

USE OF SAMPLING METHOD, I

Dmitri TERZI

CZU: 338:303.5

dgerterzi@yandex.ru

The use of the selective method is increasing in various new situations of socio-economic activity, for example, in the agricultural sector of the economy, in population statistics (for conducting a census, estimating employment), in social statistics to survey household budgets, in consumer price statistics, for a survey of wages, for tax audits, in marketing research [1].

When conducting a sample survey, the observance of certain requirements is required: the reliability of the results of the observations, the costs of the research should be in an acceptable ratio with the value of the information received, the human and material resources involved must be consistent the complexity of the study.

A mandatory feature of the sampling method is a random equiprobable order of selection of sampling units (each element of the general population has a certain fixed probability of being selected). If the result of a sample study turned out to be far from real, then, most likely, the selection of units was limited to the most accessible part of the general population for observation, it was possible that the selection was made at random, according to an expert or performer, as the most “typical” for this population with some predetermined properties

Features of the problem being solved, originality of the general population, cause diversity types of sample observation, depending on the unit of selection, repetition of selection, method of selection of units, the number of stages of observation, sampling phases interpenetration of observations.

These features also determine the relevance of a more accurate understanding and use of the techniques, methods and principles of organizing a sample survey, mastering the technique of forming a sample population and calculating sample indicators, which we will consider using a number of examples.

1. Selective control method, for example, in customs statistics. The purpose of control is to assess the extent of shadow imports and the implementation, on the basis of such estimates, of “adding up” data to the actual level of indicators. For a sample survey (control) is recommended to select several dozens of specific import flows. A sample survey reveals trends in understating the customs value by manipulating product codes and importing consignments of goods without customs clearance, overstatement of the price when the volume is underestimated, which is advantageous in relation to such goods for which customs duties depend not on the cost, but on the weight of the cargo. Illegal deliveries according to “black” (illegal, contraband) schemes are typical for such goods as clothes, shoes, knitwear.

In commodity flows with a predominance of “gray” (semi-legal) import schemes, work in the legal field, i.e. “white imports”, makes up some (possibly smaller) part (determined by analyzing sample data on imports) in comparison with the value of officially registered imports (for goods such as pharmaceuticals and preparations, haberdashery leather goods).

2. On the organization of sample research. At the stage of designing a sample observation, one of the issues to be solved is the determination of the requirements for the degree of reliability of the survey materials (acceptable margins of error of representativeness, the probability with which this result is guaranteed). The decision on these issues is made taking into account the permissible risks of error in the final indicators. The closer the connection of the investigated problem with vital processes, with phenomena that can lead to negative (catastrophic) consequences, the more cruel should be the requirements for the accuracy of the information obtained during the examination.

This applies, in particular, to the accuracy of economic indicators in statistics of public opinion. The difficulty in deciding on the permissible limits of sampling errors lies in the fact that several users of the resulting information may have different requirements for the degree of accuracy and reliability of the materials. Depending on the permissible limits of the selection errors, the size of the sample population is determined. When estimating the sample size n for the

marginal error Δ and confidence coefficient t for sociological research using random sampling, the $n = 0,25t^2 / \Lambda^2$ formula is recommended, those. The estimate is not lower than the maximum possible in order to avoid the risk of underestimating the required sample size.

The greatest difficulties arise when calculating the required sample size due to the lack of reliable data on the variance of the studied characteristics. In this case, they resort to determining the magnitude of the variance by conducting a pilot survey to obtain its approximate value. You can also apply expert estimates of variance based on data from previous periods for indicators that change little in time or by analogy with the known indicators of other, unrelated features.

3. Models for the formation of a sample for a survey of small enterprises. Small business as an element of a market economy, including a set of small enterprises, peasant (farmer) enterprises and individual entrepreneurs, is a dynamic sector of the economy and therefore, its careful monitoring is necessary.

Features of the model for the formation of a sample of small enterprises are as follows: a) the purpose of conducting statistical monitoring of the activities of small enterprises is the formation of indicators of their activity in the context of areas, sectors of the economy, patterns of ownership; b) criteria for the identification of small enterprises are developed taking into account the legal status, number of employees and the structure of the authorized capital; c) for stratified random sampling with a total sample size of n units, the sample size for individual layers is calculated as follows:

$$n_h = nm_h\sigma_h / \sum_{h=1}^H m_h\sigma_h ,$$

where n_h – the number of units selected from the hth layer; m_h – volume of the hth layer; σ_h^2 – variance of the hth layer,

$$\sigma_h^2 = \sum_i^{m_h} (y_{hi} - y_h)^2 / (m_h - 1),$$

H – the number of layers, y_{hi} – the base value of the attribute of the i th unit of observation of the h th layer; y_h – the average value of the characteristic in the h th layer.

One of the problems in the survey of small enterprises is the insufficient level of reliability of sample data due to the failure to provide statistical reporting forms and distortion of the value of reporting indicators in order to reduce the tax burden. In the statistical processing of non-responses, the weighting method is used or a method for replacing missing parameters. The method of weighting is that the development of the results is carried out according to the submitted reports, and when the results of the sample survey are distributed to the general population, the share of each layer is determined by the sum of the submitted reports and non-answers. Another way to fill in non-responses is to replace the missing parameters from the aggregate of reporting enterprises. Missing information is “modeled” and assigned to an unanswered company (individual indicator) using one of the following methods: a) random selection of a donor enterprise from the same layer; b) logical selection of data in case of non-answers to individual questions in the submitted reports; c) replacing an unanswered enterprise with a nearest neighbor from the same layer; d) modeling the value of missing indicators in reporting enterprises using regression or other mathematical models.

Selected subjects small businesses are then grouped by industry for a more complete study. Thus, a list (catalog) of objects of statistical observation is formed.

A sample study of individual entrepreneurship.

The peculiarity of individual entrepreneurship is as follows. The main source of information on the activities of individual entrepreneurs is the data of state registration, declarations available to the tax authorities. A significant part of individual enterprises is engaged in retail and wholesale, repair work of various profiles, and the provision of transportation services.

Statistical monitoring of the activities of individual entrepreneurs – car owners. When organizing this survey, statistical authorities

interact with the transport inspection of their area. On the basis of the registers of entrepreneurs (owners of cars entitled to engage in activities for commercial purposes), a sample is formed in the form of a surname list of entrepreneurs in alphabetical order. Each entrepreneur is assigned a number from 1 to N (N is the total number of entrepreneurs of this profile in the region). To form a sample population from this list, random selection is made with the return of the selected units in the general population. The sample size in each time period is defined as the product of the total population and the proportion of selection.

Sample surveys of individual entrepreneurs operating in retail. Not all enterprises fall into the field of view of statistics, but only with predetermined annual turnover, book value of fixed assets and the number of employees hired.

Hence, new tasks arise for organizing the accounting of this sector of small business, adapted to the specifics of individual entrepreneurship, for the solution of which sample surveys are acceptable.

Of particular interest are models of the formation of a sample of individual entrepreneurs. The calculation of the sample size for each district can be made according to the formula

$$n_h = n(x_h^a \cdot v(y_h)) / \sum_h (x_h^a \cdot v(y_h)),$$

where n is the given total sample size as a whole for the study area; n_h – is the sample size in the h th region; x_h is the indicator of the turnover of individual trade entrepreneurs in the h th region according to the results of the previous survey cycle; assessment of the coefficient of variation of the turnover indicator of individual trade entrepreneurs in the h th subject, based on the results of the previous observation cycle; a – parameter of the degree of influence of the regional total indicator on the placement of the sample, $0 < a < 1$.

Based on information about the required sample size in the districts, a combined sampling frame is compiled, consisting of a territorial and payroll base. Lists of observation units are formed:

individual entrepreneurs in the sample, (on a list basis) and trading places belonging to individual entrepreneurs in the sample (on a territorial basis). The proportion between the two types of observation units is taken to be 30% and 70% (respectively for the payroll base and the territorial base) of the total sample size in the region.

After the survey is completed, an array of information is adjusted taking into account non-answers and identified atypical units of the population. The procedure for disseminating data from a combined sample involves determining the total value of the indicator separately for each sample. The two estimates obtained (on a territorial and on a payroll basis) indicators are used to calculate a single weighted average score:

$$Y = p \cdot Y_T + (1 - p) \cdot Y_C$$

$$Y_R = \sum_{k=1}^{n_R} w_k \cdot y_k,$$

where Y – is the total indicator of the survey program for the sample as a whole; Y_R – the total indicator of the survey program for the sample on a territorial (at $R = T$) or list (at $R = C$) basis; k – serial number of the unit of observation in the sample; w_k – sample weight (distribution factor) calculated at the sampling planning stage for the k th unit obtained during the observation; y_k – is the value of the characteristic of the k th unit obtained during the observation; n_R – is the number of observation units in the sample population on a territorial ($n_R = n_T$) or list ($n_R = n_C$) basis.

The combined sample formation model is being tested by a pilot survey in selected territories. With such an examination, the procedures and algorithms for generating a combined sample are tested, the reliability of the methodology for disseminating sample data to the general population.

References:

1. NEWBOLD, P., CARLSON, W.L., THORNE, B. *Statistics for Business and Economics*. Published by Pearson Education, 2013.