The Korean Communications
in Statistics Vol. 11 No. 1, 2004
pp. 1-12

# Centralization of Yates' End Corrections Method[1]

**Hyuk Joo Kim[2]**

## Abstract

In this paper, Yates' end corrections method is centralized to estimate the mean of a population which has a linear trend in the case of $k$ (the reciprocal of the sampling fraction) even. The efficiency of the resultant method is compared with that of existing methods.

*Keywords* : linear trend, centered systematic sampling, end corrections, infinite superpopulation model

## 1. Introduction

It is often of interest to estimate the mean of a statistical population. Suppose that the population has an increasing or decreasing linear trend. It is well known that systematic sampling is efficient in such a case. Systematic sampling includes ordinary systematic sampling(OSS), centered systematic sampling(CSS), balanced systematic sampling(BSS) and modified systematic sampling(MSS). CSS(Madow(1953)), as the name implies, comes from the centralization of OSS. BSS(Sethi(1965) and Murthy(1967)) and MSS(Singh et al.(1968)) are sampling methods made by giving some modifications to OSS. CSS, BSS and MSS were found to be more efficient than OSS. Kim(1985) proposed centered balanced systematic sampling(CBSS) and centered modified systematic sampling(CMSS). These two sampling methods, obtained by centralizing BSS and MSS respectively, turned out to be more efficient than BSS and MSS when $n$(the sample size) is an odd integer and $k$(the reciprocal of the sampling fraction) is an even integer.

On the other hand, Yates(1948) proposed a method called end corrections(EC) method. This involves the same sampling method as OSS. But the population mean is estimated by a weighted mean, not by the simple mean, of the sample values. This method was found to be more efficient than OSS as the linear trend in the population becomes stronger.

In this paper, we apply the concept of centralization to Yates' EC method. So the resultant method is obtained by using the same technique as in EC for the sample selected by CSS.

---

1) This paper was supported by Wonkwang University in 2003.

2) Professor, Division of Mathematics & Informational Statistics and Institute of Basic Natural Sciences, Wonkwang University, Iksan, Jeonbuk 570-749, Korea.
E-mail : hjkim@wonkwang.ac.kr

## 2. Centralization of end corrections

Let $N$ and $n$ denote the population size and the sample size, respectively. $k=N/n$ is supposed to be an integer. We denote the $N$ units of the population as $U_1, U_2, \cdots, U_N$. The population is partitioned into $k$ clusters $S_1, S_2, \cdots, S_k$, where the $i$th cluster $S_i$ is defined by

$$S_i=\{U_{i+(j-1)k}: j=1,2,\cdots,n\}.(i=1,2,\cdots,k)$$

Throughout this paper the following notation will be used:

$y_i$ : value for the $i$th 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 $j$th unit in $S_i$ $(i=1,2,\cdots,k;\ j=1,2,\cdots,n)$,

that is, $y_{ij}=y_{i+(j-1)k}$

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

Let us briefly review Yates' EC. A sample of size $n$ is drawn by OSS, that is, one of $S_1, S_2, \cdots, S_k$ is selected with respective probability $1/k$. If the selected cluster is $S_i$ then $\overline{Y}$ is estimated by giving the weights $\frac{1}{n}+\frac{2i-k-1}{2k(n-1)}$ and $\frac{1}{n}-\frac{2i-k-1}{2k(n-1)}$ to the first and the last units in the sample respectively instead of the usual weight $1/n$, that is, the estimator of $\overline{Y}$ is

$$\overline{y}_{EC}=\overline{y}_i+\frac{2i-k-1}{2k(n-1)}(y_{i1}-y_{in}). \tag{2.1}$$

Just as CSS was obtained by centralizing OSS, a new method can be obtained by centralizing EC. So the resultant method consists of drawing a sample of size $n$ by CSS and estimating the population mean $\overline{Y}$ using EC. The procedure is described specifically as follows. If $k$ is odd, then $S_{(k+1)/2}$ is selected as a sample with probability one and $\overline{Y}$ is estimated by $\overline{y}_{(k+1)/2}$, that is, the simple sample mean. So the method is the same as CSS in this case. If $k$ is even, either $S_{k/2}$ or $S_{k/2+1}$ is selected with equal probability. If $S_{k/2}$ is selected, $\overline{Y}$ is estimated by

$$\overline{y}^*_{k/2}=\overline{y}_{k/2}-\frac{1}{2k(n-1)}(y_{k/2,1}-y_{k/2,n}). \tag{2.2}$$

If $S_{k/2+1}$ is selected, then $\overline{Y}$ is estimated by

$$\bar{y}^*_{k/2+1} = \bar{y}_{k/2+1} + \frac{1}{2k(n-1)}(y_{k/2+1,1} - y_{k/2+1,n}). \tag{2.3}$$

Let us denote the method described above as CEC and the resultant estimator of $\bar{Y}$ as $\bar{y}_{CEC}$. Then it is obvious that the bias and the mean square error of $\bar{y}_{CEC}$ are as follows:

$$Bias(\bar{y}_{CEC}) = \frac{1}{2}(\bar{y}^*_{k/2} + \bar{y}^*_{k/2+1}) - \bar{Y} \tag{2.4}$$

$$MSE(\bar{y}_{CEC}) = \frac{1}{2}\{(\bar{y}^*_{k/2} - \bar{Y})^2 + (\bar{y}^*_{k/2+1} - \bar{Y})^2\}. \tag{2.5}$$

## 3. Expected bias and expected mean square error of the estimator

In this section, we will derive the expected bias and the expected mean square error of $\bar{y}_{CEC}$ by using the infinite superpopulation model. This model, due to Cochran(1946), regards the finite population as a sample drawn from an infinite superpopulation. First, as a general case, we set up the model as

$$y_i = \mu_i + e_i \quad (i = 1, 2, \cdots, N), \tag{3.1}$$

where $\mu_i$ 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 $\mu$ and $e$ the same style of notation as adopted for $y$ will be used. That is

$$\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mu_i,$$
$$\mu_{ij} = \mu_{i+(j-1)k},$$
$$\bar{\mu}_i = \frac{1}{n}\sum_{j=1}^{n}\mu_{ij},$$

and so on.

The following theorem is important in evaluating the efficiency of $\bar{y}_{CEC}$. Since CEC is the same as CSS when $k$ is an odd number, it is sufficient to consider only the case when $k$ is an even number from now on.

**Theorem 3.1.** Under the model expressed as (3.1), the expected bias and the expected mean square error of $\bar{y}_{CEC}$ for $k$ even are as follows:

$$\begin{aligned} EBias(\bar{y}_{CEC}) = & \frac{1}{2}\{\bar{\mu}_{k/2} - \frac{1}{2k(n-1)}(\mu_{k/2,1} - \mu_{k/2,n}) \\ & + \bar{\mu}_{k/2+1} + \frac{1}{2k(n-1)}(\mu_{k/2+1,1} - \mu_{k/2+1,n})\} - \bar{\mu}, \end{aligned} \tag{3.2}$$

$$
\begin{aligned}
&EMSE(\bar{y}_{CEC}) \\
&= \frac{1}{2}\{(\bar{\mu}_{k/2}-\bar{\mu})^2+(\bar{\mu}_{k/2+1}-\bar{\mu})^2\}-\frac{1}{2k(n-1)}(\bar{\mu}_{k/2}-\bar{\mu})(\mu_{k/2,1}-\mu_{k/2,n}) \\
&\quad+\frac{1}{2k(n-1)}(\bar{\mu}_{k/2+1}-\bar{\mu})(\mu_{k/2+1,1}-\mu_{k/2+1,n}) \\
&\quad+\frac{1}{8k^2(n-1)^2}\{(\mu_{k/2,1}-\mu_{k/2,n})^2+(\mu_{k/2+1,1}-\mu_{k/2+1,n})^2\} \\
&\quad+\frac{\sigma^2}{2k^2(n-1)^2}+\sigma^2\left(\frac{1}{n}-\frac{1}{N}\right).
\end{aligned} \tag{3.3}
$$

Proof. (3.2) is derived by straightforward calculation. As for (3.3), substituting (2.2) and (2.3) into (2.5), we have

$$
\begin{aligned}
&MSE(\bar{y}_{CEC}) \\
&= \frac{1}{2}[\{(\bar{y}_{k/2}-\bar{Y})-\frac{1}{2k(n-1)}(y_{k/2,1}-y_{k/2,n})\}^2 \\
&\quad+\{(\bar{y}_{k/2+1}-\bar{Y})+\frac{1}{2k(n-1)}(y_{k/2+1,1}-y_{k/2+1,n})\}^2] \\
&= \frac{1}{2}\{(\bar{y}_{k/2}-\bar{Y})^2+(\bar{y}_{k/2+1}-\bar{Y})^2\}-\frac{1}{2k(n-1)}(\bar{y}_{k/2}-\bar{Y})(y_{k/2,1}-y_{k/2,n}) \\
&\quad+\frac{1}{8k^2(n-1)^2}(y_{k/2,1}-y_{k/2,n})^2 \\
&\quad+\frac{1}{2k(n-1)}(\bar{y}_{k/2+1}-\bar{Y})(y_{k/2+1,1}-y_{k/2+1,n}) \\
&\quad+\frac{1}{8k^2(n-1)^2}(y_{k/2+1,1}-y_{k/2+1,n})^2.
\end{aligned}
$$

Taking expectation over the infinite superpopulation and using (3.1), we have the following (for $i=k/2, k/2+1$):

$$
\begin{aligned}
E\{(\bar{y}_i-\bar{Y})^2\} &= E[\{(\bar{\mu}_i-\bar{\mu})+(\bar{e}_i-\bar{e})\}^2] \\
&= (\bar{\mu}_i-\bar{\mu})^2+E\{(\bar{e}_i-\bar{e})^2\} \\
&= (\bar{\mu}_i-\bar{\mu})^2+E(\bar{e}_i^2)-2E(\bar{e}_i\bar{e})+E(\bar{e}^2) \\
&= (\bar{\mu}_i-\bar{\mu})^2+\frac{\sigma^2}{n}-2\cdot\frac{\sigma^2}{N}+\frac{\sigma^2}{N} \\
&= (\bar{\mu}_i-\bar{\mu})^2+\sigma^2\left(\frac{1}{n}-\frac{1}{N}\right), \\
E\{(\bar{y}_i-\bar{Y})(y_{i1}-y_{in})\} &= E[\{(\bar{\mu}_i-\bar{\mu})+(\bar{e}_i-\bar{e})\}\{(\mu_{i1}-\mu_{in})+(e_{i1}-e_{in})\}] \\
&= (\bar{\mu}_i-\bar{\mu})(\mu_{i1}-\mu_{in})+E\{(\bar{e}_i-\bar{e})(e_{i1}-e_{in})\} \\
&= (\bar{\mu}_i-\bar{\mu})(\mu_{i1}-\mu_{in}),
\end{aligned}
$$

(since the second term is easily shown to be zero)

$$
\begin{aligned}
E\{(y_{i1}-y_{in})^2\} &= E[\{(\mu_{i1}-\mu_{in})+(e_{i1}-e_{in})\}^2] \\
&= (\mu_{i1}-\mu_{in})^2+E\{(e_{i1}-e_{in})^2\} \\
&= (\mu_{i1}-\mu_{in})^2+2\sigma^2.
\end{aligned}
$$

(3.3) is derived by using the above formulas.

Now let us consider the case when the population has a linear trend. The linear trend is represented by $\mu_i = a + bi$, where $a$ and $b$ are constants with $b \neq 0$. Thus the assumed model is

$$y_i = a + bi + e_i \qquad (i = 1, 2, \cdots, N). \tag{3.4}$$

In this case, we get the following formulas:

$$\begin{aligned} \mu_{ij} &= a + b\{i + (j-1)k\} \\ \bar{\mu}_{k/2} &= a + (\frac{b}{2})kn \\ \bar{\mu}_{k/2+1} &= a + (\frac{b}{2})(kn+2) \\ \bar{\mu} &= a + (\frac{b}{2})(kn+1). \end{aligned}$$

Substituting these formulas into (3.2) and (3.3) gives the following theorem:

**Theorem 3.2.** For a population having a linear trend represented by (3.4), the expected bias and the expected mean square error of $\bar{y}_{CEC}$ for $k$ even are as follows:

$$EBias(\bar{y}_{CEC}) = 0, \tag{3.5}$$

$$EMSE(\bar{y}_{CEC}) = \sigma^2(\frac{1}{n} - \frac{1}{N}) + \frac{\sigma^2}{2k^2(n-1)^2}. \tag{3.6}$$

## 4. Comparison of efficiency with existing methods

### 4.1 Comparison with methods using simple mean of the sample

Let us compare the efficiency of CEC with the methods that draw a sample by SRS, SSRS, OSS, CSS, BSS, MSS, CBSS, CMSS, CBS, CMS and TES, and estimate $\bar{Y}$ with the simple sample mean. Here SRS represents simple random sampling, SSRS represents stratified simple random sampling with one unit per stratum, and CBS, CMS and TES represent centered balanced sampling, centered modified sampling and two-end sampling, respectively. The latter three sampling methods are due to Fountain and Pathak(1989). SSRS is such that the $j$th stratum $(j = 1, 2, \cdots, n)$ consists of units $U_{1+(j-1)k}$, $U_{2+(j-1)k}, \cdots, U_{jk}$. From each stratum one unit is selected at random. Since all strata are of equal size and one unit is selected from each stratum, the estimator of $\bar{Y}$ simplifies to the sample mean. Bellhouse and Rao(1975) also gave discussions on comparisons of the performances of OSS, CSS, BSS, MSS and EC.

When $k$ is an even number, the common form of the expected mean square error of the estimators for $\bar{Y}$ by those methods is

$$EMSE(\hat{\bar{Y}}) = \sigma^2(\frac{1}{n} - \frac{1}{N}) + b^2 f(n,k), \tag{4.1}$$

where $f(n,k)$ is a function of $n$ and $k$. The specifications for each method are as follows:

$$f(n,k)=\begin{cases} \frac{(nk+1)(k-1)}{12} & \text{(SRS)} \\ \frac{k^2-1}{12n} & \text{(SSRS)} \\ \frac{k^2-1}{12} & \text{(OSS)} \\ \frac{1}{4} & \text{(CSS)} \\ 0 \quad (n:\text{ even}) & \text{(BSS, MSS, CBSS, CMSS, CBS, CMS, TES)} \\ \frac{k^2-1}{12n^2} \quad (n:\text{ odd}) & \text{(BSS, MSS)} \\ \frac{1}{4n^2} \quad (n:\text{ odd}) & \text{(CBSS, CMSS, CBS, CMS, TES)} \end{cases} \tag{4.2}$$

Efficiency of CEC and these methods can be compared ,using (3.6), (4.1) and (4.2). The necessary and sufficient condition for $EMSE(\bar{y}_{CEC})$ to be less than $EMSE(\hat{\bar{Y}})$ is

$$\sigma^2 < 2b^2k^2(n-1)^2 f(n,k). \tag{4.3}$$

This means that CEC becomes especially efficient as compared with these methods as $\sigma^2$, the variance of the random error term in (3.3), becomes smaller.

Of course it should be noted that when $n$ is even, $f(n,k)=0$ for BSS, MSS, CBSS, CMSS, CBS, CMS and TES, so that CEC is less efficient than these seven methods. Hence CEC is worth using when $k$ is even and $n$ is odd.

As a numerical example, consider the case when $N=500, n=25, k=20$, and the slope of the linear trend is $b=0.8$. Then the conditions for CEC to be more efficient than (i) SRS, (ii) OSS, (iii) SSRS, (iv) CSS, (v) BSS and MSS, (vi) CBSS, CMSS, CBS, CMS and TES are, respectively, that (i) $\sigma^2 < 233,938,944$, (ii) $\sigma^2 < 9,805,824$, (iii) $\sigma^2 < 392,232.96$, (iv) $\sigma^2 < 73,728$, (v) $\sigma^2 < 15,689.3184$, and (vi) $\sigma^2 < 117.9648$.

Breidt(1995) studied Markov chain designs, a wide class of methods for one-per-stratum selection from a finite population. Given a doubly stochastic transition probability matrix $P$, a Markov chain sample is given by

$$S=\{U_{R_1}, U_{k+R_2}, \cdots, U_{(n-1)k+R_n}\},$$

where $R_1,\cdots,R_n$ is the Markov chain defined by $P$ and $R_1$ is uniformly distributed over $\{1,2,\cdots,k\}$. Breidt obtained the result that in estimating $\bar{Y}$ under the linear trend model (3.4), any Markov chain design has

$$EMSE(\bar{y}_{MC}) = \sigma^2\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{b^2}{n^2}\right)V_{MC}\left(\sum_{i=1}^{n} R_i\right). \tag{4.4}$$

The Markov chain designs include SSRS, OSS, BSS and MSS as special cases. From (3.6) and (4.4) we see that the condition for CEC to be more efficient than a Markov chain design is that

$$\sigma^2 < \frac{2b^2k^2(n-1)^2}{n^2} V_{MC}(\sum_{i=1}^{n} R_i). \tag{4.5}$$

In (4.5), $V_{MC}(\sum_{i=1}^{n} R_i)$ denotes the variance of $\sum_{i=1}^{n} R_i$ under the given Markov chain design. For example,

$$V_{SSRS}(\sum_{i=1}^{n} R_i) = \frac{n(k^2-1)}{12},$$

$$V_{OSS}(\sum_{i=1}^{n} R_i) = \frac{n^2(k^2-1)}{12},$$

$$V_{BSS}(\sum_{i=1}^{n} R_i) = V_{MSS}(\sum_{i=1}^{n} R_i) = \begin{cases} 0 & \text{for } n \text{ even}, \\ \frac{k^2-1}{12} & \text{for } n \text{ odd}. \end{cases}$$

### 4.2 Comparison with methods using weighted mean of the sample

The following methods estimate $\overline{Y}$ with a weighted mean of the sample values.

(1) End corrections method (EC) (Yates, 1948)

$$EMSE(\overline{y}_{EC}) = \sigma^2\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{(k^2-1)\sigma^2}{6k^2(n-1)^2} \tag{4.6}$$

(2) Method using BSS and interpolation (BI) (Kim, 2000b)
(3) Method using BSS, interpolation and extrapolation (BIE) (Kim, 1999)
(4) Method using MSS and interpolation (MI) (Kim, 1998)
(5) Method using MSS, interpolation and extrapolation (MIE) (Kim and Oh, 2002)
(6) Method using CBSS and interpolation (CBI) (Kim and Seok, 2000)
(7) Method using CBSS, interpolation and extrapolation (CBIE) (Kim, 2000a)
(8) Method using CMSS and interpolation (CMI) (Kim and Choi, 2002)
(9) Method using CMSS, interpolation and extrapolation (CMIE) (Kim and Jung, 2002)

Methods (2) through (9) are defined for the case of $k$ even and $n$ odd. The common form of $EMSE(\widehat{\overline{Y}})$ resulting from these methods is

$$EMSE(\widehat{\overline{Y}}) = \sigma^2\left(\frac{1}{n} - \frac{1}{N}\right) + \sigma^2 g(n,k), \tag{4.7}$$

where $g(n,k)$ is a function of $n$ and $k$. So $EMSE(\widehat{\overline{Y}})$ does not depend on $b$, the slope of the linear trend. Details of the form of $g(n,k)$ for methods (2) through (9) are as follows:

$$g(n,k)=\begin{cases}\frac{1}{2n^2}(1-4A_k+2kB_k) & \text{(BI)}\\ \frac{1}{2n^2}(1-\gamma-2\ln 2+C_k) & \text{(BIE)}\\ \frac{1}{12n^2}(4-12A_k+6kB_k) & \text{(MI)}\\ \frac{k^2-1}{6n^2k^2} & \text{(MIE)}\\ \frac{1}{2n^2(k+1)^2} & \text{(CBI)}\\ \frac{1}{4n^2}\left\{\frac{1}{(k-1)^2}+\frac{1}{(k+1)^2}\right\} & \text{(CBIE)}\\ \frac{1}{4n^2}\left\{\frac{1}{k^2}+\frac{1}{(k+1)^2}\right\} & \text{(CMI)}\\ \frac{1}{2n^2k^2} & \text{(CMIE)}\end{cases} \tag{4.8}$$

where

$$A_k=\frac{1}{2}\left\{\Psi(k+\frac{1}{2})-\Psi(\frac{k+1}{2})\right\}$$

$$B_k=-\frac{1}{4}\left\{\Psi^{(1)}(k+\frac{1}{2})-\Psi^{(1)}(\frac{k+1}{2})\right\}$$

$$C_k=\frac{k}{8}\left\{\pi^2-2\,\Psi^{(1)}(k+\frac{1}{2})\right\}-\Psi(k+\frac{1}{2})$$

$\Psi(x)=\frac{d}{dx}\ln\Gamma(x)\ (x>0)$ : the polygamma function

$\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt\ (x>0)$ : the gamma function

$$\Psi^{(1)}(x)=\frac{d}{dx}\Psi(x)$$

$\gamma=0.577215\cdots\cdots$ : the Euler constant

From (3.6) and (4.6) we obtain

$$EMSE(\bar{y}_{EC})-EMSE(\bar{y}_{CEC})=\frac{(k^2-4)\sigma^2}{6k^2(n-1)^2}.$$

This is always positive if $k$ is an even number greater than 2. Hence CEC is more efficient than EC when $k>2$. It is obvious that CEC and EC are the same methods when $k=2$.

Tables 1, 2 and 3 present values of $EMSE(\widehat{Y})/\sigma^2$ by methods (1) through (9) and CEC for some values of $n$ in the case of $k=8,12,20$. As we see from these tables, the five methods, namely, CEC, CBI, CBIE, CMI and CMIE, are more efficient than EC, BI, BIE, MI and MIE (although the differences become smaller as $n$ increases with $k$ fixed). This fact tells us that centralization is a useful tool when we estimate the mean of a population which has a linear trend.

In the tables the values of $1/n-1/N$ are also presented for given $k$ and $n$. $\sigma^2(1/n-1/N)$ is the value of $EMSE(\widehat{Y})$ resulting from CSS when $k$ is odd, and from BSS and MSS when $n$ is even. CSS when $k$ is odd, and BSS and MSS when $n$ is even, are known as completely trend-free sampling methods. Since there is little difference between $EMSE(\bar{y}_{CEC})/\sigma^2$ and $1/n-1/N$ as we see from the tables, we can say that CEC, together with CBI, CBIE, CMI and CMIE, is nearly optimal for estimating the mean of a population with a linear trend.

Table 1: Values of $EMSE(\widehat{Y})/\sigma^2$ when $k=8$

| $n$ | EC | BI | BIE | MI | MIE | CBI | CBIE | CMI | CMIE | CEC | $\frac{1}{n}-\frac{1}{N}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | .1853 | .1773 | .3065 | .1794 | .1816 | .1752 | .1753 | .1753 | .1753 | .1755 | .1750 |
| 25 | .0353 | .0351 | .0403 | .0352 | .0353 | .0350 | .0350 | .0350 | .0350 | .0350 | .0350 |
| 55 | .0160 | .0159 | .0170 | .0159 | .0160 | .0159 | .0159 | .0159 | .0159 | .0159 | .0159 |
| 105 | .0083 | .0083 | .0086 | .0083 | .0083 | .0083 | .0083 | .0083 | .0083 | .0083 | .0083 |

Table 2: Values of $EMSE(\widehat{Y})/\sigma^2$ when $k=12$

| $n$ | EC | BI | BIE | MI | MIE | CBI | CBIE | CMI | CMIE | CEC | $\frac{1}{n}-\frac{1}{N}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | .1936 | .1855 | .4054 | .1877 | .1899 | .1835 | .1835 | .1835 | .1835 | .1836 | .1833 |
| 25 | .0370 | .0368 | .0456 | .0369 | .0370 | .0367 | .0367 | .0367 | .0367 | .0367 | .0367 |
| 55 | .0168 | .0167 | .0185 | .0167 | .0168 | .0167 | .0167 | .0167 | .0167 | .0167 | .0167 |
| 105 | .0087 | .0087 | .0092 | .0087 | .0087 | .0087 | .0087 | .0087 | .0087 | .0087 | .0087 |

Table 3: Values of $EMSE(\widehat{Y})/\sigma^2$ when $k=20$

| $n$ | EC | BI | BIE | MI | MIE | CBI | CBIE | CMI | CMIE | CEC | $\frac{1}{n}-\frac{1}{N}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | .2004 | .1923 | .5993 | .1945 | .1967 | .1900 | .1901 | .1900 | .1901 | .1901 | .1900 |
| 25 | .0383 | .0381 | .0544 | .0382 | .0383 | .0380 | .0380 | .0380 | .0380 | .0380 | .0380 |
| 55 | .0174 | .0173 | .0207 | .0173 | .0174 | .0173 | .0173 | .0173 | .0173 | .0173 | .0173 |
| 105 | .0090 | .0090 | .0100 | .0090 | .0090 | .0090 | .0090 | .0090 | .0090 | .0090 | .0090 |

### 4.3 A numerical example

Consider the following artificial population of size $N=36$. A sample of size $n=9$ is to be drawn from this population.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 45 | 44 | 42 | 39 | 40 | 38 | 37 | 33 | 35 | 36 |
| 35 | 32 | 30 | 31 | 27 | 27 | 24 | 23 | 25 | 27 |
| 23 | 19 | 18 | 18 | 17 | 13 | 15 | 13 | 16 | 15 |
| 14 | 10 | 11 | 12 | 9 | 8 | | | | |

The mean of this population is $\overline{Y}=25.0278$. This population is exhibiting a decreasing linear trend. The mean square errors of the estimators of $\overline{Y}$ resulting from existing methods are

$MSE(\overline{y}_{SRS})=10.2928$, $MSE(\overline{y}_{SSRS})=0.4205$, $MSE(\overline{y}_{OSS})=1.9282$,
$MSE(\overline{y}_{CSS})=0.2693$, $MSE(\overline{y}_{BSS})=0.1566$, $MSE(\overline{y}_{MSS})=0.0394$,
$MSE(\overline{y}_{CBSS})=0.0965$, $MSE(\overline{y}_{CMSS})=0.0471$, $MSE(\overline{y}_{CBS})=0.3233$,
$MSE(\overline{y}_{CMS})=0.1088$, $MSE(\overline{y}_{TES})=0.9576$, $MSE(\overline{y}_{EC})=0.0995$,
$MSE(\overline{y}_{BI})=0.2610$, $MSE(\overline{y}_{BIE})=0.1711$, $MSE(\overline{y}_{MI})=0.0544$,
$MSE(\overline{y}_{MIE})=0.0776$, $MSE(\overline{y}_{CBI})=0.1267$, $MSE(\overline{y}_{CBIE})=0.1323$,
$MSE(\overline{y}_{CMI})=0.0313$, $MSE(\overline{y}_{CMIE})=0.0256$.

On the other hand, if we estimate the population mean by using the method CEC, $\overline{Y}$ is estimated as one of the following two values:

$$\overline{y}_2^*=25.1667,$$
$$\overline{y}_3^*=25.1823.$$

Hence the mean square error of our estimator $\overline{y}_{CEC}$ is obtained as

$$MSE(\overline{y}_{CEC})=0.0216.$$

This shows that CEC is the most efficient of the methods considered.

## 5. Concluding remarks

When a population under consideration has a linear trend, the population parameter can be efficiently estimated by using well-devised methods. In this paper, a new method denoted by CEC has been suggested to estimate the mean of a population which has a linear trend. When $k$ is an odd number, CEC has little meaning because it is the same as CSS, which, resulting in a nonrandom sample, is not appropriate to be used in practice. So CEC is actually suggested for use in the case of $k$ even.

Finally, it is to be noted that the expected mean square errors in equations such as (3.6), (4.1), (4.6) and (4.7) are not anticipated mean square errors that were obtained by a model-assisted method, but they were obtained by using a model-based method.

## References


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

[2] Breidt, F. J.(1995), Markov chain designs for one-per-stratum sampling, *Survey Methodology*, **21**, 63-70.

[3] 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.

[4] 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.

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

[6] Kim, H. J.(1998), Estimation of population mean using interpolation in modified systematic sampling, *Korean Annals of Mathematics*, **15**, 217-231.

[7] Kim, H. J.(1999), A study on estimating population mean by use of interpolation and extrapolation with balanced systematic sampling, *Journal of the Korean Data & Information Science Society*, **10**, 91-102.

[8] Kim, H. J.(2000a), A new estimator of population mean based on centered balanced systematic sampling, *Journal of the Korean Data & Information Science Society*, **11**, 91-101.

[9] Kim, H. J.(2000b), Estimation of mean using balanced systematic sampling and interpolation for population with linear trend, *Journal of the Korean Statistical Society*, **29**, 455-471.

[10] Kim, H. J. and Choi, B. C.(2002), Efficient estimation of population mean using centered modified systematic sampling and interpolation, *The Korean Communications in Statistics*, **9**, 175-185.

[11] Kim, H. J. and Jung, S.-J.(2002), On the effectiveness of centering, interpolation and extrapolation in estimating the mean of a population with linear trend, *Journal of the Korean Data & Information Science Society*, **13**, 365-379.

[12] Kim, H. J. and Oh, E.-S.(2002), A study on estimation of the population mean using modified systematic sampling, interpolation and extrapolation, *Korean Annals of Mathematics*, **19**, 251-267.

[13] Kim, H. J. and Seok, E.-Y.(2000), Efficient estimation of the mean for populations with a linear trend: an extension of systematic sampling, *The Korean Journal of Applied Statistics*, **13**, 457-476.

[14] 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.

[15] Murthy, M. N.(1967), Sampling Theory and Methods, Statistical Publishing Society, Calcutta, India.

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

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

[18] Yates, F.(1948), Systematic sampling, *Philosophical Transactions of the Royal Society of London*, **A241**, 345-377.

[ Received August 2003, Accepted January 2004 ]