Multi-circular Systematic Sampling Design

N. Uthayakumaran

NIE, ICMR, Chennai, Tamilnadu, India

Abstract: ***For population units arranged in a single or multiple dimensions, several estimators have been developed to estimate the population parameter when the population size is a multiple of sample size. Only limited numbers of estimators have been introduced to estimate population parameter when the population size is not a multiple of sample size. Also, there are few estimators that have been introduced for unequal multi-dimensional populations. To resolve these matters, a design called multi-circular systematic sampling for an unequal multi-dimensional population when $N_1 \neq n_1k_1$, $N_2 \neq n_2k_2$,…, $N_h \neq n_hk_h$ in turn $(N_1 \times N_2 \times N_3 \ldots\ldots \times N_h) \neq (n_1 \times n_2 \times n_3 \ldots\ldots \times n_h) \times (k_1 \times k_2 \times k_3 \ldots\ldots \times k_h)$ is described for the population exhibiting linear trend.***

Keywords: Multi-circular: Identifying labels in a cyclic fashion in multiple dimensions, **Multi-dimension:** Multiple dimensions, **Linear trend:** Uniformly increasing trend

1. Introduction

Most of the systematic sampling methods such as linear systematic sampling (Madow and Madow, 1944), balanced systematic sampling (Sethi, 1965), centered systematic sampling (Singh, Jindal and Garg, 1968), square grid systematic sampling (W.G Cochran, 1977), cuboidal systematic sampling (Uthayakumaran, 2015), multi-label systematic sampling (Uthayakumaran, 2015) etc., have been introduced to develop estimators of the population parameters when the population size is a multiple of sample size. The major limitation of these systematic sampling methods is that the estimators are developed when the population size is a multiple of sample size.

In real life situations, sample size is determined by some fundamental assumptions pertaining to contact variables, study result, type-I error and power. Other important considerations such as man power, time frame, feasibility, problem of non-response and cost also play a vital role in deciding a sample size. Hence, population size may not always be a multiple of sample size. To overcome this limitation, circular systematic sampling method (Lahiri, 1951), balanced and centered circular systematic sampling methods (Uthayakumaran, 1998) for single dimensional population, dual circular systematic sampling method (Uthayakumaran and Venkatasubramanian, 2013) for two dimensional population, UV cubical circular systematic sampling method (Uthayakumaran and Venkatasubramanian, 2015) for three dimensional population have been introduced to generate estimators when the population size is not a multiple of sample size.

In these dual and cubical circular systematic sampling methods, only equal dimensional populations are dealt for estimation. To resolve this matter, a design called multi-circular systematic sampling, which gives a reliable estimator to the population mean for the population arranged in multiple unequal dimensions and its corrected estimator, which may coincide with the population mean in the presence of a linear trend has been described.

2. Multi-Circular Systematic Sampling Design

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

A multi-circular systematic sample is selected by drawing multiple independent starting coordinates r_g at random, each between 1 to N_g respectively. A sample of size $(n_1 \times n_2 \times\ldots\times n_h)$ contains all units whose coordinates are of the form

$$\begin{aligned} &\{r_g + \gamma k_g\} && \text{if } 1 \leq r_g + \gamma k_g \leq N_g \\ &\{r_g + \gamma k_g - N_g\} && \text{if } r_g + \gamma k_g > N_g, \\ &\gamma = 0, 1,\ldots, (n_g-1);\ g = 1, 2,\ldots\ldots, h && \end{aligned} \tag{2.1}$$

Estimation of population mean of study variable

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

$$\left(\bar{y}_{Multi-circular}\right)_{r_g} = \frac{1}{\prod_{g=1}^{h} n_g} \left\{ \prod_{g=1}^{h} \sum_{i_g=1}^{n_g} \left(Y_{i_g} \right)_{r_g} \right\} \tag{2.2}$$

The variance of the above estimator can be established by considering

$$V(\bar{y}_{Multi-circular})_{r_g} = \frac{1}{\prod_{g=1}^{h} N_g} \prod_{g=1}^{h} \sum_{r_g=1}^{Ng} \left\{ (\bar{y}_{Multi-circular})_{r_g} - \bar{Y} \right\}^2 \quad (2.3)$$

where population mean of study variable

$$\bar{Y} = \frac{1}{\prod_{g=1}^{h} N_g} \prod_{g=1}^{h} \sum_{i_g=1}^{Ng} Y_{i_g} \quad (2.4)$$

Sampling of six units (h=2):

To start with, study variable of multi-dimensional population exhibiting linear trend

$$Y_{i_g} = \sum_{g=1}^{h} i_g \quad (2.5)$$

is considered for two unequal dimensional population in the following illustration.

To estimate population mean of study variable, if areas like group of taluks (I_1) and group of villages (I_2) are identified according to its increasing order of geographical size, it is possible to arrive at reliable estimate under the assumption that study variable exhibits linear trend. In this illustration, row dimension is 5 groups of taluks and column dimension is 8 groups of villages. The technique of multi-circular systematic sampling can be illustrated by applying it to the case of sampling n_1 x n_2 = 2 x 3 = 6 units from the population arranged in two- unequal dimensions of N_1 x N_2 = 5 x 8 = 40 units.

Step 1:
A random number is drawn from 1 to 5 of $\mathbf{I_1}$ (say 5). Sampling interval $k_1 = N_1/n_1 = 5/2 = 2$, using (2.1), the following i_1^{th} label of sampling units are selected.

{5, 2}

Step 2:
A random number is drawn from 1 to 8 of $\mathbf{I_2}$ (say 8). Sampling interval $k_2 = N_2/n_2 = 8/3 = 2$, using (2.1), the i_2^{th} label of sampling units are selected.

{8, 2, 4}

Step 3:
The following sample units of two-unequal dimensional population are selected using the labels selected from the step1and step 2.

(5, 8), (5, 2), (5, 4), (2, 8), (2, 2), (2, 4)

Illustration:
This sampling design is discussed theoretically with study variable exhibiting linear trend through the model (2.5)

$Y_{i_g} = i_1 + i_2$, $i_1 = 1,.., 5$; $i_2 = 1,.., 8$.

I_1 I_2	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	2	3	4	5	6	7	8	9
(2)	3	4	5	6	7	8	9	10
(3)	4	5	6	7	8	9	10	11
(4)	5	6	7	8	9	10	11	12
(5)	6	7	8	9	10	11	12	13

In the above illustration, if it is assumed that sample survey is carried out to find out the estimate sample mean in the randomly selected 2 x 3 sampling units, it can be noted that Y_{i_g} deducted from the selected cells are 4, 6, 10, 7, 9 and13.

Using (2.2), the estimated sample mean is 49/6 = 8.17, which is close to the population mean of (N_1 + N_2 + 2)/2 = 7.5.

The possible 40 samples together with the sample means of multi- circular systematic sampling design is shown in Table-1.

Table-1: Sample mean under multi- circular systematic sampling design (h=2)

Random starts	Sample composition	Probability	Sample mean
1,1	(1,1),(1,3),(1,5) (4,1),(4,3),(4,5)	1/40	$(Y_{11}+\ldots\ldots+Y_{45})/6$
1,2	(1,2),(1,4),(1,6) (4,2),(4,4),(4,6)	1/40	$(Y_{12}+\ldots\ldots+Y_{46})/6$
…	…	…	…
…	…	…	…
5,8	(5,8),(5,2),(5,4) (2,8),(2,2),(2,4)	1/40	$(Y_{58}+\ldots\ldots+Y_{24})/6$

The expected value of sample mean of the multi-circular systematic sampling design is just the simple average of column (4) in Table-1, which turns out to be population mean of study variable.

Similarly, this exercise may be repeated for more than two (h > 2) dimensional populations.

3. Corrected Estimator

It is well known that linear systematic sampling scheme performs better than simple random sampling in the presence of linear trend. Yates (1948) suggested a modification over the usual expansion estimator in order to estimate the population mean of study variable in the presence of linear trend without any error. Proceeding along these lines, Bellhouse and Rao (1975) suggested a corrected estimator meant for circular systematic sampling to estimate the population mean of study variable in the presence of linear trend without any error. Motivated by this, Uthayakumaran (1998) suggested a new estimating procedure that is quite general in nature and applicable under various systematic sampling methods for population viewed in single dimension. Uthayakumaran and Venkatasubramanian (2013, 2015) extended this procedure for population arranged in two and three dimensions according to the associated factors.

In the light of the above development, the following procedure is introduced for the population arranged in multiple dimensions.

Let the sampled units be arranged in an original serial order of the population. A multiple dimensional sample units may be represented Y_{j_g}, $j_g = 1, 2,\ldots, n_g$; $g = 1, 2,\ldots, h$. Here, Y_{j_g} is the value of the h dimensional elements. The values of the elements which have equal array labels are given a unique weight R. The weight R is selected in such a way that the corrected sample mean of study variable coincides with the population mean of study variable under the linear trend (2.5).

This can be achieved by equating the corrected sample mean of study variable to the population mean of study variable.

$$\bar{y}_c = \frac{1}{\prod_{g=1}^{h} n_g}\left\{\prod_{g=1}^{h} \sum_{\substack{j_g=1 \\ \neq(j_1=j_2=\ldots=j_h)}}^{n_g} (Y_{j_g})_{r_g} + R\prod_{g=1}^{h} \sum_{\substack{j_g=1 \\ =(j_1=j_2=\ldots=j_h)}}^{n_g} (Y_{j_g})_{r_g}\right\} \tag{3.1}$$

coincides with the population mean of study variable. That is, unique weight R can be chosen in such a way that $\bar{y}_c = \bar{Y}$,

where $\bar{Y} = (N_1 + N_2 + \ldots + N_h + h)/2$ (3.2)

By solving equations (3.1) and (3.2), we get

$$R = \frac{\prod_{g=1}^{h} n_g \,(N_1 + N_2 + \ldots + N_h + h)/2 - \prod_{g=1}^{h} \sum_{\substack{j_g=1 \\ \neq(j_1=j_2=\ldots=j_h)}}^{n_g} (Y_{j_g})_{r_g}}{\prod_{g=1}^{h} \sum_{\substack{j_g=1 \\ =(j_1=j_2=\ldots=j_h)}}^{n_g} (Y_{j_g})_{r_g}} \tag{3.3}$$

It may be noted that the unique weight R = 9/13 in the above mentioned illustration when the corrected estimator $\bar{y}_c = \bar{Y} = (N_1 + N_2 + 2)/2 = 7.5$ coincides with the population mean of the study variable.

4. Discussion

Estimate will be reliable by the adoption of the above methodology as care is being taken to spread the sampling units into the population as much as possible. In routine, it is not often to get the population in equal number of cells with the associate factors and also it is recognized that population size is not always multiple of sample size. To overcome these difficulties, multi-circular systematic sampling design is explained through the linear trend (2.5) for the population arranged as (N_1 x N_2 x ….x N_h) cells, where (N_1 x N_2 x ….N_h) ≠ (n_1 x n_2 x ….x n_h) x (k_1 x k_2 x ….x k_h).

In the case of corrected estimator in multi-circular systematic sampling design, the values of the equal labeled elements are given a unique weight R to estimate the population mean of study variable in the presence of linear trend without any error.

5. Conclusion

For the population size not a multiple of sample size, the suggested design is useful in selecting the appropriate samples from an unequal number of multi-dimensional cells. The estimation of parameter is likely to be more reliable if the sample is identified in accordance with the associated factors. This multi-circular systematic sampling approach has successfully incorporated this flexibility by ensuring an equal probability sample where it is not necessary to have a complete sampling frame.

In real life, too many levels of dimension will give inconsistency and uncertainty. Generally, three dimensional sampling methods are adequate for large scale surveys. The researchers can decide the levels of dimension to attain a reliable estimator using proper decision, existing knowledge and classification of associate factors.

The suggested corrected sample mean (3.1) is equal to the population mean under multi-dimensional population exhibiting linear trend.

If ($N_1 = N_2 = \ldots = N_h = N$) and ($N^h \neq n^h k^h$) then this sampling design yields an equal probability sample with cyclic fashion for equal multi-dimensional population. If population is a multiple of sample size in some dimensions and not in other dimensions and (N_1 x N_2 x N_3 …….x N_h) ≠ (n_1 x n_2 x n_3…..x n_h) x (k_1 x k_2 x k_3…..x k_h) then also the sampling design yields an equal probability sample with linear and cyclic composition of selection.

This sampling design may be applied by adopting different sampling procedures instead of systematic sampling procedures for the selection of array labels of sampling units in different levels. However, accuracy and reliability of the estimator by this attempt may be examined by further research.

References

[1] Bellhouse, D.R. and Rao, J.N.K., (1975): *Systematic sampling in the presence of trend*, Biometrika, 62, 694-697.
[2] W.G Cochran, (1977): *Sampling Techniques*, Third Edition, Wiley Eastern, P227-
[3] Lahiri D.B., (1951): *A method for selection providing unbiased estimates*, Int. Stat. Ass. Bull, 33, 133-140.
[4] Lahiri D.N., (1954): *Technical paper on some aspects of the development of the sample design, Sankhya, Vol. 14, P332-362.*
[5] Leslie Kish, (1987): Statistical Design For Research, John Wiley & Sons., P33-
[6] Madow, W.G. and L.H. Madow, (1944): *On the theory of systematic sampling*, Ann. Math. Stat., 15, 1-24.
[7] Sethi, V.K., 1965: *On optimum pairing of units*, Sankhya, B 27, 315-320.
[8] Singh, P., Jindal, K.K., and Garg, J.N., 1968: *On modified systematic sampling*, Biometrika, 55, 541-546.
[9] N. Uthayakumaran, (1998): *Additional circular systematic sampling Methods*, Biometrical Journal, 40, 4, 467-474.
[10] 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.
[11] 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.
[12] N. Uthayakumaran, (2015): *Cuboidal systematic sampling method: An unbiased estimator in the presence of linear trend*, International journal of statistics, Accepted for publication.
[13] N. Uthayakumaran, (2015): *Multi-label systematic sampling design: An unbiased estimator in the presence of linear trend*, International journal of science and research, Vol. 4, Issue 8, 1343-1345.
[14] 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.