

농가경제조사를 위한 표본설계

신민웅¹⁾ 이계오²⁾ 홍기학³⁾ 이기재⁴⁾

요 약

본 논문에서는 급변하는 농촌의 환경을 충분히 반영할 수 있도록 1997년도에 설계되어 사용되고 있는 현행의 농가경제조사를 개선하였다.

새로운 표본 조사구를 선정하기 위하여 1999년도와 2000년도 농가경제조사 조사데이터와 2000년에 실시된 농어업총조사 결과를 심도 있게 분석하였다. 이를 기초로 현재의 농촌 실정에 적합한 조사모집단을 새롭게 구성하였고, 현재의 농촌 환경을 반영할 수 있는 층화 기준을 마련하여 표본 조사구를 추출하였다. 또한, 논벼를 비롯한 6개 주요작물들에 대한 농산물생산비조사의 정도(精度) 향상을 위해서 각 작물별 주산지별 표본 조사구를 선정하였다.

주요 용어 : 불균등확률추출, 복합표본설계, 사후층화, 상대표준오차, 설계 가중치

1. 서 론

농가경제조사는 다항목 복합표본조사로 농가경제와 관련된 여러 항목들을 조사하며, 동일한 표본 가구를 대상으로 5년간에 걸쳐 매년 조사되는 패널조사이다. 또한 농가경제조사의 표본은 농산물생산비조사에 동시에 이용되는 다목적 표본이다. 따라서 농가경제조사를 위한 표본설계에서는 이러한 조사의 특수성들이 반영될 수 있도록 층화와 불균등확률추출 등이 복합적으로 사용된 표본설계를 사용해야 한다. 또한 표본의 모집단에 대한 대표성을 높이고, 추정량의 精度를 높이기 위해서 모집단의 정보를 사후층화(post-stratification) 단계에서 이용하는 방안을 함께 검토해야 한다.

현행의 농가경제조사는 1997년도에 설계되어 그 동안의 급변하는 농촌의 환경을 충분히 반영하는데 미흡하고, 특히 2000년도에 실시된 농어업총조사를 통해서 새로운 모집단 자료를 얻게 됨에 따라 현재의 농촌 환경을 올바르게 반영할 수 있는 새로운 표본설계가 필요하게 되었다. 새로운 표본 설계의 목적은 표본 개편뿐만 아니라 조사현장에서 발생하는 문제점들을 보완할 수 있는 연구를 함께 함으로써 농가경제조사의 정도(精度)를 향상시키고자 함에 있다.

2. 현행 농가경제조사 조사 데이터 분석

현재 농가경제조사에서 사용되고 있는 표본은 1995년 농업총조사 및 지역조사를 기초로 작성된 모집단을 이용하여 추출되었다. 현행 표본설계의 조사모집단은 단독농가와 20호 미만의 조사구를 모집단에서 제외하는 등의 일련의 작업을 통해서 정비되었는데, 우리나라의 전체 농가

-
- 1) 한국외국어대학교 정보통계학과 교수
 - 2) 한국 Gallup조사연구소
 - 3) 동신대학교 컴퓨터학과 교수
 - 4) 한국방송통신대학교 정보통계학과 교수

농가경제조사를 위한 표본설계

인 1,500,740호에서 단독농가와 20호 미만의 조사구에 속한 257,251호를 제외한 1,243,491호이다.

현행 표본설계에서는 각 도를 부차모집단으로 설정하여 동일한 층화 기준을 이용하여 표본설계 하였고, 표본크기 및 표본배정을 위해서 고려된 변수는 농가소득이다. 각 지역별 표본크기는 지역별 농가소득의 상대표준오차(표본변이계수)가 균등하도록 결정되었다.

표본 조사구 및 조사대상가구 선정 방법은 각 지역별 표본배정을 마친 후 농가 수에 따라 확률비례추출법으로 표본 조사구를 추출하였고, 선정된 표본 조사구에 대하여 계통추출법으로 10호씩 표본농가를 선정하였다.

농가경제통계를 위한 각 모수의 추정을 위한 가중치는 매년 변화된 농업구조에 맞도록 매년 농업기본통계조사의 전·겸업 형태별, 지역별 농가수(단독가구 제외)를 사용하였다. 표본농가의 영농형태 유형은 1년간 농업 경영성과(농업조수입과 농외수입)를 집계하여 사후에 구분하였다.

3. 모집단 분석

새로운 표본설계의 모집단은 2000년 농업총조사에서 조사된 전체 농가로 서울특별시 및 부산, 대구, 인천, 광주, 대전, 울산 등의 광역시에 속한 개인농가를 포함해서 1,383,468호의 개인농가이다. 1995년 농업총조사 결과와 비교할 때, 2000년 농업총조사에 나타난 개인농가는 약 7.8% 줄어든 반면, 전업농가의 비중은 약 6.2% 정도 늘어났다.

단독 농가와 20호 미만의 조사구를 모집단에서 제외하는 등의 일련의 작업을 통하여 최종적으로 정비된 조사모집단의 총 농가 수는 1,115,252호이다. 현행 표본 설계와 비교해 볼 때, 총 농가 수는 1,383,468호에서 1,115,252호로 줄어들었다. 농업 의존도에서 보면 1종 겸업농가와 2종 겸업농가의 비중이 각각 18.0%와 20.6%로서 조사모집단에서 2종 겸업농가가 차지하는 비중이 적지 않음을 알 수 있었다

농가경제조사의 조사구로는 부락과 농업조사구, 그리고 인구주택조사구 등을 이용할 수 있다. 따라서 이 세 가지 조사구들의 장·단점을 분석하여 적절한 형태의 조사 지구를 결정할 필요가 있다. 새로운 표본설계의 조사구는 현행 표본설계와 마찬가지로 시 지역과 제주도는 농업조사구를 조사구로 결정하였고, 군 지역은 부락(리)를 조사구로 결정하였다.

4. 새로운 표본설계

4.1 층화

새로운 표본설계의 최종 층화 방법은 각 층에 속한 조사구 수, 총 농가 수, 동일농가비율 등에 대한 분석을 통해서 다음과 같은 층화 방법을 결정하였다.

즉, 현행 표본설계의 층화 방법을 따르되 논비 층을 해당 조사구의 논비 농가 비율에 따라서 70% 이상 층과 50-70% 층으로 구분하고, 해당 영농형태 층으로 분류되기 위한 최소 기준(10% 이상)을 주는 방법을 고려하였다. 다음 <표 4.1>은 새로운 표본설계의 최종 층화 기준을 정리한 것이다.

<표 4.1> 새로운 표본설계의 층화 기준

구 분	층 화 기 준
논벼지구1	전업 또는 1종 겸업농가이면서 논벼 농가의 비율이 70%이상인 지구
논벼지구2	전업 또는 1종 겸업농가이면서 논벼 농가의 비율이 50%-70%인 지구
과수지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 과수 농가가 최대인 지구(10% 이상)
채소지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 채소 농가가 최대인 지구(10% 이상)
특작지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 특작 농가가 최대인 지구(10% 이상)
화훼지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 화훼 농가가 최대인 지구(10% 이상)
전작지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 전작 농가가 최대인 지구(10% 이상)
축산지구	논벼 및 2종 겸업농가의 비율이 50%이하이고, 축산 농가가 최대인 지구(10% 이상)
기타지구	위의 영농특성 층이나 2종겸업 층에 속하지 않는 지구
2종겸업	2종 겸업농가의 비율이 50%이상인 지구

새로운 표본설계에서 사용된 층화 방법의 특징을 정리하면 다음과 같다.

첫째, 현행 표본설계에서 사용된 층화 방법과 마찬가지로 영농형태를 기준으로 층화함으로써 통계조사 결과의 연속성 유지에 바람직할 것이다.

둘째, 논벼 층을 논벼 농가 비율에 따라서 세분하여 층화함으로써 표본의 모집단에 대한 대표성을 높였고, 표본의 구성이 논벼 비율에 대해서 안정적일 것으로 기대한다.

셋째, 현행 표본설계의 층화 방법의 문제점을 개선하였다. 현행 층화 방법은 논벼 비율과 2종 겸업농가 비율이 각각 40% 이상인 경우에는 극히 소수 가구의 영농형태에 따라서 구분되게 된다. 새로운 표본설계의 층화는 현행의 방법을 유지하되 해당 영농형태로 분류할 때, 최소한의 기준(해당 영농형태의 농가 비율이 10% 이상)을 설정함으로써 이러한 문제점을 개선하였다.

새로운 표본설계에서 사용된 층화 방법을 적용하여 각 층별 조사구 수, 총 농가 수, 동일농가 비율 보면 전국을 기준으로 보면 57.7%로 나타났고, 동일농가 비율이 가장 높은 층은 논벼지구 1이고, 가장 낮은 층은 축산 층으로 나타났다.

4.2 지역별 표본수 결정 및 주산지 결정

농가경제조사는 지역, 영농형태, 전·겸업 구분 등의 다양한 영역에 대한 통계 작성을 필요로 한다. 새로운 표본설계의 표본배정은 농가경제조사의 이러한 특징을 반영할 수 있어야 한다.

새로운 표본설계에서의 표본 조사구 수는 현행 조사의 314개 표본 조사구와 농산물생산비조사를 위해서 추가된 6개의 주산지 표본 조사구를 합해서 모두 320개 조사구로 결정되었다. 내용을 보면, 주요 작물에 대한 생산비 통계의 정도(精度)를 향상시키기 위해서 각 작물에 대한 주산지 표본 조사구 수를 2개로 결정하였다. 따라서 전체 표본 조사구는 12개의 주산지 표본 조사구와 308개의 표본 조사구로 구성된다.

지역별 표본크기의 결정은 각 도에 대해서 30개의 조사구를 우선 배정하고, 나머지 표본은 비례배정법에 따라서 추가로 배정하는 방법에 기초하되 추가로 늘어나는 주요 작물에 대한 주산지 표본 조사구를 반영하여 결정하였다.

각 주요작물에 대한 2개의 주산지 표본 조사구는 유의추출법으로 선정하는 것을 원칙으로 하였다. 주산지를 포함한 지역별 최종표본 크기의 현황은 다음 <표4.2>와 같다.

<표 4.2> 주산지를 포함한 지역별 최종표본 크기 현황

	모집단 농가수 (%)	현행 (%)	새로운 배정	주산지 표본	최종 표본 (%)
경기	146,442 (13.1)	37 (11.8)	38		38(0) (11.9)
강원	64,842 (5.8)	36 (11.5)	34		34(0) (10.6)
충북	74,488 (6.7)	36 (11.5)	34		34(1) (10.6)
충남	162,044 (14.5)	37 (11.8)	38		38(0) (11.9)
전북	101,263 (9.1)	39 (12.4)	36		36(1) (11.3)
전남	177,708 (15.9)	38 (12.1)	39	2	41(3) (12.8)
경북	207,039 (18.6)	36 (11.5)	40	2	42(4) (13.1)
경남	146,532 (13.1)	38 (12.1)	38	1	39(2) (12.2)
제주	34,894 (3.1)	17 (5.4)	17	1	18(1) (5.6)
전국	1,115,252 (100)	314 (100)	314	6	320(12) (100%)

단, 최종표본에서 괄호 안의 수치는 주산지 시·군의 수임.

4.3 표본추출

본 연구에서는 기본적으로 추출된 표본 조사구가 조사구 내 동일농가비율을 높이고, 변화된 영농형태를 반영할 수 있도록 노력하였다. 따라서 층화에서 논벼 농가의 비율을 좀더 세분화 하였고, 농가경제조사를 위한 표본농가가 동시에 농산물생산비조사의 표본농가라는 점을 고려 하여 표본조사구를 추출하였다. 이를 위해 각 지역에서 층별로 배정된 표본 조사구 수만큼을 조사구 내의 농가수에 비례하는 확률비례계통추출법을 적용하여 추출하는 것을 원칙으로 하였다. 이렇게 추출된 표본 조사구들은 각 지역별 통계청 시·군 출장소에 따라서 집계하여 출장소별 조사 업무량과 현재의 인력을 감안하여 조정되었다.

조사농가의 선정은 표본 조사구 내의 모든 농가들의 영농형태를 파악하여 영농형태별 구성비를 산정 하고, 그 구성비에 따라서 10호의 조사농가들을 계통추출하는 방안을 제시하였다.

5. 추정

농가경제조사는 층화2단추출법에 의해서 표본 농가를 선정하여 조사한다. 층화는 전국의 9개 각 지역에 대해서 영농형태에 따라서 10개 층으로 구분하였다. 1차추출단위는 조사구이고, 2차추출단위는 각 농가이다. 농가경제조사의 주요 추정 내용을 살펴보면, 전국 농가소득, 농업소득, 농가부채 등의 추정 및 지역 별 추계, 그리고 전국 각 영농형태별 농가소득, 농업소득, 농가부채 등의 추정 등을 들 수 있다.

5.1 가중치 산정

농가경제조사의 표본설계는 층화, 2단계 집락추출, 가중치 등 표본설계의 여러 측면이 종합적으로 사용된 복합표본설계(complex sample design)이다. 일반적으로 추정 과정에서 가중치를 무시하고 분석하면 모수 추정에 심각한 편향(bias)이 발생할 수 있다. 농가경제조사와 같이 표본 규모가 큰 대규모 조사에서 문제가 되는 것은 추정량의 편향이기 때문에 추정 과정에서 반드시 가중치를 이용해야 한다.

일반 표본농가 및 주산지 표본조사구 내의 표본농가에 대한 설계 가중치는 각각 식 (5-1)과 (5-2)와 같이 주어진다.

$$w_{dhij} = \frac{N_{dh}S_{dh}}{n_{dh}S_{dhi}} \times \frac{M_{dhi}}{m_{dhi}} \quad (5-1)$$

$$w_{dhij} = \frac{M_{dhi}}{m_{dhi}} \quad (5-2)$$

여기서, d : 지역을 나타내는 첨자, $d=1, 2, \dots, 9$

h : 층을 나타내는 첨자, $h=1, 2, \dots, 10$

i : 표본 조사구를 나타내는 첨자, $i=1, 2, \dots, n_{dh}$

j : 표본 조사구 내에서 표본 농가를 나타내는 첨자, $j=1, 2, \dots, m_{dhi}$

w_{dhij} : 농가경제조사를 위한 설계 가중치

S : 표본추출에 사용된 크기 측도(표본설계 당시의 조사구 내 농가 수)

N_{dh} : 모집단에 있는 지역 d 의 층 h 내의 전체 조사구 수

n_{dh} : 지역 d , 층 h 내의 표본 조사구 수

M_{dhi} : 지역 d , 층 h 의 i 번째 표본 조사구 내의 전체 농가 수

(현지 조사 결과에 의한 조사구 내 단독농가를 제외한 전체 농가 수)

m_{dhi} : 지역 d , 층 h 의 i 번째 표본 조사구에서 조사 완료된 농가 수

(이 값은 $m_{dhi}=10$ 으로 고정됨)

지역 d 에서 전·겸업(전업농가, 1종 겸업농가, 2종 겸업농가) 구분에 따라서 추정된 값을 X 라 놓고, 조사모집단의 값을 X 라고 정의하면, 사후층화 조정된 가중치는 다음과 같다.

$$w_{dhij}' = w_{dhij} \times (X/\bar{X}) \quad (5-3)$$

이렇게 해서 구해진 사후층화 조정된 가중치 w_{dhij}' 를 간단한 표현을 위해서 w_{dhij} 로 표시하도록 한다. 이후의 각종 특성치 Y 에 대한 가중표본평균은 w_{dhij} 를 이용해서 구한다.

5.2 전국에 대한 추정

농가경제조사에서 각종 모집단에 대한 평균은 표본가중평균(標本加重平均)을 이용한다. 표본가중평균 및 분산 추정값은 다음과 같이 주어진다.

$$\bar{y} = \frac{\sum_{d=1}^9 \sum_{h=1}^L \sum_{i=1}^{n_{dh}} \sum_{j=1}^{m_{dhi}} w_{dhij} y_{dhij}}{\sum_{d=1}^9 \sum_{h=1}^L \sum_{i=1}^{n_{dh}} \sum_{j=1}^{m_{dhi}} w_{dhij}} \quad (5-4)$$

$$\widehat{var}(\bar{y}) = \frac{\sum_{d=1}^9 \sum_{h=1}^L \frac{n_{dh}}{n_{dh}-1} \sum_{i=1}^{n_{dh}} \left[W_{dhi} (\bar{y}_{dhi} - \bar{y}) - \frac{1}{n_{dh}} \sum_{s=1}^{n_{dh}} W_{dhs} (\bar{y}_{dhs} - \bar{y}) \right]^2}{\left(\sum_{d=1}^9 \sum_{h=1}^L \sum_{i=1}^{n_{dh}} W_{dhi} \right)^2} \quad (5-5)$$

여기서, $W_{dhi} = \sum_{j=1}^{m_{dhi}} w_{dhij}$ 이고, $\bar{y}_{dhi} = \sum_{j=1}^{m_{dhi}} w_{dhij} y_{dhij} / \sum_{j=1}^{m_{dhi}} w_{dhij}$ 이다.

지역별 추정 과 영농형태별 추정도 비슷한 방법으로 구할 수 있다.

6. 요약 및 결론

본 연구에서는 현행 농가경제조사에서 나타나는 미비점을 개선하여, 조사항목에 대한 통계를 안정적으로 산출할 수 있도록 새롭게 설계하였다. 새로운 표본설계의 특징을 요약하면 다음과 같다.

첫째, 전체적으로 모집단의 대표성을 높여주고, 표본의 구성을 안정되게 해주기 위하여 논벼 층을 논벼 농가 비율에 따라서 세분하여 기존의 논벼 층을 논벼 농가 비율이 70%이상인 논벼 지구1과 논벼 비율이 50-70%인 논벼지구2로 세분하여 층화하였다.

둘째, 현행 층화방법에서는 논벼 비율과 2종 겸업농가의 비율이 모두 50% 미만인 조사구들에 대해서 가장 다수를 차지하는 영농형태로 해당 조사구를 층화하고 있다.

셋째, 농산물 생산비조사를 위한 주산지를 추가적으로 표본 조사구에 포함시켰다.

넷째, 농가경제조사 항목에 대한 추정치의 정도(精度)가 지역별로 일정 수준을 유지하도록 표본 수 결정과 표본배정법에 대하여 연구하였다.

이상과 같은 작업을 통하여 농가경제조사 표본지구와 추정량을 제시하였다. 바라건대 본 연구에서 제시한 새로운 표본설계와 체계적인 조사를 통해서 농가경제조사의 신뢰도 높은 통계 생산을 기대한다.

참 고 문 헌

- [1] 박재수(1989). 「표본조사법(이론과 실제)」, 박영사.
- [2] 박홍래(1989). 「통계조사론」, 영지문화사.
- [3] 서울대학교 통계연구소(1997). 「농가경제조사 표본설계 용역 최종보고서」.
- [4] Cochran, W. G.(1977). Sampling Technique, 3rd edition, John Wiley & Sons, New York, U.S.A.
- [5] Hansen, M. H., Hurwitz, W. N. and Madow, W. G.(1953). Sample Survey Methods and Theory, New York: Wiley, U.S.A.
- [6] Kish, L.(1965). Survey Sampling, New York: Wiley, U.S.A.
- [7] Lehtonen, R. and Pahkinen, E. J.(1995). Practical Methods for Design and Analysis of Complex Surveys, New York: Wiley, U.S.A.
- [8] Lessler, J. T. and Kalsbeek, W. D. (1992), Nonsampling error in surveys, John Wiley and Sons, U.S.A.
- [9] Little, R. J. A. and Rubin, D. B. (1987), Statistical Analysis with missing data, John Wiley and Sons, U.S.A.
- [10] Yates, F.(1981). Sampling Methods for Censuses and Surveys, 4th edition, U.S.A.