

## 가축통계 표본조사설계

윤기중<sup>1)</sup> 박상언<sup>2)</sup>

### 요약

1995년 가축통계 총조사자료에 기초하여 1998년부터의 향후 5년간의 표본설계를 하였다. 먼저 본 설계에서는 시도별 변이계수가 목표인 점을 고려하여 농가별 사육마리수를 전체사육마리수로 나눈 사육비율을 구한 다음 이를 전수기준과 총화에 이용하였다. 전수농가기준은 전수농가를 제외한 후 조사구의 축종별 사육비율의 분산이 일정하게 하는 기준을 정하였고, 총화는 네축종중 최대사육비율이 비슷한 조사구를 complete linkage 방법에 의해 총화하였다. 층의 개수는 총화에 따른 표본개수를 최소화하는 층의 수를 선정하였다. 또한 표본설계과정상의 해결하기 힘들었던 문제점을 열거하였다.

### 1. 서론

우리 나라 농촌은 급격히 변화하는 전환기에 처해 있다. 산업의 고도화로 농업은 상대적으로 쇠퇴하고 농촌 인구의 향도 이농 현상이 진행되는 과정에서 WTO시대를 맞아 모든 농산물 시장의 개방으로 농산품도 공산품과 같이 국제적인 경쟁시대를 맞고 있다. 정부는 이러한 전환기에 농업의 보호 육성을 위하여 오래 전부터 농어촌 후계자 육성과 지원에 많은 예산을 투입할 뿐만 아니라 농업의 기계화, 품종개량, 농업 경영 방법의 개선 지도, 작목 선정의 지도 그리고 농산품의 가공 산업 육성에도 적극 지원해 왔다. 농업은 농지경작업과 축산업 2대 분야로 나눌 수 있는데, 이 가운데 축산업은 여러 조건으로 보아 한국 축산업의 국제 경쟁력이 약한 편이다. 이런 이유로 축산업에서 주종을 이루는 한우, 젓소, 돼지 그리고 닭에 대한 가격 안정과 생산증진 그리고 국제경쟁력의 강화를 위해서 이들의 동태를 정확히 파악해야 함이 요구되고 있다. 더욱이 지방자치제에 따라 지방의 기초단체에서도 자발적으로 농업발전을 꾀하고 있어 축산업의 정확한 동태 파악을 절실히 요구하고 있다.

현행의 가축통계조사를 위한 표본은 1990년도 농업총조사 결과에 의해서 설계된 것이므로 그 표본이 노후화 되어 정도의 저하를 초래케 할 것으로 예상된다. 이러한 정도의 저하를 막고 보다 효과적인 표본에 의해서 조사되어야 함이 요구되고 있어 1995년도 농업총조사에서 얻은 축산농가와 축종별 사육현황자료를 토대로 정도가 높은 표본조사를 위하여 새로운 표본설계를 하여야 한다. 1995년 총 조사에서 얻은 전국의 축종별 가축통계에 의해서 모집단 규모와 조사구 그리고 조사구별 특성을 파악하였다. 각 조사구에서 전수조사 대상과 표본조사 모집단을 구분해야 하는데 그 방법은 광역시와 도별 축종별로 표본조사 대상 조사구의 분산을 같게 하는 방법으로 구획하였다. 또 표본 추출을 위해 조사구를 사육마리수가 비슷하게 총화하고 표본수를 최소화할 수 있는 층의 수를 결정하였다. 시도별 표본수는 시별 변이계수는 5%, 도별 변이계수는 3%를 넘지 않도록 결정하였다.

1) (120-749) 서울 서대문구 신촌동 134, 연세대학교 상경대학 응용통계학과, 명예교수

2) (120-749) 서울 서대문구 신촌동 134, 연세대학교 상경대학 응용통계학과, 조교수

## 2. 1995년 총 조사자료에 대한 축종별 사육현황

다음에 요약할 모집단자료는 1995년 농업총조사 자료에 기초하고 총 조사이후 경상남도 울산시가 광역시로 승격됨으로 울산시를 경상남도에서 분리한 전국 16개 시도(서울, 6개 광역시, 9개도)의 가축사육농가를 모집단으로 한다.

### 2.1. 축종별 가축사육 농가수 및 마리수

축종별 사육농가수는 모두 감소하였고 특히 돼지, 닭의 경우는 절반 수준으로 감소하였다. 따라서 농가당 축종별 사육마리수는 대폭 증가하였다. 즉 사육형태가 대규모화된다는 것을 의미한다.

표 2.1: 축종별 사육호수 및 사육마리수의 비교

	한우		젖소		돼지		닭	
	호수	마리수	호수	마리수	호수	마리수	호수	마리수
서울	30	557	14	374	27	2290	56	54212
부산	980	3724	183	3437	229	30260	449	259660
대구	3522	25554	296	6481	219	38725	1021	425686
인천	2236	21082	394	9231	1625	111915	2489	966709
광주	1292	7059	121	1749	112	14763	503	306520
대전	1551	5814	56	1203	49	8283	620	184053
울산	5487	32448	92	2160	182	32156	537	498317
경기	24686	223970	9249	206939	5343	1524126	9266	27752614
강원	38132	172979	1040	23025	1061	227889	12714	4027384
충북	38645	168051	1328	28301	1083	2668817	6934	5142424
충남	68093	377334	3539	77082	12022	1093070	11208	12283110
전북	41805	204249	1369	36682	5830	525196	8665	11001695
전남	99035	441722	1472	34809	8832	511797	15936	7736577
경북	99577	504469	1718	42934	3298	691667	17734	12803624
경남	86568	336114	1536	38911	6233	722739	12806	6733321
제주	3299	29185	120	3920	370	203118	233	923788
합	514938	2554311	22527	517238	46515	6006811	101171	911099694

### 2.2. 농가별 사육마리 분포

농가별 사육마리를 시도별로 boxplot에 의해서 표시해 보았다. X축의 숫자는 서울, 6개 광역시와 9개 도를 지역코드로 표시했고 Y축은 사육 마리 수를 표시한 것이다. 이를 살펴

보면 한우를 제외한 나머지 축종에 대해서는 대규모 사육 농가가 많고 또한 시도별로 대규모 사육규모가 상당히 상이함을 알 수 있다.

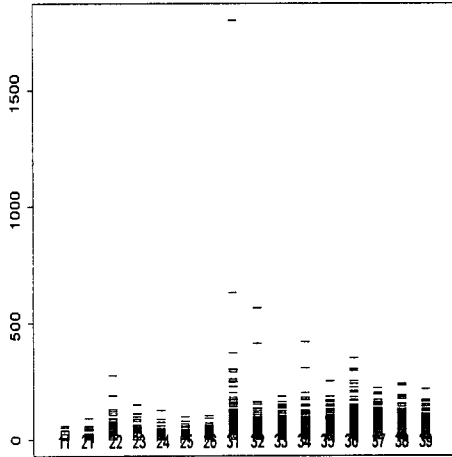


그림 2.1: 한우 사육마리수

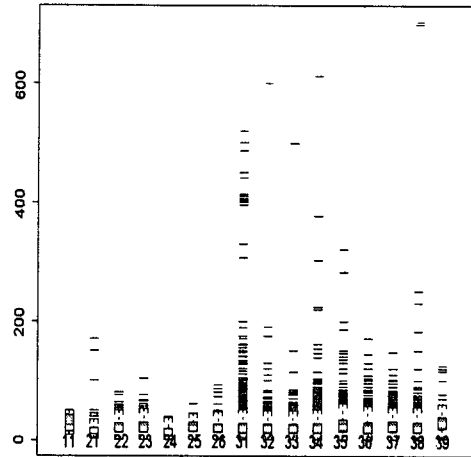


그림 2.2: 젖소 사육마리수

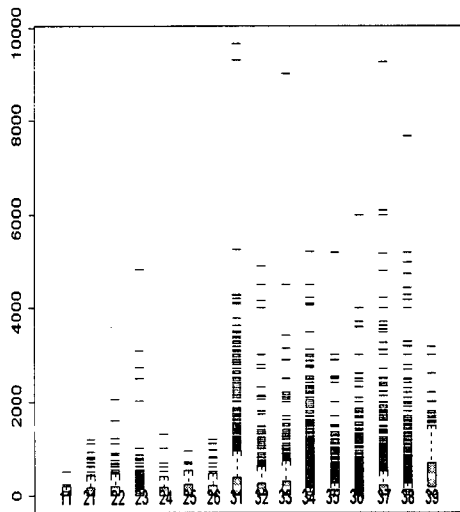


그림 2.3: 돼지 사육마리수

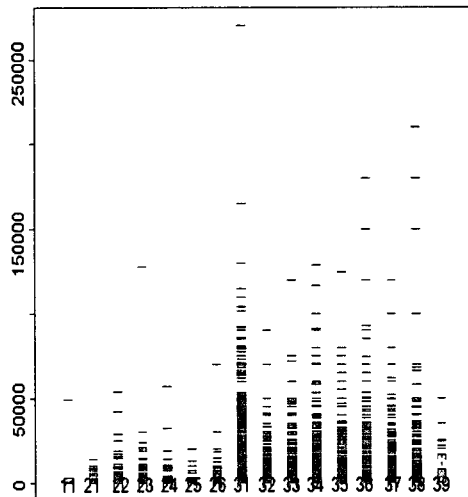


그림 2.4: 닭 사육마리수

### 3. 과거표본설계와 본 설계의 설계목표

가축통계표본조사는 1980년, 1990년 총 조사에 기초하여 두 번의 표본설계가 행해졌고 현재 실시중인 표본조사는 1990년 총 조사자료에 따른 표본설계에 의해 이루어지고 있다. 두 번의 표본설계는 일정규모 이상을 사육하는 농가의 사육마리를 조사하는 전수조사와 전수조사농가를 제외한 농가를 조사구로 묶어 그 중 표본을 추출하여 조사하는 표본조사로 구성되어 있다. 전수조사는 설계시 기준이 설정되고 표본조사는 설계시 조사구가 설정된다. 즉 전수조사와 표본조사는 그 조사단위가 다르고 또한 조사방법 역시 다르다. 전수조사는 설계시 설정된 기준이상을 사육하는 농가를 매 조사시 신고 등의 간접적 방법에 의해 파악하여 해당축종만의 사육마리수를 조사하는 것이고 표본조사는 설계시 설정된 해당조사구를 직접 방문하여 조사구내 모든 축종의 사육마리수를 파악하는 것이다. 따라서 표본설계에서는 전수농가의 기준과 표본조사구의 개수를 설정하는 것이 가장 중요한 부분이라고 할 수 있다.

#### 3.1. 과거의 표본설계

1980년, 1990년 총 조사에 따른 두 번의 표본설계에서 전수기준과 표본추출단위는 모두 6개시 및 9개 도로 같으나 표본조사구 추출방법은 1980년에는 단순임의추출을, 1990년에는 층화추출을 하였다. 또한 1980년의 표본추출조사구는 농업조사구이며 1990년의 경우는 농업조사구가 세분화된 단위인 인구주택조사구이다. 1990년의 표본설계는 1980년의 표본설계보다 전수조사농가수를 대폭 줄였으나 표본조사에 따른 업무량은 비슷하게 유지되어 전체적으로 업무량이 많이 감소하였다.

#### 3.2. 제약조건

본 논문에서의 표본설계는 전수기준은 시도단위의 농가, 표본추출단위는 시도단위의 인구주택조사구, 표본추출방법은 층화추출법, 표본수는 네 축종의 변이계수가 서울 및 6개 광역시에서는 5%, 9개 도에서는 3%내가 되도록 하는 제약조건을 따르도록 되어있다. 따라서 설계내역의 주요부분이 이미 지침으로 정해져 있어 지침 내에서의 효율적인 전수기준과 층화에 따른 표본추출만이 정해져야할 부분이라고 할 수 있다. 또한 현행 업무량 내에서 조사 가능한 표본설계라는 현실적인 지침으로 인해 전수기준도 현행에서 크게 벗어나지 못하고, 시도별 표본수 역시 변화한 모집단 특성에 맞게 할당하는 것이 제한되어 있을 뿐만 아니라 보다 큰 문제는 표본추출단위는 시도이고 업무량단위는 시군이어서 시군 표본수가 제한되는 해결하기 힘든 표본설계를 해야한다는 것이다.

### 4. 본 설계내역

본 설계는 1995년 총 조사자료에 기초한 것으로 모집단은 울산광역시를 따로 분리한 16개 시도 가축사육농가이다. 전체적인 설계내역은 앞서 밝힌 제약조건으로 인해 현행설계에서 크게 벗어나지 않았다. 표본추출의 효율성을 위해 시도별 사육마리수에 따라 조사

구들을 1차 층화만 한다. 표본에는 임의표본과 유의표본이 있으며 임의표본은 각 층에 해당하는 조사구들 중 일부만을 표본 추출하여 조사하여 추계하는 것이고 유의표본은 사육마리수가 일정규모이상인 층에서는 층내 분산이 커져 층내 조사구를 모두 조사해야 하므로 이러한 층들을 하나의 별도 층으로 묶어 해당조사구 전부를 모두 조사한다. 따라서 추계는 임의표본에 의한 추계치에 유의표본조사치와 전수농가조사치를 합하게 된다.

#### 4.1. 축종별 전수기준 및 층화기준

표본설계에서 일부 축종에 대해 층화하여 표본설계를 하게되면 축종들에 대한 변이계수가 비슷하게 유지되지 않게 된다. 즉 그 중에서 가장 큰 변이계수를 목표 변이계수내로 하게되면 지나치게 많은 표본수가 필요하고 다른 축종의 변이계수는 매우 낮게 된다. 또한 전수기준 역시 일부 축종에 대해 지나치게 낮게 하면 표본추출대상인 조사구의 축종별 분산이 다르게 되어 전수기준이 상대적으로 높은 축종에 의해 표본수가 결정되게 되는 문제점이 있다.

이러한 문제점을 극복하기 위해서는 먼저 단위수가 크게 다른 네 축종의 사육마리수를 표준화하여야 하는데 이를 위해 농가별 사육마리수를 전체사육마리수로 나누고 이러한 표준화된 단위인 사육비율을 전수기준 및 층화기준에 사용하였다. 이는 변이계수는 이러한 사육비율의 표준편차와 같기 때문이다.

전수기준 : 전수 규모는 낮을수록 좋으나 그로 인해 업무량이 증가하게 된다. 전수기준이 일부 축종에 치우치게 되면 조사구의 축종별 사육비율의 분산의 차이가 심하게 되어 표본수가 필요이상으로 커지게 된다. 따라서 전수기준은 전수농가를 제외한 조사구들의 축종별 사육비율의 분산이 비슷하게 하는 기준을 정하였다. 물론 표본추출이 층화추출이어 층화까지 고려한 분산을 계산해야하나 이러한 계산이 계산시간상의 이유로 사실상 불가능함을 밝혀둔다. 네 축종의 일정한 분산을 너무 낮게 하면 전수농가수가 많아지므로 전수농가수에 따른 일정한 분산기준은 현행전수농가수를 기준하였다. 따라서 전남의 경우처럼 현행 설계에서 타 시도에 비해 전수 농가수가 상당히 작으나 이를 대폭 증가시키지는 않았다.

층화기준 : 전수기준이 설정된 후 조사구들을 축종별 사육비율이 비슷한 조사구끼리 층을 만들어야 한다. 이에 대한 층화방법은 complete linkage 방법을 이용하였다. 물론 K-means 방법이 좀 더 적합한 것 같으나 시간이 많이 걸리는 이유로 시간이 적게 걸리는 complete linkage 방법으로 층화하였다. 이러한 층화방법의 층화 key로는 네 축종의 사육비율을 동시에 고려해야 하지만 본 설계에서는 조사구내 네 축종의 사육 비율중 최대값 하나를 층화 key로 사용하였다. 그 이유는 각 층별 기준을 요구하는 지침을 만족시켜주기 위해서이다. 즉 표 4.2과 같은 시도별 층화기준표를 작성하기 위해서는 어쩔 수가 없었던 점을 밝혀둔다.

층의 개수 결정 : 층의 수를 늘려가며 필요한 표본수를 Neyman 할당법에 따라 계산한 후 그 중 표본수가 가장 작은 층의 개수를 선택하였다. 층내 조사구 수보다 많은 표본을 필

요로 하는 층들을 한데 묶어 유의표본층으로 구분하였다. 대부분 도의 경우 최소의 표본수를 갖는 층의 개수가 20을 넘어서게 되는데 이 경우 층의 개수를 20을 넘지 않게 임의로 제어하였다. 시간의 변화에 따른 층내 변동에 민감하지 않기 위해 층의 수를 가급적 작게 하려 하였으나 이로 인해 시도 표본수가 커지고 따라서 실제 업무 단위인 시군 표본이 현행보다 많게 될 우려가 있어 시도표본수가 대폭 감소하는 20개로 설정할 수밖에 없었던 현실적인 점을 밝혀둔다.

#### 4.2. 표본설계결과

전수조사대상규모 및 1995년 현재 전수조사대상 농가수 : 표본결과의 전수기준은 표 4.1와 같고 젓소의 경우 현행보다 낮게 설정되었고 다른 축종의 경우 현행보다 높게 설정되었다.

표 4.1: 축종별 전수규모 비교

	현행설계				본설계			
	한우	젓소	돼지	닭	한우	젓소	돼지	닭
서울	40	40	100	5000	30	30	100	3000
부산	30	30	100	5000	20	25	300	2500
대구	40	20	100	7000	40	40	250	2000
인천	20	40	300	7000	50	50	500	6000
광주	30	30	100	10000	30	25	300	4000
대전	30	30	100	7000	35	25	200	6000
울산					50	25	300	4000
경기	60	60	500	10000	60	60	800	25000
강원	50	30	100	7000	100	25	500	12000
충북	50	30	300	7000	150	50	500	10000
충남	60	60	300	7000	300	55	800	10000
전북	60	60	300	7000	200	45	300	21000
전남	60	60	300	7000	300	40	550	20000
경북	60	60	300	7000	200	40	850	8000
경남	60	60	300	7000	130	40	650	8000
제주	50	40	300	7000	150	20	1100	5500

시도별 층화기준 및 표본수 : 조사구의 층화기준은 네 축종의 사육비율중 최대값에 따라 층화하였고 이를 다시 사육마리수로 역산하면 층별 기준을 알기 쉽게 나타낼 수 있다. 여기서는 서울의 경우만 예를 들어 설명한다. 표 4.2에는 각 층과 층에 따른 축종별 수가 표시되어 있는데 해당축종의 사육마리수가 해당 층의 수보다 같거나 크게되면 해당 층에 속

하게 된다. 따라서 하나의 조사구에 대해 네 축종의 사육마리수에 따라 각각 4개의 층이 나오게 되는데 그러면 그 중 최대 층이 해당조사구의 층이 된다. 예를 들어 만일 서울의 조사구내 사육마리수가 한우 3마리, 젓소 5마리, 돼지 100마리, 닭 1000마리이면 한우는 1층의 기준인 0보다 크고 2층의 기준인 7보다 작으므로 1층, 젓소는 2층의 기준인 5와 같고 3층의 기준인 12보다 작으므로 2층, 같은 방식으로 돼지는 3층, 닭은 2층이 된다. 따라서 해당 조사구의 층은 3층이 된다. 표 3에서 3층에서는 8개 조사구중 7개를 표본조사하여야 하는데 이러한 경우는 1개를 추가하여 8개 모두를 유의표본 처리하는게 더 현실적일수도 있으나 표본조사구를 줄인다는 지침에 따라 그대로 두기로 하였음을 밝혀둔다.

표 4.2: 서울 층 기준 및 표본크기

층	한우	젓소	돼지	닭	조사구	표본
1	0	0	0	0	45	4
2	7	5	30	500	11	6
3	17	12	70	1600	8	7
유의	35	25	150	3000	4	4
계					68	21

### 4.3. 표본 설계 결과비교

표 4.3의 표본설계에 의한 표본크기를 보면 서울, 6개 광역시중 조사구수가 대폭 증가된 시는 표본크기가 증가하고 그러하지 않은 시는 표본크기가 대폭 감소되었다. 또한 9개 도에서 다같이 표본크기가 상당히 감소되었다. 유의표본크기는 현행 설계에서보다 증가하는데 이것은 유의표본을 확대함으로써 표본에 따른 변동을 줄이게 되는 효과가 있다.

표 4.4의 목표변이계수는 대부분의 시도에서 낮아지고 네 축종에 대한 변이계수가 비교적 고르게 되어 있다. 이는 본 설계에서 전수기준, 표본조사구추출이 어느 한 축종에 기준하지 않고 모든 축종의 변동을 고르게 반영해 주고 있음을 말한다.

## 5. 표본추출, 검증 및 관리

표본추출 및 배정 : 1차년도 표본은 시도별로 층내에서 난수를 발생시켜 추출하였다. 시도별 표본추출에 따른 시군별 표본수를 임의적으로 제어할 수 없어 경기 6개, 충북 2개, 충남 1개 군의 경우 현 인력으로 감당할 수 없는 표본수가 할당되었다. 따라서 이러한 시군 표본의 일부를 자의적으로 표본크기가 작은 군으로 이동시켰음을 밝혀 둔다.

표본에 의한 검증 : 표 5.1은 추출된 표본에 따라 축종별로 추정한 마리수와 변이계수이다. 단 모집단이 1995년 총 조사때와 같이 고정되어 있을 경우이다. 따라서 1998년 조사

표 4.3: 표본크기 비교

	현행설계				본 설계			
	조사구	층	표본	유의	조사구	층	표본	유의
서울	70	7	40	8	68	3	21	4
부산	258	9	53	5	335	7	76	23
대구	325	10	64	8	502	16	95	30
인천	276	11	83	11	585	15	127	31
광주	399	11	81	10	325	12	66	17
대전	367	11	82	8	313	15	64	20
울산					488	17	85	24
경기	6193	15	830	54	5559	20	795	80
강원	3117	14	433	55	3004	20	350	55
충북	2926	14	420	24	2916	19	413	61
충남	5250	14	593	53	5179	20	567	87
전북	4429	15	547	19	3996	20	453	93
전남	6869	12	684	28	6407	20	488	176
경북	7475	12	872	33	7040	20	606	127
경남	6112	12	682	47	5427	20	468	140
제주	840	9	94	10	772	16	93	20
전국	44906		5558	373	42916		4767	988

시점에서의 추정마리수와 변이계수는 다를 것이고 시간이 감에 따라 변이계수는 점차 커질 것으로 생각된다.

표본관리 : 먼저 전수농가는 기준이상의 농가가 모두 파악이 되어야 한다. 따라서 매 번 조사할 때마다 전수농가가 달라지므로 이에 따라 층내의 동질성이 변하게 되는데 이러한 변동을 층내 표본조사구의 조사에서 파악되어야 한다. 즉 층내 조사구의 이동은 절대 하여서는 안되고 이는 유의표본층의 경우도 마찬가지이다. 예를 들어 표본조사구의 사육마리수가 많아져 유의표본층 기준을 넘어서는 경우 현재는 해당조사구를 유의표본층으로 이동하고 해당 층에서 표본조사구를 추가하여 조사하고 있는데 이 결과로 표본에 의한 변이계수가 실제보다 낮아질 우려가 있다.

## 6. 개선점

본 가축통계조사 표본설계에서 가장 중점을 두어야 할 사항은 고정된 과거시점 모집단에 기초한 전수농가 및 표본조사구수의 감소가 아니라 시간의 변화에 따른 모집단의 변동



표 4.4: 목표변이계수

	현행설계				본 설계			
	조사구	총	표본	유의	조사구	총	표본	유의
서울	0.045	0.037	0.015	0.002	0.048	0.026	0.044	0.027
부산	0.045	0.044	0.044	0.024	0.050	0.047	0.039	0.025
대구	0.047	0.050	0.038	0.015	0.044	0.042	0.038	0.018
인천	0.049	0.048	0.043	0.017	0.044	0.044	0.041	0.030
광주	0.050	0.045	0.049	0.019	0.047	0.042	0.040	0.027
대전	0.048	0.048	0.019	0.036	0.044	0.032	0.031	0.036
울산					0.042	0.037	0.042	0.030
경기	0.030	0.022	0.028	0.009	0.024	0.025	0.023	0.022
강원	0.026	0.030	0.020	0.014	0.026	0.026	0.027	0.021
충북	0.027	0.030	0.024	0.014	0.027	0.030	0.026	0.020
충남	0.029	0.025	0.025	0.014	0.027	0.030	0.028	0.018
전북	0.030	0.029	0.025	0.011	0.027	0.028	0.025	0.023
전남	0.029	0.029	0.025	0.027	0.025	0.026	0.027	0.018
경북	0.025	0.029	0.025	0.012	0.025	0.025	0.028	0.016
경남	0.029	0.028	0.027	0.010	0.025	0.026	0.027	0.016
제주	0.045	0.050	0.046	0.031	0.049	0.023	0.048	0.028
전국	0.010	0.012	0.011	0.005	0.009	0.012	0.010	0.008

에 민감하지 않은 표본설계라고 생각된다. 하지만 설계당시의 제약조건이 너무 세부적이어서 표본설계를 재량껏 할 수 없었고 설계 중 본 설계자들의 주관적인 견해에 따라 제약조건을 변경토록 하는 것이 바람직하다고는 생각되지 않았다. 여기서는 제약조건들에 대한 본 설계자들의 주관적인 견해를 서술하였고 다음 설계자들이 표본설계시 참고가 되었으면 한다.

### 6.1. 교체표본제에 따른 추정

시간의 변화에 따른 변동을 고려하기 위해 교체표본제에 따른 composite 추정을 생각할 수 있으나 본 가축통계의 경우 시간에 따른 심한 변동으로 인해 그러한 교체표본제가 오히려 역효과를 낼 수도 있다고 생각되어 이는 교체표본제에 따른 표본자료를 면밀히 검토한 후 판단해야할 점이라고 생각된다. 이러한 이유로 본 설계에서는 교체표본제가 제약조건이 되어 이에 따라 표본을 추출하였으나 추정방법은 composite 추정을 도입하지 않았다. 하지만 차후 교체표본제에 따른 표본자료를 얻을 수 있어 이에 대한 연구가 가능하다고 생각된다.

표 5.1: 표본에 의한 검증

		사육마리수				추계오차			
		한우	젓소	돼지	닭	한우	젓소	돼지	닭
서울	1995년	557	374	2290	54212	0.048	0.026	0.044	0.027
	1차년도	556	360	2351	53484	0.040	0.024	0.040	0.009
부산	1995년	3724	3437	30260	259660	0.050	0.047	0.039	0.025
	1차년도	3588	3387	30666	267304	0.052	0.050	0.041	0.035
대구	1995년	25554	6481	38725	425686	0.044	0.042	0.038	0.018
	1차년도	26224	6160	37767	436525	0.043	0.038	0.037	0.025
인천	1995년	21082	9231	111915	966709	0.044	0.044	0.041	0.030
	1차년도	22939	8898	116172	941074	0.042	0.051	0.041	0.028
광주	1995년	7059	1749	14763	306520	0.047	0.042	0.041	0.027
	1차년도	6956	1714	15001	295332	0.056	0.037	0.040	0.025
대전	1995년	5814	1203	8283	184053	0.044	0.032	0.031	0.036
	1차년도	5476	1168	8743	185544	0.052	0.037	0.042	0.033
울산	1995년	32448	2160	32156	498317	0.042	0.037	0.042	0.030
	1차년도	32698	2063	31991	490056	0.041	0.039	0.045	0.026
경기	1995년	223970	206939	1524126	27752614	0.024	0.025	0.023	0.021
	1차년도	219918	215492	1525939	28501180	0.024	0.025	0.022	0.022
강원	1995년	172979	23025	227889	4027384	0.026	0.026	0.027	0.021
	1차년도	174513	22849	226570	4059754	0.027	0.025	0.028	0.021
충북	1995년	168051	28301	268817	5142424	0.030	0.029	0.028	0.020
	1차년도	167918	27000	273757	5154671	0.026	0.030	0.027	0.019
충남	1995년	377334	77082	1093070	12283110	0.027	0.028	0.028	0.018
	1차년도	374676	75582	1070184	12261588	0.026	0.029	0.028	0.018
전북	1995년	204249	36682	525196	11001695	0.027	0.028	0.025	0.023
	1차년도	201744	37731	505206	11229652	0.028	0.028	0.024	0.023
전남	1995년	441722	34809	511797	7736577	0.025	0.026	0.027	0.018
	1차년도	441087	34646	520408	7840505	0.026	0.026	0.029	0.019
경북	1995년	504469	42934	691667	12803624	0.025	0.025	0.028	0.016
	1차년도	487376	42050	727683	13132852	0.027	0.024	0.028	0.017
경남	1995년	336114	38911	722739	6733321	0.025	0.026	0.027	0.016
	1차년도	335751	59652	722823	6556588	0.024	0.026	0.026	0.013
제주	1995년	29185	3920	203118	923788	0.049	0.023	0.048	0.028
	1차년도	28611	3871	214925	965108	0.055	0.020	0.051	0.035
전국	1995년	2554311	517238	6006811	91099694	0.009	0.012	0.010	0.008
	1차년도	2530031	522623	6030186	92371217	0.009	0.012	0.010	0.008

## 6.2. 업무량 우선의 설계

시도단위의 표본추출에 따라 실제 업무단위인 시군의 출장소별 업무량이 이를 감당할 수 없게 되는 현실에서 이를 무시한 최적의 표본설계는 그 실효성이 없다고 판단된다. 하지만 업무량단위는 시군이고 표본추출 단위는 시도인 현 제약조건으로는 해결하기 곤란한 문제이다. 또한 표본설계에서 최우선 고려대상인 목표변이계수도 유지하고 현재의 업무량도 감안하여야 한다. 하지만 이 두 가지를 동시에 만족시킬 수 없는 시도 또는 시군 단위가 있다. 따라서 표본추출 단위를 현재 시도 단위에서 시군별로 하고 시군별 목표변이계수가 얼마라는 설계보다는 시군별 현재업무량 한 도내에서 최소의 목표변이계수를 갖는 표본설계로 정의하는 것이 바람직하다고 생각된다. 표본추출을 시군 단위로 하는 것은 지방자치체에 따라 자연스러운 것이고 이러한 시군 업무량 한도내 설계결과로 시군 변이계수는 시군 별로 차이가 많이 나게되는데 이는 시군 인력의 상대적인 차이로 인한 것을 나타낸다.

## 6.3. 총화의 문제점

총화추출방법은 시간에 따른 조사구내 사육마리수의 변동 및 전수농가의 변동 등으로 총내의 변동이 생겨 총내 동질성이 점차 떨어질 우려가 있다. 따라서 현재 채택하고 있는 총화추출법의 효율성이 시간경과에 따라 어느 정도 유지되는지 과거자료로 연구해 볼 필요가 있으며 총화추출법이 시간이 감에 따라 비효율적이라면 단순임의추출법이 이러한 변동을 고려하지 않아도 되므로 고려대상이 될 수 있다. 현행설계보고서에서 단순임의추출 방법을 택한 과거의 표본설계와의 비교에서 과거표본설계에서 22335개 농업조사구에서 2004개의 표본이 필요했으므로 이를 인구주택조사구를 대상으로 단순비례인 5883개라고 산정하였다. 이는 표본설계에 따른 업무량의 비교이고 실제 표본추출의 효율성을 비교하기 위해서는 인구주택조사를 대상으로 단순임의 추출한 표본수와 비교가 이루어져야한다. 즉 표본추출 모집단 대상크기가 늘어나면 표본추출비율이 낮아지므로 만일 농업조사구대신 인구주택조사구를 대상으로 단순임의 추출하면 5883개보다 작아지게 된다. 따라서 총화추출방법의 현행표본설계는 단순임의추출의 과거 표본설계에 비해 전수조사 대상농가를 대폭 줄였으나 표본조사구수는 실제로 증가했다고 보아진다.

## 6.4. 표본설계의 효율성 유지여부 점검

변이계수는 추계치의 정확도를 나타내는 것으로 변이계수가 낮을수록 추계치는 실제 모집단의 사육마리수에 근접해 있으므로 변이계수가 낮게 유지되면 표본설계가 효율적으로 유지되고 있다고 할 수 있다. 만일 표본설계과정상의 잘못으로 변이계수가 실제보다 잘못 계산되는 경우에는 표본에 의한 변이계수가 낮다고 표본설계가 효율적으로 유지되고 있다고 판단할 수 없는 것은 지극히 당연하다. 즉 표본조사결과에 의한 추계치가 변이계수를 고려했을 때 실제 모집단의 사육마리수와 일치하느냐에 의해 표본설계과정상의 정당성을 먼저 판단해야 한다. 표본설계과정이 정당한지는 조사기간중 모집단조사가 이루어지는 총조사시점에서의 총조사결과와 표본조사결과를 비교해 봄으로 확인할 수 있다.

표 6.1은 1995년 12월 총조사 시점에서의 총조사 결과와 표본조사결과에 의한 전국 추

계를 비교한 것이다. 표본조사결과를 보면 전국의 변이계수가 최대 2.1%로 어느 정도 효율성이 유지가 되고 있는 것으로 판단되지만, 표본조사결과에 의한 추계치와 변이계수를 고려한 모집단의 사육마리수에 대한 신뢰구간을 총조사결과와 비교해보면 한우를 제외한 나머지 축종들의 경우 표본에 의한 추계결과가 총조사결과와 전혀 일치하지 않는다는 것을 보여주고 있다. 예를 들어 닭의 경우 표본조사결과 모집단의 추계치는 8580만 마리고 변이계수가 0.8%라는 것은 95%신뢰수준에서 모집단 사육마리수가 8446만 마리와 8716만 마리 사이에 있다는 것으로 추계된다. 하지만 실제 모집단의 사육마리수는 9110만 마리로 그 차이가 너무 심하다고 할 수 있다. 이 경우 변이계수가 적어도 3%가 되어야만 표본결과에 의한 추계가 모집단의 사육마리수와 일치하게 되는 것으로 나타난다. 따라서 이러한 불일치에 대해 조사방법, 추계과정 등을 면밀히 검토하여 그 원인을 조사해야 하므로 그러한 변이계수를 기준하여 표본조사가 효율적으로 유지되고 있는지를 판단하는 것은 바람직하지 않다. 즉 표본설계의 효율성을 변이계수로 판단하기 이전에 먼저 표본설계과정이 정당한지를 실제 모집단과의 일치여부를 비교하여 점검하는 것이 바람직하다.

표 6.1: 1995년 12월 총조사와 표본조사비교 (단위:천마리)

축종	총조사결과	표본조사결과	표본변이계수	표본변이계수에 따른 95% 신뢰구간	95% 신뢰구간 (변이계수 3%)
한우	2554	2594	1.9%	2497 - 2691	2441 - 2747
젖소	517	553	2.1%	530 - 576	520 - 586
돼지	6007	6461	1.3%	6296 - 6626	6081 - 6841
닭	91100	85800	0.8%	84455 - 87145	80755 - 90845

## 7. 결론

가축통계조사는 이로 인해 파악되는 가축현황에 기초하여 가축사육에 대한 정책들이 수립되므로 보다 정확한 표본설계를 할 책임이 있다고 생각된다. 하지만 이러한 책임을 느끼면서도 본 설계가 본 설계자들의 짧은 지식과 제약조건등의 이유 등으로 인해 미흡한 점이 많다고 생각되며, 다음 번 가축통계조사를 설계하는 관련통계학자들에 참고가 되어 미흡한 점들이 개선된 표본설계가 되었으면 하는 바람에서 이번 표본설계의 내역을 서술하였다. 가축통계조사에 대해 표준화되는 표본설계는 관련 통계학자들의 공통된 의견을 따른 것이어야 한다고 생각된다. 이를 위해서는 관련학자들의 과거 및 본 설계내역에 대한 비판과 제언을 필요하다고 생각되며 그러한 논의가 다음 설계자들이 보다 나은 설계를 하는데 도움이 되리라고 생각한다.

## 참고문헌

- [1] 서울대학교 통계연구소 (1992), <가측통계조사 표본설계>.
- [2] 박홍래 (1987), 농업기본통계 및 가측통계조사의 표본설계 연구, <응용통계연구>, 제 1권 2호. 12-20.
- [3] Fong, W. K. (1990), Bayesian approach to successive sampling with partial replacement of units on two occasions, *Biometrika*, 77. 383-388
- [4] Williams, B. (1978), *A sampler on sampling*, New York. Wiley.
- [5] Wolter, K. M. (1979), Composite estimation in finite populations, *Journal of the American Statistical Association*, 74. 604-613

[ 1998년 2월 접수, 1998년 6월 최종수정 ]

## A Sampling Design for the Livestock (Korean Native Beef Cattle, Milk Cow, Pig, Chicken) Statistics

Kijung Yoon<sup>1)</sup> Sangun Park<sup>2)</sup>

### ABSTRACT

We made a sample design for next 5 years, based on the population as of 1995, for livestock statistics. In the sample design, we used the stratified one stage sampling method where the sample size depends on the prefixed coefficient of variation. In stratifying the population, we considered the complete linkage method, and decided the number of strata to be the one which yields the minimum sample size. We listed here some difficulties we had for the better sample design in the future.

---

1) Emeritus Professor, Department of Applied Statistics, Yonsei University, Shinchon Dong 134, Seoul, 120-749 Korea.

2) Assistant Professor, Department of Applied Statistics, Yonsei University, Shinchon Dong 134, Seoul, 120-749 Korea.