PRELIMINARY PROGRAM:
Workshop on Statiscial Disclosure Control for Data Confidentiality

10-12 November 2004, Madison, Wisconsin

Overview

Statistical Disclosure Control is an important topic in the production process of statistical offices. Modern production processes in these offices are capable of producing large statistical databases and detailed tables. This is very useful for the users of the statistical information like policy makers and researchers. Nowadays the users of statistical information have the capacity of handling large amounts of data and can perform complex analyses on their own computers. However, the other side of the coin is that there is a real risk of breaking the privacy of the respondents. And safeguarding the privacy of the respondents is vital for Statistical Offices. Failure to do so will be not keeping the promises made when we collected the data but also will lead to a suspension of the cooperation in the participation of our surveys and censuses. This awareness has a.o. lead to several national and international initiatives to research this topic and also to develop software like ARGUS to implement the results of this research. The state of the art is now that the results of this research are available and that the ARGUS software has been released to implement the theory of SDC.

The course will cover a state of the art overview of disclosure control covering both microdata protection and the protection of tabular data, the two major forms of output of a statistical office. Both for tabular data as well as for micro data we will show where the risks of disclosure are, which methods are available to measure the risks of disclosure and which methods can be applied to protect the data against disclosure. The course will be a combination of teaching sessions, where the theory will be explained, practical exercises, where the participants will work with (small) examples, which will show them the disclosure risks and make them more aware of the problems and sessions to train them in the use of the ARGUS software, both μ-ARGUS and τ-ARGUS.

The target audience will be the people actually working on these surveys as well as the survey managers. It is in our vision essential that also managers should have a profound knowledge of the issues of SDC.

The course leaders will provide handouts of all the material used during the lectures as well as the manuals for the ARGUS software.

Location
The location of the course will be the University of Wisconsin – Madison.

Results
As a result of this course the participants will have a good knowledge of the state-of-the-art in SDC and know how to implement this in their practical work, a.o. by applying the ARGUS software. The course will also provide a basis for continued study of the topic of SDC through further reading of SDC literature.

Preliminary Schedule
The schedule on the following page is preliminary. We will begin with a light breakfast each morning, and then will start with presentations promptly at 9:00 AM. The computer lab will be available to participants until 6:00 PM each day, but formal presentations will end at about 5:00 PM. We will end at 4:00 PM on Friday. Theoretical presentations will be interspersed with demonstrations, discussion, and practical exercises. Participants are encouraged to bring along their own data, a brief (1-2 page) description of a study with which they are working, and to prepare questions and problems. A portion of Friday's program will be devoted to a discussion of participants' work and to the ways in which SDC methods and the ARGUS software might be used to estimate disclosure risk and to protect data confidentiality.

Preliminary Schedule

Location: The Pyle Center on the University of Wisconsin – Madison campus.
Conference Hotel: Howard Johnson Plaza Hotel, in downtown Madison.
Lectures: A.J. Hundepool, E. Schulte Nordholt and P.P. de Wolf (all Statistics Netherlands)

Wednesday, 10 November
Light Breakfast
Introduction of course, lecturers, participants
General introduction to SDC
Break
Theory/methods of SDC concerning microdata (General)
Break
Theory/methods of SDC concerning microdata (Methods)
Lunch
Theory/methods of SDC concerning microdata (Methods) (continued)
Break
Exercises microdata I
Break
Demonstration and exercises with μ-ARGUS

Thursday, 11 November
Light Breakfast
Exercises microdata II
Break
Theory/methods of SDC concerning tabular data (General)
Break
Theory/methods of SDC concerning tabular data (Methods)
Lunch
Theory/methods of SDC concerning tabular data (Methods) (continued)
Break
Exercises tabular data

Friday, 12 November
Light Breakfast
Demonstration and exercises with τ-ARGUS
Break
Legal Issues
Break
Onsite facilities
Lunch
User cases studies
Break
Evaluation and conclusion

About the lecturers

The lectures in this course will be Anco Hundepool, Eric Schulte Nordholt and Peter-Paul de Wolf, all of Statistics Netherlands.

Anco Hundepool studied mathematics at Leyden University and subsequently he joined Statistics Netherlands. He started his career in the department for Statistical Methods. His main interests were seasonal adjustment, compilation of price index series and a pilot study on purchasing power statistics. After that he was involved in the development of the Blaise system, became a project-leader for the Abacus tabulation package and the STATview dissemination package. Within the SDC-project, partially funded by the 4th Framework he was the project-leader for the development of the ARGUS package for statistical disclosure control. In the TADEQ project for the documentation of electronic questionnaires (also partially funded by the 4th Framework) he was responsible for the development of the TADEQ-software. In the 5th Framework CASC project he is the overall project leader and continues the development of the ARGUS software. Anco Hundepool presented his work on various conferences and has published in different refereed journals.

Eric Schulte Nordholt studied mathematics at the University of Utrecht and econometrics at the Erasmus University Rotterdam. He started his career as a researcher at the Department of Statistical Methods of Statistics Netherlands. In 1995 he worked as a detached national expert at the European Community Household Panel (ECHP) team of Eurostat in Luxembourg. Subsequently, he became senior researcher at the Department of Employees of the Division of Socio-economic Statistics. Since 2000 he is a senior researcher and project leader in the division of Social and Spatial Statistics. He is also Statistics Netherlands' advisor on the statistical disclosure control of social data. Recently he was project leader of the Dutch Virtual Census of 2001 and together with his team he wrote a book about the analysis and methodology of this Virtual Census. His main interests are the analysis of panel data, census data, data editing and imputation, and statistical disclosure control. Eric Schulte Nordholt presented his work on various conferences and has published in different refereed journals.

Peter-Paul de Wolf studied mathematics at the Technical University of Delft (1986 – 1991). He started his PhD in mathematical statistics in 1991, at the same university, on extreme value estimation. In 1999, already working for Statistics Netherlands, he successfully defended his PhD.
In 1996 he joined the research department of Statistics Netherlands and worked on several subjects: sample survey design, editing and imputation, lifetime of capital stock, statistical disclosure control. In the recent years he has specialized in the field of statistical disclosure control and at present he co-ordinates the research in that area at the Research Department of Statistics Netherlands. Peter-Paul de Wolf presented his work on various conferences and has published in different refereed journals.