STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working paper No. 7
English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Data Editing
(27 – 29 May 2002, Helsinki, Finland)

Topic (i): Planning and management of statistical data editing

PLANNING AND MANAGEMENT OF STATISTICAL DATA EDITING

Contributed Paper

Submitted by the National Statistical Institute, Bulgaria [1]

Abstract

The planning and management of statistical data editing are very important for the improvement of the quality of statistical information. The National Statistical Institute of the Republic of Bulgaria is carrying out the editing of statistical data as part of data processing using different in-house developed software products. The efforts of the NSI are mainly directed to the imputation of the missing data. The National Statistical Institute of the Republic of Bulgaria has begun the work on building up an Integrated Statistical Information System (ISIS).

I. INTRODUCTION

A. Statistical data editing

1. The concept of statistical data editing is very wide. Data editing can significantly influence the quality of statistical data as well as the cost efficiency of statistical production. The editing process is truly effective as a part of the continued improvement cycle of the whole survey process. The stages of data editing are: survey management (questionnaire design, methodology of the survey); data collecting (for comprehensive and sample surveys), ensuring control during data collection process; data entry and data review (detection of statistical data errors by suitable software and correcting them); data adjustment.

2. Data editing is still the most central and extensive operative task in the whole statistical production process. The level of complexity of the data editing process depends primarily on the following two basic factors:
- Technical and technological approaches to the data editing process in statistical production;
- Organizational issues, in connection with the contact with the reporting units, decisions about sampling, the possibility of obtaining data from other sources (e. g. administrative registers and other public evidence) [1].

B. Planning of data editing processes

3. Data editing could be rendered inefficient because of poor planning. Planning of data editing is greatly influenced by survey specific aspects but in addition to those general methods and procedures should

[1] Prepared by Penka Dimitrova (pdimitrova@nsi.bg).

be implemented to achieve more efficient survey operations and standardization of planning activities with the benefit of better comparability of surveys. The planning activities should take into account new developments in the data collection process. As the planning of editing consists mainly of different activities, the assessment of detailed plans would be facilitated by benchmarks, which should enable comparisons between surveys [1].

C. Management of data editing processes

4. The management activities consist of a periodic observation of the error detection and correction. The management of data editing requires information about the state of error correction, real time effort and costs. The following are essential for the management of data editing:
- the definition of data quality which sets the aims for data editing activities;
- the way in which data editing activities are organized, and
- criteria and data used for the judgment of the efficiency of the data editing activity [2].

5. User demands for data quality should be collected as well. With regard to planning of data editing, the focus could be put on the commonly used quality criteria, such as timeliness, accuracy and clarity. Clarity is an important criterion for the planning of data editing because data editing must in many cases provide information for informative and user-friendly quality indicators. A continuous contact with users and relevant measures for data quality improvement used in interaction with users should be maintained [2].

6. The achieved "accuracy" of statistical data, needed runtimes and process costs from essential criteria for the management of data editing processes, measured by real data of data editing processes and compared with planned data coming from the planning of data editing [2].

II. NATIONAL STATISTICAL INSTITUTE PRACTICE

7. The overall activity of the National Statistical Institute (NSI) is based on the worked out Strategy for Development of Statistics for the period 2000 – 2006.

8. On the basis of this strategy, the National Statistical System formulates its policy at the stage of the country's preparation for accession to the European Union during the period 2000 – 2006 and also carries out the concrete tasks, implemented in the Law of Statistics and other normative documents. The basic points of the strategy are:
- To describe the essence and characteristics of the statistical system;
- To review the fundamental principles of the official statistics;
- To identify the problems and measures, needed for their solution;
- To formulate the objectives, tasks and priority as well as the necessary organisation and providing resources.
- To plan the work on harmonisation with the EC legislation and schedule the basic surveys and activities by areas for harmonisation and improvement until the year 2006. Meanwhile, on 22 May 2000 Bulgaria became a member of the General Data Dissemination System (GDDS) of the International Monetary Fund.

9. The main goals of the statistical practice in the process of harmonisation according to the requirements of the European statistical system are:
- improving the quality of the statistical information;
- improving efficiency of the statistical activity;
- improving the public confidence in statistical information;
- improving the cooperation between the participants in the statistical activity – users and respondents.

10. NSI is considering elaborating the following programmes to realize the main goals:
- National Programme for Adoption of the Acquis (NPPA);
- National Programme for Improvement of Quality of the Statistical Information, etc.

11. The National Statistical Institute of Bulgaria is a centrally organized institution and therefore statistical surveys are carried out in direct communication with reporting units. Data are collected by 28 Regional Statistical Offices at the regional level and are aggregated in NSI. Data collection is usually undertaken using the classical method of paper-based questionnaires. This means a huge number of statistical reports. In compliance with the above, technological and organizational approaches are set up to perform data editing for each individual survey.

12. Different software products, developed mainly in-house and continuously improved in accordance with Eurostat requirements, are used for data processing (including data checking and correction).

A. Census of Population, Housing Fund and Agricultural Farms in 2001

13. The National Statistical Institute carried out Census of Population, Housing Fund and Agricultural Farms in 2001. The statistical data editing of the Census data followed the traditional steps: questionnaire design, elaboration of survey methodology; data collection and control during the data collection process; data entry and error detection and their correction. Editing of the statistical data is ensured by manual review of the data by controllers before data entry in the Regional Statistical Offices and use of developed user-friendly software providing efficient facilities for statistical data entry editing. The software is accomplished by data control in the questionnaire (logical rules) and by access to the implemented classifications - ESCO, NACE, Rev. 1 International Standard Classification of Occupations (ISCO '88) and nomenclatures. For entry of data about enterprises and location, administrative registers such as BULSTAT Register, Register of the Settlements (EKTTE), civil register (PIN code) are used. The next stage is the aggregation and data processing in NSI.

III. DATA EDITING CONNECTED WITH IMPUTATION OF MISSING DATA

14. Another way to improve the quality of the statistical data is the imputation of missing data.

A. Continuing Professional Training Survey

15. In May 2000, NSI conducted a sample survey of the continuing professional training, which enterprises provided for their workers and employees in 1999. The survey was within the framework of the Project CVTS 2 organized by Eurostat simultaneously in the 15 member-countries of the European Union and for a number of east European countries candidates for membership, Bulgaria among them. The survey is coordinated with the methodological requirements and criteria of Eurostat. The data collection method was to interview the 3000 enterprises. Collection, data entry and data editing (primary control and correction of errors) were undertaken in 28 Regional Statistical Offices. Data were aggregated in NSI.

16. According to the Eurostat methodology for imputation of missing data, the following were used:
- second contact with the enterprises;
- using data from other surveys, compatible with adopted by CVTS 2 definitions;
- automated correction, included in the software used for data entry.

17. According to the Eurostat requirements the evaluation of data quality was realized by SAS-MACRO software CLAN97, elaborated by Statistics Sweden.

18. The sample survey of continuing professional training in 1999 was carried out successfully. Bulgaria is among the countries that finished the survey in time and it is one of those that have the highest results in respect of the registered norm of response (above 80 %) and quality of collected information.

B. Improving Statistics on the Distributive Trade in the Central European Countries

19. In 1996, NSI of the Republic of Bulgaria participated in the pilot project for "Improving Statistics on the Distributive Trade in the Central European Countries". This project was incorporated into the technical and financial assistance forwarded to the national statistical institutes of the CEC. A different system of validation and controls of the data, imputation of partial responses to the questions of the data, and extrapolation of some non-responses were techniques used to ensure the good quality of results.

20. The efforts of the NSI are directed to the survey of seasonal product and imputation of the missing prices in the survey of Producer Price Index, Consumer Price Index and External Trade Price Index.

C. Producer Price Index

21. The samples introduced so far in the PPI revision include seasonal products. These products are sampled because it is important to ensure the seasonal samples that are available throughout the year. For example, in textiles and apparel manufacturing, it is important to have representative samples for spring/summer, fall/winter, and year-round products. This involves close work with the respondents in order to identify the representative transaction for each season, to be included in the sample. This enables the collection of prices for year-round specifications on a monthly basis and prices for the spring/summer or fall/winter specifications when they are in production.

22. It is important that replacements for the seasonal specifications are not made for products that are out of season. Such replacements can only take place at the beginning of the seasonal production period. When the price for a seasonal specification is not available, an imputation is necessary. Imputations are made by using the short-term price change from a matched sample of specifications in the same product group or enterprise. If no prices are available within the product group or enterprise, the general approach is to use the short-term price change in the industry to impute the missing prices. However, there are circumstances where different product trends have been noted within the industry. In this situation, the short-term price trend of a similar product is used to make the imputation. In such cases, it is important that the same approach for imputation be applied consistently over time. If the product imputation approach is chosen for a particular PRODCOM group, then this method is always used for that group when there are no prices available.

23. The additional development of export unit values as part of a PHARE project are closely coordinated with a related PHARE project proposed for the development of an industrial export price index in the PPI. Also, the next phase of development should include a set of systematic edits for computing unit values at the detailed product level and reduce the reliance solely on analyst judgment.

24. The following approach has been established for the measurement of the PPI for non-domestic industrial output. A long-term strategy for improving the calculation of the PPI for non-domestic industrial output is now being implemented in the NSI. A key element of the strategy is contacts that will be established with enterprises that export commodities with a large share in the PPI for the non-domestic market. The best methods for providing information to the NSI should be discussed, and information should then be obtained from these enterprises, initially on an experimental basis.

25. In the short and medium term we plan to improve PPIs first on an experimental basis using different sources:

- true export price indices (via surveys of only the most relevant exporters initially);
- including additional questions in the new questionnaires for PPI-domestic;

- using unit value indices (especially for the more homogeneous, low SITC categories);
- using domestic PPIs (where there are indications that no significant differences exist between domestic prices and export prices); and
- using comparable price developments, including possibly the foreign exchange rate (there should be clear evidence that the selected price development is relevant for the export commodity considered).

D. Consumer Price Index

26. Imputations are made for all missing prices (both regular and seasonal) using short-term price relatives from a matched sample of collected prices as presented in the section on seasonal prices and imputations - from a matched sample and short-term for all items (excluding seasonal) beginning 2000, and from a matched sample and short-term for seasonal beginning 2001. Also, the NSI CPI division is working and will continue to work on evaluating the effects of quality differences and making quality adjustments when replacement varieties are introduced. Tests for alternative methods for estimating quality differences and making quality adjustments in the CPI should be completed and a programme for ongoing quality adjustment implemented.

IV. INTEGRATED STATISTICAL INFORMATION SYSTEM (ISIS)

27. The National Statistical Institute of the Republic of Bulgaria began work on building up an Integrated Statistical Information System (ISIS).

28. In the current year it is planned to provide the pre-design investigation. In the work on investigation it is expected to describe and order systematically the existing surveys, procedures of surveys, data processing, information aggregation, information flows, statistical infrastructure, statistical tools, statistical surveys' methods and organization.

29. As a result it is expected to define the vision for ISIS, for it's subsystems, for it's building stages by using contemporary information technologies. The purpose is to achieve better organization and coordination of statistical activities; unifying the statistical tools and implementing contemporary information technologies for data collecting, reduction of response burden, speeding up of data processing, quick and facilitated access to the statistical information (statistics). Work is expected to begin on System for administration of classifications and nomenclatures, Unified System of Statistical Indicators and Register of the statistical units (with included subsystems for Register usage as a tool for statistical activities planning and coordination).

30. One of the main issues of the system investigation and of the strategy for building up the ISIS is defining the procedures for data collection, processing, editing and dissemination of the statistical data as a result of statistical activities of the NSI and other bodies.

V. CONCLUSIONS

31. As a result of continuously increasing user demand, NSI directed its efforts to the improvement of the quality of the statistical information that led to a new request for data editing methods and management techniques.

32. The strategy for building up the ISIS is defining the procedures for data collection, processing, editing and dissemination of the statistical data as a result of the statistics activities of NSI and other bodies.

33. The NSI experts are very interested in being acquainted more deeply both with the Eurostat recommendations on this topic and with the best practices implemented by the national statistical offices of the countries – members of the European Union. We also are interested in receiving more information about

all seminars, meetings, etc. that are connected to data editing and imputations missing data which would contribute to the better understanding of data editing planning and management, methods and procedures which are of primary importance for the further improvement of the quality of the statistical information.

References

[1] Elmar Wein, The Planning of Data Editing, UNECE Work Session on Statistical Data Editing, 2000
[2] Elmar Wein, The Role of Error Detection and Correction Systems in Survey Management and Optimisation, 2001.