How a New Statistical Infrastructure Induced a New Computing Trend in Data Analysis

A. Ciok [1)], T. Kowalczyk[1)], E. Pleszczynska[1,2)]

[1)] Institute of Computer Science, Polish Academy of Sciences 01-237 Warsaw, Ordona 21, Poland
[2)] eple@ipipan.waw.pl

Abstract:

A statistical infrastructure based on the concentration measures is presented. It aims at being a consistent system of descriptive parameters with a clear interpretation, summarizing knowledge on distributions of variables in populations of objects. The system requires unified formulas for sets of mixed continuous-categorical variables. These demands are met e.g. by the formulas given in Sec. 3 for the bivariate dependence measures, presented as suitably weighted averages of concentration indices for pairs of conditional distributions. Further, in Sec. 4, a concentration oriented modification of the statistical procedure called correspondence analysis is mentioned and exemplified by applying it to data from the Polish Parliament Elections in 1993 and 1997. Sec. 5 contains remarks on creating an inference and computing system adjusted to this new concentration - based approach to statistical descriptive parameters.