

국민건강영양조사 표본설계를 위한 추출틀 구축

박진우¹ · 변종석² · 박민규³

¹수원대학교 통계정보학과, ²한신대학교 정보통계학과, ³고려대학교 통계학과

(2010년 8월 접수, 2010년 9월 채택)

요약

제 5기 국민건강영양조사(2010년-2012년)를 위한 표본설계의 여건은 과거에 비해 불리한데, 무엇보다 과거와 같은 추출틀 마련이 여의치 않다는 점이 두드러진다. 제5기 조사를 위한 표본설계 시기가 인구주택총조사 주기와 맞지 않음으로 인해 과거의 총조사 자료를 그대로 추출틀로 사용한다면 어쩔 수 없이 추출틀오차의 문제가 발생하게 될 것이기 때문이다. 이를 극복하기 위한 현실적인 방안을 마련하는 것은 실용적인 면에서 매우 절실한 문제라고 할 수 있다. 본 연구의 목적은 인구주택총조사 자료가 낡아 현재의 모집단 상황을 제대로 대표할 수 없는 상황에서 이를 극복할 수 있는 효과적인 추출틀을 마련하는 데 있다. 본 연구에서는 추출틀의 포함률을 높이기 위한 방안으로 아파트 가구와 일반가구(단독주택, 연립주택, 다세대주택 등)를 분리하여 각각에 대해 서로 다른 자료를 활용하는 방안을 도입한다. 다시 말해 표본설계를 위해 여러 개의 추출틀을 사용하는 방안을 제시한다.

주요용어: 추출틀, 추출틀오차, 포함률.

1. 서론

국민건강영양조사는 국민의 보건수준을 정확하게 진단하기 위해 건강관련 의식 및 행태, 질환의 유병 및 관리 현황, 식품 및 영양섭취실태 등을 측정하여 국가 대표통계를 산출하는 조사이다. 조사 결과는 국민건강증진 종합계획의 목표 설정 및 평가, 건강증진 프로그램 개발 등 국가 보건정책에 필요한 기초자료로 활용된다(질병관리본부, 2008). 이계오와 박진우(2007)에 의하면, 2007년부터 2009년에 걸쳐 조사된 제 4기 국민건강영양조사를 위한 표본설계에서는 국내 최초로 순환표본추출(rolling sampling) 방식이 적용되었다. 순환표본을 도입함으로써 전문화된 조사요원들에 의해 상시조사가 진행되는 연중조사 체계가 정착되었고, 조사자들의 수준 차이로 인해 생길 수 있는 비표본오차의 발생을 방지할 수 있게 되었다. 2010년부터 2012년 사이에 수행될 제 5기 조사를 위한 표본설계에서도 제 4기와 마찬가지로 순환표본설계 방식을 채택한다(한국통계학회, 2010).

그런데 제 5기 조사를 위한 순환표본설계의 여건은 제4기 표본설계에 비해 여러 면에서 불리하다. 무엇보다도 중요한 차이는 추출틀 마련을 위한 여건이다. 우리나라에서 대부분의 전국단위 또는 대형 가구 조사의 경우 통계청(2005)에서 제공하는 인구주택총조사 90% 자료를 표본추출틀로 사용하고 있다(강현철 등, 2009). 제 4기 국민건강영양조사를 위한 표본설계는 2005년 인구주택총조사 결과가 공표되던 시기에 이루어졌기에 모집단 현황을 나타내기 위한 최신의 자료로 인구주택총조사를 활용할 수가 있었다. 즉, 인구주택조사구를 1차추출단위(primary sampling unit; psu)로 정한 후, 2005년 인구주택총조사의 조사구별로 집계된 연령대별 인구수, 주택형태별 가구수 자료들을 담은 추출틀을 마련하였다. 그

¹교신저자: (445-743) 경기도 화성시 봉담읍 와우리, 수원대학교 통계정보학과, 교수.

E-mail: jwpark@suwon.ac.kr

리나 제 5기 표본설계가 이루어지는 2009년 현재 시점에서는 2005년 인구주택총조사 자료를 추출틀로 사용할 경우 모집단에 대한 대표성이 떨어져 심각한 추출틀오차(frame error)가 생기게 될 것이므로 이러한 추출틀오차를 최소화시킬 수 있는 새로운 방안을 마련하는 것이 필요하게 되었다. 표본조사에서 복수의 추출틀을 활용하는 주제에 대해서 Hartley (1962, 1972), Iachan과 Dennis (1993) 등이 연구한 바 있다.

우리나라의 경우 많은 표본조사에서 추출틀로 5년에 한번 시행되는 인구주택총조사 자료를 활용하게 된다. 만일 조사를 위한 표본설계 시기가 제 5기 국민건강영양조사와 같이 총조사 주기와 맞지 않은데도 불구하고 과거의 총조사 자료를 그대로 추출틀로 사용한다면 어쩔 수 없이 추출틀오차의 문제에 부딪히고 만다. 따라서 이를 극복하기 위한 현실적인 방안을 마련하는 것은 실용적인 면에서 매우 절실한 문제인데도 이에 대한 연구는 찾아보기 어려운 실정이다. 본 연구의 목적은 인구주택총조사 자료가 낡아 현재의 모집단 상황을 제대로 대표할 수 없는 상황에서 이를 극복할 수 있는 효과적인 추출틀을 마련하는 데 있다. 본 연구에서는 추출틀의 포함률(coverage rate)을 높이기 위한 방안으로 아파트가구와 일반가구(단독주택, 연립주택, 다세대주택 등)를 분리하여 각각에 대해 서로 다른 자료를 활용하는 방안을 도입한다. 다시 말해 표본설계를 위해 여러 개의 추출틀을 도입하여 사용하는 방안을 제시한다. 2절에서는 2009년 시점에서 인구주택조사구 추출틀을 사용하려 할 때의 문제점을 지적하고, 3절에서는 새로운 추출틀 구축에 대해 설명한다. 마지막으로 4절에서는 본 연구의 결과를 요약, 정리한다.

2. 인구주택조사구 추출틀의 문제점

제 4기 국민건강영양조사 표본설계를 위한 추출틀로는 2005년 인구주택총조사의 동읍면별 자료가 활용되었다 (이계오와 박진우, 2007). 제 5기 조사를 위한 표본설계를 하게 되는 2009년 현재 시점에서 활용 가능한 최신의 인구주택총조사 자료는 지난 4기 때와 마찬가지로 2005년의 결과뿐이다. 그런데 지난 3~4년 사이 모집단의 상황에는 많은 변화가 생겼을 것을 예상할 수 있다. 참고로 다음의 표 2.1에 2005년 인구주택총조사 자료와 2009년 6월의 주민등록자료에 나타난 우리나라 인구수 현황을 비교하여 나타내었다.

표 2.1의 결과를 보면 2009년 6월 주민등록 인구수는 2005년 총조사 인구수에 비해 약 8.5%인 약 390만 명이 늘어난 것으로 나타났다. 이는 2005년 총조사 자료를 추출틀로 사용할 경우 모집단의 8.5% 가량이 추출틀에 포함되지 못하게 된다는 것을 시사한다. 시도별로 정도의 차이는 있지만 비슷한 양상을 나타내고 있다. 물론 인구주택총조사 인구와 주민등록 인구가 정확하게 일치하는 것은 아니지만 거의 비슷한 수준이므로 인구주택총조사 추출틀의 불완전성에 대해 문제를 제기하는 것은 전혀 무리한 일이라고 할 수 없다.

다음으로 살펴볼 것은 아파트 가구수이다. 참고로 2005년 우리나라 전체 가구수는 약 1,250만 호 가량이었는데 그 중 아파트는 약 660여 만 호로 전체 가구 중 53.0%를 차지하였다. 다음의 표 2.2에는 시도별 2005년 인구주택총조사 아파트 주택수와 2005년 11월 이후 2008년 3월까지 신축된 아파트 세대수가 나와 있다. 이 표는 2005년 총조사는 2005년 10월까지의 아파트 주택수 현황을 나타낸 것이므로 그 이후 얼마나 많은 아파트 주택이 추가 되었는지를 보여준다. 달리 말하자면, 2005년 인구주택총조사 자료를 그대로 추출틀로 할 경우 미포함(undercoverage)이 어느 정도 되는지를 시사해주는 결과라고 말할 수 있다.

표 2.2를 보면, 전국적으로 2005년 11월 이후부터 2008년 3월 사이 총 아파트 수의 약 8.4%에 해당하는 55.6만 세대가 늘어났음을 알 수 있다. 시도별로 보면 서울은 9.1%가 증가하였으며, 충남(14.5%), 대구(11.8%), 대전(10.5%) 지역이 상대적으로 많이 증가한 편임을 알 수 있다. 아파트 수가 총 주택수 중

표 2.1. 시도별 모집단 현황

시도	2005년 총조사인구(A)	2009년 6월 주민등록인구(B)	차이비율 (%)
전국	45,737,011	49,642,216	8.5
서울	9,541,445	10,226,685	7.2
부산	3,428,470	3,553,483	3.6
대구	2,401,409	2,491,478	3.8
인천	2,471,073	2,697,339	9.2
광주	1,371,442	1,427,194	4.1
대전	1,397,798	1,482,157	6.0
울산	1,017,051	1,113,761	9.5
경기	10,091,129	11,359,140	12.6
강원	1,403,986	1,509,484	7.5
충북	1,398,991	1,522,927	8.9
충남	1,792,913	2,024,318	12.9
전북	1,717,527	1,854,180	8.0
전남	1,756,153	1,915,226	9.1
경북	2,479,717	2,670,040	7.7
경남	2,950,140	3,233,094	9.6
제주	517,767	561,710	8.5

표 2.2. 2005년 총조사 이후 아파트 세대수의 변화

시도	2005년 총조사 신규분양 세대수	2005.12-2008.3 증가율(%)	아파트세대
전국	6,628,993	556,416	8.4
서울	1,218,779	110,652	9.1
부산	520,856	40,881	7.8
대구	356,812	42,168	11.8
인천	381,050	36,438	9.6
광주	260,726	12,622	4.8
대전	234,075	24,573	10.5
울산	168,478	11,009	6.5
경기	1,641,505	140,677	8.6
강원	197,686	16,895	8.5
충북	202,990	15,567	7.7
충남	225,113	32,543	14.5
전북	253,607	15,860	6.3
전남	197,949	6,669	3.4
경북	310,112	15,531	5.0
경남	424,420	32,583	7.7
제주	34,835	1,748	5.0

차지하는 비율이 50%를 넘는다는 사실을 고려할 때, 2005년 총조사 자료를 추출틀로 활용할 경우 모집단 가구의 미포함율이 너무 커져 이로 인한 비표본오차가 생길 것을 우려하지 않을 수 없는 실정이다.

이 밖에도 2005년 인구주택총조사를 기초로 작성된 조사구 요도 및 조사구 내 가구 정보가 낡았다는 문제점을 지적할 수 있다. 제 4기 국민건강영양조사를 실시한 결과 이사, 재개발 등의 다양한 사유로 인해

2005년 표본조사구에 속한 가구 중 5~15% 정도의 가구가 변동된 것이 발견되었다. 그러므로 2005년 인구주택총조사를 추출틀로 사용할 경우 기본단위인 조사구에 대한 정보 자체가 일정 부분 현재의 모집단을 제대로 반영하지 못한다고 할 수 있다.

이상에서 지적한 여러 문제점들을 생각할 때 2009년 표본설계에서 2005년 인구주택총조사 자료를 새로운 표본설계를 위한 추출틀로 사용하는 것은 무리한 일이라고 할 수 있다.

2.1. 새로운 추출틀 구축

국민건강영양조사의 목표모집단(target population)과 조사모집단(target population)은 대한민국에 거주하는 모든 가구의 대한민국 국민으로 개념상 일치된다. 그런데 현실적으로 시간이 흐름에 따라 모집단인 국민의 상황은 늘 변하고 있어 제 4기 조사의 표본설계에서 사용된 2005년 인구주택총조사 자료를 추출틀로 사용하면 포함률(coverage rate)이 떨어진다는 문제를 안게 된다. 이는 표본의 모집단에 대한 대표성을 떨어뜨리는 문제이므로 심각하다고 할 수 있다.

포함률을 가능한 높일 수 있는 추출틀 마련을 위해 먼저 전국의 가구를 일반가구와 아파트가구로 구분하여 고려하기로 한다. 아파트의 경우, 새롭게 건설, 분양되는 단지들에 대한 정보를 파악할 수 있으므로 이를 활용하면 2005년 인구주택총조사 기준시점 이후에 모집단에 추가된 세대들을 추출틀에 추가시킬 수 있기 때문이다. 한편 일반가구의 경우, 행정안전부의 주민등록자료를 추출틀로 활용하려 한다. 각각의 경우에 대해 구체적으로 생각해보자.

2.2. 일반가구 추출틀

(1) 추출단위

일반가구 추출틀은 주민등록 자료를 활용한다. 행정안전부에서 관리하는 주민등록통계는 국민 개인의 신고에 의해 작성 관리되는 보고통계로서 거의 실시간으로 모집단의 변동 상황이 갱신되는 대표적인 세대 및 인구 통계이다. 본 연구에서는 질병관리본부 만성병조사팀의 협조를 얻어 2009년 6월의 주민등록통계자료를 활용하였다. 주민등록인구 자료에는 통/반/리별 인구수와 가구수 정보가 들어 있는데, 이것을 기초로 추출단위를 결정하고 추출틀을 구축한다. 인구주택총조사 자료에서는 인구주택조사구가 별도로 구성되어 있었으나 주민등록자료에는 그러한 구분이 존재하지 않는다. 직관적으로 일반가구의 1차추출단위 후보로 주민등록자료에서 제공되는 최소 행정단위인 통/반/리를 생각할 수 있다. 그런데 인구주택조사구와는 달리 통/반/리는 통계적 목적을 위해 구분된 단위가 아니므로 기본적인 사항들을 점검해볼 필요가 있다. 주민등록자료를 통/반/리 단위로 정리하여 기초통계를 구한 결과가 다음의 표 2.3에 나와 있다.

표 2.3을 살펴보면, 전국의 통/반/리는 약 38만 개이며, 통/반/리 당 평균 가구수는 50가구 내외이나 표준편차가 62.2가구로 매우 큰 편이다. 이것은 통/반/리의 규모가 일정하지 않음을 나타낸다. 시도별로 구분하여 보면, 평균가구수나 표준편차의 크기가 매우 다르다. 통계청에서 작성한 인구주택조사구가 60가구 내외로 구성되어 상대적으로 표준편차가 작은 것에 비해 볼 때 통/반/리의 경우는 지역별 일관성이 떨어져서 그대로 1차추출단위로 활용하기에는 무리가 있다.

(2) 추출틀 양식

본 조사를 통해 얻어지는 건강 및 영양 관련 변수들은 대부분 성, 연령과 연관성이 높은 변수들이므로 가급적이면 모집단의 성별, 연령별 구성을 잘 반영하는 추출틀 정보를 마련하는 것이 바람직하다. 현재 확보된 일반가구 추출틀인 주민등록자료의 통/반/리 정보에는 인구수와 가구수 현황만 들어있다. 이게

표 2.3. 2005년 시도별 세대수와 통/반/리 수

지역	세대수	통/반/리 수	평균	표준편차
전국	19,035,656	381,735	50	62.2
서울	4,094,558	98,963	41	21.7
부산	1,311,210	26,835	49	24.3
대구	899,329	20,946	43	38.8
인천	1,013,931	18,675	54	43.2
대전	533,923	12,882	41	21.0
광주	514,393	10,391	50	26.0
울산	393,328	7,964	49	70.6
경기	4,296,503	63,324	68	97.9
강원	607,816	18,146	33	32.0
충북	588,455	12,439	47	49.7
충남	813,470	10,511	77	121.9
전북	713,782	18,244	39	29.1
전남	773,939	14,090	55	83.8
경북	1,058,375	21,302	50	76.2
경남	1,207,210	23,843	51	78.0
제주	215,434	3,180	68	132.1

표 2.4. 일반가구 추출틀 양식

시도	시군구	읍면동	통반리	남자 비율	14세이하 비율	66세이상 비율	가구수	인구수	조사구
경기	가평군	가평읍	경반리	0.508	0.199	0.145	234	439	1
경기	가평군	가평읍	경반리	0.508	0.199	0.145	234	439	2
경기	가평군	가평읍	경반리	0.508	0.199	0.145	234	439	3
경기	가평군	가평읍	금대리	0.508	0.199	0.145	124	243	1
경기	가평군	가평읍	금대리	0.508	0.199	0.145	124	243	2
경기	가평군	가평읍	금대리	0.508	0.199	0.145	124	243	3
경기	가평군	가평읍	금대리	0.508	0.199	0.145	124	243	4
경기	가평군	가평읍	달전1리	0.508	0.199	0.145	173	319	1
경기	가평군	가평읍	달전1리	0.508	0.199	0.145	173	319	2
경기	가평군	가평읍	달전2리	0.508	0.199	0.145	374	691	1
경기	가평군	가평읍	달전2리	0.508	0.199	0.145	374	691	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
서울	강남구	신사동	11통 1, 2반	0.468	0.137	0.071	77	170	1
서울	강남구	신사동	11통3반	0.468	0.137	0.071	61	116	1
서울	강남구	신사동	11통4, 5반	0.468	0.137	0.071	84	187	1
서울	강남구	신사동	11통6반	0.468	0.137	0.071	72	151	1

오와 박진우 (2007)는 순환표본들 사이의 변동을 최소화하기 위한 방안으로 동/읍/면별 연령별 인구수 비율을 층화변수로 활용하여 층화하는 방안을 도입하였다. 이 방법을 그대로 활용하기 위해서는 새롭게 구성된 1차추출단위인 조사구별로 보조정보를 마련해야 한다.

본 연구에서 제시한 1차추출단위는 기본적으로 최소 행정단위인 통/반/리를 기초로 하고 있다. 그러므로 각 조사구가 속한 동/읍/면별 2005년 인구주택총조사의 동/읍/면별 성별, 연령별 구성비율을 그대

표 2.5. 아파트가구 추출률 기초통계

지역	세대수	아파트 단지수	조사구수 조사구수	조사구당 평균세대수	단지당 평균조사구	단지당 평균세대수
전국	6,040,995	13,587	45,558	132.6	3.4	444.6
서울	1,178,160	2,790	9,068	129.9	3.3	422.3
부산	478,875	1,086	3,593	133.3	3.3	441.0
대구	343,719	773	2,565	134.0	3.3	444.7
인천	393,848	927	3,000	131.3	3.2	424.9
광주	225,729	521	1,654	136.5	3.2	433.3
대전	231,785	390	1,789	129.6	4.6	594.3
울산	132,497	309	1,000	132.5	3.2	428.8
경기	1,675,268	3,415	12,646	132.5	3.7	490.6
강원	142,667	371	1,038	137.4	2.8	384.5
충북	150,630	360	1,120	134.5	3.1	418.4
충남	190,274	424	1,423	133.7	3.4	448.8
전북	192,295	466	1,422	135.2	3.1	412.7
전남	105,117	249	764	137.6	3.1	422.2
경북	223,723	569	1,647	135.8	2.9	393.2
경남	359,218	877	2,711	132.5	3.1	409.6
제주	17,190	60	118	145.7	2.0	286.5

로 부가하기로 한다. 이런 방식으로 만든 일반가구 추출률 양식의 예가 다음의 표 2.4에 나와 있다.

2.3. 아파트가구 추출률

(1) 추출단위

아파트가구에 이용되는 2008년 KB 국민은행 아파트 시세조사용 자료에는 세대수, 단지별 평수, 단지별 시세 매매가격 정보가 들어 있다. 또한 인구주택조사구와는 달리 가구 수가 아닌 주택세대수 정보가 담겨 있다. 주택세대수는 물리적인 주택수인데 비해 가구 수는 인구학적 개념의 가구에 대한 수이므로 서로 다르다. 일반적으로 가구 수는 주택세대수보다 큰 편이다.

단지별 총 세대 수를 기준으로 120-240가구를 하나의 조사구로 하는 것을 원칙으로 하여 240가구가 넘는 경우는 2개의 조사구로 분리하였고, 120가구 미만의 단지의 경우는 주변의 단지와 통합하여 하나의 조사구로 간주하였다. 표 2.5에는 아파트가구 추출률에 대한 기초통계가 나와 있는데, 총 세대 수는 6,040,995개, 아파트 단지는 총 13,587개, 인위적으로 구분한 아파트 조사구 수는 총 45,558개이다. 단지당 평균 세대 수는 444.6개이고, 조사구 당 평균 세대수는 132.6개이며, 한 단지 당 조사구 수는 평균 3.4개이다.

다음의 표 2.6에는 아파트 추출률에 있는 정보를 활용하여 추가적인 면적 및 가격 관련 기초통계들을 정리한 결과가 소개되어 있는데, 단지별 평균 면적은 전국 평균 $98.5m^2$ 이고, 서울의 경우 평균 $104.4m^2$ 으로 평균면적이 가장 넓은 편이다. 단지별 $3.3m^2$ 당 가격은 시세 매매일반가(60%), 시세 매매상한가(20%), 시세 매매하한가(20%)를 이용하여 가중평균으로 계산하였는데, 전국의 평균시세는 815천원이었다. 서울이 1,685천원으로 가장 높았고 전남이 235천원으로 가장 낮았다. 단지 내 $99.17m^2$ 이상 세대 비율을 구한 결과 전국 비율은 52.1%였고, 서울이 가장 높은 60.1%, 충북이 가장 낮은 40.1%로 나타났다. 이 결과들은 모두 세대수를 가중값으로 사용하여 분석된 결과이다.

표 2.6. 지역별 아파트 면적 및 3.3m²당 가격 현황

지역	99.17m ²	99.17m ²	3.3m ² 당 가격(천원)			단지당 평균면적(m ²)		
	미만세대수	이상세대수	mean	SD	CV(%)	mean	SD	CV(%)
전국	2,894,752	3,146,243	815.00	14887.05	18.27	98.5	537.9	5.46
서울	470,449	707,711	1685.82	20165.79	11.96	104.4	560.8	5.37
부산	225,387	253,488	451.68	3848.01	8.52	100.1	513.7	5.13
대구	160,183	183,536	464.89	2993.00	6.44	100.7	510.8	5.07
인천	211,873	181,975	674.47	4316.43	6.40	92.8	459.7	4.95
광주	113,652	112,077	307.50	2128.86	6.92	97.5	505.6	5.19
대전	93,511	138,274	483.68	4532.89	9.37	100.6	579.1	5.76
울산	73,320	59,177	453.91	3467.36	7.64	94.1	379.3	4.03
경기	760,711	914,557	941.97	10147.82	10.77	100.9	599.8	5.94
강원	84,511	58,156	301.39	2004.75	6.65	90.9	468.2	5.15
충북	90,170	60,460	339.59	2721.63	8.01	89.6	476.1	5.31
충남	100,089	90,185	391.25	3224.41	8.24	92.3	494.0	5.35
전북	106,076	86,219	290.83	2191.89	7.54	93.2	484.1	5.20
전남	57,818	47,299	235.13	1595.38	6.78	90.0	456.7	5.07
경북	131,640	92,083	321.11	2229.18	6.94	89.7	402.8	4.49
경남	206,423	152,795	412.05	3681.54	8.93	1.5	423.4	4.63
제주	8,939	8,251	390.69	2670.12	6.83	94.0	465.4	4.95

아파트가구의 경우 아파트 세대수 80가구 내외가 되도록 하는 가상적인 조사구를 구성하였다. 단일 아파트 단지의 세대수가 80가구 내외이면 독립적인 조사구로 간주한다. 은마아파트와 같이 한 아파트 단지 내의 세대수가 큰 경우는 총 동수를 감안하여 여러 개의 조사구로 분할한다. 이런 식으로 구성된 아파트 조사구가 1차추출단위가 된다. 2차추출단위는 아파트 조사구 내의 개별 세대이다.

(2) 추출틀 양식

앞서 언급한 대로 일반가구의 경우는 2009년 6월 주민등록 통/반/리별 목록 자료를 추출틀로 사용하고, 아파트가구의 경우는 KB 국민은행 아파트 시세조사용 목록 자료(2008년 4월자료)를 추출틀로 사용한다. 본 조사를 통해 얻어지는 건강 및 영양 관련 변수들은 대부분 성, 연령과 연관성이 높은 변수들이므로 가급적이면 모집단의 성별, 연령별 구성을 잘 반영하는 추출틀 정보를 마련하는 것이 바람직하다. 현재 확보된 일반가구 추출틀인 주민등록자료의 통/반/리 정보에는 인구수와 가구수 현황만 들어있다. 이것에 2005년 인구주택총조사에서의 동/읍/면별 성별, 연령별 구성비를 적용시켜 추가하였다. 다음의 표 2.7은 일반가구 추출틀 양식의 예이다.

한편, 아파트가구 추출틀 자료인 KB 아파트 시세자료에는 인구특성별 인구수 정보가 전혀 들어 있지 않다. 따라서 아파트가구의 경우에는 인구사회적 특성을 상당 부분 반영한다고 여겨지는 아파트 단위면적당, 단지당 평균면적, 단지 내 99.17m²이상 비율 등의 변수를 활용하기로 한다. 또한 지역별 통계를 생산하기 위해 지역도 함께 고려한다. 아파트 추출틀 양식의 예가 다음의 표 2.8에 소개되어 있다.

3. 맺음말

국민건강영양조사는 3년을 주기로 진행되는 조사인 까닭에 3년마다 표본설계도 갱신된다. 표본설계를 위한 추출틀로는 통계청의 인구주택총조사 조사구별 자료가 활용되었는데, 인구주택총조사가 5년 주기로 이루어지다보니 국민건강영양조사와 주기가 서로 맞물리지 않는다. 그러다보니 표본설계 시점에 따

표 2.7. 일반가구 추출틀 양식

시도	시군구	읍면동	단지명	평당가격	평균평수	99.17m ² 이상(%)	세대수	조사구
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	1
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	2
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	3
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	4
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	5
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	6
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	7
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	8
경기	의정부시	가능동	금융	364.21	28.92	0.56	150	1
경기	의정부시	가능동	신도	384.85	27.11	0.44	135	1
경기	의정부시	가능동	신동아파라디움	543.89	33.80	0.87	195	1

표 2.8. 아파트 추출틀 양식

시도	시군구	읍면동	단지명	3.3m ² 당 면적	평균면적	99.17m ² 이상	세대수	조사구
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	1
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	2
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	3
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	4
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	5
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	6
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	7
경기	의정부시	가능동	SK뷰	823.02	31.44	0.75	987	8
경기	의정부시	가능동	금융	364.21	28.92	0.56	150	1
경기	의정부시	가능동	신도	384.85	27.11	0.44	135	1
경기	의정부시	가능동	신동아파라디움	543.89	33.80	0.87	195	1

라 인구주택총조사 추출틀이 낡은 것이어서 포함률이 낮은 경우가 생기는데 2010년부터 시행될 제5기 조사가 이 경우에 해당된다. 인구주택총조사 자료를 추출틀로 활용하는 표본설계에서 설계 시기가 총조사 주기와 맞아 떨어지지 않는 경우 이런 문제는 늘 발생하게 된다.

본 연구는 표본설계 시점이 총조사 실시 시기와 매우 다른 경우 낡은 추출틀을 보완할 수 있는 현실적인 대안을 제시하는데 의의를 두고 있다. 행정자료로서 거의 실시간으로 갱신이 이루어진다고 볼 수 있는 주민등록통계와 분양이 이루어진 신축 아파트 정보를 수집함으로써 모집단에 대한 추출틀을 대표성을 높일 수 있었다. 주민등록통계에서는 통/반/리가, 아파트 세대의 경우 단지가 1차추출단위가 되어야 하는데, 이것들은 인구주택총조사의 조사구와는 달리 통계적 목적에 의해 구분된 단위가 아니라는 문제점이 있다. 이를 보완하기 위해 각 기초단위들의 특성에 따라 가상적인 조사구를 만들 것을 제안하였다.

본 연구에서 사용한 통계 외에 각각의 조사마다 추출틀을 보완하는데 도움이 되는 참고자료들을 구할 수 있는 경우가 대부분이다. 각 경우마다 본 연구에서 제시한 아이디어를 활용한다면 보다 효과적인 해법을 찾을 수 있을 것이다.

참고문헌

- 강현철, 박승열, 김지연, 김인수, 이동수, 황재일, 박민규 (2009). 2008년 서울서베이 추출틀 구축 및 표본추출 사례 연구, <조사연구>, 10, 157-172.

- 이계오, 박진우 (2007). 제4기 국민건강영양조사를 위한 순환표본 설계연구, <조사연구>, **8**, 67-89.
- 질병관리본부 (2008). <2007 국민건강통계>, 질병관리본부.
- 통계청 (2005). <인구주택총조사>, 통계청.
- 한국통계학회 (2010). <2009년 국민여행실태조사 표본설계 및 모수추정 용역 최종보고서>, 한국통계학회.
- Hartley, H. O. (1962). Multiple frame surveys, In *Proceeding of the Social Statistics Section, American Statistical Association*, 203-169.
- Hartley, H. O. (1972). Multiple frame methodology and selected applications, *Sankhya, Series C*, **36**, 99-118.
- Iachan, R. and Dennis, M. L. (1993). A multiple frame approach to sampling the homeless and transient population, *Journal of Official Statistics*, **9**, 747-764.

Construction of Sampling Frames for the 5th Korea National Health and Nutrition Examination Survey

JinWoo Park¹ · Jongseok Byun² · Minkyu Park³

¹Department of Applied Statistics, University of Suwon; ²Department of Statistics, Hanshin University

³Department of Statistics, Korea University

(Received August 2010; accepted September 2010)

Abstract

One of the problems in designing the fifth Korean National Health and Nutrition Examination Survey(KNHNES) is the lack of an appropriate sampling frame. Due to the significant time difference, we expect eight severe sampling frame errors if we use the sampling frame obtained from the latest population and housing census that was conducted in 2005. Thus, the construction of an appropriate sampling frame for the fifth KNHNES is crucial for a successful survey. We considered the construction of a sampling frame that overcomes the limitations of the 2005 population and housing census based frame. For the construction of eight new sampling frames, we considered the use of multiple sampling frames in which the frame for the apartment households and the frame for the general households are obtained from different sources.

Keywords: Sampling frame, frame error, coverage rate.

¹Corresponding author: Professor, Department of Applied Statistics, University of Suwon, Suwon 445-743, Korea. E-mail: jwpark@suwon.ac.kr