



**Centro de Informática**

★ • • • • • • • • • • UFPE

# Aprendizagem de Dados Simbólicos e/ou Numéricos

*Francisco de A.T. de Carvalho*

# Funções de Proximidade: Dados Simbólicos Modais 1/

➡ Coeficiente de Afinidade (Bacelar-Nicolau).

➡  $E = \{1, \dots, N\}$ : conjunto de indivíduos

➡  $Y = \{Y_1, \dots, Y_p\}$ : conjunto de variáveis modais

➡  $X = (\xi_{kj})_{N \times p}$ : matriz de dados simbólicos com  $N$  objetos  $k$

# Funções de Proximidade: Dados Simbólicos Modais 2/

➤ Coeficiente de Afinidade (Bacelar-Nicolau).

➤  $Y_j$ : variável modal categórica com  $m_j$  modalidades  
 $\{1, \dots, m_j\}$

➤  $\xi_{kj} = (n_{kj1}, \dots, n_{kjm_j})$  onde  $n_{kjv}$  retorna o número de indivíduos que partilham a categoria  $v$  da variável  $Y_j$

➤  $\mathbf{x}_k = (\xi_{k1}, \dots, \xi_{kp})$ : objeto simbólicos modal

# Funções de Proximidade: Dados Simbólicos Modais 3/

➡ Coeficiente de Afinidade (Bacelar-Nicolau).

<b>X</b>	...	$Y_j$	...	$\Sigma$
...	...	...	...	...
<b>k</b>	...	$(n_{kj1}, \dots, n_{kjmj})$	...	$n_{k..}$
...	...	...	...	...
<b>k'</b>	...	$(n_{k'j1}, \dots, n_{k'jmj})$	...	$n_{k'..}$
...	...	...	...	...
$\Sigma$	...	$(n_{.j1}, \dots, n_{.jmj})$	...	$n_{..}$

# Funções de Proximidade: Dados Simbólicos Modais 4/

➡ Coeficiente de Afinidade (Bacelar-Nicolau).

➡ Coeficiente de afinidade parcial

$$\text{aff}(\xi_{kj}, \xi_{k'j}) = \sum_{l=1}^{m_j} \sqrt{\left( \frac{n_{kjl}}{n_{kj\cdot}} \right) \left( \frac{n_{k'jl}}{n_{k'j\cdot}} \right)}$$

$$n_{kj\cdot} = \sum_{l=1}^{m_j} n_{kjl} \quad n_{k'j\cdot} = \sum_{l=1}^{m_j} n_{k'jl}$$

# Funções de Proximidade: Dados Simbólicos Modais 5/

➤ Coeficiente de Afinidade (Bacelar-Nicolau).

➤ Coeficiente de afinidade parcial

➤  $n_{kj}$  : número de indivíduos pertencentes ou observados no item (grupo)  $k$  para os quais a variável  $Y_j$  foi observada

# Funções de Proximidade: Dados Simbólicos Modais 6/

➤ Coeficiente de Afinidade (Bacelar-Nicolau).

➤ Coeficiente de afinidade parcial

➤ Propriedades

➤  $0 \leq \text{aff}(\xi_{kj}, \xi_{k'j}) \leq 1$

➤  $\text{aff}(\xi_{kj}, \xi_{k'j}) = 1$  se  $\xi_{kj}$  e  $\xi_{k'j}$  forem idênticos ou proporcionais

➤  $\text{aff}(\xi_{kj}, \xi_{k'j}) = 0$  se  $\xi_{kj}$  e  $\xi_{k'j}$  forem ortogonais

# Funções de Proximidade: Dados Simbólicos Modais 7/

➡ Coeficiente de Afinidade (Bacelar-Nicolau).

➡ Coeficiente de afinidade

$$aff(k, k') = \sum_{j=1}^p w_j aff(\xi_{kj}, \xi_{k'j}) = \sum_{j=1}^p w_j \sum_{l=1}^{m_j} \sqrt{\left( \frac{n_{kjl}}{n_{kj\bullet}} \right) \left( \frac{n_{k'jl}}{n_{k'j\bullet}} \right)}$$

$$0 \leq w_j \leq 1 \quad \sum_{j=1}^p w_j = 1$$

# Funções de Proximidade: Dados Simbólicos Modais 8/

➡ Outras funções de proximidade

➡ Distancia  $L_1$

$$d(k, k') = \sum_{j=1}^p w_j \sum_{l=1}^{m_j} \left| \left( \frac{n_{kjl}}{n_{kj\bullet}} \right) - \left( \frac{n_{k'jl}}{n_{k'j\bullet}} \right) \right|$$

$$0 \leq w_j \leq 1 \quad \sum_{j=1}^p w_j = 1$$

# Funções de Proximidade: Dados Simbólicos Modais 9/

➡ Outras funções de proximidade

➡ Distancia  $L_2$

$$d(k, k') = \sum_{j=1}^p w_j \sqrt{\sum_{l=1}^{m_j} \left| \left( \frac{n_{kjl}}{n_{kj\bullet}} \right)^2 - \left( \frac{n_{k'jl}}{n_{k'j\bullet}} \right)^2 \right|}$$

$$0 \leq w_j \leq 1 \qquad \sum_{j=1}^p w_j = 1$$

# Distancia $L_2$

$$d(k, k') = \sum_{j=1}^p w_j \left[ \sum_{l=1}^{m_j} \left( n_{kjl} n_{k'jl} \right) - \frac{\left( \sum_{l=1}^{m_j} n_{kjl} \right) \left( \sum_{l=1}^{m_j} n_{k'jl} \right)}{m_j} \right] - \left[ \sum_{l=1}^{m_j} \left( n_{kjl}^2 \right) - \frac{\left( \sum_{l=1}^{m_j} n_{kjl} \right)^2}{m_j} \right] \left[ \sum_{l=1}^{m_j} \left( n_{k'jl}^2 \right) - \frac{\left( \sum_{l=1}^{m_j} n_{k'jl} \right)^2}{m_j} \right]$$

$$0 \leq w_j \leq 1$$

Francisco de A.T. de Carvalho,  
CIn/UFPE

$$\sum_{j=1}^p w_j = 1$$

# Funções de Matching:

## Dados Simbólicos Booleanos 1/27

### ➤ Matching.

- compara a descrição de uma classe  $C$  (modelo) com a descrição de um individuo  $k$ 
  - objetivo: estabelecer se esse individuo pode ser considerado uma instancia dessa classe

### ➤ Matching Canônico

- Existe um matching perfeito entre as estruturas?

# Funções de Matching:

## Dados Simbólicos Boleanos 2/27

### ➤ Matching Flexível

➤ Qual o grau de matching entre as estruturas?

### ➤ Matching Canônico

➤ Seja  $S$  o espaço de objetos simbólicos booleanos

➤ O matching canônico é uma função

$$\textit{Match} : S \times S \rightarrow \{0,1\}$$

# Funções de Matching:

## Dados Simbólicos Boleanos 3/27

### ➤ Matching Canônico

➤ que associa a cada par de objetos simbólicos booleanos

$$a = [y_1 R_1 A_1] \wedge \dots \wedge [y_p R_p A_p]$$

$$b = [y_1 R_1 B_1] \wedge \dots \wedge [y_p R_p B_p]$$

# Funções de Matching: Dados Simbólicos Booleanos 4/27

## ➡ Matching Canônico

➡ os valores 1 e 0 segundo

$$Match(a, b) = \begin{cases} 1, & \text{se } B_j \subseteq A_j, \forall j \\ 0, & \text{senão} \end{cases}$$

# Funções de Matching:

## Dados Simbólicos Booleanos 5/27

### ➤ Matching Canônico

#### ➤ Exemplo

$$➤ a = [ \text{idade } R_1 [30,50] ] \wedge [ \text{altura } R_2 [160,180] ]$$

$$➤ b_1 = [ \text{idade} = 35 ] \wedge [ \text{altura} = 164 ]$$

$$➤ b_2 = [ \text{idade} = 55 ] \wedge [ \text{altura} = 170 ]$$

$$➤ \text{Match}(a, b_1) = 1 \qquad \text{Match}(a, b_2) = 0$$

# Funções de Matching: Dados Simbólicos Booleanos 6/27

## ➤ Matching Flexível

- Seja  $S$  o espaço de objetos simbólicos booleanos
- O matching flexível é uma função

$$\textit{Match} : S \times S \rightarrow [0,1]$$

- que associa a cada par de objetos simbólicos booleanos

# Funções de Matching: Dados Simbólicos Boleanos 7/27

## ➤ Matching Flexível

$$a = [y_1 R_1 A_1] \wedge \dots \wedge [y_p R_p A_p]$$

$$b = [y_1 R_1 B_1] \wedge \dots \wedge [y_p R_p B_p]$$

➤ um valor real no intervalo  $[0,1]$  tal que

$$Match(a,b) = \begin{cases} 1, & \text{se } Match(a,b) = 1 \\ \in [0,1), & \text{senão} \end{cases}$$

# Funções de Matching:

## Dados Simbólicos Booleanos 8/27

### ➤ Matching Flexível: Modelo

➤ Seja um objeto simbólico booleano  $b'$  obtido de um objeto  $b$  através de mudanças nos valores dos atributos tal que

➤  $\text{Match}(a, b') = 1$

➤ Seja  $P(b/b')$  a probabilidade condicional de observar  $b$  dado que a observação original era  $b'$ .

# Funções de Matching: Dados Simbólicos Booleanos 9/27

## ➡ Matching Flexível: Modelo

➡ Seja  $S_a = \{b' \in S / \text{Match}(a, b') = 1\}$

$$F - \text{Match}(a, b) := \underset{b' \in S_a}{\text{Max}} P(b / b'), \forall a, b \in S$$

# Funções de Matching:

## Dados Simbólicos Boleanos 10/27

### ➤ Matching Flexível: Modelo

➤ Seja  $b = b_1 \wedge \dots \wedge b_p$  e  $b' = b'_1 \wedge \dots \wedge b'_p$  onde

➤  $b_j = [y_j = z_j]$  e  $b'_j = [y_j = z'_j]$

$$P(b / b') = \prod_{j=1}^p P(b_j / b') = \prod_{j=1}^p P(b_j / b'_1 \wedge \dots \wedge b'_p)$$

(independência)

# Funções de Matching:

## Dados Simbólicos Boleanos 11/27

### ➤ Matching Flexível: Modelo

#### ➤ Simplificação:

$$P(b_j / b') = P(b_j / b'_j)$$

#### ➤ Quanto mais similar os eventos

$$➤ b_j = [y_j = z_j] \text{ e } b'_j = [y_j = z'_j]$$

#### ➤ maior a probabilidade

$$P(b_j / b'_j) = P([y_j = z_j] / [[y_j = z'_j]$$

# Funções de Matching: Dados Simbólicos Boleanos 12/27

## ➤ Matching Flexível: Modelo

### ➤ Dispondo-se

- da distribuição de probabilidade do domínio  $O_j$ 
  - correspondente a variável aleatória  $y_j$
- de uma função de dissimilaridade  $\delta_j$  definida em  $O_j$

# Funções de Matching:

## Dados Simbólicos Boleanos 13/27

### ➤ Matching Flexível: Modelo

➤ Pode-se definir

$$\begin{aligned} P(b_j / b'_j) &:= P([y_j = z_j] / [[y_j = z'_j]]) \\ &= P(\{X / \delta_j(z'_j, X) \geq \delta_j(z'_j, z_j)\}) \end{aligned}$$

➤  $P(b_j / b'_j) = P(b_j / b'_j)$  é a probabilidade de observar uma distorção entre  $X$  e  $z'_j$  que é maior do que a existente entre  $z_j$  e  $z'_j$ .

# Funções de Matching:

## Dados Simbólicos Boleanos 14/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

##### ➤ Variáveis contínuas

$$\delta_j(x, y) = |x - y|$$

##### ➤ Variáveis nominais

$$\delta_j(x, y) = \begin{cases} 0, & \text{se } x = y \\ 1, & \text{se } x \neq y \end{cases}$$

# Funções de Matching: Dados Simbólicos Boleanos 15/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Variáveis ordinais

$$\delta_j(x, y) = |\text{ord}(x) - \text{ord}(y)|$$

#### ➤ Exemplo

➤  $O_1 = \{\text{amarelo, branco, azul, verde, vermelho, preto}\}$

➤  $O_2 = \{\text{pequeno, médio, grande}\}$

# Funções de Matching:

## Dados Simbólicos Boleanos 16/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

#### ➤ Exemplo

$$➤ a = [ \text{cor } R_1 \{ \text{azul, verde} \} ] \wedge [ \text{tamanho } R_2 \{ \text{médio, grande} \} ]$$

$$➤ b = [ \text{cor} = \text{vermelho} ] \wedge [ \text{tamanho} = \text{pequeno} ]$$

➤  $Y_1 = \text{cor}$  distribuição uniforme em  $O_1$ :

$$➤ P(Y_1 = x) = 1/6$$

# Funções de Matching: Dados Simbólicos Boleanos 17/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Exemplo

➤  $Y_2 =$  tamanho

➤  $P(Y_2 = \text{pequeno}) = 0.3$

➤  $P(Y_2 = \text{médio}) = 0.6$

➤  $P(Y_2 = \text{grande}) = 0.1$

# Funções de Matching: Dados Simbólicos Boleanos 18/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Exemplo

➤ Para  $Y_1$

$$\delta_j(x, y) = \begin{cases} 0, & \text{se } x = y \\ 1, & \text{se } x \neq y \end{cases}$$

#### ➤ Para $Y_2$

➤  $\delta(\text{pequeno}, \text{grande}) \geq \delta(\text{médio}, \text{pequeno}) \geq \delta(\text{médio}, \text{grande})$

# Funções de Matching: Dados Simbólicos Booleanos 19/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Exemplo

##### ➤ Para $Y_1$

➤ Passo 1:  $Y_1 = \text{vermelho} \notin \{\text{azul}, \text{verde}\}$

➤ Passo 2: Candidatos

$\{ b_1' = [Y_1 = \text{azul}], b_1'' = [Y_1 = \text{verde}] \}$

# Funções de Matching:

## Dados Simbólicos Boleanos 20/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

#### ➤ Exemplo

#### ➤ Para $Y_1$

➤ Passo 1:  $Y_1 = \text{vermelho} \notin \{\text{azul}, \text{verde}\}$

➤ Passo 2: Candidatos

$\{ b_1' = [Y_1 = \text{azul}], b_1'' = [Y_1 = \text{verde}] \}$

# Funções de Matching:

## Dados Simbólicos Boleanos 21/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

#### ➤ Exemplo

#### ➤ Para $Y_1$

#### ➤ Passo 2:

$$b_1' = [Y_1 = \text{azul}]$$

$$P(X/\delta_1(\text{azul}, X) \geq \delta_1(\text{azul}, \text{vermelho})) =$$

$$P(X/\delta_1(\text{azul}, X) \geq 1 - P(X \neq \text{azul}) = 1 - 1/6 = 5/6$$

# Funções de Matching: Dados Simbólicos Booleanos 22/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Exemplo

##### ➤ Para $Y_1$

##### ➤ Passo 2:

$$b_1' = [Y_1 = \text{verde}]$$

$$P(X \neq \text{verdel}) = 1 - 1/6 = 5/6$$

Então  $P = 5/6$  e  $z_1 = \text{verde}$

# Funções de Matching:

## Dados Simbólicos Booleanos 23/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

#### ➤ Exemplo

#### ➤ Para $Y_2$

#### ➤ Passo 1:

$Y_2 = \text{pequeno} \notin \{\text{médio}, \text{grande}\}$

#### ➤ Passo 2:

Candidatos  $\{ b_2' = [Y_2 = \text{médio}], b_2'' = [Y_2 = \text{grande}] \}$

# Funções de Matching:

## Dados Simbólicos Booleanos 24/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

#### ➤ Exemplo

#### ➤ Para $Y_2$

#### ➤ Passo 1:

$Y_2 = \text{pequeno} \notin \{\text{médio}, \text{grande}\}$

#### ➤ Passo 2:

Candidatos  $\{ b_2' = [Y_2 = \text{médio}], b_2'' = [Y_2 = \text{grande}] \}$

# Funções de Matching:

## Dados Simbólicos Boleanos 25/27

### ➤ Matching Flexível: Modelo

#### ➤ Exemplos de distâncias

#### ➤ Exemplo

#### ➤ Para $Y_2$

#### ➤ Passo 2:

$$b_2' = [Y_2 = \text{médio}]$$

$$P(X/\delta_1(\text{médio}, X) \geq \delta_1(\text{pequeno}, \text{médio})) =$$

$$P(X = \text{pequeno} \text{ ou } X = \text{grande}) = 0.3 + 0.1 = 0.4$$

# Funções de Matching: Dados Simbólicos Boleanos 26/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Exemplo

#### ➤ Para $Y_2$

#### ➤ Passo 2:

$$b_2' = [Y_2 = \text{grande}]$$

$$P(X/\delta_1(\text{grande}, X) \geq$$

$$\delta_1(\text{pequeno}, \text{grande})) = P(X = \text{pequeno}) = 0.3$$

# Funções de Matching: Dados Simbólicos Boleanos 27/27

## ➤ Matching Flexível: Modelo

### ➤ Exemplos de distâncias

#### ➤ Exemplo

➤ Então  $P=0.4$  e  $z_2 = \text{grande}$

$$F\text{-Match}(a,b) := \underset{b' \in S_a}{\text{Max}} P(b/b') = 5/6 * 0.4 = 0.33$$

# Funções de Dissimilaridade: SODAS 1/

➔ Entrada: Matriz de Dados Simbólicos

SOEditor - WV.FIL - OLEOS.SDS

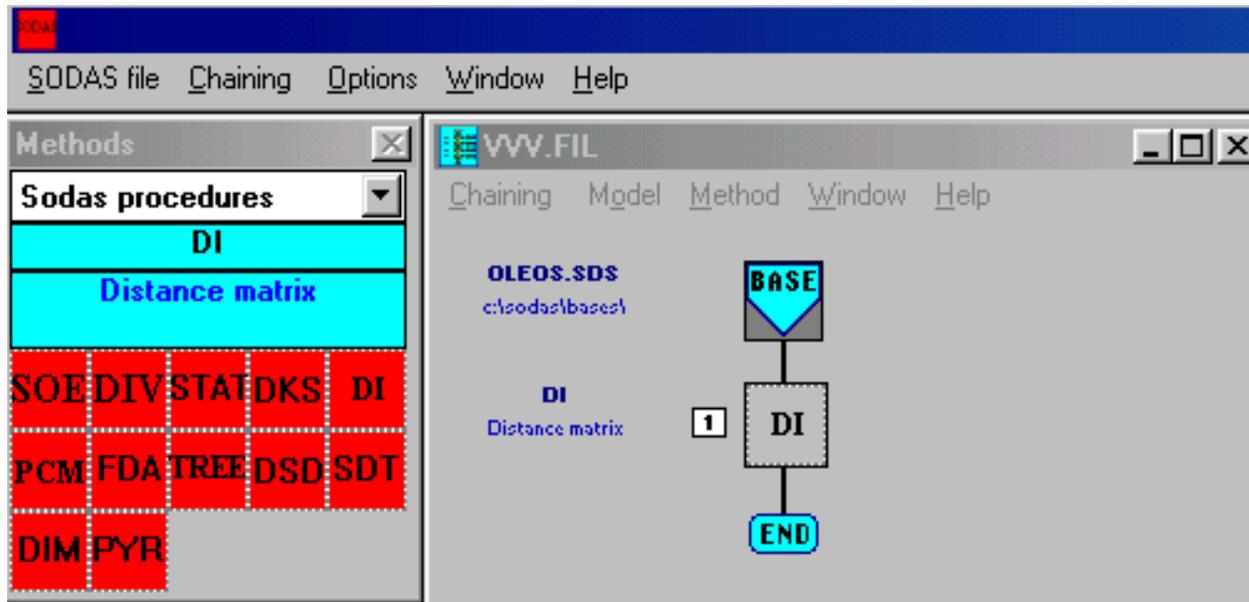
File Edit View Selection Graphic Window Help

Table

	specific gravit	freezing point	iodine value	saponification	Major Fatty aci
	[ 0.93 : 0.94 ]	[ -27.00 : -18.00 ]	[ 170.00 : 204.00 ]	[ 118.00 : 196.00 ]	Linoleic Acid, Linolenic Acid, Oleic Acid, Palmitic Acid, Myristic Ac
	[ 0.93 : 0.94 ]	[ -5.00 : -4.00 ]	[ 192.00 : 208.00 ]	[ 188.00 : 197.00 ]	Linoleic Acid, Linolenic Acid, Oleic Acid, Palmitic Acid, Searic Ac
	[ 0.92 : 0.92 ]	[ -6.00 : -1.00 ]	[ 99.00 : 113.00 ]	[ 189.00 : 198.00 ]	Linoleic Acid, Oleic Acid, Palmitic Acid, Myristic Acid, Searic Ac
	[ 0.92 : 0.93 ]	[ -6.00 : -4.00 ]	[ 104.00 : 116.00 ]	[ 187.00 : 193.00 ]	Linoleic Acid, Oleic Acid, Palmitic Acid, Searic Acid, Arachid Ac
	[ 0.92 : 0.92 ]	[ -21.00 : -15.00 ]	[ 80.00 : 82.00 ]	[ 189.00 : 193.00 ]	Linoleic Acid, Oleic Acid
	[ 0.91 : 0.92 ]	[ 0.00 : 6.00 ]	[ 79.00 : 90.00 ]	[ 187.00 : 196.00 ]	Linoleic Acid, Oleic Acid, Palmitic Acid, Searic Acid
	[ 0.86 : 0.87 ]	[ 30.00 : 38.00 ]	[ 40.00 : 48.00 ]	[ 190.00 : 199.00 ]	Oleic Acid, Palmitic Acid, Myristic Acid, Searic Acid, Capric Ac
	[ 0.86 : 0.86 ]	[ 22.00 : 32.00 ]	[ 53.00 : 77.00 ]	[ 190.00 : 202.00 ]	Linoleic Acid, Oleic Acid, Palmitic Acid, Myristic Acid, Searic Acid, Laur

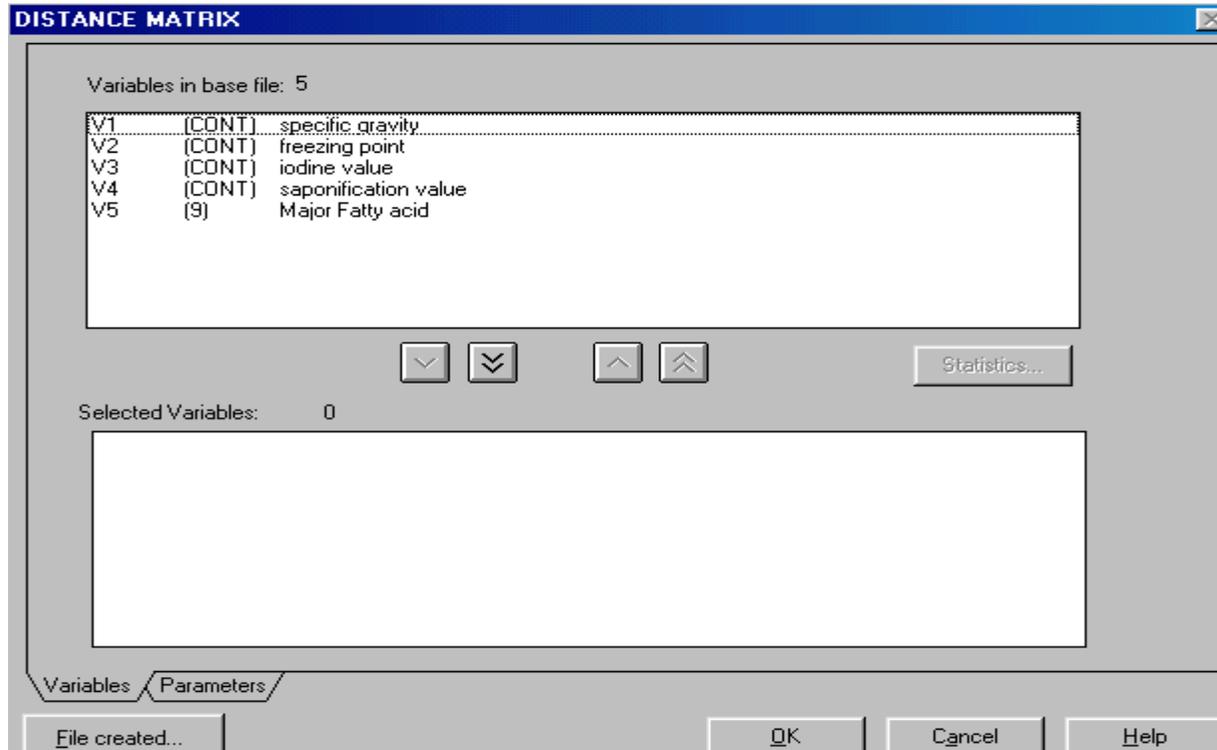
# Funções de Dissimilaridade: SODAS 2/

➡ Entrada: Comando DI



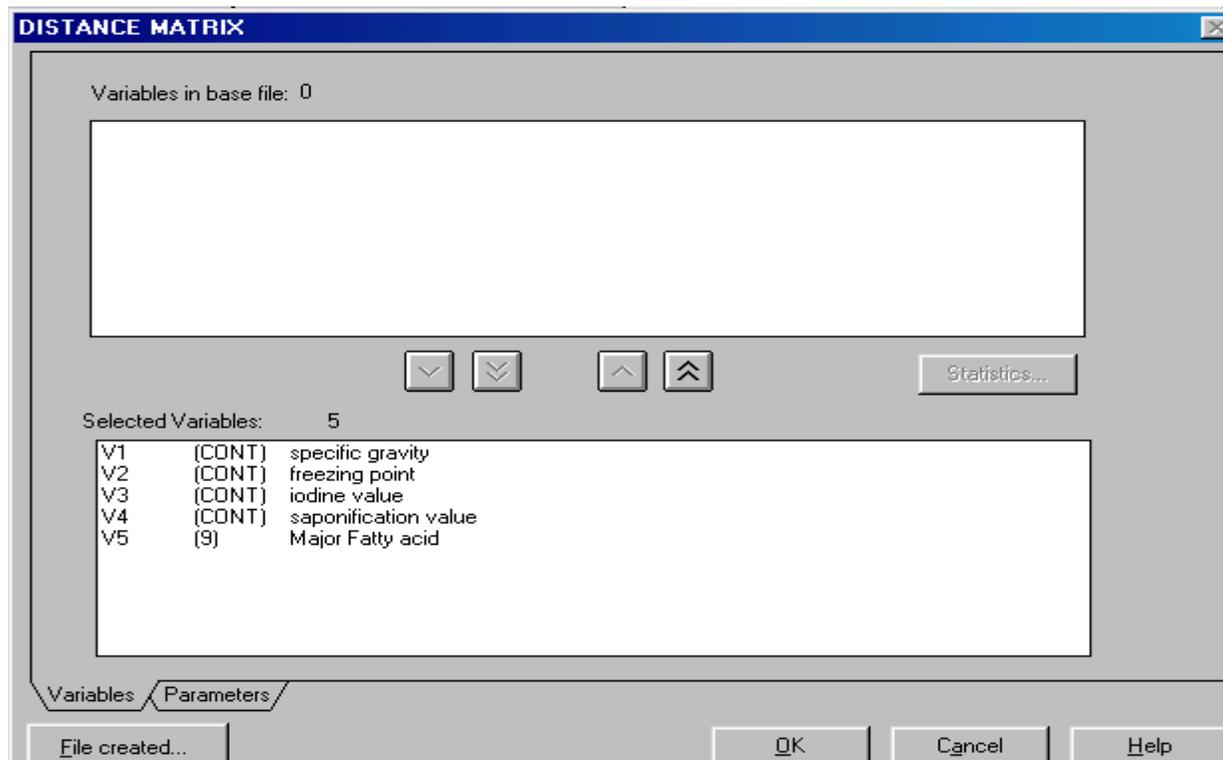
# Funções de Dissimilaridade: SODAS 3/

➡ Entrada: Comando DI: Seleção de Variáveis



# Funções de Dissimilaridade: SODAS 4/

➡ Entrada: Comando DI: Seleção de Variáveis



# Funções de Dissimilaridade: SODAS 5/

➡ Entrada: Comando DI: Parâmetros

**DISTANCE MATRIX**

Parameters

choice of a function of similarity      choice of a comparison function

Gamma: 0.5      Order of the distance function: 2

Weights

Uniform weights  
 List      Define

Class BSD number: 2

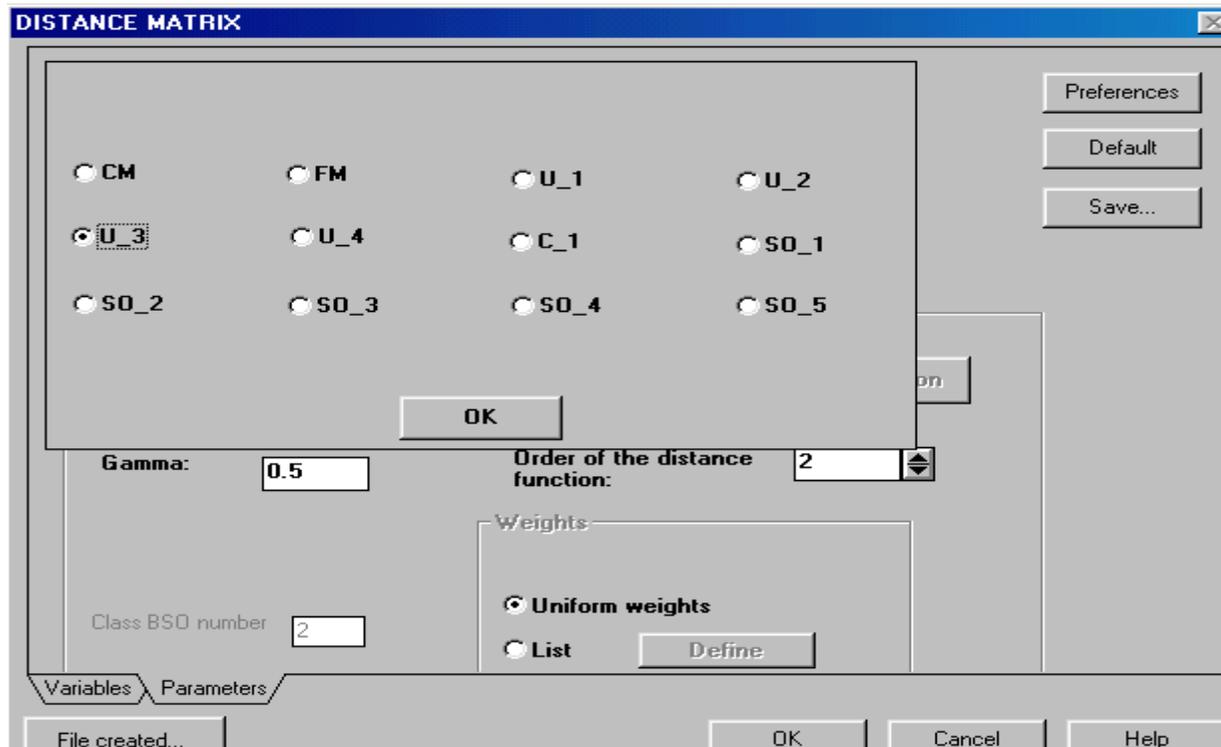
Variables    Parameters

File created...      OK      Cancel      Help

Preferences  
Default  
Save...

# Funções de Dissimilaridade: SODAS 6/

➡ Entrada: Comando DI: Parâmetros



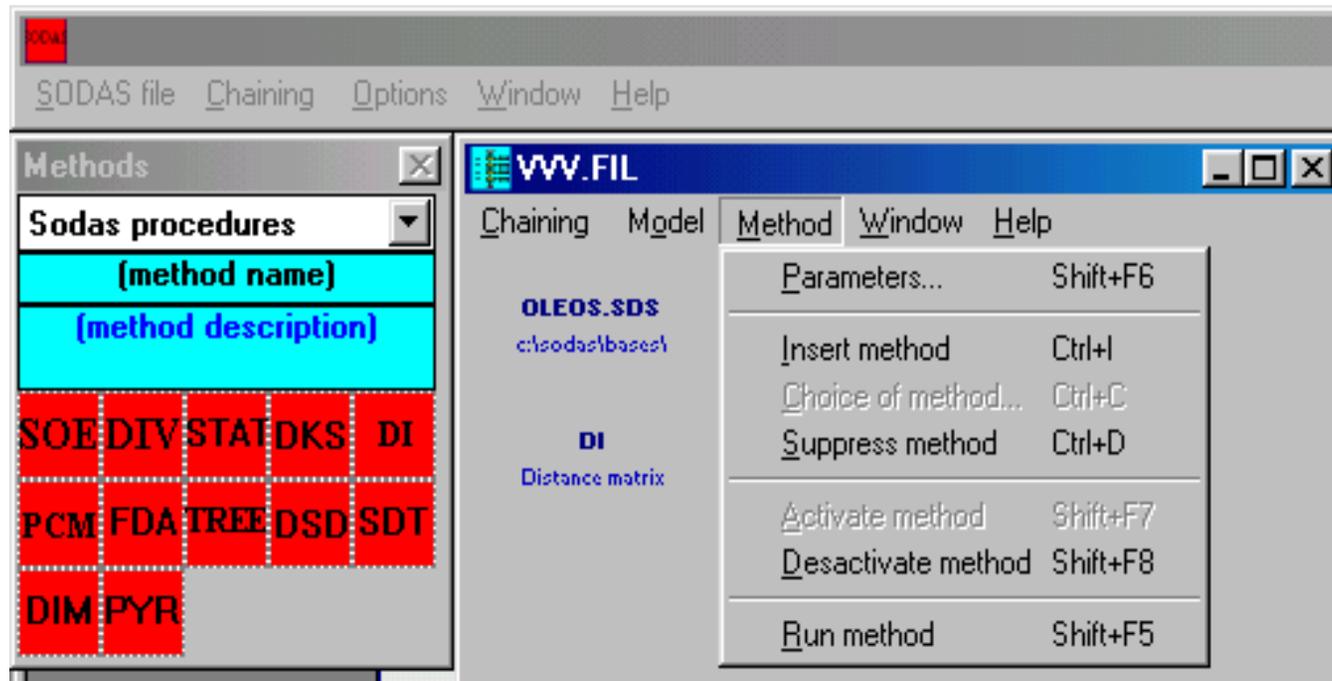
# Funções de Dissimilaridade: SODAS 7/

➡ Entrada: Comando DI: Parâmetros

<i>Dissimilarity measure</i>	<i>Parameters</i>	<i>Constraints</i>	<i>Default</i>
U_1 (Gowda & Diday)	none		
U_2 (Ichino & Yaguchi)	Gamma	[0 .. 0.5]	0.5
	Order of power	1 .. 10	2
U_3 (Normalized Ichino & Yaguchi)	Gamma	[0 .. 0.5]	0.5
	Order of power	1 .. 10	2
U_4 (Weighted Normalized Ichino & Yaguchi)	Gamma	[0 .. 0.5]	0.5
	Order of power	1 .. 10	2
	List of weights per variable	Sum(weights) = 1.0	Equal weights
C_1 (Normalized De Carvalho)	Comparison function	$D_1, D_2, D_3, D_4, D_5$	$D_1$
	Order of power	1 .. 10	2
SO_1 (De Carvalho)	Comparison function	$D_1, D_2, D_3, D_4, D_5$	$D_1$
	Order of power	1 .. 10	2

# Funções de Dissimilaridade: SODAS 8/

➡ Entrada: Comando DI: Saída



# Funções de Dissimilaridade: SODAS 9/

➡ Entrada: Comando DI: Saída

The screenshot displays the SODAS 9.0 software interface. On the left, a 'Methods' window lists 'Sodas procedures' with a table of method names and descriptions. The main window, titled 'VVV.FIL', shows a flowchart for the 'DI' (Distance matrix) procedure. The flowchart starts with 'OLEOS.SDS' (c:\sodas\bases\), leading to a 'DI' box (Distance matrix) which is connected to 'BASE' and 'END' boxes. A green box with the number '1' is also present next to the 'DI' box.

Sodas procedures				
(method name)				
(method description)				
SOE	DI	STAT	DKS	DI
PCM	FDA	TREE	DSD	SDT
DIM	PYR			

# Funções de Dissimilaridade: SODAS 10/

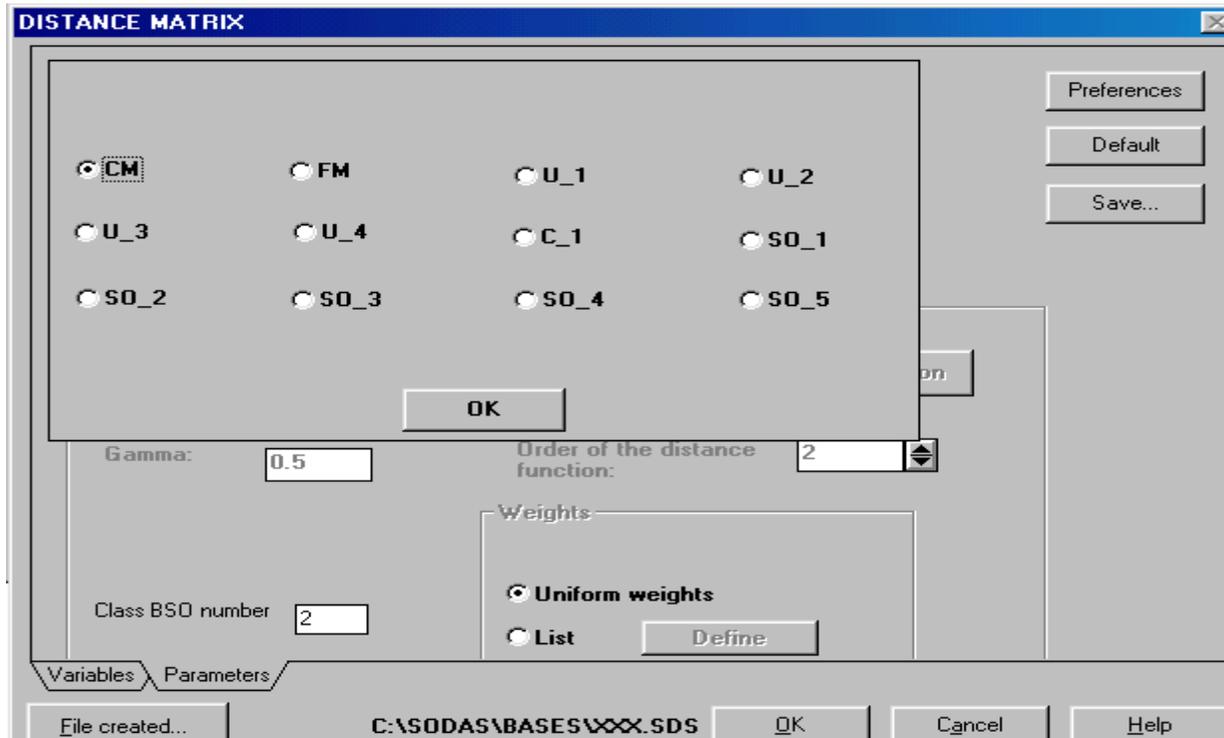
## 👉 Entrada: Comando DI: Saída

```
*****D I S T A N C E   M E A S U R E S*****  
  
Data Information:  
  
      Input Sodas File: C:\SODAS\BASES\OLEOS.SDS  
  
8 Boolean Symbolic Objects (BSOs) read.  
5 Variables selected for each BSO: 1 -- 5  
  
Selected Distance Function: U_3   Normalized Ichino & Yaguchi  
      Gamma:    0.5  
      Power:    2  
  
Distance Matrix  
  
      BSO          1          2          3          4  
      1              0  
      2          0.5232          0  
      3          0.7467          0.6084          0  
      4          0.7281          0.5639          0.1447          0  
      5          0.8159          0.7873          0.3172          0.3232
```

# Matching Canônico

## SODAS 1/

➡ Entrada: Comando DI: Parâmetros



# Matching Canônico

## SODAS 2/

➡ Entrada: Comando DI: Saída

```
***** C A N O N I C A L   M A T C H I N G   *****
```

### Data Information

```
8 Boolean Symbolic Objects (BSOs) read.  
2 : Boolean Symbolic Object (BSO) selected as class.  
5 Variables selected for each BSO: 1 -- 5
```

### Matching Vector

```
BSO          2  
  
1      No Match  
2      Match  
3      No Match  
4      No Match  
5      No Match
```

Page 2

SODAS

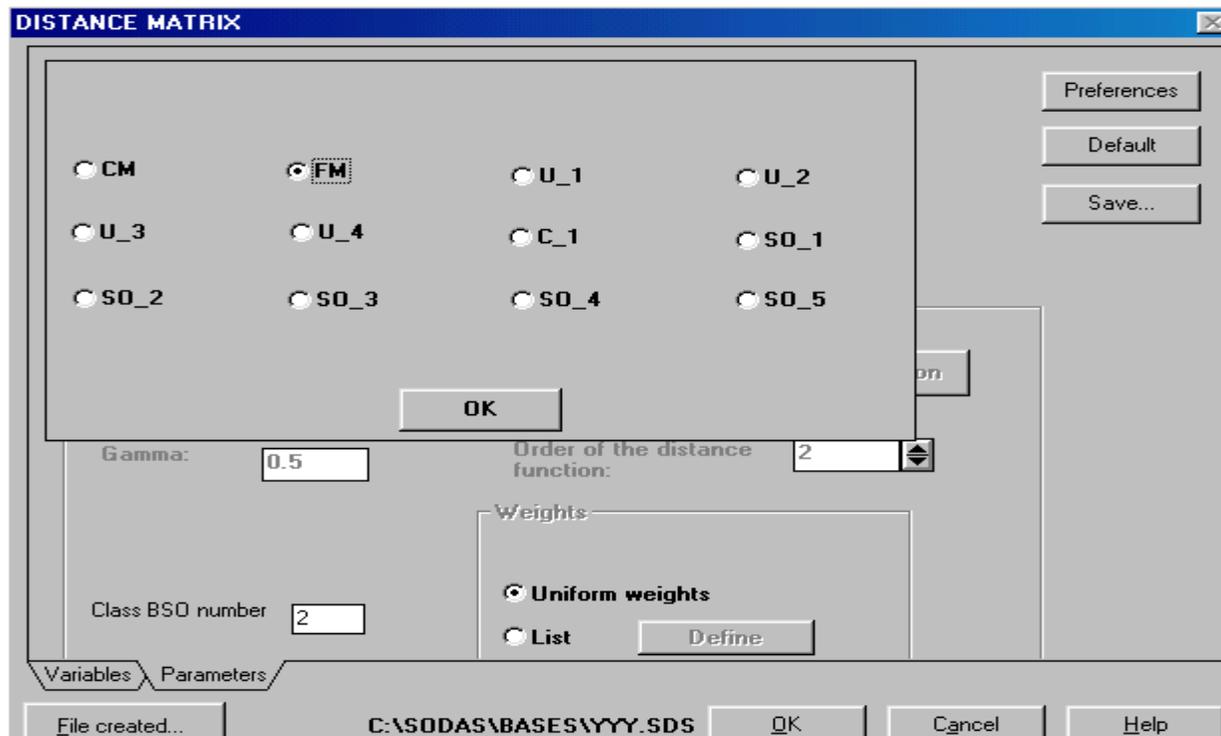
11/04/03

```
BSO          2  
  
6      No Match  
7      No Match  
8      No Match
```

# Matching Flexível

## SODAS 1/

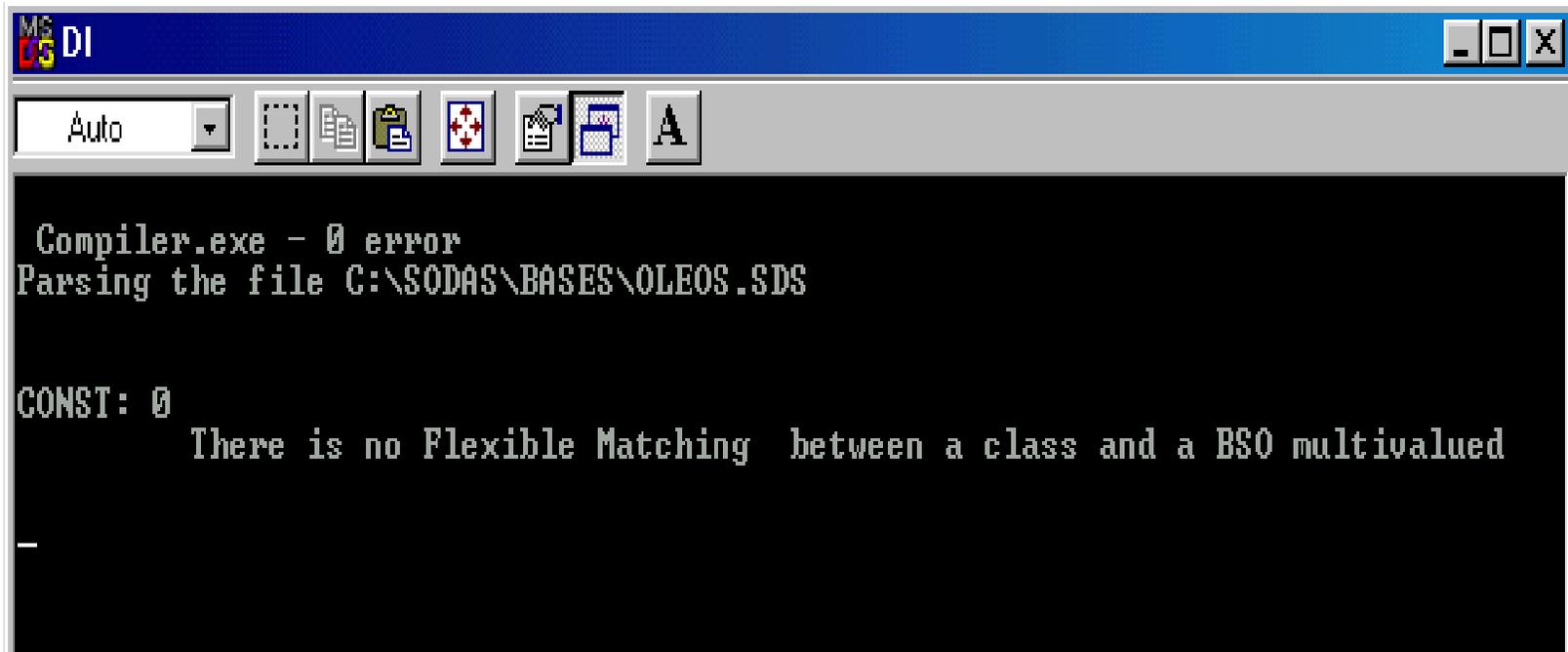
➡ Entrada: Comando DI: Parâmetros



# Matching Flexível

## SODAS 2/

➡ Entrada: Comando DI: Saída



```
MS-DOS DI
Auto
Compiler.exe - 0 error
Parsing the file C:\SODAS\BASES\OLEOS.SDS

CONST: 0
    There is no Flexible Matching between a class and a BSO multivalued
-
```

# Funções de Similaridade: SODAS 1/

➔ Entrada: Matriz de Dados Simbólicos Modais

	sex	age	
Student / masc.	masc. (1.00)	15-19 years (0.68), 20-24 years (0.31), 25-34 years (0.02)	seconde
Retired / fem.	fem. (1.00)	+ 65 years (0.60), 45-64 years (0.39), 35-44 years (0.01)	seconde
Employed / fem.	fem. (1.00)	15-19 years (0.05), + 65 years (0.01), 45-64 years (0.25), 35-44 years (0.26), 20-24 years (0.13), 25-34 years (0.30)	seconde
Small Indep. /	fem. (1.00)	+ 65 years (0.03), 45-64 years (0.37), 35-44 years (0.34), 20-24 years (0.08), 25-34 years (0.18)	seconde
Housewife / fem	fem. (1.00)	+ 65 years (0.23), 45-64 years (0.68), 35-44 years (0.04), 25-34 years (0.04)	
Medium Staff /	masc. (1.00)	15-19 years (0.02), + 65 years (0.02), 45-64 years (0.18), 35-44 years (0.24), 20-24 years (0.22), 25-34 years (0.31)	seconde
Ind. Worker / m	masc. (1.00)	15-19 years (0.08), 45-64 years (0.22), 35-44 years (0.17), 20-24 years (0.18), 25-34 years (0.36)	seconde
Small Indep. /	masc. (1.00)	15-19 years (0.02), + 65 years (0.06), 45-64 years (0.41), 35-44 years (0.18), 20-24 years (0.11), 25-34 years (0.22)	seconde
Intelec / cient	masc. (1.00)	45-64 years (0.21), 35-44 years (0.28), 20-24 years (0.05), 25-34 years (0.46)	
Student / fem.	fem. (1.00)	15-19 years (0.74), 20-24 years (0.23), 25-34 years (0.03)	
Other / fem.	fem. (1.00)	15-19 years (0.13), 45-64 years (0.35), 35-44 years (0.11), 20-24 years (0.19), 25-34 years (0.23)	seconde
Employed / masc	masc. (1.00)	15-19 years (0.08), + 65 years (0.01), 45-64 years (0.16), 35-44 years (0.27), 20-24 years (0.29), 25-34 years (0.20)	seconde
Other / masc.	masc. (1.00)	15-19 years (0.09), + 65 years (0.04), 45-64 years (0.30), 35-44 years (0.15), 20-24 years (0.13), 25-34 years (0.28)	seconde
Dir. / Lib. / m	masc. (1.00)	+ 65 years (0.03), 45-64 years (0.16), 35-44 years (0.37), 20-24 years (0.05), 25-34 years (0.39)	
Intelec / cient	fem. (1.00)	+ 65 years (0.03), 45-64 years (0.15), 35-44 years (0.23), 20-24 years (0.10), 25-34 years (0.49)	

# Funções de Similaridade: SODAS 2/

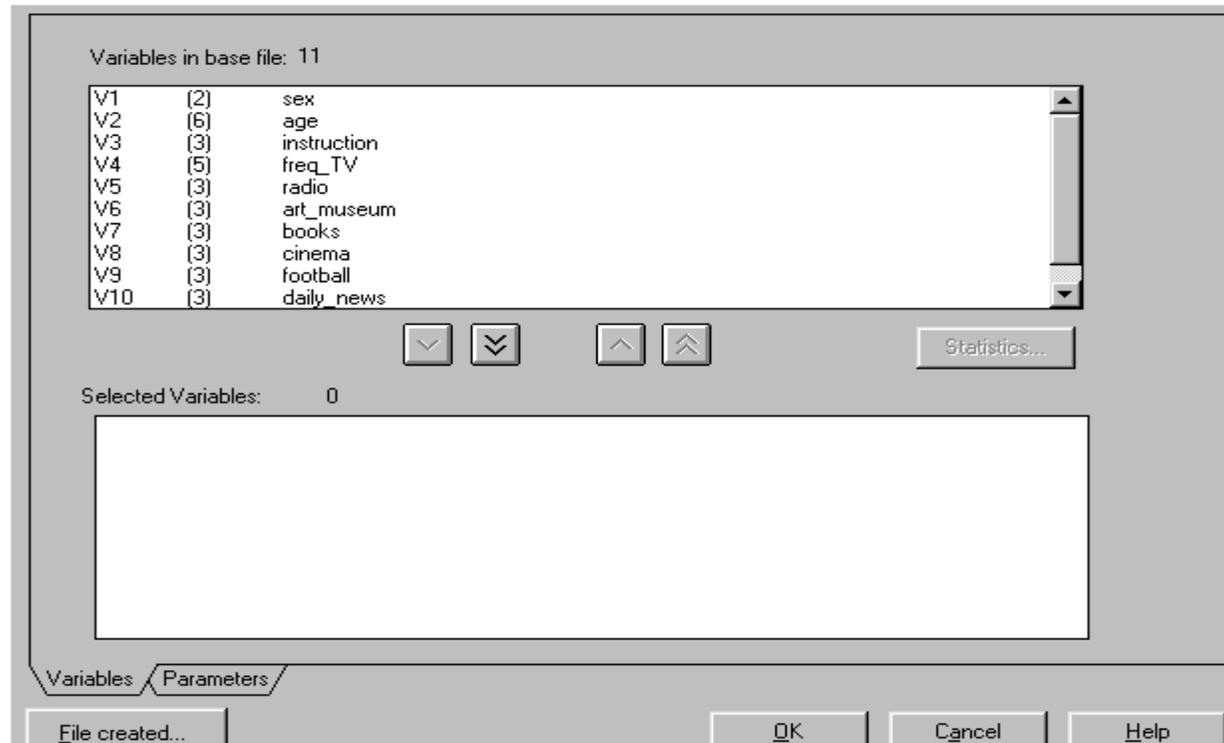
➔ Entrada: Comando DIM

The screenshot displays the SODAS 2.0 software interface. The main window is titled 'YYY.FIL' and contains a menu bar with 'Chaining', 'Model', 'Method', 'Window', and 'Help'. The main area shows a flow diagram with three boxes: 'BASE' (top), 'DIM' (middle), and 'END' (bottom), connected by vertical lines. A small box with the number '1' is positioned to the left of the 'DIM' box. Text labels include 'QUERY2.SDS' with the path 'c:\sodas\bases\', and 'DIM' with the description 'Dissimilarity similarity matrix'. On the left side, a 'Methods' panel is open, showing a list of 'Sodas procedures'. The 'DIM' procedure is highlighted in blue, with the description 'Dissimilarity similarity matrix'. Below this, a grid of other procedures is visible, including SOE, DIV, STAT, DKS, DI, PCM, FDA, TREE, DSD, SDT, and DIM, PYR.

Sodas procedures				
DIM				
Dissimilarity similarity matrix				
SOE	DIV	STAT	DKS	DI
PCM	FDA	TREE	DSD	SDT
DIM	PYR			

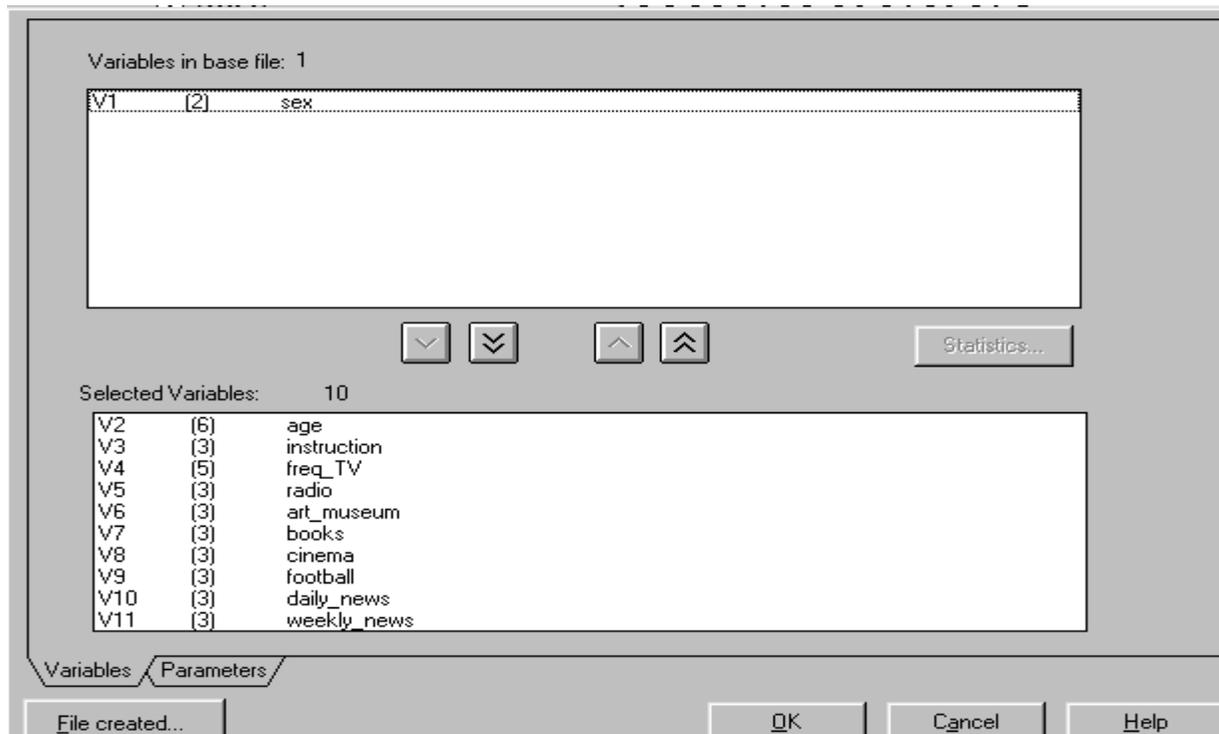
# Funções de Similaridade: SODAS 3/

➡ Entrada: Comando DIM: Seleção de Variáveis



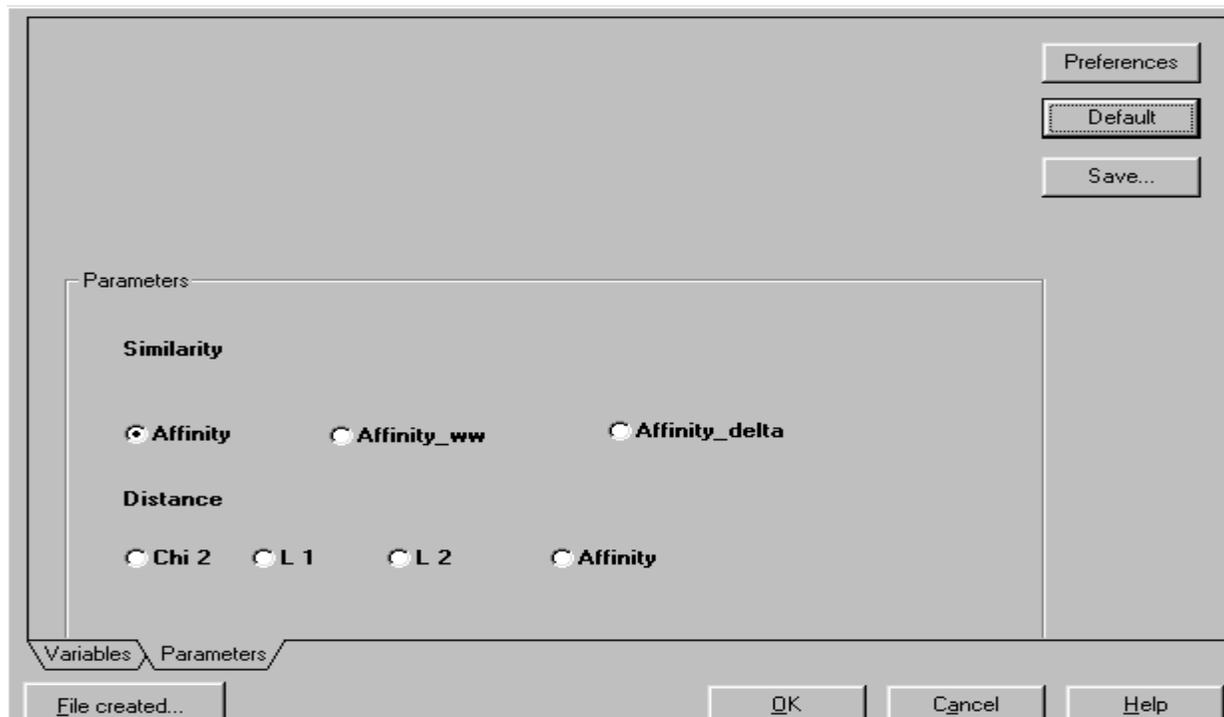
# Funções de Similaridade: SODAS 4/

➡ Entrada: Comando DIM: Seleção de Variáveis



# Funções de Similaridade: SODAS 5/

➡ Entrada: Comando DIM: Parâmetros



# Funções de Dissimilaridade: SODAS 6/

➡ Entrada: Comando DIM: Saída

The screenshot displays the SODAS 6 software interface. On the left, a 'Methods' panel shows a list of 'Sodas procedures' with 'DIM' selected and highlighted in cyan. Below this, a grid of other procedures is visible, including SOE, DIV, STAT, DKS, DI, PCM, FDA, TREE, DSD, SDT, and DIM, PYR. The main window, titled 'YYY.FIL', shows a flow diagram for the 'DIM' command. The diagram starts with a cyan arrow-shaped box labeled 'BASE' (with path c:\sodas\base.ch) pointing to a red box labeled 'DIM' (with description 'Dissimilarity similarity matrix'). A green box with the number '1' is positioned to the left of the 'DIM' box. A yellow document icon is to the right of the 'DIM' box. The flow ends with a cyan box labeled 'END'.

# Funções de Similaridade: SODAS 7/

➡ Entrada: Comando DI: Saída

```
-----  
Learning Set      :    21  
Number of variables :    10  
Criterion coding  :    1  Affinity Similarity  
  
GROUP OF SELECTED VARIABLES :  
(  2 ) age                6 MODALITIES  
(  3 ) instruction        3 MODALITIES  
(  4 ) freq_TV           5 MODALITIES  
(  5 ) radio              3 MODALITIES  
(  6 ) art_museum         3 MODALITIES  
(  7 ) books              3 MODALITIES  
(  8 ) cinema             3 MODALITIES  
(  9 ) football          3 MODALITIES  
( 10 ) daily_news         3 MODALITIES  
( 11 ) weekly_news       3 MODALITIES  
  
LIST OF SYMBOLIC OBJECTS IN THE SET :  
  1 : "Student / masc."      2 : "Retired / fem."  
  3 : "Employed / fem."     4 : "Small Indep. / fem."  
  5 : "Housewife / fem."    6 : "Medium Staff / masc."  
  7 : "Ind. Worker / masc." 8 : "Small Indep. / masc."  
  9 : "Intelec / cient. / masc." 10 : "Student / fem."  
 11 : "Other / fem."        12 : "Employed / masc."  
 13 : "Other / masc."      14 : "Dir. / Lib. / masc."
```

# Funções de Similaridade: SODAS 8/

➡ Entrada: Comando DIM: Saída

```
SIMILARITY MATRIX
 1 : "Student / masc."          1.000000
 2 : "Retired / fem."          0.782577  1.000000
 3 : "Employed / fem."         0.891505  0.922334  1.000000
 4 : "Small Indep. / fem."     0.862770  0.930632  0.987971  1.000000
 5 : "Housewife / fem."        0.802363  0.985166  0.951437  0.955792  1.000000
 6 : "Medium Staff / masc."    0.921931  0.833959  0.950396  0.947837  0.863181
 7 : "Ind. Worker / masc."     0.917204  0.884639  0.982102  0.968087  0.922634
 8 : "Small Indep. / masc."    0.905061  0.901688  0.973946  0.980620  0.928293
 9 : "Intelec / cient. / masc." 0.814324  0.760442  0.860654  0.875890  0.782832
10 : "Student / fem."          0.984010  0.782443  0.895466  0.860095  0.805666
11 : "Other / fem."            0.889596  0.923082  0.987144  0.971847  0.955262
12 : "Employed / masc."        0.939074  0.865882  0.974236  0.968329  0.896419
13 : "Other / masc."           0.909634  0.906785  0.978847  0.976565  0.930979
14 : "Dir. / Lib. / masc."     0.813245  0.757813  0.853252  0.868034  0.777105
15 : "Intelec / cient. / fem." 0.821716  0.762966  0.861623  0.873743  0.779076
16 : "Retired / masc."         0.820738  0.977044  0.932559  0.950770  0.973725
17 : "Medium Staff / fem."     0.875678  0.875123  0.966164  0.968390  0.900646
18 : "Dir. / Lib. / fem."      0.843952  0.811861  0.881206  0.883540  0.815157
19 : "Ind. Worker / fem."      0.891335  0.916147  0.990595  0.977322  0.951825
20 : "Manager / fem."          0.713527  0.774905  0.816550  0.827012  0.783618
21 : "Manager / masc."         0.859947  0.764354  0.867403  0.877792  0.791650

Average          : 0.900562
Standard Deviation : 0.072362
```