



STATISTICAL
RESEARCH
CENTER

Master Sample and Its Application in Sample Surveys Particularly in the Household Surveys

(Abbas Gerami, Mohammad Bagher Sekhvat,
Fatemeh Harandi, Zohreh Fallah Mohsenkhani)

**Technical Designs and Statistical Methods Research Group
Spring 2002**

Master Sample and Its Application in Sample Surveys Particularly in the Household Surveys

1. Introduction

Estimation of a statistical population's parameters, covering various characteristics is very costly through conduct of statistical surveys. This is due to many factors including construction of the frame, listing, and geographical level of the required estimates. Construction of frame is generally based on a census the implementation of which is complicated and expensive. Also, number of information items collected from a census is limited. A census is usually executed every 5 or 10 years. For this reason, sample surveys are designed and conducted as a scientific and practical procedure in order to estimate the population's parameters in different years. In a sample survey there is usually a contrast between the estimation error and the survey's cost so that the more we are willing to have a smaller estimation error the more cost we should incur (though there may be no exact reverse relation between the two). On the other hand, we may not reach so accurate estimates though enjoying a considerable fund. Therefore, a balance should always be made between them or one sacrificed for the other. It is evident that applying appropriate sampling methods and considering other executive problems can have a vital role in reducing the costs and in improving the accuracy of the estimates.

One of the ways to reduce costs is to divide such costs as preparation of a frame between several surveys or different survey periods. The other way is to select a fixed set of the sample units referred to as Master Sample for use in various surveys or different rounds of a same survey. In applying the Master Sample, some of initial steps together with a heavy workload can be taken adequately or inadequately for various surveys in a mixed procedure. In addition, it is possible to establish a relation between the estimates resulting from such surveys by applying this procedure. The use of Master Sample was first focused on by a certain number of countries during the second half of the 20th century and it is now being used by them. The present study has considered the matter briefly.

The Statistical Center of Iran as the major organization producing and coordinating the official statistics in the country has carried out many sample surveys in addition to various censuses since its establishment. There are some common aspects in a number of such surveys, particularly those dealing with households' characteristics, in which it is possible to use the Master Sample.

In this connection, a project, with an aim to study the possibility of using the Master Sample in household surveys, ordered by the Statistical Research Center was defined and carried out. The report of this study has been presented in three parts.

The first part introduces the Master Sample including its history, applications, designing and updating the Master Sample Frame as well as defining the Master Sample and its key characteristics, advantages, disadvantages and limitations and finally use of the Master Sample in the periodic and multi-round surveys. The necessary recommendations for the Master Sample have been made at the end of this part.

The second part with the title, “Designing a Master sample for Use in the Statistical Center of Iran’s Household Surveys” deals with the preparation of the samples required for the Household Surveys of the Center during 2002-2006. In this respect, first use of the 1996 General Census of Population and Housing was evaluated for designing the Master Sample Frame, however the data of the census were not found reliable for the MSF because of their being old and unupdated. Then, the frame obtained from the 2002 General Census of Establishments was studied as a substitute procedure for the MSF, but this frame was also found lacking in the essential attributes for this purpose. Furthermore, the current household surveys of Statistic Center of Iran were reviewed and it was concluded as such that except for the household’s income and expenditure survey, it is not possible in other cases to forecast the titles of the household surveys and their characteristics for a period of several years. Finally, it was, based on the studies carried out, evident that lack of an integrated household surveys program is the major issue for failure in designing the Master Sample.

Considering the studies presented in the second part of the report and with regard to the necessity of cost-effectiveness in application of the Master Sample, this research has not been successful as expected in terms of an adequate designing of a MS with an aim to draw the samples required for the Statistical Center of Iran’s household surveys during 2002-2006 and its comparison with the existing methods for designing of the samples. One of the important obstacles is the absence of a long-run extensive program for the Center’s current household surveys. Apart from this major issue, this research presents next the integrated programs for the design and conduct of the survey. This has been discussed in the third part.

The third part entitled, “Designing the Integrated Household Surveys Programs”, describes the factors involved in designing such programs in addition to the points which should be considered when planning for the surveys and the matter of cost-effectiveness in applying MS to the related programs.

2. Design of the Master Sample using the data from the 2002 General Census of Establishments as a frame

a. The Master Sample Frame and duration of its use

The list of blocks and villages throughout the country obtained from the 2002 General Census of Establishments in urban and rural areas constitutes the MSF. This list is made from the data included in form 1 (listing form).

In addition to the above file, the maps of the enumeration districts are also accessible. These maps have been segmented in blocks in the urban areas and mapped rural areas. In the other rural areas the location of a village has been marked with a dot. MSF is used for four years (2003-2006).

b. Uses of MS

1. Employment and Unemployment Characteristics Surveys (making seasonal estimates annually with rotation sampling and distribution of seasonal samples between the three months of each season).
2. The Urban Households’ Income and Expenditure Survey (once a year with distribution of annual samples between the twelve months).

3. The Rural Households' Income and Expenditure Survey (once a year with distribution of annual samples between the twelve months).
4. The Household's Socio-Economic Characteristics Survey (a follow-up survey beginning from 2004 and a two-round visit in the subsequent two years).
5. Population Changes Measurement (once in a period of four years and a two-round visit during sixth months).
6. Ad hoc household surveys

The implementation of the first three surveys is definite but the rest indefinite.

c. Geographical level of estimates

The estimates are made at the urban and rural areas level in each province for such surveys as employment and unemployment characteristics, households income and expenditure, and population changes, and at total country level for the household's socio-economic characteristics surveys. However, the geographical level of estimates for the ad hoc surveys is not determined.

d. Primary Sampling Units

The Primary Sampling Unit (PSU) is composed of blocks or villages or a combination of small blocks or villages in a neighboring enumeration district, or a part of large blocks and unmapped large villages with A households at the minimum and B households at the maximum.

To construct PSU of small blocks or villages, the integration is made to reach the minimum number of the households using the blocks or villages within an enumeration district and, if required, in neighboring enumeration districts.

However, the village's blocks are used in case of the mapped villages for this purpose.

The ways to construct PSU for large blocks or unmapped large villages are as follows:

1. Blocks or unmapped villages with at least A households and at most B households would constitute a PSU. The optimal values A and B are determined with regard to the number of households in blocks and villages throughout the country in 2002, pattern of sample rotation in the periodic employment and unemployment characteristics surveys as well as correlation coefficient within PSUs.
2. Blocks or unmapped villages with over B households will be subjectively divided in several segments all with an equal size. Each segment is regarded as a PSU so that the whole objective or subjective PSUs would cover at least A households and at most B households.

After construction of PSUs in the frame (whether objective or subjective), the units can be divided into strata in terms of urban and rural areas, the population in the cities and Dehestans (administrative subdivisions), and number of their households. The boundary of the strata may be determined by quantiles of attributes distribution obtained from the results of the 2002 General Census of Establishments.

- Ultimate Sampling Unit: the sample household's housing unit

e. Sampling Design

Two-stage cluster sampling for all surveys

f. Type and rate of overlapping of PSUs and USUs

PSU: if possible the same for all designs

USU:

- Various USUs are employed for the employment and unemployment characteristics, the household's income and expenditure, the household's socio-economic characteristics and ad hoc surveys.
- The same USUs are used for the employment and unemployment characteristics and the population growth measurement surveys in the related years.
- A part of the USUs remains constant and another part changes due to the pattern of the rotation of the sample in various seasons of the employment and unemployment characteristics survey.

g. How to select the units of each stage

First, the PSUs are selected by the PPS method in a required number within each stratum. Then, sample PSUs in each stratum are divided into a number of random sub-groups (in a systematic manner). The number of the sub-groups is mostly determined according to the number of sample PSUs needed for the employment and unemployment characteristics survey in each year. The first sub-group is used in the first year of MS application. The housing units and households are listed in each sample PSU with the selection of random starting point. In listing process, such data as job specifications and literacy of the head of the household are asked so as to group PSU households appropriately. In cases where a PSU is subjective (a part of a large block or an unmapped village), the total number of housing units and households of the block or village relating to this PSU will be counted. Then, with reference to the division number of the sample PSUs in a block or village, the part of households pertaining to the sample PSU can be determined and listed. Considering the fact that most large villages are mapped, it is mostly the case with the blocks covering housing complexes whose households can be counted rather easily.

As already mentioned it may not be necessary to list again the housings units and households of sample PSU at least in the first year of the duration of MS application when the existing lists from the census can be of use. Particularly knowing that the USU is the housing unit of the sample household not the household itself, the probable changes in the lists can only be due to the change of the sample PSU housing units. Of course, in the latter case, a part of the survey's efficiency that arises from grouping of PSU households based on the characteristics of the head of the household and the other household's socio-economic characteristics and that can be achieved together with the listing of sample PSU households might be lost.

The households listed in the sample PSU in each stratum are divided into a number of clusters systematically with regard to the average size of the cluster in the stratum. If the households' socio-economic characteristics are recorded in listing process, it is then possible to arrange the PSU households based on such characteristics in an appropriate procedure and to build the sample cluster too. The average size of the

cluster in strata can be determined according to the intracorrelation coefficients of PSUs of each stratum and the components of the survey's cost in the same stratum. Depending on the results obtained the average size of the cluster in all strata can be assumed equal or different. In the next process, the clusters made are divided into a number of random subgroups. The number of the random subgroups are assigned according to the sample rotation model in the employment and unemployment characteristics survey, the number of the sample clusters required for the same survey, the household's income and expenditure survey and the number of additional clusters predicted for the household's socio-economic characteristics survey and the ad hoc surveys during the MS application in each year. For each survey, the household's income and expenditure survey as an example, one or more random subgroups are employed to provide the required samples.

h. Master Sample's Size

The number of the sample PSUs are assigned as follows:

First, the number of the sample PSUs required for each survey is separately determined. Then, the final number of the sample PSUs needed for each survey is specified according to the rounds of conduct of surveys during a 4-year period. Finally, considering the way PSUs are overlapped in various surveys, the number of the sample PSUs of a survey that enjoys the maximum sample size is regarded as the basis. A percent is added to this number for satisfying any probable needs, thus the number of final sample PSUs can be obtained.