응용통계연구 제 **14**권 1호, 2001년, pp. 111-120

절사계통추출법의 효율성에 관한 연구

이계오 [1] 최정배 [2] 석영우 [3]

요 약

하나의 업종에 대한 경영실태조사에서 모집단을 구성하는 기업체들의 규모가 큰 차이가 없다고 판단되면 계통추출법이나 층화확률추출법을 주로 적용할 수 있으나 일부기업체의 합계가 모총계의 상당히 큰 부분을 차지하는 경우에는 절사계통추출법이 효율적이다. 본고에서는 위 세 가지 추출법에 의한 모총계 추정량과 모총계 추정량의 분산의 추정법을 살펴보고, 세 가지 추출법을 비교하여 절사계통추출법의 효율성을 실제 자료인 벌목업 경영실태 조사자료를 통해서 입증하였다.

주요용어: 계통추출법, 층화확률추출법, 절사계통추출법.

1. 서론

특정업종의 미지의 모수(총고용인원, 총생산량, 총판매액 등)의 추정에 대해 관심이 있을 때는 정확히 파악하기 위해서 전수조사를 고려할 수 있으나 현실적인 제약조건(비용, 시간 등) 때문에 표본조사에 의해 모수를 추정하게 된다. 그런데 표본을 추출할 때는 직접 추출하지 않고 관심 있는 변수와 상관관계가 클 것으로 생각되는 보조자료를 이용하여 사전에 모집단을 분석하고 추출틀을 작성한 후에 표본추출법을 적용한다. 이때 모집단의 추출단위들을 보조자료의 크기에 따라 정렬할 수 있고 보조자료와 조사변수간의 상관성이 클 경우에는 계통추출법이나 층화확률추출법의 적용을 생각할 수 있다(Kish,1965). 그러나 일부 대규모 기업체의 총계가 모집단 총계의 대부분을 차지하는 경우에는 대규모 기업체들을 모두 조사대상으로 선정하고 나머지 기업체 중에서 일부를 표본으로 선정하는 절사계통추출법이 효율적이다(이계오 외 3인, 2000). 구체적으로 크기 N인 모집단의 조사단위들이 보조자료 값의 크기에 따라 정렬될 수 있다는 조건 하에서, 크기 n인 표본을 추출하여 모총계를 추정하는 방법을 살펴보기 위해 편의상 표 1.1과 같은 배열을 고려한다. 여기서 y_{ij}는 i번째 층의 j번째 조사단위의 조사대상 값으로, 조사하기 전에는 미지이며 조사 가능한 실수 값이라고 가정하다.

표 1.1에서 $N = n \cdot k$라고 가정하면

$$\boldsymbol{y}_{i\cdot} = \sum_{j=1}^{k} y_{ij} \ , \quad 1 \le i \le n,$$

1) (363-849) 충북 청원군 남일면 쌍수리, 공군사관학교 전산통계학과, 교수
E-mail: kayolee@hanimail.com
2) (363-849) 충북 청원군 남일면 쌍수리, 공군사관학교 전산통계학과, 조교수
3) (143-747) 서울특별시 광진구 군자동 98, 세종대학교 응용수학과, 교수
E-mail: sokyu@sejong.ac.kr

$$y_{\cdot j} = \sum_{i=1}^{n} y_{ij} \ , \quad 1 \le j \le k,$$

$$y_{\cdot\cdot} = \sum_{i=1}^{n} y_{i\cdot} = \sum_{j=1}^{k} y_{\cdot j} = \sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}$$

이다.

표 1.1: 자료의 배열

층 \ 단위	1	2	$\cdots$	j	$\cdots$	k	합
1	y_{11}	y_{12}	$\cdots$	y_{1j}	$\cdots$	y_{1k}	$y_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2j}	$\cdots$	y_{2k}	$y_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{ik}	$y_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
n	y_{n1}	y_{n2}	$\cdots$	y_{nj}	$\cdots$	y_{nk}	$y_{n\cdot}$
합	$y_{\cdot 1}$	$y_{\cdot 2}$	$\cdots$	$y_{\cdot j}$	$\cdots$	$y_{\cdot k}$	$y_{\cdot\cdot}$

크기 N인 모집단에서 크기 n인 표본을 추출하고자 할 때, 선형계통추출법은 첫 번째 층에서 k개의 단위 중 하나의 단위를 랜덤하게 선정한 후 동일한 열에 해당되는 단위를 선정하여 n개의 조사단위를 조사하고, 층화확률추출법은 각 층에서 오직 1개씩만을 추출하기 때문에 모총계의 불편추정량은 구할 수 있으나 모총계의 추정량의 분산에 대한 불편 추정값을 구할 수 없다.(박홍래, 2000, p212). 절사계통추출법도 일부 대규모 기업체들을 무조건 조사하고 표 1.1과 같이 배열하여 선형계통추출을 시행하기 때문에 유사한 문제점을 갖는다.

그런데 각 층에서 2개 이상의 단위를 선정하는 선형계통추출법이나, 각층에서 2개 이상의 표본을 추출하는 층화확률추출법은 모총계의 불편추정량 뿐 아니라 모총계의 추정량의 분산의 불편추정량을 구할 수 있으므로 본고에서는 세 가지 추출법에서 각 층에서 2개 이상의 표본을 추출한다는 가정하에서 모총계 추정에서 효율성을 비교 분석하고자 한다.

제 2장에서는 추출법별 모총계의 추정량 및 모총계 추정량의 분산의 추정량에 대한 일반적인 형식의 계산식을 제시하고 제 3장에서는 임업 업종별 경영실태조사중 벌목업의 실제자료에서 각 추출법에 대해 500회의 표본추출을 실시하여 효율성을 수치적으로 비교하며 마지막으로 제 4장에서 결언을 언급하겠다.

2. 모총계의 추정

모총계를 추정할 때 추출법에 따라 서로 다른 형식의 불편추정량을 제시할 수 있으나 불편추정량의 분산을 추정할 수 없다면 추출법의 효율성 비교가 곤란해진다. 본 절에서는 선형계통추출법, 층화확률추출법 및 절사계통추출법에 의한 각각의 모총계 추정량과 그들의 분산의 불편추정량을 설명하고자 한다.

2.1. 선형계통추출법(Linear Systematic Sampling)에 의한 추정

크기 N인 모집단의 조사단위들이 보조자료의 크기 순으로 정렬이 가능하다면 표 1.1에서 $y_{ij}(1 \le i \le n,\ 1 \le j \le k)$들의 배열은 다음과 같은 성질은 갖는다. 만약 $i \le r \le n$이거나 $i = r \le n$이고 $j \le s \le k$이면

$$y_{ij} \ge y_{rs} \tag{2.1}$$

이다. 다시 말하면 층의 번호가 증가하면 보조변수의 값은 감소하고 동일 층 번호 내에서는 단위번호가 증가하면 보조변수의 값은 감소한다. 그런데 표 1.1과 같이 배열하여 k개 단위 중 하나의 단위만 선택하여 n개의 표본을 추출하면 모총계의 불편추정량을 구할 수 있으나 모총계 추정량의 분산의 불편추정량은 구할 수 없으므로 층별 단위 수를 lk개로 확장하고 층수를 n/l개로 축소하여 표 2.1과 같이 배열한 후 각 층에서 $l\ (\ge 2)$개의 단위를 선택한다면 크기 n인 표본을 추출할 수 있고 분산의 불편추정값도 계산할 수 있을 것이다.

표 2.1: 자료의 배열

층 \ 단위	1	2	$\cdots$	j	$\cdots$	k	$\cdots$	lk	합
1	y_{11}	y_{12}	$\cdots$	y_{1j}	$\cdots$	y_{1k}	$\cdots$	$y_{1(lk)}$	$\boldsymbol{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2j}	$\cdots$	y_{2k}	$\cdots$	$y_{2(lk)}$	$\boldsymbol{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{ik}	$\cdots$	$y_{i(lk)}$	$\boldsymbol{y}_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
n/l	$y_{\frac{n}{l}1}$	$y_{\frac{n}{l}2}$	$\cdots$	$y_{\frac{n}{l}j}$	$\cdots$	$y_{\frac{n}{l}k}$	$\cdots$	$y_{\frac{n}{l}(lk)}$	$\boldsymbol{y}_{\frac{n}{l}\cdot}$
합	$\boldsymbol{y}_{\cdot 1}$	$\boldsymbol{y}_{\cdot 2}$	$\cdots$	$\boldsymbol{y}_{\cdot j}$	$\cdots$	$\boldsymbol{y}_{\cdot k}$	$\cdots$	$\boldsymbol{y}_{\cdot lk}$	$\boldsymbol{y}_{\cdot\cdot}$

표 2.1에서 추출방법을 자세하게 살펴보면 첫 번째 층에서는 1에서 lk까지 정수 중에서 랜덤하게 상이한 l개의 정수 $i_1, i_2, \cdots, i_l$를 선정하여 l개 해당 단위를 표본으로 추출하고 다른 층에서는 해당 위치의 단위를 표본으로 선정한다면 크기 n인 표본이 결정된다. 그러면

l개 단위에 해당되는 총계 $\boldsymbol{y}_{\cdot i_1}, \boldsymbol{y}_{\cdot i_2}, \cdots, \boldsymbol{y}_{\cdot i_l}$들이 결정되기 때문에 일종의 단순확률비복원 추출법으로 생각할 수 있다.

크기 lk인 모집단 $\{\boldsymbol{y}_{\cdot 1}, \boldsymbol{y}_{\cdot 2}, \cdots, \boldsymbol{y}_{\cdot lk}\}$에서 단순확률비복원추출법으로 선정한 임의의 표본을 $\{y_{\cdot 1}, y_{\cdot 2}, \cdots, y_{\cdot l}\}$라고 표시한다면 표본총계는 다음과 같이 나타낼 수 있다.

$$y_{\cdot\cdot} = \sum_{j=1}^{l} y_{\cdot j} \ , \quad \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{l}.$$

그러면 $\{i_1, i_2, \cdots, i_l\} \subset \{1, 2, \cdots, lk\}$에 대해서

$$\Pr(y_{\cdot 1} = \boldsymbol{y}_{\cdot i_1}, y_{\cdot 2} = \boldsymbol{y}_{\cdot i_2}, \cdots, y_{\cdot l} = \boldsymbol{y}_{\cdot i_l}) \ = \ \frac{1}{\binom{lk}{l}}$$

이다. 모총계 $\boldsymbol{y}_{\cdot\cdot}$의 추정량을 아래와 같이 표현하면

$$\widehat{\boldsymbol{y}}_{\cdot\cdot y} \ = \ lk \cdot \bar{y}_{\cdot\cdot}$$

$\widehat{\boldsymbol{y}}_{\cdot\cdot y}$는 $\boldsymbol{y}_{\cdot\cdot}$의 불편추정량이고 $\widehat{\boldsymbol{y}}_{\cdot\cdot y}$의 분산은

$$Var(\widehat{\boldsymbol{y}}_{\cdot\cdot y}) \ = \ lk(k-1)\sigma_y^2$$

이다. 여기서

$$\sigma_y^2 \ = \ \frac{\sum_{j=1}^{lk}(\boldsymbol{y}_{\cdot j} - \bar{\boldsymbol{y}})^2}{lk-1} \ , \quad \bar{\boldsymbol{y}} \ = \ \boldsymbol{y}_{\cdot\cdot}/lk$$

이다. 그런데 $Var(\widehat{\boldsymbol{y}}_{\cdot\cdot y})$는 전체를 조사하지 않으면 구할 수 없기 때문에 조사 가능한 l개의 표본으로부터 추정해야 한다. 이를 위해 표본분산을 다음식으로 표현한다면

$$\widehat{\sigma_y^2} \ = \ \frac{\sum_{j=1}^{l}(y_{\cdot j} - \bar{y_{\cdot\cdot}})^2}{l-1}$$

이므로 다음식을 만족한다.

$$E(\widehat{\sigma_y^2}) \ = \ \sigma_y^2$$

결론적으로

$$\widehat{Var}(\widehat{\boldsymbol{y}}_{\cdot\cdot y}) \ = \ lk(k-1)\widehat{\sigma_y^2}$$

은 $Var(\widehat{\boldsymbol{y}}_{\cdot\cdot y})$의 불편추정량이다.

선형계통추출법에 의한 모총계와 모총계 추정량의 분산의 추정절차를 다음과 같이 요약할 수 있다.

(1) 크기 N인 모집단의 조사단위를 보조자료의 크기에 따라 표 2.1과 같이 배열한다.

이 때, n/l이 정수가 아니면 순환계통추출법을 적용할 수 있지만 본고에서는 정수인 경우로 가정한다.

(2) 1에서 lk까지의 정수 중에서 단순확률비복원추출에 의해서 l개의 정수 $i_1, i_2, \cdots, i_l$를 선택한다.

(3) 선택된 l개의 단위에 각각 해당되는 값들 $y_{\cdot i_1}, y_{\cdot i_2}, \cdots, y_{\cdot i_l}$을 조사하여 그들의 합 $y_{\cdot\cdot}$, 평균 $\bar{y}_{\cdot\cdot}$ 및 분산 $\widehat{\sigma_y^2}$을 계산한다.

(4) 모총계의 추정값은 $\widehat{\boldsymbol{y}}_{\cdot\cdot y} = k \cdot y_{\cdot\cdot}$로 하고 모총계 추정량의 분산의 추정값은

$$\widehat{Var}(\widehat{\boldsymbol{y}}_{\cdot\cdot y}) \;=\; lk(k-1)\widehat{\sigma_y^2}$$

로 한다.

2.2. 층화확률추출법(Stratified Random Sampling)에 의한 추정

모집단의 구성이 표 2.1의 배열과 같을 때 각 층에서 l개의 조사단위를 랜덤하게 추출하여 전체적으로 n개의 조사단위로 표본을 구성한다. 선형계통추출과 마찬가지로 단순확률비복원추출로 생각할 수 있다. 표기의 편의를 위해서 조사단위의 구분을 지수(Index)로 나타낼때 $1 \le i \le n/l$에 대해 $s_i = \{i_1, i_2, \cdots, i_l\}$ 은 $\{1, 2, \cdots, lk\}$ 중에서 단순확률비복원추출될 크기 l인 표본이라 가정하고,

$$y_{i\cdot} = \sum_{j \in s_i} y_{ij} \;,\quad \bar{y}_{i\cdot} \;=\; \frac{y_{i\cdot}}{l}$$

이라 하자.

그러면 $ky_{i\cdot}$는 i번째 층의 총합 $\boldsymbol{y}_{i\cdot}$의 불편추정량이고 $ky_{i\cdot}$의 분산은

$$Var(ky_{i\cdot}) \;=\; lk(k-1)\,\sigma_i^2$$

이다. 여기서

$$\sigma_i^2 \;=\; \frac{\sum_{j=1}^{lk}(y_{ij} - \bar{\boldsymbol{y}}_{i\cdot})^2}{lk-1} \;,\quad \bar{\boldsymbol{y}}_{i\cdot} \;=\; \boldsymbol{y}_{i\cdot}/lk$$

이되고, 또한 i 층 내에서 추출된 표본의 분산을

$$\widehat{\sigma_i^2} \;=\; \frac{\sum_{j \in s_i}(y_{ij} - \bar{y}_{i\cdot})^2}{l-1}$$

이라 하면

$$E(\widehat{\sigma_i^2}) \;=\; \sigma_i^2$$

이다. 그러므로 모총계 $\boldsymbol{y}_{..}$를 추정하기 위해서

$$\widehat{\boldsymbol{y}}_{..t} \;=\; k\sum_{i=1}^{n/l} y_{i\cdot}$$

라 하면 $\widehat{\boldsymbol{y}}_{..t}$는 $\boldsymbol{y}_{..}$의 불편추정량이고

$$Var(\widehat{\boldsymbol{y}}_{..t}) \;=\; lk(k-1)\sum_{i=1}^{n/l}\sigma_i^2$$

이므로

$$\widehat{Var}(\widehat{\boldsymbol{y}}_{..t}) \;=\; lk(k-1)\sum_{i=1}^{n/l}\widehat{\sigma_i^2}$$

은 $Var(\widehat{\boldsymbol{y}}_{..t})$의 불편추정량이다.

실제 추정절차는 선형계통추출법과 대동소이하므로 선형계통추출의 절차를 준용할 수 있을 것이다.

2.3. 절사계통 추출법(Cut-off Systematic Sampling)에 의한 추정

절사계통추출법은 모집단의 조사단위를 보조자료의 크기 순으로 정렬하였을 때 소수의 대규모 기업체로 인하여 분포가 한쪽으로 심하게 기울어져 있어 전체합계의 대부분 비중이 소수 대규모 기업체에 의존하는 경향이 높은 경우에 소수 대규모 기업체는 전수조사 대상으로 하고 나머지 기업체들을 표본조사 대상으로 해서 여기에 계통추출법을 적용하여 표본을 추출 조사한다. 소수 대규모기업체 선정에 관한 조건은 이계오 외 3인(2000)이 제시하여 모총계와 모총계의 분산을 직접 계산하는 방법을 제안하였다. 그러나 전체 자료가 없는 경우는 모총계 추정량의 분산을 추정할 수 없기 때문에 다음과 같은 모총계 추정량의 분산을 추정하는 방법을 제안한다.

크기 N인 모집단의 조사단위 중에서 크기 n인 표본을 추출하고자 할 때 보조자료 값이 적정 조건을 만족시키는 n_1개 $(y_1, y_2, \cdots, y_{n_1})$는 전수 조사를 하고 나머지 $N^*(=N-n_1)$개의 조사단위에서 계통추출에 의해 $n^*(=n-n_1)$개를 추출한다. 크기 N^*인 모집단의 조사단위에서 크기 n^*인 표본을 선형계통 추출하는데 별도의 기호 정의 없이 선형계통추출에서 사용한 모든 기호 상단의 오른 쪽에 "*" 표시하여 이용하기로 한다(예: N 대신 N^* , $y_{\cdot j}$ 대신 $y^*_{\cdot j}$). 그러면 모총계 $\boldsymbol{y}_{..}$를 추정하기 위해

$$\hat{y}_{..c} \;=\; \sum_{j=1}^{n_1} y_j \;+\; k^*\sum_{j=1}^{l^*}\boldsymbol{y}^*_{\cdot j}$$

라 하면 $\hat{\boldsymbol{y}}_{..c}$는 $\boldsymbol{y}_{..}$의 불편추정량이고

$$\widehat{Var}(\hat{\boldsymbol{y}}_{..c}) = \widehat{\sigma_c^2} = l^*k^*(k^*-1)\frac{\sum_{j=1}^{l^*}(y_{.j}^* - \bar{y}_{..}^*)^2}{l^*-1}$$

이다. 지금까지 세 가지 추출법에 대한 모총계의 추정량과 추정량의 분산 추정에 대해서 설명하였다. 추출법의 효율성에 대한 수치적인 비교 분석을 위해 임업 업종중에서 벌목업의 경영실태조사를 위한 표본설계시 이용한 정보와 모집단의 특성을 살펴보고 표본추출법을 적용한 후에 결과를 분석하였다.

3. 효율성의 수치적 비교분석

임업경영실태조사에서 벌목업에 종사하는 가구 또는 법인은 1,050 조사단위로 이들의 벌목 면적을 보조정보로 사용하여 표본설계를 하였다. 벌목업에 종사하는 조사단위의 연간 수입을 추정하고자 한다면 벌목면적을 보조변수로 이용할 수 있을 것이므로 벌목면적의 분포형태와 모집단의 특성값을 알아보기 위해 계산한 기술통계량들이 표 3.1에 주어졌다.

왜도가 24.626이므로 모집단의 분포형태는 오른쪽으로 긴 꼬리를 갖는 왼쪽으로 치우친 모양을 갖고 있을 것이며 또한 제1사분위수와 제3사분위수가 각각 1과 8이지만 평균이 11.78인 점으로 보아 벌목면적이 큰 몇 개의 조사단위들의 벌목면적 합계가 전체 벌목면적의 대부분을 차지함을 보이고 있음을 알 수 있다.

표 3.1: 벌목면적의 기술통계량

단위 : ha

모집단크기	모총계	평균	표준편차	왜도	CV	Q_1	중앙값	Q_3
1,050	12378.58	11.78	68.002	24.626	576.886	1	3	8

표본의 크기는 편의상 모집단의 약 5%인 50조사단위로 하여 다음과 같이 표본추출과정을 시행하였다. 선형계통추출과 층화확률추출은 $N = 1,050$, $n = 50$, $k = 21$, $l = 2$로 하였으며 절사계통추출은 벌목면적이 80ha이상인 16개의 조사단위는 전수조사 대상 단위로 분류하고 나머지 1,034개의 조사단위에서 34개의 표본을 선형계통추출을 해야 하나 정수분할이 불가하므로 순환계통추출방법을 적용해야 할 것이지만 컴퓨터계산의 편의를 위해서 자료값이 작은 14개는 제거하였다. 즉 $N = 1,050$, $n = 50$, $n_1 = 16$, $N^* = 1,020$, $n^* = 34$, $k^* = 30$, $l^* = 2$로 하였다.

위와 같이 배열한후 3가지 방법에 대한 모총계 추정값과 모총계 추정량의 분산의 추정값을 계산하기 위해서 각각 표본추출을 500회씩 실시하여 500개 모총계와 분산의 추정값

들의 평균을 계산하여 표 3.2이 주어졌다.

표 3.2: 표본추출법에 대한 모총계 및 모총계 추정량의 분산의 추정값

구분 / 추출법	선형계통추출법	층화확률추출법	절사계통추출법
모총계의 추정값	12243.11	12362.45	12335.97
추정값의 편차	135.47	16.13	472.61
모총계추정량의 분산의 추정값	7.74×10^7	7.89×10^7	7.44×10^5
CV	71.858	71.581	6.992

표 3.2에서 보면 선형계통추출법과 층화확률추출법의 변동계수(CV)는 비슷하지만 추정값의 편차는 선형계통추출법이 상대적으로 크다는 것을 알 수 있다. 절사계통추출법의 CV는 6.992로서 선형계통추출법과 층화확률추출법의 CV의 1/10정도로 효율성과 추정의 안정성에서 월등하게 우수함을 알 수 있다.

4. 결론

기업체의 경영실태조사 등과 같이 기업체를 조사대상으로 하는 많은 조사들은 광공업 총조사 등과 같은 총조사를 실시한 후에 표본조사를 계획하는 경우가 적지 않다. 모집단의 구성단위들에 대한 보조정보를 확보할 수 있으므로 선형계통추출법이나 층화확률추출법을 이용하여 표본기업체를 추출할 수 있을 것이며 또한 소수의 대규모 기업체들의 합계가 모집단 총계의 대부분을 차지한다는 정보가 주어진 경우에는 절사계통추출법이 월등하게 우수함을 수치적인 예를 통해서 보였다. 절사계통추출법의 우수성의 정도를 이론적으로 설명할 수는 없으나 모집단의 분포형태가 한쪽으로 많이 치우친 경우에는 절사계통추출법의 적용이 타당할 것이라는 주장은 일리가 있다. 절사계통추출법을 적용하는데 절사점을 결정하는 절차에 대한 이론적인 연구는 없었지만 수치적인 절차들은 참고문헌에서 참고할 수 있을 것이다([3]).

참고문헌

[1] 이계오, 류제복, 유정빈, 정연수. (2000). 절사계통추출법, 한국통계학회 2000년 춘계학술발표회 논문집, pp1-6.

[2] 박홍래. (2000). <통계조사론>, 영지문화사, 서울.

[3] 통계청. (1991). 절사법 표본설계 응용, 통계청 연구자료(91-06-023).

[4] Leslie Kish. (1965). Survey Sampling, John Wiley and Sons Inc., New York.

[2000년 10월 접수, 2001년 1월 채택]

A Study on Efficiency of the Cut-off Systematic Sampling

Kay-O Lee [1] JungBai Choi [2] YongU Sok [3]

ABSTRACT

Either systematic sampling or stratified sampling is usually applied to the business conditions survey when companies don't have much difference in their size. But the cut-off systematic sampling is an efficient method when only a few companies are so large that the total of them almost equals to the total of whole companies. Throughout this paper, three estimators of total and their variance estimations depending on three kinds of sampling schemes are discussed, and are compared with them via their variances. It is proved that the cut-off systematic sampling is most efficient by using a real data of the logging business conditions survey.

Keywords: Systematic Sampling; Stratified Random Sampling; Cut-off Systematic Sampling.

1) Professor, Department of Computer Science and Statistics, Korea Air Force Academy.
E-mail: kayolee@hanimail.com
2) Assistant Professor, Department of Computer Science and Statistics, Korea Air Force Academy.
3) Professor, Department of Applied Mathematics, SeJong University.
E-mail: sokyu@sejong.ac.kr