Symbolic Data Analysis : The Field

Paula Brito

Fac. Economia & LIAAD-INESC TEC, Universidade do Porto
mpbrito@fep.up.pt
http://www.fep.up.pt/docentes/mpbrito/

ECI 2015 - Buenos Aires
T3: Symbolic Data Analysis:
Taking Variability in Data into Account

Outline

Outline

1 **General Summary**

2 Symbolic Data Analysis: the field

3 Software and References

Concluding remarks

- From micro-data to macro-data:
- Interval, Distribution-valued data
- Take variability into account
- Several methodologies already developed
  that do take into account the variability of the data
- New problems / challenges for the 21st century:
  intervals and distributions are not real numbers !

Concluding remarks

"Distributions are the numbers of the future"

(Schweizer, 1984)

Outline

1. General Summary

2. Symbolic Data Analysis: the field

3. Software and References

Symbolic Data Analysis

Symbolic Data Analysis was introduced by Prof. Edwin Diday in the late eighties of the 20th century, in a seminal paper :

"The symbolic approach in clustering and relating methods of data analysis: The basic choices."
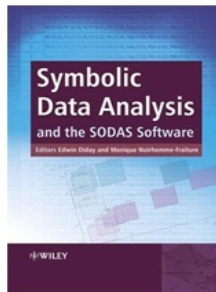In : Proc. 1st IFCS Conference (Aachen, Germany, 1987), pp. 673-684, 1988.
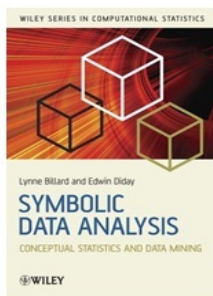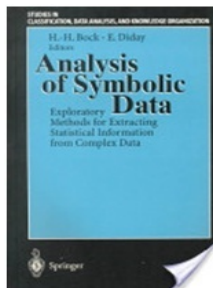
Symbolic Data Analysis

In these past (almost) 30 years, this research field has known a great development, with

- Three books published
- Many papers in journals and conference proceedings
- Three journal's Special Issues
- Workshops, tutorials and special sessions at conferences all over the world
- Two European projects, gathering teams from several different countries
- A large number of PhD and Master thesis

Symbolic Data Analysis

Books :

Symbolic Data Analysis

Nowadays :

- Established Workshop series
  Next : Orléans (France) Nov. 2015 :
  http://www.univ-orleans.fr/MAPMO/colloques/sda2015/
- Increasing number of publications in a wide variety of journals
- Journals' special issues
- Website (wiki):
  http://vladowiki.fmf.uni-lj.si/doku.php?id=sda
- Linkedin SDA Group : for discussions, announcements...

The SDA community is growing world-wide !

Outline

1. General Summary

2. Symbolic Data Analysis: the field

3. Software and References

Available Software

- SODAS: free, registration required -
  http://www.info.fundp.ac.be/asso/sodaslink.htm
- Syrokko: commercial software - http://syrokko.com/
- R packages available in CRAN (increasing...), so far mainly for
  interval data :
  RSDA ; iRegression, MAINT.Data

Books and Main Papers

Bock, H.-H.; Diday, E. (2000): *Analysis of Symbolic Data: Exploratory methods for extracting statistical information from complex data.* Berlin-Heidelberg: Springer-Verlag.

Billard, L., Diday, E. (2007): *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Wiley.

Diday, E., Noirhomme-Fraiture, M. (2008): *Symbolic Data Analysis and the SODAS Software*. Wiley.

Special Issue of *Statistical Analysis and Data Mining*. Vol. 4, Issue 2, April 2011. Wiley, on behalf of the American Statistical Association.

Billard, L. and Diday, E. (2003). From the statistics of data to the statistics of knowledge: Symbolic Data Analysis. *Journal of the American Statistical Association*, 98 (462), pp. 470487.

Noirhomme-Fraiture, M. and Brito, P. (2011). Far beyond the classical data models: Symbolic data analysis. *Statistical Analysis and Data Mining*, 4(2), 157–170.

Brito, P. (2014). Symbolic Data Analysis: another look at the interaction of Data Mining and Statistics. WIREs Data Mining and Knowledge Discovery, Volume 4, Issue 4, July/August 2014, 281-295.