


SISTEMAS DE APOIO À DECISÃO MÉDICA - DIAGNÓSTICO  
DIFERENCIAL E GERAÇÃO DE PLANOS DE INVESTIGAÇÃO

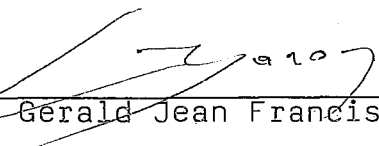
- RICARDO JOSÉ MACHADO -

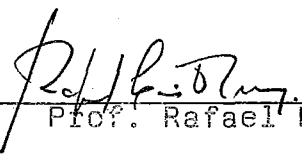
TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE  
PÓS GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE  
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO  
DO GRAU DE DOUTOR EM CIÊNCIAS (D.SC.) EM ENGENHARIA DE SISTEMAS  
Aprovada por:

  
Prof. João Lizardo R.H. de Araújo  
(Orientador)

  
Prof. Basílio de Bragança Pereira

  
Prof. Flávio Fonseca Nobre

  
Prof. Gerald Jean Francis Banon

  
Prof. Rafael Leite Luna

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 1985

---

**MACHADO, RICARDO JOSÉ**

Sistemas de Apoio à Decisão Médica - Diagnóstico Diferencial e Geração de Planos de Investigação [Rio de Janeiro] 1985.

xvi , 421 p. 29,7cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas, 1985)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Sistemas de Apoio à Decisão I. COPPE/  
UFRJ II. Título (Série)

---

À Sheila e  
as minhas filhas Vanessa e Adrienne

## AGRADECIMENTOS

Ao Prof. João Lizardo R.H. de Araújo pela orientação e pelo incentivo sempre presente.

Ao cardiologista Rogério Brant M. Chaves pela co-orientação na parte médica.

A minha esposa pelo apoio, renúncia e compreensão.

À COPPE pela oportunidade.

Ao IASERJ pelo fornecimento dos dados médicos.

À IBM Brasil pela utilização de suas facilidades computacionais.

Aos meus ex-gerentes Carlos A.G. Cordovil e Herbert P. Fortes F<sup>o</sup> pelo estímulo.

Aos prof<sup>s</sup> B. McA. Sayers (Imperial College) e F.T. de Dombal (Universidade de Leeds) pelas sugestões apresentadas.

A Stavna Uchoa pelo excelente trabalho datilográfico.

Resumo da Tese Apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

**SISTEMAS DE APOIO À DECISÃO MÉDICA - DIAGNÓSTICO  
DIFERENCIAL E GERAÇÃO DE PLANOS DE INVESTIGAÇÃO**

**Ricardo José Machado**

**Março de 1985**

**Orientador:** João Lizardo R.H. de Araújo

**Programa :** Engenharia de Sistemas e Computação

O presente trabalho visa estudar a aplicação de técnicas de apoio à decisão em medicina, particularmente nas áreas de diagnóstico diferencial e geração de planos de investigação médica. Selecionamos para estudo experimental a relevante e complexa área representada pela doença aterosclerótica coronariana.

Empregamos modelos bayesianos de classificação diagnóstica visando extrair o máximo de informação dos dados médicos de pacientes com suspeita de doença coronariana. Para tanto, as interações entre sintomas foram consideradas através da poderosa e flexível aproximação baseada na formação de aglomerados compostos por descritores interdependentes, a qual permitiu a obtenção de significativos ganhos de desempenho. Algoritmos eficientes, baseados na representação esparsa das probabilidades condicionais conjuntas, foram projetados e testados para estimação de parâmetros e inferência diagnóstica. A eficiência e o espaço de memória destes algoritmos depende basicamente do tamanho do banco de dados de treinamento, não sendo afetados, em situações práticas, pelo crescimento exponencial do número de probabilidades condicionais, experimentado quando consideramos interações mais fracas entre sintomas.

Propomos finalmente um gerador de planos de investigação médica, baseado em técnicas de programação heurística, capaz de selecionar a melhor seqüência de exames para um paciente específico, de acordo com a estratégia de investigação selecionada pelo médico e a precisão requerida para o diagnóstico.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

**MEDICAL DECISION SUPPORT SYSTEMS - DIFFERENTIAL DIAGNOSIS  
AND CREATION OF INVESTIGATION PLANS**

**Ricardo José Machado**

**March - 1985**

**Chairman:** João Lizardo R. H. de Araújo

**Department:** Engenharia de Sistemas e Computação

The application of decision support techniques in medicine is examined focusing particularly the areas of differential diagnosis and the development of optimal investigation plans for patients suspected of coronary artery disease.

Bayesian diagnostic models are employed aiming to extract the maximum information from patients medical data. Symptoms interactions are considered through the flexible and powerful approximation of clusters formed by interdependent variables, allowing significant gains in the classification performance. Efficient algorithms based on the sparse representation of joint conditional probabilities are designed and tested for parameters estimation and diagnostic inference. The efficiency and memory space of these algorithms depend on the training data set size, being not affected, in practical situations, by the exponential growth of the number of parameters experienced when weaker interactions between symptoms are considered.

Finally it is proposed an algorithm to create medical investigation plans based on heuristic programming techniques, capable to select the best sequence of tests to a specific patient, according to the investigation strategy selected by the physician and the required diagnosis accuracy.



II.3.1.4 - Funções Discriminantes .....	19
II.3.1.5 - Critérios de Decisão .....	20
II.3.1.6 - Principais Estudos/ Desempenho .....	21
II.3.1.7 - Bancos de Dados com Funções Analíticas.....	23
II.4 - SADD's Baseados em Métodos Seqüenciais de De- cisão .....	24
II.4.1 - Potencialidade dos Métodos da Teo- ria da Decisão em Medicina .....	24
II.4.2 - Classificação dos Modelos de Apli- cação da Teoria da Decisão .....	25
II.4.3 - Modelos de Diagnóstico Seqüencial .....	26
II.4.3.1 - Tipos de Regras de De- cisão .....	27
II.4.3.2 - Principais Trabalhos .....	28
II.4.4 - Modelos para Seleção de Tratamen- to .....	28
II.4.4.1 - Pressuposições .....	30
II.4.4.2 - Árvore de Decisão .....	30
II.4.4.3 - Estrutura de Utilidades .....	32
II.4.5 - Modelos Integrados de Diagnóstico e Tratamento .....	35
II.4.5.1 - Árvore de Decisão .....	36
II.4.5.2 - Estratégia .....	37
II.4.5.3 - Desempenho .....	37
II.4.5.4 - Dificuldades .....	38
II.5 - SAD's Baseados em Técnicas de Inteligência Artificial .....	39
II.5.1 - Estrutura da Base de Conhecimento .....	40
II.5.2 - Estrutura do Sistema .....	42
II.5.3 - Sistemas de Consulta Médica Usan- do Técnicas de I.A .....	44
II.5.4 - Comentários .....	48



II.6 - Discussão .....	49
II.6.1 - Obstáculos .....	49
II.6.2 - Pontos de Vista .....	50
II.6.3 - Pré-requisitos para o Sucesso .....	50
II.6.3.1 - Desempenho .....	51
II.6.3.2 - Aceitação .....	51
II.6.4 - Conclusões .....	53

### III. MODELOS BAYESIANOS DE CLASSIFICAÇÃO DIAGNÓSTICA

III.1 - Introdução .....	55
III.2 - Formulação .....	55
III.2.1 - Notação e Pressuposições .....	56
III.2.2 - Descrição da Regra de Diagnóstico .....	60
III.2.3 - Características / Pontos de Discussão .....	61
1. Flexibilidade .....	61
2. Estabilidade das Distribuições de Dados Diagnósticos .....	61
3. Vício de Seleção .....	62
4. Método de Multi-elegibilidade .....	62
5. Paradigma Diagnóstico .....	63
III.3 - Estimação de Parâmetros .....	63
III.3.1 - Número de Estimativas .....	63
III.3.2 - A Pressuposição de Independência dos Descritores .....	65
III.3.3 - Estimação de Parâmetros a Partir de um Banco de Dados Real .....	66
- Deficiências de Bancos de Dados Reais.....	67
- Anomalias na Estimação de Parâmetros .....	69

III.3.4 - Modelos com Classes Aglutinadas.....	70
III.3.5 - Estratégias de Calibração e Teste.....	71
III.4 - Influência Potencial da Dependência entre Descritores sobre o Desempenho .....	71
III.5 - Modelo Bayesiano de Classificação Diagnóstica Baseado em Aglomerados de Descritores Interdependentes .....	76
III.5.1 - Formação de Aglomerados de Descritores .....	77
III.5.2 - Estimação de Probabilidades Condicionais de Aglomerados .....	80
- Anomalias na Estimação com Banco de Dados Real .....	81
III.5.3 - Algoritmo de Estimação das Probabilidades Condicionais Conjuntas dos Aglomerados Baseado numa Representação Esparsa .....	82
- Algoritmo de Estimação .....	83
- Considerações quanto a Eficiência Computacional e Espaço em Memória .....	84
III.5.4 - Algoritmos de Inferência Diagnóstica em Modelos com Aglomerados de Descritores Baseados numa Representação Esparsa.....	85
- Algoritmo DIAGPREP - Inferência Diagnóstica .....	86
- Algoritmo DIAGNPREP - Estimação e Inferência Diagnóstica .....	88
- Eficiência Computacional dos Algoritmos de Inferência .....	90
III.5.5 - Considerações Quanto ao Desempenho de Modelos Baseados no Uso de Aglomerados de Descritores .....	91

## IV. GERADOR HEURÍSTICO DE PLANOS DE INVESTIGAÇÃO MÉDICA

IV.1 - Introdução .....	93
IV.2 - Objetivo do Gerador Heurístico de Planos de Investigação Médica.....	93
IV.3 - Atributos de Custo de Exames .....	94
IV.3.1 - Adição de Utilidades .....	95
IV.4 - Estratégias de Investigação .....	96
- Estratégias Elementares .....	96
- Estratégia Robusta .....	96
- Estratégias Compostas .....	98
IV.5 - Estrutura do Sistema .....	98
IV.6 - Notação .....	100
IV.7 - Pressuposições Básicas .....	101
IV.8 - Propriedades do Diagnóstico Bayesiano Se- qüencial .....	102
IV.9 - Descrição do Método de Geração de Planos de Investigação Médica .....	103
IV.9.1 - Programação Heurística - O Algorit <u>o</u> mo A*.....	103
- Grafo do Problema.....	103
- Admissibilidade .....	104
- Busca Heurística .....	105
- Algoritmo A* .....	106
- Teorema de Hart .....	106
IV.9.2 - Representação do Espaço de Estado no Problema de Investigação Diag- nóstica .....	107

IV.9.3 - Descrição do Estado do Problema.....	110
- Representação Computacional .....	111
IV.9.4 - Descrição do Estado Inicial .....	112
IV.9.5 - Operador $T$ de Geração de Estados	
Sucessores .....	112
IV.9.5.1 - Características Adici_	
onais do Modelo Incor_	
poradas ao Operador $T$ .....	113
IV.9.5.2 - Heurísticas para Limi_	
tação na Geração de	
Sucessores .....	114
Critérios Heurísticos	
para $\lambda_j$ .....	114
IV.9.6 - Definição dos Estados Terminais.....	116
IV.9.6.1 - Proposta de Alperovitch .....	117
- Estimacão do Parâmetro $\pi_{j\alpha}$ ..	117
IV.9.6.2 - Proposta Baseada no Banco	
de Dados .....	119
- Algoritmo de Determinacão de	
$p_{rev}(x)$ .....	119
IV.10 -Construcão da Funçãõ de Mérito para o Algorit-	
mo $A^*$ .....	120
IV.10.1 - Funçãõ Heurística para o Problema de	
Investigacão Diagnóstica .....	121
- Consistência da Funçãõ Heurística	
Proposta .....	123
IV.11 Caráter Iterativo do Gerador de Planos de In-	
vestigacão Médica .....	125
IV.12 - Características do Sistema .....	126
IV.13 - Aplicacões do Gerador de Planos de Investiga-	
çãõ .....	129

## V. DESCRIÇÃO DOS EXPERIMENTOS

V.1 - Introdução .....	131
V.2 - Coleta de Dados .....	132
- Formulário de Coleta de Dados .....	134
- Considerações sobre a Qualidade do Arquivo de Pacientes.....	138
V.3 - Construção de um Sistema de Suporte à Decisão.....	139
V.3.1 - Ambiente Computacional .....	139
- APLDI II .....	140
V.3.2 - Estrutura do Sistema de Suporte à De- cisão .....	141
V.3.2.1 - Subsistema de Diálogo .....	143
V.3.2.2 - Subsistema de Dados .....	143
V.3.2.2.1 - Principais Entidades .....	143
A - Pacientes .....	144
B - Descritores .....	145
C - Aglomerados de Descritores..	146
D - Observações.....	147
E - Exames .....	149
F - Variáveis Diagnósticas e Prognósticas.....	150
G - Classes Diagnósticas ou Prognósticas .....	150
H - Modelos Bayesianos com Pressuposição de Indepen- dência dos Descritores.....	154
I - Modelos Bayesianos Basea- dos em Aglomerados de Des- critores Interdependentes...	154
J - Geradores de Planos de In- vestigação Médica.....	155
K - Estratégias Elementares de Investigação.....	155
L - Estratégias Cômpostas de Investigação.....	156

V.3.2.3 - Subsistema de Modelos .....	156
V.3.2.3.1 - Funções Especialmente Desenvolvidas para o Sistema de Suporte à Decisão.....	157
A - Geração / Atualização da Base de Dados para uma Aplicação Médica Genérica .....	157
B - Funções Gerais .....	158
C - Gerenciamento dos Mo- delos Bayesianos de Classificação Diagnós- tica com Pressuposição de Independência dos Descritores .....	159
D - Gerenciamento de Mode- los Bayesianos de Clas- sificação Diagnóstica Baseados em Aglomera- dos de Descritores In- terdependentes .....	159
E - Gerenciamento de Gera- dores de Planos de In- vestigação Médica .....	160
F - Aplicação do Sistema na Prática Médica .....	161
V.4 - Experimentação .....	161
A - Exploração Inicial dos Dados .....	161
B - Definição dos Conjuntos de Calibração e Teste .....	162
C - Construção de Modelos Bayesianos de Clas- sificação Diagnóstica com Pressuposição de Independência dos Descritores .....	163
D - Estimação de Probabilidade a Priori .....	164
E - Estudo dos Modelos Bayesianos de Classi- ficação Diagnóstica com Pressuposição de	

de Independência dos Descritores .....	165
F - Estudo das Associações entre Descritores de Pacientes com Suspeita de Doença Coronariana .....	166
G - Construção de Modelos Bayesianos de Classificação Diagnóstica Usando Aglomerados de Descritores .....	167
H - Estudo dos Modelos Bayesianos de Classificação Diagnóstica Usando Aglomerados de Descritores Interdependentes .....	168
I - Construção de Geradores de Planos de Investigação Médica .....	169

## VI. RESULTADOS - APRECIÇÃO E ANÁLISE

VI.1 - Valor Diagnóstico dos Descritores .....	170
VI.2 - Estudo do Modelo Bayesiano de Classificação Diagnóstica Baseado na Pressuposição de Descritores Independentes .....	174
VI.2.1 - Estimativas de Probabilidades a Priori das Classes Diagnósticas.....	174
VI.2.2 - Estimativas de Probabilidades Condicionais das Observações.....	175
VI.2.3 - Avaliação do Desempenho dos Modelos com Pressuposição de Independência.....	175
VI.2.4 - Influência da Ecocardiografia sobre o Desempenho .....	176
VI.2.5 - Influência da Modalidade de Calibração e Teste.....	177
VI.2.6 - Desempenho em Função do Número de Classes Diagnósticas.....	178
VI.2.7 - Influência da Aglutinação de Classes a Posteriori.....	178
VI.2.8 - Avaliação do Desempenho de Modelos Construídos Segundo o Critério de Multielegibilidade .....	179

VI.2.9 - Credibilidade das Probabilidades a Posteriori.....	179
VI.2.10 - Influência do Valor da Probabilidade Condicional Mínima .....	180
VI.2.11 - Influência das Estimativas de Probabilidades a Priori sobre o Desempenho.....	180
VI.2.12 - Desempenho em Função do Número de Descritores Utilizados.....	181
VI.2.13 - Influência do Tamanho do Banco de Dados de Treinamento no Desempenho.....	182
VI.3 - Estudo das Associações entre Descritores de Pacientes com Suspeita de Doença Coronariana.....	183
VI.4 - Estudo de Modelos Bayesianos de Classificação Diagnóstica usando Aglomerados de Descritores Interdependentes .....	184
VI.4.1 - Avaliação do Desempenho na Estratégia "O Banco de Dados como Universo".....	184
VI.4.2 - Avaliação do Desempenho na Estratégia "O Banco de Dados como Amostra".....	185
VI.4.3 - Estimativas das Probabilidades Condicionais de Configurações de Observações .....	186
VI.4.4 - Avaliação do Desempenho de uma Bateria de Modelos com $\alpha$ Crescente.....	186
VI.4.5 - Influência da Probabilidade Condicional Mínima .....	187
VI.5 - Parâmetros para Construção de um Gerador Preliminar de Planos de Investigação .....	188



## VII. CONCLUSÕES E RECOMENDAÇÕES

VII.1 - Conclusões .....	298
VII.1.1 - Valor Diagnóstico dos Descritores Usados na Investigação da Doença Coronariana.....	299
VII.1.2 - Modelos Bayesianos de Classificação Diagnóstica com Pressuposição de Independência .....	299
VII.1.3 - Modelos Bayesianos de Classificação Diagnóstica Baseados em Aglomerados de Descritores Interdependentes .....	302
VII.1.4 - Geração de Planos de Investigação Médica .....	305
VII.2 - Recomendações .....	306
APÊNDICE 1 - Arquivo de Pacientes com Suspeita de Doença Coronariana.....	307
APÊNDICE 2 - Descrição da Base de Dados e Principais Programas Desenvolvidos .....	318
BIBLIOGRAFIA .....	402

---

---

## **I - INTRODUÇÃO**

---

---

## I.1 - AMBIENTE

Os problemas de decisão em medicina têm se tornado cada vez mais complexos nos recentes anos, tendo em vista a rápida expansão do conhecimento, a utilização de novas e revolucionárias tecnologias, bem como os altos valores colocados em jogo, seja em termos financeiros, seja em termos de vidas humanas. Assim se tínhamos 3000 doenças e síndromes conhecidas no início do século, segundo estimativas de Grémy [76], já em 1975 este número saltara para 30000, o número de medicamentos comercializados estava no entorno de 8000, crescendo a uma taxa de 300 anualmente, e o vocabulário médico correspondia a 200000 palavras ou expressões; processo este que tem resultado numa crescente especialização da medicina. O médico moderno tem a sua disposição mais instrumentos para combater as enfermidades do que em qualquer época da história, mas freqüentemente o conhecimento de como melhor usar estes instrumentos está faltando. Por outro lado, os crescentes custos da medicina moderna, aliados ao aumento da demanda por assistência médica tem criado graves problemas para a gestão dos sistemas nacionais de saúde no que tange à alocação de escassos recursos financeiros, à falta e má distribuição de pessoal e ao número inadequado de instalações. Outrossim, existe uma desproporção gritante entre o esforço dispendido no estudo do sistema-objeto (o paciente) e o esforço gasto para conhecer o sistema-sujeito (sistema-saúde).

Neste trabalho vamos focar o sistema sujeito, em particular quanto ao processo decisório e estudar em detalhe algumas técnicas de suporte à decisão médica usando computadores.

### I.1.1 - Níveis de Decisão no Sistema Saúde

Podemos classificar as decisões em medicina de acordo com os níveis de organização do sistema-saúde, a saber: níveis operacional, tático e estratégico. As decisões operacionais correspondem àquelas tomadas rotineiramente pelo médico no gerenciamento de pacientes, a saber: diagnóstico, prognóstico e sele

ção de terapia. As decisões táticas são tomadas nos níveis intermediários da organização, podendo ser por exemplo: vigilância epidemiológica, avaliação da eficácia de tratamentos alternativos, controle de qualidade, avaliação da produção de serviços de saúde, investigação de riscos associados a ações médicas, definição de normas, etc. As decisões estratégicas correspondem àquelas tomadas no nível hierárquico mais alto da organização, por exemplo: projeção da necessidade de recursos para uma população, avaliação do nível de saúde de uma população, planejamento estratégico etc.

Vamos nos dedicar nesse trabalho ao estudo das decisões a nível operacional ligadas ao gerenciamento de pacientes.

### I.1.2 - O Processo Decisório no Gerenciamento de Pacientes

Por **gerenciamento de pacientes** entendemos o processo de decisões seqüenciais que devem ser tomadas por um médico (ou outros profissionais de saúde) na determinação do curso de ação a ser seguido no tratamento de um paciente com um conjunto particular de sinais e sintomas.

Cada uma dessas decisões pode ser descrita usando-se o modelo de Simon [165] que subdivide o processo decisório em três fases:

- |   |              |  |
|---|--------------|--|
| } | INTELIGÊNCIA | - Pesquisa de dados e indícios que indiquem problemas.                                 |
|   | DELINEAMENTO | - Compreensão do problema, geração de cursos de ação e averiguação de sua viabilidade. |
|   | ESCOLHA      | - Seleção de um curso de ação particular para implementação.                           |

## SEQÜÊNCIA NORMAL DE DECISÖES

De modo geral a primeira e por vezes crucial decisão cabe ao paciente: decidir que está enfermo e procurar o médico. O médico procurado avaliará inicialmente sinais e sintomas, recorrendo então a exames laboratoriais de rotina. Estes estreitam a faixa de possibilidades diagnósticas e podem mesmo permitir que ele estabeleça um diagnóstico. Freqüentemente, contudo, o diagnóstico ainda não estará claro, e testes adicionais poderão ser indicados. Todavia estes testes costumam acarretar maiores riscos e desconfortos que os procedimentos iniciais de investigação, tendo o médico de decidir se a informação a ser ganha é suficiente para justificar a sujeição do paciente a tais riscos e custos. Presumindo-se que tenha estabelecido um diagnóstico com um grau razoável de certeza, o médico deverá selecionar um tratamento. Novamente ele terá de pesar vários fatores, alguns envolvendo a chance de melhoria ou cura, outros a possibilidade de dano. Depois de uma variedade de considerações incluindo julgamento de valores, ele tomará uma decisão quanto à terapia.

Como vemos o processo decisório em medicina envolve quase sempre o binômio risco-incerteza, em decorrência da imperfeição dos procedimentos.

### I.1.3 - Informação em Medicina

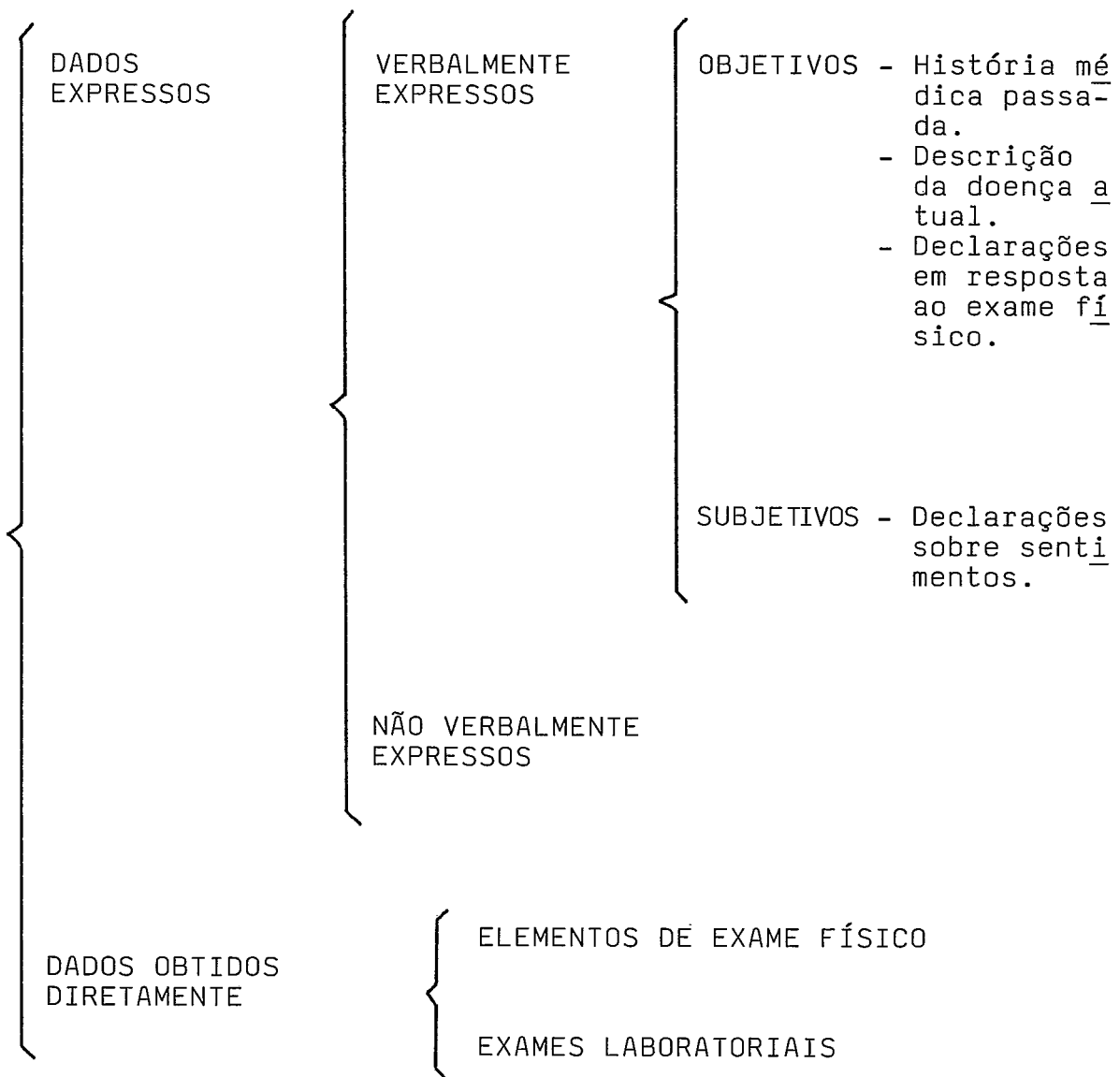
A informação desempenha um papel fundamental no processo decisório. Podemos classificar as informações envolvidas no gerenciamento de pacientes em **dados primários** e **dados secundários**.

Os dados primários são os que alimentam o processo decisório e correspondem a medições ou observações realizadas sobre o paciente. Os dados secundários correspondem a informações geradas em decorrência do processo decisório, tais como, diagnóstico, prognóstico, retroprognósticos, síndromes ou conclu-

sões terapêuticas.

### DADOS PRIMÁRIOS EM MEDICINA

Os dados primários em medicina podem ser classificados conforme o esquema a seguir:



Os dados primários em medicina, como por exemplo resultados de exames, raramente são patognomônicos. De um modo geral envolvem um maior ou menor grau de incerteza caracteriza

da através de parâmetros de **sensibilidade** (probabilidade de o teste apresentar resultado positivo dado que o paciente tem a doença) e **especificidade** (probabilidade de o teste apresentar resultado negativo dado que o paciente não tem a doença). Uma maneira interessante de visualizar esse fato é a chamada curva **ROC** ("Receiving Operating Characteristic"), ilustrada na figura (I.1).

A curva ROC é um gráfico que plota o percentual de falsos positivos contra o percentual de falsos negativos para diversos limiares de discriminação de um teste particular.

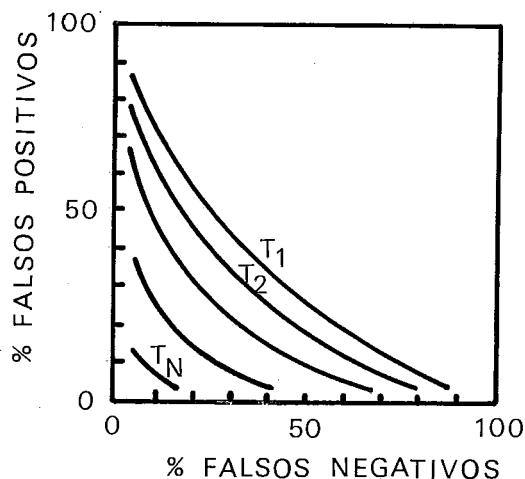


FIGURA I.1 CURVAS ROC PARA VÁRIAS TECNOLOGIAS

A localização da curva ROC no gráfico representa o poder discriminante do teste, sendo tanto maior este quanto mais próxima da origem dos eixos estiver a curva ROC (teste perfeito). O ponto sobre a curva representa o critério individual do observador ou a calibração do instrumento para positividade. Todas as tecnologias imperfeitas contêm uma necessidade inerente de balancear falsos positivos contra falsos negativos. O ponto ótimo de operação do observador pode ser determinado a partir dos custos dos diferentes tipos de erros de interpretação (teoria de detec

ção de sinais) [172].

#### I.1.4 - A Consulta Médica

A abordagem empregada pelos médicos no processo diagnóstico durante a consulta médica nunca chegou a ser um assunto profundamente analisado e conhecido. Estudos cognitivos recentes estão contudo permitindo elucidar um mecanismo do processo diagnóstico que parece ser válido consistentemente para toda a população médica.

Numa primeira fase da consulta, o médico experiente gera hipóteses diagnósticas específicas, bem antes de ter colhido a maioria dos dados de um caso particular. Por vezes, sexo, idade e queixa principal lhe são suficientes. O número de hipóteses é limitado entre 3 a 5 hipóteses segundo Elstein [46] e 4 a 11 segundo Gorry [93]. Em vez de convergir progressiva e sistematicamente na formulação do problema através de uma série de questões constritivas, o médico experiente parece saltar diretamente para esse pequeno arranjo de hipóteses, já no início da consulta. Essas hipóteses são ordenadas informalmente pelo médico, influenciando nessa ordenação critérios de probabilidade, gravidade, tratabilidade e novidade da doença, esta última funcionando como um elemento motivacional [46].

A segunda fase consiste na obtenção de informações adicionais que confirmem ou afastem a hipótese mais importante gerada na 1ª fase. A experiência clínica permite que o médico escolha diretamente os itens mais importantes de informação [60]. Esta segunda fase também é chamada interrogatório seletivo-seqüencial.



## I.2 - DECISÃO MÉDICA ASSISTIDA POR COMPUTADOR

A partir da década de 50 começaram a surgir estudos sobre aplicações de computadores em decisão médica, inicialmente na área de diagnóstico, e posteriormente em prognóstico e seleção de terapia.

Técnicas de diversas áreas foram utilizadas como Teoria Estatística da Decisão, Reconhecimento de Padrões, Álgebra de Boole, Taxonomia Numérica, Heurística e Inteligência Artificial.

Uma revisão sumarizada do trabalho desenvolvido nessa área será apresentada no Capítulo II. A seguir listamos os principais objetivos dos sistemas de apoio à decisão médica, seus benefícios potenciais e justificativas de emprego.

### I.2.1 - Benefícios Potenciais

O principal elemento de motivação para essas pesquisas sempre foi o grande potencial de benefícios que adviria caso fosse possível implementar sistemas eficazes de apoio à decisão médica. Croft [33] alinha entre os benefícios potenciais um vasto impacto sobre o custo e a disponibilidade de saúde, tanto em países desenvolvidos como nos países em desenvolvimento. Centros diagnósticos operados principalmente por pessoal paramédico usando modelos automáticos de diagnóstico poderiam processar milhares de pacientes por ano a um baixo custo por paciente. Tais clínicas se concentrariam na identificação de categorias abrangentes de doenças para grandes segmentos da população. Modelos de diagnóstico poderiam ser usados para calcular as probabilidades da presença de doenças em casos onde os relacionamentos entre doenças e sintomas são tão complexos que nenhum diagnóstico claro pode ser feito hoje. Szolovits [170] sugere que uma possibilidade radical para melhoria do sistema de assistência médica, seria utilizar o computador como um instrumento intelectual e dedutivo, como um consultor incorporado à própria

estrutura do sistema de saúde.

### I.2.2 - Objetivos/Justificativas de Sistemas de Apoio à Decisão

Os objetivos de sistemas de apoio à decisão médica são:

- A - Aumentar a exatidão da diagnose clínica através de enfoques que sejam sistemáticos, completos, e que possam integrar dados de diversas fontes (exames clínicos, sintomas, etc)
- B-- Aumentar a confiabilidade de decisões clínicas ao evitar que haja polarizações errôneas nas conclusões, ou que sejam descartadas hipóteses por falha de memória, e ao fazer com que critérios de decisão [8] sejam explícitos e, desta forma, reproduzíveis.
- C - Aumentar a eficiência de custo de testes e terapias, ao permitir pesar as despesas financeiras, de tempo e de inconvenientes, contra os benefícios e riscos de ações definitivas.
- D - Aumentar a compreensão sobre a estrutura do conhecimento médico e identificar inconsistências e inadequações nesse conhecimento.
- E - Aumentar nosso conhecimento do processo de decisão clínica a fim de melhorar o ensino médico e fazer programas de computadores mais efetivos e fáceis de entender.

Do ponto de vista da Medicina as razões imediatas que justificam o estudo da utilização de computadores são:

- A - O volume cada vez maior de conhecimento médico, que torna difícil a um clínico ter em mente todas as alternativas relevantes a serem consideradas durante o processo de consul-

ta.

- B - O tempo muito longo necessário para se obter uma formação médica especializada (estudo e experiência) bem como para manter-se atualizado.
- C - O desenvolvimento significativamente pior do médico inexperiente quando comparado com o de um especialista em sua área.
- D - A distribuição geográfica desigual dos profissionais médicos e das especialidades com concentração nos grandes centros, deixando vastas áreas sem assistência médica adequada.
- E - A existência de relativamente poucos profissionais competentes, os quais poderiam ser melhor utilizados se seu conhecimento acumulado pudesse ser "gravado" e distribuído por um sistema automatizado de consulta.
- F - O desmesurado encarecimento dos serviços médicos.
- G - A falta de capacidade humana para manipular num senso estatístico grandes quantidades de dados e para descobrir padrões complexos em situações multivariadas com muitas interdependências [51, 141].
- H - As limitações da experiência própria do médico individual [104, 146].
- I - As limitações cognitivas do ser humano que tornam a experiência recente mais acessível que a experiência distante [99, 146].
- J - A polarização motivacional resultando numa tendência a reconhecer evidência que suporte uma posição já tomada (Ex.: hipóteses geradas no início da consulta) e desprezar evidência que a contraria, especialmente se a posição foi adotada após um período de incerteza prolongado [99].

K - As limitações da memória humana de curto prazo que limita o número de diferentes hipóteses com que o ser humano pode trabalhar simultaneamente [99].

L - O conservantismo do ser humano ao agregar dados probabilísticos fazendo com que uma pessoa típica desperdice 50 a 80% de informação [43, 44].

### I.3 - OBJETIVOS DO TRABALHO DE TESE

O presente trabalho tem por objetivo explorar algumas técnicas de apoio à decisão médica na área de diagnóstico diferencial. Esta é uma situação freqüente em medicina onde temos um conjunto delimitado de classes diagnósticas exaustivas, numa das quais o paciente que estamos estudando deverá ser enquadrado. Vamos focar situações de difícil diferenciação entre as classes diagnósticas e que envolvam um conjunto razoavelmente grande de exames com uma ampla variedade em atributos como custo financeiro, riscos, sofrimento, etc.

O trabalho é semi-experimental, tendo sido escolhida como área médica de teste, a área de doença coronariana. Alguns objetivos específicos que buscamos são os seguintes:

- Exploração de modelos bayesianos de diagnóstico visando a um melhor aproveitamento da informação contida nos dados médicos.
- Proposta de um gerador heurístico de planos de investigação médica capaz de selecionar os melhores exames para um paciente específico, de acordo com a estratégia selecionada pelo médico e a precisão requerida para o diagnóstico.
- Definição e construção de um Sistema de Suporte à Decisão na área de Doença Coronariana envolvendo a construção de um banco de dados real de pacientes e modelos matemáticos em computador.

- Avaliação prática do desempenho de modelos bayesianos de diagnóstico e estudo de seu comportamento em pacientes com doença coronariana.

### **I.3.1 - Considerações sobre a Área Médica Selecionada: Doença Coronariana**

A doença coronariana ou doença cardíaca isquêmica é uma das principais doenças fatais que aflige a humanidade. Sua incidência tem crescido dramaticamente nos últimos 50 anos, sendo hoje reconhecida como a principal causa de mortes nos países desenvolvidos (aproximadamente 500000 óbitos por ano nos EUA) [163]. No Brasil a doença coronariana ocupa posição de destaque, intimamente relacionada com o desenvolvimento da região geográfica sob enfoque. Assim, no município de São Paulo (7247 óbitos/ano) ela é a principal causa de mortes e nos municípios do Rio de Janeiro e Porto Alegre representa a 2ª causa, segundo dados do IBGE de 1979.

A doença afeta homens na flor da idade; a idade média por ocasião do 1º infarto do miocárdio está nos meados dos 50. As mulheres são poupadas por cerca de 10 anos em relação aos homens. A importância da doença coronariana ultrapassa as altas taxas de morbidade e mortalidade a ela associadas, isto porque suas manifestações clínicas são imprevisíveis ou mesmo ausentes e de ocorrência aparentemente aleatória, porque a sua evolução é variável, e porque em 1/3 a 1/2 dos pacientes a morte é súbita e inesperada [163].

Em cerca de 99% dos casos, a doença coronariana se deve a alterações ateroscleróticas (Aterosclerose Coronariana). Outras causas de doença coronariana incluem várias formas de arterite, embolia coronária e lupus eritematoso e doenças similares do tecido conjuntivo. Em alguns casos o espasmo coronário sozinho pode ser a causa da isquemia coronariana, embora mais frequentemente o espasmo coronário seja apenas um complicador da aterosclerose coronariana [163].

A aterosclerose coronariana é uma condição patológica das artérias coronárias caracterizada pela acumulação anormal de tecido lipídico e fibroso na parede dos vasos (ateromas) com resultante desorganização da arquitetura e da função dos vasos e com redução variável do fluxo de sangue para o miocárdio. A aterosclerose é conhecida há séculos e até muito recentemente a doença era considerada um componente necessário do processo de envelhecimento. Estudos epidemiológicos dos últimos 30 anos, entretanto, têm revelado grandes diferenças regionais das taxas de incidência e prevalência da aterosclerose e têm mostrado que aterosclerose significativa não é um componente necessário do processo de envelhecimento. O reconhecimento de que fatores genéticos, ambientais e outros podem acelerar o processo aterosclerótico tem feito do envelhecimento um fator importante mas não o único determinante das mudanças patológicas, as quais são influenciadas em diferentes graus por muitos fatores, alguns dos quais já foram identificados. Toda evidência presente indica que a aterosclerose coronariana seja uma doença multifatorial [163].

Tem havido uma grande expansão nos métodos de investigação clínica para a doença coronariana. A base para o diagnóstico clínico, contudo, continua sendo a história de angina pectoris, infarto do miocárdio, insuficiência cardíaca ou arritmias. A arteriografia coronariana pode ser usada se necessário, para estudar a anatomia das artérias coronárias. O cineangiograma ventricular esquerdo permite a estimacão do tamanho e da função do ventrículo esquerdo, e da movimentação das paredes. Estudos do eletrocardiograma sob exercício representam um meio de induzir a isquemia do miocárdio para fins de diagnóstico. A ecocardiografia é uma técnica não invasiva que permite estimar o tamanho e a configuração das dimensões ventriculares, identificar anormalidades de contração, determinar a fração de ejeção e diagnosticar ou afastar a doença mitral ou aórtica, a pericardite e o mixoma atrial esquerdo. Estes e outros estudos hoje disponíveis fazem da avaliação do paciente com doença coronariana um processo altamente complexo e sofisticado que envolve na sua expressão ideal equipes de médicos e técnicos especialmente treinados, incluindo fisiologistas, cardiologistas, radiologistas, anesthesiologistas e cirurgiões [163].

### I.3.2 - Descrição do Trabalho

O trabalho foi organizado em 7 capítulos, o 1º correspondendo a esta introdução. O capítulo seguinte (II) apresenta, de forma sumariada, uma revisão da literatura sobre sistemas de apoio à decisão médica. O Capítulo III trata de modelos bayesianos de classificação diagnóstica. O modelo clássico baseado na suposição de independência dos descritores é revisto e um modelo mais avançado baseado no uso de aglomerados de descritores interdependentes, capaz de aproveitar melhor a informação contida nos dados médicos, é discutido. No Capítulo IV, uma proposta para um gerador de planos de investigação médica, baseado no uso de programação heurística, é apresentada. O Capítulo V descreve os experimentos realizados incluindo o trabalho de construção de um banco de dados real de pacientes com doença coronariana, a estrutura de dados definida, a construção de modelos, a descrição das ferramentas utilizadas e o trabalho de investigação executado. O Capítulo VI apresenta e analisa os resultados obtidos no trabalho experimental descrito no capítulo anterior. Finalmente, o Capítulo VII é reservado para as conclusões e recomendações. Nos dois apêndices podemos ver o arquivo de pacientes utilizado e os programas de computador desenvolvidos.

---

---

## **II - REVISÃO DA LITERATURA**

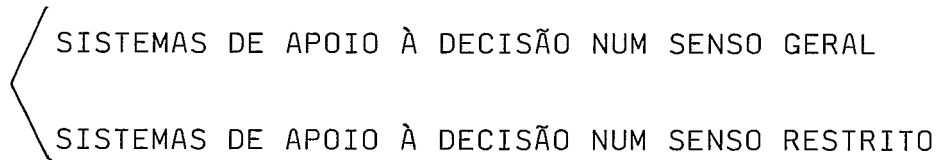
---

---



## II.1 - CLASSIFICAÇÃO DOS SISTEMAS DE APOIO À DECISÃO NO GERENCIAMENTO DE PACIENTES

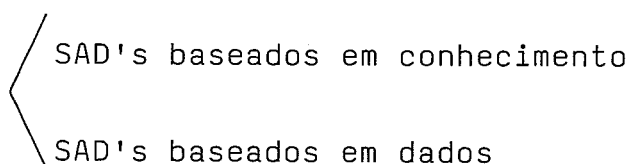
Podemos classificar os sistemas de apoio à decisão no gerenciamento de pacientes em duas grandes categorias:



Os sistemas de apoio à decisão num senso geral suportam a fase de inteligência do processo de tomada de decisão permitindo colher, checar, armazenar, recuperar e comunicar dados e produzir relatórios. O seu objetivo é obter uma melhoria na qualidade dos dados utilizados em medicina, o barateamento de seu custo e diminuição de riscos. Entre eles se incluem "check lists", questionários, sistemas de anamnese por computador [116, 179, 180], protocolos e a aplicação de técnicas de bioengenharia.

Os sistemas de apoio a decisão num senso restrito, que são objetivo de nossa atenção nesse trabalho, suportam principalmente as fases de delineamento e seleção do processo decisório. Dentro do processo de gerenciamento de pacientes eles suportam especificamente as decisões de diagnose, prognose, seleção de exames e seleção de terapia. Os sistemas de apoio à decisão num senso restrito, serão chamados daqui para frente simplesmente de sistemas de apoio à decisão e abreviados por SAD.

Diferentes formas de classificar os SAD's foram publicadas na literatura [33, 76, 85, 161, 170]. Vamos adotar as classificações de Shorflife [161] e Szolovits [170] que apresentam duas classes extremas de um espectro de soluções a saber



Existem pelo menos 4 tipos de conhecimento que podem ser diferenciados de dados estatísticos puros, a saber:

- a - Conhecimento derivado da análise de dados (principalmente numérico)
- b - Conhecimento subjetivo ou julgamental
- c - Conhecimento científico ou teórico
- d - Conhecimento de alto nível estratégico ou auto conhecimento.

De um modo geral os modelos baseados em conhecimento utilizam o que Szolovits chama de raciocínio categórico enquanto os modelos baseados em dados utilizam o raciocínio probabilístico [170].

Uma decisão categórica tipicamente depende de relativamente poucos fatos, sua conveniência é fácil de julgar e seu resultado, é inambíguo. Uma decisão categórica é simples de fazer e a regra que forma a sua base é também usualmente simples de descrever (embora sua validade possa ser complicada de justificar). Os médicos muito freqüentemente trabalham com decisões categóricas. Infelizmente nem todas decisões podem ser desse tipo. Não existe uma regra simples para decidir se se realiza uma biópsia de medula óssea ou quando dar alta a um paciente de uma unidade coronariana. Para tomar estas decisões deve-se pesar cuidadosamente toda a evidência disponível. Desconhece-se como os médicos pesam a evidência que favorece e a que opõe-se às diversas hipóteses ou cursos de ação. Vários esquemas formais, no entanto, existem para pesar evidência, sendo mais amplamente usado o probabilístico. Outros enfoques menos convencionais são o de funções de credibilidade e as aplicações da teoria de conjuntos nebulosos [90].

Se existe uma cronologia no campo desde os últimos 25 anos ela reflete a tendência progressiva de sistemas com menos dependência de dados puros e mais ênfase em conhecimento simbólico inferido dos dados primários.

## II.2 - SAD's BASEADOS EM CONHECIMENTO (SADC)

Vamos apresentar a seguir dois casos puros de SAD's baseados em conhecimento.

### II.2.1 - Algoritmos Clínicos

Algoritmos clínicos são fluxogramas ou árvores de decisão projetados por médicos especialistas onde é codificado diretamente o conhecimento médico pertinente. Os nós intermediários do fluxograma correspondem a questões e os nós terminais a resultados como por exemplo um diagnóstico, a referência do paciente a um especialista ou uma terapia selecionada. Embora cada decisão num algoritmo clínico seja categórica, o desenvolvimento de seu fluxograma pode muitas vezes ter sido baseado em computações probabilísticas extensas.

As áreas de aplicação dos algoritmos clínicos tendo em vista suas limitações, estão restritas a campos de problema que só requerem conhecimento superficial como triagem, sua área de maior sucesso, ou problemas que requerem uma grande quantidade de conhecimento de um campo médico limitado e suficientemente bem entendido no nível clínico [113, 152, 183].

Entre as limitações dos algoritmos clínicos citamos sua inflexibilidade e extrema dependência do contexto decisório montado nos estágios iniciais do processo de consulta. Tais sistemas são sensíveis aos dados iniciais e não podem trabalhar num contexto de informações parciais, incorretos ou inexatos, que é a realidade de trabalho de um clínico. Sua exatidão é limitada àquela do especialista que criou o algoritmo.

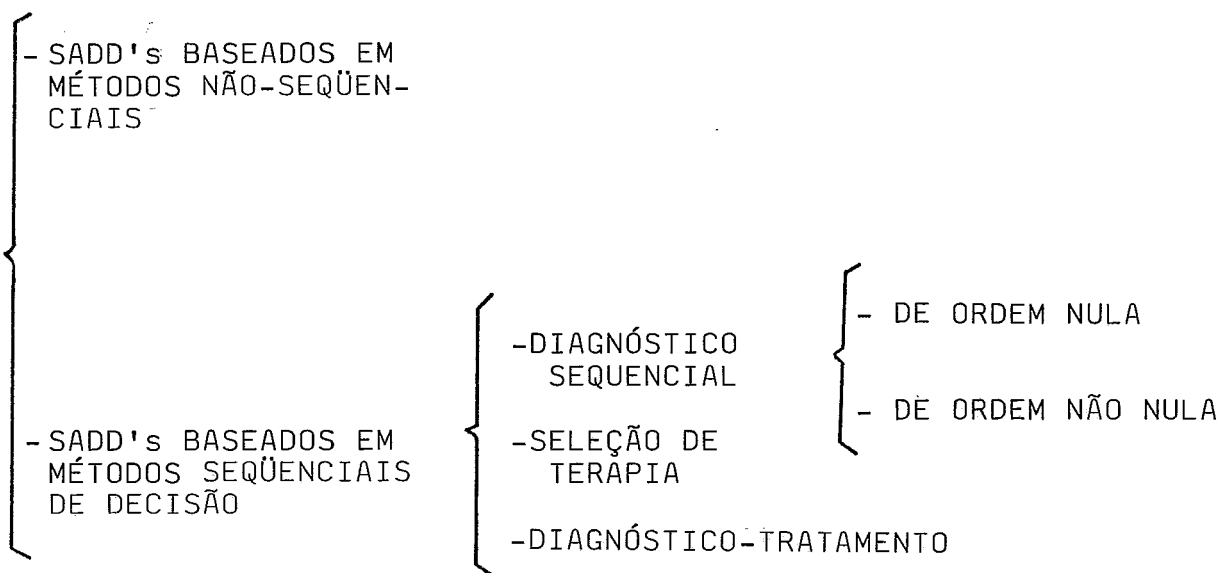
### II.2.2 - Modelos Matemáticos de Processos Físicos

Num número limitado de áreas clínicas os processos fisiopatológicos podem ser descritos precisamente por meio de fór

mulas matemáticas. Isto tem levado ao desenvolvimento de sistemas de computador para auxílio à diagnose nestas áreas. Estes programas pedem as informações, fazem os cálculos apropriados e fornecem uma análise clínica ou uma recomendação terapêutica [123]. O sistema mais conhecido é o de Bleich [17, 18] em Equilíbrio Ácido-Base do sangue. Uma vantagem destes sistemas é a possibilidade de se montar programas eficientes de computador com base em modelos matemáticos claros e bem definidos. Por outro lado, a maior deficiência desse tipo de enfoque é que poucas áreas da Medicina são passíveis de uma descrição quantitativa exata e não ambígua.

### II.3 - SAD's BASEADOS EM DADOS (SADD)

Sobre os SAD's baseados em dados (SADD) foi concentrada a maior parte do esforço desenvolvido na área de decisão médica assistida por computador. Apresentamos a seguir uma classificação dos SADD's em função das técnicas empregadas, que tenta refletir a perspectiva histórica de utilização e sua importância na área.



## II.3.1 - SADD's Baseados em Métodos Não-Sequenciais

### II.3.1.1 - Introdução

Os métodos classificatórios permitem construir casos puros de SAD's baseados em dados. Aplicam-se especialmente a decisões de diagnose e prognose e foram os primeiros a serem usados em Medicina. Também são chamados de Reconhecimento Estatístico de Padrões [128, 132]. Pressupõem que a cada indivíduo esteja associada uma bem definida classe (diagnóstica ou prognóstica) e uma coleção de observações diagnósticas.

As observações diagnósticas são valores realizados de variáveis chamadas **Descritores**, ou seja, peças de informação médica ou não médica de relevância diagnóstica ou prognóstica potencial, tais como sintomas, sinais, história, resultados de exames, etc.

Os descritores podem ser contínuos ou discretos, estes últimos podendo ser classificados em dicotômicos ou politômicos.

O objetivo do sistema é, conhecidas as **observações** obtidas pela mensuração dos descritores em um paciente, classificá-lo em uma de um conjunto de **classes diagnósticas** ou **prognósticas** que descrevem o seu estado de saúde. As classes diagnósticas podem ser doenças, síndromes, estados clínicos ou estados fisiológicos, enquanto as classes prognósticas correspondem a estados futuros do paciente, decorrentes de um processo de doença e seu tratamento, podendo ser por exemplo, faixas de sobrevivência, tais como "menos de 5 anos", "mais de 5 anos de sobrevivência".

### II.3.1.2 - O Paciente como um Sistema

Para descrever um paciente como um sistema, o número necessário de classes e descritores é significativamente grande. Bleich estimou que são necessárias cerca de 6000 classes e 20000

descritores [132]. Os médicos resolvem o problema dessa enorme dimensão formando subsistemas, cada um contendo um subconjunto de classes e seus correspondentes descritores. Os subsistemas podem ser definidos como subsistemas orgânicos, por exemplo, renal, cardiovascular, pulmonar ou endócrino, ou como subsistemas de doenças, por exemplo, doenças infecciosas, abdominais agudas, diabetes, doenças cardiovasculares, cardiopatias congênitas, etc. Outra abordagem na definição de subsistemas é a chamada abordagem sintomática, baseada em sinais e sintomas, por exemplo, pacientes em coma.

### II.3.1.3 - O Paciente num Estado Fixo

Para aplicação dos métodos classificatórios freqüentemente se adota o que é chamado: "O Modelo do Paciente num Estado Fixo". Esse modelo impõe uma série de pressuposições, mas corresponde aproximadamente a uma situação básica, recorrente em medicina, a do diagnóstico diferencial [132].

### II.3.1.4 - Funções Discriminantes

A classificação é feita através de funções **discriminantes**, uma por classe, dependentes dos valores dos descritores. Diferentes tipos de modelos são empregados levando a diferentes formas de funções discriminantes. A diversidade entre os modelos decorre de diferentes pressuposições quanto aos perfis de sintomas e/ou critérios de tomada de decisão.

Sistemas reconhedores de padrões antes de serem utilizados para classificação precisam ser calibrados (treinados) usando-se normalmente para isso um **conjunto de calibração (treinamento)** formado por indivíduos cuja classificação é conhecida (aprendizado supervisionado).

### II.3.1.5 - Critérios de Decisão

Todos os modelos de reconhecimento de padrões realizam a classificação de indivíduos de acordo com algum critério de decisão. Isso significa que as funções discriminantes do modelo deveriam fazer diagnósticos bons sob certo sentido. Mas o que é um bom conjunto de diagnósticos. Três dos critérios melhor conhecidos e mais atraentes, intuitivamente empregados em problemas de reconhecimento de padrões são discutidos abaixo.

#### A - CRITÉRIO DE BAYES

O critério de Bayes é baseado na noção de que o melhor conjunto de alocações de indivíduos é aquele resultante da mínima perda esperada decorrente de classificações errôneas. O ponto central desse critério é a especificação do custo que incorremos quando um sujeito de uma classe é incorretamente alocado à outra classe (também chamado de perda).

#### B - CRITÉRIO DE "CLUSTERING"

É baseado na noção de que indivíduos densamente aglomerados no espaço de padrões devem ser classificados numa mesma categoria. O conceito de proximidade de pontos num espaço vetorial exige que as variáveis sejam contínuas ou, quando discretas, que pelo menos apresentem uma ordenação de seus valores, o que freqüentemente não é possível em medicina, como por exemplo na variável "raça".

#### C - CRITÉRIO DE MÍNIMO ERRO MÉDIO QUADRÁTICO

É baseado na técnica de análise de regressão. Um conjunto de funções discriminantes funciona bem nesse sentido se o erro quadrático esperado entre os valores assumidos pelas funções e os valores pré-determinados for pequeno.

Croft [33] comparou o desempenho de 10 modelos diferentes de reconhecimento de padrões, incluindo modelos não paramétricos como o método de Fix-Hodges e o do "Mais Próximo Vi-

zinho", testando-os com um mesmo problema médico. Chegou à conclusão que não existe diferença significativa de desempenho entre eles, tendo o modelo Bayesiano com descritores discretos supostos independentes apresentado o melhor desempenho.

### II.3.1.6 - Principais Estudos/Desempenho

A maioria das aplicações de sistemas de apoio à decisão na área de diagnóstico médico foi baseada no critério bayesiano de decisão, tendo-se assumido, como meio de reduzir o número de parâmetros a estimar, a independência estatística dos descritores dentro de cada classe diagnóstica.

A Tabela (II.1) mostra alguns dos principais trabalhos que utilizaram o método bayesiano em diagnóstico, suas características e a taxa de classificação correta obtida. O padrão para aferição do modelo nem sempre foi objetivo (ex: biopsia, necrópsia, diagnóstico cirúrgico). Em muitos casos o padrão foi o diagnóstico do especialista, o que obviamente pode introduzir consideráveis distorções. Em alguns trabalhos foi possível estabelecer uma comparação de desempenhos do sistema e dos médicos, conforme mostra a tabela. Outros estudos baseados em métodos classificatórios podem ser encontrados em [1, 16, 21, 22,34,40,47,48,54,58,62,88,117,118,124,145,157 e 167].

Vamos comentar em mais detalhe o trabalho de Dombal [39] que estudou a exatidão do diagnóstico na síndrome de dor abdominal aguda, comparando o desempenho de diversas classes de profissionais médicos e o de um computador usando o modelo Bayesiano (calibrado com um banco de dados de 700 casos). 552 casos de pacientes novos com dor abdominal aguda de menos de uma semana, encaminhados ao Departamento de Cirurgia na Universidade de Leeds tiveram os diagnósticos pré-operatórios atribuídos pelos médicos e o calculado pelo computador, comparados com o diagnóstico exato pós-operatório. O estudo considerava 7 classes diagnósticas possíveis e 29 diferentes descritores. Os seguintes resultados foram obtidos:



Autores	Área Médica	Comentários	Taxa class. correta modelo	Taxa class. corr. médica	Padrão aferição	Nº Classes	Nº Descrit.	Tamanho Con.	
								Trein.	Teste
Overall-Williams[131]	Função da Tireóide	Independência Condicional	96,3%		Médico	3	21	450	-
Warner-Toronto-Veasy [181]	Cardiopatas Congênitas	Indep. Cond. com Aprox.				35	53		-
Wilson [187]	Úlceras gástricas		100%		Cirurg.	19	39	1196	-
Lodwick [109]	Tumores ósseos primários	Indep. condic. com elim.sint.corr.	77,9%		Histol.	8	71 ACH	77	-
Knill-Jones [98]	Icterícia	Indep. Condic.	71%			11	45	320	65
Boyle [19]	Bócio não tóxico	Indep. Condic.	82%	77%	Biopsia	3	30	155	88
de Dombal [41]	Síndrome dor abdom. aguda	Indep. Condic.	91%	81%	Cirurg.	7	29	700	552
Horrocks - de Dombal[86]	Dispepsia	Indep. Condic.	87,7%	70%		5	27	212	-
Wilson, de Dombal[188]	Dor abdom.aguda em emerg.	Indep. Condic.	65%	65,6%	Mixto Clin. Cirurg.	19	39	1196	-
Mc Neil [121]	Dor pleurítica torácica	Indep. Condic.	73%			3	11	97	-
Burbank [23]	Doenças do Fígado	Indep. Condic.	58,5%		Patol.	6	144 ACH		
Gustafson[78]	Doenças da Tireóide	Indep. Condic.	87,6%			3	15	259	-
Croft [33]	Doenças do Fígado	Indep. Condic.	64%		Biopsia	20	50	1991	437

Tabela II.1

<u>TIPO DE DIAGNÓSTICO</u>	<u>EXATIDÃO</u>
Admissão	42%
"House Surgeon"	71%
"Registrar"	79%
"Senior Clinician"	81%
Computador	91%

### II.3.1.7 - Bancos de Dados de Casos Médicos com Funções Analíticas

Uma categoria diferente de SADD's baseia-se na interação direta com um banco de dados de casos médicos. Os primeiros bancos de dados em medicina serviam apenas para armazenar informações clínicas, à semelhança dos prontuários de pacientes. Mais tarde foram desenvolvidos sistemas consultores (principalmente para doenças crônicas) com esquemas de recuperação de dados de pacientes que apresentem determinada configuração de atributos (grupos de semelhança), o que permite realizar diretamente a inferência diagnóstica. Além disso esses sistemas permitem calcular correlações entre variáveis, medir indicadores de prognóstico e comparar as reações a diferentes terapias. Muitos deles têm regras diagnósticas embutidas que permitem fazer sugestões diagnósticas ao médico. Também representam um poderoso suporte para decisões de cunho tático-estratégico, para pesquisa e para educação médica bem como para administração de sistemas de assistência médica.

Entre os sistemas mais conhecidos estão o HELP [142, 182], ARAMIS [122], o Banco de dados cardiovasculares da Universidade de Duke [146,147, 166], o de Feinstein [100], o de Wiener [186], o de Collen [27,28,30,31], o PROMIS [179], o MARIS [162] de Slamecka, o COSTAR [11], o CARE [70,71,151], o TOD [61], o CDSS [69,70].

O problema fundamental é que seus processos de decisão não são capazes de montar diagnoses em situações complexas, uma vez que o conhecimento armazenado não é estruturado, é basi

camente empírico, não refletindo os processos julgamentais do clínico. Acrescente-se os problemas de aquisição e construção do Banco de Dados, que só será capaz de produzir resultados após um longo e, por vezes, não muito motivante, trabalho de minuciosa coleta de casos.

## II.4 - SADD's BASEADOS EM MÉTODOS SEQUENCIAIS DE DECISÃO

A partir do final da década de 60 passou a existir um interesse crescente sobre os meios de como utilizar as técnicas da teoria da decisão em medicina e em implementá-las em computador. Algumas boas revisões sobre o assunto podem ser encontradas em [67,92,164].

Em termos gerais a análise de decisão pode ser vista como uma tentativa de considerar valores associados às escolhas com probabilidades a fim de avaliar os processos pelos quais as decisões são tomadas ou deveriam ser tomadas. Um conjunto de axiomas delimitando a situação de decisão para um único decisor, juntamente com algumas noções intuitivas de "racionalidade" conduzem a uma bem definida prescrição para a realização ótima de decisões - o assim chamado "Paradigma Bayesiano". Este paradigma requer que sejam atribuídos valores numéricos (utilidades) a todas as conseqüências que podem advir de uma decisão clínica particular.

### II.4.1 - Potencialidade dos Métodos da Teoria da Decisão em Medicina

Os métodos classificatórios descritos anteriormente são aplicáveis a situações essencialmente estáticas, usando toda a informação do paciente de uma vez só, mesmo quando ela foi colhida seqüencialmente ao longo do tempo. Os métodos da teoria da decisão permitem representar o processo seqüencial de deci-

são típico do gerenciamento de pacientes, seja quanto a decisões diagnósticas, seja quanto a ações terapêuticas. Os métodos da teoria da decisão jogam com a maioria dos fatores que influem no processo decisório de um clínico, tais como:

- Riscos de complicação associados com um exame diagnóstico ou um tratamento, incluindo a possibilidade de morte.
- Custo financeiro dos testes e tratamentos.
- O valor diagnóstico do teste, ou seja a quantidade de informação com que o exame contribuirá para a identificação de uma entidade específica de doença.
- Características do paciente tais como sexo, idade, situação econômica, estado geral de saúde.
- Mudanças espontâneas no estado do paciente seguindo-se a uma ação médica ou decorrentes do curso natural da doença quando nenhuma ação é tomada.
- As preferências do paciente, do médico ou outros decisores quanto às diversas conseqüências possíveis, tanto em termos de saúde como em termos econômicos.

#### **II.4.2 - Classificação dos Modelos de Aplicação da Teoria da Decisão**

Podemos identificar 3 tipos de áreas de aplicação de técnicas de teoria da decisão no processo de gerenciamento de pacientes a saber:

- Modelos de Diagnóstico Seqüencial
- Modelos de Seleção de Tratamento
- Modelos Integrados Diagnóstico-Tratamento

### II.4.3 - Modelos de Diagnóstico Seqüencial

Os modelos de diagnóstico seqüencial têm como objetivo indicar, com base na informação colhida até o momento sobre o paciente, quais os exames que devem ser realizados nas próximas etapas de investigação, obedecendo a um critério de minimização do custo global da decisão. Trata-se de um problema crítico e "on line" associado à rotina diária do médico. Os métodos de teoria de decisão podem ajudar a responder questões do tipo:

"Que exames devem ser realizados e em que ordem?"

"Quando uma bateria de exames deveria ser executada simultaneamente e quando seqüencialmente?"

"Quando usar uma bateria completa de exames, e quando aplicar apenas uns poucos?"

"Em que testes os riscos e custos são maiores que a informação significativa que os testes podem fornecer?"

"Quão eficientemente estão organizados os procedimentos de rotina de exame?"

"Existem circunstâncias em que os médicos deveriam obter certos exames antes de examinar o paciente ou existem circunstâncias em que o exame físico deve proceder a tomada da história?"

"Que testes diagnósticos de rotina deveriam ser realizados de rotina em um paciente assintomático e qual o seu valor médico?"

O modelo de diagnóstico seqüencial utiliza como infraestrutura o esquema usado no método classificatório bayesiano. Leva em consideração os custos dos exames, necessitando também de uma **matriz de perdas** (que podem ser interpretados como utilidades negativas) para representar as preferências de decisor pelas diferentes conseqüências da decisão diagnóstica. A matriz de perdas representa o custo de diagnosticar uma doença  $d_i$  quando o paciente tem realmente a doença  $d_j$ . Sem a matriz de perdas seria impossível avaliar o custo da conseqüência de interromper o processo de investigação e diagnosticar uma doença com base na informação acumulada até o estágio atual. Os custos dos exames e as perdas devem ser obrigatoriamente expressos nas mesmas

unidades.

### II.4.3.1 - Tipos de Regras de Decisão

Podemos encarar o problema de buscar uma seqüência ó-tima de exames de dois modos distintos, quanto ao horizonte de tempo de realização dos exames, gerando dois tipos diferentes de regras de decisão, a saber:

/ REGRAS DE DECISÃO DE ORDEM NÃO NULA  
 \ REGRAS DE DECISÃO DE ORDEM NULA

Nas regras de decisão de ordem não nula os valores das variáveis de decisão dependem de valores de variáveis aleatórias (resultados de testes) que são observados durante o horizonte de tempo. Ou seja, a regra de decisão no  $k^{\text{º}}$  estágio é conditionada à informação acumulada depois de realizar os  $k-1$  exames precedentes. Apenas no  $1^{\text{º}}$  estágio temos certeza absoluta da ação a realizar. Note-se que a regra de decisão resultante terá em geral uma estrutura bastante complexa tal como: "Realize primeiramente o teste  $S_i$ ; se o resultado  $s_{i1}$  for observado então pare de testar e diagnostique a doença  $d_j$ ; se  $s_{i2}$  for observado então realize o teste  $s_k$  como próximo passo"; e assim por diante.

Na regra de decisão de ordem nula os valores das variáveis de decisão devem ser determinados no começo do horizonte de tempo. Portanto, temos já desde o início certeza sobre todos os testes a realizar sem necessidade de esperar pelos resultados dos exames. É provável que o médico se interesse mais por uma solução desse tipo (ordem nula) pois não há restrições na ordem em que os testes devem ser feitos, e o tempo total de investigação será menor. Além disso o planejamento para execução de todos os exames da seqüência pode desde já ser iniciado. Este planejamento pode envolver o preparo do paciente para os exames e o agendamento das atividades necessárias para realização dos exames (como reserva de sala cirúrgica, etc).

Peters apresenta em [139] os métodos bem como abordagens heurísticas para a obtenção de regras ótimas de decisão de ordem não nula e de ordem nula.

#### II.4.3.2 - Principais Trabalhos

A eficiência dos modelos de diagnóstico seqüencial tem sido constatada em diversos estudos. Gorry [72] conseguiu, usando os dados de cardiopatias congênitas de Warner [181] obter um desempenho equivalente ao conseguido por Warner, porém usando um número significativamente menor de exames. Em média o modelo de Gorry utilizou 6,9 testes para estabelecer o diagnóstico, um número muito menor que os 31 testes empregados por Warner. Gleser [68] conseguiu criar uma regra de decisão de ordem não nula capaz de detectar pacientes do banco de dados do "Kaiser Permanente Multiphasic Screening Program" com Diabetes Mellitus não reconhecida em check-ups anteriores, utilizando uma árvore de decisão com apenas 5 níveis. Peters [139] usando os dados de Flehinger [64] sobre doenças hematológicas, determinou regras de decisão de ordem nula e não nula para um subconjunto de 13 doenças.

#### II.4.4 - Modelos para Seleção de Tratamento

O dilema de administrar ou não um certo medicamento ou executar uma determinada operação em um paciente sem um diagnóstico estabelecido é familiar aos médicos. Em muitas situações clínicas existe considerável incerteza sobre a presença ou ausência de uma determinada doença porque não existem mais exames diagnósticos confirmatórios disponíveis. Dada esta incerteza, administrar o tratamento eficaz para a doença sob consideração será benéfico se a doença realmente estiver presente, porém será prejudicial se a doença estiver ausente. Por outro lado, deixar de administrar o tratamento pode ser catastrófico se a doen

ça estiver presente, porém, não se estiver ausente. Consideremos, por exemplo, o problema comum referente a dor inexplicada de curta duração no quadrante abdominal inferior direito. Suponhamos que o médico tenha exaurido todos os estudos diagnósticos e tenha concluído que a apendicite aguda é bastante provável. Que fatores estão envolvidos na decisão de operar imediatamente ou esperar e observar o paciente? O benefício da apendicectomia é curar o paciente se ele tiver realmente apendicite, mas se o paciente não tiver a doença, o custo é um risco finito de morbidade e mortalidade. Ao invés, evitar a operação pode resultar numa pena grave (perfuração, peritonite e septicemia), porém é a melhor decisão se o paciente estiver com gastroenterite em lugar de apendicite. Esta é uma decisão relativamente simples que não envolve os desejos do paciente e a habilidade do cirurgião. Os riscos são baixos e os benefícios potenciais elevados. Existem no entanto situações mais complexas como o caso de decidir entre uma cirurgia de ponte de safena e a terapia médica conservadora para um paciente com doença coronariana. Nesse caso a cirurgia traz alívio da dor (ao menos a curto prazo), mas os riscos operatórios são significativos e os benefícios a longo prazo sobre a expectativa de vida e alívio da dor ainda não foram provados.

A partir do final da década de 60 tais complexos problemas clínicos passaram a ser abordados através da aplicação de técnicas de teoria da decisão. O primeiro trabalho de peso no assunto foi o de Ginsberg [67] na Rand Corporation. Outros trabalhos reportados são os de Gorry [73] no gerenciamento da insuficiência renal aguda, Pauker [137] no estudo da cirurgia das artérias coronárias, Giaugue [65], Tompkins [177] e Murphy [127] na avaliação de estratégias de tratamento da faringite e prevenção da febre reumática aguda, Pliskin [140] no gerenciamento de pacientes com insuficiência renal crônica, Conklin [32] na tomada de decisão em radioterapia, Krischer [101] no estudo de defeitos congênitos do lábio e palato e Schwartz [155] no estudo da hipertensão arterial. Forst [57] sugeriu a aplicação da teoria de decisão em julgamentos de litígios por imperícia médica.



#### II.4.4.1 - Pressuposições

As pressuposições comumente adotadas para aplicação de técnicas de teoria da decisão na seleção de terapia são as seguintes:

1. O paciente só apresenta uma doença
2. Existe um tratamento indicado claramente benéfico e bem definido para cada classe diagnóstica
3. Só é possível selecionar uma das classes de tratamento disponíveis em cada estágio da decisão
4. São conhecidas as conseqüências possíveis para o paciente, decorrentes de cada ação terapêutica
5. São conhecidas as probabilidades das diversas conseqüências possíveis decorrentes de cada ação terapêutica
6. São conhecidas funções de utilidade que exprimem as preferências do decisor quanto às diversas conseqüências possíveis tanto em termos de estado de saúde como custos financeiros.

#### Árvore de Decisão

A aplicação do método exige, antes de executar qualquer ação, a construção de uma árvore de decisão formada por nós de decisão, onde o clínico deve escolher uma dentre um conjunto de ações, e nós do acaso onde o resultado é uma resposta probabilística do paciente a uma ação realizada. Analisando o processo de decisão o médico deverá, ao construir a árvore teórica de decisão, delinear antecipadamente todos os nós pertinentes de decisão e do acaso, todas as conseqüências plausíveis, bem como os caminhos pelos quais essas conseqüências podem ser atingidas. A complexidade da árvore de decisão reflete a complexidade do problema e o conhecimento do decisor sobre a área em questão. Podemos ter árvores simples como no problema clássico [135] de decisão TRATAR X NÃO TRATAR com apenas um es

tágio de decisão, como mostrado na figura (II.1). Em situações mais complicadas, altamente dinâmicas, podemos necessitar de árvores envolvendo muitos níveis de decisão.

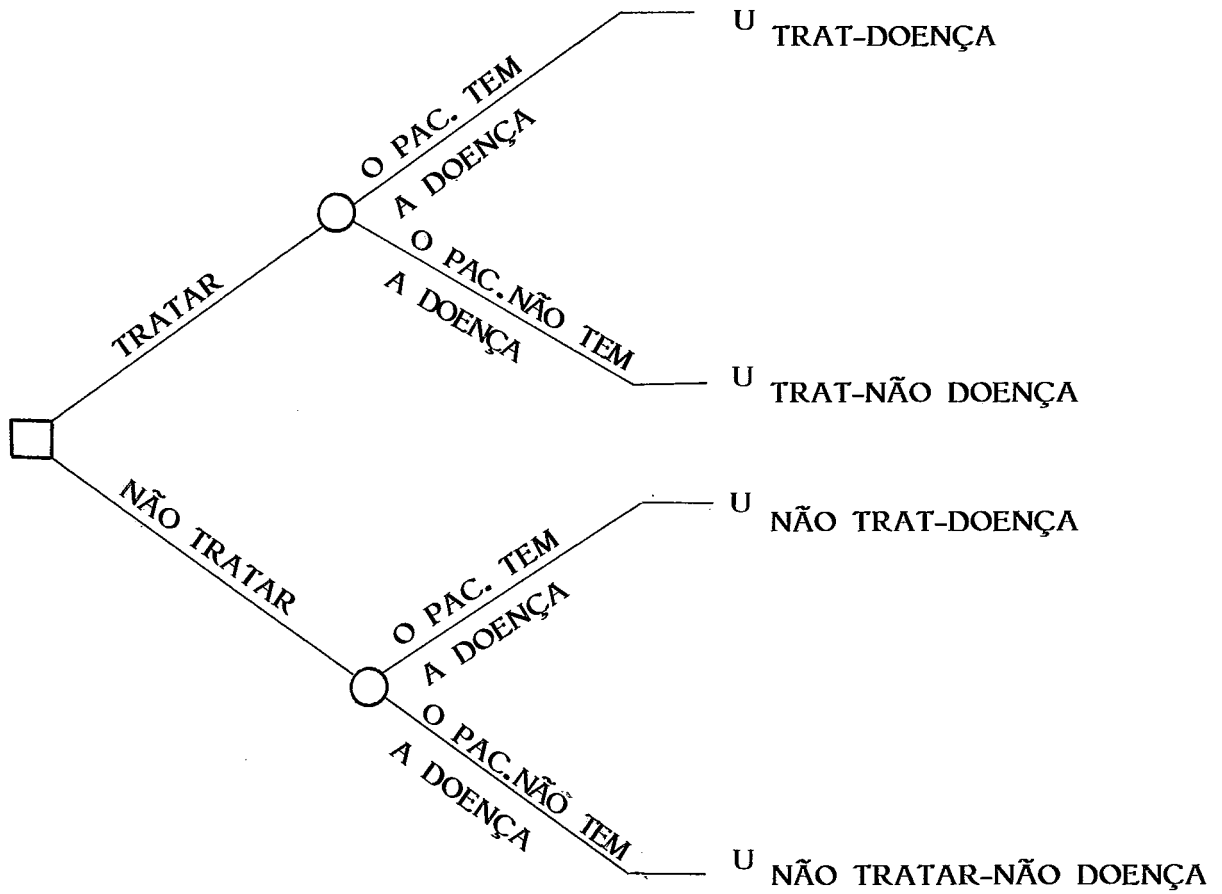


Figura II.1 - Árvore de decisão para o problema Tratar x Não tratar.

Nesses casos mais complexos, a árvore teórica pode assumir proporções consideráveis, fazendo-se então necessário podá-la. A poda da árvore é feita geralmente usando o conhecimento do especialista e levando em conta as peculiaridades do paciente sob análise.

#### II.4.4.3 - Estrutura de Utilidades

Um ponto crucial para a aplicação de técnicas de teoria da decisão em tratamento é a construção de uma estrutura que relacione as conseqüências finais sobre o paciente, de cada ação, bem como sua utilidade para o decisor. Normalmente as conseqüências são definidas como o estado de saúde resultante do paciente.

Existem 3 modos de definir a estrutura de utilidades, descritas a seguir

##### A - Estrutura com um Único Atributo Contínuo

Nesse caso se enquadram todos os Índices de Status de Saúde desenvolvidos para a área de planejamento e descrição de populações, os quais visam atribuir um valor não monetário para a vida humana. Torrance [178] apresentou uma formulação matemática unificada que enquadra 14 desses índices partindo do princípio de que um Índice de Status de Saúde deve ser formado pelo produto dos fatores: Quantidade de Vida e Qualidade de Vida.

Outra medida utilizada no contexto de operações de alto risco do tipo cura ou morte é a Expectativa de Vida Residual. Sacket e Torrance [164] e Murphy [120] usaram diretamente como utilidade a expectativa de vida residual em seus estudos.

Tendo em vista que a utilidade da vida humana certamente não é linear, e variará de acordo com o grupo de pacientes em estudo (avessos a risco ou propensos a risco) Spiegelhalter e Smith [164] discutem formas adequadas de funções de utilidade que transformem a expectância de vida na verdadeira utilidade

para o paciente sob análise.

## B - Estrutura Formada por uma Lista Finita de Conseqüências

Neste caso simplesmente relacionamos todas as conseqüências possíveis numa lista. As utilidades podem ser atribuídas a cada componente da lista através de jogos de loteria. Citamos alguns exemplos de trabalhos que usaram esta abordagem. Gorry [73] e Betaque [15] empregaram para o gerenciamento da insuficiência renal aguda, as seguintes conseqüências: "Melhorado"; "Inalterado" e "Piorado". Ledley [107] sugeriu as seguintes categorias "Curado", "Bem", "Mal" e "Crítico". Spiegelhalter [164] sugeriu a seguinte lista de conseqüências para o tratamento de doenças graves do fígado, a saber:

1. Morte imediata ou peri-operatória
2. Morte em 1 ano, confinado ao leito
3. Morte em 1 ano, móvel mas incapaz de trabalhar
4. Morte em 5 anos, móvel mas incapaz de trabalhar
5. Morte em 5 anos, capacitado para o trabalho
6. Sobrevivência maior que 5 anos, capacitado para o trabalho.

Safran [149] na avaliação do valor da linfoangiografia em pacientes com a doença de Hodgkin empregou como atributos os estágios de classificação de tumores de Ann Arbor, a saber: Estágios I+II, Estágio III e Estágio IV.

## C - Estrutura com Múltiplos Atributos

Esta abordagem é a mais realista uma vez que os resultados dos tratamentos médicos são em geral multifacetados. A determinação direta de uma utilidade unidimensional é bastante complexa. Por isso os pesquisadores tem procurado sempre que possível investigar condições de independência de modo a poder empregar algum dos resultados da teoria de utilidade multiatributos que permitam trabalhar com as funções de utilidade de cada atributo individualmente [94].

Considerável atenção foi dirigida ao modelo bidimensional com os atributos:

DURAÇÃO DA VIDA  
MEDIDA DE QUALIDADE DE VIDA

Entre os trabalhos que utilizaram o modelo bidimensional, citamos os de Pauker [137], Weinstein [164] e Nomura [128].

Krischer [101] em seu estudo sobre a correção cirúrgica de anomalias congênitas do lábio e palato, utilizou os seguintes atributos:

- Inteligibilidade da fala
- Atributo Cosmético
- Problemas de Audição
- Atributo Financeiro

Krischer concluiu que para 60% dos entrevistados, os atributos acima são preferencialmente independentes e sugeriu o uso da forma multiplicativa da função de utilidade. Giaouque [65] em seu estudo sobre estratégia de tratamento da dor de garganta e febre reumática usou os seguintes atributos:

- Custo para o paciente
- Custo (ou lucro) para o médico
- Custo para o sistema de saúde pública ou previdência
- Método de medicação (oral/injetável)
- Imunidade desenvolvida a espécie infectante de estrep<sub>t</sub>ococo
- Dias doente devido à infecção estreptocócica
- Reação a antibiótico
- Gravidade da febre reumática aguda ou do episódio ag<sub>u</sub>do de glomerulonefrite
- Tipo de medicação pós-reumática
- Efeitos a longo prazo da febre reumática ou glomérulo<sub>-</sub>nefrite.

Giauque concluiu preliminarmente que os atributos acima são preferencialmente independentes e independentes sob o ponto de vista de utilidade, sugerindo o uso da forma multiplicativa da função de utilidade.

#### II.4.5 - Modelos Integrados de Diagnóstico e Tratamento

Trata-se de uma fusão dos modelos de diagnóstico - e de tratamento permitindo abranger de forma integrada todas as etapas do processo de gerenciamento do paciente. Os modelos de diagnóstico padecem muitas vezes do defeito de encarar a identificação da doença como um fim em si mesmo, perdendo a perspectiva real do problema, na qual o diagnóstico é um passo intermediário para a seleção da terapia que solucionará o problema do paciente. Na realidade, o diagnóstico unitário nem sempre é possível ou necessário para o início do tratamento [83]. Só faz sentido continuar o processo de investigação diagnóstica solicitando novos exames se a sua realização for capaz, dependendo de seus resultados, de alterar uma decisão terapêutica.

O modelo de diagnóstico seqüencial procurava levar em consideração estes fatores através da matriz de perdas, que no entanto tem a desvantagem de ser muito difícil de construir. Alguns trabalhos envolvendo modelos integrados de diagnóstico e tratamento podem ser encontrados na literatura, entre os quais estão:

- Betaque e Gorry [15,73] com seu modelo para gerenciamento de pacientes com problemas renais.
- Safran [149] com seu modelo para avaliação do linfogramma no gerenciamento de pacientes com a doença de Hodgkin.
- Nomura [128] com seu modelo para o estudo de doenças pancreáticas.

### II.4.5.1 - Árvore de Decisão

Apresentamos na figura [II.2] um esquema de árvore de decisão adequado a modelos de diagnóstico e tratamento, a qual apresenta primeiramente os nós de decisão de seleção de exames e posteriormente nós para seleção de terapia

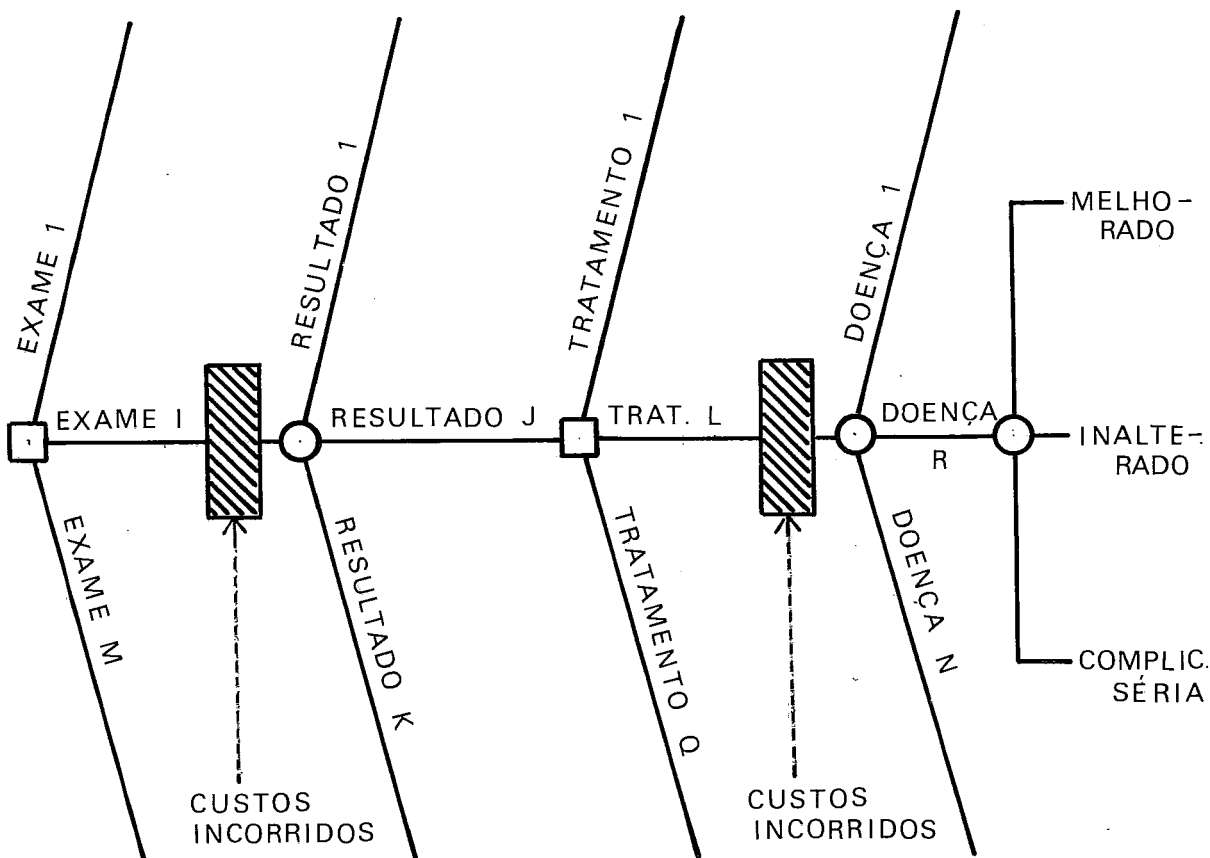


Figura II.2 - Árvore de decisão para o modelo integrado Diagnóstico-Tratamento.

#### II.4.5.2 - Estratégia

Citamos a estratégia empregada no modelo de Gorry (máxima utilidade esperada). Primeiramente o sistema determina qual o tratamento ótimo na hipótese de não haver mais testes a realizar e calcula a sua máxima utilidade esperada. Antes de recomendar este tratamento, o sistema avalia se os exames ainda disponíveis trazem vantagens. Para tanto, para cada exame e para cada resultado, o sistema simula a mudança decorrente na distribuição de probabilidades de doenças, determina qual seria o tratamento ótimo com essas novas probabilidades e calcula a sua utilidade esperada. Incorpora a este último valor o custo de realização do exame e determina se o valor combinado (exame + tratamento) é maior que o do tratamento imediato.

#### II.4.5.3 - Desempenho

Num teste realizado com o sistema de Gorry de gerenciamento de problemas renais foram avaliados 18 casos nos quais o diagnóstico era incerto, mas para os quais 2 especialistas em nefrologia desejavam realizar decisões gerenciais. Em 14 dos casos o programa selecionou o mesmo plano terapêutico e testes diagnósticos que foram escolhidos pelos especialistas. Para 3 dos 4 casos restantes, a decisão do programa correspondeu à 2ª escolha dos médicos, sendo de acordo com seu sentimento um plano alternativo razoável. No último caso os médicos também aceitaram a decisão do programa como razoável, embora não tenha correspondido as suas duas primeiras escolhas. O bom desempenho do programa de Gorry, a despeito de sua edificação sobre estimativas subjetivas de especialistas, serve para enfatizar a importância da análise clínica que alicerça a abordagem teórica de decisão. Shortlife [161] atribui a excelência deste resultado à maior participação de conhecimento clínico, através da destilação do domínio médico na árvore formal de decisão, transformando este programa em um SAD intermediário aos casos puros de SADD's e



SADC's.

#### II.4.5.4 - Dificuldades

Vários críticos apontam, como principal problema da a bordagem da teoria da decisão, a dificuldade em atribuir valores numéricos à vida humana e a relutância das pessoas em trabalhar com esses números. Acrescente-se o problema de expressar em uma escala comum os valores de atributos originalmente medidos em es calas diferentes. Outra dificuldade corresponde a atribuição de probabilidades aos resultados das ações terapêuticas. Tais probab bilidades dependem de características do paciente tais como sexo, idade, peso, status cardiovascular etc, bem como de fatores ex ternos como experiência do cirurgião, habilidade do anestesiolo gista, disponibilidade de staff de boa qualidade para cobertura pós-operatória.

Em suma, encontramos pontos de vista opostos quanto a utilidade das técnicas de teoria da decisão em medicina. Os pessi mistas, amparados pela atitude cautelosa do mundo dos negócios quanto a essa teoria, suspeitam que árvores de decisão, avaliaç ão de probabilidades e utilidades, etc podem ser inadequadas pa ra lidar com os importantes e sutis julgamentos que entram na decisão real. Os otimistas esperam que as técnicas de teoria da decisão venham a fazer mais gerenciáveis e racionais as decisões freqüentemente complexas que o médico precisa tomar a cada dia.

Mesmo que as dificuldades citadas anteriormente não ve nham a ser plenamente solucionadas, a teoria da decisão pode tra zer uma importante contribuição ao pensamento clínico:

- A construção da árvore força o médico a fazer uma tarefa muitas vezes ignorada, qual seja, antecipar os possíveis resultados de uma ação médica.
- Tentando estimar as probabilidades de cada nó do acaso os médici cos terão de revisar a literatura médica, o que provavelmente revelará vazios no conhecimento sugerindo futuras linhas de

pesquisa.

- Tentando atribuir valores de utilidades às conseqüências, os médicos terão de examinar não só os importantes aspectos intangíveis, mas também os conflitos legítimos de valores de diferentes decisores.
- Estudando coortes de pacientes com grupos específicos de valores observados, poderão ser definidas práticas recomendadas por grupo e desenvolver protocolos suficientemente estruturados para uso por pessoal paramédico em situações de rotina.

## II.5 - SAD's BASEADOS EM TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Por volta do início da década de 70 os pesquisadores na área de sistemas de apoio à decisão médica começaram a perceber a existência de situações complexas para as quais os SAD's baseados em formalismos não apresentavam uma resposta satisfatória. Parecia faltar a estes sistemas um verdadeiro "entendimento" da medicina. Faltava-lhes claramente o bom senso quando falhavam, o que representava um fator altamente frustrante para os médicos usuários. A partir de então, pesquisadores de várias instituições começaram a investigar a aplicação clínica potencial de técnicas de raciocínio simbólico, retiradas do ramo da ciência de computação conhecido como **Inteligência Artificial (I.A)**. O termo "Inteligência Artificial" é geralmente aceito como aquele que inclui aplicações de computadores que envolvem inferência simbólica em vez de cálculos numéricos estritos. Exemplos incluem programas que raciocinam sobre exploração mineral, química orgânica ou biologia molecular; programas que conversam em inglês e entendem sentenças em linguagem natural e programas que geram teorias a partir de observações. Tais programas ganham sua potência de julgamentos experimentais e qualitativos codificados nas assim chamadas **heurísticas**, em contraste com programas de cálculo nu-

mérico cujo poder provém das equações analíticas empregadas. As heurísticas enfocam a atenção do programa de raciocínio em partes do programa que parecem ser mais críticas e partes da base de conhecimento que parecem ser mais relevantes. Elas também guiam a aplicação do campo de conhecimento a um caso individual, pela retirada de itens de consideração e pelo enfoque em outros itens. O resultado é que o programa segue uma linha de raciocínio, ao contrário de seguir uma seqüência pré-definida de passos, liberando o programador da tarefa virtualmente impossível de especificar todas as contingências.

O objetivo dos SAD's baseados em técnicas de inteligência artificial é o de reproduzir o comportamento inteligente e o desempenho de um médico competente em sua área de trabalho. Tais programas correspondem a uma abordagem intermediária se vistas a luz da classificação dos SAD's em SAD's baseados em dados e SAD's baseados em conhecimento, utilizando à semelhança do clínico, tanto regras de raciocínio categórico como probabilístico. O foco da pesquisa de aplicação de técnicas de IA em medicina é achar meios de combinar estas formas de raciocínio para criar programas competentes que exibam perícia médica [161,170]. Foram utilizadas técnicas provenientes dos principais campos da IA: representação de conhecimento, programação heurística, compreensão e geração de linguagem natural e modelos de processo de pensamento.

### II.5.1 - Estrutura da Base de Conhecimento

Os subsistemas baseados em técnicas de IA podem ser contrastados aos métodos anteriores pela maior complexidade de conhecimento médico que explicitamente representam no computador. Todos têm estruturas de dados que permitem a expressão de relacionamentos semânticos entre os fatos que constituem suas bases de conhecimento, indo além dos pesos probabilísticos ou heurísticos que eram usados como ligações nos modelos anteriores. No entanto, os SAD's baseados em IA diferem vastamente entre si, em seus campos de utilização e nas formas de representar o co-

nhhecimento médico. As idéias centrais são apresentadas a seguir.

Ao reconhecer uma enfermidade como uma instância específica de uma doença e ao selecionar um tratamento apropriado, o sistema consultor está aplicando várias normas ou regras de interpretação. Basicamente podem ser distinguidas duas classes bastante distintas de regras:

- A - **Regras Táticas** - aquelas que interpretam conjuntos de fatos específicos em termos de hipóteses clínicas.
- B - **Regras Estratégicas** - aquelas que interpretam as condições ou o contexto no qual é apropriado aplicar as regras táticas para derivar interpretações de mais alto nível ou sugerir ações.

Os vários tipos de regras existentes que coordenam o processo decisório são conhecidas pelo nome global de **CONHECIMENTO NORMA  
TIVO**.

Diferentes regras estratégicas, que fazem parte integrante deste conhecimento normativo, são necessárias para:

- A - Focalizar um conjunto de problemas ou hipóteses diag  
cas (focalização).
- B - Efetuar um raciocínio discriminatório para chegar a um diag  
nóstico diferencial (diagnose).
- C - Medir o grau de confirmação das hipóteses (confirmação).
- D - Extrapolar os possíveis cursos futuros que a enfermidade po  
derá ter (prognose).
- E - Planejar os tratamentos (terapia).
- F - Explicar os processos de raciocínio do programa ao usuário (explanação).

Regras normativas táticas podem ser expressas logica-

mente como relações entre combinações de sintomas e hipóteses; entre hipóteses e outras hipóteses; e entre combinações de sintomas com hipóteses e tratamentos que deverão controlá-los. Existem usualmente medidas de confirmação associadas a cada uma dessas regras.

O **Conhecimento Descritivo** refere-se a fatos médicos sobre tratamento, enfermidades, doenças e sintomas e é usualmente expresso por relacionamentos causais, taxonômicos, temporais, ou associacionais que se observou existirem entre esses conceitos generalizados. A maneira pela qual tal conhecimento generalizado deve ser aplicado, ao fazer-se inferência e planos específicos relativos a pacientes individuais é determinada pelas estratégias de um sistema determinado. A figura (II.3) resume o exposto acima.

## II.5.2 - Estrutura do Sistema

Vamos descrever a seguir uma arquitetura geral para sistemas consultores baseados em técnicas de inteligência artificial conforme apresentado por Elfes [45]. Os seguintes módulos ou componentes podem ser identificados:

- Base de conhecimento médico
- Módulos de aquisição de conhecimento médico
- Dados do paciente
- Módulo de consulta, subdividido em
  - Aquisição e crítica de dados de pacientes
  - Focalização
  - Inferência
  - Planejamento
  - Explicação
- Modelo da Situação do Paciente
- Programa de Diálogo
- Módulo de Aprendizado

No módulo de consulta o programa de focalização in-

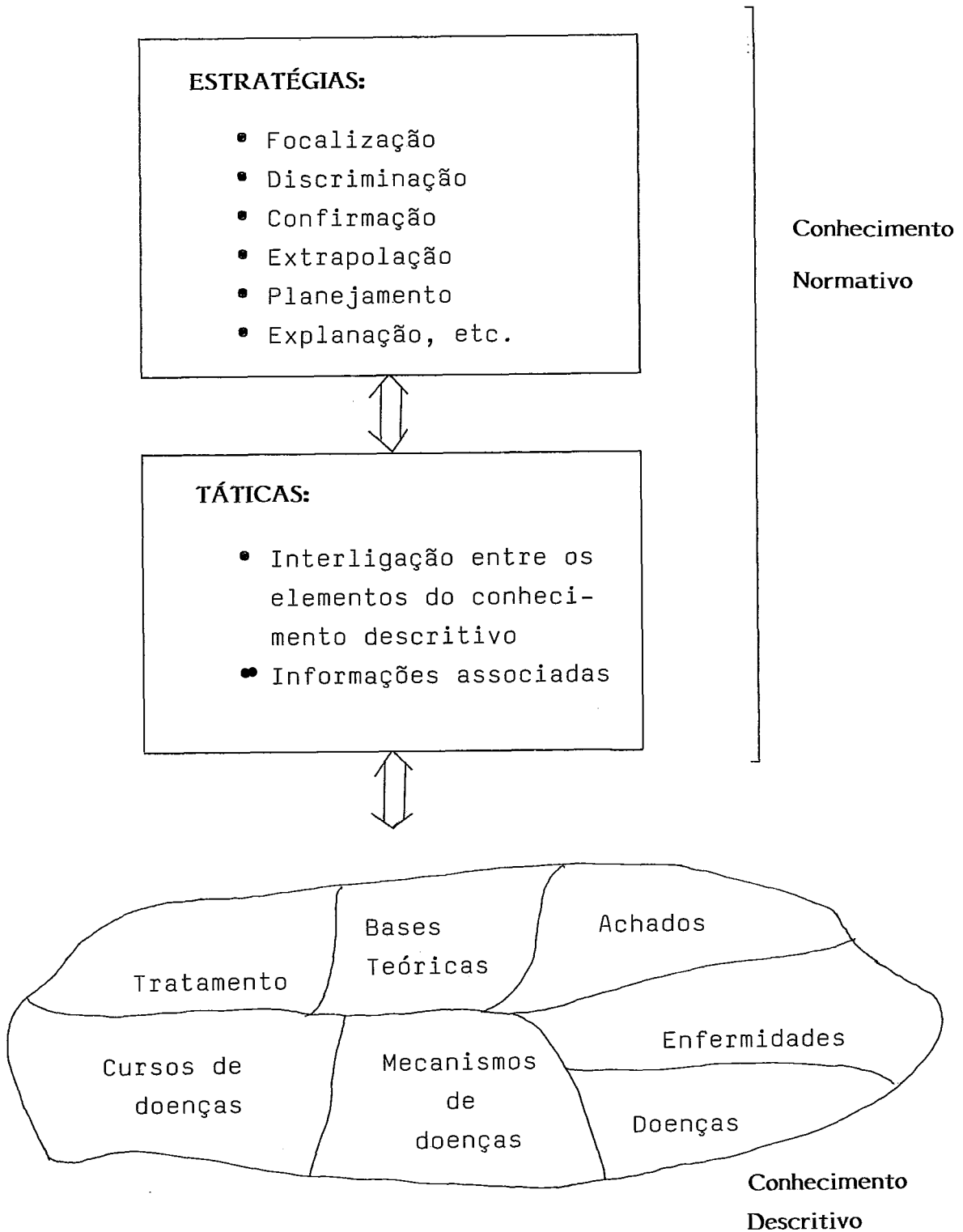


Figura II.3 - Componentes principais da base de conhecimentos de sistemas de Inteligência Artificial para Consulta Médica.

cumbe-se de dirigir o raciocínio do sistema de forma a sempre concentrar-se naquelas sub-áreas específicas da base de conhecimento que sejam mais relevantes à interpretação das observações do paciente. O programa de inferência gerencia a aplicação de táticas e estratégias para obter a interpretação final da situação do paciente; o programa de planejamento fornece as recomendações de tratamento e o módulo de explanação é capaz de fornecer ao usuário explicações do raciocínio do programa para chegar à solução apresentada. O modelo da situação do paciente, representa a interpretação do caso pelo sistema, englobando as hipóteses diagnósticas validadas até o momento bem como os possíveis cursos futuros das enfermidades específicas (prognose) e sugestões de ações terapêuticas. O programa de diálogo é o responsável pela interação com o usuário, a qual em alguns sistemas implementados é feita em linguagem natural. O módulo de aprendizado deveria ser capaz de, a partir dos resultados obtidos pelo sistema de consulta e informações fornecidas pelo usuário sobre eventuais discrepâncias com resultados "corretos", rever o processo de raciocínio e efetuar modificações, corrigindo o comportamento do sistema.

### **II.5.3 - Sistemas de Consulta Médica Usando Técnicas de I.A.**

Diversos sistemas de apoio a decisão médica usando técnicas de inteligência artificial foram desenvolvidos depois dos trabalhos pioneiros de Kleinmuntz [95] e Wortman [189], a saber: o CASNET [185] na área de glaucoma, o MYCIN [158,160,161,175] na área de diagnóstico e terapia antimicrobiana, o INTERNIST [170] abrangendo a área de medicina interna, o MEDAS [15] destinado à área de emergência, o PIP [136] para doenças renais.

As tabelas (II.2), (II.3) e (II.4) apresentam as características principais destes sistemas [45].

Nome do Sistema	Área Médica	Ano do 1º Protótipo	Instituição / Projeto	Pesquisadores	Esquema de Funcionamento
CASNET	Glaucoma	1971	Rutgers Univ./ Rutgers Research Resource on Com- puters in Biome- dicine	Weiss, Kulikowski, Safir e outros	Rede semântica causal manipulada por algo- ritmos de várias es- tratégias de raciocí- nio.
MYCIN	Doenças infeccio- sas: Bac- teremias	1972	Stanford Univ./ SUMEX-AIM	Shotliffe, Buchanan, Davis, Cohen, e outros.	Árvore de contexto e regras de produção.
INTERNIST I	Medicina interna	1973	Pittsburgh Univ./ Clinical Decision Laboratory	Pople, Myer, e outros.	Rede hierárquica e al- goritmos heurísticos.
PRESENT ILLNESS PROGRAM (PIP)	Medicina interna: proble- mas re- nais	1974	MIT - Tufts Medi- cal School/Clini- cal Decision Making Group	Pauker, Gorry, Kassirer, Schwartz, Szolovits e outros	Quadros e heurísti- cas para processamen- to de hipótese.

Tabela II.2 - Principais Sistemas de Consulta Médica utilizando técnicas de Inteligência Artificial.



SISTEMA	REPRESENTAÇÃO DE CONHECIMENTO	
	DESCRITIVO	NORMATIVO
CASNET	Rede causal-associacional, com vários níveis descritivos.	Ligações de implicação entre achados, hipóteses, e tratamentos.
MYCIN	Árvore de contexto e propriedades das estruturas de dados	Regras de produção
INTERNIST I	Rede hierárquica (taxonômica) com algumas ligações causais.	Ligações de implicação entre achados e hipóteses.
PIP	Quadros ("frames") para descrição de enfermidades. Estrutura de memória de longo prazo.	Restrições lógicas dentro de e entre quadros.

**Tabela II.3** - Representação de conhecimento em sistemas de Inteligência Artificial em Medicina.

SISTEMAS	ESTRATÉGIAS DE RACIOCÍNIO
CASNET	Gerenciamento global da rede para escolha de alternativas; avaliação de hipóteses intermediárias e de alto nível para diagnose; utilização dos trajetos da rede causal para prognose; regras de produção para avaliação de planos de tratamento; explanação feita por meio dos trajetos da rede específicos ao paciente.
MYCIN	Árvore de contexto guia a escolha de alternativas; utilização de regras de produção para alcançar alvos inferiores a partir de alvos superiores, num processo de encadeamento de regras orientado pelo conseqüente, a fim de obter diagnose, prognose e terapia; explanação feita por meio de regras de produção específicas ao paciente.
INTERNIST I	Heurísticas de particionamento para escolha de hipóteses; heurísticas combinam pesos das evidências para avaliação de hipóteses (diagnose).
PIP	Focalização é feita por meio de heurísticas para ativar hipóteses, trazendo-as para a memória de curto prazo; heurísticas de avaliação combinam pesos de evidências para diagnose.

Tabela II.4 - Estratégias de raciocínio em sistemas de Inteligência Artificial em Medicina.

#### II.5.4 - Comentários

Os sistemas baseados em técnicas de IA parecem ter um futuro promissor pela capacidade de combinar as técnicas usadas em sistemas puros baseados em dados e baseados em conhecimento, contornando contudo as dificuldades inerentes a esses métodos. Tais sistemas podem ter uma abrangência bem maior, englobando mesmo, alguns como o INTERNIST ou o CADUCEUS, especialidades inteiras. Possuem uma capacidade praticamente ilimitada de ampliar sua base de conhecimento e poderão ter seu desempenho melhorado pela representação e utilização de metaconhecimento de alto nível, que lhes permita entender suas próprias limitações e estratégias de raciocínio. O conhecimento de seu campo de atuação pelo programa é particularmente desejável em áreas onde o conhecimento tende a ser altamente julgamental e pouco quantitativo [161].

Comparados com o médico competente, no entanto, os melhores sistemas que usam técnicas de IA ainda apresentam muitas deficiências entre as quais:

- a) Programas que lidam com áreas relativamente amplas como o INTERNIST ou o PIP têm critérios inadequados para decidir quando um diagnóstico está completo. Não existe percepção de quando os problemas diagnósticos principais já foram resolvidos, restando apenas pontos secundários. O programa continua a explorar hipóteses adicionais cada vez menos coerentes até que o usuário se canse da consulta.
- b) Devido ao fato de a estratégia inicial do programa usar cada nova observação significativa como um indício para a possibilidade de enfermidades associadas, e porque essa estratégia permanece, ao longo de toda a operação, novas hipóteses são continuamente geradas.
- c) O foco do programa pode, em casos extremos, mudar inteiramente de questão para questão o que é desconcertante para o usu

ário médico, habituado que está a desenvolver parte da rotina numa ordem sistemática [170].

## II.6 - DISCUSSÃO

Até o momento os experimentos na área de decisão médica assistida por computador têm sido freqüentemente protótipos. Uma proporção insignificante de sistemas saíram do ambiente experimental para o uso numa base de rotina, mesmo quando seu desempenho foi considerado excelente.

### II.6.1 - Obstáculos

Os principais obstáculos apontados já desde a década de 60 pelos pesquisadores tem sido:

- A - A falta de definições médicas padronizadas para doenças e sintomas. Na profissão médica, hoje, cada hospital ou clínica usa suas próprias definições para termos médicos, sintomas, doenças, faixas de normalidade de resultados de exames de laboratório, etc o que impede a construção de sistemas de âmbito regional ou a transferência de sistemas de uma instituição para outra [33,80].
- B - A falta de bancos de dados médicos grandes e confiáveis dos quais parâmetros para modelos matemáticos possam ser extraídos.
- C - A qualidade precária dos dados clínicos sujeitos a elevados níveis de erro e variabilidade de interpretação.

### II.6.2 - Pontos de Vista

Duas opiniões complementares existem sobre a utilização prática de sistemas de apoio a decisão em medicina: a de Card, segundo a qual a transferência de atividades da medicina clínica para computadores deve esperar a remoção dos obstáculos citados, exigindo portanto pesquisa básica e detalhada sobre o processo diagnóstico em si; e a opinião de Edwards, que embora aceite a necessidade de formalização do diagnóstico, em termos gerais considera que existem muitas áreas da medicina onde benefícios poderiam ser obtidos pela imediata aplicação do conhecimento e técnicas disponíveis [39]. A experiência tem revelado que essas duas visões não são só complementares, mas muitas vezes concorrentes, de tal modo que um desenvolvimento em um campo de atividade tem um efeito realimentador em ambos.

### II.6.3 - Prerequisitos para o Sucesso

Existem dois quesitos básicos e recorrentes que são a chave para a plena utilização dos sistemas por nós tratados. São eles:

**DESEMPENHO** - como se pode projetar sistemas que produzam decisões melhores, mais confiáveis e a um custo razoável numa ampla gama de áreas de aplicação?

**ACEITAÇÃO** - Como podemos efetivamente encorajar o uso de tais sistemas pelos médicos e outros usuários?

### II.6.3.1 - Desempenho

O quesito desempenho pode ser visto de vários ângulos: O ponto de vista **Exatidão** sempre foi o que mereceu maior preocupação pelos pesquisadores tendo sido amplamente discutido nesse capítulo para cada tipo de técnica. Sob o ângulo de **Economicidade** seria necessário comparar os custos dos sistemas de apoio a decisão ("hardware", "software", operação, manutenção, etc) com os benefícios potenciais, pelo menos os tangíveis (redução de pacientes-dia internados, menos gasto de tempo do médico e outros profissionais, redução no número de exames onerosos, etc). Raríssimos estudos de economicidade foram realizados. Citamos o caso de de Dombal da Universidade de Leeds com seu sistema para diagnóstico na síndrome de dor abdominal aguda, que estimou o custo por paciente em 10 pence e comparou com o benefício esperado de evitar anualmente 34 laparatomias negativas e as hospitalizações pós-operatórias envolvendo um custo de £200 por paciente [39]. De qualquer forma, a tendência sob o ponto de vista de economicidade é altamente favorável se lembrarmos os padrões de evolução dos custos da tecnologia de processamento de dados e dos custos crescentes dos serviços de saúde.

### II.6.3.2 - Aceitação

Trata-se de um ponto tão ou mais importante que o de desempenho.

#### Como o Pessoal Médico vê o Computador?

Existem dados para suportar a visão extrema que as polarizações do pessoal médico contra computadores são tão fortes que os sistemas de apoio a decisão serão inevitavelmente rejeitados independentemente de seu desempenho [115,161,168]. Provavelmente esse é um ponto de vista exagerado. Todavia não há co-

mo negar a posição altamente cautelosa assumida pela classe médica quanto a sistemas de apoio a decisão. Existe uma dificuldade de grande de encorajar os médicos, principalmente os mais antigos, a usarem diretamente o computador. Feinstein [50] afirmava que o médico teme ser substituído pelo computador, teme que o computador venha a demolir com a arte da medicina clínica ou que o computador venha a alterar a prática da medicina. Na verdade a prática da medicina tem sido alterada a cada avanço tecnológico — a imprensa, o microscópio, o raio X, o laboratório químico, etc. O computador, no entanto representa um tipo diverso de alteração pois pela 1ª vez um avanço tecnológico (ainda misterioso para muitos médicos) compete com a habilidade humana de pensar, assumindo por isso uma conotação profissionalmente ameaçadora.

### **Como os Sistemas de Apoio a Decisão Atendem às Necessidades Básicas dos Usuários?**

Esta é uma maneira diferente e provavelmente de maior peso para explicar a baixa aceitação dos SAD's pela classe médica. Consideramos alguns fatores nem sempre lembrados na seleção de técnicas e no desenvolvimento de sistemas.

- A - A classe médica é carente de formação matemática o que multiplica a dificuldade já por si grande de lidar com distribuições de probabilidades, de quantificar noções subjetivas como indesejabilidade, risco de exames, eficácia de tratamento, etc.
  
- B - Os médicos são muitas vezes extremamente relutantes em engajar-se em qualquer computação numérica envolvendo a verossimilhança de uma doença ou o prognóstico de um tratamento. Os médicos certamente têm uma forte impressão de sua confiança no diagnóstico e tratamento, mas essa impressão deve advir mais de reconhecimento de uma situação típica ou da comparação do presente caso com experiências passadas, do que de qualquer computação formal de verossimilhanças [170].

- C- A maioria dos sistemas realizados têm um escopo consideravelmente estreito limitando a oportunidade de aplicação em pacientes na rotina diária e portanto reduzindo sua utilidade prática.
- D- Frequentemente os sistemas não são de uso cordial — exigem habilidades não médicas especializadas, são ditatoriais e inflexíveis ao estilo particular do clínico, parecem arbitrárias em seu comportamento ou são incompreensíveis aos olhos do clínico. É importante considerar que a maioria dos profissionais, e em particular, os médicos, não gostam que consultores (computadores ou pessoas) lhes digam o que fazer em cada passo do processo de tomada de decisão. Preferem tomar a maioria das decisões eles mesmos e chamar o consultor só quando alcançam um beco sem saída, ou se sentem inseguros sobre a rota que estão tomando.
- E - Muitos sistemas não atendem a restrições de tempo e local. O sistema não deve gastar mais tempo do que os médicos usualmente gastam para uma decisão devido às grandes pressões de tempo sob as quais o médico trabalha. Além disso o sistema deve estar disponível no local de trabalho do médico, evitando que o mesmo tenha de se deslocar.

#### II.6.4 - Conclusões

As pesquisas nos campos de Decisão Médica Assistida por Computador e também de Sistemas de Informação Médica tem criado frentes de choque da medicina com outras áreas tais como ciência de computação e estatística. Esse choque é benéfico pois pressiona pela revisão da metodologia médica atual, pela melhoria da qualidade de descritores de pacientes, pela implementação de abordagens mais racionais com relação à informação médica tais como a organização do prontuário médico por problemas [184] e a criação de bancos de dados médicos de boa qualidade com



abrangência possivelmente regional; assim como aponta para brechas no conhecimento médico e oferece uma série de ferramentas inovadoras capazes de revolucionar o sistema de saúde quanto à disponibilidade e qualidade de serviços à população. O impacto dessas ferramentas só será pleno, no entanto, se houver uma reformulação no ensino médico e se conseguirmos implementá-las em larga escala a nível operacional, ou seja na prática médica diária. Para tanto, faz-se necessário vencer os desafios de desempenho e aceitação por nós discutidos.

---

### **III - MODELOS BAYESIANOS DE CLASSIFICAÇÃO DIAGNÓSTICA**

---

### III.1 - INTRODUÇÃO

A abordagem estatística para o problema do diagnóstico médico sempre atraiu muita atenção de pesquisadores médicos e estatísticos. Já em 1959 Ledley e Lusted [106] propunham uma formulação bayesiana para o diagnóstico médico diferencial.

A realização de estudos de diagnóstico diferencial em qualquer área da medicina está cercada por dificuldades desde o nível elementar. Precisamos colher dados sobre doenças e sintomas de pacientes. A primeira dificuldade a ser vencida é a falta generalizada de definições precisas para doenças, sintomas e procedimentos de medição, todos os quais apresentam uma notória variação entre diferentes observadores. Uma vez que as doenças são processos dinâmicos, podem existir mudanças importantes com o tempo e com o tratamento, nas quantidades que estamos tentando medir. Embora teoricamente seja possível incorporar a influência dessas mudanças, frequentemente nos estudos de diagnóstico elas são ignoradas ou controladas, de modo que sintomas e doenças possam ser considerados atributos bem definidos de um indivíduo (recaindo no modelo "O Paciente num Estado Fixo"). Além disso, existe o problema do erro de observação [26,42,56,64,66,77,99,111,112] inerente às ciências biológicas o qual pode ter um efeito devastador sobre a exatidão do diagnóstico. Esta é uma área em que, como nos mostra Gill [66], pode-se conseguir substanciais melhorias através de atenção meticulosa e detalhada na predefinição dos termos a serem usados na coleta de dados. Vamos a seguir rever o modelo bayesiano de classificação diagnóstica e suas principais características.

### III.2 - FORMULAÇÃO

O objetivo do modelo bayesiano de classificação diagnóstica é o de alocar o paciente sob análise a uma, de um conjunto finito de classes diagnósticas ou prognósticas, com base

nos sinais e sintomas apresentados pelo paciente.

É pressuposição básica que a cada indivíduo esteja associada uma bem definida classe (diagnóstica ou prognóstica)  $d$  e uma coleção de observações diagnósticas  $\underline{s}$ . Nem  $\underline{s}$  nem  $d$  variam significativamente para um indivíduo e podem ser determinados sem erro por hipótese. Nós consideramos o  $(\underline{s}, d)$  de um indivíduo como uma realização de variáveis aleatórias  $(\underline{S}, D)$  tendo uma distribuição conjunta sobre a população. Os componentes de  $\underline{S}$  são chamados **Descritores**, enquanto o termo **Valor Observável**, indica um possível valor realizado de um descritor. Os descritores usados em medicina são freqüentemente variáveis discretas.

Nós, de forma semelhante, distinguimos a variável aleatória  $D$  (**Diagnose ou Prognose**) e uma **Classe Diagnóstica ou Prognóstica**  $d$  como um valor realizado de  $D$ .  $D$  é usualmente uma variável discreta com uma classe diagnóstica ou prognóstica representando o estado de saúde do paciente.

Um banco de dados colhido pelo centro médico acumulará informações sobre os descritores e as classes de doença de um grande número de indivíduos. Estes dados serão usados para fazer inferência sobre a distribuição  $(\underline{S}, D)$  podendo ser analisados por métodos estatísticos mais ou menos formais para criar uma regra de diagnóstico. Tal regra será aplicada ao complexo  $\underline{s}$  de valores observáveis de um novo paciente para inferirmos a classe de doença  $d$  não observada.

### II.2.1 - Notação e Pressuposições

A Tabela (III.1) apresenta a notação utilizada neste trabalho e a Tabela (III.2) as pressuposições básicas exigidas pelo modelo bayesiano de classificação diagnóstica.

TABELA III.1 - PARTE 1 - NOTAÇÃO

$D$	= Variável aleatória DIAGNOSE (ou PROGNOSE) com espaço amostral $\Delta = \{d_1, d_2, \dots, d_N\}$
$d_i$	= $i^{\text{a}}$ classe diagnóstica
$N$	= Número de classes diagnósticas possíveis
$\tilde{S}$	= Vetor de descritores do paciente com componentes $(S_1, S_2, \dots, S_M)$
$M$	= Número de descritores
$S_i$	= $i^{\text{o}}$ descritor com espaço amostral $\Sigma_i = \{s_{i1}, s_{i2}, \dots, s_{i, n(i)}\}$
$s_{ij}$	= $j^{\text{o}}$ valor observável do descritor $S_i$
$n(i)$	= Número de diferentes valores observáveis do descritor $S_i$
$\tilde{s}$	= Perfil do paciente = vetor formado pelos valores observados de um paciente
$s_q$	= Valor observado do $q^{\text{o}}$ teste de uma seqüência de testes realizados sobre o paciente
$p(d_i)$	= Probabilidade a priori que o paciente tenha a doença $d_i$
$p(s_{jq}/d_i)$	= Probabilidade condicional de obter $s_{jq}$ dado que o paciente tem a doença $d_i$
$p(\tilde{s}/d_i)$	= Probabilidade condicional de obtermos o perfil $\tilde{s}$ dado que o paciente tem a doença $d_i$
$p(d_i/\tilde{s})$	= Probabilidade que o paciente tenha a doença $d_i$ após observarmos o perfil $\tilde{s}$
$\tilde{p}(d/\tilde{s})$	= Vetor de probabilidade $[p(d_1/\tilde{s}), p(d_2/\tilde{s}), \dots, p(d_N/\tilde{s})]$

TABELA III.1 - PARTE 2 - NOTAÇÃO

$\tilde{p}(d/s_1, s_2, \dots, s_n)$  = Vetor de probabilidades das doenças  $d_1, \dots, d_N$  após obtermos os valores observados  $s_1, s_2, \dots, s_n$

$\tilde{L}$  = Matriz de perdas  
 $l_{ij}$  = custo de diagnosticar  $d_i$  quando o paciente tem realmente  $d_j$

$\tilde{L}_i$  =  $i^{\text{a}}$  linha de  $\tilde{L}$

$A_{ij}$  =  $j^{\text{o}}$  aglomerado de descritores interdependentes da doença  $d_i$

$a_{ijk}$  =  $k^{\text{a}}$  configuração de observações do aglomerado de descritores  $A_{ij}$

$p(a_{ijk}/d_i)$  = probabilidade condicional de ocorrência da configuração de observações  $a_{ijk}$  dado que o paciente tem a doença  $d_i$ .

$U$  = Matriz de similaridade entre descritores

$u_{ij} = 1$  se  $S_i$  similar a  $S_j$   
 $= 0$  em caso contrário.

TABELA III.2

- PRESSUPOSIÇÕES DO MODELO BAYESIANO DE  
CLASSIFICAÇÃO DIAGNÓSTICA

1. O conjunto  $\{d_1, \dots, d_N\}$  de classes diagnósticas é conhecido e fixo, e todas as classes são diferentes.
2. O conjunto  $\{S_1, S_2, \dots, S_M\}$  de descritores disponíveis é conhecido e fixo, e todos os descritores são diferentes. Cada descritor é caracterizado por um número de valores observáveis possíveis.
3. As classes diagnósticas são mutuamente exclusivas, ou seja, o paciente tem apenas uma doença.
4. As classes diagnósticas são exaustivas.
5. Cada par de doenças difere ao menos por uma das observações, a qual é revelada ao menos por um descritor de  $\{S_1, \dots, S_M\}$
6. A distribuição a priori de probabilidades das classes diagnósticas é conhecida
7. A probabilidade de obter alguma observação para cada descritor, dado que o paciente tem uma doença, é conhecida para todas as doenças e valores observáveis.
8. Para cada par ordenado de doenças  $i, j$  os custos de diagnóstico errônea da doença  $j$  em vez de  $i$  são conhecidos
9. Para cada descritor  $S_i$  existe um exame capaz de mensurá-lo sobre um paciente qualquer
10. Os resultados de realização dos testes podem ser determinados com certeza

### III.2.2 - Descrição da Regra de Diagnóstico

Dado um paciente com o vetor de observações  $\underline{s}$  pela regra de Bayes podemos calcular as probabilidades a posteriori:

$$p(d_i/\underline{s}) = \frac{p(d_i) \cdot p(\underline{s}/d_i)}{p(\underline{s})} \quad i = 1, \dots, N \quad (\text{III.1})$$

onde

$$p(\underline{s}) = \sum_{i=1}^N p(d_i) p(\underline{s}/d_i) \quad (\text{III.2})$$

$p(\underline{s})$  funciona como uma constante de normalização.

O custo esperado de diagnóstico errôneo por categoria é:

$$c(d_i, \underline{s}) = p(d/s) \cdot \underline{L}_i \quad (\text{III.3})$$

A regra clássica de decisão de mínima perda preconiza que o paciente seja alocado à classe que apresentar menor custo de diagnóstico errônea. A matriz de perdas  $\underline{L}$  em geral é assimétrica para a situação da medicina [2].

Quando se define  $\underline{L}$  com

$$\begin{cases} l_{ij} = 0 & \text{se } i = j \\ l_{ij} = 1 & \text{se } i \neq j \end{cases} \quad \text{a equação (III.3) reduz-se à}$$

regra de decisão de mínima probabilidade de erro. A regra mínima de probabilidade de erro foi a mais comumente implementada. O médico recebe como resultado da análise uma lista das doenças mais plausíveis com as probabilidades a posteriori calculadas pelo sistema.



## CARACTERÍSTICAS / PONTOS DE DISCUSSÃO

### 1. Flexibilidade

A regra de Bayes apresenta como vantagens o fato de poder processar observações em qualquer ordem e poder trabalhar com subconjuntos de descritores, o que lhe dá um caráter iterativo que é de importância fundamental para os métodos seqüenciais de apoio ao diagnóstico.

### 2. Estabilidade das Distribuições de Dados Diagnósticos

Dawid [35] apresenta uma interessante discussão sobre a estabilidade da distribuição  $(\underline{S}, D)$  com relação à variação da estrutura da população. Vários autores [19,35] pressupõem que a distribuição  $p(\underline{s}/d)$  pode ser considerada fixa, o que repousa no conceito de doença como causa, e sintomas como efeito. Porém nem todos componentes de  $\underline{S}$  são logicamente posteriores à doença. Descritores como sexo, cor, idade, status sócio-econômico, etc são logicamente anteriores à doença invalidando portanto essa pressuposição. Apenas a distribuição  $p(m/d, c)$  pode ser considerada estável, onde  $m, c$  são os fatores logicamente posteriores (médicos) e anteriores à doença, respectivamente. Segundo este raciocínio Dawid sugere que quantidades do tipo

$$\frac{p(d_1/s_1) \cdot p(d_2/s_2)}{p(d_1/s_2) \cdot p(d_2/s_1)}$$

são invariantes da distribuição  $(\underline{S}, D)$  mesmo sob mudanças da estrutura da população e portanto deveriam ser publicadas na literatura em lugar de  $p(\underline{s}/d)$ .

Shea [156] mostrou que mesmo a estabilidade da distribuição de sintomas médicos é insustentável quando lidamos com

doenças de múltiplos estágios.

### 3. Vício de Seleção

As probabilidades a priori  $p(d_i)$  dependem profundamente das características da população alvo e do processo de seleção de pacientes. Podemos considerar três mecanismos de seleção os quais conduzem a diferentes estimativas de  $p(d_i)$ , a saber: [156]

A - Procedimento de "Screening" da população

B - Indivíduos que procuram voluntariamente assistência médica

C - Pacientes referidos pela clínica geral ou emergência para clínicas ou hospitais especializados.

Devido a dificuldade de estimar as probabilidades  $p(d_i)$  de uma população particular, alguns autores adotam alternativas como usar estimativas subjetivas fornecidas pelo médico, ou a hipótese de Laplace de uniformidade das probabilidades a priori, o que transmuta o modelo Bayesiano no método de verossimilhança relativa.

### 4. Método de Multi-Eligibilidade

Uma crítica freqüente ao método Bayesiano refere-se às pressuposições 3 e 4 de exclusividade e exaustividade difíceis de sustentar, quando áreas mais amplas da medicina são consideradas. Uma solução proposta para o caso de pacientes com mais de uma doença é a definição de complexos de doenças. Para evitar a explosão combinatorial de classes daí decorrente, teríamos de nos limitar apenas aos complexos de ocorrência mais freqüente [106, 170].

Uma solução engenhosa para as limitações impostas pelas pressuposições 3 e 4 é o Método de Multi-Eligibilidade usado por Flehinger [55] e Ben-Bassat [12] nos sistemas HEME e MEDAS.

DAS. Esse método cria para cada doença uma estrutura bayesiana própria particionando a população de pacientes em cada estrutura  $i$  em pacientes com a doença  $d_i$  e pacientes sem a doença  $d_i$ .

## 5. Paradigma Diagnóstico

Ao contrário do paradigma amostral, como é chamado por alguns autores o modelo até aqui apresentado, no paradigma diagnóstico as probabilidades  $p(d_i/s)$  são estimadas diretamente a partir do banco de dados de pacientes.

O paradigma diagnóstico possibilita a obtenção de bons resultados com pressuposições mais fracas e menos informação que os estudos costumeiros usando o paradigma amostral [35]. Theater [176] usando apenas 7 dos 30 descritores empregados por Knill-Jones [98] obteve um desempenho equivalente (85% de exatidão) na alocação cirúrgica de pacientes com icterícia. David [35] demonstra que o processo de seleção não afeta as distribuições diagnósticas  $p(D/S)$ . Isto é muito conveniente pois estas distribuições são justamente as que procuramos quando nos defrontamos com um paciente cujos dados observamos.

### III.3 - ESTIMAÇÃO DE PARÂMETROS

#### II.3.1 - Número de Estimativas

Vamos nos concentrar em modelos bayesianos de classificação diagnóstica que usem como descritores apenas variáveis discretas. Isto corresponde à situação mais comum em estudos de diagnóstico diferencial onde uma grande parte dos descritores costuma ser de variáveis discretas. As variáveis contínuas eventualmente existentes podem ser discretizadas usando limites de normalidade freqüentemente disponíveis na literatura. A

seleção de limites de normalidade em medicina é discutida em [29,81,96,114 e 119].

O processo de estimação de parâmetros de um modelo bayesiano de classificação diagnóstica também chamado de processo de Calibração ou Treinamento é feito a partir de um banco de dados com informações colhidas sobre os sintomas e a classe diagnóstica de um grande número de indivíduos atendidos por um ou mais centros médicos. Chamamos a esse banco de dados de **Banco de Dados de Calibração ou Treinamento**.

Temos de estimar dois tipos de parâmetros: as probabilidades a priori  $p(d_i)$  e as probabilidades condicionais  $p(\underline{s}/d_i)$ . O número de estimativas de probabilidades a priori é  $N$  (número de classes diagnósticas). As probabilidades a priori podem ser estimadas facilmente calculando as proporções de ocorrência das diversas classes diagnósticas numa amostra realizada sobre a população em foco. Devem ser evitadas estimativas nulas de probabilidades a priori pois isto eliminaria da análise as classes diagnósticas correspondentes, independentemente da quantidade de evidência colhida em seu favor.

No modelo bayesiano completo, também chamado de **Atual**, a quantidade de estimativas de probabilidades condicionais do tipo  $p(s_{jq}/d_i)$  é de

$$N \cdot \prod_{j=1}^M n(j)$$

ou de  $N \cdot n^M$  se o número de valores observáveis por descritor for constante e igual a  $n$ . Efetivamente precisamos calcular

$$N \cdot \left[ \prod_{j=1}^M n(j) \right] - N \text{ estimativas.}$$

As probabilidades condicionais podem ser estimadas diretamente

dividindo o número de indivíduos do banco de dados de treinamento que possuam a doença  $d_i$  e que estejam contidos na célula  $s = \{s_1, s_2, s_3, \dots, s_M\}$  pelo número total de pacientes com a doença  $d_i$  no banco de dados.

$$\hat{p}(s_1, s_2, \dots, s_M / d_i) = \frac{\text{Número de pacientes com a doença } d_i \text{ nos quais ocorreu a configuração } i \text{ de observações } s_1, s_2, \dots, s_M}{\text{Nº total de pacientes com a doença } d_i}$$

$i = 1, \dots, N$  (III.4)

### III.3.2 - A Pressuposição de Independência dos Descritores

Um dos maiores problemas do método bayesiano sempre foi a sua voracidade por dados para a estimação das probabilidades condicionais, que crescem exponencialmente em número com a quantidade de descritores. Para apreciarmos a magnitude do problema basta dizer que no estudo de dor abdominal aguda citado na seção II.3.1.6 seriam necessárias  $3,75 \cdot 10^9$  estimativas usando descritores binários, e para o ser humano como sistema, descrito por Bleich,  $5,58 \cdot 10^{605}$  estimativas.

Tendo em vista a impossibilidade de determinar tal volume de estimativas, tornou-se muito popular adicionar ao modelo a pressuposição de independência dos descritores dada uma doença. Dizemos que dois descritores são estatisticamente independentes se a probabilidade de obtermos um determinado resultado na mensuração do primeiro descritor não for alterada pelo conhecimento do valor do 2º descritor e vice-versa. Com a pressuposição de independência entre descritores basta estimar as probabilidades marginais  $p(s_{kj} / d_i)$   $i=1, \dots, N$  o que reduz o número de

estimativas necessárias drasticamente para  $N.M(n-1)$ , o que corresponderia a 203 parâmetros no estudo de dor abdominal aguda e  $1,2.10^8$  no exemplo de Bleich.

A estimativa da probabilidade  $p(s_{jk}/d_i)$  pode ser obtida pela proporção dos pacientes do banco de dados de treinamento com a doença  $d_i$  que apresentaram o valor observado  $s_{jk}$  por ocasião da mensuração do descritor  $S_j$ . Com essa pressuposição calcula-se  $p(\underline{s}/\underline{d}_j)$  por

$$p(\underline{s}/\underline{d}_j) = \prod_{i=1}^M p(s_i/d_j) \quad (\text{III.5})$$

Para a obtenção dessa significativa simplificação basta que os descritores sejam estatisticamente independentes dentro de cada uma das classes diagnósticas e não para toda população. Daí o nome freqüentemente encontrado na literatura de Pressuposição de Independência Condicional de Descritores.

### III.3.3 - Estimação de Parâmetros a partir de um Banco de Dados Real

Como dissemos, para a estimacão dos parâmetros em um modelo bayesiano de classificacão diagnóstica, necessitamos de um banco de dados de calibracão, o que representa uma grande dificuldade para o pesquisador, face à quase inexistência de banco de dados em medicina.

Uma alternativa adotada por muitos pesquisadores foi utilizar como estimativas das probabilidades condicionais, probabilidades subjetivas fornecidas por especialistas na área médica em estudo. No entanto, Leaper e de Dombal [104] demonstraram em seu trabalho sobre a síndrome de dor abdominal aguda que o emprego de probabilidades subjetivas fez com que a exatidão do diagnóstico de seu modelo caísse de 91% para 82%, um valor

bastante próximo aos 79% obtidos por especialistas. Leaper suspeita que o uso bastante disseminado de probabilidades subjetivas possa ter sido a causa do desempenho sofrível de muitas aplicações do método bayesiano. Sugere que os sistemas diagnósticos devem basear-se em dados cuidadosamente colhidos em larga escala em estudos reais, em vez de estimativas de médicos.

O banco de dados de treinamento pode ser construído prospectivamente ou retrospectivamente. No 1º caso, temos um maior grau de controle na seleção das variáveis e na qualidade dos dados, mas também possivelmente um custo elevado e um longo prazo de acumulação de dados se os casos médicos em estudo tiverem uma freqüência de ocorrência baixa. No 2º caso podemos aproveitar toda a experiência passada do hospital, registrada em seu arquivo médico, porém praticamente não temos controle sobre a qualidade dos dados.

## DEFICIÊNCIAS DE BANCOS DE DADOS REAIS

Ao final do trabalho de colheita de dados teremos construído um banco de dados real que apresentará em maior ou menor grau as seguintes imperfeições, do ponto de vista do objetivo de estimação de probabilidades condicionais.

### A - Insuficiência do Número de Casos

O conjunto de treinamento deve ter um tamanho suficiente para proporcionar estimativas estáveis das probabilidades. O tamanho dependerá em última análise do número de estimativas a serem realizadas, a qual é função do nº de classes, nº de descritores, nº de observações, bem como da aproximação utilizada para representar as dependências entre descritores [89]. Nem sempre é possível obter um número adequado de casos, principalmente quando dentre as classes diagnósticas constam doenças de ocorrência pouco freqüente ou estamos considerando interações

de ordem elevada entre descritores.

## **B - Imprecisão dos Dados Médicos**

A imprecisão dos dados médicos tem sido apontada na literatura e advém de uma série de fatores tais como:

- Falta de padronização nas definições de dados médicos.
- Variabilidade inter e intra-observadores no processo de medição.
- Problemas técnicos na realização de exames.
- Erros de lançamento.
- Ilegibilidade dos registros médicos.

Este problema tende a ser mais grave nos bancos de dados construídos retrospectivamente.

Dependendo da área médica em consideração, nem sequer conseguiremos conhecer o diagnóstico definitivo de cada paciente incluído no bando de dados de treinamento, tendo de nos contentar apenas com a opinião diagnóstica do especialista, a qual pode estar sujeita a considerável margem de erro.

## **C - Registros Incompletos**

Nem sempre é possível colher dados de todos os descritores sobre um paciente em função de fatores tais como:

- Exames não indicados a um paciente por razões médicas.
- Óbito do paciente.
- Não disponibilidade temporária de um ou mais exames no hospital.
- Esquecimento da colheita de um dado.
- Esquecimento de lançamento de resultados no prontuário do paciente.



- Extravio de resultados de exames.
- Ilegibilidade.

## ANOMALIAS NA ESTIMAÇÃO DE PARÂMETROS

Todas as imperfeições do banco de dados de treinamento citadas na seção anterior prejudicam a exatidão das estimativas de parâmetros.

Vamos discutir a seguir algumas anomalias na estimação de parâmetros, que podem ocorrer com bancos de dados de treinamento pequenos e com registros incompletos, e propor soluções para corrigir seus efeitos no modelo bayesiano com suposição de independência de descritores. Consideramos as seguintes anomalias:

1. Não ocorrência da observação  $s_{jk}$  no banco de dados de treinamento para todos os pacientes com a doença  $d_i$ .
2. Não ocorrência da observação  $s_{jk}$  para todos os pacientes do banco de dados de treinamento.
3. Não mensuração do descritor  $S_j$  em todos os pacientes do banco de dados de treinamento.

No 1º caso teremos uma estimativa nula para a probabilidade  $p(s_{jk}/d_i)$ . Isto implica que um paciente novo no qual tenha ocorrido a observação  $s_{jk}$ , terá um valor nulo para a probabilidade a posteriori da doença  $d_i$ , sendo portanto esta doença retirada de consideração na análise, independentemente da quantidade de evidência que exista em seu favor nas demais observações realizadas.

No 2º caso as probabilidades condicionais  $p(s_{jk}/d_i)$   $i=1,2,\dots,N$  serão nulas introduzindo problemas computacionais

no cálculo das probabilidades a posteriori de pacientes para os quais tenha ocorrido a observação  $s_{jk}$ . Teremos a ocorrência de divisões zero/zero, em todas as classes e como resultante a perda de toda a informação existente.

A solução proposta para os problemas 1 e 2 é a de impedir a ocorrência de probabilidades condicionais  $p(s_{jk}/d_i)$  nulas substituindo-se o valor zero por um valor  $p_{\min}$  não nulo arbitrariamente pequeno. Esta solução impede no 1º caso a exclusão completa de uma classe diagnóstica da competição com as demais. No 2º caso o valor  $p_{\min}$  é eliminado automaticamente por ocorrer no numerador e denominador de (III.1) durante o cálculo das probabilidades a posteriori, evitando-se desse modo as divisões zero/zero e a perda de toda a evidência acumulada. Voltaremos a essa proposta mais adiante na apreciação dos resultados experimentais.

No 3º caso teremos a ocorrência de divisões zero/zero na estimação das probabilidades condicionais  $p(s_{jk}/d_i)$   $i=1,2,\dots,N$ . Nesse caso devido à falta total de informações, propomos atribuir valores uniformes às probabilidades condicionais

$$p(s_{jk}/d_i) = \frac{1}{n(k)} \quad i = 1,2,\dots,N \quad (\text{III.6})$$

#### III.3.4 - Modelos com Classes Aglutinadas

Existem situações em que podemos desejar aglutinar 2 ou mais classes em uma super classe diagnóstica. Tal é o caso por exemplo, do método de multi-elegibilidade, citado anteriormente, onde criamos modelos formados por 2 classes: uma classe correspondente à doença  $d_i$  e uma outra classe, aglutinando-se as demais doenças.

Flehinger [55] demonstrou que a aglutinação de classes quebra a condição de independência entre os descritores existentes anteriormente, a não ser em circunstâncias especiais descritas a seguir:

A condição de independência entre os descritores é mantida, apenas quando as probabilidades condicionais de valores observados se mantiverem constantes no processo de aglutinação, ou seja:

$$p(s_{ij}/d_{\text{aglutinada}}) = p(s_{ij}/d_{\text{compon.1}}) = \dots = p(s_{ij}/d_{\text{compon.k}})$$

(III.7)

Na prática essa condição é muito restritiva, não sendo portanto obedecida. Como consequência, temos, além da quebra da suposição de independência, inconsistência entre os valores das probabilidades a posteriori calculados no modelo primitivo e no novo modelo.

Para evitar este tipo de inconsistência, sugerimos que as probabilidades a posteriori seja calculadas no modelo primitivo e que as classes sejam aglutinadas depois disso, somando-se então os valores calculados de probabilidades a posteriori de acordo com o esquema de aglutinação desejado. Chamamos a esses modelos de Modelos com Classes Aglutinadas a Posteriori.

### III.3.5 - Estratégias de Calibração e Teste

Podemos distinguir 2 conjuntos de dados na aplicação de modelos bayesianos de classificação diagnóstica, a saber:

- O Banco de Dados de Treinamento, já citado, destinado à estimação dos parâmetros do modelo.
- O Banco de Dados de Teste, destinado a testar o desempenho do modelo.

Jacquez [89] sugeriu três estratégias diferentes para o processo de calibração e teste de modelos bayesianos de classificação diagnóstica, a saber:

#### **A - O Banco de Dados como Universo**

Nesse caso assume-se que a população é finita e limitada ao conjunto de treinamento, cujos dados são também utilizados para testar o modelo. Isto proporciona uma informação muito interessante, qual seja, a melhor performance que o método pode obter, já que as estimativas dos parâmetros são exatas por definição.

#### **B - O Banco de Dados como uma Amostra de um Universo Estacionário**

O banco de dados é considerado como uma amostra de um universo estacionário mas desconhecido. O desempenho do modelo de classificação diagnóstica nessa estratégia será naturalmente inferior ao da situação anterior.

#### **C - O Banco de Dados como uma Amostra de um Universo Variável no Tempo**

Nesse caso, considera-se que as probabilidades a priori das doenças variam ao longo do tempo em função de surtos epidêmicos, sazonalidades e tendências a longo prazo. Nesse caso o banco de dados para estimar as probabilidades  $p(d_i)$  deve ser constituído apenas pelas amostras mais recentes, havendo a necessidade de balancear a precisão das estimativas (tanto maior quanto maior for o banco de dados) com sua sensibilidade a mudanças [89,106].

### **III.4 - INFLUÊNCIA POTENCIAL DA DEPENDÊNCIA ENTRE DESCRITORES SOBRE O DESEMPENHO**

A pressuposição de independência entre os descritores foi usada na maioria dos estudos de modelos bayesianos de clas-

sificação diagnóstica, como forma de reduzir o número de estimativas a serem calculadas, mesmo em situações que violavam claramente esta condição. De fato, é natural esperar que numa situação complexa como é a do diagnóstico médico, ocorram muitas interações entre as variáveis. Flehinger [55] aponta que a dependência entre descritores pode aparecer de 3 formas:

- A - Mecanismos interrelacionados que se desenvolvem no mesmo cenário genético e ambiental.
- B - Descritores que só são mensurados quando certas observações específicas estão presentes.
- C - Mudanças na organização de doenças feita pelo homem.

Muitos dos trabalhos que usaram a pressuposição de independência obtiveram bons resultados. Uma pergunta colocada por alguns pesquisadores [63,129] é se esses resultados não poderiam ser melhorados através da consideração das interações entre descritores? E se podemos atribuir algumas das situações em que resultados sofríveis foram obtidos à falta de independência? O potencial de melhoria de desempenho pode ser apreciado no seguinte simples exemplo:

		$s_i$		$s_1 s_2$			
		0	1	00	01	10	11
D=1		30	30	30	0	0	30
D=2		20	20	0	20	20	0

Tabela III.3

A Tabela (III.3) mostra 2 descritores  $S_i$   $i=1,2$  que individualmente não permitem nenhuma discriminação entre as doenças 1 e 2, mas que analisados conjuntamente permitem discriminação perfeita

Dos problemas relevantes em diagnose o de interações entre descritores certamente é um dos problemas mais sérios. Embora a maioria dos autores não tenha conscientemente dado atenção à este problema, podemos encontrar na literatura várias aproximações usadas para lidar com o problema da interação entre descritores. Entre os principais, citamos:

- A - Grupar descritores dependentes, de modo que apenas uma observação seja apresentada – utilizada por Warner [181].
- B - Eliminar descritores que são altamente correlacionados com descritores já mensurados – utilizada por Nugent [129] e Ben-Bassat [12]. Apresenta o grave inconveniente de desperdiçar informação.
- C - Ortogonalização de Graham Schmidt [129]
- D - Aproximação de Lincoln-Parker [129,63]
- $$p(s/d_i) = p(s_1/d_i) \cdot p(s_2/s_1 \cap d_i) \cdot p(s_3/s_2 \cap d_i) \dots p(s_n/s_{n-1} \cap d_i)$$
- (III.8)
- E - Aproximação de Fryback [63]. Equivalente à anterior com a ordenação dos descritores por um procedimento heurístico baseado em correlações.
- F - Criação por classe diagnóstica, de Aglomerados ("Clusters") condicionalmente independentes formados por descritores que apresentam dependência entre si, proposta por Norusis e Jacquez [130].

Norusis e Jacquez [129,130], usando um banco de dados de doenças cardiovasculares, realizaram uma investigação visan-

do estabelecer as conseqüências da pressuposição de independência quando a distribuição de probabilidade conjunta é considerada conhecida (Estratégia — "O Banco de Dados como Universo"), bem como para avaliar vários modelos estabelecidos para estimação de probabilidades conjuntas que incorporam interações entre variáveis. Seus resultados suportam a conjectura de que a pressuposição de independência pode reduzir substancialmente o desempenho do sistema bayesiano de classificação, mesmo quando as correlações entre sintomas são pequenas e principalmente no caso de doenças moderada ou dificilmente diferenciáveis. Concluíram também que modelos que incluem dependências de 2ª ordem (tais como função discriminante de Fisher, árvores ótimas de dependência, expansão de Bahadur de 2ª ordem) conduzem a resultados imprevisíveis quando comparados com o modelo de independência, e sugeriram que interações de ordem mais elevada (até  $0,5 \times n^{\circ}$  de descritores) sejam incorporadas nos procedimentos de estimação. Analisando a influência do crescimento do Banco de Dados de Treinamento, verificaram que o desempenho dos modelos Função Discriminante e Independência foi pouco afetado. O desempenho do modelo atuarial no entanto, melhorou rapidamente com o crescimento do banco de dados, igualando o desempenho do modelo de independência para classes facilmente diferenciáveis e superando-o amplamente para classes moderadamente separáveis, quando o  $n^{\circ}$  de casos por doença atingiu  $1/2$  do  $n^{\circ}$  total de vetores de sintomas possíveis  $n^M$ . O modelo com aglomerados de descritores, proposto pelos autores, apresentaram taxas de desempenho que se comparam muito favoravelmente com as obtidas pelos modelos de independência, função discriminante linear, árvore ótima de dependência e expansões de Bahadur de várias ordens. Enfatizaram eles contudo, que para amostras extremamente pequenas mesmo na presença de interações entre descritores, o modelo de independência pode ser uma solução prática.

### III.5 - MODELO BAYESIANO DE CLASSIFICAÇÃO DIAGNÓSTICA BASEADO EM AGLOMERADOS DE DESCRITORES INTERDEPENDENTES

Dentre as aproximações já propostas para tratar das interações entre descritores, a de utilização de aglomerados de descritores interdependentes parece-nos a mais conveniente devido a sua robustez, simplicidade e baixa exigência por dados. Vamos dedicar a ela o restante deste capítulo, revendo inicialmente a proposta de Norusis [130].

Em muitas situações é razoável postular a existência de diferentes interrelacionamentos entre variáveis como já comentamos anteriormente. No contexto do diagnóstico médico parece natural que grupos distintos de descritores interdependentes possam existir e que diferentes grupos ocorram em diferentes doenças. As observações de um mesmo sistema orgânico estarão mais intimamente associadas entre elas mesmas, do que com resultados de outros sistemas.

O método proposto por Norusis consiste em formar, para cada classe de doença  $d_i$ , aglomerados  $A_{ij}$  de descritores que estejam estatisticamente associados. Para cada doença  $d_i$  teremos um conjunto de aglomerados  $\{A_{i1}, A_{i2}, \dots, A_{i.g(i)}\}$ . Os aglomerados são estatisticamente independentes de modo que a estimação das probabilidades condicionais é dada por

$$p(s/d_i) = \prod_{j=1}^{g(i)} p(a_{ijk}) \quad i=1, \dots, N \quad (\text{III.9})$$

onde  $a_{ijk} = k^a$  configuração de observações do aglomerado  $A_{ij}$ .

O modelo baseado em aglomerados de descritores é uma combinação dos modelos de independência e atuarial. Se todos os descritores forem independentes teremos  $M$  aglomerados individuais e recaímos no modelo de independência. Se todos os descritores estiverem associados formando um único aglomerado, recaímos



remos no modelo atuarial.

A maior vantagem do modelo baseado em aglomerados está na redução do número de estimativas de probabilidades conjuntas que precisam ser obtidas do banco de dados. Por exemplo, se tivermos 10 variáveis dicotômicas e formarmos dois aglomerados de 5 descritores cada, apenas  $64=(2^5+2^5)$  probabilidades precisarão ser estimadas em lugar das 1024 exigidas pelo modelo completo. Trata-se de uma redução significativa. O número real de estimativas a serem feitas é uma função do número e do tamanho de aglomerados. Basta, no entanto, particionar um descritor independente para cortar pela metade o número de estimativas (se o descritor for binário). Esta redução é de particular importância para bancos de dados pequenos, o que corresponde à situação prática mais comum.

Entre outras vantagens do modelo baseado em aglomerados, citamos:

1. Inclusão de interações entre descritores de ordem mais elevada sem computações complexas de muitos parâmetros.
2. Implementação flexível já que diferentes aglomerados podem ser escolhidos para cada classe diagnóstica em consideração.
3. Aplicação tanto a dados dicotômicos como politômicos.

### III.5.1 - Formação de Aglomerados de Descritores

Diversas estratégias podem ser usadas para delinear os aglomerados de descritores. Caso exista conhecimento a priori sobre interrelacionamentos biológicos que definam grupos naturais de variáveis, podemos usá-los na formação de aglomerados. Outra possibilidade é a definição de aglomerados com base em critérios subjetivos por equipes de médicos, usada por Gustafson [78], e que apresenta a vantagem de dispensar a existência de um banco de dados. Finalmente, os aglomerados de descritores podem ser estabelecidos matematicamente a partir de um

banco de dados. Norusis [130] agrupou os sintomas de uma forma simples com base na magnitude dos coeficientes de correlação de Pearson. Em nosso trabalho de pesquisa também empregamos um procedimento matemático para determinação dos aglomerados de descritores, porém, baseado em testes de hipótese quanto à interação entre pares de descritores.

O procedimento utilizado é descrito a seguir e deve ser aplicado para cada classe diagnóstica.

#### PROCEDIMENTO PARA FORMAÇÃO DE AGLOMERADOS DE DESCRITORES PARA A DOENÇA $d$

1. Selecionar no banco de dados todos os casos pertencentes à classe diagnóstica  $d$

Chamamos de  $h$  o número de casos obtidos.

2. Para cada par de descritores  $(i, j)$   $i, j = 1, 2, \dots, M$   
 $i \neq j$

- Calcular a matriz de contingência  $X$

$X$  tem dimensão  $n(i) \times n(j)$

- Calcular a estatística  $V_{ij}$

$$V_{ij} = \sum_{k=1}^{n(i)} \sum_{e=1}^{n(j)} \frac{(x_{ke} - h \hat{w}_k \hat{z}_e)^2}{h \cdot \hat{w}_k \cdot \hat{z}_e}$$

onde

$$\hat{w}_k = \frac{1}{h} \sum_{g=1}^{n(j)} x_{kg} \quad k \in \{1, \dots, n(i)\}$$

$$\hat{z}_e = \frac{1}{h} \sum_{g=1}^{n(i)} x_{ge} \quad e \in \{1, \dots, n(j)\}$$

$$V_{ij} \sim \chi^2_{[n(i)-1] \cdot [n(j)-1]}$$

- Com base num nível de significância  $\alpha$  previamente fornecido realizar o teste de hipótese quanto à in dependência dos descritores  $i$  e  $j$ , comparando  $V_{ij}$  com o valor  $\chi^2_{\text{crítico}}$ .
- Armazenar o resultado do teste numa matriz de simi laridade  $U$  tal que

$$u_{ij} = 0 \quad \text{- não existe associação entre } i \text{ e } j$$

$$u_{ij} = 1 \quad \text{- existe associação entre } i \text{ e } j$$

3. Aplicar o método de ligação simples (método do mais próximo vizinho) sobre a matriz de similaridade  $U$  para formar os aglomerados de descritores.

No nosso caso, isto corresponde a dizer que os descritores  $i$  e  $j$  pertencem ao mesmo aglomerado se  $u_{ij} = 1$ .

4. Fim.

O procedimento de formação de aglomerados proposto a presenta as seguintes vantagens:

- Aproveita ao máximo as interações existentes na formação dos aglomerados, isto devido à propriedade de encadeamento, típica do método do mais próximo vizinho.

- Permite determinar diferentes estruturas de aglomerados em função do nível de significância adotado nos testes de hipótese. Verifica-se o crescimento do tamanho médio dos aglomerados com o aumento de  $\alpha$ .
- Permite determinar em função do tamanho do banco de dados de treinamento disponível, qual o valor do nível de significância (e conseqüentemente qual a estrutura de aglomerados) que propicia o máximo desempenho do modelo de classificação diagnóstica em estudo.

### III.5.2 - Estimação de Probabilidades Condicionais de Aglomerados

Seja  $A_{ij}$  o  $j^{\circ}$  aglomerado de descritores correspondentes à doença  $d_i$ . Teremos de estimar

$$\prod_{\ell \in A_{ij}} n(\ell)$$

probabilidades condicionais referentes ao aglomerado  $A_{ij}$ , que pode ser encarado como uma nova variável discreta politômica.

Chamamos  $a_{ijk}$ ;  $k = 1, 2, \dots, \prod_{\ell \in A_{ij}} n(\ell)$

às configurações de valores observáveis pertencentes ao aglomerado  $A_{ij}$ . A estimação das probabilidades condicionais conjuntas é feita facilmente por

$$\hat{p}(a_{ijk}/d_j) = \frac{\text{n}^{\circ} \text{ de ocorrências de pacientes com a doença } d_i \text{ nos quais ocorreu a configuração de observações } a_{ijk}}{\text{N}^{\circ} \text{ total de pacientes com a doença } d_i}$$

## ANOMALIAS NA ESTIMAÇÃO COM BANCO DE DADOS REAL

Tendo em vista as deficiências dos bancos de dados reais apontadas anteriormente, podem ocorrer as seguintes anomalias no processo de estimação e inferência:

1. Nenhuma mensuração do conjunto de descritores do aglomerado  $A_{ij}$  foi realizada em pacientes com a doença  $d_i$ .
2. A configuração de observações  $a_{ijk}$  não ocorreu em nenhum paciente com a doença  $d_i$ .
3. Ocorrência do 2º tipo de anomalia concomitantemente em todas classes diagnósticas.

As soluções propostas são semelhantes às aquelas do modelo de independência.

No 1º caso temos divisões zero/zero e nenhuma informação. A proposta é de distribuir o peso probabilístico uniformemente entre as configurações de observações:

$$\hat{p}(a_{ijk}/d_i) = \frac{1}{\prod_{\ell \in A_{ij}} n(\ell)} \quad (\text{III.11})$$

No 2º caso a estimativa de  $p(a_{ijk}/d_i)$  apresenta valor nulo, o que implica no afastamento de consideração da doença  $d_i$  para pacientes com  $a_{ijk}$ , independentemente da quantidade adicional de evidência que eventualmente exista em favor de  $d_i$ . Este problema ocorrerá com maior frequência para bancos de dados pequenos e aglomerados de grande dimensão.

No 3º caso teremos divisões 0/0 no cálculo das probabilidades a posteriori com perda total da informação.

As anomalias 2 e 3 podem ser solucionadas substituindo

do as estimativas de valor nulo por um valor  $p_{\min}$  não nulo arbitrariamente pequeno. Assim, nenhuma classe diagnóstica será afastada definitivamente da competição e as divisões 0/0 no 3º problema serão evitadas aproveitando-se a informação disponível.

### III.5.3 - Algoritmo de Estimação das Probabilidades Condicionais Conjuntas dos Aglomerados Baseado numa Representação Esparsa

É fundamental para o projeto de um algoritmo de estimação reconhecer que, em situações práticas na medicina a desproporção provável entre o tamanho da amostra e o número potencial de configurações de valores observáveis por aglomerado é de tal ordem que fará com que a maior parte das estimativas de probabilidades condicionais sejam nulas. Um banco de dados médicos raramente ultrapassa algumas centenas de casos por classe diagnóstica, ao passo que um aglomerado formado por apenas 15 descritores dicotômicos já apresenta 32768 células.

Tendo em vista este fato podemos adotar algumas orientações visando a melhoria radical da eficiência do algoritmo e economia de memória, entre as quais:

- Usar uma representação esparsa só calculando e armazenando as estimativas de probabilidades condicionais maiores que zero.
- Não gerar todas as configurações de valores observáveis para então verificar se elas ocorreram no banco de dados, mas trabalhar diretamente apenas com as existentes no banco de dados uma vez que só estas terão estimativas de probabilidades não nulas.
- Ler o banco de dados uma única vez.

A seguir apresentamos um algoritmo que segue estas orientações:

## ALGORITMO DE ESTIMAÇÃO

### 1. Set up

- Criar lista de classes diagnósticas.
- Criar lista de aglomerados por classe diagnóstica.
- Estabelecer contadores de ocorrências de mensurações dos aglomerados com valores iniciais nulos.
- Criar listas vazias de imagens de configurações de resultados e também correspondentemente de contadores de ocorrências, para cada aglomerado de cada classe diagnóstica:

### 2. Para cada caso do banco de dados de treinamento faça:

- Identificar a classe diagnóstica  $i$  do caso.
- Para cada aglomerado  $j$  pertencente a classe diagnóstica  $i$  faça:
  - Verificar se todos os descritores do aglomerado  $A_{ij}$  foram mensurados.
  - Se o aglomerado  $A_{ij}$  foi mensurado então faça:
    - Ler a configuração  $a_{ijk}$  de valores observados no caso.
    - Incrementar o contador de ocorrências de aglomerados  $A_{ij}$ .
    - Varrer a lista de configurações de  $A_{ij}$  procurando encontrar a configuração  $a_{ijk}$ .  
Se encontrar, incrementar o contador de configurações correspondente de uma unidade.  
Se não encontrar, acrescentar a configuração  $a_{ijk}$  a esta lista e estabelecer o contador de configurações associado com valor inicial 1.

### 3. Para cada classe diagnóstica $i$ faça:

- Para cada aglomerado  $j$  pertencente à doença  $i$  faça:
  - Ler sua contagem de ocorrência  $T$
  - Se sua  $T > 0$  então:
    - Para cada configuração de observações  $a_{ijk}$  pertencente ao aglomerado  $A_{ij}$  faça:
      - Ler contagem  $C$  de ocorrências da configuração em questão.

- Estimativa  $\leftarrow C \div T$
- Publicar e armazenar a estimativa  $p(a_{ijk}/d_i)$ .

4. Fim.

**Observação:**

No momento de consultarmos o arquivo gerado pelo algoritmo de estimação, caso a estimativa da probabilidade condicional da configuração desejada não for encontrada no arquivo, podemos ter 2 situações possíveis:

- A - Se o contador de ocorrências do aglomerado for zero, adotaremos como estimativa  $1/N^{\circ}$  total de configurações possíveis.
- B - Se o contador for maior que zero a estimativa é nula e adotaremos como valor  $p_{\min}$ .

**CONSIDERAÇÕES QUANTO A EFICIÊNCIA COMPUTACIONAL E ESPAÇO EM MEMÓRIA**

O trabalho do algoritmo de estimação é da ordem

$$O [N_{ad} \cdot N_{ea} (N + N_c)]$$

onde

$N_c$  = Número de casos do banco de dados de treinamento.

$N_{ad}$  = Número de aglomerados por doença.

$N_{ea}$  = Número de estimativas não nulas por aglomerado.

$N$  = Número de doenças.



Levando em consideração que

$$N_{ea} \leq \frac{N_c}{N} \quad (\text{III.12})$$

$$N_{ad} \leq M \quad (\text{III.13})$$

$$N \ll N_c$$

$$\text{Trabalho} = O(N_c^2 \cdot \frac{M}{N}) \quad (\text{III.13})$$

onde M = número de descritores

O espaço em memória pode ser expresso por:

$$O(N \cdot N_{ad} \cdot N_{ea}) = O(N_c \cdot M) \quad (\text{III.14})$$

Conseguimos, portanto, obter um algoritmo de estimação cujos trabalho e espaço de memória requerido são proporcionais ao número de descritores, e independentes do tamanho dos aglomerados tornando-se desta forma o problema da explosão combinatorial de estimativas. A eficiência e o espaço de memória requerido dependem do tamanho do banco de dados de treinamento, que em última análise, representa o conhecimento que dispomos sobre o problema médico em questão.

#### III.5.4 - Algoritmos de Inferência Diagnóstica em Modelos com Aglomerados de Descritores Baseado numa Representação Esparsa

Apresentamos 2 algoritmos para o cálculo das probabilidades a posteriori das doenças a partir do vetor de observações do paciente. Ambos incorporam as sugestões para correção das anomalias discutidas na seção III.5.2.

O 1º algoritmo – DIAGPREP – utiliza as estimativas de probabilidades condicionais calculadas pelo método da seção III.5.3. Tem como principal vantagem a maior rapidez de processamen-

to.

O 2º algoritmo – **DIAGNPREP** – realiza além da inferência diagnóstica, também, a estimação dos parâmetros necessários para tanto, a partir do banco de dados. É mais lento que o primeiro em função deste trabalho adicional. Suas principais características e vantagens são:

- Estima apenas as probabilidades condicionais das configurações observadas no paciente.
- Realiza apenas uma leitura do banco de dados.
- Não dispense memória em arquivos permanentes de estimativas.
- Trabalha com aglomerados de qualquer dimensão.
- Apresenta características de auto-aprendizado. Qualquer caso que seja adicionado ao banco de dados, estará automaticamente contribuindo para a análise do próximo paciente.

Um ponto que deve ser previsto no projeto dos algoritmos de inferência é a capacidade de lidar com dados incompletos de pacientes. Nessa situação teremos eventualmente aglomerados com alguns ou mesmo todos seus descritores não mensurados. O critério adotado nesses casos foi o de reduzir o tamanho do aglomerado, de forma a englobar apenas os descritores mensurados. No algoritmo **DIAGPREP** é necessário chamar uma rotina especial para realizar a estimativa da configuração.

Apresentamos a seguir os 2 algoritmos:

### **ALGORITMO DIAGPREP - INFERÊNCIA DIAGNÓSTICA**

1. Ler o vetor de observações  $\underline{s}$  do paciente.
2. Criar o vetor **PC** de dimensão **N** com valores iniciais unitários.
3. Para cada classe diagnóstica **d** faça:

• Para cada aglomerado  $A_{dj}$  da doença  $d$  faça

- Ler a configuração  $a_{djk}$  de observações do paciente
- Se todos descritores de  $A_{dj}$  foram mensurados no paciente.

Então se contador de ocorrências de  $A_{dj} = 0$

$$\text{Então } \bullet \text{ PC}(d) \leftarrow \text{PC}(d) \cdot \frac{1}{\prod_{l \in A_{dj}} n(l)}$$

Se não • Procurar  $a_{djk}$  no arquivo de configurações de  $A_{dj}$ .

• Se for encontrado  $a_{djk}$

Então • Ler  $p(a_{djk}/d)$

•  $\text{PC}(d) \leftarrow \text{PC}(d) \cdot p(a_{djk}/d)$

Se não •  $\text{PC}(d) \leftarrow \text{PC}(d) \cdot p_{\min}$

Se não • Se pelo menos um descritor foi medido no paciente.

Então • Elimine temporariamente os descritores não mensurados da definição do aglomerado  $A_{dj}$ .

• Estime  $p(a_{ijk}/d)$  diretamente do banco de dados.

•  $\text{PC}(d) \leftarrow \text{PC}(d) \cdot p(a_{ijk}/d)$

4. Para cada doença  $d$  faça:

- Calcular a probabilidade a posteriori através de

$$p(d/s) \leftarrow \frac{p(d) \cdot PC(d)}{\sum_{d' \in \Delta} p(d') \cdot PC(d')}$$

5. Fim.

### ALGORITMO DIAGNPREP - ESTIMAÇÃO E INFERÊNCIA DIAGNÓSTICA

1. Ler vetor de observações  $\underline{s}$  do paciente Z.
2. Para cada classe diagnóstica defina as seguintes listas:
  - Lista NAGLOM de contadores de ocorrências de medições de aglomerados  $A_{dj}$ , com valores iniciais nulos.
  - Lista NCONFIG de contadores de ocorrências das configurações  $a_{djk}$  dos sintomas observados no paciente, com valores iniciais nulos.
  - Lista MARCA de indicação de não mensuração de aglomerados no paciente, com valores iniciais nulos.
3. Para cada caso c do banco de dados de treinamento faça:
  - Ler classe diagnóstica  $d$  do caso.
  - Para cada aglomerado  $A_{dj}$  da classe diagnóstica  $d$  faça:
    - Ler configuração  $a_{djk}$  de observações apresentadas pelo paciente Z.
    - Se nenhum descritor de  $A_{dj}$  foi mensurado no paciente Z. Então  $MARCA_d(j) \leftarrow 1$ .

Se não • Se um ou mais descritores de  $A_{dj}$  não foram mensurados no paciente Z

Então elimine temporariamente os desc $\bar{r}$ i não mensurados no paciente Z da definição do aglomerado  $A_{dj}$ .

• Se todos os descritores pertencentes ao aglomerado  $A_{dj}$  foram mensurados no caso C do banco de dados.

Então • Incremente de 1 a  $j^{\circ}$  contador fa lista  $NAGLOM_d$

• Ler a configuração  $a_{djz}$  de sintomas do caso C do banco de dados.

• Se a configuração  $a_{djk}$  for igual a  $a_{djz}$

Então incremente de 1 o  $j^{\circ}$  contador da lista  $NCONFIG_d$

4. Definir o vetor PC de dimensão N, com valores iniciais unitários.

5. Para cada classe diagnóstica d, faça:

• Para cada aglomerado  $A_{dj}$  da classe diagnóstica d faça

• Se  $MARCA_d(j) = 1$

Então • Se  $NAGLOM_d(j) = 0$

Então •  $PC(d) \leftarrow PC(d) \cdot \frac{1}{\prod_{\ell \in A_{dj}} n(\ell)}$

Se não •• Se  $NCONFIG_d(j) = 0$

Então  $PC(d) \leftarrow PC(d) \cdot P_{\min}$

Senão  $PC(d) \leftarrow PC(d) \cdot \frac{NCONFIG_d(j)}{NAGLOM_d(j)}$

6. Para cada classe diagnóstica  $d$  faça:

- Calcular a probabilidade a posteriori através de

$$p(d/s) \leftarrow \frac{p(d) \cdot PC(d)}{\sum_{d'=1} p(d') \cdot PC(d')}$$

7. Fim

#### EFICIÊNCIA COMPUTACIONAL DOS ALGORITMOS DE INFERÊNCIA

O trabalho do algoritmo DIAGPREP pode ser expresso por:

$$O(N \times N_{ad} \times N_{ea} + N \times N_{ad}) \quad (III.15)$$

Usando as simplificações de (III.12) temos:

$$\text{Trabalho de DIAGPREP} = O(M \cdot N_c) \quad (III.16)$$

O trabalho do algoritmo DIAGNPREP pode ser expresso por:

$$O(N_c \cdot N_{ad} \cdot N_{va} + N \cdot N_{ad}) \quad (III.17)$$

Onde  $N_{va}$  = Número de descritores por aglomerado.

Considerando  $N_{ad} \cdot N_{va} \cong M$  e as simplificações de (III.12), temos

$$\text{Trabalho de DIAGNPREP} = O(M \cdot N_c) \quad (\text{III.18})$$

Comparando as primeiras parcelas das expressões originais de trabalho dos algoritmos DIAGPREP com DIAGNPREP (III.15) e (III.17) temos:

$$1^{\text{a}} \text{ parc. trab. de DIAGPREP} \cong N \cdot N_{ad} \cdot N_{ea} \leq N_c \cdot M \cong N_c \cdot N_{ad} \cdot N_{va} \cong 1^{\text{a}} \text{ parc. trab. de DIAGNPREP}$$

Conforme se verifica o algoritmo DIAGNPREP é menos eficiente que o algoritmo DIAGPREP em função do trabalho adicional de estimação executado. A eficiência de ambos algoritmos é praticamente independente do tamanho dos aglomerados, dependendo basicamente do tamanho do banco de dados de pacientes (bastante limitado na prática) e do número de descritores.

Conseguiu-se portanto, obter algoritmos altamente eficientes que conseguem mediante uma exploração inteligente das peculiaridades do problema de classificação diagnóstica, contornar uma situação em que o número de estimativas é tipicamente exponencial.

### III.5.5 - Considerações quanto ao Desempenho de Modelos Baseados no Uso de Aglomerados de Descritores

Para tecer considerações quanto ao desempenho de modelos bayesianos de classificação baseado no uso de aglomerados de descritores interdependentes, é fundamental distinguir qual a estratégia utilizada para treinamento e teste.

## ESTRATÉGIA - O BANCO DE DADOS COMO UNIVERSO

Nesta estratégia a incidência das doenças e as probabilidades condicionais são conhecidas e idênticas às do banco de dados. É de se esperar uma taxa de classificações corretas crescente na medida que cresça o tamanho dos aglomerados, chegando ao máximo desempenho quando tivermos um único aglomerado abarcando todos os descritores. Nesta situação o modelo com aglomerados se confunde com o modelo atuarial, que é ótimo na estratégia "O Banco de Dados como Universo". A taxa de classificações corretas pode mesmo chegar a 100% se não existirem no banco de dados casos com classes distintas, ocupando o mesmo ponto do espaço de padrões. Paradoxalmente isso é tão mais provável quanto menor for o Banco de Dados de Treinamento. O modelo na situação limite se assemelha à busca do casamento perfeito entre um padrão e um elemento de um banco de dados de protótipos.

## ESTRATÉGIA - O BANCO DE DADOS COMO AMOSTRA

Nesta situação mais realista, é de se esperar que a curva de desempenho do modelo em função do tamanho dos aglomerados assumam um formato diferente bem como os resultados sejam inferiores ao da estratégia anterior. Numa 1ª fase do crescimento dos aglomerados, a taxa de classificações corretas deverá ser crescente pois o modelo passa a aproveitar quantidades cada vez maiores da informação contida nas interdependências dos descritores. Porém com a continuação do crescimento dos aglomerados, devido ao tamanho finito do banco de dados de treinamento e ao crescimento exponencial experimentado pelo número de estimativas, verifica-se rapidamente uma rarefação dos casos nas células e uma conseqüente queda no desempenho. Pode-se obter experimentalmente qual o valor ótimo do nível de significância e conseqüentemente qual a estrutura ideal de aglomerados para um modelo, os quais dependerão do tamanho do banco de dados disponível.



---

---

**IV - GERADOR HEURÍSTICO DE PLANOS  
DE INVESTIGAÇÃO MÉDICA**

---

---

#### IV.1 - INTRODUÇÃO

Devido ao arsenal crescente de exames diagnósticos disponíveis para o clínico, o maior desafio da diagnose está deixando de ser o ato intelectual de decisões explanatório-dedutivas para se transformar no ato gerencial de selecionar exames para realização no paciente. A seleção de exames complementares é um problema complexo porque os exames desde uma simples cultura até uma toracotomia diferem amplamente em termos de riscos de morbidade e mortalidade, custo financeiro, complexidade de interpretação, campo de aplicação, duração e sofrimento imposto ao paciente. Torna-se claro, portanto, que para cada paciente existe de acordo com suas hipóteses diagnósticas, uma bateria de exames a serem executados numa ordem definida, que corresponde a uma atitude ótima, qual seja, a de mínimo "custo". ("Custo" devendo ser interpretado num sentido amplo). Não existe razão para aceitar a priori que as seqüências de exames aplicados hoje rotineiramente sejam ótimas neste sentido. As seqüências de exames serão consideradas ótimas sempre com relação a uma estratégia que o médico tenha selecionado em relação ao paciente, em função de seu estado de saúde, da gravidade e urgência do caso, da necessidade de um diagnóstico preciso, de fatores sócioeconômicos e restrições de várias ordens. Vamos nesse capítulo apresentar a proposta de um sistema de apoio à decisão capaz de sugerir planos de investigação médica baseado no emprego de técnicas de programação heurística.

#### IV.2 - OBJETIVO DO GERADOR HEURÍSTICO DE PLANOS DE INVESTIGAÇÃO MÉDICA

O objetivo do modelo proposto é a determinação do plano ótimo de investigação médica visando estabelecer o diagnóstico com uma precisão definida, segundo a estratégia selecionada pelo médico. O plano gerado é específico para o paciente sob a-

nálise. O gerador de planos de investigação proposto baseia-se na visualização do problema de seleção de exames como um problema de busca do caminho de custo mínimo em um grafo, bem como na utilização de técnicas de programação heurística, as quais garantem a obtenção do caminho ótimo fazendo o melhor uso da informação heurística disponível. Formulação análoga, porém sem o uso de programação heurística, foi proposta por Alpérovitch [3].

O gerador de planos de investigação fornecerá como resultado da análise dos dados do paciente, a seqüência ótima de exames a serem realizados. Trata-se portanto, de um gerador de planos de ordem nula, capaz de informar também a seqüência ideal. Comparada com as árvores de decisão, produzidas como resultado da aplicação do método da Teoria da Decisão, a seqüência de exames possui várias vantagens:

- Simplicidade de apresentação
- Facilidade de interpretação pelo médico
- Maior operacionalidade - Podemos desde o início planejar a aplicação de todos os exames da seqüência e realizar tarefas como o preparo do paciente e o agendamento de recursos.

Na construção do modelo proposto, evitamos a aplicação do modelo clássico da Teoria da Decisão em função da dificuldade em atribuir valores numéricos monetários à vida humana e principalmente em função da extrema relutância dos médicos em trabalharem com esses números, o que praticamente inviabiliza o seu uso na rotina médica.

#### IV.3 - ATRIBUTOS DE CUSTO DOS EXAMES

Uma etapa básica no projeto de um gerador de planos de investigação médica consiste em escolher quais os tipos de custos de exames com que vamos trabalhar. Os custos podem ser vistos como utilidades negativas sendo os principais tipos rela

cionados a seguir:

- Custo técnico de realização do exame
  - Custo financeiro
  - Dispêndio de investimentos logísticos
- Morbidade (física e psicológica)
- Mortalidade
- Tempo de hospitalização
- Número de dias fora da vida normal do paciente
- Tempo para obtenção de resultados
- Sofrimento induzido pelo exame

Os diversos tipos de custos deverão ser mensurados em unidades apropriadas. Assim os custos técnicos podem ser medidos em cruzeiros, unidades de serviço, homens/hora; os custos que envolvem tempo são medidos naturalmente em unidades de tempo, morbidade e mortalidade são probabilidades e o sofrimento induzido ao paciente pode ser medido numa escala subjetiva de 0 a 10, por exemplo, ou através do tempo passado sob sofrimento nulo, leve, moderado ou intenso.

#### IV.3.1 - Adição de Utilidades

Para cada atributo de custo de exame devemos definir um operador de adição. Para os atributos medidos em unidades monetárias ou unidades de tempo o operador adição aritmética se aplica. Atributos probabilísticos de risco como mortalidade e morbidade, no entanto, exigem um operador diferente. Assumindo para fins de simplicidade que os danos em termos de mortalidade e morbidade ocasionados ao paciente pelos exames são independentes estatisticamente, temos:

$$\text{Risco} ( \{e_1, e_2, \dots, e_i\} ) = 1 - \prod_{j=1}^i (1-r_j) \quad (\text{IV.1})$$

onde  $r_j$  = variável de risco do exame  $e_j$  (mortalidade ou morbidade).

#### IV.4 - ESTRATÉGIAS DE INVESTIGAÇÃO

O modelo proposto calcula o plano ótimo de investigação de acordo com uma estratégia selecionada pelo médico, a qual corresponde diretamente a uma variável de custo dos exames, que será usada pelo algoritmo de determinação do caminho de custo mínimo.

Diversos tipos de estratégias podem ser vislumbrados, a saber:

##### A - Estratégias Elementares

Uma estratégia elementar é aquela que visa minimizar diretamente um dos tipos de custos de exames relacionados na seção anterior. Assim teremos estratégias elementares de mínimo custo financeiro, mínima morbidade, mínimo sofrimento, etc.

##### A - Estratégia Robusta

Um plano de investigação segundo a estratégia robusta é obtido pela superposição dos planos obtidos com cada estratégia elementar. Nesse caso vamos verificar em cada etapa de investigação o exame mais indicado pelas diversas estratégias, escolhendo-o, caso não tenha sido usado em etapas anteriores do plano. A figura (IV.1) ilustra a determinação de um plano segundo a estratégia robusta.

# PLANOS DE INVESTIGAÇÃO

ESTRATÉGIA					
	MÍNIMO RISCO	MÍNIMO SOFRIMENTO	MÍNIMO CUSTO	MÍNIMA DURAÇÃO	ROBUSTA
1º EXAME	A	A	A	A	<b>A</b>
2º EXAME	B	B	B	D	<b>B</b>
3º EXAME	C	E	C	H	<b>C</b>
4º EXAME	D	C	D	E	<b>D</b>
5º EXAME	E	F	G	B	-

FIGURA IV.1- DETERMINAÇÃO DO PLANO DE INVESTIGAÇÃO DIAGNÓSTICA SEGUNDO A ESTRATÉGIA ROBUSTA

## C - Estratégias Compostas

As estratégias compostas permitem aos médicos elaborarem planos de investigação que sejam ótimos de acordo com uma combinação que englobe vários pontos de vista simultaneamente. As estratégias compostas são obtidas por combinações de vários atributos de custo dos exames, produzindo como resultado uma única variável numa escala contínua. Para simplificar o processo de determinação de combinações consistentes de atributos, podemos buscar auxílio na Teoria de Utilidades Multi-Atributos. Esta teoria busca examinar pressuposições de independência e balanceamento entre os atributos com vistas a deduzir formas de combinação de fácil avaliação [94].

## IV.5 - ESTRUTURA DO SISTEMA

A Figura (IV.2) mostra o esquema do sistema de apoio à decisão proposto com o objetivo de gerar planos de investigação diagnóstica. Como se observa, o sistema tem dois componentes básicos:

A - Um modelo bayesiano de classificação diagnóstica que contém a descrição da área médica para a qual pretendemos gerar os planos de investigação. Este modelo permite calcular, a partir dos resultados dos exames já realizados e dados básicos do paciente, os valores das probabilidades a posteriori das diversas classes diagnósticas existentes os quais serão utilizados pelo 2º componente do sistema. O modelo bayesiano utilizado no sistema obedece às características definidas no capítulo III sendo a pressuposição de independência dos descritores de uso recomendado.

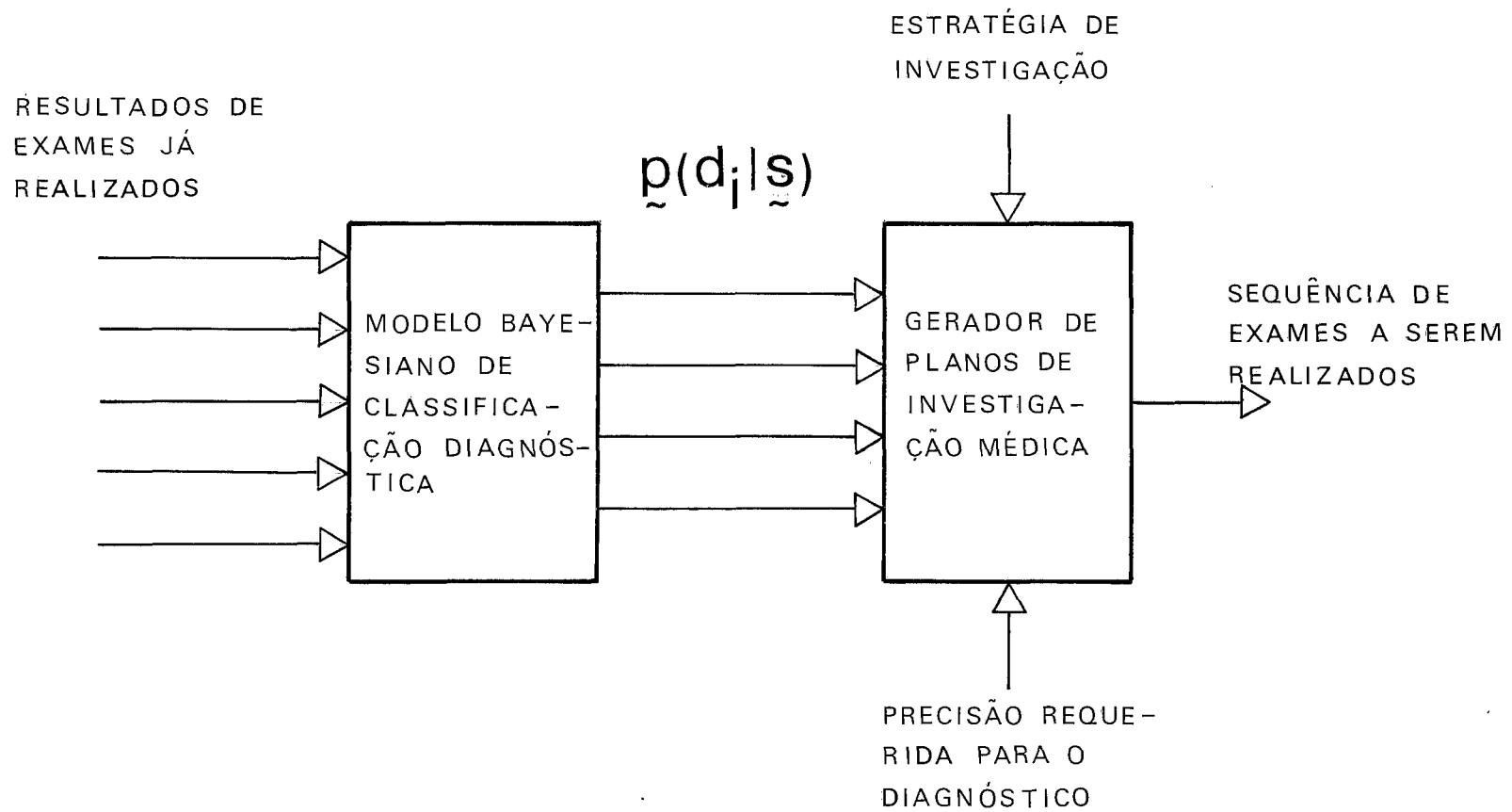


FIGURA IV.2 - ESTRUTURA DO SISTEMA DE GERAÇÃO DE PLANOS DE INVESTIGAÇÃO DIAGNÓSTICA



B - Um gerador de planos de investigação diagnóstica conforme exposto neste capítulo. O gerador de planos, com base nas probabilidades a posteriori das classes diagnósticas, na precisão requerida para o diagnóstico e na estratégia selecionada pelo médico, irá calcular uma seqüência ótima de exames adicionais dentre o conjunto de exames ainda não realizados no paciente.

#### IV.6 - NOTAÇÃO

Apresentamos a seguir a notação utilizada neste capítulo a qual deve ser somada a notação já apresentada na tabela (III.1).

- $X_i$  - Representação do estado do problema de investigação médica caracterizado pelos exames já realizados.
- $X_0$  - Conjunto de estados iniciais
- $x_0$  - Estado inicial  $\in X_0$
- $X_t$  - Conjunto de estados terminais ou desejados
- $x_t$  - Estado terminal  $\in X_t$
- $\pi_{j\alpha}$  - Probabilidade de o exame/descritor  $j$  revelar a doença  $\alpha$
- $q_{j\alpha}$  - Probabilidade que o exame  $j$  venha a ser realizado no processo de investigação dado que o paciente tem a doença  $\alpha$ .
- $c_j$  - Custo de realizar o exame  $j$

- $c_j$  - Custo de realizar o exame  $j$  dado que o paciente tem a doença  $\alpha$
- $r_j$  - Risco associado ao exame  $j$
- $u_{x/\alpha}$  - Probabilidade de que os exames (ou descritores) realizados até o estado  $x$  tenham revelado a doença  $\alpha$
- $\tilde{p}(d/x_0) = \{p(d_1/x_0), \dots, p(d_N/x_0)\}$  - Probabilidade a posteriori das diversas classes diagnósticas conhecidos os resultados dos exames realizados no estado inicial  $x_0$ .
- $F(x)$  - Operador de geração de sucessores usado na definição implícita do grafo de espaço de estado.
- $E = \{e_1, \dots, e_N\}$  - Conjunto de exames médicos.
- $e_i$  - Exame médico  $\in E$  corresponde a um conjunto de descritores.
- $p_{rev}(x)$  - Probabilidade de que alguma doença tenha sido revelada no estado  $x$ .

#### IV.7 - PRESSUPOSIÇÕES BÁSICAS

Relacionamos a seguir as pressuposições básicas referentes ao modelo de geração de planos de investigação exposto nesse capítulo:

11. A probabilidade de se obter o resultado de um teste é independente da seqüência de testes já realizados sobre o paciente.
12. O estado de saúde do paciente é estacionário durante o processo de investigação.
13. O tratamento do paciente é iniciado depois de uma doença ter sido diagnosticada.

14. Todos os exames permanecem disponíveis para execução por um período ilimitado de tempo.
15. O tempo total de processamento de cada exame (consistindo de seu tempo de execução e o tempo de espera do resultado) bem como os custos de sua realização são fixos.

Tendo em vista a fundação bayesiana utilizada devemos acrescentar a estas pressuposições as apresentadas na tabela (III.2).

#### IV.8 - PROPRIEDADES DO DIAGNÓSTICO BAYESIANO SEQÜENCIAL

As duas propriedades seguintes enunciadas por Peters [139] são fundamentais nos modelos de diagnóstico seqüencial nos quais nos enquadraremos.

Consideremos  $n$  observações  $s_1, s_2, \dots, s_n$ . Vamos denotar com  $R$  o conjunto de todas as permutações possíveis das  $n$  observações. Portanto  $R$  consiste de  $n!$  elementos  $R_1, R_2, \dots, R_n!$ , cada elemento representando uma diferente permutação de  $s_1, \dots, s_n$ .

##### Propriedade I

Se a pressuposição 11 é válida

$$p(d_i/R_j) = p(d_i/R_k) \quad \forall R_j, R_k \in R \quad (\text{IV.2})$$

$$i = 1, 2, \dots, N$$

ou seja as probabilidades a posteriori das doenças não dependem da ordem em que os descritores foram mensurados.

## Propriedade II

Se a pressuposição 11 é válida, a probabilidade de obter um resultado  $s_{jq} \notin \{s_1, \dots, s_n\}$  do descritor  $S_j$ , depois de ter observado os resultados  $s_1, \dots, s_n$ , não depende da ordem em que estes resultados foram obtidos.

$$p(s_{jq}/R_j) = p(s_{jq}/R_k) \quad \forall R_j, R_k \in R \quad (\text{IV.3})$$

## IV.9 - DESCRIÇÃO DO MÉTODO DE GERAÇÃO DE PLANOS DE INVESTIGAÇÃO MÉDICA

O método proposto baseia-se na visualização do problema da determinação de planos ótimos de investigação médica para um paciente como um problema de busca do caminho de custo mínimo em um grafo de espaço do estado.

A solução proposta está baseada no emprego de técnicas de programação heurística, ou mais especificamente, na aplicação do algoritmo  $A^*$  proposto por Hart e Nilsson [79]. Transcrevemos para benefício do leitor não familiarizado com o algoritmo  $A^*$  um sumário com os principais resultados, elaborado por Araújo [7].

No restante desta seção a solução proposta será desenvolvida com a apresentação e discussão de cada elemento requerido pela arquitetura da programação heurística.

### IV.9.1 - Programação Heurística - O Algoritmo $A^*$

#### Grafo do Problema

Os problemas considerados são aqueles para os quais nós conhecemos um conjunto  $X_0$  de estados iniciais, podemos identificar um conjunto  $X_t$  de estados desejados (estados solu

ção ou terminais), e dispomos de um conjunto (finito)  $T$  de operações sobre os estados, ou seja, se  $x$  for um estado do problema e  $\gamma \in T$  for uma operação válida sobre  $x$ , então  $\gamma x = \{y, c\}$  onde  $y$  será um estado do problema e  $c$  será o custo de aplicar  $\gamma$  a  $x$ . Sob estas condições, o grafo (colorido e orientado)  $G_{X_0} = \{X_0, T\}$ , definido por:

- i)  $X_0$  está contido no conjunto de nós de  $G_{X_0}$ , denotado como  $V(G_{X_0})$
- ii)  $n_i \in V(G_{X_0})$ ,  $\gamma \in T$ , e  $\gamma n_i = (n_j, c_{ij}) \rightarrow n_j \in V(G_{X_0})$  e  $c_{ij}$  representa o custo do arco  $n_i n_j$ , é chamado grafo gerado por  $X_0$  e  $T$ .

No caso (ii) acima  $n_j$  é chamado sucessor de  $n_i$ . A solução de um problema pode então ser vista como a exploração parcial de  $G_{X_0}$ .

### Admissibilidade

Consideraremos somente grafos com custos de arcos superiores a um número positivo; nos referiremos a eles como grafos- $\delta$ . Vamos também pressupor que os custos são aditivos, ou seja, se  $\eta = (n_0, n_1, \dots, n_k)$ , onde  $n_{i+1}$  é um sucessor de  $n_i$ , for um caminho de  $n_0$  a  $n_k$ , ele terá um custo

$$k_{\eta}(n_0, n_k) = \sum_{i=1}^k c_{i-1, i} \quad (\text{IV.4})$$

Nós denotamos por  $k(n_0, n_k)$  o custo mínimo entre todos os caminhos de  $n_0$  a  $n_k$ .

Algumas notações particularmente úteis são as seguintes:

Se  $n$  é um nó de  $G_{x_0}$  nós escrevemos

$$g(n) = \min_{x_0 \in X_0} k(x_0, n) \quad (\text{IV.5})$$

$$h(n) = \min_{x_t \in X_t} k(n, x_t) \quad (\text{IV.6})$$

$$h(X_0) = \min_{x_0 \in X_0} h(x_0) \quad (\text{IV.7})$$

O mínimo custo de um caminho de  $X_0$  a  $X_t$ , obrigado a passar através do nó  $n$ , é denotado como

$$f(n) = g(n) + h(n) \quad (\text{IV.8})$$

Uma **solução** para o problema é um caminho de  $X_0$  a  $X_t$ . Uma solução é dita **preferida** se seu custo for  $f(X_0)$ . Um algoritmo de busca é dito ser **admissível** se para cada grafo que possua uma solução finita, puder garantir encontrar a solução preferida num número finito de passos.

## Busca Heurística

Nós vamos usar o termo "busca heurística" para significar o uso de informação adicional sobre o problema para ajudar a guiar a busca. Uma boa discussão sobre este assunto pode ser achada em Nilsson [125,126]. Para os nossos objetivos, precisamos assumir somente que, para cada nó  $n$  já visitado ou sucessor de um nó já visitado, nós possamos calcular uma "função de mérito"  $\hat{f}(n)$ , a qual pode depender tanto de informação externa como do status da pesquisa. Para cada nó, armazenamos:

- i) Descrição do estado
- ii) Indicador de nó  $F$  (aberto) ou 1 (fechado)
- iii) Valor de  $\hat{f}$
- iv) Apontador  $P$  para seu melhor predecessor

(O significado do F e P ficará claro a seguir).

#### ALGORÍTMO A\*

1. Para todo  $x_0 \in X_0$ , calcular o armazenar  $\hat{f}(x_0)$  e fazer  $F(x_0) = 0$ ,  $P(x_0) = 0$ ;
2. Escolher o nó  $n$  aberto que tenha o melhor mérito. Resolva os empates do modo que preferir, porém sempre em favor de nós de  $X_t$ .
3. Se  $n \in X_t$  faça  $F(n) = 1$  e pare. A solução é obtida seguindo os apontadores para trás.
4. Em caso contrário, faça  $F(n) = 1$  e gere todos os sucessores de  $n$ ; para cada sucessor  $m$  calcule  $\hat{f}(m)$ . Se  $m$  for melhor que o antigo, faça  $P(m)=n$ ,  $F(m)=0$  e armazene  $\hat{f}(m)$ .
5. Vá para 2.

Neste algoritmo,  $\hat{f}(n)$  é usualmente uma estimativa de  $f(n)=g(n)+h(n)$ . Um importante caso especial é:

$$\hat{f}(n) = \hat{g}(n) + \hat{h}(n) \quad (\text{IV.9})$$

onde

$\hat{g}(n)$  é o mínimo custo, achado até agora, do caminho de  $X_0$  até  $n$ ;  $\hat{h}(n)$  é uma estimativa de  $h(n)$ . Neste caso temos o seguinte teorema:

#### Teorema (Hart et alii)

Se  $\hat{f}(n) = \hat{g}(n) + \hat{h}(n)$  e  $\hat{h}(n) \leq h(n)$  para todos os nós

de  $G_{X_0}$ , então  $A^*$  é admissível. Além disso, somente nós tendo  $\hat{f}(n) \leq f(X_0)$  serão fechados.

Daqui para frente usaremos sempre  $\hat{f}(n)$  nesta forma e chamaremos  $\hat{h}(n)$  de "função heurística".

Se além de ser um valor inferior a  $h(n)$ ,  $\hat{h}(n)$  também satisfizer:

$$\hat{h}(m) - \hat{h}(n) \leq k(m,n) \quad (\text{IV.10})$$

para qualquer  $m, n \in V(G_{X_0})$ , e nós terminais  $t \in T$  que tenham  $\hat{h}(t)=0$  nós dizemos que a heurística é **consistente**. Para o caso de heurísticas consistentes, Hart et al demonstraram que o algoritmo  $A^*$  faz uso ótimo da informação disponível, no sentido de que minimiza o número de nós visitados, ao mesmo tempo que garante a admissibilidade. Além disso, quanto mais próximo  $\hat{h}(n)$  for de  $h(n)$  melhor será o desempenho obtido pelo algoritmo.

#### IV.9.2 - Representação do Espaço de Estado no Problema de Investigação Diagnóstica

A investigação diagnóstica é um processo seqüencial de decisão caracterizado por etapas intercaladas de tomada de decisão e realização de exames até o estabelecimento do diagnóstico do paciente. Podemos, portanto, definir naturalmente como **estado do problema** de investigação diagnóstica, o conjunto  $x = \{e_1, e_2, \dots, e_i\}$  de exames realizados até o momento no paciente.

Tendo em vista as propriedades I e II enunciadas por Peters podemos afirmar que a ordem em que foram realizados os exames pertencentes ao conjunto  $x$  é irrelevante para a definição de estado do problema.

Para cada estado teremos um conjunto de operações possíveis que correspondem à realização de cada um dos exames do conjunto  $E-x$  de exames ainda não realizados sobre o paciente.



A operação  $\gamma_k$  (realização do exame  $e_k$ ) aplicada sobre um estado  $x = \{e_1, \dots, e_i\}$  o transforma no estado  $y = \{e_1, \dots, e_i, e_k\}$  com um custo  $c_k$ , o qual é definido pela estratégia selecionada pelo médico. Caso se trate de uma estratégia elementar, o custo  $c_k$  será exatamente o valor do atributo de custo do exame  $e_k$ , segundo a estratégia em questão. Caso se trate de uma estratégia composta, o custo  $c_k$  será obtido por uma combinação definida pelo usuário de atributos de custo do exame  $e_k$  de diversas estratégias.

Definido um conjunto de estados iniciais a aplicação sucessiva dos operadores gerará um espaço de estado do problema de investigação diagnóstica. Sem nos preocupamos ainda com a definição de estados terminais e considerando como estado inicial  $x_0 = \emptyset$  podemos afirmar que o grafo de espaço de estado terá  $2^M$

nós e  $\sum_{i=1}^M A_M^i$  seqüências possíveis de exames. Trata-se de um

um problema combinatório que assume rapidamente grandes proporções com o crescimento do número de exames. Para melhor apreciação apresentamos a seguir as dimensões do problema para conjunto de  $M = 10$  e  $M = 15$  exames.

Nº de Exames	Nº Máximo de Estados	Nº Máximo de Seqüências de Exames
10	1.024	6.235.300
15	32.768	$2,246 \cdot 10^{12}$

A representação explícita do grafo de espaço de estado para o caso de 4 exames é ilustrada na figura (IV.3).

A representação implícita do grafo de espaço de estado é dada pelos seguintes elementos:

- Descrição do estado do problema.

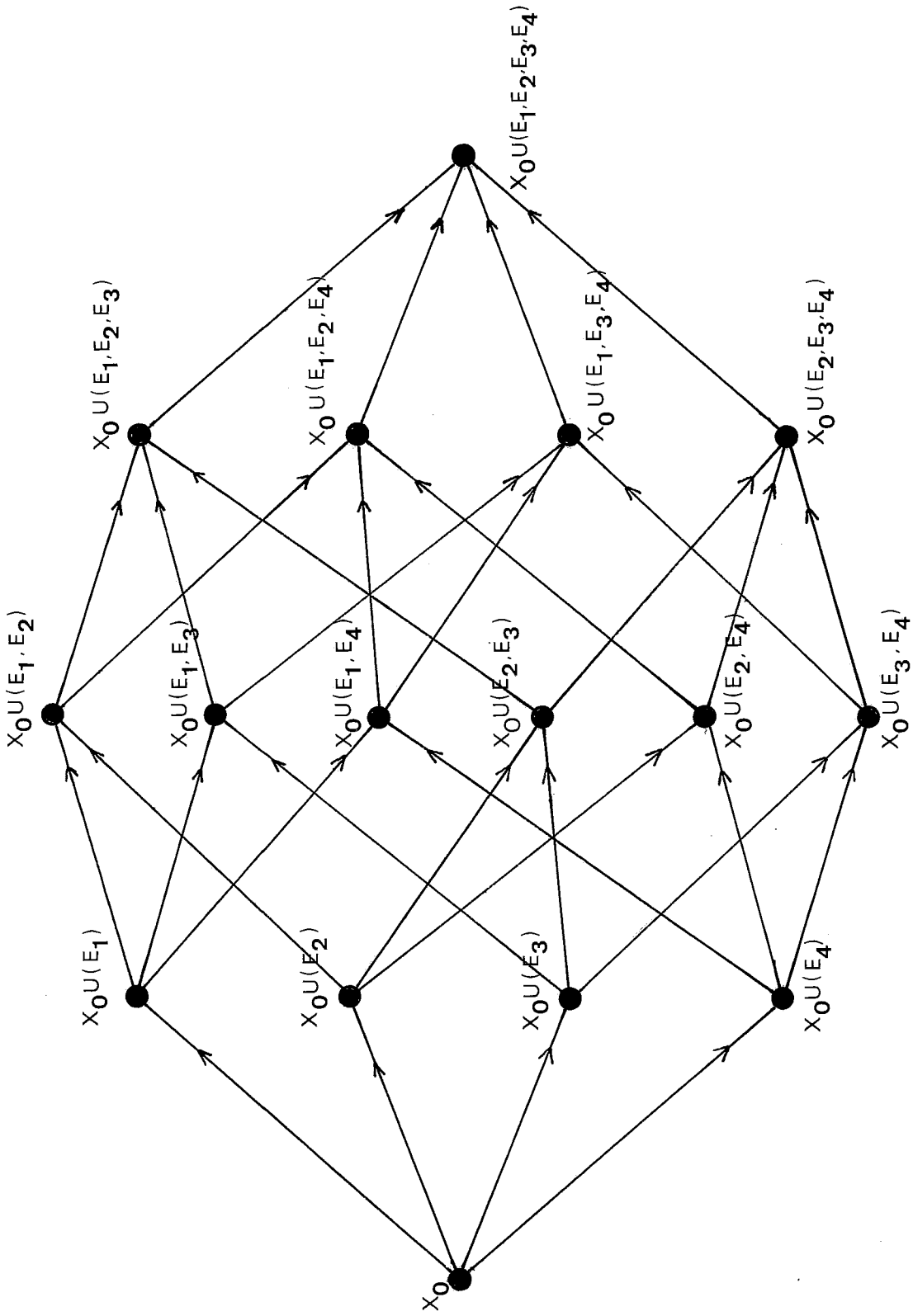


FIGURA IV.3 - GRAFO DE ESPAÇO DE ESTADO PARA O PROBLEMA DE INVESTIGAÇÃO DIAGNÓSTICA COM 4 EXAMES

- Descrição dos estados iniciais
- Operador  $T$  de geração de estados sucessores
- Propriedade que descreve os estados terminais (ou estados desejados)

que serão detalhados nas próximas seções.

#### IV.9.3 - Descrição do Estado do Problema

A descrição do estado do problema de investigação diagnóstica, conforme já dissemos, é dada pelo conjunto de exames realizados até o momento considerado, sobre o paciente. Um aspecto importante a ser considerado é o fato que freqüentemente os exames são procedimentos que envolvem a mensuração de mais de um descritor. Denotamos isto através de

$$e = \{S_1, S_2, \dots, S_i\}$$

Assim é o caso da história clínica, do exame físico, de baterias de testes laboratoriais, do eletrocardiograma, etc. No eletrocardiograma, por exemplo, podemos sumarizar as informações fornecidas pelo exame através dos seguintes descritores dicotômicos:

Isquemia inferior  
Isquemia lateral  
Isquemia anterior  
Outras isquemias  
Necrose inferior  
Necrose lateral  
Necrose anterior  
Outras necroses  
Hipertrofia ventricular  
Arritmias  
Alterações inespecíficas da repolarização ventricular  
Bloqueios

Portanto, o projetista do sistema se defrontará com duas alternativas para definição da descrição do estado, a saber:

1. Descrição baseada nos descritores mensurados
2. Descrição baseada nos exames realizados.

A 1ª alternativa oferece algumas dificuldades tais como:

- Grande crescimento da complexidade do problema de busca no grafo de espaço de estado.
- Artificialismo dos planos de investigação produzidos pelo sistema que freqüentemente pedirão a mensuração de apenas um descritor, que normalmente seria colhido com vários outros dentro do mesmo procedimento de exame.
- Dificuldade na distribuição do custo de realização do exame entre os descritores associados. Devemos alocar todo o custo do exame ao descritor solicitado, ou proporcionalmente ao número total de descritores?

A descrição de estado baseado em exames realizados não apresenta os problemas citados anteriormente. Envolve a agregação das informações fornecidas pelos descritores individuais, o que pode ser feito facilmente dentro da formulação bayesiana.

### REPRESENTAÇÃO COMPUTACIONAL DA DESCRIÇÃO DE ESTADO

A representação computacional da descrição de estado pode ser feita alternativamente através de:

Um vetor binário indicando os exames ou descritores já mensurados

Ex: 1 0 1 1 0 1 0 0 0 0

Uma lista contendo os exames ou descritores já mensurados

Ex:  $e_1 e_2 e_4 e_6$

#### IV.9.4 - Descrição do Estado Inicial

No nosso modelo teremos apenas um estado inicial  $x_0$  no conjunto  $X_0$ , o qual espelhará os exames realizados sobre o paciente antes da aplicação do gerador de planos de investigação médica. Tipicamente o estado  $x_0$  englobará dados de identificação do paciente e os exames de história clínica e opcionalmente o exame físico. Com base nas observações realizadas nos exames do estado inicial  $x_0$ , calcularemos as probabilidades  $p(d/x_0)$  que serão usadas pelo gerador de planos de investigação como um perfil do estado de saúde do paciente.

#### IV.9.5 - Operador $T$ de Geração de Estados Sucessores

O operador  $T$  é um operador que aplicado a um estado  $x$  gera todos os seus sucessores. Na sua acepção mais simples o operador  $T$  terá que executar apenas o trabalho de oferecer como resultado o conjunto

$$T(x) = E - x \quad (IV.11)$$

$E$  = conjunto universo de exames.

Em versões mais elaboradas o operador  $T$  pode ser projetado de forma a incorporar características importantes do modelo de geração de planos de investigação médica, bem como implementar heurísticas destinadas à realização de poda horizontal no grafo de espaço de estado, conforme descrito a seguir.

Nessas versões teremos:  $T(x) \subseteq E - x$

#### IV.9.5.1 - Características Adicionais do Modelo Incorporadas ao Operador $T$

Através do registro das datas de realização de exames, agendas de disponibilidade e características de interação entre exames, podemos adequar o operador  $T$  à complexa realidade da rotina de trabalho em um centro médico.

Entre as características adicionais que podem ser incorporadas ao modelo através de modificações na construção do operador  $T$ , citamos:

- Inclusão de interações entre exames
  - Incompatibilidade entre exames
  - Latências (exames que impedem a realização de outros exames por determinado tempo)
  - Pré-requisitos (exames que exigem a realização de outros como pré-requisitos)
- Contra indicações de exames em função do estado de saúde do paciente.
- Prazo de validade de exames
- Restrição quanto à duração máxima da investigação
- Disponibilidade efetiva de exames na instituição.

#### IV.9.5.2 - Heurísticas para Limitação na Geração de Sucessores

A incorporação de heurísticas no projeto permite obter um operador  $T$  mais informado capaz de expandir apenas os estados sucessores mais promissores. Isto propicia um aumento considerável da eficiência do sistema, porém sem garantia de obtenção da solução ótima.

Para tanto devemos definir um critério heurístico  $\lambda_j$  que permita ao operador  $T$  medir o valor informativo de um exame  $e_j$  ou descritor  $S_j$ . O critério heurístico  $\lambda_j$  será aplicado uma única vez no estado  $x_0$  sobre os descritores não mensurados. Estes descritores serão ordenados em uma lista

$$S_{r(1)}, S_{r(2)}, \dots, S_{r(k)}$$

de tal forma que

$$\lambda_{r(1)} \geq \lambda_{r(2)} \geq \dots \geq \lambda_{r(k)}$$

Duas estratégias podem ser adotadas para escolha dos descritores mais promissores:

1. Escolher apenas os  $q$  primeiros descritores da lista
2. Determinar o maior  $q^*$  de tal forma que  $\lambda_{r(q^*)} \geq \beta$  ( $\beta = \underline{\text{limiar especificado}}$ ). Então escolher apenas os  $q^*$  primeiros descritores da lista.

#### CRITÉRIOS HEURÍSTICOS PARA $\lambda_j$

##### A - Proposta de Peters [139]

Definamos um subconjunto  $\Delta_j$  do conjunto  $\Delta$  de doenças,

tal que para todo elemento  $d \in \Delta_j$  exista pelo menos um dos resultados possíveis  $s_{j1}, s_{j2}, \dots, s_{j,n(j)}$  do descritor  $S_j$  que apresente  $p(s_{jr}/d) > \gamma$  onde  $\gamma =$  limiar pré-definido. Se  $\gamma$  for próximo a 1,  $\Delta_j$  consistirá geralmente de poucos elementos pertencentes a  $\Delta_j$ . A probabilidade condicional que  $\Delta_j$  seja verdadeiro, dada a informação capturada no estado inicial  $x_0$  é dada por

$$p(\Delta_j/x_0) = \sum_{i \in I_j} p(d_i/x_0) \quad (\text{IV.12})$$

sendo utilizada como critério  $\lambda_j$ .

## B - Teoria da Informação

Gleser [68] e Taylor [174] usaram como critério heurístico  $\lambda_j$  o ganho esperado de informação propiciado pelo descritor  $S_j$ . A função de entropia de Shannon mede a incerteza de uma decisão e pode ser calculada para o nó  $x_0$  através de

$$H(x_0) = - \sum_i p(d_i/x_0) \log p(d_i/x_0) \quad (\text{IV.13})$$

O ganho esperado de informação é dado pela diferença da entropia no estado  $x_0$  e o valor esperado da entropia dado que o descritor  $S_j$  foi mensurado.

$$\lambda_j = \Delta H = H(x_0) - \sum_{q=1}^{n(j)} p(s_{jq}/x_0) H(x_q) \quad (\text{IV.14})$$

Uma modificação empregada por Taylor [179] é o ganho esperado de informação por unidade de custo

$$j = \frac{\Delta H}{c_j}$$



onde  $c_j$  = custo de  $S_j$ .

### C - Mudança Cumulativa Esperada de Probabilidades

Rector [144] propôs como critério heurístico  $\lambda_j$ , a mudança cumulativa esperada de probabilidades decorrentes da mensuração do descritor  $S_j$ . A mudança cumulativa esperada de probabilidades decorrentes do achado  $s_{jq}$  é dada por

$$\delta_{jq} = \sum \left| p(d_i/x_o) - p(d_i/x_q) \right| \quad (\text{IV.15})$$

A mudança esperada cumulativa de probabilidades para o descritor  $S_j$  é

$$\lambda_j = \sum_{q=1}^{n(j)} p(s_{jq}/x_o) \cdot \delta_{jq} \quad (\text{IV.16})$$

#### IV.9.6 - Definição dos Estados Terminais

O conjunto  $X_t$  de estados terminais ou estados desejados é definido através de uma propriedade relacionada ao objetivo final do sistema, qual seja, o estabelecimento de um diagnóstico verdadeiro. Definimos como estado terminal todo estado no qual a probabilidade de que alguma doença tenha sido revelada, seja maior que um valor limiar  $p_*$ , fornecido pelo médico usuário como sendo o nível de certeza mínimo requerido para o diagnóstico.

Apresentamos a seguir duas propostas alternativas para o cálculo da probabilidade de revelação de alguma doença no estado  $x$  denotada como  $p_{rev}(x)$ .

#### IV.9.6.1 - Proposta de Alperovitch

Esta primeira proposta apresentada por Alperovitch [3] é a mais eficiente do ponto de vista computacional dentre as duas que apresentamos. Baseia-se no emprego de um novo parâmetro  $\prod_{j\alpha}$  definido como a probabilidade de que o exame  $j$  (ou o descritor  $j$ , dependendo do caso) revele a doença  $\alpha$ . A probabilidade de que alguma doença tenha sido revelada no estado  $x$  é calculada através de

$$p_{\text{rev}}(x) = \sum p(d_{\alpha}/x_0) \cdot u_{x/\alpha} \quad (\text{IV.17})$$

onde

$u_{x/\alpha}$  = probabilidade que os exames (ou descritores se for o caso) realizados até o estado  $x$  tenham revelado a doença  $\alpha$

$$u_{x/\alpha} = 1 - \prod_{j=1}^i (1 - \pi_{j\alpha}) \quad (\text{IV.18})$$

#### ESTIMAÇÃO DO PARÂMETRO $\prod_{j\alpha}$

##### A - Caso de Descrições de Estado Baseadas em Descritores Mensurados

O parâmetro  $\pi_{j\alpha}$  corresponde nessa situação à sensibilidade do descritor  $S_j$  para a doença  $d_{\alpha}$  e sua estimação pode ser feita a partir do banco de dados de treinamento através de:

$$\pi_{j\alpha} = \frac{\text{Nº de pacientes com } d_{\alpha} \text{ que tiveram resultado positivo na } \alpha \text{ mensuração do descritor } S_j}{\text{Nº total de pacientes com } d_{\alpha}}$$

(IV.19)

Naturalmente, o conceito de resultado positivo na mensuração de  $S_j$  variará para cada doença  $d_{\alpha}$ .

#### B - Caso de Descrições de Estados Baseados em Exames Realizados

Esta situação é um pouco mais complexa havendo a necessidade de agregar a informação dos vários descritores pertencentes a cada exame. Isto pode ser feito através da criação de modelos bayesianos de classificação diagnóstica usando apenas os descritores de um exame específico  $e_j$  conforme mostra a figura (IV.4).

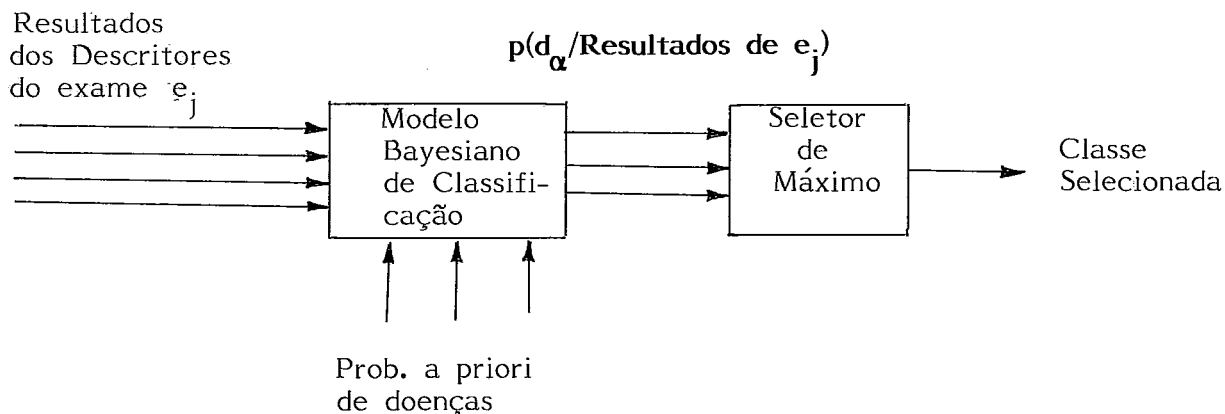


Figura IV.4 - Esquema para estimativa de  $\pi_{j\alpha}$  no caso de descrições de estados baseados em exames realizados.

As probabilidades a priori podem ser as da população ou probabilidades uniformes (modelo de verossimilhança relativa). Os casos do banco de dados com a doença  $d_\alpha$  nos quais foi realizado o exame  $e_j$  são analisados pelo modelo de classificação, sendo a classe selecionada registrada juntamente com a classe real de cada caso. O parâmetro  $\pi_{j\alpha}$  pode então ser estimado através da expressão:

$$\pi_{j\alpha} = \frac{\text{Nº de pacientes com } d_\alpha \text{ que foram diagnosticados corretamente pelo modelo}}{\text{Nº Total de pacientes com a doença } d_\alpha}$$

(IV.20)

#### IV.9.6.2 - Proposta Baseada no Banco de Dados

A primeira proposta produz valores  $p_{rev}(x)$  que se aproximam muito rapidamente de 1 na medida que o estado  $x$  abranje maior número de exames realizados, o que não corresponde muito bem à realidade.

Uma maneira alternativa que permite calcular valores mais realistas de  $p_{rev}(x)$  está baseada no uso direto do banco de dados, sendo seu algoritmo o seguinte.

#### ALGORITMO DE DETERMINAÇÃO DE $p_{rev}(x)$

1. Criar um modelo bayesiano de classificação diagnóstica com pressuposição de independência, envolvendo apenas os descritores pertencentes aos exames do estado  $x$  em consideração.

2. Aplicar o modelo a um conjunto de casos de um banco de dados, registrando a classe diagnóstica selecionada pelo modelo e a classe real de cada caso.
3. Estimar diretamente a partir dos resultados do passo 2 o valor de  $u_{x/\alpha}$

$$u_{x/\alpha} = \frac{\text{Número de casos com a doença } \alpha \text{ diagnosticados corretamente pelo modelo}}{\text{Número de casos com a doença } \alpha \text{ no banco de dados}}$$

4. Calcular  $p_{rev}(x)$  através de

$$p_{rev}(x) = \sum_{\alpha=1}^N p(d_{\alpha}/x_0) \cdot u_{x/\alpha}$$

Esta 2ª proposta de cálculo de  $p_{rev}(x)$  é bastante custosa em termos computacionais. Assim sendo, só a recomendamos para situações nas quais sejam usadas descrições de estados baseadas em exames realizados no paciente e nas quais o número de exames seja pequeno.

#### IV.10 - CONSTRUÇÃO DA FUNÇÃO DE MÉRITO PARA O ALGORÍTMO A\*

Conforme vimos na seção IV.9.1 o algoritmo A\* utiliza uma função de mérito para ordenar os estados mais promissores para expansão, que possui a seguinte forma:

$$\hat{f}(x) = \hat{g}(x) + \hat{h}(x)$$

A função  $\hat{g}(x)$  é definida como a estimativa do custo do caminho ótimo do estado  $x_0$  até o estado  $x$  e é expresso por:

$$\hat{g}(x) = \sum_{j \in x-x_0} c_j \quad (\text{IV.21})$$

No caso de estratégias de mínimo risco de morbidade ou mortalidade, assumindo-se que os danos ocasionados ao paciente pelos diversos exames sejam estatisticamente independentes, a função  $\hat{g}(x)$  será expressa por

$$\hat{g}(x) = 1 - \prod_{j \in x-x_0} (1-r_j) \quad (\text{IV.22})$$

onde  $r_j$  = risco de morbidade ou mortalidade associado ao exame  $e_j$ .

A função  $\hat{h}(x)$  ou função heurística é a que permite fornecer informação externa sobre o problema para o algoritmo.  $\hat{h}(x)$  representa uma estimativa do custo do caminho ótimo do estado  $x$  ao conjunto de estados terminais. Apresentamos a seguir nossa proposta de função heurística para o problema da investigação diagnóstica.

#### IV.10.1 - Função Heurística para o Problema de Investigação Diagnóstica

Partimos do princípio de utilizar a experiência passa-

da sobre o emprego dos diversos exames no processo de investigação diagnóstica como informação heurística para o algoritmo. Para tanto definimos um novo parâmetro  $q_{j\alpha}$  capaz de expressar para cada doença quais os exames mais promissores.  $q_{j\alpha}$  é definido como a probabilidade de o exame  $e_j$  ser realizado, dado que o paciente tem a doença  $\alpha$ .  $q_{j\alpha}$  pode ser estimado a partir da experiência passada de investigação diagnóstica em pacientes, armazenada num banco de dados, ou obtida através de probabilidades subjetivas fornecidas por especialistas médicos. A função heurística proposta para atributos que não sejam do tipo risco é definida por:

$$\hat{h}(x) = \sum_{j \in E-x} p(e_j) \cdot c_j \quad (\text{IV.23})$$

onde

$p(e_j)$  = probabilidade de que o exame  $e_j$  venha a ser realizado nas etapas restantes da investigação a partir do estado  $x$ .

$$p(e_j) = \sum_{\alpha=1}^N p(d_\alpha/x_0) \cdot q_{j\alpha} \quad (\text{IV.24})$$

Pela apreciação da condição de admissibilidade do algoritmo  $A^*$  para o nosso caso

$$\sum_{j \in E-x} p(e_j) \cdot c_j \leq h(x) \quad (\text{IV.25})$$

podemos concluir que face à estrutura probabilística usada não

podemos garantir a admissibilidade com certeza. Podemos, porém diminuir tanto quanto queiramos o risco da condição (IV.25) não ser verdadeira, utilizando um banco de dados com casos reais não viciados e, por outro lado, usando estimativas  $q_{j\alpha}$  mais conservadoras.

Para as estratégias de mínimo risco (morbidade ou mortalidade) a função heurística poderá ser expressa por:

$$\hat{h}(x) = 1 - \prod_{j \in E-x} [1 - p(e_j)r_j] \quad (\text{IV.26})$$

Outra expressão alternativa para a função heurística em estratégias de mínimo risco é:

$$\hat{h}(x) = 1 - \prod_{j \in E-x} (1 - r_j)^{p(e_j)} \quad (\text{IV.27})$$

### CONSISTÊNCIA DA FUNÇÃO HEURÍSTICA PROPOSTA

Para estudar a consistência da função heurística proposta em (IV.23), (IV.36) e (IV.27) convém apreciar a seguinte propriedade do problema de investigação médica.

"Se um nó  $n$  do grafo de espaço de estado for descendente de um nó  $m$ , então o conjunto de exames realizados no estado  $m$  está contido no conjunto de exames realizados no estado  $n$ ".

### CONSISTÊNCIA PARA ESTRATÉGIAS QUE NÃO ENVOLVAM RISCO - EQ. (IV.23)

Sejam dois estados  $m$  e  $n$  para os quais  $k(m,n)$  exista. Pela propriedade apresentada anteriormente, podemos afirmar que o caminho de  $m$  a  $n$  compreende o conjunto de exames  $\epsilon = n-m$  com



$$\text{custo } k(m,n) = \sum_{j \in \epsilon} c_j$$

$$\hat{h}(m) - \hat{h}(n) = \sum_{j \in E-m} p(e_j) c_j - \sum_{j \in E-n} p(e_j) c_j$$

$$\hat{h}(m) - \hat{h}(n) = \sum_{j \in \epsilon} p(e_j) c_j \leq \sum_{j \in \epsilon} c_j$$

Portanto  $\hat{h}(m) - \hat{h}(n) \leq k(m,n)$  e a função heurística proposta satisfaz a desigualdade (IV.10); porém, não é consistente visto que um nó terminal  $n$  pode ter  $\hat{h}(n) > 0$ . No entanto, a ordenação que ela induz vai no mesmo sentido, e é de se esperar boa eficiência de busca.

#### CONSISTÊNCIA PARA ESTRATÉGIAS DE MÍNIMO RISCO - EQ.(IV.26)

Sejam dois estados  $m$  e  $n$  quaisquer para os quais exista um caminho de  $m$  para  $n$ . Vamos comparar a estimativa  $\hat{h}(m)$  de risco do caminho ótimo do nó  $m$  a  $X_t$  com a estimativa  $h^*$  do caminho ótimo de  $m$  a  $X_t$  passando pelo nó  $n$ .

Podemos expressar:

$$\hat{h}(m) = 1 - \prod_{j \in E-m} [1 - p(e_j) \cdot r_j]$$

$$k(m,n) = 1 - \prod_{j \in n-m} (1 - r_j)$$

$$\hat{h}(n) = 1 - \prod_{j \in E-n} [1 - p(e_j) \cdot r_j]$$

$$h^* = 1 - [1 - \hat{h}(n)] \cdot [1 - k(m,n)]$$

$$h^* = 1 - \prod_{j \in E-n} [1 - p(e_j)r_j] \cdot \prod_{j \in n-m} (1-r_j)$$

Desdobrando  $\hat{h}(m)$  em

$$\hat{h}(m) = 1 - \prod_{j \in E-n} [1 - p(e_j) r_j] \cdot \prod_{j \in n-m} [1-p(e_j) r_j]$$

e comparando com  $h^*$  conclui-se que  $\hat{h}(m) \leq h^*$ . Portanto, a função heurística proposta em (IV.26) para estratégias de mínimo risco satisfaz a desigualdade (IV.10). Através do mesmo raciocínio, podemos provar o mesmo para (IV.27). No entanto, não é constante porque nós terminais podem ter  $\hat{h}(x) > 0$ . No entanto, a ordenação que ela induz vai no mesmo sentido, e é de se esperar boa eficiência de busca.

#### IV.11 - CARÁTER ITERATIVO DO GERADOR DE PLANOS DE INVESTIGAÇÃO MÉDICA

Desejamos ressaltar o caráter iterativo do sistema de geração de planos de investigação diagnóstica proposto, que sugere o seu emprego de forma "on line" em centros médicos.

O sistema, alimentado com os dados iniciais da anamnese do paciente, calculará as probabilidades das diversas classes diagnósticas e elaborará um plano de investigação, de acordo com a estratégia selecionada e o nível de precisão requerido pelo médico. Com base no plano sugerido pelo sistema, o médico selecionará exames para realização, cujos resultados poderão ser também introduzidos no sistema. Repete-se o ciclo com uma

reavaliação das probabilidades das classes diagnósticas e, caso o médico assim o decida, com a geração de um novo plano de investigação segundo a estratégia anterior ou outra. A interrupção do processo é feita quando o médico tiver estabelecido um diagnóstico para iniciar o tratamento do paciente. A figura (IV.5) ilustra este processo.

#### IV.12 - CARACTERÍSTICAS DO SISTEMA

Relacionamos a seguir as principais características do sistema proposto para que o leitor possa apreciar melhor o seu poder de abrangência.

##### **Características:**

- Determinação eficiente de planos ótimos de investigação para fins de diagnóstico diferencial com base em técnicas de programação heurística.
- Planos de investigação sob forma de seqüências de exames, de fácil interpretação e utilização pelos médicos.
- Caráter iterativo com possibilidade de uso "on line".
- Permite ao médico selecionar a estratégia desejada e o nível mínimo de precisão requerido para o diagnóstico.
- Estratégias
  - elementares
  - compostas
  - robusta
- Múltiplos atributos de custo associados aos exames
  - Custo financeiro
  - Morbidade
  - Mortalidade

# MÉDICO

# SISTEMA

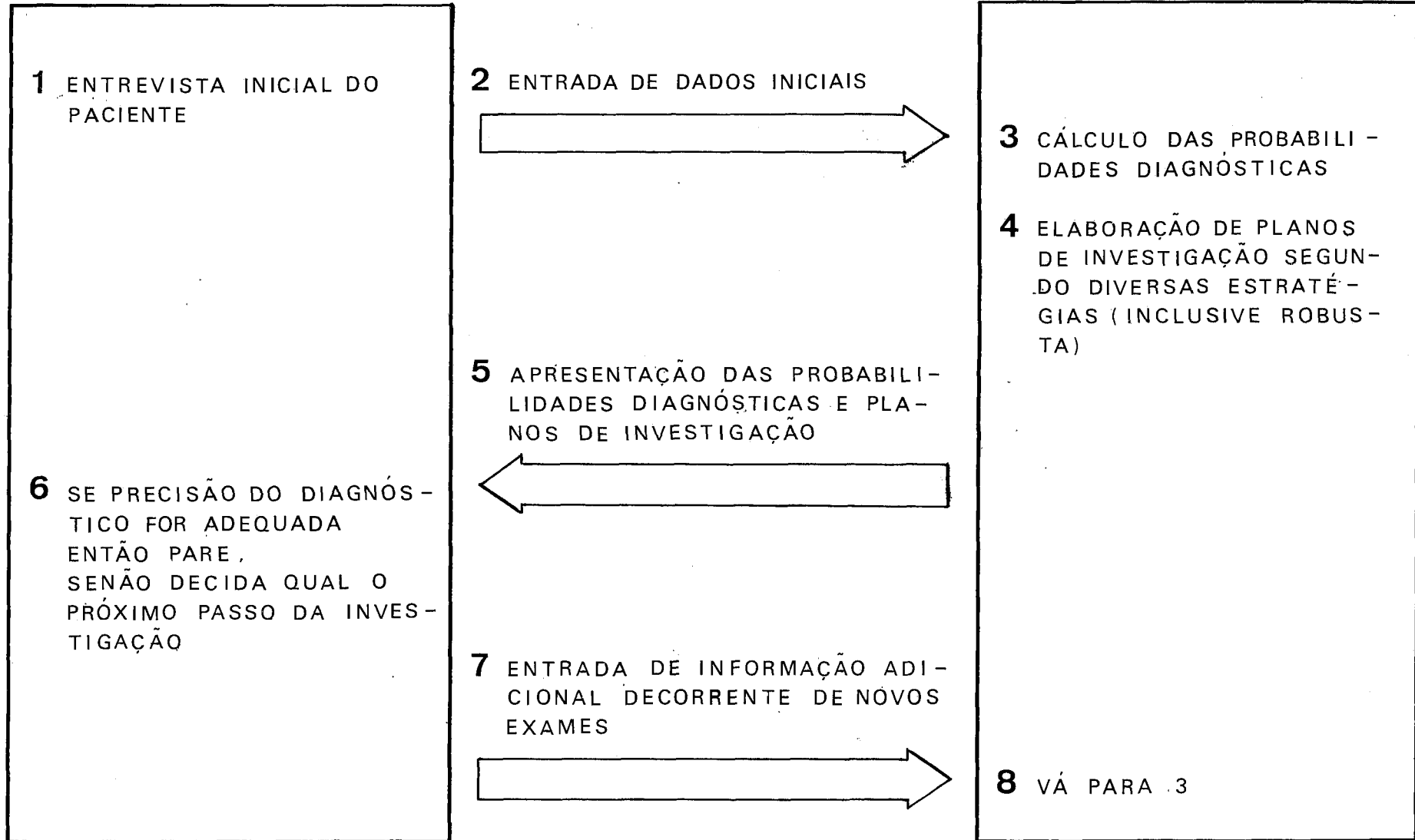


FIGURA IV.5

- Tempo de hospitalização
  - Tempo para obtenção de resultados
  - Sofrimento induzido pelo exame
- 
- Descrições dos estados do problema em termos de descritores mensurados ou exames realizados.
  - Emprego ótimo da informação heurística sobre os exames mais promissores para cada doença.
  - Função heurística consistente.
  - Possibilidade de poda horizontal no grafo de espaço de estado.
  - Alta flexibilidade devido ao embasamento bayesiano.
  - Bancos de dados com
    - Informações epidemiológicas sobre prevalência de doenças.
    - Informação médica sobre
      - relacionamento entre doenças e sintomas
      - interações entre sintomas
      - relações entre exames e descritores
    - Dados de pacientes
  - Calibração do modelo bayesiano de classificação e estimação de parâmetros do gerador de planos de investigação, a partir do banco de dados de pacientes.
  - Auto-aprendizagem através de calibrações periódicas dos modelos.
  - Interações logísticas entre exames
    - Incompatibilidades
    - Latências
    - Pré-requisitos

- Contra indicações de exames.
- Prazo de validade da informação de exames.
- Restrições quanto a duração máxima da investigação.
- Disponibilidade efetiva dos exames no centro médico.

#### IV.13 - APLICAÇÕES DO GERADOR DE PLANOS DE INVESTIGAÇÃO

Dentre as principais aplicações do modelo discutido neste capítulo, citamos:

- Emprego do sistema na prática médica como sistema de apoio à decisão no planejamento da investigação diagnóstica.
- Modelagem do processo de investigação diagnóstica.
- Utilização no ensino médico.
- Análise de restrições/condicionantes impostos em serviços de assistência médica.
- Utilização em administração hospitalar.

O sistema é especialmente indicado para áreas da medicina de grande complexidade no diagnóstico e onde os exames variem amplamente em termos de riscos e desconforto para o paciente, custos e durações.

As áreas médicas recomendadas pelo Dr. B. McA.Sayers do Imperial College para utilização deste modelo são:

- Doenças hepáticas
- Doenças renais

- Doenças metabólicas
- Doenças congênitas
- Neoplasias
- Doenças cardiovasculares
- Doenças endócrinas
- Efeitos tóxicos
- Defeitos em variáveis bioquímicas
- Estudos com traçadores.

---

---

## **V - DESCRIÇÃO DOS EXPERIMENTOS**

---

---



## V.1 - INTRODUÇÃO

O trabalho experimental realizado teve como objetivo explorar as técnicas de apoio à decisão médica apresentadas nos capítulos III e IV. Escolhemos como área médica para o estudo, o diagnóstico diferencial em doença aterosclerótica coronariana. Esta doença, além de sua elevada importância intrínseca como uma das principais doenças fatais que aflige a humanidade, nos ofereceu a oportunidade de lidar com:

- um espectro de variáveis diagnósticas que abrange desde situações relativamente simples como a diferenciação entre ausência x presença de lesões significativas, até situações muito complexas como a diferenciação entre configurações de lesões nos vasos coronarianos.
- variáveis prognósticas tais como tipo de tratamento e tempo de sobrevivência (esta última não implementada).
- um número elevado de descritores envolvendo fatores de risco e sintomas da doença.
- uma gama razoavelmente grande de procedimentos de exame que apresentam diferenças amplas em atributos como custo financeiro, morbidade, mortalidade, sofrimento imposto ao paciente, etc.
- o diagnóstico exato do paciente quanto à anatomia, através dos resultados da cineangiocoronariografia.

Como base para a pesquisa realizada, definimos e construímos um banco de dados de pacientes com suspeita de doença coronariana e implementamos em computador os modelos discutidos nos capítulos III e IV. A descrição deste trabalho básico e dos experimentos realizados é o objetivo deste capítulo. As

técnicas do 3º capítulo puderam ser aplicadas e exploradas em profundidade. Restrições de prazo para o término do estudo de tese limitaram o trabalho com geradores de planos de investigação diagnóstica ao delineamento de dois modelos preliminares. Ao longo de todo o trabalho experimental contamos com a orientação médica do Dr. Rogério Brant M. Chaves, Cardiologista do IASERJ.

## V.2 - COLETA DE DADOS

O banco de dados de pacientes foi construído usando-se informações de pacientes com suspeita de doença coronariana que realizaram o exame de cateterismo cardíaco (cineangiocoronariografia) no Hospital Central do Instituto de Assistência aos Servidores do Estado do Rio de Janeiro, uma entidade autárquica que tem por objetivos prestar assistência médica, serviços suplementares de saúde e serviço social aos funcionários, dependentes e aposentados do Estado e do Município do Rio de Janeiro. Para acelerar o processo optamos por uma abordagem retrospectiva de coleta de dados, o que permitiu aproveitar as informações de pacientes atendidos a partir de 1971, registradas em prontuários médicos e laudos de exames arquivados nos serviços de Documentação Médica e de Cardiologia.

### ESPECIFICAÇÃO DA POPULAÇÃO ESTUDADA

Foram selecionados apenas pacientes que realizaram o exame de cineangiocoronariografia, pois este é o exame capaz de fornecer a classificação diagnóstica do paciente quanto a situação de anatomia coronariana e função ventricular, o que é fundamental para a calibração de modelos bayesianos de classificação diagnóstica.

Foram colhidos 142 casos que apresentaram a seguinte

distribuição:

A - Quanto à presença de lesões ateroscleróticas coronarianas significativas

- 31 casos sem lesões significativas
- 111 casos com lesões significativas.

B - Quanto à configuração de lesões nos principais vasos coronarianos

- 31 Sem lesões significativas
- 12 Com lesões de Coronária Direita isolada
- 19 Com lesões de Descendente Anterior isolada
- 6 Com lesões de Coronária Circunflexa isolada
- 14 Com lesões de Coronária Direita e Descendente Anterior
- 10 Com lesões de Coronária Direita e Circunflexa
- 21 Com lesões de Descendente Anterior e Circunflexa
- 29 Com lesões nos 3 vasos.

C - Quanto à Função Ventricular

- 75 Com função ventricular normal
- 67 Com função ventricular alterada.

D - Quanto à presença de lesões e função ventricular

- 30 Sem lesões significativas e com função ventricular normal
- 1 Sem lesões significativas e com função ventricular alterada
- 45 Com lesões significativas e com função ventricular normal
- 66 Com lesões significativas e com função ventricular alterada.

E - Quanto ao tipo de tratamento

- 97 Tratamento clínico
- 21 Tratamento cirúrgico
- 24 Não informado

O banco de dados construído representa uma amostra de uma população de pacientes com forte suspeita de terem doença coronariana (suspeita tão forte que o exame de cateterismo chegou a ser realizado). Portanto devemos ter o cuidado de extrapolar as conclusões e inferências realizadas a partir deste banco de dados apenas a pacientes que se enquadrem nesta população.

#### FORMULÁRIO DE COLETA DE DADOS

A figura (V.1) apresenta uma cópia do formulário utilizado na coleta de dados dos pacientes selecionados para integrarem o banco de dados. O formulário foi projetado pelo cardiologista Dr. Rogério Brant M. Chaves e validado pela equipe do Serviço de Cardiologia do IASERJ. Para reduzir o seu tamanho foram selecionados apenas dados médicos de reconhecido valor diagnóstico tendo se procurado sintetizar os resultados de exames em poucas variáveis em geral de tipo discreto.

Conforme podemos observar, o formulário está dividido em seções de identificação de paciente, dados de história e exame físico, 7 seções de exames complementares e uma seção de tratamento. Os dados constantes do formulário podem ser classificados em:

- Dados de cunho administrativo
- Dados médicos  $\left\{ \begin{array}{l} \text{Descritores} \\ \text{Variáveis diagnósticas e prognósticas} \end{array} \right.$

O exame de cateterismo cardíaco nos oferece as duas



# \* — EXAMES COMPLEMENTARES —

## LAB.

NUM.:

(25) DT. PEDIDO...:  DIA - MÊS

(26) DT. EXEC.:  DIA - MÊS

(27) COLESTEROL

MG%

(28) TRIGLICERÍDEOS

MG%

(29) GLICEMIA

MG%

(30) ÁCIDO ÚRICO

MG%

## ECG

NUM.:

(31) DT. PEDIDO...:  DIA - MÊS

(32) DT. EXEC.:  DIA - MÊS

— RESULTADO —

## RX DE TÓRAX

NUM.:

(33) DT. PEDIDO...:

(34) DT. EXEC.:

— RESULTADO —

## ERGOMETRIA

NUM.:

(35) DT. PEDIDO...:

(36) DT. EXEC.:

— RESULTADO —

## FONO

NUM.:

(37) DT. PEDIDO...:

(38) DT. EXEC.:

— RESULTADO —

## ECO.

NUM.:

(39) DT. PEDIDO...:

(40) DT. EXEC.:

— RESULTADO —

CONTINUAÇÃO - CASO Nº

### CATETERISMO

NUM.:

④1

DT. PEDIDO....

DIA - MÊS

④2

DT. EXEC.:

DIA - MÊS

RESULTADO

\*

### TRATAMENTO

④3

TIPO.....

CLÍNICO

CIRÚRGICO

EVOLUÇÃO

FIGURA V.1

variáveis diagnósticas:

- Anatomia coronariana - uma variável discreta de 8 classes que expressa a configuração de lesões nos 3 vasos coronarianos.
- Função ventricular, uma variável dicotômica que expressa a ausência ou presença de anormalidades como acinesia, discinesia e hipocinesia moderada a grave.

A última seção do formulário nos oferece uma variável prognóstica ou seja, o tipo de tratamento que pode ser clínico ou cirúrgico.

Os dados dos pacientes foram colhidos a partir dos registros hospitalares, e transcritos para os formulários por médicos residentes. Os laudos de exames foram transcritos sob forma de texto livre e depois codificados pelo Dr. R.B.M. Chaves. Procedeu-se então à digitação e conferência dos dados para um arquivo seqüencial de registros de tamanho fixo e posteriormente à inversão deste arquivo para utilização pelo "software" APL - Data Interface II.

### CONSIDERAÇÕES SOBRE A QUALIDADE DO ARQUIVO DE PACIENTES

O banco de dados construído, pelo fato de ter sido colhido retrospectivamente sobre os registros do serviço de rotina de uma instituição oficial ao longo de uma década, é afetado pela maioria dos problemas que prejudicam a qualidade dos dados médicos, discutidos no Capítulo III. Temos um número reduzido de casos e a presença de muitas brechas no banco de dados. Assim, por exemplo, o descritor Colesterol foi medido em 64,7% dos casos, Triglicerídeos em 47,18%, Ácido Úrico em 32,3%, o exame de Ergometria foi realizado em apenas 30,9% dos casos, o de Ecocardiografia em 43,6% e o de Fonocardiografia teve de



ser eliminado do estudo por ter sido feito em apenas 2 casos. Em suma, acreditamos que o banco de dados obtido representa um exemplar típico do que podemos obter se decidirmos usar o patrimônio representado pelos dados registrados em nossos hospitais.

### V.3 - CONSTRUÇÃO DE UM SISTEMA DE SUPORTE À DECISÃO

Para facilitar e acelerar o trabalho de pesquisa desenvolvido procuramos montar um sistema de suporte à decisão dentro da conceituação de Sprague [165] ("Um sistema baseado em computador que auxilia decisores a enfrentarem problemas complexos através de interação com procedimentos de gestão e análise de dados"). As capacidades buscadas para esse sistema foram:

- Sistema interativo
- Facilidade de uso
- Suporte a decisões semi-estruturadas e não-estruturadas
- Suporte a todas as fases do processo de tomada de decisão (inteligência, delineamento e escolha)
- Construção de modelos suficientemente robustos para lidar com bancos de dados médicos com as imperfeições já comentadas.
- Generalidade visando a seu aproveitamento em outras áreas da medicina.

#### V.3.1 - Ambiente Computacional

Utilizamos um computador IBM 4341 e posteriormente um IBM-4381 dentro do ambiente operacional VM-CMS ("IBM Virtual Machine Facility" e "Conversational Monitor System").

As duas ferramentas básicas utilizadas foram:

- VSAPL (Linguagem APL em versão para o CMS)
- APLDI II (APL - Data Interface II)

Outras ferramentas complementares foram

- Full Screen Editor and Manager
- APL Statistical Library
- ADRS II BG (ADRS II Business Graphics)
- GDDM (Graphical Data Display Manager)

A título de informação vamos descrever resumidamente o APLDI-II.

#### APLDI-II

O APL Data Interface II [6] é uma facilidade de consulta de finalidade geral que fornece um meio relativamente simples, porém muito poderoso e econômico de interativamente obter acesso, analisar, manipular e gerar relatórios sobre informações armazenadas em arquivos "on line".

A flexibilidade destas funções permite ao usuário selecionar dados "on line" usando lógica "e/ou", termos relacionais tais como "menor que", "igual a" e valores computados, referenciando os campos através de nomes simbólicos. Além de prover capacidades potentes de seleção existem diversas funções que operam sobre os dados extraídos dos arquivos. É possível sumarizar, contar, obter distribuições de freqüências, acumular estatísticas, classificar, fazer tabulação cruzada, ordenar, subtotalizar e extrair dados para processamento por programas em APL do usuário. Os resultados de saída são exibidos em terminais de vídeo ou impressos caso seja necessário. O conceito de arquivo invertido é usado para aumentar a velocidade de acesso ao arquivo. Esta organização permite que somente os campos de interesse sejam lidos ignorando-se todos os demais campos.

O procedimento básico de consulta segue a seguinte seqüência de perguntas colocadas pelo sistema ao usuário:

FNAME FTYPE FMODE(s):

SELECTION:

FUNCTION:

FIELDS:

### V.3.2 - Estrutura do Sistema de Suporte à Decisão

A estrutura do sistema de suporte à decisão desenvolvido para o trabalho de pesquisa é formada por 3 subsistemas, a saber:

- Subsistema de Diálogo
- Subsistema de Dados
- Subsistema de Modelos

A figura (V.2) de Sprague [165] ilustra a composição de um sistema de suporte à decisão típico, seguido em nosso trabalho.

# SSD

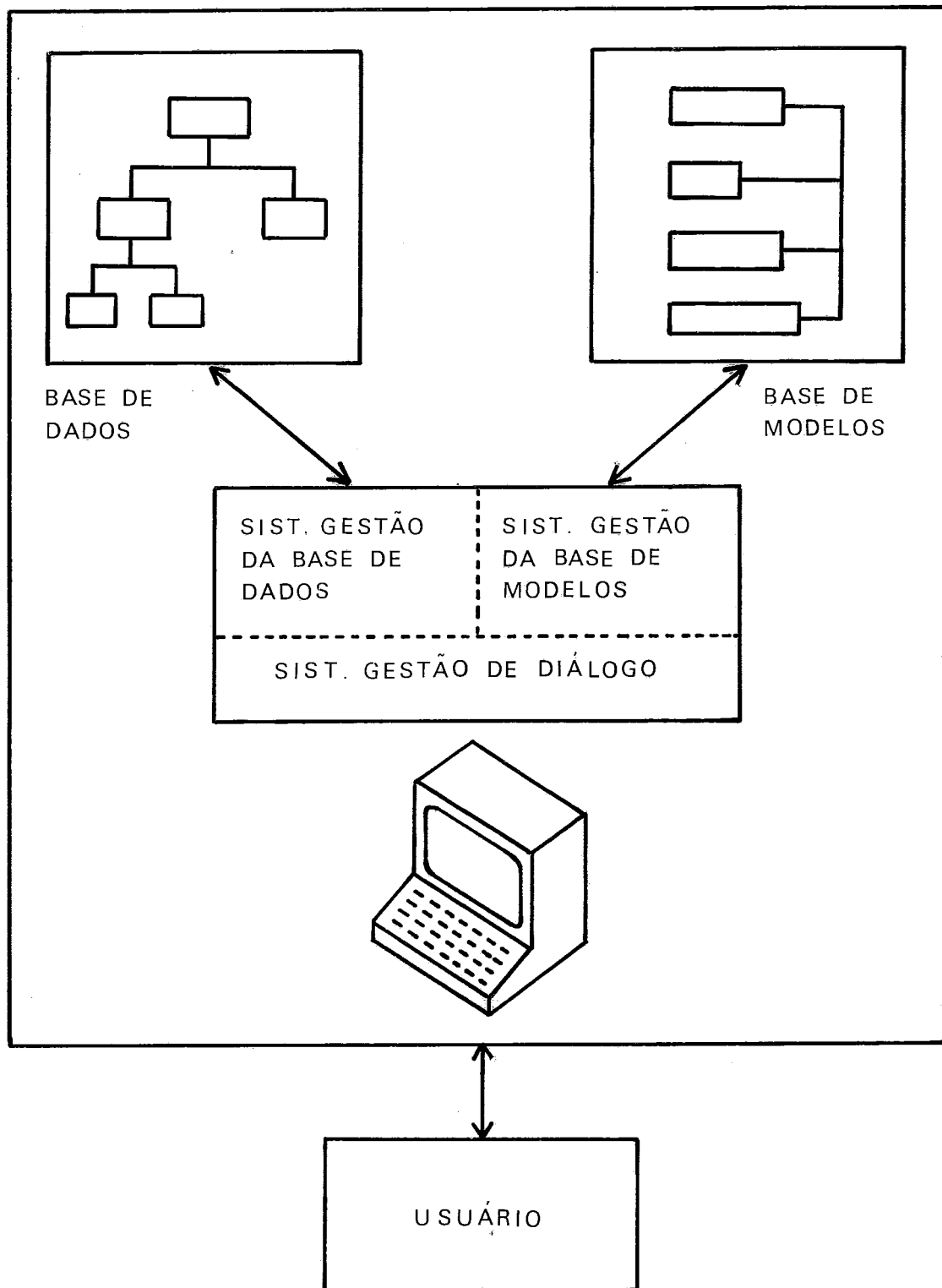


FIGURA V.2 COMPONENTES DO SISTEMA DE SUPORTE À DECISÃO

### **V.3.2.1- Subsistema de Diálogo**

O componente de diálogo do sistema de suporte à decisão é o software e hardware que provê a interface entre o usuário e o sistema. Muitas das características de flexibilidade e facilidade de uso do sistema advém das características de interação com o usuário. O sistema de diálogo utilizado envolve o uso de vídeo e impressora para apresentação de dados e o teclado com teclas de função para o controle do sistema e a entrada de dados e parâmetros. Os estilos de diálogo adotados foram o de menus e o de perguntas e respostas.

### **V.3.2.2 - Subsistema de Dados**

O subsistema de dados é formado pela base de dados (uma coleção de dados armazenados em computador) e um sistema de gestão da base de dados (um conjunto de programas usados para criar, manter, ter acesso, atualizar e proteger a base de dados).

A gestão da base de dados foi formada pelos programas do APLDI e por programas especialmente desenvolvidos pelo autor em APL.

A base de dados foi constituída pelos arquivos do APLDI e por arquivos armazenados em Workspaces do APL. Uma descrição detalhada da estrutura de dados pode ser encontrada no apêndice 2. O principal componente da base de dados foi o banco de dados de pacientes descrito na seção V.2. Porém diversos outros tipos de dados e parâmetros também constituem parte fundamental da base de dados. A seguir descrevemos as principais entidades representadas na base de dados.

#### **V.3.2.2.1 - Principais Entidades**

São as seguintes as principais entidades representa-

das na base de dados:

## A - Pacientes

São representados pelos registros do Banco de Dados de Pacientes. Seus atributos no arquivo construído para o APLDI são aqueles apresentados no formulário de coleta de dados. Para os demais programas desenvolvidos em APL os pacientes tem seus dados médicos agrupados em uma matriz que é armazenada nos próprios "workspaces" do APL. Esta matriz chamada MCASOS é apresentada no apêndice 1. A matriz MCASOS é formada apenas por descritores discretos. Para tanto, discretizamos os descritores contínuos conforme as seguintes faixas:

### Colesterol

$\leq 250$ mg%	(normal)
$> 250$ mg%	(anormal)

### Triglicerídeos

$\leq 160$ mg%	(normal)
$> 160$ mg%	(anormal)

### Glicemia

$\leq 110$ mg%	(normal)
$> 110$ mg%	(anormal)

### Ácido Úrico

$\leq 7,5$ mg%	(normal)
$> 7,5$ mg%	(anormal)

### Idade

$< 40$ anos
40 a 50 anos
$\geq 60$ anos

## B - Descritores

São variáveis de cunho médico que descrevem o estado de saúde do paciente sendo utilizadas para inferência sobre o diagnóstico ou prognóstico do paciente. Seus principais atributos são:

- Código
- Nome do descritor
- Tipo de descritor
  - Anterior à doença (fator de risco)
  - Posterior à doença (manifestação da doença)
- Exame ao qual pertence
- Lista de valores observáveis possíveis.

Apresentamos a seguir a lista dos descritores separados por tipo.

### DESCRITORES ANTERIORES À DOENÇA

<b>Código</b>	<b>Nome</b>
1	Cor
2	Sexo
3	Idade
11	Hipertensão Arterial
12	Atividade Física
13	Tabagismo
14	História Familiar de Coronariopatia
17	Colesterol > 250 MG por DL
18	Triglicerídeos >160 MG por DL
19	Glicemia >110 MG por DL
20	Ácido Úrico >7.5 MG por DL
31	Hipertrofia Ventricular
36	Hipertrofia Concêntrica

## DESCRITORES POSTERIORES À DOENÇA

<b>Código</b>	<b>Nome</b>
4	Dor Precordial
5	Infarto do Miocárdio Prévio
6	Capacidade Funcional
7	Tonteira
8	Síncope
9	Palpitação
10	Dispnéia aos Esforços
15	Presença de Arritmia Cardíaca
16	Presença de B4
21	Raios X de Tórax
22	Ergometria
23	Isquemia Inferior
24	Isquemia Lateral
25	Isquemia Anterior
26	Outras Isquemias
27	Necrose Inferior
28	Necrose Lateral
29	Necrose Anterior
30	Outras Necroses
32	Arritmias (ECG)
33	Alterações Inespecíficas da Repolarização Ventricular
34	Bloqueios
35	Ecocardiografia

### C - Aglomerados de Descritores

São conjuntos de descritores que apresentam interdependência estatística para a população de pacientes pertencentes a uma classe diagnóstica. Seus atributos são:

- Lista de descritores interdependentes
- Classe diagnóstica considerada.



## D - Observações

São os valores observáveis que podem ser obtidos da mensuração dos descritores sobre o paciente. Seus principais atributos são:

- Código da observação
- Nome da observação
- Descritor ao qual pertence
- Positividade quanto à doença coronariana
  - 0 = negativo
  - 1 = positivo
  - 1 = indiferente ou não aplicável

Apresentamos a seguir a lista de descritores e as observações associadas. O critério de positividade das observações com relação à doença coronariana é mostrado na coluna da direita.

Cód.	Descritor	Observação	Cód. Obs.	Crit. Posit.
1	Cor	Branca	1	1
		Negra	2	0
		Parda	3	-1
2.	Sexo	Masculino	1	1
		Feminino	2	0
3.	Idade	<40 anos	1	0
		40 a 59 anos	2	1
		>60 anos	3	1
4.	Dor Precordial	Sem dor	0	0
		Dor típica	1	1
		Dor atípica	2	0
5.	Infarto do Miocárdio prévio	Não	0	0
		Sim	1	1
6.	Capacidade Funcional	Um	1	0
		Dois	2	1
		Três	3	1
		Quatro	4	1

Cód.	Descritor	Observação	Cód. Obs.	Crit. Posit.
7	Tonteira	Não Sim	0 1	0 1
8	Síncope	Não Sim	0 1	0 1
9	Palpitação	Não Sim	0 1	0 1
10	Dispnéia aos Esforços	Não Sim	0 1	0 1
11	Hipertensão Arterial	Não Sim	0 1	0 1
12	Atividade Física	Sedentária Moderada Intensa	1 2 3	1 1 0
13	Tabagismo	0 a 20 Cig/dia >20 cig./dia	1 2	0 1
14	História Familiar de Coronariopatia	Não Sim	0 1	0 1
15	Presença de Arritmia cardíaca	Não Sim	0 1	0 1
16	Presença de B4	Não Sim	0 1	0 1
17	Colesterol >250 MG%	Não Sim	0 1	0 1
18	Triglicerídeos >160 MG%	Não Sim	0 1	0 1
19	Glicemia > 110 MG%	Não Sim	0 1	0 1
20	Ácido Úrico >7.5 MG%	Não Sim	0 1	0 1
21	Raios X de Tórax	Normal DCT 55 a 65% DCT 65 a 70% DCT >70%	1 2 3 4	0 1 1 1
22	Ergometria	Negativa Positiva	1 2	0 1
23	Isquemia Inferior	Não Sim	0 1	0 1

Cód.	Descritor	Observação	Cód. Obs.	Crit. Posit.
24	Isquemia Lateral	Não	0	0
		Sim	1	1
25	Isquemia Anterior	Não	0	0
		Sim	1	1
26	Outras Isquemias	Não	0	0
		Sim	1	1
27	Necrose Inferior	Não	0	0
		Sim	1	1
28	Necrose Lateral	Não	0	0
		Sim	1	1
29	Necrose Anterior	Não	0	0
		Sim	1	1
30	Outras Necroses	Não	0	0
		Sim	1	1
31	Hipertrofia Ventricular	Não	0	0
		Sim	1	1
32	Arritmias	Não	0	0
		Sim	1	1
33	Alter.Inespecíf. da Repolariz. Ventric.	Não	0	0
		Sim	1	1
34	Bloqueios	Não	0	0
		Sim	1	1
35	Ecocardiografia	Negativa	0	0
		Positiva	1	1
36	Hipertrofia Concêntrica	Não	0	0
		Sim	1	1

## E - Exames

Os principais atributos dos exames são:

- Código do exame
- Nome do exame
- Lista de descritores associados
- Utilidade para cada tipo de atributo de custo considerado.

- Lista de contraindicações
- Lista de exames conflitantes
- Lista de latências para cada outro exame
- Lista de exames pré-requisitos

Apresentamos a seguir os exames catalogados no sistema e os seus descritores associados.

<b>Código</b>	<b>Exame</b>	<b>Lista de Descritores Associados</b>
1	História	4,5,6,7,8,9,10,12,13,14,15
2	Exame físico	11,16
3	Laboratório	17,18,19,20
4	Raios-X de Tórax	21
5	Eletrocardiograma	23,24,25,26,27,28,29,30, 31,32,33,34
6	Ergometria	22
7	Ecocardiografia	35,36
8	Cateterismo Cardíaco	38,40

#### **F - Variáveis Diagnósticas e Prognósticas**

São as variáveis que temos interesse em inferir ou prever. Seus principais atributos são:

- Código da variável
- Nome da variável
- Lista de classes diagnósticas associadas
- Tipo de variável - diagnóstica ou prognóstica

#### **G - Classes Diagnósticas (Doença) ou Prognósticas**

São os valores que podem ser assumidos pelas variáveis diagnósticas ou prognósticas. Seus atributos são:

- Código da classe diagnóstica
- Nome da classe
- Probabilidade a priori de ocorrência na população

Listamos a seguir as variáveis diagnósticas e suas classes.

### G.1 - Variáveis Primárias

Duas variáveis diagnósticas e uma prognóstica nos são fornecidas diretamente a partir do formulário de coleta de dados.

#### ANATOMIA CORONARIANA

Expressa a configuração de lesões nos vasos coronarianos. Suas classes são:

Descrição da classe	Código
- Sem lesões significativas	1
- Com lesões na Coronária Direita Isolada	2
- Com lesões na Descendente Anterior Isolada	3
- Com lesões na Coronária Circunflexa Isolada	4
- Com lesões na Coronária Direita e Descendente Anterior	5
- Com lesões na Coronária Direita e Circunflexa	6
- Com lesões na Descendente Anterior e Coronária Circunflexa	7
- Com lesões nos 3 vasos	8

#### FUNÇÃO VENTRICULAR (OU VENTRICULOGRAFIA)

Expressa a presença de anormalidades como acinesia, discinesia e hiposinesia moderada a grave. Suas classes são:

<b>Descrição da Classe</b>	<b>Código</b>
- Função ventricular normal	1
- Função ventricular anormal	2

#### **TIPO DE TRATAMENTO**

Expressa a conduta de tratamento adotada para o paciiente. Suas classes são:

<b>Descrição da Classe</b>	<b>Código</b>
- Tratamento clínico	1
- Tratamento cirúrgico	2

#### **G.2 - Variáveis Derivadas**

A partir das variáveis diagnósticas primárias podemos derivar uma série de outras variáveis diagnósticas úteis a través de processos de aglutinação de classes, produto cartesiano e outros. Apresentamos algumas variáveis derivadas:

#### **PRESENÇA DE LESÕES SIGNIFICATIVAS (OU PRESENÇA DE DOENÇA CORONARIANA)**

Expressa a presença ou ausência de doença coronariana no paciente.

<b>Descrição da Classe</b>	<b>Código</b>
- Sem lesões significativas	1
- Com lesões significativas	2

## NÚMERO DE VASOS LESADOS

Expressa a quantidade de vasos coronarianos com lesões significativas:

Descrição da classe	Código
- Zero vasos lesados	0
- Um vaso lesado	1
- Dois vasos lesados	2
- Três vasos lesados	3

## PRESENÇA DE LESÕES x FUNÇÃO VENTRICULAR

É formada pelo produto cartesiano entre as variáveis Presença de Lesão e Função Ventricular.

Descrição da Classe	Código
- Anatomia e F. Ventricular normais	1
- Anatomia normal e F. Ventric. anormal	2
- Anatomia anormal e F. Ventric. normal	3
- Anatomia e F. Ventricular anormais	4

## PERFIL DO PACIENTE

Formada pela variável anterior excluindo-se a classe 2.

## CORONÁRIA DIREITA

Expressa a ausência (=1) ou presença (=2) de lesões significativas na Coronária Direita.

## **DESCENDENTE ANTERIOR**

Expressa a ausência (=1) ou presença (=2) de lesões significativas na Descendente Anterior.

## **CORONÁRIA CIRCUNFLEXA**

Expressa a ausência (=1) ou presença (=2) de lesões significativas na Coronária Circunflexa.

## **H - Modelos Bayesianos de Classificação Diagnóstica com Pressuposição de Independência dos Descritores**

Correspondem aos modelos discutidos no Capítulo III. Seus atributos são:

- Nome do modelo
- Código do modelo
- Variável diagnóstica (com suas classes)
- Lista de descritores (com seus valores observáveis)
- Probabilidades condicionais de observações conhecidas a classe diagnóstica
- Probabilidades a priori das classes diagnósticas
- Indicador de que o modelo é formado por aglutinação de classes de outra variável diagnóstica e que a inferência deve ser feita a partir do modelo original.
- Código do modelo original se for o caso
- Esquema de aglutinação de classes empregado

## **I - Modelos Bayesianos de Classificação Diagnóstica Baseados em Aglomerados de Descritores Interdependentes**

Correspondem aos modelos discutidos no capítulo III.



Seus atributos são:

- Nome do modelo
- Código do modelo
- Variável diagnóstica
- Lista de aglomerados de descritores
- Probabilidades condicionais de configurações de observações conhecida a classe diagnóstica
- Probabilidades a priori das classes diagnósticas.

### J - Geradores de Planos de Investigação Médica

Correspondem ao modelo proposto no capítulo IV. Seus principais atributos são:

- Nome do gerador
- Código do gerador
- Modelo bayesiano de classificação diagnóstica associado
- Lista de exames permitidos
- Definição do estado inicial (exames já realizados)
- Nome do operador  $T$  de geração de sucessores
- Nome do programa que calcula a probabilidade de revelação de alguma doença
- Valor do limiar para definição de estado terminal
- Função heurística usada
- Matriz de probabilidade  $\prod_{j\alpha}$
- Matriz de probabilidade  $q_{j\alpha}$
- Estratégia selecionada
- Lista de utilidades dos exames
- Operador de adição de utilidades

### K - Estratégias Elementares de Investigação

Seus atributos são:

- Nome da estratégia elementar
- Lista de utilidades por exame
- Fórmula do operador de adição de utilidades

## L - Estratégias Compostas de Investigação

Seus atributos são:

- Nome da estratégia composta
- Fórmula de composição
- Operador de adição de utilidades

### V.3.2.3 - Subsistema de Modelos

O terceiro componente do Sistema de Suporte à Decisão visa permitir a aplicação de modelos matemáticos e funções analíticas sobre os dados da base de dados para a tomada de decisão. Um aspecto muito poderoso dos Sistemas de Suporte à Decisão é a sua capacidade de integrar o acesso aos dados e aos modelos. Isto permite ao usuário empregar os modelos de decisão sem se preocupar com a compatibilização dos dados de entrada e saída referentes a cada modelo específico.

Podemos distinguir no sistema de modelos uma base de modelos e funções de gestão desta base de modelos. No nosso caso a base de modelos é constituída por funções analíticas sob forma de programas de computador bem como de especificações de modelos bayesianos de classificação diagnóstica e geradores de planos de investigação médica. Os primeiros correspondem às funções já prontas do APLDI e do APL Statistical Library bem como às funções programadas pelo autor. Os últimos, na verdade são dados, e como tal são armazenados na base de dados, conforme já citamos.

O sistema de gestão da base de modelos corresponde a um conjunto de programas que permite gerar, reestruturar, atualizar, exhibir as características, arquivar, recuperar modelos

bem como exibir resultados da operação dos modelos.

### **V.3.2.3.1 - Funções Especialmente Desenvolvidas para o Sistema de Suporte à Decisão**

Listamos a seguir as principais funções desenvolvidas em APL para o sistema de suporte à decisão. Os programas de computador correspondentes podem ser encontrados no Apêndice 2.

#### **PRINCIPAIS FUNÇÕES**

##### **A - Geração/Atualização da Base de Dados para uma Aplicação Médica Genérica**

Definição, aquisição, arquivamento das ocorrências que farão parte do sistema, referentes às seguintes entidades:

- Pacientes
- Exames
- Descritores
- Observações
- Variáveis diagnósticas
- Classes diagnósticas
- Estratégias elementares de investigação
- Estratégias compostas de investigação

Estabelecimento dos relacionamentos entre entidades da base de dados:

- Exames x Descritores
- Descritores x Observações
- Variáveis diagnósticas x classes
- Exames x Estratégias elementares
- Estratégias compostas x Estratégias elementares
- Pacientes x Observações
- Pacientes x Classes diagnósticas

Apresentação da lista de ocorrências por entidade.

Apresentação de uma ocorrência específica de uma entidade.

Atualização de ocorrências de entidades arquivadas na base de dados.

Apresentação dos relacionamentos entre entidades.

## B - Funções Gerais

- Seleção de casos do arquivo de pacientes
- Definição de conjuntos de casos para treinamento e teste de modelos bayesianos segundo as modalidades:
  - Seleção aleatória
  - Os n primeiros
  - Seleção arbitrária
  - Seleção complementar
- Seleção de descritores
- Apresentação de casos e descritores selecionados
- Determinação do poder informativo de cada descritor individual com relação a uma variável diagnóstica especificada calculando-se:
  - Sensibilidade
  - Especificidade
  - Intervalo de confiança clássico para sensibilidade
  - Intervalo de confiança clássico para especificidade
  - Valor preditivo
  - Probabilidade total de erro
  - Chi-quadrado da tabela de contingência
  - Realização de teste de hipótese de associação
- Cálculo da matriz de associação entre descritores (expressa pelos  $\chi^2$  das tabelas de contingência ou opcionalmente pelos coeficientes de Cramer)
- Determinação da estrutura de aglomerados de descritores para cada classe diagnóstica.
- Apresentação da lista de modelos arquivados na base de dados
- Estimação de probabilidades a priori de classes diagnósticas a partir do banco de dados.

### C - Gerenciamento dos Modelos Bayesianos de Classificação Diagnóstica com Pressuposição de Independência dos Descritores

- Criação de modelos diretos
- Criação de modelos com classes aglutinadas a posteriori
- Atualização
- Calibração
- Teste
- Apresentação descritiva do modelo
- Armazenamento/recuperação/eliminação de modelo no arquivo
- Aceita probabilidades a priori fornecidas pelo usuário ou calculadas a partir do arquivo de pacientes
- Definição do valor mínimo  $p_{\min}$  para probabilidades condicionais
- Inferência diagnóstica
- Relatório analítico dos resultados do teste do modelo contendo
  - Tabela de contingência da classe diagnóstica calculada versus classe real do caso
  - Índice porcentual de diagnósticos corretos ou eficiência
  - Chi-quadrado da tabela de contingência
  - Entropia média
  - Ganho médio de informação com relação à situação inicial definida pelas probabilidades a priori e por probabilidades uniformes.

### D - Gerenciamento de Modelos Bayesianos de Classificação Diagnóstica Baseados em Aglomerados de Descritores Interdependentes

- Criação de modelos
- Atualização

- Calibração
- Teste de modelo previamente calibrado
- Teste de modelo não calibrado previamente com estimação concomitante das probabilidades condicionais de configurações de observações
- Apresentação do modelo
- Armazenamento/recuperação/eliminação de modelo no arquivo
- Aceita probabilidades a priori fornecidas pelo usuário ou calculadas a partir do arquivo de pacientes
- Inferência diagnóstica
- Definição do valor mínimo  $P_{\min}$  para probabilidades condicionais
- Relatório analítico dos resultados do teste do modelo

#### E - Gerenciamento de Geradores de Planos de Investigação Médica

- Criação de um gerador
- Atualização de um gerador
- Apresentação descritiva de um gerador
- Armazenamento/recuperação/eliminação de um gerador no arquivo
- Seleção de estratégia de investigação
- Geração de planos de investigação para um paciente específico segundo estratégias elementares, compostas ou robusta e o nível de precisão requerido para o diagnóstico
- Entrada/atualização de parâmetros  $q_{j\alpha}$  para a função heurística
- Estimação dos parâmetros  $q_{j\alpha}$  a partir do arquivo de pacientes
- Entrada/atualização de parâmetros  $\Pi_{j\alpha}$
- Determinação da probabilidade de revelação de alguma doença usando os parâmetros  $\Pi_{j\alpha}$  ou diretamente

do arquivo de pacientes.

## **F - Aplicação do Sistema na Prática Médica**

- Colheita de dados básicos do paciente
- Arquivamento/recuperação/eliminação de registro num arquivo de pacientes
- Atualização de dados do paciente
- Apresentação dos dados do paciente e impressão se necessário
- Avaliação diagnóstica do paciente pela aplicação de um ou mais modelos bayesianos de classificação diagnóstica
- Geração de plano de investigação diagnóstica de acordo com a estratégia selecionada pelo médico e com a precisão requerida do diagnóstico
- Acréscimo de resultados de exames ao registro do paciente
- Alteração das probabilidades a priori das classes diagnósticas.

## **V.4 - EXPERIMENTAÇÃO**

Descrevemos a seguir o trabalho semi-experimental desenvolvido com a base de dados e o sistema de suporte à decisão construídos.

### **A - Exploração Inicial dos Dados de Pacientes**

Além das distribuições em frequência das diversas variáveis diagnósticas e da determinação das principais falhas no arquivo de pacientes, procedemos a estudos visando investigar o poder informativo de cada descritor.

Com relação à variável "Presença de Lesões Significativas" e com relação à variável "Função Ventricular" determinamos:

- Tabela de contingência
- Chi-quadrado
- Sensibilidade
- Especificidade
- Intervalo de Confiança para Sensibilidade
- Intervalo de Confiança para Especificidade
- Valor Preditivo
- Probabilidade total de erro
- Teste de hipótese de associação ao nível de significância de 5%
- Teste de hipótese da diferença de médias nas variáveis contínuas

Com relação à variável "Anatomia Coronariana" determinamos para cada descritor:

- Tabela de contingência
- Chi quadrado
- Teste de hipótese de associação ao nível de significância de 5%

## **B - Definição dos Conjuntos de Calibração e Teste**

Decidimos estudar os modelos bayesianos de classificação diagnóstica segundo as estratégias já apresentadas:

1. "O Banco de Dados como Universo"
2. "O Banco de Dados como Amostra de um Universo Estacionário"

No 1º caso os conjuntos de treinamento e de teste são idênticos e abrangem todos os casos do arquivo de pacientes.



Para 2º caso temos de particionar o arquivo de pacientes em 2 conjuntos disjuntos. Dedicamos 90 casos ao conjunto de treinamento que foram selecionados aleatoriamente usando o gerador de números randômicos do APL. São eles:

1 2 4 7 8 9 11 12 15 16 19 22 23 25 26 27  
29 33 35 36 37 39 40 41 42 44 45 49 50 52 53 55  
56 57 58 61 63 64 65 66 67 68 69 70 72 73 74 75  
76 77 78 79 80 81 82 83 85 87 91 93 95 98 102 105  
107 108 109 110 111 113 114 116 117 119 120 123 124 125 126 128  
130 131 133 134 136 138 139 140 141 142

Os casos restantes são dedicados ao conjunto de teste, a saber:

3 5 6 10 13 14 17 18 20 21 24 28 30 31 32 34  
38 43 46 47 48 51 54 59 60 62 71 84 86 88 89 90  
92 94 96 97 99 100 101 103 104 106 112 115 118 121 122 127  
129 132 135 137

### **C - Construção de Modelos Bayesianos de Classificação Diagnóstica com Pressuposição de Independência de Descritores**

Construímos os seguintes modelos com pressuposição de independência de descritores:

- Presença de Lesões Significativas
- Número de Vasos lesados
- Anatomia Coronariana
- Função Ventricular
- Presença de Lesões x Função Ventricular
- Perfil
- Tipo de Tratamento (não incluindo o exame de cateterismo)
- Tipo de Tratamento (incluindo o exame de cateterismo).

Modelos segundo o esquema de multi-elegibilidade foram criados para avaliação da anatomia coronariana. São eles:

- Presença de Lesões Significativas na Coronária Direita
- Presença de Lesões Significativas na Descendente Anterior
- Presença de Lesões Significativas na Coronária Circunflexa

Construímos também dois modelos baseados em aglutinação de classes do modelo Anatomia Coronariana, a saber:

- Presença de Lesões Significativas
- Número de Vasos Lesados

O esquema de aglutinação de classes foi:

Anatomia Coronariana	Presença de Lesões	Número de Vasos Lesados
1	1	0
2	2	1
3	2	1
4	2	1
5	2	2
6	2	2
7	2	2
8	2	3

Os modelos referidos acima foram calibrados nas duas modalidades (B.D. como Universo, B.D. como Amostra).

#### D - Estimação de Probabilidades a Priori

Foram usados 3 conjuntos de valores de probabilidades a priori para o teste dos modelos bayesianos de classificação

ção diagnóstica, a saber:

- Estimativas calculadas a partir do arquivo de pacientes
- Estimativas subjetivas realizadas por um grupo de cardiologistas para as classes das variáveis Presença de Lesões Significativas, Número de Vasos Lesados e Função Ventricular
- Probabilidades a priori uniformes

#### F - Estudo dos Modelos Bayesianos de Classificação Diagnóstica com Pressuposição de Independência dos Descritores

Para avaliar o desempenho dos modelos construídos realizamos testes segundo as estratégias "O Banco de Dados como Universo" e "O Banco de Dados como Amostra". Os testes foram realizados usando probabilidades a priori uniformes e posteriormente usando as probabilidades a priori estimadas a partir do arquivo de pacientes. Os testes foram realizados duas vezes. Na 1ª utilizamos  $p_{\min}=0$  e na 2ª  $p_{\min}=10^{-6}$ . Para cada execução do procedimento de teste construímos a tabela de contingência da classe diagnóstica calculada x classe real do caso, calculamos seu Chi-quadrado, realizamos testes de hipótese de associação ao nível de significância de 0,05 e 0,01. Além disso calculamos o percentual de diagnósticos corretos (também chamado de eficiência do modelo), a entropia média e o ganho médio de informação com relação à situação inicial.

Relacionamos a seguir uma série de ensaios realizados sobre o comportamento dos modelos bayesianos de classificação diagnóstica com pressuposição de independência para a área de doença coronariana.

- Avaliação do desempenho dos modelos em função da faixa de probabilidade da classe diagnóstica de maior probabilidade a posteriori.

- Avaliação do porcentual de diagnósticos corretos nas  $n$  classes de maior probabilidade a posteriori.
- Avaliação do desempenho em função do valor mínimo permitido para as probabilidades condicionais ( $p_{\min}$ )
- Avaliação do desempenho em função dos valores das probabilidades a priori.
- Avaliação do desempenho em função do tamanho do banco de dados de calibração.
- Avaliação do desempenho em função do número de descritores utilizados
- Avaliação do desempenho em função do limiar para o estabelecimento do diagnóstico
- Avaliação do efeito da inclusão e exclusão do exame de ecocardiografia sobre o desempenho dos modelos
- Avaliação do desempenho dos modelos usando apenas descritores que foram mensurados em mais de 90% dos casos do arquivo de pacientes.

#### **F - Estudo das Associações entre Descritores com Suspeita de Doença Coronariana**

O objetivo deste estudo foi determinar, seguindo o esquema sugerido na seção III.5.1, os aglomerados de descritores interdependentes necessários à construção de modelos bayesianos de classificação diagnóstica. Determinamos estruturas de aglomerados para as classes das seguintes variáveis diagnósticas:

- Presença de Lesões Significativas
- Anatomia Coronariana
- Função Ventricular
- Coronária Direita

Construímos estruturas de aglomerados para diversos níveis de significância, a saber:

$$\alpha = 10^{-5} / 5 \cdot 10^{-4} / 7,5 \cdot 10^{-4} / 10^{-3} \\ 2,5 \cdot 10^{-3} / 5 \cdot 10^{-3} / 7,5 \cdot 10^{-3} / \\ 0,01 / 0,025 / 0,05 / 0,075 / \\ 0,1 / 0,25$$

#### G - Construção de Modelos Bayesianos de Classificação Diagnóstica Usando Aglomerados de Descritores

Construímos modelos para os níveis de significância citados no item anterior e para as seguintes variáveis diagnósticas:

- Presença de Lesões Significativas
- Anatomia Coronariana
- Função Ventricular
- Coronária Direita

Os modelos citados foram calibrados dentro da modalidade "O Banco de Dados como Amostra de Um Universo Estacionário". Os modelos do 1º tipo (Presença de Lesões Significativas) também foram calibrados na modalidade "O Banco de Dados como Universo".

## H - Estudo dos Modelos Bayesianos de Classificação Diagnóstica Usando Aglomerados de Descritores Interdependentes

Os modelos construídos no ítem anterior foram testados segundo as modalidades em que foram calibrados. Foi utilizado como valor mínimo permitido para probabilidades condicionais,  $p_{\min} = 10^{-6}$ . O mesmo tipo de análise de resultados usado para os modelos de independência, foi realizado também aqui.

### BATERIA DE MODELOS COM NÍVEL DE SIGNIFICÂNCIA CRESCENTE

Uma modalidade diferente de teste foi aplicada ao banco de dados de teste. Cada caso foi submetido seqüencialmente à análise por uma bateria de modelos com nível de significância  $\alpha$  crescente. A decisão só foi tomada quando se obteve uma probabilidade a posteriori superior a um limiar especificado, por exemplo: 0,99, ou quando se esgotou a bateria de modelos. Aplicamos este tipo de teste às variáveis diagnósticas Presença de Lesões Significativas, Anatomia Coronariana e Função Ventricular. Reaplicamos o procedimento para um subconjunto de teste formado apenas por pacientes que realizaram o exame de ecocardiografia.

Os seguintes estudos de comportamento dos modelos bayesianos com aglomerados de descritores na área de doença coronariana foram realizados:

- Avaliação do desempenho do modelo Presença de Lesões Significativas em função do valor mínimo permitido para as probabilidades condicionais de configurações de observações.
- Avaliação do desempenho do modelo Presença de Lesões Significativas em função do limiar para o estabelecimento do diagnóstico.

I - Construção de Geradores de Planos de Investigação Médica

Construímos um gerador simplificado para a variável diagnóstica Presença de Doença Coronariana.

---

---

**VI - RESULTADOS - APRECIÇÃO E  
ANÁLISE**

---

---



Os objetivos deste capítulo são apresentar e discutir os principais resultados dos experimentos realizados. Muitos dos resultados serão apresentados sob forma de tabelas e gráficos que por razões de conveniência serão concentrados no final do capítulo e referenciados sob o nome geral de figuras. No decorrer deste capítulo vamos ter a oportunidade de realizar diversas inferências de cunho médico na área de doença coronariana. É essencial neste caso, que não percamos de vista o tipo de população ao qual elas se aplicam, e o grau limitado de confiança que o tamanho de nossa amostra permite.

## VI.1 - VALOR DIAGNÓSTICO DOS DESCRITORES

As figuras (VI.1) a (VI.10) ilustram os principais resultados obtidos no trabalho inicial de investigação do poder informativo dos descritores quando analisados individualmente com relação às variáveis diagnósticas Presença de Doença Coronariana, Anatomia Coronariana e Função Ventricular. Os resultados foram separados em dois grupos de acordo com o tipo dos descritores anteriores à doença (fatores de risco) e posteriores à doença (manifestação da doença). São apresentados para cada descritor o número de casos usado para calcular os parâmetros, a sensibilidade, especificidade, valor preditivo,  $\chi^2$  da tabela de contingência, graus de liberdade de  $\chi^2$ , teste de hipótese de associação, o percentual total de erros de classificação e intervalos de confiança para sensibilidade e especificidade

As figuras (VI.11) a (VI.14) apresentam os gráficos ROC dos descritores anteriores e posteriores à doença com relação às variáveis Presença de Lesões Significativas e Função Ventricular. O gráfico ROC é uma maneira interessante de visualizar o poder informativo dos descritores. Trata-se de um gráfico sensibilidade x especificidade onde cada descritor será representado por um ponto. No gráfico ROC a diagonal representada pela reta

$$\text{sensibilidade} + \text{especificidade} = 1$$

contém os descritores que não trazem nenhuma informação diagnóstica, ou seja, descritores que operam baseados na sorte [53]. O ponto (0,5 0,5), por exemplo, corresponde ao lançamento de uma moeda, o ponto (0 1) corresponde a decidir sempre pela presença da doença e o ponto (1 0) a decidir sempre pela ausência da doença. A magnitude do desvio do ponto representativo de um descritor com relação à esta diagonal é uma indicação do conteúdo de informação do descritor

Da análise dos resultados constantes, nas figuras (VI.1) a (VI.14) pode-se notar que:

- Existe consistência entre os resultados dos testes de hipótese sobre associações de descritores com variáveis diagnósticas e o afastamento com relação a diagonal de não informação nos gráficos ROC.
- Os descritores usados na investigação de doença coronariana, vistos isoladamente com relação às variáveis diagnósticas, são surpreendentemente pouco informativos. Os descritores estão situados em sua vasta maioria num entorno muito próximo da diagonal de não informação ocorrendo, no caso dos descritores posteriores à doença, uma polarização em um grupo de alta sensibilidade com baixa especificidade e outro grupo de baixa sensibilidade com alta especificidade.
- Poucos descritores apresentaram associações significativas com as variáveis diagnósticas. Destacamos as associações encontradas ao nível de significância de 5%, a seguir:

--Associações com relação à Presença de Doença Coronariana

- Variáveis anteriores à doença

- Cor

- Sexo

- Variáveis posteriores à doença

- Dor Precordial

- Infarto do Miocárdio Prévio

- Presença de B4

- Necrose Inferior
  - Necrose Anterior
  - Outras Necroses
  - Ecocardiografia
- Associações com relação à Função Ventricular
- Variáveis anteriores à doença
    - Sexo
  - Variáveis posteriores à doença
    - Infarto do Miocárdio Prévio
    - Presença de Arritmia Cardíaca
    - Raios X de Tórax
    - Isquemia Anterior
    - Necrose Inferior
    - Necrose Anterior
    - Outras Necroses
    - Arritmias (ECG)
    - Alterações Inespecíficas da Repolarização Ventricular
    - Ecocardiografia
- Associações com relação à Anatomia Coronariana
- Variáveis anteriores à doença
    - Sexo
    - Cor
  - Variáveis posteriores à doença
    - Infarto Prévio do Miocárdio
    - Capacidade Funcional
    - Necrose Inferior
    - Necrose Anterior
    - Outras Necroses
    - Ecocardiografia
- O melhor descritor para as 3 variáveis diagnósticas referidas foi sem dúvida a Ecocardiografia. Com relação à Presença de Doença Coronariana, este descritor revelou-se um teste patognômico (Sensibilidade e especificidade 100%). Convém, no entanto, frisar que a técnica de interpretação do exame de eco-

cardiografia é de difícil aplicação em pacientes hipertensos ou que apresentem hipertrofia concêntrica, o que reduz o seu potencial.

- Existe uma notável semelhança na distribuição dos descritores nos gráficos ROC para as variáveis Presença de Doença Coronariana e Função Ventricular.
- A eficiência dos descritores (100 - Porc. erro) nas figuras (VI.1) a (VI.4) pode ser usada como um índice de quão bem os descritores classificam os pacientes, embora não como um padrão absoluto de eficiência, pois isto só seria válido caso os custos e benefícios das conseqüências fossem iguais. Os 5 descritores mais eficientes para classificar a Presença de Doença Coronariana são:

Ecocardiografia	100,00
Cor	83,90
Sexo	76,06
Dor Precordial	75,00
Idade	74,47

Os 5 descritores mais eficientes para classificar a Função Ventricular são:

Ecocardiografia	66,13
Presença de Arritmias (ECG)	64,12
Infarto do Miocárdio Prévio	62,86
Necrose Inferior (ECG)	62,41
Raios-X do Tórax	60,66

Observe-se que 3 dos melhores descritores para classificação de Presença de Doença Coronariana são os fatores de risco: Sexo, Cor e Idade. É bom não esquecer que a eficiência, ao contrário da sensibilidade e especificidade, depende também da prevalência da doença, estando por isso profundamente ligada à população amostrada.

## VI.2 - ESTUDO DO MODELO BAYESIANO DE CLASSIFICAÇÃO DIAGNÓSTICA BASEADO NA PRESSUPOSIÇÃO DE DESCRITORES INDEPENDENTES

Os principais resultados desta etapa do trabalho são apresentados nas figuras (VI.15) a (VI.53). Passamos a comentar e discutir alguns pontos relevantes.

### VI.2.1 - Estimativas de Probabilidades a Priori das Classes Diagnósticas

As estimativas de probabilidade a priori das classes diagnósticas, calculadas a partir do banco de dados e exibidas na figura (VI.15), são as melhores que podemos usar nos testes a serem realizados nessa pesquisa, por serem exatas para a população sob investigação.

As estimativas subjetivas de probabilidades a priori realizadas por um grupo de cardiologistas (figura (VI.16)) divergem das estimativas calculadas a partir do banco de dados contemplando a classe de pacientes normais com um acréscimo de aproximadamente 20%. Isto é explicável, pois a população visada pelos médicos, ou seja, a de pacientes com suspeita de doença coronariana, abrange os conjuntos de pacientes que fizeram ou não fizeram o exame de cateterismo. Justamente no grupo dos pacientes que não fizeram o exame de cateterismo é esperável uma proporção maior de normais.

A classe 2 da variável Presença de Lesões x Função Ventricular (correspondente à ausência de lesões significativas com função ventricular anormal) apresenta uma probabilidade a priori extremamente baixa, o que já era esperado pois no processo de evolução da doença coronariana a alteração da função ventricular é um fenômeno que decorre do aparecimento de lesões significativas nas artérias coronárias, e não o contrário.

### VI.2.2 - Estimativas de Probabilidades Condicionais das Observações

As figuras (VI.17) a (VI.26) mostram os valores estimados para as probabilidades condicionais dos valores observáveis dos descritores com relação às principais variáveis diagnósticas.

Podemos observar a existência freqüente de estimativas nulas de probabilidades condicionais, responsáveis por anomalias na análise dos dados dos pacientes, conforme já discutido. O número de estimativas nulas cresce com a diminuição do número médio de casos de treinamento por classe conforme ilustrado na figura (VI.27) para a família de modelos derivados a partir da variável Anatomia Coronariana. Podemos portanto afirmar que a recomendação de uso de  $p_{\min} \neq 0$  cresce em importância para modelos com grande número de classes e pequenos bancos de dados de treinamento.

### VI.2.3 - Avaliação do Desempenho dos Modelos com Pressuposição de Independência

As figuras (VI.28) e (VI.29) mostram os resultados dos testes de desempenho de diversos modelos bayesianos de classificação diagnóstica baseados na pressuposição de independência, sob uma variedade de condições quanto a: probabilidades a priori das classes diagnósticas, modalidade de calibração e teste e probabilidade condicional mínima utilizada.

Os resultados obtidos foram de um modo geral muito bons. Obtivemos associações significativas entre as decisões realizadas pelos modelos e as classes reais dos pacientes ao nível de 1% para todos os 8 modelos quando treinados na modalidade "O Banco de Dados como Universo" (figura (VI.28)). Já na modalidade "O Banco de Dados como Amostra", obtivemos:

- associações significativas ao nível de 1% para os modelos:
  - Presença de Lesões Significativas
  - Número de Vasos Lesados
  - Presença de Lesões x Função Ventricular
  
- associações significativas ao nível de 5% para os modelos
  - Função Ventricular
  - Perfil
  - Tratamento incluindo dados de cateterismo

Os modelos Anatomia Coronariana e Tipo de Tratamento sem dados do cateterismo não apresentaram associações significativas nem sequer ao nível de 10%.

O modelo Presença de Doenças Coronariana apresentou em suas decisões

Sensibilidade	=	95,23%
Especificidade	=	80,00%
Valor Preditivo positivo	=	95,23%
Valor Preditivo negativo	=	80,00%

quando treinado na modalidade "O Banco de Dados como Amostra", usando  $p_{\min} = 10^6$  e com probabilidades a priori estimadas do banco de dados. Seu desempenho só foi superado pelo descritor Ecocardiografia que é patognomônico e foi realizado em 43,6% dos casos. Podemos tentar estabelecer uma comparação com o desempenho dos médicos no que tange ao valor preditivo positivo. Dos 142 pacientes enviados para a realização da cineangiocoronariografia, 31 se mostraram normais, o que dá à decisão médica um valor preditivo de 78,71% bastante inferior aos 95,23% de modelo. Com relação ao valor preditivo negativo, a comparação não é possível pois desconhecemos o número e a distribuição de pacientes atendidos que não realizaram o cateterismo.

#### VI.2.4 - Influência da Ecocardiografia sobre o Desempenho

Para responder à pergunta se o bom desempenho dos modelos não estaria fundamentado exclusivamente nos 43,6% dos casos em que foi realizada a ecocardiografia, criamos e testamos modelos para Presença de Doença Coronariana, Número de Vasos Lesados e Anatomia Coronariana excluindo-se os resultados deste exame (Figura (VI.30)). Constatamos uma queda considerável no desempenho, que não impediu que obtivéssemos associações significativas ao nível de 1% para os modelos treinados na modalidade "O Banco de Dados como Universo" e para o modelo Presença de Doença Coronariana na modalidade "O Banco de Dados como Amostra".

O modelo Presença de Lesões Significativas ("O Banco de Dados como Amostra",  $p_{\min} = 10^{-6}$ , probabilidades a priori estimadas do banco de dados) apresentou:

Sensibilidade	=	97,61%
Especificidade	=	30,00%
Valor preditivo positivo	=	85,41%
Valor preditivo negativo	=	75,00%

Estes resultados de fato são comparáveis aos dos melhores descritores, por exemplo: Cor do paciente. Revelam uma incapacidade específica do modelo "Presença de Doença Coronariana" com pressuposição de independência de melhorar o desempenho além do apresentado pelos descritores mais eficientes, através da agregação da informação de descritores menos eficientes. Teremos a oportunidade de voltar a este tema mais adiante.

### VI.2.5 - Influência da Modalidade de Calibração

O porcentual de diagnósticos corretos (eficiência) é maior quando usamos a estratégia de calibração e teste "O Banco de Dados como Universo", o que é naturalmente esperado (ocorreu apenas uma exceção). A entropia média tende a ser menor para a estratégia "O Banco de Dados como Amostra", pois nesta modalidade verifica-se um maior número de diagnósticos de certeza decorrente do elevado número de estimativas nulas de probabilidades



condicionais.

### **VI.2.6 - Desempenho em Função do Número de Classes Diagnósticas**

A eficiência dos modelos de classificação diagnóstica tende a diminuir com o crescimento do número de classes. Assim por exemplo, para os modelos da família da variável Anatomia Coronariana podemos constatar, na medida que vamos desdobrando a classe de Presença de Lesões, uma progressiva deterioração do desempenho (92,31% → 38,46% → 34,6% na modalidade "O Banco de Dados como Amostra"). Isto pode ser explicado neste exemplo pela diminuição do número médio de casos de treinamento por classe bem como pelo crescimento da dificuldade de diferenciação entre as classes que na variável Anatomia Coronariana atinge um elevado grau de complexidade.

### **VI.2.7 - Influência da Aglutinação de Classes a Posteriori**

A construção de modelos com classes aglutinadas a posteriori foi recomendada no capítulo III como meio de evitar o aparecimento de inconsistências entre as probabilidades a posteriori de diferentes modelos, devido às alterações introduzidas pelo processo de aglutinação na estrutura de dependências dos descritores. A figura (VI.31) mostra que em 3 de 4 situações foi possível obter ganho de eficiência com a implementação desta recomendação.

### **VI.2.8 - Avaliação do Desempenho de Modelos Construídos Segundo o Critério de Multi-Elegibilidade**

O modelo Anatomia Coronariana foi desdobrado em 3 modelos usando o critério de multi-elegibilidade, cujos desempenhos são apresentados na figura (VI.32). Considerando-se os resultados produzidos pelos modelos Coronária Direita, Descendente Anterior e Coronária Circunflexa em cada paciente do conjunto de teste, e comparando-os com a configuração real de lesões do paciente, conhecida através do cateterismo cardíaco, obtivemos uma eficiência global de 45,77% para a modalidade "O Banco de Dados como Universo" e 34,62% para a modalidade "O Banco de Dados como Amostra". A queda de desempenho verificada na primeira modalidade (69,01 para 45,77) recomenda a não utilização de modelos construídos sob o critério de multi-elegibilidade.

### **VI.2.9 - Credibilidade das Probabilidades a Posteriori**

Uma questão interessante corresponde a investigar até que ponto a probabilidade a posteriori da classe selecionada pelo modelo como decisão realmente expressa a probabilidade de o paciente pertencer a esta classe. Para estudar esta questão usamos os modelos da família da variável Anatomia Coronariana. Através dos resultados apresentados nas figuras (VI.33) a (VI.35) podemos constatar que na modalidade "O Banco de Dados como Universo" existe excelente aderência entre o percentual de diagnósticos corretos obtidos pelos modelos e a faixa da probabilidade a posteriori com que foi tomada a decisão. As figuras (VI.36) a (VI.38) revelam que na modalidade "O Banco de Dados como Amostra" existe uma perda considerável da aderência nos modelos Número de Vasos Lesados e Anatomia Coronariana, possivelmente devido à queda experimentada pelo número médio de casos de treinamento por classe. Os resultados obtidos sugerem, em suma, que

podemos acreditar nos valores, de probabilidades a posteriori fornecidos pelo modelo, em situações em que tenhamos estimativas exatas ou quase exatas das probabilidades condicionais (bancos de dados de treinamento suficientemente grandes).

#### **VI.2.10 - Influência do Valor da Probabilidade Condicional Mínima**

A adoção de  $p_{\min} = 10^{-6}$  (probabilidade condicional mínima) no teste dos modelos da figura (VI.28), como meio de contornar as anomalias discutidas no capítulo III, permitiu obter na modalidade "O Banco de Dados como Amostra" pequenas melhorias em 5 de 8 modelos e piora em 1 situação. Na modalidade "O Banco de Dados como Universo" não se verificou melhorias dada a inexistência de anomalias inerente a esta estratégia.

Uma questão interessante corresponde à determinação dos valores de  $p_{\min}$  que não prejudicam o desempenho dos modelos bayesianos de classificação diagnóstica. Os resultados do estudo da variação do desempenho do modelo Presença de Doença Coronariana em função de  $p_{\min}$ , apresentados nas figuras (VI.40) e (VI.41), sugerem que devemos adotar valores iguais ou menores que  $10^{-6}$ .

As figuras (VI.40) e (VI.41) mostram-nos também que a entropia média é um parâmetro muito mais sensível que a eficiência para detectar alterações no comportamento dos modelos bayesianos de classificação diagnóstica.

#### **VI.2.11 - Influência das Estimativas de Probabilidades a Priori sobre o Desempenho**

A utilização de estimativas de probabilidades a priori calculadas a partir do banco de dados em lugar de probabilidades a priori uniformes tende a produzir melhorias no desempe

nho, tanto maiores quanto maiores forem as divergências entre estes dois conjuntos de estimativas, conforme podemos observar nas figuras (VI.28) e (VI.15). Isto é natural pois no último caso temos um conjunto exato de estimativas da população sob teste. De qualquer forma as melhorias verificadas são geralmente pequenas, o que demonstra um alto grau de robustez dos modelos bayesianos de classificação diagnóstica com relação às estimativas de probabilidades a priori das doenças. Isto pode ser melhor observado nas figuras (VI.42) e (VI.43) que apresentam um amplo platô para a eficiência máxima do modelo Presença de Doença Coronariana no entorno do valor da probabilidade a priori de presença de lesões estimado do banco de dados. Mesmo para variações muito amplas da probabilidade a priori de presença de lesões (na faixa 0,2 a 0,98) o modelo manteve sua eficiência acima de 84% (figura (VI.42)). Portanto, estimativas subjetivas feitas por médicos para probabilidades a priori com relação a uma população específica podem ser usadas com baixo risco, o que vem a facilitar a transferência do sistema entre instituições.

#### **VI.2.12 - Desempenho em Função do Número de Descritores Utilizados**

Nas figuras (VI.44) a (VI.49) podemos estudar a variação do desempenho dos modelos da família da variável Anatomia Coronariana em função dos dados dos pacientes acumulados nas várias fases de uma seqüência natural de investigação clínica, a saber:

- Identificação
- História
- Exame físico
- Eletrocardiograma
- Laboratório
- Raios X de tórax
- Ergometria
- Ecocardiografia

Constatamos que o modelo Presença de Doença Coronariana já parte de um patamar muito elevado de eficiência: 83,10% (de fato os dados de identificação incluem o 2º, 3º e o 5º descritores mais eficientes para a variável Presença de Lesões Significativas). O modelo, conforme já comentamos, apresenta dificuldades em elevar o desempenho além deste nível, o que só consegue com a inclusão dos dados da Ecocardiografia. No entanto o gráfico da entropia média (VI.45) permite-nos constatar que a História e o Exame Físico mesmo sem reflexo na eficiência trazem quantidades sensíveis de informação propiciando diagnósticos mais seguros.

Para os modelos mais complexos como o de Número de Vasos Lesados e Anatomia Coronariana, ficam bem evidenciados os incrementos de eficiência e ganhos de informação trazidos pelos exames em cada etapa de investigação, o que demonstra a utilidade de prática de todos os exames considerados neste trabalho.

#### **VI.2.13 - Influência do Tamanho do Banco de Dados de Treinamento no Desempenho**

As figuras (VI.50) a (VI.53) ilustram a variação do desempenho dos modelos Presença de Doença Coronariana e Anatomia Coronariana em função do tamanho do banco de dados de treinamento. O conjunto de 90 casos de treinamento foi particionado aleatoriamente em 9 conjuntos de 10 elementos que foram acumulados ao longo de uma seqüência de etapas de treinamentos e testes. Os testes foram realizados sobre o conjunto padrão de casos de teste. Não conseguimos identificar, com este número limitado de casos, um padrão de crescimento no desempenho dos modelos, o que possivelmente viria a acontecer com amostras maiores.

### VI.3 - ESTUDO DAS ASSOCIAÇÕES ENTRE DESCRITORES DE PACIENTES COM SUSPEITA DE DOENÇA CORONARIANA

Sumarizamos o resultado desta etapa do trabalho nas estruturas de aglomerados de descritores interdependentes por classe das variáveis diagnósticas Presença de Doença Coronariana, Anatomia Coronariana e Função Ventricular, apresentadas nas figuras (VI.54) a (VI.65). As estruturas de aglomerados da variável Presença de Doença Coronariana para diversos níveis de significância são apresentadas sob a forma de dendrogramas por classe diagnóstica nas figuras (VI.54) e (VI.55). As estruturas de aglomerados para as classes das demais variáveis são apresentadas sob forma de listas.

Conforme podemos observar nestas figuras no nível de significância 0,1 trabalhamos com interações fracas entre os descritores, verificando-se poucos aglomerados (1 a 4). Na medida que diminui o valor do nível de significância, interações de maior intensidade são exigidas verificando-se um processo de fragmentação dos aglomerados iniciais. Passamos neste processo da situação extrema do modelo atuarial ou quase atuarial para o outro extremo representado pelo modelo com (ou quase com) pressuposição de independência dos descritores.

O nosso banco de dados é muito reduzido para determinação de interações entre descritores para variáveis com grande número de classes diagnósticas, como é o caso da Variável Anatomia Coronariana onde temos em média apenas 17,7 casos por classe. Para as classes "2" das variáveis Presença de Lesões Significativas e Função Ventricular os casos (111 e 67) com que contamos representam um número razoável. Nessas duas classes diagnósticas observamos interações fortes que tendem a ocorrer em ambas e são facilmente interpretáveis até pelo leigo em medicina, a saber:

(Capacidade Funcional, Atividade Física)

(Isquemia Lateral, Isquemia Anterior, Ergometria)

(Tonteira, Síncope)

(Presença de Arritmia Cardíaca, Arritmia-ECG, Palpitação)

#### VI.4 - ESTUDO DE MODELOS BAYESIANOS DE CLASSIFICAÇÃO DIAGNÓSTICA USANDO AGLOMERADOS DE DESCRITORES INTERDEPENDENTES

Com o objetivo de realizar este estudo construímos modelos para todas as estruturas de aglomerados de descritores apresentadas nas figuras (VI.54) a (VI.65). Passamos a discutir os principais resultados obtidos.

##### VI.4.1 - Avaliação do Desempenho na Estratégia "O Banco de Dados como Universo"

As figuras (VI.66) e (VI.67) apresentam a eficiência e a entropia média do modelo Presença de Doença Coronariana, treinado e testado na modalidade "O Banco de Dados como Universo" em função do nível de significância usado na definição dos aglomerados de descritores.

Constatamos conforme já previsto no capítulo III um desempenho crescente com o nível de significância (e por conseguinte com o tamanho dos aglomerados). O modelo consegue atingir a eficiência ótima de 100% para o nível de significância  $\alpha = 0,05$ , portanto bem antes de se transformar no modelo atuarial o que só ocorre para  $\alpha = 0,25$ . Naquela altura temos dois grandes aglomerados com 31 e 26 descritores nas classes 1 e 2 respectivamente. Na modalidade "O Banco de Dados como Universo" a obtenção de uma eficiência de 100% é quase uma fatalidade, pelo menos para situações com muitos descritores, aglomerados muito grandes e bancos de dados muito pequenos. (Quanto maior a rarefação dos casos nas células melhor, o que bem caracteriza a artificialidade desta modalidade).

#### VI.4.2 - Avaliação do Desempenho na Estratégia "O Banco de Dados como Amostra"

O desafio agora consiste em tentar, para a situação mais realista representada pela modalidade "O Banco de Dados como Amostra", obter melhorias no desempenho dos modelos bayesianos de classificação diagnóstica aproveitando a informação eventualmente existente em decorrência de interdependências dos descritores usados na área de doença coronariana. Trabalhamos, para estudar esta questão, com os modelos Presença de Doença Coronariana, Anatomia Coronariana e Função Ventricular sendo os resultados obtidos, apresentados nas figuras (VI.68) a (VI.74).

A figura (VI.74) estabelece uma comparação entre o desempenho dos modelos de classificação diagnóstica baseados na presuposição de independência com o desempenho dos modelos ótimos baseados no emprego de aglomerados de descritores interdependentes. Podemos constatar uma melhoria geral em todos os parâmetros de desempenho em favor dos modelos com aglomerados de descritores. A melhoria de eficiência foi maior nos modelos com maior número de classes. No modelo Anatomia Coronariana onde a diferenciação entre configurações de lesões nos vasos coronarianos é um problema de grande dificuldade, a eficiência saltou de 34,62% (não significativa ao nível de 10%) para um desempenho de 51,92%, significativo ao nível de 0,1%, o que pode configurar um modelo de aplicação prática imediata. Ressalte-se o fato de que este desempenho foi conseguido com um banco de dados de treinamento de tamanho muito reduzido (Número médio de casos por classe = 11,25), abrindo portanto um amplo potencial de melhoria através da simples ampliação deste banco de dados.

As curvas de eficiência obtidas apresentam aproximadamente a forma prevista no capítulo III podendo-se identificar um valor ótimo de nível de significância e em decorrência uma estrutura ótima de aglomerados de descritores. Obtivemos os seguintes valores ótimos de  $\alpha$ :



MODELO	$\alpha$ ótimo
Presença de Doença Coronariana	0,01
Anatomia Coronariana	0,025
Função Ventricular	0,0075 (0,0025)

As curvas de entropia média seguem de forma inversa as curvas de eficiência apresentando seu ponto de mínimo no valor  $\alpha$  ótimo (com uma pequena divergência no modelo Função Ventricular).

#### VI.4.3 - Estimativas das Probabilidades Condicionais de Configurações de Observações

- As figuras (VI.75) a (VI.86) apresentam as estimativas não nulas das probabilidades condicionais das configurações de observações diagnósticas dos modelos na presença de Doença Coronariana, Anatomia Coronariana e Função Ventricular utilizando as estruturas ótimas de aglomerados determinados a partir do nosso banco de dados de treinamento (90 casos).

A representação esparsa permite um ganho considerável de memória. Para melhor apreciarmos este fato tomemos como exemplo o modelo Presença de Doença Coronariana  $\alpha = 0,01$ . Este modelo necessita 19745 estimativas de probabilidades condicionais. Entretanto na representação esparsa precisamos de espaço de memória para acomodar apenas 112 estimativas não nulas se calibrarmos o modelo com os 90 casos do banco de dados de treinamento da modalidade "O Banco de Dados como Amostra" ou 121 estimativas se usarmos todos os 142 casos para calibração.

#### VI.4.4 - Avaliação do Desempenho de uma Bateria de Modelos com $\alpha$ Crescente

Construímos e testamos o desempenho de baterias de mo

delos com valores crescentes de  $\alpha$  no entorno de  $\alpha_{\text{ótimo}}$ . Neste procedimento uma decisão de classificação só é tomada quando a probabilidade a posteriori da melhor classe ultrapassar um valor limiar pré-estabelecido; em caso contrário passa-se para o modelo seguinte da bateria. Aplicando este procedimento ao conjunto de casos de teste, obtivemos para a bateria do modelo Anatomia Coronariana constituído pelos valores  $\alpha: \{0,0075/0,01/0,025/0,05\}$  com valor limiar 0,99.

- Eficiência = 59,62%
- Entropia Média = 0,131066
- $\chi^2 = 148,13$

Observe-se que a eficiência obtida é bem superior a do modelo individual ótimo. Este foi o melhor resultado conseguido. Nos demais modelos não obtivemos melhoria de desempenho com o procedimento de baterias de modelos.

A figura (VI.87) mostra as etapas de aplicação da bateria para Anatomia Coronariana em 3 casos que obtiveram classificação correta. No caso 84 observamos a convergência para a classe correta a partir de uma situação inicial de erro. No caso 3 uma tendência em uma direção errônea é subitamente corrigida. No caso 18 presenciamos a convergência para a classe correta com uma oscilação no caminho. O limiar não chega a ser ultrapassado.

#### IV.4.5 - Influência da Probabilidade Condicional Mínima

A recomendação de emprego de probabilidade condicional mínima  $p_{\text{min}} \neq 0$  é de importância vital nos modelos com aglomerados de descritores. Nestes modelos a explosão combinatória de estimativas ocorre justamente com as células vazias, resultando num número insuportável de anomalias.

As figuras (VI.88) e (VI.89) mostram uma avaliação

da influência do valor  $p_{\min}$  sobre o desempenho do modelo Presença de Doença Coronariana  $\alpha = 0,005$  treinado na modalidade "O Banco de Dados como Amostra". Repete-se aqui o resultado já obtido com a pressuposição de independência, segundo o qual valores de  $p_{\min}$  menores ou iguais a  $10^{-6}$  não interferem no desempenho do modelo, nem sob o ponto de vista de entropia média e muito menos sob o ponto de vista de eficiência.

#### VI.5 - PARÂMETROS PARA CONSTRUÇÃO DE UM GERADOR PRELIMINAR DE PLANOS DE INVESTIGAÇÃO

A título de contribuição para pesquisas futuras, apresentamos nas figuras (VI.90) e (VI.91) alguns parâmetros para construção de um gerador de planos de investigação diagnóstica para Presença de Doença Coronariana visando testar os programas desenvolvidos e estudar o comportamento do modelo proposto. A figura (VI.90) mostra os valores dos atributos custo, duração, sofrimento, morbidade e mortalidade, correspondentes aos diversos exames. Na figura (VI.91) temos os valores dos parâmetros  $\pi_{j\alpha}$  estimados através de construção de modelos de classificação diagnóstica para cada exame específico conforme sugerido no capítulo IV e os parâmetros  $q_{j\alpha}$  estimados diretamente a partir do banco de dados. Os modelos construídos para estimar os parâmetros  $\pi_{j\alpha}$  empregaram probabilidades a priori uniformes para as classes diagnósticas. Os valores obtidos para os parâmetros  $q_{j\alpha}$  demonstraram que os pacientes do grupo normal, pertencentes ao banco de dados, foram submetidos com maior frequência a exames complementares. As estimativas  $q_{j\alpha}$  certamente merecem ser revistas por um especialista.

\*\*\*\*\*  
 CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIABEL:  
 PRESENCA DE LESAO CORONARIANA  
 N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD G.LIB PORC.ERRO  
 DESCRITOR

COR	138	84.11	29.03	84.11	24.9794*Σ	2	16.10
SEXO	142	82.88	51.61	85.98	15.5258*Σ	1	23.94
IDADE	141	93.64	6.45	78.03	2.8616	2	25.53
HIPERTENSAO ARTERIAL	140	41.28	51.61	75.00	.4972	1	56.43
ATIVIDADE FISICA	134	41.90	44.83	73.33	4.5335	2	57.46
TABAGISMO	137	32.08	70.97	79.07	.1031	1	59.12
HISTORIA FAMILIAR DE CORONARIOPATIA	137	42.99	56.67	77.97	.0011	1	54.01
COLESTEROL >250 MG POR DL	92	25.00	85.00	85.71	.8685	1	61.96
TRIGLICERIDEOS >160 MG POR DL	67	59.26	30.77	78.05	.4387	1	46.27
GLICEMIA >110 MG POR DL	121	22.34	77.78	77.78	.0002	1	65.29
ACIDO URICO >7.5 MG POR DL	46	23.68	75.00	81.82	.0063	1	67.39
HIPERTROFIA VENTRICULAR	141	9.09	83.87	66.67	1.2602	1	74.47
HIPERTROFIA CONCENTRICA	68	3.85	75.00	33.33	6.8056*Σ	1	79.41

\*\*\*\*\*  
 \*Σ = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE 0.05

Figura VI.1 - Parâmetros dos descritores anteriores à doença com relação à Presença de Lesões Significativas.

```

*****
CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIABEL:
PRESENCA DE LESAO CORONARIANA
*****
DESCRITOR          N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD  G.LIB PORC.ERRO
-----
DOR PRECORDIAL      140  84.40  41.94  83.64  11.9450*S  2  25.00
INFARTO DO MIDCARDIO PRE
VIO                 140  36.70  93.55  95.24  10.5140*S  1  50.71
CAPACIDADE FUNCIONAL
                    131  86.14  13.33  76.99  7.6329  3  30.53
TONTTEIRA          140  22.02  70.97  72.73  .6591  1  67.14
SINCOPE            140  3.67  93.55  66.67  .4553  1  76.43
PALPITACAO        141  22.73  80.65  80.65  .1604  1  64.54
DISPNEIA AOS ESFORCOS
                    140  41.82  53.33  76.67  .2263  1  55.71
PRESENCA DE ARRITMIA CAR
DIALA              131  24.51  86.21  86.21  1.5045  1  61.83
PRESENCA DE B4    141  82.73  41.94  83.49  8.3846*S  1  26.24
RAIOS X DE TORAX  122  22.92  88.46  88.00  1.9184  3  63.11
ERGOMETRIA        44  87.50  16.67  73.68  .1287  1  31.82
*****
*S = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE 0.05

```

Figura VI.2 - Parâmetros dos descritores posteriores à doença com relação à Presença de Lesões Significativas.

```

*****
CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIABEL:
PRESENCA DE LESAO CORONARIANA
*****
DESCRITOR          N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD  G.LIB PORC.ERRO
-----
ISQUEMIA INFERIOR  141  6.36  90.32  70.00  .4030  1  75.18
ISQUEMIA LATERAL  141  10.91  96.77  92.31  1.7057  1  70.21
ISQUEMIA ANTERIOR 141  10.00  93.55  84.62  .3638  1  71.63
ISQUEMIA: OUTRAS  141  2.73  96.77  75.00  .0218  1  76.60
NECROSE INFERIOR  141  28.18  93.55  93.94  6.3704*§  1  57.45
NECROSE LATERAL  141  3.64  100.00  100.00  1.1602  1  75.18
NECROSE ANTERIOR  141  20.00  100.00  100.00  7.3462*§  1  62.41
NECROSE - OUTRAS  141  15.45  100.00  100.00  5.4477*§  1  65.96
ARRITMIAS        141  9.09  93.55  83.33  .2164  1  72.34
ALTER. INESPECIF. DA REPOL  141  10.91  80.65  66.67  1.5491  1  73.76
ARIZ. VENTRIC.    141  10.91  83.87  70.59  .6215  1  73.05
ECCARDIOGRAFIA   62  100.00  100.00  100.00  62.0000*§  1  .00

```

```

*****
*§ = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE 0.05
*****

```

Figura VI.2 (cont.) - Parâmetros dos descritores posteriores à doença com relação A Presença de Lesões Significativas.

\*\*\*\*\*  
 CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIABEL:  
 VENTRICULOGRAFIA

\*\*\*\*\*  
 DESCRITOR N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD G.LIB PORC.ERRO  
 -----

COR	138	82.54	12.00	48.60	3.7233	2	48.31
SEXO	142	86.57	34.67	54.21	8.5907*§	1	40.85
IDADE	141	90.91	4.00	45.45	2.8928	2	55.32
HIPERTENSAO ARTERIAL	140	37.88	52.70	41.67	1.2637	1	54.29
ATIVIDADE FISICA	134	49.21	59.15	51.67	1.6119	2	45.52
TABAGISMO	137	34.38	71.23	51.16	.4980	1	45.99
HISTORIA FAMILIAR DE COR ONARIOPATIA	137	41.54	55.56	45.76	.1177	1	51.09
COLESTEROL >250 MG POR D L	92	26.19	80.00	52.38	.4966	1	44.57
TRIGLICERIDEOS >160 MG P OR DL	67	60.00	37.84	43.90	.0326	1	52.24
GLICEMIA >110 MG POR DL	121	24.53	79.41	48.15	.2667	1	44.63
ACIDO URICO >7.5 ME POR DL	46	26.09	78.26	54.55	.1195	1	47.83
HIPERTROFIA VENTRICULAR	141	7.58	86.67	33.33	1.2242	1	50.35
HIPERTROFIA CONCENTRICA	68	.00	84.62	.00	4.8933*§	1	51.47

\*\*\*\*\*  
 \*§ = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE 0.05

Figura VI.3 - Parâmetros dos descritores anteriores à doença com relação à Função Ven-tricular.

\*\*\*\*\*  
 CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIABEL:  
 VENTRICULOGRAFIA  
 N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD G.LIB PORC.ERRO  
 DESCRITOR

DOR PRECORDIAL	140	85.07	27.40	51.82	5.2276	2	45.00
INFARTO DO MIOCARDIO PRE VIO	140	42.42	81.08	66.67	9.1783*S	1	37.14
CAPACIDADE FUNCIONAL	131	88.33	15.49	46.90	2.5899	3	51.15
TONTEIRA	140	22.73	75.68	45.45	.0494	1	49.29
SINCOPE	140	6.15	97.33	66.67	1.0323	1	45.00
PALPITACAO	141	28.79	84.00	61.29	3.3471	1	41.84
DISPNEIA AOS ESFORCOS	140	48.48	62.16	53.33	1.6148	1	44.29
PRESENCIA DE ARRITMIA CAR DIACA	131	35.00	88.73	72.41	10.6259*S	1	35.88
PRESENCIA DE B4	141	83.33	28.00	50.46	2.5702	1	46.10
RAIOS X DE TORAX	122	30.51	88.89	72.00	8.1062*S	3	39.34
ERGOMETRIA	44	92.31	16.13	31.58	.5536	1	61.36

\*\*\*\*\*  
 \*S = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE 0.05

Figura VI.4 - Parâmetros dos descritores posteriores à doença com relação à Função Ventricular.



```

*****
CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIABEL:
VENTRICULOGRAFIA
*****
DESCRITOR          N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD  G.LIB PORC.ERRO
-----
ISQUEMIA INFERIOR      141    4.55  90.67  30.00    1.2213    1    49.65
ISQUEMIA LATERAL      141   13.64  94.67  69.23    2.8916    1    43.26
ISQUEMIA ANTERIOR     141   15.15  96.00  76.92    5.2160*S   1    41.84
ISQUEMIA: OUTRAS      141    3.03  97.33  50.00    .0168     1    46.81
NECROSE INFERIOR      141   34.85  86.67  69.70    9.0652*S   1    37.59
NECROSE LATERAL      141    1.52  96.00  25.00    .7864     1    48.23
NECROSE ANTERIOR     141   24.24  92.00  72.73    7.0333*S   1    39.72
NECROSE - OUTRAS     141   18.18  93.33  70.59    4.3903*S   1    41.84
ARRITMIAS            141   13.64  96.00  75.00    4.1868*S   1    42.55
ALTER.INESPECIF.DA REPOL  141    6.06  81.33  22.22    5.0096*S   1    53.90
  ARIZ.VENTRIC.
  BLOQUEIOS
ECCOCARDIOGRAFIA     62  100.00  36.36  58.00   13.0764*S   1    33.87

```

```

*****
*S = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE 0.05
*****

```

Figura VI.4 (Cont.) - Parâmetros dos descritores posteriores à doença com relação à Função Ventricular.

DESCRIPTOR	INTERV. CONF. SENSIBILIDADE	INTERV. CONF. ESPECIFICIDADE
1	77.1855	45.0111
2	75.8757	69.2051
3	89.0746	15.0998
11	32.0414	69.2051
12	32.4671	62.9281
13	23.1895	86.9466
14	33.6102	74.3992
17	14.9979	100.0000
18	46.1538	55.8587
19	13.9200	93.4596
20	10.1666	100.0000
31	3.7185	96.8184
36	.0000	96.2176

INTERVALOS DE CONFIANCA DAS SENSIBILIDADE E ESPECIFICIDADE  
 DOS DESCRITORES COM RELACAO A VARIAVEL DIAGNOSTICA  
 PRESENCA DE LESAO CORONARIANA  
 (ALFA=0.05)

Figura VI.5 - Intervalos de confiança para sensibilidade e especificidade dos descritores anteriores à doença (com relação à Presença de Lesões Significativas).

```

*****
INTERVALOS DE CONFIANCA DAS SENSIBILIDADE E ESPECIFICIDADE
DOS DESCRITORES COM RELACAO A VARIAVEL DIAGNOSTICA
PRESENCA DE LESAO CORONARIANA
(ALFA=0.05)
*****
DESCRITOR   INTERV. CONF. SENSIBILIDADE   INTERV. CONF. ESPECIFICIDADE
4           77.5923      91.2150      24.5646      59.3063
5           27.6489      45.7456      84.9002      100.0000
6           79.3996      92.8776      1.1689      25.4977
7           14.2392      29.7975      54.9889      86.9466
8           .1400        7.1995      84.9002      100.0000
9           14.8957      30.5588      66.7373      94.5530
10          32.6002      51.0362      35.4809      71.1858
15          16.1620      32.8576      73.6565      98.7573
16          75.6630      89.7915      24.5646      59.3063
21          14.5090      31.3244      76.1809      100.0000
22          76.0412      98.9588      .0000        37.7529
*****

```

Figura VI.6 - Intervalos de confiança para sensibilidade e especificidade dos descritores posteriores à doença (com relação à Presença de Lesões Significativas).

```

*****
INTERVALOS DE CONFIANCA DAS SENSIBILIDADE E ESPECIFICIDADE
DOS DESCRITORES COM RELACAO A VARIÁVEL DIAGNOSTICA
PRESENÇA DE LESÃO CORONARIANA
(ALFA=0.05)
*****
DESCRITOR INTERV. CONF. SENSIBILIDADE INTERV. CONF. ESPECIFICIDADE
23 1.8019 10.9254 79.9149 100.0000
24 5.0831 16.7351 90.5544 100.0000
25 4.3936 15.6064 84.9002 100.0000
26 .0000 5.7711 90.5544 100.0000
27 19.7744 36.5892 84.9002 100.0000
28 .1381 7.1346 100.0000 100.0000
29 12.5249 27.4751 100.0000 100.0000
30 8.6994 22.2097 100.0000 100.0000
32 3.7185 14.4633 84.9002 100.0000
33 5.0831 16.7351 66.7373 94.5530
34 5.0831 16.7351 70.9235 96.8184
35 100.0000 100.0000 100.0000 100.0000
*****

```

Figura VI.6 - Continuação.

\*\*\*\*\*  
 INTERVALOS DE CONFIANÇA DAS SENSIBILIDADE E ESPECIFICIDADE  
 DOS DESCRITORES COM RELACAO A VARIÁVEL DIAGNOSTICA  
 VENTRICULOGRAFIA  
 (ALFA=0.05)

\*\*\*\*\*  
 DESCRITOR INTERV. CONF. SENSIBILIDADE INTERV. CONF. ESPECIFICIDADE  
 \*\*\*\*\*

1	73.1653	91.9141	4.6454	19.3546
2	78.4017	94.7326	23.8958	45.4375
3	83.9734	97.8448	.0000	8.4350
11	26.1757	49.5819	41.3271	64.0783
12	36.8611	61.5516	47.7211	70.5888
13	22.7385	46.0115	60.8484	81.6173
14	29.5584	53.5185	44.0777	67.0335
17	12.8933	39.4876	68.9126	91.0874
18	42.4692	77.5308	22.2106	53.4650
19	12.9447	36.1119	69.8011	89.0224
20	8.1411	44.0328	61.4037	95.1181
31	1.1918	13.9597	78.9732	94.3601
36	.0000	.0000	73.2916	95.9392

\*\*\*\*\*

Figura VI.7 - Intervalos de confiança para sensibilidade e especificidade dos descritores anteriores à doença (com relação à Função Ventricular).

```

*****
INTERVALOS DE CONFIANCA DAS SENSIBILIDADE E ESPECIFICIDADE
DOS DESCRITORES COM RELACAO A VARIAVEL DIAGNOSTICA
VENTRICULOGRAFIA
(ALFA=0.05)
*****
DESCRIPTOR INTERV. CONF. SENSIBILIDADE INTERV. CONF. ESPECIFICIDADE
4 76.5420 93.6072 17.1661 37.6284
5 30.5005 54.3479 72.1573 90.0048
6 80.2103 96.4563 7.0763 23.9096
7 12.6168 32.8377 65.9002 85.4512
8 .3116 11.9961 93.6871 100.0000
9 17.8643 39.7115 75.7029 92.2971
10 36.4274 60.5423 51.1121 73.2123
15 22.9310 47.0690 81.3774 96.0874
16 74.3421 92.3245 17.8382 38.1618
21 18.7593 42.2576 81.1284 96.6494
22 77.8223 100.0000 3.1816 29.0765

```

\*\*\*\*\*

Figura VI.8 - Intervalos de confiança para sensibilidade e especificidade dos descritores posteriores à doença (com relação à Função Ventricular).

```

*****
INTERVALOS DE CONFIANCA DAS SENSIBILIDADE E ESPECIFICIDADE
DOS DESCRITORES COM RELACAO A VARIAVEL DIAGNOSTICA
VENTRICULOGRAFIA
(ALFA=0.05)
*****
DESCRIPTOR INTERV. CONF. SENSIBILIDADE INTERV. CONF. ESPECIFICIDADE
23 .0000 9.5709 84.0830 97.2503
24 5.3570 21.9158 89.5813 99.7520
25 6.5012 23.8019 91.5650 100.0000
26 .0000 7.1660 93.6871 100.0000
27 23.3527 46.3443 78.9732 94.3601
28 .0000 4.4623 91.5650 100.0000
29 13.9033 34.5816 85.8601 98.1399
30 8.8766 27.4871 87.6879 98.9788
32 5.3570 21.9158 91.5650 100.0000
33 .3040 11.8172 72.5149 90.1518
34 3.1773 18.0348 78.9732 94.3601
35 100.0000 100.0000 19.9507 52.7765
*****

```

Figura VI.8 - Continuação.

Descritores Anteriores à Doença	$\chi^2$	G.L.	Signif.
- Cor	30,834	14	S
- Sexo	21,771	7	S
- Idade	10,508	14	
- Hipertensão Arterial	9,379	7	
- Atividade Física	10,390	14	
- Tabagismo	6,495	7	
- História Familiar de Coronariopatia	2,963	7	
- Colesterol 250mg/dl	4,070	7	
- Triglicerídeos 160mg/dl	6,417	7	
- Glicemia > 110mg/dl	3,631	7	
- Ácido Úrico > 7.5mg/dl	2,489	7	
- Hipertrofia Ventricular	4,714	7	
- Hipertrofia Concêntrica	8,956	7	

**Figura VI.9** - Testes de hipótese sobre associação de descritores anteriores com relação à Anatomia Coronariana ( $\alpha = 0,05$ ).





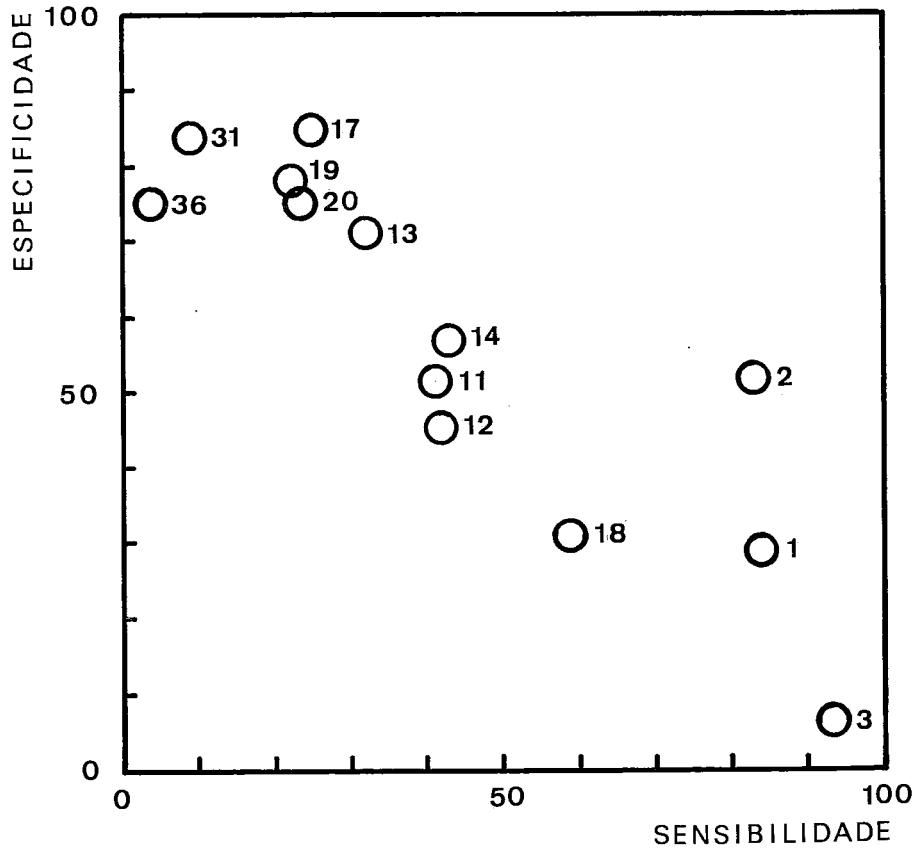


Figura VI.11 - Gráfico ROC dos descritores anteriores à doença com relação à Presença de Lesões Significativas.

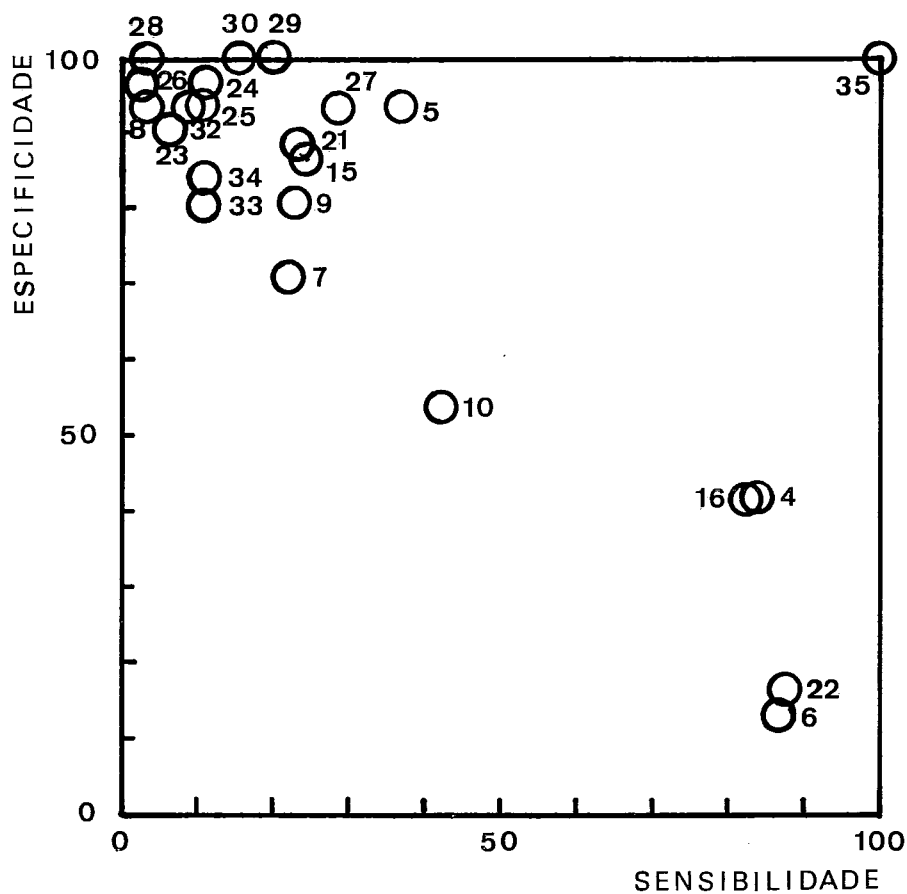


Figura VI.12 - Gráfico ROC dos descritores posteriores à doença com relação à Presença de Lesões Significativas.

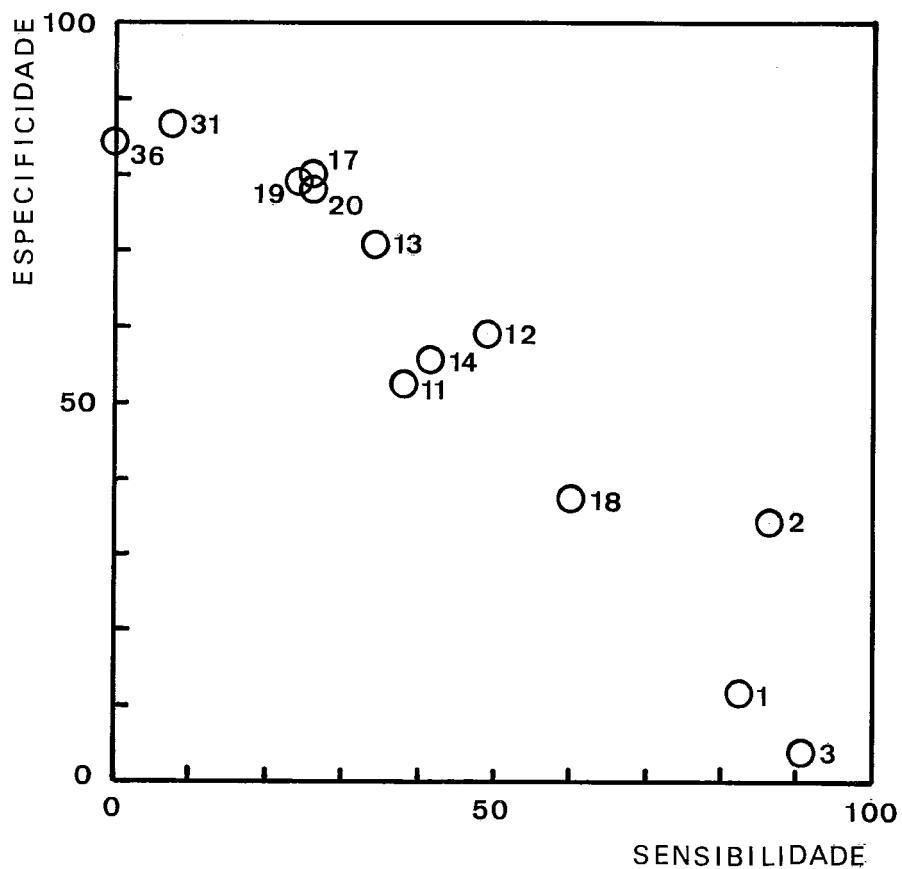


Figura VI.13 - Gráfico ROC dos descritores anteriores à doença com relação à Função Ventricular.

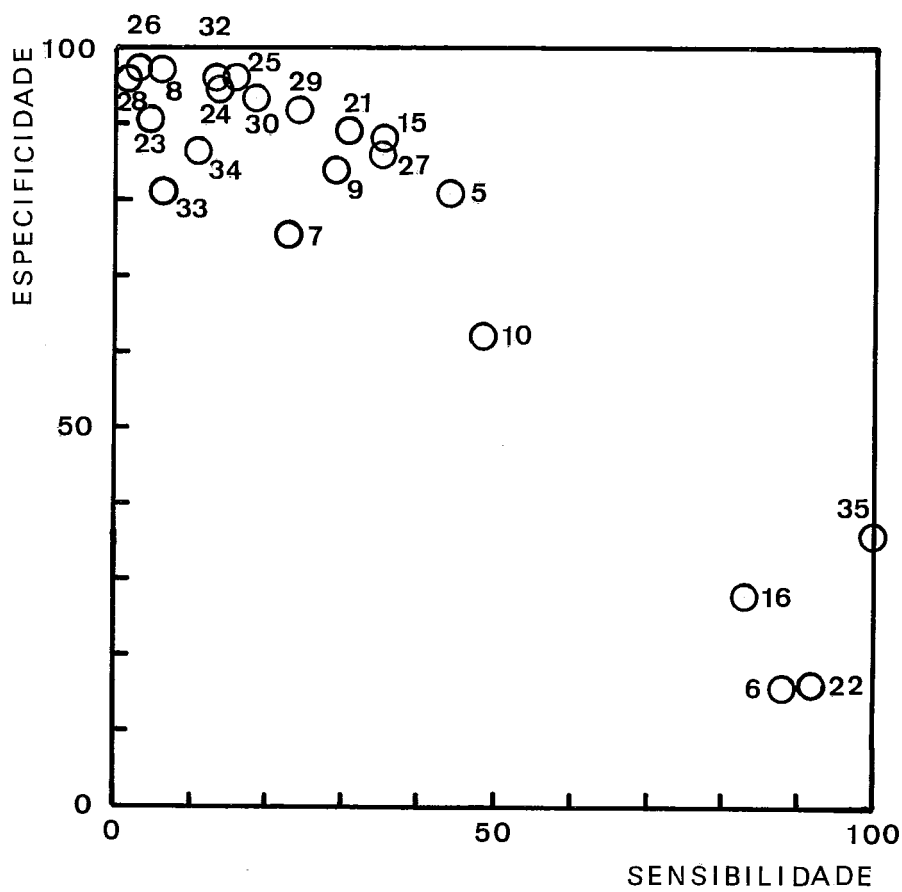


Figura VI.14 - Gráfico ROC dos descritores posteriores à doença com relação à Função Ventricular.

VARIÁVEL DIAGNÓSTICA	CÓDIGO DA CLASSE DIAGNÓSTICA								
	0	1	2	3	4	5	6	7	8
Anatomia Coronariana	-	0,21830	0,08460	0,13380	0,04220	0,09680	0,07040	0,14790	0,20420
Função Ventricular	-	0,52817	0,47183	-	-	-	-	-	-
Tipo de Tratamento	-	0,82203	0,17797	-	-	-	-	-	-
Presença de Doença Coronariana	-	0,21830	0,78170	-	-	-	-	-	-
Número de vasos Lesados	0,26056	0,31690	0,20423	0,21831	-	-	-	-	-
Presença de Lesões x Função Ventricular	-	0,21127	0,00704	0,31690	0,46479	-	-	-	-
Perfil do Paciente	-	0,21280	0,31910	0,46810	-	-	-	-	-
Coronária Direita	-	0,54225	0,45775	-	-	-	-	-	-
Descendente Anterior	-	0,54225	0,45775	-	-	-	-	-	-
Coronária Circunflexa	-	0,41549	0,58451	-	-	-	-	-	-

Figura VI.15 - Probabilidades a priori das classes diagnósticas estimadas a partir do banco de dados de pacientes.

VARIÁVEL DIAGNÓSTICA	CLASSE DIAGNÓSTICA				
	0	1	2	3	4
Presença de Doença Coronariana	-	0,5	0,5	-	-
Número de Vasos Lesados	0,5	0,2	0,2	0,1	-
Função Ventricular	-	0,7	0,3	-	-

Figura VI.16 - Probabilidades a priori de classes diagnósticas estimadas por Cardiologistas:

DESC V.OBS		CLASSES		DESC V.OBS		CLASSES	
1	2	1	2	1	2	1	2
1	1	.5483870968	.8411214953	8	1	.0645161290	.0366972477
1	2	.2903225806	.0186915888	9	0	.8064516129	.7727272727
1	3	.1612903226	.1401869159	9	1	.1935483871	.2272727273
2	1	.4838709677	.8288288288	10	0	.5333333333	.5818181818
2	2	.5161290323	.1711711712	10	1	.4666666667	.4181818182
3	1	.0645161290	.0636363636	11	0	.5161290323	.5871559633
3	2	.8709677419	.7454545455	11	1	.4838709677	.4128440367
3	3	.0645161290	.1909090909	12	1	.5517241379	.4190476190
4	0	.0000010000	.0000010000	12	2	.3448275862	.5428571429
4	1	.5806445806	.8440358532	12	3	.1034482759	.0380952381
4	2	.4193544194	.1559631468	13	1	.7096774194	.6792452830
5	0	.9354838710	.6330275229	13	2	.2903225806	.3207547170
5	1	.0645161290	.3669724771	14	0	.5666666667	.5700934579
6	1	.1333332000	.1386138614	14	1	.4333333333	.4299065421
6	2	.7999992000	.5544554455	15	0	.8620689655	.7549019608
6	3	.0666666000	.2871287129	15	1	.1379310345	.2450980392
6	4	.0000010000	.0198019802	16	0	.4193548387	.1727272727
7	0	.7096774194	.7798165138	16	1	.5806451613	.8272727273
7	1	.2903225806	.2201834862	17	0	.8500000000	.7500000000
8	0	.9354838710	.9633027523	17	1	.1500000000	.2500000000

Figura VI.17 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Presença de Lesões Significativas, estimadas na modalidade "O Banco de Dados como Universo".



DESC V.OBS		CLASSES		DESC V.OBS		CLASSES	
1	2	1	2	1	2	1	2
18	0	.3076923077	.4074074074	27	0	.9354838710	.7181818182
18	1	.6923076923	.5925925926	27	1	.0645161290	.2818181818
19	0	.7777777778	.7765957447	28	0	.9999990000	.9636363636
19	1	.2222222222	.2234042553	28	1	.0000010000	.0363636364
20	0	.7500000000	.7631578947	29	0	.9999990000	.8000000000
20	1	.2500000000	.2368421053	29	1	.0000010000	.2000000000
21	1	.8846136154	.7708333333	30	0	.9999990000	.8454545455
21	2	.1153843846	.1979166667	30	1	.0000010000	.1545454545
21	3	.0000010000	.0208333333	31	0	.8387096774	.9090909091
21	4	.0000010000	.0104166667	31	1	.1612903226	.0909090909
22	1	.1666666667	.1250000000	32	0	.9354838710	.9090909091
22	2	.8333333333	.8750000000	32	1	.0645161290	.0909090909
23	0	.9032258065	.9363636364	33	0	.8064516129	.8909090909
23	1	.0967741935	.0636363636	33	1	.1935483871	.1090909091
24	0	.9677419355	.8909090909	34	0	.8387096774	.8909090909
24	1	.0322580645	.1090909091	34	1	.1612903226	.1090909091
25	0	.9354838710	.9000000000	35	0	.9999990000	.0000010000
25	1	.0645161290	.1000000000	35	1	.0000010000	.9999990000
26	0	.9677419355	.9727272727	36	0	.7500000000	.9615384615
26	1	.0322580645	.0272727273	36	1	.2500000000	.0384615385

Fig. VI.17 (Continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Presença de Lesões Significativas, estimadas na modalidade "0 Banco de Dados como Universo".

DESC	V.OBS	CLASSES			0
		1	2	3	
1	1	.833332500000	.822222222222	.884614500000	.548387096774
1	2	.000001000000	.044444444444	.000001000000	.290322580645
1	3	.166666500000	.133333333333	.115384500000	.1612903222581
2	1	.702702702703	.888888888889	.896551724138	.483870967742
2	2	.297297297297	.111111111111	.103448275862	.516129032258
3	1	.081081081081	.045454545455	.068965517241	.064516129032
3	2	.729729729730	.772727272727	.724137931034	.870967741935
3	3	.189189189189	.181818181818	.206896551724	.064516129032
4	0	.000001000000	.000001000000	.000001000000	.000001000000
4	1	.810810000000	.863635500000	.857142000000	.580644580645
4	2	.189189000000	.136363500000	.142857000000	.419354419355
5	0	.702702702703	.613636363636	.571428571429	.935483870968
5	1	.297297297297	.386363636364	.428571428571	.064516129032
6	1	.142857142857	.150000000000	.115384500000	.133333200000
6	2	.628571428571	.550000000000	.461538000000	.799999200000
6	3	.200000000000	.275000000000	.423076500000	.066666600000
6	4	.028571428571	.025000000000	.000001000000	.000001000000
7	0	.702702702703	.844444444444	.777777777778	.709677419355
7	1	.297297297297	.155555555556	.222222222222	.290322580645
8	0	.972972972973	.954545454545	.964285714286	.935483870968
8	1	.027027027027	.045454545455	.035714285714	.064516129032
9	0	.864864864865	.822222222222	.571428571429	.806451612903
9	1	.135135135135	.177777777778	.428571428571	.193548387097
10	0	.702702702703	.533333333333	.500000000000	.533333333333
10	1	.297297297297	.466666666667	.500000000000	.466666666667
11	0	.638888888889	.600000000000	.500000000000	.516129032258
11	1	.361111111111	.400000000000	.500000000000	.483870967742

Figura VI.18 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Número de Vasos Lesados estimados na modalidade "0 Banco de Dados como Uni verso".

DESC	V.OBS	CLASSES		
		1	2	3
12	1	.400000000000	.395348837209	.481481000000
12	2	.571428571429	.534883720930	.518518000000
12	3	.028571428571	.069767441860	.000001000000
13	1	.722222222222	.613636363636	.730769230769
13	2	.277777777778	.386363636364	.269230769231
14	0	.638888888889	.534883720930	.535714285714
14	1	.361111111111	.465116279070	.464285714286
15	0	.852941176471	.682926829268	.740740740741
15	1	.147058823529	.317073170732	.259259259259
16	0	.108108108108	.250000000000	.137931034483
16	1	.891891891892	.750000000000	.862068965517
17	0	.791666666667	.769230769231	.681818181818
17	1	.208333333333	.230769230769	.318181818182
18	0	.444444444444	.380952380952	.400000000000
18	1	.555555555556	.619047619048	.600000000000
19	0	.852941176471	.750000000000	.708333333333
19	1	.147058823529	.250000000000	.291666666667
20	0	.857142857143	.666666666667	.750000000000
20	1	.142857142857	.333333333333	.250000000000
21	1	.878786121212	.756756756757	.653845500000
21	2	.121211878788	.189189189189	.307692000000
21	3	.000001000000	.027027027027	.038461500000
21	4	.000001000000	.027027027027	.000001000000
22	1	.181818181818	.166666666667	.000001000000
22	2	.818181818182	.833333333333	.999999000000
23	0	.918918918919	.909090909091	.999999000000
23	1	.081081081081	.090909090909	.000001000000

Figura VI.18 (Continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Número de Vasos Lesados estimados na modalidade "0 Banco de Dados como Universo".

DESC	V.OBS	CLASSES				0
		1	2	3	0	
24	0	.97297272973	.818181818182	.896551724138	.967741935484	
24	1	.027027027027	.181818181818	.103448275862	.032258064516	
25	0	.918918918919	.863636363636	.931034482759	.935483870968	
25	1	.081081081081	.136363636364	.068965517241	.064516129032	
26	0	.945945945946	.977272727273	.999999000000	.967741935484	
26	1	.054054054054	.022727272727	.000001000000	.032258064516	
27	0	.783783783784	.772727272727	.551724137931	.935483870968	
27	1	.216216216216	.227272727273	.448275862069	.064516129032	
28	0	.945945945946	.954545454545	.999999000000	.999999000000	
28	1	.054054054054	.045454545455	.000001000000	.000001000000	
29	0	.810810810811	.795454545455	.793103448276	.999999000000	
29	1	.189189189189	.204545454545	.206896551724	.000001000000	
30	0	.756756756757	.886363636364	.896551724138	.999999000000	
30	1	.243243243243	.113636363636	.103448275862	.000001000000	
31	0	.945945945946	.909090909091	.862068965517	.838709677419	
31	1	.054054054054	.090909090909	.137931034483	.161290322581	
32	0	.945945945946	.931818181818	.827586206897	.935483870968	
32	1	.054054054054	.068181818182	.172413793103	.064516129032	
33	0	.945945945946	.863636363636	.862068965517	.806451612903	
33	1	.054054054054	.136363636364	.137931034483	.193548387097	
34	0	.945945945946	.909090909091	.793103448276	.838709677419	
34	1	.054054054054	.090909090909	.206896551724	.161290322581	
35	0	.000001000000	.000001000000	.000001000000	.999999000000	
35	1	.999999000000	.999999000000	.999999000000	.000001000000	
36	0	.999999000000	.999999000000	.875000000000	.750000000000	
36	1	.000001000000	.000001000000	.125000000000	.250000000000	

Figura VI.18 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Número de Vasos Lesados estimadas na modalidade "0 Banco de Dados como Universo".

DESC V.OBS		CLASSES							
		1	2	3	4	5	6	7	8
1	1	.54839	.81818	.89474	.66667	.78571	.90000	.80952	.88461
1	2	.29032	.00000	.00000	.00000	.07143	.10000	.00000	.00000
1	3	.16129	.18182	.10526	.33333	.14286	.00000	.19048	.11538
2	1	.48387	.75000	.68421	.66667	.78571	.90000	.95238	.89655
2	2	.51613	.25000	.31579	.33333	.21429	.10000	.04762	.10345
3	1	.06452	.16667	.05263	.00000	.00000	.10000	.05000	.06897
3	2	.87097	.58333	.73684	1.00000	.85714	.80000	.70000	.72414
3	3	.06452	.25000	.21053	.00000	.14286	.10000	.25000	.20690
4	0	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
4	1	.58064	.75000	.89474	.66667	.84615	.90000	.85714	.85714
4	2	.41935	.25000	.10526	.33333	.15385	.10000	.14286	.14286
5	0	.93548	.66667	.68421	.83333	.71429	.30000	.70000	.57143
5	1	.06452	.33333	.31579	.16667	.28571	.70000	.30000	.42857
6	1	.13333	.00000	.11111	.50000	.07692	.25000	.15789	.11538
6	2	.80000	.72727	.72222	.16667	.46154	.50000	.63158	.46154
6	3	.06667	.18182	.16667	.33333	.46154	.25000	.15789	.42308
6	4	.00000	.09091	.00000	.00000	.00000	.00000	.05263	.00000
7	0	.70968	.58333	.78947	.66667	.64286	.80000	1.00000	.77778
7	1	.29032	.41667	.21053	.33333	.35714	.20000	.00000	.22222
8	0	.93548	.91667	1.00000	1.00000	.84615	1.00000	1.00000	.96429
8	1	.06452	.08333	.00000	.00000	.15385	.00000	.00000	.03571
9	0	.80645	.83333	.84211	1.00000	.78571	.90000	.80952	.57143
9	1	.19355	.16667	.15789	.00000	.21429	.10000	.19048	.42857
10	0	.53333	.75000	.68421	.66667	.42857	.60000	.57143	.50000
10	1	.46667	.25000	.31579	.33333	.57143	.40000	.42857	.50000
11	0	.51613	.75000	.72222	.16667	.64286	.70000	.52381	.50000
11	1	.48387	.25000	.27778	.83333	.35714	.30000	.47619	.50000

Figura VI.19 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Anatomia Coronariana estimadas na modalidade "0 Banco de Dados como Universo".

CLASSES

DESC	V.OBS	1	2	3	4	5	6	7	8
12	1	.55172	.54545	.33333	.33333	.42857	.40000	.36842	.48148
12	2	.34483	.45455	.61111	.66667	.57143	.50000	.52632	.51852
12	3	.10345	.00000	.05556	.00000	.00000	.10000	.10526	.00000
13	1	.70968	.72727	.78947	.50000	.61538	.40000	.71429	.73077
13	2	.29032	.27273	.21053	.50000	.38462	.60000	.28571	.26923
14	0	.56667	.66667	.63158	.60000	.64286	.60000	.42105	.53571
14	1	.43333	.33333	.36842	.40000	.35714	.40000	.57895	.46429
15	0	.86207	.63636	.94118	1.00000	.69231	.42857	.76190	.74074
15	1	.13793	.36364	.05882	.00000	.30769	.57143	.23810	.25926
16	0	.41935	.16667	.10526	.00000	.23077	.10000	.33333	.13793
16	1	.58065	.83333	.89474	1.00000	.76923	.90000	.66667	.86207
17	0	.85000	.90000	.75000	.66667	.60000	.83333	.80000	.68182
17	1	.15000	.10000	.25000	.33333	.40000	.16667	.20000	.31818
18	0	.30769	.50000	.42857	.33333	.40000	.80000	.18182	.40000
18	1	.69231	.50000	.57143	.66667	.60000	.20000	.81818	.60000
19	0	.77778	.75000	.93750	.83333	.76923	.66667	.76471	.70833
19	1	.22222	.25000	.06250	.16667	.23077	.33333	.23529	.29167
20	0	.75000	.83333	1.00000	.75000	.50000	.75000	.66667	.75000
20	1	.25000	.16667	.00000	.25000	.50000	.25000	.33333	.25000
21	1	.88461	.81818	.88235	1.00000	.92857	.75000	.60000	.65385
21	2	.11538	.18182	.11765	.00000	.00000	.25000	.33333	.30769
21	3	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.03846
21	4	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
22	1	.16667	.00000	.28571	.00000	.20000	.00000	.20000	.00000
22	2	.83333	1.00000	.71429	1.00000	.80000	1.00000	.80000	1.00000
23	0	.90323	.83333	.94737	1.00000	.92857	.80000	.95000	1.00000
23	1	.09677	.16667	.05263	.00000	.07143	.20000	.05000	.00000

Figura VI.19 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Anatomia Coronariana estimadas na modalidade "0 Banco de Dados como Universo".

CLASSES

DESC	V.OBS	1	2	3	4	5	6	7	8
24	0	.96774	1.00000	.94737	1.00000	.78571	.80000	.85000	.89655
24	1	.03226	.00000	.05263	.00000	.21429	.20000	.15000	.10345
25	0	.93548	1.00000	.89474	.83333	.78571	1.00000	.85000	.93103
25	1	.06452	.00000	.10526	.16667	.21429	.00000	.15000	.06897
26	0	.96774	.91667	.94737	1.00000	.92857	1.00000	1.00000	1.00000
26	1	.03226	.08333	.05263	.00000	.07143	.00000	.00000	.00000
27	0	.93548	.66667	.84211	.83333	.85714	.40000	.90000	.5172
27	1	.06452	.33333	.15789	.16667	.14286	.60000	.10000	.4828
28	0	1.00000	1.00000	.94737	.83333	.92857	1.00000	.95000	1.00000
28	1	.00000	.00000	.05263	.16667	.07143	.00000	.05000	.00000
29	0	1.00000	1.00000	.63158	1.00000	1.00000	.90000	.60000	.79310
29	1	.00000	.00000	.36842	.00000	.00000	.10000	.40000	.20690
30	0	1.00000	.91667	.63158	.83333	.92857	.90000	.85000	.89655
30	1	.00000	.08333	.36842	.16667	.07143	.10000	.15000	.10345
31	0	.83871	.91667	1.00000	.83333	.85714	.90000	.95000	.86207
31	1	.16129	.08333	.00000	.16667	.14286	.10000	.05000	.13793
32	0	.93548	.91667	.94737	1.00000	.92857	1.00000	.90000	.82759
32	1	.06452	.08333	.05263	.00000	.07143	.00000	.10000	.17241
33	0	.80645	1.00000	.94737	.83333	.78571	.80000	.95000	.86207
33	1	.19355	.00000	.05263	.16667	.21429	.20000	.05000	.13793
34	0	.83871	.91667	1.00000	.83333	1.00000	.90000	.85000	.79310
34	1	.16129	.08333	.00000	.16667	.00000	.10000	.15000	.20690
35	0	1.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
35	1	.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
36	0	.75000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	.87500
36	1	.25000	.00000	.00000	.00000	.00000	.00000	.00000	.12500

Figura VI.19 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Anatomia Coronariana estimadas na modalidade "O Banco de Dados como Universo".

CLASSES		CLASSES			
DESC V.OBS	1	2	DESC V.OBS	1	2
1	.7333333333	.8253968254	8	.0266666667	.0615384615
1	.1200000000	.0317460317	9	.0400000000	.7121212121
1	.1466666667	.1428571429	9	.1600000000	.2878787879
2	.6533333333	.8656716418	10	.6216216216	.5151515152
2	.3466666667	.1343283582	10	.3783783784	.4848484848
3	.0400000000	.0709090909	11	.5270270270	.6212121212
3	.8266666667	.7121212121	11	.4729729730	.3787878788
3	.1333333333	.1969696970	12	.4084507042	.4920634921
4	.0000010000	.0000010000	12	.5211267606	.4761904762
4	.7260266712	.8507454179	12	.0704225352	.0317460317
4	.2739723288	.1492535821	13	.7123287671	.6562500000
5	.8108108108	.5757575758	13	.2876712329	.3437500000
5	.1891891892	.4242424242	14	.5555555556	.5846153846
6	.1549295775	.1166666667	14	.4444444444	.4153846154
6	.6478873239	.5666666667	15	.8873239437	.6500000000
6	.1830985915	.3000000000	15	.1126760563	.3500000000
6	.0140845070	.0166666667	16	.2800000000	.1666666667
7	.7567567568	.7727272727	16	.7200000000	.8333333333
7	.2432432432	.2272727273	17	.8000000000	.7380952381
8	.9733333333	.9384615385	17	.2000000000	.2619047619

Figura VI.20 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Função Ventricular, estimadas na modalidade "0 Banco de Dados como Univer<sub>so</sub>".





DESC V.OBS		CLASSES		DESC V.OBS		CLASSES	
1	2	1	2	1	2	1	2
1	.7578947368	1	.8095230000	10	0	.6041666667	.4000000000
1	.0947368421	2	.0000010000	10	1	.3958333333	.6000000000
1	.1473684211	3	.1904760000	11	0	.5937500000	.4285714286
2	.7422680412	1	.7142857143	11	1	.4062500000	.5714285714
2	.2577319588	2	.2857142857	12	1	.4782608696	.4736842105
3	.0625000000	1	.0000010000	12	2	.4673913043	.4736842105
3	.7812500000	2	.8571420000	12	3	.0543478261	.0526315789
3	.1562500000	3	.1428570000	13	1	.6774193548	.7142857143
4	.0000010000	0	.0000010000	13	2	.3225806452	.2857142857
4	.7835043711	1	.8947359474	14	0	.5212765957	.7142857143
4	.2164946289	2	.1052630526	14	1	.4787234043	.2857142857
5	.6947368421	0	.6666666667	15	0	.7415730337	.7500000000
5	.3052631579	1	.3333333333	15	1	.2584269663	.2500000000
6	.1136363636	1	.1904760000	16	0	.2395833333	.1904761905
6	.6136363636	2	.6666660000	16	1	.7604166667	.8095238095
6	.2500000000	3	.1428570000	17	0	.8208955224	.5384615385
6	.0227272727	4	.0000010000	17	1	.1791044776	.4615384615
7	.7500000000	0	.7500000000	18	0	.3571428571	.6000000000
7	.2500000000	1	.2500000000	18	1	.6428571429	.4000000000
8	.9687500000	0	.9500000000	19	0	.7619047619	.7894736842
8	.0312500000	1	.0500000000	19	1	.2380952381	.2105263158
9	.8041237113	0	.7500000000	20	0	.7352941176	.6666666667
9	.1958762887	1	.2500000000	20	1	.2647058824	.3333333333

Figura VI.21 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Tipo de Tratamento estimadas na modalidade "0 Banco de Dados como Univer<sub>SO</sub>".

CLASSES		CLASSES				
DESC V.OBS	1	DESC V.OBS	2			
21	.8170723537	.7368406316	31	.0937500000	1	.1904761905
21	.15853364268	.2631573684	32	.9166666667	0	.8571428571
21	.0243902195	.0000010000	32	.0833333333	1	.1428571429
21	.0000010000	.0000010000	33	.8541666667	0	.9047619048
22	.1612903226	.0000010000	33	.1458333333	1	.0952380952
22	.8387097774	.9999990000	34	.8541666667	0	.8571428571
23	.9166666667	.9523809524	34	.1458333333	1	.1428571429
23	.0833333333	.0476190476	35	.2857142857	0	.0000010000
24	.8958333333	.9047619048	35	.7142857143	1	.9999990000
24	.1041666667	.0952380952	36	.8750000000	0	.9999990000
25	.9270833333	.7619047619	36	.1250000000	1	.0000010000
25	.0729166667	.2380952381	38	.2989690722	1	.0000010000
26	.9583333333	.9999990000	38	.1030927835	2	.0000010000
26	.0416666667	.0000010000	38	.1237113402	3	.1428567143
27	.7812500000	.8095238095	38	.0412371134	4	.0000010000
27	.2187500000	.1904761905	38	.0721649485	5	.1428567143
28	.9895833333	.9047619048	38	.0721649485	6	.0952378095
28	.0104166667	.0952380952	38	.1340206186	7	.1904756190
29	.8750000000	.8571428571	38	.1546391753	8	.4285701429
29	.1250000000	.1428571429	40	.6185567010	1	.3809523810
30	.8958333333	.9523809524	40	.3814432990	2	.6190476190
30	.1041666667	.0476190476				
31	.9062500000	.8095238095				

Figura VI.21 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Tipo de Tratamento estimadas na modalidade "O Banco de Dados como Universo".

CLASSES		CLASSES			
DESC V.OBS	1	2	DESC V.OBS		
1	.6666666667	.8484848485	8	.0952380952	.0441176471
1	.2380952381	.0151515152	9	.7619047619	.8260869565
1	.0952380952	.1363636364	9	.2380952381	.1739130435
2	.4285714286	.7826086957	10	.5000000000	.5942028986
2	.5714285714	.2173913043	10	.5000000000	.4057971014
3	.0476190476	.0579710145	11	.4761904762	.5820895522
3	.9047619048	.7391304348	11	.5238095238	.4179104478
3	.0476190476	.2028985507	12	.6315789474	.3846153846
4	.0000010000	.0000010000	12	.2631578947	.5692307692
4	.5238090000	.8260861304	12	.1052631579	.0461538462
4	.4761900000	.1739128696	13	.7142857143	.6716417910
5	.9523809524	.6666666667	13	.2857142857	.3283582090
5	.0476190476	.3333333333	14	.6000000000	.6363636364
6	.1499997000	.1774193548	14	.4000000000	.3636363636
6	.8499983000	.5322580645	15	.8500000000	.7936507937
6	.0000010000	.2741935484	15	.1500000000	.2063492063
6	.0000010000	.0161290323	16	.5238095238	.1617647059
7	.8095238095	.7681159420	16	.4761904762	.8382352941
7	.1904761905	.2318840580	17	.9230769231	.7777777778
8	.9047619048	.9558823529	17	.0769230769	.2222222222

Figura VI.22 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Presença de Lesões Significativas estimadas na modalidade "O Banco de Dados como Amostra".

CLASSES		CLASSES	
DESC	V.OBS	1	2
18	0	.1428571429	.3611111111
18	1	.8571428571	.6388888889
19	0	.6842105263	.7666666667
19	1	.3157894737	.2333333333
20	0	.6666666667	.7600000000
20	1	.3333333333	.2400000000
21	1	.8333316667	.7966101695
21	2	.1666663333	.1694915254
21	3	.0000010000	.0169491525
21	4	.0000010000	.0169491525
22	1	.2000000000	.1600000000
22	2	.8000000000	.8400000000
23	0	.9523809524	.9130434783
23	1	.0476190476	.0869565217
24	0	.9523809524	.8985507246
24	1	.0476190476	.1014492754
25	0	.9523809524	.9130434783
25	1	.0476190476	.0869565217
26	0	.9999990000	.9710144928
26	1	.0000010000	.0289855072
27	0		
27	1		
28	0		
28	1		
29	0		
29	1		
30	0		
30	1		
31	0		
31	1		
32	0		
32	1		
33	0		
33	1		
34	0		
34	1		
35	0		
35	1		
36	0		
36	1		

Figura VI.22 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Presença de Lesões Significativas estimadas na modalidade "0 Banco de Dados como Amostra".

DESC V.ORS	CLASSES			
	1	2	3	0
1	.769230000000	.920000000000	.866665800000	.666666666667
1	.000001000000	.040000000000	.000001000000	.238095238095
1	.230769000000	.040000000000	.133333200000	.095238095238
2	.666666666667	.840000000000	.882352941176	.428571428571
2	.333333333333	.160000000000	.117647058824	.571428571429
3	.074074074074	.000001000000	.117647058824	.047619047619
3	.777777777778	.759999240000	.647058823529	.904761904762
3	.148148148148	.239999760000	.235294117647	.047619047619
4	.000001000000	.000001000000	.000001000000	.000001000000
4	.777777000000	.879999120000	.823528588235	.523809000000
4	.222222000000	.119999880000	.176470411765	.476190000000
5	.740740740741	.600000000000	.647058823529	.952380952381
5	.259259259259	.400000000000	.352941176471	.047619047619
6	.200000000000	.181818000000	.133333200000	.149999700000
6	.520000000000	.590908500000	.466666200000	.849998300000
6	.240000000000	.227272500000	.399999600000	.000001000000
6	.040000000000	.000001000000	.000001000000	.000001000000
7	.703703703704	.800000000000	.823529411765	.809523809524
7	.296296296296	.200000000000	.176470588235	.190476190476
8	.962962962963	.958333333333	.941176470588	.904761904762
8	.037037037037	.041666666667	.058823529412	.095238095238
9	.925925925926	.880000000000	.588235294118	.761904761905
9	.074074074074	.120000000000	.411764705882	.238095238095
10	.777777777778	.480000000000	.470588235294	.500000000000
10	.222222222222	.520000000000	.529411764706	.500000000000
11	.653846153846	.560000000000	.500000000000	.476190476190
11	.346153846154	.440000000000	.500000000000	.523809523810

Figura VI.23 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Número de Vasos Lesados estimadas na modalidade "0 Banco de Dados como Amostra".

DESC	V.OBS	CLASSES			
		1	2	3	0
12	1	.4000000000000	.3750000000000	.374999625000	.631578947368
12	2	.5600000000000	.5416666666667	.624999375000	.263157894737
12	3	.0400000000000	.0833333333333	.000001000000	.105263157895
13	1	.769230769231	.5416666666667	.705882352941	.714285714286
13	2	.230769230769	.4583333333333	.294117647059	.285714285714
14	0	.653846153846	.608695652174	.647058823529	.600000000000
14	1	.346153846154	.391304347826	.352941176471	.400000000000
15	0	.8800000000000	.681818181818	.812500000000	.850000000000
15	1	.1200000000000	.318181818182	.187500000000	.150000000000
16	0	.074074074074	.250000000000	.176470588235	.523809523810
16	1	.925925925926	.750000000000	.823529411765	.476190476190
17	0	.7500000000000	.875000000000	.692307692308	.923076923077
17	1	.2500000000000	.125000000000	.307692307692	.076923076923
18	0	.384615384615	.357142857143	.333333333333	.142857142857
18	1	.615384615385	.642857142857	.666666666667	.857142857143
19	0	.8400000000000	.666666666667	.785714285714	.684210526316
19	1	.1600000000000	.333333333333	.214285714286	.315789473684
20	0	.777777777778	.714285714286	.777777777778	.666666666667
20	1	.222222222222	.285714285714	.222222222222	.333333333333
21	1	.916664833333	.809523000000	.571428000000	.833331666667
21	2	.083333166667	.142857000000	.357142500000	.166666333333
21	3	.000001000000	.000001000000	.071428500000	.000001000000
21	4	.000001000000	.047619000000	.000001000000	.000001000000
22	1	.181818181818	.250000000000	.000001000000	.200000000000
22	2	.818181818182	.750000000000	.999999000000	.800000000000
23	0	.888888888889	.880000000000	.999999000000	.952380952381
23	1	.111111111111	.120000000000	.000001000000	.047619047619

Figura VI.23 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Número de Vasos Lesados estimadas na modalidade "0 Banco de Dados como Amostra".

DESC	V.OBS	CLASSES			
		1	2	3	0
24	0	.962962962963	.800000000000	.941176470588	.952380952381
24	1	.037037037037	.200000000000	.058823529412	.047619047619
25	0	.888888888889	.920000000000	.941176470588	.952380952381
25	1	.111111111111	.080000000000	.058823529412	.047619047619
26	0	.962962962963	.960000000000	.999999000000	.999999000000
26	1	.037037037037	.040000000000	.000001000000	.000001000000
27	0	.814814814815	.760000000000	.588235294118	.904761904762
27	1	.185185185185	.240000000000	.411764705882	.095238095238
28	0	.962962962963	.999999000000	.999999000000	.999999000000
28	1	.037037037037	.000001000000	.000001000000	.000001000000
29	0	.814814814815	.840000000000	.764705882353	.999999000000
29	1	.185185185185	.160000000000	.235294117647	.000001000000
30	0	.777777777778	.960000000000	.882352941176	.999999000000
30	1	.222222222222	.040000000000	.117647058824	.000001000000
31	0	.925925925926	.920000000000	.823529411765	.809523809524
31	1	.074074074074	.080000000000	.176470588235	.190476190476
32	0	.962962962963	.920000000000	.764705882353	.904761904762
32	1	.037037037037	.080000000000	.235294117647	.095238095238
33	0	.925925925926	.800000000000	.823529411765	.714285714286
33	1	.074074074074	.200000000000	.176470588235	.285714285714
34	0	.999999000000	.960000000000	.823529411765	.857142857143
34	1	.000001000000	.040000000000	.176470588235	.142857142857
35	0	.000001000000	.000001000000	.000001000000	.999999000000
35	1	.999999000000	.999999000000	.999999000000	.000001000000
36	0	.999999000000	.999999000000	.999999000000	.636363636364
36	1	.000001000000	.000001000000	.000001000000	.363636363636

Figura VI.23 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Número de Vasos Lesados estimadas na modalidade "O Banco de Dados como Amostra".



## CLASSES

DESC	V.OBS	1	2	3	4	5	6	7	8
1	1	.66667	.66667	.86667	.60000	.90000	.83333	1.00000	.86667
1	2	.23810	.00000	.00000	.00000	.00000	.16667	.00000	.00000
1	3	.09524	.33333	.13333	.40000	.10000	.00000	.00000	.13333
2	1	.42857	.71429	.66667	.60000	.70000	1.00000	.88889	.88235
2	2	.57143	.28571	.33333	.40000	.30000	.00000	.11111	.11765
3	1	.04762	.14286	.06667	.00000	.00000	.00000	.00000	.11765
3	2	.90476	.57143	.80000	1.00000	.80000	.83333	.66667	.64706
3	3	.04762	.28571	.13333	.00000	.20000	.16667	.33333	.23529
4	0	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
4	1	.52381	.71428	.86667	.60000	.90000	1.00000	.77778	.82353
4	2	.47619	.28571	.13333	.40000	.10000	.00000	.22222	.17647
5	0	.95238	.85714	.66667	.80000	.70000	.33333	.66667	.64706
5	1	.04762	.14286	.33333	.20000	.30000	.66667	.33333	.35294
6	1	.15000	.00000	.14286	.60000	.00000	.50000	.22222	.13333
6	2	.85000	.66667	.64286	.00000	.55555	.50000	.66667	.46667
6	3	.00000	.16667	.21429	.40000	.44444	.00000	.11111	.40000
6	4	.00000	.16667	.00000	.00000	.00000	.00000	.00000	.00000
7	0	.80952	.57143	.80000	.60000	.60000	.83333	1.00000	.82353
7	1	.19048	.42857	.20000	.40000	.40000	.16667	.00000	.17647
8	0	.90476	.85714	1.00000	1.00000	.88889	1.00000	1.00000	.94118
8	1	.09524	.14286	.00000	.00000	.11111	.00000	.00000	.05882
9	0	.76190	1.00000	.86667	1.00000	.80000	1.00000	.88889	.58824
9	1	.23810	.00000	.13333	.00000	.20000	.00000	.11111	.41176
10	0	.50000	.85714	.73333	.80000	.20000	.50000	.77778	.47059
10	1	.50000	.14286	.26667	.20000	.80000	.50000	.22222	.52941
11	0	.47619	.71429	.78571	.20000	.60000	.66667	.44444	.50000
11	1	.52381	.28571	.21429	.80000	.40000	.33333	.55556	.50000

Figura VI.24 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Anatomia Coronariana estimadas na modalidade "O Banco de Dados como Amostra".

## CLASSES

DESC	V.OBS	1	2	3	4	5	6	7	8
12	1	.63158	.50000	.35714	.40000	.50000	.33333	.25000	.37500
12	2	.26316	.50000	.57143	.60000	.50000	.50000	.62500	.62500
12	3	.10526	.00000	.07143	.00000	.00000	.16667	.12500	.00000
13	1	.71429	.83333	.80000	.60000	.66667	.16667	.66667	.70588
13	2	.28571	.16667	.20000	.40000	.33333	.83333	.33333	.29412
14	0	.60000	.71429	.60000	.75000	.60000	.66667	.57143	.64706
14	1	.40000	.28571	.40000	.25000	.40000	.33333	.42857	.35294
15	0	.85000	.71429	.92308	1.00000	.66667	.25000	.88889	.81250
15	1	.15000	.28571	.07692	.00000	.33333	.75000	.11111	.18750
16	0	.52381	.14286	.06667	.00000	.22222	.16667	.33333	.17647
16	1	.47619	.85714	.93333	1.00000	.77778	.83333	.66667	.82353
17	0	.92308	.80000	.83333	.60000	.75000	1.00000	.87500	.69231
17	1	.07692	.20000	.16667	.40000	.25000	.00000	.12500	.30769
18	0	.14286	.50000	.33333	.33333	.25000	.66667	.28571	.33333
18	1	.85714	.50000	.66667	.66667	.75000	.33333	.71429	.66667
19	0	.68421	.71429	.92308	.80000	.66667	.50000	.75000	.78571
19	1	.31579	.28571	.07692	.20000	.33333	.50000	.25000	.21429
20	0	.66667	.66667	1.00000	.66667	1.00000	.50000	.75000	.77778
20	1	.33333	.33333	.00000	.33333	.00000	.50000	.25000	.22222
21	1	.83333	.85714	.92308	1.00000	1.00000	.80000	.50000	.57143
21	2	.16667	.14286	.07692	.00000	.00000	.20000	.33333	.35714
21	3	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.07143
21	4	.00000	.00000	.00000	.00000	.00000	.00000	.16667	.00000
22	1	.20000	.00000	.28571	.00000	.25000	.00000	.50000	.00000
22	2	.80000	1.00000	.71429	1.00000	.75000	1.00000	.50000	1.00000
23	0	.95238	.71429	.93333	1.00000	.90000	.66667	1.00000	1.00000
23	1	.04762	.28571	.06667	.00000	.10000	.33333	.00000	.00000

Figura VI.24 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Anatomia Coronariana estimadas na modalidade "0 Banco de Dados como Amostra".

CLASSES

DESC	V.OBS	1	2	3	4	5	6	7	8
24	0	.95238	1.00000	.93333	1.00000	.80000	.66667	.88889	.94118
24	1	.04762	.00000	.06667	.00000	.20000	.33333	.11111	.05882
25	0	.95238	1.00000	.86667	.80000	.80000	1.00000	1.00000	.94118
25	1	.04762	.00000	.13333	.20000	.20000	.00000	.00000	.05882
26	0	1.00000	.85714	1.00000	1.00000	.90000	1.00000	1.00000	1.00000
26	1	.00000	.14286	.00000	.00000	.10000	.00000	.00000	.00000
27	0	.90476	.71429	.80000	1.00000	.90000	.50000	.77778	.58824
27	1	.09524	.28571	.20000	.00000	.10000	.50000	.22222	.41176
28	0	1.00000	1.00000	.93333	1.00000	1.00000	1.00000	1.00000	1.00000
28	1	.00000	.00000	.06667	.00000	.00000	.00000	.00000	.00000
29	0	1.00000	1.00000	.66667	1.00000	1.00000	1.00000	.55556	.76471
29	1	.00000	.00000	.33333	.00000	.00000	.00000	.44444	.23529
30	0	1.00000	1.00000	.66667	.80000	1.00000	1.00000	.88889	.88235
30	1	.00000	.00000	.33333	.20000	.00000	.00000	.11111	.11765
31	0	.80952	.85714	1.00000	.80000	.90000	1.00000	.88889	.82353
31	1	.19048	.14286	.00000	.20000	.10000	.00000	.11111	.17647
32	0	.90476	1.00000	.93333	1.00000	.90000	1.00000	.88889	.76471
32	1	.09524	.00000	.06667	.00000	.10000	.00000	.11111	.23529
33	0	.71429	1.00000	.93333	.80000	.70000	.83333	.88889	.82353
33	1	.28571	.00000	.06667	.20000	.30000	.16667	.11111	.17647
34	0	.85714	1.00000	1.00000	1.00000	1.00000	.83333	1.00000	.82353
34	1	.14286	.00000	.00000	.00000	.00000	.16667	.00000	.17647
35	0	1.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000
35	1	.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
36	0	.63636	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
36	1	.36364	.00000	.00000	.00000	.00000	.00000	.00000	.00000

Figura VI.24 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Anatomia Coronariana estimadas na modalidade "O Banco de Dados como Amostra".

DESC V.OBS		CLASSES		DESC V.OBS		CLASSES	
1	2	1	2	1	2	1	2
1	1	.8000000000	.8108108108	8	1	.0200000000	.1025641026
1	2	.0800000000	.0540540541	9	0	.8600000000	.7500000000
1	3	.1200000000	.1351351351	9	1	.1400000000	.2500000000
2	1	.6000000000	.8250000000	10	0	.6530612245	.4750000000
2	2	.4000000000	.1750000000	10	1	.3469387755	.5250000000
3	1	.0400000000	.0750000000	11	0	.5102040816	.6153846154
3	2	.8000000000	.7500000000	11	1	.4897959184	.3846153846
3	3	.1600000000	.1750000000	12	1	.4347826087	.4473684211
4	0	.0000010000	.0000010000	12	2	.4782608696	.5263157895
4	1	.6799993200	.8499991500	12	3	.0869565217	.0263157895
4	2	.3199996800	.1499998500	13	1	.7142857143	.6410256410
5	0	.8600000000	.5750000000	13	2	.2857142857	.3589743590
5	1	.1400000000	.4250000000	14	0	.5957446809	.6666666667
6	1	.1956521739	.1388887500	14	1	.4042553191	.3333333333
6	2	.6304347826	.5833327500	15	0	.8750000000	.7142857143
6	3	.1521739130	.2777775000	15	1	.1250000000	.2857142857
6	4	.0217391304	.0000010000	16	0	.3000000000	.1794871795
7	0	.7800000000	.7750000000	16	1	.7000000000	.8205128205
7	1	.2200000000	.2250000000	17	0	.8333333333	.7857142857
8	0	.9800000000	.8974358974	17	1	.1666666667	.2142857143

Figura VI.25 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Função Ventricular estimadas na modalidade "0 Banco de Dados como Amostra".

CLASSES		CLASSES	
DESC	V.OBS	DESC	V.OBS
18	0	.3181818182	.3333333333
18	1	.6818181818	.6666666667
19	0	.7608695652	.7272727273
19	1	.2391304348	.2727272727
20	0	.7333333333	.7500000000
20	1	.2666666667	.2500000000
21	1	.8292666341	.7777777778
21	2	.1707313659	.1666666667
21	3	.0000010000	.0277777778
21	4	.0000010000	.0277777778
22	1	.2000000000	.1000000000
22	2	.8000000000	.9000000000
23	0	.9200000000	.9250000000
23	1	.0800000000	.0750000000
24	0	.9600000000	.8500000000
24	1	.0400000000	.1500000000
25	0	.9999990000	.8250000000
25	1	.0000010000	.1750000000
26	0	.9800000000	.9750000000
26	1	.0200000000	.0250000000
27	0	.9000000000	.6250000000
27	1	.1000000000	.3750000000
28	0	.9800000000	.9999990000
28	1	.0200000000	.0000010000
29	0	.9000000000	.8000000000
29	1	.1000000000	.2000000000
30	0	.9600000000	.8250000000
30	1	.0400000000	.1750000000
31	0	.8400000000	.9250000000
31	1	.1600000000	.0750000000
32	0	.9400000000	.8500000000
32	1	.0600000000	.1500000000
33	0	.7400000000	.9250000000
33	1	.2600000000	.0750000000
34	0	.9200000000	.9250000000
34	1	.0800000000	.0750000000
35	0	.3181818182	.0000010000
35	1	.6818181818	.9999990000
36	0	.8461538462	.9999990000
36	1	.1538461538	.0000010000

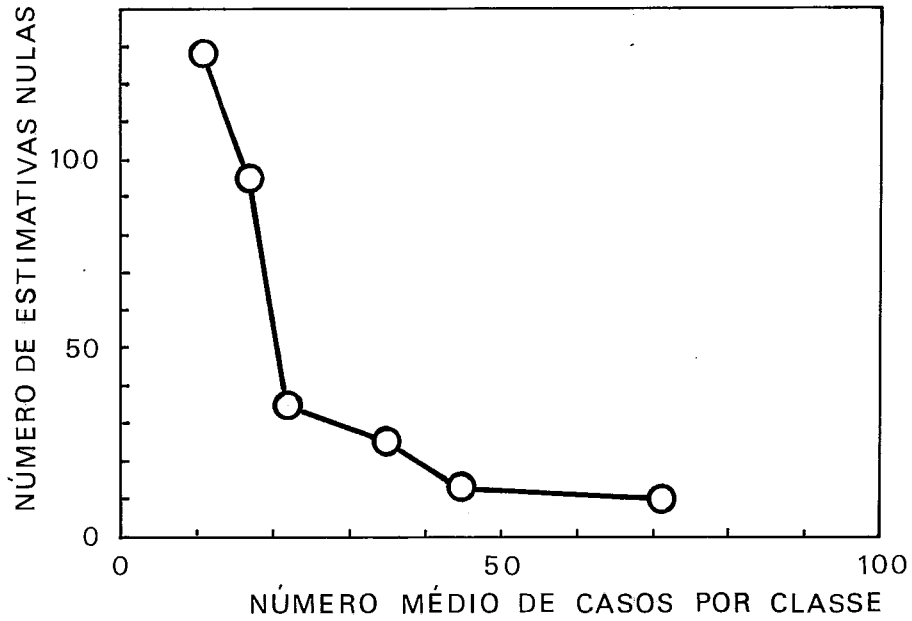
Figura VI.25 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Função Ventricular estimadas na modalidade "0 Banco de Dados Como Amostra".

CLASSES		CLASSES			
DESC V.DES	1	2	DESC V.DES	1	2
1	.8035714286	.7142850000	10	.6607142857	.2857142857
1	.0892857143	.0000100000	10	.3392857143	.7142857143
1	.1071428571	.2857140000	11	.5714285714	.5000000000
2	.6842105263	.5714285714	11	.4285714286	.5000000000
2	.3157894737	.4285714286	12	.4905660377	.4615384615
3	.0350877193	.0000100000	12	.4528301887	.4615384615
3	.7894736842	.8571420000	12	.0566037736	.0769230769
3	.1754385965	.1428570000	13	.6785714286	.7142857143
4	.0000100000	.0000100000	13	.3214285714	.2857142857
4	.7543852105	.8571420000	14	.6000000000	.7857142857
4	.2456137895	.1428570000	14	.4000000000	.2142857143
5	.7368421053	.7142857143	15	.7735849057	.7692307692
5	.2631578947	.2857142857	15	.2264150943	.2307692308
6	.1568627451	.1428570000	16	.2678571429	.2142857143
6	.6078431373	.7142850000	16	.7321428571	.7857142857
6	.2156862745	.1428570000	17	.8500000000	.5714285714
6	.0196078431	.0000100000	17	.1500000000	.4285714286
7	.7719298246	.7142857143	18	.2857142857	.5000000000
7	.2280701754	.2857142857	18	.7142857143	.5000000000
8	.9464285714	.9285714286	19	.7254901961	.7692307692
8	.0535714286	.0714285714	19	.2745098039	.2307692308
9	.8596491228	.7857142857	20	.6521739130	.9999990000
9	.1403508772	.2142857143	20	.3478260870	.0000010000

Figura VI.26 - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Tipo de Tratamento, estimadas na modalidade "O Banco de Dados como Amostra".

DESC V.OBS		CLASSES		DESC V.OBS		CLASSES	
1	2	1	2	1	2	1	2
21	1	.8367338571	.7499985000	31	1	.1228070175	.1428571429
21	2	.1428570000	.2499995000	32	0	.8947368421	.8571428571
21	3	.0204081429	.0000010000	32	1	.1052631579	.1428571429
21	4	.0000010000	.0000010000	33	0	.7894736842	.8571428571
22	1	.2000000000	.0000010000	33	1	.2105263158	.1428571429
22	2	.8000000000	.9999990000	34	0	.8947368421	.9285714286
23	0	.9122807018	.9285714286	34	1	.1052631579	.0714285714
23	1	.0877192982	.0714285714	35	0	.3043478261	.0000010000
24	0	.8947368421	.9285714286	35	1	.6956521739	.9999990000
24	1	.1052631579	.0714285714	36	0	.8518518519	.9999990000
25	0	.9473684211	.7857142857	36	1	.1481481481	.0000010000
25	1	.0526315789	.2142857143	38	1	.3333333333	.0000010000
26	0	.9649122807	.9999990000	38	2	.0877192982	.0000010000
26	1	.0350877193	.0000010000	38	3	.1754385965	.2142850714
27	0	.8070175439	.8571428571	38	4	.0526315789	.0000010000
27	1	.1929824561	.1428571429	38	5	.1052631579	.1428567143
28	0	.9999990000	.9285714286	38	6	.0526315789	.1428567143
28	1	.0000010000	.0714285714	38	7	.0701754386	.1428567143
29	0	.8947368421	.8571428571	38	8	.1228070175	.3571417857
29	1	.1052631579	.1428571429	40	1	.6666666667	.4285714286
30	0	.9122807018	.9999990000	40	2	.3333333333	.5714285714
30	1	.0877192982	.0000010000				
31	0	.8771929825	.8571428571				

Figura VI.26 (continuação) - Probabilidades condicionais dos valores observáveis dos descritores com relação à variável Tipo de Tratamento, estimadas na modalidade "O Banco de Dados como Amostra".



**Figura VI.27** - Variação do número de estimativas nulas de probabilidades condicionais em função do número médio de casos de calibração por classe diagnóstica em modelos derivados da variável Anatomia Coronariana.



MODELO	O BANCO DE DADOS COMO AMOSTRA DE UM UNIVERSO ESTACIONÁRIO				O BANCO DE DADOS COMO UNIVERSO			
	Prob. a Priori Uniformes		Prob. a Priori Calculadas do DB		Prob. a Priori Uniformes		Prob. a Priori Calculadas do DB	
	Porcentual Diagn. Corretos	Entropia média	Porcentual Diagn. Corretos	Entropia média	Porcentual Diagn. Corretos	Entropia média	Porcentual Diagn. Corretos	Entropia média
Presença de Lesões Significativas	88,46 **	0,078827	92,31 **	0,080082	88,73 **	0,132082	91,55 **	0,130781
Número de Vasos Lesados	40,38 **	0,539042	38,46 **	0,549430	64,08 **	0,645907	66,90 **	0,657397
Anatomia Coronariana	34,62	0,552558	34,62	0,548118	67,61 **	0,712457	69,01 **	0,701025
Função Ventricular	65,38 *	0,259600	65,38 *	0,264111	76,06 **	0,340483	76,06 **	0,342268
Tratamento (sem catet)	76,60	0,237406	78,72	0,163534	83,90 **	0,344476	84,75 **	0,256087
Presença de Lesões x Função Ventricular	53,85 **	0,351752	59,62 **	0,338727	72,54 **	0,480444	76,06 **	0,483779
Perfil	53,85 **	0,351737	59,62 *	0,338727	72,34 **	0,483852	75,89 **	0,487210
Tratamento (com catet) (Vars 1 a 36,38,42)	70,21	0,179031	80,85 *	0,185809	80,51 **	0,246376	88,14 **	0,232172

\* Significativo ao nível de 5%

\*\* Significativo ao nível de 1%

Figura VI.28 - Avaliação do desempenho dos modelos bayesianos de classificação diagnóstica na área de doença coronariana com  $p_{\min} = 10^{-6}$ .

MODELO	O BANCO DE DADOS COMO AMOSTRA DE UM UNIVERSO ESTACIONÁRIO				O BANCO DE DADOS COMO UNIVERSO			
	Prob. a Priori Uniformes		Prob. a Priori Calculadas do DB		Prob. a Priori Uniformes		Prob. a Priori Calculadas do DB	
	Porcentual Diagn. Corretos	Entropia média	Porcentual Diagn. Corretos	Entropia média	Porcentual Diagn. Corretos	Entropia média	Porcentual Diagn. Corretos	Entropia média
Presença de Lesões Significativas	86,54 **	0,10223	92,31 **	0,08983	88,73 **	0,13169	91,55 **	0,13064
Número de Vasos Lesados	38,46 **	0,62385	36,54 **	0,62846	64,08 **	0,64566	66,90 **	0,65721
Anatomia Coronariana	38,46 -	0,78930	38,46 -	0,75611	67,61 **	0,71235	69,01 **	0,70094
Função Ventricular	63,46 -	0,27127	63,46 -	0,25561	76,06 **	0,34047	76,06 **	0,34226
Tratamento (sem catet)	76,60 -	0,237173	78,72 -	0,163448	83,90 **	0,344382	84,75 **	0,256065
Presença de Lesões x Função Ventricular	51,92 *	0,45588	57,69 **	0,41132	72,54 **	0,48033	76,06 **	0,483701
Perfil	51,92 **	0,42809	57,69 **	0,40800	72,34 **	0,483738	75,89 **	0,486478
Tratamento (com catet) (Vars 1 a 36,38,42)	70,21 -	0,17873	80,85 *	0,18430	80,51 **	0,24614	88,14 **	0,232113

\* Significativo ao nível de 5%

\*\* Significativo ao nível de 1%

Figura VI.29 - Avaliação do desempenho dos modelos bayesianos de classificação diagnóstica na área de doença coronariana com  $p_{\min}=0$ .

MODELO	O BANCO DE DADOS COMO AMOSTRA		O BANCO DE DADOS COMO UNIVERSO	
	Porcentual de Diagnós- ticos Cor- retos	Entropia Média	Porcentual de Diagnós- ticos Cor- retos	Entropia Média
Presença de Lesões Significativas	84,62 **	0,093508	83,80 **	0,223343
Número de Vasos Lesados	34,69 -	0,606184	63,38 **	0,751547
Anatomia Coronariana	30,77 -	0,578656	65,49 **	0,792131

\* Significativo ao nível  $\alpha = 0,05$

\*\* Significativo ao nível  $\alpha = 0,01$

Figura VI.30 - Avaliação do desempenho de modelos bayesianos de classificação diagnóstica excluindo-se os dados do exame de ecocardiografia.

(Pressuposição de independência,  $p_{\min} = 10^{-6}$ ).

MODELO	Porcentual de Diagnósticos Corretos	
	O Banco de Dados como Amostra	O Banco de Dados como Universo
Presença de Lesões Significativas	90,38 **	92,95 **
Número de Vasos Lesados	42,30 **	73,23 **

\*\* Significativo ao nível = 0,01

Figura VI.31 - Avaliação do desempenho de modelos bayesianos com classes diagnósticas aglutinadas a posteriori.

MODELO	O Banco de Dados como Amostra		O Banco de Dados como Universo	
	Porcentual de Diagnós- ticos corre- tos	Entropia Média	Porcentual de Diagnós- ticos corre- tos	Entropia Média
Presença de lesões na Coronária Direita	67,31 *	0,311092	76,76 **	0,381079
Presença de Lesões na Descendente Anterior	65,38	0,237655	76,76 **	0,352033
Presença de Lesões na Coronária Circunflexa	63,46	0,279325	73,24 **	0,357175

\* Significativo ao nível =5%

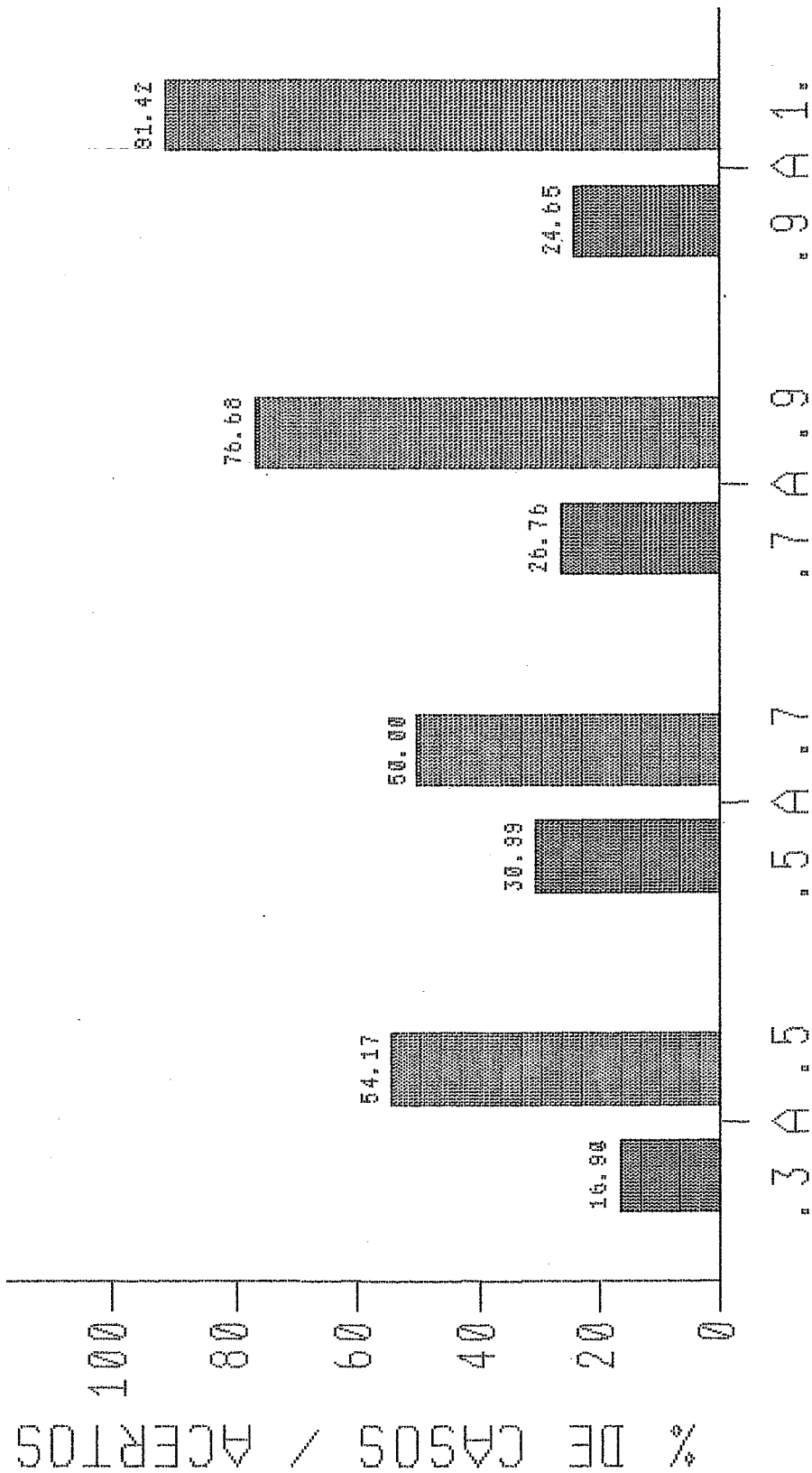
\*\* Significativo ao nível =1%

Figura VI.32 - Avaliação do desempenho dos modelos de classificação diagnóstica Coroná-  
ria direita, Descendente Anterior e Coronária Circunflexa.  
(Probabilidades a priori estimadas do banco de dados,  $p_{\min}=0$ ).



MODELO PRESENÇA DE DOENÇA CORONARIANA - O BANCO DE DADOS COMO UNIVERSO

Figura VI.33 - Eficiência do modelo Presença de Doença Coronariana em função da faixa de probabilidade a posteriori da classe selecionada como decisão (Modelo com pressuposição de independência, probabilidades a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ )



### FAIXA DE PROBABILIDADE

MODELO NUMERO DE VASOS LESADOS - O BANCO DE DADOS COMO UNIVERSO

Figura VI.34 - Eficiência do modelo Número de Vasos Lesados em função da faixa de probabilidade a posteriori da classe selecionada como decisão. (Modelo com pressuposição de independência, probabilidades a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ )

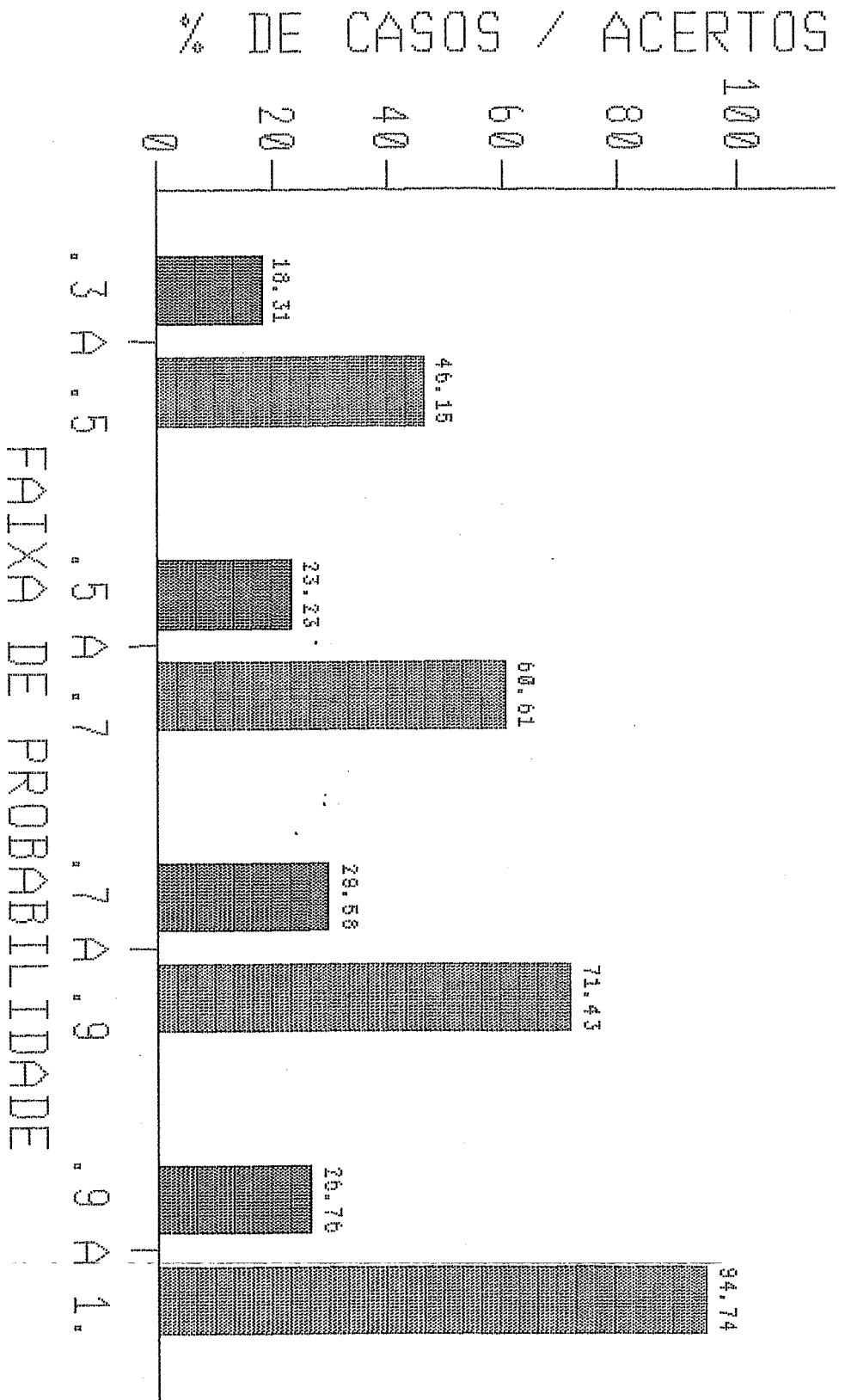
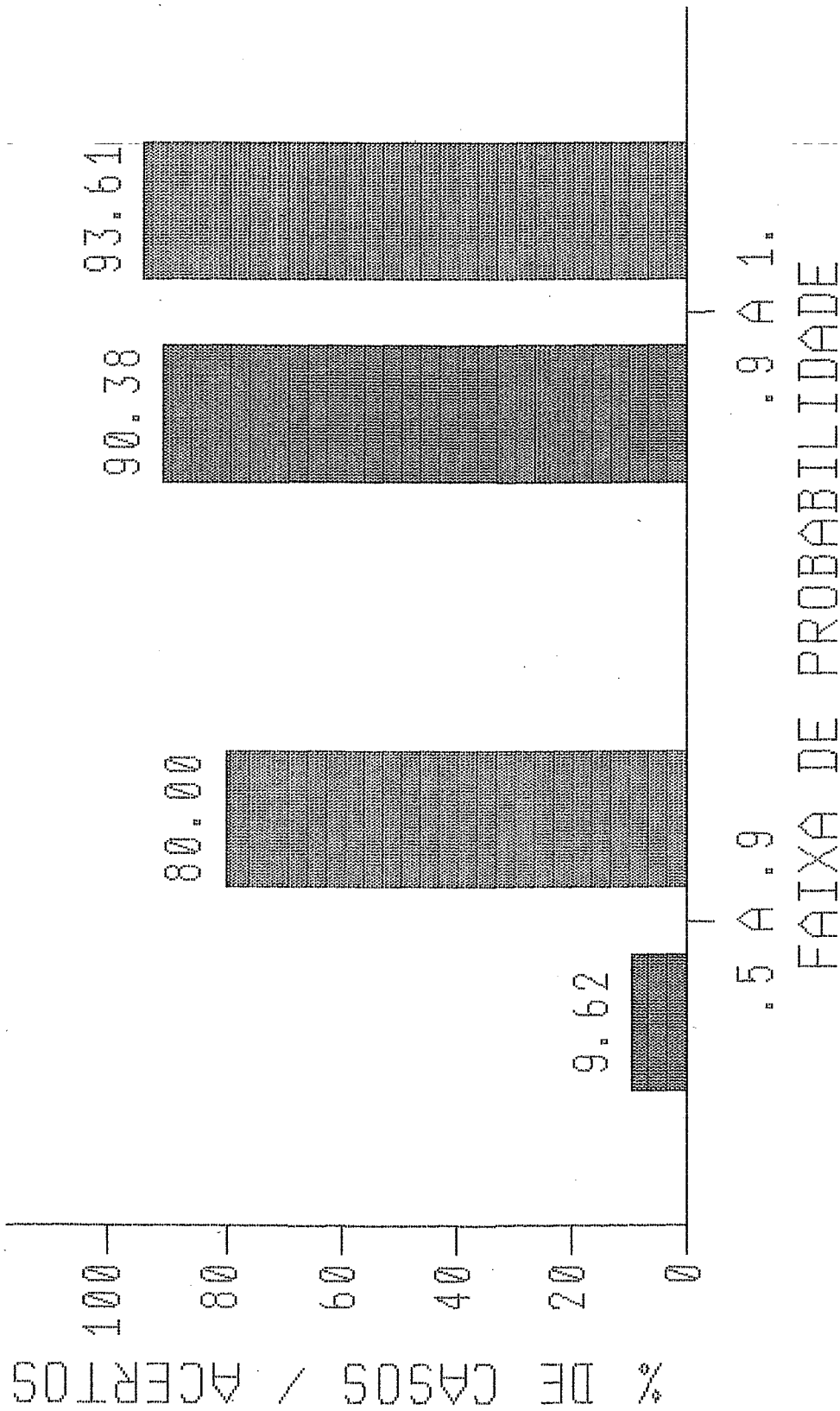


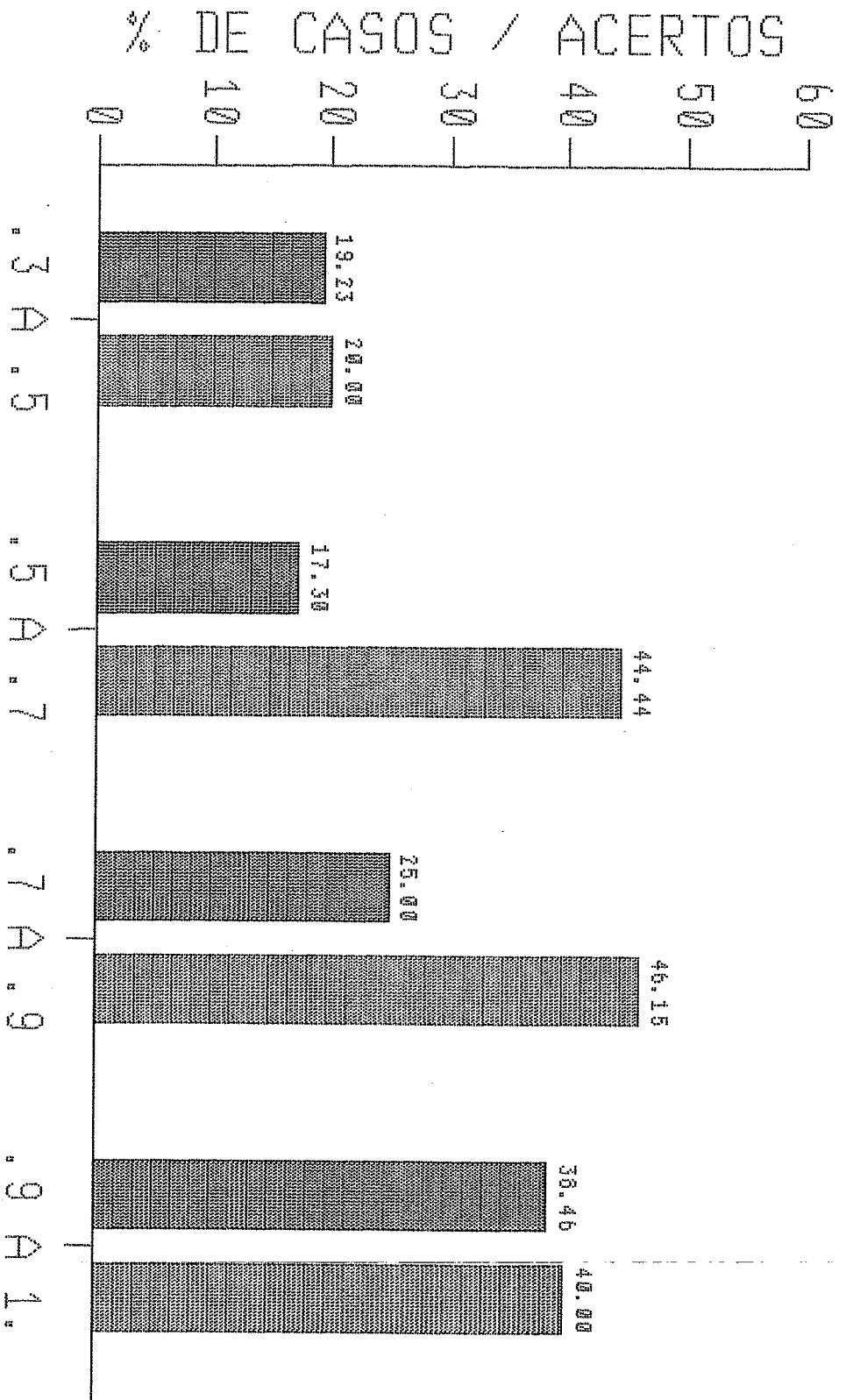
Figura VI.35 - Eficiência do modelo Anatomia Coronariana em função da faixa da probabilidade a posteriori da classe selecionada como decisão. (Modelo com pressuposição de independência,  $P_{\min} = 10^{-6}$ )





MODELO PRESENCIA DE DOENÇA CORONARIANA - O BANCO DE DADOS COMO AMOSTRA

Figura VI.36 - Eficiência do modelo Presença de Doença Coronariana em função da faixa da probabilidade a posteriori da classe selecionada como decisão (Modelo com pressuposição de independência, probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ )



MODELO NUMERO DE VASOS LESADOS - O BANCO DE DADOS COMO AMOSTRA

Figura VI.37 - Eficiência do modelo Número de Vasos Lesados em função da faixa da probabilidade a posteriori da classe selecionada como decisão (Modelo com pressuposição de independência, probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ ).



MODELO ANATOMIA CORONARIANA - O BANCO DE DADOS COMO AMOSTRA

Figura VI.38 - Eficiência do modelo Anatomia Coronariana em função da faixa da probabilidade a posteriori da classe selecionada como decisão. (Modelo com pressuposição de independência, probabilidades a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ )

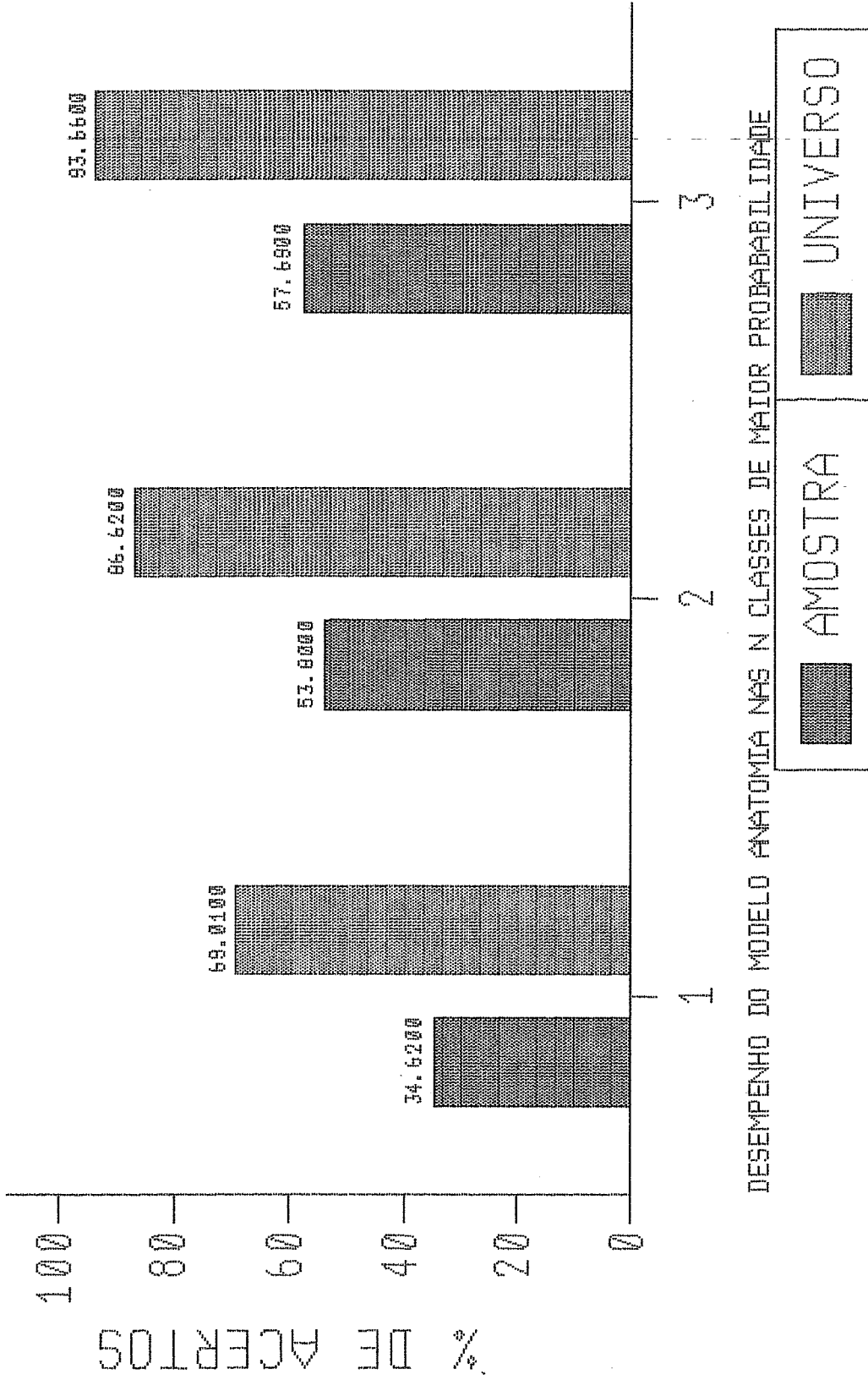


Figura VI.39 - Desempenho do modelo Anatomia Coronariana nas n classes de maior probabilidade de a posteriori. (Modelo bayesiano com pressuposição de independência, probabilidades a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ ).

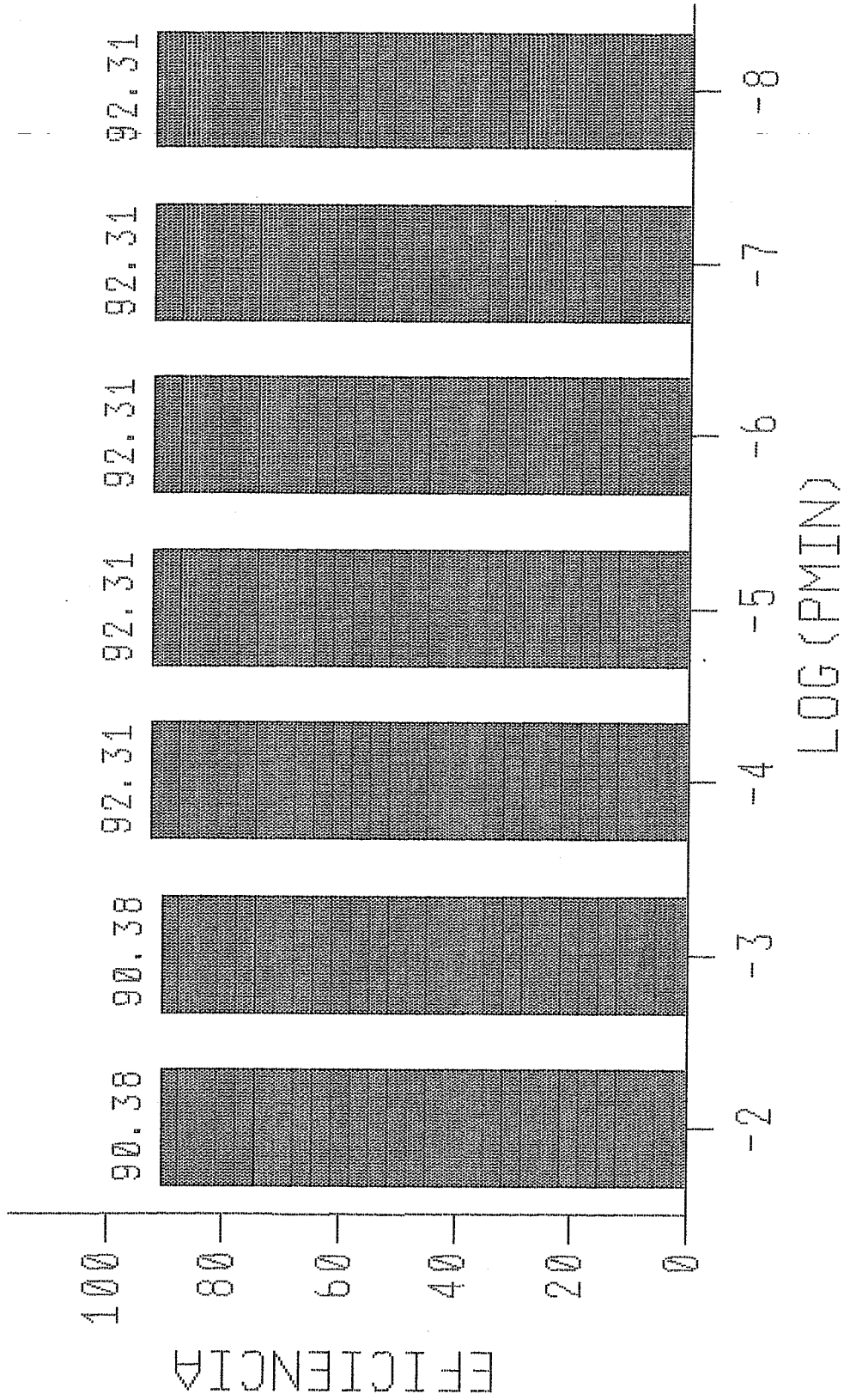


Figura VI.40 - Eficiência do modelo bayesiano de classificação diagnóstica Presença de Doença Coronariana em função do valor da probabilidade condicional mínima ( $p_{min}$ ).

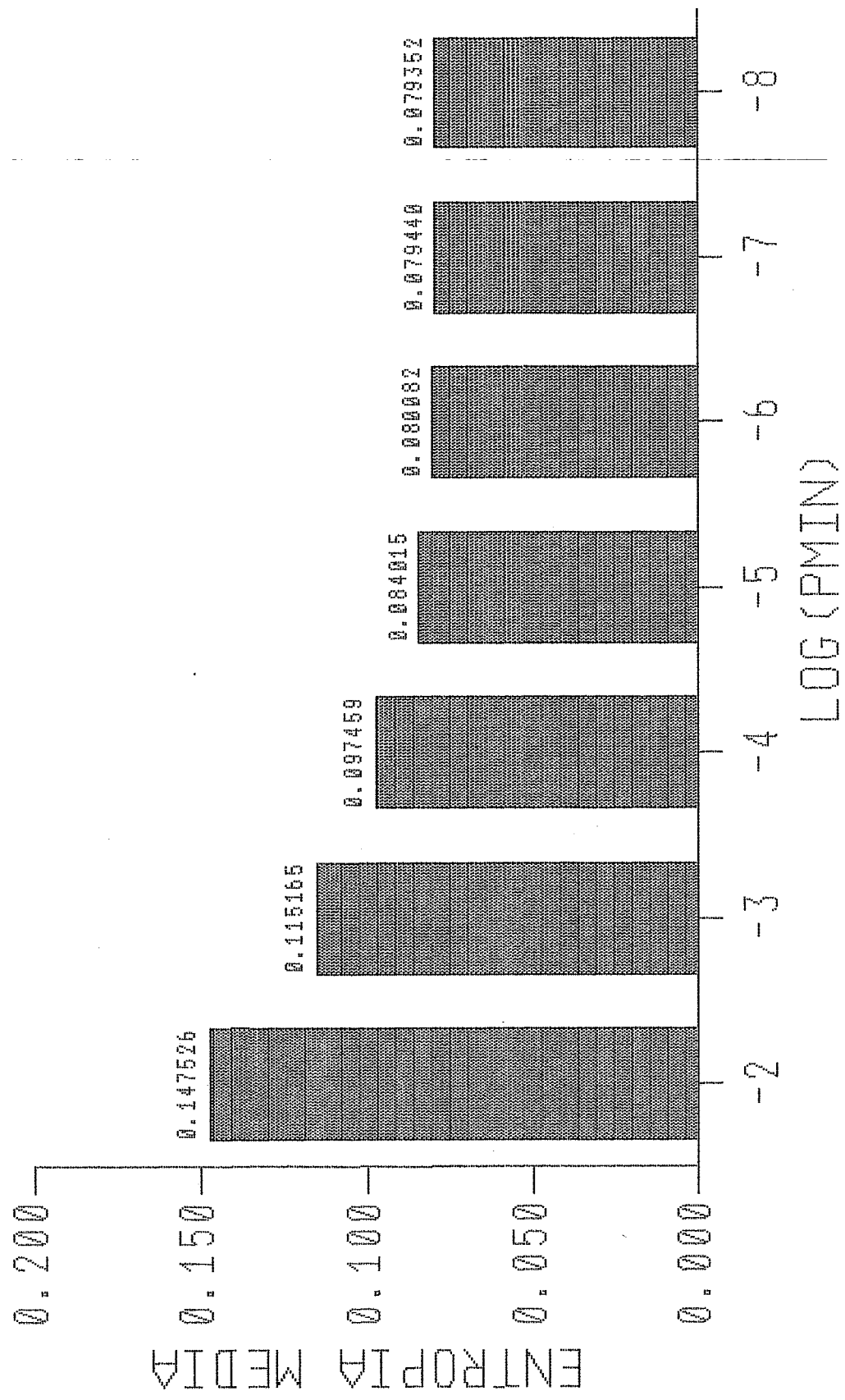


Figura VI.41 - Entropia média do modelo bayesiano de classificação diagnóstica Presença de Doença Coronariana em função do valor da probabilidade condicional: mínima ( $P_{min}$ ).

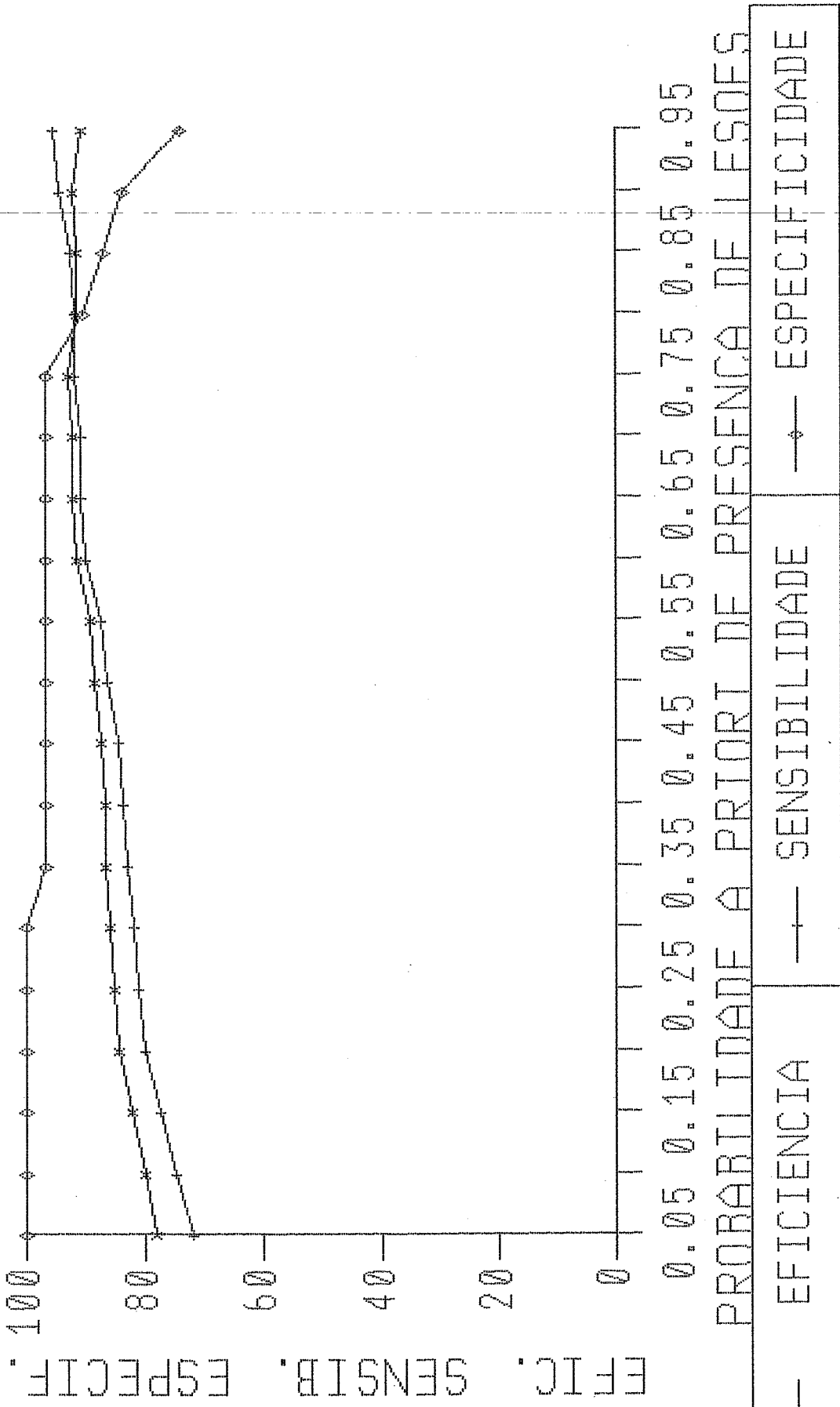


Figura VI.42 - Desempenho do modelo bayesiano de classificação diagnóstica da Presença de Doença Coronariana em função do valor da probabilidade a priori de existência de lesões significativas. (Pressuposição de independência, "0 Banco de Dados com o Universo").

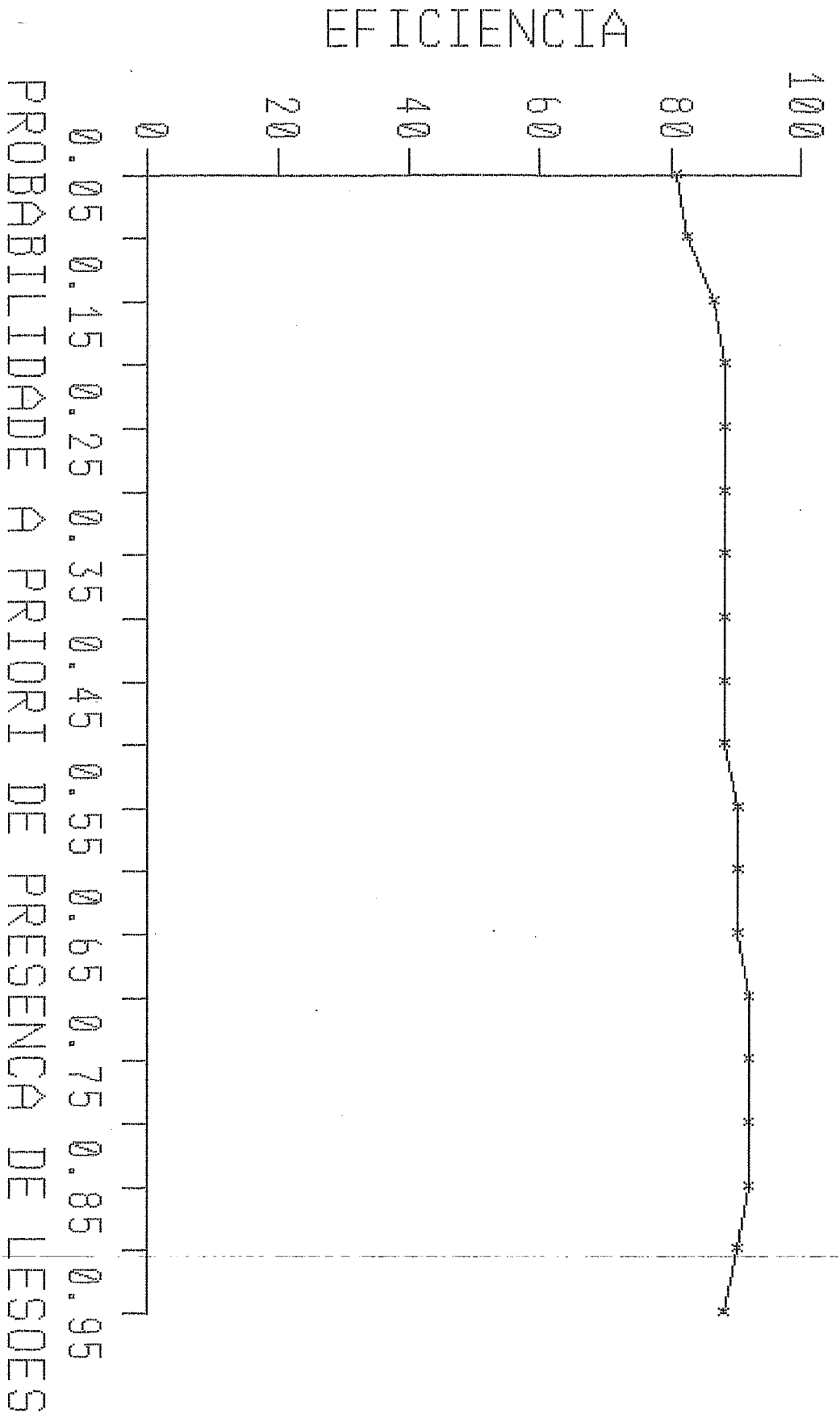
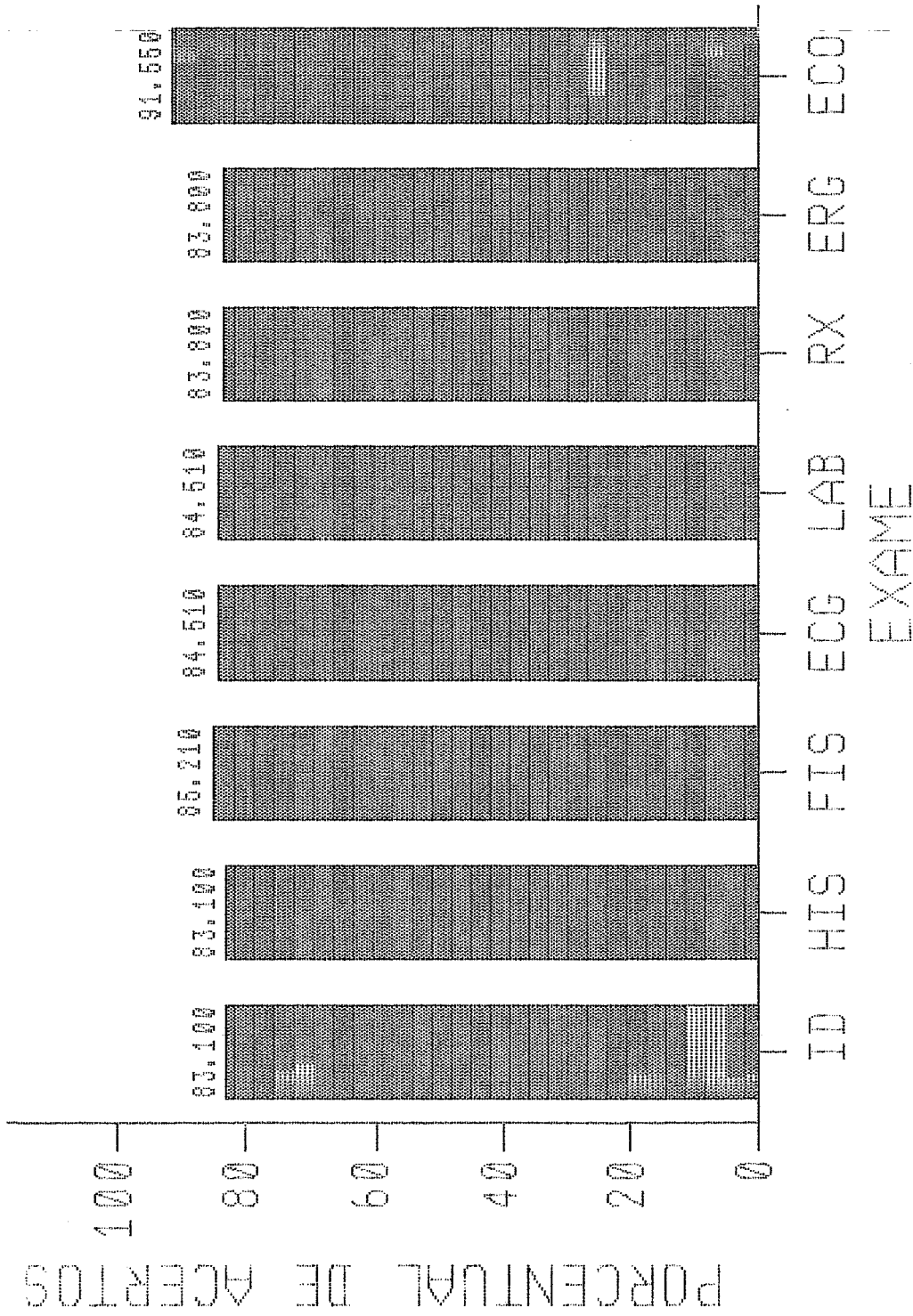


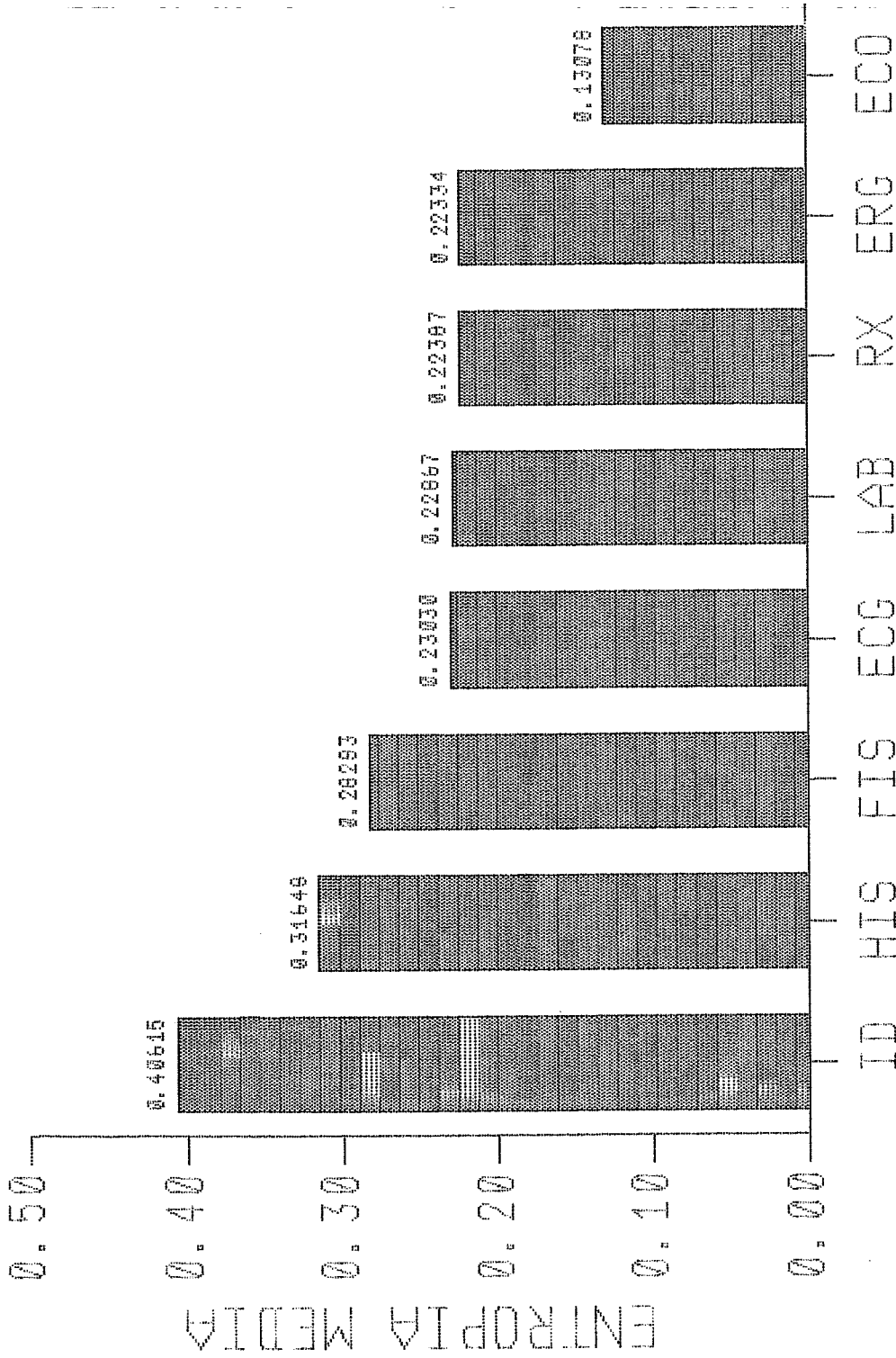
Figura VI.43 - Desempenho do modelo bayesiano de classificação diagnóstica Presença de Doença Coronariana em função da probabilidade a priori de existência de lesões significativas. (Pressuposição de independência, "0 Banco de Dados como Amostragem").





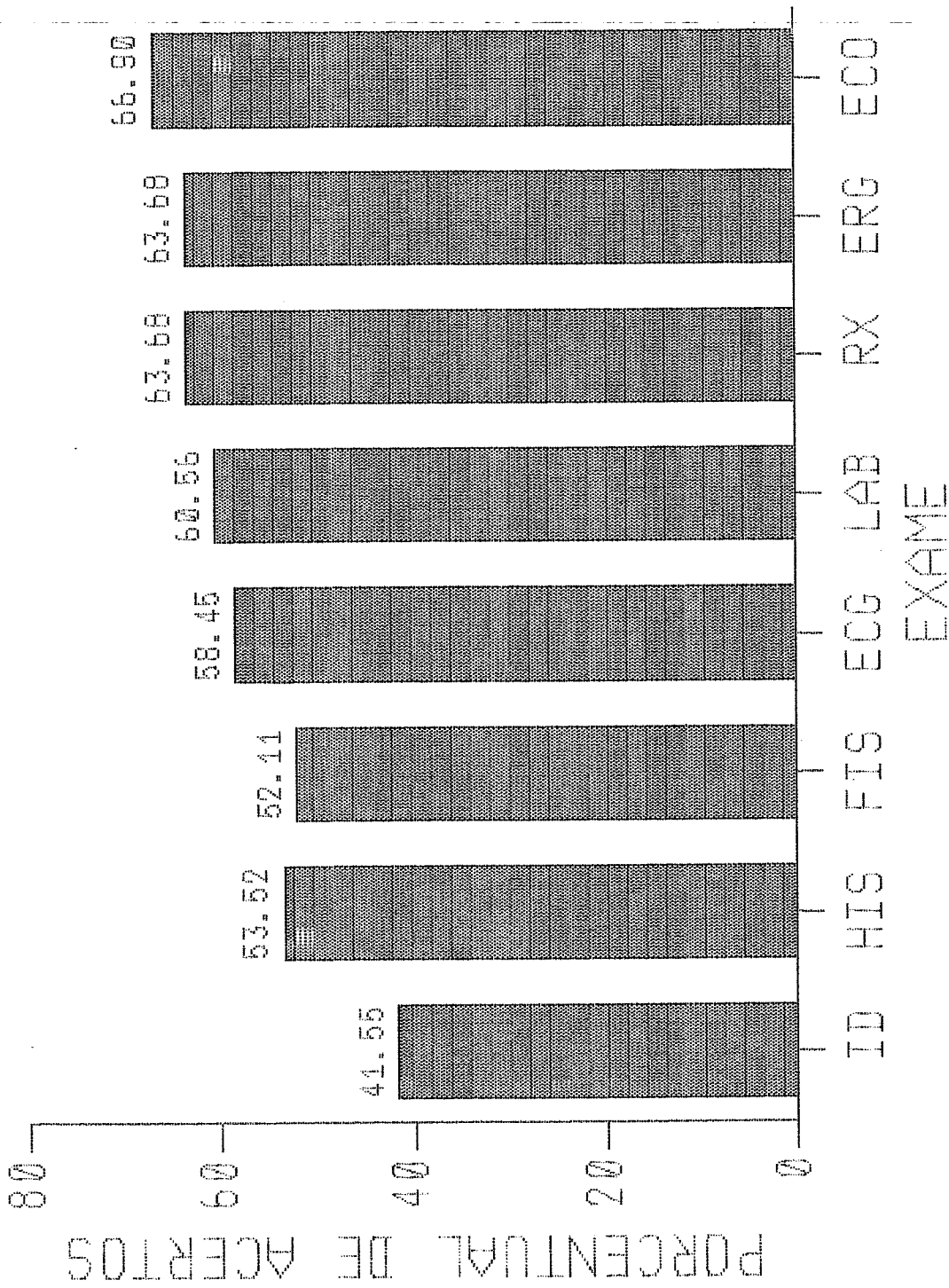
MODELO PRESENÇA DE LESÃO - O BANCO DE DADOS COMO UNIVERSO

Figura VI.44 - Eficiência do modelo de classificação diagnóstica Presença de Doença Coronariana em função do número de descritores utilizados. (Pressuposição de independência,  $p_{\min} = 10^{-6}$ ).



MODELO PRESENÇA DE LESÃO - O BANCO DE DADOS COMO UNIVERSO

Figura VI.45 - Entropia média do modelo de classificação diagnóstica Presença de Doença Coronariana em função do número de descritores utilizados. (Pressuposição de independência,  $P_{min} = 10^{-6}$ ).



MOD. NUMERO DE VASOS LESADOS - O BANCO DE DADOS COMO UNIVERSO

Figura VI.46 - Eficiência do modelo de classificação diagnóstica Número de Vasos Lesados em função do número de descritores utilizados. (Pressuposição de independência,  $P_{\min} = 10^{-6}$ ).

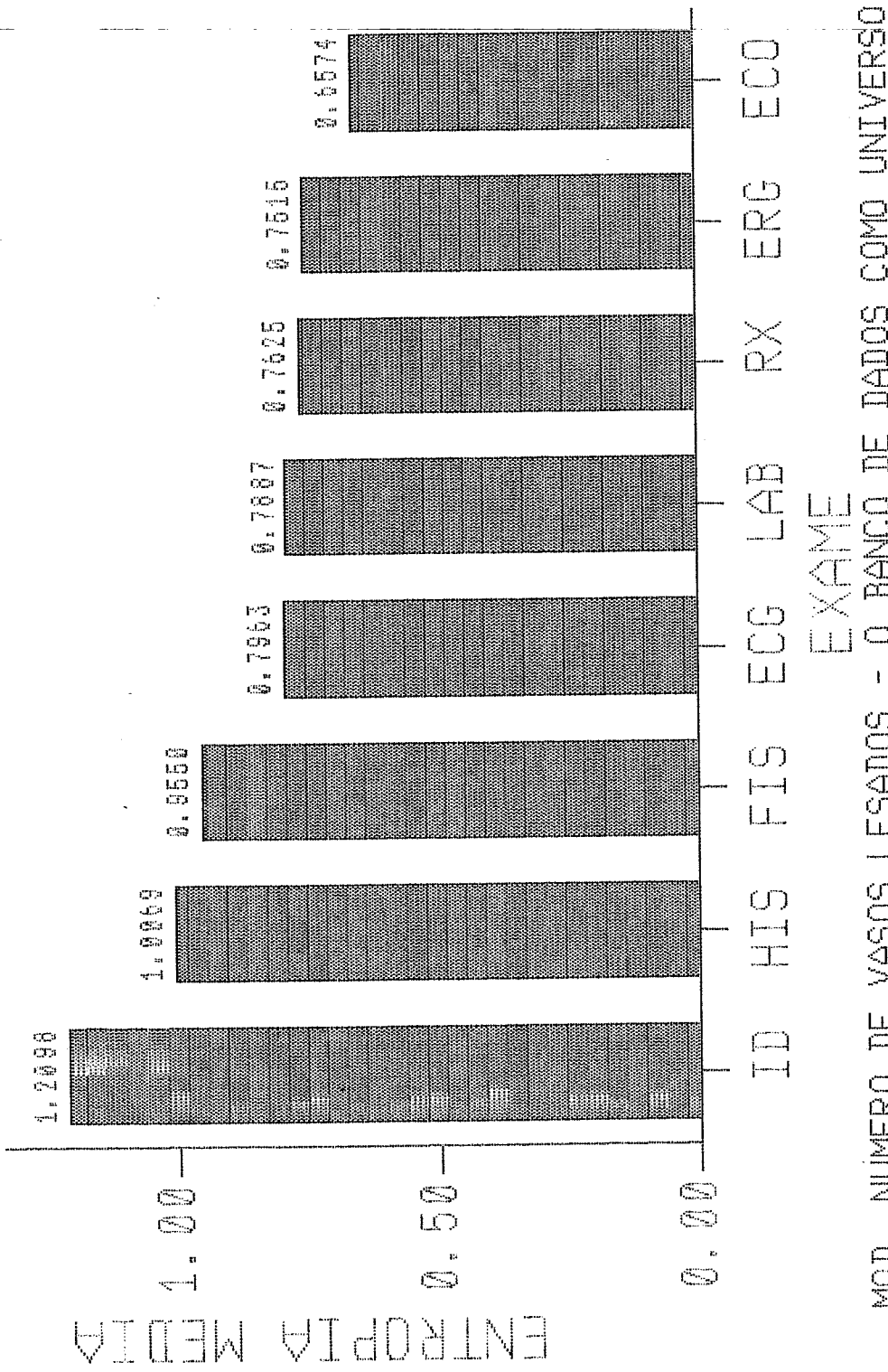
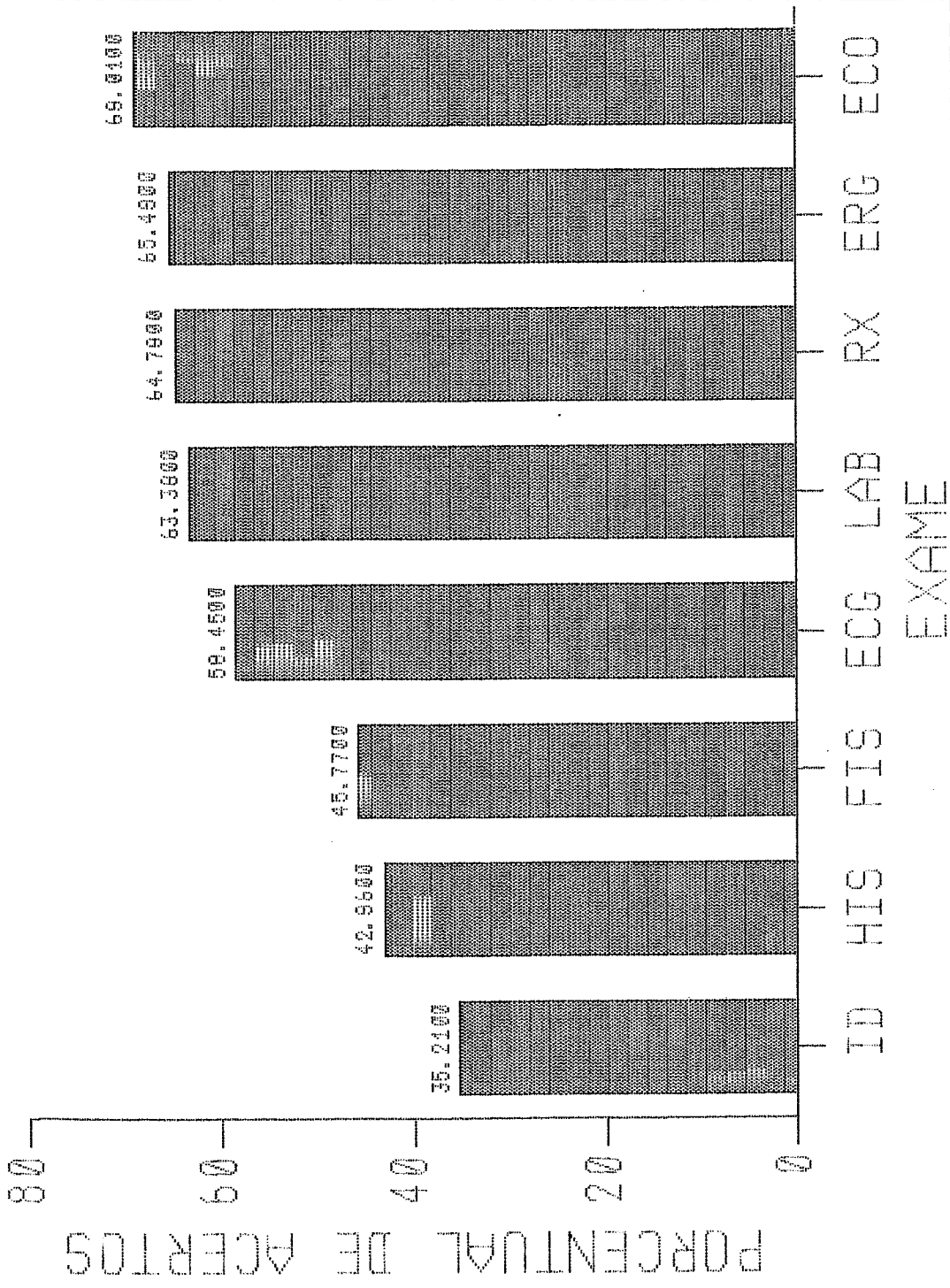


Figura VI.47 - Entropia média do modelo de classificação diagnóstica Número de Vasos Lesados em função do número de descritores utilizados. (Pressuposição de independência,  $P_{\min} = 10^{-6}$ ).



MODELO ANATOMIA - O BANCO DE DADOS COM UNIVRSO

Figura VI.48 - Eficiência do modelo de classificação diagnóstica Anatomia Coronariana em função do número de descritores utilizados. (Pressuposição de independência,  $P_{\min} = 10^{-6}$ ).

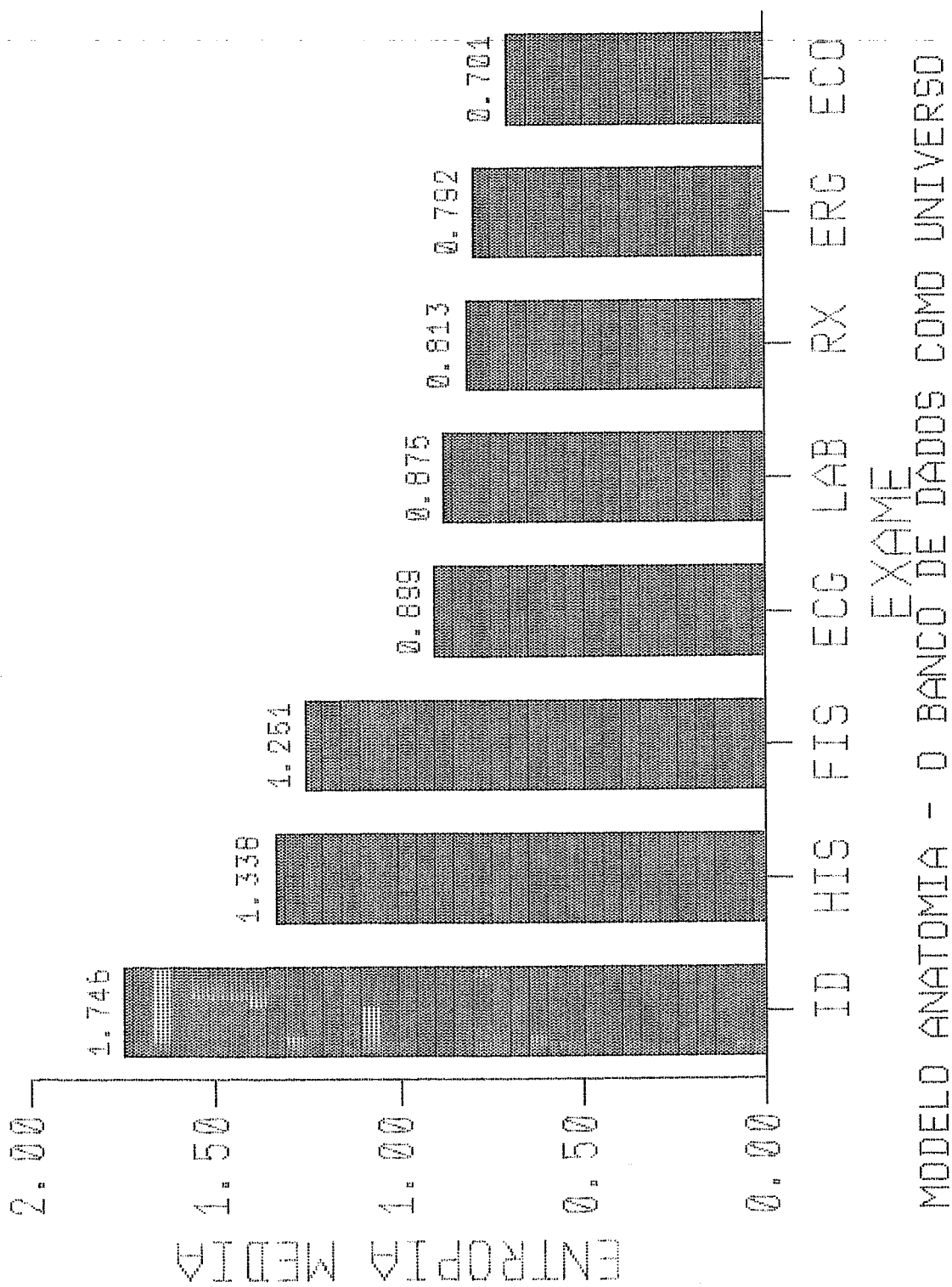
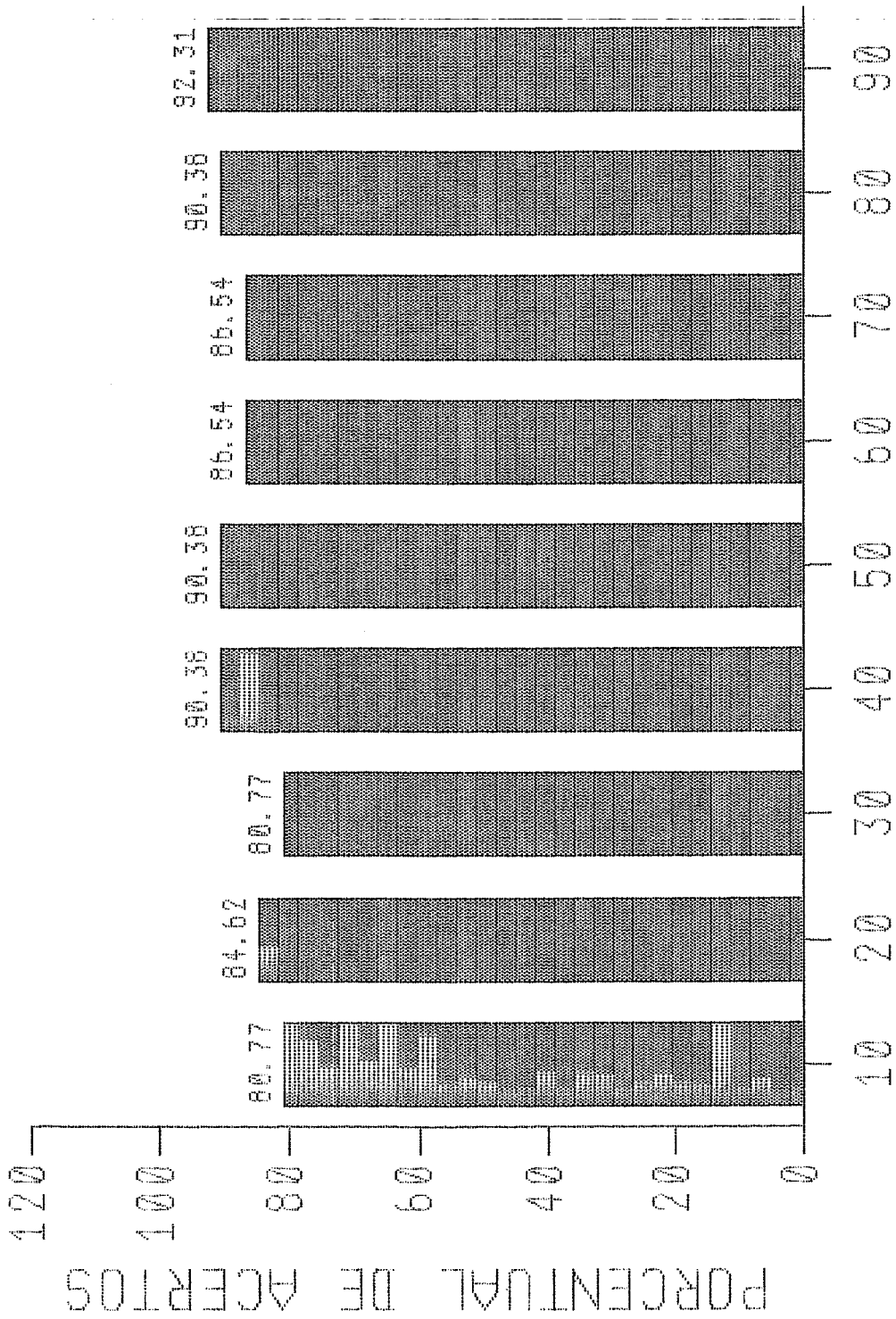


Figura VI.49 - Entropia média do modelo de classificação diagnóstica Anatomia Coronariana em função do número de descritores utilizados. (Pressuposição de independência,  $P_{\min} = 10^{-6}$ ).



TAMANHO DO BANCO DE DADOS  
MODELO PRESENÇA DE LESÃO - O BANCO DE DADOS COMO AMOSTRA

Figura VI.50 - Eficiência do modelo de classificação diagnóstica Presença de Doença Coronariana em função do tamanho do banco de dados de calibração. (Pressuposição de independência,  $p_{\min} = 10^{-6}$ ).

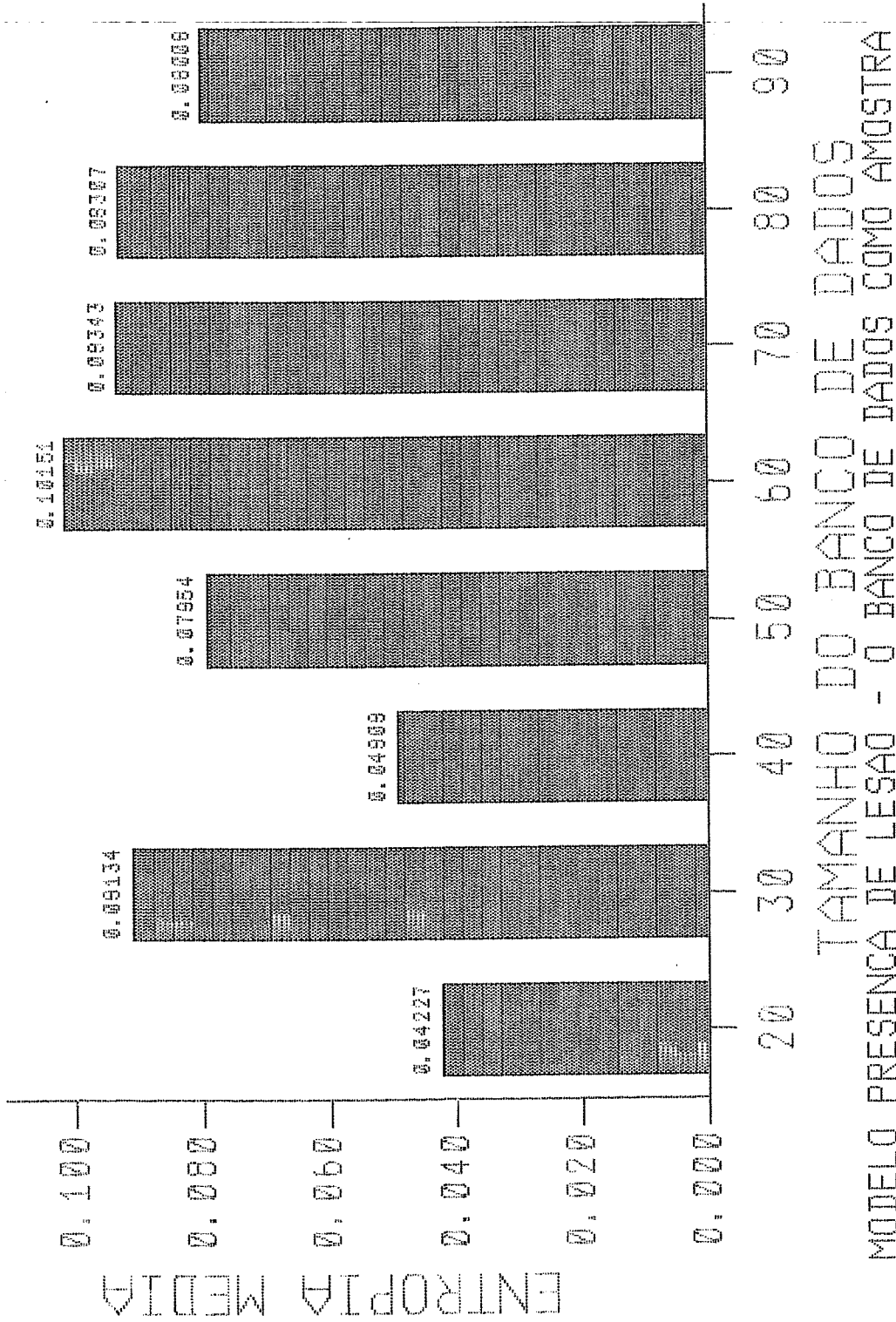


Figura VI.51 - Entropia média do modelo de classificação diagnóstica Presença de Doença Coronariana em função do tamanho do banco de dados de calibração. (Pressuposição de independência,  $P_{\min} = 10^{-6}$ ).



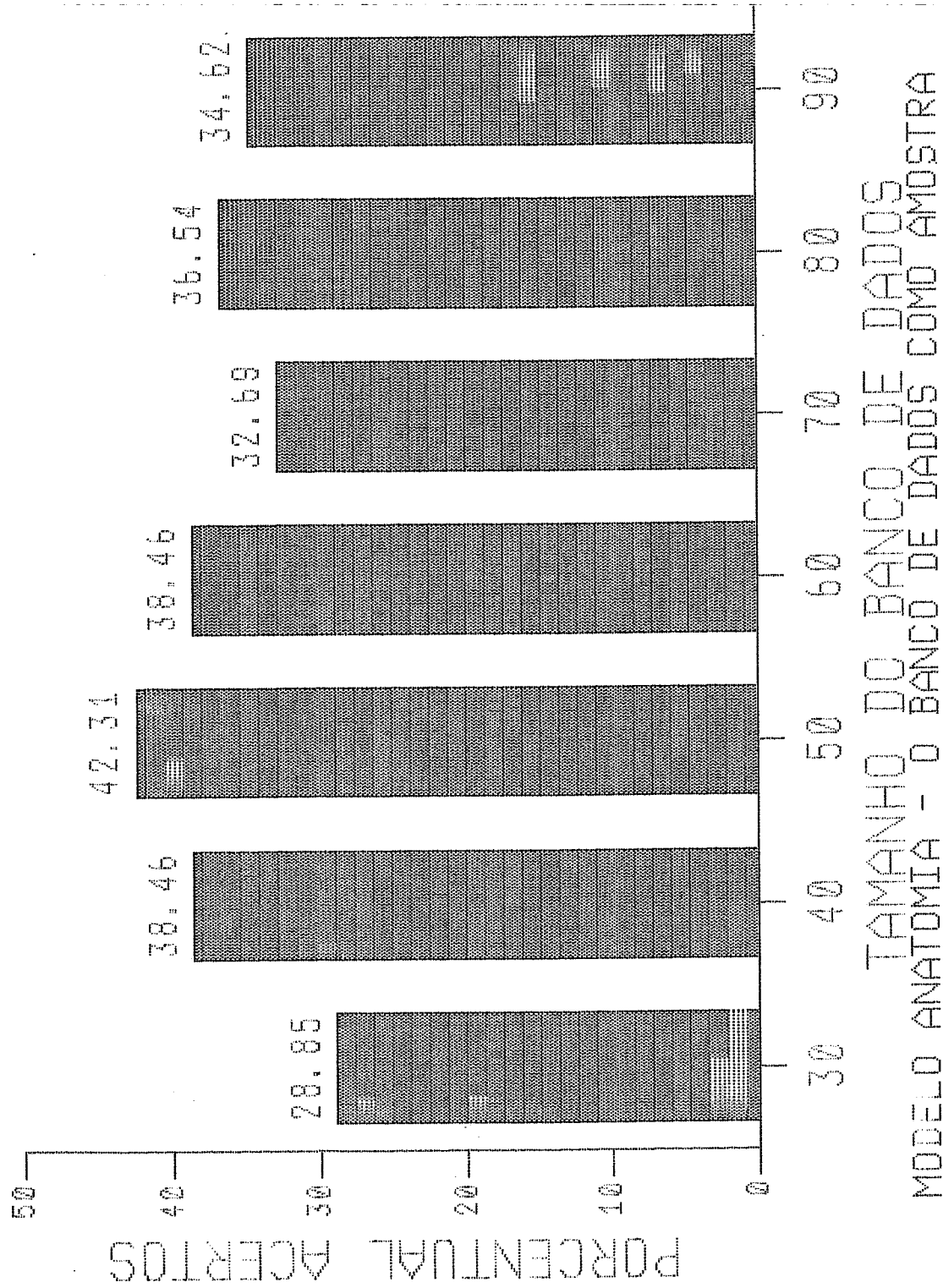


Figura VI.52 - Eficiência do modelo de classificação diagnóstica Anatomia Coronariana em função do tamanho do banco de dados de calibração. (Pressuposição de independência,  $p_{\min} = 10^{-6}$ ).

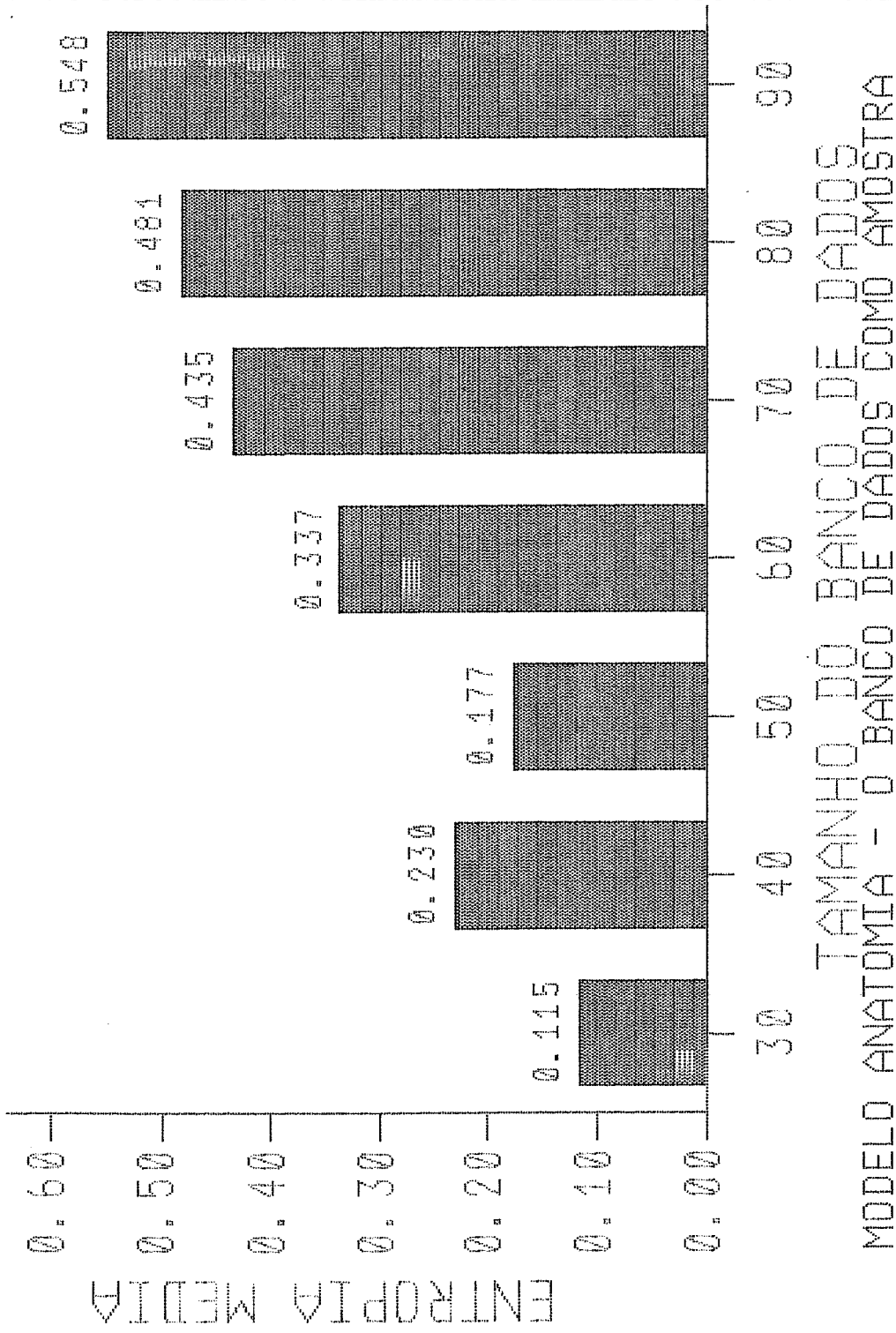


Figura VI.53 - Entropia média do modelo de classificação diagnóstica Anatomia Coronariana em função do tamanho do banco de dados de calibração. (Pressuposição de independência,  $P_{\min} = 10^{-6}$ ).

NÍVEL DE SIGNIFICÂNCIA

0,25    0,1    0,075    0,05    0,025    0,01    0,0075    0,005    0,0025    0,001    0,00075    0,0005    0,00001

DESCRITORES

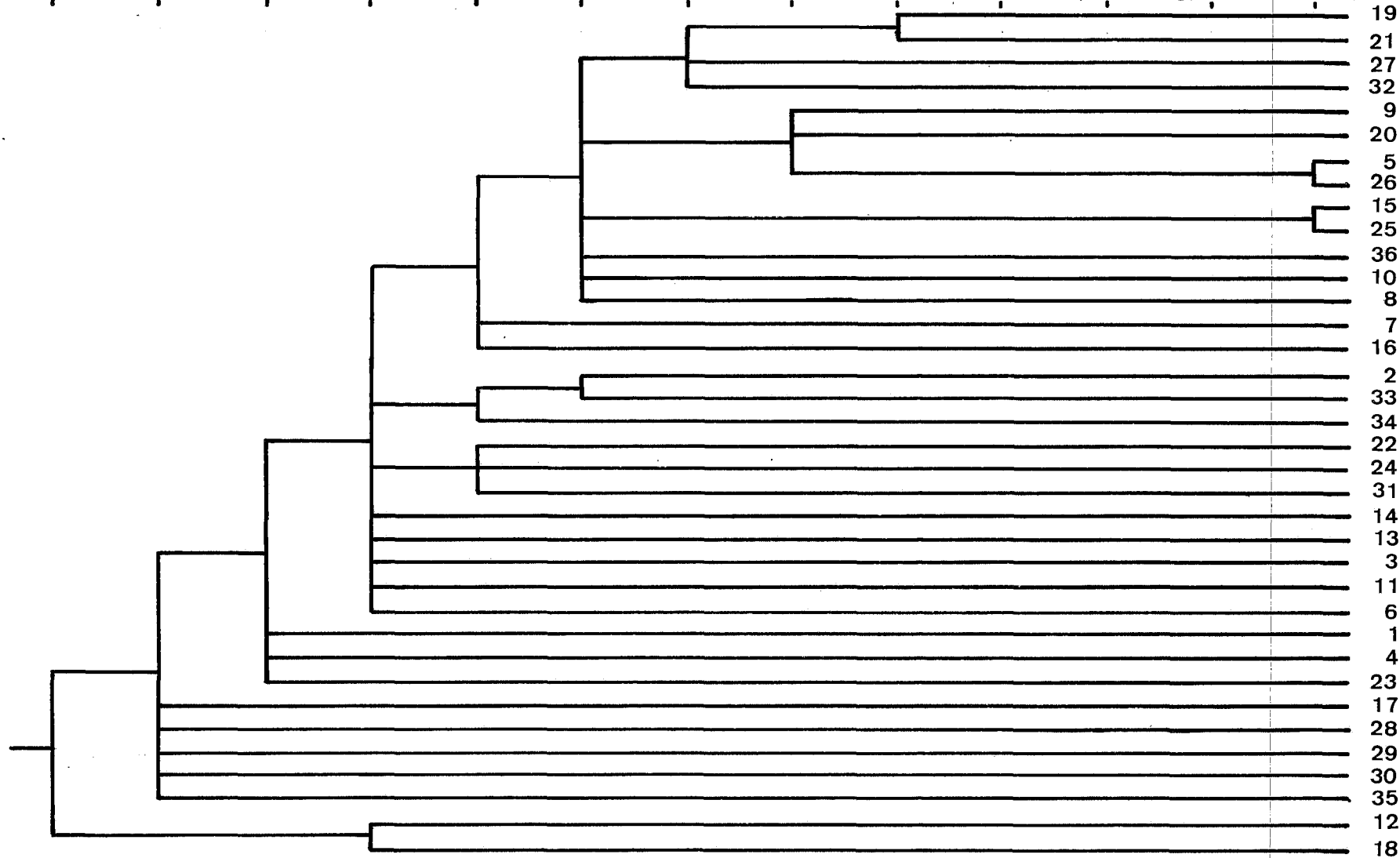
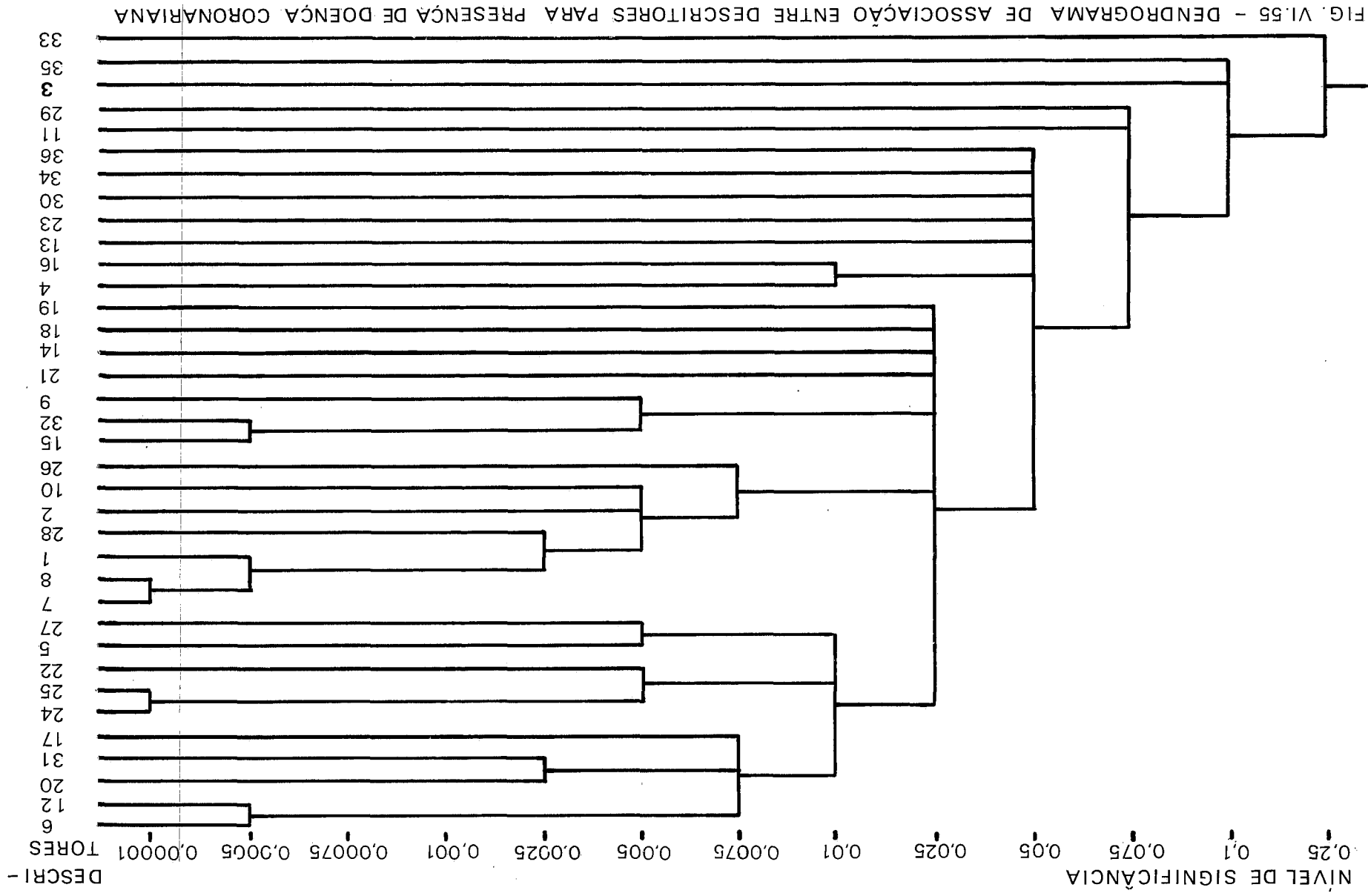


FIG. VI. 54 - DENDROGRAMA DE ASSOCIAÇÃO ENTRE DESCRITORES PARA AUSÊNCIA DE DOENÇA CORONARIANA



NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE 1 DESCRITOR
0,0005	(5,26), (15,25)
0,00075	(5,26), (15,25)
0,001	(5,26), (15,25)
0,0025	(5,26), (15,25), (19,21)
0,005	(5,9,20,26), (15,25), (19,21)
0,0075	(5,9,20,26), (15,25), (19,21,27,32)
0,01	(2,33), (5,8,9,10,15,19,20,21,25,26,27,32,36)
0,025	(2,33,34), (5,7,8,9,10,15,16,19,20,21,25,26,27,32,36), (22,24,31)
0,05	(2,3,5,6,7,8,9,10,11,13,14,15,16,19,20,21,22,24,25,26,27,31,32,33,34,36), (12,18)
0,075	(1 a 11, 13 a 16, 19 a 27, 31 a 34, 36), (12,18)
0,1	(1 a 11, 13 a 17, 19 a 36), (12, 18)

**Figura VI.56** - Aglomerados de descritores interdependentes para a classe 1 (Sem lesões significativas) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	-
0,00075	(8,26)
0,001	(8,26)
0,0025	(8,17,26)
0,005	(8,17,26), (10,14)
0,0075	(8,17,26), (10,14)
0,01	(8,17,26), (10,14)
0,025	(1,2,4), (6,11,21), (8,17,26), (9,32,34), (10,14), (20,23)
0,05	(1,2,4), (5,27), (6,9,10,11,12,14,15,20,21,23,31,32,34), (8,17,26), (22,35,36)
0,075	(1 a 4, 6, 8 a 15, 17, 19 a 23, 26, 30, 31, 32, 34,35,36), (5,27)
0,1	(1 a 36)

**Figura VI.57** - Aglomerados de descritores interdependentes para a classe 2 (com lesões na Coronária Di<sub>i</sub>reita isolada) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	(7,11), (15,33), (19,24)
0,00075	(7,11), (9,16), (15,33), (19,24)
0,001	(7,11), (9,16), (15,33), (19,24)
0,0025	(7,11), (9,16), (15,33), (19,24)
0,005	(1,15,33), (4,32), (7,11), (9,16,23), (19,24,25)
0,0075	(1,15,33), (4,32), (7,11), (9,16,23), (19,24,25)
0,01	(1,15,33), (4,32), (7,11), (9,16,23), (19,24,25)
0,025	(1,15,33), (2,12,30), (3,4,32), (7,11), (9,16,23), (19,24,25)
0,05	(1,2,7,9 a 13, 15,16,23,26,28,30,33), (3,4,32), (19,24,25)
0,075	(1,2,6,7,9 a 16, 19 a 26, 28,30,33), (3,4,32)
0,1	(1 a 4, 6 a 17, 19 a 28, 30 a 36)

**Figura VI.58** - Aglomerados de descritores interdependentes para a classe 3 (Com lesões na Descendente Anterior isolada) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	-
0,00075	-
0,001	-
0,0025	-
0,005	-
0,0075	-
0,01	-
0,025	(2,4,7), (5,30), (19,33), (25,31), (27,28,34)
0,05	(1,20,25,31), (2,4,7), (5,30), (6, 27,28,34), (19,33)
0,075	(1,20,25,31), (2,4,7), (5,30), (6,12,27,28,34), (8,9,11,15,16,21,22,23,24,26,29,32,35,36), (19,33)
0,01	(1,3,5,6,8 a 18, 20 a 32,34 a 36), (2,4,7), (19,33)

**Figura VI.59** - Aglomerados de descritores interdependentes para a classe 4 (com lesões na Coronária Circunflexa isolada) da variável Anatomia Coronariana.



NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	(2,19), (21,30)
0,00075	(2,19), (21,30)
0,001	(1,28), (2,19), (21,30)
0,0025	(1,28), (2,19), (21,30)
0,005	(1,28), (2,19), (21,30)
0,0075	(1,28), (2,3,19,33), (7,9), (8,16), (21,30)
0,01	(1,28), (2,3,19,33), (7,9), (8,16), (21,30)
0,025	(1,4,7,8,9,16,23,28), (2,3,19,33), (6,11), (21,30), (27,32)
0,05	(1 a 4, 7,8,9,16,19,23 a 28,32,33), (6,11,18,31), (10,17), (21,30)
0,075	(1 a 4, 6 a 9, 11,13,16,18,19,23 a 28, 31,32,33), (10,17), (21,30)
0,1	(1 a 36)

**Figura VI.60** - Aglomerados de descritores interdependentes para a classe 5 (Com lesões na Coronária Direita e Descendente Anterior) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	-
0,00075	-
0,001	-
0,0025	(2,30,31), (4,29), (16,34)
0,005	(2,30,31), (4,29), (16,34)
0,0075	(2,30,31), (3,9), (4,29), (16,34)
0,01	(2,30,31), (3,9), (4,29), (16,34)
0,025	(1,3,9), (2,30,31), (4,17,29), (5,27), (16,34)
0,05	(1,3,4,7,9,11,12,13,17,20,23,24,29,33), (2,30,31), (5,27), (16,34), (18,19)
0,075	(1 a 7, 9 a 15, 17, 20 a 24, 27,29,30,31,33), (16,34), (18,19)
0,01	(1 a 36)

**Figura VI.61** - Aglomerados de descritores interdependentes para a classe 6 (Com lesões na Coronária Direita e Circunflexa) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	-
0,00075	-
0,001	-
0,0025	(2,32)
0,005	(2,32), (5,24)
0,0075	(,32), (5,24,25)
0,01	(2,32), (5,24,25), (6,12), (9,15)
0,025	(2,32), (4,6,12,31), (5,9,11,15,20,21,23,24,25,27), (17,28), (30,33)
0,05	(1,3 a 6, 9 a 12, 14,15,17,20 a 25, 27,28,31), (2,32), (30,33)
0,075	(1,3 a 6, 9 a 15, 17, 20 a 25, 27,28,31), (2, 18,19,20,32), (30,33)
0,1	(1 a 15, 17 a 33, 35,36)

**Figura VI.62** - Aglomerados de descritores interdependentes para a classe 7 (Com lesões na Descendente Anterior e Circunflexa) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	-
0,00075	-
0,001	-
0,0025	(1,2), (15,32)
0,005	(1,2,8), (15,32), (16,29)
0,0075	(1,2,8), (12,34), (15,32), (16,29), (20,31)
0,01	(1,2,8), (12,34), (15,32), (16,29), (20,31)
0,025	(1,2,8), (7,18), (12,34), (13,16,29), (14,20,31), (15,32)
0,05	(1,2,8,15,32), (4,13,16,29,30), (6,10,11,12, 14,17,20,21,24,31,33,34), (7,18)
0,075	(1,2,4 a 21, 24,25,27,29 a 34)
0,1	(1 a 36)

**Figura VI.63** - Aglomerados de descritores interdependentes para a classe 8 (Com lesões nos 3 vasos) da variável Anatomia Coronariana.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	(5,27), (9,36)
0,00075	(5,27), (9,36)
0,001	(5,27), (7,9,36), (8,28)
0,0025	(5,27,29), (7,9,36), (8,28)
0,005	(4,16), (5,27,29), (7,9,36), (8,28)
0,0075	(4,16), (5,27,29,30), (7,9,36), (8,28), (18,19)
0,01	(2,7,9,10,33,36), (4,16), (5,27,29,30), (8,28), (18,19)
0,025	(1,2,4,7 a 11, 16, 28,33,36), (5,27,29,30,35), (13,18,19)
0,05	(1,2,4 a 13,15,16,18,19,20,23,27 a 30, 33 a 36), (14,22), (17,21,24,25,26,31)
0,075	(1 a 31,33 a 36)
0,1	(1 a 31, 33 a 36)

**Figura VI.64** – Aglomerados de descritores interdependentes para a classe diagnóstica Função Ventricular Normal.

NÍVEL DE SIGNIFICÂNCIA	AGLOMERADOS FORMADOS POR MAIS QUE UM DESCRITOR
0,0005	(2,8), (6,12), (15,32), (24,25)
0,00075	(2,8), (6,12), (9,15,32), (22,24,25)
0,001	(2,8), (6,12), (9,15,32), (22,24,25)
0,0025	(2,8), (6,12), (9,15,32), (20,31), (22,24,25)
0,005	(2,8), (6,12), (9,15,32), (20,31), (22,24,25)
0,0075	(2,7,8,18), (6,12,22,23,24,25), (9,15,32), (20,31)
0,01	(1,2,6,7,8,12,18,22 a 26), (9,15,32), (20,31)
0,025	(1,2,3,5 a 20, 22 a 27,31,32,33)
0,05	(1 a 27,31 a 34)
0,075	(1 a 34)
0,1	(1 a 36)

**Figura VI.65** - Aglomerados de descritores interdependentes para a classe diagnóstica Função Ventricular Anormal.

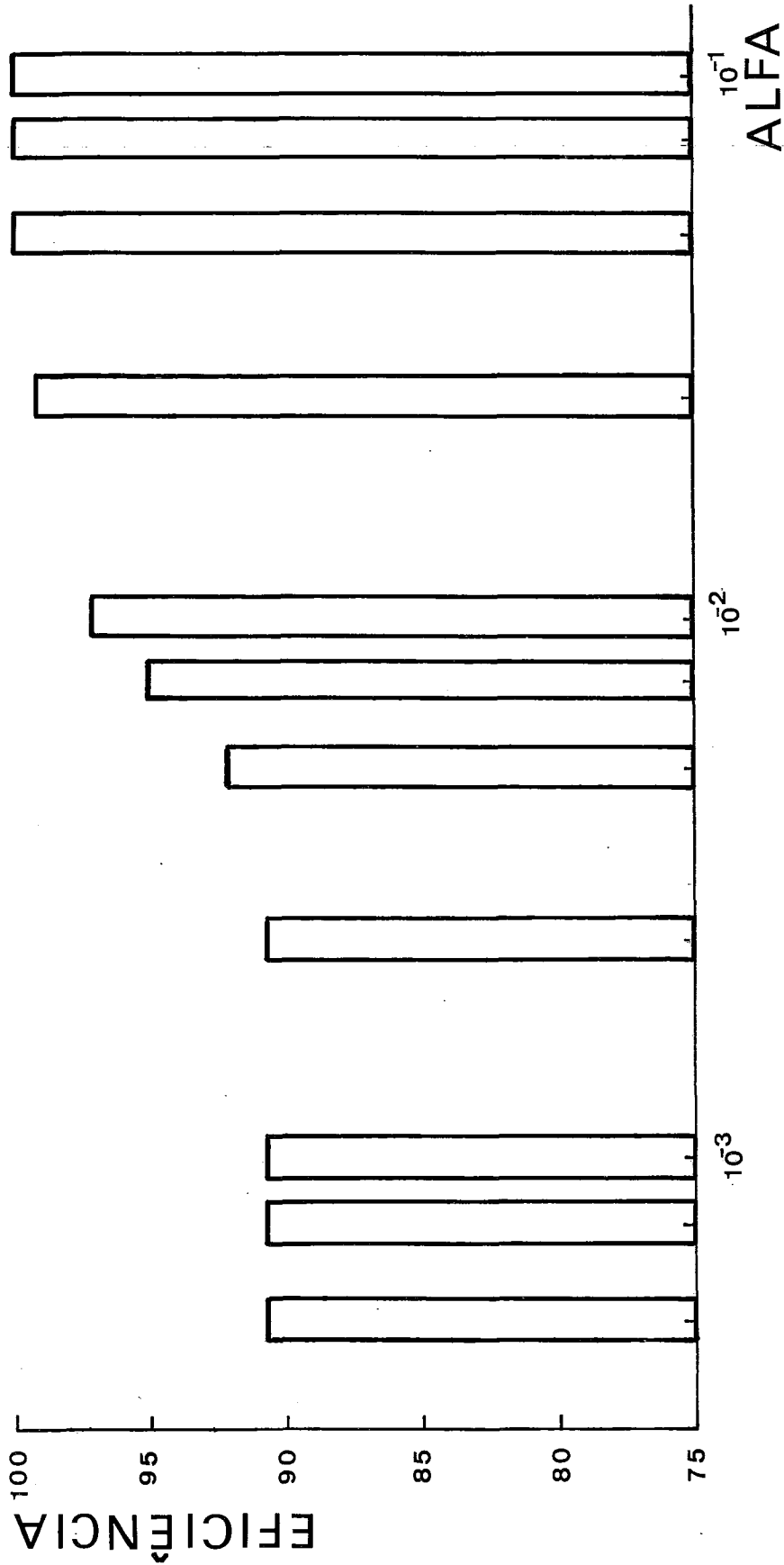


Figura VI.66 - Eficiência do modelo Presença de Doença Coronariana em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade "0 Banco de Dados como Universo", probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ )

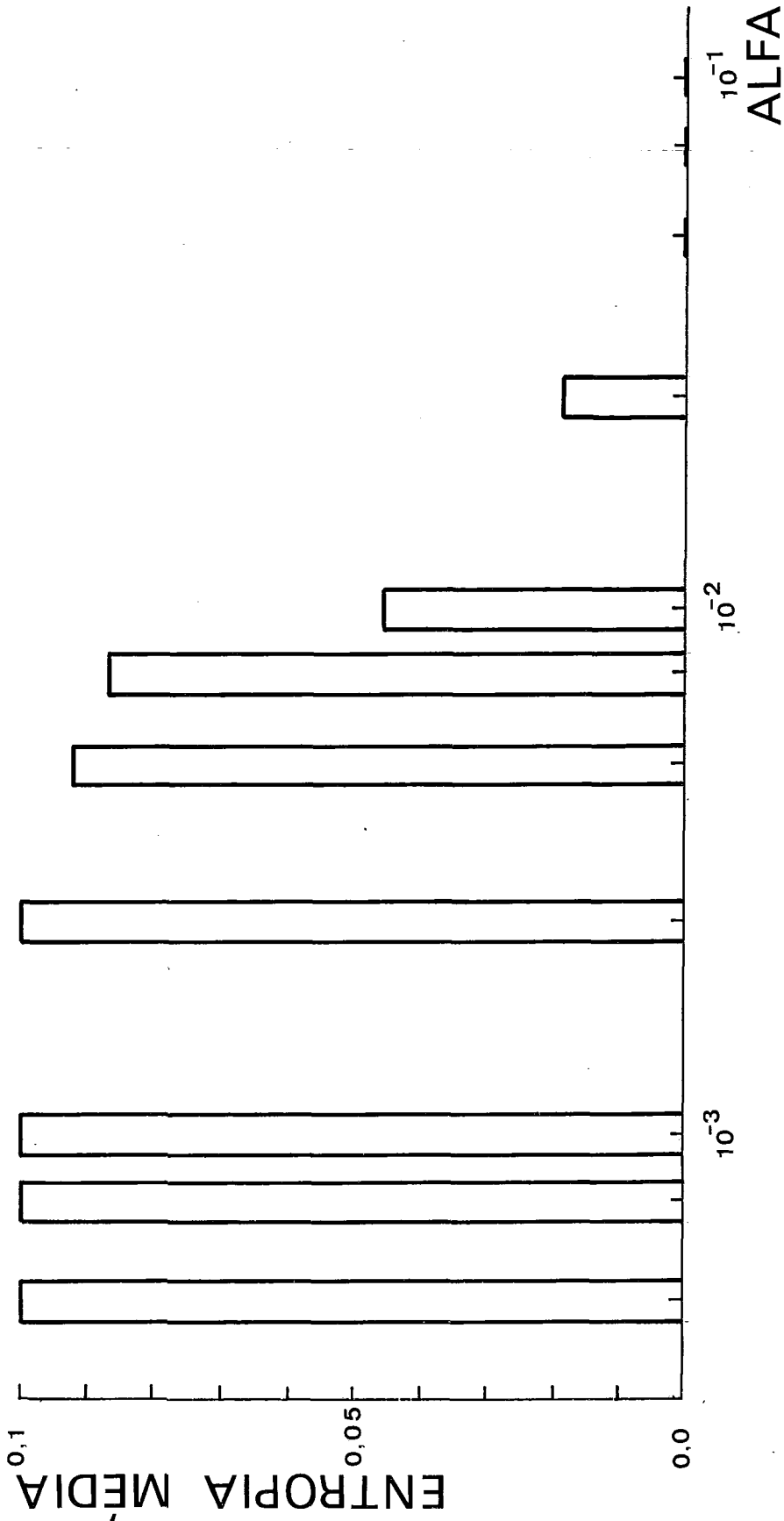


Figura VI.67 - Entropia média do modelo Presença de Doença Coronariana em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade "O Banco de Dados como Universo", probabilidades a priori estimadas do banco de dados,  $p_{min} = 10^{-6}$ ).



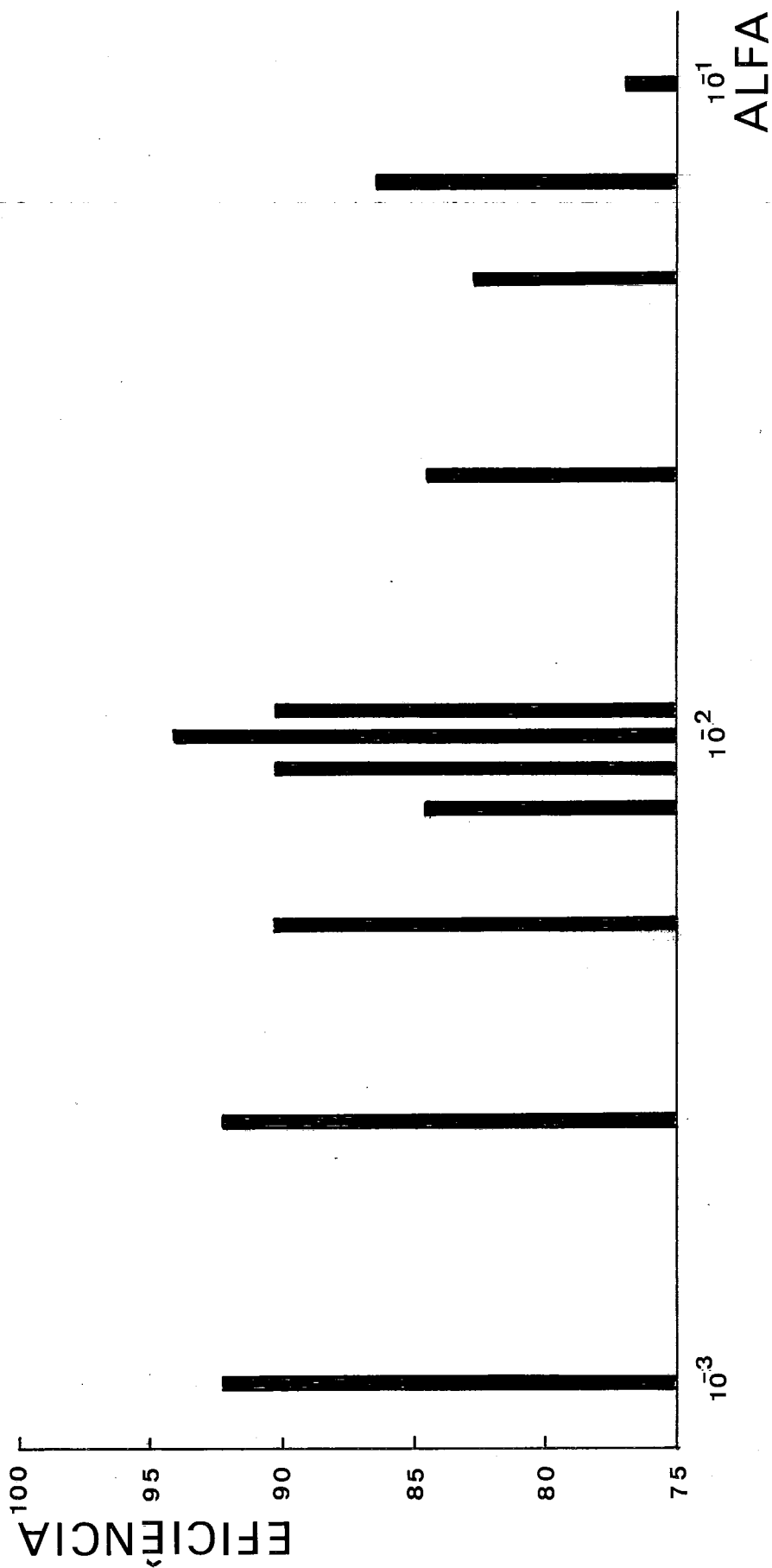


Figura VI.68 - Eficiência do modelo Presença de Doença Coronariana em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade: "0 Banco de Dados como Amostra", probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ ).

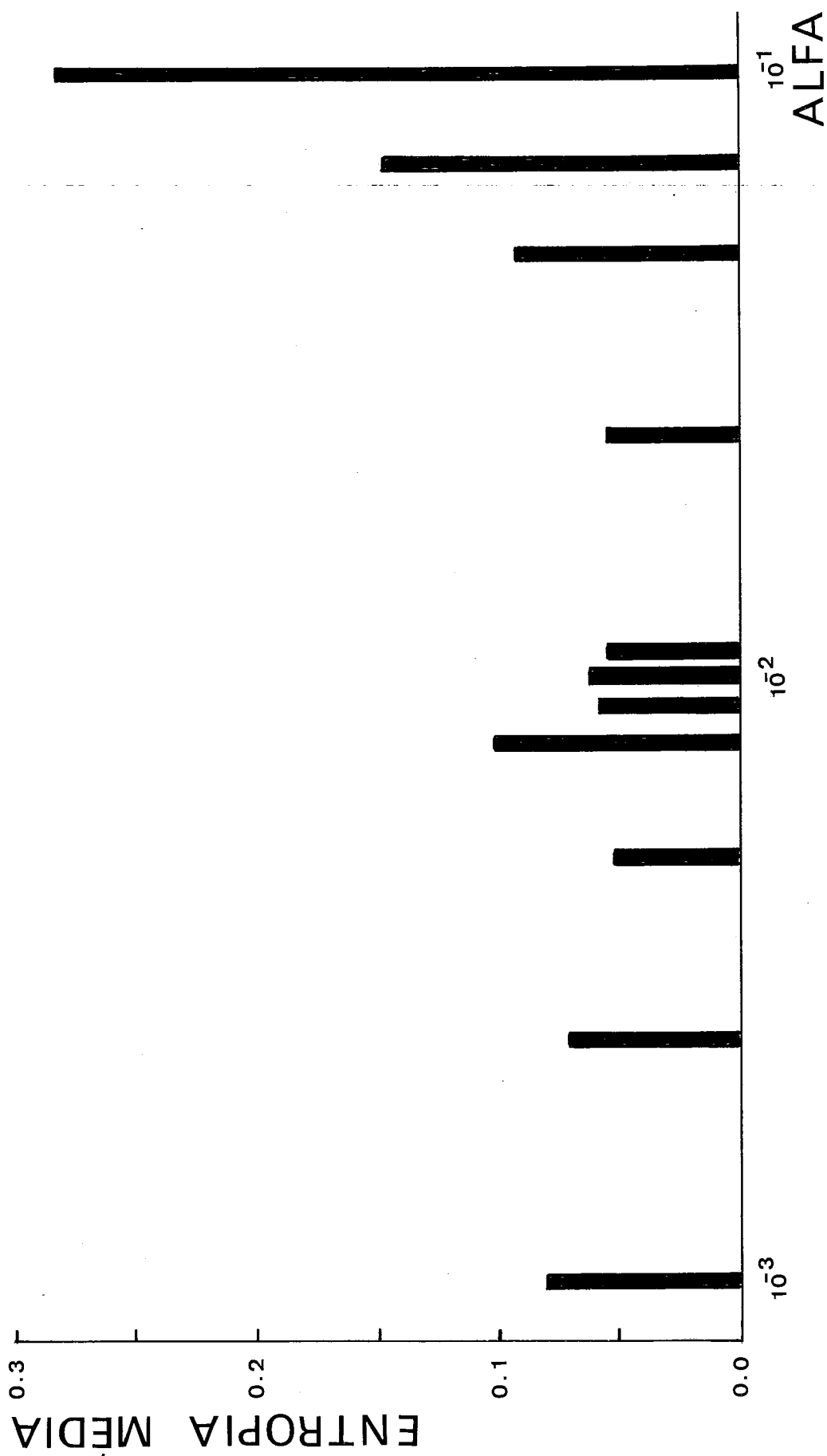


Figura VI.69 - Entropia média do modelo Presença de Doença Coronariana em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade "0 Banco de Dados como Amostra", probabilidades a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ ).

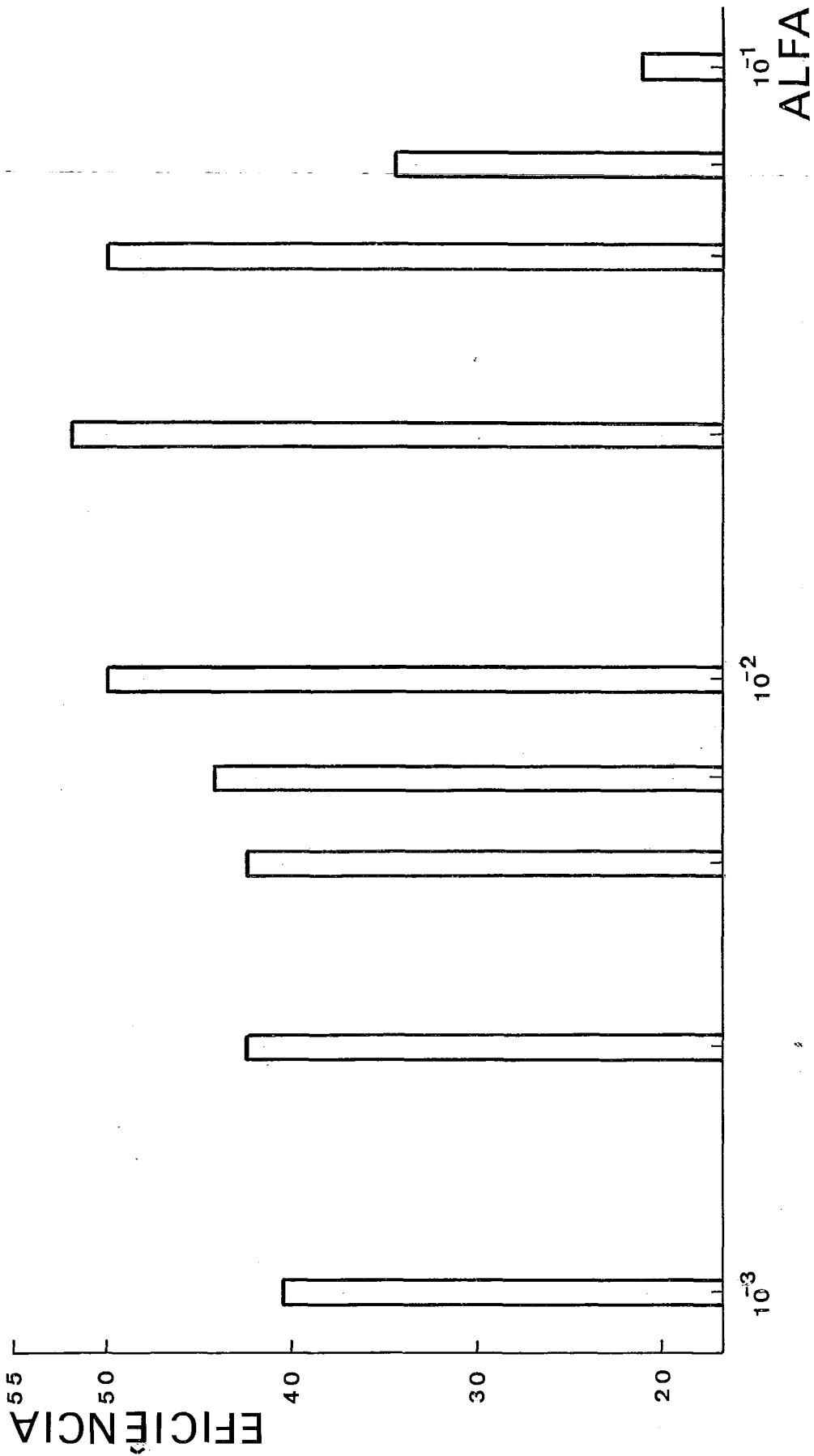


Figura VI.70 - Eficiência do modelo Anatomia Coronariana em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade "0 Banco de Dados como Amostra", probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ ).

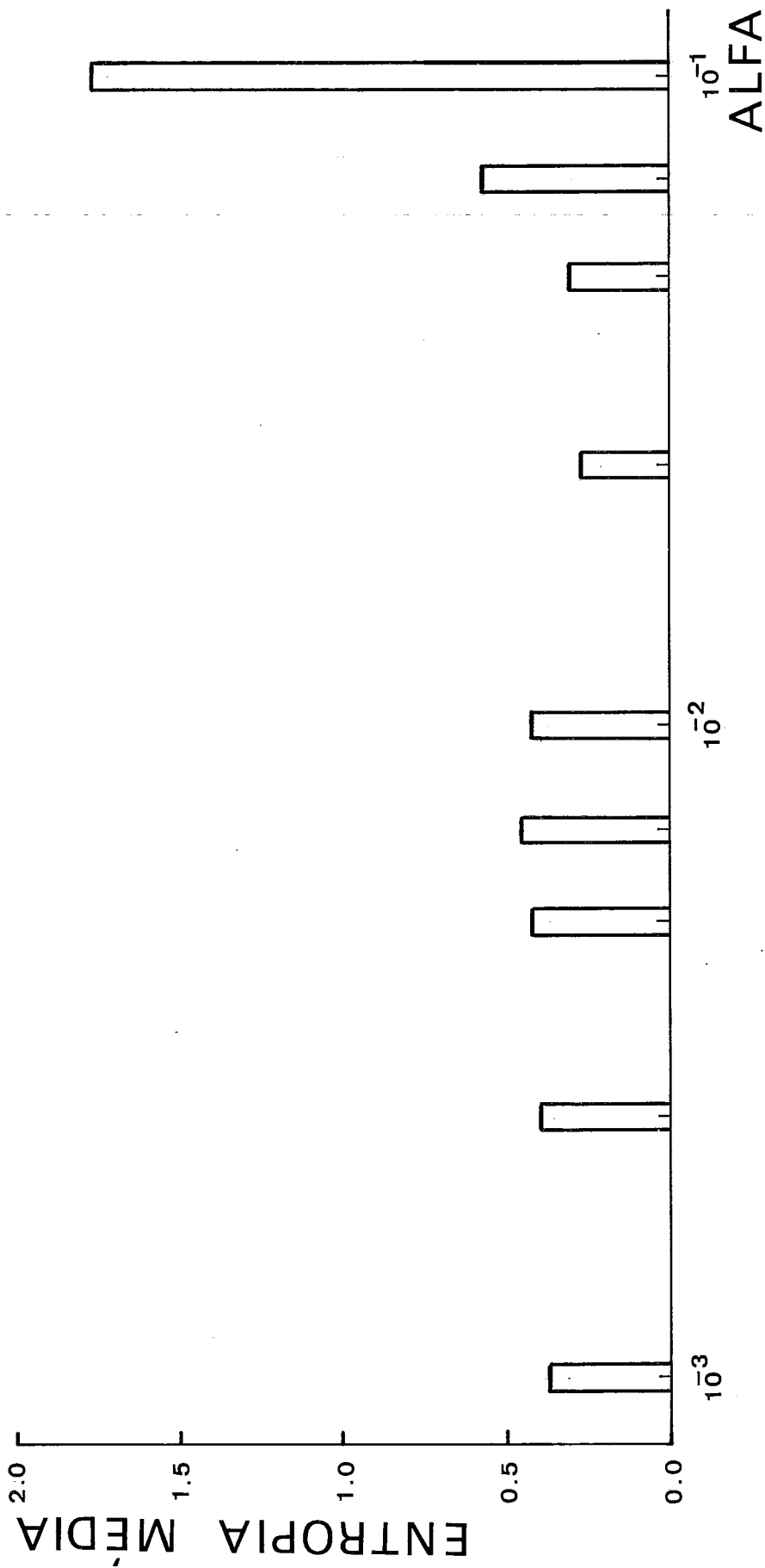


Figura VI.71 - Entropia média do modelo Anatomia Coronariana em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade "0 Banco de Dados como Amostra", probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ ).

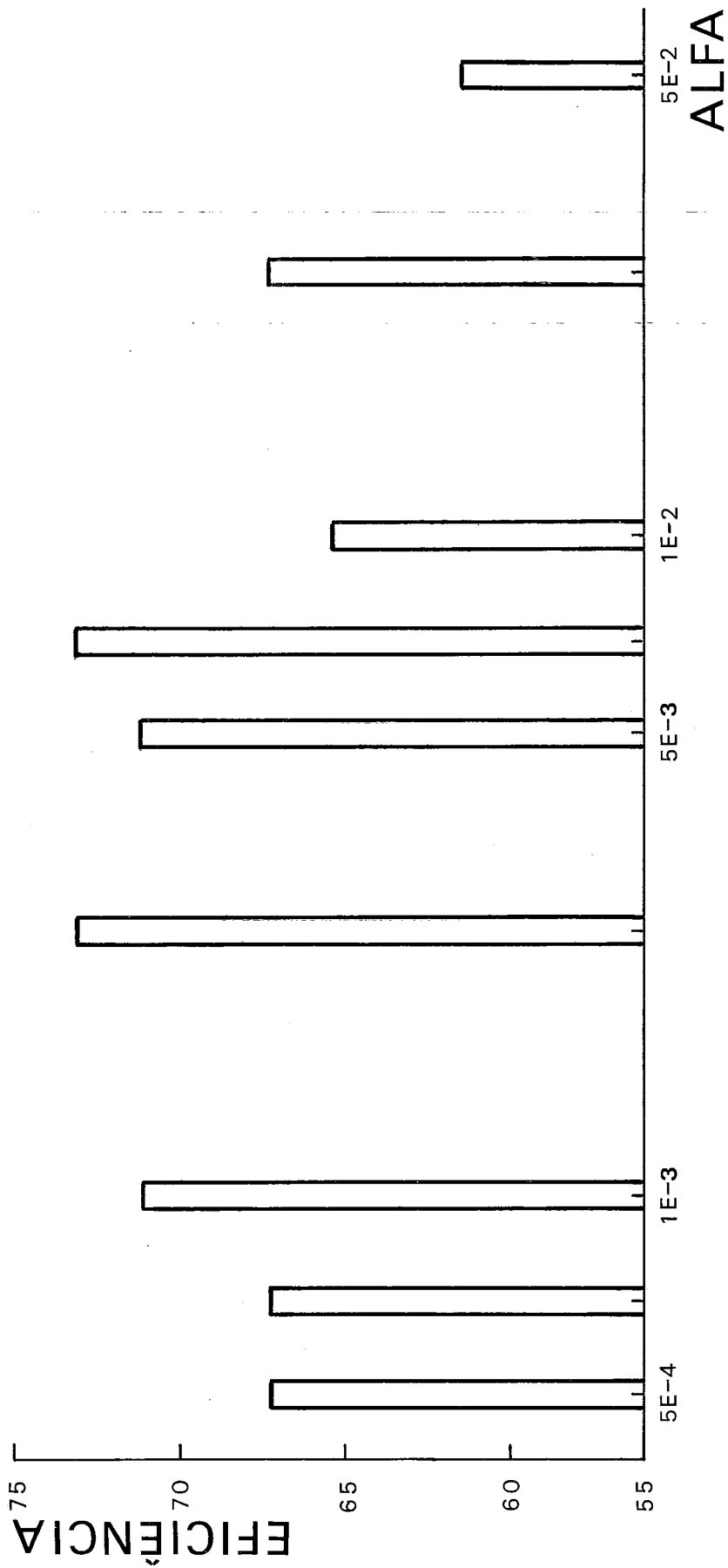


Figura VI.72 - Eficiência do modelo Função Ventricular em função do nível de significância usa do para definição de aglomerados de descritores. (Modalidade "0 Branco de Dads" como Amostra", probabilidades a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ ).

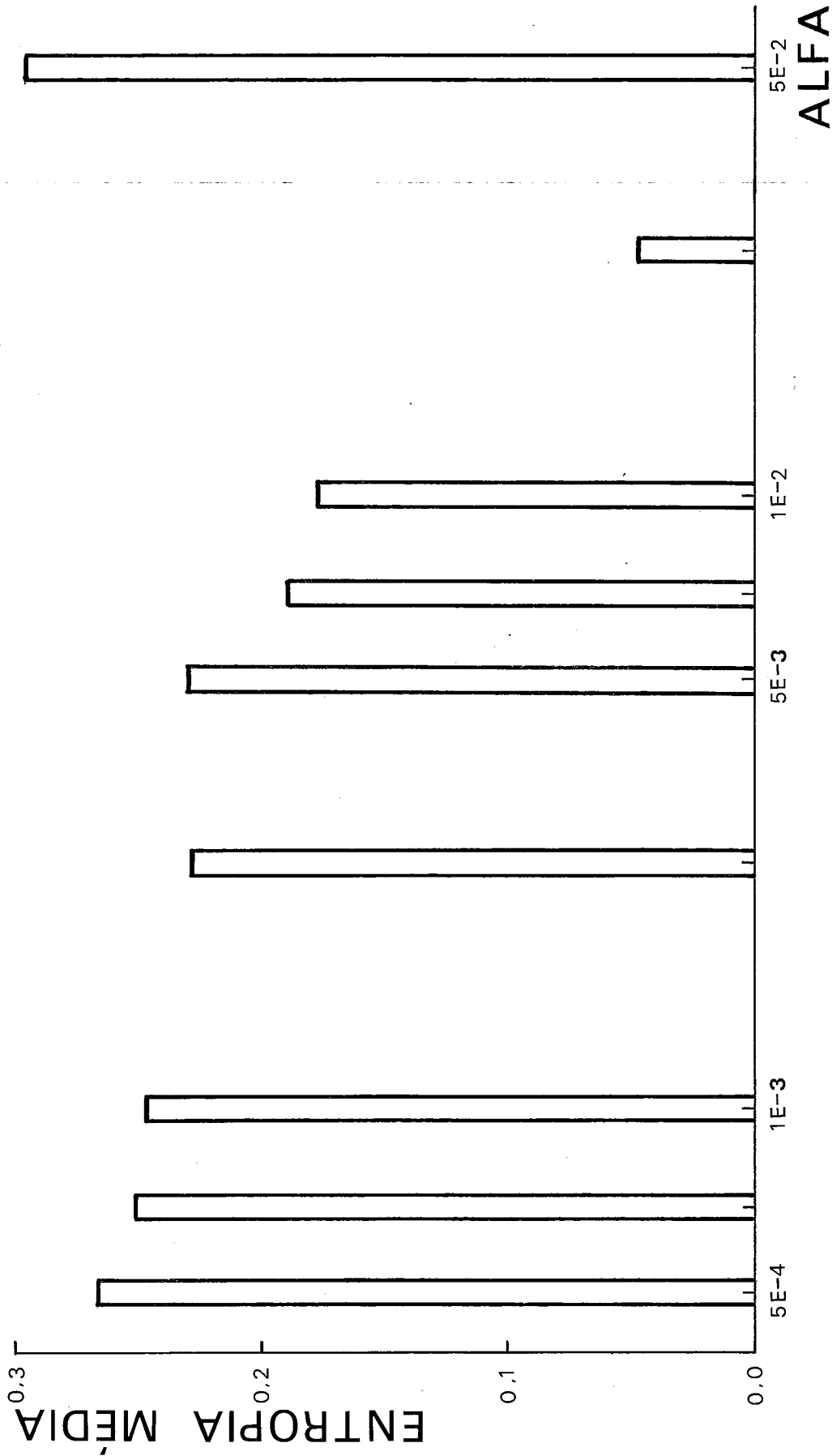


Figura VI.73 - Entropia média do modelo Função Ventricular em função do nível de significância usado para definição de aglomerados de descritores. (Modalidade "0 Banco de Dados como Amostra", probabilidades a priori estimadas do banco de dados,  $p_{min}=10^{-6}$ ).

MODELOS	Pressuposição de Independência			Aglomerados de Descritores Interdependentes		
	Eficiência	Entropia Média	X <sup>2</sup> Tab. Contingência	Eficiência	Entropia Média	X <sup>2</sup> Tab. Contingência
Presença de Doença Coronariana	92,31	0,080083	29,4360 ***	94,23	0,063838	37,1428 ***
Anatomia Coronariana	34,62	0,548118	55,2040 -	51,92	0,276254	132,6102 ***
Função Ventricular	65,38	0,264112	4,87653 *	73,08	0,189881	11,18109 ***

- \* Significativo ao nível  $\alpha = 5\%$
- \*\* Significativo ao nível  $\alpha = 1\%$
- \*\*\* Significativo ao nível  $\alpha = 0,1\%$

Figura VI.74 - Comparação dos desempenhos do modelo de classificação diagnóstica baseado na pressuposição de independência e do modelo baseado em aglomerados de descritores interdependentes.  
(Modalidade: "O Banco de Dados como Amostra", probabilidades a priori estimadas do banco de dados,  $P_{\min} = 10^{-6}$ ).

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 1
MODELO: PRES LESAO ALFA=0.01
*****
AGLOMERADO [1]: (2)=0.2903225806 / (3)=0.1612903226 /
(1)=0.5483870968
AGLOMERADO [33 2]: (0 1)=0.4838709677 / (0 2)=0.3225806452 /
(1 2)=0.1935483871
AGLOMERADO [3]: (1)=0.06451612903 / (3)=0.06451612903 /
(2)=0.8709677419
AGLOMERADO [4]: (2)=0.4193548387 / (1)=0.5806451613
AGLOMERADO [36 32 27 26 25 21 20 19 15 10 9 8 5]:
(0 0 1 0 0 1 0 1 0 0 0 0 0)=0.1666666667 /
(1 0 0 0 0 1 0 0 0 1 0 1 0)=0.1666666667 /
(0 0 0 0 0 1 0 0 0 0 0 0 0)=0.3333333333 /
(1 0 0 0 0 2 1 1 0 1 1 0 0)=0.1666666667 /
(0 0 0 0 0 1 0 0 1 1 0 0 0)=0.1666666667
AGLOMERADO [6]: (3)=0.06666666667 / (2)=0.8 /
(1)=0.13333333333
AGLOMERADO [7]: (1)=0.2903225806 / (0)=0.7096774194
AGLOMERADO [11]: (1)=0.4838709677 / (0)=0.5161290323
AGLOMERADO [12]: (2)=0.3448275862 / (3)=0.1034482759 /
(1)=0.5517241379
AGLOMERADO [13]: (2)=0.2903225806 / (1)=0.7096774194
AGLOMERADO [14]: (1)=0.43333333333 / (0)=0.5666666667
AGLOMERADO [16]: (0)=0.4193548387 / (1)=0.5806451613
AGLOMERADO [17]: (1)=0.15 / (0)=0.85
AGLOMERADO [18]: (0)=0.3076923077 / (1)=0.6923076923
AGLOMERADO [22]: (1)=0.1666666667 / (2)=0.8333333333
AGLOMERADO [23]: (1)=0.09677419355 / (0)=0.9032258065
AGLOMERADO [24]: (1)=0.03225806452 / (0)=0.9677419355
AGLOMERADO [28]: (0)=1
AGLOMERADO [29]: (0)=1
AGLOMERADO [30]: (0)=1
AGLOMERADO [31]: (1)=0.1612903226 / (0)=0.8387096774
AGLOMERADO [34]: (0)=0.8387096774 / (1)=0.1612903226
AGLOMERADO [35]: (0)=1
*****

```

Figura VI.75 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 1 do modelo Presença de Doença Coronariana  $\alpha = 0,01$ .



```

*****
AGLOMERADO [28 26 10 8 7 2 1]:
(0 0 1 0 1 1 2)=0.009708737864 /
(0 0 0 0 1 2 1)=0.01941747573 /
(0 0 0 0 0 2 1)=0.03883495146 /
(0 1 1 0 1 1 1)=0.009708737864 /
(1 0 1 0 0 1 1)=0.009708737864 /
(0 0 1 0 0 1 3)=0.04854368932 /
(0 0 1 0 0 2 3)=0.01941747573 /
(0 0 1 0 1 2 3)=0.01941747573 /
(0 0 1 0 0 2 1)=0.009708737864 /
(0 1 0 0 0 2 1)=0.009708737864 /
(0 0 0 0 1 1 1)=0.03883495146 /
(1 0 0 0 0 1 1)=0.009708737864 /
(0 0 0 0 0 2 3)=0.01941747573 /
(0 0 1 1 1 2 1)=0.01941747573 /
(1 0 0 1 1 1 2)=0.009708737864 /
(0 0 1 0 1 2 1)=0.01941747573 /
(0 0 1 0 1 1 1)=0.03883495146 /
(0 0 0 0 0 1 3)=0.03883495146 /
(1 0 1 0 1 2 1)=0.009708737864 /
(0 0 0 0 0 1 1)=0.4174757282 / (0 0 1 0 0 1 1)=0.1844660194
AGLOMERADO [3]: (3)=0.1909090909 / (2)=0.7454545455 /
(1)=0.06363636364
AGLOMERADO [16 4]: (0 1)=0.1111111111 / (0 2)=0.06481481481 /
(1 2)=0.09259259259 / (1 1)=0.7314814815
AGLOMERADO [31 27 25 24 22 20 17 12 6 5]:
(0 0 0 0 2 1 0 2 3 1)=0.1111111111 /
(0 0 0 0 2 0 1 2 1 0)=0.1111111111 /
(0 0 0 0 2 0 0 2 3 0)=0.1111111111 /
(0 0 0 0 2 0 0 2 2 1)=0.1111111111 /
(0 0 0 0 2 0 1 1 2 0)=0.1111111111 /
(1 0 1 0 2 1 1 1 3 0)=0.1111111111 /
(0 1 0 0 2 0 0 2 1 1)=0.1111111111 /
(1 0 0 0 2 0 0 1 3 0)=0.1111111111 /
(1 0 1 1 2 1 1 2 1 0)=0.1111111111
AGLOMERADO [32 15 9]: (1 1 0)=0.01 / (1 0 0)=0.03 /
(1 1 1)=0.06 / (0 0 1)=0.13 / (0 1 0)=0.13 / (0 0 0)=0.59 /
(0 1 1)=0.05
AGLOMERADO [11]: (1)=0.4128440367 / (0)=0.5871559633
AGLOMERADO [13]: (2)=0.320754717 / (1)=0.679245283
AGLOMERADO [14]: (0)=0.5700934579 / (1)=0.4299065421
AGLOMERADO [18]: (1)=0.5925925926 / (0)=0.4074074074
AGLOMERADO [19]: (0)=0.7765957447 / (1)=0.2234042553
AGLOMERADO [21]: (4)=0.01041666667 / (3)=0.02083333333 /
(1)=0.7708333333 / (2)=0.1979166667
AGLOMERADO [23]: (1)=0.06363636364 / (0)=0.9363636364
AGLOMERADO [29]: (1)=0.2 / (0)=0.8
AGLOMERADO [30]: (1)=0.1545454545 / (0)=0.8454545455
AGLOMERADO [33]: (1)=0.1090909091 / (0)=0.8909090909
AGLOMERADO [34]: (1)=0.1090909091 / (0)=0.8909090909
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (1)=0.03846153846 / (0)=0.9615384615
*****

```

Figura VI.76 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 2 do modelo Presença de Doença Coronariana  $\alpha = 0,01$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 1
MODELO: ANATOMIA ALFA=.025
*****
AGLOMERADO [1]: (2)=0.2380952381 / (3)=0.09523809524 /
(1)=0.6666666667
AGLOMERADO [34 33 2]: (0 0 1)=0.4285714286 /
(0 1 2)=0.1428571429 / (0 0 2)=0.2857142857 /
(1 1 2)=0.1428571429
AGLOMERADO [3]: (3)=0.04761904762 / (1)=0.04761904762 /
(2)=0.9047619048
AGLOMERADO [4]: (2)=0.4761904762 / (1)=0.5238095238
AGLOMERADO [36 32 27 26 25 21 20 19 16 15 10 9 8 7 5]:
(0 0 1 0 0 1 0 1 0 0 0 0 0 0 0)=0.2 /
(1 0 0 0 0 1 0 0 1 0 1 0 1 0 0)=0.2 /
(0 0 0 0 0 1 0 0 0 0 0 0 0 0 0)=0.2 /
(1 0 0 0 0 2 1 1 1 0 1 1 0 0 0)=0.2 /
(0 0 0 0 0 1 0 0 1 1 1 0 0 0 0)=0.2
AGLOMERADO [6]: (2)=0.85 / (1)=0.15
AGLOMERADO [11]: (1)=0.5238095238 / (0)=0.4761904762
AGLOMERADO [12]: (2)=0.2631578947 / (3)=0.1052631579 /
(1)=0.6315789474
AGLOMERADO [13]: (2)=0.2857142857 / (1)=0.7142857143
AGLOMERADO [14]: (1)=0.4 / (0)=0.6
AGLOMERADO [17]: (1)=0.07692307692 / (0)=0.9230769231
AGLOMERADO [18]: (0)=0.1428571429 / (1)=0.8571428571
AGLOMERADO [31 24 22]: (0 0 1)=0.1 / (1 0 1)=0.1 /
(0 0 2)=0.8
AGLOMERADO [23]: (1)=0.04761904762 / (0)=0.9523809524
AGLOMERADO [28]: (0)=1
AGLOMERADO [29]: (0)=1
AGLOMERADO [30]: (0)=1
AGLOMERADO [35]: (0)=1
*****

```

Figura VI.77 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 1 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 2
MODELO: ANATOMIA ALFA=.025
*****
AGLOMERADO [4 2 1]: (2 2 3)=0.3333333333 /
(1 1 1)=0.6666666667
AGLOMERADO [3]: (3)=0.2857142857 / (1)=0.1428571429 /
(2)=0.5714285714
AGLOMERADO [5]: (1)=0.1428571429 / (0)=0.8571428571
AGLOMERADO [21 11 6]: (1 0 2)=0.5 / (1 1 2)=0.1666666667 /
(1 0 4)=0.1666666667 / (2 1 3)=0.1666666667
AGLOMERADO [7]: (1)=0.4285714286 / (0)=0.5714285714
AGLOMERADO [26 17 8]: (1 1 1)=0.2 / (0 0 0)=0.8
AGLOMERADO [34 32 9]: (0 0 0)=1
AGLOMERADO [14 10]: (1 1)=0.1428571429 / (0 0)=0.7142857143 /
(1 0)=0.1428571429
AGLOMERADO [12]: (2)=0.5 / (1)=0.5
AGLOMERADO [13]: (2)=0.1666666667 / (1)=0.8333333333
AGLOMERADO [15]: (1)=0.2857142857 / (0)=0.7142857143
AGLOMERADO [16]: (0)=0.1428571429 / (1)=0.8571428571
AGLOMERADO [18]: (0)=0.5 / (1)=0.5
AGLOMERADO [19]: (0)=0.7142857143 / (1)=0.2857142857
AGLOMERADO [23 20]: (1 1)=0.3333333333 / (0 0)=0.6666666667
AGLOMERADO [22]: (2)=1
AGLOMERADO [24]: (0)=1
AGLOMERADO [25]: (0)=1
AGLOMERADO [27]: (1)=0.2857142857 / (0)=0.7142857143
AGLOMERADO [28]: (0)=1
AGLOMERADO [29]: (0)=1
AGLOMERADO [30]: (0)=1
AGLOMERADO [31]: (0)=0.8571428571 / (1)=0.1428571429
AGLOMERADO [33]: (0)=1
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****

```

Figura VI.78 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 2 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 3
MODELO: ANATOMIA   ALFA=.025
*****
AGLOMERADO [33 15 1]: (1 1 3)=0.07692307692 /
(0 0 3)=0.07692307692 / (0 0 1)=0.8461538462
AGLOMERADO [30 12 2]: (0 2 2)=0.07142857143 /
(0 3 2)=0.07142857143 / (1 1 1)=0.07142857143 /
(0 2 1)=0.2142857143 / (0 1 1)=0.07142857143 /
(1 2 1)=0.2857142857 / (0 1 2)=0.2142857143
AGLOMERADO [32 4 3]: (1 2 2)=0.06666666667 /
(0 2 1)=0.06666666667 / (0 1 3)=0.1333333333 /
(0 1 2)=0.7333333333
AGLOMERADO [5]: (0)=0.6666666667 / (1)=0.3333333333
AGLOMERADO [6]: (1)=0.1428571429 / (3)=0.2142857143 /
(2)=0.6428571429
AGLOMERADO [11 7]: (0 0)=0.7857142857 / (1 1)=0.2142857143
AGLOMERADO [8]: (0)=1
AGLOMERADO [23 16 9]: (0 1 1)=0.06666666667 /
(1 0 1)=0.06666666667 / (0 1 0)=0.8666666667
AGLOMERADO [10]: (0)=0.7333333333 / (1)=0.2666666667
AGLOMERADO [13]: (1)=0.8 / (2)=0.2
AGLOMERADO [14]: (1)=0.4 / (0)=0.6
AGLOMERADO [17]: (0)=0.8333333333 / (1)=0.1666666667
AGLOMERADO [18]: (0)=0.3333333333 / (1)=0.6666666667
AGLOMERADO [25 24 19]: (1 1 1)=0.07692307692 /
(1 0 0)=0.07692307692 / (0 0 0)=0.8461538462
AGLOMERADO [20]: (0)=1
AGLOMERADO [21]: (2)=0.07692307692 / (1)=0.9230769231
AGLOMERADO [22]: (1)=0.2857142857 / (2)=0.7142857143
AGLOMERADO [26]: (0)=1
AGLOMERADO [27]: (1)=0.2 / (0)=0.8
AGLOMERADO [28]: (0)=0.9333333333 / (1)=0.0666666667
AGLOMERADO [29]: (0)=0.6666666667 / (1)=0.3333333333
AGLOMERADO [31]: (0)=1
AGLOMERADO [34]: (0)=1
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****

```

Figura VI.79 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 3 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```
*****  
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS  
CONJUNTAS DA CLASSE 4  
MODELO: ANATOMIA -- ALFA=,025  
*****  
AGLOMERADO [1]: (3)=0.4 / (1)=0.6  
AGLOMERADO [7 4 2]: (0 1 1)=0.6 / (1 2 2)=0.4  
AGLOMERADO [3]: (2)=1  
AGLOMERADO [30 5]: (1 1)=0.2 / (0 0)=0.8  
AGLOMERADO [6]: (1)=0.6 / (3)=0.4  
AGLOMERADO [8]: (0)=1  
AGLOMERADO [9]: (0)=1  
AGLOMERADO [10]: (0)=0.8 / (1)=0.2  
AGLOMERADO [11]: (0)=0.2 / (1)=0.8  
AGLOMERADO [12]: (2)=0.6 / (1)=0.4  
AGLOMERADO [13]: (1)=0.6 / (2)=0.4  
AGLOMERADO [14]: (1)=0.25 / (0)=0.75  
AGLOMERADO [15]: (0)=1  
AGLOMERADO [16]: (1)=1  
AGLOMERADO [17]: (1)=0.4 / (0)=0.6  
AGLOMERADO [18]: (1)=0.66666666667 / (0)=0.33333333333  
AGLOMERADO [33 19]: (1 1)=0.2 / (0 0)=0.8  
AGLOMERADO [20]: (1)=0.33333333333 / (0)=0.66666666667  
AGLOMERADO [21]: (1)=1  
AGLOMERADO [22]: (2)=1  
AGLOMERADO [23]: (0)=1  
AGLOMERADO [24]: (0)=1  
AGLOMERADO [31 25]: (1 1)=0.2 / (0 0)=0.8  
AGLOMERADO [26]: (0)=1  
AGLOMERADO [34 28 27]: (0 0 0)=1  
AGLOMERADO [29]: (0)=1  
AGLOMERADO [32]: (0)=1  
AGLOMERADO [35]: (1)=1  
AGLOMERADO [36]: (0)=1  
*****
```

Figura VI.80 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 4 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 5
MODELO: ANATOMIA   ALFA=.025
*****
AGLOMERADO [28 23 16 9 8 7 4 1]:
(0 0 1 1 0 1 1 1)=0.1111111111 /
(0 0 1 0 0 1 1 3)=0.1111111111 /
(0 1 0 0 0 0 2 1)=0.1111111111 /
(0 0 0 0 1 1 1 1)=0.1111111111 /
(0 0 1 0 0 0 1 1)=0.5555555556
AGLOMERADO [33 19 3 2]: (0 1 2 2)=0.1111111111 /
(1 1 3 2)=0.2222222222 / (0 0 2 1)=0.5555555556 /
(1 0 2 1)=0.1111111111
AGLOMERADO [5]: (1)=0.3 / (0)=0.7
AGLOMERADO [11 6]: (1 2)=0.4444444444 / (0 2)=0.1111111111 /
(0 3)=0.4444444444
AGLOMERADO [10]: (0)=0.2 / (1)=0.8
AGLOMERADO [12]: (1)=0.5 / (2)=0.5
AGLOMERADO [13]: (2)=0.3333333333 / (1)=0.6666666667
AGLOMERADO [14]: (1)=0.4 / (0)=0.6
AGLOMERADO [15]: (1)=0.3333333333 / (0)=0.6666666667
AGLOMERADO [17]: (1)=0.25 / (0)=0.75
AGLOMERADO [18]: (1)=0.75 / (0)=0.25
AGLOMERADO [20]: (0)=1
AGLOMERADO [30 21]: (0 1)=1
AGLOMERADO [22]: (1)=0.25 / (2)=0.75
AGLOMERADO [24]: (1)=0.2 / (0)=0.8
AGLOMERADO [25]: (1)=0.2 / (0)=0.8
AGLOMERADO [26]: (1)=0.1 / (0)=0.9
AGLOMERADO [32 27]: (1 1)=0.1 / (0 0)=0.9
AGLOMERADO [29]: (0)=1
AGLOMERADO [31]: (0)=0.9 / (1)=0.1
AGLOMERADO [34]: (0)=1
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****

```

Figura VI.81 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 5 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```
*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 6
MODELO: ANATOMIA   ALFA=.025
*****
AGLOMERADO [9 3 1]: (0 3 2)=0.1666666667 /
(0 2 1)=0.8333333333
AGLOMERADO [31 30 2]: (0 0 1)=1
AGLOMERADO [29 17 4]: (0 0 1)=1
AGLOMERADO [27 5]: (0 1)=0.1666666667 / (1 1)=0.5 /
(0 0)=0.3333333333
AGLOMERADO [6]: (2)=0.5 / (1)=0.5
AGLOMERADO [7]: (1)=0.1666666667 / (0)=0.8333333333
AGLOMERADO [8]: (0)=1
AGLOMERADO [10]: (1)=0.5 / (0)=0.5
AGLOMERADO [11]: (1)=0.3333333333 / (0)=0.6666666667
AGLOMERADO [12]: (1)=0.3333333333 / (2)=0.5 /
(3)=0.1666666667
AGLOMERADO [13]: (1)=0.1666666667 / (2)=0.8333333333
AGLOMERADO [14]: (1)=0.3333333333 / (0)=0.6666666667
AGLOMERADO [15]: (0)=0.25 / (1)=0.75
AGLOMERADO [34 16]: (1 0)=0.1666666667 / (0 1)=0.8333333333
AGLOMERADO [18]: (0)=0.6666666667 / (1)=0.3333333333
AGLOMERADO [19]: (1)=0.5 / (0)=0.5
AGLOMERADO [20]: (1)=0.5 / (0)=0.5
AGLOMERADO [21]: (1)=0.8 / (2)=0.2
AGLOMERADO [22]: (2)=1
AGLOMERADO [23]: (0)=0.6666666667 / (1)=0.3333333333
AGLOMERADO [24]: (1)=0.3333333333 / (0)=0.6666666667
AGLOMERADO [25]: (0)=1
AGLOMERADO [26]: (0)=1
AGLOMERADO [28]: (0)=1
AGLOMERADO [32]: (0)=1
AGLOMERADO [33]: (1)=0.1666666667 / (0)=0.8333333333
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****
```

Figura VI.82 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 6 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 7
MODELO: ANATOMIA   ALFA=.025
*****
AGLOMERADO [1]: (1)=1
AGLOMERADO [32 2]: (1 2)=0.1111111111 / (0 1)=0.8888888889
AGLOMERADO [3]: (2)=0.6666666667 / (3)=0.3333333333
AGLOMERADO [31 12 6 4]: (0 1 3 1)=0.125 / (1 3 1 2)=0.125 /
(0 2 2 2)=0.125 / (0 2 2 1)=0.375 / (0 1 2 1)=0.125 /
(0 2 1 1)=0.125
AGLOMERADO [27 25 24 23 21 20 15 11 9 5]:
(0 0 0 0 2 0 0 1 0 0)=0.5 / (0 0 0 0 1 0 0 1 0 0)=0.5
AGLOMERADO [7]: (0)=1
AGLOMERADO [8]: (0)=1
AGLOMERADO [10]: (1)=0.2222222222 / (0)=0.7777777778
AGLOMERADO [13]: (1)=0.6666666667 / (2)=0.3333333333
AGLOMERADO [14]: (1)=0.4285714286 / (0)=0.5714285714
AGLOMERADO [16]: (0)=0.3333333333 / (1)=0.6666666667
AGLOMERADO [28 17]: (0 0)=0.875 / (0 1)=0.125
AGLOMERADO [18]: (0)=0.2857142857 / (1)=0.7142857143
AGLOMERADO [19]: (0)=0.75 / (1)=0.25
AGLOMERADO [22]: (1)=0.5 / (2)=0.5
AGLOMERADO [26]: (0)=1
AGLOMERADO [29]: (1)=0.4444444444 / (0)=0.5555555556
AGLOMERADO [33 30]: (1 1)=0.1111111111 / (0 0)=0.8888888889
AGLOMERADO [34]: (0)=1
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****

```

Figura VI.83 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 7 do modelo Anatomia Coronariana  $\alpha = 0,025$ .



```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 8
MODELO: ANATOMIA   ALFA=.025
*****
AGLOMERADO [8-2-1]: (1 2 1)=0.06666666667 /
(0 1 3)=0.06666666667 / (0 2 3)=0.06666666667 / (0 1 1)=0.8
AGLOMERADO [3]: (1)=0.1176470588 / (3)=0.2352941176 /
(2)=0.6470588235
AGLOMERADO [4]: (2)=0.1764705882 / (1)=0.8235294118
AGLOMERADO [5]: (1)=0.3529411765 / (0)=0.6470588235
AGLOMERADO [6]: (3)=0.4 / (2)=0.4666666667 /
(1)=0.1333333333
AGLOMERADO [18 7]: (0 1)=0.1111111111 / (1 0)=0.6666666667 /
(0 0)=0.2222222222
AGLOMERADO [9]: (1)=0.4117647059 / (0)=0.5882352941
AGLOMERADO [10]: (1)=0.5294117647 / (0)=0.4705882353
AGLOMERADO [11]: (0)=0.5 / (1)=0.5
AGLOMERADO [34 12]: (0 1)=0.1875 / (0 2)=0.625 /
(1 1)=0.1875
AGLOMERADO [29 16 13]: (1 1 2)=0.05882352941 /
(1 1 1)=0.05882352941 / (0 1 2)=0.1176470588 /
(0 0 1)=0.05882352941 / (1 0 2)=0.1176470588 /
(0 1 1)=0.5882352941
AGLOMERADO [31 20 14]: (0 1 0)=0.1111111111 /
(0 0 0)=0.2222222222 / (0 0 1)=0.5555555556 /
(1 1 0)=0.1111111111
AGLOMERADO [32 15]: (1 0)=0.0625 / (1 1)=0.1875 / (0 0)=0.75
AGLOMERADO [17]: (1)=0.3076923077 / (0)=0.6923076923
AGLOMERADO [19]: (0)=0.7857142857 / (1)=0.2142857143
AGLOMERADO [21]: (1)=0.5714285714 / (3)=0.07142857143 /
(2)=0.3571428571
AGLOMERADO [22]: (2)=1
AGLOMERADO [23]: (0)=1
AGLOMERADO [24]: (1)=0.05882352941 / (0)=0.9411764706
AGLOMERADO [25]: (1)=0.05882352941 / (0)=0.9411764706
AGLOMERADO [26]: (0)=1
AGLOMERADO [27]: (1)=0.4117647059 / (0)=0.5882352941
AGLOMERADO [28]: (0)=1
AGLOMERADO [30]: (1)=0.1176470588 / (0)=0.8823529412
AGLOMERADO [33]: (1)=0.1764705882 / (0)=0.8235294118
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****

```

Figura VI.84 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 8 do modelo Anatomia Coronariana  $\alpha = 0,025$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 1
MODELO: FUNCAO VENTRICULAR   ALFA= .0075
*****
AGLOMERADO [1]: (2)=0.08 / (3)=0.12 / (1)=0.8
AGLOMERADO [2]: (1)=0.6 / (2)=0.4
AGLOMERADO [3]: (3)=0.16 / (1)=0.04 / (2)=0.8
AGLOMERADO [16 4]: (1 2)=0.16 / (0 2)=0.16 / (0 1)=0.14 /
(1 1)=0.54
AGLOMERADO [30 29 27 5]: (0 1 0 0)=0.02 / (1 1 0 0)=0.02 /
(1 0 0 0)=0.02 / (0 0 0 1)=0.02 / (0 0 1 1)=0.06 /
(0 0 1 0)=0.04 / (0 1 0 1)=0.06 / (0 0 0 0)=0.76
AGLOMERADO [6]: (4)=0.02173913043 / (3)=0.152173913 /
(2)=0.6304347826 / (1)=0.1956521739
AGLOMERADO [36 9 7]: (1 1 1)=0.03846153846 /
(0 1 1)=0.03846153846 / (1 0 0)=0.07692307692 /
(0 0 1)=0.1923076923 / (1 1 0)=0.03846153846 /
(0 0 0)=0.6153846154
AGLOMERADO [28 8]: (0 1)=0.02 / (1 0)=0.02 / (0 0)=0.96
AGLOMERADO [10]: (0)=0.6530612245 / (1)=0.3469387755
AGLOMERADO [11]: (1)=0.4897959184 / (0)=0.5102040816
AGLOMERADO [12]: (2)=0.4782608696 / (3)=0.08695652174 /
(1)=0.4347826087
AGLOMERADO [13]: (2)=0.2857142857 / (1)=0.7142857143
AGLOMERADO [14]: (1)=0.4042553191 / (0)=0.5957446809
AGLOMERADO [15]: (0)=0.875 / (1)=0.125
AGLOMERADO [17]: (1)=0.1666666667 / (0)=0.8333333333
AGLOMERADO [19 18]: (0 0)=0.3333333333 / (1 1)=0.3333333333 /
(0 1)=0.3333333333
AGLOMERADO [20]: (1)=0.2666666667 / (0)=0.7333333333
AGLOMERADO [21]: (2)=0.1707317073 / (1)=0.8292682927
AGLOMERADO [22]: (1)=0.2 / (2)=0.8
AGLOMERADO [23]: (1)=0.08 / (0)=0.92
AGLOMERADO [24]: (1)=0.04 / (0)=0.96
AGLOMERADO [25]: (0)=1
AGLOMERADO [26]: (1)=0.02 / (0)=0.98
AGLOMERADO [31]: (1)=0.16 / (0)=0.84
AGLOMERADO [32]: (1)=0.06 / (0)=0.94
AGLOMERADO [33]: (0)=0.74 / (1)=0.26
AGLOMERADO [34]: (0)=0.92 / (1)=0.08
AGLOMERADO [35]: (1)=0.6818181818 / (0)=0.3181818182
*****

```

Figura VI.85 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 1 do modelo Função Ventricular  $\alpha = 0,0075$ .

```

*****
ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS
CONJUNTAS DA CLASSE 2
MODELO: FUNCAO VENTRICULAR   ALFA= .0075
*****
AGLOMERADO [1]: (2)=0.05405405405 / (3)=0.1351351351 /
(1)=0.8108108108
AGLOMERADO [18 8 7 2]: (0 0 1 1)=0.1 / (0 0 1 2)=0.05 /
(0 1 1 1)=0.05 / (0 0 0 1)=0.1 / (1 0 0 2)=0.1 /
(1 0 0 1)=0.5 / (0 0 0 2)=0.05 / (1 1 1 2)=0.05
AGLOMERADO [3]: (1)=0.075 / (3)=0.175 / (2)=0.75
AGLOMERADO [4]: (2)=0.15 / (1)=0.85
AGLOMERADO [5]: (1)=0.425 / (0)=0.575
AGLOMERADO [25 24 23 22 12 6]: (0 0 0 2 2 3)=0.2857142857 /
(0 0 0 2 2 2)=0.1428571429 / (1 1 0 1 1 2)=0.1428571429 /
(1 0 0 2 1 2)=0.1428571429 / (0 0 0 2 1 2)=0.1428571429 /
(1 0 0 2 1 3)=0.1428571429
AGLOMERADO [32 15 9]: (1 1 0)=0.02857142857 /
(1 0 0)=0.02857142857 / (1 1 1)=0.1142857143 /
(0 0 1)=0.1714285714 / (0 1 0)=0.1428571429 /
(0 0 0)=0.5142857143
AGLOMERADO [10]: (1)=0.525 / (0)=0.475
AGLOMERADO [11]: (0)=0.6153846154 / (1)=0.3846153846
AGLOMERADO [13]: (2)=0.358974359 / (1)=0.641025641
AGLOMERADO [14]: (1)=0.3333333333 / (0)=0.6666666667
AGLOMERADO [16]: (0)=0.1794871795 / (1)=0.8205128205
AGLOMERADO [17]: (1)=0.2142857143 / (0)=0.7857142857
AGLOMERADO [19]: (0)=0.7272727273 / (1)=0.2727272727
AGLOMERADO [31 20]: (0 1)=0.125 / (0 0)=0.75 / (1 1)=0.125
AGLOMERADO [21]: (4)=0.02777777778 / (3)=0.02777777778 /
(1)=0.7777777778 / (2)=0.1666666667
AGLOMERADO [26]: (1)=0.025 / (0)=0.975
AGLOMERADO [27]: (1)=0.375 / (0)=0.625
AGLOMERADO [28]: (0)=1
AGLOMERADO [29]: (1)=0.2 / (0)=0.8
AGLOMERADO [30]: (1)=0.175 / (0)=0.825
AGLOMERADO [33]: (1)=0.075 / (0)=0.925
AGLOMERADO [34]: (0)=0.925 / (1)=0.075
AGLOMERADO [35]: (1)=1
AGLOMERADO [36]: (0)=1
*****

```

Figura VI.86 - Estimativas não nulas de probabilidades condicionais conjuntas dos aglomerados da classe 2 do modelo Função Ventricular  $\alpha = 0,0075$ .

Caso	Classe	$\alpha = 0,075$	$\alpha = 0,01$	$\alpha = 0,025$	$\alpha = 0,05$
84	1	7.883446005E-8	6.335670789E-7	1.426205759E-11	2.833516321E-10
	2	0,02924686336	0.02738830135	0.0388185931	3.450592955E-6
	3	0,04185294063	0.03919329523	0.01769785781	0.00002345364332
	4	2.521158392E-13	2.360945341E-13	<b>0.648110805</b>	<b>0.99997691997</b>
	5	0.1468113019	0.1718522824	5.11195217E-6	0.00001642842901
	6	0.06664716812	0.06241191412	2.273713757E-7	9.42575757592E-18
	7	0.1718967199	0.190149511	5.590146738E-6	0.0001738458097
	8	<b>0.5435449272</b>	<b>0.5090040624</b>	0.2953618104	0.00001362153864
3	1	0.007381210782	0.0001366970537	1.371911751E-8	1.589151873E-6
	2	1.565532801E-11	1.577188304E-11	3.872635222E-16	9.76920855E-20
	3	0.3905001878	0.3934074894	3.9168343E-6	1.691025503E-9
	4	2.249219064E-23	2.26596466E-23	3.143051973E-33	1.684830132E-39
	5	0.00009428103302	7.915246834E-10	1.98280402E-9	1.822497049E-8
	6	0.0001984333332	0.0001962838812	0.0001505486216	8.471715492E-9
	7	0.00005025151383	4.100676834E-9	0.07402768999	<b>0.9999983808</b>
	8	<b>0.6017792355</b>	<b>0.6062595247</b>	<b>0.9258178288</b>	1.62256382E-9
18	1	<b>0.8735766858</b>	<b>0.8792830372</b>	<b>0.5234207629</b>	<b>0.9866849776</b>
	2	0.1234792504	0.1179058047	0.4765792267	0.01331501708
	3	8.364576318E-13	7.987026956E-13	7.841703677E-17	8.424916089E-13
	4	2.691285384E-17	2.569809646E-17	5.069790415E-22	2.633640227E-20
	5	5.667037833E-8	3.382029033E-8	1.214629466E-10	9.737225947E-13
	6	1.536721167E-25	1.467358647E-25	5.69149701E-25	6.207446272E-20
	7	8.708904416E-18	6.139408979E-18	1.076129499E-21	1.379992339E-12
	8	0.002944007093	0.002811124331	1.033360988E-8	5.304986678E-9

Figura VI.87 - Exemplos de aplicação de uma bateria de modelos Anatomia Coronariana com  $\alpha$  crescente.

(Bateria:  $\alpha = 0,0075/0,01/0,025/0,05$ ; Limiar = 0,99, "O Banco de Dados como Amostra", prob. a priori estimadas do banco de dados,  $p_{\min} = 10^{-6}$ ).

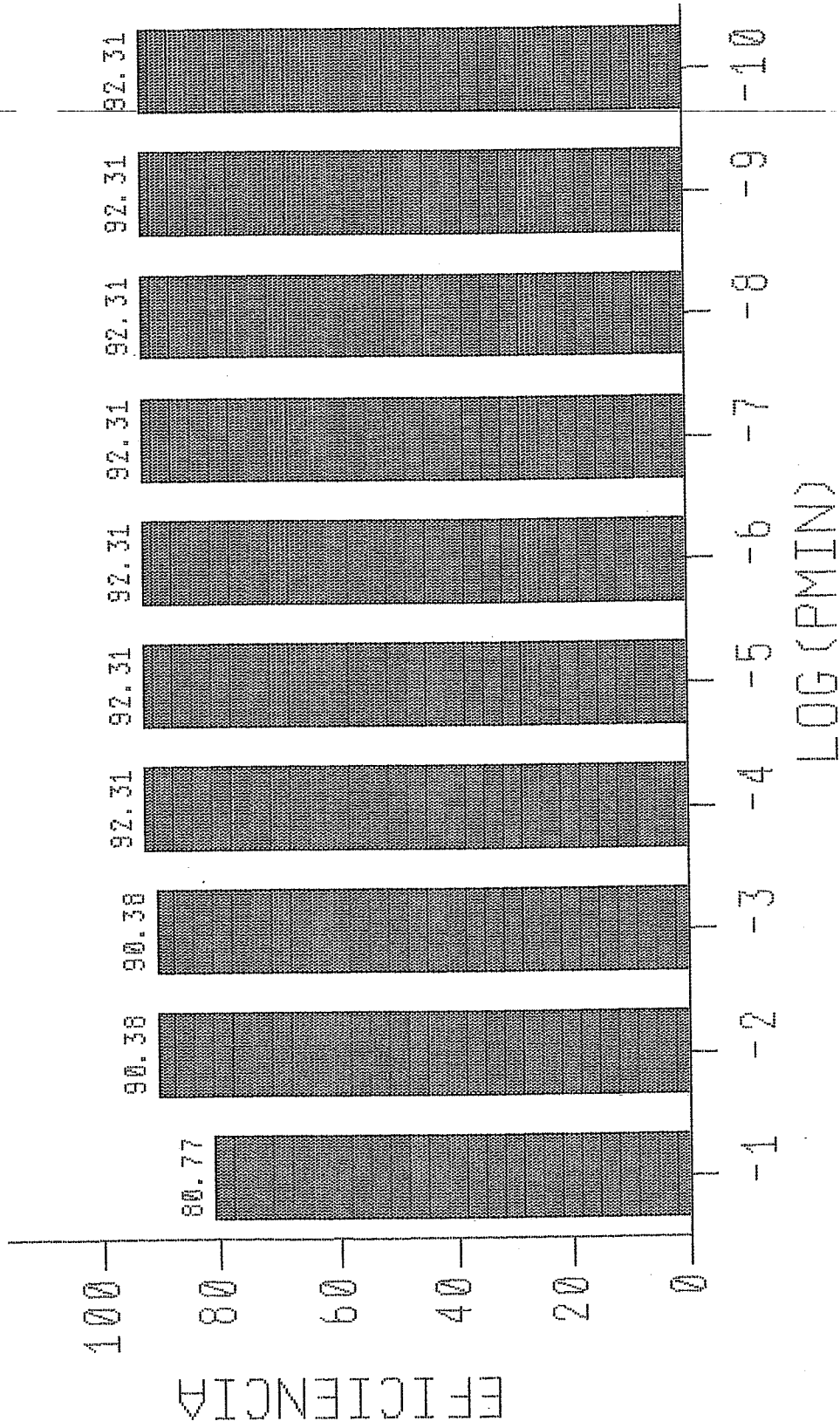


Figura VI.88 - Eficiência do modelo Presença de Doença Coronariana com aglomerados de descritores  $\alpha = 0,005$  em função do valor mínimo adotado para as probabilidades condicionais ( $p_{min}$ ). ("0 Banco de Dados como Amostra").

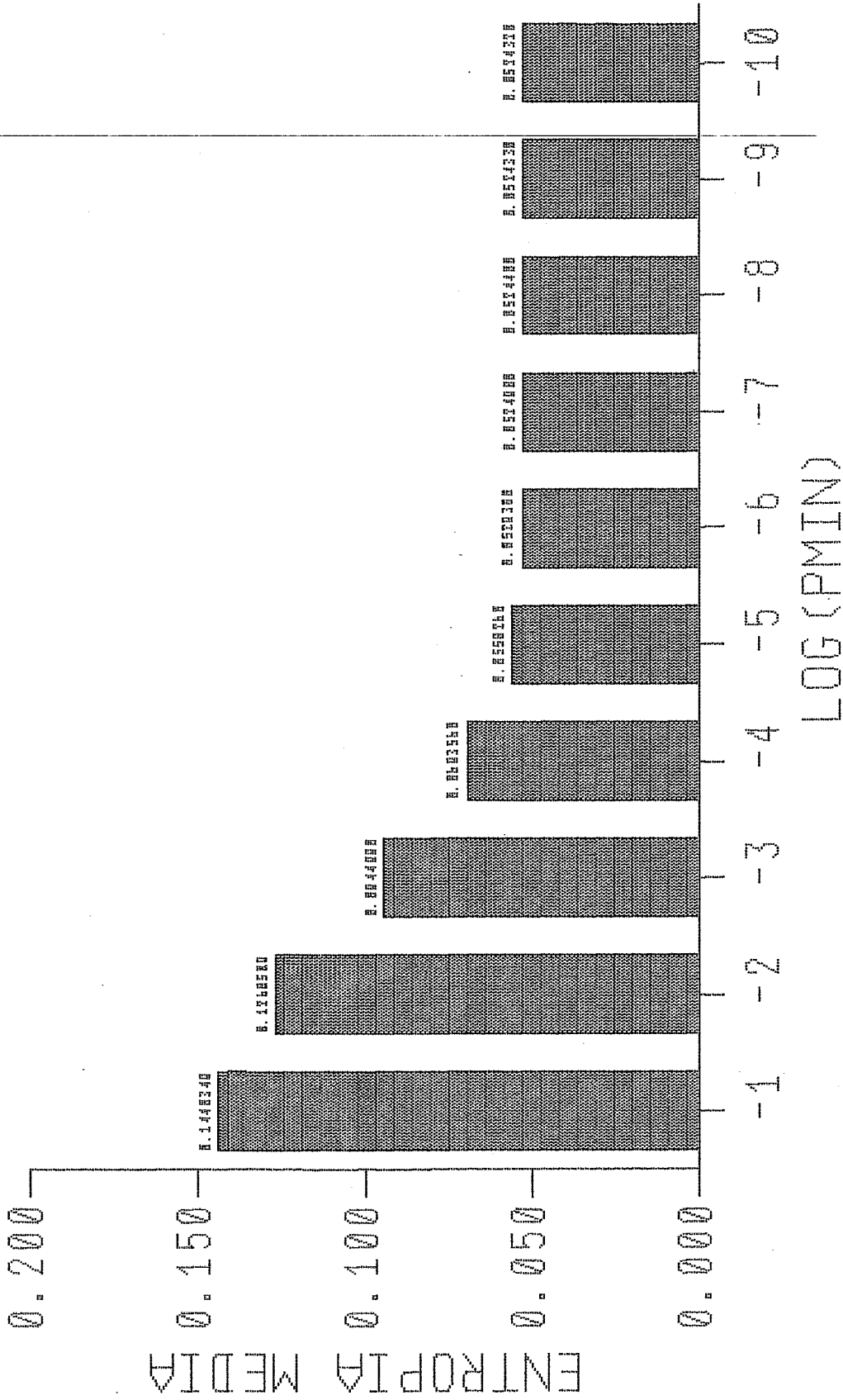


Figura VI.89 - Entropia média do modelo Presença de Doença Coronariana com aglomerados de descrito res  $\alpha = 0,005$  em função do valor mínimo adotado para as probabilidades condicionais ( $p_{\min}$ ). ("0 Banco de Dados como Amostra").

E X A M E	CUSTO (Unidades de Serviço)	DURAÇÃO (Horas)	SOFRIMENTO	MORBIDADE		MORTALIDADE	
				Paciente sem Lesões	Paciente com Lesões	Paciente sem Lesões	Paciente com Lesões
História	10	0,25	0	0	0	0	0
Exame Físico	10	0,25	0	0	0	0	0
Laboratório	50	48	3	0	0	0	0
Raios-X de Tórax	50	24	2	$2 \cdot 10^{-6}$	$2 \cdot 10^{-6}$	0	0
ECG	20	0,15	1	0	0	0	0
Ergometria	120	24	6	$10^{-3}$	$10^{-2}$	0	$2 \cdot 10^{-4}$
Ecocardiografia	150	0,50	2	0	0	0	0
Cateterismo	600	48	10	$3 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$5 \cdot 10^{-5}$	$5 \cdot 10^{-4}$

Figura VI-90 - Atributos de utilidade dos exames na área de doença coronariana.

E X A M E	$\pi_{j1}$	$\pi_{j2}$	$q_{j1}$	$q_{j2}$
História	0,6452	0,7207	1,0	1,0
Exame Físico	0,4194	0,8198	1,0	1,0
Laboratório	0,7143	0,4021	0,9032	0,8793
Raios-X de Tórax	0,8846	0,2292	0,8387	0,8649
ECG	0,9032	0,5909	1	0,9909
Ergometria	0,1667	0,8750	0,3871	0,2883
Ecocardiografia	1,0	0,9615	0,5161	0,4685
Cateterismo	1,0	1,0	-	-

Figura VI.91 - Parâmetros  $\pi_{j\alpha}$  e  $q_{j\alpha}$  estimados a partir do banco de dados com relação a variável Presença de Doença Coronariana.



---

---

## **VII - CONCLUSÕES E RECOMENDAÇÕES**

---

---

## VII.1 - CONCLUSÕES

Durante o trabalho de pesquisa realizado tivemos a oportunidade de constatar que a maioria dos dados médicos usados no estudo da doença aterosclerótica coronariana em pacientes, quando analisados isoladamente, apresenta baixo valor diagnóstico.

A agregação das informações destes descritores individuais, através do modelo bayesiano de classificação diagnóstica com pressuposição de independência, permitiu-nos obter elevados graus de decisões corretas em situações relativamente simples como a diferenciação entre ausência e presença de lesões significativas. Para situações mais complexas, como a diferenciação entre as configurações de lesões nos vasos coronarianos, o desempenho revelou-se insatisfatório.

Conseguimos demonstrar que mesmo nesses casos, é possível obter quantidades adicionais significativas de informação dos dados médicos na área de doença coronariana através de modelos mais sofisticados capazes de explorar as interações entre descritores.

Conseguimos também definir algoritmos eficientes para estimação e inferência no modelo bayesiano de classificação diagnóstica baseado em aglomerados de descritores interdependentes, capazes de em situações práticas reais, contornar o grave problema da explosão combinatorial das estimativas de probabilidades condicionais conjuntas.

Finalmente delineamos um modelo baseado em técnicas de programação heurística capaz de gerar o plano ótimo de investigação diagnóstica para um paciente específico, de acordo com a estratégia e o nível de precisão requeridos pelo médico.

Detalhamos a seguir as principais conclusões do trabalho.

### VII.1.1 - Valor Diagnóstico dos Descritores Usados na Investigação da Doença Coronariana

Os descritores usados na investigação da doença coronariana, quando analisados isoladamente com relação a variáveis diagnósticas como Presença de Lesões Significativas, Função Ventricular ou Anatomia Coronariana, revelaram-se surpreendentemente pouco informativos, a maioria deles não tendo apresentado associações significativas (ao nível de 0,05) com estas variáveis. Os descritores freqüentemente se situaram num entorno muito próximo da diagonal de não informação nos gráficos ROC para Presença de Lesões Significativas e Função Ventricular, comportando-se aproximadamente como se fossem baseados na sorte. O melhor descritor foi sem dúvida a Ecocardiografia que para Presença de Lesões Significativas apresentou sensibilidade e especificidade de 100%. Outros descritores que apresentaram associações significativas com todas as variáveis citadas foram: Sexo, Infarto do Miocárdio Prévio e as Necroses Inferior, Anterior e Outras.

### VII.1.2 - Modelos Bayesianos de Classificação Diagnóstica com Pressuposição de Independência

A agregação das informações dos descritores individuais, através do modelo bayesiano com pressuposição de independência, apresentou de um modo geral bons resultados. Assim, na modalidade de calibração e teste "O Banco de Dados como Universo", cujo desempenho pode ser interpretado como uma espécie de limite superior, para o qual tenderíamos na situação real com o crescimento do banco de dados de treinamento, obtivemos associações significativas das decisões de todos os modelos (figura VI.28) ao nível de 0,01. Para citar dois exemplos, encontramos uma eficiência de 91,55% no modelo Presença de Doen-

ça Coronariana e de 69,01% para o modelo Anatomia Coronariana. Na modalidade de treinamento e teste "O Banco de Dados como Amostra", pudemos constatar uma queda de desempenho em todos os modelos, com exceção do modelo Presença de Doença Coronariana (eficiência de 92,31%). O modelo mais complexo (Anatomia Coronariana) teve sua eficiência diminuída para 34,62%, suas decisões não se associando significativamente às classes reais, nem ao nível de 0,1.

Pudemos realizar as seguintes constatações válidas para os modelos bayesianos de classificação diagnóstica na área de doença coronariana com pressuposição de independência dos descritores.

- A - Estudando uma seqüência típica de investigação em pacientes, constatamos que os exames usados na área de doença coronariana na adicionam no modelo mais complexo (Anatomia Coronariana), em cada etapa, quantidades significativas de informação. Estas quantidades incrementais de informação decaem na medida que reduzimos o número de classes através do processo de aglutinação. O modelo Presença de Doença Coronariana, por já começar de um patamar muito elevado de eficiência (83,10%), apresenta dificuldade em elevá-la além deste nível, o que só consegue efetivamente com os dados da ecocardiografia.
- B - O valor preditivo positivo do modelo Presença de Doença Coronariana na modalidade "O Banco de Dados como Amostra", estimado em 95,23% com os dados da ecocardiografia e em 85,41% sem estes dados, superou o valor preditivo positivo da decisão médica, estimada em 78,17% a partir dos dados de indicação para cineangiocoronariografia. A comparação dos valores preditivos negativos não foi possível pois não dispomos deste parâmetro para a decisão médica.
- C - Existe excelente aderência entre a eficiência dos modelos bayesianos e a faixa de probabilidade com que é tomada a decisão na modalidade "O Banco de Dados como Universo", o mesmo se verificando para o modelo Presença de Doença Coronaria

na na modalidade "O Banco de Dados como Amostra". Estes resultados sugerem que podemos acreditar nos valores de probabilidades a posteriori fornecidas pelos modelos, para situações em que tenhamos estimativas exatas ou quase exatas das probabilidades condicionais (ou seja, bancos de dados de treinamento suficientemente grandes).

- D - Os modelos bayesianos de classificação diagnóstica demonstraram um alto grau de robustez quanto à estimativa das probabilidades a priori das classes diagnósticas. Isto sugere que a estimação destas probabilidades (justamente as mais ligadas à população sob análise) possa ser feita sem risco pelos médicos, facilitando assim, a transferência do sistema entre instituições.
- E - Estimativas nulas de probabilidades condicionais de observações ocasionam anomalias no processo de inferência tais como a eliminação de uma classe da análise diagnóstica ou a perda total da informação sobre o paciente. O número de estimativas nulas de probabilidades condicionais de observações aumenta na razão inversa do número médio de casos de treinamento por classe. O emprego de um valor mínimo não nulo para as probabilidades condicionais permite solucionar as duas anomalias citadas. A adoção do valor mínimo de  $10^{-6}$  para as probabilidades condicionais permitiu, em geral a obtenção de pequenas melhorias de desempenho nos modelos da figura (VI.29).
- F - A aglutinação de classes a posteriori, recomendada como meio de evitar o aparecimento de inconsistências entre os resultados de diferentes modelos, devido às alterações introduzidas pelo processo de aglutinação na estrutura de dependências dos descritores, permitiu na maioria das vezes a obtenção de pequenos ganhos de eficiência.
- G - Modelos construídos segundo o critério de multi-elegibilidade para refletir a configuração de lesões nos vasos coronarianos apresentaram queda acentuada de desempenho quando

comparados com o modelo global (Anatomia coronariana).

A perda de eficiência constatada na comparação dos resultados de testes dos modelos Coronária Direita, Descendente Anterior e Coronária Circunflexa com os resultados do modelo primitivo Anatomia Coronariana sugere que o critério de Multielegibilidade não deva ser adotado na construção de modelos de diagnóstico.

### **VII.1.3 - Modelos Bayesianos de Classificação Diagnóstica Baseado em Aglomerados de Descritores Interdependentes**

Como estágio final do caminho visando obter o melhor aproveitamento das informações contidas nos dados médicos, pesquisamos as interações entre descritores, potencialmente capazes de acomodar quantidades adicionais de informação. Trata-se de um problema de grande relevância na área de diagnóstico, devido à elevada frequência com que encontramos interações entre as variáveis médicas.

Dentre as aproximações propostas visando a redução do número de estimativas de probabilidades condicionais conjuntas a serem calculadas, a proposta do emprego de aglomerados de descritores interdependentes nos parece a mais potente e conveniente, em função das seguintes vantagens:

- Permite a inclusão de interações de qualquer ordem entre descritores, sem computações complexas de muitos parâmetros.
- Implementação flexível já que diferentes estruturas de aglomerados podem ser escolhidas para cada classe diagnóstica.

O modelo matemático para formação de aglomerados, proposto com base em testes de hipótese de associação entre pares de descritores e a aplicação do método do mais próximo vizinho,

permite aproveitar ao máximo as interações existentes e determinar diferentes estruturas de aglomerados em função do nível de significância pré-estabelecido. Com o crescimento deste valor podemos gerar um espectro de estruturas, que vão desde o modelo com pressuposição de independência até o modelo atuarial.

~~Mesmo com a redução substancial conseguida no número~~ de estimativas de probabilidades condicionais conjuntas a serem calculadas, ainda continuamos com o problema da explosão combinatorial destes parâmetros dentro dos próprios aglomerados. Para contornar este problema, que pode rapidamente assumir dimensão elevada quando exploramos interações fracas entre descritores, propusemos em nosso trabalho uma solução baseada no uso de representação esparsa para as probabilidades condicionais conjuntas. Esta é uma abordagem pragmática, facilmente justificável quando constatamos que a maior parte das estimativas a serem calculadas tende a ser nula, devido à desproporção entre o tamanho dos bancos de dados médicos, extremamente limitados na prática (em geral algumas centenas de casos por classe diagnóstica), e o número potencialmente explosivo de configurações de observações nos aglomerados de descritores. Partindo deste princípio, obtivemos algoritmos de estimação e inferência cujo custo, em termos computacionais (trabalho e espaço de memória requerido), é independente da estrutura de aglomerados utilizada, vinculando-se basicamente ao tamanho do banco de dados de treinamento, que em última análise representa o conhecimento médico disponível. O algoritmo de estimação proposto apresenta:

$$\text{Trabalho} = O(N_c^2 \cdot \frac{M}{N})$$

$$\text{Espaço de Memória} = O(N_c \cdot M)$$

e o algoritmo de inferência:

$$\text{Trabalho} = O(N_c \cdot M)$$

( $N_c$  = nº de casos do conjunto de treinamento,  $M$  = nº de descritores,  $N$  = nº de classes diagnósticas).

O emprego de um valor mínimo não nulo para as probabilidades condicionais é de importância fundamental nos modelos baseados em aglomerados. Experimentos com o modelo Presença de Doença Coronariana sugerem que valores menores ou iguais a  $10^{-6}$  para a probabilidade condicional mínima não afetam o comportamento dos modelos bayesianos, quer baseados em aglomerados, quer baseados na pressuposição de independência.

Constatamos nos testes realizados que o uso de aglomerados de descritores interdependentes consegue ampliar substancialmente o desempenho dos modelos bayesianos de classificação diagnóstica na área de doença coronariana, mesmo quando as interações entre descritores são fracas, e principalmente no caso de doenças moderada ou dificilmente separáveis. Assim, no modelo mais complexo (Anatomia Coronariana) conseguimos passar de uma eficiência de 34,62% (associação não significativa ao nível de 0,1), obtida com a pressuposição de independência na modalidade "O Banco de Dados como Amostra", para uma eficiência de 51,92% (associação significativa ao nível de 0,001) ou de 59,62% com o emprego de uma bateria de modelos com  $\alpha$  crescente. Já no modelo mais simples (Presença de Doença Coronariana) o ganho de eficiência foi reduzido (92,31% para 94,23%).

Na estratégia de treinamento e teste "O Banco de Dados como Universo" o modelo baseado em aglomerados de descritores apresenta eficiência crescente com o nível de significância adotado na formação dos aglomerados, até atingir a eficiência máxima (que poderá ser de 100% caso não existam pacientes de classes diferentes ocupando o mesmo ponto do espaço de padrões).

Na estratégia mais realista "O Banco de Dados como Amostra", o modelo tende a apresentar, com o crescimento do nível de significância, de início, valores crescentes de eficiência em função do aproveitamento de quantidade cada vez maiores da informação contida nas interdependências dos descritores, e posteriormente, valores decrescentes em função da rarefação dos casos nas células. É possível, portanto, determinar experimentalmente qual o valor ótimo do nível de significância, e em consequência, qual a estrutura ótima de aglomerados para um mo



delo de classificação diagnóstica em função do banco de dados de treinamento disponível. Os valores ótimos de  $\alpha$  calculados com base no nosso banco de dados (90 casos) foram de 0,01 para o modelo Presença de Lesão Coronariana, 0,025 para Anatomia Coronariana e 0,0075 para Função Ventricular. Existe potencial para obtenção de melhorias significativas de desempenho nos modelos mais complexos estudados neste trabalho através da simples expansão do banco de dados existente.

#### VII.1.4 - Geração de Planos de Investigação Médica

Conseguimos delinear um modelo capaz de determinar o plano ótimo de investigação médica para um paciente específico, de acordo com a estratégia e o nível de precisão requeridos pelo médico. O gerador de planos de investigação proposto baseia-se na visualização do problema de seleção de exames como um problema de busca do caminho de custo mínimo em um grafo, bem como na utilização de técnicas de programação heurística, as quais garantem a obtenção do caminho ótimo fazendo o melhor uso da informação heurística disponível. Os planos gerados pelo sistema são apresentados sob a forma de uma seqüência de exames a serem realizados no paciente, possuindo vantagens de simplicidade e operacionalidade.

O gerador de planos de investigação proposto utiliza como infraestrutura o modelo bayesiano de classificação diagnóstica. Juntos formam um sistema de caráter iterativo bastante abrangente, capaz de integrar muitas das características da área de diagnóstico médico, tais como estratégias de investigação de vários tipos, múltiplos atributos de exames, informação epidemiológica e de relacionamento entre doença e sintomas, registro médico do paciente, informação heurística sobre a área médica enfocada, interações logísticas entre exames e fatores médico-administrativos.

## VII.2 - RECOMENDAÇÕES

Apresentamos a seguir algumas recomendações com relação ao prosseguimento da pesquisa desenvolvida neste trabalho.

Uma primeira recomendação é a melhoria do desempenho dos modelos de classificação diagnóstica, construídos para a área de doença coronariana, com vista a sua aplicação imediata na prática médica. Isto pode ser conseguido através de:

- Expansão do banco de dados de pacientes com suspeita de doença coronariana, se necessário agregando dados de vários hospitais.
- Determinação das estruturas ótimas de aglomerados para os diversos modelos de classificação diagnóstica.
- Construção de modelos bayesianos de classificação diagnóstica baseados nas estruturas ótimas de aglomerados de descritores.

Uma recomendação adicional é a generalização da aplicação dos modelos bayesianos baseados em aglomerados de descritores para outras áreas da medicina. Poderíamos iniciar por áreas já estudadas, de forma a compararmos os desempenhos obtidos.

No que concerne ao gerador de planos de investigação médica posposto, sugerimos que o seu funcionamento seja pesquisado experimentalmente, com vistas ao seu aperfeiçoamento e aplicação em áreas tais como suporte à decisão na prática médica, modelagem do processo de investigação diagnóstica, utilização no ensino médico, análise de restrições/condicionantes impostos em serviços de assistência médica e utilização em administração hospitalar.

## **APÊNDICE 1**

---

**ARQUIVO DE PACIENTES COM SUSPEITA  
DE DOENÇA CORONARIANA**

---

Apresentamos a seguir, sob forma de uma matriz (142 X 42) os dados médicos dos pacientes utilizados neste trabalho.

As variáveis são representadas como colunas. Suas características foram descritas na seção V.3.2.2.

As colunas 1 a 36 representam os descritores apresentados na seção V.3.2.2.

As colunas 37 a 42 representam as seguintes variáveis diagnósticas:

<u>Coluna</u>	<u>Variável diagnóstica</u>
37	Presença de Doença Coronariana
38	Anatomia Coronariana
39	Número de Vasos Lesados
40	Função Ventricular
41	Presença de Lesões X Função Ventricular
42	Tipo de Tratamento

Os dados não colhidos sobre o paciente são representados como -99.

MATRIZ MCASOS

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	1	2	1	0	1	0	0	0	1	0	1	1	0	1	1	0	1	0	0	1
2	1	2	1	0	2	0	0	1	1	1	1	1	1	0	1	0	1	1	1	2
3	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	2
4	1	1	2	1	0	1	0	0	1	1	1	1	0	0	1	0	1	1	1	2
5	1	1	3	1	1	2	0	0	0	1	1	1	0	0	1	0	1	0	1	1
6	1	1	2	1	0	1	0	0	0	1	2	2	0	0	1	1	0	0	1	1
7	1	2	2	1	1	2	1	0	1	1	1	1	2	0	1	1	1	0	1	1
8	1	2	2	1	0	2	0	0	1	0	3	1	0	0	0	1	1	0	1	1
9	1	1	2	1	0	1	0	0	0	0	3	2	0	1	1	0	1	0	1	2
10	1	1	2	1	1	3	0	0	1	0	2	1	1	0	1	0	0	0	1	1
11	3	2	2	0	2	1	0	0	0	1	1	1	2	1	0	0	1	0	1	1
12	1	2	1	0	1	0	0	0	1	1	1	1	1	0	0	0	1	0	1	1
13	1	1	2	1	0	3	0	0	0	1	1	1	1	1	1	1	1	0	1	1
14	3	1	2	1	0	1	0	1	0	0	2	1	0	1	1	1	1	0	1	1
15	1	1	2	1	0	3	0	0	1	0	2	1	0	1	1	1	0	0	1	1
16	1	1	2	1	0	3	0	0	0	1	1	1	1	1	1	0	1	1	0	2
17	1	2	1	0	2	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1
18	3	2	3	2	0	2	1	0	0	1	1	1	1	1	0	0	1	0	1	1
19	2	1	2	1	0	2	1	0	1	0	1	1	0	0	1	0	1	1	1	1
20	1	1	2	1	1	3	0	0	1	0	1	1	1	1	1	0	1	0	0	2
21	1	1	3	2	0	3	1	0	1	1	1	1	1	1	1	0	1	0	0	2
22	1	1	2	1	0	2	0	0	1	0	1	1	1	1	1	0	1	0	1	1
23	1	2	2	2	0	3	1	0	1	1	1	1	0	0	1	0	1	0	0	1
24	2	1	2	2	0	2	1	1	0	0	2	1	1	0	0	0	0	0	0	1
25	1	2	3	1	1	1	1	1	1	0	1	2	0	1	0	0	1	1	1	1
26	1	1	2	2	0	2	0	0	1	0	2	2	0	0	1	0	1	0	0	1
27	2	2	1	1	0	1	0	0	0	0	2	1	1	0	0	0	0	0	0	1
28	2	1	2	2	0	2	0	0	0	0	1	2	0	0	0	1	0	0	1	1
29	1	1	3	1	1	2	0	0	0	1	1	2	0	0	0	0	0	0	0	1
30	1	1	1	1	0	2	0	0	0	0	1	1	1	0	1	1	1	0	0	1



MATRIZ MCASOS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
61	-	3	1	2	1	0	3	0	0	0	1	1	1	0	0	1	1	1	0	1	1
62	-	1	1	2	1	1	1	0	1	1	0	2	1	0	1	1	1	0	0	1	2
63	-	1	2	2	1	0	2	0	0	1	1	1	1	0	0	1	0	1	1	0	1
64	-	1	1	2	1	0	2	0	0	1	1	1	2	1	-99	1	0	1	0	-99	1
65	-	1	1	2	1	1	2	0	0	1	0	2	1	-99	0	0	0	0	0	-99	4
66	-	1	1	2	1	1	2	0	0	0	0	-99	1	0	-99	1	1	1	0	0	1
67	-	-99	1	3	1	1	2	1	1	0	0	2	1	0	0	1	1	0	1	-99	1
68	-	1	1	2	1	0	2	0	0	0	0	2	1	0	0	1	-99	-99	-99	-99	1
69	-	3	2	2	2	0	2	1	0	0	1	0	1	1	0	1	0	0	0	0	1
70	-	1	1	2	1	0	2	0	0	1	0	2	2	0	-99	0	-99	-99	1	-99	1
71	-	3	2	2	1	0	3	0	0	1	1	1	1	1	0	1	0	1	1	-99	1
72	-	1	1	2	1	1	2	0	0	1	1	2	2	1	-99	1	-99	-99	-99	-99	1
73	-	1	1	2	1	0	2	0	0	0	0	1	1	1	1	1	1	1	0	0	1
74	-	1	1	3	1	1	2	0	0	0	1	2	1	1	1	1	0	1	-99	-99	2
75	-	3	1	2	1	0	2	0	0	1	0	1	1	0	0	1	-99	-99	0	-99	1
76	-	1	1	2	2	0	2	0	0	0	0	2	1	1	0	1	-99	-99	0	-99	1
77	-	1	1	2	1	0	2	0	0	0	1	-99	1	0	0	1	-99	-99	-99	-99	1
78	-	1	1	3	1	0	2	0	0	0	0	1	1	0	0	1	-99	-99	0	-99	1
79	-	1	1	2	2	1	0	0	1	0	1	3	1	0	0	0	-99	-99	0	-99	-99
80	-	1	1	2	1	0	2	0	0	0	1	1	1	1	1	1	-99	-99	0	-99	1
81	-	3	2	2	1	0	2	0	0	0	0	1	1	0	0	0	-99	-99	-99	-99	1
82	-	1	1	2	1	1	2	0	0	0	0	1	1	0	0	1	0	0	0	0	1
83	-	1	1	1	2	1	2	0	0	0	0	2	1	0	0	1	0	1	0	0	1
84	-	1	1	2	1	0	2	0	0	1	1	2	2	1	0	1	0	-99	0	0	1
85	-	1	1	2	1	1	3	1	0	0	1	2	1	1	0	1	0	0	0	-99	1
86	-	1	1	3	1	0	2	1	0	0	1	2	1	1	0	1	-99	-99	0	-99	1
87	-	3	1	2	1	1	3	0	1	1	1	2	2	1	0	1	0	1	0	0	1
88	-	1	1	2	1	0	2	0	0	0	0	2	1	0	0	0	0	1	1	0	1
89	-	1	1	2	1	0	3	0	1	0	0	1	1	0	1	0	-99	-99	1	-99	1
90	-	1	1	2	2	1	2	1	0	0	0	2	1	0	1	1	1	0	0	0	1





MATRIZ MCASOS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
121 -	1	1	2	1	1	2	0	0	0	0	0	1 <sup>-99</sup>	1	0	0	1	1	0 <sup>-99</sup>	1		
122 -	2	2	2	1	0	2	1	0	0	1	1	2	1	1	1	1 <sup>-99</sup>	1 <sup>-99</sup>	0 <sup>-99</sup>	1		
123 -	1	2	2	2	0	1	1	0	0	1	2	1 <sup>-99</sup>	0	1	1	1	1	1	0	1	
124 -	2	2	2	2	1	2	0	1	1	1	1	1	1	0	1	1	0 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	2	
125 -	1	1	3	1	1	3	0	0	1	0	2	2	0	0	1	0	1	1	1	1	
126 -	1	1	2	1	0	1	0	0	1	0	0	2	1	1	0	1	0 <sup>-99</sup>	0	0	1	
127 -	3	1	2	2	0	2	0	0	0	1	0	2	2	0	0	0	1	0 <sup>-99</sup>	1		
128 -	1	1	2	1	0	3	1	0	1	1	0	2	2	1	0	1 <sup>-99</sup>	1 <sup>-99</sup>	0 <sup>-99</sup>	1		
129 -	1	2	2	1	1	2	1	0	0	0	0	2	1	0	0	1	0	1	0 <sup>-99</sup>	1	
130 -	1	1	2	1	1	2	0	0	0	0	0 <sup>-99</sup>	2	0	0	0	0	1	0	1 <sup>-99</sup>		
131 -	1	2	2	1	0	1	0	0	0	0	0	3	2	1	0	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1	
132 -	2	1	2	1	0	3	1	0	0	0	0	2	1	0	0	1	0	1	0 <sup>-99</sup>	1	
133 -	1	1	2	1	0	2	0	0	0	1	2	1 <sup>-99</sup>	0	1	0	1	0	1	0	0 <sup>-99</sup>	
134 -	3	1	2	1	0	1	0	0	0	0	2	1	1	1	0	1	0 <sup>-99</sup>	0 <sup>-99</sup>	1 <sup>-99</sup>	1	
135 -	1	1 <sup>-99</sup>	1	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	0	0	0	0	0	2	1	1	0	0	1	1	1 <sup>-99</sup>	1 <sup>-99</sup>	
136 -	2	1	3	1	1 <sup>-99</sup>	1 <sup>-99</sup>	1	0	0	1	0	1	2	1	1	1	0	0 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	
137 -	1	1	1	1	0	1	0	0	0	0	0	2	1	0	0	1	0	1	0	0 <sup>-99</sup>	
138 -	1	1	2	2	0	2	1	0	1 <sup>-99</sup>	1	1	2	1	0	1	0	1	0	1	0	1 <sup>-99</sup>
139 -	1	1	1	1	1	0 <sup>-99</sup>	0	0	0	0	0	2	1	0	0	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	
140 -	3	2	2	1	0	2	0	0	0	1	0	2	1	0	1	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	
141 -	1	2	2	2	0	2	0	0	0	0	0	1	1	1	0	0 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1 <sup>-99</sup>	1	
142 -	-99	1	1	2	0	2	0	0	1	1	1	1	2	1	0	0	0	1	0	0 <sup>-99</sup>	

MATRIZ MCASOS

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
1	-	2	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	1	1
2	-	-99	0	0	0	0	0	0	0	0	0	1	1-99	1	1	1	1	0	1	1	1
3	-	-99	0	1	0	0	0	0	0	0	0	0	0	1	2	7	3	2	2	4	4
4	-	-99	0	0	0	0	0	0	0	0	0	0	1-99-99	0	2	8	3	3	2	4	4
5	-	2	0	0	0	1	0	0	0	0	0	0	1-99-99	2	8	8	3	1	3	3	2
6	-	2	0	1	0	0	0	0	0	1	0	0	0	1	2	5	2	1	1	3	2
7	-	2	0	0	0	0	1	0	0	0	0	0	0	0	2	3	1	1	3	3	2
8	-	2	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1
9	-	-99	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2	2	4	2
10	-	-99	0	0	0	0	0	0	0	0	0	0	0	1	2	6	2	2	1	1	1
11	-	2	0	0	0	0	0	0	0	0	0	0	0	0	2	6	2	2	1	3	1
12	-	-99	0	0	0	0	0	0	0	0	0	1	0-99	1	1	1	0	1	1	1	1
13	-	-99	0	0	0	0	0	1	0	0	0	0	1	1	2	7	2	2	2	4	1
14	-	-99	0	0	0	0	0	1	0	0	0	0	0	0	2	7	2	2	2	4	1
15	-	2	0	0	0	0	0	0	0	1	0	0	0	1	2	5	2	2	1	3	1
16	-	2	0	0	0	0	0	0	0	1	0	0	0-99-99	2	2	2	1	1	3	1	1
17	-	-99	0	0	0	0	0	1	0	0	0	0	0-99-99	2	2	8	3	2	4	1	1
18	-	-99	0	0	0	0	0	0	0	1	0	0	0-99-99	2	1	1	0	1	1	1	1
19	-	-99	0	1	0	0	0	0	0	1	0	0	0-99-99	1	1	1	0	1	1	1	1
20	-	-99	0	0	0	0	0	0	0	0	0	0	0-99-99	1	1	0	0	2	4	2	2
21	-	-99	0	0	0	0	0	0	0	0	0	0	1-99-99	2	2	3	1	2	4	1	1
22	-	-99	0	0	0	0	0	0	0	0	0	0	0-99-99	2	2	2	1	2	4	2	2
23	-	-99	0	0	0	0	0	0	0	0	0	0	0-99-99	2	2	5	2	2	4	2	2
24	-	-99	0	0	0	0	0	0	0	0	0	0	0-99-99	2	2	4	1	1	3	1	1
25	-	2	0	0	0	0	1	0	0	0	0	0	0-99-99	2	2	5	2	2	3-99	3-99	1
26	-	1	0	0	0	0	0	0	0	0	0	1	0-99-99	2	2	5	2	2	4	1	1
27	-	2	0	0	0	0	0	0	0	1	0	0	0-99-99	1	1	1	0	1	1-99	1	1
28	-	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1
29	-	2	0	0	0	0	0	0	0	0	0	0	0-99-99	1	1	1	0	1	1	1	1
30	-	-99	0	0	0	0	0	1	0	0	0	0	0	0	2	8	3	1	3	2	2
													0-99-99		2	6	2				1



MATRIZ MCASDS

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
61	-	2	0	0	1	0	0	0	0	1	0	0	0	0	0	2	4	1	2	4	1
62	-	-99	0	0	0	0	0	0	0	1	1	0	0	0	0	2	8	3	2	4	2
63	-	-99	0	0	0	0	0	0	0	0	1	0	0	0	0	2	7	2	2	4	2
64	-	-99	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0	1	1	1
65	-	-99	0	0	0	1	0	1	0	0	0	0	0	1	0	2	7	2	2	4	-99
66	-	-99	0	0	0	1	0	0	0	0	0	0	0	0	0	2	3	1	2	4	1
67	-	-99	0	0	1	1	0	0	0	0	0	0	0	1	0	2	2	1	2	4	1
68	-	-99	0	0	0	1	0	0	1	0	0	0	0	0	0	2	3	1	2	4	-99
69	-	-99	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	1	2	4	1
70	-	-99	0	0	0	0	0	0	0	0	0	0	1	1	0	2	6	2	2	4	2
71	-	-99	0	0	0	0	0	0	0	0	0	0	1	-99	1	2	8	3	1	3	1
72	-	-99	0	1	0	0	1	0	0	0	0	0	0	1	0	2	6	2	2	4	-99
73	-	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8	3	2	4	-99
74	-	1	0	1	0	1	0	0	0	0	0	0	0	0	0	2	7	2	1	3	1
75	-	2	0	0	1	0	0	0	0	0	0	0	0	0	0	2	3	1	2	4	2
76	-	-99	0	0	0	0	0	0	0	0	0	0	0	1	0	2	7	2	1	3	2
77	-	-99	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	1	1
78	-	-99	1	0	0	0	0	0	0	0	0	0	0	0	0	2	2	1	1	3	-99
79	-	-99	0	0	0	0	0	0	0	1	0	0	0	0	0	2	7	2	1	3	-99
80	-	1	0	1	1	0	0	0	0	0	0	0	0	0	0	2	5	2	2	4	1
81	-	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1
82	-	-99	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	1	1	3	-99
83	-	-99	2	0	0	0	0	0	0	0	1	0	0	1	0	2	7	2	1	3	-99
84	-	-99	0	0	0	1	1	0	0	0	0	0	0	0	0	2	5	2	2	4	1
85	-	1	0	0	0	0	1	0	0	0	0	0	1	1	0	1	1	0	1	3	1
86	-	-99	0	0	0	1	0	1	0	0	0	0	0	1	0	2	3	1	1	3	1
87	-	-99	0	0	0	1	0	0	0	0	0	0	0	1	0	2	3	1	1	3	1
88	-	-99	0	0	0	0	0	0	0	0	0	0	0	1	0	2	8	3	2	4	1
89	-	-99	0	0	0	0	0	0	0	0	0	0	0	1	0	2	2	1	1	3	1
90	-	-99	0	0	0	1	0	1	0	0	0	0	0	0	0	2	8	3	1	3	1
														0	0	2	6	2	2	4	1

MATRIZ MCASOS

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
91	-99	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	5	2	1	3	2
92	-99	0	0	0	0	0	0	0	1	0	0	0	-99-99	2	2	3	3	1	1	3	1
93	-	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1
94	-99	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	2	1	1	3	1
95	-	1	0	0	0	0	0	0	1	0	0	0	0	0	0	3	3	1	1	3	1
96	-99	0	0	0	0	1	0	0	0	0	1	0	0	0	2	2	1	2	4	1	1
97	-99	0	0	0	0	0	0	0	1	0	0	0	0	1	2	5	2	2	4	1	1
98	-	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1
99	-99	0	0	0	0	0	1	0	0	0	0	0	0	0	2	5	2	2	4	-99	-99
100	-99	0	0	0	0	0	0	0	1	0	0	0	0	0	2	7	2	2	4	-99	-99
101	-99	0	0	0	0	1	0	0	1	1	0	0	0	0	2	6	2	2	4	1	1
102	-99	0	0	0	0	1	0	0	0	0	1	0	0	0	2	8	3	2	4	2	2
103	-	2	0	0	0	0	0	0	0	0	0	0	1	0	2	7	2	1	3	1	1
104	-99	1	0	1	0	0	0	0	0	0	0	0	0	0	2	7	2	1	3	1	1
105	-99	0	0	0	0	1	0	0	0	0	1	0	0	0	2	5	2	2	4	1	1
106	-99	0	0	0	0	0	0	1	0	0	1	0	0	0	2	7	2	2	4	1	1
107	-99	0	0	0	0	1	0	0	1	0	0	0	0	0	2	8	3	2	4	-99	-99
108	-99	0	0	0	0	0	0	1	0	0	0	0	0	0	2	8	3	2	4	-99	-99
109	-	2	0	0	0	0	0	0	0	0	0	0	0	0	2	5	2	2	4	-99	-99
110	-99	0	1	0	0	0	0	0	0	0	0	0	0	0	2	3	1	2	4	1	1
111	-99	0	0	0	0	1	0	0	0	0	1	0	0	0	2	3	1	2	4	1	1
112	-99	0	0	0	0	0	0	1	1	0	0	0	0	0	2	3	1	2	4	-99	-99
113	-99	0	0	0	0	1	0	0	0	0	0	1	0	0	2	8	3	1	3	1	1
114	-99	1	0	0	0	0	0	0	0	0	0	0	0	0	2	3	1	1	3	1	1
115	-99	0	0	0	0	0	0	1	0	0	0	0	0	0	2	3	1	2	4	-99	-99
116	-99	0	0	0	0	0	0	1	1	0	0	0	0	0	2	3	1	1	3	1	1
117	-99	0	0	0	0	0	0	0	0	0	1	0	0	0	2	3	1	1	3	1	1
118	-	2	0	0	0	0	0	0	0	0	0	0	0	0	2	8	3	2	4	1	1
119	-	2	0	0	0	1	0	0	0	0	0	1	0	0	2	6	2	2	4	1	1
120	-99	0	0	0	0	0	0	1	0	0	0	0	0	0	2	7	2	2	4	-99	-99

MATRIZ MCASOS

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
121	-	2	0	0	0	0	0	1	1	0	0	1	0	1	0	2	8	3	1	3	1
122	-	-99	0	0	1	0	0	0	0	0	0	0	0	0	-99	1	1	0	1	1	1
123	-	2	0	0	0	0	0	0	0	0	0	1	0	1	0	2	4	1	1	3	-99
124	-	-99	0	0	1	0	1	0	0	0	1	0	0	0	-99	1	1	0	2	2	-99
125	-	2	0	0	0	0	0	1	1	0	0	0	0	1	0	2	8	3	2	4	1
126	-	-99	0	0	0	0	1	0	0	0	0	0	0	0	-99	2	3	1	2	4	-99
127	-	-99	0	0	0	0	0	0	0	0	0	0	1	-99	2	7	2	2	4	4	1
128	-	-99	0	0	0	0	0	0	0	0	0	0	0	-99	2	5	2	2	1	3	-99
129	-	-99	0	0	0	0	1	0	0	0	0	0	0	1	0	2	2	1	1	3	1
130	-	-99	0	0	0	0	0	1	1	0	0	1	0	0	-99	2	7	2	2	4	1
131	-	2	0	0	0	0	0	0	0	0	0	0	0	0	-99	2	3	1	1	3	1
132	-	-99	1	0	0	0	0	0	0	0	0	0	0	0	-99	2	7	2	1	3	1
133	-	-99	0	0	0	0	0	0	0	0	0	0	0	0	-99	1	1	0	1	1	1
134	-	2	0	0	0	0	0	0	0	0	0	0	0	0	-99	2	7	2	1	3	1
135	-	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	-99	2	4	1	1	3	1
136	-	-99	1	1	0	0	0	0	0	0	0	0	0	1	0	2	7	2	2	4	1
137	-	-99	1	0	0	0	0	0	0	0	0	0	0	0	0	2	6	2	2	4	1
138	-	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1
139	-	2	0	0	0	0	0	0	0	0	0	0	0	-99	1	1	0	1	1	1	1
140	-	2	0	0	0	0	0	0	0	0	0	1	0	1	0	2	8	3	2	4	-99
141	-	-99	0	0	0	0	0	0	0	0	0	1	0	1	0	2	3	1	1	3	2
142	-	-99	0	0	0	0	1	0	0	0	0	1	0	-99	1	1	0	1	1	1	1
																2	8	3	2	4	-99

---

## **APÊNDICE 2**

---

### **DESCRIÇÃO DA BASE DE DADOS E PRINCIPAIS PROGRAMAS DESENVOLVIDOS**

---

## 2.1 - ORGANIZAÇÃO

O sistema de suporte à decisão utilizado neste trabalho foi desenvolvido no ambiente APL. Tendo em vista o tamanho limitado dos arquivos de dados, os mesmos foram armazenados nas próprias workspaces sob forma de variáveis.

A seguir descrevemos as workspaces empregadas no sistema.

### 2.1.1 - Workspace CARDIO

. OBJETIVO - Estudo de modelos bayesianos de classificação diagnóstica baseados na pressuposição de independência dos descritores.

. FUNÇÃO QUE CHAMA O MENU PRINCIPAL - SADM.

. PRINCIPAIS PROGRAMAS

AFER, ANAERRO, ATUEST, AUX, CARACVAR, CHIQ, CONTING, CRIABASE, CRIAMOD, DIAGNOSE, ENTRATESTE, ENTRAVAR, ENTROP, EXIBCASO, EXPMCASOS, GRAVAMOD, LEDADO, LELIN, LELIST, LEMOD, LISTMOD, LISTTESTE, LISTVAR, LISTVOBS, MENU, MODELS, MONTAPERFIL, MSSG, PAGE, PAUSE, SADM, SELEÇÃO, TESTAMOD, TESTEAGLUT, TITULO, TREINAMOD.

. PRINCIPAIS VARIÁVEIS

AGLUTIN, ANTPOST, CLASSE, COLR, DCLASSE, DIAG, DTESTE, DVAR, DVOBS, ELDESC, ELOBS, LVOBS, MCASOS, NCASOS, NCL, NDESC, NOMEMOD, NOMSIST, NTESTE, NVAR, PCOBS, PERFIL, POSIT, PPOST, PPRI, PROBPOST, Q, SDESC, SEL, STEST, STREIN, TIPVAR, TSTEXEC, VOBS, VORIG, YAGLUTIN, YCOLR, YDCLASSE, YNCL, YNOMEMOD, YPCOBS, YPPRI, YSDESC, YVCLASSE, YVETCONV, ZDESCR.



### 2.1.2 - Workspace CARDIAG

. OBJETIVO - Aplicação dos modelos bayesianos de classi  
ficação diagnóstica na prática médica.

---

. FUNÇÃO QUE CHAMA O MENU PRINCIPAL: SADMD.

. PRINCIPAIS PROGRAMAS

COLHE, DIAGLUT, DIAGNOSE, ELIMINE, ENTROP, GERAPAC,  
LEDADO, LELIST, LEMODELO, LOGOT, MENU, MONTAP, MSSG,  
NUMERO, PAC, PACIENTES, PAGE, PAUSE, RELAT, SADMD,  
TITULO.

. PRINCIPAIS VARIÁVEIS

AGLUTIN, ANTPOST, COLR, DCLASSE, DTESTE, DVAR, DVOBS,  
ELDESC, ELOBS, EXAMES, LDESC, LVOBS, NCL, NDESC,  
NOMEMOD, NOMES, NPAC, NTESTE, NUMPAC, NVAR, NVOBS,  
PERFIS, PPOST, PPRI, Q, SDESC, TIPVAR, TSTEXEC,  
VETCONV, VOBS, YAGLUTIN, YCOLR, YDCLASSE, YNCL,  
YNOMEMOD, YPCOBS, YPPRI, YSDESC, YVCLASSE, YVETCONV,  
ZDESCR, ZTESTE.

### 2.1.3 - Workspace CLUSTER

. OBJETIVO - Estudo de modelos bayesianos de classifica  
ção diagnóstica baseados em aglomerados de  
descritores interdependentes.

. FUNÇÃO QUE CHAMA O MENU PRINCIPAL: SADMC.

. PRINCIPAIS PROGRAMAS

AGLOMERA, ANACASO, ATUMODC, BASEMOD, CHIQ, CONTING,  
CRIAMODC, DIAGNPREP, DIAGPREP, DIAGSUPER, ENTROP,  
ESTIMA, EXIBELIST, EXPCLU, GERCLUST, GRAVAMODC,

LEDADO, LELIN, LELIST, LEMODC, LISTMODC, MAISMODELOS,  
MENU, MODELSC, MSSG, MUDACLUS, PAGE, PAUSE, PROSSIGA,  
PROVISORIA, SADMC, SELEÇÃO, TESTAMODC, TITULO,  
TREINAMODC.

---

. PRINCIPAIS VARIÁVEIS

CLASSE, COLR, DCLASSE, DIAG, DTESTE, DVAR, DVOBS,  
ECLUD, ELDESC, ELOBS, EPRCLU, EVARCLU, LCLUD, LDESC,  
LPRCLU, LVOBS, MCASOS, MODS, NCASOS, NCL, NDESC,  
NOMEMOD, NOMSIST, NVAR, NVOBS, OCORR, PINF, PMIN,  
PPOST, PPRI, PPRIANAT, PRCLU, SDESC, SEL, STEST,  
STREIN, TIPVAR, TMOD, TSTEXEC, VARCLU, VCLASSE, VOBS,  
VOBSCLU, XMOD, YCOLR, YDCLASSE, YECLUD, YEPRCLU,  
YEVARCLU, YLCLUD, YLPRCLU, YNCL, YNOMEMOD, YOCORR,  
YPPRI, YPRCLU, YSDESC, YVARCLU, YVCLASSE, YVOBSCLU,  
YZCLUVAR, ZCLUVAR, ZDESCR, ZTESTE, NPAC, PERFIL.

2.1.4 - Workspace CARDIAN

. OBJETIVO - Realização de análises diversas sobre os  
dados médicos do banco de dados de pacien-  
tes incluindo a determinação de aglomera-  
dos de descritores interdependentes.

. FUNÇÃO QUE CHAMA O MENU PRINCIPAL: SADMA.

. PRINCIPAIS PROGRAMAS

ASSOC, ASSOC1, CDTR, CHIQ, CHIQUAD, CHISQUARE, CLUST,  
CONFIANÇA, CONTING, CRAMER, EXPLORER, IMPRITESE,  
IMPRMATR, LELIST, PAGE, PAUSE, PESQ, SADMA, SIGNASSOC,  
TABCONT, TESTEMOD, TITULO.

. PRINCIPAIS VARIÁVEIS

ANTPOST, DVAR, DVOBS, ELDESC, ELOBS, LDESC, LVOBS,  
MASS, MATRIZES, MCASOS, NCASOS, NOMSIST, NVAR, NVOBS,  
POSIT, Q, SD, SDES, SDESC, SEL, STREIN, TIPVAR,  
VCLASSE, VOBS, VSEL.

2.1.5 - Workspace GPIM

. OBJETIVO: Estudo de geradores heurísticos de planos  
de investigação diagnóstica.

. FUNÇÃO QUE CHAMA O MENU PRINCIPAL: SADMG

. PRINCIPAIS PROGRAMAS

AESTRELA, APLICGER, APRESPLAN, ARQGER, ATUBAYES,  
ATUGER, ATUGPIM, ATUPIQUE, AVALDIAGN, CALCF, CALCQUE,  
CALCRISC, COLHE, COIAPAC, CRIAGER, CRITFIM,  
CRITFIMPLUS, DIAGNOSE, ELIMGER, ENTRAAB, ENTRAEX,  
ENTRAPAC, ESTCOMP, EXIBELIST, EXIBGER, EXIBPLANO,  
EXIBROB, GAMATESTE, GERCOMP, GERGPIM, GRAVAGER,  
HCHAPI, LEGER, LELIST, LEMODGER, LISTMOD, MANIGPIM,  
MENU, MONTAP, MONTAPERFIL, MSSG, PAGE, PAUSE,  
PREFERE, PROC, RELAT, REMOVEMIN, ROBUSTA, SADMG,  
TESTAGER, TITGER, TITULO.

. PRINCIPAIS VARIÁVEIS

ADICOMP, ADIUTIL, ADIUTS, AGLUTIN, ANTPOST, CLASSE,  
COLR, CRITERM, DCLASSE, DESTCOMP, DIAG, DTESTE, DVAR,  
DVOBS, ELDESC, ELOBS, EXPERM, FORMCOMP, GAMA, HCHAP,  
LATENCIA, LDESC, LIMTERM, LVOBS, MCASOS, MOD, MORBID,  
MORTAL, NCASOS, NCL, NDESC, NGER, NOMEMOD, NOMESTRAT,  
NOMESTS, NOMGER, NOMSIST, NOMUT, NTESTE, NUT, NVAR,  
NVOBS, PCOBS, PERFIL, PI, PLANO, PLANS, PLROB,  
PPOST, PPRI, PREREQ, Q, QUE, RISCO, ROB, SDESC, SEL,

STEST, STREIN, TESTE, TIPVAR, TSTEXEC, UTEX, UTIL,  
VCLASSE, VETCONV, VOBS, X, XO, YAGLUTIN, YCOLR,  
YCRITERM, YDCLASSE, YEXPERM, YGAMA, YHCHAP,  
YLIMTERM, YMOD, YMDRBID, YMORTAL, YNCL, YNOMEMOD,  
YNOMGER, YPCOBS, YPI, YPPRI, YQUE, YSDESC, YVCLASSE,  
YVETCONV, YXO, ZDESCR, ZTESTE.

## 2.2 - DESCRIÇÃO DAS PRINCIPAIS VARIÁVEIS

Apresentamos a seguir as descrições das principais variáveis utilizadas para a base de dados do sistema construído ordenadas alfabeticamente. A seguinte convenção foi usada para descrição das variáveis:

$d \times ( \dots, N_i, \dots )$

onde

$d$  = tipo do arranjo  
0 = escalar  
1 = unidimensional  
2 = bidimensional  
-----

$X$  = N: Variável numérica  
= A: Variável alfanumérica  
= B: Variável binária

$N_i$  = número de elementos na  $i^a$  dimensão

As variáveis com nome iniciado por Y são utilizadas para construção de arquivos de dados (obtidos das variáveis de mesmo nome porém sem o carácter inicial Y).

Utilizamos também as seguintes abreviaturas:

m.b.c.d. = modelo bayesiano de classificação diagnóstica.

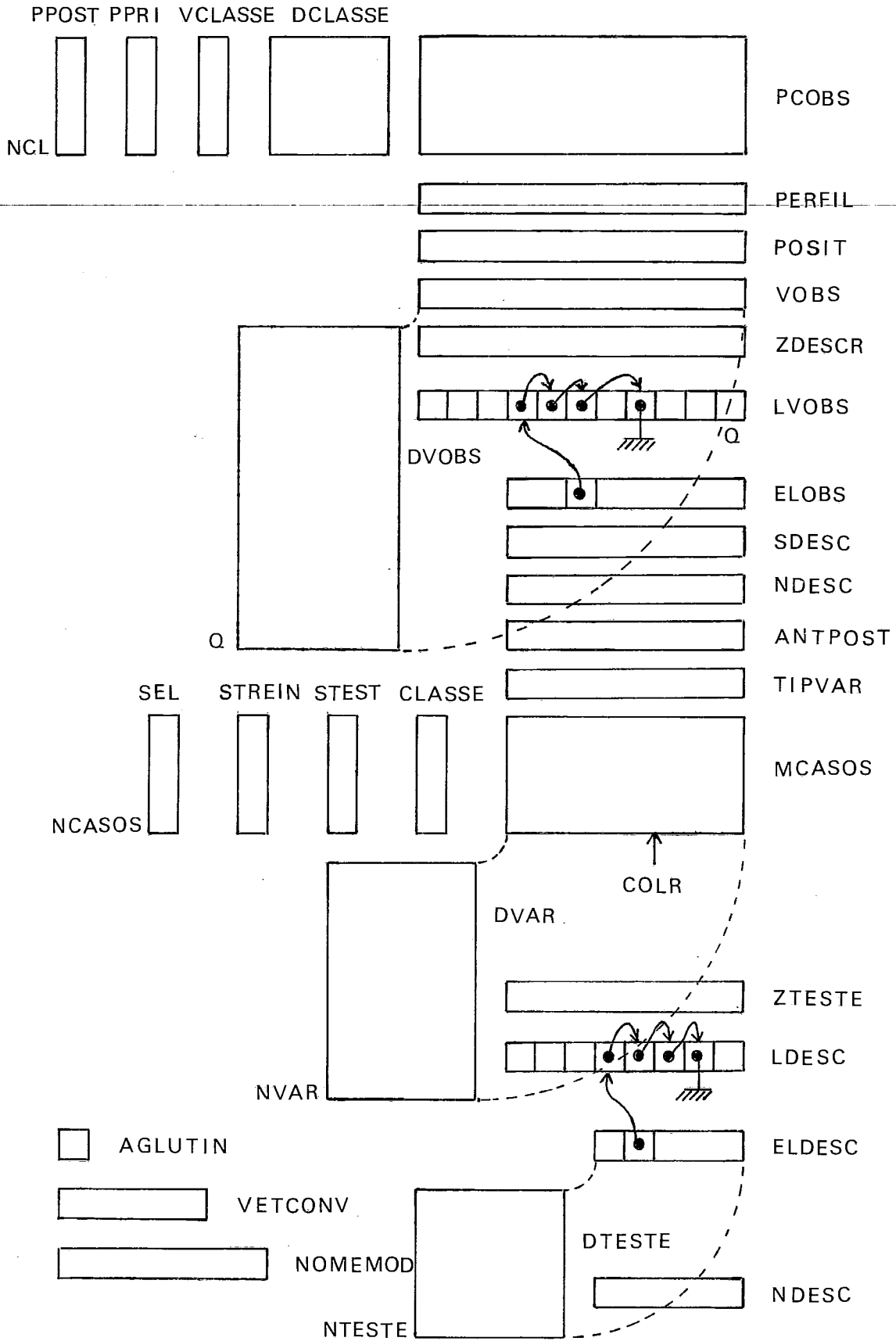


FIGURA 1.1 - ESTRUTURA DE DADOS DE M.B.C.D. COM PRESSUPOSIÇÃO DE INDEPENDÊNCIA

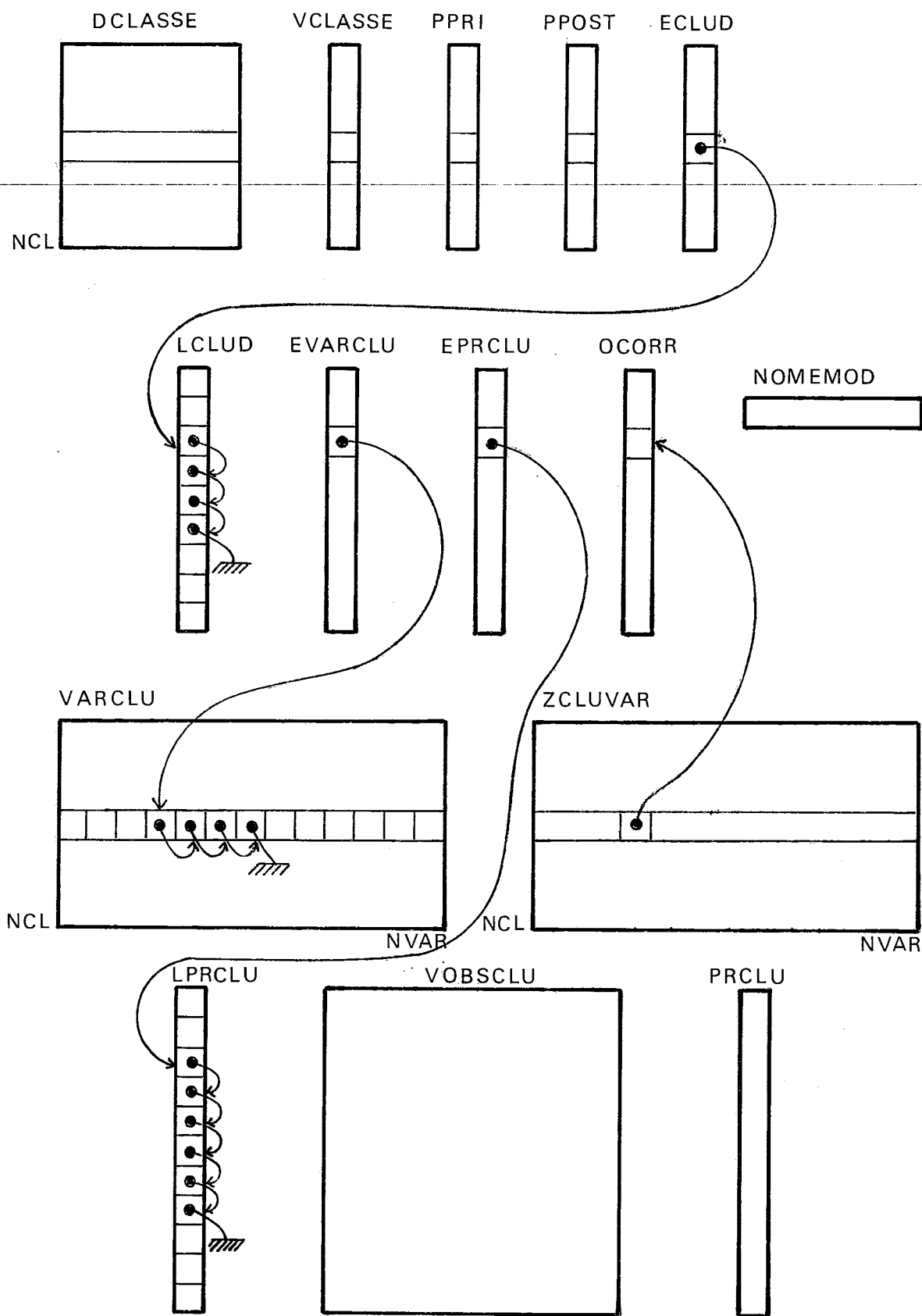


FIGURA 1.2 - ESTRUTURA DE DADOS DOS M.B.C.D. BASEADOS EM AGLOMERADOS

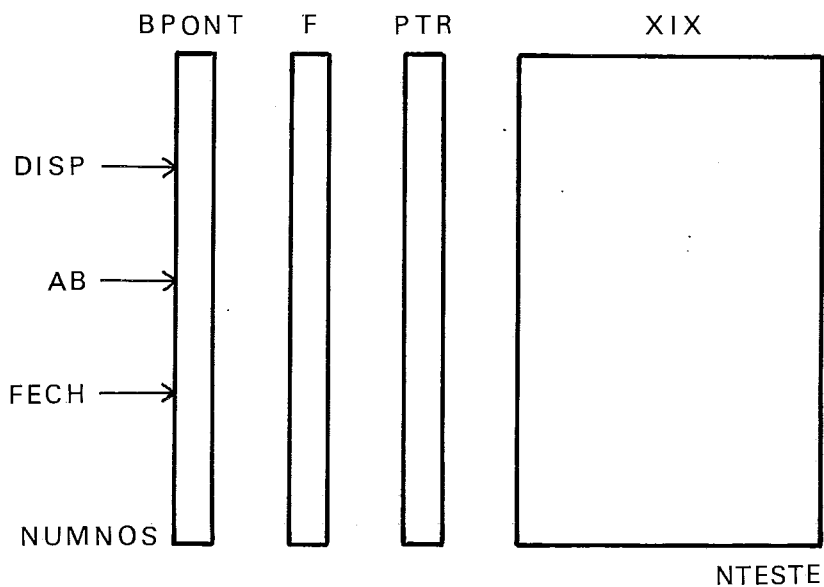
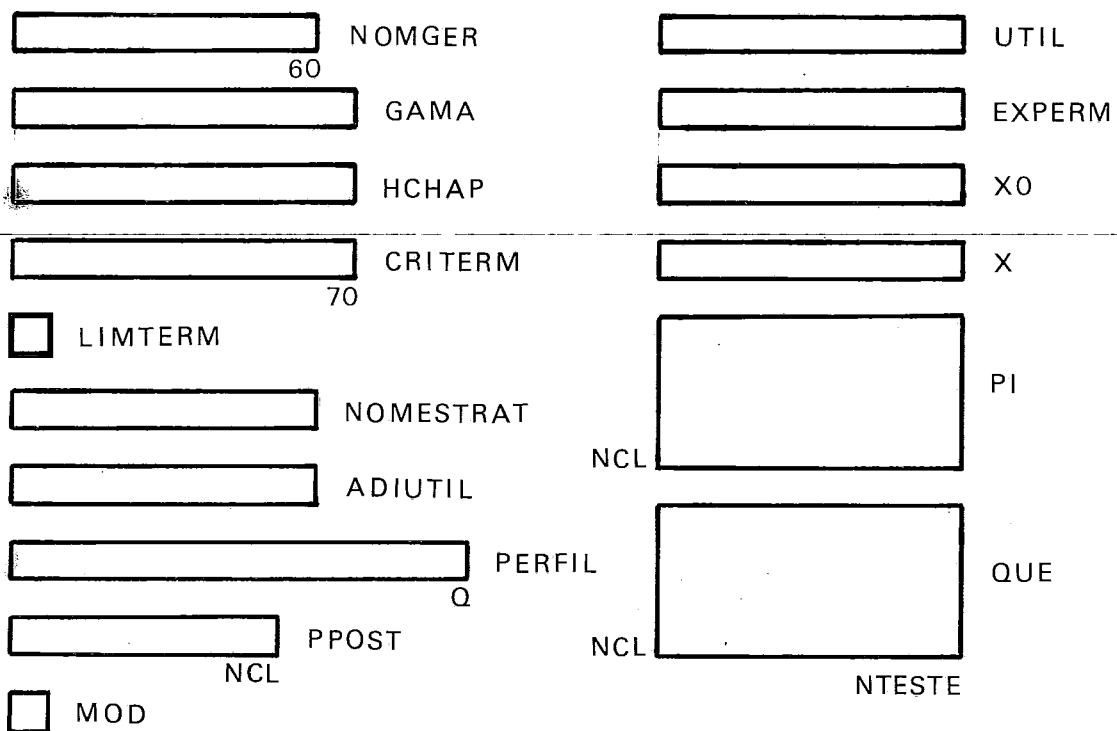


FIGURA 1.3 - ESTRUTURA DE DADOS DE UM G.P.I.M.

g.p.i.m. = gerador de planos de investigação médica.

VARIÁVEIS

- 
- ADICOMP - Operadores de adição de utilidades de exames para as diversas estratégias compostas.  
2 A (Nº strat. compostas, 70).
- ADIUTIL - Operador de adição de utilidades do g.p.i.m. corrente.  
1 A (70).
- ADIUTS - Operadores de adição das utilidades nas diversas estratégias elementares.  
2 A (NUT, 30).
- AGLUTIN - Indicador de que o m.b.c.d. corrente utiliza classes aglutinadas a posteriori a partir do modelo AGLUTIN se AGLUTIN  $\neq$  0.  
0 N.
- ANTPOST - Indicador do tipo de descritor (A = anterior à doença, P = posterior à doença).  
1 A (NVAR).
- CLASSE - Classes diagnósticas reais dos casos de MCASOS.  
1 N (NCASOS).
- COLR - Apontador da coluna de MCASOS que contém a classe diagnóstica real dos casos.  
0 N.
- CRITERM - Expressão do critério de definição de estado Terminal no grafo de espaço de estado.  
1 A (70).



- DCLASSE - Nome das classes diagnósticas de um m.b.c.d.  
2 A (NCL, 25).
- DESTCOMP - Nomes das estratégias de investigação compos  
tas.  
2 A (Nº strat. compostas, 60).
- DTESTE - Nomes dos exames.  
2 A (NTESTE, 25).
- DVAR - Nomes das variáveis (colunas de MCASOS).  
2 A (NVAR, 40).
- DVOBS - Nomes dos valores observáveis dos descrito-  
res.  
2 A (Q, 25).
- ECLUD - Ponteiros para listas de aglomerados por do-  
ença.  
1 N (NCL).
- ELDESC - Ponteiros para listas de descritores por exa  
me.  
1 N (NTESTE).
- ELOBS - Ponteiros para listas de valores observáveis  
por descritor.  
1 N (NVAR).
- EPRCLU - Ponteiros para listas de configurações de ob  
servações por aglomerado.  
1 N (Nº total de aglomerados).
- EVARCLU - Ponteiros para listas de descritores por a-  
glomerado.  
1 N (Nº total de aglomerados).

- EXAMES - Relações de exames já realizados nos pacientes cadastrados no arquivo de pacientes.  
2 N (NPAC, NTESTE).
- 
- EXPERM - Relação dos exames permitidos para inclusão em planos de investigação.  
1 B (NTESTE).
- FORMCOMP - Fórmulas de composição das utilidades dos a tributos em estratégias compostas.  
2 A (NUT, 60).
- GAMA - Operador de geração de sucessores usado na representação implícita do grafo de espaço de estado.  
1 A (70).
- HCHAP - Função heurística  $\hat{h}$ .  
1 A (70).
- LATENCIA - Latências entre exames ( $t_{ij}$  representa o tempo antes do qual não é possível realizar  $e_j$  dado que realizamos  $e_i$  neste momento).  
2 N (NTESTE, NTESTE).
- LCLUD - Listas de aglomerados de descritores por doença.  
1 N (Nº Total de aglomerados).
- LDESC - Listas de descritores por exame.  
1 N (NVAR).
- LIMTERM - Limiar para o critério de definição de estado Terminal.  
0 N.

- LPRCLU - Listas de configurações de observações por aglomerado.  
1 N (Nº total de estimativas não nulas de probabilidades condicionais conjuntas).
- 
- LVOBS - Listas de valores observáveis por descritor.  
1 N (Q).
- MASS - Matriz de associação entre descritores ( $\chi^2$  das tabelas de contingência).  
2 N (NVAR, NVAR).
- MATRIZES - Matrizes de associação para cada classe diagnóstica.  
3 N (NVAR, NVAR, NCL).
- MCASOS - Banco de dados de pacientes.  
2 N (NCASOS, NVAR).
- MOD - Número do modelo bayesiano de classificação associado ao g.p.i.m. corrente.  
0 N.
- MORBID - Estrutura de morbidades.  
B N (NCL, NTESTE, Nº de classes de risco).
- MORTAL - Estrutura de mortalidades.  
3 N (NCL, NTESTE, Nº de classes de risco).
- NCASOS - Nº total de casos contidos em MCASOS.  
0 N.
- NCL - Nº de classes diagnósticas do m.b.c.d. corrente.  
0 N

NDESC	-	Número de descritores por exame. 1 N (NTESTE).
NGER	-	Nº de g.p.i.m.'s arquivados. 0 N.
NMOD	-	Número máximo de geradores no arquivo de g.p.i.m.'s. 0 N.
NOMEMOD	-	Nome do modelo corrente. 1 A(40).
NOMES	-	Nomes dos pacientes arquivados. 2 A (NPAC, 40).
NOMESTRAT	-	Nome da estratégia de investigação selecionada pelo médico. 2 A (60).
NOMESTS	-	Nomes das estratégias elementares de investigação. 2 A (NUT, 60).
NOMGER	-	Nome do g.p.i.m. corrente. 1 A (60).
NOMSIST	-	Nome do sistema de suporte à decisão. 1 A (60).
NOMUTS	-	Nomes dos atributos de custo dos exames. 2 A (NUT, 30).
NPAC	-	Número máximo de pacientes arquivados para uso na prática médica. 0 N.

NTESTE	- Número de exames considerados pelo sistema. 0 N.
NUMNOS	- Tamanho máximo da pilha de nós disponíveis para uso pelo algoritmo A <sup>*</sup> . 0 N.
NUMPAC	- Números dos pacientes arquivados para uso na prática médica. 1 N (NPAC).
NUT	- Número total de atributos de custo de exames. 0 N.
NVAR	- Número colunas de MCASOS (descritores e variáveis diagnósticas). 0 N.
NVOBS	- Número de valores observáveis por descritor. 1 N (NVAR).
OCORR	- Indicador de ocorrências de mensuração do aglomerado no banco de dados de Treinamento. (0 = nenhuma mensuração, 1 = ocorrência de pelo menos uma mensuração. -99= modelo não treinado). 1 N (Número total de aglomerados).
PCOBS	- Matriz de probabilidades condicionais de observações nos m.b.c.d. com pressuposição de independência. 2 N (NCL, Q).

- PERFIL - Observações realizadas sobre um paciente.  
1 B (Q).
- PERFIS - Vetores PERFIL dos pacientes arquivados para uso na prática médica.  
2 B (NPAC, Q).
- PI - Matriz de parâmetros  $\Pi_{jk}$ . ( $\Pi_{jk}$  representa a probabilidade de o exame  $j$  revelar a doença  $K$ ).  
2 N (NCL, NTESTE).
- PINF - Valor limiar para seleção de diagnóstico em uma bateria de m.b.c.d. com  $\alpha$  crescente.  
0 N.
- PLANO - Plano de investigação diagnóstica.  $PLANO(i) = k$  se diferente de zero indica que o exame  $i$  deverá ser o  $k^o$  a ser executado.  
1 N (NTESTE).
- PLANS - Planos de investigação segundo cada estratégia elementar.  
2 N (NUT, NTESTE).
- PLROB - Plano segundo a estratégia robusta.  
1 N (NTESTE).
- POSIT - Critério de positividade de valores observáveis quanto à doença. (0 = negativo, 1 = positivo, -1 = não aplicável).  
1 N (Q).
- PPOST - Probabilidades a posteriori das classes diagnósticas.  
1 N (NCL).

- PPRI - Probabilidades a priori das classes diagn-  
nósticas.  
1 N (NCL).
- 
- PRCLU - Probabilidades condicionais conjuntas não  
nulas de aglomerados.  
1 N (Nº total de estimativas não nulas de  
prob. condic. conjuntas).
- PREREQ - Pré-requisitos de exames.  
2 B (NTESTE, NTESTE).
- Q - Número total de valores observáveis de to  
dos descritores.  
0 N.
- QUE - Matriz de parâmetros  $q_{jk}$ . O elemento  $(j k)$   
representa a probabilidade de o exame j  
ser realizado na investigação de um pacient  
te com a doença k.  
2 N (NCL, NTESTE).
- RISCO - Classificação do paciente numa faixa de  
risco (baixo, médio, alto).  
0 N.
- ROB - Seleção de estratégias elementares para e-  
laboração do plano robusto.  
1 N (NUT).
- SDESC - Descritores selecionados.  
1 B (NVAR).
- SEL - Casos de MCASOS selecionados.  
1 B (NCASOS).

- STEST - Casos de MCASOS selecionados para o conjunto de teste.  
1 B (NCASOS).
- 
- STREIN - Casos de MCASOS selecionados para o conjunto de treinamento.  
1 B (NCASOS).
- TIPVAR - Tipo de variável em MCASOS (I = identificação, H = história, F = Ex. físico, E = exames complementares, T = Tratamento, D = diagnóstico, N = número).  
1 A (NVAR).
- TMOD - Número de elementos do modelo no vetor YLPRCLU.  
1 N (10).
- TSTEXEC - Exames realizados sobre o paciente.  
1 B (NTESTE).
- UTEX - Utilidades dos exames para cada tipo de atributo com exceção de morbidade e mortalidade.  
2 N (NUT-2, NTESTE).
- UTIL - Utilidades dos exames para a estratégia selecionada para o g.p.i.m. corrente.  
1 N (NTESTE).
- VARCLU - Matriz contendo para cada doença as listas de descritores por aglomerado.  
2 N (NCL, NVAR).
- VCLASSE - Código das classes diagnósticas de um m.b. c.d..  
1 N (NCL).



- VETCONV - Esquema de aglutinação de classes para modelos com classes aglutinadas a posteriori.  
1 N (NCL do modelo primitivo).
- VOBS - Códigos dos valores observáveis de descritores.  
1 N (Q).
- VOBSCLU - Configurações de observações de aglomerados de descritores com estimativa não nula de probab. condicional.  
2 N (Nº estimat. não nulas, 15).
- VSEL - Variável diagnóstica selecionada para análise.  
0 N.
- X - Representação do estado do problema de busca do plano ótimo de investigação. Indica quais os exames já considerados.  
1 B (NTESTE).
- XO - Estado inicial do problema de busca do plano ótimo de investigação. Indica quais os exames já realizados.  
1 B (NTESTE).
- XMOD - Endereço do 1º elemento do modelo no vetor YLPRCLU.  
1 N (10).
- ZCLUVAR - Indicador do aglomerado, ao qual pertence um descritor.  
2 N (NCL, NVAR).

- ZDESCR - Indicador do descritor ao qual pertence um valor observável.  
1 N (Q).
- 
- ZTESTE - Indicador do exame ao qual pertence um descritor.  
1 N (NVAR).
- NPAC - Número do paciente sob análise.  
0 N.
- PERFIL - Vetor contendo as observações realizadas sobre o paciente. (-99: observação não realizada).  
1 N (NVAR).

### 2.3 - PRINCIPAIS PROGRAMAS

Apresentamos a seguir as listagens dos principais programas ordenados em ordem alfabética. Ao todo foram desenvolvidos 123 programas.

```

VAESTRELALCO]V
V RESP←AESTRELA X0;AB;FECH;BPONT;DISP;XIX;PTR;F;NO;NOPAI;I;II;V;SUCX
ROBJ: CALCULAR O PLANO (PLANO) OTIMO DE INVESTIGACAO DE ACORDO COM A
A ESTRATEGIA SELECIONADA E A PRECISAO REQUERIDA PARA O DIAGNOSTICO.
A -- RESP= 0 (FRACASSO NA BUSCA) =1 (SUCESSO NA BUSCA)
PLANO←NTESTE*0
AB←FECH*0
BPONT←1+1*NUMNOS
BPONT[1*NUMNOS]←0
DISP←1
XIX←(NUMNOS,NTESTE)*0
PTR←F←NUMNOS*0
NOPAI←0
NOPAI ENTRAAB X0
A1:→F1*1*AB=0
X←REMOVEDMIN
BPONT[NO]←FECH
FECH←NO
→S1*1*INTERMS*ACRITERM, ' X '
SUCX←1*GAMA, ' X '
NOPAI←NO
II← 1 0 /ρSUCX
I←0
C1:→A1*1*II<I+1
NOPAI PROC SUCX[I,]
→C1
F1:RESP←0
→0
S1:RESP←1
V←NO LELIST PTR
V←((ρV)-1)*V
X←X0
I←1+ρV
D1:→K1*10≥I+I-1
PLANO[(X≠NTESTE*ρXIX[V[II,]])/(NTESTE)]←(ρV)+1-I
X←NTESTE*ρXIX[V[II,]]
→D1
K1:→0
V

```

```

VANACASO[0]V
V ANACASO;F;TXT;I;TIT;J;K;S;MODS;PINF;AA;BB;V;VV
#OBJ: APLICACAO DE BATERIAS DE MODELOS BAYESIANOS DE CLASSIFICACAO
# DIAGNOSTICA COM ALFA CRESCENTE (BASEADOS EM AGLOMERADOS) A
# NOVOS PACIENTES
X1:TIT←'APLICACAO DE MODELOS DIAGNOSTICOS A UM PACIENTE'
TXT← 4 40 p' .
F←40p' .
TXT[1;]←40p'ABERTURA DO REGISTRO DE UM NOVO PACIENTE',F
TXT[2;]←40p'ENTRADA DE DADOS DO PACIENTE',F
TXT[3;]←40p'ANALISE DIAGNOSTICA',F
TXT[4;]←40p'APRESENTACAO DOS DADOS DO PACIENTE',F
I←2 MENU TXT
A1→A',TI
A0:→0
A1:PAGE
TXT[1;]
79p'#:
PERFIL←NVARp-99
'NUMERO DO PACIENTE?'
NEAC←0
'SEXO? 1=MASC. 2=FEM.'
J←0
→E1x1(J6 1 2)≠1
PERFIL[2]←J
'COR? 1=BRANCA 2=PARDA 3=NEGRA'
J←0
→E1x1(J6 1 2 3)≠1
PERFIL[1]←J
'IDADE? ANOS'
J←0
→B1x1J<40
→B2x1J>60
PERFIL[3]←2
→X1
B1:PERFIL[3]←1
→X1

```

```

[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]

```

```

[36] E2:PEREIL[3]←3
[37] →X1
[38] A2:PAGE
[39] TXTL2:]
[40] 79P' #'
[41] 'DESCRITORES JA MENSURADOS'
[42] (PEREIL#-99)/\NVAR
[43] 10
[44] 'SELECCION DESCRITORES PARA ENTRADA DE DATOS:'
[45] ((PEREIL=-99)^SDESC=1) EXIBELIST DVAR
[46] S←,0
[47] PAGE
[48] J←0
[49] K1:→K2x1(PS)<J←J+1
[50] 10
[51] 'DESCRITOR: ',DVARE[SJJ:]
[52] 79P'-'
[53] V←ELOBS[SJJ] LELIST LVOBS
[54] V←ψV
[55] (2 0 V((PV),1)PVOBS[V]),((PV),1)P' '),DVOBS[V:]
[56] 10
[57] 'VALOR?'
[58] K←0
[59] →E1x1(K←VOBS[V])=0
[60] PEREIL[SJJ]←K
[61] →K1
[62] K2: 'COLETA DE DATOS TERMINADA'
[63] PAUSE
[64] →X1
[65] A3:PAGE
[66] AA←TXT[3:],(15P' '), 'PACIENTE ',VREAS
[67] AA
[68] 79P' #'
[69] BB← 2 40 P' '
[70] BB[1:]←40P'ANATOMIA CORONARIANA',F
[71] BB[2:]←40P'FUNCAO VENTRICULAR',F
[72] 'SELECCION'

```

```

[73] 1 - ',BB[1;]
[74] 2 - ',BB[2;]
[75] K←0
[76] →E1X1(K←1 2)=0
[77] →D1X1K=2
[78] MODS←, 2 1 3 4
[79] PINF←0.9999
[80] →D2
[81] D1:MODS←,5
[82] PINF←0.9999
[83] D2:PAGE
[84] AA
[85] BBK;/J
[86] 79P'!*'
[87] MODS DIAGSUPER EEREIL
[88] 10
[89] ' CLASSE',(20P''),'PROBABILIDADE'
[90] 40P'!'
[91] (2 0 T(NCL,1)PVCCLASSE),(NCL,1)P' '),DCLASSE, 12 10 T(NCL,1)PPOST
[92] 10
[93] PAUSE
[94] →X1
[95] A4:PAGE
[96] TXT[4;],' ',TNEAC
[97] 79P'!*'
[98] 10
[99] V←(EEREIL≠'99')/INVAR
[100] I←0
[101] G2:→G1X1(PV)←I+1
[102] VV←ELOBS[V[I]] LELIST LVOBS
[103] K←VOBS[VV](EEREIL[V[I]])
[104] DVAR[V[I]];',' ',DVOBS[VV[K]];]
[105] →G2
[106] G1:PAUSE
[107] →X1
[108] E1:MSSG 'VALOR INVALIDO'
[109] →X1

```

▽

```

VAPLICGER[0]V
V APLICGER;TIT;R;TXT;I;RESP
#OBJ: APLICACAO DO GERADOR DE PLANOS DE INVESTIGACAO CORRENTE A UM
# PACIENTE ESPECIFICO
NUMNDS←500
TIT←70P 'APLICACAO DO GERADOR DE PLANOS DE INVESTIGACAO A UM PACIENTE',30P
,
[1] TXT← 8 60 P ' '
[2] TXTL1;J←60P 'ZERAR O REGISTRO DO PACIENTE',60P ' '
[3] TXTL2;J←60P 'ENTRAR/ATUALIZAR OS DADOS INICIAIS DO PACIENTE',60P ' '
[4] TXTL3;J←60P 'ENTRAR DADOS DE EXAMES DO PACIENTE',60P ' '
[5] TXTL4;J←60P 'COPIAR DADOS INICIAIS DE UM CASO DO BANCO DE DADOS',60P ' '
[6] TXTL5;J←60P 'GERAR PLANO DE INVESTIG. SEG. ESTRATEGIA ELEMENTAR',60P ' '
[7] TXTL6;J←60P 'GERAR PLANO DE INVESTIG. SEG. ESTRATEGIA ROBUSTA',60P ' '
[8] TXTL7;J←60P 'GERAR PLANO DE INVESTIG. SEG. ESTRATEGIA COMPOSTA',60P ' '
[9] TXTL8;J←60P 'EXIBIR DADOS DO PACIENTE',60P ' '
[10] X1:R←0 MENU TXT
[11] A1←A',TR
[12] A0:→0
[13] A1:PAGE
[14] NUM←0
[15] PERFIL←Q*0
[16] TSTEXED←X0
[17] MSSG 'PERFIL DO PACIENTE ZERADO!'
[18] 10
[19] →X1
[20] A2:TITGER
[21] ENTRAPAC
[22] →X1
[23] A3:ENTRAEX
[24] →X1
[25] A4:PAGE
[26] 'NUMERO DO CASO?'
[27] NUM←J←0
[28] COFIAPAC J
[29] 'CASO COPIADO'
[30] PAUSE

```

```

[35] →X1
[36] A5:PAGE
[37] 'SELECCIONE A ESTRATEGIA DESEJADA: '
[38] (NUTP1) EXIBELIST NOMESTS
[39] R←D
[40] DIAGNOSE PERFIL
[41] NOMESTRAT←NOMESTS[R;J]
[42] ADIUTIL←ADIUTSR;J]
[43] →K1×R$2
[44] UTIL←UTEX[R-2;J]
[45] →K2
[46] K1:RISCO←CALCRISC
[47] →K3×R=2
[48] I←0
[49] T2:→K2×\NTESTE<I←I+1
[50] UTIL[[I]←+/PPOST×MORBIDL;I;RISCO]
[51] →T2
[52] K3:I←0
[53] T3:→K2×\NTESTE<I←I+1
[54] UTIL[[I]←+/PPOST×MORTALL;I;RISCO]
[55] →T3
[56] K2:RESP←AESTRELA TSTEXEC
[57] →G1×\RESP=1
[58] PAGE
[59] AVALDIAGN
[60] \0
[61] 'NAO FOI POSSIVEL ELABORAR PLANO DE INVESTIGACAO USANDO O PARAMETRO '
[62] 'DE DEFINICAO DE ESTADO TERMINAL =',\LIMTERM
[63] \0
[64] 'ESTRATEGIA... ',NOMESTS[R;J]
[65] PAUSE
[66] →X1
[67] G1:PAGE
[68] AVALDIAGN
[69] PAUSE
[70] EXIBFLAND
[71] PAUSE

```



```

[72] →X1
[73] A6:PAGE
[74] RISCO←CALCRISC
[75] DIAGNOSE PERFIL
[76] ROBUSTA
[77] AVALDIAGN
[78] PAUSE
[79] PAGE
[80] EXIBROB
[81] PAUSE
[82] →X1
[83] 'SELECCION A ESTRATEGIA DESEJADA:'
[84] A7:PAGE
[85] 'SELECCION A ESTRATEGIA DESEJADA:'
[86] ((1 0 /PDESTCOMP)P1) EXIBELIST DESTCOMP
[87] R←D
[88] DIAGNOSE PERFIL
[89] RISCO←CALCRISC
[90] GERCOMP R
[91] RESP←AESTRELA TSTEXEC
[92] →H1X1RESP=1
[93] AVALDIAGN
[94] 10
[95] 'NAO FOI POSSIVEL ELABORAR PLANO DE INVESTIGACAO USANDO O PARAMETRO DE'
[96] 'DEFINICAO DE ESTADO TERMINAL.: ',VLIMITM
[97] 10
[98] 'ESTRATEGIA ',DESTCOMP,R,]
[99] PAUSE
[100] →X1
[101] H1:AVALDIAGN
[102] PAUSE
[103] EXIBPLAND
[104] PAUSE
[105] →X1
[106] A8:NOME←30P'
[107] RELAT 0P0
[108] →X1

```

```

VASSOC[0]V
V SEL ASSOC SDESC;I;J;TAB;PARES;Z;C;D;DD;CC;Z;NEL
#OBJ: CALCULAR A MATRIZ (MASS) DE ASSOCIACAO ENTRE DESCRITORES PARA
A UMA CLASSE DIAGNOSTICA
D+(SDESC=1)/INVAR
C+(SEL=1)/INCASOS
NEL+MASS+(2*PD)*0
I+0
A1:→A3*(PD)<I+I+1
J+I-1
A2:→A1*(PD)<J+J+1
PARES+(((PC),1)*MCASOS[C;DLIJJ]),((PC),1)*MCASOS[C;DLJJ]
Z+(PARES;I;J#99)^(PARES;2;J#99)
PARES+Z/[I] PARES
CC+VOBSCELOBS[DLIJ] LELIST LVOBSJ
DD+VOBSCELOBS[DLJJ] LELIST LVOBSJ
TAB+CC CONTING DD
MASS[I;JJ+MASS[J;IJ]+CHI* TAB
NELLI;JJ+NELLJ;IJ+//+/TABZ+((1 0 /PTAB)L 0 1 /PTAB)-1
→A2
A3:TITULO
1 0
'MATRIZ -MASS- DE CHI QUADRADOS CALCULADA'
▽
VCALCF[0]V
V F←CALCF X;G;H
#OBJ: CALCULAR A ESTIMATIVA DA FUNCAO F PARA O ESTADO X
G←ADITIL,' UTIL(EXPERM*X←X0)/INTESTEJ'
H←HCHAP,' X'
F←ADITIL,' G,H'
#F←ADITIL,' F'
▽
VCLUST[0]V
V CLUST LIG;CLU;N;I;V;OPEN
#OBJ: DETERMINAR OS CLUSTERS DE DESCRITORES INTERDEPENDENTES PELA

```

```

[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[1]
[2]
[3]
[4]
[5]
[1]

```

```

[2] * APLICACAO DO METODO DO MAIS PROXIMO VIZINHO A MATRIZ DE SIMI-
[3] * LARIDADE (LIG)
[4] V←SDESC/1pSDESC
[5] CLU←(pV)p0
[6] OPEN←(pV)p1
[7] N←I←0
[8] A2:→0x1(pV)<I←I+1
[9] →A2x1OPEN[I]=0
[10] N←N+1
[11] CLU[I]←1
[12] '-----CLUSTER ',(FN),50p'...'
[13] PESO I
[14] 'VARIAVEIS COMPONENTES DO CLUSTER ',FN
[15] CLU/V
[16] CLU←(pV)p0
[17] →A2
▽
▽COLHELDJ▽
▽ RET←COLHE VAR;M;J;K;N;L;I;A
*OBJ: COLHER RESULTADO DA MENSURACAO DE UM DESCRITOR NO PACIENTE
[1] ABORTO←0
[2] X1:RET←0
[3] DISPLAY 'ANAMN1'
[4] 'CAB1' SWRI 'EXAME',(24p'.'),(.''),25pDTESTELZTELVAR;J
[5] 'CAB2' SWRI 'DESCRITOR',(20p'.'),(.''),40pDVAREVAR;J
[6] 'NUM' SWRI INVOBSE[VAR]
[7] K←0p0
[8] J←ELOBS[VAR]
[9] A2:→A1x1J=0
[10] K←K,J
[11] J←LVOBSE[J]
[12] →A2
[13] A1:→C1x10=+/PERFILCK]
[14] 'CAB3' SWRI 'VALOR ATUAL = ',TVOBSE[PERFILCK]=1)/KJ
[15] →A7
[16] C1:'CAB3' SWRI 'SELECCIONE O VALOR OBSERVADO:'

```



```

[5] DEL←1.96×(SS×(1-SS)÷ND)×0.5
[6] CONFSL;1]←100×SS-DEL
[7] CONFSL;2]←100×SS+DEL
[8] DEL←1.96×(EE×(1-EE)÷ND)×0.5
[9] CONFEL;1]←100×EE-DEL
[10] CONFEL;2]←100×EE+DEL

```

```

▽CONTING[0]▽
▽ TAB←CLL CONTING CLC;K;L;M;N
ROBJ: CALCULAR A TABELA DE CONTINENCIA ENTRE 2 VARIAVEIS DISCRETAS
A -- CUJAS CLASSES SAO APRESENTADAS EM CLL E CLC
A -- CLL=CLASSES VAR. DE LINHAS TAB
A -- CLC=CLASSES VAR. DE COL. TAB
A -- PARES=MATRIZ 2XN CONTENDO VARS LIN E COL TAB
ATAB=T.CONTING
N←1.0 /PPARES
TAB←((PCLL),PCLC)P0
K←0
B1:→OX\N(K←K+1
L←CLL\PARESEK;1]
M←CLC\PARESEK;2]
TABLL;M]←TABLL;M]+1
→B1
▽

```

```

▽CRIAGER[0]▽
▽ CRIAGER
ROBJ: CRIACAO DE UM GERADOR DE PLANOS DE INVESTIGACAO DIAGNOSTICA
TITGER
'CRIACAO DE ',CABEC
'COM BASE NO ALGORITMO A-ESTRELA'
79' '*
'NOME DO GERADOR?'
'↓',(589' '),↓'
NOMGER←60PB,60P'
'QUAL O NUMERO DO MODELO BAYESIANO DE DIAGNOSE ASSOCIADO?'

```

```

[103] (YNCL>0) EXIBELIST YNMEMOD
[113] MOD<0
[123] 'CRITERIO DE ESTADO TERMINAL?'
[133] CRITERM<70p0,70p'
[143] 'VALOR DO PARAMETRO DO CRITERIO DE ESTADO TERMINAL?'
[153] LINTERM<0
[163] 'OPERADOR GAMA DE GERACAO DE SUCESSORES NO ESPACO DE ESTADO?'
[173] GAMA<70p0,70p'
[183] 'FUNCAO HEURISTICA H?'
[193] HCHAP<70p0,70p'
[203] 'INDIQUE OS EXAMES PERMITIDOS:'
[213] (NTESTE01) EXIBELIST DTESTE
[223] X<X0<EXPERM<NTESTE0
[233] EXPERM,01<1
[243] 'INDIQUE OS EXAMES QUE DEFINEM O ESTADO INICIAL (EXAMES REALIZADOS)'
[253] 'A PRIORI)?'
[263] (NTESTE01) EXIBELIST DTESTE
[273] X0L,01<1
[283] UTIL<NTESTE00
[293] ADIUTIL<70p'
[303] NOMESTRAT<60p'
[313] LEMODGER MOD
      v
vCRIAMOD[0]v
v CRIAMOD;J;K;L;R
*OBJ: CRIACAO DE UM MODELO BAYESIANO DE CLASSIFICACAO DIAGNOSTICA
* COM PRESSUPOSICAO DE INDEPENDENCIA DOS DESCRITORES
      v
[1] TITULO
[2] 'CRIACAO DE UM MODELO BAYESIANO'
[3] 'NOME DO MODELO?'
[4] NOMEMOD<40p0,40p'
[5] 'NUMERO DE CLASSES?'
[6] AGLUTIN<0
[7] VETCONV<10p'99
[8] NCL<0
[9] VCLASSE<PPRI<PPOST<NCL0
[10]
[11]

```

```

[12] DCLASSE←(NCL,25)P' '
[13] PCOBS←(NCL,Q)P0
[14] COLR←0
[15] 'A CLASSIFICACAO DIAGNOSTICA DESSE MODELO CORRESPONDE A ALGUMA '
[16] ' DAS COLUNAS DA MATRIZ DE CASOS? (S/N)'
[17] →T1X1 'N'=140
[18] 'QUAL COLUNA?'
[19] COLR←0
[20] CLASSE←MCASOSL;COLR
[21] T1:70P'+'
[22] 'ESPECIFICACAO DAS CLASSES'
[23] 70P'+'
[24] PAUSE
[25] J←0
[26] A2:PAGE
[27] →A1X1NCL<J←J+1
[28] 70P'+'
[29] 'CLASSE DIAGNOSTICA NUMERO ',TJ
[30] 70P'+'
[31] 'DESCRICA0?'
[32] 25P'-'
[33] DCLASSE[J;J←25P0,25P' '
[34] 'CODIGO DA CLASSE NUMERO ',TJ
[35] VCLASSE[J]←0
[36] D3:'PROBABILIDADE A PRIORI DA CLASSE ',TJ
[37] PPRICJ]←0
[38] →E1X1<PPRICJ]1>VPPRICJ]←0
[39] →A2
[40] A1:→E2X1<+/PPRI)≠1
[41] D4:'SELECIONE NA LISTA ABAIXO OS DESCRITORES QUE FAZEM PARTE DO MODELO:'
[42] 'DESCRITORES'
[43] 11P'-'
[44] ' '
[45] K←0
[46] B3:→B1X1NVAR<K←K+1
[47] →B3X1<+/TD'ETIPVAR[K])=1
[48] (3 0 TK),( ' '),DVAR[K];' ' ,TIPVAR[K]

```

```

[49] →B3
[50] E1:L←1,0
[51] SDESC←NVAR*0
[52] SDESCLLJ←1
[53] 10
[54] 'O MOD. EM CRIACAO SERA OBTIDO POR AGLUTINACAO DE CLASSES'
[55] 'DE UM MODELO MAIOR? (S/N)'
[56] R←0
[57] →0\ 'N'=1+R
[58] 'QUAL O NUMERO DO MODELO MAIOR'
[59] AGLUTIN←0
[60] 'FORNECA O VETOR DE CONVERSÃO DAS CLASSES DO MODELO MAIOR PARA'
[61] 'O MODELO EM QUESTÃO.'
[62] '(NOVA CLASSE ← VETCONVCENDER. DA CLASSE NO VET. VCLASSE DO MOD. MAIOR)'
[63] 'EX. CONVERSÃO: ANATOMIA → NUM VASOS LESADOS'
[64] 'VETCONV=0 1 1 1 2 2 2 3'
[65] 'VETCONV←10*0,10*~99
[66] →0
[67] E1:'VALOR ERRADO DE PROBABILIDADE'
[68] →D3
[69] E2:'ERRO --- A SOMA DAS PROB A PRIORI DEVE SER IGUAL A 1'
[70] →D4
      V
VCRIAMODC[D]V
V CRIAMODC;J,K;L;M;N
#OBJ: CRIACAO DE UM MODELO BAYESIANO DE CLASSIFICACAO DIAGNOSTICA
# BASEADO EM AGLOMERADOS DE DESCRITORES INTERDEPENDENTES
#
# TITULO
# CRIACAO DE MODELO BAYESIANO C. CLUSTERSDE DESCRITORES'
79*'*'
'NOME DO MODELO?'
'↓',(38*'-'),'↓'
NOMEMOD←40*0,40*1
'NUMERO DE CLASSES?'
NCL←0
VCLASSE←PPRI←PPOST←ECLUD←NCL*0

```



```

[12] DCLASSE←(NCL,25)P' '
[13] COLR←0
[14] ZCLUVAR←(NCL,NVAR)P0
[15] VARCLU←(NCL,NVAR)P~99
[16] DCCOR←LCLUD←EVARCLU←EPRCLU←0P0
[17] 'A CLASSIFICACAO CORRESPONDE A ALGUMA DAS COLUNAS DA MATRIZ DE CASOS?'S/N'
[18] →T1X1'N'=1↑,D
[19] 'QUAL COLUNA?'
[20] COLR←D
[21] →T4
[22] T1: 'VOCE DEVE CRIAR UM VETOR →CLASSE← CONTENDO OS DIAGNOSTICOS CORRES--'
[23] 'PONDENTES AOS CASOS DA MATRIZ MCASOS'
[24] PAUSE
[25] T4: PAGE
[26] 'INDIQUE QUAIS OS DESCRITORES SELECIONADOS DA LISTA ABAIXO.'
[27] 10
[28] 'DESCRITORES'
[29] 11P'-'
[30] 10
[31] K←0
[32] B3: →B1X(NVAR<K+K+1
[33] →B3X1(+/'TD'€TIPVAREK])=1
[34] (3 0 TK), ' ,DVAREK/],, ' ,TIPVAREK]
[35] →B3
[36] B1: L←1, D
[37] SDESC←NVARP0
[38] SDESC[L]←1
[39] PAUSE
[40] 70P' *'
[41] 'ESPECIFICACAO DAS CLASSES E DOS CLUSTERS DE DESCRITORES ASSOCIADOS'
[42] 70P' *'
[43] PAUSE
[44] J←0
[45] A2: PAGE
[46] #ESPECIF CLASSES
[47] →A1X(NCL<J←J+1
[48] 70P' +'

```

```

[49] 'CLASSE DIAGNOSTICA NUMERO ',TJ
[50] 700'+1
[51] 'DESCRICAO?'
[52] '↓',(230'-'),'↓'
[53] DCLASSELJ;J<2500,250' '
[54] G1:'CODIGO DA CLASSE NUMERO ',TJ
[55] K<0
[56] →E3x1v/VCLASSE=K
[57] VCLASSELJ;K
[58] D3:'PROBABILIDADE A PRIORI DA CLASSE ',TJ
[59] PPRILJ;+0
[60] →E1x1(PPRILJ)>1)→PPRILJ;<0
[61] AGLOMERA J
[62] →A2
[63] A1:→E2x1(+/PPRI)≠1
[64] LPRCLU→PRCLU<0≠0
[65] VDBSCLU← 0 15 ≠0
[66] J<0
[67] Y4:→Y1x1NCL<J←J+1
[68] N<0≠0
[69] M←NVAR≠0
[70] L←ECLUDLJ] LELIST LCLUD
[71] K<0
[72] Y3:→Y2x1(ϕL)<(K←K+1
[73] N←N,EVARCLU[LKJ] LELIST VARCLUCLJ;J
[74] →Y3
[75] Y2:M[NJ]+1
[76] →E4x1~^/M=SDESC
[77] →Y4
[78] Y1:'GERACAO DO MODELO TERMINADA'
[79] →0
[80] E1:'VALOR ERRADO DE PROBABILIDADE'
[81] →D3
[82] E2:'ERRO-----A SOMA DAS PROBAB. A PRIORI DEVE SER IGUAL A 1'
[83] 'FAVOR CORRIGIR'
[84] →0
[85] E3:'CODIGO JA EXISTENTE'

```

```

[86] →G1
[87] E4: 'ERRO: VODE NAO ESPECIFICOU TODOS OS DESCRITORES NA ESTRUT. DE CLUSTERS'
      ▽
      VCRITFIM[0]V
      ▽ P←CRITFIM X;V;I;K;J
[1] #OBJ: CALCULO DA PROBABILIDADE DE QUE OS EXAMES CONSIDERADOS
[2] # NO ESTADO X TENHAM REVELADO ALGUMA DOENCA
[3] # (BASEADO NOS PARAMETROS PI)
[4] REV←NCLP0
[5] #V←X^~X0
[6] V←X
[7] I←0
[8] A1:→A2X\NCL<I←I+1
[9] K←1
[10] J←0
[11] B1:→B2X\ANTESTE<J←J+1
[12] →B1X\VC[J]=0
[13] K←KX(1-PI[I;J])
[14] →B1
[15] B2:REVC[I]←1-K
[16] →A1
[17] A2:P←+/PPOSTXREV
      ▽
      VCRITFIMPLUS[0]V
      ▽ P←CRITFIMPLUS X;XR;J;I;ACERT;DIAG;C;PERFIL;PPOS;STEST;CAS;T;MEDIDA;PFR;
      SDESC
[1] #OBJ: ESTIMATIVA DA PROBABILIDADE DE QUE OS EXAMES CONSIDERADOS
[2] # NO ESTADO X TENHAM REVELADO ALGUMA DOENCA
[3] # (ESTIMATIVA DIRETA A PARTIR DO BANCO DE DADOS)
[4] STEST←NCASOSP1
[5] PPOS←PPOST
[6] PPR←PPRI
[7] PPRI←1+NCL
[8] SDESC←NVARP0
[9] XR←X

```

```

[10] J←0
[11] B2:→B1x\NTESTE<J←J+1
[12] →B2x\XRLJ]=0
[13] I←ELDESCLJ] LELIST LDESC
[14] SDESCLJ]←1
[15] →B2
[16] B1:ACERT←CAS+NCLP0
[17] C←0
[18] C1:→C2x\NCASOS<C←C+1
[19] →C1x\STESTEC]=0
[20] T←0
[21] MEDIDA←NTESTE0
[22] L1:→L2x\NVAR<T←T+1
[23] →L1x\SDESCIT]=0
[24] →L1x\MCASOS<T]=T-99
[25] MEDIDAZTESTE[T]]←1
[26] →L1
[27] L2:→C1x\1≠^/MEDIDA=XR
[28] MONTAPERFIL C
[29] DIAGNOSE PERFIL
[30] K←PPOST
[31] DIAG←MCASOS<C;COLR]
[32] CASLDIAG]←CASLDIAG]+1
[33] →C1x\VCLASSE[K]]#DIAG
[34] ACERTLDIAG]←ACERTLDIAG]+1
[35] →C1
[36] C2:P←+/PPOS×ACERT←CAS
[37] PPOST←PPOS
[38] PPRI←PPR

```

```

VDIAGNSELDJ]V
V DIAGNOSE PERF;PERFIL;PC;J;K;I
#OBJ: CALCULO DAS PROBABILIDADES A POSTERIORI (PPOST) DAS CLASSES
# DIAGNOSTICAS NO MODELO COM PRESSUPOSICAO DE INDEPENDENCIA
PERFIL←PERF
PC←NCLP0
A9:J←0

```

```

[6] A2:→A1x\NCL<J+J+1
[7] PCLJ]←x/PERFIL/PCOBS[J;]
[8] →A2
[9] A1:→A3x1(+/PC)>0
[10] J←0
[11] K←0*0
[12] B2:→B9x\NCL<J+J+1
[13] I←0
[14] C1:→B2x1(≠PERF)<I+I+1
[15] →C1x1PERFIL[I;]=0
[16] →C1x1PCOBS[J;I]>0
[17] K←K,I
[18] →B2
[19] B9:PERFIL[K]←0
[20] 'ELIMINEI OS DESCRITORES: ',YZDESCR[K]
[21] →A9
[22] A3:PC←PCxPPRI
[23] PPOST←PC++/PC
      v
vDIAGNPREP[C]v
v DIAGNPREP PERF, MARCA, OCORRTOT, OCORRCLU, J, S, L, NC, I, PCONDCLU, PC
ADBJ: ESTIMACAO DE PARAMETROS E INFERENCIA DIAGNOSTICA (CALCULO DE PPOST)
a EM MODELOS BAYESIANOS DE CLASSIFICACAO DIAGNOSTICA BASEADOS EM
a AGLOMERADOS DE DESCRITORES INTERDEPENDENTES
MARCA←OCORRTOT←OCORRCLU←(NCL, NVAR)*0
PCONDCLU←NCL*0
ASEL DE CASO
J←0
A1:→A2x\NCASOS<J+J+1
→A1x1STREIN[J]=0
#IDENT DA DOENÇA
D←VCLASSE\MCASOS[J;COLR]
NC←0
#SEL CLUSTER
C←ECLUDE[D]
B3:→A1x1C=0

```

```

[16] NC←NC+1
[17] VAR←EVARCLUCJ LELIST VARCLUCD;J
[18] S←PERFLVARJ#~99
[19] VAR←S/VAR
[20] →X1x10<pVAR
[21] MARCALD;NCJ←1
[22] X1:→B2x1v/MCASOSEJ;VARJ=~99
[23] OCCORRTOTD;NCJ←OCCORRTOTD;NCJ+1
[24] →B2x1~^/MCASOSEJ;VARJ=PERFLVARJ
[25] OCCORRCLUCD;NCJ←OCCORRCLUCD;NCJ+1
[26] B2:C←LCLUCD
[27] →B3
[28] #CALC PROB CONDIC DO CLUSTER
[29] A2:D←0
[30] C2:→C1x\NCL<D+D+1
[31] NC←p,ECLUDED] LELIST LCLUD
[32] J←0
[33] C5:→C3x\NC<J←J+1
[34] →C5x\MARCALD;JJ=1
[35] →C4x\OCCORRTOTD;JJ=0
[36] OCCORRCLUCD;JJ←OCCORRCLUCD;JJ+OCCORRTOTD;JJ
[37] →C5x\OCCORRCLUCD;JJ≥PMIN
[38] OCCORRCLUCD;JJ←PMIN
[39] →C5
[40] #ESTIM. DEFAULT DA PROBAB DO CLUSTER J DA DOENCA
[41] C4:I←1↑(J-1)↖ECLUDED] LELIST LCLUD
[42] VAR←EVARCLUCI] LELIST VARCLUCD;J
[43] OCCORRCLUCD;JJ←1+x/NVOBS[VAR]
[44] →C5x\OCCORRCLUCD;JJ≥PMIN
[45] OCCORRCLUCD;JJ←PMIN
[46] →C5
[47] #CALC PROB.CONDICC[PERFIL/DOENCA]
[48] C3:PCONDCLUCD]←x/(NC←MARCALD;J)/NC←OCCORRCLUCD;J
[49] →C2
[50] #CALC PPOST
[51] C1:PC←PCONDCLUXPPRI
[52] →Z1x1+/PC#0

```

```

[53] PC=(PFC)*1E-6
[54] Z1:PPOST+PC+P/PC
      V
      VDIAGPREP[D]V
      V DIAGPREP PERF;MATPROB;NCLUST;MARCA/D;CONFIG/L;PC/C;VAR/M;J
[1]  *OBJ: INFERENCIA DIGNOSTICA (CALCULO DE PPOST) EM MODELOS BAYESIANOS
[2]  *      PREVIAMENTE CALIBRADOS (MODELOS DE CLASSIFICACAO DIAGNOSTICA
[3]  *      BASEADOS EM AGLOMERADOS DE DESCRITORES INTERDEPENDENTES)
[4]  MARCA+(NCL,NVAR)*0
[5]  MATPROB+(NCL,NVAR)*-999
[6]  NCLUST+NCL*0
[7]  D*0
[8]  *SEL DOENCA
[9]  A1:->A2X1NCL(D+D+1
[10] *SEL CLUSTER
[11] C+ECLUD[D]
[12] C1:->A1X1C=0
[13] NCLUST[D]←NCLUST[D]+1
[14] VAR←EVARCLUCC] LELIST VARCLUCD;J
[15] CONFIG←PERF[V]
[16] →R1X1V/CONFIG=-99
[17] →D2X1OCORRIC]=0
[18] L←EPRCLUCC]
[19] D1:->D9X1L=0
[20] →D3X1~^/VOBSCLUCL;J=15*CONFIG,15*0
[21] MATPROB[D;NCLUST[D]]←PRCLUCL]
[22] C5:C←LCLUDEC]
[23] →C1
[24] D3:L←LPRCLUCL]
[25] →D1
[26] APROB. CASO N/ MAJA 1 ESTIM=DEFAULT)
[27] D2:MATPROB[D;NCLUST[D]]←1+x/NVOBSSEVAR]
[28] →C5X1MATPROB[D;NCLUST[D]]≥PMIN
[29] MATPROB[D;NCLUST[D]]←PMIN
[30] →C5
[31] D9:MATPROB[D;NCLUST[D]]←PMIN

```

```

[32] →C5
[33] R1:M←CONFIG#~99
[34] VAR←M/VAR
[35] CONFIG←NVOBS[VAR]
[36] →R2×10=ρVAR
[37] MATPROB[D;NCLUST[D]]←D ESTIMA VAR
[38] →C5
[39] R2:MARCA[D;NCLUST[D]]←1
[40] →C5
[41] ρCALC.PROB. CODIC.CLUSTER
[42] A2:D←0
[43] PC←NCLρ0
[44] A4:→A3×1NCL<D←D+1
[45] PC[D]←x/(NCLUST[D]←~MARCA[D;])/NCLUST[D]←MATPROB[D;]
[46] →A4
[47] A3:PC←PC×PFRI
[48] →Z4×1+PC≠0
[49] PC←(ρPC)ρ1E-6
[50] Z4:PPOST←PC++/PC

```

```

▽DIAGSUPER[D]▽
▽ MODS DIAGSUPER PERF;I;K
MOBJ: APLICAR UMA BATERIA (MODS) DE MODELOS DE ALFA CRESCENTE A UM
a PACIENTE COM OBSERVAÇÕES PEREIL
ρPINF←1
I←0
A1:→0×1(ρMODS)<I←I+1
LEMODC MODSLI]
NOMEMOD
DIAGNPREP PERF
K←ρPOST
'POST: ',ρPOST
→0×1(ρPOST[K]I]ZPINF
→A1
▽

```



```

VNTRAAB[0]V
V NOPAI ENTRAAB X;N
ROBJ: ENTRAR O ESTADO X NA ESTRUTURA ABERTO
  →A1X1DISP=0
A2:ND←DISP
DISP←BPONT[DISP]
BPONT[ND]←AB
AB←ND
PTR[ND]←NOPAI
XIX[ND;J]←X
FENDJ←CALCF X
  →0
[10]
[11] A1:N←LNUMNOS+10
[12] DISP←1+φBPONT
[13] BPONT←BPONT,(φBPONT)+1+1N
[14] BPONTL←BPONTJ←0
[15] F←F,Nφ0
[16] PTR←PTR,Nφ0
[17] XIX←XIX,[1](N,NTESTE)φ0
[18] →A2
V

```

```

VESTIMALD]V
V PROB←D ESTIMA VAR;NCLU;NTOT;J
ROBJ: ESTIMAR A PROBABILIDADE CONDICIONAL DE UMA CONFIGURACAO DE
A OBSERVACOES - PERFLVAR] - DADA UMA DOENCA D
  NTOT←NCLU←0
CONFIG←PERFLVAR]
  →E1X1V/CONFIG="99
J←0
A2:→A1X1NCASOS<J+J+1
  →A2X1(SELJJ)=0)VSTREINLJJ]=0
  →A2X1MCASOSELJ;COLR]#VCLASSE[D]
  →A2X1V/MCASOSELJ;VARJ]="99
NTOT←NTOT+1
  →A2X1~^/MCASOSELJ;VARJ]=CONFIG
NCLU←NCLU+1

```

```

[14] →A2
[15] A1:→X1X\NTOT=0
[16] PROB←NCLU÷NTOT
[17] →X2
[18] X1:PROB←1÷x/NVOBS\VARJ
[19] X2:→0X1PROB2PMIN
[20] PROB←PMIN
[21] →0
[22] E1:'ERRO NOS DADOS DE ENTRADA DA FUNCAO ESTIMA'
      V
      VEXPLORE[O]V
      V SDESC EXPLORER X;F;I;J;TAB;C;CC;DD;ND;NP;S;E;NN;VP;CHI;GL;DIAG;DOENTES;
        CONTR;Z;PE;R
[1] *OBJ: ANÁLISE INDIVIDUAL DE CADA DESCRITOR SELECIONADO COM RELACAO
[2] * A VARIÁVEL DIAGNOSTICA R
[3] * - X=N.COL DA VAR RESPOSTA
[4] TITULO
[5] 'CARACTERISTICAS BASICAS DOS DESCRITORES COM RELACAO A VARIÁVEL:'
[6] (25P' '),DVAR[X;J
[7] 79P' *'
[8] F+I←0
[9] CONFES←CONFE←(NVAR,2)P0
[10] C←'N.CAS SENSIB ESPECIF V.PREDIT CHI-QUAD G.LIB PORC.ERRO'
[11] (25P' DESCRITOR',25P' '),C
[12] 79P' --'
[13] 10
[14] A2:→Z8X\NVAR(I←I+1
[15] →A2X\SDESC[LI]=0
[16] Z←MCASOSL;I]≠'99
[17] PARES←0(2,(+Z))P(Z/MCASOSL;I]),Z/MCASOSL;XJ
[18] CC←VOBS[ELOBSELI] LELIST LVOBSEJ
[19] DD←VOBS[ELOBSELX] LELIST LVOBSEJ
[20] TAB←CC CONTING DD
[21] R←(PCC)P0
[22] J←ELOBSELI]
[23] B2:L←CC\VOBSLJ]

```

```

[24] →C1x1L>PR
[25] RCLJ←POSITC[J]
[26] C1:J←LVOBS[J]
[27] →B1x1J=0
[28] →B2
[29] B1:→A2x1~((+/R=0)>0)^(+/R=1)>0)>
[30] →B3x1F=1
[31] C←(pDD)ρ0
[32] J←ELOBS[X]
[33] B4:L←DD\VOBS[J]
[34] CLLJ←POSITC[J]
[35] J←LVOBS[J]
[36] →B3x1J=0
[37] →B4
[38] B3:F←1
[39] DOENTES←(C=1)/[L2] TAB
[40] ND←+//DOENTES
[41] NP←+//<R=1>/[L1] DOENTES
[42] S←100xNP÷ND
[43] CONTR←(C=0)/[L2] TAB
[44] NC←+//CONTR
[45] NN←+//<R=0>/[L1] CONTR
[46] E←100xNN÷NC
[47] CONFIANCA
[48] DIAG←(R=1)/[L1] TAB
[49] NP←+//DIAG
[50] ND←+//<C=1>/[L2] DIAG
[51] VP←100xND÷NP
[52] CHI←CHIQ TAB
[53] GL←((pC)-1)x((pR)-1)
[54] NC←ND←J←0
[55] K1:→K2x1(pR)<J←J+1
[56] L←0
[57] K3:→K1x1(pC)<L←L+1
[58] →K3x1(R[J]=1)∨C[L]=1
[59] NC←NC+TAB[CJ;L]
[60] →K3x1R[J]=C[L]

```

```

[61] ND←ND+TABLJ;LJ
[62] →K3
[63] K2:PE←100×ND÷NC
[64] A←' '
[65] →Z9×1CHI<GL CHISQUARE 1--ALFA
[66] A←'*S'
[67] Z9:ND←(10 4 TCHI),A,(4 0 TGL), 9 2 TPE
[68] (24pDVARL;J),(5 0 T+ /TAB),(7 2 TS),(7 2 VE),(9 2 TVP),ND
[69] ' ',16424pDVARL;J
[70] →A2
[71] Z8:79p'*'
[72] '*S = ASSOCIACAO SIGNIFICATIVA AO NIVEL DE ',TALFA
[73] 10
[74] PAUSE
[75] TITULO
[76] 'INTERVALOS DE CONFIANCA DAS SENSIBILIDADE E ESPECIFICIDADE '
[77] 'DOS DESCRITORES COM RELACAO A VARIABEL DIAGNOSTICA'
[78] DVARL;J,' (ALFA=0.05)'
[79] 79p'*'
[80] 'DESCRITOR INTERV. CONF. SENSIBILIDADE INTERV. CONF. ESPECIFICIDADE
E'
[81] 10
[82] I←0
[83] J1:→J2×1NVAR(I←I+1
[84] →J1×1DESCCII]=0
[85] J←(CONFELI;J)<0)/ 1 2
[86] L←(CONFELI;J)>100)/ 1 2
[87] CONFELI;J]←0
[88] CONFELI;L]←100
[89] J←(CONFELI;J)<0)/ 1 2
[90] L←(CONFELI;J)>100)/ 1 2
[91] CONFELI;J]←0
[92] CONFELI;L]←100
[93] (5 0 TI),(6p' '), (12 4 TCONFELI;J),(8p' '), 12 4 TCONFELI;J
[94] 10
[95] →J1
[96] J2:79p'*'
v

```

```

V<GAMATESTE[0]V
V SUCX<GAMATESTE X;I;N;EXAFAZER
#OBJ: GERACAO DOS SUCESSORES DO ESTADO X
EXAFAZER<,(EXPERMA~X)/INTESTE
N<pEXAFAZER
SUCX<(N,INTESTE)pX
I<0
A1:→0x\N(I←I+1
SUCX[I;EXAFAZER[I]]+1
→A1
V

```

[1]  
[2]  
[3]  
[4]  
[5]  
[6]  
[7]  
[8]

```

V<HCHAPI[0]V
V H<HCHAPI X;XR;I;HADIC
#OBJ: ESTIMAR A FUNCAO HEURISTICA PARA O ESTADO X
XR<EXPERMA~X
I<H<0
HADIC<0p0
A1:→A2x\INTESTE<I←I+1
→A1x\XR[I]=0
HADIC<HADIC,UTIL[I]x+/PPOSTxQUEC/I]
→A1
A2:H<→ADIUTIL, ' HADIC'
V

```

[1]  
[2]  
[3]  
[4]  
[5]  
[6]  
[7]  
[8]  
[9]

```

V<MANICPIM[0]V
V MANIGPIM;R;TXT;TIT;I
#OBJ: MANIPULACAO DE GERADORES DE PLANOS DE INVESTIGACAO MEDICA
TIT<70p'MANIPULACAO DE GERADORES DE PLANOS DE INVESTIGACAO MEDICA',40p' '
TXT< 11 60 p' '
TXTE1;I<60p'CRIAR UM MODELO GERADOR DE PLANOS DE INVESTIGACAO',30p' '
TXTE2;I<60p'ATUALIZAR UM MODELO',50p' '
TXTE3;I<60p'EXIBIR CARACTERISTICAS DO MODELO CORRENTE',50p' '
TXTE4;I<60p'SELECIONAR DESCRITORES',50p' '
TXTE5;I<60p'SELECIONAR EXAMES PERMITIDOS',50p' '
TXTE6;I<60p'MANIPULAR O ARQUIVO DE ESTRATEGIAS COMPOSTAS',50p' '
TXTE7;I<60p'SELECIONAR CASOS PARA TESTE',50p' '
V

```

[1]  
[2]  
[3]  
[4]  
[5]  
[6]  
[7]  
[8]  
[9]  
[10]

```

[11] TX[18;J]+60p'GERAR PLANOS DE INVESTIGACAO PARA CASOS SELECIONADOS',40p'
[12] TX[19;J]+60p'APLICAR O MODELO CORRENTE A UM PACIENTE NOVO',50p'
[13] TX[10;J]+60p'TRABALHAR COM O ARQUIVO DE MODELOS',50p'
[14] TX[11;J]+60p'DEFINIR CARACTERISTICAS DA ESTRATEGIA ROBUSTA',50p'
[15] X1:R←0 MENU TXT
[16] A'→A',TR
[17] A0:→0
[18] A1:CRIAGER
[19] →X1
[20] A2:ATUGER
[21] →X1
[22] A3:EXIBGER
[23] PAUSE
[24] →X1
[25] A4:PAGE
[26] SDESC←NVAR←0
[27] 'SELECCIONE VARIATIVAS:'
[28] (NVAR←1) EXIBELIST DVAR
[29] SDESC[1,0]←1
[30] →X1
[31] A5:PAGE
[32] 'EXAMES'
[33] ←1
[34] (NTESTE←1) EXIBELIST DTESTE
[35] 10
[36] 'EXAMES ATUALMENTE PERMITIDOS:'
[37] EXPERM/INTESTE
[38] 10
[39] 'SELECCIONE EXAMES PERMITIDOS:'
[40] R←D
[41] →X1×10=1AR
[42] EXPERM←NTESTE←0
[43] EXPERM[R]←1
[44] →X1
[45] A6:ESTCOMP
[46] →X1
[47] A7:PAGE

```

```

[48] 'SELECCIONA CASOS PARA TESTE: '
[49] STEST←NCASOS*0
[50] STEST[A,0]←1
[51] →X1
[52] A8:TESTAGER
[53] PAUSE
[54] →X1
[55] A9:APLICGER
[56] →X1
[57] A10:ARQGER
[58] →X1
[59] A11:PAGE
[60] 'SELECCAO DE ESTRATEGIAS ELEMENTARES COMPONENTES DA ESTRAT. ROBUSTA'
[61] 79*'*'
[62] 'ESTRATEGIAS'
[63] ((1 0 /%NOMESTS)*1) EXIBELIST NOMESTS
[64] 10
[65] 'SELECCIONA ESTRATEGIAS DESEJADAS: '
[66] I←0
[67] →K1 x\I[1]=0
[68] ROB←NUT*0
[69] ROELI←1
[70] K1:10
[71] 'DEFINICAO DE PREFERENCIAS ENTRE AS ESTRATEGIAS ELEMENTARES'
[72] 'FORNECA PESOS DE 1 A 10 OU ZERO PARA DESISTIR: '
[73] I←0
[74] →X1 x\I[1]=0
[75] PREF←NUT*PI,10*1
[76] →X1
V
V MODELS[0]V
V MODELS[I;J;L;S;V;VET;VT;PMIN;F;TIT;TXT
*OBJ: GERENCIAMENTO DE MODELOS BAYESIANOS DE CLASSIFICACAO DIAGNOS--
* TICA COM PRESSUPOSICAO DE INDEPENDENCIA
F←40* '
A1:TXT← 10 60 * '

```

```

[5] TIT←'MANIPULACAO DE MODELOS BAYESIANOS DE DIAGNOSTICO'
[6] TXT[1;J←60p'CRIAR UM NOVO MODELO',F
[7] TXT[2;J←60p'ATUALIZAR O MODELO CORRENTE',F
[8] TXT[3;J←60p'EXIBIR AS CARACTERISTICAS DO MODELO CORRENTE',F
[9] TXT[4;J←60p'ESTABELEZER O CONJUNTO DE CASOS DE TREINAMENTO',F
[10] TXT[5;J←60p'TREINAR O MODELO CORRENTE',F
[11] TXT[6;J←60p'ESTABELEZER O CONJUNTO DE TESTE',F
[12] TXT[7;J←60p'TESTAR O MODELO CORRENTE',F
[13] TXT[8;J←60p'TRABALHAR COM A BASE DE MODELOS',F
[14] TXT[9;J←60p'FUNCIOES ADICIONAIS',F,F
[15] TXT[10;J←60p'TESTAR MODELO DE CLASSES AGLUTINADAS',F
[16] I←0 MENU TXT
[17] A'→B',VI
[18] B0:→0
[19] B1:CRIAMOD
[20] PAUSE
[21] →A1
[22] B3:LISTMOD
[23] PAUSE
[24] →A1
[25] B4:STREIN←'CASOS DE TREINAMENTO' SELECAO STREIN
[26] →A1
[27] B5:TREINAMOD
[28] PAUSE
[29] →A1
[30] B6:I←'A' LEDADO 'SEL.COMPLEMENTAR A DE TREEN?'
[31] →A2x1'N'=141
[32] STEST←~STREIN
[33] →A1
[34] A2:STEST←'CASOS PARA TESTE' SELECAO STEST
[35] →A1
[36] B7:TESTAMOD
[37] PAUSE
[38] →A1
[39] B9:PAGE
[40] 'VOCE DESEJA'
[41] ' 1-AJUSTAR A MATRIZ DE PROBAB.CONDIC.'

```



```

[42] ' 2-... '
[43] ' 0--TERMINAR '
[44] I←D
[45] →L1x\I=1
[46] →A1
[47] L1: 'QUAL O VAL. MIN. DE PROBAB.?'
[48] PMIN←0
[49] →E7x\ (PMIN<0)∨PMIN>1÷NCL
[50] I←0
[51] M2:→M1x\NVAR<I←I+1
[52] →M2x\ (DESCI)=0
[53] VET←ELOBS[I] LELIST LVOBS
[54] →M2x\0=ρVET
[55] K←0
[56] M3:→M2x\NCL<K←K+1
[57] MM←(ρVET)ρ1
[58] SOMA←0
[59] M←0
[60] M7:→M8x\ (ρVET)<M←M+1
[61] →M7x\ (PCOBS[K],VET[M])≥PMIN
[62] MM[M]←0
[63] SOMA←SOMA+PMIN
[64] PCOBS[K],VET[M]←PCOBS[K],VET[M]+PMIN
[65] →M7
[66] M8:→M3x\ (SOMA=0
[67] L←SOMA++/MM/PCOBS[K],VET]
[68] M9: M←0
[69] M10:→M3x\ (ρVET)<M←M+1
[70] →M10x\ (MM[M]=0
[71] PCOBS[K],VET[M]←PCOBS[K],VET[M]x(1-L)
[72] →M10
[73] M1: 'MATRIZ PROBABILIDADES AJUSTADA'
[74] PAUSE
[75] →A1
[76] E9: 'ERRO'
[77] →B9
[78] B10: TITULO

```

```

[179] 'TESTE DO MODELO ',NOMEMOD
[180] 'OBTIDO POR AGLUTINACAO DE CLASSES DO MODELO ',VAGLUTIN
[181] 799'*.
[182] 'PARA REALIZAR O TESTE EXECUTE OS SEGUINTE PASSOS:'
[183] 10
[184] 1- LER O MODELO ',VAGLUTIN
[185] 2- EXECUTAR TESTE DO MODELO ',VAGLUTIN
[186] 3- LER DU CRIAR MODELO ',NOMEMOD
[187] 4- EXECUTAR FUNCAO TESTEAGLUT'
[188] 10
[189] PAUSE
[190] →A1
[191] B8:TXT← 4 28 p' '
[192] TIT←BASE DE MODELOS BAYESIANOS DE CLASSIFICACAO '
[193] TXTL1;J←28p'LISTAR MODELOS EXISTENTES',F
[194] TXTL2;J←28p'LER UM MODELO',F
[195] TXTL3;J←28p'GRAVAR O MODELO CORRENTE',F
[196] TXTL4;J←28p'ELIMINAR UM MODELO',F
[197] I←2 MENU TXT
[198] A1→C',TI
[199] C0:→A1
[100] C1:TITULO
[101] 'MODELOS EXISTENTES'
[102] 799'*.
[103] J←YNCL>0
[104] (J/[1] 5 1 p'YNCL),(J/[1] 5 1 p' '),J/[1] YNOMEMOD
[105] PAUSE
[106] →B8
[107] C2:LEMOD
[108] →B8
[109] C3:J←GRAVAMOD
[110] →B8
[111] C4:→E1x1v/'470711'#6p('A' LEDADO 'SENHA',25p'.')
[112] I←'N' LEDADO 'NUM.DO MODELO A SER ELIMINADO '
[113] →B8x1I=0
[114] YNCLC1]←0
[115] MSSG 'MOD ',(TI),' ELIMINADO'

```

```

[1163] →B8
[1173] E1:MSSG 'SENHA ERRADA'
[1183] →B8
[1193] B2:TXT← 6 60 ' '
[1203] TIT←MODIFICACAO DO MODELO DE CLASSIFICACAO DIAGNOSTICA '
[1213] TXTL1;J←60'ALTERAR O NOME DO MODELO',F
[1223] TXTL2;J←60'MODIFICAR NOME/ CODIGO DE UMA CLASSE',F
[1233] TXTL3;J←60'ATRIBUIR VALORES AS PROBABILIDADES A PRIORI DAS CLASSES',F
[1243] TXTL4;J←60'CALCULAR AS PROB. A PRIORI A PARTIR DA MATRIZ DE CASOS',F
[1253] TXTL5;J←60'CORRIGIR PROBABILIDADES A PRIORI',F
[1263] TXTL6;J←60'ALTERAR A LISTA DE DESCRITORES',F
[1273] I←1 MENU TXT
[1283] A'→D',FI
[1293] D0:→A1
[1303] D1:L←40 LELIN 'NOME ATUAL DO MODELO= ',NOMEMOD
[1313] →B2X1(SL1]=1)∧SL2]=1
[1323] NOMEMOD←L
[1333] →B2
[1343] D2:I←'N' LEDADO 'NUM.CLASSE',Z0'.'.
[1353] I←VCLASSEI1
[1363] →B2X1(I<0)∨I>NCL
[1373] L←25 LELIN 'NOME ATUAL DA CLASSE: ',DCLASSEI;J
[1383] →A6X1(SL1]=1
[1393] DCLASSEI;J←L
[1403] A6:L←+1 LELIN 'CODIGO ATUAL DA CLASSE: ',TVCLASSEI;J
[1413] →B2X1(SL1]=1
[1423] VCLASSEI;J←L
[1433] →B2
[1443] D3:TITULO
[1453] 'ENTRE COM AS PROBABILIDADES A PRIORI DAS CLASSES:'
[1463] 10
[1473] DCLASSE
[1483] 10
[1493] V←,D
[1503] →A7X11=+//V
[1513] 'ERRD:SOMA#1'
[1523] PAUSE

```

```

[153] →B2
[154] A7:→ABX\NCL=0V
[155] 'ERR0:NUM. CLASSES#',VNCL
[156] PAUSE
[157] →B2
[158] A8:PPRI←V
[159] →B2
[160] D4:I←0
[161] Z2:→Z1X\NCL<I←I+1
[162] PPRICIJ←(+SEL/VCLASSELIJ=CLASSE)÷+/SEL
[163] →Z2
[164] Z1:MSSG 'PROBS. A PRIORI:', 6 4 TPPRI
[165] →B2
[166] D6:I←'N' LEDADO '1=NOVA LIST/ 2=+1 DESC/ 3=-1 DESC'
[167] →X1X\I=1
[168] →X2X\I=2
[169] →X3X\I=3
[170] →B2
[171] D5:V←'N' LEDADO 'VALOR MIN. DE PROBAB. A PRIORI'
[172] →X7X\ (VS0)VV21
[173] I←PPRI<V
[174] PPRICI/\NCLJ←PPRICI/\NCLJ+V
[175] J←PPRI
[176] PPRICJ<IJJ←PPRICJ<IJJ-VX+/I
[177] MSSG 'PROB.A PRIORI:', 6 4 TPPRI
[178] →B2
[179] X7:MSSG 'VALOR INVALIDO'
[180] →B2
[181] X1:TITULO
[182] 'ENTRE NUMEROS DOS DESCRITORES:'
[183] V←,D
[184] →X5X\ (+/(V<1)^(V>NVAR))#0
[185] →X5X\ (+/(TIPVARE[V]='T')>TIPVARE[V]='D')#0
[186] SDESC←NVAR#0
[187] SDESC[V]←1
[188] →B2
[189] X2:TITULO

```

```

[190] 'ENTRE NUMEROS DOS NOVOS DESCRITORES:'
[191] V←,D
[192] →B2×1V[1]=0
[193] SDESC[V]←1
[194] →B2
[195] X3:TITULO
[196] 'ENTRE NUMEROS DOS DESCRITORES A RETIRAR:'
[197] V←,D
[198] →B2×1V[1]=0
[199] SDESC[V]←0
[200] →B2
[201] X5:'ERRR'
[202] PAUSE
[203] →B2

```

```

VMODELSC[0]V
V MODELSC:TXT:F;J;TIT
*OBJ: GERENCIAMENTO DE MODELOS BAYESIANOS DE CLASSIFICACAO DIAGNOS--
* TICA BASEADOS EM AGLOMERADOS DE DESCRITORES INTERDEPENDENTES
F←40;
A1:TXT← 10 55 ' '
TX1[1;J]←55'CRIAR UM MODELO',F
TX1[2;J]←55'ATUALIZAR O MODELO CORRENTE',F
TX1[3;J]←55'EXIBIR AS CARACTERISTICAS DO MODELO CORRENTE',F
TX1[4;J]←55'SELECAO DE DESCRITORES',F
TX1[5;J]←55'ESTABELECE O CONJUNTO DE TREINAMENTO',F
TX1[6;J]←55'CALIBRAR (TREINAR) O MODELO CORRENTE',F
TX1[7;J]←55'ESTABELECE O CONJUNTO DE TESTE',F
TX1[8;J]←55'TESTAR O MODELO CORRENTE NAO CALIBRADO',F
TX1[9;J]←55'TESTAR O MODELO CORRENTE CALIBRADO',F
TX1[10;J]←55'TRABALHAR COM O ARQUIVO DE MODELOS',F
TIT←'MANIPULACAO DA DASE DE MODELOS COM AGLOMERADOS DE DESCRITORES'
I←0 MENU TXT
A'→B',VI
B0:→0
B1:CRIAMDDC

```

```

[20] PAUSE
[21] →A1
[22] B2:ATUMODC
[23] →A1
[24] B3:LISTMODC
[25] PAUSE
[26] →A1
[27] B4:SDESC←'DESCRITORES' SELECAO SDESC
[28] →A1
[29] B5:STREIN←'CASOS DE TREINAMENTO' SELECAO STREIN
[30] →A1
[31] B6:TREINAMODC
[32] →A1
[33] B7:I←'A' LEDADO 'SEL. COMPLEMENTAR A DE TREIN.?'
[34] →A2×1'N'=1↑I
[35] STEST←~STREIN
[36] →A1
[37] A2:STEST←'CASOS PARA TESTE' SELECAO STEST
[38] →A1
[39] B8:TESTAMODC 1
[40] PAUSE
[41] →A1
[42] B9:TESTAMODC 2
[43] PAUSE
[44] →A1
[45] B10:BASEMOD
[46] →A1
▽

```

```

▽MONTAP[DJ]▽
▽ VAR MONTAP VOB,J
#OBJ: LANCAR RESULTADOS DE UM DESCRITOR NO PERFIL DO PACIENTE
J←ELOBS[VAR]
A3:→0×1J=0
→A1×1VOBS[J]=VOB
PERFIL[J]←0
→A2

```

```

[7] A1: PERFIL[J]←1
[8] A2: J←LVOBS[J]
[9] →A3
  v

```

```

vMONTAPERFIL[D]v
v MONTAPERFIL C;K;J
ROBJ: MONTAR PERFIL DE UM PACIENTE A PARTIR DO SEU REGISTRO EM
a
  r
  MCASOS.
[2] PERFIL←Qp0
[3] J←0
[4]
[5] A1:→0x1NVAR<J←J+1
[6] →A1x1SDESC[J]=0
[7] →A1x1MCASOS[C;J]=~99
[8] K←ELOBS[J]
[9] A3:→A2x1VOBS[K]=MCASOS[C;J]
[10] K←LVOBS[K]
[11] →E1x1K=0
[12] →A3
[13] A2: PERFIL[K]←1
[14] →A1
[15] E1: 'ERRO EM MONTAPERFIL'
  v

```

```

vPAC[D]v
v PAC;NUM;I;NOME;K;L;MD;PERFIL;M;J;EX;PERFANT;LG;RET;R;ABORTO
ROBJ: AVALIACAO DIAGNOSTICA DE UM NOVO PACIENTE USANDO MODELOS
r
  BAYESIANOS DE CLASSIFICACAO COM PRESSUPOSICAO DE INDEPENDENCIA
  LG←LOGOT
[3] DISPLAY 'PAC'
[4] 'LOGO' SWRI 401←LG
[5] 'NOMS' SWRI(9←'),NOMSIST
[6] 'TXT' SWRI '
[7] S←SREA
[8] NUM←110←SGET 'R'
[9] XB: I←NUMPAC\NUM
[10] →T1x1I>pNUMPAC
[11]

```

```

[12] NOME←NOMESLI;J
[13] TSTEXEC←EXAMESLI;J
[14] PERFANT←PERFIL←PERFISLI;J
[15] →A2
[16] T1:I←0
[17] T2:→T3×1NPAC<I←I+1
[18] →A1×1NUMPACLIJ=0
[19] →T2
[20] A1:DISPLAY 'IDPAC'
[21] PERFANT←PERFIL←Qp0
[22] TSTEXEC←NTESTEpo
[23] S←SREA
[24] NOME←25pSGET 'NOME'
[25] J←SGET 'SEXO'
[26] J←1JLI;1J
[27] →E1×1(J6 1 2)=0
[28] 2 MONTAP J
[29] J←SGET 'COR'
[30] J←1JLI;1J
[31] →E1×1(J6 1 2 3)=0
[32] 1 MONTAP J
[33] J←13pSGET 'IDADE'
[34] →B1×1J<40
[35] →B3×1J≥60
[36] J←2
[37] →A3
[38] B1:J←1
[39] →A3
[40] B3:J←3
[41] A3:3 MONTAP J
[42] A2:DISPLAY 'ANAMN'
[43] 'CAB1' SWRI 'PACIENTE : ',NOME
[44] 'CAB3' SWRI 'SELECCIONE NOVOS EXAMES: ',(34p' '),'/EX. REALIZADOS'
[45] N←T(NTESTE,1)P(NTESTE
[46] 'NUM' SWRI NL;2J
[47] 'ITENS' SWRI(40×NTESTE)PDTESTE,(NTESTE,15)P' '
[48] N←T(NTESTE,1)PTSTEXEC

```



```

[49] 'R' SWRI NL;2]
[50] 'TXT' SWRI 'EXAMES SELECIONADOS:'
[51] 'MSG' SWRI 'PRESSIONE ENTER OU PF1:TERMINAR, PF2:ANALISE'
[52] S+SREA
[53] +0X1(SC1]=1)^SC2]=1
[54] +C1X1(SC1]=1)^SC2]=2
[55] +0X1SC1]=1
[56] EX+,A479SGET 'IN'
[57] L+0
[58] MSSG 'ENTRADA DE DADOS MEDICOS DO PACIENTE'
[59] C2:→C1X1(PEX)<L+L+1
[60] TSTEXECELEX[LJJ+1
[61] K+0P0
[62] J+ELDESCCEX[LJJ
[63] C5:→C4X1J#0
[64] K+K,J
[65] J+LDESCCE[LJ
[66] +C5
[67] C4:M+1+pK
[68] C6:→C2X10=M+M-1
[69] RET+COLHE K[M]
[70] +C1X1ABORTO=1
[71] +C6X1RET=0
[72] M+M+2
[73] +C6X1MS1+pK
[74] L+L-1
[75] +E2X1L=0
[76] K+ELDESCCEX[LJJ LELIST LDESC
[77] M+2
[78] +C6
[79] C1:MSSG 'SELECAO DE ANALISES DIAGNOSTICAS'
[80] DISPLAY 'ANAMN'
[81] 'CAB1' SWRI 'PACIENTE.....',NOME
[82] 'CAB2' SWRI 'SELECIONE MODELOS DESEJADOS'
[83] 'NUM' SWRI(YNCL>0)'/12345'
[84] 'ITENS' SWRI((+/YNCL>0)x40)p(YNCL>0)/[1] YNDMEMOD
[85] 'TXT' SWRI 'MODELOS.....:

```

```

[86] 'MSG' SWRI 'PRESSIONE ENTER OU PF1 PARA TERMINAR'
[87] S←SREA
[88] →0X1S[1]=1
[89] MD←,A47$GET 'IN'
[90] NUMPAC[1]←NUM
[91] PERFISLI;J←PERFIL
[92] EXAMESLI;J←TSTEXEC
[93] NOMESLI;J←NOME
[94] RELAT MD
[95] DISPLAY 'PAC'
[96] 'LOGO' SWRI 408PLG
[97] 'NOMS' SWRI(169, ' '), 'F I M'
[98] 'TXT' SWRI 'NUMERO DE NOVO PACIENTE.: '
[99] J←SUGESTOES, COMENTS. PODEM SER ENVIADOS PARA R.J.MACHADO OU R.B.CHAVES'
[100] 'CAB1' SWRI J
[101] I←(99, ' '), 'IBM DO BRASIL - AV.PASTEUR,138 RIO DE JANEIRO RJ'
[102] 'CAB2' SWRI I
[103] S←SREA
[104] →0X1S[1]≠0
[105] R←10$GET 'R'
[106] →0X1 ' '=1AR
[107] NUM←AR
[108] →0X1NUM=0
[109] →XB
[110] E1:MSSG 'DADOS ERRADOS'
[111] →A1
[112] T3:MSSG 'ARQUIVO DE PACIENTES LOTADO'
[113] →0
[114] E2:MSSG 'FIM DOS DADOS ALCANÇADO'
V
VPESQ[0]V
V PESQ I;J
#OBJ: FORMACAO DE UM AGLOMERADO A PARTIR DO DESCRITOR I
OPEN[1]←0
J←0
B2:→0X1(pV)(J←J+1

```

```

[5] →B2x\I=J
[6] →B2x\CLUCJ]=1
[7] →B2x\LIG[I;J]=0
[8] CLUCJ]←1
[9] 'ASSOC. SIGNIFICATIVA: ',7V[I],V[J]
[10] →B2x\OPENCJ]=0
[11] PESQ J
[12] →B2
  ▽

```

```

▽PROC[0]▽
  ▽ NOPAI PROC X;I; IANT
  *OBJ: PROCESSAR UM SUCESSOR X DO ESTADO DE ENDERECO = NOPAI

```

```

[1] FC←CALCF X
[2] I←IANT←FECH
[3] A2:→A1x\I=0
[4] →B1x\^/X=XIX[I;J]
[5] IANT←I
[6] I←BPONTC[I]
[7] →A2
[8] B1:→0x\FC≥F[I]
[9] F[I]←FC
[10] →K1x\I=FECH
[11] FECH←BPONTC[I]
[12] →K2
[13] K1:BPONTC[IANT]←BPONTC[I]
[14] K2:BPONTC[I]←AB
[15] AB←I
[16] PTRC[I]←NOPAI
[17] →0
[18] A1: I←AB
[19] D2:→C1x\I=0
[20] →D1x\^/X=XIX[I;J]
[21] I←BPONTC[I]
[22] →D2
[23] D1:→0x\FC≥F[I]
[24] F[I]←FC
[25]

```

```

[26] PTRC[I]←NOPAI
[27] →0
[28] C1:NOPAI ENTRAAB X
      V
[1]  VREMOVEMIN[0]V
[2]  V X←REMOVEMIN;I,IANT,NOANT
[3]  AOBJ: REMOVER DE ABERTO O ESTADO COM MENOR F
[4]  NOANT←NO←I←IANT←AB
[5]  FMIN←F[AB]
[6]  X←XIX[I;]
[7]  →A9×IAB=0
[8]  →A4×I BPONT[NO]=0
[9]  A2: →A1×I=0
[10] →B1×I FCI] < FMIN
[11] A3: IANT←I
[12] I←BPONT[I]
[13] →A2
[14] B1: FMIN←FCI]
[15] NOANT←IANT
[16] NO←I
[17] →A3
[18] A1: BPONT[NOANT]←BPONT[NO]
[19] X←XIX[NO;]
[20] →0
[21] A4: AB←0
      →0
      A9: 'ABERTO ESTA VAZIA '
      V

```

```

V$ADM[0]V
V $ADM;S;I;J;F;TXT;IT
AOBJ: APRESENTAR O MENU DE FUNCOES PARA O ESTUDO DOS MODELOS
A BAYESIANOS DE CLASSIFICACAO DIAGNOSTICA COM PRESSUPOSICAO
A DE INDEPENDENCIA DOS DESCRITORES
F←40p;
X1: DISPLAY 'MNGER'

```

```

[6] 'NMS' SWRI NOMSIST
[7] S←SREA
[8] I←SGET 'I'
[9] →A0x1I[1;1]= ' '
[10] J← 6 2 p 'A0A1A2A3A4A5'
[11] A'→',J[(←I[1;1])+1;]
[12] A0:TITULO
[13] 'FIM - SADM'
[14] →0
[15] A3:MODELS
[16] →X1
[17] A4:TITULO
[18] 'USE A FUNCAO PAC DA MS: CARDIAG'
[19] PAUSE
[20] →X1
[21] A5:MSSG 'N.IMPLEM'
[22] →X1
[23] A2:TXT← 4 44 p ' '
[24] TIT←'MANIPULACAO DA ESTRUTURA DE DADOS DO SISTEMA'
[25] TXTC1;J←44p'OBTER EXPLANACAO SOBRE A ESTRUTURA DE DADOS',F
[26] TXTC2;J←44p'CONSULTAR A ESTRUTURA DE DADOS',F
[27] TXTC3;J←44p'ATUALIZAR A ESTRUTURA DE DADOS',F
[28] TXTC4;J←44p'GERAR A ESTRUTURA DE DADOS',F
[29] I←2 MENU TXT
[30] A'→B',VI
[31] B1: REXPLECT
[32] RPAUSE
[33] MSSG 'N.IMPLEM'
[34] →A2
[35] B0:→X1
[36] B4:→E1x1v/(←470711')#6p('A' LEDADO 30p'SENHA',30p'.')
[37] CRIABASE
[38] PAUSE
[39] →A2
[40] B3:→E1x1v/'470711'#6p('A' LEDADO 30p'SENHA',30p'.')
[41] ATUEST
[42] →A2

```

```

[43] E1:MSSG 'SENHA ERRADA'
[44] →X1
[45] B2:TXT← 6 60 P ' '
[46] TIT←CONSULTAS A ESTRUTURA DE DADOS'
[47] TXT[1;]J←60°'CONHECER O NUMERO DE ENTIDADES REPRESENTADAS NO SISTEMA',F
[48] TXT[2;]J←60°'LISTAR AS VARIAVEIS/ EXIBIR SUAS CARACTERISTICAS',F
[49] TXT[3;]J←60°'LISTAR VALORES OBSERVADOS/ EXIBIR SUAS CARACTERISTICAS',F
[50] TXT[4;]J←60°'LISTAR EXAMES/ EXIBIR SUAS CARACTERISTICAS',F
[51] TXT[5;]J←60°'LISTAR MODELOS DIAGNOSTICOS/ EXIBIR SUAS CARACTERISTICAS',F
[52] TXT[6;]J←60°'LISTAR UM SUMARIO DA ESTRUTURA DE DADOS',F
[53] I←1 MENU TXT
[54] A'→C',FI
[55] C0:→A2
[56] C1:QUANEST
[57] PAUSE
[58] →B2
[59] C2:LISTVAR
[60] →B2
[61] C3:LISTVOBS
[62] PAUSE
[63] →B2
[64] C4:LISTTESTE
[65] →B2
[66] C5:→X9X\0#+/YNCL
[67] MSSG 'NAO EXISTEM MODELOS'
[68] →B2
[69] X9:LEMOD
[70] LISTMOD
[71] PAUSE
[72] →B2
[73] C0:→C6
[74] C6:MSSG 'N. IMPLM'
[75] →B2
[76] A1:TXT← 8 50 P ' '
[77] TIT←'PROCESSAMENTO DA MATRIZ DE CASOS (MCASOS)'
[78] TXT[1;]J←50°'CONHECER A ESTRUTURA DA MATRIZ DE CASOS',F
[79] TXT[2;]J←50°'EXIBIR UM CASO',F

```

```

[80] TXTC3;J←50p'SELECCIONAR CASOS',F
[81] TXTC4;J←50p'LISTAR CASOS SELECCIONADOS',F
[82] TXTC5;J←50p'SELECCIONAR VARIÁVEIS',F
[83] TXTC6;J←50p'LISTAR VARIÁVEIS SELECCIONADAS',F
[84] TXTC7;J←50p'CRIAR MATRIZ DE CASOS PARA PROCESSAR EM APL',F
[85] TXTC8;J←50p'ANALISE DE DADOS',F
[86] I←0 MENU TXT
[87] A'→D',VI
[88] D0:→X1
[89] D1:EXPMCASOS
[90] PAUSE
[91] →A1
[92] D2:I←'N' LEDADO 30p'NUM. DO CASO',25p'.
[93] EXIBCASO I
[94] PAUSE
[95] →A1
[96] D3:SEL←'CASOS' SELECCAO SEL
[97] →A1
[98] D4:TITULO
[99] 'CASOS SELECCIONADOS'
[100] 79p'*.
[101] MCASOS[SEL/INCASOS;43]
[102] PAUSE
[103] →A1
[104] D5:SDESC←'VARIÁVEIS' SELECCAO SDESC
[105] →A1
[106] D6:TITULO
[107] 'VARIÁVEIS SELECCIONADAS'
[108] 79p'*.
[109] J←SDESC/INVAR
[110] I←pJ
[111] (V(I,1)pJ),((I,2)p' '),DVAR[J;]
[112] PAUSE
[113] →A1
[114] D7:'INDIQUE OS NUMEROS DAS VARIÁVEIS DESEJADAS:'
[115] J←,D
[116] MAT←MCASOS[SEL/INCASOS;J]

```

```

[117] 'FOI CRIADA 1 MATRIZ DE DADOS CHAMADA -MAT- QUE TEM POR LINHAS'
[118] 'OS CASOS E POR COLUNAS AS VARIAVEIS NA SEQUENCIA DO PEDIDO.'
[119] PAUSE
[120] →A1
[121] DB:MSSG 'USE A WORKSPACE CARDIAN'
[122] V
VSADMA[DIIV
V SADMA,TIT,TXT,R,CC,DDD,TAB,CHI,ALFA,CRIT;V,I,MASS;VV;MGRAM;A;L;CONFS;CO
NFE
[1] ROBJ: APRESENTAR O MENU DE FUNCOES DE ANALISE
[2] X0:TIT←70p'ANALISE DE DADOS MEDICOS',70p'
[3] TXT← 8 60 p'
[4] TXTE1;J←60p'INICIALIZAR O SISTEMA',60p'
[5] TXTE2;J←60p'EXIBIR A MATRIZ DE CASOS',60p'
[6] TXTE3;J←60p'SELECIONAR CASOS',60p'
[7] TXTE4;J←60p'SELECIONAR DESCRITORES',60p'
[8] TXTE5;J←60p'SELECIONAR UMA VARIAVEL DIAGNOSTICA',60p'
[9] TXTE6;J←60p'TAB. CONTING. DESCRITOR X VAR. DIAGN./ TESTE HIPOT.',50p'
[10] TXTE7;J←60p'CARAC. BAS. DE CADA DESCRITOR COM REL A VAR. DIAGN.',50p'
[11] TXTE8;J←60p'ESTUDAR ASSOCIACOES SIGNIFICATIVAS ENTRE DESCRITORES',60p'
[12] X1:TITULO
[13] TIT
[14] 79p'*'
[15] 'VOCE DESEJA'
[16] 10
[17] (2 0 7 8 1 p18),(8 1 p' '),TXT
[18] R←0
[19] A'→A',TR
[20] A0:→0
[21] A1:PAGE
[22] 'PARA INICIALIZAR ESTA MS EXECUTE AS INSTRUCCOES:'
[23] 10
[24] ' )COPY CARDIO ESTRUT'
[25] ' )SAVE'
[26] PAUSE
[27] →X1

```



```

[28] A2:TITULO
[29] 'MATRIZ DE CASOS'
[30] 799 '*'
[31] 'DESCRITORES'
[32] 119 '- '
[33] (3 0 T(NVAR,1)P(NVAR),((NVAR,1)P' '),DVAR
[34] 'INDIQUE AS LINHAS DESEJADAS:'
[35] MCASOS[A,D];]
[36] PAUSE
[37] →X1
[38] A3:PAGE
[39] 'INDIQUE OS CASOS SELECIONADOS:'
[40] CASSEL←NCASOS*0
[41] CASSEL[A,D]←1
[42] →X1
[43] A4:PAGE
[44] 'INDIQUE OS DESCRITORES SELECIONADOS:'
[45] (3 0 T(NVAR,1)P(NVAR),((NVAR,1)P' '),DVAR
[46] SDESCI(NVAR)←0
[47] SDESCI[A,D]←1
[48] →X1
[49] A5:PAGE
[50] V←TIPVAR='D'
[51] VV←+/V
[52] 'SELECIONE UMA VARIÁVEL DIAGNÓSTICA:'
[53] (2 0 T(VV,1)P(V\NVAR),((VV,1)P' '),V/[1] DVAR
[54] VSEL←0
[55] DDD←VOBS[ELOBS[VSEL]] LELIST LVOBS]
[56] →X1
[57] A6:PAGE
[58] 'SELECIONE UM DESCRITOR:'
[59] SD←D
[60] TITULO
[61] V←CASSEL^(MCASOS[;SD]#''99)^(MCASOS[;VSEL]#''99)
[62] PARES←V/[1] MCASOS[;SD,VSEL]
[63] CC←VOBS[ELOBS[SD]] LELIST LVOBS]
[64] DDD←VOBS[ELOBS[VSEL]] LELIST LVOBS]

```

```

[65] TAB<CC CONTING DDD
[66] 10
[67] (120' ), 'TABELA DE CONTINGENCIA'
[68] DVAR[SD;], 'X ', DVAR[VSEL;]
[69] 790'*.
[70] 'DESCRIPTOR / VARIABEL DIAGNOSTICA'
[71] 10
[72] V< 1 0 /PTAB
[73] ((V,15)P' ), 6 0 FTAB
[74] 10
[75] 'CLASSES DO DESCRIPTOR:', FCC
[76] 'CLASSES DA VAR. DIAGN.:', FDDD
[77] 10
[78] CHI<CHI<Q TAB
[79] 'CHI-QUADRADO= ', FCHI
[80] 10
[81] 'TESTE DE HIPOTESE? (S/N)'
[82] →X1X1'N'=1+, 0
[83] 'NIVEL DE SIGNIFICANCIA ALFA?'
[84] ALFA<0
[85] CRIT<(((PCC)-1)X((PDDD)-1)) CHISQUARE 1-ALFA
[86] 10
[87] 'VALOR CRITICO= ', FCRIT
[88] 10
[89] →K1X1CHI>CRIT
[90] 'TESTE NAO SIGNIFICATIVO'
[91] PAUSE
[92] →X1
[93] K1: 'TESTE SIGNIFICATIVO'
[94] PAUSE
[95] →X1
[96] A7: PAGE
[97] 'INDIQUE O NIVEL DE SIGNIFICANCIA ALFA:'
[98] ALFA<0
[99] SDESC EXPLORER VSEL
[100] PAUSE
[101] →X1

```

```

[102] AS:TIT←'ANALISE DE ASSOCIACOES SIGNIFICATIVAS ENTRE DESCRITORES'
[103] TXT← 6 60 p' '
[104] X2:TXT11;J←60p'CRIAR MATRIZES DE ASSOCIACAO ENTRE DESCRITORES',60p' '
[105] TXT12;J←60p'SELECCIONAR UMA CLASSE DA VAR. DIAGNOSTICA',60p' '
[106] TXT13;J←60p'EXIBIR A MATRIZ DE ASSOCIACAO DA CLASSE SELECIONADA',60p' '
[107] TXT14;J←60p'DETERM. CLUSTERS DE ASSOC. SIGNIF. ENTRE DESCRITORES',60p' '
[108] TXT15;J←60p'DETERM. A MATRIZ DE COEF. DE CRAMER P. CLASSE SELEC',60p' '
[109] TXT16;J←60p'EXIBIR A MATRIZ DE COEFICIENTES DE CRAMER',60p' '
[110] TITULO
[111] TIT
[112] 79p '* '
[113] 'VOCE DESEJA'
[114] 10
[115] V← 1.0 /pTXT
[116] (2 0 T(V,1)PIV),((V,1)p' '),TXT
[117] R←D
[118] A'→B',WR
[119] B0:→X0
[120] B1:PAGE
[121] V←TIPVAR='D'
[122] VV←+/V
[123] 'SELECCIONE A VAR. DIAGN. DENTRO DE CUJAS CLASSES SERAO ESTUDADAS AS'
[124] 'ASSOCIACOES'
[125] (2 0 T(VV,1)pV/\NVAR),((VV,1)p' '),DVARCV/\NVAR;J
[126] VSEL←D
[127] DD←VORSELEORS[VSEL] LELIST LVORSEJ
[128] SDES←SDESC
[129] LABELS←SDES/\NVAR
[130] V←+/SDES
[131] MATRIZES←(V,V,NVORSE[VSEL])p0
[132] I←0
[133] C1:→C3x(\NVORSE[VSEL]←I←I+1
[134] SEL←MCASOS[VSEL]=DD[I]
[135] SEL ASSOC SDES
[136] 'MATRIZ PARA ',DVARCVSEL;J,'=' ',(VDD[I]),' CALCULADA'
[137] MATRIZES;I]←MASS
[138] →C1

```

```

[139] C3:PAUSE
[140] →X2
[141] B2:PAGE
[142] 'INDIQUE O VALOR DA VARIÁVEL ',DVAR[VSEL];]
[143] 10
[144] V←(pDD),1
[145] (3 0 TVpDD),(Vp' '),DVOBS[LEOBS[VSEL] LELIST LVORS;]
[146] 10
[147] R←D
[148] CSEL←DD\|R
[149] →X2
[150] B3:TITULO
[151] 'MATRIZ DE ASSOCIAÇÃO PARA ',DVAR[VSEL];]','=',',VDD[CSEL]
[152] MASS←MATRIZES[;]CSEL]
[153] 7 10 4 IMPRMATR MASS
[154] PAUSE
[155] →X2
[156] B4:PAGE
[157] SDES SIGNASSOC MATRIZES[;]CSEL]
[158] PAUSE
[159] →X2
[160] B5:PAGE
[161] MCRAM←SDES CRAMER MATRIZES[;]CSEL]
[162] PAUSE
[163] →X2
[164] B6:TITULO
[165] 'MATRIZ DE COEFICIENTES DE CRAMER PARA ',DVAR[VSEL];]','=',',VDD[CSEL]
[166] 799'*'
[167] 7 10 4 IMPRMATR MCRAM
[168] PAUSE
[169] →X2

```

```

VSADMCL[0]V
V SADMC;TIT;TXT;NIV;I;F;C;D;R;S;J;K;L
#OBJ: APRESENTAR O MENU DE FUNCOES PARA O ESTUDO DOS MODELOS
# BAYESIANOS DE CLASSIFICACAO DIAGNOSTICA BASEADOS EM ALOMERADOS

```

```

[33] A DE DESCRITORES INTERDEPENDENTES
[34] F←40p' ,
[35] A1:TXT← 4 35 p' ,
[36] NIV←1
[37] TIT←'MODELOS DE DIAGNOSTICO COM DESCRITORES INTERDEPENDENTES'
[38] TXTL1;J←35p'OBTER EXPLANACoes SOBRE O SISTEMA ',F
[39] TXTL2;J←35p'GERAR A ESTRUTURA DE DADOS',F
[40] TXTL3;J←35p'MANIPULACAO DA BASE DE MODELOS',F
[41] TXTL4;J←35p'ANALISE DE UM PACIENTE',F
[42] I←2 MENU TXT
[43] 1'→B',VI
[44] B0:PAGE
[45] →0
[46] B1:EXPCLU
[47] PAUSE
[48] →A1
[49] B2:GERCLUST
[50] →0
[51] B3:MODELSC
[52] →A1
[53] B4:ANACASO
[54] →A1
[55] v
[56] vSADMDI[]v
[57] v SADMD;TIT;TXT;R;NUM;EXAME;NOME;PERFIL;V;VV;P
[58] *OBJ: APRESENTAR O MENU DE FUNCoes DE APLICACAO DE MODELOS BAYESIANOS
[59] * DE CLASSIFICACAO DIAGNOSTICA (PRESSUP. INDEPENDENCIA) NA PRATICA
[60] * MEDICA
[61] TIT←70p'AVALIACAO DIAGNOSTICA DE PACIENTES COM DOENCA CORONARIANA',30p' ,
[62] X1:TXT← 9 60 p' ,
[63] TXTL1;J←60p'EXPLANACoes SOBRE O SISTEMA',50p' ,
[64] TXTL2;J←60p'GERAR UM ARQUIVO VAZIO DE PACIENTES',50p' ,
[65] TXTL3;J←60p'LISTAR OS PACIENTES ARQUIVADOS',60p' ,
[66] TXTL4;J←60p'PESQUISAR O NUMERO DE UM PACIENTE',50p' ,
[67] TXTL5;J←60p'EXIBIR DADOS DE UM PACIENTE',50p' ,
[68] TXTL6;J←60p'ELIMINAR UM PACIENTE DO ARQUIVO',50p' ,

```

```

[12] TXTL7:]←60p'ALTERAR PROBAB. A PRIORI DE UM MODELO',50p'
[13] TXTL8:]←60p'ENTRAR/ ATUALIZAR/ ANALISAR DADOS DE UM PACIENTE',50p'
[14] TXTL9:]←60p'CRIAR PLANOS DE INVESTILICAO DIAGNOSTICA',50p'
[15] R←0 MENU TXT
[16] A'→A',TR
[17] A1:TITULO
[18] 4 1 p'
[19] 'O SISTEMA DE APOIO AO DIGNOSTICO MEDICO TEM COMO OBJETIVOS:'
[20] 10
[21] ' - AVALIACAO DAS PROBABILIDADES DE DOENCA APRESENTADAS PELO PACIENTE'
[22] ' - SEGUNDO DIVERSOS MODELOS.'
[23] ' - ELABORACAO DE PLANOS DE INVESTIGACAO SEGUNDO ESTRATEGIAS ESTABELE--'
[24] ' - DAS PELO MEDICO.'
[25] ' - MANIPULACAO DE DADOS DE PACIENTES .(COLETA,ARQUIVAMENTO,APRESEN--'
[26] ' - TACAO.'
[27] PAUSE
[28] PAGE
[29] 'UTILIZACAO DO SISTEMA'
[30] 21p'--'
[31] '←TODAS AS FUNCOES DO SISTEMA ESTAO INTEGRADAS SOB A FORMA DE MENUS IN--'
[32] ' - TERLIGADOS.'
[33] '←PARA INVOCAR O MENU PRINCIPAL USE A FUNCAO SADMD. O SISTEMA E AUTOEX--'
[34] ' - PLICATIVO.'
[35] '←OS DADOS DE PACIENTES SAO GRAVADOS NA PROPRIA WORKSPACE. CASO VOCE DE--'
[36] ' - SEJE ARQUIVA-LOS PARA USO FUTURO NAO ESQUECA DE NO FINAL DA SESSAO '
[37] ' - USAR O COMANDO: )SAVE'
[38] '←O REGISTRO DO PACIENTE E IDENTIFICADO POR UM NUMERO ARBITRARIO FOR--'
[39] ' - NECIDO PELO USUARIO.'
[40] '←O SISTEMA PERMITE PROCESSAR OS DADOS DO PACIENTE EM VARIAS SESSOES,'
[41] ' - POR EX:'
[42] ' - ENTRADA DA HISTORIA DO PACIENTE'
[43] ' - AVALIACAO DIAGNOSTICA'
[44] ' - ELABORACAO DE PLANO DE INVESTIGACAO'
[45] ' - ENTRADA DE RESULTADOS DE EXAMES'
[46] ' - NOVA AVALIACAO DIAGNOSTICA'
[47] ' - E ASSIM POR DIANTE ATE UMA DECISAO FINAL.'
[48] PAUSE

```

```

[49] PAGE
[50] 'GERACAO DO SISTEMA'
[51] 18p'-'
[52] 10
[53] '-A GERACAO PROPRRIAMENTE DITA DO SISTEMA ENVOLVENDO A DEFINICAO DOS'
[54] ' DESCRITORES, TIPOS DE DOENCAS E A CRIACAO E CALIBRACAO DE MODELOS'
[55] ' DIGNOSTICOS E FEITA EM UMA WS AUXILIAR DE NOME CARDIO.'
[56] '-UMA VEZ REALIZADO ESSE TRABALHO VOCE DEVE TRANSFERIR A BASE DE DA-'
[57] ' DOS PARA A WS EM QUE VOCE ESTA TRABALHANDO.'
[58] '-PARA TANTO USE OS COMANDOS:'
[59] ' )LOAD NOMESUAMS'
[60] ' )COPY CARDIO ESTRUT'
[61] ' )SAVE'
[62] PAUSE
[63] →X1
[64] A2:PAGE
[65] 'SENHA?'
[66] S←470714
[67] R←0
[68] →E1X1R#S
[69] GERAPAC
[70] PAUSE
[71] →X1
[72] E1:MSSG 'SENHA ERRADA'
[73] →X1
[74] A3:PAGE
[75] PACIENTES
[76] PAUSE
[77] →X1
[78] A4:PAGE
[79] 'NOME DO PACIENTE?'
[80] NUMERO 25p,0,25p'
[81] PAUSE
[82] →X1
[83] A5:PAGE
[84] 'NUMERO DO PACIENTE?'
[85] R←NUMPAC10

```

```

[86] →E2x1R>ρNUMPAC
[87] PERFIL←PERFIS[R;J]
[88] EXAME←EXAMES[R;J]
[89] NOME←NOMES[R;J]
[90] NUM←NUMPAC[R]
[91] RELAT 0p0
[92] →X1
[93] A6:PAGE
[94] ELIMINE
[95] PAUSE
[96] →X1
[97] A7:PAGE
[98] 'INDIQUE O NUMERO DO MODELO'
[99] 10
[100] V←YNCL≠0
[101] VV←(+/V),1
[102] (2 0 7VVpW/15),(VVp' '),V/[1] YNOMEMOD
[103] R←0
[104] 'PROBABILIDADES A PRIORI DO MODELO ',VR
[105] 10
[106] V←(0 1 /pYPPRI)ρ0
[107] V[1]YNCL[R]←1
[108] (V/[1] YDCLASSE[R;J]),((+/V),2)p': '), 8 6 7((+/V),1)pV/[1] YPPRI[R;J]
[109] 10
[110] 'FORNECA OS NOVOS VALORES:'
[111] '(LEMBRE-SE QUE VALORES NULOS ELIMINAM AS CLASSES CORRESPONDENTES)'
[112] P←0
[113] →E3x1(pP)≠YNCL[R]
[114] →E4x11#+/P
[115] YPPRI[R;J]←(0 1 /pYPPRI)pP,15p0
[116] 10
[117] 'OK!!'
[118] PAUSE
[119] →X1
[120] E3:MSSG 'NUMERO INCORRETO DE PROBABILIDADES'
[121] →X1
[122] E4:MSSG 'ERRO: SOMA#1'

```



```

[123] →X1
[124] AB:PAC
[125] →X1
[126] A9:MSSG 'NAO IMPLEMENTADO'
[127] →X1
[128] E2:MSSG 'NUMERO INEXISTENTE'
[129] →X1
[130] A0:→0
      V
      VSADMGLDJV
      V SADMGR;TIT,TXT
#OBJ: APRESENTAR O MENU DE FUNCOES PARA ESTUDO DE GERADORES DE PLANOS
#
# CABEC←'GERADOR DE PLANOS DE INVESTIGACAO MEDICA'
TIT←70;'GERADORES DE PLANOS DE INVESTIGACAO MEDICA',70p'
TXT← 4 60 p'
TXT1;J←60;'EXPLANACAO SOBRE O SISTEMA',40p'
TXT2;J←60;'GERAR A ESTRUTURA DE DADOS',60p'
TXT3;J←60;'CONSULTAR/ATUALIZAR A ESTRUTURA DE DADOS',40p'
TXT4;J←60;'TRABALHAR COM GERADORES DE PLANOS DE INVESTIGACAO',40p'
X1:R←1 MENU TXT
  1'→A',VR
A0:→0
A1:EXPCPIM
  →X1
A2:TITGER
  'PARA GERAR A ESTRUTURA DE DADOS VOCE DEVE PREPARAR OS MODELOS BAYE--'
  'SIANDS COM QUE VAI TRABALHAR USANDO O SISTEMA SADM.'
  'A SEGUIR DEVE EXECUTAR OS SEGUINTE COMANDOS:'
  ' )COPY WSSADM ESTRUT ESTRUT1'
  ' )GERGPIM'
  ' )SAVE'
  PAUSE
  →X1
A3:ATUGPIM
  →X1

```

```

[26] A4:MANIGPIM
[27] →X1
      v
      vSIGNASSOC[0]v
      v SDESC SIGNASSOC MASS,V,I;J;GL;LABELS;VCRITICO
      AOBJ: DETERMINAR ASSOCIACOES SIGNIFICATIVAS ENTRE DESCRITORES
[2] 'NIVEL DE SIGNIFICANCIA PARA TESTE DE HIPOTESE DE INDEPENDENCIA?'
[3] NS←0
[4] VCRITICO←15p0
[5] J←0
[6] K1:→K2x115<J←J+1
[7] VCRITICO[0]J←J CHISQUARE 1-NS
[8] →K1
[9] K2:LABELS←V←SDESC/\pSDESC
[10] LIG←(pMASS)p0
[11] I←0
[12] A1:→A3x1(pV)(I←I+1
[13] J←I-1
[14] A2:→A1x1(pV)(J←J+1
[15] GL←(NVOBS[I]-1)XNVOBS[J]-1
[16] →A2x1MASS[I;J](VCRITICO[GL]
[17] LIG[I;J]←LIG[J;I]+1
[18] →A2
[19] A3:TITULO
[20] 'MATRIZ INDICANDO AS ASSOCIACOES SIGNIFICATIVAS ENTRE DESCRITORES AD'
[21] 'NIVEL DE SIGNIFICANCIA DE ',VNS
[22] 79p'x'
[23] 14 5 0 IMPMATR LIG
[24] CLUST LIG
      v
      vTABCONT[0]v
      v TAB←ORD TABCONT PARES;K;I;C;D;N
      AOBJ: CRIAR 1 TAB. DE CONTINENCIA A PARTIR DE PARES DE VALORES DAS 2
[2] a VARIAVEIS ENVOLVIDAS.
[3] a - PARES = MATRIZ Nx2 CONTENDO OS PARES DE VALORES DAS VARIAVEIS.

```

```

[4]  A - ORD = 1 COLOCA A 1A VAR. DE PARES EM LINHAS DE TAB.
[5]  A - ORD = 2 COLOCA A 2A VAR. DE PARES EM LINHAS DE TAB.
[6]  A - TAB = TABELA DE CONTINGENCIA.
[7]  N= 1 0 /PPARES
[8]  DETERMINACAO DAS CLASSES
[9]  C←A(MNp 1 0 /PARES)
[10] C←PARESC;1]
[11] CC←(0p0),C[1]
[12] I←1
[13] A2:→A1x\N<I←I+1
[14] →A2x\C[1]=C[1-1]
[15] CC←CC,C[1]
[16] →A2
[17] A1:I←1
[18] D←A0Np 0 1 /PARES
[19] D←PARESD;2]
[20] DD←(0p0),D[1]
[21] A4:→A3x\N<I←I+1
[22] →A4x\D[1]=D[1-1]
[23] DD←DD,D[1]
[24] →A4
[25] *CONSTRUCAO DE TAB
[26] A3: TAB←((pCC),pDD)p0
[27] K←0
[28] B1:→B2x\N<K←K+1
[29] L←CC\PARESEK;1]
[30] M←DD\PARESEK;2]
[31] TAB[L;M]←TAB[L;M]+1
[32] →B1
[33] B2:→0x\ORD=1
[34] TAB← 2 1 @TAB
      v
vTESTAMOD[0]v
v TESTAMOD;J;I;IND;H;N;PARES
[1] *OBJ: TESTAR O DESEMPENHO DE UM MODELO DE CLASSIFICACAO COM PRESSUP.
[2] A DE DESCRITORES INDEPENDENTES

```

```

[3]  PROBPPOST←(0,NCL)P0
[4]  N←H←0
[5]  TITULO
[6]  'TESTE DE CLASSIFICADOR BAYESIANO: ',NOMEMOD
[7]  79P'!'
[8]  →E2X\AGLUTIN#0
[9]  10
[10] →D1X\COLR#0
[11] 'PRESSUP.CLASSES REAIS JA FORNECIDAS'
[12] →D2
[13] D1:CLASSE←MCASOSL;COLR]
[14] D2:J←0
[15] DIAG←NCASOSP0
[16] →E1X\0=+/STEST
[17] 'CASO  DIAGNOST  CL. REAL  PROBAB.CLASSES'
[18] 47P'!'
[19] A1:→A2X\NCASOS<J+J+1
[20] →A1X\SELLJ]=0
[21] →A1X\STESTCJ]=0
[22] MONTAPERFIL J
[23] DIAGNOSE PERFIL
[24] PROBPPOST←PROBPPOST,[1] PPOST
[25] I←PPPOST
[26] DIAGCJ]←VCLASSE[[1]]
[27] (4 0  VMCASOSLJ;43]),(7 0  VDIAGCJ]),(10 0  VCLASSECJ]),(5P' '), 7 3  VPPOST
[28] H←H+ENTROP PPOST
[29] N←N+1
[30] →A1
[31] A2: 3 1 P' '
[32] 'PROB. A PRIORI DAS CLASSES : ',VPPRI
[33] 10
[34] IND←(+/STEST/DIAG=CLASSE)÷+/STEST
[35] 10
[36] 'INDICE PORCENTUAL DE DIAGNOSTICOS CORRETOS= ', 7 2  V100XIND
[37] 10
[38] PARES←0(2, (+/STEST)P (STEST/DIAG),STEST/CLASSE
[39] TAB←VCLASSE[AVCLASSE] CONTING VCLASSE[AVCLASSE]

```

```

[40] 'TABELA DE CONTINGENCIA - CLASSE CALCULADA x CLASSE REAL'
[41] 10
[42] TAB
[43] 10
[44] 'CHIRUADRADO: ',(VCHIQ TAB), ' / GRAUS LIB: ',V(NCL-1)*2
[45] 10
[46] 'ENTROPIA MEDIA : ',VH+N
[47] 10
[48] 'GANHO MEDIO DE INFORMACAO'
[49] ' COM REL. A SIT. DE CLASSES EQUIPROVAVEIS : ',V(ENTROP NCL*(1+NCL))-H+N
[50] ' COM REL. A PROB. A PRIORI FORNECIDAS : ',V(ENTROP PPRI)-H+N
[51] ->0
[52] E1:'FALTA SEL CASOS P. TESTE'
[53] ->0
[54] E2:'MODELO DE CLASSES AGLUTINADAS'
[55] 'PARA TESTAR USE A OPCAO 10 DO MENU PRINCIPAL'
v
vTESTAMODCLOJv
v TESTAMODC TIP;J;I;IND;H;N;PARES
MOBJ: TESTAR O DESEMPENHO DE UM MODELO DE CLASSIFICACAO DIAGNOSTICA
a BASEADO EM AGLOMERADOS DE DESCRITORES INTERDEPENDENTES
N+H->0
TITULO
'TESTE DE CLASSIFICADOR BAYESIANO: ',NOMEMOD
'USANDO AGLOMERADOS DE DESCRITORES INTERDEPENDENTES'
799'*'
10
->D1X1COLR#0
'PRESSUP.CLASSES REAIS JA FORNECIDAS'
->D2
D1:CLASSE+MCASOSL;COLR]
D2:J+0
DIAG+NCASOSPO
->E1X10=+/STEST
'CASO DIAGNOST CL. REAL PROBAB.CLASSES'
479'-'

```

```

[18] A1:→A2×1NCASOS<J+J+1
[19] →A1×1SEL[J]=0
[20] →A1×1STEST[J]=0
[21] →B2×1TIP=1
[22] DIAEPREP MCASOS[J:]
[23] →B3
[24] B2:DIAGNPREP MCASOS[J:]
[25] B3:I→PPPOST
[26] DIAG[J]←VCLASSEL[1:]
[27] (4 0 VMCASOS[J;43]),(7 0 VDIAG[J]),(10 0 VCLASSEL[J]),(5p' '), 7 3 VPPOST
[28] H←H+ENTROP PPOST
[29] N←N+1
[30] →A1
[31] A2: 3 1 p' '
[32] 'PROB. A PRIORI DAS CLASSES : ', VPPRI
[33] 10
[34] IND←(+/STEST/DIAG=CLASSE)++/STEST
[35] 10
[36] 'INDICE PORCENTUAL DE DIAGNOSTICOS CORRETOS= ', 7 2 V100×IND
[37] 10
[38] PARES←0(2,(+/STEST))p(STEST/DIAG),STEST/CLASSE
[39] TAB←VCLASSEL[VCLASSE] CONTING VCLASSEL[VCLASSE]
[40] 'TABELA DE CONTINGENCIA - CLASSE CALCULADA x CLASSE REAL'
[41] 10
[42] TAB
[43] 10
[44] 'CHIQUADRADO: ',(VCHIQ TAB), ' / GRAUS LIB: ',V(NCL-1)*2
[45] 10
[46] 'ENTROPIA MEDIA : ',VH÷N
[47] 10
[48] 'GANHO MEDIO DE INFORMACAO'
[49] ' COM REL. A SIT. DE CLASSES EQUIPROVAVEIS : ',V(ENTROP NCLP1÷NCL)-H÷N
[50] ' COM REL. AS PROB. A PRIORI FORNECIDAS : ',V(ENTROP PPRI)-H÷N
[51] →0
[52] E1:'FALTA SEL CASOS P. TESTE'

```

```

[1] V'TREINAMOD[0]V
[2] V TREINAMOD;R;V;P;I;J;K;CCLASS;CON;END;N;L;A
[3] AOBJ: CALIBRAR UM MODELO BAYESIANO DE CLASSIFICACAO DIAGNOSTICA COM
[4] A PRESSUPOSICAO DE INDEPENDENCIA
[5] TITULO
[6] P←'TREINAMENTO DE CLASSIFICADOR BAYESIANO'
[7] R←'PRESSUPOSICAO DE INDEPEND. CONDICIONAL DE DESCRITORES'
[8] P
[9] R
[10] 79P '* '
[11] →Y1X1AGLUTIN=0
[12] 'MODELO OBTIDO POR AGLUTINACAO DE CLASSES.'
[13] 'O TREINAMENTO NAO SE APLICA'
[14] →0
[15] Y1:→B7X1COLR≠0
[16] 'VOCE DESEJA SELEC. COMO CLASSIF. DIAGNOSTICA ALGUMA VARIAVEL'
[17] ' DA MATRIZ DE CASOS ?(S/N)'
[18] B1:→B3X1'N'=1↑,0
[19] 'QUAL?'
[20] V←(TIPVAR='D')V'TIPVAR='T'
[21] ((pV),f)pv/1pv),((pV),f)p' ' '),V/[1] DVAR
[22] COLR←0
[23] CLASSE←MCASOS;COLR]
[24] →B5
[25] B3: 'FORNECA O VETOR CONTENDO A CLASSIF. DIAGNOSTICA DE CADA'
[26] ' CASO DA MATRIZ DE CASOS:'
[27] COLR←0
[28] CLASSE←,10
[29] →E4X1NCASOS≠pCLASSE
[30] →B5
[31] B7:CLASSE←MCASOS;COLR]
[32] B5:→E2X10=+/SEL
[33] →E3X10=+/STREIN
[34] PAUSE
[35] PAGE
[36] P
[37] R
[38] 70P '---'

```

```

[373] 'MODELO : ',NOMEMOD
[383] 70 p '---'
[393] 10
[403] 'CLASSES'
[413] 10
[423] (2 0  $\forall$ (NCL,1) $\rho$ VCLASSE),((NCL,1) $\rho$ ' '),DCLASSE
[433] 10
[443] 'DESCRITORES'
[453] 10
[463] (( $\forall$ (+/SDESC),3) $\rho$  3 0  $\forall$ SDESC/INVAR),((+/SDESC),2) $\rho$ ' '),DVAR[SDESC/INVAR;J
[473] 10
[483] 'PROBABILIDADES A PRIORI'
[493] 10
[503] 9 6  $\forall$ PPRI
[513] CCLASS $\leftarrow$ NCL $\rho$ 0
[523] PCOBS $\leftarrow$ (NCL,R) $\rho$ 0
[533] I $\leftarrow$ 0
[543] F1: $\rightarrow$ (NCASOS(I $\leftarrow$ I+1))/F2
[553]  $\rightarrow$ F1 $\times$ SEL[I]=0
[563]  $\rightarrow$ F1 $\times$ STREIN[I]=0
[573] L $\leftarrow$ VCLASSE\CLASSE[I]
[583] CCLASS[L] $\leftarrow$ CCLASS[L]+1
[593] J $\leftarrow$ 0
[603] G1: $\rightarrow$ F1 $\times$ INVAR $\leftarrow$ J $\leftarrow$ J+1
[613]  $\rightarrow$ G1 $\times$ SDESC[L;J]=0
[623] V $\leftarrow$ MCASOS[L;J]
[633]  $\rightarrow$ G1 $\times$ V=-99
[643] K $\leftarrow$ ELOBSE[L;J]
[653] G4: $\rightarrow$ G3 $\times$ V=VOBSE[K]
[663] K $\leftarrow$ LVOBSE[K]
[673]  $\rightarrow$ E5 $\times$ K=0
[683]  $\rightarrow$ G4
[693] G3:PCOBS[L;K] $\leftarrow$ PCOBS[L;K]+1
[703]  $\rightarrow$ G1
[713] F2:10
[723] 'CASOS DE TREINAMENTO'
[733] 10

```



```

[111] ' (V.OBS. NO EIXO VERT. E CLASSES NO EIXO HORIZ.)'
[112] 10
[113] 'DESCR. V.OBS.1 CLASSES'
[114] A+VVp 5 0 TV/ZDESCR
[115] A,[2](VVp 5 0 TV/VOBS),[2](((+/V),3)p' '),[2] 7 4 T0V/PCOBS
[116] 10
[117] →0
[118] E5: 'ERRO:NAD ACHEI VAL.OBS.'
[119] →0
[120] E2: 'FALTA SELEC.CASOS'
[121] →0
[122] E3: 'FALTA SELEC.CASOS TREINAMENTO'
[123] →0
[124] E4: 'ERRO:NUM.CASOS.INADEQ.'
▽
▽TREINAMODC[1]▽
▽ TREINAMODC;D;TC;J;C;I;VAR
*OBJ: CALIBRAR UM MODELO BAYESIANO DE CLASSIFICACAO DIAGNOSTICA
A BASEADO EM AGLOMERADOS DE DESCRITORES INTERDEPENDENTES
[1] TC+15
[2] EPRCLU+(pLCLUD)P0
[3] LPRCLU+PRCLU+0P0
[4] VOBSCLU+(0,TC)P0
[5] OCCORR+(pLCLUD)P0
[6] I+0
[7] A1:→A2x\NCASOS<I+I+1
[8] →A1x\{SELIJ=0}▽STREIN[1]=0
[9] D+VCLASSE\MCASOS[1];COLR]
[10] C+ECLUDED]
[11] B1:→A1x\IC=0
[12] VAR+,EVARCLUCC] LELIST VARCLU[D;]
[13] →B2x\V/MCASOS[1];VARJ=-99
[14] OCCORR[C]+OCCORR[C]+1
[15] V+TCpMCASOS[1];VARJ,NVARp0
[16] J+EPRCLUCC]
[17] C1:→C2x\J=0

```

```

[74] (SEL^STREIN)/INCASOS
[75] 10
[76] 'NUMERO DE CASOS POR CLASSE'
[77] 10
[78] CCLASS
[79] 10
[80] 'MATRIZ DE OCORRENCIAS: OBSERVACAO X CLASSE'
[81] 10
[82] V←ZDESCR(SDESC=1)/INVAR
[83] 'DESCR. V.OBS.1 CLASSES'
[84] VV←(+/V),5
[85] (VVP 5 0 TV/ZDESCR),[2](VVP 5 0 TV/VOBS),[2]((+/V),3)0' ',[2]TVV/PCOBS
[86] 10
[87] ACALC PROB
[88] I←0
[89] H3:→H8xINCL(I+I+1
[90] J←0
[91] H2:→H3xINVAR(J+J+1
[92] →H2x1SDESC[J]=0
[93] →H2x1NVOBS[J]=0
[94] END←CON←NVOBS[J]0
[95] N←1
[96] K←ELOBS[J]
[97] H5:END[N]←K
[98] CON[N]←PCOBS[I,K]
[99] N←N+1
[100] K←LVOBS[K]
[101] →H4x1K=0
[102] →H5
[103] H4:PCOBS[I,END]←PCOBS[I,END]++/CON
[104] →H2x11=+/PCOBS[I,END]
[105] 'ERRO SOMA PROBAB≠1 CL:',(VI),' POS:',TEND
[106] PCOBS[I,END]←(END)01+NVOBS[ZDESCR]END[I,J]
[107] 'PROBAB. AJUSTADAS PARA: ',7PCOBS[I,END]
[108] →H2
[109] H8:10
[110] 'MATRIZ DE PROBAB. CONDICIONAIS DE VALOR OBSERVADO x CLASSE'

```

```

[20] →C3x11=^/VOBSCLUJ]J]=V
[21] J+LPRCLUJ]
[22] →G1
[23] C3:PRCLUJ]J]+PRCLUJ]J]+1
[24] →B2
[25] C2:LPRCLU+LPRCLU,EPRCLUJ]
[26] VOBSCLU+VOBSCLU,G1] V
[27] EPRCLUJ]J]+LPRCLU
[28] PRCLU+PRCLU,1
[29] B2:C+LCLUJ]
[30] →B1
[31] A2:D+0
[32] A4:→A3x1NCL<D+D+1
[33] C+ECLUJ]
[34] F1:→A4x1C=0
[35] J+EPRCLUJ]
[36] G1:→F2x1J=0
[37] PRCLUJ]J]+PRCLUJ]J]+OCORR]C]
[38] J+LPRCLUJ]
[39] →G1
[40] F2:C+LCLUJ]
[41] →F1
[42] A3:OCORR(OCORR#0)/1+OCORR]J]+1
[43] MSSE 'MODELO CALIBRADO'

```

---

---

## **BIBLIOGRAFIA**

---

---

1. AITCHINSON, J., MOORE, M.F., WEST, S.A., TAYLOR, T.R. - "Consistency of Treatment Allocation in Thyrotoxicosis", Quart. J. Medicine, vol. 167, pp. 575-583, (1973).
2. ALPÉROVITCH, A., KNILL-JONES, R.P., LE MINOR, M., LACAINE, F., POYNARD, T., "Assessment of the Costs of Diagnostic Errors; Methodology and Preliminary Results" - Evaluation of Efficacy of Medical Action. Ed. Alperovitch, A.; de Dombal, F.T.; Grémy, F., North Holland, 1979.
3. ALPÉROVITCH, A., LELLOUCH, J., "Methods for Aiding Medical Decision: Application to Diagnosis of Round Application to Diagnosis of Round Intra-Thoracic X-Rays Picture", Comp. Biomed. Res., vol. 7, pp. 127-141, (1974).
4. ANDERSON, J.A., "Computer diagnosis: Statistical Aspects" Brit. Med. Bull., vol. 24, pp. 230-255, (1968)
5. ANDERSON, J., "The Computer: Medical Vocabulary and Information", Brit. Med. Bull., vol. 24, pp. 194-198, (1968)
6. APL Data Interface II, Program description/operation manual, IBM SH206147, (1981).
7. ARAÚJO, J.L., "Evaluating and Accelerating Heuristic Search: A Proposal", COPPE/UFRJ, Publicação Técnica, nº 174, 1974.
8. BAKWIN, H., "Pseudoxia Pediatrica", The N. Engl. J. Med., vol. 232, nº 24, pp. 691-697, (1945).
9. BARNETT, G.O., BAILLIEUL, J.B., FARGUHAR, B.B., "The Testing of Clinical Judgement - An Experimental Computer Based Measurement of Sequential Problem - Solving Ability", Jacques J.A. Ed. - Computer Diagnosis and Diagnostic Methods, C. Thomas, 1972.
10. BARON, D.N., FRASER, P.M., "Medical Applications of Taxonomic Methods", Brit. Med. Bull, vol. 24, pp. 236-240, (1968)

11. BEAMAN, P.D., JUSTICE, N.S., BARNETT, G.O., "A Medical Information System and Data Language for Ambulatory Practices", Computer, pp. 9-17, (Nov. 1979).
12. BEN-BASSAT et alii, "Pattern-Based Interactive Diagnosis of Multiple Disorders: The MEDAS System", IEEE Trans. PAMI, Vol. 2, n° 2, pp. 148-160, (1980).
13. BENEKEN, J.E.W., BLOM, J.A., JARRITSMA, F.F., NANDORFF, A.; SPIEDIJK, J., "Prognosis, Trend and Prediction in Patient Management", J. Biomed. Eng., vol. 1, pp. 185-200, (1979)
14. BERAUD, C., "Quality of Outcome or Efficacy of Medical Process: Which Criteria Should be Chosen - The Point of View of a Clinician Gastroenterologist", Alperovitch, de Dombal, Grémy (eds), Evaluation of Efficacy of Medical Action, North Holland, 1979.
15. BETAQUE, N.E., GORRY, G.A., "Automating Judgemental Decision Making for a Serious Medical Problem", Management Science, vol. 17, n° 8, pp. 421-434, (1971).
16. BILLEWICZ, W.Z., CHAPMAN, R.S., CROOKS, J., DAY, M.E., GOSSAGE, J., "Statistical Methods Applied to the Diagnosis of Hypothyroidism", Quart. J. Med., Vol. 38, pp. 150-155, (1969)
17. BLEICH, H.L., "Computer Based Consultation", Amer. J. Med., Vol. 53, pp. 285-291, (1972).
18. BLEICH, H.L., "Computer Evaluation of Acid-Base Disorders", The J. of Clinical Invest., vol. 48, pp. 1689-96, (1969).
19. BOYLE, J.A. et alii, "Construction of a Model for Computer - Assisted Diagnosis: Application to the Problem of Non Toxic Goitre", Quart. J. Med., vol. 140, pp. 565-588, (1966).

20. BOYLE, J.A., "Computer Diagnosis: Clinical Aspects", Brit. Med. Bull., vol. 24, pp. 224-229, (1968).
21. BRODMAN, K., van WOERKOM, A.J., ERDMAN, A.J., GOLDSTEIN, L.S., "Interpretation of Symptoms with a Data Processing Machine", AMA Arch. of Int. Med., vol. 103, n<sup>o</sup> 5, pp. 776-782, (1959).
22. BRUCE, R.A., YARNALL, S.R., "Computer-Aided Diagnosis of Cardiovascular Disorders", J. Chron. Dis., vol. 19, pp. 473-484, (1966).
23. BURBANK, F., "A Computer Diagnostic System for the Diagnosis of Prolonged Undifferentiating Liver Disease", Amer. J. Med., vol. 46, pp. 401-415, (1968).
24. BUSH, I.E., "Trouble with Medical Computers", Perspect. in Biol. and Med., pp. 600, (Summer, 1979).
25. CARDUS, D., "Quantitation in Biology and Medicine", J. Chron. Dis., vol. 19, pp. 319-324, (1966).
26. COCHRANE, A.L., GARLAND, L.H., "Observer Error in the Interpretation of Chest Films - Am. International Investigation", The Lancet, pp. 505-509, (Set. 1952).
27. COLLEN, M.F., "Periodic Health Examination Using an Automated Multitest Laboratory", J.A.M.A., vol. 195, pp. 830-833, (1966).
28. COLLEN, M.F., "Preventive Medicine and Automated Multiphasic Screening", 9<sup>th</sup> IBM Medical Symposium, pp. 81-97, (1969).
29. COLLEN, M.F., SIEGELAUB, A.B., CUTLER, J.L., EOLDBERG, R., "Aspects of Normal Values in Medicine", Ann. N.Y. Ac. Sci., vol. 161, n<sup>o</sup> 2, pp. 572-580, (1969).

30. COLLEN, M.F., "A Model for Evaluation Screening Systems", Evaluation of Efficacy of Medical Action, eds. Alperovitch, de Dombal, Grémy, North Holland, 1979.
31. COLLEN, M.F., "Multiphasic Screening Experience in the Kaiser - Permanent Program", Evaluation of Efficacy of Medical Action, eds. Alperovitch, de Dombal, Grémy, North Holland, 1979.
32. CONKLIN, J. et alii, "Computer-Aided Medical Decision Making in Radiotherapy", Radiology, vol. 123, n° 2, pp. 441-446, (1977).
33. CROFT, D.J., "Is Computerized Diagnosis Possible?", Comp. Biom. Res., vol. 5, pp. 351-367, (1972).
34. CROOKS, J., MURRAY, L.P.C., WAYNE, J., "Statistical Methods Applied to the Clinical Diagnosis of Thyrotoxicosis", Quart. J. Med., vol. 28, n° 110, pp. 211-233, (1959).
35. DAWID, A.P., "Properties of Diagnostic Data Distribution", Biometrics, vol. 32, pp. 647-658, (1976).
36. de DOMBAL, F.T., HARTLEY, J.R., SLEEMAN, D.H., "Teaching Surgical Diagnosis with the Aid of a Computer", Brit. J. Surg., vol. 56, pp. 754-757, (1969).
37. de DOMBAL, F.T., HORROCKS, J.C., STANILAND, J.R., GILL, P.W., "Simulation of Clinical Diagnosis: A Comparative Study" Brit. Med. J., vol. 2, pp. 575-577, (1971).
38. de DOMBAL, F.T., SMITH, R.B., MacGILL, V.K., "Simulation of the Diagnostic Process - A Further Comparison", Brit. J. Med. Educ., vol. 6, n° 3, pp. 238-245, (1972).
39. de DOMBAL, F.T., "Surgical Diagnosis Assisted by a Computer", Proc. R. Soc. Lond., B, vol. 184, pp. 433-440, (1973).



40. de DOMBAL, F.T. et alii, "Computer - Aided Diagnosis of Lower Gastrointestinal Tract Disorders", Gastroenterology, vol. 68, nº 2, pp. 252-260, (1975).
41. de DOMBAL, F.T., HORROCKS, J.C., "Use of Receiver Operating Characteristics (ROC) Curves to Evaluate Computer Threshold and Clinical Performance in the Diagnosis of Appendicitis", Methods of Information in Med., vol. 17, nº 3, pp. 157-161, (1978).
42. de DOMBAL, F.T., HALL, R., "The Evaluation of Medical Care from the Clinician's Point of View: What Should we Measure and Can we Thrust our Measurement", Evaluation of Efficacy of Medical Action, eds Alperovitch, de Dombal, Grémy, North Holland, 1979.
43. EDWARDS, W., et alii, "Probabilistic Information Processing Systems: Design and Evaluation", IEEE Trans. SSC., vol. 4, nº 3, pp. 248-265, (1968).
44. EDWARDS, W., "N=1: Diagnosis in Unique Cases", Jacquez, J.A. (ed). , Computer Diagnosis and Diagnostic Methods, C.C. Thomas, 1972.
45. ELFES, A., "Representação de Conhecimento e Descrição de Processo de Decisão em Diagnose Médica", Tese de Mestrado, CTA-ITA, (1980).
46. ELSTEIN, A.S. et alii, "Methods and Theory in the Study of Medical Inquiry", J. Med. Educ., vol. 47, nº 2, pp.85-92, (1972).
47. ENGLE, R.L., FLEHINGER, B.J., ALLEN, S., FRIEDMAN, R., LIOKIN, M., DAVIS, B.J., LEVERIDGE, L.L., "HEME - A Computer Aid to Diagnosis of Hematologic Disease", Bull. N.Y. Acad. Med., vol. 52, nº 5, pp. 584-600, (1976).

48. ENGLE, R.L., LIPKIN, M., FLEHINGER, B.J., "Computer Aided Differential Diagnosis of Hematologic Diseases", (ed) Jacquez, J.A. - Computer Diagnosis and Diagnostic Methods, CC-Thomas, 1972.
49. FEINSTEIN, A.R., "Taxonomy and Logic in Clinical Data", Ann N.Y. Ac. Sci., vol. 161, nº 2, pp. 450-459, (1969)
50. FEINSTEIN, A.R., "Compassion, Computers and the Regulation of Clinical Technology", Ann. Int. Med., vol. 66, pp. 789-805, (1967).
51. FEINSTEIN, A.R. et alii, "Estimating Prognosis with the Aid of a Conversational Mode Computer Program", Ann. Int. Med., vol. 76, pp. 911-921, (1972).
52. FEINSTEIN, A.R., "What Kind of Basic Science for Clinical Medicine", The N. Engl. J. Med., vol. 283, nº 16, pp. 847-852, (1970).
53. FINK, D.J., GALEN, R.S., "Probabilistic Approaches to Clinical Decision Support", Computer Aids to Clinical Decisions, Vol. II, B.T. Williams, CRC Press, Florida, EUA, (1982).
54. FITZGERALD, L.T., OVERALL, J.E., WILLIAMS, C.M., "A Computer Program for Diagnosis of Thyroid Disease", Amer. J. Roetgen., vol. 97, pp. 901-905, (1966).
55. FLEHINGER, B.J., ENGLE, R.L., "HEME - A Self-Improving Computer Program for Diagnosis - Oriented Analysis of Hematologic Diseases", IBM J. Res. Dev., pp. 557-64, (Nov. 1975) e IBM Research, RC 4550.
56. FLETCHER, C.M., "Observer Error", The Lancet, pp. 212, (Jan. 1954).

57. FORST, B.F., "Decision Analysis and Medical Mal-practice", Op. Res., vol. 22, nº 1, pp. 1-12, (1974).
58. FRASER, P.M., BARON, D.N., "Computer-assisted Classification and Diagnosis of Liver Disease", Proc. Roy Soc. Med., vol. 59, pp. 776-779 (1966).
59. FRIEDMAN, R.B. "Computers in Clinical Medicine - A Critical Review", Comp. Biom. Res., vol. 10, pp. 199-204, (1977).
60. FRIES, J.F., "Experience Counting in Sequential Computer Diagnosis", Arch. Int. Med., vol. 126, pp. 647-651, (1970).
61. FRIES, J.F., "Time Oriented Patient Records and a Computer Data Bank", JAMA, vol. 222, pp. 1536-1542, (1972).
62. FRITZ, K.J. et alii, "Computer Assisted Diagnosis for Ophthalmology", Comp. Biol. Med., vol. 8, pp. 223-228, (1978).
63. FRYBACK, D.G., "Bayes' Theorem and Conditional Non-Independence of Data in Medical Diagnosis", Comp. Biomed. Res., vol. 11, pp. 423-434, (1978).
64. GARLAND, H., "Studies on the Accuracy of Diagnostic Procedures", Amer. J. Roetgen., vol. 82, pp. 25-38, (1959).
65. GIAUQUE, W.C., PEEBLES, T.C., "Application of Multidimensional Utility Theory in Determining Optimal Test Treatment Strategies for the Streptococcal Sore Throat and Rheumatic Fever", Oper. Res., vol. 24, nº 5, pp. 933-950, (1976).
66. GILL, P.W. et alii, "Observer Variation in Clinical Diagnosis", Methods of Informat. in Med., vol. 12, nº 2, pp. 108-113, (1973).

67. GINSBERG, A.S., "The Diagnostic Process Viewed as a Decision Problem", (ed) Jacquez, J.A., Computer Diagnosis and Diagnostic Methods, CC. Thomas, 1972.
68. GLESER, M.A., COLLEN, M.F., "Towards Automated Medical Decisions", Comp. Biom. Res., vol. 5, pp. 180-189, (1972).
69. GOERTZEL, G., "Clinical Decision Support System", Ann. N.Y. Ac. Sci., vol. 161, nº 2, pp. 689-693, (1969)
70. GOLDWYN, R.M., et alii, "An Interactive System for Complex Data Analysis - Command Description", IBM Research, RC 3256, (Fev. 1971).
71. GOLDWYN, R.M., FRIEDMAN, H.P., SIEGEL, J.H., "Iteration and Interaction in Computer Data Bank Analysis: A Case Study in the Physiologic Classification Assessment of the Critically Ill", IBM Research, RC 3157, (Dez. 1970).
72. GORRY, G.A., BARNETT, G.O., "Sequential Diagnosis by Computer", J.A.M.A., vol. 205, nº 12, pp. 141-146, (1968).
73. GORRY, G.A., KASSIRER, J.P., ESSIG, A., SCHWARTZ, W.B., "Decision Analysis as the Basis for Computer" Aided Management of Acute Renal Failure, The Amer. J. Med., vol. 55, pp. 473-484, (1973).
74. GORRY, G.A., "The Personal Computer and the Clinical Practice" Arch. Int. Med., vol. 141, nº 13, pp 1745-1746, (1981).
75. GRÉMY, F., "The Future of Information Processing in Medicine and Public Health", Comp. Prog. in Biomedicine, vol. 11 pp. 71-80, (1980).
76. GRÉMY, F., GOLDBERG, M., "Decision Making Methods in Medicine", (eds) P.L. Reichertz, G.Goss, Informatics and Medicine, Springer Verlag, (1977).

77. GRUVER, R.H., FREIS, E.D., "A Study of Diagnostic Errors", Annals of Int. Med., vol. 47, pp. 108-120, (1957).
78. GUSTAFSON, D.H., et alii, "Probabilistic Information Processing: Implementation and Evaluation of a Semi PIP Diagnostic System", Comp. Biom. Res., vol. 6, pp. 355-370, (1973).
79. HART, P.E., NILSSON, N.J., "A Formal Basis for the Heuristic Determination of Minimum Cost Paths", IEEE Trans. Systems Science and Cybernetics, vol. SSC-4, n° 2, pp. 100-107, (1968).
80. HEALEY, J.R., "The Disciplining of Medical Data", Brit. Med. J., vol. 24, pp. 210-214, (1968).
81. HEALY, M.J.R., "Some Statistical Paradoxes of Normal Range," (eds) Alperovitch, de Dombal, Grémy, Evaluation of Efficacy of Medical Action, North Holland, 1979.
82. Health Screening - System and Implementation Guide, IBM, N° GE15-6014.
83. HILL, G.B., SPICER, C.C., "The Computer Surveillance of Congenital Mal-Formations", Brit. Med. Bull., vol. 24, pp. 215-218, (1968).
84. HOCKSTRA, D.J., "Sequential Games in Medical Diagnosis", Comp. Biom. Res., vol. 9, pp. 205-215, (1976).
85. HOFFER, E., "Computer Diagnosis: A Review and Discussion", IBM Cambridge Scient. C., 320-2011, (1967).
86. HORROCKS, J.C., de DOMBAL, F.T., "Computer-Aided Diagnosis of Dyspepsia", The Amer. J. Digestive Disease, vol. 20, n° 5, pp. 397-406, (1975).

87. HORVATH, W.J., "The Effect of Physician Bias in Medical Diagnosis", Behav. Sci., vol. 9, pp. 334-340, (1964).
88. HUGHES, W., et alii, "Myocardial Infarction Prognosis by Discriminant Analysis", Arch. Int. Med., vol. 111, pp. 120-127, (1963).
89. JACQUEZ, J.A., "Algorithm Diagnosis: A Review with Emphasis on Bayesian Methods", Computer Diagnosis and Diagnostic Methods, CC-Thomas, (1972).
90. KALMANSON, D., STEGALL, H.F., "Cardiovascular Investigations and Fuzzy Sets Theory", The Amer. J. Card., vol. 35, pp. 80-84, (1975).
91. KARAS, S., "Computers in the Emergency Department: Is there a Place for Computers in Clinical Medicine", ECHO's Coronado Conference, Out. (1980).
92. KASSIRER, J.P., "The Principles of Clinical Decision Making An Introduction to Decision Analysis", The Yale J. Biol. Med., vol. 49, pp. 149-164, (1976).
93. KASSIRER, J.P., GORRY, G.A., "Clinical Problem Solving: A Behavioral Analysis", Ann. Int. Med., vol. 89, pp. 245-255, (1978).
94. KEENEY, R.L., "A Survey of Multiattribute Utility", Evaluation of Efficacy of Medical Action, Ed. Alperovitch, A., de Dombal, F.T., Grémy, F., North Holland, 1979.
95. KLEINMUNTZ, B., McLEAN, R.S., "Diagnostic Interviewing by Digital Computer", Behav. Sci., vol. 13, pp. 75-80, (1968).
96. KLOETZEL, K., Raciocínio Clínico, EDART, S. Paulo Livraria Editora (1977).

97. KNAPP, R.G. et alii, "A Computer - Generated Diagnostic Decision Guide: A Comparison of Statistical and Clinical Diagnosis", Comp. Biol. Med., vol. 7, pp. 223-230, (1977).
98. KNILL-JONES, R.P. et alii, "Use of Sequential Bayesian Model in Diagnosis of Jaundice by Computer", Brit. Med. J., vol. 1, pp. 530-533, (1973).
99. KOMAROFF, A.L., "The Variability and Inaccuracy of Medical Data", Proc. of the IEEE, vol. 67, nº 9, pp. 1196-1206, (1979).
100. KOSS, N., FEINSTEIN, A.R., "Computer Aided Prognosis", Arch. Int. Med., vol. 127, pp. 448-459, (1971).
101. KRISCHER, J.P., "Utility Structure of a Medical Decision Making Problem", Oper. Res., vol. 24, nº 5, pp. 951-972, (1975).
102. KULIKOWSKY, C.A., "Pattern Recognition Approach to Medical Diagnosis", IEEE Trans SSC, vol. 6, nº 3, pp. 173-178, (1970).
103. LAURENTI, R., Situação Sanitária do Brasil - Administração de Saúde no Brasil, Livraria Pioneira Editora, 1982.
104. LEAPER, D.J. et alii, "Computer - Assisted Diagnosis of Abdominal Pain Using 'Estimates' Provided by Clinicians", Brit. Med. J., vol. 4, pp. 350-354, (1972).
105. LEAPER, D.J. et alii, "Clinical Diagnostic Process: An Analysis", Brit. Med. J., pp. 559-574, (Set. 1973).
106. LEDLEY, R.S., LUSTED, L.B., "Reasoning Foundations of Medical Diagnosis", Science, vol. 130, pp. 9-21, (1959).
107. LEDLEY, R.S., "Computer Aids to Clinical Treatment Evaluation", Operations Res., vol. 15, nº 4, pp. 694-705, (1967).

108. LEDLEY, R.S. "Syntax Directed Concept Analysis in the Reasoning Foundation of Medical Diagnosis", ed. Jacquez, J.A., Computer Diagnosis and Diagnostic Methods, CC Thomas, 1972.
109. LODWICK, G.S. et alii, "Computer Diagnosis of Primary Bone Tumours", Radiology, vol. 80, nº 2, pp. 273-275, (1973).
110. LUSTED, L.B., "Computers in Medicine - A Personal Perspective", J. Chron. Diseases, vol. 19, pp. 365-372, (1966).
111. LUSTED, L.B., "Observer Error, Signal Detectability and Medical Decision Making", ed. Jacquez, J.A. - Computer Diagnosis and Diagnostic Methods, - CC Thomas, (1972).
112. LUSTED, L.B., "Recent Developments in Diagnostic Radiology Efficacy Studies", eds. Alperovitch, de Dombal, Grémy, Evaluation of Efficacy of Medical Action, North Holland, 1979.
113. MACHADO, R.J., "Um Método Prático para Computarizar a Interpretação de Exames Médicos", XII Congresso da SUCESU, (1979).
114. MAINLAND, D., "Normal Values in Medicine", Ann. N.Y. Ac.Sci., vol. 161, nº 2, pp. 527-537, (1969).
115. MELHORN, J.M. et alii, "Current Attitudes of Medical Personnel Toward Computers", Comp. Biom. Res., vol. 12, pp. 327-334, (1979).
116. MOEHR, J.R., "Computer Assisted Medical History", Eds. P.L. Reichertz, G. Goss, Informatics and Medicine, Springer Verlag, 1977.
117. MOISEEVA, N.I., USOV, V.V., "Some Medical and Mathematical Aspects of Computer Diagnosis", Proc. IEEE, vol. 57, nº 11, pp. 1919-1925, (1969).



118. MORGAN, A.G. et alii, "Clinical Findings Early Endoscopy and Multivariate Analysis in Patients Bleeding from the Upper Gastrointestinal Tract", Brit. Med. J., vol. 2, pp. 237-240, (1977).
119. MURPHY, E.A., ABBEY, H., "The Normal Range - A Common Misuse", J. Chron. Dis., vol. 20, pp. 79-88, (1967).
120. MURPHY, J.F., "Sore Throat Management - Decision Analysis Using Pleasant Hour Equivalents", Comp. Biom. Res., vol. 12, pp. 203-219, (1979).
121. McNEIL, B., SHERMAN, H., "Example: Bayesian Calculations for the Determination of the Etiology of Pleuritic Chest Pain in Young Adults in a Teaching Hospital", Comp. Biom. Res. vol. 11, pp. 187-194, (1978).
122. McSHANE, D.J., et alii, "TOD: A Software System for the ARAMIS Data Bank", Computer, pp. 34-39, (nov. 1979).
123. NAIMARK, A., CHERNIAK, R.M., PROTT, D., "Comprehensive Respiratory Information System for Clinical Investigation of Respiratory Disease", Amer. Rev. Resp. Disease, vol. 103, pp. 229-239, (1971).
124. NEURATH, P.W. et alii, "Design of a Computer System to Assist in Differential Pre-operative Diagnosis for Pelvic Surgery", The N. Engl. J. Med., vol. 280, n° 14, pp. 745-749, (1969).
125. NILSSON, N.J., Principles of Artificial Intelligence, Springer Verlag, 1982.
126. NILSSON, N.J., Problem - Solving Methods in Artificial Intelligence, Mc Graw Hill, 1971.
127. NILSSON, N.J., "Survey of Pattern Recognition", Ann. N.Y. Ac. Sci., vol. 161, n° 2, pp. 380-401, (1969).

128. NOMURA, Y., "An Experimental Approach to Medical Decision Problems", Comp. Biom. Res., vol. 14, pp 1-18, (1981).
129. NORUSIS, M.J., JACQUEZ, J.A., "Diagnosis I. Symptom Non-Independence in Mathematical Models for Diagnosis", Comp. Biom. Res., vol. 18, pp. 156-172, (1975).
130. NORUSIS, M.J., JACQUEZ, J.A., "Diagnosis II - Diagnostic Model Base on Attribute Clusters: A Proposal and Comparisons", Comp. Biom. Res., vol. 8, pp. 173-188, (1975).
131. OVERALL, J.E., et alii, "Conditional Probability Program for Diagnosis of Thyroid Function", J.A.M.A., vol. 183, nº 5, pp. 307-313, (1963).
132. PATRICK, E.A., Decision Analysis in Medicine, CRC Press, (1979).
133. PATRICK, E.A., "Expected Outcome Loss to Evaluate Medical Diagnosis and Treatments", Comp. Biol. Med., vol. 7, pp. 1-8, (1977).
134. PATRICK, E.A., STELMACK, F.P., "Review of Pattern Recognition in Medical Diagnosis and Consulting Relative to a New System Model", IEEE Trans. SMC, vol. 4, nº 1, pp. 1-16, (1974).
135. PAUKER, S.G., KASSIRER, J.P., "Therapeutic Decision Making - A Cost Benefit Analysis", The N. Engl. J. Med., vol. 293, pp. 229-234, (1975).
136. PAUKER, S.G., GORRY, G.A., KASSIRER, J.P., SCWARTZ, W.B., "Towards the Simulation of Clinical Cognition - Taking a Present Illness by Computer", The Amer. J. Med., vol. 60, pp. 981-996, (1976).

137. PAUKER, S.G., "Coronary Artery Surgery: The Use of Decision Analysis", Annals of Int. Med., vol. 85, pp. 8-18,(1976).
138. PAYNE, L.C., "The Role of the Computer in Refining Diagnosis", The Lancet, pp. 32-35, (Jul. 1964).
139. PETERS, R.J., "Zero Order and Non Zero Order Decision Rules in Medical Diagnosis", IBM J. Res. and Development, vol. 21, n<sup>o</sup> 5, pp. 449-640, (1977).
140. PLISKINS, J.S., BECK, C.H., "A Health Index for Patients Selection: A Value Function Approach with Application to Chronic Renal Failure Patients", Manag. Sci., vol. 22, n<sup>o</sup> 9, pp. 1009-1021, (1976).
141. PREWITT, J.M.S., "Experiments with Statistical and Quasi - Statistical Methods in Diagnosis", ed Jacquez, J.A., Computer Diagnosis and Diagnostic Methods, CC. Thomas, (1972).
142. PRYOR, T.A. et alii, "Help - A Computer System for Medical Decision Making", Computer, vol. 8, n<sup>o</sup> 1, pp. 34-38, (1975).
143. RANSOHOFF, D.F., FEINSTEIN, A.R., "Is Decision Analysis Useful in Clinical Medicine", The Yale J. Biol. Med., vol. 49, pp. 165-168, (1976).
144. RECTOR, A.L., ACKERMAN, E., "Rules for Sequential Diagnosis", Comp. Biom. Res., vol. 8 pp. 146-155, (1975).
145. REGGIO, M., "Differential Diagnosis by Weighting Characteristic Signs of Two Diseases", Comp. Biom. Res. vol. 14, pp. 131-137, (1981).
146. ROSATI, R.A., WALLACE, A.G., STEAD, E.A., "The Way of the Future", Arch. Int. Med., vol. 131, pp. 285-287, (1973).

147. ROSATI, R.A. et alii, "A New Information System for Medical Practice", Arch Int. Med., vol. 135, pp. 1017-1024, (1975).
148. RUSSEK, E., "The Effect of Assuming Independence in Applying Bayes' Theorem to Risk Estimation and Classification in Diagnosis", Comp. Biom. Res., vol. 16, pp. 537-551, (1983).
149. SAFRAN, C. et alii, "Decision Analysis to Evaluate Lymphoangiography in the Management of Patients with Hodgkin's Disease", The N. Engl. J. Med., vol. 296, nº 19, pp. 1088-1092, (1977).
150. SCADDING, J.G., "Diagnosis: The Clinician and the Computer", The Lancet, pp. 877-82, (Oct. 1967).
151. SIEGEL, J.H., GOLDWYN, R.M., FRIEDMAN, H.P., "Pattern and Process in the Evaluation of Human Septic Shock", IBM Research, RC 3245, (1971).
152. SIFFERT, G., MACHADO, R.J., "A Computer Program for the Study of the Galltract Dynamics by Means of the Cholecystometric Method", VI Congr. Mundial Gastroenterologia, (1978).
153. SCHILD, W., GAVISH, B., LUNENFELD, B., "Computer Assisted Medical Diagnosis with an Application to Endocrinology", IBM Israel Scientific Center Technical Report 014, (Abr. 1974).
154. SCHOLMAN, H.M., BERNSTEIN, L.M., "Computer Use in Diagnosis Prognosis and Therapy", Science, vol. 200, pp. 926-931, (1978).
155. SCHWARTZ, W.B., "Medicine and the Computer - The Promise and the Problem of Change", The N. Engl. J. Med., vol. 283, nº 23, pp. 1257-1264, (1970).

156. SHEA, G., "An Analysis of the Bayes Procedure for Diagnosing Multistage Diseases", Comp. Biom. Res., vol. 11, pp. 65-75, (1978).
157. SHERMAN, H., "A Pocket Diagnostic Calculator Program for Computing Bayesian Probabilities for Nine Disease with Sixteen Symptoms", Comp. Biom. Res., vol. 11, pp. 177-186, (1978).
158. SHORTLIFE, E.H. et alii, "An Artificial Intelligence Program to Advise Physicians Regarding Antimicrobial Therapy", Comp. Biom. Res., vol. 6, pp. 544-560, (1973).
159. SHORTLIFE, E.H. et alii, "Computer - Based Consultations in Clinical Therapeutics: Explanation and Rule Acquisition Capabilities of the MYCIN System", Comp. Biom. Res., vol. 8, pp. 303-320, (1975).
160. SHORTLIFE, E.H., BUCHANAN, B.G., "A Model of Inexact Reasoning in Medicine", Math. Biosciences, vol. 23, pp. 351-379, (1975).
161. SHORTLIFE, E.H., BUCHANAN, B.G., FEIGENBAUM, E.A., "Knowledge Engineering for Medical Decision Making: A Review of Computer-based Clinical Decision Aids", Proc. IEEE, vol. 67 n<sup>o</sup> 9, pp. 1207-1224, (1979)
162. SLAMECKA, V., CAMP, H.N., BADRE, A.N., "MARIS: A Knowledge System for Internal Medicine", Info. Proc. & Manag., vol. 13, pp. 273-276, (1977).
163. SOKOLON, M., Clinical Cardiology, (1979).
164. SPIEGELHALTER, D.J., SMITH, F.J., "Decision Analysis and Clinical Decision", Perspective in Medical Statistics, pp. 103-131, Academic Press.
165. SPRAGUE, R.H., CARLSON, E.D., Building Effective Decision

Support Systems, Prentice Hall, 1982.

166. STARMER, C.F., ROSATI, R.A., McNEER, J.F., "A Comparison of Frequency Distribution for Use in a Model for Selecting Treatment in Coronary Artery Disease", Comp. Biom. Res., vol. 7, n<sup>o</sup> 3, pp. 278-293, (1974).
167. STARMER, C.F., LEE, K.L., "A Mathematical Approach to Medical Decision Application of Bayes' Rule to a Mixture of Continuous and Discrete Clinical Variables", Comp. Biom. Res., vol. 9, pp. 531-541, (1976).
168. STARTSMAN, T.S., ROBINSON, R.E., "The Attitudes of Medical and Paramedical Personnel Toward Computer", Comp. Biom. Res., vol. 5, pp. 218-227, (1972).
169. STERLING, T.D., NICKSON, J., POLLACK, S.V., "Is Medical Diagnosis a General Computer Problem?", J.A.M.A., vol. 198, n<sup>o</sup> 3, pp. 191-196, (1966).
170. SZOLOVITS, P., PAUKER, S.G., "Categorical and Probabilistic Reasoning in Medical Diagnosis", Artificial Intelligence, vol. 11, pp. 115-144, (1978).
171. SZOLOVITS, P., PAUKER, S.G., "Computers and Clinical Decision Making. Whether, How and Whom?", Proc. IEEE, vol. 67, n<sup>o</sup> 9, pp. 1224-1226, (1979).
172. SWETS, J.A., "Signal Detection in Medical Diagnosis", Comp. Diagnosis and Diagnostic Methods, ed. Jacquez, J. A., Charles C. Thomas, (1972).
173. TAYLOR, T.R., AITCHISON, J., MCGIRR, E.M., "Doctor as Decision-Makers: A Computer - Assisted Study of Diagnosis As a Cognitive Skill", Brit. Med. J., vol. 3, pp.35-40, (1971).

174. TAYLOR, T.R., "Study of Cost - Conscious Computer - Assisted Diagnosis in Thyroid Diseases", The Lancet, pp. 79-83, (Jul. 1972).
175. TEACH, R.L., SHORTLIFE, E.H., "An Analysis of Physician Attitudes Regarding Computer Based Clinical Consultation Systems", Comp. Biom. Res., vol. 14, nº 6, pp. 542-558 (1981).
176. TEATHER, D., "Statistical Techniques for Diagnosis", J.R. Statist., vol. 137, pp. 231-244, (1974).
177. TOMPKINS, R.K., BURNES, D.C., CABLE, W.E., "An Analysis of Cost-Effectiveness of Pharyngitis Management and Acute Rheumatic Fever Prevention", Ann. Int. Med., vol. 86, pp. 481-492, (1977).
178. TORRANCE, G.W., "Health Status Index Models: A Unified Mathematical View", Management Science, vol. 22, nº 9, pp. 990-1001, (1976).
179. WALTON, P.L., HOLLAND, R.L., wolf, L.I., "Medical Guidance and PROMIS", Computer, (Nov. 1979).
180. WARNER, H.R., et alii, "A Sequential Bayesian Approach to History Taking and Diagnosis", Comp. Biom. Res., vol. 5, pp. 256-262, (1972).
181. WARNER, H.R., TORONTO, A.F., VEASY, L.G., "Experience with Bayes' Theorem for Computer Diagnosis of Congenital Heart Disease", Ann. N.Y. Ac. Sci., vol. 115, nº 2, pp. 558-567, (1964).
182. WARNER, H.R., OLMSTED, C.M., RUTHERFORD, B.D., "HELP - A Program for Medical Decision Making", Comp. Biom. Res., vol. 5, pp. 65-74, (1972).

183. WARTAK, J., "A Practical Approach to Automated Diagnosis"  
IEEE Trans. BME, vol. 17, n<sup>o</sup> 1, pp. 37-43, (1970).
184. WEED, L.W., "Medical Records that Guide and Teach", 9<sup>th</sup>  
IBM Medical Symposium, 1968.
185. WEISS, S.M., KULIKOWSKI, C.A., AMAREL, S., "A Model Based  
Method for Computer Aided Medical Decision Making",  
Artificial Intelligence, vol. 11, pp. 145-172, (1978).
186. WIENER, F., "Computer Simulation of the Diagnostic Process  
in Medicine", Comp. Biom. Res., vol. 8, pp. 129-142,  
(1975).
187. WILSON, W.J. et alii, "The Computer Analysis and Diagnosis  
of Gastric Ulcer", Radiology, vol. 85, n<sup>o</sup> 6, pp. 1064-  
1073, (1965).
188. WILSON, D.H. et alii, "Diagnosis of Acute Abdominal Pain in  
the Accident and Emergency Department", Brit. J. Surg.,  
vol. 64, pp. 250-254, (1977).
189. WORTMAN, P.M., "Medical Diagnosis: An Information Processing  
Approach", Comp. Biom. Res., vol. 5, pp. 315-328, (1972).