Editing and evaluation of statistics based on administrative microdata – example by Norway

Aslaug Hurlen Foss and Ane Seierstad
What is administrative microdata?

- Data collected and processed for administrative purposes
- Production of statistics
  - Data has to be processed to fit the statistical target
  - Strongly dependent on ID-identification
  - Huge amount of data
- Population controls - more important than controlling the variables

<table>
<thead>
<tr>
<th>Enterprise id</th>
<th>Person id</th>
<th>Start date</th>
<th>Stop date</th>
<th>Working hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>918273645</td>
<td>987654321</td>
<td>01.01.2015</td>
<td>31.01.2015</td>
<td>40</td>
</tr>
<tr>
<td>918273645</td>
<td>123456789</td>
<td>01.01.2015</td>
<td>31.01.2015</td>
<td>40</td>
</tr>
<tr>
<td>918273645</td>
<td>728163548</td>
<td>15.01.2015</td>
<td>31.01.2015</td>
<td>23</td>
</tr>
</tbody>
</table>
Population controls

• “Missing population” - Macro-editing methods on aggregated data

• “Duplicate population”-classify different cases:
  a) Two reporting unit sends inn data with information on the same population.
  b) Two reporting unit sends inn data with information on the same population but for different periods.

• From manual processes to automatization and use of weights
## Generic solution with weights

<table>
<thead>
<tr>
<th>Unit id</th>
<th>Variable 1</th>
<th>Reporting id</th>
<th>Weight</th>
<th>Weight population</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>40 000</td>
<td>918</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>728163548</td>
<td>23 000</td>
<td>918</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>728163548</td>
<td>24 000</td>
<td>846</td>
<td>0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Total:** 23 000

<table>
<thead>
<tr>
<th>Unit id</th>
<th>Variable 1</th>
<th>Reporting id</th>
<th>Weight</th>
<th>Weight population</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>40 000</td>
<td>918</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>728163548</td>
<td>23 000</td>
<td>918</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>728163548</td>
<td>24 000</td>
<td>846</td>
<td>0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Total:** 23 500

<table>
<thead>
<tr>
<th>Unit id</th>
<th>Variable 1</th>
<th>Reporting id</th>
<th>Weight</th>
<th>Weight population</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>40 000</td>
<td>918</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>728163548</td>
<td>23 000</td>
<td>918</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>728163548</td>
<td>24 000</td>
<td>846</td>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Total:** 47 000
Evaluation

• Change of the population
  – Number of changes
  – Effects on macro level
  – Classified by different type of population “error”

• Change of variables

• Evaluation reports on aggregated level
Driller – general IT-system

• View to the data
• Search and filter
• Aggregation and drill down
• Automatization and rule updating
• Reports on evaluation with drilling functionality
• Administrative module
  – Set variables, aggregations, micro-controls
Driller – technical facts

• Main principle:
  • metadata driven processes
  • general functionality
  • self-service
  • reuse

• Traditional data table format
• Built in 2012 in Adobe Flex (Oracle) and the view for the users is in the browser FireFox
• 25 surveys used it in 2014 and one survey has 18 mil records a month
Driller – interface
Future developments in process

• Toolbox/library for macro-editing and other methods
• Graphics
• Improving the report module
• Facilitate generic population controls

• Openness for using software R in developing the methods in the library