New business survey confidentiality
software G-Confid

Joint UNECE/Eurostat Work Session
on Statistical Data Confidentiality
Tarragona, Spain, Oct. 26, 2011

Jean-Louis Tambay & Jean-Marc Fillion
What is G-Confid?

- Cell suppression software for tables of magnitude created by Statistics Canada
- Based on methodology used in old CONFID system
- Three components (require SAS 9.2)
  - PROC SENSITIVITY identifies confidential cells by the application of sensitivity measure(s)
  - Macro SUPPRESS creates a suppression pattern using the SAS/OR® LP solver
  - Macro AUDIT audits a suppression pattern
- Can handle any table size & number of dimensions subject to SAS & hardware limitations
PROC SENSITIVITY – identify confidential cells

Inputs:

- Microdata file
  - Classification variables (one per table dimension)
  - Enterprise identifier (blank for anonymous respondent)
  - Enterprise value
  - Enterprise value for a shadow variable (optional)
- Definition of hierarchy(ies) for each table dimension
  - Allows multiple decompositions of dimensions
- Code ranges for lowest level of hierarchies (optional)
- Sensitivity measure(s): (n,k), p-percent, user-defined
PROC SENSITIVITY – sample SAS code

proc sensitivity data=microfile
  outconstraint=consfile outcell=cellfile
  outlargest=largestfile
  hierarchy="0 1 2; 0 1 2 3;"
  srule="nk 1 70 2 80"
  range=";1 101 201 301: 2 102 202 302:
    3 103 203 303;"
  minresp=5;
  id Enterpriseid;
  var Income;
  shadow Profits;
  dimension Province Industry;
  by Year;
run;
Macro SUPPRESS – carry out cell suppression

- Inputs:
  - Cell & constraints files (e.g., from PROC SENSITIVITY)

- Syntax:

  ```
  %Suppress(InCell=, Constraint=, 
           CFunction1=, CFunction2=, CVar1=, 
           CVar2=, OutCell=, OutComplement=, 
           By=, ScaleCost=, DebugInfo= );
  ```

- Cost functions include SIZE (=tot), DIGITS (=log(tot+1)), CONSTANT (=1), INFORMATION (=log(tot+1)/(tot+1))

- Can use other variables as cost variables

- Can run LP process twice to reduce #suppressions (e.g., using SIZE & INFORMATION)
Macro SUPPRESS – output cell file contents

- Variables common with input cell file such as:
  - Values for each dimension (row, column...)
  - Cell total value (& total for shadow variable)
  - Number of respondents
  - Number of anonymous respondents (& their total value)
  - Cell sensitivity value
  - Cell input status (Sensitive, Variable, Suppressed, Published)

- New variables:
  - Cell output status (Suppressed, Published)
  - Net variation (largest amount cell was moved)
Macro AUDIT – validate a suppression pattern

- Calculates minimum and maximum values for each suppressed cell (& aggregate) using LP solver

Syntax:

```
%Audit(InCell=, Constraint=, OutCell=, LBFactor=, UBFactor=, By=, SasConnect=, DebugInfo=, ReportLevel= );
```

- LBFactor & UBFactor set bounds for suppressed cells in the LP solver (default bounds are $0.5\text{tot}$ & $1.5\text{tot}$)
- OutCell file provides minimum, maximum & midpoint values for suppressed cells (& aggregates)
- Summary results produced (protection achieved/not)
## Performance of G-Confid

<table>
<thead>
<tr>
<th>Run times</th>
<th>#dim</th>
<th>#cells</th>
<th>#sensitive cells</th>
<th>#complements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Phase 1 (SIZE COST)</td>
</tr>
<tr>
<td>9 sec.</td>
<td>2</td>
<td>3046</td>
<td>333</td>
<td>357</td>
</tr>
<tr>
<td>32 sec.</td>
<td>2</td>
<td>5245</td>
<td>856</td>
<td>712</td>
</tr>
<tr>
<td>6 sec.</td>
<td>3</td>
<td>1329</td>
<td>147</td>
<td>592</td>
</tr>
<tr>
<td>4 sec.</td>
<td>3</td>
<td>2149</td>
<td>69</td>
<td>230</td>
</tr>
<tr>
<td>10 sec.</td>
<td>3</td>
<td>2825</td>
<td>306</td>
<td>709</td>
</tr>
<tr>
<td>53min.</td>
<td>3</td>
<td>8074</td>
<td>608</td>
<td>2116</td>
</tr>
<tr>
<td>2h 45m</td>
<td>4</td>
<td>16992</td>
<td>2527</td>
<td>6007</td>
</tr>
</tbody>
</table>
For more information, please contact:

Jean-Louis Tambay
Jean-Louis.Tambay@statcan.gc.ca

Jean-Marc Fillion
Jean-Marc.Fillion@statcan.gc.ca

Pour plus d’information, veuillez contacter :

Jean-Louis Tambay
Jean-Louis.Tambay@statcan.gc.ca

Jean-Marc Fillion
Jean-Marc.Fillion@statcan.gc.ca