A Method of "Multi to Single" in Multipurpose Sampling

Wang Zuocheng
Statistical bureau of Henan province
No. 79 Hongzhuan Road,
Zhengzhou city, P. R. China
wangzuocheng@371.net

1. Introduction

In this paper, we give a method of transforming "multipurpose" to single purpose by using multivariate statistical analysis, This single purpose is a new one, and not anyone of origin items. We can use this new item to set up a new sampling frame.

2. Main idea of "Multi to single " Method

According to survey need, we can ascertain the multi items that we should survey, for example there are k items, we' ll take these k items as multipurpose. To transform those multi purposes to one, First step, we' ll partition the set of k items into m disjoint classes of items, by R cluster analysis, so as to ensure that items in the same class are"similar" to one another. Second, we' ll take a representative of each class by multivariate correlate analysis, so we have m items, each item is a representative of other items in the same class. At last, we' ll colligate these m items into one integrative object by the method of Principle component analysis.

3. Example and Process

We select the rural of Xinmi in Henan province of China as our investigate object, all the 307 administrative villages with next ten items composed of multipurpose sampling population. These items include: Annual per capita net income of rural residents, Yield of grain per hectare, Yield of cotton per hectare, Yield o f oil bearing crops per hectare, Yield of vegetables per hectare, Output of eggs per capita, Output of pork per capita, Output of beef per capita, Output of mutton per capita and output of meat per capita. We surveyed data numbers about above ten items from 1997 to 1999, and averaged these 3 years data number as analysis data of each item.

Using SAS software, We partitioned the 10 items into 5 disjoint classes of items by R cluster analysis as next:

1st class: Output of pork per capita, Output of beef per capita, Output of mutton per capita and output of meat per capita;

2nd class: Annual per capita net income of rural residents, Yield of grain per hectare and Yield of vegetables per hectare;

3rd class: Yield of cotton per hectare;

4^{th} class: Yield of oil bearing crops per hectare;

5^{th} class: Output of eggs per capita.

Using multivariate correlate analysis, we can get a representative in every class, each representative have the maximum correlate coefficient to other items in same class. So we got 5 representatives of 5 classes: Output of meat per capita (signed as Meat_PC), Yield of grain per hectare (signed as Grain_PH), Yield of cotton per hectare (signed as Cotton_PH), Yield of oil bearing crops per hectare (signed as Oil_PH) and Output of eggs per capita (signed as Eggs_PC).

We take principle component analysis to these 5 items by using these 307 administrative villages' data. From the result of analysis, we weighted these 5 items into a sum, named p, as follows:

$$p = b_1 * Egg\_PC + b_2 * Meat\_PC + b_3 * Grain\_PH + b_4 * Cotton\_PH + b_5 * oil\_PH$$

where $b_i = \sum_{j=1}^{m} d_j |a_{ij}|$ and $d_j = l_j \Big/ \sum_{j=1}^{m} l_j$, l_j is cumulative of jth principle component.

So we can arrange these 307 administrative villages by p, and carry out systematic sampling.

4. Conclusion

Using this "multi to single" method, we transform multipurpose sampling to single purpose sampling, and in the control of sampling error, we can make sure which item cannot survey in multipurpose sampling.