한국통계학회논문집 제3권 3호
1996년 12월 pp. 135-144

관리적 선정에 관한 연구[1)]

류 제 복[2)]

요 약

표본조사에서는 실사의 비용을 줄여주고 추정치의 정도를 높여 주는 바람직한 표본이 추출되기를 기대한다. Goodman과 Kish(1950)는 기존의 추출방법의 성질을 변화하지 않으면서 바람직한 표본의 추출확률을 높게 해주고 반면에 바람직하지 않은 표본의 추출확률을 작게 해주는 관리적 선정(controlled selection)방법을 제시하였다. 본 논문에서는 지금까지 관리적 선정에 관하여 연구되어 온 방법들을 정리, 분석하고 각 방법들을 상호 비교함으로써 관리적 선정방법이 갖고 있는 한계점과 실제조사에 이 방법을 적용할 때 발생하는 문제점을 파악하여, 향후 관리적 선정방법을 효율적으로 사용하기 위한 연구 방향과 과제를 제시하였다.

1. 머리말

넓은 지역으로 구성된 모집단에서 단순임의추출을 사용하여 표본을 추출하게 되면 추출된 표본이 연구 대상이 되는 특성의 관점에서 볼 때 중요하지 않은 표본이 될 수 있다. 또한 지리적으로 도심으로부터 멀리 떨어져 있거나 교통이 극히 불편한 지역이 추출되면 조사비용이 증가하게 되고 실사를 위한 조직의 구성과 관리가 어렵게 되어 수집된 자료가 무응답이나 연구자 편의 등에 영향을 받게 된다. 그리고 여러 종류의 작물에 대한 수확량 조사와 같은 한번의 조사에 여러 항목을 동시에 조사하고자 하는 다목적 조사에서 가급적이면 연구 대상이 되는 항목을 많이 포함하고 있는 표본이 추출되는 것이 바람직하게 된다.

따라서 표본추출시에 바람직한 표본의 추출확률을 높게 해주고 바람직하지 않은 표본의 추출확률을 작게 해 줌으로써 추정치의 정도를 높여줄 수 있게 된다. 이를 위해서 층화추출을 사용할 수 있으나 이것도 항상 충분하지는 않다. 그리고 집락추출이나 부차표본 기법을 층내에 사용할 수 있지만 이들은 연구 대상이 되는 특성치의 정도에 손실을 가져올 수 있다. Goodman과 Kish (1950)는 층화임의추출에서보다 바람직한 표본의 추출확률을 크게 해 주는 표본 추출방법으로 관리적 선정(controlled selection)방법을 제시하였다. Avadhani와 Sukhatme(1965, 1966, 1967, 1968, 1973)는 관리적 선정을 위한 여러가지 유용한 방법들을 제안하였으며, Waterton(1983)과 Lin(1992)은 관리적 선정을 할 때 어려운 문제가 되는 유의적인 표본들을 체계적이고 효율적으로 만드는 방법을 소개하였다. 그리고 Rao와 Nigam(1990, 1992)은 바람직하지 않은 표본의 추출확률을 작게 해주기 위해서 선형계획법을 사용하였으며 Sitter과 Skinner(1994)는 이들 방법을 다중

1) 이 논문은 '94-'95학년도 청주대학교 학술연구조성비(특별과제)에 의해서 연구되었음.
2) (360-764) 충청북도 청주시 상당구 내덕1동 36번지 청주대학교 응용통계학과 교수.

층화에 적용하였다. 또한 Hess, Reidel과 Fitzpatrick(1975)은 의료조사에 관리적 선정방법을 적용하였고 Causey, Cox와 Ernst(1985)는 관리적 선정방법에 수송문제를 적용하였다.

본 논문에서는 Goodman과 Kish(1950)이후 관리적 선정방법에 관하여 지금까지 연구되어 온 방법을 정리, 분석하고 각 방법들을 상호 비교함으로써 관리적 선정방법이 갖고 있는 한계점과 실제조사에 이 방법을 적용할 때 발생하는 문제점을 파악하여 향후 연구의 방향과 과제를 제시하고자 한다.

관리적 선정에서 다루는 문제는 크게 두 가지로 나누어 생각할 수 있다. 표본으로 추출되는 경우의 수(support size)를 줄이는 방법에 관해서는 2장에서 다루고 바람직하지 않은 표본의 추출확률을 작게 해주는 문제는 3장에서 다루었다. 그리고 이들 관리적 선정방법을 실제조사에 적용한 몇 가지의 사례를 4장에서 소개하였으며, 5장에서는 관리적 선정방법이 갖고 있는 문제점들을 살펴보고 이의 해결을 위한 방안과 향후 연구되어야 할 부분을 제시하였다.

2. 유의표본집합

표본추출방법은 확률추출(probability sampling)과 비확률추출(non-probability sampling)로 나누어지는 데, 관리적 선정은 개념적으로 보면 비확률추출방법인 유의추출의 확장으로 볼 수 있다. 그러나 확률추출을 하기 위하여 모집단의 모든 단위들이 하나 이상의 표본에 포함될 때까지 많은 유의표본을 뽑아야 한다. 이때 각 단위가 나타나는 표본수는 주어진 추출확률에 정확히 비례해야 한다. 이들 유의표본들로부터 임의로 뽑은 하나의 표본은 확률표본이 된다. 따라서 관리적 선정은 확률추출과 비확률추출인 유의추출을 결합한 방법이 된다.

관리적 선정을 위해서는 유의표본의 추출이 필요하게 되는 데 모집단의 구성 원소 N과 표본수 n이 조금만 크게 되도 유의표본을 만드는 일이 매우 번잡하고 어렵게 된다. 이러한 문제를 해결하기 위해서 Hess, Reidel과 Fitzpatrick(1975), Ernst(1981), Waterton(1983), 그리고 Lin(1992)들은 체계적이고 효율적인 방법들을 제안하였다.

관리적 선정에서는 표본으로 추출되는 경우의 수를 줄이는 것과 바람직하지 않은 표본의 추출확률을 작게 하는 문제를 주로 다루고 있다. 본 장에서는 표본으로 추출되는 경우의 수를 줄이는 문제에 대한 것을 다루고자 한다.

구분이 명백한 최소의 블럭을 갖는 계획을 설정하기 위하여 여러가지 불완전 블럭계획의 조합적 성질이 이용되고 있다. 구분이 명백한 최대의 블럭은 바람직한 표본과 바람직하지 않은 표본으로 구성되고, 하나의 블럭 또는 표본은 사용자가 선택한 블럭계획의 총 b개의 블럭으로부터 주어진 확률을 갖고 임의로 선택된다.

N개의 원소로 구성된 유한모집단에서 비복원으로 n개의 표본을 추출하는 경우의 수는 $\binom{N}{n}$인데 Chakrabarti(1963)는 균형불완비블럭계획(balanced incomplete block design : BIBD)을 이용하여, 추출되는 경우의 수를 b로 줄이는 추출계획을 설계하였다. 실험계획방법을 적용하기 위해서 1차 포함확률과 2차 포함확률이 각각

$$\pi_i = n / N,$$
$$\pi_{ij} = n(n-1) / \{N(N-1)\}$$

이고 BIBD (t, b, r, k, λ)에서 $t = N$이고 $k = n$을 가정한다. 이때 $n > 2$이다. BIBD의 $bk = rt$와 $\lambda(t-1) = r(k-1)$인 성질로부터 $\pi_i = r/b = n/N$, $\pi_{ij} = \lambda/b = n(n-1)/\{N(N-1)\}$가 되며, 유의표본의 크기는 $b \geq t = N$이고, 하한은 대칭 BIBD가 존재한다면 얻어진다.

Avadhani와 Sukhatame(1973)은 단순임의추출과 연관된 관리적 선정을 설계하기 위해서 BIBD를 사용하였으며 특정 BIBD가 존재하지 않는 상황에서, Foody와 Hedayat(1977) 그리고 Wynn (1977)은 반복블럭을 갖는 BIBD를 사용하였다.

한편 포함확률비례(inclusion probability proportional to size : IPPS)추출의 관리적 선정을 위해서 Gupta, Nigam 그리고 Kumar(1982)들은 모집단 총계 Y에 대한 Horvitz-Thompson추정치와 이와 관련해서 BIBD를 사용하였다.

Hedayat, Lin 그리고 Stufken(1989)은 다음 성질을 만족하는 관리적 IPPS추출계획을 설계하기 위해서 "emptying box"방법을 사용하였다.

$$0 < \pi_{ij} \leq \pi_i \pi_j, \quad i < j = 1, 2, \ldots, N.$$

이들은 Sen-Yates-Grundy분산추정치의 불편성과 음이 아닌 성질을 만족시켜준다. 한편 보다 일반화된 $c\pi_i\pi_j \leq \pi_{ij} \leq \pi_i\pi_j$, $i < j = 1, 2, \ldots, N$, $0 < c < 1$인 성질을 만족하는 관리적 IPPS 추출계획을 얻기 위해서 Nigam, Kumar 그리고 Gupta(1984)는 여러 형태의 실험계획법을 사용하였다.

Wynn(1977)은 주어진 표본설계 p_1에 대해서 p_1과 같은 π_i와 π_{ij}를 갖으면서 표본으로 추출되는 경우의 수가 $\frac{1}{2}N(N-1)$보다 크지 않는 p_2가 언제나 존재함을 보여주었다. 그러나 p_2를 만드는 방법을 제시하지는 않았다.

여기서는 관리적 선정에 BIBD를 사용한 예를 들어 설명하고자 한다(Avadhani와 Sukhatame (1973)).

[예1] 아래와 같이 배열된 7개의 마을에서 3개의 마을을 표본으로 추출하는 경우를 생각한다.

$$\begin{array}{ccccc} & 2 & & 1 & \\ 7 & & 5 & & 4 \\ & 6 & & 3 & \end{array}$$

크기 3인 $\binom{7}{3} = 35$개의 표본들 중에서 이동이나 실사작업의 편리함등을 고려하여 14개의 바람직하지 않은 표본집합을 가정한다. 35개의 표본들이 단순임의 추출에서와 같이 균등확률로 추출된다면, 바람직하지 않은 표본이 뽑힐 확률은 14/35 = 0.4가 된다.

<35개의 표본집합 : *는 바람직하지 않은 표본>

(1, 2, 3)*, (1, 2, 4), (1, 2, 5), (1, 2, 6)*, (1, 2, 7), (1, 3, 4), (1, 3, 5), (1, 3, 6)*, (1, 3, 7)*, (1, 4, 5), (1, 4, 6)*, (1, 4, 7)*, (1, 5, 6), (1, 5, 7), (1, 6, 7)*, (2, 3, 4)*, (2, 3, 5), (2, 3, 6)*, (2, 3, 7)*, (2, 4, 5), (2, 4, 6)*, (2, 4, 7)*, (2, 5, 6), (2, 5, 7), (2, 6, 7), (3, 4, 5), (3, 4, 6), (3, 4, 7)*, (3, 5, 6), (3, 5, 7), (3, 6, 7), (4, 5, 6), (4, 5, 7), (4, 6, 7)*, (5, 6, 7)

한편 모수 ($t=7, b=7, r=3, k=3, \lambda=1$)을 갖는 아래와 같은 BIBD를 사용하면 유의표본집합의 수는 7개로 줄어들게 된다. 그리고 이들로부터 표본을 추출하게 되면 단지 하나의 블럭 만이 바람직하지 않은 표본이 되므로 바람직하지 않은 표본이 선택될 확률은 1/7 = 0.14 가 되어 단순임의추출의 경우(0.4)보다 상당히 줄어들게 된다.

<BIBD>

블럭1 : (1, 2, 4)	블럭2 : (2, 3, 5)	블럭3 : (3, 4, 6)	블럭4 : (4, 5, 7)
블럭5 : (5, 6, 1)	블럭6 : (6, 7, 2)	블럭7 : (7, 1, 3)*	

3. 최적선정

관리적 선정에서 표본으로 추출되는 경우의 수를 줄이는 방법들은 2장에서 다루었고 여기서는 바람직하지 않은 표본의 추출확률을 작게 하는 최적선정의 문제를 다루고자 한다.

앞에서 다룬 예1에서는 BIBD를 사용하여 추출 경우의 수를 줄이는 관리적 선정문제를 다루었는데 이 예에서 사용한 것과는 다른 BIBD를 사용할 수 있다. 물론 추출되는 경우의 수는 변하지 않으나 바람직하지 않은 표본의 추출확률은 달라지게 되며, BIBD가 존재하지 않는 경우도 생긴다. 또한 추출확률이 같은 여러 개의 BIBD가 존재할 수도 있다.

Rao와 Nigam(1990, 1992) 그리고 Sitter과 Skinner(1994)들은 선형계획법을 이용하여 최적의 관리적 선정문제를 다루었다.

$p(s)$는 표본 s ($\in S$)가 선택될 확률이다. 이때 표본크기는 2보다 크다. S_1을 바람직하지 않은 표본들의 집합이라 하면, 최적의 관리적 방법인 $p_c(s)$는 아래의 선형제약조건하에서 목적함수 $\phi = \sum_{s \in S_1} p(s)$를 최소로 하는 해가 되며, 이는 심플렉스방법에 의해서 얻는다.

$$\text{(i)} \quad p(s) \geq 0, \quad \text{모든} \quad s \in S,$$
$$\text{(ii)} \quad \sum_{i,j \in s} p(s) = \pi_{ij} \quad (i < j = 1, \ldots, N). \qquad (3\text{-}1)$$

또한 특정한 S_1 대신에 모든 표본에 대한 비용함수 $w(s)$를 사용하여 기대비용을 최소로 하는 일반적인 목적함수 $\phi = \sum_{s \in S} w(s)p(s)$를 사용할 수도 있다. 여기서 $w(s) > 0$이다.

한편 다음과 같은 제약조건을 생각할 수 있다.

$$\text{(i)} \quad p(s) \geq 0, \quad \text{모든} \quad s \in S,$$
$$\text{(ii)} \quad \sum_{i \in s} p(s) = np_i \quad (i = 1, \ldots, N), \qquad (3\text{-}2)$$
$$\text{(iii)} \quad c(np_i)(np_j) \leq \sum_{i,j \in s} p(s) \leq (np_i)(np_j) \quad (i < j = 1, \ldots, N, \; 0 < c < 1).$$

(3-2)의 조건하에서 목적함수 $\phi = \sum_{s \in S_1} p(s)$를 최소로 하면 Nigam, Kumar와 Gupta(1984)의 IPPS추출계획과 일치하는 최적의 관리적 IPPS추출계획을 얻는다. 이 계획은 안정적이고 음이 아닌 불편분산추정치를 제공한다. 그러나 제약조건 (3-2)로부터 얻어진 최적계획에 의한 $\hat{Y}$의 분산이 복원인 PPS추출에 의한 일반적인 추정치의 분산보다 항상 작다는 것을 보장할 수는 없다.

이 방법을 예1에 적용하면 (3-1)식에서 $\pi_{ij} = n(n-1) \,/\, \{N(N-1)\} = \frac{1}{7}$이고 최적계획은 $\phi_{\min} = \frac{1}{7}$이 된다. 이는 BIBD를 사용하여 얻어진 것과 같은 값을 가진다. 그러나 여기서 얻은 해는 선형계획에 의한 하나의 해가 되지만 유일하지는 않다.

[예2] 예1에 마을 8을 첨부했을 때 구분이 명백한 블럭을 갖는 BIBD가 존재하지 않는다고 하자(Rao와 Nigam(1990)).

2 1
7 5 4
6 3
8

$N = 8$에서 $n = 3$을 추출하는 관리적 단순임의추출계획을 적용했을 때, 예1과 같은 기준으로 21개의 바람직하지 않은 표본을 가정한다. 가능한 표본 수는 56개이므로 단순임의추출에 의하면 $\phi = 0.375$가 되나, (3-1)식에 의한 최적계획은 $\phi_{\min} = 0.1607$이 된다. 이 최적계획은 사실 26개의 구분이 명확한 블럭 또는 표본을 반복 사용한 BIBD가 된다. 그러므로 표본추출 경우의 수가 56에서 26으로 줄어든다. 물론 BIBD가 존재하는 경우는 표본추출 경우의 수가 작게 되지만, 표본추출 경우의 수를 최소로하는 것이 최적의 관리적 계획을 얻는 데 반드시 필요하지는 않다.

총계 Y에 대한 일반선형추정치가 다음과 같을 때 관리되지 않은 계획하에서 $\hat{Y}$는 불편추정치가 된다.

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

가중치 $d_i(s)$는 i 또는 s에 의존하거나 둘 다에 의존한다. 예를 들면 Horvitz-Thompson추정치의 경우 $d_i(s) = \frac{1}{np_i}$이 되어 불편성을 만족하게 된다.

Rao와 Nigam(1992)은 표준선형계획을 사용한 통합접근법(unified approach)으로 최적의 관리적 선정을 얻었다. 즉, 최적의 관리적 선정방법인 $p_c(s)$는 아래의 선형제약조건하에서 목적함수 $\phi = \sum_{s \in S_1} p(s)$를 최소로 하는 해가 된다.

(i) $p(s) \geq 0$, $s \in S$,

(ii) $\sum_{s \in S} p(s) = 1$,

(iii) 불편성의 조건 : $\sum_{i \in s} d_i(s)\, p(s) = 1, \quad i = 1, \ldots, N,$ (3-3)

(iv) 분산일치조건 : $\sum_{i, j \in s} d_i(s)\, d_j(s)\, p(s) = d_{ij}^0 + 1 \; , \; i < j = 1, 2, \ldots, N.$

이때 $d_{ij} = E[d_i(s)d_j(s)] - 1$ 이고, 관리되지 않은 계획하에서 $p(s)$는 $p_0(s)$로 d_{ij}는 d_{ij}^0로 표시한다. 또한 특정한 S_1 대신에 모든 표본에 대한 비용함수를 생각하고 기대비용을 최소로 하는 일반적인 목적함수 $\phi = \sum_{s \in S} w(s)\, p(s)$를 사용할 수 있다. 단, $w(s) \geq 0$이다.

통합접근법에 의한 관리적 선정방법을 Lahari(1951), Murthy(1957) 그리고 Sampford(1967)방법에 적용하면 이들은 각각 본연의 성질을 유지하면서 바람직하지 않은 표본의 추출확률을 작게 해준다.

선형계획을 사용한 최적해 $p_c(s)$는 심플렉스방법에 의하여 얻어지나, 해가 반드시 유일하지 않으며 여러 해들 가운데 하나의 해가 이 목적을 위해 사용된다. 그리고 이 방법은 많은 수의 층으로부터 작은 수의 1단계 단위를 추출할 때 유용하다. 또한 최적 관리적계획은 누적합 또는 Lahiri(1951)방법을 사용하여 쉽게 수행될 수 있다.

4. 적용사례

관리적 선정방법을 실제조사에 적용한 사례는 많이 있으나 여기서는 대표적인 몇 가지의 사례만을 소개하고자 한다.

[사례1, Goodman과 Kish, 1950]

미국의 중북부 주들을 대표하는 21개의 일단계 추출단위들을 선정하기 위해서 이중층화추출을 사용하였다. 추출단위는 약 180만 인구로 구성된 17개의 층으로 묶고 층의 순서는 대도시, 소도시 그리고 농촌지역 순으로 하였다. 각 단위에 할당된 추출확률은 1940년 인구 수에 비례한다. 다른 하나의 층화변수로는 조사대상 주들을 4개의 그룹으로 묶어서 사용하고 여기에 관리적 선정방법을 적용하였다. 일반적인 층화추출을 사용할 경우, 관리적 선정을 위한 유의표본집합이 선택될 확률은 .0000058 인 반면에, 관리적 선정방법에서는 이들이 선택될 확률이 1이 된다. 또한 전체 분산중 급간분산이 차지하는 비율이 높은 항목에 관리적 선정을 사용함으로써 분산을 상당히 줄여 주고, 급내분산의 비중이 크지 않은 항목에 대해서도 분산의 감소가 있음을 알 수 있다.

[사례2, Avadhani와 Sukhatme, 1967]

인도의 Uttar Pradesh의 Nainital지역에서 사과, 배, 건포도, 살구, 버찌 등 5개 주요 과일의 재배 면적, 생산률 그리고 총생산량에 대한 신뢰할 수 있는 추정치를 얻기 위하여 1964-1965년에 실

시된 표본조사 자료를 사용하였다.

전체 지역을 사전정보를 이용하여 몇 개의 층으로 나누었으며 각각의 층에는 93개의 마을이 있다. 수확 지역과 비수확 지역에 따라 서로 다른 과일 나무 수에 관한 정보를 얻기 위하여 10개의 마을을 표본으로 뽑아 전수조사를 실시하였다. 비록 사과가 대부분의 마을에서 경작되지만 몇몇 지역에는 연구대상이 되는 모든 과일이 경작되지 않았다. 그러므로 가급적 선택된 마을이 연구대상이 되는 모든 과일들을 경작하고 있도록 마을을 선택하는 것이 바람직하다.

종전의 단순임의추출방법을 사용하면 바람직하지 않은 표본이 뽑힐 확률이 0.17인 반면에, 이들이 제시한 관리적 임의추출방법을 사용하였을 경우는 0.04로 상당히 작아졌다.

[사례3, Waterton, 1983]

스코틀랜드에 있는 학생들의 학업 성취도를 우편조사를 통하여 측정하기 위해서 297개의 학교로부터 학생들을 추출하고자 한다. 조사를 위해서 추출된 학생이 전체 297개 학교에서 일부에 속하게 되므로 관리적 선정방법을 사용하였다. 관리적 선정은 학생들을 추출하지 않고 학교를 추출하는 데 사용되었다. 층화를 위하여 1977년의 학교자료를 사용하였으며 층화변수로는 학교의 형태(5종류), 학교의 수준(4범위) 그리고 학교의 지리적 위치(4범위)를 사용하여 3중층화추출을 하였다. 총 80개의 셀에서 60개의 학교를 추출할 때, 세개의 다중비례층화추출은 관리적 선정방법에 비해 25%에서 38%정도 효율이 떨어졌다.

여러 개의 특성치를 추정하는 다목적조사에서 하나의 특성치를 추정하기 위한 확률군과 다른 하나의 특성치를 추정하기 위한 또 다른 하나의 확률군을 갖는 추출단위가 선정되는 것이 바람직하게 된다. 그러므로 두개의 서로 다른 표본단위들이 거의 일치하거나 가깝게 되면 실제조사에서의 조사비용을 상당히 줄일 수 있게 된다. 이런 경우에 Raj(1956)는 관리적 선정방법의 사용을 언급하였다.

미국의 '75-'76년 노동력 인구조사에서 각 주(state)에 대해서 표본의 일차추출단위(primary sampling unit) 수가 주의 인구수를 반영하도록 하기 위해서 관리적 선정방법을 사용하였다. 특정의 표본 PSU들을 선정하는 과정은 두 단계로 나누어 실시되었는데, 첫 단계에서는 컴퓨터에 의해서 표본 PSU들의 선정 패턴을 만들고, 두 번째 단계에서는 패턴들에 할당된 확률에 따라 허용패턴 가운데에서 하나를 임의로 선정하여 사용하였다.

Hess, Reidel 와 Fitzpatrick(1975)은 "Probability Sampling of Hospital and Patients"에서 1958년 미시간주에 있는 235개의 병원을 대상으로 병원과 입·퇴원 환자들에 대한 연구를 위해 관리적 선정방법을 사용하였고, 단계적으로 유의적인 추출패턴을 얻기 위한 알고리즘을 제시하였다.

5. 토의 및 향후 연구과제

조사를 계획할 때, 정확한 추정치를 얻어야 하고 조사비용도 가급적이면 줄여야 한다는 점을 항상 고려하게 된다. 이러한 점들을 충족시켜주기 위해서 Goodman과 Kish(1950)는 바람직한 표본의 추출확률을 높게 하고 그렇지 않은 표본의 추출확률을 작게 해주는 관리적 선정방법의 사용을 제시하였다.

관리적 선정에서는 표본의 추출과 관리를 용이하게 하기 위해서 표본으로 추출되는 경우의 수

를 줄여주는 문제와 바람직하지 않은 표본의 추출확률을 작게 해주는 문제를 다루고 있다. 이를 위해서 여러 번의 시행착오를 거쳐 유의적인 표본집합을 만들거나 알고리즘을 사용하고, 실험계획법과 선형계획법등을 사용하여 문제를 해결하고 있다. 그러나 실제조사에 관리적 선정방법을 적용할 경우 추출될 표본의 수가 너무 많고 알고리즘을 사용할 때 완전한 해를 구하기가 어렵게 된다. 또한 선형계획법을 사용하여 최적해를 구하게 되나 이의 계산이 복잡하고, 유일한 해도 항상 존재하지는 않게 된다.

따라서 관리적 선정을 효율적으로 수행하고 실제조사에 이를 원활히 사용하기 위해서는 다음과 같은 문제에 대한 연구가 필요하게 된다.

첫째, 궁극적으로는 바람직하지 않은 표본이 뽑힐 확률이 0이 되게 하는 rejected-sampling기법을 활용한다.

둘째, 유의표본집합을 만들기 위한 새로운 알고리즘을 개발하고 다양한 실험계획법을 사용하되 각 블럭의 추출확률이 동일하지 않은 경우도 고려한다.

셋째, 여러가지 IPPS추출 방법과 이들을 결합한 방법을 모색한다.

넷째, 보다 간편하게 최적해를 구하는 방법으로 최대 entropy이론을 사용하고 아울러 유일한 해를 구하는 문제도 생각한다.

참고문헌

[1] Avadhani, M. S. and Sukhatme, B. V. (1965). Controlled Simple Random Sampling, *Journal of the Indian Society of Agricultural Statistics*, Vol. 17, No. 1, 34-42.

[2] Avadhani, M. S. and Sukhatme, B. V. (1966). A Note on the Ratio and Regression Methods of Estimation in Controlled Simple Random Sampling, *Journal of the Indian Society of Agricultural Statistics*, Vol. 18, No. 2, 17-20.

[3] Avadhani, M. S. and Sukhatme, B. V. (1967). Controlled Sampling with Varying Probabilities with and without Replacement, *The Australian Journal of Statistics*, Vol. 9, No. 1, 8-15.

[4] Avadhani, M. S. and Sukhatme, B. V. (1968). Simplified Procedures for Designing Controlled Simple Random Sampling, *The Australian Journal of Statistics*, Vol. 10, No. 1, 1-7.

[5] Avadhani, M. S. and Sukhatme, B. V. (1973). Controlled Sampling with Equal Probabilities and without Replacement, *International Statistical Review*, Vol. 41, No. 2, 175-182.

[6] Causey, B. D., Cox, L. H., and Ernst, L. R. (1985). Application of Transportation Theory to Statistical Problems, *Journal of the American Statistical Association*, Vol. 80, 903-909.

[7] 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, No. 1, 78-85.

[8] De, K. A. and Roy, B. K. (1988). Choice of Binary Matrices under Restrictions with Application to Controlled Sampling, *Calcutta Statistical Association Bulletin*, Vol. 37, 215-225.

[9] Ernst, L. R. (1981). A Constructive Solution for Two-Dimensional Controlled Selection Problems, *ASA 1981 Proceeding of the Section on Survey Research Methods*, 61-64.

[10] Food, W. and Hedayat, A. (1977). On Theory and Applications of BIB Designs with Repeated Blocks, *The Annals of Statistics*, Vol. 5, No. 5, 932-935.

[11] Goodman, R. and Kish, L. (1950). Controlled Selection - A Technique in Probability Sampling, *Journal of the American Statistical Association*, Vol. 45, 350-372.

[12] 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, No. 1, 191-196.

[13] Hedayat, A., Lin, Bing-ying, and Stufken, J. (1989). The Construction of PPS Sampling Designs through as Method of Emptying Boxes, *The Annals of Statistics*, Vol. 17, No. 4, 1886-1905.

[14] Hess, I., Riedel, D. C., and Fitzpatrick, T. B. (1975). *Probalility Sampling of Hospitals and Patients*, 2nd ed. Ann Arbor, Michigan: Health Administration Press.

[15] Lahiri, D. B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates, *Bulletin of the International Statistical Institute*, Vol. 33, No. 2, 133-140.

[16] Lin, Ting-kwong (1992). Some Improvements on an Algorithm for Controlled Selection, *ASA 1992 Proceeding of the Section on Survey Research Methods*, 407-410.

[17] Murthy, M. N. (1957). Ordered and Unordered Estimators in Sampling without Replacement, *Sankhyā*, Vol. 18, 379-390.

[18] Nigam, A. K., Kumar, P., and Gupta, V. K. (1984). Some Methods of Inclusion Probability Proportional to Size Sampling, *Jouranl of the Royal Statistical Society*, Series B, Vol. 46, No. 3, 564-571.

[19] Raj, Des (1956). On the Method of Overlapping Maps in Sample Surveys, *Sankhyā*, Vol. 17, 89-98.

[20] Rao, J. N. K. and Nigam, A. M. (1989). Controlled Sampling with Probability Proportional to Aggregate Size, *Technical Reports* No. 133, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.

[21] Rao, J. N. K. and Nigam, A. M. (1990). Optimal Controlled Sampling Design, *Biometrika*, Vol. 77, N0. 4, 807-814.

[22] Rao, J. N. K. and Nigam, A. M. (1992). Optimal Controlled Sampling: a Unified Approach, *International Statistical Review*, Vol. 60, No. 1, 89-98.

[23] Sampford, M. R. (1967). On Sampling without Replacement with Unequal Probabilities of Selection, *Biometrika*, Vol. 54, 499-513.

[24] Sitter, R. R. and Skinner, C. J. (1994). Multi-way Stratification by Linear Programing, *Survey Methodology*, Vol. 20, No. 1, 65-73.

[25] Sukhatme, B. V. and Avadhani, M. S. (1965). Controlled Selection a Technique in Random Sampling, *Annals of the Institute of Statistical Mathmatics* (Tokyo), Vol. 17, 15-28.

[26] United State Bureau of the Census : *The Current Population Survey - Design and Methodology*, Government Printing Office, Washington D. C., 1978.

[27] Waterton, J. J. (1983). An Exercise in Controlled Selection, *Applied Statistics,* Vol. 32, No. 2, 150-164.

[28] Wynn, H. P. (1977). Convex Sets of Finite Population Plans, *The Annals of Statistics*, Vol. 5, No. 2, 414-418.