Sampling designs for small domains estimation through multi-way stratification techniques

Piero Demetrio Falorsi   Stefano Falorsi   Paolo Righi   Fabrizio Solari
falorsi@istat.it   stfalors@istat.it   parighi@istat.it   solari@istat.it
Istituto Nazionale di Statistica – Via C. Balbo 16 – 00184 Roma

The small area problem is usually thought of as one to be dealt with via estimation. However, there are opportunities to be exploited at the survey design stage. In this framework it is crucial to control the sample size for each domain of interest, so that each domain is treated at design stage as planned domain, for which it is possible to produce direct estimate with a prefixed level of precision. In general, this level of precision is useful to keep under control the variance of the direct estimator but it does not guarantee reliable direct estimates. In this paper, the small area problem is dealt with considering the design phase. Some techniques that allow to control the sample sizes for domains of interest which are defined by different partitions of the reference population are presented. Such techniques are useful when the overall sample size is relatively small and by consequence in some of the partitions there are small domains.

When the objectives of the survey is to produce estimates for two or more partitions of the population a standard solution to obtain planned sample sizes for the domains of interest is to use a stratified sample with the strata defined by cross-classification of variables defining the different partitions. In the following this design will be denoted as *cross-classification design*. In many practical situations, the cross-classification design is often unfeasible since it needs the selection of at least a number of sampling units as large as the product between the number of categories of the stratification variables. In order to overcome this problem, an easy strategy is to drop one or more stratifying variables or to group some of the categories and, consequently, small-area estimation problems become more serious since some planned domains become unplanned and some of them can have small or null sample size.

Many methods have been proposed in the literature to keep under control the sample size in all the categories for all the stratifying variables. These approaches may be roughly divided into two main categories or contexts. The first context contains the methods mostly known in the literature as *controlled selection*. They allow to satisfy the sample size planned for each domain of interest without using the cross-classification of the stratifying variables. In the second context there are methods based on *sample coordination*. A separate sample is selected for each partition trying to guarantee the maximum overlap among the different samples. The methods of both contexts avoid to fall into some of the problems previously described.

The aim of this work is to offer a general overview of the techniques allowing to control the selection on the separate stratifying variables and to give account of some recent methodological proposals. The methods described are particularly useful when dealing with small domains or small area problems. In fact, in this situation sampling from the cross-classification of the partition of domains is likely to be unfeasible because the resulting stratification defines a too fine partition of the population.