Modernization of the Statistical Production by Multivariate Statistical Methodology

Maurizio Vichi
Sapienza University of Rome, Dpt. Statistical Sciences, Rome, Italy – Maurizio.vichi@uniroma1.it

The statistics theory and methodologies for analysing data are changing in the last forty years under the strong pressure of two revolutions: technological innovations and the big data deluge. Computers and internet have deeply modified the theory, in the past dominated by the scarcity of the data (small samples) and by a limited computational power. The global connection has produced abundance of data from different sources even if unstructured and not ready for a statistical analysis. The velocity of the data production and the high computational power allow the production of new methods that consent quick decisions.
These changes have to be considered as new opportunities and challenges for the production of official statistics. The first priority regards the modernization of the data production with the integration between traditional data collection and the new data sources. This corresponds to pass from the manual data collection to the automatic one, integrated with different administrative data sources; thus, limiting the use of the manual collection which is generally highly expensive. Now computers allow to propose more complex multivariate data analysis such as data mining, knowledge discovery and statistical learning. Thus, the second modernization of the official statistics production has to regard the use of new methodologies; thus moving the focus from univariate and bivariate statistics, to multivariate statistical analysis, that allow to explore all the multidimensionality and complexity of the phenomena analyzed by the official statistics. The data communication and visualization represents the third priority for the modernization of the statistics, to allow the use of the statistics to a wider community of users.

Keywords: Modernization; multivariate statistics; data collection and integration; data communication and visualization.