

# 농업 기본통계 및 가축통계 조사 標本設計에 관한 연구

박 홍 래\*

## 요 약

본 연구는 한개의 표본에서 농업 基本統計와 家畜統計 조사를 다같이 실시하는 多目的 표본설계에 관한 연구를 목적으로 한다.

본 연구의 특징은 첫째 全國 및 道別 추정오차를 작게 하는 동시에 市郡별 추정오차를 균등하게 유지하도록 시도하였고, 둘째 任意標本과 有意標本을 결합하여 표본의 효율을 높이는 방법을 제안하였다. 유의표본은 農家 유의표본과 調查區 유의표본으로 구분하였고, 이들 유의표본 선정에 있어서 추정치의 精度를 높일 수 있는 선정 기준에 관하여 고찰하였다.

## 1. 서 론

農業 基本統計 조사는 농업 구조의 기본 사항인 농가구 수, 농가 인구, 농업 취업 인구 등에 관한 통계조사이다. 家畜統計 조사는 젖소, 한우 및 고기소, 돼지, 닭 등의 주요 가축수를 파악하기 위한 조사이다.

현행 표본은 1975년 간이 센서스 자료를 기준으로 작성되었으며 그동안 모집단 변동이 많았음을 감안할 때, 1980년 센서스 자료를 기준으로 표본을 재설계할 필요가 있다. 또한, 현재 실시하고 있는 표본은 그 규모가 너무 커서 업무량이 과다하고 비용이 과대하게 지출될 뿐만 아니라, 결과의 신뢰성에도 문제가 있다.

본 연구는 첫째 同一標本에서 농업 기본통계와 가축통계 조사를 다 같이 조사하도록 하고, 둘째 전국 및 도별 추정치의 오차를 적게 하면서 동시에 市郡별 추정치의 오차를 가능한 한 균등하게 되도록 하고, 셋째 표본의 크기를 되도록 작게 하며 신뢰성을 높일 수 있는 標本設計를 작성하는데 그 목적이 있다.

조사의 성질로 보아 기본통계와 가축통계는 별개의 표본으로 조사하는 것이 바람직하다. 그러나 여기에서는 한개의 표본으로 두가지 조사를 동시에 실시하도록 요구받고 있다. 또한, 농업 기본통계와 가축통계는 지방행정에서 이용도가 크기 때문에 市郡별 추정이 필요

\* 서울대학교 자연과학대학 계산통계학과, 서울 151.

하고, 이 추정치의 精度가 비슷해야 한다.

농림수산부에서는 추정치의 標本誤差가 市郡별로 5% 내외가 되도록 요구하였다. 그러나 이 조건은 전수조사에 가까운 대표본을 필요로 하므로, 무리한 요구가 된다. 실제로, 市郡별 추정오차를 5% 이내로 하기 위한 표본의 크기를 계산한 결과, 총 調查區數 25,969 가운데 농가수 추정에는 10,079개의 조사구가 필요하며 한우는 16,262개의 조사구, 돼지는 22,508개의 조사구, 닭은 24,215개의 조사구가 필요하게 되어 전수조사에 가까운 조사구수가 필요하게 된다. 그러므로 표본의 크기를 일정한 다른 기준에 의하여 정한 다음, 市郡별 추정오차가 비슷하게 되도록 재조정하는 방법을 채택하였다.

본 연구에서 시도한 표본설계의 특징은 다음과 같다.

첫째, 표본의 크기를 작게 하면서 정확성을 높이도록 하였다. 본 설계에서 표본의 크기는 전수조사 농가수가 4,352 가구, 표본 調查區 수가 2,118 개이다. 현재 시행중인 농림수산부의 조사에서는 전수조사 농가수는 46,762 가구, 표본 調查區 수가 4,997 개이다. 이를 대비하면 본 설계의 전수조사 농가수는 현행 농가수에 비하여  $\frac{1}{10}$  미만이며 표본 조사구 수도 40%로서 표본의 크기가 훨씬 작지만 오차는 거의 비슷하게 되었다.

둘째, 전국 및 도별 추정오차를 작게 하는 동시에 市郡별 추정오차들의 차이가 크지 않도록 하였다.

셋째, 표본오차를 감소하기 위하여 市지역에서는 층화추출을 하였다.

네째, 標本效率를 높이기 위하여 234 개의 조사구를 有意選定하였다.

다섯째, 市部에서는 경제기획원 인구주택 調查區를 추출단위로 사용하였고, 郡部에서는 調查區를 추출단위로 사용하였다.

## 2. 有意標本

유의표본은 일정한 규모 이상의 전수조사 대상 농가와 조사구 유의표본이 있다. 畜種별 전수조사 농가의 규모와 농가수는 <표 1>과 같다.

<표 1>에 의하면 본 설계의 전수조사 농가수는 4,352로서 현행 표본의 46,762에 비하면

<표 1> 全數調查農家數

本 設 計				現 行 標 本			
畜 種	規 模	農 家 數		畜 種	規 模	農 家 數	
젖 소	20마마 以上	2,143		젖 소	1마리	22,075	
고기소 및 한우	50 "	191		고기소 및 한우	10 "	4,110	
돼 지	500 "	146		돼 지	20 "	9,094	
닭	5,000 "	1,872		닭	200 "	11,483	
계		4,352				46,762	

〈표 2〉 規模別 調査區 變動係數(CV)

規 模 CV		規 模 CV		規 模 CV		規 模 CV	
農 家	0.404						
젓 소		韓 牛		돼 지		닭	
1-4	3.388	1-4	0.619	1-9	1.214	1-199	1.073
1-9	3.399	1-9	0.648	1-29	1.345	1-499	1.367
1-19	3.168	1-19	0.681	1-49	1.486	1-999	2.805
1-29	3.353	1-29	0.703	1-99	1.652	1-1,999	4.684
1-49	3.354	1-49	0.732	1-299	1.803	1-4,999	5.718
1-99	3.354	1-99	0.764	1-499	1.918	1-9,999	4.998
1-∞	5.103	1-∞	0.995	1-999	2.033	1-49,999	4.602
				1-4,999	2.525	1-∞	4.811
				1-∞	4.626		

불과 9.3% 밖에 되지 않는다. 표본 크기가 4,352인 것도 작은 수는 아니며 더 감소시키는 것이 바람직하다.

원래 전수조사는 특별히 큰 값을 유의선출하는 것이며 그 수가 많은 경우에는 좋지 않다. 왜냐하면 전수조사는 비용이 훨씬 많이 들고, 그 수가 많으면 대상 농가를 파악하기가 어렵기 때문이다.

〈표 1〉에서 농가 규모와 농가수를 결정한 기준은 다음과 같다. 우선, 전수조사의 농가 규모를 결정하기 위하여 규모별 변동계수(CV)를 산출하였다. 〈표 2〉는 축종별 규모별 調査區間 변동계수이다. 〈표 2〉의 내용을 설명하면 다음과 같다. 예를 들어 젓소의 경우에 규모 1-4의 뜻은 젓소 5마리 이상을 사육하는 농가는 전수조사를 하고, 1~4마리를 사육하는 농가에 대하여는 표본조사를 한다는 뜻이며, 이 때에 市郡內의 調査區間의 CV가 3.388임을 나타낸 것이다. 이론적으로는 규모가 커짐에 따라 CV도 커져야 하나 젓소와 닭의 경

〈표 3〉 規模別 農家數(家畜)

젓 소		한우 및 고기소		돼 지		닭	
規 模	農家數	規 模	農家數	規 模	農家數	規 模	農家數
1-4	9,322	1-4	973,089	1-9	489,898	1-199	678,001
5-9	6,639	5-9	12,357	10-29	3,807	200-499	915
10-19	3,971	10-19	3,023	30-49	3,309	500-999	1,664
20-29	1,107	20-29	573	50-99	2,333	1,000-1,999	2,967
30-49	682	30-49	323	100-299	1,191	2,000-4,999	4,065
50-99	265	50-99	149	300-499	211	5,000-9,999	1,266
100-∞	89	100-∞	42	500-999	80	10,000-49,999	594
				1,000-4,999	56	50,000 이상	12
				5,000-∞	10		

〈표 4〉 道別 有意標本調查區數

市 道 別		調 查 區 數	市 道 別		調 查 區 數
서 울		11	全 北		3
釜 山		7	全 南		13
京 畿		88	慶 北		36
江 原		16	慶 南		31
忠 北		6	濟 州		7
忠 南		16	계		234

우에 증가하는 경향이 다소 결여되어 있다. 이것은 原資料에 원인이 있다고 생각된다. 전수조사 대상이 되는 농가 규모의 결정은 이 CV의 증가율이 갑자기 커지는 점으로 하면 좋다. 따라서, 〈표 1〉과 같이 젓소는 20마리, 한우는 50마리, 돼지는 500마리, 닭은 5,000마리로 결정하였다.

〈표 1〉에 주어진 전수조사 농가수는 〈표 3〉에 주어진 축종별 규모별 농가수로부터 얻은 것이다.

다음으로, 有意標本 調查區數는 모두 234개로서, 〈표 4〉는 道別 유의표본 조사구수이다. 조사구 유의표본은 젓소, 한우, 돼지, 닭의 가축수가 많은 조사구를 유의선정하였다. 농가 유의표본이나 조사구 유의표본은 추정오차를 감소시키는 한 방법이다.

### 3. 標本의 크기

표본은 전수조사 농가수 4,352개와 표본 조사구수 2,118개로 되어 있다. 표본 조사구는 유의표본 조사구 234개, 市郡의 인구주택 조사구 392개, 郡部의 농업 조사구 1,492개로 구성된다.

〈표 5〉 目標精度(CV)와 標本의 크기

農 家 數		젓 소		한우 및 고기소		돼 지		닭	
CV	n	CV	n	CV	n	CV	n	CV	n
.004	6,829	.01	7,731	.009	3,887	.015	6,071	.04	3,527
.005	4,682	.015	6,143	.0095	3,557	.02	4,105	.045	3,071
.006	3,596	.02	4,771	.01	3,264	.025	2,898	.06	2,682
.007	2,490	.025	3,706	.0117	2,489	.027	2,489	.053	2,484
.008	2,162	.03	2,912	.015	1,588	.03	2,132	.055	2,352
.009	1,741	.033	2,493	.02	923	.035	1,625	.06	2,073
.01	1,429	.04	1,884	.025	600	.04	1,274	.065	1,836
		.045	1,552	.03	420	.045	1,024	.07	1,635
		.05	1,296	.035	310				

〈표 6〉 標本의 크기와 標本誤差

	現行標本 n=11,997	n=2,500	n=2,000	n=1,800	n=2,118
農 家 數	0.007	0.007	0.008	0.009	0.008
韓牛 및 고기소	0.014	0.012	0.013	0.014	0.014
돼 지	0.017	0.027	0.031	0.033	0.03
닭	0.014	0.052	0.062	0.067	0.012
것 소 全數調査		0.033	0.038	0.041	0.048

〈표 5〉에는 目標精度 CV와 표본의 크기  $n$ 의 값이 주어져 있다. 〈표 5〉를 이용하여 표본의 크기를 2,500, 2,000, 1,800으로 할 때의 표본오차를 산출하였고, 이 결과를 〈표 6〉에 정리하였다. 〈표 6〉에서는 현행 표본과 대비하여 효율을 비교하였다. 현행 표본은 전수조사 농가의 규모가 작고 표본 조사구수가 4,997개나 되는 대표본이다. 그러나, 본 설계에서는 표본의 크기를 2,000개로 축소하더라도 농가수나 한우의 추정은 현행 표본과 같은 精度를 갖는다. 그러므로 표본의 크기는 약 2,000개로 한다. 앞의 서론에서 언급한대로 市郡별 오차를 5% 이내로 하려면 표본의 크기는 전수조사에 가까운 대표본이 필요하다. 예를 들어, 모든 市郡에서 닭의 표본오차를 5%로 하려면 전체 조사구수 25,969개 가운데 24,215개가 필요하게 된다.

〈표 7〉은 市郡별 목표오차를 5%로 할 경우에 표본의 크기이다. 이것은 市郡을 副次母集團으로 하고 단순임의추출을 가정하였을 때의 값이다.

〈표 7〉 市郡別 標本誤差 5%일 때의 標本의 크기(표본조사구수)

	調 査 區 數	農 家 數	한우 및 고기소	돼 지	닭
서 울	394	326	363	383	364
釜 山	252	201	235	242	247
京 畿	2,585	1,040	1,833	2,400	2,474
江 原	1,783	1,023	1,281	1,707	1,753
忠 北	2,351	1,197	1,570	2,127	2,255
忠 南	3,024	946	1,727	2,563	2,888
全 北	2,632	696	1,471	2,156	2,422
全 南	4,141	1,196	2,407	3,130	3,475
慶 北	4,400	1,700	2,436	3,968	4,202
慶 南	3,045	1,076	1,756	2,597	2,824
濟 州	1,362	678	1,133	1,175	1,311
계	25,969	10,079	16,262	22,508	24,215

#### 4. 標本配定

앞에서 고찰한 바와 같이 표본의 크기를 2,000, 1,800의 두가지로 정하고, 각각에 대하

여 가축 종류별로 표본배정을 하였다. 표본배정을 한 다음에 각 市郡별 표본오차를 계산하여 오차가 큰 市郡에는 표본을 더 늘리고, 오차가 작은 市郡에는 표본을 줄이면서 각 市郡의 오차를 조정하여 특별히 큰 값이 없도록 하였다. 이와 같이 여러번의 조정 과정을 거쳐서 최종적으로 市郡별 표본수와 가축별 표본오차를 산출하였다. <표 8>은 道別家畜別 표본 오차이다.

<표 8>을 보면 5개 항목에 걸쳐 서울, 부산, 제주도의 표본오차가 크고, 나머지 지역은 비교적 잘 균형을 이루고 있다. 처음에는 서울, 부산, 제주도와 다른 지역에 있어서도 표본오차의 차이가 이보다 훨씬 컸으나, 여러번의 조정을 거쳐서 상당히 감소된 것이다. 유의표본 조사구의 기준은 젖소는 1,500마리 이상, 한우와 돼지는 15,000마리 이상, 닭은 20,000마리 이상이 되는 조사구로 하였다.

<표 8> 道別, 家畜別 標本誤差(CV)

市	道	農 家 數	젖 소	韓牛 및 고기소	돼 지	닭
서	울	0.204	0.363	0.387	0.481	0.917
釜	山	0.163	0.453	0.357	0.367	0.340
京	畿	0.018	0.063	0.033	0.071	0.035
江	原	0.030	0.245	0.043	0.162	0.262
忠	北	0.035	0.281	0.051	0.150	0.262
忠	南	0.023	0.146	0.038	0.083	0.051
全	北	0.026	0.218	0.048	0.095	0.223
全	南	0.022	0.185	0.040	0.056	0.119
慶	北	0.019	0.183	0.029	0.099	0.042
慶	南	0.023	0.204	0.037	0.078	0.180
濟	州	0.082	0.771	0.157	0.142	0.527
全	國	0.008	0.048	0.014	0.030	0.012

### 5. 市 層化抽出

기본통계에서 가장 중요한 항목은 농가수이다. 농가수 표본오차를 보면 市에서 현저히 크고 郡 지역에서는 작다. <표 9>는 市郡별 농가수 CV의 분포이다. 여기서 CV 값이 큰 것은 모두 市에서 얻어진 것이다. 따라서, 표본의 크기를 증가시키지 않고 市에서의 표본오차를 감소시키는 것이 중요한 과제가 되었으며, 이를 위하여 市에서는 층화추출을 실시하였다. 市의 층화는 가축수를 기준으로 하여 작성하였다. 층화의 효과를 보기 위하여 층화하지 않은 경우와 층화한 경우의 표본오차를 비교한 결과 층화의 효과가 거의 두배 이상인 것을 확인하였다.

〈표 9〉 農家數 CV 의 分布

CV	度 數	相對度數	CV	度 數	相對度數
0.1	74	0.35	0.45~0.5	6	0.03
0.11~0.15	51	0.24	0.05~0.55	12	0.06
0.15~0.20	12	0.06	0.55~0.60	13	0.06
0.20~0.25	4	0.02	0.60~0.65	5	0.02
0.25~0.30	4	0.02	0.65~0.70	2	0.01
0.30~0.35	2	0.01	0.70~0.8	8	0.04
0.35~0.4	7	0.03	0.8~0.9	2	0.01
0.40~0.45	3	0.01	0.9 以上	2	0.01

## 6. 標本抽出 및 標本代替

추출단위는 市에서는 인구주택 조사구이며 郡에서는 농업 조사구를 사용하였다. 추출 방법은 市에서는 層化一段抽出을 하였고, 郡에서는 단순임의 추출을 하였다.

기본통계 조사 및 가축통계 조사는 계속조사가 아니라 연 1회 또는 4회 실시하는 時點 調査이기 때문에 표본대체에서 어려운 문제는 별로 없다. 표본대체는 표본 조사구의 대체와 전수조사 농가의 대체가 있다.

표본 조사구의 대체는 별로 문제가 없으며, 대체를 필요로 할 경우에 市에서는 같은 層內의 조사구에서 임의추출하면 되고, 郡에서는 같은 郡內의 조사구에서 임의추출하여 대체한다.

전수조사 농가의 리스트는 每分期마다 작성하는 것이 원칙이지만, 이러한 작업은 쉽지 않다. 따라서, 일정한 기준에 의하여 전수조사 농가의 탈락 또는 편입을 판정한다. 또한, 전수조사 농가의 리스트를 연 1회 작성한다면 다음 작성때까지 전수조사 농가를 변경해서는 안된다. 예를 들어, 농가의 젖소가 18마리에서 25마리로 변경되어도 처음에 선정된 전수조사 농가를 계속 보유하여야 한다.

## 7. 推 定

추정 방법으로서 線型推定과 比推定, 確率比例(PPS) 推定 등을 생각할 수 있다. 선형추정은 비추정이나 PPS 추정보다 精度가 떨어지나 간편한 장점이 있다. 비추정과 PPS 추정은 효율은 높지만 1980년 센서스 자료를 계속 이용하여야 하고 장기간 보관하여야 한다. 따라서, 행정상의 편의를 고려하여 선형추정 방법을 택하는 것이 좋다. 선형추정에 의한 추정 공식은 다음과 같다.

### (1) 市의 總計

市部는 총화추출을 하였으며, 총계와 분산의 공식은 다음과 같다.

$$\hat{X}_s = \sum \frac{N_h}{n_h} \sum X_{hi}$$

다만,  $N_h$  : s 시의 h 층의 인구주택 조사구수

$n_h$  : s 시의 h 층의 표본 인구주택 조사구수

$X_{hi}$  : s 시의 h 층의 i 번째 인구주택 조사구수의 관측치

$$\bar{X}_s = \hat{X}_s / N_s$$

$$V(\hat{X}_s) = \sum N_h^2 \frac{s_h^2}{n_h} \left( \frac{N_h - n_h}{N_h} \right)$$

(2) 郡의 總計

$$\hat{X}_g = \frac{N_g}{n_g} \sum X_{gi}$$

다만,  $N_g$  : g 군의 농업 조사구 수

$n_g$  : g 군의 표본 조사구 수

$X_{gi}$  : g 군의 i 조사구의 관측치

$$V(\hat{X}_g) = N_g^2 \cdot \frac{s_g^2}{n_g} \left( \frac{N_g - n_g}{N_g} \right)$$

(3) 全國 및 道の 總計

道總計 :

$$\hat{X}_d = \sum_g \hat{X}_g + \sum_s \hat{X}_s + \text{전수조사 농가 관측치} + \text{유의표본 조사구 관측치.}$$

다만, s, g 는 d 道內 g 郡 s 市

$$V(\hat{X}_d) = \sum V(\hat{X}_g) + \sum V(\hat{X}_s)$$

全國總計 :

$$\hat{X} = \sum \hat{X}_d$$

$$V(\hat{X}) = \sum V(\hat{X}_d)$$

## 참 고 문 헌

- (1) 농수산부, 1975. 農業統計調查要領.
- (2) 한국농촌경제연구원, 1980. 農業統計改善에 관한 研究.



# A Study on a Multi-purpose Sampling Design Involving Both a Farm Population Survey and a Livestock Survey

Park, Hong Nai

## ABSTRACT

This paper aims at studying on a multi-purpose sampling design involving both a farm population survey and a livestock survey.

Some features of this design are firstly to equalize the sampling errors among Shi-Gun (county) estimates, secondly, to make an efforts to increase the precision by introducing a purpose sample into the probability sample.

There are two types of purpose samples according to sampling units; one is farm, the other is enumeration district. A criterion of demarcation for selecting the farm purpose sample was investigated.