

# Seminário de Estatística e Investigação Operacional

## SEMINÁRIO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

### Análise de Dados Simbólicos Representar e Analisar Dados com Variabilidade

Paula Brito



Faculdade de Economia / LIAAD - INESC TEC  
Universidade do Porto



Paula Brito

Faculdade de Economia / LIAAD – INESC TEC Universidade do Porto

14 janeiro 2015 | 14:30 | Edifício das Ciências Florestais  
(Departamento de Matemática) Sala F2.8

Resumo

A Análise de Dados Simbólicos, introduzida por E. Diday no fim dos anos oitenta do século vinte, tem como objective analisar dados que apresentem variabilidade de forma intrínseca, a qual deverá ser explicitamente considerada. Na Estatística clássica e na Análise Multivariada de Dados, os elementos a analisar são geralmente entidades individuais, para as quais um valor único é registado para cada variável – e.g., indivíduos, descritos pela sua idade, salário, nível de instrução, estado civil, etc.; automóveis descritos pelo seu peso, comprimento, potência, cilindrada, etc.; estudantes, para cada um dos quais as notas a diferentes disciplinas foram registadas. Mas quando os elementos de interesse são classes ou grupos de algum tipo – os cidadãos que vivem em determinadas cidades; equipas, compostas por jogadores individuais; modelos de automóveis, em vez de veículos específicos; turmas e não estudantes

individuais – então há variabilidade inerente aos dados. Reduzir esta variabilidade por medidas de tendência central – valores médios, medianas ou modas – conduz obviamente a uma importante perda de informação.

A Análise de Dados Simbólicos propõe uma estrutura que permite representar dados com variabilidade, usando novos tipos de variáveis. Foram também desenvolvidos métodos que tomam adequadamente em conta essa variabilidade. Dados simbólicos podem ser representados na forma matricial usual, onde cada entidade é representada numa linha e cada coluna corresponde a uma variável – mas agora os elementos de cada célula não são em geral simples valores reais ou categorias como no caso clássico, mas conjuntos finitos, intervalos, ou, mais geralmente, distribuições.

Nesta apresentação iremos introduzir a motivar a área da Análise de Dados Simbólicos, e apresentar com algum detalhe os novos tipos de variáveis que foram definidos para representar a variabilidade, ilustrando com alguns exemplos. Discutiremos ainda algumas questões que se levantam ao analisar dados que não se enquadram no modelo clássico, e apresentaremos modelos de representação para alguns tipos de variáveis. Se o tempo permitir, alguns métodos de análise de dados simbólicos serão referidos.

VER