The General Tool for Macroediting at SURS

Nejc Jevšnik, Zvone Klun, Rudi Seljak

Hague, 25.04.2017
SOP – GENERALISED APPROACH TO STATISTICAL DATA PROCESSING

TABULATION AND PROTECTION OF LINKED TABLES
- Excel outputs
- SI-STAT outputs
- outputs for tabular data protection

QUALITY INDICATORS
- unit non-response rate
- item non-response rate
- overcoverage rate
- imputation rate
- editing rate

MACRO EDITING
- validation of aggregates
- detection of outliers
- graphical data analysis

AGGREGATION
- statistical estimates
- standard error estimation
- primary suppression

MICRO EDITING
- sistematical corrections
- individual corrections
- imputations

DATA VALIDATION
- validation rules
BASIC ARCHITECTURE OF THE SOP SYSTEM BUILDING BLOCK
ARCHITECTURE OF THE MACRO EDITING MODULE

- USERS
  - Process
  - Business

- DEVELOPERS
  - Composite
    - Data and macrodata
    - Utility
    - Infrastructure
    - Informational

- Macro Editing
  - Macrodata validation
  - Outlier detection
  - Graphical analysis

  - Derived variables
    - IQD
    - QD
    - AUX_RAT
    - AVR_RAT
    - HIST_RAT

  - Metadata transformation
    - Metadata validation for macrodata validation
    - Metadata validation for outlier detection
    - Metadata validation for graphical analysis

  - Graphs
    - GRAPH 1
    - GRAPH 2
    - GRAPH 3
    - GRAPH 4
    - GRAPH 5
    - GRAPH 6
    - GRAPH 7
VALIDATION AT THE MACRO LEVEL

• Validation of the aggregated data
• Checking for the inconsistencies in the aggregated data
• All the aggregates are stored in one unique database (macro database)

Usage:
Comparing aggregates from different reference periods

Output:
- Frequencies of aggregates that failed the validation rules
- List of aggregates that failed the validation rules
OUTLIER’S DETECTION

• Searching for the values in the microdata that are significantly different from other values
• Outlier detection can be carried out by different methods (IQD, QD, HIST_RAT, AUX_RAT, AVR_RAT, MVAL)
• Easy to add new methods

Output:
- Number of outliers by methods and number of outliers by methods in each stratum
- List of outliers
GRAPHICAL ANALYSES

• Cross-sectional data (analyse data for one reference period)
• Longitudinal data (analyse data for several reference period)
• Each method requires a specific set of parameters
  – parameters that define the implementation of the method
  – parameters that define the current implementation of the method