응용통계연구 (2011)
24(2), 393–400

DOI: 10.5351/KJAS.2011.24.2.393

이단계 집락추출에서의 표본크기에 대한 연구

송종호[1] · 제해성[2] · 박민규[3]

[1]고려대학교 통계학과, [2]고려대학교 통계학과, [3]고려대학교 통계학과

(2011년 2월 접수, 2011년 3월 채택)

요 약

조사비용과 시간과 같은 현실적인 제약하에서 관측단위(observation unit)의 집합인 집락(cluster)을 추출하는 집락추출법은 대부분의 대형조사(large scale survey)에서 흔히 사용된다. 특별히 집락내의 관측단위가 매우 유사한 경우, 집락 내의 모든 관측치를 조사하는 대신 일부를 추출하여 조사하는 이단계 집락 추출법이 선호된다. 이단계 집락추출법의 적용시 집락인 1차추출단위(Primary Sampling Unit; PSU)와 관측단위인 2차추출단위(Secondary Sampling Unit; SSU)의 표본수 결정은 주어진 비용과 표본으로부터 계산되어지는 통계량의 정도에 의존한다. 본 연구에서는 기존의 1차추출단위의 크기가 동일하다는 가정하에서 유도된 최적 PSU와 SSU 표본크기 산출과정을 일반화하여 1차추출단위의 크기가 같지 않을 경우의 최적 표본크기를 유도하고 그 결과를 제4차 퇴원환자조사를 위한 표본추출 방안에 적용하여 기존방법과 비교하였으며 이를 바탕으로 제7차 퇴원환자조사를 위한 표본크기를 제안하였다.

주요용어: 이단계집락추출, 1차추출단위, 2차추출단위, 표본크기결정, 손상퇴원환자조사.

1. 서론

대부분의 대형조사(large scale survey)에서는 비용이나 시간과 같은 현실적인 제약 때문에 관측단위(observation unit)의 집합인 집락(cluster)을 추출하는 집락추출법이 흔히 사용된다. 일단계 집락추출법은 집락 내의 모든 관측단위를 추출하는 방법으로 집락 내의 관측단위들이 서로 이질적일수록 효과적이나, 현실적으로는 집락 내 관측단위들은 일반적으로 유사한 성격을 갖는다. 집락 내의 관측치가 서로 유사한 경우, 비록 최종표본의 관측치 수가 동일하다 할지라도 주어진 추출집락 수를 늘이고 집락 내 추출단위 수를 줄이는 이단계 추출방법이 일단계 추출방법보다 효율적임이 알려져 있다 (Sharon, 2010). 그러나 지나치게 많은 집락을 추출할 경우, 이동경비와 같은 조사비용이 늘어나는 문제점이 발생할 수 있다. 따라서 주어진 비용과 알려져 있는 집락 내 관측치들의 유사성 및 이와 밀접하게 관련되어 있는 집락 내 분산을 이용한 최적 표본 1차추출단위 및 2차추출단위의 결정은 표본설계에 있어 매우 중요하다.

Koch (1967)는 실험 디자인에서 자료의 계층구조를 표현하기 위해 랜덤효과모형을 적용하고 이를 바탕으로 모평균에 대한 불편추정량을 제시하였다. 이단계 집락추출에서 모수 추정에 대한 연구는 Fuller (1975)에 의해 Super-population medel하에서 처음으로 시도되었으며, Royall (1976)에 의하여 대안적인 추정량과 추정을 위한 효과적인 표본추출 설계에 대한 연구가 진행되었다. Richard와

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업지원을 받아 수행된 연구임(2009-0074328).

[3]교신저자: (608-736) 서울시 성북구 안암5가 314-79, 고려대학교 통계학과, 교수. E-mail: mpark2@korea.ac.kr

Robert (1979)는 Royall (1976)과 Koch (1967)에 의하여 제시된 추정량의 분산 추정량을 제시하였다. Cochran (1977)은 PSU 및 SSU 조사비용이 집락 간 동일하며 또한 모집단 집락 크기가 동일하다는 가정하에서 모평균 추정량의 분산을 주어진 비용하에서 최소화 시키고 표본 PSU와 SSU 크기를 산출하였다.

본 연구에서는 Cochran (1977)의 방법을 집락의 크기가 동일하지 않은 일반적인 상황으로 확장하여 주어진 비용하에서 PSU와 SSU의 최적 표본크기가 유도되었고 이를 질병관리본부에서 실시한 제4차 퇴원환자조사 결과에 적용함으로써 그 타당성을 제고하였다.

2. 이단계 집락추출에서의 최적 표본크기 산출

퇴원환자조사를 포함한 많은 조사를 위한 표본추출방법으로 층화집락추출법이 흔히 사용된다. 본 연구에서는 층간 표본추출이 서로 독립적임을 감안하여 1단계, 2단계 추출과정 모두 단순임의추출법을 가정하였다. 주어진 가정하에서 모집단 총계 $t\,(=\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij})$에 대한 추정량 $\hat{t}\ (=\sum_{i=1}^{N} M_i\bar{y}_i$, 여기서 $\bar{y}_i = 1/m_i\sum_{j=1}^{m_i} y_{ij})$의 분산과 비용함수는 각각 다음과 같이 정의된다.

$$\mathrm{Var}\left(\hat{t}\right) = N^2\left(1-\frac{n}{N}\right)\frac{S_t^2}{n} + \frac{N}{n}\sum_{i=1}^{N}\left(1-\frac{m_i}{M_i}\right)M_i^2\frac{S_i^2}{m_i}, \tag{2.1}$$

$$C = C_1 n + \frac{n}{N}\sum_{i=1}^{N} C_{2i}m_i, \tag{2.2}$$

여기서 N은 모집단 PSU 수, n은 표본 PSU 수, M_i는 i번째 집락의 모집단 SSU 수, m_i는 i번째 집락의 표본 SSU 수를 나타내며, $S_t^2\,(=(N-1)^{-1}\sum_{i=1}^{N}(t_i-\bar{t})^2)$은 집락 간의 총계의 분산, $S_i^2\,(=(M_i-1)^{-1}\sum_{i=1}^{M_i}(y_{ij}-\bar{y}_i)^2)$은 집락 내의 SSU의 분산을 나타낸다. 그리고 C는 전체 비용으로 PSU 조사비용 C_1과 i번째 집락의 SSU 조사비용 C_{2i}으로 표현된다. 주어진 분산 식과 비용함수를 이용하여 비용함수의 제약을 만족시키면서 분산을 최소화하는 PSU와 SSU의 크기를 결정하기 위하여 Khan 등 (2006)과 같이 라그랑지승수(Lagrange Multiplier)를 이용한 다음 목적함수 ϕ를 고려하였다.

$$\phi\,(n,m,\lambda) = N^2\left(1-\frac{n}{N}\right)\frac{S_t^2}{n} + \frac{N}{n}\sum_{i=1}^{N}\left(1-\frac{m_i}{M_i}\right)M_i^2\frac{S_i^2}{m_i} - \lambda\left(C_1 n + \frac{n}{N}\sum_{i=1}^{N} C_{2i}m_i - C\right). \tag{2.3}$$

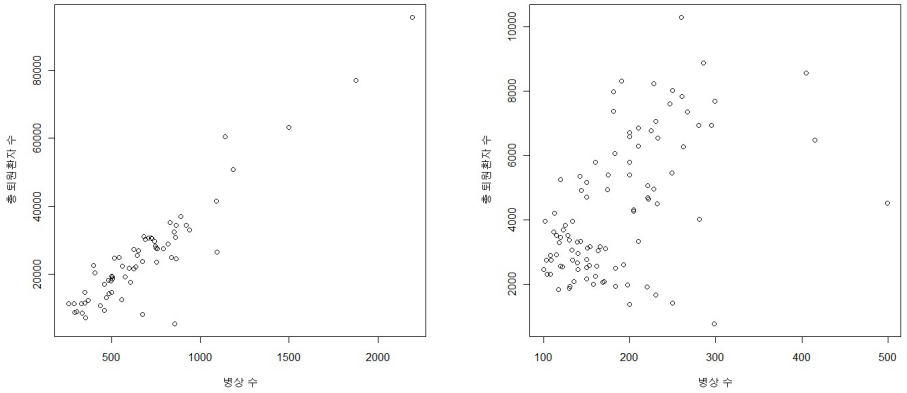
목적식 (2.3) ϕ를 최소화하는 PSU 크기 n과 SSU 크기 m_i 그리고 라그랑지 승수는 다음과 같다.

$$n_{opt} = \frac{1}{\sqrt{\lambda C_1}}\sqrt{N^2S_t^2 - N\sum_{i=1}^{N} M_iS_i^2}, \tag{2.4}$$

$$m_{i_{opt}} = \sqrt{\frac{C_1N^2M_i^2S_i^2}{\left(N^2S_t^2 - N\sum\limits_{i=1}^{N} M_iS_i^2\right)C_{2i}}}, \tag{2.5}$$

$$\frac{1}{\sqrt{\lambda}} = \frac{C}{\sqrt{\left(N^2S_t^2 - N\sum\limits_{i=1}^{N} M_iS_i^2\right)C_1} + \sum\limits_{i=1}^{N} M_iS_i\sqrt{C_{2i}}}, \tag{2.6}$$

여기서 모든 모집단과 표본의 집락 크기가 동일($M_i = M$ 그리고 $m_i = m$)하고 집락 내에서의 조사비용



(a) 전산화 병원 (b) 비전산화 병원

그림 3.1. 전산화, 비전산화 병원의 병상 수와 총 퇴원환자 수의 산점도

표 3.1. 기존의 병원별 퇴원환자의 추출

병상 수 규모	전산화	비전산화
200병상 미만	9% 추출	300명
200병상 이상 300병상 미만	9% 추출	420명
300병상 이상	9% 추출	540명

과 분산이 모두 일치($c_{2i} = c_2$, $s_i^2 = s_w^2$)하는 경우 위의 식 (2.4)와 (2.5)는 다음과 같이 표현된다.

$$n_{opt,equal} = \frac{C}{C_1 + C_2 m_{opt,equal}}, \tag{2.7}$$

$$m_{opt,equal} = \frac{S_w}{\sqrt{S_t^2/m^2 - S_w^2/M}} \sqrt{\frac{C_1}{C_2}}, \tag{2.8}$$

여기서 S_t^2은 집락 간 그리고 S_w^2은 집락 내 변이를 나타내고 있다. 식 (2.8)은 앞서 Cochran (1977)이 유도한 이단계 집락추출에서의 최적의 SSU의 크기와 일치함을 확인할 수 있다.

3. 사례연구

본 절에서는 2절에서 제시된 이단계 집락추출에서 최적의 PSU 및 SSU 크기를 질병관리본부에서 실시한 제4차 퇴원환자조사에 적용하여 사용된 표본크기의 타당성을 제고하였다. 퇴원환자조사를 위한 표본추출방법으로는 병원을 1차추출단위로 그리고 병원내 환자를 2차추출단위로 추출하는 이단계 집락추출방법이 사용되었다 (박진우와 이계오, 2007). 본 연구에서는 4차(2007년) 자료를 통하여 기존의 병원별 환자 추출률, 즉 2차 단위의 추출률과 본 연구에서 제시한 2차 단위의 추출률을 비교하였다. 표본 전산화 병원과 비전산화 병원의 병상 수와 총 퇴원환자 수의 산점도는 그림 3.1과 같고 4차 퇴원환자조사를 위한 기존의 병원별 환자 추출률은 표 3.1과 같다.

병원별 환자 추출률은 병원의 크기를 나타내는 병상 수와 전산화 여부에 따라 결정되었다. 비전산화 병원의 경우, 환자의 기록이 전산화 되어있지 않아 조사원이 환자의 차트(chart)를 직접 추출하여 필요한 정보를 모두 기록해야 함으로 전산화 병원에 비하여 자료수집 비용이 더 필요하게 된다. 표 3.1에 나타

표 **3.2.** 병상수와 전산화 유무에 따른 C_{2i}분포

병상 수 규모	전산화	비전산화
200병상 미만	-	4
200병상 이상 300병상 미만	2	8
300병상 이상 500병상 미만	3	12
500병상 이상 1000병상 미만	4	16
1000병상 이상	5	-

난 기존의 병원별 추출 환자 수를 살펴보면 전산화 병원은 병원의 병상 수와 관계없이 동일한 비율의 환자가 추출되며, 비전산화 병원은 병상 수 기준으로 3개의 범주로 분류된 병원의 크기에 따라 동일 범주 내의 병원들로 부터는 동일한 수의 환자가 추출되었음을 알 수 있다. 그러나 비전산화 병원의 경우, 병상수와 실제 입, 퇴원 환자 수가 비례관계를 나타내지 않은 경우가 빈번하게 발생하고 이로 인하여 기존의 방안은 퇴원환자 수가 표본크기 결정에 반영되지 못한 문제점을 안고 있다.

기존의 방법과 2절에서 제시된 최적 SSU 크기를 비교하기 위하여 병상 수 규모를 더욱 세분화하고 병원의 크기가 커질수록 추출 대상 환자 수가 늘어나며 비전산화 병원이 전산화 병원보다 더 많은 자료수집 비용이 요구됨을 감안하여 표 3.2와 같은 조사비용을 고려하였다. 병상수 규모의 세분화는 세분화 후 각 범주 내에서 퇴원환자 수가 일정하도록 이루어졌다. 표 3.2에 나타난 숫자는 실질적으로 상대적 크기의 의미를 나타낸다. 예를 들어 비전산화 병원의 자료수집을 위해서는 전산화 병원의 4배 정도의 비용이 소모되며, 200병상 미만의 비전산 병원의 자료수집 비용은 500병상 미만, 1,000병상 미만의 비전산화 병원에 비하여 약 1/4의 비용이 소모됨을 나타낸다. 실제 연구과정에서 표 3.2에 주어진 자료수집 비용의 가정 이외의 다른 가능한 안을 고려하였으며 그 결과는 서로 유사하였다. 전산화 200병상 미만과 비전산화 1,000병상 이상의 해당 병원은 4차 퇴원환자조사를 위한 표본에 존재하지 않아 이는 고려하지 않았다.

표 3.2와 같이 정의한 조사비용과 병원별 손상퇴원환자 여부 변수의 분산 추정량 $\hat{S}_i^2$ $(= \hat{p}_i\,(1-\hat{p}_i))$을 식 (2.5)에 적용하여 각 병원별 최적 환자 수 m_i^*을 계산하였다. 여기서 $\hat{p}_i$은 병원 i의 표본 손상퇴원환자 비율을 나타낸다.

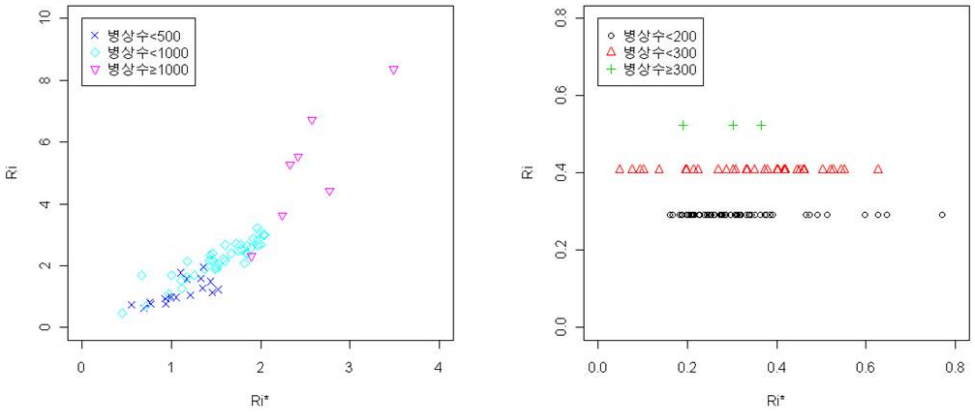
기존 추출방법과 새로 적용한 방법을 비교하기 위해 표본의 각 병원을 병원 아이디로 정렬한 후 첫 번째 병원을 기준으로 표본환자 수의 비율을 각각 식 (3.1)과 (3.2)와 같이 정의하였다.

$$R_i = \frac{m_i}{m_1}, \quad i = 2, \ldots, n, \tag{3.1}$$

$$R_i^* = \frac{m_i^*}{m_1^*}, \quad i = 2, \ldots, n, \tag{3.2}$$

여기서 i는 병원을 나타내며 R_i는 기존 방법을 통한 i병원의 첫 번째 병원 기준 표본 환자 수의 비율 그리고 R_i^*는 최적 SSU 크기 산출 방법을 통한 i병원의 첫 번째 병원 기준 표본 환자 수의 비율을 나타낸다. R_i와 R_i^*를 비교하기 위하여 작성된 전산화 병원과 비전산화 병원의 두 비율간 산점도는 다음의 그림 3.2와 같다.

$R^* \times R$ 대한 병원별 산점도가 선형의 형태로 나타난다는 것은 기존 추출방법과 제안한 추출방법이 서로 비례한 것으로 두 방법이 유사하다는 것을 의미한다. 그림 3.2(a)의 전산화 병원의 경우를 살펴보면, 기존의 9%의 추출률은 최적방법을 통하여 구한 추출률과 선형추세를 보이므로 최적방법에 가깝다고 할 수 있으나, 그림 3.2(b)의 비전산화 병원은 병상수 규모에 따라 300명, 420명, 540명으로 퇴원환자를 추출하였기 때문에 산점도가 선형이 아닌 R^*가 R에 평행한 형태로 나타났다. R^*와 R의 상관계수를



(a) 전산화 병원 (b) 비전산화 병원

그림 3.2. 전산화, 비전산화 병원의 R_i와 R_i^* 산점도

통하여 살펴본 결과, 전산화 병원의 상관계수는 0.999이고 비전산화 병원의 상관계수는 0.089로 그림 3.1과 같이 전산화 병원에서는 강한 선형관계가 비전산 병원에서는 약한 선형관계가 나타나는 것을 확인할 수 있다. 따라서 비전산화 병원의 경우 기존의 병원별 환자 추출방법의 수정이 필요할 것으로 판단된다.

살펴본바와 같이 기존의 전산화 병원의 환자 추출률 9%가 최적 환자 추출률과 유사하다는 사실을 이용하여 최적 SSU 산출식 (2.5)를 병원 내 동일 분산 가정하에서 기존 전산화 병원의 9% 환자추출률을 기준으로 비전산화 병원의 조사비용과 이에 대응되는 최적 환자 추출률을 정의하도록 하자. 전산화 병원의 환자 추출률 9%를 최적 환자 추출률로 정의하고 따라서 이를 식 (2.5)에 적용하여 아래의 식 (3.3)을 정의하였다. 여기서 constant는 식 (2.5)에서 SSU에 의존하지 않는 상수를 나타낸다.

$$\begin{aligned} m_i^* &= M_i \left(\frac{1}{\sqrt{C_{2i(\text{전산화 병원})}}} \right) \times \text{constant} \\ &= M_i \times 0.09. \end{aligned} \tag{3.3}$$

위의 식을 기초로 비전산화 병원별 환자 조사비용이 전산화 병원의 a배 만큼 크다는 가정하에서, 비전산화 병원별 환자 조사비용 $C_{2i(\text{비전산화 병원})} = a \times C_{2i(\text{전산화 병원})}$로 표현됨을 반영하면 동일한 병상 수를 갖는 비전산화 병원의 최적 추출 환자 수는 다음의 식 (3.4)와 같이 산출된다.

$$\begin{aligned} m_i^* &= M_i \left(\frac{1}{\sqrt{C_{2i(\text{비전산화 병원})}}} \right) \times \text{constant} \\ &= M_i \left(\frac{1}{\sqrt{C_{2i(\text{전산화 병원})}}} \right) \times \text{constant} \times \frac{1}{\sqrt{a}} \\ &= M_i \times 0.09 \times \frac{1}{\sqrt{a}}. \end{aligned} \tag{3.4}$$

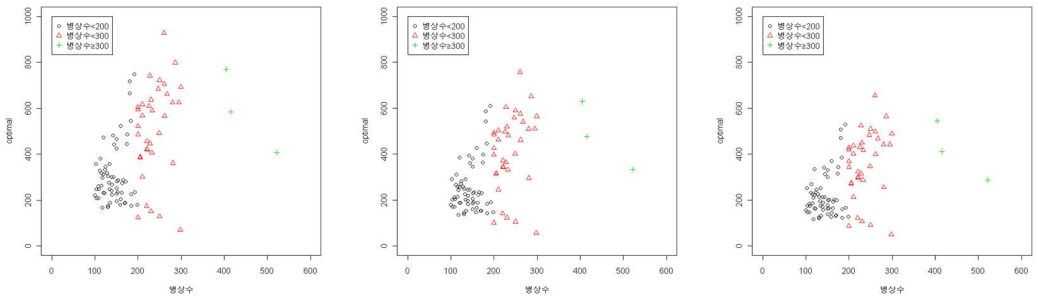
4차 퇴원환자조사의 비전산화 병원의 병원별 추출 환자 수가 환자별 조사비용을 감안하여 결정되었다는

표 **3.3.** 비전산화 병원 병상수에 따른 a_i^*의 범위

병상 수 규모	최소값	최대값	평균값	표준편차
200병상 미만	0.30	6.22	1.25	1.24
200병상 이상 300병상 미만	0.03	4.87	1.68	1.11
300병상 이상	0.57	2.03	1.26	0.74

표 **3.4.** 비전산화 병원 병상 수에 따른 최적 추출 환자 수 분포

	병상 수 규모	최소값	최대값	평균값	표준편차
$a = 1$	200병상 미만	166.6	748.4	309.2	132.2
	200병상 이상 300병상 미만	69.9	927.0	505.6	205.3
	300병상 이상	407.1	770.0	586.8	181.5
$a = 1.5$	200병상 미만	132.2	611.1	252.4	107.9
	200병상 이상 300병상 미만	57.1	756.9	412.8	167.6
	300병상 이상	332.4	628.7	479.1	148.2
$a = 2$	200병상 미만	117.1	529.2	218.6	93.5
	200병상 이상 300병상 미만	49.4	357.5	357.5	145.2
	300병상 이상	287.8	414.9	414.9	128.4



(a) $a = 1$ (b) $a = 1.5$ (c) $a = 2$

그림 **3.3.** 비전산화 병원 병상 수에 따른 최적 추출 환자수 분포

가정하에서 각 병원별 상수 a_i^*를 다음과 같이 얻을 수 있다.

$$a_i^* = \left(\frac{M_i \times 0.09}{\text{기존 비전산화 병원의 환자 추출 수}} \right)^2. \tag{3.5}$$

표 3.3은 비전산화 병원의 병상 수별 상수 a_i^*의 분포로 평균값이 1보다 약간 크게 나타나며 이는 기존 조사의 환자수 결정에 있어 유사한 병상수를 갖는 비전산화 병원의 조사비용이 전산화병원보다 평균적으로 약 1.3∼1.7배 만큼 높게 책정되었음을 의미한다.

비전산화 병원의 최적 추출 환자 수를 유도하기 위하여 a에 1, 1.5, 2의 값을 식 (3.4)에 대입하여 비전산화 병원의 최적 SSU 크기를 결정하였다. 병상 수에 따른 비전산화 병원의 최적 SSU 크기의 분포는 표 3.4와 그림 3.3과 같다. 유사한 병상 수를 갖는 병원의 경우에도 최적 추출 환자 수가 매우 다르게 나타남을 알 수 있다. 이는 실제 퇴원환자 수(M_i)와 병상 수와 상관관계가 없기 때문으로 파악된다. 따라서 비전산화 병원인 경우 기존의 200병상 미만인 병원은 300명, 200∼299병상은 420명, 300병상 이상은 540명의 주어진 환자를 추출하는 것은 바람직하지 않다. 결론적으로 비전산화 병원에서도 전산화 병

원과 마찬가지로 병원별 전체 퇴원 환자 크기에 비례하도록 표본 환자 수를 결정 하는 것이 최적 이차추출단위 크기와 유사하며 따라서 바람직한 것으로 판단된다. 현실적으로 표본추출을 위한 모집단은 해당 조사년도의 퇴원환자 수에 대한 정보가 주어지지 않은 단순 병원 리스트이며 따라서 전산화 병원에서와 같이 추출 병원의 퇴원 환자 수 결정이 조사원의 병원 방문시 즉각적으로 이루어질 수 없는 비전산화 병원의 경우, 병원 조사방문 이전에 대략적인 해당년도 전체 퇴원 환자 수를 파악하여 표본크기를 결정지어야 할 것이다.

4. 결론 및 제한점

본 논문에서는 1차추출단위의 크기가 동일하다는 가정하에서 유도된 기존의 최적 PSU와 SSU 표본크기 산출과정을 일반화하여 2차추출단위의 크기가 1차추출단위 사이에 같지 않을 경우의 최적 표본크기를 유도하였다. 유도한 결과를 바탕으로 제 4차 퇴원환자조사에 적용하여 기존방법과 비교하였으며 이를 바탕으로 전산화·비전산화 병원에서의 제 4차 퇴원환자조사를 위한 표본크기를 제안하였다. 실제 표본개편이 이루어진 7차 퇴원환자조사를 위한 표본추출을 위하여서는 전산화 유무에 관계없이 퇴원환자의 9%를 추출하는 방안이 본 연구에 기반하여 제시되었다.

본 연구에서는 병원의 규모에 상관없이 모든 비전산화 병원이 전산화 병원에 비하여 일정 비율의 비용이 더 요구된다는 가정을 하였으나 실제 비용은 각 병원에 따라 차이가 있을 수 있다. 또한 비전산화 병원의 경우, 병원의 협조와 같은 현실적인 이유로 최적 환자 수보다 적은 수의 환자가 추출될 수 있음도 감안하여야 한다.

참고문헌

박진우, 이계오 (2007). <원화자조사 표본 보정 및 가중치 산출, 최종보고서>, 질병관리본부.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed, John Willey & Sons, New York.

Fuller, W. A. (1975). Regression analysis for sample surveys, *Sankhya C*, **37**, 117–132.

Khan, M. G. M., Chand, M. A. and Ahmad, N. (2006). Optimum allocation in two-stage and stratified two-stage sampling for multivariate surveys, *American Statistical Association*, proceeding of the Survey Research Methods.

Koch, G. G. (1967). A procedure to estimate the population mean in random effect models, *Technometrics*, **9**, 577–585.

Richard, K. B. and Robert, L. S. Jr. (1979). Variance estimation based on a superpopulation model in two-stage sampling, *Journal of the American Statistical Association*, **74**, 438–440.

Royall, R. M. (1976). The linear least-squares prediction approach to two-stage sampling, *Journal of the American Statistical Association*, **71**, 657–664.

Sharon, L. L. (2010). *Sampling: Design and Analysis*, 2nd ed, Brooks/Cole, Boston.

A Study of Sample Size for Two-Stage Cluster Sampling

Jong Ho Song[1] · Hea Sung Jea[2] · Min Gue Park[3]

[1]Department of Statistics, Korea University; [2]Department of Statistics, Korea University
[3]Department of Statistics, Korea University

(Received February 2011; accepted March 2011)

Abstract

In a large scale survey, cluster sampling design in which a set of observation units called clusters are selected is often used to satisfy practical restrictions on time and cost. Especially, a two stage cluster sampling design is preferred when a strong intra-class correlation exists among observation units. The sample Primary Sampling Unit(PSU) and Secondary Sampling Unit(SSU) size for a two stage cluster sample is determined by the survey cost and precision of the estimator calculated. For this study, we derive the optimal sample PSU and SSU size when the population SSU size across the PSU are different by extending the results obtained under the assumption that all PSU have the same number of SSU. The results on the sample size are then applied to the 4^{th} Korea Hospital Discharge results and is compared to the conventional method. We also propose the optimal sample SSU (discharged patients) size for the 7^{th} Korea Hospital Discharge Survey.

Keywords: Two-stage cluster sampling, Primary Sampling Unit(PSU), Secondary Sampling Unit(SSU), sample size, Korea Hospital Discharge Survey.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2009-0074328).

[3]Corresponding author: Associate Professor, Department of Statistics, Anam-dong, Seongbuk-gu, Seoul 136-701, Korea. E-mail: mpark2@korea.ac.kr