Research on Individual Aesthetic Taste Generation of Hezhe Nationality

Yan Sun[1], Fuyun Geng[2] & Linman Li[3]

[1] Faculty of Arts, Southwest University, Chongqing, China

[2] Faculty of Education, Southwest University, Chongqing, China

[3] Chongqing Second Normal University, Chongqing, China

Correspondence: Yan Sun, Faculty of Arts, Southwest University, Beibei Direct, Chongqing, China. Tel: 86-186-1699-8911. E-mail: cxxs0828@126.com

Abstract

Proportional stratified sampling method is a frequently-used method in sample analysis. During the process of stratified sampling, the proportional stratified sampling and disproportionate stratified sampling methods adopted by us can improve representativeness of the sample and strengthen certainty of the estimated value of the population quantity index. Besides, it is also able to effectively avoid the phenomenon of centering on some characteristics or omitting some characteristics during simple random sampling. However, proportional stratified sampling method is not suitable for all samples, and different samples should be treated separately. By taking the proportional stratified sample of test scores gained by students of Hezhe nationality according to mother's education degree as an example, this paper provides beneficial references for research on individual aesthetic taste generation of Hezhe nationality.

Keywords: Hezhe nationality, individual aesthetic taste, generation, proportional stratified sampling, instance analysis

1. Introduction

Stratified sampling is a frequently-used sampling method in probability sampling. In such method, the population is classified or stratified according to major factors closely related to the research contents, and then the samples will be extracted according to the random principle in each stratum. The simplest feasible allocation mode in stratified sampling is to determine the unit quantity structure of the sample by scaling-down method. Many people hold that except following the consistence between the sample and population in unit quantity structure, the variance difference in different strata should also be considered. The minimum requirement of sampling estimator variance should be met. In short, the sample size of each stratum must be allocated under limited fund and time or other restrictions related to allocation volume of samples in each stratum, so as to minimize variance of estimators. (Henry, 2008, p. 98).

The main problem of stratified sampling in operation is how to allocate the whole sample to various stratification groups. Common allocation method includes proportional allocation and disproportionate allocation, known as proportional stratified sampling and disproportionate stratified sampling respectively. Proportional stratified sampling means to allocate samples to the stratification group according to the proportion of the individual size in the population size within the stratification group. Such method is suitable for stratified random sampling with significantly small difference in the stratification group after stratification. Disproportionate stratified sampling means to allocate samples to the stratification group according to other weights rather than the proportion of the individual size in the population size within the stratification group. Other common weight is the proportion of standard deviation of a certain stratification group in the total standard deviation of various stratification groups (Henry, 2008, p. 98).

This paper will make instance analysis by setting mother's education degree of Hezhe students as the sample through proportional stratified sampling and disproportionate stratified sampling methods. Such method will improve representativeness of the sample and strengthen certainty of the estimated value of the population quantity index. Besides, it is also able to effectively avoid the phenomenon of centering on some characteristics

or omitting some characteristics during simple random sampling. However, in the practical process of sample collection, the samples are not standard and not all samples can use proportional stratified sampling method, so different samples should be treated separately. This paper will explain planning and implementation of statistical sampling inspection by starting from the basic concept of stratified sampling and basic knowledge about mathematical statistics; besides, the application program is elaborated via examples. It has provided suggestions for selecting suitable stratified sampling methods for sample analysis in the future. In this way, the influence of incorrect sampling method on research results is avoided, and the difficulty of sample analysis is reduced. Meanwhile, beneficial empirical materials are provided for research on aesthetic taste generation of Hezhe nationality.

2. Literature Review

Statistical sampling inspection theory was proposed by Doctor Romig and Doctor Dodge in 1924; after continuous improvement and development in practice, it is still in development now. Few Chinese scholars have studied stratified sampling method.

X. Wang and J. Liu (2004) tried to solve practical problems by referring to optimization methods like nonlinear programming and nonlinear goal programming, so as to complement the defects of simple stratified sampling. Z. Yan (1999) systematically discussed multivariate simple random sampling and regression estimation method of stratified sampling; he proved that compared with non-auxiliary information sampling, regression estimation method was more effective; besides, he also gained some statistical inferences of population average. W. Fan (2013) compared the option price using variance reduction techniques with the option price not using variance reduction techniques by selecting proper options; she discovered that the option using variance reduction techniques had higher precision and smaller standard deviation, so the simulated option price possessed higher reliability.

J. Liu and G. Chen (2008) mainly studied how to properly stratify in stratified sampling, so as to increase the estimation precision of stratified sampling. Three stratification marks were studied: qualitative auxiliary information, research variable, and auxiliary variable; besides, the proper stratification limits were determined under various stratification marks, to minimize the variance of estimators. Q. Zhou (2014) revealed through sampling survey theory: efficiency of PPS sampling was better than the result of simple random sampling. He verified the sampling efficiency under small sample conditions by a case study of average and total amount estimation for fixed investments of major industries in various regions of China. According to the result, in efficiency of sampling estimation results, sampling error and interval estimation at the confidence coefficient of 95%, the efficiency of PPS estimation was higher than that of simple random sampling. H. Wang (1995) analyzed the reason for extensive application of stratified sampling and he considered that by stratifying according to the administrative subordination relations under the current system of China and carrying out sampling for different strata separately, it would provide convenience for the whole sampling work. Secondly, sampling and inference could be conducted independently for each stratum, which would meet the demands of different regions and departments. Thirdly, by proper stratification, representativeness of the sample for the population could be enhanced, the sampling error and sample size would be reduced, and the survey cost could be decreased. In stratified sampling, the study population should form various strata, and then a random sample must be selected in each stratum. The purpose of adopting stratified sampling is: to guarantee representativeness of proportion in each stratum; to reduce sampling variability; to minimize the subpopulation and thus to make the analysis more reliable.

According to the above literatures, researches on stratified sampling mainly concentrate upon method of sample selection and confidence interval; proportional stratified sampling and disproportionate stratified sampling are only used by them as a tool. No scholars have made specific analysis and comparison for proportional stratified sampling and disproportionate stratified sampling.

3. Examples of Stratified Sampling

Stratified sampling can be either proportional or disproportionate. The purpose of proportional stratified sampling is to guarantee representativeness of stratum with the same sampling fraction in each stratum.

This paper investigated 18 students of Hezhe nationality in a certain college of Heilongjiang Province, so as to study the influence of mother's education degree on student achievement. The mother's education degree was divided into "below high school", "high school and college dropout" and "undergraduate"; statistics was made for student achievement after sampling, as shown in Table 1. This variable is considered as an important predictive variable for knowledge and skill, so it is very important to make the sample represent the three groups according to a certain proportion in this variable.

3.1 Examples of Proportional Stratified Sampling

Table 1. Proportional stratified sample of test scores gained by students of Hezhe nationality according to mother's education degree

Student achievement					
A1		A2		A3	
Below high school		High school and college dropout		Undergraduate	
1	95	1	99	1	123
2	113	2	107	2	154
3	103	3	120	3	147
4	93	4	127	4	167
5	123	5	139	5	143
6	109	6	114		
7	103				
N_k	7		6		5
$\overline{X}_k$	105.0		117.7		146.8
S_k	10.34		14.31		16.13
N_k/N	0.438		0.375		0.188

N_k is the sample size of stratum;

$\overline{X}_k$ is the average value of stratum;

S_k is the standard deviation;

N_k/N is the weight based on the proportion of stratum.

Table 2. Calculation method for standard deviation of stratified sampling: by taking proportional stratified sampling as an example

Stratum for mother's education degree of students of Hezhe nationality		
15.27	34.13	52.04

$S^2_{\overline{X}} = (0.438)^2(15.27) + (0.375)^2(34.13) + (0.188)^2(52.04)$

$S^2_{\overline{X}} = 2.93 + 4.80 + 1.84$

$S^2_{\overline{X}} = 9.57$

$S_{\overline{X}} = (9.57)^{1/2} = 3.09$

In the formula:

$W_k = {N_k}/{N}$

$S^2_{\overline{X}}$ is the variance of stratum;

N_k is the sample size of stratum;

$S_{\overline{X}}$ is the overall standard error of average value.

Design effect (deff)

Calculation method for standard error: suppose that it is simple random sample.

$\overline{x} = 123.2$

$s = 13.59$

$S^2_{\overline{X}} = {13.59^2}/{18} = {184.69}/{18} = 10.26$

$S_{\overline{x}} = (10.26)^{1/2} = 3.20$

In the formula, $\overline{X}$ is the overall average value;

S is the standard deviation;

$S_{\overline{x}}$ is the standard error of simple random sample of average value.

Design effect $= \dfrac{9.57}{10.26} = 0.93$

$(deff)^{1/2} = \dfrac{3.09}{3.20} = 0.97$

When proportional stratified sampling is adopted, each stratum contains different numbers of members, so the sample sizes in various strata are also different, as shown in Table 1. For proportional stratified sampling, sampling fraction in all strata is the same. The calculation formulas for its average value, proportion and other statistics are the same with those of simple random sampling. For instance,

$$\overline{X} = \sum X_i \big/ n \tag{1}$$

However, the formula used to calculate its standard error is different from that of simple random sampling.

$$S_{\overline{X}} = \left(\sum \frac{w_k^2 s_k^2}{n} \right)^{1/2} \tag{2}$$

The weight of stratum is added into the formula of calculating standard error, $W_2 = N_k \big/ N$. In stratified sampling, N can replace n in the formula. This formula weights the square of each standard deviation ($S_k^2 \big/ n$) with the relative size of stratum in the population. Table 2 is an example of conducting calculation with this formula.

Table 3. Calculation method for standard deviation of stratified sampling: By taking alternative proportional stratification as an example

Class stratum	B1	B2	B3
	96	103	113
	122	91	147
	139	121	132
	127	103	154
	119	121	99
	107	107	120
n_k	6	6	6
x_k	118.3	107.7	127.5
S_k	15.12	11.64	20.89
N_k/N	0.333	0.333	0.333
S_k^2 / N_k	38.10	22.58	72.73

$S_x^2 = (0.333)^2 (38.10) + (0.333)^2 (22.58) + (0.333)^2 (72.73)$

$S_x^2 = 4.22 + 2.50 + 8.06$

$S_x^2 = 14.78$

$S_{\overline{X}} = (14.78)^{1/2} = 3.84$

In the formula:

$\overline{X}_k$ is the average value of stratum;

S_k is the standard deviation of stratum;

$S^{\frac{2}{x}}$ is the overall standard error of sample.

Design effect (deff)

$\text{Design effect} = \frac{14.78}{14.01} = 1.05$

$(deff)^{1/2} = \frac{3.84}{3.74} = 1.23$

Stratification can reduce standard error, which can be proved by data not stratified in this example and the standard error calculated by the formula used for simple random sample. The ratio between standard error of stratified sample and standard error of simple random sample is the square root of the design effect (deff); the

value of design effect is listed at the bottom of the table. For this example, relative gain of precision is 21% (100%-79%). The value of relative gain of precision depends on two factors:

Variability between strata: The greater the difference between the average value in various strata and the overall average value is, the higher the relative gain of precision will be;

Homogeneity within stratum: The higher the similarity within the stratum is, the higher the relative gain of precision will be.

Kish (1965) explained this problem with the concept of variance (R^2) of the research variable used to explain stratification variable. The greater the explanatory power of stratification variable is, the higher the relative gain of precision of stratification conducted according to this variable will be. The gain is in proportion to the value of the explained variance.

Table 3 uses the alternative stratified variable class in the population which is the same with that in Table 2. Gain of this stratification is 2%, and it is equal to the square root of design effect and smaller than the gain of mother's education degree stratification of Hezhe students. The square root of design effect of class stratification is 98%, while it is 79% during mother's education degree stratification of Hezhe students. By comparing mother's education degree of Hezhe students with their classes, the variance of students' test scores is explained. By comparing these two examples, we see that the interval among average values of various strata during stratification via mother's education degree of Hezhe students (30.7) is greater than the interval among average values of class stratification (16.0). Besides, we can also find that variation or standard deviation (s) within the stratum during stratification via mother's education degree of Hezhe students are both small.

Kish (1965) further pointed out that the size of stratum proportion would affect relative gain. By increasing the sample size of a small subpopulation, the estimation cannot be substantially improved no matter how much difference it has with the rest part. Finally, he also mentioned that when the percentage or proportion was the major target of analysis, gain of stratification would not be too large. When variation within the stratum (homogeneity) is small and variability between strata is high, the stratification will have a relatively great gain. Generally speaking, when the difference between strata in size proportion is huge, it will be hard to improve the precision of estimated value by stratification.

3.2 Examples of Disproportionate Stratified Sampling

If precision of the whole sample or precision of a certain subpopulation is not high enough, disproportionate stratification can be adopted. Disproportionate stratification originates from the way of using different sampling fractions for different strata. Adopting different sampling fractions will result in unequal probability of selection and disproportionate representativeness in the sample.

The advantage of disproportionate stratification is that it can increase the sampling size of a certain stratum with relatively high standard deviation, and reduce sampling variation of this stratum at the same time. By understanding why it can reduce variation, it will help us grasp the calculation formula of standard deviation.

$$S_{\overline{X}} = \left(\sum w_k^2 s_{\overline{x}}^2\right)^{1/2} = \left(w_1^2 s_{\overline{x1}}^2 + w_2^2 s_{\overline{x2}}^2 + w_3^2 s_{\overline{x3}}^2 + \ldots + w_n^2 s_{\overline{xn}}^2\right)^{1/2} \tag{3}$$

In the formula, $s_{\overline{X}}^k$ is the standard deviation of Number K stratum, and W_k is equal to N_k / N.

Except that the standard deviation of each stratum is worked out, this formula is the same with the formula of proportional stratification. Any of the formula can be used for either proportional stratification or disproportionate stratification.

We have to calculate the standard deviation of each stratum at first and then combine them into a weighted average, so the stratum with the greatest standard deviation and weight will have a huge influence on standard deviation. By increasing the sample size within the stratum with the greatest sampling error via disproportionate stratification, sampling error of this stratum can be reduced, thus sampling error of the whole sample will also be decreased. Table 4 illustrates this nature with disproportionate stratification. Stratification in this table is the same with examples in Table 2 and Table 3. Via the index of square root of design effect, we have measured that the sampling error decreases by almost 2%, and sampling error of disproportionate stratification is smaller than the corresponding sampling error of proportional stratification. For example, standard error of the efficient allocation of units in stratification according to mother's education degree is 3.48, while it is 3.56 during proportional stratification. The former one is slightly lower than the latter one, but the difference is not too significant.

Table 4. Efficient allocation of stratified sample size: variation result of sample size in stratum

Stratum of mother's education degree of Hezhe students			
S_k	10.34	15.87	16.80
n_k	6	7	5
$S_{\overline{X}} = 3.21$ $(deff)^{1/2} = 0.85$			
Class stratum			
S_k	16.90	12.76	20.77
n_k	6	5	7
$S_{\overline{X}} = 4.14$ $(deff)^{1/2} = 1.05$			

In the formula,

S_k is the standard deviation of stratum;

n_k is the size of stratum;

$S_{\overline{X}}$ is the overall standard error of average value.

In order to enhance precision of the estimated value to the largest extent, the sample size must be in proportion to standard deviation and the size of stratum:

$$n^k = \frac{n(N_k S_k)}{\sum N_k S_k} \tag{4}$$

In the formula:

n_k is the sample size of stratum;

n is the total sample size;

N_k is the sample size of stratum of the population;

S_k is the standard deviation of stratum.

This formula seems to be logical, but in fact, standard deviation of the population and stratum is unknown. Therefore, the sampling unit allocated to each stratum is not that accurate. However, when standard deviations, or their relative sizes in most cases, can be estimated, it will increase the precision of estimated value to allocate the sampling units of stratum. In the two examples listed in Table 4, we reallocated 2 sampling units among the 18 units in the first example and reallocated 1 sampling unit in the second example. In the first example, we reduced the sample units in A1 by 2, and added 1 to A2 and A3 respectively. In the second example, we took away 1 sampling unit from B2 and added it to B3. Variation of each stratum can be estimated with the same method of estimating the standard deviation of efficient sample size. Such methods include beforehand research, exploratory research, and using the range.

4. Conclusion

As stated in the above, the advantage of proportional stratified sampling is that it can enhance precision of the estimated value and guarantee proportion representativeness of stratification groups. Stratification does not require any extra cost. However, every member of the research population must be listed and classified according to variables used for stratification. It often needs a huge cost to gain such information associated with the population. Sometimes, it might be not too expensive to obtain information about the variable of stratification expected by us. For instance, from the angle of cost-effectiveness, the loss may outweigh the gain by collecting information about mother's education degree of all students. However, we have information about the region where Hezhe students live, and this can be utilized as information index of socioeconomic status. This index might be related to mother's education degree of Hezhe students.

When we need to analyze the subgroup but the standard deviation of subsample selected according to the proportion is too large, disproportionate stratified sampling should be adopted. At this time, disproportionate stratified sampling will help us increase sample size of the subpopulation without having to increase the overall

sample size proportionally. In this way, we must define the stratum in a way that can associate members of the subpopulation with this stratum. Ideal stratification of realizing this purpose happens when the stratum is composed of mutually exclusive members of subpopulation. When members of subpopulation are highly concentrated, disproportionate stratified sampling can also be used.

The main disadvantage of disproportionate stratification is that weighting is required when calculating the standard deviation. In this way, calculation of standard deviation will be inevitably more complicated. In addition, in the saved data set, a code specially used to identify stratum is required; moreover, it should also include weight on which stratification depends. Many kinds of statistical software contain programs or commands that are utilized to set estimated value and calculate standard deviation. As for the common mistake in sampling, disproportionate sampling is used during sample selection, but weighting is not conducted in the estimation process; thus such bias in the estimated value cannot be modified.

References

Fan, W. (2013). *Empirical Analysis on Simulation for Option Price via Monte Carlo Method-Stratified Sampling Method Based on Variance Reduction Technique*. Times Finance.

Henry, G. T. (2008). *Practical Sampling* (pp. 98-99). Chongqing City: Chongqing University Press.

Liu, J., & Chen, G. (2008). *Optimization Design for Stratification Bound in Stratified Sampling*. Statistics and Decision.

Wang, H. (1995). *Optimum Allocation for Sample Size of Stratified Sampling*. China Statistics.

Wang, X., & Liu, J. (2004). *New Exploration on Optimum Allocation for Sample Size of Stratified Sampling*. Statistics and Decision.

Yan, Z. (1999). *The Regression Estimation for Multivariate Simple Random Sampling and Stratified Sampling*. Journal of Shanxi University (Natural Science Edition).

Zhou, Q. (2004). *Empirical Test for Estimation Efficiency of PPS and Simple Random Sampling*. Statistics and Decision.

Copyrights