계통표집법의 특성에 관한 연구

박 진 우*, 김 영 원**

<요 약>

본 논문은 계통표집법의 장점 중 거의 주목되지 못하고 있는 한 가지 장점을 다룬다. 특정한 확률표집법에 의해 표본이 추출되었을 때 추출된 표본이 모집단을 잘 반영하고 있는지의 여부는 매우 중요한 문제이다. 따라서 표집법을 평가할 때 모집단을 잘 대표하지 못하는 바람직하지 않은 표본이 추출될 가능성을 살피는 것이 필요하다. 모의실험을 통해 모집단이 순서모집단에 가까울수록 계통표집법은 무작위표집법에 비해 바람직하지 못한 표본을 추출할 확률을 적게 하는 표집법이라는 사실을 보인다.

<Abstract>

In this paper we point out another advantage of systematic sampling over simple random sampling, which have not yet been spelled out in the literature. After a single sample is drawn by a sampling scheme, it is important to check whether the achived sample represents the population well or not. Therefore, a sampling scheme which avoids the possibility of selecting non-preferred samples is desirable. The simulation results are given to illustrate that, in the ordered population, the possibility of selecting non-preferred sample by systematic sampling is lower than that by simple random sampling.

* 경기 화성 봉담 수원대학교 통계정보학과, 445-743
** 서울시 용산구 청파동 숙명여자대학교 통계학과, 140-742

Ⅰ. 서 론

표본조사에서 계통표집법은 널리 사용되어지는 대표적인 표본추출법 중의 하나이다. Cochran(1977), Kish(1965), Bellhouse(1988), 박홍래(2000) 등 표본이론을 전문적으로 다루는 대부분의 문헌들은 계통표집법이 무작위표집법에 비해 갖는 장점들을 소개하고 있는데 그 내용을 대략적으로 요약하면 다음의 세 가지로 정리할 수 있다. 첫째 계통표집법을 이용하면 표본추출이 용이하고 편리하다. 따라서 표본추출 과정에서 편향이 발생할 우려가 적다. 둘째, 계통표집법을 이용하면 모집단의 구성 비율을 잘 반영하는 표본을 얻을 수가 있다. 적절한 보조변수를 이용하여 모집단 리스트를 정렬한 후 계통표집을 하면 모집단의 구성비율을 반영하는 표본을 얻을 수 있게 된다. 셋째, 순서 모집단(ordered population)인 경우에는 계통표집에서의 추정량의 분산이 무작위표집에서의 분산보다 더 작아져서 효율을 높일 수 있다.

현대적인 표본이론은 Neyman(1934)에 의해 그 기초가 마련되었다고 할 수 있다. Neyman은 모든 가능한 확률표본의 반복적인 추출을 가정하고 그에 따른 추정이론을 제시하였다. 한편 Holt와 Smith(1979)은 이러한 전통적인 표본이론이 표본 추출이 이루어지기 전인 설계단계에 적합한 이론이라고 지적하고 있다. 모든 가능한 표본의 반복추출을 가상한 전통적인 표본이론의 틀을 가지고 계통표집법이나 무작위표집법을 바라보면 두 표집법은 모두 바람직한 성질을 지니고 있다는 사실은 널리 알려져 있다. 그러나 실제 표본조사에서는 모든 가능한 표본이 반복해서 추출되는 것이 아니라 그 중 특정한 하나의 표본만 추출되어진다는 점을 주목할 필요가 있다. 이 경우에는 무엇보다도 추출된 표본이 모집단을 제대로 반영하는 표본인가가 중요하다. Goodman과 Kish(1950)는 무작위표집법에 의해 추출된 표본이 때로 '바람직하지 못한(non-preferred)' 표본일 수 있음을 지적하였다. 여기서 바람직하지 못하다는 용어는 여러 가지 의미를 내포할 수 있으나 본 논문에서는 추정오차를 크게 만드는 표본을 일컫는 말로 제한하여 사용하기로 한다. 실제 정부나 여러 기관에서 표본조사를 담당하는 사람들은 모집단을 제대로 반영하지 못하는 바람직하지 못한 표본이 추출되는 것을 피하기 위해 많은 주의를 기울이고 있는 점을 감안한다면 어느 표집법에 의해 바람직하지 못한 표본이 추출될 가능성이 얼마나 큰가를 살펴보는 것은 매우 필요하다. 그러나 이런 측면으로 계통표집법과 무작위 표집법의 특성을 살펴보는 연구는 거의 전무하다.

본 논문에서는 계통표집법이 무작위표집법에 비해 바람직하지 못한 표본을 추출할 가능성을 작게 해준다는 장점을 고찰한다. 물론 이러한 장점은 모집단이 어느 정도 순서모집단의 형태를 나타낼 때에만 해당되는데 대부분의 대규모 조사에서는 관심변수와 높은 상관을 갖는 보조변수들의 활용이 가능하므로 이러한 논의의 의의가 크다고 할 수 있다. 2절에서는 두

표집법의 바람직하지 못한 표본의 추출확률을 소개하고 두 가지 예를 통해 계통표집법의 장점을 나타낸다. 3절에서는 시뮬레이션을 통해 두 표집법을 비교한다. 마지막으로 4절에서는 본 연구의 결과를 요약한다.

Ⅱ. 바람직하지 못한 표본의 추출확률

본격적인 논의에 앞서 먼저 '바람직하지 못한 표본'이라는 용어에 대해 명확히 규명하는 것이 필요하다. Avadhani와 Sukhatme(1965)는 특정 표본을 통해 데이터를 수집할 경우 무응답, 조사자 편향 등의 비표본오차가 생길 가능성이 크게 하는 표본에 대해 '바람직하지 못한 표본'이라는 사용하였다. 본 논문에서는 '바람직하지 못한 표본'을 그들이 소개한 개념보다 범위를 제한하여 모집단 단위 중 한쪽으로 치우친 단위들이 표본에 과다하게 선택되어 추정오차를 크게 하는 표본이라는 의미로 사용하기로 한다. 예를 들어 $N=6$ 개의 단위로 구성된 가상적인 모집단, U={ 1,2,…,6}에서 확률추출법으로 $n=2$ 개인 표본을 뽑는다고 하자. 원래 모집단의 모평균값은 3.5이다. 그런데 추출된 표본이 {1,2} 이거나 {5,6}이면 표본평균은 각각 1.5와 5.5가 되어 오차의 크기가 2가 되어 다른 어떤 경우의 표본보다 오차가 크게 된다. 이럴 경우 오차를 가장 크게 하는 이 표본들을 바람직하지 못한 표본이라고 부르기로 한다. 따라서 가능한 한 바람직하지 못한 표본이 추출될 가능성을 줄여주는 표집법이 바람직한 표집법이 된다.

관심모수를 $\theta=f(Y_1, Y_2, \cdots, Y_n)$라고 하고, $s=(i_1, i_2, \cdots, i_n)$을 표본 조사단위들을 나타낸다고 하자. 또한 $\hat{\theta}=t(\underline{Y}, s)$를 모수 θ 의 추정량이라고 하자. 전통적인 표본이론에서는 표집법이나 그에 따른 추정량의 성질을 평가할 때 불편성(unbiasedness)이나 효율성(efficiency) 등을 중요한 요소로 고려하는데 이러한 점들은 설계단계에서 유용한 개념들이다(Holt와 Smith, 1979). 왜냐하면 설계단계에서는 모든 추출 가능한 표본들을 이론적으로 다 고려할 수 있기 때문이다. 하지만 연구자의 입장에서는 모든 가능한 표본을 다 추출하는 것이 아니라 단 하나의 표본만을 추출하여 조사하게 되므로 이 하나의 표본이 모집단을 잘 반영하는가가 매우 중요한 문제가 된다. 가령 어떤 표집법이 확률적으로는 바람직한 표본을 제공할 가능성이 높다고 해도 우연히 그 표집법에 의해 뽑힌 특정표본이 바람직하지 못한 표본일 가능성을 배제할 수는 없다. 흔히 무작위 표집법은 모집단을 잘 반영하는 표본을 추출하는 좋은 방법으로 알려져 있다. 그러나 무작위 표집법으로 뽑힌 표본이 극단적인 경우가 될 가능성 또한 있다. 위에서 소개한 가상적인 모집단의 경우를 생각한다면 무작위표집법에 의해 가장 바람직하지 못한 경우인 {1,2} 이거나 {5,6}이 표본으로 뽑힐 확률은 2/15가 된다. 그러므로 어떤 표집법의 성능을 평가하려고 할 때 바람직하지 못한 표본이 뽑힐 가능성이 얼마나

되느냐 하는 것은 현실적으로 중요한 문제가 된다. 그럼에도 불구하고 이 부분에 대한 논의가 거의 이루어지지 않고 있음은 주목할 만한 사실이다.

추정오차의 크기가 M 을 초과하게 하는 표본을 바람직하지 못한 표본으로 정의하기로 하자. 이를 달리 표현한다면 $|\theta-\hat{\theta}(s)| \geq M$ 이 되게 하는 표본 s 는 바람직하지 못한 표본이 된다. 두 가지의 표집법 a 와 b 가 있다고 하자. 표집법 a 를 사용했을 때에 바람직하지 못한 표본이 추출될 가능성을 나타내기 위해 다음과 같은 확률, $\gamma_a(M)$, 을 생각할 수 있다.

$$\gamma_a(M)=P(|\theta-\hat{\theta}_a(s)|>M).$$

표집법 a 의 가능한 최대오차의 크기를 M_a라고 표기한다. 만일 표집법 b 를 사용했을 때 M_a 보다 표본오차가 큰 표본이 추출될 확률이 양수가 된다면 즉,

$$\gamma_b(M_a)=P(|\theta-\hat{\theta}_b(s)|>M_a)>0$$

이 된다면, 표집법 b 는 표집법 a 에 비해 바람직하지 못한 표본을 추출할 가능성이 적으므로 이런 면에서 더 나은 표집법이라고 판단할 수 있다.

모평균에 대한 추정 문제에서 모수는 $\theta=\frac{1}{N}\sum_{1}^{N}y_i$이며, 그에 대한 표본의 추정량은 $\hat{\theta}=\frac{1}{n}\sum_{1}^{n}y_i$ 이다. 무작위표본에 의한 추정량과 계통표본에 의한 추정량을 구분하기 위해 추정량 아래에 첨자를 써서 각각 $\hat{\theta}_{srs}$, $\hat{\theta}_{sys}$로 나타내기로 한다. 모평균 추정을 위한 표본설계에서 무작위표집법과 계통표집법을 각각 사용할 때 바람직하지 못한 표본의 추출확률을 비교하기 위해 두 가지의 예를 고려한다. 하나는 간단한 가상모집단의 예이고 다른 하나는 통계청에서 조사하는 인천시 숙박업소와 음식점업 판매량 조사의 예이다.

<예 1> (가상 모집단의 예)

전체 $N=6$개의 단위로 구성된 가상적인 모집단 $U=\{1,2,3,4,5,6\}$ 을 가정한다. 먼저 비복원 무작위표집법에 의해 $n=2$개의 표본을 뽑는 경우를 고려해보자. 아래의 <표 2.1>의 (a)는 모든 가능한 표본의 종류와 그에 따른 표본평균을 열거하고 있으며 (b)는 표본평균, $\overline{y_{res}}$, 의 표본분포를 나타내고 있다.

<표 2.1> 무작위표집법에서의 표본과 표본분포

무작위표본	표본평균	무작위표본	표본평균
{1,2}	1.5	{2,6}	4.0
{1,3}	2.0	{3,4}	3.5
{1,4}	2.5	{3,5}	4.0
{1,5}	3.0	{3,6}	4.5
{1,6}	3.5	{4,5}	4.5
{2,3}	2.5	{4,6}	5.0
{2,4}	3.0	{5,6}	5.5
{2,5}	3.5		

(a) 모든 가능한 표본과 표본평균

$\overline{y_{srs}}$	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
$P(\overline{y_{srs}})$	1/15	1/15	2/15	2/15	3/15	2/15	2/15	1/15	1/15

(b) 표본평균의 확률분포

다음으로 $n=2$ 인 계통표집법을 생각하자. <표 2.2>의 (a)는 계통표본에서의 모든 가능한 표본과 표본분산을 나타낸 표이며, (b)는 표본평균의 확률분포를 나타낸 표이다.

<표 2.2> 계통표집법에서의 표본과 표본분포

계통표본	표본평균
{1,4}	2.5
{2,5}	3.5
{3,6}	4.5

(a) 모든 가능한 표본과 표본평균

$\overline{y_{srs}}$	2.5	3.5	4.5
$P(\overline{y_{srs}})$	1/3	1/3	1/3

(b) 표본평균의 표본분포

모평균의 값은 3.5 ($\theta=3.5$)이므로 모든 가능한 무작위표본과 계통표본들의 추정값의 표본오차, $|\theta-\hat{\theta}(s)|$, 값을 계산할 수 있다. 이 계산결과를 이용하여 오차의 크기가 M 보다 크게 하는 표본의 확률, $\gamma(M)$, 값을 나타낸 결과가 아래의 <표 2.3>이다.

<표 2.3> 두 표집법에서의 $\gamma(M)$ 계산

표본오차의 크기(M)	$\gamma_{srs}(M)$	$\gamma_{sys}(M)$
0.5	12/15	2/3
1.0	8/15	2/3
1.5	4/15	0
2.0	2/15	0

위의 표를 보면 무작위표본의 최대표본오차는 $M_{srs}=2.0$ 이며, 계통표본의 최대표본오차는 $M_{sys}=1.0$ 이다. 무작위표집법을 사용할 경우 계통표본의 최대표본오차, M_{sys}, 보다 더 큰 표본오차를 갖는 표본이 추출될 확률은 $\gamma_{srs}(M_{sys})=4/15$ 가 되는데, 이것은 무작위표집법이 계통표집법보다 바람직하지 않는 표본을 추출할 확률이 크다는 사실을 말해 준다. 그러므로 이 예의 경우 계통표집법은 무작위표집법에 비해 바람직하지 못한 표본을 추출할 가능성을 적게 한다는 측면에서 더 좋은 특성을 지닌다고 말할 수 있다.

<예 2> (인천광역시의 숙박업 및 음식점업에 대한 조사의 예)

1997년 실시한(1996년도 기준) 도·소매업 총조사 결과에서 나온 인천광역시의 숙박 및 음식점업 자료를 1998년에 실시된 도·소매업 통계조사 표본설계에 따라 표본크기를 정한 후 계통표집법과 무작위표집법을 적용시켜 표본을 추출하였을 때 각 경우에서 바람직하지 못한 표본이 선택될 확률, $\gamma(M)$,을 계산하여 두 표집법을 비교한다.

인천광역시 숙박업 및 음식점업의 모집단 크기와 표본크기 및 모평균과 모표준편차 값들(1996년 기준)을 나타낸 표가 <표 2.4>이다. 도·소매업 통계조사에서 조사변수는 각 업소들의 연간 총매출액인데 활용가능한 보조변수로 고용된 종업원수가 있다. 참고로 숙박업의 경우 조사변수와 보조변수간의 상관계수는 약 0.94로 매우 높은 편이며, 음식점업은 약 0.67인 것으로 계산되었다.

<표 2.4> 모집단 상황과 표본크기

	모 집 단		
	크기	평균	표본크기
숙박업	1582	4884.6	123
음식점업	24893	4982.4	249

먼저, <표 2.4>에 있는 표본의 크기대로 무작위표본을 추출하는 작업을 1000번 반복시행한 후 M의 값의 변화에 따라 바람직하지 못한 표본이 추출될 확률, $\gamma_{srs}(M)$, 의 변화를 계산하였다. 다음으로, 계통표집법을 적용시키기 전에 보조변수의 크기에 따라 모집단단위들을 내림차순으로 정렬시켰고, 정렬된 리스트에 의거하여 계통표본을 추출하였다. 숙박업은 추출율을 약 1/10이므로 모든 가능한 계통표본의 수가 10개이며, 음식점업은 추출율이 1/100이어서 100개의 가능한 계통표본이 존재한다. 모든 가능한 표본들의 표본오차를 구하여 $\gamma_{sys}(M)$을 계산하였다. 다음의 <표 2.5>는 숙박업과 음식점업에 대한 계산 결과들을 나타낸다.

<표 2.5> 두 표집법의 바람직하지 못한 표본이 추출될 확률 비교

(a) 숙박업

표본오차의 크기(M)	$\gamma_{srs}(M)$	$\gamma_{sys}(M)$	표본오차의 크기(M)	$\gamma_{srs}(M)$	$\gamma_{sys}(M)$
200	0.724	0.616	1200	0.038	0.000
396	0.473	0.010	1400	0.018	0.000
600	0.286	0.000	1600	0.009	0.000
800	0.153	0.000	1800	0.008	0.000
1000	0.073	0.000	2934	0.001	0.000

(b)음식점업

표본오차의 크기(M)	$\gamma_{srs}(M)$	$\gamma_{sys}(M)$	표본오차의 크기(M)	$\gamma_{srs}(M)$	$\gamma_{sys}(M)$
200	0.608	0.120	1200	0.013	0.000
400	0.315	0.050	1400	0.004	0.000
600	0.134	0.010	1600	0.001	0.000
800	0.053	0.010	1800	0.001	0.000
947	0.025	0.000	1842	0.000	0.000

숙박업의 경우 계통표본 중 최대오차의 크기는 396.4 ($M_{sys}=396.4$)인 반면 무작위표본의 최대오차의 크기는 2934.4 ($M_{srs}=2934.4$)이다. 또한 무작위표집에 의해 뽑힌 표본의 오차가 계통표본 최대오차의 크기보다 클 확률은 0.473 ($\gamma_{srs}(M_{sys})=0.473$)이다. 이상의 결과는 숙박업 조사에서 계통표집법을 사용하면 무작위표집을 사용하는 것보다 바람직하지 못한 표본을 추출할 확률을 훨씬 줄일 수 있음을 보여준다. 음식점업의 경우 계통표본 중 최대오차의 크기는 947.1 ($M_{sys}=947.1$)인데 비해 무작위표본의 최대오차의 크기는 1841.9

($M_{srs}=1841.9$) 이다. 여기서는 무작위표집에 의해 뽑힌 표본의 오차가 계통표본 최대오차의 크기보다 클 확률은 0.025 ($\gamma_{srs}(M_{sys})=0.025$)이다. 이 경우 $\gamma_{srs}(M_{sys})$값 자체는 그리 크지 않지만 전반적으로 M 값이 400을 넘을 확률이 계통표집에서는 0.01에 불과한데, 무작위표집에서는 0.315로 상당히 차이가 난다. 따라서 음식점업 조사에서도 계통표집법을 사용하면 무작위표집을 사용하는 것보다 바람직하지 못한 표본을 추출할 확률을 훨씬 줄이게 된다.

추정값과 모수값과의 차이를 손실이라고 본다면 a 라는 특정 표집법에 의해 야기될 수 있는 표본오차의 정도를 위험함수(risk function)로 설명할 수 있다. 가령 아래의 식과 같은 제곱형 손실함수(squared loss function)을 사용한다면 특정 표집법의 위험함수는 바로 추정량의 평균제곱오차(mean square error)가 된다.

$$l(\theta, \hat{\theta}_a(s)) = C(\theta - \hat{\theta}_a(s))^2 \tag{2.1}$$

위 (2.1)식과 같은 손실함수는 모든 가능한 표본들의 평균적인 위험을 나타내는데 적절한 식이다. 그러나 표본설계를 담당하는 연구자의 입장에서는 평균적인 위험보다 가장 큰 위험을 제거하는 것이 일차적으로 더 중요한 관심이 된다. 이런 경우라면 위험함수의 식은 위 (2.1)식보다는 다음의 (2.2)식이 더 적절하다고 볼 수 있다. (2.2)식

$$l(\theta, \hat{\theta}_a(s)) = \begin{matrix} 0 & \text{if } |\theta - \hat{\theta}_a(s)| \le M, \\ C & otherwise. \end{matrix} \tag{2.2}$$

의 손실함수를 사용할 경우 앞에서 소개한 $\gamma_a(M)=P(|\theta-\hat{\theta}_a(s)|>M)$ 확률을 고려하는 것이 더욱 바람직하다.

Ⅲ. 모의실험

2절에서는 두 가지 특정한 사례를 가지고 무작위표집법과 계통표집법을 비교하였는데 이 절에서는 모의실험을 통하여 더욱 다양하게 비교하고자 한다. Rao와 Sitter (1995)는 다음과 같은 모집단 모형을 제안한 바 있는데 이는 표본이론에서 자주 만나는 모집단의 형태를 반영하는 것이다.

$$y_i = \beta x_i + \sqrt{x_i} \cdot \varepsilon_i \quad ,$$

여기서 $x_i \sim \Gamma(g,h)$, $\varepsilon_i \sim N(0,\sigma^2)$을 따른다고 가정하는데 x_i 는 보조변수, y_i 는 관심변수를 나타낸다. 위 모형에 의하면 다음의 관계식들이 성립한다:

$$\mu_x = gh,\ \sigma_x^2 = gh^2,\ C_x = \sigma_x/\mu_x = 1/\sqrt{g}\ ,$$
$$\mu_y = \beta\mu_x\ ,\sigma_y^2 = \beta^2\sigma_x^2 + \mu_x\sigma_x,\ Corr(x_i, y_i) = \rho = \beta\sigma_x/\sigma_y\ .$$

$\beta=1$, $\mu_x=100$, $C_x=1$ 로 정하고 관심변수와 보조변수 사이의 상관계수인 ρ 값을 0.3에서 0.9 사이로 변화시켜 가면서 크기 $N=10,000$인 모집단을 생성하였다. 생성된 모집단을 기초로 하여 먼저는 무작위표집법으로 크기 100($n=100$)인 표본을 1000번 반복하여 추출하였다. 다음으로 모집단 단위들을 보조변수 x_i 의 크기 순으로 정렬시킨 후 계통표집법으로 크기 100인 표본을 추출하였다. 모평균은 101.5인데 추출된 표본에 의한 추정값과 모평균과의 차이를 기초로 하여 두 표집법을 비교하였는데 그 계산 결과가 다음의 <표 3.1>에 나와 있다.

<표 3.1>에 나타난 수치 중 밑줄이 쳐진 부분의 M값은 M_{srs}나 M_{sys}값이 된다. 모든 경우에 대해 $M_{srs} > M_{sys}$가 되는데 이는 계통표집법이 무작위표집법에 비해 바람직하지 못한 표본을 추출할 확률을 낮게 한다는 것을 나타낸다. 여기서는 보조변수와 조사변수의 상관계수가 변함에 따라 어떤 양상을 띠는지를 주목할 필요가 있다. ρ값이 0.7, 0.9 인 경우를 살펴보면 계통표집법을 사용할 경우 표본오차가 상당히 작은 표본의 추출 가능성이 무작위표집법에 비해 현저히 높은 한편 바람직하지 못한 표본이 추출될 가능성은 매우 낮다. ρ 값이 0.5 이하로 떨어지는 경우에는 표본오차가 작은 표본의 추출가능성은 크게 차이를 보이지 않는 반면 극단적으로 바람직하지 못한 표본의 추출가능성은 여전히 계통표집법이 낮음을 볼 수 있다. 일반적으로 대규모 조사에서 조사변수와 상관이 비교적 높은 보조변수의 활용이 가능하다는 사실을 고려한다면 이러한 결과는 계통표집법이 실용적인 측면에서 매우 유용한 표집법이라는 사실을 보여준다.

<표 3.1> 두 표집법의 바람직하지 못한 표본이 추출될 확률 비교

상관계수 (ρ)	오차 (M)	무작위표본 $\gamma_{srs}(M)$	계통표본 $\gamma_{sys}(M)$	상관계수 (ρ)	오차 (M)	무작위표본 $\gamma_{srs}(M)$	계통표본 $\gamma_{sys}(M)$
	10	0.777	0.780		10	0.608	0.530
	20	0.532	0.450		20	0.305	0.200
	30	0.354	0.270		30	0.125	0.090
	40	0.224	0.160		40	0.061	0.020
	50	0.146	0.090		50	0.019	0.010
$\rho=0.3$	60	0.093	0.090	$\rho=0.5$	59.3	0.005	0.000
	70	0.050	0.020		60	0.001	0.000
	77.4	0.025	0.000		71.8	0.000	0.000
	90	0.003	0.000				
	100	0.002	0.000				
	118.3	0.000	0.000				
	5	0.721	0.600		5	0.734	0.380
	10	0.475	0.270		10	0.455	0.030
	15	0.290	0.110		15	0.220	0.010
	20	0.171	0.030		18.6	0.098	0.000
	25	0.097	0.010		25	0.038	0.000
$\rho=0.7$	30	0.048	0.010	$\rho=0.9$	30	0.007	0.000
	35	0.020	0.010		35	0.001	0.000
	36.1	0.009	0.000		38.6	0.000	0.000
	40	0.003	0.000				
	43.7	0.000	0.000				

Ⅳ. 결 론

확률 표집법의 특성에 관한 대부분의 연구는 모든 가능한 표본의 반복추출을 전제하는 틀 위에서 이루어져왔다. 또한 서로 다른 표집법들 간의 비교는 주로 표집과정의 편리성, 추정의 효율 등의 측면에서 검토되어 왔다. 그러나 실제 표본조사에서는 어떤 표집법에 의해 단 하나의 표본만을 얻게 된다. 이 때 만일 추출된 표본이 바람직하지 못한 표본이라면 해당 표집법이 아무리 이론적으로 좋은 특성을 지닌다고 해도 소용없는 것이 될 것이다. 이런 면

에서 본다면 어떤 표집법의 성능을 평가할 때 추정의 효율성 뿐 아니라 바람직하지 못한 표본의 추출 가능성을 검토하는 것이 필요하다.

본 논문에서는 구체적인 몇 가지의 모집단들에 무작위표집법과 계통표집법을 각각 적용하여 표본을 추출했을 때 바람직하지 못한 표본이 추출될 확률이 어떻게 되는지를 살펴 보았다. 그 결과 조사변수와 비교적 상관이 높은 보조변수의 활용이 가능한 경우 계통표집법은 무작위표집법에 비해 바람직하지 못한 표본의 추출 가능성을 낮게 해주는 좋은 성질을 지니고 있음을 알 수 있었다. 대부분의 대규모 표본설계에서 실제 조사변수와 상관이 높은 보조변수들을 이용할 수 있기 때문에 위의 결과는 중요한 의미를 지닌다고 할 수 있다.

〈참 고 문 헌〉

(1) 박홍래. 2000. ≪통계조사론≫ (2판). 영지문화사.

(2) 통계청, 1997a. 《1995년 기준 도소매업 및 서비스업 통계조사보고서》, 통계청.

(3) 통계청, 1998. 《1996 도·소매업총조사보고서》, 통계청.

(4) Avadhani, M. S. and Sukhatme, B. V. 1965. Controlled Simple Random Sampling, *Journal of the Indian Agricultural Statistics,* Vol. 17, 34-42.

(5) Bellhouse, D. R. 1988. Systematic Sampling, *Handbook of Statistics,* Vol. 6, 125-145.

(6) Cochran, W. G. 1977. *Sampling Techniques (3rd ed.)*, New York: John Wiley.

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

(8) Holt, D. and Smith, T. M. F. 1979, Post Stratification, *Journal of the Royal Statistical Society A*, Vol. 142, 33-46.

(9) Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

(10) Neyman, J. 1934, On the Two Different aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection, *Journal of the Royal Statistical Society*, Vol. 97, 558-606.

(11) Rao, J. N. K. and Sitter, R. R. 1995, Variance Estimation under Two-phase Sampling with Application to Imputation for Missing Data, *Biometrika*, Vol. 82, 453-460.