Some experiments on Statistical Matching in the R environment

Marcello D'Orazio, Marco Di Zio, Mauro Scanu
(madorazi@istat.it, dizio@istat.it, scanu@istat.it)

ISTAT, Via Cesare Balbo 16, 00184 Roma, Italy

Key Words: Data Analysis, Data Processing.

In the last years, interest on Statistical Matching problems has increased (Rässler, 2002, D'Orazio *et al.* 2006). This is due to the large amount of data-sets available and, at the same time, to the need of timely and not costly information. Statistical Matching techniques aim at combining information from different sources. In particular, it is assumed that the two sources (e.g. two samples) do not observe the same set of units, so that neither merging nor record linkage techniques can be applied.

In order to explore the properties of matching techniques and therefore apply them to real data problems, a series of matching experiments have been carried out in the R environment. R has already been used for the definition of some statistical matching algorithms (Rässler 2002). In D'Orazio *et al* (2006) more algorithms have been translated in R. The codes will be available on the web page `http://www.wiley.com/go/matching`.

At first an extensive simulation study has been carried out taking into account two separate cases: (i) all the variables are continuous and (ii) all the variables are categorical. In the presence of continuous variables most of the experiments aim at the evaluation of the performances of matching techniques based on regression methods. As far as categorical variables are considered, we worked in the direction of exploring the uncertainty of the results typical of a statistical matching application. We show how the usage of some basic auxiliary information, in the form of logical constraints involving values of different variables, can reduce the uncertainty. An application of statistical matching to real data is also presented. We tried to estimate the Social Accounting Matrix by means of the fusion of the Household Balance Survey conducted by the Bank of Italy and the Household Expenditure Survey conducted by the Italian National Statistical Institute. A Social Accounting Matrix is a system of statistical information containing economic and social variables in a matrix formatted data framework. The matrix includes economic indicators such as per capita income and economic growth.

References

D'Orazio M, Di Zio M, Scanu M (2006) *Statistical Matching: theory and practice*, Wiley.

Rässler S (2002). *Statistical Matching: a frequentist theory, practical applications and alternative Bayesian applications*, Springer.