UNECE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Confidentiality

1-3 December 2021, Poland

Suppression or perturbation?

Wim Kloek (Eurostat)
wilhelmus.kloek@ec.europa.eu

Abstract

Statistical disclosure control measures come always at the cost of loss of information. Alternative protection methods have to be judged by the trade-off of risk and information loss. This is not a mathematical optimisation, as information losses can be perceived in different ways by different users depending on their needs and the remaining risks have a different nature in different protection methods. Therefore, also other considerations play a role in the final evaluation: is the method easy to implement in the statistical production process and is the protection easy to explain to the final users of the data?

In our paper, we will compare two methods for the protection of tabular data, cell suppression and perturbation based on the cell-key method. We will discuss the ultimate feasibility of suppression in the realistic case of many linked tables. An obvious advantage of the cell-key method is that it produces complete tables that can be analysed directly as a whole. However, this comes with at least three drawbacks: 1) perturbation as such could trigger a concern on the reliability of statistics, 2) in order to preserve unbiasedness in the perturbation, also safe cells have to be perturbed and 3) the method does not respect the additivity constraint. The seriousness of these issues depend on the concrete context in which the methods are applied. We will analyse frequency tables from almost full enumerations (weights close to 1), frequency tables from sample surveys, and magnitude tables (especially with skewed distributions). From the discussion, we will draw some general conclusions on what methods could be used in which context, and how to mitigate the drawbacks of the chosen solution. Will it be feasible to avoid suppression as protection method?