

Processing and estimation issues

Regional Training

Producing Register-based Population Statistics in Developing Countries

Tehran 27 – 31 October 2013

Harald Utne, Statistics Norway

harald.utne@ssb.no

From administrative to statistical data

- Editing
- Use of multiple sources

Editing by register owners

- Purpose: Improve data for administrative purposes
- Focus on variables important for administrative use
- Variable included for statistical purposes only may not be controlled as carefully
- Statisticians must have good knowledge of these editing procedures
 - What is edited and how
 - What is not edited

Editing in NSO

- Focus on statistical use of data
- Each register first controlled by using internal logical checks
 - No external sources
- Discover system errors
 - Data not meeting formal requirements
 - “Blocks” of data missing or incorrect
 - May result in contacting register owners for new delivery of corrected data
- Isolated errors may be detected and corrected
 - Possibility of contacting individuals to collect information very limited
- Main method: Computerised logical editing

Editing in NSO (cont)

- Time references must be processed to obtain information for desired point in time or periods
 - Dates of changes or events
 - Registration dates
 - Distinguish real events from corrections
- Close contact with register owners important
 - Systematic errors detected should be corrected in the administrative registers
 - Beneficial for both parties

Use of multiple sources

- Administrative data must be processed to meet statistical needs
- Creation of statistical registers often involves linkage of data on an individual level
- Find best estimator for statistical variables
- The solution may be derived variables

Derived variables

- Creation of new variables
 - Example: Current activity status
- Prioritisation of information for a given variable
 - Depending on quality and coverage
- Adjustment or correction of information
 - Obtain maximum consistency

Current activity status

- Information based on several registers
 - Employment (10+)
 - Unemployment (1)
 - Pensions (3)
 - Ongoing education (students) (3)
- Prioritization based on
 - Statistical definition of variable (by ILO)
 - ♦ Rank: Employed, unemployed, not in labour force
 - Reliability of data sources (employed –unemployed)
- Estimates for given dates or periods (weeks)

Use of combined data over time

- Often requires a complex set of rules (model)
- Detailed documentation necessary
- Allow for change over time
- Vulnerable to administrative changes
 - For instance a variable may be omitted in a register or changed in a major way
- Model may have to be adapted
- Even improvement of data may give rise to discontinuity

Updating

Key principle for registers

- Need to process data only when units or attributes are changed
- No changes: No processing

Thank you for your attention!