응용통계연구 제 12권 1호, 1999년, pp. 261-274

관리적 선정을 실제조사에 적용하기 위한 체계적 방법 *

류제복[1] 이승주[2] 김선웅[3]

요 약

층화임의추출에서 표본크기가 층의 수보다 작은 경우 또는 조사단위들이 넓게 퍼져 있어 조사비용이 증가하거나 조사의 관리가 어려운 경우에 관리적 선정을 사용한다. 그러나 관리적 선정을 실제조사에 적용할 때 많은 어려움이 있다. 이에 본 논문에서는 관리적 선정을 실제조사에 적용하기 위한 체계적 절차를 제시하였고, 바람직하지 않은 표본들을 순차적으로 제외하고 관리적 선정계획을 세우기 위해서 다양한 측도를 사용하였다. 그리고 예제를 통하여 본 논문에서 제시한 절차의 사용 적합성을 검토하고 측도들을 비교하였다.

1. 머리말

유한모집단으로부터 표본을 추출할 때 통상적으로 확률추출을 사용한다. 그러나 모집단에서 추출된 표본 중에는 연구 특성상 중요하지 않은 표본이 포함될 수 있다. 지리적으로 멀리 떨어져 있거나 교통이 극히 불편한 지역에 위치한 표본들은 조사비용을 증가시키고, 실사를 수행할 때 조직의 구성과 조사관리를 어렵게 한다. 그러므로 이런 바람직하지 않은 표본이 추출되면 조사된 자료가 무응답이나 연구자 편의 등에 의해 크게 영향을 받게 되므로 정확한 조사결과를 얻을 수 없게 된다. 이러한 문제를 해결하기 위해서 Goodman과 Kish(1950)는 비확률추출에 확률추출의 개념을 도입한 관리적 선정(controlled selection)을 제안하였다.

관리적 선정은 표본추출시 바람직한 표본의 추출확률을 높여주는 반면에 바람직하지 않은 표본의 추출확률을 작게 해주는 방법으로, 층화임의추출에서 층의 수보다 표본수가 작거나 여러 항목을 동시에 조사하는 다목적 조사에도 유용하게 사용된다.

관리적 선정에서는 표본으로 추출되는 경우의 수를 줄여주는 문제와 바람직하지 않은 표본의 추출확률을 작게 해주는 것이 주로 연구되고 있다. 전자에 대해서, Chakrabarti(1963), Avadhani와 Sukhatme(1973), 그리고 Gupta, Kumar와 Nigam(1982)등은 여러 가지 실험계획법을 사용하였으며, Hedayat. Lin, 그리고 Stufken(1989)은 상자 비우기(emptying boxes) 방법을 사용하였다. 한편 Rao와 Nigam(1990, 1992), Sitter와 Skinner(1994), 그리고 Tiwari와 Nigam(1998)은 바람직하지 않은 표본의 추출확률을 작게 해주기 위해서 선형계획법을 사용하였다. 관리적 선정을 실제조사에 효율적이고 체계적으로 적용하기 위해서 Hess, Reidel, 그리고 Fitzpatrick(1975), Ernst(1981), 그리고 Lin(1992) 등은 알고리즘

* 이 논문은 1997년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

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

을 제시하였다. Ryu(1996)는 관리적 선정을 실제조사에 적용할 때 발생할 수 있는 문제점을 지적하고, 이를 효율적으로 사용하기 위한 연구방향과 과제를 제시하였다. 또한 Ryu와 Lee(1997)는 관리적 선정계획을 얻기 위해 최대엔트로피 원리를 사용하였으며 바람직하지 않은 표본의 순위를 고려한 관리적 선정계획을 얻기 위한 알고리즘을 제시하였다.

관리적 선정이 실제조사에 필요함에도 불구하고 이제까지 잘 사용되지 않고 있는 이유는 유의표본집합을 만드는 일(표본이 선정될 확률의 계산)이 매우 번거롭고 복잡하기 때문이다. 그리고 조사비용이나 관리 등을 고려하여 일부의 표본을 제외시키고 관리적 선정계획을 세울 때 중요한 표본이 제외되기도 한다. 따라서 본 논문에서는 이러한 문제를 해결하기 위해서 관리적 선정을 실제조사에 원할하게 적용하기 위한 체계적 절차를 제시하고, 이 절차에 따라 다양한 측도를 사용하여 관리적 선정계획을 얻었다.

2장에서는 관리적 선정을 실제조사에 적용하기 위한 체계적 절차를 제시하고, Gabler (1987)가 제안하고 Adhikary(1996)가 사용한 최대근사크기비례추출계획(nearest proportional to size sampling design ; NPSSD)에 대해서는 3장에서 살펴보았다. 그리고 4장에서는 관리적 선정계획을 얻는 데 사용되는 여러 가지 측도들을 다루었으며, 5장에서는 본 논문에서 제시한 절차와 측도를 예제에 적용하여 측도들을 비교하고 관리적 선정계획을 얻었다.

2. 실용화를 위한 체계적 절차

모집단의 모든 조사단위들로부터 표본을 추출하는 경우 조사비용이나 관리 등을 고려할 때 일부의 표본은 사용하기가 바람직하지 않으므로 관리적 선정을 실제조사에 원활하게 적용하기가 쉽지 않다. 그러므로 조사비용, 조사시간 그리고 조사관리 등을 감안한 기준을 마련하여 일정 기준을 벗어나는 표본을 바람직하지 않은 표본으로 정하고, 그 기준을 벗어나는 정도에 따라 바람직하지 않은 표본들의 순위를 메겨서 제외시킬 표본을 정한 다음 표본추출확률을 조정해 주어야 한다. 물론 조사에 따라 선정 기준과 방법이 달라질 수 있으므로 각 조사에 적합한 기준을 별도로 마련해야 한다. 이때, 다양한 측도들을 사용하여 원래의 표본이 추출될 확률 $p_0(s)$에 가까운 새로운 표본추출확률, $p^*(s)$를 구한다.

본 연구에서는 기존에 사용된 측도들 외에 Adhikary, Kullback, 그리고 Metric 측도를 사용하여 관리적 선정계획을 세웠다. 관리적 선정을 실제조사에 적용할 때 편리하게 하기 위해서 다음의 절차를 생각한다.

1단계 : 층의 수에 비해 표본 수가 작은 경우나, 실시하고자 하는 조사가 비용이나, 조사관리 등을 고려할 때, 관리적 선정을 적용하는 것이 필요한 지의 여부를 결정한다.

2단계 : 관리적 선정을 사용할 때, 표본으로 선정되는 경우의 수를 줄여주는 것에 주안점을 둘지 또는 꼭 포함될 표본과 제외시킬 표본을 고려한 관리적 선정계획의 문제에 주안점을 둘지를 정한다.

3단계 : 표본으로 선정되는 경우의 수를 줄여주기 위해서는 1장에서 다룬 바와 같이 다양한 실험계획법을 적용하여 관리적 선정계획을 수립한다. 한편 바람직하지 않은 표본 중

에서 제외시킬 표본을 고려한 새로운 추출확률 $p^*(s)$을 얻기 위해서는 본 논문에서 다룬 측도를 사용한다.

4단계 : 실제조사에 앞서 바람직하지 않은 표본을 선정한 후 이들의 순위를 고려하여 제외시킬 표본을 정한다. 이때 순위는 비용이나 조사관리의 어려움 등을 고려한 별도의 기준을 마련하여 정한다.

5단계 : 3단계에서 정한 측도를 사용하여 Ryu와 Lee(1997)들이 제시한 알고리즘과 유사한 절차에 따라 새로운 관리적 선정계획 $p^*(s)$를 구한다.

6단계 : 다양한 측도를 적용하여 바람직하지 않은 표본이 추출될 확률을 구하고 각 측도에 따른 분산을 비교하여 최종적으로 사용할 측도를 정한다.

이와 같은 절차에 따라 다음 장에서 다룰 여러 가지 측도를 사용하여 바람직하지 않은 표본이 추출될 확률과 각 측도에 따른 분산을 비교하여 관리적 선정계획을 세운다.

3. 최대근사크기비례추출계획

모집단을 구성하고 있는 단위들의 추출확률이 $p_1, p_2, \cdots, p_N$, $\sum_{i=1}^N p_i = 1$ 일 때, 정해진 추출방법에 의해 모집단으로부터 크기 n인 표본 $s \in S$(S는 모든 표본들의 집합)가 추출될 확률을 $p_0(s)$라 하자. 그러나 임의의 표본 s가 추출될 확률, $p_0(s)$는 추출비용이나 조사 관리 등을 고려할 때 달라질 수 있다. 따라서 이러한 점을 고려하여 $p_0(s)$에 가장 근사한 $p^*(s)$를 구하기 위해서 Gabler(1987)는 표본추출확률 $p_0(s)$와 주어진 제약 조건들을 만족하는 임의의 표본추출계획 $\tilde{p}(s)$간의 다음과 같은 거리측도를 사용하였다.

$$\begin{aligned} D(p_0,\, \tilde{p}) &= E_{p_0}(\tilde{p}/p_0 - 1)^2 \\ &= \sum_{s \in T_{p_0}} \{\tilde{p}(s) - p_0(s)\}^2 / p_0(s) \end{aligned} \tag{3.1}$$

여기서 T_{p_0}는 $p_0(s) > 0$인 S에 있는 모든 표본 s의 집합이다. 주어진 $p_0(s)$하에서 $D(p_0, \tilde{p})$를 최소로하는 $p^*(s)$를 구한다. 즉, $D(p_0, \tilde{p})$를 최소로 하는 해는 아래의 제한조건에서 $p^*(s) = p_0(s) \sum_{i \in s} \lambda_i$ (단, $\pi_0 \lambda = \pi^*$)가 된다.

$$\text{제한조건} \; : \; \sum_{i \in s} \tilde{p}(s) = \pi_i^*, \quad i = 1, 2, \cdots, N$$

여기서 λ_i는 양수이고 π_i^*는 음이 아닌 알고 있는 수이며, $\sum_{i=1}^N \pi_i^* = n$이 된다.

한편 Adhikary(1996)는 실제 조사를 수행할 때, 조사비용이 많이 들거나 관리가 어려운 일부 바람직하지 않은 표본을 제외시키는 문제를 생각하고, Gabler(1987)가 제안한 최대근사크기비례추출계획을 이용하여 다음과 같은 표본추출계획을 제안하였다.

$$p_1(s) = \begin{cases} \dfrac{p_0(s)}{1 - \sum_{s \in S_0} p_0(s)}, & s \in S - S_0, \\ 0 \quad , & s \in S_0. \end{cases} \tag{3.2}$$

여기서 S_0는 바람직하지 않은 표본들의 집합으로 S_0에 속한 표본이 추출될 확률은 0이 된다. 그러면 $p_1(s)$는 더 이상 IPPS(inclusion probability proportional to size)설계가 아니다. 이러한 경우 가능한 한 $p_1(s)$에 가까우면서 원래의 표본추출계획인 $p_0(s)$와 동일한 IPPS 특성을 유지하는 표본추출계획 $p^*(s)$를 찾기 위해서 아래의 측도를 최소로 하는 최대근사크기비례추출계획(NPSSD)을 사용한다.

$$D(p_1, \tilde{p}) = E_{p_1}[\tilde{p}(s)/p_1(s) - 1]^2 = \sum_s \frac{\tilde{p}^2(s)}{p_1(s)} - 1 \tag{3.3}$$

여기서의 제한조건은 아래와 같다.

$$\sum_{i \in s} \tilde{p}(s) = \pi_i^* = \pi_i, \quad i = 1, 2, \cdots, N \tag{3.4}$$

$p^*(s)$가 $s \in S - S_0$에 대해 음수가 아니면 (3.3)식으로부터 $p^*(s) = (\sum_{i \in s} \lambda_i) p_1(s)$를 얻게 된다. 특히 Adhikary(1996)는 Farkar 보조정리 등을 이용하여 모든 단위들이 임의의 표본집합 S_0에 균일하게 분포하면 $\lambda_i \geq 0$이 되어 $p^*(s)(\geq 0)$를 얻을 수 있다는 것을 증명하였다.

Rao와 Nigam(1990, 1992)은 바람직하지 않는 표본의 추출확률을 최소로 하는 관리적 선정계획을 얻기 위해서, Sitter와 Skinner(1994), 그리고 Tiwari와 Nigam(1998)은 2차원으로 확장한 관리적 선정계획을 얻기 위해서 선형계획법을 사용하였다. 특히 Tiwari와 Nigam(1998)은 분산 추정량의 비음성(non-negativity)을 보장하기 위해서 기존의 제약조건에 $\pi_{ij} \leq \pi_i \pi_j$를 첨부하여 사용하였다. 그러나 관리적 선정계획을 세울 때, Ryu와 Lee(1997)가 지적한 바와 같이 중요하거나 비중이 큰 표본이 제외되는 경우가 발생하여 실제조사에 사용하기가 곤란하다. 따라서 Ryu와 Lee(1997)는 바람직한 표본은 반드시 포함시키면서 바람직하지 않은 표본들의 순위를 고려하여 가장 나쁜 표본들을 제외시키는 관리적 선정계획을 최대엔트로피를 사용하여 구하였다. 이는 결국 표본이 추출될 원래의 확률, $p_0(s)$가 변화하게 됨을 뜻한다. 그러므로 Gabler (1987)가 제시한 기존의 추출확률과의 차이를 크게 변화시키지 않으면서 중요하거나 꼭 제외시킬 표본을 고려한 관리적 선정 문제를 생각하는 것이 보다 현실적이다.

(3.1)식으로부터 유일한 해 $p^*(s)$는 λ가 음수가 아니어야 얻을 수 있으며, Adhikary(1996)는 모든 단위들이 표본에서 배제된 바람직하지 않은 표본들의 집합 S_0에 균일하게 분포하는 경우에 음수가 아닌 해를 얻을 수 있음을 보였다. 그러나 Gabler의 조건에서 λ가 음수인 경우가 생기며 또한 모든 단위들이 S_0에 균일하게 분포하지 않는 경우도 발생한다. 예를 들어 단위 i 등이 S_0에 포함되지 않는 경우는 실제 표본설계에서도 충분히 고려될 수 있는 것으로써 모집단으로부터 표본을 뽑을 때 반드시 포함되어야 하는 단위로 생각할 수 있다. Gabler 정리를 이용하는 경우 모든 단위들이 S_0에 균일하게 분포하는 경우는 물론이고 균일하게 분포하지 않는 경우에도 분산추정량의 비음성을 보장하지 못한다. 그리고 바람직하지 않은 표본 중에서 일부를 제외시키고 중요한 단위 일부를 표본에 반드시 포함시켜서 바람직하지 않은 표본 전체의 추출확률을 최소화시키는 방법을 모색하는 것이 바람직할 것이다.

앞서 설명한 거리 측도들은 모두 비선형함수이므로 비포함확률비례추출계획에서 단위 i 가 임의의 표본집합 S_0에 균등하게 분포하든 하지 않든 간에 관계없이 비선형계획법을 활용하여 관리적 선정계획 $p^*(s)$를 얻을 수 있다. 그리고 포함확률에 관한 부등식, $0 < \pi_{ij} \le \pi_i\pi_j$을 만족하는 표본추출계획도 세울 수 있다.

목적함수와 제약조건이 각각 (3.3)식과 (3.4)식으로 주어지는 Adhikary(1996)의 예제에 비선형계획법을 적용하면 Gabler 정리에 의해 얻어진 표본추출계획과 거의 동일한 결과를 얻는다.

4. KULLBACK 측도와 METRIC 측도

Shannon은 확률분포 $\mathbf{p} = (p_1, p_2, \cdots, p_n)$에 대한 불확실성의 척도로 다음과 같은 엔트로피 측도를 사용하였다(참고문헌[10]).

$$S(\mathbf{p}) = -\sum_{i=1}^{n} p_i \ln p_i \tag{4.1}$$

Ryu와 Lee(1997)는 주어진 제약조건하에서 Shannon의 엔트로피 측도를 사용하여 다음과 같은 목적함수를 최대로 하는 관리적 선정계획 $p^*(s)$를 얻었다.

$$-\sum_{s\in S_0^*} p_0(s) \ln p_0(s) \tag{4.2}$$

여기서 S_0^*는 모든 가능한 표본집합에서 바람직하지 않은 표본들의 집합, S_0중 일부를 제외시킨 표본집합이다.

한편 최소교차엔트로피 원리를 사용한 Kullback측도를 이용하여 관리적 선정계획을 세울 수 있는데, 이때 비포함확률추출계획의 목적함수를 다음과 같이 정의한다.

$$D(p_0,\ \tilde{p}) = \sum_{s\in S_0^*} \tilde{p}(s) \ln \frac{\tilde{p}(s)}{p_0(s)} \tag{4.3}$$

(4.3)식을 이용하면 Shannon의 엔트로피 원리를 이용하는 것과 달리 주어진 제약조건을 만족시키는 모든 표본추출계획 중 $p_0(s)$에 가장 가까운 표본추출계획 $p^*(s)$를 얻을 수 있다.

그리고 거리공간상에서 흔히 사용되는 Metric측도를 고려할 수 있는데, 이때는 비포함확률표본추출계획을 위해서 다음과 같은 목적함수를 사용한다.

$$D(p_0,\ \tilde{p}) = \left\{ \sum_{s\in S_0^*} (p_0(s) - \tilde{p}(s))^2 \right\}^{1/2} \tag{4.4}$$

Metric측도는 거리공간을 정의하는데 필요한 조건들인 비음성(non-negativity), 일치성(identity), 대칭성(symmetry), 삼각부등식(triangle inequality)을 만족하는 반면 Kullback 측도는 비음성과 일치성만을 만족한다.

5. 적 용

본 장에서는, 관리적 선정을 실제조사에 원할하게 활용하기 위해서 2장에서 제안한 체계적 절차와 본 논문에서 사용한 측도들을 예제에 적용하고 측도들을 비교하여 적절한 관리적 선정 계획을 얻는 문제를 다루고자 한다.

예제 5.1: $N = 7$, $n = 3$이며 p_i가 0.10, 0.12, 0.14, 0.15, 0.15, 0.16, 0.18인 모집단을 생각한다(Avadhani와 Sukhatme(1973)). 전체 35개의 표본 중 바람직하지 않은 표본 S_1이 다음과 같다고 하자.

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

그러나 실제조사시 조사비용이나 관리 등을 고려하여 이들 중 일부를 제외시키고 관리적 선정계획을 세울 때, 바람직하지 않은 표본들의 순위를 고려하여 표본설계를 하는 것이 바람직하다. 만약 제외시킬 표본 수가 최대 5개 일 때 가장 바람직하지 않은 5개의 표본 순위가 (1 3 7), (2 4 6), (1 4 7), (2 4 7), (1 2 3)이라고 가정하면 아래와 같이 하나의 표본을 제외시키는 경우에서 부터 5개까지 제외시키는 경우까지를 고려하여 관리적 선정계획을 세울 수 있다. 가장 바람직하지 않은 표본 5개를 제외시키는 경우 이들을 제외한 나머지 바람직하지 않은 표본들의 집합은 S_1^*가 된다.

i)	(1 3 7)				
ii)	(1 3 7),	(2 4 6)			
iii)	(1 3 7),	(2 4 6),	(1 4 7)		
iv)	(1 3 7),	(2 4 6),	(1 4 7),	(2 4 7)	
v)	(1 3 7),	(2 4 6),	(1 4 7),	(2 4 7),	(1 2 3)

부록의 표 1은 가장 바람직하지 않은 5개의 표본을 제외시켰을 때, 각 측도에 따른 관리적 선정계획이다. 본 논문에서는 SAS/OR을 사용하였으며, $0 < \pi_{ij} \le \pi_i \pi_j$를 만족시키기 위하여 제약조건 식에 이를 포함시켰다. 또한 여기서는 Adhikary, Kullback, Metric 측도 외에 Ryu와 Lee(1997)가 사용한 Shannon측도 그리고 Rao-Nigam(1990, 1992)의 방법도 사용하였다. Rao-Nigam(1990, 1992)의 방법을 비교한 이유는 이 방법이 바람직하지 않은 표본의 추출확률은 줄여주지만 동시에 바람직한 표본도 제외시키므로 본 연구의 의도를 잘 보여주기 때문이다.

표 5.1은 1개부터 5개까지 표본을 제외시키고 관리적 선정계획을 얻었을 때, 각 측도에 따라 얻어진 바람직하지 않은 표본 S_1^*가 뽑힐 확률 $\sum_{s \in S_1^*} p^*(s)$이다.

한편 관리적 선정계획을 얻는 데 사용되는 측도의 선정을 위해서 분산을 이용한다. 각 조사단위의 추출확률 $p_i = X_i / \sum X_i$가 보조변수로부터 얻어지고, 보조변수(X)와 연구변수(Y)와의 상관이 0.97, 0.96 그리고 0.93으로 가정했을 때 이로부터 연구변수에 대한 값을 역 추정한 3개의 가상 모집단 A, B, C를 고려한다. 그리고 비관리적 선정계획 $p_0(s)$을 얻기

모집단 A	0.20	0.40	0.40	0.50	0.55	0.60	0.70
모집단 B	0.25	0.30	0.35	0.40	0.40	0.50	0.60
모집단 C	0.65	0.90	0.95	1.05	1.15	1.00	1.40

위해 Sampford (1967)의 추출방법을 사용하였다. 각 모집단에 대해 측도에 따른 분산을 구한 것이 표 5.2에 있다.

표 5.1: 바람직하지 않은 표본 S_1^*가 추출될 확률

구분	Rao-Nigam	Shannon	Adhikary	Kullback	Metric
i)	0.275445	0.261692	0.261637	0.261719	0.262633
ii)	0.275445	0.249076	0.249031	0.249025	0.249170
iii)	0.275445	0.232974	0.233017	0.235844	0.228750
iv)	0.264590	0.207078	0.206911	0.206783	0.195804
v)	0.294381	0.193228	0.193694	0.192487	0.184389

표 5.2: 각 측도에 따른 분산

구분		Rao-Nigam	Shannon	Adhikary	Kullback	Metric
i)	모집단 A	0.073519	0.072154	0.072061	0.072074	0.073150
	모집단 B	0.021546	0.022181	0.022163	0.022152	0.022316
	모집단 C	0.066468	0.068286	0.068333	0.068301	0.068383
ii)	모집단 A	0.073519	0.073227	0.073128	0.073158	0.073869
	모집단 B	0.021546	0.022835	0.022806	0.022807	0.022865
	모집단 C	0.066468	0.069160	0.069156	0.069177	0.069438
iii)	모집단 A	0.073519	0.075220	0.075120	0.075157	0.076334
	모집단 B	0.021546	0.023619	0.023605	0.023600	0.023677
	모집단 C	0.066468	0.071910	0.072002	0.071952	0.072094
iv)	모집단 A	0.073518	0.076344	0.076284	0.076310	0.076925
	모집단 B	0.021546	0.024877	0.024869	0.024869	0.021893
	모집단 C	0.066467	0.069434	0.069972	0.069471	0.069522
v)	모집단 A	0.073519	0.072007	0.071626	0.071841	0.073025
	모집단 B	0.021546	0.023828	0.023780	0.023792	0.024012
	모집단 C	0.066468	0.070419	0.071187	0.070489	0.070385

예제 5.2: $N = 7, n = 3$이며 p_i가 0.12, 0.14, 0.15, 0.15, 0.14, 0.17, 0.13인 Adhikary(1996)의 자료를 이용한다. 바람직하지 않은 표본의 개수와 순위도 모두 예제 5.1에서와 같다고 가정한다. 표 5.3은 표 5.1에서와 같이 바람직하지 않은 표본 S_1^*가 뽑혀질 확률 $\sum_{s\in S_1^*} p^*(s)$이다.

모집단 D	0.20	0.50	0.75	0.80	0.60	1.15	0.40
모집단 E	0.40	0.45	0.60	0.70	0.50	0.90	0.35
모집단 F	0.80	0.90	0.95	1.05	1.05	1.20	0.85

예제 5.1과 같이 측도의 선정을 위해서 분산을 이용할 때, 보조변수와 연구변수와의 상관을 0.99, 0.94 그리고 0.91으로 가정하고 이로부터 연구변수에 대한 값을 역 추정한 3개의 가상 모집단 D, E, F를 고려한다. 각 모집단에 대해 측도에 따른 분산이 표 5.4에 있다.

표 5.3: 바람직하지 않은 표본 S_1^*가 추출될 확률

구분	Rao-Nigam	Shannon	Adhikary	Kullback	Metric
i)	0.174717	0.276769	0.276751	0.276823	0.277159
ii)	0.281813	0.256998	0.257063	0.257062	0.256991
iii)	0.281813	0.241747	0.241835	0.241849	0.241031
iv)	0.086819	0.216687	0.216805	0.216811	0.214411
v)	0.086819	0.197750	0.199362	0.197904	0.196288

표 5.4: 각 측도에 따른 분산

구분		Rao-Nigam	Shannon	Adhikary	Kullback	Metric
i)	모집단 D	0.516347	0.501853	0.501298	0.501501	0.505562
	모집단 E	0.154547	0.151377	0.151248	0.151294	0.152288
	모집단 F	0.032938	0.031636	0.031618	0.031636	0.031729
ii)	모집단 D	0.516343	0.491516	0.491158	0.491116	0.489502
	모집단 E	0.154546	0.147317	0.147219	0.147225	0.146426
	모집단 F	0.032938	0.033100	0.033139	0.033101	0.033244
iii)	모집단 D	0.516343	0.480127	0.479885	0.479768	0.482445
	모집단 E	0.154546	0.148954	0.148929	0.148867	0.148805
	모집단 F	0.032938	0.033485	0.033556	0.033485	0.033715
iv)	모집단 D	0.516347	0.489157	0.488382	0.488809	0.495413
	모집단 E	0.154547	0.150202	0.150496	0.150114	0.151627
	모집단 F	0.032939	0.034050	0.034164	0.034051	0.034429
v)	모집단 D	0.516347	0.480060	0.477071	0.479722	0.489297
	모집단 E	0.154547	0.150167	0.150057	0.150080	0.152329
	모집단 F	0.032939	0.031762	0.031635	0.031767	0.032078

표 5.1, 표 5.3을 살펴보면 제외시키는 표본들의 개수가 증가할수록 Metric측도가 다른 측도들에 비해 바람직하지 않은 표본들의 추출확률이 더 작아진다는 것을 알 수 있다. 부록의 표 1과 표 2에서도 Ryu와 Lee(1997)가 지적한 바와 같이 Rao-Nigam의 경우는 표본을 제외시키는 것이 달라짐에 따라 표본에 추출될 수 있는 것과 없는 것이 전적으로 변하므로

이러한 경우에 사용하기가 곤란하게 된다. 또한 표 5.2, 표 5.4로부터 보조변수와 연구변수의 상관이 높을수록 Adhikary측도가 분산이 대체로 작게 된다. 이상으로 볼 때, 바람직하지 않은 일부의 표본을 제외하고 관리적 선정계획을 세우고자 할 때는 본 논문에서 제시한 절차에 따라 제외시킬 표본을 정한 후 바람직하지 않은 표본의 추출확률과 분산을 고려하여 필요한 측도를 사용한다.

6. 맺는 말

관리적 선정을 실제조사에 적용함에 있어서 유의표본집합을 만드는 일이 매우 번거롭고 복잡하다. 그리고 실제 조사를 수행할 때 조사비용과 관리 등의 관점에서 일부의 표본을 제외시키는 경우가 생긴다. 이러한 경우 기존의 관리적 선정방법으로는 이러한 요구를 충족시킬 수 없고 오히려 중요한 표본들이 제외되는 경우가 발생한다. 따라서 본 연구에서는 중요하지 않은 표본들을 순차적으로 제외하고 관리적 선정계획을 세우는 문제를 다루었다. 그리고 관리적 선정을 실제조사에 효율적으로 사용하기 위한 체계적인 절차를 제시하고, 이 절차에 따라 Adhikary, Kullback, 그리고 Metric 측도 등을 사용하여 관리적 선정계획을 세웠다.

본 연구를 통해서 볼 때 관리적 선정에 의한 추출확률이 원래의 추출확률과 차이가 나지 않음을 알 수 있다. 그리고 제외시키는 표본들의 개수가 증가할수록 Metric측도가 다른 측도들에 비해 바람직하지 않은 표본들의 추출확률이 더 작고, 보조변수와 연구변수의 상관이 높을수록 Adhikary측도가 분산이 대체로 작게 됨을 알 수 있다. 따라서 이러한 점을 고려하면 실제조사에 관리적 선정을 사용할 때 선택의 폭을 넓힐 수 있다. 그러나 아직도 이들 문제를 동시에 만족시켜주는 관리적 선정을 수행하기가 곤란하므로 이점에 대한 연구는 향후에 다루고자 한다.

부록

표 1: 예제 5.1에서 5개의 표본을 제외시킨 경우의 관리적 선정계획

s	Rao-Nigam	Shannon	Adhikary	Kullback	Metric
$p1(1,2,3)^*$					
$p2(1,2,4)$	0.075785	0.025819	0.024333	0.026077	0.030049
$p3(1,2,5)$		0.016999	0.017595	0.017114	0.016819
$p4(1,2,6)^*$		0.020774	0.020899	0.020570	0.021492
$p5(1,2,7)$		0.033005	0.030652	0.032262	0.033573
$p6(1,3,4)$	0.016748	0.025514	0.025293	0.025430	0.027351
$p7(1,3,5)$	0.051024	0.016798	0.017012	0.016689	0.014121
$p8(1,3,6)^*$	0.023519	0.020529	0.021014	0.020403	0.019498
$p9(1,3,7)^*$					
$p10(1,4,5)$	0.005595	0.023957	0.024345	0.023887	0.023713
$p11(1,4,6)$	0.001353	0.029278	0.029518	0.029338	0.029446
$p12(1,4,7)^*$					
$p13(1,5,6)$		0.019276	0.019262	0.019254	0.016216
$p14(1,5,7)$	0.042862	0.030625	0.031516	0.030891	0.030455
$p15(1,6,7)$	0.083114	0.037426	0.038560	0.038085	0.037267
$p16(2,3,4)$	0.009167	0.032507	0.032120	0.032386	0.033017
$p17(2,3,5)$		0.021402	0.021626	0.021255	0.019788
$p18(2,3,6)^*$	0.083233	0.026155	0.026810	0.026087	0.025853
$p19(2,3,7)$	0.020008	0.041554	0.041788	0.041726	0.040796
$p20(2,4,5)$	0.037411	0.030523	0.031007	0.030502	0.030053
$p21(2,4,6)^*$					
$p22(2,4,7)^*$					
$p23(2,5,6)$		0.024559	0.024564	0.024591	0.023244
$p24(2,5,7)$	0.084952	0.039018	0.039989	0.039290	0.038908
$p25(2,6,7)$	0.049444	0.047684	0.048615	0.048140	0.046408
$p26(3,4,5)$		0.030163	0.030612	0.030253	0.030688
$p27(3,4,6)$		0.036862	0.037832	0.037074	0.037815
$p28(3,4,7)^*$	0.120705	0.058565	0.058295	0.058789	0.054914
$p29(3,5,6)$	0.052029	0.024269	0.021580	0.024331	0.024585
$p30(3,5,7)$	0.043567	0.038558	0.038425	0.038582	0.041685
$p31(3,6,7)$		0.047122	0.047593	0.046995	0.049889
$p32(4,5,6)$	0.104136	0.034613	0.034577	0.034758	0.036251
$p33(4,5,7)$	0.012175	0.054991	0.055391	0.054867	0.054071
$p34(4,6,7)^*$	0.066924	0.067205	0.066676	0.066638	0.062632
$p35(5,6,7)$	0.016248	0.044246	0.042499	0.043734	0.049402
$\sum_{s\in S_1^*} p^*(s)$	0.294381	0.193228	0.193694	0.192487	0.184389

표 2: 예제 5.2에서 5개의 표본을 제외시킨 경우의 관리적 선정계획

s	Rao-Nigam	Shannon	Adhikary	Kullback	Metric
$p1(1,2,3)^*$					
$p2(1,2,4)$	0.012726	0.038568	0.036784	0.038494	0.038828
$p3(1,2,5)$		0.023892	0.024424	0.023865	0.023917
$p4(1,2,6)^*$		0.036767	0.037114	0.036762	0.036265
$p5(1,2,7)$	0.100385	0.029912	0.028951	0.029961	0.032520
$p6(1,3,4)$		0.034199	0.034343	0.034203	0.034815
$p7(1,3,5)$	0.099690	0.021185	0.021194	0.021154	0.019550
$p8(1,3,6)^*$	0.023407	0.032602	0.033237	0.032700	0.032969
$p9(1,3,7)^*$					
$p10(1,4,5)$		0.026161	0.026609	0.026121	0.026322
$p11(1,4,6)$	0.110371	0.040260	0.040651	0.040377	0.039741
$p12(1,4,7)^*$					
$p13(1,5,6)$	0.010347	0.024940	0.024601	0.024945	0.023759
$p14(1,5,7)$	0.003074	0.020290	0.020684	0.020330	0.020014
$p15(1,6,7)$		0.031224	0.031407	0.031089	0.031301
$p16(2,3,4)$	0.121617	0.045813	0.045440	0.045887	0.044138
$p17(2,3,5)$		0.028379	0.028826	0.028405	0.028179
$p18(2,3,6)^*$	0.025821	0.043673	0.044456	0.043644	0.043711
$p19(2,3,7)$		0.035531	0.035400	0.035490	0.035751
$p20(2,4,5)$	0.013094	0.035046	0.035570	0.035074	0.034951
$p21(2,4,6)^*$					
$p22(2,4,7)^*$					
$p23(2,5,6)$	0.122542	0.033409	0.033577	0.033444	0.033807
$p24(2,5,7)$		0.027180	0.027561	0.027128	0.027267
$p25(2,6,7)$	0.023815	0.041829	0.041896	0.041847	0.040667
$p26(3,4,5)$	0.010837	0.031075	0.031352	0.031137	0.031632
$p27(3,4,6)$	0.020259	0.047823	0.048236	0.047606	0.048235
$p28(3,4,7)^*$	0.007451	0.038906	0.038758	0.038934	0.038855
$p29(3,5,6)$	0.013281	0.029624	0.027203	0.029615	0.031204
$p30(3,5,7)$	0.023628	0.024101	0.024027	0.024082	0.023245
$p31(3,6,7)$	0.104010	0.037089	0.037528	0.037143	0.037716
$p32(4,5,6)$	0.026008	0.036583	0.036383	0.036568	0.037977
$p33(4,5,7)$	0.097498	0.029763	0.030078	0.029736	0.030017
$p34(4,6,7)^*$	0.030140	0.045802	0.045797	0.045864	0.044488
$p35(5,6,7)$		0.028373	0.027911	0.028396	0.028160
$\sum_{s\in S_1^*} p^*(s)$	0.086819	0.197750	0.199362	0.197904	0.196288

참고문헌

[1] Adhikary, A. K. (1996). On non-negativity of the nearest proportional to size sampling design. *Communications in Statistics : Theory and Methods.* Vol. 25, 1757-1768.

[2] 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.

[3] 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.

[4] 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.

[5] Gabler, S. (1987). The nearest proportional to size sampling design. *Communications in Statistics: Theory and Methods.* Vol. 16, No.4, 1117-1131.

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

[7] 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.

[8] 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.

[9] Hess, I., Riedel, D. C., and Fitzpatrick, T. B. (1975). *Probability sampling of hospitals and patients*, 2nd ed. Ann Arbor, Michigan: Health Administration Press.

[10] Kapur, J. N. and Kesavan, H. K. (1992). *Entropy Optimization Principles with Applications*, Academic Press.

[11] Lin, Ting-kwong (1992). Some improvements on an algorithm for controlled selection. *ASA 1992 Proceeding of the Section on Survey Research Methods.* 407-410.

[12] Rao, J. N. K. and Nigam, A. K. (1990). Optimal controlled sampling designs. *Biometrika.* Vol. 77, 807-814.

[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 selection. *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] Sitter, R. R. and Skinner, C. J. (1994). Multi-way stratification by linear programing. *Survey Methodology*. Vol. 20, No. 1, 65-73.

[18] Tiwari, N. and Nigam, A. K. (1998). On two-dimensional optimal controlled selection. *Journal of Statistical Planning and Inference*. Vol. 69, 89-100.

[19] Winston, W. L. (1994). *Operations research*, 3rd ed. Duxbury Press, California.

[1998년 9월 접수, 1998년 11월 최종수정]

A Systematic Method for Applying the Controlled Selection to Practical Survey *

Jea-Bok Ryu[1)]Seung-Joo Lee[2)] Sun-Woong Kim[3)]

ABSTRACT

When the sample size is less than the total number of cells in stratified random sampling, the survey cost increases and supervision is difficult since sampling units are too widespread, and the muti-purpose surveys are need, the controlled selection is used. However, there are some problems in applying the controlled selection to practical surveys. In this paper, we suggest the systematic procedures for applying the controlled selection to practical survey using the several measures. And we examine the adaptability of using systematic procedures and compare the measures through numerical examples.

* The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997.

1) Professor, Department of Applied Statistics, Chongju University, 36 Naedok-dong, Sangdanggu Chongju-si, Chungbuk, 360-764, Korea.

2) Assistant Professor, Department of Applied Statistics, Chongju University, 36 Naedok-dong, Sangdanggu Chongju-si, Chungbuk, 360-764, Korea.

3) Ph. D. Candidate, Department of Statistics, Dongguk University, 26 Pildong 3-ga Chunggu Seoul, 100-715, Korea.