p)'$ é o conjunto de parâmetros de todos os itens com $\zeta'_i = (\mathbf{a}'_i, c_i, d_i)$.

Sejam

$$P_{ij} = P(X_{ij} = 1 \mid \boldsymbol{\theta}_j, \zeta_i) \quad \text{e} \quad Q_{ij} = 1 - P_{ij} = P(X_{ij} = 0 \mid \boldsymbol{\theta}_j, \zeta_i).$$

Tem-se que

$$P(X_{ij} = x_{ij} \mid \boldsymbol{\theta}_j, \zeta_i) = P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}.$$

Assim,

$$\log L(\zeta) = \sum_{j=1}^N \sum_{i=1}^p [x_{ij} \log P_{ij} + (1 - x_{ij}) \log Q_{ij}]. \quad (6.2)$$

Os estimadores de Máxima Verossimilhança para os parâmetros dos p itens são determinados quando ζ_i satisfaz a equação

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = 0, \quad i = 1, 2, \dots, p. \quad (6.3)$$

Pela condição de independência condicional a estimação é feita item a item. Desta forma, para a estimação de todos os itens o processo é simplesmente repetido p vezes. Um ponto importante é que o conhecimento das habilidades implica o conhecimento da escala (métrica) na qual elas foram geradas e assim os parâmetros dos itens estarão estimados nesta mesma métrica. No desenvolvimento da Equação 6.3 cada item possui $k + 2$ parâmetros, $k = 1, 2, \dots, m$. As expressões finais para a Equação 6.3 em relação a cada um dos $k + 2$ parâmetros são dadas por

$$\begin{aligned} \frac{\partial \log L(\zeta)}{\partial a_{ik}} &= D(1 - c_i) \sum_{j=1}^N (x_{ij} - P_{ij}) \theta_{kj} W_{ij} = 0, \\ \frac{\partial \log L(\zeta)}{\partial d_i} &= (1 - c_i) \sum_{j=1}^N (x_{ij} - P_{ij}) W_{ij} = 0, \\ \frac{\partial \log L(\zeta)}{\partial c_i} &= \sum_{j=1}^N \frac{(x_{ij} - P_{ij})}{P_{ij}^*} W_{ij} = 0, \end{aligned} \quad (6.4)$$

onde $W_{ij} = \frac{P_{ij}^* Q_{ij}^*}{P_{ij} Q_{ij}}$, com $P_{ij}^* = [1 + \exp(-D\mathbf{a}'_i \boldsymbol{\theta}_j - d_i)]^{-1}$ e $Q_{ij}^* = 1 - P_{ij}^*$.

Estas equações não possuem solução explícita e por esse motivo é necessário aplicar algum método iterativo para a obtenção das estimativas de máxima verossimilhança. Dois métodos comumente utilizados são os de Newton-Raphson e Escore de Fisher. A aplicação desses métodos exige as derivadas segundas da equação de log-verossimilhança (Equação 6.2). Tem-se então:

$$\begin{aligned}\frac{\partial^2 \log L(\zeta)}{\partial a_{ik}^2} &= D^2(1 - c_i) \sum_{j=1}^N (x_{ij} - P_{ij}) \theta_{kj}^2 W_{ij} [(1 - 2P_{ij}^*) - (x_{ij} - P_{ij}) W_{ij} (1 - c_i)], \\ \frac{\partial^2 \log L(\zeta)}{\partial d_i^2} &= (1 - c_i) \sum_{j=1}^N (x_{ij} - P_{ij}) W_{ij} [-(1 - 2P_{ij}^*) - (x_{ij} - P_{ij}) W_{ij} (1 - c_i)], \\ \frac{\partial^2 \log L(\zeta)}{\partial c_i^2} &= - \sum_{j=1}^N (x_{ij} - P_{ij})^2 W_{ij}^2 (P_{ij}^*)^{-2}, \\ \frac{\partial^2 \log L(\zeta)}{\partial a_{ik} \partial a_{it}} &= D^2(1 - c_i) \sum_{j=1}^N (x_{ij} - P_{ij}) \theta_{kj} \theta_{tj} W_{ij} [(1 - 2P_{ij}^*) - (1 - c_i)(x_{ij} - P_{ij}) W_{ij}], \\ \frac{\partial^2 \log L(\zeta)}{\partial a_{ik} \partial d_i} &= (1 - c_i) \sum_{j=1}^N (x_{ij} - P_{ij}) \theta_{kj} W_{ij} [-D(1 - 2P_{ij}^*) - (x_{ij} - P_{ij}) W_{ij} (1 - c_i)], \\ \frac{\partial^2 \log L(\zeta)}{\partial a_{ik} \partial c_i} &= \sum_{j=1}^N (x_{ij} - P_{ij}) \theta_{kj} W_{ij} [-D - (x_{ij} - P_{ij}) W_{ij} (1 - c_i) (P_{ij}^*)^{-1}], \\ \frac{\partial^2 \log L(\zeta)}{\partial d_i \partial c_i} &= \sum_{j=1}^N (x_{ij} - P_{ij}) W_{ij} [-1(x_{ij} - P_{ij}) W_{ij} (1 - c_i) (P_{ij}^*)^{-1}],\end{aligned}$$

Nas expressões acima, $t = 1, 2, \dots, m, t \neq k$.

Representando por $\hat{\zeta}_i^{(t)}$ a estimativa de ζ_i na iteração t , então, pelo algoritmo de Newton-Raphson, na iteração $t + 1$, tem-se

$$\hat{\zeta}_i^{(t+1)} = \hat{\zeta}_i^{(t)} - [\mathbf{H}(\hat{\zeta}_i^{(t)})]^{-1} \mathbf{h}(\hat{\zeta}_i^{(t)}),$$

onde $\mathbf{h}(\zeta_i) \equiv \frac{\partial \log L(\zeta)}{\partial \zeta_i}$ é a função escore e $\mathbf{H}(\zeta_i) \equiv \frac{\partial^2 \log L(\zeta)}{\partial \zeta_i \partial \zeta_i'}$ é a matriz de informação observada com $\zeta_i = (a_{i1}, a_{i2}, \dots, a_{ik}, d_i, c_i)'$.

Em Andrade, Tavares & Valle (2000), os autores definem

$$\mathbf{h}(\zeta_i) \equiv \frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^N (x_{ij} - P_{ij}) W_{ij} \mathbf{h}_{ij}$$

e

$$\mathbf{H}(\zeta_i) \equiv \frac{\partial^2 \log L(\zeta)}{\partial \zeta_i \partial \zeta_i'} = \sum_{j=1}^N (x_{ij} - P_{ij}) W_{ij} [\mathbf{H}_{ij} - (x_{ij} - P_{ij}) W_{ij} \mathbf{h}_{ij} \mathbf{h}_{ij}'],$$

com

$$\mathbf{h}_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial P_{ij}}{\partial \zeta_i} \right) \quad \text{e} \quad \mathbf{H}_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial^2 P_{ij}}{\partial \zeta_i \partial \zeta_i'} \right),$$

onde P_{ij} é dado pela Equação 3.1.

No caso multidimensional a matriz \mathbf{H}_{ij} e o vetor \mathbf{h}_{ij} são dados por

$$\mathbf{H}_{ij} = \begin{pmatrix} D^2\theta_{1j}^2\Delta & \cdot & \cdot & \cdot & \cdot & \cdot \\ D^2\theta_{1j}\theta_{2j}\Delta & D^2\theta_{2j}^2 & \cdot & \cdot & \cdot & \cdot \\ \vdots & \vdots & \ddots & \cdot & \cdot & \cdot \\ D^2\theta_{1j}\theta_{mj}\Delta & D^2\theta_{2j}\theta_{mj}\Delta & \cdots & D^2\theta_{mj}^2 & \cdot & \cdot \\ -D\theta_{ij}\Delta & D\theta_{2j}\Delta & \cdots & D\theta_{mj}\Delta & -\Delta & \cdot \\ -D\theta_{1j} & -D\theta_{1j} & \cdots & -D\theta_{mj} & -1 & 0 \end{pmatrix}$$

e

$$\mathbf{h}_{ij} = \begin{pmatrix} D(1-c_i)\theta_{1j} \\ D(1-c_i)\theta_{2j} \\ \vdots \\ D(1-c_i)\theta_{mj} \\ (1-c_i) \\ (P_{ij}^*)^{-1} \end{pmatrix},$$

respectivamente. Na matriz \mathbf{H}_{ij} , tem-se que $\Delta = (1-c_i)(1-2P_{ij}^*)$.

Para aplicar o método Escore de Fisher, substituem-se os componentes da matriz de derivadas segundas usadas no algoritmo de Newton-Raphson pelos seus valores esperados. A matriz resultante é a matriz de informação esperada, denotada por $\mathbf{H}_E(\boldsymbol{\zeta}_i)$. Como a variável X_{ij} tem distribuição Bernoulli (P_{ij}), tem-se que

$$\begin{aligned} \mathbf{H}_E(\boldsymbol{\zeta}_i) &= E(\mathbf{H}(\boldsymbol{\zeta}_i)) = \sum_{j=1}^N [E(x_{ij} - P_{ij})W_{ij}\mathbf{H}_{ij} - E(x_{ij} - P_{ij})^2W_{ij}^2\mathbf{h}_{ij}\mathbf{h}'_{ij}] \\ &= \sum_{j=1}^N [P_{ij}^*Q_{ij}^*W_{ij}\mathbf{h}_{ij}\mathbf{h}'_{ij}]. \end{aligned}$$

Assintoticamente o estimador de máxima verossimilhança $\hat{\boldsymbol{\zeta}}_i$ (sob algumas condições de regularidade) possui distribuição normal com vetor de médias $\boldsymbol{\zeta}_i$ e matriz de covariâncias dada por $[-\mathbf{H}_E(\boldsymbol{\zeta}_i)]^{-1}$.

Estimativas Iniciais

Antes do processo de estimação é necessário determinar o número de fatores do modelo. Isto é feito através da análise fatorial de informação plena introduzida no Capítulo 2. Esta análise faz uso de um processo iterativo cujas estimativas iniciais são fornecidas pelas cargas fatoriais da matriz de correlação tetracórica. Funções dessas cargas são sugeridas como estimativas iniciais para os parâmetros a 's. Para os parâmetros d 's, estimativas iniciais podem ser obtidas por funções dos pontos de corte apresentadas na Subseção 2.4.2. Para as estimativas iniciais dos parâmetros c 's podem ser utilizados os inversos do número

de alternativas para cada item. Por exemplo, se para o item i há m_i alternativas possíveis, a estimativa inicial para o parâmetro de acerto ao acaso é $c_i = 1/m_i$. A dificuldade e a discriminação multidimensionais são funções dos parâmetros a 's e d 's. Elas são obtidas, finalmente, pelas Equações 4.8 e 5.3.

6.3. Estimação das Habilidades

Na administração de testes educacionais muitas vezes o principal objetivo é obter uma medida da habilidade para os indivíduos submetidos ao teste. Para um certo indivíduo essa medida seria dada pela estimativa de máxima verossimilhança de sua habilidade (desconhecida). Para o desenvolvimento desta seção supõe-se que (a) os parâmetros dos p itens são conhecidos e (b) os indivíduos respondem ao teste independentemente, e assim, as habilidades podem ser estimadas individualmente.

A estimação da habilidade supondo os parâmetros dos itens conhecidos é um caso mais realista do que o discutido na seção anterior. Hoje, no Brasil, alguns bancos de itens estão sendo criados com o objetivo de fornecer itens já testados e calibrados (com parâmetros estimados) para avaliações educacionais do ensino fundamental e médio. Um exemplo é o Banco Nacional de Itens do Instituto Nacional de Estudos e Pesquisas Educacionais - INEP/MEC.

Novamente considerando a independência entre as respostas de diferentes indivíduos e a independência condicional, a equação de log-verossimilhança das respostas aos itens, agora como função das habilidades $\boldsymbol{\theta}$, é dada por

$$\log L(\boldsymbol{\theta}) = \sum_{j=1}^N \sum_{i=1}^p [x_{ij} \log P_{ij} + (1 - x_{ij}) \log Q_{ij}], \quad (6.5)$$

onde o vetor $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_N)'$ representa as habilidades de todos os indivíduos, sendo $\boldsymbol{\theta}'_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{mj})$, $j = 1, 2, \dots, N$, o vetor habilidade do j -ésimo indivíduo cujos componentes, θ_{kj} , $k = 1, 2, \dots, m$, representam habilidades específicas.

O estimador de máxima verossimilhança de $\boldsymbol{\theta}_j$ é o valor que maximiza a Equação 6.5. Este valor é obtido quando $\boldsymbol{\theta}_j$ satisfaz a equação

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} = 0, \quad j = 1, 2, \dots, N. \quad (6.6)$$

Em relação à métrica, tem-se que as habilidades serão estimadas na mesma métrica dos parâmetros dos itens.

Desenvolvendo a Equação 6.6 tem-se

$$\begin{aligned} \mathbf{h}(\boldsymbol{\theta}_j) &= \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} = \sum_{i=1}^p \left[x_{ij} \frac{\partial (\log P_{ij})}{\partial \boldsymbol{\theta}_j} + (1 - x_{ij}) \frac{\partial (\log Q_{ij})}{\partial \boldsymbol{\theta}_j} \right] \\ &= \sum_{i=1}^p \left[(x_{ij} - P_{ij}) \frac{W_{ij}}{P_{ij}^* Q_{ij}^*} \right] \left(\frac{\partial P_{ij}}{\partial \boldsymbol{\theta}_j} \right) \\ &= \sum_{i=1}^p (x_{ij} - P_{ij}) W_{ij} \mathbf{h}_{ij}, \end{aligned}$$

onde $\mathbf{h}_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right)$. Assim, é necessário resolver m equações da forma

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_{kj}} = \sum_{i=1}^p Da_{ik}(1 - c_i)(x_{ij} - P_{ij})W_{ij} = 0. \quad (6.7)$$

Estas equações não possuem solução explícita para θ_{kj} e, por este motivo, algum método iterativo, como o de Newton-Raphson ou o Escore de Fisher, é necessário para a obtenção das estimativas desejadas. A seguir são mostradas as derivadas segundas da equação de log-verossimilhança (Equação 6.5), necessárias à aplicação desses métodos.

A fórmula geral para o cálculo das derivadas segundas é dada por

$$\mathbf{H}(\boldsymbol{\theta}_j) \equiv \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_j'} = \sum_{i=1}^p (x_{ij} - P_{ij})W_{ij}[\mathbf{H}_{ij} - (x_{ij} - P_{ij})W_{ij}\mathbf{h}_{ij}\mathbf{h}_{ij}'],$$

com $\mathbf{h}_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right)$ e $\mathbf{H}_{ij} = (P_{ij} Q_{ij})^{-1} \left(\frac{\partial^2 P_{ij}}{\partial \theta_j \partial \theta_j'} \right)$.

Para formar a matriz \mathbf{H}_{ij} e o vetor \mathbf{h}_{ij} são necessárias as derivadas primeiras e segundas de P_{ij} em relação aos elementos do vetor $\boldsymbol{\theta}_j$. Desenvolvendo, tem-se que

$$\frac{\partial P_{ij}}{\partial \theta_{kj}} = Da_{ik}(1 - c_i)P_{ij}^* Q_{ij}^*$$

e

$$\frac{\partial^2 P_{ij}}{\partial \theta_{kj} \partial \theta_{tj}} = D^2 a_{ik} a_{it} (1 - c_i) P_{ij}^* Q_{ij}^* (1 - 2P_{ij}^*), \quad k = t = 1, 2, \dots, m.$$

Assim,

$$\mathbf{H}_{ij} = \begin{pmatrix} D^2 a_{i1}^2 \Delta & \cdot & \cdot & \cdot \\ D^2 a_{i1} a_{i2} \Delta & D^2 a_{i2}^2 \Delta & \cdot & \cdot \\ \vdots & \vdots & \ddots & \cdot \\ D^2 a_{i1} a_{im} \Delta & D^2 a_{i1} a_{im} \Delta & \cdots & D^2 a_{im}^2 \Delta \end{pmatrix}$$

e

$$\mathbf{h}_{ij} = \begin{pmatrix} Da_{i1}(1 - c_i) \\ Da_{i2}(1 - c_i) \\ \vdots \\ Da_{im}(1 - c_i) \end{pmatrix}.$$

Na matriz \mathbf{H}_{ij} , tem-se $\Delta = (1 - c_i)(1 - 2P_{ij}^*)$.

Agora, denotando por $\hat{\boldsymbol{\theta}}_j^{(t)}$ a estimativa de $\boldsymbol{\theta}_j$, $j = 1, 2, \dots, N$, na iteração t do algoritmo de Newton-Raphson, na iteração $t + 1$ tem-se

$$\hat{\boldsymbol{\theta}}_j^{(t+1)} = \hat{\boldsymbol{\theta}}_j^{(t)} - [\mathbf{H}(\hat{\boldsymbol{\theta}}_j^{(t)})]^{-1} \mathbf{h}(\hat{\boldsymbol{\theta}}_j^{(t)}).$$

Para a aplicação do método Escore de Fisher, substituem-se os componentes da matriz $\mathbf{H}(\boldsymbol{\theta}_j)$ pelos seus valores esperados e opera-se com a matriz de informação esperada, $\mathbf{H}_E(\boldsymbol{\theta}_j) = E(\mathbf{H}(\boldsymbol{\theta}_j))$. Assim, pelo método Escore de Fisher tem-se

$$\hat{\boldsymbol{\theta}}_j^{(t+1)} = \hat{\boldsymbol{\theta}}_j^{(t)} - [E(\mathbf{H}(\hat{\boldsymbol{\theta}}_j^{(t)}))]^{-1} \mathbf{h}(\hat{\boldsymbol{\theta}}_j^{(t)}).$$

Assintoticamente e sob algumas condições de regularidade, o estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}_j$ possui distribuição normal com vetor de médias $\boldsymbol{\theta}_j$ e matriz de covariâncias dada por $[-\mathbf{H}_E(\boldsymbol{\theta}_j)]^{-1}$.

Estimativas Iniciais

As estimativas iniciais para as habilidades podem ser dadas pelas estimativas dos fatores comuns do modelo da análise fatorial dada pela Equação 2.9

6.4. Estimação Conjunta dos Parâmetros dos Itens e das Habilidades

Nesta seção será abordado o caso mais comum, onde apenas é conhecida a matriz de respostas dos indivíduos; em outras palavras, não se conhece os parâmetros dos itens e nem as habilidades dos indivíduos. Nesta situação há duas abordagens possíveis: (a) a estimação conjunta dos parâmetros dos itens e das habilidades e (b) a estimação em duas etapas: primeiro a estimação dos parâmetros dos itens e, em seguida, das habilidades.

Estimação Conjunta

A estimação conjunta faz uso das duas seções anteriores. As equações a serem utilizadas foram definidas pelas Equações 6.4 e 6.7. Diferentemente das seções anteriores, onde a métrica (unidade de medida) foi estabelecida pelo conhecimento de um dos conjuntos de parâmetros (indivíduos ou itens), na estimação conjunta nenhum desses parâmetros é conhecido, portanto, não há métrica definida (veja Subseção 1.3.4). Este problema pode ser solucionado pela especificação de uma medida de posição e outra de dispersão para as habilidades. No caso multidimensional, será considerado média 0 e desvio-padrão 1 para cada uma das habilidades específicas, ou seja, para cada uma das dimensões do vetor habilidade, $\boldsymbol{\theta}$.

Considerando o modelo compensatório - MC3, cada item i , $i = 1, 2, \dots, p$, é caracterizado por um vetor de parâmetros de dimensão $m + 2$, $\boldsymbol{\zeta}_i = (a_{i1}, a_{i2}, \dots, a_{im}, c_i, d_i)'$. Como cada um dos j indivíduos, $j = 1, 2, \dots, N$, é caracterizado por um vetor habilidade de dimensão m , $\boldsymbol{\theta}_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{mj})'$, tem-se um total de $mN + (m + 2)p$ parâmetros a serem estimados simultaneamente.

Como exemplo, considere uma amostra de 300 indivíduos, $N = 300$, submetidos a um teste bidimensional, $m = 2$, contendo 40 itens, $p = 40$. Teríamos que trabalhar com a inversão de uma matriz de ordem 760. Vale ressaltar que no ENEM do ano de 1999 o número de inscritos foi de 315.960 indivíduos, ou seja, $N = 315.960$.

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

- independência condicional, implicando

$$\frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\theta})}{\partial \zeta_i \partial \zeta'_t} = 0, \quad i \neq t, \quad t = 1, 2, \dots, p;$$

- independência entre as respostas de indivíduos diferentes, implicando

$$\frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\theta})}{\partial \theta_j \partial \theta'_l} = 0, \quad j \neq l, \quad l = 1, 2, \dots, N;$$

- independência entre habilidades e itens, implicando

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

Assim, a matriz das derivadas segundas torna-se uma matriz bloco diagonal, na qual os p primeiros blocos são matrizes de ordem $(m + 2)$ relativas aos parâmetros dos itens e os N blocos restantes são matrizes de ordem m relativas às habilidades dos N indivíduos.

Embora a estrutura da matriz de derivadas segundas usadas no processo iterativo de Newton-Raphson seja simplificada pelas considerações acima, sua dimensão não é alterada.

Birbaum (1968), tratando do caso unidimensional, propôs um procedimento em dois estágios para contornar esse problema. De modo geral, esse procedimento pode ser aplicado ao caso multidimensional.

No primeiro estágio os itens são tratados independentemente e o procedimento de estimação apresentado na Seção 6.2 é realizado para cada item, assumindo as habilidades conhecidas. No segundo estágio as estimativas dos parâmetros dos itens obtidas na etapa anterior são assumidas verdadeiras e as habilidades são estimadas segundo o procedimento da Seção 6.3. Neste ponto, o processo retorna para o primeiro estágio continuando até que algum critério de parada seja alcançado.

Analisando sob o enfoque multidimensional, a grande vantagem desse método de 2 estágios é o tratamento de matrizes de ordem $(m + 2)$ no primeiro estágio e de ordem m no segundo estágio, onde m denota a dimensão do espaço $\boldsymbol{\theta}$. No caso unidimensional, $m = 1$.

Entretanto, há um sério problema inerente a este procedimento: a falta de consistência na estimação dos parâmetros dos itens (ou habilidades) na presença de um número muito grande de indivíduos (ou itens). O fato é que, para os parâmetros dos itens conhecidos, os Estimadores de Máxima Verossimilhança (EMV) das habilidades convergem para os seus verdadeiros valores quando o número de itens cresce; e com as habilidades conhecidas, os EMV dos parâmetros dos itens convergem para os seus verdadeiros valores quando o número de indivíduos cresce (Andersen, 1973; Baker, 1992).

Estimação em duas etapas: Máxima Verossimilhança Marginal

Bock & Lieberman (1970) desenvolveram um procedimento de estimação em duas etapas objetivando solucionar o problema de inconsistência na estimação conjunta. O

método assume que os indivíduos representam uma amostra de uma população em que a habilidade distribui-se de acordo com uma função densidade $g(\boldsymbol{\theta} \mid \boldsymbol{\eta})$, onde $\boldsymbol{\eta}$ é o vetor dos parâmetros da distribuição de $\boldsymbol{\theta}$.

A essência da solução é integrar a função de verossimilhança das respostas aos itens em relação a $\boldsymbol{\theta}$, removendo o efeito aleatório das habilidades. Assim, estimam-se os parâmetros dos itens pela maximização da verossimilhança marginal com o uso de algum método iterativo.

A segunda etapa consiste na estimação das habilidades dos indivíduos, que pode ser feita por máxima verossimilhança ou por método bayesiano. No caso bayesiano é utilizada a média ou a moda da distribuição posteriori condicional de $\boldsymbol{\theta}_j$ dado $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{pj})$ e $\boldsymbol{\zeta}_i = (a_{1i}, a_{2i}, \dots, a_{mi}, c_i, d_i)$, respectivamente, o vetor de respostas do j -ésimo indivíduo e o vetor de parâmetros do i -ésimo item.

Segundo Bock & Liberman (1970) a equação de verossimilhança a ser maximizada é dada por

$$\begin{aligned} L(\boldsymbol{\zeta}, \boldsymbol{\eta}) &= \prod_{j=1}^N P(\mathbf{x}_j) = \prod_{j=1}^N P(\mathbf{X}_j = \mathbf{x}_j \mid \boldsymbol{\zeta}, \boldsymbol{\eta}) \\ &= \prod_{j=1}^N \int_{\mathbb{R}^m} P(\mathbf{X}_j = \mathbf{x}_j \mid \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}) g(\boldsymbol{\theta} \mid \boldsymbol{\eta}) d\boldsymbol{\theta} \\ &= \prod_{j=1}^N \int_{\mathbb{R}^m} P(\mathbf{x}_j \mid \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} \mid \boldsymbol{\eta}) d\boldsymbol{\theta}. \end{aligned} \quad (6.8)$$

Conseqüentemente, a equação de log-verossimilhança é

$$\log L(\boldsymbol{\zeta}, \boldsymbol{\eta}) = \sum_{j=1}^N \log \left[\int_{\mathbb{R}^m} P(\mathbf{x}_j \mid \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} \mid \boldsymbol{\eta}) d\boldsymbol{\theta} \right],$$

onde $P(\mathbf{x}_j \mid \boldsymbol{\theta}, \boldsymbol{\zeta}) = \prod_{i=1}^p P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}$, com P_{ij} dado pela Equação 3.1. Por conveniência e sem perda de generalidade, o índice j foi excluído do vetor habilidade $\boldsymbol{\theta}$, pois o interesse está na distribuição da habilidade e não em valores particulares.

As equações de estimação para os parâmetros dos itens são dadas por

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

Essas equações estão desenvolvidas e bem detalhadas em Backer (1992) e Andrade Tavares & Valle (2000) utilizando o modelo logístico unidimensional de 3 parâmetros dado pela Equação 1.1. Para o caso multidimensional, utilizando o modelo compensatório (Equação 3.1), trabalha-se com integrais e somatórios múltiplos, mas as expressões finais sofrem poucas alterações devido às semelhanças entre os modelos. Por este motivo, nos restringimos a apresentar as equações finais:

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial a_{ik}} &= D(1 - c_i) \sum_{j=1}^N \int_{\mathbb{R}^m} \theta_k (x_{ij} - P_{ij}) W_{ij} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0, \quad k = 1, 2, \dots, m \\
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial d_i} &= (1 - c_i) \sum_{j=1}^N \int_{\mathbb{R}^m} (x_{ij} - P_{ij}) W_{ij} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0 \\
\frac{\partial \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial c_i} &= (1 - c_i) \sum_{j=1}^N \int_{\mathbb{R}^m} \left(\frac{x_{ij} - P_{ij}}{P_{ij}^*} \right) W_{ij} g_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0,
\end{aligned} \tag{6.9}$$

onde

$$g_j^*(\boldsymbol{\theta}) = P(\boldsymbol{\theta} | \mathbf{x}_j, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \frac{P(\mathbf{x}_j | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta})}{P(\mathbf{x}_j | \boldsymbol{\zeta}, \boldsymbol{\eta})} = \frac{P(\mathbf{x}_j | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta})}{\int_{\mathbb{R}^m} P(\mathbf{x}_j | \boldsymbol{\theta}, \boldsymbol{\zeta}) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}} \tag{6.10}$$

e

$$W_{ij} = \frac{P_{ij}^* Q_{ij}^*}{P_{ij} Q_{ij}}, \text{ com } P_{ij}^* = [1 + \exp(-D \mathbf{a}'_i \boldsymbol{\theta}_j - d_i)]^{-1} \text{ e } Q_{ij}^* = 1 - P_{ij}^*.$$

As $m + 2$ equações acima não possuem solução explícita sendo necessária a utilização de um método iterativo - Newton-Raphson, por exemplo. A aplicação do método exige a inversão da matriz de derivadas segundas. Entretanto, diferentemente da Seção 6.3, não é possível se valer da propriedade de independência condicional a fim de garantir que os parâmetros sejam estimados item a item. Desta forma, a matriz de derivadas segundas seria de ordem $p(m + 2)$.

Uma reformulação na abordagem de Bock & Liberman (1970) foi feita por Bock & Aitkin (1981). Basicamente eles introduziram a suposição de que os itens são independentes, de forma que

$$\frac{\partial^2 \log L(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \zeta_i \partial \zeta'_t} = 0.$$

Isso faz com que a matriz das derivadas segundas torne-se bloco diagonal e, conseqüentemente, que a estimação dos parâmetros dos itens possa ser feita item a item.

Vale ressaltar que a condição de independência condicional e a suposição de independência dos itens são completamente diferentes. A primeira está relacionada às respostas dos indivíduos, enquanto a segunda se refere apenas aos itens.

Outro fato é que as Equações 6.9 envolvem integração que não apresentam solução analítica e por isto algum método de aproximação deve ser empregado. Frequentemente é usado o método de Gauss-Hermite, também denominado método de quadratura gaussiana (Stroud & Secrest, 1966; Stroud, 1971).

Na utilização do método de quadratura considera-se que a função $g(\boldsymbol{\theta} | \boldsymbol{\eta})$ é contínua com integral finita. Esta função aplicada a um determinado ponto pode ser aproximada, para qualquer grau de precisão, por uma função que assume um número finito de pontos. A idéia básica é substituir a integral de uma função contínua pelo somatório das áreas de um número finito de retângulos.

O problema de integração múltipla da função $g(\boldsymbol{\theta} | \boldsymbol{\eta})$ é substituído pela obtenção sucessiva de integrais simples e, conseqüentemente, usando quadraturas o problema reduz-se

à obtenção de somatórios referentes às áreas de um número finito de retângulos. Em cada uma das m dimensões de $\boldsymbol{\theta}$ serão considerados q pontos de quadratura, $\bar{\theta}_{s_k}$, $s = 1, 2, \dots, q$ e $k = 1, 2, \dots, m$. Considerando conhecidos os pontos de quadratura (pontos médios dos retângulos) e os respectivos pesos A_{s_k} , com $A_{s_k} = g(\bar{\theta}_{s_k} | \boldsymbol{\eta}) \Delta_{s_k}$, onde $g(\bar{\theta}_{s_k} | \boldsymbol{\eta})$ e Δ_{s_k} correspondem, respectivamente, à altura e à base dos retângulos, a Equação 6.10 pode ser reescrita como

$$g_j^*(\bar{\boldsymbol{\theta}}_s) = \frac{P(\mathbf{x}_j | \bar{\boldsymbol{\theta}}_s, \boldsymbol{\zeta}) A_{s_1} A_{s_2} \cdots A_{s_m} \Delta_{s_1}^{-1} \Delta_{s_2}^{-1} \cdots \Delta_{s_m}^{-1}}{\sum_{s_1=1}^q \cdots \sum_{s_m=1}^q P(\mathbf{x}_j | \bar{\boldsymbol{\theta}}_s, \boldsymbol{\zeta}) A_{s_1} A_{s_2} \cdots A_{s_m}}. \quad (6.11)$$

Como as dimensões de $\boldsymbol{\theta}$ foram assumidas ortogonais, o peso associado a cada ponto de quadratura é dado pelo produto os pesos associados às suas coordenadas.

Para evitar o excesso de notação e simplificar as expressões das derivadas primeiras, a Equação 6.11 será redefinida por

$$g_j^*(\bar{\boldsymbol{\theta}}_s) = \frac{P(\mathbf{x}_j | \bar{\boldsymbol{\theta}}_s, \boldsymbol{\zeta}) A_{s_1} A_{s_2} \cdots A_{s_m}}{\sum_{s_1=1}^q \cdots \sum_{s_m=1}^q P(\mathbf{x}_j | \bar{\boldsymbol{\theta}}_s, \boldsymbol{\zeta}) A_{s_1} A_{s_2} \cdots A_{s_m}}. \quad (6.12)$$

As Equações 6.9 reescritas utilizando pontos de quadratura são

$$\begin{aligned} \frac{\partial \log L}{\partial a_{ik}} &= D(1 - c_i) \sum_{j=1}^N \sum_{s_1=1}^q \cdots \sum_{s_m=1}^q [\bar{\theta}_{s_k} (x_{ij} - P_{ij}) W_{is}] g_j^*(\bar{\boldsymbol{\theta}}_s), \quad k = 1, 2, \dots, m, \\ \frac{\partial \log L}{\partial d_i} &= (1 - c_i) \sum_{j=1}^N \sum_{s_1=1}^q \cdots \sum_{s_m=1}^q [(x_{ij} - P_{is}) W_{is}] g_j^*(\bar{\boldsymbol{\theta}}_s) = 0, \\ \frac{\partial \log L}{\partial c_i} &= \sum_{j=1}^N \sum_{s_1=1}^q \cdots \sum_{s_m=1}^q [(x_{ij} - P_{is}) W_{is}] (P_{is}^*)^{-1} g_j^*(\bar{\boldsymbol{\theta}}_s) = 0. \end{aligned} \quad (6.13)$$

Nas equações acima $g_j^*(\bar{\boldsymbol{\theta}}_s)$ deve ser calculado pela Equação 6.12.

Bock & Aitkin (1981) sugeriram ainda a utilização do algoritmo EM na obtenção das estimativas de máxima verossimilhança dos parâmetros dos itens. De modo geral, o algoritmo EM é um procedimento iterativo para a obtenção de estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade em situações onde há dados faltantes (Dempster, Laird & Rubin, 1977; Roderick & Donald, 1987).

No caso presente, o objetivo é obter estimativas de $\boldsymbol{\zeta}$, parâmetros dos itens, na presença das variáveis não observadas $\boldsymbol{\theta}$, habilidades dos indivíduos. Os dados incompletos estarão representados pelo vetor \mathbf{X}^o de respostas binárias dos indivíduos e os dados completos pelo vetor formato por \mathbf{X}^o e $\boldsymbol{\theta}^o$, onde o operador ‘o’ indica que uma matriz $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ é reescrita como um vetor de colunas da forma $(\mathbf{x}'_1, \dots, \mathbf{x}'_n)'$.

Assumindo a solução de Bock & Aitkin (1981), as Equações 6.13 são reescritas com as alterações necessárias para a utilização do algoritmo EM. Tem-se, então

$$\begin{aligned}
\frac{\partial \log L}{\partial a_{ik}} &= D(1 - c_i) \sum_{s_1=1}^q \cdots \sum_{s_m=1}^q \bar{\theta}_{s_k} (r_{is} - f_{is} P_{is}) W_{is} = 0, \quad k = 1, 2, \dots, m, \\
\frac{\partial \log L}{\partial d_i} &= (1 - c_i) \sum_{s_1=1}^q \cdots \sum_{s_m=1}^q (r_{is} - f_{is} P_{is}) W_{is} = 0, \\
\frac{\partial \log L}{\partial c_i} &= \sum_{s_1=1}^q \cdots \sum_{s_m=1}^q (r_{is} - f_{is} P_{is}) W_{is} (P_{is}^*)^{-1} = 0,
\end{aligned} \tag{6.14}$$

onde

$$r_{is} = \sum_{j=1}^N x_{ij} g_j^*(\bar{\theta}_s) \quad \text{e} \quad f_{is} = \sum_{j=1}^N g_j^*(\bar{\theta}_s).$$

As quantidades r_{is} e f_{is} representam, respectivamente, o número de indivíduos com habilidade $\bar{\theta}_s$ respondendo corretamente ao item i , em uma população de tamanho N , e o número de indivíduos com habilidade $\bar{\theta}_s$ respondendo ao item i , em uma população de tamanho N . Note que essas quantidades são desconhecidas e aqui faz-se útil o algoritmo EM. Essas quantidades são substituídas pelos seus valores esperados (condicionados a \mathbf{x} e ζ) nas Equações 6.14 e é dado início ao algoritmo EM. Os passos são os seguintes:

Passo E: Usar os pontos de quadratura $\bar{\theta}_{s_k}$, os pesos A_{s_k} e estimativas iniciais dos parâmetros dos itens, $\hat{\zeta}_i$, $i = 1, 2, \dots, p$, para gerar $g_j^*(\bar{\theta}_s)$ e, posteriormente, $\bar{r}_{is} = E(r_{is} | \mathbf{x}, \zeta)$ e $\bar{f}_{is} = E(f_{is} | \mathbf{x}, \zeta)$.

Passo M: Com os r 's e f 's obtidos no Passo E, resolver as equações de estimação para ζ_i , usando o algoritmo de Newton-Raphson ou Escore de Fisher através do procedimento apresentado na Seção 6.2.

Esses passos compõem cada iteração do algoritmo EM, que serão repetidas até que algum critério de convergência seja alcançado.

As habilidades dos indivíduos são obtidas após o processo descrito acima. Isto pode ser feito através do procedimento da Seção 6.3. Entretanto, Mislevy & Stocking (1989) recomendam a estimação por métodos bayesianos. A estimação das habilidades pela média da distribuição posteriori (estimação EAP) consiste em obter a esperança da posteriori, que pode ser escrita como

$$g(\boldsymbol{\theta} | \mathbf{x}_j, \zeta, \boldsymbol{\eta}) = \frac{P(\mathbf{x}_j | \boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta} | \boldsymbol{\eta})}{P(\mathbf{x}_j | \zeta, \boldsymbol{\eta})}.$$

A esperança é

$$\hat{\boldsymbol{\theta}}_j \equiv E[\boldsymbol{\theta} | \mathbf{x}_j, \zeta, \boldsymbol{\eta}] = \frac{\int_{\mathbb{R}^m} \boldsymbol{\theta} P(\mathbf{x}_j | \boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}}{\int_{\mathbb{R}^m} P(\mathbf{x}_j | \boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta} | \boldsymbol{\eta}) d\boldsymbol{\theta}}.$$

Esta forma de estimação tem a vantagem de ser calculada diretamente, não necessitando de métodos iterativos. Além disso, as quantidades necessárias para o seu cálculo são fornecidas na etapa final da estimação (via algoritmo EM) dos parâmetros dos itens descrita acima.

A estimação para os parâmetros dos itens através do algoritmo EM e as estimativas para as habilidades via método EAP estão implementadas no programa computacional TESTFACT 2.0. Este programa será utilizado para a modelagem dos dados do Exame Nacional do Ensino Médio do ano de 1999.

Capítulo 7

Aplicação a Dados Reais

7.1. Introdução

O Exame Nacional do Ensino Médio (ENEM) introduziu um novo conceito de avaliação educacional no Brasil. Diferentemente de avaliações disciplinares como, por exemplo, o Sistema Nacional de Avaliação do Ensino Básico (SAEB) e o VESTIBULAR, onde cada conhecimento é medido em testes individuais, o ENEM é um exame interdisciplinar, onde os vários conhecimentos associados aos conteúdos do ensino fundamental e médio são avaliados de uma só vez por um único teste. O grande diferencial do exame pode ser atribuído aos itens que o compõem. Cada um deles é elaborado de modo a avaliar até 5 competências, mesclando os conhecimentos de diferentes disciplinas.

As análises tradicionais de itens tornaram-se, em alguns casos, obsoletas com o surgimento e desenvolvimento da Teoria da Resposta ao Item (TRI). Entretanto, a aplicação dessa teoria estava restrita aos modelos unidimensionais. A grande importância do ENEM no contexto nacional motivou o estudo dos modelos multidimensionais da TRI.

O ENEM caracteriza-se como o exame do perfil de saída da escolaridade básica e tem como um dos principais objetivos fornecer ao participante subsídios para a sua autoavaliação (INEP, 1999; INEP 2000). Entretanto, com o aumento do número de instituições de ensino superior que vêm aderindo à utilização dos resultados do exame como parte de processos seletivos, o número de participantes tem aumentado significativamente.

Este capítulo apresenta a análise da dimensionalidade e a modelagem dos dados do Exame Nacional de Ensino Médio do ano de 1999 (ENEM-99). A verificação da dimensionalidade foi feita através da Análise Fatorial de Informação Plena introduzida no Capítulo 2. Na modelagem dos dados foi adotado o modelo compensatório com acerto casual (MC3) apresentado na Subseção 3.3.1.

7.2. Características do ENEM

O Exame Nacional do Ensino Médio é aplicado desde 1998 em todo território nacional. É constituído de um teste único contendo 63 itens de múltipla escolha e uma proposta para redação. Os itens objetivos e a redação destinam-se a avaliar as competências desenvolvidas pelos participantes ao longo da escolaridade básica.

O exame tem caráter voluntário e dele podem participar, mediante inscrição, os concluintes do ensino médio, no ano de realização do exame, e também os que já o concluíram em anos anteriores, em qualquer de suas modalidades.

O ENEM é estruturado por uma matriz de competências que define claramente os pressupostos do exame e delinea suas características operacionais. O modelo da matriz contempla a indicação das competências gerais próprias do aluno, na fase de desen-

volvimento cognitivo correspondente ao término da escolaridade básica, associadas aos conteúdos do ensino fundamental e médio.

De forma resumida, as cinco competências tratadas pelo ENEM são:

- I. Dominar linguagens;
- II. Compreender fenômenos;
- III. Enfrentar situações-problema;
- IV. Construir argumentação;
- V. Elaborar propostas.

7.3. Dados do ENEM

O ENEM é aplicado anualmente em todo território nacional a voluntários que concluíram ou estão em fase de conclusão do ensino médio. A parte objetiva do exame é composta de 63 itens. A ordem dos itens e/ou as opções de resposta correta são alteradas com o intuito de formar 4 testes “diferentes” cada um contendo os mesmos 63 itens. Esses testes são identificados pelas cores amarela, branca, rosa e verde, e a cada um deles corresponde aproximadamente 25,0% do total de inscritos. A distribuição dos testes segue uma seqüência alternada de cores de modo que todos eles estão presentes em todas as localidades de realização do exame.

O número de inscritos no ano de 1999 foi de 315.960 alunos em todo o Brasil. A distribuição para cada um dos 4 testes foi a seguinte: 80.251 para a cor amarela, 79.194 para a cor branca, 78.803 para a cor rosa e 77.712 para a cor verde.

Nesse trabalho, serão utilizados os dados referentes à parte objetiva do teste de cor amarela (teste amarelo). Como a distribuição dos testes segundo as cores é bem heterogênea em cada uma das localidades de aplicação do exame, tem-se uma amostra representativa da população e de tamanho significativo para a estimação dos parâmetros necessários à determinação da dimensionalidade dos dados e também para a estimação dos parâmetros dos itens. Para a estimação das habilidades dos participantes dispõe-se inicialmente de 63 itens, que é um número adequado - a literatura sugere um valor maior que 30.

7.4. Dimensionalidade do ENEM

O estudo da dimensionalidade do ENEM-99 foi precedido por uma análise preliminar dos 63 itens que o compõem. Estatísticas descritivas e algumas medidas da teoria clássica foram obtidas. Os resultados envolvendo o item 45 nos levaram a questionar sua utilização no estudo. Este item apresentou correlação bisserial negativa - a saber $corr_b = -0,117$. Isto significa que dos indivíduos que o responderam corretamente a maior parte foi do grupo de indivíduos de pior desempenho no exame. É provável - isto merece um estudo por especialistas em itens - que a formulação do item tenha induzido os candidatos a não optar pela alternativa correta. Além da correlação bisserial, todas as correlações tetracóricas envolvendo este item apresentaram valores negativos. Isto poderia comprometer as estimativas dos parâmetros necessários para a determinação da dimensionalidade.

Estes resultados levaram à exclusão do item 45 do estudo. Foram excluídos também 41 indivíduos que entregaram o exame sem responder nenhum item. Dessa forma, a base de dados do ENEM utilizada neste trabalho será composta por uma matriz de respostas referente a 62 itens e 80.210 indivíduos.

Continuando a análise descritiva, foram obtidos os percentuais de resposta para cada item. O objetivo foi verificar a não-resposta aos últimos itens do exame. Um percentual grande - considera-se grande qualquer valor acima de 20,0% - poderia indicar que o tempo especificado para a realização do teste foi insuficiente. Nos dados analisados o menor percentual de respostas a um item foi de 99,4%, o que sugere não ter ocorrido problema quanto ao controle de tempo no exame. Os passos seguintes referem-se propriamente à verificação da dimensionalidade.

Procedimentos

Na análise da dimensionalidade dos itens o primeiro passo foi a organização do banco de dados. O programa Statistical Package for the Social Sciences - SPSS, versão 10.0, foi utilizado para esse fim. Com ele procedeu-se à formatação da base de dados de forma conveniente à utilização dos programas BILOG 3 (Mislevy & Bock, 1990) e TESTFACT, versão 2.0.

O segundo passo foi a utilização do programa TESTFACT para determinar a dimensionalidade dos dados. Este programa tem como base teórica os artigos de Bock & Aitkin (1981) e Dempster, Laird & Rubin (1977). O TESTFACT vem sendo utilizado, principalmente, na verificação da dimensionalidade de testes. Nele se encontram implementadas a análise fatorial através da matriz de correlações tetracóricas e a análise fatorial de informação plena. O programa fornece ainda estatísticas descritivas dos itens e algumas medidas utilizadas na teoria clássica. O TESTFACT permite o uso de modelos com parâmetro de acerto casual, mas não possibilita a estimação destes parâmetros; exige a entrada destes já estimados. Utilizou-se, então, o programa BILOG 3 para este fim. O BILOG 3 é um programa computacional específico para análise de itens dicotômicos ou dicotomizados via TRI. Neste programa estão implementados os modelos unidimensionais logístico e ogiva-normal de 1, 2 e 3 parâmetros.

Análise Fatorial de Informação Plena - AFIP

A Análise Fatorial de Informação Plena será utilizada para determinar a dimensionalidade dos dados do ENEM-99 (teste amarelo), ou seja, para a determinação do número de fatores necessários ou adequados para a explicação desses dados. Uma grande vantagem da AFIP é a possibilidade de testar a significância estatística dos fatores adicionados sucessivamente ao modelo.

Os critérios utilizados para determinar a dimensionalidade são:

- i. a diferença nos valores corrigidos de qui-quadrado (veja Subseção 2.4.4);
- ii. a magnitude das cargas fatoriais após a rotação VARIMAX.

Laros et al. (2000) utilizaram um índice de unidimensionalidade definido como a razão

entre a mudança corrigida no qui-quadrado e os seus graus de liberdade. Um índice positivo maior que 2,0 implica que o modelo de $k + 1$ fatores se adequa melhor aos dados do que o modelo com k fatores. Índices positivos menores do que 2,0 indicam que o modelo de $k + 1$ fatores se ajusta melhor aos dados, mas que esta melhoria não é significativa. Índices negativos indicam que o modelo de k fatores se ajusta melhor aos dados do que o modelo de $k + 1$ fatores. Neste trabalho também será considerado tal índice - chamaremos índice de dimensionalidade (*ID*).

O programa TESTFACT utiliza as cargas fatoriais obtidas a partir da análise fatorial principal sobre a matriz de correlações tetracóricas (Divgi, 1979) como valores iniciais para o algoritmo EM na análise fatorial de informação plena. Este programa permite soluções para os principais problemas encontrados na construção da matriz de correlações tetracóricas. Essas soluções dizem respeito a possíveis correções antes da aplicação da análise fatorial. As correções são as seguintes: (i) correção para o acerto casual; (ii) correção para respostas omitidas; (iii) correção para obter uma matriz positiva definida e (iv) correção para evitar casos Heywood.

As correções para o acerto casual foram introduzidas na Subseção 2.4.2. Essas correções, entretanto, produzem não raramente um valor zero ou negativo em uma das caselas da tabela de contigência 2×2 de correlações tetracóricas. Valores negativos das proporções desta tabela também podem ser resultantes da codificação de todas as respostas omissas como respostas incorretas, após a correção para o acerto casual. O programa possibilita considerar apenas parte das respostas omissas como respostas incorretas. Isso é feito segundo as probabilidades de acerto casual. Por exemplo, se $c_i = 0,2$, então 20% das respostas omitidas no item i são codificadas como respostas corretas e 80% como incorretas. Esta correção deve preceder a correção para o acerto casual.

Mesmo com as duas correções citadas acima é possível que valores de proporções próximos de zero apareçam, ocasionando imprecisão nas correlações tetracóricas. Neste caso, o programa faz uso do método Centróide (Harmam, 1976) desenvolvido por Thurstone em 1931. Os procedimentos citados acima não garantem, entretanto, que a matriz de correlações tetracóricas seja positiva definida. O TESTFACT possibilita ainda um refinamento dessa matriz antes da análise fatorial principal ser realizada. Maiores detalhes podem ser vistos em Wilson et al. (1998). Em relação aos casos Heywood, caracterizados pelo aumento dos parâmetros a 's (veja Subseção 2.4.3), o programa permite atribuir distribuições *a priori* restrita em relação a alguns dos parâmetros dos itens. Como fase final, a matriz de correlações tetracóricas corrigida é submetida à análise fatorial principal com iterações de comunalidade. Esta análise é equivalente à análise fatorial MINRES (Harman, 1976).

Resultados

Os resultados da análise para a verificação da dimensionalidade do Exame Nacional do Ensino Médio do ano de 1999 são apresentados na Tabela 7.1.

A Tabela 7.1 mostra os valores da estatística qui-quadrado, χ^2 , para os modelos de 1 a 5 fatores com os respectivos graus de liberdade, gl , as diferenças dos qui-quadrados, χ^2_{dif} , entre um modelo de k fatores e outro de $k + 1$ fatores, $k = 1, 2, 3, 4$, os graus de liberdade

para essas diferenças, os valores corrigidos das diferenças dos qui-quadrados, χ^2_{difcorr} , e o índice de dimensionalidade, ID .

Foram ajustados sucessivamente os modelos de 1 a 5 fatores. Para cada fator adicionado ao modelo a estatística χ^2_{difcorr} foi calculada. Até o último modelo ajustado, modelo de 5 fatores, esta estatística apresentou significância, ou seja, houve melhoria no ajuste do modelo aos dados.

Tabela 7.1 - Estatísticas para o Número de Fatores do Modelo para o ENEM-99.

	Número de Fatores do Modelo				
	1	2	3	4	5
χ^2	4.006.317,82	3.992.022,79	3.981.784,75	3.976.992,41	3.974.241,79
gl	80.081	80.020	79.960	79.901	79.843
χ^2_{dif}		14.295,03	10.238,04	4.792,34	2.750,63
gl_{dif}		61	60	59	58
χ^2_{difcorr}		4.765,01	3.412,68	1.597,45	916,88
ID		78,11	56,88	27,08	15,81

Observe na Tabela 7.1 que os valores de χ^2 diminuem à medida em que são acrescentados fatores ao modelo. Conseqüentemente as diferenças corrigidas também diminuem, variando de forma decrescente de 4.765,01 a 916,88. A menor destas diferenças corresponde à diferença entre os valores de χ^2 dos modelos de 4 e 5 fatores dividida por 3 (veja Subseção 2.4.4: Teste de Ajuste do Modelo).

Segundo o índice de dimensionalidade, ID , também calculado para cada fator adicionado ao modelo, a conclusão é a mesma. Os índices decrescem de 78,11 (valor referente aos modelos de 1 e 2 fatores) até 15,81 (valor referente aos modelos de 4 e 5 fatores). Este último valor ainda está bem acima do valor crítico 2,0 citado na seção anterior, indicando a significância do modelo de 5 fatores.

O segundo critério utilizado foi a análise das cargas fatoriais. Estas cargas são apresentadas no Apêndice. As cargas fatoriais possibilitam informar o quanto da variância de cada variável (item) é explicada por cada fator e também quais itens se relacionam a cada fator - quanto maior a carga fatorial melhor é a relação entre item e fator. Neste trabalho considera-se que cargas fatoriais inferiores a 0,20 não contribuem para a mensuração dos respectivos fatores. Analisando as cargas fatoriais (após rotação VARIMAX) para cada um dos 3 primeiros fatores, nota-se que a grande maioria delas apresenta valores expressivos. Isto indica que quase todos os itens estão relacionadas a pelo menos um dos três fatores. Por exemplo, os itens $I3$, $I4$ e $I5$ estão bem relacionados aos fatores 1, 2 e 3. Já os itens $I21$ e $I27$ se relacionam melhor aos fatores 1 e 3. Para a mensuração do quarto fator contribuem quase 50% dos itens. Ao quinto fator estão relacionados apenas os itens $I9$, $I28$, $I29$, $I36$ e $I43$.

Um critério complementar para a decisão do número de fatores no modelo será o percentual de explicação da variância das variáveis (itens) devida a cada fator. A Tabela 7.2 mostra esses percentuais. O valor acumulado até o quinto fator foi de 45,33%. No

caso dos fatores não rotacionados, o primeiro fator é responsável por 39,85% e os demais fatores contribuem com percentuais inferiores a 2,06%. Sob essa análise poder-se-ia sugerir a utilização de um modelo de 1 fator (modelo unidimensional). Entretanto, a própria estrutura do exame sugere um modelo multidimensional. Objetivando uma melhor interpretação, os fatores foram submetidos a rotação VARIMAX. Com a rotação, os três primeiros fatores passaram a apresentar percentuais muito próximos. O percentual acumulado dos três primeiros foi de 38,54%. O quarto fator contribuiu com 5,55% e o quinto fator com 1,24%. Estes resultados retratam coerentemente os resultados obtidos anteriormente pela análise das cargas fatoriais.

Tabela 7.2 - Percentual da Variância Explicada por cada Fator do Modelo

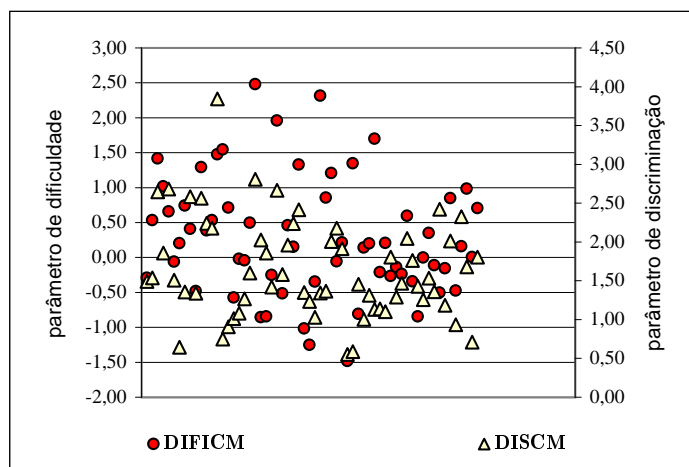
	Fatores				
	1º	2º	3º	4º	5º
Fator(es) não rotacionado(s)	39,85	2,05	1,54	1,06	0,82
Fator(es) após rotação varimax	13,49	12,31	12,74	5,55	1,24

A decisão final sobre o número de fatores do modelo, isto é, a dimensionalidade do referido instrumento de avaliação (ENEM), deve, entretanto, ser tomada em conjunto com os especialistas que o construíram.

7.5. Análise dos Itens do ENEM

Nesta seção é apresentado o resultado da estimação dos parâmetros dos itens do ENEM-99. O modelo adotado foi o MC3 (veja Subseção 3.3.1). Os itens são caracterizados pelos parâmetros de acerto casual (c), discriminação ($DISCM$) e dificuldade ($DIFICM$). No apêndice estes 3 parâmetros são apresentados para cada um dos 62 itens do exame. Os resultados estão representados graficamente na Figura 7.1.

Figura 7.1 - Representação gráfica dos parâmetros $DIFICM$ e $DISCM$ para os 62 itens do ENEM-99.



Com exceção do item 14, com parâmetro de discriminação igual a 3,84, todos os demais itens apresentaram discriminação no intervalo esperado de 0 a 3,00. Para o parâmetro de dificuldade *DIFICM*, todos os itens apresentaram valores no intervalo $-3,00$ a $3,00$, mais precisamente entre $-1,48$ e $2,48$. Na Figura 7.1 vê-se claramente que os itens estão bem distribuídos segundo esses dois parâmetros.

Os parâmetros c , d e a 's do modelo MC3 também estão presentes no Apêndice. Estes resultados permitem identificar, por exemplo, quais itens são comparáveis segundo seus parâmetros de dificuldade. Diferentemente do caso unidimensional, nem sempre a comparação é possível ou adequada (veja Capítulo 4). Para exemplificar, considere dois grupos de itens. O primeiro deles formado pelos itens *I1* e *I8* e o segundo composto pelos itens *I28* e *I29*. Analisando os valores dos parâmetros do modelo (especificamente os parâmetros a 's) é razoável comparar as dificuldades para os itens *I1* e *I8*. As dificuldades para os itens *I28* e *I29* também são comparáveis. Entretanto, as dificuldades entre itens de grupos diferentes não são comparáveis, pois os grupos não estão medindo as mesmas habilidades (fatores). Os itens *I28* e *I29* estão relacionados aos fatores 1, 2, 3 e 5, enquanto os itens *I1* e *I8* estão relacionados apenas aos fatores 1 e 2. Com esta análise é possível estruturar testes que possam medir combinações de habilidades em vários níveis de dificuldade.

Capítulo 8

Comentários e Sugestões Finais

A aplicação e o desenvolvimento da Teoria da Resposta ao Item (TRI) em muito dependem da disponibilização de programas computacionais que possam facilitar ou viabilizar sua utilização. Na Europa e nos Estados Unidos, a TRI é uma ferramenta comumente utilizada em diversas áreas profissionais, entre elas a Educacional. No Brasil a TRI é bem recente. Sua primeira aplicação foi na análise do Sistema Nacional de Avaliação Básica (SAEB) em 1995. Desde então, os órgãos governamentais, através do Ministério da Educação (MEC), vêm valorizando e incentivando o uso dessa teoria nas avaliações educacionais brasileiras. Entre outras avaliações podemos citar o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) e a Avaliação das Escolas Públicas do Estado do Rio Grande do Norte.

Este trabalho é pioneiro na utilização da TRI multidimensional no Brasil. Este estudo inicial possibilitou avaliar sob este novo enfoque o Exame Nacional do Ensino Médio caracterizado como um exame interdisciplinar e que por isto sugeria a utilização de modelos onde a habilidade fosse expressa por mais de uma dimensão. Até então, as aplicações se restringiam às avaliações unidimensionais.

Relativamente a outras áreas da psicometria, a TRI multidimensional ainda está na sua “infância”. Vários problemas precisam ser solucionados e metodologias ainda precisam ser desenvolvidas para ajudar na solução desses problemas. Contudo, os estudos têm ratificado que o relacionamento entre itens e indivíduos é bem mais complexo que o suposto nos procedimentos psicométricos usuais.

Entre as lacunas a serem preenchidas podemos citar, por exemplo, a questão do efeito sobre o escore total quando da inclusão de itens medindo múltiplas dimensões no teste (Reckase, 1989). Alguns trabalhos foram feitos no sentido de compreender melhor esta questão, mas esta pesquisa ainda está no seu início. Outro ponto pouco explorado diz respeito à equalização utilizando modelos multidimensionais. Esta é uma área de grande necessidade de pesquisa porque muitos dos testes educacionais utilizados, provavelmente, avaliam a habilidade em mais de uma dimensão e o estabelecimento de uma métrica comum no sentido de comparar os resultados desses testes ainda não está definida.

Neste trabalho apresentamos alguns dos modelos multidimensionais citados na literatura. Outros modelos estão presentes em algumas referências citadas, por exemplo, Linden et al. (1997). Há vários outros modelos ainda pouco explorados, como, por exemplo, a versão multidimensional para modelos de respostas a itens politômicos, os modelos multidimensionais longitudinais e os modelos multidimensionais multivariados. Tavares (2001) e Matos (2001) desenvolveram, respectivamente, modelos longitudinais e modelos multivariados unidimensionais.

Outra área que necessita de pesquisas adicionais é a de estimação dos parâmetros dos modelos multidimensionais. Embora bons programas existam (por exemplo, TESTFACT

e NOHAM), pouco é conhecido sobre o número de itens e/ou de respondentes necessários para a especificação adequada da dimensão do espaço latente. Perguntas como: (i) Qual é o relacionamento entre tamanho da amostra, a heterogeneidade da população de indivíduos e o número de dimensões que pode ser identificada? e (ii) O que significa dizer que duas dimensões são altamente correlacionadas mas distintas? ainda necessitam de estudos para que sejam respondidas. Sem dúvida, essa é uma área rica para futuras pesquisas.

A aplicação feita nesta dissertação considerou apenas os resultados de um dos 4 testes do Exame Nacional do Ensino Médio. Como a “diferença” nos testes está na ordem de apresentação dos itens e/ou das alternativas de resposta, uma questão importante seria avaliar se haveria mudanças significativas nos resultados obtidos para a análise da dimensionalidade dependendo do teste utilizado.

Outro ponto que necessita de pesquisas adicionais refere-se ao estabelecimento de critérios para a qualidade do item multidimensional. No caso unidimensional, um exemplo de item de má qualidade é aquele com parâmetro de discriminação negativo ou com valor positivo baixo. Para o caso multidimensional, as aplicações estão no início e um estudo detalhado faz-se necessário.

Por fim, gostaríamos de ressaltar que a aplicação apropriada dessa teoria exige fundamentalmente a integração de especialistas das áreas de estatística e educação.

Apêndice

Este Apêndice apresenta as cargas fatoriais, antes e após a rotação VARIMAX, fornecidas pelo programa TESTFACT. Os valores para os parâmetros c , d , a 's, $DIFICM$ e $DISCM$ para cada um dos itens, segundo o modelo MC3, são apresentados em seguida.

Resultados do programa TESTFACT: cargas fatoriais não rotacionadas.

ITEM	FATORES				
	1	2	3	4	5
1 I1	0.597	-0.111	-0.230	-0.110	0.015
2 I2	0.603	0.127	-0.242	-0.053	-0.091
3 I3	0.784	0.022	0.180	-0.165	-0.181
4 I4	0.676	0.111	0.158	-0.114	-0.191
5 I5	0.819	0.113	-0.018	-0.118	-0.124
6 I6	0.628	0.027	-0.021	-0.049	-0.204
7 I7	0.332	-0.101	0.016	-0.042	-0.055
8 I8	0.577	-0.127	-0.155	-0.102	-0.068
9 I9	0.794	0.011	-0.142	-0.055	0.208
10 I10	0.508	-0.320	-0.017	-0.124	-0.084
11 I11	0.794	0.243	0.023	-0.063	0.017
12 I12	0.779	0.122	-0.004	-0.077	-0.081
13 I13	0.756	-0.164	0.085	-0.121	0.039
14 I14	0.893	0.131	-0.026	-0.069	-0.128
15 I15	0.345	0.082	0.184	-0.056	-0.007
16 I16	0.456	-0.013	0.010	-0.019	0.112
17 I17	0.482	-0.139	0.056	-0.051	-0.070
18 I18	0.514	-0.101	-0.078	-0.071	0.050
19 I19	0.542	0.071	-0.236	0.004	-0.035
20 I20	0.646	0.216	-0.055	-0.046	0.006
21 I21	0.787	0.325	0.021	-0.075	0.005
22 I22	0.689	-0.268	-0.137	-0.140	0.043
23 I23	0.647	-0.308	-0.059	-0.148	-0.069
24 I24	0.615	0.117	-0.122	0.001	0.055
25 I25	0.820	0.140	-0.125	-0.039	-0.032
26 I26	0.676	0.014	-0.077	-0.019	-0.006
27 I27	0.705	0.240	-0.074	-0.010	-0.101
28 I28	0.764	-0.005	0.027	-0.075	0.206
29 I29	0.771	0.044	-0.007	-0.140	0.229
30 I30	0.578	0.091	-0.189	0.053	0.072
31 I31	0.523	-0.265	-0.008	-0.014	-0.024
32 I32	0.482	0.148	0.100	-0.006	0.063
33 I33	0.593	0.081	0.098	-0.048	0.114
34 I34	0.553	0.139	0.256	-0.015	0.014
35 I35	0.743	0.096	-0.126	-0.017	0.057
36 I36	0.755	0.033	0.179	0.000	0.131
37 I37	0.664	0.019	0.324	-0.022	0.109
38 I38	0.278	0.030	0.077	0.064	-0.085
39 I39	0.297	-0.000	-0.130	0.023	-0.020
40 I40	0.622	-0.183	0.021	-0.006	-0.037
41 I41	0.502	-0.037	0.024	-0.013	0.064
42 I42	0.549	-0.249	-0.054	-0.025	0.080
43 I43	0.517	0.028	-0.006	0.066	0.183
44 I44	0.513	0.084	0.190	0.057	0.015
45 I46	0.523	-0.098	0.088	0.066	0.004
46 I47	0.685	0.132	-0.131	0.157	-0.029
47 I48	0.574	-0.181	0.025	0.036	-0.028

ITEM		FATORES				
		1	2	3	4	5
48	I49	0.594	-0.190	0.187	0.022	-0.047
49	I50	0.759	0.065	0.014	0.052	0.086
50	I51	0.670	-0.039	-0.060	0.251	-0.028
51	I52	0.579	0.055	-0.151	0.228	-0.011
52	I53	0.551	-0.190	0.056	0.098	0.027
53	I54	0.619	-0.043	0.215	0.127	0.029
54	I55	0.567	-0.045	0.054	0.219	-0.124
55	I56	0.762	-0.068	0.011	0.287	-0.038
56	I57	0.520	0.023	0.217	0.083	-0.043
57	I58	0.760	0.005	0.017	0.056	0.048
58	I59	0.414	-0.123	-0.087	0.189	0.042
59	I60	0.771	-0.165	0.142	0.083	-0.046
60	I61	0.658	0.119	-0.090	0.179	-0.076
61	I62	0.345	-0.019	-0.040	0.163	0.002
62	I63	0.670	-0.264	-0.080	0.067	0.002

Resultados do programa TESTFACT: cargas fatoriais após rotação VARIMAX.

ITEM		FATORES				
		1	2	3	4	5
1	I1	0.438	0.452	0.122	0.125	0.089
2	I2	0.567	0.260	0.188	0.153	-0.022
3	I3	0.372	0.445	0.589	0.111	-0.107
4	I4	0.363	0.305	0.538	0.111	-0.130
5	I5	0.552	0.389	0.480	0.160	-0.042
6	I6	0.397	0.342	0.340	0.175	-0.138
7	I7	0.133	0.269	0.164	0.089	-0.017
8	I8	0.375	0.452	0.163	0.128	0.003
9	I9	0.532	0.419	0.327	0.214	0.294
10	I10	0.156	0.563	0.172	0.104	-0.020
11	I11	0.560	0.241	0.532	0.177	0.090
12	I12	0.515	0.343	0.467	0.181	-0.004
13	I13	0.300	0.551	0.433	0.160	0.118
14	I14	0.599	0.399	0.513	0.229	-0.039
15	I15	0.127	0.119	0.360	0.046	0.019
16	I16	0.236	0.245	0.250	0.133	0.159
17	I17	0.175	0.379	0.264	0.136	-0.017
18	I18	0.292	0.373	0.192	0.123	0.109
19	I19	0.489	0.250	0.134	0.190	0.028
20	I20	0.511	0.188	0.382	0.150	0.067
21	I21	0.602	0.178	0.556	0.153	0.074
22	I22	0.353	0.622	0.194	0.144	0.128
23	I23	0.268	0.637	0.221	0.129	0.011
24	I24	0.475	0.235	0.276	0.197	0.119
25	I25	0.615	0.345	0.397	0.231	0.053
26	I26	0.435	0.355	0.312	0.214	0.067
27	I27	0.569	0.198	0.404	0.207	-0.032
28	I28	0.404	0.408	0.437	0.180	0.283
29	I29	0.463	0.394	0.439	0.115	0.305
30	I30	0.474	0.225	0.189	0.238	0.137
31	I31	0.164	0.488	0.191	0.201	0.040
32	I32	0.276	0.125	0.385	0.132	0.104
33	I33	0.309	0.245	0.426	0.136	0.168
34	I34	0.218	0.163	0.542	0.144	0.056

ITEM		FATORES				
		1	2	3	4	5
35	I35	0.541	0.325	0.337	0.227	0.135
36	I36	0.314	0.344	0.554	0.238	0.200
37	I37	0.170	0.303	0.614	0.183	0.163
38	I38	0.116	0.103	0.213	0.151	-0.059
39	I39	0.244	0.161	0.058	0.129	0.017
40	I40	0.243	0.472	0.293	0.230	0.033
41	I41	0.242	0.288	0.278	0.159	0.116
42	I42	0.211	0.488	0.175	0.195	0.147
43	I43	0.283	0.213	0.273	0.225	0.235
44	I44	0.197	0.168	0.443	0.207	0.058
45	I46	0.174	0.323	0.310	0.247	0.059
46	I47	0.507	0.220	0.295	0.368	0.043
47	I48	0.207	0.432	0.266	0.252	0.037
48	I49	0.114	0.442	0.400	0.241	0.014
49	I50	0.432	0.322	0.435	0.293	0.161
50	I51	0.351	0.311	0.274	0.470	0.047
51	I52	0.409	0.203	0.188	0.409	0.055
52	I53	0.156	0.400	0.266	0.299	0.089
53	I54	0.164	0.298	0.470	0.323	0.086
54	I55	0.230	0.270	0.309	0.406	-0.063
55	I56	0.337	0.365	0.365	0.536	0.045
56	I57	0.152	0.214	0.445	0.243	0.002
57	I58	0.402	0.371	0.417	0.306	0.126
58	I59	0.189	0.258	0.094	0.334	0.093
59	I60	0.244	0.500	0.462	0.356	0.035
60	I61	0.459	0.209	0.305	0.381	-0.006
61	I62	0.181	0.148	0.130	0.273	0.041
62	I63	0.277	0.543	0.206	0.327	0.085

Os percentuais presentes na Tabela 7.2 são calculados pela fórmula $PVE_m = (\sum_{i=1}^p \lambda_{im}^2)/n$, onde PVE_m representa o percentual de variância explicada pelo m -ésimo fator, $m=1,2,3,4$ e 5, e λ_{im} denota a carga fatorial para o i -ésimo item, $i = 1, 2, \dots, n(n = 62)$, no fator m .

Parâmetros dos itens segundo o modelo MC3.

	ITEM	c	d	a1	a2	a3	a4	a5	DISCM	DIFICM
1	I1	0.139	0.437	1.137	0.701	-0.235	-0.563	0.261	1.491	-0.293
2	I2	0.207	-0.821	1.278	0.276	-0.576	-0.388	0.411	1.536	0.534
3	I3	0.142	-3.745	2.448	0.216	0.519	-0.194	0.803	2.644	1.416
4	I4	0.262	-1.885	1.732	-0.029	0.225	-0.034	0.630	1.857	1.015
5	I5	0.073	-1.760	2.446	0.359	-0.161	-0.388	0.952	2.682	0.656
6	I6	0.352	0.092	1.464	0.191	-0.087	-0.110	0.270	1.508	-0.061
7	I7	0.130	-0.131	0.572	0.249	0.115	-0.082	0.071	0.643	0.203
8	I8	0.222	-1.007	1.152	0.555	-0.095	-0.419	0.140	1.356	0.743
9	I9	0.251	-1.055	1.763	1.227	-0.342	-0.590	1.263	2.583	0.408
10	I10	0.389	0.647	1.048	0.669	0.357	-0.333	-0.096	1.339	-0.483
11	I11	0.126	-3.308	2.072	0.309	-0.287	-0.203	1.434	2.563	1.291
12	I12	0.237	-0.867	2.002	0.354	-0.160	-0.236	0.897	2.241	0.387
13	I13	0.149	-1.161	1.756	0.922	0.469	-0.335	0.692	2.178	0.533
14	I14	0.396	-5.671	3.497	0.601	-0.368	-0.359	1.384	3.843	1.476
15	I15	0.109	-1.159	0.571	0.002	0.220	0.059	0.430	0.750	1.545
16	I16	0.150	-0.648	0.642	0.423	0.024	-0.105	0.470	0.908	0.714
17	I17	0.041	0.583	0.896	0.386	0.219	-0.083	0.150	1.014	-0.575
18	I18	0.008	0.021	0.838	0.542	0.000	-0.287	0.307	1.083	-0.019

	ITEM	c	d	a1	a2	a3	a4	a5	DISCM	DIFICM
19	I19	0.025	0.053	1.005	0.387	-0.508	-0.274	0.334	1.266	-0.042
20	I20	0.225	-0.797	1.285	0.199	-0.337	-0.192	0.849	1.601	0.498
21	I21	0.234	-6.958	2.214	0.110	-0.420	-0.221	1.655	2.807	2.479
22	I22	0.064	1.736	1.510	1.133	0.137	-0.666	0.277	2.026	-0.857
23	I23	0.163	1.568	1.499	0.907	0.333	-0.512	0.039	1.856	-0.845
24	I24	0.047	0.356	1.093	0.435	-0.372	-0.199	0.673	1.420	-0.251
25	I25	0.129	-5.224	2.263	0.618	-0.553	-0.380	1.074	2.665	1.960
26	I26	0.130	0.817	1.345	0.550	-0.189	-0.201	0.560	1.581	-0.517
27	I27	0.172	-0.899	1.693	0.130	-0.499	-0.114	0.836	1.961	0.459
28	I28	0.062	-0.338	1.550	0.980	0.109	-0.349	1.226	2.236	0.151
29	I29	0.287	-3.211	1.622	0.936	0.049	-0.579	1.410	2.415	1.329
30	I30	0.127	1.376	0.971	0.528	-0.494	-0.178	0.571	1.351	-1.019
31	I31	0.136	1.544	0.964	0.722	0.219	-0.130	0.057	1.232	-1.253
32	I32	0.157	0.359	0.761	0.157	0.002	0.043	0.676	1.031	-0.348
33	I33	0.231	-3.106	0.984	0.380	0.099	-0.078	0.822	1.343	2.313
34	I34	0.164	-1.169	1.034	0.095	0.277	0.203	0.817	1.365	0.857
35	I35	0.235	-2.420	1.584	0.662	-0.416	-0.304	0.899	2.005	1.207
36	I36	0.139	0.121	1.609	0.744	0.300	0.082	1.225	2.177	-0.055
37	I37	0.161	-0.412	1.336	0.494	0.614	0.229	1.097	1.914	0.215
38	I38	0.035	0.819	0.499	0.064	0.016	0.169	0.152	0.553	-1.482
39	I39	0.026	-0.789	0.465	0.238	-0.218	-0.101	0.111	0.586	1.347
40	I40	0.112	1.178	1.234	0.699	0.185	-0.083	0.251	1.455	-0.810
41	I41	0.046	-0.142	0.784	0.442	0.057	-0.077	0.432	1.003	0.142
42	I42	0.158	-0.264	0.919	0.868	0.149	-0.237	0.220	1.313	0.201
43	I43	0.314	-1.922	0.699	0.577	-0.107	-0.003	0.668	1.131	1.699
44	I44	0.126	0.241	0.885	0.216	0.150	0.245	0.621	1.140	-0.211
45	I46	0.219	-0.226	0.897	0.509	0.146	0.126	0.334	1.101	0.205
46	I47	0.105	0.479	1.454	0.546	-0.620	0.125	0.670	1.806	-0.265
47	I48	0.074	0.176	1.074	0.667	0.150	0.004	0.211	1.290	-0.137
48	I49	0.117	0.347	1.205	0.612	0.464	0.160	0.305	1.469	-0.236
49	I50	0.160	-1.219	1.599	0.749	-0.159	0.008	1.020	2.045	0.596
50	I51	0.148	0.607	1.392	0.839	-0.371	0.369	0.438	1.763	-0.344
51	I52	0.006	1.210	1.068	0.604	-0.566	0.215	0.414	1.429	-0.846
52	I53	0.047	0.003	0.955	0.743	0.163	0.129	0.264	1.256	-0.002
53	I54	0.134	-0.534	1.173	0.574	0.275	0.395	0.644	1.533	0.348
54	I55	0.230	0.150	1.166	0.485	-0.103	0.436	0.235	1.361	-0.110
55	I56	0.080	1.218	1.928	1.135	-0.295	0.622	0.615	2.421	-0.503
56	I57	0.061	0.185	0.968	0.249	0.226	0.329	0.494	1.185	-0.156
57	I58	0.136	-1.711	1.636	0.797	-0.088	0.023	0.855	2.013	0.850
58	I59	0.041	0.445	0.616	0.643	-0.191	0.153	0.145	0.934	-0.477
59	I60	0.261	-0.373	1.962	1.006	0.372	0.267	0.580	2.325	0.161
60	I61	0.267	-1.649	1.402	0.459	-0.530	0.231	0.554	1.678	0.983
61	I62	0.015	-0.003	0.519	0.365	-0.183	0.183	0.182	0.709	0.004
62	I63	0.165	-1.272	1.398	1.120	0.028	-0.101	0.186	1.804	0.705

Referências

- [1] Andersen, E. B. (1973). Conditional Inference in multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26, 31-44.
- [2] Ackerman, T. (1996). Graphical representation of multidimensional item response theory analysis. *Applied Psychological Measurement*, 20, 311-329.
- [3] Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. 14 SINAPE.
- [4] Baker, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- [5] Birnbaum, A. (1957). *Efficient design and use of a mental ability for various decision-making problems*, (Series Report No. 58-16. Project No. 7755-23). USAF School of Aviation Medicine, Texas: Randolph Air Force Base).
- [6] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M, Lord & M.R. Novick. *Statistical Theories of Mental Test Score*. Reading, MA: Addison-Wesley.
- [7] Bock, R.D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- [8] Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an em algorithm. *Psychometrika*, 46, 443-459.
- [9] Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items and between tests. *Psychometrika*, 26, 347-372.
- [10] Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- [11] Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- [12] Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, 44, 169-172.
- [13] Fraser, C. & McDonald, R. P. (1988). *NOHARM II: A FORTRAN program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory* [Computer software]. Armidale, Australia: University of New Wngland, Centre for Behavioral Studies.
- [14] Haley, D. C. (1952). Estimation of the dosage mortality relationship when the dose is subject to error. *Technical Report No. 15*. Stanford, Calif.: Stanford University, Applied Mathematics and Statistics Labortory.
- [15] Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications* Boston: Kluwer Academic Publishers.

- [16] Harman, H. H. (1976). *Modern Factor Analysis* (3rd ed.) Chicago: University of Chicago Press.
- [17] Hattie, J. A. (1985). Methodology Review: assessing unidimensionality of tests and items: *Applied psychological Measurement*, 9, 139-164.
- [18] Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society*, Series A, 134, 486-501.
- [19] Hsu, Y. (1995). Item parameter estimation of a two-dimensional generalized MIRT model (Doctoral dissertation, University of Illinois, 1995). *Dissertation Abstracts International*, 57, 1584.
- [20] Instituto Nacional de Estudos e Pesquisas Educacionais - INEP/MEC(1999). *Exame Nacional do Ensino Médio - ENEM: Relatório Final*. Brasília: O Instituto.
- [21] Instituto Nacional de Estudos e Pesquisas Educacionais - INEP/MEC(2000). *Exame Nacional do Ensino Médio - ENEM: Documento Básico*. Brasília: O Instituto.
- [22] Jöreskog, K. G. & Sörbom, D. (1980). *EFAP II: Exploratory factor analysis program* [Computer program]. Chicago: International Educational Services.
- [23] Judge, G. G., Hill, R.C., Griffiths, W. E., Lutkepohl, H. and Lee, T. C. (1982). *Introduction to the Theory and Practice of Econometrics*. New York: John Wiley & Sons.
- [24] Laros, J. A., Pasquali, L. & Rodrigues, M. M. M. (2000). *Análise da Unidimensionalidade das Provas do SAEB*. Universidade de Brasília.
- [25] Lawley, D.N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh*, 6, 273-287.
- [26] Linden, W. J. van der & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- [27] Lim, C. (1993). An application of the joint maximum likelihood estimation procedure to a two-dimensional case of Sympon's noncompensatory IRT model (Doctoral dissertation, University of Iowa, 1993). *Dissertation Abstracts International*, 54, 2549.
- [28] Lord, F. M. (1952). A theory of test score. *Psychometric Monograph*, No. 7.
- [29] Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum Associates, Inc.
- [30] Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- [31] Matos, G. S. (2001). *Teoria da Resposta ao Item: Uma Proposta de Modelo Multivariado*. Dissertação de Mestrado. Pernambuco: DE/UFPE.
- [32] McDonald, R. P. & Ahlawat, W. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99.

- [33] Mislevy, R. J. (1986). Recent development in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11, 3-31.
- [34] Mislevy, R. J. & Stocking, M. L. (1989). A Consumer's Guide to LOGISTIC and BILOG. *Applied Psychological Measurement*, 13, 57-75.
- [35] Mislevy, R. J. & Bock, R. D. (1990). *BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models*. Chicago: Scientific Software, Inc.
- [36] Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 47, 355-366.
- [37] Pearson, K. (1900). On the correlation of characters not quantitatively measurable. *Royal Society Philosophical Transactions, Series A*, 195, 1-47.
- [38] Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9, 401-412.
- [39] Reckase, M. D. (1989, August). *Controlling the psychometric snake: Or, how I learned to love multidimensionality*. Invited address at the meeting of the American Psychological Association, New Orleans.
- [40] Reckase, M. D. & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15, 361-373.
- [41] Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika*, 1, 33-49.
- [42] Roderick J. A. L. & Donald B. R. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc.
- [43] Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. New Jersey: Prentice Hall, Englewood Cliffs.
- [44] Stroud, A. H. and Sechrest, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, New Jersey: Prentice-Hall.
- [45] Sympson, J. B. (1978). A model for testing multidimensional items. In D. J. Weiss (Ed.), *Proceedings of the 1977 computerized adaptive testing conference* (pp. 82-98). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods program.
- [46] Tavares, H. R. (2001). *Teoria da Resposta ao Item para Dados Longitudinais*. Tese de Doutorado. São Paulo: IME/USP.
- [47] Vianna, H. M. (1987). *Testes em Educação*. São Paulo: IBRASA.
- [48] Wilson, D.T., Wood, R. & Gibbons, R. (1998). *TESTFACT: Test Scoring, Item Statistics and Item Factor Analysis*. Chicago: Scientific Software, Inc.