**Statistics Finland**

# INDICATORS RELATED TO EDITING PROCESS MODEL

## Workshop of Baltic-Nordic-Ukrainian Network on Survey Statistics 2012

Saara Oinonen

25.8.2012

# INDICATORS FOR EDITING MODEL

- Control error identification and correction actions

- Analyze the effect of the editing actions

- Three goups of indicators presented:

  - Indicators describing raw data

  - Indicators related to error identification

  - Indicators related to error correction actions

# INDICATORS DESCRIBING RAW DATA

- In the beginning of the editing process
- Give information about
  - errors on data
  - errors effect on results
  - variables' and subgroups' significance
- Indicators for data missingness, effect of missing values, response, observation impact, significance and estimators sensitivity to observations

Response rate

$$\frac{\sum_i r_{ij}}{n}$$

*Statistics Finland*

# INDICATORS RELATED TO ERROR IDENTIFICATION

- Describe the amount of errors in variables and observations
- Evaluate the efficiency of error identification procedures
  - simulation test: inserting errors on edited data and testing identification methods
  - comparing results calculated from improved data
- Edit rules + flagging, data processing with functions and modelling, visual examination
- Indicators for error identification on
  - variable level (error degree on variable)
  - observation level (proportion of error occurence on all variables)
  - data level (error identification rate for certain edit rule)

Error identification rate for edit $l$

$$\frac{\sum_i e_{il}}{n}$$

# INDICATORS FOR ERROR CORRECTION

- Correction actions might include inquiring/searching the right value,  and/or imputation (cold-deck, from other data source, using statistical methods)

- Indicators describe

  - quality of data after error correction

    - response, missingness

  - amount of error correction actions

    - edit rate/ratio, proportion of corrections

  - effect on error correction on results or variables

    - edit impact, estimates sensitivity for edits

- Tagging edited observations is essential

Edit rate

$$\frac{\sum_i I(\hat{y}_{ij} \neq y_{ij})}{n}$$

# RECOMMENDATIONS FOR USE OF INDICATORS

**Quality standards of Eurostat**

- ESS Standard for Quality Reports
- ESS Handbook for Quality Reports
- Describes different dimensions of quality:
    - measurement errors
    - nonresponse errors
    - processing errors in microdata
    - imputation

## INDICATORS FOR DIFFERENT PURPOSES
## - WHAT NEEDS TO BE CONSIDERED?

- Target variables
- Subgroups
- Raw data
- Structurally missing values
- Error identification
- Error correction
- Weighting

# Statistics Finland

# THANK YOU FOR YOUR ATTENTION!

Refereces:

- EDIMBUS (2007). Recommended Practices for Editing and Imputation in Cross-sectional Business Surveys. Project report.
  http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf.
- EUREDIT / Ray Chambers (2004). Evaluation Criteria for Statistical Editing and Imputation, Project report.
- EUROSTAT (2009a). ESS Standard for Quality Reports, Luxemburg.
  http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf
- EUROSTAT (2009b). ESS Handbook for Quality Reports, Luxemburg.
  http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQR_FINAL.pdf
- OLLILA, P. (2012). Raaka-aineiston,editoinnin ja imputoinnin indikaattorit (luonnos). Unpublished memorandum. Statistics Finland.
- OLLILA, P. and ROUHUVIRTA, H (2012). Editoinnin prosessimalli (luonnos). Unpublished memorandum. Statistics Finland.
- STATISTICS CANADA (2007). Functional Description of the Banff System for Edit and Imputation, Ottawa.
- STATISTICS FINLAND (2007). Quality Guidelines for Official Statistics 2nd Revised Edition, Helsinki. http://www.tilastokeskus.fi/meta/qg_2ed_en.pdf