

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

ANDRÉ LUIZ SILVA SAMARTINI

MODELOS COM VARIÁVEIS LATENTES APLICADOS À MENSURAÇÃO DE
IMPORTÂNCIA DE ATRIBUTOS

SÃO PAULO
2006

ANDRÉ LUIZ SILVA SAMARTINI

MODELOS COM VARIÁVEIS LATENTES APLICADOS À MENSURAÇÃO DE
IMPORTÂNCIA DE ATRIBUTOS

Tese apresentada à Escola de Administração de
Empresas de São Paulo da Fundação Getúlio
Vargas.

Campo de Conhecimento: Administração,
Análise e Tecnologia de Informação

Orientador: Prof. Dr. Wilton de Oliveira Bussab

SÃO PAULO
2006

Samartini, André Luiz Silva.

Modelos com Variáveis Latentes Aplicados à Mensuração de Importância de Atributos / André Luiz Silva Samartini. - 2006.
154 f.

Orientador: Wilton de Oliveira Bussab.

Tese (doutorado) - Escola de Administração de Empresas de São Paulo.

1. Questionários - Metodologia. 2. Pesquisa – Metodologia. I. Bussab, Wilton de Oliveira, 1940- II. Tese (doutorado) - Escola de Administração de Empresas de São Paulo. III. Título.

CDU 303.62

SAMARTINI, André Luiz Silva. Modelos com Variáveis Latentes Aplicados à Mensuração de Importância de Atributos. 2006. 154 f. Tese de Doutorado (Administração de Empresas) – Escola de Administração De Empresas da Fundação Getulio Vargas, São Paulo, 2006.

Resumo: O objetivo desta tese é modelar a importância de atributos de serviços ou produtos por meio de um modelo de Desdobramento Graduado Generalizado, pertencente à família de modelos da Teoria de Resposta ao Item.

Palavras-Chave: Teoria de Resposta ao Item, ponderação pela importância, Modelo de Desdobramento Graduado Generalizado, imputação de pesos.

ANDRÉ LUIZ SILVA SAMARTINI

MODELOS COM VARIÁVEIS LATENTES APLICADOS À MENSURAÇÃO DE
IMPORTÂNCIA DE ATRIBUTOS

Tese apresentada à Escola de Administração de
Empresas de São Paulo da Fundação Getúlio
Vargas.

Campo de Conhecimento: Administração,
Análise e Tecnologia de Informação

Data de aprovação:

___/___/___

Banca examinadora:

Prof. Dr. Wilton de Oliveira Bussab
(Orientador) FGV-EAESP

Prof. Dr. Abraham Laredo Sicsú
FGV-EAESP

Prof. Dr. Francisco José Espósito Aranha
FGV-EAESP

Prof. Dr. Dalton Francisco de Andrade
UFSC

Prof. Dr. Rinaldo Artes
IBMEC-SP

AGRADECIMENTOS

Primeiramente, agradeço o apoio da EAESP na realização deste trabalho, seja através de incentivos de pesquisa do Núcleo de Pesquisas e Publicações (NPP), seja pelo ambiente que esta Escola proporciona aos seus alunos, professores e funcionários que com certeza motiva cada um a dar o melhor de si.

Durante a elaboração da tese, tive o privilégio de ser orientado pelo Prof. Bussab, com quem passei várias tardes discutindo sobre a tese. Mais que somente privilégio, foi um prazer participar destas discussões, admirar suas idéias, sugestões e soluções e aprender um pouco de tanto a ser aprendido.

Ao Prof. Francisco Aranha pela ajuda acadêmica e ao amigo Chico por toda a ajuda, incentivo, conforto e compreensão - incondicionais - do primeiro ao último instante da realização deste trabalho.

Vários outros professores contribuíram com críticas e sugestões, entre eles o Prof. Dalton Francisco de Andrade, pelas contribuições feitas desde o início do trabalho e pela disponibilidade em me ajudar mesmo em período de férias e o Prof. Abraham Laredo Sicsú, na qualificação da tese. Com certeza estas contribuições foram de inestimável valor.

A todos os colegas da EAESP, sejam alunos, professores ou funcionários, que me falavam palavras de apoio nas conversas corriqueiras no restaurante ou no corredor.

Aos meus amigos, e principalmente ao Eduardo, agradeço pela paciência e compreensão comigo ao longo deste período em que muitas vezes não estive disponível ou – pior – estressado.

À minha família, principalmente aos meus pais, que me incentivaram e me deram todas as condições de estar aqui hoje.

RESUMO

Pesquisas de opinião freqüentemente utilizam questionários com muitos itens para avaliar um serviço ou produto. Nestas pesquisas, cada respondente, além de avaliar os itens segundo seu grau de satisfação ou concordância, deve também atribuir um grau de importância ao item. Com estas duas informações disponíveis para cada item, é possível criar uma medida resumo, na forma de um escore total composto pela avaliação dos itens ponderada pela sua importância. O objetivo desta tese é **modelar a importância** dos itens por meio de um modelo de Desdobramento Graduado Generalizado, pertencente à família de modelos da Teoria de Resposta ao Item. Resultados de uma pesquisa sobre academia de ginástica mostram que o modelo tem bom ajuste neste caso, e simulações mostram que é possível, com a utilização do modelo, montar desenhos experimentais para diminuir o número de itens ou categorias de resposta a serem perguntados aos respondentes sem perda de informação.

Palavras-chave:

Teoria de Resposta ao Item, ponderação pela importância, Modelo de Desdobramento Graduado Generalizado, imputação de pesos.

ABSTRACT

Many surveys use multi-items questionnaires to make an assessment of a service or product. In these surveys, each respondent, besides having to evaluate each item (attribute) according to the degree of satisfaction or agreement, must also evaluate its importance. It is possible, with these two informations, to create a global score of the product or service weighted by the importance of the items. The goal of this thesis is to **model the importance** through a Generalized Graded Unfolding Model, which is an Item Response Theory model. Results of a survey showed a good fit of the model, and simulations showed that it is possible to design experiments to reduce the number of items or response categories asked to the respondents without loss of information if this model is used.

Key-words

Item Response Theory, importance weighting, Generalized Graded Unfolding Model, weighting imputation

SUMÁRIO

1	INTRODUÇÃO	1
1.1.	CONCEITO DE IMPORTÂNCIA E UTILIDADE	2
1.2.	DEFINIÇÃO DO PROBLEMA	4
1.3.	OBJETIVO DA TESE	6
1.4.	ORGANIZAÇÃO DA TESE	9
1.5.	JUSTIFICATIVA DA TESE	10
2	REVISÃO TEÓRICA DOS MÉTODOS DE MENSURAÇÃO DE IMPORTÂNCIA	11
2.1.	INTRODUÇÃO	11
2.2.	ERROS DE MENSURAÇÃO	11
2.3.	ESCALAS	12
2.3.1.	Principais escalas não comparativas	12
2.3.2.	Principais escalas comparativas	14
2.4.	TÉCNICAS INDIRETAS DE OBTENÇÃO DE IMPORTÂNCIA	20
2.5.	ESTIMAÇÃO DA IMPORTÂNCIA	22
3	MODELOS DA TRI E APLICAÇÃO DO MODELO DE DESDOBRAMENTO GRADUADO GENERALIZADO (GGUM) PARA MENSURAÇÃO DE IMPORTÂNCIA	24
3.1.	INTRODUÇÃO	24
3.2.	MODELOS DA TRI CUMULATIVOS	26
3.2.1.	Principais modelos unidimensionais para uma população	26
3.2.2.	Função de Informação do Item	31
3.2.3.	Suposições do modelo	34
3.2.4.	Estimação dos parâmetros	35
3.3.	MODELO DE DESDOBRAMENTO GRADUADO GENERALIZADO (GGUM)	38
3.3.1.	Introdução	38
3.3.2.	Apresentação do GGUM	40
3.3.3.	O modelo dicotômico	44
3.3.4.	GGUM Politômico	52
3.3.5.	Função de Informação do Item	55
3.3.6.	Versões do GGUM com restrições adicionais nos parâmetros	56
3.3.7.	Estimação dos parâmetros	59
3.3.8.	Itens sem resposta	61
3.3.9.	Ajuste do modelo	62
3.3.10.	O software GGUM2004	67
3.4.	UTILIZAÇÃO DO GGUM NA TESE	68
4	INVESTIGAÇÃO EMPÍRICA: SIMULAÇÃO	69
4.1.	OBJETIVO	69
4.2.	METODOLOGIA	69
4.3.	DESCRIÇÃO DOS ITENS E INDIVÍDUOS UTILIZADOS NA PRIMEIRA SIMULAÇÃO	70
4.3.1.	Estimação dos parâmetros a partir dos dados simulados	71
4.3.2.	Alternativas de cálculo para o valor previsto	73
4.4.	SIMULAÇÕES COM DADOS EXCLUÍDOS	78
4.5.	RESULTADOS	80
4.6.	CONCLUSÃO	84
5	INVESTIGAÇÃO EMPÍRICA: PESQUISA DE CAMPO	85

5.1. OBJETIVO	85
5.2. METODOLOGIA DE CONSTRUÇÃO DOS QUESTIONÁRIOS	85
5.2.1. Escolha dos atributos	85
5.2.2. Desenho do questionário	88
5.2.3. Coleta de dados.....	89
5.3. ANÁLISE DESCRITIVA	90
5.3.1. Influência da ordem de apresentação das questões na resposta.....	91
5.3.2. Análise fatorial dos itens	92
5.4. MODELO SIMPLIFICADO - RESULTADOS.....	93
5.4.1. Agrupamento das categorias de resposta observadas	94
5.4.2. Estimativa dos parâmetros.....	94
5.4.3. Ajuste do modelo.....	96
5.4.4. Interpretação do resultado	96
5.5. MODELO COM 4 CATEGORIAS - RESULTADO	100
5.5.1. Ajuste do modelo.....	100
5.5.2. Interpretação do resultado	101
5.5.3. Resultado com respostas excluídas por respondente.....	105
5.5.4. Exclusão de atributos.....	106
5.6. RELAÇÃO DO TRAÇO LATENTE COM OUTRAS VARIÁVEIS	106
5.7. COMPARAÇÃO ENTRE O MODELO DICOTÔMICO E O MODELO COM QUATRO CATEGORIAS	107
5.8. A MEDIDA DE IMPORTÂNCIA COMO PESO PARA DETERMINAÇÃO DO ESCORE GLOBAL.....	108
6 PLANEJAMENTO DE UMA PESQUISA.....	110
6.1. INTRODUÇÃO.....	110
6.2. DESENHO DO EXPERIMENTO.....	110
6.3. ITENS JÁ CALIBRADOS E QUESTIONÁRIOS ADAPTATIVOS	111
7 CONCLUSÕES E LIMITAÇÕES DO ESTUDO	113
7.1. LIMITAÇÕES DO ESTUDO	113
7.2. SUGESTÕES DE ESTUDOS FUTUROS.....	113
7.2.1. GGUM multidimensional.....	113
7.2.2. Modelo da TRI bidimensional para modelar importância e avaliação.....	114
7.2.3. Modelos com escalas comparativas.....	114
7.3. CONTRIBUIÇÕES	117
8 REFERÊNCIAS BIBLIOGRÁFICAS	120
APÊNDICES	127

1 INTRODUÇÃO

Em 2000, o Governo Federal fez uma ampla pesquisa de opinião sobre o serviço público nas áreas de saúde, previdência e educação (BRASIL, 2000). Nesta pesquisa, os entrevistados avaliaram 22 atributos de cada área. Para cada atributo, o respondente deveria atribuir uma nota de 0 a 10 para a qualidade do serviço e também atribuir um “peso” ao atributo. Este peso seria utilizado para ponderar a nota dada aos atributos na composição da nota global do serviço.

Outra pesquisa, realizada pelo instituto AKATU (2005) sobre responsabilidade social empresarial, exigia que o entrevistado ordenasse 100 itens, descritos em frases, segundo a sua importância. Mesmo a alternativa em que o entrevistado atribui notas em uma escala de 0 a 10 pontos aos itens, em vez de ordená-los, não era viável devido ao grande número de frases a analisar.

Tal como as pesquisas citadas acima, muitas pesquisas de opinião utilizam questionários com muitos itens para avaliar um serviço ou produto. Cada item pode corresponder a um atributo; é possível também utilizar vários itens para avaliar um atributo. Neste trabalho, os termos “item” e “atributo” serão intercambiáveis, e apenas um item será utilizado para avaliar cada atributo.

Na pesquisa feita pelo Governo Federal, assim como outras, o respondente, além de avaliar cada atributo segundo seu grau de satisfação ou concordância, pode também atribuir a cada atributo um grau de importância para servir, por exemplo, como fator de ponderação para a construção de um escore global.

No meio acadêmico, há vários artigos recentes que abordam o tema da mensuração da importância de atributos (CARVALHO e LEITE, 1998; CUNHA, BORGES e FACHEL, 1998; GUIZI, MERLO & NAGANO, 2004).

A mensuração da importância dos atributos de um serviço ou produto é relevante, do ponto de vista da tomada de decisão em marketing, por dois motivos principais.

Primeiro, a mensuração da importância dos atributos pode ser utilizada para criar um escore total, por exemplo, da satisfação global com determinado produto. A satisfação, lealdade ou qualidade são construtos de grande interesse para pesquisadores da área de marketing. Quando se deseja calcular um escore global para a satisfação ou qualidade de um produto ou serviço, pode-se simplesmente somar os escores individuais de cada um dos atributos mensurados (MARCHETTI, 2001), obtendo um escore não ponderado (PARASURAMAN *et al.* 1988). Este procedimento pressupõe que, além da unidimensionalidade dos atributos, cada um tem o mesmo peso na formação do escore global. Quando isto não ocorre, é usual atribuir um peso para cada atributo, de acordo com a importância percebida pelo respondente. Desta forma, pode-se utilizar a importância especificada pelo respondente como fator de ponderação para as notas dos vários atributos do produto na construção do escore global.

Segundo, a importância – ou preferência – do consumidor é uma variável relevante em segmentação de mercado. HALEY (1985) mostra que o conhecimento dos benefícios, necessidades e desejos do cliente propiciam uma melhor ação de marketing do que o conhecimento de variáveis demográficas ou sócio-econômicas. Quando se faz uma segmentação com base nestas últimas variáveis, os grupos formados são parecidos entre si com relação às motivações de compra e benefícios procurados no produto ou serviço. Desta forma, a segmentação não possibilita a elaboração de um marketing *mix* adequado para cada segmento, pois, nas características relevantes para a estratégia de marketing, os grupos não diferem muito entre si. Além disto, quando são utilizadas apenas variáveis sócio-econômicas ou demográficas, muitas conclusões são óbvias: os consumidores mais ricos consomem mais, os homens têm comportamento diferente das mulheres, e assim em diante.

1.1. CONCEITO DE IMPORTÂNCIA E UTILIDADE

Há pouca concordância sobre o significado da importância – ou peso – e como mensurá-la. ANDERSON (1971) define importância como uma medida de foco de atenção ou concentração que um indivíduo dá para um atributo, dimensão ou informação em uma tarefa de julgamento ou tomada de decisão. GREEN e KREIGER (1985) definem importância como uma medida de impacto estatístico que o atributo, dimensão ou informação tem em uma tarefa específica ou situação. KEENEY e RAIFFA (1976) argumentam que a importância não pode

ser matematicamente separada do valor da escala para aquele atributo ou dimensão dado de maneira subjetiva por um indivíduo. Esta última definição baseia-se na teoria da utilidade e áreas correlatas, como análise conjunta.

A análise conjunta é um método decomposicional em que a estrutura de preferências dos indivíduos é estimada pela avaliação global do produto ou serviço. A partir de avaliações de vários produtos com diferentes atributos presentes, é possível encontrar a importância de cada atributo para cada indivíduo. Este último conceito surgiu da teoria de demanda do consumidor, desenvolvido por LANCASTER (1966), que postula que a utilidade total pode ser decomposta em utilidades referentes às características ou benefícios que determinado bem proporciona. Um consumidor escolhe um serviço ou produto de acordo com os atributos que o produto possui e de acordo com uma escala subjetiva de valor para cada atributo em relação ao seu custo.

A teoria da utilidade postula que

“Os indivíduos pertencem a uma determinada população homogênea Q , agem racionalmente e possuem informação, isto é, eles sempre selecionam aquela opção que maximiza sua utilidade pessoal (especificados como 'Homo economicus') sujeito legal, social, físico e orçamentário.” (DOMENCICH e McFADDEN, 1975; WILLIAMS, 1977)

Matematicamente, a utilidade que o indivíduo atribui a um produto ou serviço é dada por uma avaliação global ou escore global definido por:

$$U_j = \sum_{i=1}^I w_{ij} x_{ij} \quad (1.1)$$

Em que:

w_{ij} é o peso (ou a importância) do atributo i para o indivíduo j e

x_{ij} é a resposta (avaliação) dada pelo indivíduo j ao atributo i .

Por “resposta” ou “avaliação” entende-se uma medida sobre uma escala que, por exemplo, pode ser “1: presente/ 0: ausente”, “0: inexistente/ 1: pouco/ 2: muito” ou uma nota em uma escala de 1 a 10 indicando o nível de satisfação com aquele atributo.

Diante de dois produtos ou serviços com diferentes atributos ou diante de produtos ou serviços com os mesmos atributos em níveis de avaliação diferentes, o consumidor escolhe aquele que lhe dá maior utilidade.

Operacionalmente tem-se então duas matrizes de dados, \mathbf{W} e \mathbf{X} , conforme mostra o Quadro 1.1.

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{I1} & w_{I2} & \dots & w_{In} \end{bmatrix} \text{ e } \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \dots & x_{In} \end{bmatrix}$$

Quadro 1.1 – Matriz de importância w_{ij} e matriz de notas x_{ij} .

Usualmente, w_{ij} são escolhidos de tal forma que $\sum_{i=1}^I w_{ij} = 1$.

O escore global dos indivíduos é dado pelo vetor \mathbf{U} de dimensão N , definido por $\mathbf{U} = \text{diag}(\mathbf{W}'\mathbf{X})$.

1.2. DEFINIÇÃO DO PROBLEMA

Os elementos da matriz de pesos \mathbf{W} podem ser obtidos por métodos indiretos (decomposicionais), que partem de avaliações globais de produtos ou serviços com diferentes níveis de atributos; ou diretos (composicionais), em que se pergunta ao entrevistado a importância de cada atributo considerado.

Quando a pergunta é feita diretamente, a importância pode ser medida por escalas comparativas (usualmente *rankings*) ou não-comparativas (*ratings*). Neste caso, não se costuma modelar w_{ij} . Os valores observados já são as estimativas da importância do atributo i para o indivíduo j .

Tradicionalmente, todos os respondentes avaliam todos os atributos. Neste caso, todos os elementos da matriz de importância são conhecidos. Mas quando um questionário com I atributos e N respondentes é aplicado, nem sempre é possível obter as $I \times N$ importâncias da matriz W . Isto ocorre quando o número de atributos ou itens a ser analisado é muito grande e atribuir a importância a cada um deles é uma tarefa cansativa. Neste caso, a importância w_{ij} é observada para $(i,j) \in E$ e não é observada (*missing*) para $(i,j) \notin E$, onde E é uma submatriz de uma matriz $I \times N$ e pode ser fixada a priori. Obtém-se então somente parte dos elementos da matriz W .

Há várias maneiras de estimar as importâncias w_{ij} não observadas. Pode-se, por exemplo, utilizar a importância média dada ao item, \bar{w}_i . Supondo-se que as importâncias de cada item podem ser modeladas com o mesmo valor para todos os indivíduos, pode-se utilizar a média da importância dada àquele item por todos os indivíduos. A matriz W pode neste caso ser representada por um vetor $W' = (\bar{w}_1 \quad \bar{w}_2 \quad \dots \quad \bar{w}_I)$ e o vetor de escores globais para os indivíduos é dado por $U = W'X$,

o que, em notação não matricial, corresponde a

$$U_j = \sum_{i=1}^I \bar{w}_i x_{ij} \quad (1.2)$$

Neste caso, a importância média observada do atributo i substitui na matriz W todos os elementos w_{ij} , isto é, tanto as importâncias observadas quanto as não observadas são substituídas.

Alternativamente, a importância média pode substituir somente os w_{ij} não observados. O escore global para o indivíduo j , U_j , neste caso é dado por:

$$U_j = \sum_{i \in E_j} w_{ij} x_{ij} + \sum_{i \notin E_j} \bar{w}_i x_{ij}, \quad (1.3)$$

em que E_j é o conjunto dos itens respondidos pelo indivíduo j .

O problema em se utilizar a importância média para calcular o escore global é que esta abordagem admite o mesmo padrão de importância para os indivíduos, embora estes tenham percepções diferentes sobre os atributos e portanto atribuem importância distinta a cada um deles.

Assim, o problema a ser endereçado nesta tese é encontrar um modelo para a matriz de importância que incorpore padrões de importância diferentes para cada indivíduo, mesmo quando a matriz \mathbf{W} tem caselas vazias.

1.3. OBJETIVO DA TESE

Conforme visto anteriormente, a matriz \mathbf{W} possui $I \times N$ parâmetros w_{ij} que usualmente não são descritos através de um modelo. Um primeiro objetivo é encontrar um modelo que descreva a matriz de importância \mathbf{W} (ilustrada na Figura 1.1) a partir de características do indivíduo (θ) e do atributo (ζ).

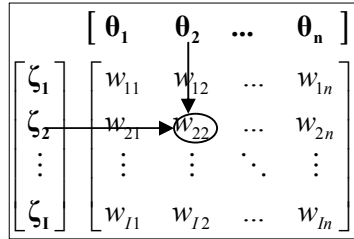


Figura 1.1 - estrutura da matriz \mathbf{W} em função de parâmetros do atributo e do indivíduo.

Isto significa que w_{ij} será obtida através de uma função que dependa de parâmetros do item i , ζ_i e do indivíduo j , θ_j , isto é, $w_{ij} = f(\zeta_i, \theta_j)$.

O objetivo da tese é mostrar que a Teoria de Resposta ao Item (TRI) pode ser utilizada para modelar a matriz de importância \mathbf{W} em função de parâmetros do item e do indivíduo.

O modelo de TRI considera que cada indivíduo é representado por um ponto em um *continuum* unidimensional que representa um “traço latente” que define sua opinião em relação à importância dos atributos. Os atributos podem ser representados neste mesmo

continuum. A importância w_{ij} é resultado de uma função derivada dos valores dos parâmetros do indivíduo e do item neste *continuum*.

Quando a matriz completa \mathbf{W} é modelada pela TRI, pode-se interpretar os itens e os indivíduos de acordo com suas posições no *continuum* latente. É possível identificar itens que não agregam informação (pois todos os indivíduos o avaliam da mesma forma, independente de sua posição no *continuum* latente) e indivíduos com comportamento muito diferente dos outros. Esta última informação é importante para segmentar os indivíduos, que podem ser descritos pela sua posição neste *continuum*.

Se as importâncias w_{ij} modeladas pela TRI se ajustarem bem às importâncias observadas, é possível estudar o que aconteceria com estas estimativas se a matriz de importância for incompleta, isto é, se houver dados faltantes.

Conforme mencionado anteriormente, quando há dados faltantes costuma-se utilizar um modelo em que w_{ij} é obtido pela resposta média ao atributo dada pelos indivíduos. O modelo de TRI, mesmo sem observar a matriz completa de importância w_{ij} , não impõe que as importâncias dadas aos atributos sejam iguais para todos os indivíduos.

Matematicamente, busca-se um modelo em que o número de parâmetros a estimar não esteja em nenhum dos dois extremos, isto é, não considere as importâncias iguais para todos os indivíduos em cada atributo (um modelo com I parâmetros a estimar) e nem considere a importância tão desestruturada a ponto de ser necessário estimar $I \times N$ parâmetros. O problema desta última abordagem é que o modelo, por ser “saturado”, não permite estimar valores faltantes na matriz, o que implica que todos os respondentes devem atribuir importância a todos os atributos.

O modelo baseado na Teoria da Resposta ao Item é um modelo intermediário aos apresentados no parágrafo anterior na medida em que o número de parâmetros a estimar não é tão grande quanto o modelo saturado e também não é tão restrito quanto o modelo de importâncias iguais para cada atributo.

Caso o modelo de TRI se ajuste à matriz de importância observada, o próximo passo é mostrar que, quando há informação incompleta da matriz de importância, as estimativas dos

parâmetros dos itens e dos indivíduos são próximos aos valores obtidos com a matriz completa. Caso isto se confirme, o modelo de TRI permitirá que as estimativas de importância dos itens sejam obtidas exigindo-se do respondente menor esforço, evitando-se assim o cansaço comum em questionários longos.

Isto ocorre pois os entrevistados respondem apenas a um subconjunto do total de atributos estudados. Por exemplo, uma pesquisa pode conter 50 atributos de determinado produto, mas cada indivíduo avalia apenas 20 itens. A TRI permite que sejam estimadas as respostas de cada indivíduo para os 30 itens não respondidos, respeitado o comportamento de resposta dado nos 20 itens respondidos.

Além disto, o modelo que utiliza TRI tem melhor aderência (respeita os *ratings* dados pelos respondentes) às importâncias observadas do que os modelos que utilizam a importância média de todos os respondentes como estimador.

Em resumo, as hipóteses da tese são que a TRI pode ser utilizada para modelar a matriz de importância \mathbf{W} e que as estimativas de w_{ij} podem ser obtidas por este modelo mesmo quando a matriz \mathbf{W} é incompleta. Evidenciada esta hipótese, a utilização do modelo permitirá um melhor entendimento do comportamento dos indivíduos, tornando a segmentação mais fácil e permitindo um menor cansaço e esforço mental por parte dos respondentes quando respondem o questionário.

Também é discutida nessa tese a criação dos pesos (importância) dos atributos, w_{ij} , através de vários métodos. Entre eles, destaca-se a utilização do valor esperado obtido pelo modelo da TRI e da probabilidade de resposta na categoria “importante” ou “muito importante”, também estimada pelo modelo da TRI. Na hipótese de se trabalhar com a probabilidade do indivíduo responder em determinada categoria, será estudada a possibilidade de agregar categorias de resposta no questionário para torná-lo mais simples, por exemplo com duas categorias, agregando “importante” e “muito importante” em uma só. Caso as estimativas de w_{ij} neste caso sejam próximas quando se utiliza probabilidade, é vantajoso trabalhar com um questionário mais simples.

Neste trabalho, por exemplo, é utilizada uma escala de importância com quatro pontos (não é importante / é pouco importante / é importante / é muito importante). Se a probabilidade (dada

pelo modelo da TRI) do indivíduo escolher uma das duas últimas categorias for, digamos, 62% e se o mesmo questionário for aplicado com apenas duas categorias (não é importante / é importante), e o modelo da TRI estimar a probabilidade do indivíduo escolher a última categoria for próxima de 62%, e vantajoso utilizar esta última escala, com apenas duas categorias.

1.4. ORGANIZAÇÃO DA TESE

No Capítulo 2, é feita uma revisão teórica de vários métodos de obtenção de importância presentes na literatura. Estes métodos podem ser divididos em dois grandes blocos: os métodos de obtenção de importância diretos e indiretos. A revisão teórica indica que ambos os métodos possuem vantagens e desvantagens, e portanto não há superioridade de um em relação ao outro. Constatado este fato, discute-se em seguida vários tipos de escalas, que são divididos basicamente entre comparativas e não comparativas.

O modelo de TRI proposto na tese para obter importância diretamente utilizando escala não comparativa é apresentado no Capítulo 3, juntamente com as definições de modelos mais tradicionais de Teoria de Resposta ao Item. Neste capítulo também são apresentados métodos de estimação, interpretação dos parâmetros, estatísticas de ajuste do modelo e a aplicação deste modelo à obtenção de importância.

No Capítulo 4 são apresentadas as simulações realizadas com objetivo de verificar o comportamento do modelo quando há dados faltantes. Várias simulações são realizadas, variando-se o número de respondentes, o número de itens e a porcentagem de dados faltantes.

No Capítulo 5 discute-se a análise de dados obtidos em uma pesquisa de campo, que ilustra uma situação real em que foram coletados dados sobre a importância de 25 atributos na escolha de uma academia de ginástica. Novamente são realizadas várias simulações, com exclusão de algumas respostas e avaliação das estimativas obtidas com o modelo de TRI. Também é estudada neste capítulo a possibilidade de diminuição do número de categorias de resposta (quatro para duas) e os impactos da redução de categorias sobre as estimativas.

O Capítulo 6 apresenta procedimentos e recomendações para se realizar uma pesquisa de campo que utilize o modelo proposto; no Capítulo 7, são apresentadas as contribuições da tese, limitações do estudo e direções para estudos futuros.

1.5. JUSTIFICATIVA DA TESE

As pesquisas de mercado feitas atualmente, devido a sua extensão, geram resultados com baixa confiabilidade e muitos questionários não respondidos ou respondidos de forma incoerente. Uma pesquisa sobre um produto que considera 25 atributos, por exemplo, torna-se extensa pois o entrevistado deve avaliar os atributos e a sua importância, gerando 50 avaliações. Além disto, costuma-se perguntar nestes questionários características sócio-econômicas e demográficas do entrevistado tornando o questionário ainda mais extenso e cansativo.

Com a aplicação da TRI, é possível desenhar questionários menores, com perda mínima de informação por um lado e ganho de confiabilidade, por exigir menor esforço do entrevistado, de outro lado.

A representação dos itens e dos indivíduos em uma escala unidimensional proporcionada pelo modelo permite também uma interpretação dos padrões de resposta que pode ser utilizada para segmentação de mercado ou outros modelos explicativos do comportamento de preferência dos indivíduos.

2 REVISÃO TEÓRICA DOS MÉTODOS DE MENSURAÇÃO DE IMPORTÂNCIA

2.1. INTRODUÇÃO

Há na literatura vários métodos destinados a mensurar importância de atributos (CARVALHO e LEITE, 1998; CRONIN e TAYLOR, 1994; PARASURAMAN, ZEITHAML e BERRY, 1994). Neste capítulo são apresentados os principais métodos de mensuração, suas vantagens e desvantagens, tipos de escalas e os erros de mensuração mais comuns.

2.2. ERROS DE MENSURAÇÃO

O principal problema relacionado aos métodos de mensuração de importância são erros de mensuração, derivados da escolha da escala e da dificuldade do respondente realizar a tarefa devido ao cansaço, falta de capacidade intelectual ou tempo requerido.

Segundo PAULHUS (1991), outros erros de mensuração são resultantes da desejabilidade social, que ocorre quando o indivíduo responde o esperado; e do viés de concordância, em que os respondentes tendem a avaliar favoravelmente todos os atributos ou a dizer coisas boas a respeito do objeto analisado.

PASQUALI (1997) acrescenta que o respondente pode também exibir um padrão sistemático de respostas, chamado de efeito halo. Isto ocorre quando o entrevistado tende a avaliar todos os atributos da mesma maneira, como se levasse em conta o produto ou serviço como um todo na avaliação dos atributos individuais. O efeito halo também pode ocorrer quando há um atributo dominante na análise do indivíduo. Isto gera uma correlação espúria entre as avaliações dos atributos, prejudicando a confiabilidade da escala.

Finalmente, também pode haver erros de tendência central, que ocorrem quando o respondente tende a colocar todos os itens no centro da escala, evitando julgamentos

extremos. Um estudo feito por CHURCHILL (1991) sobre intenção de compra mostrou que a escala com quatro pontos (sem ponto neutro) apresentou maior porcentagem de respostas negativas do que a escala com cinco pontos (com ponto neutro), indicando uma tendência das pessoas a assumir uma posição neutra em vez da negativa.

A ocorrência em maior ou menor grau de cada um destes erros depende da maneira como se pergunta sobre a importância dos atributos para os respondentes e depende do tipo de escala utilizada.

2.3. ESCALAS

Segundo GUIZI, MERLO e NAGANO (2004), uma escala é um *continuum*, no qual os objetos (nesta tese, indivíduos e atributos) são identificados de acordo com a quantidade que possuem da característica medida.

MALHOTRA (2001) classifica as escalas em dois grupos: comparativas e não comparativas. De um modo geral, as escalas não comparativas utilizam notas (*ratings*) para avaliar o atributo. Desta forma, é possível avaliar os atributos independentemente, isto é, o respondente pode avaliar cada atributo isoladamente, sem levar em consideração os demais.

As escalas comparativas comumente utilizam ordenação (*rankings*), e na maior parte das vezes o empate – ou a repetição do *ranking* – não é permitido. Desta forma, o *ranking* de um determinado atributo depende da presença de outros atributos, e com isto as respostas têm uma estrutura de dependência.

A seguir, são apresentadas as principais escalas existentes na literatura.

2.3.1. Principais escalas não comparativas

As escalas de Diferencial Semântico e de Likert, que são as escalas não comparativas mais utilizadas, são apresentadas a seguir.

2.3.1.1. Escala de Diferencial Semântico

Os pontos extremos desta escalas são caracterizados por adjetivos antônimos, por exemplo, complicado/simples, fácil/difícil, rápido/devagar e, no caso de interesse deste estudo, importante/sem importância. Entre eles são disponibilizados pontos intermediários, que podem ou não ter descrição.

O respondente assinala o ponto da escala que melhor indica a descrição do objeto analisado naquele atributo. É possível também criar uma divisão da escala em cinco ou sete pontos, caracterizados numericamente e o respondente assinala o número que mais se aproxima da descrição do objeto. Em alguns casos, pede-se ao respondente para marcar, em um segmento de reta de, por exemplo, 10cm, o ponto que indica a melhor descrição do objeto. Para quantificar a avaliação, mede-se a distância em centímetros do ponto assinalado a um dos extremos.

Segundo MALHOTRA (2001), existe controvérsia sobre se os dados devem ser tratados como provenientes de uma escala intervalar. Adicionalmente, por ser não comparativa, esta escala tem a desvantagem de não discriminar muito os atributos em relação à importância, já que o respondente tende a avaliar os atributos de modo semelhante – o chamado efeito halo.

2.3.1.2. Escala de Likert

A escala de Likert é similar à escala de Diferencial Semântico, mas os extremos da escala são “discordo” e “concordo”, isto é, o respondente deve assinalar o ponto da escala de acordo com o nível de concordância com a afirmação feita.

No caso do presente estudo, em que se deseja estudar importância de cada item, os extremos da escala poderiam ser caracterizados como “Concordo Totalmente” e “Discordo Totalmente” que o item em questão é importante.

Esta escala apresenta os mesmos problemas da escala de Diferencial Semântico, por ser não comparativa.

2.3.2. Principais escalas comparativas

Na escala comparativa, o respondente avalia dois ou mais atributos simultaneamente. Os tipos de escalas comparativas mais comuns na literatura são: comparação por pares, ordenação dos postos (*ranking*), soma constante, pontos decrescentes e *best-worst*.

2.3.2.1. Comparação de pares

Nesta técnica, o entrevistado compara um par de atributos por vez, indicando qual dos elementos do par considera mais importante. No resultado das comparações, conta-se quantas vezes cada objeto foi escolhido pelo respondente.

Esta técnica tem a desvantagem de ter uma implementação difícil quando o número de atributos é grande. Com 20 atributos, por exemplo, são necessárias 190 comparações. Uma das alternativas é considerar apenas um subconjunto das comparações, mas é recomendável fazer isto de forma balanceada para que todos os atributos sejam comparados o mesmo número de vezes.

2.3.2.2. Best-Worst

Na escala Best-Worst, também chamada de *maximum difference scaling*, o respondente deve escolher, entre vários atributos (pelo menos três), qual par é o mais dispar entre todos, isto é, o par escolhido é o que apresenta os dois atributos mais distantes na escala medida (no caso, importância).

Por exemplo, se preço, nome da marca e qualidade são os atributos considerados, o respondente deve escolher entre os pares: a) preço e nome da marca; b) preço e qualidade; c) nome da marca e qualidade; d) nome da marca e preço; e) qualidade e preço e f) qualidade e nome da marca. O par escolhido é o que, na opinião daquele respondente, possui maior diferença de importância (mais importante/ menos importante).

Esta escala assume que a diferença entre cada par não pode ser observada diretamente, mas com um planejamento de experimentos é possível obter informação sobre as diferenças entre o par Best-Worst. FINN e LOUVIERE (1993) mostraram que uma combinação da frequência de “Best” e de “Worst” de cada item pode ser utilizada para medir o grau de importância dos atributos.

Um resultado hipotético para o exemplo anterior, com 3 atributos avaliados por 50 indivíduos, é dado na Tabela 2.1.

Tabela 2.1 – Frequência de resposta “best” e “worst” para cada atributo.

Atributo	Frequência “best”	Frequência “worst”
Preço	20	15
Nome da marca	10	24
Qualidade	20	11
Total	50	50

Neste exemplo, a importância do atributo pode ser dada pela diferença entre as frequências “best” e “worst”. O atributo “preço”, por exemplo, tem importância $20-15=5$. Para facilitar a interpretação, pode-se colocar a importância w_i no intervalo entre 0 e 1 da seguinte maneira:

$$w_i = \frac{\text{frequência} \text{"best"} - \text{frequência} \text{"worst"} + N}{2N}, \quad (2.1)$$

em que N é o número de respondentes.

Tem-se então:

$$w_{\text{preço}} = 0,55;$$

$$w_{\text{marca}} = 0,36;$$

$$w_{\text{qualidade}} = 0,59.$$

Quando há muitos atributos, o número de pares que são criados torna difícil a escolha do par Best-Worst. É comum, nestes casos, a criação de subconjuntos de itens com apenas alguns atributos selecionados segundo um experimento fatorial fracionado (WINER, 1991). Por exemplo, com $I=10$ itens, há 210 combinações com 4 itens. Pode-se escolher tantos conjuntos fracionados ortogonais, a serem avaliados por cada respondente. Embora esta alternativa diminua o número de comparações, ela não é eficaz quando o número de itens é muito grande. Quando há, por exemplo, 50 atributos para analisar, o número de comparações é grande mesmo quando se utiliza um desenho fatorial fracionado e a tarefa continua cansativa para o respondente.

2.3.2.3. Ordem de posto (*ranking*) e ordenação seletiva

Neste caso, os entrevistados devem ordenar os atributos do mais importante ao menos importante. São atribuídos postos aos atributos de acordo com a ordem de escolha. Esta técnica tem como desvantagem a geração de dados ordinais, limitando as análises possíveis.

Quando o número de atributos é muito grande, a tarefa de ordenação torna-se complicada para maioria dos respondentes. Segundo KOTLER (1992), o tamanho mediano do número de alternativas que um consumidor consegue analisar é cinco. Neste caso, pede-se que o respondente ordene apenas os k atributos mais importantes, sendo k um número próximo de cinco.

CARVALHO e LEITE (1998) aplicaram um método para obtenção da importância dos atributos em que o respondente ordena, por ordem decrescente de importância, os $k=6$ atributos mais importantes dentre os $I=39$ apresentados no questionário. A seguir, levando-se em conta os dados gerados por todos os respondentes, foi computada, para cada atributo, a frequência com que ele aparece como o mais importante, o segundo mais importante, e assim por diante. Essa frequência foi multiplicada por um peso, atribuído da seguinte forma: 6 para o item mais importante, 5 para o segundo mais importante, 4 para o terceiro e assim em diante até o 6º item mais importante, que recebeu peso 1. A Tabela 2.2 mostra o exemplo, extraído do artigo, de um dos atributos:

Tabela 2.2: obtenção do valor ponderado dos atributos pela importância

Atributo 1	Frequência	Peso	Valor ponderado
Mais importante	112	6	672
2º mais importante	51	5	255
3º mais importante	25	4	100
4º mais importante	26	3	78
5º mais importante	23	2	46
6º mais importante	27	1	27
total	264		1178

Fonte: CARVALHO e LEITE, 1998.

Depois de calculado o escore ponderado para cada atributo, obtém-se a importância pela divisão do escore ponderado do atributo pela soma dos escores ponderados de todos os atributos (11025 neste exemplo):

$$\bar{w}_i^* = \frac{\sum_{r=1}^6 f_{ir} p_r}{\sum_{i=1}^I \sum_{r=1}^6 f_{ir} p_r} \quad (2.2)$$

Em que:

I = número de atributos;

r é o ranking dado pelo respondente ao atributo;

\bar{w}_i^* = importância estimada do atributo i ;

f_{ir} = frequência do atributo i como o r -ésimo mais importante;

p_r = peso associado ao r -ésimo atributo mais importante.

Neste exemplo, o atributo 1 tem importância igual a 10,68% (1178/11025). Esta importância é a mesma para todos os indivíduos, ou seja, a matriz $\mathbf{W} = (\bar{w}_i^*)$ para todo i, j .

O escore de satisfação global para o indivíduo j é dado pelo seguinte índice:

$$U_j = \sum_{i=1}^I \bar{w}_i^* x_{ij}, \quad (2.3)$$

em que x_{ij} é a resposta dada pelo indivíduo j ao atributo i .

Para ilustrar esta metodologia, é mostrado a seguir um exemplo com importâncias (\bar{w}_i^*) e avaliações (x_{ij} , em uma escala de 0 a 10) de 3 atributos para 4 indivíduos. As matrizes \mathbf{W} e \mathbf{X} são:

$$\mathbf{W} = \begin{bmatrix} 0,25 & 0,25 & 0,25 & 0,25 \\ 0,40 & 0,40 & 0,40 & 0,40 \\ 0,35 & 0,35 & 0,35 & 0,35 \end{bmatrix} \quad \text{e} \quad \mathbf{X} = \begin{bmatrix} 6 & 7 & 8 & 7 \\ 5 & 4 & 3 & 5 \\ 3 & 7 & 5 & 4 \end{bmatrix}$$

O vetor de escores globais é dado por $\mathbf{U} = [4,65 \quad 6,80 \quad 4,95 \quad 5,35]$

Esta metodologia tem a desvantagem de considerar os dados intervalares (para calcular a soma e a razão) quando na verdade os dados são ordinais e a escala escolhida para o peso associado ao r -ésimo atributo mais importante é arbitrária.

2.3.2.4. Soma constante

Nesta abordagem, pede-se ao entrevistado para distribuir 100 pontos entre todos os atributos de acordo com a sua importância. Esta abordagem foi utilizada em uma versão recente do modelo de mensuração de satisfação em Qualidade de Serviços - SERVQUAL (PARASURAMAN *et al*, 1985).

Neste modelo, os autores sugerem que a qualidade percebida de um serviço está relacionada com a diferença entre a expectativa e a performance percebida (*gap*). Se houver diferença entre a expectativa e a performance, há desconformidade de expectativas. Cinco dimensões são consideradas para formar o escore de satisfação global: tangibilidade, prontidão, empatia, garantia e confiabilidade.

Dependendo do serviço avaliado, outras dimensões podem ser consideradas. Os autores constroem uma escala global de satisfação somando todas as desconformidades (performance menos expectativas) de cada dimensão, atribuindo o mesmo peso a todas elas. A desconformidade média de cada dimensão também é calculada pela soma da desconformidade de cada item.

Os autores sugerem que o escore global pode ser calculado ponderando-se cada dimensão pela sua importância. Neste caso, é perguntado ao entrevistado qual é a importância de cada uma das cinco dimensões consideradas. ZEITHMAN *et al*. (1990) sugere que o respondente distribua 100 pontos entre as cinco dimensões determinantes da qualidade do serviço de tal forma a refletir a importância percebida da dimensão. Os atributos dentro de cada uma das dimensões têm o mesmo peso.

CARMAN (1990) sugere que o respondente dê peso para cada um dos atributos dentro de cada dimensão, mas observa que essa tarefa fica difícil quando há muitos atributos envolvidos na análise. Não se espera que uma pessoa consiga distribuir peso entre mais de 10 atributos.

Em relação à técnica de ordenação por postos, a soma constante tem a vantagem de trabalhar com uma escala intervalar.

2.3.2.5. Pontos decrescentes

Esta técnica, descrita por GHISI, MERLO e NAGANO (2004) é feita em dois estágios. No primeiro estágio, o respondente ordena os k atributos mais importantes, similar à proposta de CARVALHO e LEITE (1996). Em seguida, o respondente atribui notas decrescentes, variando de 0 a 100, aos atributos escolhidos respeitando a ordem de escolha. O atributo mais importante recebe nota 100. Em seguida, as notas atribuídas por todos os respondentes a cada um dos atributos são somadas e, por fim, divide-se a pontuação de cada atributo pela somatória da pontuação de todos os atributos, obtendo-se a importância de cada atributo.

A vantagem deste método é que a importância é medida em uma escala intervalar e os atributos são medidos tanto por uma abordagem qualitativa (primeiro estágio) quanto quantitativa (segundo passo).

2.4. TÉCNICAS INDIRETAS DE OBTENÇÃO DE IMPORTÂNCIA

A importância dos atributos também pode ser mensurada de maneira indireta. Pode-se, por exemplo, perguntar ao entrevistado qual é a satisfação global com determinado produto e qual é a sua satisfação individual com cada atributo do produto. Por meio de uma regressão múltipla obtêm-se os coeficientes relacionados a cada um dos atributos. Este coeficiente pode ser visto como a importância média daquele atributo para a avaliação geral. Na formulação da metodologia que define a divisão de classes sócio-econômicas no Brasil (Critério Brasil) com base na posse de alguns bens, utilizou-se a regressão múltipla para obter o peso (importância) de cada bem na classificação do indivíduo.

Deve-se ressaltar que, no caso da regressão múltipla, a análise da importância é agregada e não é possível segmentar os respondentes com base na importância individual dada aos atributos.

A técnica de análise fatorial também pode ser empregada para obtenção da importância. CUNHA (1998) propõe a utilização dos autovalores da análise fatorial como peso para cada dimensão da satisfação do consumidor. O escore de satisfação do indivíduo com a dimensão d é obtido a partir das variáveis com alta carga no fator. A desvantagem deste método é que o

autovalor é influenciado pelo número de itens relacionados a uma determinada dimensão do serviço. Um grande número de itens não significa que esta dimensão seja a mais importante.

A técnica de Análise Conjunta (Conjoint Analysis) permite que as importâncias dos atributos sejam estimadas de forma indireta e desagregada. O objetivo desta técnica é entender como o consumidor ou respondente reage a combinações de atributos de um produto ou serviço (GREEN, 1978). A importância de um atributo ou item é determinada pela diferença na utilidade obtida entre os níveis extremos do atributo. Exemplificando, a importância do preço do produto é obtida pela diferença na utilidade de um produto que custa \$100 (maior preço disponível) e \$10 (menor preço disponível). Este método possui a desvantagem da importância depender dos níveis extremos do atributo. Quanto maior for a diferença entre, digamos, o preço mais alto apresentado e o preço mais baixo, maior será a importância do atributo “preço” para o indivíduo.

A utilidade, por sua vez, pode ser obtida de várias maneiras. Pode-se solicitar ao respondente que ordene os estímulos (produto que combina níveis dos atributos estudados) do mais preferível ao menos preferível. Supondo-se um modelo aditivo, a utilidade pode ser obtida pela diferença do *ranking* médio de cada nível do atributo (JOHNSON, 2002). Esta análise permite que se obtenha importância individual para cada atributo.

Também é possível utilizar uma escala não comparativa, como a escala de Likert, para mensurar a utilidade de cada estímulo apresentado. Novamente a importância é obtida pela diferença na utilidade, que é obtida pela média da escala nos níveis extremos do atributo.

Um dos problemas da Análise Conjunta é não ser possível considerar um grande número de atributos, pois o número de estímulos formados aumenta exponencialmente. Este problema pode ser resolvido de duas formas: fazendo um experimento fatorial fracionado (em que apenas parte dos estímulos é apresentado ao respondente) ou utilizando a técnica de Análise Conjunta Adaptativa (Adaptative Conjoint Analysis, ACA) (JOHNSON, 2001). Esta técnica permite que seja apresentado ao usuário um estímulo com poucos atributos. Conforme se verifica que determinado atributo não é importante, ele deixa de ser relevante para a análise e outros atributos entram em seu lugar. A estimativa da importância de cada atributo é atualizada a cada comparação feita pelo respondente. Este tipo de técnica só pode ser feito via computador, pois a seqüência de estímulos apresentados ao respondente depende de sua

resposta anterior. Costuma-se perguntar a importância de cada atributo diretamente antes de iniciar o processo de comparação da ACA. Estas estimativas iniciais servem como base para a escolha dos atributos a serem incluídos nas comparações seguintes.

2.5. ESTIMAÇÃO DA IMPORTÂNCIA

O resultado obtido direta ou indiretamente fornece uma estimativa da importância de cada atributo. Esta importância pode ser estimada individualmente (w_{ij}) ou de forma agregada, como se todos os respondentes atribuísem a mesma importância para cada item ($w_{ij} = w_i$). O Quadro 2.2 mostra como obter estimativas da importância para cada método utilizado.

	Método	Obtenção da importância individual (desagregada) w_{ij}	Obtenção da importância coletiva (agregada) w_i
Métodos diretos	Likert/ Diferencial semântico	diretamente da escala	calcula-se a média do valor obtido entre os indivíduos para cada atributo
	Comparação de pares	frequência relativa do número de vezes que o atributo é escolhido	frequência relativa do número de vezes que o atributo é escolhido por todos os indivíduos
	Best / Worst	Diferença na frequência que o atributo é escolhido como “best” e “worst” pelo indivíduo	Diferença na frequência que o atributo é escolhido como “best” e “worst” por todos os indivíduos
	Ordenação de postos	posto do atributo dado pelo indivíduo	média dos postos do atributo
	Ordenação seletiva	posto do atributo dado pelo indivíduo	média dos postos do atributo
	Soma constante	ponto dado ao atributo	média de pontos dado ao atributo
	Pontos decrescentes	ponto dado ao atributo	média de pontos dado ao atributo
Métodos indiretos	Análise de Regressão	-	coeficientes da regressão
	Análise Fatorial	-	Autovalor do fator
	Análise Conjunta	diferença entre utilidade dos níveis extremos dos atributos	diferença média entre utilidade dos níveis extremos dos atributos

Quadro 2.2 – Comparação entre métodos de obtenção de importância

A revisão teórica mostra que todos os métodos apresentados possuem alguma desvantagem. Dentre as desvantagens, destacam-se os erros de mensuração, o esforço mental necessário para responder quando há muitos atributos envolvidos na análise e a possibilidade de considerar diferenças individuais para obtenção da importância dos atributos (análise desagregada), principalmente quando há valores faltantes na matriz \mathbf{W} .

No capítulo seguinte é proposto um modelo que permite modelar importâncias diferentes para cada item de acordo com o valor do indivíduo em um traço latente, que define seu comportamento em relação à importância dos itens. Esta modelagem é possível mesmo quando o indivíduo avalia a importância de um subconjunto do total de itens, o que permite criar questionários que exijam menos esforço mental do entrevistado, já que ele não precisa avaliar todos os atributos.

3 MODELOS DA TRI E APLICAÇÃO DO MODELO DE DESDOBRAMENTO GRADUADO GENERALIZADO (GGUM) PARA MENSURAÇÃO DE IMPORTÂNCIA

3.1. INTRODUÇÃO

A Teoria de Resposta ao Item (TRI) surgiu há mais de 60 anos e desde então tem sido muito utilizada na área de psicologia e educação. Os modelos da TRI modelam a probabilidade de um indivíduo dar determinada resposta a um item em função dos parâmetros deste item e de uma característica individual (traço latente) do respondente.

Na área de educação esse traço é chamado de “habilidade” ou “proficiência”; quanto maior a habilidade, maior a probabilidade de uma resposta “correta” ao item.

A TRI também tem sido muito utilizada na área de administração. Pesquisa rápida em um *site* de busca nessa área indicou mais de 5000 citações (JSTOR, 2005) com diversas aplicações, como seleção de pessoal, marketing, medidas de qualidade de vida, avaliação de comportamento, certificação e outras áreas relacionadas à administração. Muitas destas aplicações podem ser consultadas em SIJTSMA e MOLENAAR(2002).

Em marketing especificamente, os modelos da TRI têm sido utilizados em problemas em que se deseja construir uma escala de atitudes. Pode-se, por exemplo, estimar a satisfação do cliente com um determinado serviço ou produto.

No Brasil, a TRI começou a ser utilizada para analisar dados do Sistema Nacional de Ensino Básico – SAEB e no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP, ambos no final da década de 90 (ANDRADE e KLEIN, 1999).

Em estudo realizado por COSTA (2001) a TRI foi utilizada para medir o grau de satisfação do consumidor em restaurantes na cidade de Curitiba. Vários itens de serviço foram mensurados em que o entrevistado atribuiu um grau de satisfação com determinado item de serviço. A

partir da análise conjunta das respostas destes itens pela TRI é possível estimar o grau de satisfação do usuário com o serviço em geral. Mais especificamente, é possível estimar a probabilidade de um indivíduo concordar com determinado item do serviço com base tanto no seu parâmetro de satisfação quanto na “dificuldade” de concordância com o item.

Uma das vantagens da TRI é que ela permite comparação entre populações, desde que submetidas a provas ou questionários com alguns itens comuns. Aplicado ao marketing, isto significa que é possível criar vários questionários com itens (perguntas, atributos ou afirmações) comuns e fazer questionários individuais de menor tamanho, obtendo-se a partir da resposta dos entrevistados uma resposta estimada aos itens não respondidos, com pequena margem de erro.

Os modelos de TRI propostos na literatura dependem basicamente dos seguintes fatores (ANDRADE et al., 2000):

- a) da natureza do item – dicotômico ou não;
- b) do número de populações envolvidas;
- c) da quantidade de traços latentes que está sendo medida;
- d) do modelo ser linear ou não.

Outro fator importante que diferencia os modelos de TRI é o seu caráter cumulativo. Em um modelo cumulativo, a probabilidade de uma determinada resposta “correta” aumenta conforme o valor no traço latente aumenta. No modelo não cumulativo (modelo de desdobramento), a probabilidade de resposta “correta” aumenta conforme o valor do indivíduo no traço latente está próximo do valor do atributo.

Neste capítulo é feita uma revisão dos principais modelos cumulativos da TRI, principalmente o Modelo de Créditos Parciais Generalizado (MURAKI, 1992) pois este modelo serve de base para a construção do Modelo de Desdobramento Graduado Generalizado, apresentado em seguida. Este último, por ser o modelo proposto nesta tese para modelar a importância de atributos, é discutido mais detalhadamente neste capítulo.

3.2. MODELOS DA TRI CUMULATIVOS

3.2.1. Principais modelos unidimensionais para uma população

A apresentação dos modelos será dividida de acordo com a natureza do item – dicotômico ou não. No caso de não ser dicotômico, será primeiramente apresentado o modelo em que se supõe que as categorias de resposta são nominais e, em seguida, admitindo que as categorias são ordinais.

3.2.1.1. Modelos para itens dicotômicos ou de múltipla escolha corrigidos como certo/errado

Os modelos para itens dicotômicos unidimensionais para uma população foram os primeiros a serem desenvolvidos, na década de 50. Inicialmente, os modelos utilizavam a função ogiva normal, mas depois estes modelos foram substituídos pela função logística, matematicamente mais conveniente. De todos os modelos de TRI, o mais utilizado é o logístico de três parâmetros, que considera:

- a) a dificuldade do item;
- b) a discriminação do item;
- c) a probabilidade de resposta correta por indivíduos de baixa habilidade (“chute”).

Em um teste de múltipla escolha com cinco alternativas, esta probabilidade deveria ser igual a 0,20.

A probabilidade de um indivíduo j responder corretamente ao item i dada a sua habilidade θ_j é

$$P(X_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad i = 1, 2, \dots, I; \quad j = 1, 2, \dots, n \quad (3.1)$$

Em que:

X_{ij} é uma variável que assume o valor 1 se o respondente j responder corretamente à questão i e assume o valor 0 caso contrário;

θ_j é a habilidade do indivíduo j . Quanto maior for a habilidade do indivíduo, maior é a probabilidade de acerto da questão;

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

b_i é o parâmetro que define a habilidade necessária do indivíduo para que a probabilidade de resposta correta seja 50%, isto é, quando $b_i=\theta_j$, $P(X_{ij}=1|\theta_j)=0,5$.

a_i é o parâmetro de discriminação do item i . Um item com alto valor de discriminação tem grande mudança na probabilidade de acerto para pequenas variações de habilidade em determinada região do traço latente, mais especificamente em torno do valor de b_i ;

c_i é o parâmetro que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente ao item i ;

D é um fator de escala, constante e geralmente igual a 1. Se $D=1,7$ a função logística fornecerá resultados semelhantes ao da função ogiva normal;

A representação da função logística é dada no Gráfico 3.1. Esta curva é chamada de Curva Característica do Item (CCI).

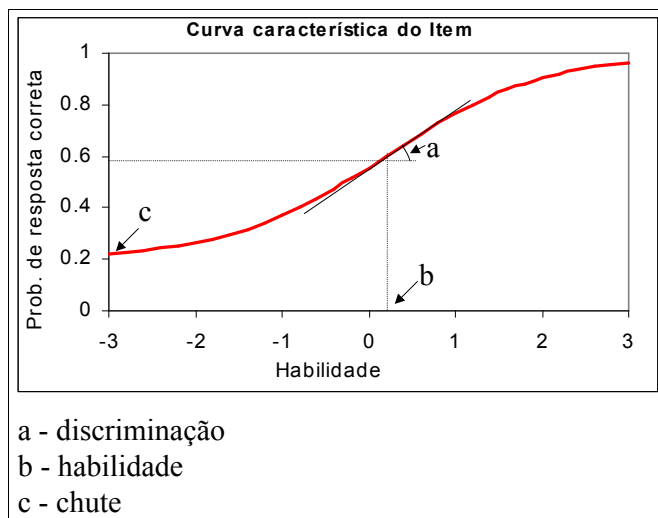


Gráfico 3.1 – Curva Característica de um item em um modelo de TRI com 3 parâmetros.

A escala de habilidade é uma escala arbitrária que apenas ordena os indivíduos do menos hábil para o mais hábil no traço latente estudado. Esta escala arbitrária pode ser definida de tal forma que a habilidade da maioria dos indivíduos assuma valores entre -3 e 3, com média zero. Para isto, basta fixar uma distribuição de probabilidades para a habilidade com média zero e desvio-padrão igual a 1.

Em alguns casos, podem ser considerados modelos mais simples. O modelo de 2 parâmetros considera que $c=0$, isto é, a probabilidade de acerto ao acaso é zero. Por exemplo, em situações com itens abertos, corrigidos como certo/errado. O modelo de 1 parâmetro, conhecido como modelo de Rasch, supõe que o parâmetro c é zero e que todos os itens tem o mesmo poder de discriminação ($a_i = a$, $i=1, \dots, I$).

3.2.1.2. Modelos para itens não dicotômicos

Na literatura há vários modelos para itens não dicotômicos. Pode-se analisar itens abertos ou itens de múltipla escolha em que a análise é feita de forma graduada (não há apenas certo e errado).

A diferença entre os principais modelos existentes está relacionada ao número de parâmetros utilizados e à hipótese das respostas serem ordinais ou nominais, isto é, se há ou não uma ordenação das categorias de resposta.

O modelo desenvolvido por BOCK (1972) de resposta nominal assume que não há nenhuma ordenação a priori das opções de resposta. Com este modelo, é possível calcular a probabilidade de resposta para cada uma das k categorias existentes. Mais detalhes sobre este modelo podem ser encontrados em AZEVEDO (2003).

O modelo de resposta gradual de SAMEJIMA (1969) assume que as categorias de resposta de um item podem ser ordenadas entre si. Um caso particular deste modelo é o modelo de escala gradual, em que se supõe que os escores das categorias são igualmente espaçados (ANDRICH, 1978).

O modelo de crédito parcial, desenvolvido por MASTERS (1982) difere do gradual pois pertence à família dos modelos de Rasch (com apenas um parâmetro). Este modelo foi aplicado pelo PISA (*Programme for International Student Assessment*), cuja principal finalidade é produzir indicadores sobre a efetividade dos sistemas educacionais, avaliando o desempenho de alunos na faixa dos 15 anos. No Brasil, o PISA é coordenado pelo Inep – Instituto Nacional de Estudos e Pesquisas Educacionais “Anísio Teixeira”. Mais informações sobre a metodologia aplicada a este estudo podem ser obtidas em ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2000).

MURAKI (1992) desenvolveu um modelo de crédito parcial generalizado, em que a suposição de igualdade do poder de discriminação dos itens foi relaxada. A probabilidade de escolha da k -ésima categoria do item i , pelo indivíduo j , é dada por:

$$P_{i,k}(\theta_j) = \frac{\exp \left[\sum_{u=0}^k Da_i(\theta_j - b_{i,u}) \right]}{\sum_{u=0}^{m_i} \exp \left[\sum_{v=0}^k Da_i(\theta_j - b_{i,v}) \right]}, \quad i = 1, 2, \dots, I; j = 1, 2, \dots, n; k = 0, 1, \dots, m_i \quad (3.2)$$

Em que:

$m_i + 1$ é o número de categorias de resposta do item i (este número pode variar para cada item);

a_i é o parâmetro de discriminação do item i . Um item com alto valor de discriminação tem grande mudança na probabilidade de acerto para pequenas variações de habilidade;

D é um fator de escala, conforme definido nos modelos anteriormente apresentados;

$b_{i,k}$ é o ponto do *continuum* latente em que as curvas de $P_{i,k-1}(\theta_j)$ e $P_{i,k}(\theta_j)$ se interceptam, isto é, se $b_{i,k} = \theta_j$, $P_{i,k-1}(\theta_j) = P_{i,k}(\theta_j)$. Em geral, $b_{i,0} = 0$.

O Gráfico 3.2 ilustra as curvas de resposta para o caso em que o item possui 3 categorias de resposta, $a_i = 1$, $b_{i,1} = -1$ e $b_{i,2} = 1$.

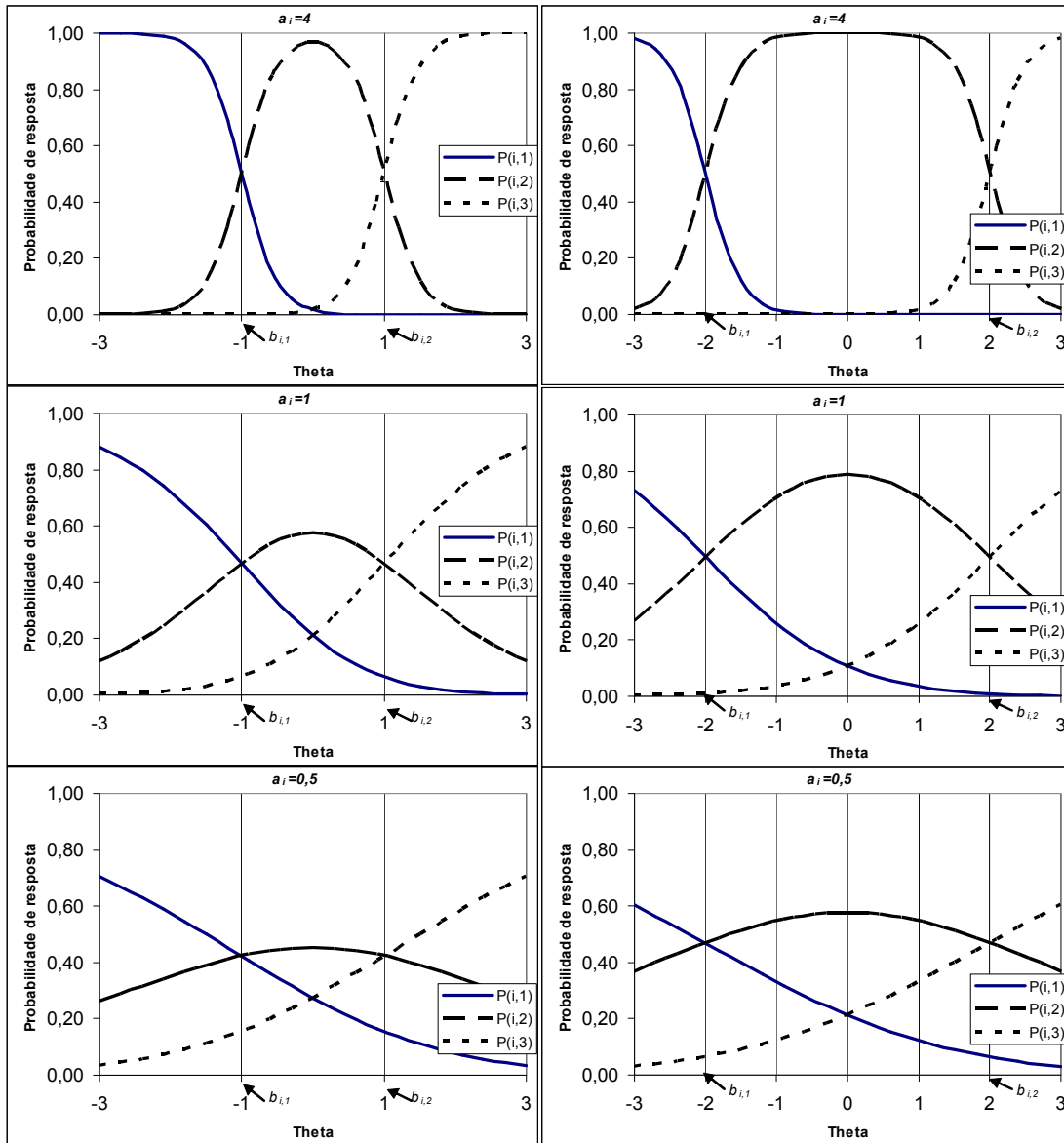


Figura 3.1 – Probabilidade de resposta do modelo de créditos parciais com 3 categorias, para alguns valores de a_i e $b_{i,k}$.

Detalhes dos modelos descritos podem ser encontrados em ANDRADE et al. (2000) e VALLE (1999).

3.2.2. Função de Informação do Item

A Função de Informação do Item permite analisar quanto o item contém de informação para estimar a habilidade do indivíduo. Para o modelo dicotômico, esta função é dada por:

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta) \right]^2}{P_i(\theta)(1 - P_i(\theta))} \quad (3.3)$$

Em que:

$I_i(\theta)$ é uma medida de informação pelo item i para a habilidade θ e

$$P_i(\theta) = P(X_{ij}=1|\theta) .$$

No modelo logístico de 3 parâmetros, a função de informação é dada por (BIRNBAUM, 1968)

$$I_i(\theta) = D^2 a_i^2 \frac{1 - P_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (3.4)$$

Esta equação permite relacionar a função de informação com os parâmetros do item. Uma análise da equação mostra que a informação é maior quando:

- b_i se aproxima de θ ;
- quanto maior for a_i ;
- quanto mais c_i se aproximar de zero.

Também pode ser definida a Função de Informação do Teste para θ , que é dada pela soma das informações de cada item:

$$I(\theta) = \sum_{i=1}^I I_i(\theta) \quad (3.5)$$

Pode-se também representar a função de informação do teste pelo erro-padrão de estimação, que é dado por

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (3.6)$$

Para ilustrar a função de informação de um item, será utilizado o modelo logístico com 3 parâmetros. A curva característica e a função de informação para quatro itens com parâmetros a e b distintos, variando-se a habilidade, são ilustradas na Figura 3.2.

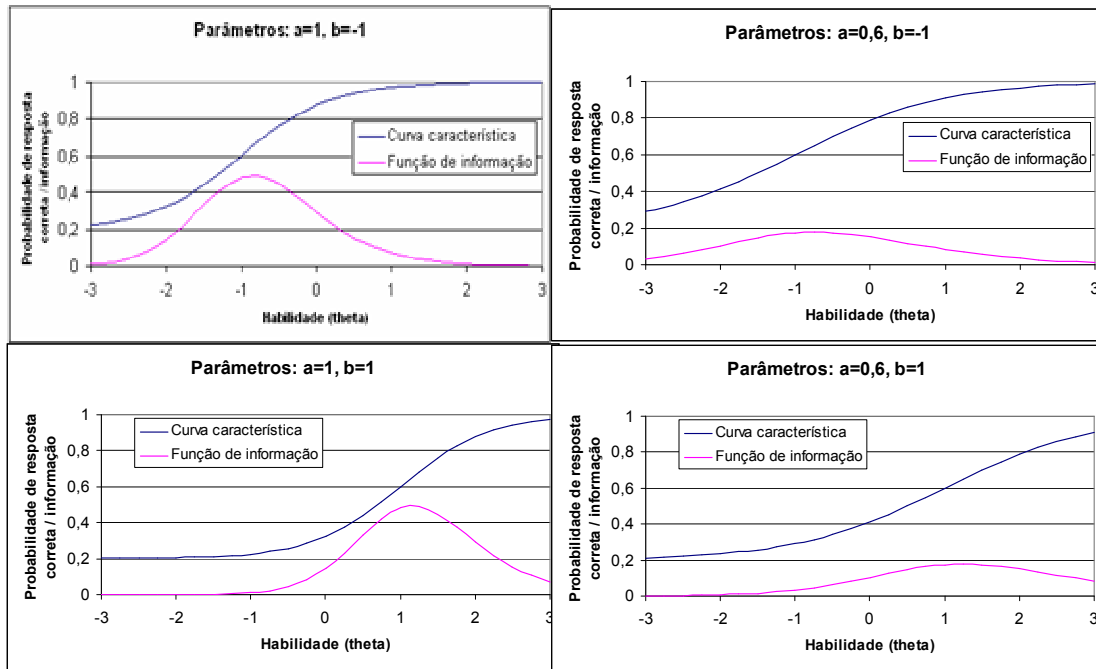


Figura 3.2 – Curva característica e função de informação para quatro itens.

Nos modelos da TRI cumulativos isto implica que, para avaliar indivíduos com grande habilidade, deve ser utilizado um item com b_i alto. Além disto, itens com a_i alto possuem maior informação, isto é, discriminam indivíduos na medida em que a diferença de probabilidade de resposta correta entre dois indivíduos com θ 's próximos a b_i é grande.

A função de informação apresentada anteriormente para o modelo dicotômico é um caso particular de uma expressão genérica para m categorias de resposta (SAMEJIMA, 1969). A informação de um item i é dada pela soma das informações associadas a cada categoria de resposta, ponderada pela probabilidade de escolha da categoria. Portanto, a função de informação do item é dada por:

$$I_i(\theta) = \sum_{k=1}^{m_i} I_{ik}(\theta) P_{ik}(\theta) \quad (3.7)$$

em que:

$I_{ik}(\theta)$ é uma medida de informação da k -ésima categoria de resposta do item i .

$P_{ik}(\theta)$ é a probabilidade da k -ésima categoria de resposta do item i ser escolhida.

A quantidade $I_{ik}(\theta) P_{ik}(\theta)$ é o quanto a k -ésima categoria de resposta contribui para a informação do item, e é dada por

$$I_{ik}(\theta) P_{ik}(\theta) = \frac{\left[\frac{\partial P_{ik}(\theta)}{\partial \theta} \right]^2}{P_{ik}(\theta)} - \frac{\partial^2 P_{ik}(\theta)}{\partial^2 \theta} \quad (3.8)$$

3.2.3. Suposições do modelo

Os modelos da TRI apresentados têm duas suposições básicas: unidimensionalidade e independência local.

A unidimensionalidade significa que o conjunto de itens mede apenas um traço latente, isto é, existe uma habilidade dominante, representada em uma dimensão, que está relacionada com a resposta dada aos itens.

Esta suposição pode ser verificada por meio de uma análise fatorial utilizando correlação tetracórica no caso de variáveis dicotômicas e correlação policórica no caso de variáveis politômicas (Likert, por exemplo). Um outro procedimento utilizado para analisar dados de resposta categórica a itens é a “full information item factor analysis” (BOCK & AITKIN, 1981; BOCK, GIBBONS & MURAKI, 1988), que pode ser conduzida com o software TESTFACT. Está implícito neste modelo que os itens medem uma ou mais variáveis latentes

contínuas, e supõe-se que a resposta atribuída pelo indivíduo ao item tem uma distribuição normal, dividida por limiares que determinam as categorias de resposta.

RECKASE (1979) argumenta que se o primeiro autovalor é muito superior aos outros e o primeiro fator responder por pelo menos 20% da variância total, então a unidimensionalidade pode ser considerada.

A suposição de independência local postula que, dada a habilidade do indivíduo, as respostas dos itens são independentes entre si. A resposta do indivíduo a um item não causa mudança valor do parâmetro do indivíduo, isto é, o indivíduo não “aprende” com o teste. Esta suposição é fundamental para a obtenção da probabilidade do vetor de respostas do indivíduo, dada pelo produtório das probabilidades de resposta a cada item. HAMBLETON e SWAMINATHAN (1991) mostraram que a unidimensionalidade implica independência local. Portanto, há apenas uma suposição a ser verificada.

3.2.4. Estimação dos parâmetros

Tanto a habilidade do indivíduo quanto os parâmetros de cada item, em geral, são desconhecidos. Na área de educação, um banco de itens já “calibrado”, isto é, com seus parâmetros já estimados, pode ser utilizado para estimar a habilidade de um novo indivíduo. Neste caso, apenas as habilidades são estimadas.

Será apresentado aqui o processo de estimação conjunta dos parâmetros dos indivíduos e dos itens para o modelo logístico unidimensional de três parâmetros para itens dicotômicos. Os outros modelos têm processo de estimação similar.

Se uma pesquisa com I itens é aplicada a um grupo de n indivíduos, há $3 \cdot I + N$ parâmetros a serem estimados. Um problema dos modelos da TRI é que o número de parâmetros cresce com o número de indivíduos na amostra. Dentre os vários métodos de estimação, o primeiro a ser utilizado foi o método de máxima verossimilhança conjunta.

Supondo independência local, a probabilidade do vetor $\mathbf{x}_j = (x_{1j}, \dots, x_{Ij})$, dada a habilidade θ_j e os parâmetros dos itens, é dada pelo produtório das probabilidades de resposta do indivíduo a cada item, como segue:

$$P_j(\mathbf{x}_j | \theta_j, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \prod_{i=1}^I P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}, \quad (3.9)$$

em que:

$X_{ij} = 1$ se a resposta do indivíduo j ao item i está correta e 0 caso contrário;

$P_{ij} = P(X_{ij}=1 | \theta_j)$ e

$Q_{ij} = 1 - P_{ij}$.

Admitindo independência de resposta entre os indivíduos, a função de verossimilhança L , baseada nas respostas dos n indivíduos, pode ser escrita como o produtório das probabilidades de resposta de cada indivíduo:

$$L(\mathbf{X} | \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \prod_{j=1}^n P_j(\mathbf{x}_j | \theta_j, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \prod_{j=1}^n \prod_{i=1}^I P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}} \quad (3.10)$$

Para obter as estimativas dos parâmetros, deriva-se e iguala-se a zero o logaritmo natural da função de verossimilhança.

Em razão do grande número de parâmetros, o problema computacional é grande. Uma saída é utilizar um procedimento em dois estágios, em que primeiro os parâmetros dos itens são estimados, assumindo-se que a habilidade dos indivíduos é conhecida. Inicialmente, utiliza-se o escore padronizado (número de acertos padronizado) de cada um dos respondentes como estimativa de sua habilidade.

No segundo estágio, as estimativas dos parâmetros dos itens, obtidas no estágio anterior, são consideradas como parâmetros conhecidos e estima-se a habilidade com base nestes valores. Forma-se um processo iterativo em que os parâmetros dos itens e da habilidade são estimados passo a passo. Quando a mudança de um estágio para outro no valor do logaritmo da verossimilhança for menor que um valor, digamos 0,00001, o processo termina.

Há um problema de identificação neste modelo, que pode ser resolvido definindo-se uma escala arbitrária para os valores das habilidades ou para os valores da dificuldade (parâmetro

b), pois eles são medidos na mesma escala. Usualmente, impõe-se que as habilidades tenham média zero e desvio-padrão 1, solucionando o problema de identificação.

Atualmente, outros métodos têm sido aplicados neste tipo de problema. Dentre estes métodos, destacam-se a Máxima Verossimilhança Marginal (BOCK e LIEBERMAN, 1970) e a estimação através de métodos Bayesianos (BAKER e KIM, 2004).

No método da Máxima Verossimilhança Marginal, primeiramente são estimados os parâmetros dos itens e, posteriormente, as habilidades dos indivíduos. A vantagem deste método é que as estimativas de máxima verossimilhança obtidas não são inconsistentes (ANDERSEN, 1980), como no método de máxima verossimilhança conjunta. Para estimar os parâmetros dos itens sem conhecer as habilidades dos indivíduos, a verossimilhança não pode depender destes parâmetros. Isto pode ser feito marginalizando a verossimilhança integrando-a em relação a θ .

Para calcular esta integral, ANDERSEN (1980) argumenta que é necessário especificar uma distribuição para θ . Considerando uma população composta por n indivíduos com habilidades θ_j , a distribuição acumulada $G(\theta)$ pode ser definida como: número de j | $\theta_j \leq \theta$ dividido por n . Para n suficientemente grande, $G(\theta)$ pode ser aproximada por uma distribuição contínua, que tem uma densidade $g(\theta)$ correspondente.

A integral resultante na função de verossimilhança, embora não apresente solução analítica, pode ser aproximada de várias maneiras. O método mais utilizado na TRI é o Hermite-Gauss, também denominado método de quadratura gaussiana. Esse método consiste em aproximar a integral por uma soma das áreas de um número finito q de retângulos. Os pontos médios de cada retângulo, $\bar{\theta}_k, k=1, \dots, q$ são denominados de pontos de quadratura. Cada ponto tem um peso no cálculo da integral de acordo com a altura $g(\bar{\theta}_k)$, que é calculada pela função definida $g(\theta)$ e o número de nós (HILDEBRAND, 1956). Para $g(\theta) \sim N(\mu; \sigma^2)$, valores calculados podem ser encontrados em STROUD e SECHEST (1966).

As estimativas de máxima verossimilhança podem ser obtidas pelo algoritmo EM (DEMPSTER, LAIRD & RUBIN, 1977), que é um processo iterativo feito em dois passos: esperança (E) e maximização (M). No primeiro passo, o objetivo é calcular a esperança do

logaritmo da função densidade conjunta dos dados dada uma estimativa dos parâmetros desconhecidos. O segundo passo é achar uma nova estimativa que maximize a função do primeiro passo. Esta maximização pode ser feita, por exemplo, pelo algoritmo de Newton-Raphson (ISAAC e KELLER, 1966) ou “Scoring” de Fisher (RAO, 1973).

O método de máxima verossimilhança pode apresentar problemas quando um item é respondido corretamente ou incorretamente por todos os indivíduos ou quando um indivíduo responde corretamente ou incorretamente a todos os itens. Neste caso, pode ser utilizado um método Bayesiano para estimar os parâmetros. O mais utilizado é a Estimação Bayesiana Marginal (MISLEVY, 1986), em que se estabelece uma distribuição *a priori* para os parâmetros de interesse e, a partir dos dados observados, chega-se a uma distribuição *a posteriori*. O estimador do parâmetro de interesse é dado por alguma característica desta distribuição, como a média ou a moda.

Maiores detalhes sobre métodos de estimação em modelos da TRI podem ser encontrados em BAKER e KIM (2004) e AZEVEDO (2003).

3.3. MODELO DE DESDOBRAMENTO GRADUADO GENERALIZADO (GGUM)

3.3.1. Introdução

O Modelo de Desdobramento Graduado Generalizado (*Generalized Graded Unfolding Model*, GGUM) pertence à família de modelos da Teoria de Resposta ao Item chamada de “modelos de desdobramento”.

A noção de “desdobramento” já existia antes dos modelos da TRI. O primeiro modelo de desdobramento (também chamado de modelo de ponto ideal) foi desenvolvido por COOMBS (1950) e aplicado a dados de preferência. Neste modelo, o ponto ideal do indivíduo está próximo dos objetos para os quais ele tem preferência. Seja $y=(y_1, \dots, y_A)$ um vetor de coordenadas no espaço Euclidiano que representa a posição do indivíduo (ponto ideal) e

$\mathbf{x}_k = (x_{k1}, \dots, x_{kA})$ o vetor de coordenadas que representa a posição do objeto i neste espaço de A dimensões. A distância entre o objeto i e o indivíduo é dada por

$$d_i(\mathbf{y}) = \sqrt{\sum_{a=1}^A (x_{ia} - y_a)^2} \quad (3.11)$$

Nesta tese, os objetos são as categorias de resposta dos itens. Em um problema que o indivíduo deve escolher uma entre várias categorias de resposta, ele escolherá aquela mais próxima neste espaço Euclidiano.

A Figura 3.3 ilustra o modelo de desdobramento de Coombs, em que as distâncias são representadas em uma dimensão. Nesta figura, há 2 indivíduos (representados pelos números 1 a 2) avaliando 4 categorias de resposta (representados pelas letras A, B, C e D).

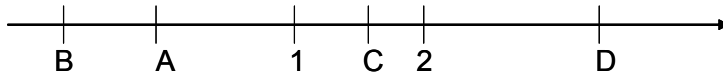


Figura 3.3 – Representação do modelo de desdobramento de Coombs em uma dimensão.

O indivíduo 1 ordena as 4 categorias de resposta pela ordem de preferência da seguinte forma: C, A, B e D, sendo a categoria “C” a mais preferida. Isto porque esta categoria possui menor distância em relação ao “ponto ideal” do indivíduo.

Se o eixo for “dobrado” no ponto ideal do indivíduo 1, como mostra a Figura 3.4, obtém-se o *ranking* de preferência deste indivíduo. O mesmo pode ser feito com o indivíduo 2.

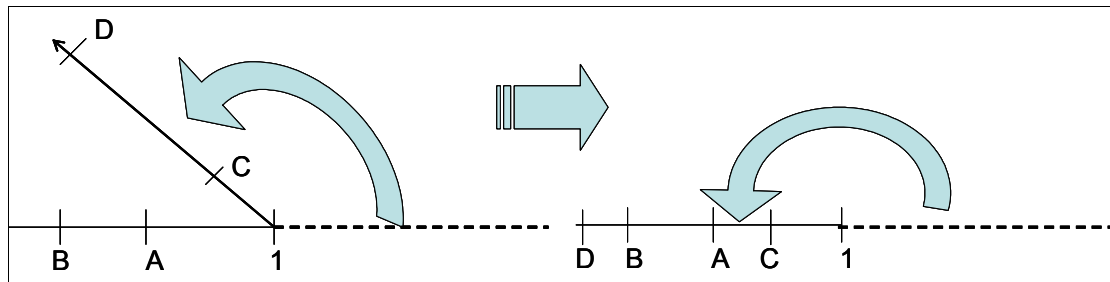


Figura 3.4 – Ilustração da “dobra” do eixo, formando a ordenação original de preferência.

Esta abordagem considera que as escolhas são feitas de modo determinístico, isto é, o indivíduo escolherá com certeza a categoria de resposta menos distante. Quando se utiliza uma dimensão, pode haver inconsistência dos dados na medida em que uma dimensão pode

não comportar as preferências dos indivíduos. Modelos de Escalonamento Multidimensional (GREEN, 1972) permitem representar os pontos em um número maior de dimensões a fim de preservar as distâncias observadas pela ordenação feita pelo respondente e representá-las no “mapa”.

Em oposição aos modelos determinísticos, os modelos de escolha, como o desenvolvido por LUCE (1959), postulam que quando um consumidor define sua preferência por uma alternativa (no caso, uma categoria de resposta) traz consigo uma probabilidade de escolha $p_i(\mathbf{y})$. Esta probabilidade pode ser dada por

$$p_i(\mathbf{y}) = \frac{e^{-d_i(\mathbf{y})}}{\sum_{i=1}^I e^{-d_i(\mathbf{y})}}, \quad (3.12)$$

em que I é o número de alternativas (categorias de resposta).

Este modelo postula que cada resposta tem uma “força”, dada por $\exp(-d_i(\mathbf{y}))$, que é uma função decrescente da distância. Uma resposta é escolhida com probabilidade proporcional às “forças” relativas de uma resposta em relação às outras. O denominador da equação anterior serve para normalizar as probabilidades, tornando sua soma igual a 1.

O modelo de desdobramento graduado generalizado, apresentado a seguir, utiliza as premissas do modelo de ponto ideal de Coombs com características dos modelos da TRI cumulativos.

3.3.2. Apresentação do GGUM

O GGUM é diferente do modelo da TRI usualmente utilizado, que tem caráter cumulativo. Isto significa, no contexto desta tese, que uma probabilidade alta de considerar o atributo importante ocorre quanto mais o parâmetro de localização do indivíduo estiver distante, à direita, do parâmetro de localização do atributo no espaço unidimensional.

Para comparar os dois modelos da TRI, será utilizado um exemplo com duas categorias de resposta para cada item, às quais serão atribuídos os valores 0 e 1. Estas duas categorias podem ser concordo/ discordo, sim/ não, certo/ errado, importante/ não importante. No exemplo, será utilizado importante/ não importante, que é o objeto de pesquisa da tese. À categoria “não importante”, será atribuído o valor 0 e à categoria “importante”, o valor 1.

O “escore” ou “valor” esperado do item para o indivíduo j é dado pela soma do produto de cada valor pela sua probabilidade, isto é, $E_{ij} = 0 \cdot P(X_{ij} = 0) + 1 \cdot P(X_{ij} = 1)$.

Esta definição pode ser estendida para casos em que há mais de duas categorias de resposta, desde que estas categorias sejam ordenadas. No caso apresentado, com duas categorias, o escore esperado é igual à probabilidade de resposta da categoria com valor 1.

O modelo GGUM incorpora o modelo de desdobramento de Coombs, em que um alto escore em um atributo é mais provável quanto mais próximo o ponto ideal do indivíduo (θ_j) estiver da localização do atributo no espaço unidimensional (δ_i) seja à esquerda ou à direita. Com isto, não é possível dizer, no modelo GGUM, que quanto maior θ_j , maior o escore do atributo.

O Gráfico 3.3 ilustra a função resposta dos dois modelos (cumulativo e desdobramento) no caso em que o parâmetro de localização do item é zero.

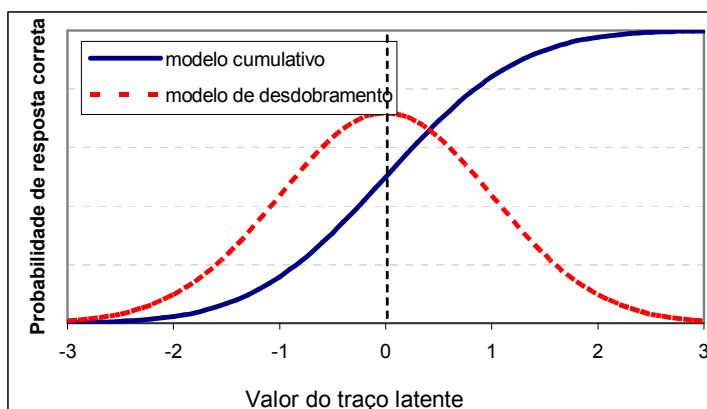


Gráfico 3.3 – Probabilidade de resposta correta em um modelo de TRI cumulativo e um modelo não cumulativo (desdobramento) para diferentes valores do traço latente.

Os modelos de desdobramento são apropriados para medir uma grande variedade de construtos. Por exemplo, atitudes medidas por escalas Thrustone ou Likert (ANDRICH, 1996; ROBERTS, 1995). Neste contexto, os modelos de desdobramento predizem um alto escore de atributo (maior probabilidade de concordância com o atributo) na medida em que o parâmetro de localização do indivíduo está próximo do parâmetro do atributo.

Foram desenvolvidos alguns modelos de desdobramentos tais como modelo de Co-seno Hiperbólico (ANDRICH & LUO, 1993), modelo Parella (HOIJTINK, 1990), o modelo de desdobramento graduado (GUM) de ROBERTS (1995) e ROBERTS & LAUGHLIN (1996ab) e finalmente o Modelo de Desdobramento Graduado Generalizado conhecido por GGUM e desenvolvido por ROBERTS *et al.* (2000).

O GGUM tem sido aplicado em situações de medidas onde os entrevistados são solicitados a indicarem seu nível de concordância com um conjunto de itens que se situam num intervalo contínuo bipolar, ou seja, que variam o seu conteúdo, do negativo, para o positivo, passando pelo neutro.

No Brasil, o GGUM foi aplicado a um estudo de satisfação em relação a um curso de tecnologia no Paraná (BORTOLOTTI, 2003).

A utilização do modelo de desdobramento graduado generalizado nesta tese decorre do fato de que o modelo da TRI cumulativo não é adequado para estimar habilidades ou traços latentes a partir de escalas Thrustone ou Likert utilizadas para medir construtos como atitude ou satisfação. ROBERTS (1995) mostrou em simulações que o modelo cumulativo não apresenta estimativas consistentes dos parâmetros e que os modelos de desdobramentos são mais adequados.

Em um estudo sobre atitude em relação ao aborto, citado em ROBERTS (2000), 50 itens (afirmações) foram apresentados a 750 estudantes de uma universidade. Os estudantes avaliaram cada item em 6 categorias: discordo totalmente, discordo, discordo em parte, concordo em parte, concordo e concordo totalmente.

Similarmente aos modelos da TRI cumulativos, o GGUM permite a representação dos itens e indivíduos em uma mesma escala unidimensional. A localização do item nesta escala é

denotada por δ_i (este parâmetro pode ser comparado ao parâmetro b_i do modelo da TRI cumulativo; a mudança foi feita para seguir notação utilizada na literatura) e a localização do indivíduo, por θ_j .

A seguir, são apresentados três itens e seus respectivos parâmetros de localização no *continuum* latente, δ_i , estimados.

- a) “O aborto é inaceitável sob quaisquer circunstâncias”; $\delta_1 = -3,5$.
- b) “O aborto deve ser ilegal exceto em casos de estupro ou incesto”; $\delta_2 = -1,0$;
- c) “O aborto deve ser legalizado sob quaisquer circunstâncias”; $\delta_3 = 2,8$.

Um indivíduo com parâmetro $\theta_j = -3$ tem grande probabilidade de concordar com a sentença a) e discordar das outras, pois $|\theta_j - \delta_1| = |-3 - (-3,5)| = 0,5$ para o item a) e a distância é grande para os outros dois itens. Já um indivíduo com $\theta_j = 3$ provavelmente discorda das duas primeiras sentenças e tem alta probabilidade de concordar com a sentença c), devido à proximidade entre o parâmetro deste item (2.8) e o parâmetro do indivíduo (3.0).

Este é um exemplo claro em que o modelo da TRI cumulativo não serve, pois este modelo pressupõe que quanto maior o parâmetro do indivíduo, maior a probabilidade dele concordar com quaisquer itens. O traço latente estudado, que representa a atitude em relação ao aborto não é cumulativo, isto é, a probabilidade de concordância com os itens não aumenta conforme aumenta θ_j .

O GGUM permite estudar construtos que apresentam este tipo de comportamento. A probabilidade de concordância com o item, dada a distância $\theta_j - \delta_i$, segue o formato apresentado no Gráfico 3.4.

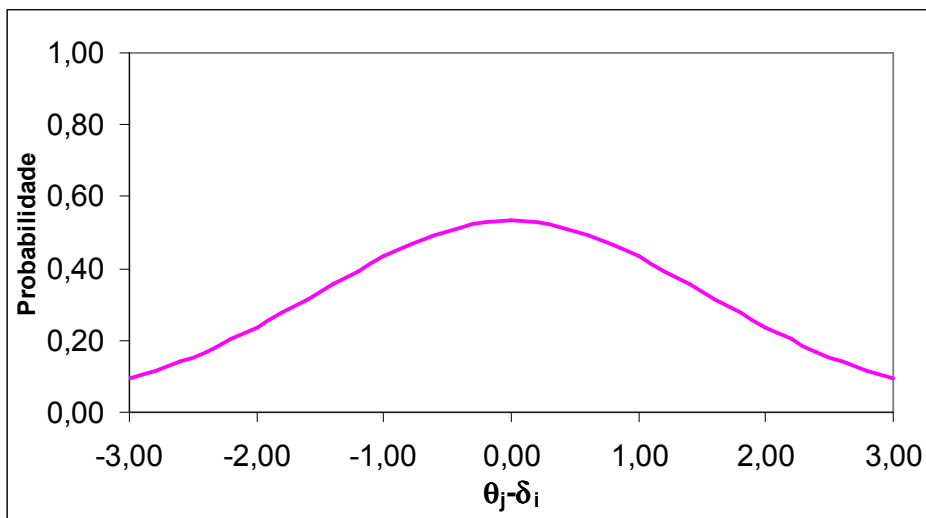


Gráfico 3.4 - Probabilidade de concordância com um item pela distância entre o item e o indivíduo no *continuum latente*.

Para introduzir o GGUM, usar-se-á o caso particular de respostas dicotômicas, que depois será estendido para o caso mais geral.

3.3.3. O modelo dicotômico

O GGUM foi construído com base em 4 premissas sobre o processo de resposta.

A primeira é que o indivíduo j tende a concordar com um item i na medida em que este está localizado próximo ao seu ponto ideal (θ_j) em um *continuum* latente, isto é, na medida em que $(\delta_i - \theta_j)$ está próximo de zero.

A segunda premissa é que o indivíduo pode concordar ou discordar de um item por dois motivos. No exemplo do aborto, uma pessoa com posição neutra sobre o assunto provavelmente discordará “acima” da primeira frase (“O aborto é inaceitável sob quaisquer circunstâncias”) e discordará “abaixo” da terceira frase (“O aborto deve ser legalizado sob quaisquer circunstâncias”). Quando o indivíduo discorda acima, significa que sua atitude em relação ao assunto é mais positiva que o item e quando ele discorda abaixo, sua atitude em relação ao assunto é mais negativa que o item.

Desta forma, há duas respostas subjetivas para cada resposta objetiva ao item. A resposta subjetiva indica se o indivíduo está acima ou abaixo do item e não é externada pelo respondente. No continuum latente, “acima” seria “à direita” do item e “abaixo”, à esquerda.

O exemplo mais simples (dicotômico) é um modelo com apenas duas categorias de resposta objetivas: discordo e concordo. Neste modelo, há 4 categorias subjetivas: discordo abaixo, concordo abaixo, concordo acima e discordo acima.

Seja Z_{ij} a resposta dada ao item i pelo indivíduo j . Z_{ij} varia entre $0, 1, 2, \dots, C$, sendo 0 o maior nível de discordância do item e C o maior nível de concordância do item. No exemplo do aborto com 2 categorias observáveis, se $Z_{ij} = 0$ significa que o indivíduo j discorda do item i e $Z_{ij} = 1$ significa que ele concorda.

Para cada uma das respostas Z_{ij} há duas respostas subjetivas. Seja Y_{ij} a resposta subjetiva ao item i dada pelo indivíduo j . Y_{ij} varia de 0 a M (0 a 3 neste caso) e portanto há $M+1$ (4, neste caso) categorias subjetivas de resposta. A Figura 3.5 mostra os valores de Y_{ij} e Z_{ij} de acordo com as respostas objetivas e subjetivas.

			discorda abaixo	concorda abaixo	concorda acima	discorda acima	
							Parâmetro do indivíduo
							Parâmetro do item
Resposta objetiva mais provável	Z_{ij}	0	1	1	0		
Resposta subjetiva correspondente	Y_{ij}	0	1	2	3		

Figura 3.5 – Relação entre valores mais prováveis de resposta subjetiva e objetiva e distância entre parâmetro do item e do indivíduo.

A terceira premissa do GGUM é que as $M+1$ respostas subjetivas seguem um modelo cumulativo de teoria de resposta ao item, mais especificamente o modelo de créditos parciais generalizado proposto por MURAKI (1992) e possui as mesmas suposições (independência local e unidimensionalidade).

A probabilidade do indivíduo j dar uma resposta subjetiva $Y_{ij}=y$ ao item i por este modelo é:

$$P(Y_{ij} = y | \theta_j) = \frac{\exp \left\{ \alpha_i \left[y(\theta_j - \delta_i) - \sum_{k=0}^y \tau_{ik} \right] \right\}}{\sum_{w=0}^M \exp \left\{ \alpha_i \left[w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik} \right] \right\}} \quad (3.13)$$

Sujeito à restrição

$$\sum_{k=0}^M \tau_{ik} = 0,$$

em que:

$w=0, \dots, M$ e $k=0, \dots, M$ representam uma resposta subjetiva qualquer;

θ_j é o parâmetro do indivíduo, e representa seu ponto ideal (localização) no traço latente considerado;

δ_i é o parâmetro de localização do item no traço latente;

α_i é um parâmetro de discriminação do item i . Sua interpretação é similar ao parâmetro a_i do modelo de créditos parciais generalizados. Quanto maior o valor de α_i , menor a mudança na probabilidade da resposta mais provável para uma pequena variação de θ_j .

τ_{ik} é o valor da distância $(\delta_i - \theta_j)$ que determina o ponto em que a k -ésima categoria de resposta subjetiva passa a ter igual probabilidade de resposta sobre $(k-1)$ -ésima categoria de resposta subjetiva para o indivíduo j no item i . τ_{i0} é, por definição, igual a zero. MURAKI (1992) mostra que este parâmetro pode ser ajustado para qualquer valor, sem mudança nas probabilidades calculadas.

Portanto, os parâmetros τ_{ik} dividem a distância entre a localização do indivíduo j e o item i no *continuum* latente em $M+1$ intervalos, nos quais uma determinada resposta subjetiva é a mais provável. O parâmetro τ_{i3} , por exemplo, é o valor da distância $(\delta_i - \theta_j)$ que determina o ponto em que a 3ª categoria de resposta subjetiva passa a ter igual probabilidade de resposta sobre 2ª categoria de resposta subjetiva para o indivíduo j no item i . Desse modo, τ_{i0} é um valor sem interpretação.

A quarta premissa do modelo GGUM é que os τ_{ik} 's são simétricos em relação ao ponto $(\delta_i - \theta_j) = 0$. Isto implica que $\tau_{iz} = -\tau_{i(M-z+1)}$, para $z \neq 0$. No caso com 2 categorias, $M=3$, e portanto $\tau_{i1} = -\tau_{i(3-1+1)} = -\tau_{i3}$. Portanto, em um modelo com 2 categorias, dado τ_{i1} , os parâmetros τ_{i2} e τ_{i3} são determinados.

Pela definição dada para τ_{ik} , tem-se:

τ_{i3} : é o valor da distância $(\delta_i - \theta_j)$ que determina o ponto em que a 3ª categoria de resposta subjetiva (concordo acima) passa a ter igual probabilidade de resposta sobre 4ª categoria de resposta subjetiva (discordo acima) para o indivíduo j no item i .

τ_{i1} : é o valor da distância $(\delta_i - \theta_j)$ que determina o ponto em que a 1ª categoria de resposta subjetiva (discordo abaixo) passa a ter igual probabilidade de resposta sobre a 2ª categoria de resposta subjetiva (concordo abaixo) para o indivíduo j no item i .

Portanto, esta premissa implica que um indivíduo tem a mesma probabilidade de concordar com o item, estando distante $-h$ ou $+h$ unidades de sua posição no *continuum* latente. Tem-se então a seguinte igualdade:

$$\sum_{k=0}^z \tau_{ik} = \sum_{k=0}^{M-z} \tau_{ik}$$

Considerando $\tau_{i1} = -1$, $\delta_i = 1$ e $\alpha_i = 1$, pela quarta premissa do modelo e pela restrição de que a soma dos valores dos limiares é zero, tem-se os valores de τ_{ik} no Quadro 3.1:

categoria	k	τ_{ik}	Soma acumulada de τ_{ik}
Discordo abaixo	0	0	0
Concordo abaixo	1	-1	-1
Concordo acima	2	0	-1
Discordo acima	3	1	0

Quadro 3.1 – Valores de τ_{ik} para um modelo com duas respostas e $\tau_{i1} = -1$, $\delta_i = 1$ e $\alpha_i = 1$.

Portanto, a probabilidade de cada resposta subjetiva dado θ_j é proporcional a:

$$\begin{aligned}
P(Y_{ij} = 0 | \theta_j) &\propto \exp\{1[0(\theta_j - 1) - (0)]\} = 1 \\
P(Y_{ij} = 1 | \theta_j) &\propto \exp\{1[1(\theta_j - 1) - (-1)]\} = \exp(\theta_j) \\
P(Y_{ij} = 2 | \theta_j) &\propto \exp\{1[2(\theta_j - 1) - (-1)]\} = \exp(2\theta_j - 1) \\
P(Y_{ij} = 3 | \theta_j) &\propto \exp\{1[3(\theta_j - 1) - (0)]\} = \exp(3\theta_j - 3)
\end{aligned} \tag{3.14}$$

As probabilidades são dadas por:

$$\begin{aligned}
P(Y_{ij} = 0 | \theta_j) &= \frac{1}{1 + \exp(\theta_j) + \exp(2\theta_j - 1) + \exp(3\theta_j - 3)} \\
P(Y_{ij} = 1 | \theta_j) &= \frac{\exp(\theta_j)}{1 + \exp(\theta_j) + \exp(2\theta_j - 1) + \exp(3\theta_j - 3)} \\
P(Y_{ij} = 2 | \theta_j) &= \frac{\exp(2\theta_j - 1)}{1 + \exp(\theta_j) + \exp(2\theta_j - 1) + \exp(3\theta_j - 3)} \\
P(Y_{ij} = 3 | \theta_j) &= \frac{\exp(3\theta_j - 3)}{1 + \exp(\theta_j) + \exp(2\theta_j - 1) + \exp(3\theta_j - 3)}
\end{aligned} \tag{3.15}$$

Embora o numerador de $P(Y_{ij}=0)$ seja constante, o denominador aumenta conforme θ_j aumenta. Isto faz com que $P(Y_{ij}=0)$ diminua conforme θ_j aumenta.

Pelos cálculos anteriores, nota-se que $P(Y_{ij}=0) > P(Y_{ij}=1)$ ocorre quando $0 > \theta_j$, isto é, quando $\theta_j < 0$ ou equivalentemente quando $(\theta_j - \delta_i) < -1 = \tau_{i1}$. Fazendo as outras duas igualdades chega-se aos outros limiares τ_{ik} .

A Tabela 3.1 exemplifica as probabilidades calculadas para as categorias subjetivas de resposta, dado θ_j .

Substituindo as fórmulas de cálculo da probabilidade subjetiva dadas na equação 3.13, tem-se:

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\left\{\alpha_i \left[z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik} \right]\right\} + \exp\left\{\alpha_i \left[(M - z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik} \right]\right\}}{\sum_{w=0}^C \left\{ \exp\left\{\alpha_i \left[w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik} \right]\right\} + \exp\left\{\alpha_i \left[(M - w)(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik} \right]\right\} \right\}} \quad (3.17)$$

O Gráfico 3.6 mostra a função de probabilidade para as 2 categorias objetivas do exemplo anterior.

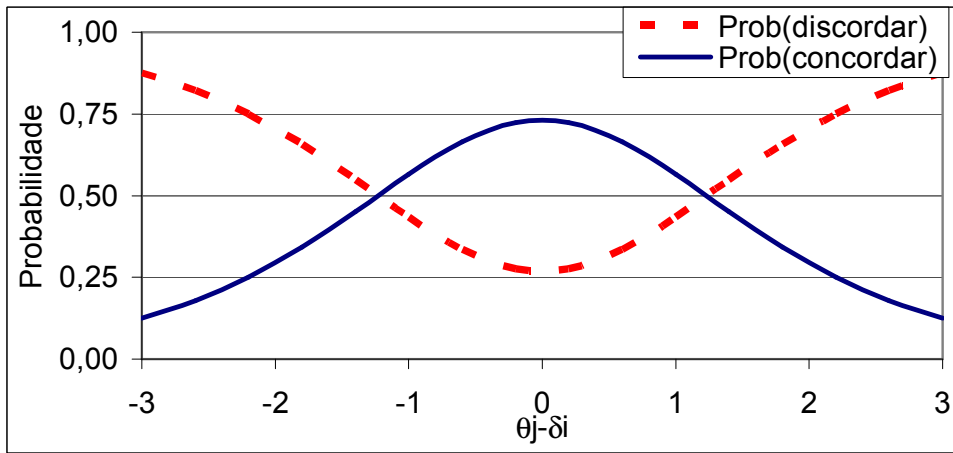


Gráfico 3.6 - Função de probabilidade das 2 categorias objetivas de resposta para $\alpha_i=1$, $\tau_{i1}=-1$, $\tau_{i2}=0$, $\tau_{i3}=1$.

Exemplificando, um indivíduo com $\theta_j=3$ terá alta probabilidade de responder que concorda com um item com $\delta_i=3$. O quão alta será esta probabilidade depende de outro parâmetro: α_i . Quanto maior for este parâmetro, maior a probabilidade deste indivíduo considerar este item importante.

Fixando-se α_i igual a 3 e mantendo τ_{ik} com os mesmos valores, nota-se também que uma pequena mudança no valor de $\theta_j - \delta_i$ provoca uma maior alteração nas curvas de probabilidade, como mostra o Gráfico 3.7. Por isto o parâmetro α_i é chamado de parâmetro de discriminação.

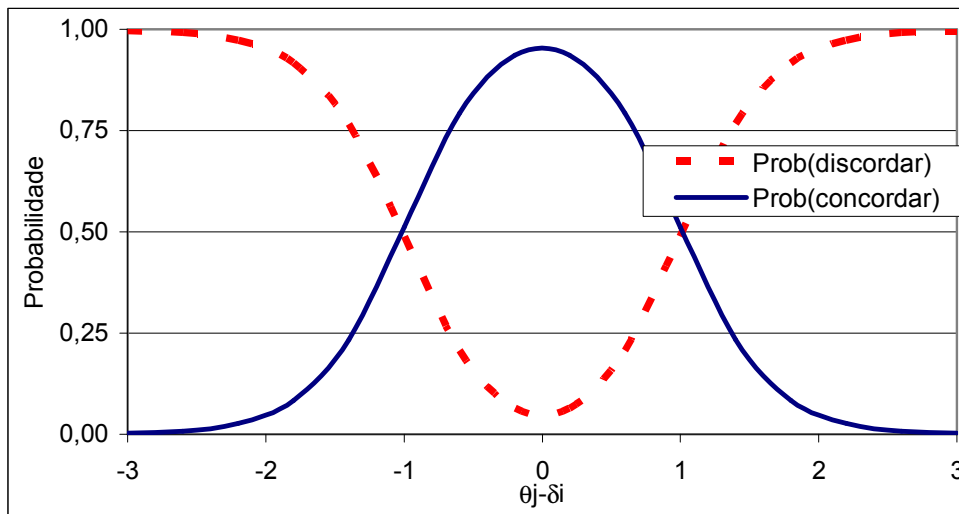


Gráfico 3.7 - Função de probabilidade das 2 categorias objetivas de resposta para $\alpha_i=3$, $\tau_{i1}=-1$, $\tau_{i2}=0$, $\tau_{i3}=1$.

Por outro lado, um valor baixo de α_i leva à quase indistinção das probabilidades qualquer que seja a distância $\theta_j - \delta_i$, como mostra o Gráfico 3.8.

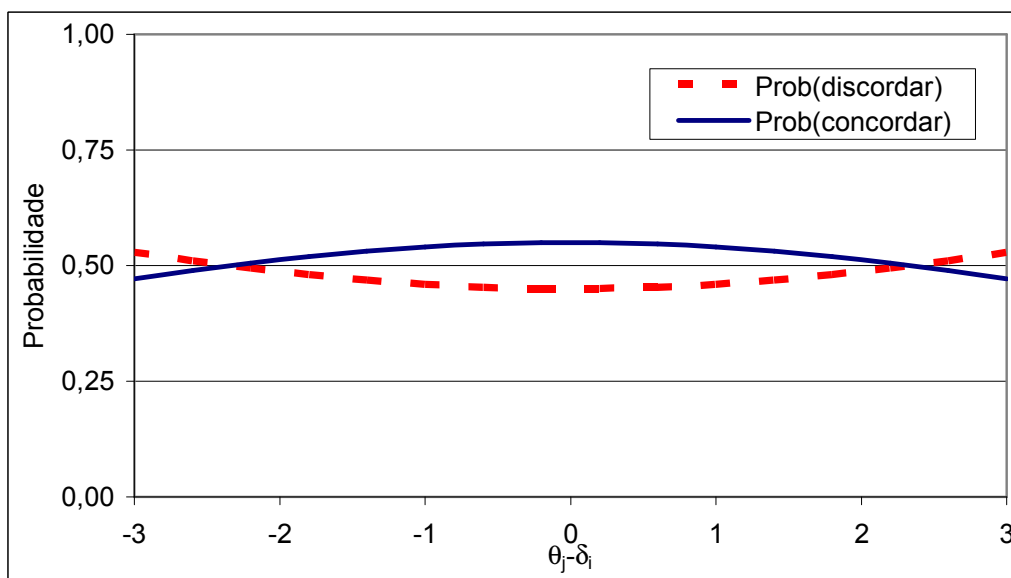


Gráfico 3.8– Probabilidade em cada categoria de resposta variando-se a distância entre o indivíduo e o atributo, para $\alpha_i=0,2$ e $\tau_{i1}=-1$.

Em resumo, valores esperados mais altos ocorrem quando $(\theta_j - \delta_i)$ se aproxima de zero e α_i é grande. A forma da distribuição de probabilidade da resposta e, conseqüentemente, do valor esperado para determinado item depende dos parâmetros α_i e τ_{ik} .

3.3.4. GGUM Politômico

Devido ao fato de, nesta tese, ser modelada a importância dos atributos e não concordância ou discordância em relação a eles, será apresentado um modelo com 4 categorias de resposta objetiva com diferentes graus de importância. O equivalente a este modelo com concordo/discordo poderia ser: discordo totalmente, discordo, concordo, concordo totalmente. Os graus de importância considerados são:

- o atributo não é importante;
- o atributo é pouco importante;
- o atributo é importante;
- o atributo é muito importante.

Estas respostas se “desdobram” em 8 categorias de resposta subjetiva. Isto ocorre porque o indivíduo pode estar “à direita” ou “à esquerda” do item no *continuum* latente.

Neste caso, Y_{ij} varia de 0 a $M=7$, onde M é o número de respostas subjetivas menos 1. Se $Y_{ij}=0$, o respondente dá o menor nível de importância abaixo do item i e se $Y_{ij}=7$ o respondente dá o menor nível de importância acima do item i . Em resumo, um indivíduo próximo do item (abaixo ou acima) tem alta probabilidade de considerá-lo importante.

A Figura 3.6 mostra os valores de Y_{ij} e Z_{ij} de acordo com as respostas objetivas e subjetivas.

			não é importante abaixo	é pouco importante abaixo	é importante abaixo	é muito importante abaixo	é muito importante acima	é importante acima	é pouco importante acima	não é importante acima	
											Parâmetro do indivíduo
											Parâmetro do item
Resposta objetiva mais provável	Z_{ij}	0	1	2	3	3	2	1	0		
Resposta subjetiva correspondente	Y_{ij}	0	1	2	3	4	5	6	7		

Figura 3.6 – Relação entre valores mais prováveis de resposta subjetiva e objetiva e distância entre parâmetro do item e do indivíduo, em um modelo com 6 categorias de resposta objetiva.

Conforme mencionado anteriormente, quanto maior for o parâmetro α_i , mais discriminatório é o item, isto é, maior a diferença na probabilidade de resposta dada a distância entre o item e o indivíduo.

No modelo com 4 categorias de resposta objetiva, há 7 parâmetros τ_{ik} . Devido às restrições impostas pelo modelo, $\tau_{i4}=0$ e, dados τ_{i1} , τ_{i2} e τ_{i3} , os outros τ_{ik} 's são determinados.

Os Gráficos 3.9 e 3.10 ilustram as probabilidades de resposta para cada categoria de dois itens, que, dado $(\theta_j - \delta_i)$, $\tau_{i1}=-2,0$, $\tau_{i2}=-1,3$ e $\tau_{i3}=-0,7$, só diferem em relação ao α_i .

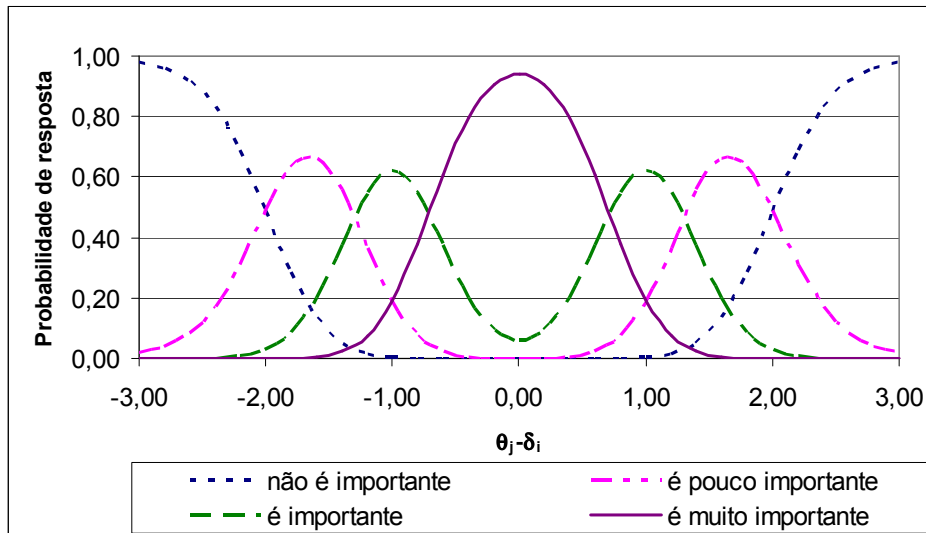


Gráfico 3.9 – Curvas de probabilidade para um item com $\alpha_i = 4$.

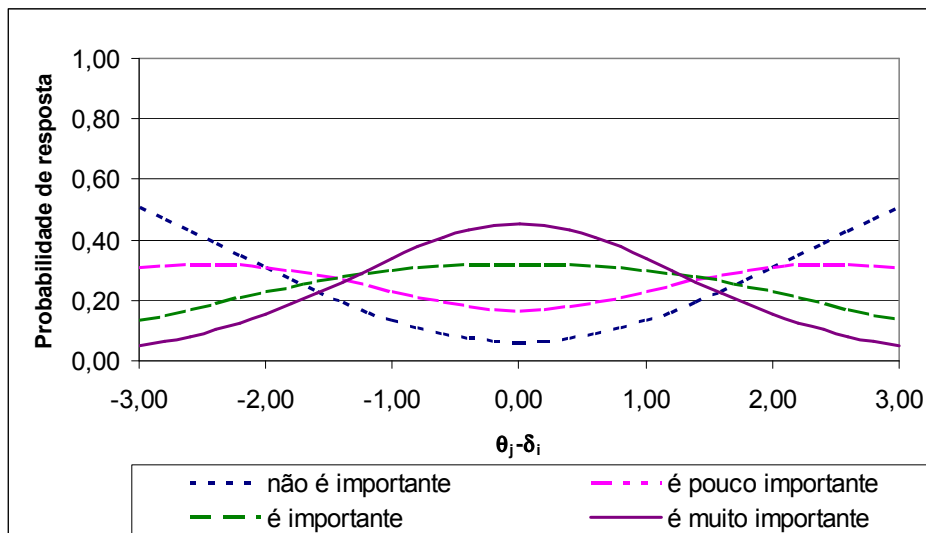


Gráfico 3.10 – Curvas de probabilidade para um item com $\alpha_i = 0,5$.

Nota-se que itens que não têm muita variabilidade na resposta (isto é, as probabilidades de resposta de cada categoria são próximas de 0,25 para qualquer valor de $\theta_j - \delta_i$) possuem um α pequeno, como o item ilustrado no Gráfico 3.10, e itens que possuem respostas concentradas em uma categoria possuem α grande, como o item ilustrado no Gráfico 3.9.

A comparação anterior foi feita com dois itens que possuem mesmos valores para τ_{ik} . Comparando itens que possuem o mesmo α , mas diferentes τ_{ik} 's, nota-se que as probabilidades de resposta de cada categoria estão mais concentradas quanto menor os valores dos limiares.

O comportamento da probabilidade de resposta da categoria “muito importante” é ilustrado no Gráfico 3.11 para α grande e pequeno e τ_{ik} concentrado em torno do zero e disperso.

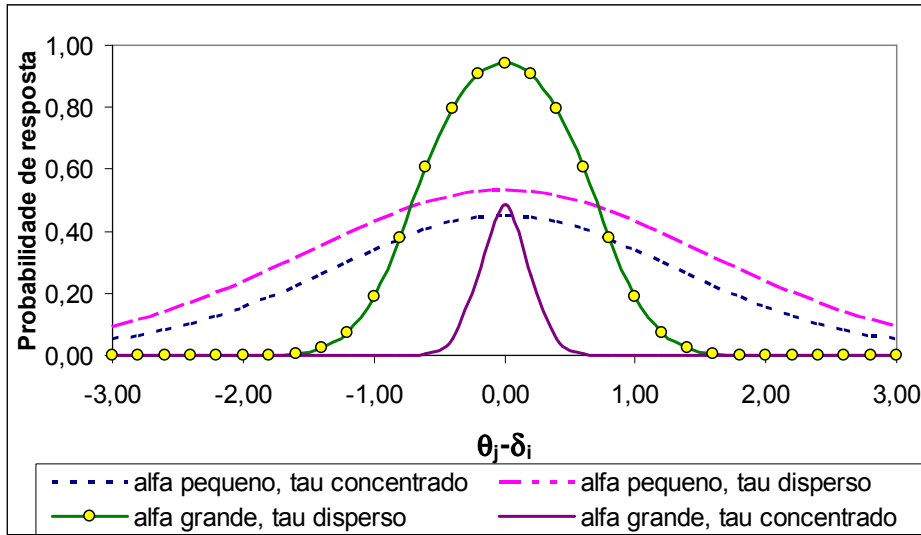


Gráfico 3.11– Probabilidades de resposta na categoria “muito importante” para diferentes valores de alfa (α) e tau(τ).

Em alguns casos, é possível que um item tenha respostas concentradas em uma só categoria mas o seu α seja grande. Isto ocorre quando δ_i é muito distante da maioria dos θ_j 's. Se a maioria dos θ_j 's estiver entre -3 e 3 e se δ_i for 10, mesmo com α grande a probabilidade de resposta concentra-se na categoria “não é importante”. Desta forma, itens com δ_i extremos também tendem a ter a resposta concentrada em uma categoria.

3.3.5. Função de Informação do Item

ROBERTS *et al.* (2000) mostram que a função de informação do item é dada por

$$I_i(\theta_j) = -E \left[\frac{\partial^2 \ln(L)}{\partial \theta_j^2} \right] = \alpha_i^2 \left[\left(\sum_{z=0}^D P(Z_{ij} = z) \sigma_{Y_{ij}|\theta_j, z}^2 \right) - \sigma_{Y_{ij}|\theta_j}^2 \right] \quad (3.18)$$

Em que:

$\sigma_{Y_{ij}|\theta_j, z}^2$ é a variância condicional da resposta subjetiva do indivíduo j ao item i dada a resposta observada do indivíduo a este item e

$\sigma_{Y_{ij}|\theta_j}^2$ é a variância da resposta subjetiva do indivíduo j ao item i .

A Figura 3.7 mostra a função de informação para um item com 6 categorias de resposta objetiva variando-se o valor de α (gráfico b) e a distância entre τ_{ik} e $\tau_{i(k-1)}$. Esta distância será denotada por Ψ .

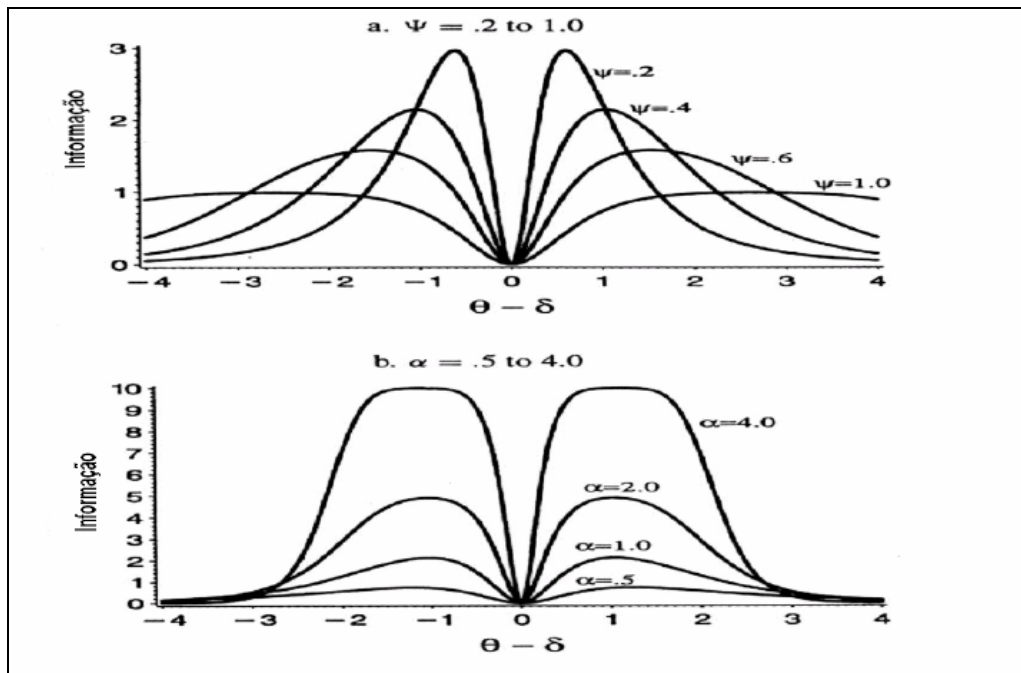


Figura 3.7 – Função de informação do item em um modelo GGUM com 6 categorias de resposta objetiva. No gráfico a, $\alpha_i = 1$ e Ψ varia de 0,2 a 1. No gráfico b, $\Psi = 0,4$ e α_i varia de 0,5 a 4. Fonte: Modificada de ROBERTS *et al.* (2000).

Observa-se que todas as funções de informação do item são bimodais e simétricas em torno de $|\theta_j - \delta_i| = 0$ e aproximam-se de zero quanto menor for a distância $|\theta_j - \delta_i|$.

A informação é maior para valores menores de Ψ , o que implica que a informação é maior quando os τ_{ik} estão concentrados em torno do zero. Em relação ao parâmetro α_i , quanto maior ele for, maior é a informação do item.

3.3.6. Versões do GGUM com restrições adicionais nos parâmetros

As restrições impostas aos parâmetros α_i e τ_{ik} formam sete submodelos do GGUM, apresentados a seguir.

Modelo 1:

Admitindo que as distâncias entre os τ_{ik} 's adjacentes são constantes e igual a $\Psi=2\lambda$ para todos os itens, isto é, $\tau_{ik} - \tau_{i(k-1)} = 2\lambda$ para todo k e todos os itens têm o mesmo valor para o parâmetro de discriminação, isto é, $\alpha_i = 1$, tem-se o modelo de “unidade constante”, descrito em ROBERTS (1995) e ROBERTS & LAUGHLIN (1996ab).

Esta restrição impõe que o número de categorias para cada itens é constante e a forma da distribuição de probabilidade do valor esperado é igual para todos os itens, dada a distância $(\theta_j - \delta_i)$. Isto significa que os intervalos de preferência são iguais para todos os itens. Este modelo pode ser escrito como

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\{z(\theta_j - \delta_i) + z(M - z)\lambda\} + \exp\{(M - z)(\theta_j - \delta_i) + z(M - z)\lambda\}}{\sum_{w=0}^C \{\exp\{w(\theta_j - \delta_i) + w(M - w)\lambda\} + \exp\{(M - w)(\theta_j - \delta_i) + w(M - w)\lambda\}\}} \quad (3.19)$$

Modelo 2:

O primeiro modelo pode ser generalizado permitindo que o parâmetro de discriminação α_i varie para cada item, obtendo-se o modelo de unidade constante generalizado (ROBERTS, 1995; ROBERTS & LAUGHLIN, 1996ab):

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\{\alpha_i [z(\theta_j - \delta_i) + z(M - z)\lambda]\} + \exp\{\alpha_i [(M - z)(\theta_j - \delta_i) + z(M - z)\lambda]\}}{\sum_{w=0}^C \{\exp\{\alpha_i [w(\theta_j - \delta_i) + w(M - w)\lambda]\} + \exp\{\alpha_i [(M - w)(\theta_j - \delta_i) + w(M - w)\lambda]\}\}} \quad (3.20)$$

Modelo 3:

Se os valores de τ_{ik} não variarem entre os itens e não houver restrição das distâncias entre as categorias adjacentes, tem-se o modelo de “rating scale”, denominado anteriormente de “graded unfolding model” ou “GUM” por ROBERTS & LAUGHLIN (1996). Se $\alpha_i = 1$ para todos os itens, tem-se:

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\left\{\left[z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_k\right]\right\} + \exp\left\{\left[(M - z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_k\right]\right\}}{\sum_{w=0}^C \left\{ \exp\left\{\left[w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_k\right]\right\} + \exp\left\{\left[(M - w)(\theta_j - \delta_i) - \sum_{k=0}^w \tau_k\right]\right\} \right\}} \quad (3.21)$$

Modelo 4:

Neste modelo, não há restrição de α_i 's iguais a 1 para todos os itens. Este modelo é chamado de “generalized rating scale”:

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\left\{\alpha_i \left[z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_k\right]\right\} + \exp\left\{\alpha_i \left[(M - z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_k\right]\right\}}{\sum_{w=0}^C \left\{ \exp\left\{\alpha_i \left[w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_k\right]\right\} + \exp\left\{\alpha_i \left[(M - w)(\theta_j - \delta_i) - \sum_{k=0}^w \tau_k\right]\right\} \right\}} \quad (3.22)$$

Modelo 5:

Se a distância entre os τ_{ik} 's adjacentes for igual a $2\lambda_i$ para o item i e $\alpha_i = 1$ para todos os itens, obtém-se o modelo de “unidade múltipla”, descrito em ROBERTS (1995) e ROBERTS & LAUGHLIN (1996ab).

Neste modelo, é permitido que o número de categorias varie de acordo com o item. A probabilidade de uma categoria objetiva é dada por:

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\{z(\theta_j - \delta_i) + z(M_i - z)\lambda_i\} + \exp\{(M_i - z)(\theta_j - \delta_i) + z(M_i - z)\lambda_i\}}{\sum_{w=0}^C \left\{ \exp\{w(\theta_j - \delta_i) + w(M_i - w)\lambda_i\} + \exp\{(M_i - w)(\theta_j - \delta_i) + w(M_i - w)\lambda_i\} \right\}} \quad (3.23)$$

Modelo 6:

O modelo de unidade múltipla pode ser generalizado permitindo que α_i varie entre os itens. Este modelo é chamado de “modelo de unidade múltipla generalizado” (ROBERTS, 1995; ROBERTS & LAUGHLIN, 1996ab):

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\{\alpha_i [z(\theta_j - \delta_i) + z(M_i - z)\lambda_i]\} + \exp\{\alpha_i [(M_i - z)(\theta_j - \delta_i) + z(M_i - z)\lambda_i]\}}{\sum_{w=0}^C \left\{ \exp\{\alpha_i [w(\theta_j - \delta_i) + w(M_i - w)\lambda_i]\} + \exp\{\alpha_i [(M_i - w)(\theta_j - \delta_i) + w(M_i - w)\lambda_i]\} \right\}} \quad (3.24)$$

Modelo 7:

Neste submodelo, chamado de “modelo de crédito parcial”, τ_{ik} varia sem restrições entre os itens mas todos α_i ’s são iguais a 1:

$$P(Z_{ij} = z | \theta_j) = \frac{\exp\left\{z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right\} + \exp\left\{(M - z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}\right\}}{\sum_{w=0}^C \left\{ \exp\left\{w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}\right\} + \exp\left\{(M - w)(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}\right\} \right\}} \quad (3.25)$$

O Modelo 8 não possui restrições e a probabilidade resposta a uma categoria objetiva é dada na equação 3.17. A Tabela 3.3 resume as restrições impostas por cada submodelo nos parâmetros α_i e τ_{ik} .

Tabela 3.3 – Características de cada submodelo do GGUM

Submodelo	α_i	τ_{ik}
1	$\alpha_i = 1$	$\tau_{ik} - \tau_{i(k-1)} = 2\lambda$ para todo i
2	α_i varia livremente	$\tau_{ik} - \tau_{i(k-1)} = 2\lambda$ para todo i
3	$\alpha_i = 1$	τ_{ik} ’s iguais para todos os itens
4	α_i varia livremente	τ_{ik} ’s iguais para todos os itens
5	$\alpha_i = 1$	$\tau_{ik} - \tau_{i(k-1)} = 2\lambda_i$
6	α_i varia livremente	$\tau_{ik} - \tau_{i(k-1)} = 2\lambda_i$
7	$\alpha_i = 1$	τ_{ik} varia livremente entre os itens
8	α_i varia livremente	τ_{ik} varia livremente entre os itens

3.3.7. Estimação dos parâmetros

Os parâmetros dos itens são estimados pelo método da verossimilhança marginal (BOCK & LIEBERMAN, 1970; BOCK & AITKIN, 1981). Seja \mathbf{X}_s um dos possíveis vetores-resposta para um indivíduo em determinado conjunto de dados e x_{si} o i -ésimo elemento de \mathbf{X}_s . Dada a suposição de independência local, podemos afirmar que

$$P[X_s | \theta] = \prod_{i=1}^I P[Z_i = x_{si} | \theta] \quad (3.26)$$

Admitindo que θ_j tem uma distribuição de probabilidade $g(\theta_j)$, a probabilidade marginal de observar um padrão de resposta \mathbf{X}_s é igual a:

$$P[X_s] = \int_{-\infty}^{+\infty} P[X_s | \theta] g(\theta) d\theta \quad (3.27)$$

Deve-se salientar que, para obter esta integral, é necessário especificar uma distribuição para θ_j . Esta distribuição restringe a locação e a escala do *continuum* latente de tal forma que o modelo fique identificável. A distribuição normal padronizada é geralmente utilizada neste caso.

Seja r_s o número de indivíduos na amostra de tamanho N com padrão de resposta \mathbf{X}_s . Então r_s tem distribuição multinomial com parâmetros N e $P(\mathbf{X}_s)$, e a função de verossimilhança é igual a

$$L = \frac{N!}{\prod_{s=1}^S r_s!} \prod_{s=1}^S P[X_s]^{r_s} \quad (3.28)$$

E a log-verossimilhança é dada por

$$\ln(L) = \ln(N!) - \ln\left(\prod_{s=1}^S r_s!\right) + \sum_{s=1}^S r_s \ln(P[X_s]) \quad (3.29)$$

As equações de verossimilhança para parâmetros τ_{ik} , δ_i e α_i são obtidas através das derivadas parciais do logaritmo da função de verossimilhança (equação 3.29) em relação a cada um dos parâmetros e igualando a zero. O algoritmo EM é utilizado para resolver as equações de verossimilhança para os parâmetros (MURAKI, 1992).

Os parâmetros θ_j dos indivíduos podem ser estimados pelo método de Esperança a Posteriori - EAP (BOCK & MISLEVY, 1982). A estimativa da EAP, é a média *a posteriori* da distribuição de θ para o *j-ésimo* indivíduo dado o seu vetor de respostas.

Tanto a estimação pela EAP quanto a máxima verossimilhança marginal utilizam uma distribuição *a priori* para θ nos seus cálculos. A integral desta distribuição é feita com uma

série de pontos de quadratura discretos, da mesma forma que é feita no modelo da TRI cumulativo. ROBERTS, DONOGHUE & LAUGHLIN (2002) mostram que tanto as estimativas de máxima verossimilhança marginal quanto a estimativas pela EAP são robustas em relação à distribuição *a priori* escolhida.

A equação para estimação dos parâmetros dos indivíduos é dada por:

$$\hat{\theta}_j = \frac{\sum_{f=1}^F V_f L_f(V_f) A(V_f)}{\sum_{f=1}^F L_f(V_f) A(V_f)} \quad (3.30)$$

Em que $L_f(V_f)$ é a verossimilhança condicional de se observar o vetor de resposta do indivíduo j , dado que o indivíduo está localizado no ponto de quadratura V_f , sendo F o número de pontos de quadratura considerado.

O desvio-padrão *a posteriori* aproximado de $\hat{\theta}_j$ é dado por (BOCK & MISLEVY, 1982):

$$\sigma_{\hat{\theta}_j} = \sqrt{\frac{\sum_{f=1}^F (V_f - \theta_j) L_j(V_f) A(V_f)}{\sum_{f=1}^F L_j(V_f) A(V_f)}} \quad (3.31)$$

3.3.8. Itens sem resposta

Quando os dados faltantes são resultantes da não apresentação do item ao indivíduo, tem-se o caso de *missing at random* (MAR) dado θ . Este caso é diferente do caso em que o item é apresentado ao indivíduo e este, por algum motivo que pode estar relacionado ao item ou ao próprio indivíduo, não o responde. Na primeira situação, pode-se calcular a verossimilhança do vetor de respostas simplesmente ignorando os itens em que não há resposta para determinado indivíduo.

Uma simulação feita por DAVIS & ROBERTS (2004) sugere que os parâmetros dos itens e dos indivíduos podem ser estimados adequadamente com 20% de dados faltantes. Esta

conclusão vale para testes que possuam pelo menos 10 itens com 6 categorias de resposta por item.

3.3.9. Ajuste do modelo

A maioria das estatísticas de ajuste dos modelos de desdobramento é derivada dos modelos de TRI cumulativos. As estatísticas de ajuste do modelo geralmente envolvem a comparação entre resposta observada e predita pelo modelo e são estatísticas que seguem uma distribuição aproximada qui-quadrado. Estas estatísticas possibilitam testar hipóteses sobre o ajuste do modelo. Nos testes qui-quadrado, as hipóteses são:

- Hipótese nula (H_0): os dados se ajustam ao modelo, isto é, o valor previsto e observado são próximos;
- Hipótese alternativa (H_A): os dados não se ajustam ao modelo, isto é, o valor previsto e observado não são próximos;

Utilizando probabilidade de erro do tipo 1 igual a α (não confundir com os parâmetros dos modelos da TRI), rejeita-se H_0 quando o p-valor do teste for menor que este valor. Em outras palavras, quando o p-valor for menor que α , há evidências estatísticas de que os dados não se ajustam ao modelo. Para maiores detalhes sobre testes de hipóteses, ver BUSSAB e MORETTIN (2002).

Em um estudo com resposta dicotômica, YEN (1981) dividiu os respondentes em $J=10$ grupos de acordo com seus valores no traço latente e calculou a diferença na proporção do grupo que respondeu ao item corretamente e a proporção prevista e em seguida elevou a diferença ao quadrado, multiplicando pelo tamanho da célula e formando uma estatística qui-quadrado.

BOCK (1972) propôs um índice similar para um modelo de resposta nominal, usando o valor previsto e observado em cada categoria de resposta, para cada grupo.

ANDRICH (1978), calculou para um modelo de “rating scale”, um escore médio previsto e observado para cada grupo e então calculou a estatística qui-quadrado. Foi designado um

valor para cada categoria de resposta ordinal (0, 1, 2 e 3, em um modelo com 4 categorias) para calcular o escore médio.

ANDRICH (1997) generalizou a estatística construída para o modelo de “rating scale” para modelos de desdobramento dicotômicos, agrupando os respondentes em J intervalos de acordo com seu traço latente. O escore médio observado e previsto dentro de cada grupo é calculado e o índice qui-quadrado resultante é dado por

$$\chi^2_i = \sum_{j=1}^J \frac{N_j (\bar{Z}_{ij} - E_{ij})^2}{Var(E_{ij})} \quad (3.32)$$

Em que:

N_j é o número de respondentes no intervalo j , $j=1, \dots, J$;

\bar{Z}_{ij} é a resposta observada média dos indivíduos no intervalo j para o item i ;

E_{ij} é a resposta prevista dos indivíduos no intervalo j para o item i , calculada por

$$E_{ij} = \sum_{z=0}^c z(P(Z_i = z | \bar{\theta}_j)) \quad \text{e} \quad Var(E_{ij}) = \sum_{z=0}^c (z - E_{ij})^2 (P(Z_i = z | \bar{\theta}_j))$$

e $\bar{\theta}_j$ é o valor médio do traço latente no intervalo j .

Pode-se também, em vez de fazer o quadrado das diferenças, calcular o log da razão do valor observado sobre o previsto:

$$G^2_i = 2 \sum_{j=1}^J \sum_{z=0}^c O_{ijz} \ln \left[\frac{O_{ijz}}{N_j P[Z_i = z | \bar{\theta}_j]} \right], \quad (3.33)$$

em que:

O_{ijz} é o número de respondentes no intervalo j que escolheram a categoria z para o item i ;

N_j é o número de respondentes no intervalo j ;

$P\left[Z_i = z \mid \bar{\theta}_j\right]$ é a probabilidade de escolher a categoria z do item i , sendo $\bar{\theta}_j$ o valor médio do traço latente no intervalo j .

A estatística apresentada na equação 3.33 possui distribuição qui-quadrado com JC graus de liberdade (MURAKI & BOCK, 1997). O ajuste do modelo como o todo, denotado por G^2 , pode ser obtido pela soma dos qui-quadrados obtidos para cada item, com graus de liberdade igual à soma dos graus de liberdade das estatísticas individuais (ROBERTS, FANG, CUI & WANG, 2004). Quanto maior o valor dessa estatística, pior o ajuste do modelo.

Alguns índices, tipicamente utilizados em modelos Rasch, não requerem que os respondentes sejam agrupados. Para cada indivíduo, é calculado o resíduo através da diferença entre as respostas observadas e previstas. Assim como no qui-quadrado de ANDRICH (1978; 1997), calcula-se a diferença entre o valor da categoria escolhida pelo indivíduo e o valor previsto (calculado pela soma do produto do valor da categoria pela sua probabilidade de ser escolhida). Os resíduos resultantes são padronizados e elevados ao quadrado.

O cálculo do qui-quadrado para os itens pode ser feito de duas maneiras: pode-se calcular a média dos resíduos padronizados dos respondentes para o item (outfit) ou fazer uma ponderação do resíduo pela variância do resíduo para o indivíduo (infit). As fórmulas, dadas por WRIGHT & MASTERS (1982) são:

$$\text{outfit}_i = \frac{1}{N} \sum_{j=1}^N \frac{(Z_{ij} - E_{ij})^2}{\text{Var}(E_{ij})} \quad (3.34)$$

$$\text{infit}_i = \frac{\sum_{j=1}^N \left(\text{Var}(E_{ij}) \frac{(Z_{ij} - E_{ij})^2}{\text{Var}(E_{ij})} \right)}{\sum_{j=1}^N (\text{Var}(E_{ij}))} \quad (3.35)$$

Em que:

N é o número de respondentes;

Z_{ij} é a resposta observada do indivíduo j para o item i ;

E_{ij} é a resposta prevista do indivíduo j para o item i ,

$$E_{ij} = \sum_{z=0}^c z(P(Z_i = z | \hat{\theta}_j)) \quad \text{e} \quad Var(E_{ij}) = \sum_{z=0}^c (z - E_{ij})^2 (P(Z_i = z | \hat{\theta}_j))$$

C é o número de categorias do item.

Nos modelos Rasch, os índices infit e outfit têm média 1. Valores maiores que 1.3 (KARABATSOS, 2000; SMITH, SCHUMACKER & BUSH, 1998) ou 1.5 (ENGELHARD, 1992; LINACRE, 1991; LUNZ, WRIGHT & LINACRE, 1990) são utilizados para destacar itens com ajuste ruim.

A distribuição das estatísticas apresentadas, bem como a probabilidade real de erro do tipo I e o poder dos testes para o GGUM é pouco conhecida (DEMARS, 2004). A autora mostra que as estatísticas de infit e outfit, mesmo utilizando probabilidade de erro do tipo I de 5%, levam à rejeição da hipótese de um bom ajuste bem menos que 5% dos itens. A estatística qui-quadrado de Andrich, por outro lado, leva à rejeição de bem mais que 5% dos itens quando se utiliza 5% de probabilidade de erro do tipo I. A estatística do log da razão de verossimilhança apresentou rejeição mais próxima da probabilidade de erro do tipo I, e portanto deve ser preferida às anteriores (DEMARS, 2004). Estes resultados mostram que estas medidas devem ser utilizadas apenas como ajuste relativo, em vez de ajuste absoluto.

Também é possível construir gráficos para verificar o ajuste do modelo para cada item. O gráfico comumente feito para este fim (HAMBLETON e SWAMINATHAN, 1985) coloca os valores esperados e observados no eixo Y, e a distância $(\delta_i - \theta_j)$ no eixo X.

Para construir este gráfico, os indivíduos são agrupados em k grupos; em seguida, calcula-se um valor médio para $(\delta_i - \theta_j)$, para o valor observado e para o valor previsto neste grupo. A Figura 3.8 mostra um exemplo (com $\delta = -0,8$ e $\alpha = 2,1$) em que $k=1$, $k=5$ e $k=20$, para uma amostra de 1000 indivíduos.

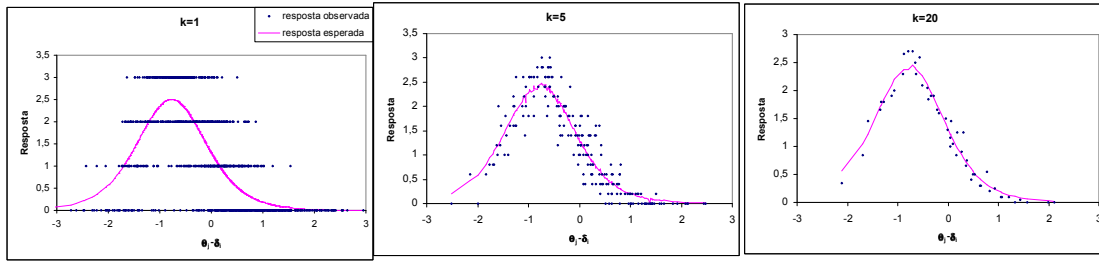


Figura 3.8 – Gráficos descritivos do ajuste do modelo para diferentes tamanhos de grupo.

Como pôde ser visto na Figura 3.8, quanto maior for o tamanho do grupo, menor será a discrepância entre o valor previsto médio e o observado médio no grupo. Este é um dos motivos pelo qual este gráfico não será a única forma de verificar o ajuste. O outro motivo é que, como este gráfico mostra o ajuste de um item, deveriam ser feitos muitos gráficos para verificar o ajuste do modelo no geral. Além disto, nesta tese são ajustados vários modelos diferentes, o que impossibilita a análise por meio deste gráfico.

Uma medida de ajuste que ao mesmo tempo é bem resumida e interpretável é o quadrado médio do erro de previsão, dado pela média dos quadrados da diferença entre o valor previsto e o observado. Esta medida será utilizada para avaliar os modelos e compará-los com casos em que o valor previsto é dado pela média do item.

Quanto menor for o quadrado médio da diferença entre os valores observados e esperados (erro), melhor o ajuste do modelo. Este quadrado médio, chamado daqui em diante de *QMErro*, é dado por

$$QMErro = \frac{1}{N} \sum_{j=1}^N (Z_{ij} - E_i)^2 \quad (3.36)$$

Em que:

E_{ij} é a resposta prevista do indivíduo j para o item i ;

N é o número de respondentes;

Z_{ij} é a resposta observada do indivíduo j para o item i .

Com isto, é possível ter uma idéia do desvio-padrão do erro de previsão (dado pela raiz quadrada do $QMErro$).

3.3.10. O software GGUM2004

Os cálculos envolvidos nos processos de estimação dos parâmetros dos itens e dos indivíduos são bastante trabalhosos e exigem recursos computacionais específicos. Para realizar estes cálculos, foi utilizado o software GGUM2004, desenvolvido por ROBERTS, DONOGHUE & LAUGHLIN (2000) e ROBERTS, FANG, CUI & WANG (2004) e está disponível para *download* gratuito em <http://www.education.umd.edu/EDMS/tutorials/FreeSoftware.html>.

O programa GGUM2004 possui algumas melhorias em relação ao anterior GGUM2000; dentre as melhorias, destaca-se a migração para o ambiente Windows, a interface gráfica, a variação do número de categorias em cada item e principalmente a possibilidade de estimar os parâmetros com dados faltantes, que são tratados como *missing at random*.

Ainda há algumas limitações no programa, que são descritas a seguir:

- a) o número máximo de indivíduos é 2000;
- b) o número máximo de itens é 100;
- c) o número máximo de categorias de resposta por item é 10;
- d) o limite do número de pontos de quadratura é 50;
- e) a distribuição a priori é limitada à distribuição normal.

Os parâmetros dos itens são estimados pelo método de máxima verossimilhança marginal e as estimativas dos parâmetros dos indivíduos são obtidas pelo procedimento de esperança *a posteriori* (EAP), detalhado no tópico de estimação dos parâmetros.

3.4. UTILIZAÇÃO DO GGUM NA TESE

Pretende-se mostrar na tese que o GGUM também pode ser utilizado para medir importância em situações em que o respondente avalia um conjunto de I atributos (itens).

A metodologia para testar as hipóteses da tese está dividida em duas partes.

Na primeira, foram simulados dados que seguem o modelo proposto, variando-se o número de itens e seus parâmetros, o tamanho da amostra e os parâmetros dos indivíduos. O objetivo da simulação é verificar como algumas medidas de ajuste se comportam perante dados que seguem o modelo e estudar este ajuste em função da quantidade de dados faltantes para cada respondente. Para realizar este último estudo, algumas respostas foram excluídas do banco de dados e estimou-se o modelo apenas com os dados restantes.

Na segunda parte, foi verificado o ajuste do modelo para dados coletados a partir de um questionário aplicado a 400 alunos do curso de graduação em Administração da EAESP/FGV sobre 25 atributos de uma academia de ginástica. O objetivo é mensurar a importância de cada atributo com base nas estimativas geradas pelo GGUM e, novamente, verificar o ajuste quando há dados faltantes. Em seguida, os resultados do modelo são interpretados e é discutida a exclusão de alguns itens não informativos.

4 INVESTIGAÇÃO EMPÍRICA: SIMULAÇÃO

4.1. OBJETIVO

O objetivo principal dessas simulações é estudar o comportamento das estimativas do parâmetro e do erro de previsão em função da quantidade de dados faltantes na matriz de importância. Neste capítulo também são discutidas várias alternativas para se obter o valor previsto de resposta para o indivíduo.

A primeira simulação foi feita com $N=1000$ observações e $I=25$ itens. Esta simulação é apresentada com mais detalhes no que diz respeito à comparação entre valores estimados e observados.

Em seguida, são apresentadas e comparadas algumas alternativas de obtenção do valor previsto utilizando o GGUM ou outras técnicas.

Por último, são feitas mais simulações variando-se o número de itens, o tamanho de amostra e a quantidade de dados excluídos.

4.2. METODOLOGIA

Os dados utilizados neste capítulo foram gerados supondo-se que o traço latente estudado é unidimensional e que os indivíduos e os itens podem ser representados nesta dimensão. O comportamento das respostas dos indivíduos depende dos parâmetros α_i , τ_{ik} , δ_i e θ e segue as premissas do GGUM.

O traço latente θ dos respondentes foi gerado aleatoriamente segundo uma distribuição normal padronizada. Os valores de α_i e τ_{ik} também variaram, sendo criados itens com α_i abaixo e acima de 1 e τ_{ik} concentrados em torno do zero e dispersos.

A partir dos parâmetros dos itens e dos indivíduos, calculou-se a probabilidade de resposta objetiva Z_{ij} em cada categoria a partir da fórmula apresentada na equação 3.17. A resposta do indivíduo foi sorteada obedecendo-se à probabilidade calculada.

4.3. DESCRIÇÃO DOS ITENS E INDIVÍDUOS UTILIZADOS NA PRIMEIRA SIMULAÇÃO

Na primeira simulação, utilizou-se $N=1000$ observações e $I=25$ itens. A Tabela 4.1 mostra os parâmetros dos itens utilizados para gerar os dados.

Tabela 4.1 – Parâmetros dos itens utilizados na primeira simulação

	α	τ_1	τ_2	τ_3	δ		α	τ_1	τ_2	τ_3	δ
item1	0,20	-3,00	-2,00	-1,00	-2,40	item14	2,00	-3,00	-2,00	-1,00	0,20
item2	0,50	-2,75	-1,83	-0,92	-2,20	item15	4,00	-2,75	-1,83	-0,92	0,40
item3	1,00	-2,50	-1,67	-0,83	-2,00	item16	0,20	-2,50	-1,67	-0,83	0,60
item4	2,00	-2,25	-1,50	-0,75	-1,80	item17	0,50	-2,25	-1,50	-0,75	0,80
item5	4,00	-2,00	-1,33	-0,67	-1,60	item18	1,00	-2,00	-1,33	-0,67	1,00
item6	0,20	-1,75	-1,17	-0,58	-1,40	item19	2,00	-1,75	-1,17	-0,58	1,20
item7	0,50	-1,50	-1,00	-0,50	-1,20	item20	4,00	-1,50	-1,00	-0,50	1,40
item8	1,00	-1,25	-0,83	-0,42	-1,00	item21	0,20	-1,25	-0,83	-0,42	1,60
item9	2,00	-1,00	-0,67	-0,33	-0,80	item22	0,50	-1,00	-0,67	-0,33	1,80
item10	4,00	-0,75	-0,50	-0,25	-0,60	item23	1,00	-0,75	-0,50	-0,25	2,00
item11	0,20	-0,50	-0,33	-0,17	-0,40	item24	2,00	-0,50	-0,33	-0,17	2,20
item12	0,50	-0,25	-0,17	-0,08	-0,20	item25	4,00	-0,25	-0,17	-0,08	2,40
item13	1,00	-0,20	-0,10	-0,05	0,00						

Dentre os 25 itens simulados, há alguns com baixo valor do parâmetro α . Espera-se que estes itens sejam pouco informativos sobre o traço latente considerado e o erro de previsão destes itens seja grande. Outros itens, com parâmetro α grande devem ter comportamento oposto.

4.3.1. Estimação dos parâmetros a partir dos dados simulados

Com os dados gerados a partir da simulação, foi ajustado o modelo do tipo 8, completo, sem restrições adicionais, descrito no Capítulo 3. Este modelo possui 125 parâmetros de itens a serem estimados (5 para cada item; 25 δ_i 's, 25 α_i 's e 75 τ_{ik} 's no total).

A tabela do Apêndice A apresenta várias estatísticas de ajuste do modelo. Em resumo, tem-se que os p-valores das estatísticas dos itens (infit, outfit e G^2) são bastante discrepantes, isto é, para o mesmo item uma estatística se apresenta não significativa enquanto outra estatística se apresenta muito significativa. Isto implica que algumas estatísticas podem mostrar falta de ajuste enquanto outras podem mostrar um bom ajuste para o mesmo item.

No geral, o modelo apresenta um bom ajuste (p-valor da estatística $G^2=0,7974$), embora alguns itens apresentem p-valor baixo, evidenciando um ajuste ruim.

Outra maneira de visualizar o ajuste do modelo é através da comparação entre o valor observado e valor previsto da resposta do indivíduo para determinado item. Uma primeira alternativa para calcular o valor previsto é pela moda. A resposta prevista será aquela com maior probabilidade de escolha.

De posse dos valores esperados e observados, é possível organizá-los numa tabela de dupla entrada com os valores observados *versus* os valores previstos. Inicialmente, para se ter uma idéia de ajuste, serão comparados os valores observados (gerados pela simulação) versus a previsão obtida utilizando os valores reais dos parâmetros $(\theta_j, \delta_i, \alpha_i, \tau_{ik})$. O resultado é mostrado na Tabela 4.2.

Tabela 4.2 – Frequência conjunta dos valores gerados (observados) e previstos pelo GGUM utilizando os valores reais dos parâmetros

		Observado				Total
		0	1	2	3	
Previsto	0	4220	1159	575	420	6374
	1	3550	2823	1734	1114	9221
	2	479	1011	1716	1379	4585
	3	271	496	1286	2767	4820
Total		8520	5489	5311	5680	25000

A diagonal principal soma 11526 respostas (46,1%) e a estatística Sommers'd, que mede a associação entre duas variáveis ordinais, apresenta um valor de 0,49 (p-valor=0). Esta estatística, assim como a correlação de Pearson, varia entre -1 e 1, sendo o valor 0 um indicador de não associação entre as variáveis e -1 e 1 um valor indicador de associação máxima entre as variáveis. Pode-se concluir portanto que há evidências estatísticas de associação entre as variáveis, e a associação é razoável.

Quando se calcula os valores previstos pelas estimativas dos parâmetros $(\hat{\theta}_j, \hat{\delta}_i, \hat{\alpha}_i, \hat{\tau}_{ik})$, obtém-se a frequência conjunta apresentada na Tabela 4.3:

Tabela 4.3 – Frequência conjunta dos valores gerados e previstos pelo GGUM utilizando os valores estimados dos parâmetros

		Observado				Total
		0	1	2	3	
Previsto	0	4243	1156	557	414	6370
	1	3603	2938	1750	1184	9475
	2	370	890	1666	1113	4039
	3	304	505	1338	2969	5116
Total		8520	5489	5311	5680	25000

Neste caso, a diagonal principal soma 11816 respostas (47,3%) e a estatística Sommers'd apresenta um valor de 0,52 (p-valor=0), novamente indicando associação razoável entre as variáveis.

É interessante notar que, quando se utilizam os valores estimados para os parâmetros, a porcentagem de acerto melhora em relação aos valores previstos com os parâmetros reais, o que é esperado, pois os valores estimados maximizam a probabilidade de observação dos valores observados. Os valores previstos neste último caso são bastante parecidos com aqueles previstos quando se utilizam as estimativas dos parâmetros. Isto pode ser visto na Tabela 4.4, que mostra que 21221 (84,9%) respostas têm a mesma previsão nos dois casos:

Tabela 4.4 – Distribuição de frequência conjunta das respostas dos valores previstos pelos parâmetros estimados e reais.

		Observado				Total
		0	1	2	3	
Previsto	0	5991	360	12	11	6374
	1	294	8195	492	240	9221
	2	57	736	2981	811	4585
	3	28	184	554	4054	4820
Total		6370	9475	4039	5116	25000

Percebe-se que o erro de previsão é grande (a porcentagem de acerto é pequena) mesmo com tamanho de amostra grande e com dados gerados pelo próprio modelo. Isto ocorre porque o número de categorias de resposta é grande (4) e alguns itens têm o parâmetro de discriminação, α_i , baixo.

4.3.2. Alternativas de cálculo para o valor previsto

No lugar da moda como valor previsto, pode-se usar o valor médio (esperado) de resposta dos indivíduos para o item i , que é dado por

$$E_{ij}=E_i=\frac{1}{N}\sum_{j=1}^N Z_{ij} \quad e$$

$$Var(E_{ij})=Var(E_i)=QMErro=\frac{1}{N}\sum_{j=1}^N (Z_{ij}-E_i)^2$$

Em que:

E_{ij} é a resposta prevista do indivíduo j para o item i ;

N é o número de respondentes;

Z_{ij} é a resposta observada do indivíduo j para o item i .

Quanto menor for o QMErro, melhor o ajuste do modelo.

Uma outra alternativa é utilizar a probabilidade de uma resposta nas categorias “importante” ou “Muito importante”, obtida pelo GGUM, e atribuí-la à importância dada ao item. Esta alternativa difere das anteriores na medida em que a importância não é mais um ponto da escala; é uma probabilidade.

Para avaliar o GGUM em comparação ao caso em que a importância não é modelada por um traço de comportamento, será estudado o caso em que a previsão é feita pela média do item e pela média geral. Tem-se, então, cinco alternativas de previsão:

- A. pela resposta média do item i dada por todos os indivíduos, \bar{Z}_i ;
- B. pela resposta média geral de todos os indivíduos, \bar{Z} ;
- C. pelo valor com maior probabilidade de escolha (obtido pelo GGUM), M_{ij} ;
- D. pelo valor esperado da resposta, dado pela multiplicação de cada valor atribuído à categoria de resposta pela sua probabilidade (obtido pelo GGUM), E_{ij} ;
- E. pela probabilidade de resposta na categoria “importante” ou “muito importante”, obtida pelo GGUM e denotada por P_{ij} .

Por exemplo, a probabilidade de cada uma das alternativas para o item 5, para um indivíduo com $\theta=0$ (e portanto $\theta_j - \delta_i = -1,6$) é dada na Tabela 4.5.

Tabela 4.5 – Probabilidade de resposta em cada categoria para o item 5.

Resposta	Não é importante	é pouco importante	é importante	é muito importante
Valor da resposta	0	1	2	3
Probabilidade calculada pelo GGUM	0,134	0,662	0,199	0,005

As alternativas A e B não utilizam o GGUM. Na alternativa A, é necessário calcular a média das respostas dadas por todos os indivíduos àquele item. No caso da simulação, a resposta média é 1,25. A alternativa B utiliza a média geral das respostas de todos os indivíduos (1,33).

Pela alternativa C, o valor previsto é 1, que é aquele com maior probabilidade de escolha. Utilizando a alternativa D, o valor previsto é

$$E_{ij} = 0 \bullet 0,134 + 1 \bullet 0,662 + 2 \bullet 0,199 + 3 \bullet 0,005 = 1,077$$

Para tornar o resultado independente dos valores dados às categorias, tanto os valores observados Z_{ij} quanto os esperados foram re-escalados para ficarem no intervalo [0;1]. Para fazer isto, basta dividir os valores originais por 3 (valor máximo da escala – valor mínimo).

Desta forma, no exemplo anterior E_{ij} será $1,077/3=0,359$, a resposta média re-escalada será 0,417 e a média geral, 0,443.

Esta normalização coloca a escala de importância entre 0 e 100%, sendo:

- 0% - Nenhuma importância
- 33% - Um pouco importante
- 67% - Importante
- 100% - Muito importante

Deve-se ressaltar que estas previsões, exceto a alternativa C, admitem uma escala de importância, que originalmente é discreta, contínua.

Para calcular o erro de previsão utilizando a alternativa E (pela probabilidade de resposta nas categorias “importante” e “Muito importante”) será considerado como “valor observado” o valor re-escalado da resposta.

Por exemplo, se um indivíduo seleciona a categoria 2 para um atributo, o valor observado re-escalado é $2/3=0,666$. O erro de previsão será zero caso a soma das probabilidades de resposta nas categorias “Importante” ou “Muito importante” seja igual a este valor.

Portanto, a importância do item 5 para este indivíduo, utilizando as probabilidades estimadas de resposta do GGUM, pode ser dada pela:

- probabilidade do indivíduo escolher a categoria 2: importante ou 3: muito importante (0,204);
- resposta esperada re-escalada (0,359);
- resposta mais provável re-escalada (0,333).

O re-escalamento foi feito para colocar as estimativas de importância no mesmo intervalo (entre 0 e 1), embora o número resultante não tenha a mesma interpretação nas cinco alternativas, pois, conforme foi mencionado anteriormente, a alternativa E produz uma importância que é dada por uma probabilidade e nas outras alternativas a importância é dada por um valor esperado ou modal calculado a partir da escala original.

No Apêndice A é mostrado o QMErro para cada alternativa, calculado com as previsões e respostas re-escaladas. Todos os itens apresentam menor QMErro na alternativa D, em que se calcula o valor previsto pela resposta esperada utilizando os parâmetros estimados pelo GGUM. Embora este resultado seja esperado, já que os dados foram gerados utilizando o próprio GGUM, é interessante observar o quanto o QMErro diminuiu. A Tabela 4.6 mostra os valores do QMErro da previsão para as cinco alternativas.

Tabela 4.6 – QMErro da previsão da simulação re-escalado com $N=1000$, $I=25$, para as cinco alternativas.

Alternativa	Descrição da alternativa	QMErro
A	a resposta média do item i dada por todos os indivíduos;	0,117
B	a resposta média geral.	0,151
C	o valor com maior probabilidade de escolha (obtida pelo GGUM);	0,119
D	o valor esperado da resposta (obtida pelo GGUM);	0,077
E	a probabilidade de resposta nas categorias "importante" e "muito importante"	0,079

O resultado mostra que a estimação pelo valor com maior probabilidade de escolha, segundo o GGUM, não produz resultados satisfatórios. Das cinco alternativas, a que produz melhor resultado (menor QMErro) é aquela em que o valor previsto é dado pelo valor esperado da resposta (alternativa D). A alternativa E, que atribui a probabilidade à importância, também apresentou um QMErro baixo, muito próxima à obtida na alternativa D.

Nota-se que itens com baixo α , nas cinco alternativas, têm maior QMErro de previsão. Isto se deve ao grande desvio-padrão de respostas para estes itens e à baixa informação proporcionada por eles, o que significa que o traço latente em questão não permite maior conhecimento da opinião do indivíduo sobre aquele atributo. Veja no Apêndice A que os itens com baixo α (itens 1, 6, 11, 16 e 21) apresentam QMErro pela alternativa D próximo de 0,13 e itens com alto α (itens 5, 10, 15, 20 e 25) apresentam QMErro menor que 0,03.

Como visto anteriormente, quanto menor o valor do parâmetro α , menor a concentração das respostas em uma categoria, o que implica que os itens com maior α (itens 5, 10, 15, 20 e 25) possuem menor QMErro de previsão.

Além do erro de previsão, são apresentados no Apêndice A erros quadráticos médios relacionados aos parâmetros α_i , θ_i e δ_i . Todos estes erros são calculados pela média do quadrado da diferença entre o valor estimado e o real.

O QMErro relacionado ao parâmetro δ_i foi o que apresentou maior valor: 0,093 (ver Apêndice A). Este valor alto se deveu muito aos itens extremos (1, 24 e 25), que possuem δ_i distante de zero. Excluindo estes itens, o $QMErro(\delta_i)$ diminui para 0,018 (50 vezes menor que o valor com todos os itens).

O item 25 também contribuiu para um aumento considerável no QMErro do parâmetro α_i (ver Apêndice A). Excluindo este item, o $QMErro(\alpha_i)$ diminui de 0,09 para 0,02. Em geral, os itens que apresentaram maior erro do parâmetro α_i são os itens com valor alto do parâmetro (itens 5, 10, 15, 20 e 25).

Em relação ao parâmetro θ_i , o QMErro obtido foi 0,036, o que implica que um intervalo com 95% de confiança para θ_i possui amplitude igual a $1,96 * \sqrt{QMErro} = 0,37$. O Gráfico 4.1 mostra que o valor estimado está próximo do real.

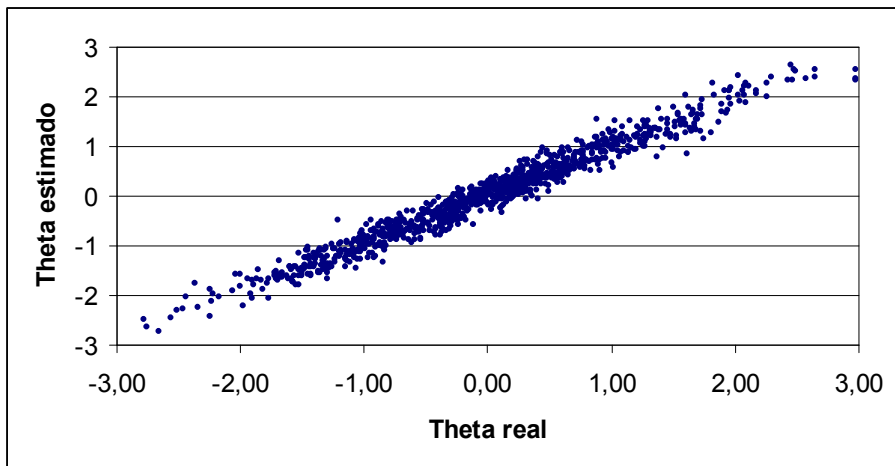


Gráfico 4.1- Gráfico de dispersão dos valores de θ estimados e reais.

4.4. SIMULAÇÕES COM DADOS EXCLUÍDOS

Para estudar o comportamento do modelo com valores faltantes, foram excluídas respostas de k itens de cada indivíduo. As k respostas apagadas foram selecionadas de acordo com um quadrado latino completo de tamanho I , onde I é o número de itens.

O quadrado latino completo (WINER, 1991) é uma matriz $I \times I$, em que I é o número de atributos, tal que cada atributo aparece uma única vez em cada linha e cada coluna. Um exemplo de um quadrado latino completo de tamanho 4 (admitindo que há 4 atributos) é dado na Tabela 4.7.

Tabela 4.7 – ordem dos atributos apresentados geradas por um quadrado latino de tamanho 4.

	coluna 1	coluna2	coluna3	coluna4
linha1	1	2	3	4
linha2	2	4	1	3
linha3	3	1	4	2
linha4	4	3	2	1

Cada linha representa um vetor de atributos apresentados a um indivíduo. Por exemplo, para o segundo indivíduo (linha 2), os atributos são apresentados na seguinte ordem: 2,4,1 e 3.

Em uma simulação com 25 atributos e 1000 respondentes, por exemplo, são necessários 40 quadrados latinos de tamanho 25, totalizando 1000 vetores ordenados dos atributos. Estes quadrados latinos foram gerados pelo programa GENSTAT. Para que os quadrados latinos não se repetissem, foram colocadas sementes diferentes em cada geração. A ilustração do primeiro quadrado latino gerado é mostrada no Apêndice N.

Exemplificando, para simular respostas faltantes a dois itens para cada indivíduo, as duas últimas colunas do quadrado latino são apagadas. Desta forma, cada item é respondido um mesmo número de vezes.

Foram feitas várias simulações, variando-se:

- o número de indivíduos (400 e 1000);
- o número de itens (25 e 50);
- o número de dados excluídos (0%, 40% e 60%);

Os parâmetros dos 25 itens constantes da Tabela 4.1 foram replicados uma vez, gerando 50 itens.

Para avaliar os modelos, foram utilizados o QMErro da previsão, o QMErro dos parâmetros e o valor da estatística G^2 .

A estimação pelo modelo 8 (sem restrições nos parâmetros) não foi possível em casos que alguma categoria de resposta em algum item teve frequência de resposta igual a zero. Neste caso, não é possível estimar o ponto do continuum latente (τ_{ik}) em que a categoria sem resposta e a categoria adjacente a ela têm a mesma probabilidade. O software acusa casos em que há alguma categoria de resposta com frequência zero em algum item e o programa é abortado. Foi utilizado, portanto, o modelo 6, que admite que as distâncias entre os τ_{ik} são constantes dentro do item. Desta forma é possível estimar todos os τ_{ik} mesmo com categorias sem resposta.

4.5. RESULTADOS

Utilizando o modelo 6, que restringe a distância entre os τ_{ik} 's, o QMErro de previsão não aumenta consideravelmente. Na simulação com 1000 observações e 25 itens, o QMErro de previsão utilizando o modelo 6 e escore médio calculado pelo GGUM (alternativa D), é 0,079 – apenas 0,002 pontos acima do QMErro obtido na simulação sem dados excluídos. A estatística G^2 também não é significativa neste caso, mostrando um bom ajuste do modelo. Isto ocorre porque os dados foram gerados a partir de itens com τ_{ik} 's equiespaçados.

Outras simulações com 80% de casos excluídos foram feitas, mas as estimativas dos parâmetros não convergiram.

A Tabela 4.8 e 4.9 mostram resultados de algumas simulações feitas variando-se o tamanho da amostra, o número de itens e a porcentagem de dados excluídos.

Tabela 4.8 – Estatísticas de ajuste dos parâmetros do modelo para cada simulação

Simulação	Modelo	tamanho da amostra	número de itens	número de respostas excluídas	QMErro theta	QMErro delta	QMErro alpha	p-valor G2
1	8	1000	50	0	0,021	0,714	0,085	0,950
2	6	1000	50	30 (60%)	0,058	0,393	0,392	0,550
3	8	1000	25	0	0,036	0,932	0,090	0,797
4	8	1000	25	15 (60%)	0,133	0,913	0,807	x
5	8	400	50	0	0,039	0,392	0,123	0,993
6	6	400	50	30 (60%)	0,218	1,420	0,634	0,075
7	8	400	25	0	0,044	0,356	0,124	1,000
8	6	400	25	10 (40%)	0,091	0,286	0,562	0,945
9	6	400	25	15 (60%)	0,175	2,373	0,432	0,518

x: valor não calculado pelo software

Tabela 4.9 – QMErro de previsão para as cinco alternativas de cálculo, para cada simulação.

Simulação	QMErro previsão alternativa A	QMErro previsão alternativa B	QMErro previsão alternativa C	QMErro previsão alternativa D	QMErro previsão alternativa E	número de respostas por item	Número de respostas por indivíduo
1	0,117	0,150	0,131	0,078	0,081	1000	50
2	0,117	0,150	0,139	0,083	0,085	400	20
3	0,117	0,151	0,119	0,077	0,079	1000	25
4	0,117	0,151	0,130	0,083	0,087	400	10
5	0,118	0,152	0,121	0,075	0,079	400	50
6	0,118	0,152	0,129	0,081	0,083	160	20
7	0,118	0,151	0,128	0,081	0,086	400	25
8	0,119	0,151	0,131	0,077	0,080	240	15
9	0,119	0,151	0,136	0,082	0,086	160	10

Sobre os resultados, pode-se dizer que:

- Os p-valores observados da estatística G^2 nas simulações estiveram acima de 0,05, embora esta estatística tenha problemas em relação à real probabilidade de erro do tipo I, o que invalida uma análise mais conclusiva sobre o resultado. Houve um caso ($I=1000$; $N=25$; 60% de dados excluídos) em que a estatística G^2 não foi apresentada pelo software, por motivo desconhecido;
- Com ou sem dados excluídos, a resposta prevista com o GGUM baseada no escore médio (alternativa D) e baseada na probabilidade de resposta na categoria 2 e 3 (alternativa E)

apresentam menor QMErro em comparação às outras alternativas. Embora isto seja esperado, já que os dados foram gerados a partir deste modelo, é interessante notar o quanto este erro é menor do que o erro quando se utiliza a média do item;

- O QMErro dos parâmetros α_i , θ_i e δ_i aumenta consideravelmente com o aumento da quantidade de respostas excluídas, com a diminuição do tamanho da amostra e com a diminuição do número de itens. Este aumento se deve a itens com valores dos parâmetros “extremos”, por exemplo δ_i distante de zero.
- O QMErro de previsão utilizando-se as alternativas D e E estão em torno de 0,08 e variam pouco em função do tamanho da amostra (400 ou 1000), do número de itens (25 ou 50) e da porcentagem de dados excluídos por respondente (0% a 60%). Este valor é 30% menor do que o QMErro de previsão quando se utiliza a média do item (sem o modelo GGUM; alternativa A) e 45% menor quando se utiliza a média geral;

O fato curioso encontrado nos resultados é que há aumento do erro das estimativas dos parâmetros do item e do indivíduo quando há respostas excluídas, mas o erro de previsão se mantém praticamente constante. Isto se deve à presença de itens com baixo valor de α na simulação. A previsão para estes itens utilizando o escore médio pouco muda para pequenas variações de θ . Isto pode ser visto no Gráfico 4.2, que mostra a resposta prevista pelo escore médio para um item com $\alpha=0,5$.

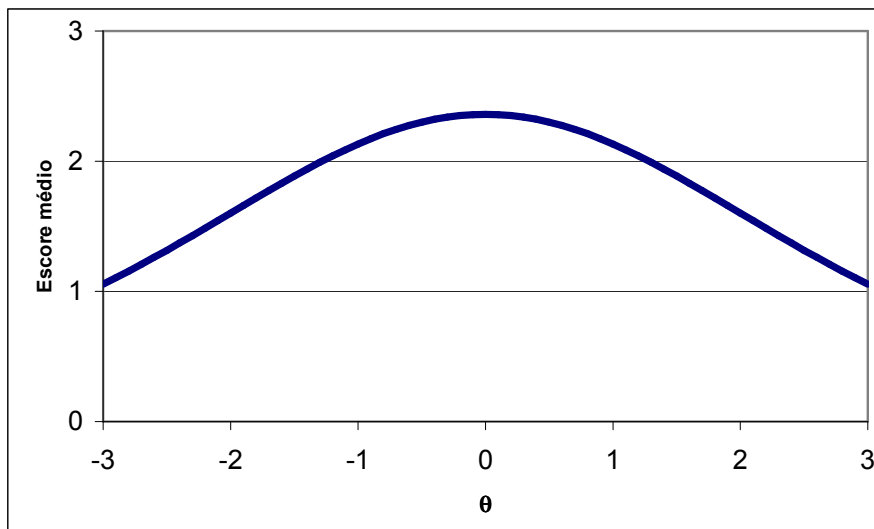


Gráfico 4.2 – Previsão pelo escore médio para um item com $\alpha_i=0,5$, $\delta_i=0$, $\tau_{1i}=-3$, $\tau_{2i}=-2$ e $\tau_{3i}=-1$.

Isto implica que um erro nas estimativas dos parâmetros θ e δ não produz grandes alterações na previsão para este caso. Um parâmetro que tem grande impacto no erro de previsão é o α . Itens com α grande possuem menor QMErro de previsão, mas este erro aumenta muito (Tabela 4.10) quando há dados excluídos. Itens com $\alpha=4$ apresentam um aumento de 40% a 80% no QMErro de previsão, enquanto que itens com $\alpha=0,2$ apresentam um aumento em torno de 1%. Apesar deste aumento, os valores obtidos do QMErro com 15 dados faltantes para estes itens (5, 10, 15, 20 e 25) ainda são menores do que o QMErro quando se utiliza a alternativa A (ver Apêndice A) – média das respostas dadas ao item.

Tabela 4.10 – Comparação entre o QMErro de previsão, utilizando o escore médio pelo GGUM (alternativa D), na simulação com $I=1000$ e $N=25$, para dados completos e com 15 respostas excluídas por respondente.

Item	α	QMErro com dados completos	QMErro com 15 dados excluídos	Diferença percentual dos QMErros
1	0,20	0,129	0,130	0,9%
2	0,50	0,102	0,104	2,7%
3	1,00	0,074	0,078	5,5%
4	2,00	0,040	0,047	17,6%
5	4,00	0,026	0,039	51,2%
6	0,20	0,131	0,131	0,0%
7	0,50	0,119	0,122	2,3%
8	1,00	0,090	0,097	8,3%
9	2,00	0,060	0,076	26,7%
10	4,00	0,034	0,063	87,8%
11	0,20	0,139	0,140	1,1%
12	0,50	0,130	0,131	1,0%
13	1,00	0,105	0,106	0,6%
14	2,00	0,029	0,034	18,7%
15	4,00	0,018	0,025	40,9%
16	0,20	0,126	0,128	1,5%
17	0,50	0,102	0,103	1,7%
18	1,00	0,078	0,085	9,0%
19	2,00	0,053	0,061	15,3%
20	4,00	0,023	0,042	80,2%
21	0,20	0,138	0,138	0,4%
22	0,50	0,102	0,104	2,1%
23	1,00	0,049	0,053	6,9%
24	2,00	0,019	0,023	18,2%
25	4,00	0,005	0,009	64,3%
Total		0,077	0,083	7,8%

No cálculo do QMErro de previsão dos dados no geral, estes itens têm pouco peso por serem baixos e, mesmo com aumento percentual significativo, continuam baixo. Este é o motivo pelo qual o QMErro apresenta pouca variação em função do número de dados excluídos por respondente. Se a simulação fosse realizada apenas com itens informativos (valor alto de α), o QMErro de previsão apresentaria maior variação em função do número de dados excluídos.

4.6. CONCLUSÃO

Como foi discutido anteriormente, pretende-se utilizar o GGUM basicamente com dois propósitos: modelar a matriz de importância (isto é, recuperar os valores observados com base em previsões do modelo) e representar o comportamento de preferência dos indivíduos em um traço latente.

A simulação realizada, apesar de ser um caso particular, mostra que não há perda de precisão do escore de importância mesmo quando se trabalha com 60% de dados faltantes para 25 e 50 itens no total, devido à presença de itens com baixo α . Os erros dos estimadores dos parâmetros dos itens e indivíduos são altos, contudo a previsão obtida possui QMErro menor do que quando se utiliza a resposta média dos indivíduos ao item. Portanto, o modelo se mostra adequado para modelar a matriz de importância em casos com dados faltantes.

Para estimar θ_j com erro pequeno, são necessárias pelo menos 20 respostas por indivíduo com tamanho de amostra 1000 e pelo menos 25 respostas por indivíduos (dados completos) com tamanho de amostra 400. Neste caso, é possível estimar as localizações dos indivíduos no *continuum* latente com pequeno erro.

5 INVESTIGAÇÃO EMPÍRICA: PESQUISA DE CAMPO

5.1. OBJETIVO

O objetivo da pesquisa de campo é ilustrar uma situação real em que o modelo proposto na tese é aplicado para modelar a matriz de importância de atributos de uma academia de ginástica. Também é feita uma análise dos resultados de um ponto de vista mais aplicado, interpretando os parâmetros dos itens e segmentando os indivíduos de acordo com o seu valor no traço latente estimado e outras características constantes no questionário.

5.2. METODOLOGIA DE CONSTRUÇÃO DOS QUESTIONÁRIOS

5.2.1. Escolha dos atributos

Para averiguar quais atributos os alunos consideram importante quando escolhem uma academia, foi feita uma pesquisa inicial com uma turma da graduação da EAESP/FGV em que se perguntou:

“Se você fosse escolher uma academia de ginástica para freqüentar, quais atributos você consideraria importantes na sua escolha?”

Dos 50 alunos da turma, 39 responderam a enquete. A tabela do Apêndice B mostra os atributos citados pelos alunos, classificados por categoria.

Foram selecionados 25 atributos a partir daqueles citados. Tanto atributos muito citados quanto atributos pouco citados e não citados foram selecionados. Atributos pouco específicos, como “ter boa infra-estrutura” não foram selecionados. Isto porque qualquer ação derivada do resultado de um atributo pouco específico é difícil de ser tomada, além do atributo poder ser interpretado de maneira muito diferente pelos respondentes.

Alguns atributos pouco específicos, como “horário flexível” foram quebrados em vários atributos mais específicos. Neste caso, “ficar aberta 24h por dia” e “ficar aberta aos domingos e feriados”. A lista de atributos selecionados é dada no apêndice C.

O próximo passo foi determinar a escala a ser utilizada para mensurar a importância de cada atributo. Na enquete passada aos alunos, foi incluída uma questão com dois tipos de escala. Dos 50 alunos, 25 responderam à pergunta “A academia estar localizada perto da sua casa:” com uma escala de diferencial semântico de 1 a 4, sendo:

- 1 – O atributo não é importante;
- 2 – O atributo é pouco importante;
- 3 – O atributo é importante;
- 4 – O atributo é muito importante.

Os outros 25 alunos responderam à mesma pergunta em uma escala de diferencial semântico de 1 a 5, sendo:

- 1 – O atributo não tem nenhuma importância;
- 3 – O atributo é razoavelmente importante;
- 5 – O atributo é muito importante.

As categorias 2 e 4 não receberam denominações.

Dos 50 alunos, 18 responderam o questionário com escala de 1 a 5 e 21 responderam o questionário com escala de 1 a 4. O objetivo desta análise foi verificar se há diferença entre as duas escalas em relação à média e ao desvio-padrão da importância. Na construção de uma escala, é importante que as respostas não fiquem concentradas em apenas um ponto; se não houver variabilidade de respostas, não há como diferenciar o comportamento dos respondentes. Da mesma forma, é desejável que a média não fique muito próxima dos extremos da escala.

Para comparar as duas escalas, elas foram transformadas em escalas de 0 a 100 pontos, conforme mostra a Tabela 5.1.

Tabela 5.1 – Transformação das escalas em intervalo de 0 a 100

Escala de 4 pontos	escala original	escala transformada
Não é importante	1	0,00
É pouco importante	2	33,33
É importante	3	66,67
É muito importante	4	100,00

Escala de 5 pontos	escala original	escala transformada
Não tem nenhuma importância	1	0
	2	25
É razoavelmente importante	3	50
	4	75
É muito importante	5	100

As Tabelas 5.2 e 5.3 mostram que ambas escalas produziram o mesmo resultado em termos de média, variância e frequência de resultados. As médias das escalas transformadas não possuem diferença estatisticamente significativa (p-valor do teste t para diferença de médias: 0,975) e as variâncias amostrais obtidas são iguais. Outro aspecto importante é que nenhuma das duas escalas apresentou maior concentração em um ponto específico (14/21 na escala 1 a 4 e 12/18 na escala 1 a 5; o que representa 2/3 em ambos os casos).

Tabela 5.2 – Média e desvio-padrão das escalas originais e transformadas

	Escala 1 a 5 original	Escala 1 a 5 transformada	Escala 1 a 4 original	Escala 1 a 4 transformada
média	4,50	87,50	3,62	87,30
Desvio-padrão	0,79	19,65	0,59	19,65
Coeficiente de variação	0,18	0,23	0,16	0,23

Tabela 5.3 – Frequência de cada categoria nas duas escalas

	Frequência	Frequência
Categoria	Escala 1 a 5	Escala 1 a 4
1	0	0
2	0	1
3	3	6
4	3	14
5	12	
Total	18	21

Dada a pouca evidência de diferenciação entre as escalas, decidiu-se utilizar a escala de 1 a 4 pontos na análise. Esta decisão também se baseou no fato de que a escala de diferencial semântico está melhor definida nesta escala (todos os pontos têm descrição) enquanto que na escala de 5 pontos há apenas 3 pontos com descrição.

Outro fator considerado nesta decisão é que o modelo da TRI que será utilizado é um modelo que considera uma escala ordinal e cada ponto da escala se “desdobra” em dois pontos. Com 4 pontos, tem-se 7 parâmetros a estimar (limiar entre as categorias). Se fosse utilizada a escala de 5 pontos, haveria 9 parâmetros a estimar sem nenhum ganho.

Também pretende-se analisar a eventual adequação de um modelo dicotômico, apenas com as categorias importante/não importante, que pode ser facilmente obtido por meio de agrupamentos de categorias de resposta de uma escala de quatro pontos.

5.2.2. Desenho do questionário

O questionário com os 25 atributos é mostrado no Apêndice D. Foram aplicados 400 questionários em várias turmas do 1º ao 5º semestre do curso de Administração da EAESP, durante as suas aulas. Cada respondente atribuiu uma nota de 1 a 4, conforme escala descrita anteriormente, para cada atributo. Também foi perguntado se os respondentes freqüentavam alguma academia, e caso não freqüentassem, se tinham interesse.

Um aspecto importante no desenho do questionário é a ordenação dos atributos. Existe a possibilidade da ordem dos atributos influenciar o resultado. Para evitar isto, foram apresentadas questões em ordem distintas para cada aluno.

A permutação das questões foi desenhada através de um quadrado latino completo de tamanho 25, descrito no capítulo 4. No problema proposto, com 25 atributos, foram gerados 16 quadrados latinos, totalizando 400 vetores ordenados dos atributos.

Os questionários foram impressos um a um, respeitando as ordens geradas pelo quadrado latino. Foi feita uma macro no software Excel para gerar os 400 diferentes questionários.

A digitação dos dados foi feita na ordem apresentada no questionário. Como todos os questionários foram numerados, foi possível, através de funções (PROCV e PROCH) do Excel, reestruturar os dados de tal forma a obter um banco de dados em que todas as respostas a um determinado atributo estivessem na mesma coluna, independente da ordem em que foi respondida.

5.2.3. Coleta de dados

A coleta dos dados terminou em novembro de 2005. Alguns alunos devolveram o questionário em branco, e estes foram novamente aplicados aos alunos da classe seguinte.

Ao final da aplicação em 11 turmas, 363 questionários estavam completamente respondidos. Dos 37 restantes, 13 foram devolvidos com alguma questão em branco ou duas respostas para a mesma questão foram descartados; os 24 restantes foram devolvidos em branco e não foram redistribuídos aos alunos de outras turmas.

5.3. ANÁLISE DESCRITIVA

A tabela do apêndice E mostra a frequência de cada categoria de resposta, para cada atributo, por ordem decrescente de média. Alguns atributos (ter disponibilidade de equipamentos, equipamentos de qualidade, limpeza, professores atenciosos, estar perto de casa e ter professores qualificados) apresentam uma média próxima de 4, indicando quase unanimidade de respostas na categoria “muito importante”. Na maioria destes atributos, ninguém respondeu a categoria “não é importante”.

O Gráfico 5.1 mostra a média e o desvio-padrão das respostas para cada respondente. A média observada (2,92) mostra que houve uma tendência dos entrevistados em considerar os atributos importantes, gerando uma baixa discriminação entre eles.

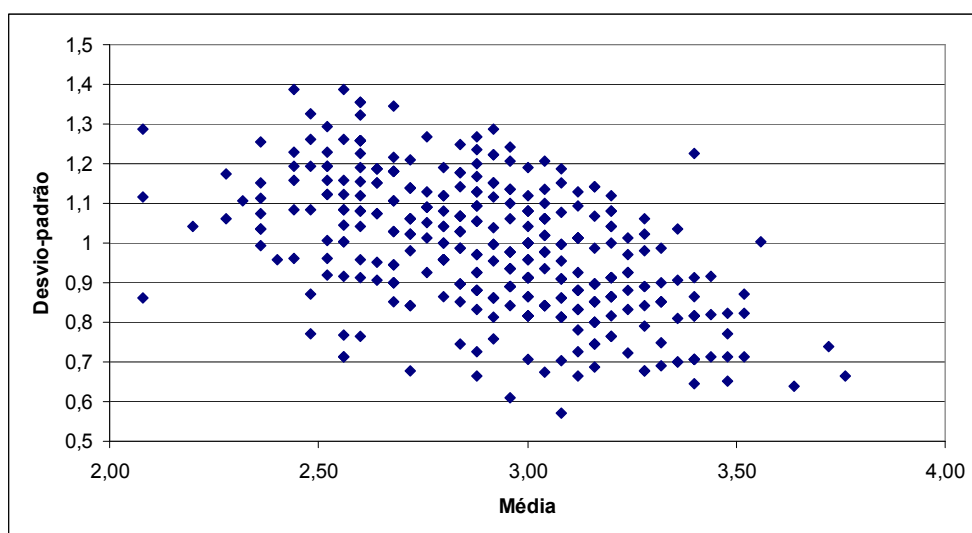


Gráfico 5.1 – Relação entre a média e desvio-padrão das respostas por indivíduo

À pergunta “Você frequenta alguma academia de ginástica?”, 140 disseram sim, 223 disseram não e 23 não responderam.

Dos que não frequentam, 52% tem interesse em frequentar, conforme mostra a Tabela 5.4.

Tabela 5.4 – Interesse em frequentar uma academia, dentre alunos que não frequentam.

Tem interesse em frequentar?	Frequência
1: definitivamente não	14
2	20
3: um pouco	68
4	50
5: definitivamente sim	66
não respondeu	5
Total	223

5.3.1. Influência da ordem de apresentação das questões na resposta

Devido ao formato de apresentação do questionário, em que a ordem das questões foi arranjada segundo um quadrado latino, foi possível verificar a influência da ordem de apresentação das questões. Isto foi possível pois todas as questões foram apresentadas em todas as possíveis ordens, o mesmo número de vezes.

Para verificar se a ordem em que a questão foi apresentada influenciou a sua resposta, foram criados dois grupos para cada item do questionário. O grupo 1 foi composto por observações do atributo i quando este era apresentado no começo do questionário (entre os 13 primeiros atributos). O grupo 2 foi composto por observações do atributo i quando este era apresentado no final do questionário (entre os 12 últimos atributos).

Foi feito um teste t-student e não foi encontrada diferença estatisticamente significativa entre as médias das respostas dos atributos nos dois grupos, exceto para o item “ter bom atendimento”, com p-valor igual a 1%. Mesmo quando as médias foram ipsatizadas (as respostas dos respondentes foram padronizadas na linha para que todos os respondentes tivessem a mesma média), não houve diferença estatisticamente significativa com 5% de probabilidade de erro do tipo I exceto, novamente, no item “ter bom atendimento”, agora com p-valor igual a 0,4%.

Foi feita uma outra análise considerando os dois grupos seguintes: o grupo 1 foi composto por observações do atributo i quando este era apresentado no começo do questionário (entre os 5

primeiros atributos). O grupo 2 foi composto por observações do atributo i quando este era apresentado no final do questionário (entre os 5 últimos atributos).

Os resultados foram similares aos anteriores. Além do atributo “ter bom atendimento”, que apresentou p-valor próximo de 1% em ambos os casos (com dados originais e ipsatizados), o atributo “ter serviço de personal trainer” também apresentou diferença estatisticamente significativa entre os dois grupos (p-valor=3% e 1%, com dados originais e ipsatizados, respectivamente).

Estas significâncias, devido à quantidade de testes (25), podem ter ocorrido ao acaso, uma vez que se espera cometer probabilidade de erro do tipo 1 em 5% das vezes. Espera-se que, a cada 20 testes, um apresente significância estatística mesmo se não houver diferença entre as médias populacionais.

Este resultado é desejado por dois motivos. Primeiro, porque o modelo da TRI assume que as respostas dadas aos itens são independentes. Segundo, porque se houvesse diferença entre as médias, os dados do modelo completo não serviriam para comparar com o modelo com dados incompletos.

5.3.2. Análise fatorial dos itens

A análise fatorial feita pelo método de componentes principais com os dados mostrou que o primeiro fator extraído corresponde a 16,2% da variância total dos dados e o segundo fator, 7,8%. O Scree plot (Gráfico 5.2) mostra o autovalor de cada um dos fatores.

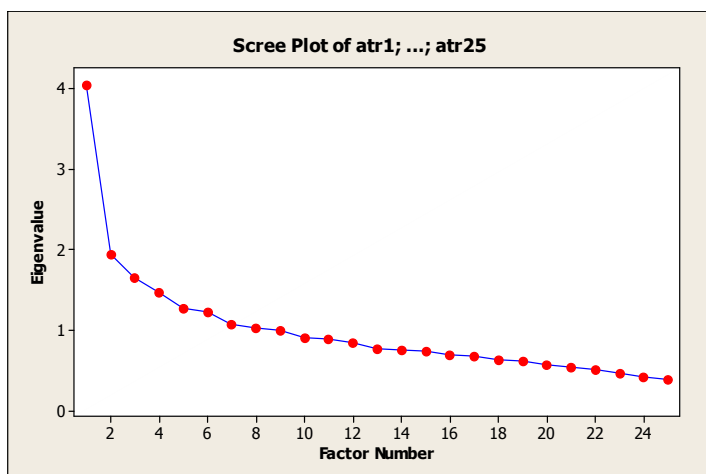


Gráfico 5.2 – Scree plot da análise fatorial

RECKASE (1979) sugere que a unidimensionalidade pode ser aceita quando o autovalor do primeiro fator for maior ou igual a 20%. Embora o valor observado esteja abaixo de 20%, será verificado o ajuste do modelo GGUM a estes dados. Para se ter uma idéia, foi feita uma análise fatorial com os dados provenientes da simulação apresentada no capítulo 4, com dados gerados a partir do GGUM unidimensional com $N=1000$ e $I=25$. O primeiro componente extraiu 25% da variância total dos dados e o segundo, 10%.

Quando se excluem os atributos “estar localizado perto de casa”, “ficar aberta 24h”, “abrir sábado e domingo” e “ter pessoas do mesmo nível sócio-econômico que o meu”, que são os itens com menor comunalidade nos dois primeiros fatores, o primeiro fator da análise fatorial com os itens remanescentes corresponde por 20% da variância total dos dados, e o segundo, 10%.

Inicialmente, será ajustado o modelo a todos os dados, sem exclusão de nenhum item. A exclusão de itens será discutida no final do capítulo.

5.4. MODELO SIMPLIFICADO - RESULTADOS

Inicialmente será apresentado, por questões de facilidade de interpretação, o GGUM com apenas duas categorias de resposta: “importante” e “não importante”.

Para aplicar este modelo aos dados, as quatro categorias de resposta originais serão agrupadas em duas.

5.4.1. Agrupamento das categorias de resposta observadas

Para obter esta simplificação, as quatro categorias serão agrupadas conforme mostra a Tabela 5.5:

Tabela 5.5 – Agrupamento das categorias de resposta

Categorias originais	Agrupamento
1 - Não é importante	0 - Não é importante
2 - É pouco importante	
3 - É importante	1 - É importante
4 - É muito importante	

A distribuição de frequência das categorias agrupadas para cada item é dada no Apêndice F.

O modelo GGUM fornecerá estimativa da probabilidade de cada item ser considerado “importante” e “não importante” para cada indivíduo. Como a soma destas probabilidades é 1, basta conhecer uma delas. A outra é obtida pelo complementar. A probabilidade do item i ser considerado importante pelo indivíduo j variará de acordo com os parâmetros do indivíduo (θ_j) e do item (δ_i , α_i , τ_{i1} , τ_{i2} , e τ_{i3}). Será utilizado o modelo GGUM sem restrições adicionais aos parâmetros (modelo 8).

5.4.2. Estimativa dos parâmetros

A Tabela 5.6 mostra as estimativas dos parâmetros de cada item por ordem crescente de δ_i .

Tabela 5.6 - Estimativas dos parâmetros de cada item por ordem crescente de δ_i .

item	Descrição	δ_i	α_i	τ_{i1}	$\delta_i - \tau_{i1}$
12	Ter aulas de lutas marciais	-8,77	0,25	-4,59	-4,18
9	Ter aulas de dança	-6,62	0,78	-5,36	-1,26
13	Abrir domingos e feriados	-6,22	0,29	-6,74	0,52
14	Ficar aberta 24h por dia	-6,01	0,42	-5,34	-0,67
21	Ter amigos ou conhecidos que freqüentam a academia	-5,94	0,31	-5,89	-0,05
11	Ter quadra poliesportiva	-5,79	1,11	-4,80	-0,99
18	Ter várias unidades (endereço que você pode freqüentar)	-5,59	0,85	-4,69	-0,90
10	Ter piscina aquecida	-4,80	0,98	-4,69	-0,11
17	Ter lanchonete	-4,79	1,38	-4,85	0,06
23	Ter orientação nutricional	-4,49	1,30	-4,95	0,46
19	Ter bom preço	-4,34	0,46	-10,9	6,56
5	Ter equipamentos de qualidade	-4,21	2,22	-7,86	3,65
15	Ter serviço de personal trainer	-4,14	0,89	-4,15	0,01
4	Ter segurança	-3,43	1,46	-5,79	2,36
7	Ter professores qualificados	-2,98	2,07	-5,52	2,54
8	Ter professores que dão atenção aos alunos	-2,56	1,39	-5,56	3,00
25	Ter avaliação física	-2,18	0,79	-4,60	2,42
24	Ter bom atendimento	-0,96	1,33	-3,51	2,55
1	Ter bastante espaço físico	-0,92	0,94	-2,99	2,07
2	Ter estacionamento gratuito	-0,89	1,18	-2,25	1,36
20	Ter pessoas do mesmo nível sócio-econômico que o meu	-0,56	0,45	0,46	-1,02
3	Ter limpeza	-0,27	1,17	-4,44	4,17
22	Ser freqüentada por gente bonita	-0,02	0,34	-0,32	0,30
6	Ter disponibilidade de equipamentos	6,20	0,40	-19,16	25,36
16	Estar localizada perto de casa	6,66	0,26	-19,36	26,02

A distribuição dos parâmetros dos indivíduos é mostrada na Figura 5.1. Conforme especificado, θ segue uma distribuição normal com média 0. Praticamente todos os θ_j estimados estão entre -3 e 3.

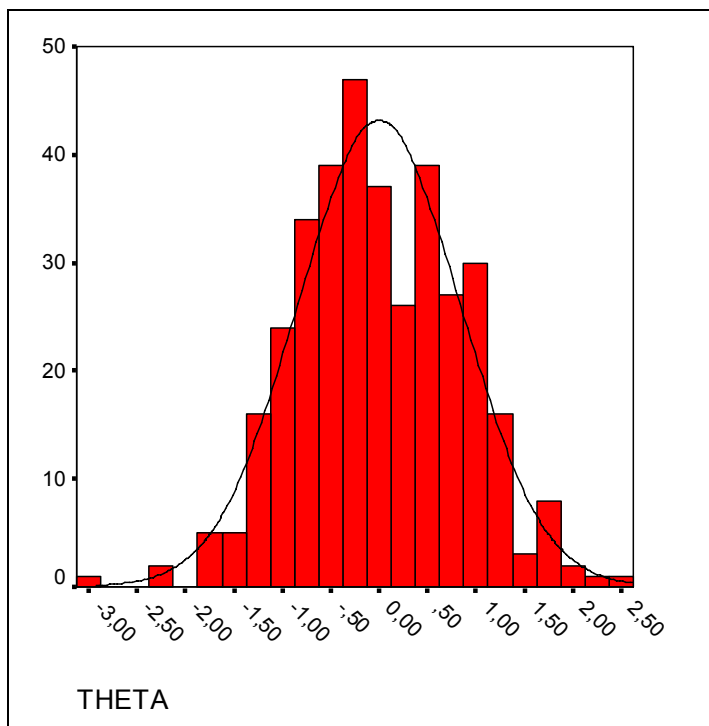


Figura 5.1 – Histograma de θ .

5.4.3. Ajuste do modelo

A estatística G^2 apresentou p-valor igual a 0,72, indicando um bom ajuste do modelo. Os p-valores das estatísticas de infit e outfit em sua maioria estão acima de 0,01. Apenas 3 itens (3: “ter limpeza”, 5: “ter equipamentos de qualidade” e 7: “ter professores qualificados”) apresentam p-valor abaixo de 1% na estatística infit. Estas três estatísticas, para cada item, podem ser encontradas no apêndice G.

5.4.4. Interpretação do resultado

A combinação das informações dos parâmetros dos itens com o parâmetro do indivíduo permite analisar o comportamento das respostas. O item 16 (“academia estar localizada perto de casa”), por exemplo, tem alta probabilidade de ser considerado importante por todos os

indivíduos da amostra. Isto ocorre sempre que $|\hat{\theta}_j - \hat{\delta}_i| = |\hat{\theta}_j - 6,66| < 19,36$. Ou, equivalentemente, $-12,70 < \hat{\theta}_j < 26,02$. Como todos θ_j estimados são menores que 3 em módulo, esta desigualdade é válida para todos os indivíduos. O Gráfico 5.3 mostra a probabilidade de cada categoria, dado θ_j , para este atributo.

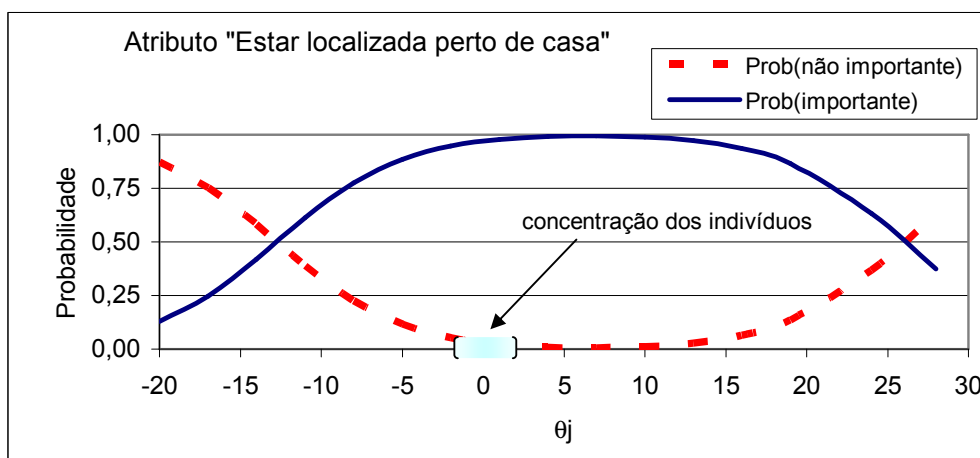


Gráfico 5.3 – Curvas de probabilidade do atributo “Estar localizada perto de casa”.

O parâmetro α_i informa se a categoria de resposta “importante” tem grande dominância sobre a outra, isto é, se a probabilidade de resposta “importante” é alta quando $\theta_j = \delta_i$ e baixa quando θ_j está distante de δ_i . No caso deste item, $\hat{\alpha}_i = 0,26$, o que mostra que o atributo não é informativo sobre o comportamento do indivíduo.

Como pode ser observado, o modelo atribui a todos os indivíduos alta probabilidade de considerar este item importante. Este resultado é coerente com o observado, em que 97% dos entrevistados consideraram este item importante.

Uma análise mais interessante pode ser feita com itens que não foram considerados importantes por todos os respondentes, como “ter lanchonete”, que obteve 51% de resposta na categoria “importante”. O Gráfico 5.4 mostra a curva de probabilidades de resposta para este item.

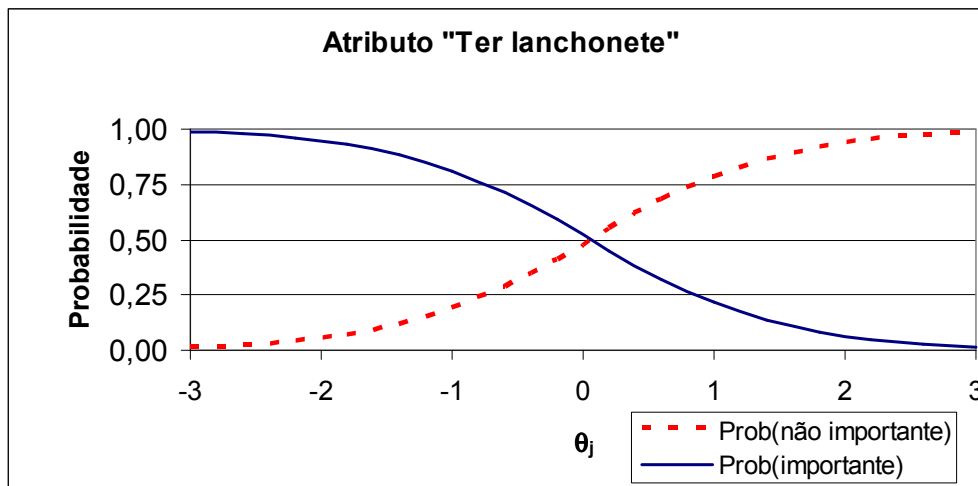


Gráfico 5.4 – Curvas de resposta do atributo “ter lanchonete”, para θ_j variando entre -3 e 3.

Pelo gráfico, nota-se que indivíduos com $|\hat{\theta}_j - \hat{\delta}_i| = |\hat{\theta}_j + 4,79| < 4,85$ ou, simplificando, $-9,64 < \hat{\theta}_j < 0,08$ têm probabilidade de escolha da categoria “importante” maior que 0,50, isto é, os respondentes com $\hat{\theta}_j$ negativo tem alta probabilidade de considerar este atributo importante.

Por fim, será analisado um atributo considerado sem importância pela maioria dos entrevistados – “ter aulas de dança”. O Gráfico 5.5 mostra a curva de resposta deste atributo.

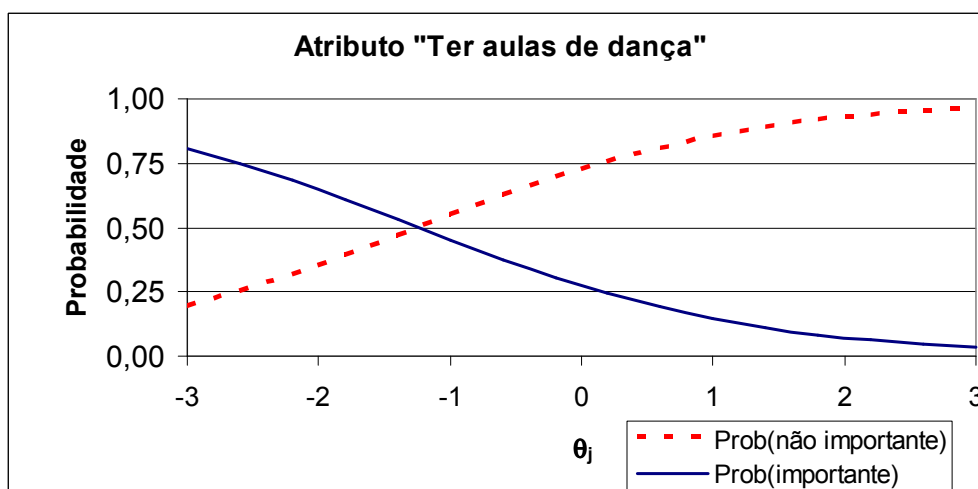


Gráfico 5.5 – Curvas de resposta do atributo “ter aulas de dança”, para θ_j variando entre -3 e 3.

Como $\hat{\delta}_i = -6,62$ e $\hat{\tau}_{i1} = -5,36$, indivíduos com $-11,98 < \hat{\theta}_j < -1,26$ terão alta probabilidade ($>0,50$) de considerar este atributo importante.

Há um caso interessante em que nenhum indivíduo tem probabilidade de resposta “importante” maior que 0,50 para um atributo. Isto ocorre quando $\tau_{i1} > 0$, que é o caso do item 20 – “ter pessoas do mesmo nível sócio-econômico que o meu”. Não existe valor de θ_j tal que $|\theta_j - \delta_i| < -\tau_{i1}$ neste caso. O Gráfico 5.6 ilustra a curva de probabilidade do item 20, cujo α_i é igual a 0,46.

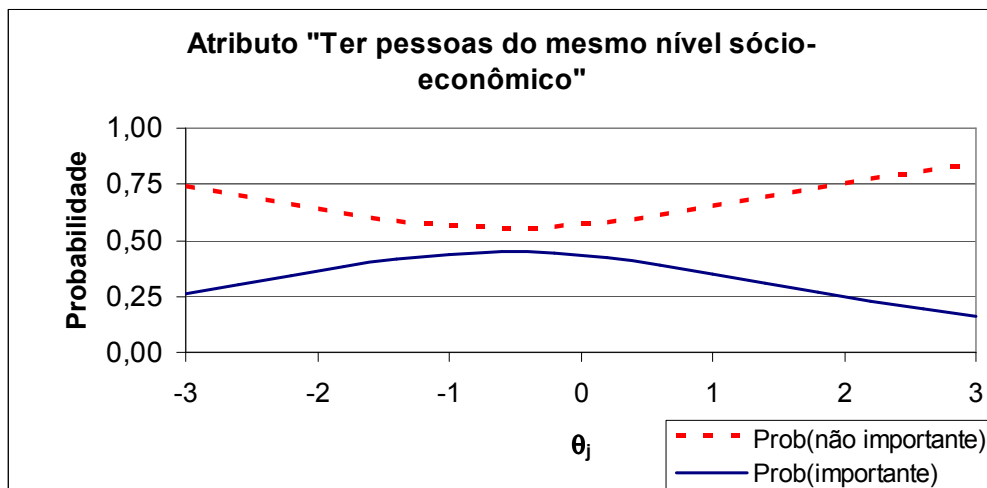


Gráfico 5.6 – Curvas de resposta do atributo “ter pessoas do mesmo nível sócio-econômico que o meu”, para θ_j variando entre -3 e 3.

O Gráfico 5.7 mostra a probabilidade de resposta na categoria “importante” para valores de θ variando entre -3 e 3. Apesar da visualização individual ser difícil, nota-se que indivíduos com θ_j próximo de -1 são os que tendem a atribuir maior importância aos itens, enquanto que os indivíduos com θ_j próximo de 3 atribuem importância a apenas alguns atributos. Os itens informativos são aqueles que têm grande variação de probabilidade da categoria “importante” neste intervalo de θ_j .

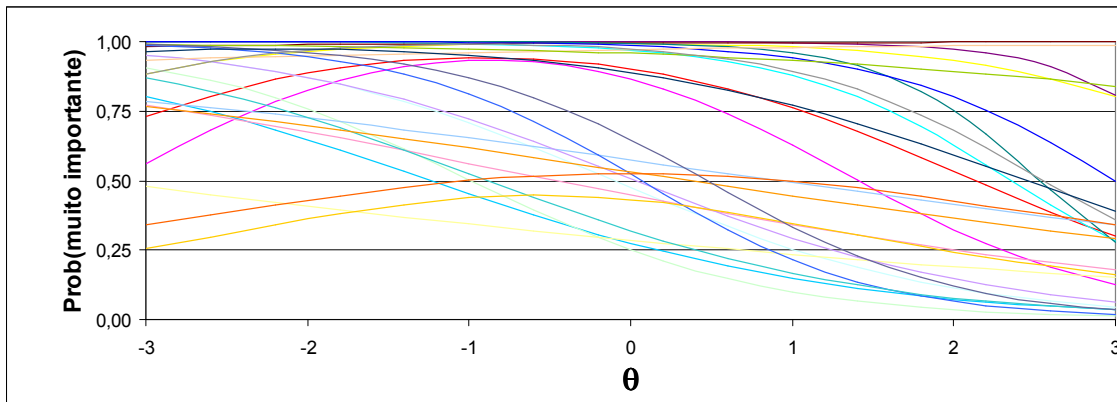


Gráfico 5.7 – Curvas de resposta para a categoria “muito importante”, para todos os atributos e para θ variando entre -3 e 3.

Como visto anteriormente, os itens informativos são aqueles que têm α_i grande e τ_{i1} próximo de δ_i . Dos 25 itens estudados, 10 têm $\hat{\alpha}_i$ maior que 1 e outros 10 têm $\hat{\tau}_{i1} - \hat{\delta}_i$ próximo de zero. Alguns itens são informativos apenas para alguns valores de θ_j .

Uma análise mais detalhada dos resultados será feita para o modelo com quatro categorias, apresentado a seguir.

5.5. MODELO COM 4 CATEGORIAS - RESULTADO

A seguir, é feita a análise do modelo com as 4 categorias originais de resposta.

5.5.1. Ajuste do modelo

Dos 25 itens, 6 não possuem resposta em pelo menos uma categoria (ver tabela no Apêndice E). Este problema também ocorreu na simulação, e isto implica que o modelo 8, que necessita todas as categorias com pelo menos uma resposta para estimar todos os τ_{ik} 's. Foi, portanto, utilizado o modelo 6, que impõe a restrição que os τ_{ik} 's são equiespaçados para todos os itens.

Os valores das estatísticas de *infit* e *outfit* para os 25 itens se mostraram não significantes ($p\text{-valor} > 1\%$), indicando um bom ajuste do modelo (ver apêndice H). Três itens tiveram $p\text{-valor}$ da estatística G^2 abaixo de 5% mas nenhum abaixo de 1%, que é o nível de significância recomendado (ROBERTS et al, 2004). A estatística G^2 do modelo no geral apresentou um $p\text{-valor}$ igual a 21%, indicando um bom ajuste do modelo aos dados.

Para avaliar a capacidade de previsão do modelo, foram utilizadas as seguintes alternativas para calcular o valor estimado da resposta do indivíduo j para o item i :

- A. utilizando a resposta média do item i dada por todos os indivíduos;
- B. utilizando a resposta média geral;
- C. utilizando o GGUM, calcular o valor com maior probabilidade de escolha;
- D. utilizando o GGUM, calcular o valor esperado da resposta, pela multiplicação de cada valor pela sua probabilidade;
- E. utilizando o GGUM, calcular a probabilidade de resposta nas categorias “Importante” e “Muito importante”.

O Apêndice H mostra o QMErro para cada atributo e para os dados gerais, considerando a resposta e a previsão reescaladas. Os resultados mostram que o QMErro com as alternativas D e E são menores, o que é coerente com o resultado obtido na simulação.

Com a aplicação do GGUM e a previsão feita pelo valor esperado (alternativa B), o QMErro da previsão é 16% menor do que utilizando a média do item (0,061 contra 0,072). A raiz do QMErro, que representa o desvio-padrão do erro, tem uma diminuição de 29%.

Quando a previsão é feita pela probabilidade da categoria “importante” ou “muito importante” (alternativa E), a raiz do QMErro de previsão é 15% menor do que quando se utiliza a média do item.

5.5.2. Interpretação do resultado

O Gráfico 5.8 mostra que os δ_i 's (deltas) estimados, em sua maioria, são negativos, e variam de -27 a +9. Os parâmetros θ_j dos indivíduos, por outro lado, estão localizados entre -3 e +3, com média zero. Os outros parâmetros dos itens são mostrados no apêndice I.

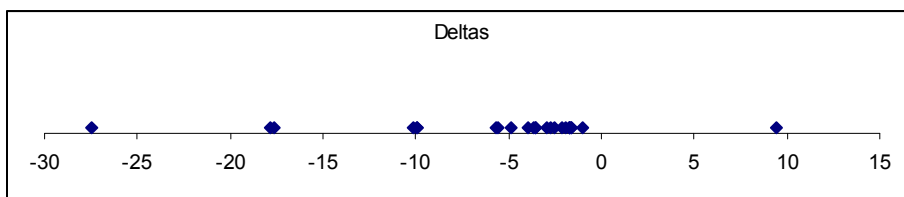


Gráfico 5.8 – Deltas (δ_i 's) estimados

A consequência desta distribuição é que, quanto maior o parâmetro θ_j do indivíduo, menor a probabilidade dele considerar a maioria dos atributos muito importantes, pois a distância $\theta_j - \delta_i$ é maior do que um indivíduo com θ_j com negativo. Com exceção do item 16 – “academia estar localizada perto de casa”, cujo $\hat{\delta}_i = 9,45$.

Espera-se, portanto, que, quanto maior θ_j , menor a resposta média do entrevistado, o que o torna mais exigente em relação à importância dos atributos. Isto pode ser visualizado no Gráfico 5.9.

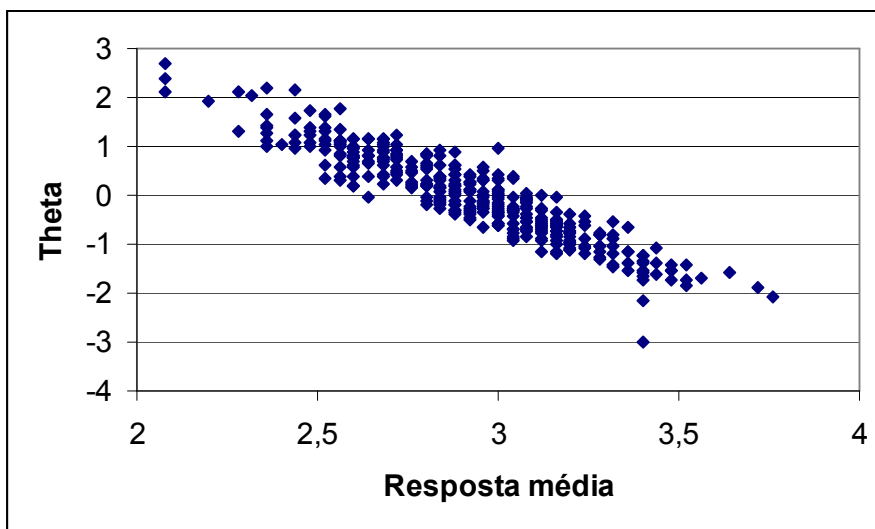


Gráfico 5.9 – Valores de θ_j em função da média das respostas dada pelo entrevistado.

O Gráfico 5.10 mostra a probabilidade de um indivíduo com traço latente θ considerar o atributo i muito importante. Foram escolhidos 5 atributos para serem ilustrados (1, 5, 9, 12 e 16, descritos no Gráfico). No eixo horizontal, tem-se o valor do parâmetro do indivíduo, θ (Theta).

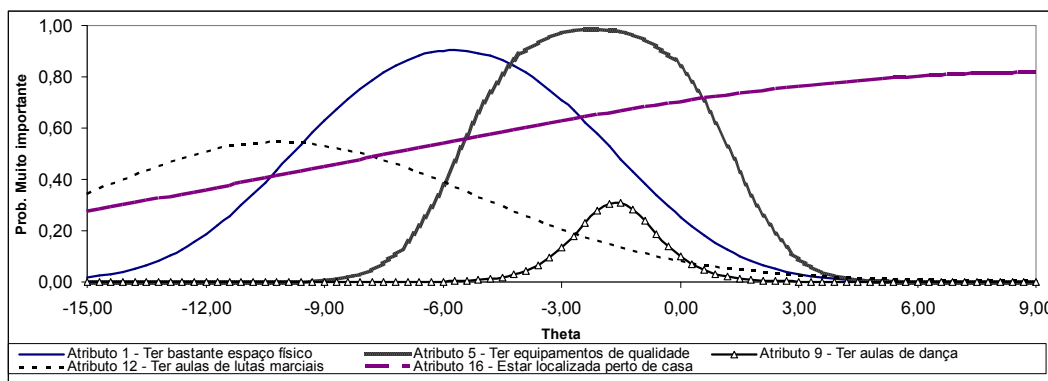


Gráfico 5.10 – Curvas de resposta para a categoria “muito importante” para 5 atributos.

O Gráfico 5.10 mostra que, para o atributo 5 (“ter equipamentos de qualidade”) e para o atributo 9 (“ter aulas de dança”), a probabilidade de um indivíduo considerá-los importante é maior nos intervalos em que θ_j está próximo de $\hat{\delta}_i$ (-2,17 e -1,63, respectivamente).

O atributo 16 – “estar localizado perto de casa” – não tem muita variação na probabilidade de escolha da categoria “muito importante” para indivíduos com θ_j entre -3 e 3. Isto ocorre porque o parâmetro $\hat{\delta}_{16}$ é 9,45 e $\hat{\alpha}_{16}$ é próximo de zero.

Os atributos 12 (“ter aulas de lutas marciais”) e 1 (“ter bastante espaço físico”), por possuírem deltas estimados distantes de zero (-10,1 e -5,72, respectivamente), apresentam baixa probabilidade de escolha da categoria “muito importante” para indivíduos com θ_j entre -3 e 3.

Os gráficos do Apêndice J mostram o escore esperado de cada item (e também o valor observado médio) em função do valor de θ_j . Além de ser outra forma de ver o ajuste do modelo (o valor esperado deve estar próximo do observado médio), este gráfico ajuda a dar uma interpretação ao *continuum* latente. Se houver muita diferença no escore esperado para um item, este é bastante informativo para definição do traço latente. A maioria dos itens têm escore médio maior para valores de θ_j negativo, exceto o item 16 (“estar localizado perto de

casa”). Portanto, os indivíduos com θ_j negativo têm maior propensão a considerar os itens importantes. Estes indivíduos tem maior probabilidade de considerar os atributos 9, 10, 11, 15, 17 e 24 (descritos na Tabela 5.7) muito importantes em comparação aos indivíduos com θ_j positivo. Os indivíduos com θ_j positivo tem baixa probabilidade de considerar estes atributos positivos, o que pode ser visto pelo baixo escore esperado destes itens neste caso.

Para ilustrar o resultado, são comparados a seguir o escore esperado para cada item, para dois indivíduos com valor do traço latente $\theta_1=-2$ e $\theta_2=2$. A Tabela 5.7 mostra a diferença do escore esperado de importância para os dois indivíduos hipotéticos. Percebe-se que os itens mencionados no parágrafo anterior são os que têm maior diferença no escore esperado e os três itens com menor $\hat{\alpha}_i$ são os que têm menor diferença dos escores esperados.

Tabela 5.7 – Diferença entre os escores esperados de importância segundo θ_j .

Atributo	Item	Escore esperado dado $\theta=-2$	Escore esperado dado $\theta=2$	Diferença dos escores esperados	$\hat{\alpha}_i$
Ter orientação nutricional	23	2,73	0,66	2,07	0,998
Ter lanchonete	17	2,40	0,60	1,80	0,689
Ter piscina aquecida	10	2,29	0,62	1,67	0,514
Ter serviço de personal trainer	15	2,36	0,69	1,67	0,526
Ter quadra poliesportiva	11	2,00	0,42	1,58	0,498
Ter aulas de dança	9	1,65	0,25	1,40	0,508
Ter avaliação física	25	2,81	1,49	1,32	0,787
Ter várias unidades (endereço que você pode frequentar)	18	1,88	0,59	1,28	0,438
Ter segurança	4	2,89	1,80	1,09	0,861
Ter bom atendimento	24	2,86	1,91	0,95	0,730
Ter bastante espaço físico	1	2,53	1,62	0,91	0,548
Ter professores qualificados	7	2,93	2,04	0,89	1,061
Ter estacionamento gratuito	2	2,40	1,52	0,88	0,320
Ter professores que dão atenção aos alunos	8	2,91	2,07	0,84	0,956
Ter limpeza	3	2,95	2,15	0,80	1,078
Abrir domingos e feriados	13	2,08	1,32	0,76	0,168
Ter equipamentos de qualidade	5	2,99	2,26	0,73	1,260
Ter amigos ou conhecidos que frequentam a academia	21	1,88	1,17	0,71	0,242
Ficar aberta 24h por dia	14	1,75	1,08	0,67	0,191
Ter aulas de lutas marciais	12	1,33	0,75	0,59	0,159
Ter pessoas do mesmo nível sócio- econômico que o meu	20	1,52	1,07	0,45	0,178
Ter disponibilidade de equipamentos	6	2,84	2,46	0,38	0,697
Ser frequentada por gente bonita	22	1,61	1,36	0,25	0,074
Ter bom preço	19	2,56	2,40	0,15	0,093
Estar localizada perto de casa	16	2,61	2,72	-0,11	0,091

5.5.3. Resultado com respostas excluídas por respondente

O resultado da simulação no Capítulo 4 mostrou que o QMErro da previsão de resposta pelo modelo GGUM utilizando o valor esperado não tem aumento significativo quando há 40% ou 60% de respostas excluídas por respondente.

Para verificar se este comportamento ocorre com os dados coletados em campo, foram excluídas 40% e em seguida 60% das respostas de cada respondente utilizando o mesmo critério estabelecido nas simulações: mantêm-se as k primeiras respostas de cada respondente, que são determinadas pelos quadrados latinos de tamanho 25 sorteados.

Ambos os modelos apresentaram bom ajuste, pois o p-valor G^2 foi maior que 5% nos dois casos e não houve nenhum item que apresentasse estatística de ajuste muito ruim (ver Apêndices K e L).

O QMErro obtido utilizando-se o valor esperado calculado pelo GGUM quando há 60% de respostas excluídas foi 0,063, contra 0,061 obtido quando os dados estão completos. Isto confirma o resultado obtido na simulação, que aponta pouco aumento no QMErro de previsão quando grande quantidade de dados é excluída.

Embora o QMErro da previsão não tenha aumentado, as estimativas dos parâmetros dos itens sofreram bastante modificações, o que também corrobora o resultado da simulação.

Em relação aos parâmetros dos atributos, a correlação entre $\hat{\theta}_j$ obtido com os dados completos e $\hat{\theta}_j$ obtido com os 10 respostas excluídas por respondente é 0,83. Quando o número de respostas excluídas é 15, a correlação diminui para 0,61. Isto sugere que os parâmetros dos indivíduos são estimados com boa precisão quando há 15 itens respondidos por indivíduo, mas quando este número cai para 10, a precisão é apenas razoável.

5.5.4. Exclusão de atributos

A exclusão dos itens pode ser feita antes do ajuste, pelo critério de itens com baixa comunalidade na análise de componentes principais com os dois primeiros componentes. Caso a comunalidade seja menor que 30%, este item deve ser excluído (RECKASE, 1979). Esta abordagem é muito utilizada para os modelos da TRI cumulativos, mas não há evidências de que este método é eficiente para o GGUM. A exclusão dos itens com menor comunalidade não resultou em um modelo com melhor ajuste.

Uma outra maneira de eliminar itens é através da curva de informação do item. Itens com baixa informação (ver Apêndice M) foram eliminados. A informação do item é muito correlacionada com o parâmetro α_i ; quanto menor α_i , menor a informação. Por simplicidade, serão excluídos os itens com menor valor estimado deste parâmetro. Os três atributos com menor valor de $\hat{\alpha}_i$ são: 16 – “estar localizada perto de casa”; 12 – “ter aulas de artes marciais” e 13 – “abrir domingos e feriados”.

A exclusão destes itens gerou um melhor ajuste (p-valor da estatística $G^2=0,34$, contra 0,21 do modelo com 25 itens). O QMErro de previsão diminuiu, utilizando a alternativa D, de 0,061 para 0,058.

É importante salientar que os itens excluídos são aqueles que não apresentam diferenças na resposta dado o valor de θ_j .

5.6. RELAÇÃO DO TRAÇO LATENTE COM OUTRAS VARIÁVEIS

Embora no questionário não houvesse nenhuma variável para segmentar os indivíduos, o exemplo seguinte mostra que o valor de θ_j está relacionado com a frequência e interesse dos respondentes em academias de ginástica (p-valor do teste F da ANOVA para diferença de médias: 1,1%). A Tabela 5.8 mostra o valor médio de θ_j para cada grupo.

Tabela 5.8 - θ_j médio para cada grupo de indivíduos.

Grupo	Média	Tamanho do grupo
Frequentam academia	0,175	140
Não frequentam, mas têm interesse	-0,141	116
Não frequentam e não têm interesse	-0,073	107

A Tabela 5.9 mostra que as pessoas que não frequentam academia, mas têm interesse, atribuem maior importância aos atributos “ter aulas de dança”, “ter piscina aquecida”, “ter quadra poliesportiva”, “ter serviço de personal trainer”, “ter lanchonete” e “ter orientação nutricional”. Isto se deve ao fato do grupo das pessoas que não frequentam academia mas têm interesse terem um θ_j médio negativo.

Para verificar este fato, foram feitas seis ANOVAs para comparar suas médias em cada grupo. Nas seis variáveis, houve significância estatística ($p\text{-valor} < 7\%$ em todos) e o grupo com menor média foi o grupo de pessoas que frequentam academia.

Uma das vantagens em se trabalhar com um traço latente interpretável é que é possível utilizá-lo em um modelo explicativo, em que outras variáveis são utilizadas para explicar o valor do traço latente para o indivíduo (BOECK, 2004).

5.7. COMPARAÇÃO ENTRE O MODELO DICOTÔMICO E O MODELO COM QUATRO CATEGORIAS

Os resultados obtidos neste trabalho sugerem que esta simplificação pode ser feita, pois os valores estimados para θ_j no modelo dicotômico e no modelo com 4 categorias tem 84% de correlação (foram comparados os modelos sem exclusão de respostas).

Para comparar se os valores previstos para a importância de cada item são iguais, será utilizada a alternativa E. No caso do modelo com 4 categorias, a importância é dada pela probabilidade do indivíduo escolher as categorias “importante” ou “muito importante”. No modelo dicotômico, a importância é dada pela probabilidade do indivíduo escolher a categoria

“importante”. A correlação entre as duas previsões foi de 97,3%. Isto significa que, se o questionário for simplificado de modo a ter apenas duas categorias de resposta, não há perda de informação sobre a importância do item.

5.8. A MEDIDA DE IMPORTÂNCIA COMO PESO PARA DETERMINAÇÃO DO ESCORE GLOBAL

Quando o objetivo da mensuração da importância é atribuir peso às notas dadas pelo indivíduo a cada atributo, é desejável que a soma destes pesos seja 1 para que o escore global seja uma “média ponderada” das notas dadas aos atributos.

Neste caso, a importância “normalizada” que o indivíduo j atribui ao item i é dada por:

$$w_{ij}^* = \frac{w_{ij}}{\sum_{i=1}^I w_{ij}} \quad (5.1)$$

O escore global do indivíduo j é dado por

$$U_j = \sum_{i=1}^I w_{ij}^* x_{ij} \quad (5.2)$$

Em que w_{ij} é a importância que o indivíduo j deu ao item i .

Se w_{ij} não for conhecida, ela pode ser estimada por qualquer um dos métodos mencionados anteriormente.

Para exemplificar, serão utilizados os pesos que o indivíduo 1 atribuiu ao item 1 na Tabela 5.9, bem como as estimativas geradas pelo GGUM. Esta tabela também contém avaliações hipotéticas que este indivíduo fez de cada atributo para uma academia, possibilitando o cálculo do escore global tanto pela média geral da avaliação, quanto pela média ponderada pelos pesos observados e estimados pelo GGUM nos casos em que há dados faltantes ou não.

Tabela 5.9 – Avaliação e Importância observada e estimada em duas situações: com dados completos e com 15 respostas excluídas, ambas pelo GGUM.

Atributo	avaliação do item (X _{1i})	COM MATRIZ COMPLETA			COM MATRIZ COMPLETA			COM MATRIZ INCOMPLETA		
		Importância observada (wi1)	Importância observada normalizada (wi1*)	(wi1*) x X _{1i}	importância estimada pelo GGUM (valor esperado) (wi1)	importância estimada normalizada (wi1*)	(wi1*) x X _{1i}	estimada pelo GGUM (valor esperado) (wi1)	estimada normalizada (wi1*)	(wi1*) x X _{1i}
1	7	2	4,4%	0,31	2,02	4,4%	0,31	2,152	4,4%	0,31
2	8	2	4,4%	0,36	2,09	4,5%	0,36	2,063	4,2%	0,34
3	9	3	6,7%	0,60	2,70	5,8%	0,53	2,821	5,8%	0,52
4	6	3	6,7%	0,40	2,39	5,2%	0,31	2,374	4,8%	0,29
5	7	2	4,4%	0,31	2,77	6,0%	0,42	2,900	5,9%	0,41
6	6	3	6,7%	0,40	2,81	6,1%	0,36	2,837	5,8%	0,35
7	8	2	4,4%	0,36	2,64	5,7%	0,46	2,895	5,9%	0,47
8	9	3	6,7%	0,60	2,64	5,7%	0,52	2,962	6,1%	0,54
9	5	3	6,7%	0,33	0,72	1,6%	0,08	0,920	1,9%	0,09
10	5	1	2,2%	0,11	1,24	2,7%	0,13	1,386	2,8%	0,14
11	4	1	2,2%	0,09	0,90	2,0%	0,08	1,017	2,1%	0,08
12	6	1	2,2%	0,13	0,97	2,1%	0,13	0,969	2,0%	0,12
13	7	2	4,4%	0,31	1,66	3,6%	0,25	1,672	3,4%	0,24
14	8	2	4,4%	0,36	1,34	2,9%	0,23	1,409	2,9%	0,23
15	4	1	2,2%	0,09	1,34	2,9%	0,12	1,633	3,3%	0,13
16	8	3	6,7%	0,53	2,68	5,8%	0,46	2,632	5,4%	0,43
17	5	1	2,2%	0,11	1,29	2,8%	0,14	1,395	2,9%	0,14
18	7	1	2,2%	0,16	1,05	2,3%	0,16	1,187	2,4%	0,17
19	8	2	4,4%	0,36	2,47	5,4%	0,43	2,509	5,1%	0,41
20	6	1	2,2%	0,13	1,26	2,7%	0,16	1,325	2,7%	0,16
21	8	0	0,0%	0,00	1,46	3,2%	0,25	1,514	3,1%	0,25
22	5	0	0,0%	0,00	1,46	3,2%	0,16	1,577	3,2%	0,16
23	7	1	2,2%	0,16	1,58	3,4%	0,24	1,840	3,8%	0,26
24	10	3	6,7%	0,67	2,41	5,2%	0,52	2,418	4,9%	0,49
25	6	2	4,4%	0,27	2,27	4,9%	0,30	2,543	5,2%	0,31
Soma		45	100,0%		46,2	100,0%		49,0	100,0%	
Média	6,76			7,13		7,10				7,07

Os valores do escore global ficaram muito próximos quando se utilizou o valor esperado (quando há dados excluídos e quando não há) e observado da importância como peso. O escore total sem ponderação pela importância apresentou diferença em relação ao escore ponderado pois este respondente deu maior peso aos atributos melhor avaliados.

6 PLANEJAMENTO DE UMA PESQUISA

6.1. INTRODUÇÃO

Neste capítulo, pretende-se discutir o planejamento de uma pesquisa que objetiva estimar a importância dada a um grande número de atributos pertencentes a um serviço ou produto para posterior análise com o modelo GGUM e entender o comportamento dos indivíduos em relação a esta atribuição da importância.

A importância será perguntada diretamente e será utilizada uma escala ordinal de 4 pontos. O planejamento não se altera caso o número de pontos da escala seja diferente de 4.

Admitindo que os itens não estejam calibrados, isto é, os parâmetros dos itens não são conhecidos, a estimação de todos os parâmetros deve ser feita conjuntamente. Caso os itens estejam calibrados, será discutida uma abordagem mais eficiente no tópico 6.3.

Também será admitido que os atributos considerados foram previamente avaliados, seja por especialistas na área, grupos de focos ou outras pesquisas anteriores. Não há restrições quanto ao número de atributos. Caso este número seja pequeno, o GGUM terá utilidade apenas para modelar os dados e representar os indivíduos num *continuum*, e eventualmente segmentá-los de acordo com a sua posição neste traço medido.

6.2. DESENHO DO EXPERIMENTO

O estudo feito na tese, em conjunto com outros estudos feitos na literatura, mostrou que são necessários no mínimo 20 respostas para cada respondente para estimar o parâmetro do indivíduo.

Dependendo do formato do questionário, caso não haja outras questões (por exemplo sócio-econômicas), pode-se aumentar este número para 20 ou 25 respostas para cada respondente, pois quando se utiliza uma escala de *ratings*, a tarefa de atribuir importância a cada item não é cansativa na opinião da maioria dos entrevistados, como mostrou a pesquisa de campo realizada nesta tese. A limitação desta conclusão se deve ao fato de que esta pesquisa foi feita com pessoas que freqüentam um curso superior e possivelmente têm maior facilidade para responder.

Sendo k o número de itens escolhidos para serem respondidos pelos indivíduos e I o número total de itens, o sorteio dos itens respondidos para cada indivíduo pode ser feito tomando-se as k primeiras colunas de um quadrado latino de tamanho I . Desta forma, nenhum item será mais respondido que outro.

Em seguida à aplicação do questionário, a matriz de importância pode ser estimada pelo software GGUM2004, juntamente com os parâmetros dos atributos e dos indivíduos.

Caso o modelo não apresente um ajuste razoável, os itens responsáveis por esta falta de ajuste podem ser eliminados da análise e serem estimados pela média do item. A falta de ajuste pode ser verificada pelas estatísticas G^2 , *infit* e *outfit* constantes na saída do software.

6.3. ITENS JÁ CALIBRADOS E QUESTIONÁRIOS ADAPTATIVOS

Se os itens considerados já estiverem calibrados por uma pesquisa anterior, existe a possibilidade de conduzir o questionário por meio de um computador. A vantagem desta abordagem é que o computador permite que os itens escolhidos para serem respondidos dêem informação sobre o parâmetro do indivíduo. Isto é feito analisando-se as respostas previamente fornecidas pelo indivíduo e pela função de informação do item.

Em um primeiro passo, admite-se que $\theta_j=0$ e escolhe-se o item que dá maior informação para um indivíduo com este valor de θ_j . Após o indivíduo responder este item, um novo θ_j é estimado e novamente procura-se um item que gere muita informação dado este novo valor de

θ_j . Isto é feito até que o desvio-padrão calculado de θ_j seja menor que um valor pré-especificado.

Estudo feito por ROBERTS, LIN & LAUGHLIN (2001) mostrou que o número médio de itens apresentados aos respondentes em um questionário adaptativo é 7,6. Este número é ainda menor do que o necessário quando as questões são apresentadas aleatoriamente aos indivíduos, como foi o caso estudado nesta tese. Este resultado sugere que aplicar questionários adaptativos é um bom método para alcançar boas estimativas dos parâmetros do modelo com poucos itens respondidos por indivíduo.

7 CONCLUSÕES E LIMITAÇÕES DO ESTUDO

7.1. LIMITAÇÕES DO ESTUDO

No presente estudo foram feitas simulações variando-se o tamanho da amostra, o número de itens e a porcentagem de dados excluídos. Outras simulações seriam necessárias, variando-se ainda mais o tamanho da amostra, o número de itens e a porcentagem de dados excluídos, para estabelecer a relação entre estas variáveis e a precisão do valor previsto pelo modelo.

No estudo de viabilidade da diminuição do número de categorias de quatro para duas, não há garantias de que o comportamento de resposta dos indivíduos ocorre da mesma maneira, isto é, não é possível dizer que as respostas nas categorias “não é importante” e “é pouco importante” seriam transferidas para a categoria “não é importante” no questionário com duas categorias.

Para averiguar isto, poderiam ser aplicados questionários diferentes, cada um com uma das escalas, comparando-se os escores médios (re-escalonados) nos dois casos. Também é possível comparar os valores dos parâmetros de localização dos itens e verificar o quanto eles são compatíveis.

7.2. SUGESTÕES DE ESTUDOS FUTUROS

As sugestões para estudos futuros estão divididas nos três tópicos seguintes.

7.2.1. GGUM multidimensional

Pode-se modelar o comportamento de resposta dos indivíduos em um número maior de dimensões de traços latentes quando o modelo unidimensional não é adequado por não

possuir um bom ajuste. Na TRI cumulativa estes modelos já existem, mas ainda precisam ser desenvolvidos para o GGUM.

7.2.2. Modelo da TRI bidimensional para modelar importância e avaliação

No cômputo do escore global, multiplica-se o valor da importância do item pela avaliação. A soma destes produtos gera o escore global. Nesta tese, a importância foi obtida pela aplicação do GGUM. Poder-se-ia também aplicar o GGUM à avaliação dos itens, admitindo que há um traço latente relacionado ao comportamento do indivíduo ao avaliar os itens.

A estimação dos parâmetros dos indivíduos nos dois traços latentes (relacionado à importância e à avaliação) pode ser realizada através de dois modelos GGUM em separado ou através de um modelo bidimensional em que as estimativas dos parâmetros dos itens e dos indivíduos são obtidas simultaneamente para ambos os traços latentes.

7.2.3. Modelos com escalas comparativas

Costuma-se também utilizar escalas comparativas (*ranking*) em questionários em que a importância é perguntada diretamente.

Pode-se perguntar aos entrevistados quais os k itens, dentre os considerados, são os mais importantes e ordená-los, como no trabalho de CARVALHO e LEITE (1996), do mais importante ao k -ésimo mais importante. Em um caso mais simples, o indivíduo apenas aponta os k itens mais importantes, sem ordená-los.

Uma sugestão é aplicar modelos da TRI a este tipo de problema. A dificuldade reside no fato que as respostas fornecidas pelos indivíduos nesta abordagem possuem correlação entre si, o que fere o pressuposto de independência local exigido nos modelos de TRI.

Devido à inexistência de independência local, as estimativas obtidas pelos modelos usuais da TRI não possuem nenhuma interpretação relevante. Em um problema em que o indivíduo

deve escolher exatamente k dos I itens disponíveis, $k < I$, o número de “acertos” ou “escolhas” é igual para todos os indivíduos, restrição esta que não existe nos modelos de TRI apresentados nesta tese.

A variável aleatória que representa as escolha do item i pelo indivíduo j é

$$X_{ij} = \begin{cases} 1 & \text{se o item } i \text{ for escolhido} \\ 0 & \text{caso contrário} \end{cases}, \text{ para } i=1,2,\dots,I \text{ e } j=1,2,\dots,n$$

Estas variáveis têm distribuição Bernoulli (p_{ij}), $i = 1, \dots, I$, onde p_{ij} é a probabilidade do item i ser escolhido como importante pelo indivíduo j .

Deve-se então encontrar a distribuição de probabilidades do vetor de respostas

$$X_j = (X_{1j}, X_{2j}, \dots, X_{Ij}) \text{ dado que } \sum_{i=1}^I x_{ij} = k.$$

A importância do item i para o indivíduo j é dada pela probabilidade do item ser considerado importante pelo entrevistado, isto é, $w_{ij} = P(X_{ij} = 1 | \theta_j)$. Esta probabilidade, que será notada por p_{ij} , depende de parâmetros do item (ζ) e do indivíduo (θ), como nos modelos considerados nesta tese.

Uma das soluções para o problema de independência entre as respostas aos itens por um indivíduo pode ser a distribuição hipergeométrica multivariada estendida (*extended multivariate hypergeometric distribution*, HARKNESS, 1965) para o vetor de respostas do indivíduo, $P_j(X_j)$, que no modelo com independência local tem distribuição multinomial.

A distribuição hipergeométrica multivariada estendida é dada por

$$P_j(x_j) = \frac{\prod_{i=1}^I t_{ij}^{x_{ij}}}{\sum_{y \in S} \left[\prod_{i=1}^I t_{ij}^{x_{ij}} \right]}, \quad (7.1)$$

e o suporte S é dado por

$$S = \left\{ \underset{\sim}{x} = (x_{1j}, x_{2j}, \dots, x_{Ij}) \mid x_{ij} \in \{0,1\} \text{ e } \sum_{i=1}^I x_{ij} = k \right\}.$$

S é, na verdade, o conjunto de vetores de respostas que o indivíduo pode dar de modo a escolher k dentre os I itens.

t_{ij} é dada por

$$t_{ij} = \frac{p_{ij}}{(1 - p_{ij})q_i} \quad (7.2)$$

onde o fator de escala q_i é ajustado de tal forma que $\sum m_i p_{ij} = k$.

A função de verossimilhança L , baseada nas respostas dos n indivíduos, pode ser escrita como o produtório das probabilidades de resposta de cada indivíduo, pois supõe-se que há independência entre os mesmos:

$$L(\mathbf{X} \mid \boldsymbol{\theta}, \boldsymbol{\varsigma}) = \prod_{j=1}^n P_j(\mathbf{x}_j \mid \boldsymbol{\theta}_j, \boldsymbol{\varsigma}) = \prod_{j=1}^n \frac{\prod_{i=1}^I t_{ij}^{x_{ij}}}{\sum_{\mathbf{y} \in S} \left[\prod_{i=1}^I t_{ij}^{y_{ij}} \right]} \quad (7.3)$$

Para obter as estimativas dos parâmetros, deriva-se e iguala-se a zero o logaritmo da função de verossimilhança. As derivadas parciais do logaritmo da função de verossimilhança em relação aos parâmetros do modelo é complicada de ser obtida, o que dificulta a criação de um algoritmo computacional para calcular as estimativas.

Uma possibilidade de pesquisa futura está no desenvolvimento da função de verossimilhança e suas derivadas e em um algoritmo computacional para obter as estimativas dos parâmetros dos itens e do traço de comportamento do indivíduo, pois não há software disponível que realize este cálculo.

7.3. CONTRIBUIÇÕES

Esta tese se baseia em duas hipóteses principais. A primeira, é que o Modelo de Desdobramento Graduado Generalizado (GGUM), da família de modelos da teoria de Resposta ao Item, pode ser utilizado para modelar a importância conferida por indivíduos a atributos de um serviço ou produto. A segunda hipótese diz que, se o modelo ajustar-se a este tipo de problema, é possível – sem incorrer em perda de informação – simplificar o questionário, ou pela redução do número de itens apresentados a cada indivíduo ou pela redução do número de categorias de resposta a cada atributo.

A primeira hipótese surgiu pois, na literatura, a importância (ou preferência) é tratada com uma abordagem de “ponto ideal”, em que um indivíduo prefere um atributo na medida em que se encontra próximo dele em um *continuum* latente, que representa tanto os indivíduos quanto os atributos. Esta é exatamente a característica do GGUM: quanto mais próximo os itens e indivíduos estiverem neste *continuum*, maior a probabilidade do atributo ser considerado “importante”.

Para testar a primeira hipótese, foi realizada uma pesquisa de campo sobre atributos de uma academia de ginástica e foi ajustado o GGUM. Embora o resultado da análise fatorial demonstre que os dados não podem ser reduzidos a uma dimensão sem perda considerável de informação, o ajuste do modelo foi satisfatório (p-valor alto indicando não haver evidência de falta de ajuste). Posterior exclusão de itens não informativos, diagnosticado através da função de informação do item, proporcionam um ajuste melhor aos dados.

Tanto a estatística de ajuste global (G^2) quanto o quadrado médio da diferença entre o valor observado e o esperado para cada resposta calculada pelo GGUM foram menores do que o quadrado médio calculado pela média do item, estratégia esta comumente utilizada na literatura. Isto significa que, se os dados de importância que um indivíduo atribuiu a um item fosse desconhecido, ele seria melhor estimado pelo GGUM do que pela importância média dada ao item.

Concluindo, a pesquisa de campo sugere que o modelo GGUM é uma alternativa melhor aos métodos de obtenção de importância de atributos existentes na literatura quando se utilizam escalas não comparativas e a importância dos atributos é obtida por método direto.

A segunda hipótese foi verificada através de simulações com dados gerados segundo o GGUM e com dados produzidos pela própria pesquisa de campo. Em ambos os casos, foram estudadas tanto medidas de ajuste quanto o erro de previsão, variando-se o número de respostas excluídas, o tamanho da amostra e o número total de itens. Observou-se, para todos os casos estudados (tamanho de amostra e número de itens), que o erro de previsão tem um aumento significativo para alguns itens quando excluem-se 60% de respostas de indivíduo; contudo este erro ainda é menor do que o erro quando se utiliza a média das respostas dos indivíduos ao item. A exclusão de respostas também provoca um aumento significativo do erro das estimativas dos parâmetros do modelo, o que impacta principalmente na previsão de itens em que o modelo é útil para prever a importância.

Quando há pelo menos 20 respostas por indivíduos, tanto os estimadores dos parâmetros dos itens quanto do parâmetro do indivíduo têm erro pequeno, independente do número total de itens. Isto torna o GGUM especialmente vantajoso quando o número de atributos é grande e deseja-se tornar a tarefa de atribuição de importância menos cansativa, através de desenhos experimentais em que o respondente avalia um subconjunto de atributos ou, eventualmente, de questionários adaptativos em que os atributos são apresentados de acordo com a quantidade de informação extraída do indivíduo sobre seu traço latente.

Uma alternativa à abordagem anterior é modelar a importância dos atributos não respondidos por determinado indivíduo pelo valor da categoria da escala com maior probabilidade de escolha no GGUM. A importância dos atributos respondidos é dada pelo valor observado da escala. Esta abordagem possui as mesmas vantagens e desvantagens da anterior, mas possui um erro na previsão da resposta do entrevistado um pouco maior.

Pode-se também definir a importância dos atributos pela soma das probabilidades de escolha das categorias “importante” e “muito importante”, que são obtidas pelo modelo GGUM. O erro proveniente desta abordagem é muito similar à obtida pelo valor esperado.

Foi estudado também um outro tipo de simplificação do questionário, obtido pela diminuição do número de categorias de resposta. A redução do número de categorias, de quatro para duas, pouco alterou os valores estimados para a localização do indivíduo no *continuum* latente, o que mostra que, sem perda de informação, um questionário mais simples poderia ter sido aplicado, apenas com as categorias de resposta importante/ não importante. Isto é corroborado pelo fato das previsões feitas com o GGUM politômico e o GGUM dicotômico terem correlação igual a 0,973.

A aplicação do GGUM contribui para o desenvolvimento de pesquisas de mercado devido às vantagens que possui em relação às escalas não comparativas para obtenção de importância através de métodos diretos comumente utilizados.

De posse das estimativas de importância w_{ij} , é recomendável, por facilidade de interpretação, fazer uma transformação tal que a soma das importâncias atribuída aos itens por um respondente seja 1. Com isto, a importância funciona como um peso dado a cada atributo para calcular um escore global de avaliação do serviço ou produto.

8 REFERÊNCIAS BIBLIOGRÁFICAS

AKATU (Instituto). *Responsabilidade social empresarial: o que o consumidor consciente espera das empresas*. São Paulo, 2005, 44p.

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

ANDERSON, N. H. *Foundation of information integration theory*. New York, Academic Press, 1971.

ANDRADE, D. F. e KLEIN, R. Métodos estatísticos para avaliação educacional: teoria da resposta ao item. *Boletim da ABE*, 43, 21-28, 1999.

ANDRADE, D. F., TAVARES, H. R. e VALLE, R.C. Teoria de Resposta ao Item: conceitos e aplicações. *ABE — Associação Brasileira de Estatística*, 4º SINAPE, 2000.

ANDRICH, D. A rating formulation for ordered response categories. *Psychometrika*, v. 43, p. 561-573, 1978.

ANDRICH, D. A general hyperbolic cosine latent trait model for unfolding polytomous responses: reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, v. 49, p. 347-365, 1996.

ANDRICH, D. A hyperbolic cosine IRT model for unfolding direct responses of persons to items. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (p.399-414).NewYork: Springer, 1997.

ANDRICH, D & LUO, G. A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement*, 17, p. 253-276, 1993.

AZEVEDO, C. L. N. *Métodos de estimação na teoria de resposta ao item*. Dissertação (Mestrado) – Instituto de Matemática e Estatística – Universidade de São Paulo. São Paulo, 2003.

BAKER, F. B. and KIM, S. *Item Response Theory: Parameter Estimation Techniques*. Statistics, New York, Dekker Series of Textbooks and Monographs, 2004.

BIRNBAUM, A. *Some latent trait models and their use in inferring an examinee's ability*. In LORD, F.M. and NOVICK, M.R., *Statistical theories of mental test scores* (cap. 17-20). Reading, MA: Addison-Wesley, 1968.

BOCK, R.D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, v. 37, p. 29-51, 1972.

BOCK, R. D., & AITKIN, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459, 1981.

- BOCK, R.D., GIBBONS, R., & MURAKI, E. Full-information item factor analysis. *Applied Psychological Measurement*, 21, 261-280, 1988.
- BOCK, R. D., & LIEBERMAN, M. Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197, 1970.
- BOCK, R. D., & MISLEVY, R. J. Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444, 1982.
- BOECK, P. D. & WILSON, M. *Explanatory item response models: a generalized linear and nonlinear approach*. New York: Springer-Verlag, 2004.
- BORTOLOTTI, S. L. V. *Aplicação de um modelo de desdobramento da teoria da resposta ao item –TRI*. Dissertação (Mestrado) - UFSC. Programa de Pós-Graduação em Engenharia de Produção. Florianópolis, 2003. 107p.
- BRASIL (País). Ministério do Planejamento, Orçamento e Gestão. *Pesquisa nacional de avaliação da satisfação dos usuários de serviços públicos*. Brasília, 2000.
- BUSSAB, W. O. e MORETTIN, P. A. *Estatística Básica*. 5ª ed. São Paulo: Saraiva, 2002.
- CARMAN, J.M. Consumer perceptions of service quality: an assessment of SERVQUAL dimensions. *Journal of Retailing*, vol. 66, n.1, p. 33-55. Spring, 1990.
- CARROLL, J. D. e CHANG, J.J. Analysis of individual differences in multidimensional scaling via an n-way generalization of ‘eckart-young’ decomposition. *Psychometrika*, vol 35, p. 283-319, 1970.
- CARVALHO, F., LEITE, V.F. Pesquisa de marketing: a ordem das questões altera o resultado? In: ENCONTRO ANUAL DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, n.20, 1996, Angra dos Reis. Anais... Angra dos Reis: Anpad, 1996. Marketing, p.511.
- CARVALHO, F., LEITE, V.F. Alternativa de ordenação da importância de atributos da qualidade de serviços: um estudo exploratório sobre o efeito do tamanho do choice set. In: ENCONTRO ANUAL DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, n.22, 1998, Foz do Iguaçu. Anais... Foz do Iguaçu: Anpad, 1998. Marketing, p.111.
- CHURCHILL, G.A. *Marketing Research: methodological foundations*. 5a ed. Orlando: The Dryden Press, 1991.
- COHEN, S. H. and NEIRA, L. Measuring preference for product benefits across countries: overcoming scale usage bias with Maximum Difference scaling. ESOMAR 2003 – Latin America Conference Proceedings. Amsterdam, Netherlands.
- COOMBS, C. H. Psychological scaling without a unit of measurement. *Psychological review*, 1950, vol 57, p. 148-158.

COSTA, M. B. F. *Técnica Derivada da Teoria de Resposta ao Item (TRI) Aplicada ao Setor de Serviços*. 2001. 134f. Dissertação – Universidade Federal do Paraná, 2001.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297-335

CRONIN, J.J.; TAYLOR, S.A. SERVPREF versus SERVQUAL: reconciling performance-based and perceptions minus expectations measurement of service quality. *Journal of Marketing*, v.58, p. 125-131, 1994.

CUNHA JR, M.V.M., BORGES JR., A. A. FACHEL, J.M. Esquema CBF para mensuração da satisfação de clientes: uma proposta conceitual e prática. In: ENCONTRO ANUAL DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, n.22, 1998, Foz do Iguaçu. Anais... Foz do Iguaçu: Anpad, 1998. Marketing, p.116.

DAVIS, M. H., & ROBERTS, J. S. Missing data effects on the MML/EAP parameter estimates in the generalized graded unfolding model. *Poster presented at the annual meeting of the American Educational Research Association*, San Diego, California, 2004.

DEMARS, C. E. Type I error rates for the generalized graded unfolding model fit indices. *Applied Psychological Measurement*, 28, 48-71, 2004.

DEMPSTER, A. P., LAIRD, N.M. and RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, v. 39, p. 1-38, 1977.

DOMENCICH, T.; McFADDEN, L. *Urban travel demand: A Behavioral Analysis*, North Holland, Amsterdam, 1975.

ENGELHARD, G., Jr. The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education*, 5, p. 171-191, 1992.

GREEN, P. E. and RAO, V. *Applied multidimensional scaling*. New York: Holt, Rinehart & Winston, 1972.

GREEN, P. E.; SRINIVASAN, V. Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, v. 5, n. 2, p. 103-123. Sep. 1978.

GREEN, P. E. and KREIGER, A.M. Attribute importance weights modification in assessing a brand's competitive potential. *Marketing Science*, 14 (3) p. 253-270, 1995.

GHISI, M. A., MERLO E. M., NAGANO, M. S. A mensuração da importância de atributos em serviços: uma comparação de escalas. In: ENCONTRO ANUAL DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, n.28, 2004, Curitiba. Anais... Curitiba: Anpad, 2004. Marketing, p.415.

HALEY, R. I. *Developing effective communications strategy: A benefit segmentation approach*. New York: John Wiley & Sons, 1985.

HAMBLETON, R. K. & SWAMINATHAN, H. *Item Response Theory Principles and Applications*. Boston: Kluwer Academic Publishers, 1985.

HAMBLETON, R.K.; SWAMINATHAN, H. e ROGERS, H.J. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, 1991, 174p.

HARKNESS, W. L. Properties of the Extended Hypergeometric Distribution. *Annals of Mathematical Statistics*, 1965, vol. 36, no. 3, pp. 938-945.

HILDEBRAND, F. B. *Introduction to numerical analysis*. New York: McGraw-Hill, 1969.

HOIJTINK, H. *Parella: Measurement of latent traits by proximity items*. Leiden, The Netherlands: DSWO Press, 1990.

ISAAC, E. e KELLER, H.B. *Analysis of numerical methods*. New York: Wiley and Sons, 1966.

JOHNSON, N. L. & KOTZ, S. *Distributions in Statistics, 1: Discrete Distributions*. Boston: Houghton Mifflin Co, 1969.

JOHNSON, R. A. e WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 5a ed. Prentice Hall, Inc. Englewood Cliffs, New Jersey, 2002.

JOHNSON, R.M. History of ACA. *Research paper series*. Sawtooth Software, Inc. Sequim: WA. 2001.

JSTOR. Disponível em

<<http://www.jstor.org/search/AdvancedSearch?si=1&hp=25&All=item+response+theory&Exact=&One=&None=&sd=&ed=&jt=&dc=Business&Search=Search>>. Acesso em 20 nov. 2005.

KARABATSOS, G. A critique of Rasch residual fit statistics. *Journal of Applied Measurement*, 1, p.152-176, 2000.

KEENEY, R. L. and RAIFFA, H. *Decisions with multiple objectives: preferences and value tradeoffs*. New York, Wiley, 1976.

KOTLER, P., LILIEN, G. L., MOORTHY, K. S. *Marketing Models*. Englewood Cliffs. Prentice Hall Intl, 1992.

LANCASTER, K. A new approach to consumer theory. *Journal of Political Economy*, n. 74, p. 132-157, 1966.

LATTIN, J ; GREEN, P. E. and CARROLL, J.D. *Analysing Multivariate Data*. Pacific Grove, CA. Thomson Learning, 2003, 556p.

LINACRE, J.M. *A user's guide to WINSTEPS and MINISTEPS computer programs*. Chicago: Author, 1991.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

LUNZ, M.E., WRIGHT, B.D., & LINACRE, J.M. Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, p.331-345, 1990.

LOUVIERE, J.J. FINN, A. Determining the appropriate response to evidence of public concerns: the case of food safety. *Journal of Public Policy and Marketing*, vol. 11 pag 12-25. 1993.

MALHOTRA, N.K. *Pesquisa de Marketing: uma orientação aplicada*. 3^a ed. Porto Alegre: Bookman, 2001, 719p.

MARCHETTI, R. e PRADO, P. H. M. Um tour pelas medidas de satisfação do consumidor. *RAE: revista de administração de empresas*, São Paulo, v.41, n. 4, p. 56-67, out./dez. 2001.

MASTERS, G. N. A rasch model for partial credit scoring. *Psychometrika*, v. 47, p. 149-174, 1982.

MISLEVY, R. J. Bayes modal estimation in item response models. *Psychometrika*, v. 51, p. 177-195, 1986.

MURAKI, E. A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement*, v. 16, p. 159-176, 1992.

MURAKI, E., & BOCK, D. PARSCALE: *IRT item analysis and test scoring for rating-scale data*. Chicago: Scientific Software, 1997.

NOJOSA, R. T. *Modelos Multidimensionais para a Teoria de Resposta ao Item*. Pernambuco, UFPE, Tese de Mestrado, 2001.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *PISA 2000 Technical Report*. Paris: OECD Publications. 2000.

ORLANDO, M. & THISSEM, D. Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64, 2000.

ORTUZAR, Juan de Dios; WILLUMSEN, Luis G. *Modeling transport*. Chichester England : John Wiley & Sons, 1990.

PARASURAMAN, A.; ZEITHAML, V. A.; BERRY, L.L. A conceptual model of service quality and its implications for future research, *Journal of Marketing*, vol. 49, 1985, p. 41-50.

PARASURAMAN, A.; ZEITHAML, V. A.; BERRY, L.L. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, v. 64, n. 2, p. 12-40, 1988.

PARASURAMAN, A.; BERRY, L.L.; ZEITHAML, V.A. Understanding customer expectations of service, *Sloan Management Review*, vol. , 1991, p. 39-48 (Spring).

PARASURAMAN, A., ZEITHAML, V.A.; BERRY, L.L. Alternative scales for measuring service quality: a comparative assessment based on psychometric and diagnostic criteria, *Journal of Retailing*, v. 70, n. 3, p. 201-230, 1994.

PARASURAMAN, A.; ZEITHAML, V. A.; BERRY, L.L. A conceptual model of service quality and its implications for future research, *Journal of Marketing*, vol. 49, 1985, p. 41-50.

PASQUALI, L. *Psicometria: teoria e aplicações*. Brasília: UNB, 1997, 289p.

PAULHUS, D.L. Measurement and control of response bias. *Measures of personality and social psychological attitudes*. Academic Press, San Diego, CA, 1991.

RAO, C. R. *Linear statistical inference and its applications*. New York: Wiley & Sons, 1973.

RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, p. 207-230, 1979.

ROBERTS, J. S. *Item response theory approaches to attitude Measurement*. (Doctoral dissertation, University of South Carolina, Columbia), Dissertation Abstracts International, 56, 7089B, 1995.

ROBERTS, J. S. A modified likelihood-based item fit statistic for the generalized graded unfolding model. *Paper presented at the annual meeting of the American Educational Research Association*, San Diego, California, 2004.

ROBERTS, J. S., DONOGHUE, J. R., & LAUGHLIN, J. E. Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement*, 26, 192-207, 2002.

ROBERTS, J. S. & LAUGHLIN, J. E. A. Unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement*, 20, p. 231-255, 1996a.

ROBERTS, J. S. & LAUGHLIN, J. E. The graded unfolding model: A Unidimensional item response model for unfolding graded responses (Research Rep. RR-96-16). Princeton N J: *Educational Testing Service*, 1996b.

ROBERTS, J.S, LAUGHLIN, J.E. e WEDELL, D.H. Validity issues in the Likert and Thrustone approaches to attitude measurement. *Educational and Psychological Measurement*, v. 59, p. 211-233, 1999.

ROBERTS, J.S, DONOGHUE, J.R. e LAUGHLIN, J.E. A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, v. 24, p. 3-32, 2000.

ROBERTS, J.S., FANG, H., CUI, W. & WANG, Y. GGUM2004: *A Windows-based program to estimate parameters of the generalized graded unfolding model*. Manuscript in preparation (2004).

ROBERTS, J. S., LIN, Y. & LAUGHLIN, J. E. Computerized adaptative testing with the generalized graded unfolding model. *Applied Psychological Measurement*, 25(2), p. 177-196, 2001.

SMITH, R. M., SCHUMACKER, R. E., & BUSH, M. J. Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement*, 2, p. 66-78, 1998.

SAMEJIMA, F. A. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 17, 1969.

SIJTSMA, K. e MOLENAAR, I. W. *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage Publications. v.5, 176p, 2002.

STROUD, A. H. and SECREST, D. *Gaussian Quadrature Formulas*. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

VALLE, R.C. *Teoria da Resposta ao Item*. Dissertação – Universidade de São Paulo – Instituto de Matemática e Estatística, 1999.

WILLIAMS, H. C. W. L. On the formation of travel demand models and economic measures of user benefit. *Environment and Planning*. 9A., p. 285-344, 1977.

WINER, B. J., BROWN, D. .R, MICHELS, K. M . *Statistical principles in experimental designs*. McGraw Hill, 1991.

ZEITHMAN, V.A., PARASURAMAN, A. & BERRY, L. I. *Delivering service quality: balancing customer's perceptions and expectations*. New York, Free Press, 1990.

APÊNDICES

APÊNDICE A –Resumo de estatísticas dos itens para o ajuste do modelo 8 para a simulação com $N=1000$ e $I=25$, sem dados excluídos.

	delta		real-estimado	alpha		real-estimado
Item	estimado	delta	delta	estimado	alpha	alpha
1	-5,98	-2,40	3,58	0,21	0,20	-0,01
2	-2,79	-2,20	0,59	0,48	0,50	0,02
3	-2,13	-2,00	0,13	0,97	1,00	0,03
4	-1,71	-1,80	-0,09	2,20	2,00	-0,20
5	-1,54	-1,60	-0,06	3,63	4,00	0,37
6	-1,39	-1,40	-0,01	0,21	0,20	-0,01
7	-1,27	-1,20	0,07	0,46	0,50	0,04
8	-0,99	-1,00	-0,01	1,00	1,00	0,00
9	-0,77	-0,80	-0,03	2,08	2,00	-0,08
10	-0,62	-0,60	0,02	3,89	4,00	0,11
11	-0,17	-0,40	-0,23	0,19	0,20	0,01
12	-0,28	-0,20	0,08	0,46	0,50	0,04
13	0,03	0,00	-0,03	0,95	1,00	0,05
14	0,20	0,20	0,00	1,99	2,00	0,01
15	0,41	0,40	-0,01	3,69	4,00	0,31
16	0,69	0,60	-0,09	0,28	0,20	-0,08
17	0,82	0,80	-0,02	0,50	0,50	0,00
18	1,03	1,00	-0,03	1,05	1,00	-0,05
19	1,28	1,20	-0,08	1,83	2,00	0,17
20	1,39	1,40	0,01	3,94	4,00	0,06
21	1,10	1,60	0,50	0,22	0,20	-0,02
22	2,09	1,80	-0,29	0,46	0,50	0,04
23	2,11	2,00	-0,11	0,94	1,00	0,06
24	4,37	2,20	-2,17	1,49	2,00	0,51
25	4,63	2,40	-2,23	2,72	4,00	1,28
QMErro			0,93	QMErro 0,09		

Item	p-valor infit	p-valor outfit	G2	Graus de liberdade G2	p-valor G2	QMErro Alternativa A	QMErro Alternativa B	QMErro Alternativa C	QMErro Alternativa D	QMErro Alternativa E
1	0,47	0,47	45,44	60	0,92	0,136	0,143	0,165	0,129	0,129
2	0,55	0,65	60,26	57	0,36	0,127	0,128	0,115	0,102	0,103
3	0,59	0,75	56,93	45	0,11	0,127	0,127	0,093	0,074	0,078
4	0,98	0,51	36,40	21	0,02	0,131	0,131	0,064	0,040	0,047
5	1,00	1,00	8,69	9	0,47	0,139	0,140	0,058	0,026	0,035
6	0,45	0,46	66,12	60	0,27	0,135	0,141	0,165	0,131	0,131
7	0,53	0,55	60,87	60	0,44	0,141	0,144	0,177	0,119	0,120
8	0,70	0,91	52,69	54	0,53	0,144	0,144	0,123	0,090	0,092
9	0,95	1,00	23,37	36	0,95	0,144	0,147	0,093	0,060	0,062
10	1,00	1,00	39,74	21	0,01	0,138	0,151	0,076	0,034	0,035
11	0,46	0,46	64,25	60	0,33	0,140	0,143	0,218	0,139	0,139
12	0,53	0,52	35,70	60	0,99	0,141	0,142	0,214	0,130	0,130
13	0,76	0,69	50,13	54	0,62	0,136	0,148	0,193	0,105	0,105
14	0,92	0,93	5,01	9	0,83	0,054	0,207	0,036	0,029	0,037
15	1,00	1,00	2,03	3	0,57	0,060	0,197	0,023	0,018	0,031
16	0,46	0,45	45,83	60	0,91	0,131	0,150	0,222	0,126	0,127
17	0,52	0,51	45,66	60	0,91	0,117	0,151	0,143	0,102	0,103
18	0,69	0,67	46,77	51	0,64	0,130	0,147	0,098	0,078	0,081
19	0,94	0,96	28,09	33	0,71	0,140	0,140	0,064	0,053	0,057
20	1,00	1,00	24,14	12	0,02	0,136	0,145	0,032	0,023	0,028
21	0,48	0,50	53,26	60	0,72	0,142	0,145	0,306	0,138	0,138
22	0,56	0,58	50,10	57	0,73	0,124	0,133	0,181	0,102	0,104
23	0,88	0,67	30,46	27	0,29	0,078	0,151	0,073	0,049	0,050
24	0,26	0,00	8,45	9	0,49	0,036	0,179	0,033	0,019	0,020
25	0,00	1,00	3,63	3	0,30	0,009	0,192	0,009	0,005	0,006
Total Geral			944,01	245	0,80	0,117	0,151	0,119	0,077	0,079

APÊNDICE B – Tabela com frequência dos atributos citados na enquete.

	<i>Frequência</i>
infra-estrutura	32
geral (sem especificar)	10
estacionamento gratuito	2
ambiente agradável	3
segurança	1
vestiário	1
limpeza	6
tamanho da piscina	2
espaço	5
disponibilidade de recursos	1
alta tecnologia	1
equipamentos	29
qualidade	15
manutenção	2
disponibilidade	5
geral	4
variedade	3
funcionários / professores	24
profissionais simpáticos	3
geral	5
professores qualificados	13
professores atenciosos	1
professores presentes	1
funcionários qualificados	1
atividades disponíveis	22
aulas de dança	2
natação / piscina	2
esportes	2
geral	2
variedade de aulas	14
preço	22
geral	22
localização	21
geral	14
próximo ao local de origem	1
próximo à casa	3
próximo à faculdade	1
próximo ao trabalho	1
próximo ao metrô ou ônibus	1

frequência	<i>13</i>
gente bonita	3
ser vazia	1
geral	5
bom relacionamento com frequentadores	1
ter amigos ou conhecidos na academia	2
quantidade de pessoas	1
horário	<i>7</i>
geral	2
flexibilidade	5
serviços	<i>5</i>
personal trainer	1
água de graça	1
atendimento	1
avaliação física	1
geral	1
outros	<i>1</i>
resultados em curto prazo	1

APÊNDICE C – Atributos selecionados para pesquisa de campo

<i>i</i>	Atributo
1	Ter bastante espaço físico
2	Ter estacionamento gratuito
3	Ter limpeza
4	Ter segurança
5	Ter equipamentos de qualidade
6	Ter disponibilidade de equipamentos
7	Ter professores qualificados
8	Ter professores que dão atenção aos alunos
9	Ter aulas de dança
10	Ter piscina aquecida
11	Ter quadra poliesportiva
12	Ter aulas de lutas marciais
13	Abrir domingos e feriados
14	Ficar aberta 24h por dia
15	Ter serviço de profissional trainer
16	Estar localizada perto de casa
17	Ter lanchonete
18	Ter várias unidades (endereços que você pode freqüentar)
19	Ter bom preço
20	Ter pessoas do mesmo nível sócio-econômico que o meu
21	Ter amigos ou conhecidos que freqüentam a academia
22	Ser freqüentada por gente bonita
23	Ter orientação nutricional
24	Ter bom atendimento
25	Ter avaliação física

APÊNDICE D – Questionário apresentado aos alunos

Nome: _____ 35

QUESTIONÁRIO - ATRIBUTOS DE UMA ACADEMIA DE GINÁSTICA

Você frequenta alguma academia de ginástica? ☐ Não ☐ Sim
Qual? _____

Se não, tem interesse em frequentar?
(marque a alternativa que melhor expressa seu interesse)

1	Definitivamente não
2	
3	Um pouco
4	
5	Definitivamente sim

Por favor, responda as perguntas abaixo **mesmo que você não tenha interesse** em frequentar uma academia e marque o tempo que você precisou (em segundos) para dar nota aos 25 atributos.

Para você, qual é a importância de cada um dos atributos abaixo na escolha de uma academia de ginástica?

Atributo	1 <i>Não é importante</i>	2 <i>É pouco importante</i>	3 <i>É importante</i>	4 <i>É muito importante</i>
Ter disponibilidade de equipamentos				
Ficar aberta 24h por dia				
Ter aulas de lutas marciais				
Ter lanchonete				
Ter equipamentos de qualidade				
Ter estacionamento gratuito				
Ter professores que dão atenção aos alunos				
Ter bom preço				
Abrir domingos e feriados				
Ter várias unidades (endereços que você pode frequentar)				
Ter limpeza				
Ser frequentada por gente bonita				
Ter professores qualificados				
Ter bom atendimento				
Ter pessoas do mesmo nível sócio-econômico que o meu				
Ter segurança				
Ter piscina aquecida				
Ter orientação nutricional				
Estar localizada perto de casa				
Ter serviço de personal trainer				
Ter aulas de dança				
Ter bastante espaço físico				
Ter amigos ou conhecidos que frequentam a academia				
Ter quadra poliesportiva				
Ter avaliação física				

Tempo necessário para dar as 25 notas (em segundos):

APÊNDICE E – Distribuição de frequências, média e desvio-padrão dos 25 itens da pesquisa de campo.

Atributo	Não é importante	Um pouco importante	Importante	Muito importante	média	desvio-padrão
Ter disponibilidade de equipamentos	1	1	69	292	3,80	0,43
Ter equipamentos de qualidade	0	1	78	284	3,78	0,42
Ter limpeza	0	6	91	266	3,72	0,49
Ter professores que dão atenção aos alunos	0	12	95	256	3,67	0,54
Estar localizada perto de casa	0	11	97	255	3,67	0,53
Ter professores qualificados	0	10	101	252	3,67	0,53
Ter bom preço	3	14	151	195	3,48	0,61
Ter bom atendimento	0	22	148	193	3,47	0,61
Ter segurança	1	23	149	190	3,45	0,63
Ter avaliação física	6	44	131	182	3,35	0,75
Ter estacionamento gratuito	33	44	122	164	3,15	0,96
Ter bastante espaço físico	8	45	217	93	3,09	0,68
Ter orientação nutricional	30	111	144	78	2,74	0,89
Abrir aos domingos e feriados	54	101	102	106	2,72	1,04
Ter amigos ou conhecidos que freqüentam a academia	48	122	149	44	2,52	0,87
Ter serviço de personal trainer	68	111	120	64	2,50	0,99
Ser freqüentada por gente bonita	55	126	134	48	2,48	0,91
Ter lanchonete	66	110	139	48	2,47	0,94
Ter piscina aquecida	82	107	118	56	2,41	1,00
Ficar aberta 24h por dia	67	132	115	49	2,40	0,94
Ter pessoas do mesmo nível sócio-econômico que o meu	58	161	123	21	2,29	0,80
Ter várias unidades (endereço que você pode freqüentar)	93	146	92	32	2,17	0,91
Ter quadra poliesportiva	118	139	66	40	2,08	0,97
Ter aulas de lutas marciais	130	128	73	32	2,02	0,96
Ter aulas de dança	173	82	64	44	1,94	1,07

Fonte: pesquisa de campo

APÊNDICE F – Distribuição de frequência dos 25 itens, com categorias de resposta agrupadas.

Atributo	Não é importante	É importante	Média
Ter disponibilidade de equipamentos	2	361	0,994
Ter equipamentos de qualidade	1	362	0,997
Ter limpeza	6	357	0,983
Ter professores que dão atenção aos alunos	12	351	0,967
Estar localizada perto de casa	11	352	0,970
Ter professores qualificados	10	353	0,972
Ter bom preço	17	346	0,953
Ter bom atendimento	22	341	0,939
Ter segurança	24	339	0,934
Ter avaliação física	50	313	0,862
Ter estacionamento gratuito	77	286	0,788
Ter bastante espaço físico	53	310	0,854
Ter orientação nutricional	141	222	0,612
Abrir aos domingos e feriados	155	208	0,573
Ter amigos ou conhecidos que freqüentam a academia	170	193	0,532
Ter serviço de personal trainer	179	184	0,507
Ser freqüentada por gente bonita	181	182	0,501
Ter lanchonete	176	187	0,515
Ter piscina aquecida	189	174	0,479
Ficar aberta 24h por dia	199	164	0,452
Ter pessoas do mesmo nível sócio-econômico que o meu	219	144	0,397
Ter várias unidades (endereço que você pode freqüentar)	239	124	0,342
Ter quadra poliesportiva	257	106	0,292
Ter aulas de lutas marciais	258	105	0,289
Ter aulas de dança	255	108	0,298

APÊNDICE G – Estatísticas de ajuste do GGUM aos dados da pesquisa de campo no modelo dicotômico.

item	p-valor G2	p-valor infit	p-valor outfit
1	0,99	0,27	0,53
2	0,38	0,47	0,91
3	0,97	0,00	0,70
4	0,22	0,26	0,67
5	0,78	0,00	1,00
6	0,97	0,31	0,79
7	0,62	0,00	1,00
8	0,48	0,05	1,00
9	0,04	0,77	0,85
10	0,36	0,94	0,93
11	0,55	0,93	1,00
12	0,41	0,49	0,48
13	0,98	0,52	0,54
14	0,56	0,60	0,62
15	0,21	0,90	0,93
16	0,83	0,40	0,54
17	0,07	0,99	1,00
18	0,69	0,85	0,90
19	0,04	0,32	0,84
20	0,70	0,53	0,54
21	0,82	0,54	0,55
22	0,98	0,46	0,46
23	0,41	0,98	1,00
24	0,85	0,05	1,00
25	0,26	0,42	0,91
total	0,72		

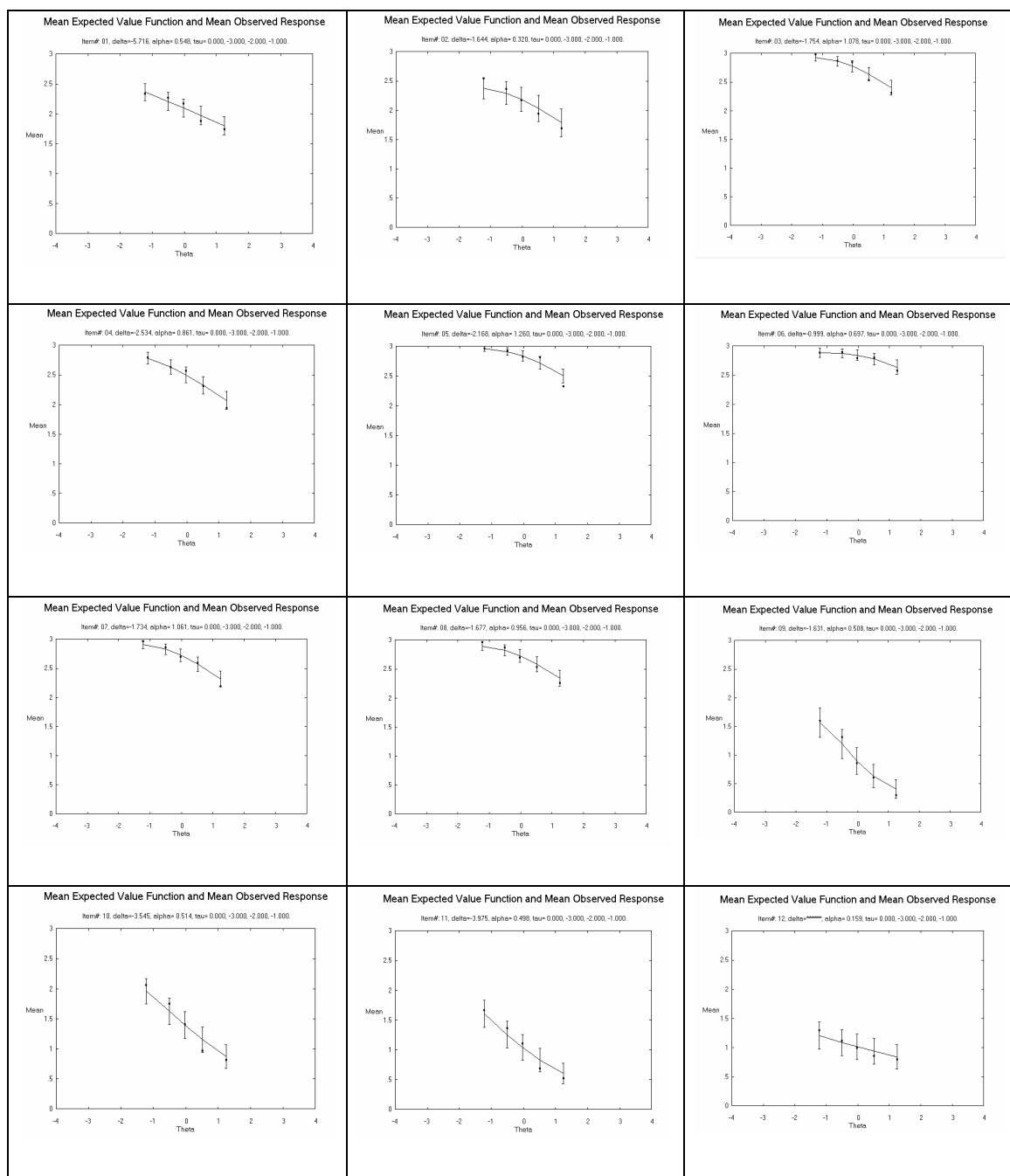
APÊNDICE H – Estatísticas de ajuste do GGUM aos dados da pesquisa de campo no modelo com 4 categorias de resposta.

item	p-valor G2	p-valor infit	p-valor outfit	QMErro previsão alternativa A	QMErro previsão alternativa B	QMErro previsão alternativa C	QMErro previsão alternativa D	QMErro previsão alternativa E
1	0,022	0,580	0,622	0,051	0,054	0,050	0,044	0,061
2	0,073	0,235	0,441	0,102	0,107	0,118	0,092	0,093
3	0,846	0,581	0,965	0,026	0,096	0,028	0,020	0,031
4	0,964	0,775	0,802	0,043	0,075	0,043	0,032	0,048
5	0,998	0,731	0,984	0,020	0,102	0,021	0,015	0,024
6	0,056	0,248	0,579	0,020	0,105	0,025	0,019	0,024
7	0,770	0,633	0,969	0,031	0,093	0,030	0,022	0,035
8	0,477	0,487	0,963	0,032	0,094	0,035	0,024	0,035
9	0,409	0,722	0,759	0,126	0,233	0,188	0,099	0,099
10	0,107	0,752	0,675	0,111	0,141	0,096	0,086	0,085
11	0,597	0,739	0,767	0,105	0,184	0,101	0,083	0,084
12	0,448	0,488	0,412	0,101	0,192	0,148	0,098	0,100
13	0,572	0,484	0,497	0,121	0,125	0,129	0,116	0,116
14	0,374	0,493	0,480	0,098	0,128	0,107	0,094	0,094
15	0,490	0,760	0,761	0,109	0,129	0,094	0,083	0,083
16	0,909	0,438	0,443	0,031	0,094	0,043	0,031	0,038
17	0,015	0,808	0,835	0,097	0,121	0,081	0,069	0,069
18	0,887	0,663	0,653	0,092	0,155	0,091	0,078	0,080
19	0,036	0,434	0,471	0,042	0,077	0,074	0,042	0,055
20	0,688	0,469	0,468	0,071	0,115	0,079	0,070	0,072
21	0,458	0,476	0,466	0,084	0,102	0,098	0,080	0,079
22	0,302	0,439	0,438	0,091	0,112	0,109	0,091	0,091
23	0,813	0,958	0,781	0,087	0,091	0,058	0,052	0,056
24	0,827	0,690	0,758	0,041	0,075	0,044	0,033	0,050
25	0,750	0,612	0,877	0,063	0,083	0,052	0,046	0,055
Total geral	0,206			0,072	0,115	0,078	0,061	0,066

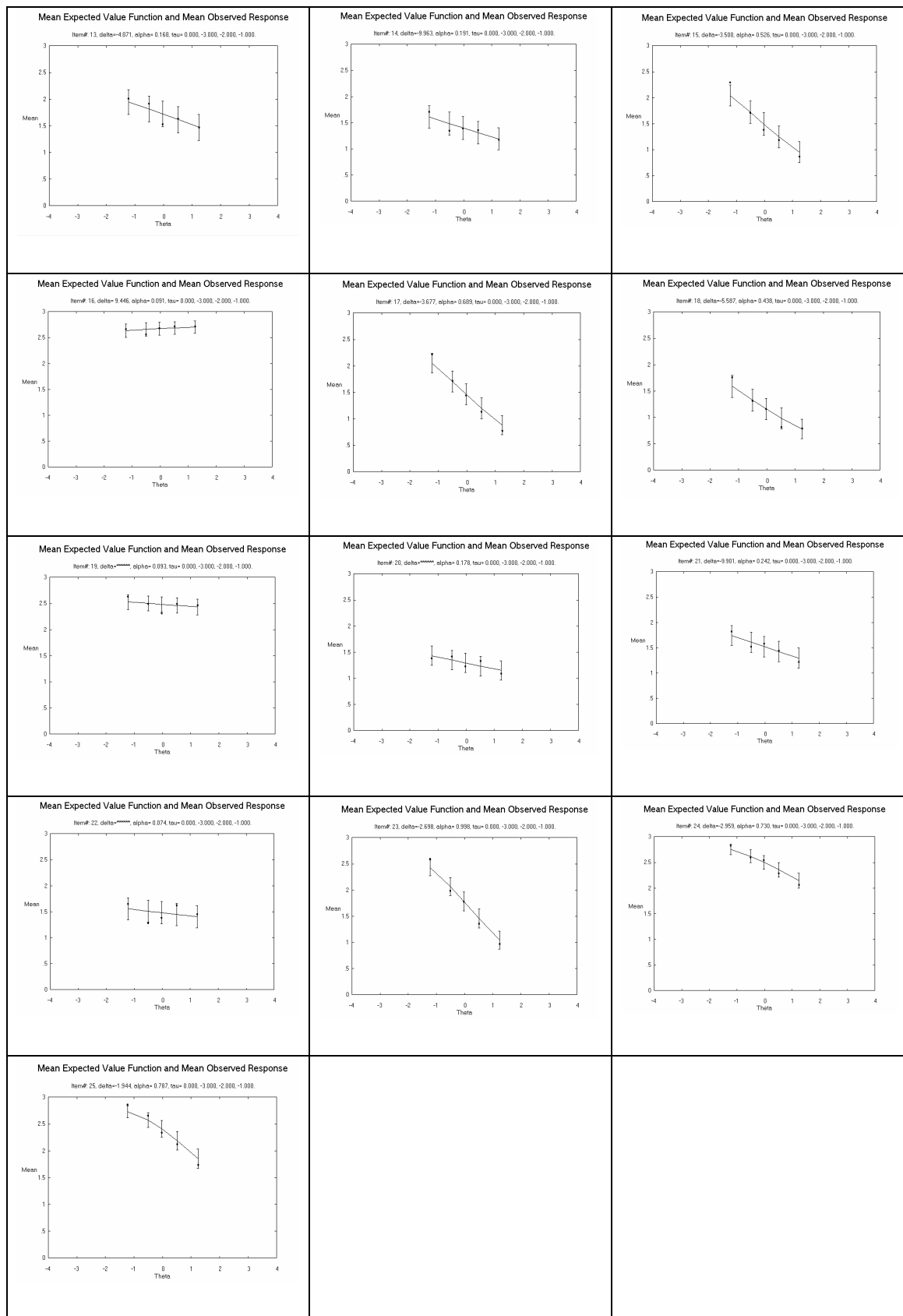
APÊNDICE I – Estimativas dos parâmetros dos itens para modelo com 4 categorias, sem respostas excluídas. Itens com α_i menor que 0,5 estão em negrito.

Item	δ estimado	α estimado	τ_1 estimado	τ_2 estimado	τ_3 estimado
1	-5,716	0,548	-12,282	-8,188	-4,094
2	-1,644	0,320	-5,180	-3,453	-1,727
3	-1,754	1,078	-8,391	-5,594	-2,797
4	-2,534	0,861	-7,984	-5,322	-2,661
5	-2,168	1,260	-10,328	-6,885	-3,443
6	-0,999	0,697	-9,196	-6,131	-3,065
7	-1,734	1,061	-7,714	-5,142	-2,571
8	-1,677	0,956	-7,766	-5,177	-2,589
9	-1,631	0,508	-0,558	-0,372	-0,186
10	-3,545	0,514	-4,824	-3,216	-1,608
11	-3,975	0,498	-4,281	-2,854	-1,427
12	-10,142	0,159	-10,030	-6,687	-3,343
13	-4,871	0,168	-8,103	-5,402	-2,701
14	-9,963	0,191	-13,733	-9,155	-4,578
15	-3,500	0,526	-5,084	-3,389	-1,695
16	9,446	0,091	-50,674	-33,782	-16,891
17	-3,677	0,689	-5,328	-3,552	-1,776
18	-5,587	0,438	-6,801	-4,534	-2,267
19	-17,647	0,093	-57,537	-38,358	-19,179
20	-17,83	0,178	-23,953	-15,968	-7,984
21	-9,901	0,242	-14,872	-9,915	-4,957
22	-27,428	0,074	-40,025	-26,683	-13,342
23	-2,698	0,998	-4,709	-3,139	-1,570
24	-2,959	0,73	-9,407	-6,271	-3,136
25	-1,944	0,787	-5,855	-3,903	-1,952

APÊNDICE J – Escore esperado dos itens no modelo com 4 categorias, sem respostas excluídas, para valores de θ_j .



(conclusão da Figura do Apêndice J)



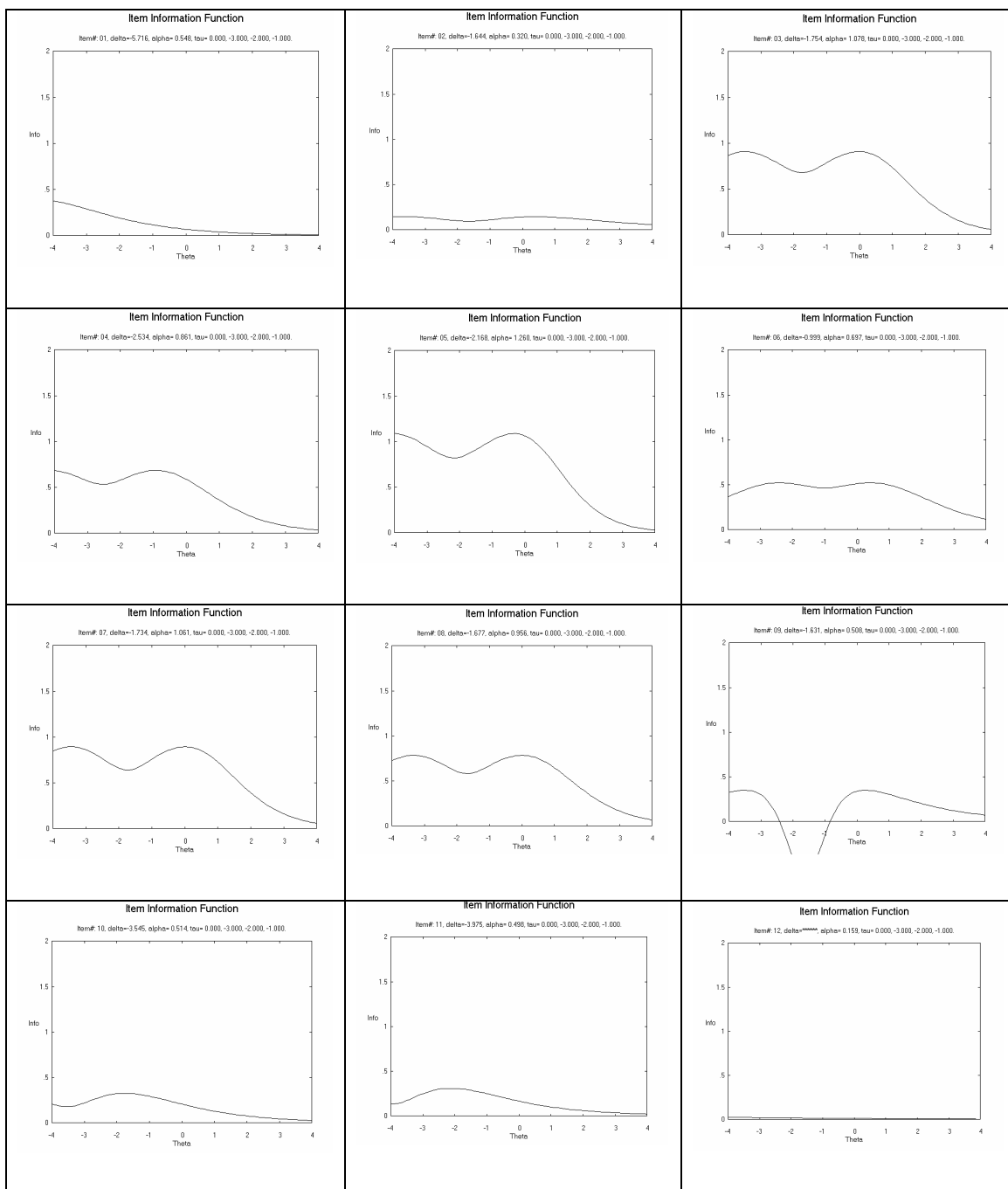
APÊNDICE K – Estatísticas de ajuste do GGUM aos dados da pesquisa de campo no modelo com 4 categorias de resposta e 10 respostas excluídas por respondente.

Item	P-valor G2	infit	outfit
1	0,072	0,407	0,405
2	0,173	0,125	0,186
3	0,950	0,429	0,782
4	0,776	0,646	0,707
5	0,977	0,699	1,000
6	0,990	0,486	0,966
7	0,600	0,695	0,012
8	0,710	0,470	0,899
9	0,760	0,639	0,488
10	0,005	0,418	0,418
11	0,055	0,456	0,454
12	0,494	0,484	0,442
13	0,720	0,706	0,710
14	0,967	0,499	0,497
15	0,257	0,616	0,628
16	0,862	0,416	0,399
17	0,031	0,801	0,783
18	0,050	0,687	0,689
19	0,956	0,440	0,449
20	0,327	0,429	0,429
21	0,072	0,404	0,404
22	0,221	0,417	0,417
23	0,982	0,999	0,999
24	0,766	0,702	0,704
25	0,404	0,914	0,978
Total	0,053		

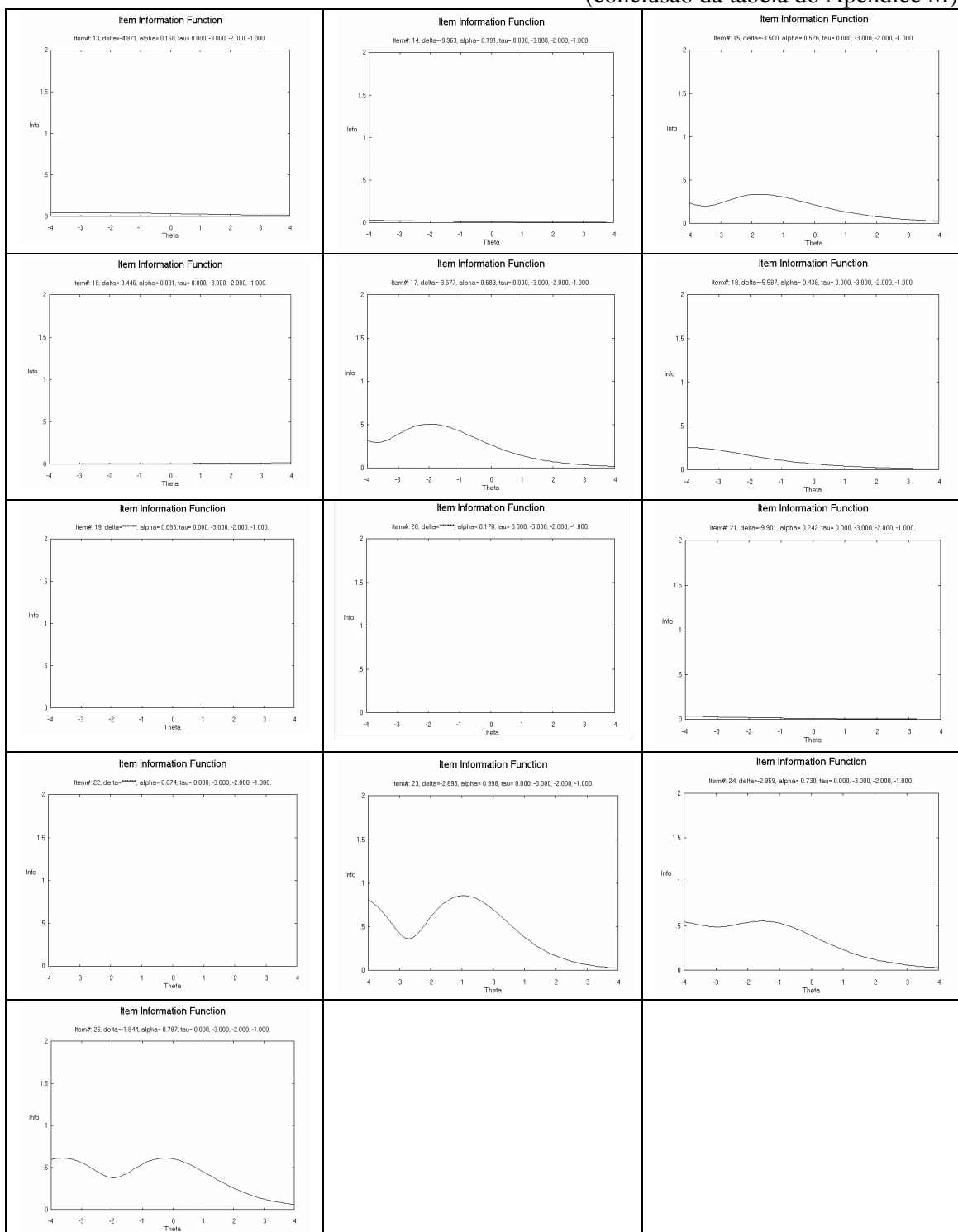
APÊNDICE L – Estatísticas de ajuste do GGUM aos dados da pesquisa de campo no modelo com 4 categorias de resposta e 15 respostas excluídas por respondente.

Item	P-valor G2	infit	outfit
1	0,804	0,638	0,652
2	0,007	0,074	0,071
3	0,990	0,664	0,998
4	0,699	0,406	0,402
5	0,939	0,096	0,112
6	0,997	0,581	0,993
7	0,639	0,112	0,016
8	0,635	0,000	0,000
9	0,940	0,731	0,633
10	0,003	0,814	0,824
11	0,047	0,613	0,631
12	0,502	0,430	0,430
13	0,746	0,480	0,487
14	0,944	0,412	0,412
15	0,653	0,417	0,417
16	0,869	0,391	0,424
17	0,174	0,400	0,400
18	0,133	0,564	0,558
19	0,974	0,508	0,533
20	0,371	0,688	0,696
21	0,222	0,454	0,455
22	0,929	0,408	0,408
23	0,314	0,414	0,413
24	0,901	0,999	1,000
25	0,222	0,596	0,672
Total	0,248		

APÊNDICE M - Função Informação dos atributos para os dados da pesquisa de campo no modelo com 4 categorias de respostas.



(conclusão da tabela do Apêndice M)



APÊNDICE N – Exemplo de quadrado latino de tamanho 25.

7	18	21	3	11	6	10	20	9	13	5	1	19	4	22	23	17	15	2	16	12	24	8	14	25
18	3	7	6	21	20	11	13	10	1	9	4	5	23	19	15	22	16	17	24	2	14	12	25	8
21	7	11	18	10	3	9	6	5	20	19	13	22	1	17	4	2	23	12	15	8	16	25	24	14
3	6	18	20	7	13	21	1	11	4	10	23	9	15	5	16	19	24	22	14	17	25	2	8	12
11	21	10	7	9	18	5	3	19	6	22	20	17	13	2	1	12	4	8	23	25	15	14	16	24
6	20	3	13	18	1	7	4	21	23	11	15	10	16	9	24	5	14	19	25	22	8	17	12	2
10	11	9	21	5	7	19	18	22	3	17	6	2	20	12	13	8	1	25	4	14	23	24	15	16
20	13	6	1	3	4	18	23	7	15	21	16	11	24	10	14	9	25	5	8	19	12	22	2	17
9	10	5	11	19	21	22	7	17	18	2	3	12	6	8	20	25	13	14	1	24	4	16	23	15
13	1	20	4	6	23	3	15	18	16	7	24	21	14	11	25	10	8	9	12	5	2	19	17	22
5	9	19	10	22	11	17	21	2	7	12	18	8	3	25	6	14	20	24	13	16	1	15	4	23
1	4	13	23	20	15	6	16	3	24	18	14	7	25	21	8	11	12	10	2	9	17	5	22	19
19	5	22	9	17	10	2	11	12	21	8	7	25	18	14	3	24	6	16	20	15	13	23	1	4
4	23	1	15	13	16	20	24	6	14	3	25	18	8	7	12	21	2	11	17	10	22	9	19	5
22	19	17	5	2	9	12	10	8	11	25	21	14	7	24	18	16	3	15	6	23	20	4	13	1
23	15	4	16	1	24	13	14	20	25	6	8	3	12	18	2	7	17	21	22	11	19	10	5	9
17	22	2	19	12	5	8	9	25	10	14	11	24	21	16	7	15	18	23	3	4	6	1	20	13
15	16	23	24	4	14	1	25	13	8	20	12	6	2	3	17	18	22	7	19	21	5	11	9	10
2	17	12	22	8	19	25	5	14	9	24	10	16	11	15	21	23	7	4	18	1	3	13	6	20
16	24	15	14	23	25	4	8	1	12	13	2	20	17	6	22	3	19	18	5	7	9	21	10	11
12	2	8	17	25	22	14	19	24	5	16	9	15	10	23	11	4	21	1	7	13	18	20	3	6
24	14	16	25	15	8	23	12	4	2	1	17	13	22	20	19	6	5	3	9	18	10	7	11	21
8	12	25	2	14	17	24	22	16	19	15	5	23	9	4	10	1	11	13	21	20	7	6	18	3
14	25	24	8	16	12	15	2	23	17	4	22	1	19	13	5	20	9	6	10	3	11	18	21	7
25	8	14	12	24	2	16	17	15	22	23	19	4	5	1	9	13	10	20	11	6	21	3	7	18