International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Multi-Label Systematic Sampling Design: A Generalized Unbiased Estimator in the Presence of Linear Trend

N. Uthayakumaran

NIE, ICMR, Chennai, Tamilnadu, India

Abstract: ***For practical utility, multi-stage sampling design is used by researchers when the population is large. This conventional sampling design omits portions of the populations in stages, as a result production of unbiased estimates to population parameter become uncertain. To overcome this limitation, a design called multi-label systematic sampling is introduced in this paper. This sampling design deals with both equal and unequal number of arrays of population with the associate factors which exist in multiple dimensions. All the array labels of sampling units are identified in a single attempt in a linear fashion by this sampling design. For population arranged as (N x N x ……h times) or (N_1 x N_2 x N_3 x ….. x N_h) cells, this sampling design is useful in selecting the samples. To estimate the population parameter of study variable in the population, if one identifies multiple numbers of associate factors of the study variable; it is possible to arrive at generalized unbiased estimate of population parameter by this sampling design for the population exhibiting linear trend.***

Keywords: Multi-label: Marker of array population in multiple dimensions, **Linear trend:** Uniformly increasing trend, **Multi-stage:** Two or more hierarchical levels, **Generalized unbiased estimator:** Unbiased in all level, **Cuboidal:** Three unequal dimensions

1. Introduction

A multi-stage sample is one in which sampling is done for larger population. It is sequentially across two or more hierarchical levels. Cochran (1939), Hansen and Hurvitz (1943) and in India Mahalanobis (1940), Sukhatme (1953), Lahiri (1954)- crop survey have found multi-stage sampling to be very useful for estimation. This conventional sampling design provides many directions. For example, there is no theoretical limit to the number of stages. In fact, too many stages will very seriously limit the ability to measure variability and also increase uncertainty. There has to be a cut-off which the researchers can decide with proper judgment, available experience and evidence. There is a possibility of bias in multi stage sampling. The multi stage sampling design is not truly random as the sample is identified in several stages omitting parts of the population in each stage. It may be noted that a technique of cuboidal systematic sampling method by Uthayakumaran (2015) has greater flexibility. This method, while removing the disadvantage of multi stage sampling i.e., omitting parts of the population in each stage of sample selection, by selecting the sample in a single attempt from the whole population arranged in three dimensions also provide unbiased estimate of the study variable. Generalizing this technique for multiple dimensions, the design called Multi-label systematic sampling, which gives the generalized unbiased estimator to the population mean for the population exhibiting linear trend has been described.

1.1 Multi-label systematic sampling design

A multiple dimensional population element may be represented by study variable Y_{i_g}, $i_g = 1,2,\ldots,N_g$; g=1,2,….h. Here, Y_{i_g} is the value of the h dimensional elements. The population contains ($N_1xN_2xN_3x\ldots xN_h$) units. The sample contains ($n_1xn_2xn_3\ldots xn_h$) units. The sampling intervals $k_1, k_2,\ldots\ldots,k_h$ are N_1/n_1, N_2/n_2, ….. N_h/n_h respectively.

A Multi-label systematic sample is selected by drawing multiple independent starting coordinates r_g at random, each between 1 to k_g respectively. A sample of size ($n_1xn_2xn_3x\ldots xn_h$) contains all units whose coordinates are of the form

$$\{r_g + \gamma k_g\}, \qquad \gamma = 0,1,\ldots,(n_g-1) \quad g = 1,2,\ldots\ldots,h \qquad (1.1)$$

Estimation of population mean of study variable

For the sampling design described above, one can estimate population mean of study variable using sample mean of study variable:

$$\left(\bar{y}_{Multi}\right)_{r_g} = \frac{1}{\prod_{g=1}^{h} n_g}\left\{\prod_{g=1}^{h}\sum_{i_g=1}^{n_g}\left(Y_{i_g}\right)_{r_g}\right\} \qquad (1.2)$$

The variance of the above estimator can be established by considering

$$V\left(\bar{y}_{Multi}\right)_{r_g} = \frac{1}{\prod_{g=1}^{h} k_g}\prod_{g=1}^{h}\sum_{r_g=1}^{kg}\left\{\left(\bar{y}_{Multi}\right)_{r_g} - \bar{Y}\right\}^2 \qquad (1.3)$$

where population mean of study variable

$$\bar{Y} = \frac{1}{\prod_{g=1}^{h} N_g}\prod_{g=1}^{h}\sum_{i_g=1}^{Ng} Y_{i_g} \qquad (1.4)$$

Theorem-1: The generalized estimator (1.2) under multi-label systematic sampling design is the generalized unbiased

estimator of the population mean for the population arranged in multiple dimensions which is exhibiting linear trend

$$Y_{i_g} = \sum_{g=1}^{h} i_g \qquad (1.5)$$

Proof: Initially, results of theoretical derivations for unbiasedness of the generalized estimator under multi-label systematic sampling design have been given for multi-dimensional equal number of array populations

$$(N \times N \times N \times \ldots..h \text{ times}) \qquad (1.6)$$

It may be noted that population size ($N_1 \times N_2 \times N_3 x \ldots\ldots x N_h$) = ($N \times N \times N \times \ldots..h$ times) = N^h; and sample size ($n_1 \times n_2 \times n_3 \ldots\ldots \times n_h$) = ($n \times n \times n \times \ldots\ldots.h$ times) = n^h .

Table 1: The generalized unbiased estimator under multi dimensional equal number of array population

h	$N_1=N_2\ldots..=N_h=N$	$E(\bar{y}_{Multi})_{r_g}$	$\bar{Y}$
2	$NxN=N^2$	2(N+1)/2=(N+1)	(N+1)
3	$NxNxN=N^3$	3(N+1)/2	3(N+1)/2
.	.	.	.
.	.	.	.
h	(NxNx..h times)=N^h	h(N+1)/2	h(N+1)/2

It is pertinent to note that for $N_1=N_2=N_3\ldots..=N_h=N$, $E(\bar{y}_{Multi})_{r_g} = h(N+1)/2 = \bar{Y}$, which is nothing but population mean under the linear trend (1.5). It has been proved that the generalized estimator under multi-label systematic sampling design is the generalized unbiased estimator of the population mean for the population arranged equal number of arrays in multiple dimensions (1.6), which is exhibiting linear trend (1.5).

Finally, results of theoretical derivations for unbiasedness of the generalized estimator under multi-label systematic sampling design have been given for multi-dimensional unequal number of array population

$$(N_1 \times N_2 \times N_3 x \ldots\ldots x N_h) \qquad (1.7)$$

for the population exhibiting linear trend (1.5).

Table 2: The generalized unbiased estimator under multi dimensional unequal number of array population

h	$N_1 \neq N_2\ldots..\neq N_h$	$E(\bar{y}_{Multi})_{r_g}$	$\bar{Y}$
2	N_1xN_2	$(N_1+N_2+2)/2$	$(N_1+N_2+2)/2$
3	$N_1xN_2xN_3$	$(N_1+N_2+N_3+3)/2$	$(N_1+N_2+N_3+3)/2$
.	.	.	.
.	.	.	.
h	$N_1xN_2x\ldots xN_h$	$(N_1+N_2+\ldots N_h+h)/2$	$(N_1+N_2+\ldots N_h+h)/2$

It is interesting to note that for $N_1\neq N_2\ldots..\neq N_h$, $E(\bar{y}_{Multi})_{r_g} = (N_1+N_2+\ldots N_h+h)/2 = \bar{Y}$, which is nothing but population mean under the linear trend (1.5). The generalized estimator under multi-label systematic sampling design is the generalized unbiased estimator of the population mean for the population arranged unequal number of arrays in multiple dimensions (1.7), which is exhibiting linear trend (1.5).

It is well known from Table-1 and Table-2, generalized estimator under multi-label systematic sampling design is generalized unbiased estimator for the population arranged in multiple dimensions which is exhibiting linear trend (1.5).

Theorem-2: Variance of the generalized estimator (1.3) under multi-label systematic sampling design is increased with an additional number of dimensions for the population arranged in multiple dimensions which is exhibiting linear trend (1.5).

Proof: Results of theoretical derivation for variance of the generalized estimator have been given for multi-dimensional equal number of array population (1.6) for the population exhibiting linear trend (1.5).

Table 3: Variance of the generalized estimator under multi dimensional equal number of array population

h	$N_1=N_2\ldots..=N_h=N$	$V(\bar{y}_{Multi})_{r_g}$
2	$NxN=N^2$	$2(K^2-1)/12 = (K^2-1)/6$
3	$NxNxN=N^3$	$3(K^2-1)/12 = (K^2-1)/4$
.	.	.
.	.	.
h	(NxNx..h times)=N^h	$h(K^2-1)/12$

Due to their complex nature, results of theoretical derivation for variance of the generalized estimator have not been given for unequal number of array population (1.7) for the population exhibiting linear trend (1.5).

It may be noted from Table-3, variance of the generalized estimator (1.3) under multi-label systematic sampling design is increased with an additional number of dimensions for the population arranged in multiple dimensions which is exhibiting linear trend (1.5).

2. Discussion

It is pertinent to note that multi-label systematic sampling design is considering the whole population, which is arranged in multiple dimensions according to its associate factors. This technique produce unbiased estimate when the population arranged in increasing order in order to exhibit linear trend for the study variable by considering the whole population arranged in multiple dimensions. The requirement and specific arrangement of the multi dimensional population in this sampling design, in effect ensure - the observance of the study variable on a much enhanced setup. The indirect use of population /geographical size variable in arranging the total population for the selection of the sample will satisfy the linear trend assumption for the study variable. Fixing the cells with the indirect use of population /geographical size information will uniquely determine the sample.

3. Conclusion

Attaining generalized unbiased estimator under multi-label systematic sampling design is absolutely assured for the population exhibiting linear trend. This suggested sampling design covers both equal and unequal number of arrays of population with the associate factors which subsist in multiple dimensions. All the array labels of sampling units are known in a single attempt in a linear fashion by this sampling design. For population arranged as (N x N x……h times) or (N_1 x N_2 x N_3 x…..x N_h) cells, the suggested sampling design is effective in selecting the samples. To estimate the population parameter of study variable in the population, if one be aware of multiple number of associate factors of the study variable; it is likely to turn up at unbiased estimate of population parameter by this sampling design for the population exhibiting linear trend.

This sampling design reveals many directions. There is no theoretical limit to the number of dimensions. However, variance of generalized unbiased estimator of this sampling design is increased with an additional number of dimensions for the population arranged in multiple dimensions. In reality, too many dimensions will give more variability and uncertainty. There has to be an optimum number of dimensions, which the researchers can choose with an appropriate decision, on hand experience and identification of accurate number of associate factors to arrive at unbiased estimator with reduced variance.

References

[1] Bellhouse, D.R. and Rao, J.N.K., (1975): *Systematic sampling in the presence of trend*, Biometrika, 62, 694-697.
[2] Cochran W.G, (1939): *Use of analysis of variance in enumeration by sampling, JASA, Vol. 34, P492-510.*
[3] W.G Cochran, (1977): *Sampling Techniques*, Third Edition, Wiley Eastern, P227-
[4] Hanif, M. and Brewer, K.R.W. (1980): Sampling with unequal probabilities without replacement: A review, In. Statist. Rev., 48, 317-335.
[5] Hansen, M.H and W.N. Hurwitz (1943): On the theory of sampling from finite populations, Ann. Math. Statist., Vol 14, P3333-362.
[6] Lahiri D.B., (1951): *A method for selection providing unbiased estimates*, Int. Stat. Ass. Bull, 33, 133-140.
[7] Lahiri D.N., (1954): *Technical paper on some aspects of the development of the sample design, Sankhya, Vol. 14, P332-362.*
[8] Leslie Kish, (1987): Statistical Design For Research, John Wiley & Sons., P33-
[9] Madow, W.G. and L.H. Madow, (1944): *On the theory of systematic sampling*, Ann. Math. Stat., 15, 1-24.
[10] Mahalanobis, P.C, (1940): Report on the sample census of jute in Bengal, Ind. Central Jute Committee.
[11] Sukhatme, P.V, (1950): Efficiency of sub sampling designs in yield surveys, J. Ind. Soc. Agr. Statist., Vol. 2, P212-228.
[12] N. Uthayakumaran, (1998): *Additional circular systematic sampling Methods*, Biometrical Journal, 40, 4, 467-474.
[13] N. Uthayakumaran, and S. Venkatasubramanian, (2013): *Dual circular systematic sampling methods for disease burden estimation*, International journal of statistics and analysis, VOL 3, NO. 3, 307-322.
[14] N. Uthayakumaran, and S. Venkatasubramanian, (2015): An alternate approach to multistage sampling: UV Cubical *circular systematic sampling method*, International journal of statistics and applications, 5(5), 169-180.
[15] N. Uthayakumaran, (2015): *Cuboidal systematic sampling method: An unbiased estimator in the presence of linear trend*, International journal of statistics, Accepted for publication.
[16] Yates, F., (1948): *Systematic sampling*, Transactions Royal Society, London, A 241, 345-377.

Author Profile

Dr. N. Uthayakumaran received his Ph.d. (Statistics) from Madras University, Chennai, Tamilnadu, India. Presently he is a Technical officer at NIE, ICMR, Chennai and has published many research articles in Internationally Reputed Journals.