

Preface

Inference control in statistical databases (also known as statistical disclosure control, statistical disclosure limitation, or statistical confidentiality) is about finding tradeoffs to the tension between the increasing societal demand for accurate statistical data and the legal and ethical obligation to protect the privacy of individuals and enterprises which are the source of data for producing statistics. To put it bluntly, statistical agencies cannot expect to collect accurate information from individual or corporate respondents unless these feel the privacy of their responses is guaranteed.

This state-of-the-art survey covers some of the most recent work in the field of inference control in statistical databases. This topic is no longer (and probably never was) a purely statistical or operations-research issue, but is gradually entering the arena of knowledge management and artificial intelligence. To the extent that techniques used by intruders to make inferences compromising privacy increasingly draw on data mining and record linkage, inference control tends to become an integral part of computer science.

Articles in this book are revised versions of a few papers selected among those presented at the seminar “Statistical Disclosure Control: From Theory to Practice” held in Luxemburg on 13 and 14 December 2001 under the sponsorship of EUROSTAT and the European Union 5th FP project “AMRADS” (IST-2000-26125).

The book starts with an overview article which goes through the remaining 17 articles. These cover inference control for aggregate statistical data released in tabular form, inference control for microdata files, software developments, and user case studies. The article authors and myself hope that this collective work will be a reference point to both academics and official statisticians who wish to keep abreast with the latest advances in this very dynamic field.

The help of the following experts in discussing and reviewing the selected papers is gratefully acknowledged:

- Lawrence H. Cox (U. S. National Center for Health Statistics)
- Gerd Ronning (Universität Tübingen)
- Philip M. Steel (U. S. Bureau of the Census)
- William E. Winkler (U. S. Bureau of the Census)

As an organizer of the seminar from which articles in this book have evolved, I wish to emphasize that such a seminar would not have taken place without the sponsorship of EUROSTAT and the AMRADS project as well as the help and encouragement by Deo Ramprakash (AMRADS coordinator), Photis Nanopoulos, Harald Sonnberger, and John King (all from EUROSTAT). Finally, the inputs by Anco Hundepool (Statistics Netherlands and co-ordinator of the EU 5th FP project “CASC”) and Francesc Sebé (Universitat Rovira i Virgili) were crucial to the success of the seminar and the book, respectively. I apologize for possible omissions.