

M

Matrix Masking

STEPHEN E. FIENBERG¹, JIASHUN JIN^{1,2}

¹Department of Statistics, Carnegie Mellon University,
Pittsburgh, PA, USA

²Department of Statistics, Purdue University,
West Lafayette, IN, USA

Synonyms

Adding noise; Data perturbation; Recodings; Sampling; Synthetic data

Definition

Matrix Masking refers to a class of statistical disclosure limitation (SDL) methods used to protect confidentiality of statistical data, transforming an $n \times p$ (cases by variables) data matrix Z through pre- and post-multiplication and the possible addition of noise.

Main Text

Duncan and Pearson [3] and many others subsequently categorize the methodology used for SDL in terms of transformations of an $n \times p$ (cases by variables) data matrix Z of the form

$$Z \rightarrow AZB + C, \quad (1)$$

where A is a matrix that operates on the n cases, B is a matrix that operates on the p variables, and C is a matrix that adds perturbations or noise.

Matrix masking includes a wide variety of standard approaches to SDL: (i) adding noise, i.e., the C in matrix masking transformation of equation [1]; (ii) releasing a subset of observations (delete rows from Z), i.e., sampling; (iii) cell suppression for cross-classifications;

(iv) including simulated data (add rows to Z); (v) releasing a subset of variables (delete columns from Z); (vi) switching selected column values for pairs of rows (data swapping). Even when one has applied a mask to a data set, the possibilities of both identity and attribute disclosure remain, although the risks may be substantially diminished.

The entry on *Statistical Disclosure Limitation For Data Access* focuses on four different matrix masking methods (i) sampling; (ii) recodings (e.g., collapsing rows or columns, sometimes referred to as global recoding; (iii) perturbation (including adding noise); and (iv) the use of synthetic data.

Cross-references

- Individually Identifiable Data
- Inference Control in Statistical Databases
- Privacy
- Randomization Methods to Ensure Data Privacy
- Statistical Disclosure Limitation for Data Access

Recommended Reading

1. Doyle P., Lane J.I., Theeuwes J.J.M., and Zayatz L. (eds.). Confidentiality, Disclosure and Data Access: Theory and Practical Application for Statistical Agencies. Elsevier, New York, 2001.
2. Duncan G.T., Jabine T.B., and De Wolf V.A. (eds.). Private Lives and Public Policies. Report of the Committee on National Statistics' Panel on Confidentiality and Data Access. National Academy Press, WA, USA, 1993.
3. Duncan G.T. and Pearson R.B. Enhancing access to microdata while protecting confidentiality: prospects for the future (with discussion). *Stat. Sci.*, 6:219–239, 1991.
4. Federal Committee on Statistical Methodology. Report on statistical disclosure limitation methodology. Statistical Policy Working Paper 22. U.S. Office of Management and Budget, WA, USA, 1994.