

Process Quality Control in the Phase of Checking: the Case of the Italian Consumer Expenditure Survey

Aspetti del Controllo del Processo nella Fase di Revisione dell'indagine Istat sui Consumi

Sabrina Barcherini, Donatella Grassi
Istat
barcheri@istat.it, dograssi@istat.it

Pinuccia Calia, Carlo Filippucci
Università di Bologna
calia@stat.unibo.it, filippucci@stat.unibo.it

Riassunto: La ricostruzione analitica del processo di produzione dei dati costituisce la fase preliminare e fondamentale per assicurare la qualità in un'ottica del miglioramento continuo. Inoltre, solo a partire da questa conoscenza di base del processo è possibile impiantare un adeguato sistema di monitoraggio del processo di produzione. In questo lavoro si tratta della rilevazione Istat dei consumi con lo scopo di pervenire alla costruzione di un sistema di controllo continuo del processo dell'indagine. In questa prospettiva si sono ricostruite tutte le operazioni e le fasi del processo di produzione e individuate le variabili chiave da tenere sotto controllo. L'analisi delle *performance* è stata utilizzata per individuare i problemi e le possibili soluzioni delineando così una esperienza concreta di miglioramento continuo della qualità dei dati.

Keywords: Data Quality, Process Control, Continuous Quality Improvement, Consumer Expenditure Survey

1. Introduction

The new perspective to quality in statistics is characterised by three basic principles: statistics have to be relevant, data quality depends on the production process, quality is achieved by means of continuous improvement of processes. These principles stress that quality stems from all operations carried out to produce data. To statistical agencies this approach has important effects. First of all, it implies to adopt a systematic approach to quality control, and management defining standard practices. Secondly, it is necessary to develop monitoring of production processes in order to discover and solve problems arising in implementing procedures and operations. Thirdly, it asks for exploring effects of changes. Finally, it is necessary spreading a new culture on quality.

In the perspective of studying and developing a quality management system, an experiment of production process control has been started for the Italian Consumer Expenditure Survey by Istat. This is a nation-wide two stage stratified survey, which involves about 25.000 households each year. Data collection is carried out through a complex organisation: Istat establishes survey sampling design, instruments, regulations and guidelines, but Statistical Offices of Municipalities are responsible for field operations. Hence it is evident that the production process control is a crucial issue.

The approach we adopted, designed according to the framework proposed by Morganstein and Marker (1997) for the Continuous Quality Improvement (CQI), is broadly outlined as follows: 1) describing the process, tracking procedures, operations and their sequence; 2) identifying key variables, i.e. factors which mostly affect quality

of data, and suitable indicators in each step of the process; 3) defining a system for continuous quality improvement, measuring effects of changes and innovations on key variables and indicators.

The paper discusses some issues and reports main findings in designing a production process control with reference to the phase of data checking and correction. A detailed analysis of the survey field operations was already performed by Filippucci and Calia (2001).

In the section 2 a brief description of the data revision process is presented, main findings from collected indicators are in the 3rd section, finally some implications for future work are drawn.

2. Data revision process

Italian Consumer Expenditure Survey is a permanent survey where data collection is carried out by Municipality Statistical Offices that have to manage: survey preliminary operations, interviewers selection and training, households contacts, effective data collection and monthly forwarding of data to Istat. Therefore accurate data revision is crucial to assure final quality and to limit or remove possible systematic errors during data processing. A brief description of the revision process is presented in order to identify fundamental operations, planned checks and associate indicators.

First manual revision is performed by municipality operators. In order to eliminate the most evident inconsistencies municipality supervisors: i) compare theoretical and actual sample size; ii) check household codes; iii) analyse the qualitative level of questionnaires.

Manual revision performed by Istat's operators. On the basis of standardised rules Istat checks: i) theoretical and actual sample size; ii) data inside the questionnaires. At this stage follow-up of municipalities or households is carried out to solve, when possible, partial non-response and inconsistencies. Questionnaires presenting non recoverable anomalies are discarded. The remaining inconsistencies are corrected by electronic checks.

The following indicators can be calculated from the checks: i) number of errors concerning the field work not identified by municipality operator, by type; ii) number of errors in field work not identified by municipality operator that produce questionnaires discarding, by type; iii) gap between theoretical and final sample size; iv) number of follow-up to households and municipalities, by type of problems.

Computer assisted data entry based on inconsistency rules and acceptable ranges helps operators avoiding typing error.

Automatic editing and imputation is carried out through several steps. First, equivalence between the number of records and the number of questionnaires is checked. Then conformity of data to manual revision rules is checked with a deterministic program (if-then), and partial non-responses and inconsistencies not treated in previous phases are solved with a complex imputation system. Different checks for qualitative and quantitative variables (expenditures) are carried out, the former by means of a probabilistic program with donor's imputation; in the second, probabilistic programs are used only when the frequency of observed expenditures for a specific item is sufficiently high to assure an adequate number of donors, otherwise, imputation is made with a random value selected in a specific range.

The following indicators can be calculated: i) percentage of records discarded due to not acceptable qualitative level (they should be discarded during the manual revision process); ii) percentage of records imputed on the basis of revision rules (not attended during the manual revision process). In order to check the effect of the automatic imputation process, other indicators are: i) n° of records by n° of imputations on qualitative variables; ii) n° of records by n° of imputations on quantitative variables; iii) discrepancies between distributions of the main variables before and after the revision.

Final check is finalised to identify households that, in spite of the treatment, still present anomalous expenditure behaviour: a smaller weight in final estimate is assigned to these households. The following indicators are useful to achieve a better comprehension of trends over the time: i) number of households with anomalous expenditure behaviour (outliers); ii) outliers mean expenditure vs. non-outliers one; iii) variance of outliers mean expenditure; iv) outliers associated weights.

3. Analysis of indicators

Indicators described in the previous section constitute the minimum set of information needed for monitoring the checking phase but not all are available at the moment: some of them arise from the current production activity but some others need new instruments and actions to be introduced into the production process. An effort to improve monitoring process of manual revision has been performed (it will be implemented from the beginning of 2002): more information on the output of checks and field feed-backs will be collected in order to recover missing information. The main checks at this stage are: i) difference between theoretical and final sample size; ii) number of discarded questionnaires; iii) errors concerning the field work not detected by municipality operator, by type.

Difference between theoretical and final sample size is about 12,3% in 2000 and 10,8% in 2001, but values are very different among Italian regions and quarters during the year. The proportion of discarded questionnaires out of those delivered to Istat is about 4,3%-4,6% in 2001 and the proportion of questionnaires not delivered to Istat ranges from 5,8% and 7,3% among quarters in the same year. Some error typologies in the field work have been found in 56% and 45% of municipalities (respectively in the first quarter of 2000 and 2001), while the proportion is lower in the remaining quarters.

Automatic editing and imputation is another crucial step of data processing, due to its complexity and impact on individual data. It was possible to calculate: i) n° of records discarded due to severe failing of editing rules and missing data; ii) n° of records by n° of imputations both on quantitative and qualitative variables. The number of records discarded range from 9 to 25 in 2000 and from 9 to 16 in 2001. This evidence shows that problems are not too relevant considering that about 6000 record are processed, but also some difficulties in the manual check. The proportion of records that do not need imputation in demographic and economic qualitative variables is about 90%-93% in 2000 and 2001 but the value diminishes to 80% for the other qualitative variables (house furniture, purchase of services and durable goods). Considering each expenditure category (everyone includes many items), the proportion of records (considered) correct is generally high (more than 90% and 95%) in each quarter in both years (2000 and 2001). This is not true only for food and another residual category, where the proportion of correct values is lower (about 87%). Nevertheless, when all single expenditures

(aside from the category to whom they belong) are considered, the proportion of records not imputed is lower (about 60-64%) in both years.

These few evidences show that some adjustments are necessary. In particular it is to be stressed the variability among quarters or between the two years considered.

Others and more specific indicators have to be calculate to get a better insight of problems in data revision. Among others: records imputed distinguishing by reason of imputation (failure of editing rule or missing data), difference in the total expenditure between raw and clean data, number of false interviews and number of interviews incorrectly carried out (information from a telephone survey conducted on a sub-sample of households).

4. Conclusions

Developing an information system to assist in monitoring and controlling of production process is fundamental to assure continuous quality improvement in statistical information. The work reviewed in this paper is the first step to this direction. Understanding the production process, outlining operations and their sequence, identifying key variables, indicators and measures of performance are the basic elements of the information system.

The second, not less important, step is implementing an effective process control. At the moment, the most part of indicators has requested ad hoc treatment of processing outputs or special recording of information collected on paper. Others and more specific indicators need to be provided, meta-data and further information collected at different levels of data processing need to be recorded and currently calculated during the production process. Moreover, continuous updating of the information system according to survey changes is needed.

References

- Ceccarelli C., Coccia G., Grassi D. Pannuzi N. (to be published) Il trattamento dei dati nel processo di produzione dell'indagine sui consumi delle famiglie, *Rivista Scritti di Statistica Economica*, Vol n° 8, Liguori editore, Napoli.
- Filippucci C., Calia P. (2001) Towards process quality control in the Italian Consumer Expenditure Survey, *Proceedings of the International Conference on Quality in Official Statistics*, May 2001, Stockholm.
- Morganstein M., Marker J. (1997) Continuous quality improvement in statistical agency, in: *Survey measurement and process quality*, Lyberg L., Biemer P., Collins M., deLeeuw E.D., Dippo C., Schwarz N., Trewin D. (Eds.), J. Wiley & Sons, 475-500.