Quality assurance and assessment of statistics from new and administrative data sources

Statistics South Africa

8 December 2022
• Background
• Example of the SVC/GSBPM mapped to SASQAF indicators
• The application of SASQAF for quality assurance and assessment of admin data sources
• Case study based on other Data Quality frameworks
• Conclusion
In pursuance of S14, Stats SA developed a quality framework for assurance and assessment of statistical products within Stats SA and in the NSS community.

The main purpose of the framework is to provide a flexible structure for the assessment of statistical products.

The SG uses the framework to evaluate the quality of statistics produced in the NSS.

SASQAF is also be used as tool to improve quality of statistical products:

- self-assessments by data producers
- Used for Independent Quality assessment (DQAT)
Background cont...

**SASQAF Dimensions – Indicators – Standards**

- **Integrity** (6 indicators 6 Std.)
- **Accuracy** (7 indicators 36 Std.)
- **Relevance** (5 indicators 5 Std.)
- **Interpretability** (3 indicators 3 Std.)
- **Comparability & Coherence** (5 indicators 12 Std.)
- **Methodological Soundness** (6 indicators 14 Std.)
- **Pre-requisites** (8 indicators 21 Std.)
- **Timeliness** (4 indicators 10 Std.)
- **Accessibility** (12 indicators 13 Std.)

**Methodological Soundness**: (6 indicators 14 Std.)

- **Integrity**: (6 indicators 6 Std.)
- **Accuracy**: (7 indicators 36 Std.)
- **Relevance**: (5 indicators 5 Std.)
- **Interpretability**: (3 indicators 3 Std.)
- **Comparability & Coherence**: (5 indicators 12 Std.)
- **Pre-requisites**: (8 indicators 21 Std.)
- **Timeliness**: (4 indicators 10 Std.)
- **Accessibility**: (12 indicators 13 Std.)
Part 1: Levels 1 and 2 of the Statistical Value Chain

Statistical Value Chain

Quality Management and Metadata Management

- **Need**
  - Determine need for information
  - Consult and confirm information requirements
  - Establish output objectives
  - Check data availability
  - Prepare business case

- **Design**
  - Define archive rules
  - Prepare statistical systems and workflow
  - Detailed project plan

- **Build**
  - Outputs
  - Data collection instrument
  - Data collection
  - Frame and sample methodology
  - Tabulation Plan / Variables

- **Collect**
  - Set up collection
  - Configure workflows
  - Process component
  - Run collection

- **Process**
  - Standardize
  - Classify and code
  - Load data into processing environment
  - Integrate data
  - Test end-to-end
  - Finalise production systems

- **Analyse**
  - Acquire ancillary information
  - Calculate aggregates
  - Validate
  - Describe and explain
  - Disclosure control & Anonymize

- **Disseminate**
  - Update output systems
  - Produce products
  - Prepare draft outputs
  - Produce Quality Statement
  - Manage release of products
  - Manage customer queries

- **Evaluate**
  - Gather inputs for programme review
  - Evaluation report
  - Quality plan
1. Problem Identification

2. Diagnostic Assessment

3. Problem Solution (Strategy & Plan)

4. Implement Solution

5. SASQAF Self Assessment

6. SASQAF Independent Assessment

Data quality improvement through technical support

Data quality Accreditation by Independent Quality Assessment

Background cont...
Background cont... SVC/GSBPM mapped to SASQAF indicators for Quality Assurance

<table>
<thead>
<tr>
<th>Phases</th>
<th>Sub-processes</th>
<th>Quality dimensions and indicators</th>
<th>Quality indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design</td>
<td>Tabulation plan/variables</td>
<td>Methodological soundness</td>
<td>8.1 Concepts, definitions, and classifications used follow accepted standards, guidelines or good practices (national, international, peer-agreed)</td>
</tr>
<tr>
<td>Design</td>
<td>Statistical processing methodology</td>
<td>Methodological soundness</td>
<td>8.3 Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), viz.:</td>
</tr>
<tr>
<td>Process</td>
<td>Integrate data</td>
<td>Comparability and coherence</td>
<td>7.4 A common set of identifiers (for the purpose of record matching) exist and have been agreed upon by data producers</td>
</tr>
<tr>
<td>Analyse</td>
<td>Validate</td>
<td>Accuracy</td>
<td>3.1 Measures of sampling errors for key variables are calculated. Amongst others these are:</td>
</tr>
</tbody>
</table>

- standard error
- coefficient of variation (CV)
- confidence interval (CI)
- mean square error (MSE)
- design effect (DEFF)
The application of SASQAF for quality assessment of outputs from admin data sources

<table>
<thead>
<tr>
<th>Quality indicators and standards that should be selected and reported</th>
<th>List of Standards</th>
<th>CPI</th>
<th>SNAP</th>
<th>FCM</th>
<th>GHS</th>
<th>QLFS</th>
<th>SAT-DTS</th>
<th>Electricity</th>
<th>FSHE</th>
<th>MACOD</th>
<th>SAPS</th>
<th>SAPAD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>List of indicators</strong></td>
<td>1.1</td>
<td>Measures of sampling errors for key variables are calculated. Amongst others these are:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- standard error</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- coefficient of variation (CV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- confidence interval (CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- mean square error (MSE) and design effect (DEFF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1.1</td>
<td>Measures of sampling errors must be calculated for the main variables. They must be available for other variables on request.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1.2</td>
<td>Measures of sampling errors must fall within acceptable standards. At a minimum the following must be calculated: standard error, coefficient of variation, confidence interval, mean square error. The low accuracy of variables (if these exist), is explained.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metrics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SE = \sqrt{\text{Var}(\hat{\theta})}$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$CV = \frac{\sqrt{\text{Var}(\hat{\theta})}}{\mu(\theta)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \mu^2(\theta)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Where $\hat{\theta}$ is an estimator of a parameter of interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1.3</td>
<td>Scientific sampling techniques must be used.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metrics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$a = \text{Design effect}(\hat{\theta}) = 1 + \hat{\theta}(n-1)$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2</td>
<td>Measures of non-sampling errors are calculated, viz.:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2.1</td>
<td>The extent of measures of non-sampling errors must be kept to an acceptable level.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metrics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Examples of Administrative data sources

- South African Police Services
  - Quality Standards to guide the collection and processing of data
- South African Reserve Bank
  - SARB – currently undertaking an online self assessment.
- Department of Home Affairs/Department of Health
  - Death Notification Forms for Mortality and causes of Death. Stats SA provide inputs into the updating the DNF Form
Case Study: Quality Assurance for administrative data using SASQAF

• Approach
  • Selection of indicators that are applicable to administrative data sources
  • Technical support giving advice on collection methods, methodology as well as specifications on what variables to include.

• Current developments
  • Research on different administrative quality assurance frameworks (other countries)
  • Consultation (National statistical system)
Case Study: Quality Assurance for administrative data using SASQAF (SVC – Implementation phase)
<table>
<thead>
<tr>
<th></th>
<th>Morocco</th>
<th>Australia</th>
<th>Canada</th>
<th>New Zealand</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Institutional environment</strong></td>
<td>Institutional environment</td>
<td></td>
<td></td>
<td></td>
<td><strong>Prerequisites of quality</strong></td>
</tr>
<tr>
<td><strong>Relevance</strong></td>
<td>Relevance</td>
<td>Relevance</td>
<td>Relevance</td>
<td>Relevance</td>
<td><strong>Relevance</strong></td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>Accuracy</td>
<td>Accuracy</td>
<td>Accuracy</td>
<td>Accuracy</td>
<td><strong>Accuracy</strong></td>
</tr>
<tr>
<td><strong>Timeliness</strong></td>
<td>Timeliness</td>
<td>Timeliness</td>
<td>Timeliness</td>
<td>Timeliness</td>
<td><strong>Timeliness</strong></td>
</tr>
<tr>
<td><strong>Accessibility</strong></td>
<td>Accessibility</td>
<td></td>
<td>Accessibility</td>
<td>Accessibility</td>
<td><strong>Accessibility</strong></td>
</tr>
<tr>
<td><strong>Coherence</strong></td>
<td>Coherence</td>
<td>Coherence</td>
<td>Coherence</td>
<td>Coherence</td>
<td><strong>Coherence</strong></td>
</tr>
<tr>
<td><strong>Interpretability</strong></td>
<td>Interpretability</td>
<td></td>
<td>Interpretability</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Coordination and Corporation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

Application of SASQAF when assessing statistics from the administrative data sources

• Status quo
  • SASQAF 2nd edition
  • Currently developing a separate checklist/framework for quality assessment of administrative data sources based on the General SASQAF.

• Lesson learnt:
  • Some of the indicators/standards are not applicable to administrative data

• Challenges
  • Takes time to identify the indicators that are not applicable

• Way forward/Recommendation
  • develop a separate SASQAF for administrative data sources
Thank You