Error Control for Rural Sampling Survey: An Application of the Agricultural Census Information

Shan Wenhao
Shanghai Statistics Bureau, China
shanwh@email.stats-sh.gov.cn

Abstract: Sampling survey tends to be the main technique on the collection of the social and economic information in rural area, along with the reforming of the Chinese rural statistical systems. The spot error occurring in the survey is becoming the important issue influencing the accuracy of rural statistics so as to decision-making. Therefore, improving survey technique timely and controlling the error of ordinary rural sampling survey effectively are of most importance for resulting reliable survey results. This paper is a study on this issue to evaluate and improve the error of the existing rural sampling survey, by utilizing the results of the First Agricultural Census.

1. The problem

As we know, spot error refers to the difference of specific indicators between the random samples and the population. The expression is,

$$Q = \sum \bar{x}_n f_n - \sum \bar{X}_N F_N$$

Here, Q is the spot error, $\bar{x}_n$ and f_n are separately the mean value of variables and weights of each sample group, $\bar{X}_N$ and F_N are separately the mean value of variables and weights of each population group. Subscripts n and N are group number of the samples and the population. $\sum$ is the sum.

For convenience, suppose that the sample variable, $\bar{x}_n$, is approximate to population variable, $\bar{X}_N$. Therefore, the difference between the above two terms would be determined mainly by the cumulating value of the weight difference between the samples and the population, f_n, and F_N. The less this value is, the more representative the samples are.

The accuracy in ordinary rural sampling survey may be improved greatly if the inference is made on the revised survey weight distribution table, F_N, which is constructed on the base of the Agriculture Census information, i.e., using $\sum \bar{x}_n F_N$, instead of $\sum \bar{x}_n f_n$. In the meantime, it makes it possible to evaluate and revise technically the results from ordinary rural sampling survey. As a conclusion, this technique may, (1) reduce the system error caused by the imperfect samples, resulting in a better inference; (2) present basic information for re-adjusting and optimization of sample structure.

The principle consideration of the above mentioned technique is to improve the sample selection, error control and revision of survey results. The compiling procedure for the weight distribution table of the rural survey, based on the Agriculture Census results, is, (1) determining the reference indicator for the survey; (2) setting proper isometric group number; (3) compiling the weight distribution table on the base of the Agriculture Census results; (4) comparing the difference between the weights for the isometric sampling and that from the Census; (5) evaluating and revising the results from the sampling survey.

2. Example

For convenience, the following table is the presentation for analyzing and evaluating the spot error in rural household survey based on the Agriculture Census information.

Based on the data above, the spot error and its composition can be calculated and evaluated. One

may find the difference between the grouped weights of the survey and the Census by comparing $\overline{F}_N$ with f_n in the above table, so as to improving the coming sample selection. One may find the amount and its causation of the survey distribution deviation by calculating the deviation, ($\overline{x}_n f_n - \overline{x}_n F_N$). One may also evaluate the spot error of various existing sampling surveys effectively and present reliable results by calculating the spot error, $(\sum \overline{x}_n f_n - \sum \overline{x}_n F_N) / \sum \overline{x}_n F_N \times 100\%$

Net income per capita yuan	Survey sample average $\overline{x}_n$	Census sample weight F_N (%)	Survey sample weight f_n (%)	Survey sample average Census weight $\overline{x}_n F_N$	Survey weighted average value $\overline{x}_n f_n$
Below 2000	1491.17	12.9	6.3	192.36	93.94
2000-3000	2548.21	34.8	13.7	886.76	349.10
3000-4000	3524.96	20.3	20.8	715.57	733.19
4000-5000	4444.81	9.2	20.0	408.92	888.96
5000-6000	5489.94	4.9	16.8	269.01	922.31
6000-7000	6522.34	2.7	6.7	176.10	437.00
Above 7000	9383.78	15.2	15.7	1426.33	1473.25
Sum	33405.21	100.0	100.0	4075.05	4897.75

Note, f_n is the survey isometric group weight calculated from average net income per capita of 600 sampled rural households. $\overline{F}_N$ is the Census isometric group weight calculated from average net income per capita of 2768 rural households. Both survey and Census are graded from low income to high income. So the data of survey and Census are consistent.

3. Suggestion

The conclusion is that the main key of using the survey weight distribution table is well connecting the sampling designs of survey and agriculture census. Should take survey and agriculture census into account together. The link up of correlated index, sampling, survey method, data processing and designing of result exporting is very important. All of the above is high concentrated for elevating quality and efficiency of statistics in Chinese rural area. Along with the economic development in Chinese rural area, deducing the collectivity weight distribution table from the agriculture census stably is not adoptable. Because the agriculture census is taken once per ten years. So, renovating and maintaining the data of agriculture census are needed. Practice proved that using survey method to renovate and maintain the important data of agriculture census timely is very efficient.