GUI Demonstration for MASSC Disclosure Limitation Method

D. H. Wilson, F. Yu, and D. Wang
Statistics Research Division, RTI International

Presentation to JSM,
Toronto, CA, August 10, 2004
The Topics

- What is the GUI For MASSC?
- MASSC background
- How does the GUI work?
- User’s input
- User’s expectation and execution result
- Future development
- Conclusions
- Example Screens
What is the GUI for MASSC

- A Graphical Interface for executing the MASSC procedures
- A Mechanism which connects users and the process (through Screens, Reports, Graphics).
- A Mechanism which logically facilitates all the procedures within MASSC
- A mechanism which manages all the MASSC treatment required resources, intermediate datasets/files and end MASSC products
What is the GUI for MASSC (Cont.)
MASSC Background

- A data disclosure treatment mechanism
- The steps include Micro-Agglomeration, Substitution, Sub-Sampling, and Weight Calibration
- MASSC reduces the data disclosure risk while controlling information loss at the same time
- Applied for patent in 2002
- MASSC application development started in 2002
- The GUI screens enable users to exercise the MASSC treatment method
How does the GUI work?

1. Facilitates all source and intermediate files
2. Controls parameters
3. Manages variables, constraints and variable recoding
4. Controls procedure execution flow
5. Manages reports (HTML format), intermediate files, and graphics for every execution step
6. Provides adequate information for users to properly examine the end product(s)
7. Repeats the treatment step(s) if necessary
How does the GUI work? (Cont.)

MASSC
Required resources

Screens

User Input

Reports/ Graphics

MASSC

End Products
User’s Input

1. The location of the source file(s)
2. Variable Recoding (if necessary)
3. Constraints
4. Control Variables, parameters, etc.
5. Other Information
6. The process needs user input, though too much input would lengthen the process and discourage the use of the GUI system
7. The MASSC GUI incorporates with the MASSC backend process to minimize the user input
User’s Expectation and the Execution Result

- The user expects a treated dataset that balances disclosure risk and information loss.
- The user is able to control the execution result by tuning the control variables, redefining the constraints, doing more recodings, etc. All can be achieved through the MASSC GUI.
- A full set of reports and graphics help the user to make a proper decision.
Each execution step can be repeated depending upon the user’s judgment.

In order to control the execution flow, the availability of each execution step is depended on the execution status of the previous step.

The MASSC GUI lets the user know the execution status of each step.
User’s Expectation and the Execution Result (Cont.)

- Screens
- User’s input
- Micro-Agglomeration
  - Satisfied?
    - Yes
    - Substitution
      - Satisfied?
        - Yes
        - Sub-Sampling
          - Satisfied?
            - Yes
            - Calibration
              - Satisfied?
                - Yes
                - Final Product

  - No
    - No
      - No
        - No
          - No
Future Development

- Web Enabled Process
- Commercial Package
- Stand-alone Application
Example Screens

Micro Agglomeration and Risk Strata Creation

Parameters:
- Preload Variable Set:
  - Core IVs
  - Noncore IVs
  - SVs
- Y Indicator
- Weight Variable
- Number of Risk Strata
- Rename Variable:

Unselect  Reset

Recoding
Example Screens (Cont.)

[Diagram of a software interface with tabs for Phase I, Phase II, Phase III Probability, and Phase III Optimal. The interface includes fields for input parameters such as # of Constraints, Stratum Span, Alpha, Lower Bound, Upper Bound, and a section for Weight for Loss Function with a Loss Function List and buttons for Add and Delete. A Run Phase III Prob button is also visible.]
Example Screens (Cont.)
Example Screens (Cont.)
Example Screens (Cont.)

Uniques distribution - original sample

The FREQ Procedure

<table>
<thead>
<tr>
<th>Unique indicator in sample 1</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core+0</td>
<td>39</td>
<td>11.85</td>
<td>39</td>
<td>11.85</td>
</tr>
<tr>
<td>Core+1</td>
<td>47</td>
<td>14.29</td>
<td>86</td>
<td>26.14</td>
</tr>
<tr>
<td>Core+2</td>
<td>52</td>
<td>15.81</td>
<td>138</td>
<td>41.95</td>
</tr>
<tr>
<td>Others</td>
<td>191</td>
<td>58.05</td>
<td>329</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Nonuniques analysis - original sample

The FREQ Procedure
Conclusions

- MASSC GUI is a user-friendly graphical interface which communicates between the users and the MASSC backend processes
- MASSC GUI is flexible
- MASSC GUI reduces the complexity of MASSC method when comes to the real world practice
- The MASSC GUI future developments are under way
References


Contributors

- Dr. Avinash C. Singh, RTI International
- Dr. David H. Wilson, RTI International
- Dr. Feng Yu, RTI International
- Dr. George H. Dunteman, RTI International
- Mr. Donghui Wang, RTI International