

A Sample Design for Forestry Management Survey

Kay-O Lee¹⁾, Jeongbin Yoo²⁾

Abstract

In this paper, a sample design is studied for 2000 forestry management survey of five types forestry ; tree felling, gathering of pine mushroom, growing of nut trees, growing of wild flowers, and lumbering industry. We introduce population stratification and a modified stratified cut-off sampling which deal with determination of sample size, sample allocation, and estimation of total and variance of estimator. Substitution of sample units and imputation of nonresponse units are discussed for reducing the nonsampling errors.

Keywords : modified stratified cut-off sampling, sample allocation, substitution

1. 서론

21세기 임업 진흥 정책과 국가 경제 정책 수립을 위하여 임업의 기본 구조 및 임업 관련 산업의 현 상태를 파악하기 위해서 1999년 11월 임업총조사([1])를 실시하여 임업 4개 대분류를 구성하는 18개 중분류업종에 대하여 기본적인 통계를 정리하였다. 이 자료를 통해서 업종별 경영실태를 정확히 파악하는데는 제한사항이 있으므로 효율적인 임업 진흥 정책을 적시에 수립하고 시행하기 위해서 업종별로 심층적인 통계가 필요하다. 18개 업종에 대한 경영실태조사를 한번에 할 수 있는 예산과 전문인력이 없으므로 우선 4년동안 연차적으로 경영실태조사를 실시할 계획으로 첫 해인 2000년에는 5개 중분류 업종인 벌목업, 송이채취업, 유실수재배업, 야생화재배업과 제재업의 경영실태 조사를 위한 표본설계를 연구하였다. 이 중 임업은 업종별 특성이 다양하고 각 업종 내에서 가구 또는 단체 규모와 생산액의 변동이 크기 때문에 각 업종별로 충화한 후에 연구변수의 특성을 분석하고, 지방자치단체별로 통계의 활용성을 고려하여 시도별 통계작성도 가능하도록 새로운 표본조사 설계를 연구하였다.

본 연구에서는 업종별로 경영실태를 잘 나타낼 수 있는 변수를 1999년 임업총조사의 조사항목에서 일부를 선택하여, 각 변수별로 특성을 분석한 뒤 이것을 각 시도별 충화변수로 사용하였다. 그리고 조사대상이 되는 조사모집단은 1999년 임업총조사에서 임업가구로 분류된 가구와 1999년 광공업 통계조사에서 제재업으로 분류된 개인이나 사업체로 정의하였다. 위의 정보들을 이용하여

1) Professor, Department of Computer Science and Statistics, Korea Air Force Academy, Chungbuk, 363-849
E-mail : kayolee@afa.ac.kr

2) Professor, Department of Applied Statistics, 231 Mochung-dong, ChungJu ChungBuk, Korea, 361-742

기업체 통계조사에서 추출단위의 규모가 다양할 때 적용하는 표본추출법 중의 하나인 절사추출법 ([2])을 적용할 수 있도록 새로운 표본추출법으로 수정된 총화절사추출법을 제안하였다. 이 논문은 다음과 같이 구성되어 있다. 제2절에서는 조사모집단 분석 및 총화를 다루고, 제3절에서는 표본설계에 관련된 내용을 언급하였으며, 제4절과 제5절에서는 각각 표본관리와 결언 및 제안을 설명하였다.

2. 모집단 분석

임업에 종사하는 개인 임가와 준임가를 목표모집단의 구성원으로 생각할 수 있으나 2000년 임업경영실태조사에서는 5개업종에 대해서만 통계조사를 하므로 부모집단을 조사모집단으로 정하였다. 조사모집단의 구성은 벌목업, 송이채취업, 유실수재배업과 야생화재배업은 1999년 임업총조사의 임가로, 제재업은 통계청의 광공업 통계조사의 개인이나 사업체로 추출단위를 정의하였다.

2.1 조사모집단 정의

1999년 11월에 실시한 임업총조사에서 조사된 임업가구와 제재업의 사업체를 포함한 31,274단위를 조사모집단으로 정의하였고, 2000년에 조사할 5개 업종은 <표 2.1>과 같이 분류하였으며, 유실수 재배업은 4개로 세분하여 8개 업종별로 조사모집단을 구성하고, 각 업종별로 표본 추출률을 작성하였다. 업종별, 시도별 조사모집단의 크기는 <표 2.1>에 주어졌다.

<표 2.1> 업종별 모집단 크기

단위: 가구수/업체수

구 분	합계	벌목업	송이 채취업	유실수 재배업				야생화 재배업	제재업
				밤나무 재배업	호두나무 재배업	대추나무 재배업	잣나무 재배업		
광역시	1,380	38	10	255	9	63	273	13	719
경기도	1,875	73	4	302	7	23	1,039	54	373
강원도	1,654	153	1,270	76	9	16	56	19	55
충청북도	1,448	104	184	318	131	359	305	4	43
충청남도	3,822	354	7	2,896	341	124	37	22	41
전라북도	1,907	135	216	796	195	302	98	26	139
전라남도	4,674	128	37	4,308	3	28	7	73	90
경상북도	5,044	87	1,324	240	88	3,142	60	16	87
경상남도	9,470	19	225	8,080	13	922	60	31	120
합계	31,274	1,091	3,277	17,271	796	4,979	1,935	258	1,667

2.2 조사모집단 총화

표본조사 업무통제의 편의성과 도시지역과 농촌지역 간의 특성 차이가 유의할 것으로 생각되고 또한 추정의 효율을 제고하기 위해서 ‘광역시’ 지역과 ‘도’ 지역으로 크게 분류하였다. 그러나 지역 간의 불균형적인 표본배정을 예방하고 차후 지역별 통계 생산([7], [8])에 대비하기 위해서 ‘도’ 지

역은 ‘경기도’, ‘강원도’, ‘충청북도’, ‘충청남도’, ‘전라북도’, ‘전라남도’, ‘경상북도’, ‘경상남도’로 구분하여 모든 업종에 대해서 9개 층으로 모집단을 구분하였다. 단, 제재업의 경우에는 광역시에 총사업체의 43% 정도가 분포되어 있으므로 ‘광역시’ 층을 ‘서울특별시’ 및 6개 광역시로 구분하여 총 15개 층으로 세분화하였고, 제주도의 조사대상 단위수는 10개로 극소수이기 때문에 전체 합계의 추정에 미치는 영향이 거의 없을뿐만 아니라 조사업무 실행과 관장의 어려움이 예상되기 때문에 조사대상에서 제외하였다.

3. 표본설계

앞 절에서 조사모집단의 특성과 층화를 다루었으며, 본 절에서는 추출단위의 추출방법을 언급하고 표본크기의 결정과 배분을 다루겠다.

3.1 표본추출 방법

각 업종별로 주요 관심변수를 산림청 담당자의 자문을 얻어 선정한 후, 그 변수에 대한 층별 합계와 분산을 계산해 본 결과, 각 층 내에서는 법인과 개인가구가 혼합되어 있을 뿐만 아니라 주요 핵심변수의 큰 이상값들이 적지 않기 때문에 절사추출법(Cut-off Sampling Method)을 적용하여 추정의 효율을 높이고자 한다.([2], [3], [4])

절사추출법은, 각 층에서 변수 값이 큰 몇 개의 추출단위들이 층 합계의 추정에 있어서 80~90%의 비중을 차지할 경우, 층 총계의 추정의 효율을 높이기 위하여 큰 변수 값을 갖는 추출단위들은 모두 조사하고, 나머지의 추출률에서 표본을 추출하는 방식을 말한다.([2], [5]) 예를 들어 벌목업의 경우, 광역시에서 8가구를 표본으로 추출하고자 한다고 가정하자. 광역시의 모집단에서 가구들을 벌목면적의 크기 순으로 나열하면 다음과 같다.

90, 50, 50, 37.8, 20, 13, 12, 10, 10, …, 0.3, 0.3, 0.1

이때, 앞에서 4개 가구의 벌목면적 합은 227.8 ha로 전체 면적의 61% 정도이다. 만일, 4개 가구들은 무조건 표본으로 선정하여 조사하고, 나머지 34 (= 38 - 4)가구 중 4개 가구를 표본가구로

추출하는 방식이 절사추출법이다. 절사추출법에서 표본의 크기가 사전에 정해지지 않고, 단지 신뢰수준과 상대허용오차만 주어졌을 때에는, 모집단을 크기 순으로 나열한 후, 값이 큰 추출단위부터 하나씩 전수조사단위로 포함시키고 나머지 단위들의 분산과 주어진 신뢰수준과 허용상대오차로부터 표본조사단위들의 수를 결정한다. 실제 표본조사단위의 수 n 은 '(전수조사단위의 수) + (표본조사단위의 수)'이므로, 신뢰수준과 상대허용오차 등을 이용하여 가장 작은 n 값을 찾아서 표본의 크기로 결정한다. 물론, 이때 전수조사단위의 수와 표본조사단위의 수는 쉽게 결정할 수 있다. 절사추출법에 관한 세부내용은 3.4 소절의 추정방법에서 언급하기로 하겠다([5], [6]).

주어진 자료는 모집단의 규모가 크기 때문에, 먼저 업종별로 적당한 크기의 목표 CV(Coefficient of Variation)를 선택하였고, 전국적인 표본크기를 산림청과 협의하여 모집단의

약 20% 내외로 결정한 후, 각 층별로 표본배정은 최적배분 방식을 적용하였다. 각 층에서 표본추출은 주요 핵심변수의 큰 이상값들이 적지 않기 때문에, 절사추출법으로 층별 전수조사단위의 수를 먼저 결정하고, 각 층에서 전수조사단위들을 제외한 추출단위들로부터 표본조사단위를 선정하였다.

3.2 표본크기 결정

표본크기를 결정하는데 고려되어야 할 핵심적 요소는 표본추출 방법, 목표허용오차, 추정방법 및 조사비용 등이다. 본 연구에서는 조사비용에 대한 제한사항은 고려하지 않지만, 조사단위 당 소요비용은 동일한 것으로 가정하며, 최대 표본크기는 모집단의 20% 정도로 한정하였다. 우선 표본추출 방법은 절사층화계통추출방법을 적용하고, 각 업종별로 먼저 절사추출법으로 표본조사가 구를 선정하기 때문에 각 층의 평균에 대한 추정량 \bar{y}_{hsy} 의 분산은 전수조사단위와 표본조사단위를 구분하여 표본조사단위에 대해서만 고려하였다. 추정량 \bar{y}_{hsy} 의 분산은

$$V(\bar{y}_{hsy}) = \frac{N_h - 1}{N_h} \frac{S_h^2}{n_h} (1 - \rho_h)$$

(단, N_h 는 h 층의 크기이고, n_h 는 h 층의 표본크기이다.)

의 형태이지만, 각 층의 급내 상관계수 ρ_h 를 계산하기가 복잡하므로 계산의 편의성을 위해 각 업종별 층화임의추출법과 절사추출법을 혼합하여 표본조사가구를 선정하는 것으로 가정하여, 모평균 추정량과 분산을 계산하였다. 여기에서 각 층별로 추출단위들의 핵심 보조변수의 크기 순으로 정렬한 후 표본조사가구를 계통추출할 경우에는 층화임의추출하는 경우보다 효율이 당연히 크므로, 요구정도를 충족하는 데에는 아무런 문제가 없었다. 각 업종별로 표본크기를 결정한 후에 전체의 표본크기가 6,000추출단위 이내(모집단의 20%이내)가 되도록 제한하였다. 각 업종별로 표본크기를 결정하는 방법은 다음과 같다.

총화임의추출법에서 허용상대오차가 주어졌을 경우, 일반적인 표본크기를 산출하는 공식을 알아보면 다음과 같다.

모평균의 추정량은 아래와 같다.

$$\widehat{Y}_{st} = \frac{1}{N} \sum_{h=1}^{N_h} N_h \widehat{Y}_h .$$

여기서 \widehat{Y}_h 는 h 번째 층의 표본평균이고, N_h 는 h 번째 층의 크기를 나타낸다.

\widehat{Y}_{st} 의 분산은 다음과 같이 표현할 수 있다.

$$V(\widehat{Y}_{st}) = \sum_{h=1}^{N_h} W_h^2 V(\widehat{Y}_h) .$$

여기서 $W_h = \frac{N_h}{N}$ 이고, $V(\widehat{Y}_h) = (\frac{1}{n_h} - \frac{1}{N_h}) S_h^2$ 이며 $S_h^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (Y_{hj} - \bar{Y}_h)^2$ 이다.

모평균 추정에서 $100(1 - \alpha)$ %의 신뢰수준에서 허용오차한계를 B 라고 한다면, 표본의 크기 n 은 다음과 같이 계산할 수 있다. 여기서 허용오차한계는 $B = z_{\alpha/2} \sqrt{V(\bar{Y}_{st})}$ 로 표현되고, $z_{\alpha/2}$ 는 표준정규분포에서 $100\left(1 - \frac{\alpha}{2}\right)\%$ 분위수이다. 각 층에 대한 표본배정은 Neyman의 최적배분법을 적용한다면, $n_h = n \frac{N_h S_h}{\sum N_h S_h}$ 로 각층의 배분 표본수를 계산할 수 있다.

$V(\bar{Y}_{st}) = \left(\frac{B}{z_{\alpha/2}}\right)^2$ 과 $n_h = n \frac{N_h S_h}{\sum N_h S_h}$ 를 이용하여 표본의 크기 n 을 구하면 다음과 같다.

$$n = \frac{\left(\sum_{h=1}^H N_h S_h\right)^2}{N^2 (B/z_{\alpha/2})^2 + \sum_{h=1}^H N_h S_h^2} \quad (3.1)$$

$B/z_{\alpha/2}$ 를 변이계수 CV 로 나타내면, $\frac{B}{\bar{Y}_{st} z_{\alpha/2}} = \frac{\sqrt{V(\bar{Y}_{st})}}{\bar{Y}_{st}}$ = CV 가 성립한다.

허용 오차한계 B 가 상대표준오차(변동계수)인 CV_0 로 주어졌다면, $B/z_{\alpha/2} = (CV_0) \bar{Y}_{st}$ 로 표현되고, 이것을 식(3.1)에 대입하면 다음의 (3.2)식을 얻을 수 있다.

$$n = \frac{\left(\sum_{h=1}^H N_h S_h\right)^2}{N^2 (CV_0 \cdot \bar{Y}_{st})^2 + \sum_{h=1}^H N_h S_h^2} \quad (3.2)$$

모든 업종에 대해 동일한 CV_0 를 적용하기 위해 표본의 크기를 계산하여 보았으나, 모집단의 CV 가 가장 큰 별목업과 가장 작은 밤나무 재배업에서 3배 정도의 CV 차이가 있으므로 단일 허용 상대오차를 적용할 수 없다고 판단하였다. 결국 표본크기가 모집단 크기의 약 20 %정도가 되도록 업종에 따라 CV_0 를 조정해서 계산한 표본의 크기는 <표 3.1>과 같았다.

<표 3.1 > 허용 상대오차와 표본크기

업 종	모집단 크기	모평균	표준편차	CV_0	n
별 목 업	1,091	11.35	66.75	0.120	216
송이채취업	3,277	49.91	118.09	0.071	658
밤나무 재배업	17,271	871.58	1,892.56	0.031	3,352
호두나무 재배업	796	183.05	405.39	0.150	157
대추나무 재배업	4,979	345.72	950.67	0.048	993
잣나무 재배업	1,935	7,931.63	17,665.76	0.080	375
야생화 재배업	258	5,865.97	32,278.74	0.300	70
제 재 업	1,667	1,543.68	6,636.92	0.150	314
합 계	31,274				6,135

<표 3.1>에 주어진 표준편차는 각 층에서 핵심변수 값이 큰 추출단위도 전수조사단위로 고려하지 않았을 때 계산한 값이다. 전주조사단위의 결정은 모든 업종에 대해서 ‘모평균 + 표준편차’보다 더 큰 핵심변수 값을 갖는 추출단위를 무조건 전수조사단위로 선정하였다. 특히 모집단의 특성상 층내에 개인과 법인이 혼합되어 큰 이상값이 적지 않은 송이 채취업, 밤나무 재배업과 대추나무 재배업의 경우 절사층의 크기를 표본 조사층의 크기와 비슷한 수준으로 조정한다면, 모평균 추정량의 표준편차는 상당히 감소할 것이다. 감소된 표준편차를 사용하여 식(3.2)로 계산된 허용상대오차와 표본의 크기는 <표 3.2>에 주어졌다. 물론, 반복적인 계산과정을 적용하여 안정적인 업종별 표본크기를 계산하기 위해서, 실제 계산에서 <표 3.2>의 표준편차와 식(3.2)에 의한 표본크기를 초기 값으로 생각하여, 1회 더 계산한 표본크기와 2회 더 계산한 표본크기는 유의할만한 차이를 보이지 않았다. 따라서 계산상의 편의를 위해 1회만 더 계산과정을 실행하여 구해진 표본크기를 업종별 최종 표본크기로 결정함을 원칙으로 하였다. 그 결과는 <표 3.2>에 주어졌다. 여기에서 n_c 는 전수조사층의 갯수, n_s 는 표본조사층의 갯수이다.

<표 3.2> 최종 허용 상대오차와 표본크기

업 종	모집단 크기	모평균	표준편차	CV_0	n_s	n_c	n
별 목 업	1,091	11.35	7.97	0.043	179	37	216
송이채취업	3,277	49.91	29.82	0.024	494	164	658
밤나무 재배업	17,271	871.58	493.87	0.015	1,314	782	2,096
호두나무 재배업	796	183.05	86.93	0.049	94	63	157
대추나무 재배업	4,979	345.72	224.19	0.020	857	136	993
잣나무 재배업	1,935	7,931.63	4,861.27	0.033	280	95	375
야생화 재배업	258	5,865.97	2,811.05	0.050	56	14	70
제재업	1,667	1,543.68	1,179.01	0.040	281	33	314
합 계	31,274				3,555	1,324	4,879

3.3 표본 배분

(1) 층별 표본 배분

광역시로 구성된 층을 제외하면 경기도, 강원도 등의 나머지 층에서는 조사단위 당 조사비용이 유사할 것이므로, 조사단위 당 조사비용이 일정하다고 가정하더라도 큰 무리는 없을 것이다. <표 3.2>에 주어진 업종별로 결정된 표본의 크기를 각 층별로 적절하게 배정하기 위해 각 층의 표준편차와 층의 크기에 비례하도록 배분함으로써 $Var(\bar{Y}_{st})$ 를 최소화하는 Neyman의 최적배분(Optimum Allocation)을 사용하였다. h 층에 배정할 표본의 크기 n_h 는 다음과 같이 표현할 수 있다.

$$n_h = n \cdot \frac{N_h S_h}{\sum N_h S_h} \quad (3.3)$$

식(3.3)은 다음과 같이 쉽게 증명될 수 있다. $n_h \propto N_h S_h$ 이므로 $n_h = k N_h S_h$ 라 놓고 $\sum_{h=1}^H n_h = n$ 을 이용하여 k 값을 구하면 $k = n / \sum N_h S_h$ 이다. 식(3.3)에 의해 업종별로 각 층에 배정된 표본의 크기는 <표 3.3>에 주어졌다.

<표 3.3> 업종별 각 층에 배정된 표본 수

(표본의 크기/표본조사층의 갯수)

층	별목업	송이 채취업	밤나무 재배업	호두나무 재배업	대추나무 재배업	잣나무 재배업	야생화 재배업	제재업
광역시	9/6	3/2	52/26	8/4	17/12	57/40	2/2	143/119
경기도	16/15	4/2	62/31	3/2	9/5	209/154	7/6	74/70
강원도	32/25	208/179	10/5	2/2	4/3	13/7	9/7	9/9
충청북도	24/19	21/16	64/32	22/15	89/69	56/44	2/2	8/8
충청남도	40/39	6/3	405/239	53/35	26/21	4/4	7/4	5/4
전라북도	30/28	30/27	111/64	46/23	55/49	16/14	7/7	28/25
전라남도	42/29	7/6	496/314	· 3/2	12/7	2/2	21/17	9/9
경상북도	15/14	326/223	24/16	17/9	614/535	9/8	8/7	18/18
경상남도	8/4	53/36	872/587	3/2	167/156	9/7	7/4	20/19
합 계	216/179	658/494	2,096/1,314	157/94	993/857	375/280	70/56	314/281

(2) 업종별 표본배정

업종별로 각 층에 대해서 전수조사 그룹과 표본조사 그룹으로 구분하는 가장 일반적인 방법은 충화변수의 값이 $\bar{Y}_{st} + \sqrt{V(\bar{Y}_{st})}$ 보다 더 큰 조사단위들을 모집단 전체에 대해 조사하여 각 층의 전수조사 그룹으로, 나머지 조사단위들은 표본조사 그룹으로 나누는 방법이다.

각 업종별 전수조사단위 구분에 적용된 절사기준(Cut-Off Criterion : $\bar{Y}_{st} + \sqrt{V(\bar{Y}_{st})}$)을 명시하면 다음과 같다. 별목업은 “별목면적 50.0ha”, 송이 채취업은 “송이 채취량 170kg 이상”, 밤나무 재배업은 “재배본수 2,700 본 이상”, 호두나무 재배업은 “재배본수 500 본 이상”, 대추나무 재배업은 “재배본수 1,300 본 이상”, 잣나무 재배업은 “재배본수 25,000 본 이상”, 야생화 재배업은 “재배면적 20,000평 이상”, 제재업은 “생산액 8,000만원 이상”에 해당하는 각 업종별 조사단위가 전수조사단위로 구분되었다.

표본추출은 전수조사 그룹내의 조사단위는 무조건 표본조사대상에 포함시키고, 나머지 표본조사 그룹에서는 계통추출법을 적용하기 위해서, h 층에서의 추출간격 $I_h (= N_h / n_h)$ 를 계산하고, 0에서 I_h 사이에서 임의의 난수 i_{h0} 를 뽑아내어 $i_{h1} = i_{h0} + 1$ 로 하여 $[i_{h1}]$ 에 해당하는 조사단위를 표본으로 선정하며, 그 이후부터는 계속해서 $[i_{h1} + I_h]$, $[i_{h1} + 2I_h]$, ..., $[i_{h1} + (n_{hs} - 1)I_h]$ 에 해당하는 조사단위를 표본조사단위로 선정하였다. 이때 $[x]$ 는 가우스 정수로 x 보다 작은 정수 중 가장 큰 정수이다. 이와 같이 충화계통추출법과 절사추출법을 결합하여 수정된 충화절사추

출법을 제안하였다.

구체적으로 벌목업의 “광역시” 층에 대해서 전수조사단위 그룹과 표본조사단위 그룹으로 구분하는 세부절차를 언급하기로 하겠다. 표본의 추출은 각 층에서 독립적으로 이루어졌다.

<표 3.4> 업종별 각 층의 크기와 배정된 표본의 크기

() : 추출률

구 분		광역시	경기	강원	충북	충남	전북	전남	경북	경남	합계
벌목업	총의크기	38	73	153	104	354	135	128	87	19	1,091
	전수	3	1	7	5	1	2	13	1	4	37
	표본	6	15	25	19	39	28	29	14	4	37
	계	9 (0.237)	16 (0.219)	32 (0.209)	24 (0.231)	40 (0.113)	30 (0.222)	42 (0.328)	15 (0.172)	8 (0.421)	216 (0.198)
송이 채취업	총의크기	10	4	1,270	184	7	216	37	1,324	225	3,277
	전수	1	2	29	5	3	3	1	103	17	164
	표본	1	2	179	16	3	27	6	223	36	494
	계	3 (0.300)	4 (1.000)	208 (0.164)	21 (0.114)	6 (0.857)	30 (0.139)	7 (0.189)	326 (0.246)	53 (0.236)	658 (0.201)
밤나무 재배업	총의크기	255	302	76	318	2,896	796	4,308	240	8,080	17,271
	전수	26	31	5	32	166	47	182	8	285	782
	표본	26	31	5	32	239	64	314	16	587	1,314
	계	52 (0.204)	62 (0.205)	10 (0.132)	64 (0.201)	405 (0.140)	111 (0.139)	496 (0.115)	24 (0.100)	872 (0.108)	2,096 (0.121)
호두 나무 재배업	총의크기	9	7	9	131	341	195	3	88	13	796
	전수	4	1	-	7	18	23	1	8	1	63
	표본	4	2	2	15	35	23	2	9	2	94
	계	8 (0.889)	3 (0.429)	2 (0.222)	22 (0.168)	53 (0.155)	46 (0.236)	3 (1.000)	17 (0.193)	3 (0.231)	157 (0.197)
대추 나무 재배업	총의크기	63	23	16	359	124	302	28	3,142	922	4,979
	전수	5	4	1	20	5	6	5	79	11	136
	표본	12	5	3	69	21	49	7	535	156	857
	계	17 (0.270)	9 (0.391)	4 (0.250)	89 (0.248)	26 (0.210)	55 (0.182)	12 (0.429)	614 (0.195)	167 (0.181)	993 (0.199)
잣나무 재배업	총의크기	273	1,039	56	305	37	98	7	60	60	1,935
	전수	17	55	6	12	-	2	-	1	2	95
	표본	40	154	7	44	4	14	2	8	7	280
	계	57 (0.209)	209 (0.201)	13 (0.232)	56 (0.184)	4 (0.108)	16 (0.163)	2 (0.286)	9 (0.150)	9 (0.150)	375 (0.194)
야생화 재배업	총의크기	13	54	19	4	22	26	73	16	31	258
	전수	-	1	2	-	3	-	4	1	3	14
	표본	2	6	7	2	4	7	17	7	4	56
	계	2 (0.154)	7 (0.130)	9 (0.474)	2 (0.500)	7 (0.318)	7 (0.269)	21 (0.288)	8 (0.500)	7 (0.226)	70 (0.271)

“광역시”의 총에서는 총 38개 단위 중에서 별목면적이 절사기준인 50.0 ha 이상인 3개 단위를 전수조사단위로 선정하였고, 나머지 35개 단위 중에서 6개 단위의 추출은 계통추출법을 적용하였다. 추출간격을 5.83으로 정하여 0에서 5.83 사이에서 하나의 난수를 발생시켜 그 난수에 1을 더한 난수의 가우스 정수, 그리고 1이 더해진 난수에 5.83을 더해 가면서 그 수의 가우스 정수에 해당하는 조사단위를 선정하였다. 예를 들면 발생시켜 1이 더해진 난수가 $i_1 = 2.30$ 이라면, 연속적으로 산출한 숫자들은 $i_2 = 8.13$, $i_3 = 13.96$, $i_4 = 19.79$, $i_5 = 25.62$, $i_6 = 31.45$ 이다. 따라서 표본단위로 선정된 단위들은 2, 8, 13, 19, 25, 31 번째 단위들이다.

별목업의 나머지 8개 층에 대해서도 같은 방법을 적용하여 전수조사단위와 표본조사단위를 선정하며, 업종별 각 층의 크기와 배정된 표본의 크기는 <표 3.4>와 <표 3.5>와 같이 주어졌다. 단, 제재업의 경우는 사업체의 소재지가 광역시에 편중되어 있기 때문에 <표 3.5>과 같이 광역시 층을 서울특별시와 부산, 대구, 인천, 광주, 대전, 울산광역시의 7개 층으로 세분화하여 층별 표본 크기를 배정하였다. 조사대상으로 선정된 조사단위들의 목록은 SAS 프로그램을 이용하여 산출하였다.

<표 3.5> 제재업의 층의 크기와 배정된 표본의 크기

구 분	총의 크기	전수조사	표본조사	표본의 크기	추출률
서울특별시	63	-	5	5	0.095
부산광역시	159	4	25	29	0.182
대구광역시	66	1	6	7	0.106
인천광역시	359	17	70	87	0.242
광주광역시	27	1	6	7	0.259
대전광역시	30	-	5	5	0.167
울산광역시	15	1	2	3	0.200
경기도	373	4	70	74	0.198
강원도	55	-	9	9	0.164
충청북도	43	-	8	8	0.186
충청남도	41	1	4	5	0.122
전라북도	139	3	25	28	0.201
전라남도	90	-	9	9	0.100
경상북도	87	-	18	18	0.207
경상남도	120	1	19	20	0.167
합 계	1,667	33	281	314	0.188

3.4 추정방법

각 업종별로 조사내용이 상이하므로 설문지 구성과 문항수가 다르다. 또한 각 문항에서 조사하는 자료의 형식이 질적인가 양적인가에 따라서 추정방법을 다르게 적용해야 하지만, 본 연구에서는 모든 업종의 각 문항에서 적용할 수 있는 일반적인 형식으로 모집단 총계의 추정량과 추정오

차에 대한 공식을 설명하겠다.

먼저, 모집단이 L 개 층으로 충화되었다면 각 층에서 층의 총계에 대한 추정량(\hat{Y}_h)의 합계로, 전체 모집단 총계 $\hat{Y} (= \sum_{h=1}^L \hat{Y}_h)$ 을 추정할 수 있다. 여기에서 \hat{Y}_h 를 전수조사부분과 표본조사부분으로 구분하여 $\hat{Y}_h = \hat{Y}_{hc} + \hat{Y}_{hs}$ 로 추정한다. 단, \hat{Y}_{hc} 는 전수조사부분에서 구해지고, \hat{Y}_{hs} 는 표본조사부분에서 추정한 값으로 $\hat{Y}_{hs} = \frac{N_{hs}}{n_{hs}} \sum_{k=1}^{n_{hs}} y_{hsk}$ 로 주어진다. 여기에서 n_{hs} 는 h 층의 표본의 크기이며, y_{hsk} 는 h 층의 표본의 k 번째 관찰값이다. 위의 식들을 종합하면 다음과 같이 표현할 수 있다.

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_{hc} + \sum_{h=1}^L \frac{N_{hs}}{n_{hs}} \sum_{k=1}^{n_{hs}} y_{hsk}, \quad (3.4)$$

식(3.4)에 주어진 모집단 총계의 추정량에 대한 추정오차는 표본조사부분에서만 발생하므로 아래 식과 같이 주어진다.

$$\begin{aligned} V(\hat{Y}) &= V\left(\sum_{h=1}^L \frac{N_{hs}}{n_{hs}} \sum_{k=1}^{n_{hs}} y_{hsk}\right) \\ &= \sum_{h=1}^L N_{hs}(N_{hs}-1) S_{hs}^2 \{1 - (n_{hs} - \rho_h)\} / n_{hs}, \end{aligned} \quad (3.5)$$

여기에서 S_{hs}^2 은 h 층의 표본조사부분의 모분산이고, ρ_h 는 h 층의 급내상관계수로서 동일한 계통표본내에 있는 조사단위들간의 상관계수이다. 실제 표본조사에서는 S_{hs}^2 과 ρ_h 를 알 수 없기 때문에 조사된 자료들로부터 추정해야 하지만 일반적인 선형계통추출법에서는 이들의 불편추정량을 계산할 수 없으므로 일반적으로 아래와 같은 분산의 추정량을 이용한다.

$$\begin{aligned} \hat{V}(\hat{Y}) &= \sum_{h=1}^L \hat{V}(\hat{Y}_{hs}) \\ &= \sum_{h=1}^L \frac{N_h^2}{n_{hs}(n_{hs}-1)} \sum_{k=1}^{n_{hs}} (y_{hsk} - \bar{y}_{hs})^2 \end{aligned} \quad (3.6)$$

예를 들어서 송이 채취량을 추정하고자 한다면 층의 수는 9개이고, 각 층의 크기와 각 층에서 전수조사부분과 표본조사부분의 크기는 실제 조사를 위해서 조사가능단위로 확인된 것을 말하며 다음 <표3.6>과 같다.

<표 3.6> 송이채취업 층별크기와 표본크기

층	광역시	경기도	강원도	충청북도	충청남도	전라북도	전라남도	경상북도	경상남도	합계
크기	10	4	1,264	183	3	209	37	1,313	224	3,247
전수부분	1	2	29	5	0	3	1	103	4	161
표본크기	2	2	179	16	3	27	6	223	36	494

강원도 층에서 강원도의 송이채취량의 추정값은 다음과 같다

$$\begin{aligned}\hat{Y}_3 &= \hat{Y}_{3c} + \hat{Y}_{3s} \\ &= \sum_{k=1}^{29} \hat{Y}_{3ck} + \frac{1235}{179} \sum_{k=1}^{179} y_{3sk}.\end{aligned}$$

강원도 송이채취량의 추정량의 분산은 아래 식으로 추정할 수 있다.

$$\begin{aligned}\widehat{V}(\hat{Y}_3) &= \widehat{V}(\hat{Y}_{3s}) \\ &= \frac{1235^2}{179 \times 178} \sum_{k=1}^{179} (y_{3sk} - \bar{y}_{3s})^2.\end{aligned}$$

송이채취업의 경우 다른 광역단체들에 대해서도 강원도와 같이 계산한 후에 이들을 식(3.4)와 식(3.6)에 대입하면 송이채취량 총합계에 대한 추정값과 분산의 추정값을 계산할 수 있다. 다른 업종에 대해서도 송이채취업과 같이 총합계와 분산의 추정값을 계산할 수 있을 것이다.

4. 표본의 관리 및 교체

4.1 표본관리

표본관리 방법에는 표본조사단위의 변경여부에 따라 크게 고정표본방식과 연동교체방식으로 나눌 수 있는데 조사된 자료의 안정성 제고와 조사원의 업무부담 및 무응답을 줄이기 위한 방법으로 표본조사단위의 고정관리 방법이 많이 사용된다. 추정오차 및 비용과 새로운 조사단위를 조사할 때 생기는 조사원의 업무량의 증가를 고려할 때 고정표본의 이용이 더욱 적절할 것으로 생각된다. 그러나 고정표본관리방식을 적용할 경우에는 비표본오차의 발생요인을 제어하여 추정값의 정확성을 높이기 위해서 아래사항들을 고려해야 할 것이다.

첫째, 고정관리되고 있는 표본조사단위가 업종변경 및 폐업, 또는 타 지역으로 전출 등의 사유로 표본에서 누락되는 경우, 비표본오차를 최소화시키기 위한 조사단위의 교체가 고려되어야 한다.

둘째, 모집단의 구조변화가 추후 연속조사에 반영되어야 한다. 즉, 업종변경 또는 폐업 등으로 조사단위의 소멸이 조사과정에서 확인될 경우 모집단의 구조에 반영되어 추후 연속조사에 활용될 수 있어야 한다.

셋째, 고정표본방식은 조사원과 응답자간의 긴밀한 친숙관계를 유지하기가 용이하기 때문에 응답자의 적극적인 협조를 받아 조사업무는 용이하게 수행할 수 있으나, 조사원들의 매너리즘으로 조사환경의 변화에 적응이 곤란하고, 조사원의 주관적인 판단으로 조사표를 기입할 수 있는 가능성이 있을 수 있다. 따라서 조사원에 대한 철저한 업무감독이 시행된다면 불성실한 조사태도로 인해 발생하는 비표본오차는 최대한 방지할 수 있을 것이다.

4.2 표본교체

표본조사과정에서 발생한 표본조사단위의 조사불가 사유를 살펴보면 업종변경, 폐업, 타지역 전출, 주거지 불분명, 기타(사망, 수몰지역)사유 등으로 나타난다. 조사과정에서 누락된 조사단위의

수가 클 경우 이를 무시한다면 표본오차가 증가하여 신뢰할 수 없는 결과를 야기할 수 있으므로, 조사의 정확도를 유지하기 위해서 조사단위를 재 선정하여 조사하는 표본교체법이 활용되어야 하며, 조사단위를 교체하는 방법을 설명하면 다음과 같다.

특정 업종의 h 층에서 총 20개의 표본조사단위 중 5개의 조사단위를 추출간격 4.0으로 계통추출한 표본조사단위를 $X_{h_3}, X_{h_7}, X_{h_{11}}, X_{h_{15}}, X_{h_{19}}$ 라 하자. 조사과정 중에서 X_{h_3} 의 조사단위가 사정에 의하여 조사가 곤란한 경우, 특성이 유사한 조사자료인 X_{h_2} 나 X_{h_4} 를 교체하여 조사한다. 만약, 과거의 조사자료를 활용할 수 있는 경우라면 X_{h_2} 또는 X_{h_4} 를 과거의 자료로 대체할 수 있다. 이러한 방법은 표본 대체법으로 많이 활용되고 있는 Cold-Deck Imputation 기법을 부분적으로 수용한 방법이다. h 층에서 전수조사단위 중 일부가 조사 불가할 경우에는 표본조사단위 중 가장 단위가 큰 유사조사단위로 교체할 수도 있다.

5. 결언 및 제언

1999년 임업총조사의 기초통계를 보완하여 임업 업종별 경영실태조사를 위해서 2000년에 5개 업종(벌목업, 송이채취업, 유실수재배업, 야생화재배업과 제재업)을 조사할 표본설계를 연구하였다. 조사모집단을 분석하여 지역별, 업종별로 총화하였고, 각 총내에서 임가의 규모가 다양하므로 규모가 큰 임가는 반드시 전수조사를 하도록 하고 나머지에서 표본조사임가를 선정하는 수정된 총화(계통)절사추출법을 적용한 표본설계를 제안하였다. 산림청의 통계조사 수행능력과 목표허용오차를 고려하여 표본크기를 산정하였고, 각 층별로 표본배분은 Neyman의 최적배분법을 적용하여 계산하였다. 그리고 제안한 표본추출법에 맞는 추정법과 추정오차 공식을 설명하였다.

앞으로 일반적인 표본설계와 임업 경영 관련 통계작성을 위해서 다음 두 가지 사항을 제안한다. 첫째는 임업경영실태조사와 같이 조사대상이 골고루 분포되어 있지 않고, 편중된 분포를 갖는 경우에는 다중추출틀(Multiple Frame)을 이용하는 통계조사방법의 적용을 제안한다. 다중추출틀을 이용한 통계조사를 적용하기 위해서는 이론적인 연구와 기존의 자료분석을 통한 실행 가능성 검토에 대한 연구가 선행되어야 할 것이며, 실제로 이 방법은 여러 통계 선진국에서 많이 이용하고 있는 통계조사 기법이다.

두 번째는 광역지방 자치단체들을 위한 통계작성을 염두에 두고 표본을 설계하고 추정기법([7], [8])에 대한 전문적인 연구를 정책적으로 진행해야 한다. 특히, 광역시와 각 도 단위의 임업경영 관련 통계는 지방자치단체에서는 필요하다. 이들 광역단체 단위의 통계를 작성할 수 있는 기법으로 사후총화 추정법([4])이나 소지역 추정법(Small Area Estimation)([8])이 있다. 미국, 캐나다 등에서 표본수가 적은 소지역의 통계작성에 활용하는 추정법이다. 현재 규모의 통계조사보다 약간만 더 조사단위 수를 증가시키고 소지역 추정법을 적용한다면 정책입안이나 정책실행의 평가자료로서 활용할 만한 수준의 각 도별 통계를 작성할 수 있을 것이다.

참고문헌

- [1] 산림청(1999), “『임업총조사』 지침서”.
- [2] 이계오 외 2인(2001), “절사계통추출법의 효율성에 관한 연구”, 응용통계연구, 14, 1, pp 111-120.

- [3] 통계청(1991), “절사법 표본설계 응용”, 통계청 연구자료(91-06-023).
- [4] Jinwoo Park(1999), "A Post-stratified Estimation in Multivariate Stratified Sampling Surveys", *The Korean Communication in Statistics*, Vol. 6, No. 3 pp 755 -760.
- [5] Leslie Kish(1965), "Survey Sampling", John Wiley & Sons Inc., New York.
- [6] Morris H Hansen, William N. Hurwitz, and William G. Madow(1993), "Sampling Survey Methods and Theory Vol I & II", John Wiley & Sons Inc., New York.
- [7] William G. Cochran(1977), "Sampling Techniques" 3rd. ed., John Wiley & Sons Inc., New York.
- [8] M. Ghosh and J.N.K. Rao(1994), "Small Area Estimation : An Appraisal", *Statistical Science*, Vol.9, No.1, 55-93.
- [9] M. P. Singh, J. Gambino and H. J. Mantel(1994), "Issues and Strategies for Small Area Data, *Survey Methodology*", Vol.20, No.1, 3-22.