

## SUMÁRIO

<b>LISTA DE TABELAS .....</b>	<b>IV</b>
<b>LISTA DE ILUSTRAÇÕES .....</b>	<b>V</b>
<b>RESUMO .....</b>	<b>VIII</b>
<b>ABSTRACT .....</b>	<b>IX</b>
<b>1. INTRODUÇÃO.....</b>	<b>1</b>
1.1 TEMAS DE ESTUDO .....	1
1.2 O PROBLEMA .....	3
1.3 OBJETIVOS .....	4
1.4 JUSTIFICATIVA .....	5
1.5 ESTRUTURA DO TRABALHO .....	7
<b>2. REVISÃO DE LITERATURA .....</b>	<b>10</b>
2.1 TEORIA CLÁSSICA DOS TESTES .....	10
2.1.1 Introdução .....	10
2.1.2 Conceção Básica .....	10
2.1.3 Confiabilidade.....	11
2.1.4 Correlação Bisserial e Bisserial por Ponto.....	12
2.2 TEORIA DA RESPOSTA AO ITEM.....	13
2.2.1 Introdução .....	13
2.2.2 Pressupostos Básicos .....	15
2.2.3 Modelo Logístico de Três Parâmetros Unidimensional Dicotômico.....	17
2.2.3.1 Curva característica do item – CCI.....	18
2.2.3.4 Estimação dos parâmetros dos itens e das habilidades dos indivíduos.....	28
2.2.3.4.1 Introdução .....	28
2.2.3.4.2 Estimação dos parâmetros dos itens .....	30
2.2.3.4.3 Estimação das habilidades dos indivíduos.....	45
2.2.3.4.3.1 Estimação por máxima verossimilhança.....	45
2.2.3.4.3.2 Estimação modal bayesiana ou maximum a posteriori (MAP). .....	47
2.2.3.4.3.3 Estimação bayesiana da média a posteriori ou expected a posteriori (EAP). .....	49
2.2.4 Cálculo do Escore Verdadeiro. ....	50
2.3 ANÁLISE FATORIAL.....	51
2.3.1 Introdução .....	51
2.3.2 Conceitos Fundamentais e Pressupostos Básicos .....	52
2.3.3 O Modelo de Análise Fatorial Ortogonal.....	58
2.3.4 Seleção do Número de Fatores. ....	62
2.3.5 Métodos de Estimação do Modelo Fatorial .....	64

2.3.5.1- Método das componentes principais.....	64
2.3.5.2 Método da máxima verossimilhança.....	66
2.3.6 Rotação de Fatores.....	69
2.3.6.1 Método de rotação Varimax.....	70
2.3.7 Escores Fatoriais.....	74
2.3.7.1 Método dos mínimos quadrados ponderados.....	75
2.3.7.2 Método da regressão.....	77
2.3.8 Considerações Adicionais.....	80
2.4 RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO.....	85
2.4.1 Introdução.....	85
2.4.2 Conceitos Gerais de Reconhecimento e Classificação.....	86
2.4.2.1 Custo esperado de erro de classificação (ECM).....	89
2.4.2.2. Probabilidade total de erro de classificação.....	92
2.4.3 Reconhecimento e Classificação para Duas Populações.....	93
2.4.3.1 Função de classificação de Anderson – populações gaussianas e matrizes de correlação iguais.....	93
2.4.3.2. Classificação quadrática – populações não necessariamente gaussianas e matrizes de correlação diferentes.....	95
2.4.3.3. Método discriminante linear de Fischer – populações não necessariamente gaussianas e matrizes de covariâncias iguais.....	96
2.4.3.4. Regressão logística: modelo para variáveis dicotômicas.....	100
2.4.3.4.1. Introdução.....	100
2.4.3.4.2. Modelo logístico linear simples.....	100
2.4.3.4.3. Modelo logístico linear múltiplo.....	101
2.4.4. Reconhecimento e Classificação para Várias Populações.....	102
2.4.4.1. Introdução.....	102
2.4.4.2 Método de máxima verossimilhança para várias populações – populações gaussianas e matrizes de covariâncias diferentes.....	103
2.4.4.3 Método linear de Fisher para várias populações – populações não necessariamente gaussianas e matrizes de covariâncias iguais.....	106
2.4.4.4 Método distância mínima – populações não necessariamente gaussianas e matrizes de covariâncias iguais.....	111
2.4.4.5 Método distância Mahalanobis – populações não necessariamente gaussianas e matrizes de covariâncias iguais.....	112
2.4.5. Avaliação de Desempenho de Classificação.....	113
<b>3 MATERIAIS E MÉTODOS.....</b>	<b>116</b>
3.1 INTRODUÇÃO.....	116
3.2 MÉTODOS ESTATÍSTICOS DE ANÁLISE.....	116
3.2.1. Teoria Clássica dos Testes.....	116
3.2.2. Teoria da Resposta ao Item.....	117
3.2.2 Análise Fatorial.....	117
3.2.3 Reconhecimento de Padrões e Classificação.....	119
3.3 COLETA DE DADOS.....	120
3.3.1 Avaliação de Candidatos ao Vestibular e Evasão de Alunos do curso de Estatística da UFPR.....	120
3.3.1.1 Coleta de dados.....	120

3.3.1.2 Procedimentos de avaliação .....	121
3.3.2 Avaliação de Fornecedores .....	121
3.3.2.1 Procedimentos de avaliação .....	121
3.4 ELABORAÇÃO DO PROGRAMA .....	121
<b>4- RESULTADOS E DISCUSSÃO.....</b>	<b>125</b>
4.1- AVALIAÇÃO DO INSTRUMENTO DO TESTE DA MATÉRIA M DO VESTIBULAR DA UNIVERSIDADE U E RANQUEAMENTO DOS ESCORES VERDADEIROS DOS CANDIDATOS UTILIZANDO A TEORIA DA RESPOSTA AO ITEM .....	125
4.1.1- Análise do Teste da Matéria M.....	125
4.1.2- Análise dos Itens do Teste da Matéria M .....	127
4.1.3- Ranqueamento dos Candidatos utilizando Escores Verdadeiros.....	137
4.2- EVASÃO DE ALUNOS DO CURSO DE ESTATÍSTICA DA UFPR.....	139
4.3- ANÁLISE DE FORNECEDORES .....	143
<b>5. CONCLUSÃO E RECOMENDAÇÕES.....</b>	<b>148</b>
<b>REFERÊNCIAS.....</b>	<b>150</b>
<b>APÊNDICES .....</b>	<b>154</b>

## LISTA DE TABELAS

TABELA 1- EXEMPLO DE ITENS POR INDIVÍDUOS .....	32
TABELA 2- EXEMPLO DE ITENS POR PADRÕES DE RESPOSTA.....	32
TABELA 3- ASSOCIAÇÃO DAS VARIÁVEIS OBSERVÁVEIS $X_1$ E $X_2$ .....	81
TABELA 4- AMOSTRAS DE FAMÍLIAS PROPRIETÁRIAS E NÃO PROPRIETÁRIAS DE CERTO EQUIPAMENTO COM BASE NA RENDA E TAMANHO DO LOTE DE SUAS MORADIAS .....	87
TABELA 5- MATRIZ DO CUSTO DE CLASSIFICAÇÃO .....	90
TABELA 6- TABELA DE CLASSIFICAÇÃO USANDO O MÉTODO DE FISHER .....	113
TABELA 7- TABELA DE CLASSIFICAÇÃO USANDO O MÉTODO DE FISHER E PROCEDIMENTO DE RETENÇÃO DE LACHENBRUCH .....	115
TABELA 8- ESTATÍSTICAS DESCRITIVAS NÚMERO DE ACERTO DOS CANDIDATOS.....	126
TABELA 9- ANÁLISE DOS ITENS DO TESTE DA MATÉRIA M DO VESTIBULAR DA UNIVERSIDADE U .....	128
TABELA 10- RANQUEAMENTO DOS ESCORES VERDADEIROS.....	138
TABELA 11- ESTATÍSTICAS DESCRITIVAS DOS ESCORES VERDADEIROS DOS CANDIDATOS .....	139
TABELA 12- TABELA DE CLASSIFICAÇÃO FEITA PELO MÉTODO LINEAR DE FISHER .....	140
TABELA 13- CLASSIFICAÇÃO DO GRUPO DE TESTE.....	142
TABELA 14- COMPARAÇÃO ENTRE OS MÉTODOS DE FISHER, DISCRIMINANTE QUADRÁTICO E REDES NEURAIS. ....	143
TABELA 15- ANÁLISE FATORIAL – MÉTODO DAS COMPONENTES PRINCIPAIS .....	145
TABELA 16- CLASSIFICAÇÃO DAS EMPRESAS FORNECEDORAS .....	146
TABELA 17- TABELA DE CLASSIFICAÇÃO USANDO O PROCEDIMENTO DE RETENÇÃO DE LACHENBRUCH.....	147

## LISTA DE ILUSTRAÇÕES

FIGURA 1- GRÁFICO DA CURVA CARACTERÍSTICA DO ITEM 1 .....	20
FIGURA 2- GRÁFICO DA CURVA CARACTERÍSTICA DO ITEM 2 .....	20
FIGURA 3- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS 1 E 2 .....	21
FIGURA 4- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS 3 E 4 .....	21
FIGURA 5- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS 1 E 5 .....	22
FIGURA 6- GRÁFICO DA CURVA CARACTERÍSTICA E DA FUNÇÃO DE INFORMAÇÃO DO ITEM 6 .....	23
FIGURA 7- GRÁFICO DA CURVA CARACTERÍSTICA E DA FUNÇÃO DE INFORMAÇÃO DO ITEM 7 .....	23
FIGURA 8- GRÁFICO DA CURVA CARACTERÍSTICA E DA FUNÇÃO DE INFORMAÇÃO DO ITEM 8 .....	24
FIGURA 9- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS DO INSTRUMENTO A .....	26
FIGURA 10- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS DO INSTRUMENTO B .....	26
FIGURA 11- INTERPRETAÇÃO GEOMÉTRICA DAS COMPONENTES PRINCIPAIS .....	54

<b>FIGURA 12- GRÁFICO DAS AMOSTRAS DE FAMÍLIAS PROPRIETÁRIAS E NÃO-PROPRIETÁRIAS COM BASE NA RENDA E TAMANHO DOS LOTES DE SUAS MORADIAS .....</b>	<b>87</b>
<b>FIGURA 13- REGIÕES DE CLASSIFICAÇÃO PARA DUAS POPULAÇÕES</b>	<b>88</b>
<b>FIGURA 14- CLASSIFICAÇÃO DAS REGIÕES PARA DUAS POPULAÇÕES .....</b>	<b>89</b>
<b>FIGURA 15- ROTAÇÃO DOS FATORES .....</b>	<b>118</b>
<b>FIGURA 16- FLUXOGRAMA DO PROGRAMA SIAVAL 1.0 .....</b>	<b>123</b>
<b>FIGURA 17- RELATÓRIO DE SAÍDA DA ANÁLISE FATORIAL.....</b>	<b>124</b>
<b>FIGURA 18- RELATÓRIO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM NA TEORIA DA RESPOSTA AO ITEM .....</b>	<b>124</b>
<b>FIGURA 19- HISTOGRAMA DO NÚMERO DE ACERTOS DOS CANDIDATOS .....</b>	<b>126</b>
<b>FIGURA 20- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 13 .....</b>	<b>131</b>
<b>FIGURA 21- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 36 .....</b>	<b>131</b>
<b>FIGURA 22- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 37 .....</b>	<b>132</b>
<b>FIGURA 23- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 43 .....</b>	<b>132</b>
<b>FIGURA 24- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 48 .....</b>	<b>133</b>

<b>FIGURA 25- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 54 .....</b>	<b>134</b>
<b>FIGURA 26- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 55 .....</b>	<b>134</b>
<b>FIGURA 27- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 60 .....</b>	<b>135</b>
<b>FIGURA 28- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 57 .....</b>	<b>136</b>
<b>FIGURA 29- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 35 .....</b>	<b>136</b>
<b>FIGURA 30- DESCRIÇÃO DOS ESCORES VERDADEIROS .....</b>	<b>138</b>

## RESUMO

Na última década, a utilização de métodos estatísticos multivariados passou a ter papel importante na análise de desempenho. Ao estudar esses métodos, verificou-se que a Teoria Clássica dos Testes (TCT) e a Teoria da Resposta aos Itens (TRI) têm um papel importante nesse tipo de análise. Quando integradas, as técnicas de Análise Fatorial e de Reconhecimento de Padrões e Classificação fornece ao pesquisador uma gama maior de opções de estudo e aplicações. Apresentam-se nesse trabalho três aplicações das técnicas estatísticas supracitadas: a primeira analisou um teste de vestibular através da TCT e TRI, sugerindo o emprego de escores verdadeiros para o ranqueamento dos candidatos, a segunda criou um procedimento para detectar alunos com possibilidades de evasão após o primeiro ano de curso de graduação aplicando o Método Discriminante Linear de Fisher e a terceira propôs uma aplicação na área da Engenharia de Produção, que atualmente tem sido assunto de grande interesse para as empresas. Ou seja, propiciando o ranqueamento de fornecedores e a possibilidade de premiação, estimulando a competitividade e a melhoria da qualidade. Para esta última aplicação utilizou-se escores fatoriais. Sugeriu-se que as empresas classifiquem seus novos fornecedores em grupos pré-definidos com a utilização do Método Discriminante Linear de Fisher. Um dos produtos resultantes deste trabalho é o programa chamado de SIAVAL 1.0, que possui uma interface amigável com o usuário e incorpora os métodos: Teoria Clássica dos Testes, Teoria da Resposta ao Item, Análise Fatorial e Reconhecimento de Padrões e Classificação. Portanto, o programa e as aplicações sugeridas podem proporcionar um avanço importante para análise de desempenho em qualquer área.

Palavras-chave: Teoria Clássica dos Testes, Teoria da Resposta ao Item, Análise Fatorial, Reconhecimento de Padrões e Classificação

## ABSTRACT

*In the last decade, the use of multivariate statistical methods started to have important paper in the acting analysis. When studying those methods, it was verified that the Classic Test Theory (CTT) and the Item Response Theory (IRT) have had an important paper in that analysis type. When integrated the techniques of Factorial Analysis and Pattern Recognition and Classification, it supplies the researcher a larger range of study options and applications. This work presents three applications of the statistical techniques mentioned above: the first analyzed a university entrance examination through CTT and IRT, suggesting the job of true scores for candidates' ranking, the second one created a procedure to detect students with drop out possibilities after the first year of degree course applying the Fisher Linear Discriminant Method and the third proposed an application in the area of the Production Engineering, that now has been subject of great interest for the companies. In other words, propitiating the ranqueamento of suppliers and the award possibility, stimulating the competitiveness and the improvement of the quality. For this last application it was used factorial scores. It suggested that the companies classify their new suppliers in pré-defined groups with the use of the Fisher Linear Discriminant Method. One of the resulting products of this work is the software call of SIAVAL 1.0, that has a friendly interface with the user and incorporates the methods: Classic Test Theory, Item Response Theory, Factorial Analysis and Pattern Recognition and Classification. Therefore, the program and the suggested applications can provide an important progress for acting analysis in any area.*

*Keywords: Classic Test Theory, Item Response Theory, Factorial Analysis, Pattern Recognition and Classification.*

# **1. INTRODUÇÃO**

## **1.1 TEMAS DE ESTUDO**

Avaliação é um processo sistemático e metodológico que através da atribuição de um juízo de valor, visa explicar e compreender fenômenos com o intuito de refletir, rever conceitos, redefinir estratégias, superar dificuldades, transformar, acompanhar, adequar necessidades ao meio, impulsionar a melhoria da qualidade, constatar conhecimentos, interpretar estruturas, verificar processos ou situações que ocorrem com indivíduos ou objetos nas várias áreas do conhecimento.

O processo avaliativo implica numa metodologia que busca, através de variáveis quantitativas e qualitativas, modelos que possibilitam inferências fidedignas.

A palavra avaliação causa sempre algum impacto nos indivíduos envolvidos no processo, ou seja, receio nos que serão avaliados e preocupação nos que avaliarão. Os primeiros, por razões óbvias, se sentem apreensivos na tentativa de atingir objetivos e inseguros quanto ao instrumento imposto pelo avaliador. Os segundos em alguns momentos não se sentem preparados para fazer a avaliação, dado a complexidade de verificar habilidades e em alguns casos o desconhecimento de técnicas estatísticas, entre outras. Na maioria das vezes o avaliador entende muito do conteúdo a ser avaliado, mas se defronta com a dificuldade de elaboração e validação do instrumento para fazê-lo. Além disso, após aplicar o instrumento o avaliador se vê com um conjunto de itens respondidos e frente ao desconhecimento das possibilidades existentes em analisá-los. Então, parte para procedimentos do tipo de ponderações dos itens, valorizando aqueles que consideram por experiência mais importante. Estas técnicas sempre são questionáveis e exigem muitas vezes a apresentação prévia dos instrumentos aos avaliados para que estes se sintam confiantes nos procedimentos adotados.

Várias áreas do conhecimento fazem avaliações e em quase todos os lugares indivíduos ou produtos estão sendo avaliados. Os modelos de análise de dados em avaliação têm origem nas áreas de psicologia e educação, mas podem ser aplicados em outros campos do conhecimento. Há vários exemplos presentes na nossa vida diária em que poderiam ser usados os métodos que serão apresentados neste trabalho, mas não o são pelas dificuldades já mencionadas. Seguem alguns exemplos.

- Ao ingressar em uma academia de ginástica, natação, musculação, etc.; é preciso fazer uma avaliação física para designar o tipo de esforço suportado pelo corpo. Essa avaliação se bem feita implica na realização além de exames médicos e de fisioterapia e também o questionamento sobre o tipo de vida do indivíduo, gerando um conjunto considerável de itens a ser analisado.
- No lançamento de um novo produto de linha branca (geladeiras, máquinas de lavar, etc.) no mercado a estratégia de avaliação do produto é selecionar pessoas ou famílias que ficarão com o produto por cerca de um ano. Neste período e imediatamente após, são feitas entrevistas e respondidos questionários. A análise deste conjunto de itens coletados proporciona aos empresários a retificação de problemas, o tipo de clientela que vai ser atingida, orientando a equipe de marketing numa estratégia de propaganda, uma estimativa da quantidade de peças que devem ser fabricadas, o nível da satisfação com o produto e a viabilidade de fabricação do mesmo.
- Na necessidade ou desejo de se fazer um regime alimentar, primeiramente é feito por um bom nutricionista uma avaliação física, de hábitos e às vezes até psicológica para verificar o tipo de tratamento indicado ao indivíduo.
- Avaliar Projetos tem sido um tema bastante atual cujos objetivos são: determinar os melhores projetos, verificar se um projeto em qualquer área está alcançando ou alcançou os resultados esperados, determinar a importância de itens como custo, tempo de execução, entre outros.
- Nos últimos anos cada vez mais as empresas avaliam seus colaboradores e algumas delas vinculam a remuneração dos funcionários ao desempenho dos mesmos, isto é, além do pagamento dos salários são incluídos pagamentos referentes a gratificações e a distribuição partes dos resultados;
- Avaliação de Fornecedores é outra atividade que vem recentemente preocupando as empresas. De modo geral somente empresas de grande porte fazem este tipo de avaliação e isto tem induzido à redução de custos e melhoria da qualidade dos produtos e processos.

Um dos objetivos deste trabalho é o de atender usuários leigos que necessitem fazer avaliação de desempenho e para isso foi desenvolvido um sistema chamado de

SIAVAL 1.0 que facilita a análise de itens de um instrumento dicotômico, com ranqueamento e reconhecimento de indivíduos (empresas ou objetos) e classificação, integrando quatro importantes áreas que são: Teoria Clássica dos Testes, Teoria da Resposta ao Item, Análise Fatorial e Reconhecimento de Padrões e Classificação. Com o intuito de exemplificar e de verificar o desempenho do programa proposto foram apresentados exemplos em duas áreas do conhecimento: educação (desempenho de candidatos em um teste de Concurso Vestibular e análise de alunos de graduação com risco de evasão) e engenharia de produção (avaliação de fornecedores de empresas).

## 1.2 O PROBLEMA

Apesar da existência de aplicações da Teoria Clássica dos Testes, Teoria da Resposta ao Item, Análise Fatorial e Reconhecimento de Padrões em avaliação, encontram-se dificuldades na sua compreensão, execução e interpretação. Estas dificuldades, só podem ser sanadas com muito estudo e amplo conhecimento de estatística. Estudiosos das áreas de aplicação que poderiam estar usufruindo destas técnicas não o fazem por estes motivos.

A Teoria Clássica dos Testes e a Teoria da Resposta ao Item (TRI), indispensáveis em qualquer tipo de avaliação, vêm sendo utilizada há pouco tempo e por poucas pessoas na UFPR, no Estado do Paraná e também no País. É preciso incentivar, no Brasil, as pessoas que trabalham com avaliação de todas as áreas a usarem estas técnicas principalmente na análise dos itens de seus instrumentos.

Muitos programas computacionais têm sido desenvolvidos para utilização da TRI nos últimos tempos, tais como BILOG, TESTFACT, PARSCALE, MULTILOG, LOGIST, etc. Aqui no Brasil, devido ao pouco tempo de estudo da teoria não existe nenhum programa brasileiro disponível no mercado. Estes programas atendem as necessidades dos atuais usuários da TRI, porém o manuseio e a análise dos dados devem ser feitas por uma pessoa especializada e, também, não existe uma interface com os sistemas de avaliação existentes nos núcleos de avaliação.

Análise Fatorial e Reconhecimento de Padrões e de Classificação são técnicas estatísticas muito conhecidas, porém sua utilização só foi possível quando os recursos

computacionais se tornaram acessíveis, velozes e com grande capacidade de memória. Existem muitos programas que fazem estas duas análises (STATGRAPHICS, STATISTICA, SPSS, R, etc.), porém ainda são muito caros e requerem um conhecimento de uso do programa e de análise dos resultados.

Muitas avaliações feitas não utilizam técnicas estatísticas nas suas análises, devido ao desconhecimento dos usuários dos métodos que podem usufruir e o difícil acesso aos programas computacionais.

Não existe um programa de avaliação que tenha integrado as três teorias para atender as necessidades do Núcleo de Concurso da Universidade Federal do Paraná e seria muito útil a ele e às outras instituições que mantenham processos de aplicação de concursos em vários níveis.

Na área da engenharia, verificou-se em visitas feitas as empresas de grande porte que várias delas fazem avaliação de fornecedores, porém sentem necessidade de que os seus fornecedores também avaliem os fornecedores deles, isto é chamado de cadeia da qualidade. Detectou-se também em pesquisas bibliográficas poucos trabalhos neste assunto e falta uma proposta de ranqueamento de fornecedores que atenda as médias e pequenas empresas .

### 1.3 OBJETIVOS

O propósito principal deste trabalho é avaliar o uso das técnicas estatísticas e apresentar, às pessoas envolvidas ou que venham a se envolver com avaliação de desempenho, estas técnicas, de forma que auxilie e facilite a análise.

Outro objetivo deste trabalho é agregar Análise Fatorial e Reconhecimento de Padrões e Classificação às técnicas já utilizadas em avaliação de desempenho, que são: Teoria Clássica dos Testes e Teoria da Resposta ao Item; com o intuito de melhorar a qualidade das avaliações no país. Assim, pretende-se construir uma ferramenta que englobe estas técnicas possibilitando ao usuário fazer análise de instrumentos de avaliação, ranqueamento de indivíduos (ou empresas) e classificar um novo indivíduo em alguma categoria pré-determinada.

Criar um sistema integrando a Teoria Clássica dos Testes, Teoria de Resposta ao Item, a Análise Fatorial e o Reconhecimento de Padrões. Implementar computacionalmente, aplicar e validar o modelo proposto.

O programa proposto pretende ser de fácil utilização tanto para os especialistas quanto para pessoas leigas e gera relatórios contendo as interpretações dos resultados e gráficos que proporcionem, aos responsáveis pela avaliação, compreensão e decisões rápidas.

Assim, com intuito de apresentar algumas possibilidades do programa, avaliou-se o desempenho dos candidatos ao vestibular de uma Universidade X, fez-se o ranqueamento dos fornecedores de uma Empresa Y e previu-se uma possível evasão de alunos a partir de variáveis observadas no término do primeiro semestre do curso de Estatística da Universidade Federal do Paraná.

#### 1.4 JUSTIFICATIVA

Nas últimas décadas, segundo notas do editor em HAMBLETON, SWAMINATHAN & ROGERS (1991, p. vii), tem-se observado uma revolução em medidas educacionais e psicológicas com a aplicação da Teoria da Resposta ao Item (TRI). Além disso, a TRI está sendo usada comumente por grandes companhias no planejamento de testes, montagem de teste, calibração de teste, construção de um banco de itens, investigação do viés dos itens do teste e outros procedimentos comuns no processo de desenvolvimento de um instrumento.

Escolheu-se neste trabalho a análise de itens com respostas dicotômicos ou dicotomizadas dado que existe uma necessidade imediata do Núcleo de Concursos da UFPR. Porém, espera-se que esta seja uma primeira versão do programa e que as subseqüentes incluam outros tipos de modelos. Esta primeira versão é experimental, mas pretende-se deixá-la na forma comercial.

A opção pelos Modelos Logísticos de Dois e Três Parâmetros Unidimensionais Dicotômicos deve-se ao fato de ser este o modelo mais usado e recomendado pelos pesquisadores. Se em algum caso não houver a possibilidade de respostas de acerto ao

acaso pode-se aplicar o Modelo Logístico de Dois Parâmetros (caso da Avaliação de Fornecedores).

A estimação dos parâmetros dos itens foi feita através do Método de Máxima Verossimilhança Marginal via algoritmo EM com as reformulações sugeridas por BOCK & AITKIN (1981, p. 443-459) que possibilita uma melhor convergência quando utilizado o método de resoluções de equações não-lineares de Newton-Raphson.

A estimação das habilidades dos indivíduos (empresas) foi realizada após a estimação dos parâmetros dos itens utilizando o estimador de Baysiano EAP (Expected A Posteriori), que no caso é o mais indicado dado que não necessita usar métodos de resolução de equações e deixa o programa mais rápido.

A Análise Fatorial foi aplicada no ranqueamento de fornecedores de uma empresa de grande porte. A partir de dados fornecidos por empresa foi avaliado o grau de relacionamento entre eles, com base nos coeficientes de correlação. Da matriz de correlação obtida foram determinados os autovalores e autovetores, dos quais se obteve o número de fatores considerando-se 90% da proporção da variância explicada. Na estimação dos fatores foi utilizado o Método de Componentes Principais, e para uma melhor interpretação dos resultados optou-se fazer uma Rotação Varimax. Os escores fatoriais foram estimados pelo Método dos Mínimos Quadrados. A diversidade de métodos e procedimentos necessários ao fazer uma análise fatorial dificulta a escolha dos mesmos, sendo assim optou-se nos que tem apresentado melhores resultados práticos.

A escolha do método de Reconhecimento de Padrões e de Classificação utilizado para alocar e reconhecer um novo indivíduo em alguma categoria predeterminada aplica-se bem as de análises pretendidas pelo programa, visto que serão analisados dados de uma mesma população. Por isso o programa contempla o Método Linear de Fisher para Várias Populações, dado não há necessidade de verificação da distribuição de probabilidade dos dados.

Nos últimos anos uma das preocupações das grandes empresas está focada no estreitamento do relacionamento com seus fornecedores, pois o fortalecimento deste vínculo, também chamado de parceria, garante às empresas uma melhoria de qualidade e preço tornando-as mais competitivas no mercado. Segundo ALVAREZ & QUEIROZ (2003, p.1), inicialmente os processos de parceria eram feitos em cima de Contratos de

Fornecimento que implicava num longo período de avaliação e qualificação do fornecedor. Isto gerava uma série de responsabilidades e compromissos impostos pelas empresas aos fornecedores e acarretava conflitos de interesses ao longo do processo. Por este motivo algumas empresas, preocupadas com estas questões, criaram processos de integração e tornaram-se mais flexíveis, de maneira que as expectativas de ambas as partes possam obter benefícios mútuos.

A avaliação de fornecedores assegura à empresa uma melhoria substancial no seu processo de produção principalmente garantindo uma melhor qualidade dos insumos e redução dos prazos de entrega. Assim, a área de Gestão Empresarial está envolvida nesta avaliação, principalmente porque o benefício maior da empresa está focado na redução de custos, portanto geralmente é ela que conduz o tema.

Procurou-se seguir no desenvolvimento dos métodos as notações mais utilizadas, principalmente no caso da Teoria da Resposta ao Item, onde as notações e as seqüências das idéias estão embasadas no livro dos autores Dalton Francisco de Andrade, Héilton Tavares e Raquel da Cunha Valle intitulado “Teoria da Resposta ao Item: conceitos e aplicações”, citados várias vezes nesta tese, dado à sua facilidade de compreensão e a intenção de preservar a mesma nomenclatura (terminologia) aqui no Brasil.

## 1.5 ESTRUTURA DO TRABALHO

O presente trabalho de tese se encontra estruturado nos capítulos da Introdução, Revisão de Literatura, Materiais e Métodos, Análise dos Dados, Conclusões e Recomendações e Referências, bem como, os Apêndices.

Na Introdução, Capítulo 1, definiu-se sucintamente avaliação, as técnicas propostas neste trabalho e os campos de aplicação em que poderiam ser usadas. As dificuldades de aplicação das técnicas se resumem no desconhecimento das mesmas e nos obstáculos das aplicações dos programas encontrados no mercado. A razão e a finalidade para propor a implementação de um programa de avaliação integrando as técnicas de Teoria da Resposta ao Item, Análise Fatorial e Reconhecimento de Padrões e Classificação, com aplicações em duas áreas (educação e engenharia de produção) são apresentadas neste capítulo.

As técnicas estatísticas dispostas nesta tese no Capítulo 2 fundamentam os meios de avaliar o instrumento, categorizar os avaliados e classificar um novo indivíduo em alguma categoria de interesse na avaliação.

Na seção 2.1, apresentam-se algumas estatísticas da Teoria Clássica dos Testes que podem auxiliar muito na avaliação, como por exemplo, a confiabilidade (fidedignidade) do instrumento, correlação bisserial e correlação bisserial por ponto.

A Teoria da Resposta ao Item que possibilita a análise dos itens do instrumento é discutida na seção 2.2. A princípio, foram definidos os pressupostos básicos, com a inclusão de um breve histórico da teoria. Em seguida, foram enunciados: o Modelo Logístico de Três Parâmetros Unidimensional Dicotômico com a respectiva Curva Característica do Item (CCI), a Função de Informação do Item, a Função de Informação do Instrumento, o método de estimação dos parâmetros dos itens e o método de estimação das habilidades dos indivíduos.

A técnica da Análise Multivariada, utilizada no ranqueamento, é apresentada na seção 2.3. Nesta seção, são retratados alguns conceitos fundamentais e pressupostos básicos, o Modelo de Análise Fatorial Ortogonal, o critério de seleção do número de fatores e dois métodos de estimação: Método das Componentes Principais e Método da Máxima Verossimilhança. Além disso, citou-se o Método de Rotação Varimax e dois métodos de obtenção do escores fatoriais: Método dos Mínimos Quadrados Ponderados e Método da Regressão.

A alocação de um novo indivíduo em uma classe pré-determinada é feita pelas técnicas de Reconhecimento de Padrões e Classificação, relatada na seção 2.4, que inicialmente apresenta os conceitos e procedimentos necessários ao seu entendimento. Continuando a seção, descreve-se o Método de Fischer para discriminação e classificação entre duas ou mais populações, o Método de Máxima Verossimilhança para várias populações supostamente Gaussianas entre outros.

No Capítulo 3 descrevem-se os materiais e métodos utilizados neste trabalho. Está dividido nas seguintes seções: Introdução, Métodos Estatísticos Utilizados, Coletas de Dados e Elaboração do Programa.

Os resultados das análises de dados reais utilizando o programa encontram-se descritos no Capítulo 4.

Finalmente as conclusões e recomendações são apresentadas no Capítulo 5 e na seqüência seguem as Referências e os Apêndices.

## 2. REVISÃO DE LITERATURA

### 2.1 TEORIA CLÁSSICA DOS TESTES

#### 2.1.1 Introdução

A análise de desempenho em testes de avaliação tem sido objeto de muitos estudos em várias áreas nas últimas décadas e nos campos da psicometria e de medidas educacionais. Há muito tempo modelos vêm sendo desenvolvidos com as seguintes finalidades: avaliar as características latentes de indivíduos, fazer comparações entre grupos de indivíduos e possibilitar a análise dos instrumentos utilizados. O modelo mais popular para testes é o da Teoria Clássica dos Testes (TCT), porém com um número de deficiências que limita a sua utilização nos tempos atuais, dado que a Teoria da Resposta ao Item (TRI) atende melhor às expectativas dos pesquisadores.

No entanto, alguns elementos da TCT ainda são importantes na interpretação dos resultados de um teste e serão, brevemente, abordados aqui.

Segundo Hwang (2002, p.27) os primeiros trabalhos desenvolvidos sobre este tema foram os de Spearman (1907, 1913), Gulliksens (1950) e posteriormente Lord and Novick (1968) fizeram uma exposição completa da teoria.

#### 2.1.2 Concepção Básica

A Teoria Clássica dos Testes (TCT) consiste em um modelo relativamente simples. Serão demonstrados aqui sumariamente os princípios desta teoria necessários à complementação dos estudos feitos através da Teoria da Resposta ao Item descrita no próximo item.

O modelo se fundamenta na observação dos escores,  $X_{ij}$ , do respondente  $i$  no item  $j$ , segundo LORD & NOVICK (1968, p.34). Assim, o modelo matemático é definido por:

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

onde a constante  $\mu_i$  é o verdadeiro escore do respondente  $i$  onde  $i = 1, 2, 3, \dots, n$  (desconhecido) e  $\varepsilon_{ij}$  é uma perturbação aleatória ou erro de observação na resposta do avaliado  $i$  no item  $j$ .

A definição do escore verdadeiro na TCT é fundamental e é dado por:

$$\mu_i = E(X_{ij})$$

Dessa forma tem-se uma distribuição de probabilidade para a variável aleatória do verdadeiro escore dada por  $f(x_{ij})$  com média  $\mu_i$  e variância  $V(X_{ij}) = V(\varepsilon_{ij})$ . Dado que  $\mu_i$  é uma constante, têm-se que o erro ou perturbação aleatória possui uma distribuição dada por  $f(\varepsilon_{ij})$  com média zero apresentada a seguir e variância  $V(\varepsilon_{ij}) = V(X_{ij}) = \sigma_i^2$ . E, sendo  $\varepsilon_{ij} = X_{ij} - \mu_i$ , tem-se que:

$$E(\varepsilon_{ij}) = E(X_{ij} - \mu_i) = E(X_{ij}) - E(\mu_i) = \mu_i - \mu_i = 0 \quad (1)$$

e

$$V(\varepsilon_{ij}) = V(X_{ij} - \mu_i) = \sigma_i^2 \quad (2)$$

O elemento chave na TCT vem da definição da mensuração paralela, especialmente na análise da precisão de mensuração. Duas medidas distintas  $X_{ij}$  e  $X_{ij^*}$  são ditas mensurações paralelas se:

$$E(X_{ij}) = E(X_{ij^*}) = \mu_i \quad (3)$$

$$V(X_{ij}) = V(X_{ij^*}) = \sigma_i^2 \quad \text{para } j \neq j^* \quad (4)$$

Assim, o mensuramento paralelo assume que os escores observados dos indivíduos tem a mesma média e mesma variância. A estimação desses parâmetros média e variância com base em uma amostra de tamanho  $n$  são respectivamente:

$$\hat{\mu}_i = \bar{X}_i = \sum_{i=1}^n \frac{X_{ij}}{n} \quad \text{e} \quad \hat{\sigma}_i^2 = \frac{\sum_{i=1}^n (X_{ij} - \bar{X})^2}{n-1} \quad (5)$$

A prova deste resultado encontra-se em LORD & NOVICK (1968, p.48).

### 2.1.3 Confiabilidade

Um conceito muito importante na TCT é o da confiabilidade, pois determina a fidedignidade com que o escore no teste é mensurado. De acordo com CHAVES NETO & TURIM (2003), a confiabilidade de um teste depende da interação entre três fatores: do próprio teste, das condições de aplicação e do grupo de respondentes. Sabe-se, portanto, que quando um grande número de itens é aplicado têm-se escores mais confiáveis, visto que há maior representatividade na amostra de conteúdos, ou seja, existe validade de conteúdo.

Neste trabalho analisou-se respostas dicotômicas, portanto, tem-se uma distribuição discreta de probabilidade do tipo Bernoulli cuja média é dada por  $\mu_i = p_i$  e a variância por  $\sigma_i^2 = p_i (1-p_i)$ , onde  $p_i$  representa a proporção dos examinados que responderam corretamente ao item  $i$ . O coeficiente de correlação de Kuder-Richardson (DOWNIE & HEATH, 1959 p. 195) é um bom estimador da confiabilidade e é dado por:

$$\rho_{KR} = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n p_i (1-p_i)}{s^2} \right) \quad (6)$$

onde:

$n$  = número de itens do teste

$p_i$  = proporção dos respondentes que acertaram o item  $i$

$s^2$  = variância dos escores do teste com base nos  $N$  escores dos respondentes

Um teste é considerado confiável quando a correlação de Kuder-Richardson for maior ou igual a 0,90 (CHAVES & TURIN, 2005 p.51).

#### 2.1.4 Correlação Bisserial e Bisserial por Ponto

A correlação bisserial, denotada por  $r_{bis}$ , é apropriada quando se tem interesse no grau de relacionamento entre dois tipos de escalas, mas por uma razão lógica uma das duas é mais sensivelmente interpretada como uma escala nominal criada artificialmente. A correlação bisserial por ponto,  $r_{bpi}$ , é apropriadamente aplicada quando a variável nominal ocorre naturalmente, por exemplo, na distribuição do gênero (masculino e feminino).

As correlações bisserial e bisserial por ponto são medidas estatísticas que medem a relação existente entre o resultado de um item em particular com o resultado do teste (escore bruto total), isto é, mede a capacidade de discriminação do item em relação ao resultado do teste.

A correlação bisserial por ponto é uma estimativa natural dada por:

$$r_{bpi} = \frac{\bar{X}_i - \bar{X}_t}{s_t} \cdot \sqrt{\frac{p_i}{1-p_i}} \quad (7)$$

onde:

$\bar{X}_i$  = média dos escores do item  $i$ .

$\bar{X}_t$  = média total dos escores.

$s_t$  = desvio padrão total dos escores.

$p_i$  = proporção dos respondentes que acertaram o item  $i$ .

O coeficiente de correlação bisserial considera  $Z \sim N(0,1)$  uma variável aleatória (não observada) associada à habilidade do respondente. Considerando que o escore bruto do respondente se associa linearmente a esta variável aleatória. De acordo com SOARES (2005, p.84), dado a linearidade aplica-se o coeficiente de correlação de Pearson resultando na equação a seguir:

$$r_{bis} = \frac{\bar{X}_i - \bar{X}_t}{s_t} \cdot \frac{p_i}{y} \quad (8)$$

onde  $y = f(z)$  é a ordenada da curva normal padrão correspondente a  $p_i$  e os outros termos são os mesmos usados na correlação bisserial por ponto. Assim, existe uma relação entre as duas correlações (bisserial e bisserial por ponto) dada por:

$$r_{bis} = r_{bpi} \cdot \frac{\sqrt{p_i \cdot (1 - p_i)}}{y} \quad (9)$$

Outros índices são estudados na TCT, como o grau de dificuldade e o índice de discriminação, ambos com base na proporção de acerto dos escores. No entanto, estes índices são tratados como parâmetros e analisados na Teoria da Resposta ao Item, descrita no item 2.2 a seguir.

## 2.2 TEORIA DA RESPOSTA AO ITEM

### 2.2.1 Introdução

A Teoria da Resposta ao Item (TRI) foi apresentada primeiramente por Lord na década de 50. Porém devido às grandes dificuldades computacionais da época passaram a ser aplicados de fato à cerca de 20 anos. No Brasil, segundo ANDRADE, TAVARES & VALLE (2000, p.5), a TRI foi usada a partir de 1995, a princípio, na análise dos dados do Sistema Nacional de Ensino Básico – SAEB.

Inicialmente a TRI foi concebida para resolver problemas na área de avaliação educacional e da psicometria, porém, tais conceitos têm despertado interesse também em outras áreas, tais como: na área médica vista no artigo *Application of the Rasch Analysis: exploring differences in depression between african-american and white children* (DeROOS & ALLEN-MEARES, 1998, p.93-107); no campo da genética no artigo de TAVARES, ANDRADE & PEREIRA (2004, p.679-685) intitulado *Detection of determinant genes and diagnostic via item response theory*; em marketing na Dissertação de Mestrado “Técnica derivada da teoria da resposta ao item aplicada ao setor de serviços” (COSTA, 2001); em gestão da qualidade no artigo de ALEXANDRE, ANDRADE, VASCONCELOS & ARAÚJO (2002, p.129-141) denominado “Uma proposta de análise de um construto para a medição dos fatores críticos da gestão pela qualidade através da teoria da resposta ao item”.

Os primeiros passos no desenvolvimento da Teoria da Resposta ao Item (TRI), segundo HISTORY OF ITEM RESPONSE THEORY disponíveis em <http://www.uic.edu/classes/ot/ot540/history.html> e acessado em abril de 2005, foram dados por Binet & Simon em 1916, quando traçaram níveis de desempenho de indivíduos com algumas variáveis independentes e usaram estas relações para o melhoramento de testes. Em 1936, Richardson introduziu a estimativa do parâmetro da TRI (modelo unidimensional com um parâmetro) que derivava dos conceitos da Teoria Clássica dos Testes (TCT). Lawley introduziu novos procedimentos para a estimação deste parâmetro alguns anos depois (1944). Contudo, na maioria das obras pesquisadas, é dado a Lord o desígnio de precursor dos estudos da TRI, pois ele apresentou o modelo unidimensional de dois parâmetros com a utilização da função distribuição da normal, em 1952 nos Estados Unidos da América.

Segundo LINDEN & HAMBLETON (1996, p. 13), por volta de 1958, Birnbaum substituiu a função distribuição da normal do modelo de Lord pela função logística, o que deu maior facilidade no tratamento matemático nos programas desenvolvidos naquela época.

Rasch em 1960, segundo ANDERSEN & OLSEN (2001, p.11-20), propôs um modelo baseado na concepção de uma distribuição de Poisson para o número de erros em leituras de textos, que com aprimoramentos, como o uso da função logística, tornou-se o

modelo conhecido como modelo logístico de um parâmetro-ML1 e é chamado de Modelo Rasch, motivando o desenvolvimento do modelo com dois parâmetros (ML2) ao modelo inicialmente proposto por Lord.

Segundo LINDEN & HAMPLETON (1996, p.13) foi Birnbaum que implementou, em 1968, um terceiro parâmetro no modelo de Rash que ficou conhecido como Modelo Logístico de Três Parâmetros – ML3. Deve-se também a Birnbaum a descrição da Função de Informação do Item (utilizando um modelo de mensuração de Fisher) e também a primeira técnica utilizando Máxima Verossimilhança para estimar os parâmetros dos itens.

O primeiro programa computacional da TRI chamado BICAL foi desenvolvido por Wright e Panchapakesan em 1969, e facilitou muito as aplicações do modelo de Rasch.

A partir daí vários modelos foram desenvolvidos tais como: os nominais, de resposta gradual, multidimensionais, longitudinais, não-paramétricos, etc. Também, alguns métodos de estimação dos parâmetros foram aparecendo, inclusive o método bayesiano. Em decorrência de tudo isto, surgiram outros programas computacionais tais como: LOGIST de Lord, BILOG de Mislevy e Bock, PARSCALE de Muraki e Bock, TESTFACT de Wilson, Wood e Gibbons, MULTILOG de Thissen entre outros. Alguns nomes importantes que contribuíram para desenvolvimento desta teoria são: Samejima, Zimowski, Lieberman, Hambleton, Baker, etc. No Brasil, destacam-se os nomes: Dalton F. de Andrade, R. Klein, Héilton R. Tavares, Raquel C.Valle, entre outros.

### 2.2.2 Pressupostos Básicos

A Teoria da Resposta ao Item (TRI) é uma técnica originária da Teoria Clássica dos Testes (TCT) a qual se fundamenta apenas nos resultados obtidos pelos respondentes do conjunto de respostas aos itens que compõe o instrumento de avaliação (escores). A TRI, a princípio desenvolvida para respostas dicotômicas, usa a probabilidade de respostas corretas aos itens de um instrumento de avaliação para determinar os seus parâmetros (poder de discriminação, dificuldade e acerto casual) e as habilidades dos respondentes, de preferência na mesma quantidade latente contínua (ou espaço latente no

caso multidimensional). Isto significa que esta teoria permite medir os indivíduos na sua característica latente estimulada pela coleção de itens do instrumento de avaliação usado, por exemplo, a habilidade em alguma área, atitude em certa situação, necessidade ou desejo de alguma coisa, satisfação, qualidade, etc. Ao mesmo tempo esta teoria analisa cada item na mesma dimensão, como ilustração disto é se ter itens fáceis e difíceis no caso de uma avaliação educacional, desfavoráveis e favoráveis em avaliação de atitudes, etc. Em suma, dependendo do estudo em que se vai aplicar a TRI os traços latentes dos respondentes e os parâmetros dos itens têm significados próprios.

Há várias propostas de modelagem da teoria que dependem principalmente de três contextos: o tipo de resposta proposta aos itens, o número de populações (ou grupos) envolvidas e a quantidade de traços latentes medidos.

Quanto ao tipo de resposta, os itens podem ter respostas dicotômicas, dicotomizadas e não-dicotômicas. As dicotômicas se referem às respostas com duas possibilidades sendo uma de sucesso e a outra de fracasso (certo ou errado, satisfeito ou insatisfeito, etc.). Para as dicotomizadas atribui-se, por exemplo, a itens com respostas de múltipla escolha que podem ser analisados como certo ou errado ou ainda quando o instrumento permitir respostas abertas (livre) e que possibilitem a *posteriori* sua dicotomização. No caso de itens com respostas de caráter não-dicotômicas, estes são principalmente empregados em análises cujas respostas não possam ser dicotomizadas como, por exemplo, as de caráter ordinal. Na atualidade existem modelos na TRI específicos para analisar instrumentos para todos os tipos de variáveis sejam elas de caráter qualitativo (nominal e ordinal) ou quantitativo (discretas e contínuas).

No que se refere ao número de populações analisadas, os modelos dividem-se em duas categorias: os envolvendo uma população e os que consideram duas ou mais populações na análise. Todo o desenvolvimento da teoria foi definido a princípio para uma população, porém quando se trata de duas ou mais populações a comparação só poderá ser feita se existirem alguns itens comuns nos instrumentos. Os modelos que envolvem mais de uma população são basicamente uma extensão dos anteriores. Uma das alternativas para comparar várias populações é, por exemplo, o uso da Equalização definida por Kolen & Brennan em 1995, que significa colocar na mesma métrica os parâmetros de itens ou populações distintas. Pode-se usar, também, o Modelo para Várias

Populações, desenvolvido por Bock & Zimowski em 1997, que na realidade é uma equalização automática (ANDRADE, TAVARES & VALLE, 2000 p.93).

A classificação dos modelos com relação à quantidade de traços latentes ou habilidades analisadas, chamados de modelos unidimensionais quando estes consideram uma habilidade e multidimensionais quando se examinam duas ou mais habilidades num mesmo instrumento. Os modelos unidimensionais são os mais utilizados na prática e toda a teoria matemática da TRI foi fundamentada com este enfoque, porém pode-se encontrar, por exemplo, em LINDEN & HAMBLETON (1996, capítulo III) referências sobre os modelos multidimensionais.

Mais recentemente outros modelos foram introduzidos como os não-paramétricos, os longitudinais, entre outros. Os modelos longitudinais, por exemplo, analisam os parâmetros dos itens e das habilidades ao longo do tempo, o que possibilita acompanhar os indivíduos em duas ou mais avaliações com base na formulação de correlações e curvas de crescimento associadas a vetores de médias, toda formulação pode ser encontrada em TAVARES, 2001.

Neste trabalho, será apresentado o desenvolvimento do Modelo Logístico de Três Parâmetros Unidimensionais Dicotômicos e a estimação dos parâmetros dos itens utilizando o Método de Máxima Verossimilhança Marginal com algoritmo EM. Deve-se notificar que o desenvolvimento feito para estimar os parâmetros do modelo escolhido é semelhante para outros modelos.

### 2.2.3 Modelo Logístico de Três Parâmetros Unidimensional Dicotômico.

O Modelo Logístico de Três Parâmetros Unidimensional proposto por Birnbaum ao descrito por Rasch é dado por:

$$P(U_{ji} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (10)$$

onde

$P(U_{ji}=1|\theta_j)$  = probabilidade do respondente j acertar o item i dado que ele possui uma habilidade  $\theta_j$ .

$U_{ji}$  = resposta dada pelo respondente j ao item i.

$\theta_j$  = habilidade ou traço latente do respondente j.

$a_i$  = parâmetro de discriminação do item i.

$b_i$  = parâmetro de dificuldade do item i.

$c_i$  = parâmetro de acerto casual do item i.

$D$  = fator de escala geralmente igual a 1,7 para que a função logística forneça resultados semelhantes ao da distribuição normal.

O parâmetro de acerto casual ( $c_i$ ) refere-se à probabilidade de um respondente com habilidade baixa de acertar o item (ao acaso). Quando não se necessita desse parâmetro, na análise, ele pode ser retirado e têm-se o Modelo Logístico de Dois Parâmetros (ML-2), e retirando ainda o parâmetro de discriminação ( $a_i$ ) tem-se o Modelo de Um Parâmetro (ML-1) chamado de Modelo de Rasch, que considera que todos os itens discriminam da mesma forma.

#### 2.2.3.1 Curva característica do item – CCI

Segundo ANDRADE, TAVARES & VALLE (2000, p.10), o modelo logístico permite que se interprete graficamente os parâmetros dos itens e se pode verificar que indivíduos com maior habilidade possuem maior probabilidade de acertar o item e constata-se, ainda, que esta relação é não-linear.

Na Figura 1, chamada de Curva Característica do Item – CCI que é obtida da equação 1, pode-se observar que o parâmetro de discriminação do item  $i$  ( $a_i$ ) é proporcional a derivada da tangente da curva no ponto de inflexão e na prática sua variação é aceita entre zero e dois (HAMBLETON, SWAMINATHAN & ROGERS, 1991, p. 15). Assim, altos valores indicam que o item tem maior poder de separar indivíduos dentro dos diferentes níveis de habilidade. O parâmetro de dificuldade do item  $i$  ( $b_i$ ), medido na mesma escala da habilidade (geralmente delimitado no intervalo da normal padrão, como por exemplo de -4 à +4), representa a habilidade necessária para uma probabilidade de resposta correta correspondente a  $(1+c_i)/2$  e em consequência disso, quanto maior o valor de  $b_i$  mais difícil é o item. O parâmetro de acerto casual ( $c_i$ ) representa a probabilidade de resposta correta para um respondente com habilidade baixa, por conseguinte quando o instrumento não permitir um acerto casual dos itens este

parâmetro será igual a zero e neste caso  $b_i$  representa o valor da habilidade do respondente com chance de acerto igual a 50%.

A análise deve ser feita observando o gráfico e os parâmetros obtidos. Os critérios de identificação do modelo dependem do tipo de avaliação que se pretenda fazer e de considerações baseadas na experiência do avaliador, porém de maneira geral pode-se considerar que para o parâmetro de discriminação têm-se um item com discriminação alta quando este valor estiver próximo ou maior que 2, médio ao redor de 1 e baixo quando menor do que 0,5; pode-se observar esse valor na inclinação da curva, assim quanto maior for a inclinação maior o poder de discriminação do item.

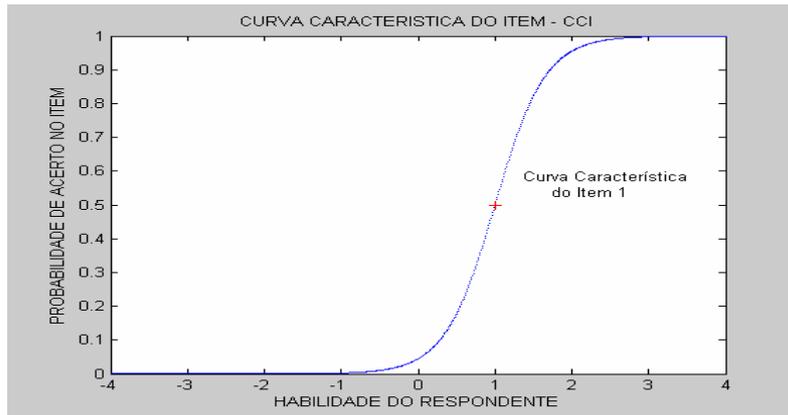
O parâmetro de dificuldade pode ser considerado alto quando assumir valores próximos ou maiores que a 4, médio em torno de zero e baixo quando for próximo ou abaixo de -4, assim no gráfico quanto mais a direita estiver a curva mais difícil é o item.

O parâmetro de acerto casual depende do número de alternativas usadas nos itens, sendo assim um item com pouca probabilidade de acerto casual deve ficar em torno de zero e com grande probabilidade de acerto casual deve ser igual ou maior que a razão (1/número de alternativas). No gráfico o parâmetro acerto casual é o valor correspondente à projeção da curva no eixo  $y$  quando ela tende a menos infinito.

Cabe aqui comentar que um item ruim é aquele que tem um parâmetro de discriminação baixo (abaixo de 0,5) e também um parâmetro de dificuldade muito alto (acima de 3,5) e muito baixo (abaixo de -3.5).

Alguns itens foram construídos para demonstrar o significado dos valores obtidos nos seus parâmetros procedendo-se as seguintes análises:

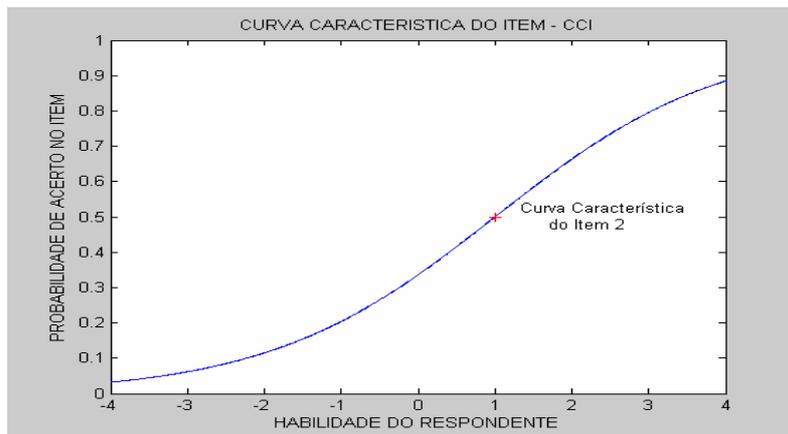
Na Figura 1 pode-se verificar que o item 1 tem discriminação alta de 1,8, dificuldade acima da média de 1,0 (ponto de inflexão denotado por uma cruz) e impossibilidade de acerto casual (zero).



FONTE: A AUTORA – SIAVAL 1.0

**FIGURA 1- GRÁFICO DA CURVA CARACTERÍSTICA DO ITEM 1**

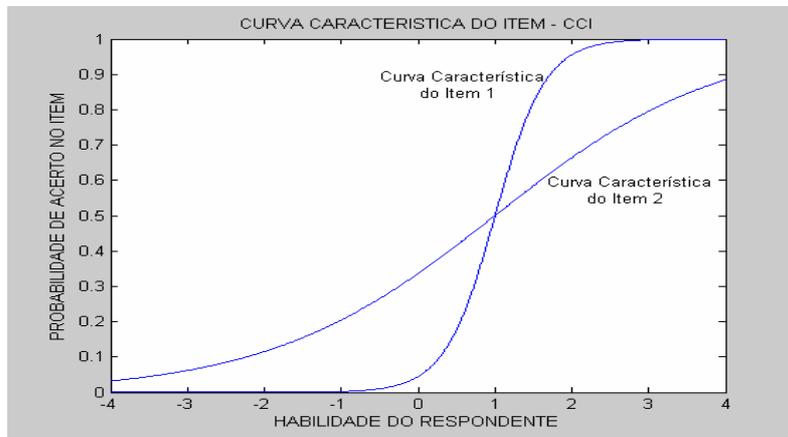
A título de comparação podemos ver que o item 2 no gráfico da Figura 2 possui o mesmo grau de dificuldade  $b = 1,0$ , nenhuma possibilidade de acerto casual  $c = 0$ , porém um índice de discriminação baixo  $a = 0,4$ . Veja a pequena inclinação da curva em contraste com a alta inclinação da curva da Figura 1.



FONTE: A AUTORA – SIAVAL 1.0

**FIGURA 2- GRÁFICO DA CURVA CARACTERÍSTICA DO ITEM 2**

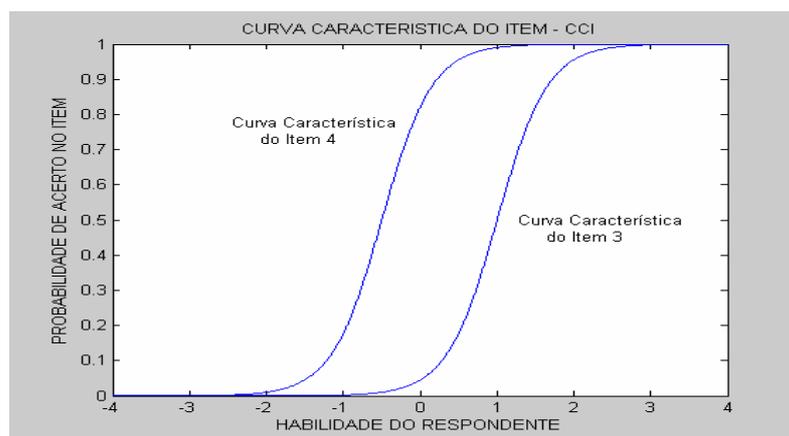
Plotando as curvas características dos dois itens juntos na Figura 3 pode-se perceber melhor que a inclinação da Curva Característica do Item 2 é menor que a do item 1. Isto significa que o item 1 é um item melhor para se aplicar do que o item 2, pois tem maior poder de discriminação.



FONTE: A AUTORA – SIAVAL

**FIGURA 3- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS 1 e 2**

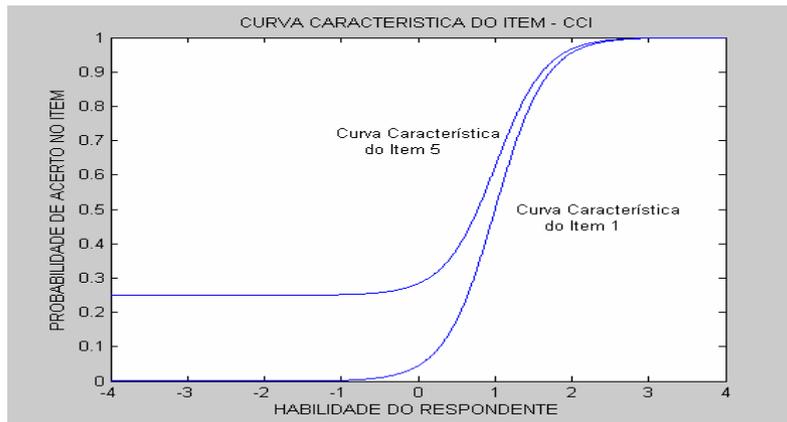
Ao se analisar outro caso com outros dois itens designados por 3 e 4, nos quais os parâmetros de discriminação sejam iguais a 1,8 e não exista a possibilidade de acerto casual ( $c = 0$ ) e com o parâmetro de dificuldade igual a 1,0 no item 3 e  $-0,5$  no item 4, observa-se na Figura 4 que a curva do item 3 está mais à direita. Assim, dado que os dois itens possuem discriminação alta, seria uma melhor escolha o item 3 para instrumentos onde é sabido que os respondentes possuem maior habilidade e o item 4 para os de menor habilidade.



FONTE: A AUTORA – SIAVAL 1.0

**FIGURA 4- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS 3 e 4**

Caso o parâmetro de acerto casual seja significativo, observa-se no Figura 5 que o item 1 já citado ( $a = 1,8$ ,  $b = 1$  e  $c = 0$ ) é comparado com um item 5 ( $a = 1,8$ ,  $b = 1$  e  $c = 0,25$ ) que possui um acerto casual de 25%.



FONTES: A AUTORA – SIAVAL 1.0

**FIGURA 5- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS 1 E 5**

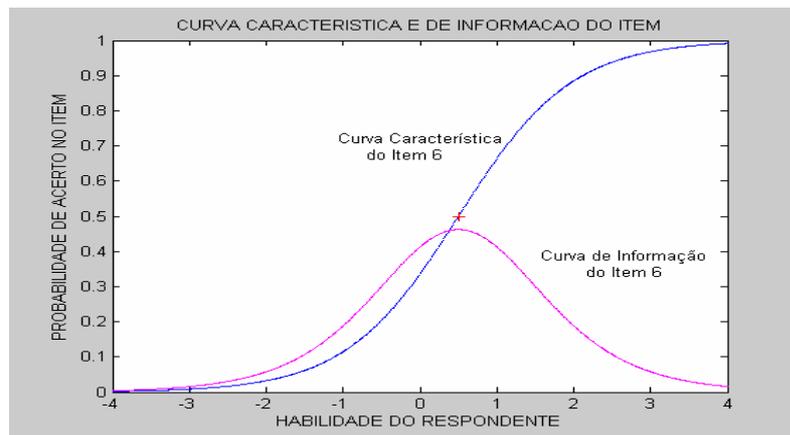
Percebe-se que o item 5 possui o parâmetro de acerto casual alto (0,25) que deforma a curva, porém continua praticamente com as mesmas características, interferindo pouco na análise.

Função de Informação do Item e do Instrumento:

Uma importante informação utilizada, conjuntamente com a CCI, é a Função de Informação do Item, denotada por  $I_i(\theta_j)$ , que no caso do modelo logístico de três parâmetros foi apresentada por Birnbaum em 1968 e citada em HAMBLETON, SWAMINATHAN & ROGERS (1991. p. 91), como:

$$I_i(\theta_j) = \frac{\left[ \frac{\partial P(U_{ji} = 1/\theta_j)}{\partial \theta_j} \right]^2}{P(U_{ji} = 1/\theta_j)[1 - P(U_{ji} = 1/\theta_j)]} = \frac{D^2 a_i^2 (1 - c_i)}{[c_i + e^{D a_i (\theta_j - b_i)}][c_i + e^{-D a_i (\theta_j - b_i)}]} \quad (11)$$

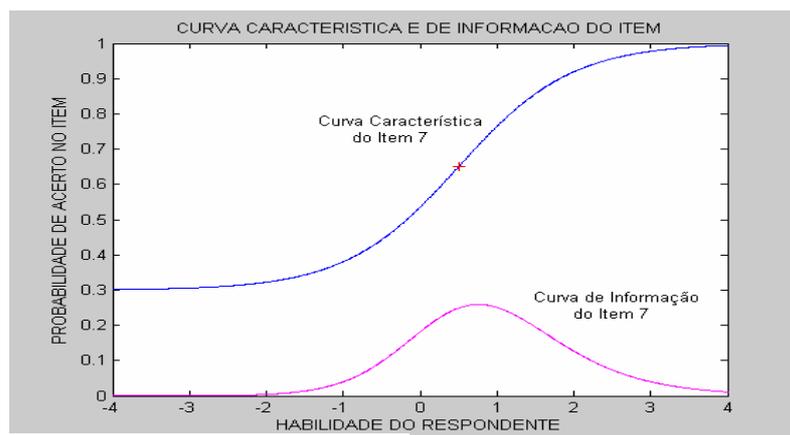
Examinando o Figura 6 a seguir, que contém a Curva Característica do item 6 ( $a = 0,8$ ,  $b = 0,5$  e  $c = 0$ ) e a sua Função de Informação, verifica-se que o valor máximo de informação do item se dá no ponto de inflexão da CCI, isto é , onde  $\theta = b$ .



FONTE: A AUTORA – SIAVAL 1.0

**FIGURA 6- GRÁFICO DA CURVA CARACTERÍSTICA E DA FUNÇÃO DE INFORMAÇÃO DO ITEM 6**

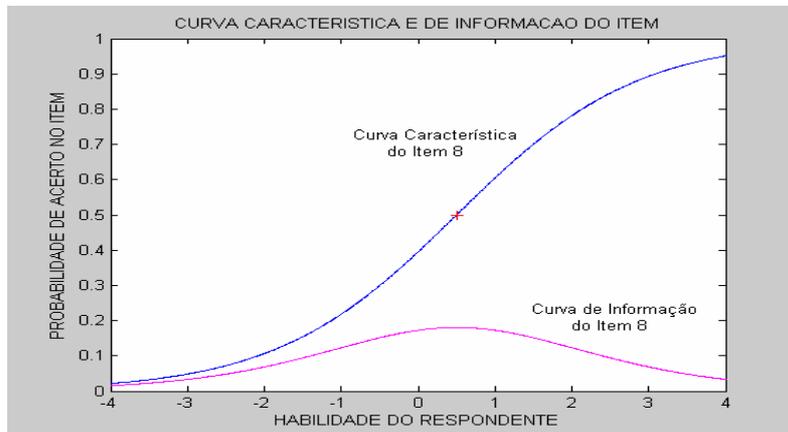
Quando o parâmetro de acerto casual é grande tem-se que o item é menos útil para determinar a habilidade, o que fica evidente no Figura 7 cujas curvas estudadas se referem ao item 7 com parâmetros:  $a = 0,8$ ,  $b = 0,5$  e  $c = 0,3$ .



FONTE: AUTORA –SIAVAL 1.0

**FIGURA 7- GRÁFICO DA CURVA CARACTERÍSTICA E DA FUNÇÃO DE INFORMAÇÃO DO ITEM 7**

Se o parâmetro de discriminação é muito pequeno tem-se pouca informação do item, como se pode observar no Figura 8 cujo item 8 analisado tem os parâmetros  $a = 0,5$ ,  $b = 0,5$  e  $c = 0$ .



FORTE: A AUTORA – SIAVAL 1.0

**FIGURA 8- GRÁFICO DA CURVA CARACTERÍSTICA e DA FUNÇÃO DE INFORMAÇÃO DO ITEM 8**

Para se determinar a informação que contém todo instrumento deve-se somar todas as informações geradas pelas funções dos itens em alguns valores de  $\theta$  de interesse, como por exemplo, no primeiro, segundo e terceiro quartis (ou outra escolha mais detalhada). Esta função é denotada por  $SI(\theta_j)$  e dada por:

$$SI(\theta_j) = \sum_{i=1}^I I_i(\theta_j) \quad (12)$$

Segundo HAMBLETON, SWAMINATHAM & ROGERS (1991, p.94), a contribuição de um item individualmente no teste pode ser determinada sem conhecimento de outro item do teste, isto é, os itens são considerados na TRI como independentes. Isto que difere da confiabilidade feita pela Teoria Clássica dos Testes (TCT), que usa a correlação bisserial de ponto. A vantagem é obtida quando se deseja trocar um dos itens, pois os outros não são afetados por esta troca. Assim, se pode direcionar a retirada deste item adequando-o para a habilidade dos indivíduos que se pretende atingir com teste, ou seja, quando um item possui parâmetro de dificuldade alto ele não discrimina indivíduos que é sabido ter habilidade baixa e vice-versa.

A quantidade de informação do teste na habilidade  $\theta_j$  é oriundo do inverso da precisão com que a habilidade é estimada no ponto  $\theta_j$  e chamada de Erro Padrão de Estimação, dada por:

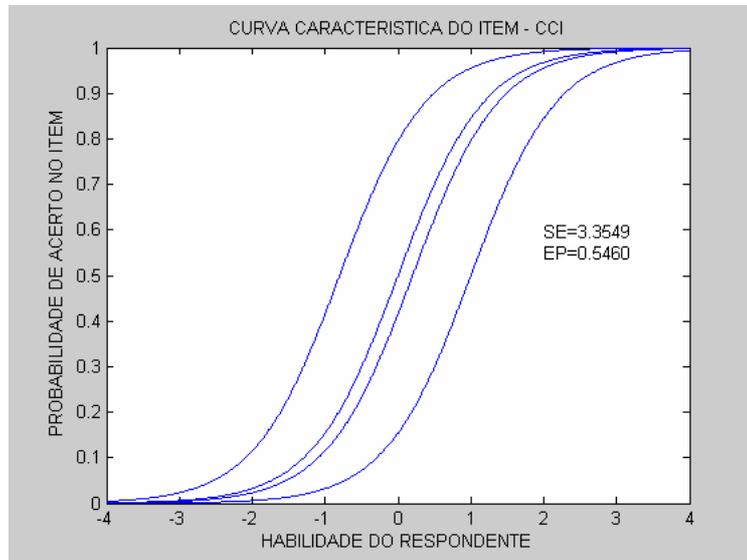
$$EP(\theta_j) = \frac{1}{\sqrt{SI(\theta_j)}} \quad (13)$$

O Erro Padrão de  $\theta_j$ ,  $EP(\theta_j)$ , é o desvio padrão da distribuição assintoticamente normal da estimativa de máxima verossimilhança da habilidade, onde:

$$S_{\theta_j} = \sqrt{\frac{1}{-E\left(\frac{\partial^2 \log L(\theta_j)}{\partial \theta_j^2}\right)}} = \sqrt{\frac{1}{\sum_{i=1}^I \left[ \frac{D^2 a_i^2 (1 - c_i)}{[c_i + e^{Da_i(\theta_j - b_i)}][1 + e^{-Da_i(\theta_j - b_i)}]^2} \right]}} = \frac{1}{\sqrt{\sum_{i=1}^I I_i(\theta_j)}} = \frac{1}{\sqrt{SI(\theta_j)}} \quad (14)$$

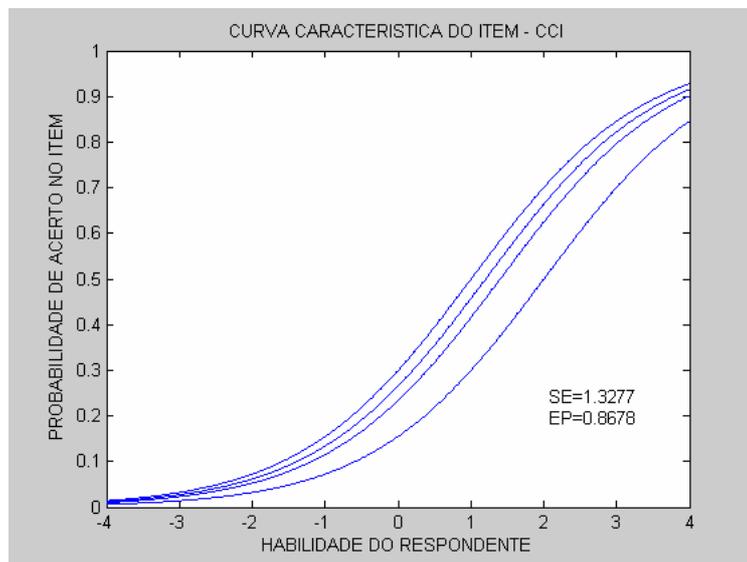
Estas funções caracterizam as informações de um instrumento que possibilitam a comparação com outros. Assim, na existência de um banco de itens, pode-se através de análises deste tipo escolher um instrumento que atenda melhor a necessidade do pesquisador.

Para melhor exemplificar, as Figuras 9 e 10 demonstram dois agrupamentos de itens diferentes (ou de dois instrumentos diferentes). No primeiro, instrumento A, a soma de todas as informações no três quartis resultou um valor de 3,3549 com um erro padrão de 0,5460. No segundo, instrumento B, a soma forneceu um valor de 1,3277 com erro padrão de 0,8678. Conseqüentemente, o instrumento A seria o escolhido, pois ele fornece melhores resultados com uma maior quantidade de informação concentrada.



FONTE: A AUTORA – SIAVAL 1.0

**FIGURA 9- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS DO INSTRUMENTO A**



**FIGURA 10- GRÁFICO DAS CURVAS CARACTERÍSTICAS DOS ITENS DO INSTRUMENTO B**

Na construção de um banco de itens se faz necessário comparar a informação de dois testes que tenham o mesmo nível de habilidade. Usa-se, então a relação denotada por

Eficiência Relativa ( $ER(\theta_j)$ ) apresentada a seguir, onde  $SI_A(\theta_j)$  representa a função de informação do teste A e  $SI_B(\theta_j)$  a do teste B.

$$ER(\theta_j) = \frac{SI_A(\theta_j)}{SI_B(\theta_j)} = \frac{3,3549}{1,3277} = 2,5268 \quad (15)$$

As análises apresentadas possibilitam selecionar itens de testes que apresentaram melhor desempenho numa habilidade determinada. Suponha, por exemplo, que se deseja aplicar um teste onde se sabe que os respondentes possuem habilidade alta. Assim, se outros testes já tenham sido feitos com outras populações, verifica-se através da eficiência relativa o melhor conjunto de itens que possibilite discriminação desta habilidade. Uma importante aplicação desta relação se dá quando se quer construir um banco de itens, pode-se, portanto, selecionar os melhores itens para cada nível de habilidade, e a construção dos testes fica facilitada.

Propriedades da TRI:

Uma propriedade importante da TRI é a invariância dos parâmetros dos itens e habilidades, que segundo HAMBLETON, SWAMINATHAN & ROGGERS (1991, p.18), é o que distingue a TRI da Teoria Clássica dos Testes. Esta propriedade sustenta que os parâmetros dos itens e a distribuição das habilidades dos respondentes são independentes. Isto significa que a CCI é a mesma para duas populações diferentes, uma vez calibrados os itens. Esta propriedade é a base da TRI e possibilita a obtenção de um banco de dados, investigação do viés do item e a adaptação do teste ao que se propõe. Deve-se ressaltar que a propriedade de invariância é válida somente quando um modelo da TRI for ajustado satisfatoriamente aos dados de interesse (calibrados).

Modelos que pressupõe um instrumento que meça somente um traço latente (habilidade) possuem uma propriedade chamada de unidimensionalidade. Na realidade isto não acontece, pois os respondentes utilizam mais de uma habilidade para responder os itens propostos. Assim, para satisfazer o postulado, segundo ANDRADE, TAVARES & VALLE (2.000, p.16), basta aceitar que nos itens selecionados, para determinado instrumento, exista uma habilidade que predomina sobre as demais, que é exatamente o fator que está sendo pesquisado. Verifica-se esta independência utilizando correlações

tetracóricas, apresentada nesta tese nos procedimentos de Análise Fatorial. Modelos que assumem mais de uma habilidade são referidos como multidimensionais e possuem estruturas mais complexas.

Quando num modelo a unidimensionalidade é adotada, assume-se outra propriedade chamada de independência local. Esta propriedade significa que fixada uma habilidade às respostas dadas aos diferentes itens do instrumento são independentes. Desta forma, está se supondo que o conjunto dos níveis de habilidade representa o espaço latente completo. A independência local é uma suposição fundamental à estimação dos parâmetros dado que a probabilidade do indivíduo com a habilidade  $\theta_j$  acertar um item ( $U_i$ ) não afeta a probabilidade dele acertar outro item ( $U_i$ ), conseqüentemente a probabilidade conjunta de um indivíduo com habilidade  $\theta_j$  acerta os n itens um instrumento é dada por:

$$P(U_1, U_2, \dots, U_n / \theta_j) = P(U_1 / \theta_j) P(U_2 / \theta_j) \dots P(U_n / \theta_j) = \prod_{i=1}^n P(U_i / \theta_j) \quad (16)$$

A presente formulação permite a estimação dos parâmetros por métodos estatísticos de bom desempenho, como o da Máxima Verossimilhança.

Uma observação importante feita por HAMBLETON, SWAMINATHAN & ROGERS (1991, p.11-12) é a de que a propriedade de independência local faz mais sentido quando todas as habilidades estão sendo medidas, por se ter realmente o espaço latente completo, logo esta propriedade pode ser aceita também para estudos onde não se considera a unidimensionalidade.

#### 2.2.3.4 Estimação dos parâmetros dos itens e das habilidades dos indivíduos

##### 2.2.3.4.1 Introdução

Os métodos de estimação dos parâmetros dos itens e das habilidades dos respondentes mais citados na literatura são:

-Método da Máxima Verossimilhança para os parâmetros dos itens e habilidades separadamente, isto é, primeiro se estima os parâmetros dos itens com a resolução de um sistema de 3 equações não-lineares e no passo seguinte as habilidades resolvendo n (número de indivíduos) equações não-lineares, onde são aplicados os métodos iterativos

de Newton-Raphson ou “Scoring” de Fisher. Este método, segundo BAKER & KIM (2004, p.53) tem mostrado para modelos abaixo de dois parâmetros que é bastante robusto e sempre converge, mas no modelo de três parâmetros não é muito robusto e não converge facilmente dado que a solução depende dos valores iniciais.

-Método de Máxima Verossimilhança Conjunta para os parâmetros dos itens e habilidades, segundo HAMBLETON, SWAMINATHAN & ROGERS (1991, p.46-47) foi muito utilizado na década de 70 nos primeiros programas da TRI (BICAL e LOGIST). Na análise de ANDRADE, TAVARES & VALLE (2.000, p.77) apresenta vários problemas de utilização (indeterminação, propriedades assintóticas, computacionalmente trabalhoso, etc.), porém serviu como base para outros procedimentos.

-Método da Máxima Verossimilhança Marginal, desenvolvido por Bock & Lieberman em 1970 para um modelo normal com dois parâmetros segundo BOCK & AITKIN (1981 p.443-459), determina uma distribuição para as habilidades e trabalha com a probabilidade marginal, considerando os itens independentes entre si. A partir desta suposição estima os itens através de um algoritmo chamado EM (Esperança e Maximização) apresentado por Dempster, Laird & Rubin em 1977, citado por BOCK & AITKIN (1981 p.444). Este método é o mais empregado, pois possui propriedades assintóticas para os parâmetros de discriminação ( $a_i$ ) e dificuldade ( $b_i$ ) e ainda os parâmetros de acerto casual ( $c_i$ ) são consistentes. No entanto, apresenta problemas de estimação dos parâmetros dos itens quando existirem itens com acerto ou erro total e também na estimação das habilidades quando respondentes acertam ou erram todos os itens.

-Método Bayesiano Conjunto (proposto por Swaminathan and Gifford, em vários artigos apresentados nos anos de 1982, 1985 e 1986, citados por SWAMINATHAN, HAMBLETON, SIRECI, XING & RIZAVI (2003, p. 27-51) assume que os parâmetros de discriminação ( $a_i$ ) tem distribuição qui-quadrado, os de dificuldade ( $b_i$ ) são independentes e identicamente distribuídos com distribuição normal e os de acerto casual ( $c_i$ ) tem distribuição beta. Na estimação utiliza-se máxima verossimilhança conjunta.

-Método Bayesiano Marginal considera que o parâmetro de discriminação ( $a_i$ ) tem distribuição lognormal, de dificuldade ( $b_i$ ) são independentes e identicamente distribuídos

com distribuição normal e o de acerto casual ( $c_i$ ) tem distribuição beta, como no método anterior. A estimação é feita usando com o algoritmo EM.

#### 2.2.3.4.2 Estimação dos parâmetros dos itens

Escolheu-se o método de máxima verossimilhança marginal com algoritmo EM no desenvolvimento deste trabalho dado que vários autores, alguns já citados, o indicam e na maioria dos artigos pesquisados, este é um dos métodos mais utilizados e apresenta bons resultados. Portanto, a seguir apresenta-se o seu desenvolvimento teórico.

Método da Máxima Verossimilhança Marginal:

O método de máxima verossimilhança é um método estatístico para estimar um vetor de parâmetros populacionais ( $\underline{\theta}$ ) de alguma distribuição de probabilidade  $f(x)$  a partir da maximização da função de verossimilhança. A função de verossimilhança de  $n$  variáveis aleatórias  $X_1, X_2, X_3, \dots, X_n$  é definida em MOOD, GRAYBILL & BOES (1974, p. 276-286), como sendo a densidade conjunta das  $n$  variáveis aleatórias denotada por  $L(\underline{\theta}; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i | \underline{\theta})$ , onde no caso  $\underline{\theta} \in \Theta$  (espaço paramétrico) e representa os parâmetros da distribuição. O estimador de máxima verossimilhança de  $\underline{\theta}$  é o vetor  $\hat{\underline{\theta}} \in \Theta$  que maximiza a função de verossimilhança  $L(\underline{\theta}, \underline{x})$ .

No caso da Teoria da Resposta ao Item, a probabilidade do respondente  $j$  acertar o item  $i$  dado que ele possui a habilidade  $\theta_j$  é modelada por uma função logística dada por:

$$P(U_{ji} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (17)$$

onde  $i = 1, \dots, I$  e  $j = 1, \dots, n$  sendo  $I$  o número de itens e  $n$  o número de indivíduos avaliados.

Assim, a Função de Verossimilhança é dada por:

$$L(\underline{\zeta}_i, \theta_j) = P(U_{..} = 1 | \theta_j, \underline{\zeta}_i) = \prod_{j=1}^n \prod_{i=1}^I P(U_{ji} = 1 | \theta_j, \zeta_i) \quad (18)$$

onde  $\underline{\zeta}_i = [a_i \ b_i \ c_i]'$  é o vetor dos parâmetros dos itens. A obtenção deste resultado foi possível devido à suposição de Independência Local.

A estimação por Máxima Verossimilhança de  $L(\underline{\zeta}_i, \theta_j)$  implica em derivar o logaritmo da equação em relação ao vetor de parâmetros dos itens  $\frac{\partial \log L(\underline{\zeta}_i, \theta_j)}{\partial \underline{\zeta}_i}$  e igualar a zero para obter, no caso, um sistema constituído de três equações. A estimativa dos parâmetros dos itens  $(a_i, b_i, c_i)$  seria imediata caso as habilidades  $\theta_j$  fossem conhecidas, e poder-se-á utilizar o método iterativo de Newton-Raphson para resolução de equações não-lineares. Porém, na maioria dos casos, as habilidades não são conhecidas, o que implica na utilização do Método de Máxima Verossimilhança Conjunta (parâmetros dos itens e habilidades), que segundo ANDRADE, TAVARES & VALLE (2.000, p.77) possui problemas de indeterminação.

Segundo BACKER & KIM (2004, p.158-174), em 1970 Bock e Lieberman apresentaram uma solução onde se assume que as habilidades dos respondentes possuem uma função distribuição empírica  $G(\theta_j)$  e se  $n$ , o número de indivíduos, for suficientemente grande pode-se supor que esta distribuição seja contínua e sua função densidade de probabilidade seja  $g(\theta_j/\underline{\tau})$ . Assim, pelo Teorema Central do Limite tem-se que  $\theta$  pode ter uma distribuição aproximadamente normal onde  $\underline{\tau}$  representa o vetor de parâmetros dado por  $\underline{\tau}' = [\mu \ \sigma^2]$ .

A estimação dos parâmetros dos itens e habilidades na mesma métrica é feita com base na distribuição normal padrão  $N(0,1)$  em ambos os casos. Para suprimir as habilidades da função adota-se a marginal da distribuição conjunta de itens e habilidades e, ainda, para garantir a propriedade de suficiência dos parâmetros estimados tem-se que:

$$P(U_j. / \underline{\zeta}_i, \underline{\tau}) = \int_{\mathfrak{R}} P(U_j. / \theta_j, \underline{\zeta}_i, \underline{\tau}) g(\theta_j, \underline{\tau}) d\theta_j \quad (19)$$

como  $P(U_j. / \theta_j, \underline{\zeta}_i, \underline{\tau})$  não depende de  $\underline{\tau} = (\mu, \sigma^2)$  e  $g(\theta_j / \underline{\tau}) \sim N(0,1)$  temos que:

$$P(U_j. / \underline{\zeta}_i, \underline{\tau}) = \int_{\mathfrak{R}} P(U_j. / \theta_j, \underline{\zeta}_i) g(\theta_j / \underline{\tau}) d\theta_j \quad (20)$$

Usando a propriedade de independência local da TRI pode-se escrever a probabilidade associada ao vetor de respostas  $U..$  como:

$$P(U_{..} / \underline{\zeta}_i, \underline{\tau}) = \prod_{j=1}^n P(U_{j.} / \underline{\zeta}_i, \underline{\tau}) = \prod_{j=1}^n \int_{\mathfrak{R}} P(U_{j.} / \theta_j, \underline{\zeta}_i) g(\theta_j / \underline{\tau}) d\theta_j \quad (21)$$

Para se obter vantagens computacionais ANDRADE, TAVARES & VALLE (2.000, p.53) recomendam que se utilize padrões de respostas ao invés de indivíduos. Isto significa dizer que  $j$  não representa mais um indivíduo e sim um padrão de resposta. Por exemplo, se 5 indivíduos tivessem respondido a 3 itens, como nas tabelas a seguir, tem-se  $j$  variando de 1 a 3.

**TABELA 1- EXEMPLO DE ITENS POR INDIVÍDUOS**

INDIVÍDUOS	ITENS		
	1	2	3
1	1	0	1
2	0	1	0
3	0	0	1
4	1	0	1
5	0	1	0

FONTE: A AUTORA

**TABELA 2- EXEMPLO DE ITENS POR PADRÕES DE RESPOSTA**

PADRÕES DE RESPOSTA	ITENS		
	1	2	3
1	1	0	1
2	0	1	0
3	0	0	1

FONTE: A AUTORA

Denota-se por  $r_j$  o número de indivíduos com padrões de respostas iguais. No exemplo, 2 indivíduos obtiveram padrão de resposta  $r_1$ , então  $r_1$  seria igual a 2 e considerando os outros padrões ter-se-ia  $r_2 = 2$  e  $r_3 = 1$ . Chamando de  $S$  o número máximo de padrões e  $s \leq \min(n, S)$  para todo  $r_j > 0$  tem-se que  $\sum_{j=1}^s r_j = n$  (número de indivíduos), ter-se-ia  $S = 3$  e  $n=5$ .

Desta forma, dada a existência de independência entre as respostas, a distribuição  $P(U_{..}/\underline{\zeta}_i, \underline{\tau})$  segue uma Distribuição Multinomial e a equação de verossimilhança é dada por:

$$L(\underline{\zeta}_i, \underline{\tau}) = \frac{n!}{\prod_{j=1}^s r_j!} \prod_{j=1}^s [P(u_{j.}/\underline{\zeta}_i, \underline{\tau})]^{r_j} \quad \text{onde } j = 1, \dots, s \quad e \quad i = 1, \dots, I \quad (22)$$

Aplicando-se o Método de Estimação de Máxima Verossimilhança tem-se que:

$$\ln L(\underline{\zeta}_i, \underline{\tau}) = \ln \left( \frac{n!}{\prod_{j=1}^s r_j!} \right) + \sum_{j=1}^s r_j \cdot \ln [P(u_{j.}/\underline{\zeta}_i, \underline{\tau})] \quad (23)$$

e

$$\frac{\partial [\ln L(\underline{\zeta}_i, \underline{\tau})]}{\partial \underline{\zeta}_i} = 0 \quad i=1, \dots, I \quad (24)$$

A resolução destas equações (ver ANDRADE, TAVARES & VALLE, 2000 p. 54-57), chega às seguintes equações de estimação dos parâmetros  $a_i$ ,  $b_i$  e  $c_i$ :

$$\frac{\partial \ln L(\underline{\zeta}_i, \underline{\tau})}{\partial \underline{\zeta}_i} = \sum_{j=1}^s \bar{r}_j \int_{\Re} \left[ (u_{ji} - P_i) \left( \frac{\partial P_i}{\partial \zeta_i} \right) \frac{W_i}{P_i^* Q_i^*} \right] g_j^*(\theta) d\theta \quad (25)$$

derivando a equação em relação a  $a_i$ ,  $b_i$  e  $c_i$  obtêm-se:

$$\begin{cases} D(1-c_i) \sum_{j=1}^s r_j \int_{\Re} [(u_{ji} - P_i)(\theta - b_i) W_i] g_j^*(\theta) d\theta = 0 & \text{para } a_i \\ -Da_i(1-c_i) \sum_{j=1}^s r_j \int_{\Re} [(u_{ji} - P_i) W_i] g_j^*(\theta) d\theta = 0 & \text{para } b_i \\ \sum_{j=1}^s r_j \int_{\Re} [(u_{ji} - P_i) \frac{W_i}{P_i^*}] g_j^*(\theta) d\theta = 0 & \text{para } c_i \end{cases} \quad (26)$$

onde:

$$P_i = c_i + (1-c_i) \cdot [1 + e^{-Da_i(\theta-b_i)}]^{-1} \quad (27)$$

$$P_i^* = [1 + e^{-Da_i(\theta-b_i)}]^{-1} \quad (28)$$

$$Q_i^* = 1 - P_i^* \quad (29)$$

$$W_i = \frac{P_i^*(1 - P_i^*)}{P_i(1 - P_i)} \quad (30)$$

$$g_j^*(\theta) = g(\theta | u_j, \zeta, \tau) = \frac{P(u_j | \theta, \zeta) g(\theta | \tau)}{P(u_j | \zeta, \tau)} \quad (31)$$

Segundo ANDRADE, TAVARES & VALLE (2000, p.59); estas equações envolvem integrais que não apresentam solução analítica e sugerem como solução uma aproximação com o Método da Quadratura Gaussiana, também chamado de Método de Hermite-Gauss. Este método, segundo BURDEN & FAIRES (1997, p.222-227), usa valores da função com igualdade nos espaços dos pontos e permite calcular por aproximação a integral de uma função contínua pela integração de funções lineares (soma de áreas de trapézios) obtidos na escolha de pontos (nós de quadratura) igualmente espaçados. Os nós, por exemplo,  $x_1, x_2, \dots, x_n$  num intervalo  $[a, b]$  e coeficientes  $c_1, c_2, \dots, c_n$  são escolhidos de tal maneira que apresentem um mínimo de erros obtidos na aproximação de função arbitrária  $f$ , dado por:

$$\int_a^b f(x) dx = \sum_{i=1}^n c_i f(x_i) \quad (32)$$

No caso da TRI, supõe-se que a função  $g(\theta / \tau)$  seja normalmente distribuída com média zero e variância igual a 1 (Normal Padrão), pode-se usar a quadratura gaussiana, pois sendo ela uma função contínua com momentos finitos permite uma aproximação com razoável grau de acurácia por uma distribuição discreta com um número finito de pontos (isto é, por um histograma). Então, segundo BAKER & KIM (2004, p.164), os nós de quadratura são dados por  $\bar{\theta}_k$ , com  $k=1, 2, \dots, q$  e considera os pesos iguais a  $A_k = g(\bar{\theta}_k / \tau) \Delta_k$  onde  $\Delta_k = \bar{\theta}_2 - \bar{\theta}_1$  e é a largura do intervalo entre dois pontos de quadratura. Os pontos de quadratura  $\bar{\theta}_k$  e os pesos  $A_k$ , de acordo com BOCK & AITKIN (1981, p.449), são obtidos resolvendo uma coleção de equações que se encontra numa tabela para distribuição gaussiana no livro de Stroud, A. H. & Secrest, D. intitulado *Gaussian Quadrature Formulas da editora Prentice-Hall* de 1966. Bock & Lieberman indicam que estes valores podem ser aproximados para uma distribuição gaussiana padrão se os nós forem multiplicados por  $\sqrt{2}$  e os respectivos pesos divididos por  $\sqrt{\pi}$ . Reescrevendo as equações em termos de quadraturas fica:

$$P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) = \prod_{i=1}^I [P_{ki}^{u_{ji}} \cdot Q_{ki}^{u_{ji}}] \quad (33)$$

fazendo

$$\begin{aligned} P_{ki} &= c_i + (1 - c_i) \cdot [1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^{-1} \\ Q_{ki} &= 1 - P_{ki} \end{aligned} \quad (34)$$

e como  $A_k = g(\bar{\theta}_k / \tau) \cdot \Delta_k$  então se pode escrever:

$$\begin{aligned} P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) \cdot g(\bar{\theta}_k / \tau) &= P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) \cdot A_k \cdot \Delta_k^{-1} \\ P(u_{j.} / \underline{\zeta}_i, \tau) &\cong \sum_{k=1}^q P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) \cdot A_k \end{aligned} \quad (35)$$

e

$$\begin{aligned} g_j^*(\bar{\theta}_k) &= P(\bar{\theta}_k / u_{j.}, \underline{\zeta}_i, \tau) = \frac{P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) g(\bar{\theta}_k / \tau)}{P(u_{j.} / \underline{\zeta}_i, \tau)} \approx \\ &\approx \frac{P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) \cdot A_k}{\sum_{k=1}^q P(u_{j.} / \bar{\theta}_k, \underline{\zeta}_i) A_k} \Delta_k^{-1} \end{aligned} \quad (36)$$

Na execução do método da máxima verossimilhança usa-se a notação:

$$\begin{aligned} P_{ki} &= c_i + (1 - c_i) \cdot [1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^{-1}, \\ Q_{ki} &= 1 - P_{ki} = 1 - \left\{ c_i + (1 - c_i) \cdot [1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^{-1} \right\} \\ P_{ki}^* &= [1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^{-1}, \\ Q_{ki}^* &= 1 - P_{ki}^* = 1 - [1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^{-1}, \\ W_{ki} &= \frac{P_{ki}^* Q_{ki}^*}{P_{ki} Q_{ki}} \end{aligned} \quad (37)$$

$$\frac{\partial \log L(\zeta_i, \tau)}{\partial \zeta_i} = \sum_{j=1}^s \sum_{k=1}^q \bar{r}_j \left[ (u_{ji} - P_{ki}) \left( \frac{\partial P_{ki}}{\partial \zeta_i} \right) \frac{W_{ki}}{P_{ki}^* Q_{ki}^*} \right] g_j^*(\bar{\theta}_k) = 0 \quad (38)$$

derivando em relação a  $a_i$ ,  $b_i$  e  $c_i$  obtêm-se:

$$\left\{ \begin{array}{l} D(1-c_i) \sum_{j=1}^s \sum_{k=1}^q r_j [(u_{ji} - P_{ki})(\bar{\theta}_k - b_i) W_{ki}] g_j^*(\bar{\theta}_k) = 0 \text{ para } a_i \\ -Da_i(1-c_i) \sum_{j=1}^s \sum_{k=1}^q r_j [(u_{ji} - P_{ki}) W_{ki}] g_j^*(\bar{\theta}_k) = 0 \text{ para } b_i \\ \sum_{j=1}^s \sum_{k=1}^q r_j \left[ (u_{ji} - P_{ki}) \frac{W_{ki}}{P_{ki}^*} \right] g_j^*(\bar{\theta}_k) = 0 \text{ para } c_i \end{array} \right. \quad (39)$$

Mais uma vez estas equações não apresentam soluções explícitas na estimação dos parâmetros dos itens, o que implica estimá-los simultaneamente gerando um grande esforço computacional (inversão de uma matriz de  $3I \times 3I$ , onde  $I$  representa o número de itens).

A reformulação de BOCK & AITKIN (1981, p. 443-459) possibilitou melhor solução para o problema assumindo que os itens são independentes entre si, os indivíduos são independentes e que os itens e indivíduos são independentes.

Esta suposição facilita os cálculos, dado que a inversão do Jacobiano da matriz do método de resoluções de equações não-lineares de Newton-Raphson, se torna bloco diagonal dado que para dois itens diferentes tem-se que:

$$\frac{\partial^2 \log L(\underline{\zeta}_i, \underline{\tau})}{\partial \zeta_i \partial \zeta_l} = 0 \quad \text{para } i \neq l \quad (40)$$

As equações em forma de quadratura são dadas a seguir derivando o logaritmo da função de verossimilhança e igualando a zero.

$$\frac{\partial \log L(\underline{\zeta}_i, \underline{\tau})}{\partial \zeta_i} = \sum_{k=1}^q \left( \sum_{j=1}^s \bar{r}_j u_{ji} g_j^*(\bar{\theta}_k) - \sum_{j=1}^s \bar{r}_j P_{ki} g_j^*(\bar{\theta}_k) \right) \left( \frac{\partial P_{ki}}{\partial \zeta_i} \right) \frac{W_{ki}}{P_{ki}^* Q_{ki}^*} = 0 \quad (41)$$

fazendo,

$$r_{ki} = \sum_{j=1}^s r_j u_{ji} g_j^*(\bar{\theta}_k) \quad (42)$$

$$f_{ki} = \sum_{j=1}^s r_j g_j^*(\bar{\theta}_k) \quad (43)$$

então:

$$\begin{cases} D(1-c_i) \sum_{k=1}^q (\bar{\theta}_k - b_i) [r_{ki} - P_{ki} f_{ki}] W_{ki} = 0 & \text{para } a_i \\ -Da_i(1-c_i) \sum_{k=1}^q [r_{ki} - P_{ki} f_{ki}] W_{ki} = 0 & \text{para } b_i \\ \sum_{k=1}^q [r_{ki} - P_{ki} f_{ki}] \frac{W_{ki}}{P_{ki}^*} = 0 & \text{para } c_i \end{cases} \quad (44)$$

As equações  $r_{ki}$  e  $f_{ki}$  são conhecidas na literatura da TRI como “dados artificiais”, e BOCK & AITKIN (1981, p.444) sugerem o emprego do algoritmo EM (Esperança e Maximização) de Dempster, Laird & Rubin para estimar os parâmetros dos itens com as devidas modificações implementadas por eles.

Algoritmo EM:

O algoritmo EM, segundo BAKER & KIM (2004, p.169), é um procedimento iterativo para encontrar os estimadores de máxima verossimilhança dos modelos probabilísticos dos parâmetros da TRI em presença de variáveis aleatórias não-observáveis  $\bar{\theta}_k$  (parâmetros de habilidade) com probabilidades  $\pi_k$  onde  $k$  varia de 1 a  $q$  seria o número de pontos de quadratura adotado. O passo E representa o passo de cálculo da Esperança e o M o passo de Maximização.

Na TRI somente o modelo de um parâmetro –ML-1 pertence à família exponencial, os modelos ML-2 e ML-3 não pertencem à família exponencial e não é possível avaliar se o estimador é suficiente. Como um substituto, o valor esperado do logaritmo da função logística condicional as variáveis não observáveis  $\bar{\theta}_k$ , são tratadas como se elas fossem conhecidas e este é o passo chamado de E.

Concebendo que as habilidades são restritas para um conjunto de valores finitos,  $q$  e estes valores ( $\bar{\theta}_k$ ) têm respectivas probabilidades  $\pi_k$ , e ainda, denotando como  $r_{ki}$  o números de acertos do item  $i$  com quadratura  $k$  e  $f_{ki}$  o número tentativas por indivíduos cada uma das  $k$  habilidades do item  $i$ . Se a habilidade ( $\bar{\theta}_k$ ) de  $n$  indivíduos são consideradas como amostras aleatórias para a distribuição anteriormente mencionada, a probabilidade conjunta é dada por uma distribuição multinomial a seguir.

$$P(f_i / \pi_k) = \frac{n!}{\prod_{k=1}^q f_{ki}!} \prod_{k=1}^q \pi_k^{f_{ki}}, \quad i = 1, 2, \dots, I \quad (45)$$

Fazendo  $\underline{f}'_i = (f_{1i}, f_{2i}, \dots, f_{qi})$  com  $\underline{f}' = (f_1, f_2, \dots, f_I)$ ,  $\underline{r}'_i = (r_{1i}, r_{2i}, \dots, r_{qi})$  com  $\underline{r}' = (r_1, r_2, \dots, r_I)$  e  $\bar{\theta}' = (\theta_1, \theta_2, \dots, \theta_q)$  a probabilidade de ocorrerem  $r_{ki}$  acertos no item  $i$  dentre as  $f_{ki}$  tentativas por respondentes com habilidade  $\bar{\theta}_k$  é dada por:

$$P(r_{ki} / f_{ki}, \bar{\theta}_k) = \binom{f_{ki}}{r_{ki}} P_{ki}^{r_{ki}} Q_{ki}^{f_{ki}-r_{ki}} \quad (46)$$

Então, a probabilidade conjunta de  $\underline{f}'$  e  $\underline{r}'$  dado  $\bar{\theta}$  e  $\pi_k$  é escrita por:

$$\begin{aligned} P(\underline{f}', \underline{r}' / \bar{\theta}', \pi_k) &= P(\underline{f}' / \bar{\theta}', \pi_k) \cdot P(\underline{r}' / \underline{f}', \bar{\theta}', \pi_k) = \\ &= P(\underline{f}' / \pi_k) \cdot P(\underline{r}' / \underline{f}', \bar{\theta}') = \\ &= \left[ \prod_{i=1}^I P(\underline{f}'_i / \pi_k) \right] \cdot \left[ \prod_{i=1}^I \prod_{k=1}^q P(r_{ki} / f_{ki}, \bar{\theta}_k) \right] \end{aligned} \quad (47)$$

Logo, o logaritmo da função de verossimilhança é dado por:

$$\begin{aligned} \ln L(\underline{\zeta}_i) &= \ln P(\underline{f}' / \pi_k) + \sum_{i=1}^I \sum_{k=1}^q \ln P(r_{ki} / f_{ki}, \bar{\theta}_k) \\ &= \ln P(\underline{f}' / \pi_k) + \sum_{i=1}^I \sum_{k=1}^q \left[ \ln \binom{f_{ki}}{r_{ki}} + r_{ki} \log P_{ki} + (f_{ki} - r_{ki}) \ln(1 - P_{ki}) \right] \end{aligned} \quad (48)$$

Fazendo  $C = \ln P(\underline{f}' / \pi_k) + \sum_{i=1}^I \sum_{k=1}^q \ln \binom{f_{ki}}{r_{ki}}$  que é constante em relação a  $\zeta$  têm-se:

$$\ln L(\underline{\zeta}_i) = C + \sum_{i=1}^I \sum_{k=1}^q \{r_{ki} \ln P_{ki} + (f_{ki} - r_{ki}) \ln(1 - P_{ki})\} \quad (49)$$

e como  $\underline{f}'$  e  $\underline{r}'$  não são observáveis e fazendo,

$$\bar{r}_{ki} = E(r_{ki} / u_{..}, \underline{\zeta}_i) \quad (50)$$

$$\bar{f}_{ki} = E(f_{ki} / u_{..}, \underline{\zeta}_i) \quad (51)$$

$$\bar{C} = E(C / u_{..}, \underline{\zeta}_i) \quad (52)$$

$$E[\ln L(\underline{\zeta}_i)] = \bar{c} + \sum_{i=1}^I \sum_{k=1}^q \{ \bar{r}_{ki} \ln P_{ki} + (\bar{f}_{ki} - \bar{r}_{ki}) \ln(1 - P_{ki}) \} \quad (53)$$

A derivada do logaritmo da equação é dada por:

$$\begin{aligned} \frac{\partial \ln L(\underline{\zeta}_i)}{\partial \underline{\zeta}_i} &= \sum_{k=1}^q \left\{ \bar{r}_{ki} \frac{1}{P_{ki}} \left( \frac{\partial P_{ki}}{\partial \underline{\zeta}_i} \right) + (\bar{f}_{ki} - \bar{r}_{ki}) \frac{1}{Q_{ki}} \left[ \frac{\partial(1 - P_{ki})}{\partial \underline{\zeta}_i} \right] \right\} = \\ &= \sum_{k=1}^q \left( \frac{\partial P_{ki}}{\partial \underline{\zeta}_i} \right) \left[ \bar{r}_{ki} \frac{1}{P_{ki}} - (\bar{f}_{ki} - \bar{r}_{ki}) \frac{1}{Q_{ki}} \right] = \\ &= \sum_{k=1}^q \left( \frac{\partial P_{ki}}{\partial \underline{\zeta}_i} \right) \left[ \frac{\bar{r}_{ki} Q_{ki} - (\bar{f}_{ki} - \bar{r}_{ki}) P_{ki}}{P_{ki} Q_{ki}} \right] = \\ &= \sum_{k=1}^q \left( \frac{\partial P_{ki}}{\partial \underline{\zeta}_i} \right) \left[ \frac{\bar{r}_{ki} (1 - P_{ki}) - (\bar{f}_{ki} P_{ki} - \bar{r}_{ki} P_{ki})}{P_{ki} Q_{ki}} \right] = \\ &= \sum_{k=1}^q \left( \frac{\bar{r}_{ki} - \bar{f}_{ki} P_{ki}}{P_{ki} Q_{ki}} \right) \left( \frac{\partial P_{ki}}{\partial \underline{\zeta}_i} \right) \end{aligned} \quad (54)$$

Usando os resultados demonstrados no Apêndice 1:

$$\frac{\partial P_{ki}}{\partial a_i} = D(1 - c_i) (\bar{\theta}_k - b_i) P_{ki}^* Q_{ki}^* \quad (55)$$

$$\frac{\partial P_{ki}}{\partial b_i} = -D a_i (1 - c_i) P_{ki}^* Q_{ki}^* \quad (56)$$

$$\frac{\partial P_{ki}}{\partial c_i} = 1 - P_{ki}^* = Q_{ki}^* \quad (57)$$

e derivando a equação em relação aos parâmetros dos itens obtêm-se:

$$\begin{aligned} \frac{\partial L(\underline{\zeta}_i)}{\partial a_i} &= \sum_{k=1}^q \left( \frac{\bar{r}_{ki} - \bar{f}_{ki} P_{ki}}{P_{ki} Q_{ki}} \right) \left( \frac{\partial P_{ki}}{\partial a_i} \right) = \sum_{k=1}^q \left( \frac{\bar{r}_{ki} - \bar{f}_{ki} P_{ki}}{P_{ki} Q_{ki}} \right) [D(1 - c_i) (\bar{\theta}_k - b_i) P_{ki}^* Q_{ki}^*] = \\ &= D(1 - c_i) \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) (\bar{\theta}_k - b_i) W_{ki} \end{aligned} \quad (58)$$

$$\begin{aligned}\frac{\partial L(\underline{\zeta}_i)}{\partial b_i} &= \sum_{k=1}^q \frac{(\bar{r}_{ki} - \bar{f}_{ki} P_{ki})}{P_{ki} Q_{ki}} \left( \frac{\partial P_{ki}}{\partial b_i} \right) = \sum_{k=1}^q \frac{(\bar{r}_{ki} - \bar{f}_{ki} P_{ki})}{P_{ki} Q_{ki}} [-Da_i(1-c_i)P_{ki}^* Q_{ki}^*] = \\ &= -Da_i(1-c_i) \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) W_{ki}\end{aligned}\quad (59)$$

$$\frac{\partial L(\underline{\zeta}_i)}{\partial c_i} = \sum_{k=1}^q \frac{(\bar{r}_{ki} - \bar{f}_{ki} P_{ki})}{P_{ki} Q_{ki}} \left( \frac{\partial P_{ki}}{\partial c_i} \right) = \sum_{k=1}^q \frac{(\bar{r}_{ki} - \bar{f}_{ki} P_{ki})}{P_{ki} Q_{ki}} Q_{ki}^* = \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) \frac{W_{ki}}{P_{ki}^*} \quad (60)$$

$$F = \begin{cases} D(1-c_i) \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) (\bar{\theta}_k - b_i) W_{ki} = 0 & \text{para } a_i \\ -Da_i(1-c_i) \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) W_{ki} = 0 & \text{para } b_i \\ \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) \frac{W_{ki}}{P_{ki}^*} = 0 & \text{para } c_i \end{cases} \quad (61)$$

Finalmente resolvendo este sistema de equações não lineares (F) através do Método de Newton-Raphson com intuito de obter as estimativas dos parâmetros dos itens, o passo M terá sido feito.

O Método de Newton-Raphson é um método iterativo e segundo BURDEN & FAIRES (1997, p.601), geralmente é esperado que dê convergência quadrática, contanto que seja fornecido um valor inicial suficientemente próximo do alvo e que o Jacobiano da matriz exista. Assim, resolve-se o sistema:

$$F = \begin{cases} F_1 \\ F_2 \\ F_3 \end{cases} \quad (62)$$

por

$$\begin{bmatrix} \hat{a}_i^{(t)} \\ \hat{b}_i^{(t)} \\ \hat{c}_i^{(t)} \end{bmatrix} = \begin{bmatrix} \hat{a}_i^{(t-1)} \\ \hat{b}_i^{(t-1)} \\ \hat{c}_i^{(t-1)} \end{bmatrix} + \begin{bmatrix} x_i^{(t-1)} \\ y_i^{(t-1)} \\ z_i^{(t-1)} \end{bmatrix} \quad (63)$$

onde

$\begin{bmatrix} \hat{a}_i^{(t-1)} \\ \hat{b}_i^{(t-1)} \\ \hat{c}_i^{(t-1)} \end{bmatrix}$  é a matriz dos valores dos parâmetros dos itens na iteração  $t-1$

e

$$\begin{bmatrix} x_i^{(t-1)} \\ y_i^{(t-1)} \\ z_i^{(t-1)} \end{bmatrix} = -\left[ J\left(\hat{a}_i^{(t-1)}, \hat{b}_i^{(t-1)}, \hat{c}_i^{(t-1)}\right) \right]^{-1} F\left(\hat{a}_i^{(t-1)}, \hat{b}_i^{(t-1)}, \hat{c}_i^{(t-1)}\right) \quad (64)$$

sendo  $J\left(\hat{a}_i^{(t-1)}, \hat{b}_i^{(t-1)}, \hat{c}_i^{(t-1)}\right)$  o Jacobiano da matriz dada por:

$$J\left(\hat{a}_i^{(t-1)}, \hat{b}_i^{(t-1)}, \hat{c}_i^{(t-1)}\right) = \begin{bmatrix} \frac{\partial F_1}{\partial a_i} & \frac{\partial F_1}{\partial b_i} & \frac{\partial F_1}{\partial c_i} \\ \frac{\partial F_2}{\partial a_i} & \frac{\partial F_2}{\partial b_i} & \frac{\partial F_2}{\partial c_i} \\ \frac{\partial F_3}{\partial a_i} & \frac{\partial F_3}{\partial b_i} & \frac{\partial F_3}{\partial c_i} \end{bmatrix} \quad (65)$$

sendo que,

$$\begin{aligned} \frac{\partial F_1}{\partial a_i} &= D(1-c_i) \sum_{i=1}^k (\bar{\theta}_k - b_i) \frac{\partial \left[ (\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki}) W_{ki} \right]}{\partial a_i} = \\ &= D(1-c_i) \sum_{i=1}^k (\bar{\theta}_k - b_i) \left[ (\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki}) \frac{\partial W_{ki}}{\partial a_i} - \bar{f}_{ki} W_{ki} \frac{\partial P_{ki}}{\partial a_i} \right] \end{aligned} \quad (66)$$

$$\begin{aligned} \frac{\partial F_2}{\partial b_i} &= -Da_i(1-c_i) \sum_{i=1}^k \frac{\partial \left[ (\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki}) W_{ki} \right]}{\partial b_i} = \\ &= -Da_i(1-c_i) \sum_{i=1}^k \left[ (\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki}) \frac{\partial W_{ki}}{\partial b_i} - W_{ki} \bar{f}_{ki} \frac{\partial P_{ki}}{\partial b_i} \right] \end{aligned} \quad (67)$$

$$\begin{aligned}
\frac{\partial F_3}{\partial c_i} &= \sum_{i=1}^k \frac{\partial \left[ \frac{(\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki}) W_{ki}}{P_{ki}^*} \right]}{\partial c_i} = \\
&= \sum_{i=1}^k \left\{ (\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki}) \left[ \frac{P_{ki}^* \frac{\partial W_{ki}}{\partial b_i} - W_{ki} \frac{\partial P_{ki}^*}{\partial c_i}}{P_{ki}^*} \right] - \frac{W_{ki}}{P_{ki}^*} \bar{f}_{ki} \frac{\partial P_{ki}^*}{\partial b_i} \right\} = \\
&= \sum_{i=1}^k \left\{ \frac{(\bar{r}_{ki} - \bar{f}_{ki} \cdot P_{ki})}{P_{ki}^*} \left[ \frac{\partial W_{ki}}{\partial b_i} - W_{ki} \bar{f}_{ki} \frac{\partial P_{ki}^*}{\partial b_i} \right] \right\}
\end{aligned} \tag{68}$$

pelos resultados obtidos nos Apêndices 1 e 2:

$$\frac{\partial P_{ki}}{\partial a_i} = D(1 - c_i)(\bar{\theta}_k - b_i)P_{ki}^*Q_{ki}^* \tag{69}$$

$$\frac{\partial P_{ki}^*}{\partial a_i} = D(\bar{\theta}_k - b_i)P_{ki}^*Q_{ki}^* = \frac{1}{(1 - c_i)} \frac{\partial P_{ki}}{\partial a_i} \tag{70}$$

$$\frac{\partial W_{ki}}{\partial a_i} = \left[ W_{ki}(1 - 2P_{ki}) + \frac{(1 - 2P_{ki}^*)}{(1 - c_i)} \right] \frac{1}{P_{ki}Q_{ki}} \frac{\partial P_{ki}}{\partial a_i} \tag{71}$$

$$\frac{\partial P_{ki}}{\partial b_i} = -Da_i(1 - c_i)P_{ki}^*Q_{ki}^* \tag{72}$$

$$\frac{\partial P_{ki}^*}{\partial b_i} = -Da_iP_{ki}^*Q_{ki}^* = -\frac{1}{(1 - c_i)} \frac{\partial P_{ki}}{\partial b_i} \tag{73}$$

$$\frac{\partial W_{ki}}{\partial b_i} = \left[ W_{ki}(1 - 2P_{ki}) + \frac{(1 - 2P_{ki}^*)}{(1 - c_i)} \right] \frac{1}{P_{ki}Q_{ki}} \frac{\partial P_{ki}}{\partial b_i} \tag{74}$$

$$\frac{\partial P_{ki}}{\partial c_i} = 1 - P_{ki}^* = Q_{ki}^* \tag{75}$$

$$\frac{\partial P_{ki}^*}{\partial c_i} = 0 \tag{76}$$

$$\begin{aligned}
\frac{\partial W_{ki}}{\partial c_i} &= \left[ W_{ki}(1 - 2P_{ki}) + \frac{(1 - 2P_{ki}^*)}{(1 - c_i)} \right] \frac{1}{P_{ki}Q_{ki}} \frac{\partial P_{ki}}{\partial c_i} = \\
&= \left[ W_{ki}(1 - 2P_{ki}) + \frac{(1 - 2P_{ki}^*)}{(1 - c_i)} \right] \frac{1}{P_{ki}}
\end{aligned} \tag{77}$$

Assim fornecendo-se um vetor inicial obtidos para cada ponto de quadratura  $k$

$\begin{bmatrix} \hat{a}_i^{(0)} \\ \hat{b}_i^{(0)} \\ \hat{c}_i^{(0)} \end{bmatrix}$  consegue-se o vetor estimado para cada item, pela resolução da equação

$$\begin{bmatrix} \hat{a}_i^{(t)} \\ \hat{b}_i^{(t)} \\ \hat{c}_i^{(t)} \end{bmatrix} = \begin{bmatrix} \hat{a}_i^{(t-1)} \\ \hat{b}_i^{(t-1)} \\ \hat{c}_i^{(t-1)} \end{bmatrix} + \begin{bmatrix} x_i^{(t-1)} \\ y_i^{(t-1)} \\ z_i^{(t-1)} \end{bmatrix} \text{ onde } t \text{ é a iteração e volta-se ao passo E até que um critério de}$$

parada seja estabelecido.

No caso deste trabalho o critério mais indicado, segundo BURDEN & FAIRES (1997, p.600) é dado por:

$$\left\| \begin{bmatrix} \hat{a}_i^{(t)} \\ \hat{b}_i^{(t)} \\ \hat{c}_i^{(t)} \end{bmatrix} - \begin{bmatrix} \hat{a}_i^{(t-1)} \\ \hat{b}_i^{(t-1)} \\ \hat{c}_i^{(t-1)} \end{bmatrix} \right\|_{\infty} = \max \left\| \begin{bmatrix} \hat{a}_i^{(t)} \\ \hat{b}_i^{(t)} \\ \hat{c}_i^{(t)} \end{bmatrix} - \begin{bmatrix} \hat{a}_i^{(t-1)} \\ \hat{b}_i^{(t-1)} \\ \hat{c}_i^{(t-1)} \end{bmatrix} \right\| < \underline{\varepsilon} \quad \text{onde } \underline{\varepsilon} = \begin{bmatrix} 10^{-2} \\ 10^{-2} \\ 10^{-2} \end{bmatrix} \quad (78)$$

Resumidamente o algoritmo EM é feito nos seguintes passos:

**E:** Usam-se os pontos de quadratura  $\bar{\theta}_k$  com os respectivos pesos  $A_k$  para fazer estimativas iniciais dos parâmetros dos itens  $\hat{\zeta}_i$  com o intuito de determinar os valores  $g_j^*(\bar{\theta}_k)$  e para gerar os valores  $\bar{r}_{ki}$  e  $\bar{f}_{ki}$ .

**M:** Resolve-se o sistema de equações usando o algoritmo de Newton-Raphson.

Os passos são repetidos até que o critério de parada seja alcançado.

Os procedimentos usados na estimação do Modelo Logístico de Dois Parâmetros é semelhante ao apresentado para o Modelo Logístico de Três Parâmetros. A seguir se fará uma breve descrição dos procedimentos.

O Modelo Logístico de Dois Parâmetros, não possui o parâmetro de acerto casual ( $c_i$ ) e é dado por:

$$P(U_{ji} = 1/\theta_j) = \frac{1}{1 + e^{a_i(\theta_j - b_i)}} \quad (79)$$

onde

$P(U_{ji}=1/\theta_j)$  = probabilidade do respondente  $j$  acertar o item  $i$  dado que ele possui uma habilidade  $\theta_j$ .

$U_{ji}$  = resposta dada pelo respondente  $j$  ao item  $i$ .

$\theta_j$  = habilidade ou traço latente do respondente j.

$a_i$  = parâmetro de discriminação do item i.

$b_i$  = parâmetro de dificuldade do item i.

Segundo, BAKER & KIM (2004, p. 333) por questões de conveniência é feita a seguinte parametrização:

$$Z_{ij} = d_i + a_i \theta_j \quad (80)$$

O implica que  $d_i$  é o intercepto e  $a_i$  o grau de inclinação. O parâmetro de dificuldade será obtido por:

$$b_i = -\frac{d_i}{a_i} \quad (81)$$

Portanto, os parâmetros estimados serão  $a_i$  e  $d_i$  e no final obtêm-se  $b_i$  através da equação 81.

O sistema de equações a ser resolvido é dado por:

$$F = \begin{cases} \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) = 0 & \text{para } a_i \\ \sum_{k=1}^q (\bar{r}_{ki} - \bar{f}_{ki} P_{ki}) X_k = 0 & \text{para } d_i \end{cases} \quad (82)$$

Sendo:

$$P_{ki} = \frac{1}{1 + e^{-d_i + a_i X_k}} \quad (83)$$

$$Q_{ki} = 1 - P_{ki} \quad (84)$$

Derivando as equações em relação a  $a_i$  e  $d_i$  obtêm-se:

$$\frac{\partial F_1}{\partial a_i} = -\sum_{k=1}^q \bar{f}_{ki} P_{ki} Q_{ki} X_k^2 \quad (85)$$

$$\frac{\partial F_1}{\partial d_i} = \frac{\partial F_2}{\partial a_i} = -\sum_{k=1}^q \bar{f}_{ki} P_{ki} Q_{ki} X_k \quad (86)$$

$$\frac{\partial F_1}{\partial d_i} = -\sum_{k=1}^q \bar{f}_{ki} P_{ki} Q_{ki} \quad (87)$$

Aplicando o algoritmo EM ,como demonstrado para três parâmetros, obtêm-se as estimativas de  $a_i$  e  $d_i$  , retoma-se a equação 80 e encontra-se o parâmetro de dificuldade  $b_i$  .

Na seção seguinte apresentam-se alguns métodos de estimação das habilidade ( $\theta_j$ ) dos indivíduos.

#### 2.2.3.4.3 Estimação das habilidades dos indivíduos.

A estimação das habilidades é feita considerando-se os parâmetros dos itens já calibrados e segundo TAVARES (2001, p. 44) os principais modelos de estimação das habilidades são o por máxima verossimilhança e os bayesianos. Apresentadas nas próximas seções.

##### 2.2.3.4.3.1 Estimação por máxima verossimilhança.

Assumindo a independência entre as respostas de diferentes indivíduos e a independência local, segundo ANDRADE, TAVARES & VALLE (2000, p. 44), a estimação por máxima verossimilhança é aplicada na equação que segue.

$$\ln L(\theta) = \sum_{j=1}^n \sum_{i=1}^I [u_{ji} \ln P_{ji} + (1 - u_{ji}) \ln Q_{ji}] \quad (88)$$

Onde  $j$  volta a ser o número de indivíduos. Derivando a equação anterior obtêm-se (resolução apresentada no Apêndice 3):

$$\frac{\partial \ln L(\theta)}{\partial \theta_j} = D \sum_{i=1}^I a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji} = 0 \quad (89)$$

E, utilizando o método de Newton-Raphson novamente, dado que a equação não apresenta solução explícita.

Fazendo,

$$F(\theta_j) = D \sum_{i=1}^I a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji} \quad (90)$$

$$\begin{bmatrix} y_1^{(t-1)} \\ y_2^{(t-1)} \\ \vdots \\ y_n^{(t-1)} \end{bmatrix} = -[J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)})]^{-1} [F(\hat{\theta}_1^{(t-1)}) \quad F(\hat{\theta}_2^{(t-1)}) \quad \dots \quad F(\hat{\theta}_n^{(t-1)})] \quad (91)$$

sendo  $J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)})$  o Jacobiano da matriz dada por:

$$J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)}) = \det \begin{bmatrix} \frac{\partial F(\theta_1)}{\partial \theta_1} & \frac{\partial F(\theta_2)}{\partial \theta_1} & \dots & \frac{\partial F(\theta_n)}{\partial \theta_1} \\ \frac{\partial F(\theta_1)}{\partial \theta_2} & \frac{\partial F(\theta_2)}{\partial \theta_2} & \dots & \frac{\partial F(\theta_n)}{\partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F(\theta_1)}{\partial \theta_n} & \frac{\partial F(\theta_2)}{\partial \theta_n} & \dots & \frac{\partial F(\theta_n)}{\partial \theta_n} \end{bmatrix} \quad (92)$$

Considerando a habilidade de um indivíduo não depende da habilidade de outro o Jacobiano da matriz se torna bloco diagonal e fica:

$$J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)}) = \det \begin{bmatrix} \frac{\partial F(\theta_1)}{\partial \theta_1} & 0 & \dots & 0 \\ 0 & \frac{\partial F(\theta_2)}{\partial \theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial F(\theta_n)}{\partial \theta_n} \end{bmatrix} \quad (93)$$

onde

$$\frac{\partial F(\theta_j)}{\partial \theta_j} = D * a_i^2 \left[ \frac{P_i(\theta_j) - c_i}{1 - c_i} \right]^2 \frac{Q_i(\theta_j)}{P_i(\theta_j)} - \frac{1}{\sigma_\theta^2} \quad (94)$$

Assim fornecendo um vetor inicial  $\begin{bmatrix} \hat{\theta}_1^{(0)} \\ \hat{\theta}_2^{(0)} \\ \vdots \\ \hat{\theta}_n^{(0)} \end{bmatrix}$  obtêm-se um vetor estimado para as

habilidades dos indivíduos pela resolução de  $\begin{bmatrix} \hat{\theta}_1^{(t)} \\ \hat{\theta}_2^{(t)} \\ \vdots \\ \hat{\theta}_n^{(t)} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_1^{(t-1)} \\ \hat{\theta}_2^{(t-1)} \\ \vdots \\ \hat{\theta}_n^{(t-1)} \end{bmatrix} + \begin{bmatrix} y_1^{(t-1)} \\ y_2^{(t-1)} \\ \vdots \\ y_n^{(t-1)} \end{bmatrix}$  até que um critério

de parada seja estabelecido, que no caso pode ser dado por:

$$\left\| \begin{bmatrix} \hat{\theta}_1^{(t)} \\ \hat{\theta}_2^{(t)} \\ \vdots \\ \hat{\theta}_n^{(t)} \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1^{(t-1)} \\ \hat{\theta}_2^{(t-1)} \\ \vdots \\ \hat{\theta}_n^{(t-1)} \end{bmatrix} \right\|_\infty = \max \left\| \begin{bmatrix} \hat{\theta}_1^{(t)} \\ \hat{\theta}_2^{(t)} \\ \vdots \\ \hat{\theta}_n^{(t)} \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1^{(t-1)} \\ \hat{\theta}_2^{(t-1)} \\ \vdots \\ \hat{\theta}_n^{(t-1)} \end{bmatrix} \right\| < \underline{\varepsilon} = \begin{bmatrix} 10^{-3} \\ 10^{-3} \\ \vdots \\ 10^{-3} \end{bmatrix}. \quad (95)$$

### 2.2.3.4.3.2 Estimação modal bayesiana ou *maximum a posteriori* (MAP).

De acordo com BAKER & KIM (2004, p. 192) a habilidade dos respondentes é obtida com o máximo da distribuição normal  $g_j^*(\theta)$  com parâmetros populacionais conhecidos dada por:

$$g_j^*(\theta_j) = \frac{P(U_{j..} | \theta_j, \zeta) g(\theta_j | \tau)}{P(U_{j..} | \zeta, \tau)} \quad (96)$$

Assim, segundo TAVARES (2001, p. 45) a equação de estimação é dada por:

$$\frac{\partial \ln g_j^*(\theta_j)}{\partial \theta_j} = \frac{\partial \ln P(U_{j..} | \theta_j, \zeta)}{\partial \theta_j} + \frac{\partial \ln g(\theta_j | \tau)}{\partial \theta_j} = 0 \quad (97)$$

Logo, pela independência local tem-se que:

$$\ln P(U_{j..} | \theta_j, \zeta) = \ln \left[ \prod_{i=1}^I P(U_{ji} | \zeta_i, \theta_j) \right] = \sum_{i=1}^I \ln P(U_{ji} | \zeta_i, \theta_j) \quad (98)$$

Derivando a equação acima em relação às habilidades  $\theta_j$  e usando conceitos e expressões citadas anteriormente na estimação dos itens tem-se uma expressão já apresentada na estimação por máxima verossimilhança dada por:

$$\frac{\partial \ln P(U_{j..} | \theta_j, \zeta)}{\partial \theta_j} = D \sum_{i=1}^I a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji} \quad (99)$$

e ainda adotando a distribuição normal a priori para  $\theta_j$ , segundo ANDRADE, TAVARES & VALLE (2000, p. 74), a segunda parcela é:

$$\frac{\partial \ln g(\theta_j | \eta)}{\partial \theta_j} = - \frac{(\theta_j - \mu_\theta)}{\sigma_\theta^2} \quad (100)$$

e a equação de estimação de  $\theta_j$  é a apresentada a seguir.

$$\theta_j : D \sum_{i=1}^I a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji} - \frac{(\theta_j - \mu_\theta)}{\sigma_\theta^2} \quad (101)$$

Utilizando, mais uma vez a resolução de Newton-Raphson do sistema de equações não-lineares tem-se que a segunda derivada da equação em relação  $\theta_j$  dada por:

Fazendo,

$$F(\theta_j) = D \sum_{i=1}^I a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji} - \frac{(\theta_j - \mu_\theta)}{\sigma_\theta^2} \quad (102)$$

$$\begin{bmatrix} y_1^{(t-1)} \\ y_2^{(t-1)} \\ \vdots \\ y_n^{(t-1)} \end{bmatrix} = -[J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)})]^{-1} [F(\hat{\theta}_1^{(t-1)}) \quad F(\hat{\theta}_2^{(t-1)}) \quad \dots \quad F(\hat{\theta}_n^{(t-1)})] \quad (103)$$

sendo  $J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)})$  o Jacobiano da matriz dada por:

$$J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)}) = \det \begin{bmatrix} \frac{\partial F(\theta_1)}{\partial \theta_1} & \frac{\partial F(\theta_2)}{\partial \theta_1} & \dots & \frac{\partial F(\theta_n)}{\partial \theta_1} \\ \frac{\partial F(\theta_1)}{\partial \theta_2} & \frac{\partial F(\theta_2)}{\partial \theta_2} & \dots & \frac{\partial F(\theta_n)}{\partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F(\theta_1)}{\partial \theta_n} & \frac{\partial F(\theta_2)}{\partial \theta_n} & \dots & \frac{\partial F(\theta_n)}{\partial \theta_n} \end{bmatrix} \quad (104)$$

Considerando a habilidade de um indivíduo não depende da habilidade de outro o Jacobiano da matriz se torna bloco diagonal:

$$J(\hat{\theta}_1^{(t-1)}, \hat{\theta}_2^{(t-1)}, \dots, \hat{\theta}_n^{(t-1)}) = \det \begin{bmatrix} \frac{\partial F(\theta_1)}{\partial \theta_1} & 0 & \dots & 0 \\ 0 & \frac{\partial F(\theta_2)}{\partial \theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial F(\theta_n)}{\partial \theta_n} \end{bmatrix} \quad (105)$$

onde

$$\frac{\partial F(\theta_j)}{\partial \theta_j} = D * a_i^2 \left[ \frac{P_i(\theta_j) - c_i}{1 - c_i} \right]^2 \frac{Q_i(\theta_j)}{P_i(\theta_j)} - \frac{1}{\sigma_\theta^2} \quad (106)$$

Assim fornecendo um vetor inicial  $\begin{bmatrix} \hat{\theta}_1^{(0)} \\ \hat{\theta}_2^{(0)} \\ \vdots \\ \hat{\theta}_n^{(0)} \end{bmatrix}$  obtêm-se um vetor estimado para as habilidades dos indivíduos pela resolução de  $\begin{bmatrix} \hat{\theta}_1^{(t)} \\ \hat{\theta}_2^{(t)} \\ \vdots \\ \hat{\theta}_n^{(t)} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_1^{(t-1)} \\ \hat{\theta}_2^{(t-1)} \\ \vdots \\ \hat{\theta}_n^{(t-1)} \end{bmatrix} + \begin{bmatrix} y_1^{(t-1)} \\ y_2^{(t-1)} \\ \vdots \\ y_n^{(t-1)} \end{bmatrix}$  até que um critério de parada seja estabelecido. Que no caso pode ser dado por:

$$\left\| \begin{bmatrix} \hat{\theta}_1^{(t)} \\ \hat{\theta}_2^{(t)} \\ \vdots \\ \hat{\theta}_n^{(t)} \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1^{(t-1)} \\ \hat{\theta}_2^{(t-1)} \\ \vdots \\ \hat{\theta}_n^{(t-1)} \end{bmatrix} \right\|_{\infty} = \max \left\| \begin{bmatrix} \hat{\theta}_1^{(t)} \\ \hat{\theta}_2^{(t)} \\ \vdots \\ \hat{\theta}_n^{(t)} \end{bmatrix} - \begin{bmatrix} \hat{\theta}_1^{(t-1)} \\ \hat{\theta}_2^{(t-1)} \\ \vdots \\ \hat{\theta}_n^{(t-1)} \end{bmatrix} \right\| < \underline{\varepsilon} = \begin{bmatrix} 10^{-3} \\ 10^{-3} \\ \vdots \\ 10^{-3} \end{bmatrix}. \quad (107)$$

#### 2.2.3.4.3.3 Estimação bayesiana da média *a posteriori* ou *expected a posteriori* (EAP).

A estimação de  $\theta_j$  neste caso é feita diretamente, sem necessitar de nenhum método iterativo. Segundo BAKER & KIM (2004, p.193) o estimador baseia-se na forma do teorema de Bayes de  $g_{\theta}^*(\theta_j)$  dado por:

$$g_j^*(\theta_j) = \frac{P(U_{j..} | \theta_j, \zeta) g(\theta_j | \tau)}{P(U_{j..} | \zeta, \tau)} \quad (108)$$

Dada a concepção de independência local à probabilidade do respondente  $j$  responder o instrumento é dado por:

$$P(U_j | \theta_j, \zeta) = \prod_{i=1}^n P_i(\theta_j)^{u_{ji}} Q_i(\theta_j)^{l-u_{ji}} \quad (109)$$

No denominador tem-se que:

$$P(U_j) = \int P(U_j | \theta) g(\theta) d\theta \quad (110)$$

Desta forma a esperança *a posteriori* de  $\theta_j$ , dado  $U_j$  é:

$$E(\theta_j | U_j, \zeta) = \frac{\int_{\mathbb{R}} \theta_j g(\theta) \cdot \prod_{i=1}^n P_i(\theta_j)^{u_{ji}} Q_i(\theta_j)^{l-u_{ji}} d\theta}{\int_{\mathbb{R}} g(\theta) \cdot \prod_{i=1}^n P_i(\theta_j)^{u_{ji}} Q_i(\theta_j)^{l-u_{ji}} d\theta} \quad (111)$$

Utilizando as aproximações das integrais na forma de quadratura obtém-se:

$$\hat{\theta}_j = E(\theta_j | U_j, \zeta) = \frac{\sum_{k=1}^{nq} X_k L(X_k) A(X_k)}{\sum_{k=1}^{nq} X_k L(X_k) A(X_k)} \quad (112)$$

onde  $A(X_k)$  são os pesos ajustados para valores fixos de  $X_k$  produzidos na estimação dos itens no final de cada etapa EM e  $L(X_k)$  é dado por:

$$L(X_k) = \prod_{i=1}^n P_i(X_k)^{u_{ji}} Q_i(X_k)^{l-u_{ji}} \quad (113)$$

#### 2.2.4 Cálculo do Escore Verdadeiro.

O cálculo do escore verdadeiro dos respondentes é obtido a partir dos parâmetros estimados da TRI ( $\hat{a}_i, \hat{b}_i, \hat{c}_i$  e  $\hat{\theta}_j$ ), onde  $j$  representa o respondente e  $i$  o item, sendo que este último varia de  $1$  a  $I-r$ , onde  $I$  corresponde ao total de itens e  $r$  o número de itens que não se ajustaram ao modelo (itens ruins) e  $j$  o respondente. Assim, o escore verdadeiro  $EV_j$  é dado por:

$$EV_j = \sum_{i=1}^{I-r} P(\hat{a}_i, \hat{b}_i, \hat{c}_i, \hat{\theta}_j) \quad (114)$$

onde,

$$P(\hat{a}_i, \hat{b}_i, \hat{c}_i, \hat{\theta}_j) = \hat{c}_i + (1 - \hat{c}_i) \frac{1}{1 + e^{-D\hat{a}_i(\hat{\theta}_j - \hat{b}_i)}} \quad (115)$$

Para uma melhor interpretação dos valores obtidos nos escores verdadeiros pode-se dividir o valor encontrado pelo número de itens que selecionados na análise, isto é, o total de  $(I-r)$  itens. Dessa maneira fica-se uma escala de valores de 0 a 1.

Como as habilidades  $\theta_j$  também são estimadas somente pelos itens considerados bons, os escores verdadeiros dos respondentes terão mais consistência e o respondente será melhor avaliado pelo teste.

## 2.3 ANÁLISE FATORIAL

### 2.3.1 Introdução

A Análise Fatorial foi desenvolvida pelo psicólogo Charles Spearman a mais de 100 anos atrás quando notou, observando escores de crianças em idade escolar, que estas tinham uma grande variedade de similaridades aparentemente sem conexão, mas que poderiam ser postuladas em uma habilidade mental geral, denotada por  $g$ , formando a base e moldando o desempenho cognitivo humano, isto é, as habilidades matemáticas, verbais, artísticas, raciocínios lógicos, etc. Assim, estas habilidades seriam explicadas por uma única medida chamada fator de inteligência geral ( $g$ ). Raymond Cattell expandiu esta idéia. A princípio desenvolveu uma teoria com dois fatores de inteligência, e depois de verificar a performance de seus próprios testes apresentou a teoria de multifatores para explicar inteligência. A teoria de Cattell propôs fatores alternativos no desenvolvimento intelectual, incluindo fatores de motivação e psicológicos. Cattell além de desenvolver vários métodos matemáticos para ajustamentos gráficos psicométricos, era defensor forte de Análise Fatorial e Psicometria. Ele acreditou que toda a teoria deveria ser derivada da pesquisa que apóia o uso contínuo de observações empíricas e testes, objetivando estudar inteligência humana (Wikipedia, [http://en.wikipedia.org/wiki/Factor\\_analysis](http://en.wikipedia.org/wiki/Factor_analysis) consultado em abril de 2005).

O termo Análise Fatorial foi introduzido primeiro por Thurstone em 1931 (*Electronic Statistics Textbook*, <http://www.statsoft.com/textbook/stfacan.html> consultado em abril de 2005). Para exemplificar o propósito desta análise suponha que se queira medir a satisfação de pessoas com suas vidas. Aplica-se um questionário de satisfação com vários itens; um dos quais seria, o quanto satisfeito ele está com suas atividades de lazer (item 1) e como intensamente ele está procurando uma destas atividades (item 2). Provavelmente, as respostas para os dois itens são altamente correlacionadas. Dado que há uma correlação alta entre os dois itens, pode-se concluir que eles são bastante redundantes e as duas variáveis correlacionadas poderiam resultar em um fator, isto ilustra bem a idéia básica da análise fatorial. Se o exemplo fosse estendido para mais variáveis, o princípio básico de expressar duas ou mais variáveis por fatores, reduz o

espaço de atributo de um número grande de variáveis para um número menor de fatores e como tal é um procedimento independente, isto é, nenhuma variável dependente é especificada.

Segundo HAIR, ANDERSON, TATHAM & BLACK (2005, p.91), a Análise Fatorial analisa a estrutura da correlação entre as variáveis em estudo (itens de testes, respostas em questionários, etc.) definindo um conjunto de dimensões latentes comuns chamadas fatores. Ela é uma técnica de interdependência na qual são consideradas todas as variáveis, de tal modo que os fatores são formados para maximizar seu poder de explicação do conjunto inteiro de variáveis.

O foco maior das técnicas de análise fatorial é condensar a informação contida em diversas variáveis originais em um conjunto menor (fatores) com uma perda mínima de informação, mas com a vantagem de não haver correlação entre os fatores. Assim, dois objetivos maiores podem ser alcançados: reduzir o número de variáveis e descobrir a estrutura nas relações entre variáveis.

No curso do desenvolvimento teórico da Análise Fatorial que segue nas próximas seções, será dada ênfase aos procedimentos utilizados neste trabalho.

### 2.3.2 Conceitos Fundamentais e Pressupostos Básicos

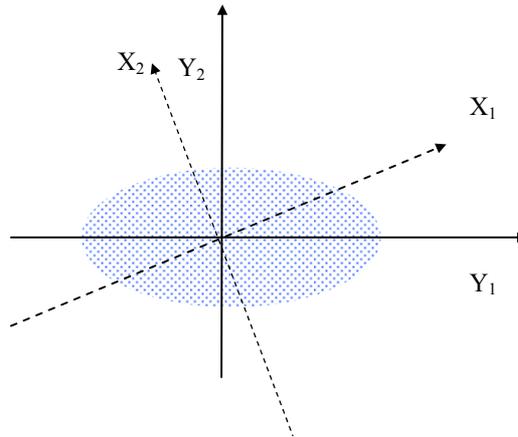
A primeira técnica em Análise Fatorial usada, originada a partir de dois trabalhos de Charles Edward Spearman: “*General intelligence objectively determined and measured*”, publicado no *American Journal of Psychology* 15, páginas 201 a 293, de 1904 e no livro “*The Abilities of Man, their nature and measurement*” da editora Macmillan de 1927, gerou várias outras. Porém a mais comumente usada é a Análise Fatorial, utilizando Componentes Principais, desenvolvida por Karl Pearson em 1901, através do artigo publicado na *Philosophical Magazine*, 2 páginas 559 a 572 intitulado “*On lines and planes of closest fit to a system of points in space*”, e completada por Harold Hotelling em “*Analysis of a complex of statistical variables into principal components*”, páginas 417 a 441 e 498 a 520, no *Journal of Educational Psychological* número 24 datado de 1933 A Análise de Componentes Principais (PCA) apresentada a seguir resolve um problema semelhante para o problema de Análise Fatorial Comum, que foi

introduzida pelo psicólogo Charles Edward Spearman, enquanto que a Análise de Componentes Principais (*PCA - Principal Components Analysis*) foi desenvolvida por um estatístico. *PCA* declara e resolve um problema estatístico bem definido, e com exceção de casos especiais sempre dá uma solução muito boa com a vantagem de possuir algumas propriedades matemáticas bem consistentes. Pode-se descrever até problemas práticos muito artificiais para os quais *PCA* provê a solução exata (FURTADO, 1999, p.37).

A Análise de Componentes Principais, segundo CHAVES NETO (2004, p.17), procura explicar a estrutura de covariância da matriz de dados ( $X$ ) através de combinações lineares ( $Y$ ) não correlacionadas das  $p$  variáveis originais. Embora  $p$  componentes sejam necessárias para reproduzir a variabilidade total do sistema, freqüentemente muito desta variabilidade pode ser explicada por um número pequeno,  $k$ , de componentes principais. Neste caso, existe quase a mesma quantidade de informação nas  $k$  componentes que nas  $p$  variáveis originais. As  $k$  componentes principais podem então substituir as  $p$  variáveis originais e, o conjunto de dados originais que consiste de  $n$  medidas das  $p$  variáveis, é reduzido para um formado por  $n$  medidas das  $k$  componentes principais.

Nas componentes principais da população, algebricamente, as componentes são combinações lineares particulares das  $p$  variáveis  $X_1, X_2, \dots, X_p$ . No entanto, geometricamente, elas representam a seleção de um novo sistema de coordenadas obtidas por rotação do sistema original com  $X_1, X_2, \dots, X_p$  como eixos. Os novos eixos  $Y_1, Y_2, \dots, Y_p$  representam as direções com variabilidade máxima e fornecem uma descrição mais simples e mais parcimoniosa da estrutura de covariância.

As componentes principais dependem da matriz de covariâncias  $\Sigma$  (ou da matriz de correlação  $\rho$ ) das v.a's  $X_1, X_2, \dots, X_p$ . O seu desenvolvimento não necessita da suposição de Gaussianidade. Por outro lado, a análise de componentes principais obtidas de populações normais multivariadas tem sua interpretação usual em termos de elipsóides de densidade constante. A figura 11 a seguir mostra as direções de eixos ortogonais originais ( $X_i$ ) e dos eixos ortogonais das componentes principais ( $Y_i$ ) colocados nas direções de maior variabilidade.



**FIGURA 11- INTERPRETAÇÃO GEOMÉTRICA DAS COMPONENTES PRINCIPAIS**

Seja o vetor aleatório denotado por  $\underline{X}' = [X_1, X_2, \dots, X_p]$  e que tem vetor de médias  $E(\underline{X}) = \underline{\mu}$  e matriz de covariância  $V(\underline{X}) = \Sigma$  com os autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , então se obtêm as combinações lineares ( $\underline{Y}$ ) dadas por:

$$\begin{aligned}
 Y_1 &= \underline{c}_1' \underline{X} = c_{11}X_1 + c_{21}X_2 + \dots + c_{p1}X_p \\
 Y_2 &= \underline{c}_2' \underline{X} = c_{12}X_1 + c_{22}X_2 + \dots + c_{p2}X_p \\
 &\vdots \\
 Y_p &= \underline{c}_p' \underline{X} = c_{1p}X_1 + c_{2p}X_2 + \dots + c_{pp}X_p
 \end{aligned} \tag{116}$$

onde  ${}_p C_p = \begin{bmatrix} c_{11} & c_{21} & \dots & c_{p1} \\ c_{12} & c_{22} & \dots & c_{p2} \\ \dots & \dots & \dots & \dots \\ c_{1p} & c_{2p} & \dots & c_{pp} \end{bmatrix}$

Logo, tem-se  $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_p \end{bmatrix} = \begin{bmatrix} c_{11} & c_{21} & \dots & c_{p1} \\ c_{12} & c_{22} & \dots & c_{p2} \\ \dots & \dots & \dots & \dots \\ c_{1p} & c_{2p} & \dots & c_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix} = {}_p C_p \underline{X}$  e das propriedades

de esperança e variância de variáveis aleatórias multivariadas tem-se que:

$$E(Y_i) = E(\underline{c}_i' \underline{X}) = \underline{c}_i' E(\underline{X}) = \underline{c}_i' \underline{\mu} \tag{117}$$

$$V(Y_i) = V(\underline{c}_i' \underline{X}) = \underline{c}_i' V(\underline{X}) \underline{c}_i = \underline{c}_i' \Sigma \underline{c}_i \tag{118}$$

$$COV(Y_i, Y_k) = \underline{c}_i' \Sigma \underline{c}_k \quad (119)$$

$$COV(\underline{Y}_p) = V(\underline{c}_p' \underline{X}) = C \Sigma C' \quad (120)$$

As componentes principais são as combinações lineares não-correlacionadas  $Y_1, Y_2, \dots, Y_p$  cujas variâncias são tão grandes quanto possíveis, assim a primeira componente principal é a combinação linear com variância máxima, ou seja, é aquela que maximiza  $V(Y_1) = V(\underline{c}_1' \underline{X})$  sujeito à restrição  $\underline{c}_1' \underline{c}_1 = 1$  (vetor de comprimento unitário); a segunda componente principal é a combinação linear que maximiza  $V(Y_2) = V(\underline{c}_2' \underline{X})$  sujeito à restrição  $\underline{c}_2' \underline{c}_2 = 1$  e assim sucessivamente até a p-ésima componente principal, que é a combinação linear que maximiza  $V(Y_p) = V(\underline{c}_p' \underline{X})$  sujeito à restrição  $\underline{c}_p' \underline{c}_p = 1$ .

Suponha que  $B_{p \times p}$  seja uma matriz positiva definida com autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  associados aos autovetores normalizados  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$ . Então

$$\underset{\underline{x} \neq 0}{\text{máx}} \frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \lambda_1 \text{ obtido quando } \underline{x} = \underline{e}_1$$

e assim sucessivamente até

$$\underset{\underline{x} \neq 0}{\text{máx}} \frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \lambda_p \text{ obtido quando } \underline{x} = \underline{e}_p$$

Além disso,

$$\underset{\underline{x} \perp \underline{e}_1, \dots, \underline{e}_k}{\text{máx}} \frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \lambda_{k+1} \text{ obtido quando } \underline{x} = \underline{e}_{k+1}, \text{ com } k = 1, 2, \dots, p-1.$$

Assim, seja  $P_{p \times p}$  uma matriz ortogonal, cujas colunas são os autovetores  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$  e seja  $\Lambda$  uma matriz diagonal com autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$  ao longo de sua diagonal. Seja  $B^{1/2} = P \Lambda^{1/2} P'$  e  $\underline{y} = P' \underline{x}$ . Conseqüentemente,  $\underline{x} \neq 0$  implica  $\underline{y} \neq 0$ . Então,

$$\frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \frac{\underline{x} B^{1/2} B^{1/2} \underline{x}}{\underline{x}' \underbrace{P P'}_{I_{p \times p}} \underline{x}} = \frac{\underline{x}' P \Lambda^{1/2} P' P \Lambda^{1/2} P' \underline{x}}{y' y} = \frac{y' \Lambda y}{y' y} = \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^p y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_1 \quad (121)$$

trocando-se

$$\underline{x} = \underline{e}_l \text{ tem-se } \underline{y} = P' \underline{e}_l = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ pois } \underline{e}_k' \underline{e}_l = \begin{cases} 1, k = l \\ 0, k \neq l \end{cases}. \quad (122)$$

Para esta escolha de  $\underline{x}$ ,  $\frac{\underline{y}' A \underline{y}}{\underline{y}' \underline{y}} = \frac{\lambda_l}{1} = \lambda_l$ , ou  $\frac{\underline{e}_l' B \underline{e}_l}{\underline{e}_l' \underline{e}_l} = \underline{e}_l' B \underline{e}_l = \lambda_l$ . Um argumento similar produz a segunda parte.

Agora  $\underline{x} = P' \underline{y} = y_1 \underline{e}_1 + y_2 \underline{e}_2 + \dots + y_p \underline{e}_p$ , como  $\underline{x} \perp \underline{e}_1, \dots, \underline{e}_k$  implica

$$0 = \underline{e}_i' \underline{x} = y_1 \underline{e}_i' \underline{e}_1 + y_2 \underline{e}_i' \underline{e}_2 + \dots + y_p \underline{e}_i' \underline{e}_p = y_i, i \leq k$$

Portanto, para  $\underline{x}$  perpendicular aos primeiros  $k$  autovetores  $\underline{e}_i$ , tem-se:

$$\frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \frac{\sum_{i=k+1}^p \lambda_i y_i^2}{\sum_{i=k+1}^p y_i^2} \quad (123)$$

Tomando  $y_{k+1} = 1, y_{k+2} = \dots = y_p = 0$ , obtem-se o máximo declarado.

Seja  $\Sigma$  a matriz de covariâncias associada ao vetor aleatório  $\underline{X}' = [X_1, X_2, \dots, X_p]$  com pares de autovalor-autovetor dados por  $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$  onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ; tem a  $i$ -ésima componente principal é dada por  $Y_i = \underline{e}_i' \underline{X}$  e tem  $V(Y_i) = \underline{e}_i' \Sigma \underline{e}_i = \lambda_i$  e  $cov(Y_i, Y_k) = 0$   $i \neq k$ . Então, se algum  $\lambda_i$  é igual a outro, na escolha do correspondente vetor de coeficientes  $\underline{e}_i$ ,  $Y_i$  então ele não é único.

Fazendo  $B = \Sigma$ , tem-se:

$$\max_{\underline{\ell} \neq 0} \frac{\underline{\ell}' \Sigma \underline{\ell}}{\underline{\ell}' \underline{\ell}} = \lambda_1 \quad (\text{obtido quando } \underline{\ell} = \underline{e}_1)$$

Se os autovetores são normalizados,  $\underline{e}_i' \underline{e}_i = 1$ , então

$$\max_{\underline{\ell} \neq 0} \frac{\underline{\ell}' \Sigma \underline{\ell}}{\underline{\ell}' \underline{\ell}} = \lambda_1 = \frac{\underline{e}_1' \Sigma \underline{e}_1}{\underline{e}_1' \underline{e}_1} = \underline{e}_1' \Sigma \underline{e}_1 = Var(Y_1)$$

Similarmente,

$$\max_{\underline{\ell} \perp \underline{e}_1, \underline{e}_2, \dots, \underline{e}_k} \frac{\underline{\ell}' \Sigma \underline{\ell}}{\underline{\ell}' \underline{\ell}} = \lambda_{k+1}, \quad k = 1, 2, \dots, p-1$$

Para a escolha de  $\underline{\ell} = \underline{e}_{k+1}$ , com  $\underline{e}_{k+1}' \underline{e}_i = 0$ , para  $i = 1, 2, \dots, k$  e  $k = 1, 2, \dots, p-1$ ,

$$\frac{\underline{e}_{k+1}' \Sigma \underline{e}_{k+1}}{\underline{e}_{k+1}' \underline{e}_{k+1}} = \underline{e}_{k+1}' \Sigma \underline{e}_{k+1} = \text{Var}(Y_{k+1}) \quad (124)$$

No entanto, da definição de autovalor e autovetor, tem-se:

$$\underline{e}_{k+1}' (\Sigma \underline{e}_{k+1}) = \underline{e}_{k+1}' (\lambda_{k+1} \underline{e}_{k+1}) = \lambda_{k+1} (\underline{e}_{k+1}' \underline{e}_{k+1}) = \lambda_{k+1} \quad (125)$$

Portanto,  $\text{Var}(Y_{k+1}) = \lambda_{k+1}$ .

Resta mostrar que  $\underline{e}_i$  é perpendicular a  $\underline{e}_k$ , dado que  $\text{Cov}(Y_i, Y_k) = 0$ . Os autovetores de  $\Sigma$  são ortogonais se todos os autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$  são distintos. Se os autovalores não são todos distintos, os autovetores correspondentes aos autovalores comuns podem ser escolhidos ortogonais. Então, para quaisquer dois autovetores  $\underline{e}_i$  e  $\underline{e}_k$ ,  $\underline{e}_i' \underline{e}_k = 0$ ,  $i \neq k$ . Desde que  $\Sigma \underline{e}_k = \lambda_k \underline{e}_k$ , a pré-multiplicação por  $\underline{e}_i'$  leva a  $\text{Cov}(Y_i, Y_k) = \underline{e}_i' \Sigma \underline{e}_k = \underline{e}_i' \lambda_k \underline{e}_k = \lambda_k \underline{e}_i' \underline{e}_k = 0$  para  $i \neq k$ , como se queria demonstrar.

As componentes principais podem ser obtidas, também, de v.a's padronizadas, ou seja, de  $Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$   $i = 1, 2, \dots, p$  que em notação matricial é  $\underline{Z} = (V^{1/2})^{-1} [\underline{X} - \underline{\mu}]$  onde

$V^{1/2}$  é a matriz desvio padrão, dada por:

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \sqrt{\sigma_{pp}} \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_p \end{bmatrix} \quad (126)$$

Verifica-se facilmente que sendo  $E(\underline{Z}) = \underline{0}$  então  $V^{1/2} \rho V^{1/2} = \Sigma$ , as componentes principais de  $\underline{Z}$  podem ser obtidas dos autovalores e autovetores da matriz de correlação  $\rho$  de  $\underline{X}$ .

Para se obter as componentes principais a partir da amostra dada por  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  que são  $n$  observações do vetor aleatório  $\underline{X}$   $p$ -dimensional com vetor de médias  $\underline{\mu}$  e matriz de covariância  $\Sigma$ , usa-se a matriz de covariância amostral  $S$  ou a matriz de correlação amostral  $R$ . Então as componentes principais da amostra são determinadas a partir dessas matrizes e são definidas como as combinações lineares que maximizam a variância amostral. Sendo assim, com  $S$  como a matriz de covariância amostral de ordem  $p \times p$  com

os pares de autovalor-autovetor  $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$  onde  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ , a  $i$ -ésima componente principal amostral é dada por

$$\hat{Y}_i = \hat{e}_i' \underline{x} = \hat{e}_{i1}x_1 + \hat{e}_{i2}x_2 + \dots + \hat{e}_{ip}x_p \quad i = 1, 2, \dots, p. \quad (127)$$

A variância amostral de  $\hat{Y}_k$  é dada por  $\hat{\lambda}_k$  onde  $k = 1, 2, \dots, p$  e valem os resultados provados para a situação populacional considerando o contexto amostral.

O modelo usado neste trabalho requer, ainda, duas condições importantes que são: linearidade e ortogonalidade. Segue o desenvolvimento do modelo ortogonal.

### 2.3.3 O Modelo de Análise Fatorial Ortogonal

Seja a variável aleatória observável  $\underline{X}$  com  $p$  componentes e oriunda de uma distribuição com média  $\underline{\mu}$  e matriz de covariância  $\Sigma$ . O modelo fatorial postula que  $\underline{X}$  é linearmente dependente sobre algumas variáveis aleatórias não observáveis  $F_1, F_2, \dots, F_m$ , chamadas fatores comuns e  $p$  fontes de variação aditivas  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , chamadas erros ou, algumas vezes, fatores específicos, dados por:

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_i - \mu_i &= \ell_{i1}F_1 + \ell_{i2}F_2 + \dots + \ell_{im}F_m + \varepsilon_i \\ &\vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (128)$$

ou em notação matricial como  $\underset{px1}{\underline{X}} - \underset{px1}{\underline{\mu}} = \underset{pxm}{L} \underset{mx1}{\underline{F}} + \underset{px1}{\underline{\varepsilon}}$

O coeficiente  $\ell_{ij}$  é chamado peso ou carregamento na  $i$ -ésima variável do  $j$ -ésimo fator, tal que a matriz  $L$  de dimensão  $(p \times m)$  é a matriz de carregamento dos fatores. Note que o fator específico ou erro  $\varepsilon_i$  é associado somente com a  $i$ -ésima resposta  $X_i$ . Os desvios  $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$  são expressos em termos de  $p+m$  variáveis aleatórias:  $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$  que não são observáveis (latentes).

Assumindo que:

$$E(\underline{F}) = \underline{0}_{m \times 1}, \quad cov(\underline{F}) = E(\underline{F}\underline{F}') = I_m \quad (129)$$

$$E(\underline{\varepsilon}) = \underline{0}_{p \times 1}, \quad cov(\underline{\varepsilon}) = E(\underline{\varepsilon}\underline{\varepsilon}') = \Psi = \begin{bmatrix} \Psi_1 & 0 & 0 & \dots & 0 \\ 0 & \Psi_2 & 0 & \dots & 0 \\ 0 & 0 & \Psi_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \Psi_p \end{bmatrix} \quad (130)$$

e com  $\underline{F}$  e  $\underline{\varepsilon}$  independentes, tem-se que  $cov(\underline{\varepsilon}, \underline{F}) = E(\underline{\varepsilon}\underline{F}') = \underline{0}_{p \times p}$  com  $m=p$ , assim com estas suposições pode-se construir o Modelo Fatorial Ortogonal com  $m$  fatores comuns como:

$$\underline{X} - \underline{\mu} = \underline{L} \underline{F} + \underline{\varepsilon} \quad (131)$$

com  $m < p$  e, onde

$\mu_i$  = média da variável  $i$ ;  $i = 1, 2, 3, \dots, p$

$\varepsilon_i$  =  $i$ -ésimo fator específico

$F_j$  =  $j$ -ésimo fator comum;  $j = 1, 2, 3, \dots, m$

$\ell_{ij}$  = carregamento da  $i$ -ésima variável no  $j$ -ésimo fator

Os vetores aleatórios não observáveis  $\underline{F}$  e  $\underline{\varepsilon}$  satisfazem as seguintes propriedades:

-  $\underline{F}$  e  $\underline{\varepsilon}$  são independentes

-  $E(\underline{F}) = \underline{0}$ ,  $Cov(\underline{F}) = I$

-  $E(\underline{\varepsilon}) = \underline{0}$ ,  $Cov(\underline{\varepsilon}) = \psi$ , onde  $\psi$  é uma matriz diagonal com a variância específica  $\psi_i$  na diagonal principal

Considerando a matriz covariância de  $\underline{X}$  e voltando ao modelo fatorial ortogonal tem-se que:

$$\begin{aligned} (\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' &= (\underline{L}\underline{F} + \underline{\varepsilon})(\underline{L}\underline{F} + \underline{\varepsilon})' \\ &= (\underline{L}\underline{F} + \underline{\varepsilon})((\underline{L}\underline{F})' + \underline{\varepsilon}') = \\ &= \underline{L}\underline{F}(\underline{L}\underline{F})' + \underline{\varepsilon}(\underline{L}\underline{F})' + \underline{L}\underline{F}\underline{\varepsilon}' + \underline{\varepsilon}\underline{\varepsilon}' \end{aligned} \quad (132)$$

a matriz de covariância de  $\underline{X}$  é :

$$\begin{aligned}
\Sigma &= \text{cov}(\underline{X}) = E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' = E[\underline{LF}(\underline{LF})' + \underline{\varepsilon}(\underline{LF})' + \underline{LF}\underline{\varepsilon}' + \underline{\varepsilon}\underline{\varepsilon}'] \\
\Sigma &= LE(\underline{FF}')L' + E(\underline{\varepsilon}\underline{\varepsilon}')L' + LE(\underline{F}\underline{\varepsilon}') + E(\underline{\varepsilon}\underline{\varepsilon}') \\
\Sigma &= \underset{p \times p}{LL'} + \underset{p \times p}{0} + \underset{p \times p}{0} + \underset{p \times p}{\Psi} \\
\Sigma &= LL' + \psi
\end{aligned} \tag{133}$$

e a matriz de covariância para o Modelo Fatorial Ortogonal é dada por:

$$\begin{aligned}
\text{cov}(\underline{X}, \underline{F}) &= E(\underline{X} - \underline{\mu})(\underline{F} - \underline{0})' = E(\underline{X} - \underline{\mu})\underline{F}' = E(\underline{\mu} + \underline{LF} + \underline{\varepsilon} - \underline{\mu})\underline{F}' \\
\text{cov}(\underline{X}, \underline{F}) &= E(\underline{LF} + \underline{\varepsilon})\underline{F}' = LE(\underline{FF}') + E(\underline{\varepsilon}\underline{F}') = LI_m + 0 \\
\text{cov}(\underline{X}, \underline{F}) &= L
\end{aligned} \tag{134}$$

Conseqüentemente temos  $V(X_i) = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 + \psi_i$

$$\text{cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \ell_{i2}\ell_{k2} + \dots + \ell_{im}\ell_{km} \tag{135}$$

$$\text{cov}(X_i, F_j) = \ell_{ij} \tag{136}$$

A porção da variância da  $i$ -ésima variável aleatória  $X_i$  advinda como contribuição dos  $m$  fatores comuns é chamada **Comunalidade** e a porção da  $V(X_i) = \sigma_i^2$  oriunda do fator específico é a **Variância Específica**, assim:

$$\begin{aligned}
V(X_i) &= V[\mu_i + \ell_{i1}F_1 + \ell_{i2}F_2 + \dots + \ell_{im}F_m + \varepsilon_i] = \\
&= \ell_{i1}^2 V(F_1) + \ell_{i2}^2 V(F_2) + \dots + \ell_{im}^2 V(F_m) + V(\varepsilon_i) = \\
&= 0 + \ell_{i1}^2 \cdot I + \ell_{i2}^2 \cdot I + \dots + \ell_{im}^2 \cdot I + \psi_i \\
V(X_i) &= \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 + \Psi_i
\end{aligned} \tag{137}$$

Assim, sendo  $V(X_i) = h_i^2 + \Psi_i$ ,  $i = 1, 2, \dots, p$  e como  $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2$ , a  $i$ -ésima comunalidade é a soma de quadrados dos carregamentos na  $i$ -ésima variável dos  $m$  fatores comuns. Portanto, a variância de qualquer variável do modelo é composta de duas partes: a comunalidade e a variância específica. Em particular, para uma variável  $i$  tem-se:

$$\text{Var}(X_i) = (\text{comunalidade } i) + (\text{variância específica } i) \tag{138}$$

Como a  $i$ -ésima comunalidade é a soma dos quadrados dos carregamentos da  $i$ -ésima variável nos  $m$  fatores comuns, ou seja,

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 \tag{139}$$

e  $\text{Var}(X_i) = \sigma_{ii}$ , então a relação resumida é  $\sigma_i^2 = \sigma_{ii} = h_i^2 + \psi_i$  com  $i = 1, 2, \dots, p$ .

O Modelo Fatorial assume que  $p + \binom{p}{2} = p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2}$  variâncias e covariâncias de  $\underline{X}$  podem ser reproduzidas pelos carregamentos fatoriais  $\ell_{ij}$  e as  $p$  variâncias específicas  $\psi_i$ . Quando  $m=p$ , qualquer matriz de covariância  $\Sigma$  pode ser reproduzida exatamente como  $LL'$ , então  $\psi$  pode ser a matriz nula. Entretanto, a análise mais usada é aquela em que  $m$  é menor do que  $p$ . Neste caso, o Modelo Fatorial produz uma simples explicação da covariância em  $\underline{X}$  com menos parâmetros em relação aos  $\frac{p(p+1)}{2}$  parâmetros de  $\Sigma$ .

Em alguns casos, quando se considera um Modelo Ortogonal Fatorial em que  $m$  é igual a 1, ou seja, tem-se apenas um fator, pode-se também obter uma única solução numérica. Entretanto, esta solução numérica pode não ser consistente com a interpretação estatística dos coeficientes, gerando, portanto, uma solução imprópria. A inconsistência da solução em relação aos coeficientes acontece quando no cálculo dos carregamentos, aparecem no modelo valores para variâncias maiores que 1 em valor absoluto, ou até mesmo valores negativos de variâncias específicas.

Para  $m > 1$ , existe sempre uma ambigüidade inerente associada com o Modelo Fatorial. Para verificar isto suponha uma matriz  $T$  de ordem  $m \times m$  ortogonal, ou seja,  $TT' = T'T = I$ . Assim, o Modelo Fatorial Ortogonal poderia ser escrito como:

$$\begin{aligned}\underline{X} - \underline{\mu} &= L\underline{F} + \underline{\varepsilon} \\ \underline{X} - \underline{\mu} &= LTT' \underline{F} + \underline{\varepsilon} \\ \underline{X} - \underline{\mu} &= (L.T).(T' \underline{F}) + \underline{\varepsilon}\end{aligned}$$

fazendo  $L.T = L^*$  e  $T' \underline{F} = \underline{F}^*$ , tem-se

$$\underline{X} - \underline{\mu} = L^*\underline{F}^* + \underline{\varepsilon} \quad (140)$$

E assim, preservam-se as suposições, mesmo com a transformação ortogonal.

$$E(\underline{F}^*) = E(T' \underline{F}) = T'E(\underline{F}) = T' \underline{0} = \underline{0} \quad (141)$$

e

$$Cov(\underline{F}^*) = Cov(T' \underline{F}) = T.Cov(\underline{F}).T = T'.I.T = T'.T = I \quad (142)$$

No modelo, não é possível distinguir os carregamentos de  $L$  dos carregamentos de  $L^*$ . Os carregamentos fatoriais de  $F$  e  $F^*$  têm as mesmas propriedades estatísticas. Mesmo que  $L$  e  $L^*$  tenham diferentes carregamentos, ambos são gerados pela mesma matriz de covariância  $\Sigma$ . Assim, não há diferença para o modelo entre  $L^* = LT$  e  $L$ , pois ambas têm a mesma representação. As comunalidades que são os elementos diagonais da matriz  $LL' = (L^*)(L^*)'$  são invariantes pela escolha de  $T$ . Logo, pode-se escrever:

$$\Sigma = LL' + \Psi = LTT'L' + \Psi = (L^*)(L^*)' + \Psi \quad (143)$$

Esta ambigüidade produz a possibilidade de se estabelecer uma rotação fatorial. Da Álgebra Matricial, sabe-se que uma transformação ortogonal corresponde a uma rígida rotação (ou reflexão) dos eixos coordenados. Mantendo inalteradas as comunalidades e as variâncias específicas, pode-se encontrar uma matriz ortogonal  $T$  de rotação que estime a estrutura de covariância da variável  $\underline{X}$  de uma forma mais simples.

#### 2.3.4 Seleção do Número de Fatores.

A adequabilidade do modelo e a interpretação dos resultados dependem significativamente do número de fatores adotado. Existem vários critérios para se determinar este valor.

O critério de Kaiser seleciona o número de fatores escolhendo aqueles que possuem os autovalores maiores que 1. Henry Kaiser, em trabalhos datados de 1960, "*The application of electronic computers to factor analysis*", *Educational and Psychological Measurement* 20, p.141-151; e 1974, "*An index of factorial simplicity*", *Psychometrika* 39, p.31-36; propõe que o número de fatores  $m$  é igual ao número de autovalores maiores que a média dos autovalores. A média dos autovalores para  $R$  é  $I$  e

para  $S$  é  $\frac{\sum_{i=1}^p \lambda_i}{p}$ . Assim como o critério é fundamentalmente heurístico e está presente em

vários programas computacionais estatísticos existem boas razões para se confirmar a sua confiabilidade, mostrando que funciona muito bem na prática. Outras linhas de pensamento surgiram posteriormente, mas desde que é suposto que a análise de componentes principais sumariza um conjunto de dados, usar um componente que

explica menos que uma discrepância de 1 é algo que não faz muito sentido. Porém, a justificativa principal de Kaiser para a regra foi que ele fez várias simulações com números diferentes de fatores, e viu que o critério proposto fez sentido.

Um método alternativo chamado Teste Scree foi sugerido por Raymond B. Cattell no artigo “*The scree test for the number of factors*” em 1966 na *Multivariate Behavioral Research*, 1(2), p.140-161. Em um gráfico plota-se o número de autovalores no eixo horizontal e a medida correspondente dos autovalores no eixo vertical. Escolhe-se o número de fatores  $m$  igual ao número de autovalores que estão na primeira linha íngreme, isto é, descartam-se os autovalores que estão na linha mais próxima do eixo horizontal. É importante ressaltar que um modelo que apresente um número muito grande de pequenos autovalores, segundo este critério, pode apresentar baixa precisão.

Outro critério, também bem parecido ao Teste Scree, mas envolve mais cálculos e menos gráficos é aquele que para cada autovalor  $\lambda$ , define-se um valor  $s$  como a soma de todos os autovalores posterior a ele. Então  $\lambda/s$  é a proporção de discrepância previamente inexplicada por  $\lambda$ . Suponha um problema com 7 variáveis e no 4º autovalor os valores posteriores fossem 0,8; 0,2; 0,15; e 0,1. A soma  $s$  resulta em 1,25 e este valor é a quantia de discrepância não explicada por um modelo de 3 fatores. Sendo  $0,8/1,25=0,64$ ; acrescentando mais um fator ao modelo de 3 fatores explicariam 64% de discrepância previamente não explicada. Um cálculo semelhante para o quinto autovalor resulta em  $0,2 / (0,2+0,15+0,1) = 0,44$ , assim o quinto componente principal explica só 44% de discrepância previamente não explicada. (DARLINGTON em <http://comp9.psych.cornell.edu/Darlington/factor.htm> consultado em setembro de 2004).

Existem outros critérios, além dos já descritos, citados na página da internet de GARSON (<http://www2.chass.ncsu.edu/garson/pa765/factor.htm> consultado em janeiro de 2005), que são:

- Alguns investigadores usam a regra de manter mais fatores, que responderiam simplesmente por 90% (às vezes 80%) da variação.

- Critério de Joliffe: uma regra menos usada, mais liberal que pode resultar no dobro de fatores obtidos no critério de Kaiser. A regra de Joliffe é escolher todos os componentes com autovalores acima de 0,7.

### 2.3.5 Métodos de Estimação do Modelo Fatorial

Dadas as observações  $x_1, x_2, \dots, x_n$  do vetor  $\underline{X}$  formado por  $p$  variáveis, geralmente correlacionadas a Análise Fatorial procura verificar se o modelo fatorial representará os dados adequadamente com um número baixo de fatores. Dado que a matriz de covariância amostral  $S$  é um estimador da matriz de covariâncias populacional desconhecida  $\Sigma$  e se os elementos fora da diagonal de  $S$  são baixos ou equivalentemente na matriz de correlação amostral  $R$  são praticamente nulos e as variáveis não são relacionadas então a Análise Fatorial não é útil. Caso contrário, o modelo fatorial pode ser usado e o problema inicial é estimar os carregamentos  $\ell_{ij}$  e as variâncias específicas  $\psi_i$ . Nessas circunstâncias, os fatores específicos realizam uma importante função na Análise Fatorial, já que a proposta principal é determinar poucos, mas importantes, fatores comuns.

Os métodos mais populares de estimação dos carregamentos são: o Método das Componentes Principais e o Método da Máxima Verossimilhança, apresentados a seguir.

#### 2.3.5.1- Método das componentes principais

Seja  $\Sigma$  a matriz de covariâncias de  $\underline{X}$ , então, dado que  $\Sigma$  é positiva definida, pode-se decompô-la, segundo o Teorema da Decomposição Espectral, na forma a seguir:

$$\Sigma = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \dots + \lambda_p e_p e_p' \quad (144)$$

ou na forma matricial

$$\Sigma = [\sqrt{\lambda_1} e_1 \quad \sqrt{\lambda_2} e_2 \quad \dots \quad \sqrt{\lambda_p} e_p] \begin{bmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \dots \\ \sqrt{\lambda_p} e_p' \end{bmatrix} = \underset{p \times p}{LL'} \quad \text{se } m = p \text{ e } \psi_i = 0 \quad \forall_i \quad (145)$$

Assim, se  $\Sigma = LL' + \psi$  tem-se  ${}_p\psi_p = {}_p\theta_p$  no ajuste do modelo fatorial. Exceto pelo escalar  $\sqrt{\lambda_j}$ , os carregamentos no  $j$ -ésimo fator são os coeficientes populacionais na  $j$ -ésima componente principal. Embora a representação de  $\Sigma = LL' + \theta = LL'$  seja exata, ela não é particularmente útil (tem muitos fatores comuns). É preferível um modelo que

explique a estrutura de covariância em termos de poucos fatores comuns, isto é,  $m < p$ . Se  $p - m$  autovalores são baixos, é possível negligenciar a contribuição de  $\lambda_{m+1} \underline{e}_{m+1} \underline{e}_{m+1}' + \lambda_{m+2} \underline{e}_{m+2} \underline{e}_{m+2}' + \dots + \lambda_p \underline{e}_p \underline{e}_p'$  para  $\Sigma$  na decomposição espectral. Assim, tem-se,

$$\Sigma \approx [\sqrt{\lambda_1} \underline{e}_1 \quad \sqrt{\lambda_2} \underline{e}_2 \quad \dots \quad \sqrt{\lambda_m} \underline{e}_m] \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1' \\ \sqrt{\lambda_2} \underline{e}_2' \\ \dots \\ \sqrt{\lambda_m} \underline{e}_m' \end{bmatrix} = LL' \quad \text{de ordem } p \times p \quad (146)$$

Esta representação aproximada assume que os fatores específicos  $\underline{e}$  são de menor importância e podem também ser ignorados na fatorização de  $\Sigma$ . Se os fatores específicos  $\underline{e}$  são incluídos no modelo, suas variâncias são os elementos da diagonal da matriz diferença  $\Sigma - LL'$  e conseqüentemente  $\psi_i = \sigma_{ii} - \sum_{j=1}^m \ell_{ij}^2$  para  $i = 1, 2, \dots, p$ .

Na prática é utilizada uma amostra das observações e é possível padronizar as variáveis. Para que isto ocorra a cada vetor de observações amostradas  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  são subtraídos o vetor médio amostral  $\bar{\underline{x}}$  divididos pelos seus respectivos desvios padrões. Assim tem-se a forma das variáveis padronizadas  $\underline{z}_j$ .

$$\underline{z}_j = \begin{bmatrix} \frac{x_{11} - \bar{x}_1}{\sqrt{S_{11}}} \\ \frac{x_{21} - \bar{x}_2}{\sqrt{S_{22}}} \\ \dots \\ \frac{x_{pj} - \bar{x}_p}{\sqrt{S_{pp}}} \end{bmatrix} \quad \text{onde } j = 1, 2, \dots, n \quad (147)$$

Para variáveis padronizadas, a matriz  $S$  que estima a matriz de covariância  $\Sigma$ , é trocada pela matriz de correlação amostral  $R$ , que estima a matriz de correlação populacional  $\rho$ . Uma vez que a matriz de correlação é a matriz de covariância de variáveis padronizadas.

A Análise de Componentes Principais de uma matriz de covariância amostral  $S$  é realizada em função de seus pares de autovalor-autovetor  $(\hat{\lambda}_1, \hat{\underline{e}}_1), (\hat{\lambda}_2, \hat{\underline{e}}_2), \dots, (\hat{\lambda}_p, \hat{\underline{e}}_p)$ , onde

$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$ . Sendo  $m$  o número de fatores comuns, a matriz dos carregamentos fatoriais estimados  $\{\hat{\ell}_{ij}\}$  é dada por:

$$\hat{L} = \left[ \sqrt{\hat{\lambda}_1} \cdot \underline{\hat{e}}_1 : \sqrt{\hat{\lambda}_2} \cdot \underline{\hat{e}}_2 : \dots : \sqrt{\hat{\lambda}_m} \cdot \underline{\hat{e}}_m \right] \quad (148)$$

As variâncias específicas estimadas são produzidas pelos elementos diagonais da matriz  $S - \hat{L}\hat{L}'$ , então

$$\hat{\Psi} = \begin{bmatrix} \hat{\Psi}_1 & 0 & \dots & 0 \\ 0 & \hat{\Psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\Psi}_p \end{bmatrix} \text{ com } \hat{\Psi}_i = s_{ii} - \sum_{j=1}^m \hat{\ell}_{ij}^2 \quad (149)$$

e as comunicações são estimadas através de  $\hat{h}_{ij}^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2$ .

A análise de componentes principais da matriz de correlação amostral é obtida substituindo  $R$  no lugar de  $S$ .

No estudo do Método de Coeficientes Principais, os primeiros coeficientes ponderados das componentes principais amostrais são os carregamentos fatoriais. Por esse fato, é que se denomina o modelo descrito como Modelo de Componentes Principais.

### 2.3.5.2 Método da máxima verossimilhança.

Segundo JOHNSON & WICHERN (1992, p. 392-394), na Análise Fatorial, se os fatores comuns  $\underline{F}$  e fatores específicos  $\underline{\varepsilon}$  são considerados normalmente distribuídos, então as estimativas dos carregamentos fatoriais e variâncias específicas podem ser obtidas pelo Método da Máxima Verossimilhança. Quando  $F_j$  e  $\varepsilon_j$  são conjuntamente normais, as observações  $X_j - \mu = LF_j + \varepsilon_j$  são também normais. Assim sendo, da verossimilhança de populações normalmente distribuídas demonstradas pelos autores no resultado 4.9 nas páginas 137 a 139, tem-se:

$$L(\underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2}} \cdot \frac{1}{|\Sigma|^{n/2}} \cdot e^{-\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( \sum_{j=1}^n (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})' \right) \right] + \frac{n}{2} (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu})} \quad (150)$$

Desmembrando as somas nos expoentes em produtos de potências na mesma base, tem-se:

$$L(\underline{\mu}, \Sigma) = (2\pi)^{\frac{-(n-1)p}{2}} \cdot |\Sigma|^{-\frac{(n-1)}{2}} \cdot e^{-\left(\frac{1}{2}\right)tr\left[\Sigma^{-1}\left(\sum_{j=1}^n (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})'\right)\right]} \cdot (2\pi)^{\frac{-p}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot e^{-\left(\frac{n}{2}\right)(\bar{\underline{x}} - \underline{\mu})'\Sigma^{-1}(\bar{\underline{x}} - \underline{\mu})} \quad (151)$$

Já que  $\Sigma = LL' + \Psi$ , a função de verossimilhança depende de  $L$  e  $\Psi$ . Pela multiplicidade das escolhas de  $L$ , possíveis por transformações ortogonais, é conveniente impor uma condição de singularidade computacional. A condição de singularidade é que

$$L' \Psi^{-1} L = \Delta \quad (152)$$

seja uma matriz diagonal.

O Método da máxima verossimilhança estima  $\hat{L}$  e  $\hat{\Psi}$  maximizando a equação (144), pela propriedade da invariância de estimadores de máxima verossimilhança. Desta forma, as comunalidades estimadas são:

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2 \quad \text{para } i = 1, 2, \dots, p \quad (153)$$

assim

$$\left( \begin{array}{l} \text{Porção do total da variância} \\ \text{amostral atribuída ao } j\text{-ésimo fator} \end{array} \right) = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \dots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \dots + s_{pp}} \quad (154)$$

Da mesma forma, como foi usada a matriz de covariância para estimar a matriz de carregamentos e a matriz de variâncias específicas, poderia se usar a matriz de correlação. Para tanto, bastava que as variáveis fossem padronizadas.

As variáveis padronizadas  $Z$  tem a forma  $Z = V^{-1/2}(\underline{X} - \underline{\mu})$ . Da relação entre as matrizes de correlação e covariância pode-se escrever:

$$\rho = V^{-1/2} \Sigma V^{-1/2} \quad (155)$$

Substituindo  $\Sigma = LL' + \Psi$  na equação anterior:

$$\rho = (V^{-1/2}L)(V^{-1/2}L)' + V^{-1/2} \Psi V^{-1/2} \quad (156)$$

Assim,  $\rho$  pode ser de maneira análoga ao Modelo Fatorial Ortogonal da matriz de covariância. Com matriz de carregamentos  $L_z = V^{-1/2}L$  e matriz de variância específica  $\Psi_z = V^{-1/2} \Psi V^{-1/2}$ . Pela propriedade da invariância dos estimadores de máxima verossimilhança, o estimador de máxima verossimilhança de  $\rho$  é

$$\hat{\rho} = (\hat{V}^{-1/2} \hat{L})(\hat{V}^{-1/2} \hat{L})' + \hat{V}^{-1/2} \hat{\Psi} \hat{V}^{-1/2} \quad (157)$$

$$\text{ou} \quad \hat{\rho} = \hat{L}_z \hat{L}_z' + \hat{\Psi}_z \quad (158)$$

Como conseqüência desta fatoração, sempre que a análise de máxima verossimilhança é aplicada à matriz de correlação, a importância dos fatores é avaliada com base nos elementos da matriz de carregamentos padronizada  $\hat{L}_z$ . Desta forma, a porção da variância atribuída a um fator é dada por:

$$\left( \begin{array}{l} \text{Porção do total da variância amostral} \\ \text{padronizada atribuída ao } j\text{-ésimo fator} \end{array} \right) = \frac{{}_z \hat{\ell}_{1j}^2 + {}_z \hat{\ell}_{2j}^2 + \dots + {}_z \hat{\ell}_{pj}^2}{p} \quad (159)$$

onde  ${}_z \hat{\ell}_{ij}$  representam os carregamentos de  $\hat{L}_z$ .

Ordinariamente as observações são padronizadas e a matriz de correlação amostral é fatorada e analisada. A matriz de correlação amostral  $R$ , é trocada por  $[(n-1)/n]S$  na função de verossimilhança (144), e as estimativas de máxima verossimilhança  $\hat{L}_z$  e  $\hat{\Psi}_z$  são obtidas usando método de numéricos de maximização. Embora a verossimilhança seja apropriada para  $S$ , não  $R$ , surpreendentemente, esta prática é equivalente a obter as estimativas de máxima verossimilhança  $\hat{L}$  e  $\hat{\Psi}$  baseadas na matriz de covariância amostral  $S$ , fixando  $\hat{L}_s = \hat{V}^{-1/2} \hat{L}$  e  $\hat{\psi}_s = \hat{V}^{-1/2} \hat{\psi} \hat{V}^{-1/2}$ . Neste caso,  $\hat{V}^{-1/2}$  é a matriz diagonal, cuja diagonal principal é formada pelo inverso dos desvios padrões amostrais, computados com o divisor  $\sqrt{n}$ .

Por outro lado, dados os carregamentos estimados  $\hat{L}_s$  e as variâncias específicas  $\hat{\psi}_s$  obtidas da matriz de correlação amostral  $R$ , as estimativas resultantes de máxima verossimilhança para a análise fatorial da matriz de covariância  $[(n-1)/n]S$  são  $\hat{L} = \hat{V}^{-1/2} \hat{L}_s$  e  $\hat{\psi} = \hat{V}^{-1/2} \hat{\psi}_s \hat{V}^{-1/2}$ , ou  $\hat{\ell}_{ij} = {}_s \hat{\ell}_{ij} \sqrt{\hat{\sigma}_{ij}^2}$  e  $\hat{\psi}_i = {}_z \hat{\psi}_i \hat{\sigma}_{ij}^2$  onde  $\hat{\sigma}_{ij}^2$  é a variância amostral calculada com divisor  $n$ . A distinção entre divisores pode ser ignorada na solução de componentes principais.

### 2.3.6 Rotação de Fatores

A rotação fatorial é uma ferramenta importante e necessária para facilitar a interpretação dos fatores. O objetivo principal é simplificar a estrutura fatorial e deixá-la teoricamente mais clara.

A rotação fatorial significa girar os eixos de referência dos fatores em torno da origem até que alguma posição mais adequada seja alcançada para que ofereça a interpretação mais adequada das variáveis examinadas. Dado que o primeiro fator explica a quantia maior da variância e os demais correspondem, na maioria dos casos, a quantia residual da variância; o efeito final de fazer a rotação é redistribuir a variância dos primeiros fatores para os últimos com o objetivo de atingir um padrão fatorial mais simples e teoricamente mais significativos. O caso mais simples, segundo HAIR, ANDERSON, TATHAM & BLACK (2005, p.105-106 ) é a rotação ortogonal, na qual os eixos são mantidos a 90 graus. Caso a rotação tenha outra angulação ter-se-á uma rotação oblíqua.

A maioria dos autores pesquisados neste trabalho reconhece que fazer rotação dos fatores melhora a interpretação e que a melhor opção é escolher um método ortogonal por duas razões. Primeiro porque os pacotes computacionais com análise fatorial contêm na sua maioria somente rotações ortogonais e segundo devido aos procedimentos analíticos para rotações oblíquas não serem bem desenvolvidos e ainda estão sujeitos a controvérsias, segundo HAIR, ANDERSON, TATHAM & BLACK (2005, p.106). Embora já tenha passado sete anos nada foi encontrado que contradiga as duas afirmações no decorrer deste estudo, além disso, nenhuma regra específica foi desenvolvida para orientar o pesquisador na escolha de uma técnica ortogonal ou oblíqua, no entanto se o objetivo da pesquisa é reduzir o número de variáveis originais, independente de quão significativos possam ser os fatores resultantes ou ele quer reduzir um número maior de variáveis a fim de obter um número menor de variáveis não-correlacionadas para serem usadas em técnicas de previsão às rotações ortogonais são mais apropriadas. Porém, se o objetivo da análise for obter diversos fatores teoricamente significativos, a solução oblíqua é a recomendada.

Para rotações ortogonais três abordagens principais foram desenvolvidas que são os métodos Quartimax, Varimax e Equimax.

O método de rotação Quartimax foi o primeiro critério analítico para a determinação psicológica de fatores interpretáveis e é devido a Carrol e citado por Henry Kaiser em um trabalho intitulado “*Note on Carrol’s analytic simple structure*”, *Psychometrika*, 21, p.89-92 em 1953. O método fundamenta-se em rotacionar o fator inicial de modo que uma variável tenha carga alta em um fator e cargas tão baixas quanto possíveis em todos os outros, assim muitas variáveis podem ter carga alta no mesmo fator. O método não foi bem sucedido na produção de estruturas mais simples.

O método Varimax tem sido muito bem sucedido como uma abordagem analítica para obtenção de uma rotação ortogonal. Por esta razão e por estar de acordo com o propósito deste trabalho, foi o método escolhido e será detalhado a seguir.

#### 2.3.6.1 Método de rotação Varimax

O método Varimax de rotação ortogonal foi proposto por Kaiser em 1958. Este nome se deve ao fato de que o método busca a máxima variância das variáveis. Por isso, o nome Varimax. Kaiser propôs por meio de sua tese de doutoramento intitulada “*The varimax method of factor analysis*”, que isto pode ser conseguido, a partir de uma maximização iterativa de uma função de carregamentos quadrados. A descrição do método é feita a seguir.

Os carregamentos obtidos a partir dos carregamentos iniciais, mediante uma transformação ortogonal têm a mesma habilidade para reproduzir a matriz de covariância ou de correlação. Da álgebra matricial, nós sabemos que uma transformação ortogonal corresponde a uma rotação rígida dos eixos coordenados.

Sendo  $L$  uma matriz ( $p \times m$ ) de carregamentos não rotacionados, e seja  $T$  uma matriz ortogonal ( $m \times m$ ), a matriz de carregamentos quadrados é  $L^* = L.T$ , ou seja,  $\ell_{ij}^*$  representa os carregamentos rotacionados da  $i$ -ésima variável no  $j$ -ésimo fator.

Segundo, FURTADO (1999, p. 85), a função  $\phi$  que o critério Varimax maximiza é a soma das variâncias dos carregamentos quadrados dentro de cada coluna da matriz de

carregamentos, onde cada linha dos carregamentos é normalizada pela comunalidade, ou seja,

$$\phi = \sum_{j=1}^m \sum_{i=1}^p (d_{ij}^2 - \bar{d}_j)^2 = \sum_{j=1}^m \sum_{i=1}^p d_{ij}^4 - p \cdot \sum_{j=1}^m \bar{d}_j^2 \quad (160)$$

onde

$$d_{ij} = \frac{\ell_{ij}^*}{h_i} \quad \text{e} \quad \bar{d}_j = p^{-1} \cdot \sum_{i=1}^p d_{ij}^2$$

Na função  $\phi$ , considerando-se todas as variáveis,  $\bar{d}_j$  representa a média dos quadrados dos carregamentos normalizados de cada fator  $j$ ;  $d_{ij}^2$  representa todos os quadrados dos carregamentos normalizados. Quando se toma o máximo de  $\phi$ , impõe-se a condição de que o quadrado dos desvios em relação à média de cada fator seja o maior valor possível. Com isso se consegue carregamentos extremos, ou seja, ou muito alto ou muito baixo. E assim, obtém-se a máxima variância.

No Critério Varimax  $\phi$  é uma função de  $T$ , e o algoritmo iterativo encontra uma matriz  $T$  que maximiza  $\phi$ .

No caso onde  $m=2$  fatores, os cálculos são simplificados. A matriz  $T$  dada por

$$T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

que faz a rotação dos eixos coordenados no sentido horário de um ângulo de medida  $\theta$ .

Assim, considerando-se apenas 2 fatores ( $m=2$ ), a matriz de carregamentos rotacionados ficaria:

$$L_{px2}^* = L_{px2} \cdot T_{2x2} \quad (161)$$

ou seja;

$$\begin{bmatrix} \ell_{11}^* & \ell_{12}^* \\ \ell_{21}^* & \ell_{22}^* \\ \vdots & \vdots \\ \ell_{p1}^* & \ell_{p2}^* \end{bmatrix} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \vdots & \vdots \\ \ell_{p1} & \ell_{p2} \end{bmatrix} \cdot \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (162)$$

assim,

$$\ell_{i1}^* = \ell_{i1} \cos \theta - \ell_{i2} \sin \theta \quad \text{e} \quad \ell_{i2}^* = \ell_{i1} \sin \theta + \ell_{i2} \cos \theta \quad \text{para } i = 1, 2, \dots, p$$

e dividindo pelas raízes quadradas das correspondentes comunalidades, isto é, normalizando as linhas, tem-se:

$$d_{i1} = \frac{\ell_{i1}^*}{h_i} = \frac{\ell_{i1} \cos \theta - \ell_{i2} \sin \theta}{h_i} \text{ e } d_{i2} = \frac{\ell_{i2}^*}{h_i} = \frac{\ell_{i1} \sin \theta + \ell_{i2} \cos \theta}{h_i}, i = 1, 2, \dots, p \quad (163)$$

de onde se obtém a matriz D(px2) dada por,

$$D = \left[ \left( \frac{\ell_{i1} \cdot \cos \theta - \ell_{i2} \cdot \sin \theta}{h_i} \right) \quad \left( \frac{\ell_{i1} \cdot \sin \theta + \ell_{i2} \cdot \cos \theta}{h_i} \right) \right], i = 1, 2, \dots, p. \quad (164)$$

Seja  $G_{a,b}$  uma função dos carregamentos não rotacionados e das comunalidades definida por:

$$G_{a,b} = \sum_{i=1}^p \frac{\ell_{i1}^a \cdot \ell_{i2}^b}{h_i^{a+b}} \quad (165)$$

Se a equação (160) para  $m=2$  é

$$\phi = \sum_{j=1}^2 \sum_{i=1}^p d_{ij}^4 - p \cdot \sum_{j=1}^2 \bar{d}_j^2 \quad (166)$$

então,

$$\phi = (d_{11}^4 + d_{21}^4 + \dots + d_{p1}^4) + (d_{12}^4 + d_{22}^4 + \dots + d_{p2}^4) - p \cdot (\bar{d}_1^2 + \bar{d}_2^2) \quad (167)$$

substituindo (163) em (160), vem

$$\begin{aligned} \phi = & \sum_{i=1}^p \left[ \left( \frac{\ell_{i1} \cos \theta - \ell_{i2} \sin \theta}{h_i} \right)^4 + \left( \frac{\ell_{i1} \sin \theta + \ell_{i2} \cos \theta}{h_i} \right)^4 \right] - \\ & - p \cdot \sum_{i=1}^p \left\{ \left[ p^{-1} \cdot \left( \frac{\ell_{i1} \cos \theta - \ell_{i2} \sin \theta}{h_i} \right)^2 \right]^2 + \left[ p^{-1} \cdot \left( \frac{\ell_{i1} \sin \theta + \ell_{i2} \cos \theta}{h_i} \right)^2 \right]^2 \right\} \end{aligned} \quad (168)$$

Desenvolvendo a equação (168) e levando-se em conta os resultados da álgebra, e agrupando em função de (166), a função  $\phi$  ficaria:

$$\phi = \frac{I}{4} \cdot [\cos(4\theta) \cdot A + \sin(4\theta) \cdot B + C]$$

ou 
$$4\phi = \cos(4\theta) \cdot A + \sin(4\theta) \cdot B + C \quad (169)$$

onde

$$A = (G_{0,4} + G_{4,0} - 6G_{2,2} - G_{0,2}^2 - G_{2,0}^2 + 2G_{0,2}G_{2,0} + 4G_{1,1}^2) \quad (170)$$

$$B = 4(G_{1,3} - G_{3,1} - G_{1,1}G_{0,2} + G_{1,1}G_{2,0}) \quad (171)$$

$$C = p(3[G_{2,0} + G_{0,2}]^2 - [3G_{0,2}^2 + 3G_{2,0}^2 + 2G_{0,2}G_{2,0} + 4G_{1,1}^2]) \quad (172)$$

Multiplicando e dividindo os 2 primeiros termos do segundo membro de (169) por  $(A^2 + B^2)^{1/2} \neq 0$ , tem-se:

$$4\phi = (A^2 + B^2)^{1/2} \left[ \cos(4\theta) \cdot \frac{A}{(A^2 + B^2)^{1/2}} + \sin(4\theta) \cdot \frac{B}{(A^2 + B^2)^{1/2}} \right] + C \quad (173)$$

fazendo  $\cos \alpha = \frac{A}{(A^2 + B^2)^{1/2}}$  e  $\sin \alpha = \frac{B}{(A^2 + B^2)^{1/2}}$ , e substituindo em (173), vem :

$$4\phi = (A^2 + B^2)^{1/2} [\cos(4\theta) \cdot \cos \alpha + \sin(4\theta) \cdot \sin \alpha] + C = (A^2 + B^2)^{1/2} \cdot \cos(4\theta - \alpha) + C \quad (174)$$

O máximo valor de  $\phi$  é obtido para  $4\theta = \alpha$ . O valor de  $\alpha$  é obtido de

$$\tan \alpha = \frac{B}{A}$$

e, ainda, para localizar o quadrante, consideram-se os sinais do seno e co-seno da seguinte forma:

- se  $A > 0$  e  $B > 0$ , então  $\alpha$  pertence ao 1° quadrante;
- se  $A < 0$  e  $B > 0$ , então  $\alpha$  pertence ao 2° quadrante;
- se  $A < 0$  e  $B < 0$ , então  $\alpha$  pertence ao 3° quadrante;
- se  $A > 0$  e  $B < 0$ , então  $\alpha$  pertence ao 4° quadrante.

No caso em que  $m > 2$  (mais do que 2 fatores), uma solução iterativa para a rotação é usada. Nessa solução os fatores são rotacionados aos pares, como no método anteriormente descrito. Assim, o primeiro e segundo fatores são rotacionados de um ângulo  $\theta$ . O novo primeiro fator é então rotacionado com o terceiro fator original, e assim por diante, até que todos os  $\binom{m}{2} = \frac{m(m-1)}{2}$  pares de fatores tenham sido rotacionados.

Esta seqüência de rotações é chamada ciclo. Desta forma, cada ciclo reúne todos os passos do Método Varimax para 2 fatores apenas. Estes ciclos são repetidos até que se tenha satisfeito algum critério de convergência predeterminado. A convergência se dará aos pares. Quando praticamente não houver mais diferença nos carregamentos de um

ciclo para o outro, a solução ótima estará próxima. Pode-se estabelecer também o número de iterações como critério de parada.

Existem outras variações do método Varimax, que estão explicadas e demonstradas por FURTADO (1999 p.87-109).

O Método Equimax é uma espécie junção dos dois primeiros métodos o Quartimax e Varimax, porém não tem obtido ampla aceitação e é pouco usado (HAIR, ANDERSON, TATHAM & BLACK, 2005, p.106).

Um método de rotação oblíqua é a chamada Rotação Oblíqua Direta, que é um método padrão quando se deseja uma solução não ortogonal, ou seja, quando é permitido que os fatores sejam correlacionados. Isto resultará em autovalores mais altos, porém, com interpretabilidade diminuída. Outro método oblíquo é a Rotação de Promax que é um método alternativo e computacionalmente mais rápido que o método de Rotação Oblíqua Direta, por esta razão às vezes é usado para conjuntos de dados muito grandes.

### 2.3.7 Escores Fatoriais

Na Análise Fatorial, segundo CHAVES NETO (2004, p. 29), o interesse está usualmente centrado nos parâmetros do modelo. Entretanto, em determinados problemas, os valores estimados dos fatores comuns, denominados escores fatoriais, podem também ser requeridos. Essas quantidades são freqüentemente usadas para propósitos de diagnósticos como também para introduzir uma análise posterior.

Escores fatoriais são estimativas dos valores para os vetores fatoriais aleatórios não observáveis  $F_j$ ,  $j = 1, 2, \dots, n$ . Assim, os escores fatoriais ( $\hat{f}_j$ ) são estimativas dos valores  $f_j$ , obtidos de  $F_j$  ( $j$ -ésimo caso).

A estimação é complicada pelo fato de que as quantidades  $f_j$  e  $\varepsilon_j$  não são observadas como  $x_j$ . Para resolver este problema, métodos heurísticos foram pesquisados e várias aproximações têm sido obtidas da estimação dos valores fatoriais. Nesta tese, dois métodos de estimação serão descritos: o Método dos Mínimos Quadrados Ponderados e o Método da Regressão.

Os dois métodos têm dois pontos em comum na estimação: tratam os carregamentos fatoriais estimados  $\hat{\ell}_{ij}$ , e as variâncias específicas  $\hat{\psi}_i$ , como se fossem os verdadeiros valores e envolvem transformações lineares dos dados originais, muitas vezes centrados ou padronizados.

Tipicamente, os carregamentos fatoriais rotacionados estimados são mais usados do que os carregamentos fatoriais não rotacionados. As fórmulas apresentadas aqui não diferenciam o uso de cada um. Assim sendo, tanto os carregamentos fatoriais rotacionados quanto os não rotacionados podem ser usados. Apresenta-se a seguir o Método dos Mínimos Quadrados Ponderados.

### 2.3.7.1 Método dos mínimos quadrados ponderados

FURTADO (1999, p.114) lembra que para este modelo a média  $\underline{\mu}$ , a matriz de carregamentos fatoriais  $L$ , e a matriz de variância específica  $\Psi$ , são conhecidas do modelo fatorial

$$\underset{(px1)}{\underline{X}} - \underline{\mu} = \underset{(pxm)}{L} \cdot \underset{(mx1)}{\underline{F}} + \underset{(px1)}{\underline{\varepsilon}} \quad (175)$$

e que os fatores específicos  $\underline{\varepsilon}' = [\underline{\varepsilon}_1, \underline{\varepsilon}_2, \dots, \underline{\varepsilon}_p]$  são erros. Desde que  $Var(\underline{\varepsilon}_i)$  não precise ser igual a  $\psi_i$ ,  $i = 1, 2, \dots, p$ , Bartlett, citado por JOHNSON e WICHERN (1992, p.410), sugeriu que os valores fatoriais comuns sejam estimados pelo Método dos Mínimos Quadrados Ponderados.

Seja a soma dos erros quadrados, ponderada pelos inversos de suas variâncias, dada por:

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_i} = \underline{\varepsilon}' \Psi^{-1} \underline{\varepsilon} = (\underline{x} - \underline{\mu} - L\underline{f})' \Psi^{-1} (\underline{x} - \underline{\mu} - L\underline{f}) \quad (176)$$

A proposta é a escolha de estimadores  $\hat{\underline{f}}$  no intuito de minimizar a expressão anterior. Pode-se demonstrar que a solução para o estimador do Método de Mínimos Quadrados Ponderados é

$$\hat{\underline{f}} = (L' \Psi^{-1} L)^{-1} L' \Psi^{-1} (\underline{x} - \underline{\mu}) \quad (177)$$

Dado que o menor erro quadrado é zero, deve-se igualar a zero a última parte da expressão da soma dos erros quadrados. Assim,

$$(\underline{x} - \underline{\mu} - L\underline{\hat{f}})' \Psi^{-1} (\underline{x} - \underline{\mu} - L\underline{\hat{f}}) = 0 \quad (178)$$

A nulidade só ocorre quando:

$$\underline{x} - \underline{\mu} - L\underline{\hat{f}} = \underline{0} \quad (179)$$

logo  $L\underline{\hat{f}} = \underline{x} - \underline{\mu}$  e multiplicando os dois lados da expressão por  $L^{-1}$  tem-se que :

$$L^{-1}(L\underline{\hat{f}}) = L^{-1}(\underline{x} - \underline{\mu}) \quad (180)$$

isolando  $\underline{\hat{f}}$  na expressão obtém-se:

$$\underline{\hat{f}} = L^{-1}(\underline{x} - \underline{\mu}) = L^{-1}(\Psi\Psi^{-1})(\underline{x} - \underline{\mu}) = L^{-1}(\Psi((L')^{-1}L')\Psi^{-1})(\underline{x} - \underline{\mu}) \quad (181)$$

pela associatividade do produto matricial, pode-se escrever :

$$\underline{\hat{f}} = (L^{-1}\Psi(L')^{-1})L'\Psi^{-1}(\underline{x} - \underline{\mu}) \quad (182)$$

Portanto,

$$\underline{\hat{f}} = (L'\Psi^{-1}L)^{-1}L'\Psi^{-1}(\underline{x} - \underline{\mu}) \quad (183)$$

que é a solução para o estimador do Método de Mínimos Quadrados Ponderados.

Tomando  $\hat{L}$ ,  $\hat{\psi}$  e  $\hat{\underline{\mu}} = \bar{x}$  para valores amostrais, o j-ésimo caso fica:

$$\hat{f}_j = (\hat{L}'\hat{\psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\psi}^{-1}(\underline{x}_j - \bar{x}) \quad (184)$$

Quando  $\hat{L}$  e  $\hat{\psi}$  são determinados pelo Método da Máxima Verossimilhança, estas estimativas devem satisfazer a condição de singularidade, isto é,  $\hat{L}'\hat{\psi}^{-1}\hat{L} = \hat{\Delta}$  deve ser uma matriz diagonal.

Sendo assim, para os carregamentos estimados pelo Método da Máxima Verossimilhança, pode-se montar o Modelo de Mínimos Quadrados Ponderados como:

$$\underline{\hat{f}}_j = (\hat{L}'\hat{\psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\psi}^{-1}(\underline{x}_j - \hat{\underline{\mu}}) = \hat{\Delta}^{-1}\hat{L}'\hat{\psi}^{-1}(\underline{x}_j - \bar{x}), \quad j = 1, 2, \dots, n \quad (185)$$

ou se a matriz de correlação for fatorada,

$$\underline{\hat{f}}_j = (\hat{L}'_z\hat{\psi}^{-1}_z\hat{L}_z)^{-1}\hat{L}'_z\hat{\psi}^{-1}_z\underline{z}_j = \hat{\Delta}^{-1}_z\hat{L}'_z\hat{\psi}^{-1}_z\underline{z}_j, \quad j = 1, 2, \dots, n \quad (186)$$

no qual  $\underline{z}_j = D^{-1/2}(\underline{x}_j - \bar{x})$  e  $\hat{\rho} = \hat{L}_z \hat{L}_z' + \hat{\psi}_z$ .

Os escores anteriores podem ser obtidos após uma rotação de fatores. Em vez de  $\hat{f}_j$  seria  $\hat{f}_j^*$ ,  $j = 1, 2, \dots, n$ .

Caso os carregamentos fatoriais fossem estimados pelo Método de Componentes Principais, para generalizar os escores fatoriais, é comum o uso de um procedimento de mínimos quadrados sem ponderação. Implicitamente assume-se que  $\psi_i$  é igual ou quase igual ao valor verdadeiro. Os escores fatoriais seriam, então, dados por:

$$\hat{f}_j = (\hat{L}'\hat{L})^{-1} \hat{L}'(\underline{x}_j - \bar{x}) \quad \text{ou, para dados padronizados} \quad \hat{f}_j = (\hat{L}_z' \hat{L}_z)^{-1} \hat{L}_z' \underline{z}_j .$$

Desde que,

$$\hat{L} = \left[ \sqrt{\hat{\lambda}_1} \cdot \hat{e}_1 \quad \sqrt{\hat{\lambda}_2} \cdot \hat{e}_2 \quad \dots \quad \sqrt{\hat{\lambda}_m} \cdot \hat{e}_m \right] \quad (187)$$

tem-se:

$$\hat{f}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{e}_1'(\underline{x}_j - \bar{x}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{e}_2'(\underline{x}_j - \bar{x}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{e}_m'(\underline{x}_j - \bar{x}) \end{bmatrix} \quad (188)$$

Para estes escores fatoriais,

$$\frac{1}{n} \sum_{j=1}^n \hat{f}_j = \underline{0} \quad (\text{média amostral})$$

e

$$\frac{1}{n-1} \sum_{j=1}^n \hat{f}_j \hat{f}_j' = I \quad (\text{covariância amostral})$$

### 2.3.7.2 Método da regressão

Considerando o Modelo Fatorial Ortogonal  $\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$ , neste método, novamente as matrizes de carregamentos  $L$  e de variâncias específicas  $\Psi$  serão tratadas

como conhecidas. Quando os fatores comuns  $\underline{F}$  e os fatores específicos (ou erros)  $\underline{\varepsilon}$  são conjuntamente normalmente distribuídos com as suposições do Modelo Fatorial Ortogonal, a combinação linear  $\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$  tem uma distribuição  $N_p(0, LL' + \Psi)$ . O resultado do Apêndice 4 pode comprovar isto. Além disso, a distribuição conjunta de  $(\underline{X} - \underline{\mu})$  e  $\underline{F}$  é  $N_{m+p}(0, \Sigma^*)$ , na qual

$$\Sigma^*_{(m+p) \times (m+p)} = \begin{bmatrix} \Sigma = LL' + \Psi & \vdots & L \\ \cdots \cdots \cdots & \vdots & \cdots \\ L' & \vdots & I \end{bmatrix} \quad (189)$$

$(pxp)$ 
 $(pxm)$ 
 $(mxp)$ 
 $(mxm)$

e  $\underline{0}$  é um vetor nulo de dimensão  $(m+p)$ . Usando o resultado do Apêndice 5, a distribuição condicional de  $\underline{F} | \underline{x}$  é multivariada normal com

$$E(\underline{F} | \underline{x}) = L' \Sigma^{-1} (\underline{x} - \underline{\mu}) = L'(LL' + \Psi)^{-1} (\underline{x} - \underline{\mu}) \quad (190)$$

e

$$\text{Cov}(\underline{F} | \underline{x}) = I - L' \Sigma^{-1} L = I - L'(LL' + \Psi)^{-1} L \quad (191)$$

As quantidades  $L'(LL' + \Psi)^{-1}$  da equação (190) são os coeficientes na regressão multivariada dos fatores nas variáveis. Estimativas destes coeficientes produzem escores fatoriais que são análogos às estimativas dos valores médios condicionados na análise de regressão multivariada. Conseqüentemente, dado qualquer vetor de observações  $\underline{x}_j$  e tomando as estimativas de máxima verossimilhança  $\hat{L}$  e  $\hat{\Psi}$  como verdadeiros valores, o  $j$ -ésimo escore fatorial é dado por

$$\hat{f}_j = \hat{L}' \hat{\Sigma}^{-1} (\underline{x}_j - \bar{\underline{x}}) = \hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1} (\underline{x}_j - \bar{\underline{x}}), \quad j = 1, 2, \dots, n \quad (192)$$

Nos Apêndices 6 e 7 encontram-se as provas dos resultados a seguir.

$$(I + \hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1} \hat{L} = I - (I + \hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \quad (193)$$

$$(\hat{L}\hat{L}' + \hat{\Psi})^{-1} = \hat{\Psi}^{-1} - \hat{\Psi}^{-1} L (I + \hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1} \quad (194)$$

Utilizando os resultados no cálculo de  $\hat{f}_j$  na equação (184) têm-se que:

$$\hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1} = \hat{L}' \hat{\Psi}^{-1} - \hat{L}' \hat{\Psi}^{-1} L (I + \hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1}$$

$$\hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1} = \hat{L}' \hat{\Psi}^{-1} [I - \hat{L}' \hat{\Psi}^{-1} L (I + \hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1}]$$

$$\hat{L}'(\hat{L}\hat{L}'+\hat{\psi})^{-1} = \hat{L}'\hat{\psi}^{-1}\{I - [I - (I + \hat{L}'\hat{\psi}^{-1}\hat{L})^{-1}]\}$$

$$\underset{(mxp)}{\hat{L}'}(\underset{(pxp)}{\hat{L}\hat{L}'} + \underset{(pxp)}{\hat{\psi}})^{-1} = (I + \underset{(mxp)}{\hat{L}'}\hat{\psi}^{-1}\hat{L})^{-1}\underset{(mxp)}{\hat{L}'}\hat{\psi}^{-1} \quad (195)$$

Esta identidade ajuda a comparar os escores fatoriais em (184), gerados por regressão, com aqueles gerados pelo Método dos Mínimos Quadrados Ponderados. Assim, denotando  $\underline{\hat{f}}_j^R$  como o escore fatorial pelo Método da Regressão, e  $\underline{\hat{f}}_j^{LS}$  como o escore fatorial pelo Método dos Mínimos Quadrados Ponderados, por (187) tem-se:

$$\underline{\hat{f}}_j^{LS} = (\hat{L}'\hat{\psi}^{-1}\hat{L})^{-1}(I + \hat{L}'\hat{\psi}^{-1}\hat{L})\underline{\hat{f}}_j^R = (I + (\hat{L}'\hat{\psi}^{-1}\hat{L})^{-1})\underline{\hat{f}}_j^R \quad (196)$$

Para as estimativas de máxima verossimilhança,  $(\hat{L}'\hat{\psi}^{-1}\hat{L})^{-1} = \hat{\Delta}^{-1}$ , e se os elementos desta matriz diagonal estão próximos de zero, o Método da Regressão e o de Mínimos Quadrados Generalizados resultarão em escores fatoriais muito próximos.

Na tentativa de reduzir os possíveis efeitos da incorreta determinação do número de fatores, práticas tendem a calcular os escores fatoriais de (188) pelo uso de  $S$ , a matriz de covariância amostral original, em vez de  $\hat{\Sigma} = \hat{L}\hat{L}'+\hat{\psi}$ . Pode-se, assim, montar o Modelo para o cálculo dos escores fatoriais por regressão da seguinte maneira.

$$\underline{\hat{f}}_j = \hat{L}'S^{-1}(\underline{x}_j - \bar{x}), \quad j = 1, 2, \dots, n \quad (197)$$

ou, se a matriz de correlação é fatorada,

$$\underline{\hat{f}}_j = \hat{L}'_z R^{-1} \underline{z}_j, \quad j = 1, 2, \dots, n \quad (198)$$

onde

$$\underline{z}_j = D^{-1/2}(\underline{x}_j - \bar{x}) \quad \text{e} \quad \hat{\rho} = \hat{L}'_z \hat{L}'_z + \hat{\psi}_z$$

Caso os carregamentos rotacionados  $\hat{L}^* = \hat{L}T$  sejam usados no lugar dos carregamentos originais, os subseqüentes escores fatoriais  $\underline{\hat{f}}_j^*$  são relacionados a  $\underline{\hat{f}}_j$  por

$$\underline{\hat{f}}_j^* = T' \underline{\hat{f}}_j, \quad j = 1, 2, \dots, n \quad (199)$$

Uma medida numérica em concordância entre os escores fatoriais dos dois diferentes métodos de cálculo é produzida pelo coeficiente de correlação amostral entre

escores do mesmo fator. Dos métodos apresentados até aqui, nenhum deles é recomendado como uniformemente superior.

### 2.3.8 Considerações Adicionais

Segundo JOHNSON & WICHERN (1992, p.415), existem muitas decisões a serem tomadas num estudo analítico de fatores. A decisão mais importante é a escolha do número de fatores  $m$ . Em seguida vem a escolha do método e do tipo de rotação. Na prática, as análises fatoriais que apresentam um bom ajuste dos dados são aquelas nas quais as rotações são aplicadas a mais de um método, e todos os resultados substancialmente confirmam a mesma estrutura fatorial.

Pela complexidade com que se estabelece a estrutura fatorial considerando todos os parâmetros envolvidos, dispondo de diferentes métodos, analisando significativamente os resultados e, ainda, levando em conta a subjetividade das interpretações, pode-se adotar um procedimento estratégico para uma melhor análise da estrutura de dados.

A escolha da Análise Fatorial de Componentes Principais é particularmente apropriada como primeiro passo na análise dos dados. Principalmente porque não requer que a matriz de correlação  $R$  ou a matriz de covariância  $S$  sejam não singulares. Em seguida, examinar-se as observações mais importantes e cria-se gráficos a fim de relacionar os escores fatoriais. Além disso, calcular-se os escores padronizados fatoriais para cada observação, estabelecendo o relacionamento entre as distâncias quadradas. Optar por uma Rotação Varimax.

Sendo  $\underline{X}' = [x_1 \quad x_2 \quad \cdots \quad x_p]$  um vetor aleatório cujas componentes são variáveis aleatórias dicotômicas, a matriz de correlação mais recomendável é a composta por coeficientes de correlação tetracórico. Este coeficiente de correlação foi introduzido por Karl Pearson e é um estimador do Coeficiente Linear de Pearson onde se supõe a existência de uma variável latente subjacente àquela observada, onde em algumas situações só é possível medir a variável observável.

Existem duas restrições quanto a sua utilização. Uma delas consiste que as variáveis latentes sejam normalmente distribuídas e a outra se refere as variáveis observadas que devem ser dicotomizadas o mais próximo possível da mediana. Então,

sendo o Coeficiente de Correlação Tetracórico um estimador do Coeficiente de Correlação Linear de Pearson o qual considera as variáveis latentes  $x_{L1}$  e  $x_{L2}$  (ambas contínuas e normais), subjacentes às variáveis dicotômicas  $x_1$  e  $x_2$  efetivamente observadas.

A idéia de Karl Pearson foi medir a associação de variáveis contínuas normais bivariadas e transformá-las em tabela 2x2. A demonstração para a obtenção da equação tetracórica, parte da transformação da distribuição normal bivariada em variáveis dicotômicas.

A equação tetracórica de transformação é obtida após a dicotomização das variáveis  $x_1$  e  $x_2$  como mostra a Tabela 3.

**TABELA 3- ASSOCIAÇÃO DAS VARIÁVEIS OBSERVÁVEIS  $X_1$  e  $X_2$**

Variáveis Observáveis		X <sub>1</sub>		Total
		0	1	
X <sub>2</sub>	0	a	B	a+b
	1	c	D	c+d
Total		a+c	b+d	n

FORTE: A AUTORA

onde:

$a, b, c, d$  são as frequências observadas;

$n = a + b + c + d$  (total de observações).

A equação tetracórica é dada por:

$$\frac{ad-bc}{yy'n^2} = \hat{\rho}_1 + \hat{\rho}_1^2 \frac{zz'}{2} + \hat{\rho}_1^3 \frac{(z^2-1)(z'^2-1)}{6} + \hat{\rho}_1^4 \frac{z(z^2-3)(z'^2-3)}{24} + \hat{\rho}_1^5 \frac{(z^4-6z^2+3)(z'^4-6z'^2+3)}{120} + \hat{\rho}_1^6 \frac{z(z^4-10z^2+15)z'(z'^4-10z'^2+15)}{720} + \hat{\rho}_1^7 \frac{(z^6-15z^4+45z^2-15)(z'^6-15z'^4+45z'^2+15)}{5040} + \dots \quad (200)$$

sendo que:

$$p = \frac{(a+b)}{n} \text{ e } q = \frac{(c+d)}{n} = 1-p ;$$

$$p' = \frac{(a+b)}{n} \text{ e } q' = \frac{(c+d)}{n} = 1 - p' ;$$

$z$  é o escore normal padrão correspondente à área menor ou igual a  $p$ . Por exemplo, se  $p = 0,50$ , então se tem que  $z = 0$  (tabela de áreas sob a curva normal);

$z'$  é o escore normal padrão correspondente à área menor ou igual a  $p'$ . Se  $p' = 0,50$ , então se tem que  $z' = 0$ ;

$y$  é o valor da ordenada no ponto  $p$  e pode ser obtida fazendo-se

$$y = f(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} . \text{ Para o exemplo citado, se } z = 0, \text{ então}$$

$$y = f(0) = \frac{e^{-0}}{\sqrt{2\pi}} = 0,39894 \text{ (tabela de ordenadas da curva normal);}$$

$y'$  é o valor da ordenada no ponto  $p'$  e pode ser obtida fazendo-se

$$y' = f(z') = \frac{e^{-\frac{z'^2}{2}}}{\sqrt{2\pi}} .$$

Uma solução aproximada do cálculo do Coeficiente de Correlação Tetracórico é ignorar-se os termos de grau superior a 2:

$$\frac{ad - bc}{yy'n^2} = \hat{\rho}_t + \hat{\rho}_t^2 \frac{zz'}{2} \tag{201}$$

onde:  $\hat{\rho}_t$  é o estimador do Coeficiente de Correlação Tetracórico;

$a, b, c, d$  são as frequências da tabela 2x2;

$z$  é o escore normal padrão correspondente à área menor ou igual a  $p$ ;

$z'$  é o escore normal padrão correspondente à área menor ou igual a  $p'$ ;

$y$  é o valor da ordenada no ponto  $p$ ;

$y'$  é o valor da ordenada no ponto  $p'$ ;

$n = (a + b + c + d)$  é o número de observações da amostra.

Chamando o primeiro termo da expressão de  $c$ ; o coeficiente de  $\hat{\rho}_t$  de  $b$ ; e  $\frac{zz'}{2}$  de  $a$ , tem-se uma equação do 2.º grau:

$$a\hat{\rho}_t^2 + b\hat{\rho}_t + c = 0$$

$$\text{que poderá ser resolvida através de: } \hat{\rho}_t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Conforme demonstrado em WONNACOTT e WONNACOTT (1978, p.424), existe uma relação entre o Coeficiente de Correlação  $\hat{\rho}$  e o  $\cos\theta$ ,  $\hat{\rho} = \cos\theta$  e  $-1 \leq \cos\theta \leq +1$ . Outra expressão é apresentada por GUILFORD (1950, p.605), utilizando o cosseno, dada por:

$$\hat{\rho}_t = \cos\left(\frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right) \quad (202)$$

Quando o produto  $bc$  é igual a  $ad$ , o ângulo é  $90^\circ$  e o cosseno é igual a zero, conseqüentemente  $\hat{\rho}_t = 0$ .

O erro padrão aproximado do estimador do Coeficiente de Correlação Tetracórico é dado por:

$$\hat{\sigma}_{\hat{\rho}_t} = \frac{\sqrt{p \times q \times p' \times q'}}{y' \times y \times \sqrt{n}} \sqrt{(1 - \hat{\rho}_t^2) \times \left[1 - \left(\frac{\text{sen}^{-1} \hat{\rho}_t}{90^\circ}\right)^2\right]} \quad (203)$$

onde:  $\hat{\sigma}_{\hat{\rho}_t}$  é o erro padrão do estimador;

$\hat{\rho}_t$  é o estimador do Coeficiente de Correlação Tetracórico;

$\text{sen}^{-1} \hat{\rho}_t$  é o arco seno de  $\hat{\rho}_t$ ;

$n = (a + b + c + d)$  é o número de observações da amostra.

Para testar a hipótese de que  $\rho_t = 0$ , é possível utilizar a estatística  $t = \frac{\hat{\rho}_t}{\hat{\sigma}_{\hat{\rho}_t}}$ . O erro padrão poderá ser calculado considerando apenas a primeira parte da expressão:

$$\hat{\sigma}_{\hat{\rho}_t} = \frac{\sqrt{p \times q \times p' \times q'}}{y' \times y \times \sqrt{n}} \quad (204)$$

O Coeficiente de Correlação Tetracórico é menos confiável que o de Pearson, sendo que sua variabilidade é cerca de 50% maior, quando  $\rho = 0$ . Para obter a mesma confiabilidade<sup>1</sup> para o Coeficiente de Correlação Tetracórico que a obtida no Coeficiente de Correlação de Pearson, é necessário o dobro do tamanho da amostra.

Para obtenção dos escores por Análise Fatorial com Componentes Principais utilizando a matriz de correlação tetracórico ( $R_{tetra}$ ) é necessário que a amostra seja grande, e os escores fatoriais são dados por:

$$\underline{\hat{f}}_{jtetra} = \hat{L}'_z R_{tetra}^{-1} \underline{z}_j, \quad j = 1, 2, \dots, n \quad (205)$$

Caso os carregamentos rotacionados  $\hat{L}^* = \hat{L}T$  sejam usados no lugar dos carregamentos originais, os subseqüentes escores fatoriais  $\underline{\hat{f}}_{jtetra}^*$  são relacionados a

$\underline{\hat{f}}_{jtetra}$  por

$$\underline{\hat{f}}_{jtetra}^* = T' \underline{\hat{f}}_{jtetra}, \quad j = 1, 2, \dots, n \quad (206)$$

Utiliza-se os escores fatoriais na obtenção do chamado escore final com intuito de fazer o ranqueamento dos vetores dos indivíduos. Assim, tomando-se a média dos escores fatoriais ponderada pelos seus autovalores e têm-se o escore final para cada indivíduo. Costuma-se colocar estes escores finais numa escala de zero a cem para uma melhor análise dos resultados.

## 2.4 RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO

### 2.4.1 Introdução

Reconhecimento de Padrões e Classificação são técnicas ou procedimentos que visam reconhecer e classificar objetos ou indivíduos em grupos com base em medidas observadas dos mesmos. As técnicas de Reconhecimento de Padrões, segundo CHAVES NETO (2003, p.4), representam um importante componente dos sistemas artificialmente inteligentes e são utilizadas para pré-processamento dos dados e tomada de decisão. Assim, o objetivo fundamental do Reconhecimento de Padrões e Classificação é num primeiro estágio categorizar este objeto (indivíduo) segundo suas características. A característica é definida como um vetor de medidas do padrão a ser classificado. Por exemplo, quando se deseja distinguir a letra **A** da letra **E** é possível comparar o número de traços inclinados, horizontais e verticais dos caracteres.

Historicamente as principais abordagens de Reconhecimento de Padrões são a estatística e a sintática (estrutural). Depois, surgiram a tecnologia de Redes Neurais e também métodos de Programação Matemática.

O Reconhecimento de Padrões e Classificação, em estatística, baseia-se na atribuição de classes para indivíduos através do processo denominado de Análise Discriminante. Segundo JOHNSON & WICHERN, 1992, p.470, Análise Discriminante são técnicas multivariadas interessadas com a separação de uma coleção de indivíduos distintos e que alocam novos indivíduos em grupos previamente definidos.

Discriminações podem ser feitas através de processos com padrões conhecidos (também chamados de supervisionados) e não conhecidos (não supervisionados). Nos processos com padrões conhecidos são tomadas amostras de cada padrão e se definem critérios ou funções discriminantes para poder diferenciá-los. Nos processos que não possuem um padrão conhecido são feitos procedimentos que agrupam os dados em padrões que tenham características semelhantes (similaridades), chamadas técnicas de agrupamento (Cluster). Os principais objetivos da Análise Discriminante, descritos por

---

<sup>1</sup>A confiabilidade, aqui, é usada como sinônimo de erro padrão.

HAIR, ANDERSON, TATHAM & BLACK (2005, p.32-33), que podem ajudar a resolver problemas em pesquisa são:

- Estabelecer a existência de diferenças estatísticas significantes entre os escores médios de dois ou mais grupos de indivíduos previamente definidos.
- Indicar quais são as variáveis independentes que explicam o máximo de diferenças entre os grupos de modo que os grupos possam ser separados tanto quanto for possível.
- Definir procedimentos para classificar novos indivíduos em grupos que melhor se assemelham, com base nos seus escores em um conjunto de variáveis independentes.
- Estabelecer o número e a composição das dimensões de discriminação entre os grupos formados a partir do conjunto de variáveis independentes.

As técnicas estatísticas de reconhecimento de padrões e classificação têm como base à estimação de funções obtidas de observações amostrais a partir de algumas suposições (independência entre variáveis, linearidade de relações, igualdade de variâncias, etc.) que resultam em regras que possibilitem a separação e alocação de objetos (indivíduos ou empresas utilizados no contexto desta tese).

Na prática as duas suposições citadas, freqüentemente, sobrepõem-se e a distinção entre separação e alocação torna-se confusa. Uma função que separa, pode servir como alocadora (classificadora), e da mesma forma uma regra alocadora, pode sugerir um procedimento discriminatório. A terminologia de “discriminar” e “classificar” foi introduzida por Fischer em 1936 (S. E. FIENBERG apud JOHNSON & WICHERN, 1992 p. 471) no primeiro tratamento moderno dos problemas de separação.

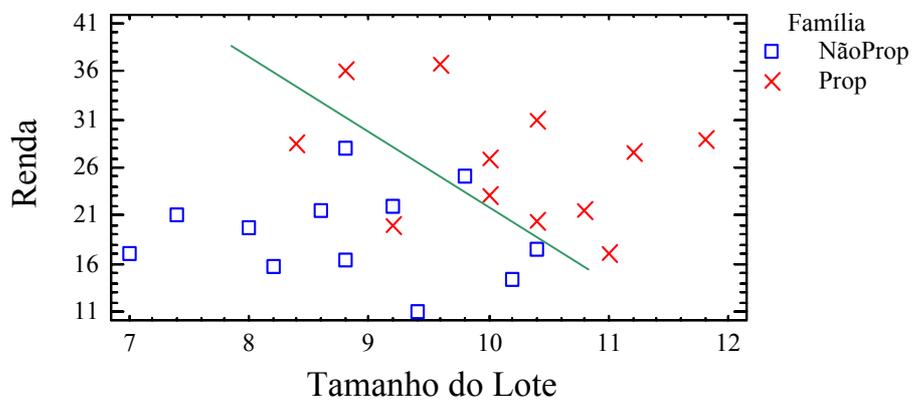
#### 2.4.2 Conceitos Gerais de Reconhecimento e Classificação

Com intuito de ilustrar e explicar os conceitos, JOHNSON & WICHERN (1992, p.481-482), considerou dois grupos em uma cidade, famílias proprietárias de certo equipamento e famílias não-proprietárias desse equipamento. A fim de identificar o melhor tipo de campanha de vendas, o fabricante do equipamento está interessado em classificar famílias como futuros compradores do equipamento ou não, com base em  $x_1$ = renda e  $x_2$  = tamanho do lote de moradia. Amostras aleatórias de  $n_1 = 12$  famílias proprietárias e  $n_2 = 12$  famílias não-proprietárias produziram os dados abaixo.

**TABELA 4- AMOSTRAS DE FAMÍLIAS PROPRIETÁRIAS E NÃO PROPRIETÁRIAS DE CERTO EQUIPAMENTO COM BASE NA RENDA E TAMANHO DO LOTE DE SUAS MORADIAS**

$\Pi_1 = \text{Proprietários}$		$\Pi_2 = \text{Não Proprietários}$	
Renda( $x_1$ )	Tamanho do Lote ( $x_2$ )	Renda( $x_1$ )	Tamanho do Lote ( $x_2$ )
20.0	9.2	25.0	9.8
28.5	8.4	17.6	10.4
21.6	10.8	21.6	8.6
20.5	10.4	14.4	10.2
29.0	11.8	28.0	8.8
36.7	9.6	16.4	8.8
36.0	8.8	19.8	8.0
27.6	11.2	22.0	9.2
23.0	10.0	15.8	8.2
31.0	10.4	11.0	9.4
17.0	11.0	17.0	7.0
27.0	10.0	21.0	7.4

FONTE: ADAPTADO DE JOHNSON & WICHERN

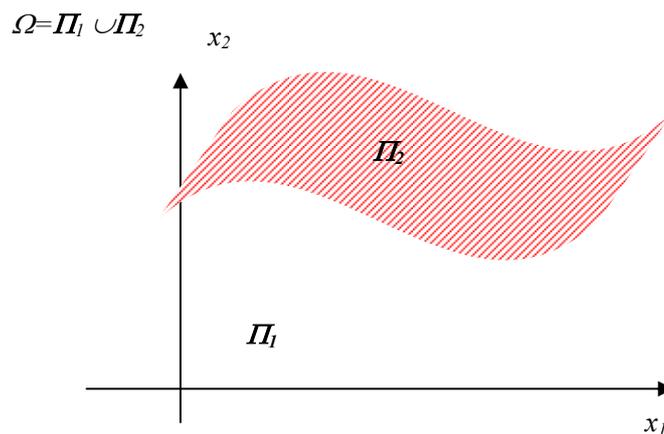


FONTE: ADAPTADO DE JOHNSON & WICHERN

**FIGURA 12- GRÁFICO DAS AMOSTRAS DE FAMÍLIAS PROPRIETÁRIAS E NÃO-PROPRIETÁRIAS COM BASE NA RENDA E TAMANHO DOS LOTES DE SUAS MORADIAS**

O gráfico da Figura 12 mostra que os proprietários tendem a ter maiores rendas e também maiores lotes, porém renda parece discriminar melhor que lote. Existe uma região onde os grupos se misturam que conseqüentemente irá gerar classificações erradas, assim pode-se criar uma regra (regiões  $R_1$  e  $R_2$ ) que minimize a chance de classificações erradas. Além disso, pode existir um grupo ou população com maior probabilidade de ocorrência do que de o outro. Uma regra de classificação ótima deve levar em conta estas duas possibilidades, isto é, as probabilidades de ocorrência “*a priori*” e o custo de classificação errada (classificar um indivíduo em  $\Pi_1$  quando na verdade ele pertencente a  $\Pi_2$  pode representar um erro mais sério do que classificar em  $\Pi_2$  quando o indivíduo pertencente a  $\Pi_1$ ).

Sejam as funções densidade de probabilidade  $f_1(\underline{x})$  e  $f_2(\underline{x})$  associadas ao vetor aleatório  $\underline{X}$  de dimensão  $p$  (número de variáveis analisadas) das populações  $\Pi_1$  e  $\Pi_2$ , respectivamente. Um indivíduo, com as medidas  $\underline{x}$ , deve ser reconhecido como de  $\Pi_1$  ou de  $\Pi_2$ . Seja  $\Omega$  o espaço amostral, isto é, o conjunto de todas as possíveis observações  $\underline{x}$ . Considerando  $R_1$  o conjunto de valores  $\underline{x}$  para os quais se classifica o indivíduo como  $\Pi_1$  e  $R_2 = \Omega - R_1$  os remanescentes valores  $\underline{x}$  para os quais se classifica os indivíduos como  $\Pi_2$ . Os conjuntos  $R_1$  e  $R_2$  são mutuamente exclusivos. Supondo  $p$  igual a 2, pode-se visualizar isto na Figura 13 que segue.



FONTE: A AUTORA

**FIGURA 13- REGIÕES DE CLASSIFICAÇÃO PARA DUAS POPULAÇÕES**

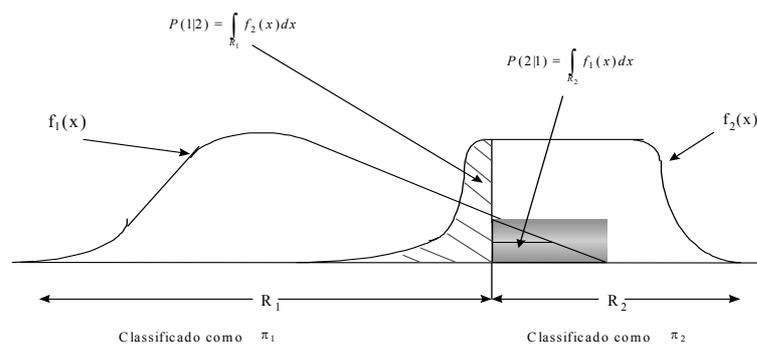
A probabilidade de classificar um indivíduo em  $\Pi_2$  quando na verdade ele é de  $\Pi_1$  é um probabilidade condicional dada por:

$$P(2|1) = P(\underline{X} \in R_2 | \Pi_1) = \int_{R_2=\Omega-R_1} f_1(\underline{x})d\underline{x} \quad (207)$$

Da mesma análoga a probabilidade de classificar um indivíduo em  $\Pi_1$  quando na verdade ele é de  $\Pi_2$  será:

$$P(1|2) = P(\underline{X} \in R_1 | \Pi_2) = \int_{R_1} f_2(\underline{x})d\underline{x} \quad (208)$$

Assim,  $P(2|1)$  representa o volume formado pela f.d.p.  $f_1(\underline{x})$  na região  $R_2$ . Se o caso fosse univariado, isto é,  $p = 1$  (uma variável sendo observada) poderia ser observado como mostra a Figura 14 a seguir.



FONTE: ADAPTADO DE JOHNSON & WICHERN

**FIGURA 14- CLASSIFICAÇÃO DAS REGIÕES PARA DUAS POPULAÇÕES**

#### 2.4.2.1 Custo esperado de erro de classificação (ECM)

Seja  $p_1$  a probabilidade “a priori” de  $\Pi_1$  e  $p_2$  a probabilidade “a priori” de  $\Pi_2$ , onde  $p_1 + p_2 = 1$ . As probabilidades de reconhecer corretamente ou incorretamente são dadas por:

$$P(\underline{X} \in \Pi_1 \text{ e é reconhecido corretamente como } \Pi_1) = P(\underline{X} \in R_1 | \Pi_1).P(\Pi_1) = P(1|1)p_1$$

$$P(\underline{X} \in \Pi_2 \text{ e é reconhecido incorretamente como } \Pi_1) = P(\underline{X} \in R_1 | \Pi_2).P(\Pi_2) = P(1|2)p_2$$

$$P(\underline{X} \in \Pi_2 \text{ e é reconhecido corretamente como } \Pi_2) = P(\underline{X} \in R_2 | \Pi_2).P(\Pi_2) = P(2|2)p_2$$

$P(\underline{X} \in \Pi_1 \text{ e é reconhecido incorretamente como } \Pi_2) = P(\underline{X} \in R_2 | \Pi_1).P(\Pi_1) = P(2|1)p_1$

As regras de classificação geralmente são avaliadas em termos destas probabilidades, mas ignora o custo de erro de classificação o que pode em alguns casos causar problemas. A Tabela 5 representa uma matriz de custo de reconhecimento.

**TABELA 5- MATRIZ DO CUSTO DE CLASSIFICAÇÃO**

Especificação		Classificado como:	
		$\Pi_1$	$\Pi_2$
População verdadeira	$\Pi_1$	0	$c(2 1)$
	$\Pi_2$	$c(1 2)$	0

FONTE: ADAPTADO DE JOHNSON & WICHERN

Um bom procedimento seria fazer com, a média, ou o custo esperado de erro de classificação *ECM* (*Expected Cost of Misclassification*), que é dado pela soma dos produtos dos elementos fora da diagonal principal pelas respectivas probabilidades, seja mínimo.

$$ECM = c(2|1)P(2|1)p(1) + c(1|2)P(1|2)p(2) \quad (209)$$

Resultado 1:

As regiões  $\Pi_1$  e  $\Pi_2$  que minimizam o *ECM* são definidas pelos valores de  $\underline{x}$  tal que valem as desigualdades:

$$\Pi_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left[ \frac{c(1|2)}{c(2|1)} \right] \cdot \left[ \frac{p_2}{p_1} \right] \quad (210)$$

$$\left[ \begin{array}{c} \text{Razão das} \\ \text{densidades} \end{array} \right] \geq \left[ \begin{array}{c} \text{Razão dos} \\ \text{custos} \end{array} \right] \cdot \left[ \begin{array}{c} \text{Razão das} \\ \text{probabilidades à priori} \end{array} \right]$$

$$\Pi_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < \left[ \frac{c(1|2)}{c(2|1)} \right] \cdot \left[ \frac{p_2}{p_1} \right] \quad (211)$$

$$\left[ \begin{array}{c} \text{Razão das} \\ \text{densidades} \end{array} \right] < \left[ \begin{array}{c} \text{Razão dos} \\ \text{custos} \end{array} \right] \cdot \left[ \begin{array}{c} \text{Razão das} \\ \text{probabilidades à priori} \end{array} \right]$$

Sendo,

$$ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

$$ECM = c(2|1)p_1 \int_{\Pi_2} f_1(\underline{x})d\underline{x} + c(1|2)p_2 \int_{\Pi_1} f_2(\underline{x})d\underline{x}$$

e como  $\Omega = \Pi_1 \cup \Pi_2$  tem-se:

$$I = \int_{\Omega} f_1(\underline{x})d\underline{x} = \int_{\Pi_1} f_1(\underline{x})d\underline{x} + \int_{\Pi_2} f_1(\underline{x})d\underline{x}$$

pode-se escrever:

$$ECM = c(2|1)p_1 \left[ I - \int_{\Pi_1} f_1(\underline{x})d\underline{x} \right] + c(1|2)p_2 \int_{\Pi_1} f_2(\underline{x})d\underline{x} \quad (212)$$

e das propriedades de integral (volume)

$$ECM = \int_{\Pi_1} [c(1|2)p_2 f_2(\underline{x}) - c(2|1)p_1 f_1(\underline{x})] d\underline{x} + c(2|1)p_1 I \quad (213)$$

e  $p_1$ ,  $p_2$ ,  $c(1|2)$ , e  $c(2|1)$  são não-negativos, e ainda  $f_1(\underline{x})$  e  $f_2(\underline{x})$  como também são não-negativos e são as únicas quantidades de  $ECM$  que dependem de  $\underline{x}$ . Assim  $ECM$  é minimizado se  $\Pi_1$  inclui esses valores  $\underline{x}$  tal que  $[c(1|2)p_2 f_2(\underline{x}) - c(2|1)p_1 f_1(\underline{x})] \leq 0$  e exclui aqueles para os quais esta quantidade é positiva. Isto é,  $\Pi_1$  deve ser o conjunto de pontos tal que:

$$c(1|2)p_2 f_2(\underline{x}) \leq c(2|1)p_1 f_1(\underline{x}) \Rightarrow \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left[ \frac{c(1|2)}{c(2|1)} \right] \left[ \frac{p_2}{p_1} \right] \quad (214)$$

e dado que  $\Pi_2$  é o complemento de  $\Pi_1$  em  $\Omega$ ,  $\Pi_2$  deve ser o conjunto de pontos  $\underline{x}$  para os quais:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} < \left[ \frac{c(1|2)}{c(2|1)} \right] \left[ \frac{p_2}{p_1} \right] \quad (215)$$

Casos Especiais de Regiões de  $ECM$ :

a) Probabilidades “a priori” iguais:  $\frac{p_2}{p_1} = 1$

$$\Pi_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{c(1|2)}{c(2|1)}; \quad \Pi_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{c(1|2)}{c(2|1)} \quad (216)$$

b) Custos de erro de reconhecimento iguais:  $\frac{c(1|2)}{c(2|1)} = 1$

$$\Pi_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{p_2}{p_1}; \quad \Pi_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{p_2}{p_1} \quad (217)$$

c) Probabilidades *a priori* iguais e custos de reconhecimento errado iguais:

$$\frac{p_2}{p_1} = \frac{c(1|2)}{c(2|1)} = 1 \text{ ou } \frac{p_2}{p_1} = \frac{1}{c(1|2)/c(2|1)}$$

$$\Pi_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq 1; \quad \Pi_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < 1 \quad (218)$$

Obs: 1) Quando as probabilidades *a priori* são desconhecidas, elas são freqüentemente tomadas como iguais (0,5) e a razão das funções densidade de probabilidade (f.d.p's) é comparada com a razão de custos de reconhecimento errado.

2) Se a razão de custo de reconhecimento errado é indeterminada, ela é usualmente tomada como 1 e a razão das f.d.p's é comparada com a razão de probabilidades *a priori*.

3) Finalmente, quando ambas: razões das probabilidades *a priori* e razões de custos são unitários ou uma razão é recíproca do outro, então as regiões de reconhecimento (classificação) ótimo são determinadas comparando-se os valores das f.d.p's.

Assim, se  $\underline{x}_0$  é uma nova observação e  $f_1(\underline{x}_0)/f_2(\underline{x}_0) \geq 1 \Rightarrow f_1(\underline{x}_0) \geq f_2(\underline{x}_0)$ , assumimos que  $\underline{x}_0 \in \Pi_1$ .

#### 2.4.2.2. Probabilidade total de erro de classificação

Outro critério, além do *ECM (Expected Cost of Misclassification)*, pode ser usado para construir procedimentos ótimos. Assim, pode-se ignorar o custo de classificação errada é escolher  $R_1$  e  $R_2$  que minimizam a probabilidade total de erro de classificação denominado de *TPM (Total Probability of Misclassification)*.

$$TPM = P(\underline{x} \in \Pi_1 \text{ e é classificada errada}) + P(\underline{x} \in \Pi_2 \text{ e é classificada errada})$$

$$TPM = p_1 \int_{R_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{R_1} f_2(\underline{x}) d\underline{x}$$

Matematicamente, isto é equivalente a minimizar *ECM* quando os custos de classificação errada são iguais. Assim, podemos alocar uma nova observação  $\underline{x}_0$  para a população com a maior probabilidade “*a posteriori*”  $P(\Pi_i | \underline{x}_0)$ , onde

$$\begin{aligned}
P(\Pi_1 | \underline{x}_0) &= \frac{P(\Pi_1 \text{ ocorre e observa-se } \underline{x}_0)}{P(\text{observa-se } \underline{x}_0)} = \\
&= \frac{P(\text{observa-se } \underline{x}_0 | \Pi_1)P(\Pi_1)}{P(\text{observa-se } \underline{x}_0 | \Pi_1)p(\Pi_1) + P(\text{observa-se } \underline{x}_0 | \Pi_2)p(\Pi_2)} = \\
&= \frac{p_1 f_1(\underline{x}_0)}{p_1 f_1(\underline{x}_0) + p_2 f_2(\underline{x}_0)} \quad (219)
\end{aligned}$$

$$e \quad P(\Pi_2 | \underline{x}_0) = 1 - P(\Pi_1 | \underline{x}_0) = \frac{p_2 f_2(\underline{x}_0)}{p_1 f_1(\underline{x}_0) + p_2 f_2(\underline{x}_0)} \quad (220)$$

A regra de classificação é dada por:

- classifica-se  $\underline{x}_0$  em  $\Pi_1$  quando  $P(\Pi_1 | \underline{x}_0) > P(\Pi_2 | \underline{x}_0)$
- classifica-se  $\underline{x}_0$  em  $\Pi_2$  quando  $P(\Pi_1 | \underline{x}_0) \leq P(\Pi_2 | \underline{x}_0)$ .

### 2.4.3 Reconhecimento e Classificação para Duas Populações

2.4.3.1 Função de classificação de Anderson – populações gaussianas e matrizes de correlação iguais.

Para o caso de duas populações gaussianas multivariadas assume-se que  $f_1(\underline{x})$  e  $f_2(\underline{x})$  são suas respectivas funções densidades multivariadas, onde a primeira com vetor de média  $\underline{\mu}_1$  e matriz de variância  $\underline{\Sigma}_1$  e a segunda com  $\underline{\mu}_2$  e  $\underline{\Sigma}_2$ , então supondo  $\underline{\Sigma}_1 = \underline{\Sigma}_2 = \underline{\Sigma}$  e a regra de classificação é feita com base no critério *ECM*.

Seja  $\underline{X}' = [X_1, X_2, \dots, X_p]$  para populações  $\Pi_1$  e  $\Pi_2$  e

$$f_i(\underline{x}) : \frac{1}{(2\pi)^{\frac{p}{2}} |\underline{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_i)' \underline{\Sigma}^{-1}(\underline{x} - \underline{\mu}_i)\right] \quad \text{para } i=1,2 \quad (221)$$

Suponha que os parâmetros  $\Pi_1$  e  $\Pi_2$  e  $\underline{\Sigma}$ , são conhecidos, tem-se as regiões de mínimo *ECM*.

$$\begin{aligned} \Pi_1: \frac{f_1(\underline{x})}{f_2(\underline{x})} &= \frac{\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma^{-1}(\underline{x} - \underline{\mu}_1)\right]}{\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma^{-1}(\underline{x} - \underline{\mu}_2)\right]} = \\ &= \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma^{-1}(\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma^{-1}(\underline{x} - \underline{\mu}_2)\right] \geq \left[\frac{c(12)}{c(21)}\right] \left[\frac{p_2}{p_1}\right] \end{aligned} \quad (222)$$

$$\Pi_2: \frac{f_1(\underline{x})}{f_2(\underline{x})} = \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma^{-1}(\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma^{-1}(\underline{x} - \underline{\mu}_2)\right] < \left[\frac{c(12)}{c(21)}\right] \left[\frac{p_2}{p_1}\right] \quad (223)$$

O procedimento de reconhecimento que minimiza *ECM* é obtido após logaritmar e fazer algumas simplificações na expressão acima e é dado por:

- reconhecer  $\underline{x}_0$  como sendo de  $\Pi_1$  se :

$$\left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \underline{x}_0 - \frac{1}{2} \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 + \underline{\mu}_2\right) \geq \ln \left[ \left( \frac{c(12)}{c(21)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (224)$$

ou como sendo de  $\Pi_2$  se:

$$\left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \underline{x}_0 - \frac{1}{2} \left(\underline{\mu}_1 - \underline{\mu}_2\right)' \Sigma^{-1} \left(\underline{\mu}_1 + \underline{\mu}_2\right) < \ln \left[ \left( \frac{c(12)}{c(21)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (225)$$

Em situações em que  $\underline{\mu}_i$ ,  $i=1,2$  são desconhecidas e  $\Sigma$  também, a regra deve ser modificada. A regra do *ECM* mínimo para duas populações normais (regra amostral) é dada por:

$$\text{Alocar } \underline{x}_0 \text{ em } \Pi_1 \text{ se } (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_p^{-1} \underline{x}_0 - \frac{1}{2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_p^{-1} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2) \geq \ln \left[ \frac{c(12)}{c(21)} \right] \left[ \frac{p_2}{p_1} \right]$$

$$\text{Alocar } \underline{x}_0 \text{ em } \Pi_2 \text{ se } (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_p^{-1} \underline{x}_0 - \frac{1}{2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_p^{-1} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2) < \ln \left[ \frac{c(12)}{c(21)} \right] \left[ \frac{p_2}{p_1} \right]$$

O primeiro termo da regra de classificação e reconhecimento,  $(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_p^{-1} \underline{x}$ , é a função linear obtida por Fisher que maximiza a variabilidade univariada entre as amostras relativamente à variabilidade dentro das amostras. A expressão inteira

$$w = (\bar{x}_1 - \bar{x}_2)' S_p^{-1} \underline{x} - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S_p^{-1} (\bar{x}_1 + \bar{x}_2) = (\bar{x}_1 - \bar{x}_2)' S_p^{-1} \left[ \underline{x} - \frac{1}{2} (\bar{x}_1 + \bar{x}_2) \right] \quad \text{é}$$

conhecida como Função de Classificação de Anderson.

2.4.3.2. Classificação quadrática – populações não necessariamente gaussianas e matrizes de correlação diferentes.

No caso de duas populações com variâncias diferentes ( $\Sigma_1 \neq \Sigma_2$ ), tem-se segundo JOHNSON & WICHERN (1992, p.492-493) a Classificação Quadrática.

Sejam as matrizes de covariância  $\Sigma_1$  para  $\underline{x} \in \Pi_1$  e  $\Sigma_2$  para  $\underline{x} \in \Pi_2$  onde  $\Sigma_1 \neq \Sigma_2$ , as regras de reconhecimento de padrões tornam-se mais complicadas.

Então para  $\underline{x} \sim N_p(\underline{\mu}_i, \Sigma_i)$ ,  $i=1,2$  tem-se que  $\underline{\mu}_1 \neq \underline{\mu}_2$  e  $\Sigma_1 \neq \Sigma_2$ . A probabilidade total de erro de classificação (TPM) e o custo esperado de erro de classificação dependem da razão das funções densidades  $\left[ \frac{f_1(\underline{x})}{f_2(\underline{x})} \right]$  ou, equivalentemente,

do logaritmo das razões das densidades  $\ln \left[ \frac{f_1(\underline{x})}{f_2(\underline{x})} \right] = \ln[f_1(\underline{x})] - \ln[f_2(\underline{x})]$

Sejam as populações  $\Pi_1$  e  $\Pi_2$  descritas por densidades normais multivariadas  $N_p(\underline{\mu}_1, \Sigma_1)$  e  $N_p(\underline{\mu}_2, \Sigma_2)$ . Então a regra de reconhecimento que minimiza o ECM é dada por:

$$\Pi_1 : -\frac{1}{2} \underline{x}'_0 (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x}_0 + \left( \underline{\mu}'_1 \Sigma_1^{-1} - \underline{\mu}'_2 \Sigma_2^{-1} \right) \underline{x}_0 - k \geq \ln \left[ \left( \frac{c(1|2)}{c(2|1)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (226)$$

$$\text{onde } k = \frac{1}{2} \ln \left( \frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} \left( \underline{\mu}'_1 \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}'_2 \Sigma_2^{-1} \underline{\mu}_2 \right)$$

$$\Pi_2 : -\frac{1}{2} \underline{x}'_0 (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x}_0 + \left( \underline{\mu}'_1 \Sigma_1^{-1} - \underline{\mu}'_2 \Sigma_2^{-1} \right) \underline{x}_0 - k < \ln \left[ \left( \frac{c(1|2)}{c(2|1)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (227)$$

Na prática, a regra de reconhecimento estabelecida é implementada substituindo-se os parâmetros  $\underline{\mu}_1, \underline{\mu}_2, \Sigma_1$  e  $\Sigma_2$  pelas suas estimativas  $\bar{x}_1, \bar{x}_2, S_1$  e  $S_2$ , tal que:

Alocamos  $\underline{x}_0$  em  $\Pi_1$  se:

$$-\frac{1}{2} \underline{x}'_0 (S_1^{-1} - S_2^{-1}) \underline{x}_0 + (\bar{\underline{x}}_1' S_1^{-1} - \bar{\underline{x}}_2' S_2^{-1}) \underline{x}_0 - k \geq \ln \left[ \left( \frac{c(1|2)}{c(2|1)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (228)$$

Alocamos  $\underline{x}_0$  em  $\Pi_2$  se:

$$-\frac{1}{2} \underline{x}'_0 (S_1^{-1} - S_2^{-1}) \underline{x}_0 + (\bar{\underline{x}}_1' S_1^{-1} - \bar{\underline{x}}_2' S_2^{-1}) \underline{x}_0 - k < \ln \left[ \left( \frac{c(1|2)}{c(2|1)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (229)$$

#### 2.4.3.3. Método discriminante linear de Fischer – populações não necessariamente gaussianas e matrizes de covariâncias iguais

Sejam  $p$  variáveis aleatórias associadas a um vetor  $E(\underline{X}/\Pi_2) = \underline{\mu}_2$ , onde as observações de valores de  $\underline{X}$  diferem de uma população  $\Pi_1$  para outra  $\Pi_2$ . Considerem-se os valores da primeira classe como sendo a população de  $\underline{X}$  valores para  $\Pi_1$  e para a segunda classe com população de  $\underline{X}$  valores para  $\Pi_2$ . Então, descrevem-se funções densidades de probabilidade  $f_1(\underline{x})$  e  $f_2(\underline{x})$  para as duas populações .

A idéia de Fisher, segundo JOHNSON & WICHERN (1992, p.471), era a transformação de observações multivariadas  $\underline{x}$  para observações univariadas  $Y$ , e que uma das funções  $Y$  derivadas das populações  $\Pi_1$  e  $\Pi_2$  sejam tão separáveis quanto possível.

Fisher sugeriu que os valores de  $y$  sejam combinações lineares, pois estas são funções de fácil manuseio matemático.

Definindo  $\mu_{1y}$  a média da combinação linear de  $\underline{x}$  pertencente a  $\Pi_1$  e  $\mu_{2y}$  a  $\Pi_2$ , Fisher selecionou as combinações lineares que maximizam a distância (quadrática) entre  $\mu_{1y}$  e  $\mu_{2y}$  relativa a variabilidade dos valores de  $\underline{Y}$ .

Seja  $Y$  a combinação linear de  $\underline{X}$  dado por  $Y = \underline{\ell}' \underline{X}$  onde  $\underline{\ell}'$  corresponde ao vetor de coeficientes.

Tem-se que:

$$\mu_{1y} = E(Y/\Pi_1) = E(\underline{\ell}' \underline{X}/\Pi_1) = \underline{\ell}' E(\underline{X}/\Pi_1) \quad (230)$$

$$\mu_{2y} = E(Y/\Pi_2) = E(\underline{\ell}' \underline{X}/\Pi_2) = \underline{\ell}' E(\underline{X}/\Pi_2) \quad (231)$$

fazendo:

$$E(\underline{X}/\Pi_1) = \underline{\mu}_1$$

e 
$$E(\underline{X}/\Pi_2) = \underline{\mu}_2$$

tem-se:

$$\underline{\mu}_{1y} = \underline{\ell}' \underline{\mu}_1 \quad (232)$$

$$\underline{\mu}_{2y} = \underline{\ell}' \underline{\mu}_2 \quad (233)$$

A matriz de covariância considerada a mesma para as duas populações é dada por:

$$\sigma_y^2 = \text{Var}(\underline{\ell}' \underline{X}) = \underline{\ell}' \text{Cov}(\underline{X}) \underline{\ell} = \underline{\ell}' \underline{\Sigma} \underline{\ell} \quad (234)$$

onde

$$\underline{\Sigma} = E(\underline{X} - \underline{\mu}_i)(\underline{X} - \underline{\mu}_i)' \quad , \quad i = 1, 2. \quad (235)$$

A melhor combinação linear é obtida da razão entre o quadrado da distância entre as médias e a variância de  $\underline{Y}$  assim:

$$\frac{(\underline{\mu}_{1Y} - \underline{\mu}_{2Y})^2}{\sigma_y^2} = \frac{(\underline{\ell}' \underline{\mu}_1 - \underline{\ell}' \underline{\mu}_2)^2}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} = \frac{\underline{\ell}' (\underline{\mu}_1 - \underline{\mu}_2) (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\ell}}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} = \frac{(\underline{\ell}' \underline{\delta})^2}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} \quad (236)$$

onde  $\underline{\delta}$  é o vetor de diferença dos valores médios, isto é:

$$\underline{\delta} = (\underline{\mu}_1 - \underline{\mu}_2)$$

Note que a matriz  $\underline{\delta} \underline{\delta}' = (\underline{\mu}_1 - \underline{\mu}_2) (\underline{\mu}_1 - \underline{\mu}_2)'$  contém os quadrados dos produtos cruzados dos diferentes componentes entre as médias das populações  $\Pi_1$  e  $\Pi_2$ .

Os coeficientes da combinação linear de Fisher  $\underline{\ell}' = [\ell_1, \ell_2, \dots, \ell_p]$  são aqueles que maximizam a razão  $\frac{(\underline{\ell}' \underline{\delta})^2}{\underline{\ell}' \underline{\Sigma} \underline{\ell}}$  :

$$\underline{\ell} = \underline{c} \underline{\Sigma}^{-1} \underline{\delta} = \underline{c} \underline{\Sigma}^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \quad \text{para} \quad \underline{c} \neq \underline{0} \quad (237)$$

Escolhendo  $\underline{c} = \underline{\ell}$  obtem-se:

$$\underline{Y} = \underline{\ell}' \underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} \underline{X} \quad (238)$$

que é conhecida como Função Discriminante Linear de Fisher e o máximo da razão é obtido por:

$$MAX_{\underline{\ell}} \frac{(\underline{\ell}' \underline{\delta})^2}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} = \underline{\delta}' \underline{\Sigma}^{-1} \underline{\delta} \quad (239)$$

Sendo a matriz de covariância uma matriz positiva definida, a prova da maximização da razão é feita aplicando a Lemma da Maximização, baseado na desigualdade de Cauchy-Schwarz (JOHNSON & WICHERN, 1992, p.473).

A Função Linear Discriminante converte as populações  $\Pi_1$  e  $\Pi_2$  que são multivariadas em populações univariadas de maneira que as médias das populações univariadas correspondentes são separadas tanto quanto possível.

A equação  $Y = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} \underline{X}$  também pode ser empregada como um dispositivo de classificação (classificador). Para alocar uma nova observação  $\underline{X}_0$ , tem-se  $Y_0 = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} \underline{X}_0$  como o valor da função discriminante e ainda:

$$m = 1/2(\mu_{1Y} + \mu_{2Y}) = 1/2(\ell' \mu_1 + \ell' \mu_2)$$

ou

$$m = 1/2(\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} (\underline{\mu}_1 + \underline{\mu}_2) \quad (240)$$

que é o ponto médio entre as duas médias populacionais univariadas. Assim pode-se mostrar que:

$$E(Y_0/\Pi_1) - m \geq 0 \quad \text{e} \quad E(Y_0/\Pi_2) - m < 0 \quad (241)$$

Desta maneira se  $\underline{X}_0$  for oriundo de  $\Pi_1$ , a esperança de  $Y_0$  é maior ou igual ao ponto médio. Caso  $\underline{X}_0$  for oriundo de  $\Pi_2$ , a esperança de  $Y_0$  é menor que o ponto médio. Então, a regra de classificação é dada por:

$$\text{alocar } \underline{X}_0 \text{ em } \Pi_1 \text{ se } Y_0 = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{X}_0 \geq m$$

ou

$$\text{alocar } \underline{X}_0 \text{ em } \Pi_2 \text{ se } Y_0 = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{X}_0 < m$$

Sabendo-se que as quantidades populacionais são raramente conhecidas, a regra acima não pode ser implementada a não ser que  $\underline{\ell}$  e  $m$  possam ser estimados por observações que já tenham sido corretamente classificadas.

Suponha que se tenha  $n_1$  observações da variável aleatória multivariada  $\underline{X}_1$  de  $\Pi_1$  e  $n_2$  observações de  $\Pi_2$ . Os vetores de dados respectivos a  $\Pi_1$  e  $\Pi_2$  são:

$$\underline{X}_1' = [X_{11}, X_{12}, \dots, X_{1n_1}]$$

$$\underline{X}_2' = [X_{21}, X_{22}, \dots, X_{2n_2}]$$

Para estes dados, os vetores amostrais de médias e as matrizes amostrais de covariância são dadas por:

$$\bar{x}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} x_{1j}$$

$$S_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)'$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j}$$

$$S_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)(x_{2j} - \bar{x}_2)'$$

Como foi assumido que as populações têm a mesma matriz de covariância  $\Sigma$ , as matrizes de covariância  $S_1$  e  $S_2$  podem então ser combinadas para se obter uma única estimativa não tendenciosa de  $\Sigma$ , dada pela matriz de covariância conjunta:

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)}$$

Substituindo as quantidades  $\underline{\mu}_1$ ,  $\underline{\mu}_2$  e  $\Sigma$  por respectivamente,  $\bar{x}_1$ ,  $\bar{x}_2$  e  $S_c$ , tem-se que a Função Discriminante Linear Amostral de Fisher para duas populações é dada por:

$$y = \hat{\ell}' \underline{x} = (\bar{x}_1 - \bar{x}_2)' S_c^{-1} \underline{x} \quad (242)$$

A estimativa do ponto médio entre duas médias amostrais univariadas,  $\bar{y}_1 = \hat{\ell}' \bar{x}_1$  e  $\bar{y}_2 = \hat{\ell}' \bar{x}_2$  é dada por:

$$\hat{m} = \frac{1}{2}(\bar{y}_1 - \bar{y}_2) = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S_c^{-1} (\bar{x}_1 + \bar{x}_2) \quad (243)$$

E a regra de classificação mostra que se deve:

- alocar  $\underline{x}_0$  em  $\Pi_1$  se:

$$y_0 = (\bar{x}_1 - \bar{x}_2)' S_c^{-1} \underline{x}_0 \geq \hat{m} \quad \text{ou} \quad y_0 - \hat{m} \geq 0$$

- alocar  $\underline{x}_0$  em  $\Pi_2$  se:

$$y_0 - \hat{m} < 0$$

2.4.3.4. Regressão logística: modelo para variáveis dicotômicas.

2.4.3.4.1. Introdução

A regressão logística, consiste em relacionar, através de um modelo, uma variável resposta  $Y$ , dicotômica, com os fatores  $(X_1, X_2, \dots, X_{p-1})$  que influenciam as ocorrências de determinado evento. Por exemplo, em um estudo para se quantificar a influência de certos fatores na ocorrência de doenças do fígado (colestase), a variável resposta  $Y$  será dicotômica, isto é, presença de câncer ( $Y=1$ ) ou presença de cálculo ( $Y=0$ ) e os fatores serão, por exemplo,  $x_1 =$  bilirrubina total,  $x_2 =$  fosfatase alcalina, etc. Os fatores são as covariáveis do modelo e correspondem aos níveis quantitativos obtidos nos exames bioquímicos, etc. (CHAVES NETO, 2004 p.61-64).

Quando a variável resposta é dicotômica ( $Y = 1$  ou  $0$ ) o Modelo Linear Geral não deve ser aplicado principalmente por duas razões:

1. produzirá valores fora do intervalo  $[0,1]$
2. as variâncias das observações não serão constantes

2.4.3.4.2. Modelo logístico linear simples

Seja o modelo linear logístico simples (uma covariável) derivado da função matemática chamada de sigmóide

$$f(y) = \frac{1}{1 + e^{-y}}, \quad y \in \mathbb{R}. \quad (244)$$

que varia monotonicamente de 0 a 1 à medida que  $y$  cresce, sendo simétrica em torno de  $y = 1/2$ .

É claro que:

$$f(y) = \frac{1}{1 + e^{-y}} = \frac{e^y}{1 + e^y} \quad (245)$$

e ainda a transformação LOGIT  $f(y) = \ln\left[\frac{f(y)}{(1-f(y))}\right] = \ln\left[\frac{(1+e^{-y})^{-1}}{1-(1+e^{-y})^{-1}}\right]$

$$\begin{aligned} \text{LOGIT } f(y) &= \ln\left[\frac{1}{1+e^{-y}} / \left(1 - \frac{1}{1+e^{-y}}\right)\right] = \ln\left[\frac{1}{1+e^{-y}} / \frac{1+e^{-y}-1}{1+e^{-y}}\right] \\ &= \ln\left[\frac{1}{1+e^{-y}} / \frac{e^{-y}}{1+e^{-y}}\right] = -\ln(1+e^{-y}) - (-y) - (-\ln(1+e^{-y})) \\ &= -\ln(1+e^{-y}) + y + \ln(1+e^{-y}) = y \end{aligned} \quad (246)$$

Então, impondo um Modelo de regressão logístico linear para estimar  $P(Y=1) = p(x)$  (tratando o caso linear simples, somente com uma variável explicativa), tem-se o modelo dado por:

LOGIT  $p(x) = \mu = \beta_0 + \beta_1 x$  que é o nosso  $y$ .

A aplicação desse modelo para  $x = 0$  resulta:

$$p(0) = P(Y = 1 | x = 0) = \frac{e^{\beta_0 + \beta_1 \cdot 0}}{1 + e^{\beta_0 + \beta_1 \cdot 0}} = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \text{ da forma } \frac{e^y}{1 + e^y}. \quad (247)$$

#### 2.4.3.4.3. Modelo logístico linear múltiplo

Quando o interesse está em se estabelecer a relação entre a variável resposta  $Y$  e as diversas covariáveis (variáveis explicativas)  $X_1, X_2, \dots, X_{p-1}$  que podem representar fatores de interesse, o modelo logístico linear múltiplo tem a forma:

$$\text{LOGIT } p(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} = \mu$$

ou

$$p(\underline{x}) = p(x_1, x_2, \dots, x_{p-1}) = e^{\mu} / (1 + e^{\mu}) = \frac{1}{(1 + e^{-\mu})} \quad (248)$$

onde  $\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1} = \underline{x}'\underline{\beta}$ .

O método de estimação de  $\underline{\beta}$  mais indicado, segundo MARQUES & LIMA (2002, p.55) é o estimador de máxima verossimilhança através do método de Levenberg-Marquart. Este método aplica-se nos extremos dos métodos de Newton e do gradiente, isto é, utiliza gradiente quando se está longe do mínimo e distante da solução caso

contrário usa o método de Newton. Combina os dois métodos através da alteração dos elementos da diagonal principal da matriz hessiana. O processo é iterativo e será finalizado quando dois critérios de parada são satisfeitos. O primeiro é obtido analisando-se as derivadas parciais da função de logverossimilhança, sendo que a soma dos valores observados deve ser igual à soma dos valores estimados. O segundo está relacionado com o fato de que uma nova solução não deve ser alterada no valor da função de logverossimilhança.

Estimados os parâmetros a regra de classificação é dada a seguir.

Seja duas populações  $\Pi_1$  e  $\Pi_2$ , a probabilidade de um indivíduo pertencer a  $\Pi_1$  é  $P(Y=0)=1-p(x)$  e a  $\Pi_2$  é  $P(Y=1)=p(x)$ . A alocação de um novo indivíduo poderá ser feita calculando  $\mu_0 = \underline{x}'_0 \hat{\beta}$  e determinando a função distribuição da logística, obtem-se probabilidade dada por

$$p(\underline{x}_0) = \frac{1}{(1 + e^{-u_0})} \quad (249)$$

que corresponde um valor entre 0 e 1. Portanto a regra de alocação é dada por:

- para  $p(\underline{x}_0) \leq 0,5$  alocar o indivíduo em  $\Pi_1 (Y=0)$
- para  $p(\underline{x}_0) > 0,5$  alocar o indivíduo em  $\Pi_2 (Y=1)$

#### 2.4.4. Reconhecimento e Classificação para Várias Populações.

##### 2.4.4.1. Introdução

Segundo CHAVES NETO (2003, p.43), a generalização de procedimentos estatísticos de reconhecimento de padrões classificação, na teoria, de 2 ou  $g \geq 2$  grupos é direta. Contudo, não se conhece muito sobre as propriedades das funções amostrais e em particular, suas taxas de erro não foram totalmente investigadas.

A robustez da função de reconhecimento de padrões linear para 2 grupos, por exemplo, em matrizes de covariâncias diferentes e distribuições não-normais pode ser estudada através de experimentos gerados por meio de computador. Para mais que duas populações esta abordagem não leva a conclusões gerais, porque as propriedades

dependem de onde os grupos estão localizados e existem muitas configurações para serem estudadas convenientemente.

#### 2.4.4.2 Método de máxima verossimilhança para várias populações – populações gaussianas e matrizes de covariâncias diferentes

O desenvolvimento do método deriva da função penalidade, que avalia as decisões incorretas (MÜLLER, 1997, p.14).

Seja  $\underline{x}$  um vetor de variáveis aleatórias pertencentes a uma classe  $\Pi_k$  (no caso cores);  $p(\underline{x}/\Pi_k)$  a probabilidade condicional de  $\underline{x}$  ocorrer nessa classe (ou população) e  $p(\Pi_k)$  a probabilidade "a priori" da classe.

Então,  $\lambda(i/k)$  é a função perda, função penalidade que designa a perda de alocação na classe  $\Pi_i$  quando de fato  $\lambda(i/k)$  pertence a  $\Pi_k$ . Sendo,  $\lambda(i/k)$  uma variável do tipo dicotômica, temos:

$$\lambda(i/k) = \begin{cases} 0 & , i = k \\ 1 & , i \neq k \end{cases} \quad i, k = 1, 2, \dots, m$$

Determinando-se a média de  $\lambda(i/k)$  e considerando todas as possibilidades de  $\Pi_k$ , obtém-se a perda média condicional apresentada abaixo para o caso multivariado, onde foi aplicado a esta variável aleatória discreta a definição de esperança matemática.

$$L_x(\Pi_i) = \sum_{\substack{k=1 \\ i \neq k}}^m \lambda(i/k) \cdot p(\Pi_k/\underline{x})$$

(250)

Esta função se refere a uma perda acumulada decorrente de valores observados de  $\underline{x}$  que venham de algumas das classes avaliadas, de maneira que se pode relacionar as funções de perda avaliadas de todas as classes para a classe  $\Pi_i$ . Então, a regra de decisão que minimiza a perda média condicional é chamada de "Classificador Ótimo", "Classificador Bayesiano" (MOIKE, 1980) ou "Otimização de Bayes" (RICHARDS; 1986) autores citados em MÜLLER (1997, p.15) dada por:

$$\begin{aligned} \underline{x} \in \Pi_i \quad \text{se } L(\underline{x}, \Pi_i) < L(\underline{x}, \Pi_k) \quad \forall \quad k = 1, 2, \dots, m \\ i = 1, 2, \dots, m \\ k \neq i \end{aligned} \quad (251)$$

Geralmente,  $p(\Pi_k/\underline{x})$  é dita probabilidade "a posteriori" e pode-se determiná-la em termos das funções distribuição de probabilidade das classes  $p(\underline{x}/\Pi_k)$ , utilizando o Teorema de Bayes como segue:

$$p(\Pi_k/\underline{x}) = \frac{p(\underline{x}/\Pi_k) \cdot p[\Pi_k]}{p(\underline{x})} \quad (252)$$

Aplicando o teorema acima na equação da perda média condicional tem-se:

$$L_x(\Pi_i) = \frac{1}{p(\underline{x})} \sum_{k=1}^m \lambda(i/k) \cdot p(\underline{x}/\Pi_k) \cdot p(\Pi_k) \quad (253)$$

fazendo:

$$\ell_x(\Pi_i) = \sum_{k=1}^m \lambda(i/k) \cdot p(\underline{x}/\Pi_k) \cdot p(\Pi_k) \quad (254)$$

fica-se com a expressão:

$$L_x(\Pi_i) = \frac{1}{p(\underline{x})} \cdot \ell_x(\Pi_i) \quad (255)$$

Dado que  $p(\underline{x})$  é comum para todas as classes, a decisão pode ser tomada somente com base em  $\ell_x(\Pi_i)$ .

Para este caso pode-se definir a função de perda como  $\lambda(i/k) = 1 - \theta_{ik}$  com  $\theta_{ik} = 1$  e  $\theta_{ik} = 0$  com  $i \neq k$ . Assim pode-se expressar:

$$\ell_x(\Pi_i) = \sum_{k=1}^m p(\underline{x}/\Pi_k) \cdot p(\Pi_k) - \sum_{k=1}^m \theta_{ik} \cdot p(\underline{x}/\Pi_k) \cdot p(\Pi_k) \quad (256)$$

sabendo ser :  $p(\underline{x}) = \sum_{k=1}^m p(\underline{x}/\Pi_k) \cdot p(\Pi_k)$ , pelo teorema da probabilidade total e fazendo:

$g_i(\underline{x}) = \sum_{k=1}^m \theta_{ik} \cdot p(\underline{x}/\Pi_k) \cdot p(\Pi_k)$  tem-se que:

$$\ell_x(\Pi_i) = p(\underline{x}) - g_i(\underline{x})$$

Novamente, pode-se retirar a probabilidade  $p(\underline{x})$  comum para todas as classes não afetando na classificação, fica-se com a igualdade:

$$\ell_x(\Pi_i) = -g_i(\underline{x})$$

Sabendo-se que  $\ell_x(\Pi_i)$  procede da expressão  $L_x(\Pi_i)$  e que representa a perda média condicional, assim o sinal negativo da expressão acima nos indica que quanto maior for  $g_i(\underline{x})$  menor será a perda, então:

$$\underline{x} \in \Pi_i \text{ se } g_i(\underline{x}) > g_j(\underline{x}) \quad \forall j \neq i \quad (257)$$

É conveniente, segundo MÜLLER (1997, p. 16), fazer  $\theta_{ik} = \delta_{ik}$  que é a função delta de Kroneker definida por:

$$\delta_{ik} = \begin{cases} 1 & \text{para } i = k \\ 0 & \text{para } i \neq k \end{cases} \quad (258)$$

Nestas condições a expressão de  $g_i(\underline{x})$  fica:

$$g_i(\underline{x}) = p(\underline{x}/\Pi_i).p(\Pi_i) \quad (259)$$

e a regra de decisão é:

$$\underline{x} \in \Pi_i \text{ se } p(\underline{x}/\Pi_i).p(\Pi_i) > p(\underline{x}/\Pi_j).p(\Pi_j) \quad \forall j \neq i \quad (260)$$

Esta regra é mais aceitável desde que  $p(\underline{x}/\Pi_k)$  sejam conhecidos através do treinamento dos dados.

Assumindo que as distribuições de probabilidade de cada classe são da forma gaussiana multivariada, tem-se que:

$$p(\underline{x}/\Pi_i) = (2\pi)^{(-n/2)} |\Sigma_i|^{-1/2} \cdot \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i)\right\} \quad (261)$$

Onde  $\underline{\mu}_i$  e  $\Sigma_i$  são respectivamente o vetor médio e a matriz de covariância dos dados da classe  $\Pi_i$  que geralmente são desconhecidos, então a partir de uma amostra de tamanho  $n_i$  da população  $\Pi_i$  pode-se obter as respectivas estimativas :

$$\bar{\underline{x}}_i = \frac{1}{n_i} \sum_{r=1}^{n_i} x_{ir} \quad (262)$$

$$S_i = \sum_{i=1}^m \sum_{r=1}^{n_i} (x_{ir} - \bar{\underline{x}}_i)(x_{ir} - \bar{\underline{x}}_i)' \quad (263)$$

Ainda, sabendo ser o resultado  $-n/2 \ln(2\Pi)$  comum para todos  $g_i(\underline{x})$ , tem-se a expressão:

$$g_i(\underline{x}) = \ln p(\Pi_i) - \frac{1}{2} \ln |S_i| - \frac{1}{2} (\underline{x} - \bar{\underline{x}}_i)' S_i^{-1} (\underline{x} - \bar{\underline{x}}_i) \quad (264)$$

Como  $p(\Pi_i)$  é suposta desconhecida, assume-se que todas as classes tenham igual probabilidade de ocorrer; pode-se então retirar esse fator, o que resultará em mais um fator comum a ser eliminado ( $1/2$ ), pois estes não afetam na decisão para a classificação. Assim a função discriminante amostral de máxima verossimilhança é dada por:

$$g_i(\underline{x}) = -\ln |S_i| - (\underline{x} - \bar{\underline{x}}_i)' S_i^{-1} (\underline{x} - \bar{\underline{x}}_i) \quad (265)$$

E a regra de decisão fica:

$$\underline{x} \in \Pi_i \quad \text{se} \quad -\ln |S_i| - (\underline{x} - \bar{\underline{x}}_i)' S_i^{-1} (\underline{x} - \bar{\underline{x}}_i) > -\ln |S_j| - (\underline{x} - \bar{\underline{x}}_j)' S_j^{-1} (\underline{x} - \bar{\underline{x}}_j) \quad (266)$$

#### 2.4.4.3 Método linear de Fisher para várias populações – populações não necessariamente gaussianas e matrizes de covariâncias iguais

O método discriminante de Fisher para várias populações é uma extensão do mesmo método para duas populações. Portanto, esta regra pode ser estendida para várias populações que não necessariamente tenham a mesma distribuição de probabilidade, entretanto necessita-se da suposição de que as matrizes das diversas populações sejam iguais; isto é,  $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$ .

Seja  $\bar{\underline{\mu}}$  o vetor médio combinado dos grupos e  $B_0$  a soma dos produtos cruzados "entre" os grupos populacionais, tais que:

$$\bar{\underline{\mu}} = \frac{1}{g} \sum_{i=1}^g \underline{\mu}_i \quad (267)$$

$$B_0 = \sum_{i=1}^g (\underline{\mu}_i - \bar{\underline{\mu}})(\underline{\mu}_i - \bar{\underline{\mu}})' \quad (268)$$

onde se considera a combinação linear  $Y = \underline{\ell}' \underline{X}$  e obtém-se o valor esperado que se segue:

$$E(Y) = \underline{\ell}' E(\underline{X} / \Pi_i) = \underline{\ell}' \underline{\mu}_i \quad (269)$$

para a população  $\Pi_i$  e a variância:

$$Var(Y) = \underline{\ell}' Cov(\underline{X}) \underline{\ell} = \underline{\ell}' \underline{\Sigma} \underline{\ell} \quad (270)$$

para todas as populações

Então, a média global para todas as populações é dada por:

$$\underline{\bar{\mu}}_y = \frac{1}{g} \sum_{i=1}^g \mu_{iY} = \frac{1}{g} \sum_{i=1}^g \underline{\ell}' \underline{\mu}_i = \underline{\ell}' \underline{\bar{\mu}} \quad (271)$$

e a razão entre a soma dos quadrados das distâncias de todas as populações para a média global de  $Y$  e a variância de  $Y$  é dada por:

$$\frac{\sum_{i=1}^g (\mu_{iY} - \bar{\mu}_y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^g (\underline{\ell}' \underline{\mu}_i - \underline{\ell}' \underline{\bar{\mu}})^2}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} = \frac{\underline{\ell}' \left( \sum_{i=1}^g (\underline{\mu}_i - \underline{\bar{\mu}})(\underline{\mu}_i - \underline{\bar{\mu}})' \right) \underline{\ell}}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} = \frac{\underline{\ell}' \underline{B}_0 \underline{\ell}}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} \quad (272)$$

Essa razão mede a variabilidade "entre" os grupos de valores relativos  $Y$  com a variabilidade comum "dentro" dos grupos.

Analogamente para o caso de 2 populações, pode-se selecionar  $\underline{\ell}$  para maximizar a razão acima. Para tanto é conveniente normalizar  $\underline{\ell}$  de tal maneira que  $\underline{\ell}' \underline{\Sigma} \underline{\ell} = 1$ .

Seja  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s > 0$ , tal que  $s \leq \min(g-1, p)$  autovalores não nulos de  $\underline{\Sigma}^{-1} \underline{B}_0$  e  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_s$  os autovetores correspondentes normalizados, tal que  $\underline{e}' \underline{\Sigma} \underline{e} = 1$ .

Então o vetor dos coeficientes  $\underline{\ell}$  que maximiza a razão:

$$\underline{\ell} = \frac{\underline{\ell}' \underline{B}_0 \underline{\ell}}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} = \frac{\underline{\ell}' \left[ \sum_{i=1}^g (\underline{\mu}_i - \underline{\bar{\mu}})(\underline{\mu}_i - \underline{\bar{\mu}})' \right] \underline{\ell}}{\underline{\ell}' \underline{\Sigma} \underline{\ell}} \quad (273)$$

é dado por  $\underline{\ell}_1 = \underline{e}_1$ . A combinação linear  $\underline{\ell}'_1 \underline{X}$  é chamada de 1º discriminante. O valor  $\underline{\ell}_2 = \underline{e}_2$  maximiza a razão anterior, sujeito a covariância  $Cov(\underline{\ell}'_1 \underline{X}, \underline{\ell}'_2 \underline{X}) = 0$  e a combinação linear  $\underline{\ell}'_2 \underline{X}$  é chamada de 2º discriminante. Continuando, tem-se que  $\underline{\ell}_k = \underline{e}_k$  maximiza a razão sujeita a  $Cov(\underline{\ell}'_k \underline{X}, \underline{\ell}'_i \underline{X}) = 0$ , para todo  $i < k$ . Aqui  $\underline{\Sigma}^{1/2} \underline{\Sigma}^{1/2} = \underline{\Sigma}$  é chamado de k-ésimo determinante e para todos os discriminante a variância é igual a unidade.

A prova desse resultado é dado pelo Teorema da Decomposição Espectral, onde temos que  $\underline{\Sigma} = P' \underline{\Lambda} P$  onde  $P$  é o vetor de autovalores e  $\underline{\Lambda}$  é uma matriz diagonal com elementos positivos,  $\lambda_i$ .

Sabendo que :

$$\Sigma^{-1/2} = P' \Lambda^{-1/2} P$$

é uma matriz raiz quadrada simétrica e que sua inversa é dada por:

$$\Sigma^{-1/2} = P' \Lambda^{-1/2} P$$

satisfaz as propriedades,

1.  $\Sigma^{1/2} \Sigma^{1/2} = \Sigma$
2.  $\Sigma^{1/2} \Sigma^{-1/2} = I = \Sigma^{-1/2} \Sigma^{1/2}$
3.  $\Sigma^{-1/2} \Sigma^{-1/2} = \Sigma^{-1}$

Fazendo  $a = \Sigma^{1/2} \underline{\ell}$  então:

$$a'a = \underline{\ell}' \Sigma^{1/2} \Sigma^{1/2} \underline{\ell} = \underline{\ell}' \Sigma \underline{\ell}$$

e

$$a' \Sigma^{-1/2} B_0 \Sigma^{-1/2} a = \underline{\ell}' \Sigma^{1/2} \Sigma^{-1/2} B_0 \Sigma^{-1/2} \Sigma^{1/2} \underline{\ell} = \underline{\ell}' B_0 \underline{\ell}$$

Conseqüentemente, o problema se reduz a maximização expressa em  $a$  :

$$\frac{a' \Sigma^{-1/2} B_0 \Sigma^{-1/2} a}{a'a} \quad (274)$$

Pelo teorema de maximização de formas quadráticas para pontos em uma unidade esférica, tem-se o máximo da razão em  $\lambda_1$  que é o maior autovalor de  $\Sigma^{-1/2} B_0 \Sigma^{-1/2}$ . Este máximo ocorre quando  $a = \underline{e}_1$ , o autovetor normalizado associado com  $\lambda_1$ . Porque se:

$$\underline{e}_1 = a = \Sigma^{1/2} \underline{\ell}_1 \Rightarrow \underline{\ell}_1 = \Sigma^{-1/2} \underline{e}_1$$

então:

$$Var(\underline{\ell}_1' X) = \underline{\ell}_1' \Sigma \underline{\ell}_1 = \underline{e}_1' \Sigma^{-1/2} \Sigma \Sigma^{-1/2} \underline{e}_1 = \underline{e}_1' \Sigma^{-1/2} \Sigma^{1/2} \Sigma^{1/2} \Sigma^{-1/2} \underline{e}_1 = \underline{e}_1' \underline{e}_1 = 1$$

Sabendo-se ser  $a$  perpendicular a  $\underline{e}_1$  e que maximiza a razão  $\frac{a' \Sigma^{-1/2} B_0 \Sigma^{-1/2} a}{a'a}$ .

Então quando  $a = \underline{e}_2$  este será o autovetor normalizado correspondente para  $\lambda_2$ . Assim:

$$\underline{\ell}_2 = \Sigma^{1/2} \underline{e}_2$$

$$Cov(\underline{\ell}'_2 \underline{X}, \underline{\ell}'_1 \underline{X}) = \underline{\ell}'_2 \Sigma \underline{\ell}_1 = \underline{e}'_2 \Sigma^{1/2} \Sigma \Sigma^{1/2} \underline{e}_1 = \underline{e}'_2 \underline{e}_1 = 0$$

desde que  $\underline{e}_2 \perp \underline{e}_1$ .

Similarmente teremos  $Var(\underline{\ell}'_2) = \underline{\ell}'_2 \Sigma \underline{\ell}_2 = \underline{e}'_2 \underline{e}_2 = 1$  e continuando neste mesmo raciocínio temos que  $a = \underline{e}_k$  maximiza a razão  $\frac{a' \Sigma^{-1/2} B_0 \Sigma^{-1/2} a}{a'a}$  sendo

$a \perp \underline{e}_1, \underline{e}_2, \dots, \underline{e}_{k-1}$  e  $\underline{\ell}_k = \Sigma^{-1/2} \underline{e}_k$  satisfaz:

$$Cov(\underline{\ell}'_k \underline{X}, \underline{\ell}'_i \underline{X}) = \underline{\ell}'_k \Sigma \underline{\ell}_i = \underline{e}'_k \Sigma^{-1/2} \Sigma \Sigma^{-1/2} \underline{e}_i = \underline{e}'_k \underline{e}_i = \begin{cases} 0 & \text{se } i < k \\ 1 & \text{se } i = k \end{cases}$$

Então, se  $\lambda$  e  $\underline{e}$  são respectivamente autovalores e autovetores de  $\Sigma^{1/2} B_0 \Sigma^{1/2}$ , tem-se:

$$\Sigma^{-1/2} B_0 \Sigma^{-1/2} \underline{e} = \lambda \underline{e}$$

e multiplicando os dois termos por  $\Sigma^{-1/2}$  fica:

$$\Sigma^{-1/2} \Sigma^{-1/2} B_0 \Sigma^{-1/2} \underline{e} = \lambda \Sigma^{-1/2} \underline{e}$$

Então,  $\Sigma^{-1} B_0$  tem os mesmos autovalores que  $\Sigma^{-1/2} B_0 \Sigma^{-1/2}$ , mas os correspondentes autovetores são proporcionais a  $\Sigma^{-1/2} \underline{e} = \underline{\ell}$ , como se afirmou.

Assim,  $\underline{\ell}' \Sigma \underline{\ell} = \underline{e}' \Sigma^{-1/2} \Sigma \Sigma^{-1/2} \underline{e} = \underline{e}' \underline{e} = 1$  e as combinações lineares  $\underline{\ell}'_1 \underline{X}, \underline{\ell}'_2 \underline{X}, \dots, \underline{\ell}'_s \underline{X}$ , correspondem a autovalores não nulos de  $\Sigma^{-1} B_0$ , formando uma série de discriminantes com :

$$Var(\underline{\ell}'_i \underline{X}) = 1$$

$$Cov(\underline{\ell}'_i \underline{X}, \underline{\ell}'_k \underline{X}) = 0 \quad \text{para } i \neq k$$

Sabendo que  $\underline{\mu}_i$  e  $\Sigma$  são geralmente desconhecidos, pode-se então treinar uma série de observações corretamente classificadas. Supondo ser a série treinada uma amostra aleatória de tamanho  $n_i$  da população  $II_i$ , onde  $i = 1, 2, \dots, g$ , tem-se o vetor de média amostral é dado por:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

e o vetor de média global é definido abaixo, oriundo do vetor médio obtido de todas as observações de uma série treinada.

$$\bar{x} = \frac{\sum_{i=1}^g n_i \bar{x}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^g n_i}$$

A matriz  $B_0$  que corresponde a matriz soma dos produtos cruzados "entre" os grupos populacionais pode ser estimada por:

$$\hat{B}_0 = \sum_{i=1}^g (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

A estimativa de  $\Sigma$  se baseia na matriz de soma dos produtos cruzados "dentro" dos grupos amostrais apresentada abaixo, onde  $S_i$  é a matriz de covariância de cada grupo:

$$W = \sum_{i=1}^g (n_i - 1) S_i = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

Então a estimativa de  $\Sigma$  chamada de matriz de covariância conjunta é dada por:

$$\hat{\Sigma} = S_c = \frac{W}{(n_1 + n_2 + \dots + n_g - g)}$$

Tendo em vista que o mesmo  $\hat{\ell}$  que maximiza  $\frac{\hat{\ell}' \hat{B}_0 \hat{\ell}}{\hat{\ell}' S_c \hat{\ell}}$  também maximiza

$\frac{\hat{\ell}' \hat{B}_0 \hat{\ell}}{\hat{\ell}' W \hat{\ell}}$ , então se pode apresentar a otimização de  $\hat{\ell}$  na forma de autovetores,  $\hat{e}_i$  de

$W^{-1} \hat{B}_0$  porque se  $W^{-1} \hat{B}_0 \hat{e} = \hat{\lambda} \hat{e}$  então:

$$S_c^{-1} \hat{B}_0 \hat{e} = \hat{\lambda} (n_1 + n_2 + \dots + n_g - g) \hat{e}$$

Portanto o discriminante amostral de Fisher para várias populações pode ser definido como:

Seja  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_s > 0$  os autovalores não nulos de  $W^{-1} \hat{B}_0$  e  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_s$  os correspondentes autovetores.

Então, o vetor de coeficientes  $\hat{\underline{\ell}}'$  que maximiza a razão  $\frac{\hat{\underline{\ell}}' \hat{B}_0 \hat{\underline{\ell}}}{\hat{\underline{\ell}}' \hat{W} \hat{\underline{\ell}}}$  é dado por  $\hat{\underline{\ell}}_1 = \hat{e}_1$  e a combinação linear  $\hat{\underline{\ell}}_1' \underline{x} = \hat{e}_1' \underline{x}$  é chamada de 1º. Discriminante Amostral;  $\hat{\underline{\ell}}_2 = \hat{e}_2$  produz o 2º. Discriminante Amostral  $\hat{\underline{\ell}}_2' \underline{x} = \hat{e}_2' \underline{x}$ , generalizando tem-se que  $\hat{\underline{\ell}}_k' \underline{x} = \hat{e}_k' \underline{x}$  é o k-ésimo Discriminante Amostral para  $k \leq s$ .

O resultado fornece subsídios para classificar novos valores a partir do cálculo das distâncias euclidianas entre o valor da função discriminante no ponto selecionado e os valores das funções discriminantes dos valores médios dos grupos (populações), então a população escolhida é a que tenha menor distância. Para tanto, segue a regra:

Alocar  $x$  na população  $\Pi_k$  se:

$$\sum_{j=1}^r (\hat{y}_j - \bar{y}_{kj})^2 = \sum_{j=1}^r [\hat{\underline{\ell}}_j' (\underline{x} - \bar{\underline{x}}_k)]^2 \leq \sum_{j=1}^r [\hat{\underline{\ell}}_j' (\underline{x} - \bar{\underline{x}}_i)]^2 \quad \forall i \neq k \quad (275)$$

onde:  $\hat{y}_j = \hat{\underline{\ell}}_j' \underline{x}$  e  $\bar{y}_{kj} = \hat{\underline{\ell}}_j' \bar{\underline{x}}_k$  com  $r \leq s$

#### 2.4.4.4 Método distância mínima – populações não necessariamente gaussianas e matrizes de covariâncias iguais

O método de distância mínima classifica um indivíduo de uma classe em particular com base na classe de média mais próxima.

Este método é um caso particular do método de máxima verossimilhança descrito na seção 2.3.4.1, este considera as matrizes de covariância de todas as classes como diagonais e iguais e ainda as variâncias de cada componente como idênticas ou relativamente próximas, assim tem-se que:

$$\Sigma_i = \sigma^2 I \quad \text{para o caso populacional}$$

$$S_i = s^2 I \quad \text{para o caso amostral}$$

onde:  $\Sigma_i$  = matriz de covariância populacional da classe  $i$

$S_i$  = matriz de covariância amostral da classe  $i$

$\sigma^2$  = variância populacional

$S^2$  = variância amostral

$I$  = matriz identidade

Dada a função discriminante de máxima verossimilhança amostral tem-se:

$$g_i(\underline{x}) = -\ln|S_i| - (\underline{x} - \bar{x}_i)' S_i^{-1} (\underline{x} - \bar{x}_i) \quad (276)$$

onde  $\bar{x}_i$  = média amostral da classe  $i$ , pode-se substituir  $S_i$  por  $s^2 I$  para todo  $i$  e tem-se:

$$g_i(\underline{x}) = -\ln|s^2 I| - (\underline{x} - \bar{x}_i)' s^2 I (\underline{x} - \bar{x}_i) \quad (277)$$

Sendo  $s^2 I$  constante para todas as classes e um fator que não discrimina, pode assim ser eliminada da função gerando o que se chama de distância quadrática que é dada pela expressão:

$$d(\underline{x}, \bar{x}_i)^2 = (\underline{x} - \bar{x}_i)' (\underline{x} - \bar{x}_i) \quad (278)$$

Desta maneira a classificação pode ser feita com base na regra:

$$\underline{x} \in \Pi_i \quad \text{se} \quad (\underline{x} - \bar{x}_i)' (\underline{x} - \bar{x}_i) < (\underline{x} - \bar{x}_j)' (\underline{x} - \bar{x}_j) \quad \forall, i \neq j$$

ou

$$\underline{x} \in \Pi_i \quad \text{se} \quad d(\underline{x}, \bar{x}_i)^2 < d(\underline{x}, \bar{x}_j)^2 \quad \forall, i \neq j \quad (279)$$

Quando se deseja alocar um novo ponto, conclui-se que este deve ser classificado na classe que possua a menor distância quadrática euclidiana.

#### 2.4.4.5 Método distância Mahalanobis – populações não necessariamente gaussianas e matrizes de covariâncias iguais

Dada a expressão  $d_i^Q(\underline{x}) = -\frac{1}{2} \ln|\Sigma_i| - \frac{1}{2} (\underline{x} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i) + \ln p_i$ ,  $i = 1, 2, \dots, g$ , assume-se variâncias iguais para as várias populações. Deve-se reconhecer  $\underline{x}$  como  $\Pi_i$  se  $-\frac{1}{2} (\underline{x} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i) + \ln p_i$  tiver o maior valor para  $\forall i$ . Se as probabilidades *a priori* são desconhecidas assume-se  $p_i = 1/g$ . A expressão  $(\underline{x} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i) = D^2$  é conhecida

como distância de Mahalanobis do vetor  $\underline{x}$  ao vetor  $\underline{\mu}_i$ . Então para o caso real (amostral)

$\underline{x}_0$  será reconhecido como de  $\Pi_k$  se  $-\frac{1}{2}D_i^2(\underline{x}_0) + \ln p_i$  for o maior  $\forall i$ .

#### 2.4.5. Avaliação de Desempenho de Classificação

Segundo MÜLLER (1997, p.30), a classificação é constituída de duas etapas, primeiramente se faz uma amostragem de dados para as várias classes de interesse, então se determina através de funções discriminantes os limites que as separam. Este processo é conhecido na área computacional como de treinamento e aprendizado, mas na realidade trata-se do ajuste de um modelo.

A técnica mais comumente usada é retirar uma parte da amostra para treinamento e outra para testar o discriminante. Elabora-se uma tabela de contingência, denominada de Tabela de Classificação e também chamada de matriz de erro de classificação ou matriz de confusão (CHAVES NETO, 2003 p. 40), porém a ideal é a de Lachenbruch. Esta tabela ou matriz, verifica o relacionamento entre os dados de referência, previamente conhecidos, com os dados obtidos pelo classificador, onde as linhas dessa matriz representam os dados de referência e as colunas os mesmos dados treinados e classificados pelo método escolhido.

Para que se possa entender melhor esta tabela, selecionou-se, como exemplo, uma amostra composta de três grupos distintos (1, 2 e 3) e aplicou-se o método de Fisher, cujo resultado aparece na Tabela 6 que segue:

**TABELA 6- TABELA DE CLASSIFICAÇÃO USANDO O MÉTODO DE FISHER**

GRUPOS DE REFERÊNCIA	GRUPOS CLASSIFICADOS PELOS MÉTODO			
	1	2	3	Total
1	15(28,85%)	0(0,00%)	0(0,00%)	15
2	5(9,62%)	20(38,46%)	0(0,00%)	25
3	0(0,00%)	0(0,00%)	12(23,08%)	12
<b>TOTAL</b>	15	20	12	<b>52</b>

Fonte: Resultados obtidos no programa SIAVAL 1.0

Na matriz de erro, obtida da Tabela 6, as colunas se referem ao número de amostras classificadas pelo método e as linhas indicam os dados nos grupos referência,

com isto os valores apresentados na diagonal principal mostra a quantidade de dados corretamente classificados.

Desta maneira a tabela ou a matriz permitem retirar várias conclusões sobre a performance, tais como o total de classificação correta para cada grupo, observado na diagonal principal da tabela e fora desta pode-se obter os erros de exclusão e inclusão como mostra a primeira linha da tabela onde o grupo 1 não foi classificada corretamente, sendo considerada pelo classificador como grupo 2 em uma das amostras, caracterizando um erro de exclusão do grupo 1 e erro de inclusão no grupo 2.

Somando-se os elementos da diagonal principal, correspondente ao total de classificação correta, e dividindo pelo número total de amostras observados, ter-se-á um valor do percentual de acerto do classificador. Que no caso do exemplo é dado por:

$$\text{Percentual de Classificação Correta} = \frac{15 + 20 + 12}{52} \times 100 = 90,38\%$$

A taxa de erro obtida dessa forma chama-se Taxa Aparente de Erro (CHAVES NETO, 2003 p. 40). Não é a mais indicada para se avaliar o desempenho do modelo, pois utiliza os mesmos dados usados no ajuste para fazer a avaliação. Um modo mais seguro de se medir a eficiência de qualquer método de reconhecimento e classificação é o Procedimento de Retenção de Lachenbruch.

O Procedimento de Retenção de Lachenbruch segundo JOHNSON & WICHERN (1992 p. 498) retêm uma observação das  $n$  disponíveis e utiliza todas as restantes na obtenção das equações do método proposto. Classifica-se essa observação num dos grupos propostos e anota-se este resultado. Repete-se este procedimento até que todas as observações forem retidas e a seguir coloca-se os resultados na tabela de classificação.

Para entender melhor o procedimento apresenta-se a seguir o algoritmo descrito em MARQUES & LIMA (2002, p.56) para dois grupos de classificação e estendida aqui para vários grupos.

- Passo 1: Inicie com as observações do grupo 1, omita uma observação desse grupo e construa as funções de classificação usando as  $(n_1-1)$  restantes do grupo 1 acrescidas das  $n_p$  observações dos  $p$  grupos que compõem a análise, .
- Passo 2: Classifique a observação retida, usando as funções construídas no passo 1.

- Passo 3: Repita os passos 1 e 2 até que todas as observações de todos os grupos sejam classificadas e construa a tabela de classificação.

Como exemplo, utilizou-se os mesmos dados da Tabela 6 e fez-se a tabela de classificação pelo Procedimento de Retenção de Lachenbruch apresentado na Tabela 7 a seguir.

**TABELA 7- TABELA DE CLASSIFICAÇÃO USANDO O MÉTODO DE FISHER E PROCEDIMENTO DE RETENÇÃO DE LACHENBRUCH**

GRUPOS DE REFERÊNCIA	GRUPOS CLASSIFICADOS PELOS MÉTODO			
	1	2	3	Total
1	<b>15(28,85%)</b>	0(0,00%)	0(0,00%)	15
2	4(7,69%)	<b>20(38,46%)</b>	1(1,92%)	25
3	0(0,00%)	2(3,85%)	<b>10(19,23%)</b>	12
<b>TOTAL</b>	19	24	11	<b>52</b>

**Fonte:** Resultados obtidos no programa SIAVAL 1.0

O percentual de classificação correta encontrado foi de 86,54%, que segundo CHAVES NETO (2003, p.41) é um percentual mais confiável e realista, dado que quando as observações foram classificadas elas não faziam parte do modelo ajustado.

### **3 MATERIAIS E MÉTODOS**

#### **3.1 INTRODUÇÃO**

O programa computacional de avaliação proposto neste trabalho é de fácil execução e possibilita fazer a análise de itens de instrumentos, ranqueamento dos indivíduos e reconhecimento de novos destes indivíduos em padrões pré-estabelecidos. Este programa é um sistema integrado de metodologias estatísticas que são: Teoria de Resposta ao Item, Análise Fatorial e Reconhecimento de Padrões e Classificação.

Com o intuito de mostrar o potencial de aplicação do programa foi proposto, neste trabalho, uma avaliação na área de educação (candidatos ao vestibular da Universidade X e alunos do curso de Estatística da Universidade Federal do Paraná) e na área de Engenharia de Produção (fornecedores de empresas).

A metodologia utilizada no programa, a coleta de dados e a construção do programa serão detalhadas nas próximas seções deste capítulo.

#### **3.2 MÉTODOS ESTATÍSTICOS DE ANÁLISE**

O programa foi concebido com intuito de facilitar o trabalho de pesquisadores na parte de estatística e por isso os métodos foram selecionados são os que apresentam melhores resultados práticos e estão relacionados a seguir

##### **3.2.1. Teoria Clássica dos Testes**

A Teoria Clássica dos Testes foi introduzida no programa para análise dos itens do instrumento, descritas nos itens 2.1.3 e 2.1.4. Utilizou-se o coeficiente de correlação de Kuder-Richardson para verificar a confiabilidade do instrumento e as correlações bisserial e bisserial por ponto que medem a capacidade de discriminação do item em relação ao resultado do teste. Além disso, outras estatísticas descritivas foram incluídas no programa como médias, desvio padrões, quartis, etc.

### 3.2.2. Teoria da Resposta ao Item

O programa foi feito para analisar somente variáveis dicotômicas ou com possibilidade de serem dicotomizadas, por isso foram utilizados os Modelos Logísticos de Dois e Três Parâmetros Unidimensionais Dicotômicos.

Selecionou-se para estimar os parâmetros dos itens o Método de Máxima Verossimilhança Marginal com algoritmo EM, e a o Método de Estimação Bayesiana da Média a *Posteriori* ou *Expected a Posteriori* (EAP) para aferir as habilidades dos respondentes. As habilidades são estimadas após a retirada dos itens ruins, isto é, que não se ajustam ao modelo e destes foram calculados os escores verdadeiros dos respondentes, conforme metodologia descrita no item 2.1.3.4.4, com os quais se procederá à classificação dos mesmos.

Após a estimação dos parâmetros pode-se obter no programa SIAVAL 1.0 os gráficos da Curva Característica e a Função de Informação dos itens.

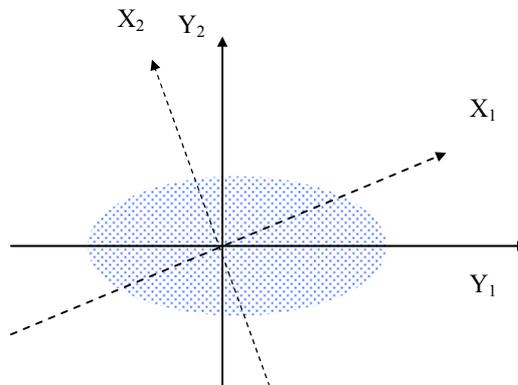
### 3.2.2 Análise Fatorial

A análise fatorial que é usada neste programa principalmente no caso de ranqueamento de dados não dicotômicos. Portanto, utiliza a matriz de correlação de Pearson e não a tetracórica, como seria o caso dicotômico.

Assim, no sistema de classificação, cada indivíduo (empresa) é representado por um conjunto de respostas dos itens analisados, vetor do tipo  $\underline{X}_j' = [X_1 \ X_2 \ \dots \ X_I]$  com  $j = 1, 2, \dots, n$ , onde  $I$  é o número de itens e  $n$  o número de indivíduos analisados.

As respostas aos itens do instrumento de cada indivíduo (empresa) são automaticamente armazenadas em um vetor  $\underline{X}_j$ , que posteriormente poderão ser sumarizadas na matriz de correlação (denotada por  $R$ ) refletindo a estrutura do relacionamento das componentes do vetor. Cada vetor (indivíduo ou empresa) será um ponto no sistema cartesiano com  $I$  eixos.

A metodologia utilizada identifica novas variáveis, chamadas de fatores, os quais, se encontram em outro sistema de eixos ortogonais nas direções de maior variabilidade dos dados, como se pode observar na Figura 15 a seguir.



FORNTE: A AUTORA

**FIGURA 15- ROTAÇÃO DOS FATORES**

Os pesos  $\underline{\ell}_{ij}$  serão obtidos dos autovetores da matriz de correlação e serão considerados os  $k < I$  fatores que realmente tiverem importância, medidos pelos autovalores da matriz de correlação  $R$  ou seja aqueles com autovalores  $\lambda_i \geq 1$ .

Cada vetor tem um escore em cada fator e que corresponde a sua posição no novo sistema de eixos  $[f_1 f_2 \dots f_k]$ .

O ranqueamento do vetor (indivíduo ou empresa) foi feito através do escore final que corresponde à média dos escores fatoriais, ponderada pelos autovalores que ditam a importância dos fatores, obtido através da equação que segue.

$$E_j = (\lambda_1 f_{1j} + \lambda_2 f_{2j} + \dots + \lambda_k f_{kj}) / (\lambda_1 + \lambda_2 + \dots + \lambda_k) \quad (280)$$

Uma vez que um indivíduo (empresa) deve ser avaliado, é atribuído a cada indivíduo (empresa), um escore final para que todos os indivíduos (empresas) avaliados sejam, então, classificados segundo uma escala de zero a cem que contenda todos os escores, permitindo assim, uma completa avaliação sob os mais variados ângulos de análise.

### 3.2.3 Reconhecimento de Padrões e Classificação

A alocação de um novo indivíduo em um dos  $g$  grupos pré-definidos é conhecido como classificação e fundamenta-se na atribuição de classes para estes indivíduos através do processo denominado de Reconhecimento de Padrões.

A Classificação se refere aos métodos de atribuição de classes a determinados conjuntos de dados através dos métodos estatísticos e o Reconhecimento de Padrões se refere a alocação de um conjunto de variáveis observáveis nas suas devidas classes correspondentes, feitas neste trabalho através de regras definidas para um do método citado a seguir.

O Método Linear de Fisher foi deduzido com o propósito de separar várias populações (classes) tanto quanto possível, não sendo necessário assumir que as populações sejam normais multivariadas; no entanto deve-se assumir que suas matrizes de covariância sejam iguais. A idéia principal deste método é de transformar observações multivariadas  $\underline{X}' = [X_1 \ X_2 \ \dots \ X_I]$  ( $I$  é o número de itens) em combinações lineares univariadas  $Y_k$  ( $k$  o número de classes), com intuito de reduzir a dimensão de um grande número de características para relativamente poucas combinações lineares.

Seja  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_s > 0$  os autovalores não nulos de  $W^{-1}\hat{B}_0$  e  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_s$  os correspondentes autovetores. Então, o vetor de coeficientes  $\hat{\ell}'$  que maximiza a razão  $\frac{\hat{\ell}'\hat{B}_0\hat{\ell}}{\hat{\ell}'W\hat{\ell}}$  é dado por  $\hat{\ell}_1 = \hat{e}_1$  e a combinação linear  $\hat{\ell}'_1 \underline{x} = \hat{e}_1' \underline{x}$  é chamada de 1º.

Discriminante Amostral;  $\hat{\ell}_2 = \hat{e}_2$  produz o 2º Discriminante Amostral  $\hat{\ell}'_2 \underline{x} = \hat{e}_2' \underline{x}$ , generalizando tem-se que  $\hat{\ell}'_k \underline{x} = \hat{e}_k' \underline{x}$  é o k-ésimo Discriminante Amostral para  $k \leq s$ .

A regra de alocar  $\underline{x}$  na população  $\pi_k$  é na menor distância euclidiana dada por:

$$\sum_{j=1}^r (\hat{y}_j - \bar{y}_{kj})^2 = \sum_{j=1}^r [\hat{\ell}'_j (\underline{x} - \bar{\underline{x}}_k)]^2 \leq \sum_{j=1}^r [\hat{\ell}'_j (\underline{x} - \bar{\underline{x}}_i)]^2 \quad \forall i \neq k \quad (281)$$

onde,  $\hat{y}_j = \hat{\ell}'_j \underline{x}$ ,  $\bar{y}_{kj} = \hat{\ell}'_j \bar{\underline{x}}_k$  e  $\hat{\ell}'_j$  são os autovetores correspondentes aos autovalores não-nulos da matriz  $W^{-1}\hat{B}_0$ , onde  $W^{-1}$  é a matriz da soma de produtos cruzados dentro

dos grupos amostrais e  $\hat{B}_0$  é a matriz soma dos produtos cruzados entre os grupos amostrais

Geralmente, quando se trabalha com reconhecimento de padrão considera-se algumas suposições acerca das populações (grupos a serem discriminados). Primeiramente, deve-se certificar se as populações seguem uma distribuição gaussiana e, a seguir, verificar se as variâncias destas populações são iguais. Neste trabalho os dados supostamente são provenientes de uma única população, então, as variâncias entre os grupos discriminados serão consideradas iguais. O método não necessita que se trabalhe com populações gaussianas, isto facilita ao usuário, pois não necessita de nenhuma intervenção, basta que ele aplique o método.

### 3.3 COLETA DE DADOS

A coleta dos dados para verificar a eficiência do programa será apresentada nas seções que seguem, detalhadas a cada um dos campos de estudo utilizados nesta tese.

#### 3.3.1 Avaliação de Candidatos ao Vestibular e Evasão de Alunos do curso de Estatística da UFPR.

##### 3.3.1.1 Coleta de dados

A partir de dados de uma Universidade U foi selecionada para análise um teste da matéria M, dado que tem um número grande de respondentes (20.550) e itens (60), sendo que cada item possui 5 alternativas.

Para determinar a evasão de alunos foram coletados dados de 163 ex-alunos (sendo 155 para classificação e 8 de controle aleatoriamente escolhidos) do curso de Estatística da UFPR que iniciaram o curso através do vestibular nos anos de 1998, 1999 e 2000. O dados se referem às notas de três disciplinas obtidas no primeiro semestre cursado, escores no vestibular, dados sócio econômicos e outras questões que se tinha disponível sobre os ex-alunos. As variáveis analisadas estão descritas no Apêndice 8.

### 3.3.1.2 Procedimentos de avaliação

Para a análise dos itens do teste da matéria M, aplicada na Universidade U, foi usado o Modelo Logístico de Três Parâmetros Unidimensional, determinando os seus parâmetros (discriminação, dificuldade e acerto ao acaso), a função de informação dos itens e a função de informação do instrumento. O coeficiente de confiabilidade e as correlações oriundas da Teoria Clássica são incluídos além das estatísticas descritivas dos itens. O ranqueamento dos candidatos foi feito utilizando a estimativa das habilidades dos mesmos após a retirada dos itens considerados ruins.

Na análise de evasão foi utilizada a técnica de Reconhecimento de Padrões e Classificação denominada Função Discriminante Linear de Fisher para várias populações (grupos). Os dados foram classificados em três grupos: não terminaram o curso (evasão), terminaram o curso com atraso e terminaram o curso no tempo certo.

### 3.3.2 Avaliação de Fornecedores

#### 3.3.2.1 Procedimentos de avaliação.

O programa SIAVAL 1.0 desenvolvido neste trabalho possibilita o ranqueamento das empresas fornecedoras, pois algumas empresas que avaliam seus fornecedores costumam divulgar e premiar os melhores fornecedores e seria interessante que elas conhecessem estas técnicas estatísticas, caso elas não a utilizem.

Além disso, pode-se verificar se um fornecedor recentemente incluído no quadro corresponderia às expectativas da empresa a longo prazo, para tanto as técnicas de Reconhecimento de Padrões e Classificação descritas neste trabalho foram aplicadas.

A avaliação de fornecedores é, hoje, uma atividade necessária às empresas e se tornará rotineira, com o tempo e a facilidade de técnicas.

## 3.4 ELABORAÇÃO DO PROGRAMA

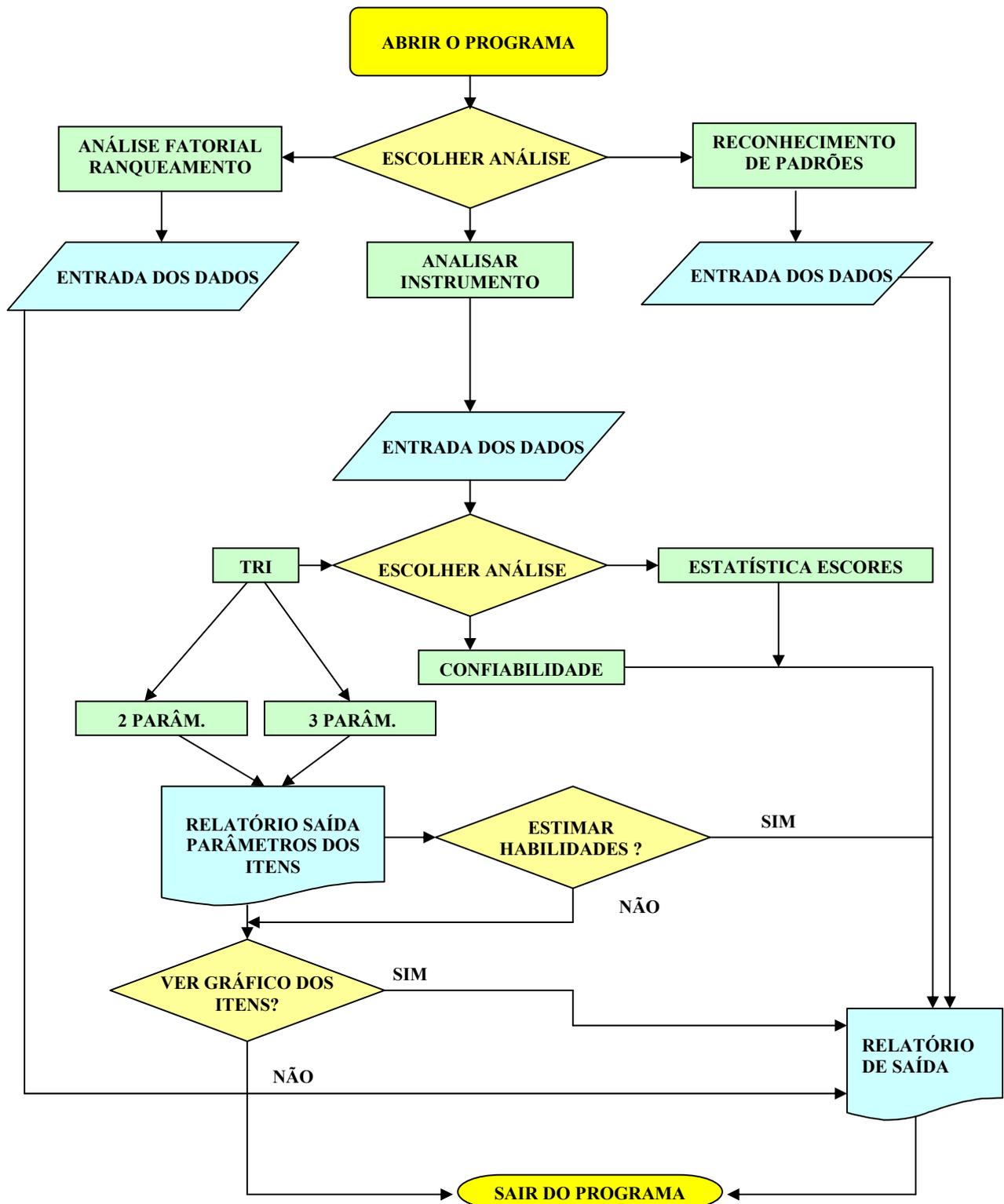
O programa SIAVAL 1.0 foi elaborado em ambiente *Windows*, linguagem computacional Visual Basic e de forma completamente genérica, ou seja, de tal modo que

um usuário leigo possa analisar instrumentos, ranquear e alocar novos indivíduos (empresas) com as variáveis de seu interesse. Assim, pesquisadores e profissionais de todas as áreas que necessitem fazer avaliação possam facilmente manusear os programas.

A entrada de dados tem um formato da planilha Excel o que possibilita uma cópia dos dados diretamente de uma planilha Excel feita pelo usuário.

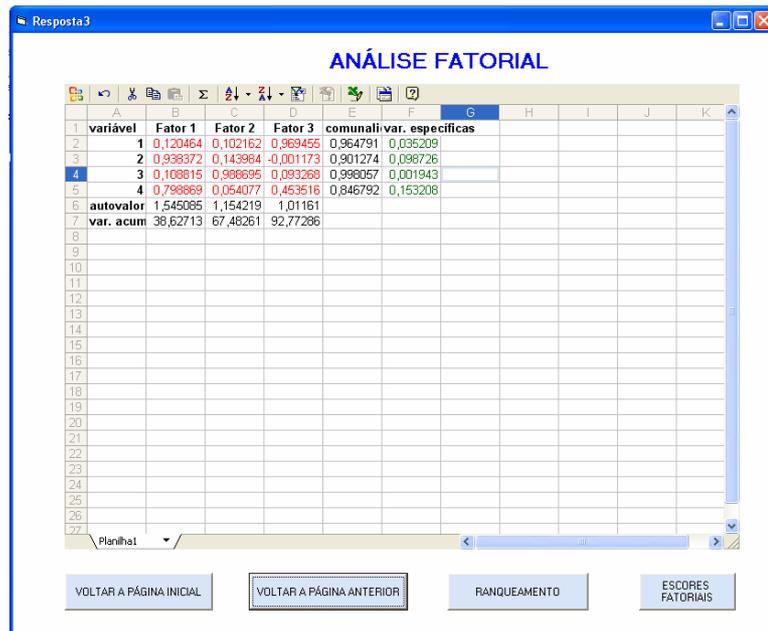
O fluxograma do programa, que a princípio denomina-se SIAVAL 1.0 (Sistema Integrado de Avaliação, versão 1.0), apresenta-se na Figura 16 a seguir.

Os relatórios de saída são apresentados em gráficos e em planilhas Excel para facilitar o relatório da análise feita pelo usuário (pesquisador), exemplificados nas Figuras 17 e 18.



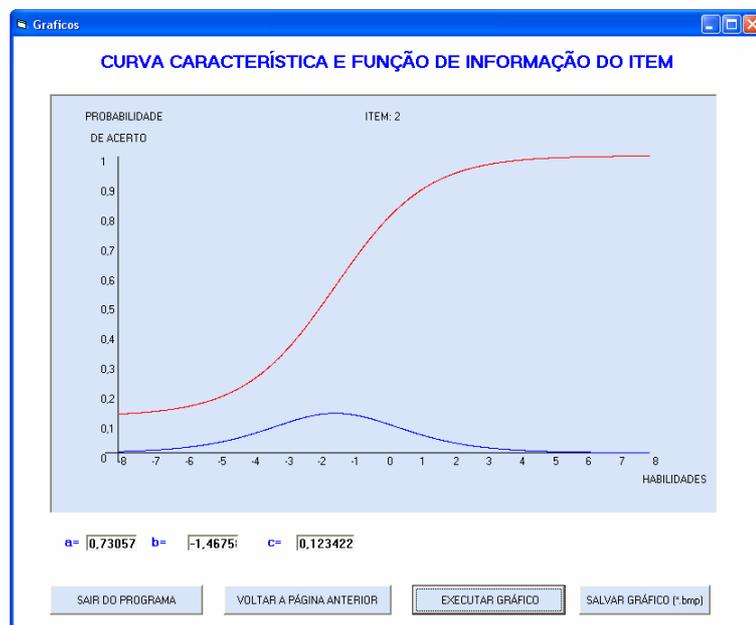
**FIGURA 16- FLUXOGRAMA DO PROGRAMA SIAVAL 1.0**

A Figura 17 mostra parte do relatório de saída da Análise Fatorial do programa.



**FIGURA 17- RELATÓRIO DE SAÍDA DA ANÁLISE FATORIAL**

Visualiza-se na Figura 18 um gráfico típico de saída da análise de itens, a Curva Característica e a Função de Informação de um item.



**FIGURA 18- RELATÓRIO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM NA TEORIA DA RESPOSTA AO ITEM**

Todas as telas do programa SIAVAL 1.0 podem ser vistas no Apêndice 11.

## 4- RESULTADOS E DISCUSSÃO

### 4.1- AVALIAÇÃO DO INSTRUMENTO DO TESTE DA MATÉRIA M DO VESTIBULAR DA UNIVERSIDADE U E RANQUEAMENTO DOS ESCORES VERDADEIROS DOS CANDIDATOS UTILIZANDO A TEORIA DA RESPOSTA AO ITEM

#### 4.1.1- Análise do Teste da Matéria M

Esse teste apresentou um índice de confiabilidade de 89,20%, calculado pelo programa SIAVAL 1.0, que utiliza a correlação de Kuder-Richardson, dada pela equação 282 a seguir, já comentada na seção 2.1.3 deste trabalho. Essa confiabilidade pode ser considerada boa, porém o ideal segundo a literatura é um valor acima de 90% (CHAVES & TURIN, p.51). Este valor da confiabilidade indica que nos 60 itens que compõem o teste existem poucos traços significativos de ambigüidade e de textos que conduzam a interpretações diferentes, por partes dos candidatos, quanto à solução dos itens (questões), ou seja:

$$\rho_{KR} = \frac{n}{n-1} \left( 1 - \frac{\sum_{i=1}^n p_i(1-p_i)}{s^2} \right) \quad (282)$$

onde:

$n$  = número de itens do teste

$p_i$  = proporção dos respondentes que acertaram o item  $i$

$s^2$  = variância dos escores do teste com base nos  $N$  escores dos respondentes.

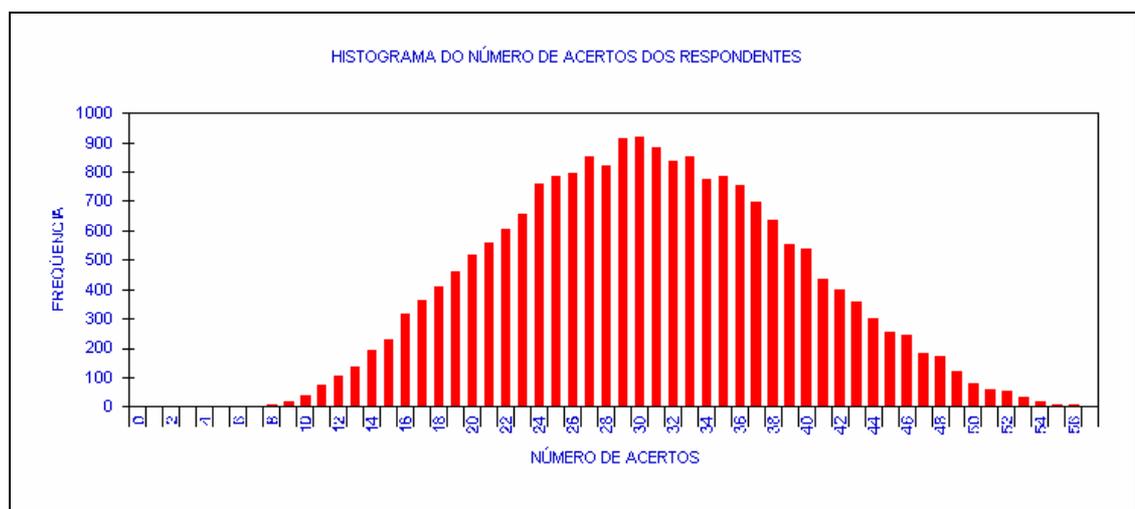
Observa-se na Tabela 8, que descreve o número de acertos dos candidatos nos 60 itens propostos no teste, que o número médio de acertos resultou em 30,39 com um desvio padrão de 8,75. O número mínimo de acertos foi 0 (zero) e o máximo de 57. Apresenta-se, também nessa tabela os quartis, sendo que o primeiro quartil (24) expressa o máximo alcançado por 25% dos candidatos que tiveram menor número de acertos. Já número máximo alcançado pela metade dos candidatos (50%) é o valor observado no 2º. Quartil, que corresponde também a mediana, e de 30 acertos. O terceiro quartil resultou no valor 37 acertos representa o máximo alcançado por 75% dos candidatos, isto significa dizer que apenas 25% dos candidatos acertaram mais de 37 dos itens aplicados.

**TABELA 8- ESTATÍSTICAS DESCRITIVAS NÚMERO DE ACERTO DOS CANDIDATOS**

ESTATÍSTICA	NÚMERO DE ACERTOS
MÉDIA ARITMÉTICA	30,39
DESVIO PADRÃO	8,75
VALOR MÍNIMO	0
VALOR MÁXIMO	57
1°. QUARTIL	24
2°. QUARTIL	30
3°. QUARTIL	37

FONTE: RESULTADO SIAVAL 1.0

A Figura 19 a seguir mostra o histograma do número de acertos dos respondentes.



FONTE: RESULTADO SIAVAL 1.0

**FIGURA 19- HISTOGRAMA DO NÚMERO DE ACERTOS DOS CANDIDATOS**

Observa-se no histograma do número de acerto dos candidatos que a distribuição dos mesmos pode ser considerada aproximadamente normal e isto indica que o instrumento utilizado no teste é considerado bom, pois em geral espera-se que a variável do número de acertos siga a curva normal.

#### 4.1.2- Análise dos Itens do Teste da Matéria M

Através do programa SIAVAL 1.0 estimou-se os parâmetros de discriminação, dificuldade e acerto aleatório dos 60 itens do teste da matéria M aplicada na Universidade U. O método de estimação utilizado pelo programa é o de Máxima Verossimilhança Marginal com algoritmo EM (Esperança e Maximização) descrito na seção 2.1.3.4.2 deste trabalho. Esse método ajusta o Modelo Logístico de Três Parâmetros Unidimensional Dicotômico apresentado a seguir pela equação 283. Utilizou-se um critério de convergência de 0,01 com restrição de parada em 10 ciclos, apresentadas na Tabela 9 a seguir.

$$P(U_{ji} = 1 / \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (283)$$

onde:

$P(U_{ji}=1|\theta_j)$  = probabilidade do respondente j acertar o item i dado que ele possui uma habilidade  $\theta_j$ .

$U_{ji}$  = resposta dada pelo respondente j ao item i.

$\theta_j$  = habilidade ou traço latente do respondente j.

$a_i$  = parâmetro de discriminação do item i.

$b_i$  = parâmetro de dificuldade do item i.

$c_i$  = parâmetro de acerto casual do item i.

$D$  = fator de escala geralmente igual a 1,7 para que a função logística forneça resultados semelhantes ao da distribuição normal.

**TABELA 9- ANÁLISE DOS ITENS DO TESTE DA MATÉRIA M DO VESTIBULAR DA  
UNIVERSIDADE U**

ITEM	DISCRIMI- NAÇÃO $a_i$	DIFICUL- DADE $b_i$	ACERTO CASUAL $c_i$	CONVER- GÊNCIA	PERCENTUAL DE ACERTOS	CORRELAÇÃO	
						BISSERIAL	BISSERIAL POR PONTO
1	1,2286	-0,1069	0,1870	0,0009	61,05	0,4393	0,3457
2	1,1872	-1,6010	0,2299	0,0071	76,02	0,4192	0,3053
3	1,5358	-0,7153	0,1880	0,0037	67,45	0,5256	0,4062
4	0,9458	2,6094	0,1764	0,0172	29,27	0,3531	0,2657
5	0,9120	3,6437	0,0440	0,0174	20,09	0,3088	0,2158
6	0,8254	2,2867	0,2310	0,0117	38,67	0,3015	0,2358
7	1,1550	0,0691	0,1902	0,0014	59,38	0,4094	0,3240
8	0,9276	-1,3386	0,3360	0,0262	75,53	0,3191	0,2359
9	0,9922	1,9018	0,2253	0,0005	40,45	0,3649	0,2872
10	1,4814	2,3644	0,2154	0,0043	35,82	0,5230	0,4040
11	1,2302	-2,1218	0,1636	0,0072	80,21	0,4315	0,3033
12	0,8219	-0,2200	0,1908	0,0025	62,29	0,2918	0,2292
13	0,4749	4,9791	0,0594	0,0105	10,21	0,1574	0,0913
14	0,5943	3,1761	0,1417	0,0328	30,48	0,2337	0,1765
15	1,1335	-0,3450	0,1852	0,0013	63,52	0,4035	0,3165
16	0,8149	3,1050	0,0599	0,0118	24,26	0,2635	0,1911
17	1,2821	3,7364	0,0430	0,0042	22,22	0,4605	0,3280
18	0,5340	3,2066	0,2726	0,0416	36,99	0,1611	0,1248
19	0,7396	1,7136	0,2102	0,0116	45,11	0,2688	0,2138
20	1,6177	-1,0255	0,2021	0,0039	70,39	0,5539	0,4219
21	1,7545	-0,9772	0,1941	0,0050	69,82	0,5873	0,4513
22	1,3890	-1,1935	0,1878	0,0083	71,91	0,4788	0,3646
23	1,5136	-2,9674	0,1829	0,0096	87,88	0,5670	0,3674
24	1,4582	-0,7928	0,2035	0,0055	68,31	0,5075	0,3901
25	1,3597	-1,3977	0,1832	0,0048	73,85	0,4723	0,3553
26	0,6612	2,2794	0,2031	0,0048	39,33	0,2392	0,1879
27	1,1874	-3,2474	0,1900	0,0228	86,55	0,4440	0,2897
28	1,2560	1,9739	0,2251	0,0067	39,46	0,4390	0,3446
29	0,8330	-3,2420	0,1658	0,0099	37,21	0,2932	0,2290
30	1,4567	-1,4620	0,1959	0,0038	74,92	0,4988	0,3733
31	1,8022	-0,6938	0,1416	0,0026	65,97	0,6104	0,4772
32	1,4029	-0,8494	0,2257	0,0098	69,06	0,4908	0,3746

33	1,6737	-2,0020	0,1518	0,0073	79,22	0,5770	0,4098
34	1,4068	-0,5598	0,1868	0,0021	65,76	0,4887	0,3815
35	1,6889	-0,5465	0,1861	0,0021	65,39	0,5763	0,4488
36	0,1413	3,3797	0,0350	5,6791	18,51	-0,0090	-0,0061
37	0,4167	4,7241	0,0508	0,0206	12,76	0,1322	0,0793
38	0,8822	-4,7017	0,1929	0,0282	93,67	0,4288	0,2365
39	1,4506	-1,6100	0,1576	0,0071	74,77	0,5005	0,3738
40	0,7896	2,0443	0,3052	0,0111	41,92	0,2334	0,1839
41	1,5962	-0,0701	0,1943	0,0010	60,77	0,5415	0,4285
42	0,7435	3,0217	0,1872	0,0046	30,23	0,2780	0,2105
43	0,2515	-1,5133	0,2153	0,3438	64,92	0,0508	0,0398
44	1,3537	2,8778	0,2118	0,0041	30,60	0,4881	0,3683
45	0,9969	-0,8188	0,1888	0,0018	68,23	0,3522	0,2705
46	1,2654	0,2733	0,1926	0,0008	57,32	0,4469	0,3549
47	1,5721	-2,8195	0,1680	0,0097	86,25	0,5891	0,3807
48	0,0343	4,4643	0,0490	0,0535	18,36	0,1635	0,1108
49	1,7870	2,2875	0,2128	0,0023	36,85	0,6097	0,4729
50	0,7024	-0,0471	0,1794	0,0012	60,28	0,2457	0,1940
51	1,3711	1,3466	0,2245	0,0051	46,10	0,4747	0,3779
52	1,0106	1,9800	0,2378	0,0049	39,08	0,3716	0,2923
53	1,5110	4,1329	0,0356	0,0034	18,36	0,5569	0,3774
54	0,3018	4,6801	0,0970	0,0382	22,22	0,1142	0,0813
55	0,2895	1,4521	0,1848	0,0068	51,20	0,1106	0,0883
56	0,5317	3,9798	0,1578	0,0079	23,97	0,1412	0,1004
57	1,4515	2,6285	0,2146	0,0043	33,17	0,5131	0,3946
58	1,3373	-0,6675	0,1891	0,0016	66,73	0,4630	0,3600
59	0,5249	3,8475	0,1842	0,0237	25,73	0,1726	0,1255
60	0,0188	4,7582	0,0431	0,1070	13,27	-0,0484	-0,0302

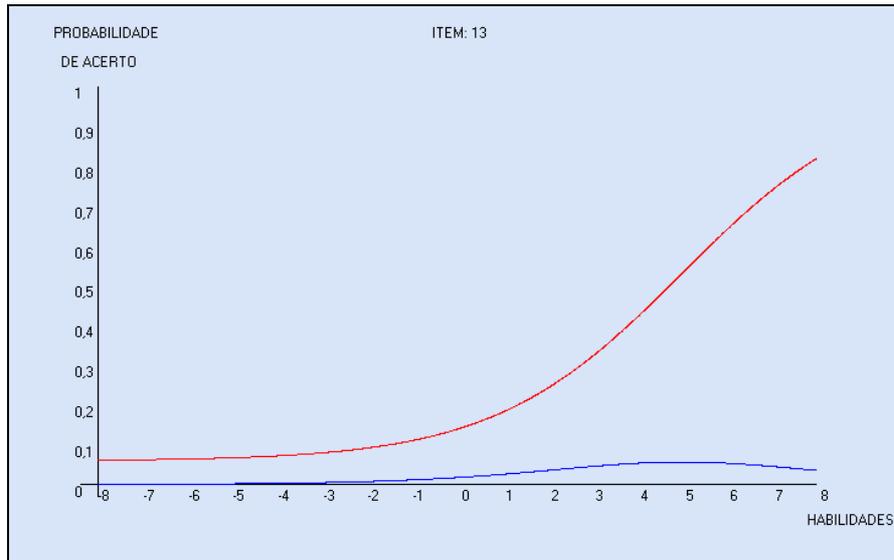
FONTE: RESULTADO DA ANÁLISE FEITA NO SIAVAL 1.0

Apresenta-se, também nessa tabela o percentual de acertos observados em cada item e as correlações bisserial e bisserial por ponto com intuito de auxiliar a análise dos itens. Além disso, consta no Apêndice 9 os gráficos da Curva Característica e da Função de Informação de cada item.

O parâmetro índice de discriminação do item, em termos práticos, varia de 0 a 2. Assim interpretou-se um baixo índice de discriminação um valor menor de 0,5. Considerou-se no parâmetro de dificuldade, um intervalo prático de -4 a +4 na sua estimação. Portanto, quanto mais próximo de -4 mais fácil o item e de +4 mais difícil e conseqüentemente os valores em torno de zero representam um item de média dificuldade. Como o parâmetro de acerto aleatório depende do número de alternativas dos itens, o valor ficou em torno de 0,20, dado que cada item possibilitou 5 alternativas de resposta.

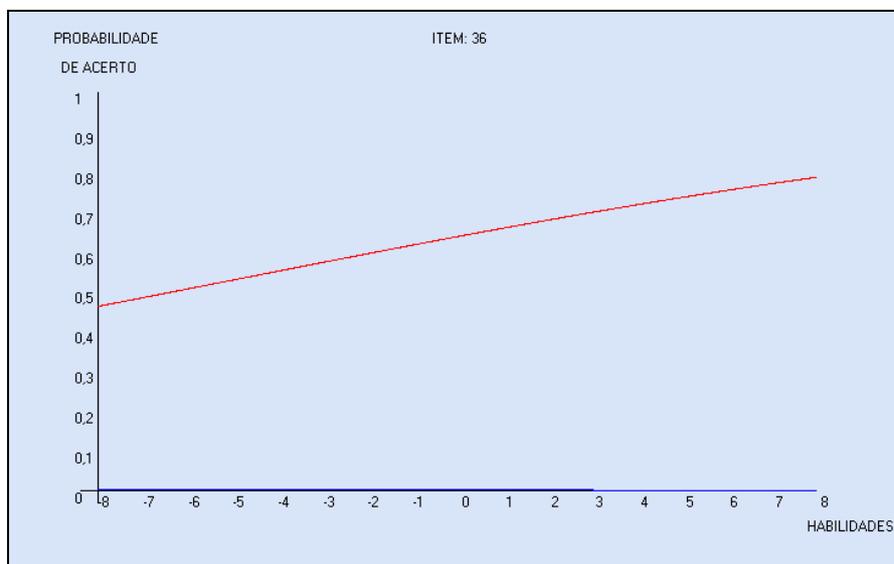
Observa-se que os itens 13, 36, 37, 43, 48, 54, 55 e 60 apresentaram discriminação abaixo de 0,50, cujos valores são respectivamente 0,4749; 0,1413; 0,4167; 0,2515; 0,0343; 0,3018; 0,2895 e 0,0188. Desses têm-se que os itens 36, 43 e 60 não se ajustaram ao modelo, pois que não convergiram corretamente. Dado que o critério de convergência foi de 0,01 os itens 36, 43 e 60 não apresentaram convergência e obtiveram um valor de respectivamente de 5,6791; 0,3438 e 0,1070. Na Teoria Clássica dos Testes, citada na seção 2.1.4, quando um item tem discriminação baixa possui correlação bisserial e bisserial por ponto também baixa, assim constata-se isso nos itens citados o que dá uma consistência a estimação dos parâmetros dos itens feita pelo programa SIAVAL 1.0.

Examinando os gráficos, expostos nas Figuras 20 a 27 correspondentes aos itens 13, 36, 37, 43, 48, 54, 55 e 60 percebe-se que não são itens bons e afetam negativamente o desempenho do instrumento.



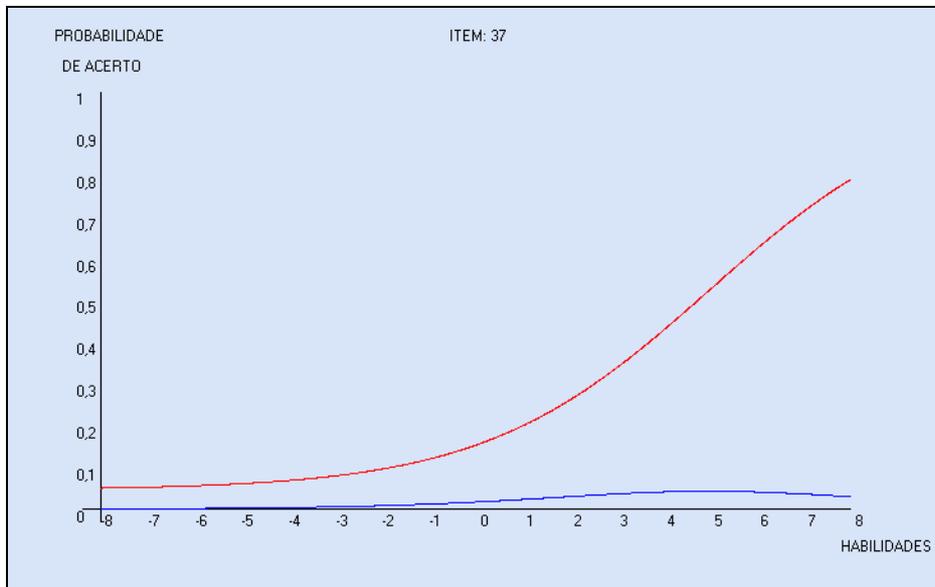
**FIGURA 20- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 13**

O item 13 além de apresentar o parâmetro de discriminação 0,4749 considerado neste trabalho como baixo tem um índice de dificuldade 4,9791, conseqüentemente é um item difícil evidenciado pelo menor percentual de acertos dos itens de 10,21% e o acerto casual foi baixo (0,0594). Olhando a Figura 20 percebe-se que a Função de Informação do Item não apresenta o formato de uma distribuição normal.



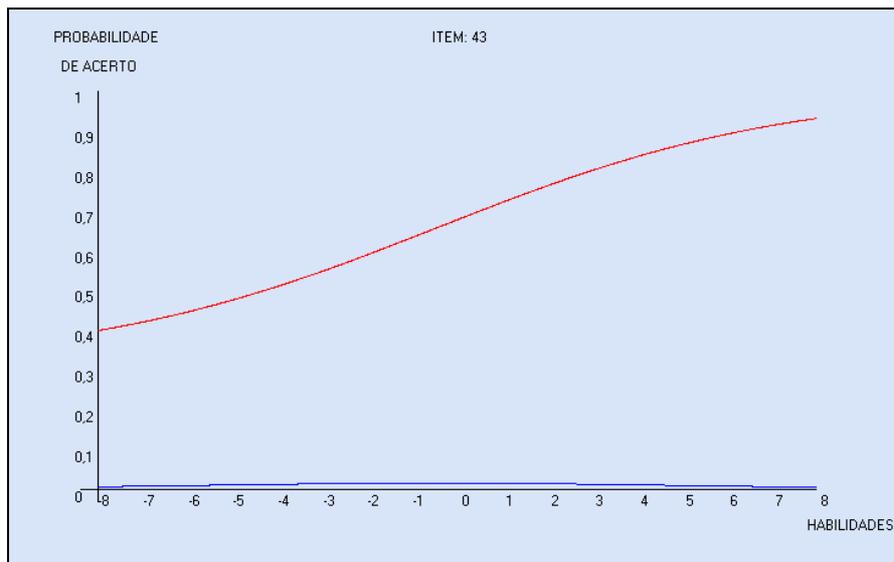
**FIGURA 21- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 36**

A Figura 21, mostra claramente que o item 36 não se ajustou ao modelo, pois apresenta a Curva Característica do Item e a Função de Informação do Item como retas, assim deve-se desconsiderar os resultados obtidos no ajuste dos seus parâmetros.



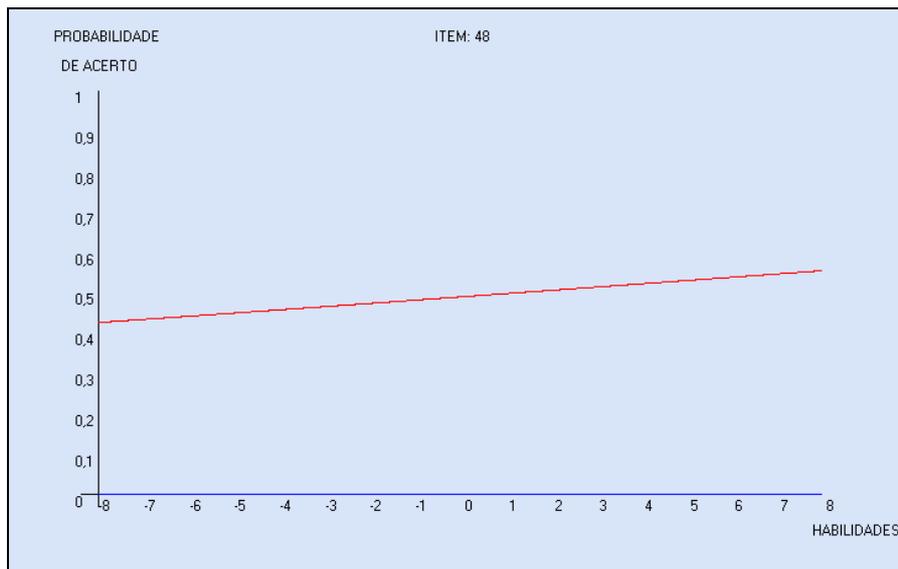
**FIGURA 22- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 37**

Apresentando parâmetro de discriminação igual a 0,4167, de dificuldade igual a 4,7241 e de acerto casual igual a 0,0508, nota-se na Figura 22 que o item 37 é muito difícil e apresenta pouca discriminação, portanto não é considerado um bom item.



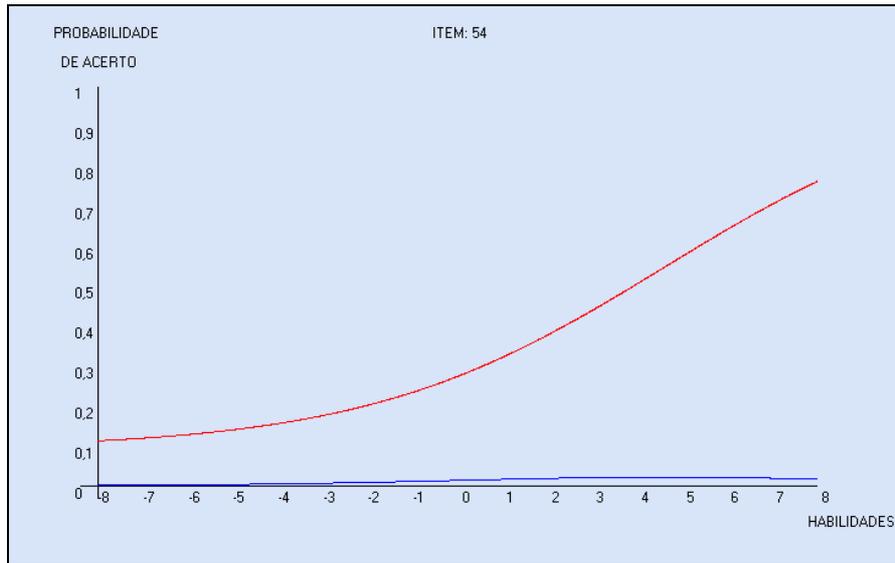
**FIGURA 23- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 43**

Apesar do item 43 ser um item relativamente fácil, dado que seu parâmetro de discriminação é igual a -1,5133, apresentou pouca discriminação (0,2515) com probabilidade de acerto casual alta (0,2153). Na Figura 23, observa-se que a Função de Informação do Item não apresenta o formato de uma distribuição aproximadamente normal, portando considera-se que o item pode prejudicar o julgamento dos candidatos.



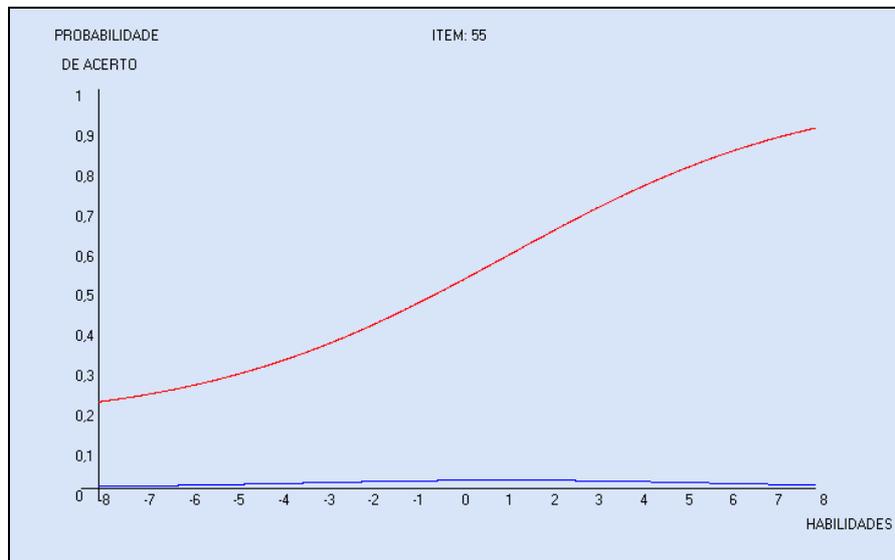
**FIGURA 24- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 48**

Os Gráficos apresentados na Figura 24 referente ao item 48 não apresentam bom comportamento e isso se verifica dado que esse item é difícil ( $b_{48} = 4,4643$ ) com pouca discriminação ( $a_{48} = 0,0343$ ) e possibilidade de acerto aleatório baixa ( $c_{48} = 0,0490$ ).



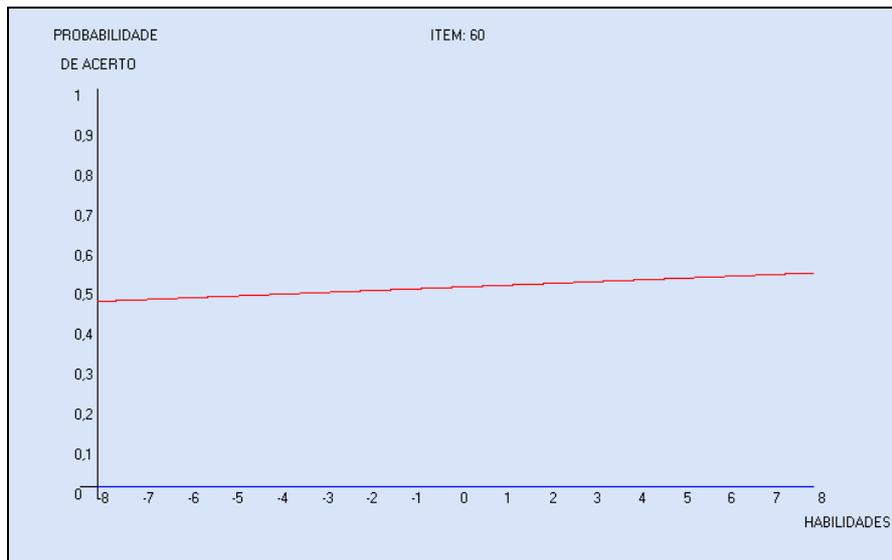
**FIGURA 25- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 54**

O item 54, caracterizado pelos gráficos da Figura 25, apresenta parâmetro de discriminação igual a 0,3018, de dificuldade igual a 4,6801 e de acerto casual de 0,0970. É um item difícil com discriminação baixa e possibilidade de acerto aleatório baixa, conseqüentemente não é um item considerado bom.



**FIGURA 26- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 55**

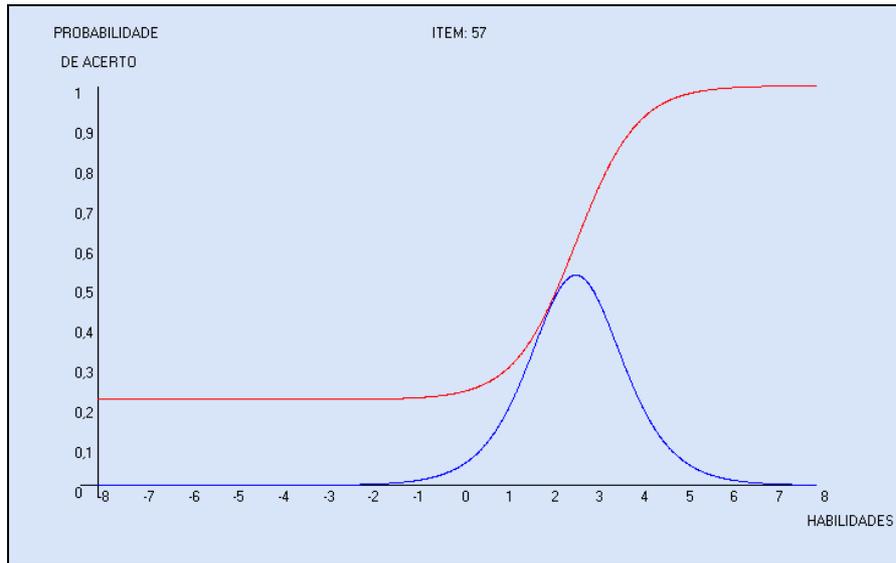
Apesar do item 55 possuir um parâmetro de dificuldade razoável ( $b_{55} = 1,4521$ ) é um item de pouca discriminação ( $a_{55} = 0,2895$ ), de modo que a sua respectiva Função de Informação não se ajustou corretamente a uma distribuição gaussiana, portanto que esse item afeta desfavoravelmente o instrumento.



**FIGURA 27- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 60**

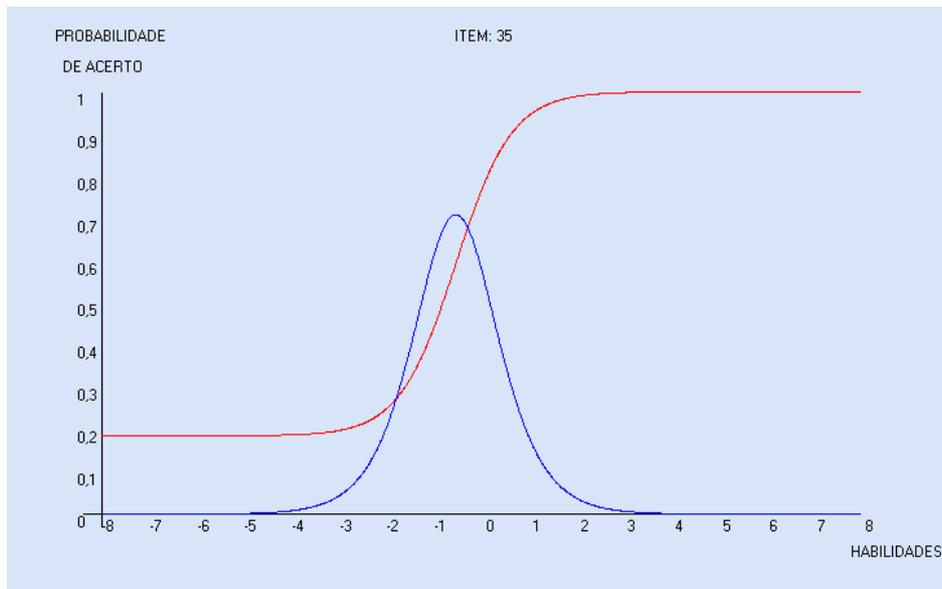
A Figura 27 representa a Curva Característica e Função de Informação do item 60. Observa-se que ele não se ajustou ao modelo, dado que não apresentou convergência, portanto deve ser retirado do instrumento para estimação do escore verdadeiro dos candidatos.

A título de exemplificação, do que seria um item de desempenho médio e bom no teste, apresenta-se a seguir as análises dos itens 57 e 35 acompanhadas seus dos gráficos.



**FIGURA 28- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 57**

O item 57 apresentou um parâmetro de discriminação 1,4515, de dificuldade igual a 2,62858 e acerto aleatório igual a 0,2146. Apesar apresentar boa discriminação o item é relativamente difícil e têm alto percentual de acerto aleatório, portanto o item possui médio desempenho.



**FIGURA 29- GRÁFICO DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DO ITEM 35**

O parâmetro de discriminação do item 35 possui um valor alto de 1,6889, média dificuldade (-0,5465) e acerto aleatório de 0,1861. Observando os gráficos da Figura 29, pode-se observar que a Curva Característica configura um curva logística perfeita e a Função Característica uma normal leptocúrtica significando que as habilidades são bem representadas pelo seu valor médio e o desvio padrão é pequeno, portanto esse item é considerado bom.

Para o cálculo das habilidades dos respondentes foram retirados os itens 13, 36, 37, 43, 48, 54, 55 e 60. A partir deste cálculo obtém-se os escores verdadeiros e foi feito o ranqueamento dos candidatos descrito na seção 4.1.3 a seguir.

#### 4.1.3- Ranqueamento dos Candidatos utilizando Escores Verdadeiros

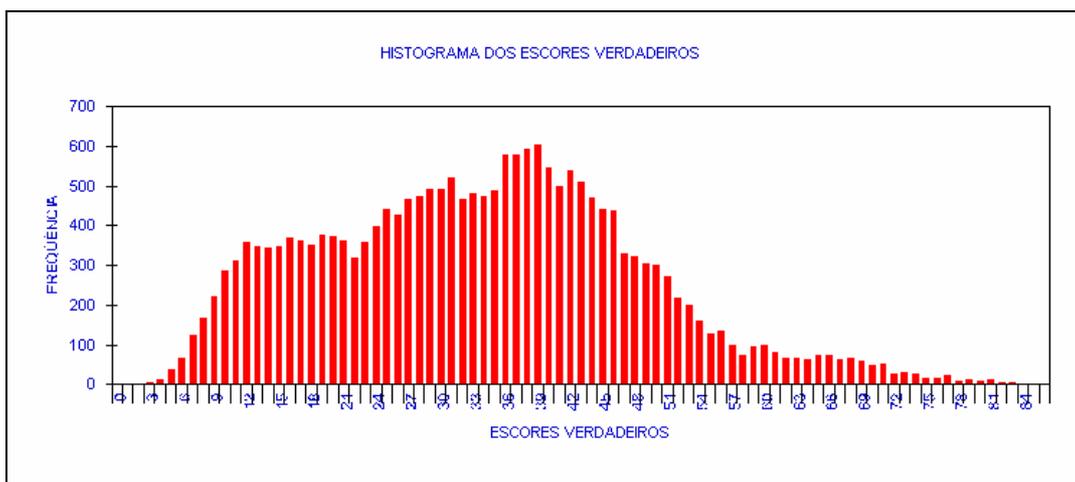
Os escores verdadeiros (colocados numa escala de 0 a 100) foram calculados utilizando os conceitos apresentados na seção 2.2.4, que consiste na soma dos itens certos de cada candidato considerando a probabilidade de acerto dos itens utilizando as habilidades obtidas pelo Método de Estimação Bayesiana da Média a *Posteriori* descrito na seção 2.2.3.4.3.3. O ranqueamento feito pelo programa SIAVAL 1.0 e parte dele é apresentado na Tabela 10 a seguir, dado que seria inviável colocar a classificação de todos os 20.550 candidatos.

**TABELA 10- RANQUEAMENTO DOS ESCORES VERDADEIROS**

RANQUEAMENTO	CANDIDATO	ESCORE VERDADEIRO
1	9730	85,67595
2	9923	85,60762
3	6665	85,44853
4	8393	84,22240
5	3773	83,56939
6	15646	83,06105
7	8821	82,94101
...	...	...
20543	18107	3,3166
20544	14347	3,1507
20545	15595	3,0599
20546	15158	2,9072
20547	13413	2,7333
20548	3729	0,9485
20549	9299	0,8884
20550	3089	0,0000

FONTE: RESULTADO DA ANÁLISE OBTIDA NO SIAVAL 1.0

O histograma dos escores verdadeiros feito no SIAVAL 1.0 pode ser visto na Figura 29 que segue.



FONTE: RESULTADOS DO SIAVAL 1.0

**FIGURA 30- DESCRIÇÃO DOS ESCORES VERDADEIROS**

Observa-se na Figura 30 que os dados encontram-se concentrados á esquerda o que indica que o escore médio é baixo. As estatísticas descritiva dos escores verdadeiros são apresentados a seguir na Tabela 11.

**TABELA 11- ESTATÍSTICAS DESCRITIVAS DOS ESCORES VERDADEIROS DOS CANDIDATOS**

<b>ESTATÍSTICA</b>	<b>ESCORES VERDADEIROS</b>
MÉDIA ARITMÉTICA	33,6204
DESVIO PADRÃO	14,6820
VALOR MÍNIMO	0
VALOR MÁXIMO	85,6759
1º. QUARTIL	22,4760
2º. QUARTIL	33,7776
3º. QUARTIL	43,1724

FONTE: RESULTADO SIAVAL 1.0

O valor máximo que poderia ser alcançado nos escores verdadeiros seria de 1, obteve-se, porém que o valor máximo alcançado no teste foi de 85,6759 e mínimo de 0. A média dos escores verdadeiros foi de 33,6204 e o desvio padrão de 14,6820. Quando aos quartis têm-se 25% dos candidatos tiveram escores abaixo de 22,4760, 50% abaixo de 33,7776 e apenas 25% acima de 43,1724.

#### 4.2- EVASÃO DE ALUNOS DO CURSO DE ESTATÍSTICA DA UFPR

No processo educacional o fenômeno de evasão é extremamente preocupante, tanto no ensino médio, quanto no ensino superior. Nas universidades brasileiras esse fenômeno atinge um índice alto nos cursos da área de Ciências Exatas. Isto ocorre em IES públicas e em IES privadas.

Segundo alguns pesquisadores do fenômeno evasão, a perspectiva de alcançar a formação superior se mantém presente na maioria dos evadidos (POLYDORO, 1995, 2000; AZZI, MERCURI e MORAN, 1996). Também, é grande o interesse da instituição em manter o aluno no curso, pois a evasão representaria uma perda social, de recursos e

de tempo de todos os envolvidos. A identificação de um aluno com chances de se evadir, logo que ele ingressa no curso, permite que a instituição promova estratégias garantindo a permanência do aluno.

Este trabalho demonstra que é possível detectar uma futura evasão a partir do término do primeiro semestre do curso frequentado. Para tanto, foi aplicado o Método Linear de Fisher para Várias Populações, como apresentado na seção 2.3.3.3.

A amostra estudada consistiu de 163 ex-alunos ingressos nos anos de 1999, 2000 e 2001, destes usou-se 155 ex-alunos para classificação e 6 para teste, Eles estavam agrupados da seguinte maneira: 92 não terminaram o curso (3 usados para teste), 56 terminaram o curso com atraso ( 3 usados para teste) e 15 terminaram no prazo certo (2 usados para teste),

As variáveis estudadas foram coletadas nos arquivos cedidos pelo Núcleo de Concurso da UFPR e nos diários de classe das disciplinas de Estatística Geral I, Cálculo de Probabilidade I e Cálculo com Geometria Analítica I, O conjunto das 45 variáveis (Apêndice 8) analisadas entraram variáveis sócio-econômicas, formação de grau médio, escores no vestibular, notas e frequência nas disciplinas citadas,

O desempenho do classificador é apresentado na tabela de classificação obtida do método linear de Fisher apresentado na Tabela 12, a seguir.

**TABELA 12- TABELA DE CLASSIFICAÇÃO FEITA PELO MÉTODO LINEAR DE FISHER**

Grupos de Referência	Grupos Classificados pelo Método			Total
	Abandonaram ou Trancaram o Curso	Não concluíram o Curso no Tempo Certo	Concluíram o Curso no Tempo Certo	
Abandonaram ou Trancaram o Curso	75(48,39%)	10(6,45%)	4(2,58%)	89
Não concluíram o Curso no Tempo Certo	6(3,87%)	40(25,81%)	7(4,52%)	53
Concluíram o Curso no Tempo certo	0(0,00%)	1(0,64%)	12(7,74%)	13
<b>Total</b>	81	53	21	<b>155</b>

Percentual de Classificação Correta: 81,94%

A soma dos valores obtidos na diagonal da tabela indica o total de ex-alunos do grupo de análise que foram classificados corretamente, Dividindo este valor pelo total (155) obteve-se um percentual de classificação correta 81,94% significando que as equações apresentadas subseqüentes classificam o grupo de teste com esta probabilidade de acerto,

$$Y_1 = -0,09202X_1 - 0,28175X_2 - 0,07758X_3 - 0,02018X_4 + 0,04222X_5 - 0,03868X_6 + 0,04829X_7 + 0,09409X_8 + 0,21174X_9 + 0,09102X_{10} + 0,03199X_{11} + 0,06741X_{12} + 0,04501X_{13} + 0,00802X_{14} - 0,03996X_{15} - 0,30428X_{16} + 0,20738X_{17} + 0,17139X_{18} - 0,16212X_{19} - 0,01622X_{20} - 0,37201X_{21} - 0,05958X_{22} + 0,15912X_{23} + 0,02403X_{24} + 0,04107X_{25} + 0,05589X_{26} - 0,09056X_{27} + 0,01441X_{28} - 0,12039X_{29} - 0,04309X_{30} - 0,00371X_{31} - 0,00019X_{32} - 0,00442X_{33} + 0,00046X_{34} - 0,00213X_{35} - 0,00063X_{36} - 0,00219X_{37} - 0,00011X_{38} - 0,00187X_{39} + 0,00082X_{40} - 0,00283X_{41} + 0,00191X_{42} - 0,01378X_{43} + 0,03926X_{44} + 0,00369X_{45}$$

$$Y_2 = + 0,4194X_1 + 0,32911X_2 - 0,12859X_3 + 0,06805X_4 + 0,10594X_5 - 0,05845X_6 + 0,08985X_7 - 0,18337X_8 + 0,12003X_9 + 0,04305X_{10} + 0,25476X_{11} + 0,0654X_{12} + 0,00266X_{13} - 0,01012X_{14} - 0,27692X_{15} - 0,01517X_{16} - 0,0928X_{17} + 0,11045X_{18} - 0,04492X_{19} + 0,0629X_{20} + 0,12851X_{21} - 0,22944X_{22} + 0,01865X_{23} - 0,0248X_{24} + 0,10266X_{25} + 0,06032X_{26} - 0,03355X_{27} - 0,18576X_{28} + 0,07316X_{29} + 0,01435X_{30} + 0,00068X_{31} - 0,00057X_{32} + 0,00041X_{33} - 0,00222X_{34} + 0,00055X_{35} + 0,0011X_{36} + 0,00104X_{37} + 0,00043X_{38} - 0,00168X_{39} + 0,00683X_{40} - 0,00162X_{41} - 0,00654X_{42} + 0,02439X_{43} - 0,00828X_{44} + 0,02536X_{45}$$

A Tabela 13 a seguir mostra o resultado alcançado com os 8 ex-alunos utilizados para teste, sendo que foi considerado grupo 1 os que evadiram(abandonaram ou trancaram o curso), grupo 2 os que não se formaram no tempo certo e grupo 3 os que concluíram o curso normalmente.

**TABELA 13- CLASSIFICAÇÃO DO GRUPO DE TESTE**

<b>Ex-Aluno</b>	<b>Grupo</b>	<b>Método de Fisher</b>
1	1	2
2	1	<b>1</b>
3	1	<b>1</b>
4	2	3
5	2	<b>2</b>
6	2	<b>2</b>
7	3	<b>3</b>
8	3	2
<b>% de Acerto</b>	-	62,5%

FONTE: Resultado da Análise feita no SIAVAL 1.0

Esta sistemática, de identificar alunos com alta probabilidade de evasão, pode ser muito útil à instituição, desde que se tenha uma política de acompanhamento e orientação desses alunos.

Para verificar a qualidade dos resultados encontrados no programa utilizou-se o programa STATGRAPHICS para comparar o método de Fisher com o método discriminante quadrático (com probabilidades *a priori* iguais e proporcionais ao grupo de classificação) e com redes neurais (com probabilidades *a priori* iguais e proporcionais ao grupo de classificação). A comparação dos resultados pode ser vista na Tabela 14 e no Apêndice 10 têm-se os resultados obtido no programa STATGRAPHICS.

**TABELA 14- COMPARAÇÃO ENTRE OS MÉTODOS DE FISHER, DISCRIMINANTE QUADRÁTICO E REDES NEURAIS.**

Percentual	Método				
	Discriminante de Fisher	Discriminante Quadrático (Prob.a priori iguais)	Discriminante Quadrático (Prob.a priori proporcionais ao grupo de classificação)	Redes Neurais (Prob.a priori iguais)	Redes Neurais (Prob.a priori proporcionais ao grupo de classificação)
Acertos no Grupo de Classificação	81,94%	81,94%	85,16%	54,19%	60,65%
Acertos no Grupo de Teste	62,50%	37,50%	37,50%	25,00%	50,00%

FONTE: RESULTADOS OBTIDOS NOS PROGRAMAS SIAVAL 1.0 E STATGRAPHICS

Os resultados apresentados na Tabela 14 indicam que o método de Fisher apresentou 81,94% acertos no grupo de classificação e 62,50% acertos no grupo de teste. O método discriminante quadrático com probabilidade *a priori* proporcionais ao grupo de classificação apesar de apontar melhor desempenho no grupo classificação (85,16%) não apresentou resultado razoável no grupo de teste (37,50%). Portanto, o método de Fisher foi uma boa escolha.

#### 4.3- ANÁLISE DE FORNECEDORES

A avaliação de fornecedores é uma atividade que vem preocupando as empresas industriais e de serviços, desde que foi intensificada a competição pelos mercados globalizados. De um modo geral, todas as empresas fazem avaliação dos seus fornecedores, seja informalmente com base na experiência do comprador ou formalmente e neste caso com base em um instrumento de avaliação cientificamente elaborado.

Essa atividade procura assegurar à empresa uma melhoria substancial no seu processo de produção, principalmente uma melhor qualidade dos insumos além da redução nos prazos de entrega. Um dos propósitos desse trabalho é facilitar a análise e interpretação de dados coletados por empresas na avaliação de seus fornecedores. O programa desenvolvido inclui técnicas de análise multivariada, tais como, análise fatorial (descritos na seção 2.2) que é usada no ranqueamento das empresas fornecedoras, e também, faz o reconhecimento e classificação de novas empresas, que venham a integrar a lista de fornecedores, utilizando o Método de Fisher para várias populações (grupos) apresentado na seção 2.3.3.3.

Trabalhou-se com dados reais de 52 fornecedores de uma empresa de grande porte que será chamada de Empresa X. Esta empresa possui procedimentos de ranqueamento e seleção de fornecedores que foram construídos ao longo de muitos anos de experiência. Os procedimentos feitos em um departamento da empresa abrangeram quatro características: suprimentos (1), qualidade (2), logística (3) e tecnologia (4). O conjunto de dados é inserido num sistema e a análise é feita de forma subjetiva.

No reconhecimento de padrões foram usadas as 52 empresas fornecedoras pesquisadas para estimar as funções discriminantes. A empresa forneceu os dados segundo seus critérios e após ser feito o ranqueamento dividiram-se os fornecedores em bom (parceiros), regular e ruim.

O programa desenvolvido, neste trabalho, está em forma genérica, ou seja, de tal modo que um usuário leigo possa fazer o ranqueamento e alocar novos fornecedores utilizando os dados obtidos no instrumento proposto pela empresa. A entrada dos dados pode ser feita por meio de uma planilha Excel.

No relatório de saída do ranqueamento é apresentada uma listagem ordenada dos fornecedores analisados com a classificação e o score obtido. Adiante, têm-se os detalhes numéricos. A classificação e alocação de novos fornecedores (empresas) foram feitas na mesma página onde são anexados os dados de entrada (em reconhecimento de padrões). Além disso, as funções lineares de classificação são apresentadas numa janela na mesma página e numa página a parte aparece a tabela de classificação.

A Tabela 15, adiante, resume os resultados obtidos com a aplicação do programa aos dados. Observa-se pelos valores das comunalidades que todas as variáveis do

questionário são importantes na análise. Três fatores resumem 92,77% das informações contidas na matriz de correlação. Porém apenas dois resumem 67,48% das informações e poderia, também, ser utilizado esse nível, isto é, utilizar os dois primeiros fatores. É interessante observar que no primeiro fator (F1) as variáveis qualidade e tecnologia são as predominantes, o que significa que essas variáveis são as mais importantes no desempenho dos fornecedores. O segundo fator (F2) é dominado pela variável suprimentos e o terceiro fator (F3) por logística.

Uma empresa fornecedora qualquer pode ter a sua avaliação quantitativa nos três fatores ou, isoladamente, em cada um dos fatores. Assim, pode-se classificar uma empresa de forma global nas três dimensões fatoriais ou em qualquer das três dimensões: “tecnologia e qualidade”, “suprimento” e “logística”. Essa informação é muito importante para a empresa manter ou substituir um fornecedor.

**TABELA 15- ANÁLISE FATORIAL – MÉTODO DAS COMPONENTES PRINCIPAIS**

Variáveis	Pesos (Carregamentos) Estimados			Comunalidades $h_i^2$	Variâncias Esp. $\Psi_i$
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>		
1. Suprimentos	0,120464	0,969454	0,102161	0,964790	0,035202
2. Qualidade	0,938372	-0,001173	0,143983	0,901274	0,098725
3. Logística	0,108815	0,093268	0,988694	0,998057	0,001942
4. Tecnologia	0,798869	0,453515	0,054077	0,846792	0,153207
<b>Autovalores (<math>\lambda_j</math>)</b>	1,545085	1,154219	1,011610		
<b>Proporção Acum.</b>	38,63%	67,48%	92,77%		

**Fonte:** Resultado da Análise

Os valores das comunalidades mostram que todas as variáveis são importantes na análise. Portanto, o instrumento de avaliação de fornecedores, construído pela Empresa X, merece confiança.

A classificação das 52 empresas fornecedoras analisadas foi baseada nos escores finais colocados na escala de 0 a 100 e o resultado está na Tabela 16, adiante.

**TABELA 16- CLASSIFICAÇÃO DAS EMPRESAS FORNECEDORAS**

Ranking	Fornecedor	Escore Final	Ranking	Fornecedor	Escore Final
1	16	93,783	27	41	80,118
2	5	93,462	28	37	79,285
3	12	92,736	29	43	77,635
4	25	92,647	30	35	76,981
5	17	92,064	31	13	74,891
6	50	92,043	32	38	74,097
7	45	91,934	33	26	73,600
8	20	91,765	34	46	73,560
9	14	90,992	35	21	73,143
10	24	90,713	36	44	71,899
11	1	90,293	37	49	69,394
12	27	89,977	38	47	69,196
13	39	89,915	39	19	68,597
14	33	89,509	40	4	67,474
15	52	89,497	41	51	66,940
16	10	88,810	42	6	66,664
17	32	88,704	43	30	60,395
18	2	87,982	44	7	55,026
19	34	87,396	45	22	54,023
20	31	87,232	46	3	53,697
21	8	86,738	47	42	52,122
22	9	85,275	48	29	50,302
23	18	85,039	49	48	47,783
24	40	84,564	50	36	34,720
25	15	82,768	51	11	31,515
26	28	80,311	52	23	23,988

Fonte: Resultados da Análise.

O ajuste do modelo de reconhecimento de padrões foi feito com base nas 52 empresas escolhidas classificadas em três grupos: Bom, Regular e Ruim. As equações obtidas do modelo linear de Fisher são apresentadas a seguir.

$$Y_1 = +0.06016 X_1 + 0.05165 X_2 + 0.07782 X_3 + 0.11007 X_4$$

$$Y_2 = +0.07989 X_1 - 0.01333 X_2 + 0.0233 X_3 - 0.11395 X_4$$

Considerou-se um fornecedor do grupo 1 como bom quando seu escore ficou acima de 85, grupo 2 como regular entre 85 e 70 e grupo 3 como ruim abaixo de 70. O programa SIAVAL 1.0 forneceu a tabela de classificação (Tabela 17) que mostra os acertos e erros de classificação dos dados analisados. Esta classificação foi feita utilizando o Procedimento de Retenção de Lachenbruch descrito na seção 2.3.5. Deve-se observar que os valores da diagonal são o das observações onde foi classificada corretamente.

**TABELA 17- TABELA DE CLASSIFICAÇÃO USANDO O PROCEDIMENTO DE RETENÇÃO DE LACHENBRUCH**

Grupos de Referência	Grupos Classificados pelo Método			Total
	1- Bom	2- Regular	3- Regular	
<b>1- Bom</b>	<b>22(42,31%)</b>	1(1,92%)	0(0,00%)	23
<b>2- Regular</b>	2(3,85%)	<b>11(21,15%)</b>	0(0,00%)	13
<b>3- Regular</b>	1(1,92%)	6(11,54%)	<b>9(17,31%)</b>	16
<b>Total</b>	25	18	9	<b>52</b>

FONTE: RESULTADO OBTIDO NO SIAVAL 1.0

Totalizando os dados observados na diagonal e dividindo pelo total de observações obtém-se o percentual de acerto de classificação que resultou um percentual de 80,77%, significando que o classificador estatisticamente é bom.

O programa pode ser uma ferramenta muito importante na análise e ranqueamento de fornecedores, principalmente para as empresas que venham iniciar a análise, pois bastaria entrar com dados que a empresa julga importante no processo e automaticamente obteria o ranqueamento dos fornecedores. Além disso, pode-se obter a classificação (por exemplo: bom, regular e ruim) de um fornecedor que tenha recentemente sido incluído no quadro de fornecedores da empresa. Outros fatores importantes a serem considerados são: o tempo para análise e a não existência de experiência do aplicador do método.

## 5. CONCLUSÃO E RECOMENDAÇÕES

Algumas aplicações inéditas foram desenvolvidas. Construiu-se, neste trabalho, um sistema integrado para o ranqueamento de fornecedores feito por Análise Fatorial e que pode facilitar muito as empresas que já fazem ranqueamento. No caso de empresas que desejam iniciar a avaliação de fornecedores o programa poderia ajudar na escolha das variáveis importantes de um instrumento feito pela empresa.

Outra aplicação importante é o procedimento de detectar alunos com possibilidade de evasão. Muitas instituições de ensino e órgãos governamentais buscam as causas de evasão para poder enfrentá-las, porém no procedimento apresentado neste trabalho, indicou-se um meio de detectar a evasão antes que ela aconteça e assim seja possível trabalhar somente com os alunos que possuem esta possibilidade. O fator mais importante do procedimento é que a análise pode ser feita sem que o aluno perceba que está sendo avaliado neste item, pois a instituição pode obter as variáveis de interesse na sua entrada na instituição, como por exemplo com os dados sócio-econômicos do vestibular e ainda através de notas e frequências nas disciplinas do primeiro semestre.

Quando se escolheu a análise de itens de uma prova de vestibular pretendeu-se incentivar instituições de ensino que analisassem os itens de suas provas e que o ranqueamento dos candidatos pudessem ser feitos utilizando escores verdadeiros. Este seria o meio mais justo de avaliar candidatos, pois itens ruins não entrariam no escore ficando somente os itens bem formulados.

Este trabalho forneceu como produto o programa SIAVAL 1.0 que é uma ferramenta muito útil para a análise de desempenho, várias áreas do conhecimento fazem este tipo de análise e não utilizam as estatísticas utilizadas no programa, principalmente no Brasil.

Como o programa possui uma comunicação facilitada com o usuário, como, por exemplo, entrada e saída de dados através de uma planilha Excel, espera-se que a comunidade de pesquisadores e analistas de desempenho tenha maior acesso a estas estatísticas e possa melhor analisar seus dados, visto que os programas encontrados no mercado são de língua estrangeira, caros e de difícil manuseio.

No entanto, o programa desenvolvido atende as pesquisas de desempenho no que se refere à Análise de Itens, visto que está operando somente com itens que possuam respostas dicotômicas ou possíveis de serem dicotomizadas. Portanto, para uma versão 2.0 poderia ser programada com o Modelo Nominal na Teoria da Resposta ao Item, de modo que o programa se tornaria mais generalista.

Verificou-se também a necessidade de construir um programa de uso nacional que contivesse todas os modelos da Teoria da Resposta ao Item com intuito de que a análise de desempenho escolar pudesse ser feita por instituições de ensino, municípios, regiões, etc.

Entende-se, portanto, que os objetivos do trabalho foram alcançados, mostrando a viabilidade e aplicações das metodologias apresentadas.

## REFERÊNCIAS

- ALEXANDRE, J.W.C., ANDRADE, D.F., VASCONCELOS, A.P. e ARAÚJO, A.M.S. Uma Proposta de Análise de um Construto para a Medição dos Fatores Críticos da Gestão pela Qualidade através da Teoria da Resposta ao Item. **Gestão & Produção**, v.9, n.2, p.129-141, 2002.
- ALVAREZ, M. P. & QUEIROZ, A. A. Aproximações dos Laços de Parcerias entre Fornecedor-Cliente na Cadeia de Suprimentos como Fonte de Competitividade. **XXIII ENEGEP**, Ouro Preto, outubro 2003.
- ANDERSEN, Erling B. & OLSEN, Lin W. *The Life of Georg Rasch as a Mathematician and as a Statistician. Essays on Item Response Theory*. New York: Lecture Notes in Statistics, p.3-24, 2001.
- ANDRADE, Dalton F.; TAVARES Héilton R. & VALLE, Raquel C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. Caxambu-MG: 14<sup>a</sup>. SINAPE-ABE, 2000.
- AZZI, R. G.; MERCURI, E.; MORAN, R. C. **Fatores que interferem na decisão de desistência de curso no primeiro ano de graduação**. Campinas (Mimeogr.), 1996.
- BAKER, Frank B. & KIM, Seock-Ho. *Item Response Theory: Parameter Estimation Techniques*. 2 ed. Lisboa: Marcel Dekker, 2004.
- BOCK, R. Darrell & AITKIN, Murray. *Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm*. **Psychometrika**, v. 46, n. 4, 443-459, December 1981
- BROWN, J. Dean. *Point-biserial correlation coefficients*. **JALT Testing & Evaluation SIG Newsletter**, v. 5, n. 3, 12-15, October 2001
- BURDEN, R. L. & FAIRES, J.D. *Numerical Analysis*. 6 ed. California: Books/Cole Publishing Company, 1997
- COSTA, M.B.F. **Técnica Derivada da Teoria da Resposta ao Item Aplicada ao Setor de Serviços**. Curitiba: Dissertação (Mestrado em Métodos Numéricos em Engenharia), 2001 – Setor de Tecnologia e Ciências Exatas, Universidade Federal do Paraná.

- CHAVES NETO, A. **Análise Multivariada Aplicada à Pesquisa**. Curitiba: UFPR, 2004, 74 p. (notas de aula).
- CHAVES NETO, A. **Introdução ao Reconhecimento de Padrões**. Curitiba: UFPR, 2003, 61 p. (notas de aula).
- CHAVES NETO, A. & TURIM M. E. **Introdução às Teorias da Avaliação**. Minicurso, Semana da Matemática UNICENTRO. Guarapuava, 2003.
- DARLINGTON RICHARD B. **Factor Analysis**. Disponível em <<http://comp9.psych.cornell.edu/Darlington/factor.htm>>. Acesso em setembro de 2004.
- DeROOS, Y., ALLEN-MEARES, P. *Application of the Rasch Analysis: exploring differences in depression between african-american and white children*. **Journal of Social Service Research**, v. 23, n. 3 e 4, 93-107, 1998.
- DOWNIE, N. M. & HEATH, R. W. (1959) – **Basic Statistical Methods**, Harper & Brothers, Publishers, N.Y.
- ELECTRONIC STATISTICS TEXTBOOK. **Principal Components and Factor Analysis**. Disponível em <<http://www.statsoft.com/textbook/stfacan.html>>. Acesso feito em abril de 2005
- FURTADO, E. M. **Automação do Ranqueamento Qualitativo de Áreas Especialmente Protegidas do Estado do Paraná através da Análise Fatorial**. Curitiba: Dissertação (Mestrado em Métodos Numéricos em Engenharia), 1999 – Setor de Tecnologia e Ciências Exatas, Universidade Federal do Paraná.
- GARSON DAVE . **Factor Analysis**. Disponível em <<http://www2.chass.ncsu.edu/garson/pa765/factor.htm>>. Acesso feito em janeiro de 2005.
- GUILFORD, J. P. **Fundamental statistics in psychology and education**. 4.ed. New York: McGraw-hill Book, 1950. 605p.
- HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L. & BLACK, W. C. **Análise Multivariada de Dados**. ; Tradução: Adonay S.Sant'anna e Anselmo Chaves Neto. 5 ed Porto Alegre: Bookman, 2005.
- HAMBLETON, Ronald K.; SWAMINATHAN H. & Rogers, H. Jane. **Fundamentals of Item Response Theory**. USA: Sage Publications, 1991.
- HISTORY OF ITEM RESPONSE THEORY <<http://www.uic.edu/classes/ot/ot540/history.html>> . Acesso feito em abril de 2005.

- HWANG, Dae-Yeop. *Classical Test Theory and Item Response Theory: Analytical and Empirical Comparisons*. Annual Meeting of Southwest Educational Research Association, 27 p.,2002.
- JOHNSON R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*.2.ed. New Jersey: Prentice Hall International,1992.
- LINDEN, Wim J. van der & HAMPLETON, Ronald K. *Handbook of Modern Item Response Theory*. USA: Springer, 1996.
- LORD, Frederic M. & NOVIK, Melvin R. *Statistical Theories of Mental Test Scores*. USA e Canadá: Addison-Wesley Publishing Company, 1968.
- MARQUES, J. M. & LIMA, J. D. A Estatística Multivariada na Análise Econômica-Financeira de Empresas. **Revista da FAE**, v. 5, n. 3, 51-59, Setembro 2002.
- MOOD, Alexander M.; GRAYBILL, Franklin A. & BOES, Duane C. *Introduction to The Theory of Statistics*. International Student Edition. USA: McGraw-Hill, 1974.
- MÜLLER, S.I.M.G. **Comparação entre os Métodos de Máxima Verossimilhança, Distância Mínima e Método de Fisher para Reconhecimento de Padrão em Imagens Coloridas**. Curitiba: Dissertação (Mestrado em Ciências Geodésicas), 1997-Setor de Tecnologia, Universidade Federal do Paraná.
- POLYDORO, S. A. J. **Evasão em uma instituição de ensino superior: desafio para a psicologia escolar** (Dissertação de Mestrado). Campinas, SP: PUCAMP, 1995.
- POLYDORO, S. A. J. **O Trancamento de matrícula na trajetória acadêmica do universitário: condições de saída e de retorno à instituição** (Tese de Doutorado). Campinas, SP: UNICAMP, 2000
- SOARES, T. Machado. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa Operacional**, v. 25, n. 1, 83-112. Rio de Janeiro, Jan./Abr. 2005.
- SWAMINATHAN, Hariharan; HAMBLETON, Ronald K.; SIRECI, Stephen G.; XING Dehui & RIZAVI, Saba M. *Small Sample Estimation in Dichotomous Item Response Models: effect of priors based on judgmental information on the accuracy of item parameter estimates*. **Applied Psychological Measurement**, v. 27, n. 1, 27-51, January 2003.

TAVARES, H.R., ANDRADE, D.F. and PEREIRA, C.A.B. *Detection of Determinant Genes and Diagnostic via Item Response Theory. **Genetics and Molecular Biology***,v 27, n.4, 679-685, 2004. TAVARES, H. R. **Teoria da Resposta ao Item para Dados Longitudinais**. São Paulo, 2001 Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística –Universidade de São Paulo.

WIKIPEDIA. **Factor analysis**.< [http://en.wikipedia.org/wiki/Factor\\_analysis](http://en.wikipedia.org/wiki/Factor_analysis) >. Acesso feito em Abril de 2005.

WONNACOTT, Ronald J.; WONNACOTT, Thomas H. *Econometria*. 2. ed. São Paulo: Livros Técnicos e Científicos, 1978. 424 p.

## **APÊNDICES**

## APÊNDICE 1

### DERIVADAS DO ML-3 EM RELAÇÃO AO PARÂMETRO DE DISCRIMINAÇÃO

$$P_{ki} = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}}$$

$$P_{ki}^* = \frac{1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}}$$

$$Q_{ki}^* = 1 - \frac{1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}} = \frac{1 + e^{-Da_i(\bar{\theta}_k - b_i)} - 1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}} = \frac{e^{-Da_i(\bar{\theta}_k - b_i)}}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}}$$

$$\frac{\partial P_{ki}}{\partial a_i} = \frac{D(1 - c_i)(\bar{\theta}_k - b_i)e^{-Da_i(\bar{\theta}_k - b_i)}}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2} = D(1 - c_i)(\bar{\theta}_k - b_i)P_{ki}^*Q_{ki}^*$$

$$\frac{\partial^2 P_{ki}}{\partial a_i^2} = D(1 - c_i)(\bar{\theta}_k - b_i) \frac{\partial P_{ki}^*Q_{ki}^*}{\partial a_i}$$

onde

$$\begin{aligned} \frac{\partial P_{ki}^*Q_{ki}^*}{\partial a_i} &= \frac{\partial}{\partial a_i} \left[ \frac{e^{-Da_i(\bar{\theta}_k - b_i)}}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2} \right] = \frac{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2 (-D)(\bar{\theta}_k - b_i)e^{-Da_i(\bar{\theta}_k - b_i)} -}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^4} \\ &\quad - \frac{e^{-Da_i(\bar{\theta}_k - b_i)} [(-2D)(\bar{\theta}_k - b_i)e^{-Da_i(\bar{\theta}_k - b_i)} + (-2D)(\bar{\theta}_k - b_i)e^{-2Da_i(\bar{\theta}_k - b_i)}]}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^4} = \\ &= \frac{(-D)(\bar{\theta}_k - b_i)e^{-Da_i(\bar{\theta}_k - b_i)}}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2} + \frac{2D(\bar{\theta}_k - b_i)e^{-Da_i(\bar{\theta}_k - b_i)}(e^{-Da_i(\bar{\theta}_k - b_i)} + e^{-2Da_i(\bar{\theta}_k - b_i)})}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^4} = \end{aligned}$$

$$\begin{aligned} &= -D(\bar{\theta}_k - b_i)Q_{ki}^*P_{ki}^* + 2D(\bar{\theta}_k - b_i)Q_{ki}^*(Q_{ki}^*P_{ki}^{2*} + Q_{ki}^{2*}P_{ki}^*) = \\ &= D(\bar{\theta}_k - b_i)Q_{ki}^*P_{ki}^*(-1 + 2Q_{ki}^*P_{ki}^* + 2Q_{ki}^{2*}) = \\ &= D(\bar{\theta}_k - b_i)Q_{ki}^*P_{ki}^*[-1 + 2(1 - P_{ki}^*)P_{ki}^* + 2(1 - 2P_{ki}^* + P_{ki}^{2*})] = \\ &= D(\bar{\theta}_k - b_i)Q_{ki}^*P_{ki}^*(1 - 2P_{ki}^*) \end{aligned}$$

então:

$$\frac{\partial^2 P_{ki}}{\partial a_i^2} = D^2(1 - c_i)(\bar{\theta}_k - b_i)^2 P_{ki}^*Q_{ki}^*(1 - 2P_{ki}^*)$$

## APÊNDICE 2

### DERIVADAS DO MODELO ML-3 EM RELAÇÃO AOS PARÂMETROS DE DIFICULDADE E ACERTO CASUAL

$$P_{ki} = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}}$$

$$P_{ki}^* = \frac{1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}}$$

$$\frac{\partial P_{ki}}{\partial b_i} = \frac{-Da_i(1 - c_i)e^{-Da_i(\bar{\theta}_k - b_i)}}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2} = -Da_i(1 - c_i)P_{ki}^*Q_{ki}^*$$

$$\frac{\partial^2 P_{ki}}{\partial b_i^2} = -Da_i(1 - c_i) \frac{\partial P_{ki}^*Q_{ki}^*}{\partial b_i}$$

onde

$$\begin{aligned} \frac{\partial P_{ki}^*Q_{ki}^*}{\partial b_i} &= \frac{\partial}{\partial b_i} \left[ \frac{e^{-Da_i(\bar{\theta}_k - b_i)}}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2} \right] = \frac{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2 Da_i e^{-Da_i(\bar{\theta}_k - b_i)} -}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^4} \\ &\quad - \frac{e^{-Da_i(\bar{\theta}_k - b_i)} [2Da_i e^{-Da_i(\bar{\theta}_k - b_i)} + 2Da_i e^{-2Da_i(\bar{\theta}_k - b_i)}]}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^4} = \\ &= \frac{Da_i e^{-Da_i(\bar{\theta}_k - b_i)}}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^2} - \frac{2Da_i e^{-Da_i(\bar{\theta}_k - b_i)} (e^{-Da_i(\bar{\theta}_k - b_i)} + e^{-2Da_i(\bar{\theta}_k - b_i)})}{[1 + e^{-Da_i(\bar{\theta}_k - b_i)}]^4} = \\ &= Da_i Q_{ki}^* P_{ki}^* - 2Da_i Q_{ki}^* (Q_{ki}^* P_{ki}^{2*} + Q_{ki}^{2*} P_{ki}^*) = \\ &= Da_i Q_{ki}^* P_{ki}^* [1 - 2Q_{ki}^* P_{ki}^* - 2Q_{ki}^{2*}] = \\ &= Da_i Q_{ki}^* P_{ki}^* [1 - 2(1 - P_{ki}^*)P_{ki}^* - 2(1 - 2P_{ki}^* + P_{ki}^{2*})] = \\ &= Da_i Q_{ki}^* P_{ki}^* (-1 + 2P_{ki}^*) \end{aligned}$$

então:

$$\frac{\partial^2 P_{ki}}{\partial^2 b_i} = -D^2 a_i^2 (1 - c_i) P_{ki}^* Q_{ki}^* (-1 + 2P_{ki}^*) = D^2 a_i^2 (1 - c_i) P_{ki}^* Q_{ki}^* (1 - 2P_{ki}^*)$$

$$\frac{\partial P_{ki}}{\partial c_i} = 1 - \frac{1}{1 + e^{-Da_i(\bar{\theta}_k - b_i)}} = 1 - P_{ki}^* = Q_{ki}^*$$

$$\frac{\partial^2 P_{ki}}{\partial^2 c_i} = 0$$

### APÊNDICE 3

#### DERIVADA DA FUNÇÃO DE VEROSSIMILHANÇA EM RELAÇÃO AOS PARÂMETROS DE HABILIDADE

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

fazendo,

$$\frac{\partial P_{ji}}{\partial \theta_j} = (1 - c_i) \frac{Da_i e^{-Da_i(\theta_j - b_i)}}{(1 + e^{-Da_i(\theta_j - b_i)})^2} = Da_i (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \frac{e^{-Da_i(\theta_j - b_i)}}{1 + e^{-Da_i(\theta_j - b_i)}} = Da_i (1 - c_i) P_{ji}^* Q_{ji}^*$$

então temos que:

$$\frac{\partial \log L(\theta)}{\partial \theta_j} = \sum_{i=1}^I \left[ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^* Q_{ji}^*} \right] Da_i (1 - c_i) P_{ji}^* Q_{ji}^*$$

e finalmente

$$\frac{\partial \log L(\theta)}{\partial \theta_j} = D \sum_{i=1}^I [a_i (1 - c_i) (u_{ji} - P_{ji}) W_{ji}]$$

**APÊNDICE 4**  
**PARÂMETROS DA NORMAL MULTIVARIADA PARA COMBINAÇÕES**  
**LINEARES**

Resultado:

Se  $\underline{X}$  tem distribuição  $N_p(\underline{\mu}, \Sigma)$ , as  $q$  combinações lineares

$$\underset{(q \times p)}{A} \underset{(p \times 1)}{\underline{X}} = \begin{bmatrix} a_{11}X_1 + \dots + a_{1p}X_p \\ a_{21}X_1 + \dots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \dots + a_{qp}X_p \end{bmatrix}$$

têm distribuição  $N_q(A\underline{\mu}, A\Sigma A')$ . Além disso, se  $\underline{d}$  é um vetor de constantes, então

$$\underset{(p \times 1)}{\underline{X}} + \underset{(p \times 1)}{\underline{d}} \text{ tem distribuição } N_p(\underline{\mu} + \underline{d}, \Sigma).$$

Prova:

Dadas as combinações lineares  $\underline{Z} = A\underline{X}$  têm

$$\underline{\mu}_Z = E(\underline{Z}) = E(A\underline{X}) = A\underline{\mu}_X$$

e

$$\Sigma_Z = \text{Cov}(\underline{Z}) = \text{Cov}(A\underline{X}) = A\Sigma_X A'.$$

Qualquer combinação linear  $\underline{b}'(A\underline{X})$  é uma combinação linear de  $\underline{X}$ , da forma  $\underline{a}'\underline{X}$  com  $\underline{a} = A'\underline{b}$ . Se  $\underline{X}$  é normalmente distribuído com  $N_p(\underline{\mu}, \Sigma)$ , então qualquer combinação linear das variáveis  $\underline{a}'\underline{X} = a_1X_1 + a_2X_2 + \dots + a_pX_p$  tem distribuição  $N(\underline{a}'\underline{\mu}, \underline{a}'\Sigma \underline{a})$ . Além disso, se  $\underline{a}'\underline{X}$  tem distribuição  $N_p(\underline{a}'\underline{\mu}, \underline{a}'\Sigma \underline{a}) \forall \underline{a} \in \mathbb{R}^p$ , então  $\underline{X}$  tem distribuição  $N_p(\underline{\mu}, \Sigma)$ . Desta forma, as conclusões de  $A\underline{X}$  estão satisfeitas.

A segunda parte da prova pode ser obtida considerando  $\underline{a}'(\underline{X} + \underline{d}) = \underline{a}'\underline{X} + (\underline{a}'\underline{d})$ , onde  $\underline{a}'\underline{X}$  tem distribuição  $N(\underline{a}'\underline{\mu}, \underline{a}'\Sigma \underline{a})$ . Sabe-se do caso univariado que a adição de uma constante  $\underline{a}'\underline{d}$  à variável  $\underline{a}'\underline{X}$  translada a média para  $\underline{a}'\underline{\mu} + \underline{a}'\underline{d} = \underline{a}'(\underline{\mu} + \underline{d})$ , sem alterar a variância. Já que  $\underline{a}$  é arbitrário,  $\underline{X} + \underline{d}$  tem distribuição  $N_p(\underline{\mu} + \underline{d}, \Sigma)$ .

## APÊNDICE 5

### PARÂMETROS DA DISTRIBUIÇÃO NORMAL MULTIVARIADA DA VARIÁVEL CONDICIONADA

Resultado:

$$\text{Seja } \underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_2 \end{bmatrix} \text{ com distribuição } N_p(\underline{\mu}, \Sigma) \text{ na qual } \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \ddots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix}, \text{ e}$$

$|\Sigma_{22}| > 0$ . Então a distribuição condicional de  $X_1$ , dado  $X_2 = x_2$ , é normal com

$$\text{Média} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \quad \text{e}$$

$$\text{Covariância} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Assim, a covariância independe do valor de  $x_2$  da variável condicionada.

Prova:

$$\text{Seja } \mathbf{A} = \begin{bmatrix} \mathbf{I} & \vdots & -\Sigma_{12}\Sigma_{22}^{-1} \\ \text{(qxq)} & \ddots & \text{qx(p-q)} \\ \cdots & \vdots & \cdots \\ \mathbf{0} & \vdots & \mathbf{I} \\ \text{(p-q)xq} & \ddots & \text{(p-q)x(p-q)} \end{bmatrix}$$

$$A(\underline{X} - \underline{\mu}) = A \begin{bmatrix} X_1 - \mu_1 \\ \cdots \\ X_2 - \mu_2 \end{bmatrix} = \begin{bmatrix} X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \\ \cdots \\ X_2 - \mu_2 \end{bmatrix}$$

é normal conjuntamente com matriz de covariância  $A\Sigma A'$  dada por

$$\begin{bmatrix} \mathbf{I} & \vdots & -\Sigma_{12}\Sigma_{22}^{-1} \\ \cdots & \ddots & \cdots \\ \mathbf{0} & \vdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \ddots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \vdots & \mathbf{0}' \\ \cdots & \ddots & \cdots \\ (-\Sigma_{12}\Sigma_{22}^{-1}) & \vdots & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \vdots & \mathbf{0}' \\ \cdots & \ddots & \cdots \\ \mathbf{0} & \vdots & \Sigma_{22} \end{bmatrix}$$

Como  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$  e  $X_2 - \mu_2$  tem covariância zero, são independentes. Além disso,  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$  tem distribuição  $N_q(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ . Dado  $X_2 = \mu_2$ ,  $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$  é uma constante. A distribuição condicionada de  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$  é a mesma da distribuição não condicionada  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$ , pois  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$  e  $X_2 - \mu_2$  são independentes. Já que a quantidade  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$  tem

distribuição  $N_q(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ , ela é o vetor aleatório  $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ , no qual  $X_2$  assume o valor particular  $x_2$ . De forma equivalente, dado que  $X_2 = x_2$ ,  $X_1$  tem distribuição  $N_q(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ .

## APÊNDICE 6

### RESULTADO USADO NO MÉTODO DE REGRESSÃO PARA DETERMINAR OS ESCORES FATORIAIS

Resultado:

Supondo não singulares as matrizes e possíveis as operações, então

$$(I + L'\Psi^{-1}L)^{-1} L'\Psi^{-1}L = I - (I + L'\Psi^{-1}L)^{-1}$$

Prova:

Partindo da igualdade entre matrizes identidades de mesma ordem

$$I = I$$

multiplicando membro a membro por  $L'\Psi^{-1}L$ , vem

$$L'\Psi^{-1}L = L'\Psi^{-1}L$$

somando e subtraindo a identidade ao segundo membro

$$L'\Psi^{-1}L = I + L'\Psi^{-1}L - I$$

$$L'\Psi^{-1}L = (I + L'\Psi^{-1}L) - (I + L'\Psi^{-1}L).(I + L'\Psi^{-1}L)^{-1}$$

fatorando a expressão :

$$L'\Psi^{-1}L = (I + L'\Psi^{-1}L). [ I - (I + L'\Psi^{-1}L)^{-1} ]$$

pré-multiplicando membro a membro por  $(I + L'\Psi^{-1}L)^{-1}$ , tem-se

$$(I + L'\Psi^{-1}L)^{-1}(L'\Psi^{-1}L) = (I + L'\Psi^{-1}L)^{-1}. (I + L'\Psi^{-1}L). [ I - (I + L'\Psi^{-1}L)^{-1} ]$$

como  $(I + L'\Psi^{-1}L)^{-1}. (I + L'\Psi^{-1}L) = I$ , pode-se escrever

$$(I + L'\Psi^{-1}L)^{-1}(L'\Psi^{-1}L) = I - (I + L'\Psi^{-1}L)^{-1}$$

## APÊNDICE 7

### RESULTADO USADO NO MÉTODO DE REGRESSÃO PARA DETERMINAR OS ESCORES FATORIAIS

Resultado:

Supondo não singulares as matrizes e possíveis às operações, então:

$$(LL' + \Psi)^{-1} = \Psi^{-1} - \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}$$

Prova:

Partindo da igualdade entre matrizes identidades de mesma ordem  $I = I$  somando e subtraindo expressões ao segundo membro, tem-se:

$$I = I + \Psi^{-1}LL' - \Psi^{-1}LL' + \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L' - \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L'$$

$$I = I + \Psi^{-1}LL' - \Psi^{-1}L.[I - (I + L'\Psi^{-1}L)^{-1}]L' - \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L'$$

e substituindo  $I = \Psi^{-1}\Psi$ , tem-se

$$I = \Psi^{-1}\Psi + \Psi^{-1}LL' - \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}LL' - \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L'$$

Fatorando

$$I = \Psi^{-1} (LL' + \Psi) - \Psi^{-1}L. [(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}LL' + (I + L'\Psi^{-1}L)^{-1}L']$$

Fatorando novamente temos:

$$I = \Psi^{-1} (LL' + \Psi) - \Psi^{-1}L. [(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}](LL' + \Psi)$$

Após multiplicar membro a membro por  $(LL' + \Psi)^{-1}$ , tem-se

$$(LL' + \Psi)^{-1} = \Psi^{-1} - \Psi^{-1}L.(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}$$

## APÊNDICE 8

### VARIÁVEIS UTILIZADAS NA ANÁLISE DE EVASÃO DE ALUNOS DO CURSO DE ESTATÍSTICA DA UFPR

1- GENERO: 1-Masculino; 2- Feminino

|1| Masculino

|2| Feminino

2- ESTADO CIVIL:

|1| Solteiro(a)

|2| Casado(a)

|3| Outro

3- SITUAÇÃO DE MORADIA:

|1| Mora em casa dos pais, quitada ou financiada

|2| Mora em casa dos pais, alugada

|3| Mora em casa própria, quitada ou financiada

|4| Mora em casa alugada, paga por você

|5| Mora em república, casa de estudante, pensão ou pensionato

|6| Mora em casa de parentes ou amigos

|7| Mora em casa alugada para você, paga por seus pais

4- NÍVEL DE INSTRUÇÃO DO PAI:

|1| Sem escolaridade

|2| Ensino fundamental incompleto

|3| Ensino fundamental completo

|4| Ensino médio incompleto

|5| Ensino médio completo

|6| Superior incompleto

|7| Superior completo

|8| Não sei informar

5- NÍVEL DE INSTRUÇÃO DA MÃE:

- |1| Sem escolaridade
- |2| Ensino fundamental incompleto
- |3| Ensino fundamental completo
- |4| Ensino médio incompleto
- |5| Ensino médio completo
- |6| Superior incompleto
- |7| Superior completo
- |8| Não sei informar

6- OCUPAÇÃO DO PAI:

- |1| Funcionário público do governo Federal, Estadual ou Municipal
- |2| Empregado de empresa
- |3| Sócio ou proprietário de empresa
- |4| Trabalho remunerado por conta própria, com auxílio de parentes e/ou de familiares
- |5| Trabalho remunerado por conta própria, com empregados
- |6| Artista (pintor, escultor, músico, cantor, ator etc.)
- |7| Trabalha em entidade, organização ou instituição não-governamental de cunho filantrópico, assistencial, religioso, de lazer ou outro
- |8| Parlamentar ou cargo eleitoral, diplomata, militar
- |9| Atleta profissional
- |10| Trabalha em casa e/ou não tem atividade remunerada
- |11| Não trabalha
- |12| Outros

7- OCUPAÇÃO DA MAE:

- |1| Funcionário público do governo Federal, Estadual ou Municipal
- |2| Empregado de empresa
- |3| Sócio ou proprietário de empresa
- |4| Trabalho remunerado por conta própria, com auxílio de parentes e/ou de familiares
- |5| Trabalho remunerado por conta própria, com empregados

- |6| Artista (pintor, escultor, músico, cantor, ator etc.)
- |7| Trabalha em entidade, organização ou instituição não-governamental de cunho filantrópico, assistencial, religioso, de lazer ou outro
- |8| Parlamentar ou cargo eleitoral, diplomata, militar
- |9| Atleta profissional
- |10| Trabalha em casa e/ou não tem atividade remunerada
- |11| Não trabalha
- |12| Outros

8- RENDA FAMILIAR:

- |1| Até R\$ 300,00
- |2| De R\$ 301,00 a R\$ 500,00
- |3| De R\$ 501,00 a R\$ 1.000,00
- |4| De R\$ 1.001,00 a R\$ 1.500,00
- |5| De R\$ 1.501,00 a R\$ 2.000,00
- |6| De R\$ 2.001,00 a R\$ 3.000,00
- |7| De R\$ 3.001,00 a R\$ 4.000,00
- |8| De R\$ 4.001,00 a R\$ 5.000,00
- |9| Acima de R\$ 5.001,00

9- NÚMERO DE PESSOAS QUE CONTRIBUEM RENDA FAMILIAR:

- |1| Uma
- |2| Duas
- |3| Três
- |4| Quatro
- |5| Cinco
- |6| Seis ou mais

10- NÚMERO PESSOAS SUSTENTADAS RENDA FAMILIAR:

- |1| Uma
- |2| Duas
- |3| Três
- |4| Quatro
- |5| Cinco
- |6| Seis ou mais

11- IDADE COMEÇOU EXERCER ATIVIDADE REMUNERADA

- |1| Antes dos 14 anos
- |2| Entre 14 e 16 anos
- |3| Entre 16 e 18 anos
- |4| Após 18 anos
- |5| Nunca trabalhei

12- DURANTE O CURSO VOCÊ TERÁ OBRIGATORIAMENTE QUE TRABALHAR

- |1| Sim, mas apenas nos últimos anos
- |2| Sim, desde o primeiro ano, em tempo parcial
- |3| Sim, desde o primeiro ano, em tempo integral
- |4| Não sei
- |5| Não

13- ESTUDO DE 1<sup>o</sup>.GRAU:

- |1| Todos em escola pública
- |2| Todos em escola particular
- |3| Maior parte em escola pública
- |4| Maior parte em escola particular
- |5| Em escolas comunitárias/CNEC ou outro

14- ANO DE CONCLUSÃO DO 2<sup>o</sup>. GRAU:

15- ESTUDO 2<sup>o</sup>. GRAU:

- |1| Integralmente em escola pública
- |2| Integralmente em escola particular
- |3| Maior parte em escola pública
- |4| Maior parte em escola particular
- |5| Em escolas comunitárias/CNEC ou outro

16- RELAÇÃO ENTRE FORMAÇÃO DE 2<sup>o</sup>. GRAU E ATIVIDADE ATUAL:

- |1| Concluiu o curso de magistério
- |2| Concluiu curso técnico (agrícola, contábil, mecânico etc.)

|3| Não se enquadra nas alternativas anteriores

17- TURNO DO 2<sup>o</sup>. GRAU:

- |1| Todo diurno
- |2| Todo noturno
- |3| Maior parte diurno
- |4| Maior parte noturno
- |5| Outro

18- CURSO PREPARATÓRIO PARA O VESTIBULAR:

- |1| Fiz apenas o "terceirão"
- |2| Fiz apenas cursinho
- |3| Fiz "terceirão" e cursinho
- |4| Não fiz nem "terceirão" nem cursinho

19- MOTIVO DO CURSO PREPARATÓRIO:

- |1| Para atualizar meus conhecimentos, porque parei de estudar há muito tempo
- |2| Para aprender "macetes"
- |3| Para complementar os conhecimentos adquiridos no colégio
- |4| Por outro motivo
- |5| Não fiz cursinho

20- NÚMERO DE VEZES QUE PRESTOU VESTIBULAR:

- |1| Uma vez
- |2| Duas vezes
- |3| Três vezes
- |4| Quatro vezes
- |5| Cinco ou mais
- |6| Nenhuma

21- NÚMERO DE VESTIBULARES QUE FEZ NO ANO DE INGRESSO:

- |1| Um

- |2| Dois
- |3| Três
- |4| Quatro
- |5| Cinco ou mais

22- INICIOU OUTRO CURSO SUPERIOR:

- |1| Sim, mas não concluí
- |2| Sim, estou cursando
- |3| Sim, mas já concluí
- |4| Não

23- MOTIVO DE TER PRESTADO VESTIBULAR NA UFPR:

- |1| Por se tratar de universidade pública e gratuita
- |2| Pela qualidade do ensino
- |3| Pelo horário do curso que pretendo fazer
- |4| Pela localização
- |5| Por ser a única na cidade que oferece o curso que desejo
- |6| Outro motivo

24- MOTIVO DE ESCOLHA DO CURSO

- |1| Mercado de trabalho e possibilidades salariais
- |2| Possibilidade de contribuir para a sociedade
- |3| Possibilidade de cursar algo de que gosta
- |4| Por ter habilidades relacionadas ao curso
- |5| Gosto pelas matérias do curso
- |6| Baixa concorrência pelas vagas
- |7| Permite conciliar aula e trabalho
- |8| Outro motivo

25- FATOR QUE TERÁ QUE ENFRENTAR DURANTE O CURSO QUE  
CONSIDERA IMPORTANTE:

- |1| Habilidades específicas exigidas pelo curso

- |2| Relacionamentos
- |3| Persistência e hábitos de estudo
- |4| Informações da profissão
- |5| Conciliar estudos com vida pessoal e familiar
- |6| Não se enquadra nas alternativas anteriores

26- O QUE ESPERA EM 1º. LUGAR DE UM CURSO UNIVERSITÁRIO:

- |1| Aquisição de cultura geral ampla
- |2| Formação profissional, voltada para o trabalho
- |3| Formação teórica, voltada para a pesquisa
- |4| Formação acadêmica para melhorar a atividade prática que já estou desempenhando
- |5| Aquisição de conhecimentos que me permitam compreender melhor o mundo em que vivemos
- |6| Aquisição de conhecimentos que permitam melhorar meu nível de instrução
- |7| Diploma de nível superior

27- ATIVIDADE DE LAZER:

- |1| TV
- |2| Religião
- |3| Teatro
- |4| Cinema
- |5| Música
- |6| Dança
- |7| Artesanato
- |8| Leitura
- |9| Esporte
- |10| Outra
- |11| Nenhuma

28- QUE MEIO QUE MAIS UTILIZA PARA SE MANTER INFORMADO:

- |1| Jornal escrito

- |2| TV
- |3| Rádio
- |4| Revistas
- |5| Outros
- |6| Nenhum

29- NÃO SE INSCREVEU PELA INTERNET...

- |1| Por não ter acesso a ela
- |2| Por não ter conhecimento do período de inscrição
- |3| Por outro motivo
- |4| Me inscrevi pela Internet em outro lugar
- |5| Me inscrevi pela Internet em casa

30- CLASSIFICAÇÃO NO VESTIBULAR

31- ESCORE NO VESTIBULAR EM PORTUGUÊS

32- ESCORE NO VESTIBULAR EM MATEMÁTICA

33- ESCORE NO VESTIBULAR EM BIOLOGIA

34- ESCORE NO VESTIBULAR EM QUÍMICA

35- ESCORE NO VESTIBULAR EM GEOGRAFIA

36- ESCORE NO VESTIBULAR EM FÍSICA

37- ESCORE NO VESTIBULAR EM HISTÓRIA

38- ESCORE NO VESTIBULAR EM LINGUA ESTRANGEIRA MODERNA

39- ESCORE NO VESTIBULAR EM REDAÇÃO

40- FREQUÊNCIA NA DISCIPLINA DE ESTATÍSTICA GERAL I:

DE 0 A 60 HORAS

41- NOTA NA DISCIPLINA DE ESTATÍSTICA GERAL I:

DE 0 A 100

42- FREQUÊNCIA NA DISCIPLINA DE CÁLCULO DE PROBABILIDADE I:

DE 0 A 60 HORAS

43- NOTA NA DISCIPLINA DE CÁLCULO DE PROBABILIDADE I:

DE 0 A 100

44- FREQUÊNCIA NA DISCIPLINA DE CÁLCULO COM GEOMETRIA:

ANALÍTICA I:

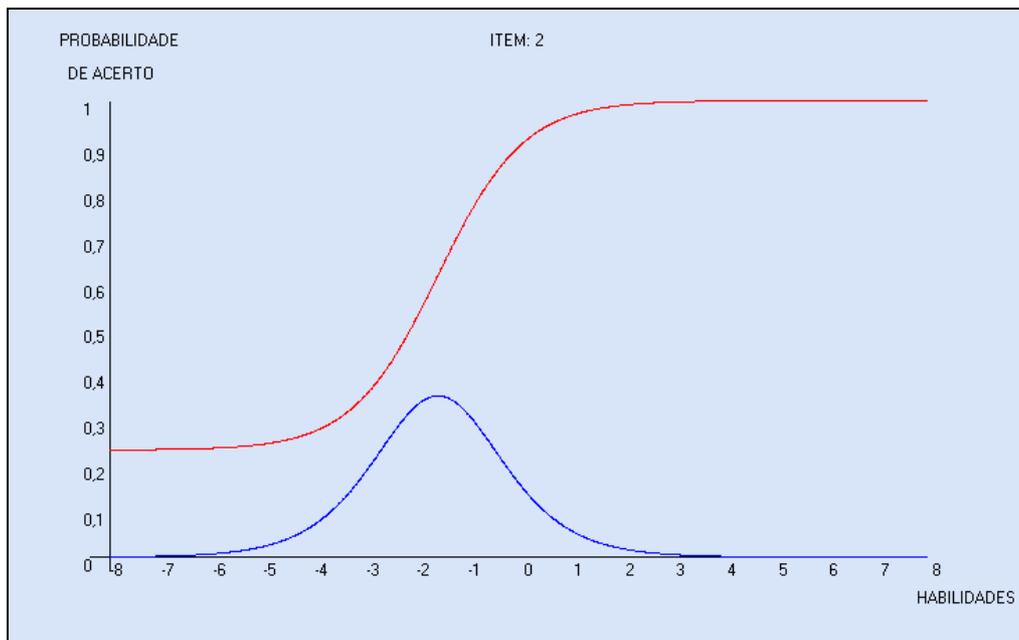
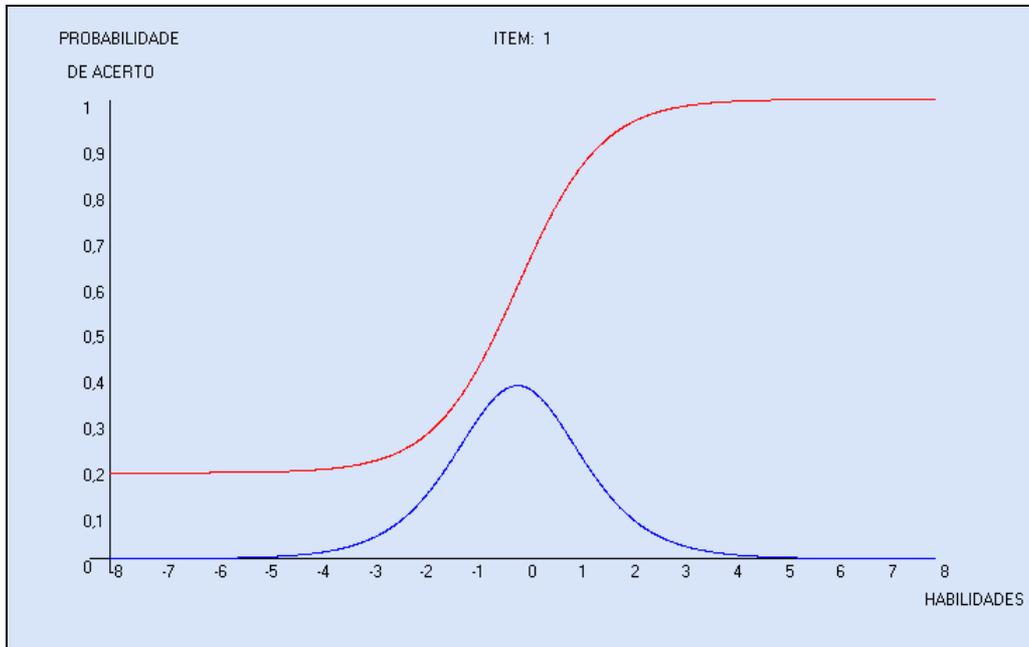
DE 0 A 60HORAS

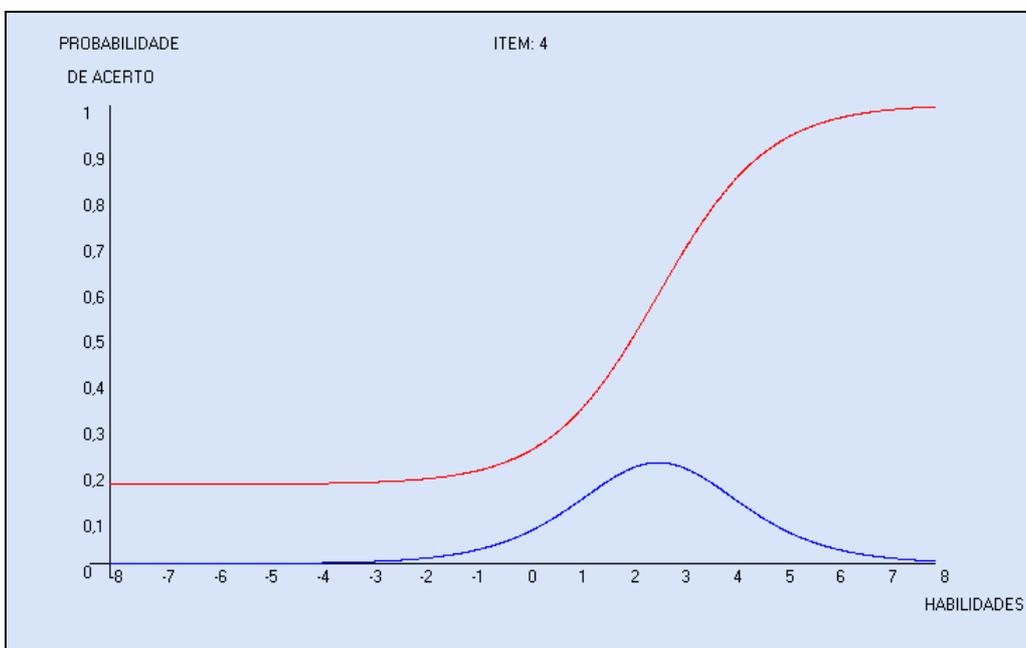
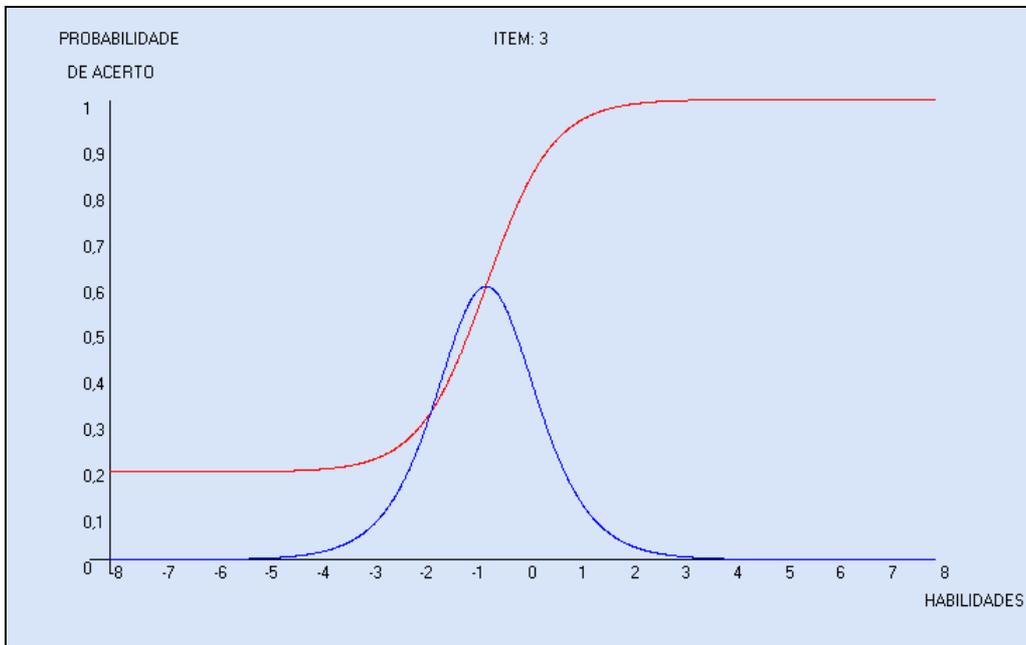
45- NOTA NA DISCIPLINA DE CÁLCULO COM GEOMETRIA ANALÍTICA I:

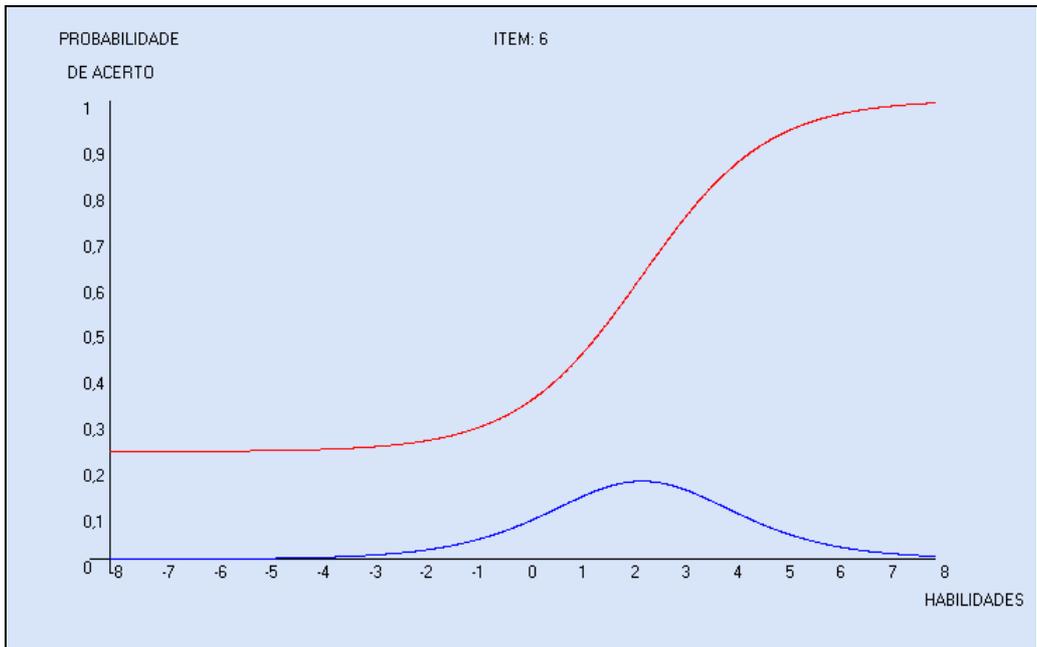
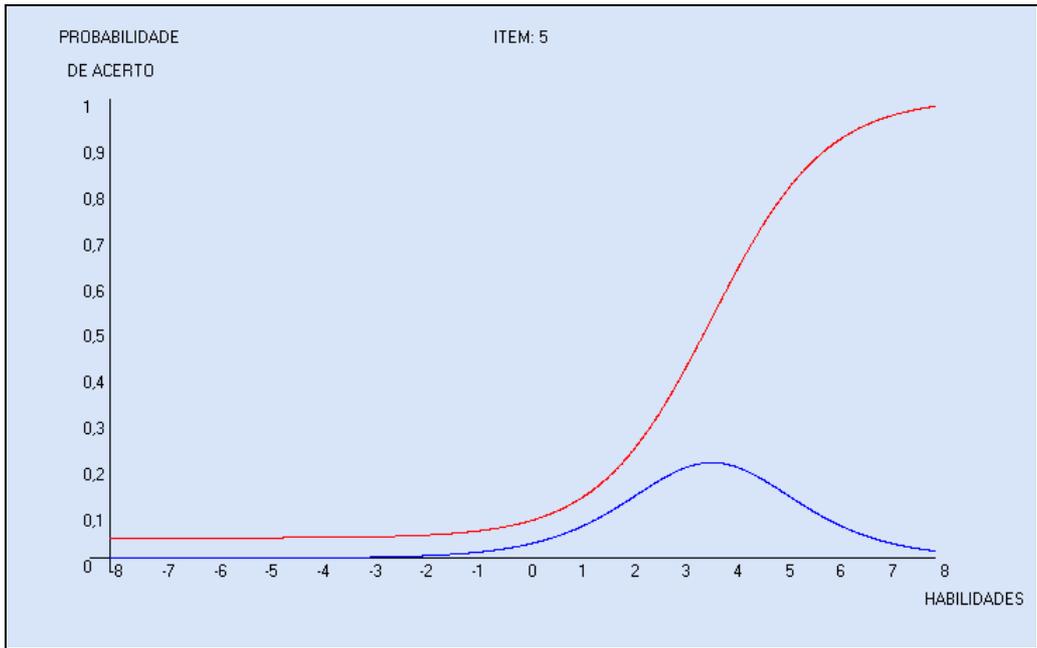
DE 0 A 100

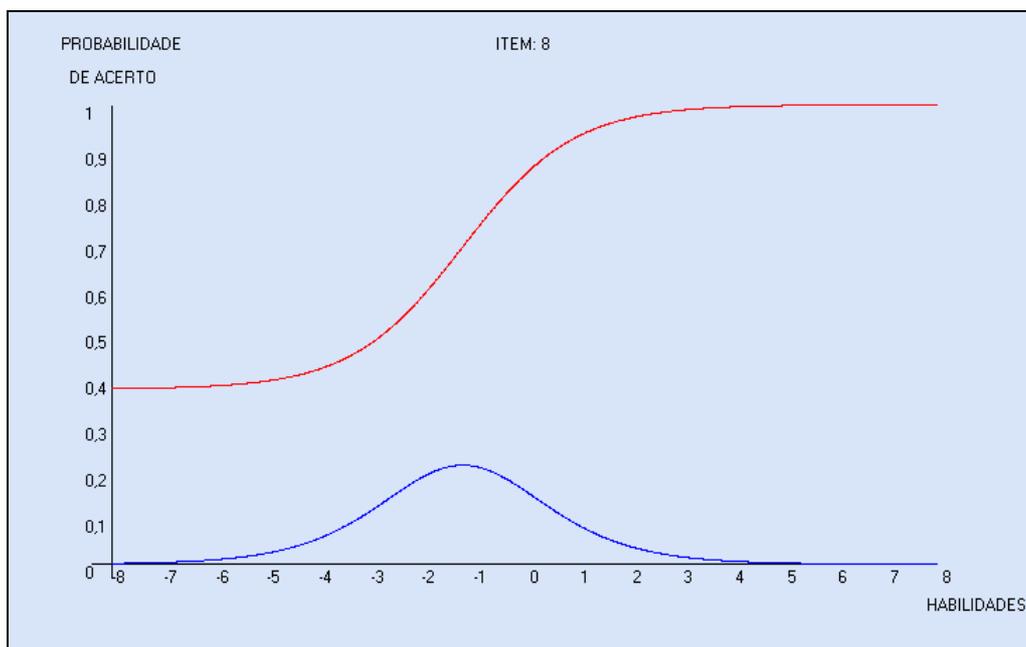
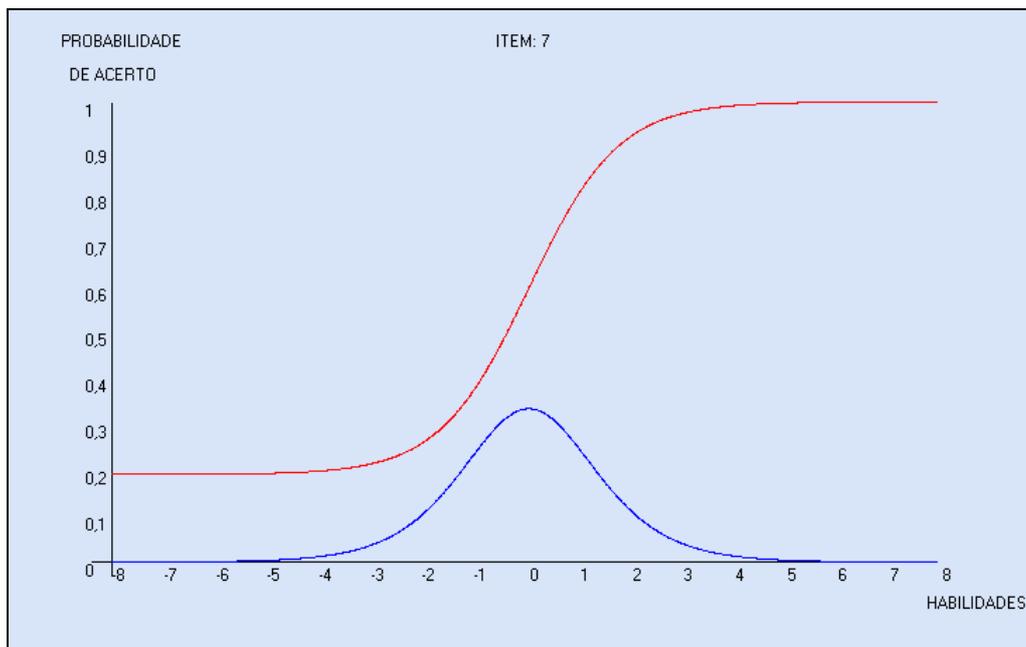
## APÊNDICE 9

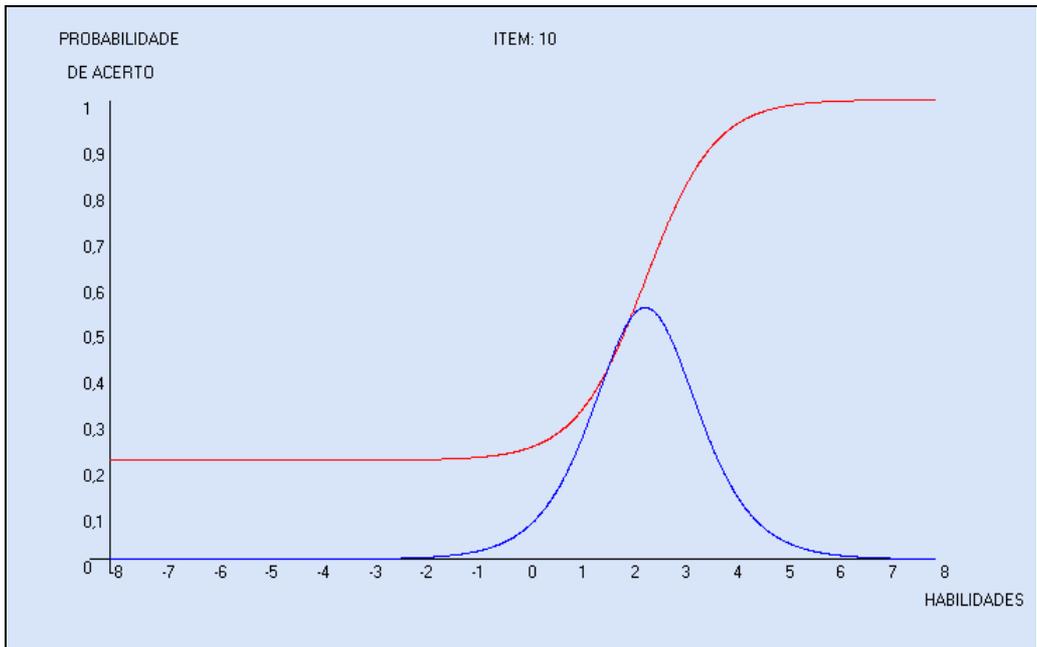
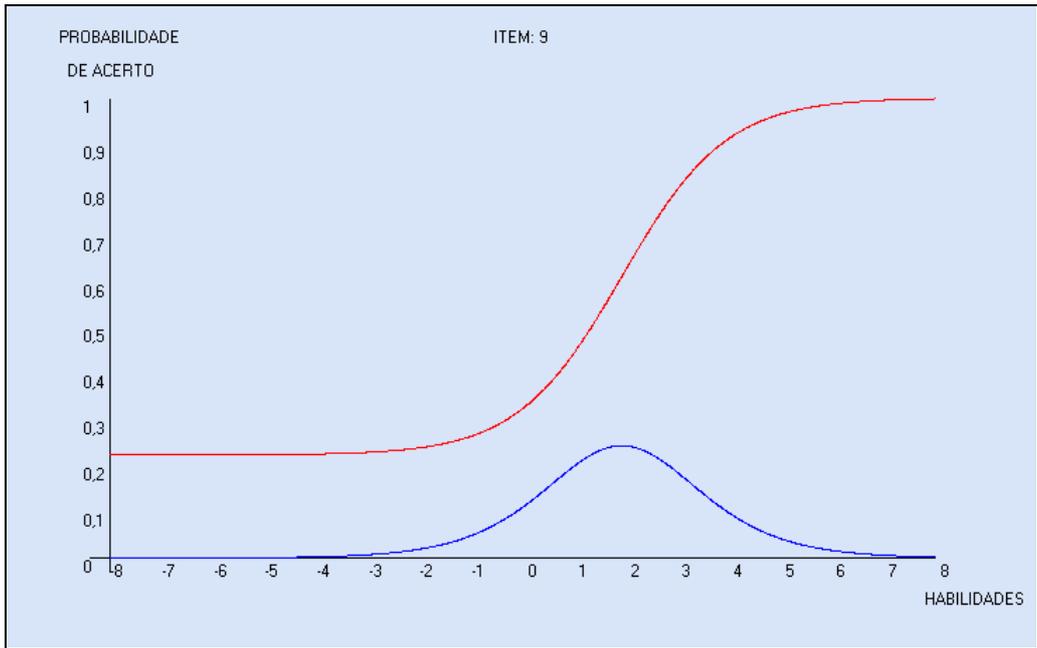
### GRÁFICOS DA CURVA CARACTERÍSTICA E FUNÇÃO DE INFORMAÇÃO DOS ITENS 1 A 60

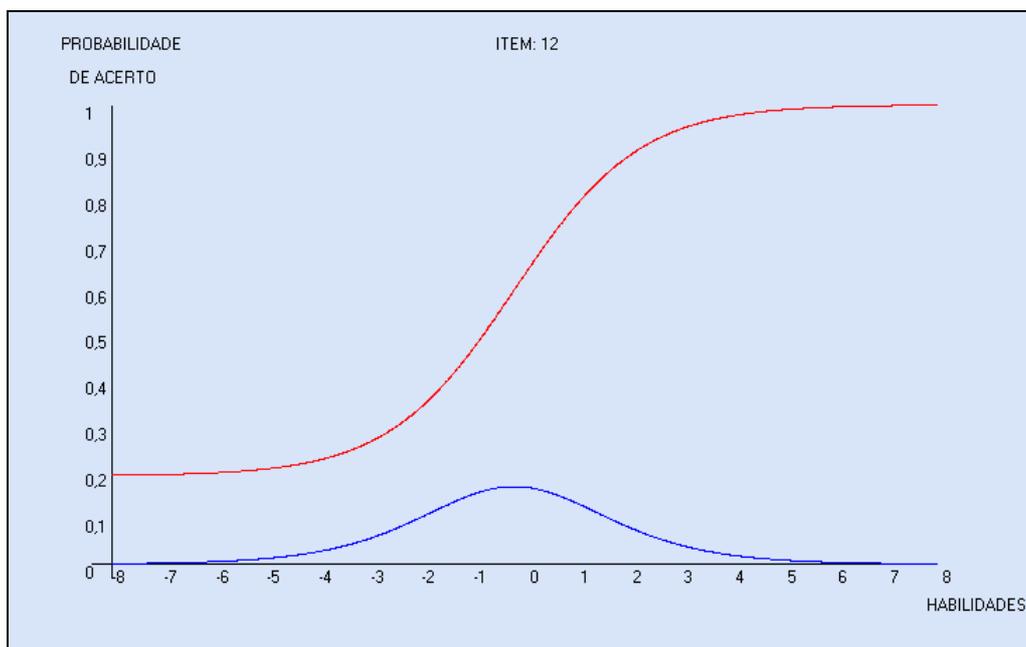
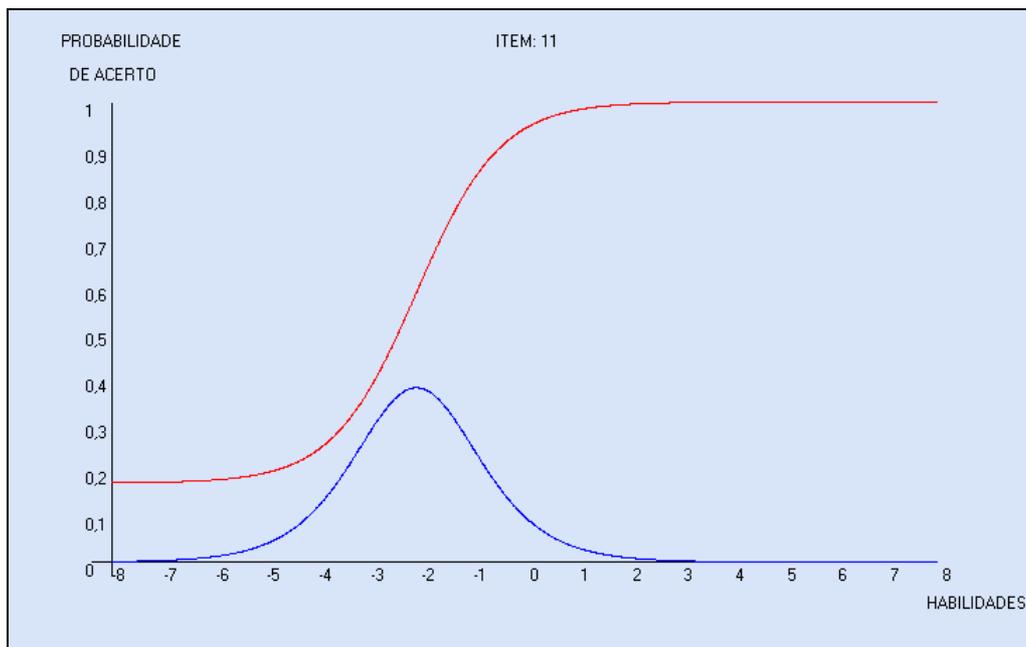


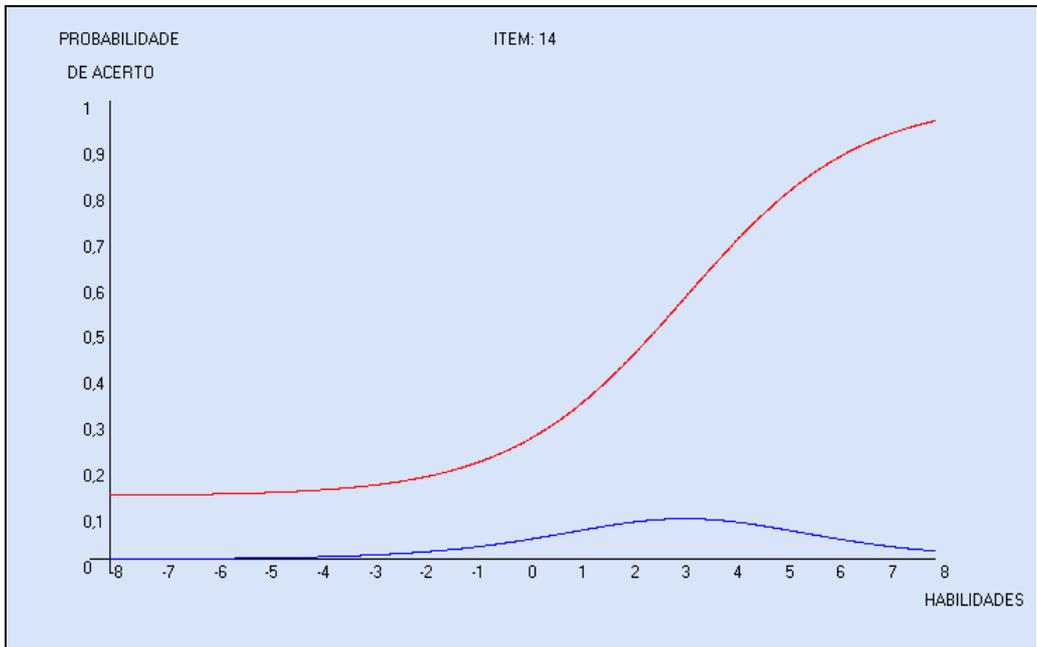
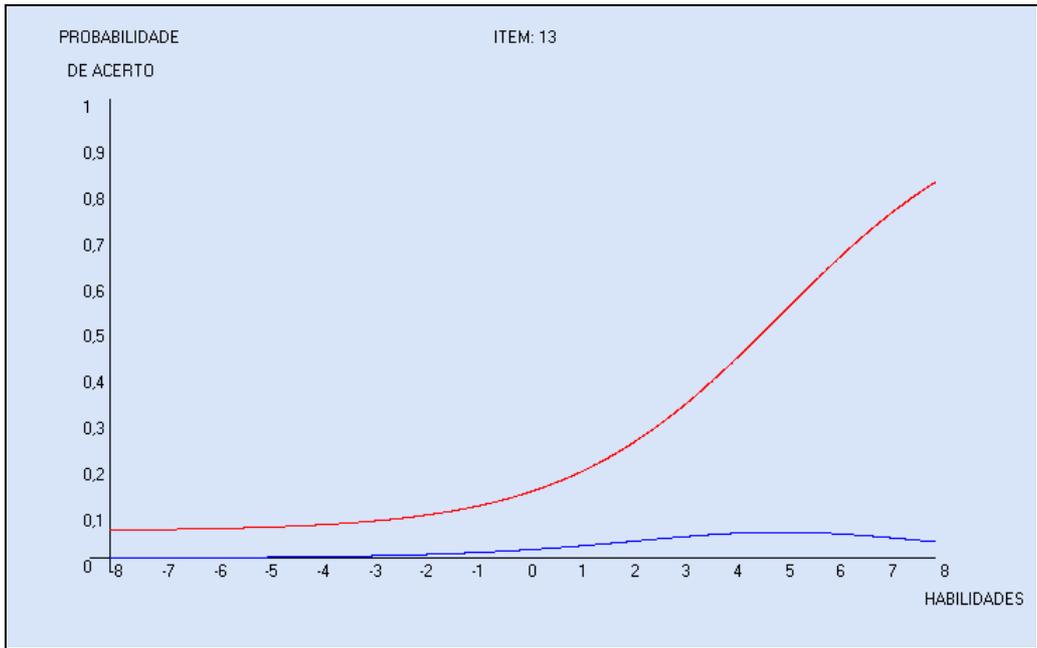


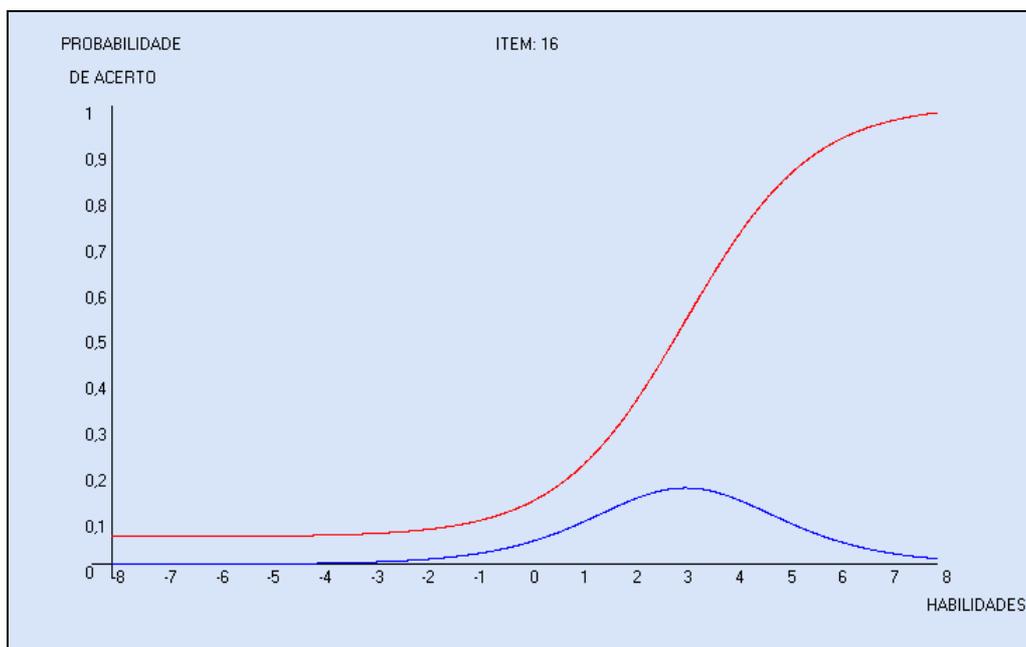
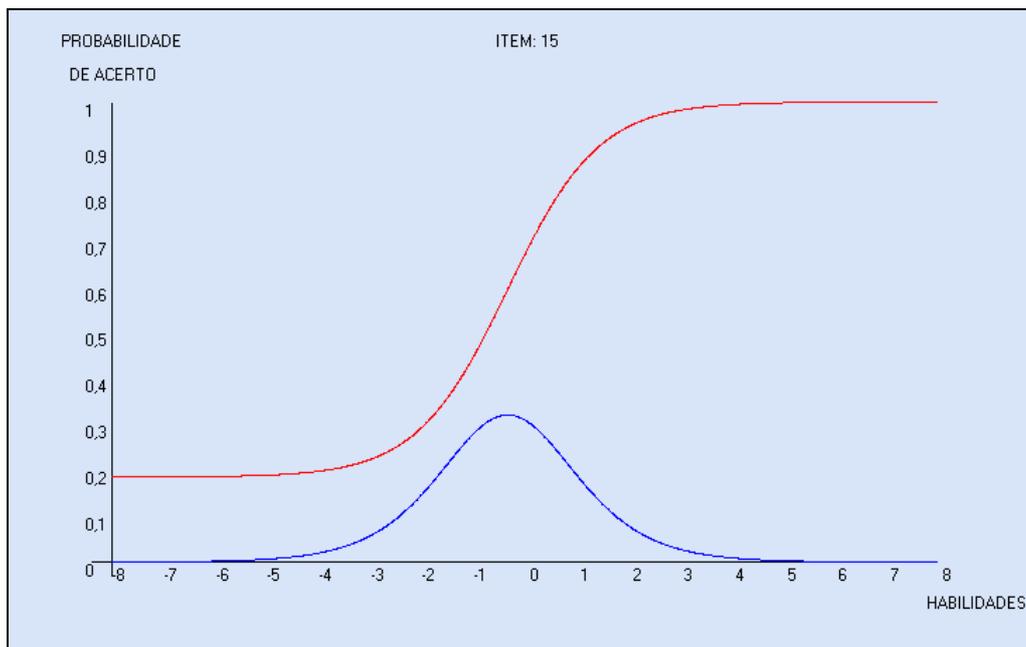


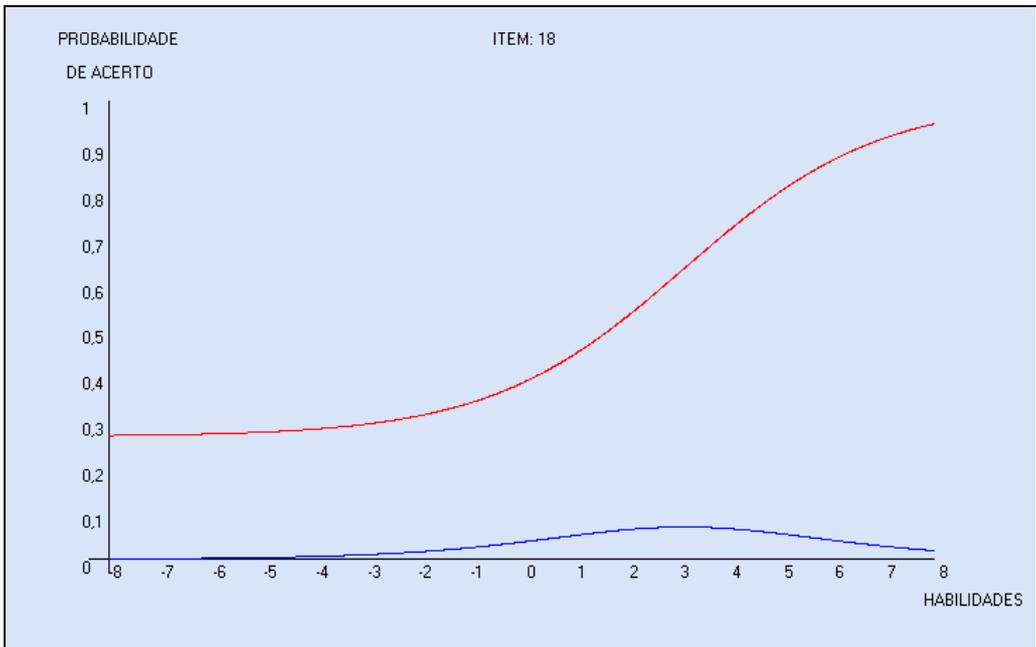
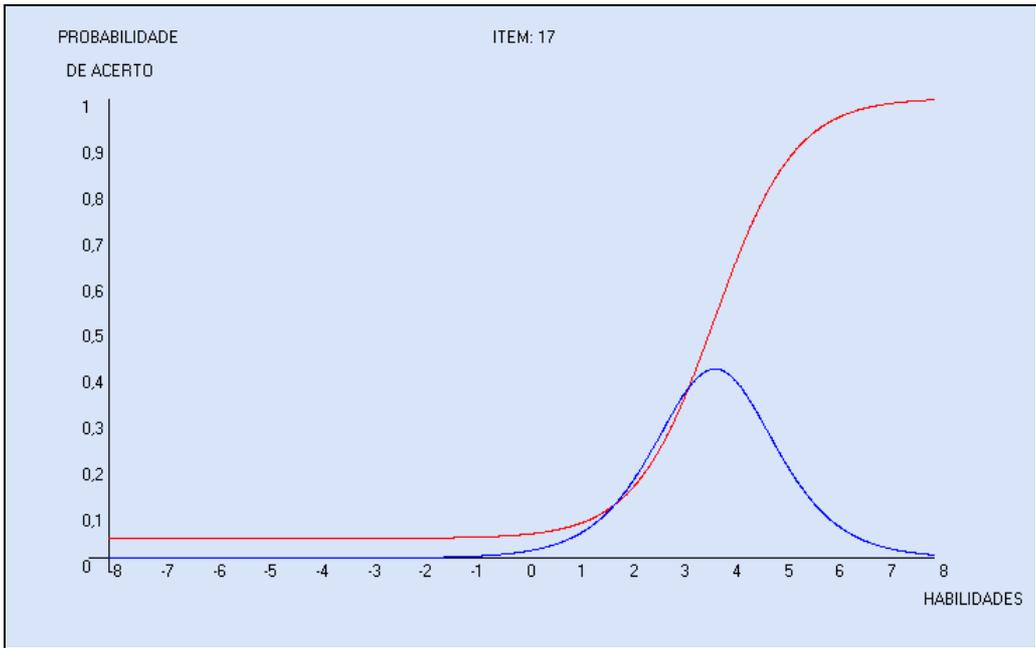


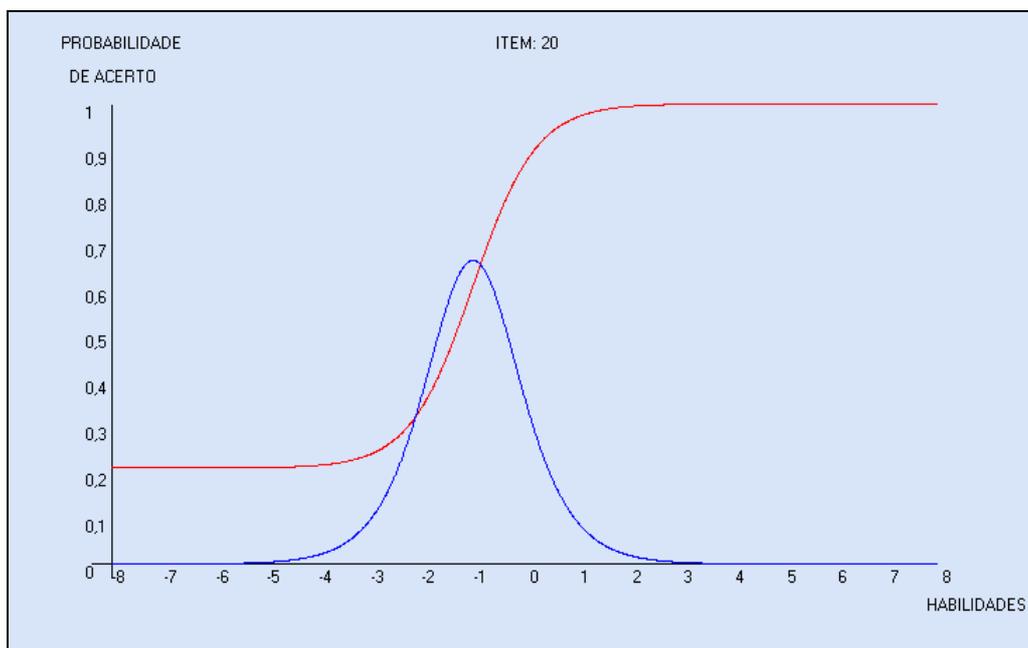
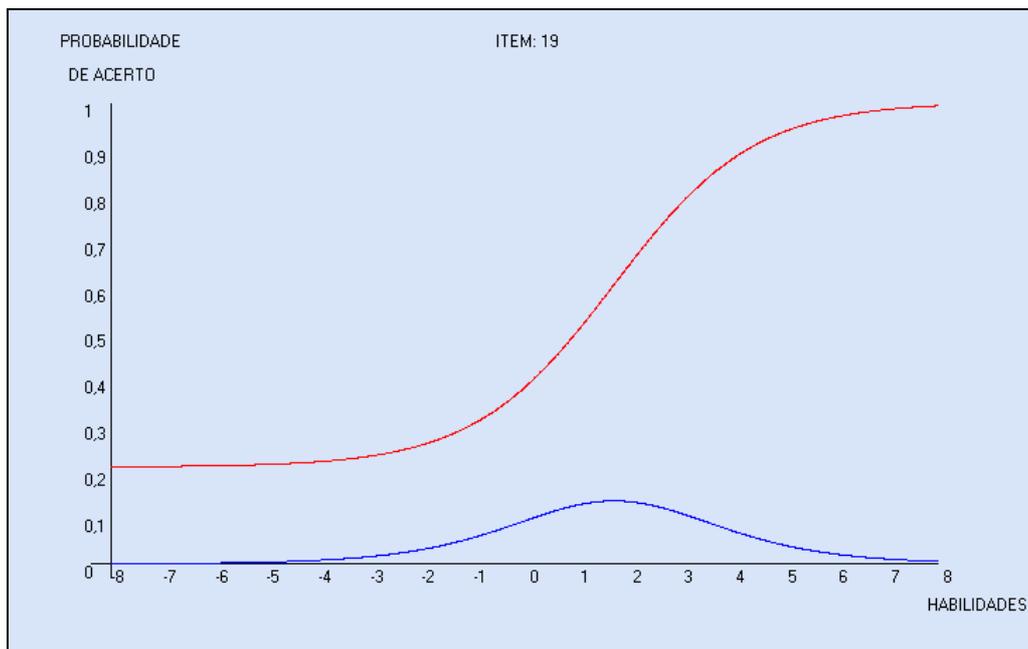


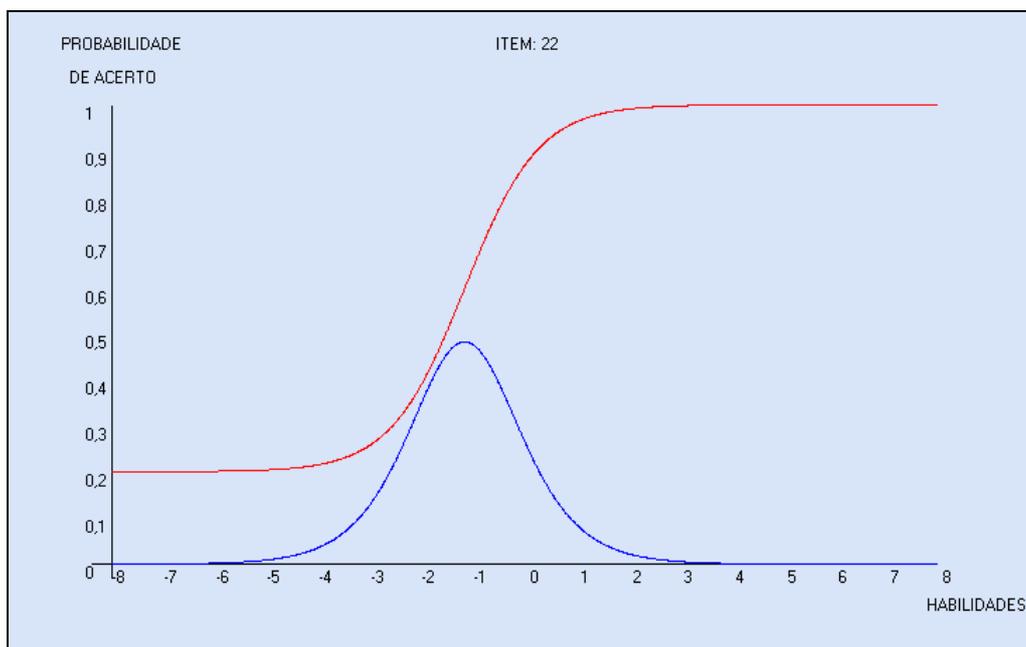
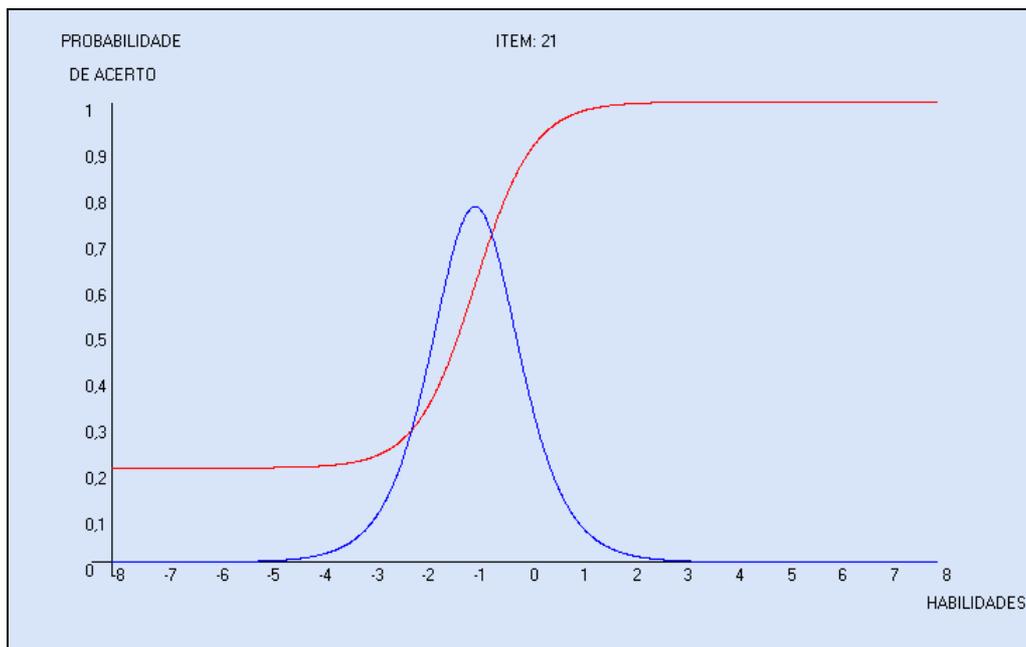


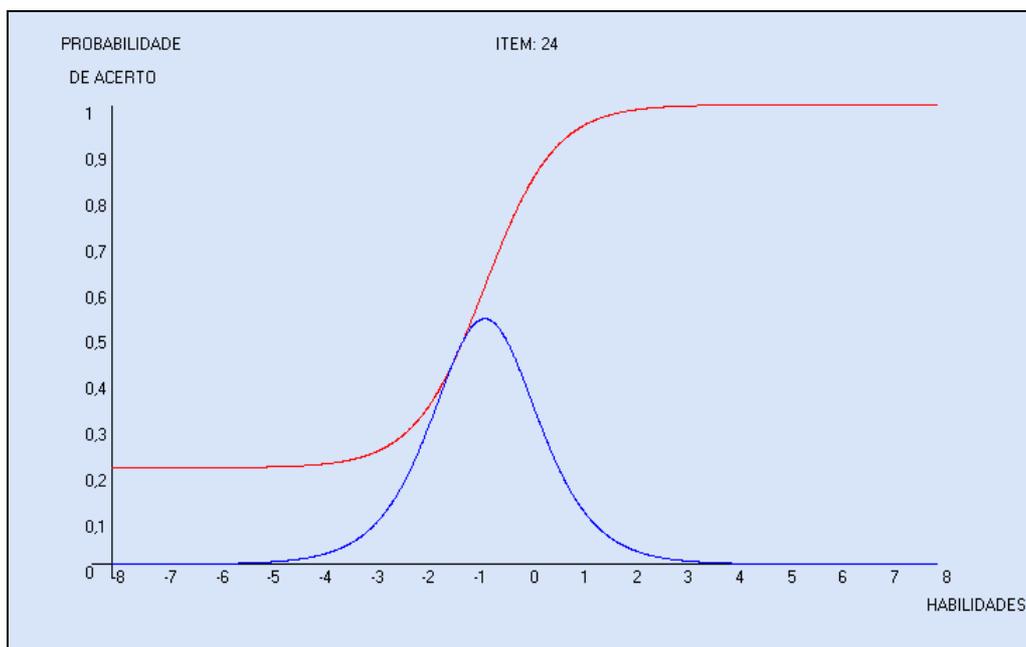
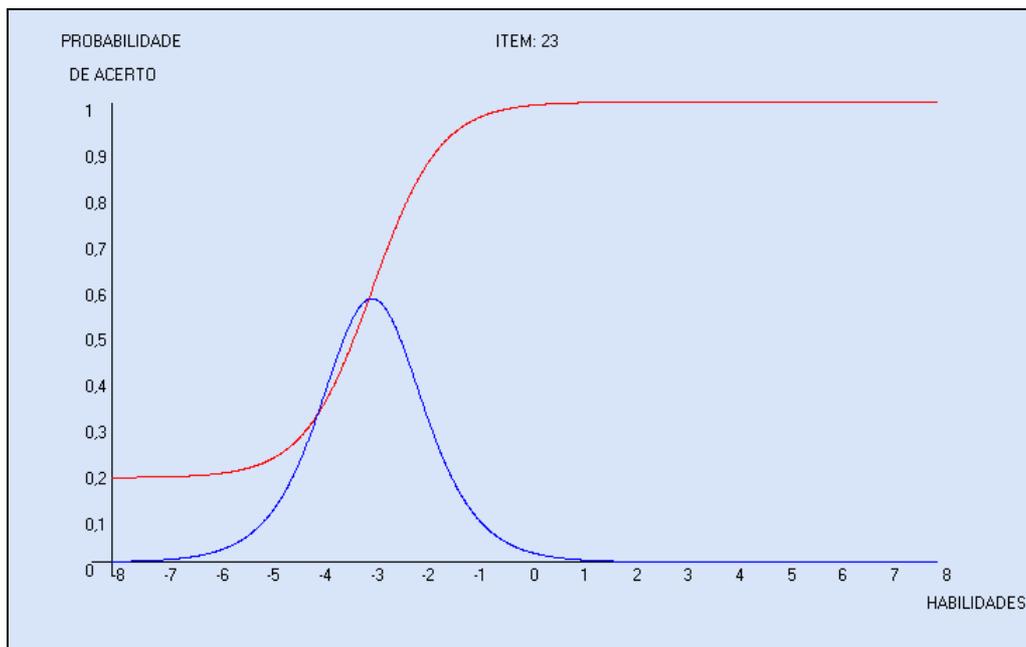


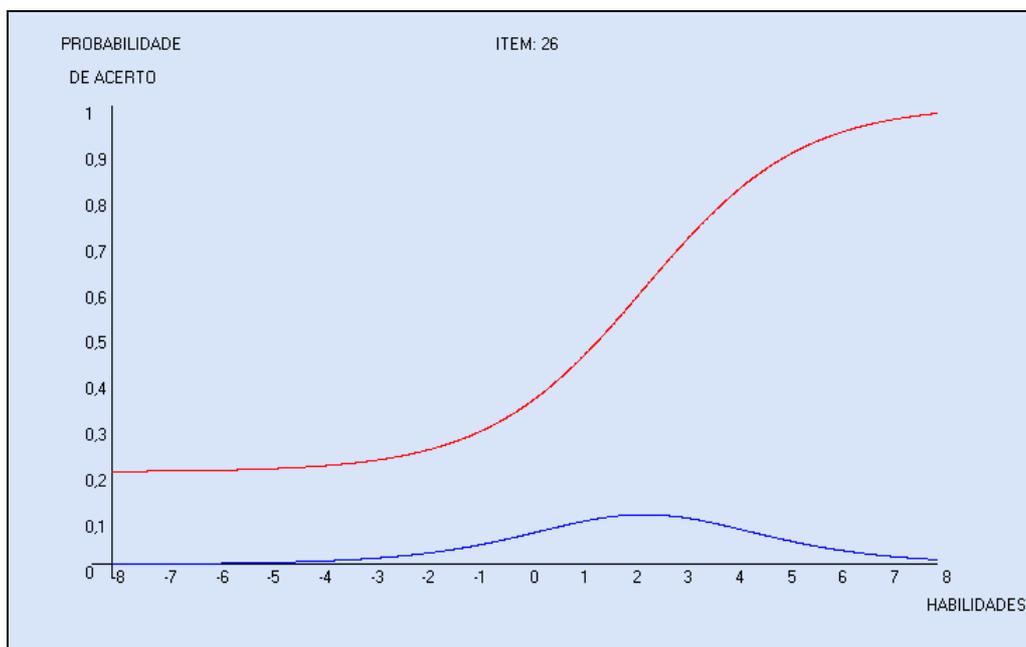
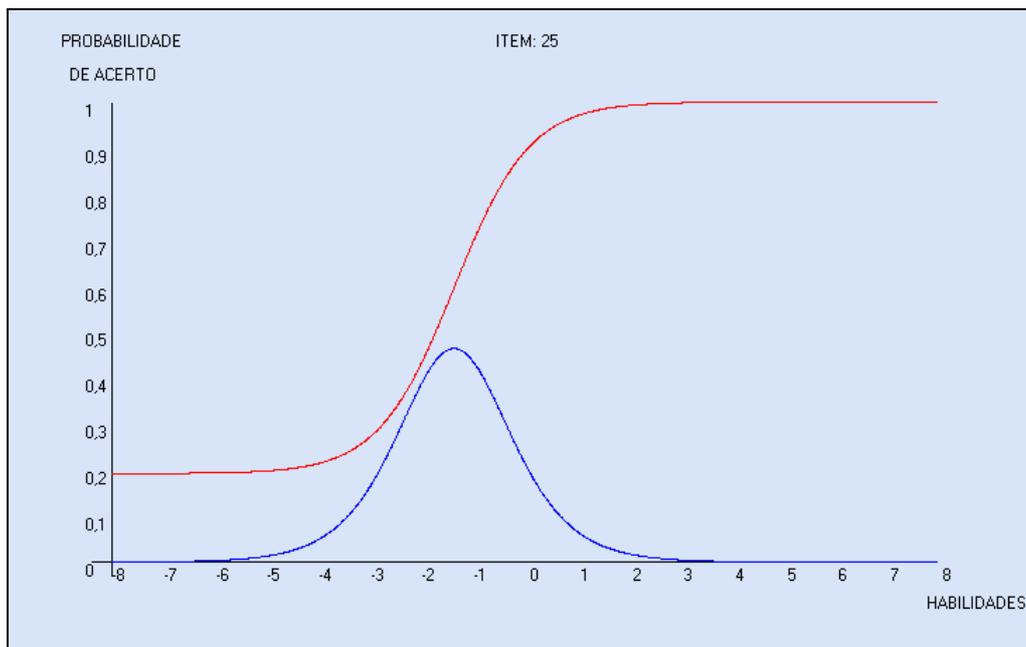


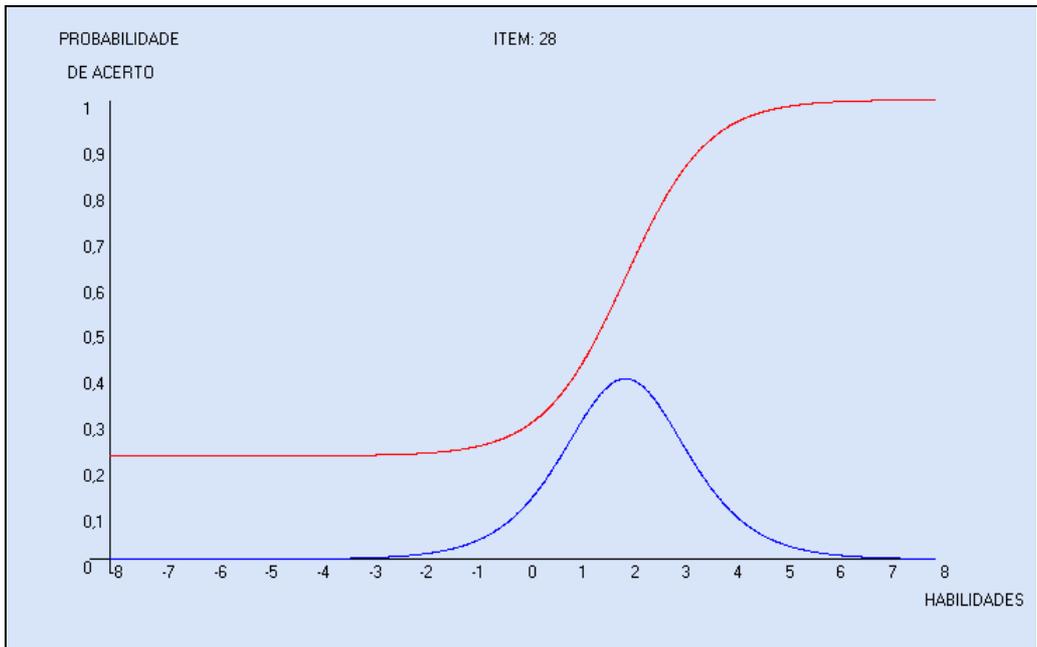
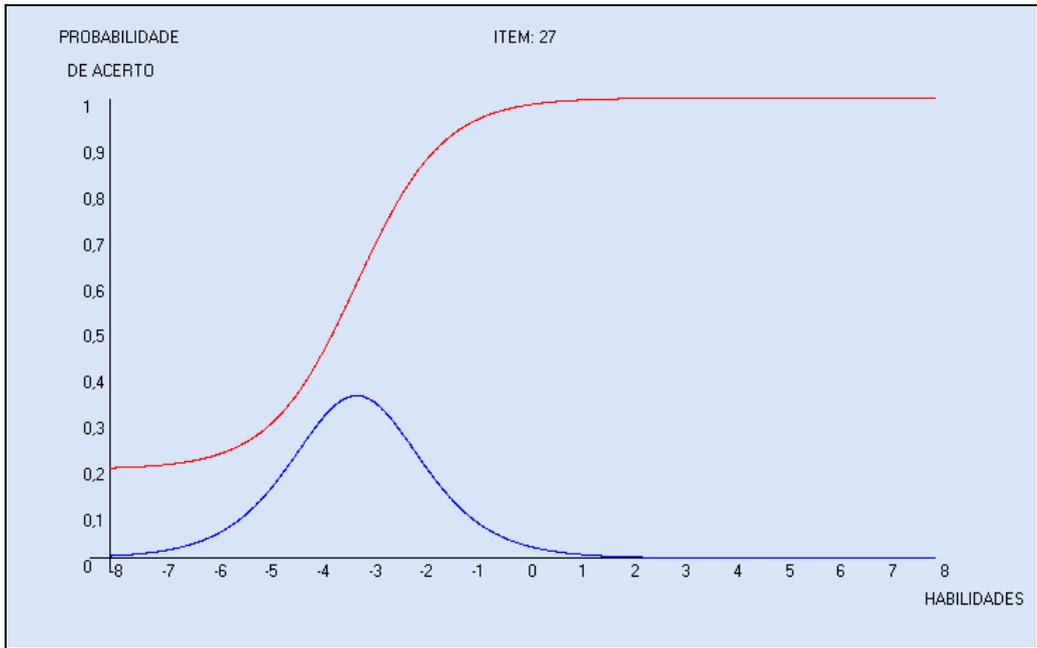


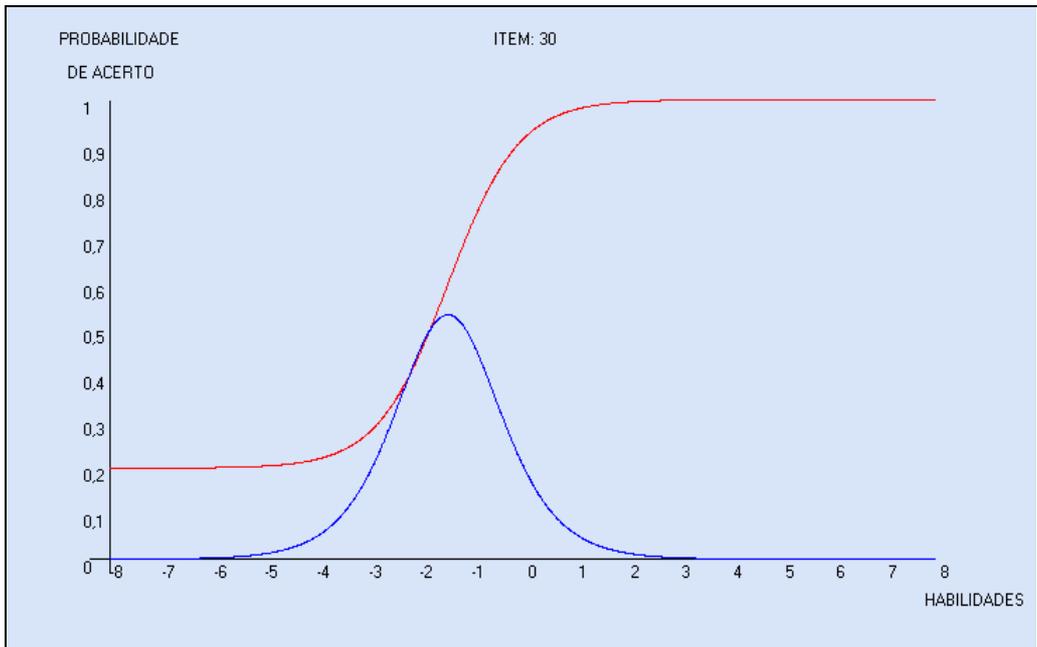
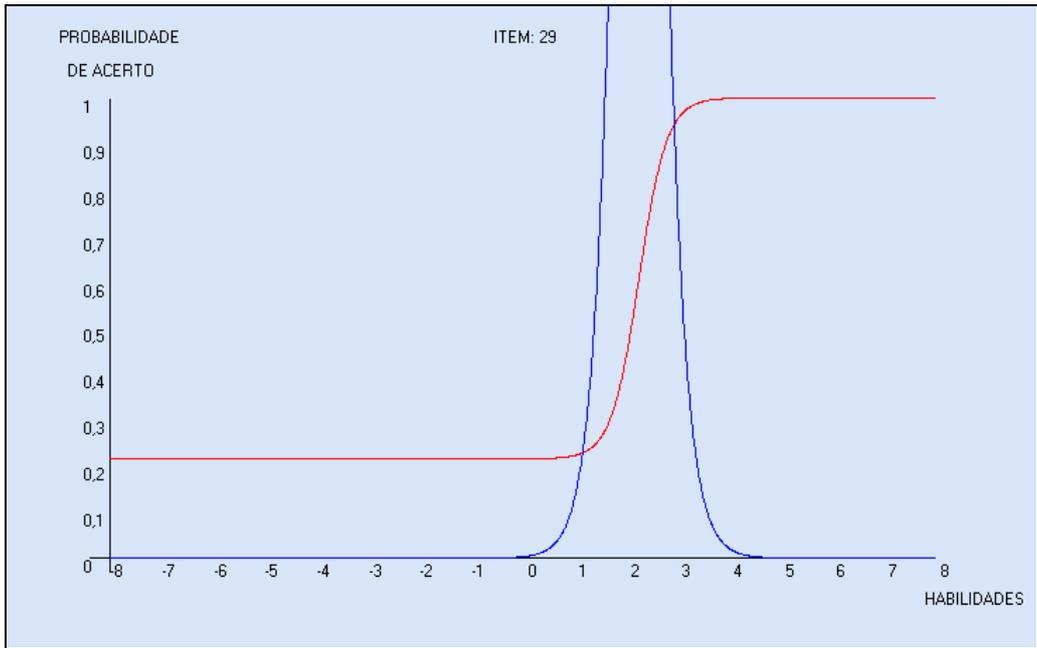


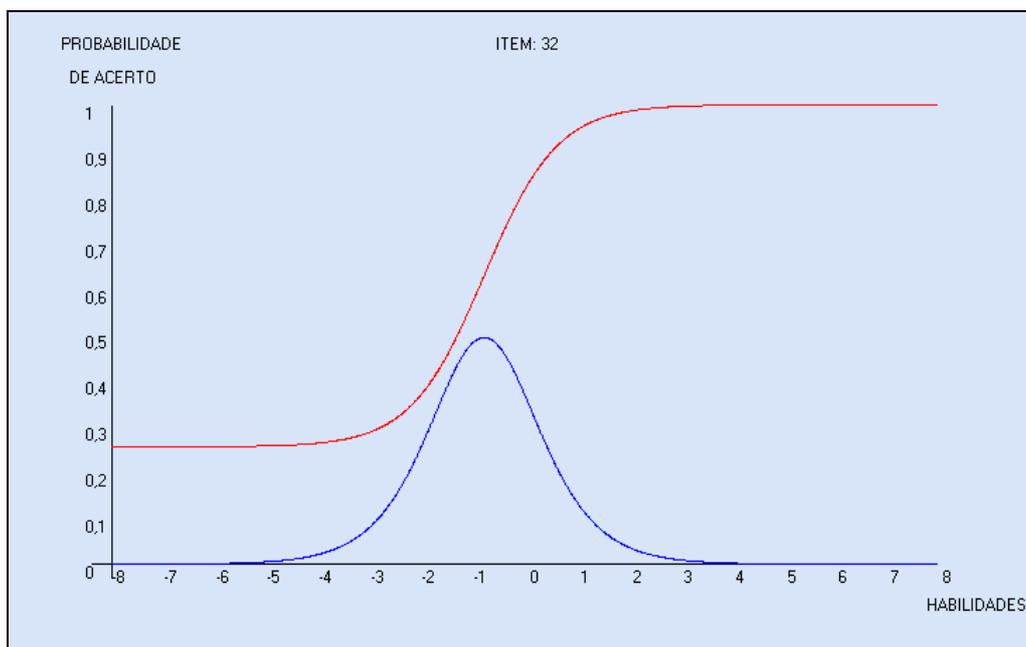
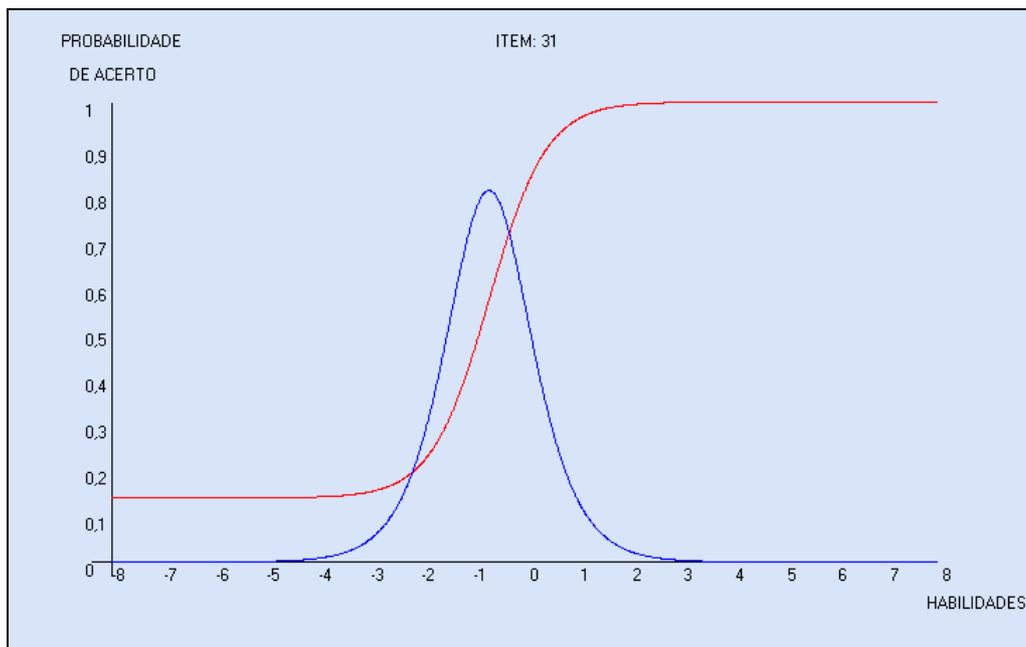


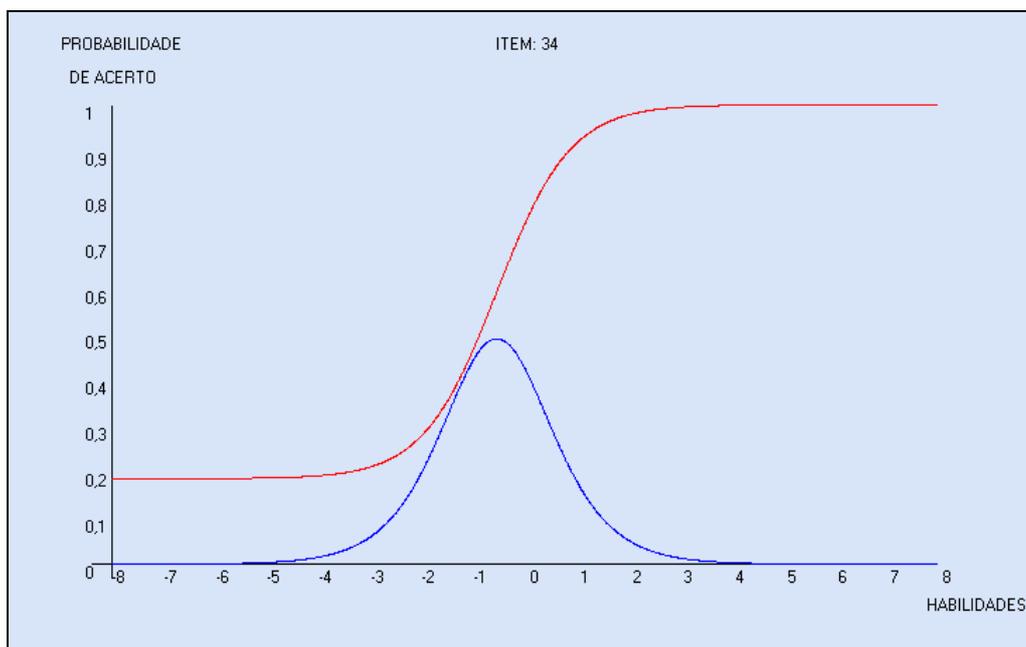
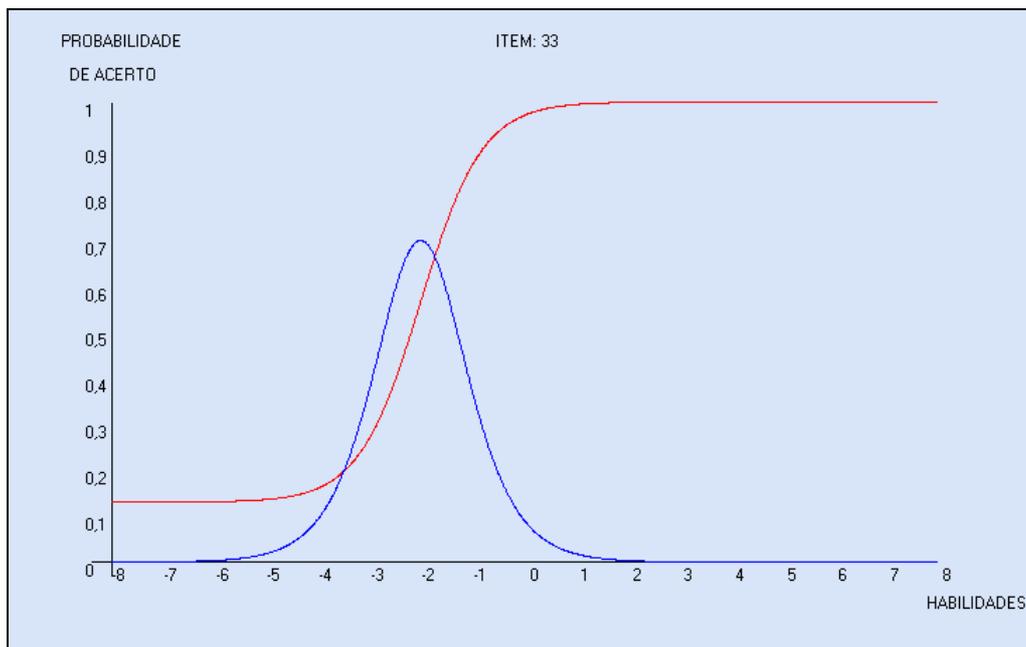


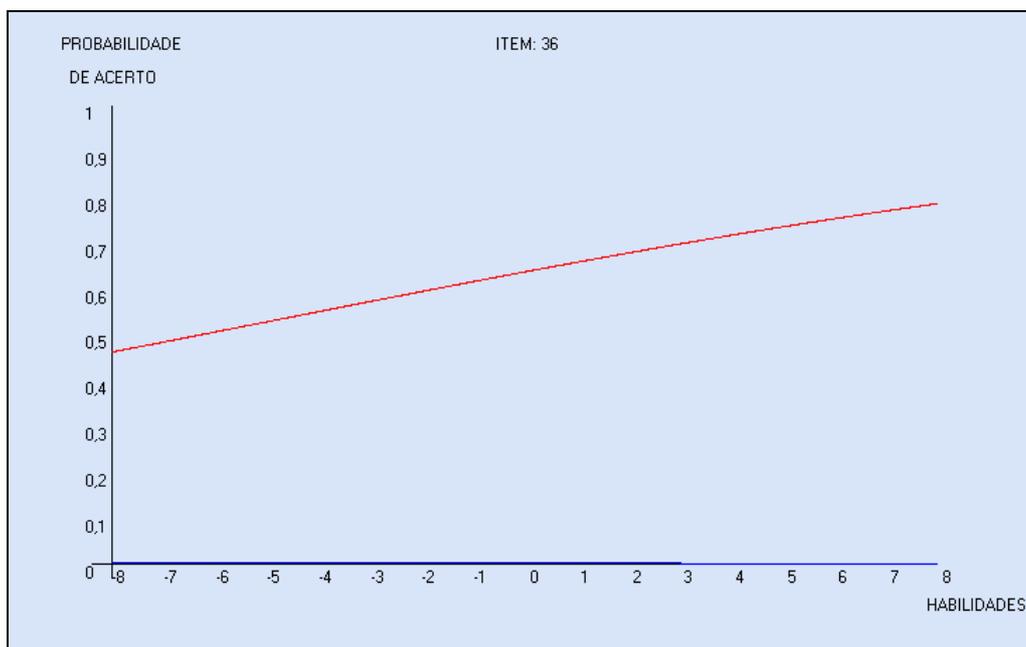
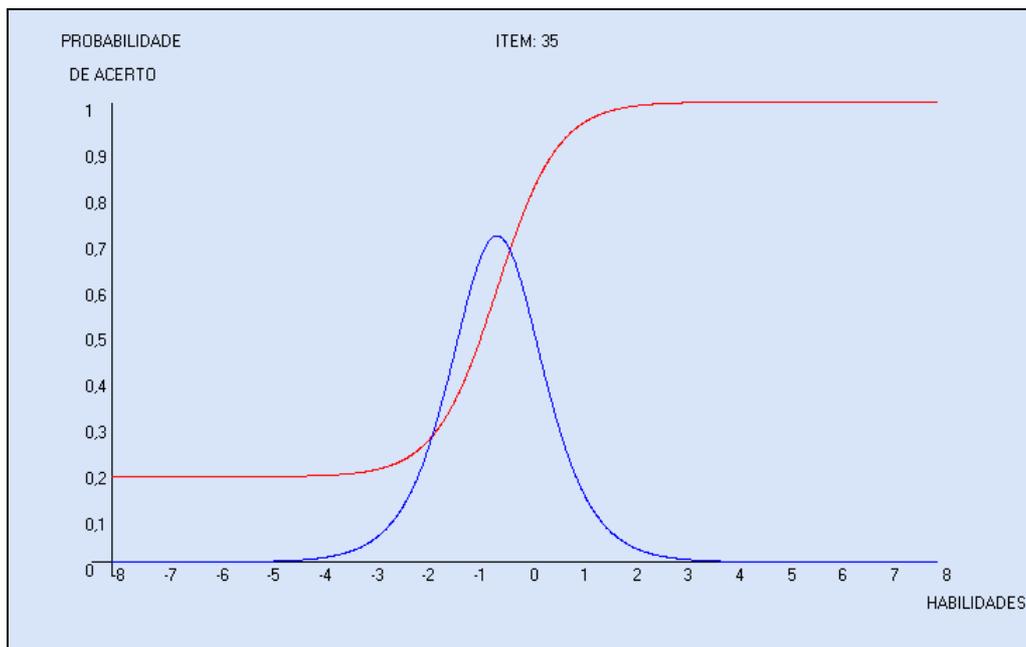


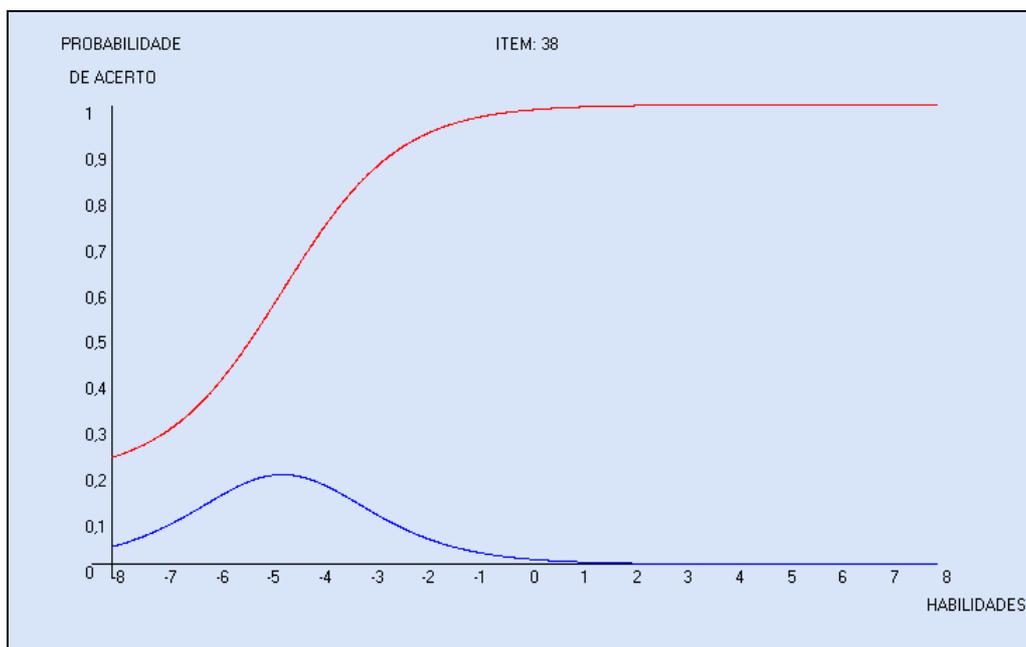
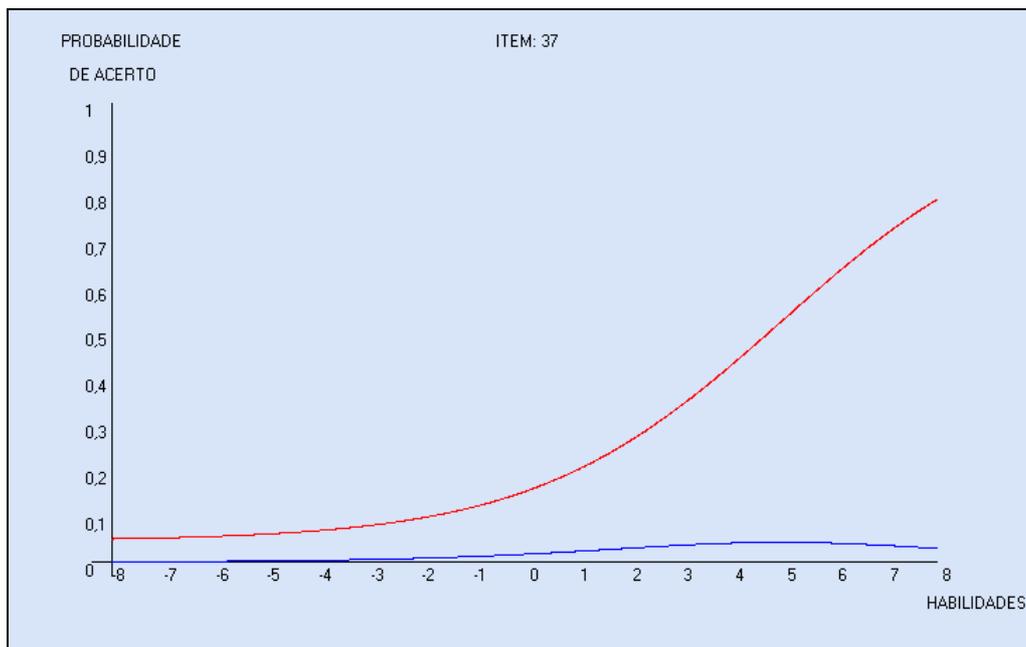


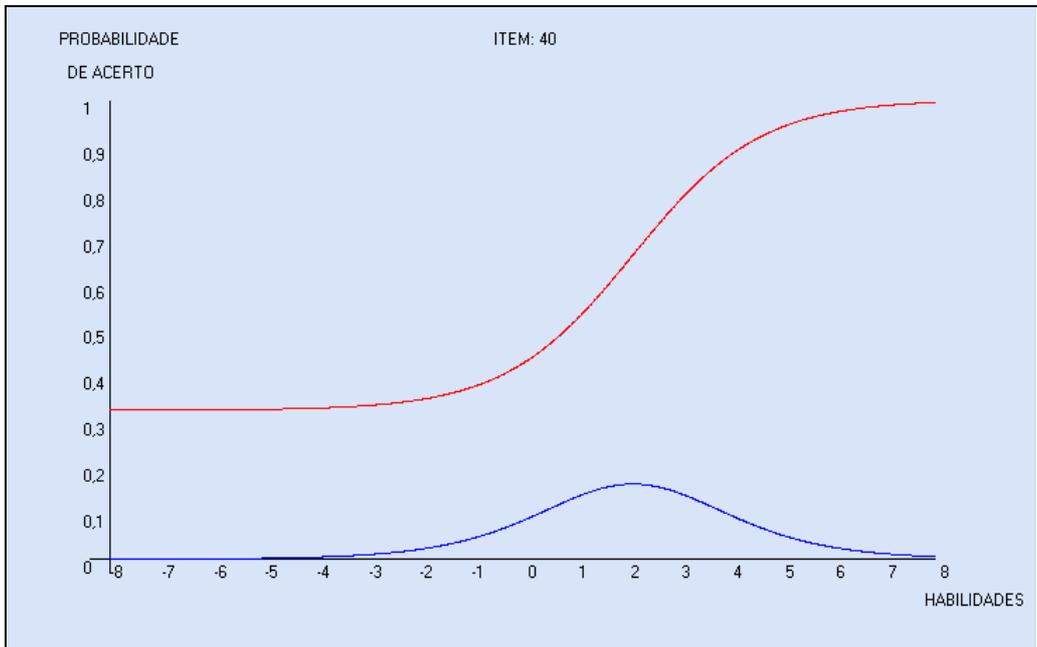
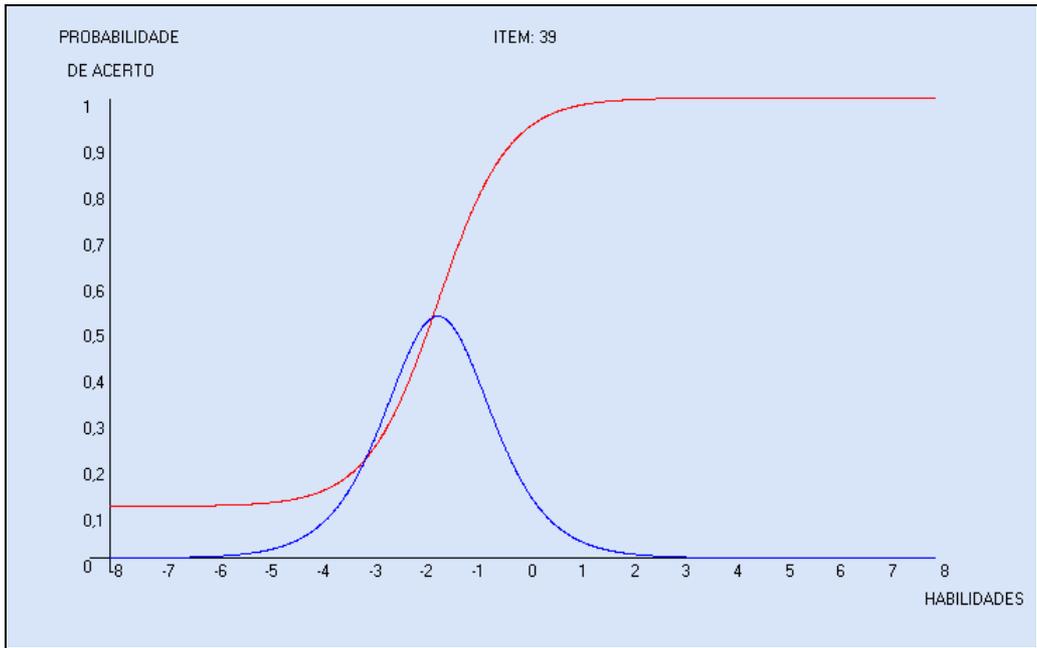


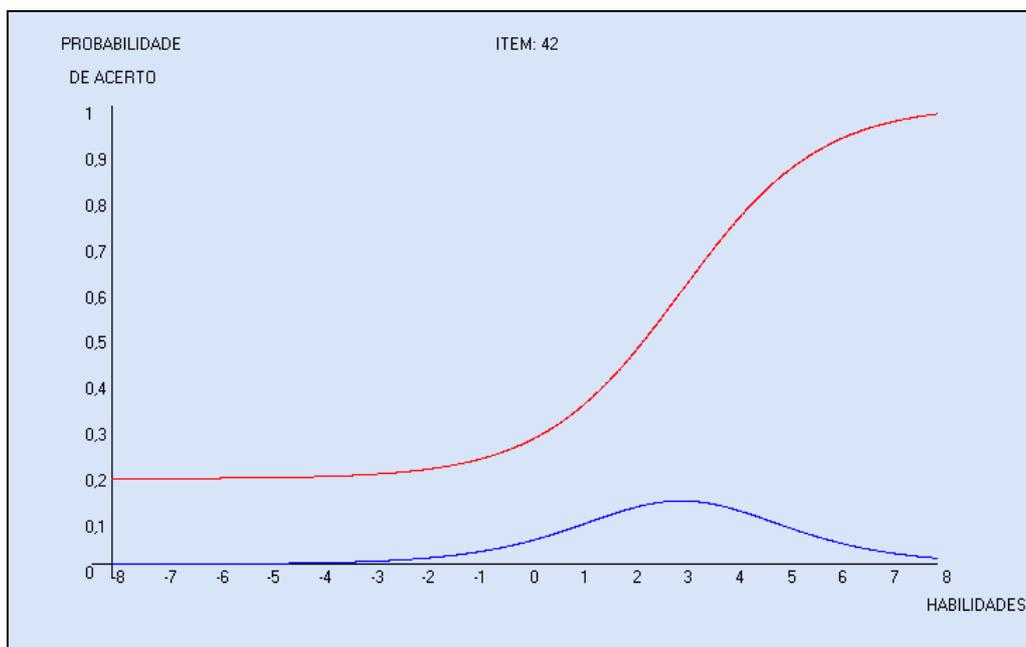
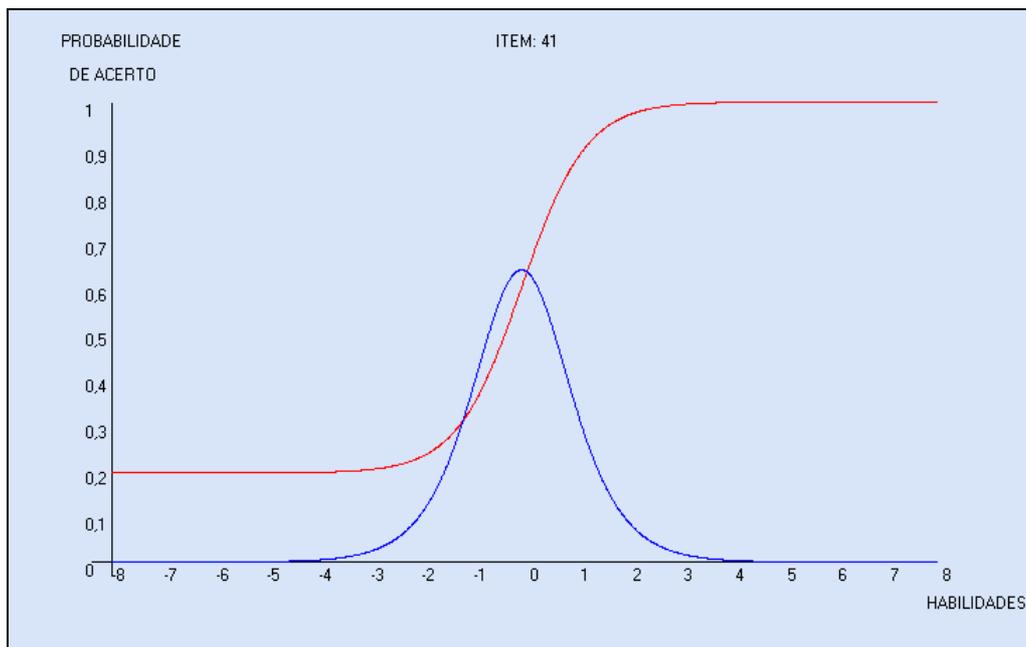


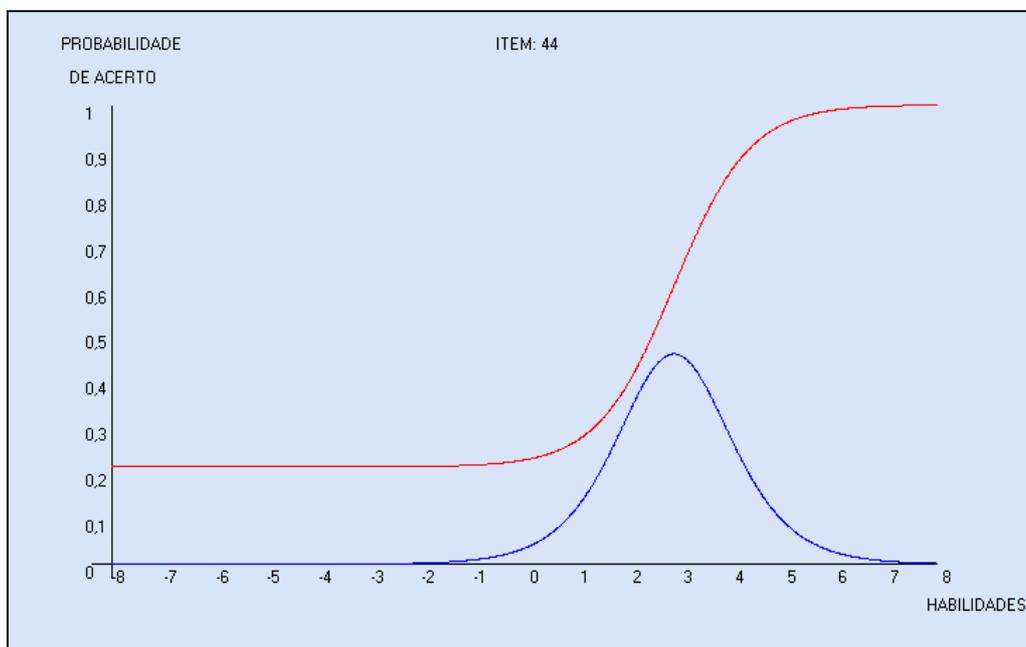
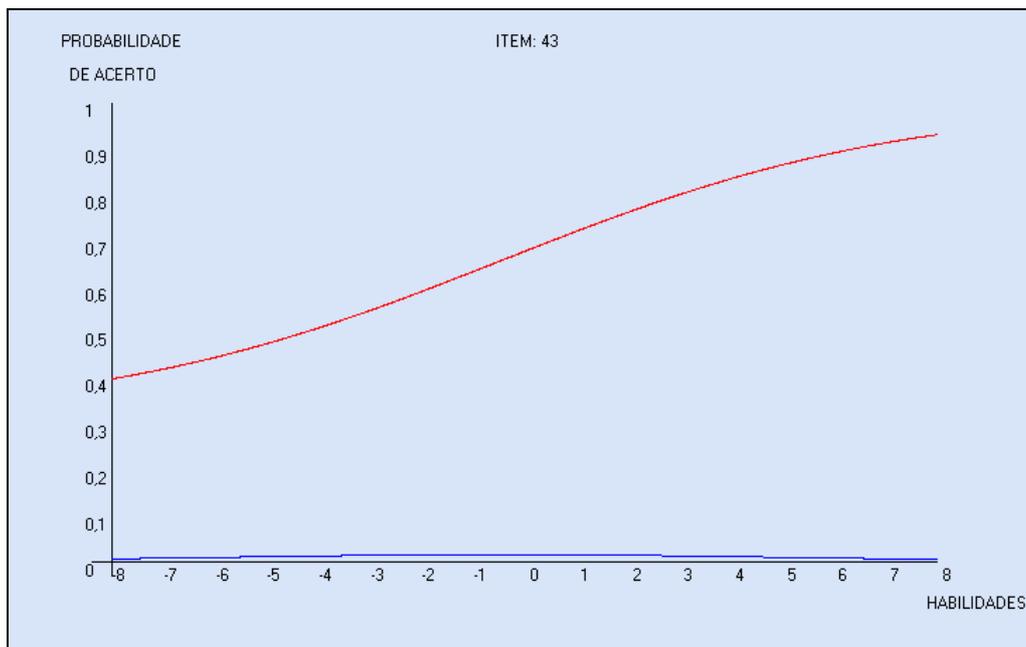


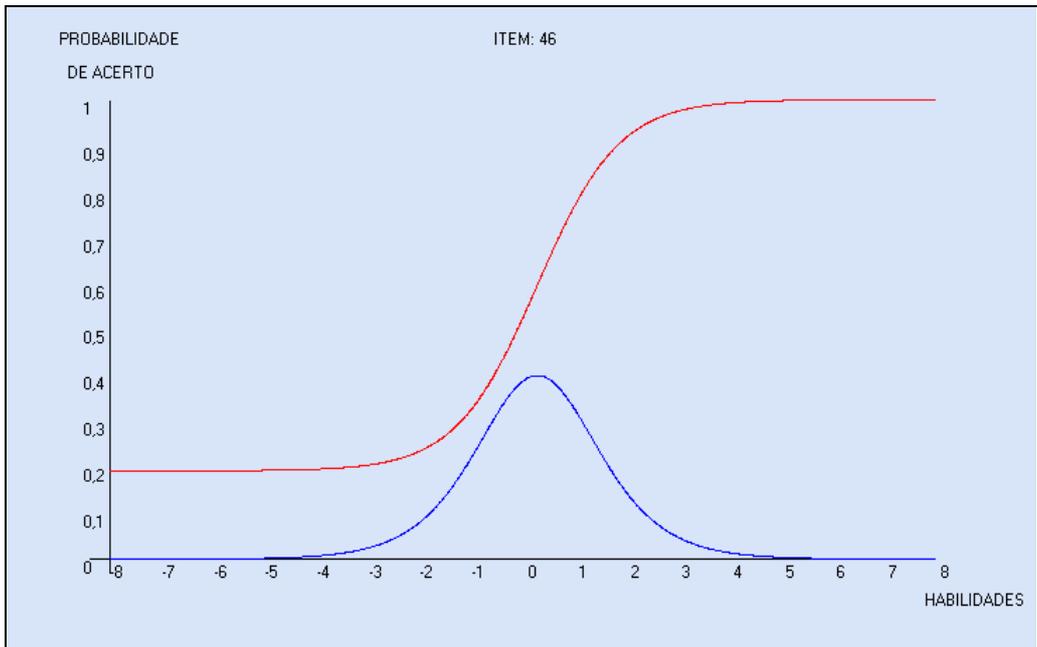
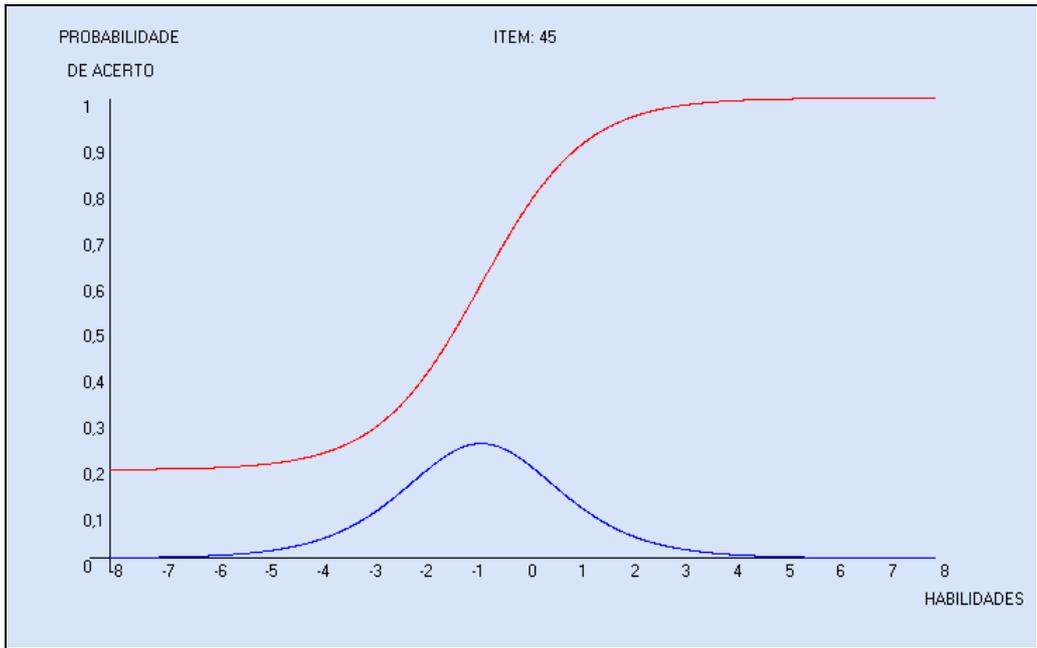


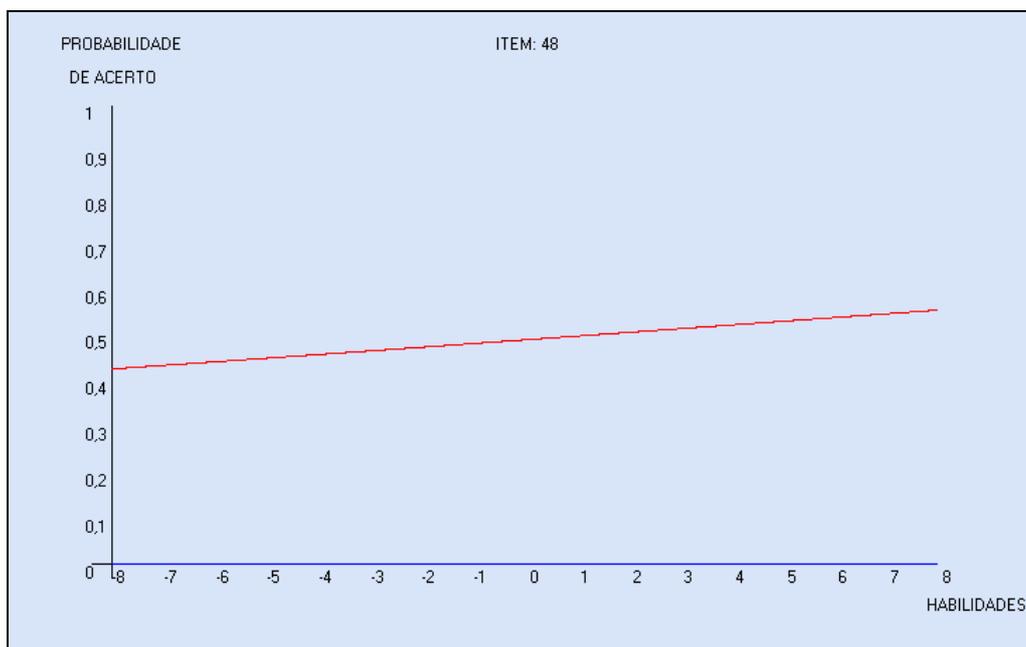
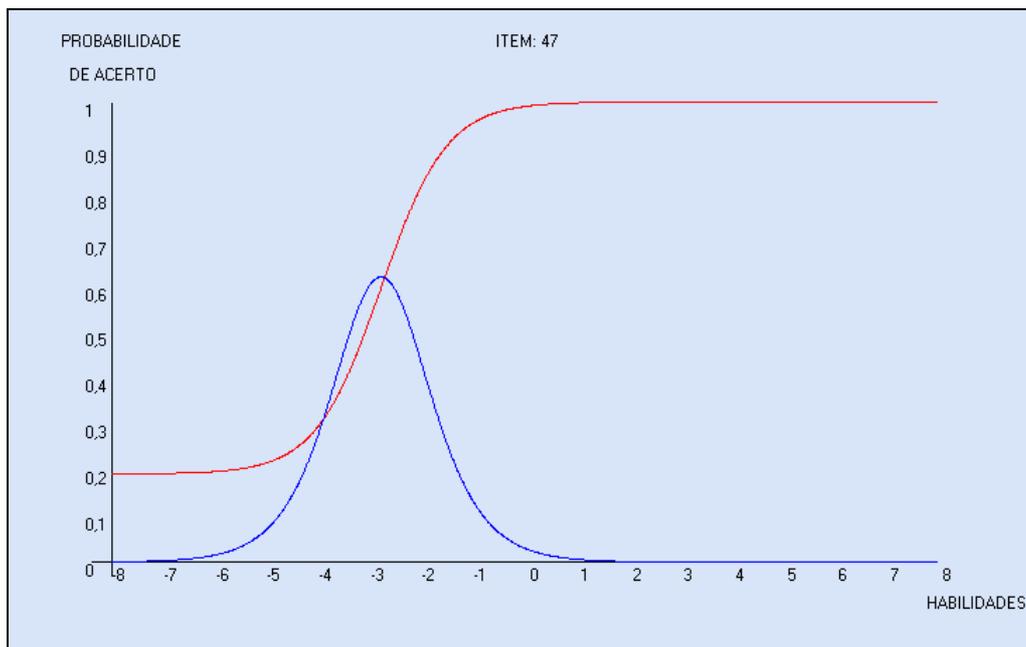


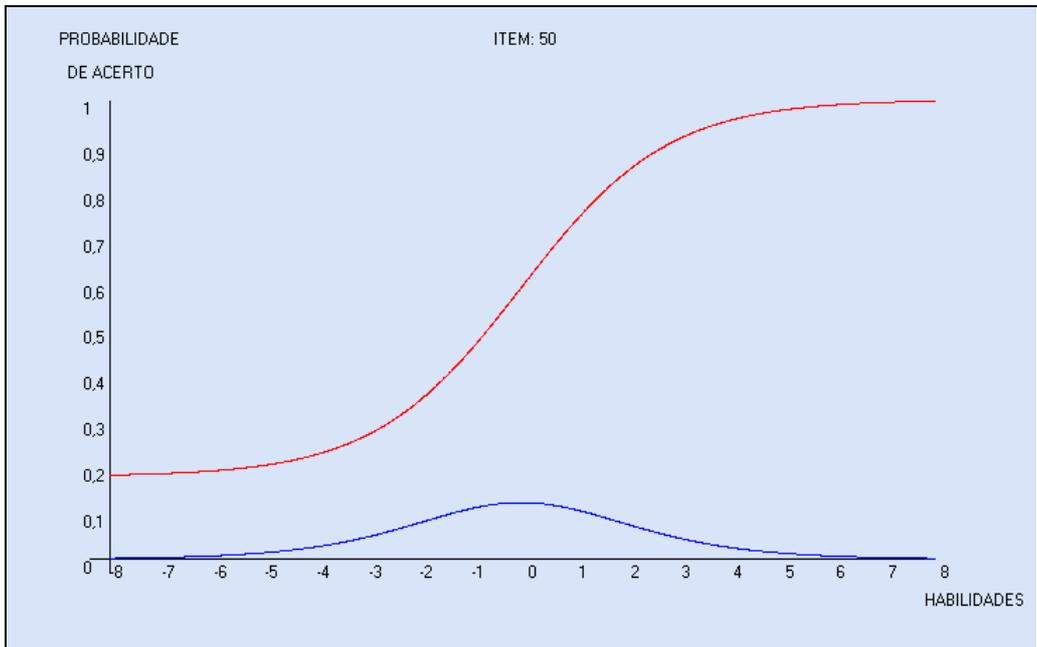
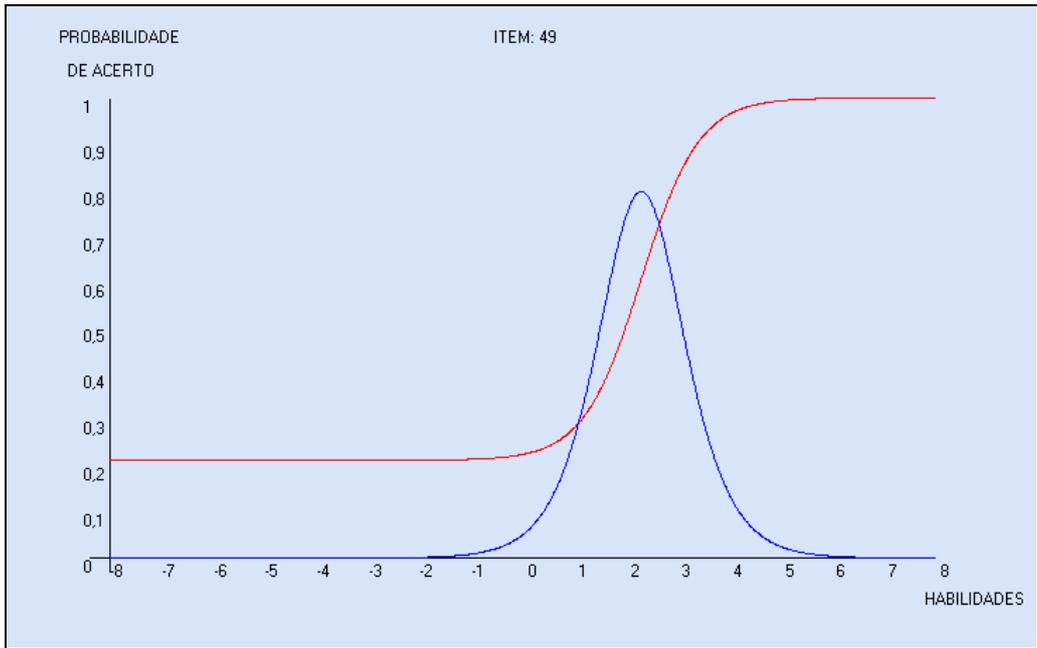


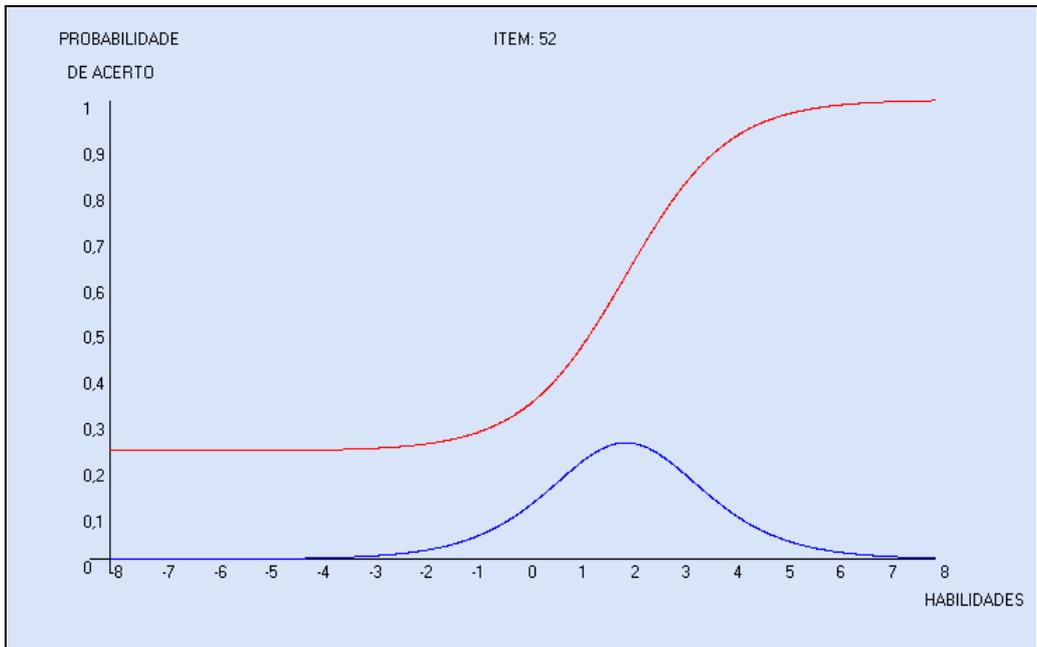
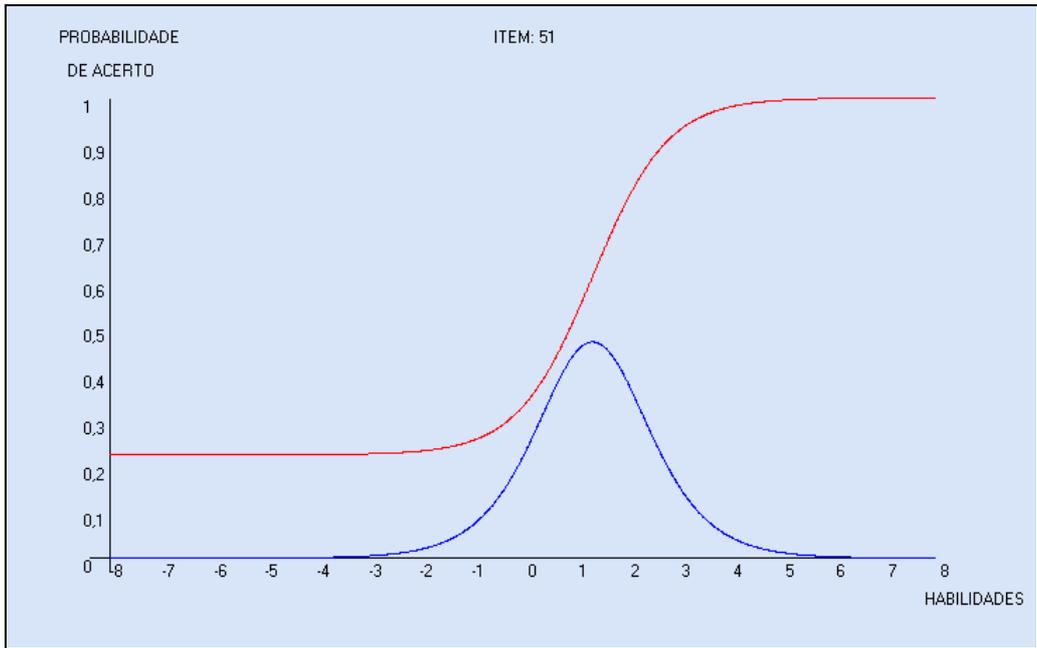


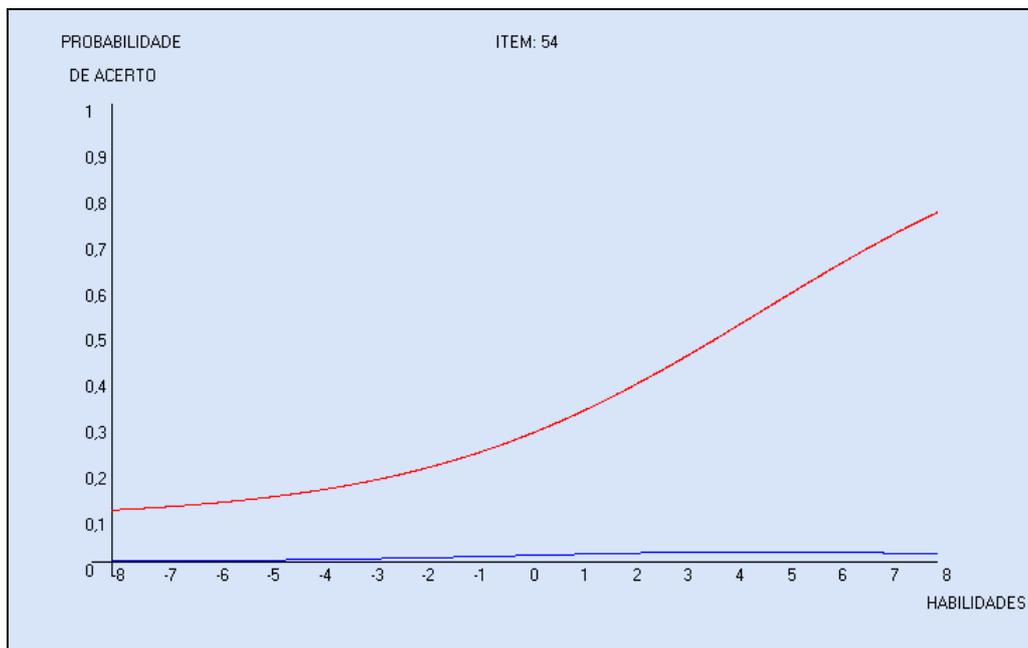
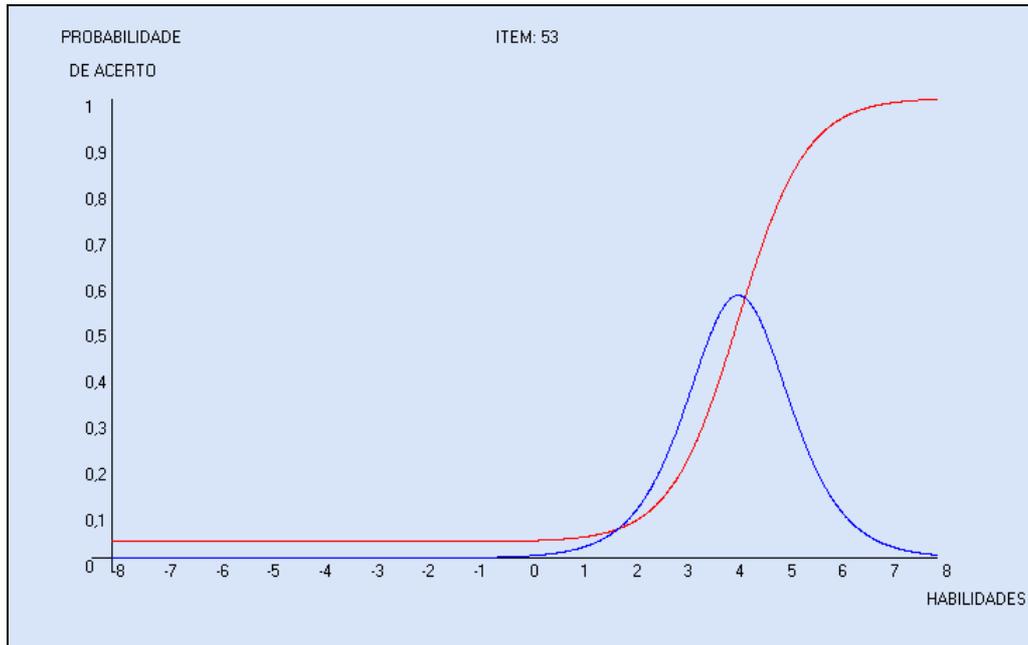


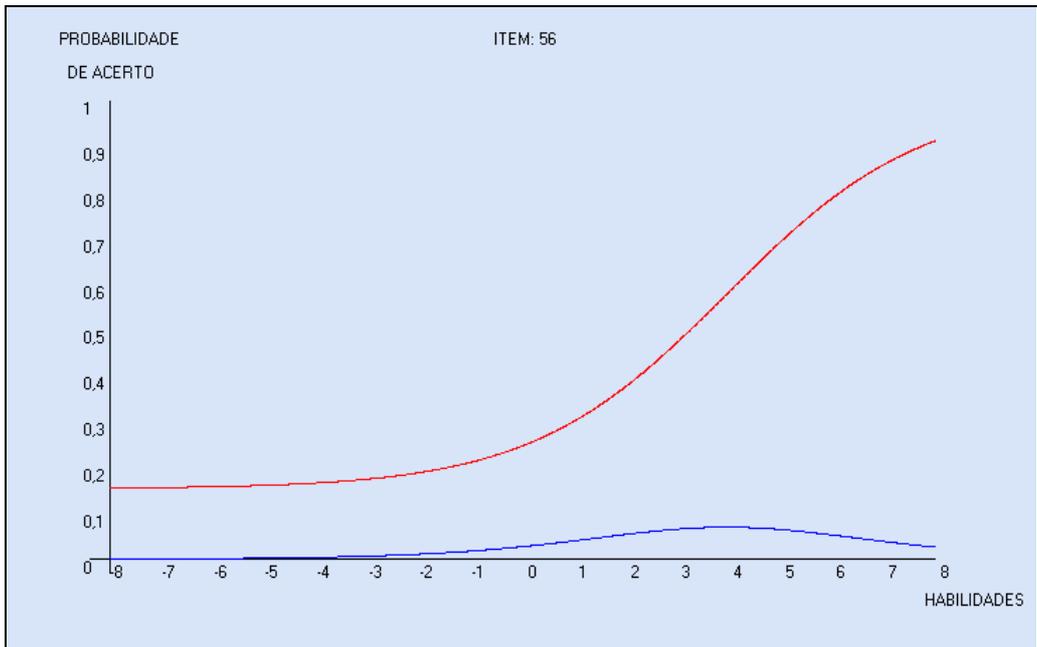
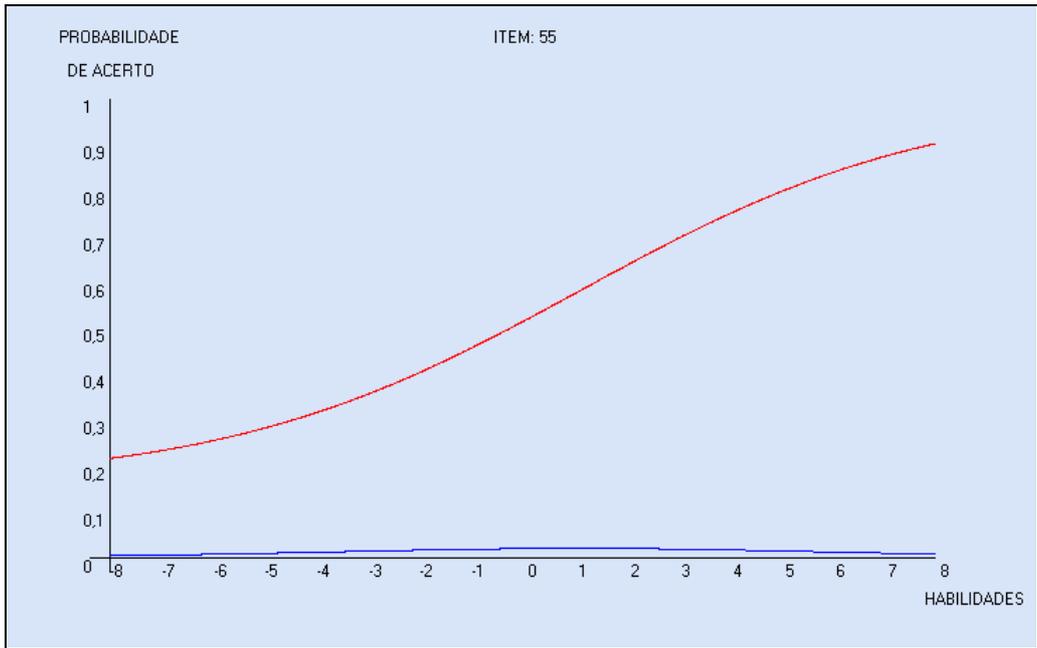


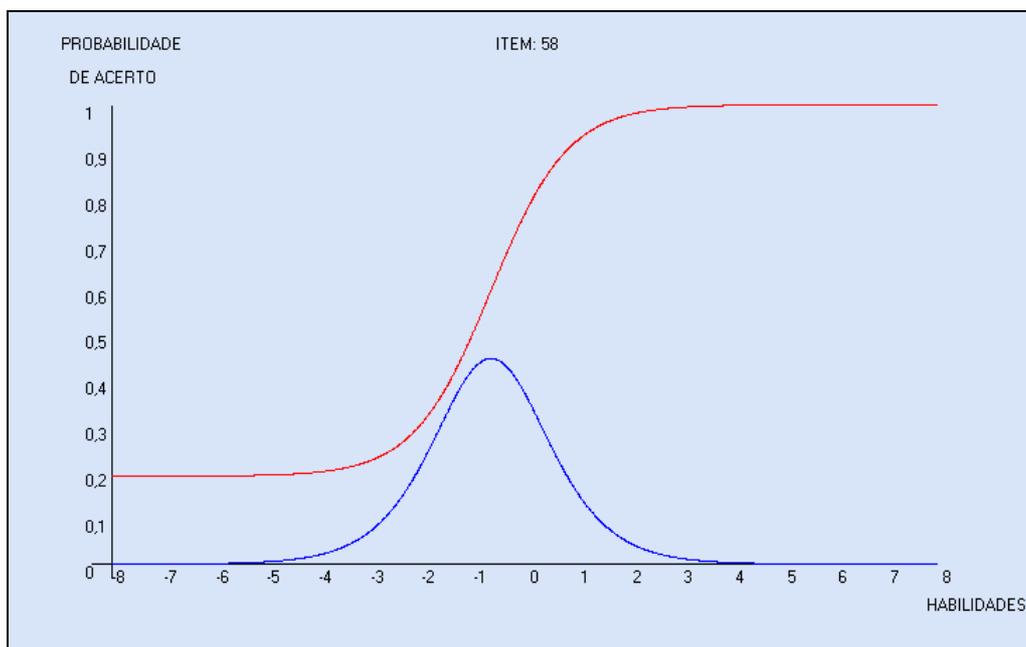
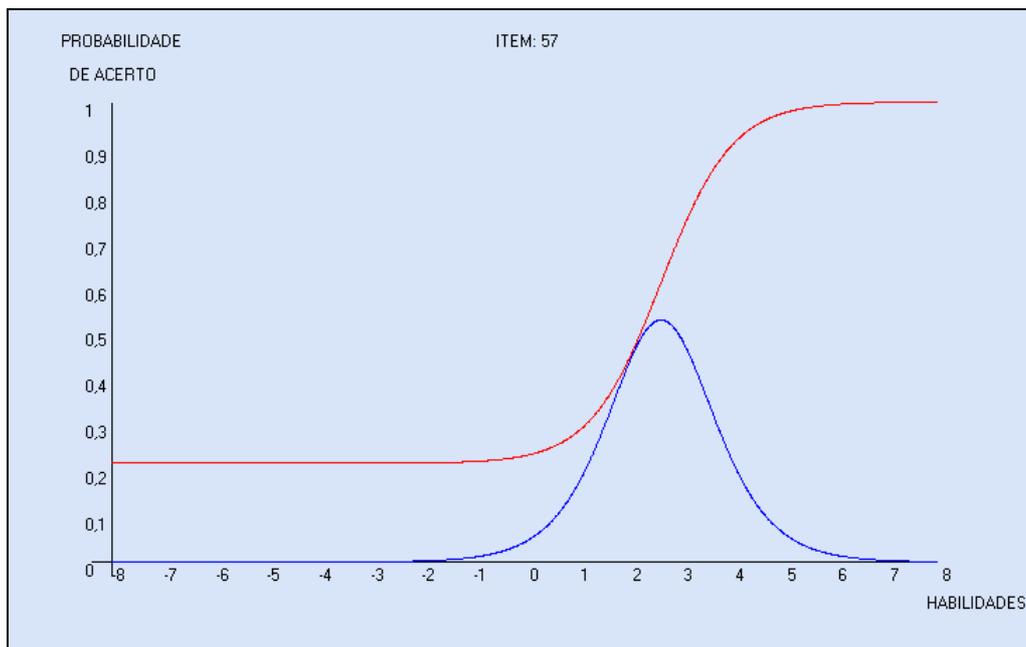


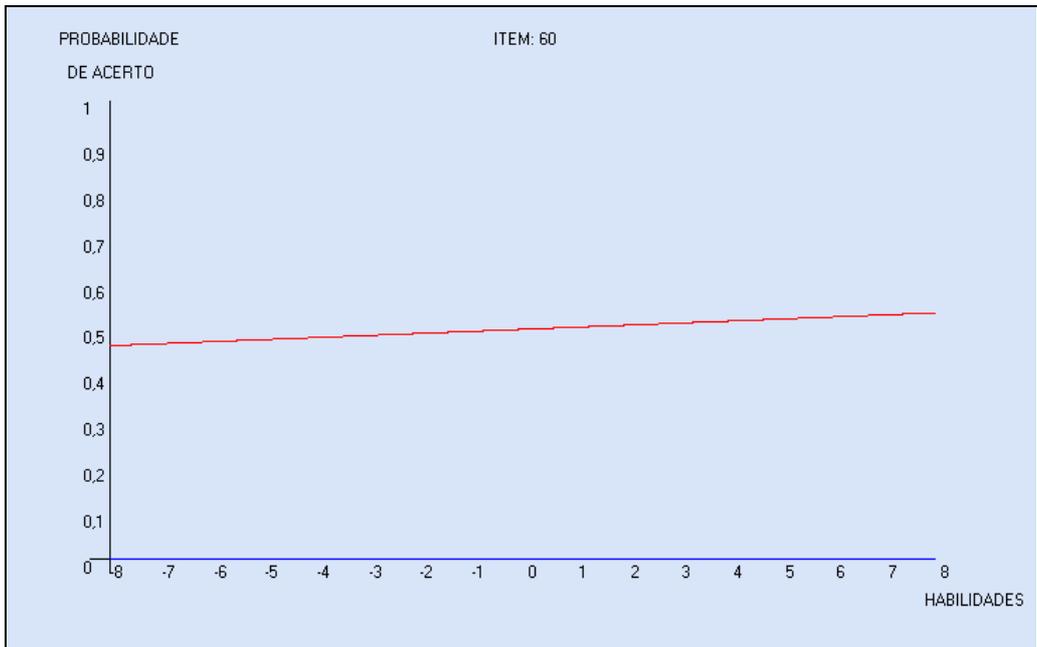
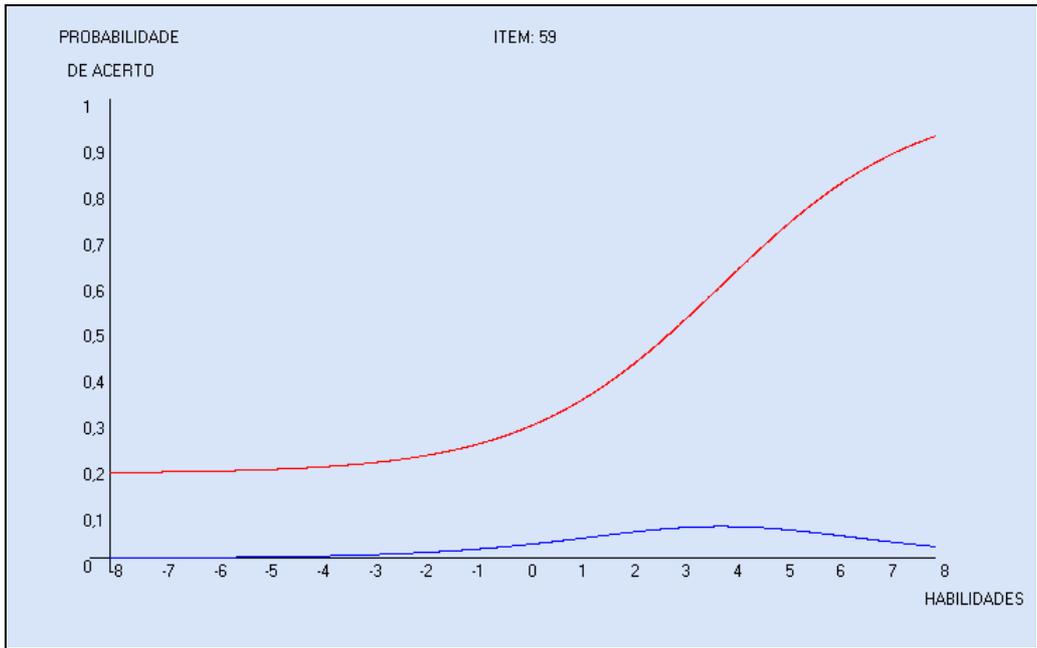












## APÊNDICE 10

**TABELA DE CLASSIFICAÇÃO FEITA PELO MÉTODO DISCRIMINANTE QUADRÁTICO  
USANDO PROBABILIDADE *A PRIORI* IGUAIS PARA OS GRUPOS DE CLASSIFICAÇÃO**

Grupos	Grupos			Total
	Abandonaram ou Trancaram o Curso	Não concluíram o Curso no Tempo Certo	Concluíram o Curso no Tempo Certo	
Abandonaram ou Trancaram o Curso	78	9	2	89
Não concluíram o Curso no Tempo Certo	8	37	8	53
Concluíram o Curso no Tempo certo	0	1	12	13
<b>Total</b>	86	47	22	<b>155</b>

Percentual de Classificação Correta: 81,94%

**TABELA DE CLASSIFICAÇÃO FEITA PELO MÉTODO DISCRIMINANTE QUADRÁTICO  
USANDO PROBABILIDADE *A PRIORI* DIFERENTES, PROPORCIONAIS AOS GRUPOS DE CLASSIFICAÇÃO**

Grupos	Grupos			Total
	Abandonaram ou Trancaram o Curso	Não concluíram o Curso no Tempo Certo	Concluíram o Curso no Tempo Certo	
Abandonaram ou Trancaram o Curso	81	8	0	89
Não concluíram o Curso no Tempo Certo	9	41	3	53
Concluíram o Curso no Tempo certo	0	3	10	13
<b>Total</b>	90	52	13	<b>155</b>

Percentual de Classificação Correta: 85,16%

**TABELA DE CLASSIFICAÇÃO FEITA POR REDES NEURAS USANDO PROBABILIDADE *A PRIORI* IGUAIS PARA OS GRUPOS DE CLASSIFICAÇÃO**

Grupos	Grupos			Total
	Abandonaram ou Trancaram o Curso	Não concluíram o Curso no Tempo Certo	Concluíram o Curso no Tempo Certo	
Abandonaram ou Trancaram o Curso	53	26	10	89
Não concluíram o Curso no Tempo Certo	13	27	13	53
Concluíram o Curso no Tempo certo	1	8	4	13
<b>Total</b>	67	61	27	<b>155</b>

Percentual de Classificação Correta: 54,19%

**TABELA DE CLASSIFICAÇÃO FEITA POR REDES NEURAS USANDO PROBABILIDADE *A PRIORI* DIFERENTES, PROPORCIONAIS AOS GRUPOS DE CLASSIFICAÇÃO**

Grupos	Grupos			Total
	Abandonaram ou Trancaram o Curso	Não concluíram o Curso no Tempo Certo	Concluíram o Curso no Tempo Certo	
Abandonaram ou Trancaram o Curso	84	5	0	89
Não concluíram o Curso no Tempo Certo	43	10	0	53
Concluíram o Curso no Tempo certo	6	7	0	13
<b>Total</b>	133	22	0	<b>155</b>

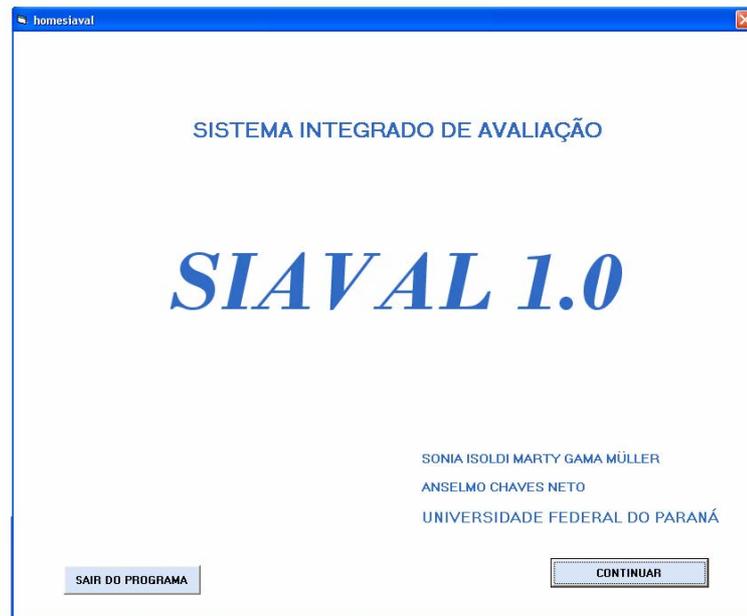
Percentual de Classificação Correta: 60,65%

**TABELA DE COMPARAÇÃO ENTRE OS MÉTODOS NA CLASSIFICAÇÃO DO GRUPO DE TESTE**

Alunos do Grupo de Teste	Método					Classificação Correta
	Fisher	Discr. Quadrático(Prop. <i>a priori</i> iguais)	Discr. Quadrático (Prop. <i>a priori</i> diferentes)	Redes Neurais(Prop. <i>a priori</i> iguais)	Redes Neurais (Prop. <i>a priori</i> iguais)	
<b>1</b>	2	1	1	2	1	1
<b>2</b>	1	1	1	1	1	1
<b>3</b>	1	2	2	1	1	1
<b>4</b>	3	1	1	1	1	2
<b>5</b>	2	1	1	1	1	2
<b>6</b>	2	2	2	3	2	2
<b>7</b>	3	1	1	2	1	3
<b>8</b>	2	2	2	2	1	3
<b>Percentual de Acertos</b>	62,50%	37,50%	37,50%	25,00%	50,00%	-

## APÊNDICE 11

### 1. TELA INICIAL



### 2. TELA DE ESCOLHA DE MÉTODOS



### 3. TELA DE ENTRADA DE DADOS PARA ANÁLISE DO INSTRUMENTO

datasiaval1

## ENTRADA DE DADOS

ENTRE COM O NÚMERO DE ALTERNATIVAS

ENTRE OS DADOS DO INSTRUMENTO COLOCANDO OS RESPONDENTES NAS LINHAS E OS ITENS NAS COLUNAS

	A	B	C	D	E	F	G	H	I	J
1	1	1	0	0	0	1	1	1	0	1
2	1	1	1	1	0	0	1	1	1	0
3	1	1	0	0	0	0	0	1	0	0
4	0	0	1	0	0	1	1	1	0	0
5	1	1	1	0	0	0	1	1	0	0
6	1	1	0	0	0	0	0	1	0	0
7	0	1	1	0	0	1	1	1	1	0
8	0	1	1	1	0	1	1	1	1	0
9	0	1	0	0	0	0	0	0	0	0
10	0	1	1	0	0	0	1	1	1	0
11	1	0	1	0	1	1	1	1	0	1
12	0	1	1	0	0	1	1	1	0	1
13	0	1	0	0	0	1	1	1	0	0
14	0	1	0	0	0	0	1	1	0	1
15	0	1	0	0	0	0	1	1	0	0
16	1	1	1	1	0	1	1	1	1	0
17	1	1	1	0	0	0	0	0	0	0
18	0	1	0	0	0	0	1	1	1	1
19	1	1	0	0	0	1	1	1	1	0
20	1	0	1	0	0	1	1	1	1	0
21	1	1	1	0	0	0	0	0	1	0

Planilha1

SAIR DO PROGRAMA      VOLTAR A PÁGINA ANTERIOR      CONTINUAR

### 4. TELA DE OPÇÕES DE ANÁLISE DO INSTRUMENTO

## ANÁLISE DO INSTRUMENTO

TEORIA DA CLÁSSICA DA MEDIDA:

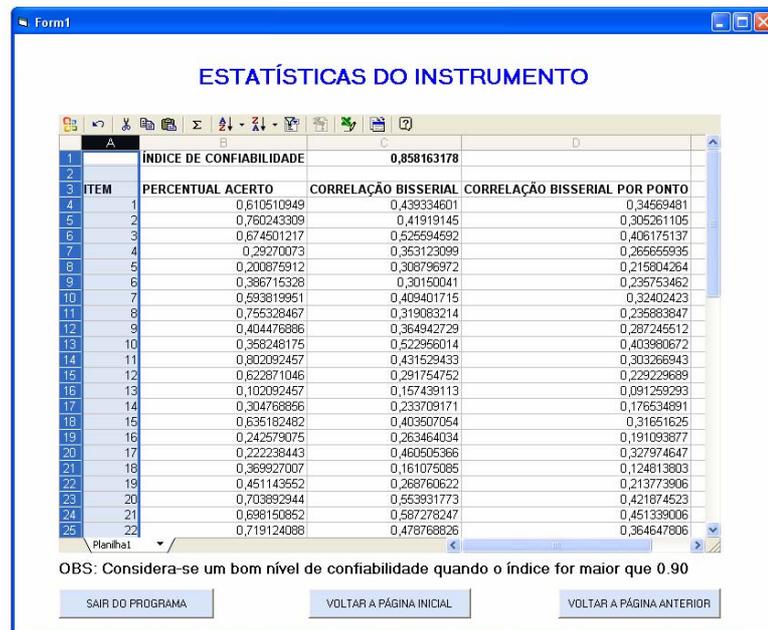
ESTATÍSTICAS DO INSTRUMENTO      ESTATÍSTICAS DESCRITIVA DO NÚMERO DE ACERTOS DOS RESPONDENTES

TEORIA DA RESPOSTA AO ITEM:

DOIS PARÂMETROS      TRÊS PARÂMETROS

VOLTAR A PÁGINA ANTERIOR

## 5. TELA RESULTADOS DAS ESTATÍSTICAS DO INSTRUMENTO



## 6. TELA DE RESULTADOS DAS ESTATÍSTICAS DESCRITIVAS DO NÚMERO DE ACERTOS DOS RESPONDENTES



## 7. TELA DE RESULTADO DA TRI PARA DOIS E TRÊS PARÂMETROS

Resposta1

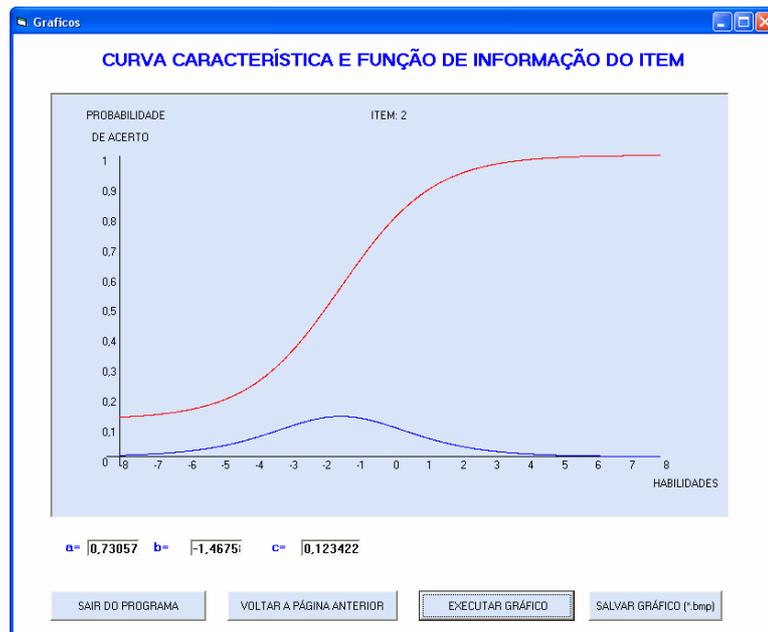
### PARÂMETROS DOS ITENS

ITEM	DISCRIMINAÇÃO	DIFICULDADE	ACERTO CASUAL	CONVERGÊNCIA
1	1,227463126	-0,106790341	0,186664388	0,000946403
2	1,188680887	-1,589558244	0,234724507	0,007148623
3	1,541212559	-0,715908408	0,190667868	0,003673375
4	0,938726664	2,620636258	0,172999412	0,017179966
5	0,911062598	3,659352779	0,04398001	0,017412424
6	0,824010432	2,309063275	0,234334633	0,011739254
7	1,154858437	0,069945052	0,190518379	0,001358621
8	0,928511739	-1,210773945	0,38109633	0,026247263
9	0,986793671	1,911225319	0,224011257	0,000513509
10	1,479663763	2,3699519	0,21531108	0,004276752
11	1,231236935	-2,10743928	0,171964005	0,007195526
12	0,822119176	-0,216749236	0,192125052	0,002463073
13	0,475415408	4,98881451	0,059110291	0,010530949
14	0,593807399	3,134949207	0,137911052	0,032776594
15	1,131353736	-0,340387613	0,184837058	0,001296878
16	0,81477493	3,107367516	0,060064718	0,011796998
17	1,282399416	3,736605167	0,043019019	0,004212379
18	0,529474676	3,149448633	0,267110646	0,041624784
19	0,738438487	1,719369888	0,209861875	0,011566924
20	1,624670148	-1,011449814	0,209548862	0,003889627
21	1,757871151	-0,971271515	0,204570502	0,004980087
22	1,391178012	-1,172894478	0,201346278	0,008264688
23	1,513678074	-2,96764183	0,183358878	0,009607553
24	1,461733341	-0,780088425	0,209864882	0,005498827

Planilha1

SAIR DO PROGRAMA    VOLTAR A PÁGINA ANTERIOR    GRÁFICOS DOS ITENS    ESTIMAR ESCORES VERDADEIROS

## 8. TELA DO GRÁFICO DA CURVA CARACTERÍSTICA DO ITEM E DA FUNÇÃO DE INFORMAÇÃO DO ITEM



## 9. TELA DO RANQUEAMENTO FEITO POR ESCORES VERDADEIROS

Resposta11

### RANKING DOS RESPONDENTES - ESCORES VERDADEIROS

RANKING	RESPONDENTE	ESCORE VERDADEIRO
1	9923	44,55149
2	6665	44,51596
3	9730	44,43324
4	8393	43,79565
5	3773	43,45608
6	15646	43,19175
7	8821	43,12932
8	5010	43,04462
9	10343	42,94034
10	8271	42,90462
11	17667	42,7632
12	7705	42,68624
13	17513	42,61797
14	17795	42,48073
15	19913	42,40113
16	11046	42,31579
17	709	42,31251
18	18860	42,30456
19	9520	42,30401
20	10772	42,25973
21	3802	42,21674
22	4885	42,21309
23	11267	42,19124
24	5295	42,04691
25	14461	42,03141
26	13197	41,95235
27	18508	41,79383
28	18653	41,65104
29	20496	41,65104

Planilha1

SAIR DO PROGRAMA    VOLTAR A PÁGINA INICIAL    VOLTAR A PÁGINA ANTERIOR    ESTATÍSTICA DAS HABILIDADES

## 10. TELA DOS RESULTADOS DAS ESTATÍSTICAS DOS ESCORES VERDADEIROS



## 11. TELA DE ENTRADA DE DADOS PARA ANÁLISE FATORIAL

**ANÁLISE FATORIAL**

ENTRE COM AS AMOSTRAS NAS LINHAS E AS VARIÁVEIS NAS COLUNAS

	A	B	C	D	E	F	G	H	I	J	K
1	70,0	92,9	90,1	90,0							
2	75,0	99,2	62,1	93,8							
3	54,0	56,4	88,4	72,5							
4	70,0	84,8	64,4	72,5							
5	88,0	94,1	77,5	93,3							
6	70,0	93,1	38,9	91,0							
7	43,8	84,8	76,0	75,0							
8	70,0	93,1	80,3	90,5							
9	76,3	90,0	72,7	90,0							
10	80,3	78,8	90,2	92,0							
11	85,0	23,1	67,1	70,0							
12	79,3	93,1	87,7	93,3							
13	81,3	63,7	75,4	83,8							
14	76,0	91,8	88,7	91,3							
15	89,0	65,4	77,7	88,8							
16	80,8	90,9	90,1	95,0							
17	82,8	91,2	88,6	93,0							
18	77,8	84,8	77,4	91,3							
19	70,0	72,5	56,8	90,0							
20	85,8	92,0	90,3	91,0							
21	70,0	90,7	52,0	90,0							
22	50,0	80,2	86,9	71,3							
23	62,5	21,6	62,0	77,5							
24	73,8	92,7	85,6	92,5							
25	77,6	96,8	85,7	91,3							

Planilha1

SAIR DO PROGRAMA      VOLTAR A PÁGINA ANTERIOR      EXECUTAR

## 12. TELA DA TABELA DE RESULTADOS DA ANÁLISE FATORIAL

**ANÁLISE FATORIAL**

	A	B	C	D	E	F	G	H	I	J	K
1	<b>variável</b>	<b>Fator 1</b>	<b>Fator 2</b>	<b>Fator 3</b>	<b>comunal</b>	<b>var. específicas</b>					
2	1	0,120464	0,102162	0,969455	0,964791	0,035209					
3	2	0,938372	0,143984	-0,001173	0,901274	0,098726					
4	3	0,108815	0,988895	0,093268	0,998057	0,001943					
5	4	0,798869	0,054077	0,453516	0,846792	0,153208					
6	<b>autovalor</b>	1,545085	1,154219	1,01161							
7	<b>var. acum</b>	38,62713	67,48261	92,77286							

Planilha1

VOLTAR A PÁGINA INICIAL      VOLTAR A PÁGINA ANTERIOR      RANQUEAMENTO      ESCORES FATORIAIS

### 13. TELA DO RANQUEAMENTO FEITO POR ANÁLISE FATORIAL

Resposta3

#### ANÁLISE FATORIAL

	A	B	C	D	E	F	G
1	1	16	93,45166				
2	2	5	93,08323				
3	3	12	92,418				
4	4	25	92,30073				
5	5	17	91,83542				
6	6	50	91,77469				
7	7	45	91,66812				
8	8	20	91,62106				
9	9	14	90,69973				
10	10	24	90,36295				
11	11	1	89,9647				
12	12	27	89,68421				
13	13	39	89,65012				
14	14	33	89,31323				
15	15	52	89,18533				
16	16	10	88,55845				
17	17	32	88,31567				
18	18	2	87,58722				
19	19	34	87,14534				
20	20	31	86,97268				
21	21	8	86,50232				
22	22	9	85,09219				
23	23	18	84,87932				
24	24	40	84,40914				
25	25	15	82,58424				
26	26	28	80,24485				
27	27	41	80,05143				

Planilha1

VOLTAR A PÁGINA INICIAL      VOLTAR A PÁGINA ANTERIOR      ESCORES FATORIAIS

### 14. TELA DE ENTRADA DE DADOS PARA RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO

datasiava14

#### RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO

1. Coloque os dados de modo que cada respondente represente  
 2. Coloque na coluna A o número correspondente ao grupo no qual o respondente pertence (1,2,3...) e nas colunas seguintes o valor das variáveis.

Dados para Análise

	A	B	C	D	E	F
1	GRUPOS	VAR 1	VAR 2	VAR 3	VAR 4	
2	1	70,0	92,9	90,1	90,0	
3	1	75,0	99,2	62,1	93,8	
4	1	88,0	94,1	77,5	93,3	
5	1	70,0	93,1	80,3	90,5	
6	1	78,3	90,0	72,7	90,0	
7	1	80,3	78,8	90,2	92,0	
8	1	79,3	93,1	87,7	93,3	
9	1	76,0	91,8	88,7	91,3	
10	1	80,8	90,9	90,1	95,0	
11	1	82,8	91,2	88,6	93,0	
12	1	85,8	92,0	90,3	91,0	
13	1	73,8	92,7	85,6	92,5	
14	1	77,6	96,8	85,7	91,3	
15	1	80,0	93,7	83,2	90,0	
16	1	71,5	93,1	78,5	91,5	
17	1	75,0	93,1	69,7	95,5	

Planilha1

Dados para Reconhecimento

	A	B	C	D	E
1	VAR 1	VAR 2	VAR 3	...	
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					

Planilha1

EQUAÇÕES LINEARES:

OBS: AS CLASSIFICAÇÕES PODEM SER VISTAS NA ÚLTIMA COLUNA DESTA PLANILHA

1    2    Procedimento de Retenção de Lachenbruch

SAIR DO PROGRAMA      VOLTAR A PÁGINA INICIAL      EXECUTAR      TABELA DE CONTINGÊNCIA

## 15. TELA DE SAÍDA DE DADOS PARA RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO

datasiaval4

### RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO

1. Coloque os dados de modo que cada respondente represente  
2. Coloque na coluna A o número correspondente ao grupo no qual o respondente pertence (1,2,3...) e nas colunas seguintes o valor das variáveis.

Dados para Análise

GRUPOS	VAR 1	VAR 2	VAR 3	VAR 4
1	70,0	92,9	90,1	90,0
1	75,0	99,2	62,1	93,8
1	88,0	94,1	77,5	93,3
1	70,0	93,1	80,3	90,5
1	78,3	90,0	72,7	90,0
1	80,3	78,8	90,2	92,0
1	79,3	93,1	87,7	93,3
1	76,0	91,8	88,7	91,3
1	80,8	90,9	90,1	95,0
1	82,8	91,2	88,6	93,0
1	85,8	92,0	90,3	91,0
1	73,8	92,7	85,6	92,5
1	77,6	96,8	85,7	91,3
1	80,0	93,7	83,2	90,0
1	71,5	93,1	78,5	91,5
1	75,0	93,1	69,7	95,5

Dados para Reconhecimento

VAR 1	VAR 2	VAR 3	...	VAR 4	grupo
70,0	92,9	90,1		90,0	1,0

EQUAÇÕES LINEARES:  
 $y_1 = +0.05707 x_1 + 0.05228 x_2 + 0.07674 x_3 + 0.10878 x_4$   
 $y_2 = +0.07856 x_1 - 0.01014 x_2 + 0.02487 x_3 - 0.12157 x_4$

OBS: AS CLASSIFICAÇÕES PODEM SER VISTAS NA ÚLTIMA COLUNA DESTA PLANILHA

1 2 Procedimento de Retenção de Lachenbruch

SAIR DO PROGRAMA VOLTAR A PÁGINA INICIAL EXECUTAR TABELA DE CONTINGÊNCIA

## 16. TELA DE RESULTADO DA TABELA DE CLASSIFICAÇÃO

Form1

### TABELA DE CLASSIFICAÇÃO

Grupo	1	2	3	Total
1	21	1	0	22
2	3	11	0	14
3	2	6	8	16
Total	26	18	8	
Percentual Classif	86,53846			

### TABELA DE CLASSIFICAÇÃO LACHENBRUCH

Grupo	1	2	3	Total
1	21	1	0	22
2	3	11	0	14
3	2	6	8	16
Total	26	18	8	
Percentual	76,92308			

### CLASSIFICAÇÃO LACHENBRUCH POR AMOSTRA

	B	C	D	E	F
1	92	90	90	1	1
2	99	62	93	1	2
3	94	77	93	1	1
4	93	80	90	1	1
5	90	72	90	1	1
6	78	90	92	1	1
7	93	87	93	1	1
8	91	88	91	1	1
9	90	90	90	1	1

SAIR DO PROGRAMA VOLTAR À PAGINA ANTERIOR