JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3764-3768
https://publishoa.com
ISSN: 1309-3452

A Study of Efficiency Calculation in Multistage Sampling and Reliability Calculations

Md Abdul Qudus Sheikh[1]

Research Scholar Department Of Mathematics, Faculty of School of Teaching department, Dr. A. P. J. Abdul Kalam University Indore, MP, India

Dr. Anjna Rajoria [2]

Research Scholar Department Of Mathematics, Faculty of School of Teaching department, Dr. A. P. J. Abdul Kalam University Indore, MP, India
[1] maqudussheikh@gmail.com,[2] anjna_rajoria@rediffmail.com

Received 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

ABSTRACT

In all the sample selection procedures discussed so far in this block, the entire investigation was based on the assumption that a usable list of units (i.e., a frame) is available from which one selects a sample. Unfortunately, not always such a list of units is available especially when we are concerned with countrywide investigations. Even existence of such a list of units under investigation (in some cases) would not give us enough scope to base our enquiry on a simple random sample because of high budgeting costs. The clean development mechanism (CDM) Executive Board (hereinafter referred to as the Board) at its fiftieth meeting approved the "General Guidelines for Sampling and Surveys for Small-Scale CDM Project Activities (sampling guideline)". Further, the Board at its sixtieth meeting agreed to set up a joint task force comprising members of the Methodologies Panel and the Small-Scale Working Group to further work on the issue to develop one set of common sampling guidelines and best practices examples covering large and small-scale projects and Programme of Activities (PoAs). It further agreed that the scope of the guidelines shall include guidance to designated operational entities (DOEs) on how to review sampling and survey designs in project design documents (PDDs) as well as how to apply sampling to validation/verification work.

KEYWORDS: Efficiency Calculation, Multistage Sampling, Reliability Calculations, small-scale projects, Programme of Activities, project design documents

INTRODUCTION

As said in the introduction, when the sampling unit is a cluster, the procedure of sampling is called cluster sampling. So, cluster sampling consists of forming suitable clusters of contiguous population units and surveying all the units in a sample of clusters selected according to some appropriate sample selection method. We consider a multi-stage cluster sampling, which can be used in large-scale sample survey in practice. At each stage, we take a fixed size sampling design and denote a sample of cluster by at the first stage and sample of element in the ith selected cluster by and so on. To estimate the population mean, we introduce homogeneous linear estimator as

$$\bar{y}_s = \sum_{i \in s^*} b_{s^*i} \bar{y}_i$$

where depends on both i and s^* and $\bar{y}_i$ is an unbiased estimator for the ith cluster mean $\bar{Y}_{i\cdot}$. The mean square error (MSE) can be calculated from the formula

$$MSE(\bar{y}_s) = MSE_1[E_2(\bar{y}_s)] + E_1[Var_2(\bar{y}_s)]$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3764-3768
https://publishoa.com
ISSN: 1309-3452

where indices denote the stages. For the unbiased case, Cochran(1977) suggested an unbiased variance estimator of the population total. However, it has a limitation because his result can not be used for the bias case. In a single stage cluster sampling, that is, the first part of the right hand side of MSE, a general formula of MSE of the homogeneous linear estimator was suggested by Rao and Vijayan (1977) and more generally by Vijayan et al. (1995). They showed that if there is a point at which MSE is zero, then the MSE is represented as

$$MSE(\bar{y}_s) = -\sum_{i=1}^{N}\sum_{j>i}^{N} d_{ij} w_i w_j \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2$$

where N is total number of clusters, $D=(d_{ij})$ is $N\times N$ symmetric constant matrix not depending on y's, and $w'=(w_1,\dots\dots,w_N)$ is such that $MSE(\bar{y}_s\,|\,w)=0$.

In the article, we will extend their approach to multi-stage sampling and formulate a nonnegative unbiased MSE estimator of the population mean.

Non-negative unbiased MSE estimator

For the ith cluster, let M_i be the size, and $\hat{\sigma}_i^2$ be unbiased variance estimator of $\bar{y}_i$. Also we define $b_i = E_1(w_{s^*i})$ and $b_{ij} = E_1(w_{s^*i} w_{s^*j})$. Then we have the following results.

Result 1: Under a fixed size design, if $MSE(\bar{y}_s)=0$ when $y=w$, then we obtain

$$MSE(\bar{y}_s) = -\sum_{i=1}^{N}\sum_{j>i}^{N} d_{ij} w_i w_j \left(\frac{\bar{Y}_i}{w_i} - \frac{\bar{Y}_j}{w_j}\right)^2 + \sum_{i=1}^{N} b_{ii}\sigma_i^2$$

where $d_{ij} = b_{ij} - b_i M_i/M - b_j M_j/M + M_i M_j/M^2$.

Result 2. Under a fixed size design we assume $E(e_{ij}(s^*)) = d_{ij}$ for all ij. Then a non-negative unbiased estimator of $MSE(\bar{y}_s)$ is given by

$$m(\bar{y}_s) = -\sum_{i\in s^*}\sum_{j>i\in s^*} e_{ij}(s^*) w_i w_j \left(\frac{\bar{y}_i}{w_i} - \frac{\bar{y}_j}{w_j}\right)^2 + \sum_{i\in s^*}\left(2w_{s^*i}\frac{M_i}{M} - \frac{1}{\pi_i}\left(\frac{M_i}{M}\right)^2\right)\hat{\sigma}_i^2$$

where π_i is the first order inclusion probability of the ith cluster.

The unbiased condition of $\bar{y}_s$ is that $b_i = M_i/M$. Therefore Variance formula can is obtained from MSE by replacing $d_{ij} = b_{ij} - M_i M_j/M^2$. In addition, an unbiased variance estimator of $\bar{y}_s$ is

$$v(\bar{y}_s) = -\sum_{i\in s^*}\sum_{j>i\in s^*} e_{ij}(s^*) w_i w_j \left(\frac{\bar{y}_i}{w_i} - \frac{\bar{y}_j}{w_j}\right)^2 + \sum_{i\in s^*}\left(w_{s^*i}\frac{M_i}{M}\right)\hat{\sigma}_i^2$$

Estimation of Population Mean

Simple random sampling, systematic sampling, and stratified sampling are various types of sampling procedures that can be applied in the cluster sampling by treating the clusters as sampling units. In this unit, however, we shall restrict to situations where clusters are selected using without replacement simple random sampling procedure. The theory of cluster sampling in its own right is rather complex, where the complexity depends on whether one takes equal or unequal-sized clusters. In general, a formulae for calculating the standard error of cluster estimates has two terns, where the first relates to the variability between cluster means (or proportions) and the second to the variability within cluster. In this unit, we start with the case of unequal clusters and then deduce from this the results about clusters of equal sizes as a special case.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3764-3768
https://publishoa.com
ISSN: 1309-3452

Case-I: Unequal clusters. Usually, in practice, clusters are of unequal sizes. For instance, households as a group of persons and villages as a group of households can be taken as clusters for the purpose of sampling. We assume (i) the population consists of N clusters, where the ith cluster has Mi elements, i = 1,2,. . . , N, and (ii) n clusters are selected from N clusters by without replacement simple random sampling procedure.

Then, an unbiased estimator of population mean $\overline{Y}$, with M_0 known, is given by the relations

$$\hat{\overline{Y}}_c = \frac{N}{nM_0}\sum_{k=1}^{n} M_k\,\overline{Y}_k = \frac{1}{\overline{M}n}\sum_{k=1}^{n} Y_k, \text{ where } \overline{M} = \frac{M_0}{N}.$$

And, the variance of the estimator $\hat{\overline{Y}}_c$ is given by the relation

$$V(\hat{\overline{Y}}_c) = \left(\frac{N-n}{N\,n\,\overline{M}^2}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \overline{Y}_c\right)^2.$$

Also, an unbaised estimator of variance is given by relations

$$\hat{V}(\hat{\overline{Y}}_c) = \left(\frac{N-n}{N\,n\,\overline{M}^2}\right)\frac{1}{n-1}\sum_{k=1}^{n}\left(Y_k - \overline{M}\,\hat{\overline{Y}}_c\right)^2 = \left(\frac{N-n}{N\,n\,\overline{M}^2}\right)\frac{1}{n-1}\left[\left(\sum_{k=1}^{n} Y_k^2\right) - n(\overline{M}\,\hat{\overline{Y}}_c)^2\right]$$

We try to understand a use of these relations with the help of the following problem.

Problem 2. For studying the cultivation practices and yield of apple, a pilot sample survey is conducted in a district of Kashmir. The yields (in kgs) of 3 clusters of trees, selected by without replacement simple random sampling, from 15 are as given in the following table.

cluster	size	yield
1	12	5.53,26.11,11.08,12.66,0.87,6.40,54.31,37.94, 7.13,3.53,14.23,1.24
2	10	4.84,10.93,0.65,32.52,3.56,11.68 35.97,47.07,17.69,40.7
3	6	15.79,11.18,27.54,28.11,21.70,1.25

With $\overline{M} = 10$, estimate the average yield per tree as well as the production of apple in the village and their standard errors.

Solution. Here, $N = 15, n = 3$ and $\overline{M} = 10$. Then, $M_0 = N\overline{M} = 150$. So, $M_1 = 12, M_2 = 10$ and $M_3 = 6$. Then, using values from table, we get

$$\hat{\overline{Y}}_c = \frac{1}{30}[Y_1 + Y_2 + Y_3] = \frac{1}{30}\left[\sum_{j=1}^{M_1} Y_{1j} + \sum_{j=1}^{M_2} Y_{2j} + \sum_{j=1}^{M_3} Y_{3j}\right] = \frac{1}{30}[181.03 + 205.61 + 105.57] = \frac{492.21}{30} = 16.407.$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3764-3768
https://publishoa.com
ISSN: 1309-3452

Thus, the average yield of apple per tree is 16.407 kgs. Also, we have

$$Y_1 = \sum_{j=1}^{M_1} Y_{1j} = 181.03,\ Y_2 = \sum_{j=1}^{M_2} Y_{2j} = 205.61,\ \text{and } Y_3 = \sum_{j=1}^{M_3} Y_{3j} = 105.57.$$

Using these values of $Y_i(1 \leq i \leq 3)$, the estimated variance $\hat{V}(\hat{\bar{Y}}_c)$ is given by

$$\begin{aligned}\hat{V}(\hat{\bar{Y}}_c) &= \left(\frac{15-3}{15 \times 3 \times 100}\right) \frac{1}{3-1} \left[\left(\sum_{k=1}^{n} Y_k^2\right) - 3(10 \times 19.709)\right] \\ &= \frac{1}{750}[(32771.86 + 42275.47 + 11145.03) - 26918.97] = 79.03\end{aligned}$$

Thus the standard error of $\hat{\bar{Y}}_c$ is $\sqrt{79.03} = 8.89$.

Efficiency of Cluster Sampling

Here, right in the beginning, we want to remark that the estimator $\hat{\bar{Y}}_c$ for equal sized clusters is based on a sample of nM units in the form of n clusters each consisting of M units. Thus, if the same number of units are selected from a population of NM units by without replacement simple random sampling procedure, then the sample mean estimator $\hat{\bar{Y}}_c$ and its variance $V(\hat{\bar{Y}})$ are given by the relations

$$\hat{\bar{Y}} = \frac{1}{nM} \sum_{k=1}^{nM} Y_k, \text{ and}$$

$$\begin{aligned}V(\hat{\bar{Y}}) &= \left(\frac{1}{n\,M} - \frac{1}{N\,M}\right) S^2 \\ &= \left(\frac{N-n}{N\,n\,M}\right) \frac{1}{N\,M-1} \left(\left(\sum_{i=1}^{NM} Y_i^2\right) - NM\,\bar{Y}^2\right), \text{ respectively.}\end{aligned}$$

Observe that the relative efficiency defined above involves value of study variable for all population units. However, in practice, the investigator has only the sample observations of n clusters of M units each. For this, he needs the estimates of two variances involved in the formulae of relative efficiency (RE).

CONCLUSION

Many estimators in survey sampling are not unbiased, for instance ratio estimator, regression estimator and Murthy's estimator. One method to handle biasd estimator is approximation to unbiased estimator, and another is to make inference about MSE of biased estimator. Our results in the article may be useful in the case of inference about biased estimators.

REFERENCES

[1]. Arnab, R (1988). Variance estimation in multi-stage sampling. The Australian Journal of Statistics, 30, 107-110.
[2]. Cochran, W. (1977). Sampling techniques. John Wiley and Sons. New Yorks.
[3]. Rao, J.N.K. and Vijayan, K. (1977). On estimating the variance in sampling with probability proportional to aggregate size. Journal of the American Statistical Association, 72, 579-584.
[4]. Vijayan, K., Mukhopadhyay, P. and Bhattachayya, S. (1995). On non-negative unbiased estimation of quadratic forms in finite population sampling. The Australian Journal of Statistics, 37, 169-178.
[5]. Carnette, I. (2010): Analysis of Complex Survey Data in Stata. Paper Present at the 2010 Mexican Stata Users Group Meeting, April 29, 2010

[6]. Carle, A. C. (2009): Fitting Multilevel Models in Complex Survey data with Design Weights: Recommendations. BMC Medical Research Methodology, 1471-2288-9-49
[7]. Cochran, W. G. (1977): Sampling Techniques. Third Edition. John Wiley and Sons New York.
[8]. Kalton, G. (1983): Introduction to Survey Sampling. Sape Publications: Los Angeles, CA.
[9]. Kerry, I. and Bland J. (2006): The Intra-cluster Correlation Coefficient in Cluster Recommendation. British Medical Journal, 316, 1455-1460
[10]. Lavallee, P. (2010): Cross Section Weighting of Longitudinal Surveys of Individuals and Households Using the Weight Share Method. Survey Methodology, 32, 165-176
[11]. Kumar, R., Singh, J.P., Srivastava, G. (2014). Altered Fingerprint Identification and Classification Using SP Detection and Fuzzy Classification. In: , et al. Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1602-5_139
[12]. Gite S.N, Dharmadhikari D.D, Ram Kumar," Educational Decision Making Based On GIS" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-1, April 2012.
[13]. Ram Kumar, Sarvesh Kumar, Kolte V. S.," A Model for Intrusion Detection Based on Undefined Distance", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1 Issue-5, November 2011
[14]. Vibhor Mahajan, Ashutosh Dwivedi, Sairaj Kulkarni,Md Abdullah Ali, Ram Kumar Solanki," Face Mask Detection Using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022
[15]. Rajawat, Anand Singh and Chauhan, Chetan and Goyal, S B and Bhaladhare, Pawan R and Rout, Dillip and Gaidhani, Abhay R, Utilization Of Renewable Energy For Industrial Applications Using Quantum Computing (August 11, 2022). Available at SSRN: https://ssrn.com/abstract=4187814 or http://dx.doi.org/10.2139/ssrn.4187814
[16]. Bedi, P., Goyal, S.B., Rajawat, A.S., Shaw, R.N., Ghosh, A. (2022). A Framework for Personalizing Atypical Web Search Sessions with Concept-Based User Profiles Using Selective Machine Learning Techniques. In: Bianchini, M., Piuri, V., Das, S., Shaw, R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_23