Computational results for various error localisation algorithms

Ton de Waal
Statistics Netherlands

UN/ECE Work Session on statistical data editing, 20-22 October, Madrid

Algorithms for error localisation

- All algorithms are based on Fellegi-Holt paradigm:
  - minimise the (weighted) number of fields to be modified so that all edits can be satisfied
- Evaluated algorithms:
  - Standard software for integer programming
  - Vertex generation (Chernikova's algorithm)
  - Non-standard branch-and-bound algorithm
  - Cutting plane algorithm

Data sets and evaluation

- Six data sets from business surveys
  - Only numerical data
  - Realistic data and edit rules
- Evaluation study restricted to comparing computing times
  - Standard software for integer programming determines only ***one*** optimal solution
  - Other algorithms determine ***all*** optimal solutions; thus allowing one to select the “best” solution using a secondary, more statistical criterion

Conclusions

- All algorithms appear sufficiently fast for use in practice at Statistics Netherlands
- Therefore other aspects more important:
  - Costs
  - Dependency on third-party software
  - Complexity of the algorithm
  - Development and implementation of the software
  - Maintenance of the software
  - Possibilities to extend the functionality of the algorithm