

COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Ana Lorga da Silva^{1,2,3}, Helena Bacelar-Nicolau^{2,3} e Gilbert Saporta⁴

¹Universidade Lusófona de Humanidades e Tecnologias,

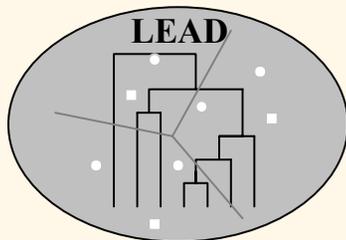
Departamento de Economia e Gestão

²Universidade de Lisboa, FPCE, Laboratório de Estatística e Análise de Dados

³Centro de Estatística e Aplicações da Universidade de Lisboa

Linha de Investigação de Análise de Dados Multivariados

⁴Departamento de Estatística, CNAM, Paris

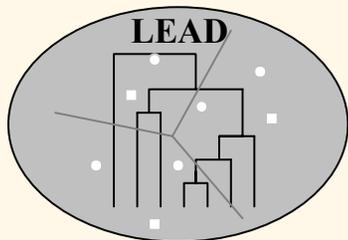


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

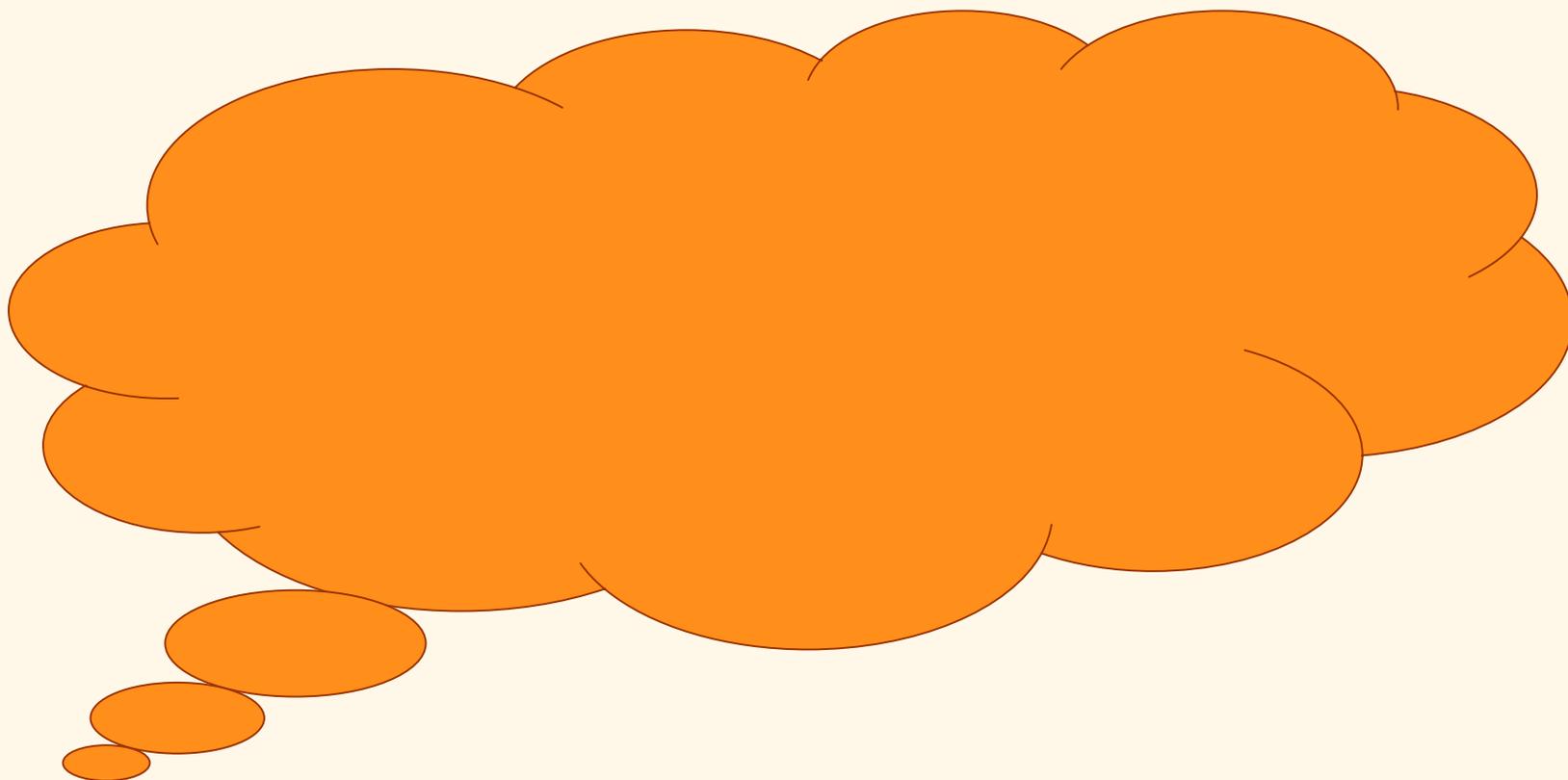
Estrutura do Trabalho:

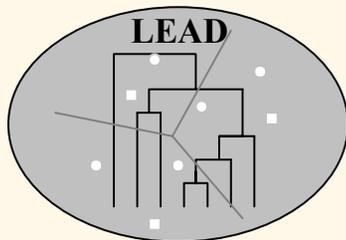




COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais





COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

MÉTODOS CLÁSSICOS

Coeficiente de Afinidade Básico

AL / AMP

SL / UIM

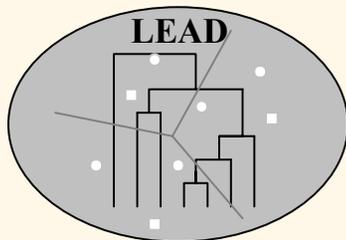
CL / USM

Coeficiente de Bravais-Pearson

AL / AMP

SL / UIM

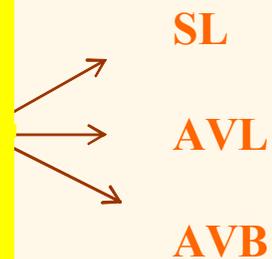
CL / USM

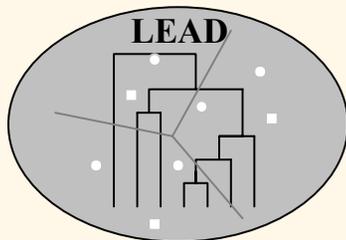


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

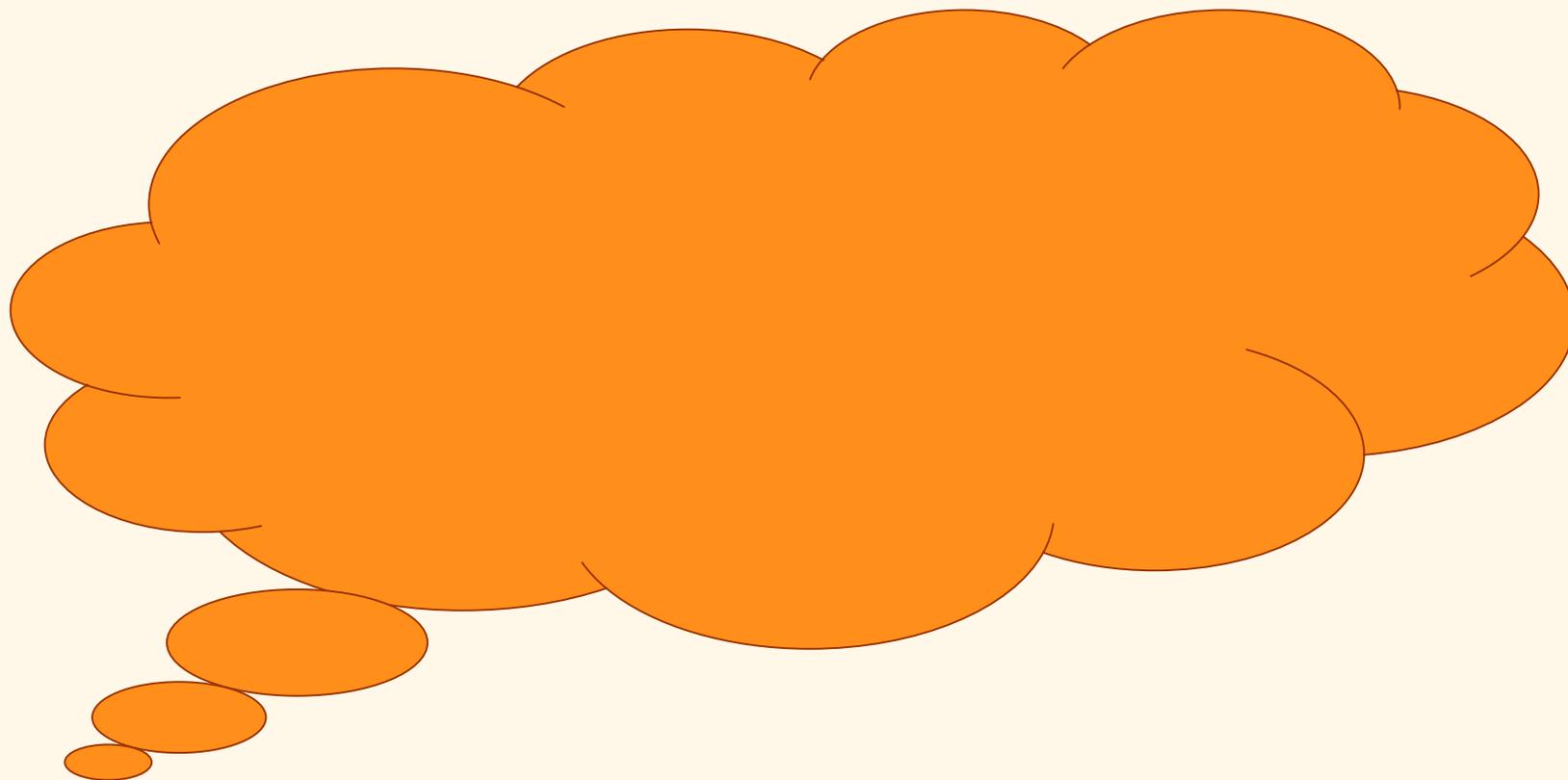
MÉTODOS PROBABILÍSTICOS

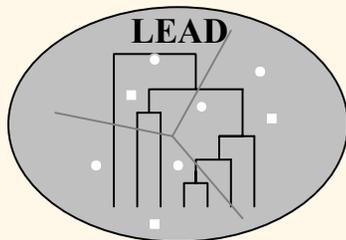




COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais





COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

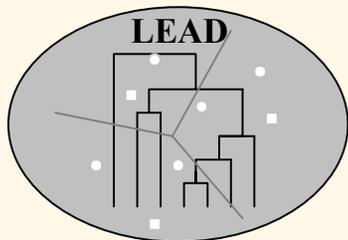
Tipo de Dados Omissos:

$$MAR: \text{Pr ob} (M | X_{obs}, X_{mis}) = \text{Pr ob} (M | X_{obs})$$

$$M = [M_{ij}]$$

é um indicador dos dados omissos,

$$M_{ij} = \begin{cases} 1, & \text{se } x_{ij} \text{ é observado} \\ 0, & \text{se } x_{ij} \text{ é omissos} \end{cases}$$

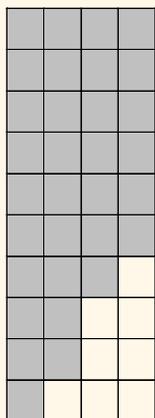


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Padrões de Dados Omissos

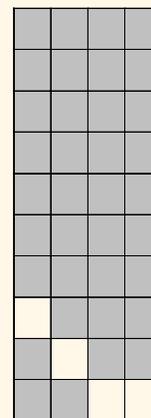
MONÓTONOS:

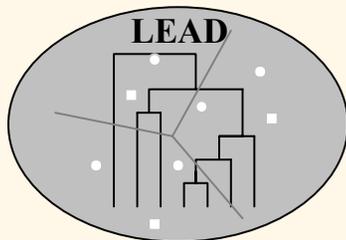


 - dado observado

 - dado omissos

NÃO MONÓTONOS:



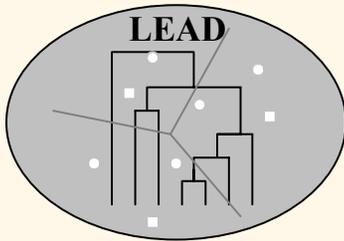


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Métodos de Imputação:



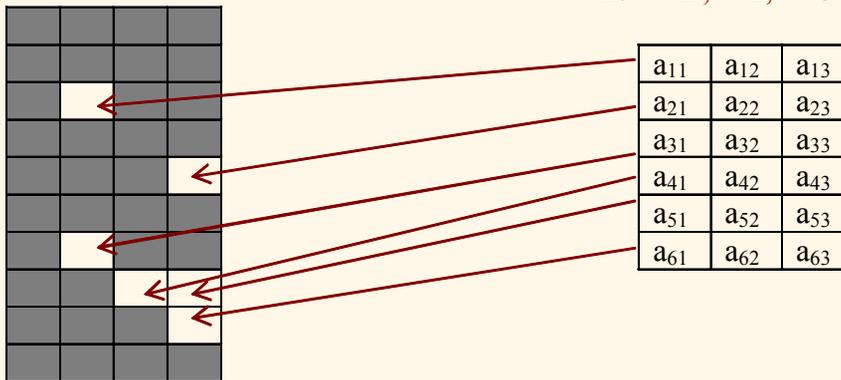


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

IMPUTAÇÃO MÚLTIPLA

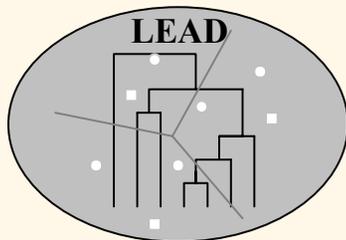
Imputações
 $m = 1, 2, 3$



	a_{11}		
			a_{21}
	a_{31}		
		a_{41}	a_{51}
			a_{61}

		a_{12}	
			a_{22}
		a_{32}	
		a_{42}	a_{52}
			a_{62}

		a_{13}	
			a_{23}
		a_{33}	
		a_{43}	a_{53}
			a_{63}



COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

IM VM

$$\bar{S} = \left(\sum_{k=1}^m S_k \right) / m$$

IM CONSENSO

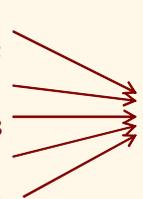
$X^1 \longrightarrow S_1$

$X^2 \longrightarrow S_2$

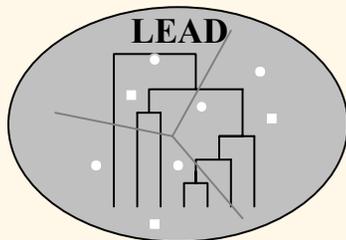
$X^3 \longrightarrow S_3$

$X^4 \longrightarrow S_4$

$X^5 \longrightarrow S_5$



$\bar{S} \Rightarrow U_s$ Dendrograma



COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

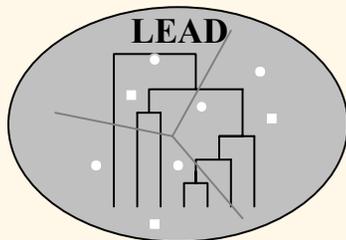
- I. **Dados Originalmente Completos**
- II. **Dados Incompletos**
- III. **Dados “completos” por Métodos de Imputação**

Coeficiente de Spearman (1%) / Comparar Ultramétricas



Comparar





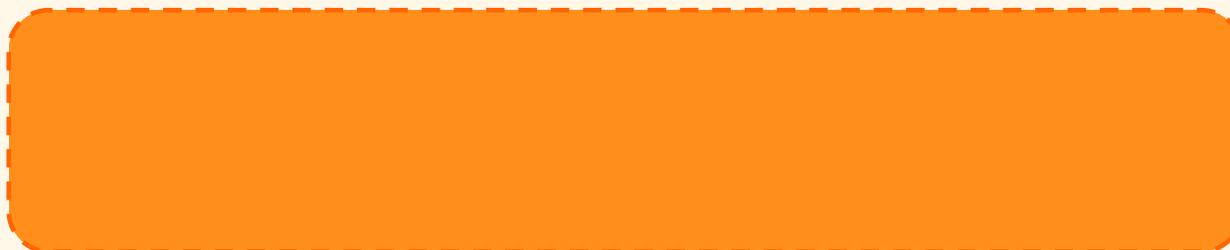
COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

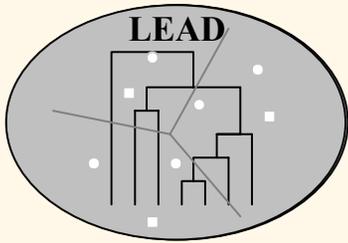
Dados Simulados

- ❑ Dados completos com distribuição Multinormal
- ❑ 1000 indivíduos, 5 variáveis

Estruturas:



- ❑ 10%, 15%, 20% (sobre o total) de Dados Omissos em duas Variáveis
- ❑ Dados Omissos, segundo a condição MAR

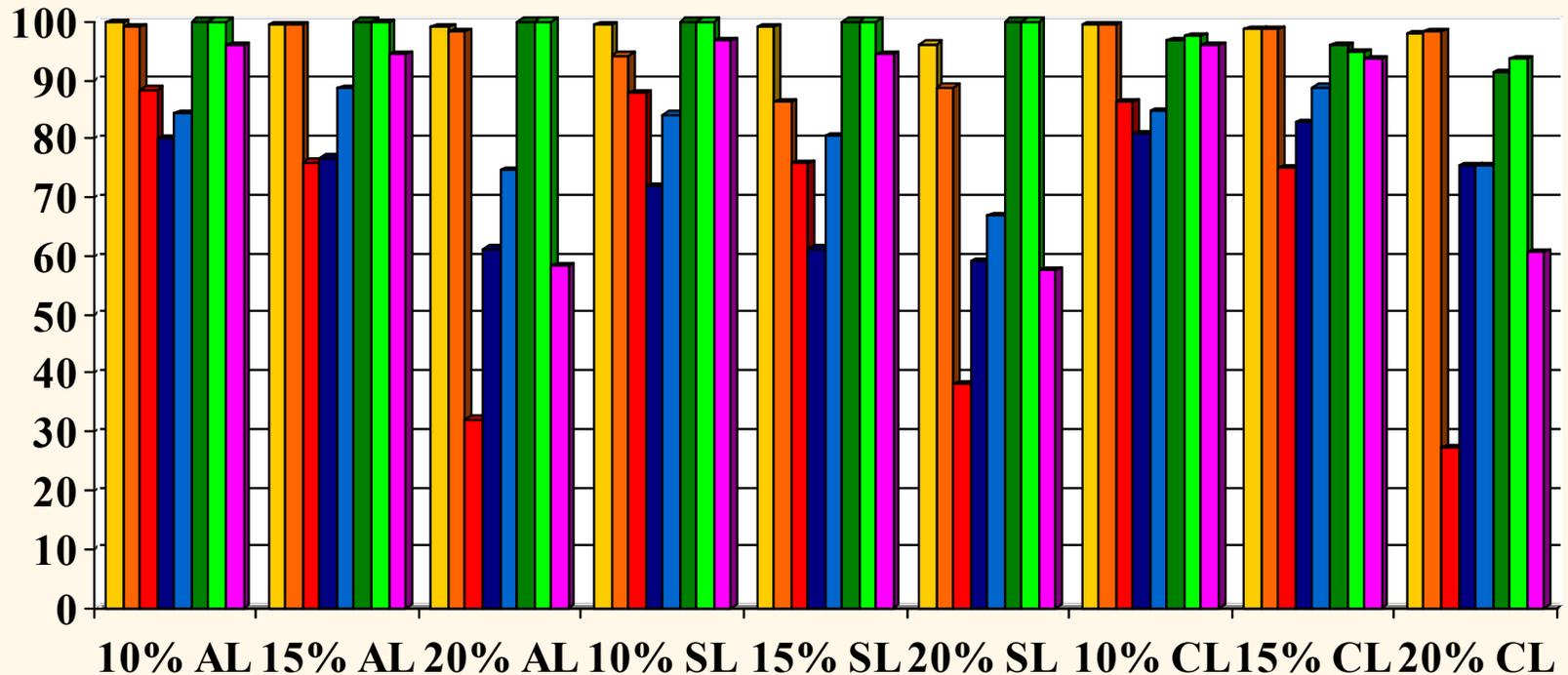


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

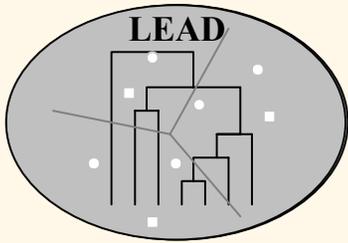
Uma Aplicação a Dados Reais

Coeficiente de Afinidade

$$r_s = 1$$



■ Listwise
 ■ Pairwise
 ■ EM
 ■ OLS
 ■ NIPALS
 ■ IMvm
 ■ IMcons
 ■ PLS2

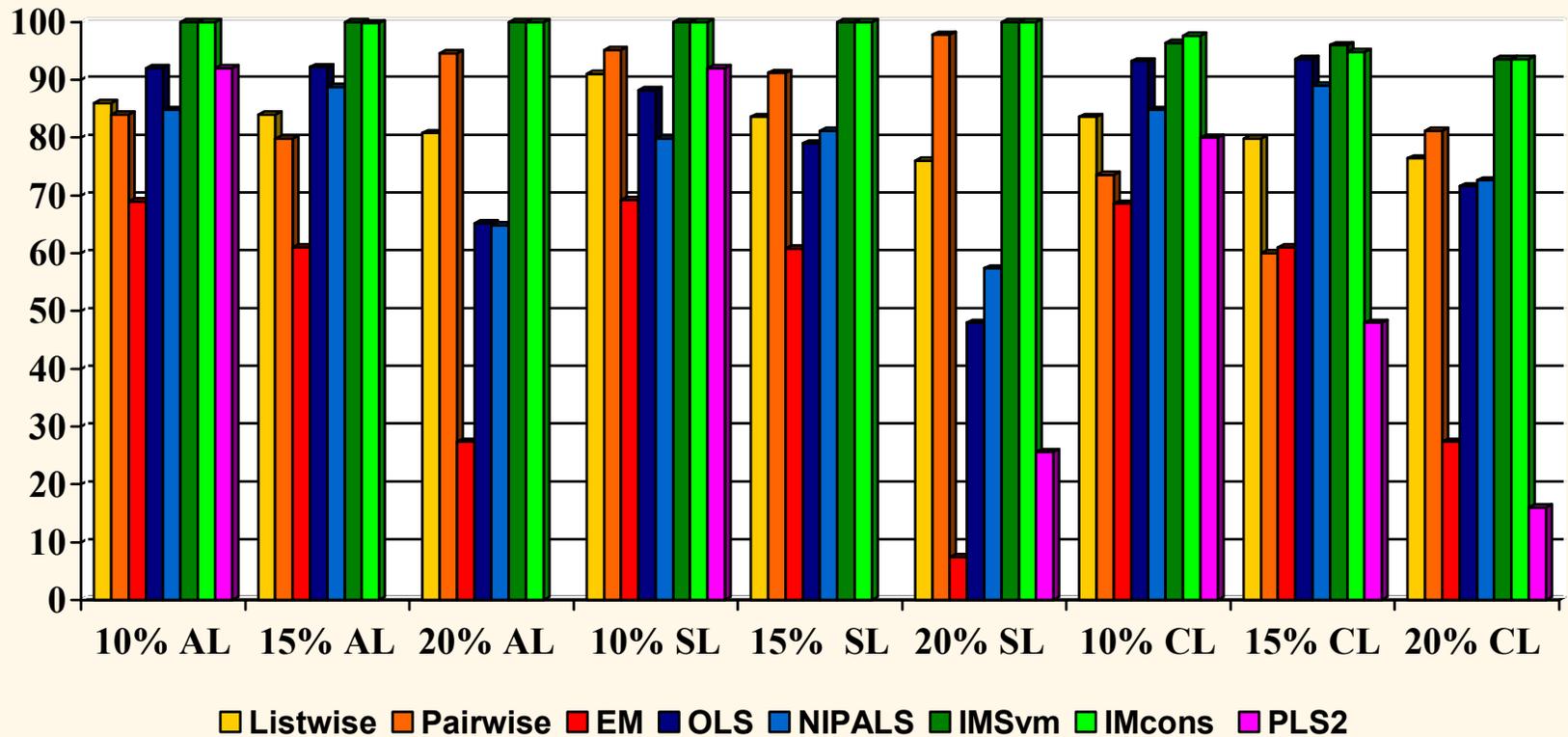


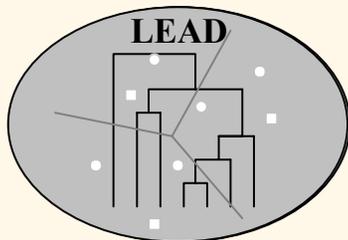
COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Coeficiente de Correlação de Pearson

$$r_s = 1$$



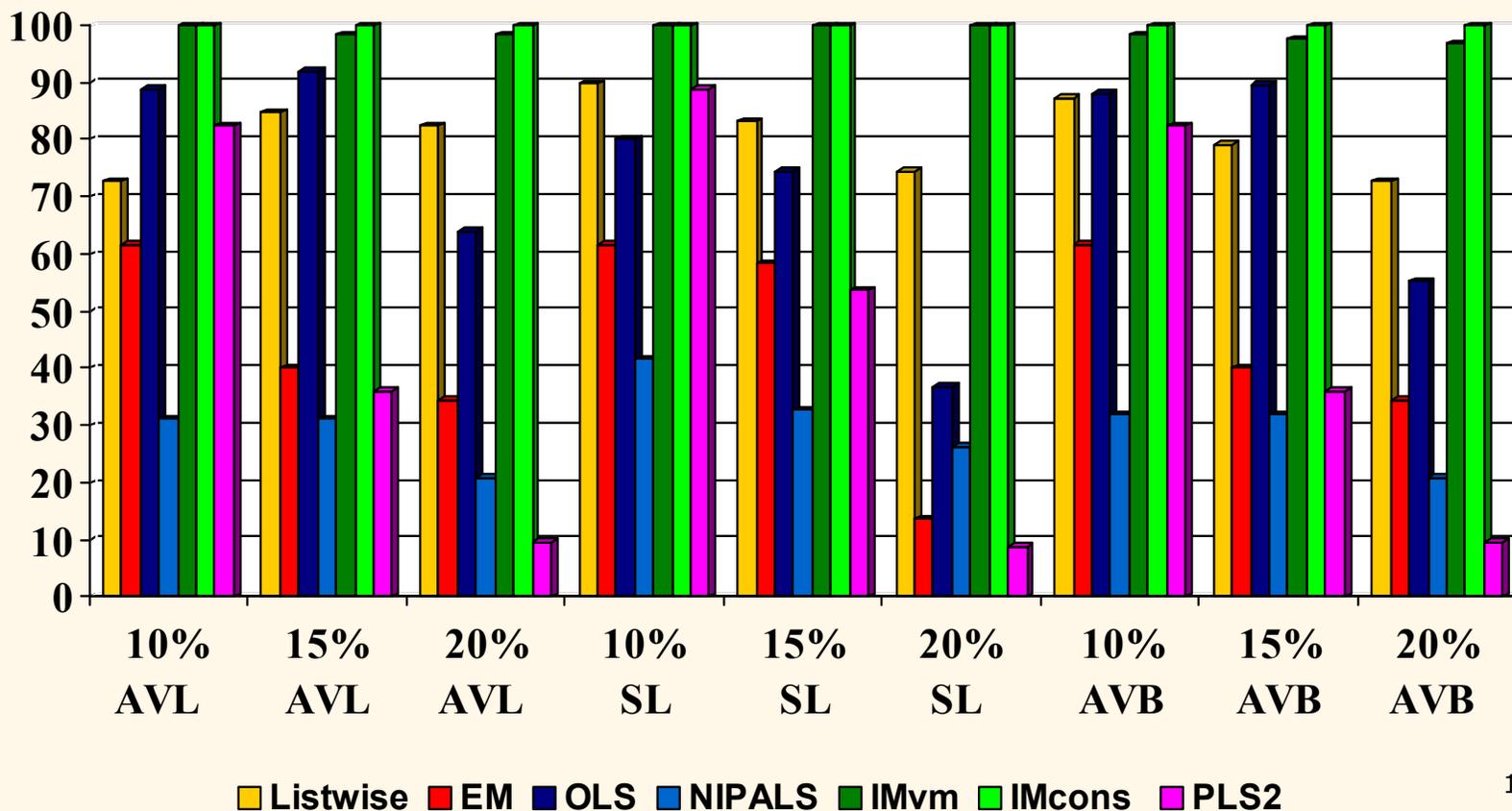


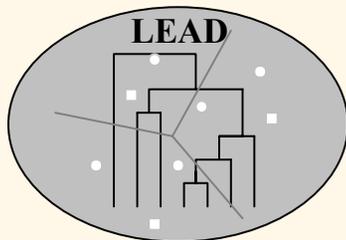
COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Coeficiente de Afinidade Centrado e Reduzido pelo Método WW

$$r_s = 1$$





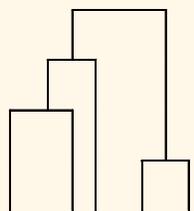
COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Dados Reais

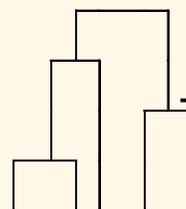
- ❑ Dados completos com distribuição Multinormal
- ❑ 181 indivíduos, 5 variáveis

Estruturas:



- B-P Mét. Clássicos
- C. A. cr WW Mét. Prob.

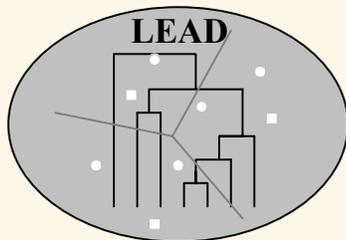
x3,x4,x5,x1,x2



- C. A. Mét. Clássicos

x3,x4,x5,x1,x2

- ❑ 10%, 15%, 20% (sobre o total) de Dados Omissos em duas Variáveis
- ❑ Dados Omissos, segundo a condição MAR



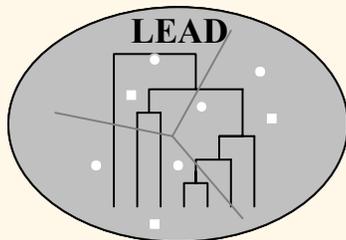
COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

Utilizam-se dados de um estudo efectuado com adolescentes Portugueses (Geada(1998)). Esse estudo pretende avaliar o efeito do clima familiar no desenvolvimento da personalidade e nos comportamentos “desviantes”, utilizando-o como prevenção do uso de drogas.

“Some longitudinal studies on deviant behaviours have shown that there is a logical continuity between instances of this type of in childhood, adolescence and adulthood” (Geada(1998))

A classificação hierárquica ascendente é uma metodologia que pode complementar o estudo efectuado.



COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

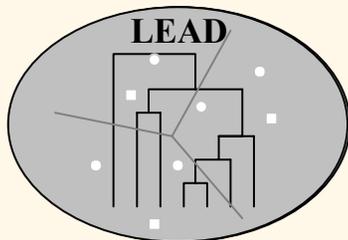
Variáveis:

- Transgressividade
- Delinquencia
- Autoconceito
- Aptidão para gerir o Stress
- Clima Familiar

GRUPOS (2):

{transgressividade, delinquência} /

{autoconceito, aptidão para gerir o stress, clima familiar}

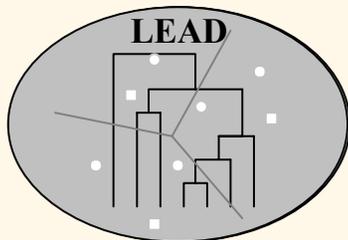


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

MÉTODOS CLÁSSICOS

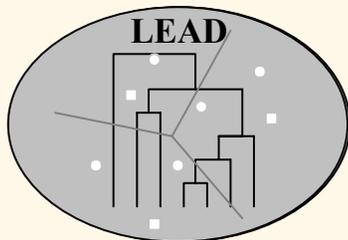
MD	métodos	Coeficiente de Afinidade			Coeficiente de Pearson		
		AL	SL	CL	AL	SL	CL
10%	listwise	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	pairwise	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s < r_s'$	$ r_s > r_s'$
	EM	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	OLS	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	IMcons	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	IMvm	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	PLS2	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$



COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

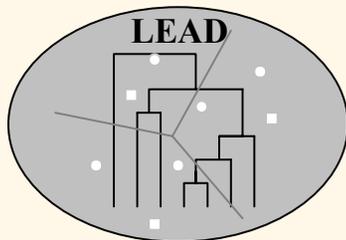
MD	métodos	Coeficiente de Afinidade			Coeficiente de Pearson		
		AL	SL	CL	AL	SL	CL
15%	listwise	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	pairwise	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	EM	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	OLS	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	IMcons	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	IMvm	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$
	PLS2	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$	$ r_s > r_s'$	$ r_s > r_s'$	$ r_s > r_s'$



COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

MD	métodos	Coeficiente de Afinidade			Coeficiente de Pearson		
		AL	SL	CL	AL	SL	CL
20%	listwise	$ r_s =1$	$ r_s =1$	$ r_s =1$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$
	pairwise	$ r_s =1$	$ r_s =1$	$ r_s =1$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$
	EM	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$
	OLS	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s =1$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$
	IMcons	$ r_s =1$	$ r_s >r_s'$	$ r_s =1$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$
	IMvm	$ r_s =1$	$ r_s =1$	$ r_s =1$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$
	PLS2	$ r_s =1$	$ r_s =1$	$ r_s =1$	$ r_s >r_s'$	$ r_s >r_s'$	$ r_s >r_s'$

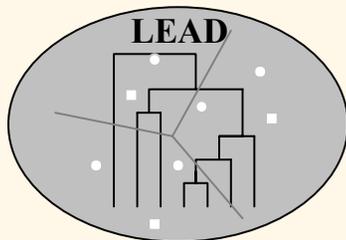


COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

MÉTODOS PROBABILÍSTICOS

MD	métodos	C. A. C.R. WW							
		AVL	SL	AVB					
10%	listwise	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$					
	EM	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$					
	OLS	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$					
	NIPALS	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$					
	IMcons	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$					
	IMvm	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$					
	PLS2	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$					
15%	listwise	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$	20%	listwise	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$
	EM	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$		EM	$ r_s < r_s'$	$ r_s < r_s'$	$ r_s < r_s'$
	OLS	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$		OLS	$ r_s < r_s'$	$ r_s < r_s'$	$ r_s < r_s'$
	NIPALS	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$		NIPALS	$ r_s < r_s'$	$ r_s < r_s'$	$ r_s < r_s'$
	IMcons	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$		IMcons	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$
	IMvm	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$		IMvm	$ r_s > r_s'$	$ r_s = 1$	$ r_s > r_s'$
	PLS2	$ r_s < r_s'$	$ r_s = 1$	$ r_s < r_s'$		PLS2	$ r_s = 1$	$ r_s = 1$	$ r_s = 1$



COEFICIENTE DE AFINIDADE EM CLASSIFICAÇÃO COM DADOS OMISSOS

Uma Aplicação a Dados Reais

CONCLUSÕES:

- O **Coeficiente de Afinidade Básico** revela-se mais robusto que o coeficiente de Pearson
- O ACHA, **SL** é muito estável em relação ao CAcr, pelo método **WW**
- O método de IM vm / consenso parece ser globalmente o melhor método de imputação
- Os resultados obtidos com os dados reais não contrariam os “*resultados esperados*”.