Workshop on Statistical Data Editing
18-20 September 2018
Neuchatel, Switzerland

Topic (ii): Quality of editing and imputation processes

Topic organizers: Sander Scholtus, CBS, Netherlands and Simona Rosati, Istat, Italy
Topic (ii): Quality of editing and imputation processes

• Editing and imputation (E&I) processes can improve data quality, but it is necessary to choose appropriate imputation methodologies in order to avoid introducing bias.
  – Reducing non-sampling errors (quality of data)
  – Providing information on the survey process (quality of process)

• Not only indicators, but also techniques that allow to evaluate the quality of E&I processes and of data that result from these processes.

• Main purposes and results: development of a conceptual framework (common language); facilitate and standardize the evaluation indicators measuring the effect on data of E&I processes.
Topic (ii): Quality of editing and imputation processes - Papers

1. Some quality indicators of editing and imputation in Household Budget Survey in Bosnia and Herzegovina – Bosnia and Herzegovina

The aim of this paper is to present some basic indicators of raw data, as measures of the quality of editing and imputation processes. A first step towards a standardization of editing and imputation process (concerning quality measuring) in statistical surveys.

2. On the development of a generalized framework to evaluate and improve imputation strategies at Statistics Canada – Canada

This paper will discuss the creation of an evaluation framework within the confines of CANCEIS that allows the user to calculate Monte Carlo estimates of bias, variance, and mean squared error for various parameters of interest, under various imputation strategies, through simulation. The user is able to perform simulation studies on either supplied real world or simulated data sets under various response mechanisms.
3. **A proposal of an evaluation framework for processes based on the use of administrative data – Italy**

This paper focuses on the key issue of the development of a new quality framework to assess the quality of Official Statistics based on a multi-source process. The aim is to propose an evaluation system framework useful to: (i) monitor the development and the final quality of new processes; (ii) help practical decisions about statistical design and monitoring.

4. **U.S. Federal Committee on Statistical Methodology Working Group on Transparent Quality Reporting of Editing and Imputation when Integrating Data from Multiple Data Sources – USA**

The paper shows processes and results of a working group. The group’s objective was to propose quality standards for use with statistical information products and services that are based on multiple data sources. Principal attention was directed toward standards related to transparency regarding the quality of input data sources, statistical methodology, and output data products and services.
Enjoy the presentations!
Topic (ii): Quality of editing and imputation processes

Discussion
Topic (ii): Quality of editing and imputation processes – Points for discussion

• Questions for discussion taken from USA’s presentation (Transparent Quality Reporting of Editing and Imputation when Integrating Data from Multiple Data Sources):
  
  – What are your experiences working with a data quality criterion and methodological tools or approaches?

  – For what applications have you found them to be useful?

  – What feedback have you received from your stakeholders?

  – Please share recommendations, software used, tech reports, and case studies.
Topic (ii): Quality of editing and imputation processes – Points for discussion

- Indicators
  - Many indicators available for quality of raw data (EDIMBUS, cf. presentation by Bosnia and Herzegovina)
    - E.g., (un)weighted item response rate, proportion of records with missing values, ...
  - Also many indicators for impact of data editing and imputation (EDIMBUS, cf. presentation by Italy)
    - E.g., proportion of values imputed, compare distributions before and after E&I, ...
    - Can be used to monitor E&I process over time (repeated surveys)
    - Mostly no direct interpretation as indicators for quality of edited/imputed data
  - Any experiences with measuring quality of edited/imputed data?
Topic (ii): Quality of editing and imputation processes – Points for discussion

• Any experiences with estimating uncertainty in final estimates due to editing and imputation?
  – Effect of overall E&I process on accuracy of output
  – Contributions of individual process steps
  – If accuracy is insufficient: where to make improvements?
Topic (ii): Quality of editing and imputation processes – Points for discussion

• Quality reporting:
  – Different types of users
    • Internal: data editors, statistical analysts, process managers, methodologists
    • External: general public, journalists, policy makers, researchers
  – Different interests, different levels of statistical knowledge
  – Therefore also different indicators required?
  – How to communicate about editing and imputation to external users?