Macro selection and micro editing: a prototype

Wim Hacking
Statistics Netherlands

Macro-editing or selective editing is a process that usually takes place after automatic corrections and derivations have been applied to the microdata. After this has been done, records have to be corrected manually; to reduce the amount of work, macro-editing is applied. The idea is to select (groups of) records that

- have suspicious values
- **and** are influential, e.g. they are contributing to the value in the aggregated (output) tables

Especially the last criterion can result in a large reduction in the effort related to manual correction of micro-data. Some simulation studies (e.g., Granquist, 1994) report an efficiency gain of 35% – 80%.

Many specific applications have been implemented, but to our knowledge no general tool has been created for macro-editing. We are currently working on a prototype for this purpose. Its main features are:

- A number of visualizations (currently a hierarchical grid, scatter/histogram plots) and operations on the data (currently aggregation and outlier detection).
- Selection steps: e.g. what is the next step when a suspicious stratum has been selected: stratifying the records from that cell on another or finer aggregation level, show a number of variables in a scatter-plot, etc.
- A connection to R for more advanced calculations, e.g. for outlier detection.
- A Blaise-like syntax specifying the macro-selection process.
- Micro-editing is based on de Blaise data-entry program.

The Blaise-like specification of the macro-selection process should render it flexible enough to be applicable for many surveys. In addition, its flexibility will allow analysts to rapidly try a number of strategies to see what works during the prototyping phase.

Literature:

Granquist, L. (1994). Macro editing: A Review of some Methods for Rationalizing the Editing of Survey Data. Statistical Data Editing, Volume No. 1: Methods and Techniques.