The Korean Communications
in Statistics Vol. 9, No. 1, 2002
pp. 175-185

Efficient Estimation of Population Mean Using Centered Modified Systematic Sampling and Interpolation

Hyuk Joo Kim[1] and Byoung Chul Choi[2]

Abstract

A method is proposed for efficiently estimating the mean of a population which has a linear trend. The proposed estimator is based on the centered modified systematic sampling method and the concept of interpolation. Using the expected mean square error criterion, it is shown that the proposed method is more efficient than conventional methods in most real cases.

Keywords : Linear trend, Centered modified systematic sampling, Interpolation, Infinite superpopulation model.

1. 서론

유한모집단의 평균을 추정하는 문제를 생각해 보자. 모집단의 단위들이 특성값에 관해 랜덤한 순서로 배열되어 있지 않고 어떤 추세(trend)를 가지고 있는 경우가 있다. 이 추세는 선형 또는 곡선형일 수도 있고 그 밖의 형태일 수도 있다.

본 논문에서는 선형 추세를 갖는 모집단의 평균을 효율적으로 추정하기 위한 새로운 방법을 제시하고자 한다. 선형 추세는 3절에서 통계적 모형에 의하여 정의될 것이다. 제시되는 방법은 표본크기 n이 3 이상의 홀수이고 추출률의 역수인 k가 짝수인 경우에 사용할 수 있는 방법이며, 대부분의 현실적인 경우에 전통적인 여러 방법들보다 효율적이다.

2. 새로운 모평균 추정 방법

크기 N인 모집단을 구성하는 N개의 단위들을 $U_1, U_2, \cdots, U_N$으로 나타내기로 하자. 이 모

1) Professor,Division of Mathematics and Informational Statistics, Wonkwang University, Iksan, Chonbuk, 570-749, Korea.
E-mail : hjkim@wonkwang.ac.kr
2) Professor, Division of Mathematics and Statistical Informatics, Chonbuk National University, Chonju, Chonbuk, 561-756, Korea.
E-mail : jbcbc@stat.chonbuk.ac.kr

집단으로부터 크기 n 인 표본을 뽑으려 한다. $N=kn$ 이라 하자.

집락 $S'_i\,(i=1,2,\cdots,k)$ 를 다음과 같이 정의하자.

n 이 짝수일 때

$$S'_i=\{\, U_{i+(j-1)k}\ : j=1,2,\cdots,n/2\,\} \cup \{\, U_{N+1-i-(j-1)k} : j=1,2,\cdots,n/2\,\}$$

n 이 홀수일 때

$$S'_i=\{\, U_{i+(j-1)k} : j=1,2,\cdots,(n+1)/2\,\} \cup \{\, U_{N+1-i-(j-1)k} : j=1,2,\cdots,(n-1)/2\,\}$$

예를들어 N=20, n=5, k=4이면 $S'_1=\{\,U_1,U_5,U_9,U_{16},U_{20}\,\}$, $S'_2=\{\,U_2,U_6,U_{10},U_{15},U_{19}\,\}$, $S'_3=\{\,U_3,U_7,U_{11},U_{14},U_{18}\,\}$, $S'_4=\{\,U_4,U_8,U_{12},U_{13},U_{17}\,\}$ 이다.

이제 Kim(1985)에 의하여 제시된 중심변형계통추출(centered modified systematic sampling: CMSS) 방법을 살펴보자. 이 방법은 Singh 등(1968)의 변형계통추출(modified systematic sampling: MSS)과 Madow(1953)의 중심계통추출(centered systematic sampling: CSS)을 결합한 것이다. k가 홀수이면 $S'_{(k+1)/2}$를 뽑는다. 따라서 이 경우 CMSS는 CSS와 같은 방법이다. k가 짝수(앞으로는 이 경우만 생각하기로 하자)이면 각각 1/2의 확률로 두 집락 $S'_{k/2}$와 $S'_{k/2+1}$ 중 하나를 뽑는다. 이 방법에 의한 모평균 $\overline{Y}$의 추정량 $\overline{y}_{cm}$ 은 다음과 같은 평균제곱오차를 갖는다.

$$MSE(\overline{y}_{cm})=\frac{1}{2}\{(\overline{y}'_{k/2}-\overline{Y})^2+(\overline{y}'_{k/2+1}-\overline{Y})^2\} \tag{2.1}$$

단 $\overline{y}'_i$는 S'_i 안의 단위들의 특성값의 평균이다($i=1,2,\cdots,k$).

이제 본 논문에서 사용될 기호들을 정의한다.

y_i : 모집단 안의 i 번째 단위 U_i 가 가지고 있는 특성값 ($i=1,2,\cdots,N$)

$\overline{Y}=\frac{1}{N}\sum_{i=1}^{N}y_i$: 추정하고자 하는 모집단 평균

y'_{ij} : S'_i안의 j 번째 단위의 특성값 ($i=1,2,\cdots,k$; $j=1,2,\cdots,n$)

즉, n 이 짝수일 때

$$y'_{ij}=y_{i+(j-1)k} \quad (j=1,2,\cdots,n/2)$$
$$y'_{ij}=y_{N+1-i-(n-j)k}=y_{1-i+jk} \quad (j=n/2+1,\ n/2+2,\ \cdots,n)$$

n 이 홀수일 때

$$y'_{ij} = y_{i+(j-1)k} \quad (j=1,2,\cdots,(n-1)/2,\ (n+1)/2)$$
$$y'_{ij} = y_{N+1-i-(n-j)k} = y_{1-i+jk} \quad (j=(n+3)/2,(n+5)/2,\cdots,n)$$

$$\bar{y}'_i = \frac{1}{n}\sum_{j=1}^{n} y'_{ij}:\ S'_i\text{안의 단위들의 특성값의 평균 } (i=1,2,\cdots,k)$$

이제 k가 짝수이고 n이 3 이상의 홀수인 경우 모평균 $\bar{Y}$를 추정하는 새로운 방법을 제시하고자 한다. 예를 들어 N=20, n=5, k=4 인 경우를 다시 생각해 보자. 우선 CMSS 를 써서 S'_2와 S'_3 중 하나의 집락을 뽑는다. $S'_2 = \{U_2, U_6, U_{10}, U_{15}, U_{19}\}$ 이고 $S'_3 = \{U_3, U_7, U_{11}, U_{14}, U_{18}\}$ 이므로, 집락 안의 단위들의 번호를 합하면 S'_2의 경우 52이고 S'_3의 경우는 53이다. 두 집락 사이에 약간의 차이가 있으므로, 단순한 표본평균 $\bar{y}'_2$나 $\bar{y}'_3$를 쓰는 것보다 수정된 추정량을 써서 $\bar{Y}$를 추정하는 것이 더욱 좋지 않을까 하는 추측을 할 수 있다. 2+6+15+19=3+7+14+18이고 10과 11에서 차이가 있으므로, S'_2가 뽑힌 경우에는 y_{10} 대신 $y_{10.5}$를 쓰고, S'_3가 뽑힌 경우에는 y_{11} 대신 $y_{10.5}$를 쓰면 균형이 이루어진다. 물론 여기서 $y_{10.5}$라는 값은 실제로는 존재하지 않는 가상적인 값이다.

S'_2의 경우에는 y_{10}과 y_{15}를 써서 $y_{10.5}$를 "추정"할 수 있다. 보간법(interpolation)을 쓰면 $y_{10.5}$는 $(1/10)(9y_{10}+y_{15})$로 추정된다. 따라서 y_{10}대신 이 값을 사용하여 $\bar{Y}$를

$$\begin{aligned}\bar{y}'^{*}_2 &= \frac{1}{5}\left\{y_2 + y_6 + \frac{1}{10}(9y_{10}+y_{15}) + y_{15} + y_{19}\right\} \\ &= \bar{y}'_2 + \frac{1}{50}(y_{15}-y_{10})\end{aligned} \tag{2.2}$$

로 추정한다. 2차원 첨자를 써서 나타내면, S'_2가 뽑힌 경우 $\bar{Y}$의 추정값을

$$\bar{y}'^{*}_2 = \bar{y}'_2 + \frac{1}{50}(y'_{24}-y'_{23}) \tag{2.3}$$

로 한다.

S'_3의 경우에는 y_7과 y_{11}에 보간법을 적용하여 $y_{10.5}$를 $(1/8)(y_7+7y_{11})$로 추정할 수 있으며, y_{11} 대신 이 값을 쓰면 $\bar{Y}$ 는

$$\begin{aligned} \bar{y}'^{*}_{3} &= \frac{1}{5}\left\{y_3 + y_7 + \frac{1}{8}(y_7 + 7y_{11}) + y_{14} + y_{18}\right\} \\ &= \bar{y}'_{3} - \frac{1}{40}(y_{11} - y_7) \end{aligned} \tag{2.4}$$

로 추정된다. 2차원 첨자를 써서 나타내면, S'_3가 뽑힌 경우 $\overline{Y}$의 추정값은

$$\bar{y}'^{*}_{3} = \bar{y}'_{3} - \frac{1}{40}(y'_{33} - y'_{32}) \tag{2.5}$$

가 된다.

이상의 내용을 일반화하면 다음과 같다. k가 짝수이고 n이 3 이상의 홀수인 경우 CMSS에 의하여 하나의 집락을 뽑는다.

즉, 각각 1/2의 확률로 두 집락 $S'_{k/2}$와 $S'_{k/2+1}$ 중 하나를 뽑는다. $\overline{Y}$ 의 추정값은 $S'_{k/2}$가 뽑힌 경우

$$\bar{y}'^{*}_{k/2} = \bar{y}'_{k/2} + \frac{1}{2n(k+1)}(y'_{k/2,(n+3)/2} - y'_{k/2,(n+1)/2}) \tag{2.6}$$

로 하며, $S'_{k/2+1}$이 뽑힌 경우에는

$$\bar{y}'^{*}_{k/2+1} = \bar{y}'_{k/2+1} - \frac{1}{2nk}(y'_{k/2+1,(n+1)/2} - y'_{k/2+1,(n-1)/2}) \tag{2.7}$$

로 한다.

이 방법을 CMI로 나타내고 CMI에 의한 $\overline{Y}$의 추정량을 $\bar{y}_{cmi}$로 표시하자. $\bar{y}_{cmi}$는 다음과 같은 편의와 평균제곱오차를 갖는다는 것을 알 수 있다.

$$B(\bar{y}_{cmi}) = \frac{1}{2}(\bar{y}'^{*}_{k/2} + \bar{y}'^{*}_{k/2+1}) - \overline{Y} \tag{2.8}$$

$$MSE(\bar{y}_{cmi}) = \frac{1}{2}\{(\bar{y}'^{*}_{k/2} - \overline{Y})^2 + (\bar{y}'^{*}_{k/2+1} - \overline{Y})^2\} \tag{2.9}$$

3. 무한초모집단 모형 하에서의 기대평균제곱오차

Cochran(1946)은 추세를 갖는 모집단에 관한 추론의 효율성을 논하는 경우 훌륭한 이론적 근거가 되는 무한초모집단 모형(infinite superpopulation model)을 제시하였다. 이것의 기본 개념은 주어진 유한모집단을 무한초모집단으로부터 뽑힌 하나의 표본으로 간주하는 것으로서 다음과 같은 모형에 의해 표시된다.

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

여기서 μ_i는 i의 함수이며, e_i는 랜덤오차항으로서 $E(e_i) = 0$, $E(e_i^2) = \sigma^2$, $E(e_i e_j) = 0$ ($i \neq j$일 때)이다. E는 무한초모집단에 걸친 기대값을 나타낸다.

이제부터 μ에 관해서도 y에 관해서 정의된 것과 같은 양식의 기호를 사용하기로 한다. 예컨대

$$\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mu_i$$

$$\mu'_{ij} = \mu_{i+(j-1)k} \quad (j = 1, 2, \cdots, (n+1)/2)\,(n: \text{홀수})$$

$$\bar{\mu}'_i = \frac{1}{n}\sum_{j=1}^{n}\mu'_{ij}$$

$$\bar{\mu}'^{*}_{k/2} = \bar{\mu}'_{k/2} + \frac{1}{2n(k+1)}(\mu'_{k/2,(n+3)/2} - \mu'_{k/2,(n+1)/2})$$

$$\bar{\mu}'^{*}_{k/2+1} = \bar{\mu}'_{k/2+1} - \frac{1}{2nk}(\mu'_{k/2+1,(n+1)/2} - \mu'_{k/2+1,(n-1)/2})$$

등이다.이러한 기호들과 위의 가정들, 그리고 식 (2.9)를 이용하면 다음과 같은 정리를 얻을 수 있다. 증명은 김혁주와 석은양(2000)에 있는 정리5.1의 증명과 유사한 방법에 의해 할 수 있으므로 생략하기로 한다.

정리 1. 모형 (3.1)을 가정할 때, $\bar{y}_{cmi}$의 평균제곱오차의 기대값은 다음과 같다. 단, 여기서 $A = \sigma^2(k-1)/N$ (이하 같음)이며, k는 짝수, n은 3 이상의 홀수이다.

$$EMSE(\bar{y}_{cmi}) = \frac{1}{2}\{(\bar{\mu}'^{*}_{k/2} - \bar{\mu})^2 + (\bar{\mu}'^{*}_{k/2+1} - \bar{\mu})^2\} + A + \frac{\sigma^2}{4n^2}\left\{\frac{1}{k^2} + \frac{1}{(k+1)^2}\right\} \tag{3.2}$$

이제 모집단에 선형추세가 존재하는 경우를 생각해 보자.
선형추세는 $\mu_i = a + bi$(a와 b는 상수, $b \neq 0$)에 의해 표시된다. 즉 다음과 같은 무한초모집단 모형을 가정한다.

$$y_i = a + bi + e_i \qquad (i = 1, 2, \cdots, N) \tag{3.3}$$

이 경우 정리1을 이용하면 다음과 같은 정리를 얻게 된다.

정리 2. 모집단이 식 (3.3)과 같은 선형 추세를 가질 때, $\bar{y}_{cmi}$ 의 기대평균제곱오차는 다음과 같다.

$$EMSE(\bar{y}_{cmi}) = A + \frac{\sigma^2}{4n^2}\left\{\frac{1}{k^2} + \frac{1}{(k+1)^2}\right\} \ (k: \text{짝수},\ n: 3 \text{ 이상의 홀수}) \tag{3.4}$$

4. 무한초모집단 모형 하에서의 효율성 비교

이 절에서는 선형 추세를 갖는 모집단의 평균을 추정하는 경우 본 논문에서 얻어진 추정량인 $\bar{y}_{cmi}$ 의 효율성을 다음과 같은 기존의 여러 방법들에 의한 추정량들과 비교해 보기로 한다.

(1) 보통의 계통추출(ordinary systematic sampling)

$$EMSE(\bar{y}_{sy}) = \frac{b^2(k^2-1)}{12} + A \tag{4.1}$$

(2) 끝값수정법(end corrections) (Yates, 1948)

$$EMSE(\bar{y}_{ec}) = A + \frac{\sigma^2(k^2-1)}{6k^2(n-1)^2} \tag{4.2}$$

(3) 변형계통추출(modified systematic sampling) (Singh et al., 1968)

$$EMSE(\bar{y}_{mod}) = \begin{cases} A & (n: \text{짝수}) \\ \dfrac{b^2(k^2-1)}{12n^2} + A & (n: \text{홀수}) \end{cases} \tag{4.3}$$

(4) 균형계통추출(balanced systematic sampling) (Sethi, 1965; Murthy, 1967)

$$EMSE(\bar{y}_{bal}) = \begin{cases} A & (n: \text{짝수}) \\ \dfrac{b^2(k^2-1)}{12n^2} + A & (n: \text{홀수}) \end{cases} \tag{4.4}$$

(5) 중심계통추출(centered systematic sampling) (Madow, 1953)

$$EMSE(\bar{y}_{cen}) = \begin{cases} \dfrac{b^2}{4} + A & (k\text{: 짝수}) \\ A & (k\text{: 홀수}) \end{cases} \tag{4.5}$$

(6) 중심변형계통추출(centered modified systematic sampling) (Kim, 1985)

$$EMSE(\bar{y}_{cm}) = \begin{cases} A & (k\text{: 짝수, } n\text{: 짝수}) \\ \dfrac{b^2}{4n^2} + A & (k\text{: 짝수, } n\text{: 홀수}) \end{cases} \tag{4.6}$$

(7) 중심균형계통추출(centered balanced systematic sampling) (Kim, 1985)

$$EMSE(\bar{y}_{cb}) = EMSE(\bar{y}_{cm}) \tag{4.7}$$

(8) 중심균형계통추출과 보간법을 이용한 방법(centered balanced systematic sampling with interpolation: CBI) (김혁주와 석은양, 2000)

$$EMSE(\bar{y}_{cbi}) = A + \frac{\sigma^2}{2n^2(k+1)^2} \quad (k\text{: 짝수, } n\text{: 5 이상의 홀수}) \tag{4.8}$$

(9) 중심변형추출(centered modified sampling) (Fountain과 Pathak, 1989)

$$EMSE(\bar{y}_{cmFP}) = \begin{cases} A & (n\text{: 짝수}) \\ A & (k\text{: 홀수, } n\text{: 홀수}) \\ \dfrac{b^2}{4n^2} + A & (k\text{: 짝수, } n\text{: 홀수}) \end{cases} \tag{4.9}$$

(10) 중심균형추출(centered balanced sampling) (Fountain과 Pathak, 1989)

$$EMSE(\bar{y}_{cbFP}) = EMSE(\bar{y}_{cmFP}) \tag{4.10}$$

(11) 양끝추출(two-end sampling) (Fountain과 Pathak, 1989)

$$EMSE(\bar{y}_{tes}) = EMSE(\bar{y}_{cmFP}) = EMSE(\bar{y}_{cbFP}) \tag{4.11}$$

여러 방법들의 효율성의 비교는 식(3.4)와 식(4.1)부터 식(4.11)까지를 사용하여 이루어질 수 있다. 이 절에서 고려하는 경우는 모두 k가 짝수이고 표본크기 n이 3 이상의 홀수인 경우이다.

먼저 본 논문에서 제시된 방법인 CMI와 보통의 계통추출법을 비교해 보자. (3.4)와 (4.1)로부

터,

$$EMSE(\bar{y}_{cmi}) < EMSE(\bar{y}_{sy}) \tag{4.12}$$

일 필요충분조건은

$$\sigma^2 < \frac{b^2 n^2 (k^2-1)}{3\{1/k^2+1/(k+1)^2\}} \tag{4.13}$$

임을 얻을 수 있다. 이것은 모형 (3.1)에서의 오차항의 분산 σ^2이 터무니없이 크지만 않으면 $\bar{y}_{cmi}$가 $\bar{y}_{sy}$보다 효율적이라는 것을 의미한다(예 참조).

다음으로 CMI와 Yates(1948)의 끝값수정법을 비교해 보자. (3.4)와 (4.2)로부터

$$\begin{aligned} &EMSE(\bar{y}_{ec}) - EMSE(\bar{y}_{cmi}) \\ &= \frac{\sigma^2}{12k^2(k+1)^2 n^2 (n-1)^2} \{ (2k^4+4k^3-6k^2-10k-5)n^2 + 3(2k^2+2k+1)(2n-1) \} \end{aligned} \tag{4.14}$$

을 얻게 되는데, 이 값은 k가 짝수이고 n이 3 이상의 홀수이면 항상 0보다 큰 값이다. 따라서 기대평균제곱오차를 기준으로 할 때 $\bar{y}_{cmi}$는 항상 $\bar{y}_{ec}$보다 효율적이라는 사실을 알 수 있다. 선형 추세를 갖는 모집단의 경우 끝값수정법이 아주 좋은 방법으로 인정받아 왔으므로, CMI가 이보다 더 효율적이라는 사실은 좋은 결과로 볼 수 있을 것이다.

이번에는 CMI를 김혁주와 석은양(2000)의 CBI와 비교해 보자. (3.4)와 (4.8)로부터

$$EMSE(\bar{y}_{cbi}) - EMSE(\bar{y}_{cmi}) = \frac{\sigma^2}{4n^2}\left\{\frac{1}{(k+1)^2} - \frac{1}{k^2}\right\} \tag{4.15}$$

이 된다. 이 값은 엄밀히 부호를 따지면 음수이긴 하나 대부분의 현실적인 경우에 0과 거의 같은 값이며(예를 들면 (a) $k=4$, $n=7$인 경우 $-1.15\times10^{-4}\sigma^2$, (b) $k=10$, $n=25$인 경우 $-6.94\times10^{-7}\sigma^2$, (c) $k=20$, $n=35$인 경우 $-4.74\times10^{-8}\sigma^2$), 더욱이 EMSE $(\bar{y}_{cmi})$와 EMSE $(\bar{y}_{cbi})$에 공통으로 포함되어 있는 항인 $A=\sigma^2(k-1)/N$의 값에 대한 비율을 생각해 보면 거의 완전히 무시할 수 있는 수준이다(절대값의 비율이 (a) $k=4$, $n=7$인 경우 1.07×10^{-3}, (b) $k=10$, $n=25$인 경우 1.93×10^{-5}, (c) $k=20$, $n=35$인 경우 1.75×10^{-6}). 따라서 CMI는 CBI와 거의 같은 정도의 효율성을 가지고 있다고 볼 수 있다.

이러한 방식으로 모든 경우에 대하여 CMI를 여러 방법들과 비교한 결과를 다음과 같이 정리

할 수 있다. 표현을 간결하게 하기 위하여 $EMSE(\bar{y}_{sy})$, $EMSE(\bar{y}_{cmi})$ 등을 각각 sy, cmi 등으로 나타내기로 한다. 따라서, 예컨대 " $cmi < sy$ " 라는 표현은 $\bar{y}_{cmi}$가 $\bar{y}_{sy}$ 보다 효율적이라는 것을 의미한다. 그리고 k가 짝수이고 n이 홀수인 모든 경우에 $cm = cb = cmFP = cbFP = tes$ 이므로 대표로 cm 한 가지만 표시하겠다. 여기서 $A_k = 1/\{1/k^2 + 1/(k+1)^2\}$ 이다.

(1) $k=2$ 이고 n 이 3 이상의 홀수인 경우

(i) $\sigma^2 < 36b^2/13$ 이면 $cmi < cm = mod = bal < cen = sy$

(ii) $36b^2/13 \le \sigma^2 < 36b^2n^2/13$ 이면 $cm = mod = bal \le cmi < cen = sy$

(iii) $36b^2n^2/13 \le \sigma^2$ 이면 $cm = mod = bal < cen = sy \le cmi$

(2) k 가 4 이상의 짝수, n 이 3 이상의 홀수이고 $n < \sqrt{(k^2-1)/3}$ 인 경우

(i) $\sigma^2 < b^2A_k$ 이면 $cmi < cm < cen < mod = bal < sy$

(ii) $b^2A_k \le \sigma^2 < b^2n^2A_k$ 이면 $cm \le cmi < cen < mod = bal < sy$

(iii) $b^2n^2A_k \le \sigma^2 < b^2(k^2-1)A_k/3$ 이면 $cm < cen \le cmi < mod = bal < sy$

(iv) $b^2(k^2-1)A_k/3 \le \sigma^2 < b^2n^2(k^2-1)A_k/3$ 이면 $cm < cen < mod = bal \le cmi < sy$

(v) $b^2n^2(k^2-1)A_k/3 \le \sigma^2$ 이면 $cm < cen < mod = bal < sy \le cmi$

(3) k 가 4 이상의 짝수, n 이 3 이상의 홀수이고 $n = \sqrt{(k^2-1)/3}$ 인 경우
(예를 들면 $k=26$, $n=15$)

(i) $\sigma^2 < b^2A_k$ 이면 $cmi < cm < cen = mod = bal < sy$

(ii) $b^2A_k \le \sigma^2 < b^2n^2A_k$ 이면 $cm \le cmi < cen = mod = bal < sy$

(iii) $b^2n^2A_k \le \sigma^2 < b^2n^2(k^2-1)A_k/3$ 이면 $cm < cen = mod = bal \le cmi < sy$

(iv) $b^2n^2(k^2-1)A_k/3 \le \sigma^2$ 이면 $cm < cen = mod = bal < sy \le cmi$

(4) k 가 4 이상의 짝수 , n 이 3 이상의 홀수이고, $n > \sqrt{(k^2-1)/3}$ 인 경우

(i) $\sigma^2 < b^2 A_k$ 이면 $cmi < cm < mod = bal < cen < sy$

(ii) $b^2 A_k \le \sigma^2 < b^2 (k^2-1) A_k/3$ 이면 $cm \le cmi < mod = bal < cen < sy$

(iii) $b^2 (k^2-1) A_k/3 \le \sigma^2 < b^2 n^2 A_k$ 이면 $cm < mod = bal \le cmi < cen < sy$

(iv) $b^2 n^2 A_k \le \sigma^2 < b^2 n^2 (k^2-1) A_k/3$ 이면 $cm < mod = bal < cen \le cmi < sy$

(v) $b^2 n^2 (k^2-1) A_k/3 \le \sigma^2$ 이면 $cm < mod = bal < cen < sy \le cmi$

예. $N=150$개의 단위들로 이루어진 모집단으로부터 크기 $n=15$인 표본을 뽑는 경우를 생각하자. 이 때 $k=10$이 된다. 선형 추세의 기울기는 $b=0.5$라고 하자. 이 경우 여러 방법들의 효율성을 비교해 보면 다음과 같다.

(i) $\sigma^2 < 13.688$ 이면 $cmi < cm < mod = bal < cen < sy$

(ii) $13.688 \le \sigma^2 < 451.697$ 이면 $cm \le cmi < mod = bal < cen < sy$

(iii) $451.697 \le \sigma^2 < 3079.751$ 이면 $cm < mod = bal \le cmi < cen < sy$

(iv) $3079.751 \le \sigma^2 < 101631.787$ 이면 $cm < mod = bal < cen \le cmi < sy$

(v) $101631.787 \le \sigma^2$ 이면 $cm < mod = bal < cen < sy \le cmi$

여기서 볼 수 있듯이, 무한초모집단 모형에서의 오차항의 분산 σ^2이 작을수록 CMI는 다른 방법들에 비해 효율적인 방법이 된다. (iii),(iv),(v)의 경우처럼 σ^2이 커지면 CMI는 상대적으로 비효율적인 방법이 되어간다. 그러나 σ^2이 매우 크다는 것은 선형 추세의 의미가 거의 없어진다는 것을 의미하므로 이러한 경우는 논의할 필요가 거의 없다고 할 수 있다.

참고문헌

[1] 김혁주, 석은양(2000), 선형추세를 갖는 모집단에 대한 효율적인 모평균 추정: 계통추출의 확장, 「응용통계연구」, 제13권 제2호, 457-476.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

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

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

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

[2001년 11월 접수, 2002년 1월 채택]