

Methods for the Reduction of Statistical Burden in Istat Surveys

Metodi per la riduzione del carico statistico per le indagini ISTAT

Di Consiglio L. Falorsi S., Solari F.
Istat, Via Magenta 4, Roma email: diconsig@istat.it

Riassunto: Il coordinamento negativo delle indagini risponde all'esigenza di introdurre nei meccanismi di estrazione casuale per la selezione dei campioni sistemi che permettano di tenere sotto controllo il numero di unità della popolazione incluse in più campioni. Scopo della riduzione della sovrapposizione tra i campioni è il controllo del disturbo statistico, ossia una miglior distribuzione tra le unità della popolazione dell'onere che ogni indagine comporta per le unità inserite nel campione (disturbo o carico statistico). Intervistare un'unità più volte in un ristretto intervallo temporale potrebbe indurre a fornire risposte imprecise quando non il rifiuto di rispondere. Una miglior distribuzione del disturbo statistico può portare notevoli benefici al sistema delle indagini sia in termini di precisione delle stime che rispetto alla diminuzione dei fattori distorsivi legati ai fenomeni della mancata risposta totale e parziale.

Keywords: Coordination of samples, Statistical Burden.

1. Introduction

Sample coordination in statistical surveys (in particular in business surveys) comes up to the need to introduce in the mechanisms of random selections some method to control the number of units included in different samples. This paper focuses on negative coordination whose goal is to minimize the overlap of different samples, in order to reduce the survey burden on sampled units. In fact, the recurring selection of the same unit in many surveys may produce an higher propensity to unit or item nonresponse or a decrease in the quality of responses, so that biases can arise in result of the surveys. For these reasons negative coordination has been recently introduced in many statistical institutes in order to distribute the burden more uniformly among units in the population of interest.

A research working group has been instituted in Istat in order to verify the possibility of reduction of the burden in the system of business surveys. To this aim the most important methods for sample coordination introduced in many different countries have been studied: the JALES method in Sweden (Atmer *et al.*, 1975), the EDS methodology in Netherlands (De Ree, 1983), the Ocean system in Australia (Cotton and Hesse 1992), the microstrata methodology (Riviere, 2001). For a detailed presentation and description of the methodologies for sample coordination, see for instance Hesse (1999).

An important outcome of the research working group is a software prototype implementing the microstrata methodology. This software has been successfully used in a simulation study carried out on a set of Italian business surveys (Istat, 2005).

2. Coordination methods

The coordination method can be classified according to their use of response burden. The first class of methods is based on the use of permanent random numbers associated to each unit in the

population. Coordination of samples is guaranteed by proper use of starting points in the selection of units.

A second class of methods, instead, makes explicit use of a function of response burden. In order to eliminate any possible bias in the selection of units due to different survey stratifications, *building blocks* (see De Ree 1982) or *microstrata* (Riviere 2001) are defined with units whose survey *history* is the same. More precisely, both building blocks and microstrata are elementary partition and all the strata in the surveys have to be union of partition sets. Then, units are ordered with respect to their cumulative burden within the same building block or microstratum. Building blocks partition is the same for all the surveys and, therefore, the survey design is constrained with stratification based on the building blocks. On the other hand, microstratification changes for each new survey to be selected and it is obtained cross-classifying the strata of the set of surveys we would like to be correlated with. This set, generally, refers to the surveys which have been carried out in a specific period of time ending at the time we have to select the new survey. The probably small size of the microstrata arising from sets referred to a large number of surveys may affect the effectiveness of the method. In order to overcome this problem, we can define several microstratifications referred to different sets of surveys. If the aim is the coordination with all the surveys in a given period of time, a full analysis should first define the microstrata on the last survey, then on the last two and so on, ending with the last microstratification based on all the surveys in the period. Of course, computer time constraints avoid us to carry out all the possible microstratifications, so that we have to restrict only to some of them. Riviere (1999) suggests to carry out only permutations based on the last survey and all the surveys carried out in the current year and in the last two years. A necessary condition to preserve the original balance is that in each permutation, microstratification and cumulative burden have to be coherent in the sense that they have to refer to same set of surveys.

3. Description of the algorithm for the microstrata method

Let A_t be the sample to be selected in t and r_t the random numbers vector associated to units in t . Furthermore, if i_t denotes the sets of the surveys selected in a given period of time, let P_{i_t} and C_{i_t} be the partition in microstrata and the cumulative burden for the surveys in i_t . The microstratification P_{i_t} is generated by the intersection of the strata for the surveys in i_t and C_{i_t} is the cumulative burden which takes into account the selections in the same set. P_{i_t} and C_{i_t} are said to be consistent if they refer to the same set of surveys. The different sets of surveys used for permutations are the previous survey $\{A_{t-1}\}$, the sets i'_t and i''_t of the surveys carried out in the current year and in the last two years, respectively. The selection of A_t is performed by means of the following steps:

1. ordering w.r.t. the previous survey A_{t-1}
 - 1.1. order units within microstrata of $P_{\{A_{t-1}\}}$ w.r.t. increasing $C_{\{A_{t-1}\}}$ and random numbers;
 - 1.2. permute r_{t-1} w.r.t. the order obtained in 1.1; let $r_{1,t-1}$ be the new random numbers vector;
2. ordering w.r.t. the surveys in i'_t
 - 2.1. order within microstrata of $P_{i'_t}$ w.r.t. increasing $C_{i'_t}$ and random numbers;
 - 2.2. permute $r_{1,t-1}$ w.r.t. the order obtained in 2.1, let $r_{2,t-1}$ be the new numbers vector;
3. ordering w.r.t. the surveys in i''_t
 - 3.1. order within microstrata of $P_{i''_t}$ w.r.t. increasing $C_{i''_t}$ and random numbers;

- 3.2. permute $r_{2,t-1}$ w.r.t. the order obtained in 3.1; let r_t be the new numbers vector;
4. selection of the sample units of A_t
the sample selection is obtained by ordering units within the strata of A_t according to increasing r_t . In each stratum of A_t , units with smaller random numbers are included in the sample.
- For all new units in t , stratum values missing (or zero) is assigned for all time periods previous to t . This implies that for the selection of A_t all the new units are included in the same microstratum. For the selection of the first survey in the system of coordination, assigning the starting random numbers r_0 to the units, being the sets A_{t-1} , i'_1 and i''_1 empty, no permutation is needed and r_1 is equal to r_0 .

4. An application

The efficacy in reduction of burden of the methodology under study depends on the effectiveness of permutations within the microstrata defined in real practice. For this reason, Istat has carried out a large scale study covering the main Istat short-term and structural business surveys for a three years period. In this study microstrata methodology has been compared with the simple random sampling and JALES techniques. The results shows that microstrata performs better in terms of reduction of statistical burden (see Istat, 2005).

Here we report a simple example in which the microstrata method and simple random sampling are compared for sample selection from an artificial population by means of a simulation study. The study has been undertaken in order to compare the performances of microstrata methodology and simple random sampling in terms of: reduction of burden (Table 1); preservation of the population renewal rate in the sample (Table 2) and bias and mean square error (MSE) of the Horvitz-Thompson estimator (Table 2).

To this aim two surveys, A and B, are supposed to be carried out during two different years. The artificial population is composed of 1,000 units for the first year and 1,050 for the second one. Survey A uses a stratification of units with two strata, survey B with three strata. Strata code is randomly assigned with probability 0.8 and 0.2 in the first case and according to a uniform distribution in the second. Strata are supposed not to change for units from year 1 to year 2.

A discrete variable y_1 and a continuous variables y_2 are associated to each unit. The discrete variable y_1 is Poisson with mean generated by a χ^2 distribution with degrees of freedom equal to 10 in stratum A1 and 500 in stratum A2. The continuous variable y_2 , instead, is lognormal with $\log(y_2)$ having mean given by the log of the mean of y_1 and square root error equal to 0.3 times its mean. Stratified sample is conducted for both surveys; sample fractions equal to 0.05 for the first stratum and 0.2 for the second one for survey A, while the sample fraction is 0.06 for all the three strata of survey B. New values are generated for the units in the second year for both variables.

A thousand samples have been drawn and for each of them sample mean for y_1 and y_2 has been evaluated. The function of statistical burden has been assumed constant equal to one.

Table 1. Distribution of frequencies for the maximum number of selection of sampling units.

Max number of selections	Microstrata	SRS
1	1000	0
2	0	195
3	0	779
4	0	26

Table 2. Renewal rates in population and average renewal rates for microstrata method and srs.

	Stratum	Population	Microstrata	SRS
survey A	A1	0,047	0,048	0,047
	A2	0,050	0,050	0,050
survey B	B1	0,044	0,045	0,044
	B2	0,047	0,047	0,049
	B3	0,051	0,053	0,053

Table 3. Bias and CV for microstrata method and srs.

		Microstrata		SRS	
		Relative Bias	Relative MSE	Relative Bias	Relative MSE
survey A	y1 year 1	0,001	0,007	0,007	0,007
	y1 year 2	0,013	0,090	0,024	0,008
survey B	y2 year 1	0,019	0,708	0,012	0,722
	y2 year 2	0,032	0,881	0,013	0,836

Table 1 reports the frequency distributions for microstrata and simple random sampling of the maximum number of times a given unit has been included in the sample over the two years period. It is clear that, while the microstrata methodology succeeds to uniformly distribute the burden over all the units, in the simple random sample case we have that in 26 runs at least one unit has been included always in the sample and that in no runs we have units include in the sample only one time. The results in table 2 show that, as well as simple random sampling, microstrata preserves the population renewal rates for all the strata in survey A and survey B. Finally table 3 displays the performance of Horvits-Thompson estimator in terms of relative bias and MSE. Since the similarity of the measures in microstrata and simple random sampling cases, we can deduce that the microstrata methodology does not introduce any kind of bias in the estimation process.

References

- Atmer J.G., Thulin G. and Backlund S. (1975) Coordination of sample with the JALES technique, *Statistik Tidskrift*, 13, 443-450.
- Cotton F, Hesse C. (1992) Co-ordinated selection of stratified samples, *Proceedings of Statistics Canada Symposium*.
- De Ree J (1983) *A system of co-ordinated sampling to spread response burden of enterprises*. Prepared by the Netherlands Central Bureau of Statistics (CES/AC 48/43) for the Conference of European Statistician (United Nations)
- Hesse C. (1999) *Sampling co-ordination: a review by country*, INSEE, working paper E9908.
- Istat (2005) *Documentazione della metodologia e della sperimentazione per il coordinamento dei campioni* Prepared for Direttiva metodi e procedure per il coordinamento dei campioni delle rilevazioni sulle imprese
- Riviere P. (1999) Coordination of samples: the microstrata methodology, *Proceedings of the 13th Roundtable on Business Survey Frames*, 27 September-1 October 1999, Paris.
- Riviere P. (2001) Coordinating samples using the microstrata methodology, *Proceedings of Statistics Canada Symposium*.