A Methodological and Computational Framework for Statistical Disclosure Control

Matthias Templ, Alexander Kowarik and Bernhard Meind[125]

Abstract

Data privacy/confidentiality is one of the core businesses in official statistics. SDC becomes more and more important in the last years because of an increase of the awareness on data privacy and because of the fact that more and more data are provided to the public or to researchers.

Before providing microdata to the public or to researchers confidentiality/privacy has to be respected. Only data with low disclosure risk can be disseminated. Confidentiality can be achieved and disclosure risk can be estimated with statistical disclosure control (SDC) methods. This contribution will give an outline of problems and possibilities in the area of statistical disclosure control on microdata.

Existing anaonymisation and disclosure risk methods are briefly reviewed and summarized and modern efficient tools are presented. The discussed methods are then applied on popular large real-world data set. The application of few selected anonymisation methods leads to well-protected anonymised data with high data utility and low information loss. Moreover, the first time it is possible to also anonymise large data sets with millions of observations with efficient ready-to-use user-friendly software tools.

Key words:

statistical disclosure control, data utility, disclosure risk, R

[25] Matthias Templ, Austrian Statistical Society & Statistics Austria & Vienna University of Technology & Palacky University Olomouc & data-analysis OG, Austria; email: matthias.templ@gmail.com

Italian Presidency of the Council of the European Union

eurostat

ESAC European Statistics Advisory Committee

Sapienza Università di Roma
Department of Statistical Sciences

SIS

Istat Italian National Institute of Statistics

CONFERENCE OF EUROPEAN STATISTICS STAKEHOLDERS

Methodologists, Producers and Users of European Statistics

Rome 24 | 25 November 2014