

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**AVALIAÇÃO DE DESEMPENHO DE SISTEMAS  
EDUCACIONAIS: UMA ABORDAGEM UTILIZANDO  
CONJUNTOS DIFUSOS**

**Tese submetida à Universidade Federal de Santa Catarina para obtenção do  
título de Doutor em Engenharia de Produção**

**Rogério Cid Bastos**

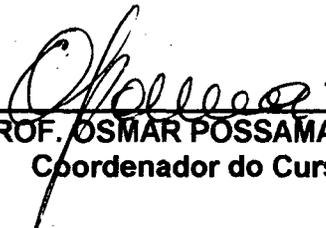
**Orientador: Prof. Ricardo Miranda Barcia, Ph.D.**

**Florianópolis, Setembro de 1994.**

**AVALIAÇÃO DE DESEMPENHO DE SISTEMAS EDUCACIONAIS: UMA  
ABORDAGEM UTILIZANDO CONJUNTOS DIFUSOS**

**ROGÉRIO CID BASTOS**

**Esta tese foi julgada adequada para obtenção do título de DOUTOR EM  
ENGENHARIA DE PRODUÇÃO e aprovada em sua forma final pelo Programa  
de Pós-Graduação em Engenharia de Produção.**

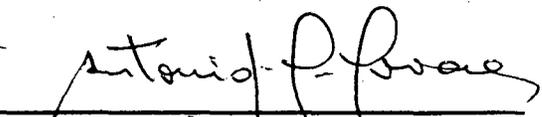
  
\_\_\_\_\_  
**PROF. OSMAR POSSAMAI, Dr.**  
Coordenador do Curso

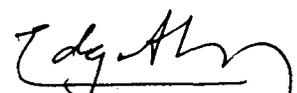
**BANCA EXAMINADORA:**

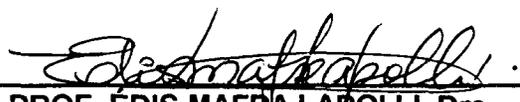
  
\_\_\_\_\_  
**PROF. RICARDO M. BARCIA, Ph.D.**  
Orientador

  
\_\_\_\_\_  
**PROF. REINALDO C. SOUZA, Ph.D.**  
Examinador Externo

  
\_\_\_\_\_  
**PROF. ISRAEL BRUNSTEIN, Dr.**  
Examinador Externo

  
\_\_\_\_\_  
**PROF. ANTÔNIO G. NOVAES, Dr.**

  
\_\_\_\_\_  
**PROF. EDGAR A. LANZER, Ph.D.**

  
\_\_\_\_\_  
**PROF. EDIS MAFRA LAPOLLI, Dra.**  
Moderadora

*À Lia*

*e*

*ao David*

## **AGRADECIMENTOS**

Ao Prof. Ricardo Miranda Barcia agradeço a orientação recebida para a elaboração deste trabalho. Tive acesso, através desta orientação, ao tema central do trabalho - Conjuntos Difusos. Os processos e relações existentes entre pessoas são de intensidade e, nessas relações, algumas pessoas merecem pertinência máxima em se tratando de reconhecimento de competência e amizade. O Prof. Barcia é, sem dúvida, um claro exemplo desses casos.

A Lia, que é outro exemplo de pessoa a quem atribuo pertinência máxima.

Aos amigos Flávio, Édis, Fernando, Ana e Masanao, agradeço a amizade e o companheirismo que demonstraram durante a elaboração do trabalho. Não está faltando o Alceu - ele é que nos faz falta.

Aos Professores Edgar A. Lanzer e Antônio G. Novaes, agradeço as inúmeras vezes que se dispuseram a discutir as idéias contidas neste trabalho. Agradeço, também, as várias sugestões e contribuições que recebi desses professores para a elaboração desta tese.

Aos Professores Reinaldo Castro Souza e Israel Brunstein, Examinadores Externos, pelas valiosas sugestões recebidas para melhoria deste trabalho.

Ao Professor Luís A. Valadares Tavares, pelas discussões realizadas no início deste trabalho e, também, por proporcionar a utilização dos dados pertencentes a base de dados do Gabinete de Estudos e Planeamento do Ministério da Educação de Portugal.

Aos professores e colegas do Departamento de Informática e de Estatística agradeço o convívio e o estímulo para o desenvolvimento de trabalhos científicos.

Ao bolsista Rodrigo Becke Cabral agradeço a dedicação com que trabalhou comigo durante o desenvolvimento deste trabalho.

Agradeço à Professora Édis Mafra Lapolli, moderadora, pela sua atuação e condução dos trabalhos de apresentação da tese.

## RESUMO

Neste trabalho são discutidos modelos para avaliação de desempenho de sistemas, mais especificamente, sistemas educacionais.

Um sistema educacional é um sistema complexo. Em sistemas educacionais, um elemento marcante é a presença viva de relações humanas. Modelos tradicionais de análise não são tão adequados na caracterização dessas relações e, conseqüentemente, a aplicação desses modelos se torna problemática.

A teoria da lógica difusa é especialmente aplicável a modelos nos quais é grande a presença de incertezas próprias do pensamento humano, processos de raciocínio, processos cognitivos, processos de percepção ou, de forma mais ampla, informação cognitiva. A incorporação destes elementos subjetivos na análise torna-se possível quando da utilização dos conceitos de conjuntos difusos.

Este trabalho apresenta, numa primeira etapa, uma descrição de modelos de análise tradicionais, discutindo-se os requisitos necessários a sua aplicação. Posteriormente, é apresentada uma análise da teoria dos conjuntos difusos, enfatizando-se as técnicas que podem ser utilizadas com sucesso na análise de desempenho de sistemas.

Um modelo teórico para o tratamento de questões relativas a escolha qualitativa, baseado na lógica difusa, é proposto.

Através de aplicações procura-se demonstrar a viabilidade da utilização dos modelos de análise de dados difusos. Entre estas aplicações, um estudo de caso, utilizando dados reais, é realizado. Neste estudo de caso demonstra-se como políticas educacionais podem ser utilizadas para aferição de

desempenho de um sistema educacional. Nestas aplicações realizadas, modelos são construídos para avaliar como fatores relacionados as características organizacionais do sistema, como por exemplo, localização de um estabelecimento de ensino ou seu estado de conservação podem vir a interferir no desempenho do sistema.

## **ABSTRACT**

In this work models for evaluating educational systems performance are approached.

In educational systems human relations are a major feature. Traditional models are not well suited for describing these systems. Fuzzy Logic is another instrument for modeling systems that contains uncertainty that are present both in the human thinking and human relations.

This work initially describes traditional analysis models. Following the fuzzy logic aspects that may be successfully usually for analyzing systems performance are presented. A model for dealing with qualitative choice models based on fuzzy logic is proposed.

Through applications the viability of using fuzzy data analysis models is demonstrated. A case study shows how educational polices can be inferred and successfully used for evaluating the performance an educational systems.

# SUMÁRIO

	pág.
<b>1. INTRODUÇÃO</b>	<b>1</b>
1.1 Identificação do Problema	1
1.2 Objetivos do Trabalho	2
1.3 Justificativa e Importância do Trabalho	3
1.4 Estrutura do Trabalho	5
<b>2. DESEMPENHO DE SISTEMAS EDUCACIONAIS</b>	<b>7</b>
2.1 Introdução	7
2.2 Sistema Educacional: Considerações	8
2.3 Fatores Intervenientes no Sistema Educacional	11
2.4 Desempenho do Sistema Educacional	12
2.5 Principais Questões Relacionadas à Caracterização do Desempenho das Escolas	13
<b>3. MÉTODOS PARA ANÁLISE DE DESEMPENHO DE SISTEMAS EDUCACIONAIS</b>	<b>16</b>
3.1 Introdução	16
3.2 Modelos para Caracterização e Análise de Sistemas Educacionais	18
3.3 Modelos Estatísticos	19
3.4 Modelos Baseados em Análise Estatística Multivariada	24

	pág.	
3.4.1	Formulação dos Modelos Baseados em Análise de Regressão	25
3.4.2	Modelos com Coeficientes Constantes para todas as Escolas, sem Identificação das Escolas	26
3.4.3	Modelo de Efeitos Fixos	27
3.4.4	Modelos com Dados Agregados ao Nível de Escolas	28
3.4.5	Modelo de Efeitos Contextuais	29
3.4.6	Modelo de Efeitos Contextuais: Componentes de Variância	30
3.4.7	Ajustamento dos Modelos, Inferência, Medidas de Discrepância e Análise dos Resíduos	32
3.5	Modelos Hierarquizados	35
3.6	Modelos Baseados em Análise Taxionómica	39
3.6.1	Medidas de Similaridade	40
3.6.2	Técnicas para Realizar Agrupamentos	41
3.7	Conclusões	43
<b>4.</b>	<b>TEORIA DOS CONJUNTOS DIFUSOS E TEORIA DAS POSSIBILIDADES</b>	<b>44</b>
4.1	Introdução	44
4.2	Considerações sobre a Lógica Clássica	45
4.3	Definição de Conjuntos Difusos	45
4.4	Definições e Operações Básicas da Teoria dos Conjuntos Difusos	47
4.5	Funções de Pertinência	51
4.6	Conexão da Função de Pertinência com os Operadores	53
4.7	Estimação de Funções de Pertinência	56
4.8	Teoria da Possibilidade	57
4.9	Regressão Possibilística	59
4.9.1	Sistemas Lineares Possibilísticos	59
4.9.2	Regressões Lineares Possibilísticas	61

	pág.	
4.9.2.1	Problema de Minimização	64
4.9.2.2	Problema de Maximização	64
4.9.2.3	Problema de Conjunção	65
4.10	Teoria de Conjuntos Difusos e Sistemas Educacionais: uma Proposta	73
<b>5.</b>	<b>MODELOS PARA CARACTERIZAÇÃO DO DESEMPENHO DAS ESCOLAS</b>	<b>77</b>
5.1	Introdução	77
5.2	O Problema do Desempenho dos Sistemas Educacionais	77
5.3	Modelagem Estatística Multivariada: Modelos Logísticos	84
5.3.1	Modelos para Análise do Desempenho das Escolas Baseados na Aprovação dos Alunos e nas suas Características Individuais	87
5.3.2	Modelo Hierarquizado para Análise do Desempenho Através das Características dos Alunos e Características das Escolas	88
5.4	Modelagem Através de Conjuntos Difusos	91
5.4.1	Introdução de um Modelo de Regressão Linear Possibilístico para o Intervalo ( 0,1)	92
5.4.2	Modelo de Regressão Possibilístico Triangular Logit	95
5.4.3	Modelos para Análise de Desempenho das Escolas Utilizando - se Conjuntos Difusos	101
5.5	Modelagem do Desempenho das Escolas Através de Análise Taxionômica	102
5.5.1	Processo de Classificação das Escolas Utilizando-se os Resultados do Modelo Logístico Tradicional	103
5.5.2	Processo de Classificação das Escolas Utilizando-se os Resultados do Modelo Logit Triangular Difuso	104
5.6	Conclusões	106

	pág
<b>6. APLICAÇÕES DOS MODELOS</b>	<b>108</b>
6.1 Introdução	108
6.2 Caracterização das Escolas	108
6.3 Aplicação dos Modelos para Análise do Desempenho das Escolas, Baseados na Aprovação dos Alunos e nas suas Características Individuais (MODELOS 1 e 2)	113
6.4 Modelo Hierarquizado - Aplicação do MODELO 3	117
6.5 Modelos para Análise do Desempenho das Escolas Baseados na Aprovação dos Alunos e nas suas Características Individuais (MODELOS 4 e 5) Utilizando-se Conjuntos Difusos	120
6.6 Modelo Difuso Hierárquico - Aplicação do MODELO 6	124
6.7 Processos para Classificação das Escolas	127
6.7.1 Análise Taxionómica: Classificação das Escolas quanto as Probabilidades de Aprovação	128
6.7.2 Classificação de Escolas Utilizando Resultados dos Modelos Difusos	131
6.8 Conclusões	140
<b>7. ANÁLISE COMPARATIVA DOS MODELOS</b>	<b>142</b>
7.1 Introdução	142
7.2 Aplicações dos Modelos de Regressão Estatística Tradicional	143
7.3 Aplicações dos Modelos de Regressão Possibilística	144
7.4 Comparação entre os Modelos Baseados na Análise de Regressão Possibilística e os Modelos de Regressão Clássicos	147
7.5 Análise de Regressão quando a Variável Dependente é um Número Difuso	149
7.5.1 Ajuste do Modelo Possibilístico Triangular Logit para a Análise da "Propensão à Evasão"	152

	pág.
7.5.1.1 Problema de Minimização	152
7.5.1.2 Problema de Conjunção	159
7.6 Conclusões	162
<b>8. CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS</b>	<b>164</b>
8.1 Conclusões	164
8.2 Recomendações para Trabalhos Futuros	168
<b>BIBLIOGRAFIA</b>	<b>172</b>

## LISTA DE FIGURAS

	pág.
FIGURA 3.1 Sistema Dinâmico	20
FIGURA 4.1 Funções Conjuntivas Estritamente Monotônicas	54
FIGURA 4.2 Funções Conjuntivas Não Estritamente Monotônicas	54
FIGURA 4.3 Funções Disjuntivas Estritamente Monotônicas	55
FIGURA 4.4 Funções Disjuntivas Não Estritamente Monotônicas	55
FIGURA 4.5 Inclusão $[A_1]_h \supseteq [A_2]_h$	62
FIGURA 4.6 Relação entre o Nível h e o Índice de Performance J (c)	73
FIGURA 5.1 Inclusão do Valor Observado no Valor Estimado a um Nível h	92
FIGURA 5.2 Relações entre $Y_i$ e a Transformada de $g(y) = X = Y_i / (1 - Y_i)$	93
FIGURA 5.3 Função de Pertinência de um Número Difuso P entre ( 0,1)	95
FIGURA 5.4 Um Critério para Classificação das Escolas quanto a Possibilidade de Aprovação	105
FIGURA 6.1 Intervalos de Confiança para Fator SSE - MODELO 2	116
FIGURA 6.2 Intervalos de Confiança para Fator INDANO - MODELO 2	117
FIGURA 6.3 Gráfico de Resíduos para Ajuste do MODELO 3: Coeficientes do INTERCEPTO	119
FIGURA 6.4 Agrupamento Realizado - Coeficientes do MODELO 1	130
FIGURA 6.5 Diferenças entre os Grupos Determinados através da Análise de Agrupamentos	131
FIGURA 6.6 Um Critério para Classificação das Escolas quanto a Possibilidade de Aprovação	137
FIGURA 6.7 Agrupamento das Escolas - Possibilidade de Aprovação	139
FIGURA 7.1 Funções de Pertinência para Possibilidade de Aprovação	144
FIGURA 7.2 Valores Corte - $\alpha = 0.8$ - Escolas de Loulé	145
FIGURA 7.3 Entrada das Variáveis do Modelo 7.1	153

	pág.
FIGURA 7.4 Apresentação dos Resultados do Modelo 7.1 - Problema de Minimização - Valores $P_0$ e $P_1$	154
FIGURA 7.5 Valores Observados e Valores Estimados - Problema de Minimização - Modelo 7.1	155
FIGURA 7.6 Grau de Pertinência do Valor Estimado e do Valor Observado - Observação 15	156
FIGURA 7.7 Gráfico dos Intervalos - 17 Observações	157
FIGURA 7.8 Gráfico dos Intervalos - 14 Observações	158
FIGURA 7.9 Quatro Observações do Segundo Ajuste	159
FIGURA 7.10 Resultados Obtidos - Problema de Conjunção	160
FIGURA 7.11 Valores Observados e Valores Estimados - Modelo 7.1 - Problema de Conjunção	161
FIGURA 7.12 Grau de Pertinência do Valor Estimado e do Valor Observado - Observação 15 - Problema de Conjunção	161
FIGURA 7.13 Gráfico dos Intervalos - 17 Observações - Problema de Conjunção	162

## LISTA DE TABELAS

	pág.
TABELA 3.1 Associação entre Algumas Distribuições e Tipos de Dados	26
TABELA 5.1 Situação Sócio - Econômica e Agrupamento Realizado	81
TABELA 5.2 Variáveis Utilizadas para Desenvolvimento dos Modelos	83
TABELA 5.3 População Sub - dividida em N Grupos	85
TABELA 6.1 Escolas Analisadas e Total de Alunos	109
TABELA 6.2 Escolas Analisadas e Total de Alunos Considerados na Aplicação, Segundo o Ano de Escolaridade	110
TABELA 6.3 Alunos Aprovados e Reprovados, por Escola	111
TABELA 6.4 Alunos por Grupos Sócio-Econômicos, por Escolas	111
TABELA 6.5 Alunos Aprovados e Reprovados do 7º Ano, por Escola, Segundo a Idade	112
TABELA 6.6 Alunos Aprovados e Reprovados do 8º Ano, por Escola, Segundo a Idade	112
TABELA 6.7 Alunos Aprovados e Reprovados do 9º Ano, por Escola, Segundo a Idade	113
TABELA 6.8 Resultados do MODELO 1 - Escolas do Distrito de Faro	115
TABELA 6.9 Resultados do MODELO 2 - Escolas do Distrito de Faro	116
TABELA 6.10 Escolas Utilizadas para Estimação dos Modelos Hierárquicos - Valores para Variáveis Contextuais	118
TABELA 6.11 Resultados do MODELO 3 - Variável Dependente: Coeficientes do INTERCEPTO	118
TABELA 6.12 Resultados do MODELO 3 - Variável Dependente: Coeficientes da Variável SEXO	119
TABELA 6.13 Resultados do MODELO 3 - Variável Dependente: Coeficientes da Variável SSE	119
TABELA 6.14 Resultados do MODELO 3 - Variável Dependente: Coeficientes da Variável INDANO	120

	pág.	
TABELA 6.15	Número de Categorias Existentes por Escolas	121
TABELA 6.16	Resultados da Aplicação do MODELO 4	122
TABELA 6.17	Resultados da Aplicação do MODELO 5	122
TABELA 6.18	Resultados da Aplicação do MODELO 5 para Alunos do Sexo Masculino	123
TABELA 6.19	Resultados da Aplicação do MODELO 5 para Alunos do Sexo Feminino	123
TABELA 6.20	Resultados da Aplicação do MODELO 6	126
TABELA 6.21	Agrupamento das Escolas Analisadas - Coeficientes do MODELO 1	129
TABELA 6.22	Classificação das Escolas - MODELO 4 - Limite Inferior	132
TABELA 6.23	Classificação das Escolas - MODELO 4 - Centróide	133
TABELA 6.24	Classificação das Escolas - MODELO 4 - Limite Superior	134
TABELA 6.25	Classificações Normalizadas por Escolas - MODELOS 4 e 5	135
TABELA 6.26	Classificações Normalizadas por Escolas - MODELOS 5 para Sexo Masculino e 5 para Sexo Feminino	136
TABELA 6.27	Hierarquizações Obtidas para as Escolas Analisadas	138
TABELA 6.28	Agrupamento das Escolas - Possibilidade de Aprovação	138
TABELA 7.1	Probabilidades de Aprovação - MODELO 1	143
TABELA 7.2	Probabilidades de Aprovação - MODELO 3	144
TABELA 7.3	Possibilidades de Aprovação - Escolas de Loulé - Aluno Exemplo	145
TABELA 7.4	Efeitos Contextuais - MODELO 6	146
TABELA 7.5	Teste t para Coeficientes INTERCEPTO	147
TABELA 7.6	Teste t para Coeficientes do Fator SEXO	147
TABELA 7.7	Teste t para Coeficientes do Fator SSE	148
TABELA 7.8	Teste t para Coeficientes do Fator INDANO	148
TABELA 7.9	Características Individuais de uma Amostra de Alunos	149
TABELA 7.10	Valores de Julgamentos Emitidos por Especialistas para a Propensão à Evasão de Alunos	151

# 1. INTRODUÇÃO

## 1.1 - Identificação do Problema

Do ponto de vista filosófico, a análise de um determinado fenômeno ou sistema pode ser efetuada sob, basicamente, duas óticas: a da argumentação e a da demonstração.

A argumentação pressupõe a comunicação, diálogo e discussão.

A demonstração pressupõe a exposição de dados e premissas, seguidas de conclusões sobre os mesmos.

As abordagens desenvolvidas, pela Teoria Geral de Sistemas (TGS) (TRONCALE, 1988) são caracterizadas pelo uso da demonstração para análise de um sistema. A questão da evolução de um sistema, através da TGS, é inicialmente tratada a partir da identificação de variáveis imprescindíveis à análise.

No caso de um Sistema Educacional, embora muitas vezes, seja necessário recorrer à argumentação - por exemplo, no debate sobre o ideal político ideológico das legislações ligadas às diretrizes e bases educacionais - as questões relativas a avaliação de sua evolução são estudadas através do uso dos paradigmas da demonstração.

Os métodos tradicionais de análise de dados podem ser classificados como sendo (TAZZAN, 1988):

- a utilização de dados disponíveis como forma de se obterem outros de relevância;

- o estudo de relações de causa-efeito, pelo qual pretende-se estabelecer a origem e a consequência dos dados disponíveis;
- estabelecer através de exemplos, modelos e analogias de uma estrutura para a realidade analisada.

Neste trabalho discutem-se diferentes abordagens metodológicas utilizadas na avaliação de Sistemas Educacionais e procura-se demonstrar que algumas das técnicas de análise tradicionalmente utilizadas podem ser substituídas, com sucesso, por técnicas que se baseiam na Lógica Multivaluada. Fenômenos sociais e suas decorrências são melhor compreendidos e estudados dentro desta ótica do que utilizando-se a lógica clássica.

O tratamento de incerteza com base apenas em modelos sustentados pela Teoria Tradicional de Probabilidades tem sido objeto de ampla discussão e debate. O problema da modelagem da incerteza tem acarretado no surgimento de novas ferramentas de análise, baseadas em diferentes raciocínios e diferentes graus de incerteza. A Teoria dos Conjuntos Difusos fornece um instrumento adequado para modelar essas situações.

## **1.2 - Objetivos do Trabalho**

O objetivo geral desse trabalho tem é o de melhorar o processo de avaliação de desempenho de sistemas educacionais através da obtenção de novas informações que não são passíveis de serem extraídas utilizando-se os processos tradicionais de análise. Essa melhoria é obtida através da utilização da teoria de conjuntos difusos e da análise de regressão possibilística.

Apresenta como objetivos específicos:

- desenvolver modelos baseados em regressão possibilística, para analisar o desempenho de sistemas educacionais;
- propor um modelo teórico baseado na teoria dos conjuntos difusos, para tratar situações de escolha qualitativa, nas quais subsistem nebulosidades ou imprecisões quanto à importância dos fatores considerados como relevantes na explicação da escolha.

### **1.3 - Justificativa e Importância do Trabalho**

Necessidades de habitação, saúde, segurança e transporte da população podem ser enquadradas no conceito genérico de dívida social. Considerando-se o volume de investimentos necessários, qualquer que seja o grau de desenvolvimento do país, ao Estado dificilmente serão dadas possibilidades de atendimento global de todos esses itens, sem que ocorra exaustão nas suas capacidades de investimento.

Uma solução simplista e que, normalmente é apresentada, é o aumento sistemático da carga fiscal. Entretanto, tal estratégia é potencialmente perigosa, pois, penaliza em excesso aquela parcela da população que pode contribuir, produz efeitos nocivos na economia do país, aumenta a tendência a sonegação. Porém, o que está longe de ser exaurido, é a capacidade do Poder Público de aplicar os recursos de que dispõe de forma eficaz (FRIEDMAN, 1980, 1985).

Os avanços tecnológicos recentes têm levado a um novo perfil de cidadão: criativo, capaz de solucionar problemas, adaptar-se às mudanças verificadas e, sobretudo, capaz de gerar, selecionar e interpretar informações.

Neste contexto, a Educação torna-se indispensável ao sucesso econômico e social de qualquer país. Esse fato tem como consequência imediata as necessidades de revisão e de questionamento dos perfis dos diferentes sistemas de ensino.

O Estado, em maior ou menor grau, consoante o nível de desenvolvimento do país, tem um papel importante na correção das desigualdades sociais. O Estado não pode e não deve abdicar de seu papel. Uma nação não pode conviver com modelos de prestação de serviços públicos educacionais ineficientes e improdutivos e deve buscar soluções que conduzam a reformulações profundas. Grandes mudanças não são feitas mediante a adoção de soluções simplistas. Programas de treinamento de professores, aumentos de salários e novas metodologias de ensino poderão resolver, de forma paliativa, alguns dos problemas existentes. Todavia, compete aos governos a proposição de soluções globais e inovadoras visando a racionalização dos equipamentos educacionais, controle do quadro de pessoal, agilização das informações e maior produtividade dos recursos públicos.

O papel do Estado deve ser o de corrigir distorções do fluxo escolar e estabelecer um processo criterioso de avaliação do sistema educacional, o qual ofereça informações sobre a qualidade do serviço prestado pelas escolas. Se a melhoria da Educação requer, por um lado, ações globais por parte dos governos, por outro, pressupõe uma mobilização da população que se utiliza do sistema educacional. Os usuários do sistema educacional devem estar prontos para participar e fiscalizar o sistema. A eficácia do sistema está, portanto, associada à presença e ao controle da sociedade.

Um Sistema Educacional eficaz e eficiente traduz-se em uma perfeita harmonia entre os atores envolvidos. O processo de alocação de recursos dá-se mediante critérios transparentes e objetivos. Existe uma delegação de responsabilidades e de autonomia de execução de atividades. Porém, a avaliação dos resultados e a cobrança do desempenho torna-se

também necessária. Assim, é essencial a um sistema educativo a existência de formas de avaliação do seu desempenho. Esta é a proposta deste trabalho.

#### **1.4 - Estrutura do Trabalho**

Este trabalho é estruturado em oito capítulos.

No primeiro capítulo é feita a identificação do problema, delinea-se os objetivos pretendidos, a justificativa e importância do trabalho e a sua estrutura.

O segundo capítulo é dedicado a análise de Sistemas Educacionais, apresentando sua descrição, suas interações, agentes intervenientes e discussão de questões relativas ao seu desempenho.

A seguir, no terceiro capítulo são apresentados e discutidos os modelos tradicionalmente utilizados na análise de desempenho de Sistemas Educacionais.

No quarto capítulo é realizado um estudo sobre a Teoria dos Conjuntos Difusos.

No quinto capítulo são propostos modelos para avaliação de desempenho, descrevendo as estruturas lógicas e os requisitos necessários para aplicação dos mesmos.

As aplicações dos modelos são apresentadas no sexto capítulo.

No capítulo subsequente, uma análise comparativa dos modelos é realizada.

O oitavo capítulo apresenta as conclusões do trabalho e recomendações para futuras pesquisas relacionadas ao tema desenvolvido.

Finalmente é listada a bibliografia utilizada, bem como, a citada no trabalho.

## **2. DESEMPENHO DE SISTEMAS EDUCACIONAIS**

### **2.1 Introdução**

A integração sócio-econômica das diferentes regiões geográficas de um país, pode ser considerada como meta e preocupação constante de governos democráticos. Disparidades econômicas entre regiões parecem estabelecer uma forte correlação com disparidades regionais escolares.

Entre as causas de disparidades regionais escolares, sem dúvida, está a oferta de escolas. Quedas nessa oferta levam a um estrangulamento do sistema educacional do país.

Admita-se, porém, que exista uma oferta razoável e que, portanto, não seja esse o motivo causador das disparidades. Nesse caso, então, a atenção do poder público deverá se concentrar em questões relativas ao desempenho do sistema, avaliando-se questões associadas à qualidade do mesmo. Fatores como, por exemplo, níveis de repetência e de evasão extremamente elevados são vistos como causas imediatas de baixa escolarização de uma região e, conseqüentemente, impulsionadores de disparidades regionais.

Estudos de caracterização realizados em um sistema educacional apresentam, portanto, como aspectos relevantes o fato de permitir o conhecimento mais detalhado a respeito da região e o de prover condições de análise comparativa em termos macro regionais. A existência de disparidades regionais pode indicar que o desempenho desigual dos serviços públicos seria a principal causa das dificuldades para integração social e econômica entre uma dada região e o restante do país. Portanto, melhorias da qualidade do ensino oferecido passam a ser uma prioridade que, entre outros aspectos, contribuirá para a redução dessas disparidades.

Estabelecimentos escolares desempenham o papel de reprodução das relações ou das contradições verificadas no contexto social em que se insere o Sistema Educacional. É, portanto, a escola que desempenha um caráter estratégico dentro de uma sociedade organizada. A partir das escolas, uma ideologia dominante é repassada ou uma nova cultura é transmitida a outras gerações.

Filósofos da Educação, quaisquer que sejam suas matizes ideológicas entendem a Educação como uma doutrina que reflete concepções de uma sociedade e que, dentro de uma realidade específica, essa doutrina é transmitida por intermédio de vetores, dentre os quais a escola é um dos principais (FREITAG, 1986).

## **2.2 Sistema Educacional: Considerações**

A maioria dos autores concordam com dois pontos em relação a conceituação de Educação e sua situação no contexto social:

- a educação apresenta uma doutrina pedagógica, que se baseia em uma filosofia de vida (homem/sociedade);
- o processo educacional, dentro de uma realidade social, é o porta-voz de uma determinada doutrina pedagógica, através de instituições específicas (escola, família, comunidade, igreja, etc.).

DURKHEIM (1972) foi um dos primeiros autores a sistematizar essa posição. Para ele, o processo educacional é conduzido basicamente pela família, mas também por instituições do Estado (escolas, universidades, etc.). Onde as gerações adultas promovem nas crianças, por meio dessas instituições, "certo número de estados físicos, intelectuais e morais, reclamados pela sociedade política no seu conjunto e pelo meio especial ao qual a criança se destina".

Nessa teoria educacional está implícita a filosofia de vida de que a experiência das gerações adultas é indispensável para a sobrevivência das gerações mais novas. A educação é um fato social. Nesse contexto, os conteúdos da educação são independentes das vontades individuais, são as normas e valores de uma sociedade num dado momento histórico, que adquirem certa generalidade e com isso uma natureza própria, tornando-se assim "algo exterior" aos indivíduos. Entretanto, é no processo educacional que essas "externidades" impostas aos indivíduos são por eles "internalizadas", de tal forma que são reproduzidas e perpetuadas na sociedade.

As linhas mestras das idéias de Durkheim também são partilhadas por PARSONS in FREITAG (1986), onde o mesmo acredita que a educação seja o mecanismo básico para constituição de sistemas sociais e de manutenção e perpetuação desses em forma de sociedades. Mas, para Parsons, existe no processo educativo uma troca de "equivalentes" em que tanto o indivíduo quanto a sociedade se beneficiam.

Tanto Durkheim como Parsons evidenciam que valores genéricos como continuidade, conservação, ordem, harmonia, equilíbrio são os princípios básicos que regem o funcionamento do sistema social. Por esse motivo, é que esses autores têm sido criticados, por seus pressupostos conservadores, pois os mesmos não vêem na educação um fator de desenvolvimento e melhorias nas estruturas sociais vigentes, mas sim o conhecimento necessário transmitido de geração para geração, perpetuando a estrutura e o funcionamento de uma dada sociedade.

DEWEY (1971) e MANNHEIM (1972) divergem dos autores acima citados. Para eles a educação não é um mecanismo de correção e ajustamento dos indivíduos a dadas estruturas sociais, mas sim, um fator de dinamização das estruturas. O indivíduo, dentro do processo educacional, é capacitado para atuar no contexto social em que vive, não somente, reproduzindo as experiências das gerações adultas mas, avaliando essas experiências, de tal forma que o mesmo

se torne habilitado a reorganizar seu comportamento e contribuir para reestruturação da sociedade. Desse modo, tanto o indivíduo como a sociedade são vistos como um processo dinâmico de constantes mudanças.

Segundo DEWEY (1971) o ato educacional consiste em fornecer ao indivíduo os subsídios necessários para a reorganização de experiências vividas, de maneira ordenada e sistematizada. Para que isso ocorra, o meio em que o processo educacional é conduzido, tem que ser organizado e reestruturado. Esse autor acredita que esse meio é a escola, e que essa deva assumir as características de uma pequena comunidade democrática. Portanto, a educação não é simplesmente um mecanismo de perpetuação de estruturas sociais anteriores, mas um mecanismo de implantação de estruturas sociais democráticas. Pressupõe indivíduos com igualdades de chances, em uma sociedade livre e igualitária, competindo por diferentes privilégios. Dessa forma, o modelo social é o de igualdade de chances, não o de igualdade entre os homens, sendo essa igualdade reconhecida e aceita pelos indivíduos que se admitem como diferentes em relação a natureza (inteligência, habilidade, etc.). Em vista disso, o modelo social não é questionado, criticado ou modificado, e sim, nele a ordem é regulamentada pela competição e os conflitos são democraticamente resolvidos. As mudanças admissíveis neste sistema se resumem no aperfeiçoamento das estruturas democráticas, e uma vez implantado o sistema, todos os mecanismos funcionarão para a sua conservação. Com isso, percebe-se que existe uma convergência com a concepção de Durkheim e Parsons quanto aos resultados dos processos educacionais.

BECKER e EDDING in FREITAG (1986) e SCHULTZ (1971), podem ser considerados os pais de disciplinas criadas paralelamente, tais como, planejamento educacional, economia da educação, etc. Essas disciplinas buscam preencher lacunas esquecidas até então pela sociologia da educação referentes aos aspectos econômicos da mesma, sendo que essas novas disciplinas tem orientado os tomadores de decisões da área educacional.

Para esses autores, a fundamentação de suas reflexões, é a

existência de uma alta correlação entre o crescimento econômico e o nível educacional dos membros que compõem uma sociedade.

### **2.3 Fatores Intervenientes no Sistema Educacional**

O Sistema Educacional não funciona isoladamente, muito pelo contrário, atua de diversas formas nos vários setores da sociedade (econômico, político e social) determinando seu perfil e em contrapartida sofrendo fortes influências e condicionamentos desses.

Quando se está interessado em proceder análises a um determinado sistema organizacional torna-se necessária uma correta caracterização qualitativa e quantitativa desse, tanto em termos de sua dinâmica interna como de seu enquadramento exterior.

Deste modo, afim de se efetuar avaliações ao Sistema Educacional, torna-se imprescindível identificar fatores que intervém no seu funcionamento e desenvolvimento.

Entre os fatores intervenientes no Sistema Educacional pode-se citar:

- aumento na procura do Sistema Educacional, pela população, na busca de melhores condições de vida e de promoção social, em função do desenvolvimento econômico, tecnológico e científico;
- procura por determinada qualificação profissional, devido as condições impostas pelo mercado em relação a mão-de-obra (lei da oferta e da procura). O Sistema Educacional necessita constantes reajustamentos de modo a satisfazer as realidades físicas e humanas, regionais e locais;
- necessidade de mobilização de recursos mais exigentes, em

termos qualitativos e quantitativos, para assegurar o desenvolvimento do setor educativo, bem como, de outros setores (habitação, saúde e alimentação) que interferem na educação;

- ausência de informações sistemáticas, qualitativas e quantitativas, de determinados indicadores educacionais e de carência de especialistas na área de planejamento educacional, dificultando o processo de planejamento;
- diversidade de fontes oficiais dos indicadores estatísticos apresentando conteúdos diversos, embora relacionados aos mesmos indicadores.

## **2.4 Desempenho do Sistema Educacional**

O objetivo principal de se efetuar a avaliação do Sistema Educacional é o fornecimento de informações referentes aos mecanismos de funcionamento da educação. Essa avaliação pode, por exemplo, descrever resultados de análises realizadas ou sintetizar dados através de indicadores que permitam avaliar o funcionamento do sistema escolar. Em resumo, permite o levantamento e acesso de todo um conjunto de informações críticas que forneçam subsídios aos tomadores de decisões, para uma reforma contínua do Sistema Educacional e para o planejamento escolar. Quanto mais aprofundadas forem as informações, mais eficazes serão as decisões. Em contrapartida, a avaliação será influenciada pela complexidade crescente das funções da educação.

Dependendo do nível no qual se deseja fazer a avaliação, diferentes interesses podem surgir, pois essa funciona como um sistema regulador e controlador, quer a nível de rendimento individual, no processo de aprendizagem, quer a nível de rendimento educacional.

A avaliação do Sistema Educacional serve como subsídio ao poder público para mostrar a adequabilidade do sistema consoante aos objetivos e perspectivas da educação na sociedade. Desse modo, torna-se necessário capacitar o sistema educacional com um conjunto de estratégias e metodologias que permitam realizar o seu próprio controle e ao mesmo tempo gerar os fundamentos teóricos que justifiquem esse controle.

Para avaliar a educação torna-se necessário um modelo teórico que explique o funcionamento do sistema escolar. Os dados para a avaliação devem ser interpretados em função desse modelo, o que leva ao desenvolvimento do modelo por si próprio, e que crescerá em complexidade. Dessa forma, a informação obtida sobre o funcionamento do sistema será cada vez mais abrangente. Deve-se considerar ainda, na avaliação da educação, variáveis exógenas determinantes, geralmente não quantificáveis.

A transferência de resultados para o planejamento educacional deve ocorrer de forma equilibrada entre estudos quantitativos e qualitativos, de acordo com as perspectivas da nova sociologia da educação e da sociologia do conhecimento (COSTA, 1981).

## **2.5 Principais Questões Relacionadas ao Desempenho das Escolas**

O Sistema Educacional é concebido, apoiado, financiado e desenvolvido pela sociedade fornecendo, em retorno, progresso humano, ético e cultural bem como capital de conhecimentos útil à vida social e econômica e aptidões para o desempenho profissional (TAVARES, 1991). Além de fornecer elementos necessários ao desenvolvimento, torna-se extremamente importante como setor econômico na medida que atua como vetor propulsor desse mesmo desenvolvimento e pelos investimentos em si realizados.

Em termos das relações "Educação-Sociedade" como estudos relevantes e de interesse continuado podem ser enumerados aqueles relativos à:

- participação e Acessibilidade;
- participação e Igualização;
- participação e Acesso.

De acordo com TAVARES (1991) o interesse no estudo desses aspectos é caracterizado por:

- **Participação e Acessibilidade:** "Estudo das relações que se estabelecem entre o sistema educacional e seus beneficiários: os alunos". Nesse sentido, são questões relevantes avaliar os benefícios advindos do sistema educacional nos diferentes estratos etários e sociais e em qual medida o sucesso e integração do corpo docente é influenciado, quer pelas características sócio econômicas desse corpo docente, quer por fatores próprios de cada estabelecimento de ensino;
- **Participação e Igualização:** É o sistema educacional um fator de igualdade social? Permite o sistema educacional que ocorram compensações entre grupos e classes sociais de diferentes origens? Nesses casos, a problemática a ser tratada refere-se a "identificar e estimar funções mais significativas, f, para explicar a variabilidade de S na população de alunos em análise, sendo S uma variável definidora do seu sucesso escolar. Essa função poderá incluir variáveis caracterizadoras do tipo de escola e qualidade de seu ensino, E, do meio familiar e social, F, do sucesso anterior, S, das características dos professores, P, etc.";
- **Participação e Acesso:** Caracterizam-se pela tentativa de estabelecer em que condições se dá a participação dos diferentes grupos sociais no sistema educacional. Além da

participação, o interesse é também voltado para explicar a intensidade dessa participação. Normalmente, tem sido utilizados modelos baseados em taxas de escolarização e, para determinação de fatores causais, modelos que infiram influências decorrentes de: condições econômicas de acesso; nível sócio econômico dos responsáveis pelo aluno; sucessos anteriores; etc.

Diversos modelos de análise tem sido propostos e utilizados para caracterizar e explicar as relações existentes entre o sistema educacional e a sociedade. Nos próximos capítulos serão apresentados alguns dos métodos de natureza estatística para caracterização dessas relações, bem como serão discutidas a viabilidade da Teoria dos Conjuntos Difusos e da Teoria das Possibilidades para essa mesma finalidade.

## **3. MODELOS PARA ANÁLISE DE DESEMPENHO DE SISTEMAS EDUCACIONAIS**

### **3.1 - Introdução**

Modelar um sistema requer um processo de autocrítica constante. É necessário verificar, a cada passo, a verossimilhança do modelo em contraposição a realidade a ser apresentada (NOVAES, 1982).

Um sistema é composto por um conjunto de elementos e suas relações. Os elementos são representados por seus atributos relevantes, isto é, características que apresentam relações de causa e efeito com outros elementos. O estado de um sistema é caracterizado por um conjunto de valores dos elementos e das relações entre eles.

Na medida que aumenta a complexidade do sistema, aumenta o número de seus estados. Assim, a análise deve ser executada em um conjunto possivelmente finito de estados representativos.

A análise de desempenho de sistemas pode ser realizada, através de modelos, qualitativamente ou quantitativamente.

Nas análises qualitativas procura-se verificar o desempenho do sistema quanto aos níveis de satisfação esperados. Nas análises quantitativas, além de níveis de satisfação, procura-se quantificar, em termos numéricos, valores relevantes para o decisor.

Os modelos destinados a análise quantitativa de sistema subdividem-se em duas classes: Modelos Analíticos e de Simulação.

A classe de modelos analíticos pressupõe representar um sistema através de equações. Essas equações permitem descrever associações existentes entre dados, relações, critérios e decisões.

O conhecimento associado aos dados pode ser classificado como (KAUFMANN, 1986):

- certo;
- probabilístico;
- incerto.

É o tipo de conhecimento associado que determina a forma do modelo. Modelos Determinísticos são baseados em uma pressuposição de certeza. Modelos Probabilísticos baseiam-se no comportamento aleatório dos dados e procuram descrever alguma regularidade existente. Os modelos determinísticos e probabilísticos baseiam-se na lógica tradicional e na teoria clássica de conjuntos.

Na construção de modelos analíticos baseados ou no conhecimento certo ou no conhecimento probabilístico dos dados, estão presentes princípios básicos da lógica e da teoria de conjuntos, como o princípio da identidade, do meio excluído e da contradição.

Modelos analíticos baseados na incerteza procuram ultrapassar as limitações derivadas do uso da lógica e da teoria de conjuntos. Procuram, portanto, evitar ambivalências decorrentes de uma lógica bivalente. Paradoxos tradicionais, como o proposto por RUSSEL, são resolvidos pela não restrição a valores binários na construção de tabelas-verdade. A Álgebra de Boole é substituída pela teoria proposta por ZADEH(1965) - Teoria dos Conjuntos Difusos - na qual a transição dos elementos entre as categorias é obtida de uma forma gradativa (e não de forma abrupta 0 ou 1).

Os modelos analíticos são construídos, muitas vezes, com um elevado grau de complexidade. A complexidade pode ser avaliada através do número de associações estabelecidas para o modelo. Nos casos de grande grau de complexidade, os usuários finais do modelo, na maioria das vezes, não conseguem aplicá-lo. Conseqüentemente, vantagens tradicionais na utilização de modelos, como abrangência plena, não se verificam.

Nos modelos de simulação, as relações são descritas, geralmente, em programas computacionais e a análise é executada sobre registros históricos ou utilizando tabelas de dados normalizados. Uma restrição que se põe à utilização desses modelos é a de que os mesmos requerem uma ampla aplicação de técnicas estatísticas (redução de variância, ajuste de modelos teóricos de comportamento, etc.) para sua efetiva utilização (FISHMAN, 1979).

### **3.2 - Modelos para Caracterização e Análise de Sistemas Educacionais**

A análise de Sistemas Educacionais tem despertado interesse continuado de pesquisadores ligados as questões educacionais, quer a nível filosófico quer a nível de elaboração de novas metodologias.

Tradicionalmente, os responsáveis pela gestão de sistemas educacionais fazem uso de indicadores para a avaliação de desempenho desses sistemas. KEEVES (1986), ressalta que, apesar da importância da educação dentro do contexto social e importância de se possuir indicadores que reflitam a eficiência do processo educacional, poucos são os indicadores que têm sido desenvolvidos nesta área. Os indicadores existentes não refletem os resultados obtidos pelo processo educacional e, pouco contribuem para esclarecer sobre progressos ocorridos. Os indicadores que têm sido desenvolvidos objetivam obter bases para comparações dentro do sistema, entre períodos de tempo ou entre sistemas. A preocupação de Keeves é a de minimizar

as falhas apontadas por KANDEL (1986) quando da utilização de índices como indicadores.

Um relatório do grupo de trabalho da ROYAL STATISTICAL SOCIETY (1985) aponta dificuldades na elaboração de alguns indicadores. Cita como exemplo, taxas de retenção que não necessariamente representam um aumento no nível de educação alcançado.

Além de simples indicadores, a avaliação do impacto das variáveis controláveis e não controláveis é proposta por MAJLUF (1988) como um argumento para aprofundar a análise das causas estruturais que levam a obtenção de um valor para um indicador educacional. Essa análise deve levar em consideração a questão de fatores que contribuem para a eficiência do sistema. SWAMINATHAN (1989) comenta sobre as dificuldades de caracterização filosófica dos conceitos de eficácia, e alerta sobre a necessidade de torná-los mais desenvolvidos e operacionalizáveis. Para que se atinjam esses objetivos e os resultados produzam impacto, salienta a necessidade de tornar a utilização das técnicas de análise mais amigáveis aos utilizadores finais.

Existe na literatura uma série de técnicas e de modelos utilizados para a análise de Sistemas Educacionais. Esses modelos englobam a utilização de Técnicas Estatísticas, de Programação Linear, de Projeção e de Simulação entre outras. Neste capítulo a atenção será dada aos modelos e métodos de natureza estatística para compreensão, caracterização e análise de sistemas educacionais.

### **3.3 - Modelos Estatísticos**

Modelos estatísticos são aqueles em que existe uma forte componente probabilística.

Na inferência estatística clássica, os modelos são utilizados para descreverem situações experimentais ou processos geradores de dados. Dentro da inferência clássica, o modelo compreende um espaço de resultados  $\Omega$  cujos elementos  $w$  são resultantes de processos observados sobre uma variável ou sobre um vetor de variáveis (MURTEIRA, 1988).

Considere-se o seguinte sistema dinâmico representado na FIGURA 3.1.

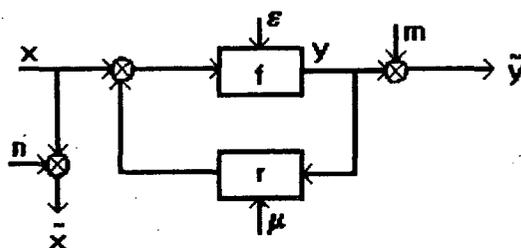


FIGURA 3.1 - Sistema Dinâmico

Fonte: BUNKE e BUNKE (1986)

onde:  $x$  e  $y$  = sinais não-observáveis de entrada e de saída

$\tilde{x}$  e  $\tilde{y}$  = sinais observáveis afetados por ruídos  $n$ ,  $m$

$\mu$  = sinal de controle

$\varepsilon$  = variáveis de erro

$f, r$  = funções de transferência

Sejam  $y_t$  o valor de  $y$  no instante de tempo  $t$  e  $y_0$  as condições iniciais do sistema;  $x^t = \{x_s \mid 0 < s < t\}$ ,  $\mu^t$  e  $\varepsilon^t$  realizações de  $x$ ,  $\mu$ ,  $\varepsilon$  durante o intervalo de tempo  $(0, t)$ . Então  $y_t$  pode ser obtido através de:

$$y_t = \phi(y_0, t, x^t, \varepsilon^t, \mu^t) \quad (3.1)$$

Entretanto, por definição,  $y_t$  e  $x_t$  são observáveis, apenas, através das variáveis  $\tilde{y}_t$  e  $\tilde{x}_t$ , onde:

$$\tilde{x}_t = x_t + n_t \quad (3.2)$$

$$\tilde{y}_t = y_t + m_t \quad (3.3)$$

Se as funções de transferência  $f$  e  $r$  são desconhecidas, então os problemas decorrentes são:

- ajustar funções de transmissão adequadas, a partir do uso das observações de  $\tilde{y}$  e  $\tilde{x}$ ;
- identificar os parâmetros adequados para as funções estabelecidas no item acima.

O sistema descrito pode ser interpretado como sendo um modelo estatístico e sua resolução será obtida através da aplicação de princípios de inferência estatística.

Uma formulação matemática mais consistente de um modelo estatístico pode ser obtida a partir das seguintes definições (BUNKE e BUNKE, 1986):

### Definição 1:

Se a variável aleatória "z" assume valores em um espaço de resultados "Z", então um modelo distribucional para "z" é o conjunto de distribuições de probabilidade sobre uma sigma-álgebra  $U_z$ , definida em Z. Cada distribuição de probabilidade é denominada uma estrutura distribucional para "z". Seja  $P_0$  a verdadeira estrutura distribucional; então o modelo  $\Gamma$  é considerado adequado se e somente se  $P_0 \in \Gamma$ .

Se  $\Gamma = \{ P_\theta \mid \theta \in \Theta \}$  e  $P_0 = P_{\theta_0}$ , então  $\theta_0$  é chamado de verdadeiro parâmetro.

**Definição 2:**

Se uma variável aleatória "z" toma valores em "Z" então uma estrutura funcional para z é um par  $[f, P]$ , onde  $f: ( \xi, U_\xi ) \rightarrow ( Z, U_Z )$  é uma função mensurável e P é uma distribuição de probabilidade sobre  $U_\xi$ . Um conjunto de estruturas funcionais é chamado um modelo funcional para "z". Seja  $[f_0, P_0]$  a estrutura funcional verdadeira (isto é  $z = f_0( \xi )$  onde  $\xi \sim P_0$ ); então, o modelo  $\Gamma$  é considerado adequado se e somente se  $[f_0, P_0] \in \Gamma$ .  $v\Gamma = \{ [f_v, P_v] \mid v \in \Theta \} = \Gamma_\Theta$  e  $[f_0, P_0] = [f_{v\theta}, P_{v\theta}]$ , então "v $\theta$ " é chamado um verdadeiro parâmetro.

Modelos  $\Gamma_\Theta$  podem ser escritos na forma:

$$z = f_v( \varepsilon ), \quad \varepsilon \sim P_v, \quad v \in \Theta \quad (3.4)$$

Nos modelos estatísticos, portanto, existem estruturas funcionais e estruturas distribucionais. O relacionamento entre os tipos de estruturas existentes se dá pelo feito de que uma estrutura funcional  $([f, P])$  gera uma estrutura distribucional  $P^Z$  em  $U_Z$ . A estrutura distribucional  $P^Z$  é definida através de:  $P^Z(A) = P(f^{-1}A) = P(\{ \varepsilon \in \xi \mid f(\varepsilon) \in A \})$  para cada  $A \in U_Z$ .

Obviamente, existem muitas estruturas funcionais que geram uma determinada estrutura distribucional. Como exemplo, considere-se que y seja  $N( \mu ; \sigma^2 )$ . Essa é uma estrutura distribucional que poderia ter sido gerada por uma das seguintes estruturas funcionais:

$$y = \mu + \sigma \varepsilon, \quad \varepsilon \sim N(0,1) \quad (3.5)$$

ou

$$y = \mu + \sigma \phi^{-1}(\eta), \quad \eta \sim R(0,1) \quad (3.5a)$$

onde:  $\phi^{-1}$  é uma função inversa da função densidade de probabilidade da normal padrão e  $R(0,1)$  é a função densidade de probabilidade para uma variável com distribuição uniforme no intervalo (0,1).

Uma estrutura funcional, portanto, fornece uma descrição mais definitiva para uma variável aleatória do que uma estrutura distribucional. Enquanto a estrutura funcional especifica quais os mecanismos geradores dos dados observados, a estrutura distribucional refere-se a distribuição probabilística dos dados.

Na análise de Sistemas Educacionais, o maior interesse é o de encontrar a melhor ou a mais adequada estrutura funcional para descrever o sistema em termos de eficiência, de desempenho, etc.

Diversos autores afirmam que um problema importante em pesquisa educacional quantitativa é a medida de eficiência da escola.

O problema, em termos estatísticos, é a descrição das relações entre uma ou mais variáveis "resultantes" (resultados de exames, "atendimento", problemas disciplinares, etc.) processos da escola e de ensino (recursos financeiros, tamanho de classe, razão aluno/professor, atitudes de professores, etc.) e características dos alunos considerados (status sócio-econômico e outras informações demográficas, habilidade individual ao entrar na escola, sexo, grupo étnico, etc.). Um importante aspecto de descrição é a representação da variabilidade das relações entre escolas.

Diferentes procedimentos de modelagem têm sido utilizados, cada um deles levando a diferentes conclusões sobre a importância de variáveis explicativas individuais e a eficiência estimada de uma escola. Na elaboração do modelo, devem ser tomadas decisões sobre o nível apropriado de análise. Se o objeto da análise é a medida da importância das variáveis a nível de escola, existe disponibilidade de resultados a nível de alunos, eles devem ser agregados ao nível de escola para análise, ou sua análise deve ser feita separadamente? Essa é uma, entre muitas questões a serem respondidas, ao se adotar um modelo estatístico para análise da eficiência do Sistema Educacional.

### 3.4 Modelos Baseados em Análise Estatística Multivariada

Quando se utiliza análise de regressão, o que se procura é o estabelecimento de uma estrutura funcional que descreva a relação causal entre  $x$  e  $y$ .

O par  $[f, w]$  é considerado uma estrutura funcional para a relação causal entre  $x$  e  $y$  se:  $f: x, \varepsilon \rightarrow y$  e  $w: x, U_\varepsilon \rightarrow [0, 1]$  são funções e se  $w(x, \bullet)$  é uma distribuição de probabilidade em  $U_\varepsilon$  para cada  $x$  fixo. A mesma estrutura funcional pode ser explicada da seguinte forma: para cada valor  $x$  da variável de entrada, a saída  $y$  é dada por:  $y=f(x, \varepsilon), \varepsilon \sim w(x, \bullet)$ .

As estruturas funcionais encontradas na literatura são da forma:

$$f(x, \varepsilon) = \varphi_\beta(x) + \varepsilon, \quad w(x, \bullet) = N(0, \sigma^2) \quad (3.6)$$

onde: a função  $\varphi_\beta$  é chamada função de regressão (ou função resposta) e representa o valor da variável de saída  $y$  como uma função do valor da variável de entrada  $x$ .

São tomadas observações  $(x_i, y_i)$ ,  $i = 1, \dots, n$  e usadas como base para a realização de inferências sobre a relação causal entre  $x$  e  $y$ . Os valores observados de:  $\xi = (x_1, x_2, \dots, x_n)$  e  $\eta = (y_1, y_2, \dots, y_n)$  e a relação causal entre  $\xi$  e  $\eta$  devem então ser relacionadas à relação causal entre  $x$  e  $y$ . Essa relação entre as duas causais pode ser descrita por modelos funcionais:  $\{(f_v, w_v) \mid v \in \Theta\}$  e  $\{(f_v^n, w_v^n) \mid v \in \Theta\}$ , onde as duas relações causais usam o mesmo espaço paramétrico e supõe-se existir um parâmetro verdadeiro  $v_0$  em  $\Theta$  de modo que  $\left[ f_{v_0}, w_{v_0} \right]$  e  $\left[ f_{v_0}^n, w_{v_0}^n \right]$  são as duas estruturas funcionais verdadeiras. Tem-se, então, o seguinte:

$$f_V(x, \varepsilon) = \varphi_\beta(x) + \varepsilon, \quad w(x, \bullet) = N(0, \sigma^2) \quad (3.7)$$

$$f_V^\eta(x, \varepsilon) = \varphi_\beta^\xi + \varepsilon^\xi, \quad w(\xi, \bullet) = N(0, \sigma^2 I_\eta) \quad (3.8)$$

onde:

$$\varphi_\beta^\xi = (\varphi_\beta(x_1) + \dots + \varphi_\beta(x_n)), \quad \varepsilon^\xi = (\varepsilon_1, \dots, \varepsilon_n) \quad (3.9)$$

### 3.4.1 Formulação dos Modelos baseados em Análise de Regressão

Para a formulação dos modelos deve-se considerar os seguintes aspectos:

- análise da distribuição de probabilidade adequada para a variável de resposta;
- exame cuidadoso dos dados, principalmente com respeito a: assimetria, natureza (contínuos ou oriundos de processo de contagens - discretos), intervalo de variação etc.;
- cuidado no processo de escolha do conjunto de covariáveis. Os termos componentes da estrutura de covariáveis podem ser: contínuos, qualitativos ou mistos. Para uma covariável contínua, em geral, corresponde apenas um único parâmetro  $\beta$ . Sendo qualitativa, obtém-se um conjunto de parâmetros  $\alpha_i$ , onde  $i$  é o índice que representa os diversos níveis do fator.

A TABELA 3.1 apresenta diversas distribuições associadas com algumas características existentes nos dados.

TABELA 3.1 - Associação entre Algumas Distribuições e Tipos de Dados

Distribuição	Associação com Dados
Gama	Dados contínuos assimétricos, apresentando coeficiente de variação constante.
Normal Inversa	Dados contínuos assimétricos.
Normal	Dados simétricos e variação no conjunto dos reais. Campo de variação entre $(0, \infty)$ e dados transformados (por exemplo, logaritmos).
Poisson	Dados na forma de contagem. Dados contínuos com variância aproximadamente igual à média. Dados com super-dispersão: $V = \gamma\mu, \gamma > 1$ . Dados com sub-dispersão: $V = \gamma\mu, \gamma < 1$ , com $\mu$ representando o parâmetro média da população.
Binomial	Dados na forma de proporções. Dados contínuos ou discretos apresentando sub-dispersão.

AITKIN e LONGFORD (1986) estudando a questão da eficiência das escolas e, conseqüentemente, do Sistema Educacional, propõem cinco modelos estatísticos baseados em análise de regressão, procurando identificar qual a melhor estrutura para o problema.

### 3.4.2 Modelos com Coeficientes Constantes para todas as Escolas, sem Identificação das Escolas

O modelo é da forma:

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_n x_{nij} + \varepsilon_{ij}, \quad (3.10)$$

$$j = 1, \dots, n_i; \quad i = 1, \dots, k$$

onde:  $x = (x_1, x_2, \dots, x_n)$  vetor de variáveis com características dos alunos

$\varepsilon_{ij} \sim N(0, \sigma^2)$ ,  $i = i$ -ésima escola,  $j = j$ -ésimo aluno

Neste modelo os alunos são considerados a partir de uma única população. Os dados são tratados como uma única amostra de  $\sum n_i$  observações sobre  $y$  (variável de resposta) e vetor  $x$ . O ajuste do modelo é realizado através do método dos Mínimos Quadrados Ordinários.

Análises sobre o desempenho das escolas são efetuadas sobre os resíduos médios de cada escola. Uma vez que não há identificação das escolas, o processo adotado para estabelecer resíduos médios das escolas é:

$$e_{(1)j} = \bar{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_{1i} - \dots - \hat{\beta}_n \bar{x}_{ni} =$$

$$= \bar{y}_i - (\bar{y} - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2 - \dots - \beta_n \bar{x}_n) - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_n x_{ni} \quad (3.11)$$

e o efeito escola é analisado sobre os coeficientes  $\beta_j$ ,  $i = 1, \dots, n$  ou seja os coeficientes das características dos alunos.

Os coeficientes  $\beta_l$ ,  $l = 1, \dots, n$  são comuns a todas escolas e são calculados, resolvendo-se o sistema de equações normais, considerando-se o valor médio de todas as escolas para as variáveis características dos alunos.

### 3.4.3 Modelo de Efeitos Fixos

O modelo é da forma:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_n x_{nij} + \varepsilon_{ij} \quad (3.12)$$

$$i = 1, \dots, k; \quad j = 1, \dots, n$$

onde:  $y_{ij}$  = variável resposta para o  $j$ -ésimo aluno da  $i$ -ésima escola

$x_{lij}$  = variável correspondente a  $l$ -ésima característica do  $j$ -ésimo aluno da  $i$ -ésima escola, com  $l = 1, \dots, n$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Neste modelo obtém-se um conjunto de regressões paralelas de  $y$  sobre o vetor de características  $x = (x_1, \dots, x_n)$  e o efeito das escolas é definido pelo valor do intercepto  $\beta_0$ . Para cálculo dos coeficientes  $\beta_0$  são utilizados os valores médios de cada escola:

$$\hat{\beta}_0 = \bar{y}_i - \beta_1 \bar{x}_{1i} - \dots - \beta_n \bar{x}_{ni} \quad (3.13)$$

Os valores  $\beta_l$ ,  $l = 1, \dots, n$  também são comuns a todas as escolas. Contudo, o sistema de equações normais é resolvido considerando o valor médio de cada escola para as variáveis características dos alunos.

#### 3.4.4 Modelos com Dados Agregados ao Nível de Escolas

Neste modelo os dados são agregados dentro das escolas. Assim, tanto os regressores quanto a variável utilizada como resposta são consideradas em termos de médias observadas na  $i$ -ésima escola. O modelo tem, então, a seguinte forma:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{1i} + \dots + \beta_n \bar{x}_{ni} + \eta_i; \quad i = 1, \dots, k \quad (3.14)$$

onde:  $\bar{y}_i$  = média da variável resposta para a  $i$ -ésima escola

$\bar{x}_{ni}$  = média da n-ésima característica observada na i-ésima escola

$\eta_{ij}$  = ruído aleatório

O ruído aleatório pode assumir diversas formas. Por exemplo, se:  $\eta_{ij} \sim N(0, \sigma^2)$  então o efeito escola é considerado invariante em relação a população de alunos presentes em cada escola. Contudo, esta suposição de homocedasticidade pode não ser plausível, dado que são verificadas diferenças claras entre, por exemplo, escolas urbanas e escolas localizadas em agregados populacionais menores. Pelo mesmos motivos, as suposições de ruídos não correlacionados são violadas. A solução mais imediata é a de considerar  $\eta_{ij} \sim N(0, \eta_i \sigma^2)$  e o modelo ser estimado utilizando Mínimos Quadrados Ponderados, com  $\eta_i$  sendo o fator de ponderação.

O modelo difere do descrito no item 3.4.3 pelo fato que o sistema de equações normais é solucionado considerando-se os desvios existentes entre a média da i-ésima escola e a média global das escolas.

### 3.4.5 Modelo de Efeitos Contextuais

No modelo procura-se captar o efeito contextual da i-ésima escola. Esse efeito é entendido como sendo o efeito produzido nos alunos pelo fato desses pertencerem a uma dada escola.

A forma do modelo é:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \delta_1 \bar{x}_{1i} + \dots + \beta_n x_{ij} + \delta_n \bar{x}_{ni} + \varepsilon_{ij} \quad (3.15)$$

A equação (3.15) difere da (3.14) pela inclusão dos termos relativos as médias das escolas para a n-ésima característica considerada.

AITKIN e LONGFORD (1986) argumentam que em modelos mais complexos (em termos do número de variáveis explicativas) os erros padrões das variáveis agregadas a nível de escola tendem a se tornar bastante elevados. Assim, a inclusão desses tipos de variáveis explicativas para caracterizar efeitos "contextuais" não conduz a interpretações confiáveis.

Este modelo, dadas as suas limitações, é interessante sob o aspecto da modelagem, pela tentativa de explicitar diretamente os efeitos contextuais das escolas.

### 3.4.6 Modelo de Efeitos Contextuais: Componentes de Variância

Admita-se que os efeitos contextuais das escolas sejam descritos por uma variável aleatória  $\xi_i$ , com  $\xi_i \sim N(0, \sigma_i^2)$  e que esses efeitos sejam incorporados no seguinte modelo:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \xi_i + \varepsilon_{ij}; \quad i = 1, \dots, k; \quad j = 1, \dots, n_i \quad (3.16)$$

onde:  $y_{ij}$  = variável resposta do  $j$ -ésimo aluno da  $i$ -ésima escola

$x_{ij}$  = variável característica do  $j$ -ésimo aluno da  $i$ -ésima escola

$\varepsilon_{ij}$  = ruído aleatório  $N(0, \sigma^2)$ , independente de  $\xi_i$

A inclusão dos efeitos "escola" pode, por outro lado, ser considerada como em 3.4.3:

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \varepsilon_{ij}; \quad j = 1, \dots, n_i; \quad i = 1, \dots, k \quad (3.17)$$

e os efeitos são avaliados através de  $\beta_{0i}$ . Evidentemente, os valores  $\beta_{0i}$  para cada escola são valores observados de variáveis aleatórias com densidade

$N(\beta_0, \sigma_1^2)$ . O valor de  $\beta_0$  representa, portanto, o intercepto para uma escola "média".

As observações  $y_{ij}$ , dentro de uma escola, não são independentes; pois,  $\text{cov}(y_{ij}, Y_{ij'}) = \sigma_1^2$  se  $j \neq j'$ . Para avaliar o grau de homogeneidade de estudantes dentro de uma mesma escola e, conseqüentemente, comparar com estudantes de outras, pode-se utilizar o coeficiente de correlação para observações dentro de uma unidade:

$$\begin{aligned} \rho &= \sigma_1^2 / (\sigma_1^2 + \sigma^2) & (3.18) \\ (y_{ij} &= \beta_0 + \beta_1 x_{ij} + \xi_i + \varepsilon_{ij} \\ \text{var}(y_{ij}) &= \text{var}(\xi_i) + \text{var}(\varepsilon_{ij}) = \\ &= \sigma_1^2 + \sigma^2 \end{aligned}$$

Os ajustes desses modelos podem ser realizados utilizando o algoritmo de Fischer (MCCULLAGH e NELDER, 1989; DOBSON, 1990), obtendo-se estimativas de máxima verosimilhança e erros padrões (via matriz de informação) para os coeficientes da regressão e para  $\hat{\sigma}_1^2$  e  $\hat{\sigma}^2$ .

O modelo difere do descrito em 3.4.4 pelo fato que o ajuste dos efeitos contextuais é avaliado a partir da inclusão de valores médios, por escola, das características dos estudantes considerados, enquanto que neste, esta inclusão é realizada através de uma variável "efeito contextual". A variável "efeito contextual" assume valores positivos se a escola em questão estiver acima de uma escola "média" e negativo caso contrário.

AITKIN e LONGFORD (1986) argumentam que em modelos mais complexos (em termos do número de variáveis explicativas) os erros padrões das variáveis agregadas a nível de escola tendem a se tornar bastante elevados. Assim, a inclusão desses tipos de variáveis explicativas para caracterizar efeitos "contextuais" não conduz a interpretações confiáveis.

Este modelo, dadas as suas limitações, é interessante sob o aspecto da modelagem, pela tentativa de explicitar diretamente os efeitos contextuais das escolas.

### **3.4.7 Ajustamento dos Modelos, Inferência, Medidas de Discrepância e Análise dos Resíduos**

O ajustamento de modelos de estatística multivariada representa o processo de estimação dos parâmetros dos modelos.

Dependendo do modelo que se está ajustando diferentes algoritmos podem ser utilizados. Por exemplo, quando o modelo a ser ajustado tem uma componente aleatória cuja função de ligação pertence à família das distribuições exponenciais, o algoritmo de estimação utiliza um processo iterativo, através do cálculo repetido de regressões lineares ponderadas. Durante o processo de estimação, o algoritmo utiliza a matriz dos valores esperados das derivadas segundas do logaritmo da verossimilhança (Matriz de Informação de Fischer) ao invés de utilizar a matriz correspondente de valores esperados.

O processo de inferência consiste, basicamente, dos seguintes passos:

- verificação da adequação do modelo como um todo e da realização de um estudo detalhado quanto as discrepâncias locais;
- verificação quanto a precisão e interdependência das estimativas dos parâmetros;
- construção de regiões de confiança;
- testes sobre os parâmetros de interesse;
- análise dos resíduos;

- realização de previsões.

O recurso à análise de gráficos é de grande importância no processo de inferência e ajustamento de modelos, por exemplo:

- **Resíduos Padronizados x Valores Ajustados:** esse gráfico, sem nenhuma tendência, é um forte indicativo de que a relação funcional variância/média proposta é satisfatória;
- **Resíduos x Covariáveis não presentes no modelo:** se nenhuma covariável adicional é necessária, então, nenhuma tendência deverá ser notada nos gráficos;
- **Exame do Comportamento dos Preditores Lineares:** fornece uma idéia bastante acessível sobre a influência de diversos níveis para uma (ou mais) covariável qualitativa.

No processo de análise de um conjunto de observações dois aspectos são de fundamental importância: a parcimônia e o ajustamento do modelo propriamente dito.

Um modelo ajustado sem parcimônia não é de grande utilidade, pois dificulta o processo de sumarização dos dados, tornando-se nada informativo. Por outro lado, pequenas discrepâncias podem ser toleradas enquanto que as grandes devem ser analisadas cuidadosamente. Elas podem ser oriundas do próprio processo de ajustamento, como podem indicar dados transcritos erroneamente ou não significativos ("outliers") ou, ainda, ser um indicativo da ineficiência do modelo reproduzir o fenômeno sob análise.

De forma clássica, importantes estatísticas são utilizadas para mensurar a discrepância existente entre os dados observados e o modelo ajustado: a função desvio; a estatística Qui-quadrado; coeficiente de determinação ajustado  $R^2$ , estatística F, estatística t de Student, etc.

A análise dos resíduos quando do processo de ajuste de um modelo de regressão é de extrema valia. Os resíduos podem ser utilizados para explorarem a adequação do ajuste com relação a:

- escolha da função variância;
- escolha da função de ligação;
- análise dos termos componentes do preditor linear;
- detecção de valores anômalos.

Dentre os vários tipos de resíduos que se pode obter destacam-se os seguintes:

- Resíduos de Pearson:

$$r_p = (y - \mu)/(V(\mu))^{1/2}, \text{ com } \sum r_p^2 \sim \chi^2 \quad (3.19)$$

- Resíduos de Ascomb:

$$A(\bullet) = \int d\mu/V^{1/3}(\mu) \quad (3.20)$$

• Resíduos Função Desvio: se a função desvio é utilizada como medida de discrepância para um modelo linear, então, cada unidade contribui com uma quantidade  $d_i$  para aquela discrepância e o resíduo é calculado como:

$$\begin{aligned} \sum d_i &= D & (3.21) \\ r_d &= \text{sign}(y - \mu)\sqrt{d_i} \\ \sum r_d^2 &= D \end{aligned}$$

no caso da distribuição Poisson, por exemplo,  $r_d$  é igual a:

$$r_d = \text{sign}(y - \mu)\{2(y \log(y/\mu) - y + \mu)\}^{1/2} \quad (3.22)$$

A utilização dos resíduos de Pearson torna-se problemática quando a normalidade dos dados não é verificada. Nesses casos, os resíduos de Ascomb procuram mecanismos de diminuir ou mesmo eliminar o problema da não normalidade. É, também, uma boa prática analisar o gráfico dos resíduos ordenados com os valores obtidos através de uma distribuição conhecida acumulada (por exemplo, resíduos de Pearson ordenados x Valores acumulados da Normal padrão): se o ajuste for bom este gráfico deve se aproximar de uma reta.

### **3.5 Modelos Hierarquizados**

A informação necessária para avaliar resultados positivos no desempenho de um sistema educacional não deve basear-se apenas em medidas de desempenho dos alunos, mas, também, no conhecimento de como esse desempenho é influenciado por fatores sobre os quais os responsáveis pela gestão do sistema educacional possuem alguma influência e controle.

É sabido que, por exemplo, condições físicas de estabelecimentos escolares exercem um efeito bastante acentuado nos alunos. Escolas com excelentes laboratórios tem mais condições de fixar a atenção dos alunos do que aquelas menos equipadas. A política adotada para distribuição das atividades escolares pode conduzir a uma maior ou menor atratividade dos alunos em permanecerem nas escolas.

Efeitos como os citados anteriormente, podem ser verificados dentro de uma classe, dentro de uma escola ou mesmo entre escolas. Identificar e prever esses efeitos exige a modelação estatística da variação e das relações de cada um dos níveis. Os modelos hierarquizados fornecem uma ferramenta analítica para estes estudos.

Do ponto de vista teórico os procedimentos baseados em modelos hierarquizados tem sido utilizados no contexto de modelos de projetos experimentais de efeitos aleatórios ou mistos, modelos de regressão com coeficientes aleatórios, estimação Bayesiana etc. (SWAMINATHAN, 1989). BURSTEIN in SWAMINATHAN (1989) foi um dos primeiros pesquisadores a propor que os coeficientes de uma equação de regressão relacionados a um resultado direto de uma variável e um conjunto de preditores pudessem, eles próprios, serem a variável resposta de outro modelo e, portanto, explicados em termos de outros preditores lineares.

O conceito de "coeficientes como resposta" é atraente e possui um grande potencial para a modelação e explicação de resultados educacionais. Sua aceitação não foi imediata devido a ausência de procedimentos adequados para a estimação de seus parâmetros. Entretanto, estudos propostos por LINDLEY e SMITH (1972), no contexto da estimação Bayesiana, forneceram bases para a solução de alguns dos problemas de estimação existentes através do uso de modelos lineares hierárquicos.

O problema da estimação em modelos hierárquicos é discutido em detalhes nos trabalhos de RAUDENBUSH e BRYK (1989) e BRYK e RAUDENBUSH (1989). Para obter-se modelos quantitativos para a estimação da eficiência das escolas, o problema da modelação multiníveis é posto da seguinte forma, admita-se que, dentro de cada escola, o modelo para o vetor de resultados  $Y_i$  é:  $Y_i = X_i\beta_i + r_i$ ,  $i = 1, \dots, k$ .

Onde:  $Y_i$  é um vetor  $n_i \times 1$  e representa as respostas de  $n_i$  estudantes na escola  $i$ ;

$X_i$  é uma matriz conhecida ( $n_i \times p$ ) incluindo informações como, por exemplo, desempenho acadêmico e informações demográficas;

$\beta_i$  ( $p \times 1$ ) vetor de parâmetros desconhecidos, específicos para a escola  $i$ , incluindo um nível base para os resultados e coeficientes de regressão os quais especificam a magnitude dos efeitos das características básicas dos estudantes sobre os resultados;

$r_i$  ( $n_i \times 1$ ) é o vetor de erros aleatórios, ortogonal a  $X_i$  com valor esperado;  $E[r_i] = 0$  e matriz de dispersão arbitrária  $E[r_i r_i'] = \Sigma_i$ .

O vetor de parâmetros  $\beta_i$  contém um intercepto e diversos coeficientes de regressão. O intercepto indica o nível base para as saídas dentro da escola  $i$  (valor médio para cada escola, independente das características consideradas). Os coeficientes de regressão indicam a extensão dos efeitos das informações básicas admitidas e um conhecimento inicial sobre os resultados. Então, o vetor  $\beta_i$  inclui todas as informações relevantes a respeito do nível médio e a distribuição dos resultados dentro da escola  $i$ .

Através das escolas, os parâmetros  $\beta_i$  são aleatórios de forma que:

$$\beta_i = W_i \Phi + Z_i \Theta + u_i \quad (3.23)$$

Cada vetor de parâmetros para uma escola varia sistematicamente como uma função de  $W_i$  e  $Z_i$ . A matriz  $W_i$  representa as características normalmente não manipuladas através de políticas, incluindo a composição de estudantes (por exemplo, nível médio e dispersão do status sócio-econômico, características étnicas, conhecimento acadêmico) e o contexto comunitário (ex., localização urbana x localização rural, nível econômico da comunidade etc.). Pode-se admitir que essas variáveis sejam consideradas como composição ou contextuais para a escola.

A matriz  $Z_i$  inclui variáveis representando políticas, práticas, características organizacionais e os processos da escola que possam influenciar  $\beta_i$ . Como exemplos, citam-se: estado de conservação; área de recreio; campos de jogos; qualidade física das instalações; taxas de ocupação, etc.

A componente erro,  $u_i$ , representa práticas, de uma única escola ou não observáveis, as quais influenciam os resultados. Esses erros, por adequação, são assumidos ortogonais a  $W_i$  e  $Z_i$ . Se alguns dos elementos do

vetor  $u_i$  forem iguais a zero, tal característica irá indicar que, a componente  $\beta_j$  correspondente é invariante através das escolas ou a variabilidade para aquela componente é determinada completamente por  $W_j$  e/ou por  $Z_j$ .

A componente  $u_i$  é assumida ter expectância  $E[u_i] = 0$  e matriz de dispersão  $E[u_i u_i'] = \delta$ .

Combinando (3.22) e (3.23), obtém-se o seguinte modelo:

$$Y_i = X_i W_i \Phi + X_i Z_i \Theta + X_i u_i + r_i \quad (3.24)$$

o qual representa as saídas de  $Y_i$  como uma função dos efeitos principais e interações envolvendo variáveis características dos estudantes,  $X_i$ ; variáveis de contexto estrutural para as escolas  $W_i$ , variáveis políticas  $Z_i$  e um erro aleatório.

O modelo pode ser reescrito:

$$Y = XW\Phi + XZ\Theta + Xu + r \quad (3.25)$$

Pré-multiplicando (3.24) por  $(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$  obtém-se a equação equivalente:

$$\hat{\beta} = W\Phi + Z\Theta + e, \quad e = u + \varepsilon \quad (3.26)$$

onde:  $\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$  é o estimador de mínimos quadrados generalizados de  $\beta$ , baseado nos dados de cada escola e  $\varepsilon = \hat{\beta} - \beta$ .

Assim, condicionado ao conhecimento de  $\Sigma$ , o vetor dos interceptos estimados e dos coeficientes  $\hat{\beta}$  contém toda a informação necessária sobre o nível médio e a distribuição dos resultados dentro das escolas. Esses interceptos estimados (assim como os coeficientes) dependem das variáveis contextuais  $W$ ,

das variáveis políticas  $Z$  e de um vetor aleatório (que se assume ter  $E[e] = 0$  e dispersão  $E[ee'] = \Delta$ , onde  $\Delta = (X'\Sigma^{-1}X)^{-1} + \varepsilon$ ).

Tecnicamente, o vetor  $\hat{\beta}$  é suficiente para  $\beta$ , condicionado ao conhecimento das matrizes de dispersão.

Definindo o efeito escola como sendo o vetor  $\tau_i$ , onde:

$$\tau_i = \beta_j - W_i\Phi = Z_i\theta + u_i \quad (3.27)$$

Isto é, o conhecimento de  $\tau_i$  poderia conter informações necessárias para o estabelecimento de como as variáveis políticas  $Z_i$  e os efeitos únicos  $u_i$  influenciam no nível médio e na distribuição dos resultados na escola  $i$ .

De outro modo,  $\tau_i$  é a componente para a escola  $i$  do vetor de parâmetros  $\beta_j$ . Esse vetor de parâmetros é resultado das políticas e práticas adotadas, em oposição aos conhecimentos dos estudantes e da composição da escola.

### 3.6 Modelos baseados em Análise Taxinômica

De acordo com TAVARES (1991) "esses modelos correspondem a uma abordagem distinta da tradicional no estudo de problemas da Educação em que se definem as classes segundo hipóteses baseadas em pressupostos teóricos e se analisam os dados para validar ou anular tais hipóteses. Pelo contrário, esses modelos permitem guiar o analista na construção das classes, o que se julga não menos útil nem profícuo para a compreensão das relações entre Sociedade e Educação".

Procedimentos exploratórios são úteis, particularmente, na compreensão da natureza complexa de relacionamentos multivariados. Pesquisar nos dados uma estrutura de agrupamento natural é uma importante técnica exploratória. Os grupos encontrados podem sugerir hipóteses bastante interessantes a respeito do relacionamento existente entre as variáveis analisadas.

A análise de grupos é uma técnica que não faz uso de qualquer hipótese sobre o número e/ou a estrutura dos grupos que serão encontrados. Os agrupamentos são realizados com base em medidas de similaridades ou distâncias. O objetivo básico da análise de grupos ou análise taxionométrica é o de descobrir um agrupamento natural de itens ou de variáveis.

### **3.6.1 Medidas de Similaridade**

Identificar ou produzir uma estrutura de grupos a partir de um conjunto complexo de dados requer uma medida de proximidade ou de similaridade.

Existe, naturalmente, uma grande parcela de subjetividade envolvida na escolha de tal medida e importantes considerações devem ser feitas com relação a natureza das variáveis, escalas de medidas e conhecimento subjetivo do assunto analisado. Porém, quase que na maioria das aplicações de análise taxionométrica, quando o interesse volta-se para agrupar itens (ou casos) uma medida de proximidade baseia-se ou é indicada por algum tipo de distância. Já quando o interesse é o de agrupar variáveis, recomenda-se a utilização de alguma medida de associação, como por exemplo, coeficiente de correlação.

Medidas de distância, baseiam-se, em geral, na métrica de Minkowski:

$$d(x, y) = \left[ \sum_{i=1}^p |x_i - y_i|^m \right]^{1/m} \quad (3.28)$$

onde  $d(x, y)$  representa a distância entre dois pontos no espaço  $p$ -dimensional. Quando  $m = 1$ , obtém-se a distância "city-block"; para  $m = 2$ , a distância calculada é a bem conhecida distância euclidiana. De modo geral, a variação de  $m$  resulta em uma maior ou menor ponderação dada as maiores (e, conseqüentemente, as menores) diferenças.

A partir das distâncias pode-se construir medidas de similaridade para itens, de modo a tornar possível compará-los, por exemplo, com base na presença ou na ausência de determinadas características. A idéia básica é a de que itens similares apresentam mais características comuns do que itens dissimilares. As medidas de similaridades para associações entre variáveis, por outro lado, normalmente são aferidas tendo-se como base coeficientes de correlações amostrais. Evidentemente, nota-se aqui a ligação existente com o tradicional teste de independência entre variáveis.

A partir da observação de medidas de distâncias ou de similaridades torna-se possível formar os grupos pretendidos. Técnicas específicas devem ser utilizadas para a realização de tal tarefa.

### 3.6.2 Técnicas para Realizar Agrupamentos

Difícilmente é possível o exame de todas as possibilidades de agrupamentos de itens ou de variáveis. Dada essa dificuldade, uma extensa variedade de algoritmos para agrupamentos tem sido propostos. As técnicas de agrupamentos existentes podem ser subdivididas em dois macros grupos: métodos hierárquicos e os métodos não-hierárquicos.

Métodos hierárquicos realizam séries sucessivas de divisões ou junções sobre os dados de modo a identificar os grupos existentes. Os métodos hierárquicos podem então ser considerados como sendo aglomerativos quando procuram agrupar os objetos sob análise, a partir do número total de objetos existentes. Desse modo, tem-se, inicialmente, tantos grupos quanto objetos. Em direção oposta, encontram-se os métodos hierárquicos divisíveis, os quais, consideram, em primeiro lugar, todos os objetos como sendo um único grupo e, mediante divisões sucessivas, pretendem estabelecer agrupamentos que reproduzam grupos com características semelhantes.

Dentre os procedimentos hierárquicos aglomerativos, destaque se dá a uma categoria específica, a qual pode ser útil, tanto para o caso de itens quanto para o caso de variáveis: os procedimentos baseados em acoplamentos de itens ou de variáveis em termos das distâncias verificadas (mínima, máxima e média). Para a distância mínima, os elementos são agrupados considerando-se em conta a distância entre os membros mais próximos. Utilizando-se a distância máxima os elementos são agrupados considerando-se a distância máxima verificada entre os seus elementos. No caso da distância média, os grupos são formados considerando-se a distância média entre pares de seus elementos. Esses métodos são, como se pode concluir, sensíveis a elementos não característicos e sua estabilidade deve ser verificada através de pequenas perturbações introduzidas nos dados iniciais. Se os grupos foram bem estabelecidos tais perturbações não deverão influir nos resultados observados.

Os métodos não-hierárquicos são mais utilizados para agrupar itens em uma coleção de  $K$  grupos. O número de grupos pode ser estabelecido ou determinado como parte dos algoritmos de agrupamento. Dadas as suas características computacionais, esses procedimentos não necessitam do armazenamento de matrizes de distâncias (ou similaridades) durante a fase de processamento, podem ser aplicados a conjuntos mais extensos de dados.

Os procedimentos não-hierárquicos são inicializados pela definição de uma partição inicial de grupos ou através de um conjunto inicial de sementes.

o qual, estabelece um núcleo para os grupos. O procedimento não-hierárquico mais popularizado é o método das k-Médias.

### 3.7 Conclusões

Neste capítulo foram discutidas questões relacionadas a modelagem de sistemas. O capítulo descreve, inicialmente, aspectos sobre a construção de modelos e sua utilização para análise de sistemas.

Com relação a análise de sistemas educacionais, é discutido a construção de modelos que forneçam maiores subsídios para interpretação de questões de fundamental importância para a compreensão desse tipo de sistema, em contraposição à utilização de apenas indicadores de desempenho.

A construção de um modelo estatístico envolve discutir aspectos relacionados a sua estrutura funcional e a sua estrutura distribucional. É a partir da estrutura funcional que se consegue analisar relações entre variáveis envolvidas no modelo. A estrutura distribucional procura explicitar, em termos probabilísticos, o comportamento das variáveis utilizadas no modelo.

Na maioria das vezes, os modelos são construídos a partir de hipóteses realizadas sobre a estrutura distribucional. A escolha de uma forma particular de distribuição para os parâmetros do modelo pode ter como consequência uma representação errônea da realidade que este se propõe estudar. Sistemas Educacionais são sistemas nos quais relações humanas estão envolvidas. As relações humanas são, em sua gênese, relações difusas.

## **4. TEORIA DOS CONJUNTOS DIFUSOS E TEORIA DAS POSSIBILIDADES**

### **4.1 - Introdução**

"Na medida em que as leis da matemática refletem a realidade, elas não são exatas. Na medida em que elas são exatas não refletem a realidade."  
(Albert Einstein).

A contradição apontada por Einstein é sentida, com maior intensidade, quando se tenta modelar sistemas, nos quais, componentes sociais, como é o caso dos Sistemas Educacionais, estão presentes. Ao engenheiro de sistemas, compete produzir ferramentas e meios de expressar aqueles relacionamentos, embora, na maioria das vezes os conceitos e termos utilizados não se apresentem da forma mais apropriada para o tratamento de incertezas associadas.

O tratamento de incerteza com base em modelos sustentados pela teoria tradicional de probabilidades tem sido objeto de ampla discussão e debate. O problema da modelação da incerteza tem acarretado no surgimento de novas ferramentas de análise baseadas em diferentes raciocínios e diferentes graus de incertezas.

A teoria dos conjuntos difusos desenvolvida por ZADEH (1965) fornece um instrumento adequado para modelar situações em que ocorram imprecisões ou incertezas. Os recentes progressos verificados nessa teoria, na teoria das possibilidades e nas teorias correlatas como, por exemplo, a teoria da evidência de Shaffer, tem contribuído para dissipar aspectos não muito claros e tornado viável o desenvolvimento de ferramentas de análise bastante poderosas.

Essas teorias, além de permitir a generalização de medidas de incerteza, possibilitam a ampliação considerável de aplicações de caráter prático.

## 4.2 - Considerações sobre a Lógica Clássica

A formalização do raciocínio dedutivo está fundamentada na lógica clássica. Assim denominada por ter sua origem no primeiro estudo sobre lógica, creditado a Aristóteles (384 - 322 A.C.). Essa lógica trabalha essencialmente com verdades e falsidades de um argumento.

A lógica clássica fundamenta-se em três princípios básicos:

- princípio da identidade

$$X = X$$

- princípio do meio excluído

Dados  $A$  e  $\sim A$ , uma é verdadeira

- princípio da contradição

Dados  $A$  e  $\sim A$ , uma é falsa

"Toda lógica tradicional habitualmente assume que símbolos precisos estão sendo aplicados. Ela não é, portanto, aplicável à vida terrestre, mas somente a uma existência celestial imaginada (...) a lógica coloca-nos mais próximo ao céu do que outros estudos" (Bertrand Russel).

## 4.3 - Definição de Conjuntos Difusos

Os conjuntos clássicos apresentam limites bem definidos. Na teoria

clássica dos conjuntos, um elemento pertence ou não pertence a um conjunto, não existindo uma situação intermediária.

Então, a noção clássica de pertinência é dada por:

Seja  $A$  um conjunto em  $U$ :  $A \subset U$ .

Pode-se indicar a pertinência de um elemento  $x$  em  $U$  que esteja em  $A$  por:  $x \in A$ .

Pode-se caracterizar essa pertinência por uma função com os seguintes valores:

$$A(x) = \begin{cases} 1, & \text{se e somente se } x \in A \\ 0, & \text{se e somente se } x \notin A \end{cases}$$

Entretanto, no mundo real, existe uma série de conjuntos que não apresentam limites bem definidos, ou seja, a pertinência de um elemento ao conjunto não pode ser especificada por um critério binário do tipo "sim" ou "não". Foi sob essas bases, que ZADEH (1965) propôs e desenvolveu a teoria dos conjuntos difusos. Essa teoria diz que um conjunto não necessariamente apresenta limites bem definidos, podendo um elemento pertencer parcialmente a ele, ou pertencer a dois conjuntos ao mesmo tempo. O que caracteriza-o é o "grau de pertinência", que é uma medida que quantifica o grau com que o elemento pertence a um dado conjunto.

Matematicamente, conjunto difuso é definido como: Se  $X$  é uma coleção de objetos denotados genericamente por  $x$ , então um conjunto difuso  $A$  em  $X$  é um conjunto de pares ordenados:  $A = \{ (x, \mu_A(x)) \mid x \in X \}$ .  $\mu_A(x)$  é chamada função de pertinência ou grau de pertinência de  $x$  em  $A$ , a qual mapeia  $X$  para o espaço de pertinência  $M$ . Quando  $M$  contém somente os dois pontos "0" e "1", o conjunto  $A$  é não difuso e a função  $\mu_A(x)$  é idêntica a função característica de um conjunto não difuso (ZIMMERMANN, 1985).

Os graus de pertinência refletem uma "ordenação" dos objetos no universo. É interessante notar que o grau de pertinência  $\mu_A(x)$  de um objeto  $x$  em  $A$  pode ser interpretado como o grau de compatibilidade do predicado associado com  $A$  e o objeto  $x$ . É também possível interpretar  $\mu_A(x)$  como o grau de possibilidade onde  $x$  é o valor de um parâmetro restringido, de forma difusa por  $A$ .

A definição de um conjunto difuso pressupõe a generalização da noção clássica de pertinência. Na teoria dos conjuntos difusos, os valores de pertinência não são mais apenas "0" e "1" e sim um conjunto de valores reais positivos.

A representação mais usual de conjunto difuso é dado por:

$\sum \mu_A(x) | X$ , para todo  $x$  no universo do conjunto, onde a função de pertinência é discreta.

$\int_x \mu_A(x) / X$ , para todo  $x$  no universo do conjunto, onde a função de pertinência é contínua.

#### 4.4 - Definições e Operações Básicas da Teoria dos Conjuntos Difusos

As definições a seguir são baseadas nas de ZIMMERMANN (1991) e DUBOIS e PRADE (1980, 1986, 1989) e são relevantes no contexto deste trabalho.

a) **Conjunto Difuso Normalizado:** Chama-se conjunto difuso normalizado todo conjunto difuso cujo maior grau de pertinência é 1.

b) **Conjunto Difuso Convexo:** é todo conjunto difuso em que, para dois pontos quaisquer de sua função de pertinência, o segmento de reta que os une também pertence ao conjunto.

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \text{MIM}(\mu_A(x_1), \mu_A(x_2)), \quad x_1, x_2 \in X \\ \lambda \in [0, 1]$$

c) **Número Difuso:** é um conjunto difuso convexo e normalizado em que:

- existe um e somente um ponto  $x_0$  para o qual o valor da função de pertinência seja 1;
- a função de pertinência que define o conjunto difuso é contínua por partes.

d) **Corte de Nível  $\alpha$ :** denomina-se corte de nível  $\alpha$  ao conjunto clássico formado pelos elementos  $x$  cuja função de pertinência seja maior que  $\alpha$ .

$$A = \{x \in X \mid \mu_A(x) \geq \alpha\}$$

e) **Suporte de um Conjunto Difuso  $A$ :** o suporte de um conjunto difuso  $A$ ,  $S(A)$ , é o conjunto de todo  $x \in X$  tal que  $\mu_A(x) > 0$ .

f) **Cardinalidade:** para um conjunto difuso finito  $A$ , a Cardinalidade  $|A|$  é definida como:

$$|A| = \sum \mu_A(x), \text{ onde } x \in X$$

g) **Igualdade de Conjuntos:** dois conjuntos difusos  $A$  e  $B$  são iguais ( $A=B$ ), se e somente se, para todo e qualquer  $x \in X$ ,  $\mu_A(x) = \mu_B(x)$ .

h) **Função de Pertinência para Intersecção de dois Conjuntos Difusos:** a função de pertinência  $\mu_C(x)$  da intersecção de  $C = A \cap B$ , é definida por:

$$\mu_C(x) = \text{MIM}(\mu_A(x), \mu_B(x)), \quad x \in X$$

i) **Função de Pertinência para União de dois Conjuntos Difusos:** a função de pertinência  $\mu_D(x)$  da união de  $D = A \cup B$ , é definida por:

$$\mu_D(x) = \text{MAX} (\mu_A(x), \mu_B(x)), x \in X$$

j) **Função de Pertinência para o Complemento de um Conjunto Difuso:** a função de pertinência do complemento de um conjunto difuso  $A$ ,  $\mu_{CA}(x)$  é definido por:

$$\mu_{CA}(x) = 1 - \mu_A(x), x \in X$$

k) **Produto Cartesiano de Conjuntos Difusos:** define-se o produto cartesiano de conjuntos difusos como segue: seja  $A_1, \dots, A_n$  conjuntos difusos em  $X_1, \dots, X_n$ . O produto cartesiano é o conjunto cartesiano no espaço produto  $X_1 * X_2 * \dots * X_n$ , com função de pertinência:

$$\mu_{(A_1 * \dots * A_n)}(x) = \text{MIN}_i \{ \mu_{A_i}(x_i) \mid x = (x_1, \dots, x_n), x_i \in X_i \}$$

l) **m-Ésima Potência de um Conjunto Difuso:** a m-ésima potência de um conjunto difuso  $A$  é um conjunto difuso com a função de pertinência dada por:

$$\mu_{A^m}(x) = [\mu_A(x)]^m, x \in X$$

m) **Soma Algébrica:** a soma algébrica (soma probabilística),  $C = A + B$  é definida como:

$$C = \{ (x, \mu_{A+B}(x)) \mid x \in X \}, \text{ onde:}$$

$$\mu_{A+B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) * \mu_B(x)$$

n) **Soma Limitada:** a soma limitada,  $C = A \oplus B$  é definida como:

$$C = \{ (x, \mu_{A \oplus B}(x)) \mid x \in X \}, \text{ onde:}$$

$$\mu_{A \oplus B} = \text{MIN} (1, \mu_A(x) + \mu_B(x))$$

o) **Diferença Limitada:** a diferença limitada,  $C = A - B$  é definida como:

$$C = \{(x, \mu_{A \oplus B}(x)) \mid x \in X\}, \text{ onde}$$

$$\mu_{A \oplus B}(x) = \text{MAX}(0, \mu_A(x) + \mu_B(x) - 1)$$

p) **Produto Algébrico:** o produto algébrico de dois conjuntos difusos,  $C = A * B$  é definido como:

$$C = \{(x, \mu_A(x) * \mu_B(x)) \mid x \in X\}$$

q) **Relação Difusa:** sejam  $X, Y \subseteq R$  conjuntos universais, então:

$$R = \{(x, y), \mu_R(x, y) \mid (x, y) \subseteq X, Y\}$$
 é chamada uma relação difusa sobre  $X, Y$

r) **Relação Difusa entre Conjuntos Difusos:** seja  $X, Y \subseteq R$  e

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

$$B = \{(y, \mu_B(y)) \mid y \in Y\}$$

dois conjuntos difusos.

Então:

$$R = \{(x, y), \mu_R(x, y) \mid (x, y) \in X, Y\}$$
 é uma relação difusa sobre  $A$  e

$B$ .

s) Sejam  $R$  e  $Z$  duas relações difusas no mesmo espaço, então:

$$\mu_{R \cup Z}(x, y) = \text{MAX}\{\mu_R(x, y), \mu_Z(x, y)\}, (x, y) \in X, Y$$

$$\mu_{R \cap Z}(x, y) = \text{MIN}\{\mu_R(x, y), \mu_Z(x, y)\}, (x, y) \in X, Y$$

t) **Projeção Difusa:**

Seja  $R = \{(x, y), \mu_R(x, y) \mid (x, y) \in X, Y\}$  uma relação binária difusa.

A primeira projeção de  $R$  é definida como:

$$R_{(1)} = \{(x, \text{MAX}_y \mu_R(x, y)) \mid (x, y) \in X, Y\}$$

A segunda projeção é definida como:

$$R_{(2)} = \{(y, \text{MAX}_x \mu_R(x, y)) \mid (x, y) \in X, Y\}$$

A projeção total é definida:

$$R_{(T)} = \text{MAX}_x \text{MAX}_y \{\mu_R(x, y) \mid (x, y) \in X, Y\}$$

u) Sejam  $R_1(x,y), (x,y) \in X.Y$  e  $R_2(y,z), (y,z) \in Y.Z$  duas relações difusas. A composição MAX - MIN,  $R_1 \text{ MAX-MIN } R_2$  é então o conjunto difuso definido por:

$$R_1 \circ R_2 = \left\{ \left[ (x,y) \text{ MAX}_y \left\{ \text{MIN} \left\{ m_{R_1}(x,y), m_{R_2}(y,z) \right\} \right\} \right] \text{ tal que } x \in X, y \in Y \text{ e } z \in Z \right\}$$

#### 4.5 Funções de Pertinência

A função de pertinência reflete o conhecimento que se tem em relação a intensidade com que um elemento pertence a um conjunto. Ela é utilizada para medir o grau de pertinência de um elemento a um conjunto. Quando o conjunto é difuso e normalizado, o grau de pertinência assume um valor no intervalo  $[0,1]$  (PAO, 1989).

O conceito de função de pertinência, não pode ser considerado como primário, porque o grau de pertinência não é um valor completamente definido. Como exemplo, pode-se citar o conceito "idade de uma pessoa": a maneira de uma pessoa perceber a idade de outra depende da idade dela própria. Em vista disso, pode-se dizer que a função de pertinência é difusa, pois sempre que tenha uma forma que satisfaça a avaliação e os princípios apresentados, poderia ser considerada como uma aproximação aceitável. Então, muitas funções de pertinência ordinárias seriam suficientes para uma representação quantitativa aproximada da noção qualitativa de pertinência gradual a uma categoria.

Para construção de funções de pertinência deve-se verificar as seguintes propriedades (DOMBI, 1990):

- todas funções de pertinência são contínuas;
- todas funções de pertinência mapeam um intervalo  $[a,b] \rightarrow [0,1] \rightarrow \mu[a,b]$

→ [0, 1];

- as funções de pertinência são:
  - monotonicamente crescentes, monotonicamente decrescentes, ou subdividida em parte crescente e parte decrescente;
- as funções de pertinência monótonas sobre um intervalo completo são:
  - funções convexas, ou
  - funções côncavas, ou
  - existe um ponto "c" no intervalo [a,b] tal que [a,c] é convexo e [c,b] é côncava.
- funções monotonicamente crescentes tem a propriedade  $\mu(a) = 0$  e  $\mu(b) = 1$ , enquanto funções monotonicamente decrescentes tem a propriedade  $\mu(a) = 1$  e  $\mu(b) = 0$ ;
- as funções de pertinência devem apresentar uma forma linear ou devem ser linearizadas.

Uma das funções de pertinência mais utilizadas, é a proposta por DOMBI (1990). Na realidade, Dombi contempla uma família de funções de pertinência definidas a partir do ajuste de quatro parâmetros básicos:  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$ .

A função  $\mu_x(x)$  é proposta considerando-se que variações na pertinência ocorrem de forma monotônica crescente e decrescente, como expressa na equação (4.1).

$$\mu_X(x) = \begin{cases} \frac{(1-\gamma)^{\delta-1}(x-\alpha)^{\delta}}{(1-\gamma)^{\delta-1}(x-\alpha)^{\delta} + \gamma^{\delta-1}(\beta-x)^{\delta}}, & \text{parte crescente} \\ \frac{(1-\gamma)^{\delta-1}(\beta-x)^{\delta}}{(1-\gamma)^{\delta-1}(\beta-x)^{\delta} + \gamma^{\delta-1}(x-\alpha)^{\delta}}, & \text{parte decrescente} \end{cases} \quad (4.1)$$

onde:  $\gamma$  = ponto de reflexão

$\alpha, \beta$  = intervalo de valores para x

$\delta$  = grau de distorção da curva

Deve-se salientar que, no caso particular de  $\gamma = 0,5$  e  $\delta = 1$  tem-se equações de retas.

Obviamente, a função de Dombi para ajustes requer a definição de dois parâmetros adicionais que definem o intervalo para o qual a função é decrescente.

Um método bastante eficaz para se processar o ajuste de uma função de pertinência, do tipo proposta por Dombi, consiste em executar o seguinte algoritmo:

- obter "graus de veracidade" sobre uma proposição (por exemplo, mediante a aplicação de questionário) quando importadas com o valor da variável.
- criar uma tabela que contenha frequências ou médias para os "graus de veracidade".
- estimar, através de mínimos quadrados os valores dos coeficientes  $\gamma$  e  $\delta$ .

#### 4.6 - Conexão da Função de Pertinência com os Operadores

De acordo com DOMBI (1982) a conexão entre a classe racional de funções de pertinência e a forma racional dos operadores agregados é dada por:

- um operador agregado é uma função contínua e estrita, monotonicamente crescente  $a: [0,1] \rightarrow [0,1]$ , satisfazendo:
  - $a(0,0) = 0$

- $a(1,1) = 1$
  - $a(x,a(y,z)) = a(a(x,y),z)$  (Associativa)
  - $a(x,y) = n(a(n(x),n(y)))$  (Distributiva)
- o símbolo  $[ ]$  tem o seguinte significado:  $[x] = \begin{cases} 1 & \text{se } x \geq 1 \\ x & \text{se } 0 < x < 1 \\ 0 & \text{se } x \leq 0 \end{cases}$
- o tipo conjuntivo (disjuntivo) de funções de pertinência são nos casos estrito e não estrito, respectivamente, figuras 4.1, 4.2, 4.4. e 4.4:

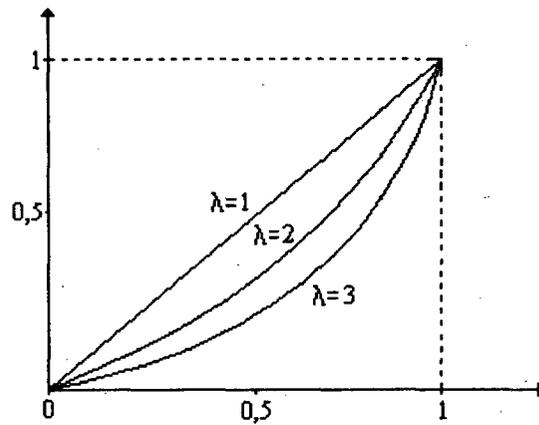


FIGURA 4.1- Funções Conjuntivas Estritamente Monotônicas

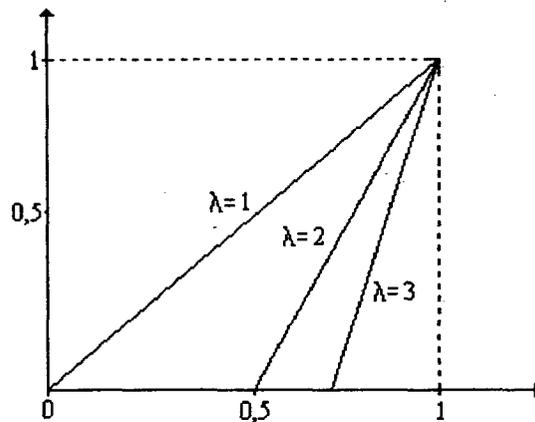


FIGURA 4.2 - Funções Conjuntivas não Estritamente Monotônicas

$$\mu_c(x) = f^{-1}\left(\lambda f\left(\frac{x-a}{b-a}\right)\right), \quad \mu_c(x) = f^{-1}\left[\lambda f\left(\frac{x-a}{b-a}\right)\right], \quad x \in [a,b]$$

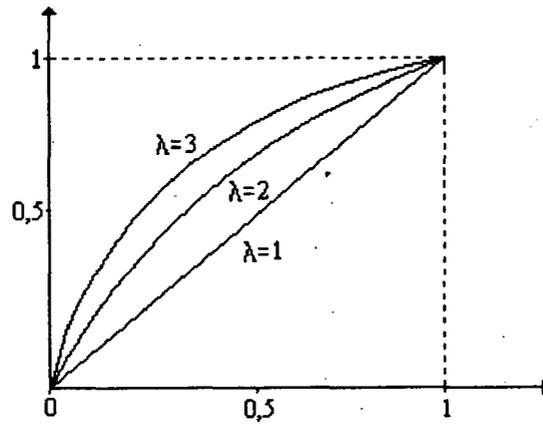


FIGURA 4.3 - Funções Disjuntivas Estritamente Monotônicas

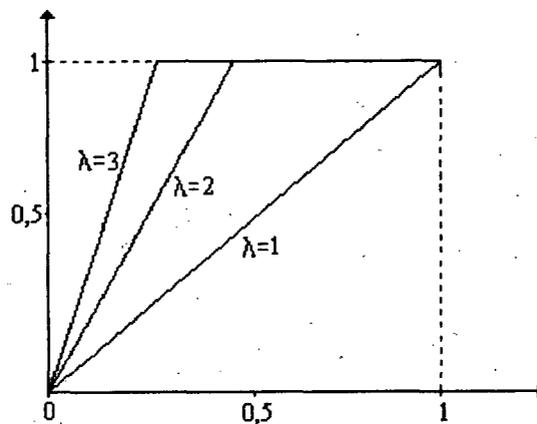


FIGURA. 4.4 - Funções Disjuntivas não Estritamente Monotônicas

$$\mu_d(x) = g^{-1}\left(\lambda g\left(\frac{x-a}{b-a}\right)\right), \quad \mu_d(x) = g^{-1}\left[\lambda g\left(\frac{x-a}{b-a}\right)\right], \quad x \in [a,b]$$

#### 4.7 - Estimação de Funções de Pertinência

Um dos tópicos importantes para a aplicação da teoria dos conjuntos difusos é a estimativa da função de pertinência.

Segundo DEVI e SARMA (1985), para estimativa da função de pertinência os pontos importantes a serem considerados são a forma, os parâmetros e o domínio da função. Entre os métodos mais utilizados pode-se citar: o baseado em histograma e o baseado no consenso de especialistas.

No método baseado em histograma, segundo DEVI e SARMA (1985), o número de classes, para um nível de significância de 0,05, é obtido através da seguinte fórmula:

$$m = 1,87(n-1)^{2/5} \quad (4.2)$$

onde:  $m$  = número de classes

$n$  = tamanho da amostra

Para estimar a função de pertinência a partir do histograma pode ser utilizada uma função racional, onde ajusta-se os parâmetros usando-se o método dos mínimos quadrados, sendo que a função deve ser normalizada.

Para obter-se a função de pertinência baseada em histograma de uma outra maneira, inicialmente, estima-se a função densidade de probabilidade ( $p(x)$ ). A função de pertinência é então estimada através da seguinte forma:

$$f(x) = \begin{cases} p(x), & \text{se } \beta \cdot p(x) \leq 1 \\ 1, & \text{se } \beta \cdot p(x) > 1 \end{cases}$$

onde a constante  $\beta$  é um parâmetro de ajuste.

No método baseado em consenso de especialistas, obtém-se a função de pertinência com base em opiniões subjetivas de especialistas da área considerada. São feitas entrevistas com especialistas com o objetivo de valorar declarações lingüísticas. Dada uma declaração do tipo "X é Y", o especialista deve manifestar sua concordância ou não com essa declaração. Essa concordância pode ser expressa na forma binária (sim ou não) ou na forma multivalorada (por exemplo, numa escala de 0 a 10).

A partir dessas respostas, são calculadas as freqüências relativas (no caso binário) ou a média dos graus de concordância (no caso multivalorado). É, então, construído um gráfico dos pontos assim obtidos, e ajustada uma função, estimando-se os parâmetros através dos mínimos quadrados, sendo essa função normalizada.

#### **4.8 - Teoria da Possibilidade**

A Teoria da Possibilidade enfoca a imprecisão intrínseca na linguagem natural, e esta assume ser antes "possibilística" do que "probabilística".

A diferença entre os conceitos de probabilidade e de possibilidade é que a probabilidade é objetiva, baseada em dados estatísticos, pode ser descrita pelas freqüências da ocorrência de eventos e é geralmente aditiva. Enquanto que a possibilidade é uma estrutura de nosso conhecimento subjetivo, não aditivo e dependente de seu meio (KANDEL, 1986), (DUBOIS e PRADE, 1986).

De acordo com ZADEH (1978), a teoria da possibilidade é definida como:

Seja Y uma variável assumindo valores no Universo U; então a

distribuição de possibilidade  $\pi_Y$ ; associada a  $Y$  pode ser vista como uma restrição difusa sobre os valores que podem ser associados a  $Y$ . Tal distribuição é caracterizada por uma função de distribuição de possibilidade  $\pi_Y: U \rightarrow [0,1]$  a qual associa cada  $u \in U$ , o "grau de liberdade" ou a possibilidade de que  $Y$  tenha  $u$  como seu valor.

Em alguns casos, as restrições sobre os valores de  $Y$  é física na origem; em muitos casos, entretanto, a distribuição de possibilidade que é associada a variável, é epistêmica antes do que física. Uma suposição básica em lógica difusa é que tais distribuições de possibilidades epistêmicas são induzidas por proposições expressas em uma linguagem natural. Essa suposição pode ser estabelecida com o seguinte postulado:

Se  $F$  é um conjunto difuso no universo  $U$  caracterizado por sua função de pertinência  $\mu_F: U \rightarrow [0,1]$ , então a proposição " $Y$  é  $F$ " induz uma distribuição de possibilidade  $\pi_Y$  que é igual a  $F$ . Equivalentemente " $Y$  é  $F$ " translada para a equação de possibilidade  $\pi_Y = F$ , que é:  $Y \text{ é } F \rightarrow \pi_Y = F$ , a qual significa que a proposição  $Y$  é  $F$  tem o efeito de restringir os valores que podem ser assumidos por  $Y$ , com a distribuição de possibilidade,  $\pi_Y$  identificada com  $F$ .

Exemplo: Seja a proposição "Maria é Gorda"; essa proposição induz a distribuição de possibilidade definida por:

$\pi_{\text{peso}}(\text{Maria}) = \text{Gorda}$ , onde "Gorda" é um conjunto difuso do universo de pesos, e  $\text{peso}(\text{Maria})$  é uma variável. Então pode-se escrever:  $\text{Poss}(\text{Peso}(\text{Maria}) = u) = \mu_{\text{Gordo}}(u)$ , onde  $\text{Poss}(\text{Peso}(\text{Maria}) = u)$  lê-se "a possibilidade de que peso de Maria seja igual a  $u$ " onde  $u$  é um valor específico da variável peso de Maria e  $\mu_{\text{Gordo}}(u)$  é o grau de pertinência  $u$  no conjunto difuso "Gordo".

## 4.9 - Regressão Possibilística

Modelos nos quais o julgamento humano exerça uma influência acentuada são modelos difusos por natureza.

Fenômenos difusos podem ser representados através de sistemas lineares possibilísticos (TANAKA, 1982). Um sistema linear possibilístico é definido por parâmetros difusos, os quais representam uma distribuição de possibilidade (DUBOIS e PRADE, 1980).

Em modelos tradicionais de regressão, os desvios são supostos decorrentes de erros de mensuração. Em modelos lineares difusos, os desvios são atribuídos a flutuações nos parâmetros do sistema. A estrutura do sistema não pode ser completamente caracterizada dado a existência de imprecisões e/ou presença de fenômenos nebulosos em sua formação.

### 4.9.1 - Sistemas Lineares Possibilísticos

Sistemas lineares possibilísticos são sistemas lineares para os quais os parâmetros são definidos por distribuições de possibilidade.

Seja  $Y$  uma variável assumindo valores em  $X$ ; então a distribuição de possibilidade,  $\pi_Y$ , associada com  $Y$  pode ser entendida como sendo uma restrição difusa sobre os valores que podem ser assumidos por  $Y$ . A distribuição de possibilidade é caracterizada por uma função distribuição de possibilidade  $\pi_Y: X \rightarrow [0,1]$  a qual associa com cada valor  $x \in X$  um "grau da possibilidade" que  $Y$  pode assumir  $x$  como um valor (KANDEL, 1986). As distribuições de possibilidade são originadas por expressões de linguagens naturais e podem ser

estabelecidas através do postulado de possibilidade (ZADEH, 1978) (KANDEL, 1986).

Uma distribuição de possibilidade não é estatística por natureza. Como consequência se  $P[Y=y]$  representa a distribuição de probabilidade de  $Y$ , então,  $\pi_Y$  e  $P_Y$  terão como única conexão o fato de que impossibilidade implica em probabilidade nula (mas, o contrário não é verdadeiro!).  $\pi_Y$  não pode ser inferido de  $P_Y$  e, tampouco,  $P_Y$  ser inferido de  $\pi_Y$ .

Uma distribuição de possibilidade  $\pi_Y$  pode ser representada por um número difuso  $A$ , o qual satisfaz as seguintes condições:

- $[A]_h = \{a | \mu_A(a) \geq h\}$  é um intervalo fechado  $\forall h \in [0,1]$
- Existe um  $a$  tal que  $\mu_A(a) = 1$
- O número difuso é convexo, isto é,  $\forall \lambda \in [0,1]$ :

$\mu_A(\lambda a_1 + (1-\lambda)a_2) \geq \mu_A(a_1) \wedge \mu_A(a_2)$ , onde  $\mu_A(a)$  é uma função de pertinência para um número difuso  $A$  e  $\wedge$  denota um mínimo. Para efeitos de análise de regressão possibilística, considere-se que  $A$  é um número difuso simétrico, representado por  $A_i = (\alpha_i; c_i)$ .

Considere-se a seguinte função  $f(x,a)$ , do conjunto  $X$  sobre o conjunto  $Y$ , onde  $x = (x_1, \dots, x_n)^T$  e  $a = (a_1, \dots, a_n)$ . Então  $f(x,a)$  é um mapeamento de  $X$  em  $Y$ . Se os parâmetros forem formados por números difusos  $A_i$ , então, através do princípio da extensão,  $f(x,a)$  é uma função difusa. O conjunto difuso  $Y$  pode ser mapeado através de  $A$  de acordo com a seguinte definição (TANAKA, UEGIMA e ASAI, 1982):

$$f: X \rightarrow \mathfrak{F}(Y); y = f(x, A) \quad (4.3)$$

onde:  $\mathfrak{J}(y)$  = conjunto de todos os subconjuntos difusos sobre Y

O conjunto difuso Y apresenta a seguinte função de pertinência:

$$\mu_Y(y) = \begin{cases} \text{Max } \mu_A(a), \{a \mid y = f(x, a)\} \neq \emptyset \\ 0, & \text{caso contrário} \end{cases} \quad (4.4)$$

Considerando-se um sistema linear possibilístico da forma (TANAKA, 1989):

$$Y = A_1x_1 + A_2x_2 + \dots + A_nx_n = Ax \quad (4.5)$$

onde:  $A_j = (\alpha_j; c_j)$

$x_i$  = número real

$Y = (\alpha x; cx)$

O problema de regressão possibilística consiste em determinar os valores de  $\alpha$  e  $c$  de modo que, para um nível  $h$ , os valores observados estejam contidos nos valores estimados (problema de minimizar a somatória dos "spreads"); ocorrer uma intersecção entre valores observados e estimados, para um nível  $h$  (problema também de minimizar "spreads", porém com restrições diferentes) e, finalmente, determinar valores de  $\alpha$  e  $c$  de modo que, para um nível  $h$ , os valores estimados estejam contidos (problema de maximizar o "spreads" dos valores estimados) nos valores observados.

#### 4.9.2 - Regressões Lineares Possibilísticas

Inicialmente considere-se a definição de inclusão de números

difusos: A inclusão de números difusos com um grau  $0 \leq h \leq 1$ , representada por  $A_1 \supseteq_h A_2$ , é definido através de  $[A_1]_h \supseteq [A_2]_h$ , onde  $[A]_h$  é o conjunto nível  $h$  de um número difuso  $A$  (TANAKA, 1989).

Sejam  $A_i = (\alpha_i, c_i)$  dois números difusos. A FIGURA 4.5 mostra a inclusão  $[A_1]_h \supseteq [A_2]_h$ .

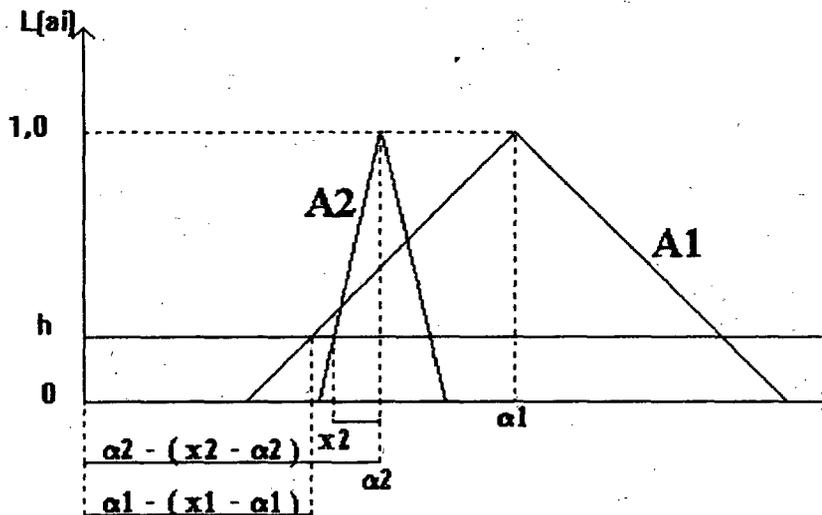


FIGURA 4.5 - Inclusão  $[A_1]_h \supseteq [A_2]_h$

O ponto  $x_2$  tem coordenadas  $(\alpha_2 - (x_2 - \alpha_2), 0)$  e o ponto  $x_1 = (\alpha_1 - (x_1 - \alpha_1), 0)$ . A inclusão  $[A_1]_h \supseteq [A_2]_h$  é, então, equivalente a:

$$\alpha_1 - (x_1 - \alpha_1) \leq \alpha_2 - (x_2 - \alpha_2) \quad (4.6)$$

$$\alpha_1 - \frac{(x_1 - \alpha_1)c_1}{c_1} \leq \alpha_2 - \frac{(x_2 - \alpha_2)c_2}{c_2}$$

$$\alpha_1 \leq \alpha_2 + |L^{-1}(h)| (c_1 - c_2) \quad (4.7)$$

De forma similar pode-se demonstrar ainda que:

$$\alpha_1 \geq \alpha_2 - \left| L^{-1}(h) \right| (c_1 - c_2) \quad (4.8)$$

Se  $[A_1]_h \supset [A_2]_h$  para  $h \in (0,1)$ , tem-se que:

$[A_1]_{h'} \supset [A_2]_{h'}$ ,  $\forall h' \leq h$ , desde que  $\hat{c}_1 \geq c_2$  e  $\left| L^{-1}(h') \right| \geq \left| L^{-1}(h) \right|$ . A inclusão de um número difuso com nível  $h$  acontece para  $h' \leq h$ .

Considere-se um sistema possibilístico  $Y_i = A x$ , onde  $Y_i = (y_i, e_i)_L$  é uma observação da variável de interesse e  $x_i = (x_{i1}, \dots, x_{in})^t$  é um vetor de variáveis explicativas para a  $i$ -ésima amostra. A solução de tal sistema possibilístico, para todas as observações disponíveis, consiste na determinação dos parâmetros difusos de modo que:

$$Y_i \subseteq_h \bar{Y}_i = \bar{A}_1 x_{i1} + \dots + \bar{A}_n x_{in} \quad (4.9)$$

$$Y_i \supseteq_h \underline{Y}_i = \underline{A}_1 x_{i1} + \dots + \underline{A}_n x_{in} \quad (4.10)$$

$$[Y_i]_h \cap [\hat{Y}_i = \hat{A}_1 x_{i1} + \dots + \hat{A}_n x_{in}] \neq 0 \quad (4.11)$$

Em (4.9) os valores observados são cobertos, a um nível  $h$ , pelos números difusos  $\{\bar{Y}_i\}$ . Em (4.10) os valores estimados estão contidos, a um nível  $h$ , nos valores observados  $Y_i$ . A relação (4.11) implica que a intersecção para um dado resultado  $Y_i$  e um número difuso  $\hat{Y}_i$  não é vazia.

A seguir são discutidas as resoluções dos sistemas (4.9), (4.10) e (4.11).

#### 4.9.2.1 - Problema de Minimização

$$\bar{A}_j = (\bar{\alpha}_j, \bar{c}_j) \quad \bar{J}(\bar{c}) = \sum \bar{c} |x_i|$$

Sujeito a:

$$Y_i \subseteq_h \bar{Y}_i = \bar{A}_1 x_{i1} + \bar{A}_2 x_{i2} + \dots + \bar{A}_n x_{in}$$

onde:  $\bar{c} |x_i|$  = dispersão para um resultado difuso estimado  $\bar{Y}_i$

O problema é, então, obter a menor dispersão de  $\sum \bar{Y}_i$  tal que

$$\bar{Y}_i \supseteq_h Y_i$$

$$\text{Min}_{\bar{\alpha}, \bar{c}} \sum \bar{c} |x_i| = \bar{J}(\bar{c}) \quad (4.12)$$

Sujeito a:

$$y_i + |L^{-1}(h)| e_i \leq \bar{\alpha} x_i + |L^{-1}(h)| \bar{c} |x_i|$$

$$y_i - |L^{-1}(h)| e_i \geq \bar{\alpha} x_i - |L^{-1}(h)| \bar{c} |x_i|$$

$$\bar{c} \geq 0, i=1, \dots, N$$

#### 4.9.2.2 - Problema de Maximização

$$\bar{A}_j = (\bar{\alpha}_j, \bar{c}_j) \quad \bar{J}(\bar{c}) = \sum \bar{c} |x_i|$$

Sujeito a:

$$Y_i \supseteq_h Y_i = A_1 x_{i1} + A_2 x_{i2} + \dots + A_n x_{in}$$

Neste caso, o problema é o de obter a maior dispersão para  $\sum Y_i$  tal que  $Y_i \subseteq_h Y_i$ .

$$\underset{\underline{\alpha}, \underline{c}}{\text{Max}} \sum \underline{c} |x_i| = \underline{J}(\underline{c}) \quad (4.13)$$

Sujeito a:

$$y_i + |L^{-1}(h)| e_i \geq \underline{\alpha} x_i + |L^{-1}(h)| \underline{c} |x_i|$$

$$y_i - |L^{-1}(h)| e_i \leq \underline{\alpha} x_i - |L^{-1}(h)| \underline{c} |x_i|$$

$$\underline{c} \geq 0, \quad i = 1, \dots, N$$

#### 4.9.2.3 - Problema de Conjunção

$$\hat{A}_j = \underset{(\hat{\alpha}_j, \hat{c}_j)}{\text{Min}} \hat{J}(\hat{c}) = \sum \hat{c} |x_i| = \underset{\hat{\alpha}, \hat{c}}{\text{Min}} \sum \hat{c} |x_i| = \hat{J}(\hat{c}) \quad (4.14)$$

Sujeito a

$$y_i + |L^{-1}(h)| e_i \geq \hat{\alpha} x_i - |L^{-1}(h)| \hat{c} |x_i|$$

$$y_i - |L^{-1}(h)| e_i \leq \hat{\alpha} x_i + |L^{-1}(h)| \hat{c} |x_i|$$

$$\hat{c} \geq 0, \quad i = 1, \dots, N$$

O problema pode ser interpretado como sendo obter a menor dispersão de  $\sum \hat{Y}_i$ , tal que  $[Y_i]_h \cap [\hat{Y}_i]_h \neq \emptyset$ .

Seja  $(Y_i^\circ, x_i^\circ)$ ,  $i = 1, \dots, N$  satisfazendo um sistema linear possibilístico:

$$Y_i^\circ = A_1^\circ x_{i1}^\circ + A_2^\circ x_{i2}^\circ + \dots + A_n^\circ x_{in}^\circ \quad (4.15)$$

**Teorema:** Se  $(Y_i^\circ, x_i^\circ)$ ,  $i = 1, \dots, N$  satisfaz (4.15), obtém-se as seguintes igualdades através da resolução dos problemas Min e Max:

$$A^\circ = \bar{A} = \underline{A} \quad \text{e} \quad Y^\circ = \bar{Y} = \underline{Y} \quad (4.16)$$

**Prova:**  $(Y_i^\circ, x_i^\circ)$  satisfaz (4.12), então:  $e_i = c^\circ \mid x_i^\circ \mid$

$$\left. \begin{aligned} y_i + \mid L^{-1}(h) \mid e_i &\leq \bar{\alpha} x_i + \mid L^{-1}(h) \mid \bar{c} \mid x_i \mid \\ y_i^\circ + \mid L^{-1}(h) \mid c^\circ \mid x_i^\circ \mid &\leq \bar{\alpha} x_i^\circ + \mid L^{-1}(h) \mid \bar{c} \mid x_i^\circ \mid \\ y_i^\circ &\leq \bar{\alpha} x_i^\circ + \mid L^{-1}(h) \mid \bar{c} \mid x_i^\circ \mid - \mid L^{-1}(h) \mid c^\circ \mid x_i^\circ \mid \\ y_i^\circ &\leq \bar{\alpha} x_i^\circ + \mid L^{-1}(h) \mid (\bar{c} - c^\circ) \mid x_i^\circ \mid \\ y_i^\circ &\geq \bar{\alpha} x_i^\circ - \mid L^{-1}(h) \mid (c^\circ - \bar{c}) \mid x_i^\circ \mid \end{aligned} \right\} \Rightarrow (\alpha^\circ, c^\circ) \text{ é uma solução admissível} \quad (4.17)$$

Admitindo  $\bar{c}'$  tal que:

$$\Sigma \bar{c}' \mid x_i^\circ \mid < \Sigma c^\circ \mid x_i^\circ \mid \quad (4.18)$$

para algum  $i$   $\bar{c}' \mid x_i^\circ \mid < c^\circ \mid x_i^\circ \mid$  (4.19), o que é uma contradição com (4.17).

Assim  $\bar{c} = c^\circ$  é uma solução mínima.

Substituindo  $\bar{c} = c^\circ$  em (4.17), obtém-se:

$$y_i^\circ = \bar{\alpha} x_i^\circ, \quad i = 1, \dots, N \quad (4.20)$$

Então, tem-se,  $\bar{\alpha} = \alpha^\circ$ , desde que é assumido genericamente que existem  $x$  valores independentes em  $\{x_1, \dots, x_N\}$ , onde  $N > n$ .

**Teorema:** Se  $(y_i^\circ, x_i^\circ)$ ,  $i=1, \dots, N$  satisfaz (4.15), então,  $\hat{J}(\hat{c}) = 0$  para o problema de conjunção.

**Prova:**  $\hat{c} = 0$  em (4.14), a condição de restrição é:

$$\begin{aligned} y_i^\circ + \left| L^{-1}(h) \right| e_i &\geq \hat{\alpha} x_i^\circ \\ y_i^\circ - \left| L^{-1}(h) \right| e_i &\leq \hat{\alpha} x_i^\circ \end{aligned} \quad (4.21)$$

$\hat{c} = 0$  é uma solução admissível e ótima,  $\hat{J}(\hat{c}) = 0$ .  $\hat{\alpha} = \alpha_i^\circ$  é uma das soluções mas não é a única.

**Teorema:** Um dado conjunto  $(y_i^\circ, x_i^\circ)$ ,  $i=1, \dots, N$  satisfazendo um sistema linear convencional  $y_i^\circ = a^\circ x_i^\circ$ , tem-se:

$$\begin{aligned} y_i &= \bar{Y}_i = \underline{Y}_i = \hat{Y}_i, \quad a^\circ = \bar{A} = \underline{A} = \hat{A} \\ \bar{J}(\bar{c}) &= \underline{J}(\underline{c}) = \hat{J}(\hat{c}) = 0 \end{aligned} \quad (4.22)$$

**Prova:** Substituindo  $c=0$  nos problemas de Minimização, Maximização e Conjunção, obtém-se  $y_i^\circ = a^\circ x_i^\circ$ . Então obtém-se (4.22) através da solução dos problemas Min, Max e Conj.

Em geral, o conjunto de dados não satisfaz um sistema linear possibilístico (4.15). Deve-se, então, discutir a existência de uma solução e propriedades mútuas para os três problemas.

**Teorema:** Existem soluções ótimas  $\bar{A} = (\bar{\alpha}, \bar{c})_L$  e  $\hat{A} = (\hat{\alpha}, \hat{c})_L$  para

todo  $h \in [0, 1)$  para os problemas Min e Conj, porém não é assegurada que exista uma solução ótima para o problema Max.

**Prova:** Assumindo  $\bar{\alpha} = 0$ , as restrições de (4.12) conduzem a:

$$- \left| L^{-1}(h) \right| (\bar{c} \mid x_i \mid - e_i) \leq y_i \leq \left| L^{-1}(h) \right| (\bar{c} \mid x_i \mid - e_i) \quad (4.23)$$

Se um número suficientemente grande para todo  $\bar{c}_i$ ,  $\bar{A} = (0, \bar{c}_i)_L$  é uma solução viável porque  $y_i$  é finito. Então, existe uma solução ótima para o problema de Minimização. O problema de Conjunção pode ser demonstrado da mesma maneira. Reciprocamente, mesmo quando toma-se  $\underline{c}_i = 0$  não assegura-se que (4.13) possua solução admissível.

**Teorema:** Existe uma solução ótima no problema de Max se, e somente se,  $\hat{J}(\hat{c}) = 0$  no problema de Conjunção.

**Prova:** Assumindo que exista uma solução ótima no problema de Maximização. Fazendo  $\hat{c} = 0$  em (4.14), obtém-se:

$$\begin{aligned} \hat{\alpha} x_i &\leq y_i + \left| L^{-1}(h) \right| e_i \\ \hat{\alpha} x_i &\geq y_i - \left| L^{-1}(h) \right| e_i \end{aligned} \quad (4.24)$$

Por hipótese, existe um conjunto admissível para (4.13), então existe um conjunto admissível para (4.24). Então,  $\hat{c} = 0$  é uma solução ótima no problema de Conjunção. Reciprocamente, assuma que  $\hat{J}(\hat{c}) = 0$ . Segue-se de (4.24) que existe um conjunto admissível para (4.13) com  $\underline{c} \geq 0$ . Existe, então, uma solução ótima para o problema Max.

**Teorema:**  $\bar{J}(\bar{c}) \geq \hat{J}(\hat{c}) \quad (4.25)$

**Prova:** Denotando os conjuntos admissíveis (4.12) e (4.14) como, respectivamente,  $\bar{D}$  e  $\hat{D}$ . De (4.12) e (4.14), obtém-se:

$$\begin{aligned} y_i - |L^{-1}(h)| e_i &\leq y_i + |L^{-1}(h)| e_i \\ &\leq \alpha x_i + |L^{-1}(h)| c |x_i| \\ y_i + |L^{-1}(h)| e_i &\geq y_i - |L^{-1}(h)| e_i \\ &\geq \hat{\alpha} x_i - |L^{-1}(h)| c |x_i| \end{aligned} \quad (4.26)$$

Desde que,  $\bar{D} \subset \hat{D}$  por (4.26),  $\bar{J}(\bar{c}) \geq \hat{J}(\hat{c})$

**Teorema:** Se os resultados difusos  $Y_i = (y_i, e_i)$ ,  $i=1, \dots, N$  tornam-se "crisp", isto é,  $e_i = 0$ ,  $i=1, \dots, N$  tem-se:

$$\bar{Y}_i = \hat{Y}_i \quad (4.27)$$

**Prova:** Substituindo  $e_i = 0$  em (4.12) e (4.14), as restrições igualam-se, então,  $\bar{Y}_i = \hat{Y}_i$ , desde que  $\bar{D} = \hat{D}$ .

Para grandes valores de  $h$  tem-se intervalos com altas possibilidades: a análise é, portanto, otimista.

Para pequenos valores de  $h$  os intervalos apresentarão baixas possibilidades: análise pessimista.

Embora, existam soluções ótimas para todo  $h \in [0,1)$ , para os problemas Max e Conj, obtém-se o seguinte teorema com relação ao problema Max.

**Teorema:** Se existe uma solução ótima para  $h'$  no problema Max,

existe uma solução ótima para  $0 \leq h \leq h'$ .

**Prova:** Seja  $(\underline{\alpha}', \underline{c}')$ <sub>L</sub> uma solução ótima para  $h'$ . Desde que,  $h' \geq h$ , assegura-se que  $\left| L^{-1}(h') \right| \leq \left| L^{-1}(h) \right|$ . A condição de restrição para  $h$  no problema de Max é rescrito como:

$$\begin{aligned} y_i + \left| L^{-1}(h) \right| e_i &\geq y_i + \left| L^{-1}(h) \right| e_i \geq \underline{\alpha}' x_i + \left| L^{-1}(h') \right| \underline{c}' x_i \\ &= \underline{\alpha}' x_i + \left| L^{-1}(h') \right| \underline{c}' x_i \left| L^{-1}(h') \right| / \left| L^{-1}(h) \right| \\ y_i - \left| L^{-1}(h) \right| e_i &\leq y_i - \left| L^{-1}(h') \right| e_i \leq \underline{\alpha}' x_i - \left| L^{-1}(h') \right| \underline{c}' x_i \\ &= \underline{\alpha}' x_i - \left| L^{-1}(h) \right| \underline{c}' x_i \left| L^{-1}(h') \right| / \left| L^{-1}(h) \right| \end{aligned} \quad (4.28)$$

A solução  $(\underline{\alpha}', \underline{c}' \left| L^{-1}(h') \right| / \left| L^{-1}(h) \right|)$ <sub>L</sub> satisfaz (4.28). Desde que existe uma solução admissível para  $h$ , existe uma solução ótima para  $h$  no problema Max.

Denotando uma solução ótima e um índice de performance para  $h$  como sendo  $(\alpha_h, c_h)$ <sub>L</sub> e  $J(h)$ , respectivamente, então:

**Teorema:** Tem-se para  $h' \leq h$ :

$$\bar{J}(h') \leq \bar{J}(h) \quad (4.29)$$

$$\underline{J}(h') \geq \underline{J}(h) \quad (4.30)$$

$$\hat{J}(h') \leq \hat{J}(h) \quad (4.31)$$

**Prova:**(4.29): condições de restrições para  $h$ :

$$\left. \begin{aligned} y_i &\leq \bar{\alpha}_h x_i + \left| L^{-1}(h) \right| (\bar{c}_h |x_i| - e_i) \\ y_i &\geq \bar{\alpha}_h x_i - \left| L^{-1}(h) \right| (\bar{c}_h |x_i| - e_i) \end{aligned} \right\} \quad (4.32)$$

Segue-se de  $\bar{Y}_i \supseteq_h Y_i$  que  $\bar{c}_h |x_i| - e_i \geq 0$ . Então:

$$\left. \begin{aligned} y_i &\leq \bar{\alpha}_h x_i + \left| L^{-1}(h') \right| (\bar{c}_h |x_i| - e_i) \\ y_i &\geq \bar{\alpha}_h x_i - \left| L^{-1}(h') \right| (\bar{c}_h |x_i| - e_i) \end{aligned} \right\} \quad (4.33)$$

Portanto  $(\bar{\alpha}_h, \bar{c}_h)_L \in \bar{D}_h$  conduz a  $\bar{J}(h') \leq \bar{J}(h)$ . (4.30) condições de restrições para h:

$$\left. \begin{aligned} y_i &\geq \underline{\alpha}_h x_i + \left| L^{-1}(h) \right| (\underline{c}_h |x_i| - e_i) \\ y_i &\leq \underline{\alpha}_h x_i - \left| L^{-1}(h) \right| (\underline{c}_h |x_i| - e_i) \end{aligned} \right\} \quad (4.34)$$

Segue-se de  $\underline{Y}_i \subseteq Y_i$  que  $\underline{c}_h |x_i| - e_i \leq 0$ . Então:

$$\left. \begin{aligned} y_i &\geq \underline{\alpha}_h x_i + \left| L^{-1}(h') \right| (\underline{c}_h |x_i| - e_i) \\ y_i &\leq \underline{\alpha}_h x_i - \left| L^{-1}(h') \right| (\underline{c}_h |x_i| - e_i) \end{aligned} \right\} \quad (4.35)$$

Portanto  $(\underline{\alpha}_h, \underline{c}_h)_L \in \underline{D}_h$  conduz a  $\underline{J}(h') \geq \underline{J}(h)$ . (4.31) condições de restrição para h:

$$\left. \begin{aligned} y_i &\geq \hat{\alpha}_h x_i - \left| L^{-1}(h) \right| (\hat{c}_h |x_i| - e_i) \\ y_i &\leq \hat{\alpha}_h x_i - \left| L^{-1}(h) \right| (\hat{c}_h |x_i| - e_i) \end{aligned} \right\} \quad (4.36)$$

Segue-se de  $\left[ Y_i \right]_h \cap \left[ \hat{Y}_i = \hat{A}_1 x_{i1} + \dots + \hat{A}_n x_{in} \right]_h$  que  $\hat{c}_h = 0$ .

Então:

$$\left. \begin{aligned} y_i &\geq \hat{\alpha}_h x_i - \left| L^{-1}(h') \right| (\hat{c}_h |x_i| - e_i) \\ y_i &\leq \hat{\alpha}_h x_i - \left| L^{-1}(h') \right| (\hat{c}_h |x_i| - e_i) \end{aligned} \right\} \quad (4.37)$$

Portanto  $(\hat{\alpha}_h, \hat{c}_h)_L \in \hat{D}_{h'}$  conduz a  $\hat{J}(h') \leq \hat{J}(h)$ .

**Teorema:** Se existe uma solução ótima no problema de Max:

$$\bar{J}(\bar{c}) \geq \underline{J}(\underline{c}), \quad \forall h \quad (4.38)$$

**Prova:** Se existe uma solução ótima para os problemas de Max e Min, as seguintes condições devem ocorrer para todo i:

$$\bar{c} |x_i| \geq e_i, \quad \underline{c} |x_i| \leq e_i \quad (4.39)$$

Então:

$$\bar{J}(\bar{c}) = \text{Min} \sum \bar{c} |x_i| \geq \text{Max} \sum \underline{c} |x_i| = \underline{J}(\underline{c}) \quad (4.40)$$

o que conduz ao teorema.

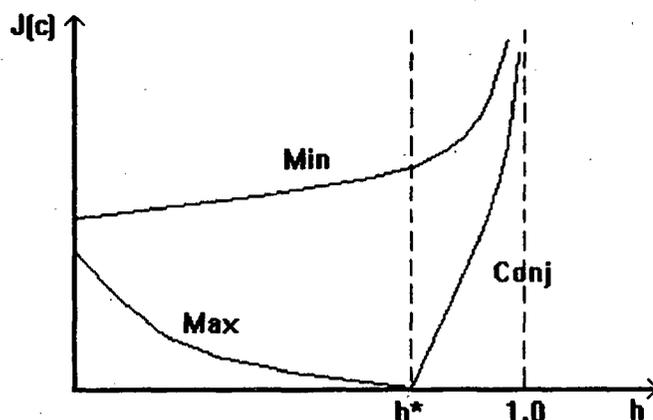


FIGURA 4.6 - Relação entre o Nível  $h$  e o Índice de Performance  $J(c)$

Se os resultados difusos forem reduzidos a resultados não difusos, o problema de Conjunção é equivalente ao problema de Minimização, embora o problema de Maximização possa não existir.

#### 4.10 - Teoria de Conjuntos Difusos e Sistemas Educacionais: uma Proposta

Uma característica da lógica difusa, a qual é de particular importância para o tratamento de incertezas é a perspectiva de se dispor de um ferramental sistemático para o tratamento com quantificadores difusos, como por exemplo, a maioria, muitos, poucos, quase todos, cerca de 80% etc.

Um sistema educacional é considerado um sistema complexo. Portanto, um modelo para analisá-lo envolve múltiplos e diferentes graus de dificuldade. Envolve também, explicitar sob que prismas se quer analisar esse sistema. Quando o objetivo primordial é o de enfocar medidas quantitativas sobre esse sistema, uma dificuldade que surge é a forma de parametrização dessas medidas.

Vários modelos para análise que, empregam medidas quantitativas,

são disponíveis para análise de sistemas educacionais. Alguns tipos de modelos, contudo, incorporam apenas o tratamento de incertezas do Tipo Um, conforme proposto por GUPTA (1991). Esse tipo de incerteza trata com informações ou fenômenos que surgem a partir do comportamento aleatório de um sistema. Existe uma teoria estatística bem estabelecida e adequada para caracterizar tais fenômenos aleatórios.

Um problema encontrado, e que requer esforços de pesquisa, é o de modelar incertezas derivadas do pensamento humano, processos de raciocínio, processos cognitivos, processos de percepção ou, de uma forma mais ampla, informação cognitiva. A esse tipo de incerteza, denominada de incerteza Tipo Dois por Gupta, os métodos tradicionais de tratamento não podem ser aplicados.

O número de alunos que ingressarão no sistema educacional nos próximos  $n$  anos pode ser previsto a partir de um modelo comum de previsão. Já uma informação sobre o estado de conservação de uma escola, certamente irá diferir, em termos cognitivos, entre um especialista em educação e um engenheiro de sistemas. Os modelos estatísticos tradicionais não permitem uma explicação cognitiva suficiente (ZINSER e HENNEMAN, 1988). Conjuntos difusos podem ser utilizados para capturar imprecisões e características nebulosas decorrentes de conceitos naturais (PENG, KANDEL e WANG, 1991).

Um problema típico de um sistema educacional é o da evasão escolar. Uma resposta de interesse para os responsáveis pelo sistema educacional pode ser a propensão que um determinado aluno apresenta em abandonar o sistema. Essa resposta pode ser investigada, em termos de juízos de valores emitidos por professores, psicólogos e pedagogos que tenham um contato mais efetivo com o aluno. Através das respostas obtidas, pode-se ajustar uma distribuição de possibilidade que descreva a propensão que o indivíduo apresenta para abandonar o sistema. A variável "propensão à evasão" pode ser explicada através de características individuais de cada aluno, por exemplo, situação sócio-econômica dos responsáveis, escolaridade dos pais, rendimento

escolar acumulado do aluno, distância casa-escola etc. Nesses casos, a variável-resposta pode ser avaliada em termos de uma função possibilística da forma:

$$pev_{ij} = A_{0j} + A_{1j}SSE_{ij} + A_{2j}ESCP_{ij} + A_{3j}RAA_{ij} + A_{4j}DCE_{ij}$$

onde:  $pev_{ij}$  = Propensão à evasão para o aluno  $i$  na escola  $j$

$SSE_{ij}$  = Situação sócio-econômica dos responsáveis do aluno  $i$  na escola  $j$

$ESCP_{ij}$  = Escolaridade dos pais do aluno  $i$  na escola  $j$

$RAA_{ij}$  = Rendimento acumulado do aluno  $i$  na escola  $j$

$DCE_{ij}$  = Distância casa-escola, percorrida pelo aluno  $i$  na escola  $j$

$A_j, j = 1, \dots, n$  = números difusos que descrevem a estrutura difusa, através da qual as variáveis exógenas atuam sobre a variável-resposta.

O modelo descreve, para uma escola particular, como características individuais de alunos podem contribuir, na avaliação da propensão ao abandono.

Para diferentes escolas, as características dessas escolas podem contribuir de forma diversa na avaliação da propensão à evasão. Por exemplo, o nível de preparação de professores e educadores, a qualidade atribuída pelos alunos à merenda escolar, a qualidade das instalações escolares, etc apresentam efeitos diferenciados quanto a avaliação da propensão à evasão de um aluno. É razoável admitir-se que quando um juízo sobre a propensão à evasão é emitido, tal juízo baseia-se em características individuais do aluno. Assim, a contribuição de cada uma das características individuais dos alunos consideradas agrega uma informação derivada do efeito contextual da escola:

$$A_{ij} = B_0 + B_1NPE_l + B_2QME_l + B_3QE_l \quad l = 1, \dots, n.$$

onde:  $NPE_l$  = nível de preparação de educadores na escola  $l$

$QME_l$  = qualidade atribuída pelos alunos à merenda escolar na escola  $l$

$QE_1$  = qualidade das instalações da escola I

A partir da resolução dos modelos propostos, é possível identificar:

- **Dentro de uma escola:** quais características individuais são consideradas relevantes na propensão à evasão e de que forma essas características atuam; dada uma escola, a partir da análise das características individuais dos alunos, propor medidas de combate à evasão escolar; compor classes de alunos com características homogêneas, de modo a realizar um trabalho preventivo no combate à evasão; etc.

**Entre escolas:** estabelecer que características são desejáveis nos estabelecimentos escolares, de modo a prevenir a evasão escolar; identificar como investimentos realizados sobre características da escola podem contribuir no combate à evasão; identificar escolas que potencialmente podem influir na propensão de um aluno com determinadas características abandonar o sistema educacional; etc.

## **5. MODELOS PARA CARACTERIZAÇÃO DO DESEMPENHO DAS ESCOLAS**

### **5.1 - Introdução**

Neste capítulo são propostas duas abordagens para o problema da caracterização do desempenho de sistemas educacionais. Uma baseada na construção de modelos estatísticos de análise e outra fundamentada nos princípios da lógica difusa.

Para a construção dos modelos que se utilizam da teoria dos conjuntos difusos é, inicialmente, derivado um modelo teórico para regressões possibilísticas restritas a um intervalo.

Para ambas abordagens, após caracterizado o problema do desempenho, os modelos são estabelecidos e, posteriormente, procedimentos de classificação das escolas de acordo com os resultados obtidos são analisados.

### **5.2 - O Problema do Desempenho dos Sistemas Educacionais**

O desempenho de Sistemas Educacionais é abordado através da modelagem das seguintes questões:

- aproveitamento escolar dos alunos;
- caracterização desses alunos e sua influência no aproveitamento escolar;

- contexto em que se inserem os estabelecimentos escolares e características desses estabelecimentos que possam ser consideradas relevantes para explicar o rendimento escolar.

Uma vez construídos tais modelos, é possível, obterem-se respostas às seguintes questões de interesse de planejadores e responsáveis pelo Sistema Educacional:

- observado um aluno qualquer, a partir do exame de suas características individuais, qual a possibilidade de vir a ser aprovado?
- observado o mesmo aluno, qual a possibilidade de vir a ser aprovado caso se matricule em uma escola específica?
- para um determinado aluno, e para um elenco de escolas, em qual das escolas o aluno apresenta maiores possibilidades de aprovação?
- admitida uma política de âmbito maior, onde as escolas sejam ordenadas de acordo com algum critério, no qual se considerem taxas de aprovação como um dos fatores intervenientes, como se enquadram as escolas dentro dessa política?

O problema é modelado através de técnicas estatísticas multivariadas e de regressão possibilística.

Como visto anteriormente, a formulação de modelos de análise de desempenho de Sistemas Educacionais considera fatores relacionados com aluno e com escolas. Para tratar com algum realismo o problema em questão, as variáveis que serão incorporadas ao modelo são variáveis disponibilizadas na base de dados do Gabinete de Estudos e Planeamento (GEP) do Ministério da Educação de Portugal, no período 1989/1990. Essas variáveis compreendem:

- sexo;

- idade (em anos completos);
- situação sócio-econômica dos responsáveis (dezesesseis categorias);
- ano escolar (atual e anterior);
- número de repetências ;
- nacionalidade do aluno (cinco categorias).

Para utilização dos dados coletados sobre os alunos, ajustes foram necessários. Por exemplo, a variável situação sócio-econômica dos responsáveis apresentava, originalmente, dezesseis categorias relativas à ocupação principal do responsável pelo aluno. De modo a obter-se uma variável que correspondesse a um grupo social, quinze categorias originais foram reduzidas, mediante análise de agrupamento, a cinco grupos sociais. Como variáveis "proxis", foram utilizadas informações censitárias (Informação do Censo de Portugal -1980, dados mais recentes não estavam disponíveis) sobre valores de aluguéis pagos segundo as categorias sócio-econômicas. A décima sexta categoria, "Desempregados", não foi considerada por dois motivos: não ser significativo o número observado e poder constituir uma situação transitória, de forma que sua inclusão poderia acarretar em algum tipo de desvio nos resultados. A TABELA 5.1 apresenta as categorias originais e o agrupamento efetuado.

O agrupamento realizado permitiu o estabelecimento de uma "métrica de grupos sócio-econômicos" para as categorias. O objetivo dessa métrica é o de estabelecer, através dos valores médios dos aluguéis pagos por grupo identificado, uma relação que determinasse o grupo social ao qual pertencem os alunos. A hipótese adotada para o estabelecimento da métrica foi a de que os valores de aluguéis são pagos consoante o rendimento do agregado familiar. Os seguintes pesos foram, então, adotados (a partir da menor média de aluguel pago pelo grupo):

- Grupo 1 - 0,35;
- Grupo 2 - 0,69;

- Grupo 3 - 1,08;
- Grupo 4 - 1.30;
- Grupo 5 - 1.59.

As variáveis Idade do Aluno, Ano Escolar (atual e anterior) e Número de Repetências, para este estudo, foram agregadas ao nível da escola. Assim, consoante os fatores idade, ano escolar e repetências, os alunos são agrupados em três categorias: "1 - Normal"; "2 - Defasado" e "3 - Muito Defasado".

No cálculo desse índice, os valores indicativos do ano escolar foram considerados, neste trabalho, como sendo igual a idade mínima necessária para a freqüência do referido ano. Assim, definiu-se o seguinte critério para o valor do índice:

- 1 se  $(\text{Idade} - \text{Ano Escolar} - \text{N}^\circ \text{ de Repetências}) \leq 1$ ;
- 2 se  $(\text{Idade} - \text{Ano Escolar} - \text{N}^\circ \text{ de Repetências}) = 2$ ;
- 3 caso contrário.

O aluno classificado como "1 - Normal" encontra-se matriculado no ano escolar que corresponde a sua idade cronológica e não apresenta (como esperado) nenhuma repetência naquele ano. A nova variável penaliza tanto repetências ocorridas como, também, desconexão entre ano escolar e idade do aluno.

TABELA 5.1 - Situação Sócio-Econômica e Agrupamento Realizado

<b>Situação Sócio-Econômica do Encarregado de Educação (Original)</b>	<b>Agrupamento Realizado</b>
Patrões agrícolas	Grupo 1
Agricultores que trabalham por conta própria	Grupo 1
Outros trabalhadores agrícolas	Grupo 2
Patrões e trabalhadores por conta própria das profissões liberais e relacionadas	Grupo 3
Patrões da indústria, construção civil, comércio, transportes e serviços	Grupo 4
Trabalhadores por conta própria, da indústria, construção civil, comércio, transportes e serviços	Grupo 1
Diretores, membros de corpos legislativos e quadros superiores dirigentes da Administração Pública e privada	Grupo 5
Empregados com profissões liberais e relacionadas	Grupo 3
Pessoal de enquadramento dos trabalhadores do comércio, serviços e Administração Pública	Grupo 4
Outro pessoal do comércio, administrativo, dos serviços e da Administração Pública	Grupo 1
Encarregados e capatazes (empregados) dos trabalhadores manuais	Grupo 1
Operários qualificados e semiquilificados	Grupo 1
Operários não qualificados	Grupo 2
Pessoal das Forças Armadas (de carreira e serviço militar obrigatório)	Grupo 1
Outras pessoas ativas não classificadas anteriormente	Grupo 4
Desempregados	

Com as duas novas variáveis estabelecidas, obtém-se trinta categorias de alunos (2 para a variável sexo x 5 para grupo sócio-econômico x 3 para o índice calculado em função da idade, ano escolar e número de repetências). Um percentual de aprovação, para cada combinação de categorias

é calculado e esse percentual de aprovação é utilizado para expressar desempenho de alunos de acordo com suas características básicas.

Informações sobre a nacionalidade dos alunos foram desconsideradas, nesse estudo, por apresentarem variações muito pequenas.

Os estabelecimentos de ensino são avaliados segundo os seguintes critérios: localização, estado de conservação, adequação das salas e taxa de ocupação. Para o cálculo desses critérios são adotadas as definições (conforme "Relatório de Avaliação do Parque Escolar da Região do Algarve", Seminário sobre Carta Escolar, Vilamoura, Nov. 1989):

- **Localização:** Avaliar os diversos aspectos ligados à localização de uma instalação: isolamento, aspectos ambientais e acesso. A fórmula de cálculo desse índice consiste de:

Isolada	Aspectos Ambientais	Acesso
Sim - 0	Sim - 0	A pé - 0
Não - 3	Não - 3	A pé e auto - 2
		Todos - 3

Esse índice representa a média ponderada desses fatores e é calculado através de:

$$\text{média} = \frac{\text{Isolada} + 2 \times \text{Aspectos Ambientais} + 0.5 \times \text{Acesso}}{3}$$

- **Estado de Conservação:** Avaliar o estado de conservação dos edifícios pertencentes a instalação. O seu cálculo é função das seguintes variáveis: área total das instalações e estado de conservação - classificado em ruína/irrecuperável, degradado, razoável e bom.
- **Adequação das Salas:** Avaliar a adequação das salas de aula de uma instalação e classificar os espaços de ensino utilizados. Essa classificação é

realizada levando em consideração os diversos tipos de espaços de ensino, como por exemplo, salas de aula normais, salas de aula de ciências, salas de meios, laboratórios, oficinas, anfiteatros e ginásios. O índice é avaliado considerando-se inadequação desses espaços em relação a fatores como área, ventilação, iluminação, etc.

- **Taxa de Ocupação:** (T02) Avaliar a qualidade do estabelecimento quanto à taxa de ocupação. É função do número de alunos matriculados sobre a capacidade ou número de lugares de aluno. Esse último termo vem da relação existente entre a área construída pela área normativa para as atividades do aluno.

A TABELA 5.2 descreve as variáveis que serão utilizadas para avaliar o desempenho do sistema educacional. A partir da definição dessas variáveis, o problema da modelagem do desempenho das escolas é resolvido pela utilização de modelos estatísticos lineares multivariados. A classe de modelos a ser adotada é a classe dos modelos logísticos.

TABELA 5.2 - Variáveis Utilizadas para Desenvolvimento dos Modelos

Variável	Notação	Descrição
Aprovação	$YP_{ij}$	Percentual de Aprovação de alunos na categoria i, para a escola j.
Sexo	$SEX_{ij}$	Sexo do aluno (1 - Masculino; 2 - Feminino)
Sit. Sócio Econômica	$SSE_{ij}$	Categoria de Situação sócio econômica
Índice Classificação	$INDANO_{ij}$	Índice de classificação do aluno
Localização	$LOC_i$	Localização dos Estabelecimentos Escolares
Adequação	$ADS_i$	Adequação das salas de aula
Conservação	$EC_i$	Estado de conservação dos estabelecimentos
Ocupação	$To2_i$	Taxa de ocupação dos estabelecimentos

### 5.3 - Modelagem Estatística Multivariada - Modelos Logísticos

Modelos para escolha qualitativa são modelos de regressão linear construídos para o tratamento de situações, nas quais a variável dependente está associada a duas ou mais escolhas qualitativas. Tais modelos encontram grande aplicação prática (PINDYCK e RUBINFELD, 1981; PATERSON, 1993).

Uma parte considerável dos modelos para escolha qualitativa pode restringir-se aos casos em que a variável dependente assume apenas dois valores:  $Y = 1$  ou  $Y = 0$ . A variável dependente assumindo o valor 1 pode ser interpretado como "o resultado desejado foi alcançado" e, naturalmente,  $Y$  igual a 0 "tal não ocorreu". Admitindo-se que  $Y$  assumo o valor 1 com uma probabilidade  $\pi$  ( $P[ Y=1 ] = \pi$ ;  $P[ Y=0 ] = 1 - \pi$ ), torna-se mais interessante modelar a probabilidade  $\pi$ .

Em modelos tradicionais, têm-se:

$$y = X^T \beta + \varepsilon_i \quad (5.1)$$

onde:  $X^T$  = matriz de variáveis explicativas

No caso da variável dependente assumir apenas dois valores implica que a mesma apresenta uma distribuição binomial com parâmetros  $N$  e  $\pi$ .

Considere-se agora uma população subdividida em  $N$  grupos (ou subgrupos) tal como apresentado na TABELA 5.3.

TABELA 5.3 - População subdividida em N Grupos.

	subgrupo			
	1	2	...	N
Sucessos	$Y_1$	$Y_2$	...	$Y_N$
Não Sucessos	$n_1 - Y_1$	$n_2 - Y_2$		$n_N - Y_N$
Total	$n_1$	$n_2$		$n_N$

A intenção é a de encontrar estimativas  $P_i = \frac{Y_i}{n_i}$ , ou seja, a proporção de "sucessos" em cada grupo. O modelo (5.1) pode, então ser rescrito, a menos do termo ruído, na forma:

$$g(\pi_i) = \mathbf{X}_i^T \boldsymbol{\beta} \quad (5.2)$$

A solução trivial para (5.2), obtém-se utilizando a ligação identidade tal que:

$$\pi_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_N X_{iN} \quad (5.3)$$

O modelo pode ser resolvido, por exemplo, pelo Método dos Mínimos Quadrados. Contudo, essa solução tem a desvantagem de permitir valores para  $\pi_i$ , fora do intervalo (0,1).

Para assegurar que  $\pi_i \in (0,1)$ , a solução usual é a de definir  $\pi_i$  como sendo o valor acumulado de uma densidade  $f_S(s)$ . Em aplicações relacionadas com análise de sobrevivência, essa densidade  $f_S(s)$  é denominada de função densidade de tolerância. Utilizando-se a função acumulada para a densidade  $f_S(s)$  como função de ligação, o modelo (5.2) pode ser apresentado na forma:

$$\pi_i = g^{-1}(X^T \beta) = \int_{-\infty}^s f_S(s) ds \quad (5.4)$$

Se a função densidade  $f_S(s)$  for a Normal, obtém-se o modelo

Probit:

$$\pi_i = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-z^2} dz \quad (5.5)$$

$$\pi_i = \Phi(z),$$

onde:  $\Phi(z)$  = Normal acumulada padrão

Como  $\Phi^{-1}(z)$ , por definição, representa a razão entre o desvio da variável aleatória em relação a sua média e a raiz quadrada de sua variância, tem-se:

$$\Phi^{-1}(z) = \frac{x - \mu}{\sigma} = \beta_1 - \beta_2 x \quad (5.6)$$

onde:  $\beta_1 = -\frac{\mu}{\sigma}$  e  $\beta_2 = \frac{1}{\sigma}$

Define-se o Modelo Logit quando:

$$\pi_i = \int_{-\infty}^s f_S(s) ds = \frac{\exp[X^T \beta]}{1 + \exp[X^T \beta]} \quad (5.7)$$

$$g^{-1}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (5.8)$$

Estas transformações  $g^{-1}(\pi_i)$  asseguram que  $P[Y=y_i] = \pi_i$  está no intervalo (0,1). Outros modelos são: modelo log-log; complemento log-log, etc..

### 5.3.1 - Modelos para Análise do Desempenho das Escolas Baseados na Aprovação dos Alunos e nas suas Características Individuais

Com a finalidade de analisar o desempenho das escolas, são definidos dois modelos:

$$\text{MODELO 1: } Y = \beta_0 + \beta_1 \text{Sexo} + \beta_2 \text{SSE} + \beta_3 \text{INDANO} + \varepsilon \quad (5.9)$$

$$\text{MODELO 2: } Y = \beta_0 + \beta_1 \text{SSE} + \beta_2 \text{INDANO} + \varepsilon \quad (5.10)$$

onde:  $Y = \log\left(\frac{Y P_{ij}}{1 - Y P_{ij}}\right)$ , calculado para cada categoria de aluno observada na  $j$ -ésima escola

Esses modelos irão capturar, para uma escola específica  $j$ , os efeitos dos fatores Sexo, SSE e INDANO sobre a aprovação dos alunos.

Com os MODELOS 1 e 2 é possível a identificação das escolas, onde, alunos com determinadas características apresentam probabilidades de aprovação melhores. Através de transformações aplicadas sobre os coeficientes estimados pode-se isolar o efeito de cada uma das características individuais dos alunos na probabilidade de aprovação. O MODELO 1 permite capturar, diretamente, o efeito do fator sexo na aprovação do aluno.

### **5.3.2 Modelo Hierarquizado para Análise do Desempenho através das Características dos Alunos e Características das Escolas**

O MODELO 1 descreve, para uma escola particular, como características individuais de alunos podem contribuir, na sua aprovação.

As diferentes características das escolas, entretanto, podem contribuir de forma diversa na aprovação de seus alunos. Por exemplo, o nível de preparação de professores e educadores, a qualidade atribuída pelos alunos a merenda escolar, a qualidade das instalações escolares, etc., apresentam efeitos diferenciados sobre a probabilidade de aprovação de um aluno. A contribuição de cada uma das características individuais dos alunos consideradas agrega informação derivada do efeito contextual da escola.

Com a introdução de informações sobre as escolas é possível, identificar os efeitos dentro de uma escola e entre escolas.

Para uma escola torna-se possível:

- identificar quais características individuais são consideradas relevantes na probabilidade de aprovação e de que forma essas características atuam;
- dada uma escola, a partir da análise das características individuais dos alunos, propor medidas para melhoria na chance de aprovação de um determinado tipo de aluno;
- compor classes de alunos com características homogêneas, de modo a realizar um trabalho preventivo.

Já, entre as diferentes escolas:

- estabelecer que características são desejáveis nos estabelecimentos escolares, de modo a melhorar o desempenho das escolas;
- identificar como investimentos realizados sobre características da escola podem contribuir na melhoria de seu desempenho;
- identificar escolas que potencialmente podem influir na aprovação de um aluno com determinadas características; etc.

A avaliação do efeito “escola” na contribuição das características individuais dos alunos para sua aprovação ( $\beta_{ji}$ ) é modelada como:

$$\text{MODELO 3: } \beta_{ji} = \gamma_{1j}\text{ADS}_i + \gamma_{2j}\text{EC}_i + \gamma_{3j}\text{LOC}_i + \gamma_{4j}\text{To2}_i + \xi; \text{ com } j = 1, \dots, 4 \\ \text{e } i = 1, \dots, N \quad (5.11)$$

onde:  $\beta_{ji}$  = coeficientes  $\beta_j$  obtidos pela aplicação do modelo (5.9) na escola  $i$ ;

$N$  = Número de escolas

Recombinando-se os modelos (5.9) e (5.11), é possível avaliar os efeitos derivados das características das escolas sobre a aprovação de um determinado tipo de aluno. Tais efeitos podem ser obtidos da seguinte forma:

- **Adequação das Salas de Aula:**

$$(\gamma_{11} + \gamma_{12}\text{IndAno} + \gamma_{13}\text{SSE} + \gamma_{14}\text{Sexo})$$

- **Estado de Conservação:**

$$(\gamma_{21} + \gamma_{22}\text{IndAno} + \gamma_{23}\text{SSE} + \gamma_{24}\text{Sexo})$$

- **Localização:**

$$(\gamma_{31} + \gamma_{32}\text{IndAno} + \gamma_{33}\text{SSE} + \gamma_{34}\text{Sexo})$$

- **Taxa de Ocupação:**

$$(\gamma_{41} + \gamma_{42}\text{IndAno} + \gamma_{43}\text{SSE} + \gamma_{44}\text{Sexo})$$

Os modelos estabelecidos em (5.11), mais os descritos em (5.9) e (5.10) irão capturar, para uma escola específica  $i$ , os efeitos dos fatores Sexo, SSE e INDANO e os efeitos de características próprias da escola na probabilidade de aprovação de um aluno. Em outras palavras, admite-se ser possível descrever, por meio de uma distribuição de probabilidades, um comportamento estocástico para, por exemplo, os efeitos do fator SSE ou do estado de conservação da escola sobre a probabilidade de aprovação de um aluno.

Contudo, existe uma dificuldade bem caracterizada e nítida de estabelecer, com precisão, uma estrutura distribucional adequada para os diversos efeitos considerados. Essa imprecisão e incerteza existente na estrutura dos coeficientes do modelo permite considerar a utilização de regressões possibilísticas.

Em situações de avaliações expressas em percentagens, normalmente o decisor utiliza um tipo de raciocínio difuso: em torno de 80%, ou seja, existe um valor para o qual o decisor credita a possibilidade máxima. Essas percentagens vão declinando até atingir valores que o próprio decisor não crê serem factíveis. Um decisor que emite um juízo valorado da seguinte forma: tal fato apresenta 80% de possibilidade, dificilmente aceitaria, sem surpresas, a não ocorrência do fato. Por outro lado, estaria mais propenso a aceitar a sua ocorrência, sem o mesmo grau de expectativa. Essas situações necessitam de um tratamento particularizado.

Uma característica da lógica difusa, a qual é de particular importância para o tratamento de incertezas, é a perspectiva de se dispor de um ferramental sistemático para o tratamento com quantificadores difusos, como, por exemplo, a maioria, muitos, poucos, quase todos, cerca de 80%, etc.

A análise de regressão difusa permite a construção de modelos em ambientes onde é grande a presença de incerteza, bem como de fenômenos difusos. A relação entre variável dependente e variável independente não é bem

estabelecida, como no caso da regressão tradicional. A suposição básica é a de que os desvios entre valores observados e preditos são devidos a imprecisões e nebulosidades existentes na estrutura do sistema. De forma mais concreta, imprecisões e nebulosidades ocorrem nos coeficientes do modelo.

#### **5.4. Modelagem através de Conjuntos Difusos**

As dificuldades em estabelecer, formas funcionais bem definidas para mensurar os efeitos de cada um dos fatores considerados na aprovação de alunos e, a impossibilidade de, através das técnicas tradicionais, avaliar esses efeitos na presença de quantificadores lingüísticos conduz à modelagem através de conjuntos difusos.

A utilização de regressão possibilística, porém, não é imediata. Considere-se o modelo (5.9). Nesse modelo, a variável que se está utilizando para mensurar o desempenho das escolas tem seu domínio definido no intervalo  $(0,1)$ . Aplicar os modelos teóricos de regressão possibilística descritos no capítulo quatro, pode conduzir a valores para a variável dependente fora do intervalo unitário. Assim, é necessário, em primeiro lugar, definir um modelo teórico de regressão possibilística, no qual os valores da variável dependente estejam restritos a um intervalo.

### 5.4.1 Introdução de um Modelo de Regressão Linear Possibilística para o Intervalo (0,1)

TANAKA, HAYASHI e WATADA (1989) estudam aplicações da regressão linear possibilística, investigando a solução para três problemas básicos: minimização, maximização e conjunção.

Na regressão possibilística, para uma função de pertinência triangular simétrica tem-se, no caso de minimização, a solução apresentada na FIGURA 5.1.

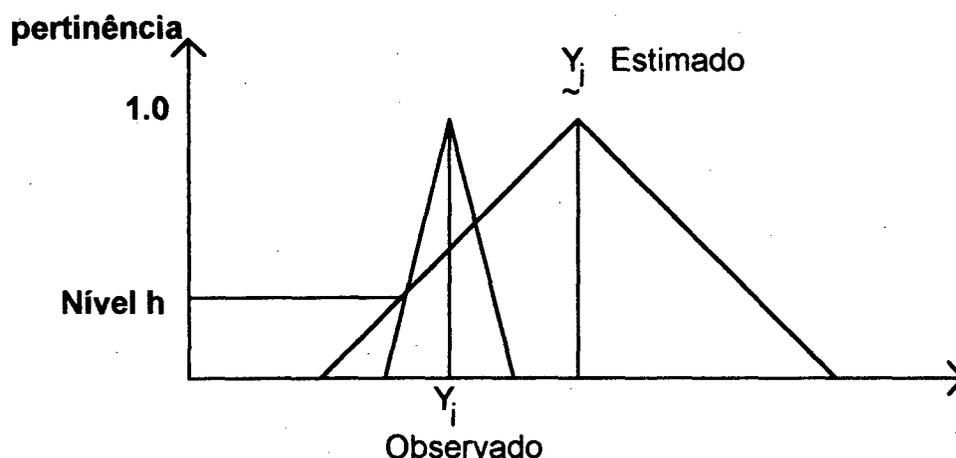


FIGURA 5.1 - Inclusão do Valor Observado no Valor Estimado a um Nível h

Para limitar-se os valores possíveis de  $Y$ , o vetor de resposta da variável dependente, ao intervalo (0,1), uma solução convencional é a utilização da transformação logarítmica para  $Y$ . Um problema que surge, entretanto, é o relacionado com o tratamento da dispersão do valor observado  $Y_i$ . A escala logarítmica não mantém a simetria admitida para o número difuso triangular. Uma solução para esse problema é admitir a simetria em uma escala logarítmica.

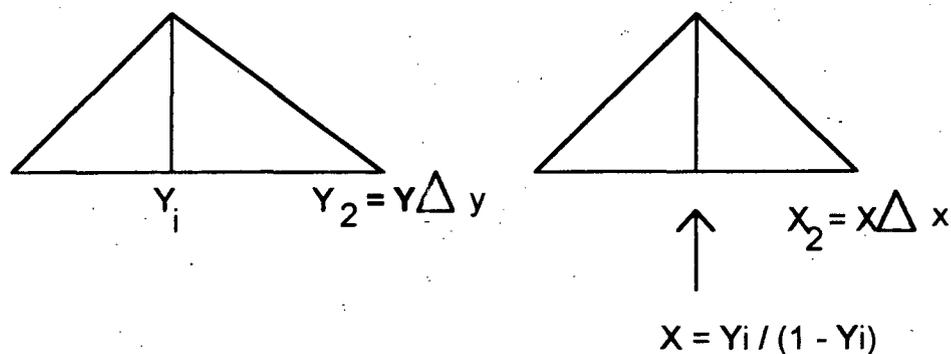


FIGURA 5.2 - Relações entre  $Y_i$  e a Transformada de  $g(y) = X = Y_i / (1 - Y_i)$

Analisando-se a FIGURA 5.2 e admitindo-se para a dispersão um efeito multiplicativo, obtém-se as seguintes relações:

$$g^{-1}(x_2) = y\Delta y \quad (5.12)$$

$$g^{-1}(x_1) = y\Delta^{-1}y \quad (5.13)$$

$$X = \frac{Y}{1 - Y} \quad (5.14)$$

$$x_2 = x\Delta x \quad (5.15)$$

$$x_1 = x\Delta^{-1}x \quad (5.16)$$

De (5.15) obtém-se:

$$g^{-1}(x_2) = \frac{x\Delta x}{1 + x\Delta x}$$

$$\frac{x\Delta x}{1 + x\Delta x} = y\Delta y$$

Rearranjando-se a expressão, tem-se:

$$\Delta x = \frac{y\Delta y}{x - yx\Delta y} \quad (5.17)$$

ou

$$\Delta x = \frac{y\Delta y}{\left(\frac{y}{1-y}\right)[1-y\Delta y]} \quad (5.18)$$

Mais especificamente:

$$\Delta x = \frac{\hat{f}(A)}{f(y)} \quad (5.19)$$

onde:  $\hat{f}(A)$  = função que transforma  $y\Delta y$  em  $x\Delta x$

$f(y)$  = função que leva  $y$  a  $x$

A regressão possibilística é efetuada, utilizando uma função triangular difusa simétrica sobre  $[\ln X; \ln \Delta x]$ . Os valores de  $\ln X$  e  $\ln X \pm \ln \Delta x$  não estão restritos ao intervalo  $(0, 1)$ . Para obter-se as estimativas de  $y$ , tem-se:

$$\text{Centróide: } \frac{e^X}{1+e^X}$$

$$\text{Limite Superior: } \frac{e^{x\Delta x}}{1+e^{x\Delta x}}$$

$$\text{Limite Inferior: } \frac{e^{x\Delta-1x}}{1+e^{x\Delta-1x}}$$

Nota-se, entretanto, que esse conjunto de equações conduz a um intervalo em  $y$  não simétrico.

### 5.4.2 Modelo de Regressão Possibilística Triangular Logit

Em várias situações de interesse prático, julgamentos são efetuados sobre uma escala definida no intervalo  $(0,1)$ . Valores próximos ao limite inferior da escala representam, para o decisor, uma possibilidade fraca sobre o fato ou situação confrontada. Valores próximos ao extremo superior, ao contrário, representam uma possibilidade forte sobre a ocorrência de tal fato.

O problema que subsiste, todavia, diz respeito a valores intermediários, dentro do intervalo  $(0,1)$ . Nesses casos, é razoável admitir um valor para o qual o decisor atribua pertinência máxima e, que, para valores à esquerda ou à direita, a pertinência vai diminuindo. Essa situação é ilustrada na FIGURA 5.3.

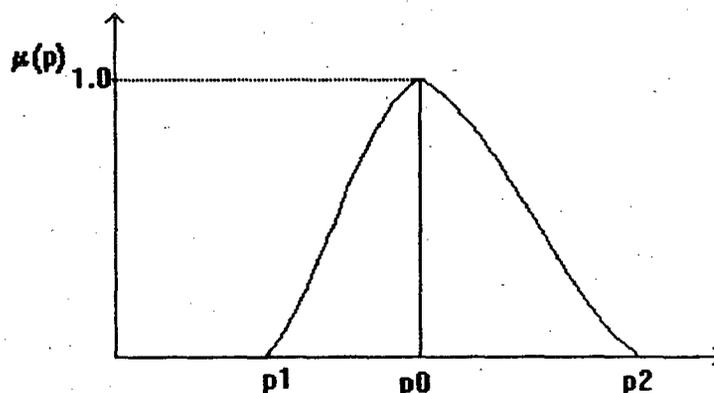


FIGURA 5.3 - Função de Pertinência de um Número Difuso  $P$  entre  $(0,1)$

Seja  $P$  um número difuso que assuma valores no intervalo  $(0,1)$ . Admita-se que  $P$  possa ser modelado através de  $X$ , um vetor de variáveis não difusas utilizadas pelo decisor para explicar  $P$ . Então, tem-se a seguinte definição:

**Definição 5.1:** Um modelo possibilístico Triangular Logit, para o número difuso  $P$ , tal como representado na FIGURA 5.3, é especificado como:

$$P = \mathfrak{F}(A, X) = \frac{1}{1 + e^{-AX}} \quad (5.20)$$

onde:  $A = (\alpha, c)$  = vetor de coeficientes difusos

$X$  = vetor de variáveis independentes

**Definição 5.2:** Uma regressão possibilística Triangular Logit consiste em determinar os valores de  $A^*$  tal que  $P^* = \mathfrak{F}(A^*, x) = \frac{1}{1 + e^{-A^*x}}$  seja uma estimativa de  $P$ .

**Definição 5.3:** O número  $P$  é um número difuso, definido por dois parâmetros:  $p_0$  - pertinência máxima - e  $p_2$  - pertinência zero à direita - com  $0 < p_0 < p_2 < 1$ ,  $p_1$  da forma:

$$p_1 = \frac{\left(\frac{p_0}{1-p_0}\right)^2 \left(\frac{1-p_2}{p_2}\right)}{1 + \left(\frac{p_0}{1-p_0}\right)^2 \left(\frac{1-p_2}{p_2}\right)} \quad (5.21)$$

e com a função de pertinência dos valores  $P$  dada por:

$$\mu(p) = \begin{cases} \frac{\ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{p_1}{1-p_1}\right)}{\ln\left(\frac{p_0}{1-p_0}\right) - \ln\left(\frac{p_1}{1-p_1}\right)} & \text{se } p_1 \leq p \leq p_0 \\ 1 - \frac{\left[\ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{p_2}{1-p_2}\right)\right]}{\left[\ln\left(\frac{p_0}{1-p_0}\right) - \ln\left(\frac{p_2}{1-p_2}\right)\right]} & \text{se } p_0 \leq p \leq p_2 \end{cases} \quad (5.22)$$

O número  $P$  é apropriado para modelar questões relacionadas a problemas de escolhas e/ou juízo de valores. A partir da definição (5.3) observa-se que são necessários apenas dois valores ( $p_0$  e  $p_2$ ) para defini-lo inequivocamente. O valor  $p_0$  representa o valor para o qual o decisor credita a possibilidade máxima. Para valores superiores, a possibilidade vai declinando até o valor  $p_2$ , a partir do qual as pertinências tem valor nulo. Os graus de pertinências à esquerda de  $p_0$  também diminuem até o limiar  $p_1$ .

**Lema 5.1:** O valor de  $p_1$  pertence ao intervalo  $(0,1)$ .

**Demonstração:** Analisando-se a expressão (5.21), verifica-se que o denominador pode ser igualado a uma constante. Assim, o valor de  $p_1$  é igual a:

$$p_1 = \frac{cte}{1+cte}, \text{ portanto } p_1 \text{ é menor do que } 1.$$

O primeiro termo do denominador de (5.21) é uma expressão quadrática, portanto positiva. O segundo termo do denominador envolve a expressão  $(1-p_2)$ . Como  $p_2$ , por hipótese, é um número entre  $(0,1)$ , então  $(1-p_2)$  também pertence a  $(0,1)$ . O produto de dois valores positivos é sempre positivo. Consequentemente,  $p_1$  é maior do que 0 e  $p_1 \in (0,1)$ .

**Lema 5.2:** Conhecidos os valores  $p_1$  - grau de pertinência zero à esquerda - e  $p_0$  - grau de pertinência máxima - o número difuso  $P$  fica caracterizado se o valor de  $p_2$  for:

$$p_2 = \frac{\left(\frac{p_0}{1-p_0}\right)^2 \left(\frac{1-p_1}{p_1}\right)}{1 + \left[\left(\frac{p_0}{1-p_0}\right)^2 \left(\frac{1-p_1}{p_1}\right)\right]}$$

**Demonstração:** O resultado é imediato isolando o valor de  $p_2$  na equação (5.21).

**Teorema 5.1:** Resolver a regressão possibilística Triangular Logit, apresentada na definição (5.2), é equivalente a resolver a regressão linear possibilística  $Y = \ln\left(\frac{P}{1-P}\right) = Ax$ .

**Demonstração:** Primeiramente será mostrado que  $Y$  é um número difuso triangular simétrico. Posteriormente demonstra-se que  $Y$  pode ser modelado como  $Y = Ax$ .

**1ª Parte:**  $Y$  é um número difuso triangular simétrico  $(\alpha, c)$ .

Aplicando-se a transformação de ligação logit para os valores  $p_0$ ,  $p_1$  e  $p_2$  dados pela definição (5.3), obtém-se, respectivamente:  $\ln\left(\frac{p_1}{1-p_1}\right)$ ,  $\ln\left(\frac{p_0}{1-p_0}\right)$  e  $\ln\left(\frac{p_2}{1-p_2}\right)$ . O valor de  $p_1$  é:

$$p_1 = \frac{\left(\frac{p_0}{1-p_0}\right)^2 \left(\frac{1-p_2}{p_2}\right)}{1 + \left(\frac{p_0}{1-p_0}\right)^2 \left(\frac{1-p_2}{p_2}\right)}$$

de onde obtém-se:

$$\ln\left[\left(\frac{p_2}{1-p_2}\right)\left(\frac{p_1}{1-p_1}\right)\right] = \ln\left(\frac{p_0}{1-p_0}\right)^2 \quad (5.23)$$

$$\ln\left(\frac{p_2}{1-p_2}\right) - \ln\left(\frac{p_0}{1-p_0}\right) = \ln\left(\frac{p_0}{1-p_0}\right) - \ln\left(\frac{p_1}{1-p_1}\right) \quad (5.24)$$

A última expressão (5.24) estabelece que a dispersão à direita e à esquerda são iguais podendo, portanto, serem representados por um único valor:

$$c = \ln\left(\frac{p_0}{1-p_0}\right) - \ln\left(\frac{p_1}{1-p_1}\right) \quad (5.25)$$

Sendo:

$$\mu(p) = \begin{cases} \frac{\ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{p_1}{1-p_1}\right)}{\ln\left(\frac{p_0}{1-p_0}\right) - \ln\left(\frac{p_1}{1-p_1}\right)} & \text{se } p_1 \leq p \leq p_0 \\ 1 - \frac{\ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{p_0}{1-p_0}\right)}{\ln\left(\frac{p_2}{1-p_2}\right) - \ln\left(\frac{p_0}{1-p_0}\right)} & \text{se } p_0 \leq p \leq p_2 \end{cases}$$

e fazendo  $\ln\left(\frac{p_0}{1-p_0}\right) = \alpha$ , então:

$$\mu\left(\ln\left(\frac{p}{1-p}\right)\right) = \mu(Y) = \begin{cases} \frac{Y - (\alpha - c)}{c} & \alpha - c \leq Y \leq \alpha \\ 1 - \left[\frac{Y - \alpha}{c}\right] & \alpha \leq Y \leq \alpha + c \end{cases} \quad (5.26)$$

$$\mu\left(\ln\left(\frac{p}{1-p}\right)\right) = \mu(Y) = \begin{cases} 1 - \frac{|\alpha - Y|}{c} & \alpha - c \leq Y \leq \alpha \\ 0 & \text{outros valores} \end{cases} \quad (5.27)$$

que é a função de pertinência para um número triangular difuso simétrico  $Y=(\alpha,c)$ .

**2ª. Parte:** Y é modelado como  $Y=Ax$ .

$P^* = \mathfrak{J}(A^*, x) = \frac{1}{1 + e^{-A^*x}}$ . Esse modelo pode ser escrito, mediante transformações

simples como  $\ln\left(\frac{p^*}{1-p^*}\right) = A^* x$ , então  $Y = A^* x$ .

**Corolário:** O grau de inclusão de P em  $P^*$  é menor ou igual a h, onde h é o nível de inclusão adotado por Tanaka no ajuste do problema de regressão linear possibilística  $Y = Ax$ .

**Demonstração:** Será feita apenas para o problema de minimização. Para os problemas de conjunção e maximização a demonstração é similar.

Por definição:

$$\begin{aligned} \mu(p) &= \mu(y) \\ \mu_{p^*}(p) &= \mu_{y^*}(y) \\ &= 1 - \frac{|y - y^*|}{c|x| - \text{spread}Y} \\ \mu_{p^*}(p) &= 1 - \frac{\ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{p^*}{1-p^*}\right)}{\left[\ln\left(\frac{p_0^*}{1-p_0^*}\right) - \ln\left(\frac{p^*}{1-p^*}\right)\right]|x| - \left[\ln\left(\frac{p_0}{1-p_1}\right) - \ln\left(\frac{p}{1-p}\right)\right]} \end{aligned} \quad (5.28)$$

### 5.4.3 - Modelos para Análise de Desempenho das Escolas utilizando-se Conjuntos Difusos

Os modelos propostos com a finalidade de analisar os desempenhos das escolas são da forma:

$$\text{MODELO 4: } Y = A_0 + A_1 \text{Sexo} + A_2 \text{SSE} + A_3 \text{INDANO} \quad (5.29)$$

$$\text{MODELO 5: } Y = A_0 + A_1 \text{SSE} + A_2 \text{INDANO} \quad (5.30)$$

onde:  $Y = \log\left(\frac{\% \text{ de aprovação na categoria}}{1 - \% \text{ de aprovação na categoria}}\right)$ , calculado para cada categoria de aluno observada na  $i$ -ésima escola

$A_i$  = um número triangular difuso simétrico,  $(\alpha, c)$

Esses modelos irão medir, para uma escola específica  $i$ , os efeitos dos fatores Sexo, SSE e INDANO sobre a possibilidade de aprovação. Porém, é importante salientar que o considerado difuso é, justamente, o efeito de cada um dos fatores. Em outras palavras, admite-se não ser possível descrever, por meio de uma distribuição de probabilidades, um comportamento estocástico para, por exemplo, os efeitos do fator SSE.

Nesses modelos, as contribuições derivadas das características individuais dos alunos conduzem a uma estrutura paramétrica difusa. Além desse fato, normalmente, os conceitos relacionados à qualidade de instalações, corpo docente, condições da escola de modo geral, apresentam-se como conceitos derivados de linguagem natural e que, portanto, podem ser considerados difusos e imprecisos. Assim, pode ser realizada a aplicação de modelos de regressão possibilística para modelos hierarquizados em dois níveis.

Pode-se, então ajustar os seguintes modelos:

#### MODELO 6:

$$A_{ji} = B_{1j}ADS_i + B_{2j}EC_i + B_{3j}LOC_i + B_{4j}To2_i; \text{ com } j = 1, \dots, 4 \text{ e } i = 1, \dots, N \text{ (5.31)}$$

onde:  $A_{ji}$  = coeficientes difusos obtidos pela aplicação do modelo (5.30) nas escolas

Recombinando-se os MODELOS 4 e 5, é possível avaliar os efeitos derivados das características das escolas sobre a possibilidade de aprovação. Tais efeitos podem ser obtidos da seguinte forma:

- **Adequação das Salas de Aula:**

$$(B_{11} + B_{12}IndAno + B_{13}SSE + B_{14}Sexo)$$

- **Estado de Conservação:**

$$(B_{21} + B_{22}IndAno + B_{23}SSE + B_{24}Sexo)$$

- **Localização:**

$$(B_{31} + B_{32}IndAno + B_{33}SSE + B_{34}Sexo)$$

- **Taxa de Ocupação:**

$$(B_{41} + B_{42}IndAno + B_{43}SSE + B_{44}Sexo)$$

## 5.5 Modelagem do Desempenho das Escolas através de Análise Taxionômica

Aplicando-se os MODELOS 1 e 5 obtém-se, em termos numéricos, probabilidades e possibilidades de aprovação quando considerados valores relativos a características individuais de alunos. Para cada tipo de aluno, as diversas escolas integrantes de um sistema educacional devem apresentar valores de probabilidade e de possibilidade de aprovação diferentes.

Uma questão que se coloca, então, é a forma de classificar essas escolas de acordo com os resultados numéricos obtidos através dos modelos.

Em termos de política de educação, pode-se, por exemplo, argumentar que escolas possam ser monitoradas segundo algum padrão idealizado. Escolas que se mantenham em determinados patamares receberão incentivos adicionais. Aquelas que, dentro de uma estratégia preestabelecida, estiverem apresentando valores abaixo da expectativa poderão receber um acompanhamento mais freqüente, de modo a ultrapassar eventuais dificuldades.

Um critério a utilizar é hierarquizar, para cada escola, os valores obtidos para a aprovação de um determinado aluno. Esse critério pode ser estendido a todas as combinações de tipos de alunos e todas as escolas. Evidentemente, o número de classificações depende diretamente do número de alunos padrão definidos. Mais razoável é, portanto, admitir para cada escola uma classificação média em termos de seus resultados.

### **5.5.1 - Processo de Classificação das Escolas utilizando-se os Resultados do Modelo Logístico Tradicional**

As escolas podem ser agrupadas de acordo com os valores de probabilidade de aprovação obtidos para cada um dos tipos de alunos identificados.

A aplicação de uma análise de agrupamento permite identificar escolas com padrões de aprovação similar e, conseqüentemente, obter uma estrutura que melhor caracterize o desempenho das escolas dentro do sistema educacional.

Uma vez identificados os grupos de escolas com seus respectivos rendimentos (expresso em termos das probabilidades de aprovação) os gestores do sistema educacional dispõem de maiores facilidades para a implantação de políticas direcionadas à manutenção da qualidade e desempenho das escolas.

Tem-se, então, a população das escolas subdivididas em subpopulações, nas quais o desempenho verificado é similar, ou seja:

$$[G_{R1}, G_{R2}, \dots, G_{RN}] \quad (5.32)$$

onde:  $G_{Ri}$  = grupo de escola com desempenho similar

### 5.5.2 - Processo de Classificação das Escolas utilizando-se os resultados do Modelo Logit Triangular Difuso

Pretendendo classificar as escolas de acordo com as suas possibilidades de aprovação em, por exemplo, três níveis “Baixa Possibilidade”, “Média Possibilidade” e “Alta Possibilidade”, um primeiro problema que se depara é o de precisar limites entre tais níveis. O conceito de variável lingüística, oriundo da teoria dos conjuntos difusos, se ajusta perfeitamente para a modelação dessa situação.

Nesse trabalho é proposto o seguinte procedimento para associar a “ordenação” de um grupo de escolas (avaliado em termos das possibilidades de aprovação) e um critério global de classificação. O critério de classificação é explicitado através de uma variável lingüística definida com os seguintes rótulos: “Abaixo da Média do Grupo”, “Na Média do Grupo” e “Acima da Média do Grupo” (ver FIGURA 5.4):

- ajustar, para cada escola do grupo, os modelos 4 e 5;

- estimar, para cada uma das categorias de alunos definidas, as possibilidades de aprovação para o ponto centróide, de acordo com os modelos 4 e 5;
- ordenar as escolas em valores crescentes das estimativas obtidas no item anterior, para cada uma das categorias de alunos definidas;
- calcular a posição global de cada escola, em termos das ordenações parciais realizadas no item anterior;
- transformar as posições globais de cada escola, através de:

$$\frac{PGR_i - P_{\text{Min}}}{P_{\text{Max}} - P_{\text{Min}}}$$

onde:  $PGR_i$  = posição global relativa da escola  $i$ ;

$P_{\text{Min}}$  = posição mínima;

$P_{\text{Máx}}$  = posição máxima.

- repetir os passos dois a cinco, para os valores (centróide - dispersão) e (centróide + dispersão) obtendo-se estimativas pessimistas e estimativas otimistas.

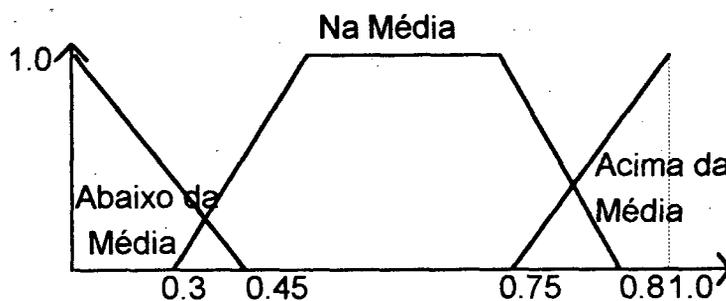


FIGURA 5.4 - Um Critério para Classificação das Escolas quanto a Possibilidade de Aprovação.

A cada uma das classificações obtidas, é atribuído um valor de pertinência, segundo o qual uma escola é classificada como: "Abaixo da Média do Grupo", "Na Média do Grupo" e "Acima da Média do Grupo". Calculando-se a média difusa das classificações obtidas e reordenando-se essas médias, obtém-se

uma hierarquização das escolas. Esse processo de ordenação tem, como vantagem, a utilização de limites difusos ou nebulosos. Em processos tradicionais, baseados na lógica clássica, os limites são arbitrários, não ocorrendo de forma gradual, suscitando controvérsias difíceis de serem sustentadas. A mesma forma de agrupamento (análise de agrupamento) das escolas, sugerido para os resultados dos modelos 1 e 2, pode ser, também, aplicado.

## 5. 6 Conclusões

Neste capítulo, foram apresentados modelos para avaliação de desempenho de sistemas educacionais, nos quais a variável dependente está restrita ao intervalo (0,1).

Os modelos são construídos considerando-se características individuais dos alunos e características dos estabelecimentos de ensino que podem afetar o desempenho dos estudantes e, conseqüentemente, do sistema educativo.

É, também, derivado um modelo de regressão possibilística, no qual a variável dependente é restrita ao intervalo (0,1). Através desse modelo é possível tratar situações de escolha qualitativa, nas quais subsistem nebulosidades ou imprecisões quanto à importância dos fatores considerados como relevantes na explicação da escolha.

Para utilização do modelo, dois pontos no intervalo (0,1) são solicitados ao decisor. Obviamente, quanto maior for o grau de incerteza presente no julgamento, maiores serão os intervalos estimados.

Os parâmetros para ajustes do modelo podem ser considerados os mesmos propostos em trabalhos sobre regressão possibilística (TANAKA, HAYASHI e WATADA. (1989), KANDEL e HASHAMANI (1991) entre outros).

Modelos para avaliação de desempenho de sistemas educacionais, principalmente, em questões relacionadas com o aproveitamento são definidos à luz da teoria dos conjuntos difusos. Tais modelos conduzem a uma abordagem diferente da abordagem tradicional, permitindo a modelagem de fenômenos difusos ou imprecisos que ocorrem nos sistemas educacionais.

São propostos modelos de classificação das escolas com base tanto na modelagem tradicional quanto na modelagem através da lógica difusa.

Com relação ao processo de hierarquização das escolas utilizando-se a teoria dos conjuntos difusos, ressalta-se a sua potencialidade de adaptação a diferentes situações, refletindo, desse modo, as disparidades regionais existentes em sistemas educacionais. Assim sendo, vantagens imediatas são obtidas pelos responsáveis pela gestão do sistema educacional, tornando-se uma importante ferramenta gerencial.

Aplicações dos modelos desenvolvidos serão apresentadas no próximo capítulo, para escolas do Distrito de Faro - Portugal.

## **6. APLICAÇÕES DOS MODELOS**

### **6.1 - Introdução**

Neste capítulo são realizadas aplicações dos modelos anteriormente propostos para escolas do Distrito de Faro - Portugal.

Para ajustar os modelos baseados na lógica difusa foi desenvolvido um sistema computacional para tratamento de dados difusos. Para ajuste dos modelos estatísticos foram utilizados programas tradicionais de análise estatística.

Os dados utilizados neste estudo de caso foram obtidos a partir de informações coletadas durante o período de matrículas de alunos do 7º, 8º e 9º ano de escolarização do ciclo secundário, no ano escolar 89/90. Essas informações pertencem à base de dados do Gabinete de Estudos e Planeamento (GEP) do Ministério da Educação de Portugal.

### **6.2 - Caracterização das Escolas**

A TABELA 6.1 apresenta as escolas analisadas e o total de alunos. A codificação adotada para identificação das escolas segue a utilizada pelo GEP. Nessa codificação, os dois primeiros dígitos referem-se ao Distrito de Faro e os dois subseqüentes, ao Concelho no qual a escola está localizada.

TABELA 6.1 - Escolas Analisadas e Total de Alunos

ESCOLAS		TOTAL DE ALUNOS
N. de Ordem	Código e Localização	
1	08010110 - Concelho de Albufeira	571
2	08020104 - Concelho de Alcoutim	93
3	08050413 - Concelho de Faro	1105
4	08050506 - Concelho de Faro	595
5	08050507 - Concelho de Faro	398
6	08050513 - Concelho de Faro	94
7	08060308 - Concelho de Lagoa	478
8	08060403 - Concelho de Lagoa	111
9	08070502 - Concelho de Lagos	293
10	08070608 - Concelho de Lagos	157
11	08080108 - Concelho de Loule	365
12	08080503 - Concelho de Loule	566
13	08080810 - Concelho de Loule	670
14	08080909 - Concelho de Loule	249
15	08090315 - Concelho de Monchique	253
16	08100208 - Concelho de Olhao	255
17	08110310 - Concelho de Portimao	753
18	08110311 - Concelho de Portimao	861
19	08110331 - Concelho de Portimao	355
20	08120114 - Concelho de S. B. de Alportel	330
21	08130206 - Concelho de Silves	242
22	08130517 - Concelho de Silves	357
23	08140607 - Concelho de Tavira	709
24	08160210 - Concelho de V.R. de S. António	1005

O número de alunos considerados, por escola, e a distribuição segundo o ano de escolarização são mostrados na TABELA 6.2.

TABELA 6.2 - Escolas Analisadas e Total de Alunos Considerados na Aplicação,  
Segundo o Ano de Escolaridade

ESCOLAS	7 ANO	8 ANO	9 ANO	TOTAL
8010110	38	129		167
8020104	29			29
8050413	45	208	276	529
8050506	16	143	128	287
8050507	20	89	101	210
8050513	35	27	27	89
8060308	3	8	42	53
8060403	27	34	25	86
8070502	7	43	61	111
8070608	115			115
8080108	84	88	23	195
8080503	177	125	66	368
8080810	21	173	115	309
8080909	203			203
8090315	72	50	34	156
8100208	22	56	43	121
8110310	37	134	196	367
8110311	82	196	181	459
8110331	16			16
8120114	75	75	42	192
8130206	43	46	44	133
8130517	98	96	44	238
8140607	29	161	127	317
8160210	30	252	192	474

Os valores apresentados na TABELA 6.2 são derivados de pré-análises efetuadas no conjunto de dados global. Foram considerados apenas alunos matriculados na mesma escola, durante dois períodos letivos consecutivos.

Na TABELA 6.3 é apresentado o número de alunos, por sexo, do 7º, 8º e 9º ano, aprovados e reprovados no ano escolar 89/90. A TABELA 6.4 apresenta a distribuição dos estudantes considerados neste estudo, segundo o grupo sócio econômico ao qual o aluno pertence. Os grupos foram definidos de acordo com a análise efetuada no capítulo anterior. As TABELAS 6.5, 6.6 e 6.7 apresentam, respectivamente, a distribuição dos alunos aprovados e reprovados, por idade, para o sétimo, oitavo e nono ano de escolaridade.

TABELA 6.3 - Alunos Aprovados e Reprovados, por Escola

ESCOLAS	APROVADOS			REPROVADOS		
	MAS.	FEM.	TOTAL	MAS.	FEM.	TOTAL
8010110	61	37	98	24	45	69
8020104	6	12	18	5	6	11
8050413	213	206	419	39	71	110
8050506	131	89	220	34	33	67
8050507	74	75	149	26	35	61
8050513	87	0	87	2	0	2
8060308	23	9	32	10	11	21
8060403	37	40	77	5	4	9
8070502	43	51	94	7	10	17
8070608	52	49	101	5	9	14
8080108	78	75	153	16	26	42
8080503	143	134	277	44	47	91
8080810	137	110	247	28	34	62
8080909	115	80	195	3	5	8
8090315	70	67	137	6	13	19
8100208	63	48	111	3	7	10
8110310	143	101	244	77	46	123
8110311	186	146	332	51	76	127
8110331	8	8	16	0	0	0
8120114	83	70	153	18	21	39
8130306	66	43	109	12	12	24
8130517	95	83	178	28	32	60
8140607	130	124	254	29	34	63
8160210	209	171	380	43	51	94

TABELA 6.4 - Alunos por Grupos Sócio Econômicos, por Escola

ESCOLAS	GRUPO 1	GRUPO 2	GRUPO 3	GRUPO 4	GRUPO 5	TOTAL
8010110	7	57	87	15	1	167
8020104	2	16	10	1		29
8050413	12	202	258	49	8	529
8050506	6	84	127	57	13	287
8050507	5	45	109	50	1	210
8050513		14	55	12	8	89
8060308	2	23	24	4		53
8060403		24	45	13	4	86
8070502	5	37	58	10	1	111
8070608	5	48	56	5	1	115
8080108		48	146	1		195
8080503	31	127	173	32	5	368
8080810	10	115	148	32	4	309
8080909	7	55	129	11	1	203
8090315	13	55	80	8		156
8100208	10	21	82	8		121
8110310	11	128	173	47	8	367
8110311	7	163	216	66	7	459
8110331		1	11	4		16
8120114	6	108	61	17		192
8130206	6	30	94	2	1	133
8130517	9	102	115	10	2	238
8140607	14	132	147	17	7	317
8160210	26	174	220	47	7	474

TABELA 6.5 - Alunos Aprovados e Reprovados do 7º Ano, por Escola, Segundo a Idade

7 ANO	APROVADOS					REPROVADOS				
	IDADE					IDADE				
ESCOLAS	12	13	14	15	≥16	13	14	15	16	≥17
8010110	3	1				8	14	7	4	1
8020104	8	3	6		1	5		5	1	
8050413		1			1	7	17	13	5	1
8050506						8	6	1	1	
8050507		1				9	5	1	2	2
8050513	32	2	1							
8060308		2							1	
8060403	16	5	1			3	1	1		
8070502						2	2	1	1	1
8070608	41	35	15	8	2	8	6			
8080108	30	16	10	5		5	9	7	1	1
8080503	28	53	23	10	2	23	21	12	3	2
8080810						6	6	6	2	1
8080909	128	37	23	6	1	4	2	2		
8090315	31	15	10	4	2	2	5	2	1	
8100208	6	6	4		1	2	2	1		
8110310						16	16	3	1	1
8110311						20	34	22	5	1
8110331	8	2	3	2	1					
8120114	22	15	8	4	3	1	13	2	4	2
8130206	15	7	7	4		4	3	1		2
8130517	37	22	4	11	3	6	11	2	1	1
8140607	2		1		1	12	7	4	1	1
8160210		1	1	1		9	8	6	3	1

TABELA 6.6 - Alunos Aprovados e Reprovados do 8º Ano, por Escola, Segundo a Idade

8 ANO	APROVADOS					REPROVADOS				
	IDADE					IDADE				
ESCOLA	13	14	15	16	≥17	14	15	16	17	≥18
8010110	40	21	20	9	4	11	9	9	4	2
8050413	90	31	24	9	3	10	21	11	7	2
8050506	81	18	13	1	5	12	3	7	1	2
8050507	31	14	12	6	2	10	5	7	2	
8050513	22	3				1	1			
8060308	2	1	1	1			1	2		
8060403	18	11	2	2					1	
8070502	16	5	14	2		2	3		1	
8080108	35	25	6	4		7	4	4	2	1
8080503	33	43	13	8	3	12	7	6		
8080810	83	36	17	7	2	8	8	7	4	1
8090315	17	16	7	2	2	2	1	2	1	
8100208	19	13	15	3	1	2	1	1	1	
8110310	51	23	14	7	1	14	9	11	4	
8110311	114	33	11	9	3	15	4	4	3	
8120114	27	24	7	4	2	1	2	5	3	
8130206	14	9	5	7	3	4	1	3		
8130517	34	18	13		1	8	13	6	2	1
8140607	66	34	20	14	3	9	6	6	2	1
8160210	90	67	39	12	7	15	13	8	1	

**TABELA 6.7 - Alunos Aprovados e Reprovados do 9º Ano, por Escola, Segundo a Idade**

9 ANO	APROVADOS					REPROVADOS				
	IDADE					IDADE				
ESCOLAS	14	15	16	17	≥18	15	16	17	18	≥19
8050413	132	68	37	21	2	4	7	1	4	
8050506	69	13	10	10		12	4	6	3	1
8050507	47	14	15	4	3	2	7	5	1	3
8050513	23	4								
8060308	14	4	4	1	2	4	6	5	2	
8060403	9	8	3	1	1	1		2		
8070502	19	21	8	8	1	1	1	2		
8080108	8	7	4	1	2		1			
8080503	31	10	16	4		1	4			
8080810	62	20	10	5	5	6	3	2	2	
8090315	15	6	5	5			1	2		
8100208	14	10	9	5	5					
8110310	65	44	28	9	2	15	18	13	2	
8110311	48	47	30	22	15	4	3	5	7	
8120114	9	14	9	4		2	2	1	1	
8130206	14	11	8	3	2		2	4		
8130517	16	7	8	3	1	2	3	3		1
8140607	70	16	17	10		2	4	6	2	
8160210	86	35	23	15	3	9	9	9	2	1

### **6.3 Aplicação dos Modelos para Análise do Desempenho das Escolas, baseados na Aprovação dos Alunos e nas suas Características Individuais (Modelos 1 e 2)**

O MODELO 1 foi aplicado nas vinte e quatro escolas do Distrito de Faro, obtendo-se os resultados mostrados na TABELA 6.8. Para a aplicação dos MODELOS 1 e 2 foram considerados os alunos agrupados em categorias, como descrito na TABELA 6.15.

Analisando-se esses resultados, verifica-se que a influência do fator INDANO é imediata sobre a probabilidade de aprovação dos alunos. Enquanto a influência do fator INDANO é sempre negativa, os fatores SEXO e SSE apresentam sinais variáveis nas diferentes escolas.

Para verificar a aleatoriedade, em relação à influência dos fatores SEXO e SSE nas diferentes escolas, dois testes não paramétricos foram efetuados. Para o fator SSE, o valor mediana obtido é igual a -0,0795 conduzindo a um valor da estatística Z igual a 0,2087, com significância de 0,8347. O mesmo teste de aleatoriedade, baseado em trocas alternadas do tipo "*up and down*" apresenta, como resposta, Z igual a 1,42661 e significância 0,1536. Para o fator SEXO, o teste de aleatoriedade baseado no valor mediana (-0,1345) resultou em um valor de Z igual a -0,2087 e significância de 0,8347. No teste de aleatoriedade do tipo "*up and down*" para o fator SEXO obteve-se a estatística Z igual a 1,93012 com significância de 0,0535. Esses resultados podem sugerir uma pequena tendência, em algumas escolas, de que exista uma maior aprovação em alunos de um determinado sexo em relação a alunos de outro sexo.

Testes não paramétricos também foram realizados sobre os valores dos coeficientes dos fatores SSE e SEXO com a finalidade de comparar as diferenças entre esses coeficientes. A hipótese básica testada foi a de que não existem diferenças, isto é, os coeficientes são alternados, porém, sem predominância de um tipo de sinal. A comparação entre os dois conjuntos de coeficientes, realizada através do teste dos sinais, apresentou um valor de Z igual a -0,204124 com significância de 0,8382. A mesma hipótese testada através do teste de "Postos", apresentou valor de Z igual a 0,3 com significância de 0,7641.

TABELA 6.8 - Resultados do MODELO 1 - Escolas do Distrito de Faro

Modelo 1								
Escolas		N. Obs	Interc.	SEXO	SSE	INDANO	r <sup>2</sup>	F
1	8010110	23	3.919	-0.336	0.180	-1.587	0.7441	18.41
2	8020104	13	3.603	-0.386	-0.688	-1.124	0.5523	3.70
3	8050413	25	5.473	-0.248	0.077	-2.013	0.7452	20.47
4	8050506	27	4.219	-0.359	-0.091	-1.514	0.6540	14.49
5	8050507	22	4.987	0.055	-0.581	-1.835	0.7884	22.35
6	8050513	6	3.489	3.282	-0.732	-2.517	0.8877	5.27
7	8060308	17	4.600	-0.582	-0.698	-1.400	0.7814	15.49
8	8060403	15	5.611	0.360	-1.826	-1.549	0.8179	16.47
9	8070502	21	3.466	-0.175	0.114	-1.006	0.5420	6.71
10	8070608	17	4.253	-0.138	-0.563	-1.177	0.4893	4.15
11	8080108	13	2.774	0.726	0.895	-1.964	0.7421	8.63
12	8080503	26	4.774	-0.060	-0.621	-1.551	0.6622	14.34
13	8080810	26	4.915	-0.007	-0.457	-1.725	0.6163	11.78
14	8080909	19	4.484	-0.691	1.038	-1.368	0.6402	8.90
15	8090315	21	4.101	-0.170	0.330	-1.396	0.6189	9.20
16	8100208	19	4.228	-0.962	0.572	-0.892	0.4479	4.06
17	8110310	28	3.836	0.477	-0.311	-1.918	0.7247	21.07
18	8110311	26	5.271	-0.873	0.953	-2.045	0.7759	25.38
19	8110331	9	1.238	0.252	0.224	-0.254	0.2066	0.43
20	8120114	20	5.799	-0.131	-0.068	-2.088	0.8959	45.66
21	8130206	20	2.641	-0.073	0.246	-0.838	0.4890	5.10
22	8130517	23	3.609	0.214	-0.475	-1.366	0.6591	12.25
23	8140607	28	3.966	-0.153	0.107	-1.329	0.4990	7.97
24	8160210	26	4.699	0.130	-0.104	-1.880	0.8157	32.45

Para o modelo 2, os resultados são descritos na TABELA 6.9. Os coeficientes do fator SSE não se mostram significativos. Na FIGURA 6.1 são apresentados os intervalos de confiança de 95% para os coeficientes do MODELO 2. Na FIGURA 6.2 são mostrados os intervalos de confiança para o coeficiente INDANO para o mesmo modelo.

TABELA 6.9 - Resultados do MODELO 2 - Escolas do Distrito de Faro

Modelo 2							
Escolas		N. Obs	Interc.	SSE	INDANO	r <sup>2</sup>	F
1	8010110	13	3.735	0.451	-1.851	0.7957	19.47
2	8020104	9	2.858	-0.609	-1.016	0.4528	2.48
3	8050413	13	5.356	0.419	-2.222	0.7701	16.74
4	8050506	14	4.170	-0.452	-1.491	0.6757	11.46
5	8050507	13	4.961	-0.492	-1.863	0.7356	13.91
6	8050513	6	6.770	-0.732	-2.517	0.8877	11.86
7	8060308	10	4.069	-0.292	-1.699	0.8085	14.78
8	8060403	9	7.062	-1.981	-1.738	0.8180	13.49
9	8070502	12	4.169	-0.246	-1.260	0.6276	7.59
10	8070608	10	4.300	-0.436	-1.289	0.4882	3.34
11	8080108	7	5.372	-0.828	-1.895	0.7742	6.86
12	8080503	14	4.656	-0.513	-1.551	0.8125	23.84
13	8080810	14	5.735	-0.316	-2.097	0.7531	16.77
14	8080909	10	4.319	1.266	-1.725	0.7317	9.55
15	8090315	11	4.905	0.130	-1.805	0.8045	16.46
16	8100208	12	4.111	-0.065	-1.220	0.4823	4.19
17	8110310	15	4.957	-0.306	-2.089	0.7819	21.51
18	8110311	14	3.992	0.877	-2.102	0.7253	14.52
19	8110331	7	1.506	0.318	-0.199	0.0683	0.15
20	8120114	11	6.582	-0.069	-2.454	0.8569	23.94
21	8130206	12	3.014	-0.178	-0.857	0.5129	4.74
22	8130517	12	4.365	-0.344	-1.587	0.8578	27.15
23	8140607	15	4.644	0.189	-1.691	0.6178	9.70
24	8160210	13	4.490	0.291	-2.129	0.8790	36.33

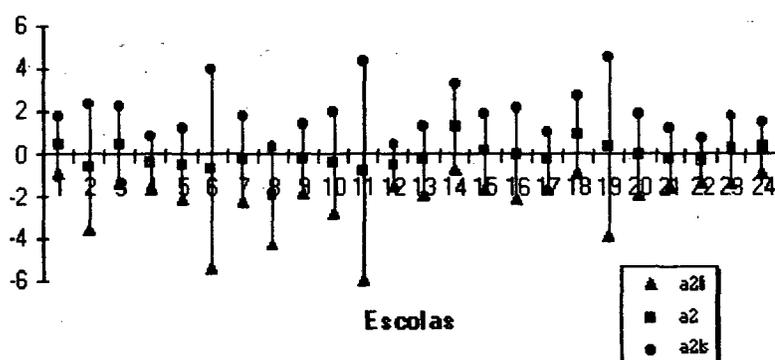
Intervalos de Confiança de 95% para os Coeficientes do Fator SSE (a<sub>2</sub>)

FIGURA 6.1 - Intervalos de Confiança para Fator SSE - MODELO 2

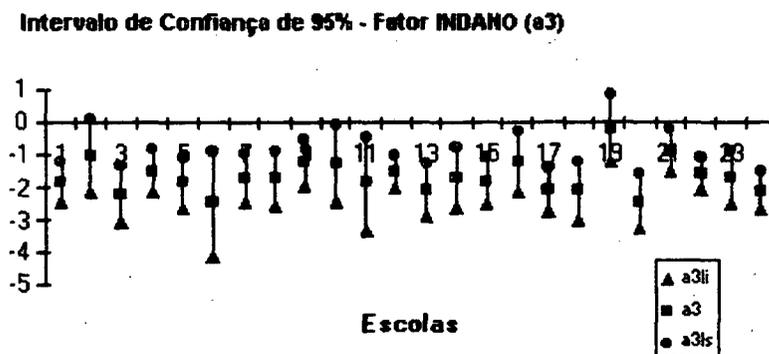


FIGURA 6.2 - Intervalos de Confiança para Fator INDANO - MODELO 2

Também, no MODELO 2, os valores relativos aos coeficientes do fator INDANO mostraram-se influentes. A exceção das escolas 2 e 19, todos os valores são estatisticamente significantes a 5%.

#### 6.4 - Modelo Hierarquizado - Aplicação do MODELO 3

Admitindo os valores dos coeficientes como sujeitos a variações devido a fatores estruturais e contextuais das escolas, a contribuição de cada uma das características individuais dos alunos agrega uma informação contextual da escola. Tais efeitos são estabelecidos através do MODELO 3.

Considerando as variáveis Adequação das Salas de Aula (ADS), Localização do Estabelecimento Escolar (LA), Estado de Conservação (EC) e Taxa de Ocupação (To2), tal como definidas no capítulo anterior, tem-se na TABELA 6.10, os valores para dezessete escolas das vinte e quatro analisadas, devido a disponibilidade de dados para as variáveis contextuais.

TABELA 6.10 - Escolas Utilizadas para Estimação dos Modelos Hierárquicos -  
Valores para Variáveis Contextuais

	Escolas	EC	LOC	ADS	To2
1	8010110	3	3	0.3	1.29
2	8020104	3	1	2.2	0.43
3	8050413	3	1	2.2	1.42
4	8050506	2	3	2.2	1.2
5	8050507	2	3	2.3	0.92
9	8070502	1	3	1.5	0.96
10	8070608	3	1	1.3	0.68
12	8080503	3	3	1.2	1.68
13	8080810	2	3	2.2	1.12
14	8080909	3	3	1.9	1.08
15	8090315	3	3	1.7	0.85
17	8110310	2	1	2.3	1.43
18	8110311	3	3	1	0.98
21	8130206	3	1	0.8	0.98
22	8130517	3	3	0.7	3.92
23	8140607	3	1	1.8	0.87
24	8160210	3	1	1.7	1.21

A aplicação do MODELO 3, utilizando-se os valores dos coeficientes do INTERCEPTO, obtidos para o MODELO 1, como variáveis dependentes conduziu aos resultados apresentados na TABELA 6.11.

TABELA 6.11 - Resultados do MODELO 3 - Variável Dependente: Coeficientes do INTERCEPTO

Efeitos Contextuais	Coeficientes	Erro Padrão	Estatística t	Valor P	95% Inferior	95% Superior
EC	0.735	0.164	4.475	0.000	0.380	1.090
LOC	0.419	0.149	2.800	0.012	0.096	0.742
ADS	0.861	0.206	4.184	0.001	0.416	1.305
To2	-0.001	0.234	-0.004	0.997	-0.506	0.504

A FIGURA 6.3. apresenta o gráfico dos resíduos para esse ajuste.

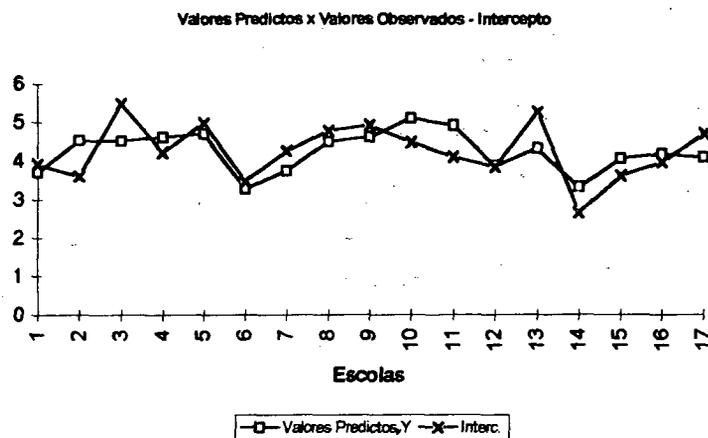


FIGURA 6.3 - Gráfico de Resíduos para Ajuste do MODELO 3: Coeficientes do INTERCEPTO

Para os valores do fator SEXO, os resultados são mostrados na TABELA 6.12. As TABELAS 6.13 e 6.14 apresentam, respectivamente, os resultados para os fatores SSE e INDANO.

TABELA 6.12 - Resultados do MODELO 3 - Variável Dependente: Coeficientes da Variável SEXO

Efeitos Contextuais	Coefficientes	Erro Padrão	Estatística t	Valor P	95% Inferior	95% Superior
EC	-0.150	0.063	-2.395	0.028	-0.286	-0.015
LOC	-0.145	0.057	-2.534	0.021	-0.268	-0.021
ADS	0.121	0.079	1.537	0.143	-0.049	0.291
To2	0.282	0.089	3.154	0.006	0.089	0.475

TABELA 6.13 - Resultados do MODELO 3 - Variável Dependente: Coeficientes da Variável SSE

Efeitos Contextuais	Coefficientes	Erro Padrão	Estatística t	Valor P	95% Inferior	95% Superior
EC	0.131	0.127	1.033	0.316	-0.143	0.406
LOC	0.144	0.116	1.243	0.231	-0.106	0.393
ADS	-0.228	0.159	-1.431	0.171	-0.571	0.116
To2	-0.274	0.181	-1.516	0.148	-0.664	0.117

**TABELA 6.14 - Resultados do MODELO 3 - Variável Dependente: Coeficientes da Variável INDANO**

Efeitos Contextuais	Coeficientes	Erro Padrão	Estatística t	Valor P	95% Inferior	95% Superior
EC	-0.218	0.090	-2.411	0.027	-0.413	-0.023
LOC	-0.106	0.082	-1.290	0.214	-0.284	0.071
ADS	-0.349	0.113	-3.085	0.007	-0.593	-0.105
To2	-0.112	0.128	-0.873	0.395	-0.390	0.165

Analisando os valores apresentados nas TABELAS 6.11 à 6.14, pode-se concluir que os efeitos contextuais relacionados ao Estado de Conservação do estabelecimento (EC) são significativos para os coeficientes SEXO e INDANO. Com relação aos efeitos Localização (LOC) e Taxa de Ocupação (To2), esses são significativos para explicar os valores do coeficiente INDANO. Já o coeficiente SSE não se mostrou significativo para nenhum dos efeitos contextuais considerados. Esse resultado era esperado, dado que, o coeficiente não foi significativo em nenhuma das escolas nas quais foram aplicados os MODELOS 1 e 2 (ver FIGURA 6.1).

### **6.5 - Modelos para Análise do Desempenho das Escolas Baseados na Aprovação dos Alunos e nas suas Características Individuais (Modelos 4 e 5) Utilizando-se Conjuntos Difusos**

O objetivo principal deste item é o de demonstrar que aplicações de técnicas derivadas da teoria de conjuntos difusos podem esclarecer e mesmo incorporar informações que não são captadas pelos modelos tradicionais de análise. Nesse contexto, os MODELOS 4 e 5, apresentados no capítulo 5, são avaliados para as escolas do Distrito de Faro. Como já salientado, esses modelos irão capturar, para uma escola específica  $i$ , os efeitos dos fatores Sexo, SSE e INDANO sobre a possibilidade de aprovação.

Os MODELOS 4 e 5 foram aplicados nas 24 escolas estudadas do Distrito de Faro. Para essas 24 escolas, o MODELO 5 foi reaplicado considerando-se, separadamente, alunos do sexo masculino e alunos do sexo feminino. Para cada uma das escolas, os alunos foram subdivididos em categorias, as quais representam os tipos padrão de aluno. O número de alunos por categoria, evidentemente, é dependente do número total de alunos na escola.

Para o MODELO 4 existem 30 categorias possíveis. Para o MODELO 5 são possíveis 15 categorias diferentes de alunos. A TABELA 6.15 mostra o número de categorias existentes por escola.

TABELA 6.15 - Número de Categorias Existentes por Escola

N. de Ordem	ESCOLA	CATEGORIAS POR MODELO (TOTAL POR ESCOLA)			
		Modelo 4	Modelo 5	Modelo 5 p/ Sexo Masc.	Modelo 5 p/ Sexo Fem.
1	08010110	23	13	12	11
2	08020104	13	9	6	7
3	08050413	25	13	12	13
4	08050506	27	14	14	13
5	08050507	22	13	12	10
6	08050513	6	6	6	Não Obs.
7	08060308	17	10	8	9
8	08060403	15	9	7	8
9	08070502	21	12	9	12
10	08070608	17	10	9	8
11	08080108	13	7	7	6
12	08080503	26	14	14	12
13	08080810	26	14	13	13
14	08080909	19	10	10	9
15	08090315	21	11	10	11
16	08100208	19	12	9	10
17	08110310	28	15	14	14
18	08110311	26	14	12	14
19	08110331	9	7	5	Não Obs.
20	08120114	20	11	9	11
21	08130206	20	12	9	11
22	08130517	23	12	11	12
23	08140607	28	15	14	14
24	08160210	26	13	13	13

Os resultados das aplicações dos MODELOS 4, 5, 5 para sexo masculino e 5 para sexo feminino são apresentados nas tabelas 6.16, 6.17, 6.18 e 6.19.

TABELA 6.16 - Resultados da Aplicação do Modelo 4

N. de Ordem	ESCOLA	INTERCEPTO ( $\alpha$ ;c)	INDANO ( $\alpha$ ;c)	SEXO ( $\alpha$ ;c)	SSE ( $\alpha$ ;c)
1	08010110	(4,7867 ; 0,9384)	(-1,6094 ; 0)	(0 ; 0)	(0,9122 ; 1,8243)
2	08020104	(2,5979 ; 0)	(-0,6945 ; 0,7457)	(-0,1493 ; 0)	(-0,1814 ; 0)
3	08050413	(9,0019 ; 1,1592)	(-2,3053 ; 0)	(-0,4090 ; 1,5643)	(-0,6359 ; 0)
4	08050506	(4,6092 ; 1,9718)	(-1,2364 ; 0)	(0,4149 ; 0,6105)	(-0,7749 ; 0)
5	08050507	(5,0484 ; 0,8750)	(-1,5677 ; 0)	(0,1543 ; 1,1293)	(0,0958 ; 0)
6	08050513	(0 ; 0,9990)	(-3,0061 ; 0,2596)	(10,5480 ; 0)	(-0,5934 ; 0)
7	08060308	(5,3013 ; 2,2951)	(-0,8366 ; 0)	(-1,0114 ; 0)	(-1,0587 ; 0)
8	08060403	(7,8377 ; 2,1474)	(-1,6856 ; 0)	(0,5481 ; 0)	(-2,3999 ; 0)
9	08070502	(4,2676 ; 1,8289)	(-1,6061 ; 0)	(-0,1771 ; 0,0201)	(0,4314 ; 0,5610)
10	08070608	(4,8977 ; 2,3894)	(-1,1982 ; 0)	(-0,0321 ; 0)	(-0,0323 ; 0,1566)
11	08080108	(4,7530 ; 2,7204)	(-1,9335 ; 0)	(1,0109 ; 0)	(0,4323 ; 0)
12	08080503	(6,9132 ; 2,2594)	(-1,4539 ; 0)	(-0,9567 ; 0,6574)	(-0,3542 ; 0)
13	08080810	(7,5927 ; 4,1705)	(-2,0907 ; 0)	(0 ; 0)	(-0,6343 ; 0,0311)
14	08080909	(5,3653 ; 0)	(-1,2413 ; 0)	(-0,7043 ; 1,1654)	(1,4647 ; 0,9282)
15	08090315	(5,9433 ; 1,5853)	(-1,6163 ; 0,3227)	(0,0602 ; 0)	(0,1771 ; 0)
16	08100208	(6,9097 ; 0)	(-1,4166 ; 0)	(-1,2043 ; 1,3499)	(0,4231 ; 0,8463)
17	08110310	(7,3590 ; 2,8328)	(-2,2247 ; 0)	(0 ; 0)	(-0,8705 ; 0,5141)
18	08110311	(6,9835 ; 1,9908)	(-2,3655 ; 0)	(-0,2533 ; 0)	(1,2872 ; 1,0706)
19	08110331	(1,6479 ; 1,0986)	(0 ; 0)	(0 ; 0)	(0 ; 0)
20	08120114	(8,3952 ; 0)	(-2,3584 ; 0,0882)	(-0,1625 ; 0)	(0,1246 ; 1,5767)
21	08130206	(4,1856 ; 1,7795)	(-1,0986 ; 0)	(0,2089 ; 0,4177)	(0 ; 0)
22	08130517	(5,8464 ; 2,5774)	(-1,4031 ; 0)	(-0,1150 ; 0)	(-0,2438 ; 0,1890)
23	08140607	(5,6292 ; 3,4783)	(-1,4751 ; 0)	(-0,0186 ; 0)	(0,1233 ; 0)
24	08160210	(7,1865 ; 0,0613)	(-2,0648 ; 0,2726)	(0,0593 ; 0,2354)	(-0,1128 ; 1,1463)

TABELA 6.17 - Resultados da Aplicação do Modelo 5

N. De Ordem	ESCOLAS	INTERCEPTO ( $\alpha$ ;c)	INDANO ( $\alpha$ ;c)	SSE ( $\alpha$ ;c)
1	08010110	(6,1274 ; 2,6504)	(-1,9459 ; 0)	(0,5377 ; 0)
2	08020104	(4,6036 ; 0)	(-1,2035 ; 0,8023)	(-0,8446 ; 0)
3	08050413	(8,0347 ; 3,6616)	(-2,5630 ; 0)	(1,1039 ; 0)
4	08050506	(5,8066 ; 2,3055)	(-1,4082 ; 0)	(-0,6517 ; 0)
5	08050507	(6,3467 ; 2,3048)	(-2,0479 ; 0)	(0,2593 ; 0,5186)
6	08050513	(10,5480 ; 0,9990)	(-3,0061 ; 0,2596)	(-0,5934 ; 0)
7	08060308	(5,4899 ; 2,1172)	(-1,5782 ; 0)	(-0,5037 ; 0)
8	08060403	(10,9849 ; 0,2301)	(2,1968 ; 0,6865)	(-3,0089 ; 0)
9	08070502	(4,3276 ; 0,1188)	(-1,0415 ; 0)	(0,5652 ; 2,4970)
10	08070608	(6,5088 ; 1,4243)	(-1,6447 ; 0)	(-0,2897 ; 1,1925)
11	08080108	(9,0964 ; 0)	(-2,1095 ; 0)	(-1,7981 ; 2,2171)
12	08080503	(6,4560 ; 1,8538)	(-1,5477 ; 0)	(-0,6754 ; 0)
13	08080810	(6,8823 ; 2,4530)	(-1,8549 ; 0)	(0,4187 ; 0,8374)
14	08080909	(5,8865 ; 1,3244)	(-1,6864 ; 0)	(1,4923 ; 1,0094)
15	08090315	(7,2066 ; 0,9346)	(-2,0872 ; 0)	(0,6118 ; 1,2236)
16	08100208	(6,5598 ; 0,7735)	(-1,6926 ; 0)	(-0,3747 ; 2,6259)
17	08110310	(8,0743 ; 2,6319)	(-2,5443 ; 0)	(-0,1724 ; 0)
18	08110311	(7,3616 ; 1,9086)	(-2,7482 ; 0)	(1,5819 ; 1,8619)
19	08110331	(1,9559 ; 1,7346)	(0 ; 0)	(0 ; 0)
20	08120114	(9,9036 ; 0)	(-2,8141 ; 0,1252)	(-0,0659 ; 1,8277)
21	08130206	(4,5694 ; 1,1661)	(-1,0944 ; 0)	(-0,1444 ; 0,7867)
22	08130517	(6,4118 ; 1,8153)	(-1,8153 ; 0)	(-0,1660 ; 0)
23	08140607	(7,4034 ; 3,2627)	(-2,0790 ; 0)	(0,3667 ; 0)
24	08160210	(7,4833 ; 0)	(-2,2169 ; 0,3929)	(0,1918 ; 0,5374)

TABELA 6.18 - Resultados da Aplicação do Modelo 5 para Alunos do Sexo Masculino

N. de Ordem	Escolas	INTERCEPTO ( $\alpha$ ;c)	INDANO ( $\alpha$ ;c)	SSE ( $\alpha$ ;c)
1	08010110	(5,8528 ; 0,9384)	(-2,1425 ; 0)	(0,9122 ; 1,8243)
2	08020104	(7,4049 ; 0,3266)	(-1,5658 ; 0,4236)	(-3,0099 ; 0)
3	08050413	(8,5929 ; 2,72360)	(-2,3053 ; 0)	(-0,6359 ; 0)
4	08050506	(4,5643 ; 0,4484)	(-1,2806 ; 0,7043)	(0,0686 ; 0)
5	08050507	(5,2028 ; 2,0043)	(-1,5677 ; 0)	(0,0958 ; 0)
6	08050513	(10,5480 ; 0,9990)	(-3,0061 ; 0,2596)	(-0,5934 ; 0)
7	08060308	(4,2247 ; 2,2757)	(-0,8001 ; 0)	(-1,1087 ; 0)
8	08060403	(8,1801 ; 0)	(-2,2580 ; 0)	(-0,8844 ; 1,4583)
9	08070502	(2,7875 ; 0)	(-1,0005 ; 0)	(1,7779 ; 1,7721)
10	08070608	(6,8648 ; 1,1309)	(-1,6380 ; 0,5161)	(-0,8860 ; 0)
11	08080108	(5,7639 ; 2,7204)	(-1,9335 ; 0)	(0,4323 ; 0)
12	08080503	(6,0535 ; 2,7068)	(-1,6629 ; 0)	(0,2633 ; 0)
13	08080810	(7,5927 ; 4,1705)	(-2,0907 ; 0)	(-0,6343 ; 0,0311)
14	08080909	(3,9118 ; 0,1702)	(-1,0759 ; 0)	(1,7755 ; 2,0028)
15	08090315	(5,3827 ; 1,7399)	(-1,6252 ; 0)	(0,8128 ; 0,5741)
16	08100208	(3,6707 ; 0)	(-0,8757 ; 0,1402)	(0,9898 ; 1,3364)
17	08110310	(7,5932 ; 2,9427)	(-2,3345 ; 0)	(-0,8448 ; 0,3342)
18	08110311	(6,6284 ; 2,1944)	(-2,3655 ; 0)	(1,5780 ; 0,4890)
19	08110331	(1,0986 ; 0)	(0 ; 0)	(0,5086 ; 1,0172)
20	08120114	(8,7839 ; 0)	(-2,4268 ; 0,1841)	(-0,3677 ; 1,0820)
21	08130206	(3,9318 ; 1,2720)	(-0,9830 ; 0,2313)	(0 ; 0)
22	08130517	(6,0131 ; 2,6415)	(-1,4694 ; 0)	(-0,4120 ; 0)
23	08140607	(5,5662 ; 3,2774)	(-1,4937 ; 0)	(0,2068 ; 0,1698)
24	08160210	(6,6454 ; 0)	(-2,0153 ; 0,1736)	(0,3138 ; 1,6739)

TABELA 6.19 - Resultados da Aplicação do Modelo 5 para Alunos do Sexo Feminino

N. de Ordem	Escola	INTERCEPTO ( $\alpha$ ;c)	INDANO ( $\alpha$ ;c)	SSE ( $\alpha$ ;c)
1	08010110	(6,4765 ; 1,7216)	(-1,8338 ; 0)	(-0,6536 ; 0)
2	08020104	(3,8217 ; 0)	(-1,0029 ; 0,1904)	(-0,7529 ; 1,5057)
3	08050413	(8,1516 ; 2,8046)	(-2,7061 ; 0)	(0,7975 ; 0,0507)
4	08050506	(5,6017 ; 0,4746)	(-2,1223 ; 0)	(0,7240 ; 1,5689)
5	08050507	(9,6842 ; 0,3789)	(-2,4673 ; 0,0749)	(-1,5468 ; 1,1664)
7	08060308	(4,9979 ; 0)	(-1,3459 ; 0,2351)	(-1,2213 ; 0,7171)
8	08060403	(8,0315 ; 1,5408)	(-1,3823 ; 0)	(-2,4104 ; 0)
9	08070502	(4,6495 ; 0,9771)	(-1,1511 ; 0,1050)	(0,0513 ; 0,9265)
10	08070608	(5,7369 ; 1,8390)	(-1,2933 ; 0)	(-0,7297 ; 0,2822)
11	08080108	(4,5751 ; 0)	(-1,8655 ; 0)	(2,3919 ; 1,8802)
12	08080503	(6,1986 ; 1,1763)	(-1,4539 ; 0)	(-1,2764 ; 1,8445)
13	08080810	(5,4700 ; 1,9451)	(-1,3719 ; 0)	(0,0927 ; 1,0092)
14	08080909	(6,0354 ; 2,3094)	(-1,9851 ; 0)	(0,8126 ; 0,5499)
15	08090315	(5,5504 ; 0,5586)	(-1,4880 ; 0,5794)	(0,1771 ; 0)
16	08100208	(5,4250 ; 0)	(-1,7956 ; 0,7579)	(0,6204 ; 1,2408)
17	08110310	(6,9336 ; 2,7932)	(-2,2192 ; 0)	(0 ; 0)
18	08110311	(6,5384 ; 1,0814)	(-2,4238 ; 0)	(1,3975 ; 1,7940)
20	08120114	(7,8969 ; 0)	(-2,1519 ; 0)	(-0,2810 ; 1,6553)
21	08130206	(2,6229 ; 0)	(-0,7621 ; 0)	(1,2106 ; 2,4213)
22	08130517	(3,8501 ; 1,1086)	(-0,9625 ; 0,2722)	(0 ; 0)
23	08140607	(6,3919 ; 2,8387)	(-1,6094 ; 0)	(-0,2976 ; 0,2287)
24	08160210	(6,4942 ; 0,3879)	(-1,6706 ; 0,1554)	(-0,5277 ; 1,2638)

O exame da Tabela 6.17 mostra que, à exceção de algumas poucas escolas, a influência do fator INDANO é imediata sobre a possibilidade de aprovação do aluno. Os resultados nulos observados para os valores das

dispersões dos coeficientes vagos INDANO podem ser explicados pelo fato das repetências ocorrerem com maior frequência para aqueles alunos classificados como sendo "muito defasados". Os valores para o coeficiente relacionado à variável SSE e o fato desses apresentarem elevadas dispersões pode levar a seguinte reflexão: nas escolas em que tal fenômeno ocorre é necessário uma atuação mais efetiva da escola no sentido de interagir e inserir-se no meio sócio-econômico que a rodeia. A escola e sua estratégia de atuação deve ser o vetor a promover tal integração. O fenômeno é mais acentuado em escolas nas quais se verifica uma maior heterogeneidade de grupos sócio-econômicos.

Resultados similares são encontrados nos demais modelos avaliados. Uma palavra adicional sobre a TABELA 6.19: enquanto a componente relativa à variável INDANO se mantém essencialmente com o mesmo padrão de comportamento, a variável SSE apresenta dispersões em maior quantidade e intensidade, levando a concluir que as alunas sofrem um maior efeito do fator sócio-econômico na possibilidade de aprovação.

No próximo item, esses resultados obtidos serão utilizados para estimar o modelo difuso hierárquico, apresentado no capítulo 5. As discussões a respeito dos resultados obtidos pela aplicação dos MODELOS 4 e 5 são apresentadas no próximo capítulo, o qual inclui, também, uma análise comparativa entre os modelos baseados na lógica tradicional e os modelos baseados na lógica difusa.

## **6.6 - Modelo Difuso Hierárquico - Aplicação do MODELO 6**

Os MODELOS 4 e 5 descrevem, para uma escola particular, como características individuais de alunos podem contribuir, na avaliação da aprovação dos mesmos.

Para diferentes escolas, as características das instalações, o nível de preparação de professores e educadores, a qualidade atribuída pelos alunos à merenda escolar, a qualidade das instalações escolares, etc., apresentam efeitos diferenciados e podem contribuir de forma diversa na avaliação da aprovação dos estudantes. Esses efeitos são vagos e imprecisos. A teoria dos conjuntos difusos permite o tratamento de tais fenômenos, provendo ferramental teórico capaz de tratar com esse tipo de imprecisão e nebulosidade. Assim, a contribuição de cada uma das características individuais dos alunos agrega uma informação difusa e imprecisa, derivada do efeito contextual da escola. Esse efeito é modelado, através de uma análise hierarquizada, baseada na lógica difusa, considerando-se as mesmas variáveis (Estado de Conservação (EC); Localização (LOC); Adequação das Salas (ADS) e Taxa de Ocupação (To2)), apresentadas anteriormente.

Os seguintes modelos foram ajustados:

$$A_{ji} = B_{1j} \text{ADS}_i + B_{2j} \text{EC}_i + B_{3j} \text{LOC}_i + B_{4j} \text{To2}_i$$

com  $j = 1, \dots, 4; i = 1, \dots, N$ . de Escolas

onde:  $A_{ji}$  = coeficientes difusos obtidos pela aplicação do MODELO 4

$\text{ADS}_i$  = Adequação das salas de aula

$\text{EC}_i$  = Estado de Conservação dos Estabelecimentos

$\text{LOC}_i$  = Localização dos Estabelecimentos Escolares

$\text{To2}_i$  = Taxa de Ocupação

Os resultados para os novos coeficientes difusos são apresentados na TABELA 6.20.

TABELA 6.20 - Resultados da Aplicação do MODELO 6

	ADS	EC	LOC	To2
Intercepto ( $A_{1i}$ )	(1.3428; 1.6302)	0.1422; 0.4001)	(0.4348; 0.4486)	(0.5945; 0)
IndAno ( $A_{2i}$ )	(-0.4443; 0.9075)	(-0.2467; 0.0712)	(-0.1895; 0)	(-0.053; 0)
SSE ( $A_{3i}$ )	(-0.1816; 0.2991)	(0.226; 0.0268)	(0.4151; 0,5514)	(-0.7414; 0)
Sexo ( $A_{4i}$ )	(0.059; 0.5422)	(0.1203; 0.411)	(-0.1214; 0)	(-0.2448; 0)

Nota-se a influência bastante nítida da variável To2 em todas as regressões avaliadas.

Para o caso do coeficiente relativo à variável INDANO, a contribuição relativa à adequação das salas apresenta um grau de imprecisão elevado. Tal fato pode indicar um "desajuste" dos alunos às condições oferecidas pelas escolas. Esse ponto é corroborado pelo sinal negativo do coeficiente, conduzindo a possibilidades menores de aprovação para valores de INDANO crescentes.

Com relação às variáveis Estado de Conservação, Localização e Taxa de Ocupação, notam-se, também, os efeitos negativos. Esses efeitos são observados em quase todas as características básicas dos estudantes. Ao contrário dos modelos tradicionais, onde o critério de agregação se dá pelo valor médio da distribuição de probabilidades, em modelos difusos, o operador de agregação (em geral operadores do tipo compensatórios) leva em conta uma quantidade maior de informação. Portanto, esse delta adicional de informação agregado pelos operadores definidos na lógica difusa, explica os resultados descritos na TABELA 6.20. Com relação a variável Estado de Conservação é interessante notar o efeito positivo sobre a variável Sexo do Aluno.

Recombinando-se os modelos é possível avaliar os efeitos derivados das características das escolas sobre a possibilidade de aprovação. Tais efeitos são da seguinte forma:

#### Adequação das Salas de Aula:

$$(B_{11} + B_{12} \text{IndAno} + B_{13} \text{SSE} + B_{14} \text{Sexo})$$

**Estado de Conservação:**

$$(B_{21} + B_{22} \text{IndAno} + B_{23} \text{SSE} + B_{24} \text{Sexo})$$

**Localização:**

$$(B_{31} + B_{32} \text{IndAno} + B_{33} \text{SSE} + B_{34} \text{Sexo})$$

**Taxa de Ocupação:**

$$(B_{41} + B_{42} \text{IndAno} + B_{43} \text{SSE} + B_{44} \text{Sexo})$$

Nessas recombinações, os coeficientes são considerados números difusos, diferindo, portanto, dos resultados obtidos pela aplicação de modelos hierarquizados tradicionais.

**6.7 - Processos para Classificação das Escolas**

Modelos de avaliação de desempenho, objetivam fornecer mecanismos para avaliar a qualidade das resposta de um sistema. Quando o sistema é composto por vários elementos, o seu desempenho passa a ser dependente do desempenho de cada um desses elementos individuais. Nesse caso, é interessante dispor de processos que permitam realizar uma classificação ou hierarquização com relação a determinadas metas e objetivos a serem alcançados.

Acompanhar o desempenho relativo das escolas fornece aos gestores do sistema educacional importantes instrumentos de gestão. Neste item, são apresentados dois processos para realização de classificação de escolas, baseados em indicadores de desempenho. O critério adotado, foi o desempenho das escolas em termos da aprovação de alunos. Os processos de classificação e de hierarquização foram discutidos no capítulo 5. O primeiro, baseia-se na

aplicação de modelos taxionômicos sobre os resultados obtidos pela aplicação de modelos tradicionais de análise. O segundo utiliza a lógica difusa no seu desenvolvimento, bem como, os resultados dos modelos difusos estimados.

### **6.7.1 - Análise Taxinômica: Classificação das Escolas quanto as Probabilidades de Aprovação**

O uso de técnicas de agrupamento, como já referido, é dependente de contexto e sujeito a critérios subjetivos. Tais critérios são utilizados, por exemplo, para definir o número de grupos que se pretende construir. Utilizando-se os resultados obtidos pela aplicação do MODELO 1, é possível, através das técnicas de agrupamento, identificar escolas com coeficientes similares e, conseqüentemente, probabilidades de aprovação similares. Com a utilização de software estatístico convencional, chegou-se ao agrupamento descrito na TABELA 6.21, para as vinte e quatro escolas analisadas.

TABELA 6.21 - Agrupamento das Escolas Analisadas-Coefficientes do MODELO 1

	ESCOLA	CLUSTER
1	08010110	5 (E)
2	08020104	5 (E)
3	08050413	5 (E)
4	08050506	5 (E)
5	08050507	5 (E)
6	08050513	1 (A)
7	08060308	5 (E)
8	08060403	2 (B)
9	08070502	5 (E)
10	08070608	5 (E)
11	08080108	3 (C)
12	08080503	5 (E)
13	08080810	5 (E)
14	08080909	4 (D)
15	08090315	5 (E)
16	08100208	4 (D)
17	08110310	5 (E)
18	08110311	5 (E)
19	08110331	6 (F)
20	08120114	5 (E)
21	08130206	5 (E)
22	08130517	5 (E)
23	08140607	5 (E)
24	08160210	5 (E)

A visualização gráfica desse agrupamento é apresentada na FIGURA 6.4.

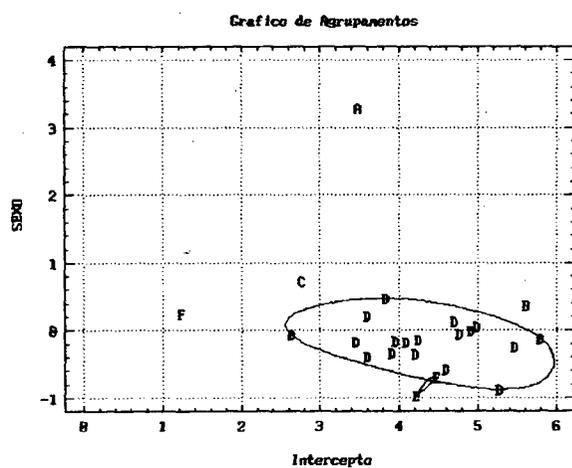
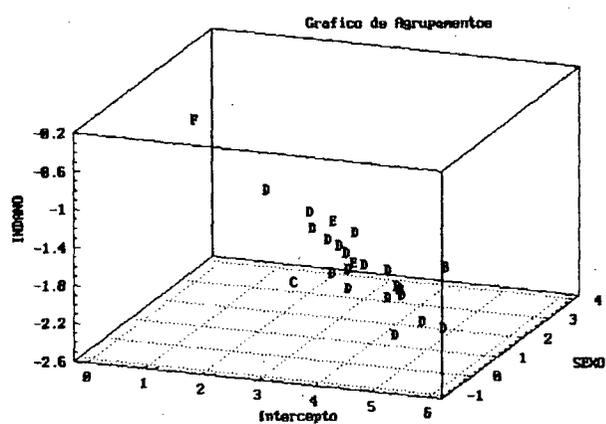
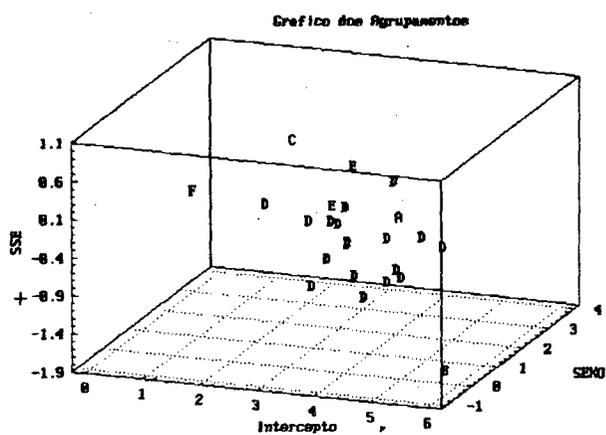


FIGURA 6.4 - Agrupamento Realizado - Coeficientes do MODELO 1

As escolas "08050513", "08060403", "08080108" e "08110331" apresentam padrão singular. As escolas "08080909" e "08100208" apresentam padrões de comportamentos similares entre si. Já as demais escolas, podem ser

consideradas como constituindo um único grupo. Esse resultado é ilustrado na FIGURA 6.5.

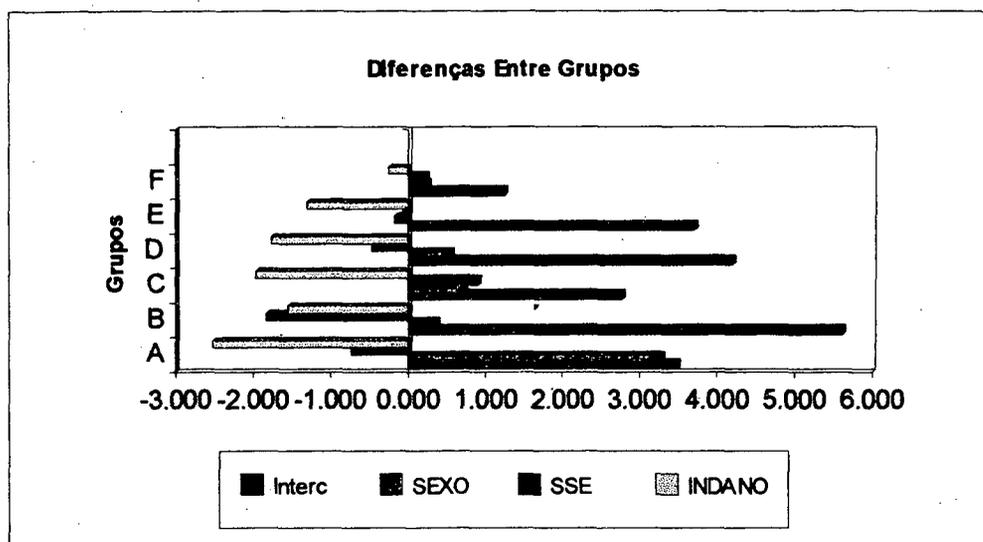


FIGURA 6.5 - Diferenças Entre os Grupos Determinados através da Análise de Agrupamentos

### 6.7.2 - Classificação de Escolas Utilizando Resultados dos Modelos Difusos

Os dados obtidos a partir dos modelos difusos, estimados anteriormente, fornecem, em termos numéricos, a possibilidade de aprovação quando considerados valores relativos as características individuais de alunos. Para cada tipo de aluno, as diversas escolas apresentam possibilidades de aprovação diferentes. A questão que se coloca é: "Como classificar as escolas de acordo com os resultados numéricos obtidos através dos modelos 4 e 5"? Aplicando-se o algoritmo proposto em 5.5, para os resultados estimados pelo MODELO 4, obtém-se, inicialmente, as classificações apresentadas nas tabelas 6.22, 6.23 e 6.24.

TABELA 6.22 - Classificação Escolas - Modelo 4 - Limite Inferior

Lim. Inf.	Aluno Padrão - Critérios															Média
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Escolas																
08010110	19	20	12	11	10	20	19	16	12	11	19	18	16	12	11	15,07
08020104	15	14	18	16	15	19	21	20	20	19	20	21	21	19	19	18,47
08050413	24	21	10	8	6	24	23	21	18	12	23	23	22	20	17	18,13
08050506	16	17	20	18	17	15	14	13	11	13	12	10	8	9	9	13,47
08050507	23	23	22	21	18	17	17	17	15	15	14	13	10	10	10	16,33
08050513	4	3	2	2	1	13	11	8	7	8	18	20	19	18	18	10,13
08060308	21	22	21	23	21	18	18	19	19	17	16	17	15	15	15	18,47
08060403	2	2	4	6	12	3	4	11	17	21	7	12	18	21	22	10,80
08070502	7	10	19	20	22	4	6	9	10	16	4	3	5	6	7	9,87
08070608	10	11	16	17	20	12	12	14	16	18	9	9	11	11	16	13,47
08080108	3	5	9	19	23	2	3	12	21	23	3	5	12	17	21	11,87
08080503	11	8	8	9	9	8	9	6	6	7	6	7	6	5	4	7,27
08080810	14	13	13	13	14	16	16	15	14	14	17	16	13	13	13	14,27
08080909	8	7	5	4	2	10	5	3	2	1	8	6	2	2	2	4,47
08090315	6	6	6	5	5	6	7	4	4	5	11	8	7	7	8	6,33
08100208	9	15	24	24	24	11	15	23	24	24	10	14	20	22	23	18,80
08110310	17	18	17	14	13	22	22	22	22	20	22	22	23	23	20	19,80
08110311	22	24	23	22	19	23	24	24	23	22	24	24	24	24	24	23,07
08110331	18	19	14	12	11	5	2	2	1	2	1	1	1	1	1	6,07
08120114	1	1	1	1	4	1	1	1	3	3	2	2	3	3	5	2,13
08130206	13	12	15	15	16	7	10	7	9	9	5	4	4	4	3	8,87
08130517	12	9	7	7	7	14	13	10	8	6	13	11	9	8	6	9,33
08140607	20	16	11	10	8	21	20	18	13	10	21	19	17	16	12	15,47
08160210	5	4	3	3	3	9	8	5	5	4	15	15	14	14	14	8,07

TABELA 6.23 - Classificação das Escolas - Modelo 4 - Centróide

Escolas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Média
08010110	15	15	12	10	11	19	19	17	14	11	19	18	14	15	14	14,87
08020104	24	24	24	24	24	22	22	24	24	23	14	16	15	17	19	21,07
08050413	7	5	3	3	3	20	16	8	6	4	22	21	17	14	10	10,60
08050506	16	19	20	21	21	9	14	19	19	19	8	9	12	12	15	15,53
08050507	20	20	19	17	14	16	17	10	9	10	12	10	7	7	7	13,00
08050513	4	3	4	6	8	12	15	18	17	18	21	22	23	23	23	14,47
08060308	22	22	23	23	23	21	21	23	23	22	18	19	21	20	20	21,40
08060403	1	2	9	15	22	1	2	11	21	24	2	5	19	22	24	12,00
08070502	18	18	14	11	10	8	7	4	4	3	4	3	3	3	3	7,53
08070608	11	11	13	12	12	6	10	9	10	13	9	12	8	10	11	10,47
08080108	3	4	11	16	19	3	5	16	20	21	5	11	16	19	21	12,67
08080503	10	14	17	19	17	5	11	13	16	16	7	7	9	11	13	12,33
08080810	6	8	7	8	6	7	6	5	5	5	13	8	6	5	5	6,67
08080909	13	6	2	2	2	11	4	2	2	1	10	4	4	2	2	4,47
08090315	8	9	5	5	5	15	9	6	7	7	17	15	10	8	8	8,93
08100208	12	13	16	14	15	10	12	14	15	15	11	13	13	13	12	13,20
08110310	14	16	15	13	13	23	23	22	22	20	23	23	24	24	22	19,80
08110311	19	12	8	4	4	24	24	20	11	9	24	24	22	21	17	16,20
08110331	23	23	22	22	20	4	3	3	3	6	1	1	1	1	1	8,93
08120114	2	1	1	1	1	2	1	1	1	2	3	2	2	4	4	1,87
08130206	21	21	21	20	18	14	13	12	12	14	6	6	5	6	6	13,00
08130517	17	17	18	18	16	18	20	21	18	17	16	17	18	16	16	17,53
08140607	5	7	6	7	7	13	8	7	8	8	15	14	11	9	9	8,93
08160210	9	10	10	9	9	17	18	15	13	12	20	20	20	18	18	14,53

TABELA 6.24 - Classificação Escolas - Modelo 4 - Limite Superior

Escolas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Média
08010110	8	9	11	11	12	8	11	13	12	12	13	16	17	15	15	12,20
08020104	24	24	24	24	24	17	22	24	24	23	3	6	10	12	14	18,33
08050413	2	1	1	2	2	3	1	2	3	7	10	9	9	7	8	4,47
08050506	10	16	17	17	19	7	12	16	16	17	6	12	15	16	17	14,20
08050507	14	15	15	12	11	9	8	9	10	10	7	8	8	8	9	10,20
08050513	6	7	14	15	15	13	19	21	22	21	23	24	24	24	24	18,13
08060308	19	21	22	22	22	19	23	23	23	22	17	19	22	22	21	21,13
08060403	1	6	16	21	23	1	4	14	20	24	1	2	14	18	22	12,47
08070502	23	19	9	9	7	22	13	8	6	2	14	7	2	1	1	9,53
08070608	13	12	10	10	10	11	10	10	11	11	11	13	11	11	11	11,00
08080108	9	13	13	13	13	12	15	15	14	14	18	17	18	17	16	14,47
08080503	12	17	19	20	20	10	16	18	19	19	9	15	19	19	19	16,73
08080810	4	3	4	4	5	4	3	3	5	6	5	3	4	5	5	4,20
08080909	16	8	6	5	4	15	6	4	4	4	16	10	5	2	2	7,13
08090315	17	14	8	8	9	21	17	11	9	9	22	18	16	14	12	13,67
08100208	15	10	7	7	6	14	7	7	7	5	15	11	6	6	4	8,47
08110310	7	11	12	14	14	16	18	19	18	18	20	23	23	23	23	17,27
08110311	11	5	2	1	1	23	14	6	2	1	24	22	13	10	6	9,40
08110331	22	23	23	23	21	6	9	12	13	13	2	1	1	4	7	12,00
08120114	5	4	3	3	3	5	5	1	1	3	8	4	3	3	3	3,60
08130206	21	22	21	18	17	20	20	17	15	15	12	14	12	13	13	16,67
08130517	18	18	20	19	18	18	21	22	21	20	19	20	21	21	20	19,73
08140607	3	2	5	6	8	2	2	5	8	8	4	5	7	9	10	5,60
08160210	20	20	18	16	16	24	24	20	17	16	21	21	20	20	18	19,40

Essas classificações podem ser normalizadas no intervalo  $[0,1]$ . O mesmo procedimento sendo aplicado aos demais modelos estimados conduz aos resultados mostrados nas TABELAS 6.25 e 6.26.

TABELA 6.25 - Classificações Normalizadas por Escolas - Modelos 4 e 5

Escolas	Modelo 4			Modelo 5		
	Mínimo	Centroide	Máximo	Mínimo	Centroide	Máximo
08010110	0.43189369	0.4418146	0.43099788	0.38216561	0.33447099	0.5095057
08020104	0.55315615	0.17159763	0.02760085	0.21974522	0.01706485	0.15969582
08050413	0.27906977	0.47534517	0.7133758	0.23566879	0.55290102	0.95057034
08050506	0.34883721	0.34319527	0.3970276	0.45859873	0.3003413	0.39543726
08050507	0.47342193	0.28994083	0.16348195	0.32165605	0.43003413	0.62357414
08050513	0.92857143	0.86193294	0.63269639	0.61783439	0.35494881	0.17110266
08060308	0.41196013	0.18343195	0	0.21974522	0	0
08060403	0.80730897	0.67455621	0.28450106	0.58598726	0.48122867	0.49429658
08070502	0.7641196	0.6765286	0.39065817	0.63057325	0.70989761	0.66159696
08070608	0.69933555	0.69428008	0.3970276	0.45859873	0.55972696	0.57794677
08080108	0.48172757	0.48717949	0.38428875	0.53503185	0.44709898	0.38022814
08080503	0.27242525	0.41420118	0.57324841	0.75477707	0.46416382	0.25095057
08080810	0	0.234714	0.83864119	0.42038217	0.75426621	0.96577947
08080909	0.84717608	1	0.84501062	0.88853503	0.8668942	0.79847909
08090315	0.77574751	0.80276134	0.55626327	0.79936306	0.63822526	0.42585551
08100208	0.60465116	0.67061144	0.6836518	0.20382166	0.41979522	0.72243346
08110310	0.06478405	0	0.1656051	0.15605096	0.08191126	0.22053232
08110311	0.35714286	0.44970414	0.49044586	0	0.2662116	0.66920152
08110331	1	0.7495069	0.15286624	0.81210191	0.63822526	0.52091255
08120114	0.87043189	0.64891519	0.13375796	1	1	1
08130206	0.71428571	0.62130178	0.38428875	0.67834395	0.43003413	0.25475285
08130517	0.60963455	0.70414201	0.59023355	0.65605096	0.19795222	0.07984791
08140607	0.33388704	0.75542406	1	0.36305732	0.63822526	0.88593156
08160210	0.6461794	0.52465483	0.21019108	0.71656051	0.35153584	0.09885932

TABELA 6.26 - Classificações Normalizadas por Escola - Modelos 5 p/ Sexo Masculino e 5 p/ Sexo Feminino

Escolas	Modelo 5 p/ Sexo Masculino			Modelo 5 p/ Sexo Feminino		
	Mínimo	Centroide	Máximo	Mínimo	Centroide	Máximo
08010110	0.17434211	0.15185185	0.39325843	0.37596899	0.11721612	0.16780822
08020104	0.39144737	0.16296296	0.04494382	0.11627907	0.08424908	0.19178082
08050413	0.76315789	0.93703704	0.93632959	0.19379845	0.43223443	0.62328767
08050506	0.37828947	0.31481481	0.50561798	0.05813953	0.03296703	0.1609589
08050507	0.49013158	0.24074074	0.16853933	0.49612403	0.35897436	0.30136986
08050513	0.67763158	0.5	0.29213483			
08060308	0.44407895	0.35555556	0.34082397	0.26356589	0	0
08060403	0.67105263	0.27777778	0	0.84496124	0.6043956	0.43493151
08070502	0.875	0.66296296	0.46441948	0.58527132	0.54578755	0.51027397
08070608	0.4375	0.54074074	0.60299625	0.5620155	0.50915751	0.45890411
08080108	0.18421053	0.15555556	0.3670412	0.74806202	0.58974359	0.47260274
08080503	0.52960526	0.68888889	0.74531835	0.04263566	0.25274725	0.56164384
08080810	0	0.34074074	0.98127341	0.24806202	0.71794872	0.80821918
08080909	1	1	0.83520599	0.49224806	0.75091575	0.86643836
08090315	0.71710526	0.67037037	0.61048689	0.54263566	0.58241758	0.54794521
08100208	0.99342105	0.8962963	0.62172285	0	0.28571429	0.70205479
08110310	0.00328947	0	0.3670412	0.18604651	0.48717949	0.70890411
08110311	0.5	0.55925926	0.65543071	0.07751938	0.34432234	0.54452055
08110331	0.95723684	0.64074074	0.26217228			
08120114	0.81907895	0.61481481	0.25093633	0.74031008	0.54578755	0.35273973
08130206	0.67434211	0.43703704	0.28464419	0.3372093	0.9010989	0.97945205
08130517	0.48026316	0.62222222	0.6329588	0.76744186	1	1
08140607	0.25328947	0.65185185	1	1	0.63736264	0.15410959
08160210	0.57236842	0.44444444	0.3670412	0.48837209	0.41391941	0.41438356

Como já citado, um critério global de classificação poderia ser explicitado através de uma variável lingüística, com os rótulos: abaixo da média, na média e acima da média.

Os responsáveis pela gestão do Sistema Educacional poderiam propor as funções de pertinência, para essa variável, que descrevam a classificação das escolas quanto a possibilidade de aprovação, apresentadas na FIGURA 6.6

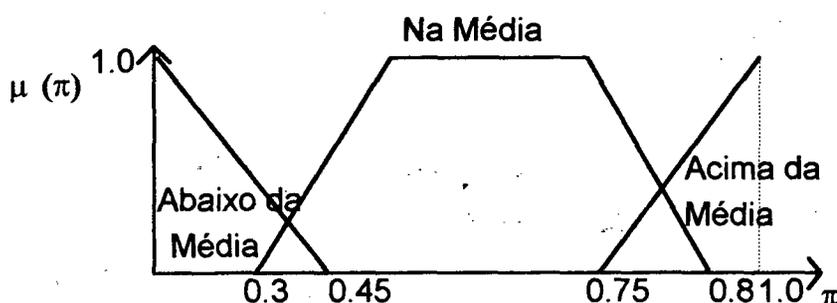


FIGURA 6.6 - Um Critério para Classificação das Escolas quanto a Possibilidade de Aprovação

A cada uma das classificações descritas nas TABELAS 6.25 e 6.26 pode ser atribuído o grau de pertinência, para o qual, a escola é classificada em abaixo da média, na média ou acima da média. Um cálculo direto da média difusa das classificações e seu ordenamento, conduzem a uma hierarquização das escolas. Esse processo de ordenamento tem como principal vantagem a inclusão, quando de sua realização, de conceitos vagos ou imprecisos. Em processos tradicionais, baseados na Lógica Clássica, o corte é arbitrado e não ocorre de forma gradual, suscitando, assim, controvérsias difíceis de serem sustentadas.

Utilizando-se os valores dos modelos nos quais as vinte e quatro escolas foram observadas, chega-se às hierarquizações, apresentadas na TABELA 6.27

Um refinamento adicional, seria o processamento de uma análise de agrupamentos, de modo a identificar as escolas com padrões semelhantes. Realizando-se essa análise, obtêm-se os quatro grupos descritos na TABELA 6.28. A FIGURA 6.7 apresenta o resultado obtido.

TABELA 6.27 - Hierarquizações Obtidas para as Escolas Analisadas

Escolas	Modelo 4	Modelo 5	Modelo 5 p/ Sexo Masc
08010110	17	18	22
08020104	22	23	23
08050413	12	8	2
08050506	19	19	17
08050507	21	12	20
08050513	3	17	13
08060308	23	24	18
08060403	11	11	19
08070502	6	5	5
08070608	7	10	12
08080108	15	15	21
08080503	16	9	6
08080810	20	3	16
08080909	1	2	1
08090315	4	4	7
08100208	8	14	3
08110310	24	22	24
08110311	18	21	10
08110331	5	6	8
08120114	13	1	11
08130206	10	13	14
08130517	9	20	9
08140607	2	7	4
08160210	14	16	15

TABELA 6.28 - Agrupamento das Escolas - Possibilidade de Aprovação

Escola	Grupo	Escola	Grupo	Escola	Grupo
08010110	A	08070502	D	08110310	A
08020104	A	08070608	B	08110311	A
08050413	B	08080108	A	08110331	D
08050506	A	08080503	B	08120114	C
08050507	A	08080810	C	08130206	B
08050513	B	08080909	D	08130517	B
08060308	A	08090315	D	08140607	D
0860403	B	08100208	B	08160210	A

O grupo A constitui-se das escolas que obtiveram uma classificação baixa em todos os modelos estudados. Escolas inseridas no grupo B podem ser consideradas escolas intermediárias, enquanto que as escolas enquadradas no grupo E são aquelas que, em termos dos modelos analisados, apresentam

melhor desempenho relativo. O grupo C é constituído de apenas duas escolas. Essas escolas apresentaram classificações heterogêneas entre os modelos analisados.

O processo utilizado para obtenção de uma classificação relativa das escolas pode ser considerado eficiente. Em apenas 8.3% dos casos, o mecanismo de ordenação adotado não conseguiu discriminar as escolas analisadas.

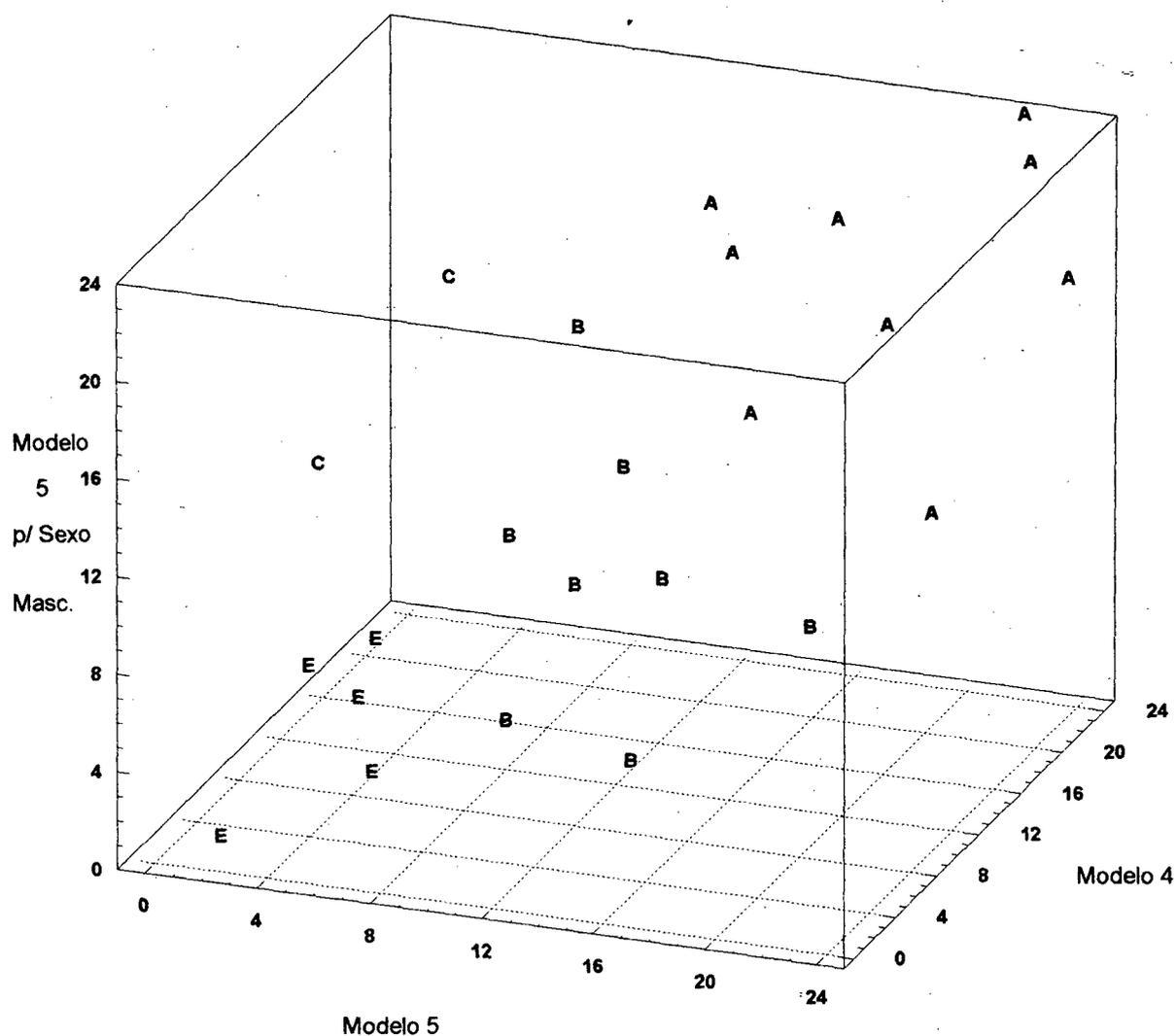


FIGURA 6.7 - Agrupamento das Escolas - Possibilidade de Aprovação

## 6.8 Conclusões

Neste capítulo foram realizados diversos ajustes de modelos para análise de sistemas educacionais.

Os modelos construídos utilizam análise de regressão tradicional e análise de regressão possibilística. Tendo-se, como base, as taxas de aprovação verificadas em escolas do Distrito de Faro - Portugal, os modelos avaliam o desempenho das escolas considerando dois níveis de informação: características dos estudantes e características das escolas. As características básicas dos estudantes examinadas foram sexo, idade, ano de escolaridade do aluno, repetências anteriores e indicador de "status" sócio econômico do aluno. Como características das escolas que influenciam no rendimento escolar foram examinadas: localização do estabelecimento de ensino, estado de conservação, adequação das salas de aula e taxas de ocupação das salas de aula.

A exceção do fator INDANO, não existe, em termos macro, predominância de algum efeito para o grupo das escolas analisadas. Assim, políticas de gestão devem ser direcionadas aos estabelecimentos de ensino, procurando sanar problemas localizados. Essa tendência é observada tanto nos modelos estimados através de regressões tradicionais quanto naqueles estimados mediante regressões possibilísticas.

Em todos os modelos testados, o coeficiente do fator INDANO, que expressa uma relação entre ano de escolaridade, idade do aluno e número de repetências anteriores, foi negativo. Considerando a definição dada para a variável INDANO, conclui-se pela criação de turmas específicas, as quais, permitam um atendimento diferenciado para aqueles alunos. Dessa forma, espera-se que os resultados venham a garantir uma maior eficiência global para o sistema educativo.

Com relação à classificação das escolas destaca-se que poucas apresentaram desempenho diferente do restante do grupo. Um exemplo é a escola 08110311. Através das classificações propostas para as escolas, estratégias de atuação e interação podem ser delineadas com o intuito de promover melhorias desejadas.

## 7. ANÁLISE COMPARATIVA DOS MODELOS

### 7.1 - Introdução

No capítulo anterior foram estimados modelos para avaliar o desempenho de escolas do Distrito de Faro, Portugal, tendo-se como base as taxas de aprovação verificadas naquelas escolas.

Os modelos desenvolvidos incluem características individuais básicas dos alunos. Foram ajustados, também, modelos que incluem informações estruturais e de contexto das escolas, seguindo uma abordagem multinível ou hierarquizada para avaliação do desempenho do sistema.

As técnicas regressão utilizadas foram de duas formas. A primeira está alicerçada na teoria estatística tradicional, segundo a qual, os parâmetros dos modelos seguem a um determinado modelo de distribuição de probabilidade. Nesse caso, modelos do tipo logit foram adotados para descrever as probabilidades de aprovação verificadas em cada uma das escolas. A segunda forma, utiliza de conceitos oriundos da teoria de conjuntos difusos, ajustando-se um modelo de regressão possibilística, para um sistema de equações possibilísticas. Nesse caso, a hipótese fundamental é bastante distinta dos modelos convencionais. Os parâmetros do sistema não seguem uma distribuição de probabilidade admitida conhecida. Ao contrário, esses coeficientes são números difusos e, portanto, são modelados através de uma distribuição de possibilidade, conforme definido por Zadeh. A técnica de regressão possibilística, ao contrário da técnica de regressão tradicional, permite a modelação de variáveis lingüísticas como variáveis dependentes. Contudo, alguns cuidados na obtenção e coleta dos dados são necessários. Uma aplicação, utilizando o sistema computacional desenvolvido, é apresentada para exemplificar esse tipo de modelagem.

Aplicações dos modelos estimados são, também, realizadas com a intenção de avaliar e comparar os resultados obtidos.

## 7.2 - Aplicações dos Modelos de Regressão Estatística Tradicional

Admita-se que um aluno, com as características dadas a seguir, deseja se matricular em uma das escolas do Concelho de Loulé:

**Idade:** 13 Anos;

**Ano Escolar:** 7º Ano;

**Repetências Anteriores:** Nenhuma;

**Sexo:** Feminino;

**Ocupação principal do Responsável:** Operário Não Qualificado.

De acordo com os resultados obtidos pela aplicação do MODELO 1, para esse aluno obter-se-iam as probabilidades de aprovação, apresentados na TABELA 7.1.

TABELA 7.1 - Probabilidades de Aprovação - MODELO 1

Escola	Probabilidades (Intervalo de 95%)		
	Limite Inferior	Valor Médio	Limite Superior
8080108	0.000298	0.946843	0.999999
8080503	0.116207	0.935483	0.999375
8080810	0.037719	0.945917	0.999872
8080909	0.038566	0.920717	0.999703

Para o MODELO 3, considerando-se os efeitos contextuais das escolas, o aluno exemplo apresenta as probabilidades de aprovação mostradas na TABELA 7.2.

TABELA 7.2 - Probabilidades de Aprovação - MODELO 3

Modelo 3	
Escola	Probabilidades
8080108	Não Obs.
8080503	0.9286
8080810	0.9427
8080909	0.9475

### 7.3 - Aplicações dos Modelos de Regressão Possibilística

De acordo com os resultados obtidos pela aplicação do MODELO 4, o aluno considerado como exemplo apresentaria as funções de pertinência relativas à possibilidade de aprovação, mostradas na FIGURA 7.1, onde:

a: 0.485687

c: -1.726598

e: -1.702826

g: -0.486515

b: 5.926487

d: 5.421802

f: 6.681092

h: 5.456001

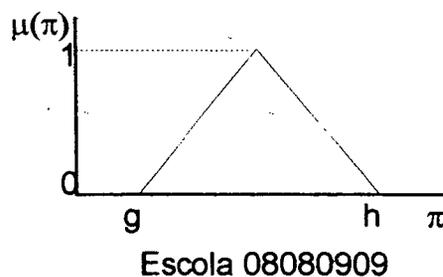
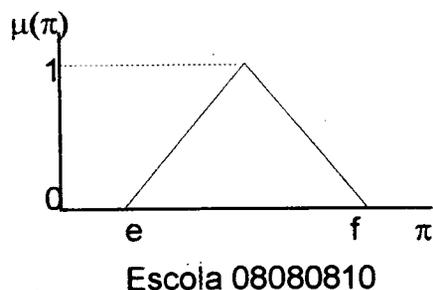
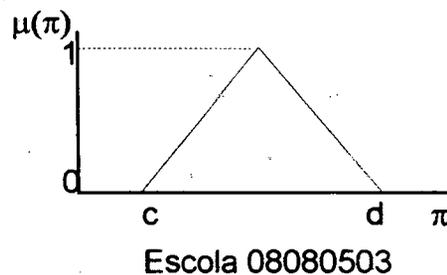
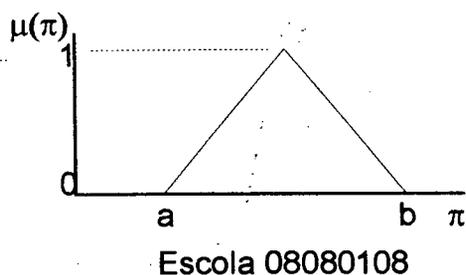


FIGURA 7.1 - Funções de Pertinência - Possibilidade de Aprovação

Estabelecendo um corte  $\alpha = 0.8$  (aprovação com grau de pertinência superior a 0.8), tem-se os valores mostrados na FIGURA 7.2, onde:

i: 2.662007

k: 1.332762

m: 1.507412

o: 1.8904914

j: 3.750167

l: 2.562442

n: 3.327524

p: 3.0789946

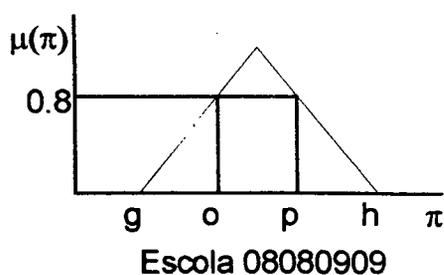
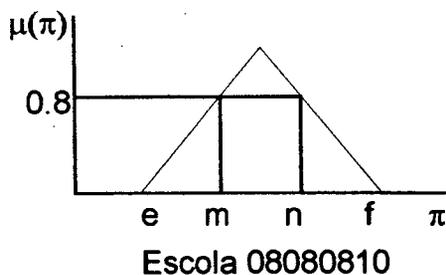
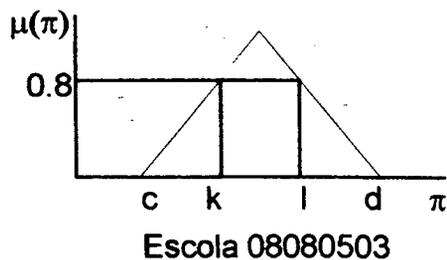
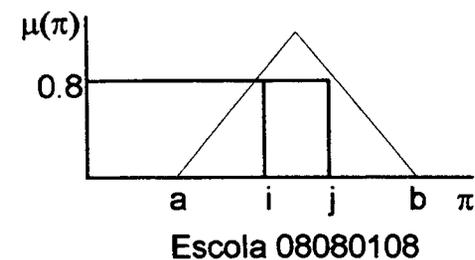


FIGURA 7.2 - Valores Corte -  $\alpha = 0.8$  - Escolas de Loulé

As possibilidades de aprovação para o aluno considerado, adotando-se um corte- $\alpha$  igual a 0.8, são apresentadas na TABELA 7.3.

TABELA 7.3 - Possibilidades de Aprovação (corte- $\alpha = 0.8$ ) - Escolas de Loulé - Aluno Exemplo

Escola	Limite Inferior	Limite Superior	Amplitude
08080108	0.9347	0.9770	0.0423
08080503	0.7563	0.9284	0.1721
08080810	0.8389	0.9654	0.1265
08080909	0.8688	0.9560	0.0872

Aplicando-se os valores resultantes do MODELO 6, para o aluno utilizado como exemplo tem-se os efeitos mostrados na TABELA 7.4.

TABELA 7.4 - Efeitos Contextuais - MODELO 6

Efeito	Valor
Adequação das Salas	(0.8912; 3.8285)
Estado de Conservação	(0.2920; 1.3120)
Localização do Estabelecimento	(0.2889; 0.8291)
Taxa de Ocupação	(-0.4597; 0)

Para alunos com características similares as do aluno exemplo, o fator mais influente na sua possibilidade de aprovação é Adequação das Salas de Aula. O fator Taxa de Ocupação é um fator que inibe a possibilidade de aprovação e, por não apresentar dispersão pode constituir-se em um forte argumento contra a massificação do ensino.

Com relação às escolas do Concelho de Loulé, o aluno teria, as seguintes possibilidades de aprovação (com grau de pertença máxima):

Escola	Possibilidade de Aprovação
08080108	Não Obs.
08080503	0.9335
08080810	0.9709
08080909	0.9812

Admitindo-se um corte -  $\alpha$  igual a 0.8, obter-se-iam as seguintes possibilidades de aprovação:

Escola	Possibilidade de Aprovação	Amplitude
08080108	Não Obs.	
08080503	(0.4727;0.9955)	0.5228
08080810	(0.5455;0.9989)	0.4534
08080909	(0.6492;0.9993)	0.3501

#### 7.4 - Comparação entre os Modelos Baseados na Análise de Regressão Possibilística e os Modelos de Regressão Clássicos

Para avaliar, globalmente, os modelos de regressão possibilística, testes baseados na estatística t foram realizados. A hipótese admitida foi a de que os modelos fundamentados na lógica difusa e os tradicionais não divergiam, significativamente, em relação aos fatores considerados nas componentes explicativas dos modelos. Os resultados obtidos são apresentados nas TABELAS 7.5, 7.6, 7.7 e 7.8.

TABELA 7.5 - Teste t para Coeficientes INTERCEPTO

	MODELO 1	MODELO 4
Média	4.141565217	5.785169565
Variância	1.100218984	3.201355772
Observações	23	23
Correlação de Pearson	-0.268454638	
Variância Comum	3.5	
Graus de Liberdade	35.52424239	
Estatística t	-3.800558852	
P(T<=t) bilateral	0.000553593	
t Crítico Bilateral	2.030110409	

TABELA 7.6 - Teste t para Coeficientes do Fator SEXO

	MODELO 1	MODELO 4
Média	0.000956522	-0.119
Variância	0.675100225	0.247911177
Observações	23	23
Correlação de Pearson	-0.503889458	
Variância Comum	3.5	
Graus de Liberdade	36.23775462	
Estatística t	0.598803389	
P(T<=t) bilateral	0.553053586	
t Crítico Bilateral	2.02809133	

TABELA 7.7 - Teste t para Coeficientes do Fator SSE

	MODELO 1	MODELO 4
Média	-0.10326087	-0.07943478
Variância	0.434813747	0.654062037
Observações	23	23
Correlação de Pearson	0.018032073	
Variância Comum	3.5	
Graus de Liberdade	42.28561484	
Estatística t	-0.109503366	
P(T<=t) bilateral	0.913324792	
t Crítico Bilateral	2.018082341	

TABELA 7.8 - Teste t para Coeficientes do Fator INDANO

	MODELO 1	MODELO 4
Média	-1.498086957	-1.5427087
Variância	0.247335538	0.331634181
Observações	23	23
Correlação de Pearson	-0.352300118	
Variância Comum	3.5	
Graus de Liberdade	43.08657799	
Estatística t	0.281243495	
P(T<=t) bilateral	0.779872615	
t Crítico Bilateral	2.016690814	

A partir do exame das TABELAS 7.5, 7.6, 7.7 e 7.8, verifica-se que com relação aos fatores SEXO, SSE e INDANO não foram encontradas diferenças significativas. Para os demais modelos estimados neste trabalho, resultados similares foram obtidos. Entretanto, os modelos de regressão possibilística permitem a análise de situações, nas quais, a variável dependente é um número difuso com função de pertinência dada por (5.22). Um exemplo de tal situação é apresentado a seguir.

## 7.5 - Análise de Regressão Quando a Variável Dependente é um Número Difuso

Uma questão relevante em um Sistema Educacional é a da evasão escolar. Uma resposta de interesse para os responsáveis pelo Sistema Educacional pode ser a propensão que um determinado aluno apresenta em abandonar o sistema. Essa resposta pode ser investigada, em termos de juízos de valores emitidos por professores, psicólogos e pedagogos que tenham um contato mais efetivo com o aluno. Através das respostas obtidas, pode-se ajustar uma distribuição de possibilidade que descreva a propensão que o indivíduo apresenta para abandonar o sistema. Essa situação não pode ser modelada pela aplicação dos MODELOS 1 e 2.

Considere-se que, para uma escola particular, tenha sido amostrado um grupo de alunos, e que estes apresentam as características individuais, mostradas na TABELA 7.9.

TABELA 7.9 - Características Individuais de uma Amostra de Alunos

Aluno	Distância Casa-Escola (DCE) (KM)	Renda do Agregado Familiar (SSE) (U.M.)	Escolaridade dos Pais (ESCP) (Anos)	Rendimento Escolar do Aluno (RAA) (Ind. Ac.)
1	7	108,819	4	15
2	5	95,812	2	13
3	4	82,913	13	8
4	9	69,866	6	9
5	5	74,581	4	9
6	4	80,199	14	17
7	5	67,752	3	19
8	5	117,402	2	11
9	4	109,737	8	20
10	6	76,466	3	18
11	9	84,705	5	25
12	5	54,702	8	16
13	9	68,066	5	18
14	7	95,654	6	12
15	7	79,923	15	15
16	6	82,883	14	12
17	8	99,273	5	13

Avaliar a propensão à evasão do Sistema Educacional para um aluno com base no exame de suas características e explicitar, de forma quantitativa, essa propensão envolve, claramente, a elaboração de um julgamento subjetivo e, portanto, sujeito a imprecisões que dificilmente podem ser explicadas através de um modelo probabilístico. Pode-se perceber que o "grau de influência" de cada um dos elementos considerados como condutores a uma propensão à evasão é difuso e não existe um modelo estocástico que o possa representar bem. Por outro lado, o juízo emitido, aqui representado pela possibilidade de evasão do Sistema Educacional, tampouco é preciso ou definitivo. É razoável admitir-se uma faixa de valores, na qual o decisor admite a sua valoração.

A variável "propensão á evasão", da forma que está estabelecida, pode ser definida no intervalo (0,1). Essa variável pode ser explicada através de características individuais de cada aluno como, por exemplo, rendimento do agregado familiar, escolaridade dos pais, rendimento escolar acumulado do aluno, distância casa-escola etc. Nesses casos, a variável-resposta pode ser avaliada em termos de uma função possibilística da forma:

$$pev_i = A_0 + A_1SSE_i + A_2ESCP_i + A_3RAA_i + A_4DCE_i \quad (7.1)$$

onde:  $pev_i$  = Propensão à evasão para o aluno  $i$

$SSE_i$  = Rendimento do agregado familiar do aluno  $i$

$ESCP_i$  = Escolaridade dos pais do aluno  $i$

$RAA_i$  = Rendimento escolar acumulado do aluno  $i$

$DCE_i$  = Distância casa-escola, percorrida pelo aluno  $i$

$A_j$  = números difusos que descrevem a estrutura difusa, através da qual as variáveis exógenas atuam sobre a variável-resposta

O modelo descreve, para uma escola particular ou um grupo de interesse, como as características individuais dos alunos podem contribuir, na avaliação da propensão à evasão.

Evidentemente, as características individuais dos alunos são valores bem definidos. Entretanto, é necessário avaliar a variável "dependente" dessas características: a propensão à evasão. Seus valores podem ser obtidos mediante um consenso entre especialistas, ou a partir da ponderação de opiniões entre os mesmos. Para a aplicação do modelo desenvolvido no capítulo 5 deste trabalho, para cada aluno são necessários apenas dois pontos: um de pertinência máxima e um limitante (à esquerda ou à direita). A TABELA 7.10 mostra um conjunto de resultados que poderiam ser obtidos a partir do exame do grupo de alunos e suas características descritas na TABELA 7.9.

TABELA 7.10- Valores de Julgamentos Emitidos por Especialistas para a Propensão à Evasão de Alunos

Aluno	Valor à Esquerda P <sub>1</sub>	Valor Central P <sub>0</sub>	Valor à Direita P <sub>2</sub>
1	0.35	0.4	0.5
2	0.35	0.4	0.6
3	0.45	0.5	0.6
4	0.5	0.6	0.7
5	0.5	0.65	0.7
6	0.45	0.5	0.6
7	0.55	0.6	0.65
8	0.2	0.25	0.35
9	0.05	0.1	0.15
10	0.45	0.5	0.55
11	0.1	0.15	0.2
12	0.4	0.5	0.6
13	0.55	0.60	0.65
14	0.45	0.5	0.6
15	0.2	0.3	0.4
16	0.3	0.35	0.4
17	0.15	0.2	0.3

Este modelo foi ajustado utilizando-se o sistema computacional desenvolvido para este trabalho.

### **7.5.1 - Ajuste do Modelo Possibilístico Triangular Logit para a Análise da "Propensão à Evasão"**

Para a estimação dos coeficientes difusos do modelo para análise da propensão à evasão, dois critérios foram utilizados.

No primeiro, procurou-se estimar um conjunto de coeficientes, para os quais as estimativas produzidas contivessem, a um nível  $h$  de inclusão, os valores arbitrados pelos especialistas. Esse critério implica na resolução do sistema possibilístico através da minimização das dispersões dos coeficientes difusos. O segundo critério, procurou determinar os valores dos coeficientes de forma que as suas estimativas apresentassem uma interseção, para o mesmo nível  $h$ , com os valores emitidos pelos especialistas. Esse segundo critério corresponde à solução do problema de "Conjunção", ou seja determinar o mínimo dispersão para os coeficientes difusos, de tal forma que a interseção entre os conjuntos difusos respeite o nível  $h$  preestabelecido.

#### **7.5.1.1 - Problema de Minimização**

Inicialmente, o modelo (7.1) foi resolvido utilizando-se os parâmetros descritos na FIGURA 7.3.



Educacional. Com relação às variáveis Distância Casa-Escola e Renda do Agregado Familiar, seus efeitos ocorrem de maneira imprecisa, acarretando nas dispersões verificadas. É evidente que novas variáveis incorporadas no modelo podem permitir uma melhor clarificação desses efeitos.

Resultados do Modelo Logit		
<b>Problema de Minimização</b>		
Valor de J(c):	25.5736	
Nível H:	0.3	
<b>Variável Dependente</b>		
Alfa	A.P02	
Spread	A.P12	
<b>Ajuste</b>		
[Localizar]		
[Ver Diagrama]		
[Cancelar]		
<b>Variável</b>	<b>Alfa</b>	<b>Spread</b>
A.N	0.0243	0.6447
A.N2	-0.3469	0.5535
A.N3	-0.4339	0
A.N4	-0.1757	0
Intercepto	-0.1648	0.5268

FIGURA 7.4 - Apresentação dos Resultados do Modelo 7.1 -Problema de Minimização - Valores  $P_0$  e  $P_1$

As estimativas obtidas e seu grau de ajuste são apresentadas na FIGURA 7.5.

Estatísticas							
Obs	Observado			Estimado			Pertinências
	Esq	Max	Dir	Esq	Max	Dir	$\hat{Y}$
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
1	0.3500	0.4000	0.4522	0.1255	0.2789	0.5103	0.30
2	0.3500	0.4000	0.4522	0.1304	0.4616	0.8306	0.84
3	0.4500	0.5000	0.5500	0.2174	0.6257	0.9096	0.68
4	0.5000	0.6000	0.6923	0.1161	0.4952	0.8799	0.74
5	0.5000	0.6500	0.7752	0.2994	0.5668	0.8003	0.30
6	0.4500	0.5000	0.5500	0.2714	0.6212	0.8784	0.61
7	0.5500	0.6000	0.6480	0.2898	0.6307	0.8773	0.89
8	0.2000	0.2500	0.3077	0.1390	0.4802	0.8409	0.30
9	0.0500	0.1000	0.1900	0.0310	0.2612	0.7960	0.30
10	0.4500	0.5000	0.5500	0.3868	0.6474	0.8424	0.30
11	0.1000	0.1500	0.2189	0.0712	0.2954	0.6962	0.30
12	0.4000	0.5000	0.6000	0.1300	0.5343	0.8980	0.92
13	0.5500	0.6000	0.6480	0.2051	0.5851	0.8852	0.96
14	0.4500	0.5000	0.5500	0.1146	0.3188	0.6285	0.30
15	0.2000	0.3000	0.4235	0.1394	0.3000	0.5313	1.00
16	0.3000	0.3500	0.4035	0.2678	0.4380	0.6241	0.30
17	0.1500	0.2000	0.2615	0.1119	0.3239	0.6456	0.34

Reprocessar   Gráficos   Gravar Est.   Impressa   Cancelar

FIGURA 7.5 - Valores Observados e Valores Estimados - Problema de Minimização - Modelo 7.1

Um aluno com características semelhantes às apresentadas pelo aluno 15 (Distância Casa - Escola em torno de 7 Km, Rendimento Familiar aproximadamente 80.000 u.m., Escolaridade dos Pais: 15 anos e Rendimento Acumulado igual a 15) apresenta uma distribuição de possibilidade para a propensão à evasão no intervalo (0,13;0.53). Graficamente esse resultado pode ser visto na FIGURA 7.6.

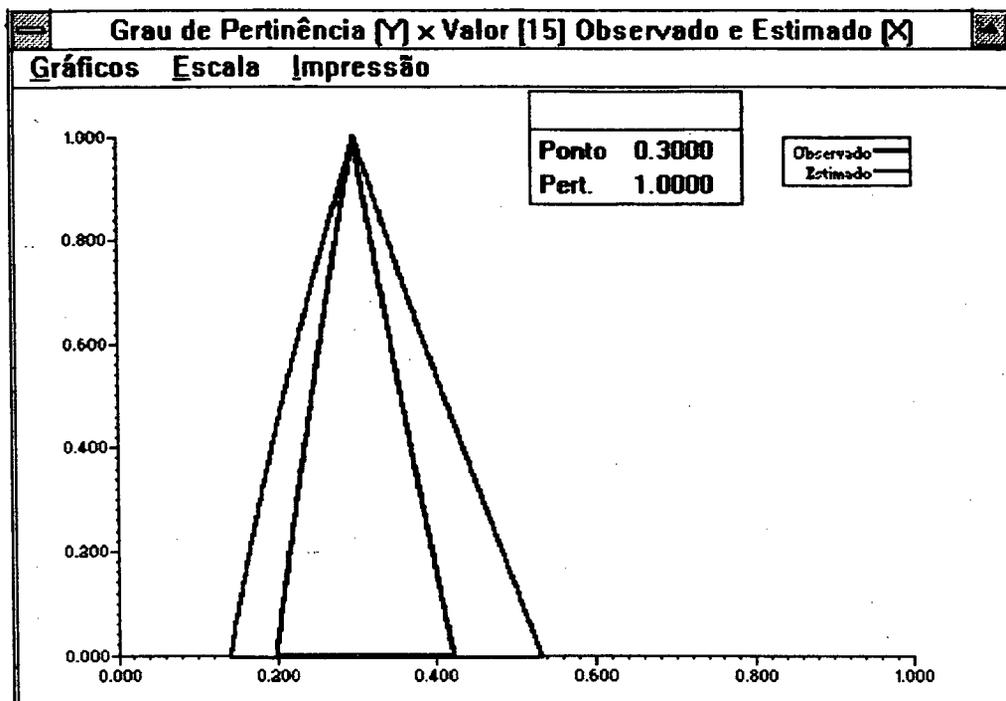


FIGURA 7.6 - Grau de Pertinência do Valor Estimado e do Valor Observado -  
Observação 15

A FIGURA 7.7 apresenta os intervalos obtidos para as 17 observações. Um novo ajuste foi efetuado, agora com a exclusão de algumas observações que apresentaram maior dispersão ( Obs. 4, 6 e 13). Os resultados obtidos foram equivalentes e estão apresentados na FIGURA 7.8. Com relação à observação 15, nota-se um pequeno acréscimo à direita do valor estimado (FIGURA 7.6 e FIGURA 7.9, quadrante direito inferior). Esse acréscimo, contudo, é compensado pelo valor mínimo observado à esquerda do ponto centróide.

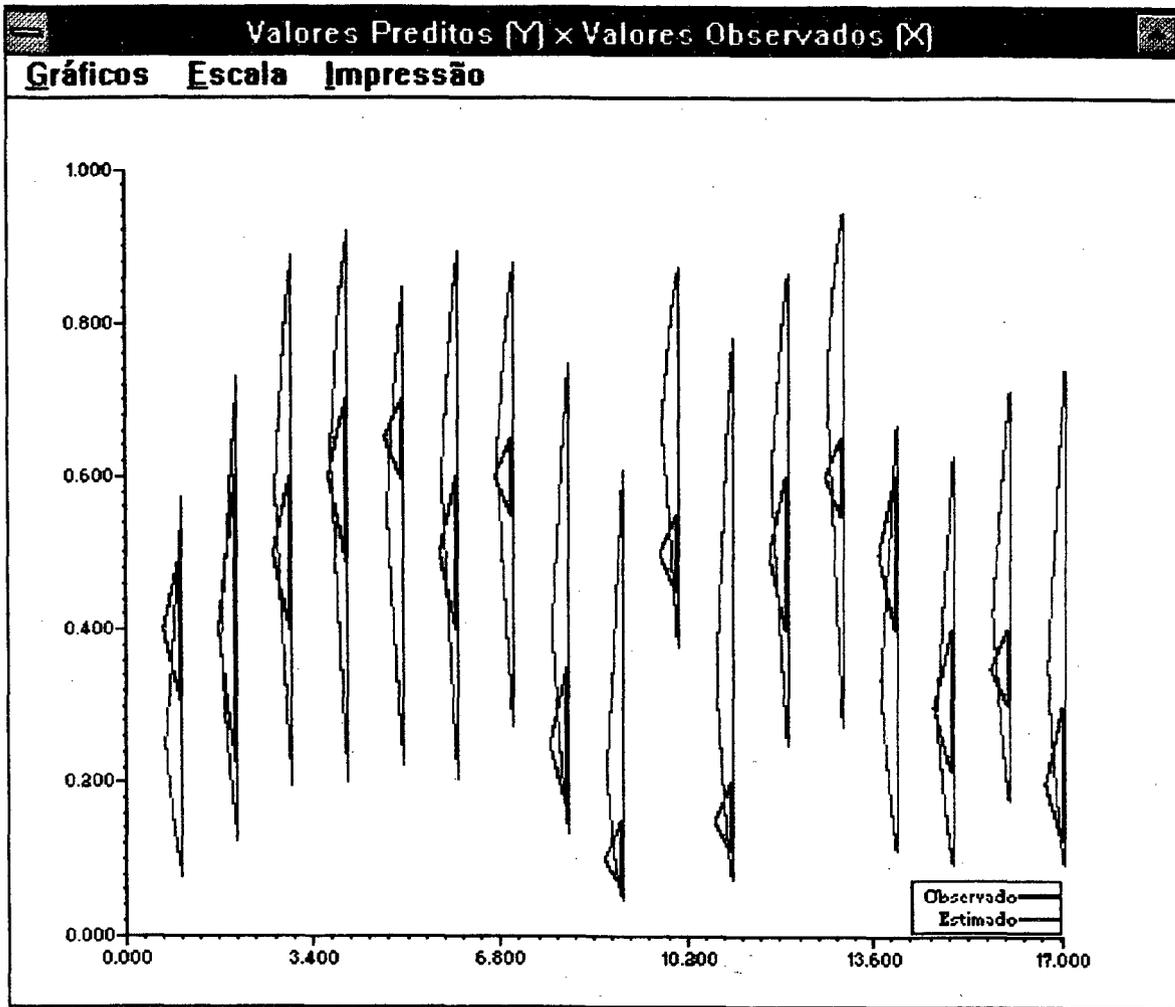


FIGURA 7.7 - Gráfico dos Intervalos - 17 observações

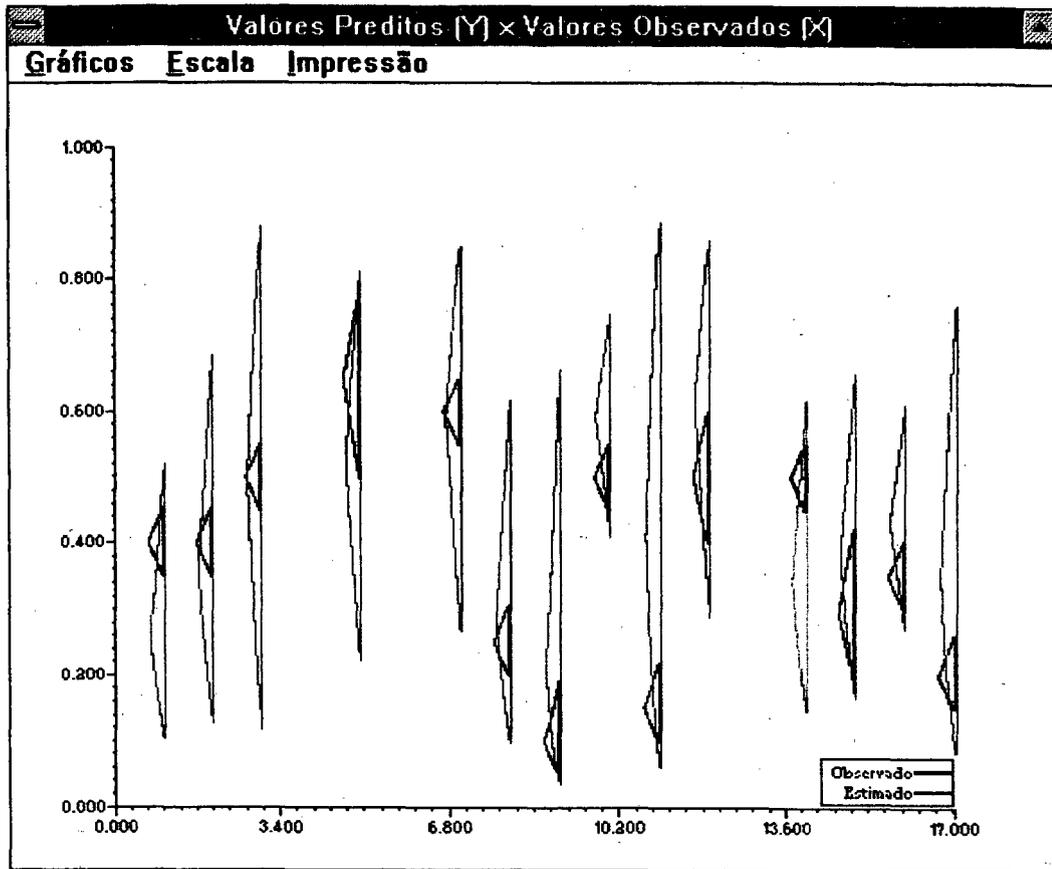


FIGURA 7.8 - Gráfico dos Intervalos - 14 observações

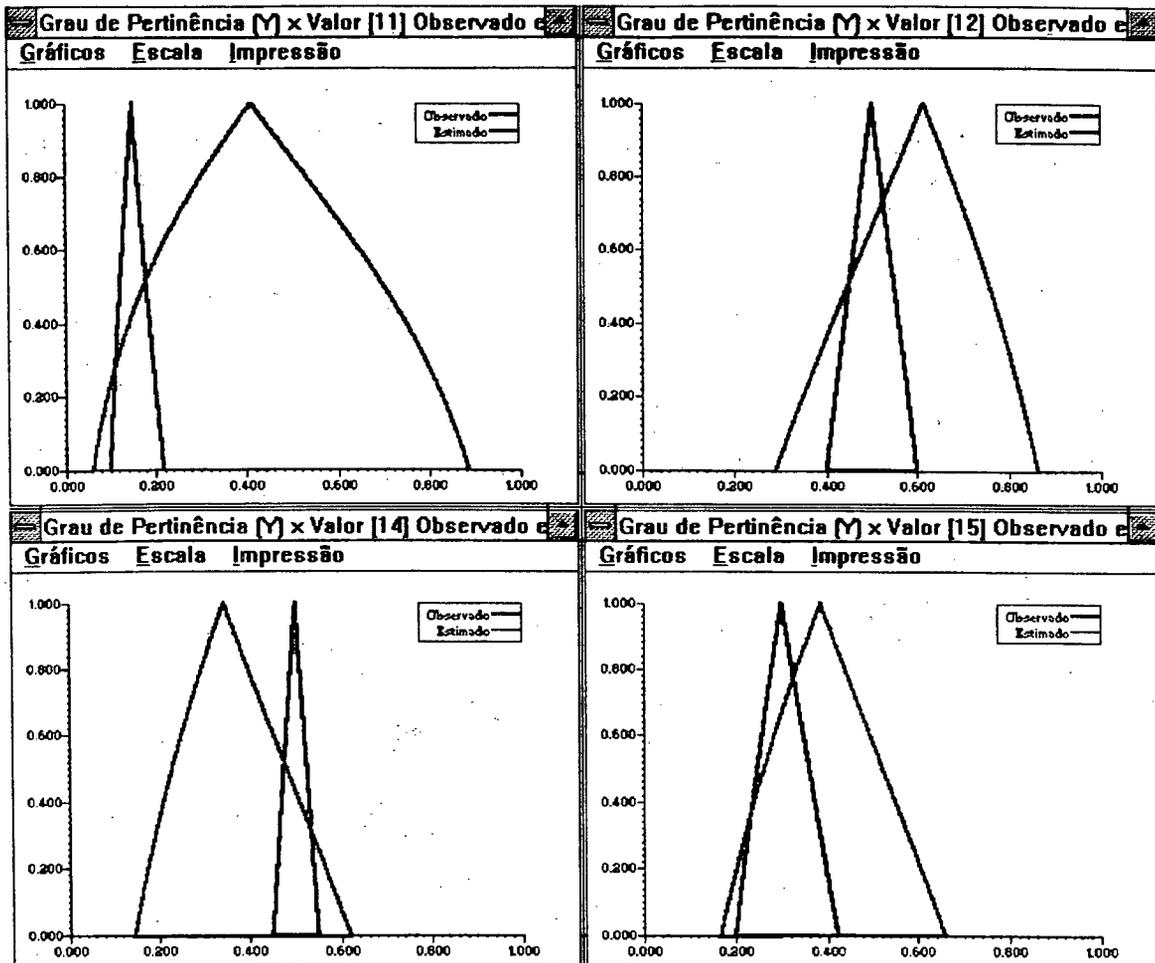


FIGURA 7.9 - Quatro Observações do Segundo Ajuste

### 7.5.1.2 - Problema de Conjunção.

O mesmo modelo (7.1) foi resolvido utilizando como critério para determinar os coeficientes a dispersão mínima que conduza a uma interseção entre valores estimados e valores observados para um nível  $h$  preestabelecido. Os resultados obtidos são apresentados na FIGURA 7.10.

Resultados do Modelo Logit		
<b>Problema de Conjunção</b>		<b>Ajuste</b>
Valor de J(c):	13.5276	<input type="button" value="Estimar"/>
Nível H:	0.3	<input type="button" value="Estimar"/>
<b>Variável Dependente</b>		<input type="button" value="Estimar"/>
Alfa	A.P02	
Spread	A.P12	
<b>Variável</b>	<b>Alfa</b>	<b>Spread</b>
A.N	0.2674	0.3869
A.N2	-0.2112	0
A.N3	-0.5422	0
A.N4	-0.1231	0
Intercepto	-0.2048	0.4686

FIGURA 7.10 - Resultados Obtidos - Problema Conjunção

Para esse caso, apenas, a variável Distância Casa-Escola manteve dispersão. Assim, para encontrar-se uma estrutura de coeficientes que ocasione a ocorrência de interseção entre estimativas e os valores arbitrados pelos especialistas para a propensão à evasão, é necessário considerar, apenas, a imprecisão derivada do efeito distância casa-escola. Modelos resolvidos por conjunção são úteis para os casos em que o decisor necessita de uma indicação sobre a direção do efeito das variáveis independentes na variável dependente. As FIGURAS 7.11, 7.12 e 7.13 apresentam, respectivamente, as novas estimativas obtidas para o modelo 7.1, o grau de ajuste para a observação 15 nesse novo modelo e os intervalos construídos para as observações do problema.

Estatísticas							
Obs	Observado			Estimado			Pertinências
	Esq	Max	Dir	Esq	Max	Dir	$\hat{Y}$
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
1	0.3500	0.4000	0.4522	0.1602	0.2673	0.4108	0.30
2	0.3500	0.4000	0.4522	0.2091	0.3531	0.5298	0.79
3	0.4500	0.5000	0.5500	0.3111	0.5368	0.7485	0.87
4	0.5000	0.6000	0.6923	0.3470	0.6110	0.8227	0.97
5	0.5000	0.6500	0.7752	0.3382	0.5133	0.6853	0.58
6	0.4500	0.5000	0.5500	0.3299	0.5582	0.7644	0.80
7	0.5500	0.6000	0.6480	0.3769	0.5553	0.7205	0.80
8	0.2000	0.2500	0.3077	0.2470	0.4038	0.5830	0.30
9	0.0500	0.1000	0.1900	0.1238	0.2662	0.4821	0.30
10	0.4500	0.5000	0.5500	0.4971	0.6214	0.7315	0.30
11	0.1000	0.1500	0.2189	0.1498	0.3425	0.6062	0.30
12	0.4000	0.5000	0.6000	0.3901	0.5691	0.7316	0.75
13	0.5500	0.6000	0.6480	0.4221	0.6834	0.8645	0.72
14	0.4500	0.5000	0.5500	0.2241	0.3557	0.5135	0.30
15	0.2000	0.3000	0.4235	0.1727	0.2852	0.4327	0.94
16	0.3000	0.3500	0.4035	0.2848	0.3980	0.5233	0.72
17	0.1500	0.2000	0.2615	0.1975	0.3690	0.5816	0.30

FIGURA 7.11 - Valores Observados e Valores Estimados - Modelo 7.1 -Problema de Conjunção

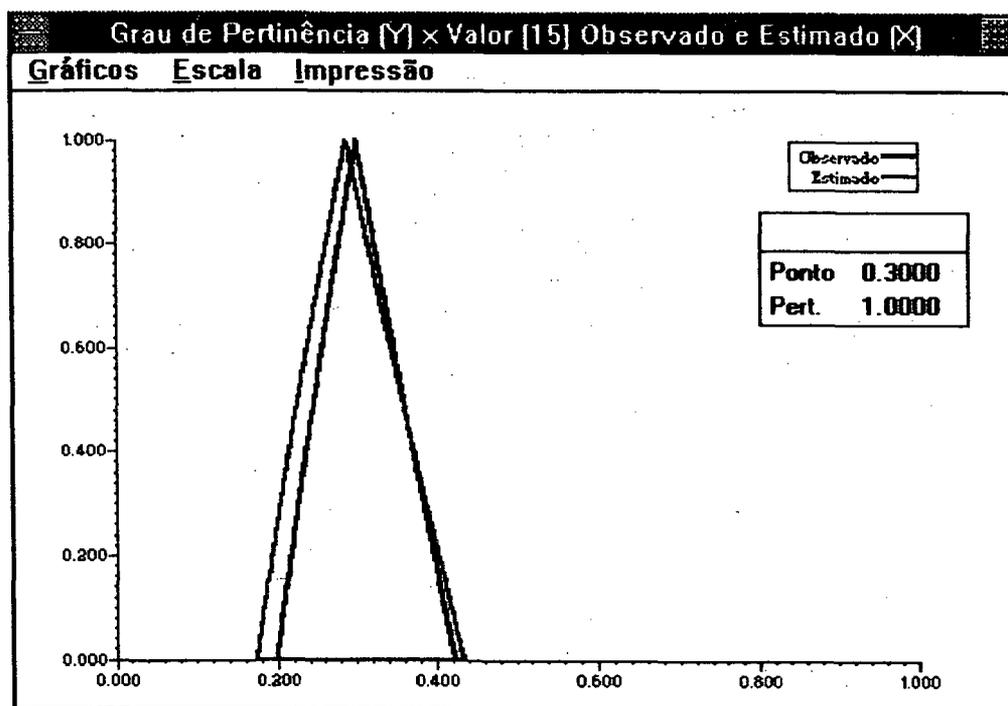


FIGURA 7.12 - Grau de Pertinência do Valor Estimado e do Valor Observado - Observação 15 - Problema de Conjunção

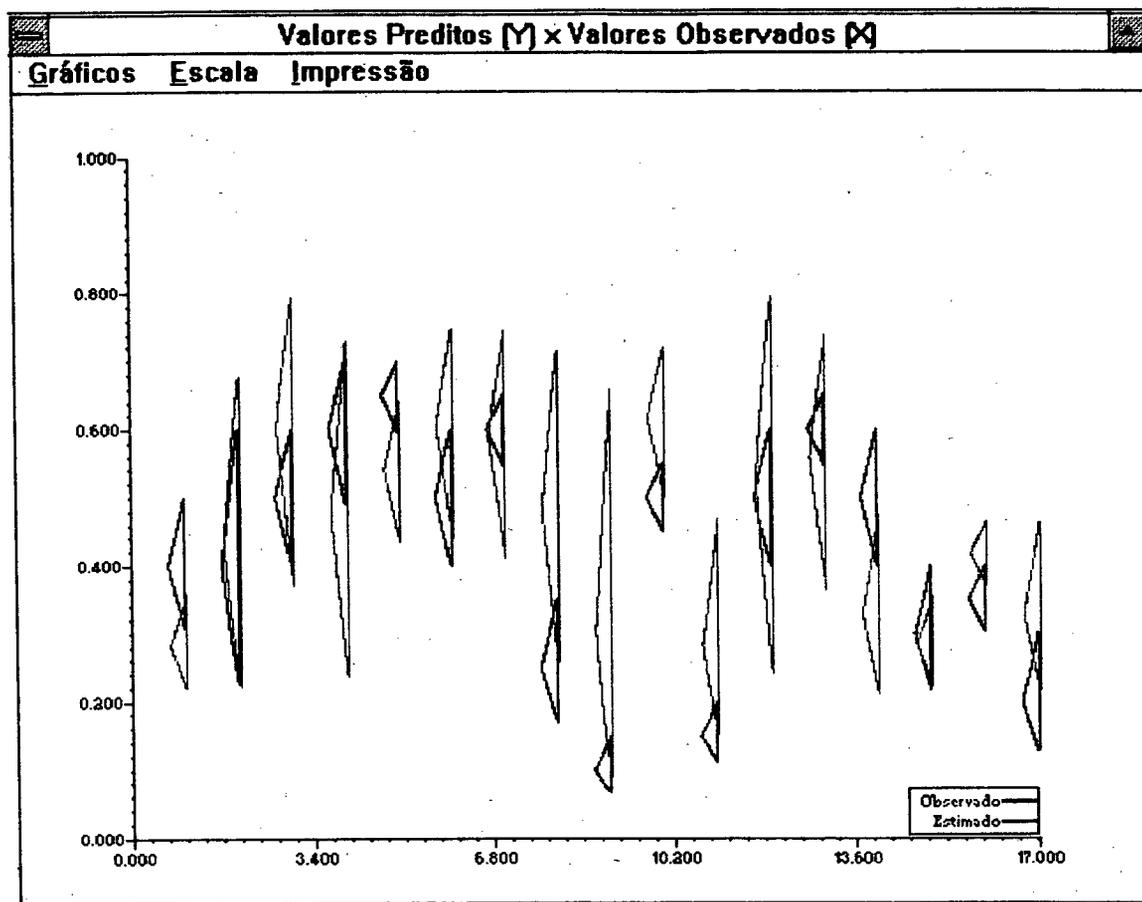


FIGURA 7.13 - Gráfico dos Intervalos - 17 Observações - Problema de Conjunção

## 7.6 - Conclusões

Este capítulo apresenta uma série de aplicações dos modelos difusos para análise de sistemas educacionais.

Questões que envolvam julgamento subjetivo, sujeito a imprecisões, podem, perfeitamente, serem tratadas através de um modelo triangular logit difuso. Esse modelo incorpora e trata relações difusas, permitindo aos agentes decisores avaliar o impacto dessas relações sobre alguma variável. Para demonstrar a viabilidade de tais modelos, um exemplo foi proposto e resolvido.

Um estudo de caso foi apresentado com o objetivo de demonstrar que os modelos difusos são, perfeitamente, aplicáveis em problemas reais. Com base nesse estudo de caso, foi possível definir uma classificação de estabelecimentos de ensino com relação a um determinado aspecto de interesse. A generalização do processo é imediata. O processo, através do qual, a classificação foi obtida, no caso de modelos difusos, ultrapassa os aspectos problemáticos de processos que se baseiam na lógica clássica, ao permitir transições suaves entre diversos atributos utilizados na classificação.

Modelos que incorporam variações dentro e entre escolas, modelos hierarquizados, também podem ser elaborados a partir da utilização de regressões possibilísticas. Nesses tipos de modelos, o uso de regressão possibilística é mais adequado, uma vez que é pouco provável que se possa estabelecer um modelo estocástico que traduza os efeitos de características de escolas sobre um grupo de alunos. O que se pode afirmar, com segurança, é que tais efeitos são, no mínimo, imprecisos e, justamente, é esse o pressuposto básico para a aplicação de regressão possibilística.

## **8. CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS**

### **8.1 - Conclusões**

A abordagem sistêmica representa uma maneira de tratar realidades de forma racional e científica. Nessa abordagem, o pensamento analítico é utilizado de forma sistemática e organizadamente. Sob o enfoque sistêmico o planejamento é encarado, sobretudo, como um processo de tomada de decisão. É através desse processo que se pretende a consecução de objetivos bem claros e definidos. Assim, a realidade é modelada com vistas ao estabelecimento de um ideal a ser alcançado.

A realização dos objetivos se dá por intermédio de alternativas colocadas ao longo do processo de tomada de decisão. Alternativas são possíveis cursos de ação a disposição do tomador de decisões. A escolha do melhor curso se dá através da escolha da melhor alternativa. Para avaliar-se alternativas são necessários critérios que possam atuar como grandezas utilizadas para efeitos de comparação. Esses critérios podem, então, ser qualitativos e/ou quantitativos.

Em sistemas complexos, como é o caso do sistema educacional, diversos são os tipos de critérios possíveis. Embora, existam alguns tipos de critério que possam aparentar uma aplicação universal, como por exemplo, custo x benefício, é razoável admitir que em sistemas educacionais os critérios sejam estabelecidos a partir de valores educacionais. Entre os valores educacionais encontram-se, por exemplo, valores sociais, valores culturais, valores humanistas, valores práticos e valores econômicos.

A aplicação de um enfoque sistêmico pressupõe definir um modelo de avaliação que permita inferir a exeqüibilidade dos objetivos e a eficiência do sistema. Esse modelo de avaliação permite, também, que o próprio sistema seja retro-alimentado de modo a garantir a continuidade do processo. Uma dificuldade encontrada em modelos de avaliação de sistemas educacionais é que educadores, de uma forma geral, tem seus interesses focalizados em resultados como: aumento da capacidade de raciocinar, grau de aprendizagem de certa matéria, desenvolvimento de capacidade intelectual, etc. Essas questões são, por natureza, imprecisas e difusas.

A informação necessária para avaliar resultados positivos no desempenho de um sistema educacional não deve basear-se apenas em medidas de desempenho dos alunos. O conhecimento de como o desempenho é influenciado por outros fatores controláveis permite a adoção de alternativas que vão de encontro, de forma mais imediata, aos objetivos idealizados. É sabido que, por exemplo, condições físicas de estabelecimentos escolares exercem um efeito bastante acentuado nos alunos. Escolas com excelentes laboratórios têm mais condições de fixar a atenção dos alunos do que aquelas menos equipadas. A política adotada para distribuição das atividades escolares pode conduzir a uma maior ou menor atratividade dos alunos em permanecerem nas escolas. Compreender a intensidade com que essa atratividade ocorre deve ser objeto permanente de análise.

Com relação a análise de sistemas educacionais, o trabalho apresenta e discute a construção de modelos para interpretação de questões de fundamental importância para a compreensão desse tipo de sistema.

É discutido no trabalho que as principais dificuldades derivadas da utilização desses modelos relacionam-se aos seguintes pontos:

- Identificar qual é a melhor estrutura que represente uma variável resposta de interesse;

- Ajustar essa estrutura de forma que os mecanismos geradores dos dados observados sejam adequadamente representados.

O capítulo três deste trabalho é dedicado ao estudo e descrição de diversos tipos de modelos que podem ser utilizados para análise de Sistemas Educacionais, sob a ótica da modelagem estatística tradicional.

Sistemas Educacionais, como já citado, são sistemas nos quais relações humanas estão envolvidas. As relações humanas são, em sua gênese, relações difusas e imprecisas. Assim, em alguns casos, ocorrem problemas que não possuem uma estrutura na matemática clássica nem na teoria das probabilidades, para serem resolvidos. Entretanto, são problemas que os seres humanos são capazes de resolver através do uso do chamado raciocínio aproximado ou difuso, baseado na teoria dos conjuntos difusos e na lógica difusa. Raciocínio difuso é o processo (ou processos) através do qual uma conclusão possivelmente imprecisa, é deduzida de uma coleção de premissas imprecisas. Tal raciocínio é, na maioria dos casos, antes qualitativo do que quantitativo, e quase sempre cai fora do domínio de aplicabilidade da lógica clássica.

Baseado na teoria dos conjuntos difusos é possível definir e operacionalizar uma técnica de análise denominada Análise de Regressão Possibilística (ou difusa). A análise de regressão possibilística permite a construção de modelos em ambientes onde é grande a presença de incerteza, bem como de fenômenos difusos. A relação entre variável dependente e variáveis independentes não é bem estabelecida, como no caso da regressão tradicional. A suposição básica é a de que os desvios entre valores observados e preditos são devidos a imprecisões e nebulosidades existentes na estrutura do sistema. De forma mais concreta, imprecisões e nebulosidades ocorrem nos coeficientes do modelo. Essa ferramenta de análise é discutida no capítulo quatro deste trabalho.

Como contribuição teórica do trabalho, uma extensão ao modelo de regressão linear possibilística é apresentada no capítulo 5 onde se estabelecem modelos para análise de sistemas educativos.

Esse novo modelo permite que situações bastante freqüentes na prática como a variável dependente sendo restrita ao intervalo  $(0,1)$ , possam ser modeladas através de um sistema possibilístico. Em situações de avaliações, expressas em percentagens, normalmente o decisor utiliza um tipo de raciocínio vago: em torno de 80%, ou seja, existe um valor para o qual o decisor credita a possibilidade máxima. Essas percentagens vão declinando, não abruptamente, até atingir valores que o próprio decisor não crê serem factíveis. Essas situações são contempladas pelos modelos apresentados e discutidos neste trabalho. Os modelos apresentados são capazes de tratar situações de escolha qualitativa em que subsistam nebulosidades ou imprecisões, próprias do raciocínio humano, quanto à importância dos fatores considerados como relevantes na explicação da escolha. Quando pretende-se modelar situações nas quais a variável dependente, além de expressar conceitos derivados da linguagem natural, está restrita a um intervalo, um modelo específico é proposto. Para utilização desse modelo, dois pontos no intervalo  $(0,1)$  são solicitados ao decisor. Obviamente, quanto maior for o grau de incerteza presente no julgamento, maiores serão os intervalos estimados. Tal fato reflete, então, a realidade verificada. Esse modelo tem importantes aplicações no campo da avaliação de desempenho de sistemas educacionais, principalmente, em questões relacionadas a aproveitamento, repetências, evasões etc. Uma aplicação do modelo é apresentada no capítulo 7 deste trabalho.

Um estudo de caso é apresentado no capítulo 6, com o objetivo de demonstrar que o ferramental baseado em lógica difusa é plenamente aplicável em problemas reais. Para efeitos de comparação, os mesmos dados são analisados utilizando conceitos e técnicas de análise estatística tradicionais.

Baseado nesse estudo de caso foi possível estabelecer uma classificação de estabelecimentos de ensino com relação a um determinado

aspecto de interesse. A generalização do processo é imediata, não se prevendo dificuldades adicionais. O processo, através do qual, a classificação foi obtida ultrapassa aspectos problemáticos de processos que se baseiam na lógica clássica, ao permitir transições suaves entre diversos atributos utilizados na classificação.

Modelos que incorporem variações dentro e entre escolas - modelos hierarquizados - também podem ser elaborados a partir da utilização de regressões possibilísticas. Talvez, nesses tipos de modelos, o uso de regressão possibilística seja mais adequado, uma vez que é pouco provável que se possa estabelecer um modelo estocástico que traduza os efeitos de características de escolas sobre um grupo de alunos. O que se pode afirmar, com segurança, é que tais efeitos são no mínimo imprecisos e, justamente, é esse o pressuposto básico para a aplicação de regressão possibilística.

## **8.2 - Recomendações para Trabalhos Futuros**

Técnicas de regressão possibilística, conduzem a uma abordagem diferente da proposta pela análise tradicional de modelos lineares. Na análise tradicional são freqüentes os recursos a testes de hipóteses sobre comportamentos de variáveis. É, portanto, bastante comum que, uma vez processado um modelo possibilístico se tente uma validação do mesmo utilizando-se critérios estabelecidos para a análise tradicional. Neste trabalho inclusive, os mesmos modelos foram gerados de acordo com as técnicas usuais de regressão, obtendo-se resultados similares. Contudo, a comparação entre estes dois tipos de resultados é, ainda, um problema em aberto.

Um dos critérios utilizados na verificação de um modelo possibilístico é o do somatório minimizado das dispersões. Contudo, a dispersão em modelos possibilísticos é diretamente proporcional à grandeza das variáveis

envolvidas. No modelo computacional desenvolvido para a realização deste trabalho são incluídos alguns tipos de transformações que podem ser utilizadas sobre variáveis utilizadas durante o processo de modelação. Novos estudos sistematizados sobre essas transformações seriam de grande utilidade para utilizações futuras.

Para fazer frente a problemas de elevado grau de complexidade presentes em sistemas reais, deve-se dispor de ferramentas adequadas. Novas ferramentas tem sido propostas para o tratamento de incerteza e de problemas complexos. Essas técnicas vem contribuindo para a substituição do paradigma da informação pelo paradigma do conhecimento.

O problema da modelagem da incerteza tem se tornado na questão principal (DUBOIS e PRADE, 1989) em diversas áreas, como por exemplo, Inteligência Aplicada, Teoria da Decisão, Teoria de Controle, etc. . Por uma questão de tradição mais do que adequação (CHOOBINEH e BEHRENS, 1992), vários modelos para tratamento de incerteza baseiam-se nos pressupostos e axiomáticas da teoria das probabilidades: existência de informação a respeito de freqüências e ocorrência relativa de resultados possíveis. Tais modelos incorporam o tratamento de incertezas do Tipo Um, conforme proposto por GUPTA (1991). Este tipo de incerteza trata com informações ou fenômenos que surgem a partir do comportamento aleatório de um sistema. Existe uma teoria estatística bastante rica para caracterizar tais fenômenos aleatórios.

O problema que se depara na modelagem probabilística da incerteza é o de estudar incertezas que surgem derivadas do pensamento humano, processos de raciocínio, processos cognitivos, processos de percepção ou, de uma forma mais ampla, informação cognitiva. A este tipo de incerteza, denominada de incerteza Tipo Dois por GUPTA, os métodos tradicionais de tratamento não podem ser aplicados. Os modelos estatísticos tradicionais não permitem uma explicação cognitiva suficiente. Conjuntos difusos podem ser utilizados para capturar imprecisões e características nebulosas decorrentes de conceitos naturais (PENG, KANDEL e WANG, 1991).

A lógica difusa apresenta como principal vantagem a possibilidade de tratamento de problemas e a obtenção de respostas em situações onde a informação disponível é difusa, ambígua, qualitativa, incompleta ou imprecisa. Por essa razão, os sistemas baseados em conjuntos difusos apresentam uma habilidade de inferir, similar ao raciocínio humano. Esses sistemas tornam-se, desse modo, compreensíveis, robustos, de fácil tratamento e manutenção. KOSKO (1994), por exemplo, enumera algumas aplicações práticas: controle de qualidade de água; sistemas automáticos de operação de trens; controle de elevadores; controle de reatores nucleares; controle de transmissão em automóveis; sistemas de hardware para controle lógico difusos; chips programados com lógica difusa etc.

A representação do conhecimento e sua modelagem através de uma ferramenta de análise permite inferir o desenvolvimento de sistemas inteligentes aplicados. Unindo duas vertentes em franco desenvolvimento na ciência atual: Hardware e Software, pode-se prever um crescimento fantástico para a melhoria de performance dos sistemas inteligentes. SOUCEK (1992) fala de um novo paradigma: "processamento e raciocínio através de associação e através de analogia".

Para SOUCEK, desenvolvimento de hardware inclui: processamento paralelo intensivo; implementação de sistemas cognitivos difusos de alta velocidade; conceitos inovadores de processadores associativos; conteúdos de memória acessíveis diretamente sem pesquisa adicional e aceleradores celulares para computadores pessoais com até 2000 MOPS (milhões de operações por segundo). A nível de software: redes neurais holográficas, invariantes, de alta-ordem e temporais; sistemas especialistas difusos; desenvolvimento de pacotes computacionais voltados para raciocínio difuso, tomada de decisão e controle; métodos de programação para visão associativa e processadores celulares; ambientes especiais de programação orientada a objetos e ferramentas adaptativas orientadas a usuários. De acordo com SOUCEK, os sistemas inteligentes poderão ser capazes de tratar com problemas altamente complexos, confusos, apresentando ruídos em

administração, finanças, reconhecimento de padrões, instrumentação adaptativa, controle de processo e robótica, diagnósticos, bases de dados inteligentes etc.

Da mesma forma que sistemas especialistas baseados em regras irão ser, provavelmente, superiores a redes neurais quando o problema sob análise for bem definido, fechado e não envolver incertezas e relações difusas, as técnicas estatísticas irão ser, provavelmente, superiores a outras técnicas sempre que a informação disponível for completa e disponível em termos de frequências de ocorrência de todos os resultados possíveis. O tratamento de problemas complexos exige, obviamente, a aplicação de ferramentas adequadas.

## BIBLIOGRAFIA

- ACKOFF, R.L., Operational Research, John Wiley and Sons, New York, 1963.
- AITKIN, M. e LONGFORD, N. Statistical Modelling Issues in School Effectiveness Studies. Journal of Royal Statistical Society, A, 149, parte 1, 1986. p. 1-43.
- AMEEN, J.R.M. e HARRISON, P.J. Discounting Bayesian Multiprocess Modelling with CUSUM. Internal Report N.34, Statistics Dept., Warwick University, Coventry, UK. 1983.
- ANDERSON, D.A. e AITKIN, M. Variance Component Models with Binary Response: Interviewer Variability. Journal of the Royal Statistical Society, B., v. 47, nº 2, 1985. p. 203-210.
- ASHTON, W.D. The Logit Transformation with Special Reference to its Uses in Biossay. London, Charles Griffin & Company Limited, 1972. 88p.
- BALDWIN, J.F. Fuzzy Sets and Expert Systems. Information Sciences, v. 36, 1985. p. 123-156.
- BANDEMÉR, H. Evaluating Explicit Functional Relationships from Fuzzy Observations. Fuzzy Sets and Systems, nº 16, 1985. p. 41-52.
- BANKS, D.L. Patterns of Oppresion: An Exploratory Analysis of Human-Rights Data. Journal of the American Statistical Association, v. 84, nº 407, set.1989. p. 674-681.
- BASTOS, L.C. e BASTOS, R.C. Programas de Análise do Parque Escolar, Lisboa, Ministério da Educação/Gabinete de Estudos e Planeamento, 1989. 91p.

- BASTOS, R.C. "Aplicação de Modelos Generalizados na Análise de Dados Educacionais" I Conferência em Estatística e Optimização, Troia, 3-5 Dezembro, 1990.
- BELLMAN, R.E. e ZADEH, L. Decision-Making in a Fuzzy Environment. Berkeley, Report nº ERL-69-8, NASA CR-1594. 59p.
- BEZDEK, J.C.; HATHAWAY, R.J.; SABIN, M.J. e TUCKER, W.T. Convergence Theory for Fuzzy c-Means: Counter Examples and Repairs. IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-17, nº 5, set/out. 1987. p. 873-878.
- BOCK, R.D. Multilevel Analysis of Educational Data. San Diego, Academic Press, 1989. 354p.
- BOGLER, P.L. Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems. IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-17, N. 6, nov/dez. 1987. p. 968-977.
- BOOCH, G. Object Oriented Design with Applications, The Benjamin/Wmings Publishing Company, Inc., California, 1991. 580 p.
- BOOKER, J.M. e MEYER, M.A. Sources of and Effects of Interexpert Correlation: An Empirical Study. IEEE Transactions on Systems, Man, and Cybernetics, v. 18, N. 1, jan/fev. 1988. p. 135-142.
- BOURDIEU, P. e PASSERON, J. C. A Reprodução - Elementos para uma Teoria do Sistema de Ensino, Ed. Francisco Alves, Rio de Janeiro, 1975.
- BOUSSOCIANE, A.; DYSON, R.G. e THANASSOULIS, E. Applied Data Envelopment Analysis. European Journal of Operational Research, N. 52, 1991. p. 1-15.
- BROOKING, A.G. The Analysis Phase in Development of Knowledge based Systems. Artificial Intelligence and Statistics. Editor: William A. Gale. Reading, Addison-Wesley, 1986. 418p.
- BRYK, A.S. e RAUDENBUSCH, S.W. Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical

Linear Model. In *Multilevel Analysis of Education Data*. Ed. by R. Darwel Bock, Academic Press, 1989. p.159-199.

BUNKE, H. e BUNKE, O. *Statistical Inference in Linear Models - Statistical of Model Building*, John Wiley and Sons, New York, 1986, 614p.

CAMERON, A.C. e TRIVEDI, P.K. *Econometric Models based on Count Data: Comparisons and Applications of some Estimators and Tests*. *Journal of Applied Econometrics*, v. 1, N. 1, jan. 1986. p. 29-53.

CARLSSON, C. *Editorial to the Feature Issue on Fuzzy Sets*. *European Journal of Operational Research*, v. 25, N. 3, jun. 1986. p. 317-319.

CARROL, R.J. e RUPPERT, D. *Transformation and Weighting in Regression*. New York, Chapman and Hall, 1988. 249p.

CHANDRASEKARAN, F.B. e GOEL, A. *From Numbers to Symbols to Knowledge Structures: Artificial Intelligence Perspectives on the Classification Task*. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 18, nº 3, maio/jun. 1988. p. 415-424.

CHARNES, A., COOPER, W.W. e RHODES, E., *Measuring The Efficiency of Decision Units*, *European Journal of Operational Research*, Vol. 6, 1978. p. 429-444.

CHEESEMAN, P. *Probabilistic versus Fuzzy Reasoning*. *Uncertainty in Artificial Intelligence*, Editor: L.N. Kanal e J.F. Lemmer. Amsterdam, Elsevier Science Publishers, 1986.

CHOOBINEH, F. e BEHRENS, A. *Use of Intervals and Possibility Distributions in Economic Analysis*. *Journal of the Operational Research Society*, v. 43, N 9, 1992. p. 907-918.

CHU, F. *Quantitative Evaluation of University Teaching Quality - An Application of Fuzzy Set and Aproximate Reasoning*. *Fuzzy Sets and Systems*, 1990. p. 1-11.

CIVANLAR, M.R. e TRUSSELL, H.J. *Constructing Membership Functions Using Statistical Data*. *Fuzzy Sets and Systems*, v. 18, 1986. p. 1-13.

CLEVELAND, W.P. The Inverse Autocorrelations of Time Series and their Applications. *Technometrics*, Vol. 14, N. 2, 1972. p. 277-298.

COMISSÃO DE REFORMA DO SISTEMA EDUCATIVO. Medidas que Promovam o Sucesso Educativo. Textos das Comunicações e Conclusões do Seminário Realizado em Braga, Ministério da Educação/Gabinete de Estudos e Planeamento, 1988. 191p.

COSTA, M.L.T. Função Social da Avaliação in Sistema de Ensino em Portugal, Coordenação de Manuela Silva e M. Isabel Tamen, Fundação Calouste Gulbenkian, Lisboa, 1981.

COURTOIS, P., On Time and Space Decomposition of Complex Structures, *Communications of ACM*, Vol. 28, N.6. 1985.

COX, D.R. The Analysis of Multivariate Binary Data. *Applied Statistics*, v. 21, N. 2, 1972. p. 113-120.

CROSBY, R.W., Toward a Classification of Complex Systems, *European Journal of Operational Research*, N. 30, 1987. p. 291-293.

CUNHA, L.V. in Sistema de Ensino em Portugal, Coordenação de Manuela Silva e M. Isabel Tamen, Fundação Calouste Gulbenkian, Lisboa, 1981.

DEMPSTER, A.P.; LAIRD, N.M. e RUBIN, D.B. Maximum Likelihood for Incomplete Data via the EM Algorithm. *Journal of the Royal Society, B*, N. 39, 1977. p. 1-38.

DEVI, B.B. e SARMA, V.V.S. Estimation of Fuzzy Memberships from Histograms. *Information Sciences*, nº 35, 1985. p. 43-59.

DEWEY, J. Vida e Educação, Melhoramentos, São Paulo, 1971, 7a. Ed.

DOBSON, A.J. An Introduction to Generalized Linear Models. London, Chapman and Hall, 1990. 174p.

DOMBI, J., Basic Concepts for a Theory of Evaluation: The Aggregative Operator, *European Journal of Operational Research*, Vol. 10, 1982. p. 282-293.

- DOMBI, J. Membership Function as an Evaluation. *Fuzzy Sets and Systems*, N.35, 1990. p. 1-21.
- DYKSTRA, R.L. e LEMKE, J.H. Duality of I Projections and Maximum Likelihood Estimation for Log-Linear Models Under Cone Constraints. *Journal of the American Statistical Association*, v. 83, N. 402, jun. 1988. p. 546-554.
- DUBOIS, D. e PRADE, H. *Fuzzy Sets and Systems: Theory and Applications*. New York, Academic Press, 1980. 393p.
- DUBOIS, D. e PRADE, H. Fuzzy Sets and Statistical Data. *European Journal of Operational Research*, v. 25, N, 3, jun. 1986. p. 345-356.
- DUBOIS, D. e PRADE, H. Fuzzy Sets, Probability and Measurement. *European Journal of Operational Research*, N. 40, 1989. p. 135-154.
- DUNN, J. E. e BERG, R.C. Effect of Iteratively Reweighted Least Squares (IRLS) on Teaching a Course in Quantal/Categorical Response. *Proceedings of Twelfth Annual SAS User's Group International Conference*, SAS Institute, Cary, N.C., 1987. p. 1047-1052.
- DURKHEIN, E. *Educação e Sociologia, Melhoramentos*, São Paulo, 1972, 8a. Ed.
- ECHENIQUE, M. "Modelos Matemáticos de la Estructura Espacial Urbana. Aplicaciones en America Latina", Buenos Aires, Ediciones Siap, 1975.
- EFRON, B. Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association*, v. 83, N. 402, jun. 1988. p. 414-425.
- ESTABLET, R. A Escola, *Tempo Brasileiro*, N. 35, Rio de Janeiro, 1973.
- EVERITT, B.S. e DUNN, G. *Advance Methods of Data Exploration and Modelling*. London, Heinemann Educational Books, 1983. 253p.
- FAIRLEY, W.B.; IZENMAN, A.J. e BAGCHI, P. Inference for Welfare Quality Control Programs. *Journal of the American Statistical Association*, v. 85, N.411, set. 1990. p. 874-890.

- FAZENDEIRO, A. et ali. Critérios de Planeamento da Rede Escolar, Lisboa, Série E: Rede Escolar, Ministério da Educação/Gabinete de Estudos e Planeamento, 1990. 61p.
- FIENBERG, S.E. The Analysis of Cross-Classified Categorical Data. Cambridge, MIT Press, 1989. 198p.
- FISHMAN, G.S. Principles of Discrete Event Simulation. Editor: Harold Chestnut, John Wiley, North Carolina, 1979. 514p.
- FOX, J. Knowledge, Decision Making, and Uncertainty. Artificial Intelligence and Statistics. Editor: William A. Gale. Reading, Addison-Wesley, 1986. 418p.
- FREITAG, B. Escola, Estado & Sociedade, 6ª Ed., São Paulo, Editora Moraes, 1986.
- FREITAS, B. Escola, Estado e Sociedade. São Paulo, Ed. Moraes, 1986. 142p.
- FRIEDMAN, M. e FRIEDMAN, R. Liberdade para Escolher, Publicações Europa-América, Lisboa, Portugal, 1980.
- FRIEDMAN, M. Capitalismo e Liberdade, São Paulo, Ed. Nova Cultural, 1985.
- GAY, L.R. Educational Research: Competencies for Analysis and Application. Columbus, Charles E. Merrill Publishing Company, 1976. 353p.
- GIL, M.G. Fuzziness and Loss of Information in Statistical Problems. IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-17, nº 6, nov/dez. 1987. p. 1016-1025.
- GILCHRIST, R. Introduction: Glim and Generalized Linear Models. Proceedings Generalized Linear Models, Lancaster, 1985. p. 1-5.
- GILULA, Z.; KRIEGER, A.M. e RITOV, Y. Ordinal Association in Contingency Tables: Some Interpretative Aspects. Journal of the American Statistical Association, v. 83, N. 402, jun. 1988. p. 540-545.

- GOLDBERGER, A.S. Structural Equation Models: An Overview. Quantitative Studies in Social Relations. Editor: Arthur S. Goldberger e Otis Duncan, New York, Seminar Press, 1973. 358p.
- GOLDSTEIN, H. Some Models for Analysing Longitudinal Data on Educational Attainment. Journal of the Royal Statistical Society, N. 142, parte 4, 1979. p. 407-442.
- GOLDSTEIN, H., Models for Multilevel Response Variables with an Application to Growth Curves , in Multilevel Analysis of Education Data. Ed.by R. Darwel Bock, Academic Press, 1989. p. 107-126.
- GOLDSTEIN, H., Multilevel Models in Educational and Social Research, London, Griffin, Oxford University Press, 1987.
- GOLDSTEIN, H., Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares, Biometrika, 73, p. 43-46.
- GORDON, T.J. Chaos in Social Systems. Technological Forecasting and Social Change, N. 42, 1992. p. 1-15.
- GRAY, H.L.; KELLEY,G.D. e McINTIRE,D.D. A New Approach to ARMA Modelling. Communications in Statistics, Vol 87, N. 1. 1978.p. 1-78.
- GRILICHES, Z. e MASON, W.M. Education, Income, and Ability. Quantitative Studies in Social Relations. Editor: Arthur S. Goldberger e Otis D. Duncan, New York, Seminar Press, 1973. 358p.
- GUPTA. M.M. Cognition, Perception and Incertainty., in Fuzzy Logic in Knowledge-Based Systems, Decision and Control. Ed. por M.M. Gupta e T. Yamakwa, North-Holland, 2. ed. 1992.
- HAND, D.J. Patterns in Statistical Strategy. Artificial Intelligence and Statistics. Editor: William A. Gale. Reading, Addison-Wesley, 1986. 418p.
- HARRISON, P.J. e STEVENS, C.F. Bayesian Forecasting (with discussion), J. Royal Statistical Soc. v. 38, N.3. 1976.

- HARTIGAN, J.A. Clustering Algorithms. New York, John Wiley, 1975. 352p.
- HAUSER, R.M. Disaggregating a Social-Psychological Model of Educational Attainment. Quantitative Studies in Social Relations. Editor: Arthur S. Goldberger e Otis D. Duncan, New York, Seminar Press, 1973. 358p.
- HESHMATY, B. e KANDEL, A. Fuzzy Linear Regression and its Applications to Forecasting in uncertain Environment. Fuzzy Sets and Systems, 15, 1985, p. 159-191.
- HOKSTAD, P. A Method for Diagnostic Checking of Time Series Models. Journal of Time Series Analysis, Vol. 4, 1983.p. 177-184.
- HOPEMAN, R.J. Análise de Sistemas e Gerência de Operações. Petrópolis, Vozes, 1974. 398p.
- IBRAHIM, J.G. Incomplete Data in Generalized Linear Models. Journal of the American Statistical Association, v. 85, N. 411, set. 1990. p. 765-769.
- INSTITUTO NACIONAL DE ESTATÍSTICA. Classificação das Actividades Económicas Portuguesas por Ramos de Actividade. Lisboa, 1973. 106p.
- INSTITUTO NACIONAL DE ESTATÍSTICA. XII Recenseamento Geral da População. XII Recenseamento Geral da Habitação 1981, Lisboa, 1984. 411p.
- INSTITUTO NACIONAL DE ESTATÍSTICA. Estatística da Educação. Lisboa, 1989. 192p.
- INSTITUTO NACIONAL DE ESTATÍSTICA. Estatísticas Demográficas Continente, Açores e Madeira 1989, Lisboa, 1990. 277p.
- ISAAC, S. e MICHAEL, W.B. Handbook in Research and Evaluation. San Diego, Edits, 1982. 234p.
- KANDEL, A. Fuzzy Mathematical Techniques with Applications. Reading, Addison-Wesley, 1986.

- KANDEL, A. e HESHMATY, B. Using Fuzzy Linear Regression as a Forecasting Tool in Intelligent Systems, in *Fuzzy Logic in Knowledge-Based Systems, Decision and Control*, ed. M.M. GUPTA e T. YAMAKAWA, North-Holland, 2d. Ed., 1991, p. 361-366.
- KASS, R.E. e STEFFEY, D. Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models). *Journal of the American Statistical Association*, v. 84, N. 407, set. 1989. p. 717-726.
- KAUFMANN, A. On the Relevance of Fuzzy Sets for Operations Research. *European Journal of Operational Research*, v.25, N. 3, jun. 1986. p. 330-335.
- KEEVES, J. Changing Standards of Performance and the Quality of Education. Paris, OCDE, Directorate for Social Affairs, Manpower and Education, SME/ET/86.19, 1986. 32p.
- KICKERT, W.J. *Fuzzy Theories on Decision Making. A Critical Review*. Leiden, 1978. 182p.
- KOSKO, B. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. New Jersey, Prentice-Hall. 1992. 449p.
- KOSKO, B. *Fuzzy Thinking: The New Science of Fuzzy Logic*. New York, Hyperion, 1994, 318p.
- KUSIAK, A. e CHOW, W.S. An Efficient Cluster Identification Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-17, N. 4, jul/ago. 1987. p. 696-699.
- LAIRD, N.; LANGE, N. e STRAM, D. Maximum Likelihood Computations with Repeated Measures Application of the EM Algorithm. *Journal of the American Statistical Association*, v. 82, N. 397, mar. 1987. p. 97-105.
- LEE, C.C. Fuzzy Logic in Control Systems: Fuzzy Logic Controller - Part I. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 20, N. 2, mar/abr. 1990. p. 404-418.

- LEE, C.C. Fuzzy Logic in Control Systems: Fuzzy Logic Controller - Part II. IEEE Transactions on Systems, Man, and Cybernetics, v. 20, N. 2, mar/abr. 1990. p. 419-435.
- LEE, H.L. Policy Decision Modeling of the Costs and Outputs of Education in Medical Schools. Operations Research, v. 35, N. 5, set/out. 1987. p. 667-683.
- LENK, P.J. The Logistic Normal Distribution for Bayesian, Non Parametric, Predictive Densities. Journal of the American Statistical Association, v. 83, N. 402, jun. 1988. p. 509-516.
- LEONARDO, R. e SANTOS, R.F. Indicadores de Custos das Infraestruturas para o Ensino Básico e Secundário. Lisboa, Série F: Custos, Valias e Financiamentos, Ministério da Educação/Gabinete de Estudos e Planeamento, 1990. 133p.
- LEVI, I. Possibility and Probability. Erkenntnis. v. 31, 1989. p. 365-386.
- LEWANDOWSKY, S. e SPENCE, I. Discriminating Strata in Scatterplots. Journal of the American Statistical Association, v. 84, N. 407, set. 1989. p. 682-688.
- LINDLEY, D.V. e SMITH, A.F.M., Bayes Estimates for Linear Model (with discussion), Journal of The Royal Statistical Society, Ser. B, 34, p. 1-41.
- LING, C. e PEN-GUOZHONG. A Predictive Method of Teacher's Structure in China's University (1985-2000). Management Science, v. 33. N. 6, jun. 1987. p. 739-749.
- LITTLE, R.J.A. e RUBIN, D.B. On Jointly Estimating Parameters and Missing Data by Maximizing the Complete Data Likelihood. The American Statistician, v. 37, N. 3, ago. 1983. p. 218-220.
- LONGFORD, N.T. Statistical Modelling of Data from Hierarchical Structures Using Variance Component Analysis. Proceedings Generalized Linear Models, Lancaster, 1985. p. 112-119.

- LOONEY, C.G. Fuzzy Petri Nets for Rule-Based Decision Making. IEEE on Transactions on Systems, Man, and Cybernetics, v. 18, N 1, jan/fev. 1988. p. 178-190.
- LOPES, E.R. et all. Portugal: O Desafio dos Anos Noventa. Lisboa, Presença, 1989. 317p.
- MABUCHI, S. An Approach to the Comparison of Fuzzy Subsets with an  $\alpha$ -cut Dependent Index. IEEE Transactions on Systems, Man, and Cybernetics, v. 18, N. 2, mar/abr. 1988. p. 264-272.
- MAJLUF, N. Education in Chile, in Operational Research'87, Ed. G.K. Rand, North Holland, 1988.
- MANNHEIM, K. e STEWART, W.A.C. Introdução à Sociologia da Educação, Ed. Cultrix, São Paulo, 1972.
- MANTON, K.G.; WOODBURY, M.A.; STALLARD, E.; RIGGAN, W.B.; CREASON, J. P. e PELLOM, A.C. Empirical Bayes Procedures for Stabilizing Maps of U.S. Cancer Mortality Rates. Journal of the American Statistical Association, v. 84, N. 407, set. 1989. p. 637-650.
- MCCULLAGH, P. e NELDER, J.A. Generalized Linear Models. London, Chapman and Hall, 1989. 511p.
- MODIS, T. e DEBECKER, A. Chaoslike States Can Be Expected Before and After Logistic Growth. Technological Forecasting and Social Change, N. 41, 1992. p. 111-120.
- MOORE, P.G. Higher Education: The Next Decade. Journal of the Royal Statistical Society, A, N. 146, Parte 3, 1983. p. 213-245.
- MORRIS, C.N. Parametric Empirical Bayes Inference: Theory and Applications. Journal of the American Statistical Association, v. 78, N. 381, mar. 1983. p. 47-65.
- MURTEIRA, B.J.F. Estatística: Inferência e Decisão. Imprensa Nacional - Casa da Moeda, Lisboa, 1988, 394 p.

- MURTHY, C.A.; PAL, S.K. e MAJUMDER, D.D. Representation of Fuzzy Operators Using Ordinary Sets. *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-17, N. 5, set/out. 1987. p. 840-847.
- MURTHY, S.V. Fuzzy Sets and Typicality Theory. *Information Sciences*, v. 51, N. 1, jun. 1990. p. 61-93.
- MUTHÉN, B. e SATORRA, A. Multilevel Aspects of Varying Parameters in Structural Models, in *Multilevel Analysis of Education Data*. Ed. by R. Darwel Bock, Academic Press, 1989. p. 87-99.
- NELDER, J.A. A Reformulation of Linear Models. *Journal of the Royal Statistical Society, A*, N. 140, Parte 1, 1977. p. 48-76.
- NERVOLE, M.; GREYER, D.M. e CARVALHO, J.L. *Analysis of Economic Time Series: A Synthesis*. New York, Academic Press, 1979.
- NOVAES, A. G. *Modelos em Planejamento Urbano, Regional e de Transportes*. São Paulo, Ed. Edgard Blücher, 1982. 220p.
- OCDE. *Papel da Educação no Desenvolvimento Sócio-Económico*. Lisboa, Ministério da Educação/Gabinete de Estudos e Planeamento, 1987. 122p.
- OLDFORD, R.W. e PETERS, S.C. *Implementation and Study of Statistical Strategy. Artificial Intelligence and Statistics*. Editor: William A. Gale. Reading, Addison-Wesley, 1986. 418p.
- PARK, K.S. Fuzzy Set Theoretic Interpretation of Economic Order Quantity. *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-17, N. 6, nov/dez. 1987. p. 1082-1087.
- PATERSON, L. *Entry to University by School Leavers in A Guide to ML3 for New Users*, 2nd. Ed., Edited by G. Woodhouse, Multilevel Models Project, Institute of Education, University of London, 1993, 124p.
- PEDRYCZ, W.; DIAMOND, J. e McLEOD, R.D. *A Fuzzy Cognitive Structure: Foundations, Applications, and VLSI Implementation in. Fuzzy, Holographic,*

and Parallel Intelligence. Editor: Branko Soucek e the IRIS GROUP. New York, John Wiley & Sons, 1992. 349p.

PELAUMER, P. Confidence Intervals for Population Projections Based on Monte Carlo Methods. *International Journal of Forecasting*, N. 4, 1988. p. 135-142.

PENG, X.T.; KANDEL, A. e WANG, P. Concepts, Rules, and Fuzzy Reasoning: A Factor, Space Approach. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 21, N. 1 jan/fev. 1991. p. 194-205.

PINDYCK, R.S. e RUBINFELD, D.L. *Econometric Models and Economic Forecasts*, McGraw Hill Book Company, New York, 1981, 629p.

POPPER, K.R. *A Lógica da Pesquisa Científica*. São Paulo, Editora Cultrix, 1972. 566p.

POULANTZAS, N. *Escola em Questão*, *Tempo Brasileiro*, N. 35, Rio de Janeiro, 1973.

RAUNDENBUSH, S.W. e BRYK, A.S. Quantitative Models for Estimating Teacher and School Effectiveness, in *Multilevel Analysis of Education Data*. Ed. by R. Darwel Bock, Academic Press, 1989. p. 205-232.

REGE, A. e AGOGINO, A.M. Topological Framework for Representing and Solving Probabilistic Inference Problems in Expert Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 18, N. 3, maio/jun. 1988. p. 402-414.

REPORT OF THE ROYAL STATISTICAL SOCIETY'S WORKING PARTY ON PROJECTIONS OF STUDENT NUMBERS IN HIGHER EDUCATION. *Journal of the Royal Statistical Society*, A, N. 148, Parte 3, 1985. p. 175-213.

RUDD, E. The Educational Qualifications and Social Class of the Parents of Undergraduates Entering British Universities in 1984. *Journal of the Royal Statistical Society*, A, N. 150, Parte 4, 1987. p. 346-372.

SAATY, T.L. Scaling the Membership Function. *European Journal of Operation Research*, N. 25, 1986. p. 320-329.

- SAKAWA, M.; YANO, H. e YUMINE, T. An Interactive Fuzzy Satisficing Method for Multiobjective Linear Programming Problems and its Application. IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-17, N. 14, jul/ago. 1987. p. 654-661.
- SÃO PEDRO, M.E. e CASTANHEIRA, C. A Origem Sócio Económica do Aluno e o Sucesso Escolar. Lisboa, Ministério da Educação/Gabinete de Estudos e Planeamento, 1989. 146p.
- SÃO PEDRO, M.E. e CASTANHEIRA, C. Progresso Escolar no Ensino Básico: Um Estudo Comparativo. Lisboa, Série C: Fluxos Escolares, Ministério da Educação/Gabinete de Estudo e Planeamento do Ministério da Educação, 1989. 72p.
- SCHULTZ, T. W. O Capital Humano - Investimentos em Educação e Pesquisa, Ed. Zahar, Rio de Janeiro, 1971.
- SHUMWAY, R.H. Applied Statistical Time Series Analysis. Englewood Cliffs, Prentice-Hall, 1988. 379p.
- SILVA, M. e TAMEN, M.I. Sistema de Ensino em Portugal. Lisboa, Fundação Calouste Gulbenkian, 1981. 696p.
- SOUCEK, B. Fuzzy, Holographic, and Parallel Intelligence. New York, John Wiley, 1992. 349 p.
- SOUZA, R.C. e BRASIL, G.H. Controle Automático da Produção Industrial pelo Modelo Bayesiano de Estados Múltiplos.com CUSUM. Anais do I SBMQC, Santa Maria, RS. 1987.
- SPIEGELHALTER, D.J. A Statistical View of Uncertainty in Expert Systems. Artificial Intelligence and Statistics. Editor: William A. Gale. Reading, Addison-Wesley, 1986. 418p.
- STEINBERG, D. Interpretation and Diagnostics of the Multinomial Logistic Regression. Proceedings of The Twelveth Annual SAS User's Group International Conference. SAS Institute, Cary, N.C., 1987. p. 1071-1076.

- STEPHANON, H.E. e SAGE, A.P. Perspectives on Imperfect Information Processing. IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-17, N. 5, set/out. 1987. p. 780-798.
- STIPANICEV, D. e EFSTATHIOU, J. Fuzzy Reasoning in Planning, Decision Making, and Control: Intelligent Robots, Vision, Natural Language. in Fuzzy, Holographic, and Parallel Intelligence. Editor: Branko Soucek e The IRIS GROUP. New York, John Wiley & Sons, 1992. p.93-132.
- STUKEL, T. Implementation of an Algorithm for Fitting a Class of Generalized Logistic Models. Proceedings Generalized Linear Models, Lancaster, 1985. p. 160-167.
- STUKEL, T.A. Generalized Logistic Models. Journal of the American Statistical Association, v. 83, N. 402, jul. 1988. p. 426-431.
- SWAMINATHAN, H. Multilevel Data Analysis: A Discussion, in Multilevel Analysis of Education Data. Ed. by R. Darwel Bock, Academic Press, 1989. p.277-282.
- TANAKA, H. Fuzzy Data Analysis by Possibilistic Linear Model. Fuzzy Sets and Systems, 24, 3, 1987. p. 363-375.
- TANAKA, H., UEJIMA, S. e ASAI, K. Linear Regression Analysis with Fuzzy Model", IEEE Transactions on System, Man and Cybernetics, V. SMC-12, N.6, Novembro/Dezembro 1982. p. 903-906.
- TANAKA, H.; HAYASHI, I. e WATADA, J. Possibistic Linear Regression Analysis for Fuzzy Data. European Journal of Operational Research, N. 40, 1989. p. 389-396.
- TAVARES, L.V. Desenvolvimento de Sistemas Educativos: Modelos e Perspectivas, GEP, Ministério da Educação, Lisboa, 1991.
- TAVARES, L.V. Investigação Operacional, AEIST, Lisboa, 1975.
- TAYLOR, P. E. e THOMAS, M.E. Short Term Forecasting: Houses of Courses. J. of Opl. Res. Soc. N.33. 1982.

- TAZZAN, M. Análise Quantitativa de Sistemas. III Escola Brasileiro-Argentina de Informática. Curitiba, 1988. 169p.
- THISTED, R.A. Representing Statistical Knowledge for Expert Data Analysis Systems. Artificial Intelligence and Statistics. Editor: William A. Gale. Reading, Addison-Wesley, 1986. 418p.
- TOMA, L. e GHEORGHE, E. Equilibrium and Disorder in Human Decision-Making Processes: Some Methodological Aspects within the New Paradigm. Technological Forecasting and Social Change, N. 41, 1992. p. 401-422.
- TOMBERLIN, T.J. Predicting Accident Frequencies for Drivers Classified by Two Factors. Journal of the American Statistical Association, v. 83, N. 402, jun. 1988. p. 309-321.
- TRONCALE, L. The Systems Sciences: What Are They? Are They One, or Many? European Journal of Operational Research, N. 37, 1988. p. 8-33.
- TUCKMAN, B.W. Conducting Educational Research, New York, Harcourt Brace Jovanovich, 1978. 479 p.
- TVERSKY, A. e KAHNEMAN, D. The Forming of Decisions and the Psychology of Choice. Science, v. 211, jan. 1981. p. 453-458.
- WEI, G.C.G. e TANNER, M.A. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. Journal of the American Statistical Association, v. 85, N. 411, set. 1990. p. 699-704.
- WHEELWRIGHT, S.C. e MAKRIDAKIS, S., Forecasting Methods for Management, 4th ed., John Wiley & Sons, Inc., Republic of Singapore, 1985.
- WHITE III, C.C. e WHITED, J. Markov Decision Process. European Journal of Operational Research, N. 39, 1989.
- WHITEHEAD, J. Fitting Cox's Regression Model to Survival Data Using GLIM. Applied Statistics, v. 29, N. 3, 1980. p. 268-275.

- WITTINK, D.R. The Application of Regression Analysis. Boston, Allyn and Bacon, 1988. 324p.
- WONG, G.Y. e MASON, W.M. The Hierarchical Logistic Regression Model for Multilevel Analysis. Journal of the American Statistical Association, set. 1985. p. 513-523.
- XAVIER, T.M.B.S. Modeles Lineaires Generalises dans l'Etude de la Pluviometrie au Nord-Est du Brésil. PUB. IRMA, v. 17, N. 5, 1989. 40p.
- YAGER, R.R. Expected Values from Probabilities of Fuzzy Subsets. European Journal of Operational Research, v. 25, N. 3, jun. 1986. p. 336-344.
- ZADEH, L. Fuzzy Sets. Information and Control. v. 8, N. 3, 1965, p. 338-353.
- ZADEH, L. Outline of a New Approach to the Analysis of Complex Systems and Decision Process, IEEE Transactions on Systems, Man and Cybernetics, v. SMC-3, N. 1, Janeiro, 1973. p.28-44.
- ZADEH, L. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning - Part I", Information Sciences, N.8, 1975. p. 199-249.
- ZADEH, L. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning - Part II, Information Sciences, N.8, 1975. p. 301-357.
- ZADEH, L. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning - Part III, Information Sciences, N.9, 1975. p. 43-80.
- ZADEH, L. Is Probability Theory Sufficient for Dealing with Uncertainty in AI: A Negative View. Uncertainty in Artificial Intelligence. Editor: L. N. Kanal e J. F. Lemmer. Amsterdam, Elsevier Science Publishers, 1986.
- ZADEH, L. e KACPRZYK, J. Fuzzy Logic for the Management of Uncertainty. New York, John Wiley e Sons. 1992. 676p.
- ZIMMERMANN, H.J. Fuzzy Set Theory and its Applications. Boston, Kluwer-Nijhoff, 1985. 363p.

ZINSER, K. E HENNEMAN, R.L. Development and Evaluation of a Model of Human Performance in a Large-Scale System. IEEE Transactions on Systems, Man, and Cybernetics, v. 18, N. 3, maio/junho 1988. p. 367-375.