Jour. Ind. Soc. Ag. Statistics
43 *(3), 1991 : 321-328*

Estimation of Population Mean in The Presence of Parabolic Trend

S. Sampath and K. Suresh Chandra*
Loyola College, Madras
(Received : November, 1989)

Summary

Construction of estimators of population mean in finite populations with trend have attracted the attention of many researchers. Recently Agrawal and Jain [1] have proposed two methods for construction of estimators, which coincide with population mean, under quadratic trend. Eventhough their estimators coincide with the population mean, they depend on some parameters which in general will not be known. In this paper two alternative estimators have been proposed which coincide with the population mean under quadratic trend and their performances have been assessed with the help of a super population model.

Key Words : Finite Population, Systematic Sampling, Super Population Models, Average Variance.

Introduction

Several sampling-estimating strategies have been introduced, by various authors for estimating the mean (total) of a finite population where the estimators coincide with the mean (total) when there is a perfect linear trend in the population. The following are some of the examples of such strategies:

(i) Linear systematic sampling with yates corrected mean as estimator;

(ii) Modified and Balanced systematic sampling schemes (Singh *et al.* [6] and Sethi [5] with sample mean as an estimator when the sample size is even; and

(iii) The centered systematic sampling with sample mean as estimator when the sampling interval $k = N/n$ is odd, N and n being the population and sample sizes respectively.

* University of Madras, Madras

Detailed comparative studies of the above strategies along with others have been carried out by Bellhouse and Rao [2] and Agrawal and Jain [1] for populations with linear and parabolic trend under certain fixed and superpopulation models. Eventhough such estimators coincide with the population mean in the above mentioned examples for populations with linear trend, it is not so in the case of quardratic trend.

Agrawal and Jain [1] have suggested the following two methods wherein the estimate coincides with the population mean for populations described by the model

$$Y_i = \mu + \theta_1 i + \theta_2 i^2, \quad i = 1, 2, \ldots, N \tag{1.1}$$

(i) They introduced a method of sampling called "extreme point sampling" where the first and last observations in the populations are taken as sample and the population mean is estimated by

$$(1 + \varphi) Y_1 + (1 - \varphi) Y_N$$

where $\varphi = (N-2)\frac{\theta_2}{3}\left[\theta_1 + (N+1)\theta_2\right]$

(ii) Choosing the sample using Linear systematic sampling and estimating the population mean using the sample mean by giving weights $1+\eta$ and $1-\eta$ for the first and last observations in the sample where n is as given in (5.1) of their paper.

It may be observed that eventhough their estimators coincide with the population mean the weights φ and η require the knowledge of the parameters involved in the model which in general will not be known. Motivated by this in the following section two alternative sampling-estimating strategies are presented for populations described by the model(1.1).

2. *Sampling-Estimating Strategies*

2.1 Alternative to extreme point sampling

As as alternative to the extreme point sampling we suggest the following three point sampling scheme:

Sampling Scheme: When the population size N is odd, choose the first, (N+1)/2-st and N-th units as sample. On the other hand if N is even, choose the first, N/2-st and N-th units as sample.

The following theorem which is easy to prove on routine algebra gives estimates which coincides with the population mean under the above scheme:

Theorem 2.1 : Under the three points sampling scheme

(i) $$\frac{1}{3}\left[\frac{N+1}{2(N-1)} Y_1 + 2\,\frac{N-2}{N-1} Y_{(N+1)/2} + \frac{N+1}{2(N-1)} Y_N\right]$$

is equal to the population mean when N is odd and

(ii) $$\frac{1}{3}\left[\frac{Y_1}{2} + 2\,\frac{N-1}{N} Y_{N/2} + \frac{N+4}{2N} Y_N\right]$$

is equal to the population mean when N is even.

2.2 Correcting the sample mean

The three point sampling scheme described above is devoid of randomisation. Here we shall develop an estimator which coincides with the population mean under linear systematic sampling. It may be noted that, linear systematic sampling can be regarded as dividing the population into k clusters of n units in a systematic manner and choosing a cluster at random.

When the ith cluster S_i is drawn as sample in linear systematic sampling and n is odd, we take

$$\bar{y}_3 = \frac{1}{n}\left[\gamma_{11} Y_i + \sum_{j=2}^{(n-1)/2} Y_{i+(j-1)k} + \gamma_{12} Y_{i+(n-1)k/2} + \sum_{j=(n+3)/2}^{n-1} Y_{i+(j-1)k} + \gamma_{13} Y_{i+(n-1)k}\right]$$

as an estimator of $\bar{Y}$. If n is even we use the estimator

$$\bar{y}_4 = \frac{1}{n}\left[\gamma_{21} Y_i + \sum_{j=2}^{(n-2)/2} Y_{i+(j-1)k} + \gamma_{22} Y_{i+(n-2)k/2} + \sum_{j=(n+2)/2}^{n-1} Y_{i+(j-1)k} + \gamma_{33} Y_{i+(n-1)k}\right]$$

Under the model (1.1) it can be seen that the estimators $\bar{y}_3$ and $\bar{y}_4$ are equal to $\bar{Y}$ under the respective cases when

$$\gamma_{11}(i) = \left\{\frac{2}{(n-2)^2}k^2\right\}\left[3\left\{i+\frac{(n-1)k}{2}\right\}\{1+(n-1)k\} + \partial_{12}(i) - \partial_{11}(i)\left\{2i+\frac{3(n-1)k}{2}\right\}\right]$$

$$\gamma_{12}(i) = \left\{\frac{4}{(n-2)^2}k^2\right\}\left[\partial_{11}(i)\{(n-1)k+2i\} - \partial_{12}(i) - 3i\{i+k(n-1)\}\right]$$

$$\gamma_{13}(i) = \left\{\frac{2}{(n-1)^2}k^2\right\}\left[\partial_{12}(i) - \partial_{11}(i)\left\{2i+\frac{k(n-1)}{2}\right\} + 3i\left\{i+\frac{k(n-1)}{2}\right\}\right] \qquad (2.1)$$

and

$$\gamma_{21}(i) = \left[(n-1)(n-2)k^2\right]^{-1}\left[3\{2i+(n-2)k\}\{i+(n-1)k\} + 2\partial_{22}(i) + \partial_{21}(i)\{(4-3n)k-4i\}\right]$$

$$\gamma_{22}(i) = 4\left[n(n-2)k^2\right]^{-1}\left[\partial_{21}(i)\{(n-1)k+2i\} - \partial_{22}(i) - 3i\{i+(n-1)k\}\right]$$

$$\gamma_{23}(i) = 2[n(n-1)k^2]^{-1}\left[\partial_{22}(i) - \partial_{21}(i)\left\{2i+\frac{k(n-2)}{2}\right\} + 3i\left\{i+\frac{k(n-2)}{2}\right\}\right]$$

where

$$\partial_{11}(i) = \frac{1}{2}[n(N+1) - k(n-3)(n-1)] - (n-3)i$$

$$\partial_{12}(i) = \frac{1}{6}[n(N+1)(2N+1)] - \frac{1}{12}[k^2(n-1)(n-3)(4n-5)] - (n-3)\{i^2 + ik(n-1)\}$$

$$\partial_{21}(i) = \frac{1}{2}[n(N+1) - (n-2)^2 k] - (n-3)\,i \qquad \text{and}$$

$$\partial_{22}(i) = \frac{1}{6}[n(N+1)(2N+1)] - \frac{1}{24}[k^2 (n-2)(8n^2 - 26n + 24)]$$

$$- (n-3)\,i^2 - (n-2)^2 ik \qquad (2.2)$$

As a result of the above discussion we infer that the estimators $\bar{y}_3$ and $\bar{y}_4$ will coincide with the population mean $\bar{Y}$ under the respective cases when the weights given in (2.1) and (2.2) are used accordingly. It is pertinent to note that the estimators developed in 2.1 as well as $\bar{y}_3$ and $\bar{y}_4$ do not require the knowledge of the parameters θ_1 and θ_2 unlike those developed by Agrawal and Jain [1].

3. *Comparison under Superpopulation Model*

In the earlier section of this paper two sampling- estimating strategies have been developed to estimate the population mean with no error when the population is modelled by (1.1). However in practice, we rarely encounter such populations. In this section we shall compare the above developed strategies with the help of the super population model

$$Y_i = \mu + \theta_1 i + \theta_2 i^2 + e_i \qquad (3.1)$$

where $E(e_i) = 0$; $\quad V(e_i) = \sigma_i^2\, g$, $\quad i = 1, 2, \ldots, N$.

Denote by M_1 and M_2 the average variance of the appropriate estimators under the three point sampling described in 2.1 and the corrected estimators developed in 2.2 respectively. It can be easily seen that

$$M_1 = \left(\frac{\sigma^2}{N^2}\right)\sum_{i=1}^{N} i^g + \left(\frac{\sigma^2}{9}\right)\left\{\left[\frac{(N+1)}{2(N-1)}\right]^2 + 4\left[\frac{(N-2)}{(N-1)}\right]^2\left[\frac{(N+1)}{2}\right]^g\right.$$

$$\left. + \left[\frac{N+1}{2(N-1)}\right]^2 N^g\right\} - \left(\frac{2\sigma^2}{3N}\right)\left\{\left[\frac{(N+1)}{2(N-1)}\right] + 2\left[\frac{(N-2)}{(N-1)}\right]\left[\frac{(N+1)}{2}\right]^g\right.$$

$$\left. + \left[\frac{N+1}{2(N-1)}\right] N^g\right\} \qquad \text{if N is odd.}$$

$$= \left(\frac{\sigma^2}{N^2}\right)\sum_{i=1}^{N} i^g + \left(\frac{\sigma^2}{9}\right)\left\{\left(\frac{1}{4}\right) + \left[\frac{2(N-1)}{N}\right]^2 \left(\frac{N}{2}\right)^g + \left[\frac{(N+4)}{2N}\right]^2 N^g\right\}$$

$$- \left(\frac{2\sigma^2}{3N}\right)\left\{\frac{1}{2} + \left[\frac{2(N-1)}{N}\right]\left(\frac{N}{2}\right)^g + \left[\frac{N+2}{2N}\right] N^g\right\} \qquad \text{if N is even}$$

$$M_2 = \left(\frac{\sigma^2}{k}\right)\sum_{i=1}^{k}\left[\frac{k-2}{n^2 k}\left\{\sum_{j=2}^{\frac{n-1}{2}} [i+(j-1)k]^g + \sum_{j=\frac{n+3}{2}}^{n-1} [i+(j-1)k]^g\right\}\right.$$

$$+ \left\{\frac{\gamma_{11}^2(i)}{n^2} - \frac{2\gamma_{11}(i)}{nN}\right\} i^g + \left\{\frac{\gamma_{12}^2(i)}{n^2} - \frac{2\gamma_{12}(i)}{nN}\right\}$$

$$\left.\left[i + \frac{(n-1)k}{2}\right]^g + \left\{\frac{\gamma_{13}^2(i)}{n^2} - \frac{2\gamma_{13}(i)}{nN}\right\}[i+(n-1)k]^g\right]$$

$$+ \frac{\sigma^2}{N^2}\sum_{i=1}^{N} i^g \qquad \text{if n is odd.}$$

$$= \left(\frac{\sigma^2}{k}\right)\sum_{i=1}^{k}\left[\frac{k-2}{n^2 k}\left\{\sum_{j=2}^{\frac{n-2}{2}} [i+(j-1)k]^g + \sum_{j=\frac{n+3}{2}}^{n-1} [i+(j-1)k]^g\right\}\right.$$

$$+ \left\{\frac{\gamma_{21}^2(i)}{n^2} - \frac{2\gamma_{21}(i)}{nN}\right\} i^g + \left\{\frac{\gamma_{22}^2(i)}{n^2} - \frac{2\gamma_{22}(i)}{nN}\right\}[i+(N/2)]^g$$

$$+ \left[\left\{\frac{\gamma_{23}^2(i)}{n^2} - \frac{2\gamma_{23}(i)}{nN}\right\}[i+(n-1)k]^g\right] + \frac{\sigma^2}{N^2}\sum_{i=1}^{N} i^g$$

if n is even.

where $\gamma_{11}(i)$, $\gamma_{12}(i)$, $\gamma_{13}(i)$ and $\gamma_{21}(i)$, $\gamma_{22}(i)$, $\gamma_{23}(i)$ are as defined in (2.1) and (2.2) respectively.

In view of the difficulties involved in comparing M_1 and M_2 theoretically, empirical comparisions have been made and the results are presented below:

Table 1. Value of the Ratio M_1/M_2

G	N = 25	N = 50		N = 100	
	n = 5	n = 5	n = 10	n = 10	n = 20
0	2.507	2.415	5.785	5.336	12.007
1	2.507	2.395	5.728	5.304	11.952
2	2.095	1.979	4.705	4.362	9.892

Eventhough both the methods of three point sampling and correcting the sample mean in linear systematic sampling estimate the population mean without any error under the fixed population model (1.1), their performance differ significantly under the super-population model (3.1). The corrected sample mean in linear systematic sampling is superior to the estimator suggested in three point sampling.

4. *Concluding Remarks*

1. Eventhough the proposed estimators, $\bar{y}_3$ and $\bar{y}_4$ have complicated forms, when compared to the simple expansion estimator based on linear systematic and sample random sampling schemes, manual computation of these estimators are not likely to be very cumbersome, because of their closed forms. It may be noted that the proposed estimators do not require the prior knowledge of certain population parameters as needed in the case of the estimator proposed by Agrawal and Jain [1].

2. There is an interesting example cited by Bellhouse [3] concerning sampling an archaeological site wherein the total number of artifacts tended to have an approximate quadratic trend. Practical situations like this enhance the applicability and usefulness of the proposed estimators.

3. Towards assessing the relative performances the proposed estimators over the mean estimator in LSS and SRS, an empirical study was carried with the data on the values of Y given in p.177, Table 5.11, (Data 1) and the first 80 values given in p.178, Table 5.12 (Data 2) of Murthy [4]. Table 2 summarized the relative performances.

   It may be noted that $\bar{y}_3/\bar{y}_4$ is not necessarily unbiased for

Table 2. Variance/Mean Square Error of Different Estimators

n	Data 1 N = 40		Data 2 N = 80			
	5	8	5	8	10	20
$\bar{y}_3/\bar{y}_4$	.08389	.196	402.4075	336.218	244.911	209.1048
$\bar{y}_{sys}$	13.4	6.925	562.5743	424.365	312.7744	218.146
$\bar{y}_{srs}$	65.16282	37.236	1163.748	698.249	543.0825	232.749

population mean if the population does not have a quadratic trend.

4. The authors are thankful to the Referee for suggesting an application of the concept of preliminary sampling to ascertain the presence of parabolic trend to make the entire exercise more meaningful.

ACKNOWLEDGEMENT

This work has been carried out under the UGC Minor Research Project in respect of the first author and under the UGC Research Scientistship in respect of the second author. The authors express their grateful thanks for the pertinent remarks by the Referee on the earlier version of this paper.

REFERENCES

[1] Agrawal, M.C. and Nirmal Jain, 1988. Comparison of some sampling strategies in the presence of trend, *Jour. Ind. Soc. Ag. Statistics*, **40(3)**, 191-201.

[2] Bellhouse, D.R. and Rao, J.N.K., 1975. Systematic sampling in the presence of trend, *Biometrika*, 694-697.

[3] Bellhouse, D.R., 1981. Spatial sampling in the presence of trend, *Journal of Statistical Planning and Inference*, **5(4)**, 365-375.

[4] Murthy, M.N., 1967. Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

[5] Sethi, V.K., 1965. On optimum pairing of units, *Sankhya*, **27(B)**, 315-320.

[6] Singh, D., Jindal K.K. and Garg, J.N., 1968. On modified systematic sampling, *Biometrika*, **55**, 541-546.