The Korean Communications
in Statistics Vol. 5, No. 2, 1998
pp. 391-401

관리적 선정 하에서 추출방법의 비교

김 종 호[1], 류 재 복[2], 김 선 웅[3]

요 약

유한모집단에서 표본을 추출할 때 표본으로 추출되는 경우의 수를 줄이거나 바람직하지 않은 표본들이 추출될 확률을 줄이기 위해서 관리적 선정 방법을 사용한다. 본 연구에서는 여러 가지 표본추출방법들을 관리적 선정에 적용하고 예를 통해서 그 효율성을 비교하였다.

1. 서 론

표본조사에서 조사비용을 증가시키고 실사시 조직의 구성과 관리를 어렵게 하는 바람직하지 않은 표본들이 있다. 이들은 무응답이나 연구자 편의 등에 영향을 주고 추정치의 정도를 낮게 함으로써 연구 특성의 관점에서 볼 때 중요하지 않은 단위나 지리적으로 서로 멀리 떨어져 있는 표본단위 등이 포함된 표본을 말한다.

관리적 선정(controlled selection) 방법은 유한모집단에서 표본추출시 바람직하지 않은 표본들의 추출확률을 작게 하거나 표본추출 경우의 수를 줄이는 방법이다.

Goodman과 Kish(1950)는 층화임의추출법보다 바람직한 표본이 뽑힐 확률을 증가시키면서 바람직하지 않은 표본의 추출확률은 작게 해주는 관리적 선정 방법을 제시하였다.

표본으로 추출되는 경우의 수를 줄이기 위해 Chakrabarti(1963)는 균형불완비블럭계획(balanced incomplete block design : BIBD)을 이용하였으며 Avadhani와 Sukhatme(1973) 등도 관리적 선정을 위해 여러 가지 BIBD를 사용하였다. 또한 Gupta, Kumar와 Nigam(1982)은 BIBD에 단위의 크기에 비례하는 확률비례추출법(inclusion probability proportional to size : IPPS)을 적용하였다.

한편 Rao와 Nigam(1992)은 BIBD 등의 실험계획법을 이용하여 표본으로 추출되는 경우의 수를 작게 하는 것이 반드시 관리적 선정 계획의 최적 기준이 될 수는 없다는 점에 착안하여 비관리적 선정 하에서의 특성을 유지하면서 바람직하지 않은 표본들의 추출 확률을 최소화시키는 통합접근법(unified approach)을 이용한 관리적 선정 방법을 제안하였다.

Ryu(1996)는 관리적 선정 방법의 한계점과 실제조사에 이 방법을 적용할 때 발생하는 문제점을 파악하고 이들을 효율적으로 사용하기 위한 연구 방향과 과제를 제시하였다. 그리고 Ryu와 Lee(1997)는 최대엔트로피 원리를 사용하여 관리적 선정 계획을 얻었고 바람직하지 않은 표본의

1) (100-715) 서울 중구 필동 3가 26번지 동국대학교 통계학과 교수
2) (360-764) 충청북도 청주시 상당구 내덕동 36번지 청주대학교 응용통계학과 교수
3) (100-715) 서울 중구 필동 3가 26번지 동국대학교 통계학과 박사과정

순위를 고려했을 때 관리적 선정 계획을 얻기 위한 알고리즘을 제시하였다.

그런데 Rao와 Nigam(1992)은 단지 각 표본 추출 방법에서 추정량의 분산과 일치하는 관리적 선정 계획을 얻기 위한 통합접근법을 제안하였을 뿐, 분산을 줄일 수 있는 표본추출계획에 대해서는 고려하지 않았다.

따라서 본 논문에서는 통합접근법을 개선하고자, 주어진 정보를 이용하여 추정량의 분산을 줄이는 표본추출방법을, 바람직하지 않은 표본들의 추출확률을 최소화시킬 수 있는 관리적 선정 방법에 적용하고, 예제를 통하여 기존의 추출방법들과 효율성을 비교하였다.

2. 통합 관리적 선정

어떤 특성의 총합 Y의 일반선형추정량은 다음과 같다.

$$\hat{Y}=\sum_{i\in s}d_i(s)y_i,\ s\in S \tag{2.1}$$

여기서 S는 모든 가능한 표본 s의 집합이고 $d_i(s)$는 i, s 또는 둘 다에 의존한다. Horvitz-Thompson 추정량의 경우 $d_i(s)=1/\pi_i$ 이다.

관리적 선정 하에서 표본 s가 추출될 확률은 $p(s)$ 이고 $p_0(s)$를 비관리적 선정 하에서 표본 s가 뽑힐 확률이라 하면 $\hat{Y}$는 $p_0(s)$하에서 Y의 불편추정량이 된다.

$\hat{Y}$의 분산은 Rao(1979)에 의해 다음과 같이 주어진다.

$$Var(\hat{Y})=-\sum_i\sum_{i<j}d_{ij}w_iw_j(y_i/w_i-y_j/w_j)^2 \tag{2.2}$$

$$d_{ij}=\sum_{s\ni i,j}d_i(s)d_j(s)p(s)-1 \tag{2.3}$$

여기서 w_i는 기지의 상수이다.

비관리적 선정 하에서 d_{ij}와 π_{ij}를 각각 d_{ij}^o, $\pi_{ij}^o=\sum_{s\ni i,j}p_0(s)$ 라 하고 S_1은 S의 부분집합으로써 바람직하지 않은 표본들의 집합이라 하면 최적 관리적 선정 계획은 조건

$$\begin{aligned}
&(\mathrm{i})\ p(s)\geq 0,\ s\in S\\
&(\mathrm{ii})\ \sum_{s\in S}p(s)=1\\
&(\mathrm{iii})\ \text{불편조건 : }\sum_{s\ni i}d_i(s)p(s)=1,\quad i=1,\cdots,N\\
&(\mathrm{iv})\ \text{분산일치조건 : }\\
&\qquad\sum_{s\ni i,j}d_i(s)d_j(s)p(s)=d_{ij}^0+1,\quad i<j=1,\cdots,N
\end{aligned} \tag{2.4}$$

하에서 $\{s\,;s\in S,\,p(s)\}$에 대한 목적함수 $\phi=\sum_{s\in S_1}p(s)$ 을 최소로 하는 해가 된다.

$p(s)$ 는 심플렉스법에 의해 쉽게 얻어지며 본 논문에서는 SAS/OR을 사용하였다.

3. 표본추출방법

Horvitz와 Thompson(1952)은 다음과 같은 모집단 총합의 추정량과 그 분산을 제안하였다.

$$\hat{Y}_{HT}=\sum_{i}^{n}\frac{y_i}{\pi_i} \tag{3.1}$$

$$V(\hat{Y}_{HT})=\sum_{1}^{N}\frac{y_i^2}{\pi_i}+2\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}-Y^2 \tag{3.2}$$

여기서 π_i는 i번째 단위가 표본에 포함될 확률, π_{ij}는 i번째와 j번째 단위가 동시에 표본에 포함될 확률이며 y_i는 i번째 단위의 y값이다.

π_i가 y_i에 비례할 경우 추정량의 분산은 0에 근사하게 된다. 그러나 y_i가 미지이므로 기지의 보조 특성치 x가 특성치 y에 비례한다고 할 때 π_i가 x_i에 비례하도록 표본을 뽑는 것이 바람직하다.

그러한 표본을 뽑는 방법은 여러 가지가 있을 수 있으며 각 방법에 따른 π_{ij}는 추정치의 분산에 영향을 준다.

다음에 논의되는 방법들은 추정량의 분산을 줄이기 위해 제안된 방법들이며 이들을 관리적 선정에 적용한다.

3.1 Raj 방법

x와 y에 대한 사전정보를 활용함으로써 분산을 최소화하는 Raj(1956)의 방법을 관리적 선정계획을 얻는 데 사용한다.

Raj는 $n=2$인 경우 식(3.2)와 같이 주어지는 분산을 최소화시키기 위해서 다음 조건

$$\begin{aligned}&\pi_{ij}\geq 0\ ,\\&\sum_{j\neq i}\pi_{ij}=(n-1)\pi_i=(n-1)np_i,\quad i=1,\cdots,N\end{aligned} \tag{3.3}$$

하에서 선형계획법을 이용하여 $\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}$ 를 최소화시키는 π_{ij}를 구하였다.

먼저 사전정보에 의하여 x와 y가 다음과 같은 선형관계를 갖는다고 가정하자. 즉

$$y=\alpha+\beta x$$

그러면 $\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}$ 는 다음과 같이 $n\geq 2$ 인 경우로 전개된다.

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}(\alpha^2+\alpha\beta x_i+\alpha\beta x_j+\beta^2 x_i x_j) \\ &= \frac{\alpha^2}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}+\frac{\alpha\beta}{2}\left(\sum_{j=1}^{N}\sum_{i\neq j}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}x_i+\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}x_j\right) \\ &\quad +\frac{\beta^2}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}x_i x_j \\ &= \frac{\alpha^2}{2}\sum\sum\frac{\pi_{ij}}{\pi_i\pi_j}+\frac{N(n-1)}{n}\alpha\beta\left(\sum^{N}x_i\right)+\frac{(n-1)}{2n}\beta^2\left(\sum^{N}x_i\right)^2\end{aligned} \tag{3.4}$$

여기서 $\pi_i=nx_i/\sum_{i=1}^{N}x_i$, $\sum_{j\neq i}\pi_{ij}=(n-1)\pi_i$ 이다.

이러한 가정 하에서는 분산을 최소화시키기 위하여 다음과 같은 목적함수를 사용할 수 있다. 이때 목적함수는 $p(s)$ 의 함수로 표시될 수 있다.

$$\begin{aligned}\phi_1 &= \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{\pi_i\pi_j}\pi_{ij} \\ &= \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{np_i\ np_j}\sum_{i,j\in s,\,s\in S}p(s) \\ &= \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{np_i\ np_j}\Big[\sum_{i,j\in s,\,s\in(S-S_1)}p(s)+\sum_{i,j\in s,\,s\in S_1}p(s)\Big]\end{aligned} \tag{3.5}$$

그리고 다음과 같은 또 다른 목적함수를 고려할 수 있다.

$$\phi_2 = \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{np_i\ np_j}\sum_{i,j\in s,\,s\in S_1}p(s) \tag{3.6}$$

결과적으로 목적함수 ϕ_1 을 이용하면 주어진 가정 하에서 분산을 최소화시키는 관리적 선정 계획을 얻을 수는 있으나 바람직하지 않은 표본들이 뽑혀질 확률을 최소화하지는 못한다. 반면에 목적함수 ϕ_2 를 이용하면 분산을 줄이는 동시에 바람직하지 않은 표본들이 뽑혀질 확률을 최소화시킬 수 있는 관리적 선정 계획을 얻을 수 있다.

3.2 Hartley와 Rao의 방법

Hartley와 Rao(1962)는 π_i 가 x_i 에 비례하는, 즉, 다음 식이 성립하는 표본을 뽑는 방법을 이용하였다.

$$(n-1)\pi_i = \sum_{j \neq i}^{N} \pi_{ij} = \gamma x_i \tag{3.7}$$

위 식이 만족되는 표본추출절차를 만들기 위한 조건들은 다음과 같다.

$$\begin{aligned} &\gamma = n(n-1)/X \\ &np_i \leq 1 \end{aligned} \tag{3.8}$$

여기서 $p_i = x_i / X$ 이며 $X = \sum_{i=1}^{N} x_i$ 이다.

따라서 식(3.8)이 만족되는 다음과 같은 표본추출절차를 고려한다.

(1) 추출단위들을 임의의 순서로 배열하고 이 임의의 순서를 $j = 1, 2, \cdots, N$ 으로 표기한다. 이 때 np_i 의 합은 다음과 같이 나타낸다.

$$\Pi_j = \sum_{i=1}^{j} np_i, \quad \Pi_0 = 0 \tag{3.9}$$

(2) 임의출발점을 선택한다. 즉, $0 \leq d < 1$인 임의의 변량 d를 선택한다. 이 때 선택되는 n 개의 단위들은 첨자 j가 다음을 만족시키는 단위들이다.

$$\Pi_{j-1} \leq d + k < \Pi_j \tag{3.10}$$

여기서 k는 0과 $n-1$ 사이의 정수이다.

$n \geq 2$ 인 경우 이러한 표본추출절차에 따라 얻어지는 π_{ij} 는 $O(N^{-4})$ 까지 전개했을 때 다음과 같이 된다.

$$\begin{aligned} \pi_{ij} \doteq\ & \frac{(n-1)}{n} \pi_i \pi_j + \frac{(n-1)}{n^2} (\pi_i^2 \pi_j + \pi_i \pi_j^2) \\ & - \frac{(n-1)}{n^3} \pi_i \pi_j \sum_1^N \pi_k^2 + \frac{2(n-1)}{n^3} (\pi_i^3 \pi_j + \pi_i \pi_j^3 + \pi_i^2 + \pi_j^2) \\ & - \frac{3(n-1)}{n^4} (\pi_i^2 \pi_j + \pi_i \pi_j^2) \sum_1^N \pi_k^2 + \frac{3(n-1)}{n^5} \pi_i \pi_j \Big(\sum_1^N \pi_k^2\Big)^2 - \frac{2(n-1)}{n^4} \pi_i \pi_j \sum_1^N \pi_k^3 \end{aligned} \tag{3.11}$$

이렇게 하여 얻어진 π_{ij} 를 식(3.2)에 대입시켜 $O(N^{-1})$ 까지 구해보면 다음과 같이 되며 이 분산은 복원 PPS 표본추출 하에서 얻어진 추정량의 분산보다 항상 작게 된다.

$$V(\hat{Y}_{HT}) \doteq \sum_1^N \left(1 - \frac{(n-1)}{n} \pi_i\right) \left(\frac{y_i}{\pi_i} - \frac{Y}{n}\right)^2 \tag{3.12}$$

그런데 앞의 식(3.2)를 Yates와 Grundy(1953)가 유도한 식으로 바꾸어 표현하면 다음과 같다.

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \tag{3.13}$$

위 식은 $w_i = \pi_i$ 인 식(2.2)와 동일한 형태이므로 다음과 같은 식을 얻는다.

$$d_{ij}^0 = \frac{\pi_{ij}}{\pi_i\pi_j} - 1$$

식(2.4)의 분산일치조건에 위 식과 $d_i^0(s) = 1/\pi_i$ 를 대입하여 정리하면 분산일치조건은 다음과 같이 간단해진다.

$$\sum_{s \ni i,j} p(s) = \pi_{ij}^0 \tag{3.14}$$

따라서 조건식 (3.14)하에서의 목적함수 $\phi = \sum_{s \in S_1} p(s)$ 를 최소로 하는 관리적 선정 계획을 세울 수 있다.

3.3 Sampford 방법

Sampford(1967)가 제안한 IPPS표본추출법에 의한 $\hat{Y}_{HT}$ 의 분산은 복원 PPS 표본추출에 의한 추정량의 분산보다 항상 작으며 다음과 같은 분산추정량의 비음성(nonnegativity) 조건을 만족한다.

$$\pi_{ij} > 0, \quad \pi_{ij} - \pi_i\pi_j \le 0 \tag{3.15}$$

이 표본추출법을 적용하는 관리적 선정은 이러한 특성들을 유지하면서 바람직하지 않은 표본이 뽑힐 확률을 최소화시킨다.

또한 이 방법에서도 $d_i^0(s) = 1/\pi_i$ 이며 분산일치조건은 Hartley와 Rao의 방법과 동일하다. 즉,

$$\sum_{s \ni i,j} p(s) = \pi_{ij}^0, \quad i < j = 1, \cdots, N \tag{3.16}$$

3.4 Midzuno-Sen 방법

이 방법은 Midzuno(1952), Sen(1953)이 고안한 것으로 총크기비례추출(probability proportional to aggregate size : PPAS)방법이라고도 한다.

이 표본추출방법에 대한 비추정량은 식(3.17)과 같으며 PPAS 표본추출 하에서 불편추정량이 되고 y_i/x_i 의 이상치에 의한 영향을 적게 받는다.

$$\hat{Y}_r = (Y_s/X_s)X \tag{3.17}$$

여기서 $p_0(s) = X_s/\left[X\binom{N-1}{n-1}\right]$, $X_s = \sum_{i\in s} x_i$, $Y_s = \sum_{i\in s} y_i$.

PPAS 표본을 뽑는 가장 간단한 방법은 첫 번째 표본 단위를 확률 $p_i = x_i/\sum_{i=1}^{N} x_i$ 로 뽑고 나머지 $(n-1)$개 단위는 단순임의 비복원추출로 뽑는다.

이 비추정량 $\hat{Y}_r$ 대해서 $d_i(s) = X/X_s$이며 분산일치조건은 다음과 같다.

$$d_{ij}^0 = \sum_{s \ni i,j} (X/X_s)^2 p_0(s) - 1 \tag{3.18}$$

이러한 조건하에서 얻어진 최적 관리적 선정은 PPAS 표본 추출 하에서 $\hat{Y}_r$의 분산에 대응하고 동시에 바람직하지 않은 표본들이 뽑힐 확률을 최소화시킨다.

3.5 BIBD를 이용한 방법

이 방법은 Gupta, Kumar와 Nigam(1982)이 제안한 것으로서 먼저 모수 $v=N, b, r, k \ge n, \lambda$를 갖는 균형불완비블럭계획을 선택한 후 다음의 확률로 b개의 블럭으로부터 1개의 블럭 s을 선택한다.

$$p_0(s) = v(r\sum_{i\in s} p_i - \lambda)/\{b(r-\lambda)\} \quad (s=1,\cdots,b) \tag{3.19}$$

만약 $k=n$일 경우 선택된 블럭 s는 크기 n인 표본이 된다. 그렇지 않은 경우는 선택된 블럭 s에 속하는 k개의 단위로부터 비복원 동일 확률로 크기 n인 부표본을 뽑는다. 이 방법은

$$\begin{aligned}\sum_{i\in s} p_i > \lambda/r &= (k-1)/(N-1) \\ &= (n-1)/(N-1) + (k-n)/(N-1)\end{aligned} \tag{3.20}$$

을 만족시키는 모집단에 대해서만 적용할 수 있으며 크기 n인 표본에 있는 i번째 단위의 포함확률은 $\pi_i = np_i$이다.

또한 균형불완비블럭계획 (v, b, r, k, λ)에서 모든 3개의 단위가 동시에 α개의 블럭에서 나타날 경우 이를 이중균형계획(doubly balanced design)이라고 한다. 이러한 계획을 3-계획(3-design)이라고도 하며 다음이 항상 성립한다.

$$\alpha(v-2) = \lambda(k-2) \tag{3.21}$$

이 이중균형계획에 의한 표본추출계획은 식(3.19)과 (3.20)에 의해 다음 식이 얻어진다.

$$\pi_{ij}^{0} = n(n-1)\{(p_i+p_j)-1/(N-1)\}(N-2)^{-1} \tag{3.22}$$

이중균형계획에 의한 표본추출계획은 복원 PPS 추출방법보다 분산이 작으며 식(3.22)에 대응하는 관리적 선정 방법은 이러한 특성을 유지할 수 있다.

4. 예 제

세 모집단 A, B, C는 각각 $N=7$인 마을로 구성되어 있다. 각 모집단은 주요 과일의 총생산량(y)을 추정하기 위해 표본조사를 실시하며 각 모집단의 경작지 면적(x)은 사전에 알려져 있다(Avadhani와 Sukhatme(1973)).

이 경우 각 모집단으로부터 비복원추출에 의해 $n=3$인 35개의 표본이 있으며 이 중 다음과 같은 14개 표본은 실사를 위한 비용과 조직관리의 어려움 등을 고려하여 바람직하지 않은 표본(S_1)으로 분류하였다.

(1 2 3) (1 2 6) (1 3 6) (1 3 7) (1 4 6) (1 4 7) (1 6 7)
(2 3 4) (2 3 6) (2 3 7) (2 4 6) (2 4 7) (3 4 7) (4 6 7)

각 모집단의 마을들과 관련한 p_i값은 다음과 같으며 y와 x 사이의 관계가 선형이라고 가정한다.

모집단	p_i값						
A	0.10	0.12	0.14	0.15	0.15	0.16	0.18
B	0.10	0.12	0.12	0.18	0.15	0.20	0.13
C	0.25	0.19	0.16	0.15	0.12	0.08	0.05

위의 모집단들의 p_i값은 각각 Rao와 Nigam(1992), Gupta, Nigam과 Kumar(1982), Nigam, Kumar와 Gupta(1984)의 논문의 예제로부터 인용하였다.

앞서 언급한 방법들에 대하여 선형계획법을 이용하여 관리적 선정 계획을 얻을 수 있다.

BIBD를 이용한 표본추출방법을 사용하기 위해 본 예제의 35개의 표본들을 BIBD의 모수로 표현하면 다음과 같은 이중균형계획이 된다.

$$N=7,\ n=3,\ b=35,\ r=15,\ \lambda=5$$

편의상 Raj 방법은 각각의 목적함수 ϕ_1, ϕ_2에 따라 Raj (1), (2)로 구분하며 Hartley와 Rao의 방법은 H-R, Sampford 방법은 S, Midzuno와 Sen 방법은 M-S, BIBD를 이용한 방법은 BIBD로 표기한다.

모집단 A, B, C에 대해 각 방법에 대응하는 관리적 선정과 비관리적 선정 하에서의 바람직하지

않은 표본들이 추출될 확률, 즉 $\phi = \sum_{s \in S_1} p(s)$ 는 아래의 표와 같이 얻어진다. 단 모집단 C는 비관리적 선정 하에서 $p(s) < 0$ 인 경우가 발생하므로 Raj 방법 (1), (2)는 $\phi = \sum_{s \in S_1} p(s)$ 를 고려할 수 없으며 BIBD를 이용한 표본추출방법의 경우 식(3.20)이 만족되지 않으므로 결과를 얻을 수 없다.

<표> 각 추출방법에 따른 관리적 선정 및 비관리적 선정 하에서 ϕ값의 비교

추출방법 / 모집단	Raj (1)	Raj (2)	H -R	S	M - S	BIBD
A	0.490000 0.385000 -27.3	0.000000 0.385000 100	0.189665 0.379729 50.1	0.192624 0.381465 49.5	0.155358 0.396667 60.8	0.202000 0.385000 47.5
B	0.270000 0.385000 29.9	0.000000 0.385000 100	0.191649 0.382580 49.9	0.195609 0.385110 49.2	0.121328 0.396667 69.4	0.196000 0.385000 49.1
C	0.490000 * *	0.040000 * *	0.301629 0.390156 22.7	0.325102 0.400466 18.8	0.157265 0.410667 61.7	* * *

(주) 각 셀의 첫 번째, 두 번째 값은 각각 관리적 선정과 비관리적 선정 하에서의 ϕ값이며 세 번째 값은 비관리적 선정에 대한 관리적 선정의 ϕ값 감소율이고, * 는 적용이 불가능한 경우임.

5. 결 론

Rao와 Nigam(1992)이 제안한 통합접근법을 이용한 관리적 선정 방법은 일반 선형 추정량의 분산과 일치하는 표본추출계획을 세우는 방법으로서, 바람직하지 않은 표본이 추출될 확률을 최소화시키는 관리적 선정 방법이다. 이 방법의 경우 분산식 중 d_{ij}^0에만 일치시켜 바람직한 특성을 유지하였으나 특성 y에 대한 보조정보 x를 충분히 활용하지는 못하였다. 그러므로 보조 특성치 x가 특성치 y에 비례하는 경우, Raj 방법과 같이 추정치의 분산에 영향을 미칠 수 있는 목적함수를 구성하여 선형계획법을 이용함으로써 분산을 줄이면서 바람직하지 않은 표본이 추출될 확률을 최소로 하는 관리적 선정 계획을 세울 수 있다.

앞의 예제를 살펴보면 관리적 선정 방법이 비관리적 선정 방법에 비해 ϕ값이 모두 작을 뿐 아니라, 이 중 Raj (2)는 모집단 A, B에 대하여 ϕ값의 감소율이 가장 크며 모집단 C에 대해서도 ϕ값이 다른 방법들 보다 작다는 것을 알 수 있다.

또한 BIBD 방법의 경우에는 비교를 위해 다른 방법들과 블록의 수가 동일한 경우만을 다루었으나 Hartley와 Rao 방법, Sampford 방법, Midzuno-sen 방법에 비해 ϕ값이 큰 차이가 나지 않았다.

따라서 보조 특성치 x에 대한 정보나 기본 가정 등을 보완하거나 표본 추출 방법을 개선시킴으로써 보다 적절한 관리적 선정 계획을 세울 수 있다.

한편 표본의 관리적 측면에서 볼 때 BIBD 방법이 다른 방법들에 비해 ϕ값이 비슷하거나 동일

하다면, 블럭의 수를 더 줄일 수 있으므로 BIBD 방법이 바람직하다. 그러므로 다양한 실험계획방법을 이용하면 간단하게 관리적 선정을 할 수 있게 될 것이다.

참고문헌

[1] Avadhani, M. S. and Sukhatme, B. V. (1973). Controlled sampling with equal probabilities and without replacement, *International Statistical Review*, Vol. 41, 175-182.

[2] Chakrabarti, M. C. (1963). On the use of incidence matrices of designs in sampling from finite populations, *Journal of the Indian Statistical Association*, Vol. 1, 78-85.

[3] Cochran, W. G. (1977). *Sampling technique*, 3rd ed. John Wiley & Sons, New York.

[4] Godambe, V. P. (1955). A unified theory of sampling from finite populations, *Journal of the Royal Statististical Society*, B, 17, 269-278.

[5] Goodman, R. and Kish, L. (1950). Controlled selection - a technique in probability sampling, *The Journal of the American Statistical Association*, Vol. 45, 350-372.

[6] Gupta, V. K., Nigam, A. K., and Kumar, P. (1982). On a family of sampling schemes with inclusion probability proportional to size, *Biometrika*, Vol. 69, 191-196.

[7] Hartley, H. O. and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement, *Annals of Mathematical Statistics*, Vol. 33, 350-374.

[8] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe, *The Journal of the American Statistical Association*, Vol. 47, 663-685.

[9] Midzuno, H. (1952). On the sampling system with probability proportional to sum of sizes, *Annals of the Institute of Statistical Mathematics*(Tokyo), Vol. 3, 99-107.

[10] Nigam, A. K., Kumar, P. and Gupta, V. K. (1984). Some mehtods of inclusion probability proportional to size sampling, *Journal of the Royal Statististical Society*, B, Vol. 46, 564-571.

[11] Raj D. (1956). A note on the determination of optimum probabilities in sampling without replacement, *Sankhyã*, Vol. 17, 197-200.

[12] Rao, J. N. K. (1979). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling, *Journal of the Indian Statististical Association*, Vol. 17, 125-136.

[13] Rao, J. N. K. and Nigam, A. K. (1992). Optimal controlled sampling : a unified approach, *International Statistical Review*, Vol. 60, 89-98.

[14] Ryu, J. B. (1996). A study on the controlled selection, *Korean communications in Statistics*, Vol. 3, No. 3, 135-144.

[15] Ryu, J. B. and Lee , S. J. (1997). Efficient controlled sampling, *Korean communications in Statistics*, Vol. 4, No. 1, 151-159.

[16] Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection, *Biometrika*, Vol. 54, 499-513.

[17] Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities,

Journal of the Indian Society of Agricultural Statistics, Vol. 5, 119-127.
[18] Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size, *Journal of the Royal Statististical Society*, B, Vol. 15. 253-261.