

## 농업기본통계조사를 위한 새로운 표본설계

남궁 평<sup>1)</sup>, 김규성<sup>2)</sup>, 변종석<sup>3)</sup>, 최경호<sup>4)</sup>

### 요 약

농업기본통계조사의 현행 표본은 1995년 농업총조사 결과를 모집단으로 선정한 것이므로 표본설계 시 고려되는 모집단의 기본 구조 변화 및 「2000년 농업총조사」 결과에서 나타난 변화된 특성을 제대로 반영하지 못하고 있으므로 「2000년 농업총조사」 결과를 바탕으로 모집단을 새로이 구성하여 변화하는 농가 구조를 반영하고, 기존 표본의 노후화를 방지하며, 기존 통계 생산을 유지하면서 신규 통계를 생산할 수 있도록 표본설계를 하였다.

주요 용어 : 농업기본통계조사, 조사구, 층화추출법, 변이계수

### 1. 서 론

농업기본통계조사는 매년 실시되는 계속조사로써 우리나라 농업의 기본 지표인 농가, 농가 인구, 영농형태 등 농업의 구조를 파악하여 필요한 통계자료를 생산하고, 농업정책 수립 및 농업관련 연구기관의 연구를 위한 기초자료를 제공하기 위해 실시되는 표본조사이다.

<표 1>과 같이 「2000년 농업총조사」 결과에 의하면 농가 및 농가 인구는 최근의 경기 부진으로 인해 감소 추세는 둔화되었지만 지속적으로 감소하고, 인구의 고령화 현상은 더욱 가속화되고 있는 반면, 농업의 경영 구조가 규모화·전문화되어 대내외 여건 변화에 빠르게 대응하는 특성이 나타나고 있다.

<표 1> 2000년 농업총조사의 농가 및 농가 인구 증감율

구 분		농업총조사 시행 연도		
		1995년	2000년	증감율(%)
가구 (단위:천가구)	전체	12,958	14,312	10.4
	농가	1,501	1,383	-7.8
인구 (단위 : 천명)	전체	45,093	47,008	4.2
	농가	4,851	4,031	-16.9

농업기본통계조사의 표본은 5년 주기의 표본 개편으로 인해 모집단 변화의 반영이 미흡하고, 현행 표본도 1995년 농업총조사 결과를 모집단으로 하여 표본조사구 내 농가를 대상으로 선정한 것이므로 표본설계 시 고려되는 모집단의 기본 구조 변화 및 「2000년 농업총조사」 결과에서

- 1) 성균관대학교 통계학과 교수, 서울 종로구 명륜동 3가 53번지
- 2) 서울시립대학교 컴퓨터·통계학과 부교수, 서울시 동대문구 전농동 90
- 3) 한신대학교 정보통계학과 부교수, 경기도 오산시 양산동 411
- 4) 전주대학교 정보기술컴퓨터공학부 부교수, 전북 전주시 완산구 효자동 3가

나타난 변화된 특성을 제대로 반영하지 못하기 때문에 현행 표본에 대한 검토가 필요하다고 생각된다.

따라서 본 연구에서는 「2000년 농업총조사」 결과를 바탕으로 현재의 농촌 실정에 맞는 모집단을 새로이 구성하여 농촌 환경을 잘 반영하면서 농가를 대표하는 새로운 표본을 추출하여 변화하는 농가의 인구사회 및 경제구조를 반영하고, 기존 표본의 노후화를 방지하며, 기존 통계 생산 단위를 가능한 유지하면서 신규 통계를 생산할 수 있도록 변화된 농가의 특성을 반영하는 표본설계를 통해 표본 개편을 검토해 보고자 한다.

## 2. 현행 표본설계 분석

현행 표본설계는 1995년 농업총조사 자료를 기초로 전국의 농가를 모집단으로 작성되었다. 시·군별 통계를 산출하기 위해 서울과 6개 광역시를 포함하는 167개 시·군(충남 계룡출장소와 논산군 통합, 충북 증평출장소와 괴산군 통합)을 부차모집단으로 설계하였으며, 표본추출단위는 인구주택조사구를 사용하였다. 그리고 각 부차 모집단내의 농가수 분포에 따라 3~5개의 층으로 층화하여 각 층마다 단순임의 추출하는 단순 층화추출법을 이용하였다.

또한 현행 표본설계는 설계 당시 구(區)가 있는 서울, 광역시 그리고 각 도의 일부 시는 농가수나 농가 인구수가 대체로 적기 때문에 구를 무시하고 하나의 부차모집단으로 시·군별로 5%이내를 원칙으로 하였다. 그러나 경기 수원시, 의정부시, 양주군을 비롯한 13개의 시·군의 경우 1997년의 설계 표본수가 1992년에 비해 급격히 증가하기 때문에 목표 정도를 완화하여 5%를 넘기도 하였다.

현행 표본설계에 대한 분석은 통계청에서 제공한 2001년 농업기본통계조사의 표본 자료를 대상으로 분석하였다. 농업기본통계조사의 현행 표본설계는 1997년에 설계되었으며, 당시 전국의 인구주택조사구 수 68,931개 중 3,059개로 구성되어 있으며, 2001년에 실시된 농업기본통계 조사에서는 농가가 있는 표본 조사구 수가 2,979개로 조사되어 설계 당시보다 80개의 조사구가 농가가 없는 조사구로 나타났다. 이는 설계 당시 표본으로 선정된 일부 조사구에서 이농현상으로 인해 농가가 없어진 조사구이거나 행정구역이 개편됨에 따라 조사구 자체가 변경되었거나 없어졌기 때문으로 생각된다.

「2000년 농업총조사」에서는 전국의 인구주택조사구 수가 72,330개로 조사되었으며, 현행 표본의 추출율은 전국 기준 4.12%이었으며, 서울과 울산을 제외한 5개 광역시와 제주도가 3%이내, 강원도가 6.67%, 나머지 시·도는 3~5%이하이다.

현행 표본설계의 추계오차를 살펴보면, 현행 표본설계는 시·군별 농가수, 농가 인구수 등을 산출할 수 있도록 설계되어 있으며, 이를 근거로 전국의 농가수, 농가 인구수 및 광역시·도별 농가수, 농가 인구수 등에 대한 추계치를 산출하고 있다. 현행 표본설계 당시 부차모집단인 시·군의 수는 167개이었으나 현재는 행정구역이 변경, 통합되어 전국이 1개 특별시, 6개 광역시, 9개도 74개 시(전남의 여천시와 여천군은 여수시로 통합), 84개 군(충남 계룡출장소는 논산시로 통합, 충북 증평출장소는 괴산군으로 통합) 등 모두 165개로 변경되었다.

<표 2> 현행 표본설계의 시·도별 표본변이계수

시·도	CV(%)	시·도	CV(%)
서울	19.00	충북	5.00
부산	7.10	충남	4.08
대구	5.74	전북	3.97
인천	6.07	전남	3.22
광주	5.09	경북	3.27
대전	5.95	경남	3.48
울산	5.16	제주	7.92
경기	3.20	전국	1.25
강원	3.86		

2001년 표본자료에 대한 시·도별 표본변이계수(CV)를 계산해 본 결과, 서울이 매우 크며, 6개 광역시가 8%이내, 도지역은 제주도를 제외한 나머지 도는 5%이내인 것으로 나타났다. 따라서 광역시는 추계오차가 비교적 크지만, 제주도를 제외한 도 지역은 도별 통계자료를 얻기 위한 현행 표본설계가 비교적 목표정도를 달성하고 있다고 볼 수 있다.

또한 현행 설계에서 부차모집단별 허용오차를 보면 165개의 시·군 중 17개 지역인 10.3%만이 목표 정도 5%이내에 속해 있고, 부차모집단의 75.2%가 목표 정도 5~10%이내에 속해 있다. 반면 부차모집단의 14.5%는 추계오차가 10%를 넘는 것으로 나타나고 있다. 이는 현행 표본설계가 1995년 농업총조사를 기초로 1997년에 설계되었으므로 2001년에는 모집단이 많이 변화되었고, 인구주택조사구 수가 72,330개로 1997년 설계에 비해 5%나 증가되었으며, 또한 농가수 및 농가 인구수의 감소 등 시간이 흐름에 따라 현재의 표본이 모집단의 변화를 반영하지 못하고 있음을 알 수 있다.

### 3. 모집단 분석

#### 1) 표본추출 조사구의 선정

농업기본통계조사에서 사용 가능한 표본추출 조사구는 인구주택조사구, 농업조사구, 행정구역의 행정리(혹은 부락)이다. 실제로 이들 3개 조사구의 관계는 매우 복잡한 관계를 나타내므로 본 설계에서는 적절한 표본추출 조사구를 선정하기 위한 기준을 마련하여 새로운 표본설계의 표본추출 조사구로 사용 가능한 인구주택조사구와 농업조사구에 대해 모집단 특성을 분석·비교한 후, 농업기본통계조사에서 사용할 수 있는 적절한 표본추출 조사구로 인구주택조사구를 선택하게 되었다.

#### 2) 인구주택조사구 분포

본 표본설계에서는 「2000년 농업총조사」 자료를 기초로 하여 전국의 농가를 모집단으로 설정하였다. 본 연구에서 고려하는 표본설계의 주된 목적이 시·군별 부차모집단의 농가수 및 농가 인구수에 대한 추계이므로 모집단 자료는 광역시 및 도의 시·군부를 부차모집단으로 분석하였다.

<표 3> 전국 시·도별 인구주택조사구 분포

시·도	조사구수	평균농가수	표준편차	CV(%)	왜도	첨도
서울특별시	838	2.52	5.88	232.9125	8.0101	78.3941
부산광역시	1,059	7.94	11.26	141.7626	1.9266	3.3468
대구광역시	2,073	6.22	10.29	165.5396	2.4313	5.6024
인천광역시	1,169	13.14	16.57	126.1285	1.2574	0.4645
광주광역시	1,417	8.44	13.88	164.4691	1.9979	2.8844
대전광역시	1,214	5.83	9.89	169.6489	2.5504	6.0679
울산광역시	1,240	9.93	12.88	129.8203	1.4947	1.2797
경기도	11,668	12.73	13.83	108.7008	1.2144	0.7906
강원도	4,795	16.15	16.45	101.8486	0.8343	-0.4013
충청북도	4,594	20.05	18.29	91.2213	0.5276	-0.9584
충청남도	7,218	24.88	25.28	101.6159	18.7860	891.9292
전라북도	6,159	22.31	19.44	87.1379	1.1489	12.3814
전라남도	7,644	30.43	19.63	64.4992	-0.0696	-1.0365
경상북도	10,468	22.57	19.22	85.1915	0.4049	-1.1425
경상남도	8,417	20.24	18.27	90.2570	0.5071	-1.0659
제주도	2,357	16.59	15.50	93.3742	0.9608	-0.0924
전국	72,330	19.13	19.21	100.4081	4.9848	272.8700

모집단 분석에서는 다음과 같은 사항을 고려하여 인구주택조사구 분포를 살펴보았다.

첫째, 구(區)가 있는 서울과 6개의 광역시, 그리고 각 도의 시지역은 농가수 및 조사구 수가 적어서 추계오차를 5% 이내로 설정하기 위해 상대적으로 많은 표본이 요구되므로 구를 무시하고 하나의 부차모집단으로 분석하였다.

둘째, 광역시의 경우 군 지역이 광역시로 편입되면서 도농 통합 광역시의 성격을 보이고 있으므로 광역시 중 도농 통합 광역시의 성격이 짙은 부산(기장군), 대구(달성군), 인천(강화군과 옹진군), 광주(광산구), 울산(울주군) 등 5개 광역시는 편입된 군 지역을 군부로 분리하여 추계하기 위해 광역시의 모집단을 시부와 군부의 부차모집단으로 나누어 검토하였다.

셋째, 충북 증평출장소와 충남 계룡출장소는 행정구역상 인접한 괴산군과 논산시에 편입시켜 하나의 모집단으로 분석하였다.

넷째, 전남에서는 여천군과 여천시가 여수시와 통합되었기 때문에 하나의 모집단으로 분석하였다.

이와 같은 사항을 고려하여 모집단을 검토한 결과, 행정구역은 서울과 6개 광역시, 74개의 시, 84개의 군으로 구성되어 있으나 본 설계에서의 부차모집단은 서울, 6개 광역시의 6개 시부와 6개의 군부, 나머지 158개의 시·군 등을 종합하면 시부 81개, 군부 90개를 합하여 모두 171개의 부차 모집단으로 구성하게 되었다.

#### 4. 새로운 표본설계

##### 1) 층화 기준 결정

본 설계에서의 층화 지표는 각 시·군별 부차모집단에서의 인구주택조사구별 농가수로 정하였다. 본 설계에서는 부차 모집단인 171개 시·군 단위의 부차모집단별 농가수 추계오차를 5% 이내로 유지하면서 현행 표본수를 늘리지 않고 시·군 통계를 정확히 내기 위해서는 층화추출 방법이 적절하므로 다음과 같은 원칙으로 층화하였다.

첫째, 시·군 단위의 추계치에 대한 분산을 최소화시키는 최적 경계점을 찾기 위해 근사법인 Dalenius 와 Hodges(1959)가 제안한 방법에 의해 조사구내의 농가수에 대한 도수의 제곱근에

대한 누적합이 가능한 같아지도록 층의 경계점을 찾아 층화하였다.

둘째, 일부 시 지역에서는 농가수가 1인 조사구의 빈도수가 많이 나타나기 때문에 층화 최적 경계점의 방법으로 층화하는 것이 불가능하다. 이러한 지역에서는 가능한 층의 범위를 넓게 하여 각 층의 가중치와 표준편차 곱인  $W_h \sigma_h$ 가 비슷해지도록 층화하였다. 농가수가 1인 조사구만을 포함하는 층을 만드는 경우 전체 층화 분산은 줄어들어 표본 오차를 줄일 수 있지만, 이런 층에서 추출된 표본은 연속조사에서 시간이 흐를수록 모집단 변동에 대해 민감해지므로 모집단 변화에 민감해지지 않도록 층의 간격을 넓혀주는 것이 타당하기 때문이다.

셋째, 부차모집단별 층의 수는 가능한 적게 정하였다. 층의 수가 많아지면 층화 효과를 증대시킬 수 있지만 변화하는 모집단의 추세를 반영하기 어려우며, 층이 세분화되어 있는 경우 모집단이 지속적으로 변화한다면 선정된 표본이 설게 당시의 층을 벗어날 수 있어 층의 분산을 오히려 크게 하여 추계치의 정도를 떨어뜨릴 수도 있다. 따라서 현행 방식과 같이 부차모집단별로 3~5개의 층으로 결정하였으며, 현행 표본수나 새로운 표본수로써 목표오차를 만족할 수 있도록 최소의 층수로 결정하였다.

넷째, 농업기본통계조사의 연속성을 가지기 위해 현행 표본설계의 층수와 현행 표본설계의 표본수로 본 설계에서의 표본수가 목표정도인 5%를 만족하도록 층의 수를 조정하거나 층의 간격을 조정하여 층화하였다. 그리하여 현행 표본설계의 표본수로써 목표정도를 달성할 수 있도록 층화하여 기준을 설정하였다.

이상의 방법을 동시에 고려하여 층화 기준을 정하였으며, 이는 현실적인 최적 층화 기준으로 판단된다.

## 2) 표본수 결정 및 표본배정 원칙

본 설계에서는 앞서 설명한 바와 같이 5개의 도농 통합 광역시를 시부와 군부로 분리하고, 개편된 행정구역에 따라 171개 부차모집단으로 재편성하였다.

표본수를 결정할 때에는 4장 1절의 층화 기준 결정에서 설명한 바와 같이 현행 표본설계의 부차모집단별 표본수보다 가능한 적은 개수의 표본으로 목표정도에 도달하도록 층 수나 층의 간격을 조정하면서 표본수를 결정하였다. 그러나 현실적으로 지나치게 적은 표본수로 추계하여 상대변이계수가 크게 나타나거나 도농 통합 광역시를 시부와 군부로 분리하여 표본수를 정하였기 때문에 현행 표본수보다 크게 결정된 지역도 나타나게 되었다.

본 설계에서의 표본수 결정 및 표본배정의 기본 원칙은 다음과 같다.

첫째, 본 설계에서의 표본배정은 네이만 배정을 원칙으로 하였다. 다만 부차모집단별 표본수가 작아 층으로 배정된 표본수가 1인 경우에는 최소 표본수를 2개로 하여 배정하였다. 이는 층내 분산 추정이 가능하도록 하기 위한 것이다. 실제로 18개의 부차모집단에서 최소 표본수의 확보를 위한 층의 표본수 조정이 있었다.

둘째, 본 설계에서는 앞서 설명한 층화 기준 결정 원칙에 따라 층수 및 층의 간격을 정하여 각 층의 가중치와 표준편차 곱인  $W_h \sigma_h$ 가 대체로 비슷해지도록 층의 경계를 정하여 표본을 배정하였다.

셋째, 본 설계에서의 부차모집단별 표본수는 부차모집단별 목표정도를 5% 이내로 하는 표본수를 계산한 후, 현행 표본수에 비해 크게 증가하는 일부 지역에서는 목표정도를 다소 완화하여 현행 표본설계에 의해 계산된 상대변이계수를 감소시키는 관점에서 표본수를 조정하는 방법으로 표본수를 결정하였다.

그리고 본 설계에서는 모집단의 변동을 대비하기 위해 층화 기준을 결정할 때 농가수가 1인 조사구만을 포함하는 층은 층의 간격을 넓게 설정하였으나 대체로 변동이 크지 않아 다른 층에 비해 표본수의 추출율이 상대적으로 낮았다.

3) 부차모집단별 표본수 및 변이계수(CV)

본 설계에서 사용한 부차모집단별 층의 수는 3~5개이며, 층의 간격은 부차모집단마다 모두 다르게 설정하였다. 그러나 본 설계의 표본수로 목표 정도 5%를 달성하지 못하는 16개의 부차모집단이 존재하게 되었는데, 이들 지역의 경우 목표 정도를 달성하기 위한 표본수가 현행 표본수보다 매우 커지기 때문에 조사원의 업무량이 급격히 증가하게 될 수 있으므로 현행 표본수를 지나치게 넘지 않도록 조정하였기 때문이다. 특히, 처음으로 광역시의 시부와 군부를 분리하여 추계하려는 목적에서 표본수를 계산하였기에 많은 표본수가 필요하게 되어 표본수의 조정이 불가피하였다.

<표 4>는 본 표본설계에서 산정된 시·도별 표본수를 1997년의 표본수, 2001년 현행 표본수와 비교하여 정리한 표이다. 본 설계의 표본수는 3,011개로 72,330개의 인구주택조사구의 4.16%에 해당하며, 시·별 표본수의 추출율은 강원도가 지역의 특수성으로 인해 6.6%로 높은 반면, 제주도는 2.0%이하로 낮았다. 그 외 시·도는 2~5%이하의 추출율을 나타내고 있다.

<표 4> 본 설계의 시·도별 표본수

시·도	총조사구수	'97 표본수	'01 표본수	본 설계 표본수 (증감)		추출율(%)
서울	838	41	36	34	(-2)	4.057
부산	1,059	32	30	37	(7)	3.494
대구	2,073	50	48	49	(1)	2.364
인천	1,169	37	35	47	(12)	3.397
광주	1,417	40	40	48	(8)	3.387
대전	1,214	28	26	27	(1)	2.224
울산	1,240	44	43	43	(0)	3.468
경기	11,668	513	487	537	(50)	4.602
강원	4,795	328	320	316	(-4)	6.590
충북	4,594	181	179	192	(13)	4.179
충남	7,218	280	277	302	(25)	4.184
전북	6,159	252	242	249	(7)	4.043
전남	7,644	369	369	309	(-60)	4.042
경북	10,468	424	420	410	(-10)	3.917
경남	8,417	379	366	365	(-1)	4.336
제주	2,357	61	61	46	(-15)	1.952
전국	72,330	3,059	2,979	3,011	(32)	4.163

- \* '01 표본수는 2001년 농업기본통계조사에서 농가가 있는 표본 조사구 수임.
- \* 표본수 증감은 본 설계 표본수에서 2001년 표본수를 차감한 수임.
- \* 추출율은 본 설계 표본수를 2000년 농업총조사의 조사구 수로 나눈 비율임.

5. 결론

최근 농가 및 농가인구는 IMF 이후 최근의 경기 부진으로 인해 감소 추세는 둔화되었지만 지속적으로 계속 감소하고 있으며, 인구의 고령화 현상은 더욱 가속화되고 있어 모집단의 변화가 빠르게 진행되고 있다. 또한 현재 농업기본통계조사의 표본은 5년 주기의 개편으로 인해 표본설계 시 고려되는 모집단의 기본 구조 변화 및 변화하는 농촌 특성의 반영이 미흡한 상황이므로 본 설계에서는 모집단의 변화를 효과적으로 반영하기 위한 표본설계를 제안하였다.

먼저 본 설계에서는 새로운 표본설계를 위한 조사구를 검토하였으나 농업조사구를 조사구로 선정하면 조사 이동 거리의 감소, 조사구의 관리와 표본 조사구 수의 감소 효과를 기대할 수 있지만, 급격한 업무량의 증가, 조사구내의 동질성 약화, 조사지역의 제한 등으로 인해 모집단

변화의 반영이 인구주택조사구보다 다소 미흡하기 때문에 인구주택조사구를 본 표본설계의 조사구로 선정하였다.

그리고 본 설계는 계속조사의 연속성을 유지하면서 변화하는 모집단의 특성을 반영하기 위해 이전의 설계 방법과 동일한 방법으로 설계하였으며, 전국을 171개의 부차모집단인 시·군으로 나누어 인구주택조사구의 농가수 분포를 살펴본 후, 변화된 농촌의 현실을 반영하는 표본설계를 수행하였다.

본 설계에서는 변화된 인구주택조사구의 농가수를 층화 기준으로 설정하여 각 부차 모집단별로 층화하여 목표정도 5%를 확보하기 위한 표본수를 결정하여 네이만 배정으로 표본을 각 층에 배정하여 새로운 표본설계에서는 전국의 표본수가 3,011개로 현행 표본에 비해 약간 증가한 것으로 나타났다. 이러한 결과는 일부 광역시에서 시부와 군부를 분리하여 추계하려는 목적 때문에 나타난 결과이다. 시·군별로 표본 조사구의 변화가 있었지만 일부 지역을 제외하면 대체로 표본수의 증감 현상은 그리 크지 않았다.

새로운 표본설계에 의한 표본 조사구 수는 3,011개로 2001년 기준 현행 표본에 비해 32개 조사구의 증가를 보였지만 1997년의 표본수에 비해 48개 조사구 만큼 감소되었다. 또한 조사 표본 농가수를 보면, 2001년 현행 표본에서는 70,948가구이었으나 새로운 표본설계에서는 70,622가구로 약간 감소하는 것으로 나타나 농가수나 농가인구수의 감소 추세를 보이며 변화하고 있는 특성을 적절히 반영한 표본설계로 보인다.

그리고 농업기본통계조사는 계속조사이므로 조사구 관리, 표본 관리 및 표본 대체가 추계에 많은 영향을 주게 되므로 조사구 관리 및 표본 관리, 조사 자료의 관리면에서 체계적으로 관리할 필요가 있다. 또한 농가수의 변화로 인해 표본 조사구의 대체도 일정 부분 지속적으로 이루어져 추계치의 표본오차가 커지지 않도록 신경을 써야 할 것이다.

### [참고문헌]

- 남궁 평(2001), “최신 표본조사 설계와 분석”, 탐진  
박홍래(2000), “통계조사론”, 영지문화사  
서울대학교 통계연구소(1992), 「농업기본통계조사 표본설계」  
서울대학교 통계연구소(1997), 「농업기본통계조사 표본설계」  
통계청(1999), 「1998년 농업기본통계조사보고서」  
통계청(2000), 「1999년 농업기본통계조사보고서」  
통계청(2001), 「2000 농업총조사 분석보고서」  
한국통계학회(2002), 「농업기본통계조사 표본설계」  
Cochran, W.G.(1977), *Sampling Techniques*, 3rd edition. John Wiley & Sons.  
Lessler, J. T. and Kalsbeek, W. D.(1992), *Nonsampling Error in Surveys*. John Wiley & Sons.  
Thompson, S.K.(1992), *Sampling*, John Wiley & Sons.