연구논문

농가경제조사를 위한 새로운 표본설계방법

A Sample Design for the Farm Household Economy Survey in Korea

김진[a] · 최통진[b]
Jin Kim · Tongjin Choi

2012년 농가경제조사의 표본개편에서는 층화2단집락추출법(stratified 2−stage cluster sampling) 대신에 통합이중추출법(integrated double sampling)을 적용하였다. 통합이중추출법이란 표본설계를 크게 2단계로 구분하여, 1단계에서는 대규모의 표본을 1차 추출(1st phase)하고, 2단계에서 1차 추출된 표본을 표본 추출틀로 하여 나머지 조사들의 표본을 2차 추출(2nd phase)하는 방법이다. 본 연구에서 1차 추출은 모집단 추출틀로부터 농업조사의 3,371개 표본 수를 추출하는 단계이고, 2차 추출은 농업조사 73,815 농가 중 농가경제조사의 2,600 표본농가를 추출하는 단계이다. 또한, 조사구를 층화추출하던 영역추출법(area frame sampling) 대신에 농가를 직접 층화추출하는 리스트 추출법(list frame sampling)을 적용하였고, 총조사 이후 농가 변동이 반영된 최신 정보를 활용하여 표본을 추출하였다. 이러한 새로운 시도를 통하여 농가경제조사의 표본 대표성과 효율성 향상을 기대할 수 있다.

주제어: 통합이중추출, 층화2단집락추출, 영역추출, 리스트추출

The stratified 2−stage cluster sampling method for the Farm Household Economy Survey(FHES) has been changed into a integrated double sampling in 2012. The integrated double sample means that the sample for the Agriculture Survey functions as a master sample for the other farm samples i.e. the FHES etc. The sample for the FHES is extracted by the double sampling and the list

a) 교신저자(corresponding author): 통계청 통계정책국 표본과 서기관 김진.
E−mail: jinknso@korea.kr

b) 통계청 통계정책국 표본과

frame sampling. These new sampling methods make the FHES more efficient than previous sampling one.

Key words: integrated double sampling, stratified 2-stage cluster sampling, list frame sampling

Ⅰ. 서 론

농가경제조사는 일정규모 이상의 농가를 대상으로 농가의 소득과 자산 및 부채를 매월 파악하는 통계조사이다. 이 조사는 1953년에 최초로 시작되고, 1962년 공식통계로 지정된 이후 5년마다 표본이 새롭게 개편되었으며, 2012년에는 12차 표본개편이 이루어졌다. 기존의 표본추출법은 공식통계로 지정된 이후로 계속 층화2단집락추출(stratified 2-stage cluster sampling)을 유지해 왔다. 층화2단집락추출은 조사모집단을 위한 표본 추출틀을 농업총조사 결과자료로 형성하여 농업총조사 조사구를 1차 추출하고, 추출된 조사구 내에서 적정 농가 수를 표본농가로 선정하는 추출법이다. 확률추출법을 도입하기 시작한 1960년대부터 70년대까지는 2차 추출 단위 수를 조사구 내의 약 15농가로 하였고, 1982년부터 2002년까지는 10농가로 축소하였다. 2007년 표본개편에서는 2차 추출 단위 수를 10농가에서 5가구로 줄여, 동일한 표본오차를 유지하면서 표본규모를 3,200가구에서 2,800가구로 줄인 바 있다(통계청 2007).

2012년 농가경제조사 표본개편의 특징은 통합이중추출(integrated double sampling)이다. 이중추출(double sampling)은 중복추출 또는 이상추출(two-phase sampling)이라고도 하며, 표본설계 과정이 다음과 같이 2단계로 구분된다(김재광 2008).

[1단계] 대규모 표본을 1차 추출(first phase sample)하여, 그 표본으로부터 x를 관측한다.

[2단계] 1차 표본에서 관측된 x의 정보를 이용하여 2차 표본(second phase sample)을 추출한다.

통계청은 농업조사, 농가경제조사, 농축산물생산비조사, 양곡소비량조사, 가축동향조사 등 5종의 농가를 통계조사를 실시하고 있다. 이 중 농업조사는 매년 약 75,000농가를 조사하는 대규모 표본조사이다(통계청 2011). 이러한 대규모 표본조사인 농업조사의 표본을 1차 추출하여 표본농가의 정보를 업데이트한 후, 농가경제조사, 농축산물생산비조사, 양곡소비량조사의 표본을 2차 추출하였다. 이때, 농가 대상 표본조사들의 표본을 대규모 조사인 농업조사 표본과 통합한다는 의미에서 통합이중추출이라 명명하였다. 이러한 이중추출법의 적용은 두 가지 측면에서 농가경제조사의 효율성을 기대할 수 있다. 첫째, 2011년 12월 기준 농업조사에서 파악된 농가의 최신 정보를 이용하여 층화 등 표본설계를 함으로써 표본의 모집단 대표성이 향상된다. 둘째, 추출단위가 조사구에서 가구로 바뀜에 따라 층 내 동질성(층화변수는 영농형태) 확보가 용이하고, 영역추출법(area frame sampling) 대신 명부추출법(list frame sampling)이 적용되어 분산의 효율성 향상을 기대할 수 있다.

본 논문에서는 새롭게 적용된 농가경제조사의 통합이중추출 방법을 소개하고자 한다. Ⅱ장에서는 기존 표본설계 및 오차분석 결과를 제시하고, Ⅲ장에서는 농업조사 및 농가경제조사의 새로운 표본설계 방법을 소개한다. 마지막으로 Ⅳ장에서는 결론 및 제언을 논하기로 한다.

Ⅱ. 현행 표본설계 및 새로운 방향 설정

1. 2007년 표본설계

2007년 표본설계에서는 2005년 농업총조사에서 조사된 농가를 조사모집단으로, 각 도를 부차 모집단으로 설정하였다. 조사모집단을 위한 표본 추출틀은 농가규모가 10가구 미만이거나, 인구주택총조사구 55개 이상인 대규모 조사구를 제외하여 37,717개 조사구(농가 1,261,469가구)로 구성하였다. 조사모집단 농가현황은 <표 1>과 같다.

〈표 1〉 2005년 모집단 농가현황 (단위: 가구)

지역	논벼	과수	특작	채소	화훼	전작	축산	기타	전체
서울	1,055	253	24	1,144	662	287	59	12	3,496
부산	3,000	614	50	4,003	453	203	383	6	8,712
대구	6,719	3,439	199	3,752	141	1,069	944	29	16,292
인천	9,236	570	102	2,522	80	1,094	668	297	14,569
광주	8,356	1,151	136	2,822	154	887	365	179	14,050
대전	3,942	1,357	270	1,326	105	1,377	270	10	8,657
울산	6,352	1,739	41	1,874	42	515	1,413	4	11,980
경기	77,927	8,131	2,271	23,996	3,380	13,593	12,041	164	141,503
강원	29,027	1,844	2,282	14,749	270	22,404	5,096	118	75,790
충북	40,130	12,608	3,862	12,578	322	11,021	4,956	246	85,723
충남	103,610	8,123	5,001	24,513	664	10,891	10,197	198	163,197
전북	80,096	5,923	3,012	15,943	828	8,847	6,829	271	121,749
전남	115,660	11,871	2,500	36,576	892	20,015	10,195	1,055	198,764
경북	88,166	50,006	5,575	39,622	303	17,595	14,332	529	216,128
경남	75,014	17,067	2,151	35,874	1,558	10,400	13,685	331	156,080
제주	9	20,540	407	8,717	342	5,315	850	38	36,218
전국	648,299 (50.9%)	145,236 (11.4%)	27,883 (2.2%)	230,011 (18.1%)	10,196 (0.8%)	125,513 (9.9%)	82,283 (6.5%)	3,487 (0.3%)	1,261,469 (100%)

농가경제조사의 표본추출법은 층화2단집락추출로 시도별 영농형태별로 층화한 후 표본 수(PSU)와 표본농가(SSU)를 추출한다. 영농형태와 전겸업을 고려하여 〈표 2〉와 같은 기준으로 조사구를 층화한다.

〈표 2〉 2007년 표본설계 조사구 층화기준

구분	층 화 기 준
논벼1	논벼농가가 70% 이상인 조사구 (화훼 20% 미만 & 특작 20%미만 & 2종겸업 50% 이하)
논벼2	논벼농가가 50~70% 이상인 조사구 (화훼 20% 미만 & 특작 20%미만 & 2종겸업 50% 이하)
과수	과수농가가 최대(20% 이상)인 조사구 (화훼 20% 미만 & 특작 20%미만 & 2종겸업 50% 이하)
채소	채소농가가 최대(20% 이상)인 조사구 (화훼 20% 미만 & 특작 20%미만 & 2종겸업 50% 이하)
특작	특작농가가 20% 이상인 조사구 (특작농가 > 화훼농가)
화훼	화훼농가가 20% 이상인 조사구 (화훼농가 >특작농가)
전작	전작 농가가 최대(20% 이상)인 조사구 (화훼 20% 미만 & 특작 20%미만 & 2종겸업 50% 이하)
축산	축산 농가가 최대(20% 이상)인 조사구 (화훼 20% 미만 & 특작 20%미만 & 2종겸업 50% 이하)
기타	위의 영농특성 층이나 2종겸업 층에 속하지 않는 조사구
2종겸업	2종 겸업농가가 50% 초과인 조사구 (화훼 20% 미만 & 특작 20% 미만)

지역 및 영농형태의 농가 수와 주요항목(농가소득, 농업소득, 자산, 부채)에 대한 상대표준오차 등을 고려하여 표본 수를 산출하고, 각 층별로 표본을 배분하였다. 산출된 표본 수 및 표본농가 수는 560개 조사구, 2,800농가이고, 표본배분 결과는 〈표 3〉과 같다.

〈표 3〉 2007년 지역별 층별 표본 조사구 배정결과 (단위: 조사구)

지역	논벼1	논벼2	과수	특작	채소	화훼	전작	축산	기타	2종	전체
경기	10	7	5	2	11	11	3	10	3	8	70
강원	6	7	2	2	12	2	15	3	2	4	55
충북	5	6	10	7	9	2	3	3	2	6	53
충남	13	9	8	9	10	3	2	7	2	5	68
전북	16	7	5	9	12	3	2	5	1	3	63
전남	18	9	7	2	12	3	10	6	1	7	75
경북	12	9	14	6	15	2	2	8	1	6	75
경남	8	7	9	2	12	10	2	12	2	6	70
제주	0	0	11	2	5	2	2	2	0	7	31
전국	88	61	71	41	98	38	41	56	14	52	560

〈표 4〉 영농형태별 농가 수 분포 (단위: 가구)

영농형태	논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	합 계
농가 수 (%)	500 (18.9)	286 (10.8)	616 (23.3)	90 (3.4)	50 (1.9)	81 (3.1)	227 (8.6)	38 (1.4)	761 (28.7)	2,,649 (100.0)

층 내의 모든 조사구를 분류기준(전업+1종겸업 농가비율, 평균 경지면적)에 의해 정렬한 후, 배정된 표본 수 수만큼을 조사구 내의 농가 수에 비례하여 확률비례계통추출법을 적용하였다. 이렇게 추출된 조사구의 전 농가를 재방문하여 농가명부를 작성한 후, 영농형태, 농업수입, 농가 수입 기준으로 농가를 정렬한 후 5가구를 계통추출하였다.

2. 표본자료 분석

2010년에 조사된 농가경제조사의 결과를 이용하여 기존 표본의 현황을 살펴보았다. 2,800 농가 중 유고사유(전업, 전출, 단독, 장기출타, 조사불능, 조사불응)가 발생한 농가는 집계대상에서 제외하였다. 영농형태별 표본농가의 분포를 살펴보면, 화훼와 기타 농가가 전체에서 차지하는 비중이 각각 1.9%, 1.4%로 표본농가 수가 적어 상대표준오차가 상당히 클 것으로 예상된다. 또한 특작, 전작, 축산 영농형태에 해당하는 농가 비중도 비교적 낮아서 영농형태별 주요항목에 대한 추정치의 상대표준오차에 있어서 문제점이 있을 것으로 판단된다.

〈표 5〉는 영농형태 조사구 특성과 동일한 성격을 갖는 표본농가의 비율을 나타낸다. 해당 층의 성격과 표본농가의 영농형태가 상당 부분 일치하는 영농형태 층은 2종겸업(63.4%), 논벼(51.6%), 채소(46.9%), 과수(42.27%)이다. 이 중 동일농가 비율이 낮은 영농형태는 특작, 전작, 축산, 기타 등이다. 즉, 2종겸업, 논벼, 과수, 채소를 제외한 나머지 영농형태에서는 층 내 동질성을 유지하지 못해서 층화의 효율성을 충족시키지 못하는 것으로 나타났다. 새로운 표본설계에서는 이러한 점들을 층화기준 설정과정에서 고려해야 한다.

〈표 5〉 **설계층 내 동일 특성농가 비율** (단위: %)

조사구충	논벼	과수	채소	특작	화훼	전작	축산	기타	2종겸업	합계
동일특성 농가비율(%)	51.6	42.3	46.9	18.2	32.1	14.4	28.6	8.7	63.4	42.7

〈표 6〉 **영농형태별 상대표준오차**(2010년) (단위: 가구, %)

영농형태	표본농가 수	농가소득(%)	농업소득(%)	농가자산(%)	농가부채(%)
논벼	500	4.3	8.5	4.7	11.3
과수	286	8.6	12.0	8.3	18.4
채소	616	4.0	7.0	5.5	11.0
특용	90	17.6	24.4	11.1	45.8
화훼	50	23.9	39.8	16.8	25.4
전작	81	14.2	32.4	11.9	29.1
축산	227	15.8	24.2	10.0	18.3
기타	38	17.3	27.7	21.9	34.6
2종겸업	761	2.9	17.8	6.1	11.9
전체	2,649	2.71	7.38	2.96	6.96

〈표 6〉은 주요항목에 대한 영농형태별 상대표준오차를 구한 결과이다. 농가 수가 많은 논벼, 과수, 채소 등에 대해서는 통계적 정도가 높은 편이지만, 화훼, 특작, 전작, 기타에 대한 상대표준오차는 다소 높게 나타났다. 새로운 표본설계 시에는 표본농가 수가 적은 영농형태에 대해서도 정확한 통계작성이 가능하도록 층별 표본배정을 고려해야 한다.

3. 새로운 표본개편 방향

과거에는 군부에 거주하는 농가의 비율이 높았으나, 2010 농업총조사 결과 군부와 시부의 농가 수가 역전되어 있는 것으로 나타났다. 즉, 전체적인 농가 수는 감소하는 반면 농가의 분포가 점차적으로 시부로 이동함을 알 수 있다. 시부에 거주하는 농가의 특성은 부락을 이루지 않고 개별적으로 흩어져 거주하기 때문에 조사구 내의 농가 수

가 점차 감소하고 있다. 따라서 5년 간 조사구 내 5농가를 유지하기 위해서 10농가 미만인 조사구를 표본 추출틀에서 제외할 때 제외비율이 점차 늘어날 수 있음을 시사한다. 또한 2005년 총조사에는 농가 수 위주의 농업특성 조사구가 별도로 설정되었으나, 2010년 농업총조사는 농가 위주의 조사구 설정이 별도로 이루어지지 않았다. 따라서 새로운 표본개편에서는 조사구를 추출하고 조사구 내 동일 수의 농가들을 추출하는 2단집락추출법을 채택하지 않고, 새로운 추출방법을 모색하기로 하였다.

현재까지의 조사구 추출 방식의 또 다른 단점으로 층화의 효율성 저하를 들 수 있다. 조사구 내 농가 중 비율이 가장 높은 영농형태를 해당 조사구의 특성으로 분류하는데, 논벼와 2종겸업을 제외한 나머지 영농형태 층에서는 층 내 동질성을 확보하기 어렵다.

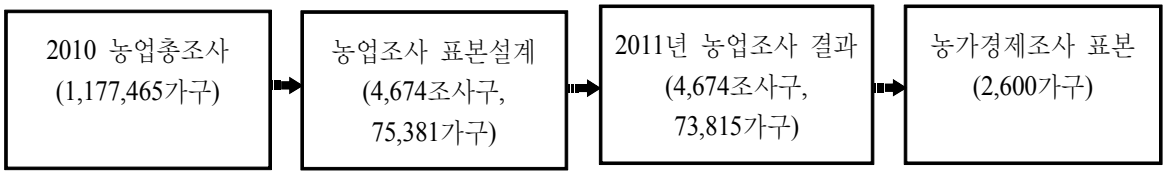
또한 층화를 위한 영농형태의 정보가 총조사 조사시점(2010년 12월 기준)으로 작성되어, 표본설계 시점에서의 농가의 이동과 영농형태의 변화 등을 적절히 반영하지 못하고 있다. 따라서 2011년 12월 기준으로 조사된 농업조사를 농가경제조사의 표본 추출틀(주표본, master sampling, 1st sample)로 활용하여 표본가구를 2차 추출하는 이중추출법(double sampling)을 적용하는 방안을 채택하였다. 이때 2차 추출에는 조사구와 같은 영역을 샘플링(Area frame sampling)하는 기법에서 가구를 직접 추출하는 리스트 샘플링(List frame sampling) 으로 변경하여 적용토록 한다.

가구를 직접 추출하는 리스트 추출법을 적용하면 표본지점이 조사구 수에서 가구수로 확대되어 같은 표본 수인 경우, 표본오차가 향상된다. 또한 층 내 동질성이 높아져 층화의 효율성 증대도 기대할 수 있다. 반면, 표본가구 간 이동거리로 인한 업무량이 증가하는 단점이 있다. 하지만, 2차 추출을 위한 표본 추출틀이 1차 추출인 농업조사의 표본 수이므로 이동범위를 다소 제한시킬 수 있으며, 주표본의 정보가 매년 업데이트되어 모집단 변동을 표본에 지속적으로 반영하는 것이 가능하다.

〈표 7〉 시 · 군별 농가 비율

(단위: 천 가구, %)

	2000		2005		2010	
시부	614	44.4	627	49.2	616	52.3
군부	770	55.6	646	50.8	561	47.7
전국	1,384	100.0	1,273	100.0	1,177	100.0



〈그림1〉 새로운 표본설계 전개도

Ⅲ. 새로운 표본설계

1. 1단계 추출: 농업조사 표본설계

농업조사는 12월1일 기준 시·군별 농가 수 및 농가인구 수를 추정하는 통계조사이며, 2010년 농업총조사 결과를 조사모집단으로 하여 표본설계하였다. 표본추출법은 층화단순집락추출이며, 표본설계를 위한 관심항목은 조사구 내의 농가 수(y)이고, 관심변수인 조사구 내 농가 수를 이용하여 시·군별로 3~4개의 층을 형성하였다.

이때 추정하고자 하는 농가 수의 모분산을 줄이기 위하여 총조사 조사구를 재설정하여 표본 추출틀을 구축하였다. 총조사 당시의 조사구는 농가 수에 대하여 최소값 1, 최대값 353, 모변동계수 132.7%를 갖는다. 이를 농가 수가 50 이상인 조사구는 분할하고 5농가 이하인 조사구는 인근 조사구에 통합하여 최소 1, 최대 49, 모변동계수 74.4%가 되도록 표본 추출틀을 재구축하였다. 모집단 조사구 수(N)는 당초 124,741개에서 분할과 통합을 거쳐 72,892개로 변경되었다.

층화변수는 조사구 내 농가 수(y)이며, 농가 수 추정을 위한 목표 상대표준오차는 3.5%, 4%, 4.5%, 층의 개수는 3~5개, 층 경계점으로는 제곱근누적도수균등화, W_hR_h 균등화, $W_h\sigma_h$ 균등화, $W_h\sigma_h$ 최소화 방법을 검토하였다(Dalenius & Hedges

〈표 8〉 표본 추출틀 재구축 전·후의 조사구 및 농가 수 분포

	농가 수	조사구 수	조사구 내 농가 수					
			평균	변동계수	왜도	최소	중앙값	최대
재구축 전	1,177,465	124,741	9.4	132.7	2.2	1	3	353
재구축 후		72,892	16.2	74.4	0.8	1	11	49

1957; Ekman 1959; Dalenius & Gurney 1951). 〈표 9〉는 목표오차별, 층 개수별, 층 경계점 등 층화방법별 표본규모이다. 층 수가 많아지면 층화효과를 증대시킬 수 있지만, 너무 세분화되면 표본설계 이후의 층 특성의 변동 등이 초래되어 효율성이 저하될 수 있다. 또한 층 개수별 모집단 분산 대비 층화분산의 감소폭이 3개 층까지 증가하지만 4개 층부터는 감소폭이 둔화되어, 분산 감소폭과 층별 모집단 수 등을 고려하여 시·군별로 3~4개 층으로 층화하였다.

시군별로 층 경계점을 검토한 결과, 대부분 $W_h\sigma_h$ 균등화 방법 또는 $W_h\sigma_h$ 최소화 방법으로 층 경계점을 정한 경우 최소 표본 수를 갖는 것을 알 수 있다. 이 두 가지 방법 중에서 층별 모집단 수 최소 10개, (네이만 배분 시) 층별 가중치 차이 최소화, 층

〈표 9〉 층화방법별 표본규모

(단위: 조사구)

목표 상대표준오차	층 경계점	층의 개수		
		3	4	5
4.5%	D-H	3,070	1,911	1,783
	WR 균등	3,038	1,883	1,783
	$W\sigma$ 최소	2,883	1,822	1,764
	$W\sigma$ 균등	2,987	1,850	1,743
4.0%	D-H	3,792	2,330	1,910
	WR 균등	3,763	2,280	1,893
	$W\sigma$ 최소	3,590	2,189	1,867
	$W\sigma$ 균등	3,710	2,227	1,843
3.5%	D-H	4,757	2,906	2,161
	WR 균등	4,733	2,900	2,141
	$W\sigma$ 최소	4,667	2,832	2,073
	$W\sigma$ 균등	4,541	2,721	2,080

별 표본 수 최소 3개 등의 조건을 만족하는 $W_h\sigma_h$ 균등화 방법으로 층 경계점을 결정하였다. 하지만 층 경계점을 D-H, W_hR_h 균등화, $W_h\sigma_h$ 최소화를 적용한 시·군도 일부 존재한다.

이와 같은 방법을 적용하여 층화한 후, 시·군별 농가 수에 대한 목표 상대표준오차를 3~5%(일부 시·군 6~7%)로 하고 네이만배분을 한 최종 표본규모는 3,371개 조사구이다. 시·군별 층별 모집단 조사구를 정해진 분류기준에 의해 오름차순으로 정렬한 후 배정된 표본 수만큼 계통 추출하여 표본 수를 추출하였다. 분류기준은 (1차) 영농형태→(2차) 전겸업→(3차) 경지면적 또는 축산형태이며, 추출된 조사구 내의 농가 수는 2010년 총조사를 기준으로 75,381농가이다.

2. 2단계 추출: 농가경제조사 표본설계

1) 조사모집단 분석

새로운 표본설계를 위한 목표모집단은 농가 정의에 부합되는 대한민국 내 전체 가구를 대상으로 한다. 조사모집단은 2011년 농업조사에서 농가로 파악된 가구이며, 표본 추출틀로는 2011년 농업조사에서 파악된 73,815농가를 사용한다.

조사모집단의 분포는 <표 10>과 같으며, 2005년 1,272,908가구에 비해 8.6% 정도 감소된 1,163,209가구이다. 2010년 1,177,465가구에 비해서는 약 1.2% 감소하였다.

농가경제조사의 조사대상은 2인 이상 농가로 제한하고 있다. 그러나 농가의 고령화 및 1인 농가의 증가 추세를 감안할 때 1인 농가를 조사대상에 포함하는 것이 필요하다. 본 조사가 시작되어 정착하던 '70년대에는 1인 농가의 비율이 2~3%로 1인 농가를 제외하여도 모집단 미포함율이 낮았지만, 1인 농가의 비율이 점차 증가하여 2010년에는 15.6%에 달하였다. <표 11>과 <표 12>를 보면 1인 농가와 2인 농가의 경지면적, 판매액, 전겸업 및 영농형태 분포가 서로 다르게 나타나, 1인 농가를 제외하는 것은 우리나라 전체 농가를 대상으로 하는 목표모집단과 표본 간의 차이를 발생시키는 것이다. 따라서, 이번 표본개편에서는 1인 농가를 포함하여 설계하였다. 1인 농가를 포함하여 공표를 하게 되면 농가소득이 낮아지는 등 시계열 단층이 예상되므로 1인 농가를 표본에 포함하되 100여 농가만 배정하고, 1인 농가의 공표방식에 대해서는 시계열 자료를 충분히 검토하는 과정을 거친 후 결정하기로 한다.

〈표 10〉 2011년 조사모집단 농가현황

(단위: 가구)

	논벼	식량 작물	채소 산나물	특약용 작물	과수	화훼	기타 작물	축산	전체
서울	792	561	1,493	68	319	220	68	20	3,541
부산	2,330	453	3,229	73	744	827	141	57	7,854
대구	5,966	1,080	3,792	436	4,623	378	213	663	17,152
인천	7,551	1,006	3,264	221	962	107	246	420	13,777
광주	6,140	536	2,495	199	1,990	490	484	146	12,481
대전	3,645	843	1,638	654	1,867	166	236	81	9,130
울산	6,138	297	2,177	179	1,461	72	51	1,398	11,772
경기	68,179	11,422	27,624	4,731	10,613	3,957	2,129	8,893	137,548
강원	22,862	16,362	19,501	5,085	2,525	440	271	4,641	71,687
충북	31,236	9,970	13,959	5,658	13,158	772	791	4,641	80,185
충남	90,405	9,199	23,681	5,087	9,081	1,897	1,916	8,941	150,207
전북	55,280	6,213	21,765	3,543	8,929	2,699	635	5,669	104,734
경북	68,995	11,901	39,015	6,019	55,919	715	1,941	12,712	197,218
경남	59,647	6,474	36,143	1,495	23,806	2,919	913	8,942	140,341
제주	-	4,760	10,073	526	21,771	173	346	849	38,497
전남	79,532	9,704	42,879	3,462	17,394	3,265	2,832	8,018	167,086
전국	508,700	90,780	252,728	37,435	175,162	19,099	13,215	66,091	1,163,209
	(43.7%)	(7.8%)	(21.7%)	(3.2%)	(15.1%)	(1.6%)	(1.1%)	(5.7%)	

〈표 11〉 가구원 특성별 평균 경지면적 및 판매액

(단위: m^2, 천원)

	경지면적			판매액		
	1인 농가	2인 농가	전체	1인 농가	2인 농가	전체
2005년 농업총조사	6,098	11,557	11,846	518	1,132	1,297
	(51.5%)	(97.6%)	(100.0%)	(39.9%)	(87.3%)	(100.0%)
2011년 농업조사	6,472	12,317	12,332	665	1,567	1,631
	(52.5%)	(99.9%)	(100.0%)	(40.8%)	(96.1%)	(100.0%)

〈표 12〉가구원 특성별 농업경영형태 (단위: 가구, %)

	1인 농가		2인 농가		전체	
전업	138,936	(79.28%)	383,505	(67.68%)	630,118	(54.17%)
1종겸업	17,289	(9.87%)	74,336	(13.12%)	181,476	(15.60%)
2종겸업	19,013	(10.85%)	108,831	(19.21%)	351,615	(30.23%)
논벼	79,410	(45.32%)	253,310	(44.70%)	508,700	(43.73%)
식량	19,414	(11.08%)	41,696	(7.36%)	90,780	(7.80%)
채소	39,281	(22.42%)	124,592	(21.99%)	252,728	(21.73%)
특약용	7,076	(4.04%)	17,460	(3.08%)	37,435	(3.22%)
과수	21,473	(12.25%)	84,564	(14.92%)	175,162	(15.06%)
화초관상	1,749	(1.00%)	7,106	(1.25%)	19,099	(1.64%)
기타	1,472	(0.84%)	5,987	(1.06%)	13,215	(1.14%)
축산	5,363	(3.06%)	31,956	(5.64%)	66,091	(5.68%)
전체	175,238	(100.0%)	566,672	(100.0%)	1,163,209	(100.0%)

2) 층화 및 표본규모

새로운 표본설계에서는 기존 표본설계와 유사하게 영농형태 및 전겸업 특성을 이용하여 층화하였다. 2011년 농업조사에서 파악된 농가를 직접 표본으로 추출하므로, 농가의 영농형태 및 전겸업별 특성에 따라 논벼, 식량, 채소, 특약용, 과수, 화훼, 기타, 축산, 2종겸업 등의 9개로 층화한다.

적정 표본규모를 산출하기 위해서는 원칙적으로 이중추출법 분산식으로부터 공식을 유도해야 한다. 하지만, 1차 표본에는 농가소득, 농업소득, 자산 및 부채 등에 대한 항목을 포함하지 않아 1차 표본단위에 대한 분산 추정이 어렵고, 1차 표본추출법이 단순추출이 아닌 층화집락추출로 이중추출 분산식 형태가 매우 복잡하며 이에 대한 선행 연구가 없는 이유로 부득이하게 기존 2단집락추출에 의한 분산식을 이용하여 표본규모를 유도하였다(Cochran 1997, 박홍래 2005).

조사구 내 최적 표본농가 수 : $$m_{opt} = \sqrt{\frac{s_w^2}{s_b^2} \cdot \frac{c_1}{c_2}}$$

최적 조사구 수 : $n_{opt} = \frac{s_b^2 + (s_w^2 / m_{opt})}{V_0} = \frac{1}{V_0} \cdot (\sqrt{s_b^2 c_1} + \sqrt{s_w^2 c_2}) \cdot \sqrt{\frac{s_b^2}{c_1}}$

여기서, s_b^2은 조사구 간 분산, s_w^2은 조사구 내 분산, c_1은 현지 확인과 조사구간 이동비용 등 1차 추출단위(조사구)당 조사에 필요한 비용, c_2은 답례품, 면접비용, 조사표류 등 2차 추출단위(가구)당 조사에 필요한 비용으로 $c_1 : c_2 = 2:1$로 가정한다. s_b^2, s_w^2은 2005~2008년 농가경제조사 원자료를 이용하여 추정하였다. 표본규모를 위한 〈표 13〉은 목표 상대표준오차를 농가소득 2.2%, 농업소득 5.3%로 정하여 조사구당 농가 수별 표본규모를 산출한 결과이다. 조사구당 1농가를 조사하는 경우 최적 표본규모는 약 2,100여 농가임을 알 수 있다.

〈표 13〉 조사구당 농가 수에 따른 표본규모 변화

(단위: 조사구, 가구)

조사구당 농가 수	농가소득		농업소득	
	조사구 수	표본규모	조사구 수	표본규모
1	2,083	2,083	2,102	2,102
2	1,111	2,222	1,146	2,292
3	787	2,361	827	2,481
4	625	2,500	668	2,672
5	528	2,640	573	2,865

〈표 14〉 2인 이상 농가 표본규모별 예상 상대표준오차

(단위: %)

항목	기존(n = 2,800)	새로운 표본규모에 따른 예상 상대표준오차			
	상대표준오차	2,100	2,300	2,500	2,800
농가소득	2.5	2.5	2.4	2.3	2.2
농업소득	6.5	6.5	6.2	6.0	5.6
농가자산	2.9	2.9	2.8	2.7	2.5
농가부채	6.7	6.7	6.4	6.1	5.8

농가경제조사는 매월 표본농가를 방문하여 조사하는 과정에서 불응, 전업, 전출 등으로 표본이 중도 탈락한다. 또한, 새로운 표본추출방식의 도입과 1인 농가의 포함으로 시계열 불안정이 우려되어 최종 표본규모를 2,600으로 결정하였다. 이중 100 농가는 1인 농가를 추출하고 나머지는 2인 이상 농가를 추출한다. 따라서, 2인 이상 농가의 표본규모는 2,500이며, 새로운 표본규모별 예상되는 상대표준오차는 〈표 14〉와 같다.

3) 표본배분 및 추출

농가경제조사는 9개 도별 통계를 생산하므로, 1차로 지역배분을 하고, 지역별 배분된 농가를 층별로 배분하게 된다. 1인 농가는 배정된 표본규모가 100가구로 크지 않기 때문에 도별 배분 후 농가를 추출하며, 지역별 1인 농가 수에 비례하여 배분하였다.

2인 이상 농가의 배분을 위해 9개 도 지역의 통계생산을 감안하여 2,500개의 표본을 지역별로 먼저 배분한다. 배분을 위해, 비례배분, 제곱근배분, 농업소득의 목표오차에 따른 표본규모 등을 고려한 후 지방청별 업무량 및 예산 등을 감안하여 배분하였다. 이때 지역별 표본 수는 최소 240개(제주도 145개)이상으로 설정하였다.

〈표 15〉 지역별 1인 농가 수 및 표본 수

(단위: 가구, %)

지역	1인 농가 수	비율(%)	표본 수
경기	14,248	8.1	9
강원	9,694	5.5	7
충북	10,462	6.0	6
충남	21,139	12.1	10
전북	16,483	9.4	9
전남	36,404	20.8	21
경북	33,758	19.3	19
경남	28,740	16.4	16
제주	4,311	2.5	3
전국	175,238	100.0	100

〈표 16〉 지역별 표본배분 결과

(단위: 가구, %)

지역	모집단 분포	상대표준오차(cv)				표본규모				최종 배분
		농가 소득	농업 소득	자산	부채	비례 배정	제곱근 배정	목표 오차	오차 배정	
경기	140,618	7.2	23.3	9.5	24.9	356	322	7.5	307	311
강원	61,993	5.3	12.4	8.7	16.6	157	215	7	135	246
충북	69,723	7.5	19.4	6.2	17.8	176	228	7	247	247
충남	138,198	8.6	26.1	6.7	17.5	350	320	7.5	332	305
전북	88,250	6.7	16.8	7.1	18.1	223	255	7.5	257	281
전남	143,163	4.7	11.1	5.9	12.9	362	325	5	343	325
경북	180,613	6.1	12.3	7.2	13.2	457	365	5	410	331
경남	131,227	7.3	13.7	8.5	17.8	332	311	7	325	304
제주	34,187	7.5	17.5	9.4	16.9	87	159	8	144	150
전국	987,972	2.5	6.5	2.9	6.7	,500	2,500		2,500	2,500

〈표 17〉 영농형태 층별 표본배분 결과

(단위: 가구, %)

영농 형태	모집단 분포		상대표준오차(cv, %)				표본배분		최종 배분
	총조사 (′10)	농업조사 (′11)	농가 소득	농업 소득	자산	부채	비례 배정	제곱근 배정	
논벼	282,316	277,105	4.3	8.5	4.7	11.3	701	515	523
전작	42,212	38,752	14.2	32.4	11.9	29.1	98	193	183
채소	135,241	147,133	4.0	7.0	5.5	11.0	372	375	385
특용	18,707	19,347	17.6	24.4	11.1	45.8	49	136	145
과수	106,837	106,630	8.6	12.0	8.3	18.4	270	319	309
화훼	9,744	9,369	23.9	39.8	16.8	25.4	24	95	94
기타	4,236	7,565	17.3	27.7	21.9	34.6	19	85	50
축산	60,197	49,384	15.8	24.2	10.0	18.3	125	217	280
2종	334,326	332,687	2.9	17.8	6.1	11.9	842	564	531

다음은 영농형태별 통계 생산을 위해 9개 영농형태 층별로 제곱근배분을 주로 고려하여 배분한 후 상대표준오차(CV)가 비교적 안정적인 논벼, 2종겸업의 표본을 상대표준오차가 높은 전작, 특용, 축산층으로 더 배분토록 하였으며, 영농형태별 표본 수가 최소 50개 이상 배분되도록 하였다. 〈표 17〉은 영농형태 층별 배분결과이다.

지역별로 배정된 2인 이상 표본 수를 각 층별로 비례배분한 후, 〈표 17〉의 층별 배정된 표본 수 합계가 맞도록 반복하여 조정한다. 이때 강원도와 제주도는 모집단 비중이 작은 일부 영농형태 층을 합하여 통합층을 형성하였고 이에 따른 최종 표본배정 결과는 〈표 18〉과 같다.

표본가구 추출을 위해서 분류기준으로 정렬한 후 층별로 배정된 표본 수만큼 계통추출한다. 분류기준은 층별 상관계수 분석을 통해 판매액과 상관관계가 높은 순으로 〈표 19〉와 같이 정하였다.

〈표 18〉 지역별 층별 최종 배분결과

(단위: 가구)

	2인 이상 농가											1인 농가	최종 표본 수
	논벼	전작	채소	특용	과수	화훼	기타	축산	2종	통합*	합계		
경기	62	16	27	12	18	28	7	34	107		311	9	320
강원	45	45	41	27	7	-	-	29	49	3	246	7	253
충북	39	29	28	28	36	4	3	28	52		247	6	253
충남	92	18	35	24	21	11	10	37	57		305	10	315
전북	80	17	42	19	24	13	3	37	46		281	9	290
전남	86	19	65	12	28	13	16	33	53		325	21	346
경북	70	14	57	16	69	3	6	36	60		331	19	350
경남	49	13	64	6	40	19	3	42	68		304	16	320
제주	-	12	26	-	66	-	-	4	39	3	150	3	153
전국	523	183	385	144	309	91	48	280	531	6	2,500	100	2,600

* 강원: 화훼+기타, 제주: 논벼+특용+화훼+기타를 통합한 층, 해당 지역의 통합된 층은 모집단 가구 수가 적어 층별 표본배분이 어려우므로 통합층을 형성하였음.

〈표 19〉 층별 분류기준

층	분류기준
논벼	논 면적, 소사육 두수, 경영주 연령, 가구원 수
채소, 과수, 특작, 기타, 2종	경지면적, 경영주 연령, 가구원 수
화훼	경영주 연령, 가구원 수, 경지면적
축산	소, 돼지, 경영주 연령, 가구원 수
1인 농가	경지면적, 경영주 연령

4) 가중치 작성 및 추정

표본농가가 이중추출법에 의해 추출되었으므로, 각 농가의 가중치는 1차 추출의 결과인 농업조사의 시·군 층별 가중치와 2차 추출의 결과인 지역 영농형태 층별 가중치의 곱으로 계산이 된다. 이러한 설계가중치를 기반으로 무응답조정을 거치고, 농업조사 결과의 농가분포를 벤치마킹하여 사후층화 조정하는 과정을 거쳐 최종 가중치를 산출하게 된다.

(1) 설계가중치(design weight)

표본 추출확률에 기초한 가중치로서 표본농가가 자신이 속하는 층에서 표본으로 추출될 확률의 역수로 설계가중치를 계산한다.

$$w_{dhi} = \frac{N_{sh1}}{n_{sh1}} \times \frac{M_{dh}}{m_{dh}}$$

여기서, s: 시군, h_1: 농업조사 층, d: 지역, h: 영농형태 층, N: 2010년 농업총조사 조사구 수, n: 농업조사 표본 수 수, M: 2011년 농업조사 결과 농가 수, m: 표본농가 수, w_{dhi}: 설계 가중치, i: 표본 가구를 나타내는 첨자, $i = 1, \cdots, m_{dh}$이다.

(2) 무응답조정 가중치(nonresponse adjustment weight)

무응답 조정은 단위 무응답 표본농가의 가중치를 해당 농가가 속한 지역의 층 내 응답농가에게 배분하는 방법으로 실시한다(송주원·안형진 2009).

$$w^{*}_{dhi} = w_{dhi} \times \frac{m_{dh}}{m_{dh(0)}}$$

여기서, w^{*}_{dhi}: 무응답 조정 가중치, m: 배정된 표본농가 수, $m_{(0)}$: 조사 완료된 표본농가 수이다.

(3) 사후층화 조정(post-stratification adjustment weight)

매년 12월에 실시하는 농업조사 자료를 벤치마킹 모집단으로 사용하여 사후층화를 실시하는데, 사후층은 2013년 이후 실제자료를 이용한 분석을 통하여 결정하도록 하고 이를 g라 하자(박홍래 2005; 김재광 2008).

$$w^{**}_{dhi} = w^{*}_{dhi} \times \frac{X_{dg}}{\hat{X}_{dg}}$$

여기서, w^{**}_{dhi}: 사후층화 조정 가중치, X: 벤치마킹 모집단 농가 수, $\hat{X}$: 해당특성으로 추정된 농가 수이다.

(4) 평균 및 분산추정식

전국에 대한 평균은 가중평균에 의해 추정하고, 분산추정은 표본농가를 1가구씩 제외하는 잭나이프 분산추정을 적용한다. 각 추정식은 다음과 같다(Kim, J.K. 외 2006; Kim, J.K. 외 2011).

$$\text{평균 : } \bar{y} = \frac{\sum_{d=1}^{9}\sum_{h=1}^{H}\sum_{i=1}^{m_{dh}} w^{**}_{dhi} \times y_{dhi}}{\sum_{d=1}^{9}\sum_{h=1}^{H}\sum_{i=1}^{m_{dh}} w^{**}_{dhi}}$$

$$\text{분산 : } \hat{v}(\bar{y}) = \sum_{d=1}^{9}\sum_{h=1}^{H} \frac{m_{dh}-1}{m_{dh}} \sum_{r=1}^{m_{dh}} (\bar{y}_{(r)} - \bar{y})^2$$

여기서 $\bar{y}_{(r)}$은 r번째 표본농가를 제외한 평균 추정값이다.

Ⅳ. 결론 및 제언

농가경제조사는 5년 주기 총조사 결과를 표본 추출틀로 하여 층화2단집락추출법으로 새로운 표본을 선정하여 왔다. 최근에는 총조사 이후 2년이 경과한 표본개편 시점에 농가의 생멸(birth and death), 영농형태 및 경지면적의 변화 등 모집단 변화가 상당부분 진행되게 된다. 따라서 이번 표본개편에서는 농업조사 표본을 1차 추출한 후, '11년 12월에 조사된 농업조사의 약 74,000여 농가를 주표본으로 하여 농가경제조사 표본농가를 2차 추출하는 이중추출법을 적용하였다. 또한 조사구를 추출하는 집락추출법보다 농가를 직접 추출하는 리스트 추출법을 적용하여 표본의 대표성과 층화 효율성을 향상시켰다. 또한, 1차 추출인 3,371개 표본 수 내의 모든 농가를 농업조사를 통해 매년 조사하므로, 주표본에 대한 변동사항이 주기적으로 관리된다. 이러한 변동사항을 표본에 매년 반영할 수 있어 표본관리 측면의 효율성도 기대된다.

하지만 새로운 표본설계에는 2013년에 실시되는 조사자료가 수집되면 해결해야 할 몇 가지 과제가 남아 있다. 첫째, 주표본인 농업조사가 단순임의추출이 아닌 층화집락추출이므로, 주표본의 설계효과와 가중치 효과를 측정하고, 이로 인한 영향력을 충분히 점검한 후 농가경제조사의 가중치 조정, 분산추정식 등 추정방법을 결정해야 한다. 둘째, 무응답 조정을 위한 무응답층의 형성과 사후층화를 위한 벤치마킹 정보의 결정에 대해서 체계적인 연구가 필요하다. 새로운 표본설계에 적합한 추정식을 체계적으로 연구·개발하여 적용함으로써, 농가경제조사 통계결과에 대한 정확도와 신뢰도가 더욱 향상될 것으로 기대한다.

참고문헌

김규성·김진석·이선순. 2001. "이중추출에서의 모평균 추정." ≪응용통계연구≫ 14(1): 13-24.

김재광. 2008. ≪표본조사론≫. 서울: 자유아카데미.

박홍래. 2005. ≪통계조사론≫. 서울: 영지문화사.

송주원·안형진. 2009. ≪무응답 자료처리 및 분석≫. 통계교육원.

통계청. 2007. ≪농가경제조사 표본설계 보고서≫.

통계청. 2011. ≪농업조사 표본설계 보고서≫.

Cochran, W.G. 1997. *Sampling Techniques*. John Wiley & Sons.

Dalenius, T. and M. Gurney. 1951. "The Problem of Optimun Stratification-II." *Skandinavisk Aktuarietidskrift* 34: 133-148.

Dalenius, T. and J.L. Hedges. 1957. "The Choice of Stratification Points." *Skandinavisk Aktuarietidskrift* 40: 198-203

Ekman, G. 1959. "An Approximation Useful in Univariate Stratification." *Annals of Mathematical Statistics* 30: 219-229.

Hidiriglou, M.A. 2001. "Double Sampling." *Survey Methodology* 27(2): 143-154.

Kim, J.K., A. Navarro, and W.A. Fuller. 2006. "Replication Variance Estimation for Two-Phase Stratified Sampling." *Journal of the American Statistical Association* 101: 312-320.

Kim, J.K. and C.L. Yu. 2011. "Replication Variance Estimation Under Two-Phase Sampling." *Survey Methodology* 37(1): 67-74.

Kish, L. 1965. *Survey Sampling*. Wiley.

Lohr, S. 1999. *Sampling: Design and Analysis*. Duxbury Press.

<접수 2013/3/6, 1차수정 2013/5/22, 게재확정 2013/5/29>