

농업 기본통계 및 가축통계 조사 표본설계

김규성¹⁾, 전중우²⁾, 박홍래³⁾

< 요약 >

본 연구에서는 농업 기본통계 조사 표본설계와 가축통계 조사 표본설계를 제안한다. 농업 기본통계는 시·군 단위의 통계를, 그리고 가축통계는 시·도 단위의 통계를 생산할 수 있도록 설계하였다. 두 설계 모두 현행 표본 보다 표본수를 줄여 표본조사의 업무량을 줄였고, 이를 위해 시·군 단위의 총화(농업 기본통계)와 시·도 단위의 총화(가축통계)를 하였으며, 특히 가축통계에서는 유의표본과 전수조사가구를 선정하여 설계의 효과를 높였다. 그 밖에 연속조사임을 감안하여 표본관리 및 표본대체 방법을 언급하였다.

1. 서론

농업 기본통계 조사는 농업 경영의 기본 요소인 농가수, 농가인구수, 경지규모, 영농형태등에 관한 통계 조사이고, 가축통계 조사는 한우 및 젖소, 돼지, 닭등의 사육 마리수를 조사하는 통계조사로 기본통계 조사는 연 1회, 가축통계 조사는 연 4회 조사하는 연속 조사이다.

현행 표본설계는 1980년 농업 총조사 자료를 바탕으로 1982년에 두 조사를 위한 다목적설계로 설계되어 현재까지 시·도별 통계 생산에 이용되고 있다. 그러나 현행 표본설계 이후의 행정구역 변경과 농촌 환경 변화를 고려할 때, 현행 표본은 그 변화를 반영하기에 미흡한 점이 많고, 또한 지방 자치제의 정착화로 늘어나는 시·군 단위 통계에 대한 요구를 충족시키기에 어려운 점이 많다. 따라서 표본의 대표성을 높이고, 시·군 단위의 통계를 위하여 표본의 재설계가 절실히 필요하게 되었다.

본 연구는 조사별로 표본설계를 하고, 시·군별 통계(또는 시·도별 통계)를 작성하기 위하여 추정 대상 범위가 되는 시·군(시·도)을 하나의 부차 모집단으로 보고 시·군(시·도)별로 총화를 하여 표본을 추출하였다. 처음에는 두 조사 모두 시·군별 통계를 위하여 시·군을 부차 모집단으로 하여 설계를 하였으나 가축통계의 경우 시·군별 표본오차를 5%이내로 하려면 현재 조사인력으로는 감당하기 어려운 다수의 표본이 요구되어 시·도로 추정 범위를 조정하였다. 가축통계에서는 임의표본 이외에 대규모 가축 사육 농가를 감안하여 유의표본과 전수조사가구를 선정하여 이용하였다. 유의표본과 전수조사가구는 현행 표본설계에서도 이용하고 있는데, 현재의 농촌 실정을 감안할 때 그 기준이 일률적이고, 대체로 낮아 전수조사가구의 수가 불필요하게 많다. 본 연구에서는 현지의 실정에 맞게 전수조사가구 및 유의표본의 기준을 시·도별로 조정하여 효율은 높이고 그 수는 줄여 설계의 효과를 높였다.

1) (151-742) 서울특별시 관악구 신림동 산 56-1 서울대학교 계산통계학과
2) (151-742) 서울특별시 관악구 신림동 산 56-1 서울대학교 계산통계학과
3) (151-742) 서울특별시 관악구 신림동 산 56-1 서울대학교 계산통계학과

2. 현행 표본설계 분석

현행 표본설계는 1980년 농업 총조사 자료를 기초로 작성 되었으며 전국 15개 시.도 (서울특별시, 5개 직할시, 9개 도)의 농가를 모집단으로 한다. 단, 가축통계조사에서 조사상의 어려움으로 울릉군은 모집단에서 제외 되었다. 표본 추출단위(sampling unit)는 농업 조사구와 인구주택 조사구로, 시 지역에서는 인구주택 조사구를, 군 지역에서는 농업 조사구를 사용하였다. 추출 방법은 시 지역은 3개의 층으로 층화한 단순임의 추출을 하였고 군지역은 단순임의 추출을 하였다. 임의표본 외에 유의표본과 전수조사가가 있다. 유의표본은 조사구내 사육 마리수가 다른 조사구의 사육 마리수에 비해 월등하게 많아 임의표본과는 별도로 분류하여 조사하는 조사구이고, 전수조사가는 축종별로 일정규모 이상 사육하는 농가를 분류하여 직접 조사하는 농가로, 한우는 20두 이상, 젓소는 10두 이상, 돼지는 100두 이상, 그리고 닭은 2,000수 이상 사육하는 농가가 전수조사 가구이다. <표1>에 시.도별 표본수 및 전수조사 농가수가 있다. 표본수는 임의 표본과 유의 표본을 합한 2,004개의 농업 조사구를 인구주택 조사구로 환산한 후의 표본수이다. 표본수 환산은 시.도별로 다음에 의하여 구한다.

$$\text{환산된표본수} = \text{농업조사구표본수} \times \frac{\text{인구주택조사구총수}}{\text{농업조사구총수}}$$

< 표 1 > 시.도별 표본 조사구수 및 전수조사 가구수

지 역	조사구 총수	표본수	전수조사 농가수	지 역	조사구 총수	표본수	전수조사 농가수
서 울	864	177	39	충 북	3,680	355	2,207
부 산	786	97	136	충 남	6,499	474	5,731
대 구	1,132	134	318	전 북	5,893	449	2,654
인 천	506	112	398	전 남	7,957	554	2,740
광 주	718	108	203	경 북	9,867	870	5,256
대 전	983	79	157	경 남	8,209	781	3,817
경 기	9,521	1,049	15,650	제 주	1,805	190	700
강 원	4,145	454	1,654	전 국	62,565	5,883	41,480

경기도의 표본수와 전수조사 가구수가 많은 것은 가축사육 농가가 많기 때문이다. 젓소와 닭의 경우 타 도에 비해 경기도에서의 사육 비율이 높아서 전수조사가가 많다.

현행 표본의 표본오차를 구하면 <표2>와 같다. 농가수에 대한 표본오차를 시.도 단위로 구하면 도 지역에서는 대체로 5% 이하이다. 그러나 현행 표본을 바탕으로 시.군별 표본오차를 구해보면 210개 시.군중 5%이하는 2개에 불과하고 5 - 10 %는 82개, 10% 이상은 126개에 달하여 현행 표본으로 시.군 단위의 통계를 작성하기에는 무리가 있다. 가축통계의 축종별 표본오차를 보면 도시 지역에서는 표본오차가 대체로 크며, 한우의 표본오차가 대체로 크게 나타난다. 젓소와 닭의 표본오차는 작게 나타나는 것은 두 축종의 전수조사 가구가 많기 때문이다. 시.도별 통계를 산출하기 위해서는 시.도 별로 축종별 CV 를 고르게 할 필요가 있고 한우의 경우는 전반적으로 표본 오차를 낮춰야 한다.

< 표 2 > 시·도별,가구수와 가축의 축종별 표본오차 (%)

지역	농가수	한우	젖소	돼지	닭	지역	농가수	한우	젖소	돼지	닭
서울	22.3	77.1	0.0	44.0	12.4	충북	2.9	6.0	4.5	6.6	1.7
부산	9.3	11.6	8.4	4.0	1.7	충남	2.9	4.7	2.4	3.2	0.3
대구	9.4	19.6	1.2	8.0	0.1	전북	3.1	6.2	2.5	3.0	0.4
인천	8.2	12.4	0.6	2.9	0.1	전남	2.3	4.6	1.8	2.7	0.8
광주	9.4	31.9	10.5	5.8	2.7	경북	2.0	3.8	1.6	2.7	0.5
대전	29.6	20.8	15.2	23.7	2.1	경남	2.7	4.3	1.2	2.6	0.4
경기	2.0	3.7	0.9	1.1	0.1	제주	5.6	14.7	2.1	2.5	1.4
강원	3.3	4.7	3.2	3.2	1.1	전국	0.9	1.6	0.6	0.8	0.1

3. 농업 기본통계 조사 표본설계

3.1 모집단

농업 기본통계 조사의 모집단은 전국의 농가로 1990년 농업 총조사를 기초로 구성한다. 시·군별 통계를 생산하기 위하여 서울, 5개 직할시를 포함한 210개 시·군을 부차 모집단으로 한다. 표본 추출 단위는 인구주택 조사구로 전국에 62,565개의 조사구가 있다. 조사항목은 농가수, 농가 인구수 등의 세부 항목이며 통계 공무원이 농가를 방문하여 면접 청취 조사를 한다. 조사회수는 연 1 회이다.

3.2 층화 및 표본 배정

각 부차 모집단 별로 농가수 표본오차를 5% 이내로 하기 위해서 표본을 층화 단순임의 추출한다. 층화를 하기 위하여 시·군별 조사구 분포를 살펴보면 크게 시 지역과 군지역의 특성이 다르게 나타난다.

첫째, 시 지역의 조사구 분포는 오른쪽으로 치우쳐진 (skewness to the right) 형태가 대부분이며 치우침의 정도에서 차이가 난다. 조사구의 평균 농가수는 10 가구 안팎이며 농가가 적은 시에서는 50 % 이상의 조사구가 5 가구 이내의 농가만 포함하고 있다. 최대 농가수는 100 가구 이내이고 평균값보다는 중앙값이 작아서 그 차이가 치우침의 정도를 나타내 주고 있다.

둘째, 군 지역의 조사구 분포는 크게 종형(Bell shape), 이봉형(Bi-modal shape) 그리고 오른쪽으로 치우친 형으로 나뉜다. 다수의 군 지역이 종형의 분포를 보이는데 좌우가 대칭인 분포보다는 오른쪽의 꼬리가 약간 긴 분포가 일반적인 형태이다. 평균값과 중앙값은 거의 비슷하며 대체로 30-40여 농가가 그 평균값이고 최대 농가수는 100 가구 이내가 대부분이나 몇 개 군에서는 150여 농가를 포함하는 조사구가 있다. 이봉형의 분포는 도시 근처의 군에서 발견되는데 도시와 인접한 조사구에서 10 농가 이내를 가지는 조사구가 다수이고 나머지 지역에서는 30-40여 농가를 갖는 조사구가 많아서 생긴다. 봉우리가 생기는 크기는 군의 특성에 따라 다르다. 오른쪽으로 치우쳐진 분포는 도시에 근접한 군에서 나타나며 심하게 치우친 분포는 많지 않고, 평균값과 중앙값이 비슷하며 분포의 폭이 크지 않다.

층화는 시 지역과 군 지역의 분포가 현저하게 구별되므로 크게 두 그룹으로 나누어서 한다.

시 지역은 오른쪽으로 치우친 분포가 대부분이므로 층의 갯수를 3 - 5 개로 하여 분산의 대부분을 차지하는 오른쪽 조사구 (농가가 많은 조사구)를 세분하여 층을 만들고 조사구수는 많지만 농가수가 적어서 분산이 크지 않은 왼쪽 조사구는 크게 묶어 층을 만든다. 형태는 비슷하지만 규모나 치우침의 정도등이 다르기 때문에 시 별로 층화 기준을 다르게 한다. 군 지역에서 오른쪽으로 치우쳐진 분포는 시 지역과 유사하게 층화를 하고 중형 분포와 이봉형 분포는 분산이 큰 양 끝을 세분하고 중앙을 크게 묶어 3 - 5 개의 층으로 층화한다. 시 지역과 마찬가지로 층화 기준은 군마다 다르다.

층화에 의해 만들어진 층에 표본배정을 한다. 표본배정은 층에 포함되는 조사구수와 표준편차의 곱에 비례하는 최적 배정을 한다. 층 1 은 농가수가 적은 조사구로 구성되어 포함되는 조사구수는 많지만 농가수의 층내 분산은 크지 않다. 반면 층4 나 층5 는 농가수가 많은 조사구들의 층으로 조사구수는 적지만 층내 분산이 크다. 처음에는 층을 3 개로 하여 환산된 시.군별 표본수 보다 적은 갯수의 표본으로 목표오차(5%)에 도달하도록 층화를 되풀이 해보고, 목표오차에 도달하지 못하면 층의 수를 늘려 층화를 다시 하였다. 210개 시.군이 층 수 3-5 개로 층화 되었다.

본 설계에서 사용한 표본수는 4,101개로 총 62,565개의 전국 인구주택 조사구의 6.5%에 해당하며 현행 표본수 5,883 개의 69.7%에 해당한다. 시.도별 표본수가 현행 표본설계의 표본수와 비교되어 <표3>에 나타나 있다.

3.3 표본 오차 (CV)

S 시(또는 군)의 농가수 총계는 다음과 같이 추정한다.

$$\hat{Y}_s = \sum_{h=1}^{L_s} N_{sh} \left(\frac{1}{n_{sh}} \sum_{i=1}^{n_{sh}} y_{shi} \right)$$

여기에서 L_s : S 시의 층의 갯수,

N_{sh} : S 시의 h 층의 조사구수,

n_{sh} : S 시의 h 층의 표본 조사구수,

y_{shi} : S 시의 h 층 i 번째 인구주택조사구의 농가수 조사값.

\hat{Y}_s 의 분산과 표본 오차는 다음과 같이 추정된다.

$$\widehat{Var}(\hat{Y}_s) = \sum_{h=1}^{L_s} N_{sh}^2 \left(\frac{S_{sh}^2}{n_{sh}} \left(1 - \frac{n_{sh}}{N_{sh}} \right) \right)$$

$$\widehat{CV}(\hat{Y}_s) = \frac{\sqrt{\widehat{Var}(\hat{Y}_s)}}{\hat{Y}_s}$$

도 단위의 추정은 시.군에서 구한 총계와 분산의 추정을 이용하여 구하는데, 도의 총계는 $\hat{Y} = \sum_s \hat{Y}_s$ 로 추정하고, 분산은 $\widehat{Var}(\hat{Y}) = \sum_s \widehat{Var}(\hat{Y}_s)$ 로 추정한다. 마찬가지로 전국의 추정은 각 도의 추정치를 이용하여 구한다.

목표 오차는 시.군별로 5%이내로 하되 표본수가 많아 업무량이 과중한 일부 시 지역에서는

현행 업무량 이내로 표본수를 줄이고 표본오차의 한계를 다소 완화하였다. 군 지역에서는 울릉군을 제외한 전 군 지역의 표본오차를 5% 이내로 하였다. 시·군의 표본오차를 5%로 한 결과 도의 표본오차는 제주도를 제외하고 모두 2% 미만이며 전국의 표본오차는 0.4% 가 된다.

< 표 3 > 시·도별 표본수 및 표본오차 (%)

시 도	표본수	표본오차		표본수	표본오차
서울	46	4.6	총 북	240	1.6
부산	52	4.2	총 남	414	1.2
대구	40	4.4	전 북	335	1.0
인천	40	4.7	전 남	512	1.1
광주	40	3.6	경 북	630	1.0
대전	40	4.7	경 남	598	1.0
경기	632	1.1	제주	78	2.7
강원	404	1.3	전 국	4,101	0.4

3.4 본 설계의 효과

본 연구에서 제시한 표본설계에서 다음과 같은 효과를 얻는다.

첫째, 본 설계는 시·군별 통계를 생산할 수 있도록 설계되어 시·군별 통계의 활용이 가능하게 되었다. 210개 시·군별 농가수에 대한 예상 표본오차는 군 지역은 울릉군을 제외한 전 지역이 5% 이내이고, 시 지역에서는 서울, 5개 직할시는 5% 이내이며 67개 일반 시 중에서는 20개 시 지역만이 5%를 초과하여 10% 이내이고 나머지 47개 시는 5% 이내이다.

둘째, 현행 표본설계의 표본수 보다 표본수가 많이 감소 되었다. 현행 표본은 농업 조사구 표본 2,004 개이며 인구주택조사구로 환산하면 5,883개에 해당하는데, 본 설계의 표본은 4,101개로 현 표본의 69.7%에 해당한다. 또한 현행 표본이 유의표본을 포함하고 있는데 반해 본 설계는 유의표본이 없이 모두 임의표본으로 표본의 대표성이 더 크다.

4. 가축통계 조사 표본설계

4.1 모집단

가축통계 조사의 모집단은 울릉군을 제외한 전국의 가축 사육 농가이며, 시·도별 통계를 위하여 15개 시·도를 부차 모집단으로 한다. 표본추출 단위는 인구주택 조사구이고, 조사항목은 한우, 젓소, 돼지, 닭등의 사육 가구수, 마리수 등의 세부 항목이고 농업기본통계와 마찬가지로 통계 공무원이 농가를 방문하여 면접 청취 조사를 한다. 조사 회수는 연 4 회이다.

4.2 전수조사가구 선정

4축종 모두 대규모 사육 농가가 있으면, 대규모 사육 농가는 표본조사보다 전수조사를 하여 추정치를 산정함으로써 표본오차를 낮추는데 효과적이다. 대규모 사육농가의 규모 및 분포는 축종에 따라, 시·도에 따라 다르다. 따라서 본 연구는 설계의 효과를 높이기 위하여 축종별, 시·도별로 전수가구 선정 기준을 정한다. '90년 농업 총조사의 규모별 축종별 사육 가구수 및 마

리수의 분포가 <표4>에 나타나 있다.

< 표 4 > 축종별 가축 사육 규모별 구성비 (%)

규 모 (마 리)	한 우		규 모 (마 리)	젓 소	
	가구수	마리수		가구수	마리수
1 - 9	94.1	68.5	1 - 9	28.7	9.3
10 - 14	3.4	11.4	10 - 14	18.9	12.9
15 - 19	1.3	6.4	15 - 19	19.1	18.3
20 - 29	0.6	3.9	20 - 29	21.1	28.3
30 - 39	0.3	3.0	30 - 39	7.3	13.9
40 - 49	0.1	1.8	40 - 49	2.4	6.0
50 - 99	0.2	3.1	50 - 99	2.0	7.1
100 이상	-	1.9	100 이상	0.5	4.2
	100.0	100.0		100.0	100.0

규 모 (마 리)	돼 지		규 모 (마 리)	닭	
	가구수	마리수		가구수	마리수
1- 29	78.3	8.1	1- 499	96.7	2.5
30- 49	5.1	3.8	500- 999	0.2	0.3
50- 99	6.7	8.9	1000- 1499	0.1	0.4
100- 299	5.5	19.6	1500- 1999	0.1	0.4
300- 499	2.5	17.9	2000- 2999	0.2	1.1
500- 999	1.5	18.7	3000- 4999	0.4	3.8
1000-4999	0.4	15.4	5000- 9999	1.0	17.8
5000-9999	-	3.6	일만-29999	1.1	42.4
10000이상	-	4.0	삼만-49999	0.1	12.1
	100.0	100.0	오만이상	0.1	19.2
				100.0	100.0

한우는 전체 사육가구의 94% 이상이 9마리 이하 사육가구로 총 마리수의 68.5%를 사육한다. 젓소는 전체 마리수의 82.7%가 40마리 미만 사육농가에서 사육한다. 이 경우 가구수 비율은 95.2%에 해당하여 젓소 역시 한우와 마찬가지로 대규모 사육 가구가 많지 않고 대규모 사육마리수도 많지 않다. 따라서 전수조사가구에서 관측한 마리수가 추정치 및 표본오차에 영향을 주기 위해서는 많은 전수조사가구가 필요하다. 한우와 젓소는 표본 관리 및 표본 대체 측면에서 전수 규모 기준을 정하는 것이 효과적이다.

돼지는 전체 마리수의 79.2%가 100 이상 사육하는 대규모 사육 농가에서 사육되며 이에 해당하는 사육농가는 전체의 9.9%에 불과하다. 따라서 돼지의 경우는 한우, 젓소와는 달리 전수 규모를 시도별로 적절하게 잡으면 추정치 및 표본오차에 상당한 영향을 줄 수 있다. 닭은 전체 마리수의 97.2%가 1,000마리 이상 사육하는 대규모 사육 농가에서 사육되며 가구수 비율은 3.1%에 불과하다. 따라서 전수조사가구로 대부분의 사육 마리수를 관찰할 수 있다.

전수조사가구의 선정 기준은 시도별로 전수조사가구를 달리 했을때의 예상 표본오차와 현재

각 도에서 가지고 있는 전수조사가구의 수를 고려하여 정한다. <표4>에서 나타나듯이 한우와 젓소는 표본 관리 측면에서 전수조사 가구를 선정하고, 돼지와 닭은 표본오차를 줄이는 측면에서 그 기준을 정한다. <표5>에 새로 정한 전수 조사 가구수가 나타나 있다.

본 설계에서 제시하는 전수조사가구는 7,667 가구이며 현행 41,480가구의 18.5 % 에 해당한다. 한우의 전수조사가구 647은 전체 한우 사육농가의 0.11%에 해당하며, 젓소 529 가구는 1.53%, 돼지 3,044가구는 2.18 %, 닭 3,447 가구는 2.61%에 해당하여 전체 가구에 대한 새로 정한 전수조사 가구수는 많지 않다. 그러나 전수 조사에서 조사되는 사육 마리수를 보면, 한우는 70,596마리(전체의 4.39%), 젓소는 46,404마리(9.28), 돼지는 2,004,679마리(44.64%), 닭은 53,945,219마리(69.85%) 를 조사하여 전수조사 가구의 효과는 크게 나타난다.

< 표 5 > 시.도별, 축종별 전수 조사 가구수

지 역	한우	젓소	돼지	닭	지 역	한우	젓소	돼지	닭
서울	1	2	11	4	충 북	30	12	135	161
부산	2	10	57	61	충 남	47	14	445	325
대구	30	45	83	19	전 북	78	46	187	463
인천	17	29	23	29	전 남	52	19	213	186
광주	10	9	47	25	경 북	107	34	340	622
대전	11	4	42	13	경 남	48	35	426	378
경기	137	187	536	913	제주	49	11	99	41
강원	28	72	400	207	전 국	647	529	3,044	3,447

4.3 총화 및 표본배정

전국의 가축 사육 농가에서 전수조사가구를 제외한 나머지 가축 사육 농가로 새로이 표본 추출을 위한 모집단을 구성한다. 전국 62,565개 인구주택조사구 중에서 전수 조사 농가를 제외했을 때 가축 사육 농가를 포함하는 인구주택조사구는 44,906 개 이다.

부차 모집단인 15개 시.도에서 각 부차 모집단별로 축종별 표본오차를 목표 오차인 3 % (8 개 도)와 5 % (서울, 직할시, 제주도)에 도달하기 위해 표본추출 방법은 총화 추출 방법을 이용한다. 총화는 부차 모집단 별로 하며 이때 총화변수는 조사구내 축종별 마리수가 된다. 총화를 위해서 조사구 분포를 분석해 보면 다음과 같은 특성을 발견할 수 있다.

첫째, 4 축종이 거의 독립적인 공분산 구조를 갖는다. 이는 총화변수로 4 축종을 모두 고려해야 함을 의미하며 주성분 분석등을 통한 차원 축소가 불가능함을 뜻한다.

둘째, 서울과 5개 직할시를 제외한 9개 도는 시 지역과 군 지역으로 구성되어 있으며 군 지역의 사육 마리수가 시 지역의 사육 마리수 보다 훨씬 많다. 또한 군 지역에서도 군별로 사육하는 마리수의 분포가 현저하게 다르며 타군에 비해 가축 사육을 많이 하는 군이 있다.

셋째, 조사구 분포를 보면 4축종 모두 오른쪽으로(숫자가 큰 쪽으로) 긴 꼬리를 가지며 돼지와 닭은 특히 심하다.

넷째, 한우는 전체 조사구중 80% 가량의 조사구에서 사육농가가 발견되며 젓소는 20% 그리고 돼지와 닭은 50% 가량의 조사구에서 사육농가가 나타난다.

이와 같은 특성을 바탕으로 서울, 5개 직할시 그리고 제주도를 제외한 8개 도는 시 지역과 군 지역으로 구분하고, 군 지역은 가축 사육을 특히 많이 하는 군과 그렇지 않은 군으로 나눈다. 이렇게 크게 3개 층을 만드는 1차 층화를 한다. 2차 층화는(서울, 직할시, 제주도 는 1차 층화) 다음의 순서로 한다.

< 표 6 > 경기도의 층 기준 및 표본수

층 번호	시 지역	군지역		조사구 수	표본 수	
		일반군	3131, 3135 3140, 3143 3144, 3146			
16	$x > 350, y > 350$ $z > 2000, w > 20000$	$x > 400, y > 350$ $z > 2000, w > 30000$	$x > 400, y > 450$ $z > 3000, w > 30000$	54	54	
15	w > 2000	w > 10000	w > 10000	128	28	
14		w > 5000	w > 5000	291	45	
13		w > 2000	y > 2000	122	23	
12		y > 100	y > 200	233	42	
11		y > 60	y > 80	450	70	
10		y > 25	y > 40	840	127	
9		y > 80	x > 90	x > 120	211	45
8		y > 30	x > 50	x > 70	353	51
7		x > 100	z > 400	x > 40	183	33
6		x > 50	z > 100	z > 400	354	44
5		z > 500	z > 50	z > 150	242	27
4		z > 70	z > 5	z > 50	516	46
3		z > 20	x > 300	z > 20	146	22
2		x > 10	x > 10	x > 10	547	56
1		나머지	나머지	나머지	1523	117
합계				6193	830	

층종별로 일정크기 이상의 조사구를 분류하여 하나의 층을 만든다. 이 층은 유의표본 층이며 포함되는 조사구는 모두 유의표본 조사구가 된다. 유의표본은 긴 꼬리를 갖는 분포에서 층화 분산을 낮추는데 효과적으로 이용될 수 있다. 층화 변수가 4개 이므로 4변수를 동시에 고려하면 한 변수로 2개 층만 만들어도 전체 층은 16개가 되어 층 수가 많아진다. 따라서 층의 수를 많지 않게 하기 위해서는 4축종을 동시에 고려하지 말고 층 하나에 변수 하나만 고려하여 층종별로 전체 분산을 제어할 수 있는 조사구들을 묶어 하나의 층으로 만든다. 분산의 대부분은 큰 수치를 갖는 조사구에서 발생하므로 이러한 조사구를 묶어 상위층을 구성한다.

닭의 분산이 가장 크므로 닭의 상위층을 먼저 만들어 닭의 분산을 제어하고 젓소, 돼지, 그리고 한우의 순으로 상위층을 만들어 층화한다. 상위층을 만드는 순서나 상위층의 크기 등은 시, 도의 특성에 맞게 한다.

층화에 의해 만들어진 상위층에 표본배정을 한다. 표본배정은 상위층에 포함되는 조사구수에 비례하여 배정하되 층종별로 대단히 이질적이어서 분산이 큰 층에는 표본수를 늘린다. 또한 표본수를 시,군별로 집계하여 현행 표본설계의 표본수 보다 그 수가 많으면 다른 군으로 표본을 옮겨 표본수를 조정한다. <표6>에 경기도의 층화 및 표본 배정의 결과가 있다.

<표6>에서 군지역의 3131,.....,3146은 군을 나타내는 코드 번호이고, 총16의 조사구는 모두 유의표본이 된다. x 는 한우의 마리수, y 는 젖소의 마리수, z 는 돼지의 마리수, w는 닭의 마리수를 나타내고, 유의표본은 4가지 조건 중 한조건 이상을 만족하면 되며, 나머지 15개 층은 층번호가 큰 순서로 우선순위를 갖는다. 즉, 시지역에서 유의표본조사구를 제외한 나머지 조사구 중에서 w>2000 인 조사구를 층10 으로 하고, 층10 을 제외한 나머지 조사구 중에서 y > 80 인 조사구를 층 9 로 한다. 이같은 방법으로 경기도의 경우 1차 층이 3개, 1차 층에서 2차층은 각각 11개, 16개, 16개를 구성하며 표본수는 830개이다. 시.도별 전수조사 가구수 및 표본수가 <표7>에 있다.

< 표 7 > 전수조사 가구수 및 표본수

지 역	전수조사가구	표본수	지 역	전수조사가구	표본수
서 울	18	40	충 북	338	420
부 산	130	53	충 남	831	593
대 구	177	64	전 북	774	547
인 천	98	83	전 남	470	684
광 주	91	81	경 북	1,103	872
대 전	70	82	경 남	887	682
경 기	1,773	830	제 주	200	94
강 원	707	433	전 국	7,667	5,558

4.4 표본 오차

S 시(또는 도)의 가축 총계의 추정은 유의표본과 전수조사가구를 고려하여 얻는다. 유의표본과 전수조사가구는 오차가 없으므로 분산에는 영향이 없다.

i) 총계 : $\hat{Y}_s = \sum_{h=1}^{L_s} N_{sh} \left(\frac{1}{n_{sh}} \sum_{i=1}^{n_{sh}} y_{shi} \right)$
 + S시의 유의표본총계 + S시의 전수조사총계

ii) 분산 : $\widehat{Var}(\hat{Y}_s) = \sum_{h=1}^{L_s} N_{sh}^2 \left(\frac{S_{sh}^2}{n_{sh}} \left(1 - \frac{n_{sh}}{N_{sh}} \right) \right)$

iii) 표본오차 : $\widehat{CV}(\hat{Y}_s) = \frac{\sqrt{\widehat{Var}(\hat{Y}_s)}}{\hat{Y}_s}$

전국의 총계는 $\hat{Y} = \sum_s \hat{Y}_s$ 로 추정하고, 이때 분산은 $\widehat{Var}(\hat{Y}) = \sum_s \widehat{Var}(\hat{Y}_s)$ 로 추정한다.

본 설계에서 제시하는 축종별, 시.도별 표본오차는 <표8>과 같다. 서울과 직할시, 그리고 제주도는 모두 5% 미만이며 나머지 8개 도는 3% 이내로 표본오차를 제어하였다. 전국의 표본오차는 4축종 모두 2% 미만이다.

< 표 8 > 시·도별, 축종별 표본 오차 (CV) (단위 : %)

지 역	한우	젓소	돼지	닭	지 역	한우	젓소	돼지	닭
서 울	4.46	3.70	1.49	0.22	충 북	2.74	2.97	2.37	1.40
부 산	4.53	4.41	4.42	2.39	충 남	2.94	2.48	2.50	1.40
대 구	4.72	4.98	3.79	1.52	전 북	2.97	2.93	2.50	1.07
인 천	4.88	4.75	4.34	1.71	전 남	2.91	2.92	2.51	2.66
광 주	4.97	4.51	4.85	1.88	경 북	2.53	2.93	2.54	1.21
대 전	4.82	4.81	1.94	3.56	경 남	2.87	2.75	2.68	0.98
경 기	2.99	2.22	2.77	0.85	제 주	4.50	4.97	4.62	3.09
강 원	2.55	2.99	2.01	1.42	전 국	1.02	1.16	1.12	0.51

4.5 본 설계 효과

본 설계에서는 현행 설계에 비해 전수조사 가구수와 표본수가 적다. 전수조사가구는 한우와 젓소의 전수조사가구를 대폭 줄이고 시·도별로 시·도의 특성에 맞게 그 기준을 조정하여 전체 전수조사 가구수를 현행 설계의 1/5 이하로 대폭 줄였다. 가축 사육 농가는 대규모 기업화 추세가 있으므로 시간이 지나면 위의 전수조사 가구수 보다는 다소 증가할 것이다. 그러나 전수조사농가 선정 기준을 도의 특성에 맞게 정하여, 사육 농가가 많은 도는 그 기준을 높게, 사육 농가가 적은 도는 그 기준을 낮게 하였기 때문에 급격한 전수 조사 가구의 증가는 일어나지 않는다. 표본수도 현행 설계에 비해 적다. 전체 표본수는 현행에 비해 560개가 줄어든 5,558 개이다. 시·도별로 보면 대전과 충청도, 전라도를 제외한 10개 시·도에서 표본수가 줄었는데 경기도, 강원도, 경상남도에서는 100여 개 이상의 표본이 줄었다. 충청도, 전라도는 표본수가 다소 증가 하였는데 그 이유는 현행 표본설계가 농업기본통계를 겸한 다목적 설계이므로 농가수 분포가 비교적 고른 이들 지역에 표본이 적게 배정되어 있었기 때문이다. 표본수는 증가하였으나, 전수조사가구수가 이들 4개 도에서는 각각 2,000 여 가구 넘게 감소 되었으므로 전체적인 업무량은 증가하지 않았다.

이와같이 본설계에서는 전체적인 업무량을 현 업무량내에서 가축 사육 마리수에 대한 추계 오차를 3% 와 5%로 하여 시·도별 통계를 산출할 수 있게 하였다.

5. 표본 관리 및 표본 대체

농업 기본통계 조사는 일년에 한 번씩, 가축 통계 조사는 일년에 네번씩 매년 실시하는 연속 조사이다. 해가 거듭될수록 모집단이 변화하므로 지속적인 표본 관리 및 표본 대체가 이루어져야 표본이 효과적으로 모집단의 변화를 반영할 수 있다.

모집단의 변화는 조사구수의 변동과 농가수, 가축사육 마리수의 변동에 의하여 생긴다. 조사구수의 변동은 행정구역 변경이나, 자연 재해등으로 생기는데 이경우 변경된 조사구수를 파악하여 총계추정의 공식을 바꾸어야 한다. 농가수의 변동은 조사구내 전,출입하는 농가의 수를 관리하여 파악하고, 전수조사가구는 매 분기마다 조사 관리하며, 유의표본의 기준을 만족하는 조

사구가 생기면 유의 표본으로 하고, 기존의 유의표본 중에서 기준에 미달하면 유의표본에서 제외한다.

표본은 원칙적으로 임의로 변경하면 안된다. 본 설계에서 제시하는 표본은 일차 표본을 선정한 후 농립수산부와 협의를 거쳐 조사가 불가능한 도서 지역이나 산간 오지 그리고 나환자촌 등에 배정된 표본은 신 표본으로 대체한 표본이다. 따라서 임의로 표본을 바꾸면 표본의 대표성도 떨어질 뿐 아니라 추정치의 오차도 커진다. 그러나 부득이한 경우, 표본내의 농가수나 가축 사육 마리수의 급격한 증가나 감소, 혹은 개발등으로 설계 당시의 모집단과는 현격한 변화가 있을 때에는 전체 조사구 목록을 보고 표본과 비슷한 규모의 새로운 조사구를 신 표본으로 대체해 준다.

6. 결 론

표본설계를 평가하는 기준은 표본오차와 표본수이다. 표본오차는 추정치의 정도를 말해주고, 표본수는 표본조사를 수행하는 작업량과 직결되어 눈에 보이지 않는 비 표본오차를 간접적으로 나타내 준다. 본 연구에서 제시한 두개의 표본 설계는 표본오차를 고르게 하면서, 현재의 표본수보다 그 수를 줄여 시·군별 농가수 및 시·도별 가축 마리수를 추정할 수 있도록 하였다. 가축은 그 속성상 사육 마리수의 변동이 심하다. 따라서 연속 조사에서는 표본관리가 중요해진다. 특히 대규모 사육 농가는 꾸준한 관리를 통하여 표본오차의 급격한 증폭을 막아야 한다.

참고문헌

- [1] 박홍래(1987), “농업기본통계 및 가축통계조사 표본설계에 관한 연구”, 응용통계연구 제1권 제2호, 12-20
- [2] 박홍래(1989), 통계조사론, 영지문화사, 서울

Sample Designs of the Farm Population Survey and the Livestock Survey

Kyu Seong Kim⁴⁾, Jong-Woo Jeon⁵⁾ and Hong Nai Park⁶⁾

Abstract

The farm population survey and the livestock survey are sample surveys related to agriculture. Two new sample designs for these surveys are considered. Shi-Gun(county) estimates in the farm population survey and Shi-Do(county) estimates in the livestock survey can be obtained. Also the sample sizes are reduced. To increase the precision of the estimates stratified simple random samples are used and particularly purposive samples are introduced in livestock survey. lastly the method of management and replacement of samples are investigated for successive occasion survey.

4) Department of Computer Science and Statistics, Seoul National University, Shinrim-Dong, Kwanak-Ku, Seoul, 151-742, Korea.

5) Department of Computer Science and Statistics, Seoul National University, Shinrim-Dong, Kwanak-Ku, Seoul, 151-742, Korea.

6) Department of Computer Science and Statistics, Seoul National University, Shinrim-Dong, Kwanak-Ku, Seoul, 151-742, Korea.