Journal of the Korean
Data & Imformation Science Society
1999, Vol. 10, No.1, pp. 91 ~ 102

A Study on Estimating Population Mean by Use of Interpolation and Extrapolation with Balanced Systematic Sampling

Hyuk Joo Kim[1]

Abstract

A new method is developed for estimating the mean of a population which has a linear trend. The suggested estimator is based on the balanced systematic sampling method and the concept of interpolation and extrapolation. The efficiency of the proposed method is compared with that of conventional methods.

Key Words and Phrases: Population with a linear trend, Balanced systematic sampling, Interpolation, Extrapolation, Infinite superpopulation model

1. Introduction

Quite often, we are interested in estimating the mean of a finite statistical population. In many cases, the numbering of the population units may be effectively random. But this is not always the case. Sometimes the population has a trend, which may be linear, curvilinear, or of other form.

In estimating the mean of a population which has a linear trend, ordinary systematic sampling (OSS) is known to be much better than simple random sampling (SRS). Several researchers have suggested sampling methods which are versions of systematic sampling. Among them, centered systematic sampling (CSS) proposed by Madow (1953), balanced systematic sampling (BSS) proposed by Sethi (1965) and named by Murthy (1967), and modified systematic sampling (MSS) proposed by Singh et al. (1968) are well-known methods.

Cochran (1946) introduced the concept of the infinite superpopulation model. Bellhouse and Rao (1975) discussed on comparisons of the performances of OSS, CSS, BSS and MSS. Iachan (1982) gave a comprehensive review of the developments in systematic sampling posterior to the review by Buckland (1951). He gave

[1]Associate Professor, Division of Mathematical Science, Wonkwang University, Iksan, Chonbuk, 570-749, Korea

emphasis to asymptotic and optimality results in the framework of superpopulation models. Fountain and Pathak (1989) discussed systematic sampling and nonrandom sampling for the case of populations with linear trends.

In this paper, a new method is developed for estimating the mean of a population which has a linear trend. The linear trend will be specified by using a mathematical expression in Section 3. The method will be developed for use in the case when the sample size n (≥ 3) is an odd number and k (the reciprocal of the sampling fraction) is an even number, and will be compared with several conventional methods under the expected mean square error criterion based on the infinite superpopulation model introduced by Cochran (1946).

2. Developing the method

Suppose we have a population of size $N = kn$, the units of which are denoted by $U_1, U_2, \cdots, U_N$. We wish to select a sample of size n from this population.

First, let us briefly review the balanced systematic sampling (BSS) method. This sampling method, proposed by Sethi (1965) and named by Murthy (1967) as stated in the previous section, was developed for populations having linear trends.

According to BSS, one of the k clusters $C_1', C_2', \cdots, C_k'$ is selected with respective probability $1/k$, and then the population mean is estimated by the sample mean, $\overline{y}_{bal}$, which is the mean of the selected cluster. Here "*bal*" represents balanced systematic sampling and the cluster C_i' is defined by

$$\begin{aligned} C_i' &= \{U_{i+2(j-1)k} : j = 1, 2, \cdots, n/2\} \cup \{U_{2jk+1-i} : j = 1, 2, \cdots, n/2\} \\ &\quad (i = 1, 2, \cdots, k) \end{aligned}$$

for n even, and

$$\begin{aligned} C_i' &= \{U_{i+2(j-1)k} : j = 1, 2, \cdots, (n+1)/2\} \cup \{U_{2jk+1-i} : j = 1, 2, \cdots, (n-1)/2\} \\ &\quad (i = 1, 2, \cdots, k) \end{aligned}$$

for n odd. For example, if $N = 20$, $n = 5$ and $k = 4$, then the four clusters are as follows :

$$\begin{aligned} C_1' &= \{U_1, U_8, U_9, U_{16}, U_{17}\} \\ C_2' &= \{U_2, U_7, U_{10}, U_{15}, U_{18}\} \\ C_3' &= \{U_3, U_6, U_{11}, U_{14}, U_{19}\} \\ C_4' &= \{U_4, U_5, U_{12}, U_{13}, U_{20}\}. \end{aligned}$$

The sample mean $\overline{y}_{bal}$ obtained by BSS is easily seen to be an unbiased estimator of $\overline{Y}$, the population mean, with variance

$$V(\overline{y}_{bal}) = \frac{1}{k}\sum_{i=1}^{k}(\overline{y}_i^{'} - \overline{Y})^2,$$

where $\overline{y}_i^{'}$ is the mean value for the units in $C_i^{'}$ $(i = 1, 2, \cdots, k)$.

Throughout this paper the following notation will be used :

y_i : value for the ith unit in the population $(i = 1, 2, \cdots, N)$,

$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$: population mean to be estimated,

$y_{ij}^{'}$: value for the jth unit in $C_i^{'}$ $(i = 1, 2, \cdots, k;\ j = 1, 2, \cdots, n)$, that is,

$$\begin{aligned} y_{ij}^{'} &= y_{i+(j-1)k}\ (j = 1, 3, 5, \cdots, n-1) \\ y_{ij}^{'} &= y_{1-i+jk}\ (j = 2, 4, 6, \cdots, n) \end{aligned}$$

for n even, and

$$\begin{aligned} y_{ij}^{'} &= y_{i+(j-1)k}\ (j = 1, 3, 5, \cdots, n) \\ y_{ij}^{'} &= y_{1-i+jk}\ (j = 2, 4, 6, \cdots, n-1) \end{aligned}$$

for n odd,

$\overline{y}_i^{'} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}^{'}$: mean for the units in $C_i^{'}$ $(i = 1, 2, \cdots, k)$.

Now we introduce a new method for estimating the population mean $\overline{Y}$. This method involves the same sampling method as BSS, but it estimates $\overline{Y}$ by an adjusted estimator, not by the sample mean itself. We only consider the case when n is an odd number $(n \geq 3)$ and k is an even number, because the method is defined and has a practical meaning in this case.

Consider again the case of $N = 20$, $n = 5$ and $k = 4$. One of $C_1^{'}, C_2^{'}, C_3^{'}, C_4^{'}$ is selected with respective probability 1/4. We notice that the sums of the numbers assigned to the units in $C_1^{'}, C_2^{'}, C_3^{'}$ and $C_4^{'}$ are, respectively, 51, 52, 53 and 54, showing differences from 1 to 3. When the population has a linear trend, it would be desirable to remove such differences. Balance would be obtained if y_{17}, y_{18}, y_{19}, or y_{20} is replaced by "$y_{18.5}$" according as $C_1^{'}, C_2^{'}, C_3^{'}$, or $C_4^{'}$ is selected. Here, of course, $y_{18.5}$ is an imaginary value which does not actually exist.

If $C_1^{'}$ is selected, then we can "estimate" $y_{18.5}$ by use of y_{16} and y_{17}. By the extrapolation method, $y_{18.5}$ is estimated by $(1/2)(5y_{17} - 3y_{16})$. Therefore, by using this value instead of y_{17}, we can estimate $\overline{Y}$ by

$$\overline{y}_1^{'*} = \frac{1}{5}\{y_1 + y_8 + y_9 + y_{16} + \frac{1}{2}(5y_{17} - 3y_{16})\}$$

$$= \overline{y}_1' + \frac{3}{10}(y_{17} - y_{16}).$$

If the selected cluster is C_2', then $y_{18.5}$ can be estimated by $(1/6)(7y_{18} - y_{15})$, and using this value instead of y_{18} we can estimate $\overline{Y}$ by

$$\begin{aligned}\overline{y}_2'^* &= \frac{1}{5}\{y_2 + y_7 + y_{10} + y_{15} + \frac{1}{6}(7y_{18} - y_{15})\} \\ &= \overline{y}_2' + \frac{1}{30}(y_{18} - y_{15}).\end{aligned}$$

Note that $\overline{y}_1'^*$ and $\overline{y}_2'^*$ can also be expressed as

$$\overline{y}_1'^* = \overline{y}_1' + \frac{3}{10}(y_{15}' - y_{14}')$$

and

$$\overline{y}_2'^* = \overline{y}_2' + \frac{1}{30}(y_{25}' - y_{24}'),$$

where 14, 15, 24 and 25 subscript to y' are two-dimensional.

Suppose now that the selected cluster is C_3'. Then we can estimate $y_{18.5}$ by using y_{14} and y_{19}. We now need to use the method of interpolation because 18.5 is between 14 and 19. Using the resultant value instead of y_{19}, we can estimate $\overline{Y}$ by

$$\begin{aligned}\overline{y}_3'^* &= \frac{1}{5}\{y_3 + y_6 + y_{11} + y_{14} + \frac{1}{10}(y_{14} + 9y_{19})\} \\ &= \overline{y}_3' - \frac{1}{50}(y_{19} - y_{14}) \\ &= \overline{y}_3' - \frac{1}{50}(y_{35}' - y_{34}').\end{aligned}$$

A similar argument enables us to estimate $\overline{Y}$ by

$$\begin{aligned}\overline{y}_4'^* &= \frac{1}{5}\{y_4 + y_5 + y_{12} + y_{13} + \frac{1}{14}(3y_{13} + 11y_{20})\} \\ &= \overline{y}_4' - \frac{3}{70}(y_{20} - y_{13}) \\ &= \overline{y}_4' - \frac{3}{70}(y_{45}' - y_{44}')\end{aligned}$$

in the case when C_4' is selected.

The above method can be generalized as follows. One of the k clusters $C_1', C_2', \cdots, C_k'$ is selected with respective probability $1/k$. If the selected cluster is C_i', then the population mean $\overline{Y}$ is estimated by $\overline{y}_i^{'*}$, where

$$\begin{aligned}\overline{y}_i^{'*} &= \frac{1}{n}\{y_{i1}^{'} + y_{i2}^{'} + \cdots + y_{i,n-1}^{'} + y_{in}^{'} + \frac{\frac{k+1}{2} - i}{2i-1}(y_{in}^{'} - y_{i,n-1}^{'})\} \\ &= \overline{y}_i^{'} + \frac{k+1-2i}{2n(2i-1)}(y_{in}^{'} - y_{i,n-1}^{'}) \end{aligned} \quad (1)$$

for $i = 1, 2, \cdots, k/2$, and

$$\begin{aligned}\overline{y}_i^{'*} &= \frac{1}{n}\{y_{i1}^{'} + y_{i2}^{'} + \cdots + y_{i,n-1}^{'} + y_{in}^{'} - \frac{i - \frac{k+1}{2}}{2i-1}(y_{in}^{'} - y_{i,n-1}^{'})\} \\ &= \overline{y}_i^{'} - \frac{2i-k-1}{2n(2i-1)}(y_{in}^{'} - y_{i,n-1}^{'}) \end{aligned} \quad (2)$$

for $i = k/2+1, k/2+2, \cdots, k$. Note that formulas (1) and (2) are identical. So we can see that the extrapolation and interpolation methods give the same formula for $\overline{y}_i^{'*}$.

Let us denote the estimator of $\overline{Y}$ resulting from the method described above by $\overline{y}_{ieb}$. Here "*ieb*" represents interpolation, extrapolation and balanced systematic sampling. Then $\overline{y}_{ieb}$ is biased for $\overline{Y}$ and it is clear that $\overline{y}_{ieb}$ has bias

$$B(\overline{y}_{ieb}) = \frac{1}{k}\sum_{i=1}^{k}\overline{y}_i^{'*} - \overline{Y}$$

and mean square error

$$MSE(\overline{y}_{ieb}) = \frac{1}{k}\sum_{i=1}^{k}(\overline{y}_i^{'*} - \overline{Y})^2.$$

3. Expected mean square error of $\overline{y}_{ieb}$

In this section, the expected mean square error of $\overline{y}_{ieb}$ is obtained using Cochran's (1946) infinite superpopulation model.

We regard the finite population as a sample from an infinite superpopulation. First, as a general case, we set up the model as

$$y_i = \mu_i + e_i \ (i = 1, 2, \cdots, N), \quad (3)$$

where μ_i is a function of i and the random error e has properties $E(e_i) = 0$, $E(e_i^2) = \sigma^2$, $E(e_i e_j) = 0$ $(i \neq j)$. The operator E denotes the expectation over the infinite superpopulation.

From now on, with regard to μ and e also we will use the same style of notation as adopted for y. That is,

$$\begin{aligned}
\overline{\mu} &= \frac{1}{N}\sum_{i=1}^{N}\mu_i, \\
\mu'_{ij} &= \mu_{i+(j-1)k}\ (j=1,3,5,\cdots,n)\ (n\ :\ \text{odd}), \\
\overline{\mu}'_i &= \frac{1}{n}\sum_{j=1}^{n}\mu'_{ij}, \\
\overline{\mu}'^{*}_i &= \overline{\mu}'_i + \frac{k+1-2i}{2n(2i-1)}(\mu'_{in}-\mu'_{i,n-1}),
\end{aligned}$$

and so on.

The following theorem is very important in evaluating the efficiency of $\overline{y}_{ieb}$.

Theorem 1. Assuming the model expressed as (3), the expected mean square error of $\overline{y}_{ieb}$ for k even and n odd ($n \geq 3$) is

$$E\{MSE(\overline{y}_{ieb})\} = \frac{1}{k}\sum_{i=1}^{k}(\overline{\mu}'^{*}_i - \overline{\mu})^2 + \frac{\sigma^2}{n}\frac{N-n}{N} + \frac{\sigma^2}{2n^2}(1-\gamma-2ln2+A_k) \qquad (4)$$

where

$$\begin{aligned}
\gamma &= 0.57721566490\cdots\ :\ \text{Euler constant} \\
A_k &= \frac{\pi^2}{8}k - \psi(k+\frac{1}{2}) - \frac{k}{4}\psi^{(1)}(k+\frac{1}{2}) \\
\psi(x) &= \frac{d}{dx}\ln\Gamma(x)\ (x>0)\ :\ \text{polygamma function} \\
\Gamma(x) &= \int_0^\infty t^{x-1}e^{-t}dt\ (x>0)\ :\ \text{gamma function} \\
\psi^{(1)}(x) &= \frac{d}{dx}\psi(x)
\end{aligned}$$

Proof. We know that

$$MSE(\overline{y}_{ieb}) = \frac{1}{k}\sum_{i=1}^{k}(\overline{y}'^{*}_i - \overline{Y})^2, \qquad (5)$$

and by (3) we obtain

$$\overline{Y} = \overline{\mu} + \overline{e}. \qquad (6)$$

On the other hand, from (3) it can be written that

$$y'_{ij} = \mu'_{ij} + e'_{ij}\ (i=1,2,\cdots,k;\ j=1,2,\cdots,n), \qquad (7)$$

from which we obtain

$$\overline{y}_i^{'*} = \overline{\mu}_i^{'*} + \overline{e}_i^{'*} \ (i = 1, 2, \cdots, k). \tag{8}$$

Substituting (6) and (8) into (5) and taking expectation, we have

$$\begin{aligned} E\{MSE(\overline{y}_{ieb})\} &= \frac{1}{k}\sum_{i=1}^{k} E[\{(\overline{\mu}_i^{'*} - \overline{\mu}) + (\overline{e}_i^{'*} - \overline{e})\}^2] \\ &= \frac{1}{k}\sum_{i=1}^{k}[(\overline{\mu}_i^{'*} - \overline{\mu})^2 + E\{(\overline{e}_i^{'*} - \overline{e})^2\}]. \end{aligned} \tag{9}$$

We also have, for $i = 1, 2, \cdots, k$,

$$\begin{aligned} E\{(\overline{e}_i^{'*} - \overline{e})^2\} &= E\{(\overline{e}_i^{'} - \overline{e} + P_i)^2\} \\ &= E\{(\overline{e}_i^{'} - \overline{e})^2\} + 2E\{(\overline{e}_i^{'} - \overline{e})P_i\} + E(P_i^2), \end{aligned} \tag{10}$$

where

$$P_i = \frac{k+1-2i}{2n(2i-1)}(e_{in}^{'} - e_{i,n-1}^{'}). \tag{11}$$

We further have, for $i = 1, 2, \cdots, k$,

$$E\{(\overline{e}_i^{'} - \overline{e})^2\} = E\{(\overline{e}_i^{'})^2\} - 2E\{(\overline{e}_i^{'})(\overline{e})\} + E\{(\overline{e})^2\} \tag{12}$$

and

$$\begin{aligned} E\{(\overline{e}_i^{'})^2\} &= E\left\{\left(\frac{1}{n}\sum_{j=1}^{n} e_{ij}^{'}\right)^2\right\} \\ &= \frac{1}{n^2}E\left\{\sum_{j=1}^{n}(e_{ij}^{'})^2 + 2\sum_{j<j'}(e_{ij}^{'})(e_{ij'}^{'})\right\} \\ &= \frac{1}{n^2}\left[\sum_{j=1}^{n} E\{(e_{ij}^{'})^2\} + 2\sum_{j<j'} E\{(e_{ij}^{'})(e_{ij'}^{'})\}\right] \\ &= \frac{\sigma^2}{n}, \end{aligned} \tag{13}$$

and similarly

$$E\{(\overline{e}_i^{'})(\overline{e})\} = E\{(\overline{e})^2\} = \frac{\sigma^2}{N}. \tag{14}$$

The second term in the rightmost side of (10) is easily shown to be zero, and the third term is also easily obtained as

$$E(P_i^2) = \frac{(k+1-2i)^2\sigma^2}{2n^2(2i-1)^2}. \tag{15}$$

Substitution of these results into (10) gives

$$E\{(\overline{e}_i^{'*} - \overline{e})^2\} = \frac{\sigma^2}{n}\frac{N-n}{N} + \frac{(k+1-2i)^2\sigma^2}{2n^2(2i-1)^2} \ (i = 1, 2, \cdots, k). \tag{16}$$

Substituting (16) into (9), we have

$$E\{MSE(\overline{y}_{ieb})\} \ = \ \frac{1}{k}\sum_{i=1}^{k}(\overline{\mu}_i^{'*} - \overline{\mu})^2 + \frac{\sigma^2}{n}\frac{N-n}{N} + \frac{\sigma^2}{2kn^2}\sum_{i=1}^{k}\left(1 - \frac{k}{2i-1}\right)^2, \tag{17}$$

and we obtain (4) by using the following well-known mathematical relations. (See, for reference, Abramowitz and Stegun (1982, pp. 258-260).)

$$\sum_{i=1}^{k}\frac{1}{2i-1} = \frac{1}{2}\{\gamma + 2ln2 + \psi(k+\frac{1}{2})\}$$
$$\sum_{i=1}^{k}\frac{1}{(2i-1)^2} = \frac{1}{8}\{\pi^2 - 2\psi^{(1)}(k+\frac{1}{2})\}$$

□

Now, let us consider the case of $\mu_i = a + bi$, where a and b are constants with $b \neq 0$. In other words, the assumed model is

$$y_i = a + bi + e_i \ (i = 1, 2, \cdots, N). \tag{18}$$

This is the case of a population which has a linear trend.

In this case, as a preparatory stage for obtaining $E\{MSE(\overline{y}_{ieb})\}$ we get the following formulas :

$$\overline{\mu} = a + \left(\frac{b}{2}\right)(N+1) \tag{19}$$

$$\overline{\mu}_i^{'} = a + \left(\frac{b}{2}\right)(N+1) + \left(\frac{b}{n}\right)\left(i - \frac{k+1}{2}\right) \tag{20}$$

$$\mu_{in}^{'} = \mu_{i+(n-1)k} = a + b\{i + (n-1)k\} \tag{21}$$

$$\mu'_{i,n-1} = \mu_{1-i+(n-1)k} = a + b\{1 - i + (n-1)k\} \tag{22}$$

Thus we have, for $i = 1, 2, \cdots, k$,

$$\begin{aligned} \overline{\mu}_i'^{*} &= \overline{\mu}'_i + \frac{k+1-2i}{2n(2i-1)}(\mu'_{in} - \mu'_{i,n-1}) \\ &= a + \left(\frac{b}{2}\right)(N+1). \end{aligned} \tag{23}$$

Now using (19), (23) and the result of Theorem 1, we obtain the following theorem:

Theorem 2. For a population characterized by (18), the expected mean square error of $\overline{y}_{ieb}$ is

$$E\{MSE(\overline{y}_{ieb})\} = \frac{\sigma^2}{n}\frac{N-n}{N} + \frac{\sigma^2}{2n^2}(1 - \gamma - 2ln2 + A_k)(k : \text{even}, n : \text{odd}, n \geq 3), \tag{24}$$

where A_k is as defined in Theorem 1.

4. Comparison with conventional methods

In this section, the efficiency of $\overline{y}_{ieb}$ is compared with that of estimators resulting from conventional methods. Let us consider SRS, OSS, MSS, BSS and CSS. Bellhouse and Rao (1975) also discussed on comparisons of the performances of OSS, MSS, BSS and CSS.

For a population characterized by the model (18), the following were obtained in Kim (1985) :

$$E\{MSE(\overline{y}_{ran})\} = \left(\frac{b^2}{12}\right)(N+1)(k-1) + \frac{\sigma^2}{n}\frac{N-n}{N} \tag{25}$$

$$E\{MSE(\overline{y}_{sys})\} = \left(\frac{b^2}{12}\right)(k+1)(k-1) + \frac{\sigma^2}{n}\frac{N-n}{N} \tag{26}$$

$$E\{MSE(\overline{y}_{mod})\} = E\{MSE(\overline{y}_{bal})\} = \left(\frac{b^2}{12n^2}\right)(k+1)(k-1) + \frac{\sigma^2}{n}\frac{N-n}{N} \quad (n : \text{odd}) \tag{27}$$

$$E\{MSE(\overline{y}_{cen})\} = \frac{b^2}{4} + \frac{\sigma^2}{n}\frac{N-n}{N} \quad (k : \text{even}) \tag{28}$$

Here $\overline{y}_{ran}, \overline{y}_{sys}, \overline{y}_{mod}, \overline{y}_{bal}$ and $\overline{y}_{cen}$ denote the samlpe mean, which is used as the estimator of $\overline{Y}$, obtained from SRS, OSS, MSS, BSS and CSS, respectively.

On the basis of formulas (24) through (28), we can arrange the methods under consideration according to the magnitude of the expected mean square error as the following theorem. For simplicity's sake, $E\{MSE(\overline{y}_{ieb})\}$ is abbreviated as "ieb", $E\{MSE(\overline{y}_{sys})\}$ as "sys", and so on. Thus, for example, "$ieb < sys$" means that our proposed method using $\overline{y}_{ieb}$ is more efficient than OSS.

Theorem 3. Let B_k denote $1 - r - 2ln2 + A_k$. For a population having a linear trend represented by (18), the following hold :

(1) The case of $k = 2$ and $n = 3, 5, 7, \cdots$

(i) If $\sigma^2 < 9b^2/10$, then $ieb < mod = bal < cen = sys < ran$.

(ii) If $9b^2/10 \leq \sigma^2 < 9b^2n^2/10$, then $mod = bal \leq ieb < cen = sys < ran$.

(iii) If $9b^2n^2/10 \leq \sigma^2 < 3b^2n^2(N+1)/10$, then $mod = bal < cen = sys \leq ieb < ran$.

(iv) If $3b^2n^2(N+1)/10 \leq \sigma^2$, then $mod = bal < cen = sys < ran \leq ieb$.

(2) The case of $k = 4, 6, 8, \cdots$, $n = 3, 5, 7, \cdots$ and $n < \sqrt{(k^2-1)/3}$

(i) If $\sigma^2 < b^2n^2/2B_k$, then $ieb < cen < mod = bal < sys < ran$.

(ii) If $b^2n^2/2B_k \leq \sigma^2 < b^2(k^2-1)/6B_k$, then $cen \leq ieb < mod = bal < sys < ran$.

(iii) If $b^2(k^2-1)/6B_k \leq \sigma^2 < b^2n^2(k^2-1)/6B_k$, then $cen < mod = bal \leq ieb < sys < ran$.

(iv) If $b^2n^2(k^2-1)/6B_k \leq \sigma^2 < b^2n^2(N+1)(k-1)/6B_k$, then $cen < mod = bal < sys \leq ieb < ran$.

(v) If $b^2n^2(N+1)(k-1)/6B_k \leq \sigma^2$, then $cen < mod = bal < sys < ran \leq ieb$.

(3) The case of $k = 4, 6, 8, \cdots$, $n = 3, 5, 7, \cdots$ and $n = \sqrt{(k^2-1)/3}$ (for example, $k = 26$ and $n = 15$)

(i) If $\sigma^2 < b^2n^2/2B_k$, then $ieb < cen = mod = bal < sys < ran$.

(ii) If $b^2n^2/2B_k \leq \sigma^2 < b^2n^2(k^2-1)/6B_k$, then $cen = mod = bal \leq ieb < sys < ran$.

(iii) If $b^2n^2(k^2-1)/6B_k \leq \sigma^2 < b^2n^2(N+1)(k-1)/6B_k$, then $cen = mod = bal < sys \leq ieb < ran$.

(iv) If $b^2n^2(N+1)(k-1)/6B_k \leq \sigma^2$, then $cen = mod = bal < sys < ran \leq ieb$.

(4) The case of $k = 4, 6, 8, \cdots, n = 3, 5, 7, \cdots$ and $n > \sqrt{(k^2-1)/3}$

(i) If $\sigma^2 < b^2(k^2-1)/6B_k$, then $ieb < mod = bal < cen < sys < ran$.

(ii) If $b^2(k^2-1)/6B_k \leq \sigma^2 < b^2n^2/2B_k$, then $mod = bal \leq ieb < cen < sys < ran$.

(iii) If $b^2n^2/2B_k \leq \sigma^2 < b^2n^2(k^2-1)/6B_k$, then $mod = bal < cen \leq ieb < sys < ran$.

(iv) If $b^2n^2(k^2-1)/6B_k \leq \sigma^2 < b^2n^2(N+1)(k-1)/6B_k$, then $mod = bal < cen < sys \leq ieb < ran$.

(v) If $b^2n^2(N+1)(k-1)/6B_k \leq \sigma^2$, then $mod = bal < cen < sys < ran \leq ieb$.

Example. Suppose that we wish to draw a sample of size $n = 25$ from a population consisting of $N = 750$ units. We have $k = 750/25 = 30$. Assume that the slope of the linear trend is $b = 0.5$. Then by use of MATHEMATICA we get $\psi(30.5) = 3.401244$ and $\psi^{(1)}(30.5) = 0.0333302$, and hence

$$\begin{aligned} A_{30} &= \frac{30}{8}\pi^2 - \psi(30.5) - \frac{30}{4}\psi^{(1)}(30.5) = 33.359796, \\ B_{30} &= 1 - \gamma - 2ln2 + A_{30} = 32.396286. \end{aligned}$$

Therefore, by (4) of Theorem 3, the efficiency of the estimation methods can be compared as follows :

(i) If $\sigma^2 < 1.1563$, then $ieb < mod = bal < cen < sys < ran$.

(ii) If $1.1563 \leq \sigma^2 < 2.4115$, then $mod = bal \leq ieb < cen < sys < ran$.

(iii) If $2.4115 \leq \sigma^2 < 722.6587$, then $mod = bal < cen \leq ieb < sys < ran$.

(iv) If $722.6587 \leq \sigma^2 < 17506.9901$, then $mod = bal < cen < sys \leq ieb < ran$.

(v) If $17506.9901 \leq \sigma^2$, then $mod = bal < cen < sys < ran \leq ieb$.

We can see from this example that our proposed method is relatively efficient as compared with other methods unless σ^2 is unrealistically large.

5. Concluding remarks

In this paper, a new method was developed for estimating the mean of a population of size $N = kn$ which has a linear trend, for the case of k even and n odd ($n \geq 3$). The method consists of selecting a sample of size n by BSS, and then computing the value of the estimator of the population mean which was developed by using the concept of interpolation and extrapolation.

The developed estimator turned out to be relatively efficient as compared with the conventional estimators if σ^2, the variance of the random error term in the infinite superpopulation model, is not unrealistically large. The proposed estimator was found to be especially efficient as the value of σ^2 becomes smaller.

References

1. Abramowitz, M. and Stegun, I. A. (1982). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, U.S. Department of Commerce.

2. Bellhouse, D.R. and Rao, J.N.K. (1975). Systematic sampling in the presence of a trend, *Biometrika*, 62, 694-697.

3. Buckland, W.R. (1951). A review of the literature of systematic sampling, *Journal of the Royal Statistical Society*, B13, 208-215.

4. Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations, *Annals of Mathematical Statistics*, 17, 164-177.

5. Fountain, R.L. and Pathak, P.K. (1989). Systematic and nonrandom sampling in the presence of linear trends, *Communications in Statistics - Theory and Methods*, 18, 2511-2526.

6. Iachan, R. (1982). Systematic sampling : a critical review, *International Statistical Review*, 50, 293-303.

7. Kim, H.J. (1985). New systematic sampling methods for populations with linear or parabolic trends, *Unpublished Master Thesis,* Department of Computer Science and Statistics, Seoul National University.

8. Madow, W.G. (1953). On the theory of systematic sampling, III. Comparison of centered and random start systematic sampling, *Annals of Mathematical Statistics*, 24, 101-106.

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

10. Sethi, V.K. (1965). On optimum pairing of units, *Sankhya*, B27, 315-320.

11. Singh, D. Jindal, K.K. and Garg, J.N. (1968). On modified systematic sampling, *Biometrika*, 55, 541-546.