

A Sampling Design of the Agricultural Machine Estimated Sales Survey

Jinwoo Park ¹⁾

Abstract

The agricultural machine estimated sales survey is a survey to estimate annual sales quantities of eight major agricultural machines such as tractor, combine, etc. The purpose of this study is to design a multipurpose sample for the agricultural machine estimated sales survey. Main achievements of this study are to present an efficient stratification criterion and to suggest a reasonable estimation method by using the concept of post-stratification.

Keywords : multipurpose sample, stratified sampling, allocation, post-stratification.

1. 서론

도시화, 산업화, 정보화 등은 우리 나라의 농업 산업에 지대한 영향을 미치고 있다. 또한 각종 농산물에 대한 수입개방이 이어짐으로 농촌 환경은 급속한 변화를 계속해가고 있다. 정부에서는 농민들이 이러한 변화의 소용돌이 속에서도 효과적으로 대처하여 경쟁력 있고 생산성 있는 영농을 해갈 수 있도록 다양한 대책들을 마련하고 있다. 농림부에서 농민들의 농업기계 구입을 위한 융자지원을 해주는 일도 그 한가지 예라고 볼 수 있다.

농업기계 구입을 위한 융자지원 예산의 규모를 과학적으로 결정하기 위해서는 주요 농기계에 대한 수요 동향을 파악하는 것이 필요하다. 연말이 되면 농협중앙회를 통해 한해 동안 판매된 농기계 판매량에 대한 정확한 통계들이 작성될 수 있다. 하지만 상반기 중에 올해의 판매량을 예측하기 위해서는 표본조사가 필요하다. 일반적으로 새로운 해의 농기계 구입 양상은 지난해와는 조금씩 달라지는 경향이 있기 때문에 전적으로 지난해의 자료만 이용해서 새로운 해의 수요량을 예측할 수는 없다. 따라서 농업기계연구소에서는 전국 농기계 취급 점포들을 대상으로 하여 트랙터, 콤바인, 경운기, 이앙기, 관리기, SS기, 농작물 건조기, 곡물 건조기 등 여덟 가지 주요 기종의 농기계 예상 판매량을 파악하기 위한 표본조사를 계획하였다.

주요 농기계들은 전국의 농기계 대리점 및 농협에서 판매되고 있는데 2000년을 기준으로 할 때 대리점의 수는 약 1379개, 농협은 855 개이다. 농기계를 판매하는 대리점이나 농협에 따라 취급하

1) Assistant Professor, Department of Applied Statistics, The University of Suwon, P.O.Box 77, Suwon, Kyunggi-Do, 445-743, Korea.
E-mail : jwpark@mail.suwon.ac.kr

는 기종들의 종류가 서로 많이 다르다. 농협과 대리점의 특성이 다를 뿐만 아니라 대리점들간 또는 농협들 간에도 취급하는 품목의 종류들도 서로 상이한 경우가 많다. 전국의 조사모집단 단위의 총 수가 2234 개이지만 각 품목별로 보면 그에 훨씬 미치지 못한다.

본 연구의 목적은 여러 가지 주요 농업기계들의 판매량에 대한 정도 높은 추정량들을 얻을 수 있도록 하는 하나의 다목적표본(multi-purpose sample)을 설계하는 것이다. 농촌기계연구소에서는 인원, 예산 등 여러 제약 조건으로 인하여 애초에 전체 표본의 크기가 가급적 250개 내외를 생각했는데 본 연구의 결과 최종 표본의 크기가 304개로 결정되었다. 전국의 대리점 및 농협들이 조사모집단이 되는데 먼저 각 조사단위에서 취급하는 기계의 종류 및 판매량의 특성을 파악하여 층화 기준을 정한다. 기종에 따라 약간의 차이가 있기는 하지만 일반적으로 특정 품목을 전혀 취급하지 않는 조사단위들이 매우 많으므로 이를 고려하여 사후층화를 사용한 추정법을 제시한다.

2절에서는 모집단 특성의 분석, 3절에서는 층화, 4절에서는 표본크기의 결정, 5절에서는 추정법에 대해 다룬다.

2. 모집단

농기계 판매량 조사에서 조사단위는 트랙터, 콤파인, 경운기, 이앙기, 관리기, SS기, 농작물건조기, 곡물건조기를 판매하는 대리점이나 농협이며, 모집단은 전국의 농업기계 대리점 및 농협이 된다. 표본설계를 위한 표집틀(sampling frame)로는 2000년 농업기계 대리점 및 농협 리스트를 사용하였는데 대리점 수는 1379개, 농협 수는 855개이었다.

각 기계별 2000년 판매량 및 취급 대리점 수를 나타낸 결과가 아래의 <표 1>에 나와있다.

<표 1> 농기계별 2000년 판매량 및 취급 점포의 수

	대리점		농협		총 판매량
	취급대리점수	판매량	취급판매대수	판매량	
트랙터	578	16,282	448	1,913	18,195
컴바인	555	9,100	439	1,373	10,473
경운기	255	2,348	515	2,931	5,279
이앙기	481	7,257	760	5,829	13,086
관리기	321	2,414	647	3,666	6,080
SS기	236	1,849	122	383	2,232
농작물관리기	542	3,752	383	1,357	5,109
곡물관리기	198	1,919	202	438	2,357
전국 점포수	1379		855		

트랙터, 콤파인 등의 큰 기종들은 대체로 대리점을 통해 판매되는 반면, 이앙기, 관리기, 경운기 등 소형기종은 상대적으로 농협에서 많이 취급되는 양상이다. 또한 트랙터, 콤파인, 이앙기, 농작물관리기를 판매하는 대리점의 수는 많으나 나머지 기종들을 취급하는 대리점은 적은 편이다. 농협에서는 이앙기, 관리기, 경운기 등을 판매하는 점포가 많은 편인데 이를 통해 대리점과 농협에

서 주로 다루는 농기계의 종류가 서로 다르다는 것을 알 수 있다. 한편 SS기나 곡물건조기는 일부 대리점에서만 취급하고 있을 뿐 대부분의 곳에서는 판매되고 있지 않는 것으로 나타났다. 대리점당 평균 취급 기종의 수는 2.30 종, 농협의 평균 취급 기종 수는 4.11 종인 것으로 나타났다. 대리점은 소수의 기종을 다량으로 판매하는 반면 농협은 다수의 기종들을 소량으로 판매한다는 특성을 지니고 있다.

3. 총 화

가장 단순한 총화기준으로 각 조사단위별 모든 기종의 총 판매량을 생각할 수 있을 것이다. 그러나 총 판매량을 총화기준으로 삼을 경우 판매량이 상대적으로 많은 콤팩트나 트랙터 위주의 설계가 이루어지게 되어 콤팩트와 트랙터에 대한 추정의 정도는 높아지는 반면 다른 기종들에 대한 추정의 정도는 매우 떨어지는 표본설계가 될 우려가 있다. 이를 피하기 위해 먼저 각 점포별로 취급하는 품목들간에 어떤 상관성이 있는 지를 고려하였다. 각 점포에서 특정 기종을 취급하면 1, 않으면 0으로 처리한 후 기종별 상관계수를 구한 결과가 <표 2>에 나와 있다.

<표 2> 농기계별 상관계수

		트랙터	컴바인	경운기	이앙기	관리기	SS기	농산물	곡물
대리점	트랙터	1.0	0.89	0.55	0.85	0.54	-0.39	-0.47	-0.34
	컴바인		1.0	0.56	0.82	0.56	-0.37	-0.45	-0.33
	경운기			1.0	0.62	0.26	-0.22	-0.18	-0.20
	이앙기				1.0	0.37	-0.33	-0.37	-0.30
	관리기					1.0	-0.25	-0.42	-0.23
	SS기						1.0	-0.36	-0.19
	농산물건조기							1.0	0.28
	곡물건조기								1.0
농협	트랙터	1.0	0.15	0.58	0.13	0.10	0.03	0.04	0.03
	컴바인		1.0	0.07	0.25	0.11	0.01	0.18	0.21
	경운기			1.0	0.12	0.05	0.07	0.05	-0.05
	이앙기				1.0	0.15	-0.02	0.17	0.13
	관리기					1.0	0.15	0.13	0.08
	SS기						1.0	-0.05	-0.07
	농산물건조기							1.0	0.18
	곡물건조기								1.0

대리점의 경우 트랙터, 컴바인, 이앙기, 관리기, 경운기들간에는 비교적 큰 양의 상관관계를 갖는데 비해 SS기, 농산물건조기, 곡물건조기는 나머지 기종들과 약한 음의 상관을 갖는다. 여기서 트랙터와 컴바인이 양의 상관을 갖는다는 것은 트랙터를 취급하는 대리점일 경우 그 곳에서 컴바

와 상관성이 높은 콤바인 등의 추정치는 별 문제가 없지만 그와 상관성이 낮은 SS기, 농산물건조기, 곡물건조기들의 추정치의 정도는 매우 나빠지게 될 것이다. 농협은 대리점과는 달리 전체적으로 기종별 상관관계가 약한 편인데 농협에서는 다양한 기종들을 소수만 판매하는 경향이 있으므로 나타나는 현상이다.

위의 사실을 토대로 트랙터, 콤바인, 이앙기 등 판매량이 많은 기종들을 중심으로 표본설계를 하되 그 밖의 기종인 SS기, 농산물건조기, 곡물건조기 등에 대해서도 어느 정도 신뢰할만한 추정값을 얻을 수 있는 표본을 마련하기 위해 다음과 같은 두 가지 총화변수를 고려하기로 한다:

- (1) 총화변수1 = 트랙터 판매량 + 콤바인 판매량
- (2) 총화변수2 = SS기판매량 + 농산물건조기판매량 + 곡물건조기 판매량.

총화변수1에 이앙기, 경운기, 관리기 등을 포함시키지 않은 까닭은 첫째 이 기종들이 트랙터와 콤바인과 상관성이 높으므로 두 기종만 고려해도 나머지 기종에 대한 고려가 이루어지기 때문이며, 둘째 상대적으로 트랙터나 콤바인이 중요한 기종이기 때문이다.

다음의 <표 3>은 위 총화변수 구분에 따른 전체 모집단의 구성을 나타낸 표이다. 대리점 중 총화변수1에 포함된 기종들을 취급하는 대리점은 832곳, 총화변수 2에 속한 기종들을 취급하는 대리점은 602곳이다. 그런데 총화변수 1과 2의 기종을 동시에 취급하는 대리점의 수는 73곳에 불과한 반면, 어느 한 쪽의 기종들만 다루는 대리점의 수는 1288 곳이나 된다. 한편 농협의 경우 대리점보다는 덜하지만 비슷한 양상이 나타남을 알 수 있다. 이러한 결과를 통해 위에 제시한 총화변수가 서로 다른 특성의 대리점이나 농협들을 효과적으로 구분하는 총화기준이 된다고 할 수 있다.

<표 3> 총화변수 구분에 따른 전체 모집단 구성

점포구분		총화변수 1을 취급않는 점포수	총화변수 1을 취급하는 점포수	합 계
대리점	총화변수 2를 취급않는 점포수	18	759	777
	총화변수 2를 취급하는 점포수	529	73	602
	합 계	547	832	1379
농 협	총화변수 2를 취급않는 점포수	116	115	231
	총화변수 2를 취급하는 점포수	210	414	624
	합 계	326	529	855

다음으로 각 총화변수 별로 층의 구분을 하기 위한 경계점을 정하는 것이 필요하다. 각 변수별로 대리점, 농협 각각에 대해 두 개의 범주로 나누어 전체 층의 수가 8이 되게 하였다. 경계점은 누적도수제공된 규칙(Cochran, 1977)을 이용하여 결정하였는데 대리점인 경우 총화변수 1의 값 35, 총화변수 2의 값 5를 경계로 하여 4개의 층으로 나누었으며, 농협인 경우 총화변수 1과 총화변수 2 모두 5를 경계로 하여 4개의 층으로 구분하였다. 8개 층별 모집단 단위수는 <표 4>에 나타나있다.

<표 4> 층화구분에 따른 층별 모집단 단위수

층		층구분 기준	모집단 단위수
대리점	층 1	층화변수1<35, 층화변수 2<5	715
	층 2	층화변수1≥35, 층화변수 2<5	362
	층 3	층화변수1<35, 층화변수 2≥5	279
	층 4	층화변수1≥35, 층화변수 2≥5	23
대리점 합 계			1379
농협	층 1	층화변수1<5, 층화변수 2<5	463
	층 2	층화변수1≥5, 층화변수 2<5	126
	층 3	층화변수1<5, 층화변수 2≥5	173
	층 4	층화변수1≥5, 층화변수 2≥5	93
농 협 합 계			'855

4. 표본배분

농촌기계연구소에서는 처음에 예산상, 시간상 제약으로 인해 표본의 크기 250 내외가 되도록 하는 설계를 고려하였다. 먼저 전체 표본크기를 250으로 하고 비례배분법을 사용하여 각 층별 표본의 크기를 정했을 때 그에 따른 주요 기종별 추정오차를 계산하여 변동계수로 표시한 결과가 <표 6>에 나와 있다. 여기서 이용한 자료는 2000년 판매량 자료이다. 각 층별 모집단 수에 비해 실제 특정 기종을 취급하는 단위의 수는 적으므로 사후층화의 개념을 도입하여 유도한 추정식을 사용하였다.

<표 6> 비례배분에 따른 표본오차

층 구분	모집단 단위수	비례배분 표본크기	트랙터		컴바인			
			취급점수	취급표본수	취급점수	취급표본수		
대리점 1	715	80	269	30	249	28		
대리점 2	362	41	7	1	7	1		
대리점 3	279	31	279	31	276	31		
대리점 4	23	3	23	3	23	3		
농 협 1	463	52	171	19	149	17		
농 협 2	126	14	51	6	55	6		
농 협 3	173	19	156	17	150	17		
농 협 4	93	10	70	8	85	9		
합 계	2234	250	1026	합	115	994	합	111
				CV	6.22%		CV	7.76%

(계속)

총 구분	모집단 단위수	비례배분 표본크기	경운기		SS기			
			취급점수	취급표본수	취급점수	취급표본수		
대리점 1	715	80	54	6	140	16		
대리점 2	362	41	5	1	96	11		
대리점 3	279	31	177	20	0	0		
대리점 4	23	3	19	2	0	0		
농 협 1	463	52	233	26	52	6		
농 협 2	126	14	69	8	33	4		
농 협 3	173	19	143	16	13	1		
농 협 4	93	10	70	8	24	3		
합 계	2234	250	770	합	87	358	합	41
				CV	13.57%		CV	17.83%

<표 6>을 보면 표본의 크기가 250일 때 기종별 전국 판매량 추정값의 변동계수(Coefficient of Variation: CV)가 6.22%에서 17.83%였다. 한편 같은 표본 크기 하에서 다목적 표본이 아니라 각각의 기종 하나만을 위한 설계를 하며 동시에 최적배분을 사용한다고 해도 CV값은 4.56%에서 10.91%에 이르는 것으로 나타났다. 그런데 사실은 다목적 설계를 해야하는 상황이므로 표본의 크기를 250으로 해서 원하는 목표 CV를 5-10% 정도가 되게 하기는 어려운 상황이다. 이상의 결과를 자료를 토대로 하여 농촌기계연구소와의 협의를 거쳐 표본의 크기를 재조정하였다. 상대적으로 취급점포가 많으며 판매량이 많은 트랙터, 컴바인, 이앙기 등의 목표정도를 5% 내외가 되게 하고, 그렇지 못한 SS기, 곡물관리기 등의 목표정도를 10% 내외가 되게 하도록 하되 기종별로 최적배분법에 의해 표본의 크기를 정한 결과 전국의 표본크기는 304개가 되었다. <표 7>에는 최종 확정된 각 층별 표본의 크기와 모집단에서 각 기종을 취급하는 점포수 및 기종별 CV의 추정값이 나와 있다. 트랙터, 컴바인, 이앙기의 CV값이 각각 4.56%, 5.57%, 5.66%로 정도가 높은 편인 반면, 경운기, SS기, 관리기, 곡물건조기 등은 CV값이 10-11%에 이른다.

<표 7> 새로운 표본의 크기, 기종별 취급 점포수 및 목표 CV

총 구분	모집단 단위수	표본 크기	트랙터	컴바인	경운기	이앙기	관리기	SS기	농산물 건조기	곡물 건조기
대리 1	715	80	269	249	54	180	177	140	245	44
대리 2	362	80	7	7	5	7	0	96	242	153
대리 3	279	70	279	276	177	271	143	0	32	1
대리 4	23	4	23	23	19	23	1	0	23	0
농협 1	463	30	171	149	233	380	312	52	130	66
농협 2	126	10	51	55	69	119	104	33	109	50
농협 3	173	10	156	150	143	169	149	13	65	35
농협 4	93	20	70	85	70	92	82	24	79	51
합 계	2234	304	1026 (4.56%)	994 (5.57%)	770 (11.21%)	1241 (5.66%)	968 (11.14%)	358 (10.91%)	925 (7.53%)	400 (10.37%)

5. 추 정

전국의 모집단 단위수는 2234 개이지만 개개의 기종별로 보면 특정 기종을 취급하지 않는 단위들의 수가 대개 반 이상이 되므로 표본 단위 중에도 특정 기종을 취급하지 않는 단위들이 많이 포함되어 있을 것이다. 따라서 특정 기종을 판매하는 점포와 그렇지 않은 점포를 두 개의 층으로 구분하는 사후층화의 개념을 사용하는 것이 필요하다. 현재 8개의 층으로 층화되어 있는데 이를 다시 각 층별로 특정 기종을 취급하는 층과 취급하지 않는 층으로 사후층화한다. 다음의 식 (1)은 층별 판매량의 추정식이며, 식 (2)는 그 분산의 추정식이다.

$$\hat{Y}_h = \frac{N_{h1}}{n_{h1}} \sum_{j=1}^{n_{h1}} y_{hj}, \tag{1}$$

$$\hat{V}(\hat{Y}_h) = N_{h1} s_{h1}^2 / f_h + \frac{1-f_h}{n_{h1}} \cdot s_{h1}^2 (N_h - N_{h1}) / f_h - N_{h1} s_{h1}^2. \tag{2}$$

여기서,

h : 층 구분 ($h=1, 2, \dots, 8$),

N_h : 모집단 단위수,

n_h : 표본단위수

N_{h1} : N_h 모집단 단위들 중 해당 기종을 다루는 단위의 수,

n_{h1} : n_h 표본 단위들 중 해당 기종을 다루는 단위의 수,

$f_h = n_h / N_h$,

$$s_{h1}^2 = \frac{1}{n_{h1} - 1} \sum_{j=1}^{n_{h1}} \left(y_{hj} - \frac{\hat{Y}_h}{N_{h1}} \right)^2$$

을 나타낸다.

위의 식 (1)과 (2)를 이용하여 전국 판매량의 추정량과 그 분산의 추정량을 구하면 각각 다음의 식 (3), (4)와 같이 나타내진다.

$$\hat{Y} = \sum_{h=1}^8 \hat{Y}_h = \sum_{h=1}^8 \frac{N_{h1}}{n_{h1}} \sum_{j=1}^{n_{h1}} y_{hj}, \tag{3}$$

$$\begin{aligned} \hat{V}(\hat{Y}) &= \hat{V}\left(\sum_{h=1}^8 \hat{Y}_h\right) = \sum_{h=1}^8 \hat{V}(\hat{Y}_h) \\ &= \sum_{h=1}^8 \left[N_{h1} s_{h1}^2 / f_h + \frac{1-f_h}{n_{h1}} \cdot s_{h1}^2 (N_h - N_{h1}) / f_h - N_{h1} s_{h1}^2 \right]. \end{aligned} \tag{4}$$

참 고 문 헌

- [1] 박홍래(2000), 「통계조사론」 (2판), 영지문화사.
- [2] 박홍래(1989), “면적조사 및 생산량조사 표본설계”, 「박홍래교수회갑기념논총」, 55-74.
- [3] Cochran, W. G.(1977), *Sampling Techniques (3rd ed.)*, New York: John Wiley.
- [4] Holt, D. and Smith, T. M. F.(1979), Post Stratification, *Journal of the Royal Statistical Society A*, Vol. 142, 33-46.