Discussion on Microediting and Software

Alexander Kowarik[1], Matthias Templ[1,2]

[1] Department of Methodology, Statistics Austria
[2] Department of Statistics and Probability Theory, TU WIEN, Austria

UNECE Worksession on data editing
Ljubljana, Mai 10, 2011

Deterministic Editing vs Statistical Estimation

- Balanced edits are necessary in questionaires and for final checks
- Soft edits should be used sparsely and carefully
- Selective editing $\simeq$ Influence function in Robust Statistics Problem: it is measured on one or few priority estimates and the influence of a statistical unit on other estimates can still be large.
- Can statistical estimation (outlier detection) help in general (problem: semi-continuous variables, i.e. high proportion of structural zeros in variables). To which extend?
- Hot-deck vs k-NN vs model-based imputation.

Deterministic Editing vs Statistical Estimation

- Balanced edits are necessary in questionaires and for final checks
- Soft edits should be used sparsely and carefully
- Selective editing $\simeq$ Influence function in Robust Statistics Problem: it is measured on one or few priority estimates and the influence of a statistical unit on other estimates can still be large.
- Can statistical estimation (outlier detection) help in general (problem: semi-continuous variables, i.e. high proportion of structural zeros in variables). To which extend?
- Hot-deck vs k-NN vs model-based imputation.

Deterministic Editing vs Statistical Estimation

- Balanced edits are necessary in questionaires and for final checks
- Soft edits should be used sparsely and carefully
- Selective editing $\simeq$ Influence function in Robust Statistics Problem: it is measured on one or few priority estimates and the influence of a statistical unit on other estimates can still be large.
- Can statistical estimation (outlier detection) help in general (problem: semi-continuous variables, i.e. high proportion of structural zeros in variables). To which extend?
- Hot-deck vs k-NN vs model-based imputation.

Deterministic Editing vs Statistical Estimation

- Balanced edits are necessary in questionaires and for final checks
- Soft edits should be used sparsely and carefully
- Selective editing $\simeq$ Influence function in Robust Statistics Problem: it is measured on one or few priority estimates and the influence of a statistical unit on other estimates can still be large.
- Can statistical estimation (outlier detection) help in general (problem: semi-continuous variables, i.e. high proportion of structural zeros in variables). To which extend?
- Hot-deck vs k-NN vs model-based imputation.

- Balanced edits are necessary in questionaires and for final checks
- Soft edits should be used sparsely and carefully
- Selective editing $\simeq$ Influence function in Robust Statistics Problem: it is measured on one or few priority estimates and the influence of a statistical unit on other estimates can still be large.
- Can statistical estimation (outlier detection) help in general (problem: semi-continuous variables, i.e. high proportion of structural zeros in variables). To which extend?
- Hot-deck vs k-NN vs model-based imputation.

Deterministic Editing vs Statistical Estimation

- Balanced edits are necessary in questionaires and for final checks
- Soft edits should be used sparsely and carefully
- Selective editing $\simeq$ Influence function in Robust Statistics Problem: it is measured on one or few priority estimates and the influence of a statistical unit on other estimates can still be large.
- Can statistical estimation (outlier detection) help in general (problem: semi-continuous variables, i.e. high proportion of structural zeros in variables). To which extend?
- Hot-deck vs k-NN vs model-based imputation.

- We saw in this session free and open-source joint development and on the other hand SAS macros.
- Why statistical offices are still afraid of free and open source software for statistics, data bases, etc, and are focused on Microsoft products (Access, Excel, . . .) and SAS?
- Should play open-source development a bigger role in editing (at the next work session?).
- CRAN Task View on Official Statistics: http://CRAN.R-project.org/view=OfficialStatistics
- Joint development of methods and software under a funded research programme? (Fellegi-Holt software, etc, under GPL and joint development?)

- We saw in this session free and open-source joint development and on the other hand SAS macros.
- Why statistical offices are still afraid of free and open source software for statistics, data bases, etc, and are focused on Microsoft products (Access, Excel, . . .) and SAS?
- Should play open-source development a bigger role in editing (at the next work session?).
- CRAN Task View on Official Statistics: http://CRAN.R-project.org/view=OfficialStatistics
- Joint development of methods and software under a funded research programme? (Fellegi-Holt software, etc, under GPL and joint development?)

Software

- We saw in this session free and open-source joint development and on the other hand SAS macros.
- Why statistical offices are still afraid of free and open source software for statistics, data bases, etc, and are focused on Microsoft products (Access, Excel, . . .) and SAS?
- Should play open-source development a bigger role in editing (at the next work session?).
- CRAN Task View on Official Statistics: http://CRAN.R-project.org/view=OfficialStatistics
- Joint development of methods and software under a funded research programme? (Fellegi-Holt software, etc, under GPL and joint development?)

Software

- We saw in this session free and open-source joint development and on the other hand SAS macros.
- Why statistical offices are still afraid of free and open source software for statistics, data bases, etc, and are focused on Microsoft products (Access, Excel, . . .) and SAS?
- Should play open-source development a bigger role in editing (at the next work session?).
- CRAN Task View on Official Statistics: http://CRAN.R-project.org/view=OfficialStatistics
- Joint development of methods and software under a funded research programme? (Fellegi-Holt software, etc, under GPL and joint development?)

Software

- We saw in this session free and open-source joint development and on the other hand SAS macros.
- Why statistical offices are still afraid of free and open source software for statistics, data bases, etc, and are focused on Microsoft products (Access, Excel, . . .) and SAS?
- Should play open-source development a bigger role in editing (at the next work session?).
- CRAN Task View on Official Statistics: http://CRAN.R-project.org/view=OfficialStatistics
- Joint development of methods and software under a funded research programme? (Fellegi-Holt software, etc, under GPL and joint development?)