Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.2 pp 365 379

On the Effectiveness of Centering, Interpolation and Extrapolation in Estimating the Mean of a Population with Linear Trend

Hyuk Joo Kim[1] and Sun-Ju Jung[2]

Abstract

We apply the techniques of interpolation and extrapolation to derive a new estimator based on centered modified systematic sampling for the mean of a population which has a linear trend. The efficiency of the proposed estimation method is compared with that of various existing methods. An illustrative numerical example is given.

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

1. 서 론

표본조사에서 중요한 문제 중의 하나는 모수의 참값을 좋은 정도(precision)로 추정하는 문제이다. 표본조사를 실시할 때, 관심의 대상인 모집단의 단위들이 특성값에 무관한 순서로 배열되어 있지 않고 어떤 추세(trend)를 가지고 있는 경우가 종종 있다.

추세를 갖는 모집단에 대한 표본추출 방법 중 많이 응용되는 것이 계통추출(systematic sampling)이다. 계통추출은 모집단이 선형 추세(linear trend)를 갖는 경우에 특히 효율적이라는 것이 알려져 있다. 계통추출에는 우리가 흔히 얘기하는 보통의 계통추출(ordinary systematic sampling: OSS) 외에도 Madow(1953)의 중심계통추출

1. Professor, Division of Mathematics and Informational Statistics & Institute of Basic Natural Science, Wonkwang University, Iksan, Jeonbuk, 570-749, Korea.
E-mail : hjkim@wonkwang.ac.kr
2. Graduate Student, Major of Mathematics Education, Graduate School of Education, Wonkwang University, Iksan, Jeonbuk, 570-749, Korea.

(centered systematic sampling: CSS), Sethi(1965)와 Murthy(1967)의 균형계통추출(balanced systematic sampling: BSS), Singh등(1968)의 변형계통추출(modified systematic sampling: MSS) 등이 포함된다.

Kim(1985)은 추출률의 역수 k가 짝수인 경우에 대해, CSS와 MSS를 결합한 중심변형계통추출(centered modified systematic sampling: CMSS)과, CSS와 BSS를 결합한 중심균형계통추출(centered balanced systematic sampling: CBSS)을 제안하였다. 이 CMSS와 CBSS는 각각 MSS와 BSS를 중심화한 것으로서, 선형 추세를 갖는 모집단의 평균을 추정하는 경우 표본크기 n이 홀수이면 MSS와 BSS보다 효율적인 방법이라는 것이 밝혀졌다. 더욱이 이 경우 OSS를 중심화한 CSS는 OSS보다 효율적이라는 사실이 이미 알려져 있으므로, 모집단에 선형 추세가 존재하는 경우 '중심화(centering)'라는 개념은 모평균 추정에 아주 유용한 도구임을 알 수 있다.

한편 n이 3이상의 홀수이고, k가 짝수인 경우 MSS에 보간법의 내삽법(interpolation)과 외삽법 (extrapolation)을 도입하여 모평균을 추정하는 방법(MIE)이 Kim과 Oh(2002)에 의해 연구되었다.
MIE는 많은 경우 MSS를 포함한 전통적인 방법들보다 효율적인 것으로 나타났다.

본 논문에서는 MIE에 중심화의 개념을 적용하여 선형 추세를 갖는 모집단의 평균을 추정하는 문제를 고찰하고자 한다. 따라서 제시될 방법은 CMSS와 내삽법 및 외삽법을 결합한 방법이며, MIE를 개량하여 더욱 효율적인 방법을 이끌어내고자 하는 의도를 가지고 자연스럽게 생각할 수 있는 방법이다.

2. 새로운 모평균 추정 방법

모집단의 크기를 N이라 하고 모집단은 N 개의 단위 $U_1, U_2, \cdots, U_N$ 으로 이루어져 있다고 하자. 이 모집단으로부터 크기 n인 표본을 뽑아, 추출된 표본에서 얻은 정보를 이용하여 모집단의 평균을 추정하려 한다. N이 n의 배수인 경우를 생각하고 $k = N/n$이라 하자.

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

n 이 짝수일 때

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

n 이 홀수일 때

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

예를 들어 N=28, n=7, k=4이면 $S'_1 = \{ U_1, U_5, U_9, U_{13}, U_{20}, U_{24}, U_{28}, \}$, $S'_2 = \{ U_2, U_6,$
$U_{10}, U_{14}, U_{19}, U_{23}, U_{27} \}$, $S'_3 = \{ U_3, U_7, U_{11}, U_{15}, U_{18}, U_{22}, U_{26} \}$,

$S'_4 = \{ U_4, U_8, U_{12}, U_{16}\ U_{17}, U_{21},$

$U_{25}\}$ 이다.

이제 CMSS방법을 살펴보자. 1절에서 언급한 것과 같이 이 방법은 Kim (1985)이 제시한 것으로서 MSS와 CSS를 결합한 것, 즉 MSS를 중심화한 것이다. k가 홀수이면 $S'_{(k+1)/2}$를 뽑는다. 이 경우 CMSS는 CSS와 정확히 일치하므로 더 이상 논의하는 것은 무의미하다. 따라서 앞으로는 k가 짝수인 경우만 생각하기로 하자. k가 짝수이면 각각 1/2의 확률로 두 집락 $S'_{k/2}$와 $S'_{k/2+1}$ 중 하나를 뽑는다. 이 방법에 의한 모평균 $\overline{Y}$의 추정량을 $\overline{y}_{cm}$ 이라 하면 $\overline{y}_{cm}$ 의 평균제곱오차는 다음과 같다.

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

단 $\overline{y}'_i$는 S'_i안의 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

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

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

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

이제 k가 짝수이고 n이 3 이상의 홀수인 경우 모평균 $\overline{Y}$를 추정하는 새로운 방법을 제시하고자 한다. N=28, n=7, k=4 인 경우를 다시 예로 들어 설명하기로 한다. 우선 CMSS를 써서 S'_2와 S'_3 중 하나의 집락을 뽑는다. S'_2와 S'_3 안의 단위들의 번호를 합하면 각각 101과 102이다. 이와 같이 두 집락 사이에는 약간의 차이가 존재하는데, 이 차이는 표본크기 n이 홀수이기 때문에 생기는 것이다. 본 논문의 아이디어

는 n이 홀수인 점을 감안하여, 선형 추세가 존재하는 모집단에 대하여 가상의 단위를 설정함으로써 추출가능한 두 집락의 단위들의 번호의 합을 같게 만드는 것이다. S_2'가 뽑힌 경우에는 y_2대신 가상의 값 $y_{2.5}$를 쓰고, S_3'가 뽑힌 경우에는 y_3대신 $y_{2.5}$를 사용하여 두 집락의 균형을 맞춘다.

S_2'의 경우에는 y_2와 y_6를 써서 $y_{2.5}$를 추정할 수 있다. 보간법의 내삽법을 쓰면 $y_{2.5}$를 $(1/8)(7y_2+y_6)$로 추정할 수 있다. 따라서 y_2대신 이 값을 사용하여 $\overline{Y}$를

$$\begin{aligned}\overline{y}'^{*}_{2} &= \frac{1}{7}\left\{\frac{1}{8}(7y_2+y_6)+y_6+y_{10}+y_{14}+y_{19}+y_{23}+y_{27}\right\} \\ &= \overline{y}'_2+\frac{1}{56}(y_6-y_2)\end{aligned} \tag{2.2}$$

로 추정한다. 2차원 첨자를 써서 나타내면, S_2'가 뽑힌 경우 $\overline{Y}$의 추정값을

$$\overline{y}'^{*}_{2}=\overline{y}'_2+\frac{1}{56}(y'_{22}-y'_{21}) \tag{2.3}$$

로 한다.

S_3'의 경우에는 y_3와 y_7에 외삽법을 적용한다. 이 경우 $y_{2.5}$는 $(1/8)(9y_3-y_7)$로 추정할 수 있으며, y_3 대신 이 값을 쓰면 $\overline{Y}$ 는

$$\begin{aligned}\overline{y}'^{*}_{3} &= \frac{1}{7}\left\{\frac{1}{8}(9y_3-y_7)+y_7+y_{11}+y_{15}+y_{18}+y_{22}+y_{26}\right\} \\ &= \overline{y}'_3-\frac{1}{56}(y_7-y_3) \\ &= \overline{y}'_3-\frac{1}{56}(y'_{32}-y'_{31})\end{aligned} \tag{2.4}$$

로 추정된다.

이상의 내용은 $N=28$, $n=7$, $k=4$인 경우에 대한 것이므로 k가 짝수이고 n이 3 이상의 홀수인 경우로 일반화될 필요가 있으며 그 내용은 다음과 같다. 우선 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

$$\overline{y}'^{*}_{k/2}=\overline{y}'_{k/2}+\frac{1}{2nk}(y'_{k/2,2}-y'_{k/2,1}) \tag{2.5}$$

로 하며, $S'_{k/2+1}$이 뽑힌 경우에는

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

로 한다.

이 방법을 CMIE('centered modified systematic sampling'과 'interpolation'과 'extrapolation'을 나타낸 것)로 나타내고 CMIE에 의한 $\overline{Y}$의 추정량을 $\overline{y}_{cmie}$로 표시하자. 제 1절에서 언급했듯이 CMIE는 Kim과 Oh(2002)의 MIE를 중심화한 것이다. $\overline{y}_{cmie}$는 다음과 같은 평균제곱오차를 갖는다는 것을 알 수 있다.

$$MSE(\overline{y}_{cmie})=\frac{1}{2}\left\{(\overline{y}'^{*}_{k/2}-\overline{Y})^2+(\overline{y}'^{*}_{k/2+1}-\overline{Y})^2\right\} \tag{2.7}$$

3. 무한 초모집단 모형 하에서의 기대평균제곱오차

이 절에서는 Cochran(1946)이 제시한 무한 초모집단 모형(infinite superpopulation model)을 이용하여 $\overline{y}_{cmie}$의 평균제곱오차의 기대값을 구한다. 무한 초모집단 모형이란, 주어진 유한모집단을 무한 초모집단에서 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 $E(e_i)=0$, $E(e_i^2)=\sigma^2$, $E(e_ie_j)=0$ ($i\neq j$일 때)이다. E는 무한 초모집단에 걸친 기대값을 나타낸다.

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

$$\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mu_i$$
$$\mu'_{ij} = \mu_{i+(j-1)k} \quad (j = 1, 2, \cdots, (n+1)/2)\ (n : \text{홀수})$$
$$\bar{\mu}'_i = \frac{1}{n}\sum_{j=1}^{n}\mu'_{ij}$$
$$\bar{\mu}'^{*}_{k/2} = \bar{\mu}'_{k/2} + \frac{1}{2nk}(\mu'_{k/2,2} - \mu'_{k/2,1})$$
$$\bar{\mu}'^{*}_{k/2+1} = \bar{\mu}'_{k/2+1} - \frac{1}{2nk}(\mu'_{k/2+1,2} - \mu'_{k/2+1,1})$$

등이다.

이러한 기호들과 위의 가정들, 그리고 식 (2.7)을 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, $\bar{y}_{cmie}$ 의 평균제곱오차의 기대값은 다음과 같다. 단, 여기서 $A = \sigma^2(1/n - 1/N)$ (이하 같음)이며, k 는 짝수, n 은 3 이상의 홀수이다.

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

이제 $\mu_i = a + bi$ (a 와 b는 상수, $b \neq 0$)인 경우, 즉 가정된 무한 초모집단 모형이

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

인 경우를 생각해 보자, 이 경우가 바로 모집단에 선형 추세가 존재하는 경우이다. 이 경우 $\overline{\mu} = a + (b/2)(N+1)$이며 또한 $\overline{\mu_i}'^{*} = a + (b/2)(N+1)\ (i = 1, 2, \cdots, k)$ 이라는 것을 유도할 수 있으므로 정리1에 의하여 다음과 같은 정리를 얻게 된다.

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

$$EMSE(\overline{y}_{cmie}) = A + \frac{\sigma^2}{2n^2k^2} \quad (k: \text{짝수},\ n: 3\ \text{이상의 홀수}) \tag{3.4}$$

4. 기존의 방법들과의 효율성 비교

이제 선형 추세를 갖는 모집단의 평균을 추정하는 경우 $\overline{y}_{cmie}$ 의 효율성을 기존의 여러 방법들에 의한 추정량들과 비교해 보자. 모평균 $\overline{Y}$를 추정하는 방법에는 크게 나눠 두 종류가 있다. 첫째는 뽑힌 표본의 단순평균으로 $\overline{Y}$를 추정하는 방법이며, 둘째는 표본의 단순평균이 아니라 가중평균으로 $\overline{Y}$를 추정하는 방법이다.

4.1 표본의 단순평균으로 모평균을 추정하는 방법들과의 비교

표본의 단순평균으로 $\overline{Y}$를 추정하는 방법에는 단순랜덤추출(simple random sampling: SRS), OSS, CSS, BSS, MSS 등의 전통적인 방법들과 Kim (1985)의 CMSS와 CBSS, 그리고 Fountain과 Pathak (1989)의 중심변형추출(centered modified sampling: CMS)과 중심균형추출(centered balanced sampling: CBS) 및 양끝추출(two-end sampling: TES) 등이 있다. k가 짝수이고 n이 홀수인 경우 이 방법들에 의한 추정량들의 기대평균제곱오차의 공통적인 형태는 다음 식과 같다.

$$EMSE(\widehat{\overline{Y}}) = A + b^2 f(n, k) \tag{4.1}$$

여기서 $f(n, k)$는 n과 k의 함수로서 각 방법에 대한 구체적인 내용은 다음과 같다.

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

이 방법들의 경우 선형 추세의 기울기 b의 값이 기대평균제곱오차에 중요한 영향을 미친다는 것을 식 (4.1)에서 알 수 있다.

CMIE와 이 방법들의 효율성은 식(3.4)와 식(4.1), (4.2)에 의하여 비교할 수 있다. $EMSE(\overline{y}_{cmie})$가 $EMSE(\widehat{Y})$보다 작을 필요충분조건은

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

임을 알 수 있다. 이것은 모형 (3.3)에서의 오차항의 분산 σ^2이 작을수록 즉 선형 추세가 강할수록 CMIE가 이 방법들보다 효율적이라는 것을 의미한다. 예를 들어 $N=300$, $n=25$, $k=12$이고 선형 추세의 기울기 b의 값은 0.7인 경우를 생각해 보자. 이 경우 CMIE가 SRS보다 효율적일 조건은 $\sigma^2 < 24,335,850$이고 OSS보다 효율적일 조건은 $\sigma^2 < 1,051,050$이며, CSS보다 효율적일 조건은 $\sigma^2 < 22,050$, MSS와 BSS보다 효율적일 조건은 $\sigma^2 < 1,681.68$, CMSS, CBSS, CMS, CBS, TES보다 효율적일 조건은 $\sigma^2 < 35.28$이다.

4.2 표본의 가중평균으로 모평균을 추정하는 방법들과의 비교

표본의 가중평균으로 $\overline{Y}$를 추정하는 방법으로는 다음 방법들이 있다. 수식은 각 방법에 의한 $\overline{Y}$의 추정량의 기대평균제곱오차이다. 이 방법들의 공통적인 특징은 기대평균제곱오차가 선형 추세의 기울기 b의 값에 무관하다는 것이다.

(1) 끝값수정법(end corrections: EC) (Yates, 1948)

$$EMSE(\overline{y}_{ec}) = A + \frac{\sigma^2(k^2-1)}{6k^2(n-1)^2} \tag{4.4}$$

(2) MSS와 내삽법을 이용하는 방법(MI) (Kim, 1998)

$$EMSE(\bar{y}_{mi}) = A + \frac{\sigma^2}{12n^2}(4 - 12A_k + 6kB_k - \frac{1}{k^2}) \tag{4.5}$$

(3) MSS와 내삽법 및 외삽법을 이용하는 방법(MIE) (Kim과 Oh, 2002)

$$EMSE(\bar{y}_{mie}) = A + \frac{\sigma^2(k^2 - 1)}{6n^2k^2} \tag{4.6}$$

(4) BSS와 내삽법을 이용하는 방법(BI) (Kim, 2000b)

$$EMSE(\bar{y}_{bi}) = A + \frac{\sigma^2}{2n^2}(1 - 4A_k + 2kB_k) \tag{4.7}$$

(5) BSS와 내삽법 및 외삽법을 이용하는 방법(BIE) (Kim, 1999)

$$EMSE(\bar{y}_{bie}) = A + \frac{\sigma^2}{2n^2}(1 - \gamma - 2\ln 2 + C_k) \tag{4.8}$$

(6) CMSS와 내삽법을 이용하는 방법(CMI) (Kim과 Choi, 2002)

$$EMSE(\bar{y}_{cmi}) = A + \frac{\sigma^2}{4n^2}\left\{\frac{1}{k^2} + \frac{1}{(k+1)^2}\right\} \tag{4.9}$$

(7) CBSS와 내삽법을 이용하는 방법(CBI) (김혁주와 석은양, 2000)

$$EMSE(\bar{y}_{cbi}) = A + \frac{\sigma^2}{2n^2(k+1)^2} \tag{4.10}$$

(8) CBSS와 내삽법 및 외삽법을 이용하는 방법(CBIE) (Kim, 2000a)

$$EMSE(\bar{y}_{cbie}) = A + \frac{\sigma^2}{4n^2}\left\{\frac{1}{(k-1)^2} + \frac{1}{(k+1)^2}\right\} \tag{4.11}$$

단, 식 (4.5), (4.7), (4.8)에서 나타난 기호들의 정의는 아래와 같다.

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

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

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

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

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

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

$$\gamma = 0.577215\cdots\cdots : \text{Euler}\text{상수}$$

<표1>, <표2>, <표3>은 각각 $k = 8, 12, 20$인 경우에 n의 몇 가지 값에 대하여 위의 여덟가지 방법과 CMIE에 의한 $EMSE(\cdot)/\sigma^2$ 의 값을 구한 것이다. 이 표들에서 볼 수 있듯이 CMI, CBI, CBIE, CMIE의 네 방법이 EC, MI, MIE, BI, BIE보다 효율적이다(k가 일정한 상태에서 n이 커지면 즉 표본추출률이 정해진 상태에서 모집단의 크기와 표본크기가 비례하여 증가하면 이 차이는 작아지지만 대소관계는 유지된다). 이로부터 우리는 선형 추세를 갖는 모집단의 평균을 추정할 때 중심화(centering)가 효율적인 추정량을 얻게 해 준다는 것을 알 수 있다. 그리고 표들에는 주어진 k와 n에 대하여 A/σ^2의 값이 구해져 있는데, 여기서 A는 정리 1에서 정의된 바와 같이 $\sigma^2(1/n - 1/N)$로서, 완전 추세무관 추출방법(completely trend-free sampling method)인, k가 홀수인 경우의 CSS와 n이 짝수인 경우의 MSS와 BSS가 갖는 기대평균제곱오차이다. 표들에서 볼 수 있듯이 CMI, CBI, CBIE, CMIE에 의한 기대평균제곱오차가 이 A의 값과 거의 차이가 없으므로 이 네 가지 방법들은 선형 추세를 갖는 모집단의 성질을 잘 이용한 최적에 가까운 추정 방법이라고 할 수 있을 것이다.

<표1> $k = 8$인 경우 $EMSE(\cdot)/\sigma^2$의 값들

n	EC	MI	MIE	BI	BIE	CMI	CBI	CBIE	CMIE	A/σ^2
5	.1853	.1794	.1816	.1773	.3065	.1753	.1752	.1753	.1753	.1750
25	.0353	.0352	.0353	.0351	.0403	.0350	.0350	.0350	.0350	.0350
55	.0160	.0159	.0160	.0159	.0170	.0159	.0159	.0159	.0159	.0519
105	.0083	.0083	.0083	.0083	.0086	.0083	.0083	.0083	.0083	.0083

<표2> $k=12$인 경우 $EMSE(\cdot)/\sigma^2$의 값들

n	EC	MI	MIE	BI	BIE	CMI	CBI	CBIE	CMIE	A/σ^2
5	.1936	.1877	.1899	.1855	.4054	.1835	.1835	.1835	.1835	.1833
25	.0370	.0369	.0370	.0368	.0456	.0367	.0367	.0367	.0367	.0367
55	.0168	.0167	.0168	.0167	.0185	.0167	.0167	.0167	.0167	.0167
105	.0087	.0087	.0087	.0087	.0092	.0087	.0087	.0087	.0087	.0087

<표3> $k=20$인 경우 $EMSE(\cdot)/\sigma^2$의 값들

n	EC	MI	MIE	BI	BIE	CMI	CBI	CBIE	CMIE	A/σ^2
5	.2004	.1945	.1967	.1923	.5993	.1900	.1900	.1901	.1901	.1900
25	.0383	.0382	.0383	.0381	.0544	.0380	.0380	.0380	.0380	.0380
55	.0174	.0173	.0174	.0173	.0207	.0173	.0173	.0173	.0173	.0173
105	.0090	.0090	.0090	.0090	.0100	.0090	.0090	.0090	.0090	.0090

4.3 모의실험을 이용한 예

모형

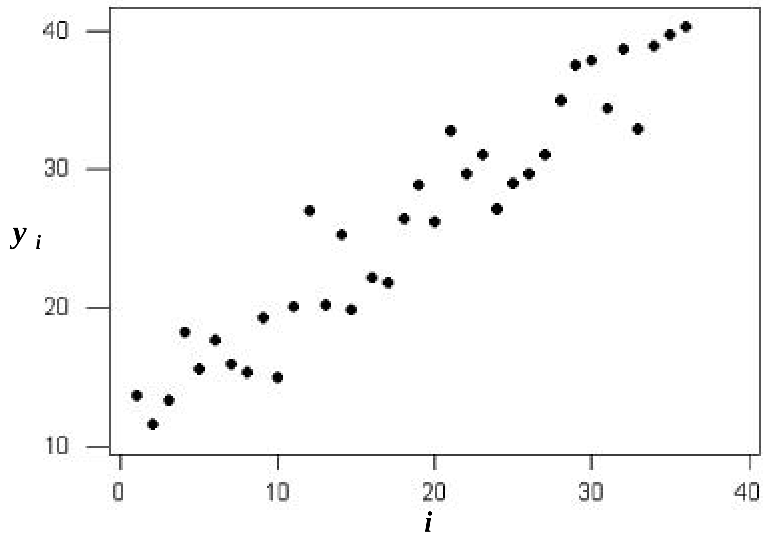
$$y_i = 12 + 0.8\,i + e_i \quad (i = 1, 2, \cdots, 36) \qquad (4.12)$$

를 설정하자(즉 $a=12$, $b=0.8$). 오차항 e_i의 값들을 실제로 발생시킴으로써 무한 초모집단으로부터 크기 $N=36$인 모집단을 생성한 다음, 다시 이 모집단으로부터 크

기 $n=9$ 인 표본을 뽑아 모평균을 추정하는 문제를 생각해 보기로 하자. $k=4$ 이며, 오차항 e_i 는 3절에서와 같은 조건을 만족시킨다. e_i 의 분산 σ^2 의 값은 $(2.5)^2$ 즉

6.25로 하였고, e_i 의 분포의 형태는 정규분포로 정하였으며, 미니탭(MINITAB)의 RANDOM 명령문을 이용하여 e_i 의 값들을 발생시켰다. 생성된 모집단은 다음과 같다.

13.6839	11.6511	13.3656	18.1942	15.5398	17.5994	15.8701
15.3858	19.2696	14.9848	20.1140	27.1042	20.2535	25.3537
19.7502	22.1680	21.8483	26.5246	28.8520	26.2472	32.8601
29.6664	31.0889	27.1769	28.9758	29.7617	31.1108	35.0991
37.6050	38.0048	34.4999	38.8084	32.9485	39.0079	39.8043
40.3434						



<그림1> 4.3절의 예에서 생성된 모집단

이 모집단의 평균은 $\overline{Y}=26.1256$ 이며, <그림1>에서 볼 수 있듯이 이 모집단은 대체적으로 증가하는 선형 추세를 가지고 있다. 4.1절과 4.2절에 열거된 기존의 열여덟 가지 방법에 의한 $\overline{Y}$ 의 추정량들의 평균제곱오차는 아래와 같다.

$MSE(\overline{y}_{srs}) = 6.2781$, $MSE(\overline{y}_{oss}) = 1.2089$, $MSE(\overline{y}_{mss}) = 0.1758$, $MSE(\overline{y}_{bss}) = 2.3746$

$MSE(\overline{y}_{css}) = 0.0438$, $MSE(\overline{y}_{cmss}) = 0.0423$, $MSE(\overline{y}_{cbss}) = 1.3592$,

$MSE(\overline{y}_{cms}) = 0.2443$

$MSE(\overline{y}_{cbs}) = 1.7606$, $MSE(\overline{y}_{tes}) = 0.0466$, $MSE(\overline{y}_{ec}) = 0.2274$, $MSE(\overline{y}_{mi}) = 0.1078$

$MSE(\overline{y}_{mie}) = 0.2006$, $MSE(\overline{y}_{bi}) = 2.0654$, $MSE(\overline{y}_{bie}) = 3.2915$, $MSE(\overline{y}_{cmi}) = 0.0478$

$MSE(\overline{y}_{cbi}) = 1.2648$, $MSE(\overline{y}_{cbie}) = 1.2323$

한편 본 논문에서 제시된 방법인 CMIE를 사용하면, 모평균 $\overline{Y}$를 다음 두 값 중 하나로 추정하게 된다(확률은 각각 1/2).

$$\overline{y}'^{*}_{2} = 25.9290$$

$$\overline{y}'^{*}_{3} = 26.0089$$

따라서 CMIE에 의한 $\overline{Y}$의 추정량 $\overline{y}_{cmie}$의 평균제곱오차는

$$MSE(\overline{y}_{cmie}) = 0.0261$$

이다. 이 값은 위의 열여덟 가지를 포함한 열아홉 가지의 방법에 의한 평균제곱오차 중 가장 작은 값이므로, CMIE가 열아홉 가지의 방법 중 가장 효율적이라는 것을 보여 준다.

5. 결 론

본 논문에서는, 선형 추세를 갖는 모집단에서 표본크기 n이 3이상의 홀수이고 추출률의 역수 k가 짝수인 경우 모평균 $\overline{Y}$에 대한 새로운 추정 방법을 제시하였다. 이 방법(CMIE)은 Kim(1985)이 제시한 CMSS(중심변형계통추출)를 써서 표본을 뽑은 뒤 내삽법과 외삽법을 사용하여 표본평균 $\overline{y}_{cmss}$보다 수정된 추정량인 $\overline{y}_{cmie}$를 써서 $\overline{Y}$를 추정하는 것이다.

Cochran(1946)의 무한 초모집단 모형에 근거를 둔 기대평균제곱오차를 기준으로 하여 CMIE와 기존의 방법들의 효율성을 비교하였다. 그 결과 CMIE는 Kim과 Oh(2002)의 MIE보다 효율적이라는 것이 밝혀졌으며, CMI, CBI, CBIE와 함께 최적에 가까운 추정 방법임을 알 수 있었다. 이는 선형 추세를 갖는 모집단의 평균을 추정하는 경우 중심화의 개념이 내삽법 및 외삽법과 함께 유용한 도구가 됨을 말해 준다.

참 고 문 헌

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. Kim, H. J. (1998), Estimation of population mean using interpolation in modified systematic sampling, *Korean Annals of Mathematics*, 15, 217-231.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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, to appear.
11. 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.
12. Murthy, M. N. (1967), *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, India.

13. Sethi, V. K. (1965), On optimum pairing of units, *Sankhya*, B27, 315-320.
14. Singh, D., Jindal, K. K. and Garg, J. N. (1968), On modified systematic sampling, *Biometrika*, 55, 541-546.
15. Yates, F. (1948), Systematic sampling, *Philosophical Transactions of the Royal Society of London*, A241, 345-377.

[2002년 9월 접수, 2002년 10월 채택]