Disclosure Control for Register Based Frequency Tables

Mats Bergdahl-Kercoff, Statistics Sweden

Head of Methodology, Social Statistics and Analysis

November 13 2023

mats.bergdahl.kercoff@scb.se

SCB · Sveriges officiella statistik

Responsibility for Statistical Secrecy

- It is the responsibility of the Statistical Agency to determine likely disclosure scenarios and principles for damage assessment with considerations and decisions according to the delegation order.

- **Individual staff should not have to make assessments of probabilities for disclosure and consequences without proper planning and routines in place.**

SCB

Main Protection Methods for Statistical Tables

Methods that do not alter data

- Aggregation

- Suppression

Methods that do alter data (perturbative methods)

- Rounding

- **Application of Noice**

SCB

Cell Key Method – CKM

Background to Decision to use CKM

- SCB applies a number of different methods for disclosure control in the register-based population statistics that is mainly presented in the form of frequency tables. The same statistics published in different products can have been treated differently for disclosure control purposes, which in practice facilitates disclosure. For a long time a common standard has been requested.

- The CKM method means that noice is applied to the published data in a controlled and automatic way for tables of frequency based on complete registers.

- The method is implemented in the Labour Market Status of the Population (BAS) product and will be used in the Census 2021.

- It has been established as a standard at SCB.

- Implementation will be made over time based on prerequisities and priority in the statistical areas.

SCB

Cell Key Method in Summary

- Noice is applied as the tables are created
  - Adds uncertainty to the data consciously, controlled and minimally
- Each individual receives a <u>Permanent Random Number (PRN)</u> in the Population Register
- A lookup table is created that determines the noice applied to each cell in the original table (noice can be 0)
- The lookup table is <u>created once </u>and is used whenever a table is created.
- The noice is <u>applied separately for each cell </u>and independently between cells.
- Cells in the protected table consists of the original value + noice.

SCB

Implications of using CKM 1

- The possibility to have a common standard for all statistics at SCB
- A consciously accepted risk of disclosure – individual assessments are not necessary
- Facilitates publication of detailed information (<u>from a disclosure perspective</u>)
- Work effort:
- Creating the lookup table and implementation (can be extensive)
- "None" when producing tables (on-the-fly)
- Complexity of the tables does not matter

SCB

Implications of Using CKM 2

- *Consistency*: the exact same protection when:
    - The same table is produced again
    - A particular cell is included in several tables
- Not too much protection (compare: suppression of table cells)
- Protects against *differentiation*

- <u>Additivity of the table is lost</u>, each cell is protected individually.

SCB

Distribution of the Noice

The Noice is designed to have certain properties:

- Possible values are all integers in the interval -X,…,X
- Zeroes are not changed
- Protected values can not be negative
- The Noice do not add bias

SCB

Further implementation at SCB

- Population statistics and Table packages

- Income and taxes

- Issues raised
    - Lack of additivity
    - Should analysis be done on original or changed data
    - Added effort to compile margins separately
    - Communication to users vital and a challenge
    - Original and changed data available at the same time
    - Legacy systems with alot of manual elements
    - Brusade och obrusade siffror i olika tabeller

- Close cooperation between subject matter and methodology is needed

- Ongoing communication to facilitate knowledge and understanding

SCB

Feedback from users so far

- The BAS product has been well received with no issues due to CKM
- Positive feedback at the user council for Population statistics
- Strong interest from other statistical agencies nationally
- Some internal concern about potential user reactions

SCB

Description of the use of CKM in BAS

- SCB is obliged to protect from the disclosure of data pertaining to individuals in the published statistics. For BAS we use a method that enables the publication of detailed statistics without having to suppress statistical values or to have to reduce the level of detail in the tables. The method implies that the statistics are supplemented with random uncertainty in a controlled manner, **without adding bias**. All statistical values, including totals, which are greater than zero have been adjusted with a small negative or positive integer. Some statistical values are left unchanged. One consuquence of the method is that the published totals not always equals the sum of their reported parts. For instance will published totals for Men and Women not necessarily correspond to the sum of the published values for Men and Women respectively. The added uncertainty is small, but its impact will be greater for lower frequencies than for higher. This will be of importance also for proportions which always should be interpreted with caution when based on low frequencies. If published statistical values are added to create new totals the uncertainty will also add on. There is therefore a risk that the new total will have larger uncertainty than what, according to the method, would have been added to a single statistical value.

SCB