

Statistical agencies collect individually identifiable data, process them, and publish statistical summaries (tables). During this process, the agencies are required to protect individually identifiable data through a variety of policies. In all cases, the scope is to provide the data users with useful statistical information, and to assure that the responses from the individuals are protected.

To this end, and due to the size of the data, combinatorial problems appear and require algorithmic approaches to find optimal or near-optimal solutions. This talk summarizes and compares the most common statistical disclosure control methods to minimize information loss while keeping small the disclosure risk from different data snoopers. A common definition of protection is first introduced.

Later the methods are described to find protected tables in accordance with the given definition. Two integer linear programming models described in the literature for the cell suppression methodology are extended to work also for the controlled rounding methodology. In addition, two relaxed variants are presented using two associated linear programming models, called partial cell suppression and partial controlled rounding, respectively.

A final discussion shows how to combine the four methods and how to implement a cutting-plane approach for the exact and heuristic resolution of the combinatorial problems in practice. The methods are in a free-and-open-source software called tau-ARGUS.

For more details we refer the reader to the book "Statistical Confidentiality: Principles and Practice", by George Duncan, Mark Elliot and Juan-Jose Salazar-Gonzalez, Springer 2011.