

등록센서스 기법개발 연구

연구용역 최종보고서

2009. 11. 30.

통 계 청
한국인구학회

등록센서스 기법개발 연구

연구용역 최종보고서

2009. 11. 30.

통 계 청
한국인구학회

제 출 문

통계청장 귀하

본 보고서를 “등록센서스 기법개발 연구” 용역
과제의 최종 연구결과물로 제출합니다.

2009년 11월 30일

계약기관 한국인구학회 전광희 ㉠

연구진

책임 연구자	전 광 희 (충남대학교 교수)
연구자	김 규 성 (서울시립대학교 교수)
	김 두 섭 (한양대학교 교수)
	김 정 석 (동국대학교 교수)
	김 태 헌 (한국교원대학교 교수)
	박 진 우 (수원대학교 교수)
	이 건 (서울시립대학교 교수)
	이 석 훈 (충남대학교 교수)
	최 진 호 (아주대학교 교수)
연구보조원	심 지 은 (충남대학교 대학원 박사과정)
	한 만 규 (충남대학교 대학원 석사과정)
	오 유 진 (충남대학교 대학원 석사과정)
	김 태 순 (충남대학교 대학원 석사과정)

최종보고서 요약문

연구과제명	등록센서스 기법개발 연구
주 제 어	등록센서스, 행정자료, 매칭, 추정기법, 보정방안, 표본조사, 대체이 용, 주민등록, 건축물/주택대장
연 구 기 간	2009년 6월 10일~2009년 11월 30일
연 구 기 관	한국인구학회
연구진구성	김두섭, 이진, 이석훈, 김규성, 전광희, 김정석, 박진우, 최진호, 김태현
(국문 요약)	
<p>이 연구는 통계청이 최근 구상해 온 2015년 등록센서스를 구체적으로 추진하는데 필요한 기법 개발에 관한 최종 결과이다. 보고서는 3개 부문으로 구성되며, 이들은 각각 (1) 등록센서스의 통계적 · 수리적 기초의 연구 (2) 등록센서스 실시로 자료원의 변동이 생겨나면서 예상되는 통계이용상의 변화 검토 (3) 2015년 등록센서스 추진을 위한 법령과 제도적 기반의 검토와 구체적인 단계적 추진절차를 제안하기 위한 연구들로 구성된다.</p> <p>제1부는 5개의 논문으로 구성된다. 처음의 3개 논문은 2010년 인구주택총조사 시험조사 자료와 행정자료로는 주민등록자료, 건축물대장자료를 이용하고 있으며, (1) 행정자료 활용을 위한 자료매칭의 최적방안을 논의하고 (2) 매칭된 자료의 불일치 항목분석과 매칭되지 않은 자료의 항목을 분석하며 (3) 지역별, 특성별 인구, 가구, 주택의 수를 추정하는 기법을 개발하는데 할애되었다. 나머지 2개 논문은 각각 행정자료와 실제 자료(총조사 시험조사 자료)의 불일치 부문에 대한 편집 및 보정방법을 구체화하고, 행정자료를 활용할 수 없는 조사항목에 대하여 그것을 조사하는 방법론을 대하여 논의한다.</p> <p>제2부는 두 개의 논문으로 구성된다. 하나는 등록센서스의 결과표 작성과 통계자료의 시계열 변동에 따른 보완방안을 논의한다. 다른 하나는 등록센서스의 실시로 모집단의 표본추출 틀이 변화하면서, 향후 어떤 방법으로 표본조사를 실시하여야 할 것이며, 현재 이용되고 있는 조사구 설정을 계속하여야 하는가 아니면 그것을 어떻게 변경하는 것이 바람직한가를 논의한다. 마지막으로 제3장은 두 개의 논문으로 구성된다. 첫 번째 논문은 현재의 통계법, 주민등록법, 건축물대장법이 등록센서스를 실시하는데 어떠한 문제점이 있으며, 그것을 향후 어떤 방향으로 개선하는 것이 바람직한가를 논의한다. 마지막 논문은 2015년의 등록센서스를 실시하는데 있어서 필요한 구체적 조치와 절차가 무엇인가(예: 센서스 표본조사의 추출률을 10%에서 20%로 확대, 등록센서스 기본법의 제정, 통계청과 여타 정부부처간의 통계조정 및 협력)를 논의한다.</p>	

(English Abstract)

This study reports a final result about the development of techniques needed to implement a Korean register-based census of 2015, which Statistics Korea has envisioned in recent years. The report consists of three parts: (1) studies on statistical and mathematical foundation of register-based census, (2) a review of expected change in use of statistics concomitant with change in data sources that takes place with the implementation of a register-based census and (3) a review of existing legal and institutional bases and specific step-by-step procedures needed to implement a Korean register-based census of 2015.

PART I consists of five separate chapters. The first three chapters use survey data from census test for the Population and Housing Census of 2010 and administrative data from population register and building register. The first chapter discusses how to develop an optimal way of matching administrative-purpose data with each other or with data from census tests or statistical surveys. The second chapter focuses on analyzing an entire set of personal information for non-matched cases as well as a set of non-matched personal information for matched cases. Lastly, the third chapter develops how to estimate the number of persons, households, and dwellings in total, or by regions and personal attributes.

In the two last chapters, PART I, the report tries to facilitate the implementation of a register-based census by discussing (1) how to edit non-matched personal information for matched cases and impute the missing cases by different methods and (2) how to obtain data on census variables which can not be directly derived from administrative-purpose data.

PART II consists of two separate chapters. The first chapter is about the generation of statistical tables from a register-based census and the harmonization of time-series census statistical data. The second chapter is to discuss how to carry out a large-scale statistical survey when it has to use the population census as a sampling frame and the census design shifts from traditional to a register-based one, and whether an integrated census, say, administrative data-based census reinforced by the use of sample survey, will continue to use the concept of census enumeration district (CED) or otherwise abolish it to develop an alternative idea.

PART III consists of two separate chapters. The first chapter discusses major problems presented by Korean current laws governing population and housing/dwelling registration and what Statistics Korea will have to

do to implement a successful register-based census of 2010 in Korea. The second chapter discusses a specific set of measures and procedures which will have to be taken into full consideration, in particular an increase of sampling fraction from 10% to 20% in a census sample survey to be carried out with a register-based census, the legislation of a new census law, and the coordination of Statistics Korea with other government ministries (Ministry of Public Administration and Security Affairs, Ministry of Land, Transport, and Maritime Affairs, and Supreme Court, etc) to facilitate a smooth transition from a traditional census to a register-based one in 2015.

보고서 목차

제1부 등록센서스의 통계적·수리적 방법론	1
제1장 행정자료 활용을 위한 자료 매칭의 최적방안 연구	3
제1절 머리말	3
제2절 분석범위와 방법론	5
제3절 매칭작업의 결과	17
제4절 인구/주택 매칭의 항목별 매칭률 평가	54
제5절 결론	69
제6절 정책적 건의사항	92
제2장 매칭결과 불일치 항목 분석 및 비 매칭자료 항목 분석	99
제1절 머리말	99
제2절 분석범위와 방법론	100
제3절 인구 매칭 자료의 불일치 항목 분석	101
제4절 주택 매칭 자료의 불일치 항목 분석	112
제5절 비 매칭 자료의 특성 분석	121
제6절 결론	129
제7절 정책적 건의사항	129
제3장 지역별, 특성별 인구, 가구, 주택 추정기법	156
제1절 머리말	156
제2절 등록센서스와 추정기법의 개관	156
제3절 추정의 범위와 방법론	157
제4절 인구의 지역별, 특성별 추정기법	158
제5절 가구수의 지역별, 특성별 추정기법	181
제6절 주택수의 지역별, 특성별 추정기법	189
제7절 결론	200

제8절 정책적 건의사항	204
제4장 매칭결과 자료 간 불일치 보완방법 및 자료보완에 따른 보정방법	206
제1절 머리말	206
제2절 레코드 매칭률 향상 방안	206
제3절 불완전 매칭시 매칭률과 영역별 규모 추정	216
제4절 불일치 자료 보완에 따른 보정 방법	223
제5절 정책적 건의사항	224
제5장 등록센서스 대체이용 불가능 조사항목의 표본조사 도입방안	225
제1절 머리말	225
제2절 센서스의 다양한 유형	228
제3절 등록센서스와 표본조사의 역할	234
제4절 현행 인구주택총조사의 조사방법론	243
제5절 행정자료 대체이용 가능성 평가 검토	252
제6절 종합 및 정책적 건의사항	263
제2부 등록센서스 이행에 따른 통계이용의 변화	273
제6장 등록센서스의 결과표 작성과 통계자료 시계열 변동 보완방안	275
제1절 논의의 출발점과 전제	275
제2절 통계품질관리기준으로서의 일관성	276
제3절 항목 및 공표결과의 시계열적 변화	280
제4절 예상결과표의 예시	288
제5절 시계열적 일관성 확보 노력	290
제7장 센서스의 표본조사 모집단 기능과 기존 마이크로데이터 활용 대체 방안	303
제1절 머리말	303
제2절 기존 인구주택총조사와 표본조사 추출틀의 역할	305
제3절 외국사례 검토	320

제4절 등록센서스 기반 추출틀 작성 방안	327
제5장 연구결과 요약	343
제6절 정책적 제안	344
제3부 등록센서스 추진을 위한 제도·법률적 기반개선과 단계적 추진계획 로드맵의 작성	347
제8장 현행 센서스와 행정자료의 주요 개념 비교분석 및 등록센서스 제도적·법률적 기반개선의 과제	349
제1절 머리말	349
제2절 행정자료와 현행 인구주택총조사 자료의 개념 비교분석	349
제3절 등록센서스 추진에 따른 행정자료의 한계성	361
제4절 등록센서스 추진을 위한 제도적·법률적 기반개선의 과제	366
제5절 연구결과의 요약 및 전반적 평가	373
제9장 2015년 등록센서스 추진을 위한 단계적 추진계획 및 구체적 과제의 검토	377
제1절 머리말	377
제2절 연구범위와 등록센서스 로드맵 작성의 절차	377
제3절 로드맵 1: 행정자료 확보와 보정방안 마련	382
제4절 로드맵 2: 관계기관(행안부, 국토해양부) 협력체제 구축	393
제5절 로드맵 3: 개인정보보호 및 행정자료 통합데이터베이스 구축	397
제6절 통계청 「2015년 등록센서스 기본계획」의 전반적 평가	399
제7절 정책적 건의사항	405
부록	411
부록 I: A-H 세계 각국의 등록센서스와 센서스 표본조사	413
A. 스페인의 행정자료와 전수조사 통합센서스	413
B. 대만의 등록센서스	416

- C. 독일의 등록센서스 416
- D. 이스라엘의 통합센서스 441
- E. 싱가포르의 등록센서스 461
- F. 덴마크와 핀란드의 등록센서스 468
- G. 영국의 One Number Census와 센서스 범위조사의 개관 483
- H. 미국의 센서스 표본조사-American Community Survey 496
- I. 프랑스의 순환센서스 표본조사 506

부록 II: 대만 등록센서스의 세부절차 514

- A. 공무등기자료 매칭작업 514
- B. 2010년 인구주택총조사 표본설계 519

부록 III: 싱가포르의 등록센서스 절차 529

- A. 싱가포르 2000년 센서스 인구주택데이터베이스 개발 529
- B. 싱가포르 2000년 센서스 표본조사 기획 및 준비 540

부록 IV: 독일의 2011년 등록센서스 관련 법률 552

- A. 독일의 2011년 등록센서스 법률 및 법적 기반 552
- B. 2011년 센서스 준비법 552
- C. 2011년 센서스 법 561

참고문헌 587

제1부

등록센서스의 통계적 · 수리적 방법론

제1장 행정자료 활용을 위한 자료 매칭의 최적방안 연구

제1절 머리말

인구센서스는 인구현상의 파악과 분석을 위한 가장 기초적인 자료로 손꼽힌다. 우리나라에서 근대적인 의미의 인구센서스는 일제하인 1925년에 『간이국세조사』라는 명칭으로 처음 시작되었으며, 2005년에 이르기까지 대부분 5년의 간격으로 총 17회 실시되었다(김민경, 2000, 2002; 권태환·김두섭, 2002, 김두섭·박상태·은기수, 2002). 센서스는 1960년부터 인구뿐만 아니라 주택에 관련되는 조사항목들을 포함하게 되었고, 전수조사(全數調査)와 아울러 표본조사를 병행하여 실시하고 있다. 그 동안 다양하게 불리던 센서스는 1990년부터 『인구주택총조사』라는 명칭으로 시행되고 있다.

일반적으로 우리나라의 센서스 자료는 조사의 완전성과 질이 매우 우수한 것으로 알려져 있다. 그리고 지난 반세기 동안 경제개발정책이나 인구정책의 수립과 수행과정에서 매우 중요한 기초자료로 활용되었다. 다양한 행정적 목적을 위해 활용되는 것과 아울러 각종 사회조사에서 표본을 산출할 수 있는 모집단을 제공하여 온 것도 센서스의 중요한 기능으로 지적될 수 있다.

그런데 그 동안의 사회적 여건의 변화로 센서스의 시행에 어려움이 가중되고 있다. 우선 조사원의 임금상승으로 인하여 조사비용이 급증하는 추세를 보인다. 주거환경과 생활양식의 변화, 그리고 사생활보호 의식이 강해지면서 조사의 수행이 어려워지고 응답률이 낮아지는 현상이 나타나고 있다.

이에 따라 전통적 방식의 센서스에 수반되는 문제점들을 극복하기 위하여 다양한 방법들이 모색되고 있다. 대표적인 예로, 북유럽의 여러 국가들은 센서스의 대안으로 행정자료를 활용하는 가상센서스(virtual census)를 1970년대 이래 시행해오고 있다. 아시아에서는 싱가포르에서 행정자료를 활용하는 방안이 시도된바 있다. 대만은 2010년 센서스를 행정자료를 활용하는 방식으로 시행하기로 결정하였다. 또한 우리나라도 2015년부터 행정자료를 기반으로 하는 등록센서스와 표본조사를 병행하여 시행하기 위하여 준비하는 과정에 있다.

우리나라는 행정자료를 기반으로 하는 등록센서스를 시행하기에 유리한 몇 가지 조건을 지니고 있다. 우선, 주민등록, 가족관계등록부, 건축물대장 등

개인, 가구 및 주택에 관하여 다양한 정보를 포함하는 행정자료가 가용하다. 아울러 지난 몇 년간 전자정부의 구현을 목표로 이루어진 종합정보망의 구축으로 행정자료의 활용이 용이해졌다는 점도 지적할 수 있다.

행정자료를 활용하여 센서스를 보완하거나 대체하기 위해서는 그 시행과정에서 수반되는 문제점에 대한 이해와, 이를 극복할 수 있는 방안의 모색이 전제되어야 한다. 이 연구는 주민등록 및 건축물대장 자료를 활용하여 등록 센서스를 시행하는 과정에서 나타나는 문제점을 파악하고, 이를 극복하기 위한 대안을 모색하는 노력의 하나로 계획되었다. 보다 구체적으로, 이 장에서는 2008년 주민등록인구와 2008년에 수행된 제2차 및 제3차 시험조사인구를 매칭하는 작업과 아울러 이에 대한 평가가 이루어진다. 그리고 주택자료를 매칭하는 작업은 건축물대장과 제2차 및 제3차 시험조사 자료를 활용하여 이루어진다.

이 연구의 계획단계에서는 주민등록과 가족관계등록부 등 행정자료 간의 매칭작업도 시도하기로 하였다. 그리고 가구(또는 세대)를 매칭하는 방안도 고려되었다. 그러나 개인정보 비밀보호 등의 이유로 인하여 이 연구에서 가족관계등록부를 활용할 수 없게 되었고, 아울러 제4차 시험조사 자료가 가용하지 않게 됨에 따라 매칭작업의 대상과 범위가 축소되었음을 밝혀둔다.

제2절 분석범위와 방법론

1. 시험조사지역의 매칭자료

가. 인구자료

이 연구에서 인구의 매칭작업을 위해 사용된 자료는 제2차 및 제3차 시험조사 자료와 주민등록 자료이다. 등록센서스를 준비하는 과정에서 사전분석의 목적을 위해 2008년 실시된 제2차 시험조사는 부산광역시, 경기도 및 강원도의 6개 지역에서 이루어졌다. 제3차 시험조사는 광주광역시, 충청북도와 경상남도의 6개 지역을 대상으로 2008년에 시행되었다. <표 1-1>에는 제2차 및 제3차 시험조사지역의 주민등록인구와 시험조사인구가 제시되어 있다.

<표 1-1>을 보면, 제 2차 시험조사의 인구는 78,359명으로 2008년 주민등록 인구보다 10.4%가 적다. 주민등록인구와의 격차는 특히 강원도 동해시 묵호동과 강원도 인제군 인제읍에서 두드러지는 것으로 확인되었다. 이들 두 지역에서 집계된 시험조사의 인구는 주민등록인구의 83.0%와 77.9%에 불과하다. 제2차 시험조사가 이루어진 6개 지역에서 주민등록인구보다 더 많은 인구가 집계된 지역은 한 곳도 없다. 이는 제2차 시험조사에서 부재 또는 불응 가구가 상당수 제외되었다는 사실과 관련이 있는 것으로 판단된다.

<표 1-1> 시험조사지역의 주민등록인구와 시험조사인구, 2008

행정구역	행정구역 코드 ¹⁾	2008년 주민등록 인구 (A)	2008년 시험조사 인구 (B)	A-B	(B/A)×100
제2차 시험조사					
부산 부산진구 범천1동	2105077	9,712	9,399	313	96.8
부산 동래구 명장1동	2106063	18,321	16,924	1,397	92.4
경기 용인시 기흥동	3119254	13,283	11,925	1,358	89.8
경기 광주시 광남동	3125053	31,772	28,709	3,063	90.4
강원 동해시 묵호동	3204059	4,786	3,971	815	83.0
강원 인제군 인제읍	3239011	9,540	7,431	2,109	77.9
합계		87,414	78,359	9,055	89.6
제3차 시험조사					
광주 북구 문흥1동	2404065	20,958	21,575	-617	102.9
광주 광산구 월곡1동	2405059	16,637	16,361	276	98.3
충북 제천시 교동	3303051	18,433	17,707	726	96.1
충북 음성군 소이면	3337031	3,423	3,176	246	92.8
경남 김해시 활천동	3807058	26,381	27,704	-1,323	105.0
경남 고성군 거류면	3834043	5,343	4,842	501	90.6
합계		91,175	91,365	-191	100.2

주: 1) 이 표에 제시된 행정구역 코드는 통계청에서 사용하는 것이며, 행정안전부의 행정동 코드와는 그 체계가 다른 것이다.
 2) 시험조사의 인구는 중복 등 오류가 포함된 건을 제거하고 작성되었음.

제3차 시험조사에서는 제2차 시험조사에 비해 상대적으로 충실하게 인구집계가 이루어진 것으로 판단된다. <표 1-1>에서 보는 것처럼, 제3차 시험조사의 인구는 주민등록인구보다 0.2%가 많은 91,366명이다. 광주광역시 북구 문흥1동과 경상남도 김해시 활천동에서 집계된 시험조사의 인구는 2008년 주민등록인구에 비해 각각 2.9%와 5.0%가 많은 것으로 확인되었다. 제3차 시험조사에서 주민등록인구에 비해 상대적으로 가장 낮은 수준의 인구집계가 이루어진 지역은 경상남도 고성군 거류면으로, 그 비율이 90.6%로 나타났다. <표 1-1>에서는 부산광역시와 광주광역시의 네 개 지역에서 시험조사 인구와 주민등록인구의 격차가 다른 지역에 비해 상대적으로 작게 나타나는 경향이 관찰된다.

주민등록인구와 센서스(총조사)인구 간에는 차이가 나타나기 마련이다. 주민등록인구는 당사자의 신고에 의해서 작성된다는 점에서 센서스인구와 구별된다. 출생, 사망, 이동에 관한 신고가 지연되거나, 연령과 거주지 등을 실수나 고의로 사실과 다르게 신고하더라도 이를 보정하기가 매우 어렵다. 그럼에도 불구하고 주민등록인구는 출생률, 사망률, 혼인율, 이혼율 등의 인구동태율을 작성할 때 기준인구로 활용되며, 장래인구를 추계하는 과정에서도 중요한 참고자료가 된다.

<표 1-2>에는 제2차 및 제3차 시험조사지역의 2005년 주민등록인구와 센서스인구가 제시되어 있다. 이는 <표 1-1>에서 나타나는 주민등록인구와 시험조사인구와의 차이를 이해하기 위한 노력의 일환으로 작성되었다. <표 1-2>를 보면, 제2차 시험조사지역과 제3차 시험조사지역에서 2005년 센서스인구가 각각 주민등록인구의 98.1%와 101.0%로 집계되었음을 확인할 수 있다. 유동인구가 많을 것으로 짐작되는 강원도 동해시 묵호동과 경상남도 고성군 거류면, 경기도 용인시 기흥동을 제외하고는 아홉 개 조사지역에서 두 인구의 격차가 7%를 넘지 않는다. 이에 따라 2008년에 실시되었던 시험조사가 2005년 센서스에 비해 그 완전도에 있어서 상대적으로 미흡하였다는 잠정적인 판단을 내릴 수 있다.

<표 1-1>과 <표 1-2>에 제시된 각 지역별 인구를 비교해보면, 일부 지역에서 2005-2008년 기간에 커다란 변화가 있었음을 확인할 수 있다. 경기도 용인시 기흥동의 인구가 급속하게 증가한 것은 이 지역에 아파트 단지가 개발되면서 유입인구가 크게 늘어난 결과로 짐작된다. 반면, 경상남도 김해시 활천동의 인구가 불과 3년 동안에 대폭적으로 줄어든 것은 이 지역에서 이루어진 시험조사가 활천동의 일부 지역(구 어방읍)만을 대상으로 이루어졌기 때문으로 판단된다. 이에 대한 자세한 설명은 시험조사지역에 대한 통계청의 내부자료를 확인하여야만 가능하다.

<표 1-2> 시험조사지역의 주민등록인구와 센서스인구, 2005

행정구역	행정구역 코드	2005년 주민등록 인구 (A)	2005년 센서스 인구 (B)	A-B	(B/A)×100
제2차 시험조사					
부산 부산진구 범천1동	2105077	10,240	10,118	122	98.8
부산 동래구 명장1동	2106063	19,529	18,963	566	97.1
경기 용인시 기흥동	3119254	8,810	9,632	-822	109.3
경기 광주시 광남동	3125053	28,671	28,383	288	99.0
강원 동해시 묵호동	3204059	4,827	3,956	871	82.0
강원 인제군 인제읍	3239011	9,730	9,213	517	94.7
합계		81,807	80,265	1,542	98.1
제3차 시험조사					
광주 북구 문흥1동	2404065	22,115	23,008	-893	104.0
광주 광산구 월곡1동	2405059	17,135	17,568	-433	102.5
충북 제천시 교동	3303051	17,228	16,179	1,049	93.9
충주 음성군 소이면	3337031	3,565	3,329	236	93.4
경남 김해시 활천동	3807058	46,165	47,960	-1,795	103.9
경남 고성군 거류면	3834043	4,889	4,171	718	85.3
합계		111,097	112,215	-1,118	101.0

주: 이 표에 제시된 행정구역 코드는 통계청에서 사용하는 것이며, 행정안전부의 행정동 코드와는 그 체계가 다른 것이다.

나. 주택자료

이 연구에서 주택의 매칭작업을 위해 사용된 자료는 제2차 및 제3차 시험조사 자료와 건축물대장이다. <표 1-3>에는 제2차 및 제3차 시험조사지역의 건축물대장과 시험조사에서 집계된 주택에 관한 자료가 제시되어 있다. 그런데 아파트는 다른 형태의 주택에 비해 조사의 편이성이나 완전도에 있어서 다른 형태의 주택과 확연하게 구분된다. 이에 따라 주택 자료를 아파트와 기타 주택의 두 범주로 구분하여 제시하였다. 기타 주택에는 단독주택(일반단독, 다가구단독, 영업겸용단독), 연립주택, 다세대주택, 비거주용 건물내 주택 등이 포함된다.

<표 1-3>을 보면, 제2차 및 제3차 시험조사지역에서 집계된 주택의 규모는 각각 건축물대장의 97.3%와 100.3%에 해당된다. 주택을 두 범주로 나누었을 때, 두 자료간 아파트 규모의 일치수준이 기타 주택의 경우보다 약간 높은 것으로 나타났다. 그러나 행정구역별로 주택의 규모를 살펴보면, 아파트는 부산광역시 동래구 명장1동, 강원도 동해시 묵호동, 강원도 인제군 인제읍, 그리고 경상남도 고성군 거류면에서 10% 이상의 격차가 발견된다. 기타 주택의 경우는 12개 시험조사 지역 중 6개 지역에서 두 주택자료의 격차가 10%를 상회한다. 특히 부산광역시 부산진구 범천1동과 강원도 동해시 묵호

동의 시험조사에서 집계된 주택의 규모는 각각 건축물대장에 등재된 주택 규모의 173.8%와 70.1%에 해당되는 것으로 밝혀졌다. 두 주택자료 간의 이 같은 커다란 격차는 매칭작업의 장애요인으로 작용하고, 이들 지역에서 주택자료 매칭률을 현저하게 낮추는 결과를 초래할 것으로 판단된다.

<표 1-3> 시험조사지역의 건축물대장과 조사 주택 현황

행정구역	아파트			기타 주택			합계		
	건축물 대장(A)	조사 자료(B)	(B/A)%	건축물 대장(A)	조사 자료(B)	(B/A)%	건축물 대장(A)	조사 자료(B)	(B/A)%
제2차 시험조사									
부산 부산진구 범천1동	1,868	1,806	96.7	1,139	1,980	173.8	3,007	3,786	125.9
부산 동래구 명장1동	889	1,208	135.9	3,240	2,893	89.3	4,129	4,101	99.3
경기 용인시 기흥동	3,468	3,182	91.8	803	695	86.6	4,271	3,877	90.8
경기 광주시 광남동	4,748	4,682	98.6	3,788	3,599	95.0	8,536	8,281	97.0
강원 동해시 묵호동	562	411	73.1	1,911	1,339	70.1	2,473	1,750	70.8
강원 인제군 인제읍	422	497	117.8	2,292	2,156	94.1	2,714	2,653	97.8
합계	11,957	11,786	98.6	13,173	12,662	96.1	25,130	24,448	97.3
제3차 시험조사									
광주 북구 문흥1동	4,882	4,881	100.0	564	589	104.4	5,446	5,470	100.4
광주 광산구 월곡1동	3,058	3,056	99.9	985	927	94.1	4,043	3,983	98.5
충북 제천시 교동	4,396	4,768	108.5	1,539	1,158	75.2	5,935	5,926	99.9
충북 음성군 소이면	456	455	99.8	1,238	1,350	109.1	1,694	1,805	106.6
경남 김해시 활천동	5,029	4,586	91.2	2,340	2,858	122.1	7,369	7,444	101.0
경남 고성군 거류면	485	534	110.1	1,923	1,808	94.0	2,408	2,342	97.3
합계	18,306	18,280	99.9	8,589	8,690	101.2	26,895	26,970	100.3

2. 자료매칭의 방법론

가. 인구자료 매칭

주민등록 자료와 시험조사 자료를 매칭하기 위해서는 두 자료에 공통적으로 담겨있는 정보를 확인하고, 이를 매칭키(matching key)로 활용하여야 한다. 이 연구를 위해 통계청으로부터 제공받은 2008년 주민등록 자료와 제2차 및 제3차 시험조사 자료에 담겨있는 인구와 가구 관련 정보들은 <표 1-4>와 같이 정리될 수 있다. 그리고 이 연구에서 두 인구자료의 매칭작업을 위해 사용된 매칭키는 주소, 성명, 생년 및 생월, 성별, 가구주와의 관계이며, 매칭작업시 고려한 매칭키의 우선고려순위를 <표 1-5>에 제시하였다.

인구자료의 매칭키로 사용된 항목들 중에서 주소, 성명, 생년 및 생월, 그리고 성별은 매칭키의 작성과정에 대해 추가적인 설명을 요하지 않는 명쾌한 개념이다. 그러나 가구주와의 관계는 통계청에서 사용하는 개념이고, 세

대주와의 관계는 법적개념으로 이 두 자료를 매칭하기 위하여 가구주와의 관계 코드는 주민등록 자료의 세대주와의 관계를 변환하여 시험조사 자료와 일치시키는 작업이 이루어졌다. 변환작업의 자세한 내용은 <표 1-6>에 제시되어 있다.

<표 1-5> 주민등록 자료와 시험조사 자료에 포함되어 있는 인구 관련 정보

주민등록 자료		시험조사 자료	
제2차 시험조사지역	제3차 시험조사지역	제2차 시험조사지역	제3차 시험조사지역
법정구역 코드	법정구역 코드	가구원 일련번호	가구원 일련번호
세대 번호	세대 번호	가구원수	가구원수
성명	성명	법정구역 코드	법정구역 코드
성별	성별	행정구역 코드	행정구역 코드
생년월일	생년월일		행정동
세대주 성명	세대주 성명	법정동	법정동
세대주 성별		법정리	법정리
세대주 생년월일	세대주 생년월일	본번 표준화	본번 표준화
세대주와의 관계		부번 표준화	부번 표준화
세대주와의 관계 코드	세대주와의 관계 코드	공동주택명	공동주택명
현주소	현주소	동명	동명 표준화
법정동		호명	호명 표준화
법정리		조사구	조사구
번지		거처 번호	거처 번호
통		가구 번호	가구 번호
건물명	건물명	가구원 번호	가구원 번호
동명	동명	일련 번호	일련 번호
호명	호명 및 호명 표준화	성명	성명
본번	본번 및 본번 표준화	성별	성별
부번	부번 및 부번 표준화	세는 나이	세는 나이 및 생년월
1년전 시군	1년전 거주지 코드	가구주와의 관계	가구주와의 관계
1년전 주소	5년전 거주지 코드	가구 구분	가구 구분
5년전 시군		주인가구 여부	주인가구 여부
5년전 주소			타지 주택 소유 여부
관할 행정동	행정동 코드		주택 점유형태
행정구역 코드	행정구역 코드		교육 수준
			혼인상태

<표 1-6> 인구자료의 매칭작업을 위해 사용된 매칭키 및 우선고려순위

항목	매칭작업		비고
	매칭키	우선고려순위	
주소	본번 표준화 (AM)	2	
	부번 표준화 (AS)	3	
	동명표준화 (AD)	3	
	호명 표준화 (AR)	3	
성명	성 (NL)	1	
	이름 (NE)	1	
생년 및 생월	생년 (BY)	2	제2차 시험조사의 세는 나이를 생년으로 변환시켰음
	생월 (BM)	2	
성별	성별	4	
가구주와의 관계	가구주 관계 코드	4	주민등록 자료의 세대주와의 관계를 변환하여 시험조사 자료와 일치시켰음

<표 1-7> 시험조사 자료와 주민등록 자료로부터 가구주와의 관계 코드를 작성하는 작업

시험조사 자료 가구주와의 관계	주민등록 자료 세대주와의 관계	가구주와의 관계 코드
가구주	본인	1
가구주의 배우자	처, 남편	2
자녀	자	3
자녀의 배우자	자부, 사위	4
가구주의 부모	부, 모, 양부, 양모	5
배우자의 부모	장인, 장모, 시부, 시모	6
손자녀, 그 배우자	손, 외손, 손부, 손서	7
증손자녀, 그 배우자	증손	8
조부모	조부, 조모, 장조부, 장조모, 시조모, 외조부, 외조모	9
형제자매, 그 배우자	형, 제, 누이, 매, 오빠, 형수, 제수, 처남, 처제, 형부, 매부, 제부, 처형, 처남댁, 시누이, 올케, 시숙, 동서	10
형제자매의 자녀, 그 배우자	조카, 처조카	11
부모의 형제자매, 그 배우자	백부, 백모, 숙부, 숙모, 고모, 고모부, 이모, 이모부, 외숙, 외숙모, 처고모, 처이모	12
기타 친인척	친척, 외가친척, 이종, 시가친척, 이질, 처가친척, 질부, 종제, 외종, 고종, 종형, 종제수, 종형수, 당질, 처이종, 종매	13
기타 동거인	동거인	14

나. 주택자료 매칭

이 연구를 위해 통계청으로부터 제공받은 2008년 건축물대장 자료와 제2차 및 제3차 시험조사 자료에 담겨있는 주택 관련 정보들은 <표 1-7>과 같이 정리될 수 있다. 그리고 이 연구에서 두 주택자료의 매칭작업을 위해 매칭키로 사용된 정보는 주소, 층, 건물명, 면적, 건축연도코드이다. <표 1-8>에는 매칭키와 매칭 작업시 고려한 매칭키의 우선고려순위를 제시하였다. 주택자료의 매칭작업과정에서 주요 특징은 동표기 변환, 건축연도 재코드화였다. 주택자료의 매칭키로 사용된 항목들 중에서 주소, 층, 건물명, 면적, 건축연도코드에서 주소의 동명/호명의 표기는 건축물대장자료와 시험조사 자료의 표기가 상이하어 이를 일치시키는 작업이 이루어졌으며, 건축물대장의 건축연도는 시험조사 자료에서 조사하는 범주와 일치하도록 변환작업을 실시하였다. 변환작업의 자세한 내용은 <표 1-9>에 제시하였다.

<표 1-8> 건축물대장과 시험조사 자료에 포함되어 있는 주택 관련 정보

건축물대장		시험조사 자료	
제2차 시험조사지역	제3차 시험조사지역	제2차 시험조사지역	제3차 시험조사지역
일련번호	일련번호	분석 일련번호	분석 일련번호
	지번별 건물수		빈집 여부
법정구역 코드	일반집합 구분	일련번호	일련번호
	법정구역 코드	법정구역 코드	법정구역 코드
	행정구역 코드	행정구역 코드	행정구역 코드
번지	번지	행정동	행정동
건물명	건물명	법정동	법정동
동명	동명	법정리	법정리
호명	호명	거주 여부	거주 여부
대지 구분	대지 구분	본번 표준화	본번 표준화
본번 표준화	본번 표준화	부번 표준화	부번 표준화
부번 표준화	부번 표준화	공동주택명	공동주택명
	호명 표준화	동명	동명 표준화
주거용 연면적	주거용 연면적	호명	호명 표준화
완공일	완공일	조사구	조사구
주택 종류	주택 종류	거처 번호	거처 번호
		거처 종류	거처 종류
		주거용 연면적	주거용 연면적
		대지 면적	대지 면적
		총방수 (방, 거실, 식사용 방)	총방수 (방, 거실, 식사용 방)
		건축연도 코드	건축연도 코드
		주거시설수 (부엌, 화장실, 독립 출입구)	주거시설수 (부엌, 화장실, 독립 출입구)

<표 1-9> 주택자료의 매칭작업을 위해 사용된 매칭키 및 우선고려순위

항목	매칭작업		비고
	매칭키	우선고려순위	
주소	본번 표준화 (AM)	1	층 매칭키는 주소, 호수 또는 거처번호를 활용하여 생성하였음.
	부번 표준화 (AS)	1	
	동명 표준화 (AD)	1	
	호명 표준화 (AR)	1	
층	층	1	
건물명	건물명	2	
면적	면적	3	
건축연도	건축연도 코드	4	

<표 1-10> 건축물대장과 시험조사 자료의 건축연도코드와 동/호명 표준화 동일화

	건축물대장	시험조사 자료	최종 변환된 건축연도코드
건축연도 코드 생성	2008년	1	1
	2007년	2	2
	2006년	3	3
	2000~2005년	4	4
	1995~1999년	5	5
	1990~1994년	6	6
	1980~1989년	7	7
	1970~1979년	8	8
	1960~1969년	9	9
	1959년이전	10	10
동(호)명 표준화	A동(호)	A동(호)	000A(호)
	1동(호)	1동(호)	0001(호)
	101동(호)	101동(호)	0101(호)

다. 세대-가구 매칭

세대 개념으로 구축된 주민등록제도와 가구 개념을 기반으로 수행된 총 조사 자료로부터 세대와 가구를 매칭하는 것은 기존의 가구개념을 적용하여 생성한 통계지표를 주민등록대장의 세대자료를 이용하여 재해석이 가능하도록 하려는 목적을 갖고 있다.

본 연구에서는 세대와 가구를 두 가지 개념으로 매칭하였다. 첫 번째는 주

민등록자료와 시험조사 자료에 공통적으로 있는 주소(본번, 부번, 동명, 호명)만을 매칭키로 고려한 것으로 이를 “주소매칭방법”이며, 두 번째는 주민등록 자료의 세대원과 시험조사 자료의 가구원이 일부라도 매칭된 경우만을 세대와 가구의 매칭으로 인정하는 “구성원매칭방법”이다.

여기서 “주소매칭방법”은 주소의 오류 등을 파악할 수 없고, 또한 주민등록에는 등재되어있으나 실존하지 않는 세대 및 주민등록에는 등재 되어있지 않지만 실존하는 가구의 매칭에 대해 파악하기가 어렵다는 문제점을 갖고 있으며, “구성원매칭방법”은 등재되어있지만 실존하지 않는 세대원(가구원)이 있거나 등재되어있지 않지만 실존하는 세대원(가구원)으로 인한 1세대 다가구매칭나 다세대 1가구매칭으로 복잡한 세대와 가구매칭 결과가 도출된다는 것이다.

그러므로 “구성원매칭방법”과 “주소매칭방법” 중 어느 것이 더 좋거나 적절하다는 것을 말하기 어렵다.

라. 행정자료 매칭

본 연구에서 사용된 행정자료는 주민등록 자료와 건축물대장 자료로 주민등록자료는 조사 자료의 인구부분과 매칭을 실시하고, 건축물대장 자료는 조사 자료의 주택부분과 매칭을 실시한 바 있다.

이에 조사 자료의 인구부분이라고 간주할 수 있는 주민등록 자료와 조사 자료의 주택부분이라고 할 수 있는 건축물대장 자료의 매칭을 통하여 기존의 인구주택총조사를 통하여 얻을 수 있었던 정보를 추출할 수 있는 자료를 생성하는 것을 목적으로 한다. 그래서 주민등록 자료와 건축물대장 자료를 공통정보인 주소와 건물명을 매칭키로 하여 매칭을 수행하였다.

마. 매칭과정의 주요 전략

이 절에서는 인구매칭, 주택매칭의 각 과정에서 행정자료와 시험조사 자료에 공통적으로 존재하는 항목을 매칭키로 하여 단계적으로 수행된 매칭작업에서 사용된 주요 전략을 토의한다. 또한 세대와 가구매칭 및 행정자료 간 매칭을 수행하는 과정과 이 때 사용된 주요 전략을 토의한다.

여러 단계에 걸쳐 매칭이 이루어진 인구부분과 주택부분의 매칭에서는 제2

차와 제3차 조사가 각각 별개로 수행되었으며, 최대한의 매칭률을 얻고자 노력하였다. 우선 첫 단계에서 매칭키가 모두 일치하는 사례를 추출하였고, 후속 단계에서는 중요도가 낮다고 판단되는 매칭키를 순서대로 제외하면서 매칭작업이 이루어졌다. 매칭작업의 하위단계에서 일부의 매칭키만 일치하는 것으로 확인된 불안한 형태의 매칭사례들은 직접 눈으로 다른 정보들을 대조하는 수작업의 과정을 통하여 매칭 여부를 다시 판단하였다. 그리고 최종 단계에 이르기까지 매칭되지 않고 남아 있는 사례들은 더 이상 매칭되지 않음을 확인하는 수작업이 이루어졌다.

인구자료의 매칭작업은 제2차 시험조사지역에서 25단계, 제3차 시험조사지역에서 19단계까지 이루어졌고, 주택자료의 매칭작업은 제2차 시험조사지역에서 58단계까지 이루어졌고, 제3차 시험조사지역에서 46단계까지 이루어졌다.

(1) 인구자료의 매칭

인구자료의 매칭에 사용한 매칭키는 주소, 성명, 출생년/월, 성별, 가구주와의 관계였는데 본 연구팀이 매칭과정에서 전략적으로 고려한 매칭항목은 성명과 출생년/월, 동명표준화에 관한 것이다.

성명의 경우에는 소양래-소양례, 강혜인-강해인, 김여숙-김려숙 등과 같이 발음이 비슷해서 발생한 오류, 코딩 시 발생할 수 있는 편칭오류 등을 찾아볼 수 있었다. 성과 이름 중 한자가 같은 사람, 이름만 같은 사람을 동일이름으로 고려하여 매칭과정에 한 가지 전략으로 포함시켰다.

출생년/월의 경우에도 세는 나이와 만 나이, 그리고 양력과 음력의 혼용에 따른 오류가 개입할 개연성이 높다. 나이가 높아질수록 실제 나이와 주민등록 나이가 차이나는 경향도 발견된다. 이를 감안하여 매칭작업의 하위단계 일부에서 출생년/월에 대한 오차범위를 최대 ± 3 년까지 허용하였다. 그리고 제2차 시험조사 자료에는 출생월의 정보 없이 세는 나이만 제공되었기 때문에 이를 출생년도로 변환시켜 매칭작업을 하였다.

제3차 시험조사 지역의 매칭에서 주소 중 동명은 101동(호)을 1동(호)로 표기 되어있는 경우가 있다. 이에 매칭의 하위 단계에서는 동명의 맨 끝자리만을 매칭키로 사용하였다.

(2) 주택자료의 매칭

매칭률 제고를 위한 주택자료 매칭 과정의 가장 중요한 전략은 건축물대장 자료와 시험조사 자료에서 주택종류가 “APT인 경우”와 “APT가 아닌 경우”로 구분하고 매칭을 진행한 것이다. 먼저 두 자료 모두 아파트라고 되어있는 개체의 매칭을 최대한 추진한 후, 매칭되지 않은 APT 자료는 “APT가 아닌 경우”의 자료와 통합하여 “매칭APT 외 주택”이라고 명명하고 매칭을 진행한다. 이러한 매칭 전략을 세운 이유는 매칭률을 최대한으로 끌어올리려고 한 것이고, 또한 매칭에 이용하는 두 자료 중 한 쪽 자료에는 주택종류가 “APT”라고 분류되어 있지만 다른 한 쪽에서는 연립주택이나 다세대 주택과 같은 집합주택 그리고 단독주택 등으로 분류될 수 있기 때문이다. (단, 3차자료에서는 조사구특성변수가 추가적으로 존재하여 주택종류로 APT를 구별하지 않고 조사구특성이 APT로 구분지어진 자료를 APT라고 정의한다.)

전략적으로 고려한 매칭항목은 “APT” 매칭에서는 건물명이었고, “매칭APT 외 주택” 매칭에서는 층번호였다. 공통적으로 사용하는 매칭항목은 “동(호)명표준화”이다.

“APT”를 매칭할 때는 “주소” 중 일부가 달라도 건물명이 완전(exact)일치하거나 부분일치하는 경우 동일 개체로 판단하여 매칭하였다.

그리고 “매칭APT 외 주택”의 매칭에는 주택종류가 단독주택, 집합주택을 포함하기 때문에 동일주소 내 거주할 수 있는 주택의 수가 2개 이상이 될 수 있기 때문에 이를 구분하고자 “층번호”를 생성하였다. 예를 들어, 일반단독주택의 2층 주택을 매칭하는 과정에서 시험조사 자료에는 한 가구가 2층 주택을 모두 사용하는 경우, 건축물이 1개로 조사되어 2층 주택이라는 정보를 얻을 수 없지만 건축물대장에는 1층과 2층으로 등록된다. 이러한 문제를 해결하기 위하여 건축물대장과 시험조사 자료에서 주소를 이용하여 층 변수를 생성하였다. 그런데 이때, 다가구와 다세대 주택에서 별도의 호명이 있으면 이를 이용하여 층 변수를 생성하였다. 시험조사 자료의 경우에는 거쳐번호가 있기 때문에 이를 활용하여 층 변수를 생성하였다.

또한 인구매칭에서와 동일하게 제3차 시험조사 자료의 매칭 시 주소 중 동명과 호명은 101동(호)을 1동(호)로 표기 되어있는 경우가 있다. 이에 매칭의 하위 단계에서는 동명과 호명의 맨 끝자리만을 매칭키로 사용하였다.

(3) 세대-가구의 매칭

세대와 가구를 “주소매칭방법”과 “구성원매칭방법”의 두 가지로 방안으로 매칭하였다. 세대와 가구 매칭에서의 전략은 “구성원매칭방법”에서는 주민등록 자료의 세대원과 시험조사 자료의 가구원이 일부라도 매칭된 경우만을 세대와 가구의 매칭으로 인정했기 때문에 매칭된 인구자료를 이용한 것이었다.

(4) 행정자료 매칭

행정자료 매칭은 주소만을 이용한 매칭으로 전략적인 내용은 없다.

제3절 매칭작업의 결과

1. 인구자료

제2차 시험조사의 단계별 매칭결과는 <표 1-10>로 작성하였고, <표 1-11>에는 이 결과를 다시 행정동별로 재정리한 것이다. 모든 매칭키가 만족하는 1단계에서 주민등록 자료 기준 31.2%(27,258건)가 매칭되었고, 그 이후에 4단계에서 주소와 성명이 같으면서 연령 차이를 ± 1 년을 허용한 결과 주민등록 자료 기준 7.8%(6,829건)가 매칭되었다. 12단계에서는 성명, 생년, 동명, 호명을 일치한 결과 주민등록 자료 기준 12%(10,516건)가 매칭되었다. 한편 정확 매칭(exact matching)된 1단계를 제외하고 2단계부터는 매칭된 건수에 대한 수작업을 함께 수행하였다. 제2차 시험조사 지역의 매칭 과정에서는 1단계부터 16단계까지는 해당단계에서 제외된 정보의 불일치 된 상황이 매칭되었다고 할 수 없다는 판단이 내려지지 않으면 매칭되었다고 하고 이를 통합하여 S(Strong)단계라고 하였다. 17단계에서 21단계까지는 해당단계에서 사용된 매칭키에 의해서 매칭된 개체들이 실제로는 매칭이 안 될 가능성이 S단계보다는 다소 높다고 보면서 수작업을 실시하였는데 이 단계를 M(Moderate)라고 하였다. 22단계 이후로는 반대로 불일치 된 정보를 중요하게 보면서 매칭되었다고 확신할 수 없다면 매칭되지 않았다고 하는 입장을 갖고 주관적인 판단으로 수작업을 하였다. 22단계부터 25단계까지는 W(Weak)단계라고 한다.

제3차 시험조사지역의 단계별 매칭결과는 <표 1-12>이고, <표 1-13>는 행정동별로 재정리한 것이다. 제3차 시험조사지역에서도 모든 키가 다 만족한 정확매칭(exact matching)된 1단계에서는 주민등록 자료 기준 46.0%(41,971건)가 매칭되었다. 그리고 주소 매칭키와, 성명의 매칭키에서 출생년/월에 오차범위 ± 3 년을 허용한 4단계에서 주민등록자료 기준 12.3%(11,245건)이 매칭되었다. 제3차 시험조사지역도 제2차 시험조사지역과 마찬가지로 1단계 정확매칭(exact matching)을 제외하고 2단계부터는 매칭된 건수에 대하여 수작업을 실시하였다. 수작업을 실시하는 과정에서 제3차 시험조사지역에서는 9단계까지를 S(Strong)단계, 10단계부터 18단계까지를 M(Moderate), 19단계를 W(Weak)단계라고 하였다.

<표 1-11> 제2차 시험조사 자료의 단계별 인구 매칭키 및 매칭작업 결과

단계	지역	주소				성명		생년		성별	가구 주와의 관계	수작 업	매칭 건수	잔여 주민등록 인구	잔여 시험조사 인구	주민등록		시험조사	
		AM	AS	AD	AR	NL	NF	BY	연령							매칭 %	매칭 누적 %	매칭 %	매칭 누적 %
		본번	부번	동명	호명	성	이름	생년											
													87,414	78,359					
1	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇		27,258	60,156	51,101	31.2	31.2	34.8	34.8	
2	◇	◇	◇	◇	◇	◇	◇	◇	◇			2,394	57,762	48,707	2.7	33.9	3.1	37.8	
3	◇	◇	◇	◇	◇	◇	◇	◇				134	57,628	48,573	0.2	34.1	0.2	38.0	
4	◇	◇	◇	◇	◇	◇	±1					6,829	50,799	41,744	7.8	41.9	8.7	46.7	
5	◇	◇	◇	◇	◇	◇	±3					1,479	49,320	40,265	1.7	43.6	1.9	48.6	
6	◇	◇	◇	◇	◇		◇					1,583	47,737	38,682	1.8	45.4	2.0	50.6	
7	◇	◇	◇	◇		◇	◇					206	47,531	38,476	0.2	45.6	0.3	50.9	
8	◇	◇		◇	◇	◇	◇					2,846	44,685	35,630	3.3	48.9	3.6	54.5	
9	◇	◇		◇	◇	◇	±1					718	43,967	34,912	0.8	49.7	0.9	55.4	
10	◇	◇		◇	◇	◇	±3					156	43,811	34,756	0.2	49.9	0.2	55.6	
11	◇	◇		◇	◇	◇	◇					487	43,324	34,269	0.6	50.4	0.6	56.3	
12	◇			◇	◇	◇	◇					10,516	32,808	23,753	12.0	62.5	13.4	69.7	
13	◇	◇	◇		◇	◇	◇					1,685	31,123	22,068	1.9	64.4	2.2	71.8	
14	◇	◇	◇		◇	◇	±3	◇				417	30,706	21,651	0.5	64.9	0.5	72.4	
15	◇	◇	◇		◇	◇	60+	◇	◇			109	30,597	21,542	0.1	65.0	0.1	72.5	
16	◇	◇	◇		◇	◇	±5 40+	◇	◇			107	30,490	21,435	0.1	65.1	0.1	72.6	
17	◇	◇			◇	◇	◇	◇				813	29,677	20,622	0.9	66.1	1.0	73.7	
18	◇	◇			◇	◇	±3	◇	◇			177	29,500	20,445	0.2	66.3	0.2	73.9	
19	◇	◇			◇	◇	60+	◇	◇			19	29,481	20,426	0.0	66.3	0.0	73.9	
20	◇	◇			◇	◇	±5 40+	◇	◇			16	29,465	20,410	0.0	66.3	0.0	74.0	
21	◇	±1			◇	◇	◇	◇	◇			113	29,352	20,297	0.1	66.4	0.1	74.1	
22	◇	◇			◇	1	◇	◇	◇			109	29,243	20,188	0.1	66.5	0.1	74.2	
23	◇	◇			◇	1	±3	◇	◇			417	28,826	19,771	0.5	67.0	0.5	74.8	
24	◇				◇	◇	◇	◇	◇			4,526	24,300	15,245	5.2	72.2	5.8	80.5	
25	◇				◇	◇	±1	◇	◇			2,153	22,147	13,902	2.5	74.7	2.7	83.3	

주 : 1) 이름 1: 이름 두 글자 중 한 글자 맞는 것을 허용하였음.
 2) 생년 ± 1, ± 3, ± 5: 생년이 최대 ± (값)까지 차이나는 것을 허용하였음.
 3) 연령 40+, 60+: 연령이 40세 및 60세 이상만을 고려하였음.

<표 1-12> 제2차 시험조사 자료의 행정구역별 인구 매칭작업 결과

	부산 부산진구 범천1동	부산 동래구 명장1동	경기 용인시 기흥동	경기 광주시 광남동	강원 동해시 북호동	강원 인제군 인제읍	전체
1단계	3,340 (34.4/35.5)	9,820 (53.6/58.0)	4,353 (32.8/36.5)	6,779 (21.3/23.6)	1,168 (24.4/29.4)	1,798 (18.8/24.2)	27,258 (31.2/34.8)
2단계	357 (3.7/3.8)	657 (3.6/3.9)	358 (2.7/3.0)	653 (2.1/2.3)	141 (2.9/3.6)	228 (2.4/3.1)	2,394 (2.7/3.1)
3단계	27 (0.3/0.3)	41 (0.2/0.2)	23 (0.2/0.2)	28 (0.1/0.1)	7 (0.1/0.2)	8 (0.1/0.1)	134 (0.2/0.2)
4단계	858 (8.8/9.1)	2,623 (14.3/15.5)	891 (6.7/7.5)	1,551 (4.9/5.4)	405 (8.5/10.2)	501 (5.3/6.7)	6,829 (7.8/8.7)
5단계	189 (1.9/2.0)	598 (3.3/3.5)	149 (1.1/1.2)	292 (0.9/1.0)	123 (2.6/3.1)	128 (1.3/1.7)	1,479 (1.7/1.9)
6단계	302 (3.1/3.2)	553 (3.0/3.3)	184 (1.4/1.5)	287 (0.9/1.0)	116 (2.4/2.9)	141 (1.5/1.9)	1,583 (1.8/2.0)
7단계	30 (0.3/0.3)	84 (0.5/0.5)	19 (0.1/0.2)	40 (0.1/0.1)	21 (0.4/0.5)	12 (0.1/0.2)	206 (0.2/0.3)
8단계	496 (5.1/5.3)	122 (0.7/0.7)	174 (1.3/1.5)	1,255 (4.0/4.4)	293 (6.1/7.4)	506 (5.3/6.8)	2,846 (3.3/3.6)
9단계	118 (1.2/1.3)	24 (0.1/0.1)	42 (0.3/0.4)	290 (0.9/1.0)	111 (2.3/2.8)	133 (1.4/1.8)	718 (0.8/0.9)
10단계	38 (0.4/0.4)	5 (0.0/0.0)	7 (0.1/0.1)	40 (0.1/0.1)	27 (0.6/0.7)	39 (0.4/0.5)	156 (0.2/0.2)
11단계	194 (2.0/2.1)	64 (0.3/0.4)	20 (0.2/0.2)	97 (0.3/0.3)	69 (1.4/1.7)	43 (0.5/0.6)	487 (0.6/0.6)
12단계	19 (0.2/0.2)	48 (0.3/0.3)	2710 (20.4/22.7)	6,900 (21.7/24.0)	60 (1.3/1.5)	779 (8.2/10.5)	10,516 (12.0/13.4)
13단계	485 (5.0/5.2)	356 (1.9/2.1)	245 (1.8/2.1)	441 (1.4/1.5)	60 (1.3/1.5)	98 (1.0/1.3)	1,685 (1.9/2.2)
14단계	114 (1.2/1.2)	96 (0.5/0.6)	57 (0.4/0.5)	107 (0.3/0.4)	18 (0.4/0.5)	25 (0.3/0.3)	417 (0.5/0.5)
15단계	24 (0.2/0.3)	27 (0.1/0.2)	11 (0.1/0.1)	8 (0.0/0.0)	26 (0.5/0.7)	13 (0.1/0.2)	109 (0.1/0.1)
16단계	19 (0.2/0.2)	47 (0.3/0.3)	6 (0.0/0.1)	16 (0.1/0.1)	8 (0.2/0.2)	11 (0.1/0.1)	107 (0.1/0.1)
17단계	277 (2.9/2.9)	15 (0.1/0.1)	34 (0.3/0.3)	323 (1.0/1.1)	3 (0.1/0.1)	161 (1.7/2.2)	813 (0.9/1.0)
18단계	52 (0.5/0.6)	5 (0.0/0.0)	10 (0.1/0.1)	79 (0.2/0.3)		31 (0.3/0.4)	177 (0.2/0.2)
19단계	2 (0.0/0.0)	1 (0.0/0.0)		4 (0.0/0.0)	7 (0.1/0.2)	5 (0.1/0.1)	19 (0.0/0.0)
20단계	1 (0.0/0.0)		1 (0.0/0.0)	8 (0.0/0.0)	3 (0.1/0.1)	3 (0.0/0.0)	16 (0.0/0.0)
21단계		11 (0.1/0.1)	7 (0.1/0.1)	92 (0.3/0.3)		3 (0.0/0.0)	113 (0.1/0.1)
22단계	16 (0.2/0.2)	18 (0.1/0.1)	16 (0.1/0.1)	44 (0.1/0.2)	3 (0.1/0.1)	12 (0.1/0.2)	109 (0.1/0.1)
23단계	58 (0.6/0.6)	134 (0.7/0.8)	43 (0.3/0.4)	86 (0.3/0.3)	42 (0.9/1.1)	54 (0.6/0.7)	417 (0.5/0.5)
24단계	68 (0.7/0.7)	118 (0.6/0.7)	327 (2.5/2.7)	2,790 (8.8/9.7)	294 (6.1/7.4)	929 (9.7/12.5)	4,526 (5.2/5.8)
25단계	21 (0.2/0.2)	14 (0.1/0.1)	443 (3.3/3.7)	1,492 (4.7/5.2)	34 (0.7/0.9)	149 (1.6/2.0)	2,153 (2.5/2.7)
전체	7,105 (73.2/75.6)	15,481 (84.5/91.5)	10,130 (76.3/84.9)	23,702 (74.6/82.6)	3,039 (63.5/76.5)	5,810 (60.9/78.2)	65,267 (74.7/83.3)

주: 첫 번째 숫자는 매칭이 이루어진 사례수; 괄호 안의 첫 번째 숫자는 주민등록대장을 기준으로 한 매칭달성 백분율; 괄호 안의 두 번째 숫자는 시험조사 자료를 기준으로 한 매칭달성 백분율.

<표 1-13> 제3차 시험조사 자료의 단계별 인구 매칭키 및 매칭작업 결과

단계	지역	주소				성명		생년/월		성별	가구주와의 관계	수작업	매칭 건수	잔여 주민등록 인구	잔여 시험조사 인구	주민등록		시험조사	
		AM	AS	AD	AR	NL	NF	BY	BM							매칭 %	매칭 누적 %	매칭 %	매칭 누적 %
		본번	부번	동명	호명	성	이름	생년	생월										
													91,175	91,365					
1	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇		41,971	49,204	49,394	46.0	46.0	45.9	45.9	
2	◇	◇	◇	◇	◇	◇	◇	◇	◇			3,654	45,550	45,740	4.0	50.0	4.0	49.9	
3	◇	◇	◇	◇	◇	◇	◇	◇				164	45,386	45,576	0.2	50.2	0.2	50.1	
4	◇	◇	◇	◇	◇	◇	±3					11,245	34,141	34,331	12.3	62.6	12.3	62.4	
5	◇	◇	◇	◇			◇	◇				1,917	32,224	32,414	2.1	64.7	2.1	64.5	
6	◇	◇	◇	◇	◇		◇					481	31,743	31,933	0.5	65.2	0.5	65.0	
7	◇	◇	◇	◇		◇	◇					59	31,684	31,874	0.1	65.2	0.1	65.1	
8	◇	◇		◇	◇	◇	◇					814	30,870	31,060	0.9	66.1	0.9	66.0	
9	◇	±1		◇	◇	◇	◇					867	30,003	30,193	1.0	67.1	0.9	67.0	
10	◇	◇	◇	2	◇	◇	◇					4,890	25,113	25,303	5.4	72.5	5.4	72.3	
11	◇	◇	◇		◇	◇	◇					3,246	21,867	22,057	3.6	76.0	3.6	75.9	
12	◇	◇		2	◇	◇	+1					294	21,573	21,763	0.3	76.3	0.3	76.2	
13	◇	◇		2	◇	◇	-1					80	21,493	21,683	0.1	76.4	0.1	76.3	
14	◇	◇		2	◇	◇	+2					81	21,412	21,602	0.1	76.5	0.1	76.4	
15	◇	◇		2	◇	◇	-2					19	21,393	21,583	0.0	76.5	0.0	76.4	
16	◇	◇			◇	◇	◇	◇				245	21,148	21,338	0.3	76.8	0.3	76.6	
17	◇	◇			◇	◇	±3		◇			547	20,601	20,791	0.6	77.4	0.6	77.2	
18	◇	◇	◇		◇	1	±3		◇			556	20,045	20,235	0.6	78.0	0.6	77.9	
19	◇				◇	◇	◇	◇				2,308	17,737	17,927	2.5	80.5	2.5	80.4	

주: 1) 본번 ± 1: 본번이 ± 1 차이나는 것을 허용하였음.
 2) 동명 2: 동명 표준화에서 뒤의 2자리 숫자만 사용하였음.
 3) 이름 1: 이름 두 글자 중 한 글자 맞는 것을 허용하였음.
 4) 생년 +1, -1, +2, -2, ± 3: 생년이 최대 ± (값)까지 차이나는 것을 허용하였음.

<표 1-14> 제3차 시험조사 자료의 행정구역별 인구 매칭작업 결과

	광주 북구 문흥1동	광주 광산구 월곡1동	충북 제천시 교동	충북 음성군 소이면	경남 김해시 활천동	경남 고성군 거류면	전체
1단계	11,269 (53.8/52.2)	8,608 (51.7/52.6)	10,599 (57.5/59.9)	589 (17.2/18.5)	9,935 (37.7/35.9)	971 (18.2/20.1)	41,971 (46.0/45.9)
2단계	1,104 (5.3/5.1)	781 (4.7/4.8)	930 (5.0/5.3)	97 (2.8/3.1)	653 (2.5/2.4)	89 (1.7/1.8)	3,654 (4.0/4.0)
3단계	65 (0.3/0.3)	24 (0.1/0.1)	30 (0.2/0.2)	2 (0.1/0.1)	39 (0.1/0.1)	4 (0.1/0.1)	164 (0.2/0.2)
4단계	2,684 (12.8/12.4)	2,636 (15.8/16.1)	2,382 (12.9/13.5)	452 (13.2/14.2)	2,583 (9.8/9.3)	508 (9.5/10.5)	11,245 (12.3/12.3)
5단계	470 (2.2/2.2)	441 (2.7/2.7)	386 (2.1/2.2)	41 (1.2/1.3)	526 (2.0/1.9)	53 (1.0/1.1)	1,917 (2.1/2.1)
6단계	96 (0.5/0.4)	115 (0.7/0.7)	80 (0.4/0.5)	15 (0.4/0.5)	129 (0.5/0.5)	46 (0.9/1.0)	481 (0.5/0.5)
7단계	16 (0.1/0.1)	9 (0.1/0.1)	13 (0.1/0.1)	2 (0.1/0.1)	16 (0.1/0.1)	3 (0.1/0.1)	59 (0.1/0.1)
8단계	50 (0.2/0.2)	37 (0.2/0.2)	203 (1.1/1.1)	154 (4.5/4.8)	159 (0.6/0.6)	211 (3.9/4.4)	814 (0.9/0.9)
9단계	592 (2.8/2.7)	18 (0.1/0.1)	59 (0.3/0.3)	64 (1.9/2.0)	5 (0.0/0.0)	129 (2.4/2.7)	867 (1.0/0.9)
10단계			1 (0.0/0.0)	40 (1.2/1.3)	4,847 (18.4/17.5)	2 (0.0/0.0)	4,890 (5.4/5.4)
11단계	655 (3.1/3.0)	251 (1.5/1.5)	123 (0.7/0.7)	91 (2.7/2.9)	1,950 (7.4/7.0)	176 (3.3/3.6)	3,246 (3.6/3.6)
12단계	4 (0.0/0.0)	2 (0.0/0.0)	23 (0.1/0.1)	18 (0.5/0.6)	221 (0.8/0.8)	26 (0.5/0.5)	294 (0.3/0.3)
13단계	1 (0.0/0.0)	5 (0.0/0.0)	3 (0.0/0.0)	9 (0.3/0.3)	60 (0.2/0.2)	2 (0.0/0.0)	80 (0.1/0.1)
14단계	1 (0.0/0.0)		8 (0.0/0.0)	11 (0.3/0.3)	56 (0.2/0.2)	5 (0.1/0.1)	81 (0.1/0.1)
15단계			2 (0.0/0.0)	1 (0.0/0.0)	16 (0.1/0.1)		19 (0.0/0.0)
16단계	24 (0.1/0.1)	6 (0.0/0.0)	3 (0.0/0.0)	20 (0.6/0.6)	163 (0.6/0.6)	29 (0.5/0.6)	245 (0.3/0.3)
17단계	122 (0.6/0.6)	54 (0.3/0.3)	12 (0.1/0.1)	53 (1.5/1.7)	251 (1.0/0.9)	55 (1.0/1.1)	547 (0.6/0.6)
18단계	103 (0.5/0.5)	88 (0.5/0.5)	77 (0.4/0.4)	31 (0.9/1.0)	214 (0.8/0.8)	43 (0.8/0.9)	556 (0.6/0.6)
19단계	287 (1.4/1.3)	254 (1.5/1.6)	376 (2.0/2.1)	384 (11.2/12.1)	248 (0.9/0.9)	759 (14.2/15.7)	2,308 (2.5/2.5)
전체	17,543 (83.7/81.3)	13,329 (80.1/81.5)	15,310 (83.1/86.5)	2,074 (60.6/65.3)	22,071 (83.7/79.7)	3,111 (58.2/64.3)	73,438 (80.5/80.4)

주: 첫 번째 숫자는 매칭이 이루어진 사례수; 괄호 안의 첫 번째 숫자는 주민등록대장을 기준으로 한 매칭달성 백분율; 괄호 안의 두 번째 숫자는 시험조사 자료를 기준으로 한 매칭달성 백분율.

<표 1-15> 주민등록 자료와 시험조사 자료를 활용한 매칭 결과, 2008

행정구역			주민등록 자료	시험조사 자료	매칭수	주민등록 인구 기준	시험조사 인구 기준
제2차 시험조사							
부산	부산진구	범천1동	9,712	9,399	7,105	73.2	75.6
부산	동래구	명장1동	18,321	16,924	15,481	84.5	91.5
경기	용인시	기흥동	13,283	11,925	10,130	76.3	84.9
경기	광주시	광남동	31,772	28,709	23,702	74.6	82.6
강원	동해시	목호동	4,786	3,971	3,039	63.5	76.5
강원	인제군	인제읍	9,540	7,431	5,810	60.9	78.2
합계			87,414	78,359	65,267	74.7	83.3
제3차 시험조사							
광주	북구	문흥1동	20,958	21,575	17,543	83.7	81.3
광주	광산구	월곡1동	16,637	16,361	13,329	80.1	81.5
충북	제천시	교동	18,433	17,707	15,310	83.1	86.5
충주	음성군	소이면	3,423	3,176	2,074	60.6	65.3
경남	김해시	활천동	26,381	27,704	22,071	83.7	79.7
경남	고성군	거류면	5,343	4,842	3,111	58.2	64.3
합계			91,175	91,365	73,438	80.5	80.4

제2차 시험조사 지역은 총 25단계의 매칭단계를 거쳐 65,267건이 매칭되었다. 이는 주민등록 자료 기준으로 74.7%, 시험조사 자료 기준으로 83.3%의 매칭률을 나타낸다. 특히 부산 동래구 명장1동에서 15,481건이 매칭되었으며, 이는 주민등록 자료 18,321건에서 84.5%, 시험조사 자료 15,481건에서 91.5%의 매칭률로 제2차 시험조사 지역에서 가장 높은 매칭률을 보이고 있다.

제3차 시험조사 지역에서는 19단계의 매칭단계를 통하여 73,438건이 매칭되었으며, 이는 주민등록 자료 기준으로 80.5%, 시험조사 자료 기준으로 80.4%의 매칭률을 나타낸다. 제3차 시험조사 지역에서 광주 북구 문흥1동에서 17,543건이 매칭되어 주민등록 자료 기준으로 83.7%, 시험조사 자료 기준으로 81.3%가 매칭되어 제3차 시험조사 지역에서 가장 높은 매칭률을 나타냈다.

특이사항으로, 제2차 시험조사 지역 매칭과정에서는 1단계 매칭과정에서 27,258건이 매칭되어 행정자료 기준 31.2%, 조사자료 기준 34.8%가 매칭되었으며, 부산 동래구 명장1동이 9,820건이 매칭되어 행정자료 기준 53.6%, 조사자료 기준 58.0%가 매칭되어 다른 행정동에 비해 많이 매칭되었다고 할 수

이다. 모든 매칭키가 사용된 1단계에서 높은 매칭률을 보인 명장1동은 행정자료와 조사자료가 다른 행정동에 비해 많이 일치 하고 있다고 볼 수 있다.

1단계에 이어서 많은 건수가 매칭된 12단계에서는 경기용인시 기흥동이 2,710건이 매칭되었는데 그중 2,632건이 ‘한보라마을휴먼시아’아파트 이다. 한보라마을휴먼시아의 본번은 행정자료에서 674번지, 679번지라고 표기되었으며, 시험조사자료에서는 73, 74, 90, 450번지라고 표기되어 있다. 또한 부번은 행정자료에서는 모두 0000이고 시험조사자료에서는 0000,1,2,15,19, 24, 33, 38, 135라고 표기되어 있다. 경기 광주시 광남동의 6,900건 중에는 쌍용스윗닷홈 1,771건, 성원아파트 1,045건, 우남퍼스트빌 697건, 광주태전 e-편한세상 아파트 679건으로 나타났다. 제2차 시험조사 지역의 행정자료와 조사자료에서 아파트와 같이 집단적으로 주소가 틀린 경우를 많이 발견되었다. 제2차 시험조사 지역이 주소에 대한 일치 정도가 많이 낮다고 판단되어 행정동별로 살펴 보았다. <표1-15>에서 보는 바와 같이 기흥동과 광남동의 일치 비율이 다른 행정동에 비해 낮게 나타나고 있다.

<표 1-15> 제2차 조사지역의 행정동별 본번표준화 일치

행정동	본번표준화		전체	일치 비율
	일치	불일치		
범천동	7,425	4,581	12,006	61.84
명장동	15,305	4,459	19,764	77.44
기흥동	6,717	8,361	15,078	44.55
광남동	13,324	23,455	36,779	36.23
목호동	2,705	3,013	5,718	47.31
인제읍	4,291	6,870	11,161	38.45
총합	49,767	50,739	100,506	49.52

제3차 시험조사 지역에서는 모든 매칭키를 사용한 1단계 매칭에서 41,971건이 매칭되어 행정자료 기준으로 46.0%, 조사자료 기준으로 45.9%의 매칭율을 차지하여 제2차 시험조사 지역보다 1단계 매칭(exact matching)의 비율이 높게 나타났다.

19단계에서는 2,308건이 매칭이 되었는데, 이는 제3차 시험조사 지역의 행정자료 기준으로 2.5%, 시험조사 자료 기준으로 2.5%의 매칭률이다. 하지만 충북 음성군 소이면은 19단계에서 384건이 매칭되었으며, 이는 행정자료 기준 11.2%, 조사자료 기준 12.1%가 매칭된 것이며, 경남 고성군 거류면은 759

건이 매칭되어 행정자료 기준 14.2%, 조사자료 기준 15.7%가 매칭된 것이다. 소이면과 거류면이 다른 행정동에 비해 낮은 매칭률을 보이지만 19단계에서는 다른 행정동에 비해 높은 매칭률이 나타나는데 19단계에서는 주소 매칭키를 사용하지 않고 성명과 생년월일로만 매칭하였기 때문이다. 소이면과 거류면은 제3차 시험조사지역의 다른 행정동에 비해 본번표준화와 부번표준화의 일치 정도가 낮게 나타나기 때문에 주소를 사용하지 않았을 때 매칭이 잘 된다고 볼 수 있다.

2. 주택자료

제2차 시험조사의 단계별 매칭결과는 <표 1-16>로 작성하였고, <표 1-17>는 이 결과를 다시 행정동별로 재정리한 것이다.

주택종류가 APT인 자료만 추출하여 모든 매칭키가 만족하는 1단계에 건축물대장 자료 기준 20.6%(5,174건)가 매칭되었고, 이후 3단계는 건물명을 제외해나가며 매칭을 실시하여 건축물대장 자료 기준 7%(1,748건)가 매칭되었다.

아파트지만 호명이 공란으로 되어 있어 매칭되지 않은 것을 매칭시키기 위해 4단계는 건축물대장 중 호명표준화가 '0000'인 것만 선택하여 매칭에 이용하였고, 5~7단계에서는 건축물대장에서 동명표준화가 '0000'인 것만을 선택하여 매칭에 이용하여 건축물대장 자료 기준 1.01%(253건)를 매칭하였고, 8단계부터는 4~7단계에서 매칭되지 않은 모든 자료를 이용하여 부번표준화부터 제외시켜가며 매칭을 실시하고 16~18단계까지는 수작업을 병행하여, 건물명이 부분적으로 일치하는 경우에 수정하여 판단매칭을 진행하였다. 16~18단계에서 건축물대장 자료 기준 15.54%(3,905건)가 매칭되었다.

예) 건물명 수정

건축물대장 자료	건축물대장 자료	건물명 수정
금광포란재아파트	금강포란재아파트	금강포

19단계부터는 18단계까지 매칭되지 않은 APT자료를 APT가 아닌 자료와 합하여 주택종류가 다른 개체도 매칭될 수 있도록 하였다. 여기서 우선 건물명이 존재하는 개체만을 이용하여 단계 19를 진행하였는데, 여기에서는 추가적으로 층번수를 포함한 모든 매칭키를 사용하여 매칭에 이용한다. 19단계에

서는 건축물대장 자료 기준 2.7%(670건)가 매칭되었다. 이후 26단계까지 중요도가 낮은 매칭키를 제외하며 매칭하였다. 27-31단계는 건축물대장 자료에서 동명표준화가 '0000'인 자료를 이용하고 반대로 시험조사 자료에서 동명표준화가 '0000'인 자료를 이용 32-33단계를 진행하여 건축물대장 자료 기준 0.29%(74건)가 매칭되었다. 이후 34단계부터 건물명이 존재하지 않는 개체도 포함하여 다시 층변수를 포함한 모든 매칭키를 사용하였다. 34단계에서는 건축물대장 자료 기준 5.3%(1,337건)가 매칭되었다. 그리고 이후 중요도가 낮은 매칭키를 제외하며 매칭시켰고, 면적과 건축연도를 제외한 38단계에서 건축물대장 자료 기준 6.5%(1,626건)가 매칭되었다. 그리고 매칭단계 중 42, 46, 47, 48, 52, 54, 58단계에서 중복이 발생하여 수작업을 통해 매칭여부를 단계별로 판단하였다.

제3차 시험조사의 단계별 매칭결과는 <표 1-18>로 작성하였고, <표 1-19>는 이 결과를 다시 행정동별로 재정리한 것이다.

주택종류가 APT인 자료만 추출하여 모든 매칭키가 만족하는 1단계에 건축물대장 자료 기준 11.2%(3,000건)가 매칭되었고, 건물명과 건축연도를 제외한 4단계에서는 건축물대장 자료 기준 30.6%(8,223건)이 매칭되었다. 이후 8단계까지 중요도가 낮은 매칭키를 제외해 나가면서 매칭을 실시하였다. 9단계부터는 재 표준화된 동명표준화를 이용하는데, 9단계에서는 모든 매칭키를 사용하여 건축물대장 자료 기준 0.4%(97건)를 매칭하였고 면적을 제외하여 11%(2,012건)이 매칭되었다. 이후, 19단계까지 중요도가 낮은 매칭키를 제외해 나가면서 매칭을 실시하였다.

예) 재 표준화된 동명표준화

동명표준화		재 표준화된 동명표준화
0101	→	01(뒤의 두 자리)

20단계에서부터 APT에서 매칭되지 않은 자료와 APT 외로 구분되어진 자료를 합쳐서 매칭을 실시하였다. 다시 모든 매칭키가 사용된 20단계에서는 6.3%(1,690건)이 매칭되었고, 27단계까지는 중요도가 낮은 매칭키를 제외해 나가며 매칭을 실시하였는데 면적과 건축연도를 제외한 24단계에서 6.5%(1,747건)이 매칭되었다. 28단계부터에서는 수작업 병행하여 호명표준화가 누락되어 있지만 다른 매칭키가 일치한 자료를 매칭하여 건축물대장 자

료 기준 0.1%(25건)이 매칭되었다. 29단계부터는 재 표준화된 동명표준화를 사용하며, 중요도가 낮은 매칭키를 제외해 나가며 매칭을 실시하였다. 44단계에서는 수작업 후 건물명을 변경하여 매칭을 실시하여 총 46단계 거쳐 매칭을 하였다.

<표 1-17> 제2차 시험조사 자료의 단계별 주택 매칭키 및 매칭작업 결과

구분	단계	지역	주소				건물명	면적	건축연도	층	수작업	매칭건수	잔여행정자료	잔여시험조사	건축물대장				시험조사						
			AM 본번	AS 부번	AD 동명	AR 호명									구분매칭%	구분누적%	전체%	전체누적%	구분매칭%	구분누적%	전체%	전체누적%			
												11,957	11,786												
APT	1	◇	◇	◇	◇	◇	◇	◇			5,174	6,783	6,612	43.3	43.3	20.6	20.6	43.9	43.9	21.2	21.2				
	2	◇	◇	◇	◇	◇		◇			5	6,778	6,607	0.0	43.3	0.0	20.6	0.0	43.9	0.0	21.2				
	3	◇	◇	◇	◇		◇	◇			1,748	5,030	4,859	14.6	57.9	7.0	27.6	14.8	58.8	7.1	28.3				
	4	◇	◇	◇	◇		◇	◇			5	5,025	4,854	0.0	58.0	0.0	27.6	0.0	58.8	0.0	28.4				
	5	◇	◇	◇	◇	◇	◇				228	4,797	4,626	1.9	59.9	0.9	28.5	1.9	60.8	0.9	29.3				
	6	◇	◇	◇	◇	◇		◇			6	4,791	4,620	0.1	59.9	0.0	28.5	0.1	60.8	0.0	29.3				
	7	◇	◇	◇	◇			◇			14	4,777	4,606	0.1	60.0	0.1	28.6	0.1	60.9	0.1	29.4				
	8	◇	◇		◇	◇	◇	◇			345	4,432	4,261	2.9	62.9	1.4	29.9	2.9	63.8	1.4	30.8				
	9	◇	◇		◇	◇		◇	◇		80	4,352	4,181	0.7	63.6	0.3	30.3	0.7	64.5	0.3	31.1				
	10	◇		◇	◇	◇	◇	◇			1,473	2,879	2,708	12.3	75.9	5.9	36.1	12.5	77.0	6.0	37.1				
	11	◇		◇	◇	◇		◇	◇		66	2,813	2,642	0.6	76.5	0.3	36.4	0.6	77.6	0.3	37.4				
	12	◇			◇	◇	◇	◇	◇		1,031	1,782	1,611	8.6	85.1	4.1	40.5	8.7	86.3	4.2	41.6				
	13	◇			◇	◇		◇	◇		609	1,173	1,002	5.1	90.2	2.4	42.9	5.2	91.5	2.5	44.1				
	14	◇			◇	◇	◇				1	1,172	1,001	0.0	90.2	0.0	42.9	0.0	91.5	0.0	44.1				
	15	◇	◇		◇	◇	◇	◇			117	1,055	884	1.0	91.2	0.5	43.4	1.0	92.5	0.5	44.6				
	16	◇	◇		◇	△				◇	44	1,011	840	0.4	91.5	0.2	43.6	0.4	92.9	0.2	44.8				
	17	◇			◇	△		◇	◇		138	873	702	1.2	92.7	0.5	44.1	1.2	94.0	0.6	45.3				
	18	◇			◇	△				◇	1	872	701	0.0	92.7	0.0	44.1	0.0	94.1	0.0	45.3				
APT 외 연계 APT 외			APT외는 APT가 아닌 모든 주택을 뜻하며 18단계까지 APT매칭을 하고 남은 건축물대장 자료(872)와 시험조사 자료(701)를 합쳐 매칭 APT 외라 한다. 19단계부터 매칭APT외 자료(건축물대장(14,045), 시험 조사 자료(13,363))를 사용하여 매칭을 실시한다.										13,173	12,662											
												14,045	13,363												
	19	◇	◇	◇	◇	◇	◇	◇	◇		670	13,375	12,693	4.8	4.8	2.7	46.8	5.0	5.0	2.7	48.1				
	20	◇	◇	◇	◇	◇	◇		◇		3	13,372	12,690	0.0	4.8	0.0	46.8	0.0	5.0	0.0	48.1				
	21	◇	◇	◇	◇	◇	◇	◇	◇		94	13,278	12,596	0.7	5.5	0.4	47.2	0.7	5.7	0.4	48.5				
	22	◇	◇	◇	◇	◇		◇	◇		56	13,222	12,540	0.4	5.9	0.2	47.4	0.4	6.2	0.2	48.7				
	23	◇	◇	◇	◇	◇	◇		◇		30	13,192	12,510	0.2	6.1	0.1	47.5	0.2	6.4	0.1	48.8				
	24	◇	◇	◇	◇	◇		◇	◇		1	13,191	12,509	0.0	6.1	0.0	47.5	0.0	6.4	0.0	48.8				
	25	◇	◇	◇	◇	◇		◇	◇		57	13,134	12,452	0.4	6.5	0.2	47.7	0.4	6.8	0.2	49.1				
	26	◇	◇	◇	◇	◇			◇		9	13,125	12,443	0.1	6.6	0.0	47.8	0.1	6.9	0.0	49.1				
	27	◇	◇	◇	◇	◇	◇	◇	◇		35	13,090	12,408	0.2	6.8	0.1	47.9	0.3	7.1	0.1	49.2				
	28	◇	◇	◇	◇	◇		◇	◇		27	13,063	12,381	0.2	7.0	0.1	48.0	0.2	7.3	0.1	49.4				
	29	◇	◇	◇	◇	◇		◇	◇		0	13,063	12,381	0.0	7.0	0.0	48.0	0.0	7.3	0.0	49.4				
	30	◇	◇	◇	◇	◇			◇		4	13,059	12,377	0.0	7.0	0.0	48.0	0.0	7.4	0.0	49.4				
	31	◇	◇	◇	◇	◇		◇	◇		0	13,059	12,377	0.0	7.0	0.0	48.0	0.0	7.4	0.0	49.4				

<표 1-17> 제2차 시험조사 자료의 단계별 주택 매칭키 및 매칭작업 결과

구분	단계	지역 행정동	주소				건물 면적	건축 연도	층	수작 업	매칭 건수	잔여 행정자료	잔여 시험조사	건축물대장				시험조사			
			AM	AS	AD	AR								구분 매칭 %	구분 누적 %	전체 %	전체 누적 %	구분 매칭 %	구분 누적 %	전체 %	전체 누적 %
			본번	부번	동명	호명															
31단계까지 매칭 후 남은 건수											13,059	12,377									
	32	◇	◇	◇	◇	◇	◇	◇	◇	7	13,052	12,370	0.0	7.1	0.0	48.1	0.1	7.4	0.0	49.4	
	33	◇	◇	◇	◇	◇	◇	◇	◇	1	13,051	12,369	0.0	7.1	0.0	48.1	0.0	7.4	0.0	49.4	
	34	◇	◇	◇	◇	◇	◇	◇	◇	1,337	11,714	11,032	9.5	16.6	5.3	53.4	10.0	17.4	5.5	54.9	
	35	◇	◇	◇	◇	◇	◇	◇	◇	142	11,572	10,890	1.0	17.6	0.6	54.0	1.1	18.5	0.6	55.5	
	36	◇	◇	◇	◇	◇	◇	◇	◇	1,028	10,544	9,862	7.3	24.9	4.1	58.0	7.7	26.2	4.2	59.7	
	37	◇	◇	◇	◇	◇	◇	◇	◇	622	9,922	9,240	4.4	29.4	2.5	60.5	4.7	30.9	2.5	62.2	
	38	◇	◇	◇	◇	◇	◇	◇	◇	1,626	8,296	7,614	11.6	40.9	6.5	67.0	12.2	43.0	6.7	68.9	
	39	◇	◇	◇	◇	◇	◇	◇	◇	17	8,279	7,597	0.1	41.1	0.1	67.1	0.1	43.1	0.1	68.9	
	40	◇	◇	◇	◇	◇	◇	◇	◇	160	8,119	7,437	1.1	42.2	0.6	67.7	1.2	44.3	0.7	69.6	
	41	◇	◇	◇	◇	◇	◇	◇	◇	35	8,084	7,402	0.2	42.4	0.1	67.8	0.3	44.6	0.1	69.7	
	42	◇	◇	◇	◇	◇	◇	◇	◇	106	7,978	7,296	0.8	43.2	0.4	68.3	0.8	45.4	0.4	70.2	
	43	◇	◇	◇	◇	◇	◇	◇	◇	2	7,976	7,294	0.0	43.2	0.0	68.3	0.0	45.4	0.0	70.2	
	44	◇	◇	◇	◇	◇	◇	◇	◇	14	7,962	7,280	0.1	43.3	0.1	68.3	0.1	45.5	0.1	70.2	
	45	◇	◇	◇	◇	◇	◇	◇	◇	2	7,960	7,278	0.0	43.3	0.0	68.3	0.0	45.5	0.0	70.2	
	46	◇	◇	◇	◇	◇	◇	◇	◇	23	7,937	7,255	0.2	43.5	0.1	68.4	0.2	45.7	0.1	70.3	
	47	◇	◇	◇	◇	◇	◇	◇	◇	16	7,921	7,239	0.1	43.6	0.1	68.5	0.1	45.8	0.1	70.4	
	48	◇	◇	◇	◇	◇	◇	◇	◇	3	7,918	7,236	0.0	43.6	0.0	68.5	0.0	45.9	0.0	70.4	
	49	◇	◇	◇	◇	◇	◇	◇	◇	13	7,905	7,223	0.1	43.7	0.1	68.5	0.1	45.9	0.1	70.5	
	50	◇	◇	◇	◇	◇	◇	◇	◇	5	7,900	7,218	0.0	43.8	0.0	68.6	0.0	46.0	0.0	70.5	
	51	◇	◇	◇	◇	◇	◇	◇	◇	2	7,898	7,216	0.0	43.8	0.0	68.6	0.0	46.0	0.0	70.5	
	52	◇	◇	◇	◇	◇	◇	◇	◇	118	7,780	7,098	0.8	44.6	0.5	69.0	0.9	46.9	0.5	71.0	
	53	◇	◇	◇	◇	◇	◇	◇	◇	12	7,768	7,086	0.1	44.7	0.0	69.1	0.1	47.0	0.0	71.0	
	54	◇	◇	◇	◇	◇	◇	◇	◇	81	7,687	7,005	0.6	45.3	0.3	69.4	0.6	47.6	0.3	71.3	
	55	◇	◇	◇	◇	◇	◇	◇	◇	30	7,657	6,975	0.2	45.5	0.1	69.5	0.2	47.8	0.1	71.5	
	56	◇	◇	◇	◇	◇	◇	◇	◇	50	7,607	6,925	0.4	45.8	0.2	69.7	0.4	48.2	0.2	71.7	
	57	◇	◇	◇	◇	◇	◇	◇	◇	4	7,603	6,921	0.0	45.9	0.0	69.7	0.0	48.2	0.0	71.7	
	58	◇	◇	◇	◇	◇	◇	◇	◇	5	7,598	6,916	0.0	45.9	0.0	69.8	0.0	48.2	0.0	71.7	

<표 1-17> 제2차 시험조사 자료의 행정구역별 주택(APT)의 매칭작업 결과

	부산 부산진구 범천1동	부산 동래구 명장1동	경기 용인시 기흥동	경기 광주시 광남동	강원 동해시 묵호동	강원 인제군 인제읍	전체
1단계	1,019 (33.9/26.9)	490 (11.9/11.9)	1,638 (38.4/42.2)	1,844 (21.6/22.3)		183 (6.7/6.9)	5,174 (20.6/21.2)
2단계	4 (0.1/0.1)			1 (0/0)			5 (0/0)
3단계		349 (8.5/8.5)	192 (4.5/5)	829 (9.7/10)	337 (13.6/19.3)	41 (1.5/1.5)	1,748 (7/7.1)
4단계	3 (0.1/0.1)		1 (0/0)	1 (0/0)			5 (0/0)
5단계	228 (7.6/6)						228 (0.9/0.9)
6단계	6 (0.2/0.2)						6 (0/0)
7단계					14 (0.6/0.8)		14 (0.1/0.1)
8단계	345 (11.5/9.1)						345 (1.4/1.4)
9단계				80 (0.9/1)			80 (0.3/0.3)
10단계			360 (8.4/9.3)	1,040 (12.2/12.6)		73 (2.7/2.8)	1,473 (5.9/6)
11단계				66 (0.8/0.8)			66 (0.3/0.3)
12단계		8 (0.2/0.2)	958 (22.4/24.7)	65 (0.8/0.8)			1,031 (4.1/4.2)
13단계				609 (7.1/7.4)			609 (2.4/2.5)
14단계		1 (0/0)					1 (0/0)
15단계	117 (3.9/3.1)						117 (0.5/0.5)
16단계			1 (0/0)			43 (1.6/1.6)	44 (0.2/0.2)
17단계				138 (1.6/1.7)			138 (0.5/0.6)
18단계				1 (0/0)			1 (0/0)
19단계	48 (1.6/1.3)	506 (12.3/12.3)	47 (1.1/1.2)	49 (0.6/0.6)		20 (0.7/0.8)	670 (2.7/2.7)
20단계	2 (0.1/0.1)	1 (0/0)					3 (0/0)
21단계	3 (0.1/0.1)	52 (1.3/1.3)		1 (0/0)	12 (0.5/0.7)	26 (1/1)	94 (0.4/0.4)
22단계		20 (0.5/0.5)	21 (0.5/0.5)	11 (0.1/0.1)		4 (0.1/0.2)	56 (0.2/0.2)
23단계	1 (0/0)	7 (0.2/0.2)	3 (0.1/0.1)	4 (0/0)		15 (0.6/0.6)	30 (0.1/0.1)
24단계			1 (0/0)				1 (0/0)
25단계		3 (0.1/0.1)	14 (0.3/0.4)	7 (0.1/0.1)		33 (1.2/1.2)	57 (0.2/0.2)
26단계		1 (0/0)	1 (0/0)	1 (0/0)		6 (0.2/0.2)	9 (0/0)
27단계		10 (0.2/0.2)	25 (0.6/0.6)				35 (0.1/0.1)
28단계		1 (0/0)	12 (0.3/0.3)	3 (0/0)		11 (0.4/0.4)	27 (0.1/0.1)
29단계							0 (0/0)
30단계		2 (0/0)	1 (0/0)	1 (0/0)			4 (0/0)

주: 첫 번째 숫자는 매칭이 이루어진 사례수; 괄호 안의 첫 번째 숫자는 건축물대장을 기준으로 한 매칭달성 백분율; 괄호 안의 두 번째 숫자는 시험조사 자료를 기준으로 한 매칭달성 백분율.

<표 1-17> 제2단계 시험조사 자료의 행정구역별 APT 외 주택의 매칭작업 결과

	부산 부산진구 범천1동	부산 동래구 명장1동	경기 용인시 기흥동	경기 광주시 광남동	강원 동해시 묵호동	강원 인제군 인제읍	전체
31단계							0 (0/0)
32단계		7 (0.2/0.2)					7 (0/0)
33단계		1 (0/0)					1 (0/0)
34단계	113 (3.8/3)	1,028 (24.9/25.1)	70 (1.6/1.8)	44 (0.5/0.5)	11 (0.4/0.6)	71 (2.6/2.7)	1,337 (5.3/5.5)
35단계	19 (0.6/0.5)	43 (1/1)	12 (0.3/0.3)	19 (0.2/0.2)	16 (0.6/0.9)	33 (1.2/1.2)	142 (0.6/0.6)
36단계	196 (6.5/5.2)	290 (7/7.1)	28 (0.7/0.7)	97 (1.1/1.2)	185 (7.5/10.6)	232 (8.5/8.7)	1,028 (4.1/4.2)
37단계	2 (0.1/0.1)	371 (9/9)	69 (1.6/1.8)	180 (2.1/2.2)			622 (2.5/2.5)
38단계	191 (6.4/5)	145 (3.5/3.5)	28 (0.7/0.7)	144 (1.7/1.7)	605 (24.5/34.6)	513 (18.9/19.3)	1,626 (6.5/6.7)
39단계		9 (0.2/0.2)		8 (0.1/0.1)			17 (0.1/0.1)
40단계	6 (0.2/0.2)	42 (1/1)	3 (0.1/0.1)	99 (1.2/1.2)	10 (0.4/0.6)		160 (0.6/0.7)
41단계		15 (0.4/0.4)		19 (0.2/0.2)	1 (0/0.1)		35 (0.1/0.1)
42단계		8 (0.2/0.2)	24 (0.6/0.6)	40 (0.5/0.5)		34 (1.3/1.3)	106 (0.4/0.4)
43단계				1 (0/0)		1 (0/0)	2 (0/0)
44단계			1 (0/0)	2 (0/0)		11 (0.4/0.4)	14 (0.1/0.1)
45단계						2 (0.1/0.1)	2 (0/0)
46단계				21 (0.2/0.3)		2 (0.1/0.1)	23 (0.1/0.1)
47단계			8 (0.2/0.2)	8 (0.1/0.1)			16 (0.1/0.1)
48단계				3 (0/0)			3 (0/0)
49단계				13 (0.2/0.2)			13 (0.1/0.1)
50단계				5 (0.1/0.1)			5 (0/0)
51단계				2 (0/0)			2 (0/0)
52단계		7 (0.2/0.2)		91 (1.1/1.1)		20 (0.7/0.8)	118 (0.5/0.5)
53단계				11 (0.1/0.1)		1 (0/0)	12 (0/0)
54단계				54 (0.6/0.7)		27 (1/1)	81 (0.3/0.3)
55단계				26 (0.3/0.3)		3 (0.1/0.1)	30 (0.1/0.1)
56단계				50 (0.6/0.6)			50 (0.2/0.2)
57단계				4 (0/0)			4 (0/0)
58단계				5 (0.1/0.1)			5 (0/0)
전체	2,303 (76.6/60.8)	3,417 (82.8/83.3)	3,519 (82.4/90.8)	5,697 (66.7/68.8)	1,191 (48.2/68.1)	1,405 (51.8/53)	17,532 (69.8/71.7)

주: 첫 번째 숫자는 매칭이 이루어진 사례수; 괄호 안의 첫 번째 숫자는 건축물대장을 기준으로 한 매칭달성 백분율; 괄호 안의 두 번째 숫자는 시험조사 자료를 기준으로 한 매칭달성 백분율.

<표 1-18> 제3차 시험조사 자료의 단계별 주택 매칭키 및 매칭작업 결과

구분	단계	지역	주소				건물명	면적	건축연도	층	수작업	매칭건수	잔여행정자료	잔여시험조사	건축물대장				시험조사										
			AM 본번	AS 부번	AD 동명	AR 호명									구분매칭 %	구분누적 %	전체 %	전체누적 %	구분매칭 %	구분누적 %	전체 %	전체누적 %							
												18,306	18,280																
APT	1	◇	◇	◇	◇	◇	◇	◇			3,000	15,306	15,280	16.4	16.4	11.2	11.2	16.4	16.4	11.1	11.1								
	2	◇	◇	◇	◇	◇	◇				1	15,305	15,279	0.0	16.4	0.0	11.2	0.0	16.4	0.0	11.1								
	3	◇	◇	◇	◇	◇		◇			45	15,260	15,234	0.2	16.6	0.2	11.3	0.2	16.7	0.2	11.3								
	4	◇	◇	◇	◇	◇		◇			8,223	7,037	7,011	44.9	61.6	30.6	41.9	45.0	61.6	30.5	41.8								
	5	◇	◇	◇	◇	◇					17	7,020	6,994	0.1	61.7	0.1	42.0	0.1	61.7	0.1	41.8								
	6	◇	◇	◇	◇	◇		◇			279	6,741	6,715	1.5	63.2	1.0	43.0	1.5	63.3	1.0	42.9								
	7	◇	◇	◇	◇	◇		◇			2,652	4,089	4,063	14.5	77.7	9.9	52.9	14.5	77.8	9.8	52.7								
	8	◇	◇	◇	◇	◇					95	3,994	3,968	0.5	78.2	0.4	53.2	0.5	78.3	0.4	53.1								
	9	◇	◇	◇	△	◇	◇	◇		◇	97	3,897	3,871	0.5	78.7	0.4	53.6	0.5	78.8	0.4	53.4								
	10	◇	◇	◇	△	◇	◇	◇		◇	2,012	1,885	1,859	11.0	89.7	7.5	61.1	11.0	89.8	7.5	60.9								
	11	◇	◇	◇	△	◇		◇		◇	129	1,756	1,730	0.7	90.4	0.5	61.5	0.7	90.5	0.5	61.4								
	12	◇	◇	◇	△	◇	◇			◇	214	1,542	1,516	1.2	91.6	0.8	62.3	1.2	91.7	0.8	62.2								
	13	◇	◇	◇	△	◇				◇	39	1,503	1,477	0.2	91.8	0.1	62.5	0.2	91.9	0.1	62.3								
	14	◇	◇		△	◇	◇	◇		◇	65	1,438	1,412	0.4	92.1	0.2	62.7	0.4	92.3	0.2	62.5								
	15	◇	◇		△	◇	◇	◇		◇	251	1,187	1,161	1.4	93.5	0.9	63.7	1.4	93.6	0.9	63.5								
	16	◇	◇		△	◇	◇			◇	1	1,186	1,160	0.0	93.5	0.0	63.7	0.0	93.7	0.0	63.5								
	17	◇		◇	△	◇		◇	◇	◇	251	935	909	1.4	94.9	0.9	64.6	1.4	95.0	0.9	64.4								
	18	◇		◇	△	◇		◇		◇	2	933	907	0.0	94.9	0.0	64.6	0.0	95.0	0.0	64.4								
	19	◇			△	◇	◇			◇	460	473	447	2.5	97.4	1.7	66.3	2.5	97.6	1.7	66.1								
APT	APT외는 APT가 아닌 모든 주택을 뜻하며 18단계까지 APT매칭을 하고 남은 건축물대장 자료(473)와 시험조사 자료(447)을 합쳐 매칭 APT 외라 한다. 19단계부터 매칭 APT외 자료(건축물대장 자료(9,062), 시험조사 자료(9,137))를 사용하여 매칭을 실시한다.											8,589	8,690																
	20	◇	◇	◇	◇	◇	◇	◇	◇	◇	1,690	7,372	7,447	18.6	18.6	6.3	72.6	18.5	18.5	6.3	72.4								
	21	◇	◇	◇	◇	◇	◇	◇		◇	157	7,215	7,290	1.7	20.4	0.6	73.2	1.7	20.2	0.6	73.0								
	22	◇	◇	◇	◇	◇	◇	◇		◇	1,344	5,871	5,946	14.8	35.2	5.0	78.2	14.7	34.9	5.0	78.0								
	23	◇	◇	◇	◇	◇		◇	◇	◇	564	5,307	5,382	6.2	41.4	2.1	80.3	6.2	41.1	2.1	80.0								
	24	◇	◇	◇	◇	◇	◇			◇	1,747	3,560	3,635	19.3	60.7	6.5	86.8	19.1	60.2	6.5	86.5								
	25	◇	◇	◇	◇	◇		◇		◇	4	3,556	3,631	0.0	60.8	0.0	86.8	0.0	60.3	0.0	86.5								
	26	◇	◇	◇	◇	◇			◇	◇	613	2,943	3,018	6.8	67.5	2.3	89.1	6.7	67.0	2.3	88.8								
	27	◇	◇	◇	◇	◇				◇	57	2,886	2,961	0.6	68.2	0.2	89.3	0.6	67.6	0.2	89.0								
	28	◇	◇	◇	◇	◇	◇	◇	◇	◇	25	2,861	2,936	0.3	68.4	0.1	89.4	0.3	67.9	0.1	89.1								
	29	◇	◇	◇	△	◇	◇	◇	◇	◇	27	2,834	2,909	0.3	68.7	0.1	89.5	0.3	68.2	0.1	89.2								
	30	◇	◇	◇	△	◇	◇			◇	4	2,830	2,905	0.0	68.8	0.0	89.5	0.0	68.2	0.0	89.2								
	31	◇	◇	◇	△	◇		◇	◇	◇	15	2,815	2,890	0.2	68.9	0.1	89.5	0.2	68.4	0.1	89.3								

<표 1-18> 제3차 시험조사 자료의 단계별 주택 매칭키 및 매칭작업 결과

구분	단계	지역	주소				건물명	면적	건축연도	층	수작업	매칭건수	잔여행정자료	잔여시험조사	건축물대장				시험조사			
			AM	AS	AD	AR									구분매칭%	구분누적%	전체누적%	구분매칭%	구분누적%	전체누적%		
			본번	부번	동명	호명																
			31단계까지 매칭 후 남은 자료										2,815	2,890								
	32	◇	◇	◇	△	◇				◇	◇	1	2,814	2,889	0.0	68.9	0.0	89.5	0.0	68.4	0.0	89.3
	33	◇	◇	◇		◇	◇	◇	◇	◇	19	2,795	2,870	0.2	69.2	0.1	89.6	0.2	68.6	0.1	89.4	
	34	◇	◇	◇		◇	◇		◇	◇	1	2,794	2,869	0.0	69.2	0.0	89.6	0.0	68.6	0.0	89.4	
	35	◇	◇	◇		◇		◇	◇	◇	147	2,647	2,722	1.6	70.8	0.5	90.2	1.6	70.2	0.5	89.9	
	36	◇	◇	◇		◇	◇	◇	◇	◇	1	2,646	2,721	0.0	70.8	0.0	90.2	0.0	70.2	0.0	89.9	
	37	◇	◇	◇		◇			◇	◇	24	2,622	2,697	0.3	71.1	0.1	90.3	0.3	70.5	0.1	90.0	
	38	◇	◇	◇	△	◇			◇	◇	23	2,599	2,674	0.3	71.3	0.1	90.3	0.3	70.7	0.1	90.1	
	39	◇	◇	◇	△	◇				◇	1	2,598	2,673	0.0	71.3	0.0	90.3	0.0	70.7	0.0	90.1	
	40	◇	◇	◇		◇	◇	◇	◇	◇	11	2,587	2,662	0.1	71.5	0.0	90.4	0.1	70.9	0.0	90.1	
	41	◇	◇		◇	◇		◇	◇	◇	1	2,586	2,661	0.0	71.5	0.0	90.4	0.0	70.9	0.0	90.1	
	42	◇			◇	◇	◇		◇	◇	17	2,569	2,644	0.2	71.7	0.1	90.4	0.2	71.1	0.1	90.2	
	43	◇			◇	◇	◇			◇	6	2,563	2,638	0.1	71.7	0.0	90.5	0.1	71.1	0.0	90.2	
	44	◇		◇		◇	△			◇	80	2,483	2,558	0.9	72.6	0.3	90.8	0.9	72.0	0.3	90.5	
	45	◇	◇		◇	◇	△			◇	18	2,465	2,540	0.2	72.8	0.1	90.8	0.2	72.2	0.1	90.6	
	46	◇			△	◇	△			◇	3	2,462	2,537	0.0	72.8	0.0	90.8	0.0	72.2	0.0	90.6	

<표 1-19> 제3차 시험조사 자료의 행정구역별 주택(APT)의 매칭작업 결과

	광주 북구 문흥1동	광주 광산구 월곡1동	충북 제천시 교동	충북 음성군 소이면	경남 김해시 활천동	경남 고성군 거류면	전체
1단계	2,437 (44.7/44.6)		507 (8.5/8.6)		56 (0.8/0.8)		3,000 (11.2/11.1)
2단계	1 (0/0)						1 (0/0)
3단계	43 (0.8/0.8)		1 (0/0)		1 (0/0)		45 (0.2/0.2)
4단계	2,122 (39/38.8)	553 (13.7/13.9)	3,549 (59.8/59.9)		1,712 (23.2/23)	287 (11.9/12.3)	8,223 (30.6/30.5)
5단계	3 (0.1/0.1)		11 (0.2/0.2)		3 (0/0)		17 (0.1/0.1)
6단계					279 (3.8/3.7)		279 (1/1)
7단계	24 (0.4/0.4)	2,469 (61.1/62)	4 (0.1/0.1)		153 (2.1/2.1)	2 (0.1/0.1)	2,652 (9.9/9.8)
8단계	3 (0.1/0.1)	34 (0.8/0.9)	32 (0.5/0.5)		18 (0.2/0.2)	8 (0.3/0.3)	95 (0.4/0.4)
9단계				97 (5.7/5.4)			97 (0.4/0.4)
10단계					2,012 (27.3/27)		2,012 (7.5/7.5)
11단계				129 (7.6/7.1)			129 (0.5/0.5)
12단계				190 (11.2/10.5)	24 (0.3/0.3)		214 (0.8/0.8)
13단계				39 (2.3/2.2)			39 (0.1/0.1)
14단계					65 (0.9/0.9)		65 (0.2/0.2)
15단계					251 (3.4/3.4)		251 (0.9/0.9)
16단계					1 (0/0)		1 (0/0)
17단계	247 (4.5/4.5)		3 (0.1/0.1)		1 (0/0)		251 (0.9/0.9)
18단계	1 (0/0)				1 (0/0)		2 (0/0)
19단계			283 (4.8/4.8)			177 (7.4/7.6)	460 (1.7/1.7)
20단계	208 (3.8/3.8)	445 (11/11.2)	205 (3.5/3.5)	6 (0.4/0.3)	744 (10.1/10)	82 (3.4/3.5)	1,690 (6.3/6.3)
21단계	14 (0.3/0.3)	35 (0.9/0.9)	36 (0.6/0.6)	11 (0.6/0.6)	25 (0.3/0.3)	36 (1.5/1.5)	157 (0.6/0.6)
22단계	120 (2.2/2.2)	284 (7/7.1)	391 (6.6/6.6)	108 (6.4/6)	242 (3.3/3.3)	199 (8.3/8.5)	1,344 (5/5)
23단계	1 (0/0)				563 (7.6/7.6)		564 (2.1/2.1)
24단계	43 (0.8/0.8)	77 (1.9/1.9)	302 (5.1/5.1)	485 (28.6/26.9)	96 (1.3/1.3)	744 (30.9/31.8)	1,747 (6.5/6.5)
25단계	1 (0/0)				3 (0/0)		4 (0/0)
26단계			371 (6.3/6.3)		226 (3.1/3)	16 (0.7/0.7)	613 (2.3/2.3)
27단계	2 (0/0)		6 (0.1/0.1)		48 (0.7/0.6)	1 (0/0)	57 (0.2/0.2)
28단계	20 (0.4/0.4)	1 (0/0)			4 (0.1/0.1)		25 (0.1/0.1)
29단계					27 (0.4/0.4)		27 (0.1/0.1)
30단계					4 (0.1/0.1)		4 (0/0)

주: 첫 번째 숫자는 매칭이 이루어진 사례수; 괄호 안의 첫 번째 숫자는 건축물대장을 기준으로 한 매칭달성 백분율; 괄호 안의 두 번째 숫자는 시험조사 자료를 기준으로 한 매칭달성 백분율.

<표 1-19> 제3차 시험조사 자료의 행정구역별 APT 외 주택의 매칭작업 결과

	광주 북구 문흥1동	광주 광산구 월곡1동	충북 제천시 교동	충북 음성군 소이면	경남 김해시 활천동	경남 고성군 거류면	전체
31단계					15 (0.2/0.2)		15 (0.1/0.1)
32단계					1 (0/0)		1 (0/0)
33단계					19 (0.3/0.3)		19 (0.1/0.1)
34단계					1 (0/0)		1 (0/0)
35단계					147 (2/2)		147 (0.5/0.5)
36단계					1 (0/0)		1 (0/0)
37단계					9 (0.1/0.1)	15 (0.6/0.6)	24 (0.1/0.1)
38단계					23 (0.3/0.3)		23 (0.1/0.1)
39단계					1 (0/0)		1 (0/0)
40단계					11 (0.1/0.1)		11 (0/0)
41단계					1 (0/0)		1 (0/0)
42단계				17 (1/0.9)			17 (0.1/0.1)
43단계				6 (0.4/0.3)			6 (0/0)
44단계					80 (1.1/1.1)		80 (0.3/0.3)
45단계			18 (0.3/0.3)				18 (0.1/0.1)
46단계					3 (0/0)		3 (0/0)
전체	5,290 (97.1/96.7)	3,898 (96.4/97.9)	5,719 (96.4/96.5)	1,088 (64.2/60.3)	6,871 (93.2/92.3)	1,567 (65.1/66.9)	24,433 (90.8/90.6)

주: 첫 번째 숫자는 매칭이 이루어진 사례수; 괄호 안의 첫 번째 숫자는 건축물대장을 기준으로 한 매칭달성 백분율; 괄호 안의 두 번째 숫자는 시험조사 자료를 기준으로 한 매칭달성 백분율.

제2차 시험조사 지역은 총 58단계의 매칭단계를 거쳐 17,532건이 매칭되었다. 이는 건축물대장 자료 기준으로 69.8%, 시험조사 자료 기준으로 71.7%의 매칭률을 나타낸다. 특히 부산 동래구 명장1동에서 3,417건이 매칭되어 건축물대장 자료 기준으로 82.8%로 가장 높은 매칭률을 보이며 경기 용인시 기흥동 시험조사 자료 기준에서는 이 3,519건이 매칭되어 90.8%로 가장 높은 매칭률을 보인다. 반면에 강원도 동해시 묵호동에서 1,191건이 매칭되었으며, 이는 건축물대장 자료 기준에서 48.2%의 매칭율로 가장 낮으며 강원 인제군 인제읍이 시험조사 자료 기준에서 53.0%로 가장 낮은 매칭율을 보인다.

제3차 시험조사 지역에서는 46단계의 매칭단계를 통하여 24,433건이 매칭

되었으며, 이는 건축물대장 자료 기준으로 90.8%, 시험조사 자료 기준으로 90.6%의 매칭률을 나타낸다. 제3차 시험조사 지역에서 광주 광산구 월곡1동에서 3,898건이 매칭되어 건축물대장 자료 기준으로 96.4%, 시험조사 자료 기준으로 97.9%가 매칭되어 제3차 시험조사 지역에서 가장 높은 매칭률을 나타냈다. 반면에 충북 음성군 소이면은 1,088건이 매칭되어 건축물대장 자료 기준에서 64.2%, 시험조사 자료 기준에서 60.3%로 가장 낮은 매칭률을 보인다.

특이사항은 제2차 시험조사 지역의 묵호동의 경우, 1단계에서 매칭 되는 것이 없었고, 건물명을 제외한 3단계에서 매칭이 시작된다. 즉, 강원도 묵호동의 경우 건물명을 재정비할 필요가 있다.

본번표준화를 제외한 10단계에서는 광남동의 매칭률이 높게 나타나는데 경기도 광남동의 본번표준화의 차이는 10이하로 같은 주택으로 볼 수 있다고 판단되었다.

본번표준화와 부번표준화를 제외한 12단계에서는 경기도 기흥시의 매칭률이 높게 나타났다.

34단계부터는 건물명이 없는 자료와 그 전 단계까지 매칭되지 않은 자료를 모두 합쳐서 다시 모든 매칭키를 사용하여 매칭을 실시한 단계로 부산 명장1동의 매칭률이 높게 나타났는데 이는 일반주택의 비율이 70%이상으로 높게 나타났다. 면적을 제외한 36단계에서는 전체적으로 1~762까지의 차이를 나타내는데 1~10 차이의 비율이 40%정도이며 그 중 부산 명장1동과 강원도 인제읍의 경우는 50%이상의 높은 비율을 보인다. 부산 범천1동의 경우 1~10 차이의 비율은 15%정도로 낮으며 전체적으로 퍼져 있다. 건물명을 제외한 37단계 매칭에서는 부산 명장1동이 매칭률이 높는데 이는 건축물대장의 건물명 누락으로 발생한 것이다. 면적과 건축연도를 제외한 38단계에서는 강원도 묵호동과 강원도 인제읍에서 매칭률이 높게 나타났는데 면적의 차이가 1인 비율은 높지만 1~959까지 전체적으로 퍼져 있다. 건축연도의 경우 50%이상 이 건축연도코드가 1차이를 나타낸다.

제3차 자료에서는 모든 매칭키를 고려하여 매칭한 1단계에서 광주 광산구 월곡1동과 충북 음성군 소이면, 경남 고성군 거류면은 매칭되는 자료가 없는 것으로 나타난다. 4단계인 건물명을 제외한 매칭에서 많은 자료가 매칭되는데 이는 두 자료의 건물명이 '문흥동명지아파트'와 '명지아파트'와 같이 한쪽 자료에 지역이름을 포함한 건물명으로 표기된 경우가 대부분이었다.

건물명, 동명 과 면적을 제외한 7단계에서는 광주시 월곡1동이 높은 매칭

을을 보이는데, 면적의 경우 1,2 차이를 보이며 건물명은 4단계와 마찬가지로 지역이름을 포함한 APT명과 지역이름을 포함하지 않은 APT명에 인한 차이이다. 특히 광주 광산구 월곡 1동에서 2개의 APT가 이와 같은 현상을 나타낸다.

동명표준화를 제외한 9단계에서는 충북 소이면에서만 매칭이 이루어졌는데, 건축물대장 자료에는 1동으로 기록되어있고 시험조사 자료에는 101동으로 기록되어 있었다. 동명표준화의 규칙적인 차이로 생기는 비매칭을 피하고자 이후 단계부터는 동명표준화를 재표준화 하여 이용하였다.

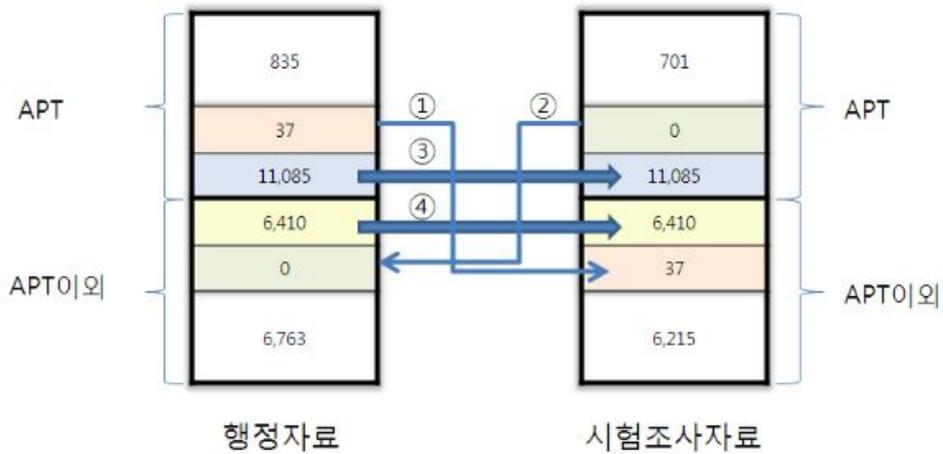
9단계에서부터는 동명표준화를 재 표준화하여 사용하였으며 면적을 제외한 10단계에서 경남 김해시 활천동이 높은 비율로 매칭이 되었다. 면적차이는 10이하의 차이가 대부분으로 조사되었다.

이 후 “매칭아파트 외 주택”에서 건물명을 제외한 23단계에서 경남 활천동의 매칭률이 매우 높는데 이는 건축물대장의 건물명이 누락된 자료와 건물명의 완전일치가 되지 않은 사례가 많은 것으로 나타났다. 예를 들어 보면 건축물대장에서 “경화2차타운”이면 시험조사 자료에선 “경화타운”으로 조사되었으며 또한 건축물대장에서 “~빌라”이면 시험조사 자료에선 “~빌”, “~빌리지”, “연립주택”등으로 조사되었다.

면적과 건축연도코드를 제외한 24단계에서 충북소이면과 경남거류면의 매칭률이 높게 나타났다. 이는 면적의 경우 충북소이면과 경남거류면의 시험조사 자료의 누락이 많이 존재하였으며 1~10차이의 비율이 30%정도이며 전체적으로 1~894까지 전체적으로 분포하였다. 건축연도의 경우 건축연도코드가 전체적으로 1차이 나는 비율이 38%정도이며 충북소이면과 경남거류면의 경우 시험조사 자료에서 누락이 존재하며 1차이 뿐 만 아니라 전체적으로 차이를 보이고 있다.

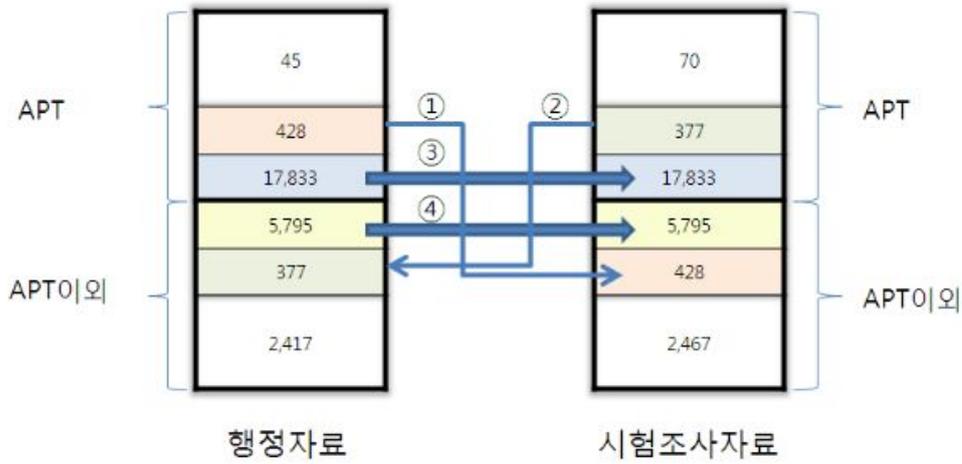
건물명과 면적을 제외한 26단계에서는 충북 교동과 경남 활천동의 매칭률이 높게 나타났다. 충북교동의 경우 23단계와 같은 형태로 건물명이 다르며 건축물대장의 면적이 누락된 자료이다. 활천동은 면적이 1~10차이가 70%이상으로 나타났다.

매칭결과, 주택유형이 상이한 경우도 매칭이 되었는데, 대표적으로 APT와 APT 이외의 주택으로 구분하여 매칭된 형태를 보면 <그림 1-1>과 <그림 1-2>와 같다. 또한 APT와 APT 이외의 주택으로 구분한 매칭률이 <표 1-20>와 <표 1-21>에 있다. 주택유형이 APT인 경우에 행정자료 기준으로 제2차에서 93%, 제3차에서 99.8%로 높은 매칭률을 보이고 있다.



- ① 건축물대장 자료에서 APT이나 시험조사 자료에서 APT이외의 주택과 매칭된 결과
- ② 시험조사 자료 APT이나 건축물대장 자료에서 APT이외의 주택과 매칭된 결과
- ③ 건축물대장 자료에서 APT이며 시험조사 자료 APT와 매칭된 결과
- ④ 건축물대장 자료에서 APT이외이며 시험조사 자료 APT이외와 매칭된 결과

<그림 1-1> 2차 자료 매칭과정



- ① 건축물대장 자료에서 APT이나 시험조사 자료에서 APT이외의 주택과 매칭된 결과
- ② 시험조사 자료 APT이나 건축물대장 자료에서 APT이외의 주택과 매칭된 결과
- ③ 건축물대장 자료에서 APT이며 시험조사 자료 APT와 매칭된 결과
- ④ 건축물대장 자료에서 APT이외이며 시험조사 자료 APT이외와 매칭된 결과

<그림 1-2> 3차 자료 매칭 과정

<표 1-20> 건축물대장과 시험조사 자료를 활용한 매칭 결과, 2008

행정동	전체				APT				APT 외			
	전체수	매칭수	실거주수	매칭율1	전체수	매칭수	실거주수	매칭률1	전체수	매칭수	실거주수	매칭률1
2차 시험조사 지역												
전체	25,130	17,532	0	69.8	11,957	11,122	0	93.0	13,173	6,410	0	48.7
범천1동	3,007	2,303	0	76.6	1,868	1,723	0	92.2	1,139	580	0	50.9
명장1동	4,129	3,417	0	82.8	889	873	0	98.2	3,240	2,544	0	78.5
기흥동	4,271	3,519	0	82.4	3,468	3,155	0	91.0	803	364	0	45.3
광남동	8,536	5,697	0	66.7	4,748	4,679	0	98.5	3,788	1,018	0	26.9
목호동	2,473	1,191	0	48.2	562	351	0	62.5	1,911	840	0	44.0
인제읍	2,714	1,405	0	51.8	422	341	0	80.8	2,292	1,064	0	46.4
3차 시험조사 지역												
전체	26,895	24,433	23,466	90.8	18,306	18,261	17,648	99.8	8,589	5,818	5,539	67.7
문흥1동	5,446	5,290	5,279	97.1	4,882	4,881	4,876	100.0	564	403	403	71.5
월곡1동	4,043	3,898	3,850	96.4	3,058	3,056	3,026	99.9	985	824	824	83.7
교동	5,935	5,719	5,417	96.4	4,396	4,390	4,111	99.9	1,539	1,306	973	84.9
소이면	1,694	1,088	766	64.2	456	455	227	99.8	1,238	539	541	43.5
활천동	7,369	6,871	6,773	93.2	5,029	5,005	4,953	99.5	2,340	1,820	1,869	77.8
거류면	2,408	1,567	1,381	65.1	485	474	455	97.7	1,923	926	929	48.2

매칭률1 : 매칭수/전체수

참고) 행정자료 기준

<표 1-21>건축물대장과 시험조사 자료를 활용한 매칭 결과, 2008

행정동	전체						APT						APT 외						
	전체 수	매칭 수	다가구 수	실거주 수	매칭률1	매칭률2	전체 수	매칭 수	다가구 수	실거주 수	매칭률1	매칭률2	전체 수	매칭 수	다가구 수	실거주 수	매칭률1	매칭률2	
2차 시험조사 지역	전체	24,448	17,532	38	0	71.7	71.9	11,786	11,085	0	0	94.1	94.1	12,662	6,447	38	0	50.9	51.2
	범천1동	3,786	2,303	7	0	60.8	61.0	1,806	1,722	0	0	95.3	95.3	1,980	581	7	0	29.3	29.7
	명장1동	4,101	3,417	3	0	83.3	83.4	1,208	848	0	0	70.2	70.2	2,893	2,569	3	0	88.8	88.9
	기흥동	3,877	3,519	6	0	90.8	90.9	3,182	3,150	0	0	99.0	99.0	695	369	6	0	53.1	54.0
	광남동	8,281	5,697	12	0	68.8	68.9	4,682	4,674	0	0	99.8	99.8	3,599	1,023	12	0	28.4	28.8
	목호동	1,750	1,191	1	0	68.1	68.1	411	351	0	0	85.4	85.4	1,339	840	1	0	62.7	62.8
	인제읍	2,653	1,405	9	0	53.0	53.3	497	340	0	0	68.4	68.4	2,156	1,065	9	0	49.4	49.8
3차 시험조사 지역	전체	26,970	24,433	16	23,472	90.6	90.7	18,657	18,551	0	17,933	99.4	99.4	8,313	5,882	16	5,539	70.8	70.9
	문흥1동	5,470	5,290	0	5,279	96.7	96.7	4,882	4,881	0	4,876	100.0	100.0	588	409	0	403	69.6	69.6
	월곡1동	3,983	3,898	0	3,850	97.9	97.9	3,056	3,056	0	3,026	100.0	100.0	927	842	0	824	90.8	90.8
	교동	5,926	5,719	0	5,417	96.5	96.5	4,729	4,728	0	4,444	100.0	100.0	1,197	991	0	973	82.8	82.8
	소이면	1,805	1,088	4	767	60.3	60.5	454	454	0	226	100.0	100.0	1,351	634	4	541	46.9	47.2
	활천동	7,444	6,871	3	6,775	92.3	92.3	4,991	4,957	0	4,906	99.3	99.3	2,453	1,914	3	1,869	78.0	78.1
	거류면	2,342	1,567	9	1,384	66.9	67.3	545	475	0	455	87.2	87.2	1,797	1,092	9	929	60.8	61.3

매칭률1 : 매칭수/전체수

매칭률2 : 다가구포함매칭수/전체수 (다가구포함 = 매칭수 + 다가구수)

참고) 시험조사 자료 기준

(APT와 APT 외 구분은 거처종류일치 변수를 사용하여 '2'이면 APT, 그 외는 APT 외로 구분하였다)

시험조사 자료와 건축물대장 자료를 매칭할 때 매칭률을 높이고자 층변수를 생성하였다. 여기서 시험조사 자료에서는 거처번호 중심으로 층번호를 생성하였기 때문에 2가구 1주택인 경우는 1개 주택이 매칭건수로 매칭률에 포함되었다. 그런데 가구입장에서 보면 2가구가 매칭된 것이므로 다가구를 포함한 매칭수로 1개 주택에 매칭된 가구들을 모두 매칭건수로 매칭률에 포함하는 또 하나의 매칭률을 계산한다.

그리고 또한 제3차 시험조사의 경우에 시험조사 자료와 건축물대장 자료와 매칭하여 실제 사람이 살고 있는 주택과 사람이 살고 있지 않은 주택으로 구분할 수 있어서 실거주수를 표에 제시하였다.

3. 세대-가구매칭자료

“주소매칭방법”과 “구성원매칭방법”으로 매칭된 세대-가구수를 <표 1-22>에 나타냈다.

주소매칭에서는 제2차 시험조사 자료는 주민등록 자료의 33,756세대 중 58.1%, 시험조사 자료의 27,862가구 중 60.6%의 매칭률을 보였으며 행정동별로는 주민등록 자료 기준 세대의 매칭률은 범천 1동 73.5%, 명장 1동 94.2%, 목호동 71.1%의 비교적 높은 매칭률을 보인 반면에 기흥동은 50.7%, 인제읍 51.0%를 보이고 광남동의 경우는 35.4%로 저조한 매칭률을 보였다.

제3차 시험조사 자료는 주민등록 자료의 31,471세대 중 79.9%, 시험조사 자료의 32,041가구 중 72.2%의 매칭률을 보였으며 행정동별로는 주민등록 자료 기준 세대의 매칭률은 문흥1동, 월곡1동, 교동은 90% 이상의 높은 매칭률을 보였고, 소이면, 활천동, 거류면는 각각 68.1%, 61.3%, 55.7% 매칭률을 보였다.

구성원매칭에서는 제2차 시험조사 자료는 주민등록 자료의 33,756세대 중 66.9%, 시험조사 자료의 27,862가구 중 77.4%의 매칭률로 주소매칭방법으로 매칭된 것보다 다소 높다. 행정동별로는 주민등록 자료 기준 세대의 매칭률은 명장1동, 범천 1동은 86.8%, 73.6%의 높은 매칭률을 보이고 기흥동, 목호동, 광남동, 인제읍은 각각 68.3%, 60.3%, 60.1%, 47.7%의 매칭률을 보였다.

제3차 시험조사 자료는 주민등록 자료의 31,471세대 중 78.9%, 시험조사 자료의 32,041가구 중 76.0%의 매칭률을 보였으며 행정동별로는 주민등록 자료 기준 세대의 매칭률은 동단위인 문흥1동, 월곡1동, 교동, 활천동은 78.1~86.8%의 높은 매칭률을 보인 반면 면단위인 소이면은 55.7%, 거류면은 46.6%의 비교적 저조한 매칭률을 보였다.

<표 1-22> 주소매칭/구성원매칭의 매칭건수와 매칭률

(단위 : 빈도, %)

		세대수			가구수			
		전체	매칭수	매칭률(%)	전체	매칭수	매칭률(%)	
주소매칭	전체	33,756	19,626	58.1	27,862	16,889	60.6	
	2차 시험조사 지역	범천1동	4,605	3,386	73.5	4,697	3,077	65.5
		명장1동	6,356	5,988	94.2	5,880	5,561	94.6
		기흥동	5,320	2,696	50.7	3,951	2,177	55.1
	주소매칭 지역	광남동	11,532	4,079	35.4	8,830	3,444	39.0
		목호동	2,211	1,572	71.1	1,776	1,221	68.8
		인제읍	3,732	1,905	51.0	2,728	1,409	51.6
	전체	31,471	25,145	79.9	32,041	23,143	72.2	
	3차 시험조사 지역	문흥1동	6,631	6,004	90.5	6,718	5,814	86.5
		월곡1동	5,767	5,434	94.2	5,454	4,962	91.0
		교동	6,486	6,026	92.9	6,036	5,483	90.9
		소이면	1,527	1,040	68.1	1,433	747	52.1
		활천동	8,671	5,311	61.3	10,308	5,124	49.7
		거류면	2,389	1,330	55.7	2,092	1,013	48.4
전체	33,756	22,575	66.9	27,862	21,552	77.4		
구성원매칭	2차 시험조사 지역	범천1동	4,605	3,388	73.6	4,697	3,296	70.2
		명장1동	6,356	5,514	86.8	5,880	5,394	91.7
		기흥동	5,320	3,634	68.3	3,951	3,394	85.9
	주소매칭 지역	광남동	11,532	6,927	60.1	8,830	6,497	73.6
		목호동	2,211	1,333	60.3	1,776	1,265	71.2
		인제읍	3,732	1,779	47.7	2,728	1,706	62.5
	전체	31,471	24,841	78.9	32,041	24,357	76.0	
	3차 시험조사 지역	문흥1동	6,631	5,754	86.8	6,718	5,697	84.8
		월곡1동	5,767	4,504	78.1	5,454	4,386	80.4
		교동	6,486	5,383	83.0	6,036	5,232	86.7
		소이면	1,527	850	55.7	1,433	800	55.8
		활천동	8,671	7,237	83.5	10,308	7,173	69.6
		거류면	2,389	1,113	46.6	2,092	1,069	51.1

세대-가구의 매칭결과 1세대-1가구 유형만 나타나지 않았다. 이에 세대-가구의 매칭결과를 1세대-1가구 유형, 1세대-다가구 유형, 다세대-1가구, 다세대-다가구 유형으로 구분하여 매칭결과 분석한다. 세대-가구의 매칭유형을 아래 표에 나타냈다. 여기서 다세대-1가구 유형은 #150인 가구원은 #50인 세대의 세대원과 매칭되었다. 그런데 #50인 세대원의 일부는 #160인 가구의 가구원과 매칭되었다. 또다시 #160인 가구의 가구원 중 일부는 #70인 세대의 세대원과 매칭되었다. 이런 유형을 다세대-다가구유형으로 분류한다.

<표 1-23> 세대번호-가구번호 매칭 유형

유형	세대 번호	가구 번호	매칭된 세대수	매칭된 가구수
1세대-1가구	#10	#110	1	1
1세대-다가구	#20	#120	1	2
	#20	#130	1	2
다세대-1가구	#30	#140	2	1
	#40	#140	2	1
다세대-다가구	#50	#150	1	2
	#50	#160	3	2
	#60	#160	3	1
	#70	#160	3	1
	#80	#170	1	2
	#80	#180	2	2
	#90	#180	2	1

그런데, “주소매칭방법”과 “구성원매칭방법”에 따라 매칭결과 특정현상들이 발생할 수 있다. 우선 “주소매칭방법”에서 발생할 수 있는 특정현상에서는 1세대-다가구, 다세대-1가구유형을 발생시키고 “구성원매칭방법”에서 발생할 수 있는 특정현상에서는 다세대-다가구유형이 발생한다. 이러한 예를 아래 표에 요약하였다.

“주소매칭방법”은 주소만을 매칭키로 하기 때문에 구성원의 정보에 대한 검토의 여지가 없다. 예를 들면, 행정자료에 세대 #1과 시험조사 자료의 가구 #10, #20, #30의 주소일치로 세대-가구를 매칭시킴으로써 결과적으로 1세대-다가구의 매칭이 된다. 그러나 이를 “구성원매칭방법”을 통하면 1세대-1가구의 매칭이 되는 것이다. 또한 이는 다세대-1가구의 매칭이 되는 경우도 발생하게 된다.

“구성원매칭방법”은 최대한 매칭률을 높인 인구매칭결과를 이용하기 때문에 근본적으로 구성원 매칭의 오류로 인한 다세대-다가구의 매칭이 발생할 수 있다. 또한, 세대 #1의 세대원은 가구 #10의 가구원과 매칭되었는데 가구 #10의 일부가구원은 다시 세대 #2와 매칭되었다. 그리고 세대 #2의 일부 세대원은 #30 가구의 가구원과 매칭됨으로써 복잡한 형태의 다세대-다가구유형으로 분류된다.

<표 1-24> 주소매칭방법과 구성원매칭방법의 차이

		세대번호	행정자료	조사자료	관계	가구번호	
주소매칭	1세대-다가구	#1	홍길동	홍길동	세대주	#10	
			박양치	박양치	배우자		
			홍다홍	홍다홍	자녀		
			홍치마	홍치마	자녀	#20	
			이미지		동거인		
					최강	세대주	#30
					김현모	배우자	
				최하나	자녀		
	다세대-1가구	#1	김기수		세대주	#30	
			양하나		배우자		
#3		홍길동	홍길동	세대주			
구성원매칭	다세대-다가구	#1	홍길동	홍길동	세대주	#10	
			김하나	김하나	자녀		
			김두리	김두리	자녀		
		#2	최고집	최고집	세대주	#30	
			김현모	김현모	배우자		
			최일번	최일번	자녀		
			최이번	최이번	자녀		
	다세대-다가구 (인구자료 매칭오류)	#1	홍길동	홍길동	세대주	#10	
			박양치	박양치	배우자		
			홍다홍	홍다홍	자녀		
#2		홍치아	홍치마	자녀(매칭 오류)	#30		
		홍고집	홍고집	세대주			
		김현모	김현모	배우자			
		이모친	이모친	어머니			

세대-가구의 4가지 유형에 따른 매칭률은 <표 1-25>에 정리하였다.

우선 “주소매칭”를 보면 제2차 시험조사 자료는 19,626 세대 중 1세대-1가구가 49.3%, 다세대-1가구 21.0%, 다세대-다가구 28.4%를 차지한 반면 제3차 시험조사 자료는 매칭된 25,145세대 중 1세대-1가구가 66.3%, 다세대-1가구 12.1%, 다세대-다가구 20.1%의 매칭률을 보였다.

제2차 시험조사 지역별로 보면, 경기도(기흥동, 광남동)는 1세대-1가구가 63.3%, 61.3%, 다세대-1가구는 31.6%, 27.9%로 대부분을 차지하고, 부산광역시(범천 1동, 명장 1동)는 1세대-1가구가 46.8%, 42.2%, 다세대-다가구는 37.4, 49.9%를, 다세대-1가구 13.7%, 7.1% 순으로 차지한다. 반면, 강원도(목호동, 인재읍)는 1세대-1가구 44.2%, 34.6%를, 다세대-1가구 30.3%, 39.9%, 다세대-다가구 22.5%, 22.7%의 매칭률을 보였다. 제3차 시험조사 지역별로 보면 활천동의 경우 1세대-1가구의 비율이 84.3%로 대부분을 차지하고 문흥1동은 1세대-1가구 70.3%, 다세대-다가구 18.8%, 다세대-1가구 7.2% 순으로 차지한 반면 교동은 1세대-1가구 71.6%, 다세대-1가구 17.4%, 다세대-다가구 10.3%를 차지하였으며, 월곡 1동은 1세대-1가구 54.5%, 다세대-다가구 35.1%로 대부분을 차지하였다. 반면 면단위의 경우 다세대-다가구의 비율이 53.7, 35.1%로 가장 높고 1세대-1가구 22.1, 34.3%, 다세대-1가구 21.8, 27.3%로 분포됨을 볼 수 있다. 소이면과 거류면은 1세대-다가구, 다세대-1가구, 다세대-다가구의 모든 비율이 타지역보다 높게 나타나고 있다.

이어서 “구성원매칭”의 경우에 제2차 시험조사 자료는 매칭된 인구자료 22,575세대 중 1세대-1가구가 89.2%, 다세대-1가구가 9.5%를 차지한 반면, 제3차 시험조사 자료는 매칭된 인구자료 24,841세대 중 1세대-1가구가 94.7%, 다세대-1가구가 4.5%를 차지함을 볼 수 있다. 매칭률을 지역별로 보면, 제2차 시험조사 지역은 1세대-1가구가 86.0~93.2%로 대부분을 차지하며, 기흥동, 광남동, 목호동, 인재읍은 다세대-1가구가 9.1~12.5% 정도 차지한다. 제3차 시험조사 지역은 1세대-1가구가 88.0~97.3%로 대부분을 차지하며, 면단위에서 다세대-1가구가 11.8%, 8.3% 나타난다.

<표 1-25> 주소매칭/구성원매칭방법의 행정동별 매칭된 세대-가구 유형별 매칭률

(단위 : 빈도, %)

		1세대-1가구		1세대-다가구		다세대-1가구		다세대-다가구		
		세대	가구	세대	가구	세대	가구	세대	가구	
주소 매칭 방법	전체	9670	9670	261	665	4120	1673	5575	4881	
		49.3	57.3	1.3	3.9	21.0	9.9	28.4	28.9	
	2차 시험 조사 지역	범천1동	1584	1584	70	174	464	194	1268	1125
			46.8	51.5	2.1	5.7	13.7	6.3	37.4	36.6
		명장1동	2526	2526	50	107	426	196	2986	2732
			42.2	45.4	0.8	1.9	7.1	3.5	49.9	49.1
		기흥동	1706	1706	6	18	853	372	131	81
			63.3	78.4	0.2	0.8	31.6	17.1	4.9	3.7
		광남동	2499	2499	35	122	1140	489	405	334
			61.3	72.6	0.9	3.5	27.9	14.2	9.9	9.7
		목호동	695	695	48	119	476	157	353	250
			44.2	56.9	3.1	9.7	30.3	12.9	22.5	20.5
	인제읍	660	660	52	125	761	265	432	359	
		34.6	46.8	2.7	8.9	39.9	18.8	22.7	25.5	
	전체	16660	16660	396	854	3038	1277	5051	4352	
		66.3	72.0	1.6	3.7	12.1	5.5	20.1	18.8	
	3차 시험 조사 지역	문흥1동	4223	4223	223	461	432	165	1126	965
			70.3	72.6	3.7	7.9	7.2	2.8	18.8	16.6
월곡1동		2961	2961	41	104	526	233	1906	1664	
		54.5	59.7	0.8	2.1	9.7	4.7	35.1	33.5	
교동		4313	4313	47	103	1047	455	619	612	
		71.6	78.7	0.8	1.9	17.4	8.3	10.3	11.2	
소이면		230	230	25	53	227	88	558	376	
		22.1	30.8	2.4	7.1	21.8	11.8	53.7	50.3	
활천동		4477	4477	16	37	443	197	375	413	
		84.3	87.4	0.3	0.7	8.3	3.8	7.1	8.1	
거류면	456	456	44	96	363	139	467	322		
	34.3	45.0	3.3	9.5	27.3	13.7	35.1	31.8		
구성원 매칭 방법	전체	20134	20134	107	216	2140	1029	194	173	
		89.2	93.4	0.5	1.0	9.5	4.8	0.9	0.8	
	2차 시험 조사 지역	범천1동	3029	3029	40	80	228	107	91	80
			89.4	91.9	1.2	2.4	6.7	3.2	2.7	2.4
		명장1동	5139	5139	32	65	298	147	45	43
			93.2	95.3	0.6	1.2	5.4	2.7	0.8	0.8
		기흥동	3162	3162	4	8	455	212	13	12
			87.0	93.2	0.1	0.2	12.5	6.2	0.4	0.4
		광남동	6059	6059	6	12	841	408	21	18
			87.5	93.3	0.1	0.2	12.1	6.3	0.3	0.3
		목호동	1147	1147	14	29	157	77	15	12
			86.0	90.7	1.1	2.3	11.8	6.1	1.1	0.9
	인제읍	1598	1598	11	22	161	78	9	8	
		89.8	93.7	0.6	1.3	9.1	4.6	0.5	0.5	
	전체	23527	23527	84	169	1115	546	115	115	
		94.7	96.6	0.3	0.7	4.5	2.2	0.5	0.5	
	3차 시험 조사 지역	문흥1동	5526	5526	23	46	159	78	46	47
			96.0	97.0	0.4	0.8	2.8	1.4	0.8	0.8
월곡1동		4174	4174	24	49	280	137	26	26	
		92.7	95.2	0.5	1.1	6.2	3.1	0.6	0.6	
교동		5017	5017	16	32	325	159	25	24	
		93.2	95.9	0.3	0.6	6.0	3.0	0.5	0.5	
소이면		748	748	2	4	100	48	0	0	
		88.0	93.5	0.2	0.5	11.8	6.0	0.0	0.0	
활천동		7044	7044	16	32	159	79	18	18	
		97.3	98.2	0.2	0.4	2.2	1.1	0.2	0.3	
거류면	1018	1018	3	6	92	45	0	0		
	91.5	95.2	0.3	0.6	8.3	4.2	0.0	0.0		

4. 행정자료 매칭

향후 매칭된 행정자료를 활용하기 위한 준비단계로서 본 연구에서 사용하고 있는 주민등록 자료와 건축물대장 자료를 매칭한 결과는 <표 1-26>에 요약하였다. 주민등록 자료 기준으로 세대 매칭률을 언급하고, 건축물대장 자료 기준 주택수 매칭률을 언급한다.

먼저 세대 매칭률을 본다. 제2차 시험조사 자료의 세대는 33,756세대 중 73.5%(24,817세대)가 매칭되었다. 지역별로는 명장1동이 92.9%, 기흥동이 84.7%의 높은 매칭률을 보이는 반면 범천1동 54.7%, 인제읍 61.7%의 낮은 매칭률을 보였다. 제3차 시험조사 자료의 세대는 31,471세대 중 82.8%(26,053세대)가 매칭되었다. 지역별로는 동단위의 지역은 활천동을 제외하고 90% 이상의 매칭률을 보이는 반면, 면단위의 지역의 경우 50%대의 매칭률을 보인다. 이를 통해 유추해 볼 때, 면단위에서 행정자료의 주소 불일치 사례가 많을 것으로 판단된다.

이어서 건축물 매칭률을 본다. 제2차 시험조사 자료의 주택은 25,210건 중 76%(19,153건)이 매칭되었다. 지역별로는 명장1동이 90.6%, 기흥동이 87.4%, 인제읍이 73.9%, 광남동이 73.8%, 목호동이 62.8%, 범천1동이 59.1%의 매칭률을 보였다. 제3차 시험조사 자료의 주택은 26,895건 중 78.7%(21,158건)이 매칭되었다. 지역별로는 동단위의 지역이 활천동을 제외하고 90% 근처의 매칭률을 보였고, 면단위 지역에서는 40%수준의 매칭률을 보였다.

<표 1-26> 행정자료 매칭수와 매칭률

(단위 : 빈도, %)

		전체	세대수 매칭수	매칭률(%)	전체	주택수 매칭수	매칭률(%)
전체		33,756	24,817	73.5	25,210	19,153	76.0
2차 시협 조사 지역	범천1동	4,605	2,519	54.7	3,085	1,822	59.1
	명장1동	6,356	5,904	92.9	4,131	3,743	90.6
	기흥동	5,320	4,505	84.7	4,271	3,732	87.4
	광남동	11,532	7,866	68.2	8,536	6,299	73.8
	목호동	2,211	1,721	77.8	2,473	1,552	62.8
인제읍	3,732	2,302	61.7	2,714	2,005	73.9	
전체		31,471	26,053	82.8	26,895	21,158	78.7
3차 시협 조사 지역	문흥1동	6,631	6,335	95.5	5,446	5,102	93.7
	월곡1동	5,767	5,493	95.2	4,043	3,758	93.0
	교동	6,486	6,089	93.9	5,935	5,235	88.2
	소이면	1,527	811	53.1	1,694	572	33.8
	활천동	8,671	5,912	68.2	7,369	5,332	72.4
거류면	2,389	1,413	59.1	2,408	1,159	48.1	

제2차 시험조사 자료에서 1세대-1주택 비율이 세대 54.9%, 주택 71.1%로 가장 높으며 다음으로 다세대-1주택이 세대 37.3%, 주택 18.5%로 나타남을 볼 수 있다. 제3차 시험조사 자료에서 1세대-1주택의 비율이 세대 67.7%, 주택 83.4%로 가장 높으며 다음으로 다세대-1주택의 비율이 세대 28.3%, 주택 12.1%로 나타남을 볼 수 있다. 이를 다르게 표현하면 건축물 대장에 등재된 주택의 약 12.1%의 주택은 2세대 이상과 매칭되었음을 의미한다. 역으로 주민등록 세대 중 0.6%(149세대)는 2개 이상의 주택과 매칭되었음을 알 수 있다.

제2차 시험조사 자료를 세대-주택 유형에 대해 행정동별로 요약해 보면 범천1동, 기흥동, 광남동은 주민등록에 등재된 세대 중 51.1%에서 68.8%가 건축물대장의 주택 중 70.7%에서 83.1%의 주택에 1:1로 매칭되어 전반적으로 높은 매칭률을 보이고 다세대-1주택의 비율은 세대 중 28.9%에서 41.2%가 주택 중 15.2%에서 17.8%의 주택에 매칭되어 두 유형이 대부분을 차지함을 볼 수 있는 반면 명장1동은 세대 42.6%의 세대가 주택 67.2%에 1:1로 매칭되었고 다세대-1주택 비율이 세대 56.2%가 주택 31.2%를 차지함을 볼 수 있다. 또한 강원도(목호동, 인제읍)는 1세대-1주택 비율이 세대는 40.0%, 43.15%가 주택 44.4%, 49.5%로 매칭되고 다세대-1주택은 세대 20.2%, 35.0%가 주택 9.4%, 14.9%로 매칭되며 다세대-다주택은 세대 36.3%, 18.2%가 주택 36.2%, 26.5%로 매칭되어 유형별로 분산됨 볼 수 있다.

제3차 시험조사 자료를 세대-주택 유형에 대해 행정동별로 요약해 보면 동의 경우 주민등록에 등재된 세대 중 53.1%에서 84.5%가 건축물대장의 주택 중 77.6%에서 93.6%에 1:1로 매칭되어 전반적으로 높은 매칭률을 보이고 다세대-1주택의 비율은 세대 중 23.9%에서 34.0%가 주택 중 10.5%에서 22.4%의 주택에 매칭되어 두 유형이 대부분을 차지함을 볼 수 있다. 반면에 면의 경우 1세대-1주택의 비율은 세대 중 27.4%, 38.4%가 주택 중 38.8%, 46.9%에 매칭되어 절반을 넘지 못하는데 다세대-다주택으로 세대 25.8%, 33.3%가 주택 23.2%, 28.0%로 매칭되어 동 단위 행정구역에 비하여 두드러진 것이 특징이다. 이는 다음의 유형별 주택 수 비율 표를 보건데 동의 경우 아파트나 연립주택과 같은 집합 건물이 상당수 존재하여 1세대-1주택에서 대부분이 나타남을 볼 수 있는 반면 면의 경우 일반주택이 많은 부분을 차지하여 네 유형에 분산됨을 볼 수 있다.

<표 1-27> 행정자료 매칭 행정동별 세대-주택 유형

(단위 : 빈도, %)

	1세대-1주택		1세대-다주택		다세대-1주택		다세대-다주택		
	세대	주택	세대	주택	세대	주택	세대	주택	
2차 시험 조사 지역	전체	13613 54.9	13613 71.1	202 0.8	460 2.4	9256 37.3	3537 18.5	1746 7.0	1543 8.1
	범천1동	1288 51.1	1288 70.7	21 0.8	45 2.5	1037 41.2	325 17.8	173 6.9	164 9.0
		명장1동	2517 42.6	2517 67.2	6 0.1	12 0.3	3316 56.2	1169 31.2	65 1.1
	기흥동	3100 68.8	3100 83.1	10 0.2	21 0.6	1304 28.9	569 15.2	91 2.0	42 1.1
		광남동	5027 63.9	5027 79.8	20 0.3	44 0.7	2447 31.1	1030 16.4	372 4.7
	목호동		689 40.0	689 44.4	60 3.5	155 10.0	347 20.2	146 9.4	625 36.3
	인제읍	992 43.1	992 49.5	85 3.7	183 9.1	805 35.0	298 14.9	420 18.2	532 26.5
	전체	17637 67.7	17637 83.4	149 0.6	338 1.6	7377 28.3	2566 12.1	890 3.4	617 2.9
	3차 시험 조사 지역	문흥1동	4566 72.1	4566 89.5	0 0.0	0 0.0	1769 27.9	536 10.5	0 0.0
월곡1동			2915 53.1	2915 77.6	0 0.0	0 0.0	2578 46.9	843 22.4	0 0.0
교동		4398 72.2	4398 84.0	18 0.3	36 0.7	1456 23.9	632 12.1	217 3.6	169 3.2
		소이면	222 27.4	222 38.8	43 5.3	100 17.5	276 34.0	90 15.7	270 33.3
활천동		4993 84.5	4993 93.6	5 0.1	11 0.2	876 14.8	309 5.8	38 0.6	19 0.4
거류면		543 38.4	543 46.9	83 5.9	191 16.5	422 29.9	156 13.5	365 25.8	269 23.2

제2차 시험조사 자료를 집합일반 구분별 주택 수에 대해 요약해 보면 1세대-1주택은 86.4%가 집합주택이 차지함을 볼 수 있으며, 1세대-다주택, 다세대-다주택은 일반주택이 97.0, 100%로 대부분을 차지하며, 다세대-1주택은 집합주택 38.4%, 일반주택 61.6%를 차지함을 볼 수 있다. 제3차 시험조사 자료를 집합일반 구분별 주택 수에 대해 요약해 보면 1세대-1주택은 92.9%가 집합주택이 차지함을 보이며, 1세대-다주택과 다세대-다주택은 일반주택이 100%이고, 다세대-1주택은 집합주택이 27.0%, 일반주택이 73.0%를 차지함을 볼 수 있다.

제2차 시험 조사자료를 주택유형별 주택 수 비율에 대해 요약해 보면 1세대-1주택은 집합주택 중 APT가 68.7%, 다세대주택이 12.9%를 차지함을 볼 수 있고, 일반단독주택이 11.0%를 차지함을 볼 수 있다. 반면 1세대-다주택과

다세대-다주택은 일단단독주택이 86.7, 93.2%로 대부분을 차지함을 볼 수 있고, 다세대-1주택은 집합주택 중 아파는 29.4%, 다세대주택 7.2%, 비거주용 건물 내 주택이 6.2%를 차지하고 일반주택 중 일반단독주택이 45.5%, 영업겸용단독주택이 7.8%를 차지함을 볼 수 있다.

제3차 시험조사 자료를 주택유형별 주택 수 비율에 대해 요약해 보면 1세대-1주택은 집합주택 중 APT가 84.2%, 다세대주택이 6.4%를 차지함을 볼 수 있고, 일반주택은 일반단독주택이 5.0%를 차지함을 보이며, 1세대-다주택과 다세대-다주택은 일반주택이 100%이며, 이중 일반단독이 95.9, 95.5%를 차지함을 볼 수 있고, 다세대-1주택은 집합주택 중 APT가 23.5%이고 일반주택 중 일반단독주택 28.2, 영업겸용단독주택 24.4, 다가구단독주택 14.3, 비거주용 건물 내 주택 6.2%를 차지함을 볼 수 있다.

<표 1-28> 행정자료 매칭 유형별 주택유형 주택수 비율

(단위 : 빈도, %)

	일반단독	다가구 단독	영업겸용 단독	아파트	연립주택	다세대 주택	비거주용 건물내	기숙사/ 사회시설
전체	4948	149	487	10392	726	2015	434	2
	25.8	0.8	2.5	54.3	3.8	10.5	2.3	0.0
1세대-1주택	1500	34	159	9352	657	1753	157	1
	11.0	0.2	1.2	68.7	4.8	12.9	1.2	0.0
2차 1세대-다주택	399	18	15	0	5	9	14	0
	86.7	3.9	3.3	0.0	1.1	2.0	3.0	0.0
다세대-1주택	1611	75	275	1040	64	253	218	1
	45.5	2.1	7.8	29.4	1.8	7.2	6.2	0.0
다세대-다주택	1438	22	38	0	0	0	45	0
	93.2	1.4	2.5	0.0	0.0	0.0	2.9	0.0
전체	2524	404	833	15454	428	1189	326	0
	11.9	1.9	3.9	73.0	2.0	5.6	1.5	0.0
1세대-1주택	888	35	194	14850	392	1137	141	0
	5.0	0.2	1.1	84.2	2.2	6.4	0.8	0.0
3차 1세대-다주택	324	2	2	0	0	0	10	0
	95.9	0.6	0.6	0.0	0.0	0.0	3.0	0.0
다세대-1주택	723	366	625	604	36	52	160	0
	28.2	14.3	24.4	23.5	1.4	2.0	6.2	0.0
다세대-다주택	589	1	12	0	0	0	15	0
	95.5	0.2	1.9	0.0	0.0	0.0	2.4	0.0

제2차 시험조사 자료를 매칭된 유형별 건물명에 대해 요약해 보면 1세대-1주택의 경우 43.0%가 행정자료에서 건물명이 정확히 일치하며, 건물명이 없는 경우가 14.5%, 주민등록 누락이 0.1%(14건), 건축물대장 누락이 10.0%(1,360건) 나타남을 볼 수 있으며 또한 32.4%가 건물명을 일치로 판정

하였으며 건물명이 상이하다 판정되는 경우도 1건 발생 했다. 1세대-다주택의 경우 건물명 없음이 97.0%를 차지하며 건축물대장 누락이 2.0%(9건) 발생하였고 건물명이 상이하다 판정된 건도 4건이 발생하였다. 다세대-1주택의 경우 14.1%가 행정자료에서 건물명이 정확히 일치하였으며, 건물명 없는 경우가 69.5%이고 주민등록누락이 0.1%(5건), 건축물대장 누락이 3.8%(353건)이고 또한 12.5%가 건물명을 일치로 판정하였으며 건물명이 상이하다 판정된 경우도 4건이 발생하였다. 다세대-다주택의 경우는 모두 건물명이 없는 경우이다.

제3차 시험조사 자료를 매칭된 유형별 건물명에 대해 요약해 보면 1세대-1주택의 경우 58.9%는 행정자료에서 건물명이 정확히 일치하며, 건물명이 없는 경우가 7.1%로 나타나고 주민등록이나 건축물대장 어느 한쪽의 누락도 0.2%(27건), 1.4%(242건)이 발생함을 볼 수 있다 또한 32.5%가 건물명을 일치로 판정하였으며 건물명이 상이하다 판정되는 경우도 1건이 발생함을 볼 수 있다. 1세대-다주택이나 다세대-다주택의 경우는 모두 건물명 없음으로 나타났으며, 다세대-1주택의 경우는 행정자료에서 건물명이 정확히 일치하는 경우 12.2%이고, 건물명이 없는 경우가 80.8%로 나타났으며, 누락된 경우도 주민등록 0.1%(6건), 건축물대장 0.3%(25건)이고, 또한 6.5%가 건물명을 일치로 판정하였으며 건물명이 상이한 경우도 0.1%(4건) 발생함을 볼 수 있다.

<표 1-29> 행정자료 매칭 유형별 건물명 정보

(단위 : 빈도, %)

유형	건물명						합계 (세대*주택)	
	일치	불일치		건축물		없음		
		일치 판정	불일치 판정	주민등록 누락	대장누락			
전체	7,155 20.6	5,573 16.0	9 0.0	19 0.1	1,724 5.0	20,291 58.4	34,771	
2차	1세대-1주택	5,851 43.0	4,415 32.4	1 0.0	14 0.1	1,360 10.0	1,972 14.5	13,613
	1세대-다주택	1 0.2	0 0.0	4 0.9	0 0.0	9 2.0	446 97.0	460
	다세대-1주택	1,303 14.1	1,158 12.5	4 0.0	5 0.1	353 3.8	6,433 69.5	9,256
	다세대-다주택	0 0.0	0 0.0	0 0.0	0 0.0	2 0.0	11,440 100.0	11,442
	전체	11,287 38.5	6,207 21.2	5 0.0	33 0.1	267 0.9	11,521 39.3	29,320
3차	1세대-1주택	10,385 58.9	5,724 32.5	1 0.0	27 0.2	242 1.4	1,258 7.1	17,637
	1세대-다주택	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	338 100.0	338
	다세대-1주택	902 12.2	483 6.5	4 0.1	6 0.1	25 0.3	5,957 80.8	7,377
	다세대-다주택	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	3,968 100.0	3,968
	전체	11,287 38.5	6,207 21.2	5 0.0	33 0.1	267 0.9	11,521 39.3	29,320

NOTE: 건물명이 불일치인 경우 판정은 눈으로 직접 확인한 결과

참고)부가적인 정보로 활용한 건물명이 매칭을 위하여 정확한 정보로서 역할을 하기 위해서는 건물명의 통일이 필요하다. 표에서 나타는 사례를 보면 다음과 같다.

<표 1-30> 건물명 일치, 불일치 사례

주소	주민등록대장 건물명	건축물대장 건물명	세대번호	주택번호	판정결과
777-1번지 123-456	문흥금호타운	금호타운	#11	#1001	일치
	현대아파트	금호타운	#12		불일치
888-1번지 605호	대현아트빌	대현아트빌라	#21	#2001	일치
888-1번지 606호	성진빌라	대현아트빌라	#22	#2002	불일치
333-1번지 225-789		주공아파트	#31	#3001	누락
555-1번지 654-321	성도빌라		#41	#4001	누락

제4절 인구/주택 매칭의 항목별 매칭률 평가

지금까지는 인구, 주택 자료를 단계별로 매칭시키는 과정을 설명하고 나타난 특징을 열거하였다. 본 절에서는 이와 같은 단계별 작업을 통하여 얻은 매칭 결과를 요약한다. 그런데 이 때 매칭률은 인구의 경우 3단계(S, M, W), 주택의 경우 2단계(S, M)로 요약하는데 S단계는 매칭키가 기타 주요 매칭키를 포함하면서 주소 정보 4항목 중 최소 2개 이상의 항목을 포함한 단계로서 모든 연구자들이 불일치 항목들은 오류로서 판단할 수 있다는 합의로 도출된 매칭결과들로 구성된다.

같은 맥락에서 W단계는 주소정보를 최대 한 개, 성명과 생년/월의 오차까지는 허용하고 신중한 수작업을 통하여 최종 매칭 여부를 결정해야 하는 매칭 결과들로 이루어져 있다.

개별단계들과 S, M, W단계의 관계는 다음과 같다.

<표 1-31> 인구, 주택 자료의 매칭 단계

	2차		3차	
	인구	주택	인구	주택
S	1~16단계	1~7, 19~41단계	1~9단계	1~13, 20~40단계
M	17~21단계	8~18, 42~58단계	10~18단계	14~19, 41~46단계
W	22~25단계		19단계	

1. 인구

매칭작업은 인구와 주택은 매칭관계의 강도에 따라서 S(Strong), M(Moderate), W(Weak)로 단계를 나누어 결과를 얻었는데 S, S+M, S+M+W 단계로 결과를 살펴본다.

가. S 단계

성별에 관한 매칭률에서 남녀 간에는 큰 차이가 없는데 여자가 다소 높다. 이 현상은 묵호동을 제외한 모든 행정동에서 나타나고 있다.

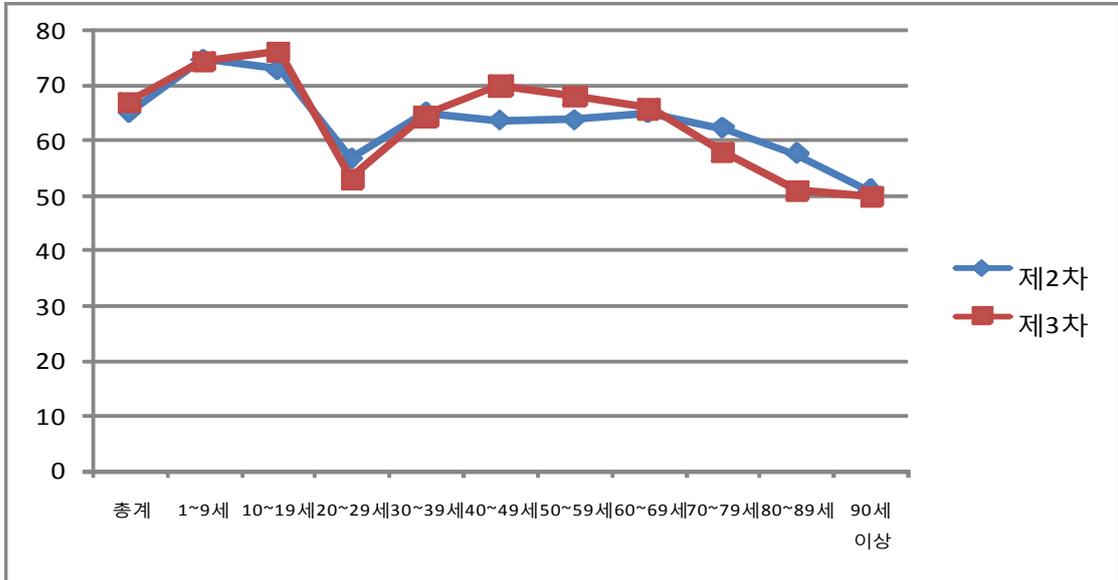
<표 1-32> 주민등록 자료기준 성별 매칭건수 및 매칭률

(단위 : 명, %)

	전체			남자			여자		
	전체	매칭수	매칭률	계	매칭수	매칭률	계	매칭수	매칭률
전체	87,414	56,924	65.1	44,406	28,101	63.3	43,008	28,823	67
범천1동	9,712	6,610	68.1	4,785	3,149	65.8	4,927	3,461	70.2
명장1동	18,321	15,165	82.8	9,121	7,387	81	9,200	7,778	84.5
2차 기흥동	13,283	9,249	69.6	6,811	4,633	68	6,472	4,616	71.3
광남동	31,772	18,784	59.1	16,344	9,367	57.3	15,428	9,417	61
목호동	4,786	2,653	55.4	2,382	1,327	55.7	2,404	1,326	55.2
인제읍	9,540	4,463	46.8	4,963	2,238	45.1	4,577	2,225	48.6
전체	91,175	61,172	67.1	45,952	30,185	65.7	45,223	30,987	68.5
문흥1동	20,958	16,346	78	10,334	7,979	77.2	10,624	8,367	78.8
월곡1동	16,637	12,669	76.1	8,450	6,266	74.2	8,187	6,403	78.2
3차 교동	18,433	14,682	79.7	9,213	7,259	78.8	9,220	7,423	80.5
소이면	3,423	1,416	41.4	1,778	710	39.9	1,645	706	42.9
활천동	26,381	14,045	53.2	13,377	6,991	52.3	13,004	7,054	54.2
거류면	5,343	2,014	37.7	2,800	980	35	2,543	1,034	40.7

연령에 관한 매칭률은 <표 1-33>에서 볼 수 있듯이 80대 이상의 인구를 제외하고 제2차 시험조사 지역에서는 1-9세가 74.7%로 가장 높았으며, 제3차 시험조사 지역에서는 10-19세가 76.1%로 가장 높았다. 또, 제2차 제3차 모두 20-29세가 56.6%, 53.1%로 가장 낮았다. 이러한 현상은 12개 행정동 모두에서 공통적으로 나타나는데 특히 제3차 시험조사 지역인 소이면과 거류면은 다른 행정동에 비해 20-29세, 30-39세의 매칭률이 모두 낮게 나타난 것이 특징이라고 할 수 있다. 제2차 시험조사 지역의 기흥동은 1-9세의 매칭률이 83.7%로 높게 나타났으며, 인제읍의 20-29세는 26.2%로 낮게 나타났다. 제3차 시험조사 지역에서는 교동의 1-9세의 인구가 89.4%의 매칭률을 보이며, 소이면과 거류면의 20-29세 인구의 매칭률은 17.9%, 18.5%로 다른 행정동에 비해 낮은 매칭률을 보이고 있다. 그림으로 살펴보면 제2차 시험조사 지역과 제3차 시험조사 지역의 S단계의 연령에 관한 매칭률을 비슷한 값을 보이고

있다.



(단위 : %)

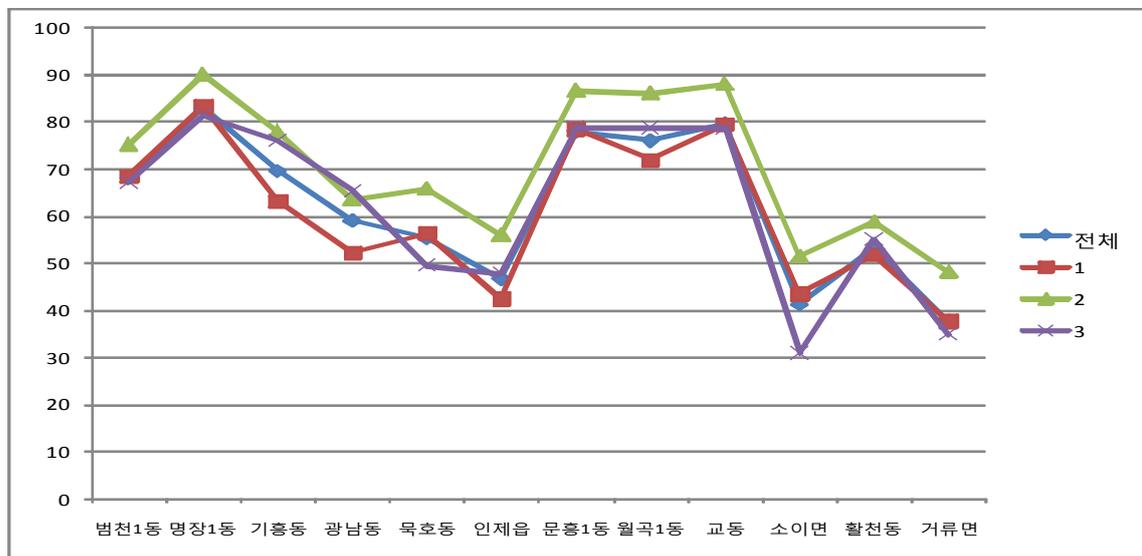
<그림 1-3> 연령별 매칭률

<표 1-33> 주민등록 자료기준 연령대 매칭건수 및 매칭률

(단위 : 명, %)

		전체	1~9세	10~19세	20~29세	30~39세	40~49세	50~59세	60~69세	70~79세	80~89세	90세 이상
전체	전체	87,414	9,237	10,253	12,590	16,394	14,382	11,713	7,715	3,766	1,234	130
	매칭수	56,924	6,904	7,466	7,129	10,651	9,166	7,474	5,011	2,348	709	66
	매칭률	65.1	74.7	72.8	56.6	65	63.7	63.8	65	62.3	57.5	50.8
법천 1동	전체	9,712	488	872	1,861	1,588	1,462	1,545	1,081	638	165	12
	매칭수	6,610	305	653	1,204	983	998	1,065	801	482	114	5
	매칭률	68.1	62.5	74.9	64.7	61.9	68.3	68.9	74.1	75.5	69.1	41.7
명장 1동	전체	18,321	1,045	2,576	3,167	2,371	3,381	3,194	1,672	653	243	19
	매칭수	15,165	906	2,289	2,321	1,867	2,808	2,770	1,455	548	187	14
	매칭률	82.8	86.7	88.9	73.3	78.7	83.1	86.7	87	83.9	77	73.7
기흥 동	전체	13,283	2,071	1,371	1,687	3,471	1,906	1,206	918	482	154	17
	매칭수	9,249	1,733	1,044	946	2,526	1,254	755	589	303	93	6
	매칭률	69.6	83.7	76.1	56.1	72.8	65.8	62.6	64.2	62.9	60.4	35.3
광남 동	전체	31,772	4,272	3,852	4,044	6,805	5,410	3,819	2,297	928	307	38
	매칭수	18,784	3,095	2,625	2,070	4,157	3,040	1,906	1,244	479	152	16
	매칭률	59.1	72.4	68.1	51.2	61.1	56.2	49.9	54.2	51.6	49.5	42.1
북호 동	전체	4,786	297	379	561	692	607	780	813	493	151	13
	매칭수	2,653	205	215	255	357	294	443	502	294	78	10
	매칭률	55.4	69	56.7	45.5	51.6	48.4	56.8	61.7	59.6	51.7	76.9
인제 읍	전체	9,540	1,064	1,203	1,270	1,467	1,616	1,169	934	572	214	31
	매칭수	4,463	660	640	333	761	772	535	420	242	85	15
	매칭률	46.8	62	53.2	26.2	51.9	47.8	45.8	45	42.3	39.7	48.4
전체	전체	91,175	8,803	16,316	12,396	14,578	18,026	10,460	5,541	3,726	1,173	156
	매칭수	61,172	6,552	12,415	6,580	9,404	12,613	7,121	3,653	2,157	599	78
	매칭률	67.1	74.4	76.1	53.1	64.5	70	68.1	65.9	57.9	51.1	50
문흥 1동	전체	20,958	1,866	4,213	3,256	2,898	4,038	2,573	1,200	654	222	38
	매칭수	16,346	1,528	3,582	2,115	2,121	3,287	2,148	978	442	127	18
	매칭률	78	81.9	85	65	73.2	81.4	83.5	81.5	67.6	57.2	47.4
월곡 1동	전체	16,637	1,669	3,359	2,491	2,734	3,374	1,734	676	431	143	26
	매칭수	12,669	1,365	2,811	1,588	1,954	2,626	1,379	516	309	104	17
	매칭률	76.1	81.8	83.7	63.7	71.5	77.8	79.5	76.3	71.7	72.7	65.4
교동	전체	18,433	2,196	2,992	2,147	3,357	3,450	2,155	1,160	719	216	41
	매칭수	14,682	1,963	2,653	1,179	2,732	2,918	1,683	897	500	133	24
	매칭률	79.7	89.4	88.7	54.9	81.4	84.6	78.1	77.3	69.5	61.6	58.5
소이 면	전체	3,423	153	298	363	307	475	525	546	551	186	19
	매칭수	1,416	85	128	65	90	166	222	279	294	78	9
	매칭률	41.4	55.6	43	17.9	29.3	34.9	42.3	51.1	53.4	41.9	47.4
활천 동	전체	26,381	2,619	4,969	3,448	4,515	5,936	2,767	1,263	672	175	17
	매칭수	14,045	1,474	2,989	1,505	2,273	3,360	1,404	660	309	65	6
	매칭률	53.2	56.3	60.2	43.6	50.3	56.6	50.7	52.3	46	37.1	35.3
거류 면	전체	5,343	300	485	691	767	753	706	696	699	231	15
	매칭수	2,014	137	252	128	234	256	285	323	303	92	4
	매칭률	37.7	45.7	52	18.5	30.5	34	40.4	46.4	43.3	39.8	26.7

<표 1-34>와 <그림 1-4>는 세대주 관계의 매칭건수와 매칭률을 요약하였다 <그림 1-4>에서 볼 수 있듯이 12개 행정동별 매칭률과 각 지역의 세대주(1), 배우자(2), 자녀(3)의 매칭률이 거의 비슷한 값을 보이는데 배우자의 매칭률이 높게 나타났다. 배우자의 매칭률은 제2차 시험조사 지역의 명장1동에서 90.1%로 가장 높게 나타났다. 자녀의 매칭률에서는 제3차 시험조사 지역의 소이면과 거류면이 30.9%, 35.0%로 다른 행정동에 비해 낮은 매칭률이 나타났다.



(단위 : %)

<그림 1-4> 행정동별 세대주 관계 매칭률

<표 1-34> 주민등록 자료기준 가구주관계 매칭건수 및 매칭률

(단위 : 명, %)

		전체	1	2	3	4	5	6	7	8	9	10	11	12	13	14
전체	전체	87,414	33,755	17,350	29,883	588	2,095	281	1,507	2	64	1,008	309	11	44	517
	매칭수	56,924	20,664	12,450	20,445	345	1,169	141	898	1	33	437	118	6	11	206
	매칭률	65.1	61.2	71.8	68.4	58.7	55.8	50.2	59.6	50	51.6	43.4	38.2	54.5	25	39.8
범천1동	전체	9,712	4,604	1,579	2,758	54	200	23	169	0	7	176	27	1	3	111
	매칭수	6,610	3,154	1,188	1,853	35	130	11	85	0	3	91	12	1	0	47
	매칭률	68.1	68.5	75.2	67.2	64.8	65	47.8	50.3	0	42.9	51.7	44.4	100	0	42.3
명장1동	전체	18,321	6,357	3,699	7,011	127	399	34	322	0	8	207	49	1	8	99
	매칭수	15,165	5,296	3,333	5,703	94	294	17	235	0	7	107	21	0	3	55
	매칭률	82.8	83.3	90.1	81.3	74	73.7	50	73	0	87.5	51.7	42.9	0	37.5	55.6
2차 기흥동	전체	13,283	5,320	2,625	4,362	76	373	43	158	0	23	152	68	6	6	71
	매칭수	9,249	3,360	2,048	3,317	47	229	21	97	0	7	59	28	3	2	31
	매칭률	69.6	63.2	78	76	61.8	61.4	48.8	61.4	0	30.4	38.8	41.2	50	33.3	43.7
광남동	전체	31,772	11,532	6,614	11,445	187	766	141	422	0	19	350	125	3	22	146
	매칭수	18,784	6,024	4,206	7,490	105	368	83	246	0	9	152	50	2	4	45
	매칭률	59.1	52.2	63.6	65.4	56.1	48	58.9	58.3	0	47.4	43.4	40	66.7	18.2	30.8
북호동	전체	4,786	2,210	907	1,213	51	102	9	189	0	4	50	13	0	5	33
	매칭수	2,653	1,244	597	602	19	53	4	99	0	4	11	3	0	2	15
	매칭률	55.4	56.3	65.8	49.6	37.3	52	44.4	52.4	0	100	22	23.1	0	40	45.5
인제읍	전체	9,540	3,732	1,926	3,094	93	255	31	247	2	3	73	27	0	0	57
	매칭수	4,463	1,586	1,078	1,480	45	95	5	136	1	3	17	4	0	0	13
	매칭률	46.8	42.5	56	47.8	48.4	37.3	16.1	55.1	50	100	23.3	14.8	0	0	22.8
전체	전체	91,175	31,469	16,875	35,438	1,079	2,065	383	1,105	9	42	1,032	381	13	532	752
	매칭수	61,172	20,553	12,621	24,397	709	1,036	157	625	2	17	435	142	8	200	270
	매칭률	67.1	65.3	74.8	68.8	65.7	50.2	41	56.6	22.2	40.5	42.2	37.3	61.5	37.6	35.9
문흥1동	전체	20,958	6,631	3,689	8,852	254	491	146	196	0	6	318	114	4	157	100
	매칭수	16,346	5,197	3,197	6,975	190	248	66	131	0	3	177	49	2	62	49
	매칭률	78	78.4	86.7	78.8	74.8	50.5	45.2	66.8	0	50	55.7	43	50	39.5	49
월곡1동	전체	16,637	5,767	2,778	6,854	223	316	53	142	0	10	209	77	5	78	125
	매칭수	12,669	4,149	2,392	5,391	165	191	25	102	0	7	103	39	4	40	61
	매칭률	76.1	71.9	86.1	78.7	74	60.4	47.2	71.8	0	70	49.3	50.6	80	51.3	48.8
3차 교동	전체	18,433	6,485	3,654	7,005	210	470	62	241	5	13	110	44	1	59	74
	매칭수	14,682	5,146	3,215	5,509	175	295	28	173	2	4	53	18	1	24	39
	매칭률	79.7	79.4	88	78.6	83.3	62.8	45.2	71.8	40	30.8	48.2	40.9	100	40.7	52.7
소이면	전체	3,423	1,527	689	835	59	109	5	124	0	1	28	3	0	11	32
	매칭수	1,416	666	355	258	25	47	1	44	0	1	6	2	0	2	9
	매칭률	41.4	43.6	51.5	30.9	42.4	43.1	20	35.5	0	100	21.4	66.7	0	18.2	28.1
활천동	전체	26,381	8,670	5,112	10,453	243	556	102	259	2	10	325	127	3	205	314
	매칭수	14,045	4,494	3,003	5,761	125	208	33	135	0	2	91	32	1	71	89
	매칭률	53.2	51.8	58.7	55.1	51.4	37.4	32.4	52.1	0	20	28	25.2	33.3	34.6	28.3
거류면	전체	5,343	2,389	953	1,439	90	123	15	143	2	2	42	16	0	22	107
	매칭수	2,014	901	459	503	29	47	4	40	0	0	5	2	0	1	23
	매칭률	37.7	37.7	48.2	35	32.2	38.2	26.7	28	0	0	11.9	12.5	0	4.5	21.5

1:가구주 2:가구주의 배우자 3:자녀 4:자녀의 배우자 5:세대주의 부모 6:배우자의 부모 7:손자녀, 그 배우자 8:증손자녀, 그 배우자 9:조부모 10:형제자매, 그 배우자 11:형제자매의 자녀, 그 배우자 12:부모의 형제자매, 그 배우자 13:기타 친인척 14:기타동거인

나. S+M 단계

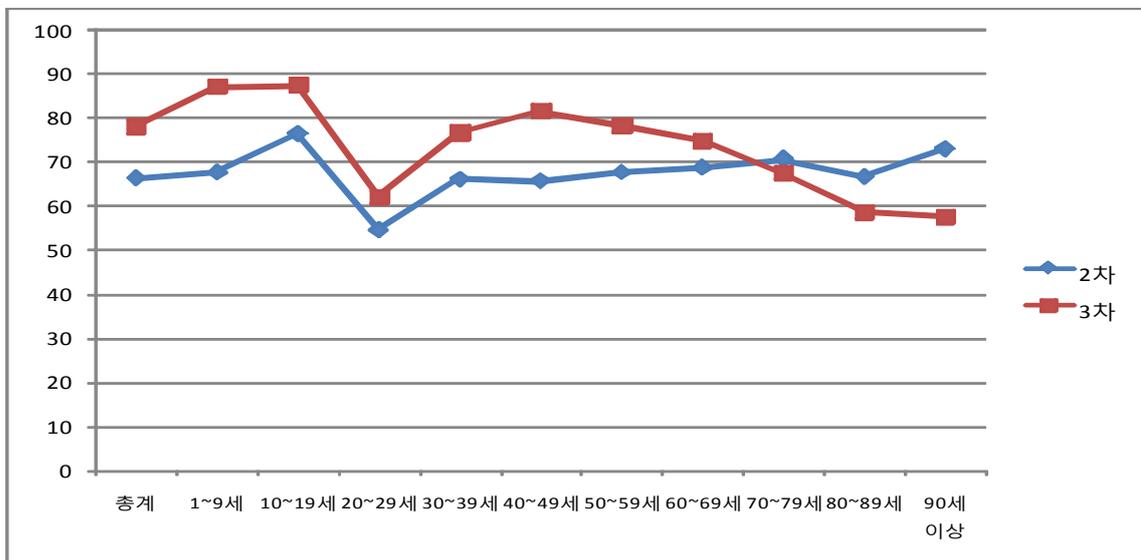
S+M단계도 S단계와 마찬가지로 남녀간에 큰 차이를 나타내지 않았으며, 제2차 시험조사 지역의 묵호동을 제외하고 여자의 매칭비율이 다소 높게 나타났다.

<표 1-35> 행정자료기준 성별 매칭건수 및 매칭률

(단위 : 명, %)

	전체			남자			여자		
	전체	매칭수	매칭률	전체	매칭수	매칭률	전체	매칭수	매칭률
전체	87,414	58,062	66.4	44,406	28,675	64.6	43,008	29,387	68.3
2차									
범천1동	9,712	6,942	71.5	4,785	3,309	69.2	4,927	3,633	73.7
명장1동	18,321	15,197	82.9	9,121	7,406	81.2	9,200	7,791	84.7
기흥동	13,283	9,301	70	6,811	4,661	68.4	6,472	4,640	71.7
광남동	31,772	19,290	60.7	16,344	9,623	58.9	15,428	9,667	62.7
묵호동	4,786	2,666	55.7	2,382	1,333	56	2,404	1,333	55.4
인제읍	9,540	4,666	48.9	4,963	2,343	47.2	4,577	2,323	50.8
3차									
전체	91,175	71,130	78	45,952	35,174	76.5	45,223	35,956	79.5
문흥1동	20,958	17,256	82.3	10,334	8,409	81.4	10,624	8,847	83.3
월곡1동	16,637	13,075	78.6	8,450	6,467	76.5	8,187	6,608	80.7
교동	18,433	14,934	81	9,213	7,386	80.2	9,220	7,548	81.9
소이면	3,423	1,690	49.4	1,778	849	47.8	1,645	841	51.1
활천동	26,381	21,823	82.7	13,377	10,908	81.5	13,004	10,915	83.9
거류면	5,343	2,352	44	2,800	1,155	41.3	2,543	1,197	47.1

S+M단계의 연령의 매칭률을 살펴보면, <그림 1-5>에서 볼 수 있듯이 제3차 시험조사 지역이 제2차 시험조사 지역보다 대부분 높은 일치 정도를 보이고 있으나 70대 이상부터는 제2차 시험조사 지역의 매칭률이 높게 나타났다. 또, S단계와 마찬가지로 20-29세에서는 제2차 시험조사 지역에서 54.5%, 제3차 시험조사 지역에서 62.1%로 연령대에서 가장 낮은 매칭률을 나타나고 있다. 특히, 인제읍에서 26.0%, 소이면 24.2%, 거류면 24.6%로 20-29세 중 낮은 매칭률을 나타내고 있다. 10-19세의 매칭률은 제2차 시험조사 지역은 76.5%, 제3차 시험조사 지역은 87.4%로 높게 나타났으며, 그 중에서 활천동이 91.0%로 가장 높은 매칭률을 보였다.



(단위 : %)

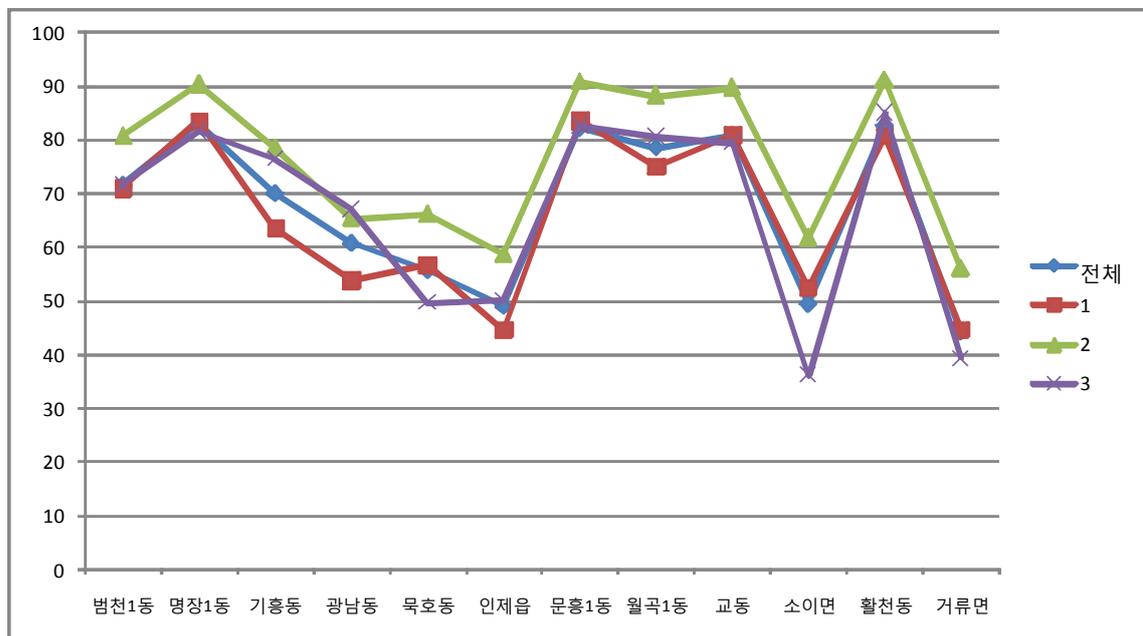
<그림 1-5> 행정동별 연령 매칭률

<표 1-36> 행정자료기준 연령별 매칭건수 및 매칭률

(단위 : 명, %)

		전체	1-9세	10-19세	20-29세	30-39세	40-49세	50-59세	60-69세	70-79세	80-89세	90세 이상	
2차	전체	87,414	9,237	10,253	12,590	16,394	14,382	11,713	7,715	3,766	1,234	130	
	매칭수	58,062	6,246	7,846	6,858	10,852	9,464	7,915	5,299	2,664	823	95	
	매칭률	66.4	67.6	76.5	54.5	66.2	65.8	67.6	68.7	70.7	66.7	73.1	
	법천 1동	전체	9,712	488	872	1,861	1,588	1,462	1,545	1,081	638	165	12
	매칭수	6,942	304	678	1,147	1,136	1,020	1,104	866	541	137	9	
	매칭률	71.5	62.3	77.8	61.6	71.5	69.8	71.5	80.1	84.8	83	75	
	명장 1동	전체	18,321	1,045	2,576	3,167	2,371	3,381	3,194	1,672	653	243	19
	매칭수	15,197	759	2,186	2,336	1,912	2,662	2,895	1,590	616	219	22	
	매칭률	82.9	72.6	84.9	73.8	80.6	78.7	90.6	95.1	94.3	90.1	115.8	
	기흥동	전체	13,283	2,071	1,371	1,687	3,471	1,906	1,206	918	482	154	17
	매칭수	9,301	1,606	1,117	843	2,513	1,363	798	600	333	116	12	
	매칭률	70	77.5	81.5	50	72.4	71.5	66.2	65.4	69.1	75.3	70.6	
	광남동	전체	31,772	4,272	3,852	4,044	6,805	5,410	3,819	2,297	928	307	38
	매칭수	19,290	2,786	2,901	1,972	4,139	3,321	2,096	1,318	564	173	20	
	매칭률	60.7	65.2	75.3	48.8	60.8	61.4	54.9	57.4	60.8	56.4	52.6	
	목호동	전체	4,786	297	379	561	692	607	780	813	493	151	13
	매칭수	2,666	182	225	230	363	300	427	508	330	89	12	
	매칭률	55.7	61.3	59.4	41	52.5	49.4	54.7	62.5	66.9	58.9	92.3	
	인제읍	전체	9,540	1,064	1,203	1,270	1,467	1,616	1,169	934	572	214	31
	매칭수	4,666	609	739	330	789	798	595	417	280	89	20	
	매칭률	48.9	57.2	61.4	26	53.8	49.4	50.9	44.6	49	41.6	64.5	
	3차	전체	91,175	8,803	16,316	12,396	14,578	18,026	10,460	5,541	3,726	1,173	156
		매칭수	71,130	7,665	14,257	7,704	11,166	14,716	8,185	4,144	2,515	688	90
		매칭률	78	87.1	87.4	62.1	76.6	81.6	78.3	74.8	67.5	58.7	57.7
문흥 1동		전체	20,958	1,866	4,213	3,256	2,898	4,038	2,573	1,200	654	222	38
매칭수		17,256	1,618	3,767	2,251	2,285	3,476	2,232	1,009	465	134	19	
매칭률		82.3	86.7	89.4	69.1	78.8	86.1	86.7	84.1	71.1	60.4	50	
월곡 1동		전체	16,637	1,669	3,359	2,491	2,734	3,374	1,734	676	431	143	26
매칭수		13,075	1,417	2,871	1,641	2,055	2,697	1,407	538	325	106	18	
매칭률		78.6	84.9	85.5	65.9	75.2	79.9	81.1	79.6	75.4	74.1	69.2	
교동		전체	18,433	2,196	2,992	2,147	3,357	3,450	2,155	1,160	719	216	41
매칭수		14,934	1,977	2,673	1,199	2,775	2,961	1,737	921	524	141	26	
매칭률		81	90	89.3	55.8	82.7	85.8	80.6	79.4	72.9	65.3	63.4	
소이면		전체	3,423	153	298	363	307	475	525	546	551	186	19
매칭수		1,690	95	151	88	106	206	267	328	341	96	12	
매칭률		49.4	62.1	50.7	24.2	34.5	43.4	50.9	60.1	61.9	51.6	63.2	
활천동		전체	26,381	2,619	4,969	3,448	4,515	5,936	2,767	1,263	672	175	17
매칭수		21,823	2,406	4,523	2,356	3,681	5,077	2,194	976	498	102	10	
매칭률		82.7	91.9	91	68.3	81.5	85.5	79.3	77.3	74.1	58.3	58.8	
거류면		전체	5,343	300	485	691	767	753	706	696	699	231	15
매칭수		2,352	152	272	169	264	299	348	372	362	109	5	
매칭률		44	50.7	56.1	24.5	34.4	39.7	49.3	53.4	51.8	47.2	33.3	

세대주 관계 항목의 매칭률을 살펴보면, S단계와 마찬가지로 12개 행정동별 매칭률과 각 지역의 세대주(1), 배우자(2), 자녀의(3) 매칭률이 거의 비슷한 값을 보이면서 배우자의 매칭률이 높게 나타났다. 세대주 관계에서는 배우자의 매칭률이 73.3%로 높게 나타났는데, 제2차 시험조사 지역의 명장1동이 90.3%, 제3차 시험조사 지역의 활천동이 91.1%로 다른 행정동에 비해 높은 비율을 보이고 있다. 또, 제3차 시험조사 지역에서 자녀에 관한 매칭률은 79.5%를 보이고 있지만 소이면과 거류면은 36.2%, 39.3%로 낮은 매칭률이 나타났다.



(단위 : %)

<그림 1-6> 행정동별 세대주 관계 매칭률

<표 1-37> 행정자료기준 가구주관계 매칭건수 및 매칭률

(단위 : 명, %)

		전체	1	2	3	4	5	6	7	8	9	10	11	12	13	14
전체	전체	87414	33755	17350	29883	588	2095	281	1507	2	64	1008	309	11	44	517
	매칭수	58062	21077	12719	20842	348	1194	146	909	1	33	445	120	6	11	211
	매칭률	66.4	62.4	73.3	69.7	59.2	57	52	60.3	50	51.6	44.1	38.8	54.5	25	40.8
범천1동	전체	9712	4604	1579	2758	54	200	23	169	0	7	176	27	1	3	111
	매칭수	6942	3262	1273	1972	36	136	13	88	0	3	95	14	1	0	49
	매칭률	71.5	70.9	80.6	71.5	66.7	68	56.5	52.1	0	42.9	54	51.9	100	0	44.1
명장1동	전체	18321	6357	3699	7011	127	399	34	322	0	8	207	49	1	8	99
	매칭수	15197	5306	3340	5713	94	296	17	238	0	7	107	21	0	3	55
	매칭률	82.9	83.5	90.3	81.5	74	74.2	50	73.9	0	87.5	51.7	42.9	0	37.5	55.6
2차 기흥동	전체	13283	5320	2625	4362	76	373	43	158	0	23	152	68	6	6	71
	매칭수	9301	3380	2057	3334	47	231	23	98	0	7	60	28	3	2	31
	매칭률	70	63.5	78.4	76.4	61.8	61.9	53.5	62	0	30.4	39.5	41.2	50	33.3	43.7
광남동	전체	31772	11532	6614	11445	187	766	141	422	0	19	350	125	3	22	146
	매칭수	19290	6209	4322	7669	107	382	84	250	0	9	155	50	2	4	47
	매칭률	60.7	53.8	65.3	67	57.2	49.9	59.6	59.2	0	47.4	44.3	40	66.7	18.2	32.2
북호동	전체	4786	2210	907	1213	51	102	9	189	0	4	50	13	0	5	33
	매칭수	2666	1254	599	603	19	53	4	99	0	4	11	3	0	2	15
	매칭률	55.7	56.7	66	49.7	37.3	52	44.4	52.4	0	100	22	23.1	0	40	45.5
인제읍	전체	9540	3732	1926	3094	93	255	31	247	2	3	73	27	0	0	57
	매칭수	4666	1666	1128	1551	45	96	5	136	1	3	17	4	0	0	14
	매칭률	48.9	44.6	58.6	50.1	48.4	37.6	16.1	55.1	50	100	23.3	14.8	0	0	24.6
전체	전체	91175	31469	16875	35438	1079	2065	383	1105	9	42	1032	381	13	532	752
	매칭수	71130	23969	14694	28179	810	1239	183	721	6	22	527	178	8	247	347
	매칭률	78	76.2	87.1	79.5	75.1	60	47.8	65.2	66.7	52.4	51.1	46.7	61.5	46.4	46.1
문흥1동	전체	20958	6631	3689	8852	254	491	146	196	0	6	318	114	4	157	100
	매칭수	17256	5543	3351	7321	204	263	71	141	0	3	188	54	2	65	50
	매칭률	82.3	83.6	90.8	82.7	80.3	53.6	48.6	71.9	0	50	59.1	47.4	50	41.4	50
월곡1동	전체	16637	5767	2778	6854	223	316	53	142	0	10	209	77	5	78	125
	매칭수	13075	4317	2449	5527	171	201	26	105	0	8	113	45	4	44	65
	매칭률	78.6	74.9	88.2	80.6	76.7	63.6	49.1	73.9	0	80	54.1	58.4	80	56.4	52
3차 교동	전체	18433	6485	3654	7005	210	470	62	241	5	13	110	44	1	59	74
	매칭수	14934	5248	3277	5570	179	311	28	175	3	6	54	18	1	24	40
	매칭률	81	80.9	89.7	79.5	85.2	66.2	45.2	72.6	60	46.2	49.1	40.9	100	40.7	54.1
소이면	전체	3423	1527	689	835	59	109	5	124	0	1	28	3	0	11	32
	매칭수	1690	800	425	302	29	59	2	49	0	1	8	2	0	2	11
	매칭률	49.4	52.4	61.7	36.2	49.2	54.1	40	39.5	0	100	28.6	66.7	0	18.2	34.4
활천동	전체	26381	8670	5112	10453	243	556	102	259	2	10	325	127	3	205	314
	매칭수	21823	6994	4658	8894	192	351	52	204	2	4	158	57	1	110	146
	매칭률	82.7	80.7	91.1	85.1	79	63.1	51	78.8	100	40	48.6	44.9	33.3	53.7	46.5
거류면	전체	5343	2389	953	1439	90	123	15	143	2	2	42	16	0	22	107
	매칭수	2352	1067	534	565	35	54	4	47	1	0	6	2	0	2	35
	매칭률	44	44.7	56	39.3	38.9	43.9	26.7	32.9	50	0	14.3	12.5	0	9.1	32.7

1:가구주 2:가구주의 배우자 3:자녀 4:자녀의 배우자 5:세대주의 부모 6:배우자의 부모 7:손자녀, 그 배우자 8:증손자녀, 그 배우자 9:조부모 10:형제자매, 그 배우자 11:형제자매의 자녀, 그 배우자 12:부모의 형제자매, 그 배우자 13:기타 친인척 14:기타동거인

다. S+M+W단계

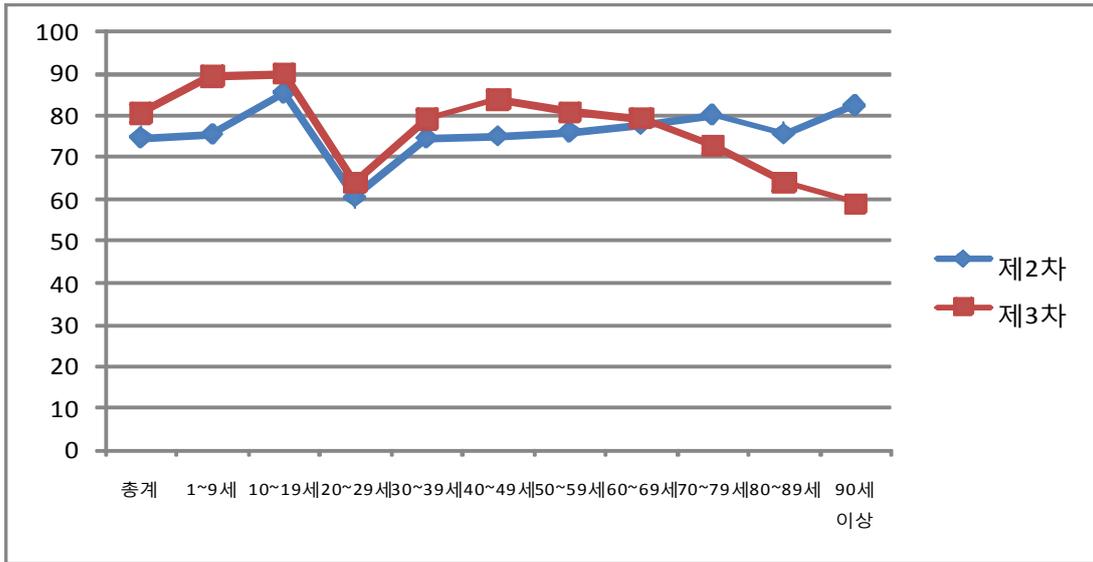
S+M+W단계에서도 성별에 관한 매칭률은 남녀에 관하여 큰 차이가 보이지 않으며, 대부분의 지역에서 여자에 매칭률이 높게 나타났다.

<표 1-38> 주민등록 자료기준 성별 매칭건수 및 매칭률

(단위 : 명, %)

	전체			남자			여자		
	전체	매칭수	매칭률	전체	매칭수	매칭률	전체	매칭수	매칭률
전체	87,414	65,267	74.7	44,406	32,330	72.8	43,008	32,937	76.6
2차									
법천1동	9,712	7,105	73.2	4,785	3,385	70.7	4,927	3,720	75.5
명장1동	18,321	15,481	84.5	9,121	7,545	82.7	9,200	7,936	86.3
기흥동	13,283	10,130	76.3	6,811	5,067	74.4	6,472	5,063	78.2
광남동	31,772	23,702	74.6	16,344	11,876	72.7	15,428	11,826	76.7
목호동	4,786	3,039	63.5	2,382	1,515	63.6	2,404	1,524	63.4
인제읍	9,540	5,810	60.9	4,963	2,942	59.3	4,577	2,868	62.7
3차									
전체	91,175	73,438	80.5	45,952	36,397	79.2	45,223	37,041	81.9
문흥1동	20,958	17,543	83.7	10,334	8,552	82.8	10,624	8,991	84.6
월곡1동	16,637	13,329	80.1	8,450	6,593	78	8,187	6,736	82.3
교동	18,433	15,310	83.1	9,213	7,574	82.2	9,220	7,736	83.9
소이면	3,423	2,074	60.6	1,778	1,063	59.8	1,645	1,011	61.5
활천동	26,381	22,071	83.7	13,377	11,037	82.5	13,004	11,034	84.9
거류면	5,343	3,111	58.2	2,800	1,578	56.4	2,543	1,533	60.3

연령에 관한 매칭률을 보면 70대 이상 인구에서의 매칭률이 S단계와, S+M단계와 달리 크게 나타났다. 70-79세에서는 제2차 시험조사 지역은 80.1%, 제3차 시험조사 지역은 72.7%이고, 80-89세에서는 제2차 시험조사 지역 75.6%, 제3차 시험조사 지역 63.9%, 90세이상은 제2차 시험조사 지역 82.3%, 제3차 시험조사 지역은 59.0%가 나타났다. S+M+W단계에서도 20-29세 인구의 매칭률은 낮은 비율을 나타내고 있다.



(단위 : %)

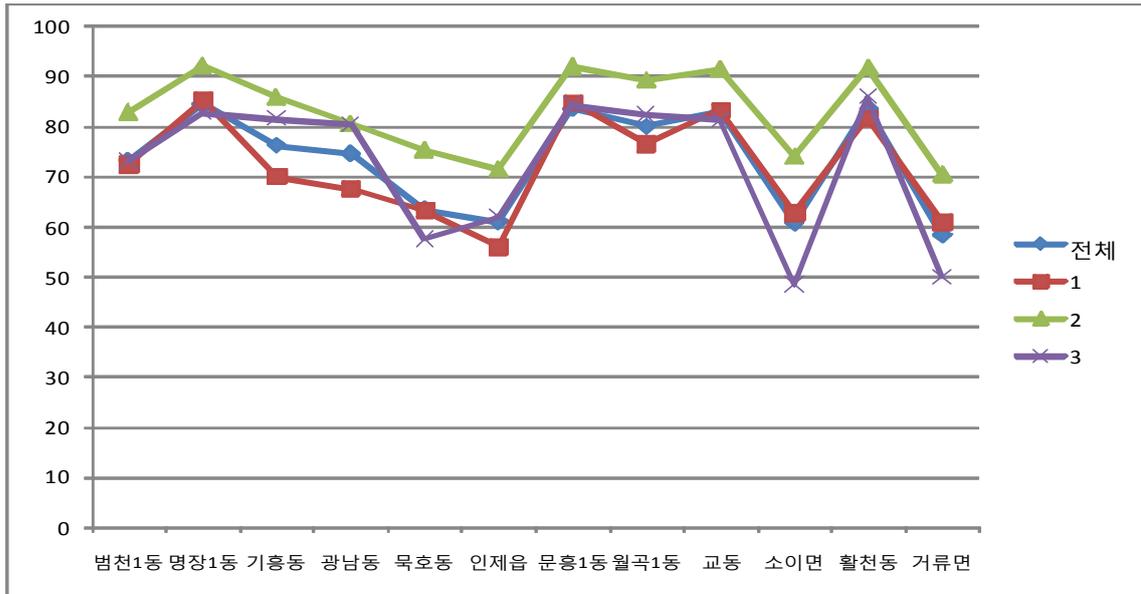
<그림 1-7> 연령별 매칭률

<표 1-39> 주민등록 자료기준 연령별 매칭건수 및 매칭률

(단위 : 명, %)

		전체	1-9세	10-19세	20-29세	30-39세	40-49세	50-59세	60-69세	70-79세	80-89세	90세 이상	
전체	전체	87,414	9,237	10,253	12,590	16,394	14,382	11,713	7,715	3,766	1,234	130	
	매칭수	65,267	6,964	8,741	7,605	12,210	10,783	8,906	6,000	3,018	933	107	
	매칭률	74.7	75.4	85.3	60.4	74.5	75	76	77.8	80.1	75.6	82.3	
범천1동	전체	9,712	488	872	1,861	1,588	1,462	1,545	1,081	638	165	12	
	매칭수	7,105	313	697	1,160	1,154	1,041	1,138	896	554	142	10	
	매칭률	73.2	64.1	79.9	62.3	72.7	71.2	73.7	82.9	86.8	86.1	83.3	
명장1동	전체	18,321	1,045	2,576	3,167	2,371	3,381	3,194	1,672	653	243	19	
	매칭수	15,481	767	2,230	2,368	1,939	2,715	2,944	1,632	639	224	23	
	매칭률	84.5	73.4	86.6	74.8	81.8	80.3	92.2	97.6	97.9	92.2	121.1	
2차	기흥동	전체	13,283	2,071	1,371	1,687	3,471	1,906	1,206	918	482	154	17
	매칭수	10,130	1,687	1,186	920	2,737	1,520	884	675	380	127	14	
	매칭률	76.3	81.5	86.5	54.5	78.9	79.7	73.3	73.5	78.8	82.5	82.4	
광남동	전체	31,772	4,272	3,852	4,044	6,805	5,410	3,819	2,297	928	307	38	
	매칭수	23,702	3,253	3,440	2,501	5,021	4,150	2,710	1,679	703	221	24	
	매칭률	74.6	76.1	89.3	61.8	73.8	76.7	71	73.1	75.8	72	63.2	
목호동	전체	4,786	297	379	561	692	607	780	813	493	151	13	
	매칭수	3,039	212	267	259	413	344	479	577	375	101	12	
	매칭률	63.5	71.4	70.4	46.2	59.7	56.7	61.4	71	76.1	66.9	92.3	
인제읍	전체	9,540	1,064	1,203	1,270	1,467	1,616	1,169	934	572	214	31	
	매칭수	5,810	732	921	397	946	1,013	751	541	367	118	24	
	매칭률	60.9	68.8	76.6	31.3	64.5	62.7	64.2	57.9	64.2	55.1	77.4	
전체	전체	91,175	8,803	16,316	12,396	14,578	18,026	10,460	5,541	3,726	1,173	156	
	매칭수	73,438	7,860	14,672	7,918	11,513	15,091	8,448	4,388	2,707	749	92	
	매칭률	80.5	89.3	89.9	63.9	79	83.7	80.8	79.2	72.7	63.9	59	
문흥1동	전체	20,958	1,866	4,213	3,256	2,898	4,038	2,573	1,200	654	222	38	
	매칭수	17,543	1,658	3,856	2,275	2,334	3,534	2,248	1,015	470	134	19	
	매칭률	83.7	88.9	91.5	69.9	80.5	87.5	87.4	84.6	71.9	60.4	50	
월곡1동	전체	16,637	1,669	3,359	2,491	2,734	3,374	1,734	676	431	143	26	
	매칭수	13,329	1,444	2,956	1,667	2,093	2,749	1,423	545	327	107	18	
	매칭률	80.1	86.5	88	66.9	76.6	81.5	82.1	80.6	75.9	74.8	69.2	
3차	교동	전체	18,433	2,196	2,992	2,147	3,357	3,450	2,155	1,160	719	216	41
	매칭수	15,310	2,017	2,740	1,235	2,840	3,012	1,786	957	546	150	27	
	매칭률	83.1	91.8	91.6	57.5	84.6	87.3	82.9	82.5	75.9	69.4	65.9	
소이면	전체	3,423	153	298	363	307	475	525	546	551	186	19	
	매칭수	2,074	118	214	105	130	274	340	385	385	110	13	
	매칭률	60.6	77.1	71.8	28.9	42.3	57.7	64.8	70.5	69.9	59.1	68.4	
활천동	전체	26,381	2,619	4,969	3,448	4,515	5,936	2,767	1,263	672	175	17	
	매칭수	22,071	2,427	4,567	2,394	3,732	5,126	2,215	991	506	103	10	
	매칭률	83.7	92.7	91.9	69.4	82.7	86.4	80.1	78.5	75.3	58.9	58.8	
거류면	전체	5,343	300	485	691	767	753	706	696	699	231	15	
	매칭수	3,111	196	339	242	384	396	436	495	473	145	5	
	매칭률	58.2	65.3	69.9	35	50.1	52.6	61.8	71.1	67.7	62.8	33.3	

세대주 관계의 매칭률은 배우자의 매칭률이 제2차 시험조사 지역에서 82.6%, 제3차 시험조사 지역에서 89.3%로 높게 나타났다. 특히 명장1동이 92.1%, 문흥1동이 91.9%, 교동이 91.3%, 활천동이 91.7%로 90%가 넘는 매칭률이 나타났다. 전체 행정동별 매칭률과 비교 해보았을 때, 묵호동의 자녀에 관한 매칭률이 낮게 나타났음을 그림을 통해서 볼 수 있다.



(단위 : %)

<그림 1-8> 행정동별 세대주 관계 매칭률

<표 1-40> 주민등록 자료기준 가구주관계 매칭건수 및 매칭률

(단위 : 명, %)

		전체	1	2	3	4	5	6	7	8	9	10	11	12	13	14
전체	전체	87,414	33,755	17,350	29,883	588	2,095	281	1,507	2	64	1,008	309	11	44	517
	매칭수	65,267	23,751	14,338	23,186	420	1,388	167	1,055	1	39	511	141	8	13	249
	매칭률	74.7	70.4	82.6	77.6	71.4	66.3	59.4	70	50	60.9	50.7	45.6	72.7	29.5	48.2
범천 1동	전체	9,712	4,604	1,579	2,758	54	200	23	169	0	7	176	27	1	3	111
	매칭수	7,105	3,331	1,308	2,015	37	143	14	93	0	3	96	14	1	0	50
	매칭률	73.2	72.4	82.8	73.1	68.5	71.5	60.9	55	0	42.9	54.5	51.9	100	0	45
명장 1동	전체	18,321	6,357	3,699	7,011	127	399	34	322	0	8	207	49	1	8	99
	매칭수	15,481	5,413	3,406	5,803	96	304	21	241	0	7	108	22	0	3	57
	매칭률	84.5	85.2	92.1	82.8	75.6	76.2	61.8	74.8	0	87.5	52.2	44.9	0	37.5	57.6
2차 기흥 1동	전체	13,283	5,320	2,625	4,362	76	373	43	158	0	23	152	68	6	6	71
	매칭수	10,130	3,726	2,250	3,554	53	258	24	107	0	11	71	33	4	3	36
	매칭률	76.3	70	85.7	81.5	69.7	69.2	55.8	67.7	0	47.8	46.7	48.5	66.7	50	50.7
광남 1동	전체	31,772	11,532	6,614	11,445	187	766	141	422	0	19	350	125	3	22	146
	매칭수	23,702	7,796	5,316	9,199	139	493	98	319	0	11	195	60	3	5	68
	매칭률	74.6	67.6	80.4	80.4	74.3	64.4	69.5	75.6	0	57.9	55.7	48	100	22.7	46.6
목호 1동	전체	4,786	2,210	907	1,213	51	102	9	189	0	4	50	13	0	5	33
	매칭수	3,039	1,396	683	698	29	65	4	121	0	4	16	5	0	2	16
	매칭률	63.5	63.2	75.3	57.5	56.9	63.7	44.4	64	0	100	32	38.5	0	40	48.5
인제 읍	전체	9,540	3,732	1,926	3,094	93	255	31	247	2	3	73	27	0	0	57
	매칭수	5,810	2,089	1,375	1,917	66	125	6	174	1	3	25	7	0	0	22
	매칭률	60.9	56	71.4	62	71	49	19.4	70.4	50	100	34.2	25.9	0	0	38.6
전체	전체	91,175	31,469	16,875	35,438	1,079	2,065	383	1,105	9	42	1,032	381	13	532	752
	매칭수	73,438	24,932	15,075	28,899	839	1,279	187	796	8	24	546	195	8	276	374
	매칭률	80.5	79.2	89.3	81.5	77.8	61.9	48.8	72	88.9	57.1	52.9	51.2	61.5	51.9	49.7
문흥 1동	전체	20,958	6,631	3,689	8,852	254	491	146	196	0	6	318	114	4	157	100
	매칭수	17,543	5,624	3,392	7,453	206	268	72	150	0	3	193	58	2	72	50
	매칭률	83.7	84.8	91.9	84.2	81.1	54.6	49.3	76.5	0	50	60.7	50.9	50	45.9	50
월곡 1동	전체	16,637	5,767	2,778	6,854	223	316	53	142	0	10	209	77	5	78	125
	매칭수	13,329	4,411	2,477	5,647	174	203	26	106	0	8	114	46	4	46	67
	매칭률	80.1	76.5	89.2	82.4	78	64.2	49.1	74.6	0	80	54.5	59.7	80	59	53.6
3차 교동	전체	18,433	6,485	3,654	7,005	210	470	62	241	5	13	110	44	1	59	74
	매칭수	15,310	5,398	3,337	5,695	181	321	29	190	5	6	55	22	1	28	42
	매칭률	83.1	83.2	91.3	81.3	86.2	68.3	46.8	78.8	100	46.2	50	50	100	47.5	56.8
소이 면	전체	3,423	1,527	689	835	59	109	5	124	0	1	28	3	0	11	32
	매칭수	2,074	959	509	405	35	62	2	74	0	1	9	2	0	4	12
	매칭률	60.6	62.8	73.9	48.5	59.3	56.9	40	59.7	0	100	32.1	66.7	0	36.4	37.5
활천 동	전체	26,381	8,670	5,112	10,453	243	556	102	259	2	10	325	127	3	205	314
	매칭수	22,071	7,085	4,689	8,981	195	358	53	206	2	4	163	61	1	121	152
	매칭률	83.7	81.7	91.7	85.9	80.2	64.4	52	79.5	100	40	50.2	48	33.3	59	48.4
거류 면	전체	5,343	2,389	953	1,439	90	123	15	143	2	2	42	16	0	22	107
	매칭수	3,111	1,455	671	718	48	67	5	70	1	2	12	6	0	5	51
	매칭률	58.2	60.9	70.4	49.9	53.3	54.5	33.3	49	50	100	28.6	37.5	0	22.7	47.7

1:가구주 2:가구주의 배우자 3:자녀 4:자녀의 배우자 5:세대주의 부모 6:배우자의 부모 7:손자녀, 그 배우자 8:중손자녀, 그 배우자 9:조부모 10:형제자매, 그 배우자 11:형제자매의 자녀, 그 배우자 12:부모의 형제자매, 그 배우자 13:기타 친인척 14:기타동거인

S, S+M, S+M+W의 단계별 결과들은 아래<표 1-41>와 같다.

단계별로 가장 큰 차이를 보이는 행정동은 활천동으로 S단계에서 매칭률이 53.2%이었으나 S+M단계에서는 29.5%P가 늘어나 82.7%가 되었으며, S+M+W단계에서 83.7%의 매칭률이 나타났다. 또, 명장1동은 S단계에서 82.8%의 매칭률이 나타났으며 S+M+W단계에서는 1.7%P가 늘어난 84.5%의 매칭률을 나타냈다.

<표 1-41> 각 단계별 매칭결과

(단위 : 명, %)

	전체	S단계			S+M단계				S+M+W단계			
2 차	전체	87,414	56,924	65.1	58,062	(1,138)	66.4	(1.3)	65,267	(8,343)	74.7	(9.6)
	범천1동	9,712	6,610	68.1	6,942	(332)	71.5	(3.4)	7,105	(495)	73.2	(5.1)
	명장1동	18,321	15,165	82.8	15,197	(32)	82.9	(.1)	15,481	(316)	84.5	(1.7)
	기흥동	13,283	9,249	69.6	9,301	(52)	70	(.4)	10,130	(881)	76.3	(6.7)
	광남동	31,772	18,784	59.1	19,290	(506)	60.7	(1.6)	23,702	(4,918)	74.6	(15.5)
	목호동	4,786	2,653	55.4	2,666	(13)	55.7	(.3)	3,039	(386)	63.5	(8.1)
	인제읍	9,540	4,463	46.8	4,666	(203)	48.9	(2.1)	5,810	(1,347)	60.9	(14.1)
3 차	전체	91,175	61,172	67.1	71,130	(9,958)	78	(10.9)	73,438	(12,266)	80.5	(13.4)
	문흥1동	20,958	16,346	78	17,256	(910)	82.3	(4.3)	17,543	(1,197)	83.7	(5.7)
	월곡1동	16,637	12,669	76.1	13,075	(406)	78.6	(2.5)	13,329	(660)	80.1	(4.)
	교동	18,433	14,682	79.7	14,934	(252)	81	(1.3)	15,310	(628)	83.1	(3.4)
	소이면	3,423	1,416	41.4	1,690	(274)	49.4	(8.)	2,074	(658)	60.6	(19.2)
	활천동	26,381	14,045	53.2	21,823	(7,778)	82.7	(29.5)	22,071	(8,026)	83.7	(30.5)
	거류면	5,343	2,014	37.7	2,352	(338)	44	(6.3)	3,111	(1,097)	58.2	(20.5)

2. 주택

가. S 단계

매칭작업은 주택에서는 인구와 달리 매칭관계의 강도에 따라서 S(Strong), M(Moderate)로 단계를 나누어 결과를 얻었다.

즉, 주택에서의 S단계는 행정동/본번표준화/부번표준화 까지는 일치하는 것을 뜻하며 S+M단계는 행정동을 제외한 나머지 매칭키가 단계적으로 제외되어가면서 수작업을 통해 매칭여부를 확인한다.

따라서 S단계는 단계적 매칭과정 중 부번표준화가 제외되기 전까지의 매칭과정을 뜻한다.

또한 행정자료 기준으로 표를 작성하였다.

<표 1-42> 자료의 단계 설명

자료	APT구분	단계	
2차	APT	S	1~7단계
		S+M	1~18단계
	매칭 APT 외	S	19~41단계
		S+M	19~58단계
3차	APT	S	1~13단계
		S+M	1~19단계
	매칭 APT 외	S	20~40단계
		S+M	20~46단계

<표 1-43>에서는 행정동별로 APT와 APT외로 나누어 매칭건수와 매칭률을 요약하였다. <그림 1-9>에서 볼 수 있듯이 12개 행정동 모두 APT가 APT외에 비해 매칭률이 높은 것을 알 수 있다.

APT와 APT외의 매칭률 형태가 비슷하나 2차 자료에서는 광남동의 경우 APT의 매칭률에 비해 APT외의 매칭률이 낮은 특성을 보이며 3차 자료에서는 소이면이 APT 매칭률에 비해 APT외의 매칭률이 낮게 나타났다.

2차 자료에서는 명장동이 APT와 APT외의 매칭률이 다른 행정동에 비해 높게 나타났으며 3차 자료에서는 거류면의 경우 APT와 APT외 매칭률이 다른 행정동에 비해 낮게 나타났다.

전체적으로 3차 자료의 매칭률이 2차 자료의 매칭률보다 높은 것으로 나타났다.



(단위 : %)

<그림 1-9> 행정동별 APT, APT 외 매칭률

*2차자료 : 범천동, 명장동, 기흥동, 광남동, 묵호동, 인제읍

3차자료 : 문흥1동, 월곡1동, 교동, 소이면, 활천동, 거류면

<표 1-43> 행정동별 APT 여부

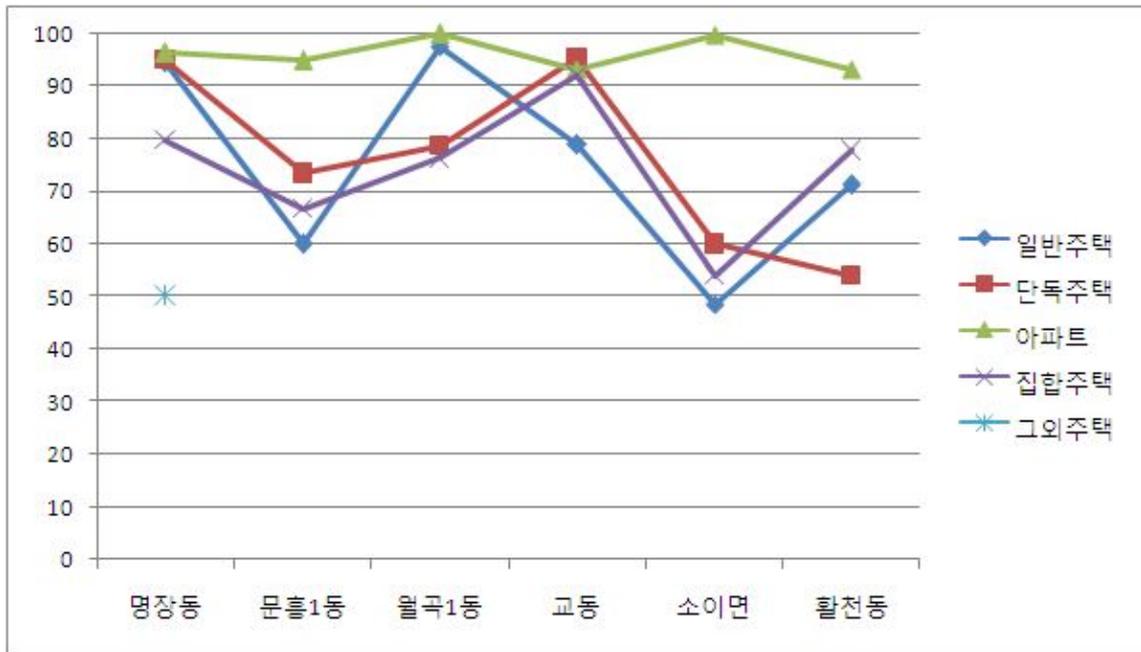
(단위 : 사례수, %)

		전체	APT	APT 외
전체	전체	25,130	11,957	13,173
	매칭수	13,141	7,207	5,934
	매칭률	52.3	60.3	45.0
부산 부산진구 범천1동	전체	3,007	1,868	1,139
	매칭수	1,841	1,261	580
	매칭률	61.2	67.5	50.9
부산 동래구 명장1동	전체	4,129	889	3,240
	매칭수	3,393	857	2,536
	매칭률	82.2	96.4	78.3
2차 경기 용인시 기흥동	전체	4,271	3,468	803
	매칭수	2,166	1,835	331
	매칭률	50.7	52.9	41.2
경기 광주시 광남동	전체	8,536	4,748	3,788
	매칭수	3,362	2,678	684
	매칭률	39.4	56.4	18.1
강원 동해시 묵호동	전체	2,473	562	1,911
	매칭수	1,191	351	840
	매칭률	48.2	62.5	44.0
강원 인제군 인제읍	전체	2,714	422	2,292
	매칭수	1,188	225	963
	매칭률	43.8	53.3	42.0
전체	전체	26,895	18,306	8,589
	매칭수	23,278	17,230	6,048
	매칭률	86.6	94.1	70.4
광주 북구 문흥1동	전체	5,446	4,882	564
	매칭수	5,042	4,633	409
	매칭률	92.6	94.9	72.5
광주 광산구 월곡1동	전체	4,043	3,058	985
	매칭수	3,898	3,056	842
	매칭률	96.4	99.9	85.5
3차 충북 제천시 교동	전체	5,935	4,396	1,539
	매칭수	5,415	4,104	1,311
	매칭률	91.2	93.4	85.2
충북 음성군 소이면	전체	1,694	456	1,238
	매칭수	1,065	455	610
	매칭률	62.9	99.8	49.3
경남 김해시 활천동	전체	7,369	5,029	2,340
	매칭수	6,468	4,685	1,783
	매칭률	87.8	93.2	76.2
경남 고성군 거류면	전체	2,408	485	1,923
	매칭수	1,390	297	1,093
	매칭률	57.7	61.2	56.8

<표 1-44>에서는 행정동별로 주택종류를 나누어 매칭건수와 매칭률을 요약하였다. <그림 1-10>은 주택종류별 매칭률 차이가 큰 행정동만 선택하였다. 그림에서 볼 수 있듯이 2차 자료에서는 명장동이 집합주택과 그 외 주택이 다른 주택에 비해 낮은 매칭률을 보였으며 3차 자료에서는 문흥1동, 소이면, 활천동의 경우 다른 주택에 비해 APT 매칭률이 높게 나타나며 월곡1동의 경우 APT와 일반주택이 다른 주택에 비해 높은 매칭률을 보인다. 교동의 경우 일반주택이 다른 주택에 비해 낮은 매칭률을 보인다.

전체적으로 APT 매칭률이 높게 나타나지만 동일 행정동에서 APT에 비해 다른 주택의 매칭률이 낮은 지역을 살펴 볼 수 있었다.

그림에서 제외된 2차 자료의 행정동은 주택종류별로 매칭률의 차이가 크지는 않지만 전체적으로 낮은 매칭률을 보인다.



(단위 : %)

<그림 1-10> 주택종류별 매칭률 차이가 나는 행정동

* 2차 자료(명장동), 3차 자료(문흥1동, 월곡1동, 교동, 소이면, 활천동)

* 일반주택(일반단독), 단독주택(다가구 단독, 영업 겸용 단독), 아파트(아파트), 집합주택(연립주택, 다세대주택, 비거주용 내 건물), 그 외 주택(기숙사 및 특수시설, 공동주택 등)

<표 1-44> 행정동별 주택종류

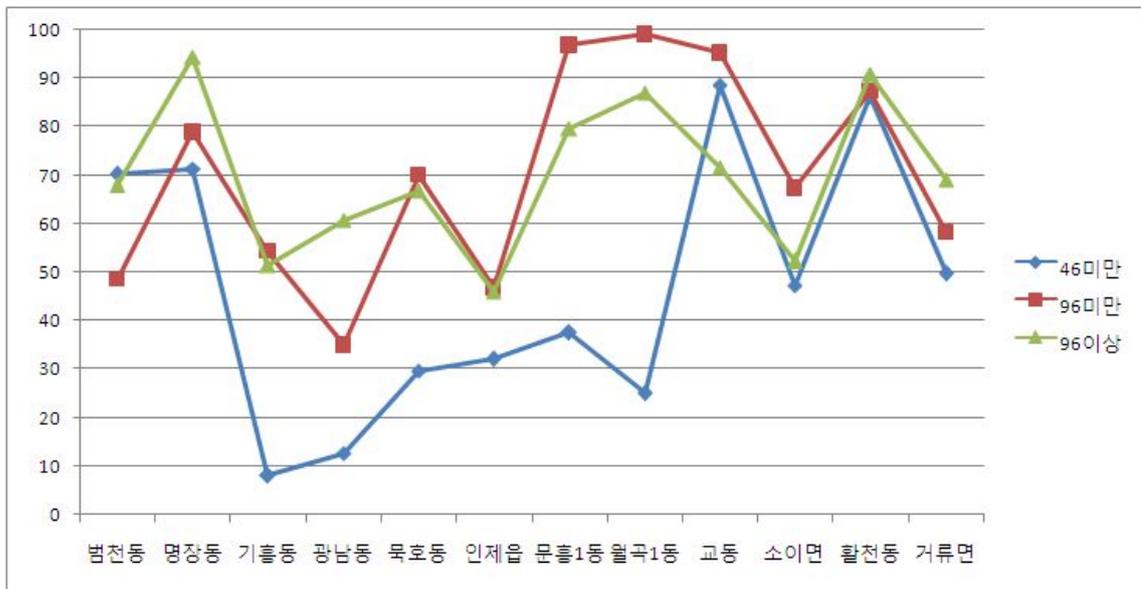
(단위 : 사례수, %)

		전체	일 반 단 독	다 가 구 단 독	영 업 겸 용 단 독	아 파 트	연 립 주 택	다 세 대 주 택	비 거 주 용 건 물	기 숙 사 및 특 수 시 설	공 동 주 택	미 상	
2차	전체	전체	25,130	6,545	200	540	11,957	778	3,636	530	2	935	7
		매칭수	13,141	3,298	101	402	7,207	187	1,122	332	0	492	0
		매칭률	52.3	50.4	50.5	74.4	60.3	24.0	30.9	62.6	0.0	52.6	0.0
	범천1동	전체	3,007	703	1	142	1,868		68	190		35	
		매칭수	1,841	314	1	98	1,261		29	106		32	
		매칭률	61.2	44.7	100.0	69.0	67.5		42.6	55.8		91.4	
	명장1동	전체	4,129	1,131	42	262	889	66	690	164		884	1
		매칭수	3,393	1,069	39	250	857	16	570	148		444	0
		매칭률	82.2	94.5	92.9	95.4	96.4	24.2	82.6	90.2		50.2	0.0
	기흥동	전체	4,271	294	15	16	3,468	128	338	10	2		
		매칭수	2,166	128	5	3	1,835	13	180	2	0		
		매칭률	50.7	43.5	33.3	18.8	52.9	10.2	53.3	20.0	0.0		
	광남동	전체	8,536	776	34	57	4,748	453	2,376	92			
		매칭수	3,362	246	6	17	2,678	119	261	35			
		매칭률	39.4	31.7	17.6	29.8	56.4	26.3	11.0	38.0			
	북호동	전체	2,473	1,758	43	33	562	34		43			
		매칭수	1,191	754	24	15	351	23		24			
		매칭률	48.2	42.9	55.8	45.5	62.5	67.6		55.8			
	인제읍	전체	2,714	1,883	65	30	422	97	164	31		16	6
		매칭수	1,188	787	26	19	225	16	82	17		16	0
		매칭률	43.8	41.8	40.0	63.3	53.3	16.5	50.0	54.8		100.0	0.0
3차	전체	전체	26,895	4,350	428	954	18,306	545	1,837	475			
		매칭수	23,278	2,725	328	718	17,230	522	1,409	346			
		매칭률	86.6	62.6	76.6	75.3	94.1	95.8	76.7	72.8			
	문흥1동	전체	5,446	30	177	342	4,882			15			
		매칭수	5,042	18	121	260	4,633			10			
		매칭률	92.6	60.0	68.4	76.0	94.9			66.7			
	월곡1동	전체	4,043	355	167	408	3,058			55			
		매칭수	3,898	347	148	305	3,056			42			
		매칭률	96.4	97.7	88.6	74.8	99.9			76.4			
	교동	전체	5,935	850	49	84	4,396	422		134			
		매칭수	5,415	671	49	78	4,104	401		112			
		매칭률	91.2	78.9	100.0	92.9	93.4	95.0		83.6			
	소이면	전체	1,694	1,124	1	24	456	8	38	43			
		매칭수	1,065	547	0	15	455	8	14	26			
		매칭률	62.9	48.7	0.0	62.5	99.8	100.0	36.8	60.5			
	활천동	전체	7,369	205	25	73	5,029	99	1,777	161			
		매칭수	6,468	146	6	47	4,685	98	1,374	112			
		매칭률	87.8	71.2	24.0	64.4	93.2	99.0	77.3	69.6			
	거류면	전체	2,408	1,786	9	23	485	16	22	67			
		매칭수	1,390	996	4	13	297	15	21	44			
		매칭률	57.7	55.8	44.4	56.5	61.2	93.8	95.5	65.7			

<표 1-45>에서는 행정동별로 면적을 나누어 매칭건수와 매칭률을 요약하였다. <그림 1-11>은 면적의 범위가 넓어 재 구분하여 행정동별로 매칭률을 살펴보았다.

그림에서 보듯이 행정동에서 면적에 따라 매칭률 차이가 나타나는데 3차 자료의 활천동의 경우 전 면적에서 고른 매칭률을 보이는 반면 2차 자료에서는 기흥동, 광남동, 묵호동이 면적에 따라 큰 차이를 보이고 3차 자료에서는 문흥1동, 월곡1동이 면적에 따라 차이를 보인다.

46^m²미만은 2차 자료의 범천동과 명장동, 3차 자료의 교동과 활천동을 제외한 행정동에서 50%미만의 낮은 매칭률을 보여 다른 면적에 비해 낮은 매칭률을 보인다. 96^m²미만은 2차 자료가 3차 자료에 비해 낮은 매칭률을 보이는데 특히 3차 자료의 문흥1동, 월곡1동, 교동은 매칭률이 95%이상으로 매우 높게 나타났다. 96^m²이상에서는 2차 자료에서 명장동이 95% 정도의 매칭률로 다른 행정동에 비해 높게 나타났으며 3차 자료에서는 활천동이 90% 정도의 매칭률로 다른 행정동에 비해 높게 나타났다.



(단위 : %)

<그림 1-11> 행정동별 면적

*2차자료 : 범천동, 명장동, 기흥동, 광남동, 묵호동, 인제읍
 3차자료 : 문흥1동, 월곡1동, 교동, 소이면, 활천동, 거류면

<표 1-45> 행정동별 면적

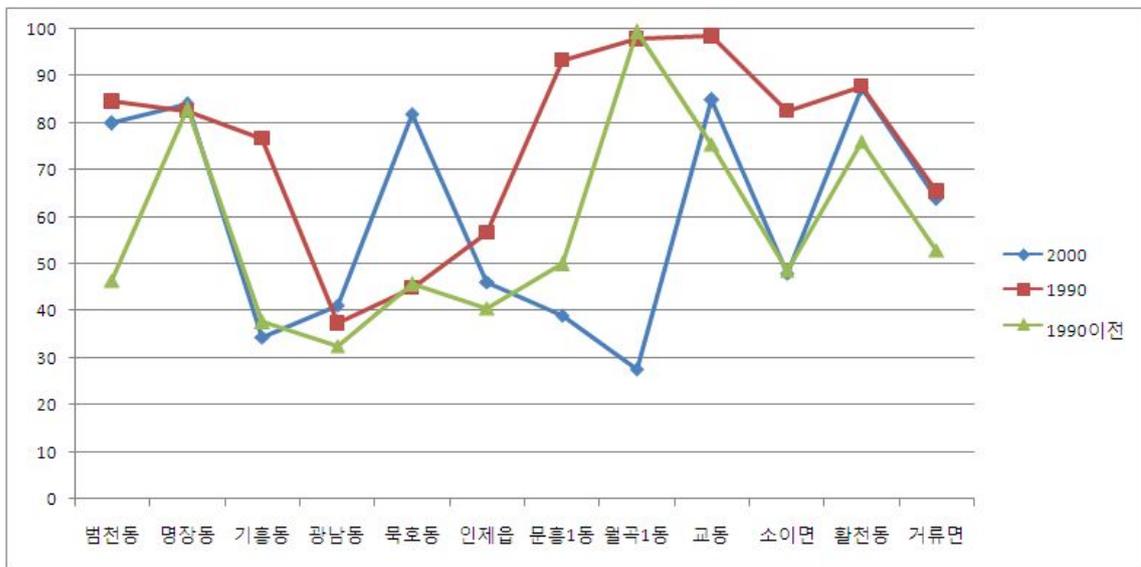
(단위 : 사례수, %)

		전체	23	30	46	63m ²	96m ²	129	162	228	327	327		
			m ²	m ²	m ²	미만	미만	m ²						
			미만	미만	미만			미만	미만	미만	미만	이상		
2차	전체	전체	25,130	655	1,317	2,657	6,448	9,660	1,968	1,370	772	196	87	
		매칭수	13,141	226	781	1,141	2,598	5,428	1,291	1,074	462	108	32	
		매칭률	52.3	34.5	59.3	42.9	40.3	56.2	65.6	78.4	59.8	55.1	36.8	
	범천1동	전체	3,007	240	794	493	250	977	125	66	42	16	4	
		매칭수	1,841	139	622	311	163	434	86	44	31	9	2	
		매칭률	61.2	57.9	78.3	63.1	65.2	44.4	68.8	66.7	73.8	56.3	50.0	
	명장1동	전체	4,129	10	23	597	912	1,386	453	462	259	25	2	
		매칭수	3,393	3	11	435	697	1,114	409	447	250	25	2	
		매칭률	82.2	30.0	47.8	72.9	76.4	80.4	90.3	96.8	96.5	100.0	100.0	
	기흥동	전체	4,271	12	10	288	1,975	1,505	93	128	150	86	24	
		매칭수	2,166	3	3	19	589	1,305	50	99	51	39	8	
		매칭률	50.7	25.0	30.0	6.6	29.8	86.7	53.8	77.3	34.0	45.3	33.3	
	광남동	전체	8,536	32	40	236	1,911	4,587	871	601	197	31	30	
		매칭수	3,362	3	5	31	478	1,793	551	419	65	10	7	
		매칭률	39.4	9.4	12.5	13.1	25.0	39.1	63.3	69.7	33.0	32.3	23.3	
	목호동	전체	2,473	297	346	687	699	308	75	24	19	10	8	
		매칭수	1,191	60	108	227	492	213	46	20	12	8	5	
		매칭률	48.2	20.2	31.2	33.0	70.4	69.2	61.3	83.3	63.2	80.0	62.5	
	인제읍	전체	2,714	64	104	356	701	897	351	89	105	28	19	
		매칭수	1,188	18	32	118	179	569	149	45	53	17	8	
		매칭률	43.8	28.1	30.8	33.1	25.5	63.4	42.5	50.6	50.5	60.7	42.1	
	3차	전체	전체	26,895	102	202	1,459	10,292	10,473	2,259	729	995	255	129
			매칭수	23,278	46	99	1,107	9,395	9,117	1,842	595	872	169	36
			매칭률	86.6	45.1	49.0	75.9	91.3	87.1	81.5	81.6	87.6	66.3	27.9
문흥1동		전체	5,446		2	6	1,524	2,592	714	94	346	101	67	
		매칭수	5,042		0	3	1,522	2,463	576	91	290	72	25	
		매칭률	92.6		0.0	50.0	99.9	95.0	80.7	96.8	83.8	71.3	37.3	
월곡1동		전체	4,043			4	1,575	1,614	124	173	447	79	27	
		매칭수	3,898			1	1,571	1,587	111	162	413	49	4	
		매칭률	96.4			25.0	99.7	98.3	89.5	93.6	92.4	62.0	14.8	
교동		전체	5,935	13	24	548	3,252	1,274	446	239	116	20	3	
		매칭수	5,415	4	12	502	3,197	1,109	308	151	112	17	3	
		매칭률	91.2	30.8	50.0	91.6	98.3	87.0	69.1	63.2	96.6	85.0	100.0	
소이면		전체	1,694	21	33	210	856	421	128	12	13			
		매칭수	1,065	11	11	103	651	209	70	7	3			
		매칭률	62.9	52.4	33.3	49.0	76.1	49.6	54.7	58.3	23.1			
활천동		전체	7,369	23	42	364	2,510	3,471	641	181	62	50	25	
		매칭수	6,468	14	29	327	2,117	3,110	628	164	47	30	2	
		매칭률	87.8	60.9	69.0	89.8	84.3	89.6	98.0	90.6	75.8	60.0	8.0	
거류면		전체	2,408	45	101	327	575	1,101	206	30	11	5	7	
		매칭수	1,390	17	47	171	337	639	149	20	7	1	2	
		매칭률	57.7	37.8	46.5	52.3	58.6	58.0	72.3	66.7	63.6	20.0	28.6	

<표 1-46>에서는 행정동별로 건축연도를 나누어 매칭건수와 매칭률을 요약하였다. <그림 1-12>은 건축연도를 2000년대, 1990년대, 1990이전 년대로 하여 행정동별로 매칭률에 대해 살펴보았다.

2차 자료에 대해 살펴보면 범천동은 1990년이전 주택의 매칭률이 낮으며 기흥동은 1990년대 매칭률만 높고 2000년대와 1990년 이전연도의 주택의 매칭률은 낮게 나타났다. 광남동의 경우 전체적으로 30~40%의 매칭률로 다른 행정동에 비해 낮은 매칭률을 보이며 목호동의 경우 2000년대 주택의 매칭률만 높은 것으로 나타났다. 명장동은 전체적으로 80%이상의 매칭률로 높게 나타났으며 인제읍은 전체적으로 40~60%의 매칭률로 광남동과 더불어 낮은 매칭률을 보인다.

3차 자료에 대해 살펴보면 문흥1동과 월곡1동의 경우 2000년대 주택의 매칭률이 낮은 특징을 보인다. 또한 문흥1동은 1990년이전 주택의 매칭률도 낮은 반면 월곡1동의 경우 1990이전 주택의 매칭률도 매우 높게 나타났다. 소이면의 경우 1990년대 주택의 매칭률을 높은 반면 다른 연도의 주택은 50%이하의 낮은 매칭률을 보인다. 3차 자료의 행정동에서는 전체적으로 1990년대 주택이 다른 연도의 주택보다 매칭률로 높은 것으로 나타났다.



(단위 : %)

<그림 1-12> 행정동별 건축연도

*2차자료 : 범천동, 명장동, 기흥동, 광남동, 목호동, 인제읍

3차자료 : 문흥1동, 월곡1동, 교동, 소이면, 활천동, 거류면

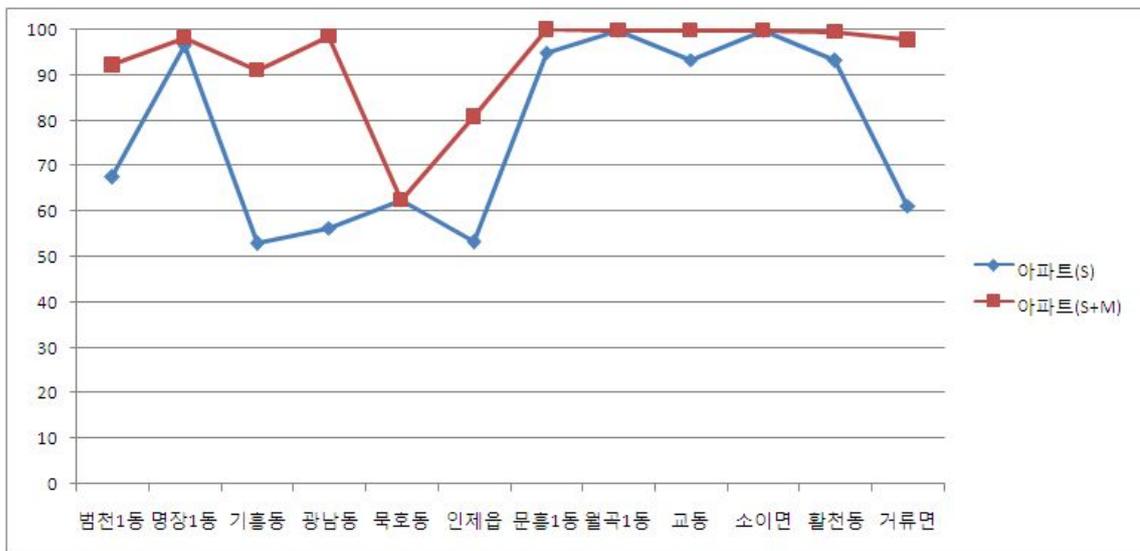
<표 1-46> 행정동별 건축연도코드

(단위 : 사례수, %)

		전체	2008	2007	2006	2000- 2005	1995- 1999	1990-1 994	1980- 1989	1970- 1979	1960- 1969	1959년 이전	결측	
2차	전체	전체	25,130	74	313	2,336	8,555	4,493	1,589	4,120	1,042	250	249	2,109
		매칭수	13,141	3	59	338	4,937	2,597	990	2,692	710	107	70	638
		매칭률	52.3	4.1	18.8	14.5	57.7	57.8	62.3	65.3	68.1	42.8	28.1	30.3
	범천1동	전체	3,007				1,340	28	56	796	301	217	137	132
		매칭수	1,841				1,071	28	43	399	150	92	32	26
		매칭률	61.2				79.9	100.0	76.8	50.1	49.8	42.4	23.4	19.7
	명장1동	전체	4,129				236	711	383	2,105	638	5	10	41
		매칭수	3,393				198	679	225	1,770	514	4	2	1
		매칭률	82.2				83.9	95.5	58.7	84.1	80.6	80.0	20.0	2.4
	기흥동	전체	4,271	2	20	1,458	933	1,102	526	102	36	1	36	55
		매칭수	2,166	0	2	0	828	861	389	39	12	0	15	20
		매칭률	50.7	0.0	10.0	0.0	88.7	78.1	74.0	38.2	33.3	0.0	41.7	36.4
	광남동	전체	8,536	67	193	437	5,149	1,993	301	101	10	11	53	221
		매칭수	3,362	1	27	6	2,372	724	132	35	4	2	16	43
		매칭률	39.4	1.5	14.0	1.4	46.1	36.3	43.9	34.7	40.0	18.2	30.2	19.5
	목호동	전체	2,473	1	7	362	96	352	78	854	31	4	1	687
		매칭수	1,191	1	4	313	64	134	59	387	18	1	1	209
		매칭률	48.2	100.0	57.1	86.5	66.7	38.1	75.6	45.3	58.1	25.0	100.0	30.4
인제읍	전체	2,714	4	93	79	801	307	245	162	26	12	12	973	
	매칭수	1,188	1	26	19	404	171	142	62	12	8	4	339	
	매칭률	43.8	25.0	28.0	24.1	50.4	55.7	58.0	38.3	46.2	66.7	33.3	34.8	
3차	전체	전체	26,895	259	441	425	2,796	6,941	11,786	1,462	402	242	1,921	220
		매칭수	23,278	14	333	223	2,522	6,191	11,100	1,302	254	140	979	220
		매칭률	86.6	5.4	75.5	52.5	90.2	89.2	94.2	89.1	63.2	57.9	51.0	100.0
	문흥1동	전체	5,446	1	3	3	62	1,452	3,911	2	4		8	
		매칭수	5,042	0	0	1	26	1,116	3,892	1	4		2	
		매칭률	92.6	0.0	0.0	33.3	41.9	76.9	99.5	50.0	100.0		25.0	
	월곡1동	전체	4,043	4	3	6	85	4	3,116	825				
		매칭수	3,898	0	1	1	25	4	3,047	820				
		매칭률	96.4	0.0	33.3	16.7	29.4	100.0	97.8	99.4				
	교동	전체	5,935	215	385	10	1,408	2,146	1,109	414	112	27	109	
		매칭수	5,415	0	310	9	1,394	2,132	1,071	338	88	18	55	
		매칭률	91.2	0.0	80.5	90.0	99.0	99.3	96.6	81.6	78.6	66.7	50.5	
	소이면	전체	1,694	10	17	11	39	115	596	33	165	97	611	
		매칭수	1,065	5	8	5	19	64	523	15	85	52	289	
		매칭률	62.9	50.0	47.1	45.5	48.7	55.7	87.8	45.5	51.5	53.6	47.3	
	활천동	전체	7,369	7	9	4	953	3,037	2,926	94	16	21	82	220
		매칭수	6,468	0	3	3	844	2,750	2,486	86	13	12	51	220
		매칭률	87.8	0.0	33.3	75.0	88.6	90.5	85.0	91.5	81.3	57.1	62.2	100.0
거류면	전체	2,408	22	24	391	249	187	128	94	105	97	1,111		
	매칭수	1,390	9	11	204	214	125	81	42	64	58	582		
	매칭률	57.7	40.9	45.8	52.2	85.9	66.8	63.3	44.7	61.0	59.8	52.4		

나. S+M 단계

<표 1-47>에서는 행정동별로 APT와 APT외로 나누어 매칭건수와 매칭률을 요약하였다. S단계와 비교하여 살펴보았으며 아래의 그림과 같이 APT에서는 S단계와 S+M단계의 매칭률 차이가 나타나는 것을 알 수 있다. 아래 <그림 1-13>에서 볼 수 있듯이 3차 자료의 APT 매칭률이 2차 자료의 매칭률 보다 높은 것을 알 수 있으며 2차 자료에서 명장동과 묵호동을 제외한 행정동에서 큰 파이를 보이고 있다. 그에 반해 3차 자료에서는 거류면을 제외한 행정동에서 비슷한 매칭률을 보인다. 이는 2차 자료에서 S단계에서 매칭되지 않은 자료가 S+M단계에서 매칭이 많이 되었다는 사실을 알 수 있으며 3차 자료의 경우 거류면이 같은 영향을 받았다고 볼 수 있다. 이는 아래의 <표 1-40>의 주택종류에서도 비슷한 경향을 나타나며 다른 주택에서는 S단계와 S+M단계의 매칭률의 큰 차이는 나지 않지만 APT의 매칭률만 S와 S+M단계의 매칭률의 차이가 나는 것으로 보인다.



(단위 : %)

<그림 1-13> 행정동별 APT, APT 외 매칭률

*2차자료 : 범천동, 명장동, 기흥동, 광남동, 묵호동, 인제읍
 3차자료 : 문흥1동, 월곡1동, 교동, 소이면, 활천동, 거류면

<표 1-47> 행정동별 APT 구분

(단위 : 사례수, %)

		전체	APT	APT 외	
2차	전체	전체	25,130	11,957	13,173
		매칭수	17,532	11,122	6,410
		매칭률	69.8	93.0	48.7
	부산 부산진구 범천1동	전체	3,007	1,868	1,139
		매칭수	2,303	1,723	580
		매칭률	76.6	92.2	50.9
	부산 동래구 명장1동	전체	4,129	889	3,240
		매칭수	3,417	873	2,544
		매칭률	82.8	98.2	78.5
	경기 용인시 기흥동	전체	4,271	3,468	803
		매칭수	3,519	3,155	364
		매칭률	82.4	91.0	45.3
	경기 광주시 광남동	전체	8,536	4,748	3,788
		매칭수	5,697	4,679	1,018
		매칭률	66.7	98.5	26.9
강원 동해시 묵호동	전체	2,473	562	1,911	
	매칭수	1,191	351	840	
	매칭률	48.2	62.5	44.0	
강원 인제군 인제읍	전체	2,714	422	2,292	
	매칭수	1,405	341	1,064	
	매칭률	51.8	80.8	46.4	
3차	전체	전체	26,895	18,306	8,589
		매칭수	24,433	18,261	6,172
		매칭률	90.8	99.8	71.9
	광주 북구 문흥1동	전체	5,446	4,882	564
		매칭수	5,290	4,881	409
		매칭률	97.1	100.0	72.5
	광주 광산구 월곡1동	전체	4,043	3,058	985
		매칭수	3,898	3,056	842
		매칭률	96.4	99.9	85.5
	충북 제천시 교동	전체	5,935	4,396	1,539
		매칭수	5,719	4,390	1,329
		매칭률	96.4	99.9	86.4
	충북 음성군 소이면	전체	1,694	456	1,238
		매칭수	1,088	455	633
		매칭률	64.2	99.8	51.1
경남 김해시 활천동	전체	7,369	5,029	2,340	
	매칭수	6,871	5,005	1,866	
	매칭률	93.2	99.5	79.7	
경남 고성군 거류면	전체	2,408	485	1,923	
	매칭수	1,567	474	1,093	
	매칭률	65.1	97.7	56.8	

<표 1-48> 행정동별 주택종류

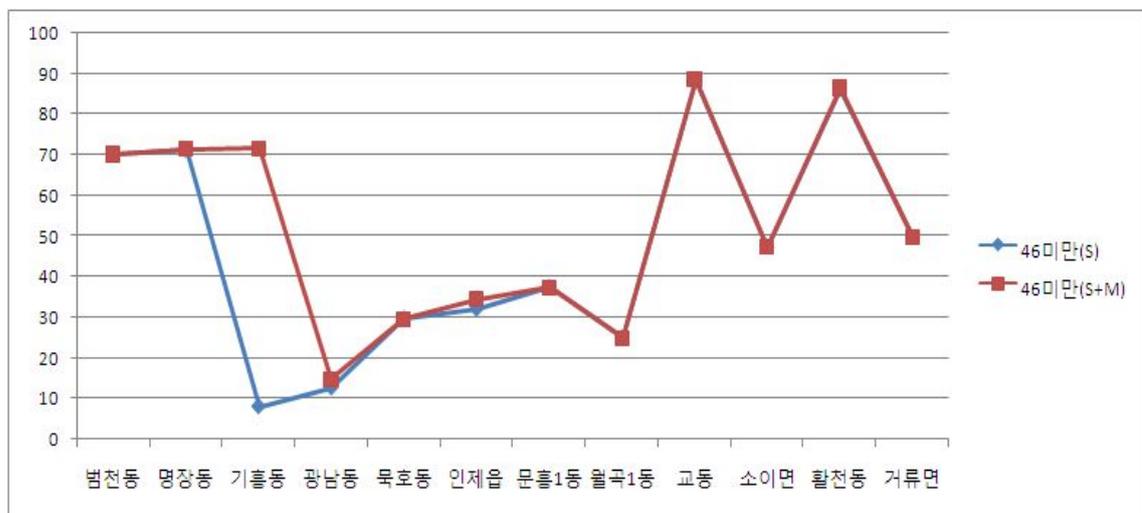
(단위 : 사례수, %)

		전체	일반단 독	다가구 단독	영업겸 용단독	아파트	연립주 택	다세대 주택	비거주 용건물	기숙사 및특수 시설	공동주 택	미상	
2차	전체	전체	25,130	6,545	200	540	11,957	778	3,636	530	2	935	7
		매칭수	17,532	3,298	101	402	11,122	270	1,509	332	0	492	6
		매칭률	69.8	50.4	50.5	74.4	93.0	34.7	41.5	62.6	0.0	52.6	85.7
	범천1동	전체	3,007	703	1	142	1,868		68	190		35	
		매칭수	2,303	314	1	98	1,723		29	106		32	
		매칭률	76.6	44.7	100.0	69.0	92.2		42.6	55.8		91.4	
	명장1동	전체	4,129	1,131	42	262	889	66	690	164		884	1
		매칭수	3,417	1,069	39	250	873	16	578	148		444	0
		매칭률	82.8	94.5	92.9	95.4	98.2	24.2	83.8	90.2		50.2	0.0
	기흥동	전체	4,271	294	15	16	3,468	128	338	10	2		
		매칭수	3,519	128	5	3	3,155	16	210	2	0		
		매칭률	82.4	43.5	33.3	18.8	91.0	12.5	62.1	20.0	0.0		
	광남동	전체	8,536	776	34	57	4,748	453	2,376	92			
		매칭수	5,697	246	6	17	4,679	131	583	35			
		매칭률	66.7	31.7	17.6	29.8	98.5	28.9	24.5	38.0			
	북호동	전체	2,473	1,758	43	33	562	34	0	43			
		매칭수	1,191	754	24	15	351	23	0	24			
		매칭률	48.2	42.9	55.8	45.5	62.5	67.6		55.8			
인제읍	전체	2,714	1,883	65	30	422	97	164	31		16	6	
	매칭수	1,405	787	26	19	341	84	109	17		16	6	
	매칭률	51.8	41.8	40.0	63.3	80.8	86.6	66.5	54.8		100.0	100.0	
3차	전체	전체	26,895	4,350	428	954	18,306	545	1,837	475			
		매칭수	24,433	2,725	328	718	18,261	540	1,515	346			
		매칭률	90.8	62.6	76.6	75.3	99.8	99.1	82.5	72.8			
	문흥1동	전체	5,446	30	177	342	4,882			15			
		매칭수	5,290	18	121	260	4,881			10			
		매칭률	97.1	60.0	68.4	76.0	100.0			66.7			
	월곡1동	전체	4,043	355	167	408	3,058			55			
		매칭수	3,898	347	148	305	3,056			42			
		매칭률	96.4	97.7	88.6	74.8	99.9			76.4			
	교동	전체	5,935	850	49	84	4,396	422		134			
		매칭수	5,719	671	49	78	4,390	419		112			
		매칭률	96.4	78.9	100.0	92.9	99.9	99.3		83.6			
	소이면	전체	1,694	1,124	1	24	456	8	38	43			
		매칭수	1,088	547	0	15	455	8	37	26			
		매칭률	64.2	48.7	0.0	62.5	99.8	100.0	97.4	60.5			
	활천동	전체	7,369	205	25	73	5,029	99	1,777	161			
		매칭수	6,871	146	6	47	5,005	98	1,457	112			
		매칭률	93.2	71.2	24.0	64.4	99.5	99.0	82.0	69.6			
거류면	전체	2,408	1,786	9	23	485	16	22	67				
	매칭수	1,567	996	4	13	474	15	21	44				
	매칭률	65.1	55.8	44.4	56.5	97.7	93.8	95.5	65.7				

<표 1-49>에서는 행정동별로 면적을 나누어 매칭건수와 매칭률을 요약하였다. 아래의 그림들은 면적의 범위가 넓어 재 구분하여 행정동별로 매칭률을 살펴보았다. <그림 1-14>는 46 m^2 미만의 S단계와 S+M단계의 매칭률 차이에 대해 살펴보았으며 <그림 1-15>는 96 m^2 미만의 S단계와 S+M단계, <그림 1-16>은 96 m^2 이상의 S단계와 S+M단계의 매칭률 차이를 행정동별로 살펴보았다.

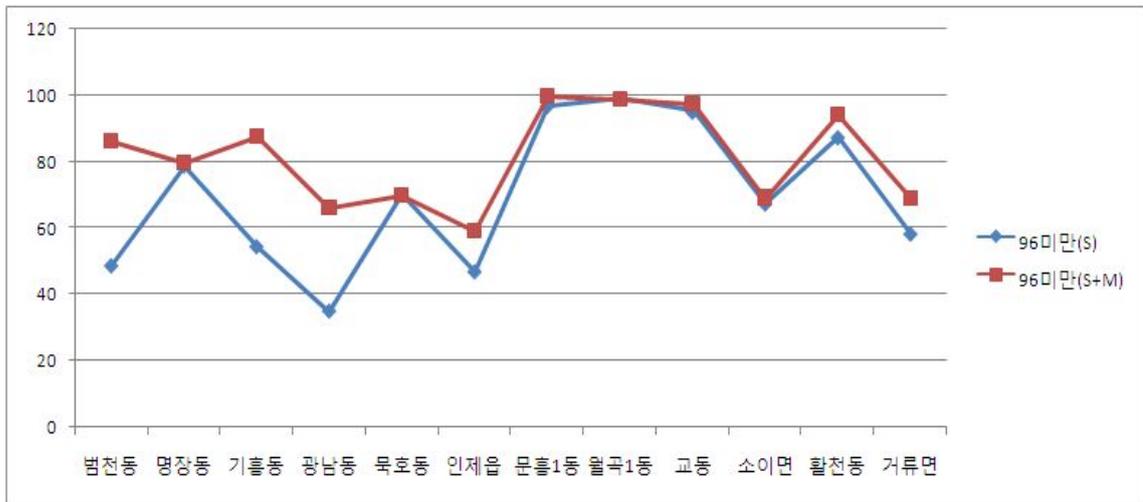
전체적으로 S단계와 비슷한 형태의 매칭률을 보이며 면적에 따라 행정동에서 S단계와 S+M단계의 매칭률의 차이가 있는 것은 아래 그림과 같이 나타났다.

46 m^2 미만에서는 2차 자료의 기흥동의 매칭률이 S+M단계에서 S단계에 비해 매칭이 많이 되었으며 96 m^2 미만에서는 2차 자료의 범천동, 기흥동, 광남동이 S단계에 비해 S+M 단계에서 매칭률이 높아짐을 알 수 있었다. 96 m^2 이상에서는 2차자료의 광남동과 3차 자료의 문흥1동과 교동이 S단계와 차이가 있음을 보인다.



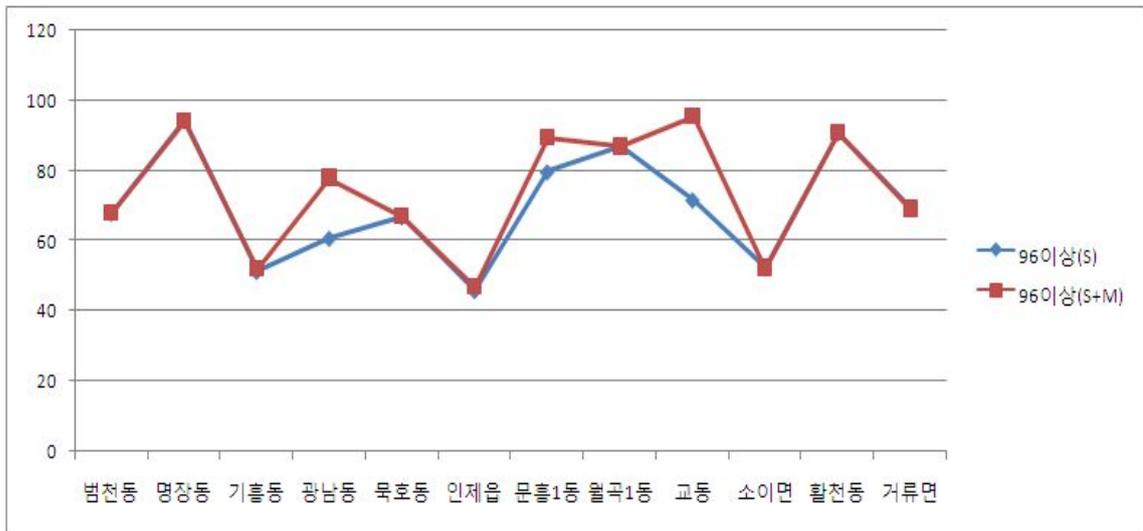
(단위 : %)

<그림 1-14> 행정동별 면적 46 m^2 미만의 매칭률 비교



(단위 : %)

<그림 1-15> 행정동별 면적 96m²미만의 매칭률 비교



(단위 : %)

<그림 1-16> 행정동별 면적 96m²이상의 매칭률 비교

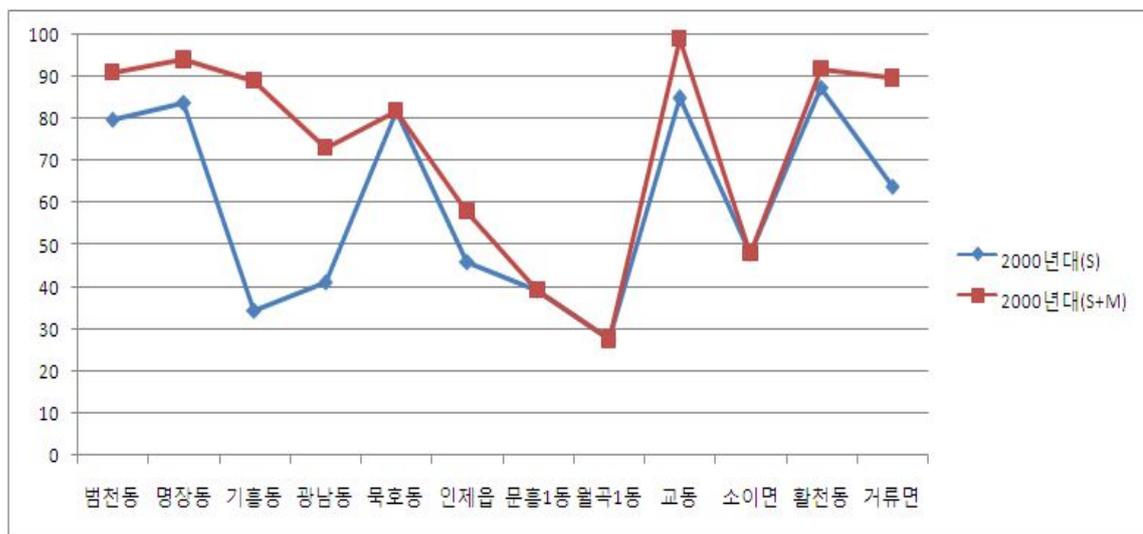
<표 1-49> 행정동별 면적

(단위 : 사례수, %)

		전체	23m ² 미만	30m ² 미만	46m ² 미만	63m ² 미만	96m ² 미만	129 m ² 미 만	162 m ² 미 만	228 m ² 미 만	327 m ² 미 만	327 m ² 이 상	
2차	전체	전체	25,130	655	1,317	2,657	6,448	9,660	1,968	1,370	772	196	87
		매청수	17,532	227	781	1,359	4,394	7,498	1,489	1,180	464	108	32
		매청률	69.8	34.7	59.3	51.1	68.1	77.6	75.7	86.1	60.1	55.1	36.8
	범천1동	전체	3,007	240	794	493	250	977	125	66	42	16	4
		매청수	2,303	139	622	311	163	896	86	44	31	9	2
		매청률	76.6	57.9	78.3	63.1	65.2	91.7	68.8	66.7	73.8	56.3	50.0
	명장1동	전체	4,129	10	23	597	912	1,386	453	462	259	25	2
		매청수	3,417	4	11	436	699	1,134	409	447	250	25	2
		매청률	82.8	40.0	47.8	73.0	76.6	81.8	90.3	96.8	96.5	100.0	100.0
	기흥동	전체	4,271	12	10	288	1,975	1,505	93	128	150	86	24
		매청수	3,519	3	3	216	1,720	1,327	50	100	53	39	8
		매청률	82.4	25.0	30.0	75.0	87.1	88.2	53.8	78.1	35.3	45.3	33.3
	광남동	전체	8,536	32	40	236	1,911	4,587	871	601	197	31	30
		매청수	5,697	3	5	38	947	3,355	743	524	65	10	7
		매청률	66.7	9.4	12.5	16.1	49.6	73.1	85.3	87.2	33.0	32.3	23.3
	북호동	전체	2,473	297	346	687	699	308	75	24	19	10	8
		매청수	1,191	60	108	227	492	213	46	20	12	8	5
		매청률	48.2	20.2	31.2	33.0	70.4	69.2	61.3	83.3	63.2	80.0	62.5
	인제읍	전체	2,714	64	104	356	701	897	351	89	105	28	19
		매청수	1,405	18	32	131	373	573	155	45	53	17	8
		매청률	51.8	28.1	30.8	36.8	53.2	63.9	44.2	50.6	50.5	60.7	42.1
3차	전체	전체	26,895	102	202	1,459	10,292	10,473	2,259	729	995	255	129
		매청수	24,433	46	99	1,108	9,770	9,572	2,088	673	872	169	36
		매청률	90.8	45.1	49.0	75.9	94.9	91.4	92.4	92.3	87.6	66.3	27.9
	문흥1동	전체	5,446		2	6	1,524	2,592	714	94	346	101	67
		매청수	5,290		0	3	1,522	2,583	704	91	290	72	25
		매청률	97.1		0.0	50.0	99.9	99.7	98.6	96.8	83.8	71.3	37.3
	월곡1동	전체	4,043			4	1,575	1,614	124	173	447	79	27
		매청수	3,898			1	1,571	1,587	111	162	413	49	4
		매청률	96.4			25.0	99.7	98.3	89.5	93.6	92.4	62.0	14.8
	교동	전체	5,935	13	24	548	3,252	1,274	446	239	116	20	3
		매청수	5,719	4	12	502	3,205	1,209	426	229	112	17	3
		매청률	96.4	30.8	50.0	91.6	98.6	94.9	95.5	95.8	96.6	85.0	100.0
	소이면	전체	1,694	21	33	210	856	421	128	12	13		
		매청수	1,088	11	11	103	667	216	70	7	3		
		매청률	64.2	52.4	33.3	49.0	77.9	51.3	54.7	58.3	23.1		
	활천동	전체	7,369	23	42	364	2,510	3,471	641	181	62	50	25
		매청수	6,871	14	29	328	2,468	3,161	628	164	47	30	2
		매청률	93.2	60.9	69.0	90.1	98.3	91.1	98.0	90.6	75.8	60.0	8.0
	거류면	전체	2,408	45	101	327	575	1,101	206	30	11	5	7
		매청수	1,567	17	47	171	337	816	149	20	7	1	2
		매청률	65.1	37.8	46.5	52.3	58.6	74.1	72.3	66.7	63.6	20.0	28.6

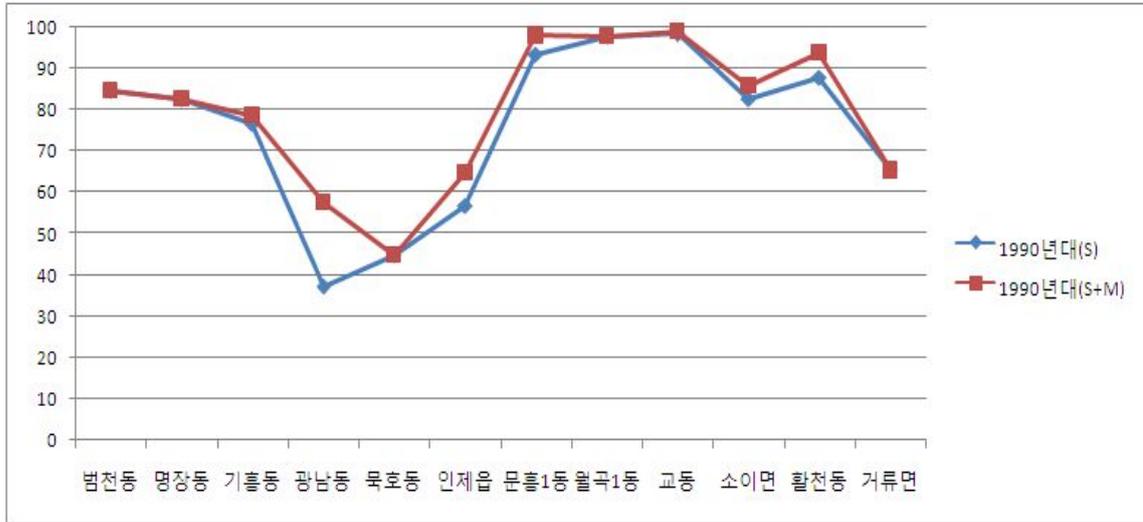
<표 1-50>에서는 행정동별로 건축연도를 나누어 매칭건수와 매칭률을 요약하였다. 아래의 그림들은 건축연도의 범위가 넓어 재 구분하여 행정동별로 매칭률을 살펴보았다. <그림 1-17>은 2000년대, <그림 1-18>은 1990년대, <그림 1-19>는 1990이전 연대에서 S와 S+M단계의 건축연도 매칭을 행정동별로 비교하여 나타냈다. 전체적으로 S단계와 비슷한 형태의 매칭률을 보이며 면적에 따라 행정동에서 S단계와 S+M단계의 매칭률의 차이가 있는 것은 아래 그림과 같이 나타났다.

2000년대에는 2차자료의 묵호동과 3차자료의 문흥1동, 월곡1동, 소이면을 제외한 나머지 행정동에서 S+M단계에서 매칭이 많이 일어 나는 것을 알 수 있다. 특히 2차 자료의 기흥동과 광남동이 S단계 이후에 매칭이 많이 되는 것을 알 수 있다. 1990년대에는 2차 자료의 광남동만 S단계 이후 매칭이 많이 되는 것을 알 수 있으며 다른 행정동은 큰 차이를 나타내지 않는다. 1990년대 이전의 경우는 2차 자료의 범천동과 인제읍이 S단계에 비해 매칭률이 높아짐을 알 수 있다.



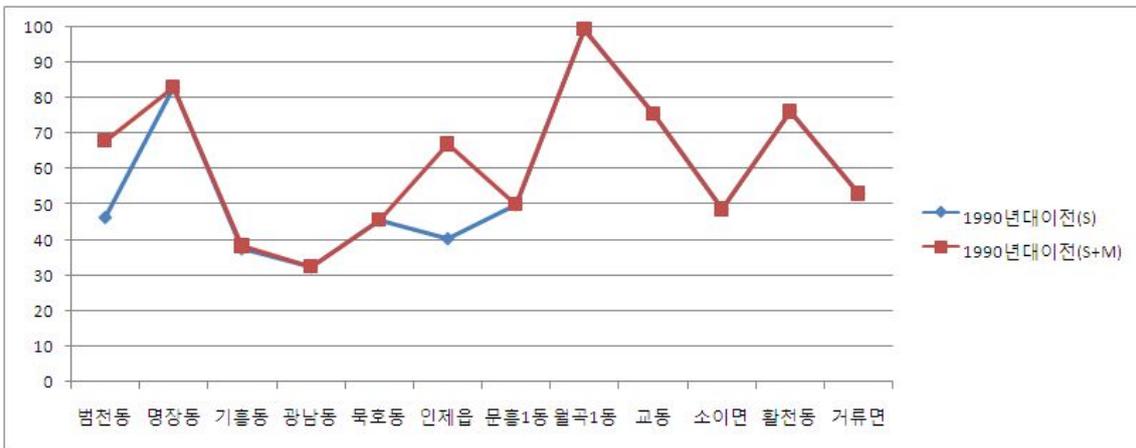
(단위 : %)

<그림 1-17> 행정동별 건축연도 2000년대 주택 매칭률 비교



(단위 : %)

<그림 1-18> 행정동별 건축연도 1990년대 주택 매칭률 비교



(단위 : %)

<그림 1-19> 행정동별 건축연도 1990이전 년대 주택 매칭률 비교

<표 1-50> 행정동별 건축연도코드

(단위 : 사례수, %)

		전체	2008	2007	2006	2000- 2005	1995- 1999	1990-1 994	1980- 1989	1970- 1979	1960- 1969	1959 년이 전	결측	
2차	전체	전체	25,130	74	313	2,336	8,555	4,493	1,589	4,120	1,042	250	249	2,109
		매칭수	17,532	3	103	1,962	6,747	3,091	1,040	3,061	710	107	70	638
		매칭률	69.8	4.1	32.9	84.0	78.9	68.8	65.4	74.3	68.1	42.8	28.1	30.3
	범천1동	전체	3,007				1,340	28	56	796	301	217	137	132
		매칭수	2,303				1,221	28	43	711	150	92	32	26
		매칭률	76.6				91.1	100.0	76.8	89.3	49.8	42.4	23.4	19.7
	명장1동	전체	4,129				236	711	383	2,105	638	5	10	41
		매칭수	3,417				222	679	225	1,770	514	4	2	1
		매칭률	82.8				94.1	95.5	58.7	84.1	80.6	80.0	20.0	2.4
	기흥동	전체	4,271	2	20	1,458	933	1,102	526	102	36	1	36	55
		매칭수	3,519	0	2	1,319	828	871	412	40	12	0	15	20
		매칭률	82.4	0.0	10.0	90.5	88.7	79.0	78.3	39.2	33.3	0.0	41.7	36.4
	광남동	전체	8,536	67	193	437	5,149	1,993	301	101	10	11	53	221
		매칭수	5,697	1	71	311	3,892	1,190	132	35	4	2	16	43
		매칭률	66.7	1.5	36.8	71.2	75.6	59.7	43.9	34.7	40.0	18.2	30.2	19.5
	목호동	전체	2,473	1	7	362	96	352	78	854	31	4	1	687
		매칭수	1,191	1	4	313	64	134	59	387	18	1	1	209
		매칭률	48.2	100.0	57.1	86.5	66.7	38.1	75.6	45.3	58.1	25.0	100.0	30.4
	인제읍	전체	2,714	4	93	79	801	307	245	162	26	12	12	973
		매칭수	1,405	1	26	19	520	189	169	118	12	8	4	339
		매칭률	51.8	25.0	28.0	24.1	64.9	61.6	69.0	72.8	46.2	66.7	33.3	34.8
3차	전체	전체	26,895	259	441	425	2,796	6,941	11,786	1,462	402	242	1,921	220
		매칭수	24,433	227	406	400	2,565	6,472	11,468	1,302	254	140	979	220
		매칭률	90.8	87.6	92.1	94.1	91.7	93.2	97.3	89.1	63.2	57.9	51.0	100.0
	문흥1동	전체	5,446	1	3	3	62	1,452	3,911	2	4		8	
		매칭수	5,290	0	0	1	26	1,364	3,892	1	4		2	
		매칭률	97.1	0.0	0.0	33.3	41.9	93.9	99.5	50.0	100.0		25.0	
	월곡1동	전체	4,043	4	3	6	85	4	3,116	825				
		매칭수	3,898	0	1	1	25	4	3,047	820				
		매칭률	96.4	0.0	33.3	16.7	29.4	100.0	97.8	99.4				
	교동	전체	5,935	215	385	10	1,408	2,146	1,109	414	112	27	109	
		매칭수	5,719	213	383	9	1,394	2,132	1,089	338	88	18	55	
		매칭률	96.4	99.1	99.5	90.0	99.0	99.3	98.2	81.6	78.6	66.7	50.5	
	소이면	전체	1,694	10	17	11	39	115	596	33	165	97	611	
		매칭수	1,088	5	8	5	19	64	546	15	85	52	289	
		매칭률	64.2	50.0	47.1	45.5	48.7	55.7	91.6	45.5	51.5	53.6	47.3	
	활천동	전체	7,369	7	9	4	953	3,037	2,926	94	16	21	82	220
		매칭수	6,871	0	3	3	887	2,783	2,813	86	13	12	51	220
		매칭률	93.2	0.0	33.3	75.0	93.1	91.6	96.1	91.5	81.3	57.1	62.2	100.0
	거류면	전체	2,408	22	24	391	249	187	128	94	105	97	1,111	
		매칭수	1,567	9	11	381	214	125	81	42	64	58	582	
		매칭률	65.1	40.9	45.8	97.4	85.9	66.8	63.3	44.7	61.0	59.8	52.4	

<표 1-51>는 주택 매칭의 S단계와 S+M단계의 매칭수와 매칭률에 대해 요약한 자료이다. 또한 증가율은 S단계에 비해 S+M단계의 매칭률이 얼마나 증가되었는지를 나타내준다.

2차 자료의 경우 S+M단계는 S단계에 비해 17.5%의 증가율을 보이며 3차 자료의 경우 4.2%의 증가율을 보여 2차 자료가 3차 자료에 비해 S+M단계에서의 매칭이 많이 이루어지는 것을 알 수 있다. 이는 2차 자료에서 S단계에서 매칭이 많이 되지 않고 수작업을 통해 매칭이 된 자료가 2차 자료에 많이 있음을 나타낸다.

<표 1-51> 각 단계별 매칭 결과

(단위 : 사례수, %)

	전체	S 단계	매칭률 (%)	S+M 단계	매칭률 (%)	증가율 (%)	
전체	25,130	13,141	52.3	17,532	69.8	17.5	
2차	범천동	3,007	1,841	61.2	2,303	76.6	15.4
	명장동	4,129	3,393	82.2	3,417	82.8	0.6
	기흥동	4,271	2,166	50.7	3,519	82.4	31.7
	광남동	8,536	3,362	39.4	5,697	66.7	27.3
	목호동	2,473	1,191	48.2	1,191	48.2	0
	인제읍	2,714	1,188	43.8	1,405	51.8	8
전체	26,895	23,278	86.6	24,433	90.8	4.2	
3차	문흥1동	5,446	5,042	92.6	5,290	97.1	4.5
	월곡1동	4,043	3,898	96.4	3,898	96.4	0
	교동	5,935	5,415	91.2	5,719	96.4	5.2
	소이면	1,694	1,065	62.9	1,088	64.2	1.3
	활천동	7,369	6,468	87.8	6,871	93.2	5.4
	거류면	2,408	1,390	57.7	1,567	65.1	7.4

제5절 결론

이 장에서는 행정자료와 시험조사 자료의 매칭을 실시하는 과정과 매칭작업에 사용된 주요전략과 작업과정에서 고려한 사항을 기술하였고, 매칭단계별, 시험조사지역별로 매칭결과를 요약하여 인구주택총조사를 수행하지 않았을 경우에 행정자료로부터 얻을 수 있는 인구 주택에 관한 정보를 어느 정도 얻을 수 있는지를 조사하였다.

또한, 향후 행정자료 활용을 위하여 조사자료의 인구에 해당하는 주민등록 자료와 조사자료의 주택에 대응되는 건축물대장 자료의 매칭을 실시하여 기존의 인구주택총조사에서 얻을 수 있는 정보에 대응되는 자료를 생성하였다.

본 연구에서 수행한 매칭과정을 요약해보면 다음과 같다.

모든 매칭결과는 모든 매칭키가 일치한 정확매칭과 일부의 키가 불일치 하지만 매칭되었다고 판단한 판단매칭으로 이루어지는데 이때 판단의 정도에 따라 각 단계를 S, M, W로 그룹화 하였다. 여기서 S는 정확매칭을 포함하여 일부항목의 불일치는 무시해도 된다고 판단하는 그룹이고 M은 일부항목의 불일치가 상당부분 오류로 해석되는 그룹이고 W는 판단자가 개별적인 아이체킹(수작업)을 수행한 어느 정도 주관에 포함된 단계를 나타낸다.

1. 주민등록자료(행정자료)와 시험조사 자료의 인구부분(인구)

인구부분의 매칭에서는 주소, 성명, 출생 년/월, 성별, 가구주와의 관계를 이용하였다.

▶ 첫 번째 단계에서는 모든 매칭키를 이용하여 정확매칭(exact matching)을 한다.

▶ 두 번째 단계에서는 매칭키 중 우선고려순위가 낮은 것부터 하나씩 제외해 나가며 판단매칭과 병행하여 실시한다.

▶ 두 번째 단계부터는 정확매칭이 아니기 때문에 중복이 발생할 수 있으므로 중복매칭에 대한 검토 후 다음 단계로 진행한다.

-성명의 경우 성, 1st 이름, 2nd 이름으로 구분하여 이용한다.

-출생 년/월에 대한 오차범위를 최대 ±3까지 허용한다.

-매칭의 하위단계에서는 주소 중 동명의 경우 끝자리 한자리만 매칭에 이용한다.

매칭작업은 2차 자료의 경우는 매칭이 25단계까지로 이루어지며, 주민등록자료 기준으로 74.7%의 매칭률을 보였으며, 시험조사 자료 기준으로 83.3%의 매칭률을 보였다. 3차 자료의 경우는 매칭이 19단계까지로 이루어지며, 주민등록자료 기준으로 80.5%의 매칭률을 보였으며, 시험조사 자료 기준으로 80.4%의 매칭률을 보였다.

2. 주민등록자료(행정자료)와 시험조사 자료의 인구부분(가구(세대))

가구(세대)의 매칭은 매칭된 인구자료를 이용하는 구성원매칭방식과 본 자료의 주소와 건물명만을 이용하여 매칭하는 주소매칭방식으로 구분하였다. 그런데 “주소매칭방식”의 경우는 주소 오류의 파악이 어렵고 행정자료로는 등록되어 있지 않지만, 조사되는 자료의 매칭에 대한 파악이 어렵고, “구성원매칭방식”의 경우는 행정자료로는 등록되어 있지만 조사되는 자료는 없거나, 행정자료로는 등록되어 있지 않지만, 자료되는 자료가 있는 경우 1세대-다가구, 다세대-1가구와 같은 복잡한 매칭결과가 도출되거나 다세대-다가구와 같은 오류가 발생할 수 있으므로 주의해야한다.

▶ 주소매칭방식은 주소와 건물명이 모두 일치하는 정확매칭(exact matching)을 한다.

결과적으로, 주소매칭방식에서 2차 자료의 경우는 세대기준 79.9%, 가구기준 72.2%로 나타났고, 3차 자료의 경우는 세대기준 79.9%, 가구기준 72.2%로 나타났다. 구성원매칭방식에서 2차 자료의 경우는 세대기준 66.9%, 가구기준 77.4%로 나타났고, 3차 자료의 경우는 세대기준 78.9%, 가구기준 76%로 나타났다.

가구(세대)매칭에서는 1세대-1가구매칭뿐 아니라 1세대-다가구, 다세대-1가구, 다세대-다가구로 분류할 수 있었다.

3. 건축물대장자료(행정자료)와 시험조사 자료의 주택부분

주택부분의 매칭에서는 주소, 건물명, 건축연도, 면적을 이용하였다.

▶ APT라고 구분되어 있는 자료를 먼저 이용하여, 첫 번째 단계에서 모든 매칭키를 이용하여 정확매칭(exact matching)을 하고, 두 번째 단계에서 우선고려순위가 낮은 것부터 하나씩 제외해 나가며 판단매칭과 병행하여 매칭한다.

▶ APT라고 구분되어 있는 자료 중 매칭되지 않는 자료와 매칭에 이용하지 않은 APT로 구분되지 않은 자료를 합하여 이용하여, 다시 모든 매칭키를 이용하여 정확매칭(exact matching)을 하고, 이어서 우선고려순위가 낮은 것부터 하나씩 제외해 나가며 판단매칭과 병행하여 매칭한다.

▶ 정확매칭 단계가 아닌 부분에서는 중복이 발생할 수 있으므로 중복매칭에 대한 검토 후 다음 단계로 진행한다.

-성명의 경우 성, 1st 이름, 2nd 이름으로 구분하여 이용한다.

-출생 년/월에 대한 오차범위를 최대 ±3까지 허용한다.

-매칭의 하위단계에서는 주소 중 동명의 경우 끝자리 한자리만 매칭에 이용한다.

결과적으로, 2차 자료의 경우는 매칭이 58단계까지로 이루어지며, 건축물대장자료 기준으로 69.8%의 매칭률을 보였으며, 시험조사 자료 기준으로 71.7%의 매칭률을 보였다. 3차 자료의 경우는 매칭이 46단계까지로 이루어지며, 건축물대장자료 기준으로 90.8%의 매칭률을 보였으며, 시험조사 자료 기준으로 90.6%의 매칭률을 보였다.

4. 주민등록자료(행정자료)와 건축물대장자료(행정자료)

주민등록자료와 건축물대장자료의 행정자료 간 매칭에서는 주소와 건물명이 이용되었다.

▶ 주소매칭방식은 주소와 건물명이 모두 일치하는 정확매칭(exact matching)을 한다.

결과적으로, 2차 자료에서 세대는 73.5%, 주택은 76%가 매칭되었으며, 3차 자료에서 세대는 82.8%, 주택은 78.7%가 매칭되었다.

제6절 정책적 건의사항

본 장에서 시도한 매칭작업은 제2차, 제3차 시험조사 자료와 이 조사지역에 해당하는 행정자료에 의하여 이루어졌다. 이 작업을 수행하면서 얻은 연구결과는 크게 매칭방법 연구와 매칭률 두 가지다. 매칭방법은 효율적, 표준적 방법의 분석과 제안이 아니라 시험 조사자료가 실제 존재하는 개체(인구, 주택)를 모두 조사한 것으로 가정하고 행정자료가 최대 얼마나 이들을 포함하고 있는지를 알 수 있도록 하는 방법을 고안하였다. 따라서 본 연구에서 제시한 방법론은 표준화 정도나 효율성 정도의 관점에서는 평가될 수 없는 것이 약점이라고 할 수 있다. 그러나 본 매칭 연구의 목적이 향후 새주소 체계가 완성되고 건축물 대장 DB가 구축되어 행정자료가 현재보다 더 정확도가 높아졌을 경우에 예상되는 매칭률에 대한 하나의 기준정보를 얻는 것이라는 면에서 본 연구에서 사용한 방법이 비효율적이고, 상당량의 수작업과 어느 정도의 주관적 판단 등을 포함하는 가능한 모든 경우를 검토하는 방법이라는 장점이 있다. 이와 같이 본 연구의 목적과 결과를 정리한 본 연구팀은 이제 다음과 같은 정책적 과제를 제안한다.

첫째는 본 연구를 통해서 습득된 다양한 매칭경험에 대한 세밀한 검토를 통하여 효율성제고 방안을 모색 할 필요가 있다. 특히 이 모색작업은 현시점의 행정자료를 기준으로 하는 것과 2015년에 사용될 새주소 체계와 완성된 건축물 대장DB를 전제하면서 하는 두 가지 작업이 가능하다고 본다. 특별히 후자의 경우를 연구할 때는 표준화의 의지를 갖고 수행되어서 표준화된 매칭 알고리즘의 개발을 목표로 하여야 한다. 한편 본 연구에서는 사용하지 못한 GIS와 매칭방안을 반드시 검토하여야 한다고 본다. 여기에는 다시 주소 DB, 주소체계의 문제가 제기되는데 비록 이들의 정확성이나, 구축수준이 현재에는 사용할 만큼 안정적이지 않다고 하더라도 향후 안정된 상황을 대비

하여 시험적인 시도를 하여야 한다. 끝으로 개인정보를 포함하고 있는 중요한 자료를 다루는 연구 작업이기 때문에 본 연구과정에서는 동반되었다. 그러나 본 연구팀의 입장에서는 연구의 효율성을 지나치게 제한할 만큼 너무 과도한 자료 관리 시스템이 작동된 점도 있었다고 본다. 이 역시 등록센서스를 수행하기 위하여 고민하여야 할 주제로서 본 연구의 중요한 결과물이었다고 본다. 따라서 매칭작업 과정에서 개인정보의 완벽한 보호와 원활한 연구수행이라는 두개의 목표를 원만히 달성할 수 있는 자료의 관리시스템에 대한 연구가 필요하다고 본다.

<참고> 싱가포르와 이스라엘의 자료매칭

** 우리나라의 2015년 등록센서스 추진을 위하여, 현재 이용하고 있는 주민등록자료, 건축물/주택대장과 제2차 및 제3차 시험조사의 결과를 토대로 매칭률의 평가와 자료매칭의 최적방안을 모색하였다.

** 우리나라에서 2015년 등록센서스 실시가능성에 대한 잠정적 판단을 포함하여 결론을 내리고, 매칭률을 개선하기 위한 technical 수준과 legal and institutional한 수준의 대응방안이 제1장의 본문에는 포함되어 있지는 않지만, 최근 등록센서스를 실시하였거나 실시계획을 가지고 있는 외국(싱가포르, 이스라엘, 독일, 대만-자세한 내용은 부록을 참고할 것)에 대한 매칭이나 포괄범위 사례를 통하여, 2015년 등록센서스 추진의 효율성을 극대화할 수 있을 것으로 판단하고 있다.

① 싱가포르의 행정자료(가구등록데이터베이스)의 자료 완전성 진단결과

싱가포르는 인구총조사와 주택총조사를 동시에 실시하여, 이러한 총조사를 등록센서스로 실시하기 위한 주축 행정자료는 가구등록데이터베이스와 전국주택데이터베이스(NDD, National Database on Dwelling)이다. 이것은 우리의 인구주택총조사의 주축 행정자료가 주민등록과 건축물/주택대장인 것과 마찬가지로이다.

가구등록데이터베이스(HRD, Household Registration Database)의 주요 조

조사항목(한글)	조사항목(영문)	1995년 12월	2000년 6월
고유식별번호	Unique ID Number(UIN)	100.0%	100.0%
거주상태	Residential status	100.0%	100.0%
성명	Name	100.0%	100.0%
성별	Sex	100.0%	100.0%
출생연월일	Date of birth	100.0%	100.0%
혼인상태	Marital status	92.7%	99.5%
인종/종족집단	Ethnic group	99.9%	100.0%
방언집단	Dialect group	99.9%	100.0%
출생국	Country of birth	100.0%	100.0%
시민권 여부	Citizen group	100.0%	100.0%
주소	Address	92.9%	99.5%

사항목의 품질은 핵심적 조치를 실시한 후에, 완전성의 면에서 대단히 우수

한 것임이 판명이 되었다. 대부분의 조사항목은 그 품질평가에서 집계범위가 100%에 이르렀다. 데이터베이스에 조사항목별로 개별기록에서 결측값(missing value)은 존재하지 않았다.

1995년 12월과 2000년 6월의 행정자료 품질진단에서 혼인상태나 주소에 대한 조사항목은 집계범위가 100%에 이르지 않았다. 그러나 이들 조사항목은 1995년 가구등록데이터베이스가 도입된 후, 모든 거주자에 대해서 완전성이 93%에서 99.5%로 개선되었다. 주소지의 집계범위가 100%에 이르지 않는 것은 등록시점에 정확한 거주지를 기재할 수 없는 새로운 영주권자들 때문에 그런 것으로 확인되었다. 혼인상태에 대해서, 집계범위가 100%에 이르지 않는 이유는 해외결혼과 같은 결혼건수는 관련기관에 등록되지 않은 사실에 기인하는 것으로 밝혀졌다.

이상의 사실을 바탕으로 하여, 싱가포르 통계국(Singapore Department of Statistics)은 행정자료를 이용하여 정확한 인구규모와 인구관련 기본특성에 대한 정보를 제공할 수 있다고 확신하면서 등록센서스를 실시하기로 결정하였다. 싱가포르에서 법률적 환경이나 자료의 비밀준수 조항은 관련 행정정보의 공유를 허용한다. 더욱이, 이러한 접근방법의 채택으로 발생하는 비용절감은 상당한 수준에 이르는 것으로 파악되고 있다.

싱가포르 통계국은 2000년대 이후에는 가구등록데이터베이스와 전국주택데이터베이스의 기록들을 추가적으로 통합하여 인구의 연속측정(continuous measurement) 시스템을 사정권으로 하고 있다. 규칙적으로 실시되는 소규모 통계조사의 시스템을 확립하여, 행정자료에서 얻을 수 없는 자료를 획득하고 어느 시점에서 관심대상이 되는 인구현상과 사회적 동향을 모니터링 하는 것이 싱가포르 통계국의 주된 관심이다.

참고문헌:

Singapore Department of Statistics (2003), Administrative Report, Census of Population and Housing, Singapore Department of Statistics, Ministry of Trade and Industry.

② 이스라엘의 시험조사 결과

이스라엘은 2008년 12월에 등록센서스의 형태로 센서스를 실시하여, 최근에 센서스 표본조사를 종료하였다고 발표하였다. 이스라엘 통계청은 1948년 인구대장을 수립하고, 초대 국회의원선거를 위한 신분증을 발급하기 위하여,

인구센서스를 실시하였으나, 예산이 제때에 지급되지 않는 인구센서스를 전통적 방식에서 등록센서스와 표본조사를 결합하는 방식으로 실시하는 이른바 “센서스의 대개혁”을 추진하였다.

이스라엘의 등록센서스 시험조사는 2002년, 2004년, 2006년의 3차례에 걸쳐서 실시되었는데, 제1장의 자료매칭과 관련된 추후의 연구과제에서, 최근의 2006년 시험조사, 2008년 표본조사를 실시하기 전의 “시범조사”(dress rehearsal) 결과를 토대로 2015년 등록센서스 추진의 타당성에 대한 설명을 추가하여야 할 것이다. 최근의 이스라엘 센서스 시험조사의 목적은 센서스 방법론, 센서스에 사용될 기술적/통계적 방법론을 점검하는 과정, 센서스 통계의 공표방식, 조사원 훈련방식 및 보수지급체계를 결정하기 위한 것이었다. 조사대상에 포함된 객체는 이스라엘 보완인구대장(Improved Population Register)에 포함된 25개 지역의 4만 명의 인구, 1만 4,000 가구이며, 여기에는 16개 거주시설(residential institutions) (1개의 교도소를 포함됨) 도 포함되어 있다. 2006년 시험조사 결과는 다음과 같다.

- 행정가족(우리나라의 주민등록세대 개념과 유사함)
90% 가량이 정확하게 구성되고 있으며, 0.5%가 부당한 방법으로 가구가 분할되었으며, 1.4%는 같은 주소지에 거주하고 있음에도 불구하고 실제 가족과 결합되지 않는 상태, 3%는 잘못된 방법으로 하나의 가족인 것처럼 통합되어 있음. 4.5%는 주소지의 오류
- 출국자
0.04%(19명)가 출국자인데, 그렇게 표시되지 않으며, 0.02%(9명)가 출국자가 아닌데, 출국자로 표시되어 있음
- 과소범위(undercoverage parameter) 평가
약 4만 명이 표본조사 주소지에 등록되어 있었고, 2.72만 명(68%)이 과소범위 추정 표본조사에서 조사객체에 포함되었으며. 4, 300명(11%)이 등록된 주소에 거주하였지만, 과소범위 표본조사에서 누락되었고. 5,700명(14%)이 상이한 주소에 거주하고 있었다. 한편 7%는 접촉이 불가능하였다. 89%가 현지실사 절차를 위한 집계범위의 최대치라고 볼 수 있다.
- 과다범위(overcoverage parameter) 평가
행정가구는 1만 1360가족으로서, 1,413가구가 전화번호가 불확실하고,

22,809명의 인구 중에서 2,082명이 전화번호가 불확실하고, 13,754 명이 완전한 전화면접(72%), 1,524명(7%)이 부문 면접 또는 잘못된 매칭, 4,326명(21%)이 전화조사가 불가능하였다.

이스라엘 통계청(Central Bureau of Statistics)은 2009년 7월 2일자로, 웹사이트에 2008년 인구센서스 표본조사 전화확인조사를 통한 자료수집이 종결되었다고 발표했다. 해당 표본조사는 두 단계로 진행되었다. 제1단계는, 현장실지조사로서, 1:1 면접조사가 실시되었다. 이 단계는 2008년 12월 28일부터 3월 8일까지 100일 동안 계속되었다. 이 기간 중, 대략 2,000명의 현장조사원이 동원되어 33만 가구의 130만 명에 해당하는 인구를 대상으로 현장실지조사를 실시하였다.

제2단계는 필요정보를 확인하기 위한 조사로 전화면접으로 실시되었다. 이 단계에서, 이스라엘 통계청은 센서스 표본조사의 표집대상으로 제1단계에서 면접불응으로 현장 실지조사 대비 인구대장 과다등록으로 판단한 10만 가구에 대하여 전화확인조사를 실시하였다. 전화확인조사는 2009년 3월 25일에 실시하여, 2009년 7월 2일에 종료하였다.

이스라엘 통계청은 센서스 표본조사에서 획득한 정보를 이용하여, 이스라엘의 인구와 그 인구의 인구학적, 경제적, 사회적 특성에 대하여 신뢰도가 높은 통계를 작성하기 위하여 노력하고 있다. 최종적으로, 현장 조사원들에 협력하여 조사가 성공적으로 끝나도록 도와준 이스라엘 국민들에게 감사를 드린다는 이야기를 하고 있다.

③ 독일의 등록센서스 시험조사

- 목적: 인구대장에 이중등록자(double entries)의 수를 추정하고, 매칭 방법론을 개선하기 위한 것
- 표본추출방법: 1월 1일, 5월 15일, 9월 15일 또는 생년월일이 불확실한 날짜에 출생한 모든 사람을 표본대상으로 하였음
- 절차: 모든 기초지방자치단체로부터 표본으로 추출된 모든 사람들로부터 인구대장(population register) 기록을 요구하고, 중복집계(duplicates)를 확인하기 위하여 6개의 상이한 매칭기법(matching techniques)을 구사하여, 중복기재를 점검하였으며, 전화조사나 현장실사에 의하여 불분명한 경우를 확인하고자 노력하였다.

자치단체 인구 규모	등록인구*	중복기재					
		순과다집계**		이중기재 점검을 통해서 삭제		과다집계 제거 불 능	
		1,000명	%	1,000명	%	1,000명	%
10,000명 미만	23,071.0	459.5	2.0	149.9	0.7	309.6	1.4
10,000-50,000명	26,928.1	643.4	2.5	153.3	0.6	490.1	1.9
50,000 - 80,000명	24,839.1	801.6	3.4	139.3	0.6	662.3	2.8
80,000명 이상	7,342.0	416.3	6.0	43.0	0.6	373.3	5.4
합계	82,180.0	2 320.8	2.9	485.5	0.6	1 835.3	2.3

* 시설가구 거주자를 포함 **일시적 과다집계는 제외함

- 2001년 시험조사 결과에 대한 결론: 인구센서스의 결과, 주민 수는 기초지자체에 자금을 분배하는데 기초가 되기 때문에, 고도의 정확성이 확보된 센서스 결과가 필요하다.

시험조사 결과는 수정된 인구대장에서 도출되는 주민 수는 등록센서스의 실시예 필요한 정확성의 기준을 충족시키지 못한다는 것을 보여주었다. 과다집계율(overcount rate)와 과소집계율(undercount rate)을 소도시와 대도시 사이에 큰 차이가 있다. 따라서 독일 등록센서스는 표본조사를 통하여 과다집계와 과소집계를 점검하는 것(95%의 확률로 1% 이내의 정확성을 갖는 추정인구를 제시하는 것이 최종목적이 되어야 한다.

제2장 매칭자료의 불일치 항목 분석 및 비(非)

매칭자료 항목 분석

제1절 머리말

제1장에서 수행한 매칭작업은 매칭관계의 강도에 따라서 인구는 S(Strong), M(Moderate), W(Weak)로 단계를 나누고, 주택은 S(Strong), M(Moderate)으로 단계를 나누어 결과를 얻었다. 이로부터 매칭된 세대원과 주택의 특성들을 다각적으로 분석하여 행정 자료에 있는 인구나 주택에 관한 정보 중 조사를 통하여 파악되는 개체(세대원, 주택)의 특성을 추출해 보았다.

반면에 제2장에서는 제1장에서 매칭되었다고 판단한 개체들이 어떤 항목에서 얼마나 상이한 값을 갖고 있는지를 특성별로 분석하여 어떤 집단에서 어떤 항목의 불일치가 어떤 형태로 나타나고 있는지를 밝혔다. 또한 매칭되지 않은 개체들에 대해 항목별 분석을 통해 매칭되지 않은 개체들의 특성을 밝혔다.

시험조사 자료에 포함된 개체들이 실제로 존재한다는 가정을 해본다면, 이 장에서의 분석내용은 행정정보의 품질에 관한 특성별 입장을 수립하여 차후보다 효과적으로 행정자료를 활용할 수 있도록 하고자 하는데 목적이 있다고 할 수 있다.

제2절 분석범위와 방법론

분석의 기본 방향은 매칭되었다고 판단한 개체들 간의 정보 중에서 불일치된 것을 단계별로 분석 하고, 또한 매칭되지 않은 개체들의 정보를 파악한다. 따라서 제3절과 제4절에서는 매칭되었다고 강하게 생각되는 S단계와 매칭정도가 중간정도인 M단계는 묶어서 두 개체의 매칭이 맞다는 전제하에 자료에 포함되어 있는 항목오류를 파악하는 입장에서 수행되었다고 볼 수 있다. 또한 매칭되었다고 생각되는 정도가 낮은 W단계에서 불일치 항목분석은 두 개체를 매칭되었다고 할 수 있는가를 검토할 목적으로 수행되었다.

5절에서는 매칭되지 않은 개체들의 정보를 탐색하여, 조사자료에는 존재(실존)하지만 등록되지 않은 개체의 특성과 행정자료에는 등록되어있지만 조사자료에는 없는(실존하지 않는) 개체의 특성을 파악하도록 하였다.

구체적인 방법을 기술한다면 다음과 같다. 제3절과 제4절에서는 매칭된 개체들의 매칭항목별로 행정자료와 시험조사 자료의 불일치 패턴을 파악하고자 인구부분과 주택부분으로 나누어서 시험조사 자료의 차수별, 행정동별로 교차분석을 실시한다. 여기서 인구부분은 시험조사 자료의 인구부분과 주민등록자료를 매칭한 것이고 주택부분은 시험조사 자료의 주택부분과 건축물대장자료를 매칭한 자료로 인구부분에서는 성별/연령/세대주와의 관계/이름의 항목 불일치에 대해 살펴보고, 주택부분에서는 주택종류/면적/건축연도의 항목 불일치에 대해 살펴본다.

제5절에서는 행정자료와 시험조사 자료로 구분하여 차수별, 각 항목별로 빈도분석을 실시한다. 단, 제3절의 불일치 항목 분석에서 결측(missing)이 있는 항목의 매칭에 대해서는 불일치라고 정의하지 않았으므로 자료의 개수가 동일하지 않을 수 있다. 그리고 제3절과 제4절까지의 불일치 사례에 해당하는 표는 모두 행정자료 기준으로 작성하였다.

제3절 인구 매칭 자료의 불일치 항목 분석

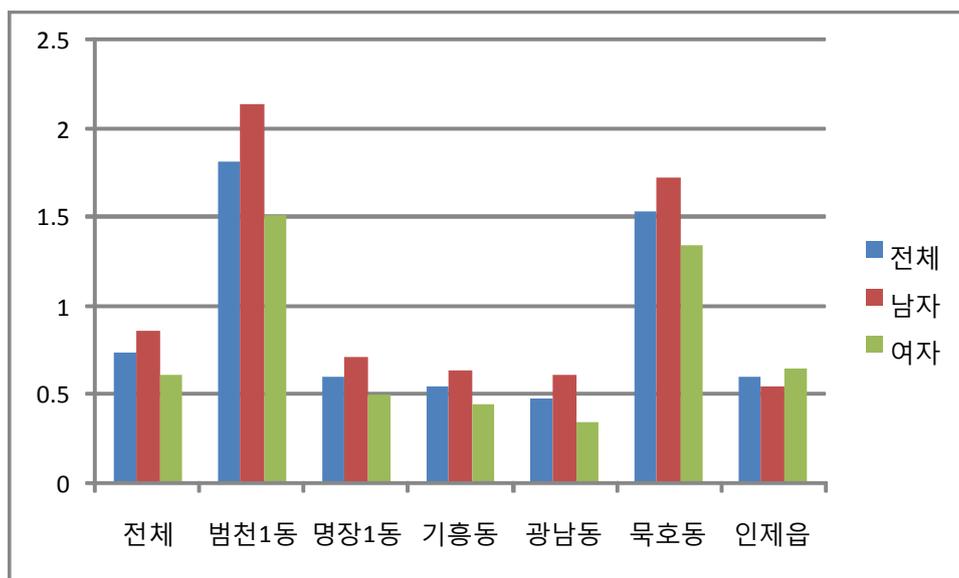
1. S+M단계

제1장에서 제2차 시험조사 자료의 매칭은 총 25단계까지 수행된 매칭 작업 중 21단계까지를 S+M단계라고 정의하였고, 제3차 시험조사 자료의 매칭은 총 19단계까지 수행된 매칭 작업 중 18단계까지를 S+M단계라고 정의하였다. S+M단계까지만 매칭을 인정한다면 제2차 시험조사 자료는 주민등록 자료 기준으로 총 87,414건 66.4%(58,062건)가 매칭되었고, 제3차 시험조사 자료는 주민등록 자료 기준으로 S+M 단계에서는 주민등록 자료 기준으로 총 91,175건 78.0%(71,130건)이 매칭되었다.

가. 성별

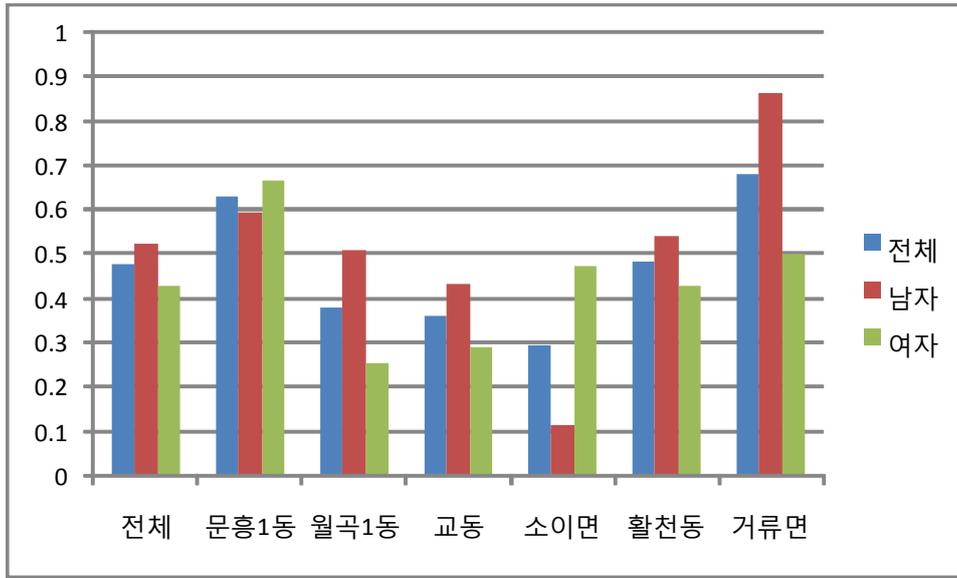
제2차 시험조사 자료의 매칭에서 성별의 불일치는 0.7%(431건)로 나타나고 있다. 제3차 시험조사 자료의 매칭에서 성별의 불일치는 0.5%(340건)로 2차에 비하여 다소 높게 나타나고 있다.

행정동별로는 제2차 자료의 (부산)범천1동에서 1.8%(126건), (강원)목호동에서 1.5%(41건)로 다른 행정동에 비해 높게 나타나고 있다.



(단위 : %)

<표 2-1> 2차 자료의 행정동별 성별 불일치 비율



(단위 : %)

<그림 2-2> 3차 자료의 행정동별 성별 불일치 비율

<표 2-1> 2차 자료의 행정동별 성별 불일치 사례

		전체			남자			여자		
		전체	불일치	불일치율	전체	불일치	불일치율	전체	불일치	불일치율
2차 시험 조사 자료	전체	58,062	431	0.7	28,675	249	0.9	29,387	182	0.6
	범천1동	6,942	126	1.8	3,309	71	2.1	3,633	55	1.5
	명장1동	15,197	92	0.6	7,406	53	0.7	7,791	39	0.5
	기흥동	9,301	51	0.5	4,661	30	0.6	4,640	21	0.5
	광남동	19,290	93	0.5	9,623	59	0.6	9,667	34	0.4
	목호동	2,666	41	1.5	1,333	23	1.7	1,333	18	1.4
인제읍	4,666	28	0.6	2,343	13	0.6	2,323	15	0.6	
3차 시험 조사 자료	전체	71,130	340	0.5	35,174	185	0.5	35,956	155	0.4
	문흥1동	17,256	109	0.6	8,409	50	0.6	8,847	59	0.7
	월곡1동	13,075	50	0.4	6,467	33	0.5	6,608	17	0.3
	교동	14,934	54	0.4	7,386	32	0.4	7,548	22	0.3
	소이면	1,690	5	0.3	849	1	0.1	841	4	0.5
	활천동	21,823	106	0.5	10,908	59	0.5	10,915	47	0.4
	거류면	2,352	16	0.7	1,155	10	0.9	1,197	6	0.5

나. 생년/월

생년/월 항목 불일치는 제2차 시험조사 자료의 매칭에서 17.3%(10,027건)으로 나타나고, 제3차 시험조사 자료의 매칭에서는 20.4%(14,532건)였다. 전체적으로 연령대가 높아질수록 생년/월 불일치율이 높아지는 패턴을 보이고 있다. 행정동별로는 2차 자료의 (경기)기흥동은 12.6%(1,174건), (경기)광남동은 12.4%(2,395건)로 생년월 개월 수 불일치율이 낮고, 3차 자료의 (충북)소이면은 39.4%(666건), (경남)거류면이 35.1%(825건)로 생년/월 개월 수 불일치율이 높았다.

<표 2-2> 행정동별 생년/월 차이 불일치 패턴

		전체	0~ 9	10~ 19	20~ 29	30~ 39	40~ 49	50~ 59	60~ 69	70~ 79	89~ 89	90~ 90~	
전체	사례수	10,027	257	681	638	1,282	2,358	2,231	1,470	829	253	28	
	불일치율	17.3	4.1	8.7	9.3	11.8	24.9	28.2	27.7	31.1	30.7	29.5	
2차 시험 조사 자료	범천1동	사례수	1,415	21	74	127	166	281	297	249	158	40	2
		불일치율	20.4	6.9	10.9	11.1	14.6	27.5	26.9	28.8	29.2	29.2	22.2
	명장1동	사례수	3,426	42	209	257	339	851	969	504	179	69	7
		불일치율	22.5	5.5	9.6	11.0	17.7	32.0	33.5	31.7	29.1	31.5	31.8
	기흥동	사례수	1,174	59	66	59	239	263	216	125	104	42	1
		불일치율	12.6	3.7	5.9	7.0	9.5	19.3	27.1	20.8	31.2	36.2	8.3
	광남동	사례수	2,395	89	227	135	388	646	461	271	132	43	3
		불일치율	12.4	3.2	7.8	6.8	9.4	19.5	22.0	20.6	23.4	24.9	15.0
	목호동	사례수	728	11	32	27	44	102	150	184	142	32	4
		불일치율	27.3	6.0	14.2	11.7	12.1	34.0	35.1	36.2	43.0	36.0	33.3
인제읍	사례수	889	35	73	33	106	215	138	137	114	27	11	
	불일치율	19.1	5.7	9.9	10.0	13.4	26.9	23.2	32.9	40.7	30.3	55.0	
전체	사례수	14,532	611	1,759	1,430	2,443	3,625	1,970	1,220	1,136	301	37	
	불일치율	20.4	8.0	12.3	18.6	21.9	24.6	24.1	29.4	45.2	43.8	41.1	
3차 시험 조사 자료	문흥1동	사례수	3,164	134	443	380	460	805	452	251	187	48	4
		불일치율	18.3	8.3	11.8	16.9	20.1	23.2	20.3	24.9	40.2	35.8	21.1
	월곡1동	사례수	2,946	173	445	349	495	703	387	171	165	51	7
		불일치율	22.5	12.2	15.5	21.3	24.1	26.1	27.5	31.8	50.8	48.1	38.9
	교동	사례수	2,644	111	234	195	554	664	377	241	199	54	15
		불일치율	17.7	5.6	8.8	16.3	20.0	22.4	21.7	26.2	38.0	38.3	57.7
	소이면	사례수	666	14	37	24	37	71	85	129	203	59	7
		불일치율	39.4	14.7	24.5	27.3	34.9	34.5	31.8	39.3	59.5	61.5	58.3
	활천동	사례수	4,287	150	511	434	829	1,283	540	284	208	46	2
		불일치율	19.6	6.2	11.3	18.4	22.5	25.3	24.6	29.1	41.8	45.1	20.0
거류면	사례수	825	29	89	48	68	99	129	144	174	43	2	
	불일치율	35.1	19.1	32.7	28.4	25.8	33.1	37.1	38.7	48.1	39.4	40.0	

※3차시험조사 자료 결측 1건 존재해서 의 사례수 14,532가 됨

*사례수 : 불일치 사례수를 뜻함

*불일치율 : 해당 연령대의 불일치 사례수/해당 연령대의 사례수

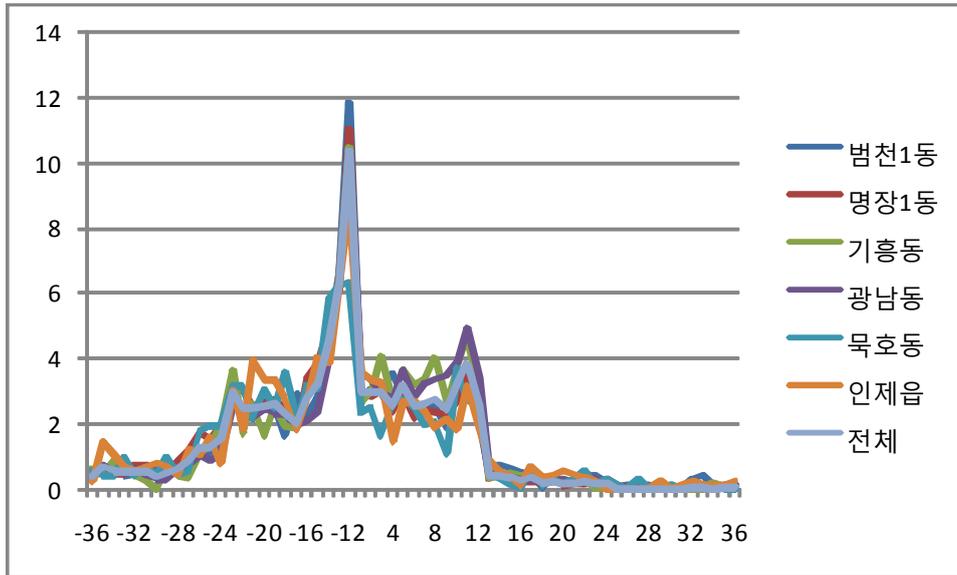
그런데 제1장의 매칭과정의 특성 중에 하나는 생년/월의 매칭키의 이용

에서 생년의 차이를 최대 ± 3 까지 허용한 것이다. 이러한 허용이 위 표의 불일치율을 높인 것으로 본다. 이러한 허용이 이루어진 단계는 제2차 시험조사 자료에서는 4,5,9,10,14,15,16,18,19,20,23,25단계에서 고려하였고, 제3차 시험조사 자료에서는 4,12,13,14,15,17,18단계에서 고려하였다. 생년의 차이 허용이 이루어진 단계의 매칭률은 2차, 3차 자료가 14.41%, 14.06%였다. 시험조사 자료와 행정자료의 생년월로부터 연령을 계산하여 매칭에 이용한바 두 자료간의 연령차이를 개월 수의 차이로 계산해 보았다. 그런데 제2차 시험조사 자료의 경우에는 주민생년만을 응답받았기 때문에 주민생년월은 연앙값인 6월로 대체하여 시험조사 자료와 주민등록 자료의 연령 차이 불일치 사례수를 얻었다. 제2차 시험조사 자료, 제3차 시험조사 자료의 행동구역별 연령차이 불일치 사례는 행정동별로 비슷한 패턴을 보였다. 전체적으로 11개월 이하로 차이가 나는 사례가 가장 많았으며, 12~24개월 사이와 24~36개월 사이에 연령차이 불일치 사례가 증가하는 패턴을 보인다. 또한 개월 차이별로 불일치 사례 패턴을 보면, 12개월 차이가 가장 많았다. 특히, 제3차 시험조사 자료는 생년/월까지 조사된 것으로 1개월 차이도 많이 나고 60%가량이 11개월 이하의 차이인 것으로 나타났는데 이는 음력과 양력의 혼용으로 인한 결과일 것으로 추측된다.

<표 2-3> 행정동별 생년월 개월수 불일치 건수

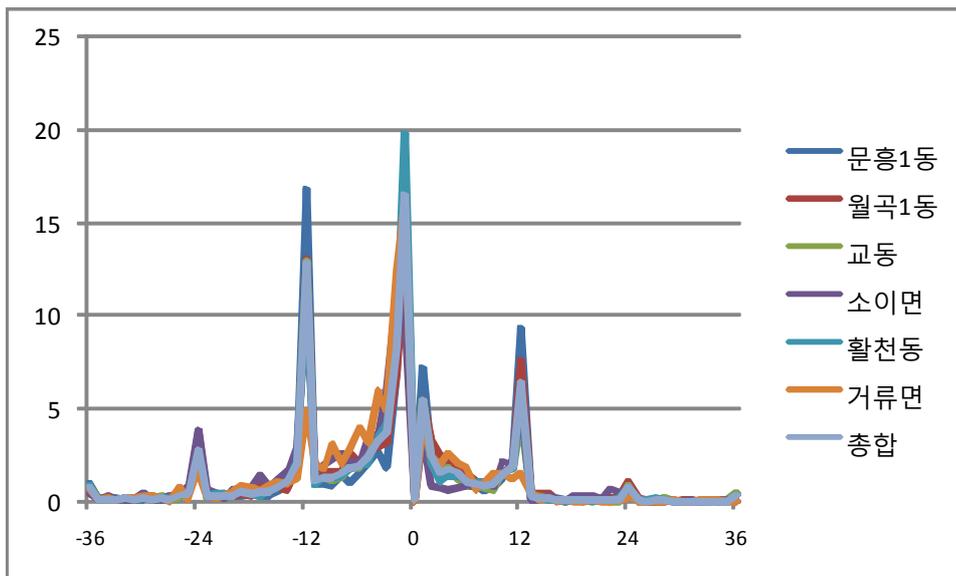
	전체	불	불	11	12	23	24	35	36	37	
		일치 건수	일치율 (%)	개월 이하	개월	개월 이하	개월	개월 이하	개월	개월 이상	
	전체	58,062	10,027	17.3	3,248	1,300	3,824	175	897	49	534
2차 시험 조사 자료	범천1동	6,942	1,415	20.4	449	195	527	23	127	6	88
	명장1동	15,197	3,426	22.5	1,012	461	1,354	75	339	14	171
	기흥동	9,301	1,174	12.6	439	157	429	23	84	9	33
	광남동	19,290	2,395	12.4	907	329	844	31	178	12	94
	목호동	2,666	728	27.3	188	64	297	16	71	4	88
	인제읍	4,666	889	19.1	253	94	373	7	98	4	60
	전체	71,129	14,532	20.4	9223	2811	1359	520	454	163	2
3차 시험 조사 자료	문흥1동	17,256	3,164	18.3	1814	826	252	140	89	43	
	월곡1동	13,075	2,946	22.5	1836	607	282	115	80	24	2
	교동	14,933	2,644	17.7	1696	459	268	98	89	34	
	소이면	1,690	666	39.4	409	107	92	29	26	3	
	활천동	21,823	4,287	19.6	2813	759	393	123	145	54	
	거류면	2,352	825	35.1	655	53	72	15	25	5	

※(충북)교동의 결측 1건 존재해서 3차시험조사 자료의 사례수 14,532가 됨



(단위 : %)

<그림 2-3> 2차 자료의 행정동별 연령별 연령차이 불일치 패턴



(단위 : %)

<그림 2-4> 3차 자료의 행정동별 연령별 연령차이 불일치 패턴

다. 세대주와 관계

세대주는 주민등록상에 나타나는 법적인 개념이고, 가구주는 세대주와는 관계없이 가구를 실질적으로 대표하는 개념으로 정의된다. 시험조사 자료와 주민등록 자료를 매칭시키는 작업에서는 가구주와 세대주가 같다는 전제하에 세대주와 관계코드를 가구주와 관계코드(14개)로 정리하여 매칭하였다.

세대주와 관계의 불일치는 제2차 시험조사 자료의 매칭에서 9.3%(5,410건)로 나타났으며, 제3차 시험조사 자료의 매칭에서는 9.1%(6,503건)이 불일치한 것으로 나타났다.

<표 2-4> 행정동별 세대주와 관계 불일치

(단위 : 사례수, %)

	전체	불 일치 건수	불 일치 율	*1	*2	*3	*4	*5	*6	*7	*8	*9	*10	*11	*12	*13	*14	
전체	58,056	5,410	9.3	2,519	789	812	147	336	49	291		12	196	66	5	7	181	
법천1동	6,942	776	11.2	311	87	157	22	66	3	38		1	46	4	1		40	
2차 시험 조사 자료	명장1동	15,197	1,052	6.9	425	149	164	29	101	4	60		2	52	14		3	49
	기흥동	9,300	858	9.2	477	122	93	17	52	10	21		2	24	13	3	2	22
	광남동	19,286	1,928	10.0	966	332	272	36	79	30	74		5	60	29	1	1	43
	목호동	2,666	329	12.3	147	44	57	10	16	2	28		1	6	3		1	14
	인제읍	4,665	467	10.0	193	55	69	33	22		70		1	8	3			13
전체	71,125	6,503	9.1	2,860	545	752	699	317	65	348	3	9	276	119	5	217	288	
3차 시험 조사 자료	문흥1동	17,256	1,704	9.9	851	104	126	180	84	31	68			116	43	2	55	44
	월곡1동	13,074	1,287	9.8	576	111	133	152	59	10	56		4	56	29	2	42	57
	교동	14,933	1,396	9.3	631	146	176	162	70	7	96	3	2	32	10		23	38
	소이면	1,688	257	15.2	100	27	52	25	16	1	22		1	3	2		2	6
	활천동	21,822	1,632	7.5	602	144	221	157	79	13	94		2	66	35	1	93	125
	거류면	2,352	227	9.7	100	13	44	23	9	3	12			3			2	18

※2차시험조사 자료의 결측 6건 존재해서 사례수 58,056가 됨

※2차시험조사 자료의 결측 5건 존재해서 사례수 71,125가 됨

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:증손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

세대주와 관계에서 주민등록 자료상의 세대주가 시험조사 자료에서 가구주

가 아니라 한 개체가 많았는데 2차 자료의 경우, 총 불일치 건수 5,410건 중 약 47%(2,519건)에 달하며, 3차 자료의 경우, 총 불일치 건수 6,508건 중 44%(2,880건)에 달하였다. 이를 시험조사 자료와 주민등록 자료의 교차표로 종합적으로 살펴보면 다음 표와 같다. 전체적으로 행정자료의 세대주가 많은 경우에 조사자료의 가구주의 배우자, 가구주의 자녀, 세대주의 부모, 형제자매 그 배우자, 동거인으로 표기되어있고, 특히 3차 자료의 경우는 행정자료의 세대주가 조사자료의 세대주의 부모로 표기 되어있는 경우도 많은 것으로 나타났다. 이는 법적인 개념의 세대주와의 관계와 가구주와의 관계가 동일하지 않기 때문에 나타난 결과로 보인다.

<표 2-5> 세대주와 관계 불일치 패턴

(단위 : 사례수)

		세대주관계														
		**	**	**	**	**	**	**	**	**	**	**	**	**	**	
가구주관계		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
전체		19,702	13,249	21,123	365	1,290	183	1,011	13	32	527	138	13	203	207	
2차 시험 조사 자료	*1	21,077	18,558	1,010	593	50	304	40	27	1	7	198	12	5	111	161
	*2	12,718	524	11,929	47	102	80	11	2			16		2	2	3
	*3	20,837	297	21	20,025	7			342	8		49	62		17	9
	*4	348	13	119	12	201			1			1			1	
	*5	1,194	221	77	1	2	858	30			3	1			1	
	*6	146	9	4			34	97							1	1
	*7	909	5	1	275				618	3		1	5		1	
	*8	1								1						
	*9	33	4				6	2			21					
	*10	445	38	9	104	1			11		1	249	3	2	24	3
	*11	120	4		31				6				54	1	24	
	*12	6	1	1								1		1	2	
	*13	11	1	1	3							1			4	1
	*14	211	27	77	32	2	8	3	4			10	2	1	15	30
전체		22,556	16,354	28,634	183	1,419	191	693	5	35	538	167	32	156	162	
3차 시험 조사 자료	*1	23,966	21,106	1,702	464	21	348	35	13	3	119	17	7	43	88	
	*2	14,694	355	14,149	17	42	87	8			9	9	11	6	1	
	*3	28,178	362	32	27,426	4	1		199			75	55	1	15	8
	*4	810	400	283	11	111		1				2			1	1
	*5	1,239	196	76	4		922	27			11		1			2
	*6	183	10	8			44	118			2			1		
	*7	720	3		338	1			372	1			4		1	
	*8	6							3	3						
	*9	22	3				6				13					
	*10	527	60	14	180	2			4			251	5	5	6	
	*11	178	3		69				16			2	59		28	1
	*12	8	3				1					1		3		
	*13	247	7	7	39		2		80		6	61	11	2	30	2
	*14	347	48	83	86	2	8	2	6	1		18	6	2	26	59

*세대주와 관계

**가구주와 관계

2. W단계

제2차 시험조사 자료의 매칭에서는 총 25단계까지 수행된 매칭 작업 중 22 단계부터 25단계까지를 W단계라고 정의하였고, 제3차 시험조사 자료의 매칭에서는 총 19단계까지 수행한 매칭 작업 중 19단계를 W단계라고 정의하였다. 2차 자료의 W단계에서는 주민등록 자료 기준으로 총 87,414건 중 8.2%(7,205건)가 매칭되었다. 3차 자료의 W단계에서는 주민등록 자료 기준으로 총 91,175건 중 2.5%(2,308건)가 매칭되었다.

2차 시험조사 자료에서 고려한 매칭키는 본번/성명/생년/성별이고, 3차 시험조사에서 고려한 매칭키는 성명/생년월을 고려했기 때문에 2차 시험조사 자료에서는 연령과 가구주(세대주)관계에 대한 항목오류 분석을 실시하고, 3차 시험조사 자료에서는 가구주(세대주)관계에 대한 항목오류 분석을 실시한다.

가. 세대주와 관계

세대주와 관계의 불일치는 제2차 시험조사 자료의 매칭에서는 14.3%(1,027건)로 나타났고 제3차 시험조사 자료의 매칭에서는 20.6%(475건)로 나타났다.

세대주와 관계 불일치율이 S+M단계에 비하여 10%가량 높아졌지만 항목내용별 불일치율의 패턴은 유사하였다.

<표 2-6> 행정동별 세대주와 관계 불일치

(단위 : 사례수, %)

	전체	불일치 건수	불일 치율	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	전체	7,204	1,027	14.3	411	129	201	47	54	13	81	3	30	18	2	1	37
2차 시험 조사 자료	범천1동	163	23	14.1	12	1	4	0	3	1	1						1
	명장1동	284	26	9.2	10	2	3	1	2	3	3			1			1
	기흥동	829	117	14.1	55	15	17	6	4	1	6	2	1	3	1	1	5
	광남동	4,411	580	13.1	253	89	105	17	30	7	30	1	17	9	1		21
	목호동	373	73	19.6	18	5	17	8	3		15		4	2			1
	인제읍	1,144	208	18.2	63	17	55	15	12	1	26		8	3			8
전체	2307	475	20.6	194	21	85	24	20	4	50	2	1	16	17	0	26	15
3차 시험 조사 자료	문흥1동	287	66	23.0	24	4	10	2	5	1	7		5	4		4	
	월곡1동	254	52	20.5	35	1	6	3	1		1			1		2	2
	교동	376	93	24.7	40	5	15	1	7	1	11	2	1	4		4	2
	소이면	384	66	17.2	22	6	12	6			17					2	1
	활천동	247	76	30.8	24	1	16	3	3	1	2		5	4		11	6
	거류면	759	122	16.1	49	4	26	9	4	1	12	0	1	5	4		3

※2차시험조사 자료의 결측 1건 존재해서 사례수 7,204가 됨

※2차시험조사 자료의 결측 1건 존재해서 사례수 2,307이 됨

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:증손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

<표 2-7> 세대주와 관계 불일치 패턴

(단위 : 사례수)

		가구	**	**	**	**	**	**	**	**	**	**	**	**	**	
		주관	1	2	3	4	5	6	7	8	9	10	11	12	13	14
		계														
2차 시험 조사 자료	세대주 관계	전체	2,446	1,718	2,381	59	247	22	158	6	5	81	21	2	34	24
	*1	2,674	2,263	146	105	11	69	6	7		1	25	6	1	15	19
	*2	1,619	65	1,490	12	20	25	1				1			2	3
	*3	2,343	71	9	2,142	3			77	4		17	10		9	1
	*4	72	5	41	1	25										
	*5	194	30	16			140	6			1				1	
	*6	21	2	2			9	8								
	*7	146			79				65	2						
	*8															
	*9	6					3				3					
	*10	66	3	1	17				4			36	2			3
	*11	21			12				4				3			2
	*12	2	1													1
	*13	2			1											1
	*14	38	6	13	12		1	1	1			2			1	1
		전체	869	479	766	8	53	6	58		1	18	7		7	35
3차 시험 조사 자료	*1	963	769	81	53	1	23	5				8	1		2	20
	*2	381	10	360		2	8					1				
	*3	719	43	8	634				25			3	4		1	1
	*4	29	13	11		5										
	*5	40	14	5			20	1								
	*6	4	2	1			1									
	*7	75	2	1	45				25				2			
	*8	2	0		2											
	*9	2	1									1				
	*10	19	5	5	4				1			3				1
	*11	17			13				2			1			1	
	*12															
	*13	29	3	3	12				5			2			3	1
	*14	27	7	4	3		1									12

*세대주와 관계

**가구주와 관계

나. 2차 시험조사 자료의 생년/월 항목

생년/월의 불일치는 95.5%(6,879건)로 나타난다. 이는 S+M단계의 생년/월 불일치와 비교해 볼 때 크게 높아진 값이지만, 11개월 이하의 불일치건수를 보면 83%정도(행정동별로 보면 72%이상)가 11개월 이하의 불일치이다. 이는

제2차 매칭 자료의 경우 생년만 조사한 것 연앙값을 이용하여 생월을 생성하여 계산한 결과이기 때문에 생월까지 조사된 자료임을 고려한다면 11개월 이하의 불일치율이 낮아질 것으로 생각해볼 수 있다.

<표 2-8> 2차 자료 행정동별 연령차이별 불일치 건수

(단위 : 사례수, %)

	불 일 치 건수	불 일 치 율 (%)	11 개 월 이 하	12 개 월	23 개 월 이 하	24 개 월	35 개 월 이 하	36 개 월	37 개 월 이 상	
전체	7,205	6,879	95.5	5,711	135	970	4	34	1	24
범천1동	163	160	98.2	120	6	28	2	3		1
명장1동	284	275	96.8	197	9	50		12		7
기흥동	829	809	97.6	597	18	187		4	1	2
광남동	4,412	4,209	95.4	3,498	94	598		10		9
목호동	373	343	92.0	305	3	28	1	3		3
인제읍	1,144	1,083	94.7	994	5	79	1	2		2

다. 3차 시험조사 자료의 성별항목

성별의 불일치는 0.6%(15건)로 나타난다. 이는 S+M단계의 생년/월 불일치와 비교해 유사한 결과라고 할 수 있다.

<표 2-9> 3차 자료 행정동별 성별 불일치 건수

	전체	문흥1동	월곡1동	교동	소이면	활천동	거류면
전체	전체 2,308	287	254	376	384	248	759
불일치	15	1		3	1	2	8
불일치율	0.6	0.3		0.8	0.3	0.8	1.1
남자	전체 1,223	143	126	188	214	129	423
불일치	8	1		1	1	1	4
불일치율	0.7	0.7		0.5	0.5	0.8	0.9
여자	전체 1,085	144	128	188	170	119	336
불일치	7			2		1	4
불일치율	0.6			1.1		0.8	1.2

이를 종합해 볼 때, 매칭에 의문을 제기한 W단계의 항목 불일치율이 높기는 하지만 S+M단계의 불일치패턴과 유사한 것은 두 개체가 매칭되었다고 판단할 수 있는 근거라고 할 수 있다.

3. 이름항목의 불일치 사례

특히 제2차, 3차 자료 전체에서 나타나는 이름의 불일치 사례를 심층 분석하여 면접과정에서 나타날 수 있는 한글 오류 유형을 조사하였다. 아래의 표는 성과 이름을 구별하여 분석한 것으로 몇 개(이 → 김)를 제외하고 현실적으로 나타날 수 있는 오류로 판단되어 대부분이 조사 자료의 오류로 판단된다. 이름의 경우는 다양하기 때문에 모음과 자음으로 구분하여 현실적으로 나타날 수 있는 오류를 정리해 보았다.

<표 2-10> 성의 불일치 사례

행정자료	조사자료					행정자료	조사자료			
강	감	김	깁	망	장	오	고	조	우	도
곽	박					유	윤	류		
김	기	길	강	검	임	이	기	김	미	임
나	라					장	강	정	전	
류	유					정	전	점		
문	운					제	재			
박	방	백				조	고	주		
백	배					지	기			
송	성	성	홍			채	최	화		
양	양	영				최	허			
업	입					황	성	한	항	환

<표 2-11> 모음과 자음 기준의 불일치 사례

행정자료	조사자료									
ㄱ	ㅅ	ㅈ	ㄹ	ㅁ	ㅂ	표	ㅇ	ㅎ	ㅌ	
ㄴ	ㅁ	ㅇ	ㅎ	ㅈ	ㄹ	ㄷ	ㅅ			
ㄷ	ㄴ	ㄱ	ㅌ	ㅌ	ㅁ	ㅈ	ㄹ	ㅂ	ㅇ	ㅅ
ㄹ	ㄴ	ㄷ	ㅁ	ㅌ	ㅎ	ㅈ	ㅅ			
ㅁ	ㄴ	ㄹ	ㅇ							
ㅂ	ㅇ	ㄱ	ㄴ							
ㅇ	ㄴ	ㄱ	ㅁ							
ㅌ	ㅌ	ㅌ	ㅌ	ㅣ						
ㅌ	ㅋ	ㅌ	ㅋ	ㅋ						
ㅋ	ㅌ	ㅡ	ㅣ	ㄱ						
ㅋ	ㅋ	ㅣ	ㄱ	ㅋ						
ㄴ	ㅌ	ㅋ	ㅣ							
ㅌ	ㅌ	ㅌ	ㅋ							
ㅌ	ㅌ									
ㅌ	ㅌ									
ㄱ	ㅌ	ㅡ	ㅌ							
ㄱ	ㅌ	ㅌ	ㅣ							
ㅌ	ㄱ	ㅌ								
ㅡ	ㅌ	ㄱ	ㄱ							
ㅣ	ㅌ	ㄱ	ㅌ	ㅋ						

제4절 주택 매칭 자료의 불일치 항목 분석

주택 매칭 자료의 경우 S단계와 M단계로만 정의하였기 때문에 이를 통합하여 불일치분석을 한다. 2차 자료의 경우는 총 58단계까지 매칭작업으로 행정자료 기준 25,130건 중에서 69.8%(17,532건), 3차 자료의 경우는 총 46단계까지 매칭 작업으로 행정자료 기준 26,895건 중에서 90.8%(24,433건)의 매칭율이었다.

주택 관련 항목은 주택종류, 면적, 건축연도가 있는데 각 항목을 구성하는 내용이 다양하여 재 구분하였다.

주택종류는 일반단독, 다가구단독, 영업겸용단독, APT, 연립주택, 다세대주택, 비거주용 건물(상가, 공장, 여관)내 주택과 판잣집, 비닐하우스로 구분이 된다. 이를 “일반단독”, “단독주택”(다가구단독, 영업겸용단독), “APT”, “집합주택”(연립주택, 다세대주택, 비거주용 건물(상가, 공장, 여관 등)내 주택)으로 하고 판잣집과 비닐하우스는 “그 외 주택”으로 그룹화하였다.

면적은 46 m^2 미만, 46~96 m^2 미만, 96 m^2 이상으로 재구분하여 살펴보았다.

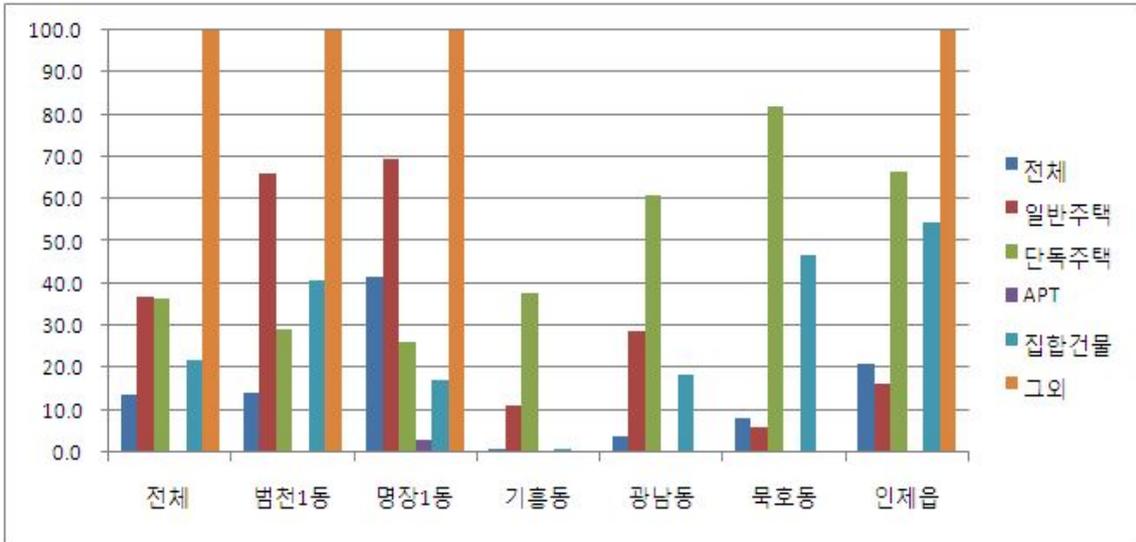
건축연도는 2000년대, 1990년대, 1990년 이전으로 구분하였다.

1. 주택부분

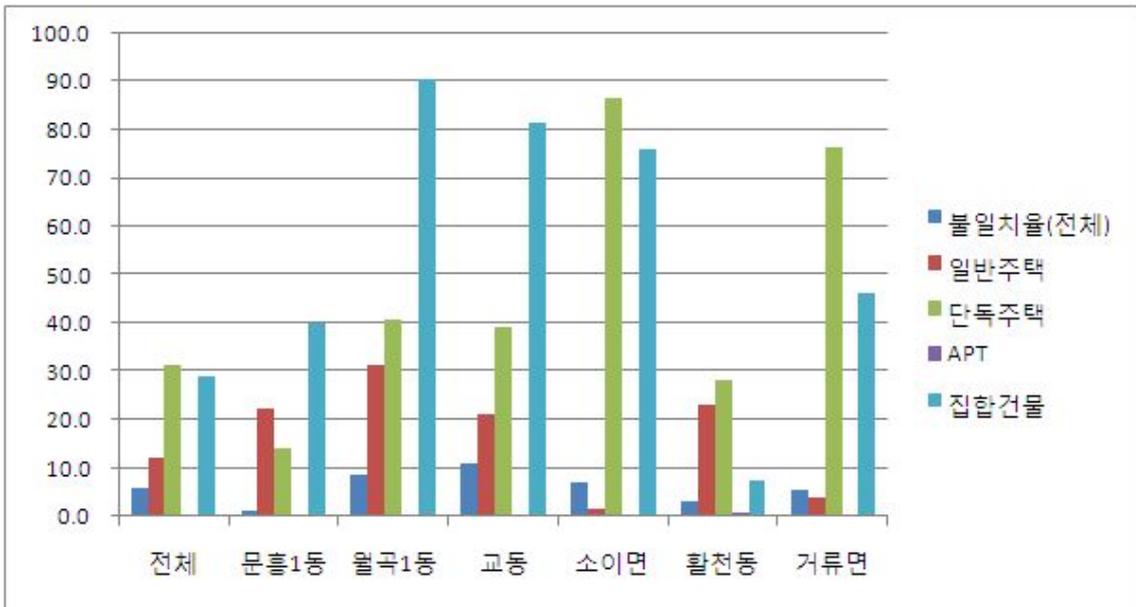
가. 주택종류

면적 불일치는 제2차 시험조사 자료에서는 총 17,532건의 매칭된 자료 중 13.6%(2,385건), 제3차 시험조사 자료에서는 24,433건의 매칭된 자료 중 5.8%(1,411건)로 나타났다. 3차 자료의 경우 전체적인 불일치는 5.8%로 2차 자료의 전체 불일치율 13.6%에 비해 낮은데 그 이유는 3차 자료의 주택종류 중 APT의 비율이 75%로 2차 자료 63%에 비해 높은 것과 2차의 행정자료에 만 있는 그 외 주택(비닐하우스 및 판잣집)의 영향으로 설명 될 수 있다고 본다.

전체적으로 APT인 경우가 주택종류 불일치율이 낮으며 APT 이외의 주택의 주택종류 불일치율이 높게 나타나고 있다.



<그림 2-5> 2차 자료 행정동별 주택종류 불일치 비율



<그림 2-6> 3차 자료 행정동별 주택종류 불일치 비율

<표 2-12> 주택종류 불일치 사례

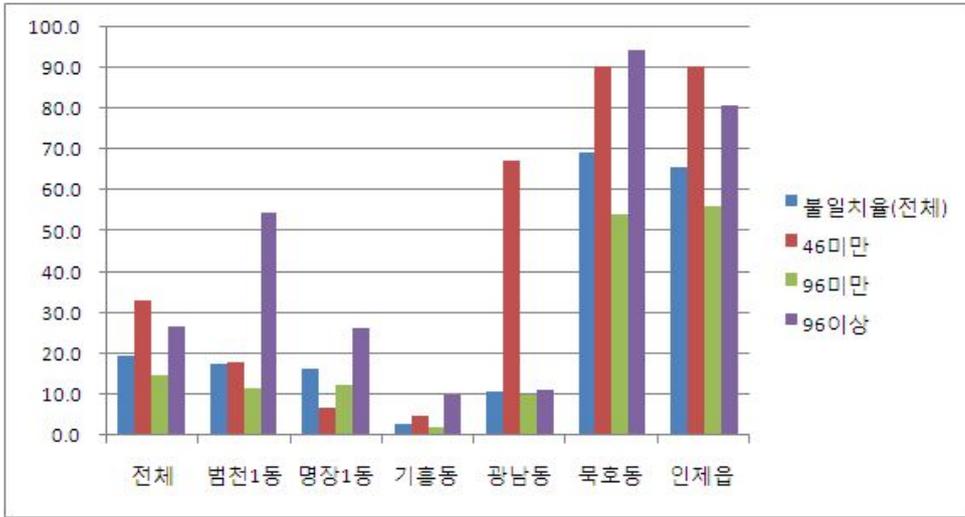
	2차 시험조사 자료							3차 시험조사 자료							
	전체	범천 1동	명장1 동	기흥 동	광남 동	목호 동	인제 읍	전체	문흥 1동	월곡 1동	교동	소이 면	활천 동	거류 면	
전체	전체건수	17,532	2,303	3,417	3,519	5,697	1,191	1,405	24,433	5,290	3,898	5,719	1,088	6,871	1,567
	불일치건수	2,385	324	1,414	24	228	99	296	1,411	62	332	627	76	226	88
	불일치율	13.6	14.1	41.4	0.7	4.0	8.3	21.1	5.8	1.2	8.5	11.0	7.0	3.3	5.6
일반	전체건수	3,298	314	1,069	128	246	754	787	2,725	18	347	671	547	146	996
주택	불일치건수	1,210	207	744	14	71	45	129	336	4	109	143	8	34	38
(단독)	불일치율	36.7	65.9	69.6	10.9	28.9	6.0	16.4	12.3	22.2	31.4	21.3	1.5	23.3	3.8
단독	전체건수	503	99	289	8	23	39	45	1,046	381	453	127	15	53	17
주택	불일치건수	183	29	75	3	14	32	30	330	54	185	50	13	15	13
	불일치율	36.4	29.3	26.0	37.5	60.9	82.1	66.7	31.5	14.2	40.8	39.4	86.7	28.3	76.5
APT	전체건수	11,122	1,723	873	3,155	4,679	351	341	18,261	4,881	3,056	4,390	455	5,005	474
	불일치건수	37	1	25	5	5	0	1	52	0	0	1	1	50	0
	불일치율	0.3	0.1	2.9	0.2	0.1	0.0	0.3	0.3	0.0	0.0	0.0	0.2	1.0	0.0
집합	전체건수	2,111	135	742	228	749	47	210	2,401	10	42	531	71	1,667	80
건물	불일치건수	457	55	126	2	138	22	114	693	4	38	433	54	127	37
	불일치율	21.6	40.7	17.0	0.9	18.4	46.8	54.3	28.9	40.0	90.5	81.5	76.1	7.6	46.3
그외	전체건수	498	32	444				22							
	불일치건수	498	32	444				22							
	불일치율	100.0	100.0	100.0				100.0							

나. 면적

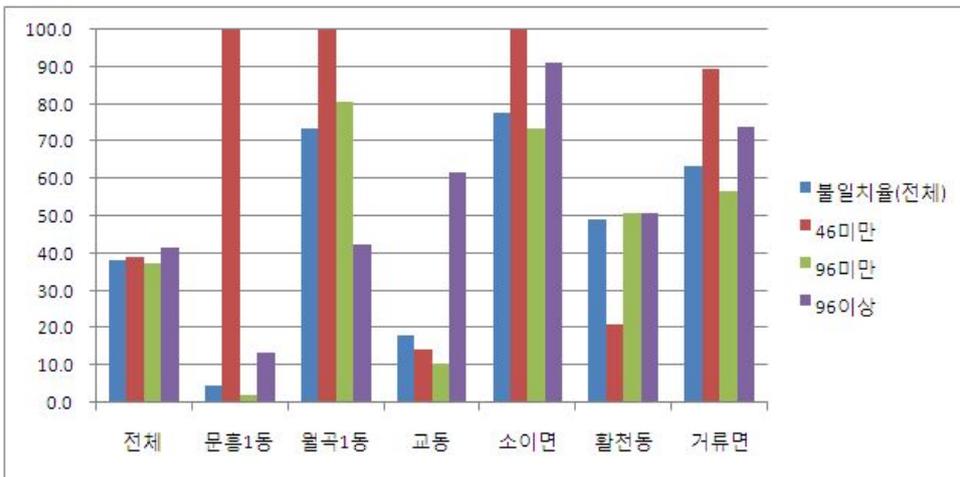
주택종류 불일치는 제2차 시험조사 자료 매칭에서는 17,532건의 매칭된 자료 중 19.6%(3,429건), 제3차 시험조사 자료 매칭에서는 24,433건의 매칭된 자료 중 38.3%(9,353건)로 나타났다.

2차 자료의 (강원)목호동과 (강원)인제읍과 3차 자료의 (충북)소이면, (경남)거류면의 주택종류 불일치율이 높게 나타났다. 이 네 지역의 특성은 모두 다른 지역에 비해 주택종류가 단독주택(일반단독, 다가구단독, 영업겸영단독)이 50%이상을 차지하는 것으로 나타났다.

(참고 : 목호동 67%(793건), 인제읍 59%(832건), 소이면 52%(562건), 거류면 65%(1,013건))



<그림 2-7> 2차 자료 행정동별 면적 불일치 비율



<그림 2-8> 3차 자료 행정동별 면적 불일치 비율

<표 2-13> 면적 불일치 사례

		전체			46m ² 미만			96m ² 미만			96m ² 이상		
		전체건수	불일치건수	불일치율	전체건수	불일치건수	불일치율	전체건수	불일치건수	불일치율	전체건수	불일치건수	불일치율
	전체	17,532	3,429	19.6	2,367	787	33.2	11,892	1,764	14.8	3,273	878	26.8
2차시 협조 사자 료	범천1동	2,303	407	17.7	1,072	192	17.9	1,059	121	11.4	172	94	54.7
	명장1동	3,417	557	16.3	451	31	6.9	1,833	227	12.4	1,133	299	26.4
	기흥동	3,519	101	2.9	222	11	5	3,047	65	2.1	250	25	10
	광남동	5,697	615	10.8	46	31	67.4	4,302	435	10.1	1,349	149	11
	목호동	1,191	827	69.4	395	358	90.6	705	383	54.3	91	86	94.5
	인제읍	1,405	922	65.6	181	164	90.6	946	533	56.3	278	225	80.9
	전체	24,433	9,353	38.3	1,253	490	39.1	19,342	7,262	37.5	3,838	1,601	41.7
3차시 협조 사자 료	문흥1동	5,290	239	4.5	3	3	100	4,105	80	1.9	1,182	156	13.2
	월곡1동	3,898	2,864	73.5	1	1	100	3,158	2,550	80.7	739	313	42.4
	교동	5,719	1,023	17.9	518	73	14.1	4,414	465	10.5	787	485	61.6
	소이면	1,088	845	77.7	125	125	100	883	647	73.3	80	73	91.3
	활천동	6,871	3,386	49.3	371	78	21	5,629	2,866	50.9	871	442	50.7
	거류면	1,567	996	63.6	235	210	89.4	1,153	654	56.7	179	132	73.7

면적 불일치 사례를 주택종류별로 살펴보면 2차 자료는 일반주택의 불일치율이 매우 높은 것을 알 수 있으며 특히 46m²미만에서 더욱 높은 불일치율을 보인다. 96m²이상에서는 집합건물의 불일치율이 높아짐을 알 수 있다.

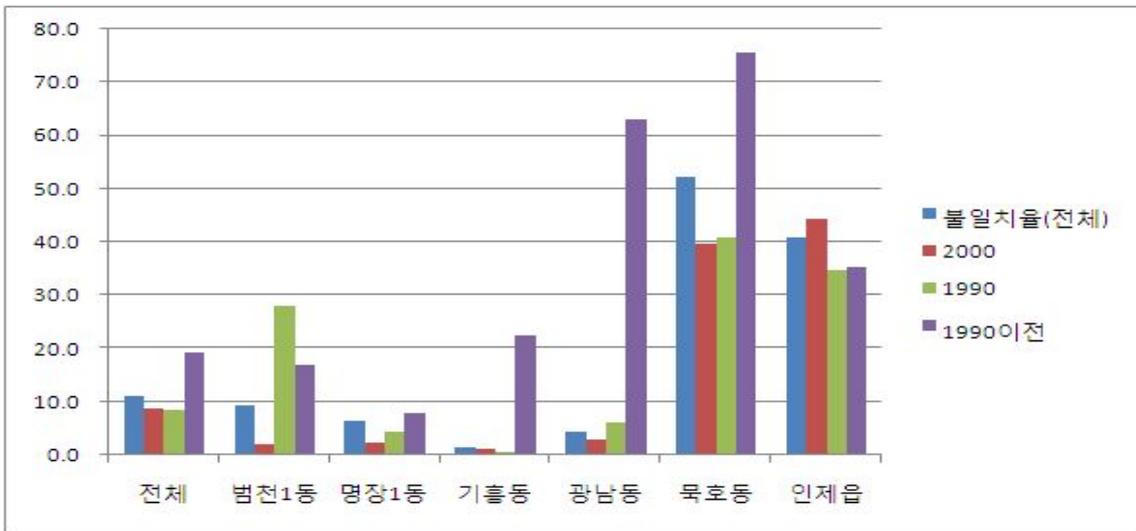
3차 자료는 APT의 불일치율이 32.6%로 높게 나타났는데, 두 자료의 면적 불일치정도가 차이가 1~3m²가 가장 많은 것으로 평수에서 m²로의 환산에서 생기는 문제로 생각된다.

<표 2-14> 주택종류에 따른 면적 불일치 사례

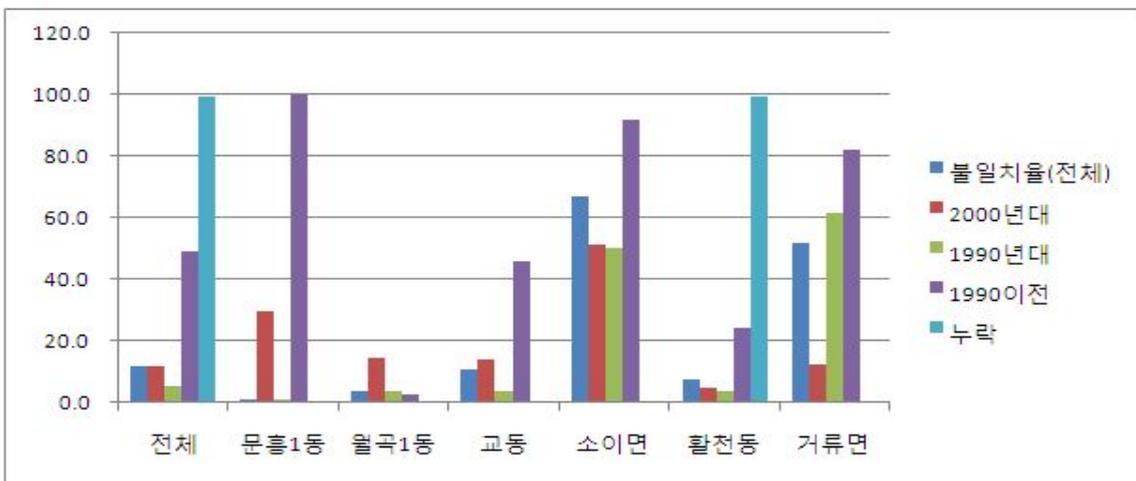
		전체	일반	단독	APT	집합	그외	
2차 시험 조사 자료	전체	전체건수	17,532	3,298	503	11,122	2,111	498
		불일치건수	3,429	2,195	260	227	712	35
		불일치율	19.6	66.6	51.7	2.0	33.7	7.0
	46m ² 미만	전체건수	2,367	734	32	1,409	162	30
		불일치건수	787	671	29	10	59	18
		불일치율	33.2	91.4	90.6	0.7	36.4	60.0
	96m ² 미만	전체건수	11,892	1,226	117	8,350	1,742	457
		불일치건수	1,764	932	62	213	545	12
		불일치율	14.8	76.0	53.0	2.6	31.3	2.6
	96m ² 이상	전체건수	3,273	1,338	354	1,363	207	11
		불일치건수	878	592	169	4	108	5
		불일치율	26.8	44.2	47.7	0.3	52.2	45.5
3차 시험 조사 자료	전체	전체건수	24,433	2,725	1,046	18,261	2,401	
		불일치건수	9,353	2,154	489	5,692	1,018	
		불일치율	38.3	79.0	46.7	31.2	42.4	
	46m ² 미만	전체건수	1,253	429	11	685	128	
		불일치건수	490	401	10	15	64	
		불일치율	39.1	93.5	90.9	2.2	50.0	
	96m ² 미만	전체건수	19,342	1,503	92	15,678	2,069	
		불일치건수	7,262	1,257	60	5,109	836	
		불일치율	37.5	83.6	65.2	32.6	40.4	
	96m ² 이상	전체건수	3,838	793	943	1,898	204	
		불일치건수	1,601	496	419	568	118	
		불일치율	41.7	62.5	44.4	29.9	57.8	

다. 건축연도

주택종류 불일치는 제2차 시험조사 자료 매칭에서는 17,532건의 매칭된 자료 중 11%(1,925건), 제3차 시험조사 자료 매칭에서는 24,433건의 매칭된 자료 중 11.9%(2,904건)로 나타났다. 건축연도에서도 단독주택(일반단독, 다가구단독, 영업겸영단독)이 50%이상을 차지하는 2차 자료의 (강원)묵호동과 (강원)인제읍과 3차 자료의 (충북)소이면, (경남)거류면의 주택종류 불일치율이 높게 나타났다.



<그림 2-9> 2차 자료 행정동별 건축연도 불일치 비율



<그림 2-10> 3차 자료 행정동별 건축연도 불일치 비율

<표 2-15> 건축연도 불일치 사례

	전체 건수	전체			2000년대			1990년대			1990이전			누락		
		불일 치건 수	불일 치율	전체 건수	불일 치건 수	불일 치율	전체 건수	불일 치건 수	불일 치율	전체 건수	불일 치건 수	불일 치율	전체 건수	불일 치건 수	불일 치율	
전체	17,532	1,925	11.0	8,815	210	2.4	4,131	350	8.5	3,948	756	19.1	638	609	95.5	
2차 시험 조사 자료	범천1동	2,303	213	9.2	1,221	0	0.0	71	20	28.2	985	167	17.0	26	26	100.0
	명장1동	3,417	224	6.6	222	4	1.8	904	39	4.3	2,290	180	7.9	1	1	100.0
	기흥동	3,519	47	1.3	2,149	4	0.2	1,283	8	0.6	67	15	22.4	20	20	100.0
	광남동	5,697	244	4.3	4,275	90	2.1	1,322	80	6.1	57	36	63.2	43	38	88.4
	목호동	1,191	622	52.2	382	27	7.1	193	79	40.9	407	308	75.7	209	208	99.5
	인제읍	1,405	575	40.9	566	85	15.0	358	124	34.6	142	50	35.2	339	316	93.2
전체	24,433	2,904	11.9	3,598	425	11.8	17,940	946	5.3	2,675	1,314	49.1	220	219	99.5	
3차 시험 조사 자료	문흥1동	5,290	67	1.3	27	8	29.6	5,256	52	1.0	7	7	100.0			
	월곡1동	3,898	146	3.7	27	4	14.8	3,051	119	3.9	820	23	2.8			
	교동	5,719	624	10.9	1,999	276	13.8	3,221	120	3.7	499	228	45.7			
	소이면	1,088	731	67.2	37	19	51.4	610	307	50.3	441	405	91.8			
	활천동	6,871	521	7.6	893	42	4.7	5,596	221	3.9	162	39	24.1	220	219	99.5
	거류면	1,567	815	52.0	615	76	12.4	206	127	61.7	746	612	82.0			

** 활천동의 경우 누락 220건 중 219건(99.5%)이 불일치임.

건축연도 불일치를 주택종류별로 살펴보면 2차 자료에서는 단독주택(일반 단독, 다가구단독, 영업겸업단독)의 건축연도가 2000년대인 개체의 불일치 비율이 높고, 3차 자료에서는 단독주택(일반단독, 다가구단독, 영업겸업단독)의 건축연도가 오래된 것의 불일치 비율이 높다.

<표 2-16> 건축연도 불일치 사례 (주택종류)

		전체	일반	단독	APT	집합	그외	
2차	전체	전체건수	17,532	3,298	503	11,122	2,111	498
		불일치건수	1,925	1,549	117	14	229	16
		불일치율	11.0	47.0	23.3	0.1	10.8	3.2
	2000년 대	전체건수	8,815	277	54	7,762	706	16
		불일치건수	210	99	16	10	81	4
		불일치율	2.4	35.7	29.6	0.1	11.5	25.0
	1990년 대	전체건수	4,131	602	99	2,471	953	6
		불일치건수	350	241	30	3	75	1
		불일치율	8.5	40.0	30.3	0.1	7.9	16.7
	1990년 대 이전	전체건수	3,948	1,808	331	889	444	476
		불일치건수	756	625	53	1	66	11
		불일치율	19.1	34.6	16.0	0.1	14.9	2.3
	누락	전체건수	638	611	19		8	
		불일치건수	609	584	18		7	
		불일치율	95.5	95.6	94.7		87.5	
3차	전체	전체건수	24,433	2,725	1,046	18,261	2,401	
		불일치건수	2,904	1,563	172	910	259	
		불일치율	11.9	57.4	16.4	5.0	10.8	
	2000년 대	전체건수	3,598	194	66	2,812	526	
		불일치건수	425	84	14	297	30	
		불일치율	11.8	43.3	21.2	10.6	5.7	
	1990년 대	전체건수	17,940	834	934	14,389	1,783	
		불일치건수	946	253	134	377	182	
		불일치율	5.3	30.3	14.3	2.6	10.2	
	1990년 대 이전	전체건수	2,675	1,697	46	840	92	
		불일치건수	1,314	1,226	24	17	47	
		불일치율	49.1	72.2	52.2	2.0	51.1	
	누락	전체건수	220			220		
		불일치건수	219			219		
		불일치율	99.5			99.5		

제5절 비 매칭 자료의 특성분석

조사자료는 실제로 존재한다는 가정을 하였으므로, 조사자료에는 존재(실존)하지만 등록되지 않은 개체의 특성을 보기위하여 조사자료에서 매칭되지 않은 자료의 인구/주택 관련 항목별 분석을 하고, 행정자료에는 등록되어있지만 조사자료에는 없는(실존하지 않는) 개체의 특성을 보기 위하여 행정 자료에서 매칭되지 않는 자료의 인구/주택 관련 항목별로 분석한다.

1. 인구

가. 비 매칭 행정자료의 항목별 특성분석

제2차 시험조사 지역의 주민등록 자료는 87,414건이며, 이 중 65,267건이 매칭되어 22,147건이 남게 되었다. 제3차 시험조사 지역의 주민등록 자료는 91,175건이며, 73,438건이 매칭되어 17,737건이 남게 되었다.

시험조사 자료와 매칭되고 남은 주민등록자료에서 성별이 2차, 3차 자료가 남자인 경우는 각각27.2%, 20.8% 여자인 경우는 23.4%, 18.1%가 등록은 되어 있으나, 실존하지 않는 개체들로 집계되었다(<표 2-17>). 연령의 경우는 20대와 70대 이상에서 등록은 돼 있으나, 실존하지 않는 개체들의 비율이 높은 것으로 집계된다. 20대의 경우는 학업이나 직업의 이유로 타 지역으로 이주 하였으나 법적인 이주를 하지 않기 때문에 발생하는 것으로 볼 수 있으며, 70대 이상 역시 20대와 비슷한 30%이상의 비매칭이 발생하였다. 행정지역에서 보면 2차 자료에서는 묵호동, 인제읍. 3차 자료에서는 소이면, 거류면이 높ی 나타났고 이 지역에서의 20대는 다른 지역보다 더욱 높이 나타났다. (<표 2-18>). 세대주와의 관계는 배우자의 부모, 형제자매 그 배우자, 형제자매의 자녀, 기타친인척, 기타동거인 경우에 비 매칭된 전체의 40%이상이 높은 비율인 것이 특징이다(<표 2-19>).

<표 2-17> 행정자료의 성별 비 매칭 현황

		총계			남자			여자		
		계	남은 건수	남은 비율	계	남은 건수	남은 비율	계	남은 건수	남은 비율
	전체	87,414	22,147	25.3	44,406	12,076	27.2	43,008	10,071	23.4
2차 조사 자료	범천1동	9,712	2,607	26.8	4,785	1,400	29.3	4,927	1,207	24.5
	명장1동	18,321	2,840	15.5	9,121	1,576	17.3	9,200	1,264	13.7
	기흥동	13,283	3,153	23.7	6,811	1,744	25.6	6,472	1,409	21.8
	광남동	31,772	8,070	25.4	16,344	4,468	27.3	15,428	3,602	23.3
	목호동	4,786	1,747	36.5	2,382	867	36.4	2,404	880	36.6
	인제읍	9,540	3,730	39.1	4,963	2,021	40.7	4,577	1,709	37.3
	전체	91,175	17,737	19.5	45,952	9,555	20.8	45,223	8,182	18.1
3차 조사 자료	문흥1동	20,958	3,415	16.3	10,334	1,782	17.2	10,624	1,633	15.4
	월곡1동	16,637	3,308	19.9	8,450	1,857	22	8,187	1,451	17.7
	교동	18,433	3,123	16.9	9,213	1,639	17.8	9,220	1,484	16.1
	소이면	3,423	1,349	39.4	1,778	715	40.2	1,645	634	38.5
	활천동	26,381	4,310	16.3	13,377	2,340	17.5	13,004	1,970	15.1
	거류면	5,343	2,232	41.8	2,800	1,222	43.6	2,543	1,010	39.7

<표 2-18> 행정자료의 연령 비 매칭 현황

		2차 시험 조사 지역							3차 시험 조사 지역						
		전체	범천	명장	기흥	광남	목호	인제	전체	문흥	월곡	교동	소이	활천	거류
총계	계	87,414	9,712	18,321	13,283	31,772	4,786	9,540	91,175	20,958	16,637	18,433	3,423	26,381	5,343
	남은건수	22,147	2,607	2,840	3,153	8,070	1,747	3,730	17,737	3,415	3,308	3,123	1,349	4,310	2,232
	남은비율	25.3	26.8	15.5	23.7	25.4	36.5	39.1	19.5	16.3	19.9	16.9	39.4	16.3	41.8
1~9세	계	9,237	488	1,045	2,071	4,272	297	1,064	8,803	1,866	1,669	2,196	153	2,619	300
	남은건수	1,190	97	111	217	518	53	194	943	208	225	179	35	192	104
	남은비율	12.9	19.9	10.6	10.5	12.1	17.8	18.2	10.7	11.1	13.5	8.2	22.9	7.3	34.7
10~19세	계	10,253	872	2,576	1,371	3,852	379	1,203	16,316	4,213	3,359	2,992	298	4,969	485
	남은건수	1,659	161	221	247	619	103	308	1,644	357	403	252	84	402	146
	남은비율	16.2	18.5	8.6	18	16.1	27.2	25.6	10.1	8.5	12	8.4	28.2	8.1	30.1
20~29세	계	12,590	1,861	3,167	1,687	4,044	561	1,270	12,396	3,256	2,491	2,147	363	3,448	691
	남은건수	4,352	565	775	589	1,291	272	860	4,478	981	824	912	258	1,054	449
	남은비율	34.6	30.4	24.5	34.9	31.9	48.5	67.7	36.1	30.1	33.1	42.5	71.1	30.6	65
30~39세	계	16,394	1,588	2,371	3,471	6,805	692	1,467	14,578	2,898	2,734	3,357	307	4,515	767
	남은건수	4,120	520	490	737	1,594	278	501	3,065	564	641	517	177	783	383
	남은비율	25.1	32.7	20.7	21.2	23.4	40.2	34.2	21	19.5	23.4	15.4	57.7	17.3	49.9
40~49세	계	14,382	1,462	3,381	1,906	5,410	607	1,616	18,026	4,038	3,374	3,450	475	5,936	753
	남은건수	3,746	385	519	505	1,484	274	579	2,935	504	625	438	201	810	357
	남은비율	26	26.3	15.4	26.5	27.4	45.1	35.8	16.3	12.5	18.5	12.7	42.3	13.6	47.4
50~59세	계	11,713	1,545	3,194	1,206	3,819	780	1,169	10,460	2,573	1,734	2,155	525	2,767	706
	남은건수	3,348	443	386	381	1,329	287	522	2,012	325	311	369	185	552	270
	남은비율	28.6	28.7	12.1	31.6	34.8	36.8	44.7	19.2	12.6	17.9	17.1	35.2	19.9	38.2
60~69세	계	7,715	1,081	1,672	918	2,297	813	934	5,541	1,200	676	1,160	546	1,263	696
	남은건수	2,016	238	190	259	741	231	357	1,153	185	131	203	161	272	201
	남은비율	26.1	22	11.4	28.2	32.3	28.4	38.2	20.8	15.4	19.4	17.5	29.5	21.5	28.9
70~79세	계	3,766	638	653	482	928	493	572	3,726	654	431	719	551	672	699
	남은건수	1,181	140	89	150	347	173	282	1,019	184	104	173	166	166	226
	남은비율	31.4	21.9	13.6	31.1	37.4	35.1	49.3	27.3	28.1	24.1	24.1	30.1	24.7	32.3
80~89세	계	1,234	165	243	154	307	151	214	1,173	222	143	216	186	175	231
	남은건수	468	51	52	55	125	72	113	424	88	36	66	76	72	86
	남은비율	37.9	30.9	21.4	35.7	40.7	47.7	52.8	36.1	39.6	25.2	30.6	40.9	41.1	37.2
90세 이상	계	130	12	19	17	38	13	31	156	38	26	41	19	17	15
	남은건수	67	7	7	13	22	4	14	64	19	8	14	6	7	10
	남은비율	51.5	58.3	36.8	76.5	57.9	30.8	45.2	41	50	30.8	34.1	31.6	41.2	66.7

<표 2-19> 행정자료의 세대주(가구주) 비매칭 현황

	2차 시험 조사 자료							3차 시험 조사 자료						
	전체	법천	명장	기흥	광남	묵호	인제	전체	문흥	월곡	교동	소이	활천	거류
계	87,414	9,712	18,321	13,283	31,772	4,786	9,540	91,175	20,958	16,637	18,433	3,423	26,381	5,343
남은건수	22,147	2,607	2,840	3,153	8,070	1,747	3,730	17,737	3,415	3,308	3,123	1,349	4,310	2,232
남은비율	25.3	26.8	15.5	23.7	25.4	36.5	39.1	19.5	16.3	19.9	16.9	39.4	16.3	41.8
1 계	33,755	4,604	6,357	5,320	11,532	2,210	3,732	31,469	6,631	5,767	6,485	1,527	8,670	2,389
남은건수	10,004	1,273	944	1,594	3,736	814	1,643	6,537	1,007	1,356	1,087	568	1,585	934
남은비율	29.6	27.6	14.8	30	32.4	36.8	44	20.8	15.2	23.5	16.8	37.2	18.3	39.1
2 계	17,350	1,579	3,699	2,625	6,614	907	1,926	16,875	3,689	2,778	3,654	689	5,112	953
남은건수	3,012	271	293	375	1,298	224	551	1,800	297	301	317	180	423	282
남은비율	17.4	17.2	7.9	14.3	19.6	24.7	28.6	10.7	8.1	10.8	8.7	26.1	8.3	29.6
3 계	29,883	2,758	7,011	4,362	11,445	1,213	3,094	35,438	8,852	6,854	7,005	835	10,453	1,439
남은건수	6,697	743	1,208	808	2,246	515	1,177	6,539	1,399	1,207	1,310	430	1,472	721
남은비율	22.4	26.9	17.2	18.5	19.6	42.5	38	18.5	15.8	17.6	18.7	51.5	14.1	50.1
4 계	588	54	127	76	187	51	93	1,079	254	223	210	59	243	90
남은건수	168	17	31	23	48	22	27	240	48	49	29	24	48	42
남은비율	28.6	31.5	24.4	30.3	25.7	43.1	29	22.2	18.9	22.0	13.8	40.7	19.8	46.7
5 계	2,095	200	399	373	766	102	255	2,065	491	316	470	109	556	123
남은건수	707	57	95	115	273	37	130	786	223	113	149	47	198	56
남은비율	33.7	28.5	23.8	30.8	35.6	36.3	51	38.1	45.4	35.8	31.7	43.1	35.6	45.5
6 계	281	23	34	43	141	9	31	383	146	53	62	5	102	15
남은건수	114	9	13	19	43	5	25	196	74	27	33	3	49	10
남은비율	40.6	39.1	38.2	44.2	30.5	55.6	80.6	51.2	50.7	50.9	53.2	60.0	48.0	66.7
7 계	1,507	169	322	158	422	189	247	1,105	196	142	241	124	259	143
남은건수	452	76	81	51	103	68	73	309	46	36	51	50	53	73
남은비율	30	45	25.2	32.3	24.4	36	29.6	28.0	23.5	25.4	21.2	40.3	20.5	51.0
8 계	2						2	9			5		2	2
남은건수	1						1	1						1
남은비율	50						50	11.1						50.0
9 계	64	7	8	23	19	4	3	42	6	10	13	1	10	2
남은건수	25	4	1	12	8			18	3	2	7		6	
남은비율	39.1	57.1	12.5	52.2	42.1			42.9	50.0	20.0	53.8		60.0	
10 계	1,008	176	207	152	350	50	73	1,032	318	209	110	28	325	42
남은건수	497	80	99	81	155	34	48	486	125	95	55	19	162	30
남은비율	49.3	45.5	47.8	53.3	44.3	68	65.8	47.1	39.3	45.5	50.0	67.9	49.8	71.4
11 계	309	27	49	68	125	13	27	381	114	77	44	3	127	16
남은건수	168	13	27	35	65	8	20	186	56	31	22	1	66	10
남은비율	54.4	48.1	55.1	51.5	52	61.5	74.1	48.8	49.1	40.3	50.0	33.3	52.0	62.5
12 계	11	1	1	6	3			13	4	5	1		3	
남은건수	3		1	2				5	2	1			2	
남은비율	27.3		100	33.3				38.5	50.0	20.0			66.7	
13 계	44	3	8	6	22	5		532	157	78	59	11	205	22
남은건수	31	3	5	3	17	3		256	85	32	31	7	84	17
남은비율	70.5	100	62.5	50	77.3	60		48.1	54.1	41.0	52.5	63.6	41.0	77.3
14 계	517	111	99	71	146	33	57	752	100	125	74	32	314	107
남은건수	268	61	42	35	78	17	35	378	50	58	32	20	162	56
남은비율	51.8	55	42.4	49.3	53.4	51.5	61.4	50.3	50.0	46.4	43.2	62.5	51.6	52.3

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:중손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13: 기타 친인척, *14:기타동거인

나. 비 매칭 시험조사 자료의 항목별 특성분석

매칭에 이용된 제2차 시험조사 자료는 78,359건이며, 이 중 65,267건이 매칭되어 13,092건이 남게 되었다. 매칭에 이용된 제3차 시험조사 자료는 91,365건이며, 73,438건이 매칭되어 17,927건이 남게 되었다.

매칭에 이용된 시험조사 자료에서 성별이 2차, 3차 자료가 남자인 경우는 각각 16.4%, 20.4% 여자인 경우는 16.8%, 18.8%가 실존하지만 등록되지 않은 개체들로 집계되었다(<표 2-20>). 연령의 경우는 행정자료에서와 같이 20대와 70대 이상에서 실존하지만 등록되지 않은 개체들의 비율이 20.8%, 24.1%로 높은 것으로 집계되었다. 이는 등록되었지만 실존하지 않는 개체들인 행정자료의 비매칭을 34.6%, 31.4%보다는 낮지만 주목할 만한 특징이라고 할 수 있다. 비율보다는 낮은 것으로 나타났다(<표 2-21>). 세대주와의 관계는 기타 친인척, 기타 동거인 경우에 실존하지만 등록되지 않은 개체들의 비율이 높은 것으로 집계되었다(<표 2-22>).

<표 2-20> 시험조사 자료의 성별 비매칭 현황

		총계			남자			여자			미상		
		계	남은 건수	남은 비율	계	남은 건수	남은 비율	계	남은 건수	남은 비율	계	남은 건수	남은 비율
	전체	78,359	13,092	16.7	38,576	6,316	16.4	39,660	6,663	16.8	123	113	91.9
2차 조사 자료	범천1동	9,399	2,294	24.4	4,387	1,018	23.2	5,010	1,274	25.4	2	2	100
	명장1동	16,924	1,443	8.5	8,180	649	7.9	8,696	746	8.6	48	48	100
	기흥동	11,925	1,795	15.1	6,016	958	15.9	5,908	837	14.2	1		
	광남동	28,709	5,007	17.4	14,316	2,467	17.2	14,344	2,498	17.4	49	42	85.7
	목호동	3,971	932	23.5	1,922	412	21.4	2,044	515	25.2	5	5	100
	인제읍	7,431	1,621	21.8	3,755	812	21.6	3,658	793	21.7	18	16	88.9
	전체	91,327	17,899	19.6	45,688	9,329	20.4	45,639	8,570	18.8			
3차 조사 자료	문흥1동	21,566	4,028	18.7	10,620	2,064	19.4	10,946	1,964	17.9			
	월곡1동	16,345	3,017	18.5	8,120	1,543	19	8,225	1,474	17.9			
	교동	17,703	2,393	13.5	8,790	1,225	13.9	8,913	1,168	13.1			
	소이면	3,176	1,102	34.7	1,653	588	35.6	1,523	514	33.7			
	활천동	27,701	5,630	20.3	13,911	2,886	20.7	13,790	2,744	19.9			
	거류면	4,836	1,729	35.8	2,594	1,023	39.4	2,242	706	31.5			

<표 2-21> 시험조사 자료의 연령 비매칭 현황

		2차 시험 조사 지역							3차 시험 조사 지역						
		전체	법천	명장	기흥	광남	묵호	인제	전체	문흥	월곡	교동	소이	활천	거류
총계	계	78,359	9,399	16,924	11,925	28,709	3,971	7,431	91,365	21,575	16,361	17,707	3,176	27,704	4,842
	남은건수	13,092	2,294	1,443	1,795	5,007	932	1,621	17,927	4,032	3,032	2,397	1,102	5,633	1,731
	남은비율	16.7	24.4	8.5	15.1	17.4	23.5	21.8	19.6	18.7	18.5	13.5	34.7	20.3	35.7
1~9세	계	7,972	386	942	1,940	3,626	245	833	8,920	1,938	1,638	2,264	156	2,642	282
	남은건수	999	72	173	255	368	30	101	1,079	286	198	253	38	217	87
	남은비율	12.5	18.7	18.4	13.1	10.1	12.2	12.1	12.1	14.8	12.1	11.2	24.4	8.2	30.9
10~19세	계	9,486	779	2,347	1,276	3,774	313	997	16,089	4,225	3,240	2,951	241	5,009	423
	남은건수	736	83	107	87	334	48	77	1,400	364	278	205	27	443	83
	남은비율	7.8	10.7	4.6	6.8	8.9	15.3	7.7	8.7	8.6	8.6	6.9	11.2	8.8	19.6
20~29세	계	9,614	1,914	2,542	1,217	3,113	323	505	12,213	2,893	2,235	1,495	179	4,821	590
	남은건수	2,003	760	184	290	593	65	111	4,301	617	568	263	74	2,429	350
	남은비율	20.8	39.7	7.2	23.8	19	20.1	22	35.2	21.3	25.4	17.6	41.3	50.4	59.3
30~39세	계	14,475	1,676	2,070	3,087	5,961	510	1,171	14,446	3,026	2,678	3,257	240	4,505	740
	남은건수	2,385	528	165	372	995	99	226	3,029	716	607	435	112	801	358
	남은비율	16.5	31.5	8	12.1	16.7	19.4	19.3	21.0	23.7	22.7	13.4	46.7	17.8	48.4
40~49세	계	12,801	1,325	2,915	1,756	5,052	459	1,294	18,350	4,431	3,445	3,479	458	5,920	617
	남은건수	2,133	301	263	233	918	125	293	3,309	913	699	482	187	800	228
	남은비율	16.7	22.7	9	13.3	18.2	27.2	22.6	18.0	20.6	20.3	13.9	40.8	13.5	37.0
50~59세	계	11,127	1,380	3,259	1,139	3,629	646	1,074	10,769	2,922	1,791	2,182	559	2,716	599
	남은건수	2,088	226	269	243	872	167	311	2,229	654	352	373	218	471	161
	남은비율	18.8	16.4	8.3	21.3	24	25.9	29	20.7	22.4	19.7	17.1	39.0	17.3	26.9
60~69세	계	7,359	1,076	1,796	829	2,163	749	746	5,509	1,258	687	1,108	540	1,249	667
	남은건수	1,398	176	151	169	503	182	217	1,137	239	150	160	157	260	171
	남은비율	19	16.4	8.4	20.4	23.3	24.3	29.1	20.6	19.0	21.8	14.4	29.1	20.8	25.6
70~79세	계	4,097	665	763	505	1,025	554	585	3,756	652	461	717	586	665	675
	남은건수	988	108	89	111	311	163	206	1,006	176	128	153	198	150	201
	남은비율	24.1	16.2	11.7	22	30.3	29.4	35.2	26.8	27.0	27.8	21.3	33.8	22.6	29.8
80~89세	계	1,287	185	260	157	327	158	200	1,154	201	152	213	198	154	236
	남은건수	328	37	37	30	99	52	73	374	59	37	59	86	49	84
	남은비율	25.5	20	14.2	19.1	30.3	32.9	36.5	32.4	29.4	24.3	27.7	43.4	31.8	35.6
90세 이상	계	141	13	30	19	39	14	26	159	29	34	41	19	23	13
	남은건수	34	3	5	5	14	1	6	63	8	15	14	5	13	8
	남은비율	24.1	23.1	16.7	26.3	35.9	7.1	23.1	39.6	27.6	44.1	34.1	26.3	56.5	61.5

<표 2-22> 시험조사 자료의 세대주 비매칭 현황

	2차 시험 조사 지역							3차 시험 조사 지역						
	전체	범천	명장	기흥	광남	목호	인제	전체	문흥	월곡	교동	소이	활천	거류
계	78,190	9,386	16,862	11,907	28,661	3,966	7,408	89,712	21,535	16,261	17,681	3,132	26,295	4,808
남은건수	12,930	2,281	1,381	1,778	4,964	927	1,599	16,280	3,992	2,933	2,372	1,060	4,226	1,697
남은비율	16.5	24.3	8.2	14.9	17.3	23.4	21.6	18.1	18.5	18	13.4	33.8	16.1	35.3
계	27,749	4,685	5,837	3,948	8,799	1,773	2,707	30,582	6,689	5,405	6,018	1,397	9,001	2,072
1 남은건수	5,601	1,458	578	584	1,775	462	744	7,157	1,586	1,311	950	489	2,122	699
남은비율	20.2	31.1	9.9	14.8	20.2	26.1	27.5	23.4	23.7	24.3	15.8	35	23.6	33.7
계	17,906	1,691	3,834	2,754	6,783	947	1,897	20,335	4,967	3,445	4,303	781	5,798	1,041
2 남은건수	2,939	312	327	391	1,262	218	429	3,502	979	608	596	252	749	318
남은비율	16.4	18.5	8.5	14.2	18.6	23	22.6	17.2	19.7	17.6	13.9	32.3	12.9	30.5
계	25,967	2,254	6,203	4,007	10,472	849	2,182	32,701	8,770	6,413	6,312	484	9,794	928
3 남은건수	2,463	243	317	376	1,177	138	212	3,301	1,049	645	558	75	766	208
남은비율	9.5	10.8	5.1	9.4	11.2	16.3	9.7	10.1	12	10.1	8.8	15.5	7.8	22.4
계	560	42	114	76	238	42	48	288	48	45	58	23	66	48
4 남은건수	136	7	19	20	64	13	13	97	17	10	18	11	20	21
남은비율	24.3	16.7	16.7	26.3	26.9	31	27.1	33.7	35.4	22.2	31	47.8	30.3	43.8
계	2,151	172	337	358	911	134	239	2,077	385	309	501	183	574	125
5 남은건수	614	28	64	99	308	46	69	605	121	103	113	78	151	39
남은비율	28.5	16.3	19	27.7	33.8	34.3	28.9	29.1	31.4	33.3	22.6	42.6	26.3	31.2
계	280	23	30	42	154	12	19	295	80	46	54	11	99	5
6 남은건수	75	5	3	9	48	2	8	98	28	14	18	7	29	2
남은비율	26.8	21.7	10	21.4	31.2	16.7	42.1	33.2	35	30.4	33.3	63.6	29.3	40
계	1,357	141	286	160	487	131	152	935	177	138	196	87	221	116
7 남은건수	188	18	27	24	85	18	16	184	47	20	33	21	34	29
남은비율	13.9	12.8	9.4	15	17.5	13.7	10.5	19.7	26.6	14.5	16.8	24.1	15.4	25
계	25	2	2	3	6	6	6	9	3	1	1	4	1	
8 남은건수	6			1	2	2	1	4		3	1			
남은비율	24			33.3	33.3	33.3	16.7	44.4		100.0	100.0			
계	53	7	7	11	18	5	5	46	7	17	8	2	11	1
9 남은건수	16	2	1	3	9	1	1	10	1	4	2	1	2	
남은비율	30.2	28.6	14.3	27.3	50		20	21.7	14.3	23.5	25	50	18.2	
계	897	183	110	161	378	24	41	870	249	213	100	26	264	18
10 남은건수	289	98	14	41	105	10	21	314	97	81	34	14	76	12
남은비율	32.2	53.6	12.7	25.5	27.8	41.7	51.2	36.1	39	38	34	53.8	28.8	66.7
계	210	28	30	37	89	16	10	240	46	61	39	3	78	13
11 남은건수	51	9	12	8	13	5	4	66	12	20	10	2	17	5
남은비율	24.3	32.1	40	21.6	14.6	31.3	40	27.5	26.1	32.8	25.6	66.7	21.8	38.5
계	19	1	3	5	10			42	13	8	13		7	1
12 남은건수	4			2	2			10	5	3	1		1	
남은비율	21.1			40	20			23.8	38.5	37.5	7.7		14.3	
계	348	32	39	65	177	12	23	276	69	56	39	6	93	13
13 남은건수	111	10	11	20	49	5	16	113	29	32	18	3	27	4
남은비율	31.9	31.3	28.2	30.8	27.7	41.7	69.6	40.9	42	57.1	46.2	50	29	30.8
계	668	125	30	280	139	15	79	1,016	35	102	39	129	285	426
14 남은건수	437	91	8	200	65	8	65	819	21	79	20	107	232	360
남은비율	65.4	72.8	26.7	71.4	46.8	53.3	82.3	80.6	60	77.5	51.3	82.9	81.4	84.5

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:증손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

2. 주택

가. 비 매칭 행정자료의 항목별 특성분석

제2차 시험조사 자료와 매칭되는 건축물대장 자료는 25,130건이며, 이 중 17,890건이 매칭되어 7,240건이 남게 되었다. 제3차 시험조사 자료와 매칭되

는 건축물대장 자료는 24,448건이며, 24,433건이 매칭되어 2,462건이 남게 되었다. 시험조사 자료와 매칭되고 남은 건축물대장 자료는 주택종류가 아파트인 경우를 제외한 경우, 면적은 46m²미만, 건축연도는 90년대 전인 경우에 등록은 돼있으나 실존하지 않은 비율이 높게 나타나고 있다(<표 2-23>).

<표 2-23> 행정자료의 비매칭 주택항목특성별 현황

		2차시험조사자료							3차시험조사자료								
		전체	법천동	명장동	기흥동	광남동	목호동	인제읍	전체	문흥동	월곡동	교동	소이면	활천동	거류면		
주택 종류	전체	계	25,130	3,007	4,129	4,271	8,536	2,473	2,714	26,895	5,446	4,043	5,935	1,694	7,369	2,408	
		남은건수	7,240	704	357	752	2,836	1,282	1,309	2,462	156	145	216	606	498	841	
		남은비율	28.8	23.4	8.6	17.6	33.2	51.8	48.2	9.2	2.9	3.6	3.6	35.8	6.8	34.9	
	일반 주택	계	6,545	703	1,131	294	776	1,758	1,883	4,350	30	355	850	1,124	205	1,786	
		남은건수	3,247	389	62	166	530	1,004	1,096	1,625	12	8	179	577	59	790	
		남은비율	49.6	55.3	5.5	56.5	68.3	57.1	58.2	37.4	40.0	2.3	21.1	51.3	28.8	44.2	
	단독 주택	계	740	143	304	31	91	76	95	1,382	519	575	133	25	98	32	
		남은건수	237	44	15	23	68	37	50	336	138	122	6	10	45	15	
		남은비율	32.0	30.8	4.9	74.2	74.7	48.7	52.6	24.3	26.6	21.2	4.5	40.0	45.9	46.9	
	APT	계	11,957	1,868	889	3,468	4,748	562	422	18,306	4,882	3,058	4,396	456	5,029	485	
		남은건수	835	145	16	313	69	211	81	45	1	2	6	1	24	11	
		남은비율	7.0	7.8	1.8	9.0	1.5	37.5	19.2	0.2	0.0	0.1	0.1	0.2	0.5	2.3	
	집합 주택	계	4,944	258	920	476	2,921	77	292	2,857	15	55	556	89	2,037	105	
		남은건수	2,827	123	175	248	2,169	30	82	456	5	13	25	18	370	25	
		남은비율	57.2	47.7	19.0	52.1	74.3	39.0	28.1	16.0	33.3	23.6	4.5	20.2	18.2	23.8	
그외 주택	계	944	35	885	2			22									
	남은건수	94	3	89	2			0									
	남은비율	10.0	8.6	10.1	100.0			0.0									
면적	전체	계	25,130	3,007	4,129	4,271	8,536	2,473	2,714	26,895	5,446	4,043	5,935	1,694	7,369	2,408	
		남은건수	7,240	704	357	752	2,836	1,282	1,309	2,462	156	145	216	606	498	841	
		남은비율	28.8	23.4	8.6	17.6	33.2	51.8	48.2	9.2	2.9	3.6	3.6	35.8	6.8	34.9	
	46m ² 미만	계	4,629	1,527	630	310	308	1,330	524	1,763	8	4	585	264	429	473	
		남은건수	2,221	455	138	88	262	935	343	510	5	3	67	139	58	238	
		남은비율	48.0	29.8	21.9	28.4	85.1	70.3	65.5	28.9	62.5	75.0	11.5	52.7	13.5	50.3	
	96m ² 미만	계	16,108	1,227	2,298	3,480	6,498	1,007	1,598	20,765	4,116	3,189	4,526	1,277	5,981	1,676	
		남은건수	3,927	168	179	433	2,193	302	652	1,423	11	31	112	394	352	523	
		남은비율	24.4	13.7	7.8	12.4	33.7	30.0	40.8	6.9	0.3	1.0	2.5	30.9	5.9	31.2	
	96m ² 이상	계	4,393	253	1,201	481	1,730	136	592	4,367	1,322	850	824	153	959	259	
		남은건수	1,092	81	40	231	381	45	314	529	140	111	37	73	88	80	
		남은비율	24.9	32.0	3.3	48.0	22.0	33.1	53.0	12.1	10.6	13.1	4.5	47.7	9.2	30.9	
	건축 연도	전체	계	25,130	3,007	4,129	4,271	8,536	2,473	2,714	26,895	5,446	4,043	5,935	1,694	7,369	2,408
			남은건수	7,240	704	357	752	2,836	1,282	1,309	2,462	156	145	216	606	498	841
			남은비율	28.8	23.4	8.6	17.6	33.2	51.8	48.2	9.2	2.9	3.6	3.6	35.8	6.8	34.9
2000 년대		계	11,278	1,340	236	2,413	5,846	466	977	3,921	69	98	2,018	77	973	686	
		남은건수	2,462	119	14	264	1,570	84	411	323	42	71	19	40	80	71	
		남은비율	21.8	8.9	5.9	10.9	26.9	18.0	42.1	8.2	60.9	72.4	0.9	51.9	8.2	10.3	
1990 년대		계	6,082	84	1,094	1,628	2,294	430	552	18,727	5,363	3,120	3,255	711	5,963	315	
		남은건수	1,798	13	39	345	970	237	194	787	107	69	34	101	367	109	
		남은비율	29.6	15.5	3.6	21.2	42.3	55.1	35.1	4.2	2.0	2.2	1.0	14.2	6.2	34.6	
1990 이전		계	5,661	1,451	2,758	175	175	890	212	4,027	14	825	662	906	213	1,407	
		남은건수	1,548	466	303	108	118	483	70	1,352	7	5	163	465	51	661	
		남은비율	27.3	32.1	11.0	61.7	67.4	54.3	33.0	33.6	50.0	0.6	24.6	51.3	23.9	47.0	
누락		계	2,109	132	41	55	221	687	973								
		남은건수	1,432	106	1	35	178	478	634								
		남은비율	67.9	80.3	2.4	63.6	80.5	69.6	65.2								

나. 비 매칭 시험조사 자료의 항목별 특성분석

제2차 시험조사 자료와 매칭되는 건축물대장 자료는 24,448건이며, 이 중

17,890건이 매칭되어 6,558건이 남게 되었다. 제3차 시험조사 자료와 매칭되는 건축물대장 자료는 26,970건이며, 24,433건이 매칭되어 2,537건이 남게 되었다. 건축물대장 자료와 매칭되고 남은 시험조사 자료는 주택종류가 아파트인 경우를 제외한 경우, 면적은 46m²미만, 건축연도는 90년대 전인 경우에 실존하나 등록되지 않은 비율이 높게 나타나고 있다(<표 2-24>). 단, 제3차 시험조사 자료에서 실존하나 등록되지 않은 비율은 다소 낮게 나타나고 있다.

<표 2-24> 조사자료의 비 매칭 주택항목특성별 현황

		2차시험조사자료							3차시험조사자료								
		전체	범천동	명장동	기흥동	광남동	북호동	인계읍	전체	문흥동	월곡동	교동	소이면	활천동	거류면		
주택 종류	전체	계	24,448	3,786	4,101	3,877	8,281	1,750	2,653	26,970	5,470	3,983	5,926	1,805	7,444	2,342	
		남은건수	6,558	1,483	329	358	2,581	559	1,248	2,537	180	85	207	717	573	775	
		남은비율	26.8	39.2	8.0	9.2	31.2	31.9	47.0	9.4	3.3	2.1	3.5	39.7	7.7	33.1	
	일반 주택	계	3,931	320	380	227	546	1,161	1,297	4,238	44	395	741	1,250	174	1,634	
		남은건수	1,725	205	29	111	349	416	615	1,552	5	14	163	679	48	643	
		남은비율	43.9	64.1	7.6	48.9	63.9	35.8	47.4	36.6	11.4	3.5	22.0	54.3	27.6	39.4	
	단독 주택	계	2,352	692	1,069	29	213	89	260	1,617	471	497	337	32	215	65	
		남은건수	723	360	29	15	134	33	152	358	122	68	35	21	77	35	
		남은비율	30.7	52.0	2.7	51.7	62.9	37.1	58.5	22.1	25.9	13.7	10.4	65.6	35.8	53.8	
	APT	계	11,786	1,806	1,208	3,182	4,682	411	497	18,657	4,882	3,056	4,729	454	4,991	545	
		남은건수	343	84	5	32	5	60	157	106	1	0	1	0	34	70	
		남은비율	2.9	4.7	0.4	1.0	0.1	14.6	31.6	0.6	0.0	0.0	0.0	0.0	0.7	12.8	
	집합 주택	계	5,256	233	1,404	423	2,714	61	421	2,400	60	31	115	63	2,063	68	
		남은건수	2,775	100	256	191	2,006	26	196	495	47	3	6	13	413	13	
		남은비율	52.8	42.9	18.2	45.2	73.9	42.6	46.6	20.6	78.3	9.7	5.2	20.6	20.0	19.1	
	그외 주택	계	1,123	735	40	16	126	28	178	58	13	4	4	6	1	30	
		남은건수	992	734	10	9	87	24	128	26	5	0	2	4	1	14	
		남은비율	88.3	99.9	25.0	56.3	69.0	85.7	71.9	44.8	38.5	0.0	50.0	66.7	100.0	46.7	
면적	전체	계	24,448	3,786	4,101	3,877	8,281	1,750	2,653	26,970	5,470	3,983	5,926	1,805	7,444	2,342	
		남은건수	6,558	1,483	329	358	2,581	559	1,248	2,537	180	85	207	717	573	775	
		남은비율	26.8	39.2	8.0	9.2	31.2	31.9	47.0	9.4	3.3	2.1	3.5	39.7	7.7	33.1	
	46m ² 미만	계	3,578	1,929	564	250	223	216	396	1,144	36	4	494	61	363	186	
		남은건수	1,596	977	57	27	158	102	275	201	31	1	25	42	28	74	
		남은비율	44.6	50.6	10.1	10.8	70.9	47.2	69.4	17.6	86.1	25.0	5.1	68.9	7.7	39.8	
	96m ² 미만	계	15,995	1,478	2,348	3,255	6,137	1,253	1,524	20,127	4,116	3,114	4,450	1,056	5,860	1,531	
		남은건수	3,672	348	235	205	1,884	359	641	1,447	15	6	128	445	383	470	
		남은비율	23.0	23.5	10.0	6.3	30.7	28.7	42.1	7.2	0.4	0.2	2.9	42.1	6.5	30.7	
	96m ² 이상	계	4,875	379	1,189	372	1,921	281	733	4,318	1,293	804	657	256	1,042	266	
		남은건수	1,290	158	37	126	539	98	332	555	123	76	38	126	114	78	
		남은비율	26.5	41.7	3.1	33.9	28.1	34.9	45.3	13	10	9	6	49	11	29	
	건축 연도	전체	계	24,448	3,786	4,101	3,877	8,281	1,750	2,653	1,381	25	61	325	432	179	359
			남은건수	6,558	1,483	329	358	2,581	559	1,248	334	11	2	16	104	48	153
			남은비율	26.8	39.2	8.0	9.2	31.2	31.9	47.0	24.2	44.0	3.3	4.9	24.1	26.8	42.6
		2000 년대	계	11,459	1,890	236	2,234	5,655	472	972	26,970	5,470	3,983	5,926	1,805	7,444	2,342
			남은건수	2,581	659	12	82	1,391	58	379	2,537	180	85	207	717	573	775
			남은비율	22.5	34.9	5.1	3.7	24.6	12.3	39.0	9.4	3.3	2.1	3.5	39.7	7.7	33.1
1990 년대		계	6,134	168	1,127	1,469	2,261	382	727	3,951	61	83	1,761	197	1,045	804	
		남은건수	1,733	75	53	185	946	144	330	461	33	56	17	98	94	163	
		남은비율	28.3	44.6	4.7	12.6	41.8	37.7	45.4	11.7	54.1	67.5	1.0	49.7	9.0	20.3	
1990 이전		계	6,328	1,639	2,692	157	231	868	741	18,635	5,372	2,996	3,231	760	5,849	427	
		남은건수	1,866	661	252	81	152	333	387	1,006	132	21	35	278	399	141	
		남은비율	29.5	40.3	9.4	51.6	65.8	38.4	52.2	5.4	2.5	0.7	1.1	36.6	6.8	33.0	
누락		계	527	89	46	17	134	28	213	3,029	12	840	609	416	399	753	
		남은건수	378	88	12	10	92	24	152	735	4	6	137	237	33	318	
		남은비율	71.7	98.9	26.1	58.8	68.7	85.7	71.4	24.3	33.3	0.7	22.5	57.0	8.3	42.2	

제6절 결론

1장에서 얻은 매칭 결과 중 정확 매칭된 개체들을 제외한 매칭되었다고 판단된 개체들이 어떤 항목에서 어떠한 형태의 상이한 값을 갖고 있는지를 특성별, 집단별로 분석하였다. 결과, 인구부분의 S+M단계 항목별 불일치(2차 자료/3차 자료)는 성별은 0.7%/0.5%, 생년에서는 17.3%/20.4%, 세대주와의 관계에서는 9.3%/9.1%로 나타났고, W단계에서는 항목별 불일치는 2차 자료에서는 세대주와의 관계가 14.3%, 생년에서는 2차 자료에서는 95.5%로 나타났고, 3차 자료는 세대주와의 관계가 20.6%, 성별이 0.6%로 나타났다.

주택부분의 S+M단계 항목별 불일치(2차 자료/3차 자료)는 주택종류는 13.6%/5.8%, 면적은 19.6%/38.3%, 건축연도는 11%/ 11.9%로 나타났다.

분석 목적과 방향은 시험조사 자료에 포함된 개체들이 실제로 존재하고 그들에 대한 정보가 맞다고 보았을 때 행정정보의 품질에 관한 항목별 특징을 파악하여 차후 효과적으로 행정자료를 활용할 수 있도록 하고자 하였다.

한편 행정자료, 조사자료에서 매칭되지 않은 개체들 각각에 대해 항목별 분석을 실시하여 매칭되지 않은 개체들의 특성을 밝혔다.

우선 행정자료부분의 인구부분 검토 결과를 보면 2차 자료의 비매칭은 25.3%로 나타났다. 성별에서는 남자 27.2%, 여자 23.4%로 나타났고, 연령대가 높을수록 비매칭률이 다소 높게 나타났다. 세대주와의 관계는 기타 친인척과 동거인의 경우 높은 비매칭률을 보였다.

3차 자료의 행정자료의 비매칭은 19.5%로 나타났다. 성별은 남자 20.8%, 여자 18.1%로 나타났고, 연령대가 높을수록 비매칭률이 다소 높게 나타났다. 세대주와의 관계는 세대주의 배우자, 자녀, 증손자녀, 그 배우자의 경우는 비매칭률이 다소 낮았고 그 외의 경우는 비매칭률이 비교적 높게 나타났다.

이어서 행정자료부분의 주택부분 검토 결과를 보면 2차 자료는 28.8%, 3차 자료는 9.2%로 나타났고, 주택종류가 일반주택인 경우 비매칭률이 높았고, 면적에서는 면적이 작을수록 비매칭률이 높았다. 건축연도는 오래된 것일 수록 비매칭률이 높게 나타났다.

제7절 정책적 건의사항

본 장에서는 매칭되었다고 판단된 자료에서 나타나는 항목의 불일치 내용을 분석하여 행정자료의 오류유형, 조사항목에 대한 행정적 개념의 차이 등을 분석하고 특히 이러한 현상이 지역별로 어떠한 차이가 나는지를 조사하고자 하였다. 또한 매칭되지 않는다고 판단된 행정자료를 분석하여 행정자료에 포함되어 있으나 조사자료를 기준으로 할 때 해당지역에 존재하지 않는 개체들의 특성을 조사하였다. 조사결과 면단위와 동단위에서 다소간의 차이가 발생한 것으로 보이는데 연구대상지역이 12개 읍면동으로 작은 수이기 때문에 규칙성까지 찾아낼 수 없었다. 따라서 1장에서 제안한 효율적 매칭작업에 대한 연구가 수행된다면 매칭결과 역시 본 연구를 수행할 때 보다 빨리 얻을 수 있기 때문에 매칭대상지역을 확장하는 것을 검토할 필요가 있다. 그런데 이 때 중요한 이슈는 매칭대상지역을 선정하는 기준, 방법이 된다. 본 연구의 대상이었던 12개 읍면동의 경우도 이들의 선정 기준에 대한 구체적인 정보가 많지 않았기 때문에 불일치, 비(非) 매칭 특성과 지역의 특성 사이의 관계를 지역 크기인 읍, 면, 동 특성만 사용하게 되었다. 따라서 매칭대상지역을 선정하는 기준이 검토될 필요가 있는데 이는 3장에서 다루는 추정을 위한 표본조사와도 관련되어 있어서 함께 검토되어야 한다고 본다.

다른 한 가지 이슈는 불일치 항목의 참값에 대한 사후조사의 필요성이다. 본 연구에서는 처음에는 조사자료가 참값이라고 가정을 갖고 진행하였으나 진행과정에서 꼭 그렇지는 않다는 정황을 많이 발견할 수 있었다. 따라서 행정자료를 기반으로 센서스가 진행될 경우 행정자료의 항목별 값에 대한 신뢰도를 위하여 매칭된 것으로 판단된 자료 중 불일치 항목을 포함한 개체에 대한 실제조사를 통하여 확인해 볼 수 있을 것이라고 본다.

<참고>

- 비매칭 자료의 교차표

비매칭된 자료들에 대하여 항목들의 교차표를 살펴봄으로써 앞에서의 불일치 분석을 구체적으로 살펴본다.

1. 인구

가. 2차 시험조사 지역 조사자료

인구에서의 비매칭 개체들에 대하여 연령대와 성별, 가구주관계와 연령대, 성별과 가구주관계에 대하여 교차표를 그려 보았다.

연령대와 성별을 살펴보면 비매칭된 개체들은 여자가 남자에 비해 많이 보이고 있다. 하지만 남자 1-9세, 10-19세, 30-39세, 40-49세는 여자보다 남자의 비율이 높게 나타났다.

가구주 관계와 연령의 교차표를 살펴보면 비(非) 매칭 비율이 가장 높은 가구주관계14(기타 동거인)에서는 20대가 가장 많이 나타났다. 가구주 관계에서 가장 매칭이 잘 된 가구주 관계코드 3에서도 20대가 가장 많이 나타났다. 비매칭된 자료의 수가 가장 많은 가구주관계1(가구주)을 보면 50대가 가장 많고 그다음으로 40대 30대가 많이 나타났다.

가구주 관계와 성별은 가구주관계 별로 성별의 비율이 다르게 나타나고 있다. 가구주관계1의 세대주는 남자의 비율이 높고, 가구주관계2의 세대주의 배우자는 여자의 비율이 높게 나타난다. 가구주관계3(자녀)은 남자의 비율이 높으며, 가구주 관계와 연령대표를 보면 자녀 중에서는 20대에서 가장 많이 나타났다.

연령대와 성별

(N, 비매칭기준 %, 전체기준 %)

	미상	남자	여자	총합계
1~9세	109 (0.8) (0.1)	500 (3.8) (0.6)	390 (3.0) (0.5)	999 (7.6) (1.3)
10~19세		374 (2.9) (0.5)	362 (2.8) (0.5)	736 (5.6) (0.9)
20~29세	1 (0.0) (0.0)	910 (7.0) (1.2)	1,092 (8.3) (1.4)	2,003 (15.3) (2.6)
30~39세		1,318 (10.1) (1.7)	1,067 (8.2) (1.4)	2,385 (18.2) (3.0)
40~49세		1,102 (8.4) (1.4)	1,031 (7.9) (1.3)	2,133 (16.3) (2.7)
50~59세	1 (0.0) (0.0)	1,009 (7.7) (1.3)	1,078 (8.2) (1.4)	2,088 (15.9) (2.7)
60~69세	2 (0.0) (0.0)	615 (4.7) (0.8)	781 (6.0) (1.0)	1,398 (10.7) (1.8)
70~79세		382 (2.9) (0.5)	606 (4.6) (0.8)	988 (7.5) (1.3)
80~89세		98 (0.7) (0.1)	230 (1.8) (0.3)	328 (2.5) (0.4)
90세 이상		8 (0.1) (0.0)	26 (0.2) (0.0)	34 (0.3) (0.0)
총합계	113 (0.9) (0.1)	6,316 (48.2) (8.1)	6,663 (50.9) (8.5)	13,092 (100.0) (16.7)

가구주관계와 연령대

(N, 비매칭기준 %, 전체기준 %)

	1~9세	10~19세	20~29세	30~39세	40~49세	50~59세	60~69세	70~79세	80~89세	90세 이상	총합계
1	111 (0.8) (0.1)	5 (0.0) (0.0)	661 (5.0) (0.8)	1,080 (8.2) (1.4)	1,143 (8.7) (1.5)	1,154 (8.8) (1.5)	742 (5.7) (0.9)	535 (4.1) (0.7)	162 (1.2) (0.2)	8 (0.1) (0.0)	5,601 (42.8) (7.1)
2	42 (0.3) (0.1)	1 (0.0) (0.0)	170 (1.3) (0.2)	576 (4.4) (0.7)	754 (5.8) (1.0)	743 (5.7) (0.9)	438 (3.3) (0.6)	191 (1.5) (0.2)	24 (0.2) (0.0)		2,939 (22.4) (3.8)
3	521 (4.0) (0.7)	636 (4.9) (0.8)	735 (5.6) (0.9)	446 (3.4) (0.6)	96 (0.7) (0.1)	24 (0.2) (0.0)	5 (0.0) (0.0)				2,463 (18.8) (3.1)
4	3 (0.0) (0.0)		26 (0.2) (0.0)	70 (0.5) (0.1)	28 (0.2) (0.0)	7 (0.1) (0.0)	1 (0.0) (0.0)	1 (0.0) (0.0)			136 (1.0) (0.2)
5	4 (0.0) (0.0)				4 (0.0) (0.0)	93 (0.7) (0.1)	160 (1.2) (0.2)	214 (1.6) (0.3)	121 (0.9) (0.2)	18 (0.1) (0.0)	614 (4.7) (0.8)
6						10 (0.1) (0.0)	24 (0.2) (0.0)	28 (0.2) (0.0)	11 (0.1) (0.0)	2 (0.0) (0.0)	75 (0.6) (0.1)
7	104 (0.8) (0.1)	54 (0.4) (0.1)	26 (0.2) (0.0)	3 (0.0) (0.0)	1 (0.0) (0.0)						188 (1.4) (0.2)
8	3 (0.0) (0.0)	2 (0.0) (0.0)	1 (0.0) (0.0)								6 (0.0) (0.0)
9								5 (0.0) (0.0)	5 (0.0) (0.0)	6 (0.0) (0.0)	16 (0.1) (0.0)
10	2 (0.0) (0.0)	11 (0.1) (0.0)	119 (0.9) (0.2)	90 (0.7) (0.1)	45 (0.3) (0.1)	12 (0.1) (0.0)	5 (0.0) (0.0)	5 (0.0) (0.0)			289 (2.2) (0.4)
11	12 (0.1) (0.0)	11 (0.1) (0.0)	22 (0.2) (0.0)	5 (0.0) (0.0)	1 (0.0) (0.0)						51 (0.4) (0.1)
12	1 (0.0) (0.0)				1 (0.0) (0.0)	1 (0.0) (0.0)	1 (0.0) (0.0)				4 (0.0) (0.0)
13	1 (0.0) (0.0)	4 (0.0) (0.0)	36 (0.3) (0.0)	25 (0.2) (0.0)	18 (0.1) (0.0)	14 (0.1) (0.0)	10 (0.1) (0.0)		3 (0.0) (0.0)		111 (0.8) (0.1)
14	40 (0.3) (0.1)	12 (0.1) (0.0)	205 (1.6) (0.3)	88 (0.7) (0.1)	41 (0.3) (0.1)	29 (0.2) (0.0)	12 (0.1) (0.0)	8 (0.1) (0.0)	2 (0.0) (0.0)		437 (3.3) (0.6)
	155 (0.0) (0.0)		2 (0.0) (0.0)	2 (0.0) (0.0)	1 (0.0) (0.0)	1 (0.0) (0.0)		1 (0.0) (0.0)			162 (0.0) (0.0)
총합계	999 (7.6) (1.3)	736 (5.6) (0.9)	2,003 (15.3) (2.6)	2,385 (18.2) (3.0)	2,133 (16.3) (2.7)	2,088 (15.9) (2.7)	1,398 (10.7) (1.8)	988 (7.5) (1.3)	328 (2.5) (0.4)	34 (0.3) (0.0)	13,092 (100.0) (16.7)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:중손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

가구주관계와 성별

(N, 비매칭기준 %, 전체기준 %)

	미상	남자	여자	총합계
1	14	3,875	1,712	5,601
	(0.1)	(29.6)	(13.1)	(42.8)
	(0.0)	(4.9)	(2.2)	(7.1)
2		213	2,726	2,939
		(1.6)	(20.8)	(22.4)
		(0.3)	(3.5)	(3.8)
3	3	1,405	1,055	2,463
	(0.0)	(10.7)	(8.1)	(18.8)
	(0.0)	(1.8)	(1.3)	(3.1)
4		25	111	136
		(0.2)	(0.8)	(1.0)
		(0.0)	(0.1)	(0.2)
5		128	486	614
		(1.0)	(3.7)	(4.7)
		(0.2)	(0.6)	(0.8)
6		8	67	75
		(0.1)	(0.5)	(0.6)
		(0.0)	(0.1)	(0.1)
7		84	104	188
		(0.6)	(0.8)	(1.4)
		(0.1)	(0.1)	(0.2)
8	1	3	2	6
	(0.0)	(0.0)	(0.0)	(0.0)
	(0.0)	(0.0)	(0.0)	(0.0)
9		2	14	16
		(0.0)	(0.1)	(0.1)
		(0.0)	(0.0)	(0.0)
10		129	160	289
		(1.0)	(1.2)	(2.2)
		(0.2)	(0.2)	(0.4)
11		24	27	51
		(0.2)	(0.2)	(0.4)
		(0.0)	(0.0)	(0.1)
12	1	2	1	4
	(0.0)	(0.0)	(0.0)	(0.0)
	(0.0)	(0.0)	(0.0)	(0.0)
13		60	51	111
		(0.5)	(0.4)	(0.8)
		(0.1)	(0.1)	(0.1)
14		317	120	437
		(2.4)	(0.9)	(3.3)
		(0.4)	(0.2)	(0.6)
(비어있음)	94	41	27	162
	(0.7)	(0.3)	(0.2)	(1.2)
	(0.1)	(0.1)	(0.0)	(0.2)
총합계	113	6,316	6,663	13,092
	(0.9)	(48.2)	(50.9)	(100.0)
	(0.1)	(8.1)	(8.5)	(16.7)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:증손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

나. 2차 시험조사 지역 행정자료

연령대와 성별의 교차표를 보면, 시험 조사자료에서는 여자의 비매칭이 남자의 비매칭 보다 많이 나타났지만 행정자료에서는 남자의 비매칭이 더 많이 나타나고 있다. 연령대 중 비매칭 자료의 수가 가장 많은 20대를 살펴보면 남자가 2,653건(4,352건 중 61%), 여자가 1,699건(4,352건 중 39%)으로 남자의 비매칭이 많이 나타났다.

가구주 관계와 연령대를 살펴보면, 20대 행정자료 전체 인구 중에서 34.6%가 매칭이 되지 않았다. 매칭되지 않은 20대 인구는 가구주와의 관계가 3인 경우가 가장 많이 나타났으며 그 다음으로는 가구주관계가 1인 경우가 많이 나타났다. 가구주관계가 1인 경우는 40대, 50대, 30대가 많이 나타났다.

가구주관계와 성별은 시험 조사자료와 마찬가지로 가구주관계 별로 성별에 차이를 보이고 있다. 가구주관계가 3인 자녀를 살펴보면 자녀 중 남자는 4,027건(6,697건 중 60.1%), 여자는 2,670건(6,697건 중 39.9%)으로 자녀는 남자의 비율이 많이 보이고 있다.

연령대와 성별

(N, 비매칭기준 %, 전체기준 %)

	남자	여자	총합계
1~9세	638	552	1,190
	(2.9)	(2.5)	(5.4)
	(0.7)	(0.6)	(1.4)
10~19세	866	793	1,659
	(3.9)	(3.6)	(7.5)
	(1.0)	(0.9)	(1.9)
20~29세	2,653	1,699	4,352
	(12.0)	(7.7)	(19.7)
	(3.0)	(1.9)	(5.0)
30~39세	2,447	1,673	4,120
	(11.0)	(7.6)	(18.6)
	(2.8)	(1.9)	(4.7)
40~49세	2,113	1,633	3,746
	(9.5)	(7.4)	(16.9)
	(2.4)	(1.9)	(4.3)
50~59세	1,780	1,568	3,348
	(8.0)	(7.1)	(15.1)
	(2.0)	(1.8)	(3.8)
60~69세	979	1,037	2,016
	(4.4)	(4.7)	(9.1)
	(1.1)	(1.2)	(2.3)
70~79세	460	721	1,181
	(2.1)	(3.3)	(5.3)
	(0.5)	(0.8)	(1.4)
80~89세	125	343	468
	(0.6)	(1.5)	(2.1)
	(0.1)	(0.4)	(0.5)
90세 이상	15	52	67
	(0.1)	(0.2)	(0.3)
	(0.0)	(0.1)	(0.1)
총합계	12,076	10,071	22,147
	(54.5)	(45.5)	(100.0)
	(13.8)	(11.5)	(25.3)

가구주관계와 연령대

(N, 비매칭기준 %, 전체기준 %)

	1~9세	10~19세	20~29세	30~39세	40~49세	50~59세	60~69세	70~79세	80~89세	90세 이상	총합계
1	2 (0.0) (0.0)	16 (0.1) (0.0)	973 (4.4) (1.1)	2,080 (9.4) (2.4)	2,370 (10.7) (2.7)	2,251 (10.2) (2.6)	1,319 (6.0) (1.5)	713 (3.2) (0.8)	253 (1.1) (0.3)	27 (0.1) (0.0)	10,004 (45.2) (11.4)
2		1 (0.0) (0.0)	94 (0.4) (0.1)	608 (2.7) (0.7)	854 (3.9) (1.0)	821 (3.7) (0.9)	437 (2.0) (0.5)	175 (0.8) (0.2)	20 (0.1) (0.0)	2 (0.0) (0.0)	3,012 (13.6) (3.4)
3	958 (4.3) (1.1)	1,408 (6.4) (1.6)	2,882 (13.0) (3.3)	1,071 (4.8) (1.2)	305 (1.4) (0.3)	68 (0.3) (0.1)	5 (0.0) (0.0)				6,697 (30.2) (7.7)
4			17 (0.1) (0.0)	80 (0.4) (0.1)	52 (0.2) (0.1)	19 (0.1) (0.0)					168 (0.8) (0.2)
5			1 (0.0) (0.0)	1 (0.0) (0.0)	14 (0.1) (0.0)	84 (0.4) (0.1)	187 (0.8) (0.2)	238 (1.1) (0.3)	156 (0.7) (0.2)	26 (0.1) (0.0)	707 (3.2) (0.8)
6						18 (0.1) (0.0)	36 (0.2) (0.0)	33 (0.1) (0.0)	22 (0.1) (0.0)	5 (0.0) (0.0)	114 (0.5) (0.1)
7	170 (0.8) (0.2)	147 (0.7) (0.2)	117 (0.5) (0.1)	17 (0.1) (0.0)	1 (0.0) (0.0)						452 (2.0) (0.5)
8	1 (0.0) (0.0)										1 (0.0) (0.0)
9								6 (0.0) (0.0)	13 (0.1) (0.0)	6 (0.0) (0.0)	25 (0.1) (0.0)
10	2 (0.0) (0.0)	17 (0.1) (0.0)	139 (0.6) (0.2)	184 (0.8) (0.2)	93 (0.4) (0.1)	42 (0.2) (0.0)	16 (0.1) (0.0)	4 (0.0) (0.0)			497 (2.2) (0.6)
11	40 (0.2) (0.0)	45 (0.2) (0.1)	67 (0.3) (0.1)	13 (0.1) (0.0)	3 (0.0) (0.0)						168 (0.8) (0.2)
12				1 (0.0) (0.0)	1 (0.0) (0.0)	1 (0.0) (0.0)					3 (0.0) (0.0)
13	6 (0.0) (0.0)	3 (0.0) (0.0)	8 (0.0) (0.0)	5 (0.0) (0.0)	1 (0.0) (0.0)	5 (0.0) (0.0)	1 (0.0) (0.0)		1 (0.0) (0.0)	1 (0.0) (0.0)	31 (0.1) (0.0)
14	11 (0.0) (0.0)	22 (0.1) (0.0)	54 (0.2) (0.1)	60 (0.3) (0.1)	52 (0.2) (0.1)	39 (0.2) (0.0)	15 (0.1) (0.0)	12 (0.1) (0.0)	3 (0.0) (0.0)		268 (1.2) (0.3)
총합계	1,190 (5.4) (1.4)	1,659 (7.5) (1.9)	4,352 (19.7) (5.0)	4,120 (18.6) (4.7)	3,746 (16.9) (4.3)	3,348 (15.1) (3.8)	2,016 (9.1) (2.3)	1,181 (5.3) (1.4)	468 (2.1) (0.5)	67 (0.3) (0.1)	22,147 (100.0) (25.3)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:중손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

가구주관계와 성별

(N, 비매칭기준 %, 전체기준 %)

	남자	여자	총합계
1	6,968 (31.5) (8.0)	3,036 (13.7) (3.5)	10,004 (45.2) (11.4)
2	94 (0.4) (0.1)	2,918 (13.2) (3.3)	3,012 (13.6) (3.4)
3	4,027 (18.2) (4.6)	2,670 (12.1) (3.1)	6,697 (30.2) (7.7)
4	31 (0.1) (0.0)	137 (0.6) (0.2)	168 (0.8) (0.2)
5	162 (0.7) (0.2)	545 (2.5) (0.6)	707 (3.2) (0.8)
6	23 (0.1) (0.0)	91 (0.4) (0.1)	114 (0.5) (0.1)
7	247 (1.1) (0.3)	205 (0.9) (0.2)	452 (2.0) (0.5)
8		1 (0.0) (0.0)	1 (0.0) (0.0)
9		25 (0.1) (0.0)	25 (0.1) (0.0)
10	293 (1.3) (0.3)	204 (0.9) (0.2)	497 (2.2) (0.6)
11	89 (0.4) (0.1)	79 (0.4) (0.1)	168 (0.8) (0.2)
12		3 (0.0) (0.0)	3 (0.0) (0.0)
13	12 (0.1) (0.0)	19 (0.1) (0.0)	31 (0.1) (0.0)
14	130 (0.6) (0.1)	138 (0.6) (0.2)	268 (1.2) (0.3)
총합계	12,076 (54.5) (13.8)	10,071 (45.5) (11.5)	22,147 (100.0) (25.3)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:증손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

다. 3차 시험조사 지역 조사자료

3차 조사지역의 시험 조사자료에서 비매칭 된 자료의 수는 17,927건이며 이를 성별과 연령대로 살펴보았다. 비매칭 된 총 자료의 수 중 남자가 9,329건(17,927건 중 52.0%)을 차지하여 남자의 비매칭 자료의 수가 많이 나타나고 있다. 비매칭 자료의 수가 가장 많은 20대를 살펴보면 가구주 관계가1(가구주)인 경우가 많으며 그 다음으로 가구주관계3(자녀)가 많이 나타나고 있다. 가구주 관계가 1(가구주)인 경우는 40대가 가장 많으며 30대가 그 다음으로 나타났다. 가구주 관계와 성별을 살펴보면 세대주와 세대주의 배우자 다음으로 남녀의 비율의 차이가 큰 경우는 가구주 관계가 14(기타 동거인)인 경우이며 가구주 관계가 14(기타 동거인)의 남자는 638건(819건 중 77.9%), 여자는 181건(819건 중 22.1%)의 차이를 보이고 있다.

연령대와 성별

(N, 비매칭기준 %, 전체기준 %)

	남자	여자	(비어있음)	총합계
1~9세	564	504	11	1,079
	(3.1)	(2.8)	(0.1)	(6.0)
	(0.6)	(0.6)	(0.0)	(1.2)
10~19세	724	675	1	1,400
	(4.0)	(3.8)	(0.0)	(7.8)
	(0.8)	(0.7)	(0.0)	(1.5)
20~29세	2,177	2,119	5	4,301
	(12.1)	(11.8)	(0.0)	(24.0)
	(2.4)	(2.3)	(0.0)	(4.7)
30~39세	1,810	1,215	4	3,029
	(10.1)	(6.8)	(0.0)	(16.9)
	(2.0)	(1.3)	(0.0)	(3.3)
40~49세	1,849	1,457	3	3,309
	(10.3)	(8.1)	(0.0)	(18.5)
	(2.0)	(1.6)	(0.0)	(3.6)
50~59세	1,268	961		2,229
	(7.1)	(5.4)		(12.4)
	(1.4)	(1.1)		(2.4)
60~69세	506	631		1,137
	(2.8)	(3.5)		(6.3)
	(0.6)	(0.7)		(1.2)
70~79세	320	685	1	1,006
	(1.8)	(3.8)	(0.0)	(5.6)
	(0.4)	(0.7)	(0.0)	(1.1)
80~89세	91	282	1	374
	(0.5)	(1.6)	(0.0)	(2.1)
	(0.1)	(0.3)	(0.0)	(0.4)
90세 이상	20	41	2	63
	(0.1)	(0.2)	(0.0)	(0.4)
	(0.0)	(0.0)	(0.0)	(0.1)
총합계	9,329	8,570	28	17,927
	(52.0)	(47.8)	(0.2)	(100.0)
	(10.2)	(9.4)	(0.0)	(19.6)

가구주관계와 연령대

(N, 비매칭기준 %, 전체기준 %)

	1-9세	10-19세	20-29세	30-39세	40-49세	50-59세	60-69세	70-79세	80-89세	90세 이상	총합계
1	16 (0.1) (0.0)	24 (0.1) (0.0)	1,152 (6.4) (1.3)	1,457 (8.1) (1.6)	1,887 (10.5) (2.1)	1,317 (7.3) (1.4)	602 (3.4) (0.7)	509 (2.8) (0.6)	173 (1.0) (0.2)	20 (0.1) (0.0)	7,157 (39.9) (7.8)
2	7 (0.0) (0.0)	2 (0.0) (0.0)	261 (1.5) (0.3)	781 (4.4) (0.9)	1,135 (6.3) (1.2)	728 (4.1) (0.8)	347 (1.9) (0.4)	216 (1.2) (0.2)	22 (0.1) (0.0)	3 (0.0) (0.0)	3,502 (19.5) (3.8)
3	844 (4.7) (0.9)	1,146 (6.4) (1.3)	916 (5.1) (1.0)	307 (1.7) (0.3)	64 (0.4) (0.1)	18 (0.1) (0.0)	2 (0.0) (0.0)			4 (0.0) (0.0)	3,301 (18.4) (3.7)
4			27 (0.2) (0.0)	43 (0.2) (0.0)	19 (0.1) (0.0)	7 (0.0) (0.0)	1 (0.0) (0.0)				97 (0.5) (0.1)
5	1 (0.0) (0.0)	1 (0.0) (0.0)	1 (0.0) (0.0)		10 (0.1) (0.0)	59 (0.3) (0.1)	139 (0.8) (0.2)	222 (1.2) (0.2)	153 (0.9) (0.2)	19 (0.1) (0.0)	605 (3.4) (0.7)
6				1 (0.0) (0.0)	3 (0.0) (0.0)	7 (0.0) (0.0)	15 (0.1) (0.0)	45 (0.3) (0.0)	20 (0.1) (0.0)	7 (0.0) (0.0)	98 (0.5) (0.1)
7	101 (0.6) (0.1)	58 (0.3) (0.1)	22 (0.1) (0.0)	3 (0.0) (0.0)							184 (1.0) (0.2)
8	2 (0.0) (0.0)		2 (0.0) (0.0)								4 (0.0) (0.0)
9								3 (0.0) (0.0)	5 (0.0) (0.0)	2 (0.0) (0.0)	10 (0.1) (0.0)
10	1 (0.0) (0.0)	18 (0.1) (0.0)	169 (0.9) (0.2)	82 (0.5) (0.1)	30 (0.2) (0.0)	10 (0.1) (0.0)	1 (0.0) (0.0)	3 (0.0) (0.0)			314 (1.8) (0.3)
11	16 (0.1) (0.0)	22 (0.1) (0.0)	19 (0.1) (0.0)	6 (0.0) (0.0)	2 (0.0) (0.0)	1 (0.0) (0.0)					66 (0.4) (0.1)
12			4 (0.0) (0.0)	2 (0.0) (0.0)	3 (0.0) (0.0)	1 (0.0) (0.0)					10 (0.1) (0.0)
13	10 (0.1) (0.0)	16 (0.1) (0.0)	42 (0.2) (0.0)	18 (0.1) (0.0)	15 (0.1) (0.0)	7 (0.0) (0.0)	3 (0.0) (0.0)	1 (0.0) (0.0)	1 (0.0) (0.0)		113 (0.6) (0.1)
14	9 (0.1) (0.0)	8 (0.0) (0.0)	421 (2.3) (0.5)	224 (1.2) (0.2)	89 (0.5) (0.1)	49 (0.3) (0.1)	17 (0.1) (0.0)	2 (0.0) (0.0)			819 (4.6) (0.9)
(비어 있음)	72 (0.4) (0.1)	105 (0.6) (0.1)	1,265 (7.1) (1.4)	105 (0.6) (0.1)	52 (0.3) (0.1)	25 (0.1) (0.0)	10 (0.1) (0.0)	5 (0.0) (0.0)		8 (0.0) (0.0)	1,647 (9.2) (1.8)
총합계	1,079 (6.0) (1.2)	1,400 (7.8) (1.5)	4,301 (24.0) (4.7)	3,029 (16.9) (3.3)	3,309 (18.5) (3.6)	2,229 (12.4) (2.4)	1,137 (6.3) (1.2)	1,006 (5.6) (1.1)	374 (2.1) (0.4)	63 (0.4) (0.1)	17,927 (100.0) (19.6)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:중손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

가구주관계와 성별

(N, 비매칭기준 %, 전체기준 %)

	남자	여자	(비어있음)	총합계
1	5250 (29.3) (5.7)	1903 (10.6) (2.1)	4 (0.0) (0.0)	7157 (39.9) (7.8)
2	331 (1.8) (0.4)	3168 (17.7) (3.5)	3 (0.0) (0.0)	3502 (19.5) (3.8)
3	1718 (9.6) (1.9)	1581 (8.8) (1.7)	2 (0.0) (0.0)	3301 (18.4) (3.6)
4	23 (0.1) (0.0)	74 (0.4) (0.1)		97 (0.5) (0.1)
5	107 (0.6) (0.1)	497 (2.8) (0.5)	1 (0.0) (0.0)	605 (3.4) (0.7)
6	18 (0.1) (0.0)	80 (0.4) (0.1)		98 (0.5) (0.1)
7	86 (0.5) (0.1)	98 (0.5) (0.1)		184 (1.0) (0.2)
8	3 (0.0) (0.0)	1 (0.0) (0.0)		4 (0.0) (0.0)
9	1 (0.0) (0.0)	9 (0.1) (0.0)		10 (0.1) (0.0)
10	146 (0.8) (0.2)	168 (0.9) (0.2)		314 (1.8) (0.3)
11	31 (0.2) (0.0)	35 (0.2) (0.0)		66 (0.4) (0.1)
12	4 (0.0) (0.0)	6 (0.0) (0.0)		10 (0.1) (0.0)
13	59 (0.3) (0.1)	54 (0.3) (0.1)		113 (0.6) (0.1)
14	638 (3.6) (0.7)	181 (1.0) (0.2)		819 (4.6) (0.9)
(비어있음)	914 (5.1) (1.0)	715 (4.0) (0.8)	18 (0.1) (0.0)	1647 (9.2) (1.8)
총합계	9329 (52.0) (10.2)	8570 (47.8) (9.4)	28 (0.2) (0.0)	17927 (100.0) (19.6)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:중손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

라. 3차 조사지역의 행정자료

행정자료에서 매칭되지 않은 자료의 수에 대해 연령대와 성별에 관한 교차표에서 비 매칭 수가 가장 많은 20대에서 남자가 2,731건(4,478건 중 61.0%), 여자가 1,747건(4,478건 중 39.0%)로 남자가 여자에 비해 많이 남아 있다. 하지만 시험 조사자료의 남은 자료의 수에서는 이와 같이 큰 차이가 나지 않기 때문에 이는 남자가 여자보다 적게 매칭되었다라고 하기보다 행정자료와 시험 조사자료의 차이 때문에 발생하였다고 할 수 있다.

가구주관계와 연령의 교차표에서는 가구주관계3(자녀)의 비 매칭 자료의 수가 가장 많으며 여기에서 여기서 20대는 가구주관계3(자녀)의 4,478건 중 3,274건(4,478건 중 73.1%)를 차지한다.

연령대와 성별

(N, 비매칭기준 %, 전체기준 %)

	남자	여자	총합계
1~9세	485 (2.7) (0.5)	458 (2.6) (0.5)	943 (5.3) (1.0)
10~19세	915 (5.2) (1.0)	729 (4.1) (0.8)	1,644 (9.3) (1.8)
20~29세	2,731 (15.4) (3.0)	1,747 (9.8) (1.9)	4,478 (25.2) (4.9)
30~39세	1,790 (10.1) (2.0)	1,275 (7.2) (1.4)	3,065 (17.3) (3.4)
40~49세	1,643 (9.3) (1.8)	1,292 (7.3) (1.4)	2,935 (16.5) (3.2)
50~59세	1,090 (6.1) (1.2)	922 (5.2) (1.0)	2,012 (11.3) (2.2)
60~69세	472 (2.7) (0.5)	681 (3.8) (0.7)	1,153 (6.5) (1.3)
70~79세	305 (1.7) (0.3)	714 (4.0) (0.8)	1,019 (5.7) (1.1)
80~89세	112 (0.6) (0.1)	312 (1.8) (0.3)	424 (2.4) (0.5)
90세 이상	12 (0.1) (0.0)	52 (0.3) (0.1)	64 (0.4) (0.1)
총합계	9,555 (53.9) (10.5)	8,182 (46.1) (9.0)	17,737 (100.0) (19.5)

가구주관계와 연령대

(N, 비매칭기준 %, 전체기준 %)

	1~9세	10~19세	20~29세	30~39세	40~49세	50~59세	60~69세	70~79세	80~89세	90세 이상	총합계
1	4 (0.0) (0.0)	23 (0.1) (0.0)	667 (3.8) (0.7)	1,433 (8.1) (1.6)	1,788 (10.1) (2.0)	1,318 (7.4) (1.4)	634 (3.6) (0.7)	496 (2.8) (0.5)	159 (0.9) (0.2)	15 (0.1) (0.0)	6,537 (36.9) (7.2)
2			71 (0.4) (0.1)	369 (2.1) (0.4)	582 (3.3) (0.6)	402 (2.3) (0.4)	239 (1.3) (0.3)	121 (0.7) (0.1)	16 (0.1) (0.0)		1,800 (10.1) (2.0)
3	750 (4.2) (0.8)	1,371 (7.7) (1.5)	3,272 (18.4) (3.6)	833 (4.7) (0.9)	248 (1.4) (0.3)	56 (0.3) (0.1)	8 (0.0) (0.0)	1 (0.0) (0.0)			6,539 (36.9) (7.2)
4			19 (0.1) (0.0)	101 (0.6) (0.1)	77 (0.4) (0.1)	32 (0.2) (0.0)	6 (0.0) (0.0)	5 (0.0) (0.0)			240 (1.4) (0.3)
5			2 (0.0) (0.0)	1 (0.0) (0.0)	10 (0.1) (0.0)	66 (0.4) (0.1)	183 (1.0) (0.2)	301 (1.7) (0.3)	193 (1.1) (0.2)	30 (0.2) (0.0)	786 (4.4) (0.9)
6					4 (0.0) (0.0)	14 (0.1) (0.0)	56 (0.3) (0.1)	75 (0.4) (0.1)	36 (0.2) (0.0)	11 (0.1) (0.0)	196 (1.1) (0.2)
7	120 (0.7) (0.1)	92 (0.5) (0.1)	92 (0.5) (0.1)	5 (0.0) (0.0)							309 (1.7) (0.3)
8	1 (0.0) (0.0)										1 (0.0) (0.0)
9								5 (0.0) (0.0)	9 (0.1) (0.0)	4 (0.0) (0.0)	18 (0.1) (0.0)
10	2 (0.0) (0.0)	29 (0.2) (0.0)	137 (0.8) (0.2)	159 (0.9) (0.2)	106 (0.6) (0.1)	49 (0.3) (0.1)	3 (0.0) (0.0)	1 (0.0) (0.0)			486 (2.7) (0.5)
11	34 (0.2) (0.0)	63 (0.4) (0.1)	66 (0.4) (0.1)	17 (0.1) (0.0)	5 (0.0) (0.0)	1 (0.0) (0.0)					186 (1.0) (0.2)
12				1 (0.0) (0.0)	2 (0.0) (0.0)	1 (0.0) (0.0)			1 (0.0) (0.0)		5 (0.0) (0.0)
13	22 (0.1) (0.0)	34 (0.2) (0.0)	56 (0.3) (0.1)	68 (0.4) (0.1)	42 (0.2) (0.0)	25 (0.1) (0.0)	3 (0.0) (0.0)		4 (0.0) (0.0)	2 (0.0) (0.0)	256 (1.4) (0.3)
14	10 (0.1) (0.0)	32 (0.2) (0.0)	96 (0.5) (0.1)	78 (0.4) (0.1)	71 (0.4) (0.1)	48 (0.3) (0.1)	21 (0.1) (0.0)	14 (0.1) (0.0)	6 (0.0) (0.0)	2 (0.0) (0.0)	378 (2.1) (0.4)
총합계	943 (5.3) (1.0)	1,644 (9.3) (1.8)	4,478 (25.2) (4.9)	3,065 (17.3) (3.4)	2,935 (16.5) (3.2)	2,012 (11.3) (2.2)	1,153 (6.5) (1.3)	1,019 (5.7) (1.1)	424 (2.4) (0.5)	64 (0.4) (0.1)	17,737 (100.0) (19.5)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:증손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

가구주관계와 성별

(N, 비매칭기준 %, 전체기준 %)

	남자	여자	총합계
1	4,277 (24.1) (4.7)	2,260 (12.7) (2.5)	6,537 (36.9) (7.2)
2	1 (0.0) (0.0)	1,799 (10.1) (2.0)	1,800 (10.1) (2.0)
3	4,071 (23.0) (4.5)	2,468 (13.9) (2.7)	6,539 (36.9) (7.2)
4	119 (0.7) (0.1)	121 (0.7) (0.1)	240 (1.4) (0.3)
5	145 (0.8) (0.2)	641 (3.6) (0.7)	786 (4.4) (0.9)
6	41 (0.2) (0.0)	155 (0.9) (0.2)	196 (1.1) (0.2)
7	161 (0.9) (0.2)	148 (0.8) (0.2)	309 (1.7) (0.3)
8		1 (0.0) (0.0)	1 (0.0) (0.0)
9	1 (0.0) (0.0)	17 (0.1) (0.0)	18 (0.1) (0.0)
10	285 (1.6) (0.3)	201 (1.1) (0.2)	486 (2.7) (0.5)
11	106 (0.6) (0.1)	80 (0.5) (0.1)	186 (1.0) (0.2)
12		5 (0.0) (0.0)	5 (0.0) (0.0)
13	145 (0.8) (0.2)	111 (0.6) (0.1)	256 (1.4) (0.3)
14	203 (1.1) (0.2)	175 (1.0) (0.2)	378 (2.1) (0.4)
총합계	9,555 (53.9) (10.5)	8,182 (46.1) (9.0)	17,737 (100.0) (19.5)

*1:세대주, *2:세대주의 배우자, *3:자녀, *4:자녀의 배우자, *5:세대주의 부모, *6:배우자의 부모, *7:손자녀, 그 배우자, *8:중손자녀, 그 배우자, *9:조부모, *10:형제자매, 그 배우자, *11:형제자매의 자녀, 그 배우자, *12:부모의 형제자매, 그 배우자, *13:기타 친인척, *14:기타동거인

2. 주택

가. 제 2차 행정자료

전체적으로 2차 행정자료는 7,240건(28.8%)의 비매칭을 보이며 주택종류에서는 일반주택, 집합주택 순으로 비매칭률이 높게 나타나며 면적은 46~96 m^2 미만, 46 m^2 미만의 순으로 높은 비매칭률을 보인다. 건축연도에서는 2000년대가 가장 높은 비매칭률을 보이며 그 외 건축연도에서는 비슷한 비매칭률을 보인다.

주택종류별 면적으로 살펴보면 집합주택이면서 46~96 m^2 미만인 곳이 가장 높은 비매칭률을 보이며 일반주택의 46 m^2 미만인 곳이 두 번째로 높은 비매칭률을 보인다.

주택종류별 건축연도로 살펴보면 일반주택이면서 건축연도가 누락된 자료의 비매칭률이 가장 높으며 집합건물 중에는 2000년대 건축연도인 자료의 비매칭률이 높게 나타났다. 일반주택의 경우 건축연도가 1990년 이전이거나 누락된 자료의 비매칭률이 높았으며 집합주택은 반대로 2000년대와 1990년대의 건축연도를 가진 자료의 비매칭률이 높게 나타났다.

면적별 건축연도를 살펴보면 46 m^2 미만이면서 건축연도가 1990년 이전이거나 누락된 자료의 비매칭률이 높았으며 면적이 46~96 m^2 미만인 곳은 2000년대와 1990년 대의 비매칭률이 높았다.

주택종류와 면적

(N, 비 매칭 기준 %, 전체기준 %)

주택종류	46m ² 미만	96m ² 미만	96m ² 이상	총합계
일반주택	1,517	1,068	662	3,247
	(21.0)	(14.8)	(9.1)	(44.8)
	(6.0)	(4.2)	(2.6)	(12.9)
단독주택	39	43	155	237
	(0.5)	(0.6)	(2.1)	(3.3)
	(0.2)	(0.2)	(0.6)	(0.9)
APT	316	419	100	835
	(4.4)	(5.8)	(1.4)	(11.5)
	(1.3)	(1.7)	(0.4)	(3.3)
집합주택	269	2,386	172	2,827
	(3.7)	(33.0)	(2.4)	(39.0)
	(1.1)	(9.5)	(0.7)	(11.2)
그외주택	80	11	3	94
	(1.1)	(0.2)	(0.0)	(1.3)
	(0.3)	(0.0)	(0.0)	(0.4)
총합계	2,221	3,927	1,092	7,240
	(30.7)	(54.2)	(15.1)	(100.0)
	(8.8)	(15.6)	(4.3)	(28.8)

주택종류와 건축연도

(N, 비 매칭 기준 %, 전체기준 %)

주택종류	건축연도				총합계
	2000년대	1990년대	1990년이전	누락	
일반주택	482	321	1,050	1,394	3,247
	(6.7)	(4.4)	(14.5)	(19.3)	(44.8)
	(1.9)	(1.3)	(4.2)	(5.5)	(12.9)
단독주택	108	37	71	21	237
	(1.5)	(0.5)	(1.0)	(0.3)	(3.3)
	(0.4)	(0.1)	(0.3)	(0.1)	(0.9)
APT	483	302	50		835
	(6.7)	(4.2)	(0.7)		(11.5)
	(1.9)	(1.2)	(0.2)		(3.3)
집합주택	1,388	1,137	286	16	2,827
	(19.2)	(15.7)	(4.0)	(0.2)	(39.0)
	(5.5)	(4.5)	(1.1)	(0.1)	(11.2)
그외주택	1	1	91	1	94
	(0.0)	(0.0)	(1.3)	(0.0)	(1.3)
	(0.0)	(0.0)	(0.4)	(0.0)	(0.4)
총합계	2,462	1,798	1,548	1,432	7,240
	(34.0)	(24.8)	(21.4)	(19.8)	(100.0)
	(9.8)	(7.2)	(6.2)	(5.7)	(28.8)

면적과 건축연도

(N, 비 매칭 기준 %, 전체기준 %)

면적	건축연도				
	2000년대	1990년대	1990년이전	누락	총합계
46m ² 미만	283	239	814	885	2,221
	(3.9)	(3.3)	(11.2)	(12.2)	(30.7)
	(1.1)	(1.0)	(3.2)	(3.5)	(8.8)
96m ² 미만	1,689	1,201	566	471	3,927
	(23.3)	(16.6)	(7.8)	(6.5)	(54.2)
	(6.7)	(4.8)	(2.3)	(1.9)	(15.6)
96m ² 이상	490	358	168	76	1,092
	(6.8)	(4.9)	(2.3)	(1.0)	(15.1)
	(1.9)	(1.4)	(0.7)	(0.3)	(4.3)
총합계	2,462	1,798	1,548	1,432	7,240
	(34.0)	(24.8)	(21.4)	(19.8)	(100.0)
	(9.8)	(7.2)	(6.2)	(5.7)	(28.8)

나. 제2차 시험 조사자료

전체적으로 2차 시험 조사자료는 6,558건(26.8%)의 매칭이 되지 않으며 거처종류에서는 집합주택, 일반주택 순으로 비매칭률이 높게 나타나며 면적은 46~96m²미만, 46m²미만의 순으로 높은 비매칭률을 보인다. 건축연도에서는 2000년대가 가장 높은 비매칭률을 보이면 그 외 누락된 건축연도를 제외하고는 비슷한 비매칭률을 보인다.

거처종류별 면적으로 살펴보면 집합주택과 일반주택 모두 46~96m²미만인 곳이 가장 높은 비매칭률을 보인다.

거처종류별 건축연도로 살펴보면 일반주택이면서 건축연도가 1990년 이전인 자료의 비매칭률이 높으며 집합주택 중에는 2000년대, 1990년대 건축연도인 자료의 비매칭률이 높게 나타났다.

면적별 건축연도를 살펴보면 면적이 46~96m²미만인 곳은 2000년대와 1990년대의 비매칭률이 높았다.

거처종류와 면적

(N, 비매칭기준 %, 전체기준 %)

종류	면적			총합계
	46m ² 미만	96m ² 미만	96m ² 이상	
일반주택	271	889	565	1,725
	(4.1)	(13.6)	(8.6)	(26.3)
	(1.1)	(3.6)	(2.3)	(7.1)
단독주택	97	268	358	723
	(1.5)	(4.1)	(5.5)	(11.0)
	(0.4)	(1.1)	(1.5)	(3.0)
APT	57	282	4	343
	(0.9)	(4.3)	(0.1)	(5.2)
	(0.2)	(1.2)	(0.0)	(1.4)
집합주택	190	2,226	359	2,775
	(2.9)	(33.9)	(5.5)	(42.3)
	(0.8)	(9.1)	(1.5)	(11.4)
기타 주택	981	7	4	992
	(15.0)	(0.1)	(0.1)	(15.1)
	(4.0)	(0.0)	(0.0)	(4.1)
총합계	1,596	3,672	1,290	6,558
	(24.3)	(56.0)	(19.7)	(100.0)
	(6.5)	(15.0)	(5.3)	(26.8)

거처종류와 건축연도

(N, 비매칭기준 %, 전체기준 %)

거처종류	건축연도				
	2000년대	1990년대	1990년 이전	누락	총합계
일반주택	388 (5.9) (1.6)	372 (5.7) (1.5)	953 (14.5) (3.9)	12 (0.2) (0.0)	1,725 (26.3) (7.1)
단독주택	136 (2.1) (0.6)	161 (2.5) (0.7)	418 (6.4) (1.7)	8 (0.1) (0.0)	723 (11.0) (3.0)
APT	124 (1.9) (0.5)	81 (1.2) (0.3)	130 (2.0) (0.5)	8 (0.1) (0.0)	343 (5.2) (1.4)
집합주택	1,284 (19.6) (5.3)	1,114 (17.0) (4.6)	362 (5.5) (1.5)	15 (0.2) (0.1)	2,775 (42.3) (11.4)
기타 주택	649 (9.9) (2.7)	5 (0.1) (0.0)	3 (0.0) (0.0)	335 (5.1) (1.4)	992 (15.1) (4.1)
총합계	2,581 (39.4) (10.6)	1,733 (26.4) (7.1)	1,866 (28.5) (7.6)	378 (5.8) (1.5)	6,558 (100.0) (26.8)

면적과 건축연도

(N, 비매칭기준 %, 전체기준 %)

	건축연도				
	2000년대	1990년대	1990년 이전	누락	총합계
46m ² 미만	717 (10.9) (2.9)	105 (1.6) (0.4)	420 (6.4) (1.7)	354 (5.4) (1.4)	1,596 (24.3) (6.5)
96m ² 미만	1,403 (21.4) (5.7)	1,174 (17.9) (4.8)	1,074 (16.4) (4.4)	21 (0.3) (0.1)	3,672 (56.0) (15.0)
96m ² 이상	461 (7.0) (1.9)	454 (6.9) (1.9)	372 (5.7) (1.5)	3 (0.0) (0.0)	1,290 (19.7) (5.3)
총합계	2,581 (39.4) (10.6)	1,733 (26.4) (7.1)	1,866 (28.5) (7.6)	378 (5.8) (1.5)	6,558 (100.0) (26.8)

다. 제3차 행정자료

전체적으로 3차 행정자료는 2,462건(9.2%)이 매칭되지 않았으며, 주택종류에서는 일반주택의 비매칭률이 다른 주택에 비해 높게 나타나며 면적은 46~96 m^2 미만이 다른 면적에 비해 높은 비매칭률을 보인다. 건축연도에서는 1990년대 이전의 자료가 가장 높은 비매칭률을 보인다.

주택종류별 면적으로 살펴보면 일반주택이면서 46~96 m^2 미만인 곳이 가장 높은 비매칭률을 보이며 46 m^2 미만인 곳이 두 번째로 높은 비매칭률을 보인다.

주택종류별 건축연도로 살펴보면 일반주택이면서 건축연도가 1990년 이전 자료의 비매칭률이 가장 높으며 집합건물 중에는 2000년대 건축연도인 자료의 비매칭률이 높게 나타났다. 일반주택의 경우 건축연도가 1990년 이전이나 누락된 자료의 비매칭률이 높았으며 집합주택은 반대로 2000년대와 1990년대의 건축연도를 가진 자료의 비매칭률이 높게 나타났다.

면적별 건축연도를 살펴보면 46 m^2 미만이면서 건축연도가 1990년 이전이거나 누락된 자료의 비매칭률이 높았으며 면적이 46~96 m^2 미만인 곳은 2000년대와 1990년대의 비매칭률이 높았다.

주택종류와 면적

(N, 비매칭기준 %, 전체기준 %)

주택종류	면적			총합계
	46m ² 미만	96m ² 미만	96m ² 이상	
일반주택	462	983	180	1,625
	(18.8)	(39.9)	(7.3)	(66.0)
	(1.7)	(3.7)	(0.7)	(6.0)
단독주택	9	50	277	336
	(0.4)	(2.0)	(11.3)	(13.6)
	(0.0)	(0.2)	(1.0)	(1.2)
APT		44	1	45
		(1.8)	(0.0)	(1.8)
		(0.2)	(0.0)	(0.2)
집합주택	39	346	71	456
	(1.6)	(14.1)	(2.9)	(18.5)
	(0.1)	(1.3)	(0.3)	(1.7)
총합계	510	1,423	529	2,462
	(20.7)	(57.8)	(21.5)	(100.0)
	(1.9)	(5.3)	(2.0)	(9.2)

주택종류와 건축연도

(N, 비매칭기준 %, 전체기준 %)

주택종류	건축연도			총합계
	2000년대	1990년대	1990년이전	
일반주택	111	223	1,291	1,625
	(4.5)	(9.1)	(52.4)	(66.0)
	(0.4)	(0.8)	(4.8)	(6.0)
단독주택	132	181	23	336
	(5.4)	(7.4)	(0.9)	(13.6)
	(0.5)	(0.7)	(0.1)	(1.2)
APT	7	38		45
	(0.3)	(1.5)		(1.8)
	(0.0)	(0.1)		(0.2)
집합주택	73	345	38	456
	(3.0)	(14.0)	(1.5)	(18.5)
	(0.3)	(1.3)	(0.1)	(1.7)
총합계	323	787	1,352	2,462
	(13.1)	(32.0)	(54.9)	(100.0)
	(1.2)	(2.9)	(5.0)	(9.2)

면적과 건축연도

(N, 비매칭기준 %, 전체기준 %)

면적	건축연도			총합계
	2000년대	1990년대	1990년이전	
46m ² 미만	23	23	464	510
	(0.9)	(0.9)	(18.8)	(20.7)
	(0.1)	(0.1)	(1.7)	(1.9)
96m ² 미만	107	516	800	1,423
	(4.3)	(21.0)	(32.5)	(57.8)
	(0.4)	(1.9)	(3.0)	(5.3)
96m ² 이상	193	248	88	529
	(7.8)	(10.1)	(3.6)	(21.5)
	(0.7)	(0.9)	(0.3)	(2.0)
총합계	323	787	1,352	2,462
	(13.1)	(32.0)	(54.9)	(100.0)
	(1.2)	(2.9)	(5.0)	(9.2)

라. 3차 조사자료

전체적으로 3차 조사 자료는 2,537건(9.4%)의 비매칭을 보이며 거처종류에서는 일반주택의 비매칭률이 높게 나타나며 면적은 46~96m²미만의 자료가 높은 비매칭률을 보인다. 건축연도에서는 1990년대가 가장 높은 비매칭률을 보인다.

거처종류별 면적으로 살펴보면 일반주택이면서 46~96m²미만인 곳이 가장 높은 비매칭률을 보인다.

거처종류별 건축연도로 살펴보면 일반주택이면서 건축연도가 1990년대 이전 자료의 비매칭률이 가장 높게 나타났다.

면적별 건축연도를 살펴보면 46~96m²미만이면서 건축연도가 1990년대 자료의 비매칭률이 높았다.

거처종류와 면적

(N, 비매칭기준 %, 전체기준 %)

거처종류	면적				
	46	96	96이상	누락	총합계
일반주택	136	933	220	263	1,552
	(5.4)	(36.8)	(8.7)	(10.4)	(61.2)
	(0.5)	(3.5)	(0.8)	(1.0)	(5.8)
단독주택	16	58	282	2	358
	(0.6)	(2.3)	(11.1)	(0.1)	(14.1)
	(0.1)	(0.2)	(1.0)	(0.0)	(1.3)
APT	1	80	18	7	106
	(0.0)	(3.2)	(0.7)	(0.3)	(4.2)
	(0.0)	(0.3)	(0.1)	(0.0)	(0.4)
집합주택	48	376	34	37	495
	(1.9)	(14.8)	(1.3)	(1.5)	(19.5)
	(0.2)	(1.4)	(0.1)	(0.1)	(1.8)
그외주택			1	25	26
	(0.0)	(0.0)	(0.0)	(1.0)	(1.0)
	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)
총합계	201	1,447	555	334	2,537
	(7.9)	(57.0)	(21.9)	(13.2)	(100.0)
	(0.7)	(5.4)	(2.1)	(1.2)	(9.4)

거처종류와 건축연도

(N, 비매칭기준 %, 전체기준 %)

거처종류	건축연도				
	2000년대	1990년대	1990년이전	누락	총합계
일반주택	196	422	670	264	1,552
	(7.7)	(16.6)	(26.4)	(10.4)	(61.2)
	(0.7)	(1.6)	(2.5)	(1.0)	(5.8)
단독주택	149	152	55	2	358
	(5.9)	(6.0)	(2.2)	(0.1)	(14.1)
	(0.6)	(0.6)	(0.2)	(0.0)	(1.3)
APT	61	39		6	106
	(2.4)	(1.5)		(0.2)	(4.2)
	(0.2)	(0.1)		(0.0)	(0.4)
집합주택	54	393	10	38	495
	(2.1)	(15.5)	(0.4)	(1.5)	(19.5)
	(0.2)	(1.5)	(0.0)	(0.1)	(1.8)
그외주택	1			25	26
	(0.0)			(1.0)	(1.0)
	(0.0)			(0.1)	(0.1)
총합계	461	1,006	735	335	2,537
	(18.2)	(39.7)	(29.0)	(13.2)	(100.0)
	(1.7)	(3.7)	(2.7)	(1.2)	(9.4)

면적과 건축연도

(N, 비매칭기준 %, 전체기준 %)

면적	건축연도				
	2000년대	1990년대	1990년이전	누락	총합계
46m ² 미만	29	67	105		201
	(1.1)	(2.6)	(4.1)		(7.9)
	(0.1)	(0.2)	(0.4)		(0.7)
96m ² 미만	228	695	523	1	1,447
	(9.0)	(27.4)	(20.6)	(0.0)	(57.0)
	(0.8)	(2.6)	(1.9)	(0.0)	(5.4)
96m ² 이상	204	243	107	1	555
	(8.0)	(9.6)	(4.2)	(0.0)	(21.9)
	(0.8)	(0.9)	(0.4)	(0.0)	(2.1)
누락		1		333	334
		(0.0)		(13.1)	(13.2)
		(0.0)		(1.2)	(1.2)
총합계	461	1,006	735	335	2,537
	(18.2)	(39.7)	(29.0)	(13.2)	(100.0)
	(1.7)	(3.7)	(2.7)	(1.2)	(9.4)

제3장 지역별 특성별 인구, 가구, 주택 추정기법

제1절 머리말

제3장에서는 주민등록 자료, 건축물 대장 두 개의 행정자료를 이용하여 지역에 의해서 결정되는 최소의 추정지역인 통계 구역에 거주하는 인구와 가구 그리고 구역 내 주택의 수를 추정할 수 있는 방법론을 검토한다. 또, 제2차 제3차 시험조사를 실시한 12개 읍면동의 자료에 이들 방법론을 적용하며 이 과정에서 발생하는 문제점과 향후 개선, 발전의 방향을 설정하기 위한 기초 자료가 되도록 한다.

등록센서스를 기반으로 인구를 추정하고 있는 대표적인 예는 이스라엘이 최근 등록센서스에 포획·재포획 모형(CR: Capture-Recapture Model)을 변형, 적용해서 거주인구를 추정하는 것을 생각할 수 있다. CR 모형은 센서스 자료의 과소집계 부분의 추정을 통하여 실제 거주인구를 추정하기 위하여 도입된 것으로 미국, 영국 등에서 이미 적용하여온 상당히 검증된 모형이다. 그러나 이러한 모형을 적용하기 위해서는 세밀한 표본조사를 필요로 하는데, 예컨대 행정자료에는 나타나 있으나 조사자료에 없는 개체들을 다시 재조사하여 이들이 추정대상지역내에 존재하는지 아니면 추정대상지역을 벗어난 상태인지를 확인하는 수준까지 조사가 수행되어야 한다.

제2절 등록센서스와 추정 기법의 개관

가능한 모든 유용한 행정자료가 매칭된 상태에서 그동안의 인구 주택 총조사 결과가 제공해왔던 지역적 특성과, 인구, 가구, 주택 특성에 따른 빈도(frequency)를 추정하는 것을 목표로 한다. 이를 위하여 본 보고서에서는 1차 단계로서 제2차, 제3차 시험조사 결과에 활용될 수 있는 방법론의 특징을 검토하고, 실제 적용 결과를 비교하면서 향후 추정방법의 최종 결정을 위한 방안을 제시하는 작업 과정을 기술한다.

추정 기법에 사용되는 모형은 크게 회귀모형과 CR모형을 이용하는 추정법과 그리고 다소 직관적인 비 추정 방법이다. 그런데 이들이 효과적으로 이용

되기 위해서는 전국의 행정동을 어떻게 세분화하느냐 하는 것이 중요한 관건이 되기 때문에, 세분화 과정도 추정기법과 함께 검토하게 된다.

추정 방법론에 따라 다음과 같은 전제가 공통적으로 또는 개별적으로 요구된다.

- 행정자료의 과소집계(under-enumeration), 과다집계(over-enumeration)를 추정하기 위하여 설계된 표본 자료가 있다.
- 행정자료와 표본 자료의 매칭 결과 매칭된 개체(주민, 가구, 주택)와 매칭되지 않은 개체가 확정되었다고 가정한다. 그리고 매칭된 개체의 경우에는 행정자료의 내용을 그 개체의 정보로 간주한다.

제3절 추정의 범위와 방법론

1. 추정범위와 추정기법

본 장의 추정대상과 추정범위, 그리고 그에 따른 추정기법은 다음과 같다.

<표 3-1> 추정범위와 추정기법

추정대상	추정범위	추정기법
인구	성별/연령별/세대주 관계	회귀/CR/비추정
가구	가구종류	회귀/비추정
주택	주택종류/면적	회귀/CR

2. 분석자료 및 비교 평가 분석 내용

위에서 열거한 추정대상은 아래의 자료를 통해서 추정되고, 그 평가 내용은 다음의 표와 같다.

<표 3-2> 분석자료 및 평가 내용

분석자료	평가 내용	계획
인구 2-1차 자료	회귀추정의 개선 가능성 검토, 가정 검토	
인구 2, 3차 자료	CR추정, 비추정	
2005년 자료	균집화, 회귀추정, 비추정	
가구 2-1차 자료	회귀추정의 개선 가능성 검토, 가정 검토	
가구 3차 자료	비추정	회귀추정
주택 2-1차 자료	회귀추정의 개선 가능성 검토, 가정 검토	
주택 2, 3차 자료	CR추정, 비추정	회귀추정

* 2-1차 자료는 2008년 추정 시안에 사용한 자료.

제4절 인구의 지역별, 특성별 추정기법

1. 회귀모형을 이용한 방법

가. 2008년 초안

통계청 인구조사과에서는 2008년에 실시한 2010 인구총조사를 위한 제2차 시험조사자료 중 4개 읍면동의 자료를 이용하여 인구, 가구수, 주택수를 추정하기 위한 초안을 작성한 바 있다. 대단히 직관적이고, 통계학적으로는 검토되어야 할 사항이 많이 있지만 추정만을 목적으로 하는 상황에서는 좋은 방향성을 제공하고 있다고 생각된다. 이 초안의 기본 아이디어를 인구추정에 대하여 정리하면 다음과 같다. 4개 읍면동의 두 개 자료를 행정동별(4), 성별(2), 연령대별(10), 가구주와의 관계별(14)로 나누어 총 1120부분군으로 나누어 이중 행정자료, 조사자료 모두에서 0이 나오는 부분군을 제외한 590개 부분군의 조사인구를 종속변수로 하고 각 부분군을 나타내는 더미변수와 주민

등록인구를 독립변수로 회귀추정식을 산출하고, 이를 이용하여 각 부분군의 실제인구를 추정하였다. 이 아이디어는 추정이라는 나름대로의 의미는 있지만 통계학적인 오차개념을 도입하기에는 몇 가지 문제점을 내포하고 있다. 하나는 590개 자료가 상호의존적이라는 점이고, 두 번째는 590개 중에서 상당수가 조사인구 또는 주민등록인구가 10이하이기 때문에 소수의 큰 인구자료에 의한 영향력이 클 수 있다는 점이다. 따라서 추정된 모형을 적용할 때 오차가 크게 될 가능성이 많은 것이 문제점이라고 할 수 있다.

나. 2008 초안 아이디어를 적용하는 방안

2008년 초안 아이디어를 실제로 적용해 보기위하여 이견외(2007)가 분석한 바 있는 2005년 인구 주택총조사를 활용하여 보았다. 적용과정은 다음과 같다.

단계 1. 이상치 판별

인구로 이상치를 제거하는 것보다 행정자료와 조사자료의 비를 이용하여 이상치를 제거하는 것이 더 좋다고 본다. 인구는 지역(동읍면)의 규모를 구분하는데 이용한다.

우선, 각 행정동별로 총조사인구 대비 주민등록인구 비를 구하여 기술통계량을 구하였다.

<표 3-3> 3,572개 동읍면별 총조사인구 대 주민등록인구 비의 기술통계량

	평균	평균의 95% 신뢰구간		중위수	표준 편차	최소값	최대값	범위
		하한	상한					
통계량	0.937	0.933	0.941	0.948	0.120	0.332	5.846	5.514

Tukey의 이상치 판별을 적용하여 0.7285이하와 1.1326이상 되는 동읍면은 제외시키다. 이때 Tukey의 Hinges는 다음과 같다.

제 1사분위수 : 0.8801, 제 3사분위수 0.981086, 사분위범위 : 0.101

Tukey의 이상점 필터

일반적으로, 데이터 표본은 이상점을 포함하고, 표본규모가 클 수록, 최소한 하나 이상의 이상점의 생길 확률이 커진다. 이상점이 존재하면, 평균이나 표준편차의 추정치는 편향된 수치가 된다. 따라서, 이들 추정치를 가지고, 집중경향이나 데이터의 분산에 대한 추론을 하거나 데이터 분포의 집중경향에 대한 신뢰구간을 계산해서는 안 된다.

이상치가 존재하는 경우, 두가지 방법을 통상적으로 채택한다. 하나는 사분편차(또는 사분편차를 바탕으로 하는 신뢰구간, 가령 중앙값의 부호테스트 Sign Test)와 같은 측정치를 이용하여, 이상점에 대한 Sensitivity를 줄일 수 있다. 둘째는 검색된 이상치를 제거하여, 평균, 표준편차, 신뢰구간의 추정치를 계산하는 것이다

Tukey는 다음의 기준에 의거하여 이상치를 배제하는 방법을 제안하였다. 가령, $Y < (Q1 - 1.5 IQR)$ or $Y > (Q3 + 1.5 IQR)$ 라는 조건을 충족하면, 그 값을 평균이나 표준편차에서 제외하는 것이다. 아래식에서 $Q1 =$ 제1사분위점수, $Q3 =$ 제3사분위점수, $IQR = (Q3 - Q1)$ 로서 사분편차를 의미한다. 상자도표를 그려서 Tukey의 규칙에 따른 이상치 절단점(cutoff) 또는 울타리(fence)로 식별하도록 하고 있다.

총조사 인구 대비 주민등록인구비의 이상치에 해당되는 자료를 보면 총 41개 지역으로 다음 표와 같으며 이들에 대하여는 지역적 특징, 사회적 환경등을 고려하여 개별적으로 추정하여야 한다고 본다.

광역 시도별로 전체/정상군/이상치군의 읍면동 분포를 보면 다음 표와 같다. 이상치는 동에서 27, 읍에 1, 면에 13곳으로 총 41 지역이 제외된다.

<표 3-4> 이상 구역

광역시도	동읍면		총조사 등록 민표	조사 대 표 자	주민 등록 자료	총조사 자료
서울특별시	송파구	잠실1동	5.8462		26	152
충청북도	충주시	홍덕구 달천동	1.6977		5,124	8,699
강원도	원주시	홍업면	1.6718		6,402	10,703
인천광역시	서구	석남1동	1.6554		13,913	23,031
경상북도	구미시	북구 공단2동	1.6155		4,491	7,255
경상북도	경산시	북구 북부동	1.6152		11,646	18,811
경기도	안성시	수지구 대덕면	1.5888		8,922	14,175
경상북도	구미시	북구 진미동	1.4805		15,504	22,954
경기도	용인시	기흥구 서농동	1.4757		9,625	14,204
강원도	삼척시	정라동	1.4176		8,537	12,102
부산광역시	영도구	동삼2동	1.3701		4,972	6,812
전라북도	익산시	덕진구 신동	1.3539		18,079	24,478
강원도	강릉시	내곡동	1.3534		10,955	14,827
강원도	춘천시	동산면	1.3345		1,701	2,270
전라북도	익산시	덕진구 중앙동	0.728		5,769	4,200
전라북도	부안군	덕진구 위도면	0.7269		1,721	1,251
강원도	강릉시	왕산면	0.7247		1,831	1,327
경상북도	김천시	북구 부항면	0.7238		2,270	1,643
광주광역시	동구	충장동	0.7235		6,351	4,595
전라남도	광양시	신안군 지도읍	0.7141		5,484	3,916
서울특별시	종로구	종로1.2.3.4가동	0.7093		6,564	4,656
전라남도	순천시	장천동	0.7088		3,867	2,741
전라북도	김제시(진안군)	덕진구 상전면	0.7078		1,109	785
경기도	시흥시	일산서구과림동	0.7044		2,574	1,813
인천광역시	옹진군	영흥면	0.6969		3,986	2,778
제주도	(서귀포시)	북제주군우도면	0.691		1,796	1,241
경기도	안산시	단원구 원곡2동	0.6837		8,531	5,833
서울특별시	용산구	한강로2동	0.6685		4,073	2,723
전라남도	여수시	묘도동	0.6614		1,468	971
인천광역시	서구	석남2동	0.6473		20,410	13,211
경상남도	진주시	대평면	0.6438		1,325	853
전라북도	(김제시)	임실군 덕진구 운암면	0.638		2,141	1,366
경기도	성남시	분당구 판교동	0.6341		902	572
강원도	삼척시	성내동	0.6263		12,769	7,997
대전광역시	동구	중앙동	0.616		3,297	2,031
인천광역시	중구	용유동	0.6114		3,559	2,176
서울특별시	중구	을지로동	0.591		1,780	1,052
경상남도	양산시	합천군 덕곡면	0.5306		1,713	909
전라북도	군산시	덕진구 옥도면	0.5208		4,395	2,289
서울특별시	성북구	월곡3동	0.3319		702	233
서울특별시	송파구	잠실2동	.		0	15

<표 3-5> 광역시도별 동읍면 분포

광역시도	전체				정상군				이상치군			
	동	읍	면	전체	동	읍	면	전체	동	읍	면	전체
강원도	74	24	89	187	71	24	86	181	3	0	3	6
경기도	384	30	110	524	380	30	109	519	4	0	1	5
경상남도	115	22	177	314	115	22	175	312	0	0	2	2
경상북도	100	35	203	338	97	35	202	334	3	0	1	4
광주광역시	91	0	0	91	90	0	0	90	1	0	0	1
대구광역시	134	3	6	143	134	3	6	143				
대전광역시	80	0	0	80	79	0	0	79	1	0	0	1
부산광역시	222	2	3	227	221	2	3	226	1	0	0	1
서울특별시	522	0	0	522	516	0	0	516	6	0	0	6
울산광역시	46	4	8	58	46	4	8	58				
인천광역시	120	1	19	140	117	1	18	136	3	0	1	4
전라남도	70	31	198	299	68	30	198	296	2	1	0	3
전라북도	85	14	145	244	83	14	141	238	2	0	4	6
제주도	31	7	5	43	31	7	4	42	0	0	1	1
충청남도	39	24	146	209	39	24	146	209				
충청북도	50	13	90	153	49	13	90	152	1	0	0	1
전체	2,163	210	1,199	3,572	2,136	209	1,186	3,531	27	1	13	41

※ 행정조사자료에 경기도 파주시 수지구 진동면이 더 있음.

단계 2. 읍면동 군집화

전체 읍면동을 동일한 식으로 추정하는 것은 오차가 너무 크게 되므로 가급적 유사한 읍면동으로 형성된 군집을 구축하는 것이 필요하다고 본다. 여기서는 방법론의 고찰이라는 입장에서 서울특별시에 국한하여 토의하겠다. 서울특별시는 “동”만 존재하며, 2163개의 “동”중에서 24.1%(522개)를 차지한다.

<표 3-6> 광역시도별 동면읍 분포

(단위 : 개)

광역시도	전체				정상군				이상치군			
	동	읍	면	전체	동	읍	면	전체	동	읍	면	전체
서울특별시	522	0	0	522	516	0	0	516	6	0	0	6
전체	2,163	210	1,199	3,572	2,136	209	1,186	3,531	27	1	13	41

군집화의 기준으로 행정자료의 연령별 인구를 이용하였다. 이는 인구의 크기와 동시에 연령별 분포를 함께 고려할 수 있다고 보았다. 아래의 표는 총 516개 동이 20개의 군으로 군집화된 것을 보이고 그림은 이들 군집의 2차원적 표현이다.

<표 3-7> 군집의 중심값과 포함된 동의 갯수

군집	1	2	3	4	5	6	7
연령9	3528.15	4080.60	4262.75	3418.17	1954.00	4905.50	2828.61
연령19	3981.74	3377.00	6267.75	5755.17	2458.00	5568.75	4523.72
연령29	5491.91	3874.20	6411.75	4621.33	10372.00	5400.25	3901.89
연령39	6217.29	6216.40	6892.75	5125.67	8734.00	7926.50	4292.78
연령49	5349.53	4523.60	7623.75	6849.33	4363.00	6979.00	5397.22
연령59	3780.59	2780.80	4626.25	3607.17	3720.00	3733.00	2941.67
연령69	2248.15	1782.60	2449.00	1655.67	1987.00	1984.00	1320.22
연령79	924.35	878.80	1339.00	976.67	674.00	990.75	697.72
연령89	288.26	263.80	481.75	336.17	284.00	348.50	258.00
연령90	34.44	36.80	56.75	39.67	48.00	41.50	29.72
동수	34	5	4	6	1	4	18

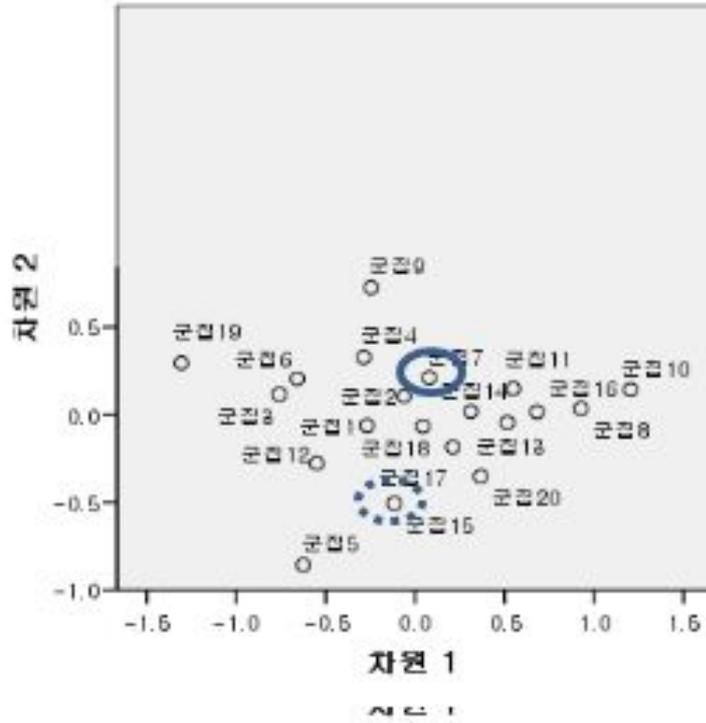
군집	8	9	10	11	12	13	14
연령9	1099.45	3533.50	533.41	1625.82	3600.75	1810.15	2368.44
연령19	1291.95	7319.50	707.39	3027.71	4115.00	2027.94	2844.06
연령29	2165.45	3096.00	1211.20	2895.94	7501.00	3685.31	3750.35
연령39	2267.09	4549.00	1204.65	2743.18	7683.25	3740.13	4235.69
연령49	1853.29	7948.50	1084.39	3604.18	5487.25	2922.93	3771.28
연령59	1372.57	2554.50	844.96	2236.71	3851.25	2251.19	2627.23
연령69	992.16	1380.00	646.49	1070.88	1916.00	1500.12	1581.45
연령79	427.39	837.50	289.98	556.12	735.50	590.90	648.55
연령89	129.04	255.50	92.16	206.76	266.25	184.10	209.64
연령90	15.47	22.50	10.88	25.41	35.00	23.78	25.73
동수	75	2	49	17	4	67	78

군집	15	16	17	18	19	20	
연령9	2290.00	1686.70	2073.47	2719.57	6787.00	1955.00	
연령19	2368.00	1876.59	2379.47	3333.66	6456.00	5362.75	
연령29	7249.60	2763.52	5138.21	4763.13	7624.00	3185.00	
연령39	6262.80	3130.42	4836.00	5022.98	10417.50	2506.13	
연령49	3464.60	2605.34	3421.37	4449.70	8419.50	5726.38	
연령59	3099.40	1908.65	2646.16	3274.11	4917.50	2801.63	
연령69	1941.80	1266.04	1601.95	1945.60	2590.50	898.38	
연령79	662.40	526.41	580.11	786.79	1253.50	556.38	
연령89	218.60	157.51	187.95	251.60	376.50	222.25	
연령90	30.00	17.90	24.16	31.36	37.00	26.00	
동수	5	71	19	47	2	8	516

26개(5%)이상의 동일 포함하고 있는 군집이 7개이고 이들 중 52개(10%)이상 동을 포함한 대형 군집은 4개로 나타났다. 이들을 다차원척도법을 이용하여 이차원 그래프로 나타내면 다음과 같다.

대상포인트

공동공간



<그림 3-30> 군집의 2차원적 표현

단계 3. 학습표본 추출 및 회귀식 추정1

7번 군집 자료 중에 무작위로 30%를 추출하여 30%의 행정동(6)개의 각각의 자료로부터 10세 단위의 연령별 인구를 계산하여 총 60개의 자료에 회귀모형을 적용시킨다.

<표 3-8> 회귀분석 결과 1

	비표준화계수		t	P	F	P	R ²	D-W
	B	SE						
(상수)	8.53	19.08	.45	0.66	25382.7	0.00	0.998	1.741
행정자료	.95	.01	159.32	0.00				

이 결과를 이용하여 이들 6개 행정동과 추정식을 산출하는데 사용되지 않았던 동일 군집내의 14개동에 적용한 결과는 다음과 같다. 여기서는 추정치와 총조사자료의 차이와 행정자료로만 대처하는 경우인 총조사자료와 행정자료의 차이를 비교한다.

<표 3-9> 각 동의 추정 결과 1

	동	추정치	총조사 자료	행정 자료	A	B	A/B
학습표 본	가락본동	27640.5	27,052	28,941	-588.5	-1889	0.3115
	고덕1동	21138.5	21,550	22,112	411.5	-562	-0.7322
	길2동	26470.4	26,531	27,712	60.6	-1181	-0.0513
	목1동	27162.6	26,741	28,439	-421.6	-1698	0.2483
	상봉1동	23782.5	24,145	24,889	362.5	-744	-0.4872
	쌍문4동	21894.5	22,070	22,906	175.5	-836	-0.2099
평가표 본	개포1동	21837.4	23,529	22,846	1691.6	683	2.4767
	도곡2동	23769.2	23,026	24,875	-743.2	-1849	0.4019
	둔촌2동	27066.4	27,133	28,338	66.6	-1205	-0.0553
	목4동	27562.4	26,615	28,859	-947.4	-2244	0.4222
	상계9동	25214.5	25,608	26,393	393.5	-785	-0.5013
	송파2동	22544.8	22,324	23,589	-220.8	-1265	0.1745
	신내1동	27848.1	28,089	29,159	240.9	-1070	-0.2251
	신내2동	26116.2	26,882	27,340	765.8	-458	-1.6721
	신정1동	24250	23,664	25,380	-586	-1716	0.3415
	신정6동	26916.9	26,304	28,181	-612.9	-1877	0.3265
	일원본동	23078	22,748	24,149	-330	-1401	0.2355
	하계2동	26116.2	26,515	27,340	398.8	-825	-0.4834

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

나타난 결과는 학습표본에 사용된 동에 대해서는 A/B가 대부분 1이하로 추정치가 조사치와 근접하고, 다른 동에 적용한 결과도 개포1동과 산내2동을 빼고는 대부분 행정자료로부터 얻는 값을 직접 사용하는 것 보다는 좋다고 본다. 한편 이 회귀추정식을 7번 군집과 다소 차이가 보이는 15번 군집에 적용한 결과도 유사한 현상을 보인다.

<표 3-10> 군집15의 행정동 추정 결과 1

동	추정치	총 조사자료	행정 자료	A	B	A/B
논현1동	26623.7	25,812	27,873	-811.7	-2061	0.3938
봉천11동	27976.6	28,358	29,294	381.4	-936	-0.4075
봉천6동	29516.2	29,997	30,911	480.8	-914	-0.5260
신림2동	24336.7	26,841	25,471	2504.3	1370	1.8280
신림본동	23304.6	24,177	24,387	872.4	-210	-4.1543

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

단계 4. 학습표본 추출 및 회귀식 추정2

단계 3에서는 추출된 6개의 행정동을 60개의 소집단으로 분리하여 회귀 분석 대상 자료를 생성한데 반하여 이 단계에서는 7번 군집 자료 중에 무작위로 추출된 30%의 행정동(6)의 자료를 통합하여 10세 단위의 연령별 인구를 계산하고 이들 10개의 자료로부터 회귀모형을 적용시킨다.

<표 3-11> 회귀 분석 결과 2

	비표준화계수		t	P	F	P	R ²	D-W
	B	SE						
(상수)	-20.24	187.87	-0.11	0.92	9351.0	.000	0.999	2.948
행정자료	0.96	0.01	96.70	0.00				

이 결과로부터 추정치와 총조사자료의 자료의 차이와 총조사자료와 행정자료의 차이를 비교한다.

<표 3-12> 각 동의 추정 결과 2

동	추정치	총 조사 자료	행정 자료	A	B	A/B	
개포1동	21654.9	23,529	22,846	1874.1	683	2.7439	
도곡2동	23596.1	23,026	24,875	-570.1	-1849	0.3083	
둔촌2동	26909.2	27,133	28,338	223.8	-1205	-0.1857	
목4동	27407.7	26,615	28,859	-792.7	-2244	0.3533	
상계9동	25048.4	25,608	26,393	559.6	-785	-0.7129	
평가표 본	송파2동	22365.7	22,324	23,589	-41.7	-1265	0.0330
신내1동	27694.7	28,089	29,159	394.3	-1070	-0.3685	
신내2동	25954.4	26,882	27,340	927.6	-458	-2.0253	
신정1동	24079.2	23,664	25,380	-415.2	-1716	0.2420	
신정6동	26759	26,304	28,181	-455	-1877	0.2424	
일원본동	22901.5	22,748	24,149	-153.5	-1401	0.1096	
하계2동	25954.4	26,515	27,340	560.6	-825	-0.6795	

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

이 회귀추정식을 단계 3에서와 마찬가지로 15번 군집에 적용해본다.

<표 3-13> 군집15의 행정동 추정 결과 2

동	추정치	총 조사자료	행정자료	A	B	A/B
논현1동	26464.4	25,812	27,873	-652.4	-2061	0.3165
봉천11동	27823.9	28,358	29,294	534.1	-936	-0.5706
봉천6동	29370.9	29,997	30,911	626.1	-914	-0.6850
신림2동	24166.3	26,841	25,471	2674.7	1370	1.9523
신림본동	23129.2	24,177	24,387	1047.8	-210	-4.9895

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

단계 5. 종합분석

초안에서는 4개 지역으로부터 추정식을 유도하여 그들 지역에 대한 추정을 수행한 후 조사자료와 추정 결과의 차이를 기술적으로 비교하였다. 이 방법을 보다 구체적으로 등록센서스 상황에 활용할 때 검토할 사항들을 고찰하기 위하여 서울특별시의 500여 동에 대한 2005년 인구 총조사자료와 그 조

사 시기에 근접한 시점의 주민등록 자료를 이용하여, 군집화, 표본 추출, 자료 생성, 회귀 분석, 추정의 과정을 보였다. 이때 군집화는 각 동의 연령대별 인구수를 변수로 사용하여 20개의 군집으로 형성했고 그 중에 18개의 동으로 형성된 군집에 관하여 추정 작업을 두 가지 방법에 적용해 본다.

먼저 18개 동으로 형성된 군집에서 6개 행정동을 임의로 추출하고, 이들 표본으로 추출된 6개 각 행정동으로부터 구성된 연령대별 부분군의 행정인구와 조사인구 60개의 자료를 생성하여 회귀모형을 세웠다.

두 번째는 임의로 추출되었던 6개동의 인구를 통합하여 연령대별 부분군의 행정인구와 조사인구를 집계하여 10개의 자료를 생성하였다. 두 번째 회귀모형의 생성 동기는 각 동을 세분화할 경우 독립성에 문제가 제기될 것으로 예상되어서 이를 통합한 경우와 비교하고자 함이다.

추정 단계에서는 학습 표본으로부터 추정식을 도출하여 추정에 사용되지 않은 동일 집락내 14개동의 행정자료를 이용한 추정값과 그들의 조사값을 비교하였는데, 두 가지 방법 사이에는 차이가 나타나지 않았다. 학습 표본에 대한 추정치는 행정자료에 비하여 6개동 모두 조사자료에 근접하였다. 평가 표본을 하고 있는 14개동 중에서는 개포 1동과 신내 2동에서 1이상의 값이 나오고, 나머지 12개 동에서는 행정자료보다 조사자료에 더 가까운 값이 나왔다.

또한 이 추정식을 7번 군집과 다소 차이가 있는 15번 군집에 적용해 본 결과 조사자료와 행정자료의 차이가 작은 신림 본동에서 차이의 4배가 되는 오차를 보이고, 신림 2동에서도 차이의 비가 1.8로서 크게 나타났다.

2. 포획 재포획 모형

가. 모형소개

이 모형은 미국 등에서 센서스의 과소집계를 추정하기 위하여 사용되어 왔던 모형으로 최근에는 2001년 영국의 One Number 센서스에서 과소집계 보정을 위하여 적용되었고 2008이스라엘 인구조사에서도 대안으로 고려되었다. 그런데 Glickmen(2003)등은 이 모형이 센서스에서 나타나는 과다집계의 오차를 설명하지 못하고 있음을 지적하면서 모형의 확장을 시도하였다. 그의 모형에서 사용된 분포에 대한 가정은 다항분포(multinomial distribution)와 포아송분포(Poisson distribution)이다.

N 명의 인구를 갖는 추정대상 특정지역이 있는데, 이 지역에 두 개의 조사

a, b가 진행된다고 하자. 이 지역이 M 개의 부분 지역(Plot)으로 분할(partition)되었다고 하고 N_i 는 i 번째 Plot의 인구로서 $N_i = U^{11}(i) + U^{12}(i) + U^{21}(i) + U^{22}(i)$ 와 같이 표현된다.

여기서 확률 변수 $U^{ab}(i)$ 는 plot i 에 거주하면서 a, b 조건을 만족하는 인구이다. $a=1, b=1$ 이라 함은 두 개의 조사에서 모두 확인된 것을 의미하는데 예를 들면, $a=1, b=2$ 는 첫 번째 조사 a(우리 경우는 행정자료)에서 확인되었으나 두 번째 조사 b(우리 경우는 시험조사 자료)에서는 조사되지 않은 사람 수이다. 다른 경우도 유사하게 해석한다. $U^{22}(i)$ 는 첫 번째 조사와 두 번째 조사에서 모두 조사되지 않는 사람 수로 관찰 할 수 없는 값이다.

<표 3-14> 자료의 재구성

	거주상태	행정자료	시험조사 자료
U^{11}	Y	Y	Y
U^{12}	Y	Y	N
U^{21}	Y	N	Y
U^{22}	Y	N	N

Y : 존재함

N : 존재하지 않음

일반적으로 이런 상황에서 적용되는 확률 모형은 다항분포와 포와송분포인데 우리는 다항분포만 사용하여 다음과 같은 모수로서 표현된다.

$$U_i \sim Mult(N_i; p_{1+}p_{+1}, p_{1+}(1-p_{+1}), (1-p_{1+})p_{+1}, (1-p_{1+})(1-p_{+1}))$$

이때 p_{1+} 는 시험조사 자료에 있는 거주하는 사람 중 행정자료와 매칭된 사람 비율이며, p_{+1} 는 행정자료에 있는 거주하는 인구 중에서 시험조사 자료와 매칭된 사람비율이다. 여기서 p_{1+} 나 p_{+1} 이 plot i 에 의존하지 않는 것은 모두 plot에서 동일한 확률을 가정하는 단순 모형이다.

이 모형은 앞에서 언급했듯이 센서스의 사후조사를 통해 과소집계 문제들을 해결하기 위하여 주로 사용되었다. 그러나 이 모형은 plot i 에 거주하는 사람들만을 구별할 것이기 때문에 과다집계 즉, 첫 번째 자료인 행정자료에는 있으나 두 번째 자료인 시험조사 자료에 포함되지 않았고, 또한 이 지역에 거주하지 않는다는 사실이 확인된 인구를 모형이 포함하고 있지 않다.

이런 인구를 $X(i, j, k)$ 라는 새로운 확률 변수로 나타내고 모수 λ 인 포아송분포(Poisson distribution)를 따른다고 가정하였다.

나. 자료와 기호

1. 최소 추정 지역구(여기서는 읍, 면, 동)를 생각하자. 이 지역구내에서 과소집계, 과다집계를 추정하기 위하여 조사구 기반으로 조사하는 표본조사와 행정자료 기반으로 조사하는 표본 조사를 수행한다. (동일 표본에서 두 가지 조사가 모두 가능할 수 있다.) 이 때, 이 지역구를 M 개의 조사구로 분할하였다고 하자. 그리고 M 개의 조사구 중 m 개 조사구를 표본으로 추출하였다 하자. 그리고 지역구내의 최소 추정 단위를 예컨대 성별(j), 연령대(k), 인구 $N(j, k)$ 라고 하자.

2. 표본으로 추출된 m 개 조사구에 해당되는 행정자료와 시험조사 자료로부터 다음의 값을 구한다.

$U^{11}(i, j, k)$: 표본 조사구 i 에 있는 성별 j 이고 연령대 k 인 사람들로 행정자료와 시험조사 자료 모두에 있는 사람 수

$U^{12}(i, j, k)$: 표본 조사구 i 에 있는 성별 j 이고 연령대 k 인 사람들로 행정자료에는 있으나 시험조사 자료에는 없는 사람 수

$U^{21}(i, j, k)$: 표본 조사구 i 에 있는 성별 j 이고 연령대 k 인 사람들로 행정자료에는 없으나 시험조사 자료에는 있는 사람 수

$U^{22}(i, j, k)$: 표본 조사구 i 에 있는 성별 j 이고 연령대 k 인 사람들로 행정자료와 시험조사 자료 모두에 없는 사람 수

$X(i, j, k)$: 표본 조사구 i 에 해당되는 행정자료에는 있으나 시험조사 자료에도 없고 또한 추적조사 결과 이 지역에 살지 않는 자로 판명된 성별 j , 연령대 k 인 사람 수

이를 일반화한다면 다음과 같이 쓸 수 있다.

$$U^{ab}(i, j, k), \quad a=1, 2; \quad b=1, 2; \quad i=1, \dots, m; \quad j=1, 2; \quad k=1, \dots, 10;$$

3. 모수추정식

m 개 조사구로부터 과소집계 모수 \widehat{p}_{1+}

$$\widehat{p}_{1+}(j, k) = \frac{\sum_{i=1}^m U^{11}(i, j, k)}{\sum_{i=1}^m (U^{11}(i, j, k) + U^{21}(i, j, k))}$$

\widehat{p}_{1+} 은 시험조사 자료 중 시험조사 자료와 행정자료에 모두 포함된(매칭된) 사람 수의 비율을 나타낸다.

m 개 조사구로부터 과다집계 모수 λ

$$\widehat{\lambda}(j, k) = \frac{\sum_{i=1}^m X(i, j, k)}{\sum_{i=1}^m \frac{(U^{11}(i, j, k) + U^{12}(i, j, k))}{\widehat{p}_{1+}(j, k)}}$$

$\widehat{\lambda}$ 는 행정자료에 포함된 거주자의 수를 시험조사 자료 정보를 이용하여 보정한 후에 계산된 이 지역에 거주하는 인구 중에서 행정자료에 존재하나 거주하고 있지 않은 사람 수의 비율이다.

4. 이들로부터 이 지역에 해당되는 행정자료에 있는 성별 j 이고, 연령대가 k 인 개체들에게 다음의 가중치를 부여한다.

$$w(j, k) = \frac{1}{\widehat{p}_{1+}(j, k) + \widehat{\lambda}(j, k)}$$

5. 이 가중치를 이용하여 이 지역의 성별 j 이고, 연령대가 k 인 인구 $N(j, k)$ 를 다음과 같이 추정한다.

$$\widehat{N}(j, k) = \sum_{i=1}^M w(j, k) (U^{11}(i, j, k) + U^{12}(i, j, k) + X(i, j, k))$$

이때의 표준오차는 다음과 같다.

$$\sqrt{N \left(\frac{(1-p_{1+})(1-p_{+1})}{p_{1+}p_{+1}} + \frac{1-f}{f} \frac{(1-p_{1+})}{p_{1+}p_{+1}} + \frac{1-f}{f} \frac{\lambda}{p_{1+} + \lambda} \left(\frac{p_{1+}}{p_{1+} + \lambda} - \frac{(1-p_{1+})}{p_{1+}} \right) \right)}$$

여기서 $f = \frac{m}{M}$ 이다.

다. 자료생성

12개 행정동 전체의 행정자료와 시험조사 자료 건수의 성별 연령별 인구는 다음과 같다.

<표 3- 15> 12개 행정동의 성별 연령별 인구

(단위 : 명)

연령대	남자		여자	
	행정	조사	행정	조사
0~9	8,867	8,808	8,090	7,974
10~19	14,031	13,348	12,685	12,200
20~29	12,771	10,474	11,582	11,347
30~39	15,604	14,495	15,304	14,638
40~49	16,712	15,907	15,843	15,406
50~59	11,655	11,062	11,059	10,608
60~69	6,557	6,227	7,000	6,694
70~79	3,261	3,198	4,664	4,520
80~89	815	767	1,759	1,616
90~	85	86	245	208
총합	90,358	84,372	88,231	85,211

제2차 시험조사 지역(6개 행정동)과 제3차 시험조사 지역(6개 행정동)의 12개 행정동에서 3개의 행정동(2차 부산시 동래구 명장1동, 3차 충북 제천시 교동, 3차 경남 고성군 거류면)을 표본으로 추출하여 자료를 생성하였다.

1장과 2장에서 행정자료의 개체에 관한 기술은 “조사자료와 매칭되었다.” “조사자료와 매칭되지 않았다.”의 두 가지로 분류하였지만, CR 모형을 적용하기 위해서는 “시험조사 자료와 매칭되지 않았다.”를 “시험조사 자료와 매칭되지 않았으나 구역 내에 거주하고 있다.”와 “시험조사 자료와 매칭되지 않았고, 구역 내에 거주하고 있지 않다.”로 다시 분류하여야 한다.

그러나 우리의 매칭결과에서는 이러한 분류를 할 수가 없으므로 시험조사 자료와 매칭되지 않은 행정자료의 개체들(U^{12})중에서 임의의 5%와 1%를

U^{12} 에 할당하고 5%와 1%가 아닌 나머지 인구를 X 에 할당하여 추정식을 계산하였다.

먼저 5%(1%)자료를 생성하는 방법은 SAS를 이용하여 난수를 생성하고 U^{12} 인구 중 난수값이 0.05(0.01)보다 작은 인구를 U^{12} 라하고, 나머지 인구를 X 라 하였다.

이와 같이 하여 3개 행정동의 건수를 U^{12} 가 5%일 경우와 U^{12} 가 1% 일 경우로 $U^{11}, U^{12}, U^{21}, X$ 를 정의 한 인구는 아래와 같다.

<표 3-16>생성된 자료 표본의 성별 연령별 인구

(단위 : 명)

	5%				1%				
	U^{11}	U^{12}	U^{21}	X	U^{11}	U^{12}	U^{21}	X	
남자	0~9	1,528	11	267	190	1,528	2	267	199
	10~19	2,774	11	214	335	2,774	3	214	343
	20~29	1,846	64	459	1,240	1,846	11	459	1,293
	30~39	2,538	43	597	807	2,538	4	597	846
	40~49	2,985	34	550	689	2,985	4	550	719
	50~59	2,565	20	413	528	2,565	6	413	542
	60~69	1,578	16	190	241	1,578	2	190	255
	70~79	707	12	156	140	707	2	156	150
	80~89	165	5	45	45	165	2	45	48
	90~	11	0	6	6	11	0	6	6
	총합	16,697	216	2,897	4,221	16,697	36	2,897	4,401
여자	0~9	1,452	11	194	182	1,452	1	194	192
	10~19	2,535	16	181	257	2,535	7	181	266
	20~29	1,999	45	338	787	1,999	11	338	821
	30~39	2,625	26	361	514	2,625	6	361	534
	40~49	3,138	35	423	556	3,138	8	423	583
	50~59	2,601	26	390	451	2,601	8	390	469
	60~69	1,506	13	292	324	1,506	1	292	336
	70~79	951	16	286	320	951	3	286	333
	80~89	354	6	134	148	354	0	134	154
	90~	44	3	21	22	44	1	21	24
	총합	17,205	197	2,620	3,561	17,205	46	2,620	3,712
전체	33,902	413	5,517	7,782	33,902	82	5,517	8,113	

라. 추정

12개 행정동중 3개의 행정동의 표본 인구를 통해 CR모형으로 추정을 수행하였다.

위의 3개 행정동의 U^{12} 가 5%인 경우, U^{12} 가 1%인 경우의 인구를 통해 과소집계, 과다집계 모수의 추정치와 부분그룹별 가중치, 그리고 이를 이용한 12개 전체 동에 대한 성별 연령별 추정치와 표준오차는 다음과 같다.

<표 3-17> 12개 행정동 인구 추정 (U^{12} 가 5%인 경우)

(단위 : %, 명)

	p1+	p+1	람다	가중치	추정	표준오차	
남자	0~9	0.85	0.99	0.11	1.05	9,272	84.22
	10~19	0.93	1	0.11	0.96	13,491	81.75
	20~29	0.8	0.97	0.52	0.76	9,669	107.84
	30~39	0.81	0.98	0.25	0.94	14,683	126.97
	40~49	0.84	0.99	0.19	0.96	16,114	121.25
	50~59	0.86	0.99	0.18	0.96	11,236	96.49
	60~69	0.89	0.99	0.13	0.97	6,382	64.94
	70~79	0.82	0.98	0.16	1.02	3,332	57.09
	80~89	0.79	0.97	0.21	1.01	820	31.06
	90~	0.65	1	0.35	1	85	12.18
	총합	0.85	0.99	0.21	0.94	84,857	277.13
여자	0~9	0.88	0.99	0.11	1.01	8,156	73.13
	10~19	0.93	0.99	0.09	0.97	12,347	74.19
	20~29	0.86	0.98	0.33	0.84	9,776	98.06
	30~39	0.88	0.99	0.17	0.95	14,581	105.21
	40~49	0.88	0.99	0.15	0.97	15,298	105.73
	50~59	0.87	0.99	0.15	0.98	10,854	91.13
	60~69	0.84	0.99	0.18	0.98	6,888	79.64
	70~79	0.77	0.98	0.25	0.98	4,558	75.74
	80~89	0.73	0.98	0.3	0.98	1,718	49.91
	90~	0.68	0.94	0.32	1.01	247	20.81
	총합	0.87	0.99	0.18	0.96	84,397	261.34
전체	0.86	0.99	0.2	0.95	169,265	381.08	

<표 3- 18> 12개 행정동 인구 추정 (U^{12} 가 1%인 경우)

(단위 : 명,%)

	p1+	p+1	람다	가중치	추정	표준오차	
남자	0~9	0.85	1.00	0.11	1.04	9,218	84.27
	10~19	0.93	1.00	0.11	0.96	13,452	82.00
	20~29	0.80	0.99	0.56	0.74	9,401	104.95
	30~39	0.81	1.00	0.27	0.93	14,462	125.72
	40~49	0.84	1.00	0.20	0.95	15,954	120.85
	50~59	0.86	1.00	0.18	0.96	11,176	96.42
	60~69	0.89	1.00	0.14	0.96	6,326	65.13
	70~79	0.82	1.00	0.17	1.01	3,285	56.78
	80~89	0.79	0.99	0.23	0.99	806	30.69
	90~	0.65	1.00	0.35	1.00	85	12.18
	총합	0.85	1.00	0.22	0.93	83,954	276.14
여자	0~9	0.88	1.00	0.12	1.00	8,100	73.32
	10~19	0.93	1.00	0.10	0.97	12,303	74.57
	20~29	0.86	0.99	0.35	0.83	9,614	97.14
	30~39	0.88	1.00	0.18	0.95	14,471	105.22
	40~49	0.88	1.00	0.16	0.96	15,168	105.82
	50~59	0.87	1.00	0.16	0.97	10,779	91.14
	60~69	0.84	1.00	0.19	0.98	6,834	79.44
	70~79	0.77	1.00	0.27	0.96	4,497	74.91
	80~89	0.73	1.00	0.32	0.96	1,690	49.07
	90~	0.68	0.98	0.36	0.96	236	19.78
	총합	0.87	1.00	0.19	0.95	83,664	261.08
전체	0.86	1.00	0.21	0.94	167,633	380.23	

U^{12} 가 1%인 인구가 U^{12} 가 5%인 인구보다 과다집계가 많다고 의미하므로 추정치가 다소 작아지는 경향을 보이고 있다.

<표 3-19> 추정치의 비교

(단위 : 명, %)

				5%			1%			
	행정	조사	B	추정치	A	A/B	추정치	A	A/B	
남 자	0~9	8,867	8,808	-59	9,272	-464	7.86	9,218	-410	6.94
	10~19	14,031	13,348	-683	13,491	-143	0.21	13,452	-104	0.15
	20~29	12,771	10,474	-2,297	9,669	805	-0.35	9,401	1,073	-0.47
	30~39	15,604	14,495	-1,109	14,683	-188	0.17	14,462	33	-0.03
	40~49	16,712	15,907	-805	16,114	-207	0.26	15,954	-47	0.06
	50~59	11,655	11,062	-593	11,236	-174	0.29	11,176	-114	0.19
	60~69	6,557	6,227	-330	6,382	-155	0.47	6,326	-99	0.3
	70~79	3,261	3,198	-63	3,332	-134	2.12	3,285	-87	1.39
	80~89	815	767	-48	820	-53	1.11	806	-39	0.81
	90~	85	86	1	85	1	1	85	1	1
	총합	90,358	84,372	-5,986	84,857	-485	0.08	83,954	418	-0.07
여 자	0~9	8,090	7,974	-116	8,156	-182	1.57	8,100	-126	1.09
	10~19	12,685	12,200	-485	12,347	-147	0.3	12,303	-103	0.21
	20~29	11,582	11,347	-235	9,776	1,571	-6.68	9,614	1,733	-7.38
	30~39	15,304	14,638	-666	14,581	57	-0.08	14,471	167	-0.25
	40~49	15,843	15,406	-437	15,298	108	-0.25	15,168	238	-0.55
	50~59	11,059	10,608	-451	10,854	-246	0.55	10,779	-171	0.38
	60~69	7,000	6,694	-306	6,888	-194	0.63	6,834	-140	0.46
	70~79	4,664	4,520	-144	4,558	-38	0.27	4,497	23	-0.16
	80~89	1,759	1,616	-143	1,718	-102	0.72	1,690	-74	0.52
	90~	245	208	-37	247	-39	1.04	236	-28	0.76
	총합	88,231	85,211	-3,020	87,520	814	-0.27	83,664	1,547	-0.51
전체	178,589	169,583	-9,006	172,360	318	-0.04	167,633	1,950	-0.22	

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

추정치와 시험조사 결과를 비교하기 위하여 시험조사 자료와 추정치 차이와 시험조사 자료와 행정자료 차이의 비를 구하여 살펴보면 남자 0-9세, 70-79세, 여자 20-29세가 시험조사 값과 추정치의 차이가 행정자료의 차이보다 크게 나왔다. 이유를 조사해 본 결과는 이들 부분군에서 매칭되지 않은 시험조사 자료가 크거나, 또는 매칭되지 않은 행정자료 중에서 구역 밖으로 나간 인구(X)가 큰 상황에서 발생한 것이다.

3. 비 추정을 이용한 기법

보통 표본론에서 다루는 회귀추정의 개념을 이용하여 12개 행정동에서 3개의 행정동을 선택하여 행정자료와 시험조사 자료로부터 다음과 같이 구한다.

행정동($i = 1, \dots, m$), 성($j = 1, 2$), 연령대($k = 1, \dots, K$)에 따라 행정자료(x_{ijk})와 조사자료(y_{ijk})로부터 인구를 구하고 다음과 같이 집계한다.

$t_{xjk} = x_{1jk} + x_{2jk} + x_{3jk}$: 성별과 연령대가 같은 3개 행정동의 행정자료의 합

$t_{yjk} = y_{1jk} + y_{2jk} + y_{3jk}$: 성별과 연령대가 같은 3개 행정동의 시험조사 자료의 합

$$\frac{t_{yjk}}{t_{xjk}} = b_{jk} : \text{위 두 개의 비율}$$

행정자료와 시험조사 자료의 비율(b_{jk})를 이용하여 전체 12개 행정동의 행정자료를 이용하여 성별 연령별 인구를 추정한다.

T_{xjk} : 성별 j , 연령 k 인 12개 행정동의 인구

$\widehat{T}_{yjk} = T_{xjk} * b_{jk}$: 추정 인구

표준오차를 구하는 공식은 다음과 같다.

$$Se^2 = \sum_{k=1}^3 (y_k - b * x_k)^2 / n - 1$$

$$V(\widehat{T}_y) = N^2 * \frac{Se^2}{n} * \frac{N-n}{N}$$

이 방법을 제2차 시험조사 자료 지역과 제3차 시험조사 자료 지역 12개 행정동에 적용한 결과는 아래 표와 같다.

<표 3-20> 비추정 추정치 결과

(단위 : 명, %)

	표본			전체				A	B	A/B	
	행정	조사	b	행정	조사	추정	표준 오차				
남 자	0~9	1,729	1,785	1.03	8,867	8,778	9,154	392.1	-376	-89	4.23
	10~19	3,120	2,980	0.96	14,031	13,309	13,401	702.0	-92	-722	0.13
	20~29	3,150	2,290	0.73	12,771	10,445	9,284	4141.7	1161	-2326	-0.50
	30~39	3,388	3,106	0.92	15,604	14,387	14,305	1599.0	82	-1217	-0.07
	40~49	3,708	3,497	0.94	16,712	15,799	15,761	1150.8	38	-913	-0.04
	50~59	3,113	2,999	0.96	11,655	11,162	11,228	630.8	-66	-493	0.13
	60~69	1,835	1,780	0.97	6,557	6,227	6,360	280.0	-133	-330	0.40
	70~79	859	894	1.04	3,261	3,289	3,394	189.8	-105	28	-3.75
	80~89	215	217	1.01	815	785	823	20.7	-38	-30	1.25
	90~	17	16	0.94	85	83	80	6.7	3	-2	-1.50
	총합	21,134	19,564	0.93	90,358	84,264	83,645	8177.5	619	-6094	-0.10
여 자	0~9	1,645	1,651	1.00	8,090	7,991	8,120	231.4	-129	-99	1.30
	10~19	2,808	2,740	0.98	12,685	12,264	12,378	507.7	-114	-421	0.27
	20~29	2,831	2,337	0.83	11,582	11,373	9,561	3049.8	1812	-209	-8.67
	30~39	3,165	2,961	0.94	15,304	14,526	14,318	1191.2	208	-778	-0.27
	40~49	3,729	3,513	0.94	15,843	15,347	14,925	1161.6	422	-496	-0.85
	50~59	3,078	3,040	0.99	11,059	10,730	10,922	382.9	-192	-329	0.59
	60~69	1,843	1,791	0.97	7,000	6,638	6,802	285.8	-164	-362	0.45
	70~79	1,287	1,259	0.98	4,664	4,560	4,563	132.5	-3	-104	0.02
	80~89	508	491	0.97	1,759	1,655	1,700	76.1	-45	-104	0.43
	90~	69	68	0.99	245	215	241	13.5	-26	-30	0.88
	총합	20,963	19,851	0.95	88,231	85,299	83,551	6681.8	1748	-2932	-0.60
전체	42,097	39,415	0.94	178,589	169,563	167,211	13726.3	2352	-9026	-0.26	

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

대체적으로 추정치가 행정자료에 비하여 시험조사 자료에 근접한 것으로 나타났으나, 남자 0-9세, 여자 20-29세는 표본조사 결과와 전체 행정동의 상황이 다르게 나타났기 때문이다. 따라서 행정동의 세분화 작업의 필요성을 시사하는 상황이다.

제5절 가구수의 지역별, 특성별 추정기법

1. 추정연구 방향

등록자료로부터 가구수를 추정하는 문제의 핵심은 주민등록 자료의 세대개념과 시험조사 자료의 가구개념의 차이를 어떻게 처리하느냐 하는 것이다.

이 문제는 이미 1장의 가구매칭 과정에서 검토된 바 있는 데 가장 단순한 논리는 주민등록명부를 세대로 하고 이 명부와 동일 주소를 갖고 있는 모든 가구를 이 세대와 매칭된 가구로 간주하는 것이다. 한편 본 보고서에서는 최소한 한 명 이상의 구성원이 매칭된 상황에서만 세대와 가구가 매칭된 것으로 보는 관점을 제안한 바 있고 그 차이를 검토한 바 있다.

추정에 사용되는 기법으로 인구 추정기법에서 활용하였던 것처럼 회귀모형을 이용하는 방법과 CR모형 방법을 생각할 수 있다.

회귀모형을 이용하는 방법은 인구추정에서와 같은 독립성의 문제와 오차추정의 문제에 대한 검토가 조금 더 필요한 상태로 계속 연구할 필요가 있다고 보는 반면 CR모형 방법은 중복 매칭이 많기 때문에 인구 추정에 사용한 기법을 변형할 수 있는지에 관하여 검토하였으나 1세대 다가구, 다세대 1가구가 적지 않아 이 역시 심층연구가 추후에 더욱 필요하다고 본다.

본 보고서에서 고려해 본 방법은 매칭된 자료로부터 얻은 세대 당 가구수를 이용하는 것이다.

2. 추정기법1-세대 당 가구수 이용방안

먼저 1장에서 얻은 주소매칭방법에 의한 세대와 가구의 매칭 결과, 구성원 매칭 방법에 근거한 세대와 가구 매칭 결과를 이용하여 얻은 세대수 대 가구수별 세대건수와 그 비율을 다음 표에서 정리하였다.

<표 3-21> 세대-가구 유형별 세대수

유형	주소매칭		구성원매칭	
	세대수	%	세대수	%
1-1	16,660	66.3	23,527	94.7
1-2	353	1.4	83	0.3
1-3	29	0.1	1	0.0
2-1	2,006	8.0	1,046	4.2
3-1	555	2.2	69	0.3
4-1	184	0.7	0	0.0
*기타	5,358	21.3	115	0.5
총계	25,145	100.0	24,841	100.0

이 결과를 이용하여 행정자료에 있는 전체 세대수에 <표 3-21>에서 구한 세대수 대 가구수별 비율을 적용하여 행정동별 가구수를 추정하였다. 그 결과는 다음의 표와 같다. 여기서 가중치는 해당 유형에 속한 세대가 매칭된 가구수를 의미한다.

<표 3-22> 주소매칭 가구수 추정

행정구역	세대수	유형	1-1	1-2	1-3	2-1	3-1	4-1	기타	추정 가구수
		% 가중치	66.3 1.0	1.4 2.0	0.1 3.0	8.0 0.5	2.2 0.3	0.7 0.3	21.3 0.8	
문흥1동	6,631		4,393	186	23	265	49	12	1,175	6,103
월곡1동	5,767		3,821	162	20	230	42	11	1,022	5,308
교동	6,486		4,297	182	22	259	48	12	1,149	5,970
소이면	1,527		1,012	43	5	61	11	3	271	1,405
활천동	8,671		5,745	243	30	346	64	16	1,537	7,981
거류면	2,389		1,583	67	8	95	18	4	423	2,199
합계	31,471		20,851	884	109	1,255	232	58	5,577	28,965

<표 3- 23 > 구성원 매칭 가구수 추정

행정구역	세대수	유형	1-1	1-2	1-3	2-1	3-1	4-1	기타	추정 가구수
		%	가중치							
			94.7	0.3	0.0	4.4	0.3	0.0	0.5	
			1.0	2.0	3.0	0.5	0.3	0.3	0.0	
문흥1동	6,631		6,280	44	1	140	6	0	31	6,502
월곡1동	5,767		5,462	39	1	121	5	0	27	5,655
교동	6,486		6,143	43	1	137	6	0	30	6,360
소이면	1,527		1,446	10	0	32	1	0	7	1,497
활천동	8,671		8,212	58	1	183	8	0	40	8,502
거류면	2,389		2,263	16	0	50	2	0	11	2,342
합계	31,471		29,806	210	4	663	30	0	146	30,858

위의 두 표에 나타난 세대별 가구수 계산과정에서 소수처리문제 때문에 추정 가구수의 합과 차이가 나타난다. 이 두 개의 추정치를 실제 시험조사 자료와 비교한 결과를 다음의 두 개의 표에 나타내었다.

<표 3- 24> 주소매칭 가구수 추정 비교

행정구역	세대수	가구수	추정 가구수	A	B	A/B
문흥1동	6,631	6,718	6,103	615	87	7.07
월곡1동	5,767	5,454	5,308	146	-313	-0.47
교동	6,486	6,036	5,970	66	-450	-0.15
소이면	1,527	1,433	1,405	28	-94	-0.29
활천동	8,671	10,308	7,981	2,327	1637	1.42
거류면	2,389	2,092	2,199	-107	-297	0.36
합계	31,471	32,041	28,965	3,076	570	5.40

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

<표 3- 25> 구성원 매칭 가구수 추정 비교

행정구역	세대수	가구수	추정 가구 수	A	B	A/B
문흥1동	6,631	6,718	6,502	216	87	2.49
월곡1동	5,767	5,454	5,655	-201	-313	0.64
교동	6,486	6,036	6,360	-324	-450	0.72
소이면	1,527	1,433	1,497	-64	-94	0.68
활천동	8,671	10,308	8,502	1,806	1,637	1.10
거류면	2,389	2,092	2,342	-250	-297	0.84
합계	31,471	32,041	30,858	1,183	570	2.08

A : 조사자료 - 추정치

B : 조사자료 - 행정자료

A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

위에서 보면 문흥1동과 활천동은 구성원 매칭 방식을 사용한 것이 시험조사 자료에 근접하고 다른 지역은 두 경우 모두 유사하다고 볼 수 있다.

3. 추정기법2-가구주수 이용 방안

또 다른 기법으로 가구주의 수를 추정함으로써 가구수를 추정하는 것을 제안한다.

인구매칭에서 얻은 결과로부터 주민등록 자료에 나타난 세대주관계 중 가구주로 매칭된 인구 비율을 이용하는 것인데 6개 행정동으로부터 얻은 결과는 다음의 표와 같다.

<표 3-26> 행정동별 관계에 따른 가구주 현황

(단위 : 명. %)

행정구역		세대주	세대주의 배우자	자녀와 그 배우자	세대주나 그 배우자의 부모	기타	합계
문흥1동	총계	5,543	3,351	7,525	334	503	17,256
	가구주	4,692	76	179	58	26	5,031
	비율	84.6	2.3	2.4	17.4	5.2	
월곡1동	총계	4,317	2,449	5,698	227	383	13,074
	가구주	3,741	78	143	37	29	4,028
	비율	86.7	3.2	2.5	16.3	7.6	
교동	총계	5,248	3,277	5,748	339	321	14,933
	가구주	4,617	93	170	42	17	4,939
	비율	88	2.8	3	12.4	5.3	
소이면	총계	798	425	331	61	73	1,688
	가구주	698	9	36	13	1	757
	비율	87.5	2.1	10.9	21.3	1.4	
활천동	총계	6,993	4,658	9,086	403	682	21,822
	가구주	6,391	96	208	46	48	6,789
	비율	91.4	2.1	2.3	11.4	7	
거류면	총계	1,067	534	600	58	93	2,352
	가구주	967	3	26	10	6	1,012
	비율	90.6	0.6	4.3	17.2	6.5	
합계	총계	23,966	14,694	28,988	1,422	2,055	71,125
	가구주	21,106	355	762	206	127	22,556
	비율	88.1	2.4	2.6	14.5	6.2	

6개 행정동의 주민등록 자료 중 시험조사 자료와 매칭된 세대원은 총 71,125명인데, 이 중 23,966명이 세대주로 되어 있다. 시험조사 자료에서 확인한 바 이들 중 88.1%인 21,106명만이 가구주로 되어있다. 이 기법을 검증하기 위하여 먼저 각 구역을 제외한 나머지로 부터 비율을 얻고, 이를 제외한 구역에 적용하여 추정을 시도한다.

<표 3-27> 각 행정동을 제외한 비율 추정

행정구역		세대주	세대주의 배우자	자녀와 그 배우자	세대주나 그 배우자의 부모	기타	합계
문흥1동 제외	총계	18,426	11,343	21,464	1,088	1,553	53,874
	가구수	16,414	279	583	148	101	17,525
	비율	89.1	2.5	2.7	13.6	6.5	
월곡1동 제외	총계	19,652	12,245	23,291	1,195	1,672	58,055
	가구수	17,365	277	619	169	98	18,528
	비율	88.4	2.3	2.7	14.1	5.9	
교동 제외	총계	18,721	11,417	23,240	1,083	1,735	56,196
	가구수	16,489	262	592	164	110	17,617
	비율	88.1	2.3	2.5	15.1	6.3	
소이면 제외	총계	23,169	14,269	28,658	1,361	1,983	69,440
	가구수	20,408	346	726	193	126	21,799
	비율	88.1	2.4	2.5	14.2	6.4	
활천동 제외	총계	16,975	10,036	19,903	1,019	1,374	49,307
	가구수	14,715	259	554	160	79	15,767
	비율	86.7	2.6	2.8	15.7	5.7	
거류면 제외	총계	22,902	14,160	28,389	1,364	1,963	68,778
	가구수	20,139	352	736	196	121	21,544
	비율	87.9	2.5	2.6	14.4	6.2	

문흥1동의 경우 위에서 문흥1동을 제외하고, 5개 구역으로부터 추정된 세대주 관계별 가구주 비율을 문흥1동의 행정자료에 나타난 세대주 관계별 인구에 적용한 값이 아래의 표에 나타난 바와 같이 총 6,390가구로 추정된다. 구체적으로 보면, 문흥1동에 세대주가 6,631명 있었는데, 이중 5,907명이 가구주일 것으로 추정하고, 행정 자료에 부모로 되어 있는 637명 중 87명이 가구주일 것으로 추정한다.

<표 3-28> 각 행정동의 가구수 추정

행정구역		세대주	세대주의 배우자	자녀와 그 배우자	세대주나 그 배우자의 부모	기타	합계
문흥1동		6,631	3,689	9,106	637	895	20,958
	추정	5,907	91	247	87	58	6,390
월곡1동		5,767	2,778	7,077	369	646	16,637
	추정	5,096	63	188	52	38	5,437
교동		6,486	3,654	7,215	532	547	18,434
	추정	5,713	84	184	81	35	6,096
소이면		1,527	689	894	114	199	3,423
	추정	1,345	17	23	16	13	1,413
활천동		8,671	5,112	10,696	658	1,245	26,382
	추정	7,517	132	298	103	72	8,121
거류면		2,389	953	1,529	138	334	5,343
	추정	2,101	24	40	20	21	2,205

이상과 같은 3개의 가구수 추정치를 다음의 표로 비교하였다.

<표 3-29> 가구수 추정치 비교

행정구역	세대수	가구수	추정치1-1	추정치1-2	추정치2
문흥1동	6,631	6,718	6,103	6,509	6,390
월곡1동	5,767	5,454	5,308	5,661	5,437
교동	6,486	6,036	5,970	6,367	6,096
소이면	1,527	1,433	1,405	1,499	1,413
활천동	8,671	10,308	7,981	8,512	8,121
거류면	2,389	2,092	2,199	2,345	2,205
합계	31,471	32,041	28,965	30,894	29,661

추정치1-1 : 주소 매칭을 통한 가구수 추정

추정치1-2 : 구성원 매칭을 통한 가구수 추정

추정치 2 : 가구주수를 이용한 가구수 추정

3개의 추정치가 큰 차이를 나타내지는 않는다. 특이한 사항은 활천동 지역에 대해서는 3개의 추정치 모두 상대적으로 시험조사 자료와 큰 차이를 나타내고 있으므로, 이 지역에 대한 특성을 좀 더 조사할 필요가 있다.

제6절 주택수의 지역별 특성별 추정기법

건축물대장으로부터 주택수를 추정하는 기법 역시 회귀모형과 CR모형을 이용한 기법으로 추진할 수 있다고 본다. 회귀추정기법은 2008년 시안에서 소개하고 있기 때문에 여기서는 CR모형을 이용한 기법을 검토한다.

1. CR모형을 이용한 기법

가. 자료 생성

인구 추정에서와 마찬가지로 2, 3차 시험조사 대상인 12개 행정동의 주택종류별 면적별 주택수의 추정을 최소추정대상군으로 한다. (여러번 언급하지만 조사구로 구성된 행정구역의 주택종류별 면적별 주택수 추정의 구도로 보면 된다.) 인구추정에서 수행한 바와 같이 3개 행정동을 추출하여 이들로부터 각 부분군의 매칭된 건축물대장 건수를 U^{11} , 매칭되지 않은 건축물대장 건수 중 5%를 구역 내에 있으나 조사에서 찾지 못한 건축물이라고 가정하면서 이를 U^{12} 로 하고, 나머지를 건축물대장에는 있으나 존재하지 않는 주택수 X 로 나타내며 집계하였다. 그리고 시험조사 자료에는 존재하나 건축물대장에 존재하지 않는 건수를 U^{21} 이라 한다.

나. 추정

표본으로부터 얻은 U^{11}, U^{12}, U^{21} 와 X 로부터 과소집계 모수(p_{1+})를 추정하고, 이를 이용하여 과다집계 모수(λ)를 추정하여 각 부분군에 속한 개체(건축물대장의 건축물)에 가중치(ω)를 부여하였다.

12개 행정동 전체의, 주택종류(j), 면적(k)을 이용하여 건축물 대장의 총 건수(T_{xjk})에서 실제 건축물 건수(T_{yjk})의 추정치는 $\widehat{T}_{yjk} = T_{xjk} * \omega_{jk}$ 를 통하여 얻고, 그에 따른 표준오차를 산출하였다.

한편, 3개 행정동의 표본자료로부터 건축물대장이나 매칭되지 않은 시험조사 자료에 관찰치가 없는 부분군이 발생하는 경우 모수와 표준오차가 추정되지 않기 때문에 그런 경우에는 건축물 대장의 값으로 대체하였다.

다. 분석

건축물대장에서 주택종류는 8개(11, 12, 13, 2, 3, 4, 5, 9)의 항목이 있는 반면 시험조사 자료에는 4개가 추가되어 12개(11, 12, 13, 2, 3, 4, 5, 6, 7, 8, 9, 10)가 있다. 따라서, 건축물대장을 기준으로 추정치를 유도하는 과정에서는 시험조사 자료의 주택종류 6, 7, 8, 10을 제외하였다.

(주택종류 11 : 단독주택-일반가구, 12 : 단독주택-다가구 단독, 13 : 단독주택-영업겸용 단독, 2 : 아파트

3: 연립주택, 4 : 다세대 주택, 5 : 비거주용 건물 (상가, 공장, 여관 등)내 주택, 6 : 오피스텔, 7 : 호텔, 여관 등 숙박업소의 객실, 8 : 기숙사 및 특수 사회시설, 9 : 판잣집, 비닐하우스, 10 : 기타)

대체적으로 적절한 추정치를 얻었으나, 특이한 현상으로 주택종류가 아파트이며, 면적이 $46m^2$ 이상 $63m^2$ 미만인 경우와, $63m^2$ 이상 $96m^2$ 미만인 경우에 시험조사 자료에 비해 347개, 1,526개가 더 크게 추정되었다. 이러한 현상은 표본으로 뽑은 3개의 행정동에서 아파트이면서 면적이 $46m^2$ 이상 $63m^2$ 미만과 $63m^2$ 이상 $96m^2$ 미만의 시험조사 자료에서 109개, 190개가 행정자료와 매칭이 되지 않는 것으로 나타났기 때문에 CR모형의 특성상 실제 주택수의 추정치가 크게 된 것이다. 그런데 아파트가 이와 같이 다량 매칭이 되지 않는 점이 의심이 되어 매칭 과정을 재점검한 결과 해당되는 아파트가 건축물대장에서 주택종류가 판잣집, 비닐하우스로 분류된 사실을 발견하였다. 재매칭 작업결과 실제 매칭되지 않은 건수 2개와 14개로 파악되었다. 이에 따라 주택종류가 아파트인 경우와 판잣집, 비닐하우스의 추정치들을 재계산 한 결과 이들 역시 합리적인 추정결과를 보인다.

<표 3-30> 주택 CR추정

주택종류	면적	전체		표본				가 중 치	추 정 치	표준 오차
		행정자료	조사자료	U^{11}	U^{12}	U^{21}	X			
단독주택-일반가구	1	589	190	19	3	5	36	0.48	282	18.19
	2	784	353	59	3	6	69	0.52	409	19.32
	3	1,769	1,027	219	10	88	195	0.76	1,339	44.37
	4	1,922	1,530	420	16	161	254	0.87	1,680	49.91
	5	2,858	3,021	857	16	339	324	1.02	2,909	65.77
	6	1,351	1,396	488	3	65	68	1	1,345	30.12
	7	710	682	421	0	12	19	0.98	699	11.99
	8	808	751	236	0	10	9	1	811	13.62
	9	78	69	14	0	1	5	0.79	62	6.75
	10	26	18	3	0	1	1	1	26	5.84
합계	10,895	9,037	2,736	51	688	980	0.93	10,088	107.91	
단독주택-다가구단독	1	0	6	0	0	0	0		0	
	2	18	26	0	0	0	0		18	
	3	0	33	0	0	1	0		0	
	4	3	65	0	0	4	0		3	
	5	8	108	1	0	11	0	12	96	56.28
	6	35	114	11	0	10	1	1.75	61	12.93
	7	41	64	17	0	4	2	1.11	45	6.44
	8	306	349	53	0	6	1	1.09	334	11.39
	9	78	88	6	0	3	0	1.5	117	13.25
	10	139	136	4	0	5	4	1.13	156	20.26
합계	628	989	92	0	44	8	1.36	854	36.28	
단독주택-영업검용	1	11	12	0	0	4	0		11	
	2	15	16	1	0	0	2	0.33	5	1.83
	3	47	63	8	0	4	3	1.09	51	9.3
	4	62	97	6	0	8	2	1.75	109	19.66
	5	229	265	73	0	16	9	1.09	249	14.78
	6	221	290	104	0	8	7	1.01	223	9.37
	7	264	295	77	0	6	2	1.05	277	9.15
	8	407	392	54	0	7	3	1.07	436	15.02
	9	194	177	18	0	1	0	1.06	205	5.84
	10	44	32	0	0	0	0		44	
합계	1,494	1,639	341	0	54	28	1.07	1,599	32.23	
아파트	1	88	93	1	0	1	0	2	176	22.98
	2	630	558	1	0	1	0	2	1,260	61.48
	3	1,692	1,543	810	2	40	11	1.04	1,752	18.06
	4	11,618	11,622	3,029	1	109	17	1.03	11,969	38.53
	5	12,873	12,991	1,470	0	190	2	1.13	14,517	75.37
	6	2,302	2,333	294	0	28	0	1.1	2,521	26.84
	7	916	893	132	0	0	0	1	916	0
	8	81	50	0	0	0	0		81	
	9	63	34	0	0	0	0		63	
	10	0	0	0	0	0	0		0	
합계	30,263	30,117	5,737	3	369	30	1.06	32,042	81.56	

<표 3-30> 주택 CR추정

주택 종류	면적	전체		표본				가중치	추정치	표준 오차
		행정자 료	조사자 료	U^{11}	U^{12}	U^{21}	X			
연립 주택	1	3	4	0	0	0	0		3	
	2	4	11	0	0	0	0		4	
	3	34	106	0	2	24	14		34	
	4	842	880	367	0	127	34	1.23	1037	35.03
	5	354	533	83	0	7	4	1.03	366	11.69
	6	22	122	0	0	1	0		22	
	7	3	5	0	0	0	0		3	
	8	54	47	0	0	0	0		54	
	9	7	2	0	0	0	0		7	
	10	0	0	0	0	0	0		0	
합계		1323	1710	450	2	159	52	1.21	1606	44.52
다세 대주 택	1	31	17	1	0	0	0	1.00	31	0.00
	2	45	31	4	0	0	1	0.80	36	4.16
	3	363	242	49	0	2	6	0.93	337	11.61
	4	1911	1436	230	2	21	53	0.89	1698	33.98
	5	3085	3224	311	0	51	50	1.00	3094	49.18
	6	21	128	4	0	3	1	1.40	29	8.19
	7	7	22	0	0	0	0		7	
	8	8	13	0	0	1	0		8	
	9	0	3	0	0	0	0		0	
	10	2	6	0	0	0	0		2	
합계		5473	5122	599	2	78	111	0.95	5221	61.58
비거 주용 건물	1	32	34	4	0	4	1	1.60	51	12.14
	2	15	15	4	0	0	1	0.80	12	2.40
	3	71	68	10	0	7	5	1.13	80	12.90
	4	99	97	26	0	6	7	0.97	96	10.02
	5	258	243	97	0	12	16	0.96	249	13.04
	6	236	194	83	0	4	13	0.91	214	10.09
	7	157	135	49	0	3	12	0.85	134	9.12
	8	101	94	26	0	2	4	0.93	94	7.18
	9	31	27	5	0	0	2	0.71	22	3.68
	10	5	6	0	0	0	0		5	
합계		1005	913	304	0	38	61	0.94	942	26.22
관갓 집, 비 닐하 우스	1	3	34	0	0	1	3		3	
	2	8	5	1	0	0	2	0.33	3	1.33
	3	140	25	13	3	0	112	0.13	18	2.40
	4	283	172	163	2	0	110	0.60	170	11.06
	5	466	286	267	10	0	170	0.62	289	14.29
	6	33	5	0	1	0	27		5	
	7	1	0	0	0	0	0		0	
	8	1	0	0	0	0	0		0	
	9	0	0	0	0	0	0		0	
	10	0	0	0	0	0	0		0	
합계		935	527	444	16	1	424	0.52	488	19.16

*면적- 1 : $23m^2$ 미만, 2 : $23m^2$ 이상 $30m^2$ 미만, 3 : $30m^2$ 이상 $46m^2$ 미만, 4 : $46m^2$ 이상 $63m^2$ 미만,
 5 : $63m^2$ 이상 $96m^2$ 미만, 6 : $96m^2$ 이상 $129m^2$ 미만, 7 : $129m^2$ 이상 $162m^2$ 미만,
 8 : $162m^2$ 이상 $228m^2$ 미만, 9 : $228m^2$ 이상 $327m^2$ 미만, 10 : $327m^2$ 이상

<표 3-31> 주택 CR추정 결과

주택 종류	면적	A	B	A/B	주택 종류	면적	A	B	A/B
단독 주택 - 일반 가구	1	-92	-399	0.23	연립 주택	1	1	1	1
	2	-56	-431	0.13		2	7	7	1
	3	-312	-742	0.42		3	72	72	1
	4	-150	-392	0.38		4	-157	38	-4.14
	5	112	163	0.69		5	167	179	0.93
	6	51	45	1.14		6	100	100	1
	7	-17	-28	0.6		7	2	2	1
	8	-60	-57	1.06		8	-7	-7	1
	9	7	-9	-0.82		9	-5	-5	1
	10	-8	-8	1		10	0	0	
합계	-1,051	-1,858	0.57	합계	104	387	0.27		
단독 주택 - 다가 구 단독	1	6	6	1	다세 대주 택	1	-14	-14	1
	2	8	8	1		2	-5	-14	0.36
	3	33	33	1		3	-95	-121	0.78
	4	62	62	1		4	-262	-475	0.55
	5	12	100	0.12		5	130	139	0.94
	6	53	79	0.67		6	99	107	0.92
	7	19	23	0.81		7	15	15	1
	8	15	43	0.34		8	5	5	1
	9	-29	10	-2.9		9	3	3	1
	10	-20	-3	6.79		10	4	4	1
합계	135	361	0.37	합계	-99	-351	0.28		
단독 주택 - 영업 검용	1	1	1	1	비거 주용 건물	1	-17	2	-8.6
	2	11	1	11		2	3	0	
	3	12	16	0.73		3	-12	-3	4.16
	4	-12	35	-0.33		4	1	-2	-0.5
	5	16	36	0.46		5	-6	-15	0.39
	6	67	69	0.97		6	-20	-42	0.47
	7	18	31	0.57		7	1	-22	-0.05
	8	-44	-15	2.9		8	0	-7	0.04
	9	-28	-17	1.63		9	5	-4	-1.21
	10	-12	-12	1		10	1	1	1
합계	40	145	0.27	합계	-29	-92	0.31		
아파 트	1	-83	5	-16.6	관갓 집, 비닐 하우 스	1	31	31	1
	2	-702	-72	9.75		2	2	-3	-0.78
	3	-209	-149	1.4		3	8	-115	-0.07
	4	-347	4	-86.73		4	2	-111	-0.02
	5	-1,526	118	-12.93		5	-3	-180	0.02
	6	-188	31	-6.07		6		-28	
	7	-23	-23	1		7		-1	
	8	-31	-31	1		8		-1	
	9	-29	-29	1		9		0	
	10	0	0			10		0	
합계	-1,925	-146	13.19	합계	39	-408	-0.1		

*A : 조사자료 - 추정치

*B : 조사자료 - 행정자료

*A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

*면적- 1 : 23m²미만, 2 : 23m² 이상 30m²미만, 3 : 30m² 이상 46m²미만, 4 : 46m² 이상 63m²미만,
5 : 63m² 이상 96m²미만, 6 : 96m² 이상 129m²미만, 7 : 129m² 이상 162m²미만,
8 : 162m² 이상 228m²미만, 9 : 228m² 이상 327m²미만, 10 : 327m² 이상

앞에서 언급한 아파트의 면적(4)-46m² 이상 63m²미만, 면적(5)-63m² 이상 96m²미만의 문제는 건축물 대장의 오류에 의하여 야기된 것이지만 추가적으로 CR모형이 갖고 있는 표본의 동질성 가정의 관점에서 검토하여 보았다.

검토결과는 표본으로 추출된 지역에서 얻은 $(U^{11}(i), U^{12}(i), U^{21}(i), X(i)), i = 1, 2, 3$ 을 통합하는 과정에서 이들의 동질성 검정을 수행하지 않고 통합시켰다는 것이다.

추정에 사용된 표본 행정구역에서 얻은 자료로부터 동질성 검정을 위하여 각 행정동별로 자료를 추출하였다. 면적(4)- $46m^2$ 이상 $63m^2$ 미만, 면적(5)- $63m^2$ 이상 $96m^2$ 미만 모두에서 명장1동의 U^{21} 이 109건, 181건으로 다른 두 개의 행정동 0건, 0건과, 0건 9건에 비하여 상당히 큰 값이 나타났다. (실제는 자료오류에 의한 것임) 따라서 이런 경우에는 명장1동을 제외시키고 다른 지역을 추가로 표집하는 것이 좋다고 본다. 표본 교체시에 어떤 변화가 나타나는 지를 보기 위하여 명장1동 대신에 범천1동을 표본으로 추출하여 CR모형을 다시 수행하였다. 새로운 표본으로 계산된 CR모형에서는 A/B값이 이전의 표본보다 작아진 것을 볼 수 있다. 하지만 여전히 주택종류(2)아파트의 면적(4)- $46m^2$ 이상 $63m^2$ 미만인 주택의 A/B값이 크게 나타난다. 이는 U^{21} , 즉 시험조사 자료에는 존재하나 건축물 대장에는 아파트로 존재하지 않는 주택이 존재하기 때문에 U^{21} 값이 크게 나타났다. 새로 뽑은 표본의 범천1동에서도 이와 같은 현상이 나타나 명장1동에 비해서는 많은 건수가 나타나지 않았다.

주택종류-아파트, 면적- $46m^2$ 이상 $63m^2$ 미만

	U^{11}	U^{12}	U^{21}	X
명장1동	226	0	109	3
교동	2,743	0	0	5
거류면	60	1	0	9

주택종류-아파트, 면적- $63m^2$ 이상 $96m^2$ 미만

	U^{11}	U^{12}	U^{21}	X
명장1동	293	0	181	0
교동	803	0	0	1
거류면	374	0	9	1

주택종류=아파트, 면적=46m² 이상 63m²미만

	U ¹¹	U ¹²	U ²¹	X
범천1동	79	2	52	22
교동	2,743	0	0	5
거류면	60	1	0	9

주택종류=아파트, 면적=63m² 이상 96m²미만

	U ¹¹	U ¹²	U ²¹	X
범천1동	780	1	10	14
교동	803	0	0	1
거류면	374	0	9	1

주택종류	면적	전체		표본				가치	추정치	표준오차
		행정자료	조사자료	U ¹¹	U ¹²	U ²¹	X			
2	4	11,618	11,622	2,882	3	52	36	1.0	11,555	32.08
	5	12,873	12,991	1,957	1	19	16	1.0	12,854	26.07

주택종류	면적	조사-추정(A)	조사-행정(B)	A/B
2	4	67.1	4	16.8
	5	137.5	118	1.2

2. 비 추정을 이용한 기법

인구 추정과 마찬가지로 12개의 행정동에서 선택된 3개의 행정동의 건축물 대장 자료와 시험조사 자료로부터 다음과 같은 자료를 구한다.

행정동($i = 1, \dots, m$), 주택종류($j = 11, 12, 13, 2, \dots, 10$), 면적($k = 1, \dots, 10$)에 따라 행정자료(x_{ijk})와 조사자료(y_{ijk})로부터 다음의 주택수를 집계한다.

$$t_{xjk} = x_{1jk} + x_{2jk} + x_{3jk} : \text{주택종류와 면적이 같은 3개 행정동의 행정자료의 합}$$

$$t_{yjk} = y_{1jk} + y_{2jk} + y_{3jk} : \text{주택종류와 면적이 같은 3개 행정동의 조사자료의 합}$$

$$\frac{t_{yjk}}{t_{xjk}} = b_{jk} : \text{행정자료의 합과 조사자료 합의 비율}$$

위의 주택수와 비율을 이용하여 전체 건축물 대장의 주택수로부터 추정을 수행하였다.

수행한 결과는 <표3-32>에 요약하였는데 조사자료와 추정치, 행정자료의 차이를 각각 A, B로 하여 A/B를 통하여 추정치가 행정자료를 직접 사용하는 것보다 더 바람직한지를 조사하였다. 기술적으로는 비 추정치가 행정자료를 직접 사용하는 것보다 더 좋다고 할 수는 없다. 그러나 표본의 대표성 문제가 내재된 상황이기 때문에 아직은 비추정이 유용하지 않다는 결론을 내릴 수 없다. 두드러진 대표성문제는 주택종류(12)-단독주택-다가구 단독의 면적(4)-46m² 이상 63m²미만, 면적(5)-63m² 이상 96m²미만에서 표본자료와 전체자료간의 괴리에서 보인다.

이들을 자세히 조사한 결과 주택종류(12)의 면적(4)-46m² 이상 63m²미만, 면적(5)-63m² 이상 96m²미만에 속하는 조사자료를 각 행정동별로 살펴보았다. 그 결과 이 부분군의 주택수 129건과 290건의 대부분이 범천1동에서 89건 113건이 발생한 것을 발견하였다.

또한 주택종류(12)-단독주택-다가구 단독인 상황은 모든 면적에서 조사자료가 건축물 대장에 비하여 많은 것으로 나타났다. 이는 조사에서 확인된 실질적 다가구가 건축물대장 자료에서 단독으로 분류되어 있다고 보인다.

<표 3-32> 주택 비 추정 결과

	표본			전체			표준오차	A	B	A/B	
	행정	조사	b	행정	조사	추정치					
단독주택 - 일방구	1	58	11	0.19	589	61	112	176.48	-51	-528	0.1
	2	131	23	0.18	784	131	138	413.15	-7	-653	0.01
	3	424	192	0.45	1,769	540	801	1,057.52	-261	-1229	0.21
	4	690	475	0.69	1,922	1,288	1,323	1,179.23	-35	-634	0.06
	5	1,197	1,151	0.96	2,858	3,194	2,748	252.48	446	336	1.33
	6	559	357	0.64	1,351	1,337	863	995.15	474	-14	-33.87
	7	440	121	0.28	710	380	195	1,494.46	185	-330	-0.56
	8	245	69	0.28	808	550	228	773.87	322	-258	-1.25
	9	19	8	0.42	78	79	33	44.65	46	1	46.16
	10	4	15	3.75	26	44	98	162.11	-54	18	-2.97
합계	3,767	2,422	0.64	10,895	7,604	7,005	5,698.11	599	-3291	-0.18	
단독주택 - 다가구 단독	1	0	3		0	15	0		15	15	1
	2	0	2		18	30	18		12	12	1
	3	0	7		0	67	0		67	67	1
	4	0	21		3	129	3		126	126	1
	5	1	118	118	8	290	944	38,975.87	-654	282	-2.32
	6	12	260	21.67	35	401	758	19,568.30	-357	366	-0.98
	7	19	339	17.84	41	438	732	23,697.55	-294	397	-0.74
	8	54	227	4.2	306	534	1,286	3,103.36	-752	228	-3.3
	9	6	30	5	78	124	390	415.48	-266	46	-5.78
	10	8	14	1.75	139	143	243	43.96	-100	4	-25.06
합계	100	1,021	10.21	628	2,171	6,412	1,996.15	-4241	1543	-2.75	
단독주택 - 영업용	1	0	5		11	15	11		4	4	1
	2	3	0	0	15	14	0	7.35	14	-1	-14
	3	11	9	0.82	47	70	38	18.48	32	23	1.37
	4	8	29	3.63	62	120	225	260.52	-105	58	-1.81
	5	82	119	1.45	229	347	332	258.93	15	118	0.12
	6	111	131	1.18	221	393	261	115.59	132	172	0.77
	7	79	78	0.99	264	273	261	24.80	12	9	1.37
	8	57	54	0.95	407	371	386	60.24	-15	-36	0.4
	9	18	18	1	194	152	194	6.00	-42	-42	1
	10	0	5		44	38	44		-6	-6	1
합계	369	448	1.21	1,494	1,793	1,814	336.95	-21	299	-0.07	
아파트	1	1	2	2	88	95	176	12.73	-81	7	-11.57
	2	1	1	1	630	558	630	0.00	-72	-72	1
	3	823	842	1.02	1,692	1,525	1,731	145.83	-206	-167	1.23
	4	3,047	3,107	1.02	11,618	11,231	11,847	2,612.73	-616	-387	1.59
	5	1,472	1,572	1.07	12,873	12,849	13,748	1,179.30	-899	-24	37.44
	6	294	203	0.69	2,302	2,226	1,589	705.30	637	-76	-8.38
	7	132	50	0.38	916	811	347	479.68	464	-105	-4.42
	8	0	0		81	52	81		-29	-29	1
	9	0	0		63	35	63		-28	-28	1
	10	0	0		0	2	0		2	2	1
합계	5,770	5,777	1	30,263	29,384	30,300	6,582.70	-916	-879	1.04	

*A : 조사자료 - 추정치

*B : 조사자료 - 행정자료

*A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

*면적- 1 : 23m²미만, 2 : 23m² 이상 30m²미만, 3 : 30m² 이상 46m²미만, 4 : 46m² 이상 63m²미만,
 5 : 63m² 이상 96m²미만, 6 : 96m² 이상 129m²미만, 7 : 129m² 이상 162m²미만,
 8 : 162m² 이상 228m²미만, 9 : 228m² 이상 327m²미만, 10 : 327m² 이상

<표 3-32> 주택 비 추정 결과(계속)

	표본			전체			표준오차	A	B	A/B	
	행정	조사	b	행정	조사	추정치					
연립주택	1	0	0	3	4	3		1	1	1	
	2	0	2	4	15	4		11	11	1	
	3	16	39	2.44	34	140	83	325.26	57	106	0.54
	4	401	300	0.75	842	642	630	1,604.76	12	-200	-0.06
	5	87	369	4.24	354	941	1,501	5,919.13	-560	587	-0.95
	6	0	6		22	206	22		184	184	1
	7	0	0		3	19	3		16	16	1
	8	0	1		54	51	54		-3	-3	1
	9	0	0		7	4	7		-3	-3	1
	10	0	0		0	0	0		0	0	0
합계	504	717	1.42	1,323	2,022	1,882	2,168.01	140	699	0.2	
다세대주택	1	1	1	31	19	31	0.00	-12	-12	1	
	2	5	0	45	26	0	21.21	26	-19	-1.37	
	3	55	48	0.87	363	204	317	55.62	-113	-159	0.71
	4	285	245	0.86	1,911	1,261	1,643	291.25	-382	-650	0.59
	5	361	340	0.94	3,085	3,186	2,906	169.59	280	101	2.78
	6	5	11	2.2	21	168	46	90.34	122	147	0.83
	7	0	2		7	23	7		16	16	1
	8	0	3		8	24	8		16	16	1
	9	0	1		0	5	0		5	5	1
	10	0	0		2	5	2		3	3	1
합계	712	651	0.91	5,473	4,921	5,004	1,196.54	-83	-552	0.15	
비주요용건	1	5	9	1.8	32	28	58	40.75	-30	-4	7.4
	2	5	2	0.4	15	12	6	12.09	6	-3	-2
	3	15	10	0.67	71	48	47	40.47	1	-23	-0.03
	4	33	12	0.36	99	59	36	72.17	23	-40	-0.58
	5	113	44	0.39	258	142	100	243.71	42	-116	-0.36
	6	96	49	0.51	236	111	120	191.60	-9	-125	0.08
	7	61	27	0.44	157	82	69	142.58	13	-75	-0.17
	8	30	22	0.73	101	65	74	54.05	-9	-36	0.25
	9	7	3	0.43	31	26	13	19.99	13	-5	-2.54
	10	0	2		5	7	5		2	2	1
합계	365	180	0.49	1,005	580	496	644.10	84	-425	-0.2	
관갓비합우스	1	3	1	0.33	3	41	1	11.31	40	38	1.05
	2	3	0	0	8	0	0	12.73	0	-8	0
	3	128	0	0	140	0	0	543.06	0	-140	0
	4	275	0	0	283	1	0	1,166.73	1	-282	0
	5	447	0	0	466	0	0	1,896.46	0	-466	0
	6	28	0	0	33	0	0	118.79	0	-33	0
	7	0	0		1	0	1		-1	-1	1
	8	0	0		1	0	1		-1	-1	1
	9	0	0		0	0	0		0	0	
	10	0	0		0	0	0		0	0	
합계	884	1	0	935	42	1	3,750.49	41	-893	-0.05	

*A : 조사자료 - 추정치

*B : 조사자료 - 행정자료

*A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

*면적 -1 : 23m²미만, 2 : 23m² 이상 30m²미만, 3 : 30m² 이상 46m²미만, 4 : 46m² 이상 63m²미만,

5 : 63m² 이상 96m²미만, 6 : 96m² 이상 129m²미만, 7 : 129m² 이상 162m²미만,

8 : 162m² 이상 228m²미만, 9 : 228m² 이상 327m²미만, 10 : 327m² 이상

앞에서 연구한 바와 같이 주택종류(11)-단독주택-일반가구, 주택종류(12)-단독주택-다가구 단독, 주택종류(13)-단독주택-영업겸용 단독의 구분이 건축물대장 자료와 조사자료에서 크게 다르다. 따라서 이를 하나로 통합하여 단독주택으로 하여 추정하였으며, 이는 <표 3-33>에 요약하였는데 면적(9)-228m² 이상 327m²미만과 면적(10)-327m² 이상을 제외하고는 대체적으로 좋은 결과를 보여준다.

<표 3-33> 주택 비 추정 결과(단독주택)

	표본			전체			표준오차	A	B	A/B	
	행정	조사	b	행정	조사	추정치					
단독 주택	1	58	19	0.33	600	91	197	167.93	-106	-509	0.21
	2	134	25	0.19	817	175	152	417.61	23	-642	-0.04
	3	435	208	0.48	1,816	677	868	1052.03	-191	-1139	0.17
	4	698	525	0.75	1,987	1,537	1,495	1060.97	42	-450	-0.09
	5	1,280	1,388	1.08	3,095	3,831	3,356	575.7	475	736	0.65
	6	682	748	1.1	1,607	2,131	1,763	357.63	368	524	0.7
	7	538	538	1	1,015	1,091	1,015	21.63	76	76	1
	8	356	350	0.98	1,521	1,455	1,495	107.14	-40	-66	0.61
	9	43	56	1.3	350	355	456	93.25	-101	5	-20.16
	10	12	34	2.83	209	225	592	218.67	-367	16	-22.95
합계	4,236	3,891	0.92	13,017	11,568	11,957	2309.81	-389	-1449	0.27	

*A : 조사자료 - 추정치

*B : 조사자료 - 행정자료

*A/B : (조사자료 - 추정치) / (조사자료 - 행정자료)

*면적- 1 : 23m²미만, 2 : 23m² 이상 30m²미만, 3 : 30m² 이상 46m²미만, 4 : 46m² 이상 63m²미만, 5 : 63m² 이상 96m²미만, 6 : 96m² 이상 129m²미만, 7 : 129m² 이상 162m²미만, 8 : 162m² 이상 228m²미만, 9 : 228m² 이상 327m²미만, 10 : 327m² 이상

제7절 결론

이 장에서는 주민등록 자료, 건축물 대장 두 개의 행정자료를 이용하여 지역에 의해서 결정되는 최소의 추정지역인 통계 구역에 존재하면서 특정한 성격을 갖는 인구와 가구수 그리고 구역 내의 주택수를 추정할 수 있는 방법론을 검토하였다. 또, 제2차 제3차 시험조사를 실시한 12개 읍면동의 자료에 이들 방법론을 적용하며 이 과정에서 발생하는 문제점과 향후 개선, 발전의 방향을 설정하기 위한 기초 자료가 되도록 하였다. 구체적인 추정방법으로는 행정자료와 조사자료의 연계 결과를 이용하지 않는 회귀모형을 이용하는 방법과 두 자료간의 연계결과를 이용하는 포획 재포획(CR : Capture-Recapture Model)을 제안하고 추가적으로 비 추정 방법과 그 외 직관적인 방안도 추가하였다.

구체적으로 얻은 결과를 요약하면 다음과 같다. 인구추정에서는 회귀모형과 CR모형 그리고 비(Ratio)모형을 적용하였는데 CR모형이 20대 인구 추정을 제외한 대부분의 연령대에서 다른 모형에 비하여 작은 상대오차를 보였다. 가구수 추정에서는 가구주 연계를 이용한 추정이 월곡동, 교동, 소이면에서 작은 오차를 보인 반면 다른 지역은 단순히 행정자료로 추정하는 것이 오차가 제일 작은 것으로 나타났다. 한편 가구주 연계방식이나, 행정자료와 같이 직관적이지 않은 방법으로는 회귀모형을 사용하는 것인데 각 행정지역 별로 보면 오차가 가장 작은 것은 아니지만 모든 지역에서 받아들일 수 있는 수준의 상대오차를 보였다. 주택수 추정에서는 주택종류별로 추정할 경우에는 회귀추정과 비추정이 CR추정에 비하여 작은 오차를 보였으며, 주거면적별로 추정한 경우에는 면적에 따라 회귀, 비, CR추정이 각각 우수한 것으로 나타났다. 이들 결과를 정리하면 어떠한 방법이 전체적으로 우수하다고 할 수 있는 상황은 아닌 것으로 나타났다. 따라서 각 모형이 갖고 있는 장점과 단점을 상호 보완하는 방향으로 추정결과를 결합시키는 것이 바람직하다고 본다. 예컨대 CR추정 방법이 연계결과의 정확성에 의존하는 단점을 여기에 의존하지 않는 회귀추정방법과 병행하여 사용하고, 가구수 추정에서는 가구주연계결과를 이용하는 직관적 방법을 일반적인 회귀추정방법에 대한 보충 방법으로 사용하는 것이 하나의 방안이라고 본다. 한편 본 장의 목적 중 하나인 추정방법의 개선 및 연구방향 모색의 결과는 다음과 같다.

첫째는 최소추정대상의 확정이 요구된다. 본 연구에서는 주어진 자료의 특

성에 따라 회귀모형의 인구 추정의 경우는 행정동이 최소추정지역이 되고 성, 연령의 인구특성을 고려하여 행정동별 남녀별, 연령대별 추정과정을 보였다. 반면 CR모형을 적용하기 위해서는 연계자료가 행정동별로 이루어 졌기 때문에 이들 행정동이 조사 추출 단위가 되어 12개 행정동 전체에 대한 성별, 연령별, 추정이 이루어 졌다. 따라서 CR모형을 활용하여 행정동별, 인구특성별 추정을 위해서는 행정동을 분할한 조사구의 표본조사 및 연계결과를 필요로 하게 된다. 또한 이러한 상황 때문에 본장에서 얻은 결과들은 실제보다 다소 오차가 작게 나온 면이 있다고 볼 수 있다.

둘째는 회귀모형을 이용하기 위한 준비과정으로 전국 3,572개의 행정읍면동에 대한 군집화(층화)과정의 연구가 요망된다. 본 연구에서는 12개 읍면동을 사용하였기 때문에 이에 대한 연구는 별로 이루어지지 않았다.

셋째는 자료연계에 필요한 실제 조사를 위하여 표본으로 추출될 조사지역을 어떤 수준(행정읍면동 수준 조사구 같은 소구역 수준)에서 어떤 표본추출 방법을 사용할 것인가 하는 문제와 이들 조사 방법(사후확인 조사 등)에 관한 연구가 심각하게 이루어져야 한다고 판단한다.

인구, 주택수, 가구수에 관한 추정방법들의 결과는 아래 표들과 같다.

<표 3-34> 인구 추정 결과 비교(연령대)

연령대	회귀자료		CR		비추정		추정				
	행정	조사	행정	조사	행정	조사	회 귀 추 정	CR 5% 추정	CR 1% 추정	비추정	
0~9	16,957	17,399	16,957	16,782	16,957	16,769	16,824	17,428	17,318	17,274	
10~19	26,716	25,199	26,716	25,548	26,716	25,573	26,248	25,838	25,755	25,779	
20~29	24,353	21,199	24,353	21,821	24,353	21,818	20,507	19,445	19,015	18,845	
30~39	30,908	28,880	30,908	29,133	30,908	28,913	29,668	29,264	28,933	28,623	
40~49	32,555	30,947	32,555	31,313	32,555	31,146	31,725	31,412	31,122	30,686	
50~59	22,714	21,454	22,714	21,670	22,714	21,892	22,378	22,090	21,955	22,150	
60~69	13,557	12,585	13,557	12,921	13,557	12,865	13,638	13,270	13,160	13,162	
70~79	7,925	7,430	7,925	7,718	7,925	7,849	7,984	7,890	7,782	7,957	
80~89	2,574	2,250	2,574	2,383	2,574	2,440	2,702	2,538	2,496	2,523	
90세이상	330	248	330	294	330	298	326	332	321	321	
총합계	178,589	167,591	178,589	169,583	178,589	169,563	172,000	169,254	167,618	167,196	

<표 3-35> 인구 추정 결과 비교(성별)

성별	회귀자료		CR		비추정		추정				
	행정	조사	행정	조사	행정	조사	회 귀 추 정	CR 5% 추정	CR 1% 추정	비추정	
남자	90,358	83,142	90,358	84,372	90,358	84,264	86,436	84,857	83,954	83,645	
여자	88,231	84,449	88,231	85,211	88,231	85,299	85,564	84,397	83,664	83,551	
총합계	178,589	167,591	178,589	169,583	178,589	169,563	172,000	169,254	167,618	167,196	

<표 3-36> 가구수 추정 결과 비교

행정동	회귀분석 원자료		추정				회귀추정 외	
	행정	조사	회귀추정	주소연계 추정	가구원연계 추정	가구주수 추정	행정	조사
문흥동	6,631	6,687	6,281	6,103	6,509	6,390	6,631	6,718
월곡동	5,767	5,408	5,382	5,308	5,661	5,437	5,767	5,454
교동	6,485	6,009	6,101	5,970	6,367	6,096	6,486	6,036
소이면	1,527	1,396	1,188	1,405	1,499	1,413	1,527	1,433
활천동	8,671	9,000	8,235	7,981	8,512	8,121	8,671	10,308
거류면	2,389	2,070	2,033	2,199	2,345	2,205	2,389	2,092
총합계	31,470	30,570	29,220	28,965	30,894	29,661	31,471	32,041

<표 3-37> 주택수 추정 결과 비교(주택종류)

주택종류	회귀자료		CR		비추정		추정		
	행정	조사	행정	조사	행정	조사	회 귀 추정	CR 추정	비추정
일반단독주택	6,315	7,576	10,895	9,037	10,895	7,604	6,263	10,088	7,005
다가구단독주택	542	2,156	628	989	628	2,171	3,108	854	6,412
영업겸용단독	1,310	1,786	1,494	1,639	1,494	1,793	1,897	1,599	1,814
아파트	30,473	29,779	30,263	30,117	30,263	29,384	30,037	32,042	30,300
연립주택	1,776	2,054	1323	1710	1,323	2,022	1,510	1,606	1,882
다세대주택	5,571	4,915	5473	5122	5,473	4,921	4,974	5,221	5,004
비거주용건물	691	573	1005	913	1,005	580	285	942	496
오피스텔	0	648	0	0	0	0	0	0	0
숙박업소	0	5	0	0	0	0	0	0	0
기숙사 및 사회시설	2	0	0	0	0	0	0	0	0
공동주택	0	1	935	527	935	42	0	488	1
기타	0	3	0	0	0	0	2	0	0
총합계	46,680	49,496	52,016	50,054	52,016	48,517	48,076	52,840	52,914

<표 3-38> 주택수 추정 결과 비교(면적별)

연면적	회귀자료		CR		비추정		추정		
	행정	조사	행정	조사	행정	조사	회 귀 추정	CR 추정	비추정
23.1미만	482	317	757	390	757	278	265	557	392
23.1~29.7미만	1,104	1,013	1,519	1015	1,519	786	1,120	1,747	796
29.7~46.2미만	3,021	3,439	4,116	3,107	4,116	2,594	3,189	3,611	3,017
46.2~62.7미만	15,555	14,491	16,740	15,899	16,740	14,731	15,311	16,762	15,707
62.7~95.7미만	18,778	21,058	20,131	20,671	20,131	20,949	18,517	21,769	22,279
95.7~128.7미만	3,664	4,847	4,221	4,582	4,221	4,842	4,109	4,420	3,659
128.7~161.7미만	1,925	2,020	2,099	2,096	2,099	2,026	2,416	2,081	1,615
161.7~227.7미만	1,587	1,645	1,766	1,696	1,766	1,647	2,112	1,818	2,118
227.7~326.7미만	382	425	451	400	451	425	849	476	700
326.7이상	182	241	216	198	216	239	188	233	392
총합계	46,680	49,496	52,016	50,054	52,016	48,517	48,076	52,840	52,914

제8절 정책적 건의 사항

본 장에서는 행정자료를 이용하여 실제 존재하는 개체(주민, 주택)의 빈도를 추정하는 방법들을 연구하였다. 고려된 방법들은 회귀추정, CR모형에 의한 추정, 비추정 등 이었는데 본 연구에서는 이들 추정법의 구현 가능성을 연구하였다. 이 방법들을 12개 읍면동에 적용한 결과는 이들 중 어느 것이 우수하다고 할 수는 없었다. 본 연구에서 얻은 정책적 제안사항은 다음과 같다.

첫째는 표본조사지역에 대한 행정자료와 조사자료의 매칭결과를 이용하여 추정을 수행하는 CR모형이용 방법을 적용하기 위해서는 매칭작업을 할 수 있도록 조사지역을 표본으로 추출하여 조사를 수행하여야 한다. 특히 본 연구의 실험결과에서 CR모형추정결과가 상대적으로 정확한 것으로 나타났고 2008년 이스라엘에서도 활용한 방법이기 때문에 정책적으로 심층 연구할 가치가 있다고 본다. 조사지역의 표본추출단위의 결정방법과 추출방법 등이 1,2장에서 언급한 연계 과제와 함께 종합적으로 연구되어야 한다고 본다. 예컨대 2009년 끝낸 시범조사 결과에 1,2,3장의 연구내용을 종합한 연구를 다른 등록센서스 추진에 관련되는 쟁점을 독립적으로 연구하는 것이 좋다고 본다.

둘째는 본 연구에서는 대체적으로 안정적인 추정 값을 제공한 회귀추정 방법에 대한 문제점을 보다 광범위한 군집분석 모의실험을 통하여 이론적으로 제기되고 있는 추정의 불안정성이 얼마나 나타나는지 확인할 필요가 있다고 본다.

<참고>

1. 제3장에서 직접 언급되지는 않았지만, 우리나라의 등록센서스 실시에 필요한 이원추정기법의 장단점은 영국의 One Number Census에서 자주 논의된다. Joint UNECE/Eurostat Meeting on Population and Housing Censuses, Geneva (13-15 May 2008)에서 발표된 논문 "United Kingdom 2011 Census Coverage Assessment and Adjustment Strategy"에서 추정방법의 3단계 전략 ① 표본추출 지역 내의 이원시스템 추정기법 적용(Dual System Estimation within Sampled Areas) ② 추정지역별 추정결과 작성(Estimation by Estimation Area)(추정지역은 집계난이도의 개념을 적용하여 센서스 본조사 집계범위를 추정하기 위하여 설정한 기초자치단체의 통계지역. 이 통계지역의 일부에서 표본을 추출하여 센서스 표본조사가 실시되도록 되어 있음) ③ 기초 자치단체 추정결과 작성(Local Authority District Estimation)을 참고할 필요가 있으며, 우리나라에도 적용가능성이 있는가에 대하여 추후의 연구과제가 될 것이다..

2. 위의 논문은 영국의 통계청(ONS)가 작성한 것으로, 2011년 영국 센서스에 대하여 센서스 포괄범위 및 보정전략을 기술한다. 영국의 센서스 프로젝트는 집계과정을 개선하고 고품질의 센서스를 생성하려는 다양한 구상을 포함하고 있다. 그러나 2011년 센서스가 여전히 인구나 가구를 누락하거나 중복하여 조사할 가능성이 있기 때문에, 사후조사("센서스 포괄범위 조사-Census Coverage Survey)"를 실시하여 센서스의 포괄범위를 평가하고, 인구나 가구를 추정하는 방법을 제시하여 영국의 센서스를 고품질 센서스로 만드는데 결정적인 기여를 하고자 하여 기획된 논문이다. 아마도, 영국의 2011년 센서스가 전통적 방식을 준수하지만, 방법론 자체의 특성은 우리나라 실시하려는 2015년의 등록센서스에서, 행정자료(주민등록, 건축물/주택대장)을 이용하여 인구, 가구, 주택의 집계결과를 개선하는데 그대로 이용될 수 있을 것이다.

제4장 매칭결과 자료 간 불일치 보완방법 및 자료보완에 따른 보정방법

제1절 머리말

제3장에서는 행정자료 활용을 위한 자료 매칭의 최적 방안 연구 및 지역별, 특성별, 인구, 가구, 주택의 추정기법을 연구하였다. 인구 부문에서는 시험조사 지역에 대하여 주소, 성명, 주민생년일, 성별, 가구주 관계를 매칭키(matching key)로 하여 주민등록 기록과 시험조사 결과를 매칭한 결과 주민등록을 기준으로 하면 매칭률이 74.7%(2차 시험조사), 80.5%(3차 시험조사)로 나타났고, 시험조사 결과를 기준으로 하면 매칭률은 83.3%(2차 시험조사), 80.4%(3차 시험조사)로 나타났다. 또한 주택 부문에서는 주소, 층, 건물명, 면적, 건축연도코드를 매칭키로 하여 건축물 대장과 시험조사 자료를 매칭한 결과 건축물 대장을 기준으로 매칭률이 69.8%(2차 시험조사), 90.8%(3차 시험조사)로 나타났고, 시험조사 자료를 기준으로 하면 71.7%(2차 시험조사), 90.6%(3차 시험조사)로 나타났다. 주택의 경우 아파트는 매칭률은 90%를 상회하는 반면 아파트 이외의 주택은 50% 미만(2차 시험조사), 72%(3차 시험조사)로 나타나 아파트가 아닌 주택의 매칭률은 낮은 것으로 나타났다.

앞 장의 연구 결과는 매칭키를 이용하여 정확 매칭(exact matching)과 판단 매칭(judgemental matching)을 거친 매칭률이기 때문에 주어진 자료(주민등록, 건축물 대장, 시험조사 결과)를 기반으로 하면 현실적으로 최선의 매칭률이라고 할 수 있다. 앞 장의 연구 결과를 토대로 본 장에서 논의할 내용은 (i) 두 자료의 매칭률을 높이는 방안으로 데이터 에디팅 방법을 소개하고, (ii) 지역별, 특성별 인구나 가구를 추정하는 방법으로 이스라엘에서 사용하는 집계범위 모형을 우리의 실정에 맞게 보정한 방법을 소개하며 (iii) 마지막으로 소지역별로 소지역 크기가 추정되었을 때 추정된 값만큼 행정데이터를 보정하는 방안으로 가중치 조정법과 대체법(inputation) 방법을 소개한다.

제2절 레코드 매칭률 향상 방안

1. 레코드 매칭 과정

가. 매칭할 파일 준비

두 자료의 레코드를 연결하기 위해서는 준비 작업이 필요하다. 먼저 연결키 변수를 정해야 하고, 각 연결키 변수에 대한 변수 표준화(standardization) 및 문법구분(parsing)을 하여 두 자료의 레코드를 자동적으로 비교할 수 있도록 하여야 한다. 앞 장에서 인구 매칭의 경우 주소(본번 표준화, 부분 표준화, 동명 표준화, 호명 표준화) 성명(성, 이름), 주민생년월(생년, 생월), 성별, 가구주와의 관계가 연결키에 해당한다.

나. 데이터 에디팅

다음 단계는 준비된 자료(주민등록, 건축물 대장, 시험조사 자료)에 대한 데이터 에디팅을 실시하는 것이다. 에디팅 과정에서는 데이터 오류 탐색과 함께 대체가 이루어진다. 데이터 에디팅을 마친 파일을 대상으로 레코드 매칭을 수행한다.

다. 레코드 매칭

레코드 연결은 두 자료를 레코드 단위로 연결하는 작업을 말한다. 레코드 연결 방법은 정확 매칭(exact matching), 판단매칭(judgemental matching), 확률매칭(probability matching) 등으로 나누어 생각할 수 있다.

정확 매칭을 수행하기 위해서는 주민등록번호나 여권번호 같은 유일한 식별자가 있어야 한다. 앞 장에서 인구의 경우 정확매칭에 해당하는 매칭키는 지역, 주소, 성명, 주민생년일, 성별, 가구주 관계가 모두 정확하게 일치하는 경우로 볼 수 있다 (단계 1).

24단계와 25단계의 경우 매칭키는 행정동, 성명, 주민생년, 성별로 주소는 고려하지 않는다. 동일한 행정동에서 주소는 다르지만 성명, 생년, 성별이 동일한 경우는 두 가지 경우로 미루어 짐작할 수 있다. 첫째는 이사를 간 경우이고, 둘째는 다른 사람일 가능성이다. 첫 번째의 경우는 주민등록 조사 시점과 시험조사 시점이 달라서 생길 가능성이 있고, 이 경우 확인이 되면 정확 매칭이라고 볼 수 있다. 두 번째 경우는 동명이인으로 다른 사람일 가능성이 있는데 이 경우는 정확매칭으로 보기보다는 확률매칭으로 분류되어야 할 것

이다.

확률 매칭은 정확 매칭과 달리 일부 매칭키 변수에 의하여 매칭을 하는 경우이다. 부분 매칭이기 때문에 완벽하게 일대일로 매칭이 되었다고 보장하기 어렵고 대신 확률적으로 매칭이 되었다고 할 수 있다. 위의 예에서 24단계와 25단계의 경우 매칭키는 행정동, 성명, 주민생년, 성별이 동일하면 주소를 고려하지 않고 매칭한다. 이 경우 이사를 가서 동일인일 수도 있고, 서로 다른 사람일 수도 있다. 동일인일 확률이 크면 매칭에는 큰 문제가 없을 것이고, 반면 동일인일 확률이 낮으면 매칭은 잘못되어 문제가 될 수 있다. 문제는 매칭된 사람이 동일인일 확률을 어떻게 계산하는가 하는 점이다.

이 문제에 대하여 펠레지-선터는 하나의 규칙을 제안하였다. 확률매칭에 대하여 확률을 계산하는 방법을 제안하고, 이에 근거하여 매칭으로 받아들일 것인지 혹은 아닌지에 대한 의사결정 방법을 제안하였다. 그 절차는 다음과 같다.

1단계. 파일A의 모든 레코드와 파일B의 모든 레코드에 대하여 에디팅을 실시한 후 표준화된 형태로 만든다.

2단계. 두 파일의 레코드를 서로 비교하여 상호 비교된 짝(pair)을 다음과 같이 구분한다.

$$M = \{(a,b) | a \in A, b \in B, (a,b) \text{는 정확 매칭된 짝}\}$$

$$U = \{(a,b) | a \in A, b \in B, (a,b) \text{는 정확 매칭되지 않은 짝}\}$$

그리고 부분 일치 패턴 γ 에 대하여 다음을 계산한다.

$$P[\gamma | \gamma \in M] = \text{정확 매칭된 짝 중에서 부분 매칭된 확률}$$

$$P[\gamma | \gamma \in U] = \text{정확 매칭되지 않은 짝 중에서 부분 매칭된 확률}$$

이제 그 비율을 구한다.

$$R = \frac{P[\gamma | \gamma \in M]}{P[\gamma | \gamma \in U]}$$

3단계. R 의 값을 보고 부분 매칭 결과를 매칭으로 받아들일지를 결정한다.

(i) 만일 R 의 값이 크면 매칭으로 받아들인다.

(ii) 만일 R 의 값이 작으면 매칭되지 않는 것으로 한다.

2. 비 매칭 원인

지리적으로 분명하게 구분이 되는 일정 지역을 가정하자. 현실적으로는 특정 조사구 또는 특정 읍면동이 될 것이다. 그리고 매칭을 위한 두 개의 파일

을 가지고 있다고 하자. 예를 들어 하나는 주민등록 기록 파일이고 다른 하나는 시험조사 파일이다. 두 파일은 만들어진 과정도 다르고 기록한 시점도 다르다. 두 파일에서 매칭에 이용되는 매칭키는 거쳐 주소, 가구, 가구 구성원 인구 속성(성명, 성별, 나이, 가구주 관계 등)이다. 이제 비매칭이 일어날 가능성에 대하여 살펴보자.

먼저 비매칭이 일어나는 형태를 크게 구분하면 첫째는 두 파일의 속성이 달라서 생기는 비매칭(비매칭 형태A)과, 낮은 자료 품질에서 오는 비매칭(비매칭 형태B)가 있다. 즉, 기록이 불분명하여 생기는 비매칭이다.

가. 비 매칭 형태 A

비매칭 형태A는 기록 시점(주민등록 기록) 혹은 조사시점(시험조사 시점)이 달라서 비매칭이 생기거나 혹은 주민등록인구와 상주인구의 개념이 달라서 생긴다. 주요 이유를 거쳐와 가구로 나누어 살펴볼 수 있다.

- 거쳐 : 한 지역에서 거처는 두 시점 간에 유지/소멸/생성된다. 건축을 새로 짓거나 기존 주택을 헐면 가구 및 가구원에 변화가 생긴다.
- 주민등록과 상주인구의 차이 : 대부분의 가구 구성원은 주민등록이 있는 거처에 거주하지만 그렇지 않은 경우도 있다. 예를 들어 주택 청약을 위하여 무주택자 가구주가 되기 위해서는 주민등록을 분리하여야 한다. 다른 가구에 세 들어 사는 독립 가구주가 되는 경우 주민등록과 실제 거주가 달라진다. 대학생인 경우와 입대를 한 경우도 비슷하다. 주민등록에는 기록이 있지만 실제 거주하지는 않기 때문이다.

나. 비 매칭 형태 B

주민등록 및 시험조사 자료에 기록이 정확하지 않으면 비 매칭이 발생한다. 표준화와 문법구분을 통하여 최대한 매칭을 하지만 원 자료가 부실하면 사후적으로 매칭을 하는 것은 한계가 있다. 이 경우는 자료의 품질을 높이는 방법을 모색하여야 한다.

3. 레코드 매칭률 향상 방안

비 매칭 형태A에서와 같이 두 레코드 파일의 생성시차가 달라서 생기는 비 매칭은 현실적으로 받아들일 수밖에 없다. 또한 주민등록 인구나 상주인구의 개념차이에서 오는 비매칭도 현실적으로 받아들이고 매칭 작업을 하여야 한다. 대신 소지역별로 상주인구를 추정할 때에 이러한 요소를 포함하는 추정식을 개발하는 것이 더 효과적이다. 이는 다음절에서 상술한다.

비매칭 형태B와 같이 자료의 낮은 품질에서 오는 비매칭은 자료의 품질을 높임으로써 매칭률을 높이는 방안을 모색할 수 있다. 이 경우 레코드 매칭률을 더 높이는 방안, 즉 불일치하는 자료의 수를 줄이는 방안은 세 가지로 나누어 생각해 볼 수 있다. 첫 번째 방안은 매칭키 자료를 더 확보하는 것이며, 두 번째 방안은 매칭 기법을 더 정밀하게 다듬는 것이고, 세 번째 방안은 원 자료(주민등록, 건축물 대장, 시험조사 자료)의 품질을 더 높이는 것이다.

첫째 방안과 둘째 방안은 현실적으로 개선하기 어려울 것으로 보인다. 왜냐하면 앞 장의 연구에서 주민등록과 건축물 대장 그리고 시험조사 자료에서 자료를 연결할 수 있는 연결키는 모두 사용했기 때문에 더 남은 연결키가 없기 때문이다. 또한 매칭 기법에서도 인구의 경우 2차 시험조사 자료는 25단계, 3차 시험조사 자료는 19단계에 걸쳐 매칭을 수행했고, 주택의 경우 아파트는 18단계(2차), 19단계(3차), 아파트가 아닌 주택은 무려 40단계(2차), 27단계(3차)에 걸쳐 매칭을 수행했기 때문이다. 앞 장에서도 자세한 설명이 있지만, 통상적으로 정밀매칭이라고 부를 수 있는 매칭은 1단계이며, 나머지 단계는 일부 매칭 변수가 일치하는 부분매칭이기 때문에 모든 경우가 정확하게 일대일로 자료가 연결되는 것은 아니다. 인구의 경우 2차 시험조사에서 25단계의 매칭 변수는 행정동과 성명, 성별 그리고 ± 1 세의 차이를 허용한 나이였다. 이제 남은 더 느슨한 매칭 변수는 행정동 안에서 성별이 다른 동일 성명, 혹은 성명이 다른 동일 성별 동일 것이다. 그러나 두 자료가 정확하다면 성명이 다른 두 사람을 성별로 매칭하는 것은 현실적으로 받아들이기 어려운 매칭이다.

이러한 현재의 상황을 감안한다면 남은 마지막 가능성은 매칭하고자 하는 원 자료(주민등록 자료, 건축물대장 자료, 시험조사 자료)의 품질을 높이는 것이다. 즉, 레코드 매칭을 수행하기 전에 매칭할 자료를 대상으로 데이터 에디팅을 실시하여 원 자료의 품질을 높인 후에 레코드 매칭을 실시하는 것이 최선의 방법으로 보인다.

4. 데이터 에디팅

가. 데이터 에디팅의 정의

데이터 에디팅(data editing)이란 데이터에서 논리적으로 일치성을 결여한 오류를 찾고 수정하는 활동을 말한다. 전통적으로 데이터 에디팅은 결측값이 포함된 원 데이터(raw data)를 정비하여 결측값이 없는 완비 데이터(complete data)를 만들기 위하여 수행하였다. 여기에 더하여 최근에는 데이터를 수집, 기록, 처리 과정에서 발견되는 데이터 오류(data error)의 양을 줄이고, 데이터 품질에 대한 정보를 제공하며, 향후 조사에 대한 개선 정보를 제공하는 역할이 강조되고 있다.

데이터 에디팅 과정은 데이터 오류를 탐색하는 과정과 탐색된 오류 데이터를 처리하는 과정으로 구분된다. 데이터 오류 탐색은 데이터가 결측값(missing value)인지, 타당하지 않은 값(invalid value)인지, 불일치하는 값(inconsistent value)인지 혹은 변칙적인 값(anomalous value)인지 점검하는 것이다. 결측값은 응답자가 응답을 하지 않아서 발생하는 값이므로 데이터 화일에서는 결측값으로 표기되어 다른 응답값과 구분되어야 한다. 타당하지 않은 값은 응답으로서 허용 가능한 범위를 벗어나는 데이터 값을 말한다. 예를 들어, 숫자를 묻는 질문에 숫자가 아닌 값이 표기되어 있으면 타당한 값이 아니다. 또한 입력된 자료가 허용된 값의 범위를 벗어나면 타당한 응답이 아니다. 불일치하는 값은 사전에 정의된 데이터 항목의 관계를 만족하지 않을 때 발생한다. 변칙적인 값은 특이한 응답이나 이상치로 나타난다.

데이터 오류 탐색에는 데이터별 전문 지식과 편집규칙(edit rule)을 활용한다. 점검규칙(check rule)이라고도 하는 편집규칙은 만일 데이터가 올바르다면 반드시 만족해야 하는 데이터 항목 값 사이의 논리적 조건이나 제약조건을 말한다. 데이터 오류 탐색 후 데이터를 최종 데이터로 채택할 수 있는 데이터와 검토가 필요한 의심스러운 데이터로 구분한 후 표기한다.

오류 데이터를 처리하는 과정에서는 결측값, 타당하지 않은 값, 혹은 불일치하는 값으로 표기된 값을 제거하고 수정한다.

나. 데이터 오류 탐색

오류를 발견하기 위해서는 예상되는 오류의 종류가 무엇인지, 내재되어 있

는 오류생성 메커니즘이 무엇인지를 알고 있어야 한다. 예를 들어, 단위측정 오류와 같은 특별한 체계적 오류 메커니즘을 염두에 두면 범위 점검이나 비율 점검 등을 통하여 체계적 오류를 찾아낼 수 있다. 질문 항목에 대한 특별한 지식이 없는 경우, 대체로 오류를 찾아내기 어렵다. 만일 예상되는 오류에 대한 사전 이해가 없을 때에는 범위 점검(range test), 비율 점검(ratio test) 방법을 사용하여 오류를 탐색한다.

범위 점검에서는 응답 값의 범위를 사전에 정한 후 응답이 범위를 벗어나면 데이터 항목을 재점검한다. 범위 점검 방법은 범주형 변수와 수치형 변수 모두에 적용할 수 있다.

가구원의 성별을 묻는 질문에서 응답은 0='무응답', 1='남성', 2='여성' 중의 하나로 받는다고 하자. 따라서 응답 집합 {0, 1, 2}에 포함되지 않는 응답은 오류로 간주한다.

'월'로 응답을 해야 하는 질문에 대하여 응답 집합을 {01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12}로 하였다고 하자. 이 경우 응답 집합을 벗어나는 응답은 오류로 간주하고 데이터를 재점검한다.

비율 점검을 위한 편집규칙에 포함되는 변수는 두 개이다. 두 변수를 x, y 라고 할 때 두 변수의 비율에 대한 상한(U)과 하한(L)을 사전에 지정하고, 비율 점검 결과 상한과 하한을 벗어나는 레코드는 오류가 있는 것으로 간주한다.

- 비율 점검 범위 : $L < \frac{y}{x} < U$

만일 두 변수에 대하여 단위측정오류 같은 응답 오류나 데이터 입력 오류가 데이터에 포함되면 y/x 비율에 즉각적인 영향을 미치게 되고, 그 값은 상한 값(U) 혹은 하한 값(L)을 벗어나기도 한다. 이러한 경우 데이터 재점검을 통하여 오류를 발견할 수 있다.

다. 오류위치포착 및 수정

편집규칙을 데이터에 적용하여 에디팅을 수행하면 편집규칙을 위반하는 불일치 레코드를 찾을 수 있다. 그 다음에는 불일치 레코드 안에서 어느 데이터가 오류 데이터인지 검토하게 된다.

불일치 레코드 안에서 어느 변수에 오류가 있는지 확인하는 활동을 '오류위치포착(error localization)'이라고 한다. 보통 오류위치포착은 쉽지 않다.

불일치 레코드에서 잘못된 변수의 위치를 일반적인 가이드라인에 의하여

포착한다. 일반 가이드라인 중에서 가장 널리 알려진 것이 펠레지-홀트(Fellegi-Holt 1976) 패러다임이다. 펠레지-홀트 패러다임은 자동 에디팅과 대체에 관한 세 가지 원칙 중 첫 번째 원칙이다.

- (i) 레코드 안에 있는 데이터는 가능한 최소 항목을 변경하여 모든 편집규칙을 만족하도록 한다.
- (ii) 데이터 파일의 빈도 구조는 가능한 한 유지한다.
- (iii) 대체 규칙(imputation rule)은 명시적인 설명 없이 대응되는 편집규칙에서 유도한다.

첫 번째 원칙은 펠레지-홀트 패러다임으로 알려져 있으며, 오류위치포착에 대한 원칙을 설명하고 있다. 두 번째와 세 번째 원칙은 랜덤 오류 탐색 후 자동 대체와 관련된 것이다. 세 번째 원칙은 오류위치포착과 대체는 서로 관계되어 있고 결합하여 적용되어야 함을 의미한다.

다음과 같은 조사데이터를 고려하자. 조사 대상은 10세이상이고, 가구주와의 관계 변수의 수준은 1=가구주, 2=배우자, 3=자녀, 4=기타이며, 학력 변수의 수준은 0=무학, 1=초등학교, 2=중고등학교, 3=대학이상이고, 혼인상태 변수의 수준은 1=미혼, 2=기혼 값을 가지는 데이터이다.

<표 4-1> 조사 데이터 (에디팅 전)

일련번호	가구번호	가구주와의 관계	연령	학력	혼인상태	편집규칙
1	1	1	33	2	1	
2	1	4	56	2	2	
3	1	2	11	3	1	위반
4	1	4	27	3	2	
5	7	4	67	1	2	
6	7	1	34	3	1	
7	7	3	16	3	2	위반
8	10	2	30	3	2	

위 데이터에서 3번째와 7번째 데이터는 편집규칙에 의해 편집규칙 위반되었다고 판단되는 레코드이다.

- 편집규칙1 : [가구주와의 관계=(가구주 혹은 배우자)] 그리고 [연령<15]
⇒ 규칙위반
- 편집규칙2 : [연령<17] 그리고 [혼인상태=기혼]
⇒ 규칙위반
- 편집규칙3 : [가구주와의 관계= 배우자] 그리고 [혼인상태=미혼]
⇒ 규칙위반

펠레지-홀트 방법에 의하여 오류를 수정한 데이터는 아래와 같다.

<표 4-2> 조사 데이터 (에디팅 후)

일련번호	가구번호	가구주와의 관계	연령	학력	혼인상태
3	1	2 ⇔ 3*	11	3	1
7	7	3	16	3	2 ⇔ 1*

라. 오류데이터 대체 처리

오류 데이터를 탐색하여 오류가 있는 것으로 여겨지는 레코드가 선정되면 다음 단계에서는 해당 레코드에서 오류가 있을 것으로 여겨지는 항목을 찾는다. 그리고 오류 항목 값을 그럴듯한 값으로 교체하여 오류 데이터를 처리한다.

오류 데이터를 처리하는 방법으로는 사후 확인 후 수정을 하는 방법과 대체(imputation) 처리를 하는 방법이 있다. 사후 확인 후 수정을 하는 방법은 응답자를 재접촉하거나 전문지식을 이용하여 오류 데이터를 수정하는 것이고, 대체 처리법은 오류 항목 값이나 결측값을 레코드나 변수 수준에서 추정치로 대체하는 것을 말한다. 대체 방법은 네 종류로 분류된다.

- 규칙 기반 대체(rule based imputation)
- 연역적 대체(deductive imputation)
- 모형 기반 대체(model based imputation)
- 기부자 기반 대체(donor based imputation)

규칙기반 대체는 전문 지식을 이용하여 오류 값을 규칙에 의하여 대체한다. 규칙 기반 대체는 일반적으로 “If ..., then...”과 같은 편집규칙을 사용하여 대체한다. 예를 들면 만일 가구원 수가 1이고 가구주와의 관계가 결측

이면, 가구주와의 관계는 본인이다. 즉,

‘가구원수=1, 가구주 관계=?’ ⇨ ‘가구주 관계=본인’

으로 대체한다.

연역적인 대체는 논리적이고 수학적인 추론에 기초하여 오류값을 대체한다. 대체된 항목이 모든 편집규칙을 만족시키는 유일한 값일 때 적용된다. 균형 편집규칙(balanced edit rule)을 이용하면 연역적 대체가 가능하다. 예를 들어 계속 조서에서 한 가정의 자녀의 수를 조사하였다. 자녀의 수 응답에서 1년차 2명, 2년차 무응답, 3년차 2명이었다. 연역적 대체법을 사용하면 논리적으로 2년차 자녀의 수는 2명으로 대체된다.

모형기반 대체에서는 명시적인 대체 모형을 이용하여 오류값을 대체한다. 대체 모형 구축에는 다음과 같은 정보를 이용한다. 모형기반 대체 방법에는 다음과 같은 방법이 많이 쓰인다.

- 회귀대체(regression imputation)
- 비대체(ratio imputation)
- 평균대체(mean imputation)
- 로지스틱 회귀 대체(logistic regression imputation)

기부자 기반 대체에서는 편집규칙을 만족하지 않은 레코드의 오류 항목에 기부자 레코드의 응답값을 대입한다. 기부자 집합은 현행조사의 응답자이거나 다른 조사의 응답자이다.

- 핫덱(hot deck) 대체: 기부자를 현행 조사 응답자 중에서 선택
- 콜드덱(cold deck) 대체: 기부자를 과거 조사 응답자나 다른 자료 응답에서 선택

기부자를 선택하는 방법은 랜덤하게 기부자를 선택하거나 최근방 기부자를 선택하는 방법 등이 있다.

- 랜덤 기부자 대체 : 기부자 집합에서 기부자를 랜덤하게 선택
- 최근방 기부자 대체 : 수령자와 거리가 가장 가까운 기부자를 선택. 만일 둘 이상의 기부자가 동일한 최소 거리를 가지면 그 중 한 기부자를 랜덤하게 선택.

제3절 불완전 매칭시 매칭률과 영역별 규모 추정

본 절에서는 파일 레코드 매칭을 수행한 후 매칭되지 않은 레코드를 어떻게 보정해야 하는지를 살펴본다. 먼저 두 조사자료(주민등록 혹은 건축물대장 자료, 시후조사 자료)를 독립적인 자료로 간주하면 센서스 사후조사 검증에서 통상적으로 많이 활용하는 이원시스템추정법(dual system estimation)을 활용하여 지원범위(coverage)를 추정할 수 있다. 그러나 이 방법은 앞 장에서도 설명이 있는 바와 같이 과소추정 문제 해결에는 적절하지만 과다 추정 문제 해결에는 적합하지 않다. 다른 방법으로 만일 사후조사 자료를 참값으로 간주하면 사후 조사자료를 기초로 주민등록 자료 혹은 건축물대장 자료의 과소집계(under-coverage) 비율, 과다집계(over-coverage) 비율을 계산하여 모집단 규모를 추정하는 것이 더 현실적이다.

먼저 다음의 예제를 보자.

1. 매칭 예제

어느 지역에 10가구가 있다고 하자. 시점1의 자료(주민등록 자료)와 시점2의 자료(시험조사 자료)를 가지고 매칭을 시도해 본다. 주소, 성명, 성별, 나이가 동일하면 정확매칭(○) 하고, 주소, 나이, 성별은 틀리더라도 성명이 동일하면 판단매칭(△)으로 간주한다.

비매칭은 가구 및 가구원 변동에서 오는 비매칭(비매칭 A)와 기록 부실에서 오는 비매칭(비매칭 B)로 구분한다.

<표 4-3> 매칭 예제

번호	시점1 (주민등록)				시점2 (시험조사)				가구 매칭	가구원	
	주소	성명	성별	나이	주소	성명	성별	나이		매칭	비매칭
1	HU1	AA1	남	63	HU1	AA1	남	63	정확	○	
2		AA2	여	60		AA2	여	60	매칭	○	
3	HU2	BB1	남	70	HZ2	BC1	남	72	비매칭 (기록 부실)		B
4		BB2	여	32		BC2	여	31			B
5		BB3	남	35		BC3	남	33			B
6	HU3	CC1	여	70	HU3	CC1	여	69	판단 매칭	△	
7		CC2	여	30							A
8	HU4	DD1	남	66	HZ4	DJ1	남	66	비매칭 (기록 부실)		B
9		DD2	여	62							A
10		DD3	남	20							A
11	HU5	EE1	여	70	HK5	EE1	여	62	판단 매칭	△	
12		EE2	남	33		EE2	남	33			△
13						EE3	여	31			A
14	HU6	FF1	남	34	HG6	FF1	남	34	정확 매칭	○	
15		FF2	여	33		FF2	여	33			○
16						FF3	남	1			A
17						FF4	여	21			A
18	HU7	GG1	남	31	HK7	GK1	남	31	판단 매칭	△	
19		GG2	여	30		GG2	여	30			△
20						GG3	여	71			A
21	HU8	JJ1	남	29					비매칭 (이사감)		A
22	HU9	KK1	남	45					비매칭 (이사감)		A
23					HU10	LL1	남	40	비매칭 (이사움)		A
24				LL2		여	35			A	
25				LL3		남	6			A	
26	HU11				HR11	MM1	남	69	판단 매칭		A
27		MM2	여	66		MM2	여	67			△
28		MM3	남	21							A

주)○: 정확매칭, △: 판단매칭, A : 비매칭 형태 A, B : 비매칭 형태 B

- 번호1, 2 : 정확매칭(○)

- 번호3, 4, 5 : 주소, 성명, 나이가 틀리므로 기록 부실 비매칭(비매칭 B)
- 번호6 : 나이(± 1) 틀리지만 판단매칭(Δ)
- 번호7 : 딸(CC2)사 시집가서 분가. 가구 변동으로 비매칭(비매칭 A)
- 번호8 : 주소, 성명이 틀리므로 비매칭(비매칭 B).
- 번호9 : 가구주 부인이 큰아들 집 방문하러 감 (비매칭 A).
- 번호10 : 아들이 다른 도시에 있는 학교로 감 (비매칭 A).
- 번호11, 12 : 주소, 나이 틀리지만 성명으로 판단매칭
- 번호13 : 며느리 들어옴(비매칭 A)
- 번호14, 15 : 정확매칭(\circ)
- 번호16 : 아들 태어남 (비매칭 A)
- 번호17 : 동생이 대학이 입학하여 HG6으로 전입(비매칭 A)
- 번호18 : 성명, 주소 다르지만 부인 성명이 동일한 것 보고 판단매칭 (Δ)
- 번호19 : 주소 다르지만 성명 같으므로 판단 매칭(Δ)
- 번호20 : 어머니가 아들 집에 방문 옴(비매칭 A)
- 번호21 : 가구 HU8이 이사감 (비매칭 A)
- 번호22 : 가구 HU9가 이사감 (비매칭 A)
- 번호23, 24, 25 : 가구 HU10이 HU8, HU9가 살던 주택으로 이사옴(비매칭 A)
- 번호26 : 가구주(MM2)의 남편이 다른 지역의 아들 집에 주민등록되어 있음. 거주는 HU11에 함. (비매칭 A)
- 번호27 : 주소, 나이 다르지만 성명 동일하므로 판단 매칭(Δ).
- 번호28 : HU11에 주민등록은 되어 있지만 군입대로 현재 상주하지는 않음. (비매칭 A)

위의 결과를 바탕으로 가구 매칭률을 계산하여 본다. 가구 매칭률은 주소만을 기준으로 하면 2가구(HU1, HU3)만이 매칭이 된다. 그러나 가구원 성명을 보고 성명이 동일하면 동일인으로 간주하면 매칭이 되는 가구는 6가구가 된다. 즉 HU1와 HU3는 정확매칭이며 HU5=HK5, HU6=HG6, HU7=HK7, HU10=HR10의 4가구는 동일한 가구로 판단되는 판단 매칭이다.

- 주민등록 가구 = 10가구
- 시험조사 가구 = 9가구
- 정확 매칭 및 판단 매칭 가구 = 6가구

- 매칭률 : 주민등록 기준 - $6/10=60\%$, 시험조사 기준 - $6/9=66.6\%$
- 비 매칭률 : 주민등록 기준 - $4/10=40\%$, 시험조사 기준 - $3/9=33.3\%$

가구원 매칭률은 다음과 같다.

- 주민등록 인구 = 20명
- 시험조사 인구 = 22명
- 정확 매칭 및 판단매칭 가구 = 6가구
- 정확 매칭 및 판단매칭 가구원 = 10명

- 매칭률 : 주민등록 기준 - $10/20=50.0\%$, 시험조사 기준 - $10/22=45.5\%$
- 비매칭률 : 주민등록 기준 - $10/20=50.0\%$, 시험조사 기준 - $10/22=45.5\%$

2. 영역별 규모 추정

위 예제를 정리하면 다음과 같다.

<표 4-4> 주민등록과 시험조사 매칭 후 가구 및 가구원 비교1

주민등록	시험조사 가구		
	기록 있음	기록 없음	합계
기록 있음	6(매칭)+ 2(비매칭 B)	2(비매칭 A)	10
기록 없음	1(비매칭 A)	-	1
합계	9	2	11
	시험조사 가구원		
기록 있음	10(매칭)+ 4(비매칭 B)	6(비매칭 A)	20
기록 없음	8(비매칭 A)	-	8
합계	22	6	28

그런데 비매칭 B의 경우는 실제로는 주민등록과 시험조사에 모두 기록이 있지만 매칭을 하지 못하는 경우이므로 현실적인 데이터 분석에서는 비매칭 A와 구분이 되지 않는다. 따라서 현실적으로 사용 가능한 비교는 다음 표가 될 것이다.

<표 4-5> 주민등록과 시험조사 매칭 후 가구 및 가구원 비교2

주민등록	시험조사 가구		
	기록 있음	기록 없음	합계
기록 있음	6(매칭)	2(비매칭 A)+ 2(비매칭 B)	10
기록 없음	1(비매칭A)+ 2(비매칭 B)		
합계	9		
	시험조사 가구원		
기록 있음	10(매칭)	6(비매칭A)+ 4(비매칭B)	20
기록 없음	8(비매칭A)+ 4(비매칭B)	-	
합계	22		

Wolter의 논문(JASA 1986, 338-346)에서는 포함범위 오차 모형(coverage error model)을 설정할 때에 100% 매칭률과 두 시점 간에 모집단 변화는 없다고 가정한다. 즉, 두 조사는 동일한 모집단에 대한 독립적인 조사로서 조사의 불완비성 때문에 조사 과정에서 누락이 발생하는 것으로 본다. 그러나 우리의 문제에서는 매칭률이 100%가 아니며, 시점 차이와 조사의 특성 때문에 누락이 발생한다.

가. 주민등록의 과소포함률(under-coverage rate)

주민등록의 과소포함률(p)는 다음과 같이 정의된다.

$$\text{주민등록 과소포함률: } p = \frac{\text{상주 인구 중 주민등록인구}}{\text{상주 인구}}$$

위 예제에서 가구와 가구원의 주민등록 과소포함률은 다음과 같다.

- 가구 과소 포함률: $\hat{p} = \frac{6}{9} = 66.7\%$
- 가구원 과소 포함률: $\hat{p} = \frac{10}{22} = 45.5\%$

나. 주민등록의 과다포함률(over-coverage rate)

$$\text{주민등록 과다포함률: } \lambda = \frac{\text{주민등록 인구 중 비상주 인구}}{\text{상주 인구}}$$

위 예제에서 가구와 가구의 주민등록 과다 포함률은 다음과 같다.

- 가구 과다 포함률: $\hat{\lambda} = \frac{4}{9} = 44.4\%$
- 가구원 과다 포함률: $\hat{\lambda} = \frac{10}{22} = 45.5\%$

다. 상주인구 추정

현재 상주인구는 다음의 공식을 이용하여 추정한다.

$$\hat{N} = \frac{\text{주민등록 인구}}{\hat{p} + \hat{\lambda}}$$

- 가구수 추정 : $\hat{N} = \frac{10}{0.667 + 0.444} = 9.0009 \approx 9$
- 인구 추정 : $\hat{N} = \frac{20}{0.455 + 0.455} = \frac{20}{0.91} = 21.98 \approx 22$

라. 추정 공식의 합리적인 해석

독립적인 두 조사(주민등록, 시험조사)를 실시하고 두 조사 데이터를 매칭하였다. x_1 은 주민등록에만 있는 인구이고, x_2 는 주민등록 및 시험조사에 있는 인구, x_3 는 시험조사에만 있는 인구이다. 만일 매칭률이 100%라고 한다면 x_2 는 주민등록과 시험조사에서 모두 발견되고 연결될 것이다. 그러나 매칭률이 100% 미만이면 이 중 일부는 연결되지 않고 주민등록과 시험조사에서 각각 기록된 것으로 나타날 것이다. 이를 Δ 라고 하자. 즉, 비매칭으로 인하여 $(x_2 - \Delta)$ 만이 매칭되어 공통으로 기록되고, $(x_1 + \Delta)$ 는 주민등록에만 기록되고, $(x_3 + \Delta)$ 는 시험조사에만 기록된다.

<표 4-6> 주민등록과 시험조사 자료의 매칭 패턴

	주민등록 에만 있음(x_1)	주민등록 및 시험조사에 있음(x_2)		시험조사 에만 있음(x_3)	
	주민등록에만 기록됨 (비매칭)	공통 기록됨 (매칭)	시험조사에만 기록됨 (비매칭)		
주민등록	x_1	Δ	$x_2 - \Delta$		
시험조사			$x_2 - \Delta$	Δ	x_3

과소포함률 p 는 상주 인구 중 주민등록 기록이 있는 인구이므로 시험조사 기록 중 주민등록 기록의 비율로 추정된다. 즉,

$$\hat{p} = \frac{x_2 - \Delta}{(x_2 - \Delta) + (x_3 + \Delta)} = \frac{x_2 - \Delta}{x_2 + x_3}$$

또한 과다추정인구는 주민등록에는 기록이 있고 상주인구가 아닌 인구의 상주인구에 대한 비율이므로 다음으로 추정된다.

$$\hat{\lambda} = \frac{x_1 + \Delta}{(x_2 - \Delta) + (x_3 + \Delta)} = \frac{x_1 + \Delta}{x_2 + x_3}$$

그런데 두 비율이 합은 결국 주민등록 인구의 상주인구 비가 된다.

$$\hat{p} + \hat{\lambda} = \frac{x_1 + x_2}{x_2 + x_3} \approx \frac{\text{주민등록 인구}(M)}{\text{상주 인구}(N)}$$

따라서 상주인구는 주민등록 인구를 두 비율의 합으로 나눈 값으로 추정 가능하다.

$$\hat{N} = \frac{M}{\hat{p} + \hat{\lambda}}$$

재미있는 사실은 매칭률이 떨어지면 과소포함률 \hat{p} 는 감소하고, 과다추정률 $\hat{\lambda}$ 는 증가한다. 그러나 두 비율이 서로 상쇄되어 상주인구를 추정할 때에는 비매칭률이 큰 영향을 주지 않는다는 것이다.

제4절 불일치 자료 보완에 따른 보정 방법

매칭된 데이터를 기반으로 지역별, 영역별 인구를 추정하면 당연히 추정 인구와 주민등록의 인구가 일치하지 않는 현상이 발생한다. 센서스의 장점 중의 하나가 소지역에서 세부 항목의 집계 가능하다라는 점을 상기하면, 행정자료를 이용하는 경우에도 이 같은 장점을 유지할 필요가 생긴다.

예를 들어 어느 영역의 주민등록 인구와 상주인구 추정값이 다음과 같다고 하자. 30대 남성의 경우 주민등록 인구는 110명인데 비하여 추정 상주인구는 115명이다. 따라서 주민등록 인구가 5명 작다. 세부 항목에 대한 집계를 위해서는 주민등록 인구 110명을 상주인구 115명으로 변환할 필요가 있다.

<표 4-7> 주민등록 인구와 추정 인구 및 가중치

연령	주민등록		추정값		가중치	
	남	여	남	여	남	여
20대	115	110	110	105	0.9565	0.9545
30대	110	105	115	110	1.0454	1.0476
40대	90	95	95	100	1.0556	1.0526
합계	315	310	320	315	1.0158	1.0161

이러한 문제를 풀기 위한 통상적인 보정 방법은 (i) 가중치 조정법과 (ii) 대체법이 있다. 첫 번째로 가중치 조정법은 100명의 30대 남성이 115명의 역할을 해야 하고 그 비가 1.0454이므로 30대 남성 한명이 1.0454배 만큼의 역할을 하는 것이다. 즉, 세부항목 집계를 할 때에 30대 남성의 경우는 가중치 1.0454를 곱한 후 집계를 한다.

두 번째로 대체법은 부족한 5명의 30대 남성을 새로이 생성하는 방법이다. 앞 절에서 설명한 바와 같이 이 경우에 적용할 수 있는 대체법으로는 모형 기반 대체와 기부자 기반 대체가 적절하다. 대체로 연속형 변수의 경우는 모형기반 대체가 효과적인 경우가 많은 반면, 이산형 변수인 경우에는 기부자 대체 방법이 적절하다. 영국의 원넘버 센서스에서는 기부자 대체 방법을 제안하고 있다(Steel et al., 2002). 기부자 대체 방법으로는 핫덱 대체(hot deck imputation), 콜드덱 대체(cold deck imputation), 랜덤 기부자 대체(random donor imputation), 최근방 기부자 대체(nearest neighbor donor

imputation) 등에서 세부항목의 특성에 맞는 방법을 선택하여 사용한다.

제5절 정책적 건의사항

등록센서스, 바로 센서스 정보를 행정자료로 대신하는 작업의 추진에서 (i) 데이터 매칭 (ii) 영역별 규모 추정 (ii) 불일치 데이터 보정 과정이 필수적으로 들어간다.

첫 번째 과정에서는 매칭률을 높이는 것이 무엇보다도 중요하다. 앞 장의 결과에 의하면 인구는 74% - 83%, 주택은 50%-90% 사이인 것으로 나타났다. 현실적인 조건에서 위 매칭률이 최선의 매칭률이라고 한다면, 매칭률을 더 높이는 방법으로 데이터 에디팅을 통한 주민등록 자료와 시험조사 자료 등 원 자료의 데이터 품질을 높이는 방안을 모색할 필요가 있다.

두 번째 영역별 규모를 추정하는 과정에서는 비매칭률을 고려하는 모수 추정법을 개발하여야 한다. 본 장에서 소개한 비매칭률을 고려한 영역별 추정법은 비매칭률로 인하여 과소포함률은 과소 추정하고 과다 포함률은 과다 추정하는 성향이 있는 반면, 상주인구 추정에는 비매칭률이 큰 영향을 주지 않았다.

세 번째로 영역 내에서 세부 항목 집계를 위한 주민등록 인구와 추정값의 불일치는 보정하는 방안으로 가중치 조정법과 대체법이 유용하다. 대체법은 집계하고자 하는 항목에 따라 선택 방법이 달라지는데, 기부자 대체 방법 중에서 랜덤 핫덱 방법이나 최근방 대체 방법이 유용할 것으로 보인다.

제5장 등록센서스 대체이용 불가능 조사항목의 표본조사 도입방안

제1절 머리말

우리나라 통계청은 물론 모든 국가에서 통계작성기관의 주요 책무는 관련성/완전성, 적시성/시의성, 정확성, 비교성/일관성, 접근성/명확성, 비용효율성, 응답자 부담경감 등의 기준에 의거하여 최고품질의 국가통계(national statistics)를 작성하는 것이다.

등록센서스의 토대자료가 될 수 있는 행정자료, 특히 행정등록대장은 통계작성을 일차적 목표로 하는 것이 아니라, 행정업무의 편의를 위하여 작성되는 것이 대부분이다. 따라서 행정자료의 개념이나 정의는 전통적 센서스에서 조사표와 현지실사(field operation)의 형태로 진행되는 통계조사의 개념이나 정의와는 달리, 국가통계의 기본요건을 당장은 충족시키지 않은 것이지만, 일상적 행정목적에는 도움이 되는 것이다.

통계작성기관은 언제나 센서스 통계작성의 과정에서 현실적 여건과 이상적 최고품질의 “통계기준” 간에 타협을 하지 않으면 안 된다. 어떤 면에서, 행정자료의 이용은 국가통계의 작성환경에 대변화를 초래하였다. 가령, 통계작성기관이 행정업무를 통하여 획득한 자료를 이용할 수 있는 경우, 그 자료를 센서스 통계작성에 이용할 수 있는가에 대하여 판단하기 위하여, 통계작성기관의 품질기준과 비교평가하지 않으면 안 될 것이다. 목표모집단의 행정적 정의(고용통계의 자영업자 조사범위), 변수정의의 방식(예: 주민등록대장의 주소지 설정방식과 인구주택총조사의 상주지 설정방식), 통계작성의 기준시점(예: 행정자료의 업데이트시점과 센서스 기준시점의 차이)이 차이가 날 수도 있으나 통계작성기관이 행정자료를 거의 자유자재로 이용할 수 있게 된다면, 행정자료의 포괄범위를 벗어난 정보를 보완하기 위하여 독자적으로 표본조사를 이용하여, 전통적 센서스에서 하는 것처럼, 필요한 보정조치를 실시하여 센서스 통계를 작성하는데 사용할 수도 있을 것이다.

등록센서스는 행정자료, 특히 행정목적의 공적장부를 통계목적의 공적장부로 변환하여 실시되는 “행정자료 센서스”를 의미한다. 물론, 이러한 센서스도 원칙적으로, 전통적 센서스의 본질적 특성인 개별성(individuality), 보편성(universality), 동시성(simultaneity), 주기성(periodicity)의 원칙을 충족시켜야

<표 5-1> 북유럽 국가의 행정자료 수립연도와 센서스 이용연도

자료유형	덴마크		핀란드		노르웨이		스웨덴	
	도 입 연 도	최 초 이 용 연 도	도 입 연도	최 초 이 용 연도	도 립 연도	최 초 이 용 연도	도 립 연도	최 초 이 용 연도
인구대장	1968	1981	1969	1970	1964	1970	1967	1975
사업체대장	1975	1981	1975	1980	1965	1980	1963	1975
주택대장	1977	1981	1980	1985	2001	2011	2008?	2011?
주택상태	1977	1981	1980	1985	2001	2011	2008?	2011?
교육	1971	1981	1970	1975	1970	1980	1985	1990
고용	1979	1981	1987	1990	1978	2001	1985	1985
가족	1968	1981	1978	1980	1964	1980	1960	1975
가구1	1968	1981	1970	1975	2001	2011	2011?	2011?
소득	1970	1981	1969	1970	1967	1980	1968	1975
완전 등록센서스		1981		1990		2011		2011?

1) 가구는 거주가구의 단위, 곧 단위거처에 거주하는 모든 사람들을 의미하며 이것은 전통적 센서스의 가사단위의 가구개념과는 다르다.

자료: UNECE (2007)과 Wallgren, A. and Wallgren, B. (2007)의 재정리

한다.

센서스는 그것이 행정자료를 이용하든, 모집단을 직접조사는 전통적 센서스이든 모두 그것이 “센서스”라는 이름으로 불리기 위해서는 이른바 “센서스 항목”에 대한 통계를 작성해야 한다. 북유럽의 국가들은 오랫동안 전통적 센서스에서 100% 완전한 등록센서스로 이행하기 위하여 노력하였으며, 과도기적으로 센서스 항목에 대한 정보를 행정자료로 대체하기가 불가능한 때는, 이들 항목에 대해서 표본조사를 도입하여 그 정보를 이용하여 센서스 통계를 작성하였다. <표 5-1>은 북유럽 국가들이 인구주택총조사를 대상으로 하여 특정의 조사항목에 대하여, 행정등록대장을 수립하고 이것을 적극적으로 센서스 통계작성에 이용하기 시작한 시점을 보여주고 있다.

등록센서스의 추진과정에서, 북유럽 국가들은 1964년부터 1969년에 이르기까지, 중앙정부가 관리하는 인구등록대장을 수립하고, 우리나라의 주민등록번호와 동일한 개인식별번호(PIN, Personal Identification Number)를 도입하였다. 행정자료가 통계작성의 원천으로 이용되기 시작한 것은 1970년대 초반부터였다. 행정등록대장이 인구통계와 소득통계를 필두로 하여, 다양한 주제영역의 통계작성에 이용되기 시작하였다. 추가적으로 북유럽의 통계청은 행정등록자료를 신규로 개발하고, 그것을 토대로 하는 통계작성업무를 개시하

였다.

등록센서스 추진의 선도적 위치에 있었던 북유럽에서도 국가별로, 주제영역별로, 행정대장을 수립하기 시작한 연도와 해당 자료를 센서스 통계목적에 최초로 사용한 연도 간의 간극은 상당한 차이가 있다. 하나의 예로 고용통계의 개발을 검토할 수 있겠는데, 덴마크, 핀란드, 스웨덴은 고용대장을 수립하고 몇 년 만에 그 자료를 센서스 통계를 작성하는데 이용하였지만, 노르웨이는 1978년에 고용대장 수립 작업을 추진하였지만 센서스 통계를 작성하는데 이용하기 시작한 것은 2001년부터였다.

북유럽의 국가의 등록센서스 추진과정에서 한 가지 동일한 점은 행정대장이 단계적으로 수립되고, 그것을 센서스 통계작성에 이용하였다는 것이다. 먼저, 다양한 영역에서 주제별 통계가 검정되고 그 결과표가 작성되기 시작했다는 것이다. 해당 국가의 통계청은 주제별 통계의 품질이 만족스러운 것이 되면, 등록대장의 변수를 센서스 통계작성에 곧바로 도입하였다. 곧 센서스에 관련되는 주제영역, 곧 인구, 가족, 가구, 교육, 고용, 소득, 주택 등 모든 분야에서 통계가 작성되면서, 100% 완전한 등록센서스로 이행하였다. 덴마크와 핀란드는 1981년과 1990년에 최초로 100% 완전한 등록센서스로 이행하였으며, 노르웨이는 2011년에 100% 완전한 등록센서스로 이행할 예정이며, 스웨덴은 2011년에 행정자료에서 가구통계를 작성할 수 있게 된다면 100% 완전한 등록센서스로 이행하게 되는 셈이다.

전통적 센서스에서 등록센서스로 이행하는 과정은 북유럽의 국가만이 아니라, 전 세계의 모든 국가를 대상으로 할 때, 그 스펙트럼은 더욱 더 다양하다. 먼저, 조사표와 현지실사라는 전통적 센서스를 버리고, 등록센서스를 추진하고자 하여도 등록센서스의 추진에 필요한 핵심적 행정자료인 인구대장이 정비되지 않아서, 현 단계에서 등록센서스로 이행을 보류하고 있는 선진국들이 있는데, 그 대표적인 경우가 영국과 미국이다.

또, 전통적 센서스에서 등록센서스로의 이행을 당장에 염두에 두지 않는 남유럽의 국가 중에는 전통적 센서스의 정보를 행정자료, 특히 인구대장의 정보를 수정하기 위하여 사용하는 경우가 있다. 스페인, 이탈리아가 그러한 경우에 속하고, 1980년대에 독일도 이 경우에 속했지만 연방재판소의 판결로 전통적 센서스 실시를 유예하였으며, 오랫동안 새로운 센서스 모형을 테스트한 결과, 인구대장을 주축대장으로 하고 필요한 경우 조사항목별로 일부는 전수조사 일부는 표본조사의 형태로 보완하는 등록센서스 모형을 개발하고, 이 모델을 테스트하고, 2011년에 센서스 본조사를 실시할 예정이다.

아시아에서는 싱가포르와 이스라엘이 2000년과 2008년에 각각 전통적 센서

스를 포기하고, 등록센서스로 발 빠른 이행을 완료하였다. 싱가포르는 1990년 센서스에서 1980년에 구축된 인구대장(People Hub)을 이용하여 현지실사를 진행하고, 개인식별번호를 토대로 센서스 결과를 행정자료와 매칭·통합하는 작업을 진행하고, 1996년에는 가구등록대장(Household Registration Database)과 전국주택대장(National Database on Dwellings)의 수립을 완료하여, 2000년에는 아시아에서 최초로 등록센서스와 20% 표본조사를 동시에 실시하였다. 한편, 이스라엘도 2008년에 인구대장(Population Register)을 토대로 보완행정파일(Improved Administrative File)을 구축하고 이 파일의 정확성을 점검하고 센서스 통계의 사회경제적 파일을 작성하기 위하여 20% 표본조사를 실시하였다. 등록센서스에서 싱가포르와 이스라엘은 모두 표본조사를 실시하였다는 점은 동일하지만, 이 표본조사가 싱가포르에서는 인구와 주택의 사회경제적 특성을 획득하는데 사용되었는데 반하여, 이스라엘은 인구대장과 표본조사의 집계결과를 매칭하여 인구대장 바로 보완행정파일의 문제점을 개선하는 도구로도 사용하였다는 점이다.

이 글에서는 먼저 센서스의 다양한 유형을 검토한다. 다음은 우리나라 현행 인구주택총조사를 개관하고, 그 조사방법론을 기술하고, 이것을 바탕으로 센서스 조사항목이 현재의 행정자료 특히 마이크로데이터를 구비한 행정자료에 의하여 어느 정도 대체가능성이 있는가를 평가하고자 한다. 다음은 외국의 센서스 표본조사 사례연구를 진행하고, 이것을 바탕으로 우리나라 2015년 등록센서스에서 표본조사가 담당해야 할 역할을 기술한다. 여기서는 센서스 표본조사의 이중적 역할, 곧 집계범위를 추정하기 위한 근거자료로 이용된다는 점과 현행 전통적 센서스 방식의 표본조사 항목 또는 전수조사 항목의 일부에 대하여 자료를 획득을 위한 조사도구로 이용된다는 점을 강조한다. 마지막으로, 센서스 표본조사의 이중적 목적을 구현하기 위하여, 우리나라의 2015년 등록센서스에서 대체 불가능인 조사항목의 현장실사는 물론 주민등록 데이터베이스와 건물/주택대장 데이터베이스의 문제점을 보완하기 위하여 현지실사를 어떻게 활용하여야 할 것인가를 설명한다.

제2절 센서스의 다양한 유형

인구센서스를 실시하는 주된 이유의 하나는 상세한 정보, 특히 소지역에 대한 상세한 정보가 필요하기 때문이다. 그러나 기존 센서스의 단점은 5년 또는 10년마다 실시되기 때문에 정보의 노후화와 비용이 많이 들고 센서스 주기를 단축하기가 힘들다는 점이다.. 또한 개인비밀보호의식의 증가 및 단

독가구의 급증 등으로 인해 조사환경이 악화되어 무응답 문제가 증가하고 있다. 이러한 문제점들을 해결하기 위해서 전통적인 센서스를 포기하거나 변형하여 실시하는 나라들이 점점 늘어나고 있다. 무응답 문제를 해결하기 위해서 행정자료를 이용하고, 센서스의 막대한 비용을 여러 해에 걸쳐 분산하기 해서 순환센서스를 실시하는 등 여러 유형의 센서스가 등장하고 있다.

UN은 이러한 여러 가지 유형을 정리하여 세계 각국의 센서스를 ① 전통적 센서스, ② 행정자료에 기반을 둔 접근법, ③ 전통적 센서스와 행정등록 센서스의 혼합형, ④ 등록센서스와 표본조사의 혼합형, ⑤ 순환 센서스, ⑥ 전통적인 센서스와 특성치의 연간 업데이트 등으로 분류하였다(UN, 2007). 2015년에 도입되어야 할 우리나라의 센서스 표본조사의 방식을 설명하기 전에 UN의 분류에 따른 해외 각국의 개괄적인 사례설명을 통해 센서스 실시 방식에 대한 해외의 전반적인 흐름을 조망하고자 한다.

가. 전통적 센서스

2000 라운드 센서스에서는 190개국 이상이 인구센서스를 실시하였고 대부분의 나라들이 전통적인 방식을 따랐다. 인구센서스는 특정 시점에서 일련의 주제에 관해 인구 및 가구에 관한 정보를 수집하는 복잡한 작업이다. 수집된 자료는 편집, 평가, 분석 및 공표과정을 통해 국가 및 지역에 관련된 인구, 경제, 사회통계를 작성하는데 이용된다. 국민들이 센서스 질문지를 직접 작성할 경우도 있고 조사원을 고용하여 정보를 얻을 수도 있다. 조사원을 고용하여 센서스를 실시할 경우, 각 조사원은 비교적 짧은 기간인 특정의 시점에서 배정된 조사구내의 모든 가구 및 개인들을 조사함으로써 보편성과 동시성이라는 센서스의 요건을 충족하게 된다. 전수조사표와 표본조사표가 모두 사용된다. 전수조사표는 모집단 전체를 대상으로 하고 표본조사표는 표본으로 선정된 가구 및 인구를 대상으로 한다. 표본조사표에는 특정한 주제에 관한 상세한 질문들이 포함된다. 표본조사 추정치는 전체 포괄범위에 기초를 둔 것은 아니지만 센서스의 결과물로 간주된다.

자료수집 방법에는 우편조사, 인편으로 조사표 전달, 전화조사, 인터넷조사, 면접조사 또는 이러한 조사방법의 혼합형 등 여러 가지가 있기 때문에 전통적인 센서스를 실시하는 나라들의 조사방법은 다양하다.

전통적인 센서스는 특정 시점에서의 인구에 대한 단면을 제공하고 소지역, 소집단에 대한 자료를 제공한다는 것이 특별한 장점이다. 전통적 센서스는 연방제를 채택하여 예산수립 및 배분에 필요한 모든 지역단위에 대해 다양

한 사회경제 특성치를 한 시점에서 동시에 생산할 것이 요구되고 있는 나라에 적합하다. 선거구 경계선 확정작업 또한 동시성을 요구하고, 이 때문에 전통적 센서스가 더 나을 것이다. 그러나 한편으로 전통적인 센서스는 센서스를 담당하는 부서가 행하는 업무 중 가장 많은 노력이 요구되며, 복잡하고, 비용이 많이 소요되는 자료수집활동이 된다. 비용문제 외에 이 복잡한 작업에는 국민들이 참여해야한다는 인식과 동의가 요구된다. 복잡성과 비용문제 때문에 이러한 센서스는 보통 매 5년 또는 10년마다 한 번씩 실시되고 따라서 센서스 자료는 여러 해가 지난 적시성/시의성이 떨어진다.

나. 행정자료로 전통적 방식을 보완하는 센서스

전통적 센서스의 각종 주제에 대한 행정자료가 균형있게 개발되어 있는 않는 경우, 모든 센서스 결과를 생산할 만큼 신뢰할 수 있는 행정자료가 없는 나라들은 행정자료에서 이용할 수 있는 정보를 센서스의 기초자료로서 사용할 수 있다. 2000 라운드 센서스에서 일부 국가는 인구대장을 토대로 거쳐주소, 거주자 이름 및 기타 정보를 사전 인쇄한 조사표를 이용하고, 해당 정보는 면접 시에 확인받거나 필요시에는 수정하고, 필요하지만 사전 인쇄되지 않은 정보는 면접 시에 수집되었다.

이러한 방법은 현지실사 조사원의 면접시간을 줄이는데 기여한다. 또 이에 따른 시간비용의 감소가 있으며, 응답자의 부담은 상당 수준 경감되고 자료품질이 개선되었다. 또한 포괄범위에 대한 평가가 인구대장과 현지실사 결과의 비교를 통해 이루어졌다.

더욱이 행정자료의 품질진단이 유사한 방식으로 이루어질 수 있다. 또한 일부 항목의 응답내용은 현지실사를 토대로 개선될 수 있다. 이 방법은 품질이 좋은 행정자료만을 이용하여 센서스 중간연도에 대한 자료 생산의 빈도를 높이는 것이 가능할 것이다. 이러한 형태의 주기적인 센서스 자료 수집은 행정자료에서 직접 획득이 가능한 인구의 수정을 가능하게 하고 이주 등과 관련된 등록되지 않는 사건이 일어남으로써 필요한 재고조사 활동으로 간주될 수 있다.

마지막으로, 이 방법은 전체 포괄범위에 기초를 두었기 때문에 전통적 센서스의 장점을 그대로 간직한다. 특정 시점에서 전체 인구에 대한 단면을 제공하고 소규모 지역, 소규모 집단에 대한 자료의 생산이 가능하다. 이러한 형태의 센서스를 시행한 경험이 있는 나라들은 전통적인 센서스의 결점이 보완되는 장점이 있다고 주장한다.

다. 행정자료와 표본조사의 통합센서스

이스라엘, 싱가포르, 네덜란드 등의 일부 국가에서는 등록센서스를 표본조사와 병행하여 실시한다. 독일도 이러한 통합센서스를 2011년에는 실시하게 될 것이고, 현재까지 알려진 바로는 대만도와 이와 유사한 통합센서스의 형태로 가고 있다. 이것은 등록센서스를 통해서 필요한 모든 정보를 얻을 수 없는 경우에 해당한다. 이 방식은 행정자료의 기록과 표본조사를 통해 얻은 자료를 결합하여 마이크로 수준 즉 개별 단위에서 매칭을 실시한다. 다음 단계에서는 통합된 데이터베이스를 이용해 통계표를 생성한다. 이 방식은 전통적 센서스의 전수/표본 형태와 유사한 점이 많다. 전수항목에 대해서는 행정자료에서 각 개인 및 가구별로 자료를 수집하고 표본항목에 대해서는 개인이나 가구의 표본조사를 이용하여 결과를 추정하는 방식이다. 따라서 『인구주택센서스의 원칙 및 권고』(Principle and Recommendation for Population and Housing Censuses)에 기술되어 있는 표본의 사용에 대한 권고가 이 모형에 적용되지만 가중치 절차(weighting procedure)에 대한 정교한 작업이 필요하다.

이 방식의 주요 장단점은 나중에 보는 100% 완전한 등록센서스와 유사하다. 가상센서스에 대한 반대는 거의 없고 무응답의 문제만이 표본조사의 자료가 사용되는 데 있어서 발생한다. 만약 무응답 문제가 표본조사에서 정정될 수 있다면 센서스에서 표본조사의 선택성 문제가 정정되는 것이 가능할 것이다. 단점은 마이크로 데이터에서 표를 생성하는데 많은 노력이 필요하다는 점이다. 왜냐하면 가중치 부여 문제가 발생하기 때문이 노력이 국민 전체를 조사객체로 하는 현지실사가 행정자료로 대체될 때, 센서스 결과에 대한 관심을 얻는 것이 줄어들 수 있다.

라. 행정자료를 토대로 하는 완전 등록센서스

행정자료를 이용하여 전통적 센서스와 비슷한 결과를 생산한다는 아이디어는 2000 라운드 센서스에 본격적으로 등장했다. 이 방식은 1970년 대 이후로 논쟁이 되었고 시험되어 왔으며 몇몇 나라에서는 이 방식을 성공적으로 수행하여 1990 라운드 센서스 자료를 생산했다. 이 개념의 배경 철학은 현존하는 행정자료를 이용하는 것이다. 즉 개인, 가구, 건물/주택에 관한 대장을 포함한 서로 다른 행정자료를 이용하는 것이다. 다음 단계에서는 이러한 것들을 개인 수준에서 기업체, 세금, 교육, 고용 및 기타 관련 행정자료와 연결시

키는 것이다. 이론적으로는 개인의 이름을 이용하여 기록들을 연결하는 것이 가능하지만, 각 개인, 가구, 거처에 대한 식별번호의 존재가 대단히 중요하다. 왜냐하면 서로 다른 행정자료의 기록들을 효율적이고 신뢰할만한 매칭과 통합을 가능하게 하기 때문이다.

이 접근법의 필수 조건은 전국 수준의 집계범위가 100%에 가까운 고품질의 인구대장을 보유하고 있어야 하며 계속적으로 보완할 수 있는 시스템을 보유해야 된다는 점이다. 지역 등록부의 경우는 지속적인 보완 및 다른 지역의 등록시스템과의 협조 관계가 잘 되어 있어야 한다. 품질평가 또한 필요하다. 만약 이러한 조건들이 충족되지 않는다면 그 나라는 인구센서스를 기준 인구통계의 자료원으로 사용해야 한다.

행정자료 센서스의 주요 장점은 센서스 비용절감 및 자료 빈도의 확대이다. 그러나 행정자료를 통계목적으로 변환하여 센서스에 이용하는데 드는 비용은 전통적 방식의 센서스보다 많이 들 수 있다. 행정등록의 필요성을 사실상 효과적인 행정 이지 통계작성이 아니기 때문에, 행정등록을 통계목적의 장부로 변환하여야 한다. 행정자료를 사용하는 데에는 고려해야 할 단점이 있다. 한 가지 단점은 등록에 기초를 둔 설명은 이용할 수 있는 명부에 기초를 두고 형성할 수 있는 정보에 전적으로 의존해야 한다는 사실이다. 더욱이, 다수의 국가에서 행정자료를 통계목적으로 사용하는 것을 법적으로 제한하고 있다. 이 점이 센서스 통계의 작성에 제약을 가져다주고 국제비교의 가능성을 저해할 수 있을 것이다.

마. 순환센서스

프랑스에서 처음으로 실시된 순환센서스(rotating census)는 전통적 센서스 방식의 대안으로서 연속적인 표본조사를 누적하는 방법이다. 이 방식은 전국을 특정 시점 또는 단기간이 아니라, 5년 또는 10년간에 걸쳐서 조사하는 방식이다. 순환센서스의 두 핵심요소는 조사기간과 표본 추출률이다. 조사기간은 자료의 보완 빈도와 관계가 있고 표본 추출률은 예산 및 공표에 필요한 지역단위와 관련이 있다. 가령, 한 번의 연간 표본조사를 통해 국가단위의 자료를 생산하고, 5년을 누적하여 소지역 통계를 생산할 목적으로 표본들을 생성하는 것이 가능하다. 연간 표본조사는 1년간에 걸쳐서 시행될 수도 있고 특정한 달 또는 좀 더 짧은 기간에 시행될 수도 있다.

센서스를 연속적인 표본조사로 이행하기 위해서는 매우 복잡한 표집 및 모형 기술, 소지역 수준의 표본틀(매년 업데이트되는 Master Address File이 필수적임), 중앙정부, 지방자치단체, 기업체 등 주요 이해 당사자의 협력자세

가 필수적이다. 이 방식의 가장 큰 장점은 자료 업데이트의 주기가 짧다는 점이다. 전통적인 센서스에서는 자료의 갱신이 5년 또는 10년 만에 이루어지지만 순환센서스에서는 매년 이루어진다. 또 다른 장점은 전통적인 센서스에서의 일시적인 큰 비용 및 인력의 부담을 완화한다는 점이다. 게다가 매년 작업의 개선을 이룰 수 있고 새로운 기술을 시험할 수 있다. 가장 큰 단점은 센서스의 동시성 원칙을 위반하여, 전체 모집단에 대한 동일 시점의 단면도를 제공하지 않고, 조사 시점의 차이 때문에 지역 간의 비교를 복잡하게 만든다는 점이다.

바. 전통적 센서스와 특성치의 연차별 업데이트

이 방식은 미국에서 2010년에서 실시될 예정으로 있으며, 이것도 전통적 센서스의 변종이라고 볼 수 있으며, 표본조사에 초점을 두고 센서스 실시연도에는 인구의 기본특성 자료만 수집하는 방식이다. 10년간에 걸쳐 매년 대규모의 가구 표본조사를 통해 상세한 인구, 사회, 경제 및 주택 자료를 수집하여 센서스 표본조사표를 대신하는 방식이다. 매년 모든 주제에 대한 자료를 수집할 필요는 없다. 왜냐하면 나라마다 필요한 자료가 다를 수 있기 때문이다. 이 표본조사에서는 매년 일정 분량의 주소지를 표본으로 선정하여 센서스 주기의 일정 기간에 걸쳐 표본조사표의 표본 추출률과 비슷하게 만든다. 소규모 행정단위에 대한 추정치의 신뢰도를 높이기 위해서 더 많은 비율의 주소지가 표본으로 추출된다. 표본은 누적되어 전통적 센서스 표본조사의 표본과 유사한 세부지역단위에서 자료를 생성하게 된다. 표본설계에 따라 표본자료들은 가중치가 부여되고, 무응답 효과를 조정하고 과소집계 또는 과다집계를 보정하게 된다. 이 마지막 가중치 조정(weighting adjustment)은 특성치에 대한 추정치가 주기적인 센서스의 표준치와 비교될 수 있도록 해준다. 마지막 가중치 작업이 완료되면 추정인구, 비율, 평균, 미디언 등의 기초통계가 생성된다.

이 방식을 추진하게 하는 주요 동기는 두 가지다. 10년에 한번 센서스를 실시했을 때 얻을 수 있는 것보다 더 시의성과 적합성이 큰 자료를 제공한다는 점과 센서스와 관련된 작업의 위험을 줄인다는 점이다. 그러나 이러한 계획은 비용이 더 많이 들고 시작하는데 있어서 기술적인 어려움이 있고, 포괄적인 계획, 개발 및 시험에 있어 장기 계획을 필요로 한다. 특히 주기적으로 인구를 전수조사에 의하여 집계할 것을 요구하는 헌법이 있는 나라에서는 이러한 방식이 하나의 선택방안이 될 수 있다.

제3절 등록센서스와 표본조사의 역할

앞 절에서는 전통적 센서스와 대안적 센서스 디자인에 대하여, 등록센서스를 포함하여 센서스의 폭넓은 스펙트럼을 검토하였다. 여기서는 독자적인 자료획득시스템을 갖춘 통계조사(statistical survey) 곧 전통적 센서스와 표본조사와 비교하여 등록기반조사의 장단점을 비교하고, 등록센서스에서 행정자료, 특히 행정자료와 표본조사의 관계에 대하여 언급하고, 다음으로, 표본조사가 우리나라의 2015년 등록센서스에서 구체적으로 어떤 역할을 하게 될 것인가를 (1) 행정자료(주민등록, 건축물/주택 또는 단위거처대장) 포괄범위의 추정을 위한 근거자료 (2) 전통적 센서스 표본항목(또는 전수항목의 일부) 자료를 획득하기 위한 조사시스템으로서의 역할에 대하여 언급하도록 한다.

1. 등록기반조사의 장단점

스웨덴 통계청은 등록기반조사의 장단점을 통계조사, 곧 독자적인 자료획득 시스템이 있는 전통적 센서스와 표본조사와 대비하여 설명하고 있다. 물론, 스웨덴 통계청은 등록기반조사를 토대로 작성되는 통계에 대하여 편견과 정당한 비판이 있다는 것을 인정한다. 그 중의 하나가 행정자료를 바탕으로 작성되는 센서스 통계는 진정한 조사통계보다 저렴하지만 품질이 떨어진다는 것이다. 정당한 비판은 통계품질 차원에서 관련성 문제, 비교가능성 문제 등이 포함될 수 있을 것이며, 따라서 통계품질을 적정수준에서 통제하기가 힘들다는 것이다. 그러나 같은 비판이 표본틀(sampling frame) 문제, 측정 문제, 무응답 문제 등이 센서스나 표본조사에도 가해질 수 있다는 것이며, 표본조사의 경우에는 표본오차와 관계없이 이러한 문제가 존재한다는 것이다.

스웨덴 통계청은 등록기반조사의 문제점에 대한 답으로서, 사회전반의 통계는 통계청이 독자적으로 획득하는 자료를 기반으로 하는 통계, 곧 조사통계와 행정목적으로 수집된 등록기반통계로 구성되어야 한다는 것이다. 따라서, 등록통계가 조사통계보다 품질이 우수하다든가 또는 조사통계가 등록통계보다 품질이 우수하다든가 하는 문제가 아니라, 상황에 따라서 등록기반통계와 통계조사에 의한 통계의 효율성이 달라질 수 있다는 것이다.

<표 5-2> 통계조사와 등록기반조사의 장점과 단점 비교

	장점	단점
통계조사 (독자적 자료 획득 시스템이 있음, 센서스와 표본조사)	<ul style="list-style-type: none"> • 조사항목을 선택이 쉽다 • 업데이트가 쉽다 	<ul style="list-style-type: none"> • 응답자에 따라서 다음과 같은 문제가 발생한다. <ul style="list-style-type: none"> -조사항목의 이해부족 -조사항목을 잊어버린다. -무응답 문제(non-response) -부주의 응답 • 응답자부담이 크다 • 고비용 문제 • (표본조사의 경우) 소집단 추정의 품질에 문제가 발생
등록기반조사	<ul style="list-style-type: none"> • 통계작성을 위한 응답자의 추가적 부담이 없다 • 비용이 적게 든다 • (영질의 등록자료) 모집단의 집계범위가 사실상 100%이다 • (양질의 등록자료) 시계열 확보가 사실 100%이다 • 응답자는 행정목적의 주요 조사항목에 대해서는 세심하게 대답한다. • 소지역통계, 지자체통계, 그리고 장기적 시계열 연구의 가능성 증대 	<ul style="list-style-type: none"> • 조사항목을 추가할 수 없다. • 행정시스템의 모집단, 조사객체, 변수 정의에 의존한다. • 행정자료 신고에 시의성에 문제가 생길 수 있다; 기준 시점과 통계목적으로 자료가 가용한 시점 간의 시차가 길 수 있다. • 행정시스템의 변화는 비교자체를 어렵게 할 수 있다. • 행정업무의 우선순위에 따라 조사항목의 품질이 달라진다. 우선순위에서 떨어지는 변수는 품질이 낮다

자료: Wallgren, A. and Wallgren, B. (2007)

2. 행정자료와 표본조사를 결합하는 통합센서스

100% 행정자료만을 이용하여 센서스 통계를 작성하는 것은 북유럽의 몇 나라, 덴마크, 핀란드, 노르웨이는 물론 현재 우리나라, 독일, 싱가포르, 이스라엘과 같이, 등록센서스를 최근 실시하였거나 추진하는 모든 국가의 최종 목표가 될 수 있을 것이다.

구체적으로, 북유럽에서는 인구통계, 사업체통계, 소득통계, 자산통계, 교육통계, 범죄통계, 거처 및 건축물/거처 통계를 작성하기 위한 행정자료를 구비하고 있으며, 필요시 유지·보완하기 위한 기초조사를 실시한다.. 가령, 행정자료의 집계범위나 포괄변수에 대하여 전반적 품질을 평가하기 위하여 다른 자료 가령 표본조사 자료를 사용할 수는 있지만, 통계작성에 이용되는 행정자료를 근본적으로 변경하기 위한 것이 아니다. 따라서 100% 완전한 등록센서스를 실시하는 국가들은 표본조사를 실시하기는 하지만, 그 역할은 상당히 제한적이다..

어떤 경우에는, 행정대장 자료와 표본조사 자료를 결합하여 센서스 통계를 작성할 필요가 있다. 가령, 가구별 소득통계이다. 가구대장이 없는 국가에서

는, 가구구성에 관한 표본조사 자료를 소득에 대한 행정자료와 결합한다. 예컨대, 덴마크의 「가계조사」(Danish Household Budget Survey)와 유럽연합 통계청(Eurostat)· 덴마크 통계청(Statistics Denmark)이 공동주관한 「소득 및 생활실태 통계」(SILC, Statistics on Income and Living Conditions)의 덴마크 부분에서, 행정 자료를 이용하여 가구를 조사객체로 실시된 표본조사 자료를 보완한다. 이 경우 사용되는 행정자료의 정보는 소득 자료와 직업, 교육, 주택에 대한 자료다.

행정자료와 표본조사의 자료를 결합할 때, 가중치(weighting)는 매크로 수준에서 정합성을 개선하는 중요한 방법이다. 가령, 네덜란드의 가상센서스(Virtual Census) 시스템에서, 매크로 데이터통합(macro-data integration)이라는 명목으로 가중치를 부여하는 방법이 포괄적으로 이용되고 있음을 볼 수 있다(Linda, 2003).

행정자료는 센서스 통계는 물론 교육, 경제활동 등 주제별 통계작성에 필요한 모든 변수를 포함하지 않는다. 이러한 경우, 센서스의 일부로서 표본조사를 통하여 추가적 정보를 획득할 필요가 있다. 마이크로 수준의 대체보정 방법(imputation by substitution)은 매크로 수준에서 통계를 결합하는 대안이다. 대체보정방법은 적합성이 있는 표본조사를 선택하여, 그 자료에 포함된 정보에서 추정치를 차용하는 것이다. 다양한 종류의 모형을 구축할 수 있으며, 이들은 대부분 일종의 층화(stratification) 개념과 각 계층 내 변수의 확률분포를 가정한다는 것이 특징이다. 이러한 대체보정방법은 그 결과가 마이크로 수준이 아니라 매크로 수준에서만 타당성이 입증될 수 있기 때문에 이용자들을 만족시킬 수 없다는 단점이 있다.

3. 행정자료의 집계범위 추정을 위한 근거자료

등록센서스의 추진과정에서 행정자료의 집계범위, 특히 과소집계를 추정하기 위하여, 표본조사를 실시하는데, 이것은 현행 인구주택조사 본조사의 실시 후에 본조사의 집계범위를 추정하기 위하여 실시하는 사후조사와 사실상 같다. 행정자료의 집계범위를 추정하기 위하여, 행정자료에 포함된 변수와 집계범위를 추정하기 위하여 실시하는 표본조사에 포함된 변수를 매칭하여, 그것을 토대로 하여, 행정자료의 포괄범위를 추정하고, 행정자료의 문제점을 보완하여 센서스통계의 정확성을 개선하고자 노력한다.

행정등록자료와 표본조사의 매칭작업에는 매칭키(matching key)로 개인식별번호(PIN)를 사용할 수 있는데, 이것이 매칭작업의 효율성을 극대화하는데

가장 좋은 변수이기는 하지만, 개인정보의 보호 때문에, 개인식별번호를 이용할 수 없는 경우에는 성명, 성별, 연령, 생년월일, 가구주와의 관계 등 다양한 변수를 이용하여, 매칭률을 극대화시키는 방법을 택하지 않으면 안 될 것이다.

대부분의 국가에서는 전통적 센서스의 집계범위를 추정하기 위하여, 센서스와 전혀 다른 사후조사시스템을 이용하여 센서스 본조사와 사후조사의 결과를 비교하는 이원시스템추정법(DSE, dual system estimation)을 사용하고 있다. 이 방법은 Chandraseka과 Deming이 1940년대에 개발한 것으로, 미국 Census Bureau이 1980년과 1990년 센서스에서 과소집계를 추정하기 위하여 이용하였다.

<표 5-3> 이원시스템 추정방법(DSE)

자료별 집계여부		행정자료의 자료		
		예	아니오	합계
센서스 표본조사	예	N ₁₁	N ₁₀	N ₁₊
	아니오	N ₀₁	N ₀₀	N ₁₊
	합계	N ₊₁	N ₊₁	N ₊₊

추정의 대상: N₊₊ (행정자료의 자료에서 등록(또는 집계)되어야 할 개인이나 가구(세대)의 총수)

이원시스템 추정법의 기본가정: 행정자료의 집계방식과 인구주택총조사 표본조사의 집계방식은 동일하지만, 완전히 별도의 시스템이다. (사실상 이 가정을 충족시키도록 주민등록법과 관련 법률을 개정하거나 이를 위한 등록센서스 준비법을 만드는 것이 대단히 중요한 과제이다)

$$N_{++} = N_{+1} \times N_{1+} \div N_{++}$$

(특정 지역의 인구 또는 가구(세대)의 총수는 인구주택총조사 표본조사 자료의 해당 총수와 주민등록 자료의 해당 총수를 두 개의 별도의 시스템에서 매칭에 성공한 총수로 나눈 것이 된다)

이원시스템추정법은 등록센서스를 실시 또는 계획하고 있는 싱가포르, 이스라엘, 독일에서도, 등록센서스의 주축자료인 인구대장의 등록범위(누락, 중복)를 점검하기 위하여 사용하였다. 또, 등록센서스로 이행을 검토하고 있는 남유럽의 스페인과 이탈리아에서 이원시스템추정법을 이용하여, 인구대장의

정확성 곧 등록범위를 점검하는 작업을 계속하고 있다. 이것은 사실상, 인구주택총조사가 제대로 조사객체를 파악하지 못할 수 있다는 개연성을 전제로 하여, 실제인구를 추정하는 절차이다. 이 방법은 이론적으로는 간단하지만, 현실적으로 발생할 수 있는 복잡한 문제는 다음과 같다

이원시스템추정법(DSE)을 이용하기 위한 전제는 행정등록자료와 인구주택총조사 표본조사에서 특정의 개인 또는 가구(세대)가 집계될 확률이 대상인구의 어떤 하위집단에서도 차이가 없다는 것이다. 2015년 등록센서스 추진에서 해결해야 할 가장 중요한 문제는 이 가정의 문제를 통계적으로 해결하는가 아니면, 제도적·법률적으로 해결하는가하는 것이다.

다음은 추정결과에 편향(偏向)이 생겨나지 않도록 하기 위해서는 행정자료 시스템과 센서스 표본조사 시스템이 특정의 개인 또는 가구를 등록 또는 집계할 확률이 통계적으로 상호독립이라고 가정한다. 그러나 이 가정도 앞에서 제시한 가정과 마찬가지로 현실적으로 문제가 있을 것으로 기대되고 있다.

마지막으로, 특정 리스트에서 집계된 개인 또는 가구가 다른 리스트에서 집계되었는가를 확인하기 위하여, 두 자료원에 대한 완전무결한 100%의 매칭작업이 필요하다. 그러나 매칭오류(matching error)는 이원시스템추정법에 편향성을 야기하는 직접적 원인이 된다.

첫 번째 가정의 타당성을 검정하기 위하여, 도시/농촌별, 응답형태(인터넷조사방문조사, 전화조사), 또는 가구규모별로 또는 성·연령(5세간격)별로 하위집단으로 분할하여, 이들 하위집단에 대하여 이원시스템추정법을 적용하여 볼 필요가 있을 것이다. 이러한 절차는 '두 통계시스템의 상호독립성'과 더불어, 두 번째 가정의 타당성도 확보될 수 있는가를 검정하기 위한 것이라고 할 수 있다. 두 번째 가정의 실패로 말미암아 생겨난 종속성의 문제가 추정결과에 주는 영향에 관한 연구들은 인구주택총조사의 집계범위와 주민등록시스템의 등록범위가 상대적으로 높은 수준에 있을 경우에는 두 시스템 간에 약간의 종속성이 있다고 하더라도, 추정방법은 비교적 양호한 결과를 도출할 수 있을 것이라고 생각한다. 그러나 문제는 '두 조사의 등록(집계)범위 수준이 높아야 한다.'는 가정설정이 사실상 실패하는 경우가 있으며, 이로 말미암아 종속성 문제를 보정하는 절차가 이원시스템추정법의 현장적용에 필수요건으로 등장하게 될 수밖에 없다.

세 번째 가정의 '매칭오류가 없어야 한다.'는 가정도 완전히 문제가 없는 것은 아니다. 그러나 두 개의 별도의 조사시스템에 대하여 최적의 매칭방안을 고안해 낼 수 있다면, 매칭오류, 특히 매칭실패의 확률을 최소화하는 방

향으로 설계되었다는 점에서, 매칭오류로 인한 추정실패는 일어나지 않아야 한다고 보아야 할 것이다.

4. 전통센서스 표본항목(또는 전수항목의 일부) 자료획득의 도구

전통적 센서스가 처음으로 도입되었을 때는 대부분의 국가에서 조사항목 자체가 몇 개에 불과하였다. 우리나라도 1925년 5월 1일, 일본 식민지 치하에서 처음으로 근대적 의미 센서스를 실시하였을 때, 조사항목은 성명, 성별, 출생연월, 혼인상태, 본적 또는 국적 등 5개로 제한되어 있었다. 해방 이전 일제 치하에서 실시된 대부분의 센서스는 조사항목이 10개미만으로 모두 전수조사로 실시되었다. 그러나 1960년 대에 들어서서 조사항목이 인구분야 18개 항목, 가구·주택분야 18개 항목으로 비약적인 증가를 하였는데, 일반적으로 끝이 '0'으로 끝나는 해에는 '5'로 끝나는 해보다 조사항목의 수가 많지만, 최근에는 그러한 경향이 다소 완화되기 시작하였다. 앞에서 지적한 것처럼, 2005년 우리나라 센서스의 조사항목은 인구분야 27개 항목, 가구·주택분야 17개 항목, 시도별 항목이 3개였으며, 2010년에는 인구분야 28개 항목, 가구·주택분야 19개 항목이고, 시도별 항목이 3개가 추가될 예정으로 되어 있다.

우리나라에서는 인구주택총조사에 표본조사가 도입되기 시작한 것은 1966년부터였는데, 당시에, 인구분야에 6개의 조사항목, 곧 총출생아수, 지난 1년간 출생아수, 취업여부, 종사상의 지위, 산업, 직업 등이 조사항목으로 도입되었다. 이러한 표본항목은 1970년에 16개, 1975년에 17개, 1980년에 19개로 증가하였다가, 1985년에는 표본항목이 폐지되었다가, 1990년에 12개, 1995년에 11개에 머물다가 2000년에는 전체 조사항목이 50개로 증가하면서 표본항목이 30개로 증가하였다. 한편 최근 인구주택총조사인 2005년의 경우에는 표본항목이 23개로 감소하였다가, 2010년에는 표본항목이 31개로 다시 증가했다.

유엔은 『인구주택센서스 원칙 및 권고안』(Principles and Recommendations for Population and Housing Censuses)에 의하면, 센서스 표본조사는 (1) 정확성과 정밀성의 관계 (2) 센서스 실시에 소요되는 인적·경비자원을 고려할 것을 권고하고 있다.

먼저, 센서스 표본조사는 표본추정치에서 어느 정도의 정밀도가 요구되는가를 염두에 둘 필요가 있다는 것이다. 높은 정밀도가 요구되면 될 수록, 표본의 규모는 커지고 절차는 복잡해지고, 소요되는 비용도 많아지게 마련이

다. 표본추정치(estimate)의 정밀도와 정확도는 그 개념에서 차이가 있다. 정밀도는 표준오차(standard error)에 의하여 측정되는데, 이것은 바로 동일한 조사여건 아래서 전수조사를 한 결과와 비교하여 표본조사를 했을 때 생겨나는 오차의 측정치가 될 수 있다. 한편, 정확도는 일반적으로 미지수에 해당하는 참값(true value)과, 실제 조사(전수조사이든 표본조사이든 관계가 없음)에서 획득한 수치 사이에 발생하는 차이에 의하여 측정된다(<그림 4-3>을 참고할 것). 정밀도와 정확도는 상호 밀접한 관계를 갖는데, 그 이유는 정밀도가 추가적 측정이 동일한 결과를 생산하는 정도를 말하는 재생가능성(reproducibility) 또는 반복가능성(repeatability)의 의미를 지니기 때문이다.

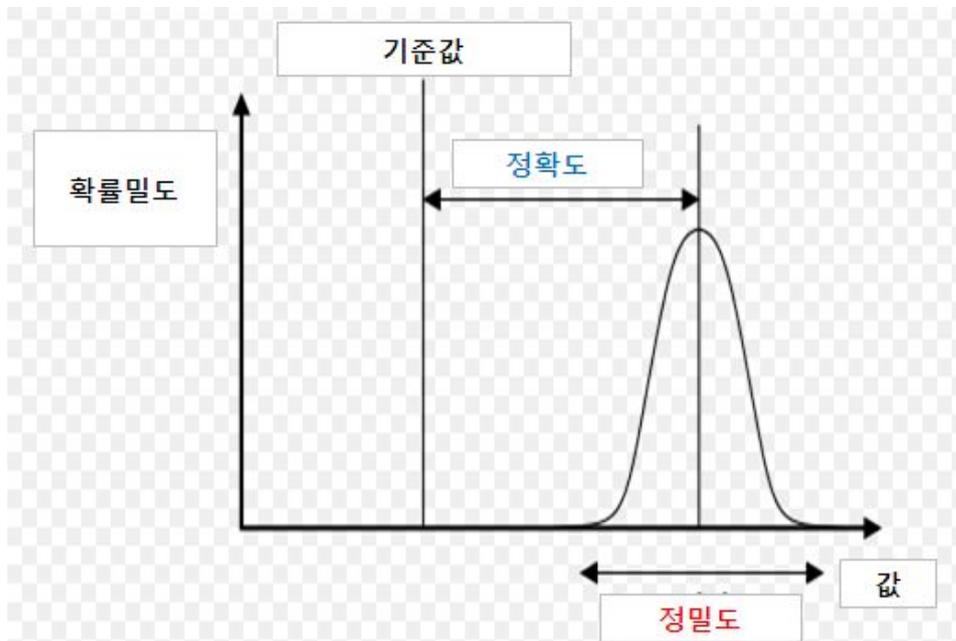
센서스에서 표본조사는 극히 예외적인 경우, 가령 시험조사를 제외하고는 확률표집(probability sampling)에 의하여 표본선정을 할 것을 권고하는데, 여기에는 판단표집(judgemental sampling), 유의표집(purposive sampling) 등의 비과학적 표본추출을 이용해서는 안 된다고 권고하고 있다. 확률표집에 의하여 성공적으로 표본을 추출하기 위해서는 과학적으로 설계된 표본선정절차를 엄격히 준수하지 않으면 안 된다. 표본선정절차는 먼저, 모집단의 조사객체 하나하나의 단위에 대하여 표본으로 선정될 확률치를 부여하는데서 시작한다. 이들 확률치의 역수(逆數, inverse)를 계산하여, 이것을 토대로 하여 모수를 추정하고 표본추정치의 정밀도, 곧 표집오차(sampling error)를 계산하여야 한다. 표본선정절차는 당초의 표본설계에 충실하여야 하며, 정해진 기준이나 지침으로부터의 이탈은 최소화되어야 마땅할 것이다.

물론, 센서스 표본조사를 토대로 하는 추정결과는 전수조사에 존재하는 비(非) 표집오차의 다양한 형태는 물론 표집오차에 기인하는 오차가 있기 마련이다. 그럼에도 불구하고, 전수조사보다 소규모의 표본조사를 실시하게 되면, 유능한 현장조사원을 이용하여, 상세한 조사항목을 개발하여 그것에 대한 응답을 획득하여 응답오류를 극소화할 수 있을 것이다. 결론적으로, 추정치의 정확도에 영향력을 미칠 수 있는 비(非) 표집오차는 전수조사보다 표본조사에서 축소될 수 있을 것이다.

센서스 표본조사의 결과를 제표화할 경우에, 적어도 주요 조사항목에 대해서는 표집오차(분산)의 추정치를 함께 제시하는 것이 마땅할 것이다. 다양한 기법이 표집오차를 계산하는데 사용될 수 있겠지만, 계산방식은 실제로 이용된 표본설계를 반영하는 것이 되어야 할 것이다.

다음은 센서스 실시와 관련된 인적·금전적 비용과 관련된 것으로, 유엔의 권고안(United Nations, 2007)은 센서스 표본조사의 효율적인 기획은 해당 국가에서 가용한 모든 전문적 지식이나 장비를 총괄적으로 이용하여야 한다

<그림 5-2> 정확도와 정밀도의 도식화



는 의견을 제시하고 있다. 동일한 목표를 지향하는 특정의 표본계획도 센서스 자원의 질이나 양에 따라서 국가별로 차이가 있을 수 있다. 전통적 센서스나 등록센서스이든 관계없이, 센서스 전체의 일부로서 표본선정에 의한 현 지실사를 기획할 때는, 가용한 예산지출이나 전문가 지식을 충분히 고려하여야 할 것이다.

센서스에서 전통적 방식이 조사비용의 문제를 야기하는 만큼이나, 표본조사의 이용은 비용문제가 대단히 중요하며, 동일한 조사항목을 전수조사를 통하여 수집하지 않기로 결정하게 되는 일차적 이유가 될 수도 있을 것이다. 다양한 요인들이 표본추출의 비용을 결정하게 되며, 특정의 조사항목을 전수조사로 할 것인가 표본조사로 할 것인가에 대해서 최종결정을 내리기 전에 다양한 요인들의 장점과 단점을 비교 검토하여야 할 것이다. 가령, 한 가지 중요한 요인은 표본의 규모와 복잡성으로 이들은 조사의 목표나 가장 효율적이라고 판단되는 절차에 의하여 결정되는 경향이 있음을 유념하여야 할 것이다.

센서스 표본조사는 유능한 통계학자의 지침 아래서 실시되어야 하며, 표본추출, 표본자료 통계분석, 표본조사의 현장운영 등의 이론에 익숙한 통계학자가 책임통계관(responsible statistician)으로서 센서스를 관리할 필요가 있음을 강조한다. 표본조사의 매 단계, 가령, 표본설계의 기획단계나 표본추정치와 오차의 계산단계에서 유능한 통계학자의 시의성 있는 자문이 절대적으

로 필요하다.

<표 5-4> 우리나라 인구주택총조사의 역사

회수	조사기준	명 칭	특 징
1	1925.10. 1.	간이국세조사	최초의 인구총조사
2	1930.10. 1.	국 세 조 사	최초로 직업등 경제활동사항 포함
3	1935.10. 1.	국 세 조 사	
4	1940.10. 1.	국 세 조 사	
5	1944. 5. 1.	인 구 조 사	
6	1949. 5. 1.	총인구조사	최초로 인구이동 사항 포함
7	1955. 9. 1.	간이총인구조사	전동천공기 도입
8	1960.12. 1.	인구주택국세조사	① 최초로 주택에 관한 조사 실시 ② 노동력 개념 설정 ③ 경제활동 및 출산력 사항의 20% 표본집계
9	1966.10. 1.	인 구 센 서 스	10% 표본조사 병행(경제활동 및 출산력)
10	1970.10. 1.	총인구 및 주택조사	10% 표본조사 병행(경제활동, 출산력, 인구이동 및 일부 주택에 관한 사항)
11	1975.10. 1.	총인구 및 주택조사	5% 표본조사 병행(경제활동, 출산력, 인구이동 및 일부 주택에 관한 사항)
12	1980.11. 1.	인구 및 주택센서스	15% 표본조사 병행(경제활동, 출산력 및 인구 이동)
13	1985.11. 1.	인구 및 주택센서스	① 전항목 전수조사 ② 성씨, 본관 및 종교에 관한 조사실시
14	1990.11. 1.	인구주택총조사	① 10% 표본조사 병행(경제활동, 출산력 및 인구이동) ② 교통관련 항목 추가 ③ 자료처리 광학판독(OMR)방식 도입
15	1995.11. 1.	인구주택총조사	① 10% 표본조사 병행(경제활동, 인구이동, 임차료 등) ② 최초로 빈집에 관한 조사실시 ③ 자료처리 광학판독(OMR)방식 활용 ④ 조사구 설정용 기본도 전산화
16	2000.11. 1.	인구주택총조사	① 10% 표본조사 병행(경제활동, 인구이동 및 임차료 등) ② 지식기반, 정보화, 복지관련 항목 신규조사 ③ 수치지도(digital map)를 이용한 조사구 설정 및 유도작성 ④ 현지분산형 PC입력방식 도입
17	2005.11. 1.	인구주택총조사	① 10% 표본조사 병행(경제활동, 주거실태, 여성·아동·고령화 등) ② 건축물대장 등 행정자료 활용을 통한 조사의 효율성 및 정확성 제고 ③ e-census 시스템 활용 및 인터넷조사 신규 도입 ④ Web에 의한 시군구 분산입력

표본선정이 표본설계의 확률표집 원칙에 의거하여 엄격하게 선정되었다는 것을 확실히 하고, 표본선택의 어떠한 편향 가능성을 배제하기 위하여, 표본

의 실제 선정 작업은 통계청 본청 또는 지방사무소에서 책임 통계관의 직접 통제 아래 이루어져야 한다는 점을 또한 강조한다.

제4절 현행 인구주택총조사의 조사방법론

1. 인구주택총조사의 간단한 역사

일본은 식민지 통치를 위하여 1919년 6월에 총독부령 제 103호로 국세조사규칙을 제정하여 일본과 동시에 1920년 국세조사를 실시하려고 했으나 1919년 3월 1일 독립운동으로 국세조사의 실시를 포기하였다. 1922년에는 법률 제 51호로 국세조사법을 개정하여 10년마다 시행하기로 되어 있는 조사주기 사이에 5년마다 간이조사를 할 수 있도록 개정했다. 1925년에는 총독부령 제 66호 「1925년 간이국세조사에 관한 건」에 의하여 10월 1일 0시 현재로 국내에 거주하고 있는 자를 대상으로 조사했는데 이 조사가 우리나라 근대적 인구센서스의 효시다. 이후 일제는 식민지인 조선에서 노동력 착취 및 경제수탈을 위해 '국세조사' 또는 '간이국세조사'라는 명칭으로 매 5년마다 인구센서스를 실시하여 왔으나 조사 결과는 엄격히 비밀로 처리되었다.

1948년 정부수립 직후 각종 정책수립을 위한 인구통계가 시급히 필요한 관계로, 정부는 1950년 실시 예정인 국세조사를 1년 앞당겨 1949년 5월 1일에 남한에서만 제1회 총인구조사를 실시했다. 조사표는 한국전쟁(1950-53)으로 소실되고, 각 특성별로 통계표를 작성하지 못하고 단지 총인구 규모 파악에 그쳤다.

1960년 UN 권고에 의한 「인구, 주택 및 농업 총조사 프로그램(World Census Programme)」을 계기로 우리나라의 센서스에도 큰 변화가 일어났다. 이 조사는 조사기획 단계부터 자료처리 및 평가에 이르기까지 명실 공히 현대적인 센서스의 면모를 갖추어 오늘날 센서스의 발전을 가져오는 밑바탕을 마련했다. 이 1960년 국세조사에서 처음으로 주택부문이 병행조사 되었으며 총조사의 정확성을 평가하기 위한 사후조사를 처음으로 도입했다. 또한, 최초로 상주인구 개념의 조사방법과 표본조사 항목을 집계하는 방안이 적용되었다.

1966년 센서스는 제1차 경제개발 5개년계획의 투자자원 확보에 따른 예산부족으로 당초 1965년에 실시하려던 계획을 1년 연기하여 1966년에 실시했는데 우리나라에서는 처음(1967년)으로 컴퓨터(IBM 1401)를 도입하여 자료를 전산처리함으로써 센서스의 통계처리에 큰 변화가 일어났다. 아울러 최초로

표본조사방법이 도입되어 한층 발전된 모습을 갖추게 되었다.

1970년 및 1975년 인구주택총조사는 각기 10월 1일 현재로 실시되었는데 기본항목은 전수조사로 나머지 항목은 표본조사방식으로 조사되었다. 1980년 총조사는 전공학과와 통근학 관련 항목들이 새로이 추가되었고 1985년 총조사에서는 전항목을 전수조사로 통합해서 실시하면서 성씨와 본관에 관한 항목을 포함했다. 1990년 총조사에서는 자료처리방식을 그동안의 키엔트리(key-entry)입력방식에서 처음으로 광학판독(OMR)기법을 도입하였는데 이는 자료처리를 정확·신속히 하여 결과를 조기 이용토록 하는데 크게 이바지했다. 2000년 총조사에서는 수치지도사용으로 조사구 설정 업무의 정확성을 제고했으며 아파트 전수지역에 대한 자계식 조사방법 적용으로 조사비용을 크게 절감했다. 자료처리에서는 지역 분산 외주 자료입력(key-entry) 방식을 도입해 자료처리기간을 단축시켰다.

가장 최근에 실시된 2005년 총조사에서는 저출산, 고령화, 주거의 질 등 복지 관련 항목을 강화하였으며 조사관리 부문에서는 e-census시스템을 도입하여 행정지원업무를 대폭 간소화하였다. 아울러 현장조사방법에서는 인터넷조사를 도입하여 1인 가구 및 주간 부재가구 등 면접이 어려운 계층의 조사 참여율을 제고하였으며 차기 센서스에서 인터넷조사 확대 도입을 위한 다양한 경험을 축적할 수 있었다. 자료 처리 면에서는 Web 분산입력방식을 채택하여 자료처리기간을 크게 단축시켰다.

2. 조사객체, 조사방법, 조사체계 및 조사내용

가. 조사대상

우리나라의 현행 인구주택총조사는 인구, 가구, 주택을 조사객체로 하여 전수조사를 실시한다. 조사객체의 경우, 사람 곧 자연인을 조사하는 방법에는 2가지가 있다. 첫째는 상주지(de jure) 방식인데 이는 평소에 살고 있는 장소에서 조사하는 방식으로 잠시 집을 떠나 있는 사람도 평소에 살고 있는 집에서 조사한다. 우리나라를 비롯하여 주로 유럽이나 북미국가에서 적용하고 있다. 다음은 현주(de facto) 개념으로 조사 당시 발견된 장소에서 조사하는 방식으로 아프리카, 남미 등에서 주로 사용하고 있다.

가구는 생계를 공동으로 하는 하나의 집합을 가구로 간주하는 가사단위 개념(housekeeping unit concept)으로 파악하는 방법과 동일주택에 거주하는 모든 사람을 하나의 가구 단위(house-dwelling unit concept)로 파악하는 방

법이 있다. 우리나라에서는 가사단위로 가구를 파악하는데 이 개념에서는 하나의 주택에 여러 가구가 살 수 있기 때문에 가구를 집계하는 과정에서 특히 유의해야 한다.

주택은 가구가 외부와 독립적으로 생활(취침, 식사, 외부로부터의 보호 등)을 할 수 있도록 구조적으로 분리된 장소를 말한다. 우리나라에서는 ① 영구건물 ② 부엌과 한 개 이상의 방 ③ 독립된 출입구를 갖고 있으면서 ④ 분리해서 사고 팔 수 있어야 주택으로 간주한다. 이러한 요건을 어떻게 적용하는가에 따라 주택 수가 달라진다.

나. 조사방법

현행 인구주택총조사의 조사방법은 조사원이 응답자에게 질문하여 조사표를 작성하는 조사원 면접방식과 응답자가 배부된 조사표를 스스로 작성하는 응답자 기입방식으로 구분한다. 응답자 기입방식은 다시 조사표를 배부하고 회수하는 방식에 따라 우편배부·우편회수(mail-out/mail-back), 조사원배부·조사원회수(enumerator-delivery/ enumerator-collect), 우편배부·조사원회수(mail-out/enumerator-collect) 등으로 구분된다. 우리나라는 전통적으로 조사원 면접방식을 사용했으나 2005년 총조사의 경우 조사원 면접방식과 더불어 응답자 기입방식의 조사원 배부·조사원 회수 방식을 적용했다. 아울러 최초의 인터넷조사도 도입하여 활용하였다.

다. 조사체계

조사체계는 통계청↔시·도↔시·군·구↔읍·면·동↔조사관리자↔조사원으로 이루어진다. 2005년 총조사의 경우 통계청은 인구주택총조사의 예산, 조사항목 등 조사기획 및 분석업무를 했고 시·도(48명) 및 시·군·구(768명)는 지역별 실시체계 구축 및 현황 파악과 조사원 교육을 담당했으며 읍·면·동(3,566명)에서는 조사관리자 및 조사원 관리와 주민 홍보를 담당했다. 조사관리자(총관리자 포함 11,259명)는 조사원(98,964명)의 현장조사를 지도했다.

라. 조사내용

조사내용은 조사항목에 따라 크게 전수지역과 표본지역으로 구분한다. 전

수지역은 조사하고자 하는 모든 조사대상에 대해 조사를 하는 반면 표본지역은 그 중 일부 지역만 뽑아 심층적인 내용을 조사한다. 이는 심층적이고 응답하기 어렵거나 개인의 프라이버시와 관련이 많은 내용은 전체 대상을 모두 조사하기보다는 표본으로 조사함으로써 국민의 응답부담을 줄이고 조사의 정확도도 높이기 위한 것이다. 우리나라는 1966년 총조사에서부터 이 방식을 도입하였는데 표본의 규모는 전체의 10%이다. 2005년 총조사의 경우 모든 가구를 조사하는 전수조사 항목은 21개였으며 이 중 10%에 대해서 추가로 조사하는 표본항목의 수는 23개였다. 표본항목 중 3개는 시도별 특성항목으로 지역통계 수요에 부응하기 위하여 각 시도에서 선정·조사하였다.

<표 5-5> 인구주택총조사 실시연도별의 조사항목수(전수항목과 표본항목)

	1970	1975	1980	1985	1990	1995	2000	2005	2010
- 계	31	28	40	28	45	28	50	44(3)	50(3)
전 수	14	11	22	28	33	17	20	21	19
표 본	17	17	18		12	11	30	23(3)	31(3)

※ ()내는 2005년과 2010년(예정) 총조사의 시도별 특성항목의 수입

2005년 총조사의 전수항목 및 표본항목은 다음과 같다.

【전수조사 항목 : 21개】

- 인구(8개) : 성명, 성별, 가구주와의 관계, 나이, 종교, 교육정도, 남북이산 가족, 혼인 상태
- 가구·주택(13개) : 가구구분, 사용방수, 주거시설 형태, 난방시설, 거주층, 점유형태, 주인가구 및 주택소유 여부, 거처의 종류 및 건물층수, 연건평, 대지면적, 총방수, 건축년도, 편익시설수

【표본조사 항목 : 23개】

- 인구(16개) : 아동보육 상태, 5년 전 거주지, 활동제약, 통근·통학여부, 통근·통학장소, 이용교통수단, 통근·통학 소요시간, 경제활동상태, 종사상지위, 산업, 직업, 근로 장소, 혼인 연월, 출생아수, 추가계획자녀수, 고령자 생활비 원천
- 가구(4개) : 거주기간, 자동차보유대수, 주차시설, 임차료
- 시도특성항목(3개) : 시도별 각 3개 항목

한편, 2010년에 실시예정인 총조사의 전수항목 및 표본항목은 다음과 같다.

【전수조사 항목 : 19개】

- 인구(8개) : 성명, 성별, 가구주와의 관계, 나이, 교육정도, 혼인상태, **국적(신규), 입국연월(신규)**
- 가구·주택(11개) : 가구구분, 사용방수, 주거시설 형태, 점유형태, 건물 및 거주층, 거처의 종류, 주거용 연면적, 건축연도, 총방수, 주거시설 수, 주인 가구 및 타지주택 소유 여부

<그림 5-3> 2010년 인구주택총조사의 조사항목(전수항목과 표본항목)(안)

	전수항목		표본항목			
	5년 주기	신규	5년 주기	10년 주기	신규	
UN 권고 항목 (40)	인구	① 성명 ② 성별 ③ 나이 ④ 가구주와의 관계 ⑤ 교육 정도 ⑥ 혼인상태	① 국적 ② 입국연월	① 5년 전 거주지 ② 경제활동상태 ③ 종사상 지위 ④ 산업 ⑤ 직업 ⑥ 근로장소 ⑦ 총 출생아 수 ⑧ 혼인 연월 ⑨ 통근학 여부 ⑩ 통근학 장소 ⑪ 이용교통수단 ⑫ 통근학 소요시간 ⑬ 활동제약	① 출생지 ② 1년 전 거주지	-
	가구	① 가구구분 ② 사용방수 ③ 주거시설형태 ④ 점유형태 ⑤ 건물 및 거주 층	-	① 난방시설 ② 주차장소 ③ 임차료	① 수도 및 식수 사용형태 ② 정보통신기기 보유 및 이용현황	① 교통수단 보유 및 이용현황
	주택	① 거처의 종류 ② 주거용 연면적 ③ 건축연도 ④ 총 방수 ⑤ 주거시설 수	-	① 대지면적		
고유 항목 (7)	인구		① 아동보육 ② 추가계획 자녀 수 ③ 고령자생활비 원천	① 현 직업 근무연수	① 사회활동	
	가구	① 주인가구 및 타지 주택 소유여부	-	① 거주기간		

【표본조사 항목 : 31개】

- 인구(20개) : 5년 전 거주지, 경제활동상태, 종사상 지위, 산업, 직업, 근로장소, 총출생아수, 혼인연월, 통근·통학 여부, 통근·통학 장소, 이용교통수단, 통근·통학 소요시간, 활동제약, 출생지, 5년전 거주지, 아동보육상태, 추가계획자녀수, 고령자 생활비 원천, 현 직업의 근무연수, **사회활동(신규)**

- 가구·주택(7개) : 거주기간, 난방시설, 주차장소, 임차료, 수도 및 식수사용 형태, 정보통신기기 보유 및 이용현황, 교통수단 보유 및 이용 현황(신규), 대지면적
- 시도특성항목(3개) : 시도별 각 3개 항목

나. 조사과정

인구주택총조사의 조사과정은 조사구 설정 및 지도제작, 준비조사, 본조사, 자료입력, 결과 집계 및 제공으로 구분된다.

1) 조사구(조사 구역) 설정 및 지도제작

조사구(ED, Enumeration district)란 우리나라의 모든 지역에 대하여, 도로, 하천 등 상대적으로 변화가 적은 지형지물을 이용하여 지도(읍면동별)상에 일정 가구수(평균 60)가 포함되도록 조사구역을 명확히 구획하는 작업이다. 조사구 설정 목적은 조사원의 조사담당 구역을 명확히 하여 조사객체(인구, 가구, 주택)의 누락 및 중복을 방지하고 조사원 업무량의 적정배분 및 향후 각종 표본조사의 추출틀로 활용하는 것이다.

조사구는 일반조사구와 특별조사구로 구분한다. 일반조사구는 일반인이 거주하는 곳으로 조사원이 자유롭게 출입할 수 있는 지역에 설정한 조사구를 말한다. 이는 다시 아파트, 보통, 섬, 기숙시설, 특수사회시설, 관광호텔 및 외국인 거주지역 조사구 등으로 세분된다. 2005년 총조사에서 일반조사구는 총 27만 3천개였다. 반면, 특별조사구는 조사원의 출입이 제한되거나 조사원에 의한 조사가 불가능한 지역으로 재외주재공관, 경찰서, 전투(의무)경찰대, 교도소(소년원), 군부대 등에 대하여 별도로 설정한 조사구이다.

조사구 설정 절차는 본조사 1년 전에 전 지역을 대상으로 경계 및 지형지물을 보완하고 그 바탕위에 변화가 예상되는 지역은 각 건물(거처)별 가구수를 파악하고 그렇지 않은 곳은 과거총조사 자료를 활용하여 20가구 단위로 거처묶음을 실시하는 기초조사를 실시한다. 기초조사 자료를 이용하여 통계청에서 조사구 설정 기준에 의해 경계구획 및 번호를 부여하는 조사구를 가설정하게 되면 지자체에서 가설정 조사구 적합성 확인 및 보완을 거쳐 통계청에서 조사구 설정 확정승인을 하게 되고 같은 해 10월 말까지 변동지역을 파악(지자체)하여 보완한 후 본조사에 활용하게 된다.

조사구설정 관련지도는 기본도1), 부분확대도2), 조사구요도3), 공동주택전개

도4)가 있는데 특히 2005년 총조사에서는 새주소지도를 도입하여 새주소 지역은 새주소 지도를 활용하고 새주소 이외 지역은 기존 수치지도를 활용하여 조사용 지도를 작성함으로써 정확성과 효율성을 향상시키고자 노력하였다.

이렇게 작성된 기본도는 조사기간(준비조사 포함) 동안 총조사 상황실 벽면에 부착되어 조사원별 담당지역 관리, 조사구별 조사 진행상황의 파악 등 조사관리에도 널리 이용되며 조사구요도는 가구명부와 함께 조사객체를 파악 관리하는데 유용하게 활용된다.

2) 준비조사

준비조사는 본조사 바로 직전에 표본지역은 2일, 전수지역은 3일간 실시하게 된다. 조사지역 확인, 조사대상 가구 리스트 작성, 조사구 요도(지도) 보완, 공동주택기본정보 수집 등과 같이 본조사를 준비하는 일을 한다.

3) 본조사

본조사는 15일간(2005년 총조사의 경우 11월1일부터 11월 15일까지) 실시된다. 이 기간 동안 조사원은 가구를 방문하여 자기에게 주어진 지역을 조사하게 된다. 2005년 총조사의 경우 전수지역 중 아파트 지역은 5개 조사구, 아파트가 아닌 지역은 3개 조사구, 표본지역은 2개 조사구가 주어졌다. 조사원은 매일매일 읍·면·동에서 지정한 장소로 출근하여 조사관리자 및 공무원의 도움을 받아 조사를 한다.

4) 자료입력

자료입력(data capture)은 조사표에 조사된 자료를 컴퓨터가 읽을 수 있는 형태로 바꾸는 작업이다. 우리나라 총조사에서는 1966년 센서스 자료를 전산

- 1) 국토지리정보원 및 지방자치단체에서 제작한 수치지형도 및 새주소지도를 이용하여 읍·면·동 단위별로 관내의 모든 지역이 포함되도록 전산으로 접합·절단하여 조사구 설정에 적합하도록 제작한 지도를 의미하며, 도시지역은 1/1,000, 도시외곽 및 평야지역 1/5,000, 산간 및 섬지역은 1/25,000을 사용하였다(일본 : 1/50,000, 미국 : 1/100,000입).
- 2) 거처(건물)가 밀집되어 식별이 곤란한 지역에 대해 조사구설정이 가능하도록 축척을 크게 하여 별도로 확대 제작한 지도를 말한다.
- 3) 거처(주택)의 중복과 누락을 방지하기 위해 기본도상에 설정한 조사구 경계, 지형지물(도로, 하천, 건물 등)과 그 명칭, 거처 및 거처번호, 거처별 가구수 등을 일정한 규격(A₃)에 표기한 약식지도를 말한다.
- 4) 아파트, 연립, 빌라 등을 정면에서 바라본 모양을 층수, 호수, 출입구수 등이 나타나도록 작성한 지도를 말한다.

으로 입력한 것이 처음이다. 이 후 1985년 인구주택총조사까지는 키 엔트리(key-entry) 방식으로 통계청에서 입력했다. 이 방식에 변화를 가져온 것은 1990년 총조사에 도입한 광학판독방식(optical marker reader: OMR)이다. 이 방식은 1995년까지 사용되었는데 자료처리의 정확성은 물론 처리기간이 단축되는 효과가 있었다. 2000년에는 자료입력방식을 경제성과 신속성을 이유로 다시 전통적인 키엔트리 방식으로 전환했는데 외주에 의한 지방 분산방식을 적용하였다. 2005년에는 Web에 의한 시군구 분산입력방식을 채택하여 자료처리 기간을 크게 단축하였다.

5) 사후조사 실시

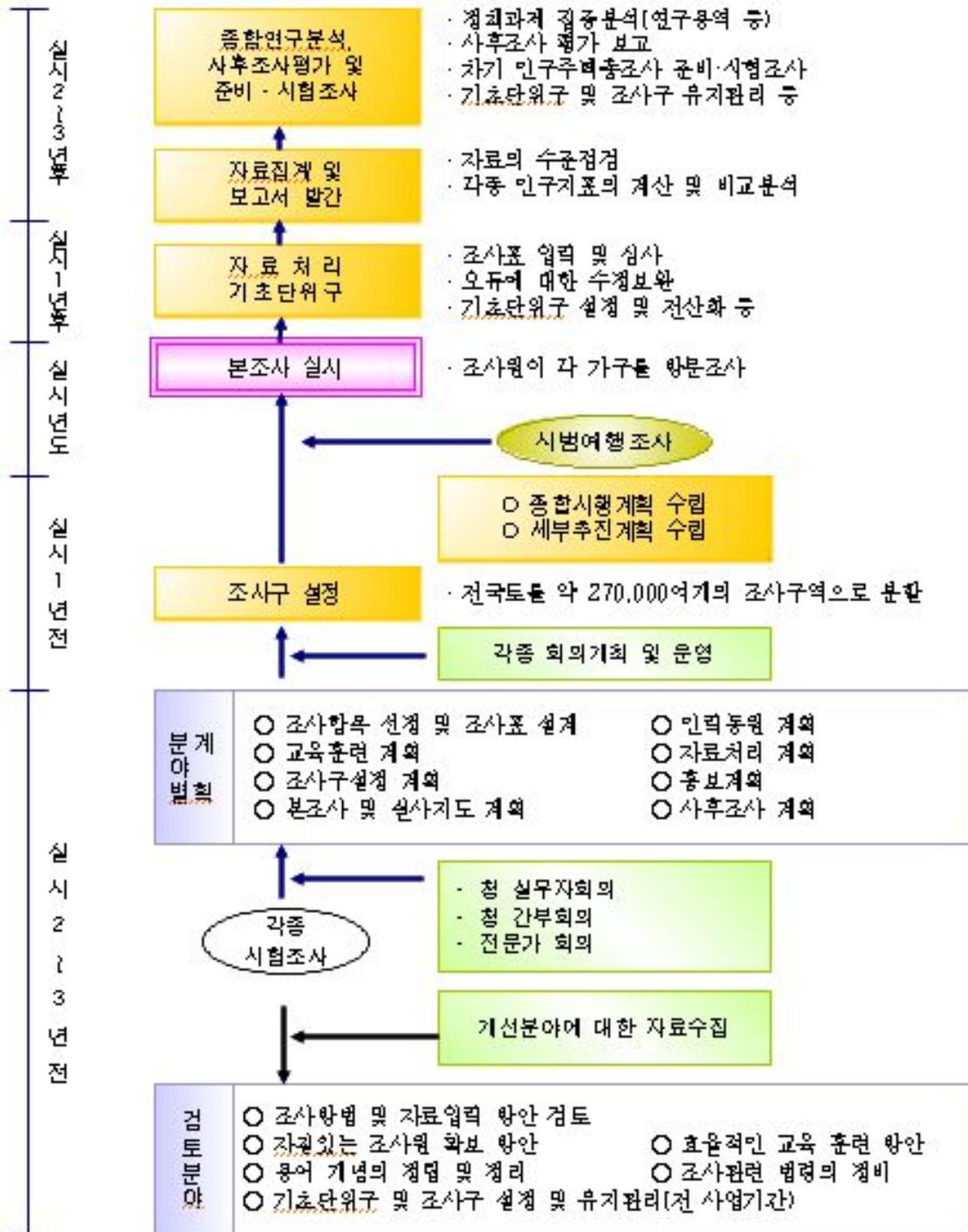
전통적 센서스의 형태로 실시되는 우리나라의 인구주택총조사의 집계오차는 자료처리 과정을 거치면서 수정되는데 수정되지 않은 오차(범주오차와 내용오차)는 총조사 결과의 질에 영향을 주기 때문에 총조사 결과를 여러 가지 방법으로 평가하게 된다. 사후조사 실시, 일치성검사, 인구학적 분석(demographic analysis), 행정자료와의 비교 등 여러 가지 방법으로 평가하는데, 사후조사(事後調査, PES, Post-Enumeration Survey)는 총조사 실시 후에 직접 오차를 측정하기 위해 실시하는 표본조사이다.

사후조사의 방법에는 종속조사방법, 독립조사방법이 있다. 종속조사방법은 총조사 조사표에서 조사된 내용 중 일부를 사후조사 조사원에게 확인하게 하는 한편, 총조사 평가에 필요한 일부사항을 추가로 조사하는 방법이다. 독립조사방법은 표본으로 추출된 조사구에 대해 이미 조사된 조사표의 내용을 조사원에게 전혀 제공하지 않고 독립적으로 별개의 조사를 실시하는 방법이다. 종속조사방법은 본조사 조사표와 사후조사 조사표와의 대조작업이 보다 수월하여 시간과 비용이 절감되는 장점이 있는 반면, 사후조사 조사원의 조사태도에 의해 사후조사의 질이 좌우하는 단점이 있다. 우리나라에서는 종속조사방법을 택하여 약 0.3%의 조사구를 추출하여 사후조사를 실시하고 있다. 향후 영국의 One Number Census와 같이 인구주택총조사의 집계수치를 하나의 최고 수치로 보정하기 위해서는 사후조사규모, 사후조사방법에 대한 연구검토가 이루어져야 한다는 의견이 있다..

6) 결과 집계 및 제공

인구주택총조사 결과는 이용자의 편의를 위해 여러 단계로 나누어 조사결

<그림 5-4> 우리나라 인구주택총조사의 업무 흐름도



과를 공표한다. 먼저 센서스는 자료량이 방대하여 조사결과가 완전하게 집계 되려면 1년이상 장기간의 시간이 소요된다. 하지만 이러한 자료는 시의성이 떨어져 정책입안 자료로 활용하기가 곤란하다. 이에 따라 조사표대신 집계표를 이용해 전국의 전체적인 인구규모만을 먼저 발표하게 되는데 이를 잠정

집계라고 한다. 잠정집계는 1995년 총조사에서는 4개월이 소요되었는데 2000년과 2005년 총조사에서는 1개월로 크게 단축되었다. 다음 단계는 조사표 자료 중 2%만을 따로 뽑아 먼저 집계하여 공표하는데 이를 2% 속보라고 한다. 이 자료는 전체적인 인구의 규모, 구조와 분포의 패턴을 파악하는데 유용하다. 하지만 2000년과 2005년 총조사에서는 전수결과를 보다 빨리 집계하기 위해서 2% 속보는 별도로 집계하여 공표하지 않았다.

조사표를 자료 처리한 결과는 조사항목이 간단한 전수조사결과부터 공표하고 다음은 주제별로 인구이동, 경제활동, 여성·아동 등의 표본조사 항목을 집계하여 최종결과를 발표한다. 2005년 총조사에서는 앞에서도 언급했듯이 Web에 의한 시군구 분산입력방식을 채택하여 전주기 대비 전수집계 결과는 3개월, 표본집계 결과는 7개월이 단축되어 시의성이 크게 제고되었다.

한편, 조사결과는 통계표는 물론 통계 모형 및 심층분석을 위해 2% 원시자료를 제공하고 있으며 2005년의 원시자료는 이용목적에 따라 1%, 5% 등 다양한 형태의 원시자료를 제공하고 있다.

제5절 행정자료 대체이용 가능성 평가 검토

본 절은 통계개발원의 센서스방법론 연구에 수록된 이내성(2008b)의 논문 “행정자료 기반의 등록센서스 작성을 위한 항목선정”⁵⁾의 내용을 검토하는 방식으로 진행된다.

1. 평가기준

이내성(2008b)의 해당 논문은 행정자료는 당초에 통계목적이 아니라 행정 목적으로 작성된 자료이기 때문에, 통계작성을 일차적 목표로 실시되는 방대한 통계조사인 인구주택총조사와는 근본적으로 차이가 있음을 인정하고 있다. 따라서 연구자는 조사항목의 선정기준으로 유엔의 권고안에 의거하여 ① 중앙정부와 지방자치단체의 필요성 ② 국제비교성 ③ 시계열 유지 ④ 항목의 응답 가능성 및 적절성 ⑤ 국가 재원의 가용성 ⑥ 자료이용의 시의성을 설정하고, 조사항목의 중요성(이용우선순위)을 ①, ②, ③의 측면에서 평가하고, 조사항목의 대체이용 가능성은 ④의 측면에서 “개념의 일치성과 포괄범

5) 이 논문은 행정자료를 이용하는 센서스의 실시 가능성을 체계적으로 검토하는 연구의 일부로서 통계개발원의 내부연구로 수행되었으며, 실제로 제5장의 연구는 이것을 구체적으로 보완하는 수준에 머문다고 보면 되겠다.

위”를 파악하여, 완전대체, 불완전대체, 불체불가 등의 3개 그룹으로 구분한다.

해당논문은 불완전대체군을 다시 보완대체군과 표본대체군으로 구분하는데, 전자는 행정자료의 일부 미비점을 보완하는 경우 완전대체가 가능할 것으로 판단되는 대체군이며, 후자는 표본대체는 행정자료의 보완이 어려워 표본수준의 자료로서만 활용가치가 있다고 평가하는 대체군이다.

해당논문은 행정자료의 대체이용 가능성 평가기준을 넓은 의미에서 해석하고 있다. 실제로, 인구주택총조사와 주민등록인구는 조사목적, 조사대상, 조사범위, 조사방법이 전혀 다르기 때문에 좁은 의미에서 본다면, 대체이용가능성은 전혀 없다고 볼 수 있을 것이다(<표 5-6>를 참고할 것). 곧 인구주택총조사의 전수조사 항목인 인구를 주민등록 시스템의 행정자료로 대체하여 등록센서스를 추진하는 것은 현재는 물론 먼 미래에서 불가능하다는 것을 의미한다.

<표 5-6> 인구총조사와 주민등록 행정자료 비교

	2005 인구총조사	주민등록 행정자료
조사목적	<ul style="list-style-type: none"> - 인구총조사는 인구규모, 분포 및 구조에 관한 제특성을 파악, 각종 정책 입안의 기초자료 제공 - 각종 가구관련 경조사 표본틀 (sampling frame)의 기초자료로 활용 	<ul style="list-style-type: none"> - 시군구의 주민을 등록하게 함으로써 주민의 거주관계 등 인구의 동태를 상시로 파악하여 주민생활의 편익을 증진시키고 행정사무의 적정한 처리 도모
법적근거	<ul style="list-style-type: none"> - 기획재정부령 제145호 - 지정 조사통계 제10101호(인구총조사) 및 제10102호(주택총조사) 	<ul style="list-style-type: none"> - 법률 제1067호(1962년 5월 10일)
조사기준일	<ul style="list-style-type: none"> - 11월 1일 0시 현재 	<ul style="list-style-type: none"> - 세대주가 그 신고사유가 발생한 날로부터 14일 이내에 신고
조사대상기간	<ul style="list-style-type: none"> - 11월 1일부터 15일까지(15일간) 	
조사항목	<ul style="list-style-type: none"> - 전수조사 42개, - 표본조사 20개, - 시도특성항목 3개 	<ul style="list-style-type: none"> - 성명, 성별, 생년월일, 세대주와의 관계, 본적, 주소, 주소이력 등
조사대상 및 조사범위	<ul style="list-style-type: none"> - 가구, 가구원 - 조사기준 시점 현재 조사지역내에 상주하는 내·외국인 ※ 조사대상에서 제외되는 인구 <ul style="list-style-type: none"> - 해외취업, 취학중인 자 - 외국외교관, 수행원, 공무로 체류 중인 외국인 및 그 가족, 국제연합소속기관 외국직원, 수행원 및 가족 - 국내 주둔 외국군인, 군속 및 그 가족 	<ul style="list-style-type: none"> - 30일 이상 거주할 목적으로 그 관할구역안에 주소 또는 거소를 가진자 ※ 주민등록 대상외의 자 <ul style="list-style-type: none"> - 외국인 - 해외이주법 제2조의 규정에 의한 해외이주자
조사방법	<ul style="list-style-type: none"> - 조사원에 의한 면접조사 	<ul style="list-style-type: none"> - 주민에 의한 등록의 신고주의 원칙

인구주택총조사와 주민등록자료 모두 우리나라 전체의 성별, 연령별 인구

를 포괄하기는 하지만 가구와 세대, 사실혼과 법률혼 등 개념과 포괄범위 등이 이해의 정도에 따라 온도차가 있으며 포괄성과 온도차가 현장조사에 반영되어 실제로 통계수치로 나타난다는 다양한 문제들 속에서 연구에서는 일부 개념이나 포괄범위의 상충문제는 통계기법 활용 등을 통해 행정자료 기반의 등록센서스 작성에 활용할 수 있다고 보았다.

2015년 등록센서스를 추진을 위하여, 행정자료의 센서스 조사항목 대체이용성을 평가하는데 있어서, 연구자는 물론 통계청의 실무자들은 다음과 같은 주장에 대하여 관심을 기울일 필요가 있다고 판단한다.

“이와 같은 광의(廣義)의 판단은 본 연구가 ‘행정자료 기반의 등록센서스 작성을 위한 항목선정’이라는 적극적 의미에서 행정자료의 미래 완결성을 목표로 접근하였기 때문이며, 이러한 입장의 견해와 달리 좁은 의미로 보는 혹자들과 이견이 있음은 명약관화하다”

<표 5-7> 센서스 조사항목의 중요도 평가

구 분	점수	조사항목의 내역
1차적 (7개)	100	성명, 성별, 연령, 가구주와의 관계, 주소(상주지), 교육정도(졸업여부), 혼인상태(법률혼 관계) <7개>
2차적 (12개)	95	-
	90	혼인상태(사실혼관계), 주거시설(부엌), 총방수
	85	주거시설(화장실), 주택소유
	80	교육정도(취학여부), 점유형태, 주거시설(목욕탕), 연건평
	75	출생지, 거처종류, 건축연도
3차적 (26개)	50~74	교육정도(문맹여부), 종교, 국적, 전거주지(1년전, 5년전), 총출생아수, 생존아수, 경제활동상태, 종사상 지위, 산업, 직업, 가구구분, 사용방수, 주거(전용)시설, 식수(상수도시설), 취사연료, 난방연료, 난방시설, 대지면적, 거주가능가구수, 동거가구수, 집 및 거주의 개수, 지붕재수, 외벽재료, 편의시설, 건물종류
4차적 (56개))	49 이하	본관, 남북이산가족, 수학연수, 전공학과, 통근·통학장소, 통근·통학시간, 이용교통수단, 초혼연령, 사망자녀수, 동거자녀수, 추가계획자녀수, 지난 1년간 출생아 수, 취직여부, 구직(활동)여부, 취업가능성, 비구직이유, 주로 한 일, 조금이라도 한 일, 근로 장소, 주부양자, 거동불편여부, 아동보육, 어머니동거여부, 컴퓨터활용상태, 인터넷활용상태, 개인휴대용통신기기, 단독주택의 종류, 다가구주택, 거주기간, 임차료, 가구심신장애자, 정보통신기기보유, 식수사용형태, 주된 연료, 야궁이형태, 조명시설, 굴뚝상태, 대청마루유무, 식수종류, 가구종류, 자동차보유상태, 주차시설, 주인가구여부, 건물층수, 거주층

2. 조사항목별 중요도 평가결과

<표 5-7>은 이내성(2008)의 논문에서 작성된 것으로, ① 중앙정부와 지방자치단체의 필요성 ② 국제 비교성 ③ 시계열 유지라는 3가지 측면에서 평가한 중요도(이용우선순위)를 보여준다. 평가결과에 의하면, 성명, 성별, 연령, 가구주와의 관계, 주소(상주지), 교육정도(최종졸업), 혼인상태(법률혼 관계) 등 7개가 3가지 측면 모두에서 중요도 100점으로 나타나는 1차적 중요성을 가지는 조사항목이다.

중요도 2차적 항목은 3가지 차원의 점수가 75점에서 100점 미만인 조사항목으로 출생지, 혼인상태(사실혼 관계), 취업여부와 거처의 종류, 총방수, 주거시설(화장실), 주거시설(목욕탕), 점유(소유)관계, 점유형태, 건축년도, 연건평 등 12개로 나타났다. 중요도 3차적 항목은 3가지 차원의 점수가 50점에서 75점 미만인 조사항목으로 교육정도(문맹여부), 종교, 국적 등 26개, 4차적 항목은 3가지 차원의 점수가 50점 미만인 조사항목으로 본관, 교육정도(수학연수) 등 55개로 나타났다.

조사항목을 3가지 차원으로 평가할 때, 2010년 인구주택총조사 항목에 신규로 포함된 국적은 중요도가 3차적인 항목으로 선정되었다. 또 입국연월은 평가대상에 포함되지 않았다. 이러한 것은 우리나라 정부가 최근에 저출산·고령사회의 결과에 대한 정책방향으로 다문화사회를 새로운 정책기조로 채택하면서, 인구주택총조사에서 외국인 노동자와 결혼이민 여성의 현황을 파악하기 위하여 채택한 조사항목이다. 이 점은 새로운 방법으로 그 중요성을 파악하는 방법이 마련되어야 할 것이다.

또 중요도가 4차적인 조사항목인 본관과 남북이산가족은 국제비교성이 없어 매우 낮게 나타났다. 본관의 경우 가족등록제도의 시행(2008. 1. 1.)으로 부계주의 원칙이 수정됨에 따라 협의시 모의 성과 본관을 사용할 수 있게 되었다. 이와 같이 기존의 호주제와 같은 가족제도에 사회적 요구가 크게 반영되어 변경되었으나, 아직까지 성씨, 본관별 인구 현황, 집성촌 실태 및 분포, 족보 연구 등과 아울러 신규 성씨와 본관의 발생 현황을 파악하는데 필요한 인구자료가 될 수도 있을 것이다.

남북이산가족은 2005년 노무현 대통령의 참여정부 아래서 신규로 포함된 조사항목으로, 세계에서 유일한 분단국가인 우리나라로서 이산가족의 규모와 경제사회학적 특성에 관한 정보 파악에 필요한 항목으로 평가되고 있다. 이 항목은 적절한 방법으로 조사객체의 응답을 획득하는 경우에는 출생지 조사항목과 더불어 출생지별 인구규모와 인구의 생애이동(lifetime migration)을

파악할 수 있는 자료이다.

3. 대체가능정도별 조사항목의 분류

대체이용가능정도의 평가결과는 국제기준 중에서 ④ 항목의 “응답가능성 및 적절성” 측면 즉, ‘개념의 일치성과 포괄범위’를 가지고 평가하였다.

전수조사 항목으로 100% 대체가능한 조사항목으로는 선정된 것이 성명, 성별, 연령, 출생지, 혼인상태(법률혼 관계), 국적, 본관, 거처종류, 건물종류, 건축년도, 대지면적, 연건평, 주택소유 등 16개 항목이었으며, 대체가능도가 50점에서 100점 미만인 이른바 “불완전대체” 조사항목은 교육정도, 난방시설, 남북이산가족, 주거시설(목욕탕), 주거시설(부엌), 점유형태, 주거용 연면적, 총방수, 편의시설, 주거시설(화장실), 취학여부, 혼인상태(사실혼 관계) 등 14개 항목으로 나타났다. 그리고 사용방수와 종교는 평가점수가 10점 미만에 이르고 있는 것에서 볼 수 있는 것처럼, 관련 행정자료가 당장에 가용하지

<표 5-8> 센서스 조사항목의 행정자료 대체정도별 상태

구 분	전수조사항목	표본조사 항목
100 <31개>	가구구분, 가구주와의 관계, 국적, 거처종류- 단독주택, 다가구주택, 건물종류- 주택용도, 건축년도, 대지면적, 본관, 상주지, 성명성별, 연건평, 연령, 주택소유, 출생지, 혼인상태(법률적 관계) <16개>	1년(5년)전거주지, 거주기간, 건물층수, 동거가구수, 동거자녀수, 별거자녀수, 사망자녀수, 생존아수, 전거주지, 지난1년간 출생아수, 초혼연령, 활동계약-거동불편여부, 가구내심신장애자, 총출생아수, 자녀거주장소, 어머니동거여부 <15개>
50~99 <38개>	교육정도(졸업여부), 난방시설, 남북이산가족, 목욕시설, 부엌시설, 점유형태, 주거용연면적, 주인가구, 총방수, 편의시설, 화장실시설, 혼인상태(사실적관계), 취학여부, 주거(전용)시설 <14개>	가구소득, 개인소득, 거주층, 경제활동상태, 근로장소, 식수종류, 식수(상수도)시설, 아동보육, 산업, 자동차보유대수, 종사상지위, 주로 한 일, 직업, 취업여부, 현직업근무년수, 취업형태, 취업기간, 주된연료-취사연료, 난방연료, 아궁이형태, 조명시설, 지붕재료, 외벽재료 <24개>
10~49 <3개>		생활비원천-생계수단, 점유형태, 주차시설 <3개>
10미만 <20개>	사용방수, 종교 <2개>	구직활동여부, 수확연수, 문맹여부, 주부양자, 추가계획자녀수, 추가취업희망, 개인휴대용통신기기, 인터넷활용상태, 컴퓨터활용상태, 대청마루, 거주가능가구수, 건물수선상태, 임차료, 굴뚝상태, 이동교통수단, 전공학과, 주차시설, 통근통학여부, 통근(학)지, 통근(학)시간 <18개>

않아서 대체할 수 없는 항목으로 평가하였다.

인구주택총조사의 표본조사 항목 중 100%에 가까이 완전하게 대체가능한 조사항목은 전거주지(5년 전, 1년 전), 거주기간, 건물층수, 동거가구수, 동거 자녀수, 별거자녀수, 사망자녀수, 생존아수, 출생아수, 초혼연령, 활동제약 등 15개 항목이며, 이들은 현재의 행정자료를 통계대장(statistical register)으로 전환하여 마이크로 데이터베이스를 구축함으로써 전수화할 수 있는 것으로 나타났다.

인구주택총조사의 표본조사항목을 총조사와 조사시점이 다른 표본조사의 자료를 가지고 대체하는 것은 아무런 의미도 없다. 행정자료를 토대로 하는 통계는 등록기반조사(register-based survey)를 이용하여 생산된 통계를 의미하며, 표본조사자료는 행정자료를 토대로 작성된 통계대장(statistical register)과 매칭하여, 등록센서스를 지원하는 이른바 “통합센서스”(integrated census)를 실시하는 과정에서 활용될 수 있다.

불완전 대체항목은 가구소득, 개인소득, 거주층, 경제활동상태, 근로장소, 아동보육, 식수시설 및 사용형태, 직업, 생활비원천, 주차시설 등 27개 항목으로 나타났다. 또한 관련 행정자료의 부족으로 대체하여 이용할 수 없는 항목은 구직활동여부, 추가계획자녀수, 추가취업희망, 개인휴대용통신기기보유, 인터넷 이용상태, 컴퓨터 활용상태 등 14개 항목으로 나타났다. 이들 조사항목에 대해서는 등록센서스에서 행정자료, 특히 주민등록대장과 건축물/주거대장의 완전성을 파악하기 위한 목적뿐만 아니라, 행정자료 대체가능성이 불완전하다고 판단하는 센서스 조사항목에 대하여 표본조사를 실시할 필요가 있을 것이다.

4. 센서스 조사항목의 중요도별 대체가능 정도의 분류

중요도별 대체이용가능정도별로 분석한 결과를 토대로 중요도와 대체이용가능정도가 50점 이상인 항목을 살펴보면 다음과 같다.(<표 5-8> 참조).

<인구 및 가구부문>에서는 가구주와의 관계, 상주지, 성명, 성별, 연령, 혼인상태(법률혼 관계), 출생지, 국적이 완전대체가 가능하며 이용 가능한 항목으로 나타난 반면, 교육정도와 ECE/Eurostat가 비핵심 항목으로 분류하고 있는 혼인상태(사실혼 관계)는 불완전대체이용가능 항목으로 나타났다.

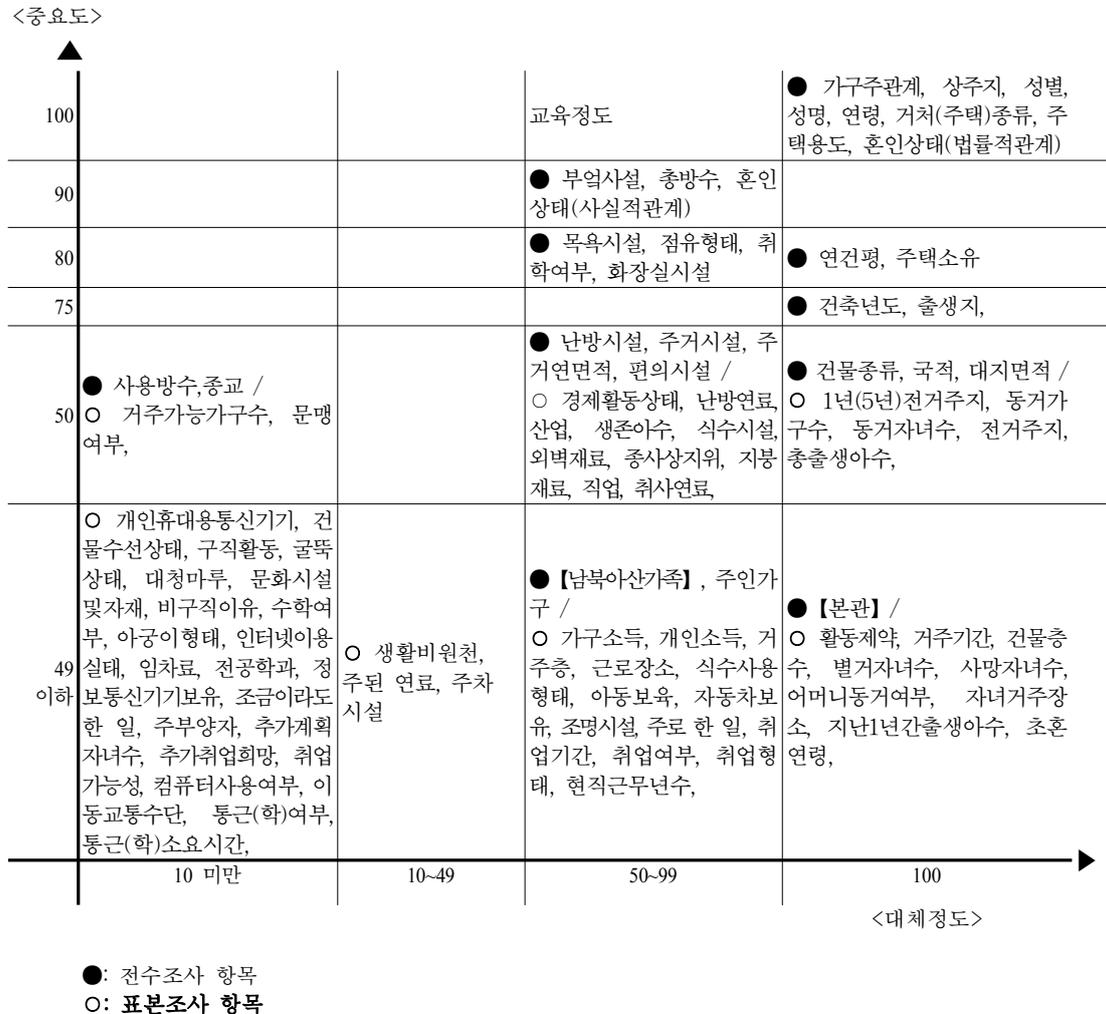
특히 국제비교성 때문에 중요도가 낮게 나타난 본관과 남북이산가족 중 본관은 완전대체가능 항목으로, 남북이산가족은 불완전대체이용가능 항목으로 나타났다.

<주택부문>에서는 거처의 종류, 주택용도, 주택의 소유여부, 건축년도, 건물의 종류, 연건평, 대지면적 등이 완전대체이용가능 항목으로 나타난 반면, 점유형태, 총방수, 부엌시설, 목욕시설, 화장실시설, 난방시설, 주거시설 등 편의시설과 주거연면적 등은 불완전대체이용가능 항목으로 나타났다

주요 이용가능 행정자료별로 살펴보면 다음과 같다.

<인구·가구부문> 중 주소(상주지), 성명, 성별, 연령은 주민등록 행정자료, 혼인상태·본적은 가족등록 행정자료, 교육정도는 교육행정정보시스템(NEIS)과 출생·사망·혼인·이혼신고서, 남북이산가족은 이산가족정보통합센터 등록 자료를 각각 활용할 수 있다.

<표 5-9> 센서스 조사항목의 중요도별 대체가능정도의 현황



<주택부문>은 건축물 관련 행정자료 중 건축물대장을 활용할 수 있다. 특

히 상기 언급한 조사항목 외에 많은 조사항목이 주민등록 행정자료와 건축물대장과 관련성이 매우 크기 때문에 반드시 이들 행정자료를 기본등록대장으로 하여야 한다.

5. 2015년 등록센서스 실시를 위한 관건

본 연구, 특히 제1장에서 제3장에 걸쳐서 작성된 등록센서스의 수리적·통계적 방법론에서는 등록기반의 센서스를 시작하는 단계에서 주민등록부와 주거부문의 거주지DB를 활용하고, 추가적인 행정자료인 가족관계등록부의 활용은 충분히 검토하지 못하였다. 그러나 시작단계에서 가장 중요한 것은 앞서 언급한 것처럼 기본등록부(주민등록부와 가족등록부의 통합과 거주지DB의 연계)의 확립과 이들 등록부의 품질을 높이는 것이다. 기본등록부가 확립된 이후 추가할 수 있는 행정자료는 대부분 인구의 특성에 관한 행정자료들이다. 이들 등록부는 주민등록번호라는 고유의 식별코드로 매칭이 원칙적으로 가능하지만, 본 연구에는 이 식별코드를 매칭작업에 직접적으로 활용하지 못하였다.

<표 5-10> 등록센서스에서 대체 가능 항목 수

구분	계	인구	가구	주택
전체 항목	18(44)	7(24)	6(14)	5(6)
행정자료 활용	11(17)	5(9)	1(2)	5(6)
	대체 가능 항목	성명 성별 나이 가구주와의관 계 혼인상태 (출생지) (1년전거주지) (5년전거주지) (혼인년월)	가구구분 (거주기간)	거처의종류 주거용연면적 총방수 건축년도 편의시설수 (대지면적)

주: 항목은 2008년 제2차 시험조사 기준이며, 괄호 안은 전수와 표본 항목의 합계임.

본 연구는 현재 고려되고 있는 기본등록부(주민등록, 가족관계등록부)와 건축물대장을 활용하는 경우 <표 5-10>에 제시된 항목들을 등록기반 센서스에서 대체할 수 있을 것이다. 이 표에서 건축물대장 대신 거주지DB를 구축하

여 활용하는 경우 주택부문의 모든 항목을 대체할 수 있음을 알 수 있다. 그러나 등록센서스를 실시중인 국가에서처럼 단위거처의 구분이 가능한 주소체계를 갖지 못한다면 가구부문의 대체율이 낮을 수밖에 없으며, 장기적으로 표본조사가 없는 완전한 등록센서스로의 이행을 위해서는 단위거처(dwelling)의 구분이 가능한 식별코드(주소체계)의 도입이 요구된다.

등록센서스 실시를 위한 바탕은 기본대장, 즉 주민등록부(가족관계등록부 포함)와 건축물대장의 품질 확보가 관건이다. 하지만 이를 위한 기본대장의 품질 평가와 개선 노력은 통계청뿐만 아니라 대법원, 행정안전부, 국토해양부 등의 관련부처와 공동으로 대처해야 할 사항이다. 다른 국가들에서도 기본등록부의 품질 향상을 위한 관련 부처와의 공동 노력이 장기간 지속되어 왔는데, 독일의 경우 2011년 등록기반의 센서스를 실시하기 위하여 1997년에 연방행정청들의 대표와 연방통계청, 각 주의 통계청으로 구성된 '유럽연합 센서스 2001' 태스크포스를 만들었다. 동시에 지자체의 인구등록부의 품질향상을 위한 태스크포스를 구성하여 등록부의 품질 개선을 위한 실행에 착수하였다.

우리나라 주민등록과 건축물대장의 경우 행정안전부와 16개시도, 국토해양부, 선거관리위원회 등 관련 부처와 통합된 TF를 구성하여 연차적으로 자료의 수준 개선을 위한 노력이 필요하다. 이와 같은 기본등록부의 품질 향상을 위한 노력은 단기간에 끝날 수 없지만, 2015년 등록센서스를 성공적으로 추진하기 위해서는 지체되어서는 안 되는 당면과제이다..

또한 주민등록부의 품질 향상을 위하여 주민 등록지를 실제 거주지와 달리 하는 위장전입을 막기 위한 대책도 동시에 필요하다. 여기에는 주민등록법에서의 사실조사와 직권조치와 같은 법적인 수단도 필요하지만, 위장전입이나 인위적인 세대분가 등의 유인을 막을 수 있는 제도적인 대책도 동시에 강구되어야 할 것이다. 왜냐하면 등록지와 실제 거주지가 다른 경우의 대부분은 부동산 매매와 주택분양, 자녀의 학군배정, 부양가족 혜택 등 사회·경제적 이익과 관련되기 때문이다.

다음으로 주민등록부의 질적 수준을 높이기 위한 노력의 일환으로 주민등록의 오류가 높을 것으로 예상되는 기숙사나 특수사회시설과 같은 집단시설 조사구에 대한 별도의 조사를 실시하여 등록내용에 대한 오차를 줄일 수 있을 것이다. 즉, 총조사 인구나 주민등록 인구를 비교할 때 전국적으로는 총조사 인구가 주민등록인구의 약 96% 수준이지만, 지역별로 보면 주로 대학

6) 여기에 국적, 본관, 총출생아수를 추가하면, 2015년 등록센서스에서 대체가능한 항목이 20개로 확대될 수 있다. 통계청의 등록센서스 기본계획은 대체가능 항목을 20개 정도로 파악하고 있다.(통계청, 2008a)

교 인근 지역에서 총조사 인구가 더 많은 것으로 집계되고 있다.⁷⁾ 해당 지역의 지자체에서도 인구수 문제에 관심을 갖고 있기 때문에, 지역학교와 협조하여 거주지 주민등록을 갖도록 계도하는 것은 등록부의 질적 개선에 도움이 될 것이다.⁸⁾ 또한 가구 측면에서 보면, 기숙사 등 집단시설은 총조사에서는 일반적으로 1개 기숙사가 1개의 집단가구로 집계되지만 주민등록상 세대에서는 1개 기숙사에 등록된 세대가 각각의 세대로 집계되어 총조사 가구와 주민등록상의 세대와 차이를 크게 하고 있다. 이에 대한 개념 통일 또는 집계기준의 재설정으로 가구수와 세대수의 차이를 일부분 좁혀 나갈 수 있을 것이다.

아울러 주민등록부의 정확성을 평가하기 위한 조사를 매년 실시하여 지자체 스스로 등록 오류율을 낮추도록 유인책을 마련하고, 전체적인 등록 오류율의 수준을 검토하여 등록센서스로의 전환이 가능한 시기를 확정하여야 할 것이다.⁹⁾

또한 우리나라에서 등록센서스를 도입하더라도 등록센서스에서 대체이용이 불가능한 항목에 대한 정보를 얻기 위하여 표본조사는 여전히 실시하지 않으면 안 되는데, 이 경우 표본조사는 전수 항목에서 얻을 수 없는 항목에 대한 자료 수집과 등록부의 완전성 평가라는 이중적 역할을 정확하게 수행하지 않으면 안 될 것이다.

주민등록의 품질 향상은 인구 뿐 만 아니라 가구 생성 측면에서도 반드시 필요하다. 우리나라는 총조사에서 가사단위의 가구개념을 사용하여 왔다. 그러나 등록센서스를 실시 중인 국가의 대부분은 거주단위의 가구 개념을 적용하여 센서스 결과를 생산해 내고 있다. 이들 국가는 단위거처까지 식별 가능한 체계적인 주소코드를 바탕으로 거주단위의 가구개념을 적용하고 있는데 반해 우리나라에서는 현실적으로 적용이 불가능하다. 따라서 우리나라에서 현장조사 없이 가구자료를 얻기 위해서는 주민등록부의 세대를 이용하는 것이 가장 현실적인데, 이를 위하여 주민등록부에서 인구와 세대부문의 품질 제고를 위한 많은 노력이 요구된다.

건축물대장의 경우는 대장 자체가 현실에 부합하지 않는 문제도 있지만, 건축물대장에서 매칭코드로 사용될 수 있는 주소가 법정동 주소로 되어 있고, 대장에서 이용할 수 있는 항목수도 제한적이다.

7) 대학 재학 중인 학생들은 기숙사나 인근 지역에 거주하지만, 주민등록을 실제 거주지로 옮겨 둔 비율은 낮을 것으로 예상된다.

8) 일부 지자체의 경우, 인구증가에 따른 조직 확대를 위하여 '내고장 주민등록 갖기'를 무리하게 추진하는 경우가 있어 이는 지양되어야 하지만, 주민등록을 현실에 맞게 등록하도록 하는 것은 등록부의 품질 향상을 위하여 반드시 필요하다.

9) 2005년 센서스에서의 총오차율은 3.88%, 순누락율(누락율-중복율)은 0.90%로 나타났다.

이와 같은 상황을 극복하기 위하여 덴마크와 핀란드의 사례에서와 같이 센서스 실시 결과를 바탕으로 하여, 주택부문의 등록부를 설립하는 방안을 검토해 볼 수 있다. 즉 2010년 인구주택총조사에서는 건축물대장 등 관련 등록부를 활용하여 전체 거처에 대한 전수조사가 실시 될 것이다. 여기에서 얻어진 자료를 바탕으로 우리나라 전체 거처에 대한 거주지DB¹⁰⁾가 구축된다면 상당한 장점이 있을 것이다. 현재까지의 인구주택총조사에서는 주소를 입력하지 않고 매 조사시마다 조사구를 새로 설정하여 과거자료를 활용할 수 없었지만, 2010년 센서스에서 모든 거처에 영구적으로 식별가능한 주소와 거처 번호를 부여하고, 주민등록부와 매칭할 수 있는 확인된 주소로 DB를 구축¹¹⁾한다면 건축물대장의 대체 등 자료의 활용도 측면에서 기여도가 커질 것이다. 그리고 그 이후의 변동사항은 국토해양부의 신축 및 멸실 건축물대장 자료를 통해 계속적으로 거주지DB가 유지·관리된다면, 이를 통해 기존의 주택DB나 건축물대장보다 더 나은 주거관련 자료를 얻을 수 있을 것이다. 아울러 기존의 조사구 설정시와 마찬가지로 수도권외의 신도시 건설사업이 동시다발적으로 전개되어 주택 신축과 합몰 등의 변동 폭이 큰 지역은 현지 실사를 통하여 거주지DB가 완전성의 면에서 정확하게 유지·관리되어야 할 것이다.

이처럼 2010년 센서스에서 주거부문의 조사결과를 바탕으로 거주지DB가 구성되어, 유지·관리가 된다면 여러 가지 장점이 있을 수 있다. 첫째, 주택부문에서는 표본조사 없이 대부분의 항목을 등록기반으로 센서스 결과의 집계 가능하다. 둘째, 모든 거처가 고유한 식별코드¹²⁾를 갖게 되어 인구부문의 주민등록 자료와 매칭이 훨씬 수월해 질 수 있다. 셋째, 등록센서스로 전환에도 불구하고 기존과 동일하게 각종 표본조사에서 표본 추출을 위한 모집단의 역할을 할 수 있다. 넷째, 건축물대장의 품질 및 정비수준에 따라 통계자료의 질이 달라질 수 있는데, 거주지DB가 구축되면 통계청의 내부자료를 이용하기 때문에, 다른 행정기관에 대한 의존도를 낮출 수 있는 장점이 있다. 마지막으로 2010년 센서스에서 추가로 거처별 거주가능 가구수를 파악하여 거주지DB에서 관리된다면, 정확한 주택보급률 산정에 기여 할 수 있을

10) 거주지DB란 기존의 주택DB와 구분하기 위하여 편의상 이름을 붙였다. 기존의 주택DB는 총조사에서 주택부문의 조사에 참고하기 위하여 구축된 것이고, 거주지DB는 2010 총조사 결과를 포괄하여 향후 주택부문의 자료가 계속적으로 업데이트 될 DB를 말한다.

11) 이러한 거주지DB를 구축하기 위해서는 2010 인구주택총조사에서 식별번호와 주소에 대한 정확한 확인이 필요하여 업무 부담이 증가하는 문제점이 있다. 물론 총조사 실시과정에서 기존의 주택DB는 동일하게 활용하여 조사해야 한다.

12) 주민등록부의 주소와 일치여부가 확인되어 인구부문과 주거부문의 매칭이 가능한 주소를 말한다.

뿐만 아니라 주민등록의 정비에도 유용하게 사용될 수 있을 것이다. 즉 특정 주택에 거주가능 가구수를 크게 초과하는 세대가 존재하는 경우 주민등록 정비에 이용 가능하여 주민등록의 품질향상에도 기여할 수 있을 것이다.

제6절 종합 및 정책적 건의사항

이 장에서는 등록센서스에서 대체이용이 불가능한 조사항목의 표본조사 도입방안을 논의하기 위하여, 센서스의 다양한 유형, 등록센서스에서 표본조사의 역할, 우리나라 현행 인구주택총조사의 조사방법론, 행정자료 대체이용 가능성 평가 검토, 우리나라 센서스 표본조사 조사방법론을 종합적으로 검토하였다.

우리나라가 2015년에 등록센서스를 추진하면서, 그것이 북유럽 특히 덴마크나 노르웨이 핀란드에서 실시되었거나 실시 예정으로 되어있는 100% 완전한 등록센서스가 아니라 다른 행정자료와 표본조사의 통합센서스, 바로 부분등록센서스(partially register-based census)라고 하는 것을 명심할 필요가 있다. 물론, 이러한 등록센서스와 표본조사를 동시에 실시하는 통합센서스는 이스라엘, 싱가포르, 네덜란드, 독일, 대만 등의 일부 국가에서 실시되었거나 실시 예정으로 되어 있다.

100% 행정자료만을 이용하여 센서스 통계를 작성하는 등록센서스에서 등록기반 조사(register-based survey)는 응답자의 부담경감, 저비용이라는 엄청난 장점이 있지만, 조사항목을 추가할 수 없고, 행정시스템의 모집단, 조사객체, 변수정의에 의존한다는 문제점이 있다. 북유럽은 오랜 경험을 통하여, 완전 등록센서스 통계를 작성하기 위하여 다양한 행정자료, 가령 인구분야만이 아니라 교육, 경제활동 등 주제 분야별로 센서스 통계를 작성하기 위하여 행정등록부를 체계적으로 이용하여 왔다. 그러나 우리나라의 경우 센서스 통계를 작성하기 위하여 상대적으로 만족스럽게 이용할 수 있는 행정자료는 주민등록, 가족관계등록부, 건축물대장이며, 그 외의 행정자료로는 현재 등록센서스의 주제별 통계를 작성하는데 이용되기는 힘들다.

우리나라의 2005년 센서스는 전수조사 항목이 인구분야 8개, 가구·주택분야 13개, 표본조사 항목이 인구 16개 가구 4개, 시도 특성항목 3개로 이루어졌다. 2010년 센서스는 전수조사 항목이 인구분야 8개, 가구·주택분야 11개, 표본조사 항목이 인구 20개 가구·주택분야 7개, 시도 특성항목 3개로 이루어졌다. 신규 조사항목에는 전수조사 항목에 국적, 입국연월, 사회활동, 교통수단 보유 및 이용 현황이 있다.

본 장에서 언급한 것처럼, 등록센서스에서 대체가능한 항목수는 17개 정도로서, 그들은 전수조사 항목의 경우 11개, 인구분야 5개(성명, 성별, 나이, 가구주와의 관계, 혼인상태), 가구분야 1개 (가구구분), 주택분야 5개(거처의 종류, 총방수, 건축연도 편익시설수)이고, 표본조사 항목의 경우, 인구분야 4개(출생지, 1년전 거주지, 5년전 거주지, 혼인연월), 가구분야 1개(거주기간), 주택분야 1개(대지면적)가 된다. 여기에 통계청의 등록센서스 기본계획에 들어 있는 국적, 본관, 총출생아수를 합하게 되면, 2015년 등록센서스에서 대체가능한 항목수는 20개 정도로 확대된다(통계청, 2008a).

우리나라가 2015년 등록센서스를 실시하여, 주민등록대장과 건축물/주택대장을 이용하여, 센서스 조사항목에서 위에서 언급한 조사항목 중에서 완전대체가능한 조사항목을 센서스 통계표로 작성할 수 있을 것이다. 그러나 현 단계로서는 두 가지의 문제를 해결해야 하는 것이다. 하나는 무엇보다도 주민등록대장과 건축물/주택대장의 집계범위를 점검하는 사후조사의 기능을 가진 표본조사를 실시하는 것이다. 다른 하나는 완전대체가 불가능한 조사항목을 표본조사로 조사하여, 그것을 바탕으로 필요한 센서스 통계를 확보하는 것이 될 것이다. 이러한 이중적 목적을 동시에 수행하기 위하여, 우리나라 2015년 등록센서스에서 중요한 기능을 수행하게 될 센서스 표본조사의 조사방법론과 관련된 몇가지 정책적 건의사항을 정리하여 보기로 한다..

1. 표본설계(디자인과 표본선정)

우리나라 센서스 표본조사의 도입방안에서 표본설계를 어떻게 할 것인가를 결정하는 것이 대단히 중요하다. 흔히들 등록센서스를 실시한다면, 등록센서스의 기본대장인 주민등록대장과 건축물/주택대장을 토대로 하여 표본설계를 하는 것이 가장 바람직하다고 할 것이다. 그러나 앞에서 지적한 것처럼, 우리나라의 센서스 표본조사는 주민등록대장이나 건축물/주택대장의 집계범위에 대한 완전성을 점검하는 기능을 가지고 있기 때문에, 표본설계 자체가 한층 더 복잡하게 되고, 이로 말미암아 그 비용도 통상적 수준보다 커질 수 있다는 점을 고려하지 않으면 안 될 것이다.

우리나라가 2015년 등록센서스를 추진하면서, 기본 통계대장에 해당하는 주민등록대장과 건축물/주택대장이 완전하지 않으며, 다양한 오류가 포함되어 있다는 것은 주지의 사실이고, 본 연구의 과정에서 증명되어질 것이다. 이 경우, 센서스 결과의 정확성을 향상시키기 위하여, 현행 인구주택총조사의 사후조사보다 훨씬 더 정밀성이나 정확성이 큰 사후조사를 센서스 표본

조사의 형태로 실시해야 할 필요가 있을 것이다. 센서스 표본조사를 독자적으로 실시하여 행정자료의 정확성을 점검하여, 인구주택총조사의 조사객체인 인구, 가구, 주택에 대해서 행정자료에 포함된 조사항목의 품질을 점검하여야 할 필요성이 크다. 일차적으로, 이러한 사후조사의 기능을 갖는 센서스 표본조사는 합리적이고 논리적으로 타당성이 있는 확률표본추출 방법을 이용하여야 할 것이다.

사후조사(PES, Post-Enumeration Survey)라는 일차적 기능을 충족하는 센서스 표본조사의 표본설계는 행정자료의 품질을 평가하기 위한 측정목적에 적절히 고려해야 한다. 이들은 행정자료의 과소집계나 과다집계를 신뢰도가 높은 수준에서 추정할 필요성을 충족시켜야 한다. 특히, 집계범위에 대한 추정치가 시군구 등의 기초자치단체, 16개 광역자치단체, 국가 전체, 도농간 비교 등 지리적 구분에 대해서도 필요할 수 있을 것이다. 이러한 요건은 센서스 표본조사의 표본설계에 중대한 영향을 주게 되며, 지역별 집계범위의 추정치가 소지역이나 소집단으로 확대되면, 표본규모는 상당히 확대되지 않으면 안 될 것이다. 일반적으로 사후조사 기능을 갖는 센서스 표본조사는 다음 사항을 유념하여 설계되어야 한다.

- ① 행정자료의 최신 업데이트 시점과 센서스 표본조사 실시시점의 시차는 축소되어야 하며, 이를 통하여 행정자료 업데이트를 소홀히 한데 따른 오류문제를 최소화하여야 한다.
- ② 센서스 표본조사는 행정자료, 특히 주민등록대장과 건축물/주택대장의 명부프레임(register frame)을 이용하지 않고 수행되어야 한다. 센서스 표본조사의 현장조사원이 주민등록대장이나 건축물/주택대장의 명부프레임이나 대장에 기재된 조사항목의 응답내용을 알고 있는 경우에는, 그 프레임이나 응답내용을 확인하는 절차 이상의 것을 수행하지 않기 때문에, 이에 대하여 각별한 주의가 요청된다.
- ③ 센서스 표본조사의 독립성을 유지하기 위하여, 표본조사의 자료수집과 자료처리는 통계등록대장의 등록기반조사와는 완전 별개의 것으로 운영되어야 한다.
- ④ 이원시스템추정법(DSE, Dual System Estimator)을 이용하여 행정자료의 집계범위를 점검하여야 한다. 다만 여기서 센서스 표본조사의 사후조사적 기능에 대한 가정은 주민등록대장이나 건축물/주택대장의 등록자료보다 센서스 표본조사의 자료가 고품질의 완전한 자료일 필요는 없으며, 해당 자료가 제2의 독립적인 표본조사에서 획득된 것이라는 가정이 필요하다.

- ⑤ 센서스 표본조사에서 사후조사의 기능을 충족시키기 위하여 면접대상이 된 가구원들과 그들의 거처를 주민등록대장과 건축물/주택대장과 1: 1로 매칭작업을 진행하여 그들이 주민등록대장과 건축물/주택대장에서 정확하게 집계되었는가를 점검하여야 한다.
- ⑥ 센서스 표본조사는 센서스 기준 시점 곧, 센서스의 본질적 특성에서 동시성을 충족시키기 위하여 설정한 기준 일자와 표본조사의 실제 실시일자 간에 이동한 인구에 대한 규칙을 설정하여야 한다. 가령, 센서스 기준 시점에서, 독립적 표본은 표본지역에 주소를 둔 사람들로서, 그들은 전출자를 포함하는 사람들이 될 수 있을 것이다.

2015년 등록센서스에서 표본조사의 두 번째 기능은 행정자료가 포괄하지 못하는 조사항목, 곧 앞에서 설명한 대체 불가능한 조사항목, 곧 완전대체항목을 제외한 모든 항목에 대하여 변수정보를 획득하는 것이다. 그러나 실제로, 주민등록과 건축물대장에 들어있어 대체이용 가능하다고 판단되는 기본적인 변수를 다시 한 번 표본조사 항목에 포함하여, 그들의 정확성을 점검하는 것도 센서스의 표본조사가 수행해야 할 기능에 포함된다는 점을 염두에 두지 않으면 안 된다.

인구주택총조사에서 인구와 단위거처의 표본에 대하여 획득하는 센서스 정보는 두 가지 방법 중에 하나로 진행된다. 하나는 센서스 표본조사를 실시할 가구를 모집단 중에서 일부를 미리 지정하는 것이다. 또, 비용이나 정밀도 기준을 염두에 둔 표본요건을 따라, 표본조사의 대상자는 전체 모집단의 25%, 20%, 10%의 비율을 구성하게 된다. 이러한 표본선정 기준에 의거하여, 모집단의 나머지 가구는 행정자료의 변수정보만을 이용하여 센서스 통계를 작성하는데 사용되며, 표본조사의 결과에 따라 센서스 모집단의 통계정보는 통계학의 각종 이론을 근거로 다양한 추정과 보정의 대상이 될 수 있다. 통계청은 이러한 표본선정 방식을 선택하게 되는 경우, 표본조사의 표집대상 가구를 사전에 선정하여 질문지를 배포하는 방법을 고안하는 것이 될 수 있겠지만, 행정자료가 완전하지 않을 뿐만 아니라, 센서스 표본조사를 주민등록의 완전성을 점검하는데 사용해야 하기 때문에 우리나라의 경우에 이 방법을 사용하는 것은 무리라고 판단한다.

또 하나의 표집방법은 센서스 표본조사표에 응답할 조사구(EA, Enumeration Area)의 표본을 사전에 지정하는 것이라고 할 수 있다. 이 방법을 사용하게 되면, 지정된 조사구의 모든 가구는 표본조사표를 받게 되고, 나머지 지역은 주민등록대장과 건축물/주택대장의 기재내용을 변수정보로

하여, 전수항목에 대한 조사를 대신하게 된다. 흔히들, 둘째 방법에 대하여 첫 번째 방법의 장점은 전체의 조사구가 표집단위로 사용되는 경우 집락효과가 표본분산을 증대시킬 수 있기 때문에, 첫 번째 방법이 두 번째 방법보다 조사결과의 표집오차를 줄이는데 기여한다는 것이다. 그러나 현재로서 우리나라의 센서스 표본조사는 표본조사표의 배포를 조사구별로 배포하는 것이 좋을 것이라고 판단하는데, 그 이유는 무엇보다도 센서스 표본조사가 사회경제적 정보뿐만 아니라, 주민등록대장과 건축물/주택대장의 정확성을 점검하는 수단이 되어야 하기 때문이다.

2. 표본추출방법

센서스 표본조사는 주민등록대장과 건축물/거처대장의 주소명부를 사용하는 것은 인구주택총조사의 등록기반 조사(register-based survey)의 범위 정확성을 점검하는데 필요한데 두 조사시스템의 추출틀(sampling frame)이 상호 독립적이어야 한다는 원칙을 위반하는 것이다. 그렇다면, 현재의 인구주택총조사에서 사용하는 조사구 방식을 이용할 수도 있을 것이고, 행정시스템의 시군구에 있는 통반 시스템을 이용할 수도 있을 것이다. 또는 지식경제부의 우편배달시스템에 사용하는 우편번호를 하나의 집락(cluster)으로 이용하여 표본조사를 실시할 수도 있을 것이다. 본 연구는 비용을 고려하여, 등록센서스에서는 전수조사를 실시하지 않기 때문에, 조사구를 별도로 설정할 필요는 없다고 판단하고 있으며, 이 때문에 조사구를 센서스 표본조사를 실시하기 위한 프레임으로 설정할 가능성을 배제한다. 따라서 현재로서는 주민등록대장과 건물/주택대장의 주소명부를 점검하는데 필요한 표본틀은 현재의 행정동 시스템의 통반 시스템을 활용하는 것이 비용 효율적(cost-effective)이라고 판단한다.

분명히, 표본을 특정방식으로 클러스터(cluster)를 단위로 선정하는 것, 가령 응답자를 특정지역을 기준으로 하여 선정하는 것은 대단히 비용 효율적이다. 물론, 시간을 단위로 하여 클러스터를 형성하는 것을 고려할 수도 있겠지만, 등록센서스도 센서스의 본질적 특성인 동시성의 원칙을 준수하기 때문에, 이 부분을 별도로 고려하지 않기로 한다.

집락표집은 2단계 표집(two-stage sampling) 또는 다단계 표집의 예에 해당한다. 2단계의 표집에서는 먼저 행정동을 선정하고, 2단계로 행정동 내의 통반을 표본으로 선정하여, 해당 통반에 있는 모든 가구와 가구에 거주하는 구성원을 조사객체로 하여 전수조사를 실시하는 것이다.

지역을 기반으로 하는 표본추출은 여행이나 여타 센서스 과정에 필요한 행정비용을 줄이는데 기여한다. 이것은 또한 목표모집단의 모든 구성요소를 기재한 표본틀(sampling frame)을 필요로 하지 않다는 것을 의미한다. 물론, 집락표집은 클러스터 내 조사가구 간 차이에 비하여 클러스터 간 조사가구가 어떻게 다른가에 따라, 일반적으로 단순무작위표집보다 표본추정치의 분산을 증가시킬 염려가 있다.

그럼에도 불구하고, 집락표집의 단점 중 하나는 실제로 어떤 클러스터, 곧 집락을 선택하는가에 따라서 표본추정치의 정밀도가 달라진다는 것이다. 선택된 클러스터가 특정의 방식으로 편향된 지역이라면, 이 표본추정치로부터 모수를 추정하게 되면, 그 추정결과는 정확성을 줄어드는 결과를 초래하게 될 것이다.

<표 5-11> ‘인구주택총조사’ 조사구와 ‘반’의 규모 비교

지 역	인구주택총조사 조사구 수 (A)	‘반’의 수			비 율 (B/A, %)
		계 (B)	도시	농촌	
전 국	273,188	480,461	367,364	113,097	75.9
서 울	56,424	101,948	101,948	-	80.7
부 산	20,348	28,121	27,271	850	38.2
대 구	13,571	21,921	20,383	1,538	61.5
인 천	14,367	21,288	19,777	1,511	48.2
광 주	7,772	11,097	11,097	-	42.8
대 전	8,195	13,674	13,674	-	66.9
울 산	5,663	9,405	7,655	1,750	66.1
경 기	56,575	86,610	70,600	16,010	53.1
강 원	9,188	20,683	13,706	6,977	125.1
충 북	8,767	17,985	8,170	9,815	105.1
충 남	11,661	24,812	14,555	10,257	112.8
전 북	11,001	23,960	13,314	10,646	117.8
전 남	11,684	22,984	7,908	15,076	96.7
경 북	16,606	38,661	15,495	23,166	132.8
경 남	18,287	32,108	18,772	13,336	75.6
제 주	3,079	5,204	3,039	2,165	69.0

*자료출처: 통계청(2005.11.1), 행정안전부(2008.1.1.)

다단계 표본추출은 클러스터 표본추출의 복잡한 형태라고 볼 수 있는데, 여기서는 2개 이상의 단위수준이 다른 수준에 포함되어 있다. 제1단계는 표

본을 추출하기 위한 클러스터를 구성하는 단계이다. 제2단계는 1차적 단위의 표본이 각각의 클러스터에서 무작위로 선정된다. 다음단계에서는, 선정된 클러스터에서 추가적인 표본이 선정된다. 마지막 단계에서는 조사객체가 모두 선정된다.

다단계 표본추출은 기본적으로, 이미 선정된 확률표본에서 무작위로 다시 확률표본을 선정하는 과정이라고 보면 되겠다. 이것은 진정한 의미의 무작위 표본만큼 효율적이지는 않지만, 무작위 표본추출에 내재하는 많은 문제점을 해결하여 준다. 특히, 이것은 "multiple randomization"의 개념에 의존하기 때문에, 효율적인 표본추출 전략t많은 그 자체로서 다. 특히, 이에서도 쉽게 이용할 수 있는 대단히 유용한 방법이라고 볼 수 있다.

다단계 표집방법은 주로 모집단 구성원의 완전한 리스트를 이용할 수 없을 때 또는 그것을 이용하는 것이 부적합하다고 할 때 이용하는 표집방법이다. 더욱이, 선정된 모든 클러스터의 표본단위를 이용하지 않음으로 해서, 단단계 표본추출은 전통적인 집락표본추출에서 볼 수 있는 불필요한 대규모의 조사 비용을 경감할 수 있는 방법으로 센서스 표본조사에 유용하게 사용될 수 있을 것이다.

2015년 등록센서스를 위하여 집락표집을 하는 기본적 방법을 예시하면 '반'을 '조사구'와 같은 개념으로 이용하여 추출하는 방식이다. 기존 총조사의 '조사구'는 평균 60가구를 기준으로 묶게 되는데 반해, 지방자치법에 의한 '반'의 구성은 대개 20~30가구를 단위로 묶되 50호를 초과하지 않는 범위 내로 한정되어 있다. 2005년 '인구주택총조사'에서 최종 확정된 조사구가 273,188개인 반면, '반'의 수는 480,461개(2008.1.1. 기준)로 '반'의 구성이 세분화되어 있어 '조사구'를 단위로 했을 때보다 더욱 정교한 표본추출이 가능할 것이다. 우리나라의 통·리·반의 설치현황은 <표 5-11>와 같다

3. 표본규모

센서스 표본조사는 앞에서 언급한 것처럼, 어느 정도의 표본규모를 택해야 할 것인가는 대단히 중요하다. 전수조사를 실시하지 않고, 표본조사를 통하여 센서스의 주요정보를 획득하는 경우, 어떤 것이 바람직한 표본규모인가에 대하여 절대적 기준은 없다. 현재, 등록센서스를 실시하는 국가의 경우, 싱가포르, 이스라엘, 대만의 경우는 센서스 모집단의 20%를 표본으로 선정하여 그들에 대한 표본규모를 선정하여 왔다. 이것은 현재로 볼 때, 우리나라의 현행 인구주택총조사의 표본조사 표본규모인 10%보다는 큰 규모이다. 이 점

은 추정결과의 정확성과 정부의 센서스 예산규모 등의 밸런스를 획득하는 방식으로 결정되어야 할 것이다.

4. 조사방법론

현행 인구주택총조사는 가구방문조사를 원칙으로 하지만, 인터넷조사, 우편조사를 이용하는 것으로 되어 있다. 우리나라의 2005년 인구주택총조사는 인터넷조사를 도입하였지만, 사실상 그것을 통한 응답자가 적었다는 점에서 실패한 것이라고 평가한 것이라고 평가할 수 있다.

2015년 인구주택총조사에서 센서스 표본조사를 실시하는 경우에, 원칙적으로 가구방문조사를 원칙으로 하고, 인터넷조사, 우편조사를 이용할 수 있을 것이라고 판단한다. 그러나 조사의 정확성을 위하여 특히 인구주택총조사에서 집계범위의 정확성을 조사한다는 점에서, 현장조사원이 어떻게 활용될 수 있는가를 확실하게 결정하지 않으면 안 될 것이다.

5. 집계난이도 개념

센서스 표본조사는 주민등록시스템의 과소집계, 과다집계를 추정할 때, 지역별 특성을 고려하는 문제를 생각할 필요가 있는가를 염두에 두어야 할 것이다. 집계난이도에 따른 특정지역의 과다추출(oversampling)은 미국의 American Community Survey에서도 볼 수 있으며, 영국의 사후조사에는 집중적으로 도입된 방법이다(부록의 해당항목, 영국의 One Number Census와 미국의 순환센서스를 참고할 것).. 집계난이도의 개념은 우리나라의 경우, 대규모의 신도시 공영개발사업이 전개되고 있는 수도권 지역에서 특히 유용한 개념이라고 생각할 수 있다. 물론, 표본조사가 이상적으로 진행되고 있다고 판단하는 지역은 독일에서 2011년에 계획된 등록센서스에서 설정하는 “특수지역”(special area)로 간주하여, 집단가구의 경우와 마찬가지로 전수조사를 하는 방법도 생각해 볼 수 있을 것이다.

6. 조사시점

현행과 같이 5년 주기의 연도에서 11월 1일을 조사시점으로 할 것인가에 대해서도 등록센서스의 실시와 함께 결정을 내리는 것이 현명하다고 생각한다. 표본조사의 실시시점이 반드시 11월 1일에 설정되어야 할 것인가에 대해서는 상당한 관심을 가지고 등록센서스를 추진하기 전에 다시 한 번 논의를

진행해야 할 사항이다.

7. 현지실사의 관리

조사 관리자, 조사원의 모집, 훈련, 조사 진행 및 경과보고, 비상대응조치 방안 등에 대하여 논의를 할 필요가 있을 것이다.

각종 비상대응조치는 두말할 필요도 없이, 조사환경의 악화에 대비하여 등록센서스로 실시될 수 없는 조사항목에 대해서 효과적으로 표본조사를 실시하기 위한 것이라 할 수 있다. 반드시 유사한 상황은 아니지만, 영국의 One Number Census 실시를 위한 사후조사 현장관리도 우리나라의 등록센서스 표본조사 현장관리에 도움이 될 수 있을 것이라고 판단한다(전광희, 2008b).

제2부

등록센서스 이행에 따른

통계 이용의 변화

제6장 등록센서스의 결과표 작성과 통계자료 시계열 변동 보완방안

제1절 논의의 출발점과 전제

행정자료를 이용한 등록센서스는 조사표와 현지실사 형태로 확보되는 전통적 센서스와 그 자료원에서 일차적인 차이를 보인다. 따라서 전통적 센서스에서 등록센서스의 전환은 다양한 형태의 난제를 낳을 수밖에 없다. 이미 앞에서 이러한 난제들에 대한 심도 있는 논의가 전개되고 있다. 여기에서는 센서스결과활용, 특히 공표자료와 관련된 이슈들을 조명하기로 한다.

논의에 앞서 이 연구에서 전제하는 바를 몇 가지 밝혀둘 필요가 있다. 우선, 이 연구에서는 통계청이 2015년에 실시하려고 구상중인 등록센서스에서 이용할 행정자료원이 '주민등록대장'과 '건축물등록대장'에 한정되었다고 가정한다. 물론, 등록센서스가 진전되면서 더 많은 형태의 행정자료원(administrative data source)(예. 자동차대장, 가족관계등록부)이 활용될 것이나, 그 출발점은 앞서 상기한 두 가지 자료원이 될 것으로 보인다. 다음으로 행정자료원을 통해 대체가능한 몇 가지 전수조사 항목 이외의 항목들에 대해서는 표본조사를 통해 접근한다고 가정한다. 전통적으로 표본조사는 총조사의 5-10%에 그치었으나, 등록센서스와 함께 진행될 표본조사는 총조사의 20%에 이를 것이라는 전망이 있다.

이 연구는 다음과 같은 세부 주제로 구성된다. 첫째, 자료원의 변동에 따라 제기될 수 있는 문제의 맥락을 통계품질관리의 측면에서 접근한다. 이를 위해서는 통계품질관리의 일반적 기준과 더불어, 인구주택총조사의 품질관리기준을 살펴볼 필요가 있다. 다음으로 그간의 인구주택총조사의 조사항목을 점검하는 한편, 자료원의 변화에 따라 대체가능한 항목을 정리한다. 이를 이어, 공표결과표의 일관성을 확보하기 위한 논의를 전개하는 한편, 등록센서스에 따른 신규통계표의 유형에 대하여 구체적인 제안을 하도록 한다. 이상의 논의들은 기존의 연구결과물, 전문가 논의 등에 기초한다.

제2절 통계품질관리기준으로서의 일관성

조사대상 전체에 관한 정보를 추구하는 센서스 또한 여타의 통계조사들과 마찬가지로 그 품질관리를 위해서 몇 가지 기준을 따르도록 요구된다. 통계 품질관리에 대한 기준은 국제기구나 국가마다 차이를 보이고 있기는 하나, 대체적으로 정확성, 적절성, 시의성, 접근성, 일관성이라는 다섯 가지 차원에서 설정될 수 있다. 다음의 <표 6-1>은 주요 국제기구와 국가의 통계품질관리 기준의 차원(dimension)을 개략적으로 보여주고 있다. 물론 각 차원이 무엇을 의미하며, 그 구성요소와 구체적인 지표가 무엇인지 또한 국제기구나 국가 간에 다소 차이를 보인다. 그럼에도 불구하고, 이러한 다섯 가지 차원이 통계품질관리의 중요 기준이라는 점에는 이의가 없을 것으로 보인다.

<표 6-1> 국제기구 및 주요 국가의 통계품질관리의 차원(기준) 비교

구분	통계품질차원(기준)					
	정확성	적절성	시의성	접근성	일관성	기타
OECD	◎	◎	◎	◎	◎	
IMF	◎	◎	◎	◎	◎	• 필수조건
Eurostat	◎	◎	◎	◎	◎	
캐나다	◎	◎	◎	◎	◎	
영국	◎	◎	◎	◎	◎	
스웨덴	◎	×	◎	◎	◎	• 보고서 내용
미국	◎	◎	◎	◎	×	

- 주: 1) 적절성은 관련성으로 표현되기도 하며, 완전성을 포괄하는 개념임.
 2) 시의성에는 국제기구나 국가에 따라 정시성, 주기성을 포괄하는 개념임.
 3) 접근성에는 국가나 국제기구에 따라 해석가능성, 이용가능성, 투명성을 포괄하는 개념임.
 4) 일관성에는 국가나 국제기구에 따라 비교성을 포괄하는 개념임.
 5) IMF의 정확성은 신뢰성을 포함하는 개념임.

출처: 김태현 외 (2004) “인구주택총조사 통계품질관리지표 개발”

이상의 다섯 가지 통계품질관리기준 중에서 공표결과물에 대한 논의는 일관성에 가장 집중될 것이다. 일관성(coherence)은 조사나 통계 작성에 있어서 공통적인 개념, 정의, 분류 등을 적용함으로써 시공간적인 비교가 가능함을 의미한다. 이러한 의미에서 스웨덴 등에서는 일관성을 비교성(comparability)과 혼용하여 적용하기도 한다. 비교성은 주로 개념 및 측정방법의 변동을 염두에 둔 것으로, 이러한 면에서의 변동이 시간과 지리상의 비교를 얼마나 확보해 줄 수 있는가를 의미한다.

국제기구나 국가마다 다소 차이가 있으나, 일관성은 크게 자료 내 일관성, 자료간의 일관성, 시계열적 일관성, 지리적 일관성 등으로 구분되어 측정된

다. 자료 내 일관성은 동일 자료 내 항목간에 동일한 개념이 적용되어 결합하여 이용될 수 있는 정도이다. 자료간 일관성은 서로 다른 자료원일지라도 개념을 공통적으로 적용하여 비교가 가능함을 의미한다. 시계열적 일관성은 서로 다른 기준시점일지라도 공통적인 개념을 적용함으로써 시간적 비교가 가능함을 의미한다. 지리적 일관성도 공간적인 비교 가능성을 의미하는 것이며, 여기에는 국제적 기준에 의거한 국제적 일관성도 포함된다.

<표 6-2> 인구주택총조사 품질관리기준: 일관성 대상 요소 및 항목

요소	항목
자료내 일관	<ul style="list-style-type: none"> • 항목들간 개념, 정의, 분류 등의 모순이 없으며, 의미 있는 결합 가능 • 동일 시계열 자료내 동일 방법에 의한 동일한 내용 비교 • 잠정통계와 최종통계간의 일관성
자료간 일관	<ul style="list-style-type: none"> • 공통적인 개념, 정의, 분류 등에 의거한 경우 자료간 차이 설명 가능 • 다른 자료원들로부터의 유사 추정치에 대한 기술 • 다른 자료원들 결과간 차이 정도 및 제시
시계열적 일관	<ul style="list-style-type: none"> • 시간에 따른 개념과 측정의 변동 및 그 효과 • 연속적 통계발행에 있어서 불일치 정도 • 시간에 따른 통계적 개념 등의 변동 및 그 영향 • 공통적인 개념, 정의, 분류 등에 의거한 경우 시간간 차이 설명 가능
지리적/국제적 일관	<ul style="list-style-type: none"> • 공통적인 개념, 정의, 분류 등에 의거한 경우 국가간 차이 설명 가능 • 개념과 정의, 범위, 분류, 기록 등은 국제적으로 인정되는 기준과 일관성 유지 • 지리적으로 특성 등에 관한 정의의 유사성 • 국내 및 국제 정의 및 기준 일치 • 국제적 규약에 일치 • 국내 및 국제 정의에 벗어난 내용 및 이유 기술
일관성 유지위한 노력	<ul style="list-style-type: none"> • 정의, 분류, 방법 등 개정관련 연구, 분석 공개 • 정기적, 투명한 계획된 개정 • 다른 자료원들과의 일관성에 관한 기술 • 수정 계획 공개 • 수정 내용 공개 • 수정분석에 대한 참고문헌 및 링크 • 이용자(특히 원시자료 이용자)를 위한 통계생산과정 문서화

출처: 김태현 외 (2004) "인구주택총조사 통계품질관리지표 개발"

이상과 같은 통계품질관리 기준에 입각하여, 인구주택총조사의 품질관리를 위한 체계를 개발하려는 노력이 있어왔다. 김태현 등(2004)에 의하면 센서스 관리기준으로서의 일관성은 대체로 다음의 <표 6-2>와 같이, 5가지 요소로 구성될 수 있다고 한다. 김태현 등(2004)은 이상과 같은 일반적인 논의에서 한 발 더 나아가, 한국인구주택총조사의 일관성을 확보하기 위해서 일관성유

지를 위한 노력을 제외한 4개 요소에 주목하는 한편, 각 요소의 구체적인 지표도 1-2개로 축소하여 다음의 <표 3>과 같이 제안한 바 있다.

<표 6-3> 인구주택총조사 일관성 체계

요소	항목
자료내 일관성	동일 인구주택총조사 내 조사항목들의 통계적 개념(정의, 분류 등) 일관성
자료간 일관성	인구주택총조사와 다른 자료원간의 통계적 개념(정의, 분류 등) 일치성
시계열 일관성	인구주택총조사간의 통계적 개념(항목, 정의, 분류 등) 일관성
지리적 일관성	인구주택총조사와 국제적 인정기준(항목, 정의, 분류, 조사방법 등)간의 일관성 국제적 기준에 따르지 않은 경우 그 내용 및 이유 제시

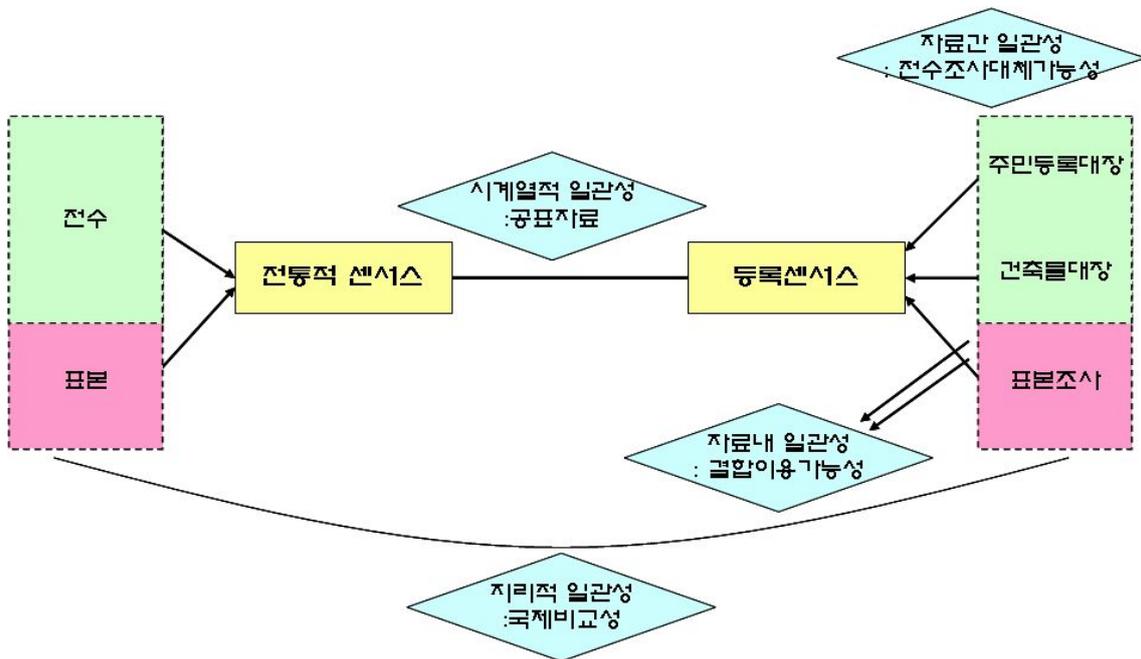
출처: 김태현 외 (2004) “인구주택총조사 통계품질관리지표 개발”

인구주택총조사의 일관성 체계에 관한 이상의 논의는 전통적인 센서스에 국한된 면이 적지 않다. 현재 논의되고 있는 등록센서스로의 이행과 관련하여 일관성의 측면을 주목한다면 다음의 <그림 6-1>, <표 6-4>와 같이 그 이슈들이 정리될 수 있다.

우선 <그림 6-1>에서 나타난 바와 같이, 전통적 센서스는 거대한 규모의 사회조사의 일종으로 단일대상, 단일조사표에 근거한다. 물론 전수조사와 표본조사로 나뉘기는 하지만 그 대상은 동일하며, 조사문항도 전수문항의 추가에 불과하다. 그런데 등록센서스는 여러 개의 행정자료원을 이용하는 동시에 표본조사를 병행함으로써, 다양한 자료원을 가진 셈이다. 따라서 <표 6-4>에서 제시된 바와 같이, 등록센서스 구축시에 여러 행정자료원과 표본조사 항목들간의 개념, 정의, 분류 등의 차이를 극복하고 비교가능성 및 일관성을 높여야 한다는 일차적인 부담을 지게 된다. 이러한 부담은 전통적인 센서스의 결과와 행정자료집계결과간에 발생하였던 자료간 일관성 문제(예. 총조사 결과 인구수와 주민등록집계상 인구수의 차이)를 등록센서스 내부에 포함시키는 것과 동일하다.

등록센서스 자료원의 다양성 문제는 자료내 일관성과도 매칭된다. 등록센서스가 거대한 단일자료로 구축되었다고 가정한다면, 자료내 일관성은 이 자료를 활용하여 사회적으로 유의미한 공표자료를 제공할 때 중요한 이슈가 될 수 있다. 특히 각 자료원으로부터 나온 항목들을 결합하여 사용하려 할 때 자료원간의 항목개념, 정의, 분류 등의 모순이 없으며, 의미있는 결합이 가능해질 것인가가 주요 관심사가 될 것이다.

<그림 6-1> 등록센서스로의 이행에 따른 일관성 이슈들



<표 6-4> 등록센서스 일관성 이슈들

요소	항목
자료내 일관성	등록센서스 이용시에 항목들간 개념, 정의, 분류 등의 모순이 없으며, 의미있는 결합이 가능한 정도
자료간 일관성	등록센서스 구축시에 여러 행정자료원 및 표본조사 항목들간 개념, 정의, 분류 등의 차이 정도
시계열 일관성	기존의 전통적 센서스와 항목, 정의, 분류 등에서의 일관성
지리적 일관성	등록센서스 및 인구주택총조사의 국제적 인정기준(항목, 정의, 분류, 조사 방법 등)과 일치하는 정도

지리적 일관성은 대체로 국제적 인정기준 혹은 권고기준에 부합되는 정도를 의미하는데, 등록센서스를 실시함에 따라 국제적 비교기준의 준거가 등록센서스를 실시하는 국가로 옮겨가야 할 가능성이 높을 것으로 보인다. 물론 국제비교가능성을 논할 때 전통적 센서스를 실시하는 국가 또한 염두에 둘 것이나 아무래도 그 비중있는 준거는 등록센서스를 실시하는 국가가 될 것이다. 그런데, 각국의 행정자료원이 해당 국가의 고유한 정책적, 행정적 맥락에서 형성되기 때문에, 전통적인 총조사에서 획득할 수 있는 국제비교가능성 보다는 기대되는 바가 적을 것으로 보인다.

시계열적 일관성은 등록센서스를 구축하는 단계에서 이슈가 되는 자료간 일관성, 등록센서스를 활용하는 단계에서 이슈가 되는 자료내 일관성 문제 이외에도 기존에 실시되었던 전통적 센서스의 항목, 정의 분류 등을 얼마나 유지하느냐를 문제를 안고 있다. 등록센서스의 자료간 일관성과 자료내 일관성이 어느 정도 담보된 상태에서라도 시계열적 일관성을 확보하는 데는 독자적인 어려움이 있을 것으로 보인다. 본 논문에서는 앞서 논의한 바와 같이, <그림 6-1>과 <표 6-4>에 제시된 자료내 일관성, 자료간 일관성 등은 어느 정도 확보되었다고 가정한다. 한편 지리적 일관성은 현 시점에서 논의할 바가 아니라고 본다. 여기에서는 시계열적 일관성이 유지하기 노력의 기준점이 무엇인지, 그리고 공표결과표의 예상형태를 논의하기로 한다.

제3절 항목 및 공표결과표의 시계열적 변화

다음의 <표 6-5>는 1960년부터 2005년까지 인구주택총조사의 조사항목을 정리하고 있다. 조사항목수와 조사내용은 사회적, 정책적 요구와 환경에 따라 변동하기도 한다. 어떤 항목은 전수에서 표본으로 이동하기도 하며, 추가, 삭제, 분류범주 변동 등이 관찰되기도 한다. 그럼에도 불구하고 기본적인 조사문항은 일관성 있게 유지되고 있다.

이 표에서 제시하지는 않았지만, 성명을 제외한 조사항목의 대부분은 관련 항목을 통해서나 집계표 등을 통해 공표되고 있다(<부표 6-1> 참조). 물론 공표자료에서는 조사항목을 원안대로 쓰지 않고, 그 일부만을 추출해서 쓰거나 가공해서 쓰는 경우도 많다. 그 대표적인 예가 연령인데, 공표자료에서는 “만나이”로 제시되고 있다.

<표 6-5> 실시 연도별 총조사 항목표(1960~2005년)

□ 인구항목

항 목	1960	1966	1970	1975	1980	1985	1990	1995	2000	2005
총 계	37	15	31	28	39	30	45	28	50	44
전 수	37	9	15	11	23	30	33	17	20	21
표 본	-	6	16	17	16	-	12	11	30	23
기본항목 계	10	9	8	6	7	7	6	7	8	8
성명	●	●	●	●	●	□	●	●	●	●
본관						□	●		●	
가구주와의 관계	●	●	●	●	●	●	●	●	●	●
성별	●	●	●	●	●	●	●	●	●	●
연령(생년월일)	●	●	●	●	●	●	●	●	●	●
교육정도		●	□	□	□	□	□	□	□	□
- 졸업여부	●	●	□	□	□	□	□	□	□	□
- 취학여부	●	●	●	□	□	□	□	□	□	□
- 수학연수	●									
- 전공학과					○				○	
- 문맹여부	●	●	●							
혼인상태	●	●	●	●	●	●	●	●	●	●
종교						●		●		●
국적	●									
남북이산가족										●
인구이동 계	1	-	2	1	6	3	7	5	7	5
출생지	●		○		○	●	●	○	●	
전 거주지										
- 1년전 거주지					○	●	○		○	
- 5년전 거주지			○	○	○	●	○	○	○	○
통근(학) 여부					○		●	□	○	○
통근(학)지					○		●	□	○	○
통근(학) 시간							●	○	○	○
이동교통수단					○		●	○	○	○

주 : ● 전수조사, ○ 표본조사
 (●)는 항목합계 수치에 불포함

□ 가구항목

항 목	1960	1966	1970	1975	1980	1985	1990	1995	2000	2005
가구 계	11	-	7	4	10	10	15	7	16	11
거처의 종류	●		●	●	●	●	●	●	●	
- 단독주택 종류									●	
가구구분						●	●	●	●	●
점유형태						●		●	●	●
거주기간	●							○	○	○
임차료						●	○	○	○	○
사용방수	●					●	●	●	●	●
점유(소유)관계	●		●	●	●	●	●			
가구 심신장애자					●					
문화시설 및 가재			●	○	●	●	●			
정보통신기기보유										
주거(전용)시설	●						●			
- 부엌시설	●						●			
- 화장실시설	●		○		●	●	●	●	●	●
- 목욕시설							●			
- 식수(상수도)시설			○				●		○	
- 식수사용형태									○	
주된 연료				○						
취사연료			○		●	●	●		○	
난방연료					●					
난방시설					(주택 쪽에 포함)	●	●		○	●
아궁이 형태	●									
조명시설			○							
굴뚝상태	●									
대청마루유무/평수	●									
식수종류(이용식수 원)	●				●		●			
가구소득							○			
자동차보유대수									○	○
주차시;설									○	○
주인가구여부 및 주택소유 여부									(주택 소유 제외)	●
거주층										○

□ 주택항목

항 목	1960	1966	1970	1975	1980	1985	1990	1995	2000	2005
주택 계										
거처의 종류										┌ ●
건물층수										┌ ●
대지면적					┌ ●		●	●	●	●
연건평			●		┌ ●	●	●	●	●	●
거주자능가구수						●	●	●	●	
총방수		┌ ●	┌ ●	┌ ●	┌ ●	┌ ●	●	●	●	●
동거가구수		┌ ●	┌ ●	┌ ●	┌ ●	┌ ●	●			
건축년도	●		●	●	●			●	●	●
건물수선상태	●									
다가구주택						(●)				
집/거주의 개수	●									
건무내 입주여부	●									
지붕재료	●		○	○	●		●			
외벽재료	●		○	○	●		●			
편의시설 (부엌,화장실 등)						●	●	●	●	●
* 주권 난방시설					●					다
주택용도			○							
건물의 종류	●		●	●						

주 : 2005년도는 '거처의 종류'항목이 주택항목으로 이동

다음의 <표 6-6>은 지난 2005년도에 실시된 인구주택총조사와 2010년에 실시예정인 인구주택총조사의 조사항목을 대조하여 제시하고 있다. 이 표에서 제시된 바와 같이, 2010년에는 남북이산가족과 종교에 관한 문항이 빠지면서 국적과 입국연월이 인구전수항목으로 추가된다. 또한 인구표본문항에는 출생지, 1년전 거주지, 현직업 근무연수 및 사회활동 항목이 추가된다. 가구항목에서는 전수조사항목의 변화는 없으며, 표본조사항목에서 자동차보유대수 대신 교통수단 보유 및 이용현황이 조사되고, 정보통신기기보유 및 이용현황, 수도 및 식수사용형태, 난방시설 등이 추가된다. 주택부문에서는 전수항목으로 있던 대지면적이 표본항목으로 옮겨지는 한편, 난방시설 등이 삭제되는 변동이 있다.

<표 6-6> 2005년과 2010년 인구주택총조사항목

구분	2005	2010	
인구	전수	①성명 ②성별 ③가구주와의 관계 ④나이 ⑤교육정도 ⑥혼인상태 ⑦남북이산가족 ⑧종교	①성명 ②성별 ③가구주와의 관계 ④나이 ⑤교육정도 ⑥혼인상태 ⑦국적 ⑧입국연월
	표본	①5년 전 거주지 ②경제활동상태 ③종사상지위 ④산업 ⑤직업 ⑥근로장소 ⑦출생아수 ⑧혼인연월 ⑨통근·통학여부 ⑩통근·통학장소 ⑪이용교통수단 ⑫통근·통학 소요시간 ⑬활동계약 ⑭아동보육상태 ⑮추가계획자녀수 ⑯고령자 생활비 원천 ⑰출생지 ⑱1년 전 거주지 ⑲현 직업 근무연수 ⑳사회활동	①5년 전 거주지 ②경제활동상태 ③종사상지위 ④산업 ⑤직업 ⑥근로장소 ⑦총 출생아 수 ⑧혼인연월 ⑨통근·통학여부 ⑩통근·통학장소 ⑪이용교통수단 ⑫통근·통학 소요시간 ⑬활동계약 ⑭아동보육 ⑮추가계획자녀수 ⑯고령자 생활비 원천 ⑰출생지 ⑱1년 전 거주지 ⑲현 직업 근무연수 ⑳사회활동
가구	전수	①가구구분 ②사용방수 형태 ③주거시설 형태 ④점유형태 ⑤거주층 ⑥주인가구 및 주택소유 여부	①가구구분 ②사용방수 형태 ③주거시설 형태 ④점유형태 ⑤건물 및 거주층 ⑥주인가구 및 타지주택소유여부
	표본	①거주기간 ②주차시설 ③임차료 ④자동차보유대수	①거주기간 ②주차장소 ③임차료 ④교통수단 보유 및 이용현황 ⑤정보통신기기보유 및 이용현황 ⑥수도 및 식수사용형태 ⑦난방시설
주택	전수	①건축연도 ②총 방수 ③연건평 ④편익시설수 ⑤거처의 종류 및 건물층수 ⑥대지면적 ⑦난방시설	①건축연도 ②총 방수 ③주거용연면적 ④주거시설수 ⑤거처의 종류
	표본		①대지면적

전통적인 인구주택총조사 항목을 대체할 수 있을 것이라 기대되는 행정자

료원의 항목들은 다음의 <표 6-7>과 같다. 이 표에 의하면 주민등록자료에서는 인구부분의 6개 항목, 가구부분의 2개 항목, 건축물등록대장에서 가구부분의 1개 항목, 주택부분의 5개 항목의 대체가능성이 높은 것으로 나타난다. 이 중에서 인구주택총조사의 가구구분은 주민등록자료의 세대주와의 관계에서 이차적으로 가공된다는 특징을 가진다.

<표 6-7> 행정자료원별 인구주택총조사 항목간 비교

행정자료 총조사	주민등록자료	건축물대장
인구	①성명(성명) ②성별(성별) ③나이(나이) ④가구주와 관계(세대주와 관계) ⑤1년 전 거주지(1년 전 주소) ⑥5년 전 거주지(5년 전 주소)	
가구	①가구구분(세대활용) ②거주기간(현주소전입일)	①주인가구 및 타지주택소유 여부 (소유자 현황)
주택		①거처의 종류(건물의 주용도, 기 타용도) ②주거용 연면적(건물의 층면적 · 호별 전유면적) ③건축년도(사용승인일자) ④총방수(총방수) ⑤주거시설수(주거시설수)

행정자료와 인구주택총조사는 자료수집의 목적 자체가 다를 뿐 아니라, 그 조사항목의 개념 또한 다르다. 다음의 <표 6-8>은 주민등록자료 및 건축물등록대장과 인구주택총조사의 항목을 비교하고 있다. 이 중에서 '가구주와의 관계(인구주택총조사)'를 대체하는 항목으로 '세대주와의 관계'를 살펴보면(물론 가구주와 세대주가 갖는 개념상, 관행상의 차이에도 불구하고), 등록센서스에서 오히려 더 세분화된 가구구분이 가능해질 것으로 보인다. 이는 주민등록에서의 세대주와의 관계가 75가지로 분류되어 인구주택총조사의 14가지 분류보다 훨씬 상세하기 때문이다. 다양한 가족형태가 출현하고 이들에 대한 사회적 이해의 필요성(예. 한부모가족, 입양가족 등)이 높아지는 상황에서 고무적이라 할 수 있다.

<표 6-8> 2005년 인구주택총조사항목과 행정자료원간 개념비교

항목구분	조사목적	개념		개념차이
		2005 총조사	행정자료 (주민등록, 건축물대장)	
성명	○ 조사대상 확인 ○ 조사 중복·누락 방지	○ 이름	○ 이름	차이 없음
성별	○ 인구구조 분석 장래 인구추계 변수	○ 남, 여	○ 남, 여	차이 없음
나이	○ 남녀 간의 인구·경제·사회적 특성 비교 ○ 학령인구, 병역인구, 생산연령인구, 기임여성 인구, 고령인구 등 인구를 특성집단별로 구분하는 변수	○ 양, 음력과 실제 생년월일	○ 주민등록에 신고된 생년월일	연령대가 높아질수록 연령 불일치율이 큼 (19세 차이내 95.3%일치)
인구 (6) 가구주와의 관계	○ 가구주의 성, 연령, 혼인상태별 특성을 기초로 장래가구수를 추계 ○ 가족구조, 가족유형, 가족의 생애주기 파악 ○ 가구·가족 특성변화, 가족 형성·해체 분석 연구	※가구주 개념 - 호주 또는 세대주와 관계없이 가구를 실질적으로 대표하는 사람 ○ 가구주와의 관계 - 가구주 - 가구주의 배우자 - 자녀 - 자녀의 배우자 - 가구주의 부모 - 배우자의 부모 - 손자녀, 그 배우자 - 증손자녀, 그 배우자 - 조부모 - 형제자매, 그 배우자 - 형제자매의 자녀, 그 배우자 - 부모의 형제자매, 그 배우자 - 기타 친인척 - 기타 동거인 등 14가지 분류	※세대주 개념 - 한 주소지에서 주거를 같이 하고 있는 세대를 대표하는 자 ○ 세대주와의 관계 - 본인 - 처 - 남편 - 자 - 자부 - 사위 - 양부 - 양모 - 모 - 부 - 시부 - 장인 - 장모 - 시모 등 75가지 분류	*가구주와 세대주 차이 : 가구주는 경제력이 있는 의사결정권자인 반면 주민등록의 세대주는 연장자인 경우가 많음 *관계차이 -> 없음 : 가구주와의 관계보다 세대주와의 관계가 분류는 더 많으나, 가구주와의 관계로 코딩이 가능

항목구분	조사목적	개념		개념차이	
		2005 총조사	행정자료 (주민등록, 건축물대장)		
인구 (6)	1년전 거주지 및 5년전 거주지	<p><상주와 현주의 개념> ※ 상주(de jure) : 평소에 살고 있는 거주지에서 조사하는 것 - 총조사 : 일정한 장소에서 거주한 기간과 거주하려는 기간을 합한 기간이 3월 이상인 경우 ※ 현주(de facto) : 조사당시 발견된 곳에서 조사하는 것 - 주민등록은 "30일 이상 거주할 목적으로 그 관할구역안에 주소 또는 거소를 가지는 경우"로 상주개념과 가깝다고 할 수 있으나, 실제로 주민등록상 등 록지와 상주지가 다른 경우가 많아 주민등록을 법적거주지라는 제3의 개념으로 파악하기도 함</p>	<p>○ 1년전, 5년전에 거주한 곳 ○ 1년전, 5년전 항목 - 태어나지 않았음 - 현재 살고 있는 집 - 같은 시군구 내 다른 집 - 다른 시군구 - 북한 또는 외국</p>	<p>○ 1년전, 5년전에 거주한 주소지 ○ 1년전, 5년전 항목- 주소지(시도, 시군구, 읍면동, 번지, 건물명 등)</p>	<p>*총조사와 주민등록은 모두 상주개념을 따르고 있어 개념차이는 없음 다만, 주민등록 주소가 거주지의 목적에 따라 실제 상주지와 다른 경우가 많아 이에 대한 시정이 필요</p>
	가구 (2)	<p>○ 가구 추계 기초 자료 ○ 가구의 유형 파악 ○ 집단가구와 일 반가구를 구분하여 일반가구에 대해서만 가구항목을 조사 ○ 인력공급전망, 가구 규모 결정요 인</p>	<p>※ 가구개념 - 1인 또는 2인 이상 이 모여서 취사, 취침 등 생계를 같이하는 생활단위 ○ 가구구분 - 가족으로 이루어진 가구 - 가족과 가족 이외의 사람이 함께 사는 가구 - 1인가구 - 가족이 아닌 남남끼리 합 계사는 5인 이하의 가구 - 가족이 아닌 남남끼리 합 계사는 6인 이상의 가구 - 그 밖에 가숙사, 보육원, 양로원, 장애인 가티복지시 설에 살고 있는 집단가구로 구분</p>	<p>※ 세대개념 - 주거 및 생계를 같이 하는 집단 (주민등록사무편 략, 2008) ○ 가구구분코드 없 음</p>	<p>*가구와 세대는 개념상 큰 차이 는 없음 다만, 개인의 특별한 목적으로 세대를 허위로 분 리, 통합하는 차 이가 발생 *주민등록에 세 대구분이란 자 료는 없으며, 세대주와의 관 계를 이용하여 가구구분을 작 성</p>
주택 (3)	주택소유여부	<p>○ 실질적인 주택 소유통계 작성 ○ 주택보급률 산정 및 주택정책 수립</p>	<p>○ 가구별 주택소유 여부를 조사</p>	<p>○ 건축물대장 소유 자 현황에 등록된 자 들을 대상으로 주 택소유여부 파악</p>	
	주거용연면적	<p>○ 주택의 규모를 파악 ○ 주택의 종류와 지역별로 파악</p>			
	건축년도	<p>○ 주택의 질이나 노후 정도를 파악 ○ 재건축, 수리 등 주택 건설 시장의 수 요 예측</p>			

제4절 예상결과표의 예시

행정자료를 이용한 등록센서스와 인구주택총조사는 그 자료원의 변화를 의미하기 때문에 기존의 인구주택총조사와 등록센서스간의 시계열적 일관성을 확보하는데는 적지 않은 노력이 필요하다. 등록센서스로의 전환을 전제로 하였을 때 무엇보다도 메타자료(Meta Data)의 생성이 요구된다. 이러한 메타자료에는 등록센서스로의 변동근거를 비롯하여, 자료원의 비교소개, 조사항목 측정과 추정방식의 소개 등이 필수적이다. 한편으로 제시될 공표결과물에서도 각주와 비교란을 적극 활용하여 시계열적 이용에 따른 유의사항을 제시하여야 할 것이다.

등록센서스의 공표자료를 염두에 두었을 때, 예상되는 결과표의 작성기준은 다음과 같다. 우선, 가능한 한 기존 센서스의 결과표 양식을 그대로 재생산한다. 이 과정에서 자료원, 개념, 분류 등의 변동이 있을 시에 해당 결과표에 이를 적시하도록 한다. 보다 구체적으로 논의하자면, 등록센서스로 대체되는 항목의 자료원, 개념정의, 분류방식 등에 대한 충분한 설명이 공표결과표에 함께 부기되어야 할 것이다. 또한 가공항목에 대해서도 가공항목이 작성되는 과정과 결과에 대한 소개가 이루어져야 할 것이다. 적시할 내용이 많은 경우, 메타자료와 매칭시키고 이를 공지한다.

다음으로 등록센서스 실시에 따라 전수항목과 표본항목간의 이동이 있을 수 있다. 등록센서스의 조사 및 공표항목은 전통적인 인구주택총조사의 항목과 관련하여 다음의 <표 6-9>와 같이 4가지로 분류가능하다. 우선 인구주택총조사의 전수항목이 등록센서스의 전수항목이 되는 경우로 성명, 성별, 세대주관계 등이 포함된다. 이 항목은 다시 항목의 내용변화가 없는 경우(예, 성별)와 항목의 개념정의나 내용변동이 발생하는 경우로 세분된다(예, 가구주와의 관계에서 세대주와의 관계). 이들 항목들은 전수라는 특성을 그대로 유지한다.

인구주택총조사의 표본항목이 등록센서스의 전수항목이 되는 경우도 있다. 예를 들어 5년 혹은 10년 주기로 표본항목으로 조사되었던 '1년전 및 5년전 거주지' 등은 이제 등록센서스에서 전수항목이 된다. 따라서 매 등록센서스에서 결과공표가 가능해진다. 이와는 다르게 인구주택총조사의 전수항목이 등록센서스에서 표본항목이 되는 경우도 있다. 가령, 교육정도는 매번의 인구주택총조사에서 전수항목이 되었으나, 등록센서스에서는 표본조사에서 조사되어야 할 항목으로 바뀌어진다. 마지막으로 인구주택총조사와 등록센서스 모두에서 표본조사항목인 경우로 다수의 인구주택총조사 표본항목이 이에

해당한다.

공표자료는 대체로 단일 항목에 대한 정보보다는 항목을 결합하여 제공된다. 표본항목이 결합항목으로 설정되어 있는 경우, 이 항목들의 자료는 전수가 아닌 표본으로 제시되게 되어 있다. 앞서 언급한 바와 같이, 등록센서스에서 전수에서 표본으로, 표본에서 전수로 이동한 항목의 경우 이들이 제시되는 바도 전수나 표본으로 바뀌게 된다. 이용자의 관점에서 전통적으로 표본편에서 제시되었던 항목이 전수편에 제시된다면(혹은 그 역으로) 혼동스러울 수 있을 것이기 때문에 이에 대한 안내가 필요할 것이다.

<표 6-9> 인구주택총조사와 등록센서스의 전수 및 표본항목변화

	인구주택총조사 전수	인구주택총조사 표본
등록센서스 전수	예) 성별, 세대주관계	예) 1년전 거주지
등록센서스 표본	예) 교육정도	예) 추가계획자녀수

인구주택총조사의 결과물은 대체로 몇 가지 형태로 공개되지만 추세는 발간책자를 CD의 형태로 담는 것이라 할 수 있다. 여기에서는 2005년도 인구주택총조사(전수, 표본) CD에 담겨진 통계표를 기준으로, 등록센서스 실시에 따라 예상되는 집계표를 상정해보기로 한다. 다음의 <표 6-10>은 2005년도 인구주택총조사의 전수항목과 표본항목 통계표 중 일부를 뽑아, 등록센서스에 나타날 형태를 보여주고 있다(전체 통계표에 대한 예상표는 <부표 2>와를 참조하기 바란다).

<표 6-10> 등록센서스 공표통계표의 예시

예제	전통적 센서스	등록센서스
1표	(인구편 전수) 행정구역 및 총조사 실시년도별 인구...	▷자료원변동/주민등록상의 인구/유의점 제시
3표	(가구편 전수) 가구의 성, 연령 및 혼인상태별 가구(일반가구)	▷표본: 혼인상태▷가구주에서 세대주로의 변동/유의점제시
1표	(고령자 표본) 행정구역 및 연령, 성별 고령자	▷전수▷연령/개념변동/유의점제시
1표	(인구이동편 표본) 현거주지, 성, 연령 및 5년 전 거주지 유형별 인구..	▷전수▷5년전거주지 개념변동

제5절 시계열적 일관성 확보 노력

행정자료를 이용한 등록센서스와 인구주택총조사는 그 자료원의 변화를 의미하기 때문에 기존의 인구주택총조사와 등록센서스간의 시계열적 일관성을 확보하는데는 적지 않은 노력이 필요하다. 이러한 노력은 등록센서스를 기획하고 준비하는 단계에서부터 이루어질 필요가 있다.

첫째, 기획하는 과정에서 항목의 정의, 분류, 추정방법 등 자료원 변동과 관련된 연구 및 분석이 이루어질 필요가 있다. 앞서 언급한 바와 같이, 여러 개의 자료원을 가지는 등록센서스가 단일 자료원을 가진 전통적 센서스의 결과물과 일관성을 갖기 위해서는 자료구축단계에서 일관성이 먼저 확보되어야 한다.

둘째, 등록센서스가 실시된 이후에, 자료원, 개념, 측정방식 변동의 효과를 분석하려는 노력이 요구된다. 특히 이 단계에서는 개념, 정의 및 분류방식이 다른 여러 개의 자료원을 통합하여 이를 한 개의 자료로 취급하기 때문에 자료내 일관성이 담보되어야 할 것이다. 이러한 자료내 일관성은 기존의 전통적 센서스와 동일한 혹은 이에 근접한 형태의 공표자료를 만들어내는 전초가 될 것이다.

셋째, 결과를 활용하는 단계에서 공표자료형태와 구성 내용이 가능하면 기존의 전통적 센서스와 유사하도록 할 필요가 있다. 이를 위해서는 최소한의 수준에서 시계열적 일관성을 보여줄 수 있는 항목과 내용이 무엇인지를 고민하여야 할 것이다. 또한 공표자료에 각주와 비고란을 적극 활용하여 시계열적 이용에 따른 유의사항을 제시하여야 할 것이다. 이와 관련하여 생각해 볼 수 있는 것이 메타자료(Meta Data)의 생성이다. 이러한 메타자료에는 등록센서스로의 변동근거를 비롯하여, 자료원의 비교소개, 조사항목측정과 추정방식의 소개 등이 필수적이다.

넷째, 기획에서 실시를 거쳐 결과공표에 이르기까지 등록센서스로의 이행에 따른 자료원, 개념, 정의, 분류방식 등의 변동을 문서화하는 작업이다. 특히 원시자료이용자들을 위해 통계생산과정을 문서로 남기고, 이를 공유하는 작업은 필수적이라 생각된다. 자료원의 변동이 있는 한, 기존의 전통적 센서스와 시계열적 일관성을 완벽하게 유지한다는 것은 불가능하다. 다만 변동의 근거와 과정 및 결과 등에 대한 성의 있는 분석을 남기고 이를 공유하는 과정에서 시계열적 일관성에 대한 이해가 높아질 것이다.

〈부표 6-2〉 2005년도 인구주택총조사 전수항목별 통계표와 등록센서스 예상표

인구편(전수)		등록센서스예상표
1표.	행정구역 및 총조사 실시년도별 인구, 가구 및 주택	전수▷자료원변동/주민등록상의 인구/유의점 제시
2표.	행정구역별 인구, 가구 및 주택	전수▷자료원 변동/주민등록상의 인구/유의점 제시
3표.	연령 및 성별 인구	전수▷주민등록상의 연령명시/유의점 제시
4표.	성, 연령 및 교육정도별 인구(6세 이상)	표본: 교육정도▷주민등록상의 연령명시/교육정도 표본조사근거
5표.	성, 연령 및 종교별 인구	표본: 종교
6표.	성, 연령 및 가족관계별 남북이산가족 인구	표본: 남북이산가족
7표.	성, 연령 및 출생지별 남북이산가족 인구	표본: 출생지, 남북이산가족
8표.	연령 및 성, 혼인상태별 인구(15세 이상)	표본: 혼인상태
9표.	성, 연령 및 가구주와의 관계별 인구	전수▷가구주와의 관계를 세대주와관계로 변동/유의점 제시
10표.	가구주와의 관계, 성 및 혼인상태별 인구	표본: 혼인상태
11표.	성, 혼인상태, 연령 및 세대구성별 인구	표본: 혼인상태
12표.	성, 혼인상태, 연령 및 교육정도별 인구(15세 이상)	표본: 혼인상태/교육정도
13표.	성 및 연령별 외국인	표본: 외국인
14표.	국적, 성 및 연령별 외국인	표본: 외국인/국적
15표.	성 및 국적별 외국인	표본: 국적
16표.	국적, 성 및 직업(대분류)별 외국인	표본: 직업
17표.	국적, 성 및 교육정도별 외국인	표본: 교육정도

가구편(전수)		등록센서스예상표
1표.	행정구역, 가구형태별 가구 및 가구원	전수▷세대주와관계를 가공/유의점 제시
2표.	행정구역, 세대구성별 가구 및 가구원	전수▷세대주와관계를 가공/유의점 제시
3표.	가구주의 성, 연령 및 혼인상태별 가구(일반가구)	표본: 혼인상태▷가구주에서 세대주로의 변동/유의점제시
4표.	가구주의 성, 연령 및 세대구성별 가구(일반가구)	전수▷세대주/세대주와 관계가공/유의점 제시
5표.	세대구성 및 가구주의 교육정도별 가구(일반가구)	표본: 교육정도
6표.	세대구성 및 가구원수별 가구(일반가구)	전수▷세대주와관계가공/세대구성원/유의점제시
7표.	가구의 크기 및 핵수별 가구(일반가구)	전수▷세대주와관계를 가공/유의점 제시
8표.	가구주의 성, 연령 및 거처의 종류별 가구(일반가구)	전수▷건축물대장의 거처종류/개념변동/유의점제시
9표.	거처의 종류 및 점유형태별 가구(일반가구)	표본: 점유형태▷건축물대장은 소유여부 파악가능
10표.	가구주의 연령 및 점유형태별 가구(일반가구)	표본: 점유형태▷세대주/건축물대장의 소유자현황/개념변동/유의점제시
11표.	거처의 종류 및 주거시설 형태별 가구(일반가구)	표본:주거시설형태
12표.	거처의 종류, 가구원규모 및 사용방수별 가구(일반가구)	전수▷건축물대장의 거처종류/세대구성원/개념변동/유의점제시
13표.	가구원 규모 및 방, 거실 및 식당수별 가구(일반가구)	전수▷건축물대장의 총방수/세대구성원/개념변동/유의점제시
14표.	세대구성 및 거처의 종류, 점유형태별 가구(일반가구)	표본: 점유형태▷세대주와관계가공/유의점제시
15표.	가구주의 연령 및 가구원수별 가구(일반가구)	전수▷세대주/세대구성원/개념변동/유의점제시
16표.	타지 주택소유 및 거처의 점유형태별 가구(일반가구)	전수▷건축물대장의 소유자현황/개념변동/유의점제시
17표.	거처의 종류 및 난방시설별 가구(일반가구)	표본: 난방시설
18표.	거주층별 가구(일반가구)	표본: 거주층
19표.	거처의 종류 및 사용방수, 주거면적별 가구(일반가구)	전수▷건축물대장의 거처종류, 연면적/개념변동/유의점제시

주택편(전수)		등록센서스예상표
1표.	주택의 종류별 주택	전수
2표.	거처의 종류 및 거처, 가구, 가구원	전수▷건축물대장의 거처종류
3표.	주택의 종류, 연건평 및 거주가구수별 주택	전수▷주택종류
4표.	주택의 종류, 연건평 및 거주인구수별 주택	전수

5표.	주택의 종류, 연건평 및 총방수별 주택	전수
6표.	주택의 종류, 연건평 및 건축년도별 주택	전수
7표.	연건평 및 편익시설수별 단독주택	표본: 편익시설수
8표.	연건평 및 대지면적별 단독주택	전수
9표.	연건평 및 거주가구수별 단독주택	표본: 거주가구수
10표.	주택의 종류 및 사유별 빈집	표본: 사유별 빈집
11표.	주택의 종류 및 기간별 빈집	표본: 기간별 빈집

〈표 6-10〉 2005년도 인구주택총조사 표본항목별 통계표와 등록센서스 예상표

인구편(전수)		등록센서스예상표
1표.	행정구역 및 총조사 실시년도별 인구, 가구 및 주택	전수▷자료원 변동/주민등록상의 인구/ 유의점 제시
2표.	행정구역별 인구, 가구 및 주택	전수▷자료원 변동/주민등록상의 인구/ 유의점 제시
3표.	연령 및 성별 인구	전수▷주민등록상의 연령명시/유의점 제시
4표.	성, 연령 및 교육정도별 인구(6세 이상)	표본: 교육정도▷주민등록상의 연령명 시/교육정도 표본조사근거
5표.	성, 연령 및 종교별 인구	표본: 종교
6표.	성, 연령 및 가족관계별 남북이산가족 인구	표본: 남북이산가족
7표.	성, 연령 및 출생지별 남북이산가족 인구	표본: 출생지, 남북이산가족
8표.	연령 및 성, 혼인상태별 인구(15세 이상)	표본: 혼인상태
9표.	성, 연령 및 가구주와의 관계별 인구	전수▷가구주와 관계를 세대주와관계로 변동/유의점 제시
10표.	가구주와의 관계, 성 및 혼인상태별 인구	표본: 혼인상태
11표.	성, 혼인상태, 연령 및 세대구성별 인구	표본: 혼인상태
12표.	성, 혼인상태, 연령 및 교육정도별 인구(15세 이상)	표본: 혼인상태/교육정도
13표.	성 및 연령별 외국인	표본: 외국인
14표.	국적, 성 및 연령별 외국인	표본: 외국인/국적
15표.	성 및 국적별 외국인	표본: 국적
16표.	국적, 성 및 직업(대분류)별 외국인	표본: 직업
17표.	국적, 성 및 교육정도별 외국인	표본: 교육정도

고령자(표본)		등록센서스예상표
1표.	행정구역 및 연령, 성별 고령자	전수 > 연령/개념변동/유의접체시
2표.	연령 및 성, 교육정도별 고령자	표본: 교육정도
3표.	연령 및 성, 혼인상태별 고령자	표본: 혼인상태
4표.	성, 연령 및 가구주와의 관계별 고령자	전수 > 연령/세대주관계/개념변동/유의접체시
5표.	세대구성 및 성, 연령별 고령자	전수
6표.	성, 연령 및 거처의 종류별 고령자	전수
7표.	연령 및 성, 거처의 점유형태별 고령자	표본: 점유형태
8표.	세대구성 및 성, 거처의 종류별 고령자(65세이상)	전수: 거처종류
9표.	세대구성 및 성, 거처의 점유형태별 고령자(65세이상)	표본: 점유형태
10표.	거처의 점유형태 및 성, 거처의 종류별 고령자(65세이상)	표본
11표.	경제활동상태 및 성, 연령별 고령자	표본
12표.	성, 경제활동상태 및 세대구성별 고령자(65세이상)	표본
13표.	산업 및 성, 종사상 지위별 고령자(65세이상)	표본
14표.	직업 및 성, 종사상 지위별 고령자(65세이상)	표본
15표.	행정구역 및 생활비원천별 고령자(65세이상)	표본
16표.	성, 연령 및 생활비원천별 고령자	표본
17표.	성, 교육정도 및 생활비원천별 고령자(65세이상)	표본
18표.	성, 세대구성 및 생활비원천별 고령자(65세이상)	표본
19표.	세대구성 및 거처의 종류별 고령자 거주가구(65세이상, 일반가구)	전수 > 세대주와 관계가함/건축물대장 거처종류
20표.	세대구성 및 거처의 점유형태별 고령자 거주가구(65세이상, 일반가구)	표본: 점유형태
21표.	거처의 점유형태 및 거처의 종류별 고령자 거주가구(65세이상, 일반가구)	표본: 점유형태
인구이동(표본)		등록센서스예상표
1표.	현거주지, 성, 연령 및 5년 전 거주지 유형별 인구(5세이상)	전수 > 5년전거주지 개념변동
2표.	현거주지, 연령 및 5년 전 거주지별 이동인구(5세이상)	전수 > 5년전거주지 개념변동
3표.	현거주지, 교육정도 및 5년 전 거주지별 이동인구(6세이상)	표본: 교육정도
4표.	현거주지, 성, 혼인상태 및 5년 전 거주지별 이동인구(15세이상)	표본: 혼인상태
5표.	현거주지, 성, 경제활동상태, 종사상 지위 및 5년 전 거주지별 이동인구(15세이상)	표본: 경제활동상태/종사상지위
6표.	현거주지, 산업 및 5년 전 거주지별 이동인구(15세이상)	표본: 산업

7표.	현거주지, 성, 직업 및 5년 전 거주지별 이동인구(15세이상)	표본: 직업
8표.	현거주지, 점유형태 및 5년 전 거주지별 이동인구(5세이상)	전수>건축물등록대장에서 점유형태
9표.	현거주지, 세대구성 및 5년 전 거주지별 이동인구(5세이상)	전수>세대주와 관계가용
경제활동(산업·직업)		등록센서스예상표
1표.	산업(소분류), 성 및 행정구역별 취업자(15세이상)	표본
1-1표.	산업(중분류) 및 시·군·구별 취업자(15세이상)(서울~울산)	표본
1-2표.	산업(중분류) 및 시·군·구별 취업자(15세이상)(경기~제주)	표본
2표.	직업(소분류), 성 및 행정구역별 취업자(15세이상)	표본
2-1표.	직업(중분류) 및 시·군·구별 취업자(15세이상)(서울~울산)	표본
2-2표.	직업(중분류) 및 시·군·구별 취업자(15세이상)(경기~제주)	표본
3표.	성, 산업(중분류) 및 직업(중분류)별 취업자(15세이상)	표본
3-1표.	성,산업(대분류), 직업(대분류) 및 시·도별 취업자(15세이상)	표본
4표.	성, 산업(중분류) 및 연령별 취업자(15세이상)	표본
5표.	성, 직업(중분류) 및 연령별 취업자(15세이상)	표본
6표.	산업(중분류) 및 성, 종사상 지위별 취업자(15세이상)	표본
7표.	직업(중분류) 및 성, 종사상 지위별 취업자(15세이상)	표본
8표.	성, 산업(중분류) 및 근로장소별 취업자(15세이상)	표본
9표.	성, 직업(중분류) 및 근로장소별 취업자(15세이상)	표본
경제활동(취업상태)		등록센서스예상표
1표.	성, 연령 및 경제활동상태별 인구(15세이상)	표본
2표.	성, 연령, 혼인상태 및 경제활동상태별 인구(15세이상)	표본
3표.	성, 교육정도 및 경제활동상태별 인구(15세이상)	표본
4표.	성, 연령 및 산업별 취업자(15세이상)	표본
5표.	성, 연령 및 직업별 취업자(15세이상)	표본
6표.	성, 직업 및 산업별 취업자(15세이상)	표본
7표.	성, 교육정도 및 산업별 취업자(15세이상)	표본
8표.	성, 교육정도 및 직업별 취업자(15세이상)	표본
9표.	현거주지, 근무지 및 산업별 취업자(15세이상)	표본

10표.	현거주지, 근무지 및 직업별 취업자(15세이상)	표본
11표.	성, 교육정도 및 종사상 지위별 취업자(15세이상)	표본
12표.	성, 연령, 혼인상태 및 종사상 지위별 취업자(15세이상)	표본
13표.	성, 종사상 지위 및 근로장소별 취업자(15세이상)	표본
14표.	성, 연령 및 근로장소별 취업자(15세이상)	표본
15표.	성, 교육정도 및 근로장소별 취업자(15세이상)	표본
16표.	세대구성 및 경제활동상태별 여성인구(15세이상)	표본
17표.	가구원 규모 및 취업자수별 가구(일반가구)	표본
18표.	가구주 부부의 경제활동상태 및 산업별 가구(일반가구)	표본
19표.	가구주 부부의 경제활동상태 및 직업별 가구(일반가구)	표본
20표.	거처의 종류 및 가구주의 경제활동상태, 종사상 지위별 가구(일반가구)	표본
여성·아동		등록센서스예상표
1표.	연령 및 출생자녀수별 기혼여성	표본
2표.	교육정도 및 출생자녀수별 기혼여성	표본
3표.	연령, 경제활동상태 및 출생자녀수별 기혼여성	표본
4표.	동거자녀수 및 출생자녀수별 기혼여성	표본
5표.	사망자녀수 및 출생자녀수별 기혼여성	표본
6표.	연령, 경제활동상태 및 동거자녀수별 기혼여성	표본
7표.	연령 및 추가계획자녀수별 기혼여성	표본
8표.	출생자녀수 및 추가계획자녀별 기혼여성	표본
9표.	교육정도 및 추가계획자녀수별 기혼여성	표본
10표.	경제활동상태 및 추가계획자녀수별 기혼여성	표본
11표.	연령 및 초혼연령별 기혼여성	표본
12표.	출생자녀수 및 초혼연령별 기혼여성	표본
13표.	교육정도 및 초혼연령별 기혼여성	표본
14표.	경제활동상태 및 초혼연령별 기혼여성	표본
15표.	어머니의 연령, 동거여부 및 보육상태별 아동인구	표본
16표.	어머니의 교육정도 및 보육상태별 아동인구	표본
17표.	어머니의 경제활동상태 및 보육상태별 아동인구	표본
18표.	세대구성 및 보육상태별 아동인구	표본

주거(주거실태)		등록센서스예상표
1표.	행정구역 및 거주기간별 가구(일반가구)	전수
2표.	거처의 종류, 점유형태 및 거주기간별 가구(일반가구)	표본: 점유형태
3표.	가구주의 성, 연령 및 거주기간별 가구(일반가구)	전수
4표.	가구주의 성, 교육정도 및 거주기간별 가구(일반가구)	표본: 교육정도
5표.	가구주의 성, 경제활동상태, 산업 및 거주기간별 가구(일반가구)	표본: 경제활동상태, 산업
6표.	가구주의 성, 경제활동상태, 직업 및 거주기간별 가구(일반가구)	표본: 경제활동상태, 직업
7표.	행정구역 및 자동차 보유대수별 가구(일반가구)	표본: 자동차보유대수
8표.	거처의 종류, 점유형태 및 자동차 보유대수별 가구(일반가구)	표본: 자동차보유대수
9표.	가구주의 성, 연령 및 자동차보유대수별 가구(일반가구)	표본: 자동차보유대수
10표.	가구주의 경제활동상태, 산업 및 자동차 보유대수별 가구(일반가구)	표본: 자동차보유대수
11표.	가구주의 경제활동상태, 직업 및 자동차 보유대수별 가구(일반가구)	표본: 자동차보유대수
12표.	세대구성 및 자동차 보유대수별 가구(일반가구)	표본: 자동차보유대수
13표.	가구원 규모 및 자동차 보유대수별 가구(일반가구)	표본: 자동차보유대수
14표.	행정구역 및 주차시설별 가구(일반가구)	표본: 주차시설
15표.	거처의 종류 및 주차시설별 가구(일반가구)	표본: 주차시설
16표.	점유형태 및 주차시설별 가구(일반가구)	표본: 주차시설
17표.	가구주의 성, 연령 및 주차시설별 가구(일반가구)	표본: 주차시설
주거(임차료)		등록센서스예상표
1표.	거처의 종류, 사용방수 및 전세금 규모별 가구(일반가구)	표본: 전세금 규모
2표.	거처의 종류, 가구주의 연령 및 전세금 규모별 가구(일반가구)	표본: 전세금 규모
3표.	거처의 종류, 가구원수 및 전세금 규모별 가구(일반가구)	표본: 전세금 규모
4표.	가구주의 연령, 혼인상태 및 전세금 규모별 가구(일반가구)	표본: 전세금 규모
5표.	거처의 종류, 월세금 및 보증금 규모별 가구(일반가구)	표본: 보증금 규모
6표.	사용방수, 월세금 및 보증금 규모별 가구(일반가구)	표본: 보증금 규모
7표.	가구주의 연령, 월세금 및 보증금 규모별 가구(일반가구)	표본: 보증금 규모
8표.	가구원수, 월세금 및 보증금 규모별 가구(일반가구)	표본: 보증금 규모
9표.	가구주의 혼인상태, 월세금 및 보증금 규모별 가구(일반가구)	표본: 보증금 규모
10표.	거처의 종류, 사용방수 및 사글세 규모별 가구(일반가구)	표본: 보증금 규모

11표.	거처의 종류, 가구주의 연령 및 사글세 규모별 가구(일반가구)	표본: 보증금 규모
12표.	거처의 종류, 가구원수 및 사글세 규모별 가구(일반가구)	표본: 보증금 규모
13표.	가구주의 연령, 혼인상태 및 사글세 규모별 가구(일반가구)	표본: 보증금 규모
1인 가구		등록센서스예상표
1표.	행정구역 및 성, 거처의 종류별 1인가구	전수
2표.	성, 연령 및 교육정도별 1인가구	표본: 교육정도
3표.	성, 연령 및 혼인상태별 1인가구	표본: 혼인상태
4표.	성, 연령 및 경제활동상태별 1인가구(15세이상)	표본: 경제활동상태
5표.	성, 연령 및 산업별 1인가구(15세이상)	표본: 산업
6표.	성, 연령 및 직업별 1인가구(15세이상)	표본: 직업
7표.	성, 연령 및 거주기간별 1인가구	전수
8표.	성, 연령 및 거처의 종류별 1인가구	전수
9표.	거처의 종류, 사용방수 및 점유형태별 1인가구	표본: 점유형태
10표.	거처의 종류 및 주거시설 형태별 1인가구	전수
11표.	거처의 종류 및 성, 점유형태별 고령자 1인가구(65세이상)	표본: 점유형태
12표.	성, 연령 및 활동제약유형별 1인가구(5세이상)	표본: 활동제약
13표.	거처의 종류 및 성, 점유형태별 활동제약자 1인가구(65세이상)	표본: 활동제약
14표.	현거주지, 성, 연령 및 5년전 거주지별 1인가구(6세이상)	전수
통근·통학		등록센서스예상표
1표.	행정구역 및 성별 통근 · 통학인구(12세이상)	표본: 통근통학
2표.	행정구역, 성 및 통근 · 통학 유형별 인구(12세이상)	표본: 통근통학
3표.	연령 및 이용교통수단별 통근 · 통학인구(12세이상)	표본: 통근통학
4표.	소요시간 및 이용교통수단별 통근 · 통학인구(12세이상)	표본: 통근통학
5표.	가구의 점유형태 및 이용교통수단별 통근 · 통학인구(12세이상)	표본: 통근통학
6-1표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구(서울특별시)(12세이상)	표본: 통근통학
6-2표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구(부산광역시)(12세이상)	표본: 통근통학
6-3표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구(대구광역시)(12세이상)	표본: 통근통학
6-4표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구(인천광역시)(12세이상)	표본: 통근통학
6-5표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구	표본: 통근통학

	통학인구(광주광역시)(12세이상)	
6-6표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구(대전광역시)(12세이상)	표본: 통근통학
6-7표.	성, 통근 · 통학지 및 이용교통수단별 통근 · 통학인구(울산광역시)(12세이상)	표본: 통근통학
7-1표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(서울, 인천, 경기)(12세이상)	표본: 통근통학
7-2표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(부산, 울산, 경남)(12세이상)	표본: 통근통학
7-3표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(대구, 경북)(12세이상)	표본: 통근통학
7-4표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(광주, 전남)(12세이상)	표본: 통근통학
7-5표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(대전, 충북, 충남)(12세이상)	표본: 통근통학
7-6표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(강원)(12세이상)	표본: 통근통학
7-7표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(전북)(12세이상)	표본: 통근통학
7-8표.	성, 현거주지 및 통근 · 통학지별 통근 · 통학 인구(제주)(12세이상)	표본: 통근통학
8표.	산업 및 이용교통수단별 통근인구(15세이상)	표본: 통근통학
9표.	성, 직업 및 이용교통수단별 통근인구(15세이상)	표본: 통근통학
10표.	성, 종사상 지위 및 이용교통수단별 통근인구(15세이상)	표본: 통근통학
11표.	성, 각급학교 및 이용교통수단별 통학인구(15세이상)	표본: 통근통학
12표.	소요시간 및 각급학교별 통학인구(12세이상)	표본: 통근통학
13표.	행정구역 및 성별 상주(야간) · 주간인구	표본: 통근통학
14표.	행정구역, 연령 및 성별 유입 · 유출인구	표본: 통근통학
활동제약		등록센서스예상표
1표.	행정구역 및 육체적 · 정신적 제약유형별 인구(5세이상)	표본
2표.	성, 연령 및 육체적 · 정신적 제약유형별 인구(5세이상)	표본
3표.	성, 교육정도 및 육체적 · 정신적 제약유형별 인구(5세이상)	표본
4표.	성, 혼인상태 및 육체적 · 정신적 제약유형별 인구(15세이상)	표본
5표.	성, 세대구성 및 육체적 · 정신적 제약유형별 인구(5세이상)	표본
6표.	성, 경제활동상태 및 육체적 · 정신적 제약유형별 인구(15세이상)	표본
7표.	성, 산업 및 육체적 · 정신적 제약유형별 인구(15세이상)	표본
8표.	성, 직업 및 육체적 · 정신적 제약유형별 인구(15세이상)	표본
9표.	행정구역 및 일상 · 사회 활동제약유형별 인구(5세이상)	표본

10표.	성, 연령 및 일상 · 사회 활동제약유형별 인구(5세이상)	표본
11표.	성, 교육정도 및 일상 · 사회 활동제약유형별 인구(5세이상)	표본
12표.	성, 혼인상태 및 일상 · 사회 활동제약유형별 인구(15세이상)	표본
13표.	성, 세대구성 및 일상 · 사회 활동제약유형별 인구(5세이상)	표본
14표.	성, 경제활동 및 일상 · 사회 활동제약유형별 인구(15세이상)	표본
15표.	성, 산업 및 일상 · 사회 활동제약유형별 인구(15세이상)	표본
16표.	성, 직업 및 일상 · 사회 활동제약유형별 인구(15세이상)	표본
17표.	성, 일상 · 사회 활동제약 및 육체적 · 정신적 제약유형별 인구(5세이상)	표본
18표.	가구원규모 및 활동 제약자 거주 가구(일반가구)	표본
19표.	가구의 성, 경제활동상태 및 활동 제약자 거주 가구(일반가구)	표본
20표.	거처의 종류 및 점유형태별 육체적 · 정신적 제약자 거주 가구(일반가구)	표본
21표.	거처의 종류 및 점유형태별 일상 · 사회 활동제약자 거주 가구(일반가구)	표본

〈부표 6-1〉 인구주택총조사 조사 및 공표항목

인구에 관한 항목	1955	1960	1966	1970	1975	1980	1985	1990	1995	2000	2005
01. 성명		●	●	●□	●□	●□	●□	●□	●□	●□	●□
02. 본관							●■	■		●■	
03. 가구주와의 관계		●	●	●■	●■	●■	●■	●■	●■	●■	●■
04. 성별		●	●	●■	●■	●■	●■	●■	●■	●■	●■
05. 연령(공표: 만나이)		●	●	●■	●■	●■	●■	●■	●■	●■	●■
06. 교육정도		●	●	●■	●■	●■	●■	●■	●■	●■	●■
07. 혼인상태		●	●	●■	●■	●■	●■	●■	●■	●■	●■
08. 종교(공표: 구시군)							●■		●■		●■
09. 출생지		본적 또는 국적		○■		○■	●■	●■	○■	●■	
10. 1년전 거주지		전입자 구분				○■	●■	○■		○■	
11. 5년전 거주지				○■	○■	○■	●■	○■	○■	○■	○■
12. 통근학 여부						○■		●■	○■	○■	○■
13. 통근학지						○■		●■	○■	○■	○■
14. 이용수단						○■		●■	○■	○■	○■
15. 소요시간(출발도착시간)								●■	○■	○■	○■
16. 초혼연령					○■	○■		○■			○■
17. 총출생아수		●	○	○■	○■	○■	●■	○■		○■	○■
18. 생존아수				○■	○■		●■	○■			
19. 사망자녀수					○■	○■	●■	○■			○■
20. 경제활동상태		●	○	○■	○■	○■	●	○■	○■	○■	○■
21. 종사상의 지위		●	○	○■	○■	○■		○■	○■	○■	○■
22. 산업		●	○	○	○■	○■	●■	○■	○■	○■	○■
23. 직업		●	○	○■	○	○■	●■	○■	○■	○■	○■
24. 개인소득					○						
25. 심신장애						●					○■
26. 전공학과						○				○	
27. 아동보육실태						○				○	○
28. 고령자 생활비 원천										○	○■
29. 컴퓨터 활용 상태(2005년)											○
30. 인터넷 활용 상태(2005년)											○
31. 개인휴대용 통신기기(2005년)											○
32. 현직업 근무 연수(2005년)											○
33. 자녀거주장소(2005년)											○

34. 거동불편 여부(2005년)											○
35. 남북이산가족(2005년)											○
36. 근로장소(2005년)											○
37. 총 출생아 수(2005년)											○
38. 추가 계획 자녀 수(2005년)											○
가구·주택에 관한 항목	1955	1960	1966	1970	1975	1980	1985	1990	1995	2000	2005
39. 거처의 종류 및 건물 층수		●		●■	●■	●■	●■	●■	●■	●■	●■
40. 가구구분	●	●	●	●■	●■	●■	●■	●■	●■	●■	●■
41. 점유형태	●	●		●■	●■	●■	●■	●■	●■	●■	●■
42. 사용방수				■	■	●■	●■	●■	●■	●■	●■
43. 식수종류(상수도)						●■		●■		○	
44. 취사연료						●■	●■	●■		○■	
45. 난방연료						●■					
46. 난방시설						●	●	●		○	●■
47. 문화시설 및 가재(정보통신기기)				●■	○■	●■	●■	●■		○	
48. 주거시설형태		●		○		●■	●■	●■	●■	●■	●■
49. 임차료							●	○	○■	○■	○■
50. 가구소득								○			
51. 외벽재료		●		○■	○■	●■		●■			
52. 지붕재료				●■	■	●■	●	●■	●		
53. 연건평				●■		●■	●■	●■	●■	●■	●■
54. 대지면적						●■		●■	●■	●■	●■
55. 건축년도				●■	●■	●■		●■	●■	●■	●■
56. 총방수				●■	●■	●■	●■	●■	●■	●■	●■
57. 동거가구수		●		●■	●■	●■	●■	●■			
58. 거주가능가구수							●	●			
59. 편익시설수							●	●	●	●	●
60. 조명시설				○							
61. 거주기간									○	○	○■
62. 거주층											●■
63. 자동차보유대수										○■	○■
64. 주차시설										○	○■
65. 주인가구 및 주택소유여부										○	●■

●: 전수 ○: 표본 ■: 공표

□: 전산미입력제6장

제7장 센서스의 표본조사 모집단 기능과 기존 마이크로데이터 활용 대체방안

제1절 머리말

나라별로 센서스를 실시하는 양상은 매우 다르다. 각 나라별 문화나 여건에 따라 센서스 방법이나 주기, 전수조사와 표본조사 병행 실시 여부 등이 서로 다른 편이다. 다음의 <표 7-1>은 OECD 주요 국가들의 센서스 실시현황을 정리한 표인데, 우리나라를 비롯하여 미국, 네덜란드, 독일, 싱가포르, 스위스, 벨기에, 오스트리아, 캐나다, 멕시코 등 여러 나라에서 전수조사와 표본조사를 병행하고 있음을 알 수 있다.

<표 7-1> 나라별 센서스 실시 현황

센서스 방법		조사주기 및 국가 명
전통 조사방법 (현장조사)	전수조사	· 5년주기 : 호주 , 뉴질랜드, 일본 · 10년주기 : 영국 , 이탈리아, 포르투갈, 슬로바키아, 터키, 폴란드, 체코, 그리스, 헝가리, 스페인 등
	전수조사 + 표본조사	· 5년주기 : 캐나다 , 한국 · 10년주기 : 멕시코
새로운 조사방법		
① 등록센서스 (행정자료 활용)		· 10년주기 : 덴마크 , 아이슬란드, 노르웨이 · 5년주기 : 핀란드 , 스웨덴
② 등록센서스+표본조사 (보완)		· 10년주기 : 네덜란드 , 독일 , 싱가폴 , 스위스, 벨기에, 오스트리아 등
③순환센서스 (Rolling Census)		· 5년주기 : 프랑스
④ 현장조사+순환센서스 (전수조사) (표본조사)		· 10년주기 : 미국

한편, 센서스에서 표본조사를 병행하든 그렇지 않든 상관없이 모든 나라에서는 센서스 자료를 기초로 하는 다양한 표본조사들을 실시하게 된다. 그러

므로 사실상 각 나라의 통계청은 전수조사와 표본조사를 동시에 수행한다고 해도 무방하다. 전수조사와 표본조사는 상호 보완적인 속성을 지니며 서로 밀접한 관련을 맺고 있다고 할 것이다.

이 장에서는 우리나라의 인구주택총조사가 기존의 전통적인 방식을 벗어나 등록센서스로 전환한다고 할 때 표본조사의 측면에서 마땅히 짚어보아야 할 요소들을 종합적으로 검토하고자 한다. 그러기 위해서는 가장 우선적으로 기존의 인구주택총조사 결과가 다른 표본조사 추출틀로서 어떻게 활용되었는지를 전반적으로 살펴보는 작업이 우선되는 것이 필요하다. 다음으로는 다른 나라들의 사례들을 살펴봄으로 새로운 시사점이나 고려해야 할 점들을 찾는 것도 의미가 있다. 이를 기초로 새롭게 등록센서스로 전환된다고 할 때 바람직한 추출틀 마련방안을 제시하고자 한다.

제2절 기존 총조사의 표본조사 추출틀로서의 역할

1. 센서스 자료를 추출틀로 활용하는 표본조사

우리나라에서 작성, 공표되는 정부승인통계는 2009년 8월 1일 현재 884종에 이른다(<그림 7-1>참조). 그중 조사통계는 368종으로 전체의 약 41.6%를 차지하고 있다. 조사통계를 다시 세분하면 전수조사와 표본조사로 구분할 수 있는데, 전수조사는 모집단에 포함되는 모든 개체를 전부 조사하여 작성되는 통계로서 인구주택총조사를 비롯한 각종 총조사가 대표적이다. 반면 표본조사는 전국 모집단을 대표하도록 표본을 설계한 후 추출된 표본에 대해 조사하여 모수를 추정하는 통계이다. 우리나라의 대표적인 표본조사로는 경제활동인구조사, 가계조사, 사회통계조사, 노동패널조사, 국민건강영양조사, 광공업·도소매업·서비스업 동태조사 등을 들 수 있다. 표본조사의 대상이 되는 단위는 개인, 가구, 사업체 등 다양한 경우가 있다.

» 통계부문별 현황
(2009. 08. 01. 현재) (단위: 종, %)

통계부문	작성통계수		작성방법별		
	통계수	구성비	조사	보고	가공
계	884	100.0	368	445	71
인 구	28	3.2	4	21	3
고 용 · 임 금	41	4.6	35	6	-
물 가 · 가 계 소 비	16	1.8	16	-	-
보 건 · 사 회 · 복 지	138	15.6	87	48	3
환 경	26	2.9	9	16	1
농 립 · 수 산	63	7.1	45	17	1
광 공 업 · 에 너 지	31	3.5	19	8	4
건 설 · 주 택 · 토 지	42	4.8	15	22	5
교 통 · 정 보 통 신	54	6.1	29	24	1
도 소 매 · 서 비 스	20	2.3	18	2	-
경 기 · 기 업 경 영	63	7.1	45	6	12
국 민 계 정 · 지 역 계 정	21	2.4	-	-	21
재 정 · 금 융	24	2.7	4	20	-
무 역 · 외 환 · 국 제 수 지	11	1.2	4	4	3
교 육 · 문 화 · 과 학	46	5.2	35	10	1
기 타	260	29.4	3	241	16

<그림 7-1> 국가승인통계 부문별 현황

표본조사에 의해 작성되는 통계들은 전국의 모집단을 대표할 수 있도록 과학적으로 설계된 표본설계를 거쳐 조사가 진행된다. 표본설계를 위해 가장 기본적인 작업은 먼저 모집단의 현황을 나타내는 좋은 추출틀(sampling

frame)을 마련하는 작업이라고 할 수 있다. 우리나라에서 일반적으로 추출틀 작성을 위해 가장 널리 쓰이는 자료가 바로 인구주택총조사 자료이다. 위의 <그림 7-1>에 나타난 바와 같이 각 부문별로 다양한 조사통계가 작성되고 있는데, 부문별로 인구주택총조사 자료를 추출틀로 활용하는 대표적인 조사의 예가 다음에 열거되어 있다.

<표 7-2> 대표적인 표본조사통계와 사용 추출틀

분야	표본조사명	사용 추출틀
고용	경제활동인구조사	10% 표본조사구
노동	사업체임금근로시간 노동패널조사	10% 표본조사구
가계	가계동향조사	10% 표본조사구
보건	국민건강영양조사	전수조사 조사구
사회	사회조사	10% 표본조사구
농림어업	경제조사 농촌생활지표조사	10% 표본조사구 동읍면별 자료
복지	한국복지패널조사	10% 표본조사구
부동산	주택가격동향조사	전수조사 조사구 + 아파트단지명부
정보통신	정보화통계조사	10% 표본조사구
관광	국민관광실태조사	전수조사 조사구

동일한 인구주택총조사 결과를 추출틀로 사용한다고 해도 조사기관이나 조사의 성격에 따라 구체적인 추출틀의 내용은 다를 수 있다. 1차추출단위를 동읍면으로 할 수도 있고 인구주택조사구를 할 수도 있으며, 추출단위에 대해 요구하는 보조변수들이 서로 다를 수가 있기 때문이다. <표 7-2>에 소개된 우리나라 대표적인 표본조사들을 보면 통계청에서 수행하는 조사도 있고 통계청이 아닌 다른 정부부처 또는 민간기관에서 수행하는 조사도 있다. 통계청이 담당하는 조사들의 경우 당연히 인구주택총조사의 10% 표본조사 자료를 추출틀로 활용하게 된다. 그러나 통계청이 아닌 외부 기관에서 수행하는 표본조사의 경우 양상은 조금 다른데, 인구주택총조사 자료라고 해도 경우에 따라 서로 다른 자료를 추출틀로 활용하게 된다. 즉, 10% 표본조사구를 활용하는 기관도 있고, 기본적인 인구, 주택 변수들만 사용하는 경우에는 전수조사의 조사구 자료를 활용하는 경우도 있으며, 아예 동읍면별 집계자료를 추출틀로 활용하는 예도 있다. 또한 <표 7-2>에 소개된 예에는 나와 있지 않지만 경우에 따라 통계청의 마이크로데이터서비스시스템인 KMDSS를 통해 제공되는 2%, 5% 데이터를 활용하는 경우도 있을 수 있다.

2. 경상조사 표본설계 사례

표본조사를 위한 기초자료로써 인구주택총조사 자료가 어떻게 활용되는지를 효과적으로 이해하기 위해 구체적인 표본설계 사례를 살펴보는 것이 도움이 될 것이다. 통계청에서 실시하는 여러 표본조사 중 가장 대표적인 조사로 경제활동인구조사, 가계조사, 집세조사 등을 들 수 있다. 이 조사를 위한 표본설계를 경상조사 표본설계라고 하는데, 기본적으로 동일한 표본에서 세 조사를 실시할 수 있도록 하는 다목적 표본설계가 이루어졌다. 이 표본조사의 사례를 통해 구체적으로 인구주택총조사 자료가 표본설계를 위해 어떻게 활용되는지를 검토하기로 한다. 다른 많은 표본설계도 크게 볼 때 이 표본설계와 유사한 형태를 띠다고 볼 수 있다.

가. 모집단 및 추출틀

가구조사 모집단은 2005년 인구주택총조사의 일반조사구내 가구 및 가구원이다. 한편 추출틀은 모집단 분석이 가능한 인구주택총조사 10% 표본조사구 중 208개 섬조사구와 7,528개 기숙시설 및 특수사회시설 조사구를 제외한 26,505개 조사구만으로 구성되었다. 실제 표본설계가 2007년에 이루어졌으므로 모집단에 대한 대표성을 높이기 위해 2005년 총조사 이후부터 표본개편 무렵인 2006년 10월말까지 신축된 아파트조사구 506개를 추출틀에 추가하였다.

나. 추출단위

경상조사 표본설계는 기본적으로 이단계 추출방식으로 이루어진다. 1차추출단위는 인구주택조사구이며, 2차추출단위는 인구주택조사구 내의 가구가 된다. 그러므로 추출틀에는 1차추출단위인 조사구별로 요약된 보조정보들이 담겨 있으면 된다. 일단 표본조사구가 추출되고 나면, 해당 조사구 내 가구 정보를 파악하여 2차추출단위인 가구를 추출하는 방식을 띠게 된다.

다. 층화

집계 정보가 정보최종 추출단위인 가구를 추출하기까지 전체적인 표본추

출은 1차추출단위인 인구주택조사구에 대해 층화가 이루어졌다. 층화변수로는 통계의 공표단위가 우선적으로 고려된다. 일차로 지역변수(7개 특별시 및 광역시, 9개 도)가 고려되었고, 각 지역 내에서는 동부와 읍면부를 구분하는 변수를 사용하여 총 25개의 층으로 구분되었다.

라. 단계별 표본추출

1단계 추출은 추출틀 내의 조사구 중 각 층에 배분된 크기만큼의 표본조사구를 추출하는 단계이다. 같은 층에 속한 조사구라고 해도 조사구의 크기가 균일하지 않으므로 조사구 내의 가구수를 크기척도(measure of size:MOS)로 하는 크기비례확률추출법(probability proportional to size sampling: pps sampling)에 의해 조사구를 확률추출하였다.

표본조사구가 추출되고 나면 조사구요도와 가구명부를 작성한 후, 조사구 내 가구들을 5가구 기준으로 묶인 조사구역으로 분할한다. 조사구내 거처들을 구역화할 때 각 구역간 거처수는 다르게 되더라도 가구수는 균일하게 하며, 조사구 내에 주택신축이 가능한 공간(택지)이 있는 경우에는 공지의 면적을 균등하게 배분시켜 설정한다. 조사구별로 4개의 조사구역을 랜덤하게 추출한 후 조사구역 내 가구들을 다 조사하게 함으로써 결과적으로 20개씩의 표본가구를 추출하게 된다.

마. 토의

위에서 경상조사 표본설계 과정을 대략적으로 검토해보았는데, 사실상 표본설계 단계에서 가장 중요한 자료는 1차추출단위인 조사구별 정보라는 사실이 두드러지게 확인된다. 다시 말하자면, 인구주택총조사의 자료는 표본설계 과정에서 조사구별로 취합된 데이터 양식으로 활용되어지는 것이다. 그러므로 인구주택총조사가 표본조사를 위한 추출틀로서의 역할을 효과적으로 감당하도록 하려면 조사구를 어떻게 합리적으로 만들 것인가 하는데 논의를 집중하는 것이 마땅하다.

3. 통계청 제공 추출틀 자료

통계청에서 제공하는 추출틀에는 10% 표본조사 자료 전부 또는 일부를 제공하는 경우와 전수조사 자료의 전부 또는 일부를 제공하는 경우로 크게 구

분할 수 있다. 다음의 가와 나 의 절은 10% 표본조사 자료 활용과 관련하여 KMDSS에 올라 있는 2005년 인구주택총조사 마이크로데이터 제공 관련 지침을 그대로 옮겨 놓은 것이다. 한편 다의 절에서는 그 밖의 인구주택총조사 활용의 사례를 정리하였다.

가. 마이크로데이터 제공 지침

① 명부자료

수 록 항 목	제공가능여부	비 고
[가구명부]		
- 거처번호	○	통계작성기관에서 통 계청이 승인한 통계조사목 적으로 사용할 경우만 제공
- 가구번호	○	
- 주 소	○	
- 가구주 성명	○	
- 가구원수	○	
- 가구의 종류	○	
- 거처의 종류	○	
- 빈집 여부	○	
- 조사구 요도	○	

② 조사표수록 항목의 마이크로데이터

통계상품으로 개발된 1% 마이크로데이터(전국), 2% 마이크로데이터 A형(시도, 가중값 있음), B형(시군구, 가중값 없음), 5% 마이크로데이터(시군구, 가중값 있음)만 제공

※ 각 통계상품별 제공항목에는 차이가 있으므로 반드시 확인 후 구입할 것

③ 미공표 집계자료 제공 기준¹⁾

전수조사 항목의 경우

1) 조사표수록 항목 중 마이크로데이터로는 제공이 불가능한 항목이나, 집계자료로는 제공이 가능한 항목으로서 통계법 제31조 제2항, 제32조 제2항 및 통계법 제33조(비밀의 보호) 고려하였음

제공 가능 항목	다른 항목과의 결합 가능여부	결합 가능 항목명	비고
전수조사항목 전체 (행정구역은 읍면동이 최소단위)	○	모든 항목 (단, 외국인, 남북이산가족, 임차료 관련항목은 결합불가)	

※ 외국인, 남북이산가족, 임차료 관련 항목은 KOSIS, 보고서 등 기 제공 통계표만 제공함

- 제공되는 집계표의 각 셀의 최소값: 5인(가구)
- 전수 항목별 세부 집계 기준

항목	세부 기준
가구주와의관계	- 가구주를 별도로 표시하여 연령과 결합하는 경우 연령하한은 15세미만
성별	
만나이	- 가구주 관련인 경우 하한 15세미만, 상한 85세이상
교육정도	- 6세이상 인구만 집계
종교	- 종교종류는 09~98은 기타로 집계
남북이산가족	
혼인상태	- 15세이상 인구만 집계
가구구분	
사용방수	- 1, 2, ..., 5, 6개이상
- 사용방수	- 1, 2, ..., 5개이상
- 거실수	- 0, 1, 2개이상
- 식당수	- 0, 1, 2개이상
주거시설형태	
거주층	거주층 구분만 집계
난방시설	
집유형태	
주인가구 및 타지	
주택소유	
거치의종류	
단독주택종류	- 다가구단독으로 분리하는 경우 시군구 단위까지만 제공
연건평	- 7평미만, 9, 14, 19, 29, 39, 49, 69, 99, 99평이상
대지면적	- 10평미만, 14, 19, 29, 39, 49, 69, 99, 149, 199, 249, 299, 299평이상
총방수	- 1, 2, ..., 7, 8개이상
- 사용방수	- 1, 2, ..., 5개이상
- 거실수	- 0, 1, 2개이상
- 식당수	- 0, 1, 2개이상
건축연도	
편의시설수	- 세부 항목(부엌수, 화장실수, 출입구수) 각 1, 2, ..., 5개이상
가구크기 (거주인수)	- 1인, 2, ..., 9인, 10인이상가구
거주가구수	- 1가구, 2, ..., 5, 6가구이상

□ 표본조사 항목의 경우2)

제공 가능 항목	다른 항목과의 결합 가능여부	결합 가능 항목명	비고
10% 표본항목 전체 (행정구역은 시군구가 최소단위)	○	모든 항목 (단, 외국인, 남북이산가족, 입차료 관련 항목은 결합불가)	

- 제공되는 집계표의 각 셀의 최소값

· 인구는 행정구역에 대한 항목만을 기준으로 상대표준오차 50% 이하만 제공

(가구는 인구 행정구역별 상대표준오차 50% 이하 기준을 사용하고, 차후 가구 상대표준오차 개발하여 적용 예정)

· 행정구역 통합지역(권역)은 행정구역중 상대표준오차(50%이하) 값이 가장 큰 지역을 기준으로 함

· 통근/통학지, 5년전 거주지별 인구는 현거주지 기준으로 적용함

- 표본 항목별 세부 집계 기준

항목	세부기준
아동보육상태, 어머니 동거여부	- 일반가구 12세이하만 집계
5년전거주지	- 5세이상만 집계
활동제약	- 5세이상만 집계
통근통학여부, 통근통학장소, 이용교통수단	- 12세이상만 집계
통근통학소요시간	- 12세이상만 집계, 15분미만, 30, 45, 60, 90, 120, 120분이상
경제활동상태, 종사상의 지위	- 15세이상만 집계
산업, 직업	- 15세이상만 집계, 소분류(3자리)까지 제공
근로장소	- 15세이상만 집계
혼인년월	- 초혼연령으로 집계 15세미만, 5세간격, 50세이상
총출생아수	- 없음(0), 1, 2, ..., 9, 10명이상
추가계획자녀수	- 없음(0), 1, 2, 3, 4, 5명이상
고령자생활비원천	- 60세이상 집계 - 이 항목들은 연령 상한을 100세이상으로 조정 가능
거주기간	
자동차 보유대수	- 총 대수 0, 1, 2, 3대이상 - 세부 항목 0, 1, 2대이상
주차시설	
입차료	- 보고서 및 KOSIS 분류기준에 따라 제공

2) 상대표준오차(통계법 제27조 제3항 참조!) 및 통계법 제31조 제2항, 제32조 제2항 및 통계법 제33조(비밀의 보호) 고려

④ 기타사항

- 외국인, 남북이산가족, 임차료 관련 통계표는 기 제공 통계표만 제공
- 시도특성항목은 자료 요청시 매 건별 제공여부 검토
- 전수항목과 표본항목을 결합시키는 경우는 표본항목 기준에 준함

나. KMDSS 신청화면

다음의 <그림 7-2>는 인구주택조사구 가구주택사항에 대한 표본 중 전국의 1% 자료 요청을 위해 제공되는 웹 화면을 나타낸 것이다. 1% 자료에서는 많은 변수들 정보를 가구단위로 제공하나 구체적인 주소변수를 포함시키지 않음을 알 수 있다. 다시 말해, 표본조사가 이루어지는 대부분의 항목 변수들의 가구별 자료를 제공하지만 가구가 속한 지역이나 조사구 변수는 제공하지 않는다. 따라서 이 자료는 전국을 대표하는 자료일 뿐 시도나 시군구 단위의 세부적인 분석을 할 수는 없는 자료라고 할 수 있다.

순번	형태	항목	순번	형태	항목	순번	형태	항목
1	T	가구일련번호	2	C	가구구분	3	C	거주기간
4	H	방수(침실)	5	H	방수(침실미와의방(서재등))	6	H	거실수(대청마루)
7	H	식당(부엌미달린식당포함)	8	C	부엌형태	9	C	부엌사용형태
10	C	수도	11	C	화장실형태	12	C	화장실사용형태
13	C	목욕시설형태	14	C	목욕시설사용형태	15	C	거주층구분
16	H	거주층수	17	H	승용차(10인승이하)	18	H	승합차(11인승이상)
19	H	화물및기타자동차	20	C	자동차없음	21	C	주차시설1
22	C	주차시설2	23	C	난방시설	24	C	주거.영업구분
25	C	점유형태	26	C	주인가구	27	C	주택소유여부
28	C	거처종류	29	C	단독주택종류	30	H	거주층수
31	H	연건평	32	H	대지면적	33	H	방수
34	H	거실(대청마루)	35	H	식당(부엌미달린식당포함)	36	C	건축년도
37	H	부엌면적시설수	38	H	화장실면적시설수	39	H	독립된출입구면적시설수

<그림 7-2> 인구주택총조사 1% 가구주택 표본자료 제공 변수

다음의 <그림 7-3>은 각각 인구주택조사구 가구주택사항에 대한 전국 2% 표본 자료 요청을 위해 제공되는 웹 화면을 나타낸 것이다. 2% 데이터에서는 1%와는 달리 시도, 가중값 등의 변수가 추가적으로 제공되고 있는 반면, 몇 가지 변수들이 누락되어 있음을 확인할 수 있다. 따라서 이 자료를 가지고는 시도별 분석을 하는 것이 가능하다.

인구주택총조사 2%(A형)_가구주택사항 2005 > 검색

순번	형태	항목	순번	형태	항목	순번	형태	항목
1	H	가구일련번호	2	C	시도	3	C	동읍면부
4	C	가구구분	5	C	거주기간	6	C	방수
7	C	거실수	8	C	식당수	9	C	부엌형태
10	C	부엌사용형태	11	C	수도	12	C	화장실형태
13	C	화장실사용형태	14	C	목욕시설형태	15	C	목욕시설사용형태
16	C	거주층구분	17	C	승용차	18	C	승합차
19	C	화물및기타자동차	20	C	자동차없음	21	C	주차시설1
22	C	주차시설2	23	C	난방시설	24	C	점유형태
25	C	주인가구	26	C	주택소유여부	27	C	거처종류
28	C	연건평	29	C	총방수(방수+거실수+식당수)	30	C	건축년도
31	C	부엌면적시설수	32	C	화장실면적시설수	33	C	독립된출입구면적시설수
34	H	가중값(weight)	35	C	세대구성	36	C	가구유형

<그림 7-3> 인구주택총조사 2% 가구주택 표본자료 제공 변수

마지막으로 <그림 7-4>는 각각 인구주택조사구 가구주택사항에 대한 전국의 5% 표본 가구주택 자료 제공 관련 화면이다. 여기서는 시도 외에 시군구까지의 정보가 제공되어 시군구별 소지역 통계 계산까지 가능해진다. 반면, 제공되는 변수의 수가 1%나 2%에 비해 현격하게 작아지는 것을 알 수 있다.

인구주택총조사 5%_가구주택샘플 2005 > 검색

순번	형태	항목	순번	형태	항목	순번	형태	항목
1	H	가구일련번호	2	C	행정구역_시도	3	C	행정구역_시군구
4	C	행정구역_동읍면부	5	C	가구구분	6	C	거주기간
7	C	방수	8	C	출차량보유대수	9	C	주차시설1
10	C	주차시설2	11	C	가구소유형태	12	C	주인가구
13	C	주택소유여부	14	C	세대구성	15	C	거처종류
16	C	연건평	17	C	건축년도	18	H	가중값

<그림 7-4> 인구주택총조사 5% 가구주택 표본자료 제공 변수

다. 기타 조사구 자료 신청

KMDSS를 통해서는 5% 조사구 자료까지 신청이 가능하다. 만일 더 자세하게 전수조사 또는 10% 표본조사의 조사구별 데이터를 활용하는 데에는 일정한 제약이 가해진다. 통계청 내부의 표본설계인 경우에는 당연히 조사구 데

이터에 접근하여 작업하는 것이 가능하다. 그러나 통계청 아닌 다른 작성기관에서 표본설계를 하고자 할 경우라면 먼저 국가승인통계 여부가 중요하다. 승인통계가 아닌 경우라면 통계청에서 공개적으로 제공하는 데이터 이상의 조사구 데이터에 접근할 수가 없다.

국가승인통계인 경우 통계청의 허가를 얻은 후에 통계청이 지정하는 제한된 장소 내의 지정된 단말기에서만 작업이 가능하다. 이 경우 원자료의 유출은 불가하고 최종 표본으로 추출된 조사구의 정보만 가져갈 수 있다. 또한 이 경우에도 통계청 자체의 표본조사 대상 조사구인 10% 조사구는 제외하며 나머지 조사구만 표본으로 추출할 수 있다는 제약을 두고 있다. 한마디로 개인이 조사구 자료에 접근하는 것은 불가능하며, 국가승인통계 표본설계인 경우에 한해 통계청에서 제공하는 장소의 단말기에서만 작업이 가능하도록 되어 있다. 사실상 외부 기관의 연구자가 조사구 원자료를 활용하는 것에는 상당한 제약이 따라 표본설계 단계에서 조사구 자료를 충분히 이용하기 어려운 실정이라고 할 수 있다.

4. 인구주택조사구

앞의 경상조사 표본설계 사례를 통해 인구주택조사구의 쓰임새를 살펴본 바가 있는데, 이와 같이 우리나라에서 인구주택총조사를 추출틀로 사용하는 대부분의 표본조사 설계를 살펴보면 1차추출단위(primary sampling unit: PSU)로 인구주택조사구를 사용하고 있음을 알 수 있다. 전국 표본설계의 효율을 고려한다면 PSU의 특성은 매우 중요하다. 그러므로 여기서는 먼저 인구주택조사구가 어떤 과정을 통해 구성되었는지를 살펴보기로 한다.

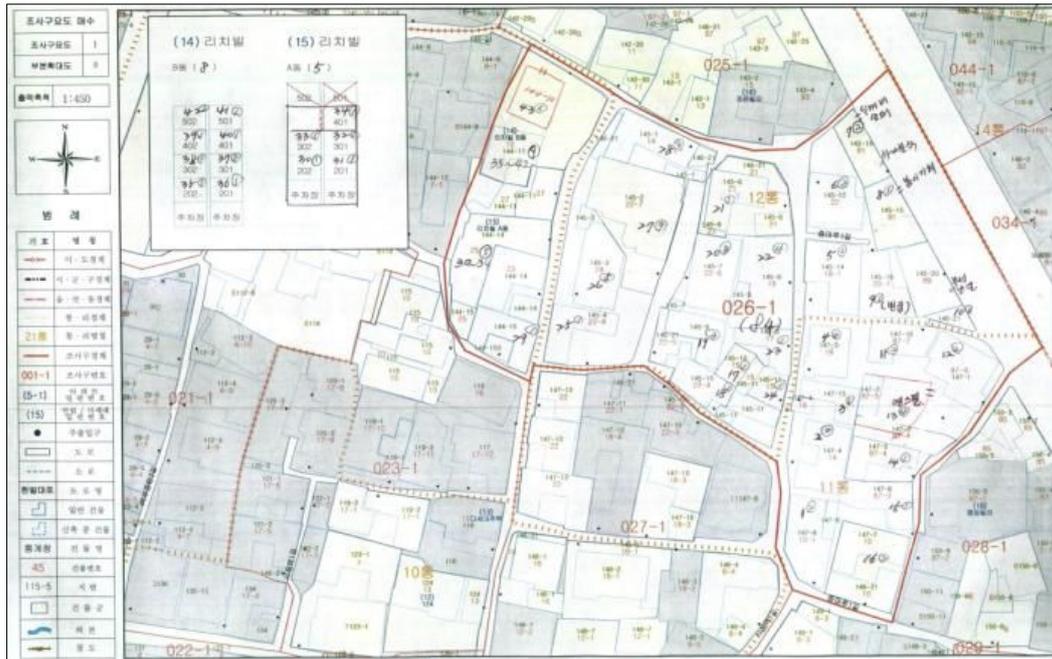
인구주택총조사의 조사과정은 조사구 설정으로부터 시작된다. 통계행정편람(2008, 통계청)에 소개된 조사구 설정 관련 내용은 비교적 간단하게만 소개하고 있는데 인용하면 다음과 같다.

“조사구란 우리나라의 모든 지역을 대상으로 도로, 하천 등 준항구적으로 변화가 적은 지형지물을 이용하여 지도(읍면동별) 상에 일정 가구수(평균 60 가구)가 포함되도록 조사구역을 명확히 구획하는 작업이다.

조사구 설정 절차는 본조사 1년 전에 전 지역을 대상으로 경계 및 지형지물을 보완하고 그 바탕 위에 변화가 예상되는 지역은 각 건물(거처)별 가구수를 파악하고 그렇지 않은 곳은 과거 총조사 자료를 활용하여 20가구 단위로 거처묶음을 실시하는 기초조사를 실시한다. 기초조사 자료를 이용하여 통계청에서 조사구 설정 기준에 의해 경계구획 및 번호를 부여하는 조사구 가

설정을 하면 지자체에서 가설정 조사구 적합성 확인 및 보완을 거쳐 통계청에서 조사구 설정 확정 승인을 하게 되고 같은 해 10월말까지 변동지역을 파악하여 보완한 후 본조사에 활용하게 된다.”

참고로 다음의 그림은 조사구의 예시도면이다. 전국의 모든 조사구들은 이와 같은 도면으로 작성되어진다.



<그림 7-5> 조사구 예시도면

출처 : 통계청 내부자료

조사구를 구성하는 주된 목적은 조사의 효과적 관리를 위한 부분이 큰 것으로 판단된다. 즉, 조사원별 조사담당 구역을 명확히 하여 조사대상(인구, 가구, 주택)의 누락 및 중복을 방지하고, 조사원의 업무량을 적정하게 배분하기 위해서이다. 그리고 덧붙여서 향후 조사구 데이터를 각종 표본조사의 추출틀로 활용할 수 있도록 제공하는 것도 하나의 목표이다.

한편 전체 조사구 중 약 10%에 해당하는 조사구는 총조사에서 10% 표본조사 대상 표본으로 지정되어 보다 자세한 항목의 조사가 이루어지고 있다. 그러므로 총조사가 끝나게 되면 전체 조사구에 대해서는 전수조사 항목 자료가, 10% 표본조사 대상 조사구에 대해서는 표본조사 항목 자료가 얻어지게 된다.

인구주택조사구의 구성 및 목표 등을 종합적으로 살펴볼 때 인구주택조사구는 특별한 행정적 의미나 통계 공표단위로서의 의미를 지니기보다는 단순

히 조사관리를 고려한 측면이 강하다는 것을 지적할 수 있다. 즉, 조사구는 공표 목적이 아닌 조사의 효율성을 목적으로 작성되는 단위로서 매년 총조사를 할 때마다 재구성되어 시계열적 의미를 지니지도 않고, 행정적 의미도 지니지 않는 다분히 임의적인 단위라고 할 수 있다. 세부적으로는 조사구의 세대수, 지리적 영역, 통계적 특성 등에 대한 이론적, 체계적 분석이 충분히 이루어진 것 같지는 않다. 또한 다분히 인구주택조사구의 구성은 통계청의 내부적 필요를 채우기 위한 측면이 많으며 외부 기관이나 개인을 위한 자료 제공 서비스 측면은 중요하게 고려되지 않은 것으로 판단된다. 끝으로 개별 인구주택조사구 단위의 통계 자체는 특별한 고려대상이 아니다. 다시 말해 동읍면 이상의 소지역별 추정 과정에서 편의상 조사구 통계가 고려될 수는 있으나 최소 공표단위로서의 의미는 지니지 않는다.

5. 기존 인구주택조사구의 문제점

여기서는 표본조사를 위한 1차추출단위로서 기존의 인구주택조사구가 지니는 문제점들을 지적하고자 한다. 전통적인 조사에서의 문제점뿐만 아니라 등록센서스와 병행되는 표본조사 상황을 전제할 때 예상되는 추가적인 문제점도 함께 살펴보기로 하겠다.

가. 조사구 설정 작업의 어려움

우리나라의 인구주택총조사 결과의 집계·공표는 행정구역 체계를 기본으로 이루어진다. 반면, 통계자료의 수집은 행정구역과는 구별되는 별도의 통계지역인 인구주택조사구가 활용된다. 조사구는 인구주택총조사가 실시되는 5년마다 새롭게 설정되며, 설정 당시의 예산이나 조사원 수 등에 따라 영향을 받는다.

통계자료 수집의 기본단위인 조사구가 5년 단위로 재설정 되면서 매회 설정을 위한 방대한 인력 및 예산을 투입해야 하는 문제점이 생긴다. 이를 보완하기 위해 통계청에서는 기초단위구라는 개념을 도입하여 설정하기도 하였다. 기초단위구는 도로, 하천, 철도, 산 능선 등과 같이 준 항구적이면서도 명확한 지형지물을 이용하여 지도상에 구획한 최소 단위구역을 의미한다. 이 구획은 인구수에 관계없이 정하는 까닭에 기초단위구별 인구 및 가구수는 균등하지 않다.

한편 통계 공표를 위한 기초단위인 행정동 단위의 규모가 너무 커서 하나

의 동질적 개념의 지역으로 간주하기 어렵다는 지적에 따라 통계청에서는 기존의 행정동 단위보다는 작고, 기초단위구보다는 크며, 기초단위구를 기본으로 하는 집계구(Enumeration District)를 설정하였다. 집계구는 평균인구규모를 500인 내외가 되게 하고 집계구 내의 주택유형, 평균지가라는 사회경제적 동질성을 고려하는 한편, 집계구의 지리적 형상이 가능한 한 정방형의 형태를 띠도록 설정되었다.

기초단위구나 집계구 개념이 도입되긴 했지만 실제 조사를 수행할 때에는 인구주택조사구가 최소의 조사단위가 되고, 표본설계를 위해 인구주택총조사 데이터를 활용할 때 사용되는 1차추출단위도 기초단위구나 집계구가 아닌 인구주택조사구 자료이므로 사실상 인구주택조사구의 설정은 매우 중요한 부분이라 할 수 있다.

이미 언급한 바와 같이 5년마다 새롭게 인구주택조사구를 설정하는 작업은 매우 방대한 일이 되어 막대한 시간과 노력이 들 뿐만 아니라 통계청 이외 다른 정부부처와의 긴밀한 협력이 요하는 까다로운 작업이다. 이렇게 해서 작성되는 조사구의 유용성에 관한 면은 논외로 친다고 하더라도 조사구 설정 작업 자체가 매우 힘들고 어렵다는 문제점을 지닌다.

나. 조사구의 임의적 성격으로 인한 문제

통계청에서 나름으로 설정한 조사구는 통계청 이외의 다른 이용자들의 입장에서 볼 때는 특별한 지리적, 행정적 의미를 지니지 못하는 임의적 성격의 단위라고 할 수 있다. 방대한 작업을 통해 작성된 인구주택조사구이긴 하지만 조사구 개념 자체가 다분히 통계청의 조사관리 편의를 고려하여 설정된 것이기 때문이다. 다시 말해 현재 편성된 조사구 구획을 일부 변경한다고 해서 별로 문제될 것이 없다. 실제 조사구는 한 번의 총조사에서만 사용되는 개념일 뿐이고 다음 번 총조사에서는 구획이 달라지므로 시계열적인 연속성이 전혀 없는 것이다.

무엇보다 조사구가 최소의 행정단위인 통/반/리와는 독립적으로 설정된 것은 재고할 여지가 있는 부분이다. 가령 최소 행정단위를 기초로 조사구가 작성된다고 하면 행정단위의 변동이 없는 한 시계열적 연속성을 지닐 수 있으며, 또한 통계청 이외의 다른 통계 작성기관이 조사구의 특성을 이해하기도 쉽기 때문이다. 아울러 이럴 경우 조사구 설정 작업의 부담을 훨씬 경감시킬 수 있을 것으로 기대된다.

현재 조사구는 60가구를 기준으로 이루어져 있는데, 도시와 시골 지역의

지리적 특성이 전혀 다른 점을 고려한다면 이와 같은 일률적인 규모의 조사구가 바람직한가에 대한 의문을 제기할 수 있다. 도시지역이라면 지리적, 사회적 특성이 비슷한 60가구가 같은 조사구에 포함될 수 있겠지만, 시골지역에서는 그렇지 않은 경우가 많이 생길 수 있을 것이다.

최적 조사구의 규모, 조사구 내 가구들을 동질적으로 할 것인가 아니면 이질적으로 할 것인가의 문제 등등 조사구 설정과 관련하여 사전에 검토되어야 할 사항들이 매우 많은데 이런 부분들에 대해 충분한 연구나 검토 없이 조사구가 작성된 측면에 대해 비판의 여지가 있다.

다. 자료이용 상의 문제점

통상적으로 조사구 단위의 자료는 총조사가 시행된 후 1년 이상 경과한 후에 활용이 가능하며, 5년에 한번만 조사될 뿐 추후 모집단에 변동이 생겨도 관리되지 않는다. 그러므로 총조사 실시년도(00년, 05년)를 고려할 때 끝자리가 02년 또는 07년에는 최신의 추출틀에 의거한 표본설계가 가능하지만 그 외의 년도일 경우 추출틀의 모집단에 대한 대표성이 떨어지게 되는데 현재로는 이에 대한 아무런 대책이 없는 실정이다.

예를 들어, 질병관리본부에서 시행하는 국민건강영양조사의 경우 표본설계 개편 주기가 3년인 까닭에 이 주기가 통계청의 인구주택총조사 주기와 일치하지 않는다. 2010년 국민건강영양조사를 위한 표본설계를 하려 할 때 2005년 인구주택총조사 결과는 너무 낡은 추출틀이어서 별도의 다른 추출틀을 찾아야만 했다.

가령, 조사구가 시계열적 분석이 가능하도록 지리적 또는 행정적으로 고정된 개념으로 작성되었다면 표본조사구에 한해서나마 모집단의 변동상황을 파악하여 전국의 모집단 변동을 추측하는 것이 가능할 텐데 현재로는 그런 가능성이 전혀 없기 때문에 인구주택총조사 결과의 추출틀로서의 효용성이 매우 제한된다고 할 수 있다.

자료 이용과 관련하여 또 다른 문제점으로는 통계청 이외의 외부기관에서 조사구를 추출틀로 활용할 경우 조사 수행 상 여러 면에서 불편을 느끼게 된다. 표본으로 추출된 조사구가 행정적 단위를 기초로 설정되어 있지 않으므로 실제 가구조사를 진행하려 할 때 뚜렷한 접촉점을 찾기가 어렵다. 실제 통장이나 반장이 지역조사의 접촉점으로 유용한데 조사구를 추출틀로 사용할 경우 그런 접촉점을 충분히 활용할 수 없게 된다.

뿐만 아니라 표본설계 단계에서 조사구를 추출틀로 활용할 경우 작업이 통

계청에서 지정하는 장소에서만 이루어져야 하는데, 계속조사의 경우 설계 단계에서 충분히 사후 관리를 위한 예비표본을 확보했다고 해도 부득불 추가적인 표본대체가 필요한 경우가 생기는데 이때마다 복잡한 행정절차를 거치는 것은 매우 불편하다.

라. 등록센서스 전환 시 예상되는 문제점

총조사가 등록센서스로 전환될 경우 기존의 인구주택조사구를 유지하는 것은 사실상 불가능할 것으로 예상된다. 무엇보다 먼저 등록자료의 경우 5년에 한번이 아니라 수시로 자료 갱신이 이루어지게 되므로 조사구를 5년에 한번씩 설정하는 방식을 더 이상 적용할 수가 없게 된다. 또한 등록자료와 매칭되지 않은 별도의 조사구 코드를 설정하여 부가하여 관리하는 것도 사실상 어렵다고 할 수 있다.

등록센서스가 시행될 경우, 새로운 조사구는 기본적으로 활용 가능한 등록자료들에 포함된 변수들 중 가장 합리적인 것을 기초로 작성되는 것이 바람직하다. 예를 들면, 주민등록의 주소 중 최종단위인 통·반·리 코드를 활용하는 방안을 생각할 수가 있을 것이다. 다시 말해 임의적 성격의 구분 코드가 아닌 지속적인 구분 코드를 기초로 하는 조사구 설정이 반드시 필요하다. 이렇게 된다면 추출틀은 등록자료의 갱신과 더불어 주기적으로 갱신되는 셈이므로 모집단의 변동 상황을 거의 실시간으로 반영할 수 있게 될 것이다.

제3절 외국사례 검토

지금까지는 우리나라의 기존 인구주택총조사의 자료가 다른 표본조사를 위한 추출틀로서 어떻게 활용되어 왔는지를 중점적으로 살펴보았는데, 여기서는 외국 나라들의 경우 어떤지를 검토해보기로 한다. 여기서 중점적으로 살펴보고자 하는 바로는 첫째 전수조사와 병행하는 표본조사의 규모가 어느 정도인가, 둘째 총조사와 병행하는 표본조사 방법은 어떠한가, 셋째 외국의 통계 관련 구역화의 특징은 무엇인가, 넷째 표본설계를 위한 마이크로데이터 정보 제공은 어떤 방식으로 이루어지고 있는가 등이다. 이러한 고려를 통해 향후 등록센서스로 전환할 때 표본조사와 관련하여 유용한 시사점들을 찾을 수 있을 것이다.

1. 표본조사 규모

각 나라들의 센서스 형태를 보면 문화와 여건에 따라 다양한 양상을 나타낸다. 나라에 따라서는 전수조사만 실시하는 나라도 있고 전수조사와 표본조사를 병행하는 나라도 있다. 또한 전수조사를 하는 경우라고 해도 전통적인 방법인 현장조사에 의해 실시하는 나라도 있고 행정자료를 이용하는 등록센서스를 도입한 나라도 있다. 한편 5년 또는 10년 주기로 실시하는 나라도 있고, 매년 부분적으로 조사를 지속하는 순환조사(rolling survey) 방식을 채택한 나라도 있다. 현재 우리는 등록센서스를 채택한다고 했을 때 표본조사를 어떻게 병행해야 하는데 관심을 두고 있기 때문에 이런 면에서 우리나라에 참고가 될 만한 나라들인 미국, 캐나다, 싱가포르, 이스라엘 등의 경우를 집중적으로 검토하고자 한다.

미국은 10년 주기로 전통적인 조사방식에 의한 전수조사(short form survey)를 실시하는 동시에 표본조사(long form survey)도 병행하는데 2000년대 들어 표본조사 방식을 일시적 조사가 아닌 순환조사 방식으로 전환하여 시행하고 있다. 미국의 표본조사 규모는 전체 가구의 1/6로 우리나라의 1/10에 비해 규모가 큰 편이다. 미국에서는 최소 통계작성 단위가 카운티(county)인데, 카운티 별 목표오차의 수준을 정하고 그에 따라 표본조사 규모를 정하였다.

캐나다의 경우, 우리나라와 마찬가지로 전통적인 조사방법에 의해 전수조사와 표본조사를 실시하고 있으며, 조사주기도 우리나라와 동일한 5년이다. 캐나다의 경우도 표본조사의 비율이 20%로 우리나라에 비해 높은 편이다.

한편, 최근에 등록센서스로 전환한 나라들인 이스라엘과 싱가포르는 센서스와 병행하여 표본조사를 실시하는데 두 나라 모두 표본조사의 비율은 20%이다.

미국, 캐나다, 싱가포르, 이스라엘 모두 우리나라와 같이 센서스와 표본조사를 병행하여 실시하고 있는데, 우리나라와 차이가 나는 점은 모두 표본조사의 비율이 우리나라보다 높다는 점이다. 각 나라별로 목표하는 바가 다르기 때문이겠지만 그럼에도 불구하고 우리나라의 표본조사 비율이 합당한 것인가 면밀히 검토해볼 필요가 있을 것이다.

2. 표본조사 방법

센서스와 병행하여 실시하는 표본조사는 전통적인 방법으로 일정 기일 내에 일괄적으로 조사를 수행하는 일시조사와 전체 표본을 독립적인 여러 순환표본(rolling sample)로 구분한 후 일상적으로 조사를 진행하는 순환조사를 들 수 있다.

가. 전통적 일시조사

전통적인 일회조사는 총조사가 실시될 때 병행하여 정해진 비율의 모집단 가구에 대해 표본조사를 실시하는 방식으로 여태까지 우리나라에서는 이 방법을 사용해왔다. 캐나다도 우리나라와 같은 방식으로 매 5년마다 20% 가구에 대해 표본조사를 실시한다. 한편, 이스라엘과 싱가포르는 최근 등록센서스로 전환하면서 그와 동시에 20% 표본조사를 일시조사 방법으로 실시한다.

전통적 일시조사의 가장 큰 약점으로 지적할 수 있는 것은 5년에 한번 통계가 작성되어 시의성이 떨어진다는 문제이다. 그밖에도 경상적인 조사가 아닌 까닭에 훈련된 조사원의 부족, 비표본오차의 문제 등 일반적인 조사방법론 상의 문제들을 생각할 수 있다.

나. 순환조사

미국의 경우 2000년 중반부터 센서스에 병행되는 표본조사인 American Community Survey (ACS)를 순환표본조사 방식으로 전환하여 실시하고 있다. 순환표본조사는 전체 모집단을 대표하는 서로 독립적인 여러 개 크기가 비슷한 순환표본을 추출하여 매년 하나의 순환표본에 대해 조사를 실시함으로써 광역단위의 통계를 매년 작성하고, 여러 순환표본이 누적될 경우 소지역단위의 통계까지 작성할 수 있게 하는 조사방법이다(Kish, 1992).

미국에서 ACS에 순환조사가 도입된 이유는 다음과 같다.

10년 동안 표본조사와 유사하도록 지리적으로 상세한 경제사회자료를 제공한다.

미국지역사회조사는 매월 행해지는 조사이므로 직원의 대다수를 종신고용으로 한다. 이를 위해서 임시직원에게 크게 의존하여 10년 간격으로 행해지는 총조사보다 높은 품질의 결과를 기대할 수 있다. 또한 데이터수집에 대한 지속적인 개량이 가능해진다.

총조사로부터 표본조사의 부분을 삭제함으로써, 운영을 간소화하고 인구조사와 포괄범위(coverage)의 향상에 집중할 수가 있다.

주(state)나 지방자치단체의 복지개혁 등에 대한 새로운 책무를 지원한다.

미국지역사회조사의 결과를 전수조사 및 다른 연방정부의 가구조사 및 행정자료 등과 통합하여, 연방정부통계시스템의 인프라를 개선한다.

순환조사의 가장 큰 장점은 기존의 일시조사와는 달리 매년 통계를 생산할 수 있어서 시의성의 문제를 해결할 수 있다는 점이다. 등록센서스로 전환될 경우 센서스 결과가 매년 공표될 터인데 이에 맞추어 표본조사 결과도 매년 갱신될 수 있으므로 센서스와 표본조사가 함께 간다는 면에서 바람직하다. 또한 조사가 항시적으로 진행되므로 훈련된 조사원에 의해 표준화된 조사를 실시함으로써 비표본오차를 최소화할 수 있다는 장점도 있다.

반면, 실제 센서스 병행 표본조사에서 순환조사를 채택한 유일한 사례가 미국의 사례라고 할 것인데 아직 경험적으로나 이론적으로 충분한 검증이 이루어지지 않았다는 것을 문제점으로 지적할 수 있다. 또한 순환조사를 위한 추출틀을 무엇으로 할 것인가에 따라 부수적으로 여러 복잡하게 해결해야 할 문제들이 생기게 된다. 구체적으로 우리나라의 현실에서 어떤 문제들이 발생할 수 있을지, 그에 대해 어떤 해결책이 있는지 등에 대해 체계적이고 충분한 논의가 필요하다고 할 것이다.

3. 통계 관련 구역화

외국에서도 사회발전에 따라 다양한 형태로 소지역통계에 대한 수요가 증가함에 따라 통계 생산과 공표를 위해 공간적으로 적절한 규모를 정하여 전국을 여러 형태의 구역으로 나누어 놓았다. 이러한 구역화는 크게 조사를 위한 조사구역과 집계 및 공표를 위한 공표구역으로 나눌 수 있다 (통계청, 2007).

<표 7-3>은 미국, 영국, 캐나다, 호주, 일본 5개국에서 사용하고 있는 통계 구역을 요약한 것이다. 전체적인 모습을 요약하면, 공표구역의 경우 일본이나 우리나라와 같이 기존의 행정구역 체계를 그대로 사용하는 경우도 있으나, 대부분의 국가에서는 행정구역과 별도로 통계의 집계 및 공표를 위해 새로운 지역 구분 체계를 사용하고 있다. 각국의 통계구역 체계는 국가에 따라 다양하게 나타나고 있기 때문에 하나로 단일하게 설명할 수는 없지만, 몇 가지 공통된 특성이 나타나고 있다.

첫째, 조사구역과 공표구역의 분리 경향이다. 대부분의 국가들, 그 가운데도 특히 영연방 국가들은 과거에 조사구역을 공표단위로도 사용하였다. 그러나 최근에는 조사구역과 공표구역을 분리하여 사용하고 있다. 전국을 작은 구역으로 나누어 관리하는 것은 매우 복잡하고 지난한 일이었지만, 최근에 GIS의 발달로 이러한 어려움이 많이 줄었기 때문에 다양한 형태의 구역을 지정하여 관리할 수 있게 되었다. 조사구역은 통계 조사의 효율성을 목적으로 하며, 공표구역은 공표를 목적으로 한다. 공표구역의 경우 개인정보보호, 통계적 유의성, 소지역 통계에 대한 수요 등을 다양한 요건을 만족시키는 기준을 마련하고 있다.

둘째, 공표 구역의 설정 조건으로는 인구 기준이 가장 중요한 요소로 작용하고 있으며, 기존의 행정구역 체계 및 기타 지역 구분 체계와의 통합가능성을 고려하고 있었다. 미국의 경우에는 블록그룹과 센서스트랙을 인구 조건만을 기준으로 설정하였으며, 영국과 캐나다의 경우는 인구 조건을 기본 조건으로 형상 조건, 내부적 동질성, 도로에의 접근성 등을 추가적으로 고려하였다.

셋째, 일반 행정, 선거구 획정, 도시지역 구분 등 에서도 원활히 사용할 수 있도록 공표구역들을 행정구역 경계를 넘나들지 않도록 작성하고 있다. 통계 기반 구역의 용도를 살펴보면 미국은 최소 공표 단위인 블록을 도시화지역, 선거구, 학군 등 다양한 지역 설정의 기본단위로 활용하고 있다. 호주는 Collection District 설정결과를 바탕으로 이들 여럿을 하나로 묶어 도시화 지

역을 설정한다. 캐나다에서는 대도시권인 Census Metropolitan Areas의 설정에서 Census Subdivision을 기본단위로 하여 통계공표지역인 Dissemination Area와 매칭하여 사용하고 있다.

<표 7-3> 외국의 통계 관련 구역의 특징*

	미 국	영 국	캐나다	호 주	일 본
센서스의 시행	10년 단위 (○○○○년)	10년 단위 (○○○1년)	5년 단위 (○○○1, ○○○6년)	5년 단위 (○○○1, ○○○6년)	5년 단위 (○○○0, ○○○5년)
최소조사 구역 및 도입년도	Block/ 1940년	Enumeration District/ 1841년	Enumeration Area/ 1961년	Census Collection District	조사구 (Enumeration District)
조사구역 설정기준	① 최소면적 약 2,780ha ② 최소폭 약21m ③ 최소1면 이상 도로에 인접	① 조사원의 업무량 표준화 ② 주요간선도로, 강, 철도 등으로 구분 ③ ward, parish의 경계 유지	① 조사원의 업무량 표준화 ② 단순한 형상 ③ 도로에 의한 접근성 ④ 타 공간단위의 경계	① 조사원1명의 10일간 업무량 ② SLA및 타지역 구분의 경계 유지 ③ 가시적 지형지물로 구분	① 조사원의 업무량 기준 ② 기본단위구를 결합하여 형성
조사구역 인구규모	제한 없음	지역별로 다양 (평균 450명)	농촌지역 125명 도시지역 650명	최소 100명 원주민 지역 80명	평균 50명
최소공표 구역 및 도입년도	Block/1940년 (주공표는 Census tract)	Output Area/ 2001년	Dissemination Area/ 2001년	Census Collection District (2011년부터Mesh Block)	cho. aza Grid Square/ 1973년
공표구역 설정기준	조사구역과 동일	①인구:최소 100명,최적 312명 ②형상 : 중심간 거리의 제곱을 최소화 ③내부 동질성: 주택유형 및 점유형태 ④ 타 공간단위의 경계	① 인구:최적500명 ② 형상:정방형우선 ③ 타공간단위의 경계 ④ 가시적 경계의 사용여부	조사구역과 동일	①국세조사: 행정구역 체계 ②Grid Square:위도 40' 경도 1°에 해당하는 격자를 80×80으로 균등 분할
가시적 경계의 사용여부	행정구역 및 도로, 철도, 하천, 능선 등의 가시적 경계 사용	행정구역 경계를 사용	행정구역 및 도로와 같은 가시적 경계를 사용	행정구역 및 도로 하천 등의 가시적 경계 사용. 농촌의 경우는 도로를 경계로 사용하지 않음	행정구역 경계를 사용
공표구역 인구규모	인구 기준 없음	평균 297명	400~700명	도시지역 평균 512명	해당사항 없음
개인정보 보호를 위한 규정	표본조사는 블록그룹(600~3,000명)단위로 공표 일부 교차표분석 자료는센서스트랙(1,500~8,000명) 단위로 공표	교차분석 자료는 1,000명 이상의 지역에 대해서만 발표	최소 공표단위의 인구 최소기준을 설정(400명 이상)	최소 공표단위의 인구 최소기준을 설정(100명 이상)	표본조사는 cho, aza의 상위체계인 시구정촌 단위로 공표
조사구과 공표구역의 관계	동일	분리. ED를 통해 수집된 자료를 개별주소 단위로 입력	분리. EA와 DA의 기본단위인 Block을 통해 매칭가능	동일	조사구역이 공표구역의 하위
도시/비도시 구분	블록을 도시화지역 설정의 기본단위로 사용	도시지역의인구추정에 OA인구를 이용	구분. 도시지역은 DA의상위체계인 CSD 단위로 설정	CD의 설정결과를 기준으로 도시화지역을 설정	구분하지 않음 행정구역에 따른 구분은 가능

출처 : 강영욱·윤은주·원종석·정재희(2007)

넷째, 소지역 통계 수요의 증가에 따라 자료의 유형에 따라 공표기준을 다르게 적용하여 다양한 형태로 통계를 공표하고 있다. 특히, 개인정보노출 위험이 적은 자료는 통계적의 의미 있는 최소단위까지 자료를 제공하여 소지역 통계의 활용도를 높이고 있다. 이와 관련하여, 자료 공표 시 우려되는 개인정보보호를 위한 제도적 장치를 도입하고 있다. 예를 들어, 미국과 일본은 표본 조사의 공표 단위를 전수 조사의 공표 단위보다 상위의 지역 단위에서만 발표하고 있으며, 영국에서는 교차표(contingency table) 분석 자료의 공표 시 인구 1000명을 기준으로 하여 최소공표단위인 Output Area 수준에서는 발표를 금지하고 있다. 호주는 공표구역설정에서 최소인구기준을 100명, 캐나다는 400명으로 정하여 구역 획정에서부터 개인정보를 보호할 수 있도록 하고 있다.

4. 마이크로데이터 제공

최근 IT 기술의 비약적인 발달로 인해 국가기관뿐 아니라 학계, 연구소 및 일반 개인들의 원자료에 대한 수요는 눈에 띄게 증가하고 있는 양상이다. 각 나라의 센서스는 기본적으로 국민들의 세금과 국민들의 협조로 이루어지므로 가능한 한 많은 사람이 센서스를 통해 생산되는 자료를 충분히 활용하게 해야 함이 마땅하다고 할 것이다. 또한 많은 이용자들이 통계자료를 활용하면 할수록 통계자료의 부가가치를 더욱 높일 수 있을 것이므로 조사기관에서는 가능한 한 다양한 자료를 제공하려 노력해야 할 필요가 있다.

그러나 센서스 마이크로데이터에는 개인정보가 담겨 있으므로 이를 무분별하게 제공할 수는 없는 실정이다. 거의 모든 국가에서는 센서스에서 조사된 개인정보의 기밀보호를 법으로 규정하고 있기도 하다. 이런 이유로 인해 마이크로데이터 제공을 할 때 부분적인 제약을 가하는 것은 어쩔 수 없는 일로 인식된다. 누구나 이용할 수 있도록 제공하는 데이터를 Public Use Microdata (PUM)라 하고, 개인정보가 담긴 데이터를 confidential data라고 구분하여 각 나라에서는 각각의 경우에 대해 서로 다른 데이터 제공 정책을 쓰고 있다. PUM의 경우 공개적으로 자료를 제공하는 반면, confidential data인 경우는 특정 이용자들에게만 까다로운 절차를 거쳐 이용을 허락하는 서비스를 별도로 운영하는 방식을 채택하고 있다.

가. Public Use Microdata

일반인들에게 PUM을 제공하는 방식은 국가마다 또는 기관의 성격에 따라 다양한데 여기서는 미국의 경우를 중점적으로 살펴보자.

미국은 센서스 표본조사(전체의 1/6 추출) 데이터를 이용하여 각각 1%와 5% 표본자료를 PUM으로 제공하고 있다. 각각의 자료는 표본조사 데이터 중에서 다시 대표성 있는 표본을 추출한 것이다.

각각의 자료는 표본조사 자료 중 일부를 다시 추출한 것이므로 자료의 통계적 정확성에 차이가 있다. 따라서 미국에서는 PUM 관련 정보를 제공하는 문서에 <그림 7-6>에 나타난 바와 같은 추정오차의 범위와 표준오차 계산 관련 정보를 별도로 제공하고 있다.

Table A. Unadjusted Standard Errors for Estimated Totals from Census 2000 5-Percent PUMS

Estimated total ¹	Size of geographic area tabulated ²							
	100,000	250,000	500,000	750,000	1,000,000	5,000,000	10,000,000	25,000,000
1,000	137	138	138	138	138	138	138	138
2,500	215	217	217	218	218	218	218	218
5,000	300	305	307	307	307	308	308	308
10,000	414	427	432	433	434	435	436	436
15,000	492	518	526	528	530	533	533	534
25,000	597	654	672	678	681	687	688	689
75,000	597	999	1,101	1,132	1,148	1,185	1,189	1,192
100,000		1,068	1,233	1,283	1,308	1,365	1,371	1,376
250,000			1,541	1,780	1,887	2,124	2,152	2,169
500,000				1,780	2,179	2,924	3,004	3,051
750,000					1,887	3,480	3,631	3,718
1,000,000						3,899	4,135	4,271
5,000,000							6,892	8,718
10,000,000								10,677

¹For estimated totals larger than 10,000,000, the standard error is somewhat larger than the table values. The formula (1) given below should be used to calculate the standard error.

²Total count of people, housing units, households, or families in the area if the estimated total is a person, housing unit, household, or family characteristic, respectively.

$$SE(\hat{Y}) = \sqrt{19(\hat{Y})\left(1 - \frac{\hat{Y}}{N}\right)} \quad (1)$$

N = Size of geographic area

\hat{Y} = Estimate of characteristic total

<그림 7-6> 미국의 PUMS 추정오차 관련 제공 정보의 예

나. Confidential Data

미국은 센서스국(Bureau of Census) 산하에 Center for Economic Studies (CES)를 설치하여 정부, 학계, 민간의 센서스 자료 활용 연구를 장려하고 있

다. CES는 미국 내에 9개의 Census Research Data Center (RDC)를 운영하고 있는데 RDC는 주로 큰 도시의 대학이나 연구소 또는 센서스국 관련기관에 설치되었다.

개인정보가 담긴 Confidential Data에 대한 접근은 지정된 RDC 공간 안에서만 가능하다. 원자료는 RDC 안에서만 다룰 수 있고 원자료를 가공하여 얻은 결과만 RDC 밖으로 가지고 갈 수 있게 하는 방식이다.

한편 RDC에 들어가서 자료를 이용할 수 있는 권한을 얻는 데는 많은 제한조건이 따른다. 엄격한 승인절차를 통과해야만 자료에 접근할 수 있는데 CES에서는 이와 관련하여 13개의 기준을 제시하고 있다(CES, 2003).

유럽의 경우 Eurostat에서 2002년 Confidential Data의 이용에 대한 규정을 마련하였는데, 여기서도 미국과 유사하게 연구계획서를 제출하여 Committee on Statistical Confidentiality의 승인을 받아야만 이용할 수 있게 하였다.

제4절 등록센서스 기반 추출틀 작성 방안

1. 쟁점

우리나라에서 지금까지 실시한 총조사에서는 표본설계를 위한 1차추출단위(PSU)로 인구주택조사구를 사용해왔다. 총조사도 크게 전수조사와 표본조사로 구분되어 시행되었는데, 일부 단순한 조사항목에 대해서만 전수조사가 이루어졌으며 다소 복잡한 나머지 조사항목들은 표본조사를 통해 자료가 수집되었다. 표본조사는 전체 인구주택조사구 중 약 10%를 표본으로 추출하여 이루어졌다.

여러 분야의 표본조사를 위한 표본설계 단계에서 추출틀을 마련하고자 할 때 일반적으로 인구주택총조사 결과를 활용하게 된다. 대부분의 표본설계에서 요구하는 추출틀 정보는 PSU 단위로 정리된 자료이다. 따라서 표본설계를 위한 자료제공의 문제를 고려할 때 PSU를 무엇으로 할 것인가 하는 문제가 가장 핵심적인 고려사항이 되어야 한다. 앞 절에서 PSU로써 기존 인구주택조사구가 지니는 문제점들을 지적한 바 있는데 여기서는 그런 점을 염두에 두고 새로운 등록센서스를 위한 PSU 작성방안을 구체적으로 검토하고자 한다.

등록센서스에 기반한 추출틀을 작성하고자 할 때 중요한 쟁점들로는 첫째

PSU를 무엇으로 결정할 것인가, 둘째 표본의 규모는 얼마가 적당한가, 셋째 표본 추출방식은 어떻게 하는 것이 좋은가 등을 들 수 있다. 각각의 쟁점에 대해 보다 상세히 검토하기로 한다.

2. 1차추출단위(PSU)

등록센서스가 실시될 경우 PSU 단위를 무엇으로 해야 할 것인가? 현실적으로 고려할 수 있는 대안으로 세 가지를 생각할 수 있는데, 첫째 인구주택조사구를 그대로 활용하는 방안, 둘째 행정구역인 통/반/리를 활용하는 방안, 셋째 지도 상 일정 면적의 지역을 추출단위로 하는 방안이 그것이다. 우리나라의 경우, 지리적 특성 상 일정 면적 단위의 구획을 PSU로 하는 방식은 적합하지 않으므로 사실상 PSU의 후보로는 인구주택조사구나 행정구역인 통/반/리를 들 수 있다. 여기서는 구체적인 자료에 근거하여 두 방안의 현황 및 가능성을 검토하고자 한다.

가. 인구주택조사구

기존의 인구주택조사구가 지니는 문제점으로 (1) 조사구 설정의 어려움 (2) 조사구의 임의적 성격 (3) 5년간 갱신되지 않아 생기는 자료 이용 상의 어려움 등을 지적한 바 있다. 등록센서스로의 전환을 염두에 둔다면 기존의 인구주택조사구가 지니는 이러한 약점들을 반드시 보완하는 것이 필요하다. 인구주택조사구를 설정할 때 등록센서스에 포함된 정보 중 유용한 항목을 이용하여 가능한 한 영구적으로 유지할 수 있으면서도 행정적, 지리적으로 의미가 있도록 설정해야 한다.

다음의 <표 7-4>는 기존의 2005년 인구주택총조사의 10% 표본조사에 나타난 인구주택조사구 현황을 정리한 결과인데, 전국의 표본조사구(Sample Enumeration District)는 34,241개이며, 표본조사구에 포함된 표본가구수는 1,591,631 가구, 조사구 당 평균 가구수는 46가구인 것으로 나타났다. 한편, 시도별 조사구 당 평균 가구수는 최저 41가구(충남)에서 최고 53가구(대구) 사이로 큰 차이가 없는 편이다. 원래 인구주택조사구를 정할 때에는 조사구 당 60가구가 기준이 되게 한다고 했으나 실제 작성된 조사구의 평균은 그에 못 미치는 46가구인 것으로 나타났다.

<표 7-4> 10% 표본조사의 조사구수 현황

시도	표본가구수	표본 조사구수	조사구당 가구수
전국	1,591,631	34,241	46
서울특별시	307,646	6,909	45
부산광역시	111,648	2,143	52
대구광역시	76,096	1,438	53
인천광역시	79,811	1,534	52
대전광역시	44,563	919	48
광주광역시	44,258	966	46
울산광역시	30,783	598	51
경기도	310,923	6,688	46
강원도	58,684	1,306	45
충청북도	55,761	1,260	44
충청남도	71,668	1,758	41
전라북도	75,146	1,696	44
전라남도	84,251	1,793	47
경상북도	109,728	2,546	43
경상남도	113,118	2,322	49
제주도	17,547	365	48

나. 행정구역상의 통/반/리

인구주택총조사의 전수조사 부문이 등록센서스로 전환될 경우 인구 및 가구 관련 정보는 행정안전부의 주민등록인구통계를 사용하게 될 것이다. 주민등록인구통계에는 인구주택조사구 코드는 없지만 거기에 가장 유사한 정보로 최소의 행정단위인 통/반/리를 고려할 수 있다. 양경진과 김황대(2008)는 주민등록자료가 표본조사를 위한 추출틀로 사용될 때 타당성, 시의성, 경제성 측면에서 유용함을 보인 바 있다.

2009년 3월의 주민등록인구통계를 활용하여 통/반/리별 현황을 정리한 것이 다음의 <표 7-5>에 나와 있는데 이 표를 보면, 2009년 3월의 우리나라 세대수는 약 1,900만 세대이며, 통/반/리 수는 약 38만 개로 통/반/리 당 평균 세대수는 50세대가량인 것을 알 수 있다. 평균 세대수로만 보면 통/반/리와 인구주택조사구는 비슷한 편이다. 한편, 시도별 평균 세대수를 살펴보

면 최저 33세대(강원)에서 최고 77세대(충남) 사이에 분포하고 있어서 앞에서 살펴보았던 인구주택조사구에 비해 시도별 편이가 다소 큰 편으로 나타나고 있다. 표준편차도 시도별로 매우 다른 양상을 띠고 있다. 대전이나 서울의 경우 표준편차가 21에 불과한데 반해 충남, 제주는 121.9, 132.1이나 된다. 즉, 통/반/리의 경우 전국적으로 비슷한 기준에 의해 구성되지 않음을 알 수 있다.

〈표 7-5〉 시도별 세대수와 통/반/리 수

지역	세대수	통/반/리 수	평균	표준편차
전국	19,035,656	381,735	50	62.2
서울특별시	4,094,558	98,963	41	21.7
부산광역시	1,311,210	26,835	49	24.3
대구광역시	899,329	20,946	43	38.8
인천광역시	1,013,931	18,675	54	43.2
대전광역시	533,923	12,882	41	21.0
광주광역시	514,393	10,391	50	26.0
울산광역시	393,328	7,964	49	70.6
경기도	4,296,503	63,324	68	97.9
강원도	607,816	18,146	33	32.0
충청북도	588,455	12,439	47	49.7
충청남도	813,470	10,511	77	121.9
전라북도	713,782	18,244	39	29.1
전라남도	773,939	14,090	55	83.8
경상북도	1,058,375	21,302	50	76.2
경상남도	1,207,210	23,843	51	78.0
제주도	215,434	3,180	68	132.1

어찌됐건 총조사를 등록센서스로 전환할 경우 현실적으로 생각할 수 있는 PSU로는 최소 행정단위인 통/반/리 단위가 가장 유력하다. 통/반/리를 PSU로 사용하게 될 때 여러 가지 면에서 장점들이 있는데 대표적인 것들로 다음 사항을 들 수 있다.

첫째, 인구주택조사구와는 달리 통/반/리는 실제적인 행정단위로서 의미를 지닌다는 사실이다. 인구주택조사구의 경우 5년에 한번 총조사를 할 때마다 새롭게 작성되는 일시적인 구분이지만, 통/반/리의 경우 항시적으로 유지된다는 점에서 매우 유용한 구분인 것이다. 그러므로 통계청 이외의 외부 기관에서 추출틀로 사용하려 할 때 쉽게 이해할 수 있으며, PSU로 어떤 통/반/리가 추출될 경우 통장/반장/이장이라는 구체적인 접촉점을 확보할 수 있어

서 매우 편리하다. 또한 특별한 사유가 없는 한 통/반/리가 지속적으로 유지될 것이므로 필요하다면 시계열 자료까지 작성하여 비교하는 것이 가능하다.

둘째, 통/반/리를 PSU로 한다면 항상 최신의 모집단 상황을 반영하는 추출틀을 제공할 수 있게 된다. 다시 말해 추출틀의 모집단에 대한 대표성을 높일 수 있다. 기존의 인구주택조사구의 경우 5년에 한번 조사가 이루어지므로 총조사 결과가 나오는 2년, 7년 차에는 비교적 모집단을 잘 대표하지만 해가 거듭될수록 추출틀의 모집단에 대한 대표성이 떨어져 포괄률(coverage rate)이 떨어지게 된다는 문제점을 지닌다. 하지만 주민등록인구통계는 매월 모집단의 변동을 반영하여 새롭게 갱신되는 통계이므로 통/반/리 단위의 추출틀 정보도 매월 새롭게 갱신되어 표본설계 시점이 언제이든 상관없이 유용하게 활용할 수 있다. 주민등록통계와 인구주택총조사 결과가 반드시 일치하는 것은 아니라는 문제점이 지적되지만 2005년의 경우 전국적 인구수 차이가 3.2%에 불과하여 추출틀로서는 전혀 문제될 것이 없음을 알 수 있다.

셋째, PSU 구성이 단순하고 간편하다. 인구주택조사구는 일시적으로 사용되는 개념임에도 불구하고 작성에 많은 노력이 필요하였다. 그러나 통/반/리의 경우 이미 있는 단위를 그대로 사용하는 것이므로 PSU 구성을 위해 별도의 추가적인 노력이 필요하지 않다.

넷째, 동일한 표본을 오랜 기간 반복적으로 조사하는 패널조사를 위한 추출틀로는 통/반/리 단위의 PSU가 효과적이다. 인구주택조사구는 모집단의 변화를 전혀 파악하지 못하므로 중간에 표본 가구나 조사구가 대체되어야 하는 상황이 생길 때 이에 적절하게 대처하기 어렵다. 하지만 통/반/리의 경우 거의 실시간으로 모집단 변화를 파악할 수 있어 적절한 대처가 가능해진다.

이상의 여러 장점들을 고려할 때 등록센서스로 전환할 경우 PSU로는 기존의 인구주택조사구보다는 통/반/리 단위가 바람직하다고 할 것이다. 다만 통/반/리 크기의 변이가 인구주택조사구에 비해 상대적으로 큰 편이므로 추가적인 고려를 통해 이를 보완하는 방안을 생각할 수 있다. 현재의 통/반/리는 전혀 통계적 목적을 고려하지 않고 단지 행정적 목적으로만 작성된 것으로 여겨진다. 가능하다면 통계청과 행정안전부가 협의하여 행정적 목적과 통계적 목적을 동시에 충족시킬 수 있도록 통/반/리를 조정하는 것이 바람직하다. 이때 인구주택조사구를 지정할 때와 같이 통/반/리별 규모를 일정한 수준으로 정할 수 있을 것이다. 만일 현재의 통/반/리를 조정하는 것이 여의치 않다면 사후적으로 인접한 통/반/리들을 필요에 따라 통합함으로써

PSU 규모의 변이를 줄일 수도 있다.

3. 표본크기

총조사가 등록센서스로 전환될 경우, 주민등록인구통계나 건축물대장 등의 활용 가능한 등록자료를 통해 확보할 수 있는 변수들은 제한적이다. 등록자료를 통해 확보하기 어려운 정보들은 모두 표본조사를 통해 보완되어야 할 것이므로 등록센서스와 병행하여 수행되는 표본조사의 중요성은 더욱 커질 것이다.

우리나라에서는 기존의 총조사에서는 전수조사와 더불어 대략 모집단의 10% 크기의 표본조사를 병행하여 실시해온 바 있다. 외국의 사례를 살펴보면, 미국은 표본조사의 규모가 모집단의 1/6이고, 캐나다는 1/5이며, 최근에 등록센서스를 도입한 이스라엘과 싱가포르 두 나라 모두 1/5이다. 또 2010 라운드에 등록센서스를 도입할 예정으로 되어 있는 독일은 건축물, 주택, 인구 주택 센서스의 표본규모를 모집단의 1/5(20%)를 상한선으로 하고 있다³⁾. 또 대만은 2010년에 처음으로 등록센서스를 실시할 때, 표본조사를 모집단의 16% 규모로 실시할 예정으로 되어 있다⁴⁾. 이와 같이, 등록센서스로 전환을 목표로 하고 있는. 외국의 사례와 비교할 때 자연스럽게 우리나라의 표본조사 규모가 작은 것 아닌가 하는 의문을 가질 수 있다. 표본조사의 규모는 조사의 목적에 따라 좌우되는데, 표본크기 결정에 가장 영향을 미치는 요소는 최소 통계작성범위와 목표오차의 수준이라고 할 수 있다. 먼저는 2005년 총조사 10% 표본의 지역별 추출률에 살펴봄으로 배분의 적절성을 검토해보고, 이어 표본크기의 적절성을 살펴보기로 하자.

다음의 <표 7-6>은 모집단의 가구수와 표본가구수 및 추출된 표본가구 비율을 나타낸다. 전국의 표본가구수는 약 159만 가구로서 전체 가구의 약 9.95%에 해당된다. 한편, 표본가구수 비율을 시도별로 살펴보면 최저 울산의 9.04%로부터 최고 전남의 12.09% 사이의 비율을 지니고 있다.

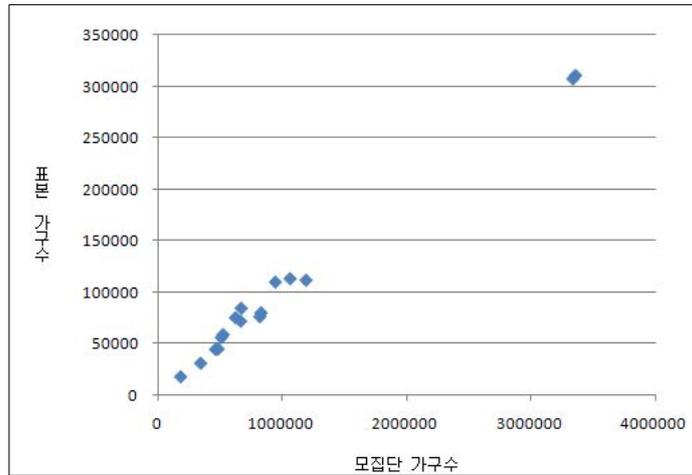
3) 본 연구의 부록 IV에 수록된 2011년 독일의 등록센서스 법(Census law) 제2절에 “표본조사의 추출률은 20%를 상한선으로 한다.”라고 규정하고 있다.

4) 본 연구의 부록 II에 수록되어 있는 대만(자유중국)의 등록센서스 실시계획에서, 최근의 시뮬레이션 테스트 결과에 의하면, 2010년 인구주택총조사의 총표본 추출율은 16%였으며, 대만지구 각 현시 추출율은 그 범위가 11%에서 29%의 사이에 있었다.

<표 7-6> 시도별 표본추출률

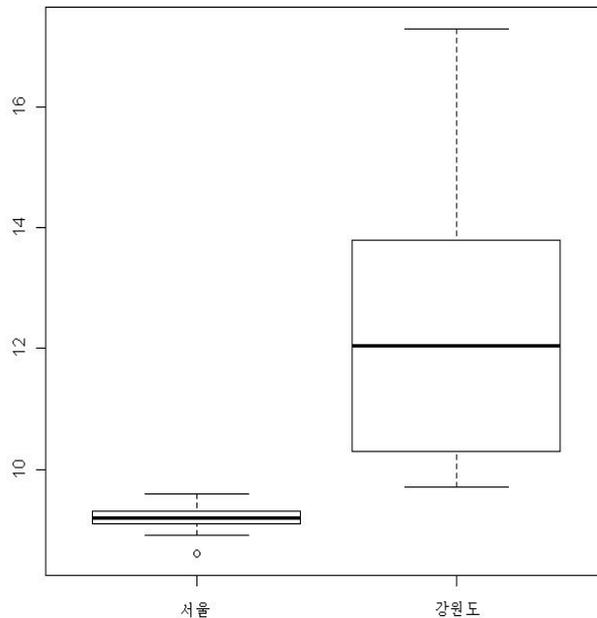
시도	일반가구수	표본가구수	표본가구 비율
전국	15,988,274	1,591,631	9.95
서울특별시	3,341,352	307,646	9.21
부산광역시	1,190,107	111,648	9.38
대구광역시	817,620	76,096	9.31
인천광역시	828,012	79,811	9.64
대전광역시	480,466	44,563	9.27
광주광역시	461,387	44,258	9.59
울산광역시	340,652	30,783	9.04
경기도	3,361,657	310,923	9.25
강원도	522,225	58,684	11.24
충청북도	507,541	55,761	10.99
충청남도	663,445	71,668	10.8
전라북도	621,735	75,146	12.09
전라남도	667,962	84,251	12.61
경상북도	943,135	109,728	11.63
경상남도	1,061,259	113,118	10.66
제주도	179,719	17,547	9.76

다음으로 시도별 모집단 가구수 크기와 표본크기와의 관계를 살펴보자. <그림 7-7>은 시도별 가구수와 표본 가구수의 관계를 산점도로 나타낸 것이다. 이 그림을 보면, 모집단 가구수와 표본 가구수는 대체로 비례하고 있음을 알 수 있다. 다시 말해, 시도를 층화변수로 간주한다면 층별 표본의 배분은 비례배분법(proportional allocation)에 의해 이루어졌다고 할 수 있는 것이다.



<그림 7-7> 시도별 모집단가구수와 표본가구수

<그림 7-8>은 서울과 강원도의 시군구별 표본비율의 분포를 나타낸 그림이다. 서울의 구들은 추출률이 전반적으로 9% 내외로 거의 비슷한 반면, 강원도는 10%에서 17.3% 사이로 나타나 서울에 비해 다소 추출률이 높은 편일뿐 아니라 시군에 따라 추출률의 차이가 다소 나는 것을 알 수 있다. 여기서는 서울과 강원도의 경우만 나타냈지만 전국 모든 지역에서 이런 양상이 나타나고 있다. 즉, 도시지역인 시나 구의 추출률은 대체로 8% ~ 11% 사이인 반면에 군지역의 추출률은 최소 10%에서 최대 27%에 이르는 것으로 나타났다.



<그림 7-8> 서울과 강원도 지역별 추출률분포

전반적으로 도시지역인 구나 시지역과 시골지역인 군지역의 추출률에는 차이가 나지만 같은 도시 내에서는 구별 가구수와 추출률 사이에 특별한 상관이 없는 것으로 드러난다. 서울을 예로 들어보자.

다음의 <표 7-7>은 서울의 구별 일반가구수와 표본가구수 및 표본비율을 나타내고 있는데, 서울에서 가장 가구수가 많은 구인 관악구(약 20만 4천 가구)의 표본비율은 9.1%인 반면, 가장 가구수가 적은 구인 중구(약 4만 8천 가구)의 표본비율은 9.2%이다. 이를 통해 서울 내에서는 구별 가구수의 크기에 상관없이 구별 표본비율은 동일한 수준으로 결정되었음을 짐작할 수 있다.

<표 7-7> 서울의 구별 표본추출률 비교

시군구	일반가구수	표본가구수	표본 조사구수	표본비율 (%)
강남구	187,294	17,115	338	9.1
강동구	143,517	13,681	301	9.5
강북구	113,037	9,766	191	8.6
강서구	178,612	16,720	344	9.4
관악구	204,392	18,538	348	9.1
광진구	127,302	12,217	342	9.6
구로구	135,436	12,638	239	9.3
금천구	87,188	7,985	172	9.2
노원구	191,785	17,886	349	9.3
도봉구	116,787	10,833	199	9.3
동대문구	132,478	11,786	278	8.9
동작구	138,179	12,900	456	9.3
마포구	136,259	12,086	243	8.9
서대문구	119,270	11,210	280	9.4
서초구	125,833	11,864	227	9.4
성동구	112,942	10,155	224	9.0
성북구	147,948	13,653	337	9.2
송파구	187,877	17,352	441	9.2
양천구	148,672	13,765	264	9.3
영등포구	139,256	12,610	293	9.1
용산구	79,340	7,146	182	9.0
은평구	145,845	13,519	328	9.3
종로구	55,748	4,990	187	9.0
중구	47,760	4,386	102	9.2
중랑구	138,595	12,845	244	9.3

한편, <표 7-8>은 서울 종로구의 동별 표본추출률을 나타내고 있는데 같은 종로구 내에서 동별로는 표본추출률이 전혀 다른 양상을 보인다. 가구수 크기와 추출률 사이에도 별다른 상관성이 존재하지 않는다. 그러므로 시군구 수준에서 추출률을 정한 이후 동 단위는 별도로 고려하지 않고 무작위로 표본조사구를 선정한 것으로 생각된다. 여기서는 종로구의 동들만 예로 들었지만 다른 지역에서도 전반적으로 마찬가지로 양상이 일어나고 있다.

<표 7-8> 서울 종로구의 동별 표본추출률 비교

동	일반가구수	표본수	추출률(%)
청운동	1,303	208	16.0
효자동	3,678	257	7.0
사직동	2,733	173	6.3
삼청동	1,268	223	17.6
부암동	3,244	268	8.3
평창동	5,387	412	7.6
무악동	1,926	248	12.9
교남동	3,301	243	7.4
가회동	1,912	217	11.3
종로1.2.3.4가동	2,163	211	9.8
종로5·6가동	2,213	183	8.3
이화동	3,648	347	9.5
혜화동	4,301	423	9.8
명륜3가동	3,272	361	11.0
창신1동	2,663	170	6.4
창신2동	3,861	334	8.7
창신3동	2,585	238	9.2
승인1동	2,311	233	10.1
승인2동	2,765	241	8.7

이상의 내용을 통해 볼 때 2005년 총조사 10% 표본조사의 시군구별 표본추출률 결정에 있어 엄밀한 통계적 고려가 다소 부족한 면이 있지 않았나 생각된다. 표본조사를 위한 표본크기 결정을 위해서는 최종 통계작성범위 별로 일정한 목표오차 수준을 정한 후 그에 적절한 표본크기를 정하는 것이 바람직하다. 이렇게 하면 상대적으로 인구가 많은 지역의 추출률은 인구가 적은 지역에 비해 낮게 나오는 것이 일반적이다. 그런데 우리나라의 경우 위에서 살펴본 바와 같이 구/시와 군의 추출률은 다소 다르지만, 같은 시도 안에서 구별 또는 시별 추출률은 가구수의 크기와 상관없이 비슷하게 정해지고 있다. 즉, 서울시 안에서는 구별 가구수의 크기와 상관없이 일률적인 추출률이 적용된 것이다. 서울시의 모든 구들의 크기가 비슷하다면 이렇게 해도 별 문제가 없다. 그러나 위의 <표 7-7>에서 알 수 있듯이 서울시의 구

들은 서로 규모가 매우 다르므로 서울 내에서도 구별 크기를 고려하여 추출률을 결정하는 것이 바람직할 것이다. 그런가 하면 같은 시/군/구 내에서 동별 추출률은 충분히 통제되지 않은 것으로 보인다.

참고로, 미국의 지역사회조사(ACS, American Community Survey)에서는 최소 통계작성범위인 county의 규모에 따라 추출률을 서로 다르게 적용하고 있다. 다음의 <그림 7-9>는 ACS의 추출률 결정 기준을 소개하는 것인데, 먼저 county의 규모와 특성에 따라 5개의 층으로 구분한 후 각 층별로 서로 다른 추출률을 적용하고 있는 것이다.

Stratum	Smallest Entity Measure of Size (SEMOS) and Tract Measure of Size (TMOS)
Blocks in large sampling entities (SEMOS >1,200) and large tracts	TMOS >2,000
Blocks in large sampling entities (SEMOS >1,200) and small tracts	TMOS ≤2,000
Blocks in small sampling entities	800 ≤SEMOS ≤1,200
Blocks in smaller sampling entities	200 ≤SEMOS <800
Blocks in smallest sampling entities.....	SEMOS < 200

Stratum	Sampling rates	
	United States	Puerto Rico
Blocks in large tracts (SEMOS >1,200, TMOS >2,000)	0.735 x BR	0.75 x BR
Blocks in small tracts (SEMOS >1,200, TMOS ≤2,000)	BR	BR
Blocks in small sampling entities (800 ≤SEMOS ≤1,200)....	1.5 x BR	1.5 x BR
Blocks in smaller sampling entities (200 ≤SEMOS <800)....	3 x BR	3 x BR
Blocks in smallest sampling entities (SEMOS <200)	10 percent	10 percent

<그림 7-9> 미국 ACS의 county별 추출률 결정 기준

위에서는 주로 지역별 표본배분(추출률)에 대한 고려가 적절했는지를 검토했는데 이제부터는 표본크기의 적절성을 살펴보기로 하자.

<그림 7-10>에는 국가통계포털(KOSIS, Korea Statistical Information System)에서 조회한 인구, 가구 총조사 관련 화면의 일부분이다. 왼쪽의 화면은 인구와 가구 부문의 메뉴이고, 오른쪽 화면은 개별 메뉴 화면을 클릭하면 나타나는 하위 단계의 메뉴이다.

<ul style="list-style-type: none"> ☐ 인구총조사 <ul style="list-style-type: none"> ☐ 인구부문 <ul style="list-style-type: none"> ☐ 총조사인구 총괄 ☐ 총조사인구(2005) <ul style="list-style-type: none"> ☐ 전수부문 ☐ 통근·통학(10% 표본) ☐ 국내인구이동통계(10% 표본) ☐ 경제활동(10% 표본) ☐ 산업, 직업(10% 표본) ☐ 활동제약(10% 표본) ☐ 여성·아동(10% 표본) ☐ 고령자(10% 표본) ☐ 가구부문 <ul style="list-style-type: none"> ☐ 총조사가구 총괄 ☐ 총조사가구(2005) <ul style="list-style-type: none"> ☐ 전수 ☐ 경제활동(10% 표본) ☐ 주거실태(10% 표본) ☐ 임차료(10% 표본) ☐ 1인가구(10% 표본) ☐ 활동제약(10% 표본) ☐ 고령자(10% 표본) 	<ul style="list-style-type: none"> ☐ 경제활동(10% 표본) <ul style="list-style-type: none"> ☐ 성/연령/혼인상태/경제활동상태별 인구(15세이상)-시도 기본조회 선택조회 ☐ 성/교육정도/경제활동상태별 인구(15세이상)-시도 기본조회 선택조회 ☐ 성/연령/산업별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/연령/직업별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/직업/산업별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/교육정도/산업별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/교육정도/직업별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 현거주지/근무지/산업별 취업자(15세이상)-시군구 기본조회 선택조회 ☐ 현거주지/근무지/직업별 취업자(15세이상)-시군구 기본조회 선택조회 ☐ 성/교육정도/종사상 지위별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/연령/혼인상태/종사상의 지위별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/종사상의 지위/근로장소별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/연령/근로장소별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/교육정도/근로장소별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 세대구성/경제활동상태별 여성인구(15세이상)-시도 기본조회 선택조회 ☐ 성/연령/경제활동상태별 인구(15세이상)-시도 기본조회 선택조회 ☐ 산업, 직업(10% 표본) <ul style="list-style-type: none"> ☐ 산업(소분류)/성/행정구역별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 직업(소분류)/성/행정구역별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/산업(중분류)/직업(중분류)별 취업자(15세이상)-전국 기본조회 선택조회 ☐ 성/산업(중분류)/연령별 취업자(15세이상)-시도 기본조회 선택조회 ☐ 성/직업(중분류)/연령별 취업자(15세이상)-시도 기본조회 선택조회
--	--

<그림 7-10> 국가통계포털(KOSIS)의 인구총조사 관련 메뉴

10% 표본조사를 통해 제공되는 통계들의 작성범위를 자세히 살펴보면 대부분 통계들의 제공범위가 시도 단위인 것을 알 수 있다. <그림 7-10>의 오른쪽 화면에서 사각형으로 표시한 '현거주지/근무지/산업별 취업자'와 '현거주지/근무지/직업별 취업자' 두 가지만 시군구별 통계가 제공되고 나머지 통계들은 모두 시도별 통계만 제공되는 상황이다. 동일한 표본조사에서 조사되는 항목들의 최소 통계작성범위가 같은 것이 바람직한데 우리나라는 그렇지 못한 실정이다.

어떤 항목에 대해서는 시도별 통계가 또 어떤 항목에 대해서는 시군구별 통계가 제공되고 하는 것은 기본적으로 작성되는 통계들의 표본오차 수준에 따라 결정된다. 다시 말해 현재 표본자료에 의해 추정했을 때 표본오차가 작은 항목들은 시군구별 세분화된 통계를 공표하는 것이고, 표본오차가 큰 항목들은 시도별 통계를 공표하는 것이다.

2005년 인구주택총조사보고서에 수록된 표본설계개요에서 표본오차 관련 설명을 살펴보면, 현재의 표본규모는 시군구별 추정값에 대한 상대표준오차 수준을 일반적으로는 25%, 경우에 따라서는 50% 상정한 표본규모임을 알 수 있다. 물론 조사변수에 따라 표본오차의 수준이 매우 다를 것이고, 변수의 특성에 따라 상대표준오차가 훨씬 적은 변수들도 많을 것인데 이를테면 시

군구별 통계를 공표하는 항목들이 그에 해당된다고 할 것이다. 반면 시도별 통계만 공표되는 항목의 경우 상대표준오차가 상대적으로 클 것이다.

2005년 표본조사의 경우 걸으로는 최소 통계작성범위가 시/군/구인 것처럼 보이지만 그것은 불과 몇 개의 항목에 국한될 뿐이고 사실상 대부분의 항목들은 최소 통계작성범위가 시도라고 할 수 있다. 그렇지 않고 명실상부하게 시/군/구를 최소 통계작성범위로 하기 위해서는 현재의 표본크기를 다소 늘리는 것이 필요하다. 앞에서 미국, 캐나다, 싱가포르, 이스라엘의 표본추출률이 우리나라의 10%보다 훨씬 높은 16% 내지 20%에 이르는 것도 이와 연관이 있는 것이다.

적절한 표본크기를 확보하기 위해서는 목표오차의 정도(precision)를 현재보다 높이는 것이 바람직하다. 보다 정확한 표본크기 계산을 위해서는 2005년 총조사의 10% 표본조사 개별 데이터를 확보하여 다양한 조사항목에 대한 시군구별 표본오차 분석을 실시하는 것이 필요하다. 경우에 따라서는 동읍면별 표본오차 분석을 실시할 수도 있을 것이다. 그러나 이러한 데이터분석은 본 과제의 범위를 벗어나는 별도의 연구과제가 될 것이므로 본 연구에서 다루기는 어렵다. 하지만 단순히 기존의 시군구별 추정값 상대표준오차 수준인 25%를 20%로 낮춘다면 현재보다 약 1.56배 이상의 표본크기가 필요하다는 계산이 나온다. 어쨌든 기존의 표본추출률 10%는 시/군/구별 추정을 위해서는 부족하며, 향후 체계적인 검토를 통해 적절한 표본규모를 마련하는 것이 필요하다.

표본크기에 관련된 이상의 검토 결과를 토대로 향후 개선방안을 요약하면 다음과 같다.

첫째, 우리나라도 미국과 같이 최소 통계작성범위인 시/군/구를 특성 및 규모에 따라 적절히 층화한 후 각 층별로 적합한 추출률을 결정하는 방안을 모색하는 것이 바람직하다. 아울러 최소 통계작성범위를 시/군/구가 아닌 동/읍/면으로 세분화할 것도 고려하여 동/읍/면 별 추출률도 비슷한 방법으로 결정하는 것이 효과적이다.

둘째, 명실상부하게 시/군/구를 최소 통계작성범위로 하기 위해서는 표본추출률을 기존의 10%보다 높이는 것이 필요하다. 표본조사 항목별로 엄밀한 표본오차 분석을 실시하여 바람직한 표본크기를 결정하는 것이 반드시 필요하다.

4. 추출방법

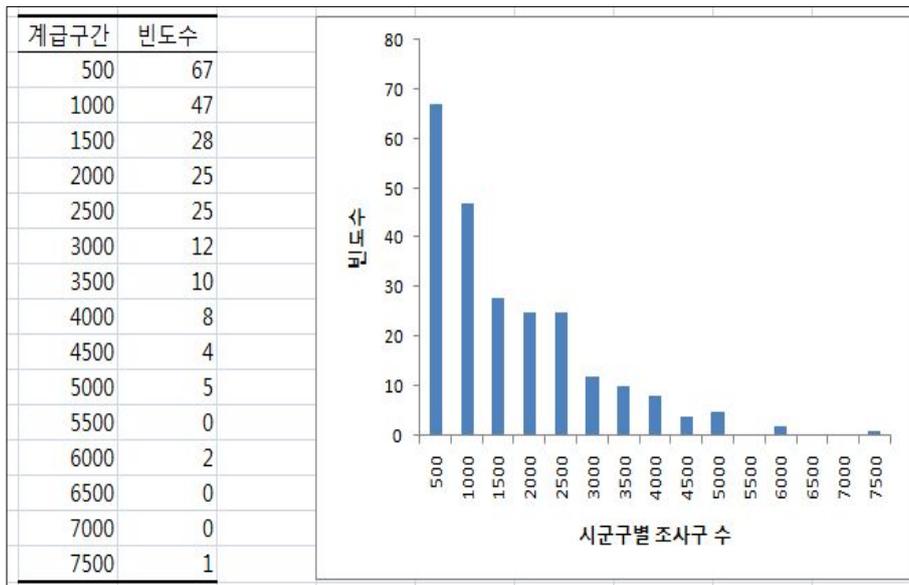
여기서는 표본조사를 위해 전국 가구의 10%를 추출하는 방안에 대해 검토해보자.

가구를 추출하는 방법으로 직접적으로 가구들 중 10%를 추출하는 방법과 조사구의 10%를 추출하는 방법을 생각할 수 있다. 참고로 2005년 총조사 때에는 조사구의 10%를 추출하는 방법을 사용하였다.

가구를 직접 추출하는 방법은 이론적으로는 가장 바람직한 방법이라 할 것이다. 하지만 이 방법은 현실적으로 여러 가지 어려움을 주게 되는데 무엇보다 먼저 조사 수행이 불편하다. 이런 경우 표본가구 추출을 위해 일반적으로 계통추출법(systematic sampling method)이 널리 쓰이고 있는데, 가구는 물리적인 건물 단위가 아니어서 동일 주택 내에도 여러 가구가 존재할 수 있으므로 계통추출법을 적용하여 조사대상 가구를 선정하는 일이 생각보다 번잡해진다. 그러다보면 조사하기 편리한 가구를 선정하는 등 다양한 형태의 비표본오차(non-sampling error)가 개입될 여지가 발생한다. 또한 지리적으로 전국 모든 지역에 표본가구가 흩어지게 되어 조사나 조사관리를 하기에 불편하다.

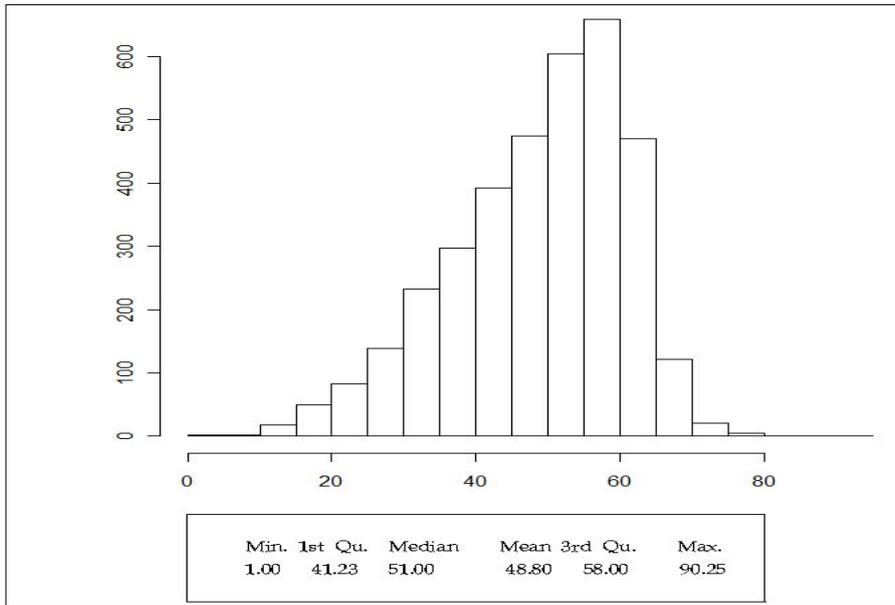
다음으로는 가구를 직접 추출하는 대신에 지역 내 조사구 중 10%를 추출하는 방법이 있다. 이 경우 표본조사구로 선정된 지역 내의 모든 가구를 다 조사하게 되므로 가구 선정에 따르는 편향의 가능성을 없앨 수 있고, 지리적인 이동거리를 줄이는 등 조사 수행 상 편리를 도모할 수 있다는 장점이 있다. 그러나 이 경우 표본크기를 결정할 때 고려한 최소 통계작성범위 지역 내의 조사구 수를 검토하는 것이 필요하다. 만일 조사구 수가 충분히 큰 경우라면 조사구의 10%를 추출하는 것이나 직접적으로 가구의 10%를 추출하는 것이나 모집단에 대한 대표성 측면에서 큰 차이가 나지 않기 때문이다.

<그림 7-11>은 우리나라 시군구별 조사구 수 크기의 분포를 나타내는 도수 분포표와 히스토그램인데, 울릉군(74개)을 제외하면 모두 100개를 넘고 중위수나 평균 모두 1000개를 넘는다. 그러므로 조사구가 적은 편인 몇몇 군을 제외하면 조사구의 10%를 추출한다고 해도 아무 문제가 되지 않는다고 할 수 있다.



<그림 7-11> 시군구별 조사구수 크기의 분포

조사구의 일부를 추출하는 방법을 채택하게 되는 경우 한 가지 주목해야 할 사실이 있는데 그것은 조사구 당 가구수가 서로 다르다는 점이다. <그림 7-12>는 2005년 총조사 표본조사구 당 가구수의 분포를 나타내고 있는데 1, 2, 3사분위수는 각각 41가구, 51가구, 58가구로 비교적 널리 퍼져 있음을 알 수 있다. 그러므로 조사구 당 가구수에 대한 고려 없이 획일적으로 조사구의 1/10을 계통추출(systematic sampling)하는 것은 바람직하지 못하다. 그보다는 조사구 크기비례 계통추출을 하되 동/읍/면이 추출을 위한 기본단위가 되게 하는 것이 효과적이다.



<그림 7-12> 표본조사구 당 가구수의 분포

제5절 연구결과 요약

각 나라별 문화와 여건에 따라 센서스 방법, 주기, 표본조사 실시 여부 등이 서로 다르다. 그러므로 우리나라에서 총조사를 등록센서스로 전환할 경우 우리 실정에 가장 적합한 방안을 마련하는 것이 필요하다. 이 장에서는 등록센서스가 실시될 경우 표본조사의 측면에서 마땅히 고려해야 할 요소들에 대해 종합적으로 검토하였는데 그 결과를 요약하면 다음과 같다.

총조사 결과는 다른 많은 표본조사를 위한 표본추출틀(sampling frame)로 활용된다. 현재 우리나라의 중요한 표본조사들에서 사용하는 추출틀은 전수조사 조사구, 10% 표본조사구, 2% 또는 5% 마이크로데이터 등이다.

전형적인 표본조사의 대표적인 예로 생각할 수 있는 통계청의 경상조사 사례를 보면, 표본설계를 위해 가장 중요한 일차적인 자료는 1차추출단위(primary sampling unit)인 조사구별 자료이다. 그러므로 추출틀의 제공이라는 측면에서 등록센서스를 바라볼 때 중요하게 고려해야 할 요소는 1차추출단위를 어떻게 구성할 것이냐 하는 점이라고 하겠다.

2005년 총조사에서는 1차추출단위로 인구주택조사구를 만들어 사용하였다. 본 연구에서는 인구주택조사구가 지니는 문제점들을 검토하였는데 ① 조사구 설정 작업의 어려움 ② 조사구 성격이 법적, 행정적으로 명확하지 않음으로 생기는 문제 ③ 자료를 이용하려 할 때 나타나는 문제들 ④ 등록센서스로 전환될 경우 적용의 어려움 등을 지적하였다.

또한 다른 나라의 사례들을 살펴봄으로 시사점을 얻을 수 있었다. 먼저 총조사와 병행하여 실시되는 표본조사의 규모를 살펴본 결과 외형적으로 우리나라의 표본추출률(sampling rate)이 낮은 것이 두드러졌다. 표본조사를 일시조사로 하는 방법과 순환조사로 하는 방법이 있을 수 있는데 각각 나름의 장단점을 지니므로 보다 면밀한 연구를 통해 우리나라 실정에 맞는 방법을 찾아야 할 것이다. 1차추출단위인 조사구를 설정하는데 있어서 각 나라는 공통적으로 ① 조사구역과 공표구역의 분리 경향 ② 공표구역 설정에서 인구가 가장 중요한 고려요소 ③ 다용도 활용을 위해 행정구역을 고려한 공표구역 설정 ④ 소지역통계 수요를 고려한 구역 설정 등의 경향을 나타내고 있다.

등록센서스에 기반한 추출틀을 작성하고자 할 때 중요한 쟁점들을 차례로 검토했는데 그 결과는 다음과 같이 요약될 수 있다.

첫째, 1차추출단위로는 기존의 인구주택조사구보다는 주민등록인구자료에서 최소 단위인 통/반/리로 하는 것이 효과적이다. 다만 현재의 통/반/리는

통계적 목적이 고려되지 않은 채 구분되었으므로 향후 통계적 목적을 반영하여 합리적으로 조정하는 것이 과제이다.

둘째, 기존의 표본조사 추출률 10%는 현재와 미래의 통계수요를 고려할 때 너무 낮은 것으로 생각된다. 최소 통계작성범위 및 목표오차 등을 합리적으로 결정한 후 그에 따라 적절한 표본크기를 새롭게 결정하는 것이 바람직한데 어쨌든 현재보다는 적어도 5% 이상은 늘려야 할 것으로 여겨진다.

셋째, 표본의 지역적 배분에 있어서도 보다 합리적인 개선이 필요하다. 현재는 공표단위별 목표오차 수준이 충분히 고려되지 않은 것으로 나타났는데 지역별 인구수 및 주요 특성을 고려하여 합리적으로 추출률을 결정할 필요가 있다.

넷째, 표본추출을 가구 수준에서 할 것인지 아니면 조사구 수준에서 할 것인지를 검토한 결과 조사구 수준에서 하는 것이 효과적일 것으로 판단되었다. 다만 조사구 크기가 충분히 크지 않은 예외적인 시군에 대해서는 별도의 고려가 필요하다. 한편, 조사구별 크기가 균등하지 않은 사정을 감안하여 조사구를 추출할 때에는 조사구 가구수 크기비례 추출법을 적용하는 것이 바람직하다.

제6절 정책적 제안

이상에서 소개한 본 장의 연구결과를 기초로 하여 다음과 같은 정책적 제안을 하고자 한다.

첫째, 표본조사의 측면에서 볼 때 1차추출단위인 조사구를 어떻게 설정하는지 하는 문제가 가장 핵심적인 사항이라고 할 수 있다. 등록센서스가 실시된다면 기존의 인구주택조사구를 설정하는 방식을 유지하는 것은 효과적이지 못하다. 그보다는 현재 주민등록인구조사 자료에 있는 최소 행정단위인 통/반/리를 활용하는 것이 여러 가지 측면에서 유용하다. 하지만 현재의 통/반/리는 행정안전부에서 나름의 기준에 의해 관리될 터인데 통계적 목적은 반영되지 않는 실정이다. 등록센서스가 본격적으로 도입될 경우, 통/반/리 구성에 있어서 통계적 목적을 반영할 수 있도록 통계청과 행정안전부가 서로 협력하는 방안을 미리 마련하는 것이 절실하게 필요하다.

둘째, 등록센서스가 도입된다면 모집단 정보가 거의 실시간으로 갱신(update)될 수 있을 것이다. 그런데 현재의 모집단 정보 관리는 기본적으로 5년 주기로 한 번씩 갱신되는 체제이다. 따라서 등록센서스 도입을 위해서는

모집단 정보 관리 차원에서 충분한 사전연구가 추가적으로 필요할 것이다. 가령, 추출틀 정보 갱신 주기, 표본조사를 순환조사로 전환하는 방안, 가중값 관리, 대표성 제고를 위한 표본보정 방안 등의 쟁점들을 예로 들 수 있다.

셋째, 등록센서스로 전환할 경우 표본조사를 통해 확보해야 할 항목들이 늘어날 가능성이 크다. 이는 표본조사의 중요성이 더욱 커진다는 말로 달리 표현할 수 있다. 그런데 현재의 표본조사 추출률인 10%는 충분하지 않은 것으로 여겨진다. 표본조사가 더욱 중요해지는 것에 걸맞게 합당한 표본조사 규모를 결정하는 것이 필요하다. 가령, 표본 추출률을 캐나다, 이스라엘, 싱가포르 등의 나라와 같이 20% 수준으로 늘리려 한다면 이에 부응하는 예산, 인력, 조직의 고려가 반드시 필요할 것이다.

제3부

등록센서스 추진을 위한
제도 · 법률적 기반개선과
단계적 추진계획 로드맵의
작성

제8장 현행 센서스와 행정자료의 주요 개념 비교분석 및 등록센서스 제도적·법률적 기반개선의 과제

제 1절 머리말

제8장의 목적은 기존의 행정자료를 이용하여 현행의 인구주택총조사를 대체할 경우에 발생할 것으로 예상되는 문제점을 명확히 하고 그 개선방법을 모색하는 것이다. 먼저 행정자료를 이용하여 인구센서스를 대체하려 할 때 두 자료간의 주요 항목의 개념의 차이가 문제가 된다. 또 행정자료와 인구센서스는 그 자료의 수집 방법이 상이한데 행정자료는 기본적으로 신고에 의한 자료이고 센서스는 조사된 자료이기 때문에 조사의 정확성에 있어서 행정자료는 한계를 지닐 수밖에 없다. 또 행정자료를 통계목적으로 활용하려 할 때 개개인의 정보 보호라는 문제와도 직면하게 된다. 따라서 이 장에서는 우선 이와 같은 여러 문제들을 검토한 후 이들 문제점을 해결할 수 있는 기술적, 행정적, 법적, 제도적 개선방법을 제시하려 한다. 또 개선방법의 모색에 있어서 우리보다 먼저 인구센서스 대신에 행정자료를 활용하고 있는 외국의 사례도 검토하게 될 것이다.

제2절 행정자료와 현행 인구주택총조사 자료의 개념 비교분석

이 절에서는 인구주택총조사를 대체할 수 있는 행정자료의 신고 항목과 인구주택 총 조사의 조사 항목간의 개념상의 괴리가 있는지를 검토하려 한다. 현재 인구 및 주택센서스에 포함되어 있는 조사 항목 중 기존의 행정자료로 대치가 가능한 항목은 대체로 14개 내외로 추정된다¹⁾. <표 8-1>은 통계청의 자체 조사 결과 현행의 행정조사자료로 인구주택 총 조사의 조사 항목을 대체할 수 있는 항목을 정리한 것이다.

1) 제5장에서는 2015년 우리나라의 등록센서스에서 행정자료를 대체할 수 있는 조사항목수를 전수조사의 경우 11개, 표본조사의 경우 6개로 모두 17개 정도로 파악하고 있다. 통계청 인구조사과의 내부자료는 대체가능 항목을 14개 내외로 보고 있어서, 약간 보수적인 대체항목을 설정하고 있다고 볼 수 있다.

<표 8-1> 등록센서스 실시가능항목(14개 항목)

구 분	인구주택총조사 항목	행정자료 항목	자료명 <기 관>
가구원 (9)	성명, 성별, 나이 가구주와 관계 1년전 거주지 5년전 거주지	성명, 성별, 나이 세대주와 관계 1년전 주소 5년전 주소	주민등록자료 <행정안전부>
	혼인상태, 본관, 출생지	혼인상태, 본관, 출생지	가족관계등록부 <대법원>
가 구 (2)	가구구분	세대주와 관계	주민등록자료 <행정안전부>
	주택소유여부	소유자 현황	건축물대장 <국토해양부>
주 택 (3)	거처의 종류 주거용 연면적 건축년도	주용도, 기타용도 주용도, 기타용도 면적 사용승인일자	건축물대장 <국토해양부>

자료: 통계청 (2008a), 등록센서스 실시 기본계획, 인구조사과 내부자료

표에서 행정자료별로 이용 가능한 항목을 보면 주민등록 자료가 7개 항목으로 성명, 성별, 나이, 가구주와의 관계, 1년전 거주지, 5년전 거주지, 가구구분 등이며, 가족관계등록부를 이용하면 혼인상태, 본관, 출생지 정보를 얻을 수 있다. 한편 주택에 관한 정보로는 건축물대장에서 거처의 종류, 주거용 연면적, 건축년도, 주택소유여부 등을 알 수 있다. 이제 이 장에서는 먼저 이들 행정조사자료의 항목과 인구주택 총 조사 항목의 개념을 비교하여 개념상의 차이점이 없는지를 검토하려 한다.

1. 행정자료 현황

가. 주민등록부

주민등록제도는 행정기관이 그 관할 구역 내에 주소 또는 거소를 둔 주민을 등록토록 함으로서 주민의 거주관계 등 이동 실태를 상시로 명확히 파악하여 주민생활의 편익을 증진시키고 행정사무의 적정한 처리를 도모하기 위한 제도이다. 따라서 주민등록제도는 주민의 거주이동 실태를 거주지에서 기록 관리하는 제도로서 국민의 혈연적 신분관계를 등록기준지에서 기록하는 가족관계등록제도와 구분된다.

이 주민등록제도의 근거가 되는 주민등록법 제 1조는 주민등록의 목적을 주민의 거주관계 등 인구의 동태(動態)를 항상 명확하게 파악하여 주민생활의 편익을 증진시키고 행정사무를 적정하게 처리하기 위한 것으로 규정하고 있다. 주민등록법은 1962년 5월에 기존의 기류법(寄留法)을 대체하여 제정되었고 1968년의 1차 개정을 시작으로 하여 2009년에 18번째로 개정되어 오늘에 이르고 있다.

현재 주민의 신고에 의해서 주민등록부에 기재되는 사항은 (1) 성명, (2) 성별, (3) 생년월일, (4) 세대주와의 관계, (5) 합숙하는 곳은 관리 책임자, (6) 가족관계의 등록 등에 관한 법률 제 10조 제 1항에 따른 등록기준지, (7) 주소, (8) 가족관계등록이 되어 있지 아니한 자 또는 가족관계등록의 여부가 분명하지 아니한 자는 그 사유, (9) 대한민국의 국적을 가지지 아니한 자는 그 국적명이나 국적의 유무, (10) 거주지를 이동하는 경우에는 전입전의 주소 또는 전입지와 해당 연월일, (11) 대통령으로 정하는 특수 기술에 관한 사항 등이다(주민등록법 제 10조).

나. 가족관계등록부

가족관계등록부는 종전의 호적법을 대신하여 2008년 1월 1일부터 발효된 가족관계의 등록 등에 관한 법에 기초하여 국민의 혈연적 신분관계를 나타내 주는 기록이다. 가족관계등록부에는 시·읍·면의 장이 관리하는 전산정보처리체계에 의하여 (1) 등록기준지, (2) 성명, 본, 성별, 출생 연월일 및 주민등록번호, (3) 출생, 혼인, 사망 등 가족관계의 발생 및 변동에 관한 사항, (4) 그 밖의 가족관계에 관한 사항을 기재하도록 되어 있다. 이전의 호적에서는 본적이 호주의 주소이기 때문에 본인의 의사와 관계없이 호주의 주소를 따랐으나 가족관계등록부에서는 본인의 선택에 의하여 등록기준지를 변경할 수 있도록 하였다.

가족관계등록부의 기록사항에 근거하여 발급이 가능한 증명서는 모두 다섯 종류인데 (1) 가족관계증명서 (본인의 등록기준지, 성명, 성별, 본, 출생 연월

일 및 주민등록번호, 부모, 양부모, 배우자, 자녀의 성명, 성별, 본, 출생 연월일 및 주민등록번호), (2) 기본증명서 (본인의 등록기준지, 성명, 성별, 출생 연월일 및 주민등록번호, 본인의 출생, 사망, 국적상실 취득 회복 등에 관한 사항), (3) 혼인관계증명서 (본인의 등록기준지, 본인 및 배우자의 성명, 성별, 본, 출생 연월일 및 주민등록번호, 혼인 및 이혼에 관한 사항), (4) 입양관계 증명서(본인의 등록기준지, 본인 및 양부모 또는 양자의 성명, 성별, 본, 출생 연월일 및 주민등록번호, 입양 및 파양에 관한 사항), (5) 친양자(親養子) 입양관계증명서(본인의 등록기준지, 본인과 친 생부모, 양부모 또는 친양자의 성명, 성별, 본, 출생 연월일 및 주민등록번호, 입양 및 파양에 관한 사항) 등이다(통계교육원, 2008).

다. 건축물대장

건축물대장은 적법하게 건축된 건축물에 대하여 건물의 소재, 번호, 종류, 구조, 건평, 소유자의 주소, 성명 등을 등록하여 건물의 상황을 명확하게 할 목적으로 작성된 공적 장부이다. 건축물대장은 건축물 1동(棟)을 단위로 하여 각 건축물마다 작성하고, 부속건축물이 있는 경우 부속건축물은 주된 건축물 대장에 포함하여 작성된다.

건축물대장은 소유권의 구분가능여부에 따라 일반건축물대장과 집합건축물대장으로 나뉘어 작성되는데, 일반건축물대장은 일반건축물에 해당하는 건축물 및 대지에 관한 현황을 기재한 건축물대장이고, 집합건축물대장은 집합건축물에 해당하는 건축물 및 대지에 관한 현황을 기재한 건축물대장이다.

건축물대장의 기재 및 관리 등에 관한 규칙 제5조에 의거하여 집합건축물대장은 표제부와 전유부(專有部)로 나뉘어 작성되고 하나의 대지에 2개 이상의 건축물(부속건축물을 제외)이 있는 경우에는 총괄표제부를 작성하도록 하고 있다. 건축물의 소유자는 건축물의 전부 또는 일부가 철거, 멸실 등으로 없어진 경우에는 건축물대장 말소신청서를 작성하고 읍·면·동장의 확인서를 첨부하여 건축물대장의 말소를 신청하여야 하는데 기존 건축물대장을 말소하거나 폐쇄하려는 때에는 건축물대장 변동사항 난에 그 사유 및 말소, 폐쇄일자를 기재한다. 이 건축물대장은 관할 시장, 군수, 구청장이 관리한다. 건축물대장에 수록되어 있는 자료는 다음의 <표 8-2>와 같다.

<표 8-2> 건축물 대장별 주요 수록자료

일반건축물대장	집합건축물대장		총괄표제부
	표제부	전유부	
·고유번호 ·대지위치 ·지번 ·명칭 및 번호 ·대지면적, 연면적, 지역, 지구, 구역, 건축면적, 용적률 산정용 연면적, 주구조, 주용도, 층수, 건폐율, 용적률, 높이, 지붕, 부속건축물의 동수와 면적, 층별 구조·용도·면적, 특이사항 ·건축주 성명 및 면허번호, 설계자 성명 및 면허번호 공사감리자 성명 및 면허번호, 공사시공자 성명 및 면허번호 ·주차장 대수와 면적(옥내/옥외, 자주식/기계식), 승강기 대수, 오수정화시설 형식과 용량, 허가일자, 착공일자, 사용승인일자 ·소유자 현황 소유자 성명·소유자 주소·소유권 지분·소유자 변동일자과 원인 ·건축물현황도	·고유번호 ·대지위치 ·지번 ·명칭 및 번호 ·대지면적, 연면적, 지역, 지구, 구역, 건축면적, 용적률 산정용 연면적, 주구조, 주용도, 층수, 건폐율, 용적률, 높이, 지붕, 부속건축물의 동수와 면적, 층별 구조·용도·면적, 특이사항 ·건축주 성명 및 면허번호, 설계자 성명 및 면허번호 공사감리자 성명 및 면허번호, 공사시공자 성명 및 면허번호 ·주차장 대수와 면적(옥내/옥외, 자주식/기계식), 승강기 대수, 오수정화시설 형식과 용량, 허가일자, 착공일자, 사용승인일자 ·변동사항 및 변동내역 ·건축물현황도	·고유번호 ·대지위치 ·지번 ·명칭 및 번호 ·대지면적, 연면적, 지역, 지구, 구역, 건축면적, 용적률 산정용 연면적, 건축물수, 주용도, 층수, 건폐율, 용적률, 층호수, 총주차대수, 부속건축물 동수와 면적 ·건축물 현황 : 건축물 명칭 건축물 주구조, 건축물지붕, 층수, 용도, 연면적, 변동일자, 변동원인 ·주차장 대수와 면적(옥내/옥외, 자주식/기계식), 승강기대수, 오수정화시설 형식과 용량 ·건축물 현황도	

자료: 통계청(2007b) 건축물대장 활용방법 연구, 인구조사과 내부자료.

2. 행정자료와 인구주택총조사 항목의 개념비교

가. 주민등록부

인구주택총조사에서 조사되는 항목 중 주민등록부를 이용하여 대체할 수 있는 자료로는 성명, 성별, 나이, 가구주와의 관계, 1년 전 거주지, 5년 전 거주지, 가구구분 등이 있다. 이 중에서 성명과 성별은 개념정의에서 차이가 없기 때문에 제외하고 여기에서는 나이와 가구주와의 관계에 대해서 각각의 정의와 개념을 서로 비교하려 한다.

1) 나이

먼저 인구주택총조사에서는 가구원 모두의 나이를 조사하는데 주민등록(또는 호적)상의 나이와 상관없이 집에서 세는 나이와 띠를 물어보고 동시에 생년월일도 조사하는데 양력인지 음력인지까지도 표시하도록 하여 매우 정확하게 현재의 나이를 파악하고 있다. 혹시 나이, 띠, 생년월일이 불일치하는 경우에는 부호집에 수록되어 있는 '연령조건표'를 참조하여 다시 확인한 후 수정한다. 이처럼 총조사에서는 인구학적 변수로서 연령이 중요하기 때문에 가능한 한 정확하게 조사하려고 한다.

이에 비해서 주민등록부상의 나이는 본인들의 신고에 의해서 기재된 나이이기 때문에 총조사에 비하면 그 신뢰성이 많이 떨어진다. 특히 과거에는 출생신고가 늦어지는 경우가 많아 실제 나이가 주민등록부에 기재된 나이보다 많은 사람들이 상당수에 달하고 있다. 실제로 통계청에서 2008년에 실시한 주민등록 확인조사 결과를 보면, 실제 연령과 주민등록 연령의 일치율은 80.4%로 나타났는데, 남자가 82.1%로서 여자의 78.8% 보다 3.3% 포인트 더 높게 나타났다(김형석 외, 2008).

주민등록부상의 나이와 실제 연령과의 일치율을 출생년도별로 살펴보면, 1950년대 이전 출생자는 60%대 일치율을 보였고, 1970년대는 85.4%, 1980년대는 95.2%를 보였으나, 특히 1990년 이후 출생자는 97.9%로서 최근 출생자의 경우는 실제 연령과 주민등록 연령이 거의 일치하는 것으로 나타났다. 또 2008년 주민등록 확인조사 결과, 연령 차이가 ± 1 세 이내는 94.4%, ± 2 세 이내는 97.8%, ± 3 세 이내는 99%로 나타났다. 전체적으로는 주민등록 연령이 실제 연령보다 많게 기재된 경우는 5.4%이고 주민등록연령이 실제 연령보다 적게 기재된 경우는 14.2%로서 월등히 더 많았다

<표 8-3> 성 및 출생년도별 주민등록 연령 일치율

(단위: %)

	계	~1919	'20~29	'30~39	'40~49	'50~59	'60~69	'70~79	'80~89	'90~07
계	80.4	68.5	65.8	60.2	66.8	66.2	72.1	85.4	95.2	97.9
남	82.1	71.1	71.1	62.6	68.3	67.1	73.8	87.6	95.5	97.8
여	78.8	67.9	63.2	58.4	65.5	65.3	70.5	83.3	95.0	98.0

자료: 김형석 외 (2008)

주: 1) 2008년 주민등록 확인조사(통계청) 결과

2) 주민등록 출생년도의 응답거부와 연령차가 7세를 초과하는 자료는 제외

<표 8-4> 주민등록 성별 연령차이 분포

(단위: 명, %)

		연령차이(주민등록연령-실제연령)												
		-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
계	83,733	107	167	343	852	2,392	8,070	67,318	3,690	404	180	101	75	34
남	40,182	40	60	109	342	1,028	3,714	33,001	1,602	147	70	31	25	13
여	43,551	67	107	234	510	1,364	4,356	34,317	2,088	257	110	70	50	21
계	100.0	0.1	0.2	0.4	1.0	2.9	9.6	80.4	4.4	0.5	0.2	0.1	0.1	0.0
남	100.0	0.1	0.1	0.3	0.9	2.6	9.2	82.1	4.0	0.4	0.2	0.1	0.1	0.0
여	100.0	0.2	0.2	0.5	1.2	3.1	10.0	78.8	4.8	0.6	0.3	0.2	0.1	0.0

자료: 김형석 외 (2008)

주: 1) 2008년 주민등록 확인조사(통계청) 결과

2) 주민등록 연령 미상과 연령차의 절대값이 7세를 초과하는 자료는 제외

2) 가구주와의 관계

인구주택총조사에서는 가구를 1인 또는 2인 이상이 모여서 취사, 취침 등 생계를 같이 하는 생활단위로 정의한다. 즉 총조사는 가구를 가사단위로 보고 가구를 정확히 구분하기 위해서 조사원 지침에서 이를 자세히 언급하고 있다 (통계청, 2005). 가구주에 대해서는 인구주택총조사에서는 호적상의 호주 또는 주민등록상의 세대주와 관계없이 가구를 실질적으로 대표하는 사람을 가구주로 하며, 한국인과 외국인이 함께 사는 경우에는 한국인이 가구주가 되도록 하고 있다. 또 혈연 또는 친족관계가 없는 사람이 2인 이상 함

게 사는 경우 그 중 한 사람(연장자 또는 대표자)을 가구주로 하고, 나머지 사람에게 대해서는 동거인으로 표시하도록 한다.

또 가구원도 한 가구에서 함께 살고 있는 모든 구성원으로 정의하면서 주민등록이 함께 되어 있더라도 같이 살고 있지 않으면 가구원에서 제외하며, 가족이라도 군복무, 취업, 교육 때문에 함께 살고 있지 않으면 가구원에서 제외한다. 거꾸로 주민등록은 함께 되어 있지 않더라도 같이 살고 있으면 가구원에 포함시키는데 예컨대 먼 친척, 친구, 가정부, 종업원 등 가족이 아닌 사람이 함께 살고 있으면 가구원에 포함시켜 조사하고 있다.

이와는 반대로 가구에서 제외되는 사람은 취업, 취학 등으로 외지에서 하숙, 자취하거나 기숙사에 있는 가족, 고아원, 양로원, 모자원, 특수병원, 부녀 보호소 등의 사회복지시설에 수용되어 있는 가족, 요양원, 기도원 등에서 장기간 요양 중인 가족, 군대, 전투경찰대에 입대한 가족, 형이 확정되어 교도소, 소년원 등에 수용된 가족, 선박, 항공기, 철도, 버스 등의 승무원으로 기숙사에서 생활하고 있는 가족, 구직, 무단가출 등으로 집을 떠난 지 1개월이 넘는 가족 등이다.

총조사에서는 이렇게 정의된 가구를 일반가구, 집단가구, 그리고 외국인가구로 나누고 있는데 일반가구는 가족으로 이루어진 가구, 가족과 5인 이하의 남남이 함께 사는 가구, 가족이 아닌 5인 이하의 남남이 함께 사는 가구, 1인가구 등이 포함된다. 집단가구는 가족이 아닌 6인 이상 19인 이하의 가구, 기숙사나 고아원·양로원·보육원 등 사회시설에 집단으로 살고 있는 가구인데 20인이 넘는 경우에는 집단가구 일지라도 기숙시설조사구로 분류된다. 외국인 가구는 외국인으로만 구성된 가구인데 한국인과 외국인이 함께 사는 가구는 일반가구로 분류된다.

이에 비해서 주민등록부에서는 가구 대신에 세대라는 용어를 사용하는데 주민등록법이나 그 시행령 어디에서도 세대를 정의하고 있지 않다. 다만 주민등록 업무지침을 보면 세대란 “주거 및 생계를 같이하는 집단을 의미하며, 세대는 일정한 친족을 중심으로 구성되며 타인(동거인 등)이 포함되는 경우도 있다”라고 규정한다(주민등록업무지침). 따라서 주민등록부의 세대 개념과 총조사의 가구개념은 동일한 주거지와 생활단위를 강조하는 측면에서는 비슷하다.

다만 주민등록의 세대는 일정한 친족을 중심으로 구성된다고 하고 있어 혈연관계를 강조한다면 총조사의 가구는 실질적인 생활단위라는 측면을 강조하는 면에서 약간의 차이점을 발견할 수 있겠다. 여기에 비해서 소득세법 시행령 등에서 사용하는 세대개념은 가족개념을 더 강조한 것으로 볼 수

있다.

한편 주민등록부의 경우에 세대주는 그 세대를 대표하는 자로서 주민의 신고에 의하여 선정이 가능한데 다만 성년이 있는 세대에서 고의적으로 미성년자를 세대주로 선정하여 기재하는 것은 금지하고 있다. 이때 세대주는 주민등록신고의 철저한 이행을 위하여 주민등록신고 의무자가 된다. 또 주민등록법 시행령 제6조 2항은 세대별 주민등록표의 등재순위를 정하고 있는데 세대주, 배우자, 세대주의 직계 존비속의 순위로 하고 그 외에는 세대주의 신고에 따르도록 하고 있다.

또 현행 주민등록제도는 세대를 주민등록의 단위로 하여 주민등록표를 작성하기 때문에 동일한 세대에 속하는 자는 모두 세대원이라고 정의한다. 주민등록의 대상자는 시·군·구 관내에 30일 이상 거주할 목적으로 주소 또는 거소를 가진 자 인데 단 30일 이상을 살아도 거주할 목적이 없으면 (예컨대 장기 업무 출장자 등) 제외된다. 주민등록 장소는 30일 이상 거주하게 될 읍·면·동인데 한 가지 특이한 것은 영내에 기거하는 군인은 그가 속한 세대의 거주지에서 등록하도록 하고 있다는 점이다. 이는 총조사에서는 영내에 거주하는 군인은 그 영내에서 조사되는 것과는 다른 점이다.

또 총조사에서 가구를 일반가구, 집단가구, 외국인가구로 구분하는데 비해서 주민등록부에서는 세대를 일반세대와 합숙사로만 구분한다. 이때 합숙사란 기숙사나 그 밖에 여러 사람이 동거하는 숙소로 정의한다(주민등록법 제12조)..

한편 주민등록부에 기재된 세대주와의 관계는 <표 8-5>에서 보는 바와 같이 총조사의 분류 코드에 맞추어서 조정이 가능하다.

<표 8-5> 세대주와의 관계 재분류

주민등록부	인구주택총조사	
세대주와의 관계	가구주와의 관계	코드
본인	가구주	①
처, 남편	가구주의 배우자	②
자	자녀	③
자부, 사위	자녀의 배우자	④
부, 모, 양부, 양모	가구주의 부모	⑤
장인, 장모, 시부, 시모	배우자의 부모	⑥
손, 외손, 손부, 손서	손자녀, 그 배우자	⑦
증손	증손자녀, 그 배우자	⑧
조부, 조모, 장조부, 장조모, 시조모, 외조부, 외조모	조부모	⑨
형, 제, 누이, 매, 오빠, 형수, 제수, 처남, 처제, 형부, 매부, 제부, 처형, 처남택, 시누이, 올케, 시숙, 동서	형제자매, 그 배우자	⑩
조카, 처조카	형제자매의 자녀, 그 배우자	⑪
백부, 백모, 숙부, 숙모, 고모, 고모부, 이모, 이모부, 외숙, 외숙모, 처고모, 처이모	부모의 형제자매, 그 배우자	⑫
친척, 외가친척, 이종, 시가친척, 이질, 처가친척, 질부, 종제, 외종, 고종, 종형, 종제수, 종형수, 당질, 처이종, 종매	기타 친인척	⑬
동거인	기타 동거인	⑭

자료: 통계청 (2007d), 2010 인구주택총조사 제1차 시험조사 행정자료 활용 결과보고, 인구조사과 내부자료

나. 가족관계등록부

인구주택총조사를 대체하기 위해서 가족관계등록부에서 얻을 수 있는 자료는 혼인상태, 본관, 출생지이다.

이중에서 본관과 출생지항목은 그 개념에 있어서 행정자료나 인구주택 총조사 자료 간에 차이가 없으나 혼인상태의 개념에는 약간의 괴리가 있다. 먼저 인

구주택총조사에서는 혼인신고와 관계없이 사실상의 혼인상태를 기준으로 조사한다. 따라서 이혼은 혼인 후 배우자와 헤어져서 재혼하지 않고 혼자 살고 있는 경우를 말하는데 이혼신고는 되어 있지 않더라도 사실상 이혼상태에 있는 별거도 이혼에 포함시키고 있다. 이에 반해서 주민등록부는 혼인이나 이혼 등 신고에 기초해서 혼인상태를 표시하기 때문에 실질적으로는 혼인이나 이혼 상태에 있지만 신고가 되어 있지 않으면 미혼이나 배우자 있음으로 분류될 수밖에 없다.

다. 건축물대장

1) 거처의 종류

한편 주택에 관한 정보로서 건축물대장으로 부터는 거처의 종류, 주거용 연면적, 건축년도, 주택소유여부 등을 알 수 있다. 먼저 인구주택총조사에서는 거처란 사람이 살고 있는 모든 장소를 의미하는데, 구조적으로 분리되고 독립된 하나의 거주단위로서 모든 거처는 크게 주택과 주택 이외의 거처로 구분한다.

주택은 가구가 독립적으로 살림을 할 수 있도록 지어진 집인데 주택으로 분류될 수 있는 요건은 (1) 영구 또는 준 영구 건물로서, (2) 부엌과 한 개 이상의 방을 갖추고 있으며, (3) 독립된 출입구가 있고, (4) 관습상 소유 또는 매매의 한 단위를 이루고 있을 것 등이다.

주택은 다시 단독주택, 아파트, 연립주택, 다세대주택, 비거주용건물 내 주택으로 분류된다. 단독주택은 일반단독, 다가구단독, 영업겸용단독을 다 포함한다. 아파트는 5층 이상의 공동주택을 말하는데 4층 이하라도 아파트로 허가 받았으면 아파트로 간주한다. 연립주택은 4층 이하의 공동주택으로 건축 당시 연립주택으로 허가 받은 주택으로 연 면적이 660 제곱미터가 넘는 주택이다.

다세대주택은 4층 이하의 공동주택으로 건축 당시 다세대주택으로 허가 받은 660제곱미터 이하의 공동주택이다. 비거주용건물 내 주택은 비거주용건물에 사람이 살되 그 거주부분이 주택의 요건(방, 부엌, 독립된 출입구) 을 갖춘 경우이다

이와 같은 총조사에서의 거처의 분류와 정의에 대해서 건축물대장은 용도별로 건축물의 종류를 구분하는데 모두 28개로 분류하고 있다. 건축물대장에서는 건축물을 토지에 정착하는 공작물중 지붕과 기둥 또는 벽이 있는 것

이에 부수되는 시설물, 지하 또는 고가의 공작물에 설치하는 사무소, 공연장, 점포, 차고, 창고 기타 대통령령이 정하는 것을 말한다(건축법 제 2조 2).

이중에서 총조사의 거처에 해당하는 건축물로는 단독주택과 공동주택의 분류가 있다. 먼저 단독주택은 다시 세분해서 (1) 단독주택, (2) 다중주택, (3) 다가구주택, (4)공관으로 구분하고 있는데 이 중에서 총조사에서는 사용하지 않는 분류인 다중주택은 여러 사람이 장기간 거주할 수 있는 구조로 되어있으나 독립된 주거의 형태를 갖추지 아니한 것으로 연면적이 330 제곱미터 이하이고 층수가 3층 이하인 건축물을 말한다(건축법 시행령 제 3조 4 관련 별표 1). 또 건축물대장에서는 공관을 별도의 범주로 구분하는 것이 특이하다.

한편 공동주택은 (1) 아파트, (2) 연립주택, (3) 다세대주택, (4) 기숙사로 구분하는데 아파트, 연립주택, 다세대주택의 정의는 총조사의 정의와 동일하다. 다만 기숙사는 학교, 또는 공장 등의 학생 또는 종업원 등을 위하여 쓰는 것으로서 공동취사 등을 할 수 있는 구조를 갖추되 독립된 주거의 형태를 갖추지 아니한 것으로 정의한다.

따라서 이와 같은 건축물대장에서의 주택의 분류를 인구주택총조사의 분류에 맞추기 위해서는 <표 8-6>에서 보는 바와 같은 약간의 재조정이 필요하다.

<표 8-6> 건축물대장의 주택 재분류

건축물 대장		인구주택총조사	
단독 주택(1)	- 단독주택(1-1) - 다가구주택(1-2) - 다중주택(1-4) - 공관(1-5)	단독 주택(1)	- 일반단독(1-1) - 다가구단독(1-2) - 영업겸용단독(1-3)
공동 주택	- 아파트(2) - 연립주택(3) - 다세대주택(4) - 기숙사(8)	아파트(2) 연립주택(3) 다세대주택(4) 비거주용건물(상가, 공장, 여관 등) 내 주택(5)	

자료: 통계청(2007b), 건축물대장 활용방법 연구, 인구조사과 내부자료

즉 건축물대장의 다중주택(1-4) 및 공관(1-5)은 인구주택총조사의 일반단독(1-1)으로 재분류하고, 비거주용건물 내 주택은 영업겸용단독(1-3) 및 비거주

용건물 내 주택(5)으로 재분류가 필요하다. 이때 건물별 주택 면적의 합이 건물 연면적의 1/2이상이면 영업겸용단독주택, 그 미만이면 비거주용건물 내 주택으로 구분한다.

따라서 인구주택총조사에서 사용하는 거처의 종류를 파악하기 위해서 건축물대장을 활용하는 데 있어서 개념상의 차이에서 오는 문제는 거의 없다고 보여 진다.

2) 건축년도

인구주택총조사에 있어서 건축년도는 신축시기를 원칙으로 하나 증축, 개축면적이 주택 총면적 (연건평) 의 반 이상이 되는 경우에는 증축, 개축시기를 표시하도록 하고 있다. 또 여러 차례 증축되었을 경우에는 가장 최근의 증축시기를 표시한다. 이에 비해서 건축물대장에서는 사용승인 일자를 건축년도로 본다. 그런데 만약 증축이나 개축이 있을 경우에는 건축물 대장에 증, 개축시기와 증, 개축 면적이 별도로 표시된다. 따라서 총조사의 정의에 맞추기 위해서는 건축물 대장에서 증, 개축 여부와 그 면적을 따져서 건축시기를 조정해야 한다.

제3절 등록센서스 추진에 따른 행정자료의 한계성

인구주택총조사 자료를 행정자료로 대체하는 데에는 앞 절에서 살펴 본 두 자료간의 개념상의 차이의외에도 다음의 두 가지 문제점이 부수된다. 우선 첫째는 기본적으로 인구주택총조사는 현지의 실사에 기초한 조사자료인데 반해서 행정자료는 행위 당사자들의 신고에 기초한 자료이기 때문에 자료 생성방법의 차이에서 발생하는 문제점이다. 둘째는 현행 행정자료의 정비와 관련된 신뢰성의 문제이다. 이제 이들 각각의 문제점에 대해서 자세히 설명함으로서 행정자료의 한계성을 살펴보려 한다.

1. 신고된 자료와 실제 조사자료 간의 불일치

현재로서는 인구주택총조사 자료를 행정자료로 대체하는데 있어서 가장 유용성이 큰 자료가 주민등록자료이다. 그런데 주민등록자료는 기본적으로 신고자료이기 때문에 여기에 따르는 많은 문제점이 많이 생겨나고 있다. 우선

가장 큰 문제는 주민등록신고지와 실제 거주지와 불일치이다. 이와 같은 불일치의 원인은 세금과 주택청약, 그리고 부동산취득 등의 이익을 위하여 허위로 등록하는 경우, 또 개인의 경우 실제로는 이동을 하였지만 신고를 하지 않아 등록이 계속 전거주지로 되어 있는 경우, 또 해외 이주자의 경우 해외에서 영주권을 취득하여 출입국관리사무소에서 명단이 통보되어서 주민등록이 말소되기 까지 계속 주민등록에 기재되어 있는 경우 등 여러 가지가 있다.

그러면 이와 같은 여러 원인에 의해서 실제로 인구주택총조사 결과와 주민등록상의 신고 인구 간에는 어느 정도의 괴리가 있는가? 우선 통계청의 조사에 의하면 2005년 11월 1일 기준의 총조사인구는 47,279천명인데 비해서 2005년 11월 30일 현재 주민등록인구는 49,210천명으로 집계되어 전국적으로 보면 주민등록인구가 총조사인구보다 1,931천명 더 많은 것으로 드러나 4.1%의 차이를 나타낸다.(통계청, 2008a).

이처럼 전국적으로 총인구에서 차이가 나는 것은 총조사에서는 누락이 있었을 것이며 주민등록인구에서는 인구센서스에서는 제외된 해외장기거주자가 계속 집계되는 것 때문이다. 실제로 최근의 한 연구는 2006년 1월 1일 현재 주민등록은 되어 있으나 해외에 거주하는 한국인이 117만 8천명에 이르고 있는 것으로 추정하였다 (김형석 외, 2008). 이들을 연령별로 보면 15세 미만이 17.5%로 가장 많고 25-29세는 11.7%, 그리고 30-34, 35-39세는 각각 11.5%를 점한다. 또한 이 연구는 출생신고의 지연으로 1만 3천명이, 또 주민등록의 직권말소로 26만 9천여 명이 2006년 1월 1일 현재 주민등록에 과소 등재되었고 거꾸로 사망신고의 지연으로는 1만 9천명이 과대 집계된 것으로 추정한다.

한편 또 다른 연구는 2005년 11월 1일 현재의 2005년 인구주택총조사 인가와 2006년 1월 1일 현재의 주민등록 인구를 비교하였는데 총조사인구는 47,041천명, 그리고 주민등록인구는 48,782 천명으로 주민등록인구가 총조사인구보다 1,741천명이 많아 주민등록인구에 대한 총조사인구의 비율이 96.4%라고 보고하였다 (통계청, 2007d). 그런데 흥미롭게도 이 주민등록인구대비 총조사인구의 비율은 그 전 2000년과 1995년에도 비슷하게 나타나 각각 96.0%, 96.9%를 기록하였다.

이 연구는 또 행정단위별로 이 주민등록인구대비 총조사인구의 비율을 비교하였는데 16개 시·도별로 보면 이 비율의 평균이 96.5%, 그리고 표준편차는 1.9%임을 보고하고 있다. 234개의 시·군·구별로 보면 평균비율은 93.8%, 표준편차는 5.2%로 시·도 수준에 비해 평균은 2.7% 포인트 감소하고 표준

편차는 3.3% 포인트 증가하였다. 시·군·구별로 보았을 때 두 인구간의 차이가 가장 큰 지역은 21.4%의 차이도 보이고 있었다.

이제 지역을 더 좁혀서 읍·면·동 수준에서 보면 주민등록인구대비 총조사인구의 평균 비율은 93.7%, 표준편차는 12.0%, 중간 값은 94.8%로 나타나 시·군·구 수준과 비교하면 평균과 중간값은 비슷하나 표준편차는 2배가량 증가하여 소지역으로 갈수록 두 인구간의 괴리가 점점 더 커지고 있음을 알 수 있다. 이 두 인구의 차이가 심한 곳은 70% 가까이 까지 차이가 나고 있었다.

이상에서 살펴 본 것처럼 전국적으로 주민등록인구와 총조사인구 간에 발생하는 괴리의 가장 큰 원인은 해외장기체류자들로서 이들은 주민등록인구에는 포함되지만 인구총조사에서는 조사되지 않기 때문이다. 즉 현재 해외에 3개월 이상 거주할 목적으로 출국하는 사람은 장기이동자로 간주되어 인구총조사에 포함되지 않는다. 반면에 주민등록에서는 유학이나 취업 등의 이유로 해외에 3개월 이상 체류하여도 해외이주법에 의하여 이민을 가지 않는 한 주민등록은 계속 유지되고 등록지에서 집계된다. 따라서 인구주택총조사의 대체자료로서 주민등록부를 사용하려면 이 해외 장기 체류자들을 별도로 확인하여 주민등록에서 이를 감안하는 작업이 필수적으로 이루어져야 한다.

<표 8-7> 주민등록세대와 총조사가구의 차이

(단위: 가구)

	주민등록세대(A) (2005.11.30일)	총조사가구(B) (2005.11.1일)	차이(A-B)
계	17,814,951	15,887,128	1,927,823 (12.1%)
- 1인가구	5,139,399	3,170,675	1,968,724 (62.1%)
- 2인가구	3,097,058	3,520,545	- 423,487
- 3인가구	3,367,149	3,325,162	41,987
- 4인가구	4,428,879	4,289,035	139,844

자료: 통계청(2008a), 등록센서스 실시 기본계획, 인구조사과 내부자료.

한편 지역단위별로 주민등록인구와 총조사인구 간에 차이가 발생하는 것은

주택청약, 부동산 취득, 가족수당 등 개인의 권리와 의무와 관련하여 거주지를 사실과 다르게 신고하는 것 때문인데 현재 이와 같은 허위신고가 얼마나 되는지에 대한 구체적인 자료는 없다. 그런데 이와 같은 허위신고는 주로 단독세대를 구성할 목적으로 이루어지는 경우가 많아 주민등록에서의 단독 세대와 인구총조사에서의 1인 가구수를 비교해 보면 대충 그 윤곽을 알 수 있다.

<표 8-7>에서 보면 2005년 11월 30일 현재 우리나라 전국의 주민등록세대는 약 1,781만세대로서 2005년 11월 1일 기준의 총조사가구인 1,589만 가구와는 약 193만의 차이를 보여 12.1%의 괴리를 나타낸다. 이는 주민등록에서는 별도의 세대를 구성하고 있는 것으로 집계되었으나 실제 총조사에서 확인한 가구수는 이보다 훨씬 못 미치고 있는 것을 보여주는 것이다. 이는 앞에서 본 바와 같이 인구의 경우는 같은 시점의 두 자료를 비교했을 때 4.1%의 차이가 났던 것과 비교하면 가구수에 있어서는 인구에 비해 세대 가까이 나 더 많은 차이가 나는 것을 알 수 있겠다.

이와 같은 차이가 발생하는 데에는 주민등록에서의 세대개념과 인구총조사에서 사용하는 가구개념과의 차이, 또 인구총조사에서의 1인가구의 누락 등도 어느 정도 영향을 미쳤을 것이지만 그 보다는 단독세대를 구성한 것처럼 허위로 신고한 것이 더 큰 영향을 미쳤을 것으로 보인다.

현행의 주민등록에서는 세대주와 친족관계에 있는 사람은 동일 주소지에서는 세대 분리를 하지 못하도록 규제하고 있다. 따라서 세금과 주택청약 등 다양한 이유로 별도의 세대구성이 필요한 사람들은 불가피하게 다른 주소지로 허위전입을 신고하고 별도의 세대를 구성할 수 밖에 없는데 이들 가공의 세대는 주로 단독 세대일 가능성이 높다. 실제로 표에서 주민등록세대수와 인구총조사 가구수와의 차이를 가구원 수별로 보면 단독 1인 가구에서 약 197만의 차이를 나타내 62.1%라고 하는 엄청나게 많은 차이를 나타내고 있다.

이와 같은 주민등록에서의 실제 거주와 신고된 거주지와의 불일치 정도는 통계청에서 2008년에 실시한 주민등록 확인조사에서도 어느 정도 그 윤곽이 드러난다. 즉 주소를 기준으로 실제 거주를 확인했을 때 전국적으로 일치하는 비율은 90.7%로 드러났다. 이를 읍·면·동 기준으로 집계하면 92.2%, 시·군·구 기준으로 집계하면 94.9%, 시·도는 97.3%로 일치율이 높아졌다. 이러한 일치율 수준은 1993년 확인조사 결과 집계된 거주지 일치율(읍·면·동 기준 82.2%, 시·군·구 기준 91.7%, 시·도 기준 95.3%)과 비교하면 크게 개선된 것으로 나타나 주민등록의 정확성이 점차로 높아지고 있는 것을 알 수 있다

(김형석 외, 2008).

그러나 아직까지도 주민등록상의 신고지와 실제 거주지와 불일치 정도는 상당한 수준에 있어 인구주택총조사를 주민등록부로 대체하는 데에 큰 걸림돌이 되고 있다. 현재 행정안전부에서는 이와 같은 불일치를 바로 잡기위해서 1년에 한번 씩은 일제 점검을 통해 이를 바로 잡으려고 하고 있으나 꼼꼼히 집 내부에서 이를 확인하는 것이 아니기 때문에 그리 효과적이지는 못하다. 또 경우에 따라서는 별도 단독 세대를 구성하면서 전입하는 사람에게 주택임대차 계약서 등의 근거 서류를 요구하기도 하지만 이 또한 허위 작성이 가능하기 때문에 별 도움이 되지 않는다고 한다. 또 허위신고가 적발된 사람에게는 처벌을 하도록 규정하고 있으나 일선 담당자들의 말을 들어보면 처벌 건수도 거의 없는 것으로 보인다. 따라서 이 허위신고문제는 향후 주민등록을 활용해서 인구총조사를 대체하려면 어떠한 방식으로든지 해결하지 않으면 안 되는 중요한 과제가 되고 있다.

2. 행정자료의 정확성 및 신뢰성

인구주택총조사를 기존의 행정자료로 대체하는 데에 발생할 수 있는 두 번째 걸림돌은 자료의 정비가 제때에 제대로 이루어 지지 않아 정확성과 신뢰성이 문제가 되는 경우이다. 현재 총조사의 주택 항목을 대체하는 데에 건축물대장이 유용하게 활용될 수 있는데 여러 연구에 의하면 현재의 건축물대장은 자료의 신뢰도에 있어서 문제가 되는 것으로 보고하고 있다.

통계청의 조사결과에 의하면 건축물대장과 인구주택 총조사자료를 1대 1로 매칭해서 일치율을 보면 일반건축물의 경우 주택종류에 있어서 일치율은 58.1%로 매우 낮게 나타난다 (통계청, 2007b). 이 불일치의 내용을 살펴보면 건축물대장의 일반단독이 총조사의 다가구 단독으로 분류된 것이 33.3%, 건축물대장에서는 영업겸용 단독으로 되어 있지만 총조사에서는 일반단독으로 분류된 것이 26.7% 등 여러 곳에서 불일치가 발견되었다. 이는 양 자료간의 주택 분류방식에 있어서의 차이와 실제사용과 대장상의 기재와의 차이에서 연유된 것으로 보인다.

또 연건평에 있어서도 두 자료간의 일치율은 60.8%로 매우 낮다. 연건평의 일치율은 주택의 규모에 따라서도 다르게 나타나는데 전체적으로 보면 연건평은 건축물대장보다 인구주택총조사에서 더 크게 조사되었다. 건축년도에 있어서는 일치율이 71.6% 이다. 전반적으로 최근에 지어진 주택일수록 일치율이 높고 60년대 이전에 건축된 주택은 일치율이 낮은 경향을 보인다. 또

단독주택의 대지면적은 건축물대장에서 약 40% 가량이 누락되어 있어 일치율을 계산하는 데에도 한계가 있었으나 매칭이 되는 자료만 놓고 본다면 일치율은 83.3%였다. 집합건축물은 일반건축물과 비교해 전반적으로 일치율이 높았는데 주택의 종류에서 일치율은 99.2% 이고, 연건평의 일치율은 93.8%였다. 또 건축년도도 98.2%의 일치율을 보이고 있었다.

현재 건축물대장이 갖고 있는 문제점으로는 건축물 대장 기재항목 가운데 주요 항목이 기재되어 있지 않은 누락이 항목별로 다르기는 하지만 상당한 정도에 이르고 있으며, 건축물대장의 총괄표제부와 동별 개요자료, 동별 개요자료와 층별 개요자료의 연계가 미흡하며, 자료별로 연계되어 있다 하더라도 연계된 정보 간 내용이 상이한 경우가 많다 (통계청, 2007d). 또 현행 건축물대장은 그 관리상에도 여러 문제를 안고 있는데 예를 들면 건축물대장의 작성에 있어서 연도별로 서식 유형이 다양하여 항목별로 기재율이 다르거나 또 건축물대장 카드 자료와 전산자료가 이중으로 관리되고 있어 업무담당자에게 부담이 되고 있다. 즉 건축물대장의 기재내용에 변동이 발생할 경우 전산대장에는 반영이 되었으나 카드대장에는 변동내용이 누락되는 경우가 있었고, 반대로 카드대장에는 수정이 되었으나 전산대장에는 반영이 안 되고 누락되는 경우 등 각 자료가 일치하지 않는 문제점이 발견되었다.

이와 같은 여러 문제점 때문에 현재 건축물대장은 단계적으로 일제 정비중에 있는데 현재의 계획으로는 2010년까지는 정비가 완료될 예정이다. 정비내용 중 주요한 정비대상으로는 구조, 용도, 대지면적 등 자료의 누락과 오류 시정, 건축물대장과 등기자료의 소유자 상이자료 정비, 집합표제부(동별 자료)는 있으나 전유부(호별자료)가 누락된 건축물대장 정비, 건축물대장은 존재하나 지번이 없는 건축물의 자료정비, 무허가건물의 정비, 건축물대장, 토지대장, 과세대장, 등기의 고유번호 일치 등이 포함되어 있다. 따라서 현재에는 인구주택총조사 중 일부 주택관련 자료를 건축물대장으로 대체하는 데에는 많은 제약이 있지만 예정대로 건축물대장이 정비되고 나면 현재보다는 그 활용가능성이 한층 높아 질 것으로 보인다.

제4절 등록센서스 추진의 제약사항과 제도적, 법률적 기반개선의 과제

기존의 행정자료를 이용하여 인구주택총조사를 대체하는 등록센서스가 본

격적으로 실시되기 위해서는 앞 절에서 지적한 바 있는 행정자료와 인구주택총조사 자료간의 개념상의 차이문제, 신고에 기초하고 있는 행정자료와 실제 조사에 기반한 인구주택 총조사자료 간에 발생하는 불일치의 문제, 그리고 행정자료의 정확성과 신뢰성에 있어서의 문제 등이 우선 해소되어야만 한다. 또 그 밖에도 행정자료 공유를 위한 행정기관 간의 협조와 조정, 점차 고조되고 있는 개인정보 보호욕구의 충족, 그리고 등록센서스에 대한 사회적 공감대 형성 등 제도적인 기반도 마련되어야 한다. 이제 이들 선결 과제들을 시간적인 순서를 고려하여 검토하려 한다.

1. 등록센서스 주축 행정자료의 정비

기존의 행정자료를 이용하여 인구주택총조사를 대체하게 되는 등록센서스를 실시하려면 우선 무엇보다도 수집된 행정자료가 어느 정도 신뢰할 수 있는지에 대한 확신이 있어야 한다. 센서스에 활용되는 행정자료의 품질을 진단하는 방법은 행정자료로 센서스통계를 작성하고 이를 최근의 센서스와 비교하는 것이다. 예컨대 네덜란드의 경우 중앙인구대장의 품질을 확인한 결과를 보면 1971년 센서스에서 인구대장의 0.02%가 과소 등록이 되었고, 0.05%는 과다 등록되었으며, 0.18%는 잘못된 지방자치단체에 등록되어 있는 것을 확인하였다. 따라서 네덜란드의 중앙인구대장은 상당히 신뢰도가 높은 행정자료임을 알 수 있다.

이제 이를 위해서 우리나라에서도 이미 수차례에 걸쳐서 주민등록자료와 건축물대장을 대상으로 센서스자료나 실제조사를 통해서 이들 행정자료의 품질을 검증한 바 있다. 그러나 그 결과는 앞 절에서 살펴본 바대로 아직까지는 만족스럽지 못해서 앞으로도 이와 같은 행정자료의 검증작업이 더 지속되어야 하겠다. 이제까지의 작업을 통해서 드러난 행정자료의 문제점과 그 개선책을 제시하면 다음과 같다.

가. 행정자료와 조사자료 간 매칭의 한계

그 동안 수차례에 걸친 실제 조사자료와 주민등록자료나 건축물대장의 매칭결과는 그리 만족스럽지 못한 것으로 드러났다. 이는 우선 일정 행정구역 내의 가구나 주택의 매칭에 있어서 매칭의 연결고리인 매칭키의 부정확성에 상당 부분 기인한다. 실제로 2007년에 실시한 통계청의 1차 시험조사 분석결

과를 보면 주소를 매칭키로 사용했을 때 매칭률은 72.2%로 드러났다(통계청, 2008a)

이때 매칭에 실패한 사유를 보면 1) 해당주소 (지번)가 없거나 일부 주소가 다른 경우, 2) 공동주택명, 동명, 호명이 없거나 다른 경우, 3) 주소하나에 지번이 두 개 이상인 경우로 일반주택의 10%는 하나의 지번에 건물이 두 개 이상 존재해 주민등록의 세대를 어떤 건물에 연결해야 할지 결정할 수 없었던 경우 등을 들 수 있다. 그런데 이와 같은 주소를 매칭키로 사용했을 때의 만족스럽지 못한 매칭 결과는 현재 행정안전부에서 추진하고 있는 새 주소 사업이 예정대로 진척되면 앞으로는 많이 개선될 것으로 기대된다.

현재 행정안전부에서 추진 중에 있는 새 주소란 도로명 사업에 의하여 새로 부여된 건물번호에 의하여 표기하는 주소를 말하는 것으로서, 이 새 주소는 건물 중심의 주소체계로 현행 지번중심의 주소체계와 달리 도로명과 건물번호로 이루어져 있다. 따라서 새 주소는 개별 건물 인식이 가능하고 각종 행정자료에서도 활용될 것이기 때문에 등록센서스 실시를 위한 주요한 행정자료 간 연계키로서 유용하게 쓰이게 될 전망이다.

현재의 계획대로라면 2009년 말까지는 모든 도로에 명판이 부착되고 2010년에는 고지단계를 거쳐 법적 주소로서의 근거를 마련하게 될 것이다. 2010년 하반기와 2011년 상반기에 걸쳐 주요한 공적장부인 주민등록부, 건축물대장, 토지, 건물, 법인 등기부, 가족관계등록부, 외국인 등록 등 주요한 공부에 새 주소를 부여하게 되며 우정사업본부에서는 우편번호와 새 주소 중심의 체계를 구축해 2012년부터는 전면적으로 사용할 예정이다.

나. 주민등록부의 정비

현재 주민등록자료의 가장 큰 문제점은 앞서서도 살펴 본 바대로 실제 거주와 신고 된 주소지와의 괴리문제이다. 이는 크게 두가지 부류의 사람들에게 의해서 발생하고 있는데 하나는 외국에 장기 거주하는 해외거주자가 주민등록상으로는 국내에 거주하는 것으로 되어 있는 것이고, 또 다른 하나는 여러 가지 목적으로 실 거주지와 주민등록상 거주지가 다른 허위거주자들의 문제이다.

이중에서 해외에 장기 체류 중인 사람들은 법무부의 출입국기록을 활용하여 주민등록이 되어 있는 거주지에서 별도로 구분해서 집계하면 문제가 해결될 것으로 보여 진다. 이 보다 더 문제가 되는 것은 단독세대 구성을 위하여 허위로 신고하는 사람들이다. 현행의 주민등록법에 의하면 주민등록 또는

주민등록증에 관하여 거짓의 사실을 신고 또는 신청한 자는 3년 이하의 징역 또는 1천만 원 이하의 벌금에 처한다고 제 37조에서 규정하고 있다.

그러나 실제로는 허위신고자에 대한 처벌은 거의 이루어지고 있지 않은데 이는 실 거주여부를 일일이 확인하기가 매우 어렵기 때문이다. 그러나 이 문제는 등록센서스를 실시하기 이전에 어떠한 방법으로든지 어느 정도는 해결하지 않으면 안 되는 과제이다. 이를 위해서는 예컨대 행정규제 철폐차원에서 과거에는 있었으나 현재에는 폐지된 전입 시 해당 거주지 통장의 확인을 받도록 하는 제도의 부활도 검토해 볼 만하다. 또 일부 지역에서 시도한 바 있는 특히 단독세대 전입의 경우 임대차 계약서를 제출하도록 해 실제 이사 여부를 확인하는 방법 등도 있을 수 있겠다.

다. 건축물대장의 정비

앞 절에서 살펴본 바대로 현재 우리나라의 건축물대장은 인구주택총조사 주택 항목을 대체하는 데에는 많은 문제점을 지니고 있다. 즉 건축물은 사용 승인 허가와 동시에 건축물 대장에 등재되고 건축물의 철거, 멸실 등의 사유로 인한 말소 신청으로 대장에서 삭제하게 되는데 현실에서는 가 사용 승인 등 건축물대장에 등재되지 않고 사용되는 건물이나 폐가, 철거 및 멸실되었는데도 건축물 대장에서는 삭제되지 않은 경우가 많다. 또 이밖에도 불법 및 무허가 건물 등이 있어 실제 건물 수와 대장상의 건물 수는 구조적으로 차이가 날 수 밖에 없다 (통계청, 2007b).

그러나 다행스럽게 현재 이 건축물대장을 관리하고 있는 국토해양부가 이 행정자료의 정비작업을 진행하고 있어 건축물대장의 정비가 완료되는 2010년에는 그 정확성이 많이 향상될 전망이다. 그러나 이와 아울러서 앞으로는 건축물대장의 신고양식도 개정하여 현재 인구주택총조사에서는 조사되고 있으나 건축물대장의 신고양식에서는 누락되어 있는 주요 주택 항목 - 예컨대 방수, 화장실 수, 부엌 및 식당 수, 독립된 출입구 여부 등 - 을 포함하여 신고하도록 하는 방안도 적극적으로 검토할 필요가 있다.

2. 행정기관 간 협력체제 구축

등록센서스가 성공하기 위해서는 통계당국과 관련 행정자료를 생성하는 행정기관간의 유기적인 협조체제의 구축이 절대적으로 필요하다. 대부분의 북유럽 국가에 있어서는 행정자료를 이용하여 통계시스템을 만드는 데에 통계

법에 근거한 국가의 지원이 있었으며, 통계작성기관이 관련 행정기관과 우호적인 그리고 밀접한 협력관계를 맺는 것은 행정자료의 효율적인 사용에 매우 중요한 일이었다.

핀란드의 경우를 들어 통계당국과 각 행정기관이 어떻게 서로 협력했는지를 살펴보면 우선 통계당국은 각 관련 행정기관과의 유기적인 협조를 위해서 담당자제도를 도입하였다. 각 행정기관 담당자의 임무는 각 관련 행정기관과 의사소통의 채널을 유지하고 관련 분야 안에서의 변화를 추적하며, 행정자료의 통계적 적용 가능성을 유지 또는 향상시키기 위한 작업을 하는 것이다. 이와 마찬가지로 각 행정기관도 한 명의 통계담당자를 지정하였다.

따라서 우리나라의 경우에도 우선 통계청과 행정안전부, 국토해양부, 그리고 법원 등의 행정기관간의 유기적인 협력체제가 구축되어야 하는데 이는 등록센서스 실시를 위한 준비단계에서부터 필요하다. 그래서 2010년에 실시될 인구주택 총조사결과를 활용하여 기존 행정자료의 품질을 점검하는 데에 유관 행정기관도 참여하여 문제점이 무엇인지를 확인하고 이를 개선할 수 있는 방안을 공동으로 마련해 나가야 할 것이다. 또 점차 등록센서스가 기틀이 잡히고 본격적으로 실시되게 되면 관련 행정기관을 더 확대하여 통계청과의 공식통계 조정시스템이 구축되어야 하겠다.

3. 개인식별시스템의 구축

행정자료를 통계목적으로 활용하기 위한 또 다른 조건 중 하나는 개별 단위가 항상 독립적으로 인식될 수 있어야 하며 서로 다른 자료원과 연계될 수 있는 통합된 식별시스템이 있어야 한다는 것이다. 북유럽 국가들은 서로 다른 행정자료를 연결하는 매칭키를 가지고 있는데 예를 들면 개인 식별 번호, 수치주소, 사업체번호 등이다. 한편 대만에서는 우리나라의 주민등록번호와 유사한 신분증통일번호(身分證統一編戶)를 사용하여 관련 공부파일을 매칭하고 있다.

우리나라에서도 주민등록부나 건축물대장, 가족관계등록부 등 서로 다른 행정자료를 연결하는 연계키가 필요한데 주민등록번호가 매우 유용하게 쓰일 수 있을 것이다. 다만 현재는 일반 국민들의 저항을 의식하여 인구주택총조사에서 주민등록번호를 조사하지 못하고 있는데 앞으로 일반 국민의 등록센서스에 대한 사회적 공감대가 형성되고 정보보호에 대한 확고한 인식이 심어진다면 주민등록번호를 연계키로 사용하는 방법을 적극적으로 모색하여야 하겠다.

4. 개인정보보호와 행정자료제공의 조화

정부 내의 각 행정기관이 고유의 업무를 위하여 수집해 놓은 행정자료를 통계목적에 위하여 이용하기 위해서는 두 가지 조건이 충족되어야 한다. 하나는 통계작성기관이 행정자료를 보유하고 있는 부서에 자료를 요청했을 때 해당부서에서 아무런 조건 없이 자료를 제공하는 법적인 근거가 마련되는 것이며, 또 다른 하나는 통계작성기관은 이러한 행정자료의 집중화로 발생할 지도 모르는 개개인의 정보누출을 방지 할 수 있는 제도적인 장치를 구비해야 한다는 것이다.

이와 같은 행정자료의 제공과 개인정보보호의 의무는 이미 등록센서스를 실시하고 있거나 현재 이를 준비하고 있는 국가들에서는 법적으로 혹은 제도적으로 확립되어 있다. 예컨대 모든 북유럽국가들의 통계법은 통계작성기관에 통계목적으로 개별단위에 대한 식별이 가능한 수준의 데이터에 대한 접근과 다른 행정자료를 연계시킬 수 있는 권한을 부여하고 있다. 여기에 한 걸음 더 나아가서 핀란드나 네덜란드 등에서는 국가 통계작성기관은 독자적으로 자료를 수집하기 전에 우선 관련 행정자료가 다른 행정기관에 존재하는 지를 먼저 검토해서 존재한다면 이를 사용할 것을 의무화하고 있다.

이와 같은 행정자료이용의 용이성의 제도적기반과 동시에 모든 북유럽국가들은 또 개인정보의 보호를 위한 법적 장치도 갖고 있다. 즉 이 법의 목적은 개인정보를 포함하고 있는 전산자료의 이용이 시민 개개인의 권리를 침해하지 않는 방식으로 이루어져야 한다는 것을 보장하기 위한 것이다. 핀란드의 개인정보보호에 관한 법률에는 행정목적에 위해서 수집된 자료는 학문적 연구와 공식 통계작성을 위한 목적으로 제 3자에게 전달 될 수 있다고 규정한다. 그러나 일단 통계작성기관에 의해 자료가 처리되면 이 자료는 통계생산과 연구 목적이외에는 사용될 수 없다.

북유럽 국가들에서 행정활동에 대한 시민들의 알 권리보장과 행정의 투명성은 매우 중요한 기본원칙이다. 그러나 정보가 통계작성을 위해 통계당국에 전달되자마자 공공의 접근가능성의 원칙은 더 이상 적용되지 않으며 비밀보호는 통계당국의 절대적인 지상과제가 된다. 정보의 흐름은 단지 한 방향으로, 즉 행정부서로부터 통계당국으로만 흐르고 반대의 흐름은 철저히 차단되는 것이다.

따라서 이와 같은 북유럽국가들의 예를 참조한다면 우리나라에서 등록센서스를 실시하기 위해서는 통계법에 위와 같은 사항을 분명하게 규정할 필요가 있다. 현재 우리나라의 통계법 제 24조는 1항에서 '중앙행정기관의 장 또

는 지방자치단체의 장은 통계의 작성을 위하여 필요한 경우에는 중앙행정기관·지방자치단체의 장에게 행정자료의 제공을 요청할 수 있다' 라고 하고 있으며, 2항에서는 '공공기관의 장은 제1항에 따라 행정자료의 제공을 요청받은 때에는 국가기밀, 개인과 기업의 중대한 비밀의 침해 등 대통령령으로 정하는 정당한 사유가 없는 한 이에 응하여야 한다.'고 규정하고 있어 행정자료의 제공요청에 이를 유보할 수 있는 여지를 남겨놓고 있다.

통계법 시행령 제39조(행정자료 제공의 예외사유)는 법 제24조제2항에서 대통령령으로 정하는 정당한 사유를 4가지로 열거하고 있는데 그중에서 제4항이 개인의 정보보호와 직접적으로 관련된 사항이다. 즉 제4항은 '개인의 정치적, 종교적 또는 성적 성향이나 생활에 관한 행정자료로서 통계의 작성을 위하여 제공되면 개인의 생명이나 신체, 재산의 보호에 현저한 지장을 줄 우려가 있다고 인정할 만한 상당한 이유가 있는 경우' 에는 자료의 제공을 거부할 수 있게 하고 있다. 따라서 향후 등록센서스의 원활한 추진과 정착을 위해서는 이 조항의 유보조건을 삭제하고 모든 경우에 이에 응하여야 하는 규정으로 개정할 필요가 있다.

한편 개인정보보호를 위해서 현행의 통계법은 제 24조 3항에서 다음과 같이 규정하고 있다 즉 '제2항에 따라 행정자료를 제공하는 경우 그 제공범위 및 방법 등에 관하여는 요청기관의 장과 제공기관의 장이 협의하여 결정하되, 제공기관의 장은 요청기관의 장에게 행정자료에 포함되어 있는 개인이나 법인 또는 단체 등의 정보를 보호하기 위하여 사용방법·사용부서나 그 밖에 필요한 사항에 대하여 제한을 하거나 행정자료의 안전성 확보를 위하여 필요한 조치(이하 "정보보호조치"라 한다)를 강구하도록 요청할 수 있다.' 또 제 4항에서는 '제2항에 따라 공공기관으로부터 제공받은 행정자료는 이를 통계작성 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다' 라고 하고 있어 개인의 정보보호를 보장하고 있다.

5. 등록센서스에 대한 사회적 공감대 형성

여러 다른 행정기관들이 수집하여 관리하고 있는 행정자료를 한군데로 모아서 통계목적으로 이용하기 위해서는 국가 통계기관이 모든 개인에 대해서 많은 것을 알게 되는 '빅 브라더 신드롬' (Big Brother Syndrome)에 대한 일반 시민들의 의구심과 불안이 불식되고 등록센서스의 필요성에 대한 사회적 공감대가 형성되어야 한다. 이미 오래전부터 행정자료를 통계적 목적에 이용하고 있는 북유럽 국가들에서도 이 문제에 대해서는 많은 논의가 있었

는데 등록센서스의 합리성과 장점을 부각시키는데 있어서 공개된 토론과 논쟁이 많은 도움이 되었다. 또 행정자료를 취급하는 기관의 개방성과 투명성도 중요한 변수로 작용한 것으로 드러나고 있다.

따라서 우리나라에서 등록센서스를 본격적으로 실시하기 위해서는 행정자료 이용의 필요성과 장점, 그리고 개인의 정보보호에 대한 확신을 시민들에게 심어주는 많은 사전 노력이 필수적으로 이루어져야 한다. 이때 행정자료를 통계목적으로 활용하면 얻을 수 있는 다음과 같은 이점을 집중적으로 홍보할 수 있겠다. 우선 통계자료 수집의 경제적 효율성과 비용 절감효과이다. 둘째는 개인, 가구, 기업의 응답 부담이 경감된다는 점이다. 셋째는 데이터를 다루는 사람의 수가 적어지고 전자적인 형태로 자료처리가 이루어지기 때문에 데이터에 대한 보안유지가 오히려 더 쉬어진다는 점 등이다. 물론 이렇게 등록센서스의 장점을 설명해도 사회적인 공감대가 형성되기까지는 많은 시간이 필요할 것이기 때문에 매우 조심스럽게 그리고 꾸준히 국민을 설득하는 작업이 이루어 져야 하겠다.

제5절 연구결과의 요약 및 전반적 평가

이 장의 목적은 기존의 행정자료를 이용하여 현행의 인구주택총조사를 대체하는 등록센서스를 실시할 경우에 발생할 것으로 예상되는 문제점을 명확히 하고 그 개선방법을 모색하는 것이었다. 이를 위해서 먼저 현재 인구주택총조사에 포함되어 있는 항목 중 기존의 행정자료로 대체가 가능한 항목을 확인하였는데 대체로 14개 정도의 항목이 확인되었다.

이들 항목을 행정자료별로 보면 주민등록 자료로 부터는 7개 항목이 가능한데 성명, 성별, 나이, 가구주와의 관계, 1년 전 거주지, 5년 전 거주지, 가구구분 등이며, 가족관계등록부를 이용하면 혼인상태, 본관, 출생지 정보를 얻을 수 있다. 한편 주택에 관한 정보로는 건축물대장에서 거처의 종류, 주거용 연면적, 건축년도, 주택소유여부 등을 알 수 있다.

이렇게 인구주택총조사의 항목 중 행정자료로 대체가 가능한 항목을 확인한 후에 먼저 이들 행정조사자료의 항목과 인구주택 총 조사 항목의 개념을 비교하여 개념상의 차이점이 없는지를 검토하였다. 그 결과 인구 부문에서는 연령과 결혼관계에서 두 자료 간에 개념상의 차이가 있는 것을 발견하였다. 즉 연령에 있어서는 인구주택총조사의 연령은 실제 태어난 생년월일을 기준으로 산정한 연령인데 비해서 주민등록부의 연령은 신고한 연령을 기재 하고 있다.

그런데 우리나라의 경우 과거에는 출생신고가 늦어지는 경우가 많아서 실제 나이가 주민등록부에 기재된 나이보다 많은 사람들이 상당수에 달하고 있었다. 즉 통계청에서 2008년에 실시한 주민등록 확인조사 결과를 보면, 실제 연령과 주민등록 연령의 일치율은 80.4%로 나타났는데, 남자가 82.1%로서 여자의 78.8% 보다 3.3% 포인트 더 높게 나타났다. 이 일치율은 출생연도에 따라 다르게 나타나는데 고 연령층에서는 일치율이 대체로 60%대에 머물며 1990년 이후 출생자의 경우는 97.9%로서 실제 연령과 주민등록 연령이 거의 일치하고 있었다.

혼인상태의 경우는 인구주택총조사에서는 혼인신고와 관계없이 사실상의 혼인상태를 기준으로 조사하고 있는데 비해서 가족관계등록부에서는 신고된 사항을 기준으로 혼인상태를 표시하고 있다. 따라서 가족관계등록부에서는 실질적으로는 혼인이나 이혼상태에 있지만 신고가 되어 있지 않으면 미혼이나 배우자 없음으로 분류되어 총조사 결과와는 약간의 괴리가 있게 된다.

한편 개념상의 차이는 아니지만 자료 항목의 분류상의 차이 때문에 행정자료와 인구주택총조사 자료간의 비교를 위해서는 약간의 조정이 필요한 항목도 있다. 즉 인구 부문에서는 가구주와의 관계인데 주민등록부의 세대주와의 관계를 총조사의 분류 코드에 맞추어 재조정해야 하며, 주택부문에서는 거처의 분류 코드가 총조사와 건축물대장간에 상이하야 약간의 조정이 필요하다.

등록센서스를 추진하는데 있어서 현행의 행정자료가 지니고 있는 한계점으로는 크게 두 가지 문제가 지적될 수 있는데 하나는 자료 생성방법의 차이에서 연유되는 자료 간 불일치의 문제이고, 다른 하나는 행정자료의 정확성과 신뢰성의 문제이다. 우선 주민등록자료는 신고에 의존하는 자료이고 총조사자료는 실제 조사된 자료이기 때문에 인구수에 있어서 많은 차이를 노정하고 있다. 그 동안에 이루어진 양 자료 간 인구수를 비교분석한 결과를 보면 전국 총인구에 있어서 주민등록인구가 총조사인구보다 대체로 약 4% 정도 많은 것으로 드러난다.

또 행정단위별로 주민등록인구 대비 총조사인구의 비율을 보면 시·도별로는 평균 96.5%, 시·군·구별로는 평균 93.8%, 그리고 가장 작은 행정단위인 읍·면·동 수준에서 두 자료를 비교하면 이 비율이 93.7%로 나타난다. 이처럼 주민등록인구가 총조사인구보다 많이 집계되는 이유는 전국 총인구의 경우는 해외에 장기간 거주하는 사람들 때문이다. 현재 주민등록제도에 의하면 해외이주자의 경우 해외에서 영주권을 취득하여 출입국사무소에서 명단이 통보되기 전까지는 주민등록에 계속 기재될 수밖에 없다. 2006년 1월 1일 현재 이처럼 주민등록은 되어 있으나 해외에 장기 거주하는 한국인은 약 118

만 명으로 추정되었다.

한편 행정구역별로 인구수에서 차이가 나는 것은 해외장기거주자 이외에도 주택청약, 부동산취득, 가족수당 등 개인적인 이득을 취할 목적으로 거주지를 사실과 다르게 신고하는 허위신고 때문이다. 이와 같은 허위신고는 특히 단독세대를 구성할 목적으로 이루어지는 경우가 많아 1인가구의 경우 주민등록세대와 총조사 가구 간에는 무려 62.1%의 차이를 나타내고 있었다. 그런데 현재로서는 이 허위신고를 바로잡을 마땅한 행정수단이 없기 때문에 향후 등록센서스를 실시하는데 있어서 이 문제가 가장 큰 걸림돌로 작용할 전망이다.

행정자료의 두 번째 한계인 자료의 정확성과 신뢰성의 문제는 특히 건축물대장에서 두드러지게 나타난다. 현재 건축물대장에는 항목별로 많은 누락이 발견되고 있으며 대장에 기재된 내용과 실제와의 괴리도 상당수 있다. 또 관리상에서도 대장의 카드자료와 전산자료가 이중으로 관리되어 양 자료 간에도 불일치가 존재하는 등의 문제를 안고 있다. 그런데 다행스럽게도 이 건축물대장은 현재 이와 같은 문제들을 정비 중에 있어 정비가 완료되는 2010년부터는 그 신뢰성이 상당히 높아질 전망이다.

이 장의 마지막 부분에서는 앞에서 지적한 현행 행정자료의 한계를 극복하기 위한 그리고 또 등록센서스가 실시될 때에 필요한 여러 가지 제도적인 방안을 제안하였다. 먼저 등록센서스의 주축이 되는 행정자료의 정비를 위해서는 앞으로도 지속적으로 총조사자료를 활용하여 행정자료의 품질을 향상시켜야 하는데 이때 2012년부터 전면적으로 시행될 새주소를 행정자료와 총조사자료 간의 매칭키로 활용하면 현재 안고 있는 불완전한 매칭의 문제는 상당 부분 해소될 것으로 예상하였다 또 현재 가장 큰 문제가 되고 있는 주민등록의 허위신고는 가능한 모든 행정력을 동원하고 또 국민을 설득하여 어떠한 형태로든지 이를 해결하지 않으면 안 되며 마찬가지로 건축물대장의 신뢰성도 꾸준히 높여 가야 할 것이다.

등록센서스를 실시하는데 있어서 통계당국자와 관련 행정기관간의 협조가 절대적으로 필요한데 이를 위해서는 등록센서스를 준비하고 있는 현 시점에서부터 우선 행정안전부, 국토해양부, 법원 등의 일부 행정기관들을 참여시키는 협력체가 구성되어야 하겠다. 또 여러 다른 행정자료원을 연계시키는데 있어서 현재로서는 가장 정확한 매칭키가 될 수 있는 주민등록번호를 사용하는 문제를 조심스럽게 검토할 필요가 있다.

향후 등록센서스의 전면 실시에 있어서 개인정보보호의무와 행정자료제공의 의무는 반드시 제도적으로 그 기반이 마련되어야 한다. 이를 위해서는

현재의 통계법 제24조 2항의 행정자료제공을 유보할 수 있는 조항을 개정하는 것이 필요하다. 그리고 마지막으로 등록센서스가 성공적으로 시행되면 대 국민 홍보를 강화하여 일반 국민들의 사생활 침해에 대한 의구심과 불안을 불식시켜 등록센서스에 대한 사회적인 공감대가 형성되도록 하여야 할 것이다.

전반적으로 볼 때 등록센서스가 우리나라에서 정착되어 성공적으로 실시되기 위해서는 극복해야 할 과제가 한두 가지가 아님을 확인하였다. 무엇보다도 행정자료의 질적인 개선 문제가 등록센서스 실시 이전에 해결되어야 하는데 이를 위해서는 많은 시간이 필요할 것으로 보인다. 무리하게 실시 시한을 정해 놓고 서두르기 보다는 2010년 인구주택총조사의 결과를 활용하여 행정자료의 품질을 꾸준히 점검하되 2010년에 완료되는 건축물대장의 정비와 2012년부터 사용될 새 주소의 효과를 보아가면서 문제를 점진적으로 개선해 나가는 자세가 요청된다.

제9장 2015년 등록센서스 추진을 위한 단계적 계획 및 구체적 과제의 검토

제1절 머리말

우리나라의 인구주택총조사는 사회적 환경의 변화로 조사비용이 빠르게 증가하고 있고, 응답자의 의식변화로 전수조사 자체가 어려워지고 있다. 이러한 환경에서 정부에서 추진하고 있는 행정자료를 이용한 등록센서스의 도입을 위한 연구가 이루어지고 있다. 이 장에서는 우리나라 인구주택총조사를 대신하는 등록센서스를 실시하는 단계적 추진계획을 마련하고, 단계별로 해결해야 할 과제를 검토하는 것이다. 그리고 앞으로 등록센서스를 성공적으로 실시하기 위한 기본계획을 수립하고 보완하는데 도움을 줄 것이다. 이것을 위하여 등록센서스 로드맵 작성절차를 개관하고, 로드맵을 3단계로 구분하여 논의하였다. 첫 번째 로드맵은 행정자료를 확보하고 보정하는 방법이고, 두 번째는 관계기관과의 협력체계를 구축하는 것이다. 끝으로 개인정보보호와 행정자료 통합 데이터베이스를 구축하는 것이다. 마지막으로 통계청의 기본 계획에 대한 전반적인 평가를 한 다음 종합적인 정책건의를 하였다.

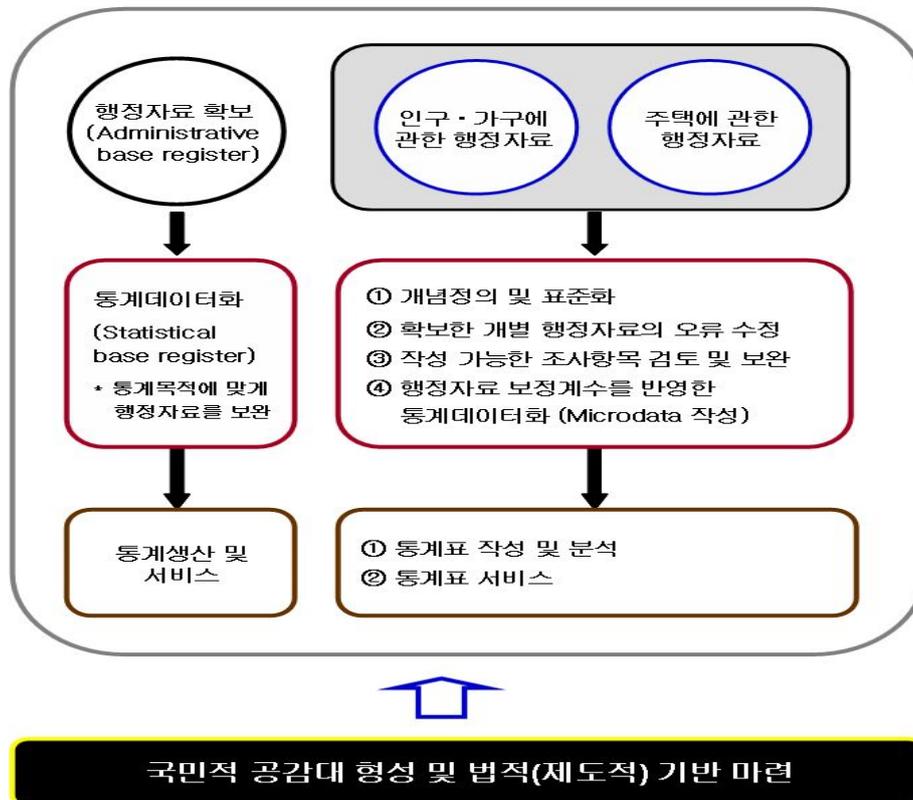
제2절 연구범위와 등록센서스 로드맵 작성의 절차

1. 등록센서스의 절차

우리나라에서 등록센서스를 실시하기 위해서는 다음과 같은 일반적인 과정에 따라 추진할 수 있다. 등록센서스를 실시하고 있는 몇 국가들의 추진 과정이 일부 차이는 있으나 먼저 행정자료를 확보하는 일이다. 이 행정자료는 등록자료(register data)를 중심으로 이루어져 있는데 우리나라의 경우 인구 부문에서는 주민등록부의 자료와 가족관계등록부의 자료라고 할 수 있고, 주택부문의 경우는 건축물대장 자료가 된다. 각 행정자료를 담당하고 있는 부서와의 협조가 이루어질 때 전 등록자료를 이용하게 되는데 이 등록자료

를 통계적 처리가 가능한 통계데이터화를 거쳐 통계목적에 맞게 변환하는 이른바 “통계등록부”(statistical register)를 작성하여야 한다. 이 때 필요한 것은 인구와 주택관련 개념의 정의를 통일하고, 용어를 표준화하여 통계처리가 가능하게 하여야 한다. 다음으로 개별 행정자료의 오류를 수정하여 통계의 정확성을 확보한다. 전통적 인구센서스의 조사항목별 통계를 최대한 확보할 수 있도록 이용 가능한 항목을 최대한 확보하는 노력이 필요하다. 끝으로 행정자료와 보완을 위한 실지조사(표본조사)자료를 매칭시켜 오차를 추정하고, 전수추계를 위한 보정계수를 추정하여 통계처리가 가능한 마이크로 데이터(micro data)를 작성한다. 이 데이터 셋(data set)을 이용하여 센서스 통계표를 만들고 기초자료를 공급하여 다양한 활용이 가능하도록 한다. 이상의 내용을 정리하면 다음 그림과 같다.

〈등록센서스 추진 과정〉



이상의 과정을 기본으로 하여, 우리나라의 2015년 등록센서스가 벤치마크

할 수 있는 이스라엘의 통합센서스의 절차를 중심으로 우리나라 등록센서스의 예상과정을 정리하였다(Israel Central Bureau of Statistics, 2005, 2007, 2008; .

A. 통계적 · 행정적 절차

가. 행정파일의 작성

등록센서스 제1단계 작업은 행정파일을 준비하는 것이다. 주민등록대장의 개인별 자료가 수록된 파일과 건축물대장의 행정파일이다. 건강보험자료와 국세자료의 활용방안을 검토하여 주민등록대장의 개인별 자료를 보완할 수 있는 방안을 강구한다.

나. 인구센서스 표본추출

등록센서스를 실시하기 위하여, 두개의 상호독립적인 표본, 곧 현장 조사구를 대상으로 한 표본과 해당지역의 행정파일(주민등록대장의 개인자료)에 등재된 모집단의 표본에 대한 조사를 실시한다.

현장 조사구 표본조사: 현장 표본조사의 표본을 선정하기 위하여, 통계청은 전국을 조사구로 구분한다. 조사구는 평균 50개의 가구로 구성된 연속적인 지리공간으로 이루어진다. 이 조사구 모집단에서, 현장 표본조사를 위한 조사구를 무작위로 추출한다. 표본추출률은 10~20%가 된다. 이 표본조사구에 대하여, 현장 조사원은 전통적 센서스 방법에 의거하여, 가구별로 모든 조사대상자를 빠짐없이 조사한다.

행정파일 모집단의 표본: 행정파일에 주소가 등재되어 있는 사람 중에서, 표본조사구에 거주하는 사람만을 대상으로 표본을 선정한다.

B. 인구센서스 표본조사 실시

가. 표본조사구의 가구집계(Under-coverage Check Sample)

이 단계에는, 현장 조사원이 모든 표본 조사구에 대하여, 모든 건물과 거처를 방문하여, 추출된 조사구별로 개인의 주택에서, 컴퓨터의 지원을 받아서 조사표를 이용하여 조사대상자로부터 필요한 자료를 얻는다.

이와 같이, 현장실지조사에 의거하여 표본 조사구의 조사대상자 자료를 토

대로 조사대상자가 집계된 주소와 행정파일 상의 주소 간에 매칭을 실시한다. 이 매칭작업의 결과에 의거하여, 행정파일 주소상의 모집단 과소집계 (Population Under-coverage)에 대한 정보를 얻는다.

나. 전화면접조사(Computer-Assisted Telephone Interviewing, CATI)

- Over-coverage Check Sample

행정파일 주소에는 등재되어 있으나 현장실지조사에는 발견되지 않은 모든 사람들은 전화로 면접을 실시하여, 그들이 실제로 거주하는 장소가 어디인가를 확인한다. 이 조사결과는 행정파일 주소에서의 모집단 과다집계 (Population Over-coverage)를 추정하는데 사용한다.

C. 센서스 통계생산

가. 등록센서스 파일(Integrated or Registered Census File, ICF)

통합센서스 파일은 행정파일에 등재된 개인의 인구학적 상세 특성으로 구성된다. 또 행정파일에 등재된 개인은 모집단에서 실제로 몇 명을 대변하는가를 명시적으로 보여주는 센서스 가중치를 포함한다. 행정파일에 등재된 개인의 센서스 가중치가 2.0이면, 그는 모집단에서 실제로 2명의 특성을 대표하게 된다. 곧, 해당 개인은 행정파일 주소지에 실제로 거주하는 사람 1명과 행정파일 주소에 누락되어 다른 주소지에 거주하는 사람 1명을 대표한다.

나. 사회경제 파일(Socioeconomic File, SEF)

사회경제파일은 현장실시조사에서 집계된 가구와 개인에 대한 자료로 이루어진다. 이 파일은 인구학적 속성과 사회경제적 속성을 모두 포함하는 파일이다. 표본조사구에서 집계된 모든 사람들은 모집단의 특성을 대표할 수 있도록 센서스 가중치를 부여받게 된다. 그 결과, 우리가 어떤 지역에서 조사된 모든 개인들의 모든 센서스 가중치를 합산하게 되면, 그 지역에서 등록 센서스 파일의 인구 총수와 사회경제 파일의 인구 총수는 동일하게 된다.

다. 센서스 통계의 이용자를 위한 서비스 제공

센서스 파일과 사회경제 파일에서 통계이용자들을 위한 다양한 센서스 통계가 작성된다. 그 중에는, 센서스 실시연도의 통계적-지리적 행정구역별 인구학적, 사회경제적 특성에 대한 통계표를 작성할 수 있는데, 여기에는 교육, 경제활동과 같은 통계표가 그 예에 해당하게 될 것이다. 또, 주소불명으로 인하여 통계처리에 어려움이 있는 사람들의 파일도 작성될 수 있을 것이다.

2. 등록센서스에 의한 인구추정의 과정

이스라엘 통합센서스에서, 인구대장의 과다집계 파라미터 람다(λ) 값은 인구대장의 순과다(net excess)에 해당하는 인구(완전히 그 지역을 떠나서 이사를 갔거나, 해외로 이민을 가서 마땅히 그 지역 인구대장에서 집계되어서는 안 되는 사람)와 현장실사 집계와 비교할 때 과다에 해당하는 인구(완전히 그 지역을 떠나서 이사를 가지는 않고 해당 행정구역 안에 거주하지만, 조사 시점에 집계되지 않았기 때문에, 인구대장에 잘못 기재되었다고 판단할 수 있는 사람)로 구분한다. 여기서 인구대장의 순과다 인구를 분자로 하고, 센서스 모집단 인구(그 참값은 아무도 모르기 때문에 추정수치를 사용함)를 분모로 하여, 그것을 계산한 결과를 과다집계 파라미터(over-coverage parameter)라고 정의한다.

가령, A지역의 주민등록 인구가 200명이고, 현장실사에서 180명으로 집계한 결과를 토대로 하여, 두 자료를 매칭하였을 때 150명이 매칭에 성공하였다면, 과소집계 파라미터(under-coverage parameter)는 $p^{1+} = 150/180 = 5/6 = 83.3\%$ 가 된다(과소집계 파라미터는 항상 현장실사를 기준으로 하고, 현장실사에서 집계된 인구가 주민등록에 모두 리스트된 사람이라면, 과소파라미터는 그 값이 100%이고, 현장실사에 집계된 인구가 주민등록에 리스트된 사람이 전혀 없다면 과소파라미터는 그 값이 0%가 된다).

다음은 과다집계 파라미터이다. 순과다를 측정하기 위하여 주민등록인구 중에서 매칭에 실패한 사람. 곧 주민등록에는 리스트되었으나 현장실사에서 집계되지 않은 사람들 50명 (주민등록인구 200명 - 150명 = 50명)의 실제 행방에 대한 재조사를 실시한다. 물론 이스라엘에서는 전화 재조사를 하는 모양인데, 우리나라에서는 어떤 방법으로 해야할 것인가에 대하여 차후에 시험조사, 시범조사, 내년 본조사에서도 충분히 논의되어야 할 사안이다. 만약, 이들 50명 중에서 30명이 A지역을 완전히 떠나버린 경우, 이들이 순과다가 된다. 따라서 순과다를 측정하는 과다집계 파라미터 람다값은 $30/(\text{센서스 모}$

집단의 인구)인데, 센서스 모집단의 인구를 그 참값을 알 수 없기 때문에, 그 값을 추정하는데, 그 공식은 주민등록에 리스트된 인구(200명)에서 순과다에 해당하는 인구(30명)를 빼서, 그 인구를 과소집계 파라미터로 나눈다. 공식을 적용하면 센서스 모집단 인구의 추정치는 $170/0.833$ 이고, 이것을 다시 공식에 적용하면 $30*(.833)/170= 14.7\%$ 가 된다.

이상의 결과를 종합하면, A지역의 과소파라미터는 0.833, 과다파라미터는 0.147이 되고, 전체집단에 대한 센서스 가중치는 $0.833+0.147= 0.980$ 의 역수인 1.020이 된다. 물론 A지역의 인구가 성별, 연령별로 동질적이라면 주민등록 인구(200명)에 1.020을 곱하여 204.0, 대략 204명이 된다.

그러나 A지역의 인구가 동질적이지 않기 때문에, 성·연령별로 나누어서 과소집계 파라미터와 과다집계 파라미터를 구하여 센서스 가중치(census weight)를 계산하고, 이것을 가지고 센서스 추정인구를 구할 필요가 있다. 연령집단을 크게 3등분을 할 수도 있고, 남녀별로 나누어서 6개 그룹으로 해서, 센서스 가중치를 구하여, 주민등록인구를 토대로 한 센서스 인구를 추정한다.

과소집계를 추정하기 위한 표본조사를 주민등록인구의 20% 정도로 실시하고, 동시에 그것을 가지고 현행 인구센서스의 전수항목에서 제외된 항목과 표본조사 항목을 집계하여, 이스라엘 식의 표본조사 데이터베이스를 구축하게 된다.

과소집계를 추정하기 위한 조사는 표본조사이기 때문에, 여기서 성별, 연령별 추정집단에 대하여 센서스 가중치를 계산하여, 그것을 전국수준 또는 필요한 경우에는 광역자치단체별로 적용하여, 전국수준의 센서스 인구 추정을 할 수 있을 것이며, 센서스 가중치를 바탕으로 하여, 가구(세대)별 통계나 사회경제적 특성에 대한 통계를 작성하면 될 것이다.

제3절 로드맵 1: 행정자료 확보와 보정방안 마련

통계청의 인구주택센서스 시험조사 자료와 등록센서스의 중심자료를 매칭하여, 센서스자료와 등록자료의 차를 보완하는 통계추정 및 보정방안의 제시하면 다음 그림과 같다.



1. 행정자료의 특성

가. 인구주택총조사 자료와 주민등록 및 가족관계등록 자료의 차이

1) 주민등록부

인구주택총조사에서 조사되는 항목 중 주민등록부를 이용하여 대체할 수 있는 자료로는 성명, 성별, 나이, 가구주와의 관계, 1년전 거주지, 5년전 거주지, 가구구분 등이 있다. 이 중에서 성명과 성별은 개념정의에서 차이가 없기 때문에 제외하고 여기에서는 나이와 가구주와의 관계에 대해서 각각의 정의와 개념을 서로 비교하고자 한다(<표 9-1> 참조).

가) 나이

먼저 인구주택총조사에서는 가구원 모두의 나이를 조사하는데 주민등록(또는 호적)상의 나이와 상관없이 집에서 세는 나이와 띠를 물어보고 동시에 생년월일도 조사하는데 양력인지 음력인지를 표시하도록 하여 매우 정확하게 현재의 나이를 파악하고 있다. 나이, 띠, 생년월일이 불일치하는 경우에는 부호집(codebook)에 수록되어 있는 '연령조건표'를 참조하여 다시 확인한 후 수정한다. 총조사에서는 인구학적 변수로서 연령이 중요하기 때문에 가능한 한 정확하게 조사하려고 한다.

주민등록부상의 나이는 본인들의 신고에 의해서 기재된 나이이기 때문에 총조사에 비하면 그 신뢰성이 많이 떨어진다. 과거에는 출생신고가 늦어지는

경우가 많아 실제 나이가 주민등록부에 기재된 나이보다 많은 사람들이 상 당수에 달하고 있다. 실제로 통계청에서 2008년에 실시한 주민등록 확인조사 결과를 보면, 실제 연령과 주민등록 연령의 일치율은 80.4%로 나타났는데, 남자가 82.1%로서 여자의 78.8% 보다 3.3% 포인트 더 높았다(김형석 외, 2008).

<표 9-1> 인구총조사 자료와 주민등록 자료의 비교

	2005 인구총조사	주민등록 자료
조사목적	<ul style="list-style-type: none"> - 인구총조사는 인구규모, 분포 및 구조에 관한 제특성을 파악, 각종 정책 입안의 기초자료 제공 - 각종 가구관련 경상조사 표집 틀 (sampling frame)의 기초자료로 활용 	<ul style="list-style-type: none"> - 시군구의 주민을 등록하게 함으로써 주민의 거주관계 등 인구의 동태를 상시로 파악하여 주민생활의 편익을 증진시키고 행정사무의 적정한 처리 도모
법적근거	<ul style="list-style-type: none"> - 재정경제부령 제145호 - 지정 조사통계 제10101호(인구총조사) 및 제10102호(주택총조사) 	<ul style="list-style-type: none"> - 법률 제1067호(1962년 5월 10 일)
조사기준일	- 11월 1일 0시 현재	<ul style="list-style-type: none"> - 세대주가 그 신고사유가 발생한 날로부터 14일 이내에 신고
조사대상기간	- 11월 1일부터 15일까지(15일간)	
조사항목	<ul style="list-style-type: none"> - 전수조사 42개, - 표본조사 20개, - 시도특성항목 3개 	<ul style="list-style-type: none"> - 성명, 성별, 생년월일, 세대주와의 관계, 본적, 주소, 주소이력 등
조사대상 및 조사범위	<ul style="list-style-type: none"> - 가구, 가구원 - 조사기준 시점 현재 조사지역내에 상주하는 내·외국인 <p>※ 조사대상에서 제외되는 인구</p> <ul style="list-style-type: none"> - 해외취업, 취학중인 자 - 외국외교관, 수행원, 공무로 체류중인 외국인 및 그 가족, 국제연합소속기관 외국직원, 수행원 및 가족 - 국내 주둔 외국군인, 군속 및 그 가족 	<ul style="list-style-type: none"> - 30일 이상 거주할 목적으로 그 관할 구역 안에 주소 또는 거소를 가진 사람 <p>※ 주민등록 대상외의 자</p> <ul style="list-style-type: none"> - 외국인 - 해외이주법 제2조의 규정에 의한 해외이주자
조사방법	- 조사원에 의한 면접조사	- 주민에 의한 등록의 신고주의 원칙

나) 가구주(또는 세대주)와 가구원(또는 세대원)

인구주택총조사에서는 “가구”를 1인 또는 2인 이상이 모여서 취사, 취침 등 생계를 같이 하는 생활단위로 정의한다. 총조사는 가구를 가사단위 (house-keeping unit)로 보고 가구를 정확히 구분하기 위해서 조사원 지침에서 이를 자세히 언급하고 있다(통계청, 2005). 인구주택총조사에서는 가구주를 호적상의 호주 또는 주민등록상의 세대주와 관계없이 가구를 실질적으로 대표하는 사람을 가구주로 하며, 한국인과 외국인이 함께 사는 경우에는 한국인이 가구주가 되도록 하고 있다. 또 혈연 또는 친족관계가 없는 사람이 2인 이상 함께 사는 경우 그 중 한 사람(연장자 또는 대표자)을 가구주로 하고, 나머지 사람에 대해서는 동거인으로 표시하도록 한다.

가구원도 한 가구에서 함께 살고 있는 모든 구성원으로 정의하면서 주민등록이 함께 되어 있더라도 같이 살고 있지 않으면 가구원에서 제외하며, 가족이라도 군복무, 취업, 교육 때문에 함께 살고 있지 않으면 가구원에서 제외한다. 거꾸로 주민등록은 함께 되어 있지 않더라도 같이 살고 있으면 가구원에 포함시키는데, 먼 친척, 친구, 가정부, 종업원 등 가족이 아닌 사람이 함께 살고 있으면 가구원에 포함시켜 조사하고 있다.

가구에서 제외되는 사람은 취업, 취학 등으로 외지에서 하숙, 자취하거나 기숙사에 있는 가족, 고아원, 양로원, 모자원, 특수병원, 부녀보호소 등의 사회복지시설에 수용되어 있는 가족, 요양원, 기도원 등에서 장기간 요양 중인 가족, 군대, 전투경찰대에 입대한 가족, 형이 확정되어 교도소, 소년원 등에 수용된 가족, 선박, 항공기, 철도, 버스 등의 승무원으로 기숙사에서 생활하고 있는 가족, 구직, 무단가출 등으로 집을 떠나서 1개월이 넘는 가족 등이다.

총조사에서는 이렇게 정의된 가구를 일반가구, 집단가구, 그리고 외국인가구로 나누고 있는데 일반가구는 가족으로 이루어진 가구, 가족과 5인 이하의 남남이 함께 사는 가구, 가족이 아닌 5인 이하의 남남이 함께 사는 가구, 1인가구 등이 포함된다. 집단가구는 가족이 아닌 6인 이상 19인 이하의 가구, 기숙사나 고아원·양로원·보육원 등 사회복지시설에 집단으로 살고 있는 가구인데 20인이 넘는 경우에는 집단가구일지라도 기숙시설조사구로 분류된다. 외국인가구는 외국인으로만 구성된 가구인데 한국인과 외국인이 함께 사는 가구는 일반가구로 분류된다.

주민등록부에서는 가구 대신에 세대(世帶)라는 용어를 사용하는데 주민등록법이나 그 시행령 어디에서도 세대를 정의하고 있지 않다. 다만 주민등록 업무지침을 보면 세대란 “주거 및 생계를 같이하는 집단”을 의미하며, 세대

는 “일정한 친족을 중심으로 구성되며 타인(동거인 등) 이 포함되는 경우도 있다.”라고 규정한다. 따라서 주민등록부의 세대 개념과 총조사의 가구개념은 동일한 주거지와 생활단위를 강조하는 측면에서는 비슷하다.

다만 주민등록의 세대는 일정한 친족을 중심으로 구성된다고 하고 있어 혈연관계를 강조한다면 총조사의 가구는 실질적인 생활단위라는 측면을 강조하는 면에서 약간의 차이점을 발견할 수 있다. 여기에 비해서 소득세법 시행령 등에서 사용하는 세대개념은 가족개념을 더 강조된 것으로 볼 수 있다(이명진·서우석, 2008)

한편 주민등록부의 경우 세대주는 그 세대를 대표하는 자로서 주민의 신고에 의하여 선정이 가능한데 다만 성년이 있는 세대에서 고의적으로 미성년자를 세대주로 선정하여 기재하는 것은 금지하고 있다. 이때 세대주는 주민등록신고의 철저한 이행을 위하여 주민등록신고 의무자가 된다. 또 주민등록법 시행령 제6조 제2항은 세대별 주민등록표의 등재순위를 정하고 있는데 세대주, 배우자, 세대주의 직계 존비속의 순위로 하고 그 외에는 세대주의 신고에 따르도록 하고 있다.

현행 주민등록제도는 세대를 주민등록의 단위로 하여 주민등록표를 작성하기 때문에 동일한 세대에 속하는 자는 모두 세대원이라고 정의한다. 주민등록의 대상자는 시, 군, 구 관내에 30일 이상 거주할 목적으로 주소 또는 거소를 가진 자 인데 단 30일 이상을 살아도 거주할 목적이 없으면(예컨대 장기 업무 출장자 등) 제외된다. 주민등록 장소는 30일 이상 거주하게 될 읍, 면, 동인데 한 가지 특이한 것은 영내에 기거하는 군인은 그가 속한 세대의 거주지에서 등록하도록 하고 있다는 점이다. 이는 총조사에서 영내에 거주하는 군인은 그 영내에서 조사되는 것과는 다른 점이다.

총조사에서 가구를 일반가구, 집단가구, 외국인가구로 구분하는데 비해서 주민등록부에서는 세대를 일반세대와 합숙사로만 구분한다. 이때 합숙사란 기숙사나 그 밖에 여러 사람이 동거하는 숙소로 정의한다(주민등록법 제 12조). 주민등록부의 세대주와의 관계는 총조사의 분류 코드에 맞추어서 조정이 가능하다.

2) 가족관계등록부

인구주택총조사를 대체하기 위해서 가족관계등록부에서 얻을 수 있는 자료는 혼인상태, 본관, 출생지이다(통계개발원, 2008b). 이 중에서 본관과 출생지 항목은 그 개념에 있어서 행정자료나 인구주택총조사 자료 간에 차이가 없

으나 혼인상태(marital status)의 개념에는 약간의 괴리가 있다. 먼저 인구주택총조사에서는 혼인신고와 관계없이 사실상의 혼인상태를 기준으로 조사한다. 따라서 이혼은 혼인 후 배우자와 헤어져서 재혼하지 않고 혼자 살고 있는 경우를 말하는데 이혼신고는 되어 있지 않더라도 사실상 이혼상태에 있는 별거도 이혼에 포함시키고 있다. 이에 반해서 주민등록부는 혼인이나 이혼 등 신고에 기초해서 혼인상태를 표시하기 때문에 실질적으로는 혼인이나 이혼 상태에 있지만 신고가 되어 있지 않으면 미혼이나 배우자 있음으로 분류될 수밖에 없다.

나. 인구주택총조사와 건축물대장자료의 비교

1) 거처의 종류

주택에 관한 정보로서 건축물대장으로 부터 거처의 종류, 주거용 연면적, 건축년도, 주택소유여부 등을 알 수 있다(건축법, 2009, 김근영 외, 2003, 양경진, 2003). 먼저 인구주택총조사에서는 거처란 사람이 살고 있는 모든 장소를 가리키는데, 구조적으로 분리되고 독립된 하나의 거주단위로서 모든 거처는 크게 주택과 주택 이외의 거처로 구분한다.

주택은 가구가 독립적으로 살림을 할 수 있도록 지어진 집인데 주택으로 분류될 수 있는 요건은 (1) 영구 또는 준(準) 영구 건물로서, (2) 부엌과 한 개 이상의 방을 갖추고 있으며, (3) 독립된 출입구가 있고, (4) 관습상 소유 또는 매매의 한 단위를 이루고 있을 것 등이다.

주택은 다시 단독주택, 아파트, 연립주택, 다세대주택, 비거주용 건물내 주택으로 분류된다. 단독주택은 일반단독, 다가구단독, 영업겸용단독을 다 포함한다. 아파트는 5층 이상의 공동주택을 말하는데 4층 이하라도 아파트로 허가 받았으면 아파트로 간주한다. 연립주택은 4층 이하의 공동주택으로 건축 당시 '연립주택' 으로 허가 받은 주택으로 연 면적이 660 제곱미터가 넘는 주택이다.

다세대주택은 4층 이하의 공동주택으로 건축 당시 다세대주택으로 허가 받은 660제곱미터 이하의 공동주택이다. 비거주용 건물내 주택은 비거주용 건물에 사람이 살되 그 거주부분이 주택의 요건(방, 부엌, 독립된 출입구) 을 갖춘 경우이다

건축물대장은 용도별로 건축물의 종류를 구분하는데 모두 28개로 분류하고 있다. 건축물대장에서는 건축물을 토지에 정착하는 공작물 중 지붕과 기둥

또는 벽이 있는 것과 이에 부수되는 시설물, 지하 또는 고가의 공작물에 설치하는 사무소·공연장·점포·차고·창고 기타 대통령령이 정하는 것을 말한다고 정의한다(건축법 제2조 2항).

이중에서 총조사의 거처에 해당하는 건축물로는 단독주택과 공동주택의 분류가 있다. 먼저 단독주택은 다시 세분해서 (1) 단독주택, (2) 다중주택, (3) 다가구주택, (4) 공관으로 구분하고 있는데 이 중에서 총조사에서는 사용하지 않는 분류인 다중주택은 여러 사람이 장기간 거주할 수 있는 구조로 되어있으나 독립된 주거의 형태를 갖추지 아니한 것으로 연면적이 330 제곱미터 이하이고 층수가 3층 이하인 건축물을 말한다(건축법 시행령 제 3조 4 관련 별표 1). 또 건축물대장에서는 공관을 별도의 범주로 구분하는 것이 특이하다.

공동주택은 (1) 아파트, (2) 연립주택, (3) 다세대주택, (4) 기숙사 로 구분하는데 아파트, 연립주택, 다세대주택의 정의는 총조사의 정의와 동일하다. 다만 기숙사는 학교, 또는 공장 등의 학생 또는 종업원 등을 위하여 쓰는 것으로서 공동취사 등을 할 수 있는 구조를 갖추되 독립된 주거의 형태를 갖추지 아니한 것으로 정의한다.

따라서 건축물대장에서의 주택의 분류를 인구주택총조사의 분류에 맞추기 위해서는 <표 9-2>에서 보는 바와 같은 약간의 재조정이 필요하다. 즉 건축물대장의 다중주택(1-4) 및 공관(1-5)은 인구주택총조사의 일반단독(1-1)으로 재분류하고, 비거주용 건물내 주택은 영업겸용단독(1-3) 및 비거주용 건물내 주택(5)으로 재분류가 필요하다. 이때 건물별 주택 면적의 합이 건물 연면적의 1/2이상이면 영업겸용단독주택, 그 미만이면 비거주용 건물내 주택으로 구분한다.

인구주택총조사에서 사용하는 거처의 종류를 파악하기 위해서 건축물대장을 활용하는 데 있어서 개념상의 차이에서 오는 문제는 거의 없다고 보여진다.

<표 9-2> 건축물대장의 주택 재분류

건축물 대장		인구주택총조사	
단독 주택(1)	- 단독주택(1-1) - 다가구주택(1-2) - 다중주택(1-4) - 공관(1-5)	단독 주택(1)	- 일반단독(1-1) - 다가구단독(1-2) - 영업겸용단독(1-3)
공동 주택	- 아파트(2) - 연립주택(3) - 다세대주택(4) - 기숙사(8)	아파트(2) 연립주택(3) 다세대주택(4) 비거주용건물(상가, 공장, 여관 등) 내 주택(5)	

자료: 통계청(2007b), 건축물대장 활용방법 연구, 인구조사과 내부자료

2) 건축년도

인구주택총조사에서 건축년도는 신축시기를 원칙으로 하나 증축, 개축면적이 주택 총면적(연건평)의 반 이상이 되는 경우에는 증축, 개축시기를 표시하도록 하고 있다. 또 여러 차례 증축되었을 경우에는 가장 최근의 증축시기를 표시한다. 이에 비해서 건축물대장에서는 사용승인 일자를 건축년도로 본다. 그런데 만약 증축이나 개축이 있을 경우에는 건축물 대장에 별도로 증, 개축시기와 증, 개축 면적이 별도로 표시된다. 따라서 총조사의 정의에 맞추기 위해서는 건축물 대장에서 증, 개축 여부와 그 면적을 따져서 건축시기를 조정해야 한다.

2. 행정자료의 일반적 매칭방법²⁾

- 자료 매칭의 일반적 원칙은 ① 상이한 자료원의 효용가치 극대화, ② 불법적 자료반출의 극소화, ③ 비밀 준수, 사생활 보호, 법률 등의 제약 속에서 통계목적으로 각종 자료에 대한 접근도 향상에 기여(영국정부 성과관리의 기준) 및 ④ 기존의 수집된 자료의 광범하고 안전한 이용의 촉진이라고 할 수 있다(United Kingdom Office of National Statistics, 2003b).

자료의 매칭과 관련된 용어의 정의를 정리하면 다음과 같다.

2) 영국 통계청의 자료매칭에 관한 규정을 중심으로 정리하였다.

① 모조 식별자(Pseudo-identifiers): 주민등록번호, 성명, 주소와 같은 일반적인 식별자를 통계담당자만 알 수 있는 통계코드로 대체하는 “모조 익명화”(Pseudo-anonymisation)의 최종산물. 유사 식별자는 통계목적, 곧 통계처리 과정에서 통계단위를 상호 구분하는 목적으로만 사용가능하다.

② 통계책임관(Responsible Statistician): 통계책임관은 통계자원과 관련하여 결정된 각종 통계정책에 대하여 최종적 책임을 지는 정부 관리다. 통계책임관은 통계처리과정에서 통계단위 관련 자료 및 정보의 비밀유지에도 최종 책임을 진다. 국가통계를 생산하는 부서, 곧 통계청의 수장(首長)이 곧 통계책임관이 된다. 통계청장은 국가통계의 공신력을 유지하기 위하여, 자신에게 유보된 권한을 타인에게 위임해서는 절대 안 된다. 통계청장은 효율성의 최고 수준에서 국가통계를 관리하기 위하여, 통계작성환경이 어떠한 관계없이, 자신이 책임지고 있는 통계자료에 대하여 공적책임을 지켜야 한다

③ 통계목적(Statistical Purpose): 통계조사의 실시나 통계 결과 또는 해석의 작성과 배포에 필요한 자료수집과 자료처리를 포함하는 제반 목적. 이 목적에 의거하여, 1) 개별 통계단위의 식별은 통계사용자의 관심거리가 아니며, 2) 식별 가능한 통계단위 정보는 해당 정보가 정보수집에 직접 관련되는 것이 아닌 경우에는 어떤 방법으로도 사용 불가능하며, 3) 가능한 모든 범위에서, 식별가능한 통계단위에 대한 정보를 누설하거나 추리할 수 있는 형태로 통계 결과표를 제시해서는 안 되며, 4) 통계목적으로 통계단위에 대한 표본을 추출하기 위하여 해당 정보를 수집하거나 이용하는 것은 통계목적의 범위를 벗어나지 않는다고 볼 수 있다.

④ 통계 자료원(Statistical Sources): 센서스, 행정자료, 표본조사 등의 통계단위에 대한 기록의 집합이다.

자료 매칭은 두 자료가 갖는 공통의 정보(또는 매칭 키)를 토대로, 둘 이상의 행정자료를 완전히 또는 부분적으로 통합하는 것이다. 이 작업을 통하여, 별도 자료원에서 획득한 자료의 효율적 이용을 촉진하고, 나아가 서로 다른 자료원의 효용가치를 증대시킨다. 자료 매칭은 추가적 자료 수집의 필요성을 감소시켜, 자료제공자의 부담을 경감시키는데도 기여한다. 그러나 자료 매칭이 동일 조사단위(개인, 가구, 회사 등)에 대한 기록들을 통합하는 것이기 때

문에, 사생활 보호와 자료제공 동의의 문제를 야기할 수 있다 또, 매칭에 사용되는 개인정보는 또한 비밀보장과 신체안전의 문제를 동시에 야기한다.

데이터 매칭 및 연계(data matching and linkage)의 유형을 분류하면 다음과 같다.

① 정확 또는 정밀매칭(Exact Matching): 행정자료 간의 “매칭키”(matching key)라 불리는 핵심변수(예: 성명 또는 주민등록번호)가 완전히 일치하여, 100% 정확한 매칭이 이루어지는 경우다. 정밀매칭이 이루어지는 자료의 비율은 행정자료의 품질과 직결되어 있으며, 많은 경우 정밀매칭과 확률매칭을 배합하여 사용하는 것이 좋다.

② 판단매칭(Judgmental Matching): 행정자료 간의 핵심변수가 완전히 일치하지 않는 경우, 판단매칭을 실시할 수 있으며, 이 경우 수동 작업을 하거나 컴퓨터 프로그램에 의하여 판단매칭을 실시한다.

③ 확률매칭(Probability Matching): 성명, 생년월일, 주소 등을 잘못 기재하여, 자료 상 약간의 불일치가 발생하는 경우, 확률매칭을 실시할 수 있다. 확률매칭을 실시하는 중에는 두 개의 행정자료를 하나 또는 둘 이상의 차단변수(blocking variables)을 이용하여, 그룹으로 묶는다. 행정자료에서 추출한 몇 개의 핵심변수를 비교하여, 일치 정도에 따라 가중치를 부여하며, 이들을 합산하여 종합 일치도(concordance)를 계산한다. 이를 바탕으로, 종합일치도의 최고치를 특정 임계치(critical value)와 비교하여 확률매칭을 실시한다.

3. 자료매칭의 방법론

가. 인구자료

주민등록 자료와 표본조사 자료를 매칭하기 위해서는 두 자료에 공통적으로 담겨있는 정보를 확인하고, 이를 매칭키(matching key)로 활용하여야 한다. 이 연구를 위해 주민등록 자료와 표본조사 자료에 담겨있는 인구자료의 매칭작업을 위해 사용된 매칭키는 주소, 성명, 생년 및 생월, 성별, 가구주와의 관계이다. 인구자료의 매칭키로 사용된 항목들 중에서 주소, 성명, 생년 및 생월, 그리고 성별은 매칭키의 작성과정에 대해 추가적인 설명이 필요하

지 않는 명쾌한 개념이다. 가구주와의 관계 코드는 주민등록 자료의 세대주와의 관계를 변환하여 표본조사 자료와 일치시키는 작업이 이루어진다.

나. 주택자료

건축물대장 자료와 표본조사 자료에 포함되어 있는 주택 관련 자료들의 매칭작업을 위해 매칭키로 사용된 정보는 주소, 층, 건물명, 면적, 건축연도이다.

다. 매칭과정

첫 단계에서 매칭키가 모두 일치하는 사례를 추출하였고, 후속 단계에서는 중요도가 낮다고 판단되는 매칭키를 순서대로 제외해 나가면서 매칭작업이 이루어진다. 매칭작업의 하위단계에서 일부의 매칭키만 일치하는 것으로 확인된 불안한 형태의 매칭사례들은 직접 눈으로 다른 정보들을 대조하는 아이매칭(eye matching)의 과정을 통하여 매칭 여부를 다시 판단한다. 그리고 최종 단계에 이르기까지 매칭되지 않고 남아 있는 사례들은 더 이상 매칭할 수 없는 자료들이다. 매칭작업의 과정에서 중복사례들이 발견되면 이들을 자료에서 제거하고 다시 첫 단계로 돌아가 반복적으로 작업이 이루어졌다.

표본조사 자료의 경우에는 조사표에 기입하거나 컴퓨터에 입력하는 과정에서 오류가 발생할 수 있다. 특히 이름의 경우에는 적지 않은 오류가 발생하는 것으로 판단된다. 이를 감안하여 발음이 비슷한 경우에는 이름 두 글자 중 한 글자만 맞아도 매칭한다. 예를 들면, 소양래-소양례, 강혜인-강해인, 김여숙-김려숙 등이 이에 해당된다. 아울러 같은 주소에서 성과 출생연도가 같은 사람, 그리고 같은 주소에서 이름과 출생연도가 같은 사람을 동일인으로 판단하여 매칭시킨다.

출생연도와 출생월의 경우에도 세는 나이(Korean age)와 만 나이(completed age), 그리고 양력과 음력의 혼용에 따른 오류가 개입할 개연성이 높다. 나이가 높아질수록 실제 나이와 주민등록 나이가 차이나는 경향도 발견된다. 이를 감안하여 매칭작업의 일부 하위단계에서 생년월에 대한 오차 범위를 ± 3 년에 이르기까지 허용한다.

주택자료의 매칭작업과정에서는 동명과 호명을 표준화시키는 작업이 이루어진다. 예를 들어, 101동은 0101, 가동은 000가, A동은 000A 로 변환한다. 일부 사례의 경우에는 이렇게 변환시킨 표준화 동명(호명) 0101 을 두 자리

의 01 로 재변환한다. 건축물대장의 건축연도는 시험조사 자료의 건축연도 범주와 일치하도록 다음과 같이 코드화 한다(① 2008년; ②2007년; ③2006년; ④2000-2005년; ⑤1995-1999년.....⑩1959년 이전).

마지막으로 지적할 것은 아파트 이외의 주택을 매칭하는 과정에서 층에 관한 변수를 생성하는 작업이 필요하다. 예를 들어, 일반단독주택의 2층 주택을 매칭하는 과정에서 표본조사 자료에는 1개 건축물로 표기되었으나, 건축물대장에는 1층과 2층으로 구분되는 경우가 적지 않다. 이러한 문제를 해결하기 위하여 건축물대장과 시험조사 자료에서 주소를 이용하여 층 변수를 생성한다. 다가구와 다세대 주택에서 별도의 호수나 주소가 있으면 이를 이용하여 층 변수를 생성한다. 표본조사 자료의 경우에는 주소와 거쳐번호를 활용하여 층 변수가 생성된 경우가 많다.

제4절 로드맵 2: 관계기관(행안부, 국토해양부) 협력체제 구축

1. 등록센서스 추진에 따른 행정자료의 한계성

인구주택총조사의 자료를 행정자료로 대체하는 데에는 두 자료간의 개념상의 차이 외에도 다음의 두 가지 문제점이 있다. 첫째는 기본적으로 인구주택총조사는 현지의 실사에 기초한 조사 자료인데 반해서 행정자료는 행위 당사자들의 신고에 기초한 자료이기 때문에 자료 생성방법의 차이에서 발생하는 문제점이다. 둘째는 현행 행정자료의 정비와 관련된 신뢰성의 문제이다.

가. 신고자료와 조사자료 간의 불일치

주민등록자료는 기본적으로 신고자료이기 때문에 여기에 따르는 많은 문제점을 노정하고 있다. 가장 큰 문제는 주민등록신고지와 실제 거주지와 불일치이다. 불일치의 원인은 세금과 주택청약, 그리고 부동산취득 등의 이익을 위하여 허위로 등록하는 경우, 또 개인의 경우 실제로는 이동을 하였지만 신고를 하지 않아 등록이 계속 전거주지로 되어 있는 경우, 또 해외 이주자의 경우 해외에서 영주권을 취득하여 출입국관리사무소에 명단이 통보되어 주민등록이 말소되기 까지 계속 주민등록에 기재되어 있는 경우 등 여러 가지가 있다.

여러 원인에 의해서 실제로 인구주택총조사 결과와 주민등록상의 신고 인

구 간에는 괴리가 나타난다. 통계청(2008a)의 조사에 의하면 2005년 11월 1일 기준의 총조사 인구는 47,279천명인데 비해서 2005년 11월 30일 현재 주민등록인구는 49,210천명으로 집계되어 주민등록인구가 총조사 인구보다 1,931천명 더 많은 것으로 드러나 4.1%의 차이가 있는 것으로 나타난다.

전국의 총인구에서 차이가 나는 것은 총조사에서는 누락이 있었을 것이며, 주민등록인구에서는 인구센서스에서 제외된 해외장기거주자가 계속 집계되는 것 때문이다. 실제로 최근의 한 연구(김형석 외, 2008)는 2006년 1월 1일 현재 주민등록은 되어 있으나 해외에 거주하는 한국인이 117만 8천명에 이르고 있는 것으로 추정하였다.

한편 또 다른 연구(통계청, 2007d)는 2005년 11월 1일 현재의 2005년 인구주택총조사 인구와 2006년 1월 1일 현재의 주민등록 인구를 비교하였는데 총조사 인구는 47,041천명이고, 주민등록인구는 48,782천명으로 주민등록인구가 총조사 인구보다 1,741천명이 많아 주민등록인구에 대한 총조사 인구의 비율이 96.4%라고 보고하였다. 이 주민등록인구대비 총조사 인구의 비율은 그 전 2000년과 1995년에도 비슷하게 나타나 각각 96.0%, 96.9%를 기록하였다.

주민등록인구와 총조사인구간에 발생하는 괴리의 가장 큰 원인은 해외장기체류자들로서 이들은 주민등록인구에는 포함되지만 인구주택총조사에서는 조사되지 않기 때문이다. 즉 해외에 3개월 이상 거주할 목적으로 출국하는 사람은 장기이동자로 간주되어 인구주택총조사에 포함되지 않는다. 반면에 주민등록에서는 유학이나 취업 등의 이유로 해외에 3개월 이상 체류하여도 해외이주법에 의하여 이민을 가지 않는 한 주민등록은 계속 유지되고 등복지에서 집계된다. 따라서 인구주택총조사의 대체자료로 주민등록부를 사용하려면 이 해외 장기 체류자들을 별도로 확인하여 주민등록에서 이를 감안하는 작업이 필수적으로 이루어져야 한다.

지역단위별로 주민등록인구와 총조사인구간에 차이가 발생하는 것은 주택청약, 부동산 취득, 가족수당 등 개인의 권리와 의무와 관련하여 거주지를 사실과 다르게 신고하는 것 때문인데 현재 이와 같은 허위신고가 얼마나 되는지에 대한 구체적인 자료는 없다. 그런데 이와 같은 허위신고는 주로 단독세대를 구성할 목적으로 이루어지는 경우가 많아 주민등록에서의 단독세대와 인구주택총조사에서의 1인 가구수를 비교해 보면 대충 그 윤곽을 알 수 있다.

현행 주민등록에서는 세대주와 친족관계에 있는 사람은 동일 주소지에서는 세대 분리를 하지 못하도록 규제하고 있다. 세금과 주택청약 등 다양한 이

유로 별도의 세대구성이 필요한 사람들은 불가피하게 다른 주소지로 허위전입을 신고하고 별도의 세대를 구성할 수밖에 없는데 이들 가공의 세대는 주로 단독 세대일 가능성이 높다. 실제로 2005년 11월 현재 주민등록세대수와 인구주택총조사 가구 수와의 차이를 가구원 수별로 보면 단독 1인 가구에서 약 197만의 차이를 나타내 62.1%라고 하는 가장 많은 괴리를 나타내고 있다(통계청, 2008).

주민등록에서의 실제 거주와 신고된 거주지와의 불일치 정도는 통계청에서 2008년에 실시한 주민등록 확인조사에서 주민등록의 정확성이 점차로 높아지고 있다(김형석 외, 2008). 그러나 아직까지도 주민등록상의 신고지와 실제 거주지와의 불일치 정도는 상당한 수준에 있어 인구주택총조사를 주민등록부로 대체하는 데에 큰 걸림돌이 되고 있다. 현재 행정안전부에서는 이와 같은 불일치를 바로 잡기위해서 1년에 한번 씩은 일제 점검을 통해 이를 바로 잡으려고 하고 있으나 꼼꼼히 집 내부에서 이를 확인하는 것이 아니기 때문에 그리 효과적이지는 못하다.

별도 단독 세대를 구성하면서 전입하는 사람에게 주택임대차 계약서 등의 근거 서류를 요구하기도 하지만 이 또한 허위 작성이 가능하기 때문에 별도움이 되지 않는다고 한다. 허위신고가 적발된 사람에게는 처벌을 하도록 규정하고 있으나 처벌 건수도 거의 없는 것으로 보인다. 따라서 이 허위신고 문제도 향후 주민등록을 활용해서 인구주택총조사를 대체하려면 어떠한 방식으로든지 해결하지 않으면 안 되는 과제가 되고 있다.

나. 행정자료의 정확성 및 신뢰성

인구주택총조사를 기존의 행정자료로 대체하는 데에 발생할 수 있는 두 번째 걸림돌은 자료의 정비가 제때에 제대로 이루어 지지 않아 신뢰성이 문제가 되는 경우이다. 현재 총조사의 주택 항목을 대체하는 데에 건축물대장이 유용하게 활용될 수 있는데 여러 연구에 의하면 현재의 건축물대장은 자료의 신뢰도에 있어서 문제가 되는 것으로 보고하고 있다.

통계청(2007d)의 조사결과에 의하면 건축물대장과 인구주택총조사자료를 1대 1로 매칭해서 일치율을 보면 일반건축물의 경우 주택종류 일치율은 58.1%로 매우 낮게 나타난다. 연건평에 있어서도 두 자료간의 일치율은 60.8%로 낮다. 연건평의 일치율은 주택의 규모에 따라서도 다르게 나타나는데 전체적으로 보면 연건평은 건축물대장보다 인구주택총조사에서 더 크게 조사되었다.

현재 건축물대장이 갖고 있는 문제점으로는 건축물 대장 기재항목 가운데 주요 항목이 기재되어 있지 않은 누락이 항목별로 다르기는 하지만 상당한 정도에 이르고 있으며, 건축물대장의 총괄표제부와 동별 개요자료, 동별 개요자료와 층별 개요자료의 매칭이 미흡하며, 자료별로 매칭되어 있다 하더라도 매칭된 정보 간 내용이 상이한 경우가 있다(통계청, 2007d). 또 현행 건축물대장은 그 관리상에도 여러 문제를 안고 있는데 예를 들면 건축물대장의 작성에 있어서 연도별로 서식 유형이 다양하여 항목별로 기재율이 다르거나 또 건축물대장 카드 자료와 전산자료가 이중으로 관리되고 있어 업무담당자에게 부담이 되고 있다. 즉 건축물대장의 변동이 발생할 경우 전산대장은 반영이 되나 카드대장은 누락되는 경우가 있으며, 반대로 카드대장은 수정이 되나 전산대장에는 반영이 안 되고 누락되는 경우 등 각 자료가 일치하지 않는 문제점이 있다.

이와 같은 여러 문제점 때문에 현재 건축물대장은 단계적으로 일제 정비 중에 있는데 현재의 계획으로 2010년 까지 정비가 완료될 예정이다. 정비내용 중 중요한 정비대상으로 구조, 용도, 대지면적 등 자료의 누락과 오류 시정, 건축물대장과 등기자료의 소유자 상이자료 정비, 집합표제부(동별 자료)는 있으나 전유부(호별 자료)가 누락된 건축물대장 정비 등이 포함되어 있다. 따라서 현재에는 인구주택총조사 중 일부 주택관련 자료를 건축물대장으로 대체하는 데에는 많은 제약이 있지만 예정대로 건축물대장이 정비되면 현재보다 그 활용가능성이 한층 높아 질 것이다.

2. 관계기관 협력체계

앞에서 논의한 행정목적의 등록부 자료의 한계를 극복하기 위하여 주민등록과 건축물대장을 관장하는 행정안전부와 국토해양부와의 협력체계가 무엇보다 필요하다.

2015년에 우리나라가 등록센서스를 추진함에 있어서, 등록센서스의 전제요건에 해당하는 (i) 등록센서스 실시를 위한 법률적 토대와 개인정보의 비밀보호 문제, (ii) 등록센서스 실시를 위한 주축대상(주민등록부, 건축물/주택대장)의 정비와 추가적인 보완 행정자료의 정비 확대 (iii) 개인식별시스템(주민등록번호)의 이용 (iv) 행정자료의 유지 보수를 위한 행정기관의 협력 등이 이루어져야 할 것이다.

제5절 로드맵 3: 개인정보보호 및 행정자료 통합데이터베이스 구축

1. 등록센서스 전환 시 조사구 설정

인구주택총조사가 전통적 디자인에서 등록센서스로 전환될 경우 기존의 인구주택조사구를 유지하는 것은 사실상 불가능할 것으로 예상된다. 등록자료의 경우 5년에 한번이 아니라 수시로 자료 갱신이 이루어지게 되므로 조사구를 5년에 한 번씩 설정하는 방식을 더 이상 적용할 수가 없게 된다. 또한 등록자료와 매칭되지 않은 별도의 조사구 코드를 설정하여 부가하여 관리하는 것도 사실상 어렵다고 할 수 있다.

등록센서스가 시행될 경우, 새로운 조사구는 기본적으로 활용 가능한 등록자료들에 포함된 변수들 중 가장 합리적인 것을 기초로 작성되는 것이 바람직하다. 예를 들면, 주민등록의 주소 중 최종단위인 통·반·리 코드를 활용하는 방안을 생각할 수가 있을 것이다. 다시 말해 임의적 성격의 구분 코드가 아닌 지속적인 구분 코드를 기초로 하는 조사구 설정이 반드시 필요하다. 이렇게 된다면 추출들은 등록자료의 갱신과 더불어 주기적으로 갱신되는 셈이므로 모집단의 변동상황을 거의 실시간으로 반영할 수 있게 될 것이다.

2. 마이크로데이터 제공시 개인정보보호

최근 IT 기술의 비약적인 발달로 국가기관뿐 아니라 학계, 연구소 및 일반 개인들의 원자료에 대한 수요는 눈에 띄게 증가하고 있다. 각 나라의 센서스는 기본적으로 국민들의 세금과 국민들의 협조로 이루어지므로 가능한 많은 사람이 센서스를 통해 생산되는 자료를 충분히 활용하게 해야 할 것이다. 또한 많은 이용자들이 통계자료를 활용하면 할수록 통계자료의 부가가치를 더욱 높일 수 있을 것이므로 조사기관에서는 가능한 한 다양한 자료를 제공하려 노력해야 할 필요가 있다.

그러나 센서스 마이크로데이터에는 개인정보가 담겨 있으므로 이를 무분별하게 제공할 수는 없는 실정이다. 거의 모든 국가에서는 센서스에서 조사된 개인정보의 기밀보호를 법으로 규정하고 있다. 이런 이유로 마이크로데이터 제공을 할 때 부분적인 제약을 가하는 것은 어쩔 수 없는 일이다. 누구나

이용할 수 있도록 제공하는 데이터를 Public Use Microdata (PUM)라 하고, 개인정보가 담긴 데이터를 Confidential Data라고 구분하여 각 나라에서는 각각의 경우에 대해 서로 다른 데이터 제공 정책을 쓰고 있다. PUM의 경우 공개적으로 자료를 제공하는 반면, Confidential Data인 경우는 특정 이용자들에게만 까다로운 절차를 거쳐 이용을 허락하는 서비스를 별도로 운영하는 방식을 채택하고 있다. 일반인들에게 PUM을 제공하는 방식은 국가마다 또는 기관의 성격에 따라 다양한데 여기서는 미국의 경우를 중점적으로 살펴 보자.

가. Public Use Microdata

미국은 센서스 표본조사(전체의 1/6 추출) 데이터를 이용하여 각각 1%와 5% 표본자료를 PUM으로 제공하고 있다. 각각의 자료는 표본조사 데이터 중 에서 다시 대표성 있는 표본을 추출한 것이다. 각각의 자료는 표본조사 자료 중 일부를 다시 추출한 것이므로 자료의 통계적 정확성에 차이가 있다. 따라서 미국에서는 PUM 관련 정보를 제공하는 문서에 추정오차의 범위와 표준 오차 계산 관련 정보를 별도로 제공하고 있다.

나. Confidential Data

미국은 센서스국 산하에 Center for Economic Studies (CES)를 설치하여 정부, 학계, 민간의 센서스 자료 활용 연구를 장려하고 있다. CES는 미국 내에 9개의 Research Data Center (RDC)를 운영하고 있는데 RDC는 주로 큰 도시의 대학이나 연구소 또는 센서스국 관련기관에 설치되었다.

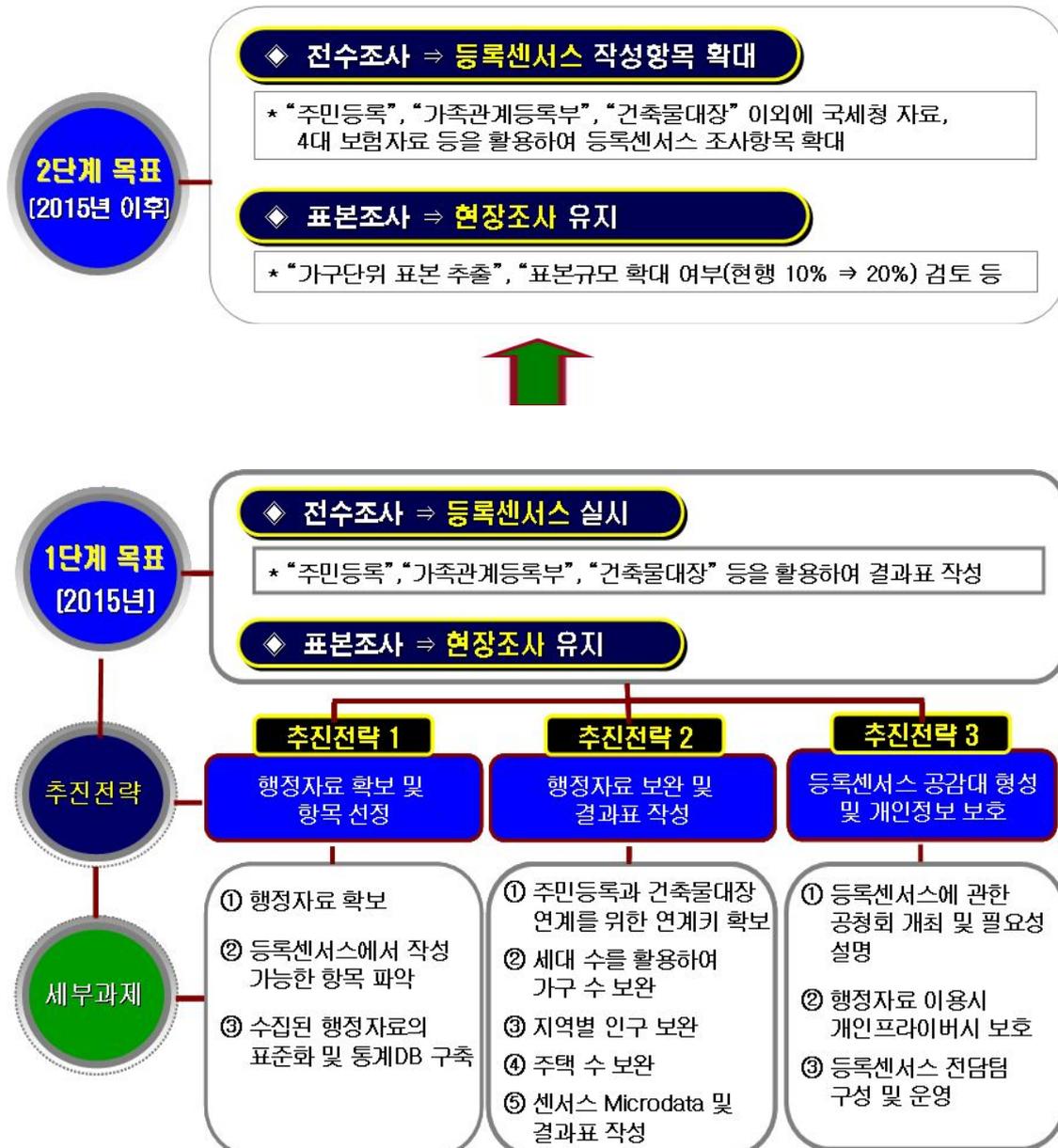
개인정보가 담긴 Confidential Data에 대한 접근은 지정된 RDC 공간 안에서만 가능하다. 원자료는 RDC 안에서만 다룰 수 있고 원자료를 가공하여 얻은 결과만 RDC 밖으로 가지고 갈 수 있게 하는 방식이다. 한편 RDC에 들어가서 자료를 이용할 수 있는 권한을 얻는 데는 많은 제한조건이 따른다. 엄격한 승인절차를 통과해야만 자료에 접근할 수 있는데 CES에서는 이와 관련하여 13개의 기준을 제시하고 있다.

유럽의 경우 Eurostat에서 2002년 Confidential Data의 이용에 대한 규정을 마련하였는데, 여기서도 미국과 유사하게 연구계획서를 제출하여 Committee on Statistical Confidentiality의 승인을 받아야만 이용할 수 있게 하였다.

제6절 통계청 “2015년 등록센서스 기본계획”의 전반적 평가

1. 통계청의 기본계획

가. 등록센서스 추진 체계도



나. 2015년 등록센서스 추진전략 및 세부과제

<추진전략 1: 행정자료 확보 및 항목 선정>

1) 행정자료 확보

등록센서스의 행정자료를 관리하고 있는 행정안전부, 대법원 및 국토해양부로부터 전국단위의 주민등록부 자료, 가족관계등록부 자료, 건축물대장 자료를 확인하고, 전 자료를 공급 받는 것이다.

우리나라 인구·가구 및 주택에 관한 전국단위 행정자료는 주민등록자료, 가족관계등록부, 건축물대장 자료이며, 등록센서스 수행을 위해서는 이들 행정자료 확보가 필수적이다.

□ 향후 추진 사항

○ 2009년~2010년: 시험조사·시범예행조사 지역 및 전국단위 주민등록자료, 가족관계등록부 및 건축물대장 자료 확보

○ 2011년 이후: 매년 11월 1일 기준 전국 주민등록자료, 가족관계등록부 및 건축물대장자료 확보. 이후 변동자료는 분기 또는 반기별로 수집하고, 장기적으로는 실시간으로 공유.

2) 등록센서스에서 작성 가능한 항목 파악

행정기관으로부터 확보한 주민등록부 자료, 가족관계등록부 자료 및 건축물대장 자료를 이용하여 통계를 생산할 수 있는 항목을 선정하는 것이다.

등록센서스를 실시하기 위해서 수집된 행정자료를 검토하여 작성 가능한 항목을 파악하는 작업이 필요하다. 즉, 주민등록자료, 가족관계등록부, 건축물대장 등 행정자료를 분석하여 성별, 나이, 주택소유 현황 등 인구주택총조사 항목을 대체할 수 있는 자료 파악하여야 한다. 만약 등록센서스에서 파악이 불가능한 항목은 표본조사 항목으로 전환하여 자료를 생산하여야 한다.

통계개발원 연구('08년 상반기) 및 인구조사과 내부검토를 통해 행정자료(주민등록, 가족관계등록부, 건축물대장)를 분석한 결과 현재 작성 가능한 항

목은 주민등록에서 8개 항목, 가족관계등록부에서 6개 항목, 건축물대장에서 4개 항목 등 총 18개 항목이 활용 가능하다(<표 9-3> 참조). 앞으로 통계청과 국토해양부와의 협력에 의해 국토해양부가 “총방수”, “주거시설수”를 제공하게 되면 작성 가능한 항목은 20개 항목으로 확대된다.

<표 9-3> 등록센서스 실시 가능 항목

행정자료명	행정자료 항목	(등록센서스 에서) 작성 가능한 항목	항목 수
주민등록 자료	<ul style="list-style-type: none"> • 성명 • 성별 • 나이 • 세대주와의 관계 • 1년전 주소 • 5년전 주소 • 현주소 전입일 	<ul style="list-style-type: none"> • 성명 • 성별 • 나이 • 가구주와의 관계 • <u>가구구분(세대 활용)</u> • 1년전 거주지 • 5년전 거주지 • 거주기간 	8개
가족관계등록부	<ul style="list-style-type: none"> • 혼인사항 • 혼인신고일 • 출생장소 • 친양자 입양사항 • 본관 • 국적 	<ul style="list-style-type: none"> • 혼인상태 • 혼인년월 • 출생지 • 총출생아수 • 본관 • 국적 	6개
건축물대장	<ul style="list-style-type: none"> • 소유자 현황 • 건물의 주용도, 기타용도 • 건물의 층면적 · 호별 전유면적 • 사용승인일자 • 총방수 • 주거시설수 	<ul style="list-style-type: none"> • 주택 소유 여부 • 거처의 종류 • 주거용 연면적 • 건축년도 • 총방수 • 주거시설수 	6개

□ 향후 추진 사항

- 2009년: 행정자료 활용을 통해 작성 가능한 조사항목(초안) 마련
- 2010년: 추가적으로 확대 가능한 항목과약

3) 수집된 행정자료의 표준화 및 통계 DB 구축

수집된 행정자료를 인구주택총조사 조사항목과 일치하도록 자료의 표준화 및 대용량 자료처리를 위한 통계 DB를 구축하는 것이다. 주민등록자료와 건

축물대장의 항목 형식과 인구주택총조사 항목이 다음과 같이 서로 다르다. 그러므로 행정자료를 활용하기 위해서는 행정자료 항목과 인구주택총조사 항목이 일치하도록 자료의 표준화 방안 마련 및 DB를 구축하여야 한다.

- * 인구주택총조사 성별 : 코드번호 1 → 남, 2 → 여
- * 주민등록자료 성별 : 주민등록번호 7번째 자리 9, 1, 3 → 남, 0, 2, 4 → 여

현재 인구주택총조사 시험조사 지역의 주민등록자료와 건축물대장을 활용하여 자료의 표준화 방안을 검토하여 시험조사 지역에 대한 DB를 구축하였다. 행정자료의 표준화 사례를 보면 <표 9-4>와 같다.

<표 9-4> 행정자료의 표준화 사례

항목	총조사 코드	행정자료 항목	행정자료의 표준화
성별	<ul style="list-style-type: none"> • 남 → 1 • 여 → 2 	<ul style="list-style-type: none"> • 주민등록번호 (13자리) 	<ul style="list-style-type: none"> • 주민등록번호 7번째 자리로 성별 구분 * 9, 1, 3 → 남자 * 0, 2, 4 → 여자
나이	<ul style="list-style-type: none"> • 주민등록 생년월일 (만 나이) 	<ul style="list-style-type: none"> • 주민등록번호 (13자리) 	<ul style="list-style-type: none"> • 주민등록번호 7번째 자리로 출생 년대를 파악하고, * 9, 0 → 1800년대 * 1, 2 → 1900년대 * 3, 4 → 2000년대 • 주민등록 1~6째 자리의 생년월일을 이용하여 만 나이 계산
거처의 종류	<ul style="list-style-type: none"> • 단독주택 * 일반단독, * 다가구단독 * 영업겸용단독 • 아파트 • 연립주택 • 다세대주택 • 비거주용 건물내 주택 등 	<ul style="list-style-type: none"> • 단독주택 * 다중주택 * 다가구주택 • 공동주택, * 아파트 * 연립, * 다세대 • 근린생활시설 등 (점포 및 주택 등) 	<ul style="list-style-type: none"> • 건물별, 층별 주용도를 분석하여 거처의 종류 구분 * “다가구주택 4가구, 2가구 주택“ ⇒ 다가구 단독으로 구분 * “점포 및 주택” ⇒ 영업겸용 단독주택 또는 비거주용 건물내 주택으로 구분

□ 향후 추진 사항

○ 2009년: 통계청, 행정안전부 및 국토해양부에서 인구주택총조사와 행정자료의 개념을 통일화

부 처	협의사항
행정안전부	<ul style="list-style-type: none"> • “가구와 세대, 표준세대”의 개념 정의 • 전입신고서에 “점유형태” 항목 추가 • 위장전입 방지 대책
국토해양부	<ul style="list-style-type: none"> • “가구와 세대, 표준세대”의 개념 정의 • “다가구 주택”의 개념 일치 및 정비 • 주택종류가 구분되지 않은 공동주택 정비

○ 2010년 이후: 매년 11월 1일 기준 주민등록자료, 가족관계등록부 및 건축물대장 자료를 활용한 등록센서스 통계 DB 구축

2. 통계청 기본계획에 대한 종합의견

등록센서스의 성공을 위한 조건은 전수추정을 위한 표본조사의 실시, 사전준비를 위한 연구와 시범실시, 정확한 행정자료의 확보·유지와 관계기관간의 긴밀한 협조체계의 확립 등이라고 할 수 있다. 이러한 기본조건 하에서만 행정자료를 근거로 정확한 통계를 추정할 수 있기 때문이다.

가. 표본규모의 확대

기존의 인구주택총조사의 기본조사항목에 대해서는 전수조사이지만 다양한 사회, 복지, 경제 항목에 대한 조사항목은 10% 표본조사에 의존하고 있다. 이 표본조사는 등록센서스에서 기본조사항목에 대한 전국 통계를 추정하기 위해서 필요하고, 전통센서스의 심층조사항목에 대한 조사를 담당하게 된다. 그러나 다음과 같은 이유로 표본규모를 종전의 10% 표본을 20% 표본으로 확대할 필요가 있다.

1) 전수자료는 행정자료와 현장조사인 표본조사 자료를 매칭하여 얻은 승수를 활용하여 추정되는 통계이므로, 비교자료의 규모를 키우면 키울수록 소지역별 전수항목에 대한 추정자료의 신뢰도를 높일 수 있을 것이다.

2) 우리나라의 센서스는 5년 주기로 전수조사를 실시하였으므로 그 시계열을 맞추어 전수자료를 조사자료에 의거 추정하자면 전 가구를 대상으로 5

년 동안 순환센서스(rolling census)를 추진할 수 있다. 이 때 매년 전 가구의 20%를 조사하는 프로그램이 완성된다면 소지역(또는 소단위 특성 등)별 전 수조사항목에 대한 자료를 매년 작성할 수 있을 것이다.

3) 위의 두 가지 이유는 등록센서스 자료를 다양하게 활용하기 위한 목적이라고 한다면, 표본규모를 늘리는 다른 목적은 등록센서스의 위험을 방지할 수 있다는 것이다. 등록센서스의 생명은 조사자료와 행정자료의 차이를 계수(승수)로 보완했을 때의 정확성인데, 처음으로 실시하는 등록센서스에서 신뢰성을 확보할 수 있는 기본이 표본추출률을 높이는 것이기 때문이다.

나. 2015년 등록센서스 준비를 위한 2010년 이후의 과정

통계청의 기본계획에서는 2015년 등록센서스 실시를 명확히 하고 있으나 2010년 이후의 단계별 세부 추진계획에 대한 사전 정리가 필요하다. 다음은 앞으로 2015년 등록센서스의 준비를 위한 최소의 단계를 정리하였다.

1) 2010년 2009 시험예행조사자료 분석

- 2009년 등록센서스를 위한 연구결과에 따라 2009.11.1.기준으로 실시한 전통적 방법의 센서스 최종 종합시험조사자료를 이용하여 특정지역을 대상으로 등록센서스 인구추정
- 등록센서스에 의한 인구추정결과를 분석하고, 필요할 경우 추정방법을 보완하여 2010 Census와 병행하여 일부지역에 대한 등록센서스 시행여부 결정

2) 2010년 11월 2010 Census 병행 일부지역 등록 센서스 시행

- 등록센서스를 2010 Census와 병행 실시하기 위한 기본계획 수립. 이 때 외국의 사례를 직접 확인하고 우리의 실정에 맞게 반영.
- 기본계획에 따라 2010.년 11월 .1일 기준 2010 Census를 실시하는 시기에 일부지역(대도시, 중소도시, 농촌 등)을 선정하여 등록센서스를 병행 실시.

3) 2011년 2010 Census 병행 등록센서스 시범실시결과 분석

- 2010 Census에 병행하여 실시한 등록센서스의 결과를 Census 결과와 비교하여 (1) 2015 Census에 등록센서스 기법을 적용할 것인지의 여부 결정 및 (2) 2015 등록센서스를 위해 연구하고, 보완해야 할 사항을 포함

한 2015 등록센서스 기본계획 수립

4) 2011~2013년 2015 등록센서스를 위해 제도개선과 함께

등록 센서스 시험적용

- 2010 Census에 병행 실시한 등록센서스의 결과에 따라 계속 추진
- 2015 등록센서스를 위한 법적 근거, 행정제도(등록자료의 협조체계 등) 개선
- 성공적인 2015 등록센서스를 위한 시험조사 실시 및 보완연구

5) 2014년 2015 등록센서스 종합시범조사 실시

6) 2015년 2015.11.1 기준 등록센서스 본조사 실시

다. 관계기관 협조체계 구축

- 등록센서스를 위한 기초자료는 행정자료이므로 관계기관간의 협조체계가 긴밀함을 물론 자료의 신뢰성을 높이기 위한 노력이 병행되어야 한다. 통계청과 각 관련행정기관에 담당자를 지정하고, 등록센서스를 위해 필요한 사항을 추진해 나가는 협조체계를 조성하는 것이다. 행정목적의 자료관리가 아니라 통계목적으로 자료관리가 궁극적으로 행정목적의 정확한 자료를 제공할 수 있다는 점을 공히 인식하고 자료의 질을 높여야 할 것이다. 즉, '정기적인 주민등록의 일체 정비'가 아니라 주민등록법에 따라 등록과 변동이 이루어질 수 있는 행정체계화가 더욱 중요하다는 것이다.

- 자료의 질을 높이는 첩경은 행정자료의 DB화이다. 개별 자료가 체계적으로 수록되어 있고, 자료의 변동이 실시간 DB에 반영될 수 있다면 등록자료를 이용한 인구와 주택 기본항목 추정의 정확성을 확보할 수 있을 것이다.

제7절 정책적 건의사항

인구주택총조사 주변 환경의 변화에 대처하고, 기초자료의 상시 확보를 위하여 지금까지 기존의 행정자료를 기초로 인구·가구·주택통계를 추정할 수 있는 통계적 방법을 제시하고, 그 가능성을 확인하였다. 여기서는 등록센서스의 성공을 위하여 필요한 건의사항을 정리하였다. 예를 들면, 소지역별 통계추정을 위한 틀을 제공하고, 다양한 항목에 대한 조사를 실시하기 위해서는 표본의 규모를 기존의 10%에서 20%로 늘리는 것이 필요하다. 이것은

앞으로 순환센서스의 기초자료로도 활용할 수 있을 것이다. 그리고 등록센서스의 발전을 위해서 등록센서스기본법(가칭)을 제정한다면 행정자료를 담당하는 부처와 통계청의 협력체계를 강화하고, 행정자료의 질을 높이는데 도움이 될 것이다. 그 외에도 행정자료의 DB화, 특수집단에 대한 전수조사의 병행, 대국민 홍보 등을 들 수 있다.

1. 등록센서스 기본법의 제정

등록센서스는 통계청의 노력만으로 성공할 수 없다. 국가의 행정자료관리 체계확립이 우선되어야 한다. 관계기관의 협조와 통계청의 방법론적인 접근이 가능할 때 성공할 수 있다. 이 시스템을 위한 법적 근거가 마련되어야 한다. 가장 좋은 것이 ‘등록센서스기본법’이라는 특별법의 제정이다. 이 법은 관련 행정자료에 대한 자료수집, 자료관리, 기관간의 협조, 자료의 질제고, 자료에 대한책임부여 등을 규정함으로써 등록센서스의 성공적인 실시와 자료의 품질 향상을 담보할 수 있는 근거가 되기 때문이다.

2. 관련 행정자료의 DB화

등록센서스의 기본은 행정자료이다. 등록센서스의 성공을 위해서는 각종 행정자료가 정의에 맞게 일정한 분류체계에 따라 실시간 정리가 필수적이다. 그러므로 등록센서스를 준비하는 과정에서 행정자료의 정비와 DB화가 사전에 이루어져야 할 것이다.

관련 행정자료(주민등록, 건축물등록 등)를 이용한 등록센서스 결과는 과거의 전통적인 방법에 의한 인구주택센서스의 결과와 연계하는데 시계열상의 문제가 발생할 수 있다. 만약 행정자료의 표준화가 이루어지고, 시계열 자료가 DB화 된다면 등록센서스 자료와 전통 센서스 자료간의 연계를 가능하게 할 수 있을 것이다.

3. 표본규모의 확대와 전수조사의 병행

표본센서스를 병행하는 목적은 (1) 등록센서스의 추정을 위한 승수를 정확히 산출하고, (2) 종전의 센서스에서 작성하였던 다양한 인구, 가구, 주택의 특성을 조사하는 것이다. 등록센서스에서 소지역(예를 들면, 읍·면·동)별 자료를 추정하기 위해서는 표본규모가 클 때 추정자료에 대한 신뢰성을 높

일 수 있을 것이다. 그리고 우리나라 센서스의 주기가 5년이므로 앞으로 다양한 자료의 정기적인 확보를 위하여 롤링센서스(순환센서스)를 고려한다면 20% 표본센서스는 5년간 전수를 대상으로 실시하게 되므로 매년 통계를 생산할 수 있다.

표본의 규모를 늘려도 자료의 신빙성을 보장할 수 없는 경우가 있다. 이 경우에는 부분적인 전수조사를 적용하여야 할 것이다. 예를 들면, 새로 개발된 지역의 경우 표본추출 자체가 어려우므로 전수조사를 병행하는 것이 좋다. 또한, 외국인조사와 같이 모집단의 분포가 명확하지 않고, 표본들을 확보하지 못할 경우에는 별도로 전수조사를 실시해야 할 것이다. 외국인조사는 다음에 별도로 정리하였다.

4. 외국인조사를 위한 별도연구

인구센서스의 기준은 국내에 거주하는 외국인을 포함하고, 국외에 장기체류하는 내국인을 제외한다. 이 경우 출입국 자료, 외국인 등록자료 등을 이용하여 전수자료를 추정할 수 있을 것이다. 이 경우에도 출입국관리 자료의 DB화가 체계적으로 정리된다면 자료의 신빙성을 확보할 수 있을 것이다. 그러나 지금까지의 경험에 의하면 실제 거주하고 있는 외국인수와 등록 외국인수의 차이가 크고, 외국인에 대한 기본 정보(개인특성, 거주지 등)를 센서스에서 조사하여야 하므로 행정자료에 의한 간접 추정과 함께 전수조사하는 방법을 고려하여 통계작성방법을 택하여야 할 것이다.

5. 표본추출틀과 순환센서스의 기법연구

표본규모를 20%로 확대하는 논의를 앞에서 하였다. 표본규모를 확대하는 것은 등록센서스 자료추정의 신뢰성을 높이는 것만이 아니라 각종 표본조사의 표본들을 제공하고, 앞으로 순환센서스의 도입을 고려할 수 있는 장점을 가지고 있다.

표본추출을 위한 틀(sampling frame)을 무엇으로 하느냐에 따라 표본의 가치가 달라질 수 있다. 전통 센서스의 경우 한 조사구의 크기를 50~60가구 내외로 전국을 구분하고 있으므로 가구를 대상으로 하는 표본조사에서 이 조사구를 표본추출틀로 이용하고, 조사대상으로 활용하였다. 그러나 등록센서스에서는 행정자료가 행정단위별로 작성되므로 표본추출 단위를 행정구

역을 기준으로 하는 것이 편리하다. 지금 예상하기로는 통·반·리와 같이 최소 행정단위를 기준으로 할 수 있다고 하지만, 지역에 따라 행정단위별 규모(인구, 지역 등)가 너무 큰 차이가 있을 때에는 효과적인 표본추출이 어렵다. 또한, 최소 행정단위별 행정자료의 정비가 잘 되어야 표본추출단위로 활용할 수 있을 것이다. 앞에서 언급한 것과 같이 관련기관간의 협조가 절대적으로 필요한 사안이다. 즉, 최소 행정단위별로 자료를 정리할 수 있도록 DB화가 이루어져야 하고, 행정단위의 경계가 명확하여야 하고, 행정단위별 규모의 차이가 적절히 관리되어야 하기 때문이다.

6. 대국민 홍보의 강화

“인구주택총조사를 폐지하고, 등록센서스를 실시한다”라고 했을 때 전통적인 인구주택센서스에 대한 인식이 왜해되고, 행정자료와 표본조사를 토대로 하여 실시되는 새로운 센서스에 대한 이해가 늦어지는데서 나타나는 문제를 최소화하기 위한 홍보가 필요하다.

기존의 센서스는 전수조사로 실시되면서 전 국민이 동참하여야 한다는 인식이 확립되어 있고, 전수조사이기 때문에 자료에 대한 신뢰가 강하게 형성되어 있는 것이 사실이다. 등록센서스로 전환될 경우 표본센서스에 대한 참여의식이 낮아질 수 있고, 더욱이 전수조사를 하지 않은 등록센서스결과에 대한 인식의 변화가 나타날 수 있기 때문이다.

7. 외국의 제도를 직접 파악 후 제도도입

전통적 센서스의 도입과정에서 방법, 정의, 절차 등을 익히기 위한 연수, 워크숍, 세미나 등에 관계직원을 파견하여 훈련을 시켰고, 필요할 경우 외국의 자문관을 초빙하여 센서스의 성공적인 실시를 위해 노력하였다. 센서스의 실시가 거듭되면서 센서스 전문가가 확보되고, 제도적 장치가 마련되어 지금은 다른 나라의 센서스를 위해 자문이나 기술적 지원을 할 수 있게 되었다. 그러나 등록센서스라는 새로운 기법을 도입하는 과정에서 자체적으로 기법을 개발하고, 시행하는데는 한계가 있다.

이미 이스라엘, 싱가포르, 독일, 대만 등에서는 행정자료를 이용한 등록센서스 실시계획을 완성하였으며, 2010 라운드 센서스에 적용할 예정으로 있다. 등록센서스의 정의와 방법도 중요하지만 등록센서스가 가능하도록 준비

한 과정과 내용이 더욱 중요할 것이다. 이것은 단기간의 시찰과 설명으로 습득할 수 없다. 관계직원이 중장기간(예를 들면, 3개월 또는 6개월) 상주하면서 추진과정을 따라 가면서 확인하고, 우리나라의 실정에 맞게 응용하자면 어떻게 해야하는지를 고민하는 충분한 시간과 노력이 필요하다. 다행히 2010년과 2011년에 수 개국에서 등록센서스를 실시하므로 통계청에서는 앞으로 등록센서스의 전담자를 양성하는 차원에서 파견하여야 할 것이다. 이와 더불어 전문가의 현장 확인은 우리나라 등록센서스의 준비와 실시를 위한 실질적인 자문이 가능하게 할 것이다.

부 록

부록 I : 세계 각국의 등록센서스와 센서스 표본조사

A. 스페인의 행정자료와 전수조사 통합센서스

스페인의 2011년 센서스는 2001년 모형을 개선한 것으로, “고전적 센서스”(곧, “고전적”이란 센서스 정보를 조사표와 현지실사에 의하여 구체적이고 포괄적으로 획득한다는 의미)와 “행정자료를 토대로 하는 센서스”를 결합하는 형태가 될 것이다. 이것은 센서스 문헌에서 볼 수 있는 3가지 센서스 유형(고전적 센서스, 100% 완전한 등록센서스, 행정자료와 표본통계조사의 통합센서스)의 요소를 포함하지만, 이들 중의 어느 것과 같다고 할 수는 없으며, “행정자료와 전수통계조사의 통합센서스”라고 명명될 수 있는 새로운 센서스 범주로 분류되어야 마땅할 것이다.

이러한 센서스 유형의 본령은 센서스 비용과 응답자 부담을 경감하기 위하여 센서스와 관련성이 있는 행정자료를 이용하고, 인구를 정확하게 집계하고 행정자료에서 획득 불가능한 센서스 변수를 조사하기 위하여 모집단에 대한 전수통계조사를 실시하는 것이다.

스페인이 2011년에 실시하는 센서스는 이와 가장 유사하다고 판단되는 행정자료와 표본조사를 실시하는 통합센서스(integrated census) 또는 부분 등록센서스(partial register-based census)와 두 가지의 주요 차이점이 있다.

- 행정자료에서 가용할 수 없는 변수는 표본조사가 아니라 고전적 센서스와 같은 모집단 전체에 대한 현지실사에 의하여 획득된다.
- 인구대장을 바탕으로 하는 인구센서스, 특히 인구 수치는 현재로서는 최고의 품질을 갖춘 통계로 간주되기 힘들다. 따라서 고전적 센서스의 전수조사를 통하여, 해당 인구 수치를 전수조사의 결과와 비교하여 점검하고 정정한다. 이러한 유형의 센서스는 인구대장의 포괄범위를 철저히 평가하고, 인구 수치를 조장하고 나아가 고전적 센서스의 특징인 과다집계와 과소집계의 주요 원인을 파악하는데 도움을 주게 될 것이다.

스페인의 이러한 센서스는 유엔이 권고하는 인구주택센서스의 4가지 핵심 원칙인 개별성(individuality), 보편성(universality), 동시성(simultaneity), 주기성(periodicity)을 충족시킨다. 특히 보편성의 원칙은 인구 수치나 기본변수의

몇 개 정도나 모든 센서스 조사항목에 대하여 충족되는 원칙이다.

2001년 센서스에서 이용된 행정자료는 인구대장과 스페인 내무부가 제공하는 부동산 및 신분증 데이터베이스였다. 2011년 센서스에서는 취업자 관련 모든 센서스 변수 자료를 획득하기 위하여 사회보장가입자 자료와 같은 행정자료를 추가하게 될 것이다. 스페인 통계청은 인구대장을 매개체로 하여, 또한 마이크로 수준의 2001년 센서스 자료를 관리하여, 센서스 자료에 분석적으로 풍요로운 장기적 관점(longitudinal perspective)을 제공하고자 하는 목표를 설정하고 있다.

스페인 센서스는 전반적으로 행정자료의 활용을 확대하고 있으며, 그것을 효율적으로 활용하는데도 관심을 기울이고 있다. 그러나 2011년 센서스에서, 행정자료만을 토대로 센서스 조사항목을 집계하는 것이 가능하다고 판단하지는 않는다.

이러한 유형의 센서스가 적합성을 가지기 위한 기술적/법률적 조건은 다음과 같은 것이 있다.

- ① 인구대장의 존재: 인구대장이 인구통계 작성목적에 100% 무결점이라고 할 만큼 신뢰도가 큰 자료일 필요는 없다. 그러나 인구대장이 센서스 과정에서 인구가 몇 명이나 되는가, 누구를 집계해야 하는가, 그리고 어디에서 집계해야 하는가에 관한 잠정적이고, 일차적인 해답을 줄 수 있는 정도의 신뢰도는 있어야 하겠다.
- ② 센서스 목적에 부합하는 별도의 행정자료를 갖는 것이 또한 바람직하다. 가령, 부동산(건축물) 대장, 징세대장, 사회보장등록, 실업자 공적장부, 최종이수학력 자료 등이 여기에 한다.
- ③ 앞의 두 개 조건이 충족된다면, 센서스 목적으로 이들 행정자료 시스템을 이용하지 않는 것이 비효율적이라고 하겠다. 이러한 센서스 유형이 특정상황에서 최적이라고 간주되기 위한 추가적인 조건이 바로 행정자료만으로 실시되는 센서스가 적어도 다음 이유들 중의 하나 때문에 타당성을 결여하게 된다는 것이다.
 - ㉠ 인구대장의 수치가 인구통계 목적으로 사용하기 위해서는 적어도 10년에 한 번씩은 구체적 현실과 비교하여 규칙적이고 포괄적으로 점검하여야 한다.

⑥ 행정자료의 결합이 센서스 정보의 기본적 요구를 포괄하지 못하며, 결측변수들이 양질의 면에서 너무나 중요한 것이기 때문에, 표본조사로서 이들을 적절하게 파악할 수는 없다.

④ 이러한 종류의 센서스는 인구대장과 관련하여, 센서스가 인구대장에 의하여 지원되는 것이냐 아니면, 인구대장도 정보를 업데이트하고 개선하기 위하여 센서스 현지실사를 이용한다는 점에서 이익이 상호보완적인 것이냐의 두 종류의 것이 있을 수 있다. 2001년 스페인 센서스를 위해서, 스페인 통계청은 다음을 필요로 하는 두 번째 유형의 센서스를 계획하였다.

① 인구대장에 관한 특정 입법이 인구대장을 업데이트하기 위하여 센서스 현지실사를 이용할 수 있도록 해야 하며, 센서스 관련 정보를 취급하는데 있어서 통계적 비밀을 엄격히 보장해야 한다.

② 적절한 기술적 조치를 취하여, 행정목적이나 통계목적으로 이용되는 인구대장과 통계적 목적으로만 이용되는 센서스 정보와는 구분될 수 있도록 해야 한다. 2001년 센서스에서는, 이러한 구분이 정보유형별로 조사표를 분리하여, 이러한 구분이 제대로 이루어지도록 노력하였다.

B. 대만의 등록센서스

대만의 2010년 인구주택센서스 실시방안

1. 배경

대만(중화민국)의 경우, 다른 나라와 마찬가지로 인구주택센서스(人口及住宅普查)는 국가의 핵심적 센서스의 하나로 통계법에 의거하여 실시된다. 인구주택센서스는 1949년의 중화민국 정부가 대만에서 수복된 후 모두 4차에 걸쳐서 실시되었다. 처음 2차례의 센서스는 1956년과 1966년에 실시되었다. 1970년과 1975년에는 연속적으로 표본조사가 2차례 실시되었다. 1980년과 1990년에는 각각 전수조사로 조사표 센서스를 실시하였다. 인구주택센서스는 내정부(內政部)의 주관 아래, 과거에 행정원(行政院)의 인구주택센서스 사무처(잠정기구)와 협력하여 실시되었다. 등록센서스 실시 준비작업의 일환으로, 호적법(戶籍法, Household Registration Law)을 1997년에 개정하였으며, 2000년 인구주택센서스는 행정원 주계처(主計處, DGBS-Director-General of Budget, Accounting, and Statistics) 통계국(統計局, Statistics Bureau)로 이관되었다. 행정원 주계처 통계국은 2000년 센서스부터 10년마다 한 번씩 인구주택총조사를 실시하도록 하였으며, 2010년 인구주택센서스는 정부수복 이후 제6차 센서스에 해당한다.

2. 개관

대만의 2010년 센서스는 공적장부, 특히 호적등기부(戶籍登記簿)를 이용하고 동시에 표본조사를 실시하여 사회경제적 인구특성, 생활상태, 주택상황을 포함하는 인구주택통계를 작성하는 통합센서스(integrated census)의 방식으로 진행된다.

이러한 통합센서스를 추진하기 위하여, 대만 통계국은 2000년 센서스의 준비과정에서 미국, 일본, 영국, 캐나다, 스웨덴, 덴마크 등의 인구통계 전문가를 초청하여 전수조사로 진행되는 센서스의 비용절감과 응답자의 부담경감을 위한 대안적 센서스 방안(alternative census designs)에 대한 연구를 시작

하였다. 또 1997년과 1998년에는 대안적 센서스 방법의 현실성이나 비용효율성에 대한 연구를 추진하면서 등록센서스로의 추진을 검토하였지만, 호적등기부 자료의 불완전성으로 등록센서스를 추진하지는 못하였다.

대만 주계처 통계국은 2010년 센서스 시험조사 자료와 호적등기부 자료의 매칭률이 75% 수준으로, 현재의 우리나라보다 행정자료와 센서스자료의 매칭률이 저조한 상황에서, 2010년부터는 전통적 센서스 방식을 포기하고 호적등기부의 인구통계 정보를 활용하고 이것을 표본틀(sampling framework)로 이용하는 표본조사에 의하여, 인구와 주택에 대한 상세정보를 획득하도록 하고 있다.

1) 센서스 기준일: 2010년 12월 26일. 직전 센서스는 2000년 12월 16일이 기준일이어서, 센서스의 주기성(periodicity) 원칙이 엄격하게 지켜지지 않는다.

2) 조사기간: 2010년 12월 26일부터 2011년 1월 22일까지, 조사기간은 4주간 계속된다.

3) 조사대상자: 센서스는 대만의 현재 국경 안에 거주하는 내국인(해외에 근무하는 대만의 공무원과 그들의 가족을 포함하지만, 대만에 파견된 외국인 공무원과 그들의 가족은 제외한다)과 외국인을 포함하여, 주택센서스의 경우 대만 내에 있는 모든 주택/주거단위(housing/dwelling units)를 포함한다.

4) 자료수집방법: 직접면접(개별방문), 우편배포/직접수거, 인터넷 온라인조사 등의 3가지 방법을 본격적으로 이용한다.

5) 조사항목:

- (1) 인구부문: 성명, 성별, 출생연월일(연령), 혼인상태, 가구등록(戶口登錄) 및 거주상태(등록지 및 상주지, 상주하지 않는 이유), 교육정도(최종학력), 현지방언 사용 여부, 5년 전 거주지, 취업상태(산업 및 직업), 근무지 및 학교의 주소, 취업인구와 학령인구의 통근·통학상태, 생존자녀수와 생존자녀의 상주지, 장기요양상태 등
- (2) 주택 및 가구 부문: 점유상태, 주택용도, 총(용적)면적(평수), 방수, 화장실수, 전입 연월일 등

6) 조사지역: 조사지역을 구분하기 위하여, 지리정보시스템(GIS, Geographic Information System)을 이용한다. 모든 조사지역은 110±30개 가구로 구성되며, 전국에서 총 6-7만 개 조사지역을 무작위로 표본 추출한다. 지리정보시스템을 이용하여, 개인면접을 위한 조사지역 기본도를 작성한다.

7) 표본추출방법론

(1) 일반가구: 도시 또는 농촌에서 진(鎭), 시(市), 구(區)를 하위 모집단(sub-population)으로 사용한다. 조사지역을 단위로 하여, '무작위 층화(또는 분층) 군집 추출법'(random stratum cluster sampling method)을 이용하여 표본을 추출한다. 총 조사지역의 16%를 표본으로 추출하며, 표본조사 대상지역의 모든 주택단위(또는 거처)와 인구를 조사대상에 포함시킨다.

(2) 집단가구: 100명 이상을 수용하는 시설이나 기타 특수 성격을 갖는 시설의 경우에는 해당감독기관에 의하여 특별조사를 실시하도록 한다.

8) 자료처리: 센서스 조사표의 응답내용을 광학문자판독장치(OCR, Optical Character Recognition)로 스캔하여 일차적으로 편집 작업을 실시한다.

9) 공적장부 파일의 내용

(1) 내국인: 가구등록(戶籍)파일, 출입국관리기록, 시설인구 및 교도소 수감인구, 학생취학명부(신입생, 재학생), 국민건강보험파일, 장애인인구 등록파일

(2) 외국인: 외국인 체류자 등록파일, 외국인 노동자 파일, 홍콩, 마타오, 중국본토에서 입국하는 외국인 인구 출입국관리기록

(3) 건축물: 진(鎭), 시(市), 구(區) 별 주소대장, 단위거처 파일, 주택세 파일 등 기타 관련 공적장부

10) 비밀보호: 인구주택총조사를 위하여 수집된 자료는 통계작성을 목적으로 하여 사용된다. 자료의 관리와 서비스는 다만 통계법과 관련 규정에 의거하여 이루어진다.

**상세자료

1. 법령근거: 2010년 대만 센서스(人口及住宅普查, 곧 인구주택총조사)는 통계법 제3조, 제4조, 제10조와 같은 법의 시행세칙 제7조, 제25조부터 제32조까지의 규정에 의거하여 실시된다.
2. 센서스의 목적 및 용도: 센서스는 현재 대만의 주권이 미치는 지구의 상주인구에 대한 기본특징, 가구구성 및 주택의 사용상황 등에 관한 자료를 수집하는 것이 주목적이며, 자료를 편집하고 통계분석을 실시한 이후, 다음의 다양한 용도로 사용한다.
 - 1) 전국 상주인구의 특징과 구조, 최신 주택관련 기초자료를 획득하여, 인구 및 주택개발정책의 근거자료를 제공한다.
 - 2) 인구의 규모와 특성, 가족구조를 파악하여, 정부의 사회복지정책 수립에 기여하고, 가족, 직장, 교육 등에 관한 정책구상의 근거자료를 제공하고, 사회안전망 구축에 도움이 되도록 한다.
 - 3) 전국 개별지역, 곧 진(鎭), 시(市), 구(區)별로 인구 및 주택분포를 파악하고, 상주인구의 통근·통학의 방향, 인구규모, 교통수단에 관한 자료를 수집하여, 도시계획, 산업발전, 공공시설, 교통운수 및 도농간 균형발전정책의 근거자료로 이용한다.
 - 4) 전국 개별지역, 곧 진(鎭), 시(市), 구(區)별로 노동력인구의 취업상태, 직업구조 등 관련 자료를 수집하여 정부가 인적자본을 강화하고, 인력운용 및 취업촉진을 위한 정책수립의 근거자료로 이용한다.
 - 5) '새로운 세기 제3기 국가건설 계획'(新世紀第三期國家建設計劃)을 추진하는데 필요한 인구 및 주택관련 자료를 획득하고, 이를 통하여 "활력창신(活力創新), 균부공의(均富公義), 영속절능(永續節能)"이라는 국가발전의 장기적 목표를 달성한다.
 - 6) 전국 개별지역, 곧 진(鎭), 시(市), 구(區)별로 상주인구와 주택총수, 그리고 센서스 조사구의 기본적 특성구조를 확정하여, 가구표본조사를 위한 표본설계와 모수추정(parameter estimation)을 위한 근거자료를 제공한다.
 - 7) 학술기관과 민간단체에 인구 및 주택문제를 연구할 수 있는 기본자료를 제공한다.
 - 8) 센서스 결과를 국제적으로 비교연구하고, 각국 인구발전의 상황을 상호 이해하며, 국가 간 협력교류 사업을 증진한다.

3. 센서스 기준시점 및 실시기간

1) 기준시점: 2010년 12월 26일을 기준일로 하고, 이 날의 0시를 기준시각으로 하며, 기준일 전 1주일 동안(12월 19일-12월 25일)을 기준주간으로 한다.

2) 실시기간: 2010년 12월 26일-2011년 1월 22일 (도합 28일간)

4. 센서스 조사범위: 대만(중화민국) 정부의 통치권에 귀속되는 대만성(臺灣省)과 복건성(福建省)을 센서스 구역범위로 하여, 대만지구 직할시(臺北, 高雄)과 대만성 각현(市)과 복건성 금문현(金門縣), 연강현(連江縣)을 포함한다.

5. 센서스 조사방식: 대만 정부의 공적장부와 보완적으로 실시하는 표본조사를 통합하는 부분 등록센서스(partially register-based census)를 실시한다. 공적등기장부를 매칭 보정하여, 대만정부의 통치권이 귀속되는 지역의 상주인구 총수와 기본 자료를 작성하고, 다시 센서스 대상지역의 16%에 해당하는 지역을 표본으로 선정하여 직접조사를 실시하고, 개별 지역(진(鎭), 시(市), 구(區))의 상주인구 총수와 센서스의 long-form questionnaire의 응답에 해당하는 상세자료를 획득한다.

6. 센서스 조사대상: 센서스 기준시점에 대만 정부의 통치권에 귀속되는 지역, 곧 센서스 범위지역 안의 상주인구(정부가 해외에 파견한 공무원과 그의 가족, 외국인 노동자와 외국인 교포는 포함하고, 대만에 주재하는 외국의 문무관 외교사절이나 그의 가족은 제외한다)와 센서스 범위지역 안의 소유주택을 모두 센서스 조사대상으로 한다.

7. 센서스 실시방식: 직접면접방식, 우편배포/직접수거방식, 인터넷조사 방식을 채택하여 16% 센서스 표본조사를 실시하며, 각 지역의 도시화 정도에 따라 조사방식을 탄력적으로 운용하여 배치하도록 한다.

8. 센서스 자료의 비밀보장: 센서스 과정에서 획득한 개별 자료는 비밀을 엄격히 보장하며, 대만 통계법과 관련 규정에 의거하여 자료의 관리와 서비스는 이루어진다.

9. 센서스 표본조사 항목

- 1) 주택 상황: 거소구분, 점유상태, 주택용도, 주택의 기본적 특성에 관한 문항
- 2) 가구 상황: 가구(戶)별 주택소유상태, 가구등록 및 거주상태 등에 관한 문항
- 3) 인구 상황: 인구의 기본특성, 상시거주 여부와 전출입 여부, 경제활동상태, 직장 및 학교 주소, 통근/통학상태, 내국인 건강상태 등에 관한 문항

10. 공적장부 조사항목

- 1) 내국인 인구: 전체 가구 호적자료 파일, 내국인 출입국 자료 파일, 시설인구 및 교도소 수감자 인구 파일, 학적자료 파일, 건강보험자료 파일, 심신장애자 대장 파일 등 대만 내국인 인구 관련 공적장부 파일
- 2) 외국인 인구: 외국인 체류자 등록 파일, 외국인 노동자 파일, 홍콩, 마카오 및 중국본토에서 입국하는 외국인 인구 등록자료 등 외국인 관련 공적장부 자료
- 3) 건축물: 진(鎭), 시(市), 구(區)의 촌리(村里) 별 주소대장, 단위거처 파일, 주택관련 재산세 파일 등 기타 관련 공적장부

11. 주관기관: 대만 행정원 주계처가 센서스 주관기관이 되어, 센서스 업무를 총괄하여 기획하고 감독한다. 주계처의 센서스 위원회는 여기에 협력하여 각종 자문과 심의 업무를 담당한다.

12. 협력기관과 업무분장

- 1) 센서스의 대상지역에 대하여 각각 특별조사와 일반조사를 관리하기 위한 주관기관을 두고, 특별조사는 전국수준의 해당 유관기관이 주관기관이 되고, 일반조사에 대해서는 지방정부가 중앙정부의 주관기관과 함께 센서스를 관리할 책임을 공유한다. 일반조사 대상지역은 지방정부가 임시로 센서스 실시조직을 설립하여, 해당조사를 책임지고 관리하는 업무를 수행하여야 한다.
- 2) 센서스 각항의 기본항목을 조사하는데 공적장부 파일이 필요한데, 이것은 내정부 호정사(戶政司), 내정부 입출국 및 이민서(移民署), 내정부 사회사(社會司), 교육부 및 각급학교, 법무부, 행정원 노공위원회(勞工委員會), 재정부 재세자료중심(財稅資料中心), 행정원 위생서(衛生署) 중앙건강보험국 등의 관련 주관기관이 책임을 지고 해당 공적장부 파일을 제공하

여야 한다.

13. 센서스 실시 임시조직 및 센서스 구획구분

1) 센서스 실시 임시조직

1. 시설인구 조사조직: 센서스에서 특별조사의 대상이 되는 시설인구, 곧 집단가구를 관리하는 관청의 경우에는, 해당 관청에 센서스 임시조직을 설치하고, 센서스 업무를 총괄적으로 지휘 감독한다.
2. 직할시, 현(시) 조직: 대만지구의 각 직할시, 대만성 각현(시)와 복건성의 금문현, 연강현 정부는 별도의 「인구및주택총조사처」(人口及住宅普查處)을 설치하고, 해당 구역의 센서스 업무를 지휘하고 감독한다.
3. 향(진, 시, 구) 조직: 각 향(진, 시, 구)에 대하여 「인구및주택총조사실」(人口及住宅普查室)을 설치하고, 각 향의 현지실사를 집행하는 책임을 맡는다.

각급의 센서스 임시조직은 업무수요를 균분하여 감독업무를 분장하고, 설치요령은 행정원 주계처가 정하며, 센서스 업무를 종료하면 임시조직은 해체된다.

- 2) 센서스조사구의 구분: 지리정보시스템(GIS, Geographic Information System)을 운용하여, 대만지구 각 직할시, 현(시) 수치화 센서스조사구를 설정하고, 복건성의 금문현, 연강현 2현에 대해서는 실지조사 방식으로 센서스조사구를 설정하여, 센서스 표본조사를 위한 현장실사작업을 진행하고, 센서스 인력을 운용한다.

14. 센서스 인원

- 1) 기획집행 및 자료처리 인원: 센서스 업무 집행기간 중, 행정원 주계처 보유인력의 동원과 감독 작업 외에, 인력수요를 감안하여 인력과견방식으로 대학 이상 미취업의 졸업생을 고용하여 센서스 업무집행, 보고서 작성, 자료처리 등의 업무에 협조하도록 하며, 센서스 업무가 완료된 후 해고하도록 한다.
- 2) 각급 센서스조직 행정인원: 내무행정, 조세행정, 사회복지행정, 군사행정, 경찰행정, 교육행정, 주계행정, 주택행정 관련 인원과 공무원을 센서스조직 행정인원으로 겸용 배치한다.
- 3) 조사인원: 각급 조사인원은 핵심요원, 지도원, 방문원 등을 포괄하며, 이

들은 공무원인원, 학교 교직원이나 민간의 유력인사를 선임하도록 한다. 공무원인원은 각급 지방 센서스 임시조직 안의 내부행정, 재무행정, 사회복지행정, 경찰행정, 교육행정, 주계행정 등 유관인원을 선발하도록 하며, 민간의 유력인사는 촌리의 장(長), 인장(隣長), 자원봉사 홈헬퍼(料理家務者), 건물관리자, 미취업의 대학 이상 졸업생 등을 적당한 인원을 선발하여 센서스를 관리하는 인원으로 하고, 센서스 업무량을 기준으로 하여 조사관련 비용을 지급하도록 한다.

15. 센서스 실시계획: 행정원 주계처가 센서스 실시계획안의 규정을 바탕으로 센서스 실시계획과 각항의 세부작업계획을 감독한다.

16. 표본설계

- 1) 일반가구: 대만의 각 향(진, 시, 구)을 하위모집단(sub-population, 副母體)으로 하여, 무작위 층화군집추출법(random stratum cluster sampling method)을 이용하여 센서스조사구를 최소추출단위로 하며, 조사구의 16%를 추출하여, 조사구 내에 상주하는 모든 개인, 가구, 주택에 대하여 조사표 집계를 실시한다.
- 2) 집단가구: 성격상 특수한 시설인구(장애자, 노인 등의 요양기관, 군대의 각종 부대)나 100명 이상을 수용하는 집단가구에 대해서는 전수조사를 실시하며, 이들 시설인구나 집단가구를 관리하는 중앙부처가 센서스의 주관기관이 된다.

17. 시험조사: 센서스 본조사의 각 단계별 작업계획, 작업방법, 조사문항, 표책 등의 타당성을 점검하기 위하여, 센서스 실시 전에, 적당한 지역과 대표성이 있는 표본을 선정하여, 단계별로 3차에 걸친 시험조사(census tests)를 실시하며, 센서스 본조사를 순조롭게 진행될 수 있도록 한다.

18. 사후조사: 센서스 본조사의 품질과 정확도를 점검하기 위하여, 사후조사(PES, Post-Enumeration Survey)를 실시하여, 집계와 누락과 중복, 응답내용의 정확성을 파악하여 향후 센서스 품질을 개선하기 위한 청사진을 마련한다.

19. 센서스 보고서 작성

1) 보고서 심의: 센서스 통계분석 결과, 행정원 주계처 센서스위원회의 심의를 거쳐서, 보고서를 편찬하여 출판한다.

2) 보고서 배포:

1. 센서스 일차보고서: 2011년 7월 22일 전 완성

2. 센서스 총괄보고서: 2012년 7월 30일 전 완성

전항의 일차보고서와 총괄보고서(분석설명을 포함)가 제출되면, 행정원이 검토하여 이를 공표한다.

20. 주요작업 진척도

1) 기획설계단계: 2007년 7월-2009년 12월

2) 조사전 준비단계: 2010년 1월-11월

3) 조사실시 및 심사단계: 2010년 12월-2011년 2월

4) 사후조사 단계: 2011년 2월-3월

5) 자료처리 및 보고서 편찬단계: 2011년 2월-2012년 7월

6) 연구분석 단계: 2012년 8월-2013년 12월

21. 성과평가: 조사인원에 대하여 협조적인 분위기와 효율적인 센서스 업무 추진을 위하여 조직과 조사인원의 업무처리 능력을 수준별로 평가하여 그에 상응하는 수당을 지급한다.

22. 센서스 비용: 2009년부터 2013년까지 예산안이 4억 대만위안(元)(한국 원화로 약 160억 원에 해당)이 행정원 주계처에 의하여 책정되었다.

23. 센서스 자료의 응용: 센서스 통계자료의 활용을 증대시키기 위하여, 간행물 보고서, CD-ROM, 온라인검색 통계결과만이 아니라, 그래프를 이용한 각 지역별 자료를 제공하고, 공적장부와 조사자료를 연계하여 주제분석을 가능하게 하며, 인터넷 연결을 통하여 분급수권방식(分級授權方式)을 보강하여 센서스 정보를 제공하여 각계각층이 센서스 자료를 광범위하게 활용할 수 있도록 하였다.

C. 독일의 등록센서스

독일의 연방 통계청(Federal Statistical Office Of Germany, Bundesbeamt) 은 1987년 이후 센서스(Volkszählung) 반대운동, 연방재판소의 위헌판결 등으로 센서스를 실시하지 못했다. 1996년 유럽연합의 센서스 실시 권고와 관련하여, 독일정부는 높은 비용과 낮은 응답을 이유로 거부권을 행사하였다. 이에 따라 독일은 유럽에서 2000년 라운드에서 센서스를 실시하지 않은 유일한 국가가 되었다.

<그림 1> 1987년 독일센서스 조사표

VOLKZÄHLUNG 1987
Personenbogen

Bitte Gemeinde angeben: _____

2 34 133 505 5

Rechtgrundlage:
Siehe Haushaltsmatrikelbogen oder Erläuterungsblatt,
die Bestandteile der Erhebungsvordrucke sind.
Stichtag: 25. Mai 1987

Falls Sie einen Abchluß an einer allgemeinbildenden bzw.
berufsbildenden Schule/Hochschule haben

1 Geburtsangaben a) Geburtsjahr _____
b) Geburtsmonat 1. Januar bis 24. Mai _____
25. Mai bis 31. Dez. _____

2 Geschlecht männlich _____ weiblich _____

3 Familienstand ledig _____
verheiratet _____
verwitwet _____
geschieden _____

4 Rechtliche Zugehörigkeit zu einer Religionsgesellschaft Römisch-katholische Kirche _____
Evangelische Kirche _____
Evangelische Freikirche _____
Jüdische Religionsgemeinschaft _____
Islamische Religionsgemeinschaft _____
andere Religionsgesellschaften _____
keiner Religionsgesellschaft rechtlich zugehörig _____

5 Welche Staatsangehörigkeit haben Sie? deutsch _____
griechisch _____
italienisch _____
übrige EG-Staaten _____
jugoslawisch _____
türkisch _____
sonstige/keine _____

6 Wird von Ihnen noch eine weitere Wohnung (Unterkunft/Zimmer) in der Bundesrepublik Deutschland einschließlich Berlin (West) bewohnt? ja _____ nein _____

Falls ja: a) Für Verheiratete, die nicht dauernd getrennt leben: ist die hierige Wohnung die vorwiegend benutzte Wohnung der Familie? ja _____ nein _____
b) Für alle übrigen Personen: ist die hierige Wohnung die vorwiegend benutzte Wohnung? ja _____ nein _____
c) Außerdem für Erwerbstätige, Schüler/Studenten: Gehen Sie vorwiegend von der hierigen Wohnung aus zur Arbeit oder Schule/Hochschule? ja _____ nein _____

7 Sind Sie erwerbstätig? Vollzeit (über 36 Std. in der Woche) _____
Teilzeit (bis zu 36 Std. in der Woche) _____

Mehrere Antworten möglich: a) Auch Landwirte, mithelfende Familienangeh., Auszubildende, Schüler, Zivildienstleistende nicht erwerbstätig _____
b) Maßgebend ist die normalerweise in der Woche getätigte Arbeit: Hausfrau, Hausmann, Schüler(in), Student(in) _____

8 Leben Sie überwiegend von Erwerbe-, Berufstätigkeit, Arbeitslosengeld, -hilfe, Rente, Pension, eigenem Vermögen, Verrentung, Verpachtung, Abanteil, Zuwendungen, Unterhalt durch Eltern, Ehegatten usw., sonstigen Unterstützungen (z. B. Sozialhilfe, BAföG) _____

9 Welchen höchsten allgemeinen Schulabschluss haben Sie? Volksschule, Hauptschule _____
Realschule/gleichwertiger Abchluß (z. B. Mittlere Reife) _____
Hochschulreife (Abitur), Fachhochschulreife _____

10 a) Welchen höchsten Abchluß an einer berufsbildenden Schule oder Hochschule haben Sie? Berufsschule (ohne Berufsschule) _____
Fachschule _____
Fachhochschule (Ing.-Schule, höhere Fachschule) _____
Hochschule (einschließlich Lehrausbildung) _____
b) Welche Hauptfachrichtung hat dieser Abchluß? _____

11 Falls Sie eine praktische Berufsausbildung (z. B. Lehre) abgeschlossen haben: a) Auf welchen Lehrberuf bezog sich diese Ausbildung? _____
b) Wie lange dauerte diese Ausbildung? (Jahre): _____

12 Bitte Name und Anschrift Ihrer Arbeitsstätte oder Schule/Hochschule angeben. Name: _____
Straße/Hausnummer: _____
PLZ: _____ Gemeinde: _____

13 Welches Verkehrsmittel benutzen Sie hauptsächlich (einzige Strecke) auf dem Hinweg zur Arbeit oder Schule/Hochschule? kein Verkehrsmittel (zu Fuß) _____
Fahrrad _____
Pkw _____
U-Bahn, S-Bahn, Straßenbahn _____
Eisenbahn _____
Bus, sonst. öffentl. Verkehrsmittel _____
sonstiges (Motorrad, Moped, Mofa) _____

14 Wieviel Zeit benötigen Sie normalerweise für den Hinweg zur Arbeit oder Schule/Hochschule? enthält, da auf gleichem Grundstück: unter 15 Minuten _____
15 bis unter 30 Minuten _____
30 bis unter 45 Minuten _____
45 bis unter 60 Minuten _____
60 Minuten und mehr _____

15 Sind Sie zur Zeit tätig als Facharbeiter(in) _____
sonstige(r) Arbeiter(in) _____
Angestellter(in) _____
Auszubildender(in) _____
Beamter(Beamtin), Richter(in), Soldat, Zivildienstmitarbeiter(in) mit bezahlten Beschäftigten _____
Selbständiger(in) ohne bezahlte Beschäftigte _____
mithelfende(r) Familienangehörige(r) _____

16 Zu welchem Wirtschaftszweig (Branche, Behörde) gehört der Betrieb (Firma, Dienststelle), in dem Sie tätig sind? _____

17 Welche Tätigkeit, welchen Beruf üben Sie aus? _____

18 Falls Sie eine Nebenberufstätigkeit ausüben, handelt es sich um eine landwirtschaftliche _____
nichtlandwirtschaftliche _____

NUR VOM STATISTISCHEN LANDESAMT AUSZUFÜLLEN
Stichtag (1 a) _____ Hauptfachrichtung (12c) _____ Erwerber Beruf (11 a) _____ Dauer (11 b) _____
2, 3 u. 4. Stelle _____
Matrikelnr., Schule/Hochschule (Firma) _____ Land (12) Gemeinde (12) _____ Straße (12) _____ Hausnummer (12) _____
Wirtschaftszweig (16) _____ Ausgeübte Tätigkeit (17) _____

현재 인구와 단위거처에 대한 자료는 1987년의 센서스 결과를 토대로 업데이트

<그림 2> 1987년 독일센서스 반대시위: 베를린장벽 위의 센서스 조사표



하여 사용되고 있다. 센서스 자료의 업데이트와 노동력조사와 같은 표본조사를 위해 1%의 '마이크로센서스' (microcensus)가 매년 수행되고 있다. 그러나 1987년 이후 독일 통일, 동독에서 서독으로의 대규모 이주, 유럽통합 등 많은 변화가 있어 과거 센서스 결과를 토대로 업데이트 된 현재의 인구수가 최소한 130만 명 이상 과다집계된 것으로 2001년과 2002년에 실시된 센서스 시험조사에서 밝혀졌다. 이와 같은 기존 센서스 결과의 업데이트 된 숫자와 이 숫자를 근거로 하여 이루어지는 표본조사의 부정확성의 증가로 인해 새로운 정확한 숫자에 대한 필요성이 제기되고 있다. 이에 따라 독일에서는 행정자료 활용과 전통적 방식의 현지실사가 결합된 새로운 모델의 2011년 센서스를 추진하고자 계획하고 있다. 이러한 센서스의 실시 결정에는 EU(European Union)의 '유럽연합 센서스' (EU-wide census) 추진도 영향을 미쳤을 것이라 판단된다.

여기에서는 독일의 2011년 등록센서스 추진계획을 구체적으로 알아보기 위해 우선 그 실시 배경과 전개과정을 우선적으로 검토하여 보겠다. 다음은 2001년의 시험조사를 살펴보고, 이에 대한 분석을 바탕으로 현재 추진되고 있는 2011년 센서스 실시 계획을 살펴보고자 한다.

1. 등록센서스의 추진 배경

독일의 등록센서스 실시 과정과 예정된 절차를 시기별로 살펴보면 다음과 같다(<표 1>). 전통적인 조사방식의 센서스가 구서독에서는 1987년에, 구동독에서는 1981년에 실시되었다. 구서독에서 당초 1981년에 실시 예정이었던 센

<표 1> 독일의 등록센서스 추진 일정

연도	주요 내용
1981/1987년	동독(1981)과 서독(1987)에서 마지막으로 전통적 센서스 실시
1992~1995년	전통적 센서스를 대체하는 새로운 센서스 Task force 조직
1996년	유럽연합센서스 가이드라인 초안을 독일 정부 거부
1997년	유럽연합 센서스 2001 Task force 조직
1997년	지자체별 인구등록대장의 품질진단 Taskforce 조직
1998년	연방정부의 행정자료를 토대로하는 센서스 방법 승인
2001년	센서스준비법 및 시험조사
2003년	시험조사 결과에 관한 보고
2006년	2011년 센서스 실시 결정
2007년	2011년 센서스실시준비법 시행
2008년	주소·건물등록대장 사전 준비 시작
2009년	2011년 등록기반 센서스법의 제정 및 공포
2011년	등록센서스 실시 예정

자료: Eppmann et al. (2006)의 재정리.

서스는 두 차례 연기되어 1987년에 실시되었다. 첫 번째 연기는 센서스의 현장 작업과 관련 소요되는 비용에 대해 지방자치단체가 연방정부의 재정적 지원을 요구하는 재정 분담 문제와 관련이 있었다. 두 번째는 1983년에 센서스법이 연방헌법재판소에 의해 위헌 판결을 받음으로써 연기되었다. 개인정보 보호와 관련 1982년 가을부터 센서스 반대 운동이 일어났으며, 약 1,200명의 시민들이 센서스법이 위헌적 요소를 포함하고 있다고 집단소송을 제기하였다. 이에 관한 연방헌법재판소의 위헌 판결의 주된 결정 사유는 센서스의 실지조사에 의하여 획득한 개인자료를 인구대장의 수정을 위해 지방자치단체에게 전달하는 것이 독일기본법(German Constitution)의 개인정보 보호 조항을 위반하는 것으로 간주되었기 때문이다. 이 판결은 여전히 유효한데, 이것은 행정자료를 행정당국이 통계청으로는 전달할 수 있지만 통계조사에 의하여 수집된 마이크로데이터를 행정당국에 다시 전달되는 것이 허용되지 않는다는 것을 의미한다(Federal Statistical Office of Germany, 1992, 2002,

2007a, 2007b).

유럽의 다른 국가들이 2000년대 센서스를 준비·실시하는 동안, 독일은 전통적 센서스의 대안을 찾는데 분주하였다. 1991년에는 조사표와 현지실사에 대한 대안을 논의하면서 등록센서스를 대안으로 모색하였다. 이러한 제안에 대해 지방자치단체의 인구등록부에서 사용 가능한 항목이 제한적이라는 점과 통계적인 목적으로 사용되기에는 등록부에 너무 많은 오류들이 포함되어 있다는 점이 문제점으로 제기되었다. 실제로 독일에는 북유럽의 등록센서스와 유사한 센서스를 실시하는데 필요한 완벽하고, 현실을 정확하게 반영하는 등록부가 존재하지 않았다. 이에 따라 현장조사를 최소화하면서 등록부의 이용을 극대화하는 센서스를 위한 모델이 개발되었다. 등록센서스 실시를 위한 기본 조건들이 호의적이지 않았기 때문에¹⁾ 모델은 점차 복잡해졌고, 많은 방법과 절차들이 독일 등록부가 가지는 특수한 상황에 대비하기 위하여 개발되어야만 했다(Eppmann et al., 2006).

1992년에는 특별연구팀이 조직되어 다양한 센서스 디자인에 대한 연구를 착수하였다. 전통적인 센서스에 비해 비용과 정보 손실을 최소화할 수 있는 행정자료 센서스의 가능성에 대해 논의가 진행되었다. 논의 중에 행정자료 센서스 실시를 위한 필수 조건으로 지방자치단체의 인구대장이 제대로 정비되어야 한다는 사실에 대해 광범위한 공감대가 형성되었다. 그러나 이러한 조정을 위해 소요될 것으로 예상되는 4억 유로의 재원은 행정자료 센서스 실시안을 부결시키게 되는 원인이 되었다. 이러한 배경 하에서, 연방 통계청과 각 주정부 통계청 책임자들이 1995년에 전수조사와 무작위 표본조사를 통합하는 센서스를 제안하였다.

1997년에는 ‘유럽연합센서스 2001(EU-wide 2001)’ 특별연구팀과 ‘지방인구대장의 품질점검’ 특별연구팀이 동시에 조직되어서 센서스 실시 방향 검토와 인구대장의 품질향상을 위한 다양한 노력이 다각적으로 수행되었다. ‘유럽연합센서스 2001’ 특별연구팀은 연방내각과 연방 통계청 및 각 주정부 통계청의 대표로 구성되었다. ‘유럽연합전체 센서스 2001’ 특별연구팀은 행정자료를 토대로 하는 센서스 모델을 대안으로 제시하였다. 이 특별연구팀은 등록센서스의 실시는 가능하지만, 전통적인 센서스에 비해 품질저하는 불가피하다고 판단하였다. 하지만 연방정부와 각 주정부 수상과 상원의원들은 1998년 11월에 개최된 회의에서 전통적인 센서스에서 행정자료를 토대로 하는 센서스로의 이행을 승인하였다.

1) 독일에는 중앙인구대장과 건물-주택대장과 같은 기본대장이 존재하지 않으며, 개인에 대한 식별번호가 없다.

등록부의 품질 개선 목적과 함께, 선거, 여권·병적의 전달 등을 위하여 업데이트가 이루어져야한다는 이유에서 지방자치단체들은 우편을 통해 인구대장을 정비하도록 요구받았다. 지역주민의 전출입, 곧 인구이동이 일어나는 경우에 지방자치단체에 전출입을 신고하는 절차들이 개선되어야 했으며, 서로 배치되는 정보가 관련 행정자료에 기재되어 있을 때 이 정보는 수정·보완되어야 했다. 이러한 노력의 궁극적 목적은 별도의 인구대장에서 거주자의 수를 계산할 수 있는 데이터의 품질을 개선하는 것이었다. 이러한 시도의 성공 여부를 추후에 센서스 테스트를 통하여 평가 작업을 실시하였으며, 결과적으로 실패작이라는 최종 결론에 도달하였다.

행정자료의 품질개선 노력과 함께, 등록센서스의 실시 이전에 관련된 등록부의 품질을 테스트하고, 서로 다른 행정자료를 매칭한 데이터가 등록센서스 실시의 요건을 충족하는가에 대하여 테스트하는 것은 필수적이다. 이 때문에, 시험조사가 2001년 6월에 '센서스 준비법(Census Preparation Law)'에 의해 규정되었다. 센서스 테스트의 조사를 위한 기준일은 2001년 12월 5일로 결정되었다. 한편 이러한 시험조사 결과를 바탕으로 2006년에는 2011년에 등록센서스를 추진하기로 결정되었다.

2. 2001년 시험조사

'유럽연합 센서스 2001' 특별연구팀이 개발한 센서스 모델은 센서스 시험조사에서 테스트되었다. 2001년 센서스 준비법을 통해서, 독일 연방 통계청은 새로운 방법의 심층 테스트를 수행할 수 있는 권한을 위임받았다. 이 법은 응답자 및 관련 기관의 정보 제공 의무, 각종 조사자료의 통합과 조사준비 및 실시를 위한 주소 제공 의무 등을 규정하였다.

시험조사에서는 특히 관련 등록부의 품질과 다양한 자료원으로부터의 데이터 매칭이 검토되어야 했는데, 세부적으로 다음 사항들이 독일 연방 통계청에 의해서 테스트되었다.

- 과다 또는 과소와 관련된 인구등록부의 품질 점검²⁾
- 인구등록부의 통계적 조정 또는 보정수단 제안을 위한 방법의 효율성
- 우편조사에 의해 건물 소유주로부터 얻은 단위거처 자료와 직접 인터뷰 방식으로 거주 가구를 통해 수집한 자료와의 차이
- 인구등록부 자료와 건물-단위거처 센서스로 수집한 자료의 결합을 통한 자

2) 과다(over estimation)는 사람이 실제로 거주하지 않는데 거주하는 것으로 추정된 것을 의미하는데 명목적(nominal)이라는 용어로 표현되기도 한다. 한편, 과소(underestimation)는 실제로 거주하는데 거주하지 않는 것으로 추정된 것을 의미하며 부족(shortage)으로 표현되기도 한다.

동화된 가구 생성 절차의 개선 가능성과 결과의 신뢰성

- 연방노동청 등록부의 품질과 활용성

가. 시험조사의 테스트 모델

Eppmann et al.(2006)에 의하면 2001년 독일의 센서스 테스트 모델은 <그림 3>과 같다.

제1단계에서는 연령, 성별, 혼인상태, 국적, 거주상황, 출생지 또는 출생국가 등과 같은 인구학적 기본변수들이 인구등록부에서 추출된다.

제2단계에서는 하나 이상의 지방자치단체에 등록되어 있는 개인을 파악함으로써 이중등록 여부를 검토한다. 이를 위해, 독일의 각 지방자치단체는 1월 1일, 5월 15일, 9월 1일에 태어난 모든 사람들과 생일이 불명확한 모든 사람들의 기록을 통계청에 송부했다. 처리절차를 통해, 하나 이상의 지방자치단체에 주요 주소를 등록한 사람들이 평가되었다. 이러한 사람들은 면접조사 후, 등록부가 수정되었다.

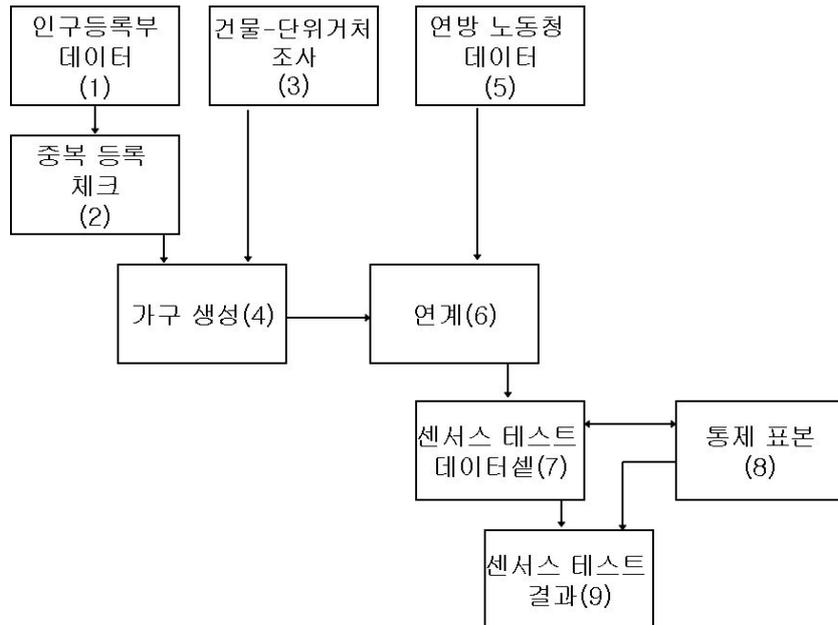
제3단계에서는 건물 소유자 또는 관리자에 대한 우편조사를 통해, 건물 및 단위거처에 대한 조사가 이루어진다. 건물 및 단위거처의 소유자, 규모, 방의 수, 난방시설, 임차료 등이 조사되었다. 전체 가구가 아닌 건물 소유자와 관리자를 대상으로 주택 부문에 대해 조사하기 때문에 응답부담과 소요비용을 경감할 것으로 예상되었다.

제4단계에서 건축물 및 주택 센서스 자료가 인구등록부의 인구관련 데이터와 매칭된다. 이러한 매칭작업은 도로명과 주택코드를 통해 주택 곧 단위거처 수준에서 이루어지는데, 같은 단위거처에 거주하는 사람들을 해당 단위거처에 할당하는 작업은 성(姓), 또는 결혼 전의 성(姓), 전입일자 등에 대한 자료를 기준으로 한다. 이러한 행정자료 데이터와 건물 및 주택센서스와의 매칭작업은 이 모델의 핵심 부분인데, 이를 통해 다음 두 가지의 작업을 수행하게 된다. 첫째 인구등록부의 정보를 건축물 및 주택 센서스의 결과와 매칭하여 가구를 생성한다. 이것은 주택, 곧 단위거처별 거주자의 정보를 획득하는 방법이다. 둘째 인구등록부의 오류가 발견될 수 있다. 예를 들어, 인구등록부에 등록된 사람이 없지만 실제로는 거주하는 사람이 거주하는 주택이 있을 수 있으며 인구등록부에 등록된 사람이 생성된 가구에 적합하지 않고 따라서 단위거처에도 연결되지 않는 경우가 발견될 수 있다. 이러한 특별한 사례들을 전화나 우편을 통해 조사하는 것은 과소등록 또는 과다등록을 줄이는데 도움이 된다.

제5단계에서는 인구대장의 데이터가 연방노동청의 고용대장의 데이터와 매

칭된다. 고용대장은 취업인구의 약 86%를 차지하고 있는 사회 보험의 대상이 되는 사람들과 훈련이나 재교육의 참여자들, 또는 실업자들에 관한 정보를 제공한다.

<그림 3> 독일의 센서스 테스트 모델



주: 괄호 안의 숫자는 단계를 의미함.

제6단계에서 연방노동청의 데이터는 주소, 출생일 또는 성명을 사용하여 생성된 가구 데이터와 매칭된다.

제7단계에서 인구관련 기본변수 데이터, 단위거처 데이터, 노동관련 데이터가 결합된 데이터베이스의 구축이 가능해진다

제8단계에서는 표본조사 결과를 행정자료와 건축물 및 주택 센서스의 자료와 비교하여 행정자료의 품질을 평가한다. 표본조사의 경우에, 건축물 및 주택에 거주하는 가구에 대하여 행정자료와 비교가 가능하도록 인구관련 기본정보와 노동 관련 정보를 수집한다. 이와 함께 건축물 및 주택 센서스와의 비교를 목적으로 관련 데이터도 동시에 요청한다. 행정자료의 집계범위를 추정하기 위하여, 표본조사는 38,000개를 전후하는 범위의 주소를 선정하였다. 이 중에서 행정자료 점검의 핵심 부분인 가구 생성과 관련 방법론적인 검토를 위해 16,000개의 주소로 구성된 하위 표본을 최종적으로 추출하였다.

나. 시험조사 결과

시험조사 결과 인구등록부는 전국 평균 4.1%의 과다등록과 1.7%의 과소등록이 있는 것으로 확인되었다. 하지만 과다·과소 등록을 통해 나타난 등록부의 품질이 연방주 별로 차이가 큰 것으로 나타났으며, 특히 인구 규모가 작은 자치단체보다 인구 규모가 큰 자치단체에서 차이가 큰 것으로 나타났다. 전출입으로 인한 일시적 오류를 줄이기 위하여 4개월 후의 등록부를 고려하였을 때 인구등록부의 과다등록의 비율은 2.9%로 낮아지고, 두 시점간의 중복점검 후에는 2.3%가 되었다. 마지막으로 가구 생성 절차를 거치면서 과다등록은 과소등록과 거의 같은 수준인 1.8%로 낮아진다. 그러나 이것은 인구규모에 따른 편차가 다르므로(자치단체 규모가 클수록 과다등록이 많음) 무시할 수준은 아니라고 판단되었다.

두 번째로 건축물 및 주택 센서스의 결과, 이 조사를 실시하는데 완벽하게 업데이트된 건물의 소유자나 관리자에 대한 주소·건물 명부가 필수적인 것을 확인하였다. 여기에서는 소유자에 대한 조사가 가구에 대한 인터뷰에서 얻는 정보와 동일한지, 동일한 단위거처를 사용하는 것으로 정의된 가구의 생성이 효율적인지, 주소를 활용하여 건물과 단위거처에 대한 우편조사와 등록부를 결합하는 것이 가능한지, 대표세입자의 이름으로 가구 생성이 가능한지 등에 관한 사항이 테스트 되었으며, 그 결과는 성공적이라고 평가되었다.

세 번째로 연방노동청의 등록부가 시험적으로 사용되어 약 90%의 자료가 인구등록부와 연결되었다. 매칭이 되지 않은 나머지의 경우도 최근에 갱신된 주소로 연결된다면 믿을만한 정보를 줄 것으로 테스트 결과 밝혀졌다.

이와 같은 2001년 시험조사에서 새로운 모델은 신뢰할만한 결과를 생산할 것이라는 것을 보여 주었다. 하지만 등록부의 보정과 더 많은 추가적인 변수들에 대한 정보의 필요성으로 인해 추가적인 표본조사가 필요한 것으로 판단되었으며, 이를 토대로 하여 2011년 센서스 실시 계획이 보완되었다.

3. 2011년 센서스 실시 계획

독일에서 계획하고 있는 2011년 센서스는 전통적 센서스 조사방식이 아니라 행정기관에서 관리되고 있는 등록데이터와 현장조사방식을 결합한 새로운 방식이다. 독일의 2011년 센서스 실시 계획은 독일 통계청의 센서스 홈페이지의 내용을 중심으로 살펴보았다. 등록기반 센서스에서 주로 이용되는 등록부는 지자체의 인구등록부와 연방노동청의 고용등록부이다. 이와 함께 건

축물과 주택에 대한 전수조사와 표본조사가 실시될 예정이다(<그림 2>를 참고할 것).

<표 2> 2011년 독일 센서스 실시 계획

구분	이용자료 또는 방법
등록센서스	<ul style="list-style-type: none"> • 약 13,500개 지자체의 인구대장 • 연방고용청의 등록부
전통적 센서스	<ul style="list-style-type: none"> • 주택 센서스 : 국가내의 건물과 단위거처의 소유자에 대한 전수조사 • 집단시설 센서스 : 기숙사 등 집단시설 센서스
표본조사	<ul style="list-style-type: none"> • 독일내 거주자의 최대 10%에 대한 표본조사

2011년 센서스 실시를 위하여 독일에서는 2007년 12월에 2011년 센서스 실시준비법(Census Preparation Law)을 제정하였는데, 이는 2011년 센서스 준비에 관한 사항으로 주소 및 건축물 등록부의 사전 준비, 지방자치단체의 인구등록부 제출 등 2011년 센서스 실시 기준일 이전의 준비사항과 관련 기관의 센서스 업무수행의 실시 근거이다.

새로운 방식의 센서스를 계획하고 있는 독일에서 2011년을 전후하여 센서스를 위하여 이용되는 자료와 그 세부적인 절차들은 다음과 같다.

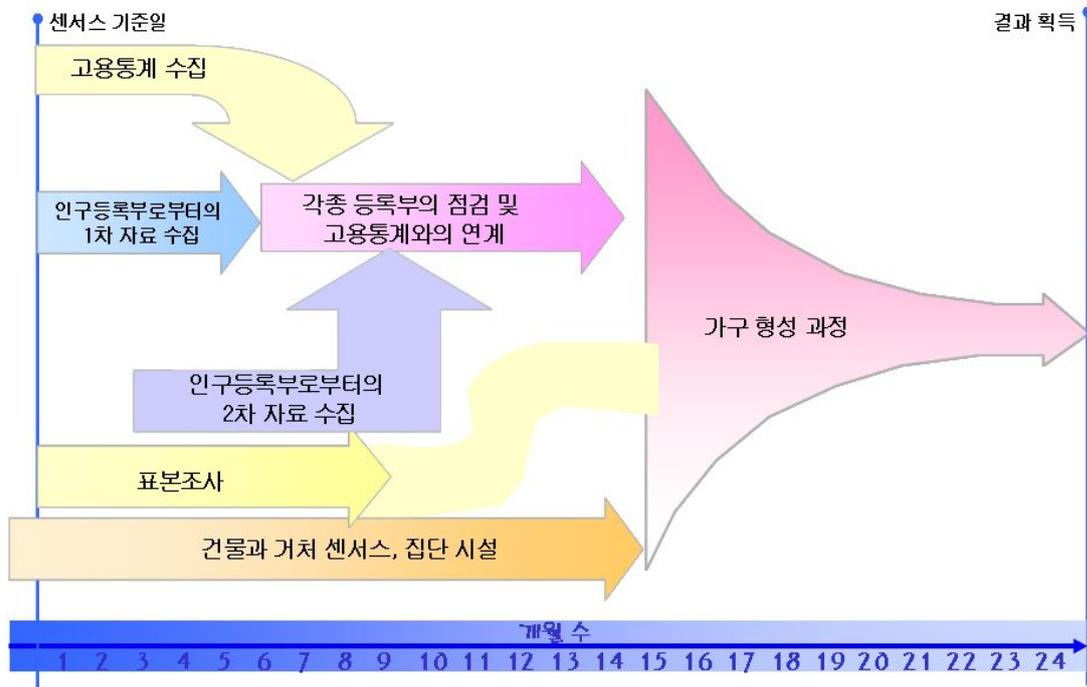
가. 주소 및 건축물 등록부의 사전 준비

등록센서스를 실시하기 위하여 사전에 반드시 준비되어야 할 것이 주소 및 건축물의 명부이다. 주소와 건축물의 명부는 건축물과 주택에 대한 우편 센서스에 사용될 뿐만 아니라 다양한 목적으로 필요하다. 이 주소 및 건축물 등록부의 목적은 1) 센서스가 포괄하는 건축물 및 주택의 완전성 점검, 2) 다양한 자료원의 데이터 매칭, 3) 건축물 및 주택에 관한 우편 센서스 과정 관리, 4) 센서스에 필요한 표본조사의 모집단 제공, 5) 센서스 결과의 소지역 추정을 위한 토대를 제공한다. 주소 및 건축물 명부를 작성하는 작업은 2008년에 미리 시작하여 2010년까지 완료될 예정이다.

나. 행정자료 이용

1) 인구등록부

<그림 4> 독일의 2011년 등록센서스의 실시절차

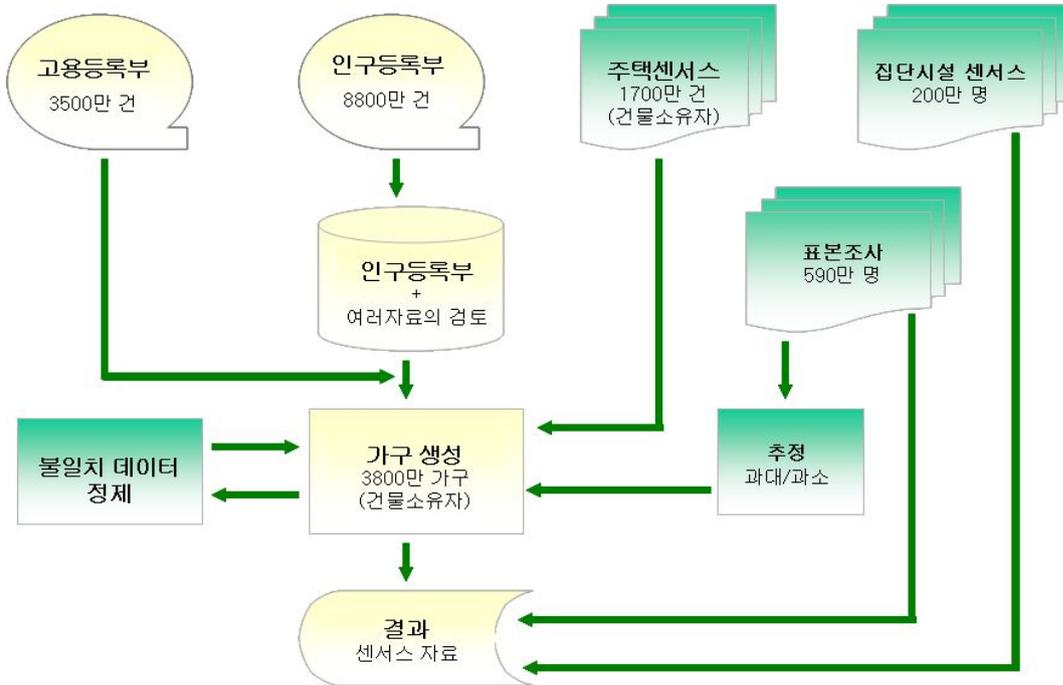


독일은 전국의 약 13,500개의 지방자치단체에서 인구등록부를 분산하여 관리하고 있다. 독일에 거주하는 모든 사람은 일정기간 안에 관할 등록사무소에 등록하므로 인구대장은 등록센서스에서 매우 중요한 자료원이다. 연령, 성, 국적, 혼인상태 그리고 거주지와 같은 기본 인적사항은 인구등록부에서 획득한다.

인구등록부의 품질은 인구대장을 유지·관리해야 하는 의무가 있는 지방자치단체의 노력뿐만 아니라 시민의 행태에도 의존한다. 어떤 사람은 동시에 여러 거처에 등록할 수도 있고, 등록의 지연이 생길 수도 있다. 따라서 자료처리 과정에서 통계당국은 성별, 생년월일, 출생지/출생국, 이름 등의 식별키를 활용하여 이중등록을 점검하고, 중복이 확인된 자료는 수정한다.

인구등록부를 센서스에 활용하기 위하여 각 지자체는 센서스 기준일에 통계청에 인구등록부 자료를 전송하도록 되어 있다. 전송자료는 누락변수 점검 등 완전성을 점검하고, 이중 등록을 점검한다. 이 작업은 8천만 건 이상의 레코드를 비교하는 복잡한 절차이므로 몇 주일이 소요될 예정이다. 센서스 기준일로부터 약 3개월 후에 각 지자체는 다시 인구대장 자료를 추출하여 전송하게 된다. 이는 센서스 기준일 이전에 발생하였으나 지연등록 등의 사

<그림 5> 독일 2011년 센서스 실시도



유로 센서스 기준일 이후에 인구대장에 포함된 행정행위를 반영하기 위한 것이다. 이와 같은 자료의 수정 후 조정된 인구대장은 연방 노동청의 자료와 통합하여 새로운 데이터베이스를 구축하게 된다.

2) 고용대장

인구의 경제활동상태는 인구주택센서스의 핵심적인 주제영역이다. 2011년 센서스에서는 경제활동 관련 항목에 대한 자료를 획득하기 위하여 연방 노동청의 고용대장을 활용할 예정이다.

연방노동청의 고용대장은 의무적인 사회보험 가입자(3,450만 명)와 공공기관 근무자(공무원, 군인, 판사 등 200만 명), 실업자, 연방노동청의 직업훈련 프로그램에 참가한 자 등에 관한 등록부로 2011년 센서스의 고용관련 데이터를 작성하기 위한 기초가 된다. 이 자료들은 인구등록부와 통합되어 새로운 데이터베이스를 구축한다. 그러나 자영업자와 가족노동자와 같은 사람들에 대한 등록부가 없기 때문에 이들에 대한 고용자료는 표본조사를 통해 획득될 것이다

다. 전통적 센서스 실시

1) 건물-단위거처 센서스

2011년 센서스는 인구와 고용형태뿐만 아니라 사람들의 주거상황에 대해서도 데이터를 수집한다. 독일에서 건축물 및 주택 센서스는 1인당 주거면적과 같이 그 결과가 인구센서스와 결합하여 사용될 수 있도록 대부분 인구센서스와 함께 실시되었다. 그러나 독일에는 모든 건물과 주택에 관한 완전한 행정자료가 없기 때문에, 센서스의 방법 변경에도 불구하고 주거관련 통계는 기본적으로 조사에 의해서만 얻어질 수 있다. 그러나 2011년 센서스에서는 1987년의 센서스와는 달리 전체 가구를 조사하는 것이 아니라 약 1,750만 명의 소유자와 관리자를 대상으로 조사한다. 건축물 및 주택 센서스는 우편조사로 실시하며, 전체 3,900만 가구를 집계범위로 하는 저비용의 대체수단이다. 이러한 방식이 독일에서 새로운 것은 아닌데, 이 방법은 1995년 구동독 지역의 건물주택센서스에서 적용되었다. 이 방법의 유용성은 2001년 시험조사에서도 확인되었으며, 주거의 구조적 특성에 관한 문항은 이전의 방법보다 그 내용이 우수하다.

2) 집단가구 센서스

2001년 시험조사 결과에 의하면 집단합숙 서비스를 제공하는 건물이나 시설, 기숙사와 같은 특별한 건물들에서 인구대장의 오류가 많았다. 이러한 오류가 센서스 결과에 영향을 주지 않도록 하기 위하여, 추가적 조사를 실시하도록 되어 있다. 그 결과는 인구대장에서 획득한 데이터를 통계적으로 수정하는데 사용되어, 신뢰할 수 있는 센서스 결과를 생산하는데 사용하게 될 것이다.

실제로 어느 정도까지 현지실사가 필요한지에 관해서는 현재 연구되고 있다. 인구대장의 데이터로 충분한 경우와 그렇지 않은 경우를 결정하기 위하여 다양한 형태의 건물에 대한 연구가 이루어져야 할 것이다.

라. 표본조사 실시

독일 연방 통계청은 등록센서스의 정확성을 확보하기 위하여 표본조사가 필요하다고 판단하고 있으며, 다음의 두 가지 이유로 전체 인구의 7~9%를 조사하는 추가 표본조사가 실시될 예정이다.

첫째는 2001년 시험조사 결과 지방자치단체에 의한 인구등록부의 품질 개선사업이 실시되었음에도 불구하고, 여전히 인구등록부가 센서스 통계 생산에 사용되기에는 부정확하다고 연방과 주의 통계청이 판단하였기 때문이다. 특히 인구규모가 큰 지방자치단체의 과다집계와 과소집계의 비율이 높았기 때문에 약 1,500개에 이르는 인구 규모 1만 명 이상의 지방자치단체에서 대한 추가적인 표본조사 실시를 계획하고 있다. 표본조사에서 얻은 정보를 바탕으로 인구등록부의 과다집계 또는 과소집계의 문제를 통계적으로 보완하여 인구등록부의 품질수준을 개선하고자 하는 것이다. 즉, 인구등록부가 실제인구를 과다집계하는 경우는 전체 데이터에서 개인데이터를 삭제하여 조정하고, 반대의 경우는 통계적 기준에 따라 개인데이터를 추가한다.

다음으로 표본조사를 통해 인구대장으로 대체이용이 불가능한 항목에 관한 자료를 획득해야 하기 때문이다. 가령, 교육, 통근 등에 관한 항목은 행정자료에서 대체이용이 불가능하기 때문에 표본조사를 통해서 자료가 수집되어야 한다. 표본조사를 통하여, 주제 분야가 교육인 경우에 응답자의 최종학력에 대한 정보를 획득할 수가 있을 것이다.

독일의 2011년 등록센서스 기준일자는 2011년 봄으로 예정되어 있는데, 이때부터 인구와 주택에 대한 자료가 수집되고, 자료처리를 거쳐 새로운 인구주택센서스의 통계를 이용하는데 2년 정도가 소요될 것으로 예상되고 있다. 앞서 살펴본 2011년 센서스의 실시절차와 기본계획은 다음과 같다(<그림 4> 과 <그림 5>를 참고할 것)

마. 가구 생성 절차

가구의 규모와 구조, 그리고 그들의 단위거처 상황에 관한 데이터를 수집하는 것은 모든 센서스의 핵심이다. 인구와 주택에 관한 센서스 초안에서도 유럽연합은 회원국에게 국제적으로 공통적으로 사용되는 가구에 관한 데이터 수집을 요청하고 있다. 그 정의에 따르면, 가구는 하나의 주택 곧 단위거처에서 함께 생활하는 사람들인 가사단위의 개념(house-keeping unit concept)으로 정의된다.

가구 내 관계에 관한 정보는 인구대장의 정보를 직접적으로 이용할 수 없으므로, 인구등록부와 건축물 및 주택 센서스의 자료를 통합함으로써 개별 가구

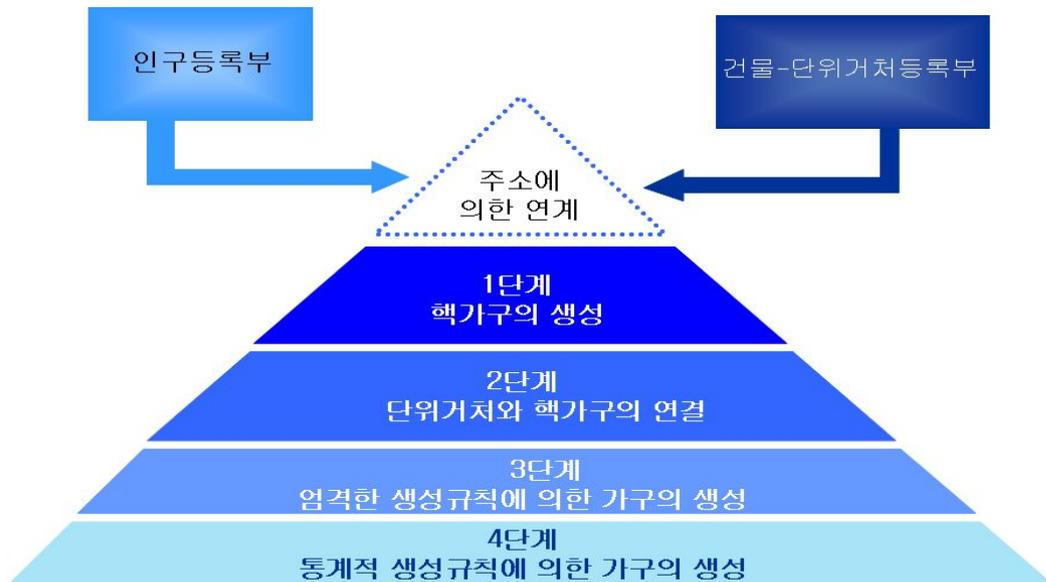
를 생성하는 절차를 밟고 있다. 이 방법으로 약 3,850만 가구가 만들어 질 것이며, 아울러 이 방법은 인구등록부의 오류 등록을 찾아내는 수단으로도 사용될 예정이다. 이 과정에서 등록 오류를 확인하기 위해 1만 명 이하의 지자체에 거주하는 약 170만 명(1인 또는 2인 가구)을 전화, 우편 또는 인터뷰 방식에 의하여 조사하게 될 것이다(Federal Statistical Office of Germany, 2007b).

가구생성절차는 다음의 네 단계로 나누어질 수 있다(그림 5>를 참고할 것).

1) 1단계 : 핵가구 생성

먼저 같은 주소에 등록된 사람들은 핵가구(core household) 생성을 위해 인구등록부의 배우자·자녀·부모의 이름 등에 의한 상호 관계에 관한 정보로 연결된다. 가구 생성 절차의 1단계에서는 단위거처와의 매칭이 이루어지지 않으며, 또한 다양한 형태의 가구를 포괄할 수 있는 가구유형이 충분한 정도까지 만들어지지 않는다.

<그림 5> 독일 2011년 센서스의 가구 생성 과정의 구조



2) 2단계 : 인구등록부와 건축물 및 주택등록부 매칭작업

가구의 단위거처 상황을 결정하는 것은 센서스에서 중요하지만 인구등록부 특정한 주소에 어떤 사람이 어느 주택 또는 단위거처에 사는지에 관한 정보를 포함하지 않고 있다. 그러므로 가구 생성 절차의 2단계에서는 인구와 주

택 또는 단위거처를 연결해야 한다. 이를 위해 특정한 주소에 등록된 사람들의 이름은 주택 또는 단위거처의 소유자 혹은 대표세입자의 이름을 조사한 건축물-주택 센서스의 결과와 비교된다. 만약 동일 이름이 인구등록부와 건축물-주택 센서스의 결과에서 발견되면, 그 사람은 해당 단위거처에 속하는 것으로 간주된다. 핵가구가 1단계에서 생성되었기 때문에 소유자나 대표세입자가 속해 있는 핵가구의 구성원들은 그 주택(단위거처)에 속하는 것으로 간주된다.

3) 3단계 : 엄격한 생성규칙에 기초한 가구 생성

부모와 자식으로 구성된 전통적 가족과 달리 현대사회에는 많은 다양한 형태의 가구가 있으므로, 현실의 가구관계를 반영하기 위해서는 보다 다양한 가구 생성 절차가 필요하다.

가구 생성 절차의 3단계에서는, 인구대장의 추가적 정보가 한 가구에 함께 사는 사람들에 관한 지표로 사용된다. 이것은 성씨, 출생시 이름, 최근 혼인 상태의 변화일, 이전 주소, 전입연월 등과 같은 항목을 비교함으로써 행해진다. 한 단위거처에 거주하는 사람의 수의 척도는 건물 및 주택 센서스 결과 산출된 단위거처에 거주하는 사람의 수이다. 가구생성의 3단계를 거침으로써, 사람들의 대부분은 특정 가구와 주택(단위거처)에 속하는 것으로 분류된다. 가구생성 절차는 건축물 및 주택 센서스에서 거주하는 것으로 파악된 특정 주소지에 더 이상 빈집이 없을 때, 그리고 인구등록부에 있는 특정 주소지에 등록된 모든 사람이 가구로 배분될 수 있을 때 완료된다. 만약에 특정한 주소지에 등록된 한 사람이라도 특정한 거처에 연결될 수 없을 때, 다음의 4단계 가구생성절차를 수행한다.

4) 4단계 : 통계적 생성규칙에 기초한 가구 생성

가구 생성 절차의 4단계에서는 특정한 주소지에 등록되어 있지만 이전 단계에서 한 개의 단위거처나 가구에 연결할 수 없는 사람들을 통계적 기준에 의하여 연결한다. 네 단계가 종료된 후 특정한 주소지에 등록된 모든 사람은 특정 가구와 거처에 배정된다.

가구생성 절차를 마침으로써 모든 사람에 대한 가구, 고용, 거주관계가 확립될 것이다. 가구 생성 절차는 표본조사 결과를 통해, 등록부의 에러를 통계적으로 보정하게 한다. 그리하여 완료된 결과는 최종 인구수로 확정된다.

그 절차는 자료가 획득된 시점부터 시작하여 약 10개월이 소요될 예정이다. 이것은 센서스 기준일로부터 약 24개월 후 이용가능하며, 그때 센서스의 결과는 공표될 것이다.

D. 이스라엘의 통합센서스

1. 통합센서스 이전의 센서스

이스라엘은 1948년 독립국가로 탄생한 후, 5차에 걸쳐 전통적 방식의 인구 센서스를 실시하였다. 최근의 센서스는 2008년에 실시되었는데 완료되었는데, 우리나라가 2015년 인구주택총조사를 실시하고자 할 때, 벤치 마크할 수 있는 방법론을 이용한 등록센서스이다. 한편, 이스라엘의 최초의 센서스는 1948년 국가수립 직후에 실시하였으며, 그 후에는 대략 10년마다 한 번씩 센서스를 실시하였다.(Glickman Hagit , Ronit Nirel, Dan Ben-Hur, 2005, Kamen, 2005; Israel Central Bureau of Statistics, 2006, 2007, 2008).

(1) 제1차 센서스(1948)

이스라엘의 제1차 센서스는 1948년 11월 국가수립 직후에 내무부의 협력을 얻어서 통계국(Central Bureau of Statistics)이 실시하였다. 제1차 센서스에서는, 인구등록대장(Population Register)을 작성하고, 초대 국회의원 선거 실시에 앞서서 주민들에게 신분증을 발급하기 위한 준비작업으로 실시되었다. 1948년 센서스는 이스라엘 국가통계의 토대를 구축하였으며, 그 후, 인구통계는 출생, 사망, 주소변경, 해외 이주민의 입출국 등의 정보를 이용하여 월 단위로 업데이트되었다.

1948년 센서스는 두 단계로 실시되었다. 제1단계로, 조사원이 센서스 기준일자 전에 가구를 방문하여, 인구학적 기본특성, 고용, 히브리어 해독능력 등의 항목에 대한 조사표를 작성하였다. 제2단계로, 센서스 기준일자인 1948년 11월 8일에는 7시간에 걸쳐 통금이 선포되었다. 통금선포의 목적은 조사원의 가구 방문 시점에 모든 가구원이 가택에 거주하고 있는가를 확인하고, 동일 가구원이 두 지역에서 중복 집계되는 것을 방지하기 위한 것이었다. 통금시간 중에, 조사원은 조사를 완료한 가구를 재방문하여, 제1단계에서 집계한 조사표 상의 기록을 다시 점검하고 제1차 방문에서 늘어나거나 줄어든 사람을 가감하였으며, 가구원 각자의 사진을 접수하고, 개인식별번호(personal identification number)가 포함된 증명서를 발급하여, 이것을 나중에 공식신분증을 발급받는데 사용할 수 있도록 하였다.

1948년 제1차 센서스 이후, 1961년, 1972년, 1983년, 1995년에 센서스가 실

시되었는데, 이들도 제1차 센서스와는 다르지만, 2단계로 실시되었으며 센서스 목적으로 통금을 선포하지는 않았다. 제1단계는 조사원이 가구원 모두에 대하여 기본적인 인구학적 정보를 획득하였다. 제2단계는 조사원이 모든 가구의 20%를 선정하여, 사회, 경제, 인구 등의 주제에 대하여 추가적인 상세 자료, 곧 long form 자료를 획득하기 위한 표본조사를 실시하였다.

(2) 제2차 센서스(1961)

1961년 제2차 센서스는 당초의 기획에 의하여 “과학적 센서스:”로 규정되었으며, 이스라엘에서 실시된 최초의 통계목적의 센서스라고 할 수 있다. 당시로서는 방법론이나 결과 자체가 선진국에서 실시된 모범적 센서스들 중의 하나였다. 또, 제2차 센서스는 인구조사를 개선할 목적으로 인구대장의 자료를 광범위하게 이용한 최초의 센서스였다. 인구대장 자료를 센서스의 현지 실사에 이용하는 관행은 그 후에 더욱 더 확대되었다.

1961년 센서스는 2단계로 실시되었다. 제1단계에는 모든 인구에 대해서, 인구관련 주제에 대해서 조사표를 완성하였으며, 제2단계에는 전체 가구의 20%에 대하여 사회경제 관련 주제를 중심으로 표본조사를 실시하였다. 이스라엘 중앙통계국에 의하면, 1961년의 센서스는 20% 표본에 대하여 long form 조사표를 이용하여 광범위한 표본조사를 실시한 세계 최초의 센서스 그룹에 속한다.

(3) 제3차 센서스(1972)

1972년 센서스는 1961년 센서스와 동일한 방법으로 실시되었다. 1972년 센서스에는 컴퓨터가 준비단계에서 사용되었으며, 인구등록대장 자료를 조사표에 사전 인쇄하였으며, 자료처리 단계에서, 센서스 결과표가 처음으로 컴퓨터에 의하여 작성되었다.

(4) 제4차 센서스(1983)

1983년 제4차 센서스에서는 주민들로부터 센서스 정보를 획득하는 방법에 변화가 있었다. 종전 센서스와 유사하게, 제4차 센서스에서 2단계에 걸쳐 조사가 진행되었다.

제1단계에는 두 종류의 조사표를 배부하였다. 하나는 인구학적 기본특성에

관한 short-form 조사표로서 응답자가 직접 기입하는 것이었는데, 이 조사표는 전체 가구의 80%에 배포되었다. 다른 하나는 전체 가구의 20%에 배포된 long form 조사표로서, 여기에는 인구학적 기본특성을 제외하고, 사회, 경제, 인구 관련 주제들에 대한 다양한 조사항목을 포함하였다.

제2단계에는 조사원이 응답자 가구를 재방문하여, 완성된 조사표를 수집하였다. 이 단계에서, 응답자의 협조를 얻어서, 조사항목에 무응답 항목에 대하여 추가질문을 하였다. 이 방법은 제1단계의 조사표 배포단계에서 모든 가구를 집계할 수 있도록 했으며, 제2단계의 조사표 수집단계에서 조사표의 무응답 항목을 줄이는 문제에 집중할 수 있도록 했다. Short form 조사표의 추가는 조사원의 부담을 경감시켰으며, 조사원들이 long form 조사표의 오류를 수정하고, 무응답 항목을 보완할 수 있는 기회를 제공하였다.

이스라엘 중앙통계국의 평가에 의하면, 집계방식의 변화는 종전 센서스와 비교할 때, long-form 조사표에서 무응답의 정도를 대폭 줄였으며, 센서스 정보의 집계범위와 품질을 개선하는데 도움을 주었다.

(5) 제5차 센서스(1995)

1995년 인구주택센서스는 이스라엘에서 실시된 제5차 센서스로서, 1983년 센서스와 유사한 점이 많다. 그러나 제5차 센서스에서는 현지실사나 조사구 작성, 그리고 OCR 시스템 도입 등에서 컴퓨터 도입으로 인한 기술혁신이 있었다. 이러한 변화는 센서스의 집계범위를 개선하고, 현지실사에서 조사원의 부담을 경감하고, 신뢰도가 높고, 고품질의 센서스 정보를 생산하는데 기여하였다.

2. 이스라엘의 통합센서스(2008)

2008년의 이스라엘 통합센서스는 다음과 같은 상호보완적인 다수의 절차를 조합하여 진행된다.

가. 통계적 · 행정적 절차

① 보완행정파일(Improved Administrative File, IAF)의 작성

통합센서스 제1단계 작업은 보완행정파일의 작성으로, 이것은 센서스 모

집단의 골격을 정의하는데 사용한다. 이 파일은 이스라엘 인구대장과 추가적인 행정파일(전력사용자 명부, 운전면허증 교부자 명부, 학생명부 등)을 대조하여 생성되는 자료로서, 인구대장의 등록누락을 최종적으로 보완한다는 의미를 지닌다.

② 인구센서스 표본추출

센서스를 실시하기 위하여, 두개의 상호독립적인 표본, 곧 현장 조사구 표본과 보완행정파일에 등재된 모집단의 표본에 대한 조사를 실시한다.

현장 조사구 표본조사: 현장 표본조사의 표본을 선정하기 위하여, 이스라엘 통계청은 전국을 조사구로 구분한다. 조사구는 평균 50개의 가수로 구성된 연속적인 지리공간으로 이루어진다. 이 조사구 모집단에서, 현장 표본조사를 위한 조사구를 무작위로 추출한다. 이러한 표본 조사구에 대하여, 현장 조사원은 전통적 센서스 방법에 의거하여, 가구별로 모든 조사대상자를 빠짐없이 조사한다.

보완행정파일 모집단의 표본: 보완행정파일에 주소가 등재되어 있는 사람 중에서, 표본 조사구에 거주하는 사람만을 대상으로 표본을 선정한다.

나. 인구센서스 표본조사 실시

① 표본 조사구의 가구집계-Undercoverage Check Sample

이 단계에는, 현장 조사원이 모든 표본 조사구에 대하여, 모든 건물과 거처를 방문하여, 추출된 조사구별로 개인의 주택에서, 컴퓨터의 지원을 받아서 조사표를 작성함으로써, 조사대상자로부터 필요한 자료를 획득한다.

이와 같이, 현장실지조사에 의거하여 표본 조사구의 조사대상자 자료를 토대로 하여, 조사대상자가 집계된 주소와 보완행정파일 상의 주소 간에 매칭을 실시한다. 이 매칭작업의 결과에 의거하여, 보완행정파일 주소상의 모집단 과소집계(Population Undercoverage)에 대한 정보를 획득한다.

② 전화면접조사(Computer-Assisted Telephone Interviewing, CATI) -Overcoverage Check Sample

보완행정파일 주소에는 등재되어 있으나 현장실지조사에는 발견되지 않은 모든 사람들은 전화로 면접을 실시하여, 그들이 실제로 거주하는 장소가 어디인가를 확인한다. 이 조사결과는 보완행정파일 주소에서의 모집단 과다집계(Population Overcoverage)를 추정하는데 사용한다.

(이스라엘의 인구센서스 표본조사)

이스라엘 중앙통계국(Central Bureau of Statistics)는 2009년 7월 2일자로, 웹사이트에 2008년 인구센서스 표본조사 전화확인조사를 통한 자료수집이 종결되었다고 발표했다. 해당 표본조사는 두 단계로 진행되었다. 제1단계는, 현장실지조사로서, 1:1 면접조사가 실시되었다. 이 단계는 2008년 12월 28일부터 3월 8일까지 100일 동안 계속되었다. 이 기간 중, 대략 2,000명의 현장조사원이 동원되어 33만 가구의 130만 명에 해당하는 인구를 대상으로 현장 실지조사를 실시하였다.

제2단계는 필요정보를 확인하기 위한 조사로 전화면접으로 실시되었다. 이 단계에서, 이스라엘 통계청은 센서스 표본조사의 표집대상으로 제1단계에서 면접불응으로 현장 실지조사 대비 인구대장 과다등록으로 판단한 10만 가구에 대하여 전화확인조사를 실시하였다. 전화확인조사는 2009년 3월 25일에 실시하여, 2009년 7월 2일에 종료하였다.

이스라엘 중앙통계국은 센서스 표본조사에서 획득한 정보를 이용하여, 이스라엘의 인구와 그 인구의 인구학적, 경제적, 사회적 특성에 대하여 신뢰도가 높은 통계를 작성하기 위하여 노력하고 있다. 최종적으로, 현장 조사원들에 협력하여 조사가 성공적으로 끝나도록 도와준 이스라엘 국민들에게 감사를 드린다는 이야기를 하고 있다.

다. 센서스 통계생산

- ① 통합센서스 파일(Integrated Census File, ICF)

통합센서스 파일은 보완행정파일에 등재된 개인의 인구학적 상세 특성으로 구성된다. 또 보완행정파일에 등재된 개인은 모집단에서 실제로 몇 명을 대변하는가를 명시적으로 보여주는 센서스 가중치를 포함한다. 보완행정파일에 등재된 개인의 센서스 가중치가 2.0이면, 그는 모집단에서 실제로 2명의 특성을 대표하게 된다. 곧, 해당 개인은 보완행정파일 주소지에 실제로 거주하는 사람 1명과 보완행정파일 주소에 누락되어 다른 주소지에 거주하는 사람 1명을 대표한다.

② 사회경제 파일(Socioeconomic File, SEF)

사회경제파일통합센서스 파일은 현장실시조사에서 집계된 가구와 개인에 대한 자료로 이루어진다. 이 파일은 인구학적 속성과 사회경제적 속성을 모두 포함하는 파일이다. 표본 조사구에서 집계된 모든 사람들은 모집단의 특성을 대표할 수 있도록 센서스 가중치를 부여받게 된다. 그 결과, 우리가 수도권(텔아비브)에서 조사된 모든 개인들의 모든 센서스 가중치를 합산하게 되면, 통합센서스 파일의 인구 총수와 사회경제 파일의 인구 총수는 동일하게 된다.

③ 센서스 통계의 이용자를 위한 서비스 제공

통합센서스 파일과 사회경제 파일에서 통계이용자들을 위한 다양한 센서스 통계가 작성된다. 그 중에는, 센서스 실시연도의 통계적-지리적 행정구역별 인구학적, 사회경제적 특성에 대한 통계표를 작성할 수 있는데, 여기에는 교육, 경제활동과 같은 통계표가 그 예에 해당하게 될 것이다. 또, 주소불명으로 인하여 통계처리에 어려움이 있는 사람들의 파일도 작성될 수 있을 것이다.

3. 통합센서스의 주요용어

용어	개념 및 정의
보완행정파일	보완행정파일은 인구센서스의 기본파일로서, 센서스 인구의 프레임 을 제공하며, 인구대장파일과 여타 행정등록파일을 통합하여 작성된 다.
센서스 표본	이스라엘 센서스에서는 두개의 독립적인 표본, 바로, 현지실사 조사 구역 표본과 보완행정파일에 등록된 모집단의 조사구역 표본이 그것 이다. 현지실사 조사구역 표본-표본선정을 목적으로, 전국을 정해진 기준 에 따라서 조사구역으로 구분한다. 조사구역은 연속적인 지리적 공 간으로 평균 50개 정도의 가구로 구성된다. 이 조사구역에서 현장실 사를 위한 조사구역이 무작위로 추출되고, 면접원은 전통적 센서스 방법으로 해당구역의 건물과 건물 안에 거주하는 모든 사람들을 예 외없이 조사한다. 모집단의 표본-보완행정파일에 있는 주소가 표본으로 추출된 현지 실사 조사구역에 있는 개인들의 표본을 가리킨다.
인구센서스 표본 조사	인구대장의 집계범위를 점검하기 위하여, 현지실사와 전화조사로 구 성되는 표본조사를 인구센서스 표본조사라고 한다. 최근 이스라엘의 통합센서스에서, 인구센서스 표본조사가 2008년 말에 시작하여 2009년 중반에 끝났다.
전화조사	통합센서스에서, 현지실사 자료와 보완인구대장 파일 자료를 매칭하 여, 표본 조사구역에서 보완인구대장 파일에는 등록되어 있으나, 현 지실사에서는 거주자로 확인되지 않은 사람들을 전화면접에 의하여 거주지 정보를 확인하는 절차이다. 이 절차는 보완인구대장 파일의 과다집계(overcoverage)를 점검하기 위하여 이용한다.
인구 과소집계	현지실사 자료를 접수하여, 응답자가 면접을 한 주소지와 보완인구 대장 파일의 주소를 매칭작업을 실시한다. 매칭결과, 보완인구대장파 일에 등록된 주소지에 기재된 인구와 현지실사에서 면접된 인구를 비교하여 모집단의 과소집계 정도를 파악한다.
인구 과다집계	보완인구대장 파일에 의거하여 인구표본에 포함되었으나 현지실사 에 면접대상이 되지 않은 모든 사람들은 그들이 실제로 거주하는 장 소가 어디인가를 확인하기 위하여 전화조사의 대상이 된다. 전화조 사의 결과, 보완행정파일에 기재된 주소지에서의 인구 과다집계에 대한 정보가 생산된다.
조사구역	이스라엘은 3000여개의 통계지역으로 구분된다. 개별 통계지역은 대 략 20개의 조사구역으로 구분되어, 인구대장 자료를 근거로 하여 개 별 조사구역에 약 50개의 가구가 포함되도록 하였다. 이들 조사구역 을 통상적으로 센서스 표본조사 조사구역으로 부른다. 모든 통계지역에서, 조사구역의 20% 정도가 표본으로 추출되고, 이 조사구역에서, 모든 주택과 가구를 전수하도록 되어 있다. 환언하 면, 표본으로 선정된 조사구역에 대해서만, 조사원이 방문하여, 그 조사구역에 있는 주택 또는 단위거처에 거주하는 인구에 대하여 조 사표를 완성한다..

<p>센서스 가중치</p>	<p>센서스 표본조사의 결과와 보완인구대장파일 자료의 매칭결과를 바탕으로, 보완인구대장 파일의 모든 조사항목은 통계처리과정에서 가중치에 해당하는 수치를 할당받는다. 이 수치는 통상 센서스 가중치라고 불린다. 이 가중치는 현지실사에서 수집된 자료와 보완인구대장파일 자료 간에 매칭결과를 토대로 하여 계산된다. 이 가중치는 모집단의 일부로서, 개인의 특성을 대표하는 수치이다.</p> <p>특정 통계지역과 연령집단에서 인구의 과소집계가 있는 경우, 곧 보완인구대장 자료에 등록된 모든 개인이 현지실사에서 집계되지 않은 경우에, 해당 개인은 1.0 보다 큰 센서스 가중치를 할당받게 된다.</p> <p>한편, 특정 통계지역과 연령집단에서 인구의 과다집계가 있는 경우 곧 보완인구대장 파일에 기재되어 있지 않은 사람이 현지실사에서 집계되는 경우에는 센서스 가중치가 0과 1의 사이에 있는 수치를 할당받게 된다.</p> <p>통계처리과정에서, 센서스 가중치는 해당 기록이 모집단에서 대표하고자 하는 사람의 수와 일치하도록 곱셈을 하는 계수를 가리킨다. 이 계수는 참값에 근접하는 자료를 획득하는데 필요한 수치이다.</p>
<p>통합센서스 파일</p>	<p>통합센서스 파일은 보완인구대장에 등록된 사람들로 이루어진다. 이 파일은 조사객체인 개인의 인구학적 특성과 센서스 가중치로 이루어지는데, 후자는 전체 모집단에서 해당 개인이 대표하는 실제의 사람 수를 의미한다.</p> <p>가령, 센서스 가중치가 2.0인 개인의 경우에, 그는 보완인구대장 주소지에서 2사람을 대표하는데, 그 중 해당 주소지에 거주하는 한 사람은 보완인구대장에 누락되어 있고 아마도 다른 주소지에 등록되어 있을 것이다. 한편, 센서스 가중치가 0.5가 되는 개인의 경우에는 해당 주소에 기재된 사람이 한 사람이 아니라 두 사람이라는 것을 말하며, 그 들 중에 한 사람은 실제로 다른 주소에 거주하고 있어야 마땅하다고 할 수 있을 것이다.</p>
<p>사회경제 파일</p>	<p>사회경제 파일은 현지실사에서 집계된 개인과 가구로 구성되며, 이들 개인과 가구의 인구학적 특성과 사회경제적 특성을 포함한다. 모든 개별기록은 그에 할당된 센서스 가중치가 있으며, 그것은 전체 모집단에서 그것이 대표하는 사람의 수를 가리킨다. 사회경제 파일의 가중치는 통합센서스 파일에 맞도록 하여, 가령, 우리가 이스라엘의 행정수도인 텔아비브의 모든 가중치를 합하게 되면, 통합센서스 파일과 사회경제 파일에서 그 수치가 동일하게 될 것이다.</p>

이스라엘 통합센서스:
인구대장 범위오차 추정 목적의 표본조사
실시를 중심으로

한국인구학회
전광희 교수
jkh96@cnu.ac.kr

머리말

◆ 통합센서스의 기본 아이디어:

- 센서스 통계작성의 토대를 조사표와 현장조사 중심의 전통적 센서스 디자인에서 행정등록자료로 대체한다는 것이다.
- 행정등록자료를 보정/보강하기 위하여 대규모 표본조사의 정보를 활용한다.

머리말

◆ 통합센서스의 예상되는 장점들:

- 센서스 통계의 품질과 시의성 개선
- 응답자의 응답부담 경감
- 센서스 통계 작성의 빈도 증가
- 센서스 비용의 감소

3

인구대장:

◆ 주요 행정자료는 전국수준의 인구대장(PR: Population Register)이다

인구대장의 주요 정보:

- ✓ 주민번호와 성명
- ✓ 주소
- ✓ 인구학적 기본변수(연령, 성별, 출생지, 입국일자, 인종/종족, 혼인상태, 종교, 가구주와의 관계)

4

인구대장:

- ◆ 인구대장의 범위오차(coverage error):

- ✓ 지역별 과소집계(undercoverage)와 과다집계(overcoverage): 인구대장의 불명확한 주소
- ✓ 전국수준 과다집계: 인구대장에 여전히 등록된 해외민자
- ✓ 전국수준 과소집계: 주민번호 없이 이스라엘에 합법, 불법으로 거주하는 사람들

범위오차는 지역이나 인구학적 특성에 따라서 차이가 있다.

5

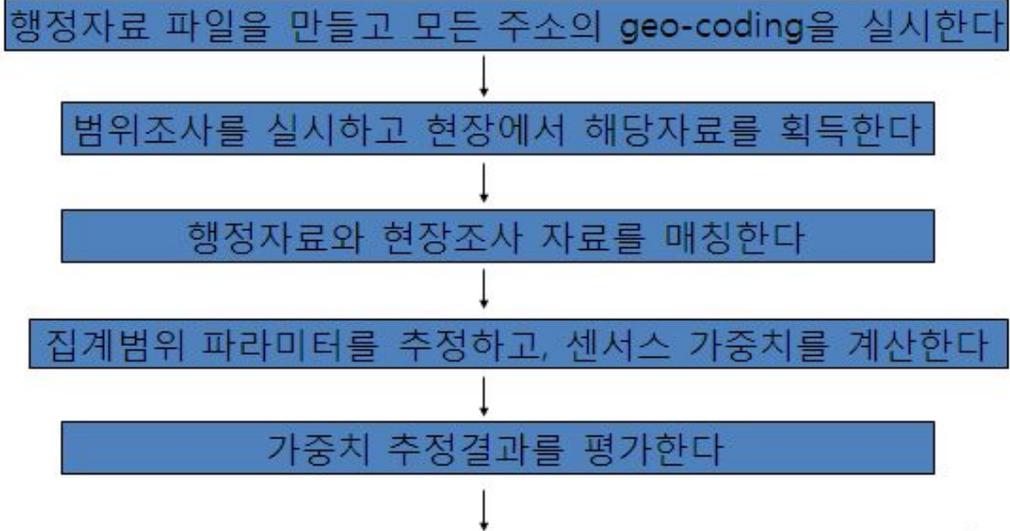
추정목표:

- ◆ 이스라엘 행정통계 시스템은 전국을 행정구역과 행정구역 내 통계지역으로 구분한다. 통계지역은 평균 4000명으로 구성되고, 1만 명 미만의 행정구역은 하나의 단일 통계지역으로 간주된다.

- ◆ 이스라엘 통합센서스의 목적은 통계지역별로 추정인구를 정확하게 작성하는 것이다.

6

통합센서스 실시과정



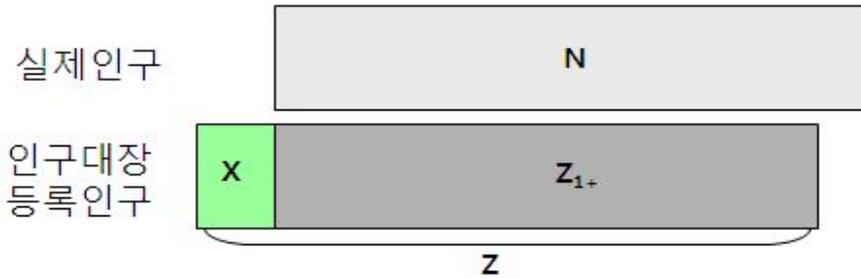
7

집계범위 모형:

- ◆ 범위오차는 통계지역별로 정의된다. 구체적으로
 - 인구대장 과소집계: 해당 지역에 거주하지만, 인구대장의 다른 주소에 등록된 사람들
 - 인구대장 과다집계: 인구대장의 해당지역 주소에 등록되어 있지만, 다른 주소(이스라엘 국내 또는 해외)에 거주하는 사람들

8

집계범위 모형



집계범위 파라미터는 다음과 같이 정의한다.

과소집계 추정공식

$$p_{1+} = EZ_{1+} / N$$

과다집계 추정공식

$$\lambda = EX / N$$

9

집계범위 모형:

과소집계 파라미터 와 과다집계 파라미터의 합은 다음과 같이 표시된다

$$p_{1+} + \lambda = EZ / N$$

그러나 문제는 과소집계 파라미터 p_{1+} 와 과다집계 파라미터 λ 는 미지수이기 때문에, 대규모 표본조사를 통하여 추정되어야 한다.

실제인구의 추정치는 다음과 같은 공식으로 계산된다.

$$\hat{N} = \frac{Z}{\hat{p}_{1+} + \hat{\lambda}}$$

10

범위파라미터 추정 표본:

- 집계범위 파라미터를 추정하기 위하여 두 개의 표본조사를 설계해야 한다.
 - ✓ 과소집계 파라미터를 추정하기 위한 지역기반 표본 선정 - U sample
 - ✓ 과다집계 파라미터를 추정하기 위한 인구대장 등록인구의 표본 - O sample

11

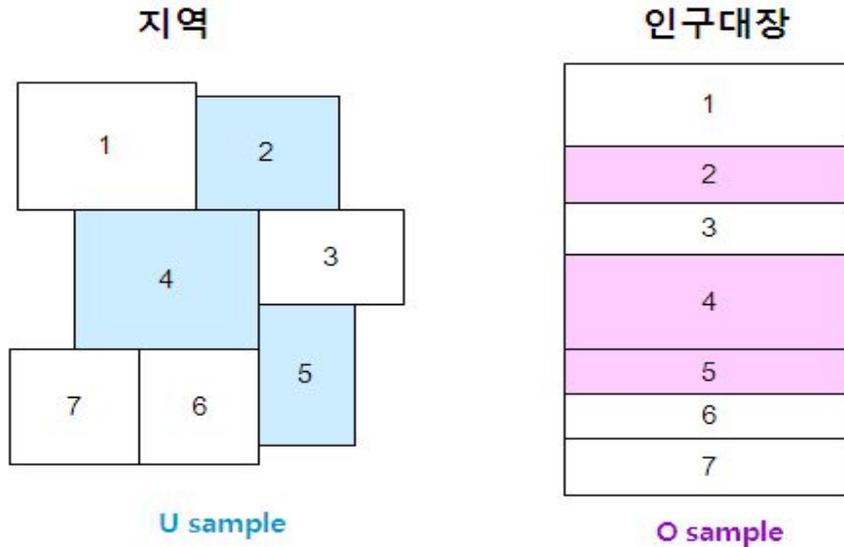
집계범위 파라미터 추정 표본:

두 종류의 표본은 다음과 같이 얻는다:

- 통계지역을 50개 가구로 이루어진 조사구(EA: Enumeration Area)로 구분한다.
- 인구대장 주소는 geo-coding에 의하여, 조사구별로 클러스터를 형성하게 한다
- 두 개의 표본을 동일한 조사구 안에서 선정한다

12

집계범위 파라미터 추정 표본



13

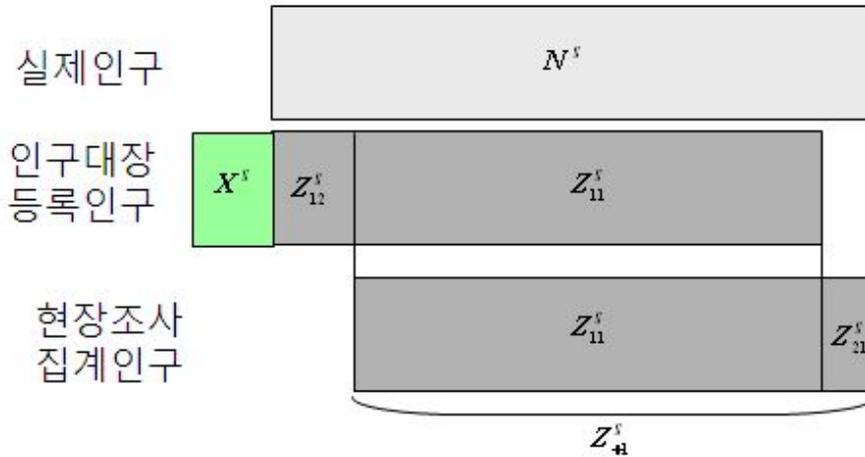
집계범위 추정 표본 - U sample

- 조사원은 표본으로 선정된 조사구를 조사하여, 모든 가구와 가구 내 모든 사람들을 집계하고자 노력하여야 한다.
- 현장조사를 통해서 생성된 리스트와 인구대장 등록 인구를 매칭한다.

매칭 (matching)은 주민번호나 연령, 성, 성명 등 인구학적 기본변수를 사용하여 실시한다.

14

범위집계 추정 표본 - U sample:



15

집계범위 추정 표본 - U sample

현장조사 집계인구

	In	Out	Total
인구대장 등록인구	Z_{11}^s	Z_{12}^s	Z_{1+}^s
	Z_{21}^s	Z_{22}^s	Z_{2+}^s
	Z_{+1}^s	Z_{+2}^s	N^s

- 과소집계 파라미터는 다음과 같은 공식에 의하여 추정된다.

$$\hat{p}_{1+} = Z_{11} / Z_{+1}$$

- 현장 면접조사와 인구대장 등록인구의 통계적 독립성이 확보되도록 면접조사가 실시되어야 한다.

16

집계범위 추정 표본 - O sample

- 과다집계 파라미터를 추정하기 위하여, 우리는 인구대장에 등록된 주소에서 집계되지 않은 사람을 찾아서, 그들이 실제로 어디에 거주하는가를 결정하여야 한다.

- ✓ 제1단계 작업으로, 조사원은 명부리스트를 가지고, 조사구로 돌아가야 한다. 본인, 친척, 이웃사람으로부터 해당 정보를 수집하여야 한다.

- ✓ 제2단계로, 성명을 CATI(Computer-Assisted Telephone Interviewing) 시스템에 이전하여, 전화면접으로 해당 정보를 수집할 수 있다.

17

실제인구의 추정

- 개별 통계지역의 실제 인구는 추정집단(estimation group)별로 구분하여 계산하며, 해당 추정집단은 범위 집계 오차가 가능한 동질적인 확률이 큰 집단으로 구분하여야 마땅하다.

- 두 개의 집계범위 파라미터 추정치는 통계지역 내 추정 집단 각각에 대하여 계산된다.

19

범위집계 추정 표본- O sample

- 지금까지 수집된 정보를 이용하여, 우리는 X^s 와 Z_{1+}^s 을 결정하고, 이것을 바탕으로 과다집계 파라미터를 추정한다

$$\hat{\lambda} = \frac{X^s}{\hat{N}^s} = \frac{X^s}{Z_{1+}^s / \hat{p}_{1+}}$$

위에서 우리는 $\lambda = EX/N$ 임을 상기할 필요가 있다.

18

추정인구:

- 센서스 가중치를 정의한다.

$$\hat{w} = \frac{1}{\hat{p}_{1+} + \hat{\lambda}}$$

- 센서스 가중치는 추정집단(estimation group)별로 인구 대장 등록인구 모두에게 하나하나 배정된다. 가중치는 실제 인구에서 등록인구가 대표하는 인구수를 반영하도록 해야 한다.
- 어떤 인구집단의 경우에도 센서스 추정치는 인구집단의 구성원에 할당된 인구대장 가중치의 합계이다.

20

**** 이스라엘의 센서스 표본조사**

이스라엘의 2008년 센서스 표본조사는 종전의 전통적 센서스 방식과 달리, 중앙통계국(Central Bureau of Statistics)이 개발한 통계적 방법에 의거하여 실시되었다. 통합센서스로 명명된 이 센서스는 응답자의 부담을 경감하고, 장래에 실시된 센서스에서 예산을 절감하는 토대를 구축하기 위한 것이다. 통합센서스는 인구대장(Population Registry)을 기본으로 하는 행정자료와 이스라엘 가구의 20% 표본을 1:1로 직접 조사하는 방식을 결합하는 것이다 (300명 미만의 소지역에 대해서는 전체 주민을 전수조사하는 방식이 사용된다. 인구대장의 과다범위와 과소범위를 추정하기 위한 현지실사와 전화면접을 센서스 표본조사라고 부른다. 이 조사는 2008년 말에 시작하여, 2009년 7월에 종료하였다.

2008년 센서스 표본조사의 시간표와 추진단계

A. 조직= 인원훈련 2,500명의 센서스 실사요원(지역사무소 책임자, 관리자, 업무조정자 및 2000명의 현지실사요원으로 구성)의 모집 및 훈련, 12개 지역사무소의 설치 및 현지실사 업무 안내, 이 단계는 2008년 8월에 시작하여 2008년 12월까지 완료되었다.

B. 표본의 건물 명부작성: 표본으로 선정된 거처와 건축물에 리스트하며, 단위거처를 직접 방문하지는 않는다. 이 단계는 2008년 7-18일의 2주간에 걸쳐 완료되었다.

C. 면접조사: 조사원은 사진이 부착된 이름표를 달고, 가구별 면접을 위하여 전 단계에서 리스트된 모든 거처를 방문한다. 인터뷰는 노트북컴퓨터를 이용하여 진행한다. 면접은 가구주가 원칙이나, 책임있는 답을 줄 수 있는 가구원이면 무방하며, 18세이상의 성인이 응답자가 되어야 한다. 인터뷰는 2008년 12월 27일 토요일 저녁의 가족상태를 기준으로 하며, 인터뷰 기간은 20-30분으로 하였다. 이 단계는 2009년 12월 28일에서 2009년 2월 말까지 진행하였다.

D. 전화조사: 표본조사의 대상이지만, 전 단계의 면접조사에서 면접이 불가능했던 사람들을 전화로 면접하는 것으로, 그들의 정확한 주소지를 확인하는

것이 목적이다. 이 단계는 2009년 3월 25일부터 2009년 7월 25일까지 진행되었다.

국민의 협조의무

2008년 센서스 표본조사는 1972년 통계법에 의거하여 실시되었다. 해당법은 표본조사의 대상자는 센서스 조사원의 조사요구에 정확한 응답을 해야 하는 책임사항을 규정하고 있다. 이 목적을 위하여, 2008년 인구센서스를 실시하기 위한 통계시행령이 국무총리의 서명으로 관보에 게시되었다.

비밀유지의 의무

주민의 비밀은 엄격히 지켜진다. 수집된 자료는 통계적 수량정보로서만 출판되며, 질문지에 개인의 신상을 파악할 개연성이 없도록 통계처리를 해야 한다. 통계국이 조사객체를 확인할 수 있는 어떠한 정보도, 정부, 공공기관, 상업기관, 민간기관에 전달하는 것은 불법이다, 통계법은 중앙통계국의 모든 직원들이 자료의 비밀을 유지하는 것을 의무화하고 있다.

지역사무소

센서스 실시시점에서, 통계국은 전국에 현지실사를 위하여 12개의 지역사무소를 설치하도록 하였다.

인원모집

센서스 실시시점에, 센서스 본부는 2500명의 직원을 다양한 지위에 고용하였다. 직원의 모집은 점진적으로 이루어져서, 관리자에서 거주자의 가정을 방문하여 직접 집계를 담당하는 조사원에 이르기까지 하향적으로 진행되었다.

센서스 자료의 대외 공표

2008년 센서스의 결과는 2010년 경에 DVD나 간행물의 형태로 또는 통계국의 웹사이트에 이스라엘 국민이 볼 수 있도록 할 예정이다.

E. 싱가포르의 등록센서스³⁾

1. 싱가포르의 2000년 이전의 센서스

(1) 식민지시대 센서스

싱가포르의 제1차 센서스는 말라카해협 식민지 센서스(Straits Settlement Census)의 일부로서 1871년 4월에 실시되었다. 그 후, 1931년까지 10년 주기로 센서스가 정례적으로 실시되었다. 제2차 세계대전의 발발로 차기 센서스가 1947년과 1957년으로까지 연기되었다. 싱가포르의 독립 후 최초의 인구센서스는 유엔의 권고안에 따라서 1970년에 실시되었다. 2000년 등록센서스로 이행하기 전에, 두 차례의 센서스가 1980년과 1990년에 실시되었다.

(2) 1970년, 1980년, 1990년 센서스

1980년 이전의 센서스는 조사표와 현지실사(field operation)로 요약되는 전통적 방식으로 실시되었다. 제1단계로 누락을 방지하기 위하여 일반주택과 임대용 단위거처를 하나씩 하나씩 집계하였다. 제2단계는 현지실사 조사원이 가구를 방문하여 조사항목별로 응답내용을 조사표에 기재하였다. 제3단계는 센서스 기준시점이 약간 지나서 조사표에 기재된 응답내용의 타당성을 확인하기 위한 절차가 진행되었다. 현지실사를 통하여 획득한 방대한 정보를 코딩, 자료입력, 확인절차 등을 거쳐서 통계표를 생성하는 절차를 수행하였다.

싱가포르 통계국이 센서스 현지실사를 위하여, 1970년에는 2,015명의 조사원을 고용하였고, 1980년에는 센서스 업무와 자료처리를 위하여 국방부의 현역군인 500명을 포함하여 2,200명의 조사원을 동원하였다.

1980년 센서스에서는 싱가포르 국민과 영주권자의 개인식별번호(UIN, Unique Identification Number)와 외국인 거주자의 외국인식별번호(FIN, Foreign Identification Number)을 정부 행정자료 데이터베이스의 개인정보 매칭작업에 활용하였다. 졸업자의 경우 최종학력과 재학자의 경우, 재학 중인 교육기관 등에 대한 정보를 해당 개인의 인구학적 기본특성과 매칭하여, 센서스 실시를 위한 사전(事前) 데이터베이스를 구축하였다. 가능한 한, 센서

3) 이 부분의 싱가포르의 등록센서스에 관한 설명들은 Singapore Department of Statistics (2002)의 2000년 인구센서스 행정보고서를 토대로 작성되어 되었다.

스 조사표는 사전 데이터베이스의 자료를 이용하여 일부 항목에 대하여 사전 인쇄를 하였으며, 이것을 응답자들이 조사시점에서 확인이 가능하도록 하였다.

싱가포르의 1990년 센서스에서 현지실사는 자료를 획득하는 주요 방법이었다. 3,500명의 센서스 조사원이 고용되었으며, 1980년 센서스와 마찬가지로, 수집된 자료는 코딩, 자료입력, 확인절차 등을 거쳐서 통계표를 생성하는 절차를 받게 되었다.

2. 2000년의 등록센서스

(1) 2000년 센서스의 기획준비

1990년 센서스 이후에, 싱가포르 통계국은 2000년 센서스의 방법과 절차를 개선하기 위하여 인구통계를 작성하는 전반적 틀을 검토하였다. 당시, 싱가포르 통계국은 인구통계의 작성에 영향을 주는 3가지 주요한 외부여건들에 대하여 주목하였다. 먼저, 인구 자료에 광범위한 수요가 증대하고 있다. 둘째, 인터넷의 활용도 증가, 데이터베이스 기술, 콜센터 통합기술 등의 정보기술(IT) 혁명은 센서스 자료의 수집과 캡처 방법에 새로운 가능성을 제시한다. 셋째는 1980년대 이후 싱가포르 통계국에서 개발된 공공 데이터베이스는 행정자료가 통계작성 목적으로 대부분 매칭될 수 있고, 업데이트될 수 있다는 것을 확인하였다.

(2) 2000년 센서스의 실시

싱가포르는 위의 3가지 외부여건을 염두에 두고, 1996년 3월에 가구등록데이터베이스(HRD, Household Registration Database)의 개발에 착수하여 센서스 간 인구의 변동에 대한 인구학적 정보를 업데이트하고자 노력하였다. 데이터베이스 구축의 최종목표는 2000년도에 행정자료를 토대로 하는 센서스를 실시하는 것이었다.

가구등록데이터베이스는 싱가포르 내 거주자 전원의 개인 및 인구관련 기초자료를 파악하고, 이들을 업데이트하기 위하여 개인식별번호를 매칭키로 하는 매칭작업을 통하여 기존 공공 데이터베이스와 매칭한다. 1990년 센서스의 경험과 IT 기술의 발전으로, 가구등록데이터베이스는 계획대로 구축되었으며, 1999년 후반에는 완전히 기능을 다하였다.

싱가포르는 단위거처의 특성을 조사하기 위하여 주택센서스를 인구센서스와 동시에 실시한다. 1980년 이후, 싱가포르 통계국은 1980년 센서스의 부산물이기는 하지만 단위거처에 대한 최신 데이터베이스를 관리하고 있다. 1996년에는 이 데이터베이스를 업그레이드하여, 그 명칭을 전국주택데이터베이스(NDD, National Database on Dwellings)로 변경하였다. 결국, 전국주택데이터베이스와 가구등록데이터베이스는 싱가포르에 있는 모든 가구에 대하여 실체적 위치, 바로 주소정보를 제공하는 역할을 한다.

센서스 자료의 가치는 집계범위의 포괄성과 기준시점의 동시성 충족여부에 있다고 할 것이다. 전통적 센서스의 전수조사를 위한 기본항목(예: 성별, 연령, 인종/종족, 국적, 주택유형)은 가구등록데이터베이스와 전국주택데이터베이스에서 동일한 기준시점에서 추출가능하기 때문에, 등록센서스를 실시하는데 아무런 문제가 없었다. 심층연구를 위한 추가적인 자료는 동일한 기준시점에서 모집단의 일부인 대규모 표본을 선정하여 이들에 대한 통계조사를 실시하여 획득하기로 하였다. 싱가포르 통계청은 종전의 센서스와 표본조사의 경험을 통하여, 20% 가구표본조사라면, 심층연구에 필요한 세부요건을 충족하고, 이용자 다수의 요구를 반영하는데 무리가 없을 것이라고 판단하였다.

인구센서스를 위하여, 공공 데이터베이스를 매칭하고 통합하는 과정에서, 센서스가 개인과 가구의 정보를 하나의 데이터베이스로 통합하는 것이라는 것을 의식하였으며, 따라서 통합된 개인자료는 엄격히 비밀이며, 센서스법 제35장의 규정에 의하여 보호되도록 하였다. 싱가포르 통계법은 싱가포르 통계국이 센서스의 조사객체인 개인, 가구에 대한 정보를 어떤 외부기관에 발설하지 못하도록 규정하고 있으며, 데이터베이스에서 공표 가능한 것은 통계 목적이나 분석목적으로 집계된 통계, 곧 마이크로데이터만으로 되어 있다.

(3) 자료수집전략

싱가포르 센서스는 20% 표본 집계에 대해서 인터넷조사, CATI (Computer-Assisted Telephone Interviewing), 현지실사 등의 3가지 방법으로 구성된, 싱가포르 통계국의 “3중 자료수집전략”(Tri-modal data collection strategy)을 구사함으로써 IT혁명이 제공하는 가능성을 100% 활용하도록 노력하였다.

① 인터넷조사

싱가포르는 2000년 센서스에서 인터넷조사로 가구관련 센서스 정보를 획득하는 국가들의 선두그룹에 서게 되었다, 당초에, 인터넷조사를 실시하는데 대하여 몇 가지 문제점과 우려사항이 제시되기도 하였다. 2000년 센서스 인터넷웹사이트를 통하여 자계식(self-enumeration) 조사에 응하는 것은 응답자에게 부담이 되는 것으로 인식되었다. 응답률을 일정 수준으로 올리기 위하여, 온라인 센서스 조사표 디자인, 인센티브, 홍보전략 등이 인터넷조사의 성공여부를 결정하는 요인이 될 것으로 판단하였다. 더욱이, 가구데이터베이스의 표본은 싱가포르 국민들이 온라인 접속을 할 수 있도록 개방되어 있기 때문에, 보안유지와 비밀보호 문제가 부당접속, 해킹, 서비스거부 등의 공격을 방지할 수 있도록 설계하지 않으면 안 되었다.

싱가포르 통계국은 인터넷조사의 장점은 잠재적 위험요인을 극소화할 수 있다면, 그 위험요인의 문제점을 능가하는 것으로 평가하였다. 응답자들은 응답내용을 조사원에게 직접 말하지 않고, 통계국 데이터베이스로 직접 전달되는 인터넷 조사의 비밀유지 조치에 만족스러워 했다. 더욱이, 인터넷을 통해 조사표를 작성하는 경험은 긍정적이고 대화를 촉진하는 것으로 파악되었다. 응답자가 센서스 웹사이트에 접속하였을 때, 사전(事前) 데이터베이스에서 추출한 기본 자료가 컴퓨터 화면에 제시되도록 되었다. User-friendly한 도우미 특성이나 설명노트도 필요한 경우에 즉시 제공될 수 있도록 하였다. 시스템은 간단한 on-line 점검을 하고, 일관성이 없거나 명백히 오류인 응답을 하는 경우 지료를 재입력하도록 하였다. 편의성을 증대시키기 위하여, 부분적으로 완성된 조사표는 저장하였다가 나중에 계속응답을 위하여 접속할 수 있도록 하였다.

② CATI

CATI, 곧 컴퓨터를 이용한 전화조사는 인터넷조사를 시작한 후에 적정시간이 경과한 후 시작하도록 하였다. 인터넷 조사와 달리, CATI는 1995년, 4만 개의 가구에 대해서 실제로 일반가계조사(General Household Survey)를 실시하는데 사용된 적이 있었던 방법으로서, 2000년 센서스에서도 아무런 문제가 없었던 것으로 확인되고 있다.

선정된 가구는 우편구역(PD, Postal Districts)에 따라 균등하게 배분되었으며, 우편구역 하나하나에 대하여, 인터넷으로 그들의 조사표를 반송하지 않은 가구들은 자동적으로 사전예약을 통하여 CATI 면접을 하도록 하였다.

③ 현지실사

몇 차례 전화접촉을 시도한 후에 CATI 방법론으로 해당가구가 접촉이 불가능한 것으로 판명되면, 그 가구는 현지실사의 대상이 되도록 하였다. 현지실사 가구는 우편구역별로 정리되었으며, 지역 센서스 사무소로 전달되었다.

현지 조사원은 1 대 1 면접을 위하여 인터넷조사나 CATI 인터뷰에 불응한 응답자들을 면접하기 위하여 현지실사를 진행하였다. 현지 조사원이 해당가구와 접촉하는데 실패하는 경우, 이들 가구에 연락처를 남겨두어서 면접을 완결하기 위하여 편리한 시간에 전화를 하도록 요청하였다.

④ 정보의 비밀 및 보안조치

3가지 자료수집 전략은 인터넷조사, CATI와 같이, 응답자 개인과 1 : 1 접촉을 피하여 개인과 가구의 센서스 정보를 수집하는 절차를 포함하였다. 싱가포르 통계국은 데이터 비밀 및 보안문제가 그 자체로서 대단히 중요한 것이라고 센서스 업무의 정상적 추진과정에 대단히 중요한 것이라고 판단하였다.

정보의 비밀을 확실히 하기 위하여, 센서스 표본조사에 선정된 모든 가구에 주택식별번호(ID)와 무작위로 생성된 비밀번호가 포함된 통지문을 보내도록 되어 있다. 주택식별번호, 비밀번호, 두 사람의 개인식별번호(UIN, Unique Identification Number)를 사용하여, 응답자는 접속하여 데이터베이스 안의 자신의 가구정보를 검색할 수 있도록 되어 있다. 비밀번호 점검은 데이터보안을 확실히 하는 방식으로 실시되었다. 이 방식은 비밀번호를 응답자의 컴퓨터에 무작위 수열(random sequence)을 생성하는 키로 사용하였다. 응답자의 무작위 수열은 인터넷으로 전달되었다. 응답자의 비밀번호는 인터넷으로 전달되지 않기 때문에, 비밀번호는 중간에 가로채거나 타인이 읽을 수 없도록 되어 있었다. Server end에서, 접수된 무작위 수열은 key server와 함께 암호를 해독한다.. 해독된 암호가 매칭에 성공하면, 응답자는 인증을 받아 접근이 허용된다.

응답자가 제공하는 모든 개인정보는 인터넷으로 128-bit로 암호화된다. 이것은 불법적으로 개인정보를 가로채는 것을 방지한다. 개인정보가 해킹되는 방지하기 위하여, 컴퓨터의 두 개 층의 방화벽을 이용하는 비무장지대(DMZ, Demilitarized Zone)이 설치되어 개인정보가 저장되는 온라인 데이터베이스

를 보호한다. 이러한 보안조치는 가장 엄격한 테스트를 받도록 되어 있으며, 싱가포르 정부기관인 정보커뮤니케이션 개발당국(IDA, Infocomm Development Authority)의 보안요구사항을 준수하도록 하였다.

CATI에 대해서는, 센서스 조사원은 전화상으로 주택식별번호를 인용하여, 표본조사를 실시하기 전에 자신이 센서스 조사업무에 종사하는 사람이라는 것을 확인한다. CATI 조사원의 신분에 의심이 있을 경우에, 응답자들은 언제나 그들의 신분을 확인하기 위하여 센서스 핫라인으로 연락을 취할 수 있도록 되어 있다.

3. 싱가포르 2000년 센서스의 의의

싱가포르 통계국은 2000년 센서스가 행정자료 센서스로서, 그 결과를 20%의 대규모 통계조사로 보완한 것에 대하여, “싱가포르 센서스의 역사에서 하나의 새로운 이정표를 세웠다.”라고 평가하고 있다. 1871년 최초의 센서스가 말라카해협의 식민지 센서스로 실시된 이후, 센서스 정보가 이제는 총인구를 대상으로 하여 현지실사 방식으로 진행되지 않다는 점을 평가하고 있다. 이러한 등록센서스 접근방식은 덴마크, 핀란드, 노르웨이, 스웨덴, 네덜란드와 같은 통계 선진국에서는 새로운 관행이 아니라는 점도 인정하고 있다.

싱가포르는 유럽 이외의 지역에서는 최초로 행정자료를 이용하여 등록센서스를 실시한 최초의 국가가 되었다. 등록센서스를 추진하기 전에, 싱가포르 통계국은 3가지 핵심쟁점을 연구하였다. 하나는 싱가포르의 행정자료의 품질이 고품질로서, 국민 개개인과 그 인구학적 기본특성을 조사하는데 아무런 문제가 없다는 것을 확인하였다. 둘째는 싱가포르의 법률적 환경과 데이터 비밀보호조치가 민감하지 않은 행정정보의 공유를 허용한다는 것이었다. 마지막으로, 등록센서스 방식으로 센서스를 실시하게 되면 엄청난 조사비용이 절약될 수 있으며, 대규모 표본조사와 함께 행정자료를 이용하게 되면 조사표와 현장실사로 실시되는 전통적 센서스 비용의 1/3에 해당하는 적은비용으로 센서스를 실시할 수 있다는 것이었다.

4. 센서스 준비·기획위원회

어느 국가나 마찬가지로, 센서스는 수많은 중앙부처를 포함하는 대규모의 통계프로젝트이기 때문에, 싱가포르는 1998년 2월에 센서스 준비·기획위원회를 설치하였다. 2000년 센서스 준비·기획위원회 위원장으로, 무역·산업

부(Ministry of Trade and Industry) 장관이 취임하였다. 새로운 등록센서스 방식과 3가지 조사방식에 대하여 최종 결정이 내려지기 전에, 싱가포르 정부의 무역·산업부와 통계국 내에서 수많은 사전검토와 토론이 전개되었음은 두말할 필요도 없다. 2000년 센서스 준비·기획위원회는 등록센서스의 성공 요인을 세심하게 점검하는데 주축의 역할을 수행하였으며, 전반적인 지침과 정책방향을 제시하였다. 핵심적인 의사결정은 2000년 센서스 준비·기획위원회에서 내려졌으며, 그 중에서 특별히 중요한 것은 2000년 센서스에서 수집되어야 할 조사항목의 수를 결정하는 것이었다.

센서스 자료의 중요성을 고려하여, 2000년 센서스 준비·기획위원회는 센서스 자료를 정책평가나 기획 작업에 이용하는 여타 정부기관에서도 참가하였다. 가령, 지역개발 및 스포츠 부(Ministry of Community Development and Sports), 인력부(Ministry of Manpower Plannin), 통신·정보기술부(Ministry of Communication and Information Technology), 내무부(Ministry of Home Affairs), 정보 및 예술 부(Ministry of Information and the Arts), 교육부(Ministry of Education), 국가발전부(Ministry of National Development), 무역·산업부(Ministry of Trade and Industry) 등이 참가하였다. 통계국장과 센서스과장은 2000년 센서스 준비·기획위원회의 상임위원으로 참가하였으며, 싱가포르 통계국은 위원회의 사무국 역할을 했다.

2000년 센서스 준비·기획위원회의 외에, 2000년 센서스를 기획하고 실시 하데 필요한 여러개의 실무위원회가 있었다. 센서스 기술자문위원회는 통계국장과 싱가포르 국립대학과 정책연구소의 학계대표들로 구성되었다. 센서스 기획, 센서스 IT자원 및 지원그룹, 표본조사 기획팀, 등의 센서스 기획집단들이 조직되었는데, 여기에는 통계국 직원을 주요 구성원으로 하였다.

F. 덴마크와 핀란드의 등록센서스

지금까지 최근에 시작한 싱가포르, 독일, 이스라엘의 등록센서스와 부가적인 표본조사를 검토하였다. 물론 이들은 부분 등록센서스로서 행정자료를 통한 센서스 항목의 조사와 행정자료에 없는 조사항목을 표본조사로 획득하는 것이 그 주요 특징이다. 여기서는 100% 완전한 등록센서스를 실시하고 있는 덴마크와 핀란드의 사례를 중심으로 등록센서스 실시를 위한 사회적 여건과 행정자료로부터 통계자료를 생산하는 일련의 과정에 대해 살펴보고자 한다.

1. 등록센서스의 실시 배경과 전개 과정

가. 덴마크

덴마크 등록센서스의 추진배경은 <표 1>에 정리되어 있다. 유럽의 국가들은 19세기부터 인구대장을 수립하기 시작했지만, 덴마크는 예외적인 경우에 속했다. 당시 납세자의 거주지가 파악되지 않았기 때문에, 지방자치단체는 조세징수와 재정지출을 예측하는데 어려움을 겪었다. 제1차 세계대전 후, 효율적 행정시스템의 확립에 여러 가지 수요가 제기되었으며, 특히 인구대장이 다양한 목적, 특히 인구통계나 선거인명부의 작성에 이용될 수 있을 것이라는 인식이 생

<표 1> 덴마크의 등록센서스 추진과정

연도	주요 내용
1924년	인구대장을 지방자치단체별로 수립
1966년	덴마크 통계법 입안
1968년	인구대장의 개혁(중앙인구대장 도입, 표준코드체계의 도입)
1969년	인구등록부로부터 통계자료 생산 시도
1970년	전통적 방식의 센서스 실시(중앙인구대장과 센서스 결과의 상호 비교)
1974년	1975년 전통적 방식의 센서스 포기 결정
1976년	부분적인 등록센서스 실시(건물/주택센서스 자료 누락)
1978년	건물/주택대장 도입
1981년	등록센서스 실시

겨났다. 덴마크는 1924년 인구대장을 도입하는 법률을 제정하고, 1924년 인구센서스를 토대로 하여 지방자치단체는 인구대장을 수립할 것을 결정하였다. 중앙인구대장을 수립하는 문제가 지방자치단체의 인구대장과는 달리 집중관리(central management)되는 행정자료이기 때문에, 당시는 막대한 행정비용 때문에 받아들여지지 못했다. 인구대장은 지방정부와 다른 행정기관 사이의 협력을 통해 오랜 시간을 두고 점진적으로 발전하였다.

덴마크 통계시스템 발전의 핵심적인 요소는 1960년대에 덴마크 의회가 등록통계시스템을 긍정적 관점에서 파악하였고, 이것이 장래의 통계발전을 위하여 필요하다고 판단했다는 점이다. 덴마크 통계청의 재정적 제약은 이러한 선택을 진척시킨 부수적인 요소였다. 덴마크 통계청이 광범위한 행정자료를 이용할 수 있는 배경에는 행정당국으로부터의 광범위한 자료 수집을 규정하고 있는 덴마크 통계법이 있다. 통계법이 제정된 1966년에는 이미 중앙인구대장의 준비하고 있었는데, 덴마크 의회는 인구대장이 통계적 목적으로 사용될 수 있는 가능성을 예상하고 있었다. 이는 통계법의 목적에 분명히 천명되었는데, 여기에는 응답부담의 경감과 미래의 센서스가 조사표 없이 전적으로 행정자료를 토대로 실시될 가능성을 언급하였다. 덴마크 의회는 대단한 선견지명으로, 덴마크 통계청은 1981년에 세계 최초로 등록센서스를 실시할 수 있었다.

1924년의 인구대장(population register)과 인구대장법을 기초로 하여 1968년에는 인구대장에 대한 개혁이 이루어졌다(Borchsenius, 1996, 2000, 2006). 이러한 개혁의 결과인 중앙인구등록부(Central Population Register: CPR) 설립의 첫 번째 그리고 가장 중요한 동기는 이중등록과 이로 인한 재원의 낭비를 줄이고자 하는 것이었다. 이는 전자적 자료 처리 기술을 사용함으로써 시스템 전반의 단순화와 효율성 증대를 목표로 하였다. 개혁의 주요 내용은 지방자치단체별로 관리되던 다수의 인구대장을 단일한 시스템의 일부로 통합한 중앙인구등록대장의 설립, 개인에 대한 고정된 식별코드와 표준화된 코드 체계의 도입이다. 이로써 행정기관마다 다르게 사용되던 기존의 코드 체계가 통일된 코드 체계로 대체되었다. 이와 같은 준비과정을 거쳐, 1969년에는 중앙인구등록부로부터 통계자료를 만들어 내는 첫 번째 시도가 이루어졌다. 이는 자료의 오류뿐만 아니라 장부를 담당하는 기관과 통계청 사이의 협력관계 상의 문제점들을 발견하는 계기가 되었다. 이 결과는 1970년에 실시된 전통적인 방식의 센서스 결과와 비교하여 행정자료의 품질개선에 기여하였다.

센서스 실시 주기에 따라 1975년에 센서스가 실시되어야 했지만 요구되는

보다 다양한 정보와 시의성 있는 정보에 대한 중요성이 강조되면서 사용자들과 덴마크 통계청 사이에 논쟁이 이루어졌다. 그러나 필요한 다양한 정보들이 서로 다른 행정자료에 포함되어 있다는 것이 명확했고, 행정자료를 통해 센서스 통계를 작성함으로써 응답부담 감소와 조사표를 보내고 모아야 하는 지방자치단체의 업무 경감이 이루어 질 수 있었다. 이러한 배경에서 1974년에 덴마크 통계청은 1975년의 전통적인 센서스를 포기할 것을 결정하고, 경제부장관(Minister of Economic Affairs)은 이를 승인하였다(Borchsenius, 2000, 2006). 이러한 결정의 전제조건은 행정목적과 통계목적으로 건물·주택대장을 작성하는 것이었다. 그러나 통계목적을 위해 사용할 수 있는 건물·주택대장의 확립은 상당한 시간이 소요되는 일이어서, 1976년의 센서스는 건물·주택대장에 관한 정보를 포함하지 않는 불완전한 등록센서스를 추진하기로 하였다(Borchsenius, 2000, 2006). 물론, 이 센서스가 100% 완전한 것이 아니었지만 다양한 행정자료를 원천자료로 하여 그들을 매칭하고 통합하는데 있어서 기술적인 경험을 축적시켰으며, 행정자료를 주기적으로 점검하는 시스템을 개발하는 계기가 되었다.

덴마크의 센서스에서 새로운 방법은 처음에는 여러 가지 문제를 야기하였다. 초기 단계의 문제는 행정자료에 산재된 자료의 결함과 오류뿐만 아니라 통계 생산과 행정자료를 관리하는 당국사이의 부정확한 커뮤니케이션 때문에 생겨난 것이었다.

1977년에는 사회부조법과 관련이 있는 통계가 만들어졌다. 이 통계는 JLAFS (Joint Local Authority Finance System)의 개인정보에 기반하고 있었는데 이 개인정보는 인구통계의 일부를 이루는 가족(가구)통계와 매칭하여 작성하였다. 이는 통계작성에 개인식별번호를 이용하는 것이 적합한지에 대한 논란이 생겨났고, 덴마크 통계청과 5개의 지방자치단체 사이에 법정소송으로 비화하였으며, 덴마크 통계청은 1980년에 최고재판소에서 승소하였다.

1977년에는 건물·주택에 대한 정보가 수집하여, 중앙인구대장과 같은 형태로 작성함으로써 해서, 1978년에 건물주택대장이 체계화되었다. 건물·주택대장은 1972년에 만들어진 전산화된 국가부동산대장(Property Register)의 확대판이었다. 이 행정자료를 수립하면서 1981년에는 세계 최초의 등록센서스 실시가 가능하였다.

나. 핀란드

핀란드에서 완전한 등록센서스가 실시된 것은 1990년으로, 세계에서 두 번

제로 등록센서스가 실시되기까지 20년이 소요되었다<표 2>. 1969년에 도입된 인구등록센터의 중앙인구대장⁴⁾에는 법적 거주자는 모두 개인식별번호에 의거하여 등록되었다. 1970년 센서스에는 중앙인구대장과 조세대장과 같은 개인자료가 이용되기 시작하였으며, 1970년대와 1980년대에 걸쳐 행정자료의 활용은 계속 확대되었다(<표 2>를 참고할 것)

<표 2> 핀란드의 등록센서스 도입 과정

연도	주요 내용
1964년	개인식별번호의 이용 사용
1969년	중앙인구대장의 도입
1970년	인구센서스에 인구대장 이용 시작
1980년	건물주택대장 기초자료 획득
1990년	완전한 등록센서스를 세계에서 2번째로 실시
1995 · 2000년	등록센서스 실시

자료: Ruotsalainen, 2004 재구성.

1970년 센서스 자료는 중앙인구등록부에서 나온 종교와 국적 자료, 조세등록부에서 나온 소득 자료와 매칭하였다. 또한 통계청은 학력 및 학위대장을 작성하였는데 이것은 1970년 센서스에서 수집된 학력에 대한 자료를 토대로 한 것이었다. 1970년 이후 이 대장은 교육기관에 의해 제공되는 정보에 기초하여 매년 수정되어 왔다. 인구등록센터와 협력하는 통계청은 건물의 좌표를 수집하고 이들 자료를 중앙인구대장에 입력하였다. 이들 자료들은 정기적으로 업데이트되어 왔고, 지역통계개발이나 지리정보시스템의 구축에 필요한 토대자료를 제공하였다.⁵⁾

1975년 간이센서스는 주택과 경제활동을 주제영역으로 하는 통계조사였는데 이것은 전통적 센서스보다 조사항목의 수가 적었다. 중앙인구대장에서 연

4) 중앙 인구등록대장은 나중에 인구정보시스템(Population Information System)으로 명칭이 변경되었다.

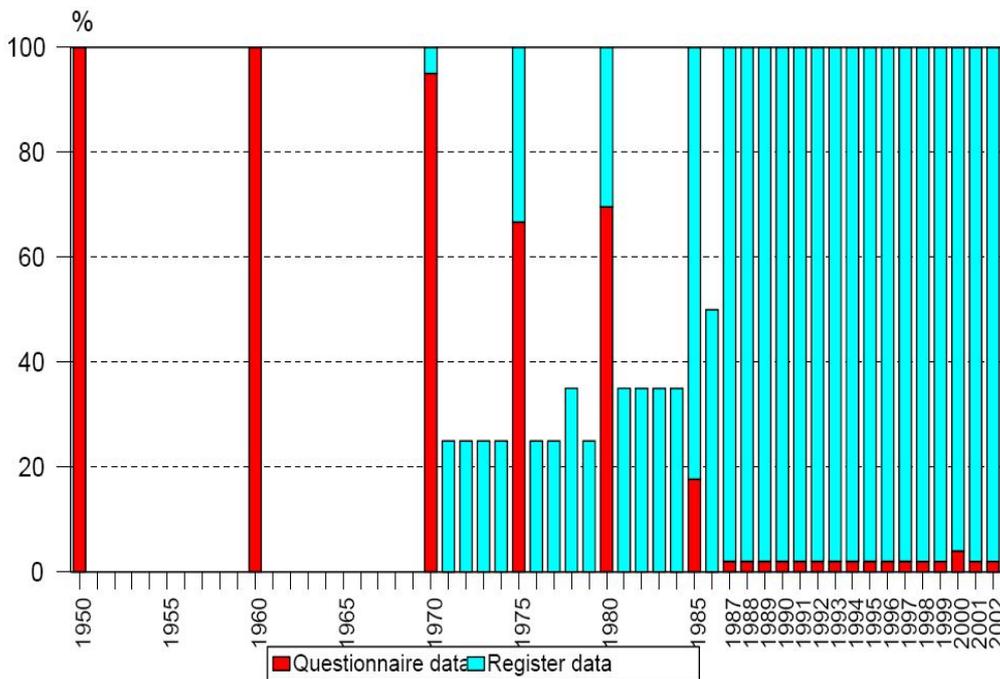
5) 인구대장을 중심으로 하는 각종 데이터베이스의 통합은 개별 자료의 절대좌표 값을 제공할 수 있는 지리정보시스템(GIS: Geographic Information System)에 의하여 효율성이 강화되었는데(김형석 외, 2003), 모든 북유럽 국가에서는 지리정보시스템의 발전과 함께 포인트를 바탕으로 하는 통계시스템을 개발하였다(UNECE, 2007). 이는 최첨단 매칭 시스템을 이용하여 개별 통계단위의 정확한 위치를 할당할 수 있도록 모든 건물의 위치가 지도 좌표를 사용해서 특정화 되어 있다는 것을 의미한다. 건물 기반의 코드체계는 좌표성정이 포함될 때, 신뢰도가 크고, 탄력적인 통계지역을 만들어 내기 위한 강력한 기반을 제공한다. 이와 같이 건물별 좌표체계의 도입이 탄력적인 통계지역의 설정을 가능하게 하는 이유는, 대부분의 GIS 패키지들에 폴리곤(통계 지역) 속의 포인트(건물)에 관한 폴리곤-포인트(point-in polygon) 분석 기능이 포함되어 있기 때문이다.

령, 성별, 혼인상태, 모국어 및 국적에 대한 자료를 획득하고, 교육관련 자료는 핀란드 통계청이 새롭게 도입한 학력 및 학위대장을 이용하였다.

1980년 인구센서스는 전통적 방식이었지만 행정자료를 더욱 더 많이 이용하였다. 1970년대 후반에 보다 정확한 통계 작성을 위한 행정자료의 필요성에 대한 인식이 확산되면서 건물·주택대장이 당시의 인구대장 법률에 의거하여, 1980년 인구주택센서스와 매칭되어 만들어졌다. 건물·주택대장은 인구등록센터에 의해 도입되었으며, 지방건물감독관(local building inspector)과 인구등록센터가 이 자료를 업데이트하여 관리하고 있다. 이후의 인구센서스에서 이 건축·주택대장은 단위거처에 대한 정보의 원천이 되었으며, 자료의 신뢰성이 개선되면서 1980년대 중반부터 건물·주택 주제통계를 매년 작성하기 시작하였다(Statistics Finland, 2004).

1990년 센서스는 전통방식이 아니라, 행정자료를 이용하여 실시되었다. 30

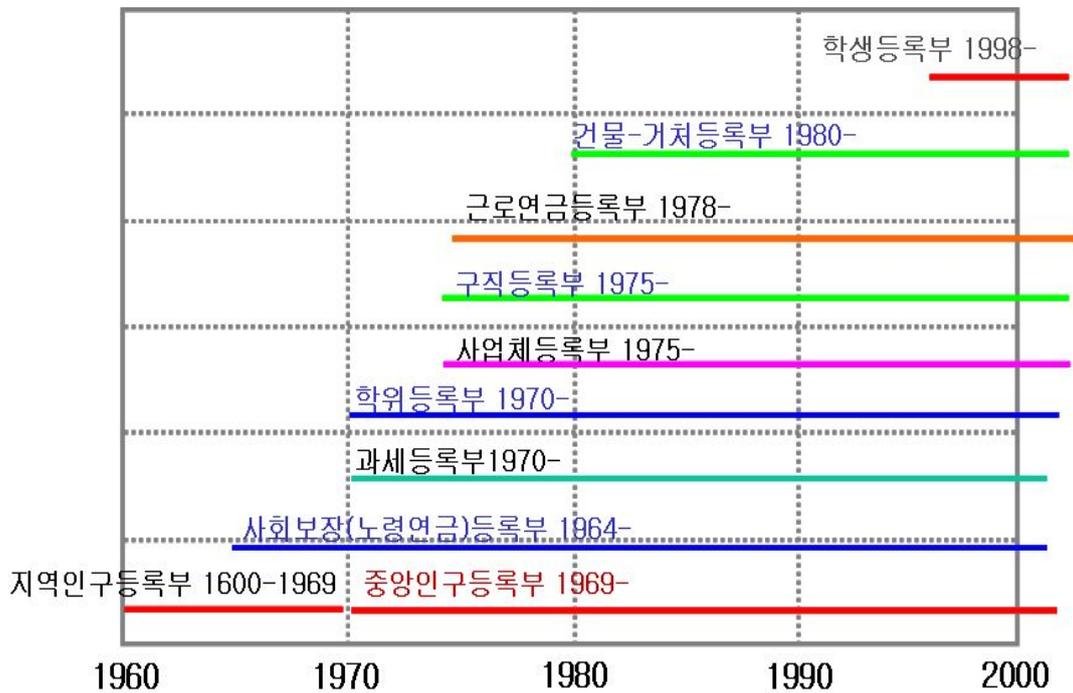
<그림 1> 핀란드 센서스와 인구·고용통계에서의 행정자료 활용비율의 변화



개의 행정자료를 이용하였는데, 대부분의 등록대장은 경제활동 부문의 센서스 통계를 추출하기 위하여 사용하였다. 주요 개별 행정대장 도입시기와 발전 과정은 <그림 4-6>에 정리되어 있다. 이처럼 1990년에 100% 완전한 등록센서스를 실시하게 됨으로써, 센서스 관련 신뢰성 연구(reliability study)를 위한 조사표를 제외하면 조사표를 우편배포하지는 경우는 없었다. 신뢰성 연

구는 행정자료와 실지조사를 비교하기 위한 독자적인 자료획득을 목적으로 수행되었는데, 등록부를 토대로 하는 행정자료가 이전 조사표를 이용한 센서스 자료의 품질만큼 우수하다는 것을 확인하였다. 등록센서스의 비용은 이전 센서스 비용의 1/10이하로 축소되었다. 동일한 자료 생산 방식이 1995년과 2000년 인구센서스에서 마찬가지로 사용되었다.

<그림 2> 핀란드의 주요 행정자료의 도입 시점



자료: Ruotsalainen, 2004.

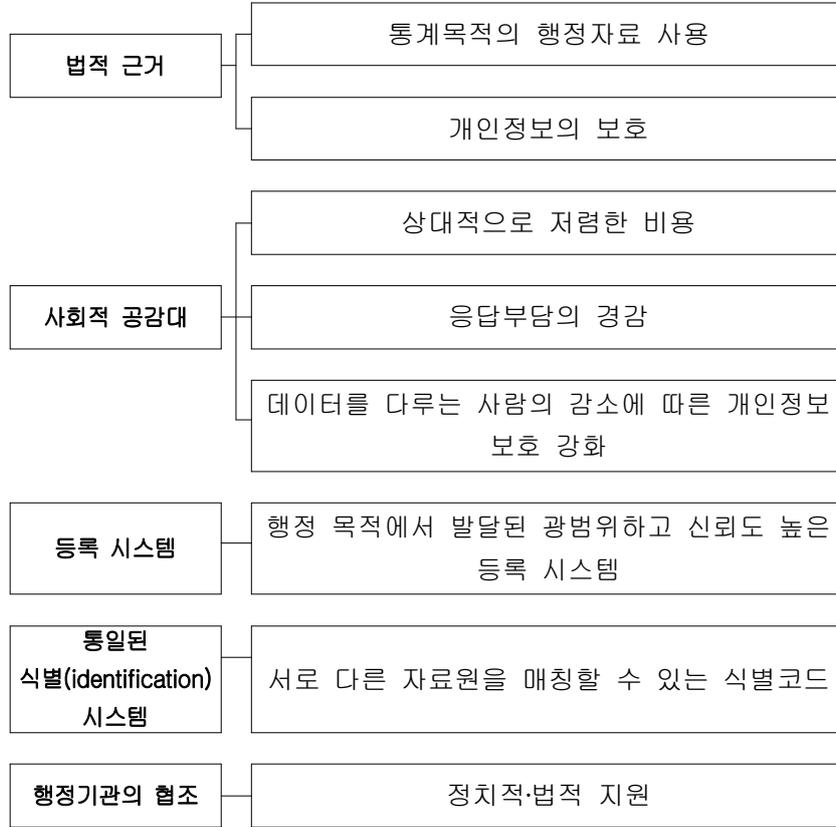
2. 완전 등록센서스 실시를 위한 사회적 여건

북유럽 국가에서 공통적으로 발견되는 등록센서스 실시를 위한 사회적 여건을 살펴보면 다음과 같다(<그림 3>을 참고 할 것). 덴마크와 핀란드에서 센서스에서의 행정자료 활용이 다른 나라보다 앞선 것은 등록센서스 실시를 위한 사회적 여건이 양호하다는 사실과 관련되는 것으로 판단된다.

가. 법률적 토대와 개인정보의 비밀보호

행정자료를 통계목적으로 이용하는데 법적 기반이 마련되어야 한다. 등록기

**<그림 3> 북유럽 국가의 등록센서스 실시 관련 사회적
여건**



반조사와 개인정보 보호를 위한 법적 근거가 필수적이다. 특히 다른 행정기관이 보유하고 있는 자료이용이 필요한 등록센서스 수행을 위해서 이러한 법률적 기반은 더욱 더 필요한 일이다. 현재 등록센서스를 실시하고 있거나 실시를 준비 중인 국가들의 경우를 보면 통계목적의 행정자료의 이용과 개인정보 보호에 관한 사항들을 법률로 규정해 놓고 있다. 법률은 통계 목적으로 자료를 수집하는 것보다 기존의 행정자료를 사용하는 것이 더 바람직하다는 견해를 반영하고 있어야 한다.

모든 북유럽 국가의 통계법은 통계작성기관에 통계 목적으로 개별 단위에 대한 식별이 가능한 수준의 데이터에 대한 접근과 다른 행정자료를 매칭시킬 수 있는 권한을 부여한다. 핀란드나 네덜란드 등에서는 다른 목적에서 수집된 데이터를 통계생산 기관이 사용해야 하는 의무가 있는데, 이 경우 국가 통계작성기관은 독자적으로 자료를 수집하기 전에 우선적으로 관련 행정자료가 존재하는지를 검토하여야 한다.

모든 북유럽 국가들은 개인정보의 보호에 대한 법률을 가지고 있다. 이 법

률의 목적은 개인정보를 포함하고 있는 전산자료의 이용과 제도가 개인정보의 보호에 관한 시민 개개인의 권리를 침해하지 않는 방식으로 이루어지는 것을 보장하는 것이다. 이 법률에 의하면 통계목적의 데이터 처리는 데이터 수집의 주된 목적이 통계 목적이 아니었다라도 허용된다. 핀란드의 개인정보 보호에 관한 법률에는 행정 목적을 위해 수집된 자료는 학문적 연구와 공식 통계 작성을 위한 목적으로 제 3자들에게 전달될 수 있다고 규정되어 있다. 이 원칙은 통계목적의 행정자료 활용이 개인의 사생활이나 권리를 위협하는 것으로 간주되지 않는다는 점에 근거하여, 통계 목적으로 행정자료를 매칭하도록 하는 EU의 자료보호지침에도 포함되어 있다(Statistics Finland, 2004). 하지만 일단 통계작성기관에 의해 자료가 처리되면 이 자료는 통계 생산과 연구 목적이외에는 사용될 수 없다('일방향의 원칙(one-way traffic)') (UNECE, 2007).

북유럽 국가들에서 행정 활동에 관한 시민들의 알 권리를 포함하여, 행정의 공개 및 투명성은 민주주의의 기본 원칙이다. 그러나 이 정보가 통계 작성을 위해 통계당국에게 전달되자마자 공공의 접근 가능성의 원칙은 더 이상 적용되지 않는다. 비밀보호는 통계청의 절대적인 지상과제이며 이는 시민과 기업으로부터 믿을 만한 정보를 얻는 데 필수적이다.

정보의 흐름은 단지 한 방향으로 즉 행정당국으로부터 통계청으로만 흐르고 반대로는 아니다. 행정당국은 통계 목적을 위해 수집하고 매칭된 정보로부터 이익을 얻을 수 있지만 통계청은 행정시스템으로 그 결과를 결코 다시 보내지 않는다. 통계법은 개인을 식별 할 수 있는 자료 없이 통계가 생산될 수 있으면 이러한 자료는 수집되지 말아야 한다는 점을 명시하고 있다. 다른 데이터 셀과 매칭이 필요한 경우는 개인 식별코드를 가진 자료를 수집할 이유가 된다고 간주된다.

대규모 행정자료의 매칭 및 대량의 데이터베이스는 개인정보 보호의 위험요소를 포함하지만, 한편 등록통계는 개인정보보호의 관점에서 장점도 있다. 등록통계에서 데이터베이스에 접근하는 사람의 수는 전통적 센서스보다 훨씬 적다. 핀란드의 경우 1980년 인구센서스의 조사표는 350개 지역 사무실에 몇 달 동안 보관되어 있었다. 약 2,200명의 임시 고용인들이 이들 조사표를 가공하였으며, 조사표의 정보는 개인의 사생활 정보를 침해할 수 있었다. 이들 자료는 부호화하여 통계청에 전달되지만, 컴퓨터이외에 어느 누구도 볼 수 없는 행정자료처럼 개인정보의 보호수준을 강화하는 것이 쉽지 않다. 핀란드에서는 개인정보를 보호하기 위하여 행정자료만이 아니라 센서스 파일에 포함되어 있는 모든 개인자료를 부호화된 파일 형태로 보관한다.

나. 행정자료 이용에 대한 사회적 분위기 형성

통계작성을 위한 행정자료 사용의 이점을 인정하고 이해하는, 사회적 공감대가 존재한다는 것은 매우 중요하다. 상황은 국가들마다 다른데, 1970년대에 노르웨이에서는 행정자료의 통계적 이용에 관한 큰 논쟁이 있었으며 스웨덴에서는 1970년대 이래 관련 논쟁이 다소 지속적으로 진행되고 있다. 반면 핀란드와 덴마크의 국민들은 등록 자료의 통계적 이용이 합리적이라고 판단하며 국가 통계기관에 강한 신뢰를 가지고 있다. 등록 자료 사용의 합리성과 장점을 설명하는데 있어 공개된 토론과 논쟁은 이러한 공감대 형성을 위한 중요한 요소이다. 한편 국가의 행정자료 관련 통계법이 시대를 선도하고, 행정자료를 취급하는 기관의 개방성과 투명성도 중요한 변수이다.

행정자료의 통계적 이용은 일반적으로 서로 다른 행정자료에서 나온 데이터를 매칭하는 것과 관련되며, 이는 국가 통계기관이 모든 개인에 대해 ‘모든 것’을 알고 있다는 빅브라더 신드롬(big brother syndrome)을 야기할 수 있다(UNECE, 2007). 반면 사람들은 통계작성기관이 통계 목적으로 수집하는 것과 같은 데이터를 행정당국이 이미 가지고 있다는 것을 안다. 가령, 국세청은 개인의 고용과 소득, 연금공단은 노인과 기타 연금 생활자에 대한 정보를 가지고 있다. 이러한 상황에서 통계 생산을 위해 사람으로부터 유사한 데이터를 수집하는 것은 어려울 수 있다.

공공의 관점에서 행정자료를 통계목적으로 이용하는데 다음과 같은 장점이 있다. 첫째는 자료수집의 경제적 효율성이다. 둘째는 개인, 가구, 기업의 응답부담 경감이다. 셋째는 데이터를 다루는 사람의 수가 감소되고, 전자적인 형태로 자료처리가 이뤄짐으로써 데이터에 대한 보안 유지가 가능하다는 점이다. 이러한 장점이 통계 목적으로 기존 등록자료를 사용하는 긍정적인 논거로 시민들에게 받아들여질 수도 있다. 그러나 통계 당국은 이와 관련 항상 조심스럽게 접근하는 것이 중요하다. 일반 대중의 신뢰를 잃기는 쉽지만 그것을 다시 회복하는 데에는 많은 노력이 들어가기 때문이다.

다. 등록센서스 주축대장 정비작업

등록센서스 실시의 핵심적 전제조건은 통계작성을 위한 고품질의 행정자료의 존재와 이용가능성이다. 센서스에 활용되는 행정자료의 품질을 진단하는 방법은 행정자료로 센서스 통계를 작성하고 이것을 최근의 센서스 결과와

비교하는 것이다. 이를 통해 어느 주체의 행정자료 시스템이 만족스러운 품질의 자료를 생산하고 있는가, 어떤 주제영역이 추가적 개선이 필요한가를 평가할 수 있을 것이다. 이러한 행정자료의 품질평가절차가 등록센서스 실시 이전에 철저하게 이루어졌음을 앞에서 살펴본 덴마크의 사례를 통해 알 수 있다.

덴마크의 사회제도는 정확한 행정자료를 수집하는 기반이 되고 있다. 각종 사회보장제도가 거주지역을 근거로 하여 운영되기 때문에 등록부의 개인정보가 정확하다. 핀란드에서는 등록부가 개인의 기본 권리와 활동의 토대로, 국민들과 행정당국 모두 등록자료에 대한 신뢰도가 높다. 국가 모든 거주지에 대한 공식주소가 행정자료를 토대로 결정되는데, 가령 인구대장의 자료가 여권, 결혼, 이혼, 사망, 상속 등에 필요한 기본서류로서 서비스된다.

네덜란드의 사례를 보면, 다소 오래된 자료이기는 하지만 중앙인구대장의 품질을 구체적인 수치를 통해 알 수 있다. 1971년 센서스에서 인구대장의 0.02%가 과소등록이고 0.05%가 과다등록되어 있으며, 0.18%가 잘못된 지자체에 등록되어 있는 것으로 확인되었다. 중앙인구대장은 지속적으로 자료의 품질이 향상되어 왔는데, 과거에도 인구등록부가 높은 수준의 품질을 보이고 있음을 알 수 있다.

인구대장이 업데이트에 문제가 없는 이유는 중앙정부의 지방교부금 분배에서 인센티브로서 작용하기 때문이다. 또 다른 이유는 네덜란드 사회에서 중앙인구대장에 등록되지 않고 생활하는 것이 힘들기 때문이다. 네덜란드에는 등록을 의무로 규정하고 있는 법은 없지만 등록하지 않고 활동하는 것은 불가능하다. 가령, 등록을 하지 않고 일하면 누구라도 최고율의 소득세를 내야 한다. 인구대장은 경찰의 입국자 자료를 포함하여 다른 여러 행정자료의 비교를 통해 정기적으로 점검된다. 지방자치단체와 국민 개개인은 모두 중앙인구대장의 품질을 훌륭하게 유지할 수 있는 충분한 동기를 가지고 있다.

라. 개인식별시스템

행정자료를 통계목적으로 이용하기 위한 핵심 조건 중 하나는 개별 단위가 항상 독립적으로 인식될 수 있어야 하며, 서로 다른 자료원과 매칭될 수 있는 “통합된 식별시스템”(unified identification system)이 있어야 한다는 것이다. 이러한 통합된 시스템이 없으면 행정자료를 토대로 하는 통계작성에 절대적으로 필요한 서로 다른 행정자료를 매칭하는 것이 힘들다. 최소 요구 조건은 기본등록부의 통합된 등록 시스템이다(UNECE, 2007). 북유럽 국가들은

행정자료를 연결하는 매칭키(linking key)를 가지고 있는데, 이 매칭키로 이용되는 것이 개인식별번호, 수치주소(numerical addresses), 사업체번호이다. 식별번호시스템을 사용하는 것이 최선이지만 개별 단위의 이름, 주소, 생년월일 등에 관한 충분한 정보가 있으면 식별코드가 없더라도 식별이 가능하다.⁶⁾

북유럽 국가 통계기관에서는 고유 식별코드를 사용한 지 오랜 시간이 지났기 때문에, 전국적인 코드시스템이 잘 정비되어 있다. 그러나 북유럽 이외의 국가들에서는 식별코드가 아직 완벽하게 통계적으로 활용되지 못하고 있다(Wallgren and Wallgren, 2007). 가형, 동일한 식별코드를 갖는 두 개의 개별 단위가 있을 수 있는데, 이것은 매칭 에러 문제가 일정 정도 항상 발생한다는 것을 의미한다.

마. 행정기관의 협력

등록기반 통계를 작성하기 위해서는 관련 행정기관 간의 긴밀한 협조뿐 아니라 책임을 명확하게 공유할 필요가 있다. 북유럽에서 등록 기반의 통계시스템을 만드는 데는 국가적 지원이 있었으며, 통계법은 이러한 정책을 반영한다. 이는 통계작성기관이 행정자료에 대한 접근에 관해 행정당국과 협상하는데 있어 강력한 지원이 있다는 것을 의미한다. 관계 당국과 우호적인 그리고 밀접한 협력을 맺는 것은 행정자료의 효율적인 사용에 중요하다. 이를 통해 등록부의 내용에 대해 실질적인 영향을 발휘하고 통계 목적을 위한 행정자료의 이용에 대한 이해를 확산시킴으로써 통계생산을 위한 자료원이 개선될 수 있다.

행정장부는 규칙과 절차가 변경할 수 있는데, 이러한 변화가 통계 생산에 미치는 결과가 고려되어야 한다. 가장 최선의 안전장치는 등록부와 통계 담당자가 긴밀한 관계를 유지하는 것이다. 이들 사이의 접촉은 기록되어야 하며, 담당자가 변경될 때 문서로 전달되어야 한다(UNECE, 2007).

덴마크 통계청은 정부와 지방 행정기관 시스템의 사용자 위원회의 일원이며, 등록부 개선 방안은 덴마크 통계청이 여러 분야의 통계 영역에 걸쳐 만들어 놓은 자문위원회에 의해 일상적으로 논의된다

Statistics Finland(2004)의 내용을 중심으로 핀란드를 사례로 행정자료의 통계적 활용을 위한 통계청과 다른 행정기관간의 협조를 좀 더 구체적으로 살펴본 결과는 다음과 같다. 재무부는 광범위한 조사표에 기반한 마지막 센서

6) 독일은 2011년 센서스에서 인구부문의 식별코드 없이, 개인의 정보(이름, 주소, 생년월일 등)를 통한 자료 간 매칭을 계획하고 있다.

스가 되어야 하는 1980년 센서스와 관련해서, 그리고 등록기반 시스템의 발전을 위해서 핀란드 통계청의 위상을 명확하게 만들었다. 이러한 재무부의 중재는 다른 행정당국에 대한 핀란드 통계청의 협상 위치 강화에 도움을 주었다. 핀란드 통계청은 각 등록당국에 대해 담당자를 지정한다. 그의 업무는 당국자와 의사소통의 개방 채널을 유지하고, 관련 분야 안에서의 변화를 모니터하며, 등록자료의 통계적 적용 가능성을 유지 또는 향상시키기 위한 작업을 하는 것이다. 또한 각 등록당국도 한 명의 통계 담당자를 지명한다. 한편 핀란드 통계청은 주요 문제를 토의하고 협력의 진행 과정을 모니터하기 위해 등록당국자의 최고위자와 연례 모임을 개최한다. 핀란드 통계청과 등록당국 사이의 이러한 협력은 핀란드 공식통계 조정시스템의 일부분이다.

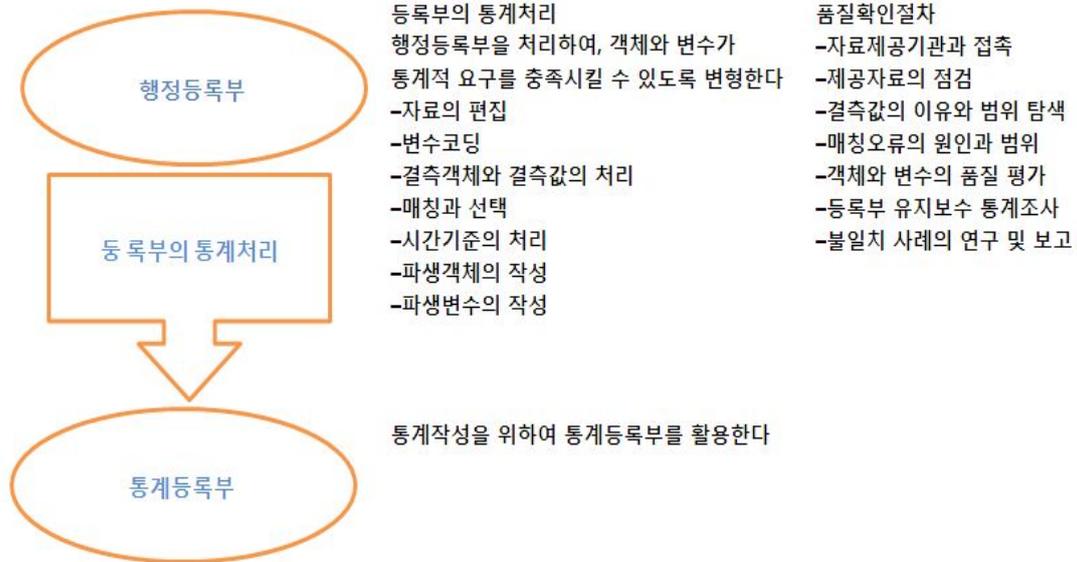
기본 대장을 유지하는데 책임 있는 정책담당자들이 '행정대장 관리연합'(the Register Pool)이라는 특별팀을 운영한다. 이 팀은 행정대장의 이용 가능성과 내적 일관성 향상, 내용·품질·접근성 개선, 효율적인 정보 시장의 창출을 위한 등록당국자들 사이의 정보 교환과 협력 증진을 목적으로 한다. 핀란드 교통부, 헬싱키 시, 국토정보원, 법무부, 특허 및 등록위원회, 지역 자치단체 연합, 통계청, 국세청, 인구등록센터 당국자들이 현재 '행정대장 관리연합'에 참여하고 있다. 행정기관의 협력은 '행정대장 관리연합'의 범위 안에서 이루어질 뿐만 아니라 정규적인 합동 프로젝트와 회의의 형태로도 이루어지고 있다.

3. 자료 생산 절차

행정자료를 관리하는 기관은 행정 목적에 맞게 자료를 제어하고, 수정하고, 기타의 처리 과정을 수행한다. 통계기관의 수요를 충족시키기 위해 행정당국이 통계작성기관에 데이터를 보낼 때 선별과 처리 과정이 추가적으로 수행되는 경우도 있지만 일반적으로 행정자료에서 직접 통계를 작성한다는 것은 바람직한 생각이 아니다(Wallgren and Wallgren, 2007). 개별 단위의 정의와 변수가 에디팅 되어야 하는 등 행정자료가 통계 목적을 충족시키기 위해서는 별도의 처리 과정이 필요하다. 이러한 "등록 장부의 통계적 처리 과정"(register-statistical processing)은 하나 혹은 여러 개의 행정대장(administrative register)을 단일한 통계등록대장(statistical register)로 전환시키는 것을 목표로 한다. 통계등록부 안에 데이터들이 조직되는데, 개별 등록부의 데이터들은 식별키를 통해 서로 매칭된다. 등록부의 통계적 처리 과정은 데이터의 편집, 변수의 코딩, 결측된(missing) 개별 단위와 변수의 처리, 매칭과 선택, 기준 시점의 처리, 파생된

(derived) 개별 단위의 생성, 파생된 변수의 생성 등의 과정을 포함한다(<그림 4>를 참고할 것)

<그림 4> 행정등록부의 통계등록부로의 전환과정



자료: Wallgren and Wallgren (2007) Register-Based Statistics, p 5, Chart 1-3.

이것은 모든 등록 장부가 개별 등록 장부로 인식되는 것이 아니라 단일한 시스템의 일부로 인식된다는 것을 의미한다. 이러한 관점은 데이터의 수집, 데이터의 처리, 품질의 제어와 보급 등 통계 생산의 모든 단계에 영향을 미친다. 이러한 통합된 시스템 아래서 모든 행정당국은 단일한 코드 체계를 사용하게 되며, 이를 통해 서로 다른 기관의 등록부가 매칭될 수 있게 된다.

한편 행정등록부에서 통계등록부로의 전환 과정에서는 행정등록부에 대한 품질 점검이 필수적인데 이를 위해서는 데이터 공급자와의 접촉이 기본적으로 필요하다. 이와 함께 입수된 데이터의 품질에 대한 점검, 결측값의 이유와 범위 파악, 잘못된 매칭의 이유와 범위 분석, 객체와 변수의 질 분석, 등록부의 관리에 관한 연구, 불일치의 관찰과 보고 등의 과정이 필요하다.

가. 행정자료의 품질진단

덴마크와 핀란드에서는 행정자료에 대한 품질진단이라는 등록센서스 실시 이전의 사전 준비가 철저하게 이루어졌다. 핀란드의 사례를 살펴보면, 등록자료의 신뢰성을 조사하고 관찰하는 연구가 등록센서스 실시의 결정에 앞서 수행되었다. 1975~1985년 센서스에서 조사표는 여전히 사용되었는데, 등록부 자료는 이 결과와는 물론 동일 시점의 다른 조사 결과와도 비교되었다. 1980년 노동력조사는 충분한 비교를 위해 약 7만 명으로 확대되었다. 등록장부를 이용하여 생산한 자료와 조사표에 의해 생산된 자료의 차이가 작아서 등록센서스가 계속 추진될 수 있었다(Statistics Finland, 2004).

핀란드에서는 1990년 센서스를 100% 완전한 등록센서스로 실시하기 위하여, 기간 행정자료에 대한 신뢰성 조사를 실시하였다. 행정자료를 바탕으로 작성된 통계등록대장 자료를 인구와 주택의 등록대장에서 약 2% 표본을 추출하여 결과를 비교 검토하였다. 그 결과 조사자료와 등록자료가 일정 정도 차이를 보였지만, 두 조사자료 사이의 차이가 조사표에 의한 두 개의 조사로부터 얻은 자료 사이의 차이보다 더 크지는 않았다.

인구자료의 정확성이 인구등록센터에 의해 확인되지만, 고용통계와 같은 등록부의 신뢰성을 점검하기 위한 품질조사가 매년 수행된다. 매년 인구등록센터는 인구정보시스템에 기록된 주소자료의 정확성을 입증하기 위한 조사를 통계청에 의뢰한다. 이를 위해 핀란드 통계청은 약 1만 명의 사람들에게 조사표를 보내는데, 2003년에는 응답자의 99%가 정보가 업데이트되어 있다고 응답하였다.

나. 행정등록부의 통계처리과정

통계등록부(statistical register)의 생성에서 가장 중요하고 자원이 많이 소요되는 과정은 기본 데이터의 입력을 확인하고 에러를 수정하는 것이다. 등록부를 운영하는 터의 입에서 수행하는 에디팅의 목적은 터의 목적을 위해 데이터의 질을 향상시키는 것이다. 이는 이러한 에디팅이 주로 터의 목적에 가장 중요한 변수에 집중되어 수행된다는 것을 의미한다. 이에 따라 통계 목적을 위해 필요한 주된 변수는 세심하게 취급되지 않을 수도 있다. 통계담당자가 이 에디팅의 과정과 한계에 대해 많은 지식을 갖는 것은 중요한 일이다.

국가통계기관에 의한 에디팅은 데이터의 통계적 사용에 그 초점이 맞추어

진다. 각각의 등록부는 에러를 제거하기 위해 내부적 로직에 의해 점검되며, 포괄범위 문제와 기타의 에러들이 등록부의 매칭 과정에서 발견될 수 있다.

통계등록부는 이상적으로는 개별 데이터들 간에 불일치가 발생하지 않기 위해서는 ‘깨끗하고(clean)’ 일관성이 있어야 한다. 이를 위해 있음직하지 않은 정보들은 검토되어야만 하는데, 이 과정에서 사용되는 주된 방법은 전산화된 논리적 에디팅(computerized logical editing)이다. 이 과정에서 행정등록부의 시스템적인 에러가 발견되는 것은 흔히 있는 일이다. 데이터가 추출되는 행정등록부에 포함되어 있는 시스템적인 에러를 수정하는 것은 중요하다. 이것이 이루어지기 위해서는 등록부를 담당하는 당국과 긴밀한 관계를 맺는 것이 중요하다. 통계 생산자에 의해 발견되어 행정등록부가 수정되는 많은 사례가 있으며, 이는 등록부와 통계 둘 다에 이익이 된다.

다. 메타 데이터의 정비

등록 기반의 데이터 시스템은 매우 복잡하며, 다양한 원천의 행정자료를 이용하여야 하기 때문에 집계범위(coverage)가 완전한 근접하고, 체계화된 메타데이터가 대단히 중요하다. 메타데이터는 통계목적으로 행정자료를 사용하는 데 있어서 핵심적인 역할을 한다. 행정당국은 정의(definition), 데이터의 형태와 질에 관한 메타데이터를 가지고 있는데, 등록부를 관리하는 주체로부터 나온 충분한 메타데이터는 통계자료의 품질을 결정하는 관건이다.

Goossens(2006)에 의하면 메타데이터는 기술적인 메타데이터, 내용 관련 메타데이터, 데이터 처리 과정에 관한 메타데이터로 구분될 수 있다. 기술적인 메타데이터는 변수의 설계서(specification of variable), 필드의 길이, 데이터의 타입 등과 관련되며, 내용 메타데이터는 데이터의 요소에 대한 정의 등과 관련된다. 과정에 대한 메타데이터는 외부 데이터를 잘 이해하고 수정하기 위해 필요한 메타데이터로, 데이터의 공급자에 의해 수행된 데이터의 변형, 계산 등에 관한 과정뿐 아니라 데이터가 수집된 이력, 즉 어떤 행정기관에 의해 그 과정이 이루어졌는지 등에 관한 과정을 다룬다.

G. 영국의 One Number Census와 센서스 범위조사의 개관

1. 배경

영국은 1801년 최초의 근대적 센서스를 실시하였으며, 그 후 1941년을 제외하고 10년마다 센서스를 실시하였다. 새로운 문항이 추가되고 불필요한 문항은 삭제되기도 하지만, 센서스 실시의 기본원칙은 변함이 없었다. 1911년까지는 영국 의회가 센서스를 실시할 때마다 새로운 센서스법을 통과시켜야 했지만, 1920년 센서스법(Census Act)의 통과로 직전 센서스가 실시된 후 5년이 경과하고, 특별 센서스의 필요성을 영국 의회가 인정하는 추가입법을 승인하게 되면, 언제라도 센서스를 실시하는 것이 가능하게 되었다. 그러나 영국의 센서스는 1939년 국민등록법(National Registration Act)이 1941년의 센서스를 대체했던 경우를 예외로 하고, 관례대로 10년마다 계속 실시되어 왔으며, 2001년은 영국 센서스 실시 200주년을 기념하는 연도였다.

영국에서 센서스의 품질을 평가하기 위하여 사후조사를 처음으로 실시한 것은 1966년의 표본센서스부터였다. 1971년 센서스에서는 사후조사가 체계적으로 이루어지지 못했다. 대신, 1981년에는 '센서스 집계조사'(CES, Census Enumeration Survey)라 불리는 사후조사가 실시되었다. 잉글랜드와 웨일즈의 경우, 당시 총인구 4,960만 명의 2.0%(1백만 명)이 센서스 본조사에서 과소집계된 것으로 나타났다. 그러나 사후조사 결과에 의하면, 순(純) 과소집계는 0.4%(21만 명)에 불과하였으며, 나머지 1.6% (80만명)는 실제로 거주하고 있으나, 조사표를 제출하지 않은 "부재가구"의 거주자로 확인되었다. 사후조사의 결과 과소집계된 인구는 1981년 센서스의 최종결과에 추가되었으며, 1981년 추계인구의 최종수치를 보정하는데 사용하였다.

1991년 사후조사의 결과에 의하면, 1991년 센서스는 총인구 5,455만 명의 3.8%(2.1백만)를 제대로 파악하지 못하였다. 그 중에서 1.6%(90만 명)는 조사 당시의 부재가구에 보정(imputation) 집계되었으나, 나머지 2.2%(1.2백만 명)는 최종집계에서도 여전히 누락된 것으로 나타났다. 이것은 등록형 또는 순환센서스로 방향을 전환하고 있는 유럽의 대륙 국가들이 영국처럼 고전적 의미의 센서스를 실시하던 때보다는 양호한 결과라고 할 수 있을 것이다. 그러나 순 과소집계가 1971년 센서스보다도 크게 늘어났고, 특정 지역이나 특정의 성·연령별 인구집단에서 훨씬 큰 것으로 나타나서 사회적 관심의 대

상이 되었다. 영국 통계청은 1991년 센서스의 현지실사에서 실제로 거주하고 있으나 응답하지 않은 가구, 이른바 ‘부재가구’를 대상으로 보정집계를 실시하였다. 이 작업에서 총인구의 1.6%인 90만 명을 파악하여 센서스의 데이터베이스에 추가하였다. 보정집계 결과는 센서스 제표작업에 포함되었으며, 따라서 센서스 본조사의 집계결과로 간주되었다.

1991년 사후조사는 ‘센서스 타당성조사’(CVS, Census Validation Survey)로 명명되었는데, 이것은 과소집계의 정도나 분포형태의 파악에서 실패작으로 취급되고 있다(United Kingdom Office for National Statistics, 1998, 2001, 2003). 그 이유는 여러 가지이다. 먼저, 사후조사가 2만 가구의 표본규모로 실시되어 지역별 과소집계의 상세내역을 파악하는데는 역부족이었다. 다음은 사후조사가 센서스 본조사와 조사설계상 완전 별도의 독자적인 조사가 되지 못했다는 점이다. 사후조사가 과소집계의 내역을 상세하게 파악하지 못함으로 해서 야기된 문제를 해결하는데 꽤나 많은 시간이 소요되었으며, 1991년 센서스 실시 후, 상당기간 영국내 거주인구의 집계수치에는 적어도 4가지 버전이 동시에 공존하고 있었다. 당초에는 총인구의 1.6%에 해당하는 부재가구의 보정집계를 포함하는 센서스 본조사의 집계결과가 발표되었다. 다음은 사후조사의 결과를 토대로, 센서스 본조사의 집계결과를 상향조정하는 결과가 발표되었다. 또 영국 통계청은 사후조사의 결과를 무시하고, 인구분석의 결과를 토대로 전국추계인구를 작성하고, 지방자치단체의 추계인구에 대해서는 수리모형을 적용하겠다고 발표하였다. 당시 인구분석은 1981년 센서스 본조사의 결과와 1981~1991년의 출생·사망·이동 등의 인구동태통계를 토대로 하여 1991년의 성·연령별 인구구조를 파악하도록 하였으며, 인구분석의 결과와 센서스의 집계결과를 비교하여, 센서스 집계의 최종수치를 확정하기로 하였다. 인구분석의 결과에 의하면, 1991년 센서스에서 총인구의 1.6%(90만 명)에 해당하는 부재가구 거주자를 제외하고도, 총인구의 2.2%(120만명)가 여전히 센서스 본조사에서 제대로 파악되지 못한 것으로 나타났다.

영국 통계청이 지방자치단체별로 과소집계의 정도를 추정하여, 그 결과를 공표하지 못하면서, 경제사회연구심의회(Economic and Social Research Council)가 프로젝트 “Estimating with Confidence”에서 작성한 추정인구가 각종 지방행정업무에 활용되기도 하였다(United Kingdom Office for National Statistics, 1998, 2001, 2003).. 결국, 1991년 센서스의 공식적 집계 결과는 1981년 센서스와 인구동태통계의 인구분석을 토대로 하여 작성된 인구 추계의 결과와 일치하지 못했으며, 이로 말미암아 인구조사의 정확성에 대하여 ‘신뢰성의 상실’ (loss of confidence)이라는 문제가 생겨났다.

영국의 통계이용자들은 2001년 센서스 실시 이전에 광범위하게 개최된 전문가회의에 참석하여, 2001년 센서스에서는 하나의 확실한, 최고의 인구를 추정하여 줄 것을 요구하였다. 더욱이, 그들은 1991년 센서스에서 영국의 통계청이 과소집계를 측정하는 문제에 큰 난관에 봉착하였음을 염두에 두고, 센서스의 최종결과에 신뢰성을 담보할 수 있도록 하기 위하여, 집계오류의 원인을 추정할 수 있는 강력한 방법론을 도입하여야 한다고 주장하였다. 영국의 통계청은 ONC 프로젝트가 이 문제에 대한 해답을 제공할 수 있다고 판단하였다. ONC 프로젝트는 사후조사를 완전히 재설계하여, 내용오류와는 별도로 과소집계의 정도를 추정하여 센서스의 개인별 데이터베이스를 보정하여, 센서스 제표작업의 최종결과는 전국수준의 추정인구와 일관성을 지닐 수 있도록 하였다.

물론, ONC 프로젝트는 종전의 센서스에서 이룩한 성과를 기반으로 하는 것이 적지 않았다. 가령, 1991년 센서스에서 부재가구(센서스 본조사의 현장 실사에서 부재중이나 실제로 거주하고 있는 것으로 파악된 가구)의 보정집계 작업을 실시하기 위하여 도입된 누락변수의 자동적인 편집 및 보정절차는 2001년의 ONC 프로젝트의 매칭작업에서 누락된 것으로 추정되는 개인과 가구를 보정하는 작업에서 지속적으로 활용되었다. 더욱이, 연앙추계인구는 ONC 프로젝트의 결과를 바탕으로 한 것이기 때문에, 이들 수치 간에는 간단한 연결 장치를 만들어야 할 필요성이 절실히 되고 있음이 확인되었다.

영국의 2001년 One Number Census 프로젝트의 궁극적 목표는 2001년 센서스 실시 후에 공존하는, 1991년과 같은 4개 버전의 추정인구가 아니라 ‘최종적인 세트의 센서스 추정인구’ (a final set of census estimates)를 작성하는 것이었다. 영국의 통계청은 2001년 센서스에서 ONC 프로젝트의 도입으로 당초의 목표를 달성하는데 100% 성공을 거두지는 못하였다는 자체평가를 내리고 있다. 본 연구는 ONC 프로젝트가 완전히 성공을 거두지 못한 이유와 프로젝트 자체가 야기한 문제점에 대해서는 제3절의 ONC 방법론에 대한 평가연구와 제4절의 2001년 ONC 프로젝트의 교훈에서 검토하기로 한다.

2. 방법론의 개관

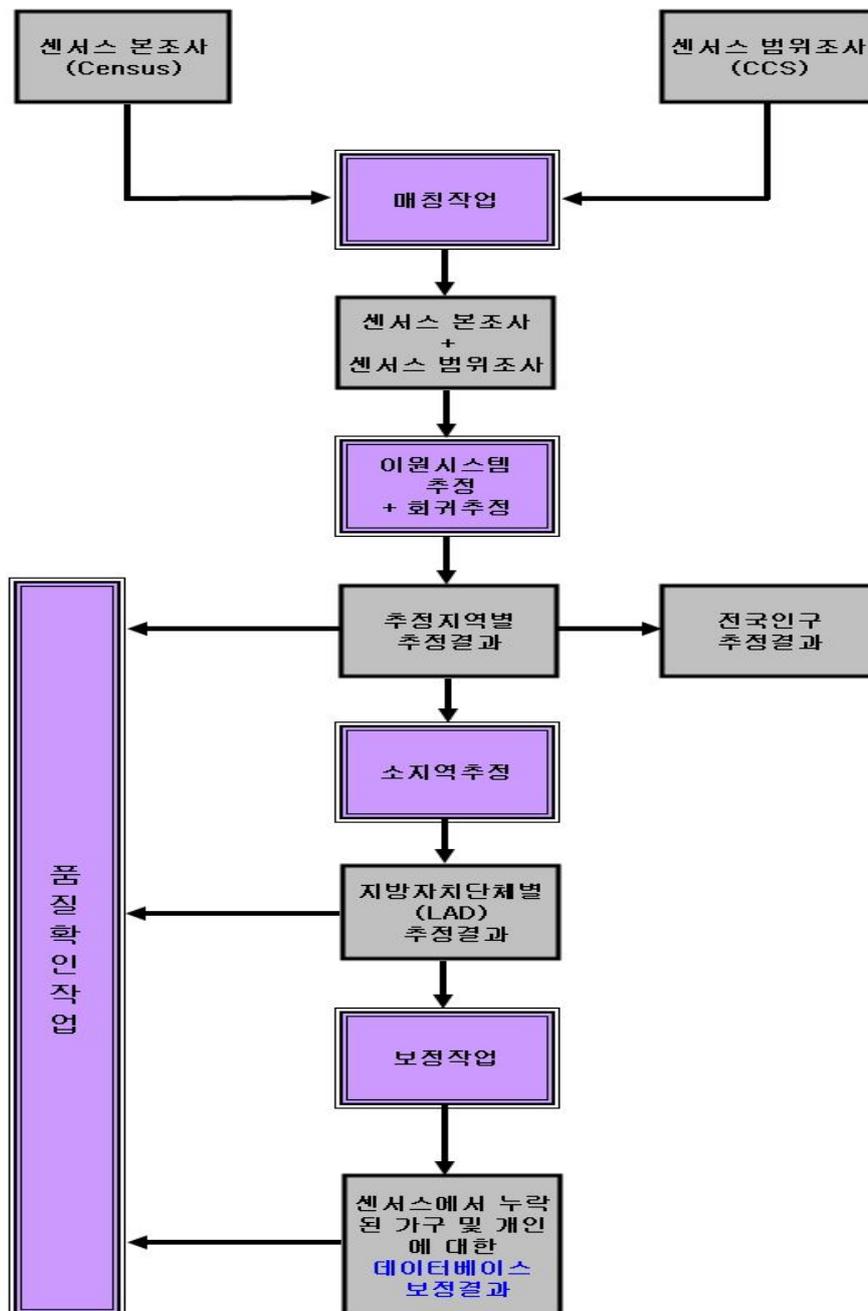
영국의 2001년 ONC 방법론은 5개의 주요단계로 구성되어 있다.

- 센서스 범위조사의 표본설계
- 센서스 본조사와 범위조사의 매칭작업

- 추정지역 및 지방자치단체별 추정작업
- 가구 및 개인에 대한 데이터베이스 보정작업
- 추정지역 및 지방자치단체에 대한 품질확인작업

<그림 1>은 ONC 방법론의 주요단계를 표본설계, 매칭작업, 추정작업, 보

<그림 1> One Number Census 방법론의 플로우차트



정작업, 품질확인작업으로 구분하여, 작업순서를 흐름도의 형태로 보여준다. 현대국가가 전수조사와 현지실사⁷⁾의 기본원칙을 준수하여 센서스를 실시하는 경우, 주권행사가 가능한 영토 내에 거주하는 개인과 가구를 100%를 정확하게 파악하기를 원한다. 그러나 어떤 국가의 센서스도 100% 완전무결한 경우는 없으며, 본조사에서 빠뜨려지는 개인과 가구가 있게 마련이다. 이 과소추계의 정도는 지역은 물론 성·연령별 인구집단의 속성에 따라 다르게 나타난다. 따라서 센서스를 실시하는 현대국가는 사후조사를 실시하여, 통계추정으로 과소집계를 평가하는 것이 일반적인 관례이다. 특히 선진국은 물론 최근의 신흥공업국에서도 센서스나 표본조사에 대한 비용답자의 비율이나 특성이 달라지고 있으며, 이에 따라 집계누락의 정도나 유형을 제대로 파악하여, 센서스 본조사의 효용가치를 증대시키는 것이 국가통계기관의 당면과제가 되고 있다.

영국 통계청은 사우스햄턴대학(Southampton University) 통계학과 교수들의 자문을 토대로 One Number Census라는 새로운 방법론을 구상하기 시작하였다. ONC 프로젝트는 1991년 센서스 평가작업을 완료한 1996년부터 개시되었으며, 센서스 통계이용자들이 제기한 과소집계의 문제를 우선적으로 해결해야 한다는 것이 주요 쟁점으로 부각되었다. 영국 통계청은 ONC 프로젝트와 관련하여, 외부전문가, 지방자치단체 대표, 통계학술기관 등을 포함하는 운영위원회를 조직하여, 프로젝트의 방법론 개발과 운영관리를 감독하고 지도하도록 하였다. 또 영국 통계청은 내부 국장급 직원과 여타 정부기관의 대표, 외부전문가들로 구성된 ONC Project Board를 설치하여 프로젝트의 진전 상황을 일일점검 하도록 하였다.

One Number Census는 이론적으로는 그리 복잡한 것이 아니었다. 사후조사를 대규모의 표본조사로 실시하여, 이것을 센서스 본조사의 집계범위를 측정하고 과소집계의 정도를 평가하기 위한 수단으로 이용하였다. 다음으로 사후조사의 결과를 개인수준에서 2001년 센서스 본조사의 결과와 짝을 맞추는 매칭작업을 실시하였다. 센서스 본조사와 사후조사의 매칭작업의 결과를 토대로, 이원시스템추정법을 사용하여 본조사에서 파악되지 못한 개인과 가구의 총수를 추정하는 작업을 실시하였다. 마지막으로, 일정기준에 따라 실사

7) 전수조사와 현지실사는 전통적 센서스의 골자라고 할 수 있는데, 2000년 라운드 센서스를 실시하는 247개 국가 중에서, 196개 국가가 이 방법을 사용하고 있는 것으로 나타난다. 등록센서스는 비용절감과 높은 빈도로 통계를 작성할 수 있다는 장점은 있으나, 현지실사라는 조건을 충족하지 못한다는 점에서 전통적 방법과 차이가 있다. 또 순환센서스는 전국의 지역들을 일정기간에 걸쳐서 순차적으로 조사를 실시하기 때문에, 센서스의 기본원칙 중에서 동시성을 충족시키지 못하고, 전체인구의 전수조사가 아니라는 점에서 보편성을 충족하지 못하는 문제점이 있다.

된 개인의 정보를 복사하는, 이른바 ‘기부자 방법’ (donor method)을 이용하여, 센서스에서 누락되었다고 추정되는 개인과 가구에 대한 정보를 보정하는 작업을 실시하였다. 영국의 ONC 프로젝트 추진방안은 개인별 데이터베이스의 보정작업을 제외한다면, 전수조사와 현지실사의 기본원칙을 준수하는 근대적 센서스를 실시하는 국가들이 대부분 채택하고 있는 방법이라고 할 수 있다. 가령, 미국의 센서스국도 1950년 이후 사후조사를 실시하여 과소집계의 정도를 파악하기 위하여 노력하여 왔다. 그러나 대부분의 국가에서 사후조사를 실시하는 주목적은 센서스 본조사의 집계결과를 보정하기 위한 것이 아니라, 센서스 본조사의 정확성에 대하여 통계이용자들의 주의를 환기하기 위한 것이었다.

영국의 ONC 프로젝트는 센서스 조사환경의 악화⁸⁾라는 문제를 염두에 둘 때, 구체적 실천 자체는 결코 간단한 문제가 아니었다. 수많은 근본적 질문, 대안, 가정들이 영국 통계청의 내검 대상과 전문가회의의 토론주제가 되었다. 프로젝트의 개발과정에서 몇 가지 대안들(예: 센서스 본조사, 범위조사, 행정자료의 동시 이용과 관련 추정방법의 개발, 센서스 조사구 수준에서 과소집계를 측정하는 수리모형의 구축 등)을 모색하였다. 본 연구는 제한된 지면 때문에 각종 대안전략, 특히 ONC 운영위원회이나 영국 국내 또는 유럽의 학술회의에서 발표된 ONC 방법론에 대하여 구체적으로 설명하지 못한다.

3. 센서스 범위조사

2001년의 센서스 범위조사는 1991년의 사후조사인 ‘센서스 타당성조사’ 사이에는 몇 가지 차이점이 있다. 우선, 범위조사는 과소집계를 측정하는데 초점을 맞추고, 본조사의 내용오류를 평가하기 위하여 별도의 독자적인 조사를 실시하기로 하였다. 둘째, 범위조사의 표본 수는 36만 가구 수준으로 대폭 증가하였다. 1991년의 타당성조사는 2만 가구 정도의 소규모 표본조사가 불가피하였는데, 그 이유는 센서스 본조사의 응답오류를 평가하는데 필요한 정보를 획득하는데 인적·물적 자원을 너무 많이 할애하였기 때문이다. 앞에서 1991년 사후조사의 실패를 지적했던 것처럼, 영국 통계청은 2만 가구

8) 센서스 환경은 사생활 의식의 강화, 보안의식의 강화, 주거환경의 변화, 생활환경의 변화, 지역공동체 의식의 약화 등을 들 수 있다. 이에 따라, 조사표의 배포방식이나 회수방식의 변화가 일어난다. 그러나 이러한 변화가 대부분의 경우, 조사표의 회수율에 심대한 영향을 미친다. 가령, 영국의 경우도, 우편배포와 우편회수의 방법을 채택하면서, 응답율의 변화가 크게 일어났다는 점을 고려해야 한다.

정도의 소규모 표본조사를 토대하여 우리나라의 시군구 급의 기초자치단체에 해당하는 소지역 단위는 물론 전국수준에서도 집계범위를 제대로 평가하는데 심각한 어려움에 직면하였다. 당시 전문가들은 센서스 본조사의 과소집계를 정확히 평가하기 위해서, 센서스 범위조사의 표본규모가 대폭적으로 확대되어야 한다고 주장하였다. 최종적으로, 센서스 범위조사는 영국 전역의 하위지역과 성·연령별 인구속성변수를 통제변수로 하여 과소집계를 가능한 정확하게 추정하여, 센서스 본조사의 집계결과를 보정하는 작업을 진행할 수 있도록 해야 한다는 것이었다.

<그림 2> 영국의 우편번호체계

A9 9AA
A99 9AA 영국의 우편번호는 5-7개의 문자 또는 숫자로 구성된다.
A9A 9AA
AA9 9AA
AA99 9AA
A99A 9AA

구성포맷	한글명칭	영문명칭	개수
A or AA	우편번호 지역	Postcode Area	124
9, 99 or 9A	우편번호 구역	Postcode District	3,064
9	우편번호 섹터	Postcode Sector	11,598
AA	우편번호 단위	Postcode Unit	1,780,000
우편번호 주소 Postcode Address			27,000,000

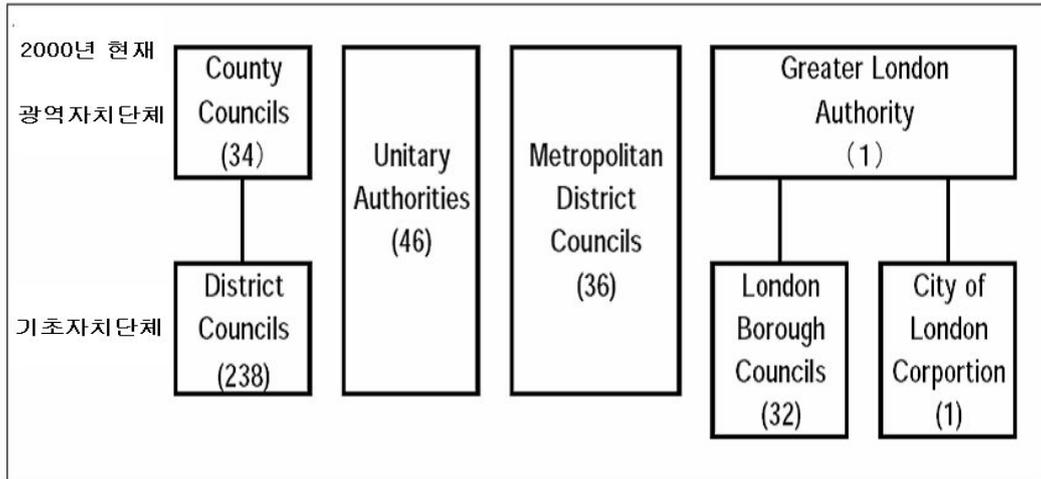
**영국의 우편번호는 대략 15개의 주소로 구성된다.

센서스 범위조사는 영국의 우정사업본부(Royal Mail)가 관리하는 영국 전역의 178만 우편번호 중에서 표본으로 추출된 19,500개의 우편번호를 대상으로 36만 가구를 현지실사하고, 개인과 가구의 수를 100% 완전무결하게 파악하는 것을 최종목표로 설정하였다(영국의 우편번호체계에 대해서는 <그림 2>를 참고할 것). 센서스 범위조사는 사후조사의 일반적 관례대로, 대규모의 공동보호시설(Communal Establishment, 예: 교도소, 병원, 장기요양소 등)을 포함하지 않았는데, 이들 시설가구 인구에 의한 추정작업의 문제점은 품질확인작업에서 검토과제로 남게 되었다. 또 센서스 범위조사의 현지실사에서 조사지역의 우편번호가 우범지역이나 불량주택지대여서 다른 곳으로 교체해야 하는 경우도 종종 있었다.

1) 추정지역

센서스 범위조사의 표본설계는 지방자치단체별로 실시되어, 이들 각각에

<그림 3> 영국 잉글랜드 지방자치단체의 구조: 2000년 현재



대해서 과소집계의 추정치를 계산할 수 있게 하는 것이 이상적일 것이다. 그러나 영국 통계청은 영국 전역에 433개의 지방자치단체⁹⁾가 있어서, 직접추정을 가능하게 하는데는 100만 가구 이상(2001년 기준 2,330만 총가구의 약 4%에 해당)의 표본이 필요한데, 이 정도로 대규모의 표본조사를 계획한다는 것은 인원배치나 재정현황을 고려하여 현실적으로 불가능한 일이라고 판단하였다. 따라서 영국 전역에 대하여 인구 50만 명을 기준으로 하여 433개 지방자치단체(단체별 평균인구 12만 명)를 112개 지역으로 분할하였다. 이들 지역은 추정지역(Estimation Area) 또는 설계집단(Design Group)으로 명명되었으며, 각각의 추정지역은 하나의 지방자치단체(LAD, Local Authority District)로 구성된 경우도 있었고, 둘 또는 셋 이상으로 구성되는 경우도 있었다. 추

9) 2001년 ONC 프로젝트에서 추정인구를 작성하기로 한 지방자치단체는 2000년 현재, 잉글랜드 353개, 웨일즈 22개, 스코틀랜드 32개, 북아일랜드 26개 등 433개로 구성되어 있다. 잉글랜드의 경우는 이층 행정구역으로, 광역자치단체와 기초자치단체의 통합적 기능을 동시에 가지는 Unitary Authority 46개, Metropolitan District Council 36개와 기초자치단체의 기능만을 가지는 District Council 238개, London Borough Council 32개 City of London Corporation 1개를 합쳐서 353개로 구성되어 있다. 잉글랜드에서 County Council 34개(우리나라의 시도급 자치단체보다는 규모가 작고, 시군구급 자치단체보다는 규모가 큼)와 Greater London Authority(런던광역시)는 광역자치단체의 기능만을 가진다. 웨일즈는 22개가 모두 1층 행정구역으로 광역자치단체와 기초자치단체의 기능이 통합되어 있다. 한편, 스코틀랜드 District Council 26개, 북아일랜드 Unitary Council 32개가 있다. 이들도 모두 광역자치단체와 기초자치단체의 기능을 겸하고 있는 일층의 통합적 지방자치단체라고 할 수 있다.

정지역의 설계 작업에서 추정지역들의 규모나 동질성과 둘 또는 세 개 이상의 지방자치단체를 하나의 추정지역으로 묶을 때 생겨나는 현실적 또는 정책적 문제점들이 충분히 조화될 수 있게끔 배려하였다. 추정지역의 설계작업을 통하여, 추정지역에 대하여 통계적 추정방법론의 적용이 가능하도록, 추정지역의 과소집계 유형에 대하여 몇 가지 가정을 설정하게 된다면, 각각의 추정지역에 대하여 직접추정을 하더라도 상당정도의 정확성이 보장될 수 있다는 사실을 확인하였다. 따라서 개별 지방자치단체에 대해서 최소한 3~4개의 표본지역을 포함할 수 있도록, 표본선정 작업에서 특별히 배려하였다.

2) 표본설계

ONC 프로젝트는 112개의 추정지역(잉글랜드와 웨일즈 101개, 스코틀랜드 7개, 북아일랜드 4개)에 대하여, 영국 우정사업본부가 관리하는 평균 15개의 주소로 구성된 178만 개의 우편번호 중에서 19,500개의 표본을 선정하였다. 그러나 실제로는 추정지역마다 별개의 독립된 표본을 선정하였다고 말해야 할 것이다. 집락표집법을 이용하여, 우편번호를 표본으로 추출하는 이유는 실용적 목적과 통계적 목적을 동시에 충족시키기 위해서였다. 구체적으로, 우편번호 주소파일(Postcode Address File)은 센서스 본조사에 이용되는 가구명부와는 달라서, 대부분의 가구주는 그들이 살고 있는 가구의 우편번호를 알고 있다는 것이다. 그러나 우편번호에서 표본을 추출할 때 일어나는 문제점은 조사대상자에 대해서 주소나 집번지를 제외하고 얻을 수 있는 유용한 정보가 별로 없다는 것이다. 이 때문에, 우편번호별로 가구수를 변수화하는 것을 제외하고, 우편번호를 직접 층화하는 것은 불가능하였다. 따라서 11개의 추정지역에 대하여 채택된 표집전략은 개별 우편번호를 1991년 센서스 조사구(CED)와 매칭하여, 해당 우편번호에 대하여 1991년 센서스 데이터베이스의 변수정보를 이용하는 2단계 표본추출법을 구사하는 것이었다.

먼저, 1991년 센서스 조사구의 표본을 선정하여, 과소집계의 정도가 조사구별로 불균등하게 분포되었을 것이라는 가정과 함께, 집계난이도 지표에 따라 센서스 조사구의 표본을 층화하였다. 센서스 범위조사의 표본설계는 당초에 우편번호를 단위로 하는 지역간 표집편차를 최소화하여, 추정인구를 상당정도 정확하게 작성하는데 도움이 되도록 하기 위한 것이었다. 1991년 센서스 조사구도 1991년 센서스 데이터베이스를 이용하여, 표본편차를 통제하기 위하여 24개의 성·연령(5세 간격)별 하위집단¹⁰⁾에 대하여 층화작업을 진행하

10) 인적·물적 자원의 제약을 고려하여, 센서스 범위조사는 24개의 성·연령집단(남녀별)과

였다.

둘째로, 1991년 센서스 조사구가 선정되면, 제2단계 표집절차는 선정된 1991년 센서스 조사구 안에서 우편번호의 가구수를 기준으로 3~5개 정도의 우편번호를 무작위로 선택하는 것이다.

3) 집계난이도 지표

집계난이도 지표는 1991년 센서스의 개인 데이터베이스를 가공하여 실업률, 외국어(출생국가) 사용여부, 다가구주택, 세입자 동거여부, 1991년 집계누락 여부 등의 구성변수를 추출하여 작성되었다. 이 지표는 과소집계의 정도와 상관관계가 높았으며, 2001년 센서스 예비조사에서도 무응답과 상관관계가 높은 것으로 확인되었다. 영국 전체를 집계난이도 지표를 기준으로 3개 집단으로 구분하였다. 가령, 잉글랜드와 웨일즈 지방에서 지표 점수를 하위 40%는 집계가 쉬운 “1점”, 다음 40%는 집계가 쉽지도 어렵지도 않은 중간 정도의 “2점”, 그리고 상위 20%는 집계가 어려운 “3점”으로 구분하였다. 집계난이도 지표를 작성하는 이유는 센서스 조사구의 모든 유형에 대하여 센서스 본조사의 모집단을 반영하도록 표본을 골고루 분산하여, 과소집계의 수준을 정확하게 파악하고자 하는 것이었다.

영국의 ONC 프로젝트는 집계난이도와 관련하여 축적된 지식이 없었기 때문에, 이 지표가 당시의 지식수준으로 구성될 수 있었던 최고의 지표라고 판단하였다. 따라서 센서스 범위조사가 집계난이도 지표의 점수가 낮은 지역과 높은 지역에서 모두 실시될 수 있도록 하는 것이 ONC 프로젝트가 해결해야 할 무엇보다도 중요한 과제였다. 집계난이도 지표의 점수가 높은 지역은 3개 층화집단 간에 집계편차의 차이를 축소하기 위하여 과다표집되었다. 집계난이도 지표는 1991~2001년의 10년간에 일어난 사생활 보호, 보안의식, 생활양식의 변화, 거주형태의 다양화, 공동체 의식의 쇠퇴 등 조사환경의 변화에 대응하여 과소집계 유형의 변화를 측정하는데 도움이 되게끔 설계하였다. 그러나 특정지역에서 발생한 대규모의 급격한 사회경제적 변화(예: 우범지대와 불량주택지역의 재개발사업)는 1991년 센서스 데이터베이스의 변수정보를 가공하여 구성한 집계난이도 지표를 그다지 설득력이 없는 것으로 만들어버렸다.

12개의 연령구간(0-4세, 5-9세, 10-14세, 15-19세, 20-24세, 25-29세, 30-34세, 35-39세, 40-44세, 45-49세, 50-54세, 55-59세, 60-64세, 65-69세, 70-74세, 75-79세, 80-84세, 85세 이상)으로 구분한다. 이러한 추출작업의 층화작업은 센서스 추정 및 보정을 위한 작업과 직접적으로 관련되어 있다.

3) 표본규모

표본규모는 센서스 범위조사의 조사설계에서 인원배치나 예산책정과 관련 하여 핵심적 관심영역이 되었다. 표본설계 작업에서 개별 추정지역 총인구에 대하여, 상당수준의 정확성, 곧 1% 이하의 상대적 표준오차(Relative Standard Error)(추정치의 표준오차를 추정치로 나누어 백분율로 표시한 수치)만을 허용하는 추정치를 얻기 위해서는 36만 가구에 해당하는 대략 4,000~5,000개 정도의 전국을 대표하는 센서스 조사구의 표본이 필요하다는 것이었다.

집계난이도 지표의 층화집단에 대하여 추출률을 결정하기 위하여 전국 수준의 표본설계에 착수하였으며, 그것을 바탕으로 센서스 조사구에 대하여 집계난이도 지표의 층화집단별로 총 표본수를 균등하게 배분하였다. 또 위에서 설명한 추출률은 표본규모를 결정하기 위하여, 112개의 추정지역 안에 있는 센서스 조사구의 가구에 대해서도 그대로 적용되었다. 따라서, 지표점수별 층화집단에 대한 추출률은 지표점수가 1점인 집단(집계가 쉬운 지역)을 3.4%, 지표점수가 2점인 집단(집계가 쉽지도 어렵지도 않은 중간지역)을 3.7%, 지표점수가 3점인 집단(집계가 어려운 지역)을 4.5%로 설정하였다.

영국의 잉글랜드와 웨일즈 지방에서 101개 추정지역의 모든 센서스 조사구(CED)에 대하여 3~5개의 우편번호가 선정되었으며, 그 결과 16,400개의 우편번호를 전체표본으로 하여 우편번호 안의 32만 가구를 조사대상으로 하는 대규모 표본조사의 골격이 갖추어지게 되었다. 한편, 영국의 스코틀랜드에서 7개의 추정지역에 대하여 2,400개의 우편번호에서 4만 가구를 추출하였으며, 북아일랜드에서는 4개의 추정지역에 대하여 700개의 우편번호에 대하여 1만 가구를 현지실사하는 별도의 센서스 범위조사가 실시되었다.

4) 조사방법론

센서스 범위조사의 현장실사는 2001년 센서스 본조사가 실시되고 3.5주 후에 실시되었다. 이것은 조사표를 토대로 하는 면접자 조사로서, 나중의 매칭 작업에 꼭 필요한 정보를 획득하고 응답자의 시간부담을 축소하기 위하여 응답자의 가구특성이나 인구학적 속성 등 센서스 본조사의 핵심적 질문만을 반복조사하였다. 조사방법은 가능한 한 조사대상자의 응답률을 높이는데 초점이 맞추어졌으며, 조사대상자와 접촉률을 극대화하기 위하여 개별주소지

를 직접 방문하는 방안을 채택하였다. 면접원의 작업분량을 조정하기 위하여, 인접 우편번호를 하나로 모아서, 1인당 90~200가구의 면접이 가능하도록 하였다. ONC 프로젝트는 센서스 범위조사의 조사원에게는 100% 완전집계가 가능하도록 종합적 면접훈련을 실시하는데도 관심을 기울였다. 센서스 범위조사의 조사원들은 대부분 현장조사 경험이 미숙한 사람들로 구성되었기 때문에, 영국 통계청에 소속된 80명 정도의 숙련된 조사원을 현지에 파견하여 조사표 작성에 협조를 하지 않는 우편번호에서 활동하는 조사원들을 도와주도록 하였다.

센서스 범위조사의 대원칙은 센서스 본조사와는 독립적으로 실시된다는 것이었다. 이것은 관련 쟁점을 추정작업에서 자세히 설명하겠지만, ONC 프로젝트에서 센서스 본조사의 과소집계를 추정하는데 필요한 최소요건이기도 하였다. 두 조사 간에 독립성이 없는 경우에는 센서스 본조사가 누락한 개인과 가구를 센서스 범위조사도 파악하지 못할 개연성이 높다. 현실적으로, 두 조사의 독립성을 극대화하기 위하여 센서스 범위조사 조사원에게 해당지역의 센서스 본조사의 가구명부에 대한 정보를 전혀 제공하지 않았다. 조사원들에게 센서스 범위조사의 현지실사를 목적으로 작성된 우편번호의 지도만을 제공하였으며, 그들은 우편번호 안에 있는 모든 주소자의 가구들을 면접하여, 해당지역을 사실상 독자적으로 재집계하려고 노력하였다. 조사원들은 우편번호 안의 가구명부를 가지고 있지 않았기 때문에, 우편번호에 포함된 가구를 모두 면접하였는가를 확인하기 위하여, 조사원들에게 우편번호의 지역경계를 최종 점검하도록 요구하였다.

추가적으로 센서스 범위조사가 센서스 본조사의 결과에 오염효과를 일으키지 않도록 하기 위하여, 센서스 범위조사가 센서스 본조사의 현지실사에서 수집된 정보를 이용하지 않도록 하기 위하여 모든 노력을 경주하였다.

- 표본으로 추출된 우편번호들의 리스트는 기밀서류로 취급하고, 그것이 무단 유출된 지역은 우편번호의 리스트를 다시 추출하도록 하였다.
- 센서스 본조사 관리자는 센서스 범위조사의 집계요원으로 참가할 수 없도록 하였다.
- 센서스 범위조사 관리자는 센서스 본조사의 집계요원으로 참가할 수 없도록 하였다.
- 센서스 본조사 집계요원이 센서스 범위조사의 집계요원으로 활동하는 것을 허용했지만 그들이 집계했던 동일지역에서 집계요원으로 활동하는 것을 허용하지 않았으며 센서스 본조사가 종료될 까지 그들이 배정받게

될 우편번호에 대하여 사전고지를 하지 않았다.

센서스 범위조사의 추정 응답률은 잉글랜드와 웨일즈에서는 확인 가능한 거주가구의 89%였으며, 추후에 우편으로 반송된 조사표를 포함하면, 실제응답률이 91%에 도달한 것으로 확인되었다. 한편 스코틀랜드와 북아일랜드의 응답률은 각각 95%와 92%에 이르렀다.

H. 미국의 센서스 표본조사-American Community Survey

미국의 American Community Survey는 연방헌법상 폐기 처분할 수 없는 short-form의 인구센서스와 함께 센서스통계의 빈번한 생산주기를 목적으로 실시되는 통계조사이다. 이 통계조사는 2015년 등록센서스를 통계청이 추진한 이후, 우리나라에서도 본격적으로 추진해 볼만한 가치가 있는 센서스 표본조사이기 때문에, 본 절에서는 미국의 2010년 센서스에서 long form을 대체하게 되어 있으며, 독자적인 조사체계의 의하여 실시되고 있는 American Community Survey 프로젝트의 조사방식을 중심으로 하여, 조사계획을 개관하기로 한다.

1. 통계조사의 개관

American Community Survey는 미국 인구주택센서스의 long form과 내용은 유사하지만, 설계는 다르다. 이 조사는 미국에서 센서스 표본조사로서 작성한 소지역통계의 연간자료를 생산하기 위하여 개발되었으며, 일련의 원차조사 표본을 사용하는데, 이것이 바로 인구주택센서스의 long form 조사표가 센서스 기준시점에 1회만 실시되는 조사와 크게 다른 점이라고 할 수 있다. 소지역 통계를 생산하기 위해서 처음에는 5년간의 표본이 필요한데 5년간의 자료가 축적되고 나면 매년 소지역에 대한 통계가 생산된다. 좀 더 규모가 큰 지역을 대상으로는 3년 자료와 연차자료를 생산하게 되며 과거의 인구주택센서스와 마찬가지로 American Community Survey의 조사대상은 일반거처(ordinary dwellings) 및 집단거주시설(group quarters)에 거주하는 사람 모두 포함한다. American Community Survey는 미국 및 푸에르토리코에서 실시되고 있으며, 푸에르토리코에서는 푸에르토리코 지역사회조사라고 불려진다(U.S. Census Bureau, 2004, National Research Council, 2006).

인구 및 주택의 특성에 관한 상세한 정보를 수집하는 대안으로서 연속적으로 자료를 수집하는 방법이 계속 연구되어 왔었지만 10년 주기의 센서스의 실질적인 대안으로는 생각되지 않았다. 그러나 1990년경에 연방정부, 주정부, 지방정부 및 민간부문의 현 시점에서 국가적으로 일관성 있는 자료에 대한 수요로 인해 정부에서는 사회통계, 경제통계 자료를 10년마다 한 번씩 수집하는 대신에 10년에 걸쳐 연속적으로 수집하는 가능성에 대해 검토하기 시작했다. 현 시점의 자료를 얻을 수 있다는 장점 및 비용절감, 계획수립, 센서스 포괄범위의 개선, 효율적인 수행 등의 예상되는 장점 때문에 센서스국은

American Community Survey를 2000년에 시작하기로 하고 준비하였으나, 신 계획의 표본설계, 조사방법, 통계작성의 미묘한 점을 이해할 필요성 때문에 American Community Survey의 시행은 2000년 센서스 이후로 연기되었다. 이후 추가의 테스트, 이해관련자들에 대한 접촉, 미국통계학회, 미국인구학회 등의 주요 이용자와의 지속적인 접촉 후에 센서스국은 가구단위에 대해서는 2005년에, 집단구역에 대해서는 2006년에 American Community Survey 본조사를 시행하게 되었다

<그림 1> 미국 Census Bureau의 American Community Survey 안내책자



American Community Survey의 역사는 4단계로 나눌 수 있다. 연속적 측정에 의한 자료수집 계획·제안 단계(1990-1993년), 몇 개의 지역을 대상으로 초기 모형 개발단계(1994-1999년), 전국을 대상으로 대규모 통계조사를 실시하고 전국, 주 및 대규모 지역들에 대한 보고서가 작성되는 시범조사 단계(demonstration stage)(2000-2004년), 연간 약 300만 주소(미국) 및 약 3만 6천 주소(푸에르토리코)의 거처단위를 대상으로 하는 완전시행 단계(2005년 1월~

현재까지)에로 구분할 수 있다.

Leslie Kish (1990, 1998, 1999)는 표본설계의 대가로서, 1940년대에 미국 센서스국에서 근무한 적이 있는 헝가리 계 이민으로서, 1981년에 새로운 센서스의 개념으로 '순환표본설계'라는 개념을 소개했다. Kish가 연구를 할 무렵에 센서스국에서도 센서스 자료의 생산주기와 공표주기를 단축해야 할 필요성에 대하여 합의가 이루어지기 시작했다. 1985년 미국의 연방하원에서는 중간연도 센서스(mid-decade census)를 승인했지만, 관련 예산을 배정하지 않는 불상사가 일어났다. 1990년대 초 연방하원은 10년 주기의 센서스에 대한 대안에 대해 다시 관심을 보이기 시작했으며, 센서스국은 Leslie Kish의 연구를 바탕으로 두고 1990년대 중반에 연속 측정(continuous measurement) 방법의 개발에 착수했다.

센서스국은 세분화된 표본자료 수집의 대안으로 연속 측정에 대한 연구계획을 마련하였고 연속측정을 위해 세 가지 초기모형을 개발했다. 실행 가능성 및 기술적인 가능성, 정책문제, 비용, 장점에 대한 직원들의 평가에 따라 센서스국은 한 개의 초기모형을 선택하였고 개발을 계속하였다.

초기모형 개발에 대해 여러 가지 결정이 이루어졌는데 비용절감을 위해서는 우편조사를 실시하고, 전화조사와 면접조사로 우편조사를 보완해야 한다는 결론에 도달하였다. 또 초기모형의 지역적 구성, 표본 추출률, 가중치 설정을 위한 연차별추계인구의 이용방안에 대한 결정도 이루어졌다.

센서스 long-form 표본조사와 같은 수준의 신뢰도를 획득할 수 있는 소지역에 대한 5년치의 자료를 생산하기 위해서 매월 5십만 크기의 거처단위 표본이 제안되었다. 그러나 이 표본의 크기를 감당하기에는 비용이 너무 많이 들게 되었다. 하지만 비표본오차(non-sampling error)를 줄임으로 해서, 25만 개의 표본으로 수용할만한 신뢰수준을 얻을 수 있을 것으로 결정되었다.

1994년에 전담 부서가 신설됨으로 해서 개발단계가 시작되었다. 전담 부서는 조사 초기모형의 개발을 계속했고 아래의 설계요소(design elements)들이 조사의 기초로서 하였다.

- ① 상호 독립적인 월차 표본을 이용하여 1년에 걸쳐 연속적으로 자료를 수집해야한다.
- ② 세 가지의 자료획득방법이 이용되어야 한다(우편조사, 무응답에 대한 전화조사, 무응답에 대한 면접조사)
- ③ 거처 점유 상태 및 다른 많은 특성을 결정하는 조사기준일은 자료가 수집되는 바로 그날이어야 한다. 어떤 항목들에 대한 조사대상 기간은 더 길어질 수 있다(예: 지난주, 지난 12개월)

④ American Community Survey 추정치들은 센서스 중간연도의 인구 및 거주 추정치에 조정되어야 한다.

⑤ 모든 추정치들은 일정기간(예: 1년)에 걸쳐 조사된 월간 표본조사에서 수집된 자료를 합쳐서 생산되어야 한다.

초기의 개발보고서는 다양한 매체로 공개되어 왔다. 연속추정계열로 알려진 20개의 보고서가 개발에 있어 아주 중요한 것들이다. 이 보고서들은 1993년에 시작되었고 최종 초기모형에 반영되었다. 연속추정 계획은 1995년에 미국통계학회가 주관하는 JSA(Joint Statistical Meetings)에 공식적으로 소개되었다. Love, Dalzell과 Alexander(1995)가 이런 조사에 필요한 전제조건에 대해서 소개하였고 Dawson, Sebold, Love과 Weidman(1995)이 전화조사에 의한 자료수집 가능성에 관해 보고서를 작성했다. 당시에 연속추정 자료로 평가된 가능한 수정안들에 대해서도 토의가 있었다.

1995년 11월에 American Community Survey의 시험조사가 4개 지역(Rockland County, NY; Brevard County, Fl; Multnomah County, Or; and Fulton County, PA)에서 개시되었다. 1996년 11월에는 시험조사가 확장되어 다양한 지리적·인구학적 특성을 가진 지역을 포함하였다. 이러한 지역에는 Harris County, TX; Fort Bend County, TX; Douglas County, NE; Franklin County, OH; Oteto County, NM이 있다. 이 시험조사는 방법과 절차의 정당성 평가와 장래 시행을 위해 조사비용을 추정하기 위하여 실시한 것이었다. 시험조사 결과에 따라 초기모형 설계를 수정하였고 추가적인 연구영역을 확인하였다. 소지역추정, 추정방법, 무응답 추적, American Community Survey 시험조사에서 가중치 부여, 항목 무응답, 무응답률 및 농촌지역 자료의 질 등 여러 분야에 걸친 더 많은 연구가 있었다.

시험조사는 계속되었고 1998년에는 South Carolina의 두개 군 및 Florida의 1개 군이 조사지역으로 추가 선정되었다. South Carolina의 두개 군은 1998년 센서스 시범조사 결과와 비교하기 위해서 1998년에만 포함되었다. 1999년에는 26개 주의 36개 군으로 시험조사가 확장되었는데, 이 지역은 전국 대표 표본으로 구성되지 않았고 군의 인구 규모, 조사의 난이도 및 1990-1995년간의 인구성장을 고려하여 선정되었다. 지리적인 다양성도 고려요인이 되었으며, 인종, 민족, 계절인구, 이동 취업자, 인디언 보호구역(Indian reservation), 경제상태의 변동, 주산업(main industry) 및 주직업(main occupation)이 고려되었다. American Community Survey 프로그램을 평가하고 개선하는데 적극적으로 참여할 수 있는 사용자들이 있는 지역을 추가로 선정하였다. 운영 시험의 결과에 근거하여 초기모형의 개선이 이루어졌고 추가적인 연구영역

이 확인되었다.

2. 표본 추출틀의 개발

미국 센서스국은 American Community Survey의 주소, 기타 인구조사의 표집 틀 및 10년 주기의 인구센서스의 주소로 사용하는 국가 마스터 주소 파일(MAF; National Master Address File)을 관리한다. MAF에는 미국 및 푸에르토리코의 일반거처, 집단거주시설 및 선택된 비거주시설(공공, 개인 및 상업용)에 대한 우편주소, 위치주소, 지리코드 및 거주구역에 관한 기타 정보가 담겨 있다. MAF는 TIGER(Topologically Integrated Geographic Encoding and Referencing) 파일과 매칭되어 있으며 TIGER는 디지털 센서스 지도 및 관련 특성을 표현할 수 있는 데이터베이스로서 자동화된 주소를 특정 지역에 부여하는 기능(geo-coding)을 보유하고 있다.

초기 MAF는 2000년 센서스를 위해 개발되었는데 1990년 주소관리파일, 미국 우정공사의 배달경로파일(Delivery Sequence File; DSF), 현장조사 확인 및 지방정부가 제공한 주소로 구성되어 있다. MAF 보완은 DSF를 이용하고, Community Address Updating System(CAUS) 등을 이용하여 자동화, 수작업, 현장 확인을 통해 이루어진다.

MAF의 각 레코드에는 다음과 같은 정보의 전부 또는 일부를 담고 있다.

- 거처의 상태: 현존, 철거, 건축 중 등
- 거주용 거처 또는 비(非) 거주용 건물
- 위도 및 경도
- 각종 지리정보(주, 카운티, 트랙트, 블록 등)

MAF 주소정보에는 도시형 주소 및 비도시형 주소를 모두 담고 있다.

- 도시형 주소: 예) 201 Main Street, Anytown, MA 01977
 - 전체 주소의 약 94.4% 차지
- 비(非) 도시형 주소: 예) 우편함 주소(PO box format) 등

3. 표본설계 및 표본선정

American Community Survey의 표본은 수도 워싱턴을 포함한 미국 내의 3,141개 군과 푸에르토리코의 78개 기초자치단체에서 각각 추출된다. 일반거

처와 집단거주시설에 대한 표본은 분리 추출된다. 일반거처에 대한 첫 번째 완전이행 표본은 2005년에 선정되었다. 매년 미국 내 약 3백만 일반거처와 푸에르토리코 내 약 3만6천 일반거처가 표본으로 선정된다. 집단거주시설에 대한 첫 번째 완전이행표본은 2006년에 선정되었으며 연간 집단거주시설 내의 약 2.5%의 사람들이 American Community Survey 표본에 포함된다. 표본의 신뢰도를 유사하게 하기 위해서 인구 규모가 큰 지역의 표본추출률은 인구 규모가 작은 지역보다 상대적으로 작으며 200가구(약 500명) 미만의 지역은 연간 10%를 표본으로 선정한다. 또한 우편 응답률이 낮은 지역의 표본추출률은 상대적으로 더 크게 한다. 5년 추정치의 표본추출오차는 전통적인 센서스 표본조사 보다 1/3 이상 클 것으로 예상이 되지만 전문조사원 채용 및 자료 시의성으로 인한 비표본오차(non-sampling error)의 감소로 인해 상쇄될 것으로 보고 있다.

일반거처에 대한 표본은 주표본과 추가표본으로 나누어진다. 주표본은 조사 전년도 8-9월에 선정되며 추가표본은 조사연도의 1-2월에 선정한다. 주표본의 표본추출률은 최근의 MAF가 이용되고 그 이후에 신축되는 거처에 대해서는 추가표본을 선정한다. 주표본으로 선정된 거처에 대해서는 12개월에 걸쳐 배분되고, 추가표본은 4월부터 12월까지 배분된다.

각 거처가 5년 동안 2번 이상 표본으로 선정되는 것을 방지하기 위해서 표본추출률은 5개의 부표본들로 나뉘며 각 거처는 이 5개의 부표본들에 랜덤하게 할당된다. 이 5개의 하위표본들 중 1개의 표본추출률을 1년간 사용하도록 되어 있다..

선택된 1개의 부표본들을 이용하여 각 카운티별로 여러 가지 특성에 따라 주소지별로 정렬한 다음 1년 동안에 사용할 표본 거처를 계통 추출한다. 1년간 사용될 표본거처들은 다시 12등분하여 조사할 월에 할당된다.

4. 조사방법

표본으로 추출된 각 거처단위 주소는 특정 월에 배정되어 조사표를 받게 된다. 우편 발송에 적합하지 않는 주소는 발송불가주소로 분류된다. 거처와 거처내의 거주자들에 대한 조사는 배정된 달부터 다음 두 달간에 걸쳐 작성 가능하다. 우편 발송된 주소에서 응답이 없고 전화번호가 있으면 CATI 담당 직원을 통해 다음 달에 전화조사를 실시한다. CATI조사에서도 무응답 가구 주소에 대해서는 이 중 표본을 선정하여 세 번째 달에 CAPI조사를 실시한다. 각 집단거주시설 표본은 특정 월에 배정되고 면접조사를 통해서 6주간에

걸쳐 조사를 하게 된다.

①) 우편조사

우편 발송이 가능한 약 95%의 가구에 대해서 매월 조사표류를 우편으로 발송하는데 ① 조사협조 공문 발송 ② 조사표류 발송 ③ 무응답 가구에 대해 조사독촉 공문 발송 ④ 조사표류 재 발송의 과정을 거친다. 우편조사에 대한 응답률은 약 50%로서 센서스 표본조사 응답률과 비슷하다.

②) CATI 조사

무응답 가구에 대해서는 우편조사 다음달에 3개의 콜 센터를 통해서 전화 조사를 약 4주간에 걸쳐서 실시하며 CATI 조사의 응답률은 약 60%이다.

③) CAPI 조사

우편조사와 CATI 조사에서 무응답 가구 및 우편발송불가 주소지에 대해서는 CATI 조사 다음 달에 면접조사를 하게 된다. 면접조사 대상 가구는 약 1/3을 표본으로 선정한다. 무응답 가구에 대해 표집하여 면접조사를 실시하는 이유는 면접조사의 비용이 우편조사 비용의 약 10배가 들기 때문이다. 면접조사는 12개 지역사무소를 통해서 실시하며 조사원은 임시직이 아닌 정규 직원이다. 또한 3년 이상의 조사 경험자를 현장관리자로 이용하며 면접조사의 응답률은 93% 이상이다.

④) 전체 응답률

전체 응답률은 95% 이상이며, 표본설계의 복잡성 때문에 전체 응답률 계산에는 가중치가 이용된다.

5. 자료처리

1년간 수집된 자료를 누적하여 당해 연도의 추정치를 계산한다. 자료처리에는 부호화(coding), 편집(editing) 및 대체-보정(imputation)의 3가지 주요 과정이 있다.

〈표 1〉 2001년 American Community Survey의 응답률

구 분	응 답 률
우편조사	51.1%
CATI	9.2%
면접조사	36.4%
무응답	3.3%

부호화는 자동화 방식의 이용 및 수기로 진행하는데 인종, 사용언어, 일하는 장소, 산업, 직업 등의 부호화 작업이 있다.

편집 과정에는 일차적으로 응답 조사표와 무응답 조사표로 분리하고 무응답으로 분류된 조사표는 버리고 나중에 가중치로 조정한다. 응답 조사표와 무응답 조사표는 조사표의 완성도에 따라 구분한다. 응답 조사표로 분류된 조사표에 대해서는 항목간의 논리가 맞지 않거나 무응답 항목이 있는 경우, 자동화된 시스템을 이용하여 찾아내어 보정한다..

보정방법은 크게 부여(assignment)와 할당(allocation)으로 나눌 수 있다. 부여는 규칙에 근거한 것으로 해당 조사표 내의 다른 항목을 이용한다. 예를 들면 사람의 이름을 사용하여 성별을 부여하는 것이다. 할당은 다른 조사표의 내용을 사용하여 대체하는 것으로 hot-deck 방법을 사용하여 이웃의 정보(neighborhood information)를 사용한다. American Community Survey의 대체율을 2000년 센서스의 대체율과 비교해 보면 매우 낮은 것을 알 수 있다.

가중치 부여는 세단계로 나눌 수 있다. 첫 번째는 표본설계시의 표본추출률이며, 이 표본추출률은 무응답률을 감안하여 조정하고, 추계인구를 이용하여 마지막으로 조정한다. 추계인구는 과거의 센서스 자료를 출생, 사망 등의 행정자료를 사용하여 보정한다. 추정치는 연간 추정치, 3년 추정치 및 5년 추정치로 분류된다. 월간 표본조사를 누적하여 연간 추정치를 계산한다. 연간 추정치는 현재의 공표 지역 단위로 추정되며, 지역의 경계가 바뀌면 바뀐 지역의 추정치를 계산한다. 3년 추정치 및 5년 추정치는 연간 추정치들의 평균으로 계산된다. 중간 값의 계산은 연간 평균치의 중간 값을 사용하지 않고 전체 데이터의 중간 값으로 한다.

6. 자료공표

자료공표는 인구 6만5천 이상의 지역을 대상으로 하며 조사 다음 해 8월에 공표한다. 인구 6만5천명 이상의 군 지역은 전체 인구의 82.5%를 차지한다.

3년 추정치의 공표는 인구 2만 이상의 지역을 대상으로 하며(2005~ 2007년의 자료는 2008년 8월에 공표), 인구 2만 이상인 군 지역은 전체 인구의 95.5%를 차지한다. 5년 추정치의 공표는 인구 2만 미만의 지역(2005~2009년의 자료는 2010년 8월에 공표)을 대상으로 한다.

표본조사의 공표지역은 우편번호, 학군, 주 행정구역, 인디언 보호구역 - PUMAs(2000년 센서스 구역 중 인구 10만 이상), 센서스 트랙트) 등으로 분류할 수 있다. American Community Survey는 표본조사이기 때문에 연차별 자료에 대해서는 점 추정치 및 신뢰구간을 발표하며 90% 신뢰구간을 사용한다.

7. 기타사항

American Community Survey는 센서스 long form 표본조사를 대체하는 것이기 때문에, 미국 연방법에 의하여 표본대상으로 선정되는 모든 시민은 응답의무가 있다. 조사응답에 대해서는 \$100~\$5,000까지의 벌금을 부여할 수 있다. American Community Survey의 조사항목 수는 60개로 센서스 표본조사의 조사항목 수와 같으며(조사항목은 <그림 2>를 참고할 것), 조사표 작성 시간은 약 38분이 소요된다. American Community Survey의 총비용은 10년간 누적하면 전통적인 센서스 비용과 거의 비슷하며 2005년 American Community Survey의 총비용은 1억4천6백만 달러이다.

<그림 2> American Community Survey의 조사항목

행정목적(센서스국 이용 항목)	경제항목(주택)
1. 성명	1. 부지내 사무실 입지여부
2. 연락처	2. 전기수도사용료
3. 주소	3. 콘도미니엄 사용료
4. 기구원수	4. 보험료
5. 면접일자	5. 이동식주택 비용
	6. 주택담보액(mortgage)
기본항목(인구)	7. 임대료
1. 연령	8. 주택소유여부
2. 성별	9. 자산가치
3. 히스패닉 여부	
4. 인종/종족	주택항목(주택)
5. 가구주와의 관계	1. 면적
	2. 농업판매
사회항목(인구)	3. 침실 수
1. 조상	4. 난방연료
2. 시민권자 여부	5. 부역시설
3. 활동제약(장애) 상태	6. 배관시설
4. 최종학력	7. 전화서비스 가용여부
5. 출산력	8. 총방수
6. 조부모의 돌봄노동 제공여부	9. 부지내 단위건축물 개수
7. 집에서 사용하는 언어	10. 가용 자동차수
8. 혼인력	11. 이사연도
9. 혼인상태	12. 건축연도
10. 군복무기간	
11. 출생지	
12. 재학여부	
13. 1년전 거주지	
14. 학부 전공분야	
15. 퇴역군인 여부	
16. 퇴역군인서비스 관련 활동제약 등급	
17. 입국일자	

I. 프랑스의 순환센서스 표본조사

가. 프랑스 센서스의 역사

프랑스에서는 나폴레옹의 동생 루시앵(Lucien)에 의해 1801년 처음으로 인구센서스가 실시되었다. 나폴레옹이 인구센서스를 실시한 주된 목적은 두 가지다. 첫째는 군사목적에 의한 것으로 전쟁을 위해 청년을 징병할 필요가 있어 각 지역의 청년수를 파악하는 것이고, 또 하나의 목적은 원로원 및 입법원의 의원 수는 나폴레옹헌법에 의해 지역의 인구수에 맞추어 정하도록 했기 때문에 각 지역별 인구수를 확정할 필요가 있었다. 인구센서스는 이 후 5년마다 실시되었다. 그러나 1939년의 2차 세계대전 발발 후 인구센서스는 부정기적으로 1946, 1954, 1962, 1975, 1982, 1990, 1999년에 실시되었다. 프랑스는 1999년 센서스 이후 전통적인 인구센서스를 포기하기로 결정하였는데 그 주된 이유는 센서스의 주기가 길어지고, 예산문제 및 지방자치제의 확산으로 인한 시의성 있는 소지역 통계에 대한 수요가 늘어났기 때문이다. 이러한 문제에 대응하기 위하여 프랑스의 순환센서스는 탄생하게 되었다.

나. 프랑스 순환센서스

1) 개관

2004년에 프랑스는 순환센서스라는 새로운 센서스 방법을 도입하였다(INSEE, 2004, 2006). 전통적인 센서스의 대안으로서 특정 시점에서 전국을 포괄하는 센서스를 실시하는 대신에 연속적인 표본조사를 실시하여 장기간에 걸쳐 전국을 포괄하는 방법을 순환센서스라고 한다. 표본조사를 하는 방식은 전국을 지역별로 나누어 돌아가면서 각 지역을 전수 조사하는 방식과 모든 지역에서 표본을 뽑아 조사하는 방법 등 여러 가지가 있을 수 있다. 순환센서스는 이런 식으로 매년 표본 조사된 자료를 누적하여 추정하는 방법을 사용한다(Kish, 1990, 1998, 1999). 현재까지는 프랑스가 전 세계에서 유일하게 순환센서스를 실시하는 나라이다.

프랑스에서 순환센서스를 도입한 이유는 두 가지가 있다. 첫 번째는 보다 최근의, 보다 정기적인 정보에 대한 수요의 증가이며, 두 번째는 전통적인 센서스에 관련되어 왔던 예산 및 인력 부담 문제를 해결하기 위해서이다

(INSEE, 2004, 2006, 西村善博, 2007a, 2007b).

시의성이 있는 자료의 필요성은 프랑스 통계청(INSEE)의 통계이용자들에게 의해서 강력하게 표명되어왔다. 센서스의 주기가 점점 더 길어짐(지난 두 센서스 간 9년)에 따라 사회나 가족의 변화든 도시주변에 정착하는 경향이든, 도시 개발이든 간에 프랑스 인구 변화를 추적하는 것이 점점 더 어렵게 되어가고 있으며 지방 분권화에 의해 권력의 점진적인 위임 때문에 센서스 자료에 대해 지역사회들은 더욱 관심을 기울이게 되었다. 예를 들면, 기초자치단체 또는 기초자치단체 연합체(municipal federation)들은 도시 교통에 대해 새로이 책임을 맡게 되었으며, 사회복지관련 부서들은 중·고등학교에 대한 투자 및 운영, 직업훈련, 지역개발계획 등의 분야에서 더욱 많은 부분을 담당하게 되었다.

프랑스는 1999년 이전까지는 전통적인 센서스를 실시하였지만 센서스 실시에 대해 헌법상에 명문 규정이 없어 센서스를 실시할 때 센서스 법률을 제정하여야 했다. 또한 프랑스 선거를 고려해야 했기 때문에 보통 7~9년 간격으로 센서스를 실시하였다. 조사주기가 긴 것 외에도 예산이나 인력 문제를 해결하기가 쉽지 않았다. 1997년에 계획된 센서스가 1999년으로 연기된 이유도 예산문제 때문이었다.

2) 표본추출틀 개발

각 대규모¹¹⁾ 기초자치단체의 표본추출틀은 관련된 건축물등록부(RIL; located building register)이며, 건축물등록부는 주거용, 행정용, 산업용, 상업용 건축물 등의 모든 건물의 목록이며 지리정보시스템을 이용하여 확인 가능하다. 건축물등록부는 1999년 센서스를 기초로 편집되었고 건축허가, 지방세, 우편주소 등의 행정자료를 사용하여 계속 보완된다. 건물등록부는 매년 기초자치단체가 점검하며 INSEE가 마지막으로 확인한다.

3) 표본설계

프랑스의 기초자치단체는 그 수도 많고 매우 다양하기 때문에 특별한 표본추출방법을 개발하였다. 프랑스에는 약 3만 7천개의 기초자치단체가 있고 그 중 절반은 인구가 400명 미만이다.

11) 인구 1만 명 이상의 자치단체를 말한다.

인구 1만 명 미만의 모든 기초자치단체에 대해서는 센서스가 실시된다. 해마다 기초자치단체 1/5에 대해서 센서스가 실시된다. 인구 만 명 미만의 35,750개의 기초자치단체는 5개의 집단으로 나누어진다. 이 기초자치단체들은 전체 인구의 절반을 차지한다. 매년 한 개의 집단에 속하는 모든 기초자치단체들에 대해서 센서스를 실시한다. 첫 번째 집단에 속하는 기초자치단체들은 5년 후에 다시 조사된다.

5개의 집단은 인구(인구수, 성별, 연령계층), 기타 주거 관련[거처(dwelling) 수, 주 거주지(principal residence) 수] 변수 등 약 10가지 기준에 의해 균형을 맞추어 형성된다. 균형은 전국단위 및 프랑스의 26개 광역자치단체 단위에서도 이루어진다.

인구 1만 명 이상의 900개 기초자치단체에서도 매년 센서스 표본조사가 이루어지지만 일부분만 조사된다. 매년 각 기초자치단체에서 8%의 거처가 조사된다. 따라서 5년 후에는 40%의 인구가 조사된다. 이것은 기초자치단체 및 기초자치단체의 각 구역에 대해 안정적인 정보를 보장하기에 충분한 표본의 크기이다.

각 기초자치단체의 센서스 표본은 다음의 표집 방법을 사용한 주거지의 부분집합에 기초를 둔다.

- (a) 집락 효과를 피하기 위해서 기초자치단체에서 가장 큰 주소들은 전부 조사한다. 이 주소들은 5개의 연간 집단으로 구분된다.
- (b) 새로운 주소들은 전부 조사한다. 표본추출에 필요한 정보가 없기 때문이다. 이 새로운 주소들도 5개의 집단으로 구분한다.
- (c) 기타 주소들은 인구 또는 주거 수를 기준으로 하여, 크기가 같은 5개의 집단으로 구분한다. 각 집단은 기초자치단체에서 골고루 분포되도록 한다.¹²⁾ 매년 조사대상이 되는 표본주소들은 올해의 집단에서 선정되며 조사될 모든 주소들은 기초자치단체의 전체 거처의 약 8%이다.

이런 식으로, 포괄적인 센서스가 매년 인구 만 명 미만의 기초자치단체의 1/5, 인구 만 명 이상의 기초자치단체 인구의 8%에 대해서 실시된다. 즉 4백 5십만 거처와 9백만 인구가 매년 조사되는 셈이다. 5년 후에는 소규모 기초자치단체 전부와 대규모 기초자치단체에 거주하는 인구의 40%가 조사되게 되며, 전체적으로 70%의 인구가 5년 주기 동안 조사된다.

마지막으로 집단가구에 거주하는 집단, 이동주택에 사는 사람들, 그리고 노

12) 따라서 동일한 거리(street)에 있는 주소들이 서로 다른 집단에 속할 수 있다.

숙자들은 대규모로 조사된다. 이런 식으로 프랑스 영역에 살고 있는 모든 개인들을 조사하게 된다.

4) 조사방법

센서스 표본조사는 매년 동일한 날에 관련된 모든 기초자치단체에서 시행된다¹³⁾. 자료수집 방법(protocol)은 조사원이 질문지를 전달하고 수거하는 방식이다. 표집방법은 혁신적으로 변화했지만 자료수집의 경우는 일반적인 센서스에서 사용되는 방식을 그대로 따른다. 즉 조사원이 전달하고 수거한다. 각 가구는 2가지 질문지를 받게 되는데, 그중 하나는 가구에 관한 것이고 다른 하나는 가구원에 관한 질문지이다. 질문지는 과거의 센서스에 사용된 질문지와 큰 차이가 나지 않는다. 가구에 관한 질문지는 가구원의 목록 및 주거의 특징과 기준에 관한 약 14개의 질문과 자동차 보유 대수로 구성되어 있다. 가구원에 관한 질문지는 나이, 성별, 출생지, 국적, 5년 전 거주지, 교육정도, 직업, 근무지 또는 통학지에 관한 25개의 문항으로 구성되어 있다.

조사가 불가능한 경우(부재, 장기간 출타, 응답거부)에는 주 거주지에 대해서 조사원이 불사불능 주택일람표(non-surveyed dwelling form)에 추정 가구원수를 기입하도록 하여 인구수를 조정하는데 사용한다..

5) 자료처리

2008년에 시작되는 세분된 파일(조사된 개인이나 가구당 기록)이 5년간의 조사된 파일을 합침으로써 매년 생산될 것이다. 따라서 일반적인 센서스처럼 교차표 생성(cross-tabulation)이 가능하다. 매년 조사된 9백만 개인 및 4백5십만 거처가 가중치 조정 후에 이 파일에 숫자로 나타난다. 따라서 세분된 파일은 모든 개인을 설명한다.

센서스 표본조사의 집계범위(만 명 미만의 기초자치단체의 경우 100%, 만 명 이상의 기초자치단체의 경우 40%) 및 정보가 개인별로 가중치가 부여된 자료들의 파일로 정리되어 있다는 사실은 아주 세분된 정보가 지리적으로 세분되고 소규모 집단에 대해서 추출될 수 있다는 것을 의미한다. 누적된 파일은 4천5백만 개인과 2천2백만 거처로 구성되어 있기 때문에 상세한 표를 만들 수 있다.

13) 1월 3번째 목요일에 시작함

6) 추정

매년 수집된 정보는 5년 주기의 중간점에서의 특정한 날에 적용하기 위해서 조정된다. 대규모 기초자치단체의 경우에는 5년간 표본의 평균값을 계산한 후 중간년도의 RIL의 거쳐 수에 맞추어 판단한다. 소규모 기초자치단체에 대해 사용되는 방법은 센서스 표본조사와 인구 기준 일간의 내삽 또는 외삽이다. 외삽은 주택세 등록자료를 이용하여 합쳐진다. 주택세 등록자료는 연간 주택 수의 변화를 나타내는 지표의 역할을 한다. 5년 순환을 이용하면, 2년이 초과할 경우에는 내삽/외삽은 불필요하다. 따라서 통계숫자에 대한 신뢰도는 상당히 높다.

인구 1만 명 이상의 기초자치단체에 사용되는 방법은 5년간의 표본에 기초한 순환평균(rolling average)이다. $y-4$ 년에서 y 년의 5개의 표본의 총합에서 거쳐 당 평균 인구수가 계산된다. 이 평균 인구수가 중간기간($y-2$ 년)을 대표한다. 이 평균 인구수에 $y-2$ 년 1월 1일 현재의 주택(단위거처)의 수(RIL을 이용)를 곱하여 기초자치단체의 인구를 구한다.

인구 1만 명 미만의 기초자치단체에 대해서는 y 년 말에 $y-2$ 년 초의 인구를 구하는 것이 필요하다. 인구 1만 명 이상의 기초자치단체와 동일한 기준일을 사용한 정보를 제공하는 것이 중요하기 때문이다. 인구 1만 명 미만의 기초자치단체들은 순환기간 전체에 걸쳐서 조사한다. 즉 $y-4$ 년에 1/5, $y-3$ 년에 1/5, 등으로 조사된다. $y-2$ 년에 조사된 기초자치단체들의 센서스 결과는 그대로 저장된다. 반면에 $y-1$ 년, y 년에 조사된 기초자치단체들의 $y-2$ 년 인구는 센서스 표본조사와 직전에 공표된 자료를 내삽(interpolation)하여 구한다. $y-4$ 년과 $y-3$ 년에 조사된 기초자치단체들에 대해서는 센서스 표본조사와 $y-2$ 년의 결과를 외삽(extrapolation)하여 구한다. 이것은 지방 주택세(local housing-tax) 자료에 기초를 둔다. 지방 주택세는 기초자치단체당 주택수의 변화를 나타내고 주택수의 변화와 거주자수의 변화간의 차이를 고려하여 조정된다. 지난 두 센서스 간의 차이는 인구의 변화를 주기위해서 주택세 자료에 의해서 측정되는 변화에 적용된다.

7) 자료 공표

이 센서스는 매년 각 기초자치단체에 대한 상세한 정보를 제공할 것이며 대규모 기초자치단체의 경우에는 통계 목적으로 합쳐진 대략 2천명의 거주자로 구성된 IRIS로 알려진 구역(blocks)에 대한 상세한 정보를 제공할 것이

다.

프랑스는 3년간의 자료수집이 끝난 시점에서 순환센서스 실시에 관한 첫 번째 중간 보고서를 작성하는 것이 가능해졌으며 반응은 전반적으로 긍정적이었다. 새로운 방법에 대한 기초자치단체 및 전 국민의 폭넓은 지지가 있었다. 2005년과 2006년 센서스에는 2004년 센서스에 반대한 2개의 기초자치단체들을 포함한 모든 기초자치단체들이 참여하였다. 프랑스의 일반 국민 쪽에서는 대규모 센서스보다 관심이 적어져 표본조사 방법에 대한 두려움이 나타나지 않았다. 이는 매우 효과적인 전국 및 지역적인 홍보 때문이었다. 마지막으로 INSEE는 인구 만 명 미만의 모든 기초자치단체에 대한 인구추정치를 처음으로 제공할 수 있게 되었다. 그리고 2004년 하반기 이후로 약 100여 개의 대규모 기초자치단체에 대한 인구추정치도 제공할 수 있게 되었다. 이러한 결과들은 인터넷을 통해 2005년 센서스 시작 이틀 전에 발표되었다. 2005년 하반기에는 2006년에 조사된 인구 만 명 미만의 기초자치단체 및 또 다른 약 100여 개의 대규모 기초자치단체로 확대되었다. 그 결과는 2006년 1월 17일에 공표되었다. 이런 식으로 초기 성과물들을 발표함으로써 새로운 방식의 효용성에 대해 사용자들을 확신시키게 되었다.

8) 프랑스 순환센서스와 전통적 센서스의 비교

순환센서스는 INSEE, 기초자치단체 그리고 응답자들에 부여하는 부담을 완화하는 수단으로서 도입되었다. 5년간의 누적표본추출률은 매우 높기 때문에(대규모 기초자치단체의 경우 40%) 정확한 결과를 보증한다. 이와는 대조적으로 전통적 센서스에서는 대개 약 15% 또는 25%의 표본을 사용한다.

표본의 사용 및 자료수집절차의 연례화(annualization)는 INSSE, 각 기초자치단체의 지역 센서스에 대한 책임이 있는 조정자들 및 자신의 임무에 집중하는 조사원들에게 도움을 주며 더 양질의 자료 수집을 가능하게 한다. 인구 만 명 이상 기초자치단체의 조사원들은 조사해야 할 주소 목록을 가지고 있다. 이것은 매우 중요한 점이다. 왜냐하면 이는 조사원들이 어디로 가야할지 정확히 알고 있다는 것을 의미하기 때문이다. 이 방법은 일반 센서스에 사용되는 지역내 모든 대상처를 모두 찾아 조사하는 방식보다 더 효율적이며, 미조사거처일람표와 결합할 경우 무응답의 문제를 줄이거나 보완하는데 도움이 된다.

순환센서스는 또한 통계작성 작업을 분산함으로써 작성비용을 분산하고 안정적인 노동력 및 예산으로 운영하는 것을 가능하게 한다. 반면에 조사규모

를 줄이기 되면서 충실한 준비가 가능해지고, 센서스 관리의 효율성을 증대시킬 수 있다.

표본조사의 다양한 편차로 인하여 조사결과가 정확성에 약간의 문제가 있기는 하지만 소지역에 대해서는 매년 수정된 정보가 노후화된 정보보다 선호도가 높다는 것이 분명하다. 순환센서스는 10년 주기의 전통적 방식보다 도시재개발 등의 당면한 정책을 수행하는데 그 유용성이 증대된다.

9) 순환센서스의 성공 요인

순환센서스의 성공여부는 여러 가지 요인에 의해 결정되는 바, 기술적인 요인(기초자치단체 하위 수준의 지리정보시스템의 사용가능성 및 보완에 필요한 행정자료의 사용가능성), 방법론적인 요인(특히 표본설계기법), 정치적인 요인(센서스 동반자로서의 기초자치단체와의 광범위한 협의 및 센서스 사용자들과의 협의)이 있다.

기술적으로 표본추출틀이 양질이어야 한다. 표본추출틀을 보완하기 위한 행정자료의 이용이 가능해야 하고 행정자료의 사용에 관한 지침을 세워야 한다.

기초자치단체와의 협동이 가장 중요하다. 왜냐하면 이러한 파일들의 내용이 지방 단위에서 정확히 점검되어야 하기 때문이다. 그리고 현장상태에 대한 지방의 이해가 훨씬 더 높기 때문에 이것은 매우 중요한 일이다. 행정자료가 없는 경우에는 프랑스의 해외에 있는 도(DOM; France's overseas department)의 경우처럼 공무원들을 현장에 보내서 주택을 확인하는 것이 필요하기 때문에 표본추출틀의 보완은 느리고 성가신 작업이 된다.

성공을 위한 또 다른 중요한 요인은 행정자료원의 사용가능성이다. 행정자료를 사용하면 표본추출틀(sampling base)을 보완할 수 있으며 필요할 경우에 조사 결과를 평가할 수 있을 것이다.

표본을 최적화하고, 조사결과를 외삽(extrapolation)하고, 자료수집 시기와 관계없이 동시성을 가지게 하기 위한 조사와 모형화에 있어 방법론적인 전문성은 필수적이다. 균형표본기법(balanced-sample technique)과 기능에 대한 지식은 표본최적화 -즉, 예산수립 및 추정치의 질- 에 있어 극히 중요하다.

순환센서스의 경우에는 센서스 업무가 연례화되는 만큼, 일정액의 예산이 충분히 확보되어야 한다. 순환센서스는 센서스 비용을 줄이지 않고 다만 예산을 장기간에 걸쳐 분산하고 정보가 매년 생산되기 때문에 더 좋은 산출물을 보장한다. 매년 프랑스의 센서스 비용은 약 일반 센서스의 1/7이다. 한

번 시작된 후에는 순환센서스의 연례적 표본조사 사업이 중도에 포기되어서는 안될 것이다. 잃어버린 1년을 복구하는 비용은 평상시의 1년간의 예산보다 훨씬 높기 때문이다.

따라서 전통적인 센서스에서 순환센서스로의 전환은 많은 중요한 전제조건들과 철저한 준비(표본 틀의 작성, 표집, 예산 협상 등)가 필요하고, 그렇기 때문에 상당한 협동적인 노력이 필요하다.

센서스 디자인의 변경을 위하여 준비를 해야 하는 중요한 단계는 센서스 결과 이용자와의 협력이다: 10년 주기 자료의 snapshot 시스템에서 일정기간의 확률표본(random sample)인 연간생산자료 시스템으로의 전환은 일조일석에 이루어질 수 있는 것은 아니다. 프랑스에서는 또한 센서스의 오랜 참여자이며 새로운 센서스의 추진과정에서 기초자치단체들의 동반자적 협력태도가 중요한 촉매역할을 했다. 이해당사자의 협력과정에는 센서스 결과를 이용하는 전국수준과 지방수준에서의 정치의사결정을 모두 고려하지 않으면 안 된다.

프랑스에서는 자료수집 및 공표를 위하여 이해당사자들의 다양한 협력이 현재 진행 중이다. 국가센서스평가위원회가 국가통계정보위원회(CNIS; National Council on Statistical Information)에 설립되었다. 상원의원이 위원장을 맡고 있고 INSEE, 기초자치단체들과 센서스 사용자들을 끌어들임으로써 국가센서스평가위원회는 자료수집 및 감독과정을 감시하고 여러 가지 프로토콜에 관해 수정안을 제시하고 센서스와 관련된 규제에 관한 수정 의견을 줄 것으로 기대된다. 자료 공표에 있어서는 INSEE와 기초자치단체들의 워킹그룹은 기초자치단체 단계 및 그 위의 단계에서의 자료 배포 및 서비스에 관한 포럼을 제공할 계획으로 되어 있다.

부록 II: 대만의 2010년 등록센서스의 세부절차¹⁴⁾

A. 공무등기자료 매칭작업

인구부문

1. 내국인 인구: 호적자료(戶籍資料, household registry data)를 중심으로, 신분증통일편호(身分證統一編戶, ID, 우리나라의 주민등록번호와 유사함)를 이용하여, 관련 공무파일을 매칭하여, 대만 내국인 인구 「공무등기자료파일」(公務登記資料檔)이라는 등록파일을 작성한다.

(1) 자료출처

1. 전체 가구 호적자료 파일(대만 내정부)
2. 내국인 출입국 자료파일 (대만 내정부 출입국 및 이민서)
3. 학적자료 파일(교육부 및 각급학교)
 - ① 학생자료 파일
 - ② 각급학교 코드(代碼) 및 주소 파일
4. 건강보험자료 파일(대만 중앙보건국)
 - ① 보험가입단위(group insurance applicants) 명부
 - ② 피보험자(insured) 명부
 - ③ 외래진료 및 입원진료 명부
5. 시설인구 및 교도소 수감인구(대만 법무부)
6. 심신장애자 파일(수첩) (대만 내정부 사회사)

(2) 정비작업:

1. 호적자료 파일: 신분증통일편호(ID)와 출생연월일(BD)에 의거하여, 중복인구를 삭제한다.
2. 출입국 파일: 인구주택총조사의 정의에 의거하여, 대만의 통치권이 귀속되는 대만지구에 상주하지 않는 내국인 인구를 제외한다.
3. 학생자료 파일: ID에 의거하여 각년도 졸업생과 신입생 자료를 정리

14) 이 자료는 통계청 인구조사과가 현지방문을 통하여 획득한 2010년의 대만 등록센서스 계획안 (中華民國 行政院, 2009a, 2009b, 2009c) 을 번역한 것으로, 우리나라의 2015년 등록센서스 추진을 위한 참고자료로 수록하는 것이다.

하고, 재학생의 학교 코드를 획득하여, 각급학교의 코드와 주소 파일에 의거하여, 각급학교가 위치하는 현의 시, 향, 진별로 촌리자료를 정비한다.

4. 건강보험자료 파일: 피보험자 파일과 외래진료/입원진료 파일을 운용하여 개인보험가입단위를 정리하고 통상적으로 내원하는 병의원의 주소 자료를 비교한다.

5. 시설인구와 교도소 수감인구 파일: 시설인구와 교도소 코드 파일과 인구자료 파일을 운용하여, 시설인구와 교도소 수감자의 시설 주소와 교도소 주소를 연결 정리한다.

6. 심신장애자 인구 파일: 2000년 대만 내정부의 정리되지 않은 심신장애자 파일을 정리하여, 파일 중에 중복, 착오 또는 누락의 상태를 점검하고, ID 점검을 통하여 중복인구를 가감한 후에, 다시 사망신고등기부(死亡申告登記簿)를 연결하여 사망자를 제외한다.

(3) 자료매칭작업:

가구자료 파일, 곧 호적 자료파일을 기초로 하여, ID+BD를 매칭키로 하여 매칭하고, 매칭작업을 완료한 후, 시험조사나 기타 관련 자료를 이용하여 공적장부 파일자료의 품질을 점검해서, 「내국인 인구 공적장부 자료파일」(本國人口公務登記資料檔)을 작성한다. 매칭작업은 아래와 같이 실시되었다.

1. 내국인 출입국 자료파일을 연결하여, 가구자료, 바로 호적 자료파일 에 “대만지구 비상주 인구”(非常住臺灣地區人口)를 주기(註記)하고 삭제한다. 2004년 가구 자료파일로 매칭상태를 보면, 매칭률은 93.9%에 이른다. 인구동태조사 자료파일과 비교 분석하여, 비상주인구가 있으면 내국인 출입국 자료파일을 보완하여, 실제로 해외에 거주하는 대만 내국인의 인구 자료를 확실하게 파악하도록 한다.

2. 학적 자료파일을 연결하여 재학생이 다니는 학교시설의 향(鄉), 시(市), 진(鎮) 별 주소지 정보를 획득하고, 다시 건강보험 자료파일을 연결하여 개인이 통상적으로 방문하는 병의원의 주소지와 보험가입단위의 주소지 정보를 획득하여, 개인이 거주 가능한 향(鄉), 시(市), 진(鎮) 별 주소지 정보를 종합 정리하여, 호적지(가구등록주소지)와 상주지(통상적 거주지)의 차이를 판별하는 보조 자료로 활용한다. 이 자료는 각

현의 향(鄉), 시(市), 진(鎭) 별 상주인구 추계를 위한 참고자료로 이용한다.

3. 시설인구와 교도소 수감인구를 연결하여, 가구 자료파일에 시설인구 또는 교도소 수감인구로 표기하고 시설 또는 교도소의 주소지를 기록한다. 2004년 가구자료로 매칭상태를 보면, 매칭률은 97.2%에 이르렀다. 인구동태조사 자료와의 비교분석에서는 자료 자체가 민감하였으며, 방문조사 대상가구의 다수가 응답을 거부하여 조사원이 상세한 질문을 하는 것이 불가능하였다. 이 때문에, 공적장부 자료파일을 보완하여, 시설인구와 교도소 수감인구에 대한 정보를 정확하게 취득할 필요가 있다.

4. 심신장애자 자료파일을 연결하여, 심신장애별 등급별 자료를 획득하여, 중복인구나 사망자 인구를 가감한 후, 2000년 인구주택총조사 자료와 매칭작업을 하였는데, 매칭률은 85%였다. 그러나 매칭된 총인원과 내정부의 발표결과 간에 차이가 있었지만, 매칭률을 개선할 필요가 있다. 이 때문에 총조사 결과에 포함하지 않고, 별도의 특별보고서로 통계결과를 공표하였다. 2003년 내정부(內政部)는 사회복지정보통합시스템(社政資訊整合系統)을 수립하고, 각 현의 사회복지관련 자료파일을 통합하여, 통일적인 표준방식을 개발하고 있기 때문에, 자료의 누락이나 중복상태를 개선하여 향후 자료의 연결 상태의 정확도를 향상시킬 것으로 기대된다.

2. 외국인 인구: 외국인 체류자, 외국인 노동자, 홍콩, 마카오, 대륙을 방문하는 대만 내국인 자료를 운용하여, 「국내거주 외국인 국적인구 공정장부 파일」(居住我國之外籍人口公務資料檔)을 수립한다.

(1) 자료출처

1. 외국인 체류자 인구파일(내정부 출입국 및 이민서)
2. 외국인 노동자 파일(노공위원회)
3. 홍콩, 마카오, 대륙방문 대만인사 파일(내정부 출입국 및 이민서)

(2) 정비작업

1. 외국인 체류자 인구 파일: 먼저 가족관계를 검토하여, 해당 가구(호적)

에 귀속시키고, 가구(호적) 귀속이 불가능한 자는 체류지를 기준으로 호적지를 설정하여 해당 가구(호적)에 귀속시키고, 나머지 가구(호적) 귀속이 불가능한 사람은 체류지에서 집계한다.

2. 외국인 노동자 파일: 외국인 국적의 남자하인이나 간호사는 고용주의 가구에 귀속시키고, 제조업이나 건설업에 종사하는 외국인 노동자는 작업장에 의거하여 해당 가구에 귀속시킨다.

3. 홍콩, 마카오, 대륙방문 대만인사 파일: 먼저 가족관계를 검토하여, 해당 가구(호적)에 귀속시키고, 가구(호적) 귀속이 불가능한 자는 체류지를 기준으로 호적지를 설정하여 해당 가구(호적)에 귀속시키고, 나머지 가구(호적) 귀속이 불가능한 사람은 체류지에서 집계한다.

(3) 자료매칭작업: 패스포트 번호(PD)를 매칭키로 하는 「외국인 자료파일」(非本國籍資料庫)을 수립하고, PD와 대만입국연월일을 기준으로, 위의 3개 자료파일을 연결하여, 중복자료를 제거한 후에, 센서스 정의에 입각하여 외국인상주인구정보를 구축하고, 센서스 결과 작성에 응용한다.

3. 내국인 인구 자료파일과 국내거주 외국인인구 자료파일을 결합하여, 「인구부문 공적장부 자료파일」(人口公務登記資料庫)를 수립하여, 인구주택총조사 기본자료 파일로 한다.

주택부문

대만지구 각 현의 향(鄉), 시(市), 진(鎮) 별 촌리 주소파일(Address Register)을 이용하여, 주택세파일(房屋稅主檔)과 주택대장파일(房屋中文主檔)을 연결하여, 주택공적장부파일(住宅公務登記資料檔)을 수립하여, 인구주택총조사의 주택관련 기본 자료로 한다.

(1) 자료출처

1. 촌리주소파일(村里門牌檔) (내정부)
2. 재산세 관련 주택파일(財稅房屋稅主檔) (재정부)
3. 각 현시 방옥과세주당(房屋課稅主檔) (재정부)

(2) 정비작업

1. 촌리주소(門牌)파일: 공식이나 수작업 방식으로, 중국어 주소지를 정리

하여 일치하도록 하고, 각 촌리의 가도일람표(街道一覽表, Street Data Sheet)를 작성하여 기타 자료파일을 검색하고 운용하는데 이용한다.

2. 주택대장파일(房屋中文主檔): 향후 주택대장의 중국어 주소지를 정리하여 1주택 1주소가 되도록 하며, 구체적으로, 각 현의 향(鄉), 시(市), 진(鎭) 별 촌리 주소파일이 완전히 정비되면, 주택과 주소는 1: 1로 대응하게 될 것이다.

3. 주택세파일(房屋稅主檔): 먼저 건축물을 용도별로 구분하고, 정거장, 엘리베이터 등의 주택과 무관한 건물은 제외하고, 다시 영업용 건물도 제외하고, 다시 자료를 재검하여 이질적 자료를 삭제 보완한다.

(3) 정비작업: 주택세 납부번호(編戶 ID)에 의거하여 주택대장 파일과 주택과세대장 파일을 연결하고, 다시 촌리 주소(門牌)대장과 연결하여 해당 건물의 현존여부를 확인하고, 「주택 공적장부 파일」(住宅公務登記資料主檔)을 수립하여 이것을 인구주택총조사에 활용한다.

B. 2010년 인구주택총조사 표본설계

1. 시뮬레이션 테스트: 표본추출(Sampling design)의 최적방안을 모색하기 위하여, 대만 행정원 주계처는 2007년부터 2000년 인구주택총조사와 호적대장자료를 이용하여, D-H 방법(Dalenius-Hodges method)과 비층화군집방법(non-hierarchical K-Means method)등의 층화법(stratification method), 모수통계와 비모수통계로 모두 활용될 수 있는 네이만 할당법(Neyman allocation) 등의 다양한 방법으로 100회에 걸친 시뮬레이션 테스트를 실시하였다. 시뮬레이션 테스트 과정에서, 각종 표본설계가 상주인구에 미치는 영향을 추정하고, 통계전문가의 의견을 지속적으로 청취하면서, 표본설계와 추정방식의 타당성을 점검하였다. 행정원 주계처가 최종적으로 채택하기로 결정된 방법은 층화군집추출법(分層集體抽樣法, stratified cluster sampling method)으로, 호적인구를 층화변수(또는 분층변수)로 하고, D-H 층화법, 모수통계의 네이만 할당법을 이용하여 각층의 표본수를 분배하는 것이다. 이러한 표본설계는 이미 전문가회의에 보고되었으며, 제42차 센서스위원회의 심의를 통과하였다.

2000년 인구주택총조사에서 각현(진, 시, 구) 센서스 조직이 센서스 대상객체의 분포 때문에 센서스 조사구를 인위적으로 확정하는 경우가 있었으며, 이 때문에, 일부 센서스조사구는 인구수가 일부 불균형을 이루는 현상이 나타났다. 2010년 인구주택총조사에서 지리정보시스템(GIS)을 운영하여, 안정되고 균질적으로 수치화된 센서스조사구를 확정하게 되기 때문에, 시뮬레이션 테스트의 조건은 2010년의 실제상황과 유사해지며, 테스트 결과는 오류를 낳을 가능성이 줄어든다. 표본추출 과정에서, 인구규모가 600명 이상이거나 200명 미만인 특수상황의 센서스 조사구는 일차적으로 배제하며, 100명을 초과하는 집단가구가 있는 센서스 조사구가 있는 경우에는 관련 주관기관이 전면 방문조사를 실시하며, 이들은 표본추출 대상에서 배제하고 다시 표본추출 시뮬레이션 테스트를 진행하였다.

최근의 시뮬레이션 테스트 결과에 의하면, 2010년 인구주택총조사의 총표본 추출률은 16%였으며, 대만지구 각 현시 추출률은 그 범위가 11%에서 29%의 사이에 있었다. 인구규모가 상대적으로 적거나 변차가 큰 향(鄉), 진(鎮), 시(市) 구는 표본수를 늘리거나 전수조사를 실시해야 이들 조사구의 상

주인구와 특성치를 추정하는 것이 가능할 것이다.

2. **표본추출방법:** 가구특성에 의거하여 일반조사구와 전수조사구로 나누어, 별도의 조사방법을 채택한다.

(1) 일반조사구: 대만의 통치권이 미치는 모든 지역의 각 향(鎭), 시(市), 구(區))을 하위모집단(sub-population)으로 하고, 층화군집추출법(分層集體抽樣法, stratified cluster sampling)에 의하여 센서스 조사구를 최소추출단위로 하고 표본을 추출한다. 호적인구수로 각각의 하위모집단 층수를 결정하며, Neyman 할당법에 의하여 각층 표본수를 결정하고, 각층을 계통표집방식(systematic sampling)으로 추출하여 조사구를 선정하며, 조사구 인구와 주택을 순서대로 방문하여 조사하는데, 총 추출률은 16%이며, 표본추출의 순서는 다음과 같다.

1. **표본추출 모집단:** GIS 통합주소시스템과 호적대장 자료를 운용하여, 대만 전국을 완전히 수치화된 조사구로 정비하고 (센서스조사구 당 가구 수는 대략 110 ± 30 가구), 이것을 바탕으로 계통표집방식에 의하여 표본추출작업을 진행하여 모수를 추출한다.

2. **추출단위:** 센서스조사구를 표본추출의 단위로 한다.

3. **층화기준:**

제1단계: 각 하위모집단(향진시구)의 호적인구수를 고려하여, 하위모집단 계층수를 결정한다. 단 최다 7계층으로 제한한다.

제2단계: 호적인구수를 층화변수(stratification variable)로 삼고, 각각의 하위모집단에 대하여 D-H(Dalenius and Hodges) 층화법을 채택하고, 적당한 층화지점(分層地點)을 선택한다.

<표 1> 2010년 센서스 각 향진지구 계층수와 할당상태 추정결과

항목별	배치층 수	향진 수	구성비
총계	-	368	100.00
호적인구수 < 5,000	1	24	6.52
5,000 ≤ 호적인구수 < 10,000	2	38	10.33
10,000 ≤ 호적인구수 < 25,000	3	83	22.55
25,000 ≤ 호적인구수 < 50,000	4	91	24.73
50,000 ≤ 호적인구수 < 100,000	5	63	17.12
100,000 ≤ 호적인구수 < 200,000	6	42	11.41
호적인구수 ≥ 200,000	7	27	7.34

4. 표본할당

(1) 각 하위모집단 추정오차율(ϵ)이 $\pm 3\%$ 가 되도록 하고, 신뢰계수 $(1-\alpha)\%$ 가 95%일 때, 식(a)에 의하여 각 하위모집단의 조사구 표본수를 계산하고, 다시 식(b)에 의하여 총 표본수를 얻는다.

(2) 계층별 표본조사 비용이 똑 같다는 전제 아래서, Neyman 할당법을 채택하여, 각 하위모집단의 각 계층별 표본규모를 식(c)에 의하여 결정한다.

$$\Pr\left(\left|\frac{\bar{y}_i - \bar{Y}_i}{\bar{Y}_i}\right| \geq \epsilon\right) = \alpha \text{ (정확도)}$$

$$n_i \geq \frac{\left(\sum_{ih} S_{ih}\right)^2}{\left(\frac{Y_i \epsilon}{Z_{1-\frac{\alpha}{2}}}\right)^2 + \sum_h N_{ih} S_{ih}^2} \dots\dots\dots(a)$$

$$n = \sum_i n_i \dots\dots\dots(b)$$

$$n_{ih} = \frac{N_{ih} S_{ih}}{\sum_h N_{ih} S_{ih}} \times n_i \dots\dots\dots(c)$$

부호설명

n_i : i 번째 하위모집단의 표본조사구 개수

S_{ih} : i 번째 하위모집단, h 번째 층의 조사구 호적인구수 표준 차이

N_{ih} : i 번째 하위모집단, h 번째 층의 조사구 총수

$$Y_i = \sum_h N_{ih} \bar{Y}_{ih}$$

: i 번째 하위모집단 호적인구 수

\bar{Y}_{ih} : i 번째 하위모집단, h 번째 층의 센서스 조사구 평균 호적인구 수

5. 표본조정: 표본할당 후, 추출률이 75%를 넘는 향진시구는 전수조사 방식을 채택하여 관리한다.

6. 표본추출: 향진 안에서 각 층을 나누어서 계통추출방식에 의하여 센서스조사구를 선정하고, 구내 인구와 주택에 관한 정보를 방문조사로 수집한다.

3. 각 현시 추정 표본조사구 및 추출률

현시명칭		표본 센서스 조사구 수	추출 률 (%)	현시명칭		표본 센서스 조사구수	추출률 (%)
總計	합계	10,250	16	屏東縣	핑동현	470	20
臺北縣	타이페이현	1,380	12	臺東縣	타이둥현	200	29
宜蘭縣	이란현	240	19	花蓮縣	화롄현	240	24
桃園縣	타오위안현	730	13	澎湖縣	핑후현	80	29
新竹縣	신주현	330	24	基隆市	지룽시	190	16
苗栗縣	마오리현	320	22	新竹縣	신주현	170	15
臺中縣	타이중현	700	17	臺中市	타이중시	470	15
彰化縣	장화현	590	18	嘉義市	자이시	90	12
南投縣	난타우현	330	23	臺南市	타이난시	290	13
雲林縣	원린현	410	21	臺北市	타이페이시	820	11
嘉義縣	자이현	320	21	高雄市	카오슝시	500	11
臺南縣	타이난현	640	20	連江縣	롄장현	30	100
高雄縣	카오슝현	610	17	金門縣	진먼현	100	36

4. 추계방법:

- (1) 표본조사방식에 의한 향진시구의 관리: 상주인구, 상주가구, 주택의 개별 특설치는 개별 비율추정법에 의하여 추정하는데, 같은 향진시구 같은 층의 표본 특성치 총수와 호적인구(혹은 호적가구) 총수의 비율을 이용하여 각 하위모집단 특성치 자료를 추정한다, 추정공식과 설명은 다음과 같다.

1. 일반조사가구:

$$r_{ih} = \frac{\sum_{j=1}^{n_{ih}} x_{ihj}}{\sum_{j=1}^{n_{ih}} y_{ihj}} = \frac{x_{ih}}{y_{ih}}$$

$$\widehat{X}_{ih} = r_{ih} \times Y_{ih}$$

$$\widehat{X}_i = \sum_{h=1}^{L_i} \widehat{X}_{ih} = \sum_{h=1}^{L_i} r_{ih} \times Y_{ih}$$

$$V(\widehat{X}_i) = \sum_{h=1}^{L_i} N_{ih}^2 \frac{N_{ih} - n_{ih}}{N_{ih}} \times \frac{1}{n_{ih}} S_{s_{ih}}^2$$

$$S_{s_{ih}}^2 = \frac{1}{N} \sum_{j=1}^{N_{ih}} (x_{ijh} - r_{jh} y_{ihj})^2$$

$$= (S_{s_{ih}}^2 + r_{ih}^2 S_{y_{ih}}^2 - 2r_{ih} \rho_{ih} S_{\xi_h} S_{y_{ih}})$$

$$\rho_{ih} = \frac{S_{xy_{ih}}}{S_{\xi_h} S_{y_{ih}}}$$

부호설명:

x_{ihj} : i 번째 하위모집단, h 번째 층. j 번째 센서스조사구 모항(某項)의 상주인구, 상주가구, 또는 상주주택 특성치의 총수 (특별조사가구는 포함하지 않음)

y_{ihj} : i 번째 하위모집단, h 번째 층. j 번째 센서스조사구 호적인구수 혹은 호적가구수

\widehat{X}_j : i 번째 하위모집단 모항(某項) 상주인구, 상주가구 혹은 상주주택 특성치의 추계 총수(특별조사가구를 포함하지 않는다)

L_i : i 번째 하위모집단의 층수

N_{ih} : i 번째 하위모집단, h 번째 층의 센서스조사구 총수

n_{ih} : i 번째 하위모집단, h 번째 층의 표본 센서스조사구 총수

2. 특별조사가구:

$$X'_i = \sum_{k=1}^{M_i^s} X_{ik}^s$$

부호설명:

X_{ik}^s : i 번째 하위모집단, k 번째 층 전수조사가구, 모항(某項)의 상주인구, 상주가구, 상주주택의 특성치 총수

X_i^s : i 번째 하위모집단에 있는 전수조사가구 모항(某項)의 상주인구, 상주가구, 상주주택의 특성치 총수

M_i^s : i 번째 하위모집단 중에서 전수조사 가구의 총수

3. 향진지구 특성치의 총수 추정: 일반조사가구의 모항 특성치 추계 총수와 특별조사가구의 모항 특성치 총수를 합하여 향진지구 모항 특성치의 총수 추정결과로 한다.

(2) 전수조사방식에 의한 향진지구의 관리: 표본할당 후, 추출률이 75%보다 큰 향진지구 및 복건성 연강현은 전수조사방식으로 관리한다. 추계

공식과 설명은 다음과 같다.

1. 일반조사가구

$$X'_i = \sum_{j=1}^{N_j} X_{ij}$$

부호설명

X_{ij} : i 번째 하위모집단, j 번째 센서스조사구 모(某) 특성의 상주인구, 상주가구, 상주주택의 특성치 총수

N_j : i 번째 하위모집단의 센서스조사구 총수

X'_i : i 번째 하위모집단 모항(某項)의 상주인구, 상주가구, 상주주택 특성치의 총수(특별조사구는 포함하지 않는다)

2. 특별조사가구

$$X_i'^s = \sum_{k=1}^{M_j^s} X_{ik}^s$$

부호설명

X_{ik}^s : i 번째 하위모집단, k 번째 특별조사가구 모(某) 항의 상주인구, 상주가구, 상주주택의 특성치 총수

M_j^s : i 번째 하위모집단 중 특별조사가구의 총가구수

$X_i'^s$: i 번째 하위모집단 특별조사가구의 모항(某項) 상주인구, 상주가구, 상주주택 특성치의 총수(특별조사구는 포함하지 않는다)

3. **향진지구 특성치 총수:** 일반조사가구의 모항 특성치 총수와 특별조사가구의 모항 특성치 총수를 합하여 향진지구 모항 특성치의 총수로 한다.

대만 2000년 센서스 조사표-뒷면

Back page

0003
0004

1. Name and Sex	<input type="text"/> <input type="checkbox"/> 1. Male <input type="checkbox"/> 2. Female	<input type="text"/> <input type="checkbox"/> 1. Male <input type="checkbox"/> 2. Female
2. Date of birth	<input type="checkbox"/> 1. After A.D. Year <input type="text"/> <input type="text"/> <input type="text"/> Month <input type="text"/> <input type="text"/> <input type="text"/> Day <input type="text"/> <input type="text"/> <input type="text"/> <input type="checkbox"/> 2. Before A.D.	<input type="checkbox"/> 1. After A.D. Year <input type="text"/> <input type="text"/> <input type="text"/> Month <input type="text"/> <input type="text"/> <input type="text"/> Day <input type="text"/> <input type="text"/> <input type="text"/> <input type="checkbox"/> 2. Before A.D.
3. Personal Identification No. <small>Foreigners should enter Nationality Code (from left to right)</small>	<input type="text"/>	<input type="text"/>
4. Living status	<input type="checkbox"/> 1. Yes (people who have lived or are expected to live in this address for three months or more) <input type="checkbox"/> 2. No, usually living in <input type="checkbox"/> (1) Taiwan-Patris area <input type="checkbox"/> (2) Non-Taiwan-Patris area (skip 5-13.) <input type="checkbox"/> (3) Road (skip 5-13.)	<input type="checkbox"/> 1. Yes (people who have lived or are expected to live in this address for three months or more) <input type="checkbox"/> 2. No, usually living in <input type="checkbox"/> (1) Taiwan-Patris area <input type="checkbox"/> (2) Non-Taiwan-Patris area (skip 5-13.) <input type="checkbox"/> (3) Road (skip 5-13.)
5. Relationship to householder	<input type="checkbox"/> 1. Householder <input type="checkbox"/> 7. (Former) Child's spouse <input type="checkbox"/> 2. Spouse/Cohabitant <input type="checkbox"/> 8. Grandchild or his/her spouse <input type="checkbox"/> 3. (Former) Parent <input type="checkbox"/> 9. Brother/sister or his/her spouse <input type="checkbox"/> 4. Spouse's (former) parent <input type="checkbox"/> 10. Other relative <input type="checkbox"/> 5. Grand parent <input type="checkbox"/> 11. Employee <input type="checkbox"/> 6. (Former) Child <input type="checkbox"/> 12. Other nonrelative	<input type="checkbox"/> 1. Householder <input type="checkbox"/> 7. (Former) Child's spouse <input type="checkbox"/> 2. Spouse/Cohabitant <input type="checkbox"/> 8. Grandchild or his/her spouse <input type="checkbox"/> 3. (Former) Parent <input type="checkbox"/> 9. Brother/sister or his/her spouse <input type="checkbox"/> 4. Spouse's (former) parent <input type="checkbox"/> 10. Other relative <input type="checkbox"/> 5. Grand parent <input type="checkbox"/> 11. Employee <input type="checkbox"/> 6. (Former) Child <input type="checkbox"/> 12. Other nonrelative
6. Marital status	<input type="checkbox"/> 1. Unmarried <input type="checkbox"/> 2. Married or cohabitant <input type="checkbox"/> 3. Divorced or separated <input type="checkbox"/> 4. Widowed	<input type="checkbox"/> 1. Unmarried <input type="checkbox"/> 2. Married or cohabitant <input type="checkbox"/> 3. Divorced or separated <input type="checkbox"/> 4. Widowed
7. Educational attainment	<input type="checkbox"/> 1. Attending school <input type="checkbox"/> (1) Primary school <input type="checkbox"/> (5) Junior college <input type="checkbox"/> 2. Graduated <input type="checkbox"/> (2) Junior high school <input type="checkbox"/> (6) University <input type="checkbox"/> 3. Attended not completed <input type="checkbox"/> (3) Senior high school <input type="checkbox"/> (7) Master's degree program completed <input type="checkbox"/> (8) Vocational senior high school <input type="checkbox"/> (8) Associate degree program <input type="checkbox"/> 4. Never or not attending school <input type="checkbox"/> (9) Child under 6 years old <input type="checkbox"/> (11) Self-educated <input type="checkbox"/> (10) Illness	<input type="checkbox"/> 1. Attending school <input type="checkbox"/> (1) Primary school <input type="checkbox"/> (5) Junior college <input type="checkbox"/> 2. Graduated <input type="checkbox"/> (2) Junior high school <input type="checkbox"/> (6) University <input type="checkbox"/> 3. Attended not completed <input type="checkbox"/> (3) Senior high school <input type="checkbox"/> (7) Master's degree program completed <input type="checkbox"/> (8) Vocational senior high school <input type="checkbox"/> (8) Associate degree program <input type="checkbox"/> 4. Never or not attending school <input type="checkbox"/> (9) Child under 6 years old <input type="checkbox"/> (11) Self-educated <input type="checkbox"/> (10) Illness
8. Place of residence 5 years ago (before Dec. 15, 1995) <small>For persons aged under 6 years old or born after Dec. 15, 1995, skip to Question 9.</small>	<input type="checkbox"/> 1. The same as the living address <input type="checkbox"/> 5. Taiwan-Patris area - Please specify <input type="checkbox"/> 2. The same as the village's where you live <input type="checkbox"/> 6. The Chinese Mainland City/County <input type="checkbox"/> 3. The same as the town/neighborhood/city/district where you live <input type="checkbox"/> 7. Overseas Code: <input type="text"/> <input type="checkbox"/> 4. The same as the city/town where you live	<input type="checkbox"/> 1. The same as the living address <input type="checkbox"/> 5. Taiwan-Patris area - Please specify <input type="checkbox"/> 2. The same as the village's where you live <input type="checkbox"/> 6. The Chinese Mainland City/County <input type="checkbox"/> 3. The same as the town/neighborhood/city/district where you live <input type="checkbox"/> 7. Overseas Code: <input type="text"/> <input type="checkbox"/> 4. The same as the city/town where you live
9. Family living supporter	<input type="checkbox"/> 1. Yes <input type="checkbox"/> 2. No (household composed of two or more unrelated individuals, skip to 8.)	<input type="checkbox"/> 1. Yes <input type="checkbox"/> 2. No (household composed of two or more unrelated individuals, skip to 8.)
10. Working status (Dec. 9 ~ 15, 2000)	<input type="checkbox"/> 1. Serviceman on active duty <input type="checkbox"/> 2. Holding a job (part-time work for family business with no pay for more than 15 hours a week) <input type="checkbox"/> 3. Holding no job <input type="checkbox"/> (1) Agriculture <input type="checkbox"/> (2) Industry <input type="checkbox"/> (3) Service	<input type="checkbox"/> 1. Serviceman on active duty <input type="checkbox"/> 2. Holding a job (part-time work for family business with no pay for more than 15 hours a week) <input type="checkbox"/> 3. Holding no job <input type="checkbox"/> (1) Agriculture <input type="checkbox"/> (2) Industry <input type="checkbox"/> (3) Service
11. Place of work or school <small>Students holding a part-time job after class should refer to their "place of work".</small>	<input type="checkbox"/> 1. Job holder, serviceman or student - Please specify <input type="checkbox"/> (1) The same as the living address City / County <input type="checkbox"/> (2) The same as the town/neighborhood/city/district where you live Town / Township City / District <input type="checkbox"/> (3) Not at the town/neighborhood/city/district where you live <input type="checkbox"/> (4) Not in Taiwan-Patris area Code: <input type="text"/> <input type="checkbox"/> 2. Holding no job or not attending school	<input type="checkbox"/> 1. Job holder, serviceman or student - Please specify <input type="checkbox"/> (1) The same as the living address City / County <input type="checkbox"/> (2) The same as the town/neighborhood/city/district where you live Town / Township City / District <input type="checkbox"/> (3) Not at the town/neighborhood/city/district where you live <input type="checkbox"/> (4) Not in Taiwan-Patris area Code: <input type="text"/> <input type="checkbox"/> 2. Holding no job or not attending school
12. Need care of others for 3 months or more since illness, injury or aging cause difficulty in or stop you from:	<input type="checkbox"/> 1. Eating <input type="checkbox"/> 5. Bathing <input type="checkbox"/> 2. Getting on/off bed <input type="checkbox"/> 6. Walking <input type="checkbox"/> 3. Dressing <input type="checkbox"/> 7. Housekeeping (cooking, laundering...) <input type="checkbox"/> 4. Going to the lavatory <input type="checkbox"/> 8. None (Multiple choice)	<input type="checkbox"/> 1. Eating <input type="checkbox"/> 5. Bathing <input type="checkbox"/> 2. Getting on/off bed <input type="checkbox"/> 6. Walking <input type="checkbox"/> 3. Dressing <input type="checkbox"/> 7. Housekeeping (cooking, laundering...) <input type="checkbox"/> 4. Going to the lavatory <input type="checkbox"/> 8. None (Multiple choice)
13. Indigenous status	<input type="checkbox"/> 1. Yes <input type="checkbox"/> (1) Ami <input type="checkbox"/> (4) Bunun <input type="checkbox"/> (7) Rukai <input type="checkbox"/> (10) Others <input type="checkbox"/> 2. No <input type="checkbox"/> (2) Atayal <input type="checkbox"/> (5) Puyuma <input type="checkbox"/> (8) Saisiyat <input type="checkbox"/> (3) Paiwan <input type="checkbox"/> (6) Tsou <input type="checkbox"/> (9) Yami	<input type="checkbox"/> 1. Yes <input type="checkbox"/> (1) Ami <input type="checkbox"/> (4) Bunun <input type="checkbox"/> (7) Rukai <input type="checkbox"/> (10) Others <input type="checkbox"/> 2. No <input type="checkbox"/> (2) Atayal <input type="checkbox"/> (5) Puyuma <input type="checkbox"/> (8) Saisiyat <input type="checkbox"/> (3) Paiwan <input type="checkbox"/> (6) Tsou <input type="checkbox"/> (9) Yami

II. Housing panel

1. Type of household	3. Occupancy status of this house	6. Ownership of this house
<input type="checkbox"/> 1. Ordinary household <input type="checkbox"/> 2. Group quarters <input type="checkbox"/> (1) Convalescent hospital <input type="checkbox"/> (3) School <input type="checkbox"/> (5) Hotel <input type="checkbox"/> (2) Hospital <input type="checkbox"/> (4) Temple <input type="checkbox"/> (6) Others	<input type="checkbox"/> 1. Occupied <input type="checkbox"/> 2. Unoccupied but for other uses (the end) <input type="checkbox"/> 3. Unoccupied and not for other uses (the end)	<input type="checkbox"/> 1. Self-owned <input type="checkbox"/> 2. Rented <input type="checkbox"/> 3. Allotted <input type="checkbox"/> 4. Other (e.g. borrowed)
2. Type of building	4. Housing usage	7. The year of moving into this house
<input type="checkbox"/> 1. Housing unit (mainly for family living, e.g. apartment, mansion, cottage...) <input type="checkbox"/> 2. Collective living quarters (mainly for institutional, commercial or industrial use, e.g. school dormitory, firm, business building, hotel, hospital, barracks...) (skip to the end) <input type="checkbox"/> 3. Other place (rent, sale, mobile, subway...) (skip to the end) <input type="checkbox"/> 4. Not exist (skip to the end)	<input type="checkbox"/> 1. Residence only <input type="checkbox"/> 2. Residence and factory <input type="checkbox"/> 3. Residence and business or service <input type="checkbox"/> 4. Residence and agriculture	Year (A.D.) <input type="text"/> <input type="text"/> <input type="text"/> <small>Please write "00" for those who move into this house</small>
5. Total number of rooms and bath and toilet of this house		Remark:
Total study, ... room(s) <input type="text"/> <input type="text"/> <input type="text"/> Living dining room(s) <input type="text"/> <input type="text"/> <input type="text"/> Total toilet(s) <input type="text"/> <input type="text"/> <input type="text"/> (excluding kitchen, garage)		<input type="text"/>

Interviewer: _____ Tel: _____ extension _____
 Enumerator: _____ (signature), Administrator: _____ (signature), Inspector: _____ (signature)

부록 III: 싱가포르의 2000년 등록센서스의 세부절차¹⁵⁾

A. 싱가포르 2000년 센서스 인구주택데이터베이스 개발

싱가포르는 아시아·태평양지역에서 처음으로 2000년 인구주택센서스를 등록센서스 방식으로 실시하였다. 2000년 6월 30일 기준 인구규모는 싱가포르 통계국의 가구등록데이터베이스(HRD, Household Registration Database)라는 행정자료를 이용하여 집계하였다. 싱가포르 통계국이 2000년에 등록센서스를 실시하기로 결정한 것은 대부분의 국가에서 조사표와 현지실시로 진행되는 전통적 방식의 센서스로부터 사고의 일대전환이라고 할 수 있다.

싱가포르에서 등록센서스는 1998년 싱가포르 내각의 승인을 얻어서 추진되었다. 등록센서스 추진을 위해서 개발된 행정자료는 가구등록데이터베이스(HRD)와 전국주택데이터베이스(NDD, National Database on Dwellings)로서, 이 등록대장은 인구와 주택에 관련된 기본정보를 제공한다. 여기서는 가구등록데이터베이스와 전국주택데이터베이스의 개발과정을 추적하고, 데이터베이스 통합과 자료의 품질에 관련된 쟁점을 논의하기로 한다.

1. 데이터베이스의 개관

가구등록데이터베이스(HRD)는 싱가포르 국민과 영주권자의 개인관련 기초정보를 제공하는 “중앙인구대장”(central population register)에 해당한다. 데이터베이스는 행정자료를 이용하여 분기별로 유지·보수되고, 업데이트되도록 되어 있다. 한편, 전국주택데이터베이스(NDD)는 싱가포르의 모든 거주자의 주소대장이다. 이것은 행정자료와 현장 표본조사를 통하여 월단위로 업데이트된다. 이 데이터베이스는 싱가포르 통계국이 실시하는 대부분의 가구조사, 특히 분기별로 실시되는 경제활동인구조사를 위한 표본틀(sampling frame)을 제공하는 중요한 자료이다.

2. 데이터베이스의 구축

15) 부록 III은 싱가포르 통계국의 2000년 인구주택센서스 행정보고서(Census of Population, 2000 Administrative Report)를 토대로 하여 작성된 것이다(Singapore Department of Statistics, 2002)

1990년대 초반, 싱가포르 통계국은 장기적으로 행정자료를 이용하는 등록 센서스로 이행하는 것이 가능하다는 판단을 하였다. 가구등록데이터베이스(RD)는 다행히, 행정자료 정비가 공공기관에서 확산되는 시점에서 구축되기 시작하였다.

가. 등록대장 이용에 대한 실험

1990년 인구센서스 기간 중에, 싱가포르 통계국은 대규모로 등록대장 이용에 대한 실험을 개시하였다. 모든 개인에 대한 정보를 관련 공공기관으로부터 획득하여, 그것을 통합하여, 센서스 조사표에 사전 인쇄하였다. 따라서 개별가구를 방문하는 조사원은 모든 가구원에 대한 기본정보가 있는 조사표를 가지고 현장조사를 하도록 되어 있었다. 응답자들은 사전 인쇄된 개인정보의 사실여부를 확인하고 편집할 수 있었다. 이 실험결과에 의하면, 싱가포르 인구의 전반적 특성이 행정자료 데이터베이스와 현지실사에 의하여 작성된 센서스 자료와 그리 다르지 않은 것으로 판명되었다. 이 사실은 바탕으로 하여, 싱가포르 통계국은 등록센서스 추진사업이 성공할 수 있을 것이라는 판단을 하였다.

나. 가구등록데이터베이스 개발

1990년 인구센서스에서 얻은 경험을 바탕으로, 싱가포르 통계국은 가구등록데이터베이스를 개발하기 위하여 적극적으로 노력하였다. 1992년 싱가포르 통계국과 무역·산업부 전산정보국 고급관리팀이 덴마크와 스웨덴을 방문하였다. 당시의 북유럽 국가들은 인구대장을 오랫동안 유지·관리하여 왔으며, 덴마크는 1970년 이후 인구주택총조사의 전통적 방식을 포기하였으며, 스웨덴은 전산화된 인구대장의 가용정보를 센서스 조사표에 사전 인쇄하여, 현지실사의 효율성을 높이는데 사용하고 있었다.

가구등록데이터베이스는 1995년 6월 말 현재 싱가포르 내무부 인민망(People Hub)과 1990년 센서스 데이터베이스에서 1회에 걸쳐 자료를 추출하여 구축하였다. 이들 데이터베이스에서 파악된 국민과 영주권자의 고유식별번호(UIN, Unique Identification Number)를 통하여, 거주자별로 개인관련 기초정보를 통합할 수 있었다. 개인관련 기초정보 등의 핵심변수로 구성되는 가구등록데이터베이스를 구축하고 나서, 다른 공공기관이 유지·관리하는 행정데이터베이스와 분기별로 매칭하여, 이 데이터베이스를 최신통합통계시스

템(integrated and updated statistical system)으로 관리되는 것을 가능하도록 하였다.

다. 전국주택데이터베이스의 개발

전국주택데이터베이스(NDD)는 1980년 인구센서스의 부산물로 개발되었다. 센서스에서 집계된 모든 단위주거의 기록을 이용하여, 주택모집단틀(Mater House Frame)을 구축하였다. 이 자료에 주택개발공사(Housing and Development Board)와 여타 공공기관의 정보를 월단위로 업데이트하였다. 1996년에 이 데이터베이스는 전국주택데이터베이스로 다시 명명되고, 주택관련 정보를 보장하였다.

3. 자료 업데이트 빈도

자료를 얼마만큼 자주 업데이트하는가? 여기에서는 가구등록데이터베이스(HRD)와 전국주택데이터베이스(NDD)의 업데이트 빈도에 대한 정보를 제공한다.

가. 가구등록데이터베이스(HRD)

가구등록데이터베이스는 정례적으로 업데이트되어 싱가포르 국민과 영주권자의 인구관련 기본정보를 종합적으로 저장하는 창고로 발전하였다. 모든 시점에서, 정상출산, 사망, 시민권 획득, 시민권 무효, 영주권 획득, 영주권 무효, 혼인, 이혼, 주소등록 변경, 재적학교, 최종학력 등에 대한 정보를 획득하

<표 1> 가구등록데이터베이스의 자료유형별 업데이트의 빈도

자료유형	빈도
정상출산	분기별
사망	분기별
입국	분기별
출국	분기별
혼인	분기별
주소변경	분기별
이혼	매년
재적학교/최종학력	매년

자료: 싱가포르 통계국(2002), Census of Population 2000 Administrative Report

도록 되어있다. 가구등록데이터베이스의 업데이트가 이루어지는 빈도는 다음과 같다.

가구등록데이터베이스 시스템은 자료의 품질과 정확성을 보장하기 위하여 업데이트 과정에서 오류수정 점검절차를 두고 있다. 타당성 또는 편집 점검 절차를 통과하지 못한 개인기록은 정밀검토의 대상이 된다. 필요한 경우에, 자료제공기관에 자문을 구하기도 한다.

나. 전국주택데이터베이스(NDD)

전국주택데이터베이스(NDD)는 개인주택과 공공주택에 대하여 주소기록을 매칭하여 자료를 업데이트한다. 거주지 주소는 전국수치주소(NCA, National Coded Address) 포맷으로 개별적으로 저장된다. NCA 포맷은 단위거처의 주소유형, 주택 또는 블록 번호, 가로 코드, 층수, 단위거처 번호 등을 확인해주는 25바이트 필드로 구성되며, 싱가포르 정부부처가 채택하는 표준 포맷이라고 보면 되겠다.

전국주택데이터베이스는 다음과 같이 업데이트된다.

<표 2> 전국주택데이터베이스의 자료유형별 업데이트의 빈도

자료유형	빈도
공공주택	매월
신규보고 민간주택	매월
점유대기 민간주택	매월
연 부동산 평가액과 부동산 세율	분기별
자동차 소유여부	분기별
지리적 경계	분기별

자료: 싱가포르 통계국(2002), Census of Population 2000 Administrative Report

4. 조사항목

가. 가구등록데이터베이스 조사항목

핵심 가구등록데이터베이스는 싱가포르 국민과 영주권자의 일차 조사항목과 보조 조사항목을 포함한다. 일차 조사항목은 다음과 같다.

- 교육식별번호(UIN, Unique Identification Number)

- 성명
- 성별
- 생년월일
- 연령
- 혼인상태
- 시민권자 여부

보조 조사항목에는 단위거처의 유형, 최종학력, 재적학교, 재적기간 등을 포함하고 있다.

나. 전국주택데이터베이스 조사항목

개별 단위거처에 대하여, 전국주택데이터베이스는 주거유형, 우편지구, 주소를 NCA 포맷으로 저장하고 있다. 이 데이터베이스는 1998년에 보장되어서, 부동산세율, 연단위 부동산 평가액, 자동차 소유여부 자료를 개별 단위거처에 대하여 기록하고 있다.

5. 데이터베이스 통합의 필요조건

싱가포르의 인구·주택 데이터베이스를 구축하기 위하여 통합 데이터베이스 시스템을 활용한 아시아 지역의 국가들 중 선두그룹에 속한다. 이들 통계 데이터베이스는 다른 정부기관과의 매칭을 통하여 규칙적으로 업데이트된다. 이러한 작업을 진행하기 위하여, 3가지 필요조건이 있는데 그 내역은 다음과 같다.

a) 개인기록을 매칭하기 위한 표준키(standard key)가 있어야 한다. 싱가포르에는 전국 표준키가 교육식별번호(UIN)로, 이것으로 싱가포르 국민과 영주권자 모두에게 부여된 9비트 필드와 점검수치로 구성되어 있다. 가구기록에 대해서는 주소 표준키가 NCA 포맷으로 정리되어 있으며, 싱가포르 정부부처가 대부분 채택하고 있는 표준키이다.

b) 효율적인 자료관리 시스템과 대용량 자료를 저장할 수 있는 능력이 필요하다. 정보통신기술의 급격한 발전은 데이터베이스의 통합과 저장을 기술적으로 가능하게 만들었다.

c) 공공기관 간의 협력이 필요하다. 이것은 공공기관이 자료 항목과 분류 시스템을 공유해야 하는 계획단계는 물론 등록센서스의 추진단계와 실시단계에서도 중요하다.

5. 데이터베이스 통합의 핵심 쟁점

대규모 데이터베이스를 구축하여 유지·보수하는데 필요한 조건이 충족되면, 통계청은 데이터베이스에 영향을 주는 두 가지 핵심 쟁점, 곧 자료보안과 자료품질의 문제에 관심을 기울여야 한다. 보안조치는 자료를 보호하기 위하여 비밀보호, 법률적, 행정적, 기술적 조치 등을 포함한다. 자료품질 문제는 포괄범위, 개념, 코딩, 분류, 기준시점 차이, 행정자료 오류, 특이자료 등을 포함한다.

가. 자료 보안·보호

통합데이터베이스 내에 포함된 정보를 부분적으로 또는 전체적으로 변경하거나 파괴하려는 불법적인 접속으로부터 통합데이터베이스를 보호하는 조치가 있어야 하겠다.

① 비밀보호 및 법률적 조치

특정 데이터베이스를 매칭하여 통합 데이터베이스를 구축하는 것이 상당한 장점이 있기는 하지만, 데이터 보안조치의 위반은 심각한 결과를 초래할 수 있다. 따라서 데이터의 비밀보호 조치를 준수할 필요가 있을 것이다. 한 가지 방법은 입법을 통한 것이다. 싱가포르의 경우, 통계국은 통계법과 센서스법에 의하여 해당권한을 부여받고 있어 해당권한법률은 통계자료의 수집, 편집, 출판에 관한 핵심 입법이며 해당법률은 비밀정보의 보호나 발표를 규제하며, 가구등록데이터베이스(HRD)와 전국주택대장(NDD)에 포함된 자료는 통계법에 의거하여 획득된 것이고 그 법에 의하여 보호를 받도록 되어 있다.

② 행정적 조치

법률 조치와 더불어, 행정조치가 정보의 무단 공표를 방지하기 위하여 취

해지도록 되어 있다. 이들은 다음과 같은 사항을 포함한다.

- a) 신참 직원들을 조직문화에 익숙하도록 하여, 데이터 보호의 중요성과 필요성을 강조한다.
- b) 주요 파일은 키와 자물쇠로 채워서 보안조치를 강화한다.
- c) 정보나 자료를 민감성이나 보안 수준에 의거하여 분류하고, 적정 수준의 보호등급을 부여한다.
- d) "알 권리"(right to know)의 원칙을 엄정하게 집행하여, 정보 또는 데이터에 대한 지식, 소유 또는 접근을 인가된 직원에게만 엄격히 제한하도록 한다.

③ 기술적 조치

정보기술의 변화추이나 데이터베이스 의존경향은 기술적 관점에서 데이터베이스에 포함된 자료를 보호하기 위한 엄격한 조치가 기술적으로 강구되어야 한다는 것을 의미한다. 다음 조치가 데이터베이스의 안전과 정보를 보호하기 위하여 구체화되어야 한다.

- a) 개인정보를 포함하는 데이터베이스는 물리적으로 안전한 사이트에 외부 네트워크와 연결되지 않은 상태에서 보관되어야 한다. 이러한 사이트는 인가도지 않은 방문객에게는 접근이 철저히 봉쇄되어야 마땅하다
- b) 컴퓨터 시스템은 이용자로 하여금 사용자 ID와 비밀번호를 이용하여 개인 신분을 확인하도록 요구해야 한다. 비밀번호는 한 달에 한 번씩은 변경되어야 한다. 통계국의 직원도 그들의 사용자 ID나 비밀번호를 노출하거나 공유해서는 안 된다.
- c) 불법침입에 대비하여 방화벽을 설치하고, 내부시스템과 외부시스템 간의 관계를 절연할 수 있는 조치를 강화한다.
- d) 자료전송은 보안장치가 마련된 이메일 네트워크나 인가된 직원의 디스

켓이나 카트리지를 통하여 실시되어야 한다.

- e) 데이터베이스 접속권한은 극소수의 인가된 통계국 직원들에게 제한되어야 한다.
- f) 자료 업데이트를 위한 IT 프로그램의 변경은 그들을 실시하기 전에 인가되어야 한다.
- g) 데이터베이스 변경내역은 보안점검을 목적으로 예외없이 상세히 기록되어야 한다.

나. 자료의 품질

데이터 통합의 핵심쟁점 중의 하나는 행정자료를 추가할 때 자료의 품질이나 정확성이 손상을 당하지 않느냐 하는 것이다. 다양한 행정자료를 원천으로 하여 정보를 통합하는 경우에, 다음 영역에 대하여 관심과 노력을 기울이고 필요가 있다.

a) 집계범위

개별 데이터베이스는 상이한 부분을 목표모집단으로 설정하고 있으며, 상이한 행정자료의 정보를 통합하는 것은 기록의 일관성 부재나 중복기재를 야기할 수 있다. 상당한 자원과 시간을 소요하여서라도 집계범위 문제를 분석하고 그것의 오류를 정정할 필요가 있다.

b) 개념과 정의

상이한 공공기관은 유사한 조사항목에 대하여 상이한 개념과 정의를 이용할 수 있다. 자료를 통합하는 기관들은 개념을 표준화 또는 조정하고 자료의 일관성을 유지하기 위하여 상당한 노력을 기울이지 않으면 안 될 것이다.

c) 코딩과 분류

상이한 데이터베이스는 상이한 코드(부호)와 분류 시스템을 채택하는 경우가 있다. 데이터를 통합하는 기관은 코드 변환표(code conversion table)를 작성하여 코드문제를 정확히 계획하여 분류상 문제가 없도록 해야 할 것이다.

d) 상이한 기준시점

데이터베이스별로 기준시점이 다를 수 있기 때문에, 데이터를 통합하는 기관은 모든 조사항목의 기준시점이 동일하도록 하는데 특별한 관심을 기울여야 한다.

e) 행정오류

데이터베이스는 정보가 오래된 것이나 결측값으로 된 것을 포함할 수 있다. 데이터 통합을 담당하는 기관은 이러한 부정합성이나 오류의 문제를 해결하지 않으면 안 될 것이다. 가령, 가구등록데이터베이스나 전국주택데이터베이스는 자동화된 오류점검 시스템을 갖추고 있어서, 다양한 정부부처의 행정자료를 통합하는데 있어서 상세한 알고리즘 체크나 표준수치점검의 과정을 포함한다.

f) 희귀 사건

데이터베이스는 어떤 경우에는 아주 희귀한 사건을 포함할 수 있다. 데이터를 통합하는 기관은 이들 기록을 점검하여 확인하는 노력을 기울여야 하며, 가능하면 자료제공기관과 협의하여 문제를 해결하여야 할 것이다.

6. 자료품질 확인방안

가구등록데이터베이스와 전국주택데이터베이스가 자료로서 정확하고 품질 수준이 우수하다는 것을 보장하기 위하여, 통계국은 규칙적인 점검방안과 개선방안을 마련하여야 한다. 데이터베이스를 점검하고 clean-up하기 위하여 다음의 3가지 방안을 제시한다.

가. 공식 추계인구와의 비교연구

싱가포르 통계국은 가구등록데이터베이스에서 획득한 인구규모를 매년 공표되는 현재추계인구(current population estimates)와 비교연구를 하도록 하고 있다. 현재인구추계인구는 1990년 센서스 기준인구에 자연증가와 국제인구이동을 합산하여 획득하는 수치이다. 당초에는 가구등록데이터베이스와 공식 현재추계인구와 상당한 차이가 있는 것으로 확인되었다. 오류의 원인을 확인하고, 자료제공기관의 도움으로 가구등록데이터베이스의 개인기록 오류를 정정함으로써 해서 그 차이를 줄이기 위한 단계적 조치가 강구되었다.

나. 정보 업데이트를 위한 소규모 표본조사의 실시

2000년 센서스를 개시하기 직전, 75세이상 노인의 “거주지” 점검은 우편조사표에 의하여 실시되었는데, 이것은 그들의 거주지와 행방을 파악하기 위한 것이었다. 고령자 관련 시설에서 나온 고령자 정보와 함께, 이 조사결과는 가구등록데이터베이스를 업데이트하는데 사용하였다. 이 통계조사는 고령자 중의 일부는 국내이동, 또는 해외이동을 하였거나 주소를 변경하여 고령자 시설에 거주하는 있는 것으로 파악되었다.

전국주택데이터베이스에 대해서는 규칙적인 통계조사를 실시하여 데이터베이스에서 단위거처의 실제적 지위를 파악하고자 노력하였다. 특정 단위거처가 없어지게 되는 경우에는 현장조사원이 그것의 진위를 파악하도록 하였다. 주요 부동산 개발업자와 연락하여, 특정의 대규모 주택프로젝트가 “임시점유허가”를 발부받았는가 또는 특정 사이트가 함몰(demolition)되었는가 등을 확인하도록 하였다.

다. 추가적인 매칭작업 및 자료 확인 작업

싱가포르 통계국은 내무부의 가사업무 종사자의 파일, 15세 미만 학령인구, 우편번호부 등을 가지고 매칭작업을 추가적으로 실시하여 가구등록데이터베이스(HRD)와 전국주택데이터베이스(NDD)의 자료에 대하여 정확성을 점검하는 조치를 실시하였다.

7. 자료의 완전성

가구등록데이터베이스의 주요 조사항목의 품질은 핵심적 조치를 실시한 후에, 완전성의 면에서 대단히 우수한 것임이 판명이 되었다. 대부분의 조사항목은 그 품질평가에서 집계범위가 100%에 이르렀다. 데이터베이스에 조사항목별로 개별기록에서 결측값(missing value)은 존재하지 않았다.

혼인상태나 주소에 대한 조사항목은 집계범위가 100%에 이르지 않았다. 그러나 이들 조사항목은 1995년 가구등록데이터베이스가 도입된 후, 모든 거주자에 대해서 완전성이 93%에서 99.5%로 개선되었다. 주소지의 집계범위가 100%에 이르지 않는 것은 등록시점에 정확한 거주지를 기재할 수 없는 새로

<표 3> 싱가포르 가구등록데이터베이스의 자료완전성 점검결과

조사항목(한글)	조사항목(영문)	1995년 12월	2000년 6월
고유식별번호	Unique ID Number(UIN)	100.0%	100.0%
거주상태	Residential status	100.0%	100.0%
성명	Name	100.0%	100.0%
성별	Sex	100.0%	100.0%
출생연월일	Date of birth	100.0%	100.0%
혼인상태	Marital status	92.7%	99.5%
인종/종족집단	Ethnic group	99.9%	100.0%
방언집단	Dialect group	99.9%	100.0%
출생국	Country of birth	100.0%	100.0%
시민권 여부	Citizen group	100.0%	100.0%
주소	Address	92.9%	99.5%

운 영주권자들 때문에 그런 영주권자들 때문에 혼인상태에 대해서, 집계범위가 100%에 이르지 않는 이유는 해외결혼과 같은 결혼건수는 관련기관에 등록되지 않은 사실에 기인하는 것으로 밝혀졌다.

이상의 사실을 바탕으로 하여, 싱가포르 통계국은 행정자료를 이용하여 정확한 인구규모와 인구관련 기본특성에 대한 정보를 제공할 수 있다고 확신하면서 등록센서스를 실시하기로 결정하였다. 싱가포르에서 법률적 환경이나 자료의 비밀준수 조항은 관련 행정정보의 공유를 허용한다. 더욱이, 이러한 접근방법을 채택하는데 있어서 비용절감은 상당한 수준에 이르는 것으로 파악되고 있다.

싱가포르 통계국은 2000년대 이후에는 가구등록데이터베이스와 전국주택데이터베이스의 기록들을 추가적으로 통합하여 인구의 연속추정시스템을 사정권으로 하고 있다. 소규모의 정례적인 통계조사의 시스템을 확립하여, 행정자료에서 얻을 수 없는 자료를 획득하고 어느 시점에서 관심대상이 되는 인

구현상과 사회적 동향을 모니터링 하는 것이 싱가포르 통계국의 관심이다.

B. 싱가포르 2000년 센서스 표본조사 기획 및 준비

싱가포르는 2000년 센서스에서 행정자료를 바탕으로 구축한 가구등록데이터베이스(HRD)와 전국주택데이터베이스(NDD)에서 가용하지 않은 추가적인 정보는 심층연구 목적으로 가구 20% 표본에서 획득이 가능하도록 하였다. 싱가포르 통계국은 센서스 표본조사의 성공을 위하여 세심한 기획과 준비가 필수적라고 판단하였다. 여기에는 센서스 조사항목의 최종 리스트를 정리하고, 싱가포르 국내에 거주하는 외국인과 해외에 거주하는 싱가포르 국민에 대한 포괄범위를 개선하고, 가구 20%를 선정하기 위한 표본추출 방법론을 설계하고, 표준코딩 및 분류시스템을 확립하고, 나아가 센서스 본조사를 실시하기 전에 시험조사 실시계획을 마련하는 것이 포함되어 있다.

1. 2000년 센서스의 조사항목

싱가포르 통계국은 1998년 5월에 2000년 인구센서스 조사항목을 기획하는 작업에 착수하였다. 다수의 정부부처와 법정기관들이 센서스 통계자료의 주이용자이기 때문에, 이들 기관에 센서스 실시기관인 싱가포르 통계국은 기존 조사항목들에 대하여 자문을 구하였다. 이들 기관은 싱가포르 통계국에 추가적으로 필요하다고 판단하는 조사항목에 대하여 답신하였다. 이러한 부처 간 대화과정에서 등록센서스의 20% 표본조사에서 획득해야 할 조사항목에 대한 리스트를 작성할 수 있었다. 싱가포르 통계청은 해당리스트를 상세히 검토하여, 조사항목 최종리스트를 작성하여 싱가포르 센서스기획위원회(Census Planning Committee)에 승인을 요청하였다.

가. 선정기준

싱가포르에서 인구센서스는 10년에 한 번씩 실시되기 때문에, 종전 센서스 조사항목과의 시계열 유지는 물론 광범한 쟁점과 주제에 대한 조사항목을 설정하는 것이 바람직하다. 센서스 현장운영과 효용극대화 간에 균형을 취하는 것도 다른 고려사항과 더불어 대단히 중요하다. 조사항목의 최종 내용은 조사표 형식, 응답내용의 정확성이나 품질, 자료처리과정, 센서스 결과공표

스케줄 등에 영향을 미치지 마련이다. 센서스 조사항목은 정부정책의 기획입안에 필요한 유용한 정보를 제공하여야 하며, 제한된 시간과 비용의 프레임 안에서 조사항목에 대한 자료를 확실하게 획득할 수 있어야 한다.

2000년 센서스 조사항목을 선정하는 과정에서, 싱가포르 통계국은 유엔 인구주택센서스의 원칙과 권고안(Principle and Recommendations for Population and Housing Censuses)을 참고하였다(United Nations, 2008). 이 권고안에는 다음 고려사항을 포함하고 있다.

- a) 조사항목은 주요 국가정책의 결정과정에 상대적으로 중요한 것을 선정하여야 마땅하다.
- b) 센서스에서 수집되는 정보는 정책기획이나 정책입안에서 정부의 당면요구를 충족시킬 수 있는 것이라야 한다.
- c) 조사항목은 심층적이거나 전문화 또는 특화된 것이 되어서는 안 된다.
- d) 조사항목이 광범위한 정밀조사(probing)를 필요로 하거나, 고도의 숙련된 조사원만이 할 수 있는 것을 포함해서는 안 된다.
- e) 조사항목은 민감한 주제나 논쟁거리가 되는 주제여서는 안 된다. 이들 주제는 인구특성이나 사회경제적 자료와 같은 센서스의 기초 조사항목의 자료 수집을 위태롭게 할 수 있다는 점을 망각해서는 안 된다.
- f) 조사항목은 너무 많은 상이한 주제를 포함해서는 안 된다.
- g) 경비나 인적자원의 활용은 허용 가능한 범위를 넘어서는 안 된다.

유엔의 권고안 외에도, 몇 가지 기준을 추가하여 정부부처가 요구하는 각종 항목의 자료수집에 대한 타당성을 평가하였다. 이들은 다음과 같은 사항을 포함한다.

- a) 응답자의 부담과 전국 센서스에서의 해당항목의 적합성

- b) 조사원의 자질
- c) 요구되는 자원(인원, 경비)

① 응답자 부담과 적합성 기준

싱가포르 통계국은 센서스에서 조사항목의 수가 계속해서 증가하였다는 사실에 주목하고 있다.

<표 4> 싱가포르 센서스의 조사항목 수: 1957-2000

센서스 연도	자료수집 항목수		
	항목 합계	전수항목	표본항목
1957	18	18	-
1970	36	17	19(10%)
1980	27	18	9(20%)
1990	45	39	6(10%)
2000*	54	8*	46(20%)

** 가구등록데이터베이스에서 추출한 조사항목

사회전반의 복잡성이 증가하면서, 센서스의 조사범위가 확대되는 것은 바람직하지만, 센서스에서 수집해야 할 자료항목의 수가 몇 개 정도가 바람직한가에 대해서는 응답자의 시간과 인내심 등을 고려하지 않을 수 없다.

현지실사에서 조사원이 가구원 개개인을 직접 면접하는 것은 운영상 별로 타당성이 없는 일이다. 특정 가구의 가구주나 여타 책임 있는 가구원이 다른 가구원을 대신하여 비교적 간단한 방법으로 응답을 할 수 있는 것이라야 한다. 추가적인 설명이 필요한 조사항목에 대해서도, 가구원이 응답하는데 너무 많은 시간이나 노력을 요구하도록 해서는 안 된다.

2000년 센서스에서, 전화조사나 현지실사 외에 조사자 직접기입 방식은 인터넷조사에 의하여 가능해졌다. 조사표의 항목이나 지침사항에 대한 관심, 이해, 해석의 수준은 응답자에 따라서 상당한 차이가 났다. 2000년 센서스에서 사용된 방법은 또한 응답자와의 접촉시간이 비교적 짧은 것을 의미하는 것이었다. 그 자체가, 응답자는 자세한 정밀조사나 성격상 민감한 조사항목(예: 가구원 중에서 신체적으로 활동제약이 있는지 없는지에 대한 문항)으로 인하여 응답에 부담을 느껴서는 안 된다는 것이었다. 사생활을 침해할 수 있는 민감한 문항은 응답자가 불평을 하거나 조사를 거부하는 사태가 일어날

수 있기 때문에, 이에 대한 특별한 조치를 할 필요가 있었다. 이러한 문제점들은 모두 예외 없이, 조사항목을 최종적으로 선정할 때 고려하도록 하였다.

② 조사원의 자질

2000년 센서스 20% 표본조사의 조사표는 80%가 전화조사나 현지실사로 수집될 것으로 싱가포르 통계국은 예상하였다. 센서스 자료수집 개시 이전에, 싱가포르 통계국은 120명의 전화조사원과 300명의 현지실사 조사원을 모집하였다. 센서스 가구의 개념, 가구 정보의 조사방법, 응답내용의 적절성, 커뮤니케이션 기법 등에 대한 교육을 싱가포르 통계국 직원들이 콜센터 훈련담당 직원을 대상으로 실시하였다. 그들은 다시 콜센터 조사원을 대상으로 적절한 훈련을 실시하여, 전화조사를 할 수 있도록 현장에 배치되었다. 콜센터가 가동 중인 기간에는 엄격한 감독을 실시하도록 하였다.

현지실사를 위하여 현장관리자를 대상으로, 현장관리보조 센서스 총감독팀(ACs, Fieldwork Assistant Census Superintendents)에 의하여 상세한 훈련이 실시되었다. 훈련과정에는 현지실사의 기획과 준비, 작업할당을 위한 현장시스템(Fieldwork System)의 이용, 현지실사 가구수의 조사원별 배당방법, 관리대상 조사원에 대한 조사지침 등을 포함하는 것이었다. 3일간에 걸친 현지실사 훈련이 현장조사원을 대상으로도 실시되었다. 현지실사가 일단 시작되면, 정례적인 모임이나 토론이 개개인의 현장관리보조 센서스 총감독(ACs)에 의하여 개별 지역사무소의 관리자나 조사원을 대상으로 실시되었다. 센서스 상의 각종 개념을 재차 강조하고, 현지실사 절차를 엄격히 집행하도록 하였다.

운영상의 효율성을 위하여, 센서스 조사항목의 포괄범위는 항목수, 항목내용, 이해수준 등에서 바람직한 수준이 될 수 있도록 조정하였다. 작업량은 콜센터 조사원이나 현지실사 조사원이 하루에 적당한 수의 가구를 효과적으로 방문하여 응답을 작성할 수 있도록 하였다.

③ 요구되는 자원

인구규모의 증가와 함께 센서스 조사항목 수의 증가는 바로, 센서스 실시에 요구되는 자원이 증가한다는 것을 의미한다. 1980년과 1990년의 인구센서스는 2,200명의 조사원과 3,500명의 조사원을 현지실사에 배치하였다.

싱가포르 통계국은 2000년의 등록센서스에서 자료수집과 자료처리에서 새

로운 IT기술의 혁신을 이용한 결과, 상당한 비용절감과 인력수요의 삭감을 가져왔다고 주장하고 있다. 그러나 가구원들의 응답을 얻는데 필요한 자료수집이나 설명에 소요되는 육체적 노력은 여전히 무시할 수 없는 수준에 있었다고 평가한다. 수집되는 자료의 양이 방대하기 때문에, 자료수집과 자료처리에 필요한 기간이 길어지고, 따라서 통계이용자에게 센서스 정보를 공표하는데 필요한 시간이 길어질 수도 있다는 것이 문제점이 되기도 하였다. 또한 센서스 자원의 적정화가 제대로 이루어지지 않을 수 있다는 문제점이 제기되기도 하였다.

나. 2000년 센서스 조사항목

충분한 검토와 토론을 거쳐서, 싱가포르 통계국은 54개 조사항목의 리스트를 최종 확정하였는데, 여기에는 인구학적 기본변수, 사회경제적 특성, 교육, 직업이동, 출산력, 주택, 수송, 가구정보 등이 포함되어 있었다. 이것은 적정 절차에 의거하여, 1999년 7월 센서스기획위원회(CPC, Census Planning Committee)에 의하여 2000년 센서스 조사항목으로 인준되었다.

다. 1990년 센서스 조사항목과의 비교

2000년 센서스 조사항목은 54개로서, 1990년 센서스의 조사항목 46개보다 8개가 증가하였다. 1990년 센서스와 비교하여, 교육개선, 직업이동, 주택개선, 해외여행, 고령자 재정지원 등에 대한 주요 정보를 획득하기 위하여 15개 항목이 추가되고, 1990년 센서스에서 두 개 항목(예: 초산연령 곧 첫째 자녀 출생시점의 모의 연령, 총 소득(원천별 합계))은 삭제하기로 하였다. 초산연령은 충분한 검토와 토론을 거쳐서, 싱가포르 통계국은 54개 조사항목의 리스트를 최종 확정하였는데, 여기에서 “초산연령”은 1967년 이후 출생한 자녀의 경우, 인구동향조사 출산부문에서 추출가능한 정보이기 때문에 삭제되었다. 총소득(원천별 합계)은 1990년 센서스에서 가구원으로부터 신뢰성 있는 정보를 획득한다는 것이 사실상 불가능하다고 판단하였기 때문에 삭제하였다.

조사항목의 최종리스트는 다음과 같다.

<표 5> 2000년 센서스 조사항목의 최종리스트와 1990년 센서스 조사항목의 비교

주제영역	조사항목	비고
개인 특수상황 및 인구관련 특성	성명, 고유 식별번호(내국인, 외국인- 외국인 등록 번호),성별, 인종/종족집단, 출생연월일, 출생국가, 시민권 여부, 거주지,	행정자료 조사 항목
	싱가포르 최초 입국연도, 혼인상태, 종교, 초혼연도, 생존자녀수, 싱가포르 국내/국외 체류여부(거주국, 해외체류 이유, 해외취업/유학 예상기간)	1990년과 2000년 센서스 공통항목
가구특성	가구주와의 관계, 배우자 매칭, 부모-자녀 매칭	1990년과 2000년 센서스 공통항목
주택 및 주택 개선 사항	현재 단위거처의 유형	행정자료 조사 항목
	현재 단위거처의 소유상태, 입주연도, 종전 단위거처의 유형, 종전 단위거처의 소유상태	신규항목
교육	재학 중인 학교수준, 학교이름, 최종학력, 전공영역(기술대학/종합대학 졸업자), 최종학력 이수 국가, 최종학력 이수연도, 숙련된 언어, 집에서 사용하는 언어(방언 포함)	1990년과 2000년 센서스 공통항목
교육개선 사항	직업관련 교육 이수여부(기술대학/종합대학 졸업생이 아닌 경우), 최종 이수 직업관련 자격 유형, 전공분야, 교육기관, 기본학위의 전공영역(졸업 후 자격을 취득한 경우)	신규항목
취업	경제활동상태, 직업상태, 직업, 산업, 근로소득, 수령 보너스액	1990년과 2000년 센서스 공통항목
	주당 실제 근로시간	신규항목
직업이동	현재 직장의 취업기간, 종전 직업, 종전 산업	신규항목
비경제활동 인구	종전 경제활동 여부, 일자리를 찾기 위하여 했던 활동, 일을 하지 않은 이유	1990년과 2000년 센서스 공통항목
통학/통근	일상적 통학 수단, 일상적 통근 수단	1990년과 2000년 센서스 공통항목
해외여행	지난 1년간 사업/여가활용 목적으로 해외여행을 한 회수 (말레이시아, 기타 국가)	신규항목
65세이상 고령자	재정지원의 주요 원천, 거동상태	신규항목

2. 싱가포르 해외국민의 집계

싱가포르는 강소국으로서, 글로벌리제이션과 지역화의 추세에 발맞추어, 해외체류 싱가포르 국민의 수가 끊임없이 증가하였다. 이러한 추세에 따라서, 해외체류 싱가포르 국민의 규모나 특성에 대한 관심이 크게 증가하여 왔다.

그러나 이러한 정보는 행정자료를 통해서는 쉽게 파악하기가 힘들다. 가구등록데이터베이스는 해외주소를 신고한 국민에 대해서는 약간의 정보를 제공하지만, 그것이 결코 완전한 정보라고 보기는 힘들다. 글로벌리제이션과 지역화의 전반적 추세는 이들을 정확하게 조사하는 업무를 더욱 더 어려운 과제로 만들어놓고 있다. 싱가포르 국민은 대단히 유동성이 강한 민족으로서, 취업 또는 유학을 목적으로 해외로 여행하는 빈도는 괄목할 수준으로 증가하여 왔다.

가. 1990년 센서스의 경험

1990년 센서스의 결과, 해외에 체류하는 싱가포르 국민은 36,179명으로 집계되었다. 이들은 주로 두 개의 집단으로 구분된다.

- a) 부재 가구원(AHM, Absentee Household Members): 가구는 국내 현지에 있는데, 가구의 구성원 곧 가구원이 해외에 체류하는 싱가포르 국민
- b) 가구 전체가 해외에 거주하는 가구원(POH, Persons living Overseas with their Household)센서스에서 수집되는 정보는 정책기획이나 정책입안에서 정부의 당면요구를 충족시킬 수 있는 것이라야 한다.

36,179명 중에서, 3,075명은 부재 가구원(AHM)이었으며, 5,426명은 (b) 가구 전체가 해외에 거주하는 가구원(POH)이었다. 부재 가구원에 대한 정보는 센서스 집계에서 가구원으로부터 정보를 획득할 수 있지만, 가구원 전체가 해외에 거주하는 가구원의 경우는 센서스 조사항목에 대한 정보를 수집하는 것이 간단하지 않았다.

1990년 센서스에서, 싱가포르 통계국은 싱가포르 직원을 두고 있는 대기업이나 다국적기업을 접촉하였다. 또 싱가포르 통계국은 싱가포르 고등위원회, 대사관, 무역대표부를 접촉하여, 학생이나 해외거주자를 조사하는데 협력을 당부하기도 하였다. 이러한 노력에도 불구하고, 사실상, 가구전체가 해외에

거주하는 가구원이 5,426명이 집계되었다. 당시 해외체류 싱가포르 국민은 36,179명이었는데, 이것은 사실상 과소집계라고 평가할 수밖에 없는 수치였다. 싱가포르 해외재단(SIF, Singapore International Foundation)은 당시 해외체류 싱가포르 국민을 10만 명 정도로 추정하고 있었다.

싱가포르 해외체류 국민 중에서 가구원이 모두 이주한 경우에는 해외공관에 등록하지 않은 경우가 많은 것으로 평가되었다. 이것은 싱가포르 해외공관이 입지하는 국가의 수도에서 멀리 떨어진 도시에 거주하는 사람들의 경우에 흔히 벌어지는 일이었다. 이러한 사람들을 해외공관이 제대로 파악한다는 것은 사실상 불가능한 일이라고 볼 수밖에 없다.

나. 해외 싱가포르 국민의 부재기간

해외체류 싱가포르 국민을 접촉하는데 있어서의 어려움만이 아니라, 개념상의 문제도 있었다. 1990년 센서스에서 해외체류 싱가포르 국민은 적어도 1년간 싱가포르를 떠나 있거나 떠나 있을 계획을 하고 있는 사람을 의미하는 것이었다.

그러나 이러한 정의는 사실상 점점 적합성을 상실하는 경우가 많았다. 해외 취업 또는 유학을 하는 싱가포르 국민은 다수가 규칙적으로 단기간 국내 여행을 한다. 결과적으로, 그들의 가구원은 시간의 상당부분을 해외에서 보내는 가구원에 대하여, “싱가포르 국내거주자”로 보고를 하는 경우도 있었다. 이러한 싱가포르 국민의 해외여행으로 인한 유동성 증가는 “해외체류 싱가포르 국민”의 개념이 변경되어야 할 필요성이 있다는 것을 의미한다.

1995년 중반의 일반가구조사(General Household Survey)에서, 최소 부재기간은 1년에서 6개월로 축소되어, “해외체류 싱가포르 국민”은 6개월 이상 싱가포르를 떠나 있는 모든 싱가포르 국민을 가리키는 것으로 되었다. 최소 6개월 이상의 부재기간이 2000년 센서스에서 “해외체류 싱가포르 국민”을 정의하기 위하여 사용되었다.

다. 2000년 센서스에 대한 접근방식

1990년 센서스의 경험에 의하면, 인구센서스는 해외체류 싱가포르 국민에 대한 정보를 획득하는 최고의 수단이 아니라는 것이 분명하다. 이것은 또한 유엔의 인구주택총조사 핸드북에 나온 평가에서도 다음과 같은 사실에 의하

여 뒷받침되고 있다.

“센서스 인구는 정의상 집계과정에서 한 국가의 모든 사람들을 포함하도록 되어 있다. 그러나 어떤 국가는 해외에 체류하는 국민들을 센서스 집계에 포함하는 경우가 있다. 이러한 맥락에서, 해외체류 내국인의 수를 정확하게 결정하는 것은 대단히 어려우며, 그들을 센서스에 포함하게 되면, 센서스 자료에 왜곡이 생겨날 수 있다는 점을 명심할 필요가 있을 것이다.”

개념상으로도 측정상의 많은 문제점이 있음에도 불구하고, 싱가포르의 독특한 사회경제적 특성 때문에 어떠한 방식으로라도 센서스에서 싱가포르 해외교민사회에 대한 정보를 획득할 필요성이 있다는 인식이 지배적이었다. 싱가포르 통계국이 센서스에서 해외체류 싱가포르 국민의 정보를 획득하고자 하는 목표는 다음과 같다.

- a) 글로벌리제이션의 충격을 측정하고, 그 추세의 강화를 모니터링하기 위하여 해외에 상주하는 싱가포르 국민의 추정치를 획득한다.
- b) 해외체류 싱가포르 국민의 기본특성을 최소한의 수준에서 파악하기 위하여, 그들의 체류목적, 체류기간, 그들의 기술 취업상태 등에 대한 특성에 대한 정보를 획득하는 것이 바람직하다.

2000년 센서스의 집계과정에서, 부재 가구원의 수와 기본특성에 대한 추정치를 획득하였다. 센서스 집계는 20% 표본조사이기 때문에, 표본조사에서 획득한 부재 가구원의 수는 승수요인을 적용하여 통계적으로 해석하지 않으면 안 될 것이다.

1990년 센서스와 마찬가지로, 2000년 센서스에서도 가구원이 모두 해외에 체류하는 부재 가구의 가구원에 대해서는 정확한 정보를 획득하는 것이 불가능했다. 이 수치는 현장 조사원이 이웃 등의 친지를 통하여 면접을 하였지만, 해외에 체류하는 싱가포르인 가구를 접촉하여 확인을 하는 것이 불가능하였기 때문에 과소 집계되었다고 할 수 밖에 없을 것이다. 가구원 전체가 부재하는 가구(POH)의 정보를 완전하고 정확하게 획득하는 과제는 당분간 수년간에 걸쳐서 싱가포르 센서스의 기획이나 준비에 엄청난 도전이 될 것이 분명하다.

3. 싱가포르 국내체류 외국인의 집계

싱가포르에서도, 국내노동력의 부족사태를 타개하기 위하여 해외에서 외국인 노동력을 수입하는 경향이 확대되고 있다. 싱가포르에 입국하는 외국인은 인력부(MOM, Ministry of Manpower)에 고용허가증을 신청하고, 취업자 비자(고급인력은 EPH, Employment Pass Holder 또는 비숙련인력은 WPH, Work Permit Holder)를 부여받는다. 고급인력의 경우나 가정부로 일하는 비숙련인력의 경우는 숙소를 제공받거나 임대주택에 거주하기 때문에, 집계과정에서 큰 문제가 발생하지는 않는다. 가정부가 아닌 비숙련 노동력, 가령 건설노동자의 경우는 대부분 비거주용 주택(단위거처) 곧 현지 생산라인/숙소에 거주하기 때문에 집계과정에 문제가 발생한다. 종전의 센서스에서, 이들 집단은 예외 없이 집계과정에서 누락이 심각한 문제로 비화하였다.

싱가포르에서 비숙련 노동력의 증가로 말미암아, 싱가포르 통계국은 2000년 센서스에서 이들을 누락이나 중복없이 집계하는 것이 필요하다고 판단하였다. 싱가포르 통계국은 일반가구에 거주하는 20%의 거주자와 마찬가지로, 이들 비숙련 노동력의 20%를 집계하여야 한다고 판단하였다. 이러한 권고안은 1999년 초반 센서스기획위원회에서 승인을 받았으며, 센서스법 제14조(1991년 수정판 제35장)는 비숙련 노동력을 고용하는 사업주가 센서스 조사원이 되는 것을 허용하기 때문에, 이러 외국인 노동자를 집계하는 접근방식은 사업주를 접촉하여 필요정보를 제공하는 것이다.

싱가포르 통계국은 20% 표본을 선정하기 위하여, 싱가포르 인력부의 협조를 얻어서 외국인 노동자를 고용하는 기업의 표본추출틀(sampling frame)을 구축하였다. 표본추출틀은 8,8000명의 비숙련 노동력을 고용하는 6,400개의 기업으로 구성되었다. 표본추출틀은 가정부가 아닌 비숙련 노동력을 고용하는 산업체를 대표하는 것이었다. 표본을 선정하고, 고용주의 연락처(주소, 전화번호)를 싱가포르 통계국의 사업체대장 시스템(Commercial Establishment Information System)에서 업데이트한 후에, 통계조사가 2000년 3월부터 2000년 8월까지 6개월 동안 우편조사로 실시되었다. 고용주에게는 사전에 인쇄된 기본 조사항목(성명, 외국인 고유번호, 민족/종족집단, 성별, 기술유형)은 물론 주택, 직업, 소득, 교육/기술, 통근수단 등에 대한 정보를 제공하고 업데이트하도록 요청을 하였다. 통계조사의 결과는 거주자의 주 표본과 통합되었으며, 이것을 통하여, 전체인구와 취업자 인구에 대한 종합적인 특성을 파악하는데 이용되었다.

4. 센서스 표본조사를 위한 지역구분

2000년 센서스의 20% 표본인 21만 8,000가구는 싱가포르 전역에 걸쳐서 계통적으로 추출된 것이다. 이들은 5개 지역, 곧 북부, 동북부, 동부, 중서부 지역으로 구분하여 현지실사 작업을 편리하도록 하였다. 각 지역에서 지역사무소(Regional Office)가 설치되었다. 각 지역사무소는 센서스 총감독 보조책임자(ACSS, Census Assistant Superintendent)가 관리하고, 레이저프린터, 고성능 컴퓨터, 이메일 시설을 갖추고 있어서 싱가포르 통계국과 안전한 방식으로 접속할 수 있도록 되어 있었다.

표본조사의 대상가구는 5개 지역에 대하여 각각 6개 그룹으로 구분되어, 모두 30개 그룹에 이르렀다. 이것은 5개월간의 자료수집기간에 각각의 하위 표본에 대하여 대표성이 있는 결과를 획득하기 위한 목적으로 규모가 비슷한 6개의 독립적인 하위표본을 선정하였다. 실용적인 관점에서, 6개의 그룹에 대하여 출발시점이 다르도록 하여, 인터넷 트래픽이 센서스 인터넷 집계 웹사이트에 균등하게 배분되도록 하기 위하여, 작업량을 분산 배치하였다. 싱가포르 통계국이 고용하는 CATI와 현지실사 조사원의 수는 이렇게 분산·배치된 작업을 효율적으로 처리할 수 있도록 하기 위한 것이다.

6개 그룹의 가구는 일차적으로 2주일 이내에 그들의 센서스 조사표를 인터넷으로 제출할 수 있다. 만약 2주일 이내에 응답이 완료되지 않으면, CATI 조사원은 그들을 3주째부터 접촉할 수 있도록 되어 있다. 해당가구는 3주에서 5주에 이르기까지 인터넷이나 CATI를 통하여 조사표를 반송할 수 있다. 그 후, 무응답 가구는 현지실사 시스템으로 이전되어 조사원이 해당가구의 가구주나 책임 있는 가구원과 1:1 면접을 진행한다. 이것은 6주째부터 10주째까지 진행될 수 있다.

5. 표본설계와 표본선정

2000년 센서스의 표본설계와 표본선정을 기술하고, 특히 표본추출의 현실적 고려사항, 원칙과 표본방법론, 20% 표본에서 획득한 자료의 신뢰성 문제를 언급한다. 표집오차와 관련된 문제는

가. 표본추출의 현실적 고려사항

가구등록데이터베이스는 인구의 총수와 기본특성에 대한 정보를 제공한다.

추가적인 상세정보를 얻기 위해서는, 비교적 규모가 큰 표본조사를 실시하는 것으로 충분하다. 싱가포르 통계청은 통계이용자의 자료에 대한 수요와 2000년에 포함된 조사항목의 범위를 고려하여, 20% 표본조사라면 대부분의 자료 요구를 충족시키는데 어려움이 없을 것이라고 판단하였다.

① 센서스 결과의 적시성

적시성에 대한 수요가 증대하고 있다. 2000년 센서스에서 조사항목의 정보를 모집단의 20% 표본조사에서 획득하고자 하기 때문에, 싱가포르 통계청은 자료수집과 자료처리가 4개월 이내에 완료되고, 센서스 기준일자로부터 6개월 이내에 공표될 수 있다.

② 비용합리성

표본은 모집단의 일부에 해당하기 때문에, 모집단을 전수 조사하는 것보다 표본조사를 통하여 정보를 획득하는 것이 비용효율성이 크다. 2000년 센서스의 경우, 싱가포르는 20% 표본조사와 함께 등록센서스를 추진하는 비용이 2,200만 달러가 소요될 것으로 추정하였다. 이것은 100%를 완전히 전수 조사하는 전통적 센서스 실시에 소요되는 총비용의 1/3에 해당한다.

나. 표본추출 방법론의 원칙

표본설계 단계에서, 상세한 연구가 진행되어서 2000년 인구센서스에 대한 표본추출 방법론이 결정되었다. 최종 표본설계는 다음의 요인을 고려하였다.

- a) 개발지도계획(도시계획지역)별로 변수의 광범한 추정치를 획득할 수 있어야 한다.
- b) 20%에서 25%로 표본규모를 쉽게 확대할 수 있도록 표본추출을 실시한다.
- c) 표본추출이 용이하도록 한다.

부록 IV. 독일의 2011년 등록센서스 관련 법률

A. 독일의 2011년 등록센서스 법률 및 법적 기반

1. 독일

2011년 센서스 준비법(Census Preparation Law)이 2007년 12월 13일 발효하였다. 2009년 7월 16일 2011년 등록기반 센서스법(Census Law)이 발효하였다. 이것은 2011년 센서스를 등록기반 방식으로 한다는 제1절의 규정과 통계법 개정을 포함한다.

2. 유럽연합

유럽연합 인구주택센서스 규정이 2008년 9월 2일에 정식으로 발효하였다.

3. 자료보호 및 통계비밀

모든 국가통계와 마찬가지로, 응답자의 마이크로 데이터는 엄격히 비밀을 보장한다는 핵심원칙은 센서스에도 그대로 적용된다. 해당 자료는 예외없이 통계적 목적에 기여해야 하며, 그것이 개인이나 공간기관에 자의적으로 배포되어서는 안 된다. 이것은 연방통계법에 의한 것이며, 추가적으로 연방자료보호법도 적용된다. 센서스 자료는 절대적으로 통계목적에 기여해야 하며, 민간기관이나 공공기관에 전달되어서는 안 된다. 이것은 연방통계법(Federal Statistics Law)에 의하여 규율되며, 연방자료보호법(Federal Data Protection Law) 규정이 적용된다.

1983년 연방헌법재판소(Federal Constitutional Court)가 1983년 12월 15일 센서스 위헌판결을 내린 후, 센서스 결과의 개인관련 마이크로 데이터가 국가통계기관에서 다른 공공기관으로 전송되어서는 안 된다고 판단하였으며, 따라서 어떤 마이크로데이터도 국가통계의 개인비밀 조항을 준수해야 하며, 그 자료가 외부에 누설될 수 없다는 것을 보장하지 않으면 안 된다.

B. 2011년 센서스 준비법(Census Preparation Law 2011)

건축물 · 주택 센서스를 포함, 등록부를 토대로 하는 2011년 센서스 준비를 위한 법률 (Gesetz zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011))

연방의회는 다음 법률을 결의한다.

내용개관

제1장 적용범위

제1절 적용범위

제2장 주소등록부, 건축물등록부, 출생지 출생국 정보의 작성

제2절 주소 및 건축물 등록부

제3절 장소 정보

제4절 측량청을 통한 자료의 전달

제5절 등록사무소를 통한 자료의 전달

제6절 연방노동청을 통한 자료의 전달

제7절 자료통합

제8절 식별번호

제9절 특별건축물

제3장 건축물 및 주택 등록부의 준비를 위한 자료의 전송

제10절 건축물 및 주택등록부에 대한 정보제공 의무자의 조사

제4장 공통의 규정 및 법률시행

제11절 비밀유지

제12절 접근 가능한 자료원의 포괄적 사용

제13절 자료전송

제14절 비용

제15절 해체

제16절 환경 및 주택통계 표본조사

제1장 적용범위

제1절 적용범위

본 법은 인구등록부와 여타 행정등록부에 포함된 자료의 이용과 보완적인 표본조사의 형태로 2011년에 실시될 인구센서스와 건축물/주택센서스를 준비를 위하여 주소등록부(address register)와 건축물등록부(building register)의 구축을 규정하기 위한 것이다,

제2장 주소등록부, 건물등록부, 출생지 및 출생국 정보의 작성

제2절 주소등록부와 건축물등록부

(1) 연방 통계청은 주소 및 건축물 등록부를 작성하여 센서스를 위한 준비를 하여야 하고, 주법(state law)에 따라 연방통계 작성업무를 수행하는 관청(곧, 주 통계청)은 주소 및 건축물 등록부의 수립에 협력하고, 센서스의 준비에 등록부를 이용할 수 있도록 한다.

(2) 주소 및 건축물 등록부는 아래의 업무에 기여한다.

1. 주소 및 건축물 센서스의 과정관리와 센서스의 일차적 통계조사의 경과조정
2. 센서스 실시와 함께 먼저 진행되는 표본조사의 준비와 기초자료 제공
3. 서로 다른 행정자료를 자료원(data source)로 하는 다양한 정보들을 취합하고, 인구, 주택, 건축물 등에 대한 자료를 센서스의 최종결과로 통합하는데 필요한 자료의 제공
4. 공간관련 분석체계의 개발과 센서스의 소지역 통계(small area statistics) 자료 작성을 위한 근거자료 제공

(3) 주소 및 건축물 등록부는 다음 정보를 수집하여야 한다..

1. 식별번호(order number)
2. 우편번호
3. 지역 또는 지자체 이름
4. 지역 또는 지자체 내 구역의 이름
5. 도로이름
6. 가옥번호(house number)

7. 주소 부대사항(address auxiliaries)
8. 건축물의 위치
9. 지자체의 행정코드
10. 지역 또는 지자체 내 구역의 코드
11. 도로코드(street key)
12. 지자체 고유의 도로코드
13. 좌표(값)와 정성적 정보(qualitative information)
14. 지자체의 규모
15. 건축물 기능
16. 건축물 상태
17. 주택의 수
18. 점유주택의 수
19. 주소 당(當) 일차 거주지(principal residence) 인구수
20. 주소 당 이차 거주지(secondary residence) 인구수
21. 주소 당 독일인 수
22. 주소 당 외국인 수
23. 주소 당 사회보험 가입의무 노동자의 수
24. 주소 당 실업자의 수
25. 조사지역의 특징
26. 표본의 특징
27. 주소 당 상이한 성씨(family names)의 수
28. 주소 당 변동률

(특수건축물에 대해서는 다음을 추가한다.)

29. 시설의 유형
30. 시설의 담당자, 소유자 또는 관리자의 이름과 주소
31. 특수건축물의 집계절차

(건축물 및 주택센서스의 보고 책임자)

32. 건축물 및 주택 소유자, 상속권자, 관리자 또는 처분권자의
성명,
33. 주소

(4) 주소 및 건축물 등록부의 작성은 등록센서스를 추진하기 위하여 늦어도 2010년 12월 31일까지 완료되어야 한다.

제3절 출생지 및 출생국 정보

(1) 연방 통계청은 제2절에 의거한 등록부와 별도로 출생지 내지 출생국가를 특정한 등록부를 작성하여야 하고, 연방 통계청과 주 통계청은 모두 이 정보를 센서스에 이용할 수 있다.

(2) 출생지 등록부는 출생지, 출생국가, 등록사무소(출생지), 출신국(독일 입국 전 거주 국가) 등의 정보를 수집한다.

제4절 측량청을 통한 정보의 전송

(1) 주법(state law)에 따라 측량 업무를 담당하는 각주의 측량청(Vermessungsbehörde, measurement authority)은 지도제작과 측량을 위하여 연방 통계청에 2007년 4월 1일 자로 작성된 다음의 전자 정보를 2008년 4월 1일까지 전송하여야 한다.

1. 식별 목적의 데이터레코드
2. 일의적인 데이터레코드 순번
3. 지자체 행정코드
4. 주 측지측량원이 부여한 지자체 내 구역별 코드
5. 주 측지측량원이 부여한 도로 코드
6. 주택번호
7. 주소 부대사항
8. 정성지표를 포함한 좌표(값)
9. 도로의 이름
10. 우편번호
11. 우편 지구의 이름과 부대사항

(2) 각 주의 측량청은 2008년부터 2010년까지 매년 4월 1일자로 제1절의 변경된 자료의 전자정보를 연방 측량청(Federal Office of Cartography and Geodesy)에 해당연도의 7월 1일까지 전송하여야 한다.

(3) 연방 측량청은 전송된 정보를 검토하고, 불완전한 정보를 보완하며, 최종적으로 완성된 정보를 연방 통계청에 전송한다.

제5절 등록사무소를 통한 정보의 전송

(1) 제2절에 따른 주소 및 건축물 등록부의 작성이나 제3절에 따른 출생지 정보를 작성을 위하여, 주법에 의거하여 인구등록을 책임진 기관(등록사무

소)은 인구등록부상의 모든 거주자에 대하여 다음 정보를 2008년 4월 1일 현재를 기준으로 주 통계청에 전송하여야 한다.

1. 지자체 코드와 현행 주소
2. (가용한 경우), 지자체 고유의 도로코드
3. 주택의 거주상태 (오직 하나인 거주지, 일차거주지, 이차거주지)

**오직 하나인 거주지-sole residence

일차거주지-primary residence

이차거주지-secondary residence

4. 주택전입일자
5. 등록사무소 등록일자
6. 성씨
7. 국적
8. 종전 주소
9. 가구주와의 관계
10. 출생년월일
11. 성별
12. 출생지
13. 출생국
14. 출생지(인구동태사무소)
15. 출발국(departure country)(독일 입국 전 거주 국가)

(2) 주 통계청에서 즉각적으로 항목 1에서 11까지의 정보를 12-15번의 정보와 별도로 분리하여 개인정보를 보호한다.

(3) 주 통계청은 제1항과 제2항에 의하여 작성된 정보를 등록사무소의 자료전송 마감일 이후 늦어도 4주 전에는 연방 통계청에 전자적으로 전송하여야 한다.

(4) 제1항 제1목의 번호 1에서 5까지와 7에서 11까지의 정보는 연방통계국이 표본조사 실시를 위한 기초자료로 별도 저장된다. 이들 자료는 연방 통계청과 주 통계청이 표본조사 계획이나 컴퓨터 예측과정을 개발하는데 응용된다.

제6절 연방노동청을 통한 정보의 전송

연방노동청(Federal Agency of Employment)은 사회보험의 적용대상인 노동자와 그렇지 못한 실업자와 관련하여, 2008년 3월 13일을 기준으로 2008년 4월 15일까지 다음의 정보를 연방 통계청에 전자적으로 전송하여야 한다.

1. 거주지, 우편번호, 지자체의 행정코드
2. 도로이름
3. 주택번호, 주소 부대사항
4. 종사상 지위(취업자, 실업자 구분)

제7절 자료통합

(1) 제5절 제1항 제1목의 1-11번과 제2항에 의거하여 전송된 정보는 제4절과 제6절에 의거하여 전송된 자료와 함께, 연방 통계청이 도로이름을 표준화하기 위하여 통합하여, 주소 참조 그룹(address reference group)으로 총괄 요약한다.

(2) 주 통계청은 업무책임상 모든 경우에 대해서 제1항에 의거하여 취합을 하는 과정에 접근한다. 주 통계청은 통합과정 특히 전송된 자료의 완전성과 매칭문제(matching problem)를 검토한다. 주 통계청은 불완전하거나 흠결이 있는 자료에 대하여 참고기준이 있는 주소들을 등록사무소에 전송한다. 등록사무소는 현재의 가용 자료를 가지고, 원래 전송된 자료가 완전한지 흠결이 있는지를 설명해야 한다. 만약, 그것이 흠결이 있는 자료인 경우에는 주 통계청은 관련 주소영역에 대하여 재차 자료를 전송하여야 한다. 검토결과는 주 통계청에 의하여 연방 통계청에 전송된다.

제8절 식별코드

(1) 주소, 건축물, 주택의 경우 예외 없이 식별코드가 부여된다. 식별코드는 지자체와 건축물에 대하여 일의적으로 연방 및 주 통계청에 의하여 부여된다. 식별코드는 제2절 제3항의 11번 항목인 도로키(street key)를 포함한다.

(2) 식별코드는 제7절에 의거한 자료 매칭이나 센서스에서 요구되는 행정 등록부 평가 자료와 보완적 조사를 통합하는데 적용된다.

제9절 특별건축물

(1) 특별건축물(special buildings)의 인구에 관한 자료를 수집하기 위하여,

주 통계청은 주소 및 건축물 등록부에 대하여 다음의 특성을 보완한다.

1. 시설의 종류
2. 시설의 담당자, 소유자, 관리자의 성명과 주소

(2) 주 통계청은 센서스가 집계하는 특별건축물의 완전성이나 제1항에 언급된 특성의 품질을 확보하여야 한다.

(3) 특별건축물은 지자체 숙박시설, 시설 관련 숙박시설, 비상 숙박시설, 호스텔, 외국군대, 외교관 또는 직업적 영사업무 대표자가 거주하는 건물을 의미한다. 지자체 또는 시설 관련 숙박시설은 단체 손님의 장기적 숙박이나 공공 목적에 기여하기 위한 시설을 말한다. 비상 숙박시설은 잡이 없는 가출자(homeless)가 등록할 수 있는 주소지가 된다.

제3장 주소 및 건축물 등록부의 준비를 위한 자료의 전송

제10절 주소 및 건축물 등록부 작성을 위한 정보제공자의 조사

(1) 주소 및 건축물 센서스를 실시하기 위하여, 주 통계청은 주소 및 건축물 등록부에 성씨, 이름, 또는 기타특성, 주소 및 건축물의 소유자, 상속권자, 관리자, 기타 처분권자의 주소를 보완하여 특정해야 하고, 인구등록사무소는 주 통계청의 요청에 따라 전자정보를 제공해야 한다.

(2) 등록사무소는 토지세 부과, 등기부나 토지대장의 운영 업무를 맡고 있는 기관이다.

제4장 공통의 규정 및 법률시행

제11절 비밀유지

개별정보의 비밀유지는 연방통계법에 의하여 규정된다.

제12절 접근 가능한 자료원의 포괄적 사용

이 법률의 시행을 위해 연방 및 주 통계청은 접근 가능한 자료원을 포괄적으로 이용할 수 있다.

제13절 정보의 전송

이 법률과 관련된 공무원은 기술적이고, 제도적인 방식을 통해 자격 없는 사람이 정보를 읽거나, 무단으로 복사하거나, 변조하지 않는 것을 보장하여야 한다.

제14절 비용

연방 통계청에 대한 정보의 전송비용(tranmission costs)은 별도로 보상하지 아니한다.

제15절 삭제

취득된 정보는 정해진 기간이 지나면 데이터베이스에서 삭제한다.

제16절 환경 및 주택통계 표본조사

- (1) 제2절 제3항 1번 20, 26, 29번의 자료는 2011년 센서스법 제6절 제2항 1번 b-f에 따른 자료와 함께 환경 및 주택통계(environmental and housing statistics) 표본조사의 추출 근거자료로 이용한다.
- (2) 제2절 제3항 1번 32, 33번의 자료는 제1항에서 선정된 주택주소와 함께, 표본조사를 준비하고 실시하는데 이용한다.
- (3) 표본조사가 더 이상 실시될 필요가 없는 경우, 제2항에 의한 자료는 삭제되어야 하며, 늦어도 2017년 3월 9일까지는 삭제되어야 한다.

2007. 12. 8. 독일센서스 준비법

C. 2011년 센서스 법(Census Law 2011)

2011년 등록기반 센서스에 관한 법률(Gesetz über den registergestützten Zensus im Jahre 2011 Zensusgesetz 2011)

2009년 7월 8일

내용개관

제1장 일반규정

- 제1절 센서스의 종류, 목적, 보고시점
- 제2절 조사단위와 개념의 규정

제2장 조사와 자료통합, 가구생성

- 제3절 등록사무소와 상급 연방관청을 통한 자료의 전송
- 제4절 연방노동청을 통한 자료의 전송
- 제5절 금융 및 개인 통계법에 따라 정보제공 의무가 주어진 부서를 통한 자료의 전송
- 제6절 건축물 및 주택 센서스
- 제7절 가구대상 표본조사
- 제8절 특별지역 주소에 대한 자료수집
- 제9절 데이터 레코드의 통합과 가구생성

제3장 조직

- 제10절 조사담당부서
- 제11절 조사요원
- 제12절 집중적 자료처리 시스템
- 제13절 식별번호

제4장 센서스 결과의 보호와 품질에 대한 조치

- 제14절 거주공간(단위거처)이 있는 건축물과 주거용 숙박시설의 주소 보완 조사
- 제15절 중복사례 조사
- 제16절 불일치 사례를 해명하기 위한 조사
- 제17절 센서스 결과의 품질평가

제5장 정보제공 의무와 데이터 보호

- 제18절 정보제공 의무와 정보제공 형태
- 제19절 삭제
- 제20절 자료전송
- 제21절 공공영역의 정보

제22절 상급연방관청, 주 관청, 지자체(또는 지자체연합) 통계부서에 대한
제표자료와 마이크로데이터의 전송

제6장 결론

제23절 건축물, 주택, 인구의 표본추출 근거제시

제24절 연방 통계청으로의 자료전송 비용

제25절 재정분배

제1장 일반규정

제1절 센서스의 종류, 목적, 보고시점

(1) 연방 및 주 통계청은 2011년 5월 9일을 기준시점(보고시점)으로 하여, 국가통계로서의 인구, 건축물 및 주택 센서스(population, building, and housing censuses)를 실시한다.

(2) 센서스에 필요한 자료는 다음과 같은 방법으로 수집된다.

1. 주법에 의하여 등록업무를 관장하는 부서(등록사무소)와 상급 연방관청의 데이터 전송 (제3절)
2. 연방노동청의 데이터 전송 (제4절)
3. 금융 및 개인통계법에 의하여 정보제공 의무가 주어진 부서의 데이터 전송 (제5절)
4. 건축물 및 주택자료의 획득을 위한 조사(제6절)
5. 인구에 대한 자료의 품질을 확보하고 보완자료를 파악하기 위한 표본 조사 실시 (제7절)
6. 지자체 조사, 시설 조사, 비상 조사, 호스텔 또는 유사 시설에 주소를 둔 거주자의 데이터 전송 (제8절)
7. 거주공간(단위거처)이 있는 건축물이나 거주자가 있는 주택의 주소에 대한 보완적 조사 (제14절)
8. 센서스 결과의 품질을 평가하기 위한 조사(제17절)

(3) 센서스는 다음 목적에 기여한다.

1. 연방, 주, 지자체 법정 인구의 확정과 두 센서스 시점 간의 법정인구를 추정하기 위한 근거자료

2. 인구구조, 특히 연방, 주, 지자체에서 인구, 경제, 사회, 주택, 공간질서, 수송, 환경, 노동시장 등의 정치적 결정을 위한 토대자료와 국가통계 시스템의 기본 자료의 획득
3. 인구주택센서스와 관련하여, 유럽의회(European Parliament)의 규정 No 763/2008과 2008년 7월 9일의 권고안에 의거한 보고책임의 완수

제2절 조사 단위와 개념의 규정

(1) 인구센서스의 조사 단위는 개인과 가구이다. 인구에는 다음을 포함한다.

1. 등록법 규정에 따라 보고시점까지 보고책임이 있는 사람
2. 연방군대, 연방경찰, 외교관(외교업무 관련 법률 제2절 참고)으로 외국에서 활동하고 있는 사람이나 이들의 부양가족

그 인구에는, 외국의 군대, 외교관, 직업적인 영사업무 종사자는 포함하지 않는다. 가구는 같이 거주하는 모든 사람을 포함한다. 혼자서 거주하는 사람은 1인 가구(one person household)를 구성한다. 복수의 주소(거주지)를 둔 사람은 각각의 장소에서 조사하여, 하나의 가구에 할당되도록 한다.

(2) 한 지자체의 법정 인구는 그 지자체에 상주지(usual residence)를 두고 있는 주민의 총수이다. 상주지는 등록법 규정에 따라, 하나의 거주지를 갖거나, 1차 거주지로 등록된 장소를 말한다. 외국에서 활동하고 있는 연방군인, 연방경찰, 외교관 그리고 그들의 가족 구성원은 거주지 대신에 거주국(country of residence)이 부여된다.

(3) 건축물 및 주택 센서스의 조사 단위는 거주공간(단위거처)이 있는 건축물, 거주자가 있는 숙박시설과 주택을 가리킨다. 여기에, 외국 또는 외국의 군대, 외교관 또는 직업적인 영사업무를 수행하는 사람들이 거주하는 주택은 제외하며, 외국과의 조약에 의하여 사용 불가능한 건축물, 숙박시설, 주택도 제외한다.

(4) “주택”은 외부로부터 차단되어, 거주목적으로 규정된, 일반적인 병렬적 공간으로서, 독자적인 가계생활을 가능하게 하며, 보고시점에 영업목적으로 완전히 이용되지 않는 것을 가리킨다. 거주목적으로 별도 증축된 지하

실이나 지상 공간(예: 다락방)도 주택에 속한다. 주택이라고 해서 반드시 부엌이나 미니주방을 포함할 필요는 없다. 주택은 입주자가 다른 가구의 주택으로 들어가지 않고 그 주택에 직접 출입할 수 있는 계단, 현관 또는 외부에서 독자적인 출입구가 있어야 한다.

(5) 특별지역(special areas)은 지자체 숙박시설, 시설관련 숙박시설, 비상 숙박시설, 호스텔 등의 유사 숙박시설 가리킨다. 지자체 숙박시설이나 시설관련 숙박시설은 일반적으로 장기적인 숙박이나 특수한 숙박목적을 가진 일련의 단체 손님들에게 봉사하는 시설을 말한다. 집이 없는 가출자(homeless)가 등록된 주소는 비상숙박시설(emergency accommodation)로 조사한다. 민감한 특별지역은 거주관련 정보가 이해당사자에게 사회적 불이익의 위험을 야기할 수 있는 영역이다. 병원, 요양원, 기타 시설에 입소하고 있는 사람들의 경우 등록의무 때문에 신고한 주소는 특별지역에 할당된다.

(6) 시군과 도시일부의 하위구역에 대한 조사와 관련해서는, 지역상황과 인구통계법 제5절에 규정된 2009년 12월 31일의 인구추계를 토대로 한다. 표본추출 시점까지 해당 주의 행정구역이 변경 조정되는 경우에는 예외로 한다.

제2장 조사와 자료통합, 가구생성

제3절 등록사무소와 상급관청을 통한 자료의 전송

(1) 2007년 12월 8일에 공포된 센서스 준비법(Census Preparation Law) 제2절에 따라, 주소 및 건축물 등록부를 구체화하고, 2011년의 등록기반 센서스를 실시하기 위하여, 등록사무소는 주 통계청(State Statistical Office)에 등록인구에 대하여 다음 항목에 관한 전자자료(electronic data)를 전송하여야 한다.

1. 인구등록부의 순번 또는 식별코드(order number)
2. 성씨, 혼전 이름(여자의 경우), 이름
3. 도로이름, 도로키, 주택번호, 주소 부대사항(address auxiliaries)
4. 거주지, 우편번호, 지자체 행정코드
5. 출생년월일
6. 등록사무소, 출생등록번호

7. 출생지와 가족관계의 특성
8. 외국 출생자의 경우: 출생국가
9. 성별
10. 국적
11. 가구주와의 관계
12. 거주상태(하나뿐인 거주지, 일차적 거주지, 이차적 거주지)
 **하나뿐인 거주지-sole, exclusive residence
 일차적 거주지: primary place of residence
 이차적 거주지: secondary place of residence
13. 전입자의 경우, 종전 지자체의 주소와 거주상태
14. 지자체 내 마지막 거주 주택 주소
15. 주택 입주일자
16. 지자체 전입일자
17. 해외 전입자의 경우: 출발국(departure country)
18. 등록사무소 등록일자
19. 거주상태 변경일자
20. 배우자 또는 등록된 동거자의 성씨, 종전 이름(여자의 경우, 이름의 변경이 있는 경우), 이름, 출생년월일
21. 미성년자의 성씨, 종전 이름, 이름, 출생년월일, 식별번호와 법적 후견인(legal representative)의 성씨, 종전 이름, 이름, 출생년월일, 식별번호
22. 마지막 혼인년월일 또는 마지막 동거관계 등록 년월일
23. 마지막 혼인 또는 마지막 동거관계 해소 년월일
24. 주택 제공자의 주소
25. 인구등록부에 자발적 등록여부에 관한 정보
26. 전송장애 및 장애의 이유
27. 종교단체에 대한 소속여부

(2) 등록사무소는 제1절에 따라 모든 경우, 다음의 시점 이후 4주 이내에 자료를 전송하여야 한다.

1. 2010년 11월 1일 마감일
2. 보고시점
3. 2011년 8월 9일

(3) 외국에 파견된 사람의 경우,

1. 연방 군대에 복무하는 사람
2. 연방 군대를 위하여 활동하고 있는 사람
3. 경찰업무에 종사하는 사람
4. 외교관 업무에 종사하는 사람

위의 경우와 독일 등록사무소에 등록되어 있지 않는 그들의 가족 구성원은 보고시점 이후 12주 이내에 연방 통계청에 다음 개인관련 정보를 전송하여야 한다.

1. 성씨, 종전 이름(여자의 경우), 이름
2. 성별
3. 출생년월일
4. 현재 체류국의 이름
5. 국내 출발 후 외국체류 개시 년월일

(4) 제3항 제1목의 1-2번에 대한 레코드의 전송은 국방부(Defense Ministry)의 소관이며, 제3항 제1목의 3번에 대한 레코드의 전송은 내무부(Interior Ministry)의 소관이며, 제3항 제1목의 4번에 대한 레코드의 전송은 외무부(Foreign Ministry)의 소관이다.

(5) 제2항 1번의 전송자료는 센서스 실시를 위한 보조 자료로 활용된다.

(6) 제2항 2-3번에 의하여 전송된 자료를 제1항 4-5번(출생년월일의 월과년의 레코드), 7-12번, 15-19번의 자료는 식별정보(identification information)로, 제1항 1-3번, 5번(출생년월일의 일자 레코드), 6번, 13번, 14번, 20-26번은 보조 자료(auxiliary data)로 사용한다.

(7) 주 통계청은 제1항의 자료를 완전성 검토 이후 제2항에 언급된 시점 이후 늦어도 8주 이내에 연방 통계청에 전송한다.

제4절 연방노동청을 통한 자료의 전송

센서스 실시를 위하여, 연방노동청은 데이터베이스에서 보고시점에 가장 근접한 보고일자에 맞추어 다음 자료를 전송하여야 한다.

1. 사회보험 가입의무가 있는 정규직 취업자와 가입의무가 없는 임시직 취업자의 경우, 보고시점 이후 늦어도 6개월 이내에 다음 항목의 자료를 제출한다.

- (a) 근로장소(지자체 행정코드)
- (b) 산업분류(industrial classification)
- (c) 사업장의 영업번호(business operation number)
- (d) 직업교육 (vocational education and training)
- (e) 종사 직업
- (f) 취업상태(정규직 또는 임시직 취업)

2. 실직자, 구직자 또는 비 활동인구(non-activity population)로 등록된 모든 사람에 대해서, 보고시점 이후 늦어도 3개월까지 다음 항목의 자료를 제출한다.

- (a) 상태(실직자, 실직은 아니지만 구직활동, 非 활동인구)
- (b) 최종 이수학력
- (c) 최종 직업교육

3. 고용촉진 프로그램 참가자로 등록된 모든 사람에 대해서, 보고시점 이후 늦어도 3개월까지는 다음 항목의 자료를 제출한다.

- (a) 프로그램의 종류 (유업자-gainful employment 파악에 의미가 있는 범위에서)
- (b) 최종 이수학력
- (c) 최종 직업교육

4. 1-3번에 언급된 사람들의 경우, 각각의 집단에 대하여 1-3번에서 언급된 기간에 대한 보조 자료로서 다음을 제출한다.

- (a) 거주지, 우편번호, 지자체 행정코드
- (b) 도로이름, 주택번호, 주소 부대상항
- (c) 성씨와 이름
- (d) 성별
- (e) 출생년월일

제5절 금융 및 개인 통계법에 따라 정보제공 의무가 주어진 부서를 통한 자료의 전송

금융 및 개인 통계법의 제11절 제2항 3번에 따라 정보제공 의무가 주어진 부서는, 그 부서가 금융 및 개인 통계법 제2절 제2항 10번에 따라 연방정부가 명목자본 또는 투표권의 과반 이상을 가지고 참가하는 연방관청이나 기타 조사단위의 경우에는, 금융 및 개인 통계법 제12절 제2항에 언급된 조사단위의 직접적 서비스 관계에 있는 사람들에 대하여 보고시점 이후 3개월 이내에 다음 자료를 전송한다.

1. 조사항목

- a) 근로 장소의 지자체 행정코드
- b) 노동청에서 데이터 전송을 위하여 부여한 사업장의 영업번호 또는 사업장의 산업분야
- c) 국가 책임영역, 지자체 책임영역 또는 지자체 재정시스템의 상품번호 (product number)
- d) 조사단위의 이름 또는 식별부호

2. 보조 자료

- a) 거주지, 우편번호, 지자체 행정코드
- b) 거리이름, 주택번호, 주소 부대사항
- c) 성씨 및 이름
- d) 출생년월일
- e) 성별
- f) 서비스의 범위
- g) 보고자 또는 서비스담당자 번호

주 통계청은 위의 금융 및 개인 통계법 제2절 제1항의 조사단위의 직접적 서비스 관계에 있는 사람들에 대하여 제1항에 정해진 기간까지 소정의 자료를 연방 통계청에 전송한다.

제6절 건축물 및 주택 센서스

(1) 센서스를 실시하기 위하여, 주 통계청은 보고시점까지 건축물과 주택에 대하여 우편조사를 실시한다.

(2) 조사항목에는

1. 건축물에 대해서
 - a) 지자체, 우편번호, 지자체 행정코드
 - b) 건축물의 유형
 - c) 소유관계
 - d) 건축물의 재료
 - e) 건축연도
 - f) 난방방식
 - g) 주택(거주공간 또는 단위거처)의 수
2. 주택에 대해서
 - a) 용도
 - b) 소유관계
 - c) (아는 경우) 등록의무가 없는 사람의 주택
 - d) 주택규모
 - f) 화장실 여부
 - g) 목욕탕 및 샤워장 설치 여부
 - h) 총방수

(3) 보조 자료의 경우

1. 정보제공 의무자의 성씨, 종전 이름, 이름, 주소
2. 정보제공 의무자 또는 재조사에 응답할 수 있는 사람의 팩스번호
3. 주택(단위거처)당 2명 이하의 주택이용자 성씨와 이름
4. (아는 경우) 주택 당 거주자의 수
5. 도로이름, 주택번호, 주택의 주소 부대사항

제7절 가구대상 표본조사

(1) 주 통계청은 보고시점까지 가구를 대상으로 표본조사를 실시한다. 표본 조사는 다음 목적에 기여한다.

1. 인구 1만 이상의 지자체, 인구 40만 이상의 도시(구역별 인구가 평균 20만 정도)에 대해서, 인구등록부에 등록된 사람이 해당 주소에 거주하고 있는가, 인구등록부에 등록되지 않은 사람이 어떤 거주지에 살고 있는가, 법정 인구가 최고 0.5%의 상대 표준오차의 범위에 있을 정도로 정확한가를 평가한다.

2. 인구 1만 이상의 지자체, 인구 40만 이상의 도시(구역별 인구가 평균 20만 정도)에 대해서 행정목적의 등록부에서 획득 불가능한 센서스 조사항목을 파악한다. 여기서 정확성의 목표치는 해당 지자체 또는 지역 단위 주민수의 최고 1%에 해당하는 상대적 표준오차(relative standard error)의 범위이다.

지자체나 도시지역의 인구를 확정하는데, 사람의 거주상태에 대하여 인구 등록부에서 전송된 자료의 보정결과는 포함하지 않는다..

(2) 제1항 제2목의 품질점검을 근거로 요구되는 표본조사의 추출률(sampling rate)은 모집단의 10%를 초과해서는 안 된다. 연방정부는 제1절 제3항과 제7절 제1항의 품질점검 목표에 도달하기 위하여 연방의회의 동의를 얻은 법 규정으로 표본조사의 과정과 범위를 규정한다. 해당 법 규정 초안은 연방의회에 2010년 3월 15일까지 제출되어야 한다.

(3) 표본조사의 조사단위는 주소 및 건축물 등록부에 따른 거주공간이 있는 주소(단위거처)이다. 주소 및 건축물 등록부에 신규로 등록된 거주공간이 있는 주소가 표본추출과 보고시점 간에 새로이 발견되면, 보완적 표본추출(additional sampling procedure)이 이루어져야 한다. 제1목과 제2목에 따른 표본조사는 특별지역(special areas)의 경우에는 제8절 제5항의 규정에 따라서 실시된다. 표본선정은 주소 및 건축물 등록부를 토대로 하여, 수리적 확률법칙(mathematical random principle) 의거하여 층화추출(stratified sampling)을 실시한다. 표본추출에는 센서스 준비법의 제5절 제4항에 따라 표본선정 데이터베이스에 입력된 레코드와 이 법의 제3절 제1항에 따라 등록사무소에서 전송된 자료를 응용한다. 표본조사의 표본은 제1항에 언급된 표본조사의 두 가지 목표에 동시에 도달할 수 있도록 선정되어야 한다. 표본선정은 지자체 수준에서 1만 명 이상의 인구를 가진 지자체, 지자체 내 구역 수준에서 1만 명 이하의 지자체, 도시 내 구역 수준의 인구가 평균 20만 명 수준이고 최소 인구가 40만 명인 도시에 대해서 추출작업이 이루어져야 한다.

(4) 조사항목은 다음과 같다

1. 거주상태
2. 성별

3. 국적
 4. 출생년월일
 5. 가구주와의 관계
 6. 혼전 거주 지자체
 7. 응답자 또는 부모가 1955년 12월 31일 이후 독일에 이주한 사람의 경우, 외국의 종전 거주지, 응답자 또는 부모의 독일 이주연도
 8. 가구원수
 9. ILO (International Labor Organization)의 노동력 개념(labor force concept)에 의거한 경제활동상태, 경제활동을 하지 않는 경우에는 마지막 활동에 일치하는 레코드, 비 경제활동 인구나 15세 미만의 모든 사람에 대해서는 보고시점 주일에 주된 활동상태를 기입함
 10. 고용상태(종사상 지위)
 11. 종사 직업
 12. 사업장의 산업분류
 13. 사업장 주소(지자체 수준)
 14. 주 활동상태
 15. 최고학력
 16. 최고 직업교육 및 훈련
 17. 실제 취학연수
 18. 종교단체 소속여부
 19. 특정종교 신앙여부, 종교관 또는 세계관(수니파 이슬람교, 시아파 이슬람교, 알레비파 이슬람교, 불교, 힌두교 등의 신앙심과 세계관)
- (5) 보조항목
1. 성씨와 이름
 2. 건물 내 주택의 주소와 위치
 3. 생년월일
 4. 정보제공 의무자 또는 재조사에 응답할 수 있는 사람의 팩스번호
 5. 활동상태에 있는 사람의 경우에는 보고시점 주일의 주 활동상태
- (6) 조사요원은 보고시점의 12주 이내에 표본조사를 완결하여야 한다. 여기에 상당한 이유가 있는 경우에는 예외로 한다.

제8절 특별지역 주소에 대한 조사

- (1) 주 통계청은 특별지역의 모든 주소에 대하여, 그 곳에 거주하는 사람들

을 확정한다. 이를 위하여, 그 곳에 거주하는 모든 사람들에 대하여 다음 자료를 수집한다.

1. 조사항목

- a) 출생연월
- b) 성별
- c) 가구주와의 관계
- d) 국적
- e) 거처입소일 또는 숙박개시일
- f) 출생국
- g) 해당자가 제2절 제2항 제4-6목이 정한 가구 주소에 실제로 살고 있는가 여부
- h) 거주상태

2. 보조자료

- a) 성씨, 종전 성씨(여자의 경우), 이름,
- b) 출생년월일
- c) 출생지
- d) 국적
- e) 거처입소일 또는 숙박개시일
- f) 출생국
- g) 해당자가 제2절 제2항 제4-6목이 정한 가구 주소에 살고 있는가 여부
- h) 거주상태

(2) 제1항에 따라 확정된 사람들에 대해서, 제3절 제1항에 따라 전송된 자료와 매칭작업(matching procedure)을 수행한다. 주 통계청은 제8절 제1항에 따라 작성된 자료를 가지고, 일차적 거주지와 이차적 거주지를 가진 사람이 어떤 거주지에서 조사되어야 할 것인가를 결정한다. 등록사무소에 해당 조사항목(survey topic)을 재조사하는 것은 허용되지 않는다.

(3) 제1항 제2목 1-g번에 따라 특정 가구에 거주하지 않는 특별지역의 사람들에 대해서는, 제1항 2번에 의하여 보조 자료에 의한 매칭작업을 즉각적으로 중지한다.

(4) 민감한 특별지역(sensitive special areas)에서, 건축물 및 주택 센서스에서 제6절 제2항의 조사항목과 성씨, 혼전 성씨(여자의 경우), 이름, 주소, 정보제공 의무자의 팩스번호만 조사한다.

(5) 민감한 특별지역에서, 제7절에 따른 표본조사는 실시하지 않는다. 제7절에 따라 선정된 나머지 특별지역에서는 거기에 거주하는 모든 사람들이 제7절의 4-5항에 대한 질문에 대하여 응답을 해야 한다.

제9절 특별지역 주소에 대한 조사

(1) 인구통계 정보와 경제통계 정보의 통합 데이터 레코드를 생성하기 위하여, 연방 통계청은 제8절, 제15절, 제16절에 따른 조사와 자료평가에 주의 하면서, 제3-5절의 데이터 레코드를 통합한다.

(2) 인구등록부의 중복사례(duplicate cases)를 확정하고, 제7절에 따른 보완적 조사항목에 대한 통합 데이터 레코드를 완성하기 위하여, 주 통계청은 제12절 제4항 제3목의 참고자료를 이용하여 제1항의 자료와 제8절의 자료를 통합한다.

(3) 주 통계청은 건축물 및 주택 센서스에서 전송된 데이터 레코드(제6절)와 제1항과 제2항에 따른 통합 데이터 레코드를 제12절 제2항, 제4-7항에 유념하여 통합한다. 각주의 통계청은 통합 데이터를 바이에른 주 통계청(Bavarian State Office for Statistics and Data Processing)으로 전송한다. 바이에른 주 통계청은 통합 데이터를 다음 특성에 따라 개인별로 주택에 매칭작업을 수행하여, 가구 데이터 레코드를 합성한다.

1. 건축물 및 주택 센서스 조사항목

- a) 주택 당 2명 이하의 주택이용자 성씨와 이름
- b) (아는 경우) 주택 당 거주자 수
- c) 이용형태
- d) 주택규모
- e) 총방수

2. 등록부 자료 항목

- a) 등록부의 식별번호(order number)
- b) 성씨, 종전 성씨(여자의 경우), 이름
- c) 출생년월일
- d) 성별
- e) 국적
- f) 가구주와의 관계
- g) 거주상태(하나뿐인 거주지, 일차적 거주지, 이차적 거주지)
- h) 전입자의 경우, 종전 지자체의 주소와 거주상태
- i) 지자체 마지막 거주 주택 주소
- j) 주택입주일자
- k) 지자체 전입일자
- l) 해외에서 전입 여부
- m) 배우자 또는 등록 동거인의 성씨, 종전 성씨(여자의 경우), 출생년월일
- n) 자녀와 법정 후견인의 성씨, 종전 성씨(여자의 경우), 출생년월일 식별번호
- o) 마지막 결혼 일자 또는 마지막 등록 동거관계의 설명
- p) 마지막 결혼 또는 마지막 등록 동거관계의 해소 일자
- q) 주택 제공자의 주소
- r) 등록부에 자발적 등록 여부에 관한 정보

제3장 조직

제10절 조사담당부서

- (1) 제6절, 제8절, 제14절, 제15절, 제16절에 따라 자료를 수집하기 위하여, 각 주는 조사담당부서를 설치할 수 있다. 이 법에 의하여 주 통계청이 수행해야 할 업무는 조사담당부서에 위탁될 수 있다.

- (2) 조사담당부서는 공간, 조직, 인원 등에서 다른 행정부서와 분리하여 설치한다. 이것은 조사담당부서의 관련문서 정보가 다른 목적에 사용되어서는 안 된다는 것을 확실히 보장하기 위한 것이다. 또, 조사담당부서에서 활동하고 있는 사람은 통계비밀을 준수하고, 조사과정에서 정보제공자에 대하여 획득한 지식을 외부에 발설하지 않는다는 것을 보장하기 위한 것이다. 조사담당부서에서 활동이 종료된 후에도 동일한 의무가 부과된다.

제11절 조사요원

(1) 제6-8절과 제14-17절에 따라 자료를 수집하기 위하여, 조사요원은 연방 통계법 제14절에 의하여 임명된다. 조사요원은 주 통계청이나 조사담당부서에 의하여 선발되어 임명된다.

(2) 연방정부와 주정부는 각주나 조사담당부서의 통계직원을 핵심관리자(core office)로 지명하며, 그들은 중요한 공무활동이 중단되지 않도록 조사요원 활동업무에서 제척한다. 특정 개인이 조사요원으로서 임명되면, 조사요원으로서의 활동 수행이 의무화된다. 임명 제척에는 건강이나 여타의 중대한 사유가 있을 때만 가능하다. 조사요원으로서 활동수행을 위한 추가적인 시민의 소환 가능성은 주법에 의하여 규정된다.

(3) 조사요원은 연방통계법의 통계비밀을 준수하고, 조사활동 중에 숙지한 사실에 대하여 외부에 발설하지 않는다는 것을 문서로 약속한다. 이러한 의무는 조사활동이 종료된 후에도 지속된다. 조사요원은 자신의 주택이 있는 지역에 임명되지 않는다. 직업상의 이유든 다른 이유이든, 특정인이 조사활동에서 숙지한 사실이 정보제공자에게 부당하게 이용될 수 있는 염려가 있는 경우에는, 조사요원으로 임명될 수 없다.

(4) 조사요원이 공익서비스(pro bono public) 차원에서 임명된 경우에는, 조사활동을 소득세법 제3절 12번 제2목의 개인비용 충당금(personal expense allowance)으로 면세 처리한다.

(5) 제6절에 따른 조사의 경우, 조사요원은 다음 목적을 위하여 임명된다.

1. 제18절 제2항의 정보제공자를 확정하기 위하여
2. 불완전하거나 자가당착적인 응답을 하는 경우, 제18절 제2항 제8목에 따라 재조사를 실시하기 위하여

(6) 제7절에 의한 가구조사에서, 조사요원은 제8절 제1항 제2목 1번의 자료와 제7절 제5항에 대한 보조 자료를 구두로 요구한다. 조사요원은 이 자료를 조사표에 직접 또는 전자적으로 기입한다. 정보제공자가 이해하는 경우, 조사표에 추가적으로 기입할 때도 동일하다.

- (7) 제8절에 따른 가구조사에서 조사요원이 임명된다. 민감하지 않은 특수지역(special areas)에서 조사요원은 제8절 제1항 제2목 1번 (a)와 (b)에 대한 조사항목과 제8절 제1항 제2절 (a)와 (b)에 대한 보조 자료의 요구를 구두로 전달한다. 조사요원은 이 자료를 조사표에 직접 기입한다. 정보제공자가 이해하는 경우, 조사표에 추가적으로 기입할 때도 동일하다.
- (8) 거주공간(단위거처)이 있는 건축물이나 거주자가 있는 숙박시설의 주소에 대한 보완적 조사를 하는 경우, 제14절 제3항에 의거한 점검조사(inspection)를 위하여 조사요원이 임명된다.
- (9) 중복사례(duplicate cases)(제15절 제3항)를 점검하는 경우에, 특히 서면조사(survey in written form)가 성공적으로 실시되지 못할 개연성이 있는 경우에 조사요원이 임명된다. .
- (10) 제16절의 매칭 불일치 사례를 해명하기 위하여, 조사요원은 제16절 제2목 1번의 (a), (b), (f)의 자료와 제16절 제2목 2번의 보조 자료를 구두로 요구할 수 있다. 조사요원은 이 자료를 조사표에서 직접 또는 전자적으로 기입할 수 있다. 정보 제공자가 이해하는 경우, 조사표에 추가적으로 기입할 때도 동일하다.
- (11) 조사요원은 제6절, 제8절, 제15-17절의 조사에서 조사활동을 지원하기 위하여, 해당 주소에 대한 축소된 인구등록부 명부를 지참한다. 이 명부는 해당 주소 아래 등록된 사람에 대하여, 성씨, 종전 이름(여자의 경우), 이름, 이름 부대사항(1세, Sr, Jr 등), 성별, 생년월일, 국적, 주소 등의 자료를 포함한다.

제12절 집중적 자료처리 시스템

- (1) 조사자료는 다음의 규정에 의거하여 집중적 자료처리 시스템(central data processing system)에 의하여 관리된다.
- (2) 연방 통계청은 센서스 실시와 관련하여, 주소 및 건축물 등록부의 관리와 이와 관련된 2011년 센서스 준비법에 명시된 책무를 완수하는데 최종 책임을 진다. 제9절의 자료통합(data integration)과 주소 및 건축물 등록부에 기입된 자료는 이용가능하다.

- (3) 연방 통계청은 센서스 메타데이터시스템(census metadata system)을 개발해야 한다.
- (4) 연방 통계청은 등록사무소에 의하여 전송된 데이터(제3절 제1항), 고용 통계 등록부에서 추출된 자료(제4-5절)를 저장하고 통합하기 위한 센서스 목적의 데이터은행시스템(data bank system) 정보기술(information technology)을 개발해야 한다. 등록사무소에서 전송된 데이터(제3절 제1항)와 고용통계 등록부의 자료(제5장 제2항)를 저장하고 통합하는 업무는 주 통계청에도 동시에 적용된다. 인구통계와 고용통계 등록부 자료는 주소 및 건축물 등록부와 연계되며, 이것은 연방 통계청이 준비해야 하는 참조데이터베이스(reference database)가 된다. 참조데이터베이스는 주 통계청과의 공동 작업에서 조사항목이나 보조항목에 대한 자동매칭작업을 통하여 조사의 완전성을 점검하고 중복사례(duplicate cases)를 결정하는데 사용된다.
- (5) 참고데이터베이스(제4항 제3목) 매칭결과에 의하여 보완된다. 여기서 확정된 불일치 사례, 특히 상이한 조사부서의 자료 간에 생겨나는 불일치 사례는 통계청(연방, 주)에 의하여 해명되어야 하며, 참고데이터베이스에 포함되어야 한다.
- (6) 연방 통계청은 이 법의 업무를 완수하기 위하여 참고데이터베이스(제4항 제2목, 제5항)에 대한 접근을 주 통계청에 허용한다. 주 통계청은 건축물 및 주택 센서스의 자료와 제7절, 제8절, 제15절 4항과 제16절의 자료를 준비하기 위하여 참고데이터베이스를 준비하여야 한다.
- (7) 주 통계청은 일차통계 수집을 위한 통계조사, 6-8절 4-5항에 의한 자료의 준비하기 위하여 집중적 자료저장과 자료처리라는 의미에서 정보 기술적 업무(information technological assignment)를 분업에 의하여 책임을 진다. 노르트하임-베스트팔렌(Nordheim-Westfalen) 주 통계청은 표본추출과 특수지역 조사(7-8절)에 대하여, 작센 주 통계청은 건축물 및 주택센서스(building-housing census)에 대하여, 가구생성(household generation)과 분석업무는 바이에른 주 통계청이 책임을 진다.
- (8) 통계청(주, 연방)의 기관장은 집중적으로 입력된 자료(제1-7항)에 대한

개인정보보호 책임을 진다. 특히, 다른 통계담당자의 자료는 이 법에 의하여 센서스 목적에만 사용되어야 한다는 것을 분명히 한다. 데이터 수령자는 자동화된 자료처리과정에서 각종 요구의 허용 여부에 책임을 진다.

제13절 식별번호

- (1) 모든 주소, 건축물, 주택, 가구, 개인에 대하여, 연방 및 주 통계청은 지자체 전역 또는 건축물 전체에 걸쳐서 순번 또는 식별번호(order number)를 부여한다.
- (2) 식별번호는 자료통합(제9절)에서 이용될 수 있다.
- (3) 식별번호는 조사항목과 함께 입력된다. 이들은 센서스 자료처리의 종결 후, 늦어도 보고시점 4년 후에는 제거되어야 한다.

제4장 센서스 결과의 보호와 품질에 대한 조치

제14절 거주공간(단위거처)이 있는 건축물과 주거용 숙박시설의 주소 보완조사

- (1) 주 통계청은 2011년 센서스 준비법에 따른 주소 및 건축물 등록부에서 연방지리측량원의 자료만으로(센서스 준비법 제4절), 등록사무소의 자료만으로(센서스 준비법 제5절), 연방노동청의 자료만으로(센서스 준비법 제6절) 채택된 주소에서, 그 주소가 거주공간(단위거처)이 있는 건축물 또는 거주자가 있는 숙박시설인가를 검토한다. 주 통계청은 이렇게 확정된 주소를 2010년 7월 30일까지 주소 및 건축물 등록부에 기입하여야 한다.
- (2) 제1항의 주소를 점검하기 위하여 사용되는 모든 문서와 일반적으로 가용한 자료원은 주 통계청 내부에서만 사용되어야 한다. 데이터베이스의 점검작업(제1항)이 만족스럽지 않은 경우에는 주 통계청은 주소점검작업을 위하여, 주법에 의거하여 도시토지이용계획(urban land use planning), 등록제도(registration system), 부동산세(real estate tax), 토지대장(land register, cadaster) 관리를 담당하는 행정부서의 문서를 조사하고 이용할 수 있다. 제2항의 관련부서는 주 통계청의 요청에 따라 자료를 전송하며, 도시이용계획 자료의 경우에는 주법이 자료전송이 허용될 때만 전송이 가능하다.

제15절 중복사례 조사

- (1) 연방 통계청은 등록사무소가 전송한 데이터를 바탕으로, 개인들이 오직 하나의 거주지(sole residence) 또는 일차거주지(primary residence) (중복사례-multiple case)보다 많이 등록되어있는가 아니면 이차거주지(secondary place residence)에만 등록되어있는가를 점검한다.
- (2) 1만 명 이상 지자체의 중복사례는 연방 통계청이 기계적 보정작업 (mechanical adjustment procedure)을 실시한다. 이 경우, 가장 중요한 판단기준은 당사자의 전입일자(date of in-migration)이다. 여기서 생성된 데이터베이스는 자료통합 (제9절 제1항)과 과다집계와 과소집계 정도의 확정 (제9절 제2항)을 위한 토대자료가 된다. 등록사무소에 소급등록(backward registration)은 허용되지 않는다.
- (3) 1만 명 이하의 지자체에서 등록된 사람들 중에서, 이차거주지만 등록되어 있는 모든 사람과 오직 하나의 거주지 또는 일차거주지보다 많이 등록되어 있는 모든 사람에 대해서는 주 통계청이 보고시점까지 거주상태를 확인한다. 등록사무소에 소급등록은 허용되지 않는다.
- (4) 제3항의 거주상태를 확인하기 위해서, 주 통계청은 해당인의 다음 레코드를 조사한다.

1. 조사항목

- (a) 출생연월
- (b) 성별
- (c) 가구주와의 관계
- (d) 국적
- (e) 해당인의 모든 주소에 대한 거주상태

2. 보조자료

- (a) 성씨, 혼전 이름(여자의 경우), 이름
- (b) 출생 일자 (연도나 월이 없는 날짜)
- (c) 출생지
- (d) 해당인의 모든 일차거주지와 이차거주지의 주소

제16절 불일치 사례를 해명하기 위한 조사

주 통계청은 1만 명 이하의 지자체에서, 거주자가 있는 주택에 대해서만 주소와 관련하여 존재하는 불일치 사례(non-matching cases) 보정작업을 실시한다.. 보정작업을 위하여, 해당 주소에서 주 통계청은 그 주소에 거주하는 모든 사람들에 대하여 다음 자료를 조사한다.

1. 조사항목

- (a) 출생연월
- (b) 성별
- (c) 가구주와의 관계
- (d) 국적
- (e) 해당 주택에 거주하는 사람의 수

2. 보조자료

- (a) 성씨, 혼전 이름(여자의 경우), 이름
- (b) 출생 일자 (연도나 월이 없는 날짜)
- (c) 주소
- (d) 해당인의 모든 일차거주지와 이차거주지의 주소

제17절 센서스 결과의 품질평가

(1) 센서스 수행의 품질을 보장하기 위하여, 조사담당부서는 조사요원의 훈련과 업무분장을 문서화(documentation)하는 작업을 행한다. 작성된 문서는 주 통계청에 제출되어 검토되어야 한다. 조사담당부서가 설치되지 않은 경우에는, 주 통계청이 문서화하는 작업을 행한다.

(2) 법정인구와 관련하여 표본조사 결과의 품질을 점검하기 위하여, 제7절 제3항에 의거하여 선택된 주소에 대하여 5-10%의 추출률로 해당 주의 관련 통계기관은 대표성이 있는 반복조사(repeat survey)를 실시한다.

(3) 1만 명 미만 지자체 법정인구의 확정에 기초가 되는 조사결과의 품질을 점검하기 위하여, 주 통계청은 주민수의 0.3% 이하를 추출률로 하는 대표성이 있는 조사를 실시한다.

(4) 제2항과 제3항에 의하여 선출된 주소에 대하여, 그 주소에 거주하는 모든 사람에 대하여 다음 자료를 조사한다.

1. 조사항목

- (a) 출생연월
- (b) 성별
- (c) 거주상태
- (d) 국적
- (e) 해당 주택에 거주하는 사람의 수

2. 보조자료

- (a) 성씨, 혼전 이름(여자의 경우), 이름
- (b) 출생 일자 (연도나 월이 없는 날짜)
- (c) 주소

(5) 연방 통계청은 주 통계청은 공동으로 2015년 12월 31일까지 센서스의 수행과 결과에 대한 품질보고서를 작성한다. 해당 보고서는 제7절 제1항의 품질 가이드라인이 어떻게 충족되고 있는가를 기술한다. 특히 보고서가 포함해야 할 사항은 다음과 같다.

- 1. 표본조사방법을 개선하기 위하여 어떤 가정에서 출발해야 하며, 그것이 어느 정도 표본조사의 결과를 통하여 확인되었는가.
- 2. 어떤 과학적 기준에 의거하여 표본조사방법이 개선되어야 할 것인가.
- 3. 표본조사의 결과는 새로운 인식에 대한 품질확보의 차원에서 어느 정도 외삽 추정법(extrapolated projection)을 도입할 근거를 제시하는가.

(6) 이 품질보고서를 작성하기 위해서, 주 통계청은 각각 자체의 책임영역에 대하여 연방 통계청에 센서스 수해에 대한 품질보고서를 2015년 3월 1일까지 제출한다. 이 보고서는 조사요원의 교육과 업무분장 및 제2-3항에 의한 점검결과에 대한 보고를 포함해야 한다.

제5장 정보제공 의무와 데이터 보호

제18절 정보제공 의무와 정보제공 형태

(1) 이 법에 의한 조사를 실시하기 위하여 정보 제공을 의무화한다. 제7절 제4항에 따른 조사항목에 대한 정보제공은 자발적인 것이어야 한다.

(2) 제6절과 제14절의 제3항이 정한 조사에 대하여 건축물이나 주택의 소유자, 관리자, 별도의 감독자, 이용자는 정보제공의 책임이 있다. 세법(tax law)의 제39절 제2항에 의하여 건축물과 주택의 이익을 향수하는 사람에게도 해당한다. 정보제공 형태에는 종이 조사표 옆에 정보제공자 본인이 직접 기입하는 방식(self-administered questionnaire)으로 온라인 조사표(on-line questionnaire)가 제시되어 있다. 사업목적의 주택을 소유한 사람을 위해서, 주 통계청은 정보제공 형태에 대하여 특별협정을 체결할 수 있다. 제6절 제2항 또는 제3항에 정한 자료를 전혀 제출할 수 없는 관리자는 소유자의 이름과 주소를 제공할 의무가 있다. 2011년 센서스 준비법의 제10절 제2항에 따라 정해진 정보제공 의무자가 같은 법에 의한 조사담당부서에서 보고시점까지 아직 완결되지 않은 재산거래를 이유로 제1항과 제2항에 정한 정보제공의무자의 범주에 들지 않는다면, 그는 통계책임부서에 재산취득자의 이름과 주소를 제공할 수 있다. 정보제공 의무자가 필요한 정보를 제공할 수 없으면, 그는 제1항과 제2항에 따라, 실제로 정보를 제공할 수 있는 정보제공 의무자를 지명한다. 이들도 필요 정보를 제공하지 못하는 경우, 건축물 또는 주택의 거주자를 정보제공 의무자로 할 수 있다.

(3) 제7절에 따른 가구 표본조사와 제17절 제2-3장에 의한 표본조사에서 정보제공 의무는 추출된 주소에 거주하는 모든 성인 또는 개별가구의 주인인 미성년자, 때로는 미성년자 가구원에게 부과된다. 성인 가구원의 경우에 자신이 정보를 제공할 수 없는 경우, 정보제공 의무가 부과된 다른 가구원에게 정보제공 의무를 부과한다. 미성년자나 자신이 정보를 제공할 수 없는 사람에 대한 정보제공 의무는 정보제공 의무자가 공지되는 날짜에만 적용된다. 정보제공이 불가능한 장애자의 경우, 그가 필요한 정보의 제공이 가능한 제3자를 지명하면, 제3자가 정보를 제공하기 때문에, 자신과 가구원 중의 정보제공 의무는 소멸한다.

(3) 조사요원이 배치되면, 정보제공 의무자는 제7절 제5항 1-3번의 보조 자료와 제7절 4항 2, 4, 8절에 대한 자료와 제7절 제4항 2, 4, 8번에 따른 자료를 필요에 따라, 정보제공 의무자와 같은 주택에 거주하는 다른 사람들

위하여 구두로 조사요원과 공유할 수 있다. 추가적인 정보제공은 구두로 조사요원에게 제공되거나, 문서나 전자적으로 제공될 수 있다. 문서나 전자적으로 정보는 정해진 기한 안에 관련 수령자에게 전달되어야 한다. 전자적 정보제공의 경우, 정보제공 의무자에 대한 자료는 처리 가능한 형태로 제공되어야 한다.

(4) 제8절 제1항에 따른 조사에 대하여 정보제공 의무자는 특수지역에 거주하는 모든 사람이며, 같은 주소에 거주하는 미성년자도 여기에 해당한다. 조사요원이 배치되면, 정부제공 의무자는 제8절 제1항 제2목 1번 (a)-(b)의 자료와 2번 (a)-(b)의 자료를 필요에 따라, 정보제공자와 같은 주택에 거주하는 사람을 위하여 구두로 조사요원과 공유할 수 있다. 자신이 정보를 제공할 수 있는 성인이나 미성년자의 경우에는 시설 관리자가 정보를 제공할 책임이 있다. 민감한 특별지역(sensitive special areas)의 사람들의 경우에는 시설 관리자가 정보를 제공할 책임이 있다. 관리자의 정보제공 의무는 그들에게 공지된 날짜에만 적용된다. 시설의 관리자가 정보제공 의무가 부과되면, 정보제공의 의무가 부과된 사람으로서 책임을 다하지 않으면 안 된다.

(6) 제15절 제4항에 의한 조사에 대하여 정보제공 의무자는 관련 당사자이다.

(7) 제16절이 정한 조사에 대하여 정보제공 의무자는 모든 성인 또는 개별 가구의 주인인 미성년자이고, 따르는 미년성년자 가구원이 정보제공 의무자가 될 수 있다. 자신이 정보를 제공할 수 없는 성인 가구원의 경우에는 정보제공 의무가 있는 다른 가구원이 정보제공 의무자가 된다. 자신이 정보를 제공할 수 없는 사람이나 미성년자에 대한 정보제공 의무는 정보제공 의무자가 공지되는 날짜에만 적용된다. 미성년자나 자신이 정보를 제공할 수 없는 사람에 대한 정보제공 의무는 정보제공 의무자가 공지되는 날짜에만 적용된다. 정보제공이 불가능한 장애자의 경우, 그가 필요한 정보의 제공이 가능한 제3자를 지명하면, 제3자가 정보를 제공하기 때문에, 자신과 가구원 중의 정보제공 의무는 소멸한다. 정보제공 의무자는 제16절 2번에 대한 보조 자료에 대한 레코드와 제16절 1번 (a)-(b0-f)에 대한 레코드를 필요에 따라, 같은 주택에 거주하는 다른 사람을 위하여 조사요원과 공유할 수 있다.

제19절 삭제

(1) 보조 자료는 가능한 한 빠른 시점에 조사항목과 분리하여 별도로 보관한다. 보조 자료는 제22절 제2항과 제23절의 자료가 아닌 경우에는 통계청에서 자료매칭이나 완전성에 대한 조사항목과 보조 자료의 점검이 종료되자마자, 삭제되어야 한다. 보조 자료는 보고시점 이후 늦어도 4년이 지나서는 삭제되어야 한다.

(2) 조사 레코드는 센서스 자료처리가 종결된 후, 늦어도 보고시점 이후 4년 이내에는 파괴해야 한다.

제20절 자료전송

(1) 제3절 제1항, 제3항, 제4항에 의한 자료전송은 기존의 관행(protocol)에 따라서 실시한다. 제5절 제1항의 자료는 새로운 조사를 실시하지 않고 기존 자료를 이용하여 작성될 수 있다.

(2) 정보보호와 자료 안전성, 특히 데이터의 신뢰성, 통일성, 순수성을 확보하기 위하여, 데이터 원격전송은 현재의 테크놀로지 수준에 부합하는 조치를 취해야 한다. 일반적으로 이용 가능한 인터넷의 경우에는, 현재의 테크놀로지 수준에 부합하는 부호화(encoding, encryption)의 방법을 적용하여야 한다.

제21절 공공영역의 정보

(1) 연방 및 주 통계청은 인터넷에 www.zensus.de라는 도메인을 확보하여 국민에게 센서스에 대하여 정보를 제공한다.

(2) 연방정부는 조사결과의 특성을 연방관보나 제1항의 인터넷 사이트를 통하여 공지한다.

제22절 상급연방관청, 주 관청, 지자체(또는 지자체연합) 통계부서에 대한 제표자료와 마이크로데이터의 전송

(1) 개별사례의 규제만이 아니라, 입법기관의 업무나 계획의 목적을 위하여, 연방 및 주 통계청은 통계표가 그 셀에 하나의 사례만 있는 경우에도, 상급 연방 관청이나 주 관청에 해당 통계표를 전달하여야 한다.

(2) 지자체의 순수한 목적을 위하여 연방 및 주 통계청은 지자체나 지자체 연합의 통계업무를 수행하는 부서를 위하여 조사항목에 대한 조사항목과 도로, 주택번호 또는 블록에 따라 통합된 마이크로데이터를 전송한다. 자료전송은 통계비밀이 법률적 조치에 의하여 보장되고, 특히 공간적, 조직적, 개인적 정보를 통계담당부서와 일반 행정부서 공유하지 않는 경우에 가능하다. 여기에도, 보조 자료는 여기서도 자료전송 이후 2년 이내에 삭제되어야 한다.

제6장 결론

제23절 건축물, 주택, 인구의 표본조사 근거제시

연방 및 주의 통계가 작성되는 건축물, 주택, 및 인구 표본조사를 위하여, 연방 및 주 통계청은 주택 및 인구의 수, 특별지역의 종류, 건축물 또는 숙박시설의 주소는 수리적 확률법칙에 따라 이 법률의 범위 안에서 조사구의 선정방식을 연구하는데 사용한다. 표본조사의 추출률은 20%를 상한선으로 한다. 이들 조사구의 특성은 목적을 달성한 후에는 즉시, 늦어도 장래의 센서스를 이용하여 상응하는 표본선택기준이 마련되는 시점에서 파괴해야 한다. 불필요한 조사구 80%는 위에서 언급한 조사구의 확정 이후 즉시, 늦어도 보고시점 4년 이후에는 삭제되어야 한다.

제24절 연방 통계청으로의 자료전송 비용

연방 통계청으로의 자료전송 비용은 별도로 보상하지 아니한다.

제25절 재정분배

연방정부는 주 정부에 2011년 등록센서스를 준비하고 수행하는데 2.5억 유로 수준의 재정을 분배한다. 재정분배는 각주의 개별비용에 따라 이루어지며, 이것은 각주 간 행정협정(administrative agreement)의 형태로 늦어도 2010년 3월 31일까지 확정되어야 한다.

참고문헌

1. 국어 문헌

- 강영옥 · 윤은주 · 원종석 · 정재희(2007), 『GIS기반 소지역 통계집계
공표구역의 획정 및 관리방안 연구』, 건설교통부
건축법(2009), “건축법 시행령, 건축물대장의 기재 및 관리 등에 관한
규칙”, 법제처 홈페이지 (www.law.go.kr)
- 국가예산정책처(2006), 『인구주택총조사사업평가』
- 권태환 · 김두섭 (2002), 『인구의 이해』, 서울대학교 출판부
- 김근영 · 김순관 · 이신해(2003), "Comparing Census Data with
Resistration Data with Respect to Population and Household in the
Seoul Metropolitan Region", 『국토계획』 38(1): 237-249
- 김두섭 · 박상태 · 은기수(편) (2002), 『한국의 인구』, 통계청
- 김민경, 2000, 『인구센서스의 이해』, 도서출판 글로벌
- _____, 2002, “인구센서스의 발전과 특징”, 김두섭 · 박상태 · 은기수(편),
『한국의 인구』, 통계청
- 김태현 외 (2004) “인구주택총조사 통계품질관리지표 개발”, 통계청 보고서
- 김형석(1999), “인구주택총조사의 문제점과 개선방안 : 행정 등록통계와의
접목 방안에 대하여”, 미발간 통계청 자료
- _____. 양경진 · 강영민(2003), “인구센서스와 행정자료의 활용.” 미발간
통계청 자료
- _____. 이태직 (2008), “주민등록자료의 평가와 보정”, 『통계연구』
13(2): 1-40, 통계청
- 법제처 홈페이지(2009a), 주민등록법, 주민등록법시행령.법제처
법령종합정보센터(www.law.go.kr)
- _____(2009b), 건축법, 건축법 시행령, 건축물대장의 기재 및
관리 등에 관한 규칙. 법제처 종합법령정보센터(www.law.go.kr)
- 변필성 · 김광익 · 안재학(2006), “소지역통계 공표구역 사례:
잉글랜드·웨일즈의 Output Area”, 『국토』 10월호. 국토연구원.
- 성남시(2008), 주민등록 업무지침, 2008
- 양경진(2003), “행정자료 활용 주택DB 구축 방안.” 「2005년
인구주택총조사 준비 제2차 워크숍」

- ____·김황대(2008), "제2장 인구주택총조사 표본조사방법 개선을 위한 사례연구 -미국, 프랑스, 네덜란드 사례를 중심으로-", 『인구센서스 방법론 연구』, 통계개발원
- 윤혁경(편) (2003), 『건축법 조례 해설』, 기문당
- 이건(2007). "읍면동 수준에서 인구주택총조사와 주민등록의 집계인구 비교", 『통계』 33(1): 37-54
- ____· 변미리 · 이명진 · 서우석(2004), 『주거부문 행정자료의 인구주택총조사 현장적용방안』, 통계청
- ____· 이명진 · 서우석 · 변미리(2006), "2005년 인구주택총조사의 조사구 설정 방법", 『조사연구』 7(1):109-129.
- ____ 등(2007), 『2010년 인구주택총조사 방법론 연구』, 통계청 연구결과보고서. 한국조사연구학회
- 이내성(2006), "주민등록 행정자료를 활용한 인구통계 조사방법 연구", 미발간 통계연구회 발표자료
- ____ (2008a), "총조사에 주민등록 행정자료의 활용을 위한 자료매칭 연구", 『조사연구』 9(2): 119-149
- ____ (2008b), "행정자료 기반의 등록센서스 작성을 위한 항목선정", 통계개발원, 『인구센서스 방법론 연구』
- 이명진 · 서우석(2008), "주민등록세대수와 인구주택총조사의 가구수 비교연구", 『통계연구』 13(1): 1-24. 통계청
- ____· 서우석· 이건(2006), "주거부문 행정자료의 인구주택총조사 활용가능성에 대한 경험적 검토", 『통계연구』 11(1): 57-87
- 이삼식 · 박종서 · 하미영(2002), 『환경 변화에 따른 인구주택총조사 개선방안』, 통계청·한국보건사회연구원
- 이지연(2007a), "2010년 인구주택총조사 방법론 연구 기초자료." 미발간 통계청 자료
- ____(2007b), "2005년 인구주택총조사의 범위오차평가와 지역별 차이 연구." 한국인구학회 춘계학술대회 발표문
- 임명선(2003), "2005년 총조사 조사구 설정방법 개선방안." 「2005년 인구주택총조사 준비 제2차 워크숍」
- 전광희(2008a), "미국 센서스의 변화와 향후 전망- 2000년의 경험과 2010년의 계획을 중심으로", 『한국인구학』 31(2): 101-132. 한국인구학회
- ____(2008b), "영국의 One Number Census 프로젝트: 2001년 방법론의

- 핵심과 2011년 계획을 중심으로“, 『인구와 사회』 4(2): 91-123.
 한양대학교 인구 및 고령사회연구소
- 통계청(2004a), 『2005 인구주택총조사 기본계획안(요약)』, 통계청.
 _____(2004b), 『지역통계 생산을 위한 도시화지역 설정 연구』, 통계청
 _____(2005), 『2005 인구주택총조사 조사지침서』, 통계청
 _____(2006a), 『2005 인구주택총조사 종합평가보고서』, 통계청
 _____(2006b), 『도시화지역 획정을 위한 시스템 개발』, 통계청
 _____(2006c), 『통계지리정보시스템 구축을 위한 ISP/BPR 사업』, 통계청
 _____(2007a), 『2010 인구주택총조사 추진기본계획 (잠정안)』, 통계청
 _____(2007b), 『건축물대장 활용방법 연구』, 통계청
 _____(2007c), 『2010년 인구주택총조사방법론 연구』, 통계청
 _____(2007d), 『2010 인구주택총조사 제1차 시험조사 행정자료활용 결과
 보고』, 통계청
 _____(2007e), 『2005년 인구주택총조사보고서(전수조사결과)』, 통계청
 _____(2007f), 『2005년 인구주택총조사 표본조사결과』, 간행물 CD
 통계청
 _____(2008a), 『등록센서스실시 기본계획』, 통계청
 _____(2008b). 『통계행정편람』, 통계청
- 통계개발원(2008), 『한국의 인구·주택-인구주택총조사 종합보고서』,
 통계청
- 통계교육원 (2008), 『인구통계 기초와 응용』, 통계청
- 통계청 인구조사과(2003), 『2005년 인구주택총조사를 위한 자료이용기관
 의견수집조사결과』, 내부자료
- 통계개발원 사회통계실(편) (2008a), 『행정자료의 통계적 활용』, 통계청
 통계개발원
 _____ (2008b), 『센서스 방법론 연구』, 통계청
 통계개발원
- 한국인구학회(1998), 『행정수요 기준설정을 위한 새로운 인구개념의 정립
 연구』
 _____ (2006), 『인구대사전』. 통계청
- 한국전산원(1996), 『건축물대장의 전산화를 위한 서식표준화 (안)
 개발보고서』
- 행정안전부(2004), 『읍면동 기능전환 관련 시군구 이관사무 처리 편람』
 _____ (2007), “2007년 주민등록인구에 관한 보도자료 (2007년 11월

19일자).”

_____ (2008), “주민등록법 일부개정법률(안) 입법예고.” 행정안전부
공고 제2008-51호(2008년 5월 28일)

홍두승 · 이진 · 이명진 · 서우석 · 장원호(2002), 『인구주택총조사 개선
방안에 관한 연구-행정자료의 인구주택총조사 활용방안』. 통계청

2. 일어문헌

日本人口學會(編)(2002), 『人口大事典』 (Encyclopedia of Population),
東京: 培風館

日本 總務省 統計局(2009) 公的統計の整備に関する基本的な計画”
(공적통계의 정비에 관한 기본계획)

(<http://www.stat.go.jp/index/index.htm>)

西村善博 (2007a), "フランス新人口センサスにおける 推計 の現段階"
(프랑스의 새로운 인구센서스 추계의 현 단계)

(http://www.hosei.ac.jp/toukei/shuppan/g_shoho36_nishimura.pdf)

西村善博 (大分大学・経済学部)

_____ (2007b), "フランス新人口センサスの基本設計の展開" (프랑스의
새로운 인구센서스의 기본설계의 전개)

(<http://ir.lib.oita-u.ac.jp/dspace/bitstream/123456789/12474/1/59%E5%B7%BB2%E5%8F%B7%EF%BC%88p99-131%EF%BC%89.pdf>)

3. 중국어문헌

中華民國 行政院 (2009a), 中華民國九十九年人口及住宅普查方案

_____ (2009b), 九十九年人口及住宅普查-公務登記資料連結備對
作業

_____ (2009c), 九十九年人口及住宅普查抽樣設計

4. 독일어문헌

Statistisches Bundesamt (1987), Gesetz über die Statistik für
Bundeszwecke Bundesstatistikgesetz - BStatG) Vom 22. Januar 1987
(BGBI. I S. 462, 565) (연방목적의 통계에 관한 법률)

_____ (2009a), Gesetz zur Vorbereitung eines
registriergestützten Zensus einschließlich einer Gebäude- und
Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011 - -
ZensVorbG 2011) Vom 8. Dezember 2007 (2011년 센서스 준비법, 곧
2011년 건축물 및 주택 센서스를 포함하여 등록부를 토대로 하는
센서스를 준비하기 위한 법률)

_____ (2009b) Gesetz über den registriergestützten Zensus im
Jahre 2011 Zensusgesetz 2011 - - ZensG 2011) Vom 8. Juli 2009
(2011년 등록기반 센서스에 관한 법률)

5. 영어문헌

- Al, Pieter G. and Jan Willem Altena (2000) "Data security, privacy and the SSB", *Netherlands Official Statistics* Vol. 15 (Summer).
- Al, Pieter G. and Bart Bakke (2000), "Re-Engineering Social Statistics by Micro-Integration of Different Sources." *Netherlands Official Statistics* Vol. 15 (Summer).
- Alexander, Charles. H (2000), "The American Community Survey and the 2010 U.S. Census." INSEE-Eurostat Seminar on Censuses after 2001.
- _____ (2002), "Still Rolling: Leslie Kish's 'Rolling Samples' and the American Community Survey." *Survey Methodology* 28: 35-41.
- Anders and Britt Wallgren. (2007). *Register-based Statistics*. New York: John Wiley & Sons.
- Arts, Arts, Bart F. M. Bakker, and Erik van Lith (2000), "Matching Administrative Registers and Household Surveys", *Netherlands Official Statistics* 15: 16-22.
- Australian Bureau of Statistics (2007), "The 2006 Australian Standard Geographical Classification and Census Geographic Areas - Diagram of Structures." (<http://www.abs.gov.au/AUSSTATS/abs@.nsf/>)

[Latestproducts/2914.0Main%20Features225002006?opendocument&tabname=Summary&prodno=2914.0&issue=2006&num=&view](http://www.iaos.org/Products/2914.0Main%20Features225002006?opendocument&tabname=Summary&prodno=2914.0&issue=2006&num=&view)).

- Bankier, Mike (2007), "Broad Issues of Editing And Imputation For Population Censuses", Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.
- Bell, William R. (2007), "A Review of Some Recent Work at the U. S. Census Bureau on Dual System Estimation of Census Coverage", Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.
- Bianchi G., Manzari A, A Pezone, A. Reale, G. Saporito. 2005, "New procedures for Editing and Imputation of Demographic Variables", Proceedings of the UN/ECE Work Session on Statistical Data Editing, Ottawa, Canada.
- Bierau, Dieter (2000), "Testing a Register-Based Census." Insee-Eurostat Seminar on the Censuses after 2001. Paris.
- Blum, Olivia (2005), "Wholistic Approach to Multifaceted Integrated Census ", Central Bureau of Statistics, Jerusalem, Israel.
(<http://www.fcs.gov/03papers/Blum.pdf>)
- Borchsenius, Lars (1996), "From a Conventional to a Register-Based Census of Population." Working Paper No.9. SCECE Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- _____ (2000), "From a Conventional to a Register-based Census of Population." INSEE-Eurostat Seminar on the Censuses after 2001. Paris.
- _____ (2006). "New Development in the Danish System for Access to Microdata." United Nations Statistical Commission and Economic Commission for Europe Conference of European Statisticians. Geneva.
- Bruhn, Åke (2001), "The 2005 Population and Housing Census in Sweden

will be Totally Register-Based." Symposium on Global Review of 2000 Round of Population and Housing Censuses. Statistics Division, Department of Economic and Social Affairs, UN Secretariat, New York, 7-10 August 2001.

Bruni R., A. Reale A., R. Torelli, 2001, "Optimization Techniques for Edit Validation and Data Imputation", Paper Presented at the Statistics Canada Symposium 2001 "Achieving Data Quality in a Statistical Agency: a Methodological Perspective", XVIIIth International Symposium on Methodological Issues.

Canberra Group, The (2001), Expert Group on Household Income Statistics: Final Report and Recommendation. Ottawa.
(<http://www.lisproject.org/links/canberra/finalreport.pdf>)

Cohen, Noam, Dan Ben-Hur and Luisa Burck (2007), "A Recursive Derivation of the Sampling Error in Multi-phase Calibration with Additive g-factors", Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.

Childs, Jennifer Hunter, and Elizabeth Nichols (2007), "A New Approach to Measuring Residence Status: A Summary of Results", Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.

Choe, Minja Kim. 2000. " Electronic Microdata of the Censuses of the Republic of Korea at the East-West Center, University of Hawaii." Pp. 123-140 in Handbook of International Historical Microdata for Population Research (Ed.) Patricia Kelly Hall, Robert McCaa, and Gunnar Thorvaldsen. Minnesota, MN: Minnesota Population Center.

Dasappa Princelle and Tjaart van der Walt (2007), "Research to Inform Census Methodologies-The South African Experience." Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and

SPSS.

- Demeny, Paul G. and Geoffrey McNicoll (2003), *Encyclopedia of Population*, New York: Thomson Gale
- Durr, Jean-Michel and J. Dumais (2002), "Redesign of the French census of Population." *Survey Methodology* 28: 43-49.
- Edmonston, Barry and Charles Schultzd (ed.) (1995), *Modernizing the U.S. Census*. Washington D.C.: National Academy Press.
- Eppmann, Helmut, Sonja Krügener, and Josef Schäfer (2006), "First German Register Based Census in 2011" *Allgemeines Statistisches Archiv* 90(3): 465-482.
- European Commission. (2002), "Commission Regulation (EC) No 831/2002 of 17 May 2002: Implementing Council Regulation (EC) No 322/97 on Community Statistics, concerning Access to Confidential Data for Scientific Purposes." *Official Journal of the European Communities* 18.5.2002: 7-9.
- Eurostat, 2007, "Towards the Adoption of EU Legislation for Population and Housing Censuses: Summarising Progress and Highlighting Issues Relevant to Register-Based Censuses", Paper Presented at the Joint UNECE-Eurostat Expert Group Meeting on the Implementation of the CES Census Recommendations for Register-Based Censuses Astana (Kazakhstan), 7-8 June 2007.
- Federal Statistical Office of Germany (1992). "Considerations of Alternatives to Censuses and Census-type Statistics: the Case of Germany." Working Paper No. 13. SCECE Work Session on Population and Housing Censuses. Geneva.
-
- _____ (2002), "Der Test eines Registergestützten Zensus." (<http://www.destatis.de>)
-
- _____ (2007a), Microdata. (<http://www.destatis.de>)
-
- _____ (2007b), "The Census Process 2011", (http://www.statistik-portal.de/Statistik-Portal/en/Zensus/en_zeitAblauf.asp).
- Fellegi, I. P. and A. B. Sunter, 1969, "A theory of Record Linkage", *Journal of American Statistical Association* 64: 183-1210.

- Gill L E. (2001), "Methods for Automatic Record Matching and Linking and Their Use in National Statistics" National Statistics Methodological Series, Great Britain, Oxford University, Number 25.
- Glickman Hagit , Ronit Nirel, Dan Ben-Hur (2005), "The Integrated Census in Israel Using Sample Surveys to Estimate Coverage Errors in Administrative Data." Central Bureau of Statistics, Israel (http://www.cbs.gov.il/mifkad/integ_census.pdf).
- Goossens, H. (2006), Metadata As a Crucial Starting Link in New Statistical Cycles, UNECE-Eurostat work session on statistical metadata Geneva, Swiss, 3-5 may, 2006.
- Hall, Kelly, Robert McCaa, and Gunnar Thorvaldsen (2000), *Handbook of International Historical Microdata for Population Research* (Ed.) Patricia Minnesota, MN: Minnesota Population Center.
- Harala, Riitta and Jari Nieminen (1998). "Use of the Building and Dwelling Register in the Production of Statistics." Statistics Finland.
- Harala, Riitta and Reinikainen Anna-Leena (1996), "Confidentiality in the Use of Administrative Data Sources." *Statistical Journal of the United Nations ECE* (1996): 361-368.
- Hawkers, William J. (1986). "Census Data Quality - A User's View." *Journal of Official Statistics* 2(4): 531-544.
- Hendricks, Coen, Paul Inge, and Harald Utne (2000). "Establishing a Dwelling Register in Norway: Completing the Register System." Insee-Eurostat Seminar on Census after 2001. Paris.
- Kawasaki, Shigeru (2006), "Use of Administrative Data for Official Statistics in Japan." (http://www.unescap.org/Stat/apex/2/APEX2_S.6_Use%20of%20Adm%20Data%20in%20Japan.pdf)
- INSEE (2004). "The New French Rolling Census." UNECE Seminar on New Methods for Population Censuses.
- _____(2006). "The Population Censuses in France : from General Censuses to 'Rolling Census'." UN Economic and Social Council. INSEE
- Israel Central Bureau of Statistics (2005), "The 2008 Israel Integrated Census of Population and Housing", Submission to the United

Nations Statistics Division Website on 2010 World Population and Housing Censuses.

(<http://unstats.un.org/unsd/Demographic/sources/census/israelpdf.pdf>)

_____ (2007), "Supporting Information Sources for Register-Based Census", Paper Prepared at the Joint UNECE-Eurostat Expert Group Meeting on the Implementation of the CES Census Recommendations for Register-Based Censuses Astana (Kazakhstan), 7-8 June 2007.

_____ (2008), "Evaluating the Integrated Census in Israel", Paper Presented at the Joint UNECE/Eurostat Meeting on Population and Housing Censuses Eleventh Meeting Geneva, 13-15 May 2008.

Jensen, Lothar (2000). "Opening Speech." Insee-Eurostat Seminar on the Censuses after 2001. Paris.

Juristisches Internetprojekt Saarbrücken (2001). Gesetz zur Vorbereitung eines Registergestützten Zensus.

Kamen, Charles S. (2005) "The 2008 Israel Integrated Census of Population and Housing: Basic conception and procedure". Central Bureau of Statistics, Israel.

Kish, Leslie (1990), "Rolling Samples and Censuses (with discussion)." *Survey Methodology* 16: 63-79.

_____ (1998), "Space/Time Variations and Rolling Samples." *Journal of Official Statistics* 14: 31-46.

_____ (1999), "Cumulating/Combining Population Surveys." *Survey Methodology* 25: 129-138.

Laihonen, Aarno (1999). "Development of the Use of Administrative Data in Population and Housing Censuses in Europe." Working Paper No. 6. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.

_____ (2000), "2001 Round Population Censuses in Europe." Insee-Eurostat Seminar on Census after 2001. Paris.

Lenk, Manuela (2008), "Methods of Register-based Census in Austria", Statistics Austria, Vienna, Austria

Leow, Bee-Geok and Eng-Chuan Koh (2001), "Combining Survey and

- Administrative Data for Singapore's Census of Population 2000."
 (<http://www.ancsdaap.org/cencon2003/Papers/Singapore/Singapore.pdf>)
- Linder, Frank (2003), "The Dutch Virtual Census 2001: A New Approach by Combining Administrative Registers and Household Sample Surveys", DIECOFIS Workshop Vienna, 13-14 November 2003
- Martin, David (1998b). "Optimizing Census Geography: the Separation of Collection and Output Geographies." *Geographical Information Science* 12(7): 673-685.
- Martin, David (2002), "Geography for the 2001 Census in England and Wales." *Population Trends* 108: 7-15.
- Martin, David (2004), "Neighborhoods and Area Statistics in the Post 2001 Census Era." *Area* 36.2: 136-145.
- Mueunnich Ralf, Siegfried Gabler, and Mathias Ganninger (2007), "Some Remarks on the Register-Based Census 2010/2011 in Germany", Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.
- Myrskylä, Pekka (1999), "New Statistics Made Possible by the Use of Registers." Working Paper No. 8. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- Myrskylä, Pekka (2000), "Effects of Changes to Registers: Experiences from Finnish Register-based Censuses." Insee-Eurostat Seminar on Census after 2001, Paris.
- National Institute of Statistics of Italy (2008) "An Overview of Editing and Imputation Methods for the Next Italian Censuses", Joint UNECE/Eurostat Meeting on Population and Housing Censuses Eleventh Meeting Geneva, 13-15 May 2008.
- National Research Council, (2004a), *Reengineering the 2010 Census: Risks and Challenges. Panel on Research on Future Census Methods*, Daniel L. Cork, Michael L. Cohen, and Benjamin F. King (eds.), Committee on National Statistics. Washington, DC: The National Academies Press.

- National Research Council (2004b), *The 2000 Census: Counting Under Adversity. Panel to Review the 2000 Census*, Constance F. Citro, Daniel L. Cork, and Janet L. Norwood (eds.), Committee on National Statistics. Washington, DC: The National Academies Press.
- National Research Council (2005), *Principles and Practices for a Federal Statistical Agency*. Third edition. Margaret E. Martin, Miron L. Straf, and Constance F. Citro (eds.), Committee on National Statistics. Washington, DC: The National Academies Press.
- National Research Council (2006). *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census*. Panel on Residence Rules in the Decennial Census, Daniel L. Cork and Paul R. Voss (Eds.), Committee on National Statistics. Washington, DC: The National Academies Press.
- National Research Council (2007), *Using the American Community Survey: Benefits and Challenges*. Panel on the Functionality and Usability of Data from the American Community.
- Nirel, Nirel and Hagit Glickman (2007), "An Integrated Census: The Statistical Perspective", Innovative Methodologies for Censuses in the New Millennium: a Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.
- Nordberg, Lief, Irmeli Penttila, and Susanna Sandstrom (2001). "A Study on the Effects of Using Interview versus Register Data in Income Distribution Analysis with an Application to the Finnish ECHP-survey in 1996." CHINTEX Working Paper No 1.
- Nordholt, E, Hartgers, M and Gircour R (2004), *The Dutch Virtual Census of 2001: Analysis and Methodology*, Netherlands Central Bureau of Statistics Monograph. Statistics Netherlands: Voorburg/Heerlen.
- Puderer, Henry (2001) "Introducing the Dissemination Area for the 2001 Census: an Update." Geography Working Paper Series No. 2000-4. Statistics Canada.
- Rahman, Nargis, and Shayla Goldring (2007), "Modelling Census Household Non-response", Innovative Methodologies for Censuses in

the New Millennium: A Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.

Ronit Nirel, Hagit Glickman, and Dan Ben-Hur (2003). "A Strategy for a System of Coverage Samples for an Integrated Census." Proceedings of Statistics Canada Symposium 2003.

Ruotsalainen, K., (2004). Use of Administrative Data in Population and Housing Census in Finland, Presentation of Seminar on the Use of Administrative Sources Paris.

Siegel, J. and David A. Swanson (eds.) (2004), *The Methods and Materials of Demography*, New York: Academic Press (Elsevier)

Singapore Department of Statistics (2002), Census of Population, 2000 Administrative Report. Department of Statistics, Ministry of Trade and Industry, Singapore.

Slagter, Herman (Statistics Netherlands) (1999), "Compiling Structure of Earning Statistics using Existing Survey Data and Register Data.". Working paper No. 7. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.

Smith, Paul (2007), "Targeting resources to achieve robust coverage adjustments to the population census." Powerpoint Presentation", Innovative Methodologies for Censuses in the New Millennium: A Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.

Spain National Office of Statistics (1999a). "Utilization of Administrative Registers in the 2001 Spanish Demographic Censuses." Working paper No. 27. Joint ECE/Eurostat Work Session on registers and Administrative Records in Social and Demographic Statistics. Geneva.

(1999b). "The 'Padron Continuo'." Working paper No. 28. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.

- Spieker, Finn (1999), "Formation of Central Variables in a Decentralised Statistical System". Working Paper No. 25. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- Statistics Canada (2004), *2001 Census Dictionary*. Ottawa: Ministry of Industry.
- Statistics Finland (2004), Use of Registers and Administrative Data Sources for Statistical Purpose.
- _____ (2006), "Access to Enterprise Data at Statistics Finland: Introducing the Practices of the Research Laboratory." OECD Conference: Assessing the Feasibility of Microdata Access, Luxembourg 26-27 October 2006.
- Statistics Netherlands (1999). "Contents and Use of Register-based Job Files." Working paper No. 32. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- _____ (2006), "The Dutch Virtual Census of 2001: A Register-Based Approach Combined with Survey Information." Paper Presented at the Conference of European Statisticians for UNESCO. Paris.
- Statistics Sweden (1999), "Registers in Official Statistics: A Swedish Perspective." Working Paper No. 14. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- Tromans Nicola and James Brown (2007), "The application of Dual System Estimation in the 2011 UK Census", Innovative Methodologies for Censuses in the New Millennium: A Satellite Meeting of the 56th Session of the ISI Conference, Lisboa, 2007 Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.
- United Kingdom Office for National Statistics (ONS) (1998), "2001 One Number Census." London: Office for National Statistics.
- _____ (2001), "Census 2001: A Guide to the One Number Census", London: Office for National Statistics

- (<http://www.statistics.gov.uk/nsbase/census2001/pdfs/oncguide.pdf>).
-
- _____ (2003a), "Alternative to a Census: Review of Interantional Approaches." Census Strategic Development Review. (<http://www.statistics.gov.uk>)
-
- _____ (2003b), "National Statistics Code of Practice Protocol of Matching", (<http://www.statistics.gov.uk>)
- United Nations (1969), *Methodology and Evaluation of Population Registers and Similar Systems*. New York: United Nations Department of Economic and Social Affairs
-
- _____ (1998), *Recommendation on Statistics of International Migration Revision 1*, New York: United Nations.
-
- _____ (2001), *Principles and Recommendations for a Vital Statistics System. Revision 2*, New York, United Nations.
-
- _____ (2008), *Principles and Recommendations for Population and Housing Censuses, Revision 2*, New York: United Nations.
-
- _____ (2009), *Fundamental Principles of Official Statistics. United Nations*. (unstats.un.org/unsd/dnss/gp/fundprinciples.aspx)
- United Nations Economic Commission for Europe (UNECE) (2006), *Conference of European Statisticians Recommendations for 2010 Censuses of Population and Housing*. New York: United Nations.
-
- _____ (UNECE) (2007), *Register-based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*. Prepared with the Statistical Office of the European Community (Eurostat), New York: United Nations.
- U.S. Census Bureau (2004), "Meeting 21st Century Demographic Data Needs - Implementing the American Community Survey Report 8: Comparison of the American Community Survey Three-Year Averages and the Census Sample for a Sample of Counties and Tracts.
- Utne, Harald (2005), *The Population and Housing Census Handbook 2001*, Statistics Norway. <http://www.ssd.no>.
- Van der Laan, Paul (2000), "The 2001 Census in the Netherlands Integration of Registers and Surveys." Insee-Eurostat Seminar on

the Censuses after 2001. Paris.

Vliegen, Mathie and Huib Van de Stadt (Statistics Netherlands) (1989),

“Is a Census Still Necessary? Experiences in the Netherlands.”

Working Paper No. 5. SCECE Seminar on the Relevance and Importance of Population and Housing Census Data. Wiesbaden, Germany.

Wagstaff, Heather and Steven Rogers (2007), “Optimizing the 2011 UK

Census Editing Strategy not reinventing the wheel.” Innovative

Methodologies for Censuses in the New Millennium: a Satellite

Meeting of the 56th Session of the ISI Conference, Lisboa, 2007

Sponsored by the ISI, IASS, IAOS, ASA, Lockheed Martin and SPSS.

Wallgren, A. and Wallgren, B. (2007), Register-based Statistics:

Administrative Data for Statistics Purpose, New York: Willey.

Wolter, Kirk M. (1986), “Some Coverage Error Models for Census Data”,

Journal of American Statistical Association 81-338-346.

