

## ECI 2015

### T3: Symbolic Data Analysis: Taking Variability in Data into Account

1. A análise de dados simbólicos levou à introdução de novos tipos de variáveis. Como se definem (formalmente) esses novos tipos de variáveis? Dê exemplos de cada caso.
2. Foram propostos diferentes métodos para regressão de variáveis intervalares. Indique os métodos estudados no curso e explique sucintamente como funcionam (em que se baseiam).
3. Efectuou-se uma classificação descendente de grupos de países (por nível de riqueza), tendo-se obtido a árvore abaixo representada. Diga qual a composição das classes formadas e como se podem descrever.

NÍVEL DE RIQUEZA	PIB/Habitante	Exportações (% PIB)	Importações (% PIB)	Investimento Directo Estrangeiro	Inflação	Crescimento da População (%)
Fraco	[180 , 1120]	[10.14 , 65.48]	[16.56 , 88.35]	[-45500100 , 108249968]	[-2.73 , 59.89]	[-0.81 , 3.10]
Medianamente Fraco	[20 , 7200]	[15.98 , 103.79]	[17.86 , 110.66]	[-437000 , 13285999616]	[-5.12 , 185.03]	[-0.52 , 3.39]
Muito Fraco	[100 , 710]	[8.31 , 92.30]	[16.05 , 84.65]	[0 , 1698300032]	[0.65 , 408.02]	[0.23 , 3.34]
Medianamente Elevado	[2920 , 9250]	[10.81 , 125.49]	[11.45 , 113.53]	[3700 ,32778999808]	[4.99 , 27.63]	[-1.59 , 2.57]
Elevado: Membros da OCDE	[8980 , 38140]	[9.98 , 88.05]	[8.44 , 84.655]	[1449000 , 133973999616]	[-1.10 , 16.27]	[0.08 , 1.25]
Elevado: Não Membros da OCDE	[10070 , 24740]	[38.99 , 179.91]	[31.18 , 161.45]	[2300 , 6390329856]	[1.71 , 25.58]	[0.13 , 3.62]

```

+---- Classe 1 (Nb_SIM=1)
!
!----2- [Crescimento_população_(%) <= 0.817500]
!
!      +---- Classe 3 (Nb_SIM =3)
!      !
!      !----3- [Investimentos Directos estrangeiros_($) <= 4918973693.000000]
!      !
!      +---- Classe 4 (Nb_NAO=1)
!
!----1- [PIB/habitante_($) <= 20477.500000]
!
+---- Classe 2 (Nb_NAO=1)

```

4. Considere um conjunto de 3 pessoas, A, B, C, para as quais se registou a quantia dispendida em várias compras, e o tipo de cartão usado para pagamento :

A [1000; 15000] {Electron, Visa, Mastercard}  
 B [2000; 2500] {Electron}  
 C [2000; 7000] {Visa, Mastercard}

- a. Qual a descrição (em intensão) do par {A,B} ? E do par {B,C} ?
- b. Algum destes corresponde a um conceito ? Justifique.

5. Pretende-se analisar um conjunto de localidades descritas por três variáveis intervalares, amplitude térmica em Janeiro e Julho, variação anual de pluviosidade. Efectuou-se uma análise em componentes principais, usando o método dos vértices. Considere o caso de New-York (JFK) : [24:7; 38:8] [66:7; 82:9] [2:70; 4:13].
- Como se representa esta observação a partir dos vértices ?
  - Suponha que na 1ª componente principal se obtiveram as seguintes coordenadas para os vértices : 1.2 , 0.6 , 2.1 , 0.9 , 2.5 , 1.4 , 0.5 , 1.8.  
Qual a representação (intervalar) de New-York (JFK) na 1ª componente principal ?
6. Como se define o modelo paramétrico apresentado para variáveis intervalares ?
- Que alternativas são consideradas ?
  - Quais as vantagens deste modelo ?
  - Que limitação tem ?