Working Paper No. 37 (Summary)
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (vi): Software tools for statistical disclosure control

SAFE – A METHOD FOR STATISTICAL DISCLOSURE LIMITATION OF MICRODATA

Contributed paper

Submitted by the Berlin State Statistical Office, Germany[1]

[1] Prepared by Jörg Höhne (j.hoehne@statistik-berlin.de).

SAFE - a method for statistical disclosure limitation of microdata

Jörg Höhne
Berlin State Statistical Office

Extended abstract of supporting paper

During the last few years, a method for statistical disclosure limitation of microdata was developed and implemented at the Berlin State Statistical Office. The method is known under the name SAFE (Appel, G.C. 1994). The projected paper will explain the principles of the method and present some test results. The primary aim of developing this method was to have a tool for disclosure protection of multiple linked multi-dimensional tables. However, as it works on the microdata level, the method is also suitable for the protection of microdata.

The paper will contain the following sections:

1. Which problems of the statistical disclosure control should / can be solved by the method?

   To ensure disclosure protection, the method creates a safe microdatafile in such a way that the following problems of statistical confidentiality are solved:
   - Any tabulations of the safe microdata file must not contain any sensitive cells: cells with less then 3 contributions, or cells that would allow for inferential disclosure of individual respondent contributions because of dominance problems.
   - Exact or inferential disclosure of individual respondent data on basis of differences between cell values should be impossible (secondary disclosure control).
   - Consistency of varying tabulations obtained from the same safe file.
   - the information loss should be minimized

2. What is the main idea of the method and which steps are taken?

   The main idea of the method is, to build a microdatafile, which consists of records only where three ore more records are identical. SAFE reduces the risk that an intruder trying to link records of the safe file to an external file is able to obtain matches, because there are at least three identical records for each combination. In addition to that, records are distorted because the values were constructed as a mean of the original values. Because all tabulations build on the microdatafile, the results are consistent.
   The basic idea (creating groups of identical records) is equal to that of microaggregation. The difference to microaggregation lies in the principle used for grouping the data. Microaggregation methods group ‘nearest’ records, according to some measure of distance. SAFE, on the other hand, constructs groups considering the impact on multidimensional distributions.

3. Short description of the method

   SAFE consists of the following steps:
   1. create a safe solution for the discrete variables
   2. match the original continuos variables to this solution
   3. find a solution without predominating units
   4. optimise the solution while keeping the confidentiality
   5. build groups of three or more records while setting values to the mean.

   The paper will explain these steps in detail.

4. Tests

   Recently, we have tested the method using data of the "monthly statistic of manufacturing and mining" ("Monatsbericht im Bergbau und Verarbeitenden Gewerbe"). The paper will give an overview of the results.

5. Outlook

The SAFE method is one of the methods considered in the German national research project "Factual Anonymisation of Business Microdata". Within this project, we compare the SAFE method to common alternative methods for statistical disclosure limitation of microdata. Some results of this study will be presented.

References:

Appel, G.C. (1994), 'Anonymization of Microdata, 1st Practical Experience with the SAFE-Programme', proceedings of the Second International Seminar on Statistical Confidentiality, Luxemburg, 1994