"Dimitrie Cantemir" Christian University
Knowledge Horizons - Economics
Volume 7, No. 1, pp. 67–70
P-ISSN: 2069-0932, E-ISSN: 2066-1061

www.orizonturi.ucdc.ro

RESEARCH BY SAMPLING, BASIC METHOD IN THE STUDY OF SOCIAL AND ECONOMIC PHENOMENA, BETWEEN TEACHING AND PRACTICE

Claudia Gabriela BENTOIU[1], Cristina ATHU[2], Graţiela GHIC[3]

[1]*Faculty of International Business and Economics, "Dimitrie Cantemir" Christian University, Bucharest, Romania, E-mail: bcg@yahoo.com*
[2]*Faculty of International Business and Economics, "Dimitrie Cantemir" Christian University, Bucharest, Romania, E-mail: cristinathu@yahoo.com*
[3]*Faculty of International Business and Economics, "Dimitrie Cantemir" Christian University, Bucharest, Romania, E-mail: gratiela72@yahoo.com*

Abstract *Statistical research by spot checks is used with great results in almost all economic and social research on a random basis. The social field and especially the living standard and the life quality scope, due to their complexity, sometimes require as the sole alternative to study phenomena – research by sampling, as sometimes there may occur situations when exhaustive statistical research is impossible to be carried out or the costs for achieving the goal of such a thorough research would be too high. Also, for the sake of clarity, we find it necessary to resort to a methodological approach, as it is both didactic and practical*

Key words:
Research by sampling, identifiers, survey errors, choice „at random"
JEL Codes:
O4, R1, C81

1. Introduction

Among the benefits providing a research by sampling we can lists:

- The time for carrying out a selective research is much smaller than the time for achieving the goal of total research;
- Use of a smaller number of qualified personnel, which means reduced costs;
- The opportunity of exercising a thorough inspection on the way of collecting and processing the information in order to remove the registration errors encountered in total research.

All these benefits represent nothing but *the research costs decrease and the efficiency increase, actually meaning that the time of the data acquisition is reduced.* Sample survey allows deepening of areas which may not be achieved only by census. However there can be made a systematic comparison of censuses and surveys, because the two ways of collecting information have complementary aspects.

The only issue of the selective research is that of scientific strictness in carrying out this research. In a selective research it is very important to define the population of interest and especially the choice of sample, in such a way that, whenever we want to characterize the entire population (in the sense of collective assembly) where from this sample was extracted, we may get its most accurate picture. It can be assumed that a population is defined by means of the four factors: *its nature* (an individual, a dwelling), *the intrinsic characteristics* (gender, dwelling type), *its location* and *the time* referred to. Any selective research may be classified as a total scientific research relying on rules and scientific methods if it starts, first of all, to establish a so-called plan of selective research. Although there are two main types of survey, *probabilistic surveys* (those for which each individual has a given likelihood, known in advance, that they belong to the sample) and *empirical surveys* (which do not allow to calculate probability of inclusion in the sample), statistics most often uses probabilistic polls because they are far more rigorous. A selective research may be considered to have achieved its purpose solely when the obtained results have been successfully expanded over the entire population in our survey. Extension can be considered to be the backbone of any statistical surveys.

2. The Survey Basis Properties

To be able to make an accurate probabilistic extraction, i.e. a drawing out where, by definition, each individual of the population has a known probability, previously fixed to be part of the survey sample, it is imperative to have, first of all, a list of all of the sampling units belonging to the survey field (able to be investigated because they belong to the population going to experience the extension). A genuine survey basis must meet a series of conditions, namely:

↳ Any survey basis must allow the unit marking with no ambiguity. A good survey basis must be a list of *high quality identifiers*. But what does actually a "statistical identifier" refer to? For example, if it is considered that statistical data of a census of the

population and dwellings represents a survey basis, then, an identifier can be a dwelling from the census file or an individual belonging to the same file. This dwelling can be identified by: locality, street, the building it belongs to and its number in the respective building. The identification data for the individual are: surname, first name and address where he lives, with the following characteristics - locality, street name, the number of home building. Otherwise we may risk to take an individual for another and to introduce a wrong estimator unknowingly, changing the likelihood of drawing out the individuals of the population initially assigned. If the identifier is not very clear, the investigator may waste a lot of time when he wants to identify him in the survey basis;

- A sample basis must be exhaustive. This means that each unit belonging to the population we want to assess must belong to a certain list of identifiers. We must be very careful to have each unit investigated.
- A sample basis must be accurate, avoiding inclusion of certain units, which for various reasons, should not be included (demolished houses, streets that have disappeared etc.);
- A good basis of sampling survey must be kept free of any repetition, thereby avoiding inclusion of a unit or a specific individual twice, or the same individual's characterisation by two different identifiers. Otherwise we obtain a changed estimator: in case the extraction is done with equal probabilities, an individual occurring more than once in the basis is more likely to be selected than the one that exists in the list only once;
- A survey basis must also be as current as possible, because, even if at the time of its composition it might have not been complete and accurate time may „erode" it.

What can we do when, regardless the reason; we don't have the basis of a good survey to allow us conduct a high quality opinion poll?

- On the one hand, we do not use a sample basis and we get outside the rigorous framework of the probabilistic survey, and therefore we make an empirical survey;
- On the other hand, we are looking for survey databases, only of the individuals (sampling units or observation units) directly able to provide the information, however going through an intermediate level of individuals' groups. Performing a drawing out on several levels and doing intermediate censuses in the selected groups we can rigorously take samples of the population units that concern us. This is frequently done on areolar type surveys, where the geographical areas are to be sampled first;
- In case we are interested in the topic that it is suitable for the information collection, through an intermediate population of the observation units, one different from the individuals of the population undergoing an extension, and for which we have a sample basis, the sampling and observation units are inevitably different in this situation and we will face a sensitive issue of correspondence between the two types of units.

To overcome the lack of exhaustiveness in an existing but incomplete database there may be used private techniques such as those of half opened intervals. The principle is simple: if people can be ordered in a natural order ascending through its identifiers and if the identifiers list is not complete, there may be imposed (when an individual has been extracted) the systematic investigation of all individuals whose on-the-spot identifiers are included between the identifier extracted from the basis (which is itself investigated) and the identifier of the basis which follows immediately the extracted identifier (which, instead, is not investigated).

3. Errors of the Research by sampling

In order to be representative for the entire population, the results of the survey there must be no errors in the process of sampling.

To ensure that the results of the survey can be used, these errors must be very insignificant as under the conditions when the survey was done correctly, its values may merely represent population approximately.

In the broadest sense, a sampling error represents the deviation between the values calculated by processing the data in the sample and the results which could have been achieved if a total observation had not been performed and the data from all the population units had been processed.

The patterns of errors encountered in the opinion polls vary from errors common to all types of observations – recording errors- to errors specific to the surveys - errors of representativeness.

As a general rule, in the opinion polls, recording errors occur in few cases, and can be easily removed by a thorough inspection. The errors of representativeness specific to surveys are of two kinds: systematic errors and random errors.

The systematic errors of representativeness can be avoided if there are strictly respected the principles of the survey theory obviating the causes leading to their occurrence.

The main causes of the *systematic errors* are:

- the choice of deliberate data wrongly considered as representative;

- the choice " at random" of the elements of opinion poll. The choice " at random" is not synonymous with random choice. The random choice involves knowledge of the approximate relative frequencies, the degree of proximity dependent on number of comments, in accordance with the law of big numbers;

The choice *"at random"* very rarely results in a selection that can be considered to be representative; it is too largely subject to the systematic errors.

- the scientist's preconceived desire to obtain a specific result;
- the substitution of a research unit or one that offers greater convenience in obtaining the data for the scientist;
- incomplete inclusion in the survey of the research units. If, for example, the dwellings where the lessee has not been found on the first visit will no longer be visited the second time, hence systematic errors without substitution will occur.

The census takers' indolence, the failure to return in time the filled-in questionnaires or the non-responses of some units included in the sample are the main causes of systematic errors. They distort the results and, in order to be avoided, it is necessary to strictly observe the rules stipulated in the opinion poll theory. However, it's worth mentioning that in a survey the systematic errors are less numerous and less serious than those occurring in the total observation.

Unlike the systematic errors, *the random selection errors* or *of representativeness* occur in the survey process itself; they are present even if it complies with the stringent principles of the selection theory, because the sample reproduces with a vague approximation the distribution of the population's variables. As a result, the arithmetical mean, or any other value of the sample will present certain deviations regarding the population parameters. To ensure that the results of the survey may be useful, the errors introduced by the selection process itself must be much smaller. The simplest process which obviates the random or representativeness errors is the increase of the sample volume and the selection of the most appropriate survey type for research purpose.

Although the errors of representativeness cannot be avoided using statistical techniques they can be calculated beforehand if this survey is probabilistic. The assessment of the population parameters will be made on the basis of the indicators resulted from the processing of the survey data, with a random error of representativeness which can be found within a certain period probabilistic calculation. Hence each derivative or synthetic indicator must be connected to its error of representativeness so that it could be generalized to the entire population.

When the sample is correctly selected (removing all causes of systematic errors), the average value of the random errors or representativeness ones of the sampling or even the expected frequency of such errors of different sizes can be calculated according to the survey actual values. The results of these calculations can be used to comparatively estimate the effectiveness of various processes applied by studying the same population and, therefore, for the sake of better research planning.

Usually in a survey the errors of representativeness can be calculated as *actual errors and prospective errors*.

The actual errors of representativeness can be calculated only for the features with certain obtained data and after a total observation. Considering that even in this case the arithmetical mean is the most representative synthetic indicator, the actual sampling error is calculated as the difference between the sample mean and the average population's mean. The calculation of the actual sampling error represents nothing but the degree of representativeness of a sample in relation to the population's structure.

Nevertheless, in practice, the reproduction of a structure identical to that of the population is only accidental, which means that the extracted samples may have different degrees of representativeness. Therefore, before processing the collected data there will be checked the sample representativeness.

The features used in the sample selection has usually various forms of manifestation, hence the sample representativeness verification is not always an easy operation.

For the verification of the sample representativeness we should start to compare the groups' structure of the sample or of the population, the so-called programmed structure. In case these structures do not differ by more than $\pm 5\%$ the sample constituted as representative is accepted.

4. Conclusions

Taking into account that the methods of research in the economic field are getting more and more refined, the analysis of the socio-economic mass phenomena acquire new dimensions. Series of data accumulated to quantify the characteristics of the socio-economic life, the progress made in the field of statistics, informatics, econometrics and data analysis have directly contributed to an increase of the researches in this field and hence in order to characterize these phenomena it is imperative to use the series of data obtained either after performing exhaustive observations or on the basis of selective and specific research, namely the research by sampling.

Bibliography:

1. Biji E.M., Gogu E (2009) : „Teoria şi tehnica sondajului statistic", Editura Oscar Print, Bucureşti
2. Institutul Naţional de Statistică (2011): *Repere economice şi sociale regionale. Statistică teritorială*
3. Institutul Naţional de Statistică (2013): Ancheta bugetelor de familie
4. *** www.insse.ro
5. *** www.gfk-ro.com
6. *** www.iccv.ro