

통계청 가구표본설계

Sample Redesign of Household Survey conducted by Korea National Statistical Office

윤연옥* · 이명호**

통계청에서 매월 실시하는 주요 가구부문 경상조사에는 경제활동인구조사와 가계조사가 있다. 이 두 조사의 표본은 인구주택총조사 자료를 표본틀로하여 5년마다 개편되고 있다. 이 논문에서는 2000년 인구주택총조사를 바탕으로 2002년에 실행한 표본개편에 관한 내용이다.

Main monthly household surveys conducted by National Statistical Office are economically active population survey and household expenditure survey. Samples of these two surveys are redesigned every 5 years based on Census. This paper is about sample redesign of household survey conducted in 2002 based on 2000 Census.

1. 서론

통계청에서 매월 경상적으로 실시하는 가구대상 표본조사는 경제활동인구조사, 가계조사, 소비자전망조사 및 집세조사가 있는데, 이들 조사는 예산과 인력을 최소화하기 위해 동일 표본에서 병행 실시되고 있다. 1969년부터 현재와 같이 인구센서스 자료를 이용하여 다목적 표본이 설계되었으며 이후 일곱 차례에 걸쳐 표본 개편을 실시하였다.

경제활동인구조사는 15세 이상인 사람을 대상으로 고용통계 관련된 조사이며 매월 15일이 포함된 1주간을 조사대상 주간으로 하여 그 다음주에 조사를 실시한다. 가계조사는 가구를 대상으로 일정기간동안의 수입과 지출을 조사하기 위하여 가계부를 가구에 배부하여 기록하게끔 한다. 조사대상기간은 매월 1일부터 말일까지 1개월간이다.

현재 가구표본은 1995년 인구주택총조사 결과자료를 바탕으로 1997년 표본개편에 따라 추출한 것으로 1998년 1월부터 사용하고 있어 표본이 노후화 되어 있다. 총조사가 이루어지는 5년 기간 중에 발생하는 모집단의 변화를 표본에 반영하기 위한 표본관리를 지속적으로 하고 있으나, 여러 조사환경으로 인해 충분한 반영은 이루어지지 못하고 있다. 따라서 신도시 개발, 인구이동 및 경제·사회의 구조변화에 따른 대표성 있는 표본을 추출하여

* 통계청 조사관리과 사무관

** 통계청 조사관리과

표본의 현실 반영도를 제고하고 표본조사의 정도를 높이기 위해 5년마다 인구주택총조사 결과를 토대로 표본을 개편하고 있다.

표본개편과정을 보면 기존조사에서 지역별 표본오차를 분석하여 새로운 표본규모를 계산, 기존의 조사구 층화 분류지표 검토, 조사구당 표본가구수 최적화 등이 있다. 이번 표본개편의 주요변경사항으로는 다음과 같다.

- 1) 가계조사의 조사대상범위를 현행 동부(도시지역)에서 읍면부까지 확대
 - 서울지역 분기별 통계자료 작성 및 공표에 추가하여 기타 시·도의 주요 항목에 대한 통계작성을 고려
- 2) 응답자의 과중한 응답부담, 모집단의 변화반영 및 조사원들의 매너리즘을 예방하기 위하여 연동표본(교체표본, Rotation Sampling) 도입
- 3) 특·광역시외의 구, 각 도의 시 및 인구 10만 이상의 군은 소지역 추정법을 적용하여 분기별로 고용통계를 작성할 수 있는 기반을 마련
 - 또한 시·도단위의 통계 중 연령계층별, 교육정도별 통계를 소지역 추정법을 적용하여 작성할 수 있는 기반을 마련

II. 기존 표본 검토

1. 표본추출

1995년 인구주택총조사 10% 표본조사구 중 보통조사구만을 표본틀로 이용하였다. 1차 층화로써 특별시와 광역시의 6대 도시와 9개 도로 나누었으며 2차 층화로 9개 도에 대하여 시부 및 군부로 나누었다. 조사구의 분류지표는 동부와 읍면부별로 다른 분류 지표를 적용하였지만 동부와 읍면부 각각에 대하여는 동일한 분류지표를 적용하였다. 이 때 동일한 분류지표를 부여함으로써 각 층별 특성 반영이 다소 미흡한 결과를 가져왔다.

지역별 표본오차와 조사인력을 고려하여 조사구당 평균 24가구를 조사한다는 전제하에 결정된 시·도별 표본조사구 수에 5년간 사용하면서 모집단의 변화를 반영하기 위해 표본을 보완한 결과 현재의 표본조사구 수는 다음과 같다.

< 표 1> 현행 표본의 표본조사구 수

시 도	경 제 활 동			도시가계	시 도	경 제 활 동			도시가계
	계	시부	군부			계	시부	군부	
전 국	1,246	1,062	184	697	강 원 총	71	50	21	27
서 울	160	160	-	127	북	66	39	27	27
부 산	92	92	-	70	충 남	72	49	23	30
대 구	74	74	-	49	전 북	63	49	14	29
인 천	95	95	-	50	전 남	67	42	25	29
광 주	89	89	-	52	경 북	73	55	18	32
대 전	65	65	-	50	경 남	62	47	15	34
울 산	42	42	-	23	제 주	34	12	14	18
경 기	121	94	27	50					

지역별 표본규모에 따라 경제활동인구조사는 조사구 가구수에 비례하는 확률비례계통으로 조사구를 추출하였으며, 경제활동인구조사의 동부 표본조사구수와 도시가계조사의 표본조사구수가 같은 지역은 경제활동인구조사의 표본조사구에서 도시가계조사를 실시하고 경제활동인구조사의 표본조사구가 많은 지역은 경제활동인구조사의 표본조사구 중에서 도시가계조사 표본조사구를 계통추출하였다.

조사구내 거처는 8가구로 이루어진 구역으로 나누어 경제활동인구조사는 임의로 추출된 1구역을 포함하여 3개의 구역을 추출하였으며, 추출된 구역내의 가구는 전수로 조사하였다. 도시가계조사의 표본조사구는 평균 10가구를 기준으로 구역을 분할한 후 경제활동인구조사의 3개 조사구역 안에 포함된 도시가계 구역 중 구역번호가 가장 빠른 한 구역을 조사 구역으로 추출하였다.

2. 추정방법

1) 경제활동인구조사

각 시·도의 어떤 특성 Y에 관한 성별 및 성별 합계의 월별 추정치를 먼저 산출한 다음 각 시·도의 월별 추정치를 합산한 결과를 전국의 월별 추정치로 공표하고 있다. 각 시·도에서 어떤 특성 Y를 갖는 성별 및 성별 합계의 월별 추정치를 sY_{qgh} , Y_{qgh} 라 하면, 이들 추정치를 산출하는 방법은 다음과 같다.

$$\begin{aligned}
 sY_{qgh} &= \sum_i sY_{qghi} \\
 &= \sum_i M_{qghi} sY_{qghi}
 \end{aligned}$$

$$Y_{qgh} = \sum_s s Y_{qgh}$$

위의 공식에서

$$M_{qghi} = sX_{qghi} / sX_{qghi}$$

$$sY_{qghi} = \sum_j sY_{qghij}$$

$$sX_{qghi} = \sum_j sX_{qghij}$$

- Y = 특성 Y를 갖는 추정인구
- Y = 표본에서 특성 Y를 갖는 것으로 조사된 인구
- M = 승수
- X = 15세 이상 추계인구
- X = 표본에서 조사된 15세 이상 인구
- s = 성별을 나타내는 첨자 ($s = 1,2$)
- q = 분기를 나타내는 첨자 ($q = 1,2,3,4$)
- g = 월을 나타내는 첨자 ($g = 1,2,3$)
- h = 시도를 나타내는 첨자 ($h = 1,2,\dots,15$)
- i = 시부 및 군부를 나타내는 첨자 ($i = 1,2$)
- j = 표본조사구를 나타내는 첨자

나) 도시가계조사

가구당 월 평균 가계수지의 분기별 공표를 위한 전 도시의 분기별 추정치는 각 시·도의 분기별 추정치를 먼저 산출한 다음 이를 다시 평균하여 작성한다. 전국의 연간 월 평균 가계수지는 각 시·도별로 분기별 월 평균을 이용하여 시·도의 연간 월 평균을 추정하고 이를 다시 이용하여 구한다. 각 지역의 어떤 특성 X를 갖는 가구당 특성 Y의 월 평균에 대한 분기별 월 평균 추정치를 \bar{Y}_{qh} 라 하면 이를 구하는 공식은 다음과 같다.

$$\begin{aligned} \bar{Y}_{qh} &= Y_{qh} / X_{qh} \\ &= \sum_i Y_{qhi} / \sum_i X_{qhi} \\ Y_{qhi} &= \sum_g g Y_{qghi}, \quad X_{qhi} = \sum_g g X_{qghi} \end{aligned}$$

- \bar{Y} = 특성 X를 갖는 가구의 특성 Y의 가구당 월평균 추정치
- X = 특성 X를 갖는 것으로 조사된 가구수
- Y = 특성 Y에 대해 조사된 값
- q = 분기를 나타내는 첨자 ($q = 1,2,3,4$)
- g = 월을 나타내는 첨자 ($g = 1,2,3$)
- h = 시도를 나타내는 첨자 ($h = 1,2,\dots,15$)
- i = 표본조사구를 나타내는 첨자

3. 표본유지 관리방법

1) 표본관리방법

표본으로 한번 선정되면 차기 표본 개편 시까지 계속 조사를 실시하는 고정표본관리방법(Fixed Sampling)을 채택하였다. 이는 표본의 교체에 따른 통계 시계열의 불안정성과 조사에 쉽게 응하지 않는 문화적 환경 때문에 이루어졌다. 고정 표본의 사용은 응답자로 하여금 무조건 5년간 가구부문의 각종 통계조사에 응하게 함으로써 응답 부담을 가중시킨다.

2) 전입 및 전출가구에 대한 조치

조사구역 내에 새로운 가구가 전입하였을 때는 전입 이후부터 조사 대상 가구로 하여 조사를 실시하였고, 전출 가구에 대하여는 조사를 중지하였다. 조사지역 내에 주택이 신축되어 가구가 전입하였을 경우에도 조사 대상 가구로 하였다.

3) 신축 아파트에 대한 조치

1995. 11. 1 이후에 준공 검사가 완료된 아파트를 포함하여 1년마다 주기적으로 전국의 신축 아파트에 대한 자료를 수집하여 정리한 후 15개의 새로운 조사구를 표본 조사구로 추가하였다. 이는 조사 모집단의 규모 증가와 기본 구조변화를 표본에 반영시키기 위해서였다.

III. 새로운 표본설계

1. 주요개선사항

대내외적 통계환경의 변화에 부응하기위해 금번 표본 개편 시에 반영한 주요사항은 1) 시군구별 통계 또는 세부 단위별 통계를 작성할 수 있는 소지역추정기법을 개발하기 위한 기반을 반영 2) 응답자의 부담, 모집단변화반영, 조사원의 매너리즘방지 등의 목적으로 기존의 고정표본에서 연동표본(rotation sampling)을 도입할 수 있는 표본구조를 반영하는 것이다.

소지역추정기법 개발 도입을 위한 기반으로 특·광역시, 구, 각 도의 시 및 인구 10만 이상인 군을 대상으로 적어도 4개 이상의 표본조사구가 추출되도록 한다. 연동표본 모형은 매월 1/36 표본이 교체되는 방식으로 응답기간은 5년에서 3년으로 감소시킨다. 교체단위는 조사구역(5가구)으로 한다.

2. 모집단

표본 설계 시의 범위(Coverage) 오류를 최소화하기 위해 모집단의 기초자료는 2000년 인구주택총조사 결과 자료를 사용하는데, 이 중 모집단의 분석이 가능한 10% 표본 자료를 모집단으로 설정하였다. 이들 자료는 평균 60가구를 중심으로 조사구(EU; Enumeration Unit)로 형성되어 있으며, 10% 표본 조사구의 종류에는 일반가구로 형성된 아파트 조사구 및 보통 조사구, 섬지역에 거주하는 가구를 묶은 섬조사구, 기숙 및 특수사회시설을 묶은 기숙 및 특수사회시설 조사구가 있다. 이들 조사구 중 실제 조사가 매우 곤란한 섬 조사구와 일반적 특성을 갖지 못하는 2,691개의 기숙시설 조사구와 1,201개의 특수사회시설 조사구를 제외한 24,998개 조사구(1,420,418가구)를 조사모집단(Survey Population)으로 사용하였다.

또 2000년 인구주택총조사 실시 이후 표본 개편 시까지의 모집단 변화를 반영 하기 위해 2000년 11월 1일부터 2001년 10월 31일까지 신축된 약 238,000가구의 아파트 자료를 보조 자료로 사용하였다.

3. 층화

층화의 기본 목적은 조사의 목적에 따라 모집단의 대표성을 높이기 위한 것으로 본 표본 설계에서는 통계의 공표 단위에 따라 층 간의 독립적 추정이 가능한 7대 도시(특별시와 광역시)와 9개 도로 우선 층화한 후 각 도에서는 동부와 읍면부의 고용 형태 및 소득·소비 구조에 차이를 보이는 점을 감안하여 다시 동부와 읍면부로 층화하여 최종적으로 전국을 25개로 층화하였다.

4. 표본규모결정을 위한 특성항목 및 목표정도 결정

표본의 크기가 필요 이상으로 크면 예산이 낭비가 되고, 반대로 너무 작으면 표본의 정도가 떨어져 좋지 않다. 따라서 표본의 크기는 표본 개편의 기본 방향과 목표 정도를 충족하는 범위 내에서 최적 규모를 산정하였다.

표본 조사의 목표 정도란 추정값의 허용 오차를 통계 이용가치의 관점에서 정하는 것으로 본 표본 설계의 허용오차는 상대표준오차(변이계수: CV)를 이용하였다. CV의 값은 추정치를 1 또는 100으로 했을 때의 표준오차의 상대적인 크기를 나타내는 것으로 표준오차의 크기를 상대적으로 비교할 때 사용한다.

1) 경제활동인구조사

표본규모 산출에 사용할 특성 항목을 정하기 위해 2000년 경제활동인구조사 결과의 연평균

시·도별 주요 항목별 상대표준오차(CV)를 분석한 결과, 대도시와 경기도의 경우 농림어업의 상대표준오차가 상대적으로 큰 것은 도시 지역의 경우 농가가 아주 희박한 요소로 농가가 많지 않으며 그나마 일부 변두리 지역에 산재하여 있기 때문에 조사구간 분산이 상당히 크기 때문이다. 경기도의 경우에는 시가 전국에서 가장 많을 뿐 아니라 시의 인접 읍면에 농어가 골고루 분포하지 않기 때문이다. 이를 표본 규모 산정의 특성 항목으로 선정할 경우에는 매우 큰 표본이 요구되므로 표본 조사의 취지에 어긋난다.

<표 2> 2000년 경제활동인구조사 주요항목별 상대표준오차(연평균)

	경제활동인구	취업자			실업자	비경제활동인구	
			농림어업	광공업			서비스업
서울	0.28	0.28	11.81	1.11	0.37	2.08	0.45
부산	0.41	0.43	7.18	1.65	0.55	2.46	0.57
대구	0.45	0.45	9.44	1.55	0.64	2.97	0.64
인천	0.36	0.38	15.48	1.20	0.55	3.02	0.56
광주	0.45	0.48	4.58	2.10	0.57	2.71	0.57
대전	0.44	0.44	6.52	2.87	0.68	3.59	0.61
울산	0.67	0.68	14.81	2.02	1.70	5.97	0.99
경기	0.36	0.37	3.95	1.28	0.57	3.03	0.57
강원	0.56	0.61	4.13	3.70	1.02	5.56	0.77
충북	0.71	0.73	2.40	2.40	1.32	4.49	1.03
충남	0.63	0.65	2.48	3.22	1.42	4.57	1.07
전북	0.57	0.58	2.00	2.75	0.95	4.33	0.80
전남	0.54	0.58	3.65	3.72	1.73	4.33	0.95
경북	0.47	0.48	2.46	2.79	1.26	4.42	0.86
경남	0.61	0.58	3.64	2.67	1.15	4.22	0.93
제주	0.57	0.57	2.72	4.05	0.88	5.46	1.18

여러 항목을 조사하는 통계의 표본 규모를 산정할 경우에는 CV가 큰 항목을 선정하는 것이 일반적이다. 그러므로 전반적으로 CV가 타 항목에 비해 큰 실업자를 표본 규모 산정의 특성 항목으로 하는 것이 바람직하다.

목표 정도는 현행 표본의 상대표준오차를 고려하여 실업자수의 연평균 상대표준오차 기준으로 전국의 목표정도를 1%이하, 시·도 단위에서는 서울시 2%이내, 광역시는 3~5%, 도의 경우는 경기 2%이내, 기타 도는 3~5%이내로 설정하였다.

2) 전국가계조사

가계조사에서 표본 규모 산출에 사용할 항목으로는 전가구의 상대표준오차와 근로자 가구의 상대표준오차가 있는데, 근로자 가구의 상대표준오차를 사용하는 것이 보다 합리적이다. 이는 가계조사에서 전가구에 대한 가계수지도 중요하지만 모든 면에서 근로자 가구의 가계수지에 더 비중을 두고 있는 현실을 고려한 것이다.

근로자 가구의 가계수지 항목 중에서는 소비 지출 항목을 소득 항목보다 중요 항목으로 다루고 있고, 현행 표본의 표본 규모 산정의 특성 항목으로 사용되었기 때문에 현행 표본과의 연계를 위해 본 표본 설계에서도 근로자 가구의 소비지출 항목을 기준으로 목표 정도를 설정하였다. 현행 상대표준오차를 고려하여 분기별 상대표준오차가 서울시는 3%이내, 광역시는 5%이내, 각 도는 7%를 유지할 수 있도록 하였다.

5. 표본규모 결정

1) 경제활동인구조사

최종 표본 가구수는 1차 추출단위(PSU)인 표본 조사구 수와 2차 추출단위(SSU)인 표본 조사구 내 표본 가구 수의 곱으로 나타난다. 따라서 표본 조사구수를 산출하기 전에 먼저 조사구당 조사구역 및 가구 수를 결정하여야 한다. 기존 표본에서는 조사구당 평균 8가구를 한 구역으로 하여 3구역을 표본 조사 구역으로 선정하고 해당 조사구역 내의 모든 가구(평균 24가구)를 표본 가구로 선정하는 방식에 의해 1,246개 표본 조사구에서 약 29,904가구를 표본으로 하고 있다.

그러나 금번 표본 개편에서는 고려해야 할 주요 사항인 시군구 고용통계를 위해 우선 적용 대상 시군구에 최소 4개의 조사구 배분과 연동표본 적용을 위해서는 각 층별로 표본 조사구 수가 9의 배수로 이루어져야 한다. 이러한 조건을 조사구당 가구수의 조정없이 그대로 적용할 경우에는 최종적인 표본 가구 수가 대폭 증가하게 되어 인력, 조사비용 등 여러 여건상 조사가 어려워지게 된다.

따라서 표본 조사구 수의 증가가 불가피한 현 상황에서 최종 표본 가구의 대폭적 증가를 줄이기 위해서는 조사구당 가구 수의 감소를 고려치 않을 수 없는데, 또 조사구당 가구 수가 너무 적은 경우에는 Cluster(집락)이라는 개념을 갖는 조사구역의 의미가 퇴색하게 된다. 실업률 2%를 기준으로 할 때 50명당 1명의 실업자가 한 조사구내에서 조사될 수 있다고 본다면 1가구당 15세 이상 인구가 약 2.5명이므로 20가구가 있으면 50명의 조사대상을 확보할 수 있게 된다.

<표 3> 경제활동인구조사 층별 표본규모 (단위 : 개, %)

시 도	시군구 수	현행 조사구수	신표본		시 도	시군구수	현행 조사구수	가 구	
			조사구수	예상CV				조사구수	예상CV
전 국	234	1,246	1,629	0.88					
서 울	25	160	207	2.00	강 원 동부 읍면부	18(7)	71 37 34	90 45 45	5.40
부 산	16	92	117	2.40	충 북 동부 읍면부	12(4)	66 35 31	81 45 36	4.49
대 구	8	74	90	2.90	충 남 동부 읍면부	16(6)	72 34 38	90 45 45	4.57
인 천	10(8)	95	108	3.02	전 북 동부 읍면부	14(6)	63 35 28	81 45 36	4.33
광 주	5	87	81	3.00	전 남 동부 읍면부	22(5)	69 39 30	90 45 45	4.30
대 전	5	65	81	3.45	경 북 동부 읍면부	23(11)	73 33 40	108 54 54	4.00
울 산	5	42	63	5.25	경 남 동부 읍면부	20(10)	62 34 28	99 54 45	3.70
경 기 동부 읍면부	31(27)	121 75 46	198 135 63	2.60	제 주 동부 읍면부	4(2)	34 20 14	45 27 18	5.30

주) ()안의 숫자는 시 또는 구의 수입

통계 이론적으로도 동일 표본 가구 수라면 조사구 수를 증대하는 것이 분산을 감소시킬 수 있으므로 조사구당 가구 수는 20가구로 결정하고, 조사구 내의 조사구역 수는 연동구조를 감안하여 1구역당 5가구씩 4구역으로 결정하였다.

최종적으로 현행 표본의 표준오차, 목표정도, 소지역 추정 기반 마련 및 연동 구조 등을 고려하여 다음과 같이 층별 표본 규모를 결정하였다.

2) 전국가계조사

각 도의 동부, 읍면부별로 조사구당 적격가구수가 동부는 평균 47.4가구인 반면 읍면부는 25.6가구로 차이가 상이하여 별도로 산출하였으며, 조사의 안정성을 확보하기 위하여 조사구당 예상 적격 가구 수가 3가구 미만인 조사구는 제외하였다.

가. 동부

가계조사의 표본 규모 산출의 특성 항목으로 결정한 근로자가구 소비지출항목의 2001년 분기별 상대표준오차를 계산한 결과는 다음과 같다.

<표 4> 근로자가구 소비지출의 상대표준오차 (단위 : %)

	1/4분기	2/4분기	3/4분기	4/4분기	평균
전국	1.46	2.08	1.60	1.35	1.62
서울	2.95	2.51	3.25	2.75	2.87
부산	3.65	3.65	3.50	3.01	3.45
대구	4.96	7.02	5.84	4.65	5.62
인천	3.8	3.03	2.95	2.59	3.09
광주	6.09	4.87	4.43	4.90	5.07
대전	6.69	5.46	6.59	5.92	6.17
울산	3.99	5.35	4.34	2.76	4.11
경기	4.67	8.91	5.20	4.15	5.73
강원	6.99	9.32	5.70	8.31	7.58
충북	6.4	5.94	5.27	5.62	5.81
충남	8.21	6.04	6.83	6.77	6.96
전북	3.94	4.97	7.06	9.27	6.31
전남	6.29	4.89	4.45	7.36	5.75
경북	6.19	3.84	3.93	4.71	4.67
경남	5.56	6.03	5.00	5.67	5.57
제주	3.72	4.66	6.14	7.39	5.48

<표 4>에 나타난 바와 같이 분기별로 다소의 차이가 있다. 이 중 각 지역별로 상대표준오차가 큰 분기를 택하여 표본 규모를 선정하게 되면 표본 규모가 상당한 규모로 커져 조사 인력 등의 문제로 이를 수용하기는 어렵다. 그래서 각 지역별 분기별 상대표준오차를 평균하여 사용하면 안정적이 되므로 본 표본설계에서는 평균 상대표준오차를 표본 규모 산출에 사용하였다. 이를 가지고 목표 상대표준오차 따른 표본규모를 다음과 같이 산출하였다.

<표 5> 목표상대표준오차에 따른 예상 표본규모

시 도	현 행		목표CV에 따른 예상 표본규모					
	조사구수	CV	2%	3%	4%	5%	6%	7%
전 국	697	1.62	4,374	1,943	1,095	695	485	360
서 울	127	2.87	262	116	65	42	29	21
부 산	70	3.45	208	93	52	33	23	17
대 구	49	5.62	387	172	97	62	43	32
인 천	50	3.09	119	53	30	19	13	10
광 주	52	5.07	334	149	84	53	37	27
대 전	50	6.17	476	211	119	76	53	39
울 산	23	4.11	97	43	24	16	11	8
경 기	50	5.73	410	182	103	66	46	34
강 원	27	7.58	388	172	97	62	43	32
충 북	27	5.81	228	101	57	36	25	19
충 남	30	6.96	363	161	91	58	40	30
전 북	29	6.31	289	128	72	46	32	24
전 남	29	5.75	240	107	60	38	27	20
경 북	32	4.67	174	78	44	24	19	14
경 남	34	5.57	264	117	66	42	29	22
제 주	18	5.48	135	60	34	22	15	11

위의 <표 5>에서 목표 정도에 맞는 표본 규모와 연동표본을 위한 9의 배수 구조에 따라 아래<표 6>와 같이 동부의 표본 규모를 결정하였다.

<표 6> 표본규모(동부)

시 도	표 본 규 모		시 도	표 본 규 모	
	조사구수	예상CV		조사구수	예상CV
전 국	765		경 기	72	4.78
서 울	126	2.88	강 원	36	6.56
부 산	72	3.40	충 북	27	5.81
대 구	54	5.35	충 남	36	6.35
인 천	54	2.97	전 북	27	6.54
광 주	54	4.98	전 남	27	5.96
대 전	54	5.94	경 북	36	4.40
울 산	36	3.29	경 남	36	5.41
			제 주	18	5.48

나) 읍면부

읍면부의 경우는 기존 조사 결과가 없어 모집단의 특성 분석을 할 수 없으므로 표본

의 대표본 이론과 연동 구조를 감안 우선 각 도별로 27개의 조사구 수를 배정한 후, 모집단의 가구 수를 고려하여 경기도의 경우는 9개의 조사구를 추가한 36개로, 제주도의 경우에는 9개의 조사구를 감소한 18개 조사구를 표본으로 조정하였다.

6. 표본조사구 추출

1) 기초자료의 집계

2000년 인구주택총조사 10% 표본조사구 중 섬조사구 및 시설단위 조사구를 제외한 24,998개의 아파트 및 보통 조사구를 추출단위 조사구로 하여, 추출단위 조사구별로 70개 특성에 대해 기초 자료를 집계하였다.

2) 특성지표의 선정

특성 지표 적용은 추출된 표본이 표본 조사의 특성을 반영할 뿐만 아니라, 모집단의 구조와 가장 흡사한 표본을 추출하기 위해 사용하였는데, 특성지표를 선정하기 위해 실업자 수와 특성 항목 간의 ANOVA 분석에 의해 상관 관계를 분석한 결과 상관이 있는 것으로 판명된 30개 항목을 특성 지표로 삼았다.

<표 7> 30개 특성지표 항목

대분류	특 성 항 목
가 구 (8개)	농어가 가구수(비농어가) 주택유형별 가구수(아파트, 단독, 다가구 및 연립) 가구구분별 가구수(월연가구) 주택점유형태별 가구수(자가) 100가구당 승용차 보유대수 100가구당 컴퓨터 보유대수
인 구 (9개)	성별 인구수(남, 여) 연령별 15세이상 인구수(15~19세, 20~29세, 30~59세, 60세이상) 교육정도별 15세이상 인구수(중졸이하, 고졸, 대졸이상)
경활상태 (13개)	경활상태별 인구수(경활, 취업자, 실업자, 비경) 산업별 취업자(농림어업, 광공업, 서비스업) 직업별 취업자(전문기술, 사무, 서비스판매) 종사상 지위별 취업자(근로자, 자영자, 사업주)

3) 표본추출단위 조사구명부 작성

계통추출에 있어 추출 단위 조사구 명부를 어떤 기준에 의해 정렬치 않고 추출하는 것 보다는 기준에 의해 명부를 정렬한 후 계통추출하는 것이 표본의 대표성을 제고할 수 있다. 분류지표 종류가 4개 이상이 되는 경우에는 각 셀에 해당하는 조사구가 적어 오히

려 비효율적이 된다. 그래서 실업자 및 가구소득과 가장 상관이 있다고 판단되는 특성을 뽑아 이들을 조합하여 각 층별로 추출 작업을 시뮬레이션하여 모집단의 특성 항목인 30개 항목별 구조에 가장 흡사한 조합을 선택하였다. 최종적으로 각 층별로 선택된 분류기준은 다음과 같다.

<표 8> 분류기준 현황

	분류기준 1	분류기준 2
1차 분류	주택특성 1. 아파트 2. 단독가구 3. 다가구, 연립 및 기타	주택특성 1. 아파트 2. 단독가구 3. 다가구, 연립 및 기타
2차 분류	산업특성 1. 농림어업 2. 광공업 3. 서비스업	산업특성 1. 농림어업 2. 광공업 3. 서비스업
3차 분류	승용차 및 컴퓨터보유율 1. 평균이상 2. 평균이하	승용차 및 컴퓨터보유율 1. 평균이상 2. 평균이하
4차 분류	행정구역 및 조사구 번호	실업자수 1. 실업자비율이 층별 평균이상 2. 실업자비율이 층별 평균이하
5차 분류	-	행정구역 및 조사구 번호
적용 층	광주, 대전, 경기동부 및 읍면부, 강원 읍면부, 충남 읍면부, 전남 읍면부, 경남 동부 및 읍면부, 제주 동부 및 읍면부(11개 층)	서울, 부산, 대구, 인천, 울산, 강원 동부, 충북 동부 및 읍면부, 전북 동부 및 읍면부, 전남 동부, 경북 동부 및 읍면부(14개 층)

조사구를 위의 분류기준에 의해 각 층별로 정렬한 표본추출단위 조사구 명부를 작성하였다.

4) 표본조사구 추출

표본조사구 추출은 각 층별로 구해진 표본 수만큼을 가구수 크기에 비례하는 비율로 추출하는 방법인 확률비례추출방법(PPS : Probability proportional to size sampling)에 의해 추출하였다. PPS 방법을 사용한 까닭은 조사구가 동일한 가구수로 형성되어 있지 않고, 각기 다른 크기의 가구로 구성되어 있기 때문이다. 다른 크기의 조사구를 추출할 때 PPS 방법을 쓰면 최종 추출되는 가구가 동일한 가중치를 갖게 되는 자체가중표본(Self-weight sampling)이 되게 된다. 자체가중표본이 되면 추정 시 아주 유리해지는 장점을 갖게 된다.

또한 모집단의 구조에 가장 유사한 표본을 추출하기 위해 시뮬레이션을 통해 최적의 표

본을 추출하였다. 이 때 가장 최적인 표본을 선택하기 위해서 다음과 같은 최적계수를 사용하였다.

$$\text{최적계수} = \sum_{i=1}^{30} (R_i - \overline{R}_i)^2 * \frac{1}{S_i^2}$$

여기에서 R_i : 해당 표본에서의 i 항목의 비율

\overline{R}_i : 모집단에서의 i 항목의 비율

S_i : 추출 가능한 모든 표본에서의 i 항목의 표준편차

5) 전국가계조사 표본조사구 추출

경제활동인구조사를 위해 추출된 표본 조사구 중에서 전국가계조사를 위한 각 층별 표본규모에 해당하는 표본을 위의 방법과 같이 시뮬레이션에 의해 모집단의 구조에 가장 유사한 표본 조사구를 선정하였다.

이 때, 가계조사의 경우에는 조사구 내의 적격 가구수가 4가구 이상이 확보되어야 바람직하나 전북 읍면부와 전남 읍면부에서는 경제활동인구조사의 표본조사구중 예상 적격 가구 비율이 30% 이상인 조사구 수가 27개 미만이어서 표본수를 조정하지 않을 수 없었다. 전남·북의 동부는 당초 27개에서 36개로, 읍면부는 27개에서 18개로 각각 조정하였다.

7. 표본조사구내 구역 및 표본가구 선정

확정된 표본 조사구에 대하여 2000년 인구주택총조사 당시의 조사구요도 및 가구명부를 기초로 현지 확인을 통하여 현 상태에 대한 조사구요도와 가구명부를 재작성하였다. 경제활동인구조사는 5가구씩 구역으로 분할하여 설정한다.

경제활동인구조사는 각 조사구별로 4개의 구역을 각각 표본으로 추출하되 중앙에서 임의로 지정한 표본구역부터 시작하여 북쪽의 구역부터 시계바늘 방향으로 서로 인접한 구역을 추출한다. 표본으로 추출된 구역에서는 경제활동인구조사를 전수조사로 실시한다.

전국가계조사를 실시하는 조사구의 경우에는 경제활동인구조사의 4개 조사구역 중 첫 번째 구역과 두 번째 구역을 조사구역으로 지정하였으며, 이 구역 내에서 단독가구, 농가 등이 제외된 적격가구만을 조사대상가구로 한다.

IV. 추정

추정식은 기존표본과 같은 비추정 방식으로 하며 연동표본 관련된 복합추정식은 현재 개발 중에 있다. 자세한 내용은 통계청 가구부문 표본개편보고서를 참고하기 바란다.

V. 결론

통계청에서 매월 실시하는 경상 가구표본은 인구주택총조사 자료를 표본틀로 하여 5년마다 표본이 개편되고 있다. 이 논문은 2000년 인구주택총조사 자료를 바탕으로 2002년에 실시된 표본개편에 관한 내용이다. 2002 표본개편에 반영된 주요 내용으로는 연동표본의 도입과 인구 10만 이상의 시군구에 대한 소지역을 추정할 수 있게끔 하기 위한 기반을 만들어 놓았다. 기존의 고정표본방식에서는 응답부담이 5년인데 반해 도입되는 연동표본에서는 3년의 응답부담으로 줄어든다. 또한 모집단을 보다 더 잘 반영하는 표본조사구를 추출하기 위하여 과거에는 동부와 읍면부로만 나누어서 동일하게 적용되던 분류지표를 층별 특성에 따라서 적용되는 방식을 취하였다. 분류지표에 적용된 변수에도 기존분류지표에 적용된 변수에 덧붙여 “승용자 및 PC 보유대수”, “실업자수”와 같은 변수들이 추가로 적용되었다. 연동표본의 적용시기는 표본이 안정되는 표본개편 1년 후부터 적용할 예정이다.

이번 표본 개편의 아쉬운 부분으로는 현재의 인력과 예산문제로 조사별로 특성에 맞는 표본을 설계하지 못하고 다목적으로 표본이 설계된 점이다. 이로 인하여 경제활동인구조사와 가계조사를 동일한 가구에서 실시하게 되고 응답부담이 많이 가중된다.

참고문헌

- 박재수. 1989. 『표본조사법(이론과 실제)』.
- 통계청. 1997. “가구부문 표본개편 보고서.”
- 통계청. 2003. “가구부문 표본개편 보고서.”
- Cochran, W. G. 1963. *Sampling Techniques*. 2nd ed., New York: John Wiley & Sons, Inc.
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons, Inc.
- Hansen, M.H., and Hurwitz, W.N., and Madow, W.G.. 1953. *Sample Survey Methods and Theory*. New York: John Wiley & Sons, Inc. Vols, I and II.