

가 구 부 문
표 본 개 편 보 고 서

2003. 1.



 통계청

85>578

목 차

I. 서 론

1. 가구부문 표본조사 개요	1
2. 표본개편 목적	6
3. 표본개편 과정	6
4. 표본개편의 전체사항	8

II. 기존 표본의 검토

1. 기존 표본의 설계 개요	9
2. 표본의 유지 관리방법	13

III. 개선사항

1. 다목적 표본설계	14
2. 소지역 추정기법 도입	15
3. 연동표본 도입	16
4. 가중치 개선	18
5. 가계조사의 대표도 제고	18

IV. 새로운 표본설계

1. 모집단	19
2. 층화	19
3. 표본규모 결정	21
4. 표본조사구 추출	31
5. 표본조사구내 구역설정	34
6. 표본가구 선정	35
7. 표본조사구 번호부여	36
8. 소비자전망조사 및 집세조사 표본가구 선정	37

V. 병행조사 및 홍보

- 1. 병행조사 42
- 2. 홍보 42

VI. 추정

- 1. 분산과 표준오차 45
- 2. 경제활동인구조사의 추정 45
- 3. 전국가계조사의 추정 51

VII. 표본의 관리

- 1. 전입 및 전출 가구의 처리 54
- 2. 불응가구의 처리 54

VIII. 제언

- 1. 다목적표본 56
- 2. 병행조사 56

부록 연동표본 모형도 57

I. 서 론

1. 가구부문 표본조사 개요

현재 통계청에서 매월 경상적으로 실시하는 가구 대상 표본조사는 경제활동인구조사, 가계조사, 소비자전망조사 및 집세조사가 있는데, 이들 조사는 예산과 인력을 최소화하기 위해 동일 표본에서 병행 실시되고 있다.

가. 경제활동인구조사

1) 목 적

조사의 목적은 국민의 경제활동 즉 취업, 실업 등과 같은 경제적 특성을 조사하여 거시경제 분석과 인력자원의 개발정책 수립에 필요한 기초자료인 노동 공급, 고용구조, 가용 노동시간 및 인력자원의 활용 정도를 측정하여 제공하고 고용창출, 직업훈련, 소득증진 등을 위한 정부정책 입안 및 평가에 필요한 자료를 제공하는데 있다.

2) 연 혁

경제활동인구에 관한 조사는 내무부 통계국에서 '노동력 조사'라는 명칭으로 1957년부터 실시되어 왔으며 1963년 통계법에 의해 지정통계로 지정, '경제활동인구조사'라는 현재의 명칭으로 개편되고 경제기획원 조사통계국(현 통계청)에서 매 분기마다 면접조사하였다.

1969년부터는 현재와 같이 인구센서스 자료를 이용하여 다목적 표본이 설계되었으며 이후 일곱 차례에 걸쳐 표본개편을 실시하였고, 1982년부터는 분기 조사에서 월별 조사로 전환되었다. 1982년 개편된 ILO 권고안에 따라 조사표를 전면 개편하였고, 1987년부터는 현재와 같이 만 15세 이상의 인구에 대해 경제활동인구조사를 실시하고 있다.

3) 조사대상

경제활동인구조사 대상자는 표본지역에 상주하는 자로서, 매월 15일 현재

만 15세 이상인 사람이다.

4) 조사항목¹⁾

가구원 성명, 가구주와의 관계 등의 인적사항(6개), 활동상태, 취업여부 등 확인 항목(5개), 취업시간, 전직 희망 여부 등의 취업자 항목(3개), 취업 가능성 여부, 구직 경로 등의 실업자 항목(5개), 취업 의사 유무, 취업 가능성 여부 등의 비경제활동인구 항목(5개), 전직 유무, 이직 유무 등의 전직 사항 항목(6개)이 있다.

5) 조사대상기간 및 시기

조사는 매월 15일이 포함된 1주간(일요일~토요일)을 조사대상 주간으로 하며, 그 다음 주간에 조사를 실시한다.

6) 조사방법

표본가구 내에 상주하는 만 15세 이상 인구에 대해 조사대상 주간 동안의 경제활동상태를 조사하며, 조사원이 대상가구를 방문하여 노트북에 조사내용을 직접 기입하는 CAPI(Computer Aided Personal Interview)방식의 면접 타계식 조사이다.

나. 가계조사

1) 목 적

서울을 비롯한 도시에 거주하는 가구를 대상으로 일정 기간 동안의 수입과 지출 상태를 조사하여 국민생활실태를 명백히 파악함으로써 국민소비 수준의 변화 측정, 소비자 물가지수의 가중치 산정, 공공사업으로 인한 세입자의 주거대책비 산정 등 여러 경제정책이나 사회복지정책의 수립에 필요한 자료를 제공하는데 목적이 있다

1) 2003년 1월부터 조사 항목이 일부 개편되어 가구주와의 관계, 성별 등 기본 항목(6개), 활동상태, 취업여부 등 확인 항목(5개), 부업여부, 취업시간 등 취업자 항목(6개), 취업 가능성, 구직경로 등 실업자 항목(7개), 취업희망여부, 취업가능성 등 비경제활동인구 항목(4개), 전직유무 및 이직시기, 산업·전직 산업 등 기타 항목(7개)으로 구성되어 있다.

2) 연 혁

도시가계조사는 1942년 일제 말기에 시작되었으나 그 기록은 찾아볼 수 없고, 1951년 한국은행에서 전시 하의 국민소비수준을 측정하기 위해 조사를 실시 하였다. 그 후 1954년 서울의 근로자 200가구를 유의(有意) 표본으로 추출하여 조사하였고 1960년부터 통계적 추출방법에 의해 선정된 가구에서 조사를 실시하였다. 1963년 한국은행에서 통계청으로 본 조사를 이관하여 1969년부터 다목적 표본으로 조사되었으며, 이후 7차에 걸쳐 표본을 개편하였다. 한편 1974년까지 식료품비에 한하여 가계부 기장방식으로 조사하던 것을 1975년부터 전 품목에 적용하였고 1995년부터는 현재와 같이 조사항목을 10대 비목으로 확대하여 실시하고 있다.

3) 조사 대상

도시에 거주하는 2인 이상의 정상적인 가계수지 파악이 가능한 가구를 대상으로 한다.

4) 조사 사항과 조사 방법

가구실태에 관한 사항은 면접 타계식, 가계의 수입과 지출에 관한 사항은 가계부 기장에 의한 자계식, 연간소득에 관한 사항은 면접 타계식으로 조사한다.

5) 조사 대상기간

가계부에 의한 조사는 매월 1일부터 말일까지 1개월간을 조사대상기간으로 하며 가계부 배부는 전월 28~30일까지 3일간 실시하고 가계부 회수는 익월 1일부터 3일까지 3일간 실시한다. 연간소득조사표에 의한 조사는 조사대상기간을 전년 1월 1일부터 전년 12월 31일까지 1년간으로 하며, 실제 조사기간은 정기 조사의 경우 1월 3일부터 10일까지, 조사대상가구로 새롭게 편입되어 가계부를 제출하는 가구에 한하여 익월 1일부터 10일까지 실시한다.

다. 소비자 전망 조사

1) 목 적

소비자들이 느끼는 경기, 생활형편, 소비지출계획 등을 조사하여 경기 동향을 분석하여 정부정책 입안 및 민간 기업에 필요한 자료를 제공하는데 목적이 있다

2) 연 혁

소비자 전망 조사는 1998년부터 분기별 조사를 시작하여 1998년 12월 부터 월별 조사로 전환하였다.

3) 조사 대상

도시에 거주하는 인구 중 만 20세 이상 가구를 조사 대상으로 약 2,000 가구를 표본 대상으로 한다.

4) 조사 사항과 조사 방법

조사 사항은 기본 사항과 세부 사항으로 나눌 수 있는데 기본사항에는 성명, 성별, 생년월일, 교육정도, 산업 및 직업이 있으며, 세부 사항에는 경기 상황, 생활형편, 소비지출, 취업 기회 여부, 향후 물가 전개 항목 등으로 구성 되어있다.

조사 방법은 지방사무소의 조사 담당 직원이 조사 대상 가구를 직접 방문 하여 면접 조사한다.

5) 조사대상 기간

매월 22일이 포함된 1주간을 조사 대상 기간으로 하여 동 기간에 실시 조사를 한다.

라. 집세조사

1) 목 적

도시에 거주하는 가구가 주거생활을 위하여 지출하는 전·월세 금액을 파악하여 소비자물가지수 작성이나 여러 경제정책이나 사회복지정책의 수립에 필요한 자료를 제공하는데 목적이 있다

2) 연 혁

집세조사는 1955년부터 조사를 실시하였다. 그 후 서울, 부산 등 36개 도시에서 경제활동인구조사의 표본대상 중 전월세 가구를 대상으로 조사를 하고 있다.

3) 조사 대상

36개 도시의 경제활동인구조사 조사구역 내에 거주하는 전월세 가구를 대상으로 한다.

4) 조사 사항과 조사 방법

집세조사 가구명부, 집세변동, 집세조사 총괄 사항 등이 있으며, 조사 직원이 각 대상가구를 방문하여 면접 타계식으로 조사한다.

5) 조사 대상기간

전월 6일에서 금월 5일까지의 1개월간을 조사 대상 기간으로 하여, 매월 5일에 조사를 실시한다.

2. 표본개편 목적

현재 가구표본은 1995년 인구주택총조사 결과자료를 이용하여 1997년 표본개편에 따라 추출한 것으로 1998년 1월부터 사용하고 있어 표본이 노후화되어 있다. 총조사가 이루어지는 5년 기간 중에 발생하는 모집단의 변화를 표본에 반영 하기 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 표본의 현실 반영도를 제고하고 표본조사의 정도를 높이기 위해 5년마다 인구주택총조사 결과를 토대로 표본을 개편하고 있다.

3. 표본개편 과정

가. 준비작업

조사별로 지역별 표본오차를 산정·분석하여 새로운 표본규모를 계산하고 기존의 조사구 층화 분류지표를 검토하고, 조사구당 표본가구 수를 최적화하는 등 본격적인 표본개편작업에 들어가기 전에 대략적인 표본개편의 틀을 구성한다.

나. 표본개편 기본 계획 수립

표본개편의 기본계획을 수립하여 일정별로 차질없이 업무수행이 가능하도록 한다.

다. 통계위원회 기준·조정 분과위원회 상정

통계위원회는 정부조직법과 통계위원회규정에 의하여 설치된 국가통계 최고책임자의 자문기관으로 통계기준설정, 지정통계 지정, 통계 개선 등에 관하여 건의하는 역할을 수행하고 있다. 통계위원회는 기준·조정분과위원회, 인구사회분과위원회, 경제분과위원회의 3개 분과위원회가 운영되고 있는데 이중 표본에 관한 사항을 관할하고 있는 기준·조정분과위원회에 표본개편기본계획을

상정하여 표본설계방법에 대한 이론적 검증과 신뢰성을 확보한다.

라. 표본설계

1) 조사구 층화 분류지표의 결정

인구주택총조사 10% 표본자료에 대해 지역별로 다양한 분류지표를 적용하여 그 중 표본조사의 특성을 반영하면서 가장 모집단비율과 적합되는 분류지표를 선정하는 작업을 수행한다.

2) 표본 규모 결정

조사별 주요 변수의 상대 표준오차와 지역별 조사인력 현황을 고려하여 표본규모를 결정한다.

3) 표본조사구 추출

지역별 분류지표가 적용된 표본 추출틀에서 지역별로 결정된 표본규모의 조사구를 추출한다.

마. 표본조사구 및 표본가구 확정

추출된 조사구에서는 가구명부와 조사구요도를 작성하고 조사구를 구역화해 표본 조사구역을 설정한 후 표본가구를 확정한다.

바. 신표본 홍보

경제활동인구조사와 가계조사를 새로 실시하게 되는 신표본에 대해 다양한 매체를 통하여 각종 홍보계획을 세우고 홍보를 실시하여 조사에 대한 이해도와 참여도를 높인다.

사. 신·구 표본병행조사

조사실시 초기에 발생하는 무응답, 허위응답 등의 비표본 오차를 줄이고 표본교체로 인한 조사결과의 시계열을 조정하기 위해 구표본과 신표본을 병행 조사한다.

4. 표본개편의 전제사항

표본개편 작업은 반영하고자 하는 개편방향에 따라 작업수행이 완전히 달라지므로, 개편 시 고려할 전제 사항들에 대하여 미리 청내 관계관 회의를 거쳐서 결정하였다. 그 주요 내용으로는 다음과 같다.

가. 표본의 운영

각종의 통계조사를 하나의 표본에서 사용하던 기존의 다목적 표본을 가급적 감소시켜 경제활동인구조사, 가계조사 등 경상조사는 동일 표본에서 조사하고, 연간 및 특별조사는 동일 표본 조사구 내 다른 조사구역에서 조사를 실시할 수 있는 다목적 표본으로 설계한다.

나. 경제활동인구조사

- 1) 시·도 단위의 월별 통계 작성이 가능하도록 한다.
- 2) 특·광역시외의 구, 각 도의 시 및 인구 10만 이상의 군은 소지역 추정법을 적용하여 분기별로 고용통계를 작성할 수 있는 기반을 마련토록 한다.
- 3) 시·도단위의 통계 중 연령계층별, 교육정도별 통계를 소지역 추정법을 적용하여 작성할 수 있는 기반을 마련토록 한다.

다. 가계조사

- 1) 가계조사는 조사 대상 범위를 현행 동부에서 읍면부까지 확대한다.
- 2) 서울지역 분기별 통계자료 작성 및 공표에 추가하여 기타 시·도의 주요 항목에 대한 통계작성을 고려한다.

라. 연동표본 도입

응답자의 과중한 응답부담과 조사원들의 매너리즘을 예방하기 위해 매월 표본의 일부를 교체하면서 전년 동월비 중복이 있는 연동표본을 도입토록 한다.

II. 기존 표본의 검토

1. 기존 표본의 설계개요

가. 표본의 추출

1) 기초자료 집계

1995년 인구주택총조사 10% 표본조사구 중 보통 조사구만을 표본틀로 이용하며 각 추출단위 조사구에는 인구주택총조사 가구수를 8로 나누어 반올림한 결과를 크기의 측도로 부여하였다.

2) 표본 추출틀 작성

표본 추출틀은 특별시와 광역시의 6대 도시와 도별 시부 및 군부별로 각각 별도로 작성하였으며, 조사구의 분류지표는 각 층에서 동부와 읍면부별로 동일한 지표를 적용하였다. 이 때 동일한 분류지표를 부여함으로써 각 층별 특성 반영이 다소 미흡한 결과를 가져왔다.

가) 동부 분류지표

① 1차 분류 : 주택특성(5)

- 30평 이상 아파트, 30평 이하 아파트, 연립다세대, 단독, 기타

② 2차 분류 : 농림어업비율(5)

- 0%, 1 ~ 10%, 11 ~ 20%, 21 ~ 30%, 31 ~ 100%

③ 3차 분류 : 광공업비율(5)

- 0%, 1 ~ 10%, 11 ~ 20%, 21 ~ 30%, 31 ~ 100%

④ 4차 분류: 행정구역번호와 조사구번호순

나) 읍면부의 분류지표

① 1차 분류 : 주택특성(5)

- 30평 이상 아파트, 30평 이하 아파트, 연립다세대, 단독, 기타

② 2차 분류 : 광공업비율(5)

- 0%, 1 ~ 10%, 11 ~ 20%, 21 ~ 30%, 31 ~ 100%

③ 3차 분류 : 서비스업비율(5)

- 0%, 1 ~ 10%, 11 ~ 20%, 21 ~ 30%, 31 ~ 100%

④ 4차 분류: 행정구역번호와 조사구번호순

3) 표본조사구의 추출

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 2.1> 연행·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	71	50	21	27
서 울	160	160	-	127	충 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 경제활동인구조사의 조사구를 크기의 축소(인구주택총조사 결과의 조사구내 가구수를 8로 나누어 반올림한 수)에 비례하는 확률로 계통추출하며, 경제활동인구조사의 시부 표본조사구수와 도시가계조사의 표본조사구수가 같은 지역은 경제활동인구조사의 표본조사구에서 도시가계 조사를 실시하고 경제활동인구조사의 표본조사구가 많은 지역은 경제활동인구조사의 표본조사구 중에서 도시가계조사 표본조사구를 계통추출하였다.

4) 표본구역의 추출

조사구 내 거처를 가구 수에 의한 크기의 측도와 같은 수의 구역으로 나누어 경제활동인구조사는 임의로 추출된 1구역을 포함하여 3개의 구역을 추출하되, 처음에 표본으로 추출된 표본 구역 북쪽 시계바늘 방향으로 인접한 구역을 추출하였고, 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.

나. 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 ${}_s\hat{Y}_{qgh}$, \hat{Y}_{qgh} 라 하면, 이들 추정치를 산출하는 방법은 다음과 같다.

$$\begin{aligned} {}_s\hat{Y}_{qgh} &= \sum_i {}_s\hat{Y}_{qghi} \\ &= \sum_i {}_sM_{qghi} {}_sY_{qghi} \\ \hat{Y}_{qgh} &= \sum_s {}_s\hat{Y}_{qgh} \end{aligned}$$

위의 공식에서

$$\begin{aligned} {}_sM_{qghi} &= {}_s\hat{X}_{qghi} / {}_sX_{qghi} \\ {}_sY_{qghi} &= \sum_j {}_sY_{qghij} \\ {}_sX_{qghi} &= \sum_j {}_sX_{qghij} \end{aligned}$$

- \hat{Y} = 특성 Y를 갖는 추정인구
- Y = 표본에서 특성 Y를 갖는 것으로 조사된 인구
- M = 승수

- \hat{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

2) 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.

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

- \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

2. 표본의 유지 관리방법

가. 표본관리방법

표본으로 한번 선정되면 차기 표본 개편 시까지 계속 조사를 실시하는 고정 표본관리방법(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.

나. 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.

다. 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. 개선사항

새로운 대내외적 통계 환경의 변화에 부응하기 위해 금번 표본 개편 시에 반영 하고자 하는 주요 개선 사항에 대하여 청내의 의견 수렴을 거친 후 통계위원회 기준조정분과위원회에 상정하여 이론적 검증을 받아 정도 높은 통계작성과 향후 점진적 발전을 위한 통계적 기반을 구축하고자 노력하였다.

《 기 본 방 향 》

- ① 다목적 표본설계로 인력수요의 최소화
- ② 소지역 추정기법 도입으로 시군구 고용통계 작성여건 마련
- ③ 연동표본관리제 도입으로 응답자 부담 경감
- ④ 모수추정방법의 개선으로 통계정확성 제고
- ⑤ 전국 가계통계조사를 위한 표본설계

1. 다목적 표본설계

현행 통계청의 조사 여건을 감안하여 동일한 표본에서 여러 조사를 실시할 수 있는 다목적 표본으로 표본을 설계하여 운영하되, 응답자의 응답부담을 고려하여 매월 조사를 실시하는 경제활동인구조사, 가계조사, 소비자전망조사 및 집계 조사 등의 정상조사는 동일 표본 가구에서 조사를 하고, 사회통계조사 등의 연간 및 특별 조사는 동일 표본 조사구 내의 다른 조사 구역에서 조사를 실시하도록 하였다.

이 때 조사별 특성 반영 미흡에 대한 단점에 대해서는 다음과 같은 다양한 모집단 분류 지표를 활용하여 해결토록 하였다.

대분류	특 성 항 목
가 구 (8개)	농어가별 가구수(비농어가) 주택유형별 가구수(아파트, 단독, 다가구 및 연립) 가구구분별 가구수(혈연가구) 주택점유형태별 가구수(자가) 100가구당 승용차 보유대수 100가구당 컴퓨터 보유대수

인 구 (9개)	성별 인구수(남, 여) 연령별 15세 이상 인구수(15~19세, 20~29세, 30~59세, 60세 이상) 교육정도별 15세 이상 인구수(중졸이하, 고졸, 대졸이상)
경 활상 태 (13개)	경 활상 태별 인구수(경 활, 취 업자, 실 업자, 비 경) 산 업별 취 업자(농 립어 업, 광 공업, 서 비스업) 직 업별 취 업자(전 문기 술, 사 무, 서 비스판 매) 종 사상 지 위별 취 업자(근 로자, 자 영자, 사 업주)

2. 소지역 추정기법 도입

지방자치제가 정착되면서 자치 시군구별 통계 수요가 대폭 증가하였다. 이에 부응하여 시군구별 통계 작성이 필요하나, 표본에 의한 조사로 시군구 통계를 작성하기 위해서는 현행 표본 규모보다 약 5 ~ 6배의 표본이 요구되므로 이는 인적, 물적 비용상 매우 실현성이 없는 일이 된다. 따라서 이를 해결하기 위해서는 통계적 추정기법인 소지역 추정기법을 개발하여 시군구별 통계 또는 세부 단위별 통계를 작성할 수 있는 기반을 금번 표본 설계 시에 마련하여 지속적으로 연구해 가는 것이 필요한 실정이다.

미국, 캐나다 등 통계 선진국에서는 이미 80년대부터 소지역 추정 기법을 연구 개발하여 일부 통계에 적용하고 있다. 우리 통계청에서는 2000년부터 연구를 시작하여 바람직한 성과가 있었는데 그 중 주요 내용은 다음과 같다.

<표3.1>에서 볼 때 여러 가지 추정법 중 직접 추정량과 합성 추정량을 가중한 복합 추정량이 가장 정도가 좋고, 직접 추정에 의해서는 표본을 다수 확대한다 할지라도 크게 개선되지 않았으며, 시군구 실업자 수의 상대표본오차가 25%이내가 되기 위해서는 최소 4개 이상이 필요한 것으로 나타났다. 여기서 상대 표본 오차 25%를 기준으로 한 것은 통계학자(J. K. Rao)의 권고사항에 따른 것이다.

따라서 금번 표본 설계 시에 반영하고자 하는 사항은 다음과 같다.

첫째, 시군구 고용통계는 분기별로 작성토록 한다. **둘째**, 특·광역시외의 구,

각 도의 시 및 인구 10만 이상인 군을 소지역 통계 적용 대상으로 한다. 셋째, 우선 적용 대상 시군구에는 기본적으로 4개 표본 조사구를 추출토록 한다.

<표 3.1> 추정기법별 상대표본오차

(단위 : %)

		조사구수		직접 추정량		합성 추정량		복합 추정량	
		기존	확대	기존 조사구	확대 조사구	기존 조사구	확대 조사구	기존 조사구	확대 조사구
광 주	동 구	10	14	34.8	27.4	10.7	10.0	10.2	9.4
	서 구	15	18	24.1	23.2	10.5	9.9	9.7	9.2
	남 구	16	18	21.5	18.2	10.5	9.9	9.5	8.7
	북 구	34	34	14.6	14.6	10.5	9.9	8.6	8.3
	광산 구	12	17	30.1	23.0	10.4	9.8	10.0	9.2
충 북	청주시	22	22	19.6	19.6	11.5	10.4	9.9	9.3
	충주시	11	16	24.9	26.8	11.9	10.7	10.7	10.0
	제천시	4	14	63.3	26.6	12.0	10.7	12.3	9.9
	보은군	2	10	86.3	23.1	29.3	15.6	27.8	13.8
	옥천군	3	10	42.2	25.3	30.5	16.2	24.9	14.0
	영동군	3	10	75.0	55.9	30.5	16.4	30.4	16.1
	괴산군	5	11	83.3	65.5	29.4	15.5	28.3	15.3
	음성군	6	11	43.4	46.9	29.3	15.4	24.6	14.6
	청원군	5	14	40.0	32.8	30.4	16.2	25.2	14.6
	진천군	1	10	-	48.5	29.8	15.7	29.8	14.9
단양군	2	10	-	46.5	29.6	15.6	29.6	14.7	

3. 연동표본(Rotation sampling) 도입

현행 표본 관리는 한번 표본으로 선정되면 다음 표본 개편 시까지 계속 조사 하는 고정 표본 관리방법을 사용하여 왔다. 따라서 고정 표본 관리에 따라 응답자는 5년간 각종 통계 조사에 응하여야 하는 응답 부담이 가중되어 이를

감소시켜 줄 필요가 있으며, 모집단 변화 반영이 시간 경과에 따라 감소되는 표본의 노후화 현상 및 전면적 표본 개편에 따른 신규 계열간의 괴리 현상을 감소하기 위해서는 매월 표본의 일부를 교체하여 주는 표본 관리 기법 도입이 필요하다.

이에 따라 99년 11월부터 2000년 3월까지의 이론적 연구에 이어 2000년 5월부터 실시한 시험조사 결과 경제활동인구조사의 응답률은 교체에 따른 차이는 거의 없었고, 가계조사의 경우 첫 3개월 간은 차이가 현저하나 그 이후에는 소폭으로 증가하는 것으로 나타났다.

<표 3.2> 경제활동인구조사 응답율

응답횟수	1	2	3	4
응답율	97.22	97.42	97.51	97.51

<표 3.3> 가계조사 가계부 회수율

응답횟수	1	2	3	4	5	6	7	8	9
회수율	36.5	51.1	57.7	58.2	60.5	55.9	59.4	61.4	63.3

이러한 분석 결과를 토대로 통계 시계열의 안정과 조사 업무 부담의 최소화에 염두를 두고 표본설계에 반영하고자 하는 연동표본의 내용은 다음과 같다.

- ① 표본의 월별 교체 단위는 조사구역 1개로 한다.
- ② 표본의 월별 교체 규모로는 1/36으로 한다. 따라서 표본으로 응답하는 기간은 기존 60개월에서 36개월로 감소한다.
- ③ 표본 교체 주기는 매월로 한다.
- ④ 교체되어 들어오는 표본 조사구는 가급적 인접한 조사구로 한다.
- ⑤ 연동표본을 시작하는 시기는 2004년 1월부터 한다.

<그림 1> 조사구 교체구조도*

최 초		교체전 조사구				교체후 조사구			
		①	②	3	4	-	-	-	-
첫 번째월	경 활		2	3	4	1			
	가 계	①	②						
두 번째월	경 활			3	4	1	2		
	가 계	①	②						
세 번째월	경 활				4	1	2	3	
	가 계		②			①			
네 번째월	경 활					1	2	3	4
	가 계					①	②		

* 조사구역 수는 4개이고 그 중 가계는 2개 구역에서 조사한다고 가정

4. 가중치 개선

기존의 모집단 추정치는 시차를 두고 하는 계속조사(Successive survey in time)일 경우 모집단의 변화를 반영하기 위해 사용하게 되는 비(Ratio) 추정법을 사용하고 있다. 이 때 사용되는 자료는 인구추계자료인데 25개 층별, 남녀별의 50개의 승수를 사용하여 모집단 특성을 추정하고 있다. 보다 더 정확한 통계를 작성하기 위해 이를 더욱 세분하여 연령대별 승수를 부여하여 모집단을 추정하고자 하였다.

5. 가계조사의 대표도 제고

가계조사의 조사대상 범위(Coverage)가 현재는 동부에 거주하는 2인 이상 가구로 되어 있으나, 금번 표본에서는 2인 이상 읍면부 지역까지 확대하여 가계 수지 통계의 대표도를 향상하고자 하였다. 이 경우 범위의 대표도가 기존의 동부근로자 가구(35.2%)에서 단독 가구를 제외한 75.6%로 대폭 향상되게 되었다.

IV. 새로운 표본설계

1. 모집단

표본설계 시의 범위(Coverage) 오류를 최소화하기 위해 모집단의 기초자료로 2000년 인구주택총조사 결과자료를 사용하였는데, 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 238,000 가구의 아파트 자료를 보조 자료로 사용하였다.

2. 층화

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

층별 조사모집단 현황은 <표 4.1>과 같다.

<표 4.1> 조사모집단 현황

시 도	조사구	가 구	시 도	조사구	가 구
전 국	24,998	1,420,418			
서 울	5,506	302,852	강 원	894	50,170
			동부	483	27,340
			읍면부	411	22,830
부 산	1,954	112,271	충 북	806	46,696
			동부	444	26,338
			읍면부	362	20,358
대 구	1,332	75,054	충 남	1,043	58,894
			동부	325	18,188
			읍면부	718	40,706
인 천	1,282	73,365	전 북	1,101	62,279
			동부	657	37,353
			읍면부	444	24,926
광 주	707	40,934	전 남	1,092	62,949
			동부	416	24,189
			읍면부	676	38,760
대 전	714	40,763	경 북	1,565	90,217
			동부	683	39,995
			읍면부	882	50,222
울 산	525	3,037	경 남	1,630	94,841
			동부	934	54,437
			읍면부	696	40,404
경 기	4,577	262,735	제 주	270	16,091
동부	3,699	213,297	동부	179	10,552
읍면부	878	49,438	읍면부	90	5,539

3. 표본규모 결정

가. 표본규모와 표본오차

표본 설계의 가장 중요한 문제 중 하나는 표본의 크기를 결정하는 것이라 할 수 있다. 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.

나. 목표정도 결정

표본 조사의 목표 정도란 추정값의 허용 오차를 통계 이용가치의 관점에서 정한다는 뜻이다. 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.

따라서 여러 가지 항목의 표본 오차를 동시에 고려해야 하는 본 표본 설계에서는 상대표준오차를 주로 사용하게 된다. 또한 CV는 시간적, 공간적으로 변동이 작으며 거의 일정한 값을 갖는다. CV값이 일정한 값으로 안정되어 있다는 것은 허용 오차가 주어지면 모수를 추정할 때 필요한 표본 크기를 구하는데 도움이 된다. 또 한 가지의 특징은 모집단의 크기가 더 확대되더라도 CV 값은 비교적 안정되어 있다는 것이다.

CV값은 추정값의 불확실성의 정도에 사용되는 값이며 예를 들면 추정값이 100,000이고 이에 대한 확실성이 95%이고, CV값이 7%라면 실제값은 86,000과 114,000사이에 있다고 볼 수 있고, 만약 추정치의 확실성을 99%라고 하고 CV 값이 7%라면 실제값은 79,000과 121,000 사이에 있다고 볼 수 있다.

1) 경제활동인구조사

표본규모 산출에 사용할 특성 항목을 정하기 위해 2000년 경제활동인구조사 결과의 연평균 시·도별 주요 항목별 상대표준오차(CV)를 분석한 결과는 <표 4.2>과 같다.

<표 4.2> 2000년 경제활동인구조사 주요항목별 상대표준오차(연평균)

	경제활동 인 구	취 업 자			실 업 자	비 경 제 활동인구	
		농림어업	광 공 업	서비스업			
서 울	0.28	0.28	11.81	1.11	0.37	2.08	0.45
부 산	0.41	0.43	7.18	1.65	0.55	2.46	0.57
대 구	0.45	0.45	9.44	1.55	0.64	2.97	0.64
인 천	0.36	0.38	15.48	1.20	0.55	3.02	0.56
광 주	0.45	0.48	4.58	2.10	0.57	2.71	0.57
대 전	0.44	0.44	6.52	2.87	0.68	3.59	0.61
울 산	0.67	0.68	14.81	2.02	1.70	5.97	0.99
경 기	0.36	0.37	3.95	1.28	0.57	3.03	0.57
강 원	0.56	0.61	4.13	3.70	1.02	5.56	0.77
충 북	0.71	0.73	2.40	2.40	1.32	4.49	1.03
충 남	0.63	0.65	2.48	3.22	1.42	4.57	1.07
전 북	0.57	0.58	2.00	2.75	0.95	4.33	0.80
전 남	0.54	0.58	3.65	3.72	1.73	4.33	0.95
경 북	0.47	0.48	2.46	2.79	1.26	4.42	0.86
경 남	0.61	0.58	3.64	2.67	1.15	4.22	0.93
제 주	0.57	0.57	2.72	4.05	0.88	5.46	1.18

<표 4.2>에 나타난 바와 같이 주요 항목의 상대표준오차에서 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 농어가가 골고루 분포하지 않기 때문이다. 이를 표본 규모 산정의 특성 항목으로 선정할 경우에는 매우 큰 표본이 요구되므로 표본 조사의 취지에 어긋난다.

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.

참고적으로 외국의 경우는 다음과 같다. 미국 CPS의 목표 상대표준오차는 6%실업률에 대해 전국의 CV는 월 1.8%, 11개 주에 대해서는 월 8%, 39개 주와 DC에 대해서는 연평균 8%를 기준으로 표본규모를 선정하고 있다. 캐나다 LFS에서는 실업자 추정치에 대한 전국 단위의 CV가 약 2%이내, 주(Province) 지역의 CV는 약 4~7%로 설정되어있으며, 고용보험(EIER) 경제구역이나 센서스 도시지역(CMA)에 대해서는 분기별 실업자의 CV가 15%이내, 경제구역(ER;Economic Region)에 대해서는 25%이내를 목표 정도로 설정하고 있다.

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.

다. 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 조사 대상을 확보할 수 있게 된다.

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 연동구조 등을 고려하여 <표 4.3>과 같이 층별 표본 규모를 결정하였다.

<표 4.3> 경제활동인구조사 층별 표본규모

(단위 : 개, %)

시 도	시군구 수	현 행 조사구수	신 표 본		시 도	시군구 수	현 행 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

()안의 숫자는 시 또는 구의 수임

2) 전국가계조사

각 도의 동부, 읍면부별로 조사구당 적격가구수가 동부는 평균 47.4가구인 반면 읍면부는 25.6가구로 차이가 상이하어 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 <표 4.4>와 같다.

<표 4.4> 근로자가구 소비지출의 상대표준오차

(단위 : %)

	1/4분기	2/4분기	3/4분기	4/4분기	평 균
전 국	1.46	2.08	1.60	1.35	1.62
서 울	2.95	2.51	3.25	2.75	2.87
부 산	3.65	3.65	3.50	3.01	3.45
대 구	4.96	7.02	5.84	4.65	5.62
인 천	3.8	3.03	2.95	2.59	3.09
광 주	6.09	4.87	4.43	4.90	5.07
대 전	6.69	5.46	6.59	5.92	6.17
울 산	3.99	5.35	4.34	2.76	4.11
경 기	4.67	8.91	5.20	4.15	5.73
강 원	6.99	9.32	5.70	8.31	7.58
충 북	6.4	5.94	5.27	5.62	5.81
충 남	8.21	6.04	6.83	6.77	6.96
전 북	3.94	4.97	7.06	9.27	6.31
전 남	6.29	4.89	4.45	7.36	5.75
경 북	6.19	3.84	3.93	4.71	4.67
경 남	5.56	6.03	5.00	5.67	5.57
제 주	3.72	4.66	6.14	7.39	5.48

<표 4.4>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 <표 4.5>와 같이 산출하였다.

<표 4.5> 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

이 때 사용한 공식은 다음과 같다

$$n_1 * CV_1^2 = n_2 * CV_2^2$$

- 여기서 n_1 : 현행 표본규모
- CV_1 : 현행 표본에서의 상대표준오차
- n_2 : 신표본에서의 표본규모
- CV_2 : 목표상대표준오차

<표 4.5>에서 목표 정도에 맞는 표본 규모와 연동표본을 위한 9의 배수 구조에 따라 <표 4.6>과 같이 동부의 표본 규모를 결정하였다.

<표 4.6> 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 표본의 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.

다) 적격가구수 추정

가계조사에서는 1인 단독가구와 농어가 등은 조사지침에 따라 부적격 가구로 조사대상에서 제외되는 가구가 있어 조사구당 최소한의 적격가구를 확보하는 것이 중요하다. 농어가가 많은 읍면부에서는 특히 주의가 요구된다.

따라서 적격가구가 3가구 미만으로 예상되는 조사구는 미리 표본들에서 제외하도록 하였다. 각 시도별 예상 적격 가구수는 조사구 수에 조사구당 표본 추출대상가구수(10가구)와 적격 가구 비율을 곱하여 산출할 수 있는 데, 동부에서는 적격 가구 비율을 $\frac{\text{조사구당 평균 적격가구수}}{\text{조사구당 평균 가구수}}$ 에 조정계수를 곱한 것을 사용하였고, 읍면부에서는 2인 이상 비농어가 비율이 30%이상인 조사구에서의 적격 가구 비율을 적용하였다.

여기에서 동부의 조정계수는 조사지침상 부적격 가구로 제외되는 음식점 등을 경영하는 겸용 주택 내의 가구, 영업상 사용인 2인 이상 동거하는 가구 등을 감안하기 위한 계수로 다음과 같다.

$$\frac{\text{현행 도시가계 적격가구 비율}(0.77)}{2000인구주택총조사 10\%표본의 동부 적격가구 비율}(0.83)$$

최종적으로 전국가계조사를 위한 표본조사구 및 적격 가구수는 <표 4.7>과 같이 결정하였다.

<표 4.7> 전국가계조사 표본조사구 및 예상 적격가구수

시 도	현행 표본			예상 표본 규모						
	조사구수	가구수	CV (%)	계		동 부			읍 면 부	
				조사구수	가구수	조사구수	가구수	CV	조사구수	가구수
전 국	697	5,087	1.62	999	7,295	765	5,856		234	1,439
서 울	127	995	2.87	126	979	126	979	2.88		
부 산	70	518	3.45	72	570	72	570	3.40		
대 구	49	357	5.62	54	420	54	420	5.35		
인 천	50	377	3.09	54	426	54	426	2.97		
광 주	52	340	5.07	54	413	54	413	4.98		
대 전	50	357	6.17	54	414	54	414	5.94		
울 산	23	190	4.11	36	274	36	274	3.29		
경 기	50	384	5.73	108	822	72	575	4.78	36	246
강 원	27	191	7.58	63	417	36	257	6.56	27	160
충 북	27	193	5.81	54	369	27	201	5.81	27	168
충 남	30	184	6.96	63	435	36	267	6.35	27	168
전 북	29	191	6.31	54	343	27*	201	6.54	27*	142
전 남	29	209	5.75	54	373	27*	203	5.96	27*	170
경 북	32	230	4.67	63	429	36	264	4.40	27	166
경 남	34	253	5.57	63	454	36	274	5.41	27	180
제 주	18	120	5.48	27	156	18	118	5.48	9	39

* 전남, 전북의 읍면부 적격 가구 비율이 30% 이상인 조사구가 적은 관계로 동부 27개, 읍면부는 18개로 조정함

4. 표본조사구 추출

가. 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.

나. 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 4.8> 특성지표 항목

대분류	특 성 항 목
가 구 (8개)	농어가 가구수(비농어가) 주택유형별 가구수(아파트, 단독, 다가구 및 연립) 가구구분별 가구수(혈연가구) 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)

다. 표본추출단위 조사구명부 작성

계통추출에 있어 추출 단위 조사구 명부를 어떤 기준에 의해 정렬치 않고 추출하는 것보다는 기준에 의해 명부를 정렬한 후 계통추출하는 것이 표본의 대표성을 제고할 수 있다. 분류지표 종류가 4개 이상이 되는 경우에는 각 셀에 해당하는 조사구가 적어 오히려 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 <표 4.9>와 같다.

<표 4.9> 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 <표 4.9>의 분류기준에 의해 각 층별로 정렬한 후 조사구내 가구수를 5로 나누어 반올림한 결과를 크기의 측도(MOS)로 부여하여 표본추출 단위 조사구 명부를 작성하였다.

라. 표본조사구 추출

표본조사구의 추출은 각 층별로 구해진 표본 수만큼을 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 표본을 추출하였다. 이 때 가장 최적의 표본을 선택하기 위해서 다음과 같은 최적계수를 사용하였다.

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

여기에서 R_i : 해당 표본에서의 i 항목의 비율

\bar{R}_i : 모집단에서의 i 항목의 비율

S_i : 추출 가능한 모든 표본에서의 i 항목의 표준편차

표준편차로 나눈 것은 30개 항목에 대한 모집단과 표본에서의 비율차이를 상대적으로 평가하기 위해 일반적인 차이거리인 유클리드(Euclidian)거리가 아닌 Mahanalobis 거리를 구하기 위해서였다.

마. 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 3가구 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.

5. 표본조사구내 구역설정

확정된 표본 조사구에 대하여 2000년 인구주택총조사 당시의 조사구요도 및 가구명부를 기초로 현지 확인을 통하여 현 상태에 대한 조사구요도와 가구명부를 재작성하였다.

가. 조사구 경계 확인

표본 조사구요도 작성시 가장 기본이 되는 사항은 경계선의 명확성이다. 따라서 표본조사구의 경계가 뚜렷한 지형지물에 의하여 설정되어 있는 지를 확인 하고 누구나 조사지역을 쉽게 식별할 수 있도록 외부 인접지역의 주요 건물, 상점, 기타 그 지역에서 특징이 있는 주택 등을 그려 넣고 그 명칭을 기입한다.

나. 조사구요도의 작성

조사구 경계확인이 완료되면 조사구요도 용지 위에 그 경계선과 도로, 하천, 교량 등의 지형지물을 그려 넣는다. 이때 조사구의 모양과 지형지물의 폭이 실제 모양에 비례되도록 작성하여야 한다. 이 때 각 건물의 명칭이나 특징이 있을 때에는 그 명칭이나 특징도 기입토록 한다.

다. 거쳐번호부여 및 가구명부작성

조사구요도의 작성이 완료되면 조사하는 순서대로 상호인접되게 거쳐번호를

부여한다. 사람의 거주 여부를 확인할 때에는 조사구요도에 출입문 표시를 하고 출입문별로 가구를 확인하여 가구 누락을 방지하여야 한다.

출입문 표시에 따른 조사동선(動線)을 확인하여 이후 조사구역 설정 시에도 참고하도록 한다. 사람이 살고 있는 것을 확인한 때에는 가구명부상의 관련 항목을 기입한다.

라. 표본조사구내 구역설정

조사구요도와 가구명부가 새로 작성된 후 경제활동인구조사는 크기의 측도만큼의 구역으로 분할하여 설정한다. 조사구 내의 거처들을 구역화할 때에, 각 구역간 거처 수는 다르게 되더라도 가구 수는 균등하게 되어야 하며, 표본 조사구내에 주택신축이 가능한 공지(택지)가 있는 경우는 공지면적을 균등하게 배분시켜 설정한다.

각 거처의 구역화가 완료되면 1번 거처가 포함된 구역부터 서로 인접되도록 2자리 숫자의 구역 번호를 부여한다.

6. 표본가구 선정

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.

이 때, 4개의 조사구역 외 향후 연간 및 특별조사를 실시할 때 경상조사의 조사구역을 사용치 않을 경우 사용할 목적으로 인접한 4개 구역을 별도로 설정하여 놓았다.

7. 표본조사구 번호부여

표본의 연동을 위하여 각 층별로 분류지표에 따라 조사구를 정렬한 후, 차례대로 연동그룹 부호와 연동그룹 내 일련 번호를 각 조사구에 부여하였다.

표본 조사구 번호는 총 8자리로 구성하였으며, 자리별 내용은 다음과 같다.

㉠	㉡	㉢	㉣	㉤	㉥	㉦	㉧
---	---	---	---	---	---	---	---

자 리	내 용	분 류
㉠~㉡	시·도	11 서울, 21 부산, 22 대구, 23 인천, 24 광주, 25 대전, 26 울산, 31 경기, 32 강원, 33 충북, 34 충남, 35 전북, 36 전남, 37 경북, 38 경남 39 제주
㉢	동부, 읍면부	1 동부, 2 읍면부
㉣~㉤	연동그룹 내에서의 일련번호	00 ~ 99
㉥	연동그룹을 표시 하는 부호	1 2004. 1월 연동교체 조사구 2 2004. 5월 연동교체 조사구 ⋮ 9 2006. 9월 연동교체 조사구
㉦	연동구조에 따라 바뀐 횟수 표시	0 표본개편당시 조사구, 1 2004. 1월 이후 바뀐 조사구, 2 2007. 1월 이후 두 번째 바뀐 조사구
㉧	연동구조 이외의 사유로 바뀐 횟수 표시	

※ 조사구 교체 시에는 조사구 번호를 재부여하며, 6번째 자리까지는 변하지 않고, 7번째와 8번째 자리만 바뀌게 된다

8. 소비자전망조사 및 집세조사 표본가구 선정

가. 소비자전망조사

1) 전제조건

소비주체가 느끼는 경제현상에 대한 인식을 조사하여 지수화하고자 하는 조사목적, 조사대상, 조사여건 및 기존 조사 등을 고려하여 일반적 인식조사를 위한 표본규모와 유사한 동부지역의 20세 이상 70세 미만의 기혼자 2,000명이 표본 규모로 결정되어졌다.

2) 표본설계

소비자전망조사의 표본은 다목적 표본의 원칙에 따라 경제활동인구조사 표본을 주표본(Master sample)으로 하여 표본을 추출하는 방식을 채택하였는데, 응답자의 부담을 고려하여 전국가계조사를 실시하지 않는 가구를 우선적으로 추출토록 하였다.

주표본에서 동부에 분포되어 있는 1,220개 조사구중 20세 이상 70세 미만 기혼자가 있는 가구 수를 조사구별로 집계하여 조사 대상 가구를 파악하였다. 조사구별로 조사 대상 가구가 8가구 이상인 조사구를 조사 가능 조사구로 간주하였다.

최종 표본은 각 조사구당 2가구, 1가구당 1명으로 하여 표본 조사구는 총 1,000개 조사구로 하였다. 시도별 표본 조사구 수는 2000년 인구주택총조사 결과 각 시도의 동부에 거주하는 20세 이상 70세 미만 기혼자 수에 비례하여 배분하되, 시·도별 조사 가능 조사구 수를 감안하여 조정하였다.

<표 4.10>과 같이 서울의 경우에 2000년 인구주택총조사 결과에 따르면 266개의 표본 조사구가 필요하나, 가계조사를 포함하여도 조사 가능 조사구가 최대 199개 밖에 없으므로 199개 전부 서울의 표본 조사구로 선정하고, 나머지 67개는 부산을 비롯한 6개 광역시에서 차례로 추가하였다. 마찬가지로 경기도의 경우 가계조사 표본을 포함한 전부를 표본으로 선정하고 부족한 61개 조사구는 기타 도에서 차례로 추가하여 표본 조사구를 조정하였다.

<표 4.10> 표본조사구 산정(안)

시 도	2000년 인구주택 총조사결과 동부의 20세 이상 70세 미만 기혼자수 (천명)	경제활동인구조사구중 조사가능조사구수 (동부의 20세이상 70세미만 기혼자 가구수가 8가구 이상인 조사구수)		표 본 조 사 구 수		
		가계조사 포 함	가계조사 제 외	인구총조사 비 레 시	조사가능 조사구수 감안조정	확 정 조사구수
전 국	18,451	1,202	1,024	1,000	0	1,000
7개광역시	10,910	712	619	591	0	591
서 울	4,914	199	172	266	△67	199
부 산	1,840	114	104	100	4	104
대 구	1,178	85	72	64	8	72
인 천	1,223	102	91	66	25	91
광 주	646	81	67	35	11	46
대 전	672	78	65	36	12	48
울 산	438	53	48	24	7	31
9개도	7,541	490	405	409	0	409
경 기	3,603	134	113	195	△61	134
강 원	445	43	33	24	9	33
충 북	425	44	34	23	10	33
충 남	300	45	38	16	7	23
전 북	614	45	39	33	6	39
전 남	405	45	37	22	10	32
경 북	658	53	47	36	11	47
경 남	916	54	42	50	4	54
제 주	175	27	22	10	4	14

표본 조사구는 경제활동인구조사 조사구 번호 순으로 정렬한 후 각 시도별로 정해진 표본 조사구만큼 시뮬레이션을 통해 20세 이상 70세 미만 기혼자가 가장 많은 조사구를 계통추출하여 표본 조사구로 선정하였다.

표본대상 가구원은 응답 부담을 고려하여 2-4-2 연동표본을 적용하여 추출 하였다. 여기서 2-4-2 연동표본체는 응답 가구의 입장에서 보면 2개월 조사-4개월 미조사-2개월 조사가 반복되는 표본 구조이다. 즉 조사구당 6명을 선정하여 2명을 조사하되 매월 1명을 교체토록 하였다. 이 때 교체되는 표본은 성별, 연령별, 교육 정도별 분포가 유사하도록 하였다. 표본 조사구 내 가구원 전체를 성별, 연령별, 교육 정도별로 정렬시킨 후 계통추출하는 방법을 시뮬레이션하여 <표 4.11>과 같이 추출된 표본의 구조가 가구원 전체의 구조와 최대한 같도록 하였다.

<표 4.11> 표본가구원의 분포현황

		1,000개 조사구		표 본		표 본 가 구 원					
		(명)	(%)	(명)	(%)	1번	2번	3번	4번	5번	6번
20세 이상 70세 미만 기혼자수		32,770	100.0	6,000	100.0	1,000	1,000	1,000	1,000	1,000	1,000
성	남 자	15,296	46.7	2,802	46.7	466	468	466	468	467	467
	여 자	17,474	53.3	3,198	53.3	534	532	534	532	533	533
연 령	20대	2,201	6.7	401	6.7	64	69	66	68	67	67
	30대	9,796	29.9	1,795	29.9	300	298	300	299	299	299
	40대	10,866	33.2	1,990	33.2	332	332	332	330	332	332
	50대	5,977	18.2	1,094	18.2	184	182	182	182	183	181
	60대 이상	3,930	12.0	720	12.0	120	119	120	121	119	121
교 육 정 도	중 졸 이 하	9,851	30.1	1,803	30.1	300	300	302	300	302	299
	고 졸	14,483	44.2	2,651	44.2	441	442	442	443	440	443
	대 졸 이 상	8,436	25.7	1,546	25.8	259	258	256	257	258	258

표본으로 선정된 가구원의 전출 등 유고 시에는 해당 조사구 내에서 성, 연령, 교육 정도가 가장 유사한 다른 가구원으로 대체하여 조사토록 하였다.

나. 집세조사

집세조사는 원칙적으로 전·월세가구를 대상으로 하는 조사인데 이를 위한 표본 추출을 하기 위한 전·월세가구에 대한 모집단 자료가 미흡할 뿐 아니라, 개개의 가구를 표본으로 추출하여 조사하는 것은 인적, 물적 비용이 많이 소요되므로 경제활동인구조사 표본 중 전·월세가구를 표본 대상으로 하도록 청내 관계관의 의견 수렴으로 결정하였다. 표본규모는 <표 4.13>과 같다.

<표 4.13> 집세조사 표본규모 현황

구 분	전 세					월 세				
	계	단 독	연 립	아파트	기 타	계	단 독	연 립	아파트	기 타
전 국	5,632	3,045	602	1,963	22	4,193	2,900	140	1,122	31
광역시	3,577	2,171	358	1,039	9	2,478	1,705	84	669	20
서울	1,273	838	127	304	4	757	508	34	215	
부산	620	392	38	190		426	344	10	72	
대구	404	320	15	69		336	253	0	83	
인천	338	98	117	118	5	185	82	23	60	20
광주	367	195	16	156		304	172	3	129	
대전	370	198	34	138		260	165	11	84	
울산	205	130	11	64		210	181	3	26	
수원	108	58	10	40		30	23	1	6	
성남	109	38	33	38		33	29	3	1	
의정부	43	32	0	11		42	41	0	1	
안양	98	51	12	35		31	29	2	0	
부천	110	42	27	36	5	26	14	3	8	1
고양	120	29	31	60		35	24	5	6	

구 분	전 세					월 세				
	계	단 독	연 립	아파트	기 타	계	단 독	연 립	아파트	기 타
춘 천	65	38	1	26		76	42	3	31	
원 주	65	27	0	38		103	55	0	48	
강 릉	63	36	0	27		83	37	0	46	
청 주	111	69	2	40		149	83	1	65	
충 주	77	23	4	50		62	36	0	25	1
천 안	79	12	7	60		108	67	0	41	
공 주	33	23	0	10		32	31	0	1	
보 령	13	9	0	4		22	18	0	3	1
전 주	79	34	5	40		82	37	3	42	
군 산	64	18	3	43		36	17	0	18	1
남 원	27	15	6	6		4	4	0	0	
목 포	56	31	1	24		41	28	0	13	
여 수	88	21	6	61		41	27	5	9	
순 천	60	16	3	41		41	18	0	23	
포 항	60	23	23	14		33	31	1	1	
경 주	78	46	2	22	8	80	75	0	0	5
안 동	55	27	0	28		108	101	0	7	
구 미	203	53	49	101		86	62	5	19	
창 원	55	35	0	20		60	56	0	4	
마 산	36	21	0	15		61	45	0	14	2
진 주	44	27	0	17		61	60	0	1	
제 주	25	8	9	8		105	72	16	17	
서귀포	31	12	10	9		44	33	8	3	0

V. 병행조사 및 홍보

1. 병행조사

가. 목적

표본의 전면 개편에 따른 신표본에서의 신속한 조사의 정상화 정착, 구표본에서의 계속적인 정확한 응답 유도 및 신·구표본 간 시계열상의 괴리를 최소화하기 위해 필요하다.

나. 기간

병행 조사 기간은 업무량 및 여러 조사 환경을 고려하여 2002년 10월부터 12월까지 3개월간으로 정하였다.

다. 업무담당

기존 조사요원이 신·구 표본을 전부 동 기간에 조사할 수 없으므로 임시 조사원을 채용하여 실시토록 하였다. 신표본은 새로이 표본으로 선정된 가구를 대상으로 조사 설득과 향후 교체될 때까지 계속 조사를 해야 하는 가구이기 때문에 조사에 경험이 많고, 해당 조사에 깊은 이해도가 있는 통계청의 정규직원이 담당하고, 이미 상당기간 조사대상으로 있어 조사에의 거부감이 없는 구표본은 임시조사원이 담당하는 것을 원칙으로 하였다.

2. 홍보

가. 목적

가구부문 경상조사의 표본으로 선정되었음과 각 통계의 중요성을 조사 대상가구에 널리 알려 조사에의 협조도로 정확한 응답을 유도하고 응답 거부 등의 열악한 조사환경을 타개하기 위해 필요하다.

나. 기본 방향

병행조사 실시 전후에 걸쳐 홍보 효과를 극대화하기 위해 다음과 같은 각종 홍보 활동을 집중적으로 전개토록 하였다.

- ① 통계조사별 팜플렛 및 포스터 제작·배부
- ② 방송, 언론, 인터넷 및 통신매체의 적극 활용
- ③ 대상가구에 유용한 「생활의 지혜」 모음집 제작·배부
- ④ 정부부처 및 유관기관에의 협조유도
- ⑤ 조사응답에 대한 답례품 지급
- ⑥ 통계퀴즈실시
- ⑦ 사무소 홍보반 편성·운영
- ⑧ 통계조사결과 요약자료 배부

다. 중점 홍보내용

홍보의 중점 내용은 첫째, 지식 사회에서의 통계의 의의 및 중요성을 강조하였다. 이를 위해 가계의 소비 및 노후 생활 설계 등 일상 생활에서의 다양한 활용성, 창업을 위한 적지 선정 등 사업계획수립에서의 통계 이용성과 통계는 의사결정의 방향을 정하는 나침반이며 국가 발전의 인프라에 해당하는 자본재임을 홍보하였다.

둘째, 비밀 보호의 제도적 장치에 관한 내용으로 조사 내용은 통계 목적에만 사용되며 개인의 비밀은 철저히 보호됨을 강조하였다. 셋째, 통계를 작성하기 위한 기획부터 자료 공표까지의 전과정을 요약하여 설명하였다. 마지막으로 정확한 통계가 작성되기 위해서는 응답자의 협조가 필수적인 요소임을 강조하여 적극적인 협조를 당부하였다.

라. 홍보방안별 실시내용

1) 팜플렛 및 포스터 제작·배부

통계의 중요성, 통계 조사 목적, 조사 방법, 주요 조사 내용, 표본 선정 배경, 조사 결과의 활용, 비밀 보호 내용 등을 압축·수룩한 팜플렛 및 포스터를 제작

하여 조사 대상 가구에 직접 배부하거나 유관 기관이나 언론 매체를 이용한 홍보 등에 기초 자료로 활용하였다.

2) 인사장 및 감사장 인쇄

병행조사 시작 시 신표본 조사 대상 가구에는 조사 협조를 부탁하는 인사장을, 구표본 조사 대상 가구에는 그 간의 조사 협조에 대한 감사의 내용을 담은 감사장을 인쇄하여 배부하였다. 이와 동시에 동사무소나 표본 조사구가 아파트 조사구일 경우에는 아파트 관리 사무소로 조사 협조를 부탁하는 협조 공문을 발송하였다.

3) 「생활의 지혜」모음집 제작·배부

매월 가계조사 대상가구에 단편적으로 배부하던 생활의 지혜 자료를 모아서 종합적으로 그 내용을 분야별, 계절별 등으로 분류 제작하여 배부하였다.

4) 답례품 지급

신표본 대상 가구에 대한 조사 환경 개척의 대화 수단, 지속적 협조 관계 유지를 위한 유인 수단 및 조사 응답에 대한 감사의 뜻 표시로 병행조사 기간 중에 지급하였다.

5) 응답자 대상 통계퀴즈

병행 조사 실시 기간 중 통계 인식을 제고하고 협조 유도를 위해 실시하였다. 참여도를 높이기 위해 홍보물 또는 우리청 홈페이지에서 찾을 수 있는 쉬운 내용을 출제하였으며, 가급적 많은 당첨자를 선정하였다.

6) 기타 홍보

이 외에 열악한 조사 환경을 극복하기 위해 통계조사직원 명함 활용, 사무소 자체 홍보반 편성·운영, 키폰을 활용한 홍보, 통계보도 요약자료 배부 등 각종 노력을 기울였다.

VI. 추정

1. 분산과 표준오차

추정치에 대한 분산은 개개의 관측값과 표본평균의 차의 제곱에 의해 구해지므로 분산의 단위는 관측값 단위의 제곱이 된다. 그러나 대부분의 경우 자료의 분포에 대한 분석을 용이하게 하기 위해 분산의 제곱근인 표준오차(Standard error)를 이용하고, 타 항목과의 비교를 위하여 표준오차를 추정치로 나누어 백분율로 표시한 상대표준오차(Relative standard error)를 구하여 이용한다.

즉, 어떤 추정치에 대한 분산을 $var(x)$ 라고 하면 이의 표준오차는 $\sqrt{var(x)}$, 상대표준오차는 $100\sqrt{var(x)} / x$ 가 된다.

2. 경제활동인구조사의 추정

경제활동인구조사의 각 특성에 대한 총수 추정에 있어 경제활동인구조사의 표본은 확률표본으로 추출되어 있어 원칙적으로 표본추출 당시의 추출률에 따라 총수를 추정하면 되지만, 이 조사는 일회성 조사(one-shot survey)가 아니고 동일한 특성에 대하여 상이한 시점에서 시차를 두고 계속 조사하는 지속 조사(successive survey) 또는 반복 조사(repeated survey)이기 때문에 모수 추정은 층화 표본에 의한 가중추정이 아닌 모집단의 보조 정보를 활용한 비추정 방법(Ratio Estimation Method)에 의하고 있다. 이 때 사용하는 보조 정보는 층별 추계 인구를 사용하고 있다. 향후 연령대별 승수를 사용할 시는 현 추정식을 준용하여 사용하면 된다.

가. 월별 총수 추정

1) 시·도별 월별 총수의 추정

월별 조사결과에서 어떤 특성 Y를 갖는 인구의 성별 추정과 성별 합계 추정에 이용된 공식은 공식(1), 공식(2)와 같다.

$${}_s \hat{Y}_{qgh} = \sum_i {}_s \hat{Y}_{qghi} \quad \text{--- (1)}$$

$$= \sum_i {}_s M_{qghi} {}_s Y_{qghi}$$

$${}_s M_{qghi} = {}_s \hat{X}_{qghi} / {}_s X_{qghi}$$

$${}_s Y_{qghi} = \sum_j {}_s Y_{qghij}$$

$${}_s X_{qghi} = \sum_j {}_s X_{qghij}$$

$$\hat{Y}_{qgh} = \sum_s {}_s \hat{Y}_{qgh} \quad \text{--- (2)}$$

위의 공식에서

\hat{Y} = 특성 Y를 갖는 추정인구

Y = 표본에서 특성 Y를 갖는 것으로 조사된 총인구

M = 승수

\hat{X} = 15세 이상 추계인구

X = 표본에서 조사된 15세 이상 총인구

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 ($h = 1, 2, \dots, 16$)

i = 동부 및 읍면부를 나타내는 첨자 ($i = 1, 2$)

j = 표본조사구를 나타내는 첨자

이들 추정치에 대한 분산은 각각 공식(3), 공식(4)와 같이 산출한다.

$$var({}_s \hat{Y}_{qgh}) = \sum_i var({}_s \hat{Y}_{qghi}) \quad \text{--- (3)}$$

$$= \sum_i {}_s M_{qghi}^2 \xi_{hi} \sum_j Z_{qghij}^2$$

$$\xi_{hi} = [(1 - f_{hi})n_{hi}] / [2(n_{hi} - 1)]$$

$$f_{hi} = n_{hi} / (10N_{hi})$$

$$Z_{qghij} = d_s Y_{qghij} - {}_s \phi_{qghi} d_s X_{qghij}$$

$$d_s Y_{qghij} = {}_s Y_{qghij} - {}_s Y_{qghij+1}$$

$$d_s X_{qghij} = {}_s X_{qghij} - {}_s X_{qghij+1}$$

$${}_s \phi_{qghi} = {}_s Y_{qghi} / {}_s X_{qghi}$$

n = 표본조사구수

N = 추출률 조사구수

$$var(\hat{Y}_{qgh}) = \sum_i var(\hat{Y}_{qghi}) \quad \text{---} \quad (4)$$

$$= \sum_i \sum_s var({}_s \hat{Y}_{qghi}) + 2 cov({}_1 \hat{Y}_{qghi} \cdot {}_2 \hat{Y}_{qghi})$$

$$cov({}_1 \hat{Y}_{qghi} \cdot {}_2 \hat{Y}_{qghi}) = {}_1 M_{qghi} \cdot {}_2 M_{qghi} \xi_{hi} \sum_j {}_1 Z_{qghij} \cdot {}_2 Z_{qghij}$$

2) 전국 월별 총수 추정

전국의 특성 Y를 갖는 인구의 성별 추정치와 성별 합계의 추정치는 시도별 해당 추정치를 합산하여 공식(5), 공식(6)과 같이 산출하고 이의 분산은 공식(7), 공식(8)과 같이 산출한다.

$${}_s \hat{Y}_{qg} = \sum_h {}_s \hat{Y}_{qgh} \quad \text{---} \quad (5)$$

$$\hat{Y}_{qg} = \sum_h \hat{Y}_{qgh} \quad \text{---} \quad (6)$$

$$= \sum_s {}_s \hat{Y}_{qg}$$

$$var({}_s \hat{Y}_{qg}) = \sum_h var({}_s \hat{Y}_{qgh}) \quad \text{---} \quad (7)$$

$$var(\hat{Y}_{qg}) = \sum_h var(\hat{Y}_{qgh}) \quad \text{---} \quad (8)$$

나. 분기별 월평균 총수 추정

1) 시도의 분기별 월평균 총수의 추정

각 시도별로 특성 Y를 갖는 인구의 성별 및 성별 합계의 분기별 월평균

추정치는 월별 추정치를 이용하여 공식(9), 공식(10)과 같이 산출한다.

$${}_s \hat{Y}_{qh} = \sum_g {}_s \hat{Y}_{qgh} / 3 \quad \text{--- (9)}$$

$$\hat{Y}_{qh} = \sum_g \hat{Y}_{qgh} / 3 \quad \text{--- (10)}$$

그리고 이들 추정치에 대한 분산은 공식(11), 공식(12)와 같이 산출한다.

$$\text{var}({}_s Y_{qh}) = \sum_g \text{var}({}_s \hat{Y}_{qgh}) / 9 \quad \text{--- (11)}$$

$$\text{var}(Y_{qh}) = \sum_g \text{var}(\hat{Y}_{qgh}) / 9 \quad \text{--- (12)}$$

2) 전국 분기별 월평균 총수의 추정

전국의 특성 Y를 갖는 인구의 성별 및 성별 합계의 분기별 월평균 추정치는 시도별 해당 추정치를 합산하여 공식(13), 공식(14)와 같이 산출하며 이의 분산은 공식(15), 공식(16)을 이용하여 추정한다.

$${}_s \hat{Y}_q = \sum_h {}_s \hat{Y}_{qh} \quad \text{--- (13)}$$

$$\hat{Y}_q = \sum_h \hat{Y}_{qh} = \sum_s {}_s \hat{Y}_q \quad \text{--- (14)}$$

$$\text{var}({}_s \hat{Y}_q) = \sum_h \text{var}({}_s \hat{Y}_{qh}) \quad \text{--- (15)}$$

$$\text{var}(\hat{Y}_q) = \sum_h \text{var}(\hat{Y}_{qh}) \quad \text{--- (16)}$$

다. 연간 월평균 총수 추정

1) 시도 단위 연간 월평균 총수의 추정

각 시도별로 특성 Y를 갖는 인구의 성별 및 성별 합계의 연간 월평균 추정치는 공식(17), 공식(18)과 같이 산출한다.

$${}_s \hat{Y}_h = \sum_q {}_s \hat{Y}_{qh} / 4 \quad \text{--- (17)}$$

$$\hat{Y}_h = \sum_q \hat{Y}_{qh} / 4 \quad \text{--- (18)}$$

그리고 이들 추정치에 대한 분산은 공식(19), 공식(20)으로 산출한다.

$$var({}_s \hat{Y}_h) = \sum_q var({}_s \hat{Y}_{qh}) / 16 \quad \text{--- (19)}$$

$$var(\hat{Y}_h) = \sum_q var(\hat{Y}_{qh}) / 16 \quad \text{--- (20)}$$

2) 전국 연간 월평균 총수의 추정

전국에서 특성 Y를 갖는 인구의 성별 및 성별 합계의 연간 월평균 추정치는 시도별 해당 추정치를 합산하여 다음 공식(21), 공식(22)와 같이 산출한다.

$${}_s \hat{Y} = \sum_h {}_s \hat{Y}_h \quad \text{--- (21)}$$

$$\hat{Y} = \sum_h \hat{Y}_h = \sum_s {}_s \hat{Y} \quad \text{--- (22)}$$

그리고 이들 추정치에 대한 분산은 공식(23), 공식(24)로 산출한다.

$$var({}_s \hat{Y}) = \sum_h var({}_s \hat{Y}_h) \quad \text{--- (23)}$$

$$var(\hat{Y}) = \sum_h var(\hat{Y}_h) \quad \text{--- (24)}$$

라. 월별 구성비의 추정

1) 시도단위 월별 구성비의 추정

각 시도별로 특성 Y를 갖는 인구에 대한 특성 U를 갖는 인구의 구성비 추정치는 공식(25), 공식(26)으로, 이에 대한 분산은 공식(27), 공식(28)과 같이 산출한다.

$${}_s \hat{R}_{qgh} = \sum_i {}_s w_{qghi} * {}_s \hat{R}_{qghi} \quad \text{--- (25)}$$

$${}_s \hat{R}_{qghi} = {}_s \hat{U}_{qghi} / {}_s \hat{Y}_{qghi}$$

$${}_s w_{qgh} = {}_s \hat{Y}_{qghi} / {}_s \hat{Y}_{qgh}$$

$$\hat{R}_{qgh} = \sum_i w_{qghi} * \hat{R}_{qghi}$$

$$= \sum_i w_{qghi} \sum_s \hat{y}_{qghi} * {}_s \hat{R}_{qghi} \quad \text{--- (26)}$$

$$w_{qghi} = \hat{Y}_{qghi} / \hat{Y}_{qgh}$$

$${}_s \hat{y}_{qghi} = {}_s \hat{Y}_{qghi} / \hat{Y}_{qghi}$$

\hat{R} = 특성 Y를 갖는 인구에 대한 특성 U를 갖는 인구의
구성비 추정치

\hat{U} = 특성 U를 갖는 인구의 추정치

$$var({}_s\hat{R}_{qgh}) = \sum_i {}_s w_{qghi}^2 var({}_s\hat{R}_{qghi}) \quad \text{--- (27)}$$

$$var({}_s\hat{R}_{qghi}) = \xi_{hi} \sum_j ({}_s Z_{qghij})^2 / ({}_s\hat{Y}_{qghi})^2$$

$${}_s Z_{qghij} = d {}_s\hat{U}_{qghij} - {}_s\hat{R}_{qghi} * d {}_s\hat{Y}_{qghij}$$

$$d {}_s\hat{U}_{qghij} = {}_s\hat{U}_{qghij} - {}_s\hat{U}_{qghij+1}$$

$$d {}_s\hat{Y}_{qghij} = {}_s\hat{Y}_{qghij} - {}_s\hat{Y}_{qghij+1}$$

$$var(\hat{R}_{qgh}) = \sum_i w_{qghi}^2 var(\hat{R}_{qghi})$$

$$= \sum_i w_{qghi}^2 [\sum_s \hat{y}_{qghi}^2 var({}_s\hat{R}_{qghi}) + 2cov({}_1\hat{R}_{qghi} \cdot {}_2\hat{R}_{qghi})]$$

$$= (1/ \hat{Y}_{qghi}) \sum_i \sum_s [{}_s\hat{Y}_{qghi} var({}_s\hat{R}_{qghi}) + 2\xi_{hi} {}_s C_{qghi}]$$

$${}_s C_{qghi} = \sum_j d {}_1 Z_{qghij} d {}_2 Z_{qghij}$$

2) 전국 월별 구성비의 추정

특성 Y를 갖는 인구에 대한 특성 U를 갖는 인구의 전국 월별 추정치는
공식(29), 공식(30)과 같이 산출한다.

$${}_s\hat{R}_{qg} = \sum_h {}_s w_{qgh} {}_s\hat{R}_{qgh} \quad \text{--- (29)}$$

$$\hat{R}_{qg} = \sum_h w_{qgh} \hat{R}_{qgh} \quad \text{--- (30)}$$

$${}_s w_{qgh} = {}_s\hat{Y}_{qgh} / {}_s\hat{Y}_{qg}$$

$$w_{qgh} = \hat{Y}_{qgh} / \hat{Y}_{qg}$$

이에 대한 분산은 공식(31), 공식(32)와 같이 산출한다.

$$var({}_s\hat{R}_{qg}) = \sum_h {}_s w_{qgh}^2 var({}_s\hat{R}_{qgh}) \quad \text{--- (31)}$$

$$var(\hat{R}_{qg}) = \sum_h w_{qgh}^2 var(\hat{R}_{qgh}) \quad \text{--- (32)}$$

3. 전국가계조사의 추정

가. 분기별 월평균추정

가구당 월평균 가계수지의 분기별 공표를 위한 전도시의 분기별 추정치는 각 시·도의 분기별 추정치를 먼저 산출한 다음, 이를 이용하여 산출한다.

1) 각 시·도의 분기별 월평균 추정

실제로 각 시·도별로 분기별 자료는 공표하지 않지만, 각 시·도의 분기별 월평균 추정치는 전 도시 분기별 자료를 공표하기 위한 추정치 산출의 기초 추정치로 사용하기 위하여 산출하여야 한다.

각 시·도에서 특성 X 를 갖는 가구의 특성 Y 의 가구당 월평균 추정치는 공식(33), 추정치에 대한 분산은 공식(34)과 같이 산출한다.

$$\begin{aligned} \bar{Y}_{qh} &= \sum_i w_{qhi} * \bar{Y}_{qhi} \\ &= \sum_i w_{qhi} * (Y_{qhi} / X_{qhi}) \\ &= \sum_i w_{qhi} (\sum_j Y_{q hij} / \sum_j X_{q hij}) \end{aligned} \quad \text{--- (33)}$$

$$Y_{qhi} = \sum_g g Y_{qhi}$$

$$X_{qhi} = \sum_g g X_{qhi}$$

w_{qhi} : 각 시도별 동부, 읍면부의 가중치

여기에서

\bar{Y} = 특성 X 를 갖는 가구의 특성 Y 의 가구당 월평균 추정치

X = 특성 X 를 갖는 것으로 조사된 총가구수

Y = 특성 Y 에 대해 조사된 값의 합

q = 분기를 나타내는 첨자 ($q = 1,2,3,4$)

g = 월을 나타내는 첨자 ($g = 1,2,3$)

h = 시도를 나타내는 첨자 ($h = 1,2,\dots,16$)

i = 동부, 읍면부를 나타내는 첨자($i = 1,2$)

j = 표본조사구를 나타내는 첨자

$$var(\bar{Y}_{qh}) = \sum_i w_{qhi}^2 [M_{qhi} \sum_j (dY_{qhij} - \bar{Y}_{qhi} dX_{qhij})^2 / X_{qhij}^2] \quad \text{--- (34)}$$

$$M_{qhi} = [(1 - f_{hi}) n_{hi}] / [2(n_{hi} - 1)]$$

$$f_{\ominus} = n_{hi} / (10N_{hi})$$

$$dY_{qhij} = Y_{qhij} - Y_{qhij+1}$$

$$dX_{qhij} = X_{qhij} - X_{qhij+1}$$

n = 표본조사구수

N = 총추출단위 조사구수

2) 전 도시 분기별 월평균 추정

전도시의 분기별 월평균 추정치와 이에 관한 분산은 다음의 공식(35), 공식(36)와 같이 산출한다.

$$\bar{Y}_q = \sum_h w_{qh} \bar{Y}_{qh} \quad \text{--- (35)}$$

$$w_{qh} = P_h X_{qh} / \sum_h P_h X_{qh}$$

P_h = 추출확률의 역수

$$var(\bar{Y}_q) = \sum_h w_{qh}^2 var(\bar{Y}_{qh}) \quad \text{--- (36)}$$

나. 연간 월평균추정

가구당 월 평균 가계 수지의 연간 공표를 위한 전 도시의 연간 추정치는 각 시·도의 연간 추정치를 먼저 산출한 다음, 이를 이용하여 산출한다.

1) 각 시·도의 연간 월평균 추정

각 시·도의 연간 월 평균 가계수지의 추정치와 분산은 공식(37)과 공식(38)과 같이 산출한다.

$$\bar{Y}_{qh} = \sum_i w_{hi} * \bar{Y}_{hi}$$

$$\begin{aligned}
&= \sum_i w_{hi} * (Y_{hi} / X_{hi}) \\
&= \sum_i w_{hi} (\sum_j Y_{hij} / \sum_j X_{hij}) \quad \text{--- (37)}
\end{aligned}$$

$$var(\bar{Y}_h) = \sum_i w_{hi}^2 [M_{hi} \sum_j (dY_{hij} - \bar{Y}_{hi} dX_{hij})^2 / X_{hij}^2] \quad \text{--- (38)}$$

$$M_{hi} = [(1 - f_{hi}) n_{hi}] / [2(n_{hi} - 1)]$$

$$f_{hi} = n_{hi} / (10 N_{hi})$$

$$dY_{hij} = Y_{hij} - Y_{hij+1}$$

$$dX_{hij} = X_{hij} - X_{hij+1}$$

n = 표본조사구수

N = 총추출단위 조사구수

2) 전도시 연간 월평균 추정

전도시의 연간 월평균 추정치와 이에 관한 분산은 다음의 공식(39), 공식(40)와 같이 산출한다.

$$\bar{Y} = \sum_h w_h \bar{Y}_h \quad \text{--- (39)}$$

$$w_h = P_h X_h / \sum_h P_h X_h$$

P_h = 추출확률의 역수

$$var(\bar{Y}) = \sum_h w_h^2 var(\bar{Y}_h) \quad \text{--- (40)}$$

VII. 표본의 관리

1. 전입 및 전출 가구의 처리

가구 표본 조사를 실시하는 조사구역에서는 항상 거처의 신축 및 철거, 가구의 전출 및 전입으로 이동이 있게 되므로 이런 변화가 표본에 반영되어야 한다.

조사구역 내에 새로운 가구가 전입하였을 때는 전입 이후부터 조사대상 가구로 하여 조사하고, 전출 가구에 대해서는 조사를 중지한다. 조사지역 내에 신축된 주택에 전입한 가구도 표본 대상 가구로 하여 조사한다.

2. 불응가구의 처리

경제활동인구조사의 한 가구는 전국의 약 440가구를 대표하게 되고 전국 가계조사는 약 1,430가구를 대표하게 되므로 한 가구의 불응은 조사 결과에 큰 영향을 미치게 된다. 현재 경제활동인구조사에서는 불응가구가 거의 없지만 전국 가계 조사에서의 불응률은 약 18% 수준이다. 이 수치도 외국에 비하면 낮은 수준이지만 여전히 응답을 유도하기 위한 다각적인 노력이 필요하다.

불응가구가 발생하면 결과적으로 표본의 크기가 줄어들게 되어 추정치의 효율을 떨어뜨릴 뿐만 아니라 완전한 자료를 바탕으로 하는 표준적인 통계적 추론을 사용하는 데도 문제가 발생할 수 있다. 더욱이 무응답자의 특성이 응답자의 특성과 체계적인 차이를 나타낸다면 이로부터 무응답 편향(Nonresponse Bias)이라는 편향이 생기게 되고, 무응답의 정확한 특성이 파악되지 않으면 이들 편향을 제거하기도 어렵다.

이러한 불응가구가 최소화하기 위해서는 부단한 예방적 노력이 필요하다. 일반적인 예방적 노력에는 다음과 같은 방법들이 있다.

1. 수 차례 재방문(More Call Attempts)
2. 조사원에 대한 업무자료의 명확화
3. 추가권유(Follow-up Reminders)
4. 인센티브(Incentives)

5. 보증(Endorsements)
6. 안내편지(Lead Letter)
7. 대리응답자 활용(Proxy Respondents) 등

그러나 이러한 예방적 노력에도 불구하고 조사 환경은 점차 악화되어 불응 가구는 점차 증가하는 추세이다. 발생한 불응가구에 대한 처리로 조사과정과 추정 과정으로 분리하여 생각해 볼 수가 있다. 조사 과정에서는 불응가구를 조사구 내 다른 조사 구역에서 가구를 대체하는 방법이 되고, 처리 과정에서는 불응가구에 대한 통계적 처리 방법이 있다. 하지만 어떤 방법을 사용하더라도 무응답 편향은 없어지지 않는다. 가구를 대체하는 방법에서는 응답을 자체는 제고될 수 있으나 불응가구와 그 특성이 완전히 일치할 수가 없으며, 중복하여 처리하는 통계적 처리 방법에서도 불응가구와 응답가구 간의 특성이 같다는 가정을 만족시켜 주지는 않는다.

대체할 가구를 불응가구의 가구원 구성, 가구주의 직업, 산업, 소득수준, 연령 및 교육정도 등의 특성과 유사한 가구를 선정할 수 있다면 편향이 완전히 제거되지는 않는다 하여도 상당히 줄일 수는 있을 것이다. 그러나 현지 조사원이 조사 시 조사 대상 가구가 비협조적일 때 설득 노력을 더 기울이지 않고 세심한 선정과정 없이 조사가 가능한 다른 가구들로 대체하는 경우에는 오히려 더 큰 편향을 가져올 수도 있다.

불응가구에 대한 추정 과정에서 응답가구의 자료를 활용하여 불응이 없는 완전자료를 만들어 추정하는 방법에서의 단점은 추정치의 분산이 과소 추정(Under-estimation)될 우려가 높다. 과소 추정된 분산 추정치를 사용하면 추정치의 변이계수나 신뢰 구간이 실제보다 작고 좁게 나타나 추론의 오류를 범하게 될 수 있으므로 세심한 주의와 연구 분석을 통해 사용하여야 한다.

VIII. 제언

1. 다목적표본

다목적 표본(Multi-purpose sample)은 인력, 예산 등의 조사 여건상 여러 조사에 대한 조사별 표본을 별도로 설계할 수 없을 때, 그에 대안으로 고안된 표본 설계기법이다. 따라서 여건이 허용한다면 표본을 달리 설계하는 것이 당연히 좋은 표본을 추출할 수가 있다. 금번 표본에서도 여러 여건상 경상 조사와 연간 및 특별조사의 조사 대상을 분리하는 정도에서 다목적 표본을 사용치 않을 수 없었다.

그러나 연간조사에는 정보화실태조사가 정보통신부로 이관되어 사회통계 조사만이 남게 되었고, 5년 주기의 특별조사로는 가구소비실태조사와 생활시간 활용조사의 2종이 있다. 그 간의 조사 경험으로 보아서 어떤 조사든 간에 응답자가 제대로 응답을 하는 데 3~4개월이 소요되는 것으로 판단이 되고, 가장 불응이 심한 조사는 가계부를 쓰는 전국가계조사로 5년 전보다도 거의 10% 정도가 응답율이 떨어지는 현실을 병행조사 기간에 볼 수 있었다.

따라서 불응이 심하지 않는 경제활동인구조사와 친숙하지 않는 연간 및 특별 조사를 하나의 표본에서 조사하는 다목적 표본으로 설계하고, 전국가계 조사는 경제활동인구조사와는 완전히 다른 표본을 추출하여 고품질의 답례품 지급 등 특별 관리하는 것이 좋겠다.

2. 병행조사

금번에 병행조사를 3개월 실시하였는데, 향후에는 조사에 대한 이해, 목적, 비밀보호 등에 관한 순수 설득 기간을 별도로 1개월을 실시하는 것이 바람직할 것으로 판단된다. 새로이 표본으로 선정된 가구에서 처음부터 조사에 들어가는 것은 많은 거부감을 야기하여 조사표가 부실하여 분석자료로 사용이 어려워 실제적으론 2개월간의 자료로 신·구표본의 시계열을 분석하기엔 시간적으로 촉박한 것으로 판단된다.

부

록

