Jour. Ind. Soc. Ag. Statistics
50(1), 1997 : 70-74

Circular Systematic Sampling with Drawback

S. Ray and M.N. Das[1]
Central Statistical Organisation, New Delhi
(Received : December, 1995)

SUMMARY

A systematic design scheme which provides unbiased estimate of population mean and also estimate of variance of the mean without putting any restriction on the population size is presented.

Key words: Circular systematic sampling, Prime numbers modified circular systematic sampling, Modified linear systematic sampling, Estimate of variance.

1. *Introduction*

In linear systematic sampling, given a sample size n, sampling is possible only if population size N is divisible by n. Even when this condition is satisfied, the scheme cannot provide estimate of variance of the sample mean. This means the precision of the estimate of mean cannot be known. But the scheme has the advantage that it is very simple to draw sample according to the scheme and the sample represents the population well, particularly when there is any linear trend of the values of an interested variable in the population. So this scheme has two drawbacks namely, given N, n has limited choice and the variance of the sample mean is not estimable. The first limitation could be removed through circular systematic sampling as suggested by Lahiri [2]. But the second one remains. This drawback could be removed by Das [1] but only under the limitation that the population size should be prime.

In the present scheme, the second limitation has been removed without any limitation on population size.

2. *Sampling Scheme*

2.1 The Scheme of Das

In circular systematic sampling first a starting unit is selected at random from N units and then an interval k = [N/n] is taken where [N/n] means the

1 I-1703, Chittranjan Park, New Delhi-110019

integral part of N/n. Then every k-th unit is taken in the sample till the required number of units in the sample is obtained. So the systematic sampling scheme has two things viz. a random start and a given interval k.

In the scheme suggested by Das [1] both the interval and the start are selected at random. The start can be any number from 1 to N while the interval can be any number from 1 to N-1. Now, this scheme can provide unbiased estimate of population mean and also unbiased estimate of variance of the sample mean, provided the population size, N is prime or prime power as discussed below.

The sample space corresponding to the scheme has $N(N-1) = b$ samples, since start can be selected in N ways and for each start there are $(N-1)$ possible samples. If the start is always a given unit, say, i-th one in m-th position then for every interval, there will be a sample containing the unit giving in all $(N-1)$ samples. For every position of the unit in the sample there will be $(N-1)$ such samples. Therefore, total number of samples in each of which the i-th unit occurs in the scheme is $n(N-1) = r$.

It can be shown through similar reasoning that the number of samples in each of which a given pair of units say i-th and j-th units, occurs together is $n(n-1) = \lambda_{ij}$. Thus both $\Pi_i^{is} = \frac{r}{b}$ and $\Pi_{ij}^{is} = \frac{\lambda_{ij}}{b}$ are constant for this scheme. As such this scheme has similar procedure of estimation of both mean and variance as in simple random sampling without replacement (SRSWOR). This scheme has, however, a limitation that N should be prime or a prime power as indicated earlier. This limitation has been removed in the following section.

2.2 A Modified Circular Systematic Sampling Scheme

The scheme as mentioned in Section 2.1 is used when the population size N is a prime or prime power. But the scheme can be used for any value of N in the following way. Suppose N is not prime, then let N = N+ where N is the largest possible prime number less than N.

Now a sample of size d is selected from the population by SRSWOR and then the units in the sample are omitted from the population. Next from the remaining N units a circular systematic sample is drawn following the method of Das. It can be shown that this scheme has also the same Π_i^{is} and Π_{ij}^{is} as in the scheme with N as prime. Thus the scheme is made to work for any population size. d should preferably be as small as possible to retain the simplicity of the scheme. The scheme is illustrated in detail through the following example.

Example 2.2.1 Let N = 15 and n be any number, say 4. Now by omitting at random two units, 13 units are left. As 13 is a prime number, systematic sampling scheme of Das is appropriately adopted to get a sample. Total number of samples following from the scheme is

$$b = \binom{N}{2}(N-2)(N-3)$$

as 2 units can be omitted in ${}^{N}C_2$ ways and for each omission total number of possible samples is $(N-2)(N-3)$. Every unit occurs in $\binom{N-1}{2} n(N-3)$ samples.

$$\Pi_i = \frac{\binom{N-1}{2} n(N-3)}{\binom{N}{2}(N-2)(N-3)}$$

$$= \frac{n}{N} = \text{the same as in SRSWOR}$$

Also the number of samples in each of which two given units, say, the i-th and j-th units occur together is

$$\lambda_{ij} = \binom{N-2}{2} n(n-1)$$

therefore, $$\Pi_{ij} = \frac{\binom{N-2}{2} n(n-1)}{\binom{N}{2}(N-2)(N-3)}$$

$$= \frac{n(n-1)}{N(N-1)} = \text{the same as in SRSWOR}$$

In general it can be shown that

$$\Pi_i = \frac{\binom{N-1}{d} n(N-d-1)}{\binom{N}{d}(N-d)(N-d-1)}$$

$$= \frac{n}{N}$$

and
$$\lambda_{ij} = \frac{\binom{N-2}{d} n(n-1)}{\binom{N}{d}(N-d)(N-d-1)}$$

$$= \frac{n(n-1)}{N(N-1)}$$

The scheme also works when the interval is chosen at random from 1 to $(N-1)/2$ or $(N+1)/2$ to N, the proof of which follows what has been given by Das [1]. In this situation Π_i becomes n/N again but Π_{ij} becomes $n(n-1)/2N(N-1)$. But this does not affect variance and its estimate, as each of them depends on N and n only.

2.3 A Circular Systematic Method of Drawing the Sample

When certain units are omitted from the population at random, drawing a circular systematic sample from the remaining units becomes somewhat difficult if the remaining units are first renumbered and then a systematic sample is drawn. A simple technique is first to draw a circular systematic sample S_0 from N_0 units numbered serially and the units are ordered by following the method of Das. Also keep the serial numbers of the units omitted from the original population handy. Now take any unit number from S_0, say, the both one and denote it by n_p. Let there be m_p numbers in the omitted sample with unit numbers less than or equal to n_p, the unit selected is $n_p + m_p$ as per original numbering of the N units ($p = 1, 2, \ldots, n$). This technique gives the same sample as obtainable by renumbering the units, drawing the sample using such numbers and then again converting these sample numbers to original unit numbers, but without renumbering the units.

2.3.1 This example will clarify the above procedure

Let $N = 77$ and $N = N_0 + d$ where $N_0 = 73$ and $d = 4$. Now 4 units at random are to be removed from the population to make its size equal to 73 which is a prime number. Let 4 units omitted be 3, 11, 13, 50. Suppose a sample of size 8 is selected by following circular systematic sampling technique of Das with a random start 3 and interval of 7. So the sample units are 3, 10, 17, 24, 31, 38, 45, 52 after renumbering the leftout units. The original numbers of the units selected are 4, 11, 20, 34, 41, 48 and 56.

3. *Modified Linear Systematic Sampling*

The type of modification as discussed in Sub-section 2.2 using circular systematic sampling scheme works out for linear systematic sampling as well, when N = nk but N = nk + d. Here first d units are omitted at random and from the remaining units a linear systematic sample of size n is drawn. This scheme provides an unbiased estimate of the population mean, but no estimate of variance is available, as unlike circular systematic sampling scheme as in Section 2.2 all Π_{ij}^{is} are not greater than zero as can be easily shown through the following example.

Example 3.1

Let N = 8, n = 3, k = 2 and d = 2

Also let the units be identified as 1, 2, 3, 4, 5, 6, 7 and 8. By omitting two units at random a sample space, in which each sample is of size 3, has been worked out. It is seen that in the sample space the units 1 and 2 do not occur in any sample of the space. Similarly there are several other pairs also which do not occur together in any sample.

REFERENCES

[1] Das, M.N., 1982. Systematic sampling without drawback. Technical report No. 8206 (Discussion series), ISI, Delhi Campus.

[2] Lahiri, D.B., 1952. NSS instructions to field workers.