

어가경제조사 표본설계에 관한 연구

김 규 성¹⁾, 전 종 우²⁾, 박 흥 래³⁾

요 약

어가경제조사는 우리나라 어가의 어가경제 및 어업경영에 관한 사항을 조사하는 통계조사이다. 본 연구는 어가경제통계를 생산할 수 있는 표본 설계를 제안한다. 연구의 기초 자료로 1990년 어업총조사 자료를 이용하였으며, 시도별 통계를 위하여 시도를 부차 모집단으로 하여 설계를 하였다. 표본은 층화 2단 추출을 하였으며, 층화를 위하여 어가소득의 소득함수를 추정하여 이용하였다. 어가의 소득은 선형 추정량을 이용하여 추정하였다.

1. 서 론

어가경제조사는 우리나라의 어가경제 및 어업경영의 실태를 파악하기 위한 통계 조사로 어가경제의 기본 사항인 어가소득, 어업소득, 기타 사업소득, 이전소득, 재산소득 등을 어가별, 어업의 생산 형태별로 조사하여 어촌경제의 구조변화를 파악한다. 조사는 일계부를 어가에 비치하여 어가의 수입과 지출을 어가가 직접 기장하는 일계부 조사와, 연 2회 조사원이 어가 재산을 직접 조사하는 원부 조사가 있는데 일계부 조사가 어가경제조사의 근간이 된다.

현행 표본설계는 '80년 어업총조사 및 '87년 어업기본통계 조사 자료를 바탕으로 작성되어 어가경제조사에 이용되고 있는데, 그 동안의 어업 환경 및 어장의 변동등 어촌의 경제 구조 변화에 따라 표본이 노후화되고 표본의 대표성이 많이 감소되었다. 또한 현재의 표본설계는 전국을 하나의 모집단으로 하여 설계되었기 때문에 시도 단위의 통계를 산출하기에는 어려움이 있다.

본 연구는 이같은 문제점을 전반적으로 검토하여, 표본의 대표성을 높이고 시도별 통계를 산출할 수 있는 표본설계를 제시한다. 표본의 대표성을 높이기 위하여 본 연구의 기초자료로 '90년 어업총조사 자료를 농림수산부로부터 제공받아 이용하였다. 또한 시도별 통계자료를 산출하기 위하여 각 도를 하나의 부차 모집단으로 하여 설계를 하였다. 본 조사의 여러 조사 항목중 중심이 되는 항목은 어가소득이다. 어가소득은 동일 어가에서도 해마다 소득의 편차가 심하고, 어가간의 소득 편차는 더욱 심하다. 따라서 시도별 추정을 위해서는 상당수 이상의 표본이 각 도에 배정되어야 하는데 현재의 표본수와 현지 조사 인력의 수를 감안하면, 시도별 표본오차를 줄이기 위하여 표본수를 과도하게 늘릴 수는 없다. 본 연구에서는 전국 어가소득의 표본오차를 3% 이내로 하고 각 도별 표본오차는 서로 비슷하게 되도록 설계를 하되, 조사 가능한 표본수 이내에서 최대한 표본오차를 줄이도록 하였다.

1) (130-743) 서울특별시 동대문구 전농동 90, 서울시립대학교 전산통계학과.
2) (151-742) 서울특별시 관악구 신림동 산 56-1, 서울대학교 계산통계학과.
3) (151-742) 서울특별시 관악구 신림동 산 56-1, 서울대학교 계산통계학과.

2. 모집단

2.1 모집단

본 연구의 모집단은 전국의 어가이며, 1990년 어업총조사 자료를 모집단의 기초 자료로 이용한다. <표 1>은 어업총조사 자료를 분석하여 얻은 모집단의 분포 형태이다. 여기에서 전업 어가는 어가소득이 모두 어업소득인 어가를 의미하며, 1종 겸업어가는 어업소득이 부업의 소득보다 많은 어가, 그리고 2종 겸업어가는 부업의 소득이 어업소득보다 많은 어가를 의미한다.

도별 어가의 분포를 보면 전남에 전국의 37.5%의 어가가 있고, 이어서 경남에 17.2%, 충남에 10.7% 순이다. 어가 형태별로 보면, 전업어가는 전체의 23.1%에 불과하고 45.7%가 1종 겸업어가이며 나머지 31.2%는 2종 겸업 어가이다. 어업 형태별로 보면 양식업을 하는 어가가 다수이고(41.0%) 동력선을 가진 어가가 29.8%, 무어선 어가는 26.8% 순이다.

<표 1> 전국 어가 분포 (전.겸업별, 어업 형태별 분포)

지역	전. 겸업별 어가수				어업 형태별 어가수			
	어가수(%)	전업	1종겸업	2종겸업	동력선	무동력선	양식업	어선비사용
부산	4901(4.0)	2639	1796	46	2346	281	1437	837
인천	2530(2.1)	632	950	948	750	28	455	1297
경기	6286(5.2)	1382	2592	2312	1071	224	2343	2648
강원	7208(5.9)	3426	3047	735	2565	324	62	4257
충남	12934(10.7)	2249	4997	5688	2809	95	7359	2671
전북	5623(4.6)	2228	2129	1266	1547	168	2005	1903
전남	45566(37.5)	6399	23384	15783	10086	563	28193	6724
경북	7451(6.1)	2703	3506	1242	2997	348	487	3691
경남	20843(17.2)	5070	10806	4967	10344	840	7310	2349
제주	8183(6.7)	1323	2303	4557	1710	74	76	6323
전국	121525	28051	55510	37964	36225	2945	49727	32628
(%)	(100.0)	(23.1)	(45.7)	(31.2)	(29.8)	(2.4)	(41.0)	(26.8)

2.2 추출틀과 조사 단위

1990년 어업총조사의 결과 전국에는 121,525가구의 어가가 있다. 어가는 시도, 시.군, 읍.면.동으로 분류되어 관리되고 있으며 읍.면.동은 다시 조사구로 세분되어 있다. 조사구에는 어업조사구와 인구주택조사구가 있는데 몇 개의 인구주택조사구를 묶은 것이 어업조사구이다. 다시 말해서 어가는 인구주택조사구로 묶여있고, 인구주택조사구 몇 개가 모여 어업조사구를 형성한다.

<표 2>는 전국 조사구 분포를 나타내고 있다. 여기에 나타난 조사구수는 어가가 있는 인구주택조사구에서 조사 불가능한 조사구를 제외한 조사구의 갯수이다. 본 연구에서 조사하는 항목은 소득에 관한 것으로 매일 조사를 해야 하기 때문에 섬 지역이나 교통이 불편해서 조사가 용이하지 않은 지역은 <표 2>에서 제외되었다. 조사 가능한 인구주택조사구는 전국에 8,540개가 있으며, 어업조사구는 2,377개가 있다.

<표 2> 전국 조사구 분포

지역	어업조사구			인구주택조사구	
	조사구수	평균어가수	평균인구주택 조사구수	조사구수	평균어가수
부산	166	29.5	6.7	1119	4.4
인천	98	25.8	6.1	596	4.2
경기	85	74.0	2.0	174	36.1
강원	155	46.5	3.9	607	11.9
충남	208	62.2	2.5	518	25.0
전북	111	50.7	3.6	395	14.2
전남	783	58.2	2.7	2098	21.7
경북	162	46.0	3.4	548	13.6
경남	484	43.1	3.6	1742	12.0
제주	125	65.5	5.9	743	11.0
	2377	51.1	3.6	8540	14.2

표본추출은 2단계로 이루어진다. 1차로 조사구를 뽑고, 뽑힌 조사구에서 2차로 어가를 추출한다. 1차로 추출할 조사구로 어업조사구와 인구주택조사구를 생각할 수 있는데, 전국 평균을 보면 3.6개의 인구주택조사구가 모여 하나의 어업조사구를 형성하므로 지역적으로 볼 때 어업조사구는 인구주택조사구보다 약 3.6배 넓다고 할 수 있다. 2차 추출단위인 어가는 1차로 추출된 조사구에서 랜덤하게 추출되므로 조사구의 면적이 넓을수록 표본어가는 더 산재하게 되고 따라서 조사원의 조사거리는 늘어나게 된다. 조사거리는 조사원의 업무량의 주요한 요소이므로 1차 추출단위로 인구주택조사구를 이용하는 것이 어업조사구를 이용하는 것보다 업무량이 더 적다. 따라서 어업조사구내의 어가 소득 분포와 인구주택조사구 내의 어가 소득 분포가 동일하다면 인구주택조사구를 1차 추출단위로 하는 것이 업무량의 측면에서 유리하다. 본 연구에서는 조사구내의 어가 소득 분포가 동일하다는 가정하에 인구주택조사구를 1차 추출 단위로 한다.

<표 3>를 보면 1어가만 포함하는 인구주택조사구수는 2,174개로 전체 조사구의 25.4%를 차지하며, 10어가 이하를 포함하는 조사구는 전체의 61.4%이다. 조사구수로는 60%를 넘지만, 이들 조사구가 포함하는 어가수는 전체 어가의 7.79%에 불과하다. 121,525 어가와 8,540개의 인구주택조사구를 모두 포함하여 모집단을 구성하는 것이 바람직하겠으나, 2단 표본 추출의 특성과 표본 관리의 어려움을 감안하여, 10어가 이하만을 포함하는 조사구는 모집단에서 제외한다. 10어가 이하의 조사구를 제외할 경우 전체 조사구수는 3,295개가 된다. 그리고 포함되는 어가수는 104,970 이다.

어가소득과 밀접한 관계가 있는 항목으로 어가 소유 선박과 양식업이 있다. 어가에서어업행위는 주로 배를 이용한 어업과 배를 이용하지 않은 어업행위, 그리고 양식장에서 행해지는 어업행위로 나눌 수 있다. 배를 이용하지 않은 어업보다는 배를 이용한 어업이 더 많은 소득을 올릴 것으로 기대된다. 그리고 대규모 양식업을 하는 어가는 다른 어가보다 더 많은 소득을

<표 3> 전국 어가수별 조사구 분포

어가수	조사구수(누적%)	어가수(누적%)	어가수	조사구수(누적%)	어가수(누적%)
1	2174(25.4)	2174 (1.79)	12	168(65.2)	2016 (16.68)
2	802(34.8)	1604 (3.11)	13	155(67.0)	2015 (18.33)
3	539(41.2)	1617 (4.44)	14	145(68.7)	2030 (20.00)
4	375(45.6)	1500 (5.67)	15	146(70.4)	2190 (21.81)
5	331(49.4)	1655 (7.04)	16	113(71.7)	1808 (23.29)
6	247(52.3)	1482 (8.26)	17	113(73.1)	1921 (24.87)
7	225(54.9)	1575 (9.55)	18	109(74.3)	1962 (26.49)
8	195(57.2)	1560 (10.83)	19	110(75.6)	1980 (28.12)
9	182(59.4)	1638 (12.18)	20	102(76.8)	2040 (29.80)
10	175(61.4)	1750 (13.62)	21이상	1980(100.0)	85314 (100.00)
11	154(63.2)	1694 (15.02)			
			합계	8540(100.0)	121525 (100.00)

올릴 것으로 예상된다. <표 4> 및 <표 5>에 조사구의 선박 보유 톤 수 및 양식장 규모 분포가 나타나 있다.

조사구의 선박 보유 분포를 보면 평균 2 톤 미만이 전체의 79.3% 를 차지하고, 5톤 이상은 5.5 %에 불과하다. 즉, 조사구의 평균 보유 톤 수는 2톤 이하가 대부분이며, 대규모로 선박을 보유한 조사구는 부산, 전남, 경북등 일부지역에 불과함을 알 수 있다. 양식장의 분포를 보면, 전체의 조사구중 34.2%는 양식장을 가지고 있지 않으며 38.9%가 1 ha 미만의 양식장을 가지고 있고, 5 ha 이상은 40 조사구에 불과하다.

<표 4> 인구주택조사구의 평균 소유 톤수 분포

규모	부산	인천	경기	강원	충남	전북	전남	경북	경남	제주	전국(%)
0 톤	0	29	32	13	109	30	77	10	19	69	388 (11.8)
0 - 1톤	15	10	57	84	146	48	424	128	190	159	1261 (38.2)
1 - 2톤	38	4	18	86	55	25	461	57	194	26	964 (29.3)
2 - 3톤	25	2	10	23	15	10	76	24	96	8	289 (8.8)
3 - 5톤	10	3	10	24	19	15	47	19	59	6	212 (6.4)
5 - 10톤	4	8	2	16	7	5	9	15	44	12	122 (3.7)
10 - 20톤	0		1	5	5	4	6	9	13	6	49 (1.5)
20 - 50톤	0						5	2			7 (0.2)
50톤 이상	2							1			3 (0.1)
합 계	94	56	130	251	356	137	1105	265	615	286	3295 (100.0)

<표 5> 인구주택조사구의 평균 양식장 규모 분포

규모	부산	인천	경기	강원	충남	전북	전남	경북	경남	제주	전국(%)
0 ha	19	36	50	208	70	37	140	165	152	25	1128 (34.2)
0 - 1 ha	25	19	43	42	193	34	424	99	367	3	1281 (38.9)
1 - 2 ha	20	1	14	1	56	23	282	1	78		476 (14.5)
2 - 3 ha	15		9		23	13	150		14		224 (6.8)
3 - 4 ha	8		5		8	18	56		4		99 (3.0)
4 - 5 ha	4		7		3	8	25				47 (1.4)
5 -10 ha	3		2		1	4	26				36 (1.1)
10 ha이상					2		2				4 (0.1)
합 계	94	56	130	251	356	137	1105	282	615	286	3295 (100.0)

3. 소득 분포 추정

1990년 어업총조사 자료에는 여러가지 항목이 조사되어 있으나 소득에 대한 항목은 조사되어 있지 않다. 본 연구에서 이용할 수 있는 소득에 대한 자료는 현행 표본자료이다.

3.1 현행 표본자료 분석

현행 표본은 1987년에 작성된 어업기본통계조사구 명부를 이용하여 전국 254개 어업기본통계조사구를 20개 층으로 나눈 후, 각 층에서 3개의 표본 조사구가 추출되었다. 조사구를 추출하는데 쓰인 추출 방법은 확률비례추출 방법으로 조사구내의 어가수에 비례하여 조사구를 추출하였다. 각 층에서 추출된 조사구에서 다시 표본어를 추출하였는데 추출된 어가수는 한 개의 표본 조사구에서 11-15 내외의 어가이며, 표본어가 추출 방법은 계통추출방법을 이용하였다. 20개 층의 층화 기준은 조사구의 어가 중에서 동력선을 가진 어가 비율, 무동력선을 가진 어가 비율, 양식업을 하는 어가 비율이고, 20개층의 모집단 어가수는 1,100-1,200 어가로 균등하게 분포되어 있으며, 표본 조사구내 어가수는 40여 가구 이상이 대부분이다.

현행 표본은 다음과 같은 특징을 갖는다.

가) 표본의 편의 (Bias) - 표본 조사구내의 어가수는 대부분 40가구를 넘어서 어가수가 상대적으로 적은 조사구의 상황을 반영하는데 미흡하다. 또한 표본이 대체로 톤수가 작은 조사를 중심으로 추출되어 있다.

나) 자료의 신뢰성 (Non-sampling error) : 조사항목이 소득이라는 점에서 조사값의 신뢰도는 다소 떨어진다. 다수의 어가에서 3개년('90,'91,'92) 소득간의 편차가 나타나고, 몇몇 어가에서는 어가의 특성 - 전업어가, 소유 배 톤수, 양식장 규모 - 과 소득이 무관하게 나타난다.

다) 표본 오차 - 표본자료의 표본오차가 다소 크게 나타난다. '92년의 경우 어가소득 - 7.8%, 어업소득 -15.4%, 어업외소득 - 7.3 % 로 계산되어 현행 표본설계를 그대로 이용하여 전국의 어가소득 표본오차를 3% 이내로 하기 위해서는 다수의 표본이 필요하다. 아래의 <표6>은 현행 표본의 시·도별, 어가 형태별 조사구 및 어가의 분포를 보여준다.

<표 6> 도별 표본수 및 어가 형태별 표본 어가수

지역	표본 조사구수	표본 어가수	전업어가	1종결업	2종결업
부산	2	25 (3.1 %)	2	22	1
인천	1	12 (1.5 %)	9	3	0
경기	5	68 (8.3 %)	9	30	29
강원	5	71 (8.7 %)	14	55	2
충남	6	83 (10.2 %)	3	55	25
전북	5	68 (8.3 %)	13	40	15
전남	18	242 (29.7 %)	6	151	85
경북	4	56 (6.9 %)	5	48	3
경남	10	135 (16.6 %)	12	107	16
제주	4	55 (6.7 %)	3	36	16
합계	60	815 (100.0 %)	76	547	192

3.2 소득 분포 추정

현행 표본 소득 자료에서는 해마다 소득의 편차가 크고, 이유가 뚜렷하지 않은 급속한 소득의 증가나 감소가 있는 어가가 발견된다. 815개 표본 어가의 소득을 연도별로 분류하여 지나치게 큰 소득의 증가와 감소가 있는 5개 어가는 본 연구의 대상에서 제외하였다.

추정치의 정도를 높이기 위해서는 세밀한 총화가 필요하다. 본 연구의 난점은 총화를 세밀하게 할 경우 표본 자료의 부족으로 총내의 평균 및 표준편차를 추정하기 어렵다는데 있다. 더

<어가소득과 총화변수간의 상관 관계>

어가소득	조사구내	조사구내	조사구내	조사구내	조사구내
	전업어가수	1종어가수	2종어가수	배 톤수	양식장규모
이상치 제거후	0.369	-0.207	-0.016	0.562	-0.224

군다나 도별 추정치를 위해서는 도별로 표본오차를 산출해야 하는데 현행 표본을 도 단위로 세분할 경우 추정치의 오차가 커져서 이용이 불가능하다. 현행표본 60개 조사구의 어가소득과 조사구내 전업어가 비율, 조사구의 평균 선박 톤수, 조사구내 양식업 어가 비율과의 상관관계가 다음에 나타나 있다.

배 톤수와 어가 소득과의 상관계수가 0.562로 가장 높게 나타나며, 비교 단위를 어가로 하면 상관계수는 더 커진다. 그러나 본 연구의 1차 추출단위가 인구주택조사구이기 때문에 조사구의 평균 소득과 배 톤수의 관계를 이용해야 한다. 양식장 규모와 소득의 상관계수가 음수로 나온 것은 양식업을 하는 어가수와 소득과의 상관계수를 구했기 때문이다. 현행표본에는 양식장의 규모가 기록된 것이 아니라 양식업 여부가 기록되어 있기 때문에 더 이상의 분석은 불가능하다. 어업 소득을 위한 총화 변수로서 조사구내의 평균 소유 톤수를 이용함이 타당하다. 양식장은 시설 규모에 따른 소득의 자료가 없기 때문에 주된 총화 변수로 이용하기는 부적절하다.

층을 나누기 위해서는 층을 나누었을 경우의 층내 평균과 분산을 알 수 있어야 한다. 60개 소득자료를 이용하여 평균 배 톤수와 이에 해당하는 소득의 평균과 표준편차를 구하여 톤수와 소득의 관계식을 구한다. 로그 변환한 평균과 표준편차를 평균 톤수와 회귀분석을 하여 다음의 결과를 얻는다.

$$\text{모형 1 : } \log(\text{mean}) = 6.958 + 0.100(t), \quad R^2 = 0.804 \quad (3.1)$$

$$\text{모형 2 : } \log(\text{std}) = 5.417 + 0.223(t), \quad R^2 = 0.972 \quad (3.2)$$

여기에서 t 는 조사구의 평균 선박 톤수이고, mean 은 조사구의 소득 평균 (단위 : 만원), 그리고 std 는 조사구의 소득 표준편차이다.

4. 새로운 표본 설계

4.1 표본 설계 선택

표본설계의 선택은 표본설계(Sample Design)의 선택과 동시에 모평균의 추정량을 고려해야 한다. 본 연구에서 추정할 대상은 전국 어가의 소득 평균과 각 도 어가의 소득평균이다. 먼저 각 도를 부차 모집단으로 하는 표본설계를 고려한다. 표본 추출 단위는 인구주택조사구가 1차 추출단위가 되며, 어가가 2차 추출 단위가 된다. 표본오차를 줄이기 위하여 각 도를 몇 개의 층으로 층화를 한다.

표본설계와 밀접한 관계를 갖는 것이 추정량이다. 본 연구에서 선택하는 추정량은 층화 표본 평균이다. 층화 표본 평균은 계산이 단순하고 표본의 소득 결과와 모집단의 조사구수 이외의 보조 변수는 이용하지 않기 때문에 관리가 쉽다.

4.2 층화

층화를 위해서는 적절한 층화변수가 필요하며, 층화의 효과를 극대화하기 위해서는 층화 분산을 최소화하도록 층을 나누어야 한다. 층화변수는 조사구내 선박 보유 톤수, 양식장 규모, 전업비율 등이다. 조사구내 선박 보유 톤수는 3.2절에서 구한 소득함수를 이용한다.

소득함수 및 양식장 규모, 전업비율을 이용하여 층화를 할 때, 층의 갯수 및 층의 경계점이 복합적으로 결정해야 할 문제이다. 층의 갯수가 많을수록 표본오차는 감소하나 표본수의 제한이 있으므로, 현실적으로는 층의 갯수를 무한히 늘릴 수는 없다. 또한 층을 세밀히 나눌 경우 층화변수의 효율이 문제되고 층의 갯수가 많으면, 층의 관리 및 표본의 관리가 복잡해진다.

각 층의 소득을 추정하기 위하여 (3.1)과 (3.2)에서 구한 소득함수를 이용하고, 양식장의 소득 평균 및 전업 비율에 따른 소득 평균을 구한다.

<표 7> 도별 총화 기준

지역	총 기준	조사구수	지역	총 기준	조사구수
경기	1. 인천:톤 ≥ 1.00	18	전남	4. 톤 ≥ 1.60	81
	2. 인천:나머지	38		5. 톤 ≥ 1.30	149
	3. 톤 ≥ 3.00	13		6. 톤 ≥ 1.10	140
	4. 톤 ≥ 1.00	28		7. 양식 ≥ 1.70	135
	5. 전업 $\geq .40$	49		8. 양식 ≥ 1.00	119
	6. 나머지	40		9. 전업 ≥ 0.30	151
강원	1. 톤 ≥ 6.00	17	경북	10. 나머지	187
	2. 톤 ≥ 3.50	22		1. 톤 ≥ 9.00	15
	3. 톤 ≥ 1.80	43		2. 톤 ≥ 5.00	13
	4. 톤 ≥ 1.20	48		3. 톤 ≥ 3.00	19
	5. 톤 $\geq .80$	45		4. 톤 ≥ 1.70	33
	6. 나머지	76		5. 톤 ≥ 1.00	47
충남	1. 톤 ≥ 5.00	12	경남	6. 전업 $\geq .90$	65
	2. 톤 ≥ 3.00	19		7. 나머지	73
	3. 톤 ≥ 1.60	30		1. 부산:톤 ≥ 2.50	26
	4. 톤 ≥ 1.10	29		2. 부산:나머지	68
	5. 양식 ≥ 1.60	35		3. 톤 ≥ 7.00	30
	6. 양식 ≥ 1.00	38		4. 톤 ≥ 4.00	45
	7. 전업 $\geq .60$	38		5. 톤 ≥ 2.70	62
	8. 전업 $\geq .10$	89		6. 톤 ≥ 2.00	75
	9. 나머지	66		7. 톤 ≥ 1.50	95
전북	1. 톤 ≥ 5.00	10	제주	8. 양식 $\geq .70$	84
	2. 톤 ≥ 3.00	15		9. 전업 $\geq .70$	103
	3. 톤 ≥ 1.80	18		10. 나머지	121
	4. 톤 ≥ 1.00	17		1. 톤 ≥ 3.50	24
	5. 전업 $\geq .80$	28		2. 톤 ≥ 1.00	34
	6. 나머지	49		3. 톤 $\geq .60$	31
전남	1. 톤 ≥ 4.50	26	4. 전업 $\geq .50$	33	
	2. 톤 ≥ 2.80	52	5. 전업 $\geq .25$	68	
	3. 톤 ≥ 2.00	65	6. 나머지	96	

총의 갯수는 최종적으로 60개로 하였다. 총을 60개로 한 이유는 표본수 120개를 염두에 두었을 경우의 총의 최대 갯수이다. <표 7>에 각 도별 총화 기준이 나타나 있다. 전국은 8개 도의

부차 모집단으로 구성되며, 각 도는 6-10개의 층으로 구성된다. 층화 기준을 보면, 경기도의 경우 인천직할시에서 조사구 평균 보유 톤수가 1 톤 이상인 조사구들로 층1 을 구성하고 인천의 나머지 조사구들로 층2 를 구성한다. 이때 층1에 해당하는 조사구수는 18개 이며, 층 2에 해당하는 조사구수는 38개 이다. 층3 은 인천을 제외한 경기도의 나머지 조사구 중에서 조사구 평균 3 톤 이상인 조사구로 구성하고, 여기에 해당되는 조사구는 13개 이다. 층 4는 1톤 이상 3 톤 미만의 조사구로 구성되며, 해당되는 조사구는 28개이다. 층5 는 경기도의 조사구중 층 1-4 에 해당하지 않는 조사구 중 전업비율이 40% 이상인 조사구로 구성되며 이에 해당하는 조사구 수는 49 개이다. 전업 비율은 조사구내의 어가 중에서 전업어가와 제1종 어가를 합한 수의 비율이다. 층 6은 경기도의 조사구중 층1 - 층5 를 제외한 나머지 조사구들로 구성한다. 선박 보유 톤수가 주된 층화 변수이며, 양식장이 많은 층남, 전남, 경남에는 양식장을 하나의 층으로 하여 층화를 하였다. 양식장의 단위는 ha 이다. 60개로 나눈 층에 표본배정을 한다. 표본을 2 개씩 배정했을 경우 전국의 표본오차는 2.67 % 가 예상되며, 각 도의 표본오차는 강원도를 제외하면 8% 이내가 된다.

<표 8> 목표 표본 오차별 표본수

지역	층수	n_1	cv_1	n_2	cv_2	n_3	cv_3	n_4	cv_4
경기	6	12	0.0773	12	0.0773	15	0.0687	19	0.0595
강원	6	12	0.0801	13	0.0764	15	0.0704	20	0.0606
충남	9	18	0.0703	18	0.0703	18	0.0703	24	0.0591
전북	6	12	0.0797	12	0.0797	15	0.0699	19	0.0597
전남	10	20	0.0695	20	0.0695	20	0.0695	27	0.0581
경북	7	14	0.0785	14	0.0785	17	0.0701	22	0.0609
경남	10	20	0.0705	20	0.0705	20	0.0705	28	0.0581
제주	6	12	0.0768	12	0.0768	14	0.0696	19	0.0593
전국	60	120	0.0267	121	0.0265	134	0.0247	178	0.0210

각 도의 목표 표본오차별 표본수가 <표 8>에 있다. 본 연구에서 제시한 여러 종류의 목표 오차별 표본수 중 전국의 표본수를 121개로 하는 것이 최종적으로 결정되었다. 이때 전국의 표본오차는 2.65%가 예상되며, 각 도는 8% 이내가 예상된다.

4.3 표본 추출

표본은 2단계로 추출된다. 전국에서 121개의 표본 조사구가 1차로 추출되고, 2차로 표본조사구에서 표본어가가 추출된다. 모집단의 각 층에서 표본은 단순임의 추출을 한다. 전국에서 121개의 표본 조사구를 추출하여 현지 출장소의 확인을 거쳐 조사 표본을 확정한다. 추출된 121개의 표본 조사구에서 어가를 추출한다. 표본어가는 표본조사구를 사전 조사하여 현재의 어촌 환경을 파악한 후 추출한다.

사전 조사는 다음과 같이 이루어진다. 121 개의 표본조사구에서 일차적으로 각 조사구의 어가를 점검한다. 새로 조사된 어가는 어가의 주소 및 어가의 요도를 새로 작성하여 각 표본조사구 별로 어가 목록을 확정하고, 각 어가에서 일년의 어가소득 및 선박 보유 톤수, 양식장 규모, 전업 여부등을 조사한다. 이 후에 표본조사구에서 조사된 각 어가의 자료를 토대로 표본 어가를 단순 임의 추출한다.

4.4 추정

평균 소득을 추정하는 추정량은 여러가지가 있다. 보편적으로 쓰이는 불편 추정량(Unbiased Estimator)을 비롯하여, 비 추정량, PPS 추정량 등이 있다. 일반적으로 쓰이는 불편추정량은 전국의 어가수 및 각 도의 어가수를 알아야 하는 단점이 있고, 비 추정량 및 PPS 추정량도 어가수 및 소득에 대한 보조 정보가 필요하다. 그러나 현실적으로 현재의 전국 어가수를 알기 어렵고, 더군다나 소득의 보조 정보는 더욱 알기 어렵다. 따라서 본 연구에서는 약간의 편의(Bias)는 포함하지만 표본 자료와 조사구수만을 가지고 결과를 산출할 수 있는 간단한 추정량을 제안한다.

$y_{A hij}$ 를 A 도 h 층 i 번째 표본 조사구의 j 번째 표본어가의 소득이라 할 때 $\bar{y}_{A hi}$ 는 i 번째 조사구의 표본어가 평균이고, $\bar{y}_{Ah} = \frac{1}{n_{Ah}} \sum_{i=1}^{n_{Ah}} \bar{y}_{A hi}$ 는 h 층내 표본조사구들의 평균값이다. 그러면 A 도의 평균어가 소득은 다음과 같이 추정할 수 있다.

$$\bar{y}_A = \sum_h \left(\frac{N_{Ah}}{N_A} \right) \bar{y}_{Ah} \quad (4.1)$$

여기에서 $N_A = \sum_{h=1}^L N_{Ah}$ 이고, N_{Ah} 는 A 도 h 층의 조사구 총수이다. 식 (4.1)의 분산은 다음과 같이 추정된다.

$$var(\bar{y}_A) = \sum_{h=1}^L \left(\frac{N_{Ah}}{N_A} \right)^2 var(\bar{y}_{Ah}) \quad (4.2)$$

$$var(\bar{y}_{Ah}) = \left(1 - \frac{n_{Ah}}{N_{Ah}}\right) \frac{s_i^2}{n_{Ah}} + \frac{1}{n_{Ah}^2} \sum_{i=1}^{n_{Ah}} \left(1 - \frac{m_{A hi}}{M_{A hi}}\right) \frac{s_{2i}^2}{m_{A hi}} \quad (4.3)$$

식 (4.3)에서 $M_{A hi}$ 는 h 층의 i 번째 표본조사구의 총 어가수이고, $m_{A hi}$ 는 h 층의 i 번째 표본조사구의 표본 어가수이다. 그리고 $s_i^2 = \sum_{i=1}^{n_{Ah}} (\bar{y}_{A hi} - \bar{y}_{Ah})^2 / (n_{Ah} - 1)$ 는 조사구 표본 분산이며, $s_{2i}^2 = \sum_{j=1}^{m_{A hi}} (y_{A hij} - \bar{y}_{A hi})^2 / (m_{A hi} - 1)$ 은 어가의 표본 분산이다. (4.1)과 (4.2)를 이용하여 A 도의 표본오차 추정량은 다음과 같이 구한다.

$$cv(\bar{y}_A) = \frac{\sqrt{var(\bar{y}_A)}}{\bar{y}_A} \quad (4.4)$$

전국 평균 어가소득은 각 도에서 구한 도 평균 추정량을 이용하여 (4.5)로 추정하고, 분산 추정은 (4.6)과 같이 한다. 그리고 전국 평균 어가소득의 표본오차 추정은 (4.7)로 한다.

$$\bar{y} = \sum_A \left(\frac{N_A}{N} \right) \bar{y}_A, \quad N = \sum_A N_A \quad (4.5)$$

$$var(\bar{y}) = \sum_A \left(\frac{N_A}{N} \right)^2 var(\bar{y}_A) \quad (4.6)$$

$$cv(\bar{y}) = \frac{\sqrt{var(\bar{y})}}{\bar{y}} \quad (4.7)$$

5. 표본 관리 및 표본 대체

소득 조사는 매일조사이며, 몇 해 동안 계속하는 연속조사이다. 모집단의 변화를 반영하기 위해서는 표본의 관리가 무엇보다 중요한데, 표본어가에서 어가소득에 관한 항목과 어가 특성을 반영할 수 있는 항목들, 즉, 어가의 선박 보유 톤수, 양식장 규모, 전업어가 여부, 표본 조사구의 전체 어가수 등을 조사하여 계속 관리하여야 한다.

모집단의 변동으로 인한 표본의 교체는 연속조사에서 불가피하다. 표본의 교체는 표본 조사구의 교체와 표본어의 교체가 있다. 표본어의 교체는 설계 당시에는 어가였으나 이후에 이전이나 전업 등으로 어가가 아닌 일반 가구가 되었을 때 행한다. 각 출장소에서는 표본 조사구에 표본어가 외에 다른 어가들의 목록을 가지고 있어야 한다. 교체 어가는 표본어가였다가 어가 환경의 변동으로 교체해야 하는 어가와 가장 유사한 어가, 즉 소득과 더불어 선박 톤수나, 양식장 규모 등이 비슷한 어가를 선출하도록 한다.

지역 개발 등으로 표본조사구 전체가 바뀌는 경우가 있다. 각 도는 6-10개의 층으로 층화가 되어 있고, 층내 조사구들은 소득이 유사한 조사구들로 구성되어 있다. 따라서 조사구 교체를 할 때에는 동일한 층에서 표본 조사구가 아닌 다른 조사구를 새로운 표본 조사구로 선정하면 된다.

표본조사는 표본오차 및 비표본오차, 연속조사에서 발생하는 시계열 오차 등을 내포하며, 이러한 오차를 최소화하는 조사를 하여야 한다. 표본의 대표성을 유지하고 오차의 한계를 계속 보장하기 위해서는 설계 이후에 모집단의 관리 및 표본의 신중한 대체가 절실히 요망된다.

6. 참 고 문 헌

- [1] 농림수산부 (1990). 「1990년 어업총조사 조사요령」.
- [2] 농림수산부 (1991). 「어가경제통계」.
- [3] 박홍래 (1989). 「통계조사론」, 영지문화사, 서울.
- [4] Cochran, W.G. (1977). *Sampling techniques*, John Wiley & Sons, New York.
- [5] Forman, E.K. (1991). *Survey sampling principles*, Marcel Dekker, New York.

A Study on the Sample Design of the Fishery Household Economy Survey

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

Abstract

The fishery household economy survey is a sample survey which produce estimates on the fishery household economy and fishery management in Korea. We propose a sample design for this survey. This design is developed based on results of 1990 fishery census and Shi-Do is assumed to be subpopulation for Shi-Do estimates. Samples are selected by stratified two-stage cluster sampling in Shi-Do and income function is found for stratification. Fishery household income is estimated by a linear estimator.

4) Department of Computer Science and Statistics, Seoul City University, 90 Jeonnong Dong, Dongdaemoon Ku, Seoul, 130-743, 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.