

지역교육청 수요자 만족도조사를 위한 표본설계에 관한 연구[†]

허순영¹ · 장덕준²

¹²창원대학교 통계학과

접수 2010년 5월 8일, 수정 2010년 7월 1일, 게재확정 2010년 7월 5일

요 약

지역교육청 수요자 만족도조사를 위한 표본설계는 경상남도의 2009년 경남지역교육청 고객만족도 조사의 표본크기에 기초하여 시·군별 지역교육청평가에 맞추어 설계하였다. 대도시의 구단위 지역교육청과 달리 지방의 시·군 교육청은 학생수와 학교수, 학급당 학생수 등의 변동이 크다. 시간·비용 등을 고려하여 전체 표본크기를 작게 하면서도 각 시·군 교육청 평가에 필요한 최소표본수를 확보하도록 설계하였다. 경상남도는 10개의 시지역과 10개의 군지역을 가지고 있고, 학생수가 상대적으로 작은 군지역교육청 평가에 필요한 최소표본수를 확보하기위해 지역별 평가에 필요한 최소표본을 우선배분한 후, 나머지는 지역별 학급수에 비례배분하였고, 표본학교는 지역과 학교설립유형별로 층화하여 학급수에 비례하여 추출하였다. 표본학교 내에서 조사대상 학생은 2단집락추출하였다. 지역별 추출율의 상이함을 보정하기 위해 가중치를 산정하였다. 조사자료의 분석은 가중치를 적용하여 가중평균, 가중총합 등을 이용하며, 분산의 추정은 통계소프트웨어에서 제공하는 균형반복복제, 잭나이프, 선형화방법 등을 사용할 수 있다.

주요용어: 가중치, 복합표본조사, 비례배분, 지역교육청, 크기확률비례추출.

1. 서 론

우리나라 교육행정조직은 중앙교육행정기관과 지방교육행정기관으로 구분되며, 중앙교육행정기관에 교육과학기술부가 있다. 지방교육행정기관은 전국의 특별시·광역시 및 도의 교육·학예에 관한 사무를 담당하기 위하여 설치된 행정기관으로 전국 16개 시·도 교육청이 설치되어 있다. 각 시·도 교육청 산하에는 시·도 교육 및 학예에 관한 사무를 분장하기 위하여 1개 또는 2개 이상의 시·군 및 자치구를 관할구역으로 하는 하급교육행정기관(지역교육청)을 두고 있다. 각 시·군·구 지역교육청은 공·사립의 유치원·초등학교·중학교·공민학교·고등공민학교 및 이에 준하는 각종학교의 운영·관리에 관한 지도·감독하는 업무를 관장한다(교육법전, 2008).

지방교육행정기관의 지도·감독을 받는 학교 교육에 대한 학생과 학부모의 교육만족도 조사, 일선 교사의 업무만족도 조사 등 학교교육과 관련된 많은 조사연구는 학교, 학년, 학급 등을 층 및 군집으로 한 표본추출을 한다. 김양분 등(2004)는 교육만족도 조사도구의 문항 적절성 및 신뢰도 및 타당도 검증을 위한 설문조사 “교육 만족도 조사 도구 개발 연구”에서 초등학교, 중학교, 일반계·실업계 고등학교에서 각각 1개 학교를 무작위추출하고 다시 각 표본학교에서 2개 학급을 무작위추출하여 조사하였다. 최명숙 등(2007)는 대구지역 학교교육 및 주요 교육시책에 대한 만족도 조사 “대구 교육 수요자 만족도

[†] 이 논문은 2009~2010년도 창원대학교 연구비에 의하여 연구되었음.

¹ 교신저자: (641-773) 경남 창원시 사림동 9번지, 창원대학교 통계학과, 부교수.

E-mail: syheo@changwon.ac.kr

² (641-773) 경남 창원시 사림동 9번지, 창원대학교 통계학과, 교수.

조사 연구”에서 초등학교, 중학교, 일반계·실업계 고등학교를 층으로 하여 학교급별 표본학교수를 결정하고 각 표본학교에서 조사대상 학년별 한 학급을 무작위추출한 후 표본 학급의 모든 학생을 조사하였다. 김양분 등 (2008)는 서울지역 초·중등학교의 학생, 학부모, 교사를 대상으로 실시한 “07 지방교육 혁신을 위한 고객만족도 조사 연구”에서 초등학교, 중학교, 일반계 고등학교, 실업계 고등학교를 층으로 하여 표본학교수를 결정한 후, 표본학교별 조사대상으로 결정된 1개 학년에 대해 두 개 학급을 무작위추출하여 표본 학급의 절반은 만족도 조사를 나머지 절반은 교육정책지도 조사를 실시하였다.

공·사립 초등학교와 중학교 학교교육의 지도·감독을 주요 업무로 하는 지역교육청 평가에 있어서 주로 초등학교와 중학교의 학생, 학부모, 교사가 조사대상이 된다. 초등학교의 경우 학교구분이 비교적 간단하여 대부분 공립과 사립으로 구분되나, 중학교의 경우는 초등학교보다 학교의 구성이 복잡하여 공·사립의 구분 외에 여자·남자·공학의 구분이 있다. 공립과 사립별로 학교교육에 대한 만족도가 다를 수 있고, 남학교, 여학교 또는 공학에 따라 만족도가 다를 수 있으므로 표본의 대표성 확보하기 위해서는 가능한 한 모든 학교유형이 표본에 포함되어야 할 것이며, 그 결과 중학교 표본설계는 초등학교의 경우보다 더 복잡해진다.

층화 (stratification), 집락 (clustering), 다단계 (multi-stage) 또는 다상 (multi-phase) 추출, 불균등추출, 다중틀 (multi-frame)추출 등을 복합적으로 적용하여 표본을 추출하는 방법을 복합표본설계 (complex sample design)라고 한다 (Lavrakas, 2008). 복합표본설계방법은 조사비용을 절감할 수 있으나 추정량과 추정량의 분산 추정방법이 복잡하다. 비록 추정방법이 복잡하다고 하나, 복합표본설계에 의한 조사자료의 추정량의 분산 추정과 서로 다른 분산추정량의 효율에 대한 비교연구가 이미 많이 행해져 있으며 (Krewski와 Rao, 1981; Binder, 1983; Shao, 1996), 컴퓨터와 통계소프트웨어의 발달로 추정량의 계산이 용이해졌다. 그 결과, 오늘날 많은 조사연구에서 복합표본설계에 의한 표본추출을 하고 있고 적용하고 있다 (김달호 등, 2009; 김양분 등, 2008).

본 연구에서는 20개 시·군 지역교육청을 가지고 있는 경상남도 교육청의 2009년 통계연보 자료를 이용하여 시·군·구 단위의 지역교육청별 중학교 표본추출을 위한 표본설계를 제안하였고, 본 표본설계는 “2009년 경남지역교육청 고객만족도 조사”에서 정한 표본크기에 기초하여 설계하였다. 본 연구에서 제안하는 표본설계는 중학교의 초등학교와 고등학교 및 타 시·도 교육청의 상황에 따라 적절하게 조정하여 적용할 수 있을 것이다.

2. 조사 모집단

경상남도 교육청에서 실시한 “2009년 경남지역교육청 고객만족도 조사”는 각 시·군 교육청별 교육 수요자 만족도조사를 위한 것으로 학생, 학부모, 교사를 대상으로 실시하였다. 이 조사의 조사모집단은 경상남도 도교육청 산하 20개 지역교육청에서 관할하는 공·사립 초등학교 5~6학년과 중학교 2~3학년 학생과 학부모로 하였고, 공·사립 초등학교와 중학교에 재직 중인 교사 전체로 하였다. 조사대상 학교에서 분교, 도서, 벽지 학교는 제외되었다. 조사의 비용과 시간 등을 고려하여 학생 표본크기는 전체 모집단의 약 2%에 해당하는 학생 3,500명으로 하였다. 학부모표본은 조사대상학생의 학부모전체로 하였고, 교사는 표본학교에 재직 중인 교사로 하였다.

표 2.1~표 2.3은 경상남도 시·군 교육청별 전체 학교수, 2~3학년 학급수와 학생수를 정리한 표이다. 이 자료는 경상남도 도교육청 홈페이지에 탑재된 2009 경남교육통계에 기초하여 작성된 것이다.

표들에서 알 수 있듯이 20개 시·군 교육청별 학교수, 학교설립유형의 구성 및 학교수, 학급수와 학생수가 매우 다양하다. 시·군별 교육청 평가를 위해서는 각 교육청별 학교유형의 다양성을 포함하면서도 동시에 교육청별 평가에 유용한 최소표본크기를 확보할 수 있어야 한다.

표 2.1 경상남도 시·군 교육청별 중학교 조사모집단 학교수

지역	모집단학교수	공립 학교수				사립 학교수			
		여자	남자	공학	합계	여자	남자	공학	합계
창원	29	2	1	22	25	1	1	2	4
마산	24	4	6	7	17	4	2	1	7
진주	20	2	4	9	15	2	2	1	5
진해	9	2	1	4	7	0	1	1	2
통영	8	2	2	2	6	1	1	0	2
사천	11	2	2	2	6	1	1	3	5
김해	29	2	3	20	25	1	0	3	4
밀양	14	1	1	5	7	1	2	4	7
거제	17	0	0	12	12	0	0	5	5
양산	14	1	2	7	10	1	0	3	4
의령	5	0	1	1	2	1	0	2	3
함안	8	0	1	1	2	1	0	5	6
창녕	10	0	1	2	3	3	1	3	7
고성	9	1	1	6	8	0	1	0	1
남해	11	1	1	5	7	0	0	4	4
하동	9	0	0	8	8	0	0	1	1
산청	6	0	0	6	6	0	0	0	0
함양	5	1	1	2	4	0	0	1	1
거창	10	1	1	5	7	1	1	1	3
합천	11	0	1	5	6	1	0	4	5
전체	259	22	30	131	183	19	13	44	76

표 2.2 경상남도 시·군 교육청별 중학교 2~3학년 조사모집단 학급수

지역	공립학교 학급수				사립학교 학급수				학급수총계
	여자	남자	공학	합계	여자	남자	공학	합계	
창원	31	20	365	416	20	20	25	65	481
마산	57	98	103	258	54	26	8	88	346
진주	41	80	44	165	45	40	11	96	261
진해	29	15	57	101	0	13	4	17	118
통영	32	39	4	75	10	8	0	18	93
사천	23	28	6	57	9	10	10	29	86
김해	28	46	340	414	9	0	20	29	443
밀양	15	16	12	43	10	16	18	44	87
거제	0	0	146	146	0	0	41	41	187
양산	16	36	100	152	21	0	38	59	211
의령	0	6	4	10	5	0	4	9	19
함안	0	12	2	14	9	0	22	31	45
창녕	0	10	13	23	16	6	7	29	52
고성	10	6	16	32	0	6	0	6	38
남해	7	8	11	26	0	0	14	14	40
하동	0	0	31	31	0	0	6	6	37
산청	0	0	22	22	0	0	0	0	22
함양	11	12	4	27	0	0	3	3	30
거창	10	12	11	33	8	10	4	22	55
합천	0	8	13	21	7	0	10	17	38
전체	310	452	1,304	2,066	223	155	245	623	2,689

표 2.3 경상남도 시·군 교육별 중학교 2~3학년 조사모집단 학생수

지역	공립학교 학생수				사립학교 학생수				총계
	여자	남자	공학	합계	여자	남자	공학	합계	
창원	1,178	820	13,421	15,419	793	808	787	2,388	17,807
마산	1,876	3,376	3,537	8,789	1,852	853	264	2,969	11,758
진주	1,752	3,136	1,310	6,198	1,892	1,572	353	3,817	10,015
진해	961	488	1,943	3,392	0	434	142	576	3,968
통영	1,123	1,366	87	2,576	340	297	0	637	3,213
사천	731	912	152	1,795	295	345	250	890	2,685
김해	909	1,515	12,726	15,150	280	0	580	860	16,010
밀양	528	540	198	1,266	354	530	424	1,308	2,574
거제	0	0	5,104	5,104	0	0	1,209	1,209	6,313
양산	534	1,242	3,473	5,249	710	0	1,309	2,019	7,268
의령	0	207	89	296	133	0	48	181	477
함안	0	345	21	366	272	0	544	816	1,182
창녕	0	298	332	630	406	164	96	666	1,296
고성	314	173	362	849	0	176	0	176	1,025
남해	215	266	205	686	0	0	273	273	959
하동	0	0	752	752	0	0	163	163	915
산청	0	0	518	518	0	0	0	0	518
함양	335	358	69	762	0	0	59	59	821
거창	319	416	179	914	252	346	140	738	1,652
합천	0	228	244	472	195	0	140	335	807
전체	10,775	15,686	44,722	71,183	7,774	5,525	6,781	20,080	91,263

3. 지역교육청별 표본크기의 결정

3.1. 학교급별 표본크기

학교급별 표본크기는 학생수에 비례하여 비례배분한다. 표본크기 3,500명을 전체 초등학교 5~6학년 학생수와 중학교 2~3학년 학생수에 기초하여 배분한다. 표 3.1은 경상남도 교육청 산하 초등학교 5~6학년 학생수와 중학교 2~3학년 학생수 및 학교급별 학생수에 비례배분한 표본크기이다.

경남지역 초등학교는 모두 남·여 공학으로 전체조사대상 473학교 중 2개교를 제외한 나머지는 공립학교이므로 표본설계가 비교적 간단하여, 본 연구는 대표성 있는 중학생 1,802명의 추출하는 것을 목표로 표본설계하였다.

표 3.1 경상남도 교육청 산하 조사대상 모집단 학생수 및 표본배분

	학생수	학생수비율	표본크기
초등학교 5~6학년	86,041	48.53%	1,698
중학교 2~3학년	91,263	51.47%	1,802
합계	177,304	100.00%	3,500

3.2. 시·군 교육청별 표본배분

(1) 지역별 표본학생수의 배분

먼저 경남지역을 20개 시·군 교육청별로 층화하였다. 각 층에 표본크기를 배분하는 방법에, 비례배분, 네이만배분, 최적배분 등이 있다. 많은 경우, 학교조사는 학교에 설문지가 우편으로 배송되고 학교의 협조를 얻어 조사가 이루어지므로 시·군별 조사비용은 동일하다고 할 수 있고, 이 경우 최적배분과

네이만배분이 동일하므로 비례배분과 네이만배분을 고려할 수 있다. 네이만배분의 경우, 주요 조사변수들에 대한 분산이나 분산추정값에 대한 정보가 필요하나 조사변수들에 대한 신뢰할 만한 분산추정값이 없는 경우가 많아 그 대안으로 비례배분을 사용한다.

각 시·군별 표본을 비례배분함에 있어서, 먼저 각 지역교육청별 전체 조사대상학 교수에 비례하여 배분하는 방법을 고려할 수 있다. 이 경우, 지역교육청별 비교적 고르게 표본이 배분되나 학급당 크기가 큰 시지역에 비해 작은 지역에 상대적으로 많은 표본이 배분된다. 예를 들어, 학생수가 2,865명인 사천시와 807명인 합천군은 학교수가 동일하므로 같은 크기의 학생수가 배정되고, 이 경우 합천군은 학생수 3,968명인 진해보다 더 많은 표본이 배분된다.

다음으로 시·군별 전체조사대상 학급수에 비례하여 표본을 배분하는 방법이 있다. 이 경우, 시지역에 비해 상대적으로 학급수가 적은 군지역은 해당 교육청을 평가하기에는 너무 작은 수의 표본이 배분된다. 예를 들어, 학급수가 상대적으로 작은 의령군이나 합천군의 경우, 단지 13명과 15명만 조사하게 되고 이러한 표본크기는 해당 지역교육청을 평가하기에 충분하다고 보기 어렵다. 따라서 이들 군에 최소 30명이상을 조사하고자 하면 전체적으로 약 2.3배에 해당하는 4,000명 정도를 조사해야한다.

세 번째로 시·군별 전체조사대상 학생수에 비례하여 표본을 배분할 수 있다. 이 경우는 학급수에 비례하여 배분하는 경우보다 지역교육청별 더 심각하게 표본크기의 변동이 크고 일부 지역교육청에 배분되는 표본수가 심각하게 작아, 학생수가 단지 477명인 의령군의 경우 9명의 표본만이 배분된다.

조사대상 모집단의 학교수, 학급수, 학생수에 비례배분하는 경우 발생하는 이와 같은 문제를 극복하기 위해, 본 표본설계에서는 통계학에서 일반적으로 표본크기가 큰지 또는 작은 지를 구분하는 데 기준이 되는 30명을 각 20개 지역교육청별로 우선배분한 후, 나머지 1,204명을 각 지역별 조사대상 학급수의 크기에 비례하여 할당한다. 아래 표 3.3의 두 번째 열 (표본학생수)은 이와 같이 결정된 지역별 표본 학생수이다.

(2) 지역별 표본학교수의 결정

경남지역 중학교는 앞의 표 2.1~표 2.3에서 보여지는 바와 같이 공·사립, 남자·여자·공학의 구분이 있고, 지역에 따라서 어떤 지역은 모든 설립형태별 학교가 존재하는 반면 어떤 지역은 그렇지 않다. 각 지역별 할당된 표본학생수에 기초하여 표본학교를 결정하는 과정에서 설립형태가 서로 다른 학교 유형들을 모두 포함시키기 위해서는 표본학교당 학생수를 크게 하기 보다는 지역별 표본학교수를 크게 할 필요가 있다. 한편, 일반적으로 시지역은 군지역보다 평균학급규모가 크다. 이러한 측면을 고려하여 모든 설립형태별 학교가 표본에 포함되기에 충분한 표본학교수를 확보하기위해 시지역은 표본학생수를 20명으로 군지역은 10명으로 각각 나누어 지역별 학교표본수를 결정하게 된다 (표 3.3의 세 번째 열). 표 3.2는 경남지역 중학교 2~3학년의 시·군지역별 학생수와 학급수의 분포 및 학급당 평균학생수를 나타낸다.

표 3.2 시·군지역별 학생수와 학급수 분포 (중학교 2~3학년)

중학교	학생수	비율	학급수	비율	학급당 학생수
시지역	81,611	89.42%	2,313	86.02%	35
군지역	9,652	10.58%	376	13.98%	26
전체	91,263	100.00%	2,689	100.00%	34

이러한 절차에 의해 표본학교수를 결정하였다 하더라도 지역별로 다양한 학교 설립형태를 모두 포함하기에 부족한 지역이 있다. 예를 들어, 진해시의 경우, 할당된 표본학교수는 4개교이나 설립형태가 서로 다른 다섯 개의 학교가 존재한다. 즉, 공립 여학교, 공립 남학교, 공립 공학교, 사립 남학교, 사립 공학교가 존재한다. 이 다섯 가지 설립형태의 학교를 모두 포함시키기 위해 진해시의 최종 표본학교수는 5개교로 한다. 이와 같은 절차에 의해 결정된 중학교 최종 표본학교수는 표 3.3의 네 번째 열 (최종 표

본학교수)에 주어져 있다.

이러한 배분방식에 의해 결정된 학교표본크기에 기초하여 각 학교설립유형별 최소 한 개 학교가 배분 되도록 할당한 후, 나머지 표본학교수는 각 지역내 학교설립유형별 학교수에 비례하여 배분한다.

(3) 표본학교당 조사대상 표본학생수의 결정

각 표본학교당 조사대상 학생수는, 각 지역별로 배분된 표본학생수를 각 지역별로 배분된 표본학교수로 나누어 결정하고, 표 3.3의 다섯 번째 열 (학교당 표본크기)에 주어져 있다. 그 결과 학교수가 더 많이 추출된 지역은 표본학교당 조사할 표본학생수가 작아진다.

최종적으로 표본학교당 중학교 2학년과 3학년 각 한 개 학급씩을 무작위 추출하여 할당된 학생수의 50%씩을 조사한다.

표 3.3 중학교 시·군 교육청별 표본 학생수와 표본 학교수

지역	표본학생수	표본학교수	최종표본 학교수	학교당 표본크기	최종 표본 학생수
창원	244	12	12	20	244
마산	184	9	9	20	184
진주	152	8	8	20	160
진해	83	4	5	18	90
통영	71	4	5	14	71
사천	68	3	6	12	72
김해	227	11	11	20	220
밀양	69	3	6	12	69
거제	113	6	6	20	120
양산	124	6	6	20	120
의령	38	4	4	10	38
함안	50	5	5	10	50
창녕	53	5	5	10	50
고성	47	5	5	10	50
남해	48	5	5	10	48
하동	46	5	5	10	50
산청	40	4	4	10	40
함양	43	4	4	10	40
거창	55	6	6	10	60
합천	47	5	5	10	50
합계	1,802	114	122		1,826

표 3.4은 최종학교표본의 학교설립유형별 분포이고, 표 3.5은 지금까지의 표본추출절차를 거쳐 최종적으로 결정된 중학교표본의 지역별 표본추출률을 나타낸다.

4. 표본의 선정

먼저 각 지역청별 표본학교의 추출은 각 학교설립유형별로 하되, 각 개별 중학교 2~3학년 총학급수에 비례하여 추출한다. 또 학급표본은 표본학교 내에서 2~3학년 각 1개 학급씩을 무작위추출한다. 각 표본학급 내에서 할당된 표본학생수만큼 단순임의무작위추출 (simple random sampling)을 한다.

5. 가중치 산정

가중치의 산정은 조사추정치 (survey estimates)의 정확도 (accuracy)를 높이기 위해 조사자료에 적용하는 통계적조정 (statistical adjustments)으로 조사연구자들이 조사자료에 가중치를 부여하는 데는

표 3.4 중학교 시·군 교육청·학교유형별 표본학교수

지역	전체 학교수	표본학교수	공립 학교수				사립 학교수			
			여자	남자	공학	합계	여자	남자	공학	합계
창원	29	12	1	1	7	9	1	1	1	3
마산	24	9	1	2	3	6	1	1	1	3
진주	20	8	1	1	3	5	1	1	1	3
진해	9	5	1	1	1	3	0	1	1	2
통영	8	5	1	1	1	3	1	1	0	2
사천	11	6	1	1	1	3	1	1	1	3
김해	29	11	1	1	7	9	1	0	1	2
밀양	14	6	1	1	1	3	1	1	1	3
거제	17	6	0	0	4	4	0	0	2	2
양산	14	6	1	1	2	4	1	0	1	2
의령	5	4	0	1	1	2	1	0	1	2
합안	8	5	0	1	1	2	1	0	2	3
창녕	10	5	0	1	1	2	1	1	1	3
고성	9	5	1	1	2	4	0	1	0	1
남해	11	5	1	1	2	4	0	0	1	1
하동	9	5	0	0	4	4	0	0	1	1
산청	6	4	0	0	4	4	0	0	0	0
함양	5	4	1	1	1	3	0	0	1	1
거창	10	6	1	1	1	3	1	1	1	3
합천	11	5	0	1	2	3	1	0	1	2
전체	260	122	13	18	48	79	13	10	19	43

표 3.5 중학교 시·군 교육청별 학교·학생 표본추출률

지역	학교수			학생수		
	전체	표본	표본 추출률	전체	표본수	표본 추출률
창원	29	12	41.4%	17,807	244	1.4%
마산	24	9	37.5%	11,758	184	1.6%
진주	20	8	38.1%	10,015	160	1.6%
진해	9	5	55.6%	3,968	90	2.3%
통영	8	5	62.5%	3,213	71	2.2%
사천	11	6	54.5%	2,685	72	2.7%
김해	29	11	37.9%	16,010	220	1.4%
밀양	14	6	42.9%	2,574	69	2.7%
거제	17	6	35.3%	6,313	120	1.9%
양산	14	6	42.9%	7,268	120	1.7%
의령	5	4	80.0%	477	38	8.0%
합안	8	5	62.5%	1,182	50	4.2%
창녕	10	5	50.0%	1,296	50	3.9%
고성	9	5	55.6%	1,025	50	4.9%
남해	11	5	45.5%	959	48	5.0%
하동	9	5	55.6%	915	50	5.5%
산청	6	4	66.7%	518	40	7.7%
함양	5	4	80.0%	821	40	4.9%
거창	10	6	60.0%	1,652	60	3.6%
합천	11	5	45.5%	807	50	6.2%
전체	259	122	46.9%	91,263	1,826	2.0%

두 가지 기본적인 이유가 있다. 하나는 추출과정에서 발생하는 불균등 추출확률을 수정하기 위함이고 다른 하나는 무응답을 조정하기 위함이다 (Lavrakas, 2008).

표 3.5에서와 같이 표본크기를 비례할당하지 않고, 표본추출에서 불균등확률추출을 적용한 결과 각

지역별 학교와 학생의 표본추출율은 매우 다르다. 즉, 각 학교와 학생들이 지역별로 표본으로 추출될 확률이 크게 다르기 때문에 표본학생들이 몇 명의 의견을 대표하는 지는 지역마다 크게 차이가 난다. 이러한 차이를 조정하기 위해 가중치를 산출하고 분석과정에서 이 가중치를 사용한다.

가중치는 기본적으로 표본을 추출하는 과정에서 결정되는 기본가중치 (base weight)와 조사과정에서 발생하는 무응답을 조정하기 위한 가중치, 최종적으로 모집단의 하위 그룹별 가중치합과 모집단과의 차이를 조정하는 사후추정가중치로 구분할 수 있다.

5.1. 기본가중치

표본을 추출하는 과정에서 결정되는 가중치가 기본가중치 (base weights)이다. h 번째 지역, i 번째 층 (학교설립유형)의 j 번째 표본학교, k 번째 표본학급, l 번째 표본학생의 기본가중치, $w_{B(hijkl)}$ 는

$$w_{B(hijkl)} = \frac{1}{n_{hi}p_{hij}n_{hij}p_{hijk}n_{hijk}p_{hijkl}}$$

이고, 여기서

n_{hi} = h 번째 지역, i 번째 층 (학교설립유형)의 표본학교수

p_{hij} = (hij) 번째 학교의 추출확률

n_{hij} = (hij) 번째 표본학교의 2학년 (또는 3학년) 표본학급수

p_{hijk} = ($hijk$) 번째 학급의 추출확률

n_{hijk} = ($hijk$) 번째 표본학급의 조사대상 표본학생수

p_{hijkl} = ($hijkl$) 번째 학생이 추출될 확률

이다.

학년별 1개 학급을 단순무작위추출하는 경우, $n_{hij=1}$ 이고 $1/p_{hijk}$ 는 학년별 학급수가 된다. 또, 학급 크기가 N_{hijk} 인 학급에서 크기 n_{hijk} 인 표본학생을 단순임의 추출하는 경우, 각 표본학급내 학생가중치는 $(n_{hijk}p_{hijkl})^{-1} = N_{hijk}/n_{hijk}$ 이 된다.

5.2. 단위 무응답조정 가중치

무응답은 크게 단위무응답과 항목무응답으로 나눈다. 단위무응답이란 표본에 포함된 기본단위로부터 정보를 얻는 데 실패하거나 응답내용 전체가 쓸모가 없어진 경우 발생한다. 단위무응답은 주로 가중치 조정에 의해서 조정하고, 항목무응답은 보정 (imputation)에 의해서 조정한다.

단위 무응답을 보정하기 위해 가장 일반적으로 사용되는 방법들 중의 하나는 무응답으로 손실된 만큼 보정하기 위해 표본학급내의 응답자들에게 동일한 값을 보정하는 방법이다 (Lavrakas, 2008).

($hijk$) 번째 표본학급에서 단위무응답이 발생하여 할당된 표본수보다 작은 m_{hijk} 명의 학생이 응답하는 경우, 그 학급내에서 응답한 학생들의 가중치는 n_{hijkl}/m_{hijkl} 만큼 증가한다. 따라서 응답한 ($hijkl$) 번째 학생의 조정된 가중치, $w_R(hijkl)$ 는

$$w_R(hijkl) = w_{B(hijkl)} * \frac{n_{hijk}}{m_{hijk}}$$

이다.

5.3. 사후층화가중치

기본 가중치에 단위무응답을 조정한 가중치는 다시 각 지역별, 세부하위 속성별 조사대상 총학생수와 일치하도록 조정하여 최종가중치를 산출한다. 즉, 속성 C 의 모집단 학생수가 N_C 이고 속성 C 를 갖는 응답학생들의 가중치합이 일치하지 않으면 일치하지 않는 정도를 조정한다. 조정된 가중치는

$$w_{hijkl} = w_{R(hijkl)} \times \frac{N_C}{\widehat{N_C}}$$

이고, 이 때 $\widehat{N_C} = \sum_{hijkl \in C} w_{R(hijkl)}$ 이다.

최종적으로 w_{hijkl} 이 분석을 위한 분석가중치가 된다.

6. 자료의 분석

복합표본설계에 의한 조사자료는 IID (independent and identically distributed)라는 가정을 만족하지 못하므로 IID를 가정하는 종래의 통계적 분석방법을 적용할 수 없다. 복합표본설계에 의해 조사된 자료의 평균이나 총합, 비율 등의 추정치는 5장에서 계산된 분석가중치를 적용하여 가중평균, 가중총합 등을 산출한다. 표본배분과 표본추출은 각 시·군 교육청별·학교설립유형별 층화하여 표본배분 및 표본추출이 이루어졌으나, 각 시·군내에서 학교설립유형별 1개 학교가 추출된 경우가 많으므로 분산추정시 모든 층을 다 고려할 수가 없으므로 추정량의 분산을 추정하기 위해 지역과 공·사립 구분을 층으로 하고 학급을 집락으로 하여 잭나이프 (Jackknife), 균형반복복제 (BRR: Balanced Repeated Replication), 붓스트랩 (bootstrap), 선형대체법 (linearization - substitution method) 등을 적용하여 추정할 수 있다. 복합표본설계에 의한 조사자료의 좀 더 자세한 분석방법은 복합표본설계에 의한 표본자료의 분석에 관한 기존의 문헌들을 참고할 수 있다 (Korn과 Graubard, 1999).

7. 결론

2009년 경남지역교육청 고객만족도 조사의 표본크기에 기초하여 시·군별 지역교육청평가에 맞추어 설계하였다. 대도시의 구단위 지역교육청과 달리 지방의 시·군 교육청은 학생수와 학교수, 학급당 학생수 등의 변동이 크다. 이 경우, 비용 등을 고려하여 전체 표본크기를 작게 하면서도 각 지역청 평가에 필요한 최소표본수를 확보하는 것이 중요하다.

2009년 경남지역교육청 고객만족도 조사는 모집단의 약 2%를 표본크기로 하였다. 각 지역청별 최소 표본크기 30명을 우선할당한 후 나머지를 각 지역별 학급크기에 비례하여 배분하였다. 이 후, 시지역은 학교당 20명을 군지역은 학교당 10명을 조사하여, 지역별 다양한 학교설립유형을 모두 포함하도록 학교표본수를 확보하였다. 표본학교에서 학년별로 1개학급을 무작위추출한 후, 표본학급에서 조사대상학생들을 정해진 수만큼 조사하였다. 학교표본은 각 학교별 총학급수에 비례하여 추출하였고, 표본학급과 표본학생은 단순임의무작위추출을 하였다.

이러한 표본설계는 시·군·구 단위 평가 및 비교에 필요한 최소한의 표본수를 만족하면서 동시에 대표성을 확보할 수 있다. 또한 유사한 다른 조사에 적용할 수 있다.

참고문헌

김달호, 김남희 (1998). 신뢰성있는 지역통계 추정을 위한 제언. <한국데이터정보과학회지>, 9, 337-344.

- 김달호, 조길호, 황진섭, 정경하 (2009). 경북인의 생활과 의식조사 표본설계. <한국데이터정보과학회지>, **20**, 1155-1167.
- 김성열, 서민원, 오세희 (2006). <2006년도 학교평가 고객만족도 조사>, 경상남도교육청.
- 김성식, 임현정, 김옥남 (2008). <2008 서울 교육 만족도 조사 연구>, 한국교육개발원.
- 김양분, 유한구, 남궁지영 (2004). <교육 만족도 조사 도구 개발 연구>, 한국교육개발원.
- 김양분 (2008). <07 지방교육 혁신을 위한 고객만족도 조사 연구>, 한국교육개발원.
- 김종태, 서효민, 이인락 (2009). 2026년까지 대구광역시와 경상북도 지역의 고등학교 3학년 학생수에 대한 예측과 대학 입학정원수와의 비교 분석. <한국데이터정보과학회지>, **20**, 159-169.
- 윤용화, 김종태, 이우동 (1998). 평활 적합도 검정에서의 분산추정의 영향. <한국데이터정보과학회지>, **9**, 189-202.
- 최명숙, 김성후, 최성열 (2007). <대구 교육 수요자 만족도 조사 연구>, 계명대학교.
- 교육법전편찬회 (2010). <교육법전>, 교학사.
- Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279-292.
- Korn, E. L. and Graubard, B. I. (1999). *Analysis of health surveys*, John Wiley & Sons, Inc., New York.
- Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, **9**, 1010-1019.
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*, Vol.2, SAGE Publication, Inc., London.
- Oh, H. I. and Scheuren, F. J. (1983). Weighting adjustment for unit nonresponse. In *Incomplete Data and Sample Surveys*, Vol. 2, Ed. W. G. Madow, I. Olkin and D. B. Rubin, 143-184. New York, Academic Press.
- Shao, J. (1996). Resampling methods in sampling surveys. *Statistics*, **27**, 203-254.

A sample survey design for service satisfaction evaluation of regional education offices

Sunyeong Heo ¹, Duk-Joon Chang ²

^{1,2}Department of Statistics, Changwon National University

Received 8 May 2010, revised 1 July 2010, accepted 5 July 2010

Abstract

A sample survey design is suggested for the service satisfaction evaluation of regional education offices based on the sample size of 2009 Gyeongnam regional education offices's customer satisfaction survey. The sample design is developed to fit the goal of evaluation of individual regional offices and allocate at least the minimum sample size to each city or county in Gyeongnam to achieve the goal of the survey. The population is stratified according to the regions and the types of schools, and the sample of schools is selected with proportional to the size of classes within each stratum. Finally, each sample student is selected according to two-stage cluster sampling within each sample school. Weighting averages, weighting totals and so on can be evaluated for analysis purposes. Their variance estimates can be evaluated using re-sampling methods like BBR, Jackknife, linearization-substitution methods, which are generally used for the data from a complex sample.

Keywords: Complex sample survey, PPS sampling, proportional allocation, regional education offices, weights.

¹ Corresponding author: Associate professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea. E-mail: syheo@changwon.ac.kr

² Professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea.