



European Conference on Quality in Official Statistics

Rome, 8-11 July 2008

Training course

STATISTICAL MATCHING: STATISTICAL METHODOLOGIES FOR MATCHING DATA FROM TWO SAMPLE SURVEYS

Instructors

Marcello D'Orazio, Italian National Statistical Institute, Istat
Marco Di Zio, Italian National Statistical Institute, Istat
Mauro Scanu, Italian National Statistical Institute, Istat

Objectives

1. Description of matching two sample surveys as a statistical problem
2. Characteristics of estimators in the statistical matching framework, needed hypotheses and accuracy evaluation
3. Examples from real life

Course Design

There are more statistical data produced today than ever before. These data are analyzed and cross-referenced for several purposes. However, many data sets have not any shared element. They are harder to combine, and therefore meaningful inference can be hardly achieved. Statistical matching is just the art of combining information from different sample surveys that contain no common unit. The course is designed for statistical methodologists from national statistical institutes and for researchers that may improve their analyses by means of the joint use of two sample surveys. Possible areas of application are, among others, joint analysis of variables available in the system of multipurpose surveys, and joint analyses on household income and expenditures. Participants' background should include maximum likelihood methods as well as basic concepts of regression analysis and of analysis of contingency tables.

The course will rely on likelihood based inference as well as on nonparametric methods.

The description of the main problems and issues concerning statistical matching will be practically discussed by means of a case study analysis. In this part of the course, a special attention will be paid to show and use software codes developed for statistical matching in the R environment. The course will also provide a critical view

of “the statistical matching problem” highlighting limits of the available statistical methodologies, it will provide as well an overview on open problems.

Course Text and Materials

D’Orazio M., Di Zio M., Scanu M. (2006) *Statistical matching: theory and practice*, John Wiley & Sons, Chichester

Program:

1. Definition of statistical matching: differences with respect to other data integration techniques and similarities with imputation problems
2. Definition of statistical matching objectives: micro and macro
3. Statistical matching under conditional independence models
4. Statistical matching methods with auxiliary information
5. Some remarks on uncertainty
6. Statistical matching in the context of sampling from finite populations
7. Accuracy of statistical matching results
8. Choice of matching variables
9. A case study: the construction of the Social Accounting Matrix