UNECE 2013 Work session on statistical data confidentiality

Ottawa, Ontario

October 28-30, 2013

Panel Discussion

Transparency Issues for Tabular Data

Lawrence H. Cox (USA)

National Institute of Statistical Sciences

cox@niss.org

SUMMARY

Statistical disclosure limitation (SDL) is necessary to protect the confidentiality of information pertaining to subjects of surveys and other statistical inquiries. SDL methods typically remove, modify or replace original data, thus distorting data, total mean square error and, potentially, inference and analytical outputs such as regression coefficients. *Transparency* in SDL refers to equipping analysts with information about the SDL method and its effects on statistical analysis--provided that this information does not lead to inferential disclosure. Thus, for example, if the SDL introduces bias, inflates variance or attenuates estimates such as for regression coefficients, then the analyst may be provided with estimates of bias, inflation or deflation for the purpose of correcting computed estimates. Most practitioners would agree that transparency is a good thing, but like truth and beauty, is in the eye of the beholder and difficult to achieve, assess and preserve. This is true for tabular data subject to suppression as these data involve complicated relationships between true and alternative values that in some situations may be deconstructed or even undone. We illustrate transparency problems for tabular data using examples.

Complementary Cell Suppression: Transparency Issues

- transparent SDL would involve revealing
  - # disclosure rule
  - # aspects of suppression rule/algorithm
- which can
  - # reduce number of alternative tables
  - # reduce number of alternative values
  - # thereby erode/threaten data security

Example 1: Sensitivity rules and data loss criteria

D_{11} (1)	18	D_{13} (6)	**25**
13	D_{22} (5)	**D_{23} (2)**	**20**
D_{31} (4)	**D_{32} (1)**	10	**15**
18	**24**	**18**	**60**

Table: 4 sensitive (bold), 6 suppressed cells

- true values of suppressions in parentheses
- **disclosure rule**: sensitive cell = 1, 2, 3, or 4
- **suppression rule**:
  - * minimize number of cells suppressed (or total value suppressed)
  - * preserve zero-cells
- 4 sensitive (**bold**), 2 complementary cells
- this pattern optimal for both number of cells (6) and total value (19) suppressed

Refresher: circuits

+/-	0	-/+
0	**-/+**	**+/-**
-/+	+/-	0

Interpretation

- 0-4 units can be *moved* in the **+** direction thru D_{11}
- 1 unit can be moved in the **–** direction
- 6 alternative values for D_{11}: D_{11} = 0,1,.., 5

Consequences

- 6 alternative tables (including original)
- intruder can construct the alternative tables

D_{11} (1)	18	D_{13} (6)	**25**
13	D_{22} (5)	**D_{23} (2)**	**20**
D_{31} (4)	**D_{32} (1)**	10	**15**
18	**24**	**18**	**60**

D_{11} = 1 (Original table and released pattern)

D_{11} (3)	18	D_{13} (4)	**25**
13	D_{22} (3)	**D_{23} (4)**	**20**
D_{31} (2)	**D_{32} (3)**	10	**15**
18	**24**	**18**	**60**

D_{11} = 3 (Alternative table and release pattern)

- all suppressions in D_{11} = 3 table are sensitive
- D_{11} = 1 and D_{11} = 3 optimal patterns are identical
- intruder cannot rule out D_{11} = 3 as true table

D_{11} (1)	18	D_{13} (6)	**25**
13	D_{22} (5)	**D_{23} (2)**	**20**
D_{31} (4)	**D_{32} (1)**	10	**15**
18	**24**	**18**	**60**

D_{11} = 1 (Original table and released pattern)

0	18	7	**25**
D_{21} (13)	6	**D_{23} (1)**	**20**
D_{31} (5)	0	**D_{33} (10)**	**15**
18	**24**	**18**	**60**

D_{11} = 0 (Alternative table and pattern)

- only 1 sensitive cell in alternative table
- only 4 suppressions in optimal pattern
- D_{11} = 0 pattern **differs from** released pattern

Consequences

- if intruder knows suppression rule, can apply rule to D_{11} = 0 table and *rule out* D_{11} = 0 table as original
- intruder can test all alternative tables in this manner

Example 2: Symmetry in sensitivity and suppression

- symmetric disclosure rules such as p/q-ambiguity lead to feasible intervals for suppressed cells that are centered on true values
- complementary suppression is based on creating algebraic circuits between suppressed entries
    - # circuits are based on “flows” of a single value that might be closely estimated
    - # circuits are vulnerable to symmetries

Examples: potential transparent releases

- controlled tabular adjustment: identify adjusted cells
- perturbation: details on perturbation mechanism
- analysis: NSO reanalyzes and reports on conformity