

STATISTICAL INTEGRATION OF DATA: RECORD LINKAGE AND STATISTICAL MATCHING

JUNE 17 - 18, 2010

Marco Di Zio and Mauro Scanu (Istat)

Waldhotel Doldenhorn, Kandersteg

Lecturers

Marco Di Zio and Mauro Scanu both have a PhD in Methodological Statistics (University of Roma "La Sapienza") and are researchers at the Italian National Statistical Institute – Istat. They are co-authors of a book on statistical matching (Statistical Matching: Theory and Practice – Wiley). M. Scanu is the coordinator and M. Di Zio a member of a Eurostat co-funded project (ESSNet on Statistical Methodology: area integration of surveys and administrative data).



Marco di Zio

Their common main research interests are statistical methodologies for data integration (record linkage and statistical matching) and non response. M. Di Zio is also interested in mixture models, data editing and imputation and M. Scanu is interested in survey methodology, graphical models, non parametric statistics, and assessment of the official statistics user needs.



Mauro Scanu

Aims and objectives

There is more statistical data produced in today's modern society than ever before. This data is analyzed and cross-referenced for innumerable reasons, for instance in the case of National Institutes of Statistics the joint analysis of two or more statistical and administrative sources is a result of a rational organization of all available informative sources and, among all, it allows the reduction of survey costs and the response burden. However, many data sets are sometimes hard to combine: errors in the record identifiers or lack of record identifiers may jeopardize any meaningful integrated use of the data sets.

The combination of different surveys or of surveys with administrative data needs to be analyzed by appropriate statistical methodologies. Broadly speaking, two main procedures can be considered:

- Record linkage: integration of different data sets composed of the same units. Complete records at unit level are obtained linking records of two or more data sets with appropriate unit identifiers. We deal with the case when the link of unit is not certain, i.e., it is probabilistic, for instance when there are errors in unit identifiers.
- Statistical matching: combination of different data sets with no unit in common, but with some variables observed in both the data sets, in this case the integration procedures should exploit the information coming from the common variables. We remark that for statistical matching, integration can be either at unit level (for instance, creation of complete – synthetic records), or at macro level (for instance, direct inferences on some target parameters that describe the joint relationship of variables not jointly observed in the data sets).

This course aims at providing the state-of-the-art of the methodologies for record linkage (first day) and statistical matching (second day). The theory will be complemented with practical examples referring to real cases and presentation of software tools.

Location

The course will be held in the Waldhotel Doldenhorn in Kandersteg, see www.doldenhorn-ruedihus.ch. The hotel offers a free shuttle service to the train station.

Date and hour

Thursday, June 17 till Friday June 18, 2010.
Course starts at 10 a.m. and ends at 5 p.m.

Course fee

- CHF 1050.-- for members of the Swiss Statistical Society, other applicants CHF 1250.--.
- Reduced course fee CHF 700.-- for students. Please send a copy of the certification.
A limited number of course places are reserved for students.

The course fee includes all meals and accommodation in single rooms.

The number of participants is limited to 20 with a minimum of 12.

Registration deadline: April 1, 2010.

Registration and further information

Swiss Statistical Society
Sabine Probst
Bergacher 8
CH-3253 Schnottwil
+41 (0)32 353 70 94

sabine.probst@stat.ch

The Swiss Statistical Society (SSS) reserves the right to cancel a course up to 14 days prior to the course due to insufficient enrollement. The SSS is not liable for any participants' expenses due to the cancellation of any booked courses. Payment of the course registration fee is required prior to the start of the course. Cancellations received in writing more than 30 days before the start of the course will be refunded 100% of the course fee. Cancellations received between 30 and 14 days prior to the course will be refunded 50% of the course fee. The SSS regrets that no refunds are allowed for cancellations received within 14 days of the course start date.