Summary Topic (v)
Software & Tools

C. Poirier[1]   M. Templ[2,3,4]

[1] Statistics Canada / Statistique Canada

[2] STATISTIK AUSTRIA — Die Informationsmanager

[3] TU WIEN

[4] data-analysis OG

UNECE Worksession on Editing, Oslo, 25/09/12

R

- We saw that the popularity of R increases ⌐‿‿‿⌐also in Official Statistics
- New Developments come almost exquisitly within R
- Exchange of scripts and code is easy
- R as mediator [Todorov and Templ, 2012, Templ and Todorov, 2012]

Note: http://cran.r-project.org/web/views/OfficialStatistics.html

R

- We saw that the popularity of R increases _____ also in Official Statistics
- New Developments come almost exquisitly within R
- Exchange of scripts and code is easy
- R as mediator [Todorov and Templ, 2012, Templ and Todorov, 2012]

Note: http://cran.r-project.org/web/views/OfficialStatistics.html

General Questions

- Prototype stage
  - Do you keep realistics objectives? Avoid open scope?
- Production stage
  - Do you keep up with maintenance/research activities?
- Foundation
  - ℝ or SAS: Do you have good reasons to use one, not the other?

TEA for editing and imputation

- Use of SQLite, R and C. Really a need to use SQLite when having a powerful server or desctop computer?
- Use of multiple imputation methods - do you call the original code of IVEWARE within R? Better to use irmi? [Templ et al., 2011]
- How about the generation of implicite edits from explicite edits. Is that considered? Which heuristic is used?
- Is TEA already used in the production process?
- Will the package be available on CRAN in future?

Treemaps and Tableplots for visual editing

Visual tools are very efficient to analyse your data - also for editing.

- treemap only for dissemination of graphics? Interactivity possible, i.e. to get some meta-data when clicking on a region?
- tableplot gives a great overview of the data matrix.
- possible improvement: if used in editing it indicates which bin is an "outlier", but there seems to be no way to interactively show the corresponding observations in the bin (relation of graphic to individuals is lost). Better to use GAUGIN for that purpose?
- why to calculate the bins only with arithmetic means?

R - X12

for batch processing of X12-ARIMA and letting the output and parameter choice be better processed and accessible.

- Many advantages in the production stage over pure X12-ARIMA
- Is there an ongoing cooperation with the CENSUS Bureau?
- Is the package already applied in production?
- What are the differences (advantages/disadvantage) to Demetra?

Screening in UNIDO

- model-based outlier detection of short time series
- pilot study if robust methods are useful
- many problems occurs in the data used: missing values and short time series. Is it really useful to automatically check the series using complex models?

editrules and deducorrect

editrules and deducorrect looks promising.

- to apply it in production, many user may need a graphical user interface. Is the development of a point and click interface in plan?
- are implicite edits generated from the explicit ones?
- which heuristic is used for the generation of implicite edits? (the problem is NP-hard)

SAS Macros, AMELIA

SAS Macros:

- Are the SAS macros available for other offices?

AMELIA2 application:

- Single imputation using expected values versus single imputation using stochastic noise.

- Also of interest to compare the results from mice [van Buuren and Groothuis-Oudshoorn, 2011], irmi [Templ et al., 2011] and mi [Yu-Sung et al.].

SAS Macros, AMELIA

SAS Macros:

- Are the SAS macros available for other offices?

AMELIA2 application:

- Single imputation using expected values versus single imputation using stochastic noise.
- Also of interest to compare the results from mice [van Buuren and Groothuis-Oudshoorn, 2011], irmi [Templ et al., 2011] and mi [Yu-Sung et al.].

References

M. Templ and V. Todorov. Official statistics and survey methodology meets R. The Survey Statistician, 66:26–34, 2012.

M. Templ, A. Kowarik, and P. Filzmoser. Iterative stepwise regression imputation using standard and robust methods. Comput Stat Data Anal, 55(10):2793–2806, 2011.

V. Todorov and M. Templ. R in the statistical office: Part 2. Working paper, United Nations Industrial Development, 2012.

S. van Buuren and K. Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3):1–67, 2011. URL http://www.jstatsoft.org/v45/i03/.

S. Yu-Sung, A. Gelman, J. Hill, and M. Yajima. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. Journal of Statistical Software, 45.