

## 5 Experimentos com Dados Reais

### 5.1 Descrição das Bases de Dados Utilizadas

As bases de dados foram extraídas do Repositório de Dados *UCI Machine Learning* (<http://archive.ics.uci.edu/ml/datasets.html>).

Visto que o uso das bases de dados abordadas nesta dissertação tem por objetivo a mera comparação entre os resultados, tocante à ordem de seleção pelo algoritmo de seleção MIFS-U, com base nos dois métodos de estimação da entropia e informação mútua apresentados, não está no escopo desta dissertação uma análise específica das bases de dados. Por essa razão, as bases de dados estão sucintamente apresentadas em tabelas com as informações relevantes.

Nas tabelas subseqüentes, pode-se observar as seguintes informações:

- a quantidade de amostras completas consideradas ( $n$ ),
- a quantidade de amostras por resultado do desfecho,
- a identificação numérica da variável,
- o tipo da variável (discreta ou contínua), e
- a identificação da variável, em que foram mantidos os termos originais, sem prejuízo à dissertação. Com a única exceção da base de dados “Câncer de Mama”, em que a identificação da totalidade das variáveis não se encontrava disponível.
- a descrição (tradução) da variável, quando possível.

### 5.1.1 Base de Dados Ecocardiograma

A base de dados Ecocardiograma refere-se a pacientes que sofreram ataques do coração em um determinado instante do passado e submeteram-se a um ecocardiograma. O desfecho diz respeito à sobrevivência do paciente ser superior a 1 ano (*alive-at-one*) ou não. Visto que os pacientes sofreram ataques cardíacos em diferentes épocas, é possível que alguns tenham sobrevivido menos de 1 ano, mas ainda estejam vivos.

Tabela 5.1 – Base de Dados ECOCARDIOGRAMA

Base de Dados		<b>ECOCARDIOGRAMA</b> <i>Echocardiogram</i>		$n = 61$
Desfecho	<i>alive-at-1</i>	17	<i>not alive-at-1</i>	44
Variável	Tipo	Identificação	Descrição	
1	Discreta	<i>still-alive</i>	paciente ainda vivo	
2	Discreta	<i>pericardial-effusion</i>	derrame pericárdio	
3	Discreta	<i>group</i>	-----	
4	Contínua	<i>survival</i>	tempo de sobrevivência do paciente	
5	Contínua	<i>age-at-heart-attack</i>	idade quando ocorreu o ataque cardíaco	
6	Contínua	<i>fractional-shortening</i>	fração de encurtamento (medida de contratilidade)	
7	Contínua	<i>epss</i>	medida de contratilidade	
8	Contínua	<i>lvdd</i>	diâmetro diastólico final do ventrículo esquerdo	
9	Contínua	<i>wall-motion-score</i>	medida de contratilidade segmentar do ventrículo esquerdo	
10	Contínua	<i>wall-motion-index</i>	índice de contratilidade segmentar (igual à variável anterior dividida pelo número de segmentos observados)	
11	Contínua	<i>mult</i>	variável derivada, que pode ser ignorada	

### 5.1.2 Base de Dados Telescópio

Os telescópios de luz Cherenkov são os mais eficientes instrumentos de medida em astronomia gama que muito têm colaborado para a compreensão dos processos de geração de raios cósmicos em núcleos ativos de galáxias e supernovas. A base de dados Telescópio refere-se a elementos caracterizadores na distinção de gamas primários (gamma) dos chuviros hadrônicos (hadron), utilizando a Técnica de Imagem (Imaging Atmospheric Cerenkov Technique), que se baseia na forma do sinal na câmara de fotomultiplicadoras.

Tabela 5.2 – Base de Dados TELESCÓPIO

Base de Dados		<b>TELESCÓPIO</b>		$n = 554$
		<i>MAGIC gamma telescope</i>		
Desfecho	<i>gamma</i>		361	<i>hadron</i>
				193
Variável	Tipo	Identificação	Descrição	
1	Contínua	$fLength$	tamanho do eixo principal da elipse	
2	Contínua	$fWidth$	tamanho do eixo secundário da elipse	
3	Contínua	$fSize$	10-log do total de fótons de todos os <i>pixels</i>	
4	Contínua	$fConc$	razão da soma dos dois <i>pixels</i> mais intensos com relação a $fSize$	
5	Contínua	$fConcl$	razão do <i>pixel</i> mais intenso com relação a $fSize$ .	
6	Contínua	$fAsym$	distância do <i>pixel</i> mais intenso ao centro, projetada sobre o eixo principal	
7	Contínua	$fM3Long$	raiz cúbica do terceiro momento com relação ao eixo principal	
8	Contínua	$fM3Trans$	raiz cúbica do terceiro momento com relação ao eixo secundário	
9	Contínua	$fAlpha$	ângulo do eixo principal com relação à origem	
10	Contínua	$fDist$	distância da origem ao centro da elipse	

### 5.1.3 Base de Dados Vinho

A base de dados Vinho refere-se à análise química de vinhos produzidos na mesma região da Itália, mas advindos de duas diferentes viniculturas. A análise determina as quantidades de 13 componentes encontrados em cada um dos dois tipos de vinho: classe 1 e classe 2.

Tabela 5.3 – Base de Dados VINHO

Base de Dados		<b>VINHO</b> <i>Wine recognition</i>		$n = 130$
Desfecho	Classe 1		59	Classe 2
Variável	Tipo	Identificação		Descrição
1	Contínua	<i>alcohol</i>		teor alcoólico
2	Contínua	<i>malic acid</i>		ácido málico
3	Contínua	<i>ash</i>		cinzas
4	Contínua	<i>alkalinity of ash</i>		alcalinidade das cinzas
5	Contínua	<i>magnesium</i>		magnésio
6	Contínua	<i>total phenols</i>		fenóis totais
7	Contínua	<i>flavanoids</i>		flavonóides
8	Contínua	<i>nonflavanoid phenols</i>		fenóis não-flavonóides
9	Contínua	<i>proanthocyanins</i>		antocianinas
10	Contínua	<i>color intensity</i>		intensidade da cor
11	Contínua	<i>hue</i>		coloração (tonalidade)
12	Contínua	<i>OD280/OD315 of diluted wines</i>		taxa de diluição
13	Contínua	<i>proline</i>		prolina

### 5.1.4 Base de Dados Dermatologia

A base de dados Dermatologia refere-se a 26 atributos clínicos considerados no diagnóstico de afecções dermatológicas: em especial, psoríase e dermatite seborréica.

Tabela 5.4 – Base de Dados DERMATOLOGIA

Base de Dados		<b>DERMATOLOGIA</b>		<i>n</i> = 171
		<i>Dermatology</i>		
Desfecho		<i>psoriasis</i>	111	<i>seboreic dermatitis</i>
				60
Variável	Tipo	Identificação		Descrição
1	Discreta	<i>erythema</i>		eritema
2	Discreta	<i>scaling</i>		escamação
3	Discreta	<i>definite borders</i>		contornos definidos
4	Discreta	<i>itching</i>		coceira
5	Discreta	<i>Koebner phenomenon</i>		fenômeno de Koebner
6	Discreta	<i>follicular papules</i>		pápulas foliculares
7	Discreta	<i>knee and elbow involvement</i>		envolvimento de joelho e cotovelo
8	Discreta	<i>scalp involvement</i>		envolvimento do couro cabeludo
9	Discreta	<i>family history</i>		histórico familiar
10	Discreta	<i>eosinophils in the infiltrate</i>		eosinófilos no infiltrado
11	Discreta	<i>PNL infiltrate</i>		infiltrado PNL
12	Discreta	<i>exocytosis</i>		exocitose
13	Discreta	<i>acanthosis</i>		acantose
14	Discreta	<i>hyperkeratosis</i>		hiperqueratose
15	Discreta	<i>parakeratosis</i>		paraqueratose
16	Discreta	<i>clubbing of the rete ridges</i>		Baqueteamento dos cones epiteliais
17	Discreta	<i>elongation of the rete ridges</i>		alongamento dos cones epiteliais
18	Discreta	<i>thinning of the suprapapillary epidermis</i>		estreitamento da epiderme suprapapilar
19	Discreta	<i>spongiform pustule</i>		pústula espongiforme
20	Discreta	<i>focal hypergranulosis</i>		hipergranulose focal
21	Discreta	<i>disappearance of the granular layer</i>		desaparecimento da camada granular
22	Discreta	<i>vacuolisation and damage of basal layer</i>		vacuolização e danos da camada basal
23	Discreta	<i>spongiosis</i>		espongiose
24	Discreta	<i>follicular horn plug</i>		tamponamento folicular
25	Discreta	<i>perifollicular parakeratosis</i>		paraqueratose perifolicular
26	Discreta	<i>inflammatory mononuclear infiltrate</i>		infiltrado mononuclear inflamatório
27	Discreta	<i>band-like infiltrate</i>		infiltrado em faixa
28	Contínua	<i>Age</i>		idade

### 5.1.5 Base de Dados Câncer de Mama

A base de dados Câncer de Mama refere-se a características de uma imagem digitalizada de um tumor do seio, que descrevem características dos núcleos celulares exibidos na imagem, podendo o tumor ser benigno ou maligno.

Tabela 5.5 – Base de Dados CÂNCER DE MAMA

Base de Dados		<b>CÂNCER DE MAMA</b> <i>Wisconsin Diagnostic Breast Cancer</i>		$n = 569$
Desfecho	<i>benign</i>	357	<i>malignant</i>	212
Variável	Tipo	Identificação		
1	Contínua	var1		
2	Contínua	var2		
3	Contínua	var3		
4	Contínua	var4		
5	Contínua	var5		
6	Contínua	var6		
7	Contínua	var7		
8	Contínua	var8		
9	Contínua	var9		
10	Contínua	var10		
11	Contínua	var11		
12	Contínua	var12		
13	Contínua	var13		
14	Contínua	var14		
15	Contínua	var15		
16	Contínua	var16		
17	Contínua	var17		
18	Contínua	var18		
19	Contínua	var19		
20	Contínua	var20		
21	Contínua	var21		
22	Contínua	var22		
23	Contínua	var23		
24	Contínua	var24		
25	Contínua	var25		
26	Contínua	var26		
27	Contínua	var27		
28	Contínua	var28		
29	Contínua	var29		
30	Contínua	var30		

### 5.1.6 Base de Dados Doenças Cardíacas

A base de dados Doenças Cardíacas refere-se a treze atributos clínicos considerados no diagnóstico: negativo (classe 1) e positivo (classe 2).

Tabela 5.6 – Base de Dados DOENÇAS CARDÍACAS

Base de Dados		<b>DOENÇAS CARDÍACAS</b> <i>Heart Disease</i>		<i>n</i> = 214
Desfecho	Classe 1		160	Classe 2
Variável	Tipo	Identificação	Descrição	
1	Discreta	<i>sex</i>	sexo	
2	Discreta	<i>cp</i>	tipo de dor no peito	
3	Discreta	<i>fbs</i>	glicose em jejum	
4	Discreta	<i>restecg</i>	resultado do eletrocardiograma em repouso	
5	Discreta	<i>exang</i>	angina induzida por exercício	
6	Discreta	<i>slope</i>	inclinação da extremidade do segmento ST no exercício	
7	Discreta	<i>ca</i>	número de vasos coloridos pela fluoroscopia	
8	Discreta	<i>thal</i>	talassemias	
9	Contínua	<i>age</i>	idade	
10	Contínua	<i>trestbps</i>	pressão sanguínea em repouso	
11	Contínua	<i>chol</i>	Soro colesterol	
12	Contínua	<i>thalach</i>	máxima taxa de batimento cardíaco atingida	
13	Contínua	<i>oldpeak</i>	depressão ST induzida por exercício relativamente sossegado	

## 5.2 Comparação dos Métodos

A comparação dos resultados da seleção pelo MIFS-U em relação aos dois métodos de estimação da entropia e informação mútua abordados nesta dissertação encontram-se nas tabelas a seguir. A análise tem seu foco nas cinco primeiras variáveis selecionadas, e, por simplificação, designar-se-á o Método Shannon / Histograma pela sigla SH, e o Método Cauchy-Schwartz / Parzen-Rosenblatt pela sigla CSPR. Vale ressaltar que os comentários são meramente especulativos, oriundos da simples observação. Para uma análise mais detalhada, seria necessária a aplicação de um classificador a fim de investigar a eficiência do grupo de variáveis selecionadas

## 5.2.1

## Resultado Comparativo da Seleção pelo MIFS-U – Ecocardiograma

Tabela 5.7 – Resultado Comparativo da Seleção pelo MIFS-U – ECOCARDIOGRAMA

Base de Dados			ECOCARDIOGRAMA		
Resultado da Seleção pelo MIFS-U					
Método Shannon / Histograma			Método Cauchy-Schwartz / Parzen-Rosenblatt		
Ordem de Seleção	Variável	IM com Desfecho	Ordem de Seleção	Variável	IM com Desfecho
1 <sup>a</sup>	4	0.8241	1 <sup>a</sup>	4	0.2463
2 <sup>a</sup>	1	0.6118	2 <sup>a</sup>	1	0.2144
3 <sup>a</sup>	2	0.0227	3 <sup>a</sup>	10	0.0523
4 <sup>a</sup>	3	0.0213	4 <sup>a</sup>	5	0.0326
5 <sup>a</sup>	10	0.2442	5 <sup>a</sup>	9	0.0307
6 <sup>a</sup>	9	0.3303	6 <sup>a</sup>	11	0.0171
7 <sup>a</sup>	11	0.2104	7 <sup>a</sup>	8	0.0285
8 <sup>a</sup>	5	0.2511	8 <sup>a</sup>	2	0.0063
9 <sup>a</sup>	6	0.1577	9 <sup>a</sup>	3	0.0128
10 <sup>a</sup>	7	0.1732	10 <sup>a</sup>	6	0.0174
11 <sup>a</sup>	8	0.2014	11 <sup>a</sup>	7	0.0125

A seleção feita pelo MIFS-U, através dos dois métodos, leva a dois conjuntos de variáveis semelhantes. Três entre as cinco primeiras selecionadas são exatamente as mesmas. As variáveis 2 e 3 apresentaram certa redundância, quase indistintamente, em relação às variáveis 5 e 9, refletida, especialmente, no SH, como pode ser observado nos resultados listados a seguir:

- Informação mútua entre as variáveis 2 e 5: 0,475(SH) e 0,007(CSPR) ;
- Informação mútua entre as variáveis 2 e 9: 0,446(SH) e 0,013(CSPR) ;
- Informação mútua entre as variáveis 3 e 5: 0,529(SH) e 0,018(CSPR) ;
- Informação mútua entre as variáveis 3 e 9: 0,529(SH) e 0,018(CSPR).

Em termos práticos, significaria que, em princípio, a permutação entre elas no grupo selecionado pouca influência teria sobre o resultado.

Note-se que as duas variáveis que trazem mais informação em relação ao desfecho são a 4 e a 1, que dizem respeito, respectivamente, ao tempo de sobrevivência do paciente após o ataque cardíaco e à informação de que o paciente ainda se encontrava vivo ou não. A variável 11, embora descrita como uma variável derivada que pudesse ser ignorada, ela figura, em ambos os métodos, como uma variável que ainda apresenta alguma informação adicional.



## 5.2.2 Resultado Comparativo da Seleção pelo MIFS-U – Telescópio

Tabela 5.8 – Resultado Comparativo da Seleção pelo MIFS-U – TELESCÓPIO

Base de Dados		TELESCÓPIO			
Resultado da Seleção pelo MIFS-U					
Método Shannon / Histograma			Método Cauchy-Schwartz / Parzen-Rosenblatt		
Ordem de Seleção	Variável	IM com Desfecho	Ordem de Seleção	Variável	IM com Desfecho
1 <sup>a</sup>	1	0.4046	1 <sup>a</sup>	9	0.0923
2 <sup>a</sup>	9	0.4046	2 <sup>a</sup>	2	0.0173
3 <sup>a</sup>	8	0.4031	3 <sup>a</sup>	1	0.0133
4 <sup>a</sup>	10	0.4046	4 <sup>a</sup>	7	0.0095
5 <sup>a</sup>	6	0.4031	5 <sup>a</sup>	3	0.0061
6 <sup>a</sup>	5	0.4046	6 <sup>a</sup>	8	0.0072
7 <sup>a</sup>	2	0.4031	7 <sup>a</sup>	6	0.0055
8 <sup>a</sup>	7	0.4046	8 <sup>a</sup>	10	0.0051
9 <sup>a</sup>	3	0.4031	9 <sup>a</sup>	5	0.0038
10 <sup>a</sup>	4	0.4031	10 <sup>a</sup>	4	0.0049

A seleção feita pelo MIFS-U, através dos dois métodos, é, também aqui, praticamente a mesma. As variáveis 2 e 8, como também as variáveis 6 e 7, apresentam relativa redundância entre si, em especial no SH. Quanto às variáveis 2 e 8, a informação mútua entre elas foi 0,776(SH) e 0,058(CSPR), e quanto às variáveis 6 e 7, 0,684(SH) e 0,093(CSPR). Destaca-se que no SH, a informação mútua com o desfecho é praticamente a mesma, o que já não ocorre com o CSPR, refletindo um maior poder discriminativo.

Note-se que as variáveis 1 e 9 destacam-se como principais, em ambos os métodos, de forma alternada. A primeira diz respeito ao tamanho do eixo principal da elipse, e a segunda, ao ângulo de tal eixo em relação à origem.

As redundâncias, destacadas acima, referem-se às seguintes variáveis:

- Tamanho do eixo secundário da elipse (var 2) e raiz cúbica do terceiro momento com relação ao eixo secundário (var 8).
- Distância do *pixel* mais intenso ao centro, projetada sobre o eixo principal (var 6) e raiz cúbica do terceiro momento com relação ao eixo principal (var 7).

### 5.2.3 Resultado Comparativo da Seleção pelo MIFS-U – Vinho

Tabela 5.9 – Resultado Comparativo da Seleção pelo MIFS-U – VINHO

Base de Dados		VINHO			
Resultado da Seleção pelo MIFS-U					
Método Shannon / Histograma			Método Cauchy-Schwartz / Parzen-Rosenblatt		
Ordem de Seleção	Variável	IM com Desfecho	Ordem de Seleção	Variável	IM com Desfecho
1 <sup>a</sup>	1	0.520	1 <sup>a</sup>	13	0.362
2 <sup>a</sup>	2	0.517	2 <sup>a</sup>	1	0.360
3 <sup>a</sup>	5	0.517	3 <sup>a</sup>	10	0.278
4 <sup>a</sup>	3	0.517	4 <sup>a</sup>	2	0.109
5 <sup>a</sup>	11	0.517	5 <sup>a</sup>	5	0.157
6 <sup>a</sup>	7	0.517	6 <sup>a</sup>	12	0.048
7 <sup>a</sup>	9	0.517	7 <sup>a</sup>	11	0.041
8 <sup>a</sup>	4	0.517	8 <sup>a</sup>	8	0.068
9 <sup>a</sup>	8	0.517	9 <sup>a</sup>	3	0.057
10 <sup>a</sup>	12	0.520	10 <sup>a</sup>	9	0.039
11 <sup>a</sup>	6	0.520	11 <sup>a</sup>	4	0.098
12 <sup>a</sup>	10	0.520	12 <sup>a</sup>	6	0.164
13 <sup>a</sup>	13	0.520	13 <sup>a</sup>	7	0.224

Novamente, a seleção feita pelo MIFS-U, através dos dois métodos, é praticamente a mesma. As variáveis 10 e 11, como também as variáveis 13 e 3, são redundantes entre si. Quanto às variáveis 10 e 11, a informação mútua entre elas foi 1,173(SH) e 0,993(CSPR), e quanto às variáveis 13 e 3, 1,130(SH) e 1,159(CSPR). Destaca-se, aqui também, que no SH, a informação mútua com o desfecho é praticamente a mesma, o que já não ocorre com o CSPR, refletindo um maior poder discriminativo.

Note-se que a variável 1, que refere-se ao teor alcoólico, aparece como principal em ambos os métodos. Embora tal variável seja a segunda selecionada no CSPR, a sua informação mútua com o desfecho é praticamente a mesma da observada para a primeira variável.

As redundâncias, destacadas acima, referem-se às seguintes variáveis:

- Intensidade da cor (var 10) e coloração (tonalidade) (var 11).
- Prolina (var 13) e cinzas (var 3)

## 5.2.4

## Resultado Comparativo da Seleção pelo MIFS-U – Dermatologia

Tabela 5.10 – Resultado Comparativo da Seleção pelo MIFS-U – DERMATOLOGIA

Base de Dados		DERMATOLOGIA			
Resultado da Seleção pelo MIFS-U					
Método Shannon / Histograma			Método Cauchy-Schwartz / Parzen-Rosenblatt		
Ordem de Seleção	Variável	IM com Desfecho	Ordem de Seleção	Variável	IM com Desfecho
1 <sup>a</sup>	16	0.833	1 <sup>a</sup>	16	0.471
2 <sup>a</sup>	23	0.818	2 <sup>a</sup>	18	0.457
3 <sup>a</sup>	28	0.409	3 <sup>a</sup>	17	0.401
4 <sup>a</sup>	27	0.004	4 <sup>a</sup>	20	0.238
5 <sup>a</sup>	22	0.004	5 <sup>a</sup>	7	0.294
6 <sup>a</sup>	24	0.009	6 <sup>a</sup>	24	0.000
7 <sup>a</sup>	6	0.004	7 <sup>a</sup>	25	0.000
8 <sup>a</sup>	25	0.009	8 <sup>a</sup>	22	0.000
9 <sup>a</sup>	10	0.151	9 <sup>a</sup>	27	0.000
10 <sup>a</sup>	2	0.016	10 <sup>a</sup>	6	0.000
11 <sup>a</sup>	1	0.005	11 <sup>a</sup>	5	0.069
12 <sup>a</sup>	4	0.092	12 <sup>a</sup>	26	0.020
13 <sup>a</sup>	9	0.064	13 <sup>a</sup>	2	0.010
14 <sup>a</sup>	26	0.031	14 <sup>a</sup>	1	0.000
15 <sup>a</sup>	11	0.008	15 <sup>a</sup>	21	0.161
16 <sup>a</sup>	5	0.195	16 <sup>a</sup>	10	0.030
17 <sup>a</sup>	21	0.330	17 <sup>a</sup>	9	0.022
18 <sup>a</sup>	19	0.136	18 <sup>a</sup>	19	0.076
19 <sup>a</sup>	13	0.033	19 <sup>a</sup>	28	0.010
20 <sup>a</sup>	3	0.362	20 <sup>a</sup>	13	0.008
21 <sup>a</sup>	14	0.136	21 <sup>a</sup>	11	0.002
22 <sup>a</sup>	20	0.416	22 <sup>a</sup>	4	0.075
23 <sup>a</sup>	15	0.248	23 <sup>a</sup>	14	0.062
24 <sup>a</sup>	7	0.452	24 <sup>a</sup>	8	0.284
25 <sup>a</sup>	8	0.396	25 <sup>a</sup>	3	0.113
26 <sup>a</sup>	12	0.601	26 <sup>a</sup>	15	0.098
27 <sup>a</sup>	17	0.762	27 <sup>a</sup>	23	0.244
28 <sup>a</sup>	18	0.777	28 <sup>a</sup>	12	0.242

A variável 16 foi a primeira selecionada em ambos os métodos. Já a variável 23, segunda selecionada no SH, é uma das últimas classificadas pelo CSPR, inversamente ocorrendo com a variável 18. Embora suas informações mútuas com o desfecho sejam significantes, há redundância entre elas. Tanto no SH quanto no CSPR, a informação mútua entre elas é considerável, a saber,

0,709(SH) e 0,380(CSPR). As variáveis 20 e 28, no SH, apresentaram certa redundância, com informação mútua 0,473 . Não refletida, porém, no CSPR, com informação mútua 0,019. Os resultados podem ser considerados semelhantes, mesmo expandindo-se a seleção para as dez primeiras variáveis.

A variável em destaque em ambos os métodos é a 1, que se refere ao baqueteamento dos cones epiteliais.

As redundâncias, citadas acima, referem-se às seguintes variáveis:

- Espongiose (var 23) e estreitamento da epiderme suprapapilar (var 18).
- Hipergranulose focal (var 20) e idade (var 28).

## 5.2.5

## Resultado Comparativo da Seleção pelo MIFS-U – Câncer de Mama

Tabela 5.11 – Resultado Comparativo da Seleção pelo MIFS-U – CÂNCER DE MAMA

Base de Dados			CÂNCER DE MAMA		
Resultado da Seleção pelo MIFS-U					
Método Shannon / Histograma			Método Cauchy-Schwartz / Parzen-Rosenblatt		
Ordem de Seleção	Variável	IM com Desfecho	Ordem de Seleção	Variável	IM com Desfecho
1 <sup>a</sup>	23	0.687	1 <sup>a</sup>	28	0.261
2 <sup>a</sup>	28	0.654	2 <sup>a</sup>	23	0.226
3 <sup>a</sup>	14	0.444	3 <sup>a</sup>	20	0.022
4 <sup>a</sup>	17	0.187	4 <sup>a</sup>	12	0.007
5 <sup>a</sup>	2	0.204	5 <sup>a</sup>	29	0.043
6 <sup>a</sup>	29	0.149	6 <sup>a</sup>	19	0.006
7 <sup>a</sup>	5	0.128	7 <sup>a</sup>	30	0.037
8 <sup>a</sup>	19	0.062	8 <sup>a</sup>	9	0.035
9 <sup>a</sup>	20	0.055	9 <sup>a</sup>	10	0.011
10 <sup>a</sup>	15	0.028	10 <sup>a</sup>	5	0.050
11 <sup>a</sup>	12	0.030	11 <sup>a</sup>	25	0.061
12 <sup>a</sup>	30	0.102	12 <sup>a</sup>	17	0.092
13 <sup>a</sup>	22	0.204	13 <sup>a</sup>	16	0.062
14 <sup>a</sup>	25	0.158	14 <sup>a</sup>	15	0.004
15 <sup>a</sup>	9	0.107	15 <sup>a</sup>	22	0.098
16 <sup>a</sup>	10	0.065	16 <sup>a</sup>	18	0.091
17 <sup>a</sup>	27	0.490	17 <sup>a</sup>	2	0.101
18 <sup>a</sup>	13	0.347	18 <sup>a</sup>	13	0.111
19 <sup>a</sup>	16	0.150	19 <sup>a</sup>	11	0.109
20 <sup>a</sup>	18	0.206	20 <sup>a</sup>	14	0.140
21 <sup>a</sup>	26	0.325	21 <sup>a</sup>	6	0.130
22 <sup>a</sup>	4	0.527	22 <sup>a</sup>	26	0.120
23 <sup>a</sup>	11	0.345	23 <sup>a</sup>	27	0.200
24 <sup>a</sup>	6	0.329	24 <sup>a</sup>	7	0.201
25 <sup>a</sup>	8	0.637	25 <sup>a</sup>	24	0.191
26 <sup>a</sup>	24	0.617	26 <sup>a</sup>	4	0.167
27 <sup>a</sup>	7	0.537	27 <sup>a</sup>	8	0.230
28 <sup>a</sup>	1	0.534	28 <sup>a</sup>	1	0.182
29 <sup>a</sup>	21	0.657	29 <sup>a</sup>	3	0.193
30 <sup>a</sup>	3	0.565	30 <sup>a</sup>	21	0.215

As variáveis 23 e 28, embora em ordem invertida, foram as primeiras classificadas em ambos os métodos. Pode-se observar, ainda, que as outras três variáveis, em cada método, exceto a variável 14 no SH, têm informações mútuas com o desfecho muito baixas, não marcando assim uma diferença. Além disso, se estender-se a seleção para as 9 primeiras classificadas, seis delas são comuns em ambos os métodos.

## 5.2.6

**Resultado Comparativo da Seleção pelo MIFS-U – Doenças Cardíacas**Tabela 5.12 – Resultado Comparativo da Seleção pelo MIFS-U  
DOENÇAS CARDÍACAS

Base de Dados		DOENÇAS CARDÍACAS			
Resultado da Seleção pelo MIFS-U					
Método Shannon / Histograma			Método Cauchy-Schwartz / Parzen-Rosenblatt		
Ordem de Seleção	Variável	IM com Desfecho	Ordem de Seleção	Variável	IM com Desfecho
1 <sup>a</sup>	9	0.0957	1 <sup>a</sup>	2	0.062
2 <sup>a</sup>	8	0.0944	2 <sup>a</sup>	8	0.036
3 <sup>a</sup>	2	0.0926	3 <sup>a</sup>	9	0.026
4 <sup>a</sup>	7	0.0672	4 <sup>a</sup>	1	0.024
5 <sup>a</sup>	5	0.0653	5 <sup>a</sup>	7	0.019
6 <sup>a</sup>	1	0.0487	6 <sup>a</sup>	6	0.020
7 <sup>a</sup>	6	0.0374	7 <sup>a</sup>	5	0.020
8 <sup>a</sup>	3	0.0114	8 <sup>a</sup>	4	0.008
9 <sup>a</sup>	4	0.0147	9 <sup>a</sup>	3	0.002
10 <sup>a</sup>	12	0.0931	10 <sup>a</sup>	11	0.006
11 <sup>a</sup>	10	0.0866	11 <sup>a</sup>	12	0.012
12 <sup>a</sup>	13	0.0769	12 <sup>a</sup>	13	0.010
13 <sup>a</sup>	11	0.0443	13 <sup>a</sup>	10	0.003

Mais uma vez, a seleção feita pelo MIFS-U, através dos dois métodos, é praticamente a mesma. As variáveis 5 e 7, embora não sejam redundantes entre si, apresentam informações mútuas com o desfecho muito próximas. A seleção de uma e de outra em cada um dos métodos torna-se, portanto, uma questão de proximidade numérica. Ademais, observa-se que, se a seleção for estendida para as 7, ou mesmo 9, primeiras classificadas, a seleção seria exatamente a mesma.

Observe-se que as três principais variáveis, em ambos os métodos, são exatamente as mesmas, embora a ordem de seleção de um em relação ao outro seja invertida. Tais variáveis são as seguintes:

- Tipo de dor no peito (var 2);
- Talassemias (var 8);
- Idade (var 9).