Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 71(1) 2017 61–65

Allocation of Sample in Stratified Sampling using Circular Systematic Sampling

Rajesh Kumar[1] and Trijya Singh[2]
ICAR-Indian Institute of Sugarcane Research, Lucknow
Le Moyne College, Syracuse, New York, USA

Received 19 January 2016; Accepted 11 January 2017

SUMMARY

In this paper, allocation of sample in the case of circular systematic sampling has been discussed when stratified sampling is used. Variances of the formulae have been discussed when the allocation is optimum or proportional. It has also been found that variances of the estimators in the case of circular systematic sampling (CSS) are more than those in the case of simple random sampling (SRS).

Keywords: Circular systematic sampling, Proportional and optimum allocation, Efficiencies.

1. INTRODUCTION

The Indian National Sample Survey (NSS) has been providing data on various demographic, economic and social aspects since its inception in 1950 through sample surveys conducted by its staff throughout rural and urban parts of India. In a single survey whose period is usually a year (called a 'round'), data on different aspects is collected by adopting a compromise sampling design which has been developed keeping in view the subjects of the survey, availability of the frame, operational requirements and errors of the estimates (Kumar and Dayal 1999 and 2000). Data are generally collected using the interview method and a moving reference period of a month or a week is usually used to reduce recall error.

The sampling design adopted by NSS is a stratified two-stage one. Each state or UT is treated as a broad stratum in NSS and within that stratum, each district is then treated as a sub-stratum. However, in urban areas, towns within a district are grouped into a number of further sub-strata according to population size. The first-stage units are villages and blocks in rural and urban areas respectively and second-stage units are generally households (Kumar and Dayal 1999 and 2000). The selection procedure of units has been changing both for rural and urban areas.

D.B. Lahiri suggested in 1952 (NSS Instructions to field workers) that the disadvantages of systematic sampling, namely the actual sample size being different from that which is required and the sample being a biased estimator of the population mean (where population size N is not a multiple of sample size n), can be overcome by adopting a device, known as circular systematic sampling Murthy (1967, Section 5.3, page 139). In fact, many of the shortcomings of systematic sampling can be overcome by CSS, including the adverse impact of stratification on systematic sampling (Dayal and Kumar 2007 and 2008).

2. PRELIMINARIES

Let a population of N units be divided into L strata. Let N_h be the number of units in the h^{th} stratum, and let Y_{hi} be the value of the study variable for the i^{th} unit in the h^{th} stratum. The population mean $\overline{Y}$ can be written as

Corresponding author: Rajesh Kumar
E-mail address: rajesh_iissr@hotmail.com

$$\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h \tag{2.1}$$

where, $\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_i} \bar{Y}_{hi}$ and $W_h = \frac{N_h}{N}$

Then, an unbiased estimator of $\overline{Y}$ can be obtained by estimating unbiasedly the stratum means ($\overline{Y}_h$) on the basis of random samples of size n_h, $h=1,2,...,L$ ($n = \sum n_h$) drawn from each stratum using a certain sampling scheme. Suppose $\overline{y}_h$ is an unbiased estimator of $\overline{Y}_h$, then an unbiased estimator of $\overline{Y}$ is given by

$$\bar{y} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{2.2}$$

and its sampling variance is

$$V_{st}\left(\bar{y}\right) = \sum_{h=1}^{L} W_h^2 V\left(\bar{y}_h\right)$$

$$V_{st}\left(\bar{y}\right) = \sum_{h=1}^{L} W_h^2 \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2 \frac{1}{(N_h - 1)n_h} \tag{2.3}$$

Also let, $S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$

and $S^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y})^2$

In the case of un-stratified population,

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

and $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2$

Dayal and Kumar (2007, 2008) considered the comparison of CSS with SRS (Simple Random Sampling) in un-stratified sampling. CSS consists in choosing a random start from 1 to *N* and selecting the unit corresponding to this random start and thereafter every k^{th} unit in a cyclical manner till a sample of *n* units is obtained, *k* being integer nearest to *N/n*, to ensure spread of the sample over the sampling frame. That is, if *r* is a number selected at random from 1 to *N*, the sample consists of units corresponding to the numbers.

$(r + jk)$, if $r + jk \le N$, $(j = 0,1,2,...,n - 1)$

and

$(r + jk - N)$, if $r + jk > N$, $(j = 0,1,2,...,n - 1)$ (2.4)

This technique ensures equal probability of inclusion in the sample for every unit. Here,

$$E(\overline{y}_{CSS}) = \frac{1}{N}\sum_{r=1}^{N}(\frac{1}{n}\sum_{i=1}^{n} y_i)_r = \frac{1}{nN}\sum_{r=1}^{N}(\sum_{i=1}^{n} y_i)_r = \bar{Y} \tag{2.5}$$

where *r* stands for the sample selected with random start. Also

$$V(\overline{y}_{CSS}) = \frac{1}{N}\sum_{r=1}^{N}(\overline{y}_r - \overline{Y})^2$$

where $\overline{y}_r = \frac{1}{n}\sum_{i=1}^{n} y_i$ (2.6)

$$= \frac{n}{n^2 N}\sum_{i=1}^{N}(\overline{y}_i - \overline{Y})^2$$

which is approximately equal to $\frac{S^2}{n}$. The variance can be checked from (12) of chapter IX of Sukhatme and Sukhatme (1970).

In the case of SRS,

$$V(\overline{y}_{SRS}) = \frac{1}{Nn}\sum_{r=1}^{k}\sum_{i=1}^{h}(y_{ri} - \overline{Y})^2 \tag{2.7}$$

which is simply $\frac{S^2}{n}$ (Murthy 1977, section 5.6). A comparison of (2.6) and (2.7) shows that V($\overline{Y}$) in the case of CCS is approximately the same as in case of SRS.

3. OPTIMUM AND PROPORTIONAL ALLOCATIONS

It can be seen from Section 2 of the paper of Dayal and Kumar (2007 and 2008) that

$$V(\bar{y}_{CSS}) = \frac{S^2}{n} \tag{3.1}$$

Therefore

$$V_{Stra}(\overline{y}_{CSS}) = \sum_{h=1}^{L}\frac{S_h^2}{n_h} \tag{3.2}$$

Using the method of Lagrange multiplier, it can be easily verified that the variance of (3.2) will be minimum and is given by

$$V_{Stra(\min)}(\overline{y}_{CSS}) = \frac{\left(\sum_{h=1}^{L} S_h\right)^2}{n} \tag{3.3}$$

The variance of the proportional allocation has already been worked out as given in Section 5.3 of Dayal and Kumar (2007, 2008). Thus,

$$V_{Prop}(\bar{y}_{css}) = \frac{1}{N}\sum_{h=1}^{L}\sum_{r=1}^{N}(\bar{y}_{rh} - \bar{Y})^2$$

or

$$V_{prop}(\bar{y}_{CSS}) = \sum_{h=1}^{L}\frac{S_h^2}{nW_h} \tag{3.4}$$

Thus, we get

$V_{prop}(\bar{y}_{CSS}) - V_{stra(\min)}(\bar{y}_{CSS})$, (which is > 0 since the second term is the smaller one).

$$= \sum_{h=1}^{L}\frac{S_h^2}{nW_h} - \frac{(\sum_{h=1}^{L}S_h)^2}{n}$$

$$< \frac{1}{n}[(\sum_{h=1}^{L}\frac{S_h}{\sqrt{W_h}} - \sum_{h=1}^{L}S_h)(\sum_{h=1}^{L}\frac{S_h}{\sqrt{W_h}} - \sum_{h=1}^{L}S_h)] \tag{3.5}$$

The second term is always positive since

$$\frac{1}{\sqrt{W_h}} > 1$$

Thus

$$V_{prop}(\bar{y}_{CSS}) > V_{stra(\min)}(\bar{y}_{CSS}) \tag{3.6}$$

It can seen from (5.35) of Cochran (1977) that

$$V_{ran} = V_{prop} + \frac{1-f}{n}\sum_{h=1}^{L}W_h(\bar{Y}_h - \bar{Y})$$

If we adopt the definition of randomness as given by Cochran (1977). This shows that V_{ran} is always greater than V_{prop}.

However, in the case of CSS, this position changes. From (3.2) and (3.4), it can be easily seen that

$$V_{ran}(\bar{y}_{css}) - V_{prop}(\bar{y}_{css}) = \sum_{h=1}^{L}(\frac{1}{n_h} - \frac{1}{n\,W_h})S_h^2$$

$$V_{ran}(\bar{y}_{css}) - V_{prop}(\bar{y}_{css}) = \sum_{h=1}^{L}\frac{1}{n_h}(1 - \frac{n_h N}{n\,N_h})S_h^2 \tag{3.7}$$

which is zero in the case where proportional allocation is adopted. Thus, proportion allocation is not of any use in the case of CSS if the definition of randomness, as given by Cochran (1977), is adopted. Another definition of randomness, as given by Dayal (1980) is worthwhile in such cases for which one may refer to section 5 of this paper. As expected, it is not always possible to adopt optimum allocation, since the values of S_h are not always known and even their estimates are biased in the case of CSS.

4. COMPARISON OF VARIANCES UNDER CSS AND SRS

A comparison of variances under the two systems is given by

$$V(\bar{y}_{css}) - V(\bar{y}_{SRS}) = \sum_{h=1}^{L}W_h S_h^2(\frac{1}{W_h^2} - 1) > 0 \tag{4.1}$$

since $W_h < 1$. Thus CSS results in higher variance of the estimator than SRS.

In the case of proportional allocation,

$$V_{prop}(\bar{y}_{css}) - V_{prop}(\bar{y}_{SRS}) = \frac{1}{n}[\sum_{h=1}^{L}\frac{S_h^2}{W_h} - \sum_{h=1}^{L}W_h S_h^2]$$

$$= \frac{1}{n}[\sum_{h=1}^{L}S_h^2(\frac{1}{W_h} - W_h)] > 0 \tag{4.2}$$

since $W_h < 1$. Thus, proportional allocation in the case of CSS results in higher variance of the estimator than that of SRS. Also, in case of optimum allocation,

$$V_{opti}(\bar{y}_{css}) - V_{opti}(\bar{y}_{SRS}) = \frac{(\sum_{h=1}^{L}S_h)^2}{n} - \frac{(\sum_{h=1}^{L}W_h S_h)^2}{n}$$

$$= \frac{1}{n}[(\sum_{h=1}^{L}S_h + \sum_{h=1}^{L}W_h S_h)(\sum_{h=1}^{L}S_h - \sum_{h=1}^{L}W_h S_h)] > 0 \tag{4.3}$$

since the second term is always positive. Thus,

$$V_{opti}(\bar{y}_{css}) > V_{opti}(\bar{y}_{SRS})$$

Thus, even in case of optimum allocation, CSS results in higher variance than SRS.

5. RANDOM ALLOCATION AS GIVEN BY COCHRAN (1977)

From (5.33) of Cochran (1977), it can be easily verified that

$$V_{ran}(\bar{y}_{css}) = \frac{S^2}{n}$$

$$= \frac{\sum_{h=1}^{L} W_h S_h^2}{n} + \frac{\sum_{h=1}^{L} W_h (\bar{Y}_h - \bar{Y})^2}{n} \quad (5.1)$$

where randomness is defined by Cochran (1977).

$V_{prop}(\bar{y}_{css})$ is available from (3.4).

Hence

$$V_{ran}(\bar{y}_{css}) - V_{prop}(\bar{y}_{css})$$

$$= \sum_{h=1}^{L} \frac{S_h^2}{n}\left(\frac{1}{W_h} - W_h\right) + \sum_{h=1}^{L} \frac{W_h}{n}(\bar{Y}_h - \bar{Y})^2 \quad (5.2)$$

Since $W_h < \frac{1}{W_h}$, the first term of (5.2) is negative while the second term is positive. Thus in the case of CSS, we cannot say that

$$V_{ran}(\bar{y}_{css}) \geq V_{prop}(\bar{y}_{css}) \quad (5.3)$$

It has already been proved that

$$V_{prop}(\bar{y}_{css}) \geq V_{opti}(\bar{y}_{css})$$

Thus, we get only

$V_{prop} \geq V_{opti}$ of CSS and not (5.3) in case the definition of randomness as given by Cochran (1977) is adopted.

6. RANDOM ALLOCATION AS GIVEN BY DAYAL (1980)

However, we define the random allocation as given by Dayal (1980) in section 3. This kind of random allocation can be described as follows. Suppose that we have L objects in a bag and we draw one object randomly from the bag. The number of draws should be n. After each draw, the object is returned to the bag. Suppose that the h^{th} object has appeared n_h times. Then we allocate a sample of size n_h to stratum h.

If $(1/L) = p$, the probabilities of having 0 to n units in a stratum under the random allocation are the terms in the expansion of $(q + p)^n$, where $q = 1 - p$. Thus n_h will be a random variable and $E(n_h) = n/L$. Variance of the mean, with fixed n_h, is given by

$$V_{ran}(\bar{y}_{st} / n_h) = \frac{1}{N}\left(\sum_{h=1}^{L} N_h^2 \frac{S_h}{n_h} - \sum_{h=1}^{L} N_h S_h^2\right)$$

In order to find the expected value of this when n_h varies as well, we are required to know E($1/n_h$). If the case in which n_h is zero is ignored, Stephan (1945) has shown that, to term of order n^{-2},

$$E\left(\frac{1}{n_h}\right) = \frac{1}{np} + \frac{1-p}{n^2 p^2} = \frac{L}{n} + \frac{1-L^{-1}}{n^2 L^{-2}}$$

$$E[V_{ran}(\bar{y}_{st} / n_h)] = \frac{1}{N^2}\left[\frac{1}{n}\sum L N_h^2 S_h^2 - \sum N_h S_h^2 + \frac{1}{n^2}\sum L(L-1) N_h^2 S_h^2\right]$$

By using (3.1) of Dayal (1980) as also given above, we find that

$$V_{ran}(\bar{y}_{css}) + V_{prop}(\bar{y}_{css})$$

$$= \left(\sum_{h=1}^{L} \frac{S_h^2}{nW_h} + \sum_{h=1}^{L} \frac{1}{n} L S_h^2\right) \sum_{h=1}^{L} \frac{1}{n} L(L-1) S_h^2$$

$$= \frac{1}{n}\left[\sum_{h=1}^{L} \frac{1}{n} L(L-1) + L - \frac{N}{N_h}\right] S_h^2 > 0$$

If there are equal number of units in each stratum or if necessary, adjusting sizes of the strata. Hence

$$V_{ran}(\bar{y}_{css}) \geq V_{prop}(\bar{y}_{css}) \geq V_{opti}(\bar{y}_{css}) \quad (6.1)$$

where random allocation is defined as described above, by Dayal (1980).

7. AN EMPIRICAL STUDY

The SERC, New Delhi and Action For Social Advancement (ASA), Bhopal, planned, organized and conducted a survey in Chhatarpur district of Madhya Pradesh. A stratified two stage design was adopted for the survey. The PACS (Poorest Area Civil Society) Programme comprised 269 DPIP. (District Poverty Initiative Programme) villages and 151 Non-DPIP villages and these two groups of villages formed the two strata.

Although villages in each stratum were selected following circular systematic sampling (CSS), only a fixed number of villages from each stratum were selected and the allocation principle was not based on proportional allocation or the minimization of overall variance. For the selection of a sample of 12 households from each sample village, the complete listing of households which had grown at least one of the four crops viz., soybean, arhar, wheat and gram, was made. Although 3 schedules were comprised, one of the schedules related to land use pattern, etc. A sample of 25 villages was selected from stratum 1 and 10 villages were selected from stratum 2. The enquiry method was adopted for the survey.

Estimated Area Productivity along with the standard error (SE) of the estimates for two crops, one each from the kharif and rabi season, are given in Table 1. Standard errors are worked out as in Dayal and Kumar, 2008, Section 6.2, since CSS has been adopted in this case.

Table 1. Estimated Area and Productivity during Kharif 2005-06 and Rabi 2006-07

Crops	DPIP Area (hectare)	Non-DPIP Area (hectare)	Projected Area (hectare)	Productivity (Q/ha) PA	Productivity (Q/ha) Non - PA
Soybean	15755.2 (785)	4457.8 (220)	20213.0	9.62 (0.5)	7.84 (0.4)
Wheat	28677.3 (1440)	29973.4 (1500)	58650.7	15.27 (0.8)	15.81 (0.8)

Note: The numbers in parentheses are SEs.

The estimates given in Table 1 are fairly satisfactory as estimates of standard errors and though they are biased, work out to be about 5% in each case.

Also, there is even no need to compare biased standard errors of stratified CSS with stratified SRS or stratified systematic sampling, as done by Dayal and Kumar (2007, 2008), Section 5, as the allocation of the sample in strata in the case of this example was done according to some criteria other than proportional allocation, as was taken by Dayal and Kumar (2007, 2008), Section 5.

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques*. Wiley Easter Ltd, New Delhi.

Dayal, S. and Kumar, Rajesh (2007, 2008). A note on circular systematic sampling. *Guj. Statist. Rev.,* **34 & 35 (1 & 2),** 12-20.

Dayal, S. (1980). On equal and random allocation of sample under stratified sampling. *J. Ind. Statist. Serv. Assoc.*, **10,** 63-69.

Kumar, Rajesh and Dayal, S. (1999, 2000). A review on sampling design of National Sample Survey. *Guj. Statist. Rev.,* **26 & 27 (1 & 2),** 5-16.

Murthy, M.N. (1967). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.

Murthy, M.N. (1977). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.

Stephen, F.F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Stat.,* **16,** 50-51.