Comparing the Cut-off and Modified Cut-off sample design on Factory Wastes Statistics

Jin, Young
National Statistical Office, Statistical Research & Development Div.
920, Dunsan-dong, Seo-gu
Taejeon, , Republic of Korea
Jiny@nso.go.kr

Lee, Sang Eun
Kyonggi University, Republic of Korea
sanglee@stat.kyonggi.ac.kr

Nam, Jae Woo
Ministry of Environment, Republic of Korea

1. Introduction

The purpose of this work is to study the cut-off and modified cut-off sampling strategies with Waste Statistics data. When you use the cut-off sampling which only counts for the "take-all" part, it may reduced the budget but the accuracy from the sample will be doubted. On the other hand, increasing the accuracy, the budget will be raised for the survey. This study should outline the best or better sampling strategies for the most important study variable.

2. Cut-off and Modified Cut-off sampling design

The cut-off sampling has been used for specially Establishment Survey in most of country. Some countries rather use the modified cut-off sampling design which takes account for " take all " and " take some "parts. Most of case, it definitely gives a better estimates than an estimates by using the cut-off sampling.

2.1 Cut-off sampling

Selection : Elements are selected by using the cut-off point which gives the minimizing the total sample size referring to the permitted error. Once we decide the cut-off point, take the all samples above the point and ignore the samples below the point.

$$\text{Estimator}: \quad \hat{X} = \sum_{i=1}^{n_c} X_i \quad \text{or} \quad \sum_{i=1}^{n_c} X_i \Big/ n_c$$

where X_c is the cut-off point

n_c is the total size of above the point

2.2 Modified Cut-off sampling

Selection : Elements are selected by using the cut-off point which gives the minimizing the total sample size referring to the permitted error. Once we decide the cut-off point, take the all samples above the point and add the some samples from below the point using systematic sampling technique.

$$\text{Estimator}: \quad \hat{X} = \sum_{i=1}^{n_c} X_i + \left(\frac{N_s}{n_s} \right) \sum_{i=1}^{n_s} X_i \quad \text{or} \quad \frac{\sum_{i=1}^{n_c} X_i}{n_c} + \frac{\sum_{i=1}^{n_s} X_i}{n_s}$$

where $n_s = \dfrac{[\alpha^2 (q \bullet cv)^2]/e^2}{1 + [\alpha^2 (q \bullet cv)^2]/e^2 \bullet N_s}$

$$q = 1 - \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{N} X_i}$$

α is the confidence level

e is the permitted error

N_s is the total size of below the point

n_s is the sample size of below the point

3. Numerical Example and conclusion

As a numerical example we use the Factory Waste Statistics which has a large skewness value so that we can see the difference between the cut-off and modified cut-off sampling effects. With two sampling methods, Table 1 gives the results of the MSE on each estimates. As we can see from the table, even we selected the over 90% of the portions from census, estimating the total using cut-off will give you the big difference from the real value. However estimating the mean won't give you the big difference compare with total estimates.

Table 1 : MSE for Cut-off and Modified Cut-off

($\alpha = 1.96,\ e = 0.05$)

	Mean Square Error (MSE)	
	Cut-off Sampling	Modified Cut-off Sampling
Total	92582884.0	488778.2
Mean	1155931.6	1167470.1

Specially in large scale such as national-wide survey, even over 90% of the dominated portions of sample were selected form the population, the cut-off sampling may has chanced to miss lead a result.

Therefore, we recommend the modified cut-off sampling design for large scale sampling even it has a large skewness value.

REFERENCE

Hidiroglou, M.A. (1986), "The Construction of a Self-Representing Stratum of Large Units in Survey," *The American Statistician.*