

# 대규모 표본설계에서 확률비례 및 단순집락추출법 비교 - 경제활동인구 표본조사 사례를 중심으로 -

윤연옥<sup>1)</sup> 이상은<sup>2)</sup>

## 요약

확률비례추출법(pps sampling)에서 표본단위의 추출확률을 결정하기 위하여 크기측도(measure of size:MOS)를 사용한다. 통계청에서 실시하는 일부 대규모조사의 경우 각 표본단위들의 크기측도를 살펴보면 서로 매우 유사한 편임을 알 수 있다. 이런 경우 조사원 업무량을 고려하여 확률비례추출법 대신 단순집락추출법의 사용 필요성이 제시되었다. 본 논문에서는 1997년 통계청의 경제활동인구조사 표본설계에 두 가지 추출법을 적용하여 평균제곱오차(MSE)와 변이계수(CV)를 비교하였다.

주요용어: 확률비례추출법, 단순집락추출법, 크기측도.

## 1. 서론

단순집락추출법은 각 집락의 추출확률을 동일하게 주는 추출법이다. 반면 집락의 크기에 비례하여 집락들의 추출확률을 적용할 경우, 일반적으로 확률비례추출법(PPS sampling, Probability Proportional to Size Sampling)을 사용하게 된다. 확률비례추출법을 사용하기 위해서 각 집락의 크기를 미리 알아야 하는데 이를 나타내기 위해 크기측도(Measure of Size: MOS)를 사용한다. 크기측도를 이용하여 표본설계를 하면 자체가중표본(self-weighting sampling)을 얻게 된다.

통계청의 가구표본설계에서는 추출단위인 집락의 크기를 가구수에 비례한 크기측도로 나타낼 수 있는데 이 때 대부분 단위들의 크기측도는 서로 비슷한 값을 가지게 된다. 이러한 경우 가구수에 비례한 크기측도를 이용하는 확률비례추출법을 사용하는 것과 단순집락추출법을 사용하는 것에 큰 차이가 없을 것으로 기대된다. 반면 단순집락추출법을 사용할 경우 조사원들의 업무량이나 조사관리 측면에서 이점이 있다.

이 논문은 대규모 표본설계 중 하나인 통계청 가구표본설계에 두 가지 추출법(확률비례추출법, 단순집락추출법)을 적용하여 주요항목에 대하여 평균제곱오차(Mean Square Error)와 변이계수(CV)를 비교하는 것을 목적으로 한다.

1) (302-701) 대전시 서구 둔산동 920 정부대전청사, 통계청 조사관리과, 표본담당 사무관

E-mail: yyoon@nso.go.kr

2) (449-791) 경기도 수원시 팔달구 의의동 산 94-6, 경기대학교 응용정보통계학과, 조교수

E-mail: sanglee@stat.kyonggi.ac.kr

## 2. 가구표본설계 개요 및 크기측도 분석

### 2.1. 가구표본설계 개요

통계청 다목적 가구표본을 이용하는 월별 경상조사는 경제활동인구조사와 도시가계조사가 있다. 97년 가구표본설계는 인구총조사 10% 표본조사구를 표본추출틀로 사용하여 각 시도별 동부와 읍면부별로 주어진 오차의 한도에서 결정된 표본수 만큼 확률비례추출한다. 현재 통계청 가구표본은 1,231개 조사구(약 3만가구)로 이루어져있다. 참고로 조사구(ED, Enumeration District)는 인구주택총조사 실시 약 2년전 부터 작성되는 것으로, 처음 설정시 60가구를 기준으로 하여 지역적, 환경적으로 분리할 수 밖에 없는 경우를 제외하고는 거의 60가구에 맞추었다. 조사구내의 가구수는 수시로 변동되기 때문에, 인구주택총조사시에는 조사구설정 후 변동된 가구수가 반영되어 조사가 된다.

가구표본설계시에는 인구주택총조사 10% 표본조사구를 표본틀로 이용하므로, 총조사 당시의 가구수가 반영된 조사구를 사용하게 된다. 결과적으로 각 조사구당 가구수는 대부분이 기준 60가구에서 약간의 편차를 가지게 된다. 각 조사구는 8가구로 구성되는 구역(Segment)을 설정하게 되며, 이에 따라 각 조사구당 평균 7~8개의 구역으로 이루어지게 된다. 이때 조사구내의 구역수를 크기측도로 이용하여 조사구 선정에 사용하며 이는 확률비례추출법이 되도록 한다. 경제활동인구조사의 경우 선정된 조사구에서 8가구로 이루어진 구역설정 후, 임의로 정한 표본구역에서 북쪽의 구역부터 시계바늘 방향으로 서로 인접한 3개의 구역을 선정하여 조사된다. 이때 한 조사구당 조사원들의 업무량을 보면, 조사하는 3개의 구역은 모든 조사구에서 같은 업무량을 제공한다(모두 24가구조사). 그러나 비조사구역에 대한 관리가 있어서 가구수가 많은 조사구는 가구수가 적은 조사구보다 많은 업무량을 요구한다. 가구의 전출입, 철폐, 및 증설등으로 인하여 조사구내의 가구수는 항상 유동적이며, 이를 감안한 조사원들의 업무량 측정이 어려운 상황이어서 일반적으로 조사구단위로 업무량 할당이 이루어지게 된다. 이러한 상황에서 조사원들의 동일 업무량 배정에 대한 요구는 필연적이다. 이에 따라 조사원들의 업무량 측정 기준이 조사구인 점을 감안하여, 조사구내에 같은 가구수가 있는 것으로 간주하고 표본설계(단순집락추출법)를 하고, 표본조사구에 같은 가구수(60가구)가 배정되도록 조사구를 재조정하는 안건이 제기된다.

### 2.2. 크기측도 분석

각 집락당 크기측도가 유사한지를 알아보기 위하여 가구표본조사의 표본틀로 쓰이는 인구주택총조사 10% 표본조사구의 각 조사구당 가구수가 평균 60가구와 같은지를 알아보았다. 귀무가설  $H_0 : \mu = 60$ , 대립가설  $H_a : \mu \neq 60$  아래에서 시·도별로 분석한 결과 서울, 부산, 대구, 광주, 경기, 충북(읍면부) 및 전북(동부)에서는 조사구당 평균 60 가구에서 많이 벗어남을 보여주고 있다. 이것은 주로 신도시 개발이나 신축 아파트건축으로 인한 가구수의 급격한 증가에 영향을 받은 소수의 조사구들에 영향을 받아 그러한 결과를 얻은 것으로 분석된다. 그러나 대부분의 조사구는 60가구를 중심으로 하여 10가구 내의 편차를 가진 것으로 나타나고 있다. (부록의 상자그림표(box-plot) 참조).

표 2.1: 시도별 가구수 분석

시·도	Z-값	P-value	시·도	Z-값	P-value
서울	-20.5037	0.0001	충남 : 동부	-2.0528	0.0412
부산	-5.3194	0.0001	: 읍면부	-0.0857	0.9317
대구	-3.6024	0.0003	전북 : 동부	-6.2373	0.0001
인천	-1.9725	0.0488	: 읍면부	-2.0996	0.0364
광주	-2.8838	0.0041	전남 : 동부	-0.5793	0.567
대전	0.2234	0.8233	: 읍면부	0.9069	0.3648
울산	0.9879	0.3237	경북 : 동부	1.3228	0.1864
경기 :동부	-3.7373	0.0002	: 읍면부	0.5474	0.5843
: 읍면부	4.5383	0.0001	경남 : 동부	1.2633	0.2069
강원 : 동부	-0.0657	0.9477	: 읍면부	0.0604	0.9519
: 읍면부	0.2538	0.7999	제주 : 동부	-1.0351	0.3021
충북 : 동부	-0.7664	0.4439	: 읍면부	-0.5503	0.5835
: 읍면부	-2.8145	0.0052			

### 3. 두 추출법에 따른 추정

#### 3.1. 확률비례추출법

약 60가구로 이루어진 각 조사구의 가구수를 8로 나누어 반올림한 값을 조사구를 분할하는 구역수로 결정한다. 예를 들면 57가구이면 7, 63가구이면 8 구역수를 부여한다. 정해진 수의 표본조사구를 구역수에 비례하는 확률로 계통추출한다. 즉 크기에 따른 확률비례 계통추출법(PPS systematic sampling)을 적용하여 표본을 추출한다. 이때  $j$  번째 표본조사구의 추출확률,  $P_{e_j}$  는 구역수에 비례하므로 다음과 같다.

$$P_{e_j} = \frac{e_j}{E/n} = \frac{e_j}{\text{추출구간}} \tag{3.1}$$

여기서  $n$  = 표본조사구수,  $N$  = 총조사구수  
 $e_j$  =  $j$ 번째 조사구의 구역수,  $E = \sum_{j=1}^N e_j$ .

표본구역은 인접한 3개의 구역을 추출하며 표본구역내에서는 전 가구를 조사한다.  $j$ 번째 조사구의 표본구역 추출확률은  $P_{s_j} = (3/e_j)$  이다. 전체 표본추출률(overall sampling probability)은  $P_{e_j} \times P_{s_j}$  이며, 정리하면

$$P_{es_j} = P_{e_j} \times P_{s_j} = \frac{e_j}{E/n} \times \frac{3}{e_j} = \frac{3n}{E} \tag{3.2}$$

으로 일정하게 되는 자체가중표본(Self-Weighting sample)이다.

이 경우 어떤 특성  $Y$ 의 추정량,  $\hat{Y}$ 은 보조변수 인구수를 이용하여 비추정식에 의해 다음과 같이 표기할 수 있다.

$$\hat{Y} = \frac{X}{x}y \quad (3.3)$$

$$x = \sum x_j, y = \sum y_j$$

여기서  $X =$  기준 인구수

$x_j = j$ 번째 표본조사구에서 조사된 사람수

$y_j = j$ 번째 표본조사구에서 특성  $Y$ 를 갖는 것으로 조사된 사람수.

### 3.2. 단순집락추출법

조사구당 구역수가 유사하다는 것을 이용하여 각 조사구의 구역수가 같다고 간주하고 표본을 추출하는 방식이다. 이 방식은 위의 각 조사구의 가구수에 따라서 다른 확률을 부여하여 조사구를 추출(PPS Sampling)하는 방식에서 벗어나서 각 조사구에 동일한 추출확률을 부여한다. 즉 정해진 표본 조사구수를 일반적인 계통추출방식으로 추출한다. 그리고 추출된 조사구에서 8가구로 이루어진 구역을 설정하고 추가로 인접한 2개의 구역을 선정한다. 이 경우  $j$ 번째 조사구의 표본구역 추출확률은

$$P'_{es_j} = P'_{e_j} \times P'_{s_j} = \frac{n}{N} \times \frac{3}{e_j} = \frac{3n}{Ne_j} \quad (3.4)$$

여기서  $N =$  총조사구수,  $n =$  표본조사구수,

$e_j = j$ 번째 조사구의 구역수이다.

즉  $P'_{es_j}$ 은 조사구의 구역수에 따라서 다르게 된다.

이 경우 어떤 특성  $Y$ 의 추정량,  $\hat{Y}$ 은 보조변수 인구수를 이용한 비추정식에 의해 다음과 같이 표기된다.

$$\hat{Y} = \frac{X}{x'}y' \quad (3.5)$$

여기서  $x' = \sum_j M_j x_j$ ,  $y' = \sum_j M_j y_j$ ,  $M_j = (Ne_j)/(3n) = ke_j$

$X =$  기준 인구수

$x_j = j$ 번째 표본조사구에서 조사된 사람수

$y_j = j$ 번째 표본조사구에서 특성  $Y$ 를 갖는 것으로 조사된 사람수

$M_j = j$ 번째 표본조사구에 적용하는 승수

$k =$  상수( $= N/3n$ )

추정식 (3.5)를 다시정리하면 다음과 같다.

$$\hat{Y} = \frac{X}{x'}y' = \frac{X \sum_j ke_j y_j}{\sum_j ke_j x_j} = \frac{X \sum_j e_j y_j}{\sum_j e_j x_j} \quad (3.6)$$

#### 4. 추출방법에 따른 MSE비교

위의 두 가지 추출방법에 따라서 경제활동인구조사 표본을 추출하기 위하여 인구주택 총조사 10% 표본조사구 자료를 분류지표에 의하여 정렬하였다. 분류지표는 표본의 대표성을 높여주기 위하여 사용하는 것으로 주택특성, 농림어업 비율, 광공업 비율, 서비스업 비율, 행정구역이 사용된다. 더 자세한 내용은 통계청 가구부문 표본개편 보고서(1997)를 참고하기 바란다. 분류지표에 의해 정렬한 후 두가지 추출방법에 따라서 계통추출로 표본을 추출하였다. 두 추출법에 의하여 추출된 표본을 비교하기 위하여 경제활동인구조사에서 많이 사용되는 경제활동인구수, 취업자수, 실업자수 3개 항목에 대하여 평균제곱오차(MSE, Mean Square Error)와 변동계수(CV)를 계산하여 보았다. CV 계산을 위하여 모집단 특성치값이  $E(\hat{Y})$  과 같다고 가정하였다 ( $E(\hat{Y}) - Y = 0$ ). 이 경우  $MSE = Var(\hat{Y})$  이 되므로 CV 계산시 MSE를 사용하였다. 7개 특별시 및 광역시, 9개 도별 동부와 읍면부 지역으로 나누어서 비교하였으며 평균제곱오차와 변동계수는 다음과 같다.

$$MSE = E(\hat{Y} - Y) = \frac{\sum_i^a (\hat{Y} - Y)^2}{a} \tag{4.1}$$

$$CV = \frac{\sqrt{MSE}}{Y} \times 100 \tag{4.2}$$

- 여기서  $\hat{Y}$  : 특성치의 추정량
- $Y$  : 특성치의 모집단 값
- $a$  : 모든 표본추출 경우의 수

MSE 계산시 단순집락 계통추출에서는 모든 표본추출 경우의 수를 고려하였으며, 확률비례계통추출에서는 비교를 위하여 단순집락계통추출과 같은 경우의 수 만큼 표본추출을 하였다. 두 추출법의 지역별 CV 차이를 보면 크지 않을 뿐 아니라 반정도는 단순집락추출법이, 다른 반정도는 확률비례추출법이 더 작은 CV값을 나타내고 있음을 알 수 있다. 즉 두 추출방법에 따라서 어느 쪽이 더 많은 편향(bias)을 가지는지는 분명치 않다. 경제활동인구수와 취업자수의 CV 차이를 보면 대부분이 2%범위내에 있으며, 실업자수의 경우 앞의 두 항목보다는 많은 차이를 보이지만, 이것 역시 특정한 한 가지 추출방법이 아닌 두 추출방법이 반반 정도 더 적은 CV를 나타내고있다.

#### 5. 결론

통계청의 표본설계와 같이 대규모의 표본이며 특히 조사원들이 이를 기준으로 5년간 같은 조사를 하게 되는 경우 조사원들의 관리는 통계생산의 질적 수준을 높이는데 큰 역할을 하게 된다. 그러므로 조사구를 단위로 조사원들의 업무량을 평가하는 현 시점에서 업무량의 형평성을 고려하기 위해 제시된 동일 크기측도를 이용하는 단순집락추출법과 기존의 확률비례추출법을 비교하였다.

표 4.1: 경제활동인구수에 대한 분석

시도	표본 조사구	$\sqrt{MSE}$		CV(%)		
		단순집락	PPS	단순집락(A)	PPS(B)	A-B
서울	160	28159.39	33636.71	0.66775	0.79643	-0.130
부산	92	20960.80	19876.95	1.36249	1.29204	0.070
대구	72	15735.51	17764.04	1.59125	1.79638	-0.205
인천	93	6478.24	13858.14	0.69162	1.47950	-0.788
광주	87	4203.66	4044.64	0.89388	0.86007	0.034
대전	65	8206.79	7835.29	1.71726	1.59558	0.122
울산	42	6640.19	4666.26	1.79445	1.78692	0.008
경기 : 동부	72	31464.05	36780.92	1.28891	1.28303	0.006
: 읍면부	46	14347.53	18162.76	2.01205	2.38792	-0.376
강원 : 동부	37	6702.63	15356.12	2.13256	4.83335	-2.701
: 읍면부	34	8449.04	8225.12	2.87041	2.86183	0.009
충북 : 동면	33	8609.47	5810.84	2.85659	1.92801	0.929
: 읍면부	31	8992.56	19048.37	3.05578	6.47287	-3.417
충남 : 동면	32	4116.50	8109.85	2.04934	4.15716	-2.108
: 읍면부	38	11765.58	19171.40	1.93033	2.53891	-0.609
전북 : 동부	35	9649.33	14797.06	2.21391	3.53822	-1.324
: 읍면부	28	11204.47	7502.36	2.97896	2.40516	0.574
전남 : 동부	37	6484.94	6495.33	2.25665	1.90849	0.348
: 읍면부	30	22569.80	14927.43	3.67467	2.05740	1.617
경북 : 동부	32	9997.06	17952.71	2.10133	4.56977	-2.468
: 읍면부	39	13195.45	13207.66	1.79696	2.52131	-0.724
경남 : 동부	34	11921.64	17090.55	1.78076	2.39238	-0.612
: 읍면부	28	15447.48	14265.65	2.66837	3.08563	-0.417
제주 : 동부	20	4120.26	2585.30	2.69456	1.69073	1.004
: 읍면부	14	2005.58	2300.07	2.11715	2.42803	-0.311

표 4.2: 취업자수에 대한 분석

시도	표본 조사구	$\sqrt{MSE}$		CV(%)		
		단순집락	PPS	단순집락(A)	PPS(B)	A-B
서울	160	33746.23	40713.58	0.86839	1.04767	-0.179
부산	92	04269.82	19433.95	1.46410	1.39001	0.074
대구	72	13065.15	18239.97	1.44555	2.01810	-0.573
인천	93	6145.87	14357.43	0.70558	1.64831	-0.943
광주	87	6149.66	4928.71	1.43543	1.15044	0.285
대전	65	7381.82	7073.26	1.68155	1.59155	0.090
울산	42	7861.87	6773.48	2.24740	2.22733	0.020
경기 : 동부	72	37043.32	42521.44	1.61651	1.49698	0.120
: 읍면부	46	16722.75	19385.09	2.49296	2.79178	-0.299
강원 : 동부	37	7286.09	14032.69	2.53500	4.83476	-2.300
: 읍면부	34	10396.26	8036.19	3.76054	3.34989	0.411
충북 : 동면	33	9235.75	7263.69	3.33084	2.61962	0.711
: 읍면부	31	9361.64	20731.16	3.33856	7.39316	-4.055
충남 : 동면	32	5001.13	7536.74	2.68359	4.24660	-1.563
: 읍면부	38	11476.46	17087.62	1.98215	2.56566	-0.584
전북 : 동부	35	9219.12	12081.25	2.32466	3.50593	-1.181
: 읍면부	28	11138.34	8597.02	3.11797	2.89401	0.224
전남 : 동부	37	5453.21	6655.52	2.09384	2.22281	-0.129
: 읍면부	30	24203.40	14896.09	4.09991	2.53461	1.565
경북 : 동부	32	10700.82	17210.24	2.44484	5.08400	-2.639
: 읍면부	39	14393.09	15192.08	2.05012	2.79172	-0.742
경남 : 동부	34	13560.91	15703.75	2.19021	2.54319	-0.353
: 읍면부	28	17646.51	17233.64	3.19643	3.50265	-0.306
제주 : 동부	20	3688.31	2740.60	2.59795	1.90595	0.693
: 읍면부	14	1792.33	2073.24	1.97502	2.28456	-0.310

표 4.3: 실업자수에 대한 분석

시도	표본 조사구	$\sqrt{MSE}$		CV(%)		
		단순집락	PPS	단순집락(A)	PPS(B)	A-B
서울	160	19818.35	18719.50	5.9885	5.65646	0.332
부산	92	7066.24	9746.96	5.0365	6.94722	-1.911
대구	72	8220.53	6665.99	9.6644	7.83681	1.828
인천	93	4703.48	4049.11	7.1656	6.16866	0.997
광주	87	2790.78	1867.04	6.6685	4.46126	2.207
대전	65	4433.77	3739.70	11.3949	8.76735	2.628
울산	42	2172.98	2597.83	10.7467	14.2753	-3.529
경기 : 동부	72	13722.64	18235.14	9.1741	8.5640	-.610
: 읍면부	46	5106.62	3586.00	12.0781	10.5824	1.496
강원 : 동부	37	1682.60	2822.80	6.2597	9.6939	-3.434
: 읍면부	34	2190.01	1936.47	12.2211	14.2262	-2.005
충북 : 동부	33	2148.03	3438.46	8.9093	14.2615	-5.352
: 읍면부	31	3225.38	3088.20	23.2544	22.2653	0.989
충남 : 동부	32	1383.91	1128.36	9.5376	7.31030	2.227
: 읍면부	38	3947.80	3942.88	12.9351	15.7494	-2.814
전북 : 동부	35	3651.42	5853.81	9.2982	12.3885	-3.090
: 읍면부	28	2529.45	2420.58	13.3904	13.6699	-0.280
전남 : 동부	37	2128.40	1445.95	7.9034	7.07389	0.830
: 읍면부	30	5553.18	3711.35	23.2740	22.62148	0.659
경북 : 동부	32	4180.43	5166.58	10.9838	14.4257	-3.442
: 읍면부	39	5016.55	2964.80	15.5504	14.0882	1.462
경남 : 동부	34	6212.11	6051.49	12.3477	10.8177	1.530
: 읍면부	28	3988.11	6128.79	14.8588	15.6875	-0.829
제주 : 동부	20	2272.89	1472.94	20.7760	13.4638	7.312
: 읍면부	14	821.20	873.03	20.6331	21.9355	-1.302



표4.1 ~ 표4.3에서 보여준 것과 같이 두 추출방법중 일방적으로 적은 CV를 제공하는 방법은 없으며, CV 차이 또한 크지 않음을 알 수 있다. 조사구당 평균 60가구의 통계적 검증에서 차이가 있음을 보여준 서울, 부산, 대구, 광주 및 경기에서도 CV에 거의 차이가 없는 것으로 나타났다. 이것은 신축 아파트나 신개발지역에 영향을 받은 소수의 조사구로 인해 통계적 검증에서 유의하였으나 표본수가 큰 관계로 영향을 받지 않은 것으로 보인다. 오히려 균일 집락으로 검증받은 각 도의 동부와 읍면부 지역에서 7개 특별시 및 광역시보다 더 큰 CV 차이를 보였는데 이는 적은 표본수와 관계가 있는 것으로 판단된다.

그러므로 표본설계시 확률비례추출을 하는 것이 원칙이나 대표본인 경우, 조사원들의 업무량 및 표본관리 등을 감안한다면 집락의 크기가 크게 다르지 않으면 단순집락추출법을 하더라도 확률비례추출법과 크게 다르지 않음을 알 수 있다.

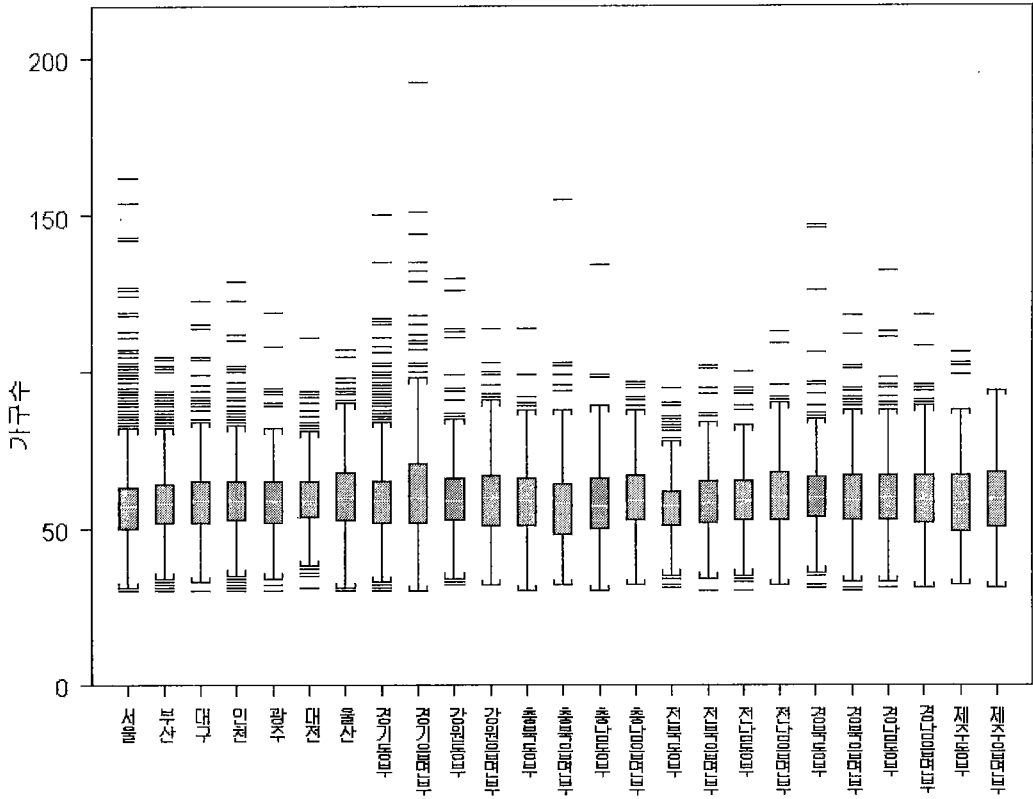
여기서 우리는 행정적 편의와 통계이론의 합의점을 찾았으며, 이를 계기로 통계이론이 통계 실무 즉 생산과정에 더욱 현실적으로 대응 할 수 있는 연구의 필요성을 제시해주었으며 또한 조사원들에 의한 실사단계에서 일어날 수 있는 비표본오차를 줄이는 효과를 기대할 수 있을 것이다.

## 참고문헌

- [1] 통계청. (1997). 가구부문 표본개편 보고서.
- [2] 박재수. (1989). 표본조사법(이론과 실제).
- [3] Cochran, W.G. (1963). Sampling Techniques, 2nd ed., New York: John Wiley & Sons, Inc.
- [4] Kish, L. (1965). Survey Sampling, New York: John Wiley & Sons, Inc.
- [5] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sample Survey Methods and Theory, New York: John Wiley & Sons, Inc. Vols, I and II.

[ 2000년 1월 접수, 2000년 8월 채택 ]

부록 : 시도별 조사구의 가구수 분포



# A Comparison of PPS and Simple Cluster Sampling in Large Scale Sampling - Based on Economically Active Population Survey Sample Design

Yeanok Yoon<sup>1)</sup> Sang Eun Lee<sup>2)</sup>

## ABSTRACT

In PPS sampling, measure of size(MOS) is used to determine the probability of selection of sampling unit. However, some large scale surveys conducted in NSO(National Statistical Office) showed that the sampling units have the similar MOS. In such case, simple cluster sampling method instead of PPS sampling is recommended to give the interviewers a similar work load. In this paper, MSE and CV of the above two sampling methods applied to the 1997 Economically Active Population Survey sample design are compared.

*Keywords:* PPS, Simple Cluster Sampling, Measure of Size.

---

1) Deputy Director, Field Management and Sampling Division, National Statistical Office.

E-mail: yyoon@nso.go.kr

2) Assistant Professor, Applied and Information Statistics Department, Kyonggi University.

E-mail: sanglee@stat.kyonggi.ac.kr