응용통계연구 제 13권 1호, 2000년, pp. 145-161

주변제약을 갖는 이차원 관리적 선정

김종호[1] 류제복[2] 김선웅[3]

요 약

조사비용을 증대시키거나 조사실시에 어려움을 주는 바람직하지 않은 표본들의 추출확률을 줄여주기 위해서 Goodman과 Kish(1950)는 관리적 선정 방법을 제시하였다. 층화추출에서 표본의 수가 셀의 수보다 작은 경우 표본 배분에 문제가 발생한다. 이러한 문제를 해결하기 위해서 관리적 선정을 적용할 수 있는데 Causey 등(1985)은 수송이론을 이용한 알고리즘을 제안하였고 Sitter와 Skinner(1994), Tiwari와 Nigam(1998)은 선형계획법을 이용하였다. 본 연구에서는 기존 방법들의 문제점들을 다루었으며 추출방법의 이론적 측면을 보완하기 위하여 표본들의 적합성을 고려한 관리적 선정을 제안한다. 아울러 분산을 최소화시키는 관리적 선정 방법과 통합 관리적 선정 방법도 제시하였다.

주요용어: 관리적 선정, 바람직하지 않은 표본, 이원층화, 비선형계획.

1. 서 론

모든 가능한 표본들을 동일한 확률로 추출하는 단순임의추출법(simple random sampling)을 표본조사에 그대로 사용할 경우 연구의 특성상 별로 중요하지 않은 단위들을 포함하고 있는 표본들이 뽑힐 수 있으며 특히 추출단위들이 너무 멀리 퍼져 있을 때 조사비용이 증가할 뿐만 아니라 실사의 감독과 조직이 어려워진다. Goodman과 Kish(1950)는 조사비용을 증가시키고 조사의 관리감독에 어려움을 주는 표본들을 바람직하지 않은 표본(non-preferred sample)이라고 하였다. 이러한 표본들의 추출을 관리하기 위해서 층화추출법, 집락추출법, 또는 다단추출법 등을 사용하나 추정치의 정도가 떨어질 수 있다. Goodman과 Kish(1950)는 바람직하지 않은 표본들의 추출확률을 줄여 주기 위하여 관리적 선정(controlled selection)이라는 새로운 방법을 처음으로 제안하였고 그 이후로 많은 학자들에 의하여 이에 대한 연구들이 수행되어 왔다.

관리적 선정은 층화와 상반되는 것은 아니며 오히려 층화를 한 후 이를 더 쉽게 행할 수 있으므로 층화시 관리적 선정을 적용할 수 있다.

이원층화(two-way stratification)시 표본의 수가 셀의 수보다 작아 표본 배분의 문제가 발생하므로 Bryant등(1960)은 이에 대한 간단한 관리적 선정 방법을 제시하였다. 한편 Jessen(1969)은 층화를 고려하지 않는 경우 추출단위의 보조정보를 이용하여 표본을 추출

1) (100-715) 서울 중구 필동 3가 26번지, 동국대학교 통계학과, 교수
2) (360-764) 충청북도 청주시 상당구 내덕동 36번지, 청주대학교 응용통계학과, 교수
E-mail: jbryu@chongju.ac.kr
3) (100-715) 서울 중구 필동 3가 26번지, 동국대학교 통계학과, 강사
E-mail: KSWYH@chollian.net

할 수 있는 4가지 방법을 제안하였으나 '방법 2'와 '방법 3'은 절차가 복잡하고 '방법 4'는 해를 얻기 위하여 임의적인 절차를 반복해야 하며 표본의 크기가 2인 경우에만 사용할 수 있다. Jessen(1970)은 '방법 2'와 '방법 3'을 층화를 고려하는 경우에 적용하였으나 이원층화인 경우에도 해를 얻지 못하는 경우가 발생하였다. Causey등(1985)은 표본의 개수를 줄이는 관리적 선정 계획을 세우기 위하여 수송이론을 이용한 알고리즘을 제안하였으며 Sitter와 Skinner(1994)는 다중층화(multiple stratification)시 관리적 선정 계획을 얻기 위하여 선형계획법을 활용하는 방법을 제안하였다. 그리고 Tiwari와 Nigam(1998)은 이원층화에 있어서 선형계획법을 이용하여 바람직하지 않은 표본들의 추출확률을 최소화 시킬 수 있는 방법을 제시하였다.

본 논문에서는 이원층화시 얻을 수 있는 모든 가능한 표본들 중 원래 자료에 가장 가까운 표본들 즉 최적 컨트롤 라운딩인 표본들이 뽑힐 확률을 높여 주면서 바람직하지 않은 표본의 추출확률을 최소화시키는 관리적 선정 계획을 얻는 방법을 제안하였다. 또한 분산을 최소화시키기 위하여 비선형계획법을 이용하여 표본추출계획을 얻는 방법과 바람직하지 않은 표본들의 추출확률은 줄이고 최적 컨트롤 라운딩인 표본들의 추출확률은 높이면서 가능한 한 분산을 최소화시킬 수 있는 통합 관리적 선정 방법도 제안하였다.

2. 최적 컨트롤 라운딩

2개의 층화변수에 의해 R행과 C열로 분할되는 2차원 배열 모집단을 고려하자. Y_{rc}는 r번째 행과 c번째 열이 서로 교차되는 셀, 즉 rc셀의 Y의 특성값을 나타내며 X_{rc}는 Y_{rc}에 관한 크기의 측도로서 Y_{rc}와 양의 상관관계가 있는 보조변수로서 $X = \sum_{r=1}^{R}\sum_{c=1}^{C} X_{rc}$이며 A_{rc}는 rc셀의 상대적 크기 측도로서 $A_{rc} = X_{rc}/X, \sum_{r=1}^{R}\sum_{c=1}^{C} A_{rc} = 1$ 이다.

이때 $R \times C$개의 셀로 부터 크기 n인 표본을 뽑는다고 할 때 nA_{rc}은 rc셀로 부터 추출되는 표본 단위의 기대수를 나타낸다. 단 제약조건들은 다음과 같다.

$$\begin{array}{lll} (i) & \text{셀 제약조건}(\textit{cell constraint}) & : \; E\{n_{rc}\} = nA_{rc}, \\ (ii) & \text{정수 제약조건}(\textit{integer constraint}) & : \; |n_{rc} - nA_{rc}| < 1, \\ (iii) & \text{주변 제약조건}(\textit{marginal constraints}) & : \; |n_{\cdot c} - nA_{\cdot c}| < 1 \\ & & \;\;\; |n_{r\cdot} - nA_{r\cdot}| < 1 \end{array} \tag{2.1}$$

여기서 $A_{\cdot c} = \sum_{r=1}^{R} A_{rc},\ A_{r\cdot} = \sum_{c=1}^{C} A_{rc}$이다.

$A_{\cdot\cdot} = \sum_{r=1}^{R}\sum_{c=1}^{C} A_{rc},\ nA_{\cdot\cdot} = n$이므로 앞으로 편의상 다음과 같이 표시하자.

$$\pi_{rc} = nA_{rc},\ \pi_{\cdot c} = nA_{\cdot c} = n\sum_{r=1}^{R} A_{rc},\ \pi_{r\cdot} = nA_{r\cdot} = n\sum_{c=1}^{C} A_{rc},\ \pi_{\cdot\cdot} = nA_{\cdot\cdot} = n\sum_{r=1}^{R}\sum_{c=1}^{C} A_{rc}$$

컨트롤 라운딩(controlled rounding)은 Cox와 Ernst(1982)가 수송이론을 이용하여 제안한 것으로서 주어진 자료를 쉽게 읽을 수 있도록 하거나 분석의 유용성을 높이기 위하여 흔

히 사용되는 방법으로서 원래 자료의 왜곡을 최소화시키면서 실수값들이 정수값으로 바꾸는 통계적 문제에도 적용될 수 있다. 표본조사에서 층화시 셀의 표본단위의 기대수가 주로 실수값이므로 표본추출을 위해서 이 기법을 활용할 수 있다.

각 셀의 π_{rc}에 대하여 다음과 같은 $(R+1)\times(C+1)$배열로 만들어 이를 Π로 표시하자.

$$\begin{matrix} (\pi_{rc})_{R\times C} & (\pi_{r\cdot})_{R\times 1} \\ (\pi_{\cdot c})_{1\times C} & (\pi_{\cdot\cdot})_{1\times 1} \end{matrix} \tag{2.2}$$

각 π_{rc}에 대한 컨트롤 라운딩 $R(\pi_{rc})$은 각 셀로 부터 추출되는 표본 단위의 개수로서 다음을 만족시켜야 하므로 영점 제한 컨트롤 라운딩(zero-restricted controlled rounding)이라고도 한다.

$$|R(\pi_{rc}) - \pi_{rc}| < 1 \tag{2.3}$$

배열 Π의 모든 가능한 컨트롤 라운딩 $R(\Pi)$ 역시 (2.2)식과 같은 배열을 만족해야 하며 이들 중 다음 측도에 대하여 최소값을 갖는 것이 최적 컨트롤 라운딩(optimal controlled rounding)이 된다.

$$D[R(\pi_{rc}),\, \pi_{rc}] = \max\{|R(\pi_{rc}) - \pi_{rc}| : 1 \le r \le R\,,\, 1 \le c \le C\} \tag{2.4}$$

3. 이원층화시 기존의 관리적 선정

3.1. 수송이론을 이용한 관리적 선정

Causey, Cox와 Ernst(1985)는 앞서 언급한 (2.1)식의 제약조건들을 만족시키는 표본추출계획을 얻기 위하여 다음과 같은 알고리즘을 제안하였다.

단계1) 배열 Π로 부터 최적 컨트롤 라운딩인 첫 번째 표본 s 즉 $R(\pi_{rc})_{R\times C}$을 얻는다.

단계2) 이 표본에 대해 $D[R(\pi_{rc}), \pi_{rc}]$을 계산하여 추출확률 $p(s) = 1 - D[R(\pi_{rc}), \pi_{rc}]$를 구한다.

단계3) 새로운 $\pi'_{rc} = R(\pi_{rc}) + [\pi_{rc} - R(\pi_{rc})]/D[R(\pi_{rc}), \pi_{rc}]$을 얻어 이로부터 최적 컨트롤 라운딩인 두 번째 표본 즉 $R(\pi'_{rc})_{R\times C}$을 얻고 다음과 같은 추출확률 $p'(s)$을 구한다.

$$p'(s) = (1 - p(s))(1 - D'[R(\pi'_{rc}), \pi'_{rc}])$$

여기서 $D'[R(\pi'_{rc}), \pi'_{rc}] = \max\{|R(\pi'_{rc}) - \pi'_{rc}| \,:\, 1 \le r \le R,\ 1 \le c \le C\}$이다.

단계4) 이 과정은 $D'[R(\pi'_{rc}), \pi'_{rc}]$이 0이 될 때까지 단계2), 단계3)을 반복하며 단계3)의 $p'(s)$에 관한 식에서 $p(s)$는 누적되어 값이 계산된다.

이 알고리즘은 $R\times C$의 크기에 상관없이 $(\pi_{rc})_{R\times C}$로 부터 얻어질 수 있는 모든 가능한 표본의 수가 비교적 큰 경우에도 쉽게 표본추출계획을 얻을 수는 있는 반면 단계1)과 단

계3)에서 최적 컨트롤 라운딩인 표본을 얻기 위하여 계속해서 Glover 등(1974)이 제안한 수송문제 알고리즘을 이용해야 하므로 실행이 복잡하다. 이때 최적 컨트롤 라운딩인 표본들이 여러 개 존재할 수 있지만 단계1)에서 얻어진 첫 번째 표본만이 원래 자료로부터 얻어질 수 있는 유일한 최적 컨트롤 라운딩이 된다.

3.2. 선형계획법을 이용한 관리적 선정

Sitter와 Skinner(1994)는 Rao와 Nigam(1990,1992)이 관리적 선정 계획을 얻기 위하여 이용한 선형계획법을 층화에 적용하여 모든 가능한 $\binom{RC}{n}$개의 표본들의 집합 S 중 $\binom{RC}{\tilde{n}}$개의 표본들의 집합 $\tilde{S}$만을 고려하여 표본추출계획을 세웠다. 여기서 $\tilde{n} = n - \sum_r \sum_c I_{rc}$, $I_{rc} = [\pi_{rc}]$이며 단 $[\pi_{rc}]$는 π_{rc}보다 작은 가장 큰 정수값이다.

$R(\pi_{rc}) - I_{rc} = \tilde{R}(\pi_{rc})$, $\pi_{rc} - I_{rc} = k_{rc}$라 할 때 관리적 선정 계획을 세우기 위한 목적함수와 제약조건은 각각 식(3.1), (3.2)와 같으며 해를 얻기 위하여 선형계획법을 이용한다.

$$\phi = \sum_{s\in\tilde{S}}\Big[\sum_{r=1}^{R}[\tilde{R}(\pi_{r\cdot}) - k_{r\cdot}]^2 + \sum_{c=1}^{C}[\tilde{R}(\pi_{\cdot c}) - k_{\cdot c}]^2\Big]p(s), \tag{3.1}$$

$$\sum_{s\in\tilde{S}} \tilde{R}(\pi_{rc})p(s) = k_{rc} \tag{3.2}$$

여기서 $k_{r\cdot} = \sum_{c=1}^{C} k_{rc}$, $k_{\cdot c} = \sum_{r=1}^{R} k_{rc}$이다.

이 방법의 문제점은 선형계획법의 계산상의 어려움을 줄이기 위하여 $\binom{RC}{\tilde{n}}$개로 구성되는 표본들의 집합 $\tilde{S}$만을 고려하여 표본추출계획을 얻으려 하지만 실제로는 그 개수가 표본들의 집합 S를 구성하는 $\binom{RC}{n}$와 동일하다. 또한 주변 제약조건을 가중치로 사용함으로서 표본들의 최적성을 충분히 고려하지 못하였으며 정수주변을 갖는 문제에 대해서는 가중치를 사용할 수 없게 된다.

한편 Tiwari와 Nigam(1998)은 모든 가능한 표본들의 집합 S 중 바람직하지 않은 표본들의 집합 S_1이 추출될 확률을 최소화시키는 관리적 선정을 제안하였다.

이 방법은 다음과 같은 목적함수를 고려한다.

$$\phi = \sum_{s\in S_1} p(s) \tag{3.3}$$

제약조건은 다음과 같으며 위의 목적함수를 최소화시키는 표본추출계획을 선형계획법을 이용하여 얻을 수 있다.

$$\begin{aligned} &(i) && \sum_{s\in S} p(s) = 1 \\ &(ii) && \sum_{rc\in s} p(s) = \pi_{rc} \\ &(iii) && p(s) \ge 0 \end{aligned} \tag{3.4}$$

이 방법으로 이원층화시 바람직하지 않은 표본들의 추출확률을 줄이는 관리적 선정 계획을 얻을 수 있으나 Sitter와 Skinner(1994)의 방법과 마찬가지로 표본들의 적합성을 고려하지는 못하였다.

4. 제안 방법

앞서 설명한 바와 같이 Causey 등(1985)이 제안한 수송이론을 이용한 관리적 선정 방법은 실행이 복잡하며 Sitter와 Skinner(1994)의 방법은 주변이 정수인 문제에 대해서는 가중치를 적용할 수 없어 표본들의 적합성을 고려할 수 없다. 또한 Tiwari와 Nigam(1998)의 방법도 표본들의 적합성을 전혀 고려하지 않았다.

먼저 본 논문에서는 이원층화시 관리적 선정에 있어서 실행의 복잡성을 줄이고 표본들의 적합성을 충분히 감안하기 위하여 최적성을 고려한 관리적 선정 방법을 제시한다. 이 방법을 사용하면 선형계획법을 이용할 수 있어 실행 절차를 간소화시킬 수 있다. 또한 원래의 모집단 자료 즉 $(\pi_{rc})_{R\times C}$에 가능한 한 가까운 표본, 즉 모든 가능한 표본들 중 최적 컨트롤 라운딩인 표본들의 추출확률을 높일 수 있으며 각 표본들의 적합성을 고려한 표본추출계획을 얻을 수 있다. 아울러 바람직하지 않은 표본들의 추출확률을 줄일 수 있다.

다음 이원층화시 추정량의 분산을 줄일 수 있는 표본추출계획을 얻기 위하여 분산 최소화를 위한 관리적 선정 방법을 고려한다. 이를 위하여 비선형 목적함수를 개발하였으며 비선형계획법을 이용하여 비교적 쉽게 해를 얻을 수 있다.

마지막으로 표본의 최적성, 바람직하지 않은 표본들의 추출확률, 추정량의 분산을 동시에 고려할 수 있는 통합 관리적 선정 방법을 제안한다. 이 방법을 이용함으로서 표본조사의 이론적인 측면과 실용적인 측면을 함께 고려할 수 있는 표본추출계획을 세울 수 있다.

4.1. 최적성을 고려한 관리적 선정

최적 컨트롤 라운딩인 표본들의 추출확률은 높이면서 바람직하지 않은 표본들의 추출확률을 줄이는 관리적 선정 계획을 얻기 위해 다음과 같은 방법을 제시한다.

실제 조사에서 있어 이원 배열 모집단의 각 셀의 추출단위의 기대수 π_{rc}가 1보다 작은 경우도 있을 수 있으나 $\pi_{rc} \geq 1$인 것이 보다 일반적인 경우라 할 수 있다. 이러한 경우 rc셀에서 $[\pi_{rc}]$개 단위는 반드시 추출되므로 이 단위들은 제외시켜 $\pi_{rc} < 1$로 만든 후 표본추출계획을 세운다. 이 때 rc셀의 추출단위의 기대수는 π^*_{rc}, 표본의 크기는 n^*라 하자.

그런데 π^*_{rc}는 rc셀의 단위의 포함확률을 나타내므로 가능한 모든 표본들 중 최적 컨트롤 라운딩인 표본들의 추출확률을 높이면서 바람직하지 않은 표본들의 추출확률을 최소화시키는 방법을 병행해야 할 것이다. 이를 위하여 다음과 같은 목적함수를 고려한다.

$$\phi_1 = \sum_{s\in S} M[R(\pi^*_{rc}), \pi^*_{rc}]p(s) + \sum_{s\in S_1} p(s) \tag{4.1}$$

여기서 메트릭 측도 $M[R(\pi^*_{rc}), \pi^*_{rc}] = \left[\sum_{rc}(R(\pi^*_{rc}) - \pi^*_{rc})^{2m}\right]^{1/2m}$ 이고 $m \geq 1$이다.

이 목적함수를 최소화시킬 때 거리 측도인 메트릭 측도로 인하여 결과적으로 얻어지는 표본추출계획에 있어 최적 컨트롤 라운딩인 표본들의 추출확률은 커지게 되고 바람직하지 않은 표본들의 집합의 추출확률인 $\sum_{s\in S_1} p(s)$은 작아지게 된다.

메트릭 측도는 다음과 같은 중요한 특성을 갖는다.

$$M[R(\pi^*_{rc}), \pi^*_{rc}] = M[R(\pi_{rc}), \pi_{rc}] \tag{4.2}$$

한편 제약조건은 다음과 같으며 목적함수 ϕ_1을 최소화시키는 관리적 선정 계획을 얻기 위하여 선형계획법을 이용할 수 있다.

$$\sum_{s\ni rc} p(s) = \pi^*_{rc} \tag{4.3}$$

4.2. 분산 최소화를 위한 관리적 선정

Jessen(1969)은 층화를 고려하지 않은 경우 모집단 총합에 대한 Horvitz-Thompson(1952) 추정량의 분산을 최소화시키기 위하여 '방법4'를 제안하였으나 이 방법을 실행하는 데에는 몇 가지 문제점이 있다. 우선 Nigam, Kumar와 Gupta(1984)가 지적한 바와 같이 이 방법은 각 추출단위의 포함확률을 만족시키는 해를 얻기 위해서 임의적인 반복 절차를 거쳐야하며 $n=2$인 경우에만 해를 얻을 수 있다. 또한 가능한 표본의 수가 비교적 큰 경우에는 포함확률을 정확히 만족시키는 해를 얻기가 힘들게 된다.

그러나 '방법4'는 층화시 관리적 선정에서 적극적으로 활용될 수 있으며 $n>2$인 경우에도 포함확률 π^*_{rc}를 만족시키는 표본추출계획을 쉽게 얻을 수 있다. 이를 위하여 다음과 같은 방법을 제시하고자 한다.

층화시 Yates-Grundy(1953) 분산은 다음과 같이 표시될 수 있다.

$$\begin{aligned} V(\hat{Y}_{HT}) &= \sum_{rc}\sum_{r'c'>rc}(\pi^*_{rc}\pi^*_{r'c'} - \pi^*_{rcr'c'})\left(\frac{y_{rc}}{\pi^*_{rc}} - \frac{y_{r'c'}}{\pi^*_{r'c'}}\right)^2 \\ &= \sum_{rc}\sum_{r'c'>rc} d_{rcr'c'}(\pi^*_{rc}\pi^*_{r'c'} - \pi^*_{rcr'c'})\left(\frac{y_{rc}}{\pi^*_{rc}} - \frac{y_{r'c'}}{\pi^*_{r'c'}}\right)^2 \\ &\quad + \sum_{rc}\sum_{r'c'>rc}(1-d_{rcr'c'})\pi^*_{rc}\pi^*_{r'c'}\left(\frac{y_{rc}}{\pi^*_{rc}} - \frac{y_{r'c'}}{\pi^*_{r'c'}}\right)^2 \end{aligned} \tag{4.4}$$

여기서

$$d_{rcr'rc'} = \begin{cases} 1, & rc, r'c' \in s \\ 0, & \text{그밖에} \end{cases},$$

기지인 π^*_{rc}는 rc셀의 단위가 표본에 포함될 확률, 미지인 $\pi^*_{rcr'c'}$는 rc셀과 $r'c'$셀의 단위가 동시에 표본에 포함될 확률이다.

또한 다음과 같은 식을 고려하자.

$$\sum_{rc}\sum_{r'c'>rc} W^*_{rcr'c'} = W_1^* + W_2^* \tag{4.5}$$

여기서 $W^*_{rcr'c'} = \pi^*_{rc}\pi^*_{r'c'} - \pi^*_{rcr'c'}$,

$$W_1^* = \sum_{rc}\sum_{r'c'>rc} d_{rcr'c'}\pi^*_{rc}\pi^*_{r'c'} - n^*(n^*-1)/2\ ,$$

$$W_2^* = \sum_{rc}\sum_{r'c'>rc} (1-d_{rcr'c'})\pi^*_{rc}\pi^*_{r'c'}\text{이다.}$$

W_2^*는 일정한 값이므로 W_1^*가 다음 $\overline{W_1}^*$와 동일하다면 분산을 최소화시킬 수 있는 관리적 선정 계획을 얻을 수 있다.

$$\overline{W}_1^* = (1/\sum_{rc}\sum_{r'c'>rc} d_{rcr'c'})\Big[\sum_{rc}\sum_{r'c'>rc} d_{rcr'c'}\pi^*_{rc}\pi^*_{r'c'} - n^*(n^*-1)/2\Big] \tag{4.6}$$

그렇지만 π^*_{rc}는 서로 값이 달라 그러한 표본추출계획을 얻는 것이 불가능하므로 다음과 같은 목적함수를 사용한다.

$$\sum_{rc}\sum_{r'c'>rc}\Big[d_{rcr'c'}(\pi^*_{rc}\pi^*_{r'c'} - \pi^*_{rcr'c'}) - \overline{W}_1^*\Big]^2 \tag{4.7}$$

식 (4.6)에서 $\overline{W}_1^*$는 상수이므로 식 (4.7)의 목적함수는 다음 목적함수를 최소화시키는 것과 동등(equivalent)하다.

$$\begin{aligned}\phi_2 &= \sum_{rc}\sum_{r'c'>rc}\Big[d_{rcr'c'}(\pi^*_{rc}\pi^*_{r'c'} - \pi^*_{rcr'c'})\Big]^2 \\ &= \sum_{rc}\sum_{r'c'>rc}\Big[d_{rcr'c'}(\pi^*_{rc}\pi^*_{r'c'} - \sum_{rcr'c'\in s, s\in S} p(s))\Big]^2\end{aligned} \tag{4.8}$$

제약조건식은 최적성을 고려한 관리적 선정과 동일하다. 또한 ϕ_2은 비선형 목적함수이므로 이를 실행하기 위해서는 비선형계획법을 이용하며 결과적으로 얻어지는 표본추출계획은 분산을 최소화시킬 수 있다.

4.3. 통합 관리적 선정

앞서 제안한 표본의 최적성과 바람직하지 않은 표본들의 추출확률을 동시에 고려한 목적함수 ϕ_1은 선형 목적함수이며 분산 최소화를 위한 목적함수는 비선형이므로 각각 선형계획법과 비선형계획법을 사용하지만 동일한 제약조건 하에서 목적함수를 최소화시키는 것은 공통적인 것이다. 따라서 컨트롤 라운딩의 최적성을 갖는 표본들의 추출확률을 높이면서 바람직하지 않은 표본들의 추출확률과 분산을 최소화시킬 수 있는 다음과 같은 목적함수를 구성할 수 있다.

$$\phi_3 = \sum_{s\in S} M\Big[R(\pi^*_{rc}), \pi^*_{rc}\Big]p(s) + \sum_{s\in S_1} p(s) + \sum_{rc}\sum_{r'c'>rc}\Big[d_{rcr'c'}\Big(\pi^*_{rc}\pi^*_{r'c'} - \sum_{s\ni rcr'c', s\in S} p(s)\Big)\Big]^2 \tag{4.9}$$

위의 목적함수로부터 얻을 수 있는 표본추출계획은 목적함수 ϕ_1을 사용할 때 보다 컨트롤 라운딩의 최적성을 갖는 표본들의 추출확률이 작아지거나 바람직하지 않은 표본들의 추출확률이 커질 수 있으며 목적함수 ϕ_2 을 사용하는 경우보다 분산이 다소 커질 수 있다. 그렇지만 실제 표본설계에 있어서 ϕ_1과 ϕ_2를 동시에 고려해야 하는 것이 바람직하므로 이 방법을 사용하여 효율적인 표본추출계획을 얻을 수 있을 것이다.

5. 예 제

다음과 같은 8×3 모집단(Causey 등, 1985)에서 각 행은 층(stratum)을 나타내고 각 열은 일정 지역의 주(state)를 의미하며 $\pi_{rc} = nA_{rc}$은 인구센서스로 부터 얻어지는 1차추출단위의 크기를 말한다. 단, 각 셀로부터 $[\pi_{rc}]$개의 단위들은 반드시 추출되므로 괄호 안의 값은 이들 단위들을 제외시킨 나머지 기대수 π^*_{rc}에 대한 y_{rc}이다.

표 5.1: 이원층화 자료

$\pi_{rc} = nA_{rc}$			Σ
0.4(0.65)	2.0	0.0	2.4
1.2(0.25)	0.0	1.0	2.2
0.2(0.30)	0.0	0.0	0.2
1.2(0.35)	0.4(0.70)	0.2(0.25)	1.8
1.0	0.6(1.05)	0.2(0.30)	1.8
0.0	0.4(0.65)	0.4(0.55)	0.8
0.0	0.2(0.20)	0.4(0.75)	0.6
0.0	0.0	0.2(0.25)	0.2
$\sum$4.0	3.6	2.4	10

위의 이원층화에서 얻을 수 있는 모든 가능한 표본의 수는 141개(부록1 참조)이며 최적 컨트롤 라운딩인 표본들은 6개(부록1 참조)이다. 또한 1열 2행으로부터 1열 5행까지 연속적으로 단위들이 할당되는 32개(부록1 참조)의 표본들을 바람직하지 않은 표본들로 간주한다.

다음 표는 본 논문에서 제안된 방법들과 기존의 방법들로부터 얻은 관리적 선정 계획(부록2 참조)을 비교한 것인데 각 방법에 대한 바람직하지 않은 표본들의 추출확률은 모두 동일하게 0.2의 값을 갖는다. 각 방법에 대한 표본추출계획을 얻기 위해서 SAS/OR(LP, NLP)을 이용하였다.

*여기서 $V = V_1 + V_2$, $V^* = V_1^* + V_2$, $V_1 = \sum_{rc} \sum_{r'c'>rc} d_{rcr'c'}(\pi^*_{rc}\pi^*_{r'c'} - \pi^*_{rcr'c'}\left(\frac{y_{rc}}{\pi^*_{rc}} - \frac{y_{r'c'}}{\pi^*_{r'c'}}\right)^2$,

$$V_1^* = \overline{W}_1^* \sum_{rc} \sum_{r'c'>rc} d_{rcr'c'}\left(\frac{y_{rc}}{\pi^*_{rc}} - \frac{y_{r'c'}}{\pi^*_{r'c'}}\right)^2, \quad V_2 = \sum_{rc} \sum_{r'c'>rc} (1 - d_{rcr'c'})\pi^*_{rc}\pi^*_{r'c'}\left(\frac{y_{rc}}{\pi^*_{rc}} - \frac{y_{r'c'}}{\pi^*_{r'c'}}\right)^2$$

표 5.2: 각 방법의 비교

구 분	Sitter	Tiwari	ϕ_1	ϕ_2	ϕ_3
최적표본들의 추출확률	0.25		0.4	0.150771	0.297661
V (V_1)	0.304687 (0.167187)		0.465625 (0.328125)	0.241118 (0.103618)	0.262528 (0.125028)
$V^*(V_1^*)$	0.184545(0.047045)				

이다.

위의 표로부터 알 수 있듯이 최적 컨트롤 라운딩인 표본들이 추출될 확률은 ϕ_1이 가장 크며 분산의 경우 ϕ_3, ϕ_2 등의 순으로 작다.

6. 결 론

본 논문에서는 이원층화시 효율적인 관리적 선정을 위하여 3가지 방법을 제시하였다. 첫째, 모든 가능한 표본들 중 최적 컨트롤 라운딩인 표본들의 추출확률을 높이면서 바람직하지 않은 표본들의 추출확률을 최소화시키는 방법을 제안하였다. 둘째, 분산을 최소화시킬 수 있는 관리적 선정 계획을 세우기 위하여 Jessen(1969)의 '방법4'를 적용하였고 이를 위하여 비선형계획법을 사용하였다. 셋째, 최적 컨트롤 라운딩인 표본들과 바람직하지 않은 표본들의 추출확률을 동시에 고려하면서 분산을 최소화시키기 위한 통합 관리적 선정 방법을 제시하였다.

제안된 방법들은 표본설계 목적에 따라 달리 사용될 수 있다. 특히 통합 관리적 선정 방법은 표본조사의 실제적인 측면과 이론적인 측면을 동시에 고려하면서도 Sitter와 Skinner(1994)의 방법과 Tiwari와 Nigam(1998)의 방법보다 효율성을 높일 수 있다.

참고문헌

[1] Bryant, E.C., Hartley, H.O. and Jessen, R.J. (1960). Design and Estimation in two-way stratification, *The Journal of the American Statistical Association*, Vol. 55, 105-124.

[2] Causey, B.D., Cox, L.H. and Ernst, L.R. (1985). Applications of transportation theory to statistical problems, *The Journal of the American Statistical Association*, Vol. 80, 903-909.

[3] Cox, L.H. and Ernst, L.R. (1982). Controlled Rounding, *INFOR*, Vol. 20, 423-432.

[4] Glover, F., Karney, D., Klingman, D. and Napier, A. (1974). A computation study on start procedures, Basic change criteria and solution algorithms for transportation

problems, *Management Sciences*, Vol. 20, 793-813.

[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] 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.

[7] Jessen, R.J. (1969). Some methods of probability non-replacement sampling, *The Journal of the American Statistical Association*, Vol. 64, 175-193.

[8] Jessen, R.J. (1970). Probability sampling with marginal constraints, *The Journal of the American Statistical Association*, Vol. 65, 776-795.

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

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

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

[12] SAS/OR Technical Report : The NLP Procedure, *SAS Institute Inc.*(1997).

[13] SAS/OR User's Guide, *SAS Institute Inc.*(1985).

[14] Sitter, R.R. and Skinner, C.J. (1994). Multi-way stratification by linear programming, *Survey Methodology*, Vol. 20, 65-73.

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

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

[17] Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata and with probability proportional to size, *Journal of the Royal Statistical Society*, B, Vol. 15, 253-261.

[1999년 5월 접수, 1999년 12월 채택]

부록1. 모든 가능한 141개 표본

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0
1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0	2 0 0
1 0 0	1 0 0	1 0 0	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0
0 0 1	0 1 0	0 1 0	0 0 0	0 0 1	0 1 0	0 1 0	0 1 0	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1	0 1 0	0 1 0
0 1 0	0 0 1	0 1 0	0 1 0	0 1 0	0 0 0	0 0 1	0 1 0	0 0 1	0 1 0	0 0 0	0 0 1	0 1 0	0 0 0	0 0 1
0 0 1	0 0 1	0 0 1	0 0 1	0 0 0	0 0 1	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1	0 0 0	0 0 0	0 0 1	0 0 0

*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0
1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1
1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0
1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0
1 0 0	1 0 0	1 0 0	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0
0 0 1	0 1 0	0 1 0	0 0 0	0 0 1	0 1 0	0 1 0	0 1 0	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1	0 1 0	0 1 0
0 1 0	0 0 1	0 1 0	0 1 0	0 1 0	0 0 0	0 0 1	0 1 0	0 0 1	0 1 0	0 0 0	0 0 1	0 1 0	0 0 0	0 0 1
0 0 1	0 0 1	0 0 1	0 0 1	0 0 0	0 0 1	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1	0 0 0	0 0 0	0 0 1	0 0 0

*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)	(45)
0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0
1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1
1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0
1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 1 0	1 1 0	1 1 0
1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 1	1 0 1	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 0 0	1 0 0	1 0 0
0 0 0	0 0 1	0 1 0	0 1 0	0 1 0	0 0 0	0 1 0	0 0 0	0 0 0	0 0 0	0 0 1	0 1 0	0 0 0	0 0 0	0 0 1
0 1 0	0 1 0	0 0 0	0 0 1	0 1 0	0 1 0	0 0 0	0 0 0	0 0 1	0 1 0	0 0 0	0 0 0	0 0 1	0 1 0	0 0 0
0 0 1	0 0 0	0 0 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1

*	*	*	*	*	*	*	*	*	*	*	*			
(46)	(47)	(48)	(49)	(50)	(51)	(52)	(53)	(54)	(55)	(56)	(57)	(58)	(59)	(60)
0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0
1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	2 0 1	2 0 1	2 0 1
1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	0 0 0	0 0 0	0 0 0
1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 0 0	1 0 0	1 0 0
1 0 0	1 0 0	1 0 0	1 0 0	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 1 0	1 1 0	1 1 0	1 0 0	1 0 0	1 0 0
0 0 1	0 0 1	0 1 0	0 1 0	0 0 0	0 0 0	0 0 0	0 0 1	0 1 0	0 0 0	0 0 0	0 0 1	0 0 1	0 1 0	0 1 0
0 0 1	0 1 0	0 0 0	0 0 1	0 0 0	0 0 1	0 1 0	0 0 0	0 0 0	0 0 0	0 0 1	0 0 0	0 1 0	0 0 1	0 1 0
0 0 0	0 0 0	0 0 1	0 0 0	0 0 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 1	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1

(61)	(62)	(63)	(64)	(65)	(66)	(67)	(68)	(69)	(70)	(71)	(72)	(73)	(74)	(75)
0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0	0 2 0
2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1	2 0 1
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 1	1 0 1	1 0 1
1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 0 0	1 0 0	1 0 0
0 0 0	0 0 1	0 1 0	0 1 0	0 1 0	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1	0 1 0	0 1 0	0 0 0	0 0 1	0 1 0
0 1 0	0 1 0	0 0 0	0 0 1	0 1 0	0 0 1	0 1 0	0 0 0	0 0 1	0 1 0	0 0 0	0 0 1	0 1 0	0 1 0	0 0 0
0 0 1	0 0 0	0 0 1	0 0 0	0 0 0	0 0 1	0 0 1	0 0 1	0 0 0	0 0 0	0 0 1	0 0 0	0 0 1	0 0 0	0 0 1

< 계속 >

(76)	(77)	(78)	(79)	(80)	(81)	(82)	(83)	(84)	(85)	(86)	(87)	(88)	(89)	(90)
020	020	020	020	020	020	020	020	020	020	020	020	020	020	020
201	201	201	201	201	201	201	201	201	201	201	201	201	201	201
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
101	101	101	101	101	101	101	101	101	110	110	110	110	110	110
100	100	101	101	110	110	110	110	110	100	100	100	100	100	100
010	010	000	010	000	000	000	001	010	000	000	001	001	001	010
001	010	010	000	000	001	010	000	000	001	010	000	001	010	000
000	000	000	000	001	000	000	000	000	001	001	001	000	000	001

(91)	(92)	(93)	(94)	(95)	(96)	(97)	(98)	(99)	(100)	(101)	(102)	(103)	(104)	(105)
020	020	020	020	020	020	020	020	020	120	120	120	120	120	120
201	201	201	201	201	201	201	201	201	101	101	101	101	101	101
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
110	110	110	110	110	110	110	110	110	100	100	100	100	100	100
100	101	101	101	101	101	110	110	110	100	100	100	101	101	101
010	000	000	000	001	010	000	000	001	001	010	010	000	001	010
001	000	001	010	000	000	000	001	000	010	001	010	010	010	000
000	001	000	000	000	000	001	000	000	001	001	001	001	000	001

(106)	(107)	(108)	(109)	(110)	** (111)	(112)	(113)	** (114)	(115)	(116)	(117)	(118)	(119)	(120)
120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
100	100	100	100	100	100	100	100	100	101	101	101	101	101	101
101	101	110	110	110	110	110	110	110	100	100	100	100	100	101
010	010	000	000	001	001	001	010	010	000	001	010	010	010	000
001	010	001	010	000	001	010	000	001	010	010	000	001	010	010
000	000	001	001	001	000	000	001	000	001	000	001	000	000	000

(121)	(122)	(123)	(124)	(125)	(126)	(127)	(128)	(129)	** (130)	(131)	(132)	** (133)	(134)	(135)
120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
101	101	101	101	101	101	110	110	110	110	110	110	110	110	110
101	110	110	110	110	110	100	100	100	100	100	100	100	101	101
010	000	000	000	001	010	000	000	001	001	001	010	010	000	000
000	000	001	010	000	000	001	010	000	001	010	000	001	000	001
000	001	000	000	000	000	001	001	001	000	000	001	000	001	000

(136)	(137)	(138)	(139)	** (140)	** (141)
120	120	120	120	120	120
101	101	101	101	101	101
000	000	000	000	000	000
110	110	110	110	110	110
101	101	101	110	110	110
000	001	010	000	000	001
010	000	000	000	001	000
000	000	000	001	000	000

* : 바람직하지 않은 표본 , ** : 최적 컨트롤 라운딩인 표본

부록2. 각 방법에 따른 관리적 선정 계획

구분	Sitter	Tiwari	ϕ_1	ϕ_2	ϕ_3
$p(1)$				0.007913	
$p(2)$					0.006555
$p(3)$					
$p(4)$				0.001718	
$p(5)$	0.100000			0.010972	0.013435
$p(6)$				0.011043	
$p(7)$				0.005936	0.017038
$p(8)$				0.026076	0.029217
$p(9)$					0.001340
$p(10)$				0.004596	0.011309
$p(11)$				0.014655	0.009410
$p(12)$				0.055856	0.065258
$p(13)$				0.007277	0.006894
$p(14)$				0.016882	0.039544
$p(15)$	0.100000		0.200000	0.037076	
$p(16)^*$				0.000195	0.002122
$p(17)^*$					0.004103
$p(18)^*$					
$p(19)^*$			0.100000		
$p(20)^*$					
$p(21)^*$					
$p(22)^*$				0.002992	
$p(23)^*$					
$p(24)^*$					0.007658
$p(25)^*$					
$p(26)^*$					
$p(27)^*$				0.026604	
$p(28)^*$				0.000373	0.019303
$p(29)^*$					
$p(30)^*$	0.100000			0.009643	
$p(31)^*$			0.10000		
$p(32)^*$				0.015868	0.003705
$p(33)^*$					
$p(34)^*$					0.006067
$p(35)^*$					
$p(36)^*$				0.000055	
$p(37)^*$					
$p(38)^*$					
$p(39)^*$					0.018297
$p(40)^*$					

< 계속 >

구분	Sitter	Tiwari	ϕ_1	ϕ_2	ϕ_3
$p(41)^*$					0.015650
$p(42)^*$				0.037628	0.021327
$p(43)^*$					
$p(44)^*$					
$p(45)^*$					
$p(46)^*$				0.025755	0.017945
$p(47)^*$				0.004465	0.008547
$p(48)^*$					
$p(49)^*$	0.050000			0.004265	0.002104
$p(50)^*$	0.050000				
$p(51)^*$					0.007641
$p(52)^*$				0.012245	
$p(53)^*$					0.001408
$p(54)^*$				0.019534	0.035954
$p(55)^*$				0.035195	0.028168
$p(56)^*$				0.004232	
$p(57)^*$				0.000952	
$p(58)$				0.004353	0.022274
$p(59)$				0.002674	
$p(60)$				0.005273	
$p(61)$					
$p(62)$					
$p(63)$					
$p(64)$				0.010413	
$p(65)$					
$p(66)$				0.015416	0.005712
$p(67)$				0.001084	0.000872
$p(68)$					
$p(69)$				0.000493	
$p(70)$					
$p(71)$					0.010767
$p(72)$				0.000656	
$p(73)$				0.001972	
$p(74)$				0.008355	
$p(75)$				0.003188	
$p(76)$				0.005634	
$p(77)$					
$p(78)$				0.002640	
$p(79)$	0.050000		0.100000		
$p(80)$					

< 계속 >

구분	Sitter	Tiwari	ϕ_1	ϕ_2	ϕ_3
$p(81)$				0.002327	0.005723
$p(82)$				0.001073	
$p(83)$					0.002852
$p(84)$				0.0027891	0.054290
$p(85)$					
$p(86)$					
$p(87)$	0.150000				
$p(88)$			0.100000	0.000139	0.002520
$p(89)$					0.012341
$p(90)$					
$p(91)$				0.018829	0.006039
$p(92)$					
$p(93)$				0.012633	0.015527
$p(94)$				0.009179	
$p(95)$					0.030714
$p(96)$					
$p(97)$					
$p(98)$				0.004427	0.030369
$p(99)$				0.061348	
$p(100)$				0.014073	
$p(101)$				0.009422	
$p(102)$				0.004229	
$p(103)$					0.003033
$p(104)$					
$p(105)$					0.018848
$p(106)$				0.006869	
$p(107)$					
$p(108)$					
$p(109)$				0.005366	0.008369
$p(110)$					
$p(111)^{**}$				0.012046	0.031773
$p(112)$				0.022489	
$p(113)$				0.040239	
$p(114)^{**}$	0.100000		0.100000	0.005096	0.065167
$p(115)$					0.008904
$p(116)$				0.008657	0.015521
$p(117)$				0.000514	0.011012
$p(118)$					
$p(119)$					
$p(120)$					0.024687

< 계속 >

구분	Sitter	Tiwari	ϕ_1	ϕ_2	ϕ_3
$p(121)$				0.014522	0.002497
$p(122)$					
$p(123)$	0.050000			0.026527	
$p(124)$	0.100000			0.019502	0.009467
$p(125)$				0.013710	
$p(126)$				0.009935	
$p(127)$					
$p(128)$					
$p(129)$					
$p(130)^{**}$					0.000768
$p(131)$					
$p(132)$					
$p(133)^{**}$				0.054225	0.069470
$p(134)$					
$p(135)$				0.000015	
$p(136)$					
$p(137)$				0.043843	
$p(138)$				0.009314	
$p(139)$					
$p(140)^{**}$				0.039797	0.012924
$p(141)^{**}$	0.150000		0.30000	0.039607	0.117559

* : 바람직하지 않은 표본의 추출확률,
** : 최적 컨트롤 라운딩인 표본의 추출확률

Two-dimensional Controlled Selection with Marginal Constraints

Jong-Ho Kim[1)] Jea-Bok Ryu[2)] Sun-Woong Kim[3)]

ABSTRACT

Goodman and Kish(1950) proposed controlled selection that reduces the selection probabilities of non-preferred samples which may not only increase the expenditure on travel but also affect the supervision and organization of fieldwork. This controlled selection can be used when the sample size is less than the number of strata in stratified sampling. Several methods of controlled selection were presented in two-way stratification. Causey et al.(1985) suggested an algorithm that uses transportation theory. Sitter and Skinner (1994), Tiwari and Nigam(1998) used a linear programming approach for controlled selection. We deal with problems of previous methods and suggest new method of controlled selection that increases the selection probabilities of optimal samples in two-dimensional case. In addition we propose controlled selection for minimizing variance and unified approach of controlled selection.

Keywords: Controlled Selection; Non-preferred Sample; Two-way Stratification; Non-linear programming.

1) Professor, Department of Statistics, Dongguk University.
2) Professor, Department of Applied Statistics, Chongju University. E-mail: jbryu@chongju.ac.kr
3) Lecturer, Depatment of Statistics, Dongguk University. E-mail: KSWYH@chollian.net