

Analyzing European Social Survey data using symbolic data methods and *Syrokko* software

Filipe Afonso*, Seppo Laaksonen**

*SYROKKO, Aéroport, Paris - Roissy CDG Cedex, France
afonso@syrokko.com

**University of Helsinki, Finland
Seppo.Laaksonen@helsinki.fi

Abstract. The paper presents an application of Symbolic Data Analysis (SDA) with SYR software of Syrokko company to the fifth round of the European Social Survey (ESS) carried out among European inhabitants. In this study, we are not interested in studying the people themselves but by the comparison of different European countries, or different regions of Europe (Western Europe, Eastern, Northern, Southern), or some groups of inhabitants by age, gender, region, etc. Here, we study, however, the 52 European countries by age groups. We describe each of them by all the results of its inhabitants using symbolic data. Symbolic Data Analysis (SDA) is proving so useful to aggregate up micro data (at the level of the inhabitants) to higher level units called concepts (the countries or European regions, for instance), using symbolic bar-chart or interval-valued variables. Using this aggregation we lose less information than occurs in classic analysis because symbolic data allow keeping the variation within the concepts. It allows keeping the variation of the results at the level of the inhabitants when they are aggregated up to their country. Hence, this could be called as “smart aggregation”.

1. Introduction

The SODAS software was developed in two EU research projects, that is, SODAS and ASSO. The newest version of the software is from 2004. The book, edited by Diday and Noirhomme (2008) was much based on this development. Since this first software for SDA such technology has been considered to be highly appreciated by data miners. As far as data mining with big data are concerned, the symbolic approach basically includes the two steps, one for aggregating smartly big micro data sets, and then for analyzing such aggregated data. This paper follows this approach so that our big micro data are the multinational survey data, but the aggregation is made by the new symbolic data software, SYR.

Our paper is next organized so that the principles of the SYR have been explained in Section 2. Our big data are the fifth round of the European Social Survey (ESS), collected late 2010 and the first half of 2011. The ESS is an academically-driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. In annex, the details of our initial data and principles of construction are given. SYR, naturally, has performed the aggregation itself. The remaining sections present examples on SYR techniques to analyze the aggregated