

2010년 인구주택총조사 자료와 행정자료의 비교분석을 통한 방법론 개발 연구 연구용역 최종보고서

2012년 11월 30일

연구책임자: 전 광 희 (충남대 사회학과)

연구원: 이 석 훈 (충남대 통계학과)

연구원: 박 진 우 (수원대 통계학과)

연구원: 박 민 규 (고려대 통계학과)

연구원: 김 달 호 (경북대 통계학과)

연구원: 은 기 수 (서울대 국제대학원)

연구원: 김 정 석 (동국대 사회학과)

통계청
한국인구학회



2010년 인구주택총조사 자료와 행정자료의 비교분석을 통한 방법론 개발 연구

연구용역 최종보고서

2012년 11월 30일

연구책임자: 전 광 희 (충남대 사회학과)
연구원: 이 석 훈 (충남대 통계학과)
연구원: 박 진 우 (수원대 통계학과)
연구원: 박 민 규 (고려대 통계학과)
연구원: 김 달 호 (경북대 통계학과)
연구원: 은 기 수 (서울대 국제대학원)
연구원: 김 정 석 (동국대 사회학과)

통계청
한국인구학회

제 출 문

통계청장 귀하

본 보고서를 “2010년 인구주택총조사 자료와 행정자료의 비교분석을 통한 방법론 개발 연구” 학술연구용역 과제의 최종 연구결과물로 제출합니다.

2012년 11월 30일

계약기관 한국인구학회 이승욱 □

연구진

책임연구자	전 광 희 (충남대학교 교수)
연구자	박 진 우 (수원대학교 교수) 이 석 훈 (충남대학교 교수) 박 민 규 (고려대학교 교수) 김 달 호 (경북대학교 교수) 은 기 수 (서울대학교 교수) 김 정 석 (동국대학교 교수)
연구보조원	신 지 은 (충남대학교 통계학 박사) 송 종 호 (고려대학교 박사학위 과정) 김 미 선 (서울대학교 박사학위 과정) 임 상 호 (충남대학교 대학원 석사과정) 이 태 호 (충남대학교 대학원 석사과정)

연구과제명	2010년 인구주택총조사 자료와 행정자료의 비교분석을 통한 방법론 개발 연구
주제어	등록센서스, 이상 시군구, 이상 읍면동, 표본설계, 리지회귀추정량, 소집단추정, 품질진단, 센서스 추정방법론, 직접추정량, 이원시스템추정, 비추정량, 포괄범위
연구기간	2011년 5월 15일-11월 30일
연구기관	한국인구학회
연구진구성	전광희, 이석훈, 박진우, 박민규, 김달호, 은기수, 김정석
2012년 등록센서스 학술연구 최종보고서는 (1) 등록자료와 조사자료의 차이에 대한 자료를 이용한 이상(異常) 시군구 탐색연구 (2) 등록센서스 표본설계 (3) 인구, 가구, 주택 부문 등록센서스 통계 작성방안 (4) 소지역 추정방법론 (5) 등록센서스 통계작성 방식의 품질진단 (6) 2010년과 2020년 라운드 등록센서스의 최근동향 (7) 센서스 추정방법론과 우리나라 2015년 등록센서스의 당면과제 등 모두 7개 개별과제로 구성되어 있다. 제1장에서는 시군구의 인구수, 가구수, 주택수를 모두 고려하여 전체 및 공표대상 범주별 조사자료와 행정자료 간 상대적 차이와 절대적 차이가 종합적으로 큰 시군구를 탐색하고 이들 시군구를 판별하는데 유용한 설명변수를 가용한 행정자료 내지는 참조자료 내에서 선택, 제안한다. 구체적으로, 우선 2011년에 제안한 등록자료와 조사자료의 차이를 나타내는 척도 RoD(상대차이)와 이 척도의 한계를 극복하기 위하여 시군구 규모 및 성별, 연령별 차이를 함께 고려하여 새롭게 제안한 SWROAD라는 척도를 이용하여 이상 시군구를 탐색하고, 그 차이의 정도(대, 중, 소)에 따라 12개 시군구를 선정하여 그 리스트를 제공한다. 다음으로 가용자료로부터 이상 시군구를 판별하는 방안을 연구하기 위하여 가용자료와 연계시킨 시군구 자료를 탐색한다. 이어서 로지스틱 회귀모형을 적용하여 인구, 가구, 주택 부문의 이상 시군구를 탐색하는데, 기존 척도인 RoD와 신규 척도인 SWROAD를 이용한다. 마지막으로 각 척도별 범주별(인구, 가구, 주택) 이상 시군구를 정리하고, 2011년 연구의 이상 읍면동과 연계시켜 연계정도를 파악한다. 제1장에서는 2011년도에 제안한 척도와 신규로 개발한 척도와의 관련성이나 지역별 구체적 특성에 대한 검토를 시도하였으나 상세한 분석이 이루어지지 않아서, 개발된 척도들 간의 관련성이나 특성 등을 추가적으로 연구할 필요가 있다. 제1장에서는 2009년 사업체조사 자료를 사용하였다. 그러나 향후 최신 자료나 더 많은 정보를 포함하고 있는 경제총조사 자료를 이용하여 이상지역을 탐색하는 연구가 필요하다. 제2장은 제1장의 연구결과와 등록센서스 담당 주부무서가 선정한 12개 대상지역의 조사구별 2010년 행정자료와 2010년 인구주택총조사 자료를 이용하여 하여 표본설계를 실시하였다. 조사구 특성을 미리 검토하였고, 기숙시설, 특수사회시설, 특수 조사구를	

전수조사로 전환시켰다. 또 이상치 조사구 중 인구가 많은 조사구만 전수조사 대상으로 하고 인구가 적은 이상치는 제외시켜 추출률을 낮추고 효율성은 키우도록 했다.

제2장은 2010년 10% 표본조사와 이전의 연구에서와 같이 시군구 단위로 목표오차를 지정하여 표본크기를 결정하는 방식을 사용하였다. 이 경우, 시군구 단위의 추정에는 효과적이었으나 읍면동 단위의 추정은 정확성의 편차가 커서 부적절하였다. 같은 시군구 안에서도 읍면동별 특성이 많이 다르므로 시군구 표본크기를 정한 후 동일 지역에서 읍면동별 추출률을 같게 하는 방안보다는 처음부터 읍면동 수준에서 표본크기를 결정하는 방식이 필요하다는 것을 알 수 있었다.

표본추출을 위한 조사구 정렬변수로 ① 인구수를 사용하는 방안, ② 인구수와 주택수를 동시에 고려하는 방안, ③ 인구수, 가구수, 주택수 모두를 고려하는 방안을 각각 적용하여 비교하였으나, 기본적으로 세 방안의 효율이 크게 다르지 않기 때문에 인구수만 고려하는 방안을 적용하였다.

표본설계의 적절성과 상관없이 행정자료와 조사자료의 근원적인 차이로 인해 생기는 추정 단계에서의 문제를 해결하기 위해 직접추정량(direct estimator)이 아닌 추정방법에 대한 별도의 연구가 필요하다.

제3장의 목적은 주어진 행정자료와 조사자료의 각 부모집단 별 인구, 가구 및 주택의 총수 정보를 이용하여 등록센서스 목적의 개인, 가구, 주택 별 가중치를 산출하고, 이를 토대로 실제 인구수, 가구수, 주택수 산출하는 것이다. 즉 행정자료를 이용한 각 관심 부모집단 별 인구수, 가구수, 주택수가 각 개인, 가구, 주택의 가중치를 1로 정의하고 산출되어짐을 감안하여 20% 표본으로부터 얻은 각 부모집단 별 불편추정량을 이용하여 각 개인, 가구, 주택 별 가중치를 보정하고 이를 인구수, 가구수, 주택수 산출을 위하여 사용하는 방안을 제안한다.

센서스 모수 추정에서 행정자료 기반 인구, 가구, 주택의 총수를 보정하기 위하여 20% 표본조사 결과를 보조정보로 사용하는 회귀추정 방안을 고려한다. 이 경우, 주어진 보조정보가 일반적으로 회귀추정량의 정의 시에 사용되는 모집단의 참값이 아닌 표본에 근거한 추정량이기 때문에 이의 변동을 고려한 추정량을 사용하여야 하기 때문에, 능형회귀추정방법을 채택하였다. 여기서는 2011년 학술연구 용역사업에서 고려된 능형회귀추정방안을 가구 및 주택의 총수 산출을 위해서도 적용했으며, 인구수 산출을 위해서는 3가지 가능한 조정 셀을 통계청과 논의하여 정의한 후 이 3가지 방안을 12개 시범지역으로 확대하여 비교/검토하였다.

추정 결과를 종합하면, 최종데이터를 이용하여 산출된 인구수 추정량들을 비교한 결과 네 가지 능형회귀 추정량의 값들은 대체로 행정자료와 조사자료 기반 추정량의 중간에 있다. 인구총수 산출을 위해 고려한 4가지 능형회귀추정방안들의 차이는 추정 셀의 정의에 따른 차이이며 네 가지 방안 중 방안 1 혹은 방안 4이 적당하다. 인구 총수 통계 공표 기준을 결정하기 위하여 살펴 본 여러 단위의 분석 결과 읍면동별(158)*성별(2)*연령5세별(21) 통계는 전체적으로 행정자료는 20%, 조사자료는 5% 그리고 능형회귀 방안 1과 방안 4는 각각 9%와 7% 정도의 상대편향을 보이며 표준편차 기준으로는 능형회귀방안이 가장 효율적이다. 그러나 특정 셀 내의 추정 인구수가 매우 적은 경우(40명 미만) 상대편향이 크게 나타나는 셀이 많고 따라서 이러한 셀에 대한 통계공표에는 주의가 필요하다.

시군구별, 성별, 연령별 통계 산출에 있어서도 능형회귀추정 방안의 적용이 적절하지만, 언급한 바와 마찬가지로 인구 총수가 적은 읍면동 혹은 고 연령대의 인구 총수 산출에는 주의가 필요하다. 가구수, 주택수 산출에 있어서도 능형회귀추정 방안이 편향과 표준오차 측면에서 통계적 정도(精度)는 적절하지만, 인구수 추정에서와 마찬가지로 공표단위 셀에 대응되는 가구수 혹은 주택수가 적은 경우 주의가 필요하다.

능형회귀 추정 방안의 효율성은 추정 셀의 정의와 더불어 능형회귀 계수를 결정하기 위하여 사용된 z 값에 의존한다. 조사자료의 통계적 효율성이 높은 공표단위 셀의 경우 신뢰계수의 값을 낮추어 행정자료 보정의 폭을 넓히는 방안을 고려할 수 있을 것이다. 제4장은 20% 표본조사 결과로부터 얻어지는 부모집단별 인구수 추정값과 행정자료를 사용하여 소지역 모형을 설정한다. 이 경우 읍면동 같은 소지역에서의 성별 연령별 부모집단에 할당되는 표본수가 상대적으로 작아서 각 부모집단별 인구수 산출시 사용되는 추정값은 신뢰할 수 없을 정도로 불안정하게 된다. 이를 극복하기 위한 방안으로 이웃의 비슷한 부모집단에서 정보를 빌려오으로써(borrowing strength) 부모집단별 추정의 효율성을 증가시키는 소지역 추정을 고려하고자 한다.

지역별 성별 연령(5세 단위)별 부모집단을 소영역(small domain)으로 설정하고 20% 표본조사에서 표본추출단위(조사구)를 고려하여 단위기반 소지역모형(unit-level small area model)을 설정한다. 지역으로는 시군구 수준과 읍면동 수준을 동시에 고려한다.

지역·성·연령별 인구수를 산출하기 위해 네 가지 추정방법 - 직접추정량(Direct estimator), 비추정량(Ratio estimator), 경험적 최량선형불편 추정량(Empirical Best Linear Unbiased Predictor; EBLUP), 계층적 베이즈 추정량(Hierarchical Bayes estimator; HB)을 사용한다. 이 때 직접추정량과 비추정량은 설계기반 추정방법, EBLUP과 HB 추정량은 모형기반 추정방법에 해당된다. 같은 방법으로 지역·가구원수별 가구수와 지역·주택유형별 주택수를 산출한다.

제4장의 주요 결론은 센서스 추정방법론에서 읍면동을 통계작성지역으로 고려할 경우 이들 소지역 추정값을 구하는 과정에서 간과해서는 안 되는 점이 있음을 상기시켜주는 것이다. 일반적으로 대지역은 소지역과는 달리, 직접추정량의 신뢰도가 상당히 높다. 그러나 소지역 추정량을 구하여 대지역의 범위에 해당하는 소지역 추정값을 합한 값은 신뢰할 만한 직접 조사 추정값과 다소 차이가 있을 수 있다.

대지역에 해당하는 소지역 추정값을 합했을 때 대지역의 직접 추정량과 같은 값을 갖도록 조정해줄 필요가 있는데 이를 벤치마킹 방법이라 한다. 능형회귀추정 방법은 회귀추정량의 벤치마킹 혹은 Calibration 성질에 의해 회귀추정량의 보조정보인 20% 표본조사 결과와 일치하게 된다. 즉, 능형회귀추정은 일종의 벤치마킹 추정이라 할 수 있다. 따라서 구한 지역·성·연령별 인구수, 지역·가구원수별 가구수, 지역·주택유형별 주택수에 대한 소지역 추정값을 벤치마킹 시킨 후 능형회귀 추정값과 비교해 보는 것이 의미가 있을 것이다.

제5장은 행정자료를 이용하여 등록센서스 통계를 작성하는 현재의 방식을 평가하는 탐색적 연구이다. 이 연구를 위해 이 연구에 함께 참여하고 있는 통계학자들이 능형회귀추정량을 이용하여 센서스 모집단(곤, 인구, 가구, 주택)의 크기를 산출하고, 이를 이용하여, 개체자료(individual records)에 가중치를 설정하는 방안을 마련하였다. 이

러한 개체자료에 대한 센서스 가중치(census weights)로 보완된 행정자료를 이용하여, 등록센서스 항목에 대하여, 네 가지의 분포표를 산출하였다.

이 연구에서 인구 부문은 등록센서스 결과 생산방식을 평가할 때, 주안점은 가중치를 적용하지 않은 행정자료와 가중치를 적용한 행정자료의 센서스 항목별 분포를 비교하는데 두었지만, 등록센서스 결과 생산방식을 평가하는 또 다른 중요한 방법은 다른 자료원의 인구자료와 행정 자료원의 인구자료를 비교하는 것이다. 이를 위해 2010 인구주택총조사에서 해당 12개 지역의 성별, 연령별, 혼인상태 분포를 구하고, 인구주택총조사의 인구분포를 기준으로 가중치 적용 전후의 행정자료들의 분포를 비교하면서 행정자료의 센서스 항목별 품질도 평가하였다.

제5장에서는 등록센서스를 실시할 때 가장 기본적인 자료가 되는 여러 형태의 행정자료 혹은 등록자료의 품질을 인구, 가구, 주택의 측면에서 평가하였다. 이를 위해 본 연구에서 선정한 12개 지역을 대상으로 등록자료를 2010년 인구주택총조사(인구센서스) 자료 및 등록자료와 조사자료에 능형회귀방정식을 적용해 도출한 능형회귀추정량(가중치)으로 산출한 분포와 비교분석하였다.

12개 지역에 대해 등록자료와 인구센서스자료와의 유사성과 차이를 인구피라미드(성별 연령별 인구구조), 연령별 성비, 성비의 차, 두 자료의 성별 연령별 오차의 비율(RLAD), 비유사성지수 등 다양한 지표를 이용해 인구, 가구, 주택부문에서 등록자료의 품질을 검토하였다.

등록자료와 인구센서스자료를 여러 각도에서 검토한 결과, 지역에 따라 등록자료가 인구센서스자료와 유사한 경우도 있었다. 하지만 등록자료와 인구센서스자료 사이에 나타나는 차이가 일관성이 적은 지역도 있었다. 두 자료의 차이를 체계적으로 유형화하기 쉽지 않은 이유는 한국사회의 지난 수십년간 장기적 변동과 지역개발, 인구유출 등 지역적 특수성 및 단기적인 요인이 복합적으로 결합하여 자료에 반영되었기 때문이다. 따라서 통계적 보완을 포함한 새로운 차원의 사회인구학적 품질평가 연구를 설계할 필요가 있다.

성 및 연령 이외의 다른 변수를 추가 조합해서 인구, 가구, 주택 등의 새로운 분포를 만들 때 나타나는 문제점을 인식하기 위해, 극단적인 경우의 하나로 자료상 정의가 완전히 다른 혼인상태별 분포를 분석하였다. 그 결과, 등록센서스를 실시한다면, 과거 인구센서스와의 시계열적 연속성을 중시할 것인지, 아니면 등록자료상의 정의를 중심으로 새로운 통계를 생산할 것인지, 통계적인 문제를 넘어선 선택의 문제가 놓여 있음을 재인식할 수 있었다.

2013년에는 인구, 가구, 주택의 다양한 항목에 대한 품질평가를 진행하여 등록센서스가 성공적으로 실시될 수 있는 발판을 마련하는 과제가 남아 있다. 제5장은 비유사성 지표 외에, 인구센서스와 행정자료, 그리고 행정자료와 센서스 가중치에 의하여 보완된 행정자료의 오차크기(error size)를 중심으로 12개 연구대상 지역간의 차이를 논의하였다. 자료간의 오차는 각 지역별로 주어진 항목에서의 오차를 구한 후, 그 절대값을 취하였다, 각 지역에서 항목에 따라 나타난 오차의 절대값을 합산하고 이 중에서 주어진 항목의 범주가 차지하는 비율을 통해 오차크기의 비율을 산정하였다. 이러한 오차크기의 비율은 지역과 항목에 따라 발생하는 전체 오차(절대값) 중에서 특정 지역과 항목에서 발생하는 오차의 상대적 크기를 제시한다.

이번 연구에서 행정자료의 혼인상태, 가구주와의 관계, 가구원수별 가구, 연면적별 주택수, 건축연도별 주택수 등에 대한 등록센서스 결과 작성방식을 평가했지만, 전반적으로 등록센서스 통계의 문제는 포괄범위를 조정하는 가중치 설정의 문제가 아니라, 행정자료 자체의 품질문제가 중요하다고 본다. 주택부문 통계를 제외하고는 혼인상태별 인구, 가구주와의 관계별 인구, 가구원수별 가구는 현재로서는 등록센서스 통계작성의 품질요건을 충족시킨다고 보기는 힘들다. 가구생성을 통하여, 행정자료의 지역별 과다범위와 과소범위를 검토하고, 필요한 경우에 그것을 보완할 수 있는 방법론을 개발할 필요가 있다고 주장한다.

제6장은 2010년과 2020년 라운드의 등록센서스 최근동향과 2015년 우리나라 등록센서스 실시에 주는 함의를 정리하고 있다. 우선 등록센서스는 단순히 등록부를 이용하는 센서스가 아니라 등록부를 주축정보(主軸情報)로 하여 통계표를 작성하는 것이라는 점에 유념한다. 물론 이러한 센서스는 표본조사를 실시하여 등록자료의 흠결을 보완하는 역할을 한다.

2020년 라운드에 실시될 예정으로 되어 있는 영국의 행정자료 옵션 센서스 중 집계수준, 개체수준, 중간수준 3개 옵션과 주소명부와 표본조사의 결합으로 이루어지는 센서스를 검토한다. 또 2020년 센서스 라운드를 기준으로 실시될 이스라엘의 순환 통합센서스(IRC)와 이탈리아의 등록 순환센서스를 검토한다.

전통적 센서스에서 등록부를 기반으로 하는 센서스로 전환하는 과정에서 다양한 어려움에 직면할 수 있다. 이 때문에, 스페인이나 독일과 같은 경우, 등록부 중에서 건물/주택대장의 불비로 전수조사를 실시하고, 이를 기반으로 새로운 등록부를 구축하는 방식을 채택하기도 하였다.

대부분의 등록센서스 실시국가에서는 Pre-Census Database 곧 통계목적의 명부를 작성하는데, 이것은 전반적으로 전통적 센서스와 마찬가지로, 개별성과 보편성의 원칙을 준수하는 방식으로 구축된다. 또 이 자료는 센서스 모수를 추정하기 위한 기초자료로서는 물론 표본 추출틀로 이용되고 있다.

영국은 2021년 이후의 센서스에서 인구대장, 곧 주민등록이 없는 상황에서 전통적 센서스를 폐기하는 방향으로 나아가고 있으며, 이것은 우리에게도 중요한 교훈을 주고 있다. 그 이유는 행정자료가 대단히 부족한 상황에서, 국민건강보험 환자등록자료나 선거인 명부를 인구부문 자료를 추출하기 위한 행정자료로 사용하고 있으며, 이를 위하여 행정자료 옵션으로 비교적 간단한 집계자료 옵션(aggregate data option)으로 전환할 가능성도 제시하고 있다. 물론, 이 경우에 영국 통계청은 포괄범위 점검조사를 과다범위와 과소범위 두 가지 면에서 실시할 계획으로 되어 있다. 또 별도의 통계조사를 실시하여, 종전의 전통적 센서스에서 획득하였던 사회경제통계 항목에 대한 자료를 획득할 계획으로 되어 있다.

이스라엘은 2008년 통합센서스의 성공을 계기로 하여, 빠른 속도로 순환 통합센서스로 나아가 있다. 특히, 주택대장을 개발하고, 교육대장, 소득대장에 대한 별도의 대장을 추가적으로 구축하여, 궁극적으로 완전 등록센서스를 목표로 하는 방향으로 나아가고 있다. 이탈리아는 2001년 센서스에서는 등록자료와 전수집계를 조합하는 방식으로 센서스를 실시하였지만, 2011년 센서스에서는 조사표를 short-form과 long-form으로 구분하여 실시하여, 2021년 센서스에서 순환 등록센서스로 이행할 준비를 하고 있다.

2021년 순환센서스에서 등록부 기준 과다와 과소를 점검하는 C-sample을 이용한 조사와, 사회경제통계를 얻기 위한 D-sample 조사를 구분할 계획이다. 이것도 중장기적으로, 우리나라의 2015년 등록센서스가 벤치 마크해야 할 바람직한 전환방향이라고 판단된다.

마지막으로, 본 연구는 등록센서스와 관련한 EU(유럽연합)의 규정을 검토하고 센서스 추정방법론을 검토한다. 여기에는 센서스 추정방법론, 특히 이원추정방법론과 소지역 추정방법의 기본원칙을 간단히 해설한다. 등록센서스 실시국가들의 이행과정에는 다양한 요인이 결부되고, 이에 따라 자료의 편집이나 삭제도 다양하게 이루어지고, 비교적 간단한 추정방법론을 채택하고 있다. 본 연구의 제7장에서는 이원시스템추정(DSE) 방법론이 일차적 관심대상이지만, 센서스를 실시하는 나라마다 직접추정, 고전적 이원시스템추정, 변형된 추정방법(예: 독일과 대만의 비추정) 등의 추정방법론이 응용되고 있으며, 이를 위하여 표본설계 적정화가 다양하게 이루어지고 있음을 알 수 있다. 그러나 등록센서스로의 전환과정에서 국가별로 다양하게 직면하게 되는 어려움이 있음을 볼 수 있고, 현재 우리나라가 추진하고 있는 센서스 추정방법론과 유사한 방식으로 능형회귀추정방법이나 소지역추정방법론에서 소개되는 다양한 추정량을 포함하여 독자적인 추정방법론을 개발하고 있음을 알 수 있다.

제7장은 우리나라의 2015년 등록센서스 추진일정을 점검하면서, 여러가지 제안을 하고 있다. 하나는 2012년의 학술연구 용역사업에서 특징적인 것은 등록센서스 담당 주무부서가 연구자들에게 조사구 단위 집계자료(aggregate data)를 제공하였다는 점이다. 실제로, 등록센서스 담당 주무부서가 개체수준의 등록자료와 조사자료를 2010년 인구주택총조사 시점에서 매칭/연계하고 있다고 판단되지만, 표본설계나 추정방법론의 개발에 있어서 세련된 방식으로 센서스 모집단을 확정하고 등록자료 기반 통계표를 작성하는데 있어서 품질점검에 필요한 지표를 개발하는데 도움이 될 수 있도록, 관련 개체수준 센서스 데이터베이스를 적극 활용하였으면 하는 기대가 있다.

2015년 등록센서스의 추정방법론 개발과정에서 행정자료는 개체수준에서 편집되고 임퓨테이션된 데이터베이스로 제공되어야 한다. 그러나 현재로서는 등록센서스의 데이터베이스를 종전의 전통적 센서스에서 사용되는 조사항목의 정의와 일관성있는 방식으로 구축하는 작업은 그리 간단하지는 않을 것이라고 판단한다. 이 때문에 등록센서스 담당부서에서는 등록센서스의 데이터베이스 구축에 필요한 기초정보를 표본조사의 수준에서 확보할 것인지 아니면 다른 방식으로 확보할 것인지에 대하여 조만간 결정을 내려야 할 것으로 판단한다.

2013년부터는 등록센서스 실시와 관련하여, 등록자료의 포괄범위를 점검하는 표본조사를 실시하여야 할 것이라고 생각한다. 표본조사에 의하여 과소범위를 조정할 수는 있지만, 별도 조사가 실시되지 않으면 과다범위를 점검하기는 쉽지는 않다. 사실, 센서스 모집단의 크기를 세련된 추정방법론을 적용하여 확정한다고 하더라도, 과다범위를 조정할 행정인구, 행정가구, 행정주택에 대한 자료를 가지고 있다면, 추가적으로 표본조사를 실시하여 과소범위를 조정하는 것이 추정결과의 신뢰성을 향상시키는데 기여한다.

인구주택총조사 목적으로, 행정자료를 사용할 때는, 목표 모집단에 대하여 행정자료의 과다집계와 과소집계의 정도를 확실하게 파악하는 것이 중요하다. 이러한 포괄범위를

조사통계의 방식으로 규명하고자 노력한 경우는 이스라엘의 2008년 통합센서스가 있고, 이탈리아가 2021년에 계획하고 있는 순환센서스에서도 포괄범위를 측정하는 문제를 포함하고 있다고 볼 수 있다. 현재, 우리나라가 시도하고 있는 추계방식은 행정자료와 조사자료의 차이를 조정하려는 노력이라고 볼 수 있겠지만, 그것이 인구추정모형의 정교한 논리에 바탕을 두어서 개선이 가능한 범위 안에서, 센서스 추정방법론을 개선할 필요성은 있다고 판단하고 있다.

보 고 서 목 차

제1장 이상 시군구 탐색 연구	1
제1절 서론	1
제2절 이상 시군구의 탐색적 연구	3
1. 이상 시군구 탐색	3
2. 가용 자료로부터 이상 시군구 판별 방안 연구	22
3. RoD를 통한 이상 시군구 탐색	36
4. SWROAD를 통한 이상 시군구 탐색	56
5. 2011년 연구의 이상읍면동과 2012년 연구의 이상 시군구 종합	68
제3절 결과의 요약	74
 [부록 1-1] 우리나라의 행정구역	75
 제2장 등록센서스를 위한 표본설계 방안 연구	77
제1절 서론	77
1. 연구목적	77
2. 연구개요	79
제2절 추출틀	81
1. 추출틀 구축	81
2. 기초분석	82
제3절 행정통계와 조사통계의 비교	86
1. 시군구	86
2. 읍면동	89
제4절 표본설계	95
1. 개요	95
2. 조사모집단의 정의	96
3. 이상치 처리	98

4. 표본크기 결정	100
5. 표본추출	102
6. 추정결과	109
제5절 결론 및 제안사항	127
1. 결론	127
2. 제안사항	128
 [부록 2]	
2-1 지역별 인구와 산점도	132
2-2 행정자료와 조사자료의 차이	138
2-3. 지역별 조사구의 인구분포	151
2-4. 읍면동 추정결과	157
 제3장 행정자료 및 표본조사자료를 이용한 인구, 가구 및 주택 통계 산출	179
제1절 머리말	179
제2절 능형회귀추정량(Ridge Regression Estimator)	181
제3절 12개 시범지역에 대한 인구, 가구 및 주택자료 분석 결과	186
1. 인구	187
2. 가구	229
3. 주택	240
제4절 최종데이터를 통한 인구, 가구 그리고 주택자료 분석 결과	251
1. 인구	251
2. 가구	268
3. 주택	281
4. 토의	294
 [부록 3-1] 인구, 가구, 주택 부문 주변분포	295
[부록 3-2] 2013년 연구의 방향	322
 제4장 소지역 추정방법론	323
제1절 지역·성·연령별 인구수 산출 방안(1)	324
1. 직접추정량	325

2. 비추정량	326
3. EBLUP 추정량	327
4. 계층적 베이지스 추정량	329
5. 추정량 비교를 위한 요약치의 계산	332
제2절 지역·성·연령별 인구 수 산출 방안(2)	341
제3절 지역·가구원수별 가구수 산출 방안	343
제4절 지역·주택유형별 주택수 산출 방안	347
제5절 향후 연구 방안	350

[부록 4-1] 인구수 추정에 대한 주변분포 (시군구)	351

제5장 등록센서스 결과(인구, 가구, 주택) 생산 방식에 대한 평가 연구 373

제1절 연구목적	373
제2절 연구방법	376
제3절 등록센서스 통계생산방식 평가: RLAD 지표를 활용한 행정자료와 인구센서스자료 비교 평가	382
1. 2010년 센서스와 행정자료의 인구	382
2. 2010년 센서스와 행정자료의 가구	419
3. 2010년 센서스와 행정자료의 주택	422
4. 소결: 2010년 센서스와 행정자료	427
제4절 등록센서스 통계생산방식 평가: 비유사성 지수의 활용	428
1. 2010년 센서스와 행정자료	428
2. 2010 인구센서스와 행정자료의 혼인상태별 비교: 센서스 가중치 적용 전후 행정자료의 인구분포 비교 1	429
3. 2010 인구센서스와 행정자료의 혼인상태별 비교: 센서스 가중치 적용 전후 행정자료의 인구분포 비교 2	434
4. 가중치 적용 전후 행정자료의 가구분포 비교	441
5. 가중치 적용 전후 행정자료의 주택분포 비교	444
제5절 자료원 차이에 따른 품질평가 요소 및 제언	446
1. 통계품질기준으로서의 일관성	446
2. 등록자료와 인구센서스 자료간 일관성: 개념 및 자료성격 차이	448

3. 등록자료내 일관성 및 시계열적 일관성 확보과제 및 제언	
제6절 종합 및 결론	457
제7절 제언: 2013년 과제	462

제6장 2010년과 2020년 라운드 등록센서스 관련 동향의 검토 46

제1절 등록부 이용 센서스 방식의 개관	477
1. 등록부 자료 + 전수집계	481
2. 등록부 자료(인구) + 전수집계(주택) + 특수목적의 표본조사	483
3. 등록부 자료 + 특수목적의 표본조사	487
4. 등록부 자료 + 기존의 표본조사	490
5. 완전 등록센서스	491
제2절 주요국의 등록센서스 추진 및 실시경과	494
1. 독일	494
2. 스위스	500
3. 스페인	509
4. 싱가포르	512
5. 대만	518
6. 폴란드	525
7. 터키	529
제3절 주요국의 2020년 라운드 센서스 계획과 준비상황	538
1. 영국의 “Beyond 2011 Programme”	538
2. 이스라엘 2020년 통합 순환센서스 운영계획	548
3. 이탈리아의 순환 등록센서스 모델	552
제4절 최근 동향의 우리나라 2015년 등록센서스에 대한 함의	562

제7장 주요국의 추정방법론과 우리나라 2015년 등록센서스의 당면과제 565

제1절 EU 센서스 규정의 목표모집단 관련 사항	565
1. 용어의 정의	566
2. 데이터 레코드의 임퓨테이션과 삭제	571
3. 포괄범위 평가작업	572
4. 센서스 가중치 설정	583

제2절 등록센서스 실시국의 표본설계와 추정방법론	584
1. 영국	587
2. 스페인	592
3. 독일	607
4. 대만	619
5. 이스라엘	624
제3절 소지역 추정의 문제	633
1. 인구학적 추정과 통계학적 추정	633
2. 소지역 추정의 문제	633
3. 추정결과의 품질진단	645
제4절 우리나라 2015년 등록센서스의 당면과제와 일정점검	647
1. 한국인구학회 주관 등록센서스 연구의 진행과정	647
2. 2015년 등록센서스의 당면과제과 일정점검	652
 <참고 자료 7-1-1> 대만의 등록센서스 보완 센서스 조사표	665
<참고 자료 7-1-2> 독일의 등록센서스 보완 센서스 조사표	667
<참고 자료 7-2> 대만의 2010년 인구주택총조사 실시계획	676

제1장 이상 시군구 탐색 연구

제1절 머리말

- 본 보고서의 이상(異常) 시군구 탐색연구는 2011년의 통계청 학술연구 용역사업 최종보고서에서와 같이 기존의 인구주택총조사의 방식대로, 읍면동의 성별, 연령별 인구, 가구의 공표를 목표로 하여 읍면동 행정자료를 가급적 사용하였다(한국인구학회, 2011). 또 행정자료와 조사자료 간의 차이가 큰 읍면동에 대하여 전수조사까지도 고려하며 표본크기를 확대하여 조사를 수행할 수도 있다는 입장에서 두 자료 간의 차이가 타 읍면동에 비하여 상대적으로 크게 나타난 지역을 “이상 읍면동”이라 정의하고 이들 이상 읍면동을 탐색하고 그 리스트를 작성하였다.
- 2011년 통계청 학술연구 용역사업의 결과 확인된 문제점은 조사자료와 행정자료의 차이를 어떻게 정의하느냐에 따라서 이상 읍면동이 다르게 선정되는 것이었다(한국인구학회, 2011). 하나는 인구규모의 상대적 차이(absolute difference)에 의한 이상지역과 절대적 차이(absolute difference)에 의한 이상지역이 일치하지 않는 문제였다. 또 다른 하나의 문제는 전체인구에서 발생하는 두 자료 간 차이에 대한 판정이 특정한 성별, 연령대의 인구차이의 크기와 무관하게 이루어지고 있었다.
- 그러나 이러한 현상은 3500여개의 읍면동과 각 읍면동에 대하여 요구되는 38개의 성별, 연령별 인구, 가구, 주택을 모두 고려하면서도 해결 할 수 없는 문제라고 판단하였다. 또한 읍면동의 조사구 단위에 이루어지는 두 자료 간의 차이에 대한 판정에서도 역시 다수의 읍면동에서 조사구의 동질성이 확보되지 않아서 문제가 제기되었기 때문에, 2012년의 통계청 학술연구 용역사업의 일

부로 진행되는 본 연구에서는 이상 읍면동보다 이상 시군구를 파악하여 보다 정밀한 표본조사가 필요한 시군구의 리스트를 제공하는 것을 목표로 설정하였다.

- 본 연구의 주 내용은 좀 더 구체적으로 말하자면, 시군구의 인구수, 가구수, 주택수 등을 함께 고려하면서 전체 및 공표대상 범주별 조사자료와 행정자료 간 상대적 차이와 절대적 차이가 종합적으로 큰 시군구를 탐색하고 이들 시군구를 판별하는데 유용한 설명변수를 가용한 행정자료내지는 참조자료 내에서 선택하고, 그것을 제안하는 것이다.
- 본 장의 구성은 다음과 같다. 제1절에서는 조사자료와 행정자료의 차이를 나타내는 효율적인 척도를 정의하기 위한 토의를 한다. 제2절에서는 제1절에서 제안한 척도를 적용하여 차이의 정도를 대, 중, 소로 구분하여 12개의 시군구를 선정하였다. 이어서 제2장에서는 참조자료를 이용하여 이상 시군구를 판별하는 방안을 연구하였고 제3장에서 2011년 학술연구 용역사업에서 판별된 이상 읍면동을 이상 시군구와 연결하였다.

제2절 이상 시군구의 탐색 연구

1. 이상 시군구 탐색

A. 조사자료와 행정자료 차이 척도 개발

- 본 장의 연구를 위하여 통계청 등록센서스 담당 주무부서가 제공한 자료에서, 시군구의 수는 조사자료 230개, 행정자료 263개였다(<표 1-1> 참고). 행정자료의 시군구는 광역시(廣域市)가 아닌 일반시(예: 고양시, 청주시, 전주시)의 33개 행정구(또는 비자치구)를 포함하고 있다(<부록 1-1>의 설명을 참고할 것). 두 자료에서, 조사인구는 50,322,927명, 행정인구는 57,293,797명이었다. 조사자료의 가구수는 18,942,123가구였고, 그 중에서 1인가구는 6,026,290가구로 약 32% 정도를 차지했다. 행정자료의 가구수는 20,591,220가구였고, 그 중 1인 가구는 4,865,161가구로 약 24% 정도를 차지했다. 조사자료의 주택수는 총 14,317,246개였고, 그 중 단독주택은 3,726,460개(약 26%), 아파트는 8,352,148개(약 58%)였고, 행정자료의 주택수는 총 16,364,613개였고, 그 중 단독주택은 4,203,792개(약 26%), 아파트는 9,918,959개(약 60%)였다.

<표 1-1> 통계청 제공 자료의 요약정리

	조사자료	행정자료
시군구 수	230	263
인구(명)	50,322,927	57,293,797
총가구(가구)	18,942,123	20,591,220
1인가구(가구)	6,026,290	4,865,161
주택수(개)	14,317,246	16,364,613
단독주택(개)	3,726,460	4,203,792
아파트(개)	8,352,148	9,918,959

- 본 장에서는 통계청 등록센서스 담당부서가 제공한 자료를 이용하여, 제1절에서는 시군구의 조사자료와 행정자료 간의 차이가 많은 지역을 찾고자 기존 척도를 검토하고 새로운 척도를 제안한다.

1. 기존 척도 특성

- 2011년 연구에서, 아래 <표 1-2>와 같이 D, AD, RoD, RoAD 및 RLAD를 이용하여 이상읍면동을 제안하였다.

<표 1-2> 2011년 연구에서 사용한 척도

RoD	$\frac{\text{조사자료} - \text{행정자료}}{\text{조사자료}}$
RoAD	$\frac{\text{건조사자료} - \text{행정자료건}}{\text{조사자료}}$
D	조사자료 - 행정자료
AD	조사자료 - 행정자료건
RLAD	$\frac{\text{건조사자료} - \text{행정자료건}}{\sum \text{건조사자료} - \text{행정자료건}}$ 여기서, $\sum$ 건조사자료 - 행정자료건는 소속지역

- 2011년 연구에서는 상대차이(RoD, RoAD)와 차이(D, AD)로 두 자료 간의 차이를 알아보았다(한국인구학회, 2011). 차이(D, AD)는 조사자료와 행정자료의 대소관계를 알기 위한 D와 대소관계가 아닌 규모의 차이를 알아보기 위한 AD로 구성되었다. 그리고 조사자료 대비 차이(D, AD)를 알아보는 상대차이

(RoD, RoAD)였으며, 여기에 추가적으로 읍면동의 차이합(sum of differences)의 상대적인 차이로 알아보았다. 그러나 위의 척도들로 조사자료와 행정자료의 차이가 큰 읍면동을 공통적으로 발견하기 어려웠으며, 연령, 성별 인구수의 차이 및 가구수, 주택수의 차이까지 모두 고려하기는 어려운 문제점을 갖고 있었다.

2. 시군구 자료와 기존 척도의 측정

- 이번 연구에서는 조사자료가 행정자료보다 많은가 또는 적은가보다 두 자료의 차이가 큰 지역을 찾는 것에 주목하였다. 그래서 기존 척도 중 AD와 ROAD만을 시군구별로 산출하였다. 그중 조사자료와 행정자료의 시군구 차이(AD, ROAD)가 큰 상위 20개의 시군구를 조사해 본 결과는 다음과 같다.
- AD가 큰 지역은 대부분 규모가 큰 ‘시’나 ‘구’ 지역이었고, RoAD가 큰 지역 대부분이 ‘군’ 지역이었다. 가령, ‘포천시’의 경우는 인구수의 AD와 RoAD가 모두 상위 20위 안에 있는 지역이었으나 다른 지역들의 경우는 AD와 RoAD의 순위가 모두 상위 20위 안에 있는 시군구는 없었다. 또한 가구 부문이나 주택 부문의 차이(AD, RoAD)도 그 순위가 모두 상위 20위 안에 있는 시군구는 없었다.

<표 1-3> 인구 차이: AD, RoAD

	시군구	조사인구	행정인구	AD	RoAD
A D 2 0 위	영등포구	374,543	428,742	54,199	0.1447
	화성시	474,160	519,966	45,806	0.0966
	평택시	381,731	422,480	40,749	0.1067
	구로구	401,239	440,153	38,914	0.0970
	안산시	699,063	736,692	37,629	0.0538
	창원시	1,047,488	1,084,904	37,416	0.0357
	포천시	136,580	167,156	30,576	0.2239
	동대문구	339,599	368,201	28,602	0.0842
	금천구	230,447	258,937	28,490	0.1236
	부천시	842,482	870,532	28,050	0.0333
	고양시	897,222	925,219	27,997	0.0312
	광주시	224,269	251,973	27,704	0.1235
	성북구	452,704	480,362	27,658	0.0611
	은평구	448,112	475,447	27,335	0.0610
	용산구	217,228	244,471	27,243	0.1254
	관악구	511,185	538,221	27,036	0.0529
	남양주시	526,639	553,513	26,874	0.0510
	파주시	323,955	350,816	26,861	0.0829
	김해시	484,244	509,525	25,281	0.0522
	수원시	1,054,053	1,078,968	24,915	0.0236
R O A D 2 0 위	울릉군	7,737	10,707	2,970	0.3839
	신안군	32,987	45,557	12,570	0.3811
	진안군	20,318	27,345	7,027	0.3459
	임실군	23,490	30,615	7,125	0.3033
	용진군	14,501	18,601	4,100	0.2827
	군위군	19,794	24,809	5,015	0.2534
	성주군	36,319	45,375	9,056	0.2493
	강서구	52,207	64,975	12,768	0.2446
	장성군	37,899	46,651	8,752	0.2309
	포천시	136,580	167,156	30,576	0.2239
	구례군	22,291	27,142	4,851	0.2176
	보성군	39,963	48,599	8,636	0.2161
	하동군	41,668	50,582	8,914	0.2139
	해남군	65,685	79,458	13,773	0.2097
	장수군	19,293	23,317	4,024	0.2086
	곡성군	26,975	32,506	5,531	0.2050
	의령군	25,249	30,378	5,129	0.2031
	순창군	25,063	30,010	4,947	0.1974
	함평군	30,770	36,787	6,017	0.1955
	강진군	34,029	40,630	6,601	0.1940

<표 1-4> 가구 차이: AD, RoAD

	시군구	조사 가구	조사 1인가구	행정 가구	행정 1인가구	가구 AD	1인 가구 AD	가구 RoAD	1인 가구 RoAD
A D 2 0 위	영등포구	138,637	34,158	162,418	61,345	23781	27187	0.1715	0.7959
	화성시	158,423	29,792	181,307	53,059	22884	23267	0.1444	0.7810
	평택시	131,814	27,617	157,738	52,735	25924	25118	0.1967	0.9095
	구로구	142,678	30,156	160,686	53,038	18008	22882	0.1262	0.7588
	안산시	249,009	61,546	268,509	90,586	19500	29040	0.0783	0.4718
	창원시	372,108	83,673	389,034	105,768	16926	22095	0.0455	0.2641
	포천시	47,413	10,140	63,288	23,902	15875	13762	0.3348	1.3572
	동대문구	128,042	36,897	148,507	59,346	20465	22449	0.1598	0.6084
	금천구	86,533	23,274	99,386	37,888	12853	14614	0.1485	0.6279
	부천시	291,513	59,244	315,584	90,334	24071	31090	0.0826	0.5248
	고양시	303,568	58,392	337,778	97,481	34210	39089	0.1127	0.6694
	광주시	72,951	11,752	90,154	27,788	17203	16036	0.2358	1.3645
	성북구	162,896	38,652	188,803	66,683	25907	28031	0.1590	0.7252
	은평구	154,923	29,795	183,494	59,625	28571	29830	0.1844	1.0012
	용산구	86,470	25,846	105,860	47,519	19390	21673	0.2242	0.8385
	관악구	217,359	84,423	236,554	109,895	19195	25472	0.0883	0.3017
	남양주시	170,933	24,346	197,340	52,393	26407	28047	0.1545	1.1520
	파주시	104,692	17,730	134,691	47,647	29999	29917	0.2865	1.6874
R O A D 2 0 위	김해시	165,258	33,100	174,206	43,568	8948	10468	0.0541	0.3163
	수원시	374,008	92,869	395,867	125,274	21859	32405	0.0584	0.3489
	울릉군	3,494	1,145	5,025	2,441	1531	1296	0.4382	1.1319
	신안군	16,114	6,070	20,952	9,430	4838	3360	0.3002	0.5535
	진안군	9,153	3,119	11,724	4,770	2571	1651	0.2809	0.5293
	임실군	10,693	3,609	13,073	5,154	2380	1545	0.2226	0.4281
	웅진군	5,952	1,954	8,701	4,150	2749	2196	0.4619	1.1238
	군위군	9,348	3,187	11,344	4,650	1996	1463	0.2135	0.4591
	성주군	15,138	4,533	19,078	7,401	3940	2868	0.2603	0.6327
	강서구	18,321	4,032	24,888	9,253	6567	5221	0.3584	1.2949
	강성군	15,071	4,642	19,525	7,695	4454	3053	0.2955	0.6577
	포천시	47,413	10,140	63,288	23,902	15875	13762	0.3348	1.3572
	구례군	9,753	3,207	11,451	4,354	1698	1147	0.1741	0.3577
	보성군	18,382	6,706	21,557	9,022	3175	2316	0.1727	0.3454
	하동군	18,345	5,973	21,369	8,115	3024	2142	0.1648	0.3586
	해남군	28,460	9,316	33,721	13,161	5261	3845	0.1849	0.4127
	장수군	8,300	2,601	9,845	3,832	1545	1231	0.1861	0.4733
	곡성군	11,704	4,246	13,581	5,462	1877	1216	0.1604	0.2864
	의령군	11,740	4,641	13,801	6,099	2061	1458	0.1756	0.3142
	순창군	10,891	3,687	12,551	4,714	1660	1027	0.1524	0.2785
	함평군	13,351	4,581	16,557	7,199	3206	2618	0.2401	0.5715
	강진군	14,985	5,262	17,542	7,078	2557	1816	0.1706	0.3451

<표 1-5> 주택 차이: AD, RoAD

	시군구	조사자료			행정자료			차이					
		주택수	단독	아파트	주택수	단독	아파트	주택수 AD	아파트 AD	단독 주택 AD	주택수 RoAD	단독 주택 RoAD	아파트 RoAD
A D 2 0 위	영등포구	87,290	19,238	58,808	88,695	18,785	58,256	1,405	552	453	0.0161	0.0235	0.0094
	화성시	143,067	24,577	106,732	150,404	27,593	110,670	7,337	3,938	3,016	0.0513	0.1227	0.0369
	평택시	116,435	29,977	69,743	122,282	29,854	72,321	5,847	2,578	123	0.0502	0.0041	0.0370
	구로구	101,947	13,935	62,302	102,932	13,039	64,476	985	2,174	896	0.0097	0.0643	0.0349
	안산시	157,912	18,337	93,072	167,547	17,340	94,230	9,635	1,158	997	0.0610	0.0544	0.0124
	창원시	286,769	77,222	192,260	293,108	77,307	193,176	6,339	916	85	0.0221	0.0011	0.0048
	포천시	43,407	19,306	16,525	40,331	16,766	15,834	3,076	691	2,540	0.0709	0.1316	0.0418
	동대문구	81,728	22,600	46,027	82,158	21,900	43,541	430	2,486	700	0.0053	0.0310	0.0540
	금천구	48,781	11,027	24,394	49,144	10,152	24,471	363	77	875	0.0074	0.0794	0.0032
	부천시	222,503	22,599	125,841	223,391	21,131	124,885	888	956	1,468	0.0040	0.0650	0.0076
	고양시	241,545	21,014	193,689	249,394	21,828	198,971	7,849	5,282	814	0.0325	0.0387	0.0273
	광주시	61,504	12,874	29,808	58,989	12,382	25,239	2,515	4,569	492	0.0409	0.0382	0.1533
	성북구	116,563	28,856	62,722	116,164	27,375	59,922	399	2,800	1,481	0.0034	0.0513	0.0446
	은평구	117,941	21,399	35,487	119,021	21,603	37,203	1,080	1,716	204	0.0092	0.0095	0.0484
	용산구	60,167	15,650	28,872	64,964	15,876	30,881	4,797	2,009	226	0.0797	0.0144	0.0696
	관악구	107,225	25,478	49,037	107,879	21,658	49,010	654	27	3,820	0.0061	0.1499	0.0006
	남양주시	159,357	14,437	120,506	163,948	14,289	121,947	4,591	1,441	148	0.0288	0.0103	0.0120
R O A D 2 0 위	파주시	96,428	23,121	65,919	96,692	20,199	68,143	264	2,224	2,922	0.0027	0.1264	0.0337
	김해시	141,199	24,794	103,445	146,719	25,831	106,800	5,520	3,355	1,037	0.0391	0.0418	0.0324
	수원시	260,423	37,723	185,720	272,065	38,288	189,924	11,642	4,204	565	0.0447	0.0150	0.0226
	울릉군	2,865	2,330	216	3,106	2,616	111	241	105	286	0.0841	0.1227	0.4861
	신안군	15,579	15,334	40	19,019	18,323	18	3,440	22	2,989	0.2208	0.1949	0.5500
	진안군	8,947	7,704	876	10,461	8,940	916	1,514	40	1,236	0.1692	0.1604	0.0457
	임실군	10,417	9,167	970	6,841	5,188	1,193	3,576	223	3,979	0.3433	0.4341	0.2299
	용진군	5,314	4,499	216	6,226	5,513	50	912	166	1,014	0.1716	0.2254	0.7685
	군위군	8,923	8,346	153	10,426	9,601	152	1,503	1	1,255	0.1684	0.1504	0.0065
	성주군	14,295	12,829	890	17,102	15,222	919	2,807	29	2,393	0.1964	0.1865	0.0326
	강서구	14,316	9,027	4,779	14,126	7,609	5,783	190	1,004	1,418	0.0133	0.1571	0.2101
	장성군	14,688	11,492	2,711	16,832	12,984	3,266	2,144	555	1,492	0.1460	0.1298	0.2047
	포천시	43,407	19,306	16,525	40,331	16,766	15,834	3,076	691	2,540	0.0709	0.1316	0.0418
	구례군	9,213	8,124	656	10,497	9,124	687	1,284	31	1,000	0.1394	0.1231	0.0473
	보성군	17,929	15,804	1,658	22,546	20,127	1,484	4,617	174	4,323	0.2575	0.2735	0.1049
	하동군	17,487	15,446	1,525	20,077	17,147	1,778	2,590	253	1,701	0.1481	0.1101	0.1659
	해남군	27,179	23,097	3,397	29,588	24,616	3,461	2,409	64	1,519	0.0886	0.0658	0.0188
	장수군	8,037	7,169	392	7,285	6,149	509	752	117	1,020	0.0936	0.1423	0.2985
	곡성군	11,207	10,137	827	12,028	10,572	960	821	133	435	0.0733	0.0429	0.1608
	의령군	11,401	10,204	807	13,379	11,822	837	1,978	30	1,618	0.1735	0.1586	0.0372
	순창군	10,560	9,240	1,000	11,864	10,253	1,046	1,304	46	1,013	0.1235	0.1096	0.0460
	함평군	13,030	12,141	544	7,073	6,068	568	5,957	24	6,073	0.4572	0.5002	0.0441
	강진군	14,353	12,466	1,062	15,883	13,554	1,204	1,530	142	1,088	0.1066	0.0873	0.1337

- 이와 같이 2011년도 연구에서 제안한 척도는 일관성 있는 결과를 얻기 힘들었다. 또한 인구, 가구, 주택 부문의 관찰범주(18범주의 성별/5세 단위 인구, 7가지의 가구원수별 가구수, 6가지 주택종류별 주택수)가 너무 많았다.
- 이러한 문제점들을 해결하기 위해서, 본 연구는 우선 범주의 개수를 줄일 수 있는 방안을 생각하였다. 첫 번째로 시군구별로 성별(남, 여) 범주를 구별할 필요가 구태여 있는지를 검토하였다. 그래서 인구 부문의 남자인구와 여자인구의 5세 단위별 절대차이(AD)와 성별을 구별하지 않은 5세 단위별 절대차이의 관계를 살펴보았다.

<표 1-6> 인구의 절대차이 상관분석 결과-성별

연령대	남-전체	여-전체	남-여	연령대	남-전체	여-전체	남-여
0-4	0.987	0.989	0.956	45-49	0.992	0.981	0.949
5-9	0.940	0.929	0.793	50-54	0.994	0.990	0.969
10-14	0.919	0.872	0.676	55-59	0.992	0.990	0.965
15-19	0.947	0.947	0.805	60-64	0.986	0.984	0.944
20-24	0.930	0.940	0.770	65-69	0.985	0.983	0.939
25-29	0.934	0.711	0.459	70-74	0.933	0.947	0.801
30-34	0.981	0.910	0.820	75-79	0.894	0.952	0.718
35-39	0.990	0.964	0.918	80-84	0.864	0.967	0.717
40-44	0.995	0.987	0.965	85+	0.947	0.988	0.891

- 상관분석(correlational analysis)의 결과에 따르면, 남자인구의 절대차이와 전체인구의 절대차이 및 여자인구의 절대차이와 전체 인구의 절대차이의 상관계수는 약 0.9정도로 높게 나타났다. 그러나 여자 25-29세에서는 0.711로 가장 낮게 나타났다.
- 또한 남자인구의 절대차이와 여자인구의 절대차이를 비교해보면 25-29세에서 상관계수가 0.459로 가장 낮았고, 이어서 10-14세가 0.676으로 나타났으나, 30대~60대까지는 상관계수가 0.9이상으로 높게 나타났다. 또한 0-4세의 상관계수도 0.956으로 높게 나타났다.

- 두 번째로 시군구별로 성별을 구별하지 않고, 특정 연령대를 묶을 수 있는지를 알아보기 위해서 인구 부문의 5세 단위별 절대차이의 관계를 살펴보았다.
- 검토 결과, 성별 5세 단위 인구의 차이(AD)는 전체 5세 단위 인구의 차이(AD)와 상관성이 높은 것으로 나타났으며, 전체 인구의 차이(AD)와 30~60대 인구의 차이(AD)와 상관성이 높은 것으로 나타났다(<표 1-7> 참고). 또한, 인구 부문뿐만 아니라 가구 부문과 주택 부문에 대하여 상관분석을 실시한 결과, 가구수 중 1인 가구의 차이(AD)는 전체 가구의 차이(AD)와의 상관계수가 0.967로 나타났으며 또한, 인구의 30~60대 인구차이(AD)와도 상관성이 높은 것으로 나타났다. 그러나 주택에서는 특이할만한 점은 발견되지 않았다.
- 그래서, 본 연구에서는 성별을 고려하지 않은 5세 단위별 인구차이(AD)를 나타내는 지표만 이용해도 무방할 것으로 판단하였다. 또한 가구수는 1인 가구수로 대표될 수 있으며, 1인 가구의 차이(AD)가 많은 지역은 인구의 차이(AD)가 많은 지역으로 인구의 차이(AD)를 나타내는 지표로 대체할 수 있다. 주택의 경우는 특이할만한 점을 찾을 수 없어 주택은 추후에 별도로 고려한다. 하지만 많은 범주를 줄일 수는 있으나 한 가지 척도를 제안하기는 어렵다. 또한 절대적인 차이와 상대적인 차이 중 어느 것도 고려하지 않을 수는 없었다. 그래서 새로운 척도를 만드는 방법을 생각하기로 하였다.

<표 1-7> 인구(5세연령 인구), 가구(1인가구), 주택(주택수, 단독주택, 아파트) 차이들의 상관분석 결과

	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)	12)	13)	14)	15)	16)	17)	18)	19)	20)	21)	22)	23)
t4Ad ¹⁾	1																						
t9Ad ²⁾	0.526	1																					
t14Ad ³⁾	0.432	0.727	1																				
t19Ad ⁴⁾	0.391	0.156	0.255	1																			
t24Ad ⁵⁾	0.300	0.222	0.212	0.618	1																		
t29Ad ⁶⁾	0.134	0.192	0.353	0.329	0.419	1																	
t34Ad ⁷⁾	0.219	0.427	0.439	0.206	0.211	0.629	1																
t39Ad ⁸⁾	0.336	0.467	0.428	0.273	0.262	0.536	0.946	1															
t44Ad ⁹⁾	0.340	0.448	0.410	0.262	0.242	0.454	0.908	0.981	1														
t49Ad ¹⁰⁾	0.241	0.474	0.398	0.166	0.170	0.406	0.872	0.939	0.957	1													
t54Ad ¹¹⁾	0.285	0.457	0.407	0.222	0.210	0.416	0.873	0.950	0.978	0.978	1												
t59Ad ¹²⁾	0.303	0.492	0.402	0.209	0.191	0.361	0.844	0.933	0.960	0.969	0.984	1											
t64Ad ¹³⁾	0.276	0.586	0.439	0.154	0.162	0.314	0.801	0.881	0.901	0.939	0.940	0.960	1										
t69Ad ¹⁴⁾	0.296	0.622	0.520	0.202	0.219	0.327	0.775	0.827	0.842	0.844	0.872	0.894	0.920	1									
t74Ad ¹⁵⁾	0.411	0.682	0.451	0.133	0.185	0.131	0.453	0.485	0.477	0.485	0.473	0.531	0.594	0.618	1								
t79Ad ¹⁶⁾	0.434	0.372	0.293	0.151	0.215	0.060	0.102	0.112	0.111	0.086	0.096	0.082	0.087	0.176	0.429	1							
t84Ad ¹⁷⁾	0.387	0.477	0.309	0.218	0.228	0.105	0.306	0.344	0.358	0.334	0.348	0.375	0.383	0.468	0.649	0.718	1						
t85Ad ¹⁸⁾	0.284	0.555	0.352	0.099	0.163	0.082	0.528	0.596	0.633	0.606	0.638	0.713	0.736	0.812	0.657	0.108	0.472	1					
가구Ad ¹⁹⁾	0.300	0.453	0.441	0.208	0.150	0.300	0.775	0.825	0.858	0.821	0.864	0.873	0.847	0.864	0.516	0.060	0.394	0.728	1				
1인가구Ad ²⁰⁾	0.425	0.459	0.412	0.265	0.191	0.259	0.749	0.845	0.883	0.818	0.872	0.883	0.844	0.843	0.527	0.120	0.429	0.737	0.967	1			
주택수Ad ²¹⁾	0.504	0.251	0.274	0.380	0.401	0.276	0.224	0.290	0.274	0.191	0.219	0.197	0.133	0.163	0.158	0.318	0.251	0.058	0.188	0.266	1		
단독주택Ad ²²⁾	-0.065	0.017	-0.026	-0.048	0.062	0.116	0.017	-0.037	-0.067	-0.037	-0.049	-0.067	-0.075	-0.021	-0.006	0.191	0.161	-0.079	-0.072	-0.089	0.338	1	
아파트Ad ²³⁾	0.414	0.257	0.331	0.413	0.341	0.130	0.248	0.330	0.338	0.246	0.310	0.278	0.220	0.238	0.189	0.255	0.249	0.149	0.348	0.416	0.569	-0.107	1

3. WROAD(i)

- *ROAD*는 특정범주가 전체 중에서 차지하는 비율을 고려하지 않는다는 점에 착안하여 해당 시군구의 전체 인구 대비 성별 *j* 연령 *k*에 해당하는 인구가 차지하는 비율($\frac{a_{ijk}}{a_i}$)을 가중치로 두는 척도를 *WROAD*라고 한다.

$$WROAD_i = \sum_j \sum_k ROAD_{ijk} \times \frac{a_{ijk}}{a_i}$$

(*i* : 시군구, *j* : 성별, *k* : 5세 단위 연령대)

a : 조사자료

- 위의 척도 *WROAD*는 인구 중 *i*번째 시군구의 성별(*j*), 연령대(*k*)의 조사인구(a_{ijk})가 많다면 *ROAD*의 수치를 크게 하기 위함이다.
- 예를 들어보면 다음 <표 1-8>과 같이 1번 시군구와 2번 시군구의 인구 차이는 각각 10이고, *ROAD*를 구하면 1/3으로 동일하다. 그러나 1번 시군구의 조사인구 전체는 30명으로 50%가 남자 20-24세인데 반해 2번 시군구는 조사인구 전체는 600명으로 5%가 남자 20-24세로 1번 시군구에 비해 상대적으로 낮은 비율을 차지하고 있다. 따라서 이를 위해 해당 시군구의 조사인구 중 특정 성별/연령대의 인구비를 가중치(weights)로 사용하여 *WROAD*를 생성하면, 0.5/3과 0.05/3으로 10배 차이로 구별할 수 있게 된다.

<표 1-8> WROAD 적용 예

시군구	조사 인구	남자 20-24세 조사인구	남자 20-24세 행정인구	ROAD	WROAD
1	60	30	20	$\frac{30-20}{30} = \frac{1}{3}$	$0.5 * \frac{1}{3}$
2	600	30	20	$\frac{30-20}{30} = \frac{1}{3}$	$0.05 * \frac{1}{3}$

4. SWROAD(i)

- WROAD는 해당 시군구의 성별, 연령별 인구의 비율을 가중치로 한 것으로 해당 시군구의 성별, 연령별 전체 규모 때문에 생기는 차이(AD)를 고려하지 않는다는 약점을 가지고 있다.
- 그래서 WROAD에 성별 j , 연령 k 에 해당하는 전체 인구 대비 i 번째 시군구의 성별 j , 연령 k 에 해당하는 인구의 비율(a_{ijk}/T_{jk})을 가중치로 하여 이상 시군구를 찾아본다. 이를 SWROAD라고 한다.

$$SWROAD_i = \sum_j \sum_k ROAD_{ijk} \times \frac{a_{ijk}}{a_i} \times \frac{a_{ijk}}{T_{jk}}$$

T_{jk} : 성별 j , 연령 k 인 인구합

(i : 시군구, j : 성별, k : 5세 단위 연령대)

- SWROAD는 해당 시군구의 규모를 위해 조정하였는데 예를 들어보면 다음 <표 1-9>와 같이 1번 시군구와 2번 시군구의 인구 차이는 각각 10과 100이고, ROAD를 구하면 1/3으로 동일하다. 또한 해당 시군구의 조사인구 중 남자 20-24세의 인구도 50%로 동일하여 WROAD가 동일하다. 그러나 1번 시군구는 20-24세 남자인구가 30명인데 반해 2번 시군구는 300명으로 시군구의 특

정 성별 .연령대의 규모를 반영하지 못하고 있다. 따라서 이를 위해 해당 시군구의 조사인구 중 특정 성별/연령대의 인구비를 가중치로 사용하여 SWROAD를 생성하면, $0.3/3 \times 30/660$ 과 $0.5/3 \times 300/660$ 으로 10배 차이로 구별할 수 있게 된다.

<표 1-9> SWROAD 적용 예

시군구	조사인구	남자 20-24세 조사인구	남자 20-24세 행정인구	ROAD	WROAD	SWROAD
1	60	30	20	$\frac{30-20}{30} = \frac{1}{3}$	$0.5 \times \frac{1}{3}$	$0.5 \times \frac{1}{3} \times \frac{30}{660}$
2	600	300	200	$\frac{300-200}{300} = \frac{1}{3}$	$0.5 \times \frac{1}{3}$	$0.5 \times \frac{1}{3} \times \frac{300}{660}$
...		...				
전체		660				

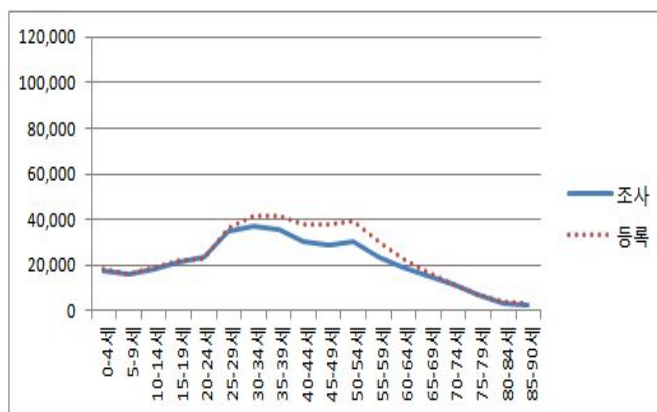
5. 시군구 자료 적용 결과 예시

- 본 연구에서는 조사자료와 행정자료의 차이에 주목하기로 하였고, 이에 기존 척도 중 AD와 ROAD를 이용하고자 하였다. 그러나 두 척도는 해당 시군구의 규모나, 연령별 성별 규모를 고려하지 않고 있다는 문제가 있었다. 그래서 우선 AD와 ROAD중 조사자료의 전체 규모를 고려한 ROAD를 이용하고 이 값에 특정범주가 전체 중에서 차지하는 비율 및 시군구의 성별, 연령별 전체 규모를 고려하여 SWROAD를 제안하였다.
- 230여개의 시군구에 적용하였을 때 기존 척도와 새로 제안된 척도를 비교한다. 인구 부문의 SWROAD를 크기 순으로 정렬하여 상위 순위 7개와 ROAD와 AD가 모두 상위 20위 안에 들었던 ‘포천시’를 <표 1-10>에 제시하였다.
- ROAD, AD가 SWROAD가 모두 상위인 영등포구와 포천시, ROAD의 순위는 높으나 AD의 순위는 낮은 편인 수원시에 대하여, 그 특성을 살펴보았다.
- 영등포구는 ROAD(인구)만 고려하면 차이는 상위 47번째에 해당하지만 AD(인구)만 고려하면 차이는 상위 1번째에 해당한다. 새로운 척도인 SWROAD(인구)로는 상위 1번째에 해당하게 된다. 수원시 같은 경우는 ROAD(인구)만 고려하면, 그 차이는 상위 201번째에 해당하지만 AD(인구)만 고려하면, 그 차이는 상위 20번째에 해당한다. 새로운 척도인 SWROAD(인구)로는 상위 7번째에 해당하게 된다. 포천시는 ROAD(인구)만 고려하면, 그 차이는 상위 10위에 해당하고, AD(인구)만 고려하면, 그 차이는 상위 7번째에 해당한다. 새로운 척도인 SWROAD(인구)로는 상위 13번째에 해당하게 된다.
- 여기서 영등포구는 연령별로 4000명 이하의 인구 분포를 보이는데, 20대에서 60대 사이의 연령에서 조사인구와 행정인구의 간격이 수원시나 포천시보다 더 차이가 많은 것을 볼 수 있다. 수원시는 60대 이하에서 4000명 이상의 인구 분포를 보이고 있으며 30대에서 50대 사이의 연령에서 조사인구와 행정인

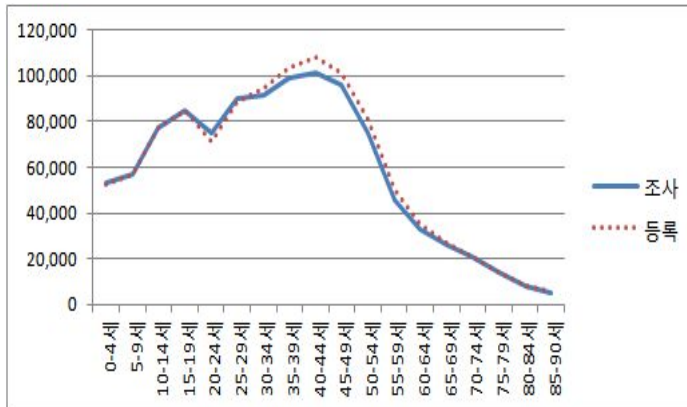
구의 차이가 있었으나 차이의 정도는 영등포구보다 적은 것으로 나타났다. 포천시는 연령별로 4000명 이하의 인구 분포를 보이고 있으며 그 차이가 영등포구나 수원시에 비해 작은 것으로 나타났다.

- 영등포구, 포천시와 같이, 그 차이(ROAD, AD)가 상위권인 경우 SWROAD도 상위권을 유지하고 있었다. 그러나 포천시의 경우는 상대적으로 인구규모가 작아서 SWROAD는 상대차이(ROAD)나 절대차이(AD)의 순위보다 하위권으로 내려가는 것으로 나타났다. 수원의 경우는 절대차이(AD)는 크지만 인구규모가 크기 때문에 상대차이(ROAD)는 낮아졌다. 그러나 인구규모 등을 고려한 SWROAD의 순위가 상위권으로 높아지는 것으로 나타났다.
- 전반적으로 평가할 때, SWROAD는 절대차이(AD)와 상대차이(ROAD) 양자를 잘 조합하여 나타내주는 척도라고 볼 수 있다.

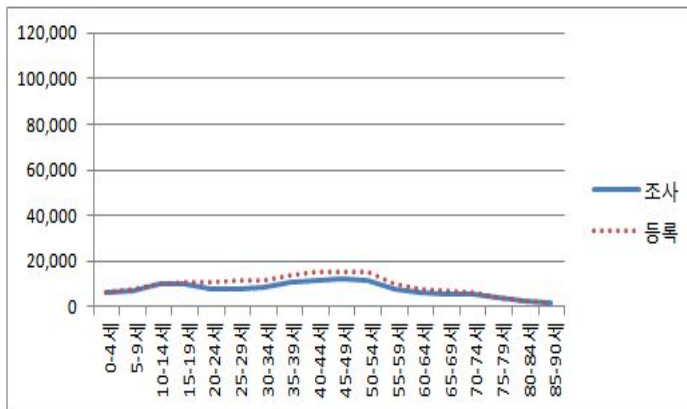
<그림 1-1> 연령별 인구(영등포구)



<그림 1-2> 연령별 인구(수원시)



<그림 1-3> 연령별 인구(포천시)



<표 1-10> 시군구 자료 적용 결과 예시(괄호안은 순위)

	시군구 ID	11190	31240	31090	11210	31070	38110	31010	31270
	시군구 명칭	영등포구	화성시	안산시	팔악구	평택시	창원시	수원시	포천시
지표 관련	인구Ad	54,199 (1)	45,806 (2)	37,629 (5)	27,036 (16)	40,749 (3)	37,416 (6)	24,915 (20)	30,576 (7)
	인구RoAD	0115	0097	0061	0053	0107	0035	0201	0024
	가구Ad	23,781 (11)	22,884 (12)	19,500 (22)	19,195 (25)	25,924 (8)	16,926 (32)	21,859 (15)	15,875 (34)
	주택수Ad	1,405 (149)	7,337 (14)	9,635 (4)	654 (194)	5,847 (28)	6,339 (23)	11,642 (3)	3,076 (84)
	sWRoAD인구	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001
	sWRoAD가구	0.0016	0.0011	0.0021	0.0024	0.0014	0.0027	0.0024	0.0008
	sWRoAD주택	0.0034	0.0016	0.0025	0.0013	0.0008	0.0018	0.002	0.0011
조사	인구	374,543	474,160	699,063	511,185	381,731	1,047,488	1,054,053	136,580
	0~4세	17,673	39,708	34,250	19,631	20,951	51,112	53,153	6,111
	5~9세	15,892	35,725	39,389	18,321	23,044	53,868	57,221	7,050
	10~14세	18,436	34,323	54,656	22,295	28,944	74,872	77,148	9,640
	15~19세	21,293	25,958	60,921	27,970	27,643	79,436	84,800	9,647
	20~24세	23,550	21,959	47,854	43,456	21,610	64,096	74,897	8,061
	25~29세	34,511	35,908	53,560	69,032	26,587	79,251	90,151	7,979
	30~34세	36,960	51,950	56,980	56,732	30,385	83,338	91,774	8,371
	35~39세	35,541	55,213	67,083	46,391	36,270	89,682	98,693	10,889
	40~44세	30,264	43,914	71,910	35,972	35,728	96,115	101,512	11,670
	45~49세	28,984	33,598	68,177	33,803	32,646	96,544	96,352	12,430
	50~54세	30,173	26,820	51,711	35,261	26,756	88,190	75,015	11,721
	55~59세	23,455	18,366	27,900	29,618	17,776	61,465	45,945	7,626
	60~64세	19,109	14,451	18,867	25,689	14,107	42,452	33,134	6,199
	65~69세	14,916	11,704	14,868	20,135	12,848	30,654	25,879	5,830
	70~74세	11,228	10,489	13,308	13,594	11,867	24,728	20,752	5,711
	75~79세	6,839	7,384	9,329	7,321	7,999	17,143	14,182	4,028
	80~84세	3,482	4,009	5,079	3,721	4,223	9,313	8,233	2,161
	85~90세	2,237	2,681	3,221	2,243	2,347	5,229	5,212	1,456
	총가구	138,637	158,423	249,009	217,359	131,814	372,108	374,008	47,413
	1인가구	34,158	29,792	61,546	84,423	27,617	83,673	92,869	10,140
	주택수	87,290	143,067	157,912	107,225	116,435	286,769	260,423	43,407
	단독주택	19,238	24,577	18,337	25,478	29,977	77,222	37,723	19,306
	아파트	58,808	106,732	93,072	49,037	69,743	192,260	185,720	16,525
등록	인구	428,742	519,966	736,692	538,221	422,480	1,084,904	1,078,968	167,156
	0~4세	18,165	39,162	33,602	19,830	21,473	49,906	52,150	6,348
	5~9세	16,330	36,095	39,209	18,590	23,568	53,891	56,926	7,468
	10~14세	18,709	34,968	54,328	22,320	29,786	74,831	76,983	10,103
	15~19세	21,984	27,001	59,689	27,224	28,856	82,811	84,550	11,053
	20~24세	22,998	24,879	47,122	37,201	23,712	68,988	71,270	11,060
	25~29세	36,273	41,546	56,627	67,253	30,478	82,238	88,448	11,851
	30~34세	41,431	58,023	62,789	60,661	34,709	86,809	94,788	11,878
	35~39세	41,886	61,625	74,153	51,317	40,720	94,372	103,782	14,121
	40~44세	38,203	49,845	79,519	41,898	40,592	101,537	107,970	15,052
	45~49세	38,128	38,391	74,188	38,840	37,055	99,732	101,129	15,442
	50~54세	39,179	31,768	57,239	40,834	31,698	92,887	80,814	15,171
	55~59세	30,053	21,708	31,600	33,532	21,028	64,684	49,843	9,789
	60~64세	23,075	16,363	20,805	28,196	16,312	43,605	35,062	7,442
	65~69세	16,752	13,124	15,770	21,861	14,421	31,659	26,954	6,710
	70~74세	11,843	11,036	13,077	14,098	12,588	25,073	20,806	6,062
	75~79세	7,099	7,541	8,968	7,578	8,140	17,046	13,784	3,940
	80~84세	3,741	4,076	4,837	3,991	4,362	9,290	8,071	2,123
	85~90세	2,893	2,815	3,170	2,997	2,982	5,545	5,638	1,543
	총가구	162,418	181,307	268,509	236,554	157,738	389,034	395,867	63,288
	1인가구	61,345	53,059	90,586	109,895	52,735	105,768	125,274	23,902
	주택수	88,695	150,404	167,547	107,879	122,282	293,108	272,065	40,331
	단독주택	18,785	27,593	17,340	21,658	29,854	77,307	38,288	16,766
	아파트	58,256	110,670	94,230	49,010	72,321	193,176	189,924	15,834

B. SWROAD를 적용한 12개 시군구 선정

- SWROAD를 이용하여, 차이의 크기에 따라 대/중/소로 분류하여 이상 시군구를 판별한다. SWROAD를 인구, 가구, 주택 세 부문에서 공통으로 상위 40위까지에 해당하는 시군구는 다음 <표 1-11>와 같다.

<표 1-11> SWROAD 공통 40위 시군구 리스트

강남구	수원시
고양시	안산시
관악구	영등포구
구로구	용산구
금천구	용인시
동대문구	은평구
마포구	종로구
부천시	창원시
서구	천안시
서초구	포천시
성남시	화성시

- SWROAD 순위를 살펴보면, 상위권의 경우에는 인구, 가구, 주택의 상위 40위에 해당하는 공통적인 시군구가 나타나지만, 중위권에 해당하는 90위-140위의 공통적인 시군구는 나타나지 않았다. 이 때문에 가구, 주택 부문의 순위범위를 늘려가면서 공통적인 시군구를 찾았다.
- 여기서, 순위범위를 늘리는 순서는 차이가 많은 주택 부문의 순위범위를 많이 늘리는 것으로 하였다.
- SWROAD를 인구 부문은 상위 90위-140위, 가구 부문은 상위 80위-140위, 주택 부문은 상위 70위-160위에서 공통으로 해당하는 시군구는 다음 <표 1-12>와 같다.

<표 1-12> SWROAD 인구 90-140위, 가구 80-140위,
주택 70-160위 시군구 리스트

강릉시	북구
강화군	사상구
광명시	수성구
김천시	의성군
남구	창녕군
남원시	충주시
달서구	하동군
동구	화순군
동래구	

- SWROAD 순위에서 중위권과 마찬가지로, 하위권은 하위 40위의 공통적인 시군구가 많지 않아 차이의 정도는 나타나지 않았다. 이 때문에 주택 부분의 순위범위를 늘려 공통적인 시군구를 찾았다.
- SWROAD를 인구 부문 하위 40위, 가구부분 하위 40위, 주택부분 하위 50위에서 공통으로 해당하는 시군구는 다음 <표 1-13>과 같다.

<표 1-13> SWROAD 인구 하위 40위, 가구 하위 40위,
주택 하위 40위 시군구 리스트

계룡시	양양군
고령군	연수구
고성군	영양군
과천시	영월군
남구	울릉군
동구	의왕시
동해시	인제군
보은군	증평군
봉화군	철원군
산청군	청양군
삼척시	화천군
양구군	

- 본 연구는 위에서 살펴본 SWORD 척도를 이용하여 각 차이의 정도(대, 중, 소)에 따라 6개의 시군구를 선정하였으며, 이 자료를 바탕으로 통계청 등록 센서스 담당 주무부서에서 2012년 학술연구 용역사업을 위한 12개의 시군구 사례연구 지역을 선정하였다. 그 결과가 <표 1-14>와 같다.

<표 1-14> 차이의 정도에 따른 시군구 구분 결과

제안	차이	통계청 확정
고양시	대	금천구
영등포구		영등포구
안산시		인천 서구
수원시		화성시
창원시		
화성시		
강릉시	중	광주 동구
남원시		나주시
김천시		김천시
동구		
남구		
부여군		
계룡시	소	양구군
영양군		증평군
양구군		임실군
과천시		신안군
울릉군		울릉군
증평군		

2. 가용 자료로부터 이상 시군구 판별 방안 연구

- 본 절에서는 표본조사에서 특별한 처리(표본설계, 크기 등에서)가 필요하다고 판단되는 이상시군구를 행정자료를 비롯한 등록센서스를 보완하는 표본조사 시점에서 가용한 자료(참조자료)를 통하여 판별하는 방안을 토의한다. 본 연구에서 가용한 참조자료는 KOSIS(한국통계포털, Korea Statistical Information Service)의 전국사업체조사에 획득한 자료를 이용하는 것으로 한다(통계청, 2012).

<그림 4-1> 참조자료: 전국사업체조사(<http://www.kosis.kr>)



- 자료연계 : 2009년 기준 시군구별 사업체 자료와 조사자료 및 행정자료를 연

계한다. 2009년 기준 시군구별 사업체 자료에는 총 241개 시군구 자료가 포함되어 있고, 조사자료에는 230개 시군구, 행정자료에는 263개 시군구 자료가 포함되어 있다. 본 연구에서는 최종적으로 211개 시군구 자료를 이용하였다.

A. 211개 시군구 자료 특성 분석

- 211개 시군구 자료에 기존의 척도를 적용하여 특성을 파악한 결과를 정리하면 아래와 같다.

1. 시군구 분석

- 211개 시군구의 총인구는 조사자료의 경우 43,285,141명, 행정자료는 45,316,464명으로 두 자료 간에는 2,031,323명(5%이내)의 차이가 있다.
- 시군구별로는 AD(조사자료-행정자료)는 최소 732명에서 최대 54,199명으로 나타났으며 평균 10,217명으로 나타났다. $ROD((\text{조사자료}-\text{행정자료})/\text{조사자료})$ 는 최소 -38.39%에서 최대 3.29%로 나타났으며 평균은 -8.98%로 나타났다.

<표 1-15> ROD와 AD의 분포

		ROD	AD
분위수	최대값	0.0329	54,199
	99%	0.0229	38,914
	95%	0.0053	27,335
	90%	-0.0107	21,720
	75%	-0.0356	12,758
	50%	-0.0734	7,638
	25%	-0.1331	5,015
	10%	-0.1829	2,794
	5%	-0.2161	2,036
	1%	-0.3459	845
	최소값	-0.3839	732
평균		-0.0898	10,217
표준편차		0.0746	8,284

- ROD가 5%이상 차이가 나는 경우는 211개 시군구 중 131개(62.1%)다. 5%이상 차이가 나는 시군구는 모두 행정자료가 많은 시군구로 이를 16개 광역시도로 분류하여 보면 다음과 같다.

<표 1-16> 211개 시군구의 Rod와 AD

코드	시군구	Rod	AD	코드	시군구	Rod	AD	코드	시군구	Rod	AD
37430	울릉군	-0.3839	2,970	35060	김제시	-0.1419	11,739	32370	화천군	-0.0925	2,036
36480	신안군	-0.3811	12,570	32350	경선군	-0.1417	5,074	33310	청원군	-0.0910	12,929
35320	진안군	-0.3459	7,027	35370	고창군	-0.1417	7,516	26310	울주군	-0.0887	16,477
35350	임실군	-0.3033	7,125	37350	영덕군	-0.1408	5,101	33350	진천군	-0.0869	5,152
23320	웅진군	-0.2827	4,100	32400	고성군	-0.1382	3,656	37080	상주시	-0.0856	8,346
37310	군위군	-0.2534	5,015	37370	고령군	-0.1381	4,281	21310	기장군	-0.0852	7,933
37380	성주군	-0.2494	9,056	11010	종로구	-0.1350	20,388	11060	동대문구	-0.0842	28,602
21120	강서구	-0.2446	12,768	38350	남해군	-0.1331	5,804	38380	함양군	-0.0842	3,175
36450	장성군	-0.2309	8,752	32380	양구군	-0.1315	2,532	21030	동구	-0.0816	7,613
36330	구례군	-0.2176	4,851	38330	창녕군	-0.1315	7,177	36020	여주시	-0.0808	21,720
36360	보성군	-0.2161	8,636	33380	단양군	-0.1314	3,680	37030	김천시	-0.0774	9,818
38360	하동군	-0.2139	8,914	38340	고성군	-0.1304	6,652	37390	철곡군	-0.0763	8,551
36400	해남군	-0.2097	13,773	36420	무안군	-0.1259	8,520	23020	동구	-0.0750	5,458
35340	장수군	-0.2086	4,024	34370	예산군	-0.1256	9,700	34390	당진군	-0.0749	10,116
36320	곡성군	-0.2050	5,531	11030	용산군	-0.1254	27,243	38060	사천시	-0.0735	7,799
38310	의령군	-0.2031	5,129	32360	철원군	-0.1249	5,384	32070	삼척시	-0.0734	4,928
35360	순창군	-0.1974	4,947	11180	금천군	-0.1236	28,490	22310	달성군	-0.0724	12,112
36430	함평군	-0.1956	6,017	37420	울진군	-0.1184	5,537	33390	증평군	-0.0715	2,231
36390	강진군	-0.1940	6,601	36370	화순군	-0.1181	7,303	21020	서구	-0.0711	8,258
36470	진도군	-0.1877	5,332	34340	서천군	-0.1171	6,281	22010	중구	-0.0680	4,925
33360	괴산군	-0.1873	5,829	34330	부여군	-0.1168	7,836	22030	서구	-0.0663	13,947
35380	부안군	-0.1829	9,250	37340	영양군	-0.1158	1,905	35310	완주군	-0.0653	5,357
36440	영광군	-0.1807	8,753	33340	영동군	-0.1139	5,238	37020	경주시	-0.0637	16,092
36380	장흥군	-0.1801	6,410	32310	홍천군	-0.1133	7,084	11040	성동구	-0.0633	18,362
36040	나주시	-0.1785	13,890	38370	산청군	-0.1130	3,583	34020	공주시	-0.0620	7,486
32320	횡성군	-0.1784	6,690	38080	밀양시	-0.1124	11,081	11220	서초구	-0.0620	24,061
23310	강화군	-0.1768	9,978	35050	남원시	-0.1108	8,688	37060	영주시	-0.0618	6,689
39020	서귀포시	-0.1749	22,693	35040	정읍시	-0.1090	11,933	31080	동두천시	-0.0614	5,551
34380	태안군	-0.1738	9,328	37410	봉화군	-0.1079	3,352	11080	성북구	-0.0611	27,658
38400	합천군	-0.1722	7,473	34310	금산군	-0.1072	5,622	11120	은평구	-0.0610	27,335
37360	청도군	-0.1677	6,363	31070	평택시	-0.1068	40,749	11140	마포구	-0.0591	21,487
36460	완도군	-0.1643	7,638	32390	인제군	-0.1055	3,017	11130	서대문구	-0.0581	17,864
36350	고흥군	-0.1636	10,337	37400	예천군	-0.1052	4,499	11070	중랑구	-0.0575	23,018
36310	담양군	-0.1633	6,652	34350	청양군	-0.1041	3,072	11090	강북구	-0.0563	18,168
11020	종구	-0.1590	18,639	37330	청송군	-0.1041	2,490	31090	안산시	-0.0538	37,629
38320	함안군	-0.1586	9,407	34060	논산시	-0.1039	12,227	35020	군산시	-0.0529	13,695
36410	영암군	-0.1579	8,788	34030	보령시	-0.1032	10,016	11210	관악구	-0.0529	27,036
33370	음성군	-0.1563	12,758	38390	거창군	-0.1026	5,859	38070	김해시	-0.0522	25,281
35330	무주군	-0.1556	3,385	33330	옥천군	-0.1002	4,937	11050	광진구	-0.0520	18,644
32340	평창군	-0.1539	5,748	38050	통영시	-0.0992	12,683	23080	서구	-0.0513	19,944
32330	영월군	-0.1538	5,365	37070	영천시	-0.0989	9,334	36030	순천시	-0.0500	12,899
23010	중구	-0.1498	12,260	11170	구로구	-0.0970	38,914				
33320	보은군	-0.1496	4,534	37090	문경시	-0.0967	6,642				
37320	의성군	-0.1455	7,425	32410	양양군	-0.0960	2,429				
11190	영등포구	-0.1447	54,199	34360	홍성군	-0.0935	7,653				

<표 1-17> 사업체조사 자료와 연계하기 전

시 코드	n ¹	n ²	%	시 코드	n ¹	n ²	%
서울(11)	25	17	68.0	강원(32)	18	12	66.7
부산(21)	16	4	25.0	충북(33)	14	11	78.6
대구(22)	8	3	37.5	충남(34)	18	13	72.2
인천(23)	10	5	50.0	전북(35)	16	14	87.5
광주(24)	5	0	0.0	전남(36)	22	20	90.9
대전(25)	5	0	0.0	경북(37)	25	21	84.0
울산(26)	5	1	20.0	경남(38)	23	19	82.6
경기(31)	51	37	72.5	제주(39)	2	1	50.0

n¹ : 해당시의 시군구 개수

n² : 해당시의 조사-인구 차이 5% 이상 시군구 개수

% : $n^2/n^1 \times 100$

<표 1-18> 사업체조사 자료와 연계한 후

시 코드	n ¹	n ²	%	시 코드	n ¹	n ²	%
서울(11)	25	17	68.0	강원(32)	18	12	66.7
부산(21)	16	4	25.0	충북(33)	12	9	75.0
대구(22)	8	3	37.5	충남(34)	16	11	68.8
인천(23)	10	5	50.0	전북(35)	14	12	85.7
광주(24)	5	0	0.0	전남(36)	22	20	90.9
대전(25)	5	0	0.0	경북(37)	23	19	82.6
울산(26)	5	1	20.0	경남(38)	18	14	77.8
경기(31)	12	3	25.0	제주(39)	2	1	50.0

n¹ : 해당시의 시군구 개수

n² : 해당시의 조사-인구 차이 5% 이상 시군구 개수

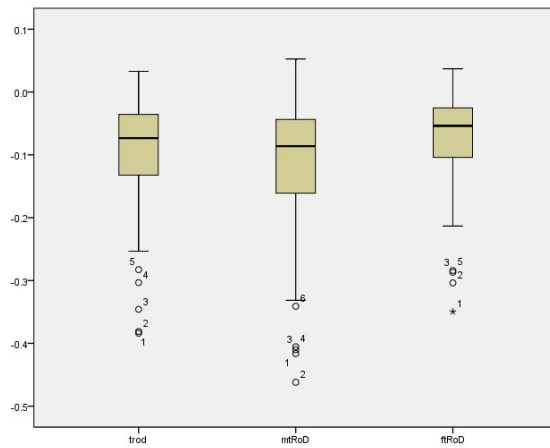
% : $n^2/n^1 \times 100$

- 본 연구의 211개 시군구를 사업체조사 자료의 시군구를 연계하였을 때, 경기도의 경우는 51개 시군구 중에서 12개 시군구만이 연계가 가능하여, 연계율(linkage rate)이 1/4 수준에 불과하였다. 현재 이용할 수 있는 자료만을 이용하여 연계된 자료를 이용하여 이후 분석을 실시한다.

2. 성별, 연령별 분석

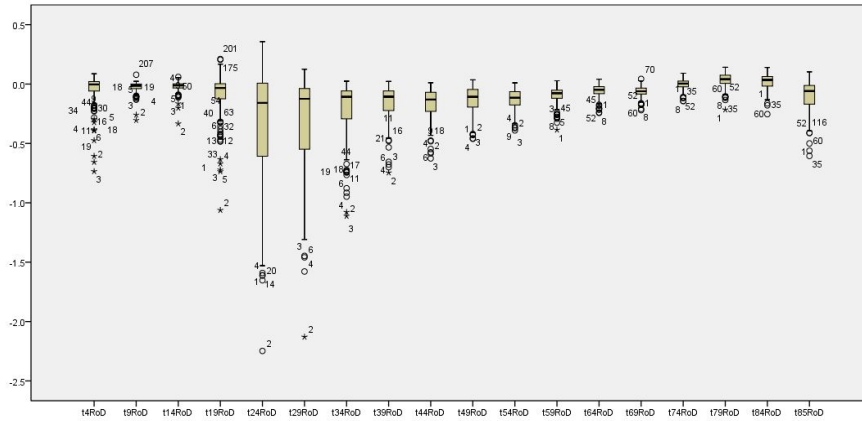
- **성별** : ROD분포는 남자가 여자보다 넓게 퍼져있고, 그 차이가 (-)인 경우 (행정자료가 조사자료보다 많은 경우)가 더 넓게 분포하고 있다.

<그림 1-5> 성별에 따른 RoD 분포



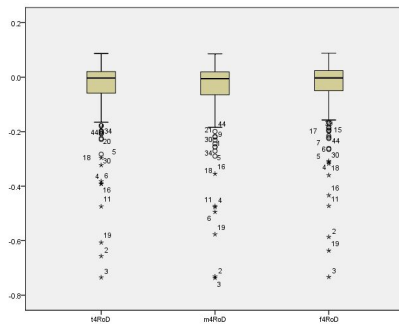
- **연령** : 연령대별로 작성된 RoD분포를 보면, 학령기의 5-9세(t9RoD) 및 10-14세(t14RoD)의 RoD분포가 한 곳에 집중되어 있는 편이었으며, 20대(20-24세 : t24RoD , 25-29세 : t29RoD)의 RoD분포가 가장 넓게 퍼져 있는 것으로 나타났다. 20대로부터 연령대가 높아짐에 따라 80-84세(t84RoD)까지는 RoD분포가 점차 줄어드는 경향을 확인할 수 있었다.

<그림 1-6> 연령대에 따른 RoD분포

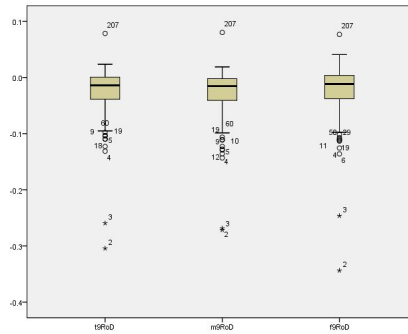


- **연령별*성별** : 연령별 ROD의 분포를 성별로 비교해보면 대부분의 경우가 비슷한 형태를 보이고 있다. 그런데, 20-24세의 경우 남자보다 여자의 분포가 넓게 퍼져있으며, 30-34세, 35-39세 및 55-59세의 경우는 여자보다 남자의 분포가 넓게 퍼져있는 것으로 나타났다.

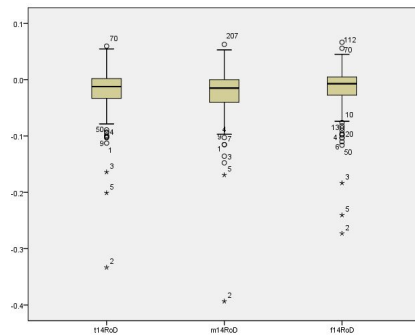
<그림 1-7> 0-4세의 성별 RoD분포



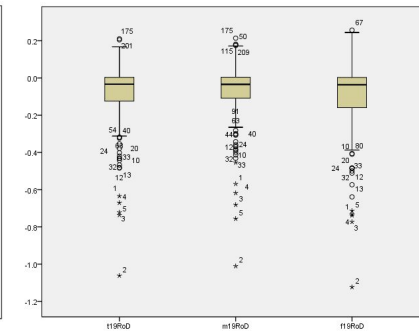
<그림 1-8> 5-9세의 성별 RoD분포



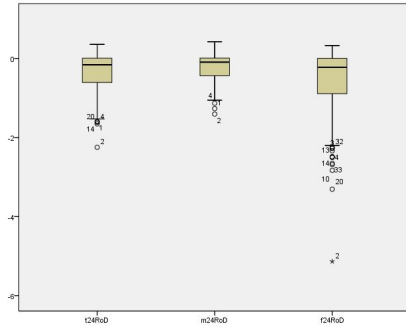
<그림 1-9> 10-14세의 성별 RoD분포



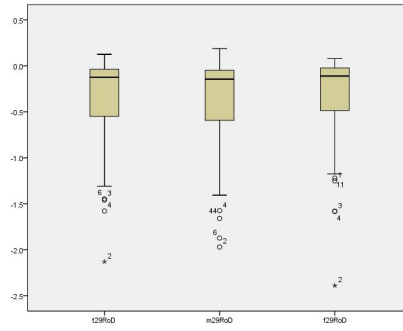
<그림 1-10> 15-19세의 성별 RoD분포



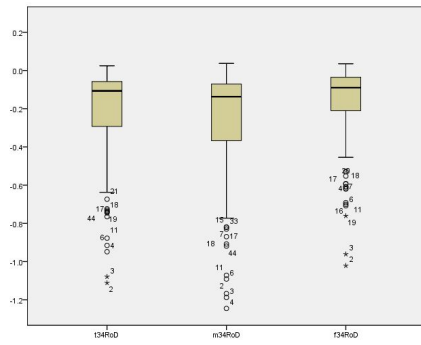
<그림 1-11> 20-24세의 성별 RoD분포



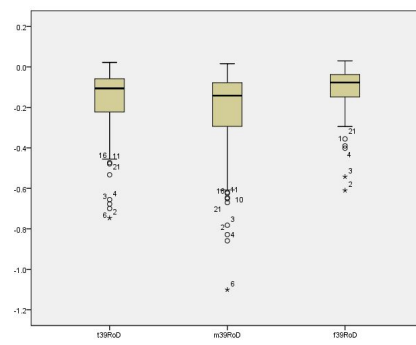
<그림 1-12> 25-29세의 성별 RoD분포



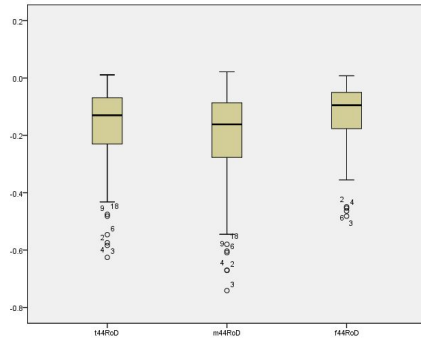
<그림 1-13> 30-34세의 성별 RoD분포



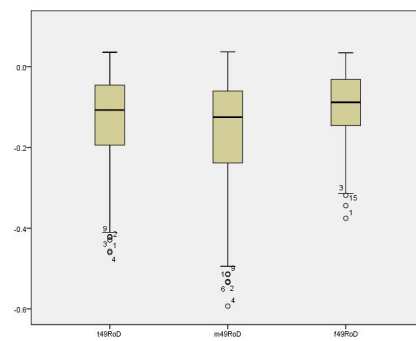
<그림 1-14> 35-49세의 성별 RoD분포



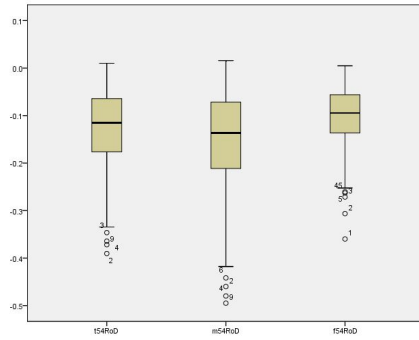
<그림 1-15> 40-44세의 성별 RoD분포



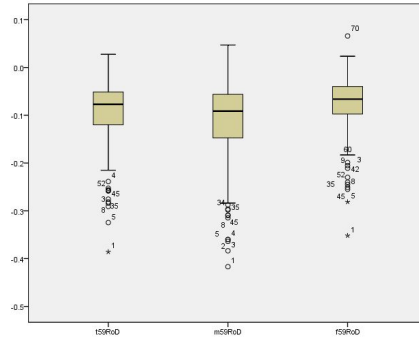
<그림 1-16> 45-49세의 성별 RoD분포



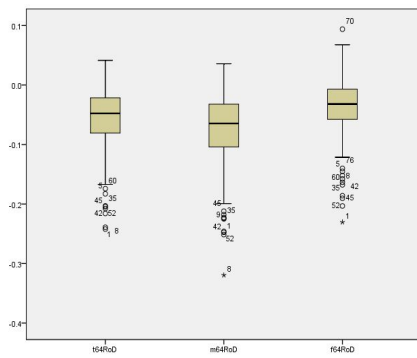
<그림 1-17> 50-54세의 성별 RoD분포



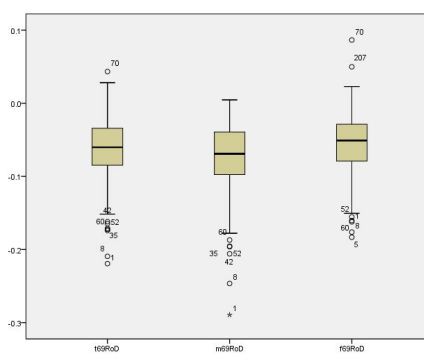
<그림 1-18> 55-59세의 성별 RoD분포



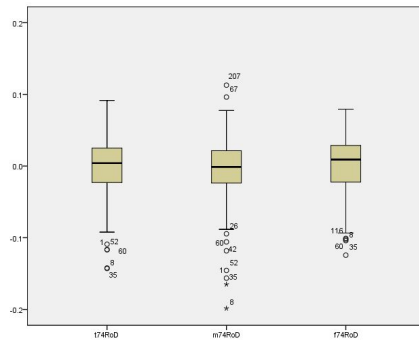
<그림 1-19> 60-64세의 성별 RoD분포



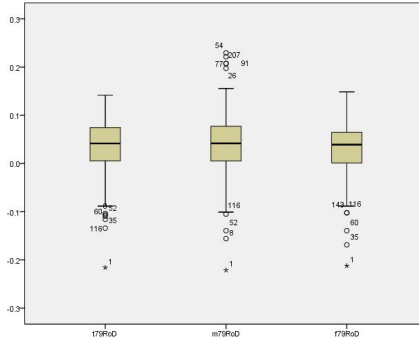
<그림 1-20> 65-69세의 성별 RoD분포



<그림 1-21> 70-74세의 성별 RoD분포



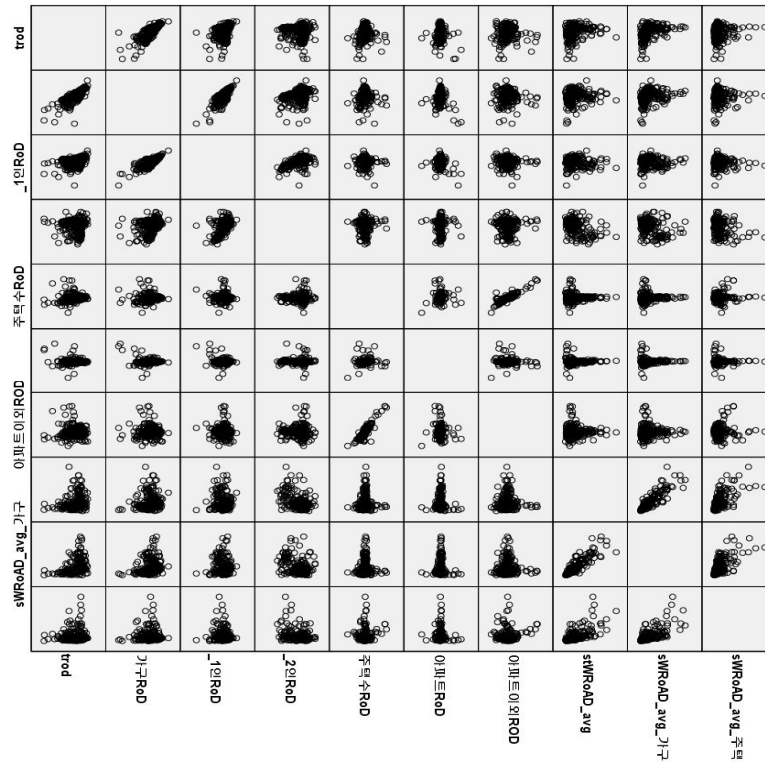
<그림 1-22> 75-79세의 성별 RoD분포



3. SWROAD와 ROD의 관계

- 여기서는 211개 시군구에 대하여 2011년 연구에서 주로 이상 읍면동을 찾는 데 사용하였던 변수인 RoD와 2012년 연구에서 제안한 SWROAD의 관계를 파악해본다(<그림 1-25> 참고).
- 아래에서 산점도와 상관분석의 결과를 살펴보면 다음과 같다. 먼저, tRoD(전체인구 RoD)와 가구RoD는 그 관계가 선형이고, 상관계수가 0.786으로 높은 것으로 나타났다. 또한 가구RoD는 1인가구ROD와 선형관계에 있고, 상관계수가 0.813으로 높게 나타나서, 조사자료와 행정자료 간의 가구수 차이는 1인가구수의 차이로 대표할 수 있을 것이다. 반면 RoD와 SWROAD(인구, 가구, 주택)은 산점도를 보면, 선형 관계가 아닌 것으로 파악되었다. 이는 SWROAD가 특성변인의 규모와 시군구의 규모를 고려하여 생성한 척도로서, RoD는 조사자료와 행정자료 간의 차이를 +와 -이로 구분하지만 SWROAD는 두 자료 간의 차이의 정도만을 알아보기 위하여 절대값을 취하는 척도이기 때문이다. 그러나 가구 SWROAD와 인구 SWROAD는 산점도를 보면 선형관계이며, 상관계수는 0.882로 나타났다. 이는 초기 상관분석의 결과와 일치하는 결과이다. 주택 SWROAD의 경우는 선형관계로 보기는 다소 어렵기는 하지만, 몇 개의 시군구를 제외한다면 선형관계로 볼 수 있을 것이다. 주택수 Rod와는 선형관련을 찾아 볼 수 없었다.

<그림 1-25> SWROAD와 RoD의 산점도



tRoD : 해당 시군구의 인구의 RoD

가구RoD : 해당 시군구의 가구의 RoD

_1인RoD : 해당 시군구의 1인 가구의 RoD

_2인RoD : 해당 시군구의 2인 가구의 RoD

주택수RoD : 해당 시군구의 주택수의 RoD

아파트RoD : 해당 시군구의 아파트 수의 RoD

아파트이외RoD : 해당 시군구의 아파트 이외 주택수의 RoD

stWRoAD_avg : 해당 시군구의 인구의 SWROAD

stWRoAD_avg_가구 : 해당 시군구의 가구수의 SWROAD

stWRoAD_avg_주택 : 해당 시군구의 주택수의 SWROAD

<표 1-19> SWROAD와 RoD의 상관계수

	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)
1) 인구 RoD	1									
2) 가구 RoD	.786 ***	1								
3) 1인 가구 RoD	.356 ***	.813 ***	1							
4) 2인 가구 RoD	-.057	.285 ***	.524 ***	1						
5) 주택수 RoD	.125	.042	-.071	.006	1					
6) 아파트 RoD	-.093	-.238 ***	-.255 ***	-.050	.120	1				
7)아파트이외 주택 ROD	.004	-.029	-.080	.017	.900 ***	-.071	1			
8) SWRoAD(인구)	.270 ***	.133	-.113	-.463 ***	.026	-.001	-.031	1		
9) SWRoAD(가구)	.419 ***	.252 ***	-.038	-.346 ***	.043	.012	-.026	.882 ***	1	
10) SWRoAD(주택)	.103	.025	-.106	-.320 ***	.145*	-.039	.103	.656 ***	.670 ***	1

*** p<.001

- 본 연구에서는 2011년의 등록센서스 학술연구 용역사업과는 달리, SWROAD를 사용하여 이상지역을 구별하고자 하였으나 SWROAD는 조사자료가 많은 지역과 행정자료가 많은 지역은 구별해 낼 수 없었다. 이에 따라 SWROAD가 큰 시군구를 이상지역으로 보는 경우와 ROD가 크거나(+) 작은(-) 시군구를 이상지역으로 보는 경우로 나누어서 이상 시군구로 판별할 수 있는지에 대해 다음의 (3)과 (4)에서 검토한다.

3. RoD를 통한 이상 시군구 탐색

A. 인구 부문

- ROD가 (+)로 너무 크거나 (-)로 너무 작아지면 이상 시군구라고 할 수 있을 것이다. 그런데 본 연구에서 사용된 시군구 단위 자료에서는 ROD가 (+)인 경우가 14개(6.6%) 지역으로 적었다. 그래서 여기서는 ROD가 (+)인 경우는 제외하고, ROD가 (-)인 경우에서 이상 시군구와 그 외(정상)로 분류하고 독립변수로 사업체조사 자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법: stepwise procedure)을 적용한다.
- 이상 시군구와 그 외(정상)의 분류는 ROD를 크기 순 정렬하여 하위 20%, 15%, 10%에 해당하는 시군구까지를 이상시군구로 하고, 그 외를 정상으로 하여 모형을 적용한다. 이때 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-20> test data와 validation data

정상군 : 이상 시군구

전체	155 : 42 (78.7% : 21.3%)	166 : 31 (84.3% : 15.7%)	176 : 21 (89.3% : 10.7%)
test data	113 : 25 (81.9% : 18.1%)	122 : 16 (88.4% : 11.6%)	128 : 10 (92.8% : 7.2%)
validation data	42 : 17 (71.2% : 28.8%)	44 : 15 (74.6% : 25.4%)	48 : 11 (81.4% : 18.6%)
이상 지역	(37430)울릉군 (36480)신안군 (35320)진안군 (35350)임실군 (23320)웅진군 (37310)군위군 (37380)성주군 (21120)강서구 (36450)장성군 (36330)구례군 (36360)보성군 (38360)하동군 (36400)해남군 (35340)장수군 (36320)곡성군 (38310)의령군 (35360)순창군 (36430)함평군 (36390)강진군 (36470)진도군 (33360)괴산군 (35380)부안군 (36440)영광군 (36380)장흥군 (36040)나주시 (32320)횡성군 (23310)강화군 (39020)서귀포시 (34380)태안군 (38400)함천군 (37360)청도군 (36460)완도군 (36350)고흥군 (36310)담양군 (11020)중구 (38320)함안군 (36410)영암군 (33370)음성군 (35330)무주군 (32340)평창군 (32330)영월군 (23010)중구		

- ROD가 상위 15%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 98.4%, 이상지역을 이상지역으로 분류할 확률은 68.8%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 90.9%, 이상지역을 이상지역으로 분류할 확률은 46.7%로 나타났다.

<표 1-21> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.002)	▶종사자_전기, 가스, 증기 및 수도사업 (0.985) ▶종사자_건설업 (1.010) ▶종사자_농업, 임업 및 어업 (0.983) ▶사업_제조업 (1.012) ▶사업_공공 행정, 국방 및 사회보장 행정 (0.708) ▶사업_보건업 및 사회복지 서비스업 (0.925) ▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.036)	▶종사자_전기, 가스, 증기 및 수도사업 (0.984) ▶사업_숙박 및 음식점업 (1.017)
test	정확 분류율	81.2%	94.9%	95.7%
	정상->정상 분류율	99%	98.4%	98.4%
	이상->이상 분류율	0%	68.8%	60%
validation	정확 분류율	74.6%	79.7%	84.7%
	정상->정상 분류율	100%	90.9%	93.8%
	이상->이상 분류율	11.8%	46.7%	45.5%

- 계속해서, 위에서 제외했던 지역인 ROD가 (+)인 14개 지역에 대해 알아보면 <표 1-22>와 같다. 대부분 시군구의 ROD가 (-)인데 ROD가 (+)인 14개의 특성을 밝혀볼 필요가 있다. 이에 ROD가 (+)인 지역이 ROD가 (-)인 지역과 구별할 수 있는 특성이 있는지 알아보고자 한다.
- 이를 위해 위와 마찬가지로 방법으로 로지스틱 회귀모형(logistic regression model)을 적용해 본다. 결과 ROD가 (-)인 지역을 ROD가 (-)인 지역으로 분류할 확률은 90.9%로 높았으나, ROD가 (+)인 지역을 (+)인 지역으로 분류할 확률은 46.7%로 낮았다.

<표 1-22> Rod가 (+)인 지역

지역 코드	지역	Rod
34010	(충남)천안시	0.0038
37100	(경북)경산시	0.0309
34040	(충남)아산시	0.0053
33010	(충북)청주시	0.0119
25040	(대전)유성구	0.0329
32010	(강원도)춘천시	0.0102
35010	(전북)전주시	0.0114
24010	(광주)동구	0.0229
24040	(광주)북구	0.0150
25030	(대전)서구	0.0046
24020	(광주)서구	0.0130
36010	(전남)목포시	0.0201
24050	(광주)광산구	0.0039
32050	(강원도)태백시	0.0164

<표 1-23> logistic regression 결과

selection=stepwise	
선택된 변수	▶종사자_전기, 가스, 증기 및 수도사업 (0.985) ▶종사자_건설업 (1.010) ▶종사자_농업, 임업 및 어업 (0.983) ▶사업_제조업 (1.012) ▶사업_공공행정, 국방 및 사회보장 행정 (0.708) ▶사업_보건업 및 사회복지 서비스업 (0.925) ▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.036)
정확 분류율	79.7%
ROD (-) 지역 -> ROD (-) 지역 분류율	90.9%
ROD (+) 지역 -> ROD (+) 지역 분류율	46.7%

B. 가구 부문

- ROD가 (+)로 너무 크거나 (-)로 너무 작아지면 이상 시군구라고 할 수 있을 것이다. 그런데 본 연구에서 사용된 시군구 단위 자료에서는 인구의 경우와 마찬가지로 ROD가 (+)인 경우가 인구부분에서와 같이 6개(2.8%) 지역으로 적었다. 그래서 여기서는 ROD가 (+)인 경우는 제외하고, ROD가 (-)인 경우에서 이상 시군구와 그 외(정상)로 분류하고 독립변수로 사업체조사 자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법: stepwise procedure)을 적용한다.
- 이상 시군구와 그 외(정상)의 분류는 ROD를 크기 순서대로 정렬하여 하위 20%, 15%, 10%에 해당하는 시군구까지를 이상 시군구로 하고, 그 외를 정상으로 하여 모형을 적용한다. 이때 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-24> test data와 validation data

정상군 : 이상 시군구

전체	164 : 41 (80% : 20%)	175 : 30 (85.4% : 14.6%)	185 : 20 (90.2% : 9.8%)
test data	121 : 24 (83.5% : 16.7%)	130 : 15 (89.7% : 10.3%)	135 : 10 (93.1% : 6.9%)
validation data	43 : 17 (71.7% : 28.3%)	45 : 15 (75% : 25%)	50 : 10 (83.3% : 16.7%)
이상 지역	(32370)화천군 (23320)용진군 (37430)울릉군 (21120)강서구 (32400)고성군 (36480)신안군 (36450)장성군 (32380)양구군 (35320)진안군 (11020)충구 (23310)강화군 (23010)충구 (11010)종로구 (37380)성주군 (32320)횡성군 (36310)담양군 (33360)괴산군 (32360)철원군 (36430)함평군 (32350)정선군 (11030)용산구 (35350)임실군 (36040)나주시 (34380)태안군 (37310)군위군 (32390)인제군 (32330)영월군 (32340)평창군 (38370)산청군 (37390)칠곡군 (35380)부안군 (31070)평택시 (36440)영광군 (33370)음성군 (35340)장수군 (36400)해남군 (11120)은평구 (32310)홍천군 (35060)김제시 (38310)의령군 (36330)구례군		

- ROD가 상위 20%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 96.7%, 이상지역을 이상지역으로 분류할 확률은 29.2%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 88.4%, 이상지역을 이상지역으로 분류할 확률은 29.4%로 나타났다.

<표 1-25> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶사업_광업 (0.873) ▶사업_운수업 (0.998) ▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.005)	▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.005)	▶사업_금융 및 보험업 (0.976) ▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.007)
test	정확 분류율	85.5%	90.3%	93.1%
	정상->정상 분류율	96.7%	100%	100%
	이상->이상 분류율	29.2%	6.7%	0%
validation	정확 분류율	71.7%	78.3%	81.7%
	정상->정상 분류율	88.4%	100%	96%
	이상->이상 분류율	29.4%	13.3%	10%

1. 1인 가구

- 가구수의 차이의 설명할 수 있는 1인가구수의 차이만을 이용하여 이상 시군구를 판별해본다.
- ROD가 (+)로 너무 크거나 (-)로 너무 작아지면 이상 시군구라고 할 수 있을 것이다. 그런데 본 연구에서 사용된 시군구 단위 자료에서는 ROD가 (+)인 경우가 2개(대전 유성구, 경북 경산시)로 나타났다. 그래서 여기서는 ROD가 (+)인 경우는 제외하고, ROD가 (-)인 경우에서 이상시군구와 그 외(정상)로 분류하고 독립변수로 사업체조사 자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법 : stepwise procedure)을 적용한다.
- 이상 시군구와 그 외(정상)의 분류는 ROD를 크기 순 정렬하여 하위 20%, 15%, 10%에 해당하는 시군구까지를 이상 시군구로 하고, 그 외를 정상으로 하여 모형을 적용한다. 이때 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-26> test data와 validation data

정상군 : 이상 시군구

전체	168 : 41 (80.4% : 19.6%)	178 : 31 (85.2% : 14.8%)	189 : 20 (90.4% : 9.6%)
test data	125 : 24 (83.9% : 16.1%)	133 : 16 (89.3% : 10.7%)	139 : 10 (93.3% : 6.7%)
validation data	43 : 17 (71.7% : 28.3%)	43 : 17 (71.7% : 28.3%)	50 : 10 (83.3% : 16.7%)
이상 지역	(32370)화천군 (21120)강서구 (32380)양구군 (37430)울릉군 (23320)웅진군 (32360)철원군 (11120)은평구 (23010)중구 (11020)중구 (32320)횡성군 (31070)평택시 (23310)강화군 (31110)파천시 (31120)구리시 (31080)동두천시 (32400)고성군 (32390)인제군 (11150)양천구 (11030)용산구 (11010)종로구 (11240)송파구 (33370)음성군 (32350)정선군 (11190)영등포구 (37390)칠곡군 (11070)중랑구 (11090)강북구 (11220)서초구 (11170)구로구 (11100)도봉구 (23020)동구 (11250)강동구 (32310)홍천군 (23080)서구 (33360)괴산군 (11080)성북구 (31060)광명시 (31030)의정부시 (32330)영월군 (11040)성동구 (38370)산청군		

- ROD가 상위 10%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 100%, 이상지역을 이상지역으로 분류할 확률은 30%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 100%, 이상지역을 이상지역으로 분류할 확률은 0%로 나타났다.

<표 1-27> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶종사자_출판, 영상, 방송통신 및 정보서비스업 (1.000) ▶사업_전기, 가스, 증기 및 수도사업 (1.190)	▶종사자_출판, 영상, 방송통신 및 정보서비스업 (1.000) ▶사업_전기, 가스, 증기 및 수도사업 (1.397)	▶종사자_부동산업 및 임대업 (0.999) ▶사업_전기, 가스, 증기 및 수도사업 (1.361) ▶사업_건설업 (1.008)
test	정확 분류율	85.2%	91.3%	95.3%
	정상->정상 분류율	99.2%	99.2%	100%
	이상->이상 분류율	12.5%	25%	30%
validation	정확 분류율	71.7%	75%	83.3%
	정상->정상 분류율	100%	100%	100%
	이상->이상 분류율	0%	0%	0%

C. 주택 부문

- 본 연구에서 사용된 시군구 단위 자료에서는 ROD가 (+)인 경우가 41개 (19.4%) 지역으로 인구나 가구에 비해 많았다. 그러나 마찬가지로 ROD가 (+)인 경우는 제외하고, ROD가 (-)인 경우에서 이상시군구와 그 외(정상)로 분류하고 독립변수로 사업체조사 자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법 : stepwise procedure)을 적용한다.
- 이상시군구와 그 외(정상)의 분류는 ROD를 크기 순 정렬하여 하위 20%, 15%, 10%에 해당하는 시군구까지를 이상 시군구로 하고, 그 외를 정상으로 하여 모형을 적용한다. 이때 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-28> test data와 validation data

정상군 : 이상 시군구

전체	136 : 34 (80% : 20%)	145 : 25 (85.3% : 14.7%)	153 : 17 (90% : 10%)
test data	100 : 17 (85.5% : 14.5%)	105 : 12 (89.7% : 12.0%)	110 : 7 (94.0% : 6.0%)
validation data	36 : 17 (67.9% : 32.1%)	40 : 13 (75.5% : 24.5%)	43 : 10 (81.1% : 18.9%)
이상 지역	(32340)평창군 (38350)남해군 (36360)보성군 (36380)장흥군 (32350)정선군 (36480)신안군 (37380)성주군 (34390)당진군 (32410)양양군 (38310)의령군 (23320)웅진군 (38400)함천군 (35320)진안군 (37310)군위군 (36370)화순군 (23010)충구 (35330)무주군 (33380)단양군 (38360)하동군 (36040)나주시 (36450)장성군 (38380)함양군 (23310)강화군 (11020)충구 (36310)담양군 (36470)진도군 (36330)구례군 (11010)종로구 (38050)통영시 (38390)거창군 (35360)순창군 (35050)남원시 (36060)광양시 (38340)고성군		

- ROD가 상위 20%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 95%, 이상지역을 이상지역으로 분류할 확률은 41.2%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 94.4%, 이상지역을 이상지역으로 분류할 확률은 47.1%로 나타났다.

<표 1-29> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶종사자_보건업 및 사회복지 서비스업 (0.999) ▶사업_전기, 가 스, 중기 및 수도사업 (0.685) ▶사업_숙박 및 음식점업 (0.997) ▶사업_교육 서 비스업 (1.030)	▶종사자_보건업 및 사회복지 서비스업 (0.997) ▶사업_건설업 (0.982) ▶사업_교육 서 비스업 (1.055)	▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.003)
test	정확 분류율	87.2%	92.3%	94.0%
	정상->정상 분류율	95%	98.1%	100%
	이상->이상 분류율	41.2%	41.7%	0%
validation	정확 분류율	79.2%	71.7%	81.1%
	정상->정상 분류율	94.4%	95%	100%
	이상->이상 분류율	47.1%	0%	0%

- 계속해서 위에서 제외했던 지역인 ROD가 (+)인 41개 지역에 대해 알아보면 <표 1-30>과 같다. 대부분 시군군의 ROD가 (-)인데 ROD가 +인 41개의 특성을 분명히 해야 할 필요가 있다. 이에 ROD가 (+)인 지역이 ROD가 (-)인 지역과 구별할 수 있는 특성이 있는지 확인하고자 한다.
- 이를 위해 위와 마찬가지로 방법으로 로지스틱 회귀모형을 적용해 본다. 결과 ROD가 (-)인 지역을 ROD가 -인 지역으로 분류할 확률은 99.4%로 높았으나, ROD가 (+)인 지역을 (+)인 지역으로 분류할 확률은 2.4%로 낮았다.

<표 1-30> Rod가 (+)인 지역

지 역 코드	지역	지 역 코드	지역	지 역 코드	지역
33350	진천군	25020	중구	34070	계룡시
37090	문경시	32390	인제군	37330	청송군
32040	동해시	21150	사상구	34310	금산군
11110	노원구	31060	광명시	35040	정읍시
11080	성북구	32360	철원군	37350	영덕군
11130	서대문구	21020	서구	32370	화천군
36010	목포시	36460	완도군	34360	홍성군
11250	강동구	35060	김제시	34030	보령시
32050	태백시	21030	동구	36440	영광군
21120	강서구	32380	양구군	35350	임실군
34380	태안군	34060	논산시	34340	서천군
35310	완주군	35340	장수군	35370	고창군
21040	영도구	32310	홍천군	36430	함평군
37060	영주시	36350	고흥군		

<표 1-31> logistic regression 결과

selection=stepwise	
선택된 변수	▶사업_건설업 (1.006) ▶사업_운수업 (0.999)
정확 분류율	80.6%
ROD (-) 지역 -> ROD (-) 지역 분류율	99.4%
ROD (+) 지역 -> ROD (+) 지역 분류율	2.4%

1. 아파트 차이

- 가구의 차이의 설명할 수 있는 1인가구의 차이만을 이용하여 이상시군구를 판별해본다.
- ROD가 (+)로 너무 크거나 (-)로 너무 작아지면 이상 시군구라고 할 수 있을 것이다. 그런데 본 연구에서 사용된 시군구 단위 자료에서는 ROD가 (+)인 경우가 45개(21.3%) 지역으로 인구나 가구에 비해 많았다. 그러나 마찬가지로 ROD가 (+)인 경우는 제외하고, ROD가 (-)인 경우에서 이상 시군구와 그 외(정상)로 분류하고 독립변수로 사업체조사 자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법 : stepwise procedure)을 적용한다.
- 이상 시군구와 그 외(정상)의 분류는 ROD를 크기 순 정렬하여 하위 20%, 15%, 10%에 해당하는 시군구까지를 이상 시군구로 하고, 그 외를 정상으로 하여 모형을 적용한다. 이때 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-32> test data와 validation data

정상군 : 이상 시군구

전체	133 : 33 (80.1% : 19.9%)	142 : 24 (85.5% : 14.5%)	150 : 16 (90.4% : 9.6%)
test data	97 : 16 (85.8% : 14.2%)	101 : 12 (89.4% : 10.6%)	107 : 6 (94.7% : 5.3%)
validation data	36 : 17 (67.9% : 32.1%)	41 : 12 (77.4% : 22.7%)	43 : 10 (81.1% : 18.9%)
이상 지역	(32340)평창군 (32410)양양군 (35340)장수군 (34340)서천군 (35350)임실군 (32350)정선군 (21120)강서구 (36450)장성군 (32400)고성군 (36350)고흥군 (38360)하동군 (36320)곡성군 (23010)중구 (34390)당진군 (36390)강진군 (37370)고령군 (38370)산청군 (34370)예산군 (34380)태안군 (11010)종로구 (37020)경주시 (37390)칠곡군 (33370)음성군 (34040)아산시 (34320)연기군 (32070)삼척시 (35370)고창군 (32020)원주시 (37360)청도군 (33320)보은군 (32060)속초시 (31080)동두천시 (22030)서구		

- ROD가 상위 20%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 99%, 이상지역을 이상지역으로 분류할 확률은 6.3%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 100%, 이상지역을 이상지역으로 분류할 확률은 11.8%로 나타났다.

<표 1-33> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶사업_숙박 및 음식점업 (0.998) ▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.004)	▶종사자_제조업 (1.000) ▶종사자_공공 행정, 국방 및 사회보장 행정 (1.000) ▶사업_운수업 (1.003)	▶사업_협회 및 단체, 수리 및 기타 개인 서비스업 (1.004)
test	정확 분류율	85.8%	88.5%	94.7%
	정상->정상 분류율	99.0%	98.0%	100%
	이상->이상 분류율	6.3%	8.3%	0%
validation	정확 분류율	71.7%	79.2%	81.1%
	정상->정상 분류율	100%	100%	100%
	이상->이상 분류율	11.8%	8.3%	0%

4. SWROAD를 통한 이상 시군구 탐색

A. 인구 부문

- SWROAD의 크기순 정렬시 상위 20%, 15%, 10%를 이상시군구로 하고, 독립 변수로 사업제조사 자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법 : stepwise procedure)을 적용한다. 위와 마찬가지로 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-34> test data와 validation data

정상군 : 이상 시군구

전체	169 : 42 (80.1% : 19.9%)	180 : 31 (85.3% : 14.7%)	190 : 21 (90% : 10%)
test data	124 : 25 (83.2% : 16.8%)	133 : 16 (89.3% : 10.7%)	139 : 10 (93.3% : 6.7%)
validation data	45 : 17 (72.6% : 27.4%)	47 : 15 (75.8% : 24.2%)	51 : 11 (82.3% : 17.7%)
이상 지역	(11190)영등포구 (11210)관악구 (31090)안산시 (38110)창원시 (11170)구로구 (31010)수원시 (31070)평택시 (11080)성북구 (31050)부천시 (11200)동작구 (11230)강남구 (31100)고양시 (11060)동대문구 (31020)성남시 (11030)용산구 (11180)금천구 (11140)마포구 (11050)광진구 (11240)송파구 (11120)은평구 (34010)천안시 (38070)김해시 (11220)서초구 (11130)서대문구 (37100)경산시 (34040)아산시 (11070)중랑구 (23060)부평구 (11010)종로구 (11150)양천구 (23080)서구 (23050)남동구 (11160)강서구 (11250)강동구 (36020)여주시 (11040)성동구 (39020)서귀포시 (11020)충구 (21050)부산진구 (25010)동구 (37020)경주시 (23030)남구		

- 결과는 다음과 같다. 공통적으로 선택된 변수로는 부동산산업 및 임대업으로 나타났다. 부동산 사업 및 임대업이 높을수록 차이가 적은(정상) 지역이 될 가능성이 높아진다. 이는 부동산이나 임대사업이 발달한 곳은 주택지구일 가능성이 높을 것으로 생각된다. 그래서 행정자료와 조사자료의 차이가 작을 것으로 생각된다.
- SWROAD가 상위 20%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 98.4%, 이상지역을 이상지역으로 분류할 확률은 88%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 88.9%, 이상지역을 이상지역으로 분류할 확률은 76.5%로 나타났다.

<표 1-35> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶종사자_전기, 가스, 증기 및 수도사업 (0.991) ▶종사자_하수 · 폐기물 처리, 원료재생 및 환경복원업 (1.015) ▶종사자_건설업 (0.998) ▶종사자_운수업 (0.999) ▶사업_농업, 임업 및 어업 (0.578) ▶사업_제조업 (0.996) ▶사업_건설업 (1.037) ▶사업_부동산업 및 임대업 (0.976) ▶사업_교육 서비스업 (1.021) ▶사업_보건업 및 사회복지 서비스업 (0.972)	▶종사자_제조업 (1.000) ▶종사자_금융 및 보험업 (1.000) ▶사업_건설업 (1.011) ▶사업_부동산업 및 임대업 (0.989)	▶종사자_전산업 (1.000) ▶종사자_부동산업 및 임대업 (1.002) ▶사업_전산업 (1.002) ▶사업_부동산업 및 임대업 (0.972)
test	정확 분류율	96.6%	94.6%	98.0%
	정상->정상 분류율	98.4%	97.7%	99.3%
	이상->이상 분류율	88.0%	68.8%	80.0%
validation	정확 분류율	85.5%	85.5%	87.1%
	정상->정상 분류율	88.9%	91.5%	94.1%
	이상->이상 분류율	76.5%	66.7%	54.5%

B. 가구 부문

- SWROAD(가구)의 크기순 정렬시 상위 20%, 15%, 10%를 이상시군구로 하고, 독립변수로 사업제조사자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법 : stepwise procedure)을 적용한다. 위와 마찬가지로 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-36> test data와 validation data

정상군 : 이상 시군구

전체	169 : 42 (80.1% : 19.9%)	180 : 31 (85.3% : 14.7%)	190 : 21 (90% : 10%)
test data	124 : 25 (83.2% : 16.8%)	133 : 16 (89.3% : 10.7%)	139 : 10 (93.3% : 6.7%)
validation data	45 : 17 (72.6% : 27.4%)	47 : 15 (75.8% : 24.2%)	51 : 11 (82.3% : 17.7%)
이상 지역	(38110)창원시 (31020)성남시 (31010)수원시 (11210)관악구 (31090)안산시 (31100)고양시 (31050)부천시 (11140)마포구 (39010)제주시 (11240)송파구 (11190)영등포구 (34010)천안시 (11230)강남구 (11050)광진구 (11080)성북구 (11060)동대문구 (11070)종로구 (11030)용산구 (35010)전주시 (37010)포항시 (31070)평택시 (11120)은평구 (11160)강서구 (37050)구미시 (11170)구로구 (11200)동작구 (23060)부평구 (11130)서대문구 (23030)남구 (11250)강동구 (11040)성동구 (31040)안양시 (11090)강북구 (34040)아산시 (21050)부산진구 (33010)청주시 (11110)노원구 (11220)서초구 (31030)의정부시 (24040)북구 (11010)종로구 (22070)달서구		

- 결과는 다음과 같다. 공통적으로 선택된 변수로는 부동산산업 및 임대업으로 나타났다. 부동산 사업 및 임대업이 높을수록 차이가 적은(정상) 지역이 될 가능성이 높아진다. 이는 부동산이나 임대사업이 발달한 곳은 주택지구일 가능성이 높을 것으로 생각된다. 그래서 행정자료와 조사자료의 차이가 작을 것으로 생각된다.
- SWROAD가 상위 20%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 97.6%, 이상지역을 이상지역으로 분류할 확률은 84%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 93.3%, 이상지역을 이상지역으로 분류할 확률은 94.1%로 나타났다.

<표 1-37> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶종사자_부동산 업 및 임대업 (1.001) ▶종사자_사업시 설관리 및 사 업지원 서비 스업 (0.999) ▶종사자_예술, 스포츠 및 여 가관련 서비 스업 (0.998) ▶사업_부동산업 및 임대업 (0.989) ▶사업_전문, 과 학 및 기술 서비스업 (1.011)	▶종사자_제조업 (1.000) ▶종사자_부동산 업 및 임대업 (1.004) ▶종사자_협회 및 단체, 수리 및 기타 개인 서 비스업 (0.997) ▶사업_건설업 (1.010) ▶사업_부동산업 및 임대업 (0.974) ▶사업_보건업 및 사회복지 서 비스업 (1.014) ▶사업_농업, 임 업 및 어업 (0.946)	▶종사자_예술, 스포츠 및 여 가관련 서비 스업 (0.999) ▶사업_운수업 (0.999)
test	정확 분류율	95.3%	96.6%	93.3%
	정상->정상 분류율	97.6%	98.5%	97.8%
	이상->이상 분류율	84.0%	81.3%	30%
validataio n	정확 분류율	93.5%	90.3%	91.9%
	정상->정상 분류율	93.3%	91.5%	96.1%
	이상->이상 분류율	94.1%	86.7%	72.7%

C. 주택 부분

- SWROAD(주택)의 크기순 정렬시 상위 20%, 15%, 10%를 이상 시군구로 하고, 독립변수로 사업제조사자료를 이용하여 로지스틱 회귀모형(logistic regression model)(방법 : stepwise procedure)을 적용한다. 위와 마찬가지로 산출된 로지스틱 회귀모형(logistic regression model)의 적절성을 평가하고자 validation 자료를 30%를 생성하여 분석을 실시하였다.

<표 1-38> test data와 validation data

정상군 : 이상 시군구

전체	169 : 42 (80.1% : 19.9%)	180 : 31 (85.3% : 14.7%)	190 : 21 (90% : 10%)
test data	124 : 25 (83.2% : 16.8%)	133 : 16 (89.3% : 10.7%)	139 : 10 (93.3% : 6.7%)
validation data	45 : 17 (72.6% : 27.4%)	47 : 15 (75.8% : 24.2%)	51 : 11 (82.3% : 17.7%)
이상 지역	(31100)고양시 (31050)부천시 (31020)성남시 (11190)영등포구 (11230)강남구 (31090)안산시 (11010)종로구 (11020)중구 (39010)제주시 (11170)구로구 (31010)수원시 (35370)고창군 (11220)서초구 (34340)서천군 (38110)창원시 (11140)마포구 (21090)해운대구 (23030)남구 (34030)보령시 (11180)금천구 (36430)함평군 (11210)관악구 (11120)은평구 (34010)철안시 (21050)부산진구 (23050)남동구 (11030)용산구 (38350)남해군 (34360)홍성군 (23010)충구 (31040)안양시 (11060)동대문구 (11160)강서구 (23070)계양구 (36440)영광군 (35040)정읍시 (23080)서구 (37010)포항시 (36010)목포시 (36360)보성군 (36350)고흥군 (34390)당진군		

- 결과는 다음과 같다. 공통적으로 선택된 변수로는 공공행정 국방 및 사회보장 행정으로 나타났다. 공공행정 국방 및 사회보장 행정 사업체수가 많을수록 차이가 적은(정상) 지역이 될 가능성이 높아진다.
- SWROAD가 상위 15%를 이상지역으로 한 경우의 분류율이 가장 높았다. test 자료에서 정상지역을 정상지역으로 분류할 확률은 98.5%, 이상지역을 이상지역으로 분류할 확률은 37.5%로 나타났다. validation 자료에서 정상지역을 정상지역으로 분류할 확률은 100%, 이상지역을 이상지역으로 분류할 확률은 53.3%로 나타났다.

<표 1-39> logistic regression 결과

selection=stepwise slentry=0.1, slstay=0.15		20%	15%	10%
선택된 변수		▶사업_출판, 영상, 방송통신 및 정보서비스업 (0.996) ▶사업_공공행정, 국방 및 사회보장 행정 (0.969)	▶사업_출판, 영상, 방송통신 및 정보서비스업 (0.994) ▶사업_공공행정, 국방 및 사회보장 행정 (0.974)	▶종사자_하수 폐기물 처리, 원료재생 및 환경복원업 (1.005) ▶종사자_출판, 영상, 방송통신 및 정보서비스업 (1.000) ▶사업_공공행정, 국방 및 사회보장 행정 (0.936)
test	정확 분류율	86.6%	91.9%	96.6%
	정상->정상 분류율	98.4%	98.5%	100%
	이상->이상 분류율	28%	37.5%	50%
validation	정확 분류율	82.3%	88.7%	93.5%
	정상->정상 분류율	95.6%	100%	100%
	이상->이상 분류율	47.1%	53.3%	63.6%

5. 2011년 연구의 이상읍면동과 2012년 연구의 이상 시군구 종합

- 2011년 연구에서 이상 읍면동을 탐색한바 있다. 탐색된 이상 읍면동이 2012년 연구에서 탐색된 이상 시군구에 속하는지 알아본다. 먼저 2012년 연구에서의 이상시군구를 RoD측면과 SWROAD측면으로 정리하면 다음 <표 1-40>과 <표 1-41>과 같다.
- <표 1-40>과 <표 1-41>에서는 RoD측면과 SWROAD측면은 인구, 가구, 주택 측면 각각에서 상위 20%에 해당하는 시군구를 체크하였다. 인구, 가구, 주택 측면에서 공통적으로 이상 시군구로 나타나는 지역이 ROD 측면에서는 14개 지역, SWROAD측면에서는 20개 지역으로 나타났다.

<표 1-40> 이상지역(20%) 종합-ROD부분

지역	인구	가구	1인가구	주택	아파트	지역	인구	가구	1인가구	주택	아파트
(11020)충구	v	v	v	v		(32370)화천군		v	v		
(23010)충구	v	v	v	v	v	(32380)양구군		v	v		
(23310)강화군	v	v	v	v		(32390)인제군		v	v		
(23320)용진군	v	v	v	v		(32400)고성군		v	v		v
(32340)평창군	v	v		v	v	(35060)김제시		v			
(35320)진안군	v	v		v		(37390)칠곡군		v	v		v
(36040)나주시	v	v		v		(38370)상정군		v	v		v
(36310)담양군	v	v		v		(32410)양양군				v	v
(36330)구례군	v	v		v		(33380)단양군				v	
(36450)장성군	v	v		v	v	(34390)당진군				v	v
(36480)신안군	v	v		v		(35050)남원시				v	
(37310)군위군	v	v		v		(36060)광양시				v	
(37380)성주군	v	v		v		(36370)화순군				v	
(38310)의령군	v	v		v		(38050)통영시				v	
(21120)강서구	v	v	v		v	(38340)고성군				v	
(32320)횡성군	v	v	v			(38350)남해군				v	
(32330)영월군	v	v	v			(38380)함양군				v	
(33360)괴산군	v	v	v			(38390)거창군				v	
(33370)음성군	v	v	v		v	(11040)성동구			v		
(34380)태안군	v	v			v	(11070)중랑구			v		
(35340)장수군	v	v			v	(11080)성북구			v		
(35350)임실군	v	v			v	(11090)강북구			v		
(35380)부안군	v	v				(11100)도봉구			v		
(36400)해남군	v	v				(11150)양천구			v		
(36430)함평군	v	v				(11170)구로구			v		
(36440)영광군	v	v				(11190)영등포구			v		
(37430)울릉군	v	v	v			(11220)서초구			v		
(35330)무주군	v			v		(11240)송파구			v		
(35360)순창군	v			v		(11250)강동구			v		
(36360)보성군	v			v		(22030)서구					v
(36380)강흥군	v			v		(23020)동구			v		
(36470)진도군	v			v		(23080)서구			v		
(38360)하동군	v			v	v	(31030)외경부시			v		
(38400)함천군	v			v		(31060)광명시			v		
(36320)곡성군	v			v		(31080)동두천시			v		v
(36350)고흥군	v			v		(31110)파천시			v		
(36390)강진군	v			v		(31120)구리시			v		
(36410)영암군	v					(32020)원주시					v
(36460)완도군	v					(32060)속초시					v
(37360)청도군	v				v	(32070)삼척시					v
(38320)함안군	v					(33320)보은군					v
(39020)서귀포시	v					(34040)아산시					v
(11010)종로구		v	v	v	v	(34320)연기군					v
(32350)경선군		v	v	v	v	(34340)서천군					v
(11030)용산구		v	v			(34370)예산군					v
(11120)은평구		v	v			(35370)고창군					v
(31070)평택시		v	v			(37020)경주시					v
(32310)홍천군		v	v			(37370)고령군					v
(32360)철원군		v	v								

<표 1-41> 이상지역(20%) 종합-SWROAD부분

지역	인구	가구	주택	지역	인구	가구	주택
(11010)충로구	v	v	v	(23080)서구	v		v
(11030)용산구	v	v	v	(11150)양천구	v		
(11060)동대문구	v	v	v	(25010)동구	v		
(11120)은평구	v	v	v	(36020)여주시	v		
(11140)마포구	v	v	v	(37020)경주시	v		
(11160)강서구	v	v	v	(37100)경산시	v		
(11170)구로구	v	v	v	(38070)김해시	v		
(11190)영등포구	v	v	v	(39020)서귀포시	v		
(11210)관악구	v	v	v	(31040)안양시		v	v
(11220)서초구	v	v	v	(37010)포항시		v	v
(11230)강남구	v	v	v	(39010)제주시		v	v
(21050)부산진구	v	v	v	(11090)강북구		v	
(23030)남구	v	v	v	(11110)노원구		v	
(31010)수원시	v	v	v	(22070)달서구		v	
(31020)성남시	v	v	v	(24040)북구		v	
(31050)부천시	v	v	v	(31030)의정부시		v	
(31090)안산시	v	v	v	(33010)청주시		v	
(31100)고양시	v	v	v	(35010)전주시		v	
(34010)천안시	v	v	v	(37050)구미시		v	
(38110)창원시	v	v	v	(21090)해운대구			v
(11040)성동구	v	v		(23010)중구			v
(11050)광진구	v	v		(23070)계양구			v
(11070)중랑구	v	v		(34030)보령시			v
(11080)성북구	v	v		(34340)서천군			v
(11130)서대문구	v	v		(34360)홍성군			v
(11200)동작구	v	v		(34390)당진군			v
(11240)송파구	v	v		(35040)정읍시			v
(11250)강동구	v	v		(35370)고창군			v
(23060)부평구	v	v		(36010)목포시			v
(31070)평택시	v	v		(36350)고흥군			v
(34040)아산시	v	v		(36360)보성군			v
(11020)중구	v		v	(36430)함평군			v
(11180)금천구	v		v	(36440)영광군			v
(23050)남동구	v		v	(38350)남해군			v

- **이상 시군구와 이상 읍면동 결합** : 본 연구에서 사용한 211개 시군구 중 이상 시군구와 2011년도 이상 읍면동의 관련성을 파악한 결과가 다음과 같다. ROD측면으로 이상 시군구를 탐색한 결과가 SWROAD 측면으로 탐색한 결과 보다 이상 읍면동을 포함하는 정도가 높게 나타났다. 이것은 2011년도 연구에서 사용한 동일한 척도를 사용했기 때문으로 생각된다.
- 또한 2011년도 연구에서 이상 읍면동 탐색방법 중 “정상여부01”이 RoD측면으로 이상 시군구를 탐색한 결과와 일치하는 정도가 39.1%로 가장 높게 나타났다으나 일치정도는 높지 않은 것으로 나타났다. 이것은 시군구 단위의 차이와 읍면동 단위의 차이를 동시에 고려하는 어려움이 있다는 것을 의미한다. 향후 시군구 단위의 차이와 읍면동 단위의 차이를 동시에 고려할 수 있는 연구가 필요하다.

<표 1-42> 이상시군구와 이상읍면동

			정상여부01		정상여부02		정상여부03		총합
			정상	이상	정상	이상	정상	이상	
ROD	인구	정상	2,018	357	1,974	401	2,223	152	2,375
		이상	180	274	277	177	413	41	454
	가구	정상	1,980	411	1,995	396	2,235	156	2,391
		이상	218	220	256	182	401	37	438
	주택	정상	2,007	421	1,988	440	2,269	159	2,428
		이상	191	210	263	138	367	34	401
SWROAD	인구	정상	1,631	562	1,811	382	2,035	158	2,193
		이상	567	69	440	196	601	35	636
	가구	정상	1,652	585	1,827	410	2,075	162	2,237
		이상	546	46	424	168	561	31	592
	주택	정상	1,751	543	1,855	439	2,134	160	2,294
		이상	447	88	396	139	502	33	535
총합			2,198	631	2,251	578	2,636	193	2,829

- <참고자료> 2011년도 연구에서 이상 읍면동을 판별하는 방안은 3가지로 제안하였다. 첫째, 전국읍면동을 Rod를 기준으로 전체 읍면동의 20%에 해당하는 읍면동을 이상 읍면동으로 판별하였다. 둘째, 행정인구 크기를 기준으로 3개 군집을 구축하여 군집별로 Rod 및 D를 기준으로 군집 내 읍면동의 20%에 해당하는 읍면동을 이상 읍면동으로 판별하였다. 셋째, 읍면동의 개수가 5개를 초과하는 시군구의 RLAD가 최대값을 갖는 경우 이상 읍면동으로 판별하였다. 이 경우를 종합하여 이상 읍면동으로 판별하는 기준으로 하여 이상 읍면동을 제외하고 조사자료와 행정자료의 차이를 비교해본 결과가 다음 참고자료와 같다.

<참고> 이상동 기준 적용 결과분석

	전국		전국			시군구			읍면동			
	조사자료	행정자료	n	AD	설명력	n	AD	설명력	n	AD	설명력	
전체	47,990,761	50,469,907	3472	2,479,146		251	2,656,718		3,472	3,285,978		
정상여부	정상군	47,414,368	49,966,446	3397	2,565,594	103.5	251	2,640,574	99.4	3,397	3,028,046	92.2
	이상군	576,393	503,461	75	86,448	3.5	75	257,932	9.7	75	257,932	7.8
정상여부01	정상군	44,596,120	46,747,659	2771	2,151,539	86.8	247	2,199,441	82.8	2,771	2,482,593	75.6
	이상군	3,394,641	3,722,248	701	327,607	13.2	166	693,885	26.1	701	803,385	24.4
정상여부02	정상군	39,237,774	40,942,661	2777	1,704,887	68.8	250	1,742,103	65.6	2,777	2,007,721	61.1
	이상군	8,752,987	9,527,246	695	774,259	31.2	194	1,111,779	41.8	695	1,278,257	38.9
정상여부03	정상군	42,665,753	44,966,330	3228	2,300,577	92.8	251	2,354,963	88.6	3,228	2,672,815	81.3
	이상군	5,325,008	5,503,577	244	178,569	7.2	244	613,163	23.1	244	613,163	18.7
정상여부12	정상군	45,534,906	47,832,522	3079	2,297,616	92.7	250	2,351,666	88.5	3,079	2,680,474	81.6
	이상군	2,455,855	2,637,385	393	181,530	7.3	135	553,698	20.8	393	605,504	18.4
정상여부13	정상군	46,749,892	49,317,427	3389	2,567,535	103.6	251	2,637,129	99.3	3,389	3,019,301	91.9
	이상군	1,240,869	1,152,480	83	88,389	3.6	83	266,677	10.0	83	266,677	8.1
정상여부23	정상군	45,003,415	47,396,915	3319	2,393,500	96.5	251	2,468,480	92.9	3,319	2,822,400	85.9
	이상군	2,987,346	3,072,992	153	85,646	3.5	153	463,578	17.4	153	463,578	14.1
정상여부2or3	정상군	36,900,112	38,512,076	2686	1,611,964	65.0	250	1,629,648	61.3	2,686	1,858,136	56.5
	이상군	11,090,649	11,957,831	786	867,182	35.0	247	1,245,438	46.9	786	1,427,842	43.5
정상여부1or3	정상군	40,511,981	42,396,562	2610	1,884,581	76.0	246	1,917,275	72.2	2,610	2,136,107	65.0
	이상군	7,478,780	8,073,345	862	594,565	24.0	247	1,020,091	38.4	862	1,149,871	35.0

정상여부 : 첫째, 둘째, 셋째 방안이 모두 만족 정상여부01 : 첫째 방안 만족(정상군은 Rod -0.2~+0.1)
 정상여부02 : 둘째 방안 만족 (1군집은 Rod가 -0.26~0.30 & D -1300~+1500, 2군집 : Rod -0.09~+0.05, 3군집 : Rod -0.07~+0.09)
 정상여부03 : 셋째 방안 만족 정상여부12 : 첫째, 둘째 방안이 만족
 정상여부13 : 첫째, 셋째 방안이 만족 정상여부23 : 둘째, 셋째 방안이 만족
 정상여부2or3 : 둘째 또는 셋째 방안이 만족 정상여부1or3 : 첫째 또는 셋째 방안이 만족

제3절 결과의 요약

- 제1장의 연구에서는 2011년 연구에서 이상 읍면동을 탐색하는 방안을 검토한 데 반하여 이상 시군구를 탐색하는 방안을 연구하였다. 이상 읍면동을 탐색하는 방안은 조사자료와 행정자료의 차이를 어떻게 정의하느냐에 따라서 이상 읍면동이 다르게 선정되는 문제를 안고 있었다. 이런 문제는 3500여개의 읍면동의 38개의 성별, 연령별 인구수, 가구수, 주택수를 모두 고려하면서는 해결할 수 없다고 판단하였다. 또한 읍면동의 조사구 단위에 이루어지는 표본설계 역시 다수의 읍면동에서 조사구의 동질성이 확보되지 않는다는 점에서 문제가 있는 것으로 판단하였다. 이 때문에, 올해의 연구는 이상읍면동보다 이상 시군구를 파악하여 정밀한 표본조사가 필요한 시군구의 리스트를 제공하는 것을 목표로 설정하였다.
- 이 장에서는 시군구의 인구수, 가구수, 주택수 등을 모두 고려하면서 전체 및 공표대상 범주별 조사자료와 행정자료 간에 존재하는 상대적 차이와 절대적 차이가 종합적으로 큰 시군구를 탐색하고 이들 시군구를 판별하는데 유용한 설명변수를 가용한 행정자료 내지는 참조자료 내에서 선택/제안하는 것이다.
- 제1장의 연구는 4가지 부분으로 이루어져 있다. 첫째로 2011년 연구에서 제안한 행정자료와 조사자료의 차이를 나타내는 척도 RoD(상대차이)와 이 척도의 한계를 극복하기 위하여 시군구 규모 및 성별, 연령별 차이를 함께 고려하여 새롭게 제안한 SWROAD라는 척도를 이용하여 이상 시군구를 탐색하였다. 이에 2011년 제안한 척도를 이용하여 이상 시군구를 탐색하고, 새로운 척도를 개발하여 차이의 정도(대, 중, 소)에 따라 12개 시군구를 선정하였다. 둘째는 가용자료로부터 이상 시군구를 판별하는 방안을 연구하기 위하여 가용자료와 연계시킨 시군구 자료를 탐색하였다. 이어서 로지스틱 회귀모형을 적용하여 인구, 가구, 주택부분의 이상 시군구를 탐색하였는데, 기존의 척도인 RoD와 새로운 척도인 SWROAD 각각에서 탐색하였다. 마지막으로 각 척도별 범주별(인구, 가구, 주택) 이상 시군구를 정리하고, 2011년 연구의 이상 읍면동과 연계시켜 연계정도를 파악하였다.

[부록 1-1] 우리나라의 행정구역

- 우리나라의 행정 구역은 1개의 특별시, 6개의 광역시, 8개의 도, 1개의 특별자치시, 1개의 특별자치도로 구성된다. 이상 총 17개의 행정구역은 광역자치단체로 분류된다.

<부록 표 1-1> 우리나라의 시도별 시군구 개수
(구의 괄호 안은 일반구 개수)

시도	시	군	구
합계	74	85	69
서울특별시	-	-	25
부산광역시	-	1	15
대구광역시	-	1	7
인천광역시	-	2	8
광주광역시	-	-	5
대전광역시	-	-	5
울산광역시	-	1	4
새종특별자치시	1	-	-
경기도	27	4	(20)
강원도	7	11	-
충청북도	3	9	(2)
충청남도	8	7	(2)
전라북도	6	8	(2)
전라남도	5	17	-
경상북도	10	13	(2)
경상남도	8	10	(5)
제주특별자치도	(2)	-	

- 특별시는 자치구로, 광역시는 자치구와 군으로, 도는 자치시와 군으로 하위 행

정구역을 둔다. 이상의 행정구역은 기초자치단체로 분류된다. 특별자치시는 그 자체로 하나의 기초자치단체를 구성한다. 특별자치도는 자치시가 아닌 행정시를 둘 수 있으며, 행정시는 특별자치도지사 직속으로 그 역할을 하며 기초지방자치단체로서의 권한이 없다. 2011년 1월 기준으로 특별시와 6개 광역시에는 총 69개의 자치구가 설치되어 있다. 또한 6개 광역시와 8개 도에는 총 74개의 자치시와 85개의 군이 설치되어 있다.

<부록 표 1-2> 우리나라의 광역/기초 자치단체와 비자치구역

광역자치단체	기초자치단체	비자치구역			
특별시	자치구		행정동	통	반
광역시	자치구			읍	
	군	면			
도	자치시	(일반구)	읍·면	통	
			행정동		
특별자치도		행정시	행 정 동 ·	통 · 행 정	
특별자치시			읍·면	리	

- 특별·광역시를 제외한 인구 50만명 이상의 시에는 일반구(또는 행정구)를 둘 수 있다. 일반구구는 기초지방자치단체로서의 권한이 없어, 특별·광역시에 설치된 자치구와 구별된다. 일반구의 구청장은 일반직 지방공무원으로 보(補)하고 시장이 임명한다. 구청장은 시장의 지휘·감독을 받아 소관 국가사무 및 지방자치사무를 맡아 처리하고 소속 직원을 지휘·감독한다.
- 시(자치시, 행정시)와 구(자치구, 일반구)는 읍·면·동으로, 군은 읍·면으로 하위 행정구역을 둔다. 다시 읍·면은 행정리로, 동은 통으로 나뉜다. 통 및 행정리는 말단 행정 구역인 반으로 나뉜다.

제2장 등록센서스를 위한 표본설계 방안 연구

제1절 서론

1. 연구목적

- 등록센서스는 기본적으로, 현지실사에 의하여 실시되는 기존의 인구주택총조사 자료를 인구, 가구, 주택에 대한 전수자료인 행정자료로 대체하려는 아이디어에서 출발한다. 그런데 우리나라의 행정자료와 인구주택총조사 자료를 비교해보면 두 자료 사이에 불일치하는 부분이 존재한다는 것이 드러난다. 이러한 차이를 보완하기 위해 과거 총조사와 같은 방법의 조사를 실시하되 모집단 전체를 대상으로 하는 전수조사 대신에 표본조사를 실시하고자 한다. 독일, 폴란드, 이스라엘, 대만, 싱가포르 등 최근 등록센서스를 도입한 여러 나라에서, 행정자료와의 불일치 정도 또는 완전성(completeness) 곧 포괄범위를 점검하고, 행정자료에서 얻을 수 없는 센서스 조사항목에 대한 자료를 얻기 위하여 표본조사를 병행하여 실시한 바 있다.
- 2010년 인구주택총조사의 결과를 검토해보면, 행정자료와 조사자료 간 차이의 양상이 지역별로 서로 다른 것을 알 수 있다. 대도시 지역의 경우, 행정자료와 조사자료 간에 통계치 차이가 그리 크지 않은 편이나 농촌 지역의 경우, 상대적으로 그 차이가 큰 편이다. 따라서 등록센서스 실시과정에서, 표본조사를 통해 행정자료를 보완하려면 지역적 특성을 감안하는 표본설계(sample design)는 필수적 절차라고 할 수 있다.
- 통계청의 등록센서스 학술연구 용역사업에서 수행된 2010년과 2011년 연구결과에 의하면 (한국인구학회, 2010, 2011), 행정자료 보완을 위한 표본설계의

추출틀(sampling frame)로는, 주민행정자료, 가족관계등록부, 건축물대장 등을 기본으로 하고 다른 관련 행정자료를 참고하여 보완한 통합행정파일이 적절한 것으로 나와 있다. 또한, 1차 추출단위(primary sampling unit, PSU)로는 통계청에서 전통적 센서스이기는 하지만, 센서스 통계의 작성이라는 목적을 염두에 두고 구축한 조사구(ED: Enumeration District)를 이용하는 것이 바람직한 것으로 추천되었다.

- 본 연구의 목적은 제1장의 연구결과와 통계청 등록센서스 주무부서가 제안한 12개 시군구를 연구대상 지역으로 정하여 등록센서스와 병행하여 실시할 표본조사를 위한 표본을 설계하는 것이다. 통계청은 2011년 등록센서스 학술연구 용역사업에서, 이미 2개 시범지역을 대상으로 이와 같은 연구를 수행한 바 있다(한국인구학회, 2011). 본 연구는 지난 연구를 기초로 하면서, 대상지역을 통계청 등록센서스 담당 주무부서가 지정한 12개 대상지역으로 확대하고, 종전에 검토되었던 인구수와 가구수 외에 주택수까지 추가하여 검토하게 된다. 12개 대상지역들은 2010년 기준 행정자료와 조사자료를 비교할 때, 그 차이가 상대적으로 큰 지역, 중간인 지역, 그리고 작은 지역들이 골고루 포함되어 있다.
- 본 연구는 구체적으로, (1) 인구수를 중심으로 표본설계 방안, (2) 인구수와 가구수를 동시에 고려하는 표본설계 방안, (3) 인구수, 가구수, 주택수 모두를 고려하는 표본설계 방안을 각각 마련하여, 3 가지 방안 중에서 어느 것이 가장 효과적인가를 검토하고자 한다. 본 연구를 통해 얻어지는 표본조사 자료는 이 보고서의 제3장의 능형회귀추정(Ridge regression estimation)과 제4장의 소지역추정(small-area estimation) 결과를 검토하는데 필요한 토대자료의 역할을 하게 될 것이다.

2. 연구개요

- 본 장의 표본설계 연구는 (1) 추출틀 구축 (2) 12개 대상지역의 2010년 행정자료와 조사자료의 비교 (3) 표본설계 방안별 결과 비교의 순서로 이루어진다. 본 연구의 단계별 절차(step-by-step procedure)을 플로차트(flow chart)로 구성한 것이 <그림 2-1>이다.

<그림 2-1> 표본설계 연구의 단계별 흐름도



- 추출틀(sampling frame)은 주민행정자료, 가족관계등록부, 건축물대장 등 행정파일을 통합, 수정하여 작성되는데, 본 연구에서는 제1장의 연구결과를 검토하여 통계청의 등록센서스 담당 주무부서가 추천한 12개 대상지역에 대해 조사구별 ID, 조사구 특성, 인구수, 가구수, 주택수 등의 변수들이 수록된 파일이 된다. 추출틀을 구축하여 조사구의 통계적 특성을 파악하게 된다.
- 본 연구는 12개 대상지역별 행정자료와 조사자료에 기초한 통계치의 차이를 시군구 및 읍면동 단위로 비교 검토한다. 검토결과는 표본설계의 방향을 정하는 기준이 된다.

- 먼저 12개 시군구를 각각 부모집단(副母集團, subpopulation)으로 간주하고, 시군구별 조사구 추출률을 20%로 할 경우의 추출오차(sampling error)를 계산한다. 이를 기초로 하여 각 시군구별로 적절한 표본크기(sample size)를 결정하게 된다.
- 시군구별 표본조사구 추출방안으로는 계통추출법(系統抽出法, systematic sampling)을 도입한다. 계통추출에서는 추출단위의 정렬 순서에 따라 효율이 달라진다. 조사구별 (1) 인구수 하나만을 기준으로 정렬하는 방안, (2) 인구수와 주택수를 동시에 고려하여 정렬하는 방안, (3) 인구, 가구수, 주택수 3가지 기준을 모두 고려하여 정렬하는 방안을 비교한 후 적절한 정렬방안을 결정한다.
- 본 연구는 최종 결정된 정렬방안에 따라 표본을 추출한 후, 지역별로 표본에 의한 추정치(sample estimate)와 모수(population parameter)를 비교함으로써 추정의 정확성을 평가한다.

제2절 추출틀

1. 추출틀 구축

- 기본적으로 표본설계를 위한 추출틀(sampling frame)은 인구, 가구, 주택과 관련되는 행정자료 파일들을 사용하여 만들어진다. 통계청의 조사자료에는 조사구 코드(ED code)가 있으나 주민행정자료나 주택 관련 데이터파일에는 조사구 코드가 없는 상태이다. 따라서 통계청에서는 주소코드를 이용하여 행정자료에 조사구 코드를 부여하는 매칭 작업을 실시하였다.
- 최종 작성한 추출틀 파일은 <그림 2-2>과 같은 형태를 띤다. 조사구가 하나의 레코드 단위이며, 조사구별 변수로는 ID, 조사구 특성, 행정자료와 조사자료에서 구한 인구수, 가구수, 주택수가 포함되도록 하였다.
- 행정자료와 조사자료 간의 코딩 상의 차이로 인해 행정자료에 조사구 코드를 부여하는 작업이 순조롭지 않았다. 본 연구가 진행되는 동안 추출틀로 사용될 자료 파일을 통합하는 과정에서 작은 에러들이 발견되어 추출틀을 구축하여 연구하는 작업을 몇 차례 반복해야 하는 시행착오가 있었다.

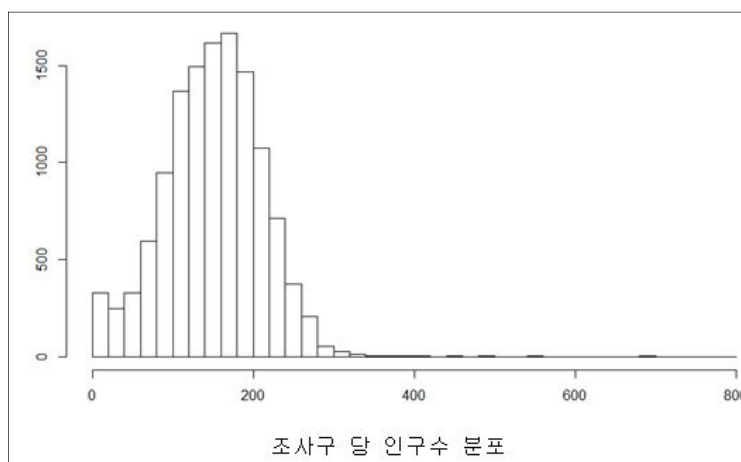
<그림 2-2> 추출틀 일부

지역	gu_code	동	조사구	특성	센서스인구	행정인구	센서스가구	행정가구	센서스주택	행정주택
김천구	11180	51	000	9	225					
김천구	11180	51	001	1	147	141	86	79	39	47
김천구	11180	51	002	4	15	13			1	1
김천구	11180	51	003	1	74	77	45	42	13	12
김천구	11180	51	004	1	109	127	51	67	19	19
김천구	11180	51	005	1	150	172	56	68	37	35
김천구	11180	51	006	1	118	115	54	56	19	19
김천구	11180	51	007	1	132	139	87	76	17	19
김천구	11180	51	008	1	140	152	106	80	13	13
김천구	11180	51	009	1	160	140	109	94	41	16
김천구	11180	51	010	1	129	160	85	121	12	12
김천구	11180	51	011	1	386	298	330	182	20	22
김천구	11180	51	012	A	239	244	70	69	70	70
김천구	11180	51	013	A	233	233	70	66	70	70
김천구	11180	51	014	A	208	211	70	68	70	70
김천구	11180	51	015	1	128	119	94	87	17	16
김천구	11180	51	016	1	172	158	143	119	10	17

2. 기초분석

- 12개 지역 전체 조사구 당 인구수의 분포가 <그림 2-3>에 제시되어 있다. 이 그림을 보면, 인구가 60명에서 260명 사이인 조사구가 대부분을 차지한다. 그러나 인구가 20명도 안 되는 조사구, 300명이 넘는 조사구 또한 존재함을 확인할 수 있다. 다시 말하자면 조사구 크기(ED size)가 균일하지 않고 다소 편차가 존재하는 편이다.

<그림 2-3> 조사구 당 인구 분포



- 이번에는 지역별 조사구 당 인구수를 살펴보자. <표 2-1>은 12개 지역별 모집단 조사구수, 조사구당 인구수의 평균과 표준편차를 나타낸 것이다. 모집단 조사구수가 가장 많은 지역은 화성시로 2,869개인 반면, 가장 적은 지역은 울릉군으로 67개에 불과하였다. 조사구 당 평균 인구수는 신안군이 92.7명으로 가장 적었고, 인천 서구가 173.2명으로 가장 많았다. 지역별 표준편차도 33.3명에서 74.4명으로 나타나 지역별 조사구 인구수의 편차가 서로 다른 것을 알

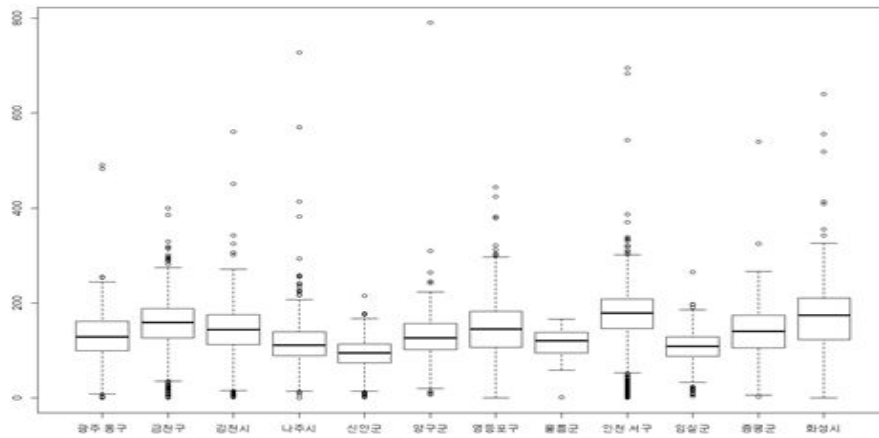
수 있었다.

<표 2-1> 지역별 조사구수, 조사구 당 인구수의 평균과 표준편차

시군구	조사구 수	평균인구	표준편차
광주 동구	799	130.2	50.6
금천구	1,492	154.5	54.2
김천시	890	142.5	54.4
나주시	660	117.9	54.2
신안군	356	92.7	33.3
양구군	146	131.9	74.4
영등포구	2,595	144.3	56.2
울릉군	67	115.5	33.4
인천 서구	2,246	173.2	61.1
임실군	219	107.3	40.0
증평군	226	138.1	60.6
화성시	2,869	165.3	65.9

- <그림 2-4>는 조사구 당 인구수의 지역별 분포를 나타내는 상자그림(box plot)이다. 지역별 상자그림을 보면, 모든 지역마다 이상치(outlier)들이 존재함을 알 수 있다. 이상치는 다시 인구가 지나치게 많은 조사구와 인구가 지나치게 적은 조사구로 구분된다. 등록센서스를 위한 이전의 연구(한국인구학회, 2011)에서는 조사구를 표본설계를 위한 1차 추출단위(PSU: primary sampling unit)로 사용하도록 추천한 바 있다. 조사구를 PSU로 사용하는 것을 전제한다면, 현재와 같이 많은 이상치 조사구들을 그대로 둔 채 표본설계를 하는 것은 효율적이지 못하다. 따라서 본격적인 표본설계를 시작하기 전에 먼저 이상치를 면밀히 검토하는 것이 바람직하다.

<그림 2-4> 지역별 조사구 당 인구 분포



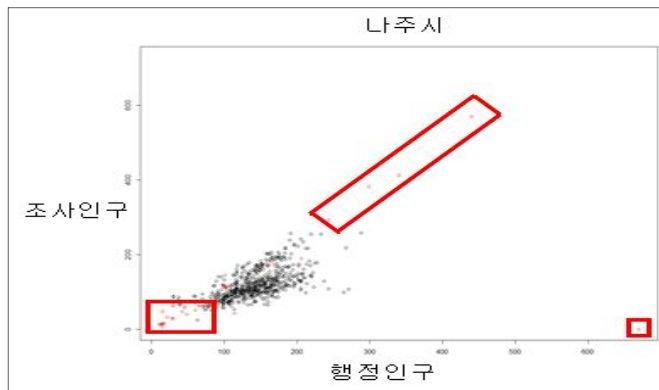
- 앞의 <그림 2-2>에 소개한 바 있는 추출틀의 변수 중에는 조사구의 특성을 나타내는 변수가 포함되어 있다. 따라서 조사구 특성별 조사구수와 이상치수를 정리하여 다음의 <표 2-2>에 나타내었다. 조사구 중 가장 수가 많은 일반 조사구와 아파트조사구, 섬조사구의 경우 이상치 비율이 낮으나, 나머지 특이인 기숙시설, 특수사회시설, 특별조사구에서는 이상치 비율이 훨씬 높다.

<표 2-2> 조사구 특성별 조사구수와 이상치수

조사구 특성	조사구수	이상치수	이상치 비율 (%)
일반	6,523	113	1.7
섬	325	1	0.3
기숙시설	215	80	37.2
특수사회시설	241	101	41.9
관광호텔 및 외국인	2	1	50.0
특별조사구	158	56	35.4
아파트	5,138	67	1.3

- 2010년의 조사구별 행정(등록)인구와 조사인구에 대한 지역별 산점도(散点圖 scatter plot)를 살펴보기로 하자. <그림 2-5>은 12개 지역 중 나주시의 산점도이다. 그림에서 사각형 표시 안에 있는 점들은 기숙시설, 특수사회시설, 특별조사구 등에 해당되는데, 이들은 대부분 이상치로 나타나고 있다.

<그림 2-5> 나주시의 행정인구와 조사인구 산점도



- 실제 등록센서스가 시행된다면 조사구의 조사인구를 파악할 수 없게 되므로 등록센서스와 병행되는 표본을 설계할 때에는 행정자료만 활용할 수 있을 것이다. 그런데 위의 <그림 2-5>를 통해 알 수 있는 바와 같이, 행정자료만 사용하여도 전부는 아니라고 해도 대부분의 이상치를 식별하는 것이 가능할 것으로 생각된다. 나주시 이외의 다른 11개 지역도 비슷한 상황인데, 나머지 지역들의 산점도는 <부록 2-1>에 수록되어 있다.

제3절 행정자료와 조사자료의 비교

1. 시군구

- 등록센서스가 본격적으로 시행되면 더 이상 조사자료를 활용할 수 없게 되므로 행정자료만을 이용하여 표본설계를 해야 한다. 그런데 현재는 12개 연구대상 지역에 대해 2010년 11월 시점의 행정자료와 조사자료가 모두 활용 가능한 상태이기 때문에, 행정자료를 이용하여 표본설계를 한 후, 해당표본에 대한 조사자료를 활용함으로써 추정치의 정확성을 미리 파악해볼 수 있다.
- 여기서는 새롭게 구축된 12개 지역의 추출틀 자료를 활용하여 행정통계와 조사통계의 차이를 비교 검토한다. 비교 변수는 인구수, 가구수, 주택수인데, 각 변수에 대해 조사인구와 행정인구의 차이의 정도를 파악하고자 한다.
- <표 2-3>에는 12개 지역 인구수의 행정통계와 조사통계가 나와 있다. 행정인구와 조사인구의 차이율¹⁾은 -32.7%에서 4.3%까지의 범위인데, 지역별로 차이율의 양상이 서로 다르다. 광주 동구만 조사인구가 행정인구보다 많은 반면, 나머지 11개 지역에서는 모두 행정인구가 조사인구보다 많다. 행정인구와 조사인구의 차이율에서, 두 자료의 인구수 차이가 가장 큰 지역은 신안군, 울릉군, 임실군, 나주시 순이며, 인구수의 차이율 절대값이 15% 이상이다. 양구군, 김천시, 증평군, 광주 서구는 차이율이 4%에서 8.5% 정도인 중간 수준이고, 나머지 화성시, 영등포구, 금천구, 인천서구는 차이율이 3% 미만으로 비교적 차이가 적은 지역이다.

1) 인구 차이율(%) = $\frac{(\text{조사인구} - \text{등록인구})}{\text{조사인구}} \times 100(\%)$

<표 2-3> 지역별 조사인구와 행정인구

지역	조사인구	행정인구	차이율(%)
신안군	32,987	43,434	-31.7
울릉군	7,737	10,111	-30.7
임실군	23,490	29,183	-24.2
나주시	77,825	89,632	-15.2
양구군	19,252	20,879	-8.5
김천시	126,823	133,523	-5.3
증평군	31,204	32,711	-4.8
화성시	474,160	487,168	-2.7
영등포구	374,543	383,435	-2.4
금천구	230,447	235,825	-2.3
인천서구	389,057	392,378	-0.9
광주동구	103,998	99,488	4.3

- <표 2-4>에는 가구수의 행정통계와 조사통계가 나와 있다. 행정가구수와 조사가구수의 차이율은 -29.7%에서 6.9%까지의 범위로 인구의 차이율과 비슷한 수준을 나타낸다. 인구와 마찬가지로 가구수도 지역별 차이율 양상이 서로 다르게 나타난다. 광주 동구만 조사가가구수가 행정가구수보다 많은 반면, 나머지 11개 지역에서는 모두 행정가구수가 조사가가구수보다 많다. 행정가구수와 조사가구수의 차이가 가장 큰 지역은 울릉군, 신안군, 양구군, 나주시, 임실군인데, 차이율 절대값이 12% 이상이다. 다음으로 영등포구, 금천구, 화성시, 광주 동구는 차이율이 5%에서 7% 사이로 중간 수준이고, 나머지 증평군, 인천서구, 김천시는 차이율 3% 미만으로 적은 지역이다. 앞의 <표 2-3>에 나온 인구수와 비교하면 약간의 차이가 있다. 가령 영등포구를 예로 들면, 인구수의 차이율은 -2.4%인데 비해 가구수의 차이율은 -5.7%여서 가구수의 차이율이 더 크다. 반대로 신안군은 인구수는 -31.7% 차이율로 가장 크지만 가구수는 -18.8%로 인구수만큼 차이율이 크지 않다.

<표 2-4> 지역별 조사가구수와 행정가구수

지역	조사가구수	행정가구수	차이율(%)
울릉군	3,494	4,533	-29.7
신안군	16,114	19,149	-18.8
양구군	6,919	8,113	-17.3
나주시	31,795	36,678	-15.4
임실군	10,693	12,046	-12.7
영등포구	138,637	146,526	-5.7
금천구	86,533	91,257	-5.5
화성시	158,423	166,406	-5.0
증평군	11,559	11,890	-2.9
인천서구	128,975	132,231	-2.5
김천시	48,843	49,529	-1.4
광주동구	42,538	39,609	6.9

- <표 2-5>에는 주택수의 행정통계와 조사통계가 나와 있다. 행정주택수와 조사주택수의 차이율은 -8.1%에서 6.9% 범위로 인구수나 가구수의 차이율과는 그 정도가 다르게 나타난다. 주택수도 지역별 차이율 양상이 서로 다른 편이다. 주택의 경우, 조사주택이 행정주택보다 많은 지역이 8개, 행정주택이 조사주택보다 많은 지역이 4개이다. 차이율의 절대값이 4%를 넘는 지역은 임실군, 양구군, 인천서구, 나주시, 광주동구이다. 나머지 금천구, 영등포구, 울릉군, 신안군, 화성시, 증평군, 김천시는 3.5% 이하이다. 주택수의 경우, 인구수나 가구수에 비해 차이율이 낮은 편이며, 지역별 차이율 양상도 인구수나 가구수와는 뚜렷한 차이를 보인다. 광주동구의 경우, 인구수나 가구수 모두 다른 지역과는 달리 조사통계가 행정통계보다 컸는데 주택수만은 조사통계가 더 작아 차이율이 음(-)으로 나타났다. 다른 지역에서도 주택수는 인구나 가구수와 다른 경향을 띠는데, 인구수와 가구수는 모두 주민행정자료 DB를 기초로 산출되지만 주택은 건축물대장 DB라는 별도의 행정 DB를 통해 산출되기 때문이라고 추측된다.

<표 2-5> 지역별 조사주택수와 행정주택수

지역	조사주택수	행정주택수	차이율(%)
임실군	12,439	13,448	-8.1
인천서구	128,715	135,316	-5.1
나주시	34,807	36,524	-4.9
광주동구	29,302	30,626	-4.5
울릉군	3,356	3,414	-1.7
화성시	153,789	155,791	-1.3
증평군	11,840	11,880	-0.3
김천시	48,542	48,663	-0.2
신안군	17,762	17,438	1.8
영등포구	100,330	97,713	2.6
금천구	52,320	50,493	3.5
양구군	7,289	6,784	6.9

2. 읍면동

- 앞에서는 12개 지역별 행정통계와 조사통계의 차이를 살펴보았다. 이번에는 같은 시군 내에서 읍면동별 차이를 살펴보기로 한다. 12개 시군구 중 행정통계와 조사통계 사이에 차이가 가장 큰 지역인 신안군과 가장 차이가 작은 지역인 인천서구 두 지역을 예로 들어 살펴보기로 한다. 나머지 10개 지역별 차이 양상도 여기에 예로 드는 두 지역과 대체로 비슷하다. 참고로 12개 지역의 읍면동별 비교 결과는 <부록 2-2>에 모두 수록되어 있다.

<표 2-6> 읍면동별 조사인구와 행정인구

군구	읍면동	조사인구	행정인구	차이율
신안군	11	4,109	4,935	-20.1
	31	1,671	2,089	-25.0
	32	3,030	3,543	-16.9
	33	1,826	2,340	-28.1
	34	3,099	3,797	-22.5
	35	2,423	3,034	-25.2
	36	2,930	4,526	-54.5
	37	1,275	2,009	-57.6
	38	1,347	1,767	-31.2
	39	1,287	1,744	-35.5
	40	2,585	3,489	-35.0
	41	936	1,183	-26.4
	42	1,745	2,173	-24.5
	43	4,724	6,805	-44.1
인천서구	51	29,074	29,325	-0.9
	53	47,399	46,864	1.1
	54	10,993	11,642	-5.9
	55	9,076	9,359	-3.1
	56	12,451	12,344	0.9
	58	24,381	24,646	-1.1
	59	16,566	17,379	-4.9
	60	17,933	18,116	-1.0
	62	13,566	14,086	-3.8
	63	23,186	23,390	-0.9
	64	20,569	20,951	-1.9
	65	12,294	12,506	-1.7
	68	39,170	40,119	-2.4
	69	23,113	23,704	-2.6
	70	22,431	21,813	2.8
	71	36,117	35,290	2.3
	72	3,475	3,401	2.1
	73	27,263	27,443	-0.7

- 먼저 인구수에 대한 읍면동별 차이를 비교해보자. 신안군과 인천서구의 읍면동별 조사인구와 행정인구의 차이를 정리한 결과가 <표 2-6>에 제시되어 있다. 앞의 <표 2-3>에 의하면, 신안군의 조사인구와 행정인구 차이율은 -31.7%였다. 읍면동별 차이율을 살펴보면 신안군 내 모든 읍면동에서 조사인구가 행정인구보다 적는데, 최소 -16.9%에서 최대 -57.6%에 이른다. 즉, 같은 신안군이라고 해도 읍면동에 따라 차이율의 편차가 큰 편이라고 할 수 있다. 한편, 차이율이 -0.9%로 가장 낮은 인천서구를 보면, 조사인구보다 행정인구가 많은 동도 있고, 반대로 조사인구가 행정인구가 많은 동도 존재한다. 읍면동별 차이율의 절대값을 보면 최소 0.7%에서 최대 5.9%까지의 범위이다. 신안군에 비해서는 차이의 정도가 적은 편이긴 하지만 동별 차이율에 편차가 존재한다는 면에서는 비슷한 양상이라고 할 수 있다.
- 다음으로 가구수에 대한 읍면동별 차이를 비교해보자. 인구와 마찬가지로 신안군과 인천서구의 읍면동별 가구수 차이를 정리한 결과가 <표 2-7>에 제시되어 있다. <표 2-4>에 의하면, 신안군의 조사가가구수와 행정가구수의 차이율은 인구의 차이율보다는 적은 -18.8%였다. 읍면동별 차이율을 살펴보면, 신안군 내 모든 읍면동에서 조사가가구수가 행정가구수보다 적는데 최소 -4.7%에서 최대 -40.9%에 이른다. 즉, 같은 신안군이라고 해도 읍면동에 따라 차이율의 편차는 큰 편이라고 할 수 있다. 한편, 차이율 -2.5%로 적은 편인 인천서구를 보면, 조사가가구수보다 행정가구수가 많은 동도 있고, 그 반대인 동도 존재한다. 읍면동별 차이율의 절대값을 보면, 최소 0.3%에서 최대 9.3%까지의 범위에 있다. 신안군에 비해서는 차이의 정도가 적은 편이긴 하지만, 동별 차이율에 편차가 존재한다는 면에서는 비슷한 양상을 보인다고 할 수 있다.

<표 2-7> 읍면동별 조사가구수와 행정가구수

군구	동	조사가구수	행정가구수	차이율
신안군	11	2,095	2,277	-8.7
	31	789	907	-15.0
	32	1,266	1,499	-18.4
	33	974	1,142	-17.2
	34	1,524	1,684	-10.5
	35	1,163	1,392	-19.7
	36	1,405	1,979	-40.9
	37	680	920	-35.3
	38	617	730	-18.3
	39	703	824	-17.2
	40	1,350	1,573	-16.5
	41	522	572	-9.6
	42	952	997	-4.7
	43	2,074	2,653	-27.9
인천서구	51	9,307	9,577	-2.9
	53	15,955	15,513	2.8
	54	4,283	4,682	-9.3
	55	3,093	3,203	-3.6
	56	4,156	4,044	2.7
	58	8,837	9,094	-2.9
	59	5,894	6,390	-8.4
	60	6,237	6,316	-1.3
	62	4,999	5,297	-6.0
	63	7,156	7,251	-1.3
	64	7,115	7,220	-1.5
	65	4,474	4,711	-5.3
	68	12,684	13,524	-6.6
	69	7,089	7,680	-8.3
	70	6,963	6,982	-0.3
	71	10,604	10,451	1.4
	72	1,084	1,087	-0.3
	73	9,045	9,209	-1.8

- 마지막으로 주택수에 대한 읍면동별 차이를 비교해보자. 이번에도 신안군과 인천서구의 경우를 예로 드는데, 읍면동별 주택수의 차이를 정리한 결과가 <표 2-8>에 제시되어 있다. <표 2-5>에 의하면, 신안군의 조사주택수와 행정주택수의 차이율은 1.8%인 반면, 인천서구의 차이율은 -5.1%여서 인구나 가구수와는 달리 주택수 차이율은 신안군이 적고 인천서구가 더 큰 것으로 나타났다. 신안군의 경우, 인구수나 가구수는 모든 읍면동에서 조사통계가 행정통계보다 적었는데 주택수는 다른 양상을 보이고 있다. 읍면동에 따라 조사주택수가 행정주택수보다 많은 곳도 있고, 그 반대인 경우도 존재한다. 두 통계의 차이율 절대값을 보면, 최소 0.5%에서 최대 11.2%의 범위에 있고, 읍면동에 따라 차이율의 편차가 큰 것으로 나타난다. 한편, 인천서구는 신안군과 마찬가지로 조사주택수가 행정주택수보다 많은 곳도 있고 적은 곳도 있는데, 대체적으로 조사주택수가 행정주택수보다 적은 동이 많은 편이다. 두 통계의 차이율 절대값을 보면, 최소 0.0%에서 최대 98.3%의 범위에 있으며, 읍면동에 따라 차이율의 편차가 매우 크게 나타난다. 특히, 54번 동은 조사주택수보다 행정주택수가 2배가량 많다는 점에서, 예외적인 지역이라고 보아야 할 것이다.

<표 2-8> 읍면동별 조사주택수와 행정주택수

군구	동	조사주택수	행정주택수	차이율
신안군	11	2,280	2,062	9.6
	31	836	804	3.8
	32	1,357	1,272	6.3
	33	1,103	1,083	1.8
	34	1,631	1,671	-2.5
	35	1,281	1,347	-5.2
	36	1,593	1,415	11.2
	37	798	853	-6.9
	38	652	656	-0.6
	39	798	794	0.5
	40	1,527	1,502	1.6
	41	571	576	-0.9
	42	1,108	1,034	6.7
	43	2,227	2,369	-6.4
인천서구	51	8,942	9,540	-6.7
	53	14,880	15,021	-0.9
	54	3,106	6,158	-98.3
	55	5,965	6,986	-17.1
	56	4,355	4,374	-0.4
	58	7,543	7,583	-0.5
	59	4,786	4,902	-2.4
	60	6,117	6,118	0.0
	62	4,790	5,084	-6.1
	63	7,200	7,227	-0.4
	64	6,688	6,834	-2.2
	65	3,561	3,517	1.2
	68	14,399	15,192	-5.5
	69	7,243	7,277	-0.5
	70	7,287	7,814	-7.2
	71	11,114	10,587	4.7
	72	2,232	2,565	-14.9
	73	8,507	8,537	-0.4

제4절 표본설계

1. 개 요

- 우리나라에서 2015년 등록센서스가 본격적으로 실시되면, 행정자료에 대한 센서스(전수집계)와 병행하여, 그것의 완전성(completeness)을 점검하고 행정자료에 없는 센서스 조사항목에 대한 표본조사가 실시되도록 되어 있다. 센서스 병행 표본조사를 위한 표본설계를 위한 추출틀로는 행정자료를 사용하게 된다. 행정자료는 실시간으로 업데이트 되므로 기존의 인구주택총조사 자료를 추출틀로 사용하는 것에 비해 최신의 모집단 상황을 반영할 수 있다는 장점이 있다.
- 통계청의 2011년 학술연구 용역사업에서 2개의 인구주택총조사 시범조사 지역에 대하여 등록센서스를 위한 표본설계 연구를 실시한 바 있는데 (한국인구학회, 2011), 본 연구에서는 이를 12개 시군구로 확대하여 적용하는 것이 그 골자(骨子)이다. 2011년의 연구를 위한 2개 지역 데이터파일에는 조사구별 인구수와 가구수에 대한 행정자료와 조사자료가 들어 있었는데 비해 2012년의 데이터파일에는 주택수 자료가 추가되며, 아울러 조사구의 특성을 나타내는 변수도 들어있다는 점이 다르다.
- 표본설계의 연구를 단계별로 구분하면 (1) 조사모집단 정의 (2) 이상치 조사구 처리 (3) 표본크기 결정 (4) 표본추출 (5) 추정의 다섯 단계로 나눌 수 있다. 이들 단계별 절차를 흐름도로 나타내면 <그림 2-6>과 같다.
- 2011년 통계청의 학술연구 용역사업에서 내려진 결론과 같이, PSU로는 통계청에서 센서스 목적으로 작성한 조사구를 사용한다. 12개 지역 각각을 독립적인 부모집단(副母集團: subpopulation)으로 간주하여 지역별로 표본크기(sample size)를 결정한다. 하지만 각 지역별 표본설계에는 동일한 방식을 적용한다.

- 행정자료에 의하여 표본을 추출한 후, 표본 조사구별 조사자료를 가지고 추정
을 한다. 추정치와 실제 모수(2010년 총조사 결과)와의 차이율 및 추정치의
상대표준오차(RSE: Relative Standard Error)를 구하여 표본설계의 적절성을
평가한다.

<그림 2-5> 표본설계 연구의 단계별 흐름도



2. 조사모집단 정의

- 본 표본설계 연구의 모집단은 12개 지역의 모든 조사구이다. 제2절의 <표
2-2>에 소개한 바와 같이, 조사구 특성에 따라 이상치 조사구 비율이 확연하
게 다르다. 일반조사구, 아파트조사구, 섬조사구인 경우 이상치 비율이 매우
낮으나, 기숙시설, 특수사회시설, 특수조사구 등은 이상치 비율이 특이하게 높
다. 2011년의 등록센서스 학술연구 용역사업에서는 이러한 조사구 특성을 반
영하지 않았기 때문에 과다한 이상치 조사구들을 전수조사 대상으로 지정하게
되었고, 그 결과 표본설계의 효율성 저하를 가져온 바 있다.
- 통계청 등록센서스 담당 주무부서와의 협의를 통해 기숙시설과 특수사회시설
은 모집단의 구성요소에서 제외하기로 했다. 모집단의 구성요소에서 제외하기
로 한 조사구들이 전체 모집단에서 차지하는 비율이 매우 낮아 포괄범위
(coverage)에 미치는 영향은 미미한 편이다. 또한 등록센서스를 실시하는 대

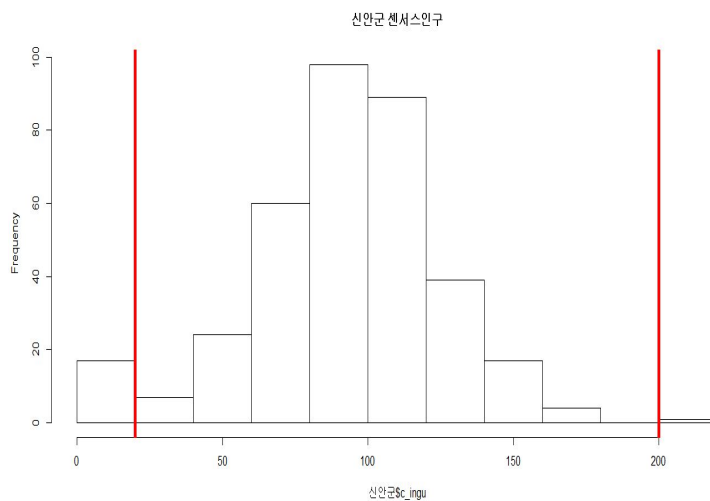
부분의 나라에서 마찬가지로, 우리나라에서도 통계청은 2015년 등록센서스 실시시점에 특수조사구에 대해서는 전수조사(full enumeration)를 실시할 계획으로 되어 있다.

- 따라서 본 연구의 조사모집단은 12개 지역 조사구들 중에서 기숙시설, 특수사회시설, 관광호텔 및 외국인 조사구들을 제외한 모든 조사구이다. 조사모집단에 포함된 특수조사구에 대해서는 전수조사를 실시하고, 나머지 조사구들에 대해서는 표본조사를 실시하게 된다.

3. 이상치 처리

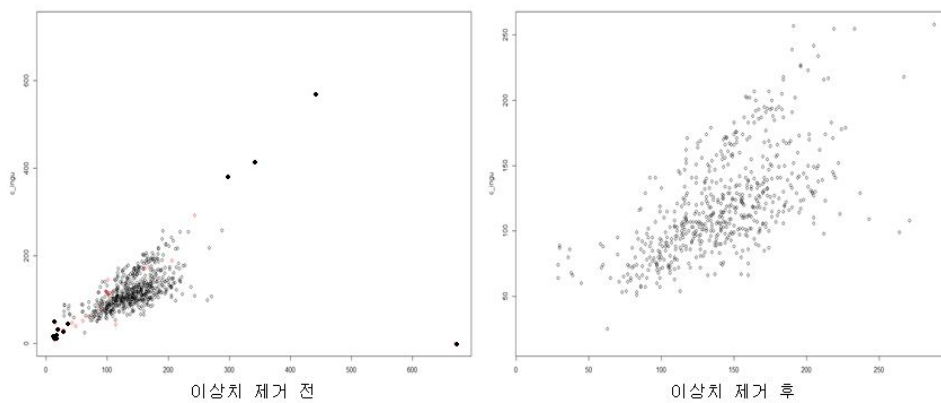
- 2011년의 연구에서는 먼저 표본크기를 결정한 후, 표본추출 단계에서 이상치를 고려하였다. 이상치 조사구에 대해서는 전수조사를 실시하였으므로 최종적인 표본크기는 설계 단계에서 정한 표본크기에 비해 커지게 되었다. 특히 인구가 너무 적어서 이상치로 판정된 조사구의 경우, 표본크기만 키울 뿐 효율적이지 못하다고 판단되었다. 이런 점들을 감안하여, 본 연구에서는 표본크기를 결정하기 전에 먼저 이상치 조사구 처리방안을 마련하고자 한다.
- 12개 지역별로 조사구 인구 분포를 파악하기 위해 지역별 히스토그램을 작성하였다. <그림 2-7>은 신안군의 히스토그램인데, 나머지 지역에 대한 히스토그램은 <부록 2-3>에 나와 있다. 그림의 좌우에 있는 세로줄은 이상치 경계선이다. 그림의 분포를 보면 이상치 기준을 상식적인 수준에서 정하는 것이 그다지 어려운 일이 아니라고 할 수 있다.

<그림 2-7> 신안군 조사구 인구 분포



- 특수조사구들을 미리 조사모집단에서 제외시켰음에도 불구하고 여전히 이상치 조사구들이 존재하므로 이에 대한 처리 기준을 미리 정하는 것이 필요하다. 본 연구에서는 인구가 적은 이상치의 경우 표본추출에서 제외시키고, 인구가 많은 이상치에 대해서는 전수조사를 실시하기로 한다. 인구가 적은 이상치의 경우 표본크기만 크게 할 뿐 추정의 효율에 미치는 영향이 미미하므로 제외하기로 한다.
- 이상치 제거에 따른 효과를 살펴보기 위해 행정인구와 조사인구에 대한 산점도를 작성하였다. <그림 2-8>은 나주시의 행정인구와 조사인구 산점도를 나타낸 것인데, 왼쪽은 이상치 제거 전 산점도이고 오른쪽은 이상치를 제거한 후의 산점도이다. 나머지 지역별 산점도는 부록에 수록되어 있다. 그림을 보면, 이상치 제거 전과 후의 산점도가 사뭇 다른데, 이를 통해 이상치를 사전에 적절히 처리함으로써 보다 효과적인 표본설계를 할 수 있다는 점을 시사한다.

<그림 2-8> 이상치 제거 전과 후의 인구 산점도(나주시)



4. 표본크기 결정

- 표본이론에 따르면, 최소 통계작성범위 단위로 표본크기를 결정하는 것이 바람직하다. 우리나라 총조사의 최소 통계작성범위는 읍면동인데, 전국의 읍면동 수는 무려 3,000개가 넘는다. 만일 읍면동별로 표본크기를 결정하여 조사를 해야 한다면 표본설계 단계가 지나치게 복잡해질뿐더러 현실적으로 조사관리가 불가능해질 것이다. 따라서 본 표본설계에서는 2011년 연구에서와 같이 시군구 단위로 표본크기를 결정하기로 한다. 이렇게 하면 같은 시군구 내의 읍면동들에 대해서는 모두 동일한 추출률을 적용하게 된다.
- 지역별 표본크기를 결정하기 위해 먼저 모든 지역에 대해 20% 추출률을 동일하게 적용하여 각 지역별 상대표준오차 (RSE: Relative Standard Error)를 계산한 결과가 <그림 2-9>에 나와 있다. 지역별 RSE는 최소 1.2%에서 최대 5.7% 범위로 나타났다.

<그림 2-9> 지역별 상대표준오차 (20% 표본추출률)

지역	모집단 (N)	표본 (n)	상대표준오차 (RSE, %)
금천구	1,423	285	1.5
영등포구	2,502	500	1.4
인천서구	2,146	429	1.2
광주동구	745	149	2.3
화성시	2,704	541	1.3
양구군	134	27	5.1
증평군	217	43	4.6
임실군	205	41	4.2
나주시	611	122	2.6
신안군	329	66	3.0
김천시	826	165	2.0
울릉군	65	13	5.7

- 이론적으로는 각 지역별 RSE 목표수준(target level)을 동일하게 하는 것이 바람직하다. 가령, 12개 지역의 RSE 목표수준을 2%로 할 경우, RSE가 2%보다 작은 지역에 대해서는 추출률을 더 낮추어야 하는 반면, 2%보다 큰 지역은 추출률을 높여야 한다. 울릉군은 20% 추출률에서 RSE가 5.7%이므로 RSE 목표수준에 맞추기 위해서는 추출률을 대폭 올려야 하므로 거의 전수조사를 하는 상황이 될 수도 있다. 현실적으로 모든 지역의 RSE 목표수준을 같게 하는 것이 어려워 일반적으로는 목표 RSE 수준을 여러 가지로 나눈다. 본 연구에서는 지역에 따라 목표 RSE를 1.5%, 2%, 2.5%, 4% 중에서 적절히 정하기로 한다.
- 지역별 RSE를 기초로 하여 적절한 수준의 RSE 목표수준을 정하고, 그에 따른 표본크기를 결정할 것이다. 참고로, RSE 목표수준에 따라 새로운 표본크기를 계산하기 위해 (2.1)의 공식을 사용한다.

$$n_{\text{new}} = n_{\text{old}} \times \left(\frac{C_{\text{new}}}{C_{\text{old}}} \right)^2 \quad (2.1)$$

- 지역별 RSE 목표수준과 식 (2.1)에 따라 계산한 표본크기 및 추출률이 다음의 <표 2-10>에 나와 있다. 제일 오른쪽 열에 나온 지역별 추출률을 보면, 인천 서구가 추출률 11.7%로 가장 낮은 반면, 울릉군이 40%로 가장 높다. 일반적으로 대도시 지역의 추출률이 비교적 낮고, 군지역의 추출률이 상대적으로 높은 편이다.
- 여기서 구한 표본크기 외에 전수조사 대상을 포함시키게 되면 지역별 추출률은 <표 2-10>에서 제시한 것보다 다소 높아질 것이다.

<표 2-10> 지역별 RSE 목표수준과 표본추출률

지역	모집단(N)	RSE (추출률 20%)	RSE 목표수준	표본 (n)	추출률 (%)
금천구	1,423	1.5	1.5	297	20.9
영등포구	2,502	1.4	1.5	417	16.7
인천서구	2,146	1.2	1.5	252	11.7
광주동구	745	2.3	2	194	26.0
화성시	2,704	1.3	1.5	400	14.8
양구군	134	5.1	4	44	32.8
증평군	217	4.6	4	57	26.3
임실군	205	4.2	4	45	22.0
나주시	611	2.6	2	198	32.4
신안군	329	3.0	2.5	94	28.6
김천시	826	2.0	2	170	20.6
울릉군	65	5.7	4	26	40.0

5. 표본추출

- 가장 효율적인 표본을 얻기 위해서는 지역별로 독립적인 표본설계를 하는 것이 바람직하다. 즉, 각 시군구별로 별도의 층화 집락추출(stratified cluster sampling) 방법을 결정할 수 있다. 하지만 전국의 수많은 시군구에 대해 독립적인 표본설계를 하는 일은 현실적으로 쉬운 일이 아니고, 설사 그렇게 했더라도 그에 따라 조사를 관리하기가 너무 힘들다. 따라서 각 지역별 표본크기만 독립적으로 결정하기로 하고 표본추출을 위해서는 하나의 표준화된 표본추출 방법을 개발하여 전국 모든 지역에 동일하게 적용하기로 한다.
- 2011년의 연구에서는 동일한 행정자료에서 인구수와 가구수 변수를 얻었으므로 조사구 인구수와 가구수의 상관계수가 높은 편이었다. 그래서 두 변수 중 인구만을 정렬변수로 하여 조사구를 정렬한 후 계통추출(系統抽出:

systematic sampling)하는 방법을 사용하였다. 그런데 본 연구에서는 인구, 가구수 외에 주택수라는 변수도 추가되었으므로 표본추출 과정에서 과거와 같이 조사구 인구수만 고려해도 되는지 아니면 다른 변수들도 동시에 고려하는 것이 더 나은지에 대해 검토하는 것이 필요하다.

- 지역별 행정자료와 조사자료 사이의 상관계수를 살펴보기로 하자. 인구수, 가구수, 주택수에 대한 상관계수 행렬(correlation matrix)을 계산한 결과가 <표 2-11>에 나와 있다. 대체로 행정자료와 조사자료 사이의 상관계수는 큰 편이나, 일부 지역에서는 상관계수가 낮은 경우도 존재한다. 행정자료와 조사자료의 상관계수가 0.5 이하인 칸에는 음영으로 표시하였다. 가령, 인천서구에서는 행정주택수와 조사주택수의 상관계수가 0.28로 매우 작다. 인천서구의 인구수와 가구수의 경우는 상관계수가 각각 0.91과 0.77로 큰데 유독 주택수의 상관계수만 작다. 인구수와 가구수는 동일한 DB에서 얻었으나 주택수는 다른 DB에서 구한 때문에 이런 현상이 나타나는 것으로 생각되지만, 이러한 예외적인 사례에 대해서는 해당지역의 이해당사자들과도 접촉하여 지역특성을 정성적으로 파악할 필요도 있다고 판단된다.
- 동일한 변수에 대한 행정자료와 조사자료 사이의 상관계수는 비교적 높은 편이지만 서로 다른 변수 간의 상관계수는 그리 크지가 않다. 금천구를 예로 들면, 행정인구수와 조사인구수의 상관계수는 0.94로 크나, 행정인구수와 조사가구수는 0.56, 행정인구수와 조사주택수는 0.49이다. 하지만 대부분의 지역에서 행정인구수와 다른 변수 사이의 상관계수가 0.5 내외가 되므로 행정인구를 정렬변수로 사용해도 무방할 것으로 여겨진다.
- 인구수, 가구수, 주택수 간의 상관계수가 그다지 크지 않다는 점에 착안하여 조사구의 행정인구수 외에 행정가구수, 행정주택수 모두를 정렬변수로 사용하면 효율을 더 높일 수 있을 지도 모른다. 그래서 본 연구에서는 (1) 행정인구수만을 정렬변수로 사용하는 방안과 (2) 행정인구수와 행정주택수 두 개를 동시에 고려하는 방안, (3) 행정인구수, 행정가구수, 행정주택수 세 변수 모두를 동시에 고려하는 방안을 사용하여 어느 것이 효과적인지를 비교하기로 한다.

<표 2-11> 지역별 행정자료와 조사자료의 상관계수

지역	행정 \ 조사	인구	가구수	주택수
금천구	인구	0.94	0.56	0.49
	가구수	0.52	0.76	0.07
	주택수	0.65	0.19	0.94
영등포구	인구	0.84	0.62	0.52
	가구수	0.55	0.66	0.22
	주택수	0.74	0.50	0.91
인천서구	인구	0.91	0.79	0.52
	가구수	0.64	0.77	0.37
	주택수	0.09	0.04	0.28
광주동구	인구	0.83	0.41	0.71
	가구수	0.51	0.46	0.40
	주택수	0.72	0.35	0.92
화성시	인구	0.85	0.62	0.67
	가구수	0.50	0.62	0.48
	주택수	0.60	0.45	0.84
양구군	인구	0.66	0.66	0.51
	가구수	0.49	0.64	0.45
	주택수	0.51	0.51	0.64
증평군	인구	0.78	0.53	0.45
	가구수	0.52	0.64	0.43
	주택수	0.55	0.60	0.90
임실군	인구	0.83	0.72	0.54
	가구수	0.69	0.80	0.64
	주택수	0.55	0.75	0.85
나주시	인구	0.64	0.57	0.41
	가구수	0.25	0.49	0.38
	주택수	0.48	0.74	0.86
신안군	인구	0.56	0.57	0.54
	가구수	0.47	0.60	0.63
	주택수	0.44	0.62	0.75
김천시	인구	0.73	0.54	0.49
	가구수	0.40	0.67	0.58
	주택수	0.48	0.61	0.91
울릉군	인구	0.51	0.42	0.15
	가구수	0.33	0.33	0.17
	주택수	0.12	0.25	0.67

- 첫 번째 추출방안은, 각 지역별 추출틀에서 특별조사구나 이상치를 제외한 후 조사구를 행정인구의 크기순으로 정렬한 후 계통추출법에 의해 지정된 표본크기(sample size)대로 표본 조사구를 추출하는 방안이다.
- 두 번째는 행정인구수와 행정주택수를 사용하는 방안인데, 각 조사구에 대해 두 변수를 사용한 마할라노비스 거리(Mahalanobis distance)²⁾를 계산한 후, 그 거리의 크기순으로 정렬한다. 이후 계통추출법으로 표본을 추출하는 방안이다.
- 마지막으로 행정인구, 행정가구수, 행정주택수를 모두 사용하는 방안으로써 두 번째 방안과 마찬가지로 마할라노비스 거리(Mahalanobis distance)를 계산한 후, 그 거리의 크기순으로 하고 계통추출법으로 표본을 추출하는 방안이다.
- 세 가지 표본추출 방안을 사용하여 얻은 표본의 조사인구를 가지고 추정을 한 결과가 다음의 <표 2-12>에 나와 있다. 추정 결과를 보면, 지역별로 양상이 서로 엇갈리게 나타난다. 다시 말해, 모든 지역에 대해 차이율의 절대값이 적으면서 동시에 RSE도 적은 방안은 존재하지 않는다. 금천구에서는 차이율과 RSE 둘 다 방안 1이 가장 낮은 것으로 나타났다. 그러나 양구군의 경우, 차이율은 방안 2가 좋고, RSE는 방안 3이 좋다. 이런 현상은 <표 2-13>에 소개된 가구수 추정 결과와 <표 2-14>의 주택수 추정 결과에서 공통적으로 나타난다.

2) 마할라노비스 거리는 군집분석(cluster analysis)에서 가장 많이 사용되는 거리 개념으로, 두 지점의 단순한 거리만이 아니라, 변수 특성을 나타내는 표준편차와 상관계수를 동시에 고려할 수 있다는 장점이 있다. 마할라노비스 거리는 각각의 케이스가 다양한 변수들의 중심값(예: 평균값)들로 이루어진 중심(centroid)에 대해서 갖는 거리를 말한다. 개념적으로 살펴보면, “multivariate test” 곧 여러 변수들을 동시에 이용하여 살펴보는 테스트 경우에 각각의 중심값을 중앙에 교차시켜 케이스 값들을 나열해보면 일종의 군집을 이루는데, 마할라노비스 거리는 특정 케이스의 값이 여기서 심하게 벗어났는가를 보기 위한 거리값이다. 이렇게 얻은 각 케이스 값의 데이터로 분포를 그려서 극한값을 찾아낼 수 있는데, 데이터의 이상치를 찾는 데 쓰이기도 한다.

요약하자면, 모든 지역에서 모든 조사변수에 대해 다른 추출방안을 압도하는 방안이 존재하지 않는다.

- 세 가지 방안 중 어느 것도 압도적으로 우월하지 않고 대체로 비슷한 수준을 나타내는 경우라면 실제 적용 상 가장 쉽고 편리한 방안이 효과적이다. 따라서 적용이 가장 쉽고 편리한 방안 1 즉, 행정인구만으로 정렬하는 방안을 사용하기로 결정한다.

<표 2-12> 세 가지 추출방안별 추정 결과 (인구수 추정)

지역	N (개)	n (개)	방안 1 (인구)		방안 2 (인구, 주택)		방안 3 (인구,가구,주택)	
			차이율 (%)	RSE (%)	차이율 (%)	RSE (%)	차이율 (%)	RSE (%)
금천구	1,492	304	-1.2	1.8	-3.6	1.8	5.0	3.3
영등포구	2,595	423	-1.0	1.8	-2.2	1.7	1.8	2.6
인천서구	2,246	267	-1.5	2.1	-1.5	2.1	2.9	1.9
광주동구	799	194	7.0	2.5	6.0	2.4	-5.7	3.3
화성시	2,869	421	-2.4	1.8	-1.2	1.8	-2.1	1.8
양구군	146	46	-6.5	7.1	0.3	7.1	17.3	5.3
증평군	226	58	-5.6	5.1	-3.2	5.0	2.7	4.2
임실군	219	46	-16.0	5.1	-30.6	4.9	-3.0	4.2
나주시	660	202	-9.6	2.7	-9.3	2.7	-6.6	1.9
신안군	356	95	-30.3	3.2	-25.5	3.2	4.2	2.8
김천시	890	172	-1.5	2.6	-3.7	2.6	-0.5	2.7
울릉군	67	26	-26.4	4.6	-31.5	4.4	1.6	4.4

<표 2-13> 세 가지 추출방안별 추정 결과 (가구수 추정)

지역	N (개)	n (개)	방안 1 (인구)		방안 2 (인구, 주택)		방안 3 (인구,가구,주택)	
			차이율 (%)	RSE (%)	차이율 (%)	RSE (%)	차이율 (%)	RSE (%)
금천구	1,492	304	-4.5	1.5	-6.6	1.5	0.6	3.2
영등포구	2,595	423	-4.3	1.3	-7.7	1.3	2.7	2.6
인천서구	2,246	267	-3.0	1.7	-2.8	1.7	0.8	1.9
광주동구	799	194	7.0	1.4	6.3	1.4	-5.8	3.3
화성시	2,869	421	-5.3	1.3	-4.2	1.3	1.2	1.8
양구군	146	46	-12.0	3.6	-12.2	3.6	8.2	4.9
증평군	226	58	-6.0	3.0	-5.9	3.0	-2.1	4.2
임실군	219	46	-5.7	3.4	-18.9	3.4	-13.6	4.0
나주시	660	202	-13.6	1.4	-13.0	1.4	-4.9	1.8
신안군	356	95	-17.9	2.1	-14.7	2.1	5.4	2.7
김천시	890	172	-0.7	1.6	-2.5	1.6	-2.9	2.7
울릉군	67	26	-22.7	2.9	-26.0	2.9	5.5	4.3

<표 2-13> 세 가지 추출방안별 추정 결과 (주택수 추정)

지역	N (개)	n (개)	방안 1 (인구)		방안 2 (인구, 주택)		방안 3 (인구,가구,주택)	
			차이율 (%)	RSE (%)	차이율 (%)	RSE (%)	차이율 (%)	RSE (%)
금천구	1,492	304	5.0	3.3	-7.2	1.5	2.5	3.2
영등포구	2,595	423	1.8	2.6	-5.7	1.3	2.8	2.6
인천서구	2,246	267	2.9	1.9	-1.9	1.7	1.4	1.9
광주동구	799	194	-5.7	3.3	6.5	1.4	-6.4	3.3
화성시	2,869	421	-2.1	1.8	-7.1	1.3	-2.1	1.8
양구군	146	46	17.3	5.3	-10.6	3.6	10.9	4.9
증평군	226	58	2.7	4.2	-2.1	3.0	1.6	4.2
임실군	219	46	-3.0	4.2	-14.7	3.3	-8.9	4.0
나주시	660	202	-6.6	1.9	-15.7	1.4	-4.9	1.8
신안군	356	95	4.2	2.8	-16.7	2.1	2.2	2.7
김천시	890	172	-0.5	2.7	-6.4	1.6	-4.9	2.7
울릉군	67	26	1.6	4.4	-34.4	2.8	0.0	4.3

- 조사구 추출틀에 별도로 표본 여부를 나타내는 변수(표본)을 만들어 기입하면 표본추출이 완료된다. <그림 2-9>는 표본여부 변수를 추가한 추출틀 일부를 나타낸 그림이다.

<그림 2-9> 표본여부를 표시한 추출틀 예시

1	지역	gu_code	동	조사구	특성	센서스인구	행정인구	센서스가구	행정가구	센서스주택	행정주택	표본
2	금천구	11180	51	000	9	225						0
3	금천구	11180	51	001	1	147	141	86	79	39	47	0
4	금천구	11180	51	002	4	15	13			1	1	0
5	금천구	11180	51	003	1	74	77	45	42	13	12	1
6	금천구	11180	51	004	1	109	127	51	67	19	19	0
7	금천구	11180	51	005	1	150	172	56	68	37	35	0
8	금천구	11180	51	006	1	118	115	54	56	19	19	0
9	금천구	11180	51	007	1	132	139	87	76	17	19	0
10	금천구	11180	51	008	1	140	152	106	80	13	13	0
11	금천구	11180	51	009	1	160	140	109	94	41	16	1
12	금천구	11180	51	010	1	129	160	85	121	12	12	0
13	금천구	11180	51	011	1	386	298	330	182	20	22	2
14	금천구	11180	51	012	A	239	244	70	69	70	70	0
15	금천구	11180	51	013	A	233	233	70	66	70	70	0
16	금천구	11180	51	014	A	208	211	70	68	70	70	0
17	금천구	11180	51	015	1	128	119	94	87	17	16	0
18	금천구	11180	51	016	1	172	158	143	119	10	17	0

6. 추정 결과

- 각 지역별 최종 표본크기는 <표 2-10>에서 정한 표본크기에 전수조사를 실시하기로 한 특별조사구와 인구가 과다하게 큰 이상치 조사구를 추가함으로 <표 2-14>와 같이 결정되었다. 최종 추출된 표본조사구에 가중치를 부여하는데, 표본조사 대상인 경우에는 설계가중치(design weights)를, 전수조사 대상인 경우에는 1을 부여하게 된다.

<표 2-14> 지역별 최종 표본크기 (전수조사 대상 포함)

지역	N	최종 표본크기 (n)			추출률 (%)
		표본	전수	합계	
금천구	1,431	297	7	304	21.2
영등포구	2,529	417	6	423	16.7
인천서구	2,201	252	15	267	12.1
광주동구	734	194	0	194	26.4
화성시	2,748	400	21	421	15.3
양구군	133	44	2	46	34.6
증평군	218	57	1	58	26.6
임실군	201	45	1	46	22.9
나주시	614	198	4	202	32.9
신안군	328	94	1	95	29.0
김천시	824	170	2	172	20.9
울릉군	64	26	0	26	40.6

- 표본으로 추출된 조사구의 조사자료를 조사변수로 간주하여 지역별 인구, 가구수, 주택수를 추정한다. 추정의 정확성을 평가하기 위해 추정치(sample estimate)와 모수(population parameter)의 차이율과 상대표준오차(RSE: Relative standard error)를 구한다. 각각의 추정식은 다음의 (2.2)~(2.5)와 같다.

$$\text{인구수 추정식: } \hat{Y} = \sum_{i=1}^n w_i \cdot y_i \quad (2.2)$$

$$\text{차이율: } \frac{|Y - \hat{Y}|}{Y} \times 100 \quad (\%) \quad (2.3)$$

$$\text{추정분산: } \hat{V}(\hat{Y}) = \frac{n(1-n/N)}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.4)$$

$$\text{상대표준오차: } RSE(\hat{Y}) = \frac{\sqrt{\hat{V}(\hat{Y})}}{\hat{Y}} \quad (2.5)$$

위 식에서, w_i 는 가중치, $y_i = w_i \cdot y_i$, $\bar{y} = \sum_{i=1}^n y_i / n$ 이다.

A. 시군구 추정 결과

(1) 인구

- 표본조사구의 자료를 이용하여 추정한 각 시군구별 추정치가 다음의 <표 2-15>에서 <표 2-17>에 나와 있다. 각각의 표에는 모집단 조사구수 (N), 표본조사구수 (n), 추출률 (f), 모집단 총계, 추정치, 차이율 (%), 상대표준오차 (RSE)가 수록되어 있다.

- <표 2-15>에 나온 시군구별 인구에 대한 추정 결과를 보면, 모수와 추정치 사이의 차이율 절대치가 최소 0.3%에서 최대 31.5% 범위로 나타난다. 양구군, 화성시, 인천서구, 영등포구, 증평군, 금천구, 김천시의 7개 지역은 차이율 절대치가 3.7% 이하로 적다. 광주동구(6%), 나주시(9.3%)로 다소 큰 편이고, 나머지 신안군(25.5%), 임실군(25.5%), 울릉군(31.5%)는 차이가 굉장히 많이 나는 지역이다. 앞의 <표 2-3>을 참고하면, 추정의 차이율이 큰 것으로 나타난 신안, 울릉, 신안 세 지역 모두 조사인구와 행정인구의 차이율 절대치가 20%를 초과한 지역이라는 사실을 알 수 있다. 표본설계를 위한 추출틀로 행정자료를 사용했기 때문에 행정자료와 조사자료 자체가 차이가 나면 최선을 다해 표본설계를 한다고 해도 추정치와 모수는 차이가 날 수 밖에 없다..

<표 2-15> 지역별 인구수 추정 결과

지역	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
금천구	1,492	304	20.4	230,447	238,844	-3.6	1.8
영등포구	2,595	423	16.3	374,543	382,778	-2.2	1.7
인천서구	2,246	267	11.9	389,057	394,975	-1.5	2.1
광주동구	799	194	24.3	103,998	97,710	6.0	2.4
화성시	2,869	421	14.7	474,160	479,820	-1.2	1.8
양구군	146	46	31.5	19,252	19,199	0.3	7.1
증평군	226	58	25.7	31,204	32,204	-3.2	5.0
임실군	219	46	21.0	23,490	30,685	-30.6	4.9
나주시	660	202	30.6	77,825	85,095	-9.3	2.7
신안군	356	95	26.7	32,987	41,409	-25.5	3.2
김천시	890	172	19.3	126,823	131,499	-3.7	2.6
울릉군	67	26	38.8	7,737	10,175	-31.5	4.4

- 이번에는 <표 2-15>의 가장 오른쪽 열에 소개된 상대표준오차(RSE)를 살펴보자. 양구군을 제외한 11개 지역의 RSE는 5% 이내로 나타나 RSE 목표수준에 근접하고 있음을 알 수 있다. 양구군만 RSE가 7.1%로 약간 큰 편인데, 양구군의 모집단 조사구수가 상대적으로 적어 추출률이 비교적 컸음에도 불가피하게 나타난 현상이다. 전반적으로 RSE 측면에서 볼 때 표본에 그리 큰 문제가 없다고 할 수 있다.
- 인구수에 대한 지역별 차이율과 상대표준오차를 검토한 결과, 추출률과는 관계 없이 표본조사를 통한 추정치와 행정자료 사이에 일정한 괴리가 생길 수 밖에 없는 지역들이 존재하는 것을 알 수 있었다. 인구가 상대적으로 적은 시골지역 중 조사자료와 행정자료 사이에 차이가 많이 나는 지역들에 이런 괴리가 집중적으로 생긴다. 따라서 이러한 괴리를 보정하기 위해서는 별도의 통계적 추정 연구를 추가하는 것이 반드시 필요하다.

(2) 가구수

- <표 2-16>에 나온 시군구별 가구수에 대한 추정 결과를 보면, 모수와 추정치 사이의 차이율 절대치가 최소 0.7%에서 최대 22.7% 범위로 나타난다. 대체로 인구에서 나타난 바와 비슷한 양상을 보이는데, 광주동구를 제외한 모든 지역에서 모집단 가구수보다 추정 가구수가 큰 것으로 나타났다. 차이율 절대치가 7%이하인 지역은 김천시, 인천서구, 영등포구, 금천구, 화성시, 임실군, 증평군, 광주동구의 8개 지역이다. 나머지 울릉, 신안, 나주, 양구군은 차이율이 10%를 초과한다. 특이한 지역으로 임실군과 양구군을 들 수 있다. 임실군은 인구수 차이율은 -30.6%로 매우 컸지만 가구수 차이율은 -5.7%에 불과하다. 반면, 양구군은 인구수 차이율은 0.3%로 매우 작았지만 가구수 차이율은 -12%로 큰 편이다. 즉, 표본자료로 지역별 가구수를 추정한 결과, 지역에 따라 인구수와 비슷한 양상을 보이는 곳도 있지만 전혀 다른 양상을 띠는 지역도 있다.

<표 2-16> 지역별 가구수 추정 결과

지역	N (개)	n (개)	f (%)	모집단 가구수	추정 가구수	차이율 (%)	RSE (%)
금천구	1,492	304	20.4	86,533	90,397	-4.5	1.5
영등포구	2,595	423	16.3	138,637	144,588	-4.3	1.3
인천서구	2,246	267	11.9	128,975	132,799	-3.0	1.7
광주동구	799	194	24.3	42,538	39,543	7.0	1.4
화성시	2,869	421	14.7	158,423	166,754	-5.3	1.3
양구군	146	46	31.5	6,919	7,748	-12.0	3.6
증평군	226	58	25.7	11,559	12,247	-6.0	3.0
임실군	219	46	21.0	10,693	11,299	-5.7	3.4
나주시	660	202	30.6	31,795	36,123	-13.6	1.4
신안군	356	95	26.7	16,114	19,002	-17.9	2.1
김천시	890	172	19.3	48,843	49,209	-0.7	1.6
울릉군	67	26	38.8	3,494	4,288	-22.7	2.9

- 이번에는 <표 2-16>의 가장 오른쪽 열에 소개된 상대표준오차를 살펴보자. 12개 전 지역의 RSE가 4% 이내로 나타나 RSE 목표수준을 달성하고 있다. 전반적으로 RSE 측면에서 볼 때 표본에 문제가 없다고 할 수 있다.
- 가구수에 대한 지역별 차이율과 상대표준오차를 검토한 결과, 추출률과는 관계 없이 표본조사를 통한 추정치와 행정자료 사이에 일정한 괴리가 생길 수 밖에 없는 지역들이 존재하는 것을 알 수 있었다. 인구수와 가구수는 같은 주민행정자료에서 얻은 자료이지만, 양구군이나 임실군과 같이 지역에 따라서는 가구수와 인구의 추정 결과가 다를 수도 있음이 흥미롭다. 이 문제에 대해서도 심층적 연구가 필요하다.

(3) 주택수

• <표 2-17>의 시군구별 주택수에 대한 추정 결과를 보면, 인구나 가구 수와는 달리 양구군을 제외한 11개 지역 모두 차이율 절대치가 7% 미만이다. 모집단 수가 추정치보다 크거나 작은 지역이 고르게 나타나는 것도 차이가 난다. 인구나 가구와 주택의 양상이 이렇게 다른 이유는 주택에 대한 행정자료가 주민등록 DB가 아닌 주택 관련 DB인 점에 기인하는 것으로 판단되며, 이들의 원인에 대한 심층적 연구가 필요하다.

<표 2-17> 지역별 주택수 추정 결과

지역	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
금천구	1,492	304	20.4	52,320	49,713	5.0	3.3
영등포구	2,595	423	16.3	100,330	98,510	1.8	2.6
인천서구	2,246	267	11.9	128,715	125,002	2.9	1.9
광주동구	799	194	24.3	29,302	30,975	-5.7	3.3
화성시	2,869	421	14.7	153,789	156,956	-2.1	1.8
양구군	146	46	31.5	7,289	6,030	17.3	5.3
증평군	226	58	25.7	11,840	11,520	2.7	4.2
임실군	219	46	21.0	12,439	12,813	-3.0	4.2
나주시	660	202	30.6	34,807	37,101	-6.6	1.9
신안군	356	95	26.7	17,762	17,020	4.2	2.8
김천시	890	172	19.3	48,542	48,795	-0.5	2.7
울릉군	67	26	38.8	3,356	3,303	1.6	4.4

- 이번에는 <표 2-17>의 가장 오른쪽 열에 소개된 상대표준오차(RSE: Relative standard error)를 살펴보자. 양구군을 제외한 11개 지역의 RSE가 5% 이내로 나타나 RSE 목표수준을 달성하고 있다. 가장 RSE가 큰 양구군도 5.3%여서 전반적으로 RSE 측면에서 볼 때 표본에 문제가 없다고 할 수 있다.
- 동일한 표본조사구를 추출한 후 인구, 가구수, 주택수를 추정하였음에도 불구하고 주택수에 대한 추정 차이율이 인구나 가구수에 비해 상대적으로 작았다. 특히, 인구가 적은 시골 군에서 인구와 가구 행정자료와 조사자료에 차이가 큰 경향이 있었으나 주택의 경우에는 행정자료와 조사자료에 차이가 훨씬 적은 것으로 나타났다.

B. 읍면동 추정 결과

- 본 표본설계에서는 시군구를 독립적인 부모집단인 층으로 간주하여 지역별 목표오차를 정한 후 표본크기를 정하였다. 따라서 시군구보다 하위인 읍면동별로는 동일한 추출률과 추출방식을 적용하였다. 이 방법은 같은 시군구 내의 읍면동들은 동일한 특성을 가지는 것을 전제한 것이므로 읍면동별 추정 결과를 검토하여 이러한 전제가 타당한 것인지를 확인해야 한다. 뿐만 아니라 시군구 단위로 RSE 목표수준을 정했을 때 읍면동 수준에서는 RSE가 얼마나 되는지도 검토하는 것이 필요하다.
- 12개 지역의 모든 읍면동의 분석 결과를 살필 경우 너무 방대해지므로 여기서는 대표로 세 지역에 대해서만 읍면동별 결과를 소개하기로 한다. 나머지 지역의 결과들은 <부록 2-4>에 수록된 표를 참고하면 된다. 대도시 지역을 대표하여 금천구를, 중소도시의 대표로는 화성시를, 시골지역의 대표로는 신안군을 택하여 각각 인구, 가구수, 주택수에 대한 추정 결과를 분석한다.

(1) 인구

① 금천구

- <표 2-18>을 보면 금천구는 조사구 추출률이 20.4%인 지역인데, 차이율 -1.2%, RSE 1.8%로 추정의 정확성이 높은 지역이다. 금천구에는 11개의 동이 있는데, 동별 차이율 절대치가 5.1% 이하이고, RSE는 8.4% 이하로서 동별 정확성도 높은 편이다.

<표 2-18> 금천구 읍면동별 인구 추정

지역	동	N	n	f	모집단 인구	추정 인구	차이율 (%)	RSE (%)
금 천 구	51	157	32	20.4	18,133	18,414	-1.6	8.4
	52	212	43	20.3	31,061	30,424	2.0	6.1
	53	142	29	20.4	22,582	23,736	-5.1	4.6
	54	196	42	21.4	27,351	27,492	-0.5	4.7
	55	117	24	20.5	18,293	18,624	-1.8	4.9
	57	197	40	20.3	29,086	29,656	-2.0	5.0
	58	134	25	18.7	24,851	24,340	2.1	4.8
	59	71	14	19.7	12,357	12,266	0.7	6.6
	60	131	28	21.4	23,361	23,951	-2.5	3.5
	61	135	27	20.0	23,372	24,349	-4.2	4.2
	전체	1,492	304	20.4	230,447	233,252	-1.2	1.8

② 화성시

- 도농복합형 도시인 경기도 화성시의 추정 결과가 <표 2-19>에 나와 있다. 화성시의 추출률은 14.7%이고, 화성시의 추정 차이율 -2.4%, RSE 1.8%로서 시 전체로는 추정의 정확성이 높은 지역이다. 그런데 읍면동별 결과를 보면, 차이율과 RSE 공통적으로 값이 적은 지역과 큰 지역이 뚜렷이 구분되는 것을 알 수 있다. 가령, 맨 위의 12번 동의 경우 차이율 1.7%, RSE 4.9%로 추정의 정확성이 높은 반면, 44번 면은 차이율 -24.8%, RSE 36.9%로 정확성이 너무

낮다. 도시의 특성을 지니는 읍면동의 경우 차이율과 RSE 모두 적은 편이나 시골 특성의 면지역에서는 차이율과 RSE가 커진다. 그러므로 지역 내 읍면동의 특성을 무시한 채 모든 읍면동에 동일한 추출률을 적용하는 것은 바람직하지 않다.

<표 2-19> 화성시 읍면동별 인구 추정

지역	동	N	n	f	모집단 인구	추정 인구	차이율 (%)	RSE (%)
화 성 시	12	367	53	14.4	63,694	62,618	1.7	4.9
	13	117	15	12.8	14,821	16,292	-9.9	10.8
	14	299	46	15.4	49,188	49,735	-1.1	5.4
	31	46	6	13.0	6,853	7,037	-2.7	10.4
	33	34	5	14.7	5,101	6,618	-29.7	14.5
	35	35	4	11.4	4,368	4,982	-14.1	19.7
	36	71	9	12.7	8,597	10,701	-24.5	11.8
	37	39	5	12.8	4,587	6,625	-44.4	15.3
	38	58	8	13.8	8,007	10,235	-27.8	11.8
	39	67	9	13.4	8,890	9,538	-7.3	14.1
	42	32	4	12.5	3,157	4,367	-38.3	19.8
	43	82	10	12.2	11,050	12,628	-14.3	15.5
	44	65	6	9.2	4,360	5,442	-24.8	36.9
	51	132	18	13.6	21,177	21,916	-3.5	8.0
	52	243	38	15.6	34,872	33,124	5.0	7.1
	53	142	20	14.1	27,098	26,655	1.6	4.5
	54	137	23	16.8	26,818	27,147	-1.2	5.4
	55	92	14	15.2	19,637	19,993	-1.8	5.9
	56	74	11	14.9	14,661	13,865	5.4	6.7
	57	131	18	13.7	22,817	23,099	-1.2	7.9
	60	166	25	15.1	32,109	31,726	1.2	4.3
	61	243	44	18.1	44,160	43,271	2.0	5.4
	62	197	30	15.2	38,138	38,036	0.3	4.7
	전체	2,869	421	14.7	474,160	485,647	-2.4	1.8

③ 신안군

- 전형적인 시골 지역인 신안군의 추정 결과가 <표 2-20>에 나와 있다. 신안군의 추출률은 26.7%인데, 추정 차이율은 -30.3%로서 매우 크지만 RSE는 3.2%로 적다. 이는 신안군의 조사인구와 행정인구 차이율 -31.7%와 유사하며, 행정자료에 기반하여 설계된 표본의 추정치가 조사인구가 아닌 행정인구를 반영하고 있음을 나타낸다. 한편 읍면별 차이율을 보면, 모든 곳에서 모집단 인구가 추정인구보다 적은 것을 알 수 있다. 차이율의 절대치도 11.6%에서 53.1% 사이이다. 읍면별 RSE는 7.7%에서 23.3% 사이로 역시 다소 큰 편이다. 신안군의 경우, 읍면별 차이율과 RSE 모두 편차가 있으므로 동일한 추출률을 적용하기보다 읍면별로 차등화 시키는 것이 더 효율적일 것으로 판단된다.

<표 2-20> 신안군 읍면동별 인구 추정

지역	동	N	n	f	모집단 인구	추정 인구	차이율 (%)	RSE (%)
신 안 군	11	43	11	25.6	4,109	4,585	-11.6	12.0
	31	18	4	22.2	1,671	2,359	-41.2	16.1
	32	30	8	26.7	3,030	3,434	-13.3	12.4
	33	21	6	28.6	1,826	2,566	-40.5	10.0
	34	29	8	27.6	3,099	3,917	-26.4	9.0
	35	24	7	29.2	2,423	3,374	-39.2	7.7
	36	34	9	26.5	2,930	4,393	-49.9	8.9
	37	17	4	23.5	1,275	1,565	-22.7	14.1
	38	13	4	30.8	1,347	1,974	-46.5	13.9
	39	14	4	28.6	1,287	1,971	-53.1	12.5
	40	32	8	25.0	2,585	3,371	-30.4	10.2
	41	12	3	25.0	936	1,204	-28.6	23.3
	42	23	7	30.4	1,745	2,051	-17.5	10.7
	43	46	12	26.1	4,724	6,230	-31.9	8.3
	전체	356	95	26.7	32,987	42,991	-30.3	3.2

(2) 가구수

① 금천구

- <표 2-21>을 보면, 금천구의 가구수 차이율 -4.5%, RSE 1.5%로 가구수 역시 인구와 마찬가지로 추정의 정확성이 높은 편이다. 금천구에는 11개의 동이 있는데, 동별 차이율 절대치는 0.7%에서 10.4% 사이인데, 인구에 비해 동별 차이율 편차가 큰 편이다. 한편, RSE는 51 번동만 9.6%로 높은 편이고 나머지 10개 동은 모두 5% 미만으로 낮다.

<표 2-21> 금천구 읍면동별 가구수 추정

지역	동	N	n	f	모집단 가구수	추정 가구수	차이율 (%)	RSE (%)
금 천 구	51	147	32	21.8	8,689	8,053	7.3	9.6
	52	200	43	21.5	12,299	12,175	1.0	4.4
	53	137	29	21.2	8,418	9,252	-9.9	3.1
	54	191	42	22.0	10,531	11,140	-5.8	3.1
	55	113	24	21.2	6,868	7,503	-9.2	3.1
	57	186	40	21.5	10,986	11,732	-6.8	3.2
	58	128	25	19.5	7,950	7,896	0.7	2.4
	59	67	14	20.9	4,177	4,403	-5.4	4.2
	60	129	28	21.7	8,211	8,969	-9.2	2.3
	61	133	27	20.3	8,404	9,276	-10.4	3.6
	전체	1,431	304	21.2	86,533	90,397	-4.5	1.5

② 화성시

- 화성시의 추정 결과가 <표 2-22>에 나와 있다. 화성시의 가구수 추정 차이율 -5.3%, RSE 1.3%로서 시 전체로는 추정의 정확성이 높은 지역이다. 읍면동별로 보면 앞의 인구와 마찬가지로 차이율과 RSE 공통적으로 값이 적은 지역과 큰 지역이 뚜렷이 구분되는 것을 알 수 있다. 인구에서와 마찬가지로 도시의 특성을 지니는 읍면동의 경우 차이율과 RSE 모두 적은 편이나 시골 특성

의 면지역에서는 차이율과 RSE가 커진다. 그러므로 지역 내 읍면동의 특성을 무시한 채 모든 읍면동에 동일한 추출률을 적용하는 것은 바람직하지 않다.

<표 2-22> 화성시 읍면동별 가구수 추정

지역	동	N	n	f	모집단 인구	추정 인구	차이율 (%)	RSE (%)
화 성 시	12	354	53	15.0	20,426	20,375	0.3	2.8
	13	103	15	14.6	5,164	6,348	-22.9	7.5
	14	294	46	15.6	17,742	17,354	2.2	4.5
	31	44	6	13.6	2,376	2,319	2.4	7.8
	33	32	5	15.6	1,827	2,576	-41.0	14.1
	35	34	4	11.8	1,591	2,001	-25.8	21.1
	36	64	9	14.1	3,244	4,171	-28.6	9.1
	37	34	5	14.7	1,766	2,846	-61.2	11.9
	38	53	8	15.1	3,175	4,644	-46.3	10.2
	39	59	9	15.3	3,234	3,941	-21.9	9.4
	42	28	4	14.3	1,208	1,927	-59.5	14.0
	43	70	10	14.3	3,943	4,806	-21.9	9.0
	44	59	6	10.2	1,414	2,021	-42.9	34.5
	51	124	18	14.5	7,262	8,119	-11.8	5.1
	52	239	38	15.9	14,133	12,921	8.6	3.5
	53	140	20	14.3	8,401	8,389	0.1	3.3
	54	135	23	17.0	8,405	8,538	-1.6	2.8
	55	87	14	16.1	5,545	6,059	-9.3	3.3
	56	72	11	15.3	4,284	4,313	-0.7	5.4
	57	125	18	14.4	7,516	8,038	-6.9	4.8
	60	163	25	15.3	9,697	9,711	-0.1	2.7
	61	240	44	18.3	13,515	13,272	1.8	4.6
	62	195	30	15.4	12,555	12,067	3.9	3.9
	전 체	2,748	421	15.3	158,423	166,754	-5.3	1.3

③ 신안군

- 전형적인 시골 지역인 신안군의 추정 결과가 <표 2-23>에 나와 있다. 신안군의 가구수 추정 차이율은 -17.9%로서 매우 크지만 RSE는 2.1%로 적다. 읍면별 차이율을 보면, 42번 면을 제외한 모든 곳에서 모집단 가구수가 추정가구수보다 적다. 차이율의 절대치는 4%에서 40.4% 범위이다. 읍면별 RSE는 41번 지역만 17%로 크고 나머지는 10% 미만이다. 신안군의 가구수 추정을 위해서도 읍면별로 동일한 추출률을 적용하기보다 차등화 시키는 것이 더 효율적이다.

<표 2-23> 신안군 읍면동별 가구수 추정

지역	동	N	n	f	모집단 가구수	추정 가구수	차이율 (%)	RSE (%)
신 안 군	11	39	11	28.2	2,095	2,181	-4.1	7.5
	31	15	4	26.7	789	1,043	-32.2	7.6
	32	28	8	28.6	1,266	1,407	-11.1	9.8
	33	20	6	30.0	974	1,229	-26.1	6.0
	34	27	8	29.6	1,524	1,764	-15.7	4.7
	35	22	7	31.8	1,163	1,498	-28.8	5.5
	36	33	9	27.3	1,405	1,747	-24.3	7.6
	37	15	4	26.7	680	728	-7.1	7.5
	38	12	4	33.3	617	837	-35.6	8.3
	39	13	4	30.8	703	987	-40.4	6.1
	40	30	8	26.7	1,350	1,554	-15.1	7.5
	41	11	3	27.3	522	613	-17.3	17.0
	42	22	7	31.8	952	914	4.0	7.1
	43	41	12	29.3	2,074	2,503	-20.7	4.9
	전체	328	95	29.0	16,114	19,002	-17.9	2.1

(3) 주택수

① 금천구

- <표 2-24>를 보면, 금천구의 주택수 차이율 5.0%, RSE 3.3%로 가구수 역시 인구와 마찬가지로 추정의 정확성이 비교적 높은 편이다. 금천구에는 11개의동이 있는데, 동별 차이율 절대치는 51번 동을 제외하고는 모두 10% 미만이다. RSE도 51번 동 이외에는 모두 10% 미만으로 낮다.

<표 2-24> 금천구 읍면동별 주택수 추정

지역	동	N	n	f	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
금 천 구	51	157	32	20.4	4,736	4,077	13.9	12.7
	52	211	43	20.4	8,711	8,313	4.6	9.2
	53	141	29	20.6	3,510	3,440	2.0	8.6
	54	195	42	21.5	5,289	5,088	3.8	8.6
	55	116	24	20.7	3,082	3,225	-4.6	9.8
	57	196	40	20.4	6,616	6,267	5.3	8.9
	58	133	25	18.8	7,804	7,335	6.0	5.3
	59	70	14	20.0	3,622	3,742	-3.3	9.2
	60	130	28	21.5	4,711	4,279	9.2	9.2
	61	134	27	20.1	4,239	3,948	6.9	9.5
	전체	1,483	304	20.5	52,320	49,713	5.0	3.3

② 화성시

- 화성시 주택수 추정 결과가 <표 2-25>에 나와 있다. 화성시 주택수 추정 차이율 -2.1%, RSE 1.8%로 시 전체로는 추정의 정확성이 높은 지역이다. 읍면동별로 보면 앞의 인구나 가구수와 마찬가지로 차이율과 RSE 공통적으로 값이 적은 지역과 큰 지역이 뚜렷이 구분되는 것을 알 수 있다. 인구수에서와 마찬

가지로 도시의 특성을 지니는 읍면동의 경우 차이율과 RSE 모두 적은 편이나 시골 특성의 면지역에서는 차이율과 RSE가 커진다. 그러므로 지역 내 읍면동의 특성을 무시한 채 모든 읍면동에 동일한 추출률을 적용하는 것은 바람직하지 않다.

<표 2-25> 화성시 읍면동별 주택수 추정

지역	동	N	n	f	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
화 성 시	12	367	53	14.4	19,751	19,874	-0.6	4.8
	13	116	15	12.9	5,174	5,651	-9.2	13.0
	14	298	46	15.4	16,359	17,246	-5.4	5.5
	31	45	6	13.3	2,357	2,427	-3.0	10.8
	33	33	5	15.2	1,866	2,096	-12.3	16.7
	35	34	4	11.8	1,685	1,785	-5.9	21.6
	36	70	9	12.9	3,604	3,792	-5.2	12.3
	37	38	5	13.2	1,913	2,339	-22.3	18.8
	38	57	8	14.0	3,226	3,840	-19.0	14.2
	39	66	9	13.6	3,526	3,265	7.4	14.0
	42	31	4	12.9	1,338	1,345	-0.5	21.1
	43	82	10	12.2	3,986	4,617	-15.8	17.6
	44	66	6	9.1	2,184	1,582	27.6	25.0
	51	131	18	13.7	7,465	7,848	-5.1	7.5
	52	242	38	15.7	9,273	9,350	-0.8	11.4
	53	141	20	14.2	8,226	8,423	-2.4	4.7
	54	136	23	16.9	8,126	8,440	-3.9	5.2
	55	91	14	15.4	5,575	5,870	-5.3	6.4
	56	73	11	15.1	4,350	4,259	2.1	6.1
	57	130	18	13.8	7,322	7,294	0.4	7.6
	60	165	25	15.2	9,724	9,442	2.9	3.8
	61	242	44	18.2	14,265	13,990	1.9	3.8
	62	196	30	15.3	12,494	12,181	2.5	3.4
	전체	2,850	421	14.8	153,789	156,956	-2.1	1.8

③ 신안군

- 전형적인 시골 지역인 신안군의 추정 결과가 <표 2-26>에 나와 있다. 신안군의 주택수 추정 차이율은 4.2%, RSE는 2.8%로 둘다 적은 편이다. 주목할 점은 차이율이다. 인구와 가구수의 경우 추정치가 모수보다 컸는데, 주택수는 그 반대이다. 또한 인구와 가구수 차이율 절대치가 30%를 넘었는데 비해 주택수는 4.2%에 불과하다. 읍면별 차이율과 RSE가 10%를 넘는 지역이 있기는 하지만 인구나 가구수보다는 많이 양호하다. 인구나 가구수에서는 행정자료와 조사자료 사이에 차이가 많이 나지만 주택수는 상대적으로 행정자료와 조사자료 사이에 차이가 적다. 신안군에서도 읍면별로 추출률을 다르게 하는 것이 필요할 것으로 판단된다.

<표 2-26> 신안군 읍면동별 주택수 추정

지역	동	N	n	f	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
신 안 군	11	42	11	26.2	2,280	1,883	17.4	10.7
	31	17	4	23.5	836	851	-1.7	16.6
	32	29	8	27.6	1,357	1,196	11.9	11.6
	33	20	6	30.0	1,103	1,166	-5.7	5.4
	34	28	8	28.6	1,631	1,684	-3.2	7.6
	35	24	7	29.2	1,281	1,278	0.3	9.9
	36	33	9	27.3	1,593	1,428	10.4	7.6
	37	16	4	25.0	798	718	10.1	13.7
	38	12	4	33.3	652	690	-5.8	7.1
	39	13	4	30.8	798	942	-18.0	5.9
	40	31	8	25.8	1,527	1,390	9.0	9.0
	41	11	3	27.3	571	595	-4.2	17.8
	42	22	7	31.8	1,108	893	19.4	5.3
	43	45	12	26.7	2,227	2,310	-3.7	9.6
	전체	343	95	27.7	17,762	17,020	4.2	2.8

C. 소결론

- 먼저 인구, 가구수, 주택수에 대한 시군구별 분석 결과를 요약하면 다음과 같다.
 - 인구, 가구수에 대한 시군구별 추정 결과에 의하면, 도시지역의 모수와 추정치 차이율은 크지 않으나 행정자료와 조사자료의 차이가 크게 나는 농촌 지역에서는 10% 이상 큰 차이가 난다. 이는 표본설계에서 추출틀로 사용한 행정자료와 조사자료 간의 차이 때문에 생기는 것이다.
 - 인구와 가구수에 비해 주택수에 대한 추정의 차이율은 낮다. 이는 인구나 가구 데이터베이스와 주택 데이터베이스가 별개의 자료이기 때문이다. 주택수 추정에서 차이가 심하게 나는 시군구는 양구군 하나 밖에 없다.
 - 시군구별 RSE는 인구, 가구수, 주택수 모두 표본설계 당시에 정한 목표오차가 안정적으로 달성되고 있다.
 - 행정인구와 조사인구의 차이가 큰 지역의 경우, 주민행정자료 대신 주택자료를 추출틀로 사용하면 인구나 가구수 차이율의 문제를 해결할 수 있을 것으로 보인다. 향후, 지역적 특성에 따라 주민등록자료 또는 주택자료를 추출틀로 선택하는 표본설계 연구가 필요할 것이다.
- 다음으로 인구, 가구수, 주택수에 대한 읍면동별 분석 결과를 요약하면 다음과 같다.
 - 동일한 시군구 내의 읍면동별 추정 차이율과 RSE 모두 편차가 큰 편이다. 다시 말해, 어느 지역의 전체 차이율이 낮다고 해도 읍면동에 따라서는 차이율이 매우 큰 곳이 존재할 수 있다.

- RSE 측면에서도 읍면동별로 균일하지 않다. 이는 본 연구의 표본설계에서 시군구 수준에서만 목표오차를 통제하고 같은 지역에 속한 읍면동에 대해서는 동일한 추출률을 적용하였기 때문에 나타난 결과이다.
- 같은 지역이라고 해도 읍면동별 추정의 정확성이 매우 다르므로 본 표본설계에서와 같이 시군구별 목표오차를 통제할 것이 아니라 읍면동별 목표오차를 통제한다면 보다 효과적일 것으로 판단된다.

제5절 결론 및 제안사항

1. 결 론

- 등록센서스와 병행하여 실시할 표본조사를 위한 표본설계에 관한 연구는 여러 해 계속 진행되어 왔다. 2009년의 등록센서스 제1차 학술연구 용역사업(한국 인구학회, 2009)에서는 기존의 10% 표본조사를 검토하여 향후 표본조사를 위한 1차 추출단위의 결정과 합리적인 표본추출률 결정을 위한 본격적인 연구의 필요성을 제안하였다. 2010년 통계청의 등록센서스 제2차 학술연구 용역사업에는 2005년 총조사의 전수조사와 표본조사 결과 비교를 통해 10% 내외에 이르는 기존의 표본크기가 불충분하며 대략 21% 정도의 추출률이 필요하다고 제안하였다(한국인구학회, 2010). 2011년에는 통계청의 제3차 등록센서스 학술연구 용역사업에서는, 2009년 시범예행조사 자료를 기초로 하여 대전시 유성구와 충남 홍성군에 대해 등록센서스 병행 표본조사를 시도하였으며, 그에 따른 각종 통계적 추정방법에 대한 연구를 진행하였다(한국인구학회, 2011).
- 과거의 축적된 연구 결과를 토대로 하여 이번의 통계청 제4차 등록센서스 학술연구 용역사업에서는 12개 지역의 조사구별 2010년 행정자료와 2010년 인구주택총조사 자료를 바탕으로 하여 표본설계를 실시하였다. 2011년에는 조사구 특성을 고려하지 않은 상태로 표본설계를 실시한 결과, 이상치 조사구로 인해 추출률에 비해 만족스럽지 않은 효율을 얻은 바 있다. 이에 따라 본 연구에서는 조사구 특성을 미리 검토하였고, 기숙시설, 특수사회시설, 특수조사구를 전수조사구(full-enumeration district)로 전환시켰다. 그리고 이상치 조사구 중 인구가 많은 조사구만 전수조사 대상으로 하고 인구가 적은 이상치는 제외시킴으로 추출률을 낮추면서도 통계적 효율성(조사비용 대비)을 높이는 방안을 마련하였다.
- 기본적으로 행정자료를 추출틀로 사용하는데, 지역별 행정자료와 조사자료의

차이 정도에 따라서 추정치와 모수의 차이가 달라지는 것으로 나타났다. 도시 지역의 경우 행정자료 기반 표본조사 차이율이 낮은 반면, 행정자료와 조사자료 차이가 큰 군 지역에서는 인구나 가구수 차이에 비해 주택수 차이가 상대적으로 작았다. 즉, 주민등록자료의 경우 의도적인 미신고나 실제 사실과 다른 신고의 개연성이 있지만 상대적으로 주택 관련 신고의 경우 그런 영향이 적은 편이다.

- 본 연구에서는 2010년 10% 표본조사와 이전의 연구에서와 같이 시군구 단위로 목표오차(target error)를 지정하여 표본크기를 결정하는 방식을 사용하였다. 이렇게 했을 때 시군구 단위의 추정에는 효과적이었으나 읍면동 단위에서는 정확성의 편차가 커서 부적절한 것으로 나타났다. 같은 시군구 안에서도 읍면동별 특성이 많이 다르므로 시군구 표본크기를 정한 후 동일 지역에서 읍면동별 추출률을 같게 하는 방안보다는 처음부터 읍면동 수준에서 표본크기를 결정하는 방식이 필요하다는 것을 알 수 있었다.
- 표본추출을 위한 조사구 정렬변수로 인구를 사용하는 방안, 인구수와 주택수를 동시에 고려하는 방안, 인구수와 가구수와 주택수 모두를 고려하는 방안을 각각 적용하여 비교하였으나 다른 방안을 압도하는 것이 없었다. 기본적으로 세 방안들의 효율이 크게 다르지 않기 때문에 적용하기 간편한 인구수만 고려하는 방안을 적용하기로 했다.
- 표본설계의 적절성과 상관없이 행정자료와 조사자료의 근원적인 차이로 인해 생기는 추정 단계에서의 문제를 해결하기 위해 직접추정량(direct estimator)이 아닌 추가적인 추정법 연구가 반드시 필요하다.

2. 제언사항

- 본 표본설계 연구는 2009년부터 2011년까지의 기존 연구의 결과를 토대로 진

행되었다. 즉, 추출단위를 조사구로 하고, 시군구 수준에서 표본크기를 결정하였다. 그런데 본 연구를 수행하는 과정에서 파악하게 된 몇 가지 개선 요망사항들이 있다. 추후 연구를 위해 이러한 제언사항들을 간단하게 제시하고자 한다.

- 조사구를 기본 단위로 하는 추출틀 작성 시 조사구별 인구, 가구수, 주택수 정보가 주어진다. 본 연구에서 가장 중요한 조사변수는 인구라고 여겨서 표본설계를 할 때 인구는 반드시 고려하였다. 그런데 행정자료와 조사자료의 차이가 많이 나는 군 지역에서도 인구를 중심으로 표본추출을 한 결과, 추정치와 모수 사이에 무시할 수 없는 차이가 발생하였다. 향후 연구에서는 행정자료와 조사자료 차이의 일정한 기준을 정하여 그 기준을 넘어서는 지역에 대해서는 인구 변수가 아닌 주택수 변수로만 표본설계를 하는 것이 효과적일 것이라 예상된다. 다시 말해 두 자료 사이의 차이가 적은 지역에서는 주민등록자료를 추출틀로 활용하도록 하고, 차이가 많은 지역에서는 주택자료를 추출틀로 사용하여 표본설계를 하는 방안을 반드시 검토해 볼 것을 제안한다.
- 본 연구에서는 1차 추출단위로 2010년 인구주택총조사에서 작성한 조사구를 사용하였다. 지역별로 조사구 크기가 균일하지 않아서 이상치 조사구가 많은 편이었다. 조사구 내 인구나 가구수가 아주 적은 조사구가 있는가 하면 반대로 한 조사구 크기가 너무 큰 조사구도 있었다. 조사구 크기 결정에서 지리적, 사회적 여건 등이 감안되었으나 통계적 효율성 측면은 적절히 고려되지 못한 것으로 여겨진다. 참고로 조사구를 합리적으로 작성함으로써 얻을 수 있는 효과를 살펴보기 위해 2010년 지역별 조사구를 일부 조정해 보았다. 가령, 어느 시군구의 조사구 당 평균 인구가 100명이라고 할 때, 조사구 인구가 200명 이상인 조사구는 가상적으로 2개의 조사구로 분리하였다. 반대로 조사구 인구가 20명 미만인 조사구들은 서로 통합하여 인구가 대략 100명 가까운 하나의 조사구인 것처럼 조정하였다. 지역별로 조사구 조정 전·후의 조사구수와 변동계수를 비교한 결과가 <표 2-27>에 나와 있다.
- <표 2-27>을 보면, 모든 지역에서 변동계수가 조정 후에 줄어든 것을 관찰할 수 있다. 동일한 추출틀을 적용한다면 조정 후의 표본오차가 조정 전보다 감

소하게 되어 보다 효율을 높일 수 있다. 이는 조사구를 설정할 때 통계적 효율을 고려하여 가급적 조사구 크기를 균일하게 작성하는 것이 효과적이라는 것을 시사한다. 그러므로 추후 조사구를 설정할 때 조사구 크기를 가능한 한 균일하게 할 것을 제안한다. 이 밖에도 한 조사구의 크기를 얼마로 하는 것이 가장 효과적인지에 대해서도 추가적인 연구를 실시하여 최적의 크기로 결정할 것을 제안한다.

<표 2-27> 지역별 조사구 조정 효과

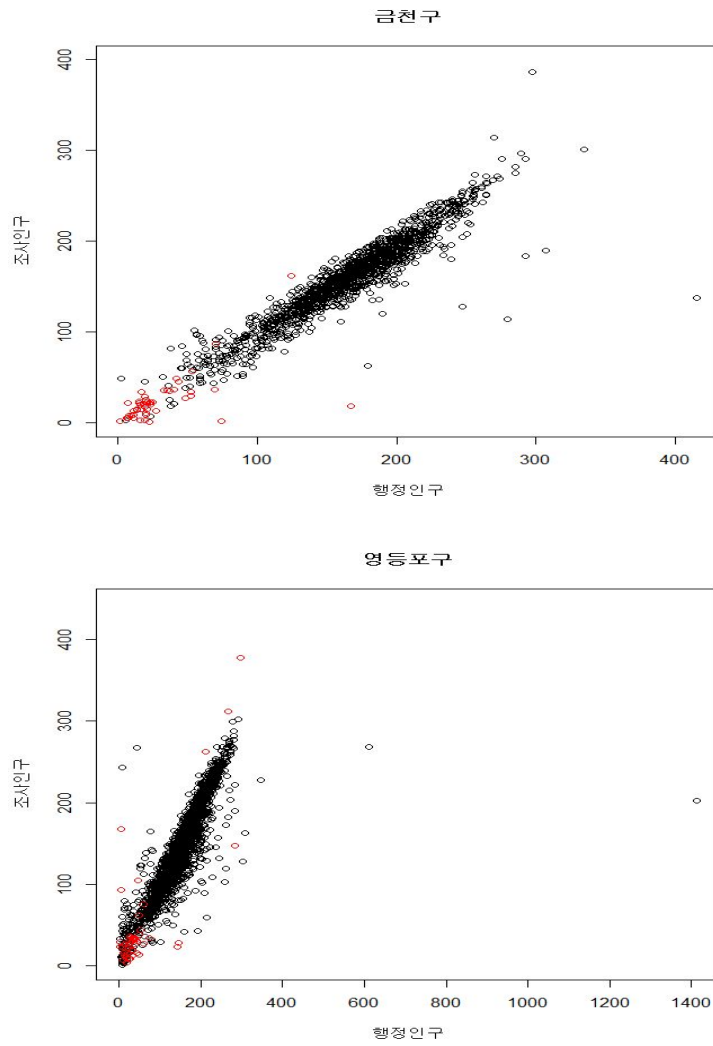
지역	조정 전		조정 후	
	조사구 수	변동계수	조사구 수	변동계수
광주 동구	799	0.48	759	0.33
금천구	1492	0.36	1580	0.26
김천시	890	0.43	871	0.26
나주시	660	0.47	646	0.28
신안군	357	0.44	352	0.31
양구군	148	0.6	148	0.3
영등포구	2597	0.4	2597	0.28
울릉군	67	0.87	72	0.33
인천 서구	2250	0.38	2337	0.23
임실군	219	0.45	227	0.3
증평군	229	0.44	228	0.25
화성시	2870	0.43	2801	0.27

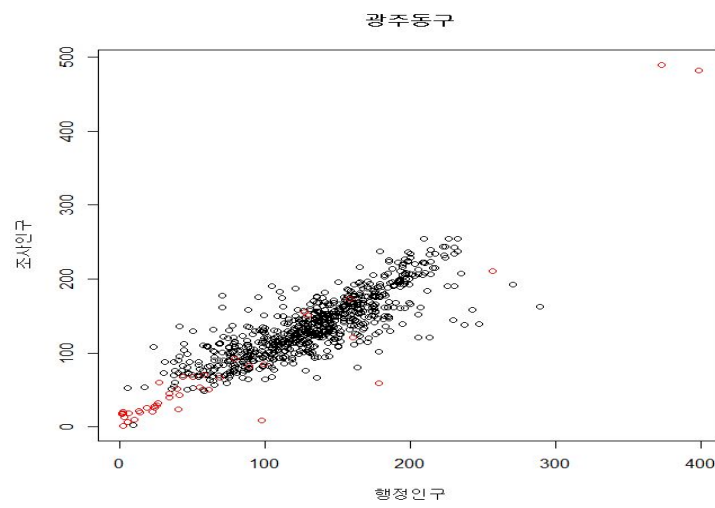
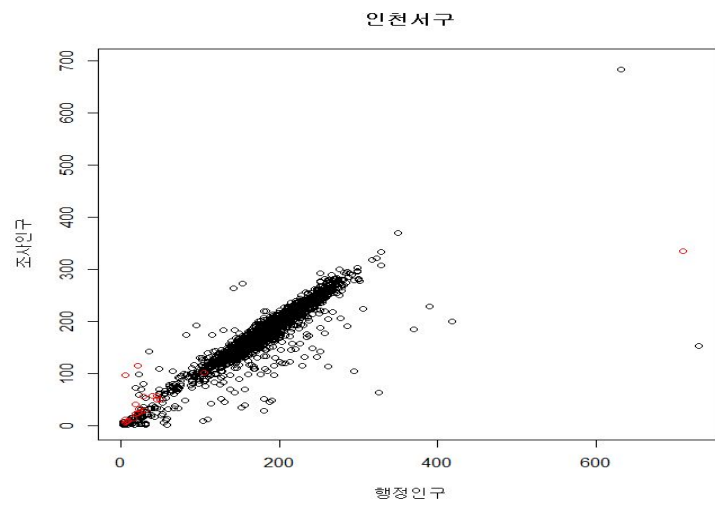
- 본 표본설계에서 표본크기 결정을 위한 최소 지역단위는 시군구이다. 먼저 시군구별 목표오차를 지정하여 그에 따른 시군구별 표본추출률을 결정한다. 그러나 실제 최소 통계작성단위는 읍면동이 된다. 동일 시군구에 속한 읍면동의 특성이 서로 비슷하다면 이 방법은 효과적이다. 그러나 분석 결과에 의하면,

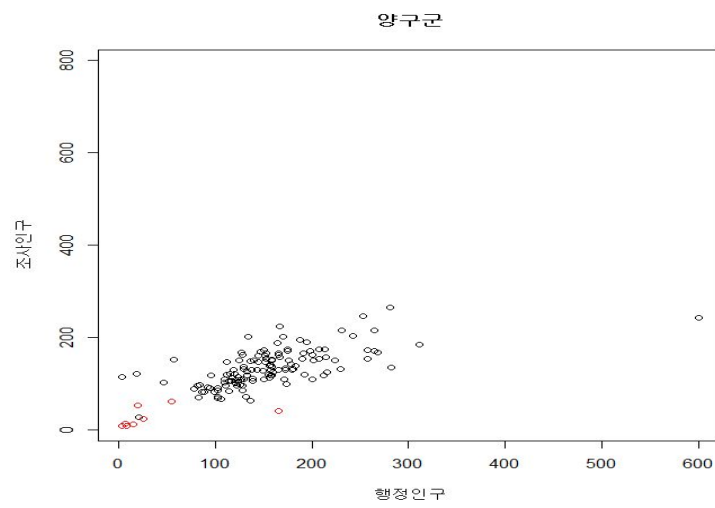
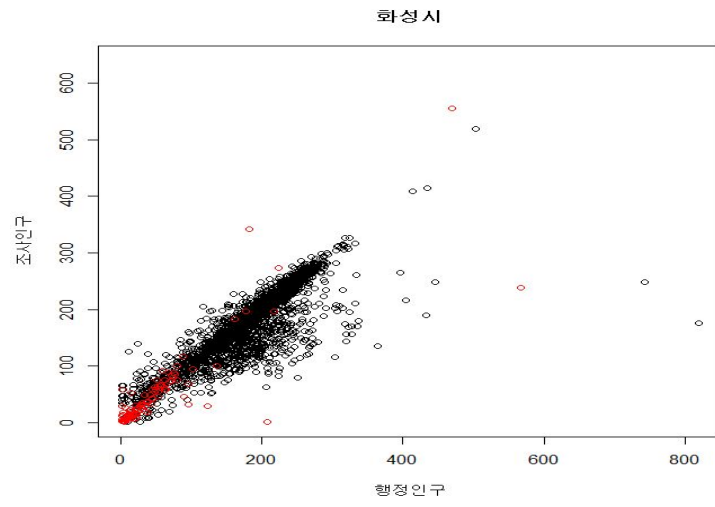
동일 지역 내 읍면동이라고 해도 차이율이나 상대표준오차의 편차가 크다는 것을 알 수 있었다. 이는 시군구 단위에서 표본크기를 결정하는 것이 효과적이지 않다는 의미이다. 따라서 추후에는 표본크기 결정을 위한 지역단위를 읍면동으로 하는 방안을 검토할 것을 제안한다. 이 경우 표본크기 결정 단위가 10배 이상 많아지게 되므로 현실적으로 관리가 가능할 것인지도 함께 충분히 고려해야 할 것이다.

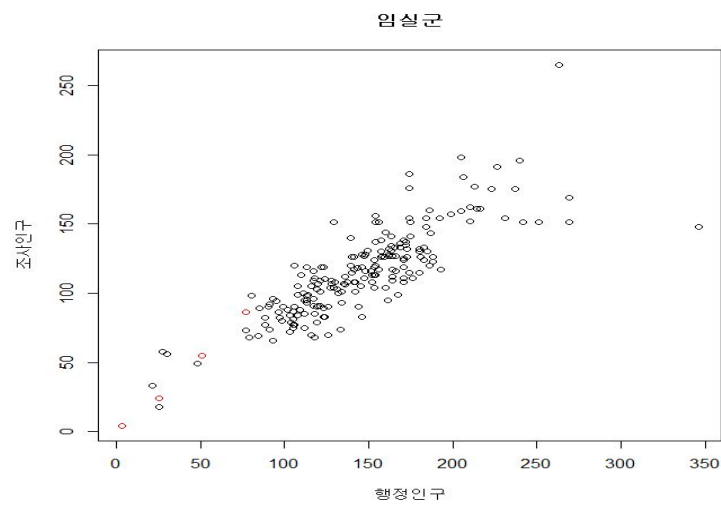
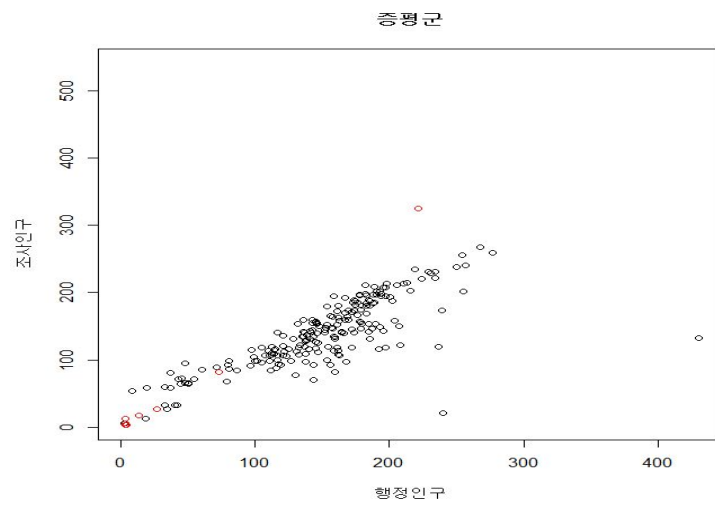
[부 록 2]

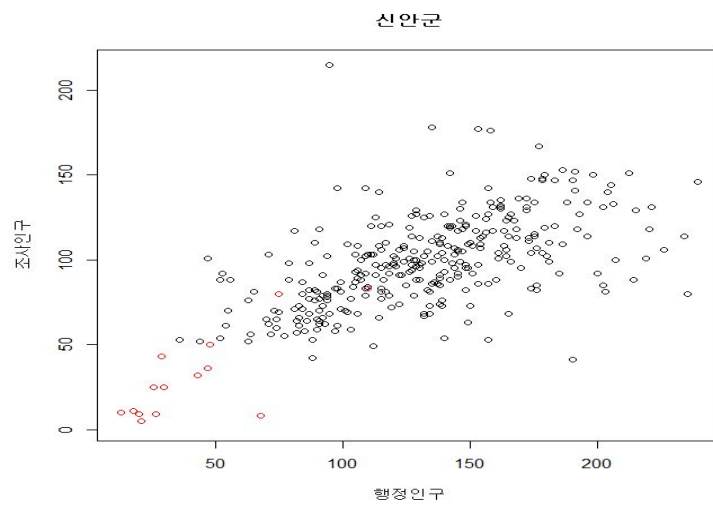
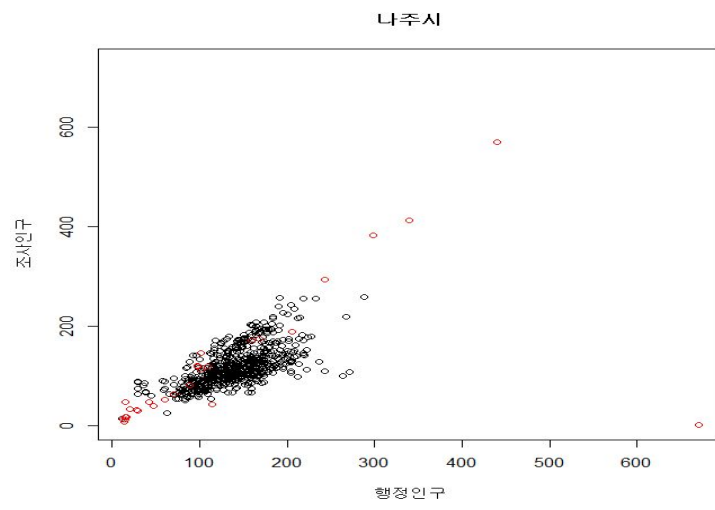
2-1. 지역별 행정인구와 조사인구의 산점도

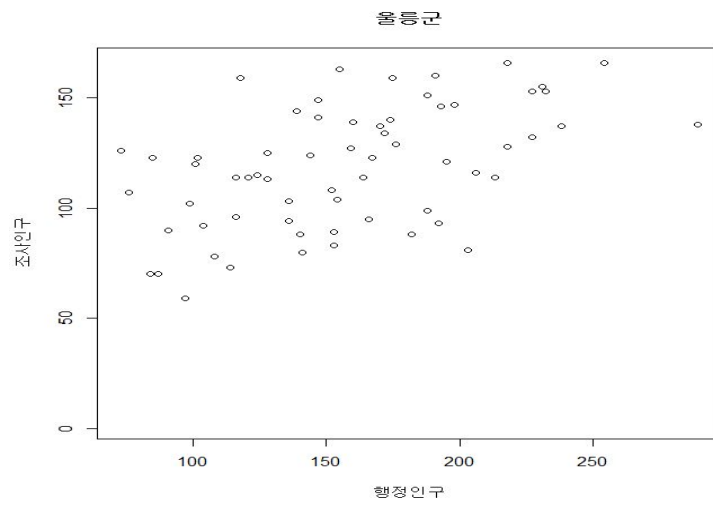
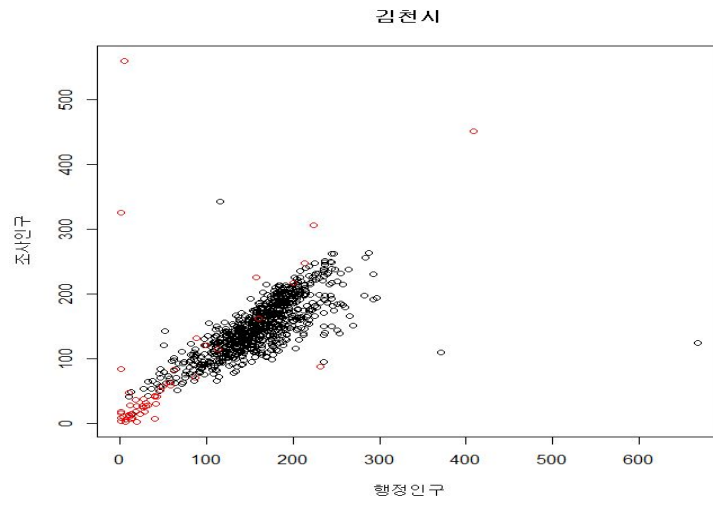












2-2. 행정자료와 조사자료의 차이

(1) 읍면동별 인구

시군구	읍면동	조사인구	행정인구	차이율
인천서구	51	29,074	29,325	-0.9
	53	47,399	46,864	1.1
	54	10,993	11,642	-5.9
	55	9,076	9,359	-3.1
	56	12,451	12,344	0.9
	58	24,381	24,646	-1.1
	59	16,566	17,379	-4.9
	60	17,933	18,116	-1.0
	62	13,566	14,086	-3.8
	63	23,186	23,390	-0.9
	64	20,569	20,951	-1.9
	65	12,294	12,506	-1.7
	68	39,170	40,119	-2.4
	69	23,113	23,704	-2.6
	70	22,431	21,813	2.8
	71	36,117	35,290	2.3
	72	3,475	3,401	2.1
광주동구	73	27,263	27,443	-0.7
	51	4,625	4,624	0.0
	54	6,153	5,594	9.1
	56	8,050	7,770	3.5
	58	9,428	9,038	4.1
	59	8,877	8,765	1.3
	61	11,128	10,538	5.3
	62	5,941	5,222	12.1
	63	6,509	6,026	7.4
	64	5,639	4,895	13.2
	68	6,676	6,879	-3.0
	71	13,598	13,124	3.5
	72	9,246	8,895	3.8
	73	8,128	8,118	0.1

시군구	읍면동	조사인구	행정인구	차이율
임실군	11	5,352	6,521	-21.8
	31	1,190	1,474	-23.9
	32	1,263	1,853	-46.7
	33	1,346	1,705	-26.7
	34	1,370	1,754	-28.0
	35	3,750	4,575	-22.0
	36	1,123	1,401	-24.8
	37	1,425	1,646	-15.5
	38	3,021	3,774	-24.9
	39	1,504	1,764	-17.3
	40	1,016	1,291	-27.1
나주시	41	1,130	1,425	-26.1
	11	7,142	8,105	-13.5
	31	2,597	3,426	-31.9
	32	2,969	3,774	-27.1
	33	1,386	1,793	-29.4
	34	2,328	2,914	-25.2
	35	2,617	3,222	-23.1
	36	3,632	4,402	-21.2
	37	1,858	2,331	-25.5
	38	4,350	5,457	-25.4
	39	4,671	5,701	-22.1
	40	3,081	4,283	-39.0
	41	1,845	2,220	-20.3
	42	3,862	5,167	-33.8
	51	7,546	7,120	5.6
	52	4,179	3,976	4.9
	54	6,411	7,120	-11.1
	55	10,152	9,989	1.6
	58	4,143	4,676	-12.9
	60	3,056	3,956	-29.5

시군구	읍면동	조사인구	행정인구	차이율
신안군	11	4,109	4,935	-20.1
	31	1,671	2,089	-25.0
	32	3,030	3,543	-16.9
	33	1,826	2,340	-28.1
	34	3,099	3,797	-22.5
	35	2,423	3,034	-25.2
	36	2,930	4,526	-54.5
	37	1,275	2,009	-57.6
	38	1,347	1,767	-31.2
	39	1,287	1,744	-35.5
	40	2,585	3,489	-35.0
	41	936	1,183	-26.4
	42	1,745	2,173	-24.5
	43	4,724	6,805	-44.1
김천시	11	7,403	8,303	-12.2
	31	2,809	3,223	-14.7
	32	2,932	3,533	-20.5
	33	2,539	2,856	-12.5
	34	3,548	3,876	-9.2
	35	5,105	5,422	-6.2
	36	2,996	3,525	-17.7
	37	3,757	4,128	-9.9
	38	1,773	2,237	-26.2
	39	2,044	2,562	-25.3
	40	2,770	3,262	-17.8
	41	1,673	1,881	-12.4
	42	1,142	1,307	-14.4
	43	1,886	2,245	-19.0
	44	990	1,175	-18.7
	55	5,335	5,493	-3.0
	56	26,291	25,976	1.2
	58	22,971	22,686	1.2
	59	10,225	10,033	1.9
	60	8,661	9,273	-7.1
	61	9,973	10,527	-5.6
울릉군	11	5,579	7,195	-29.0
	31	1,078	1,543	-43.1
	32	1,080	1,373	-27.1

(2) 읍면동별 가구수

시군구	읍면동	조사가구	행정가구	차이율
금천구	51	8,689	8,537	1.7
	52	12,299	12,359	-0.5
	53	8,418	9,076	-7.8
	54	10,531	11,370	-8.0
	55	6,868	7,631	-11.1
	57	10,986	12,001	-9.2
	58	7,950	7,950	0.0
	59	4,177	4,302	-3.0
	60	8,211	8,917	-8.6
	61	8,404	9,114	-8.4
영등포구	54	10,789	10,547	2.2
	55	7,411	8,096	-9.2
	56	11,629	11,141	4.2
	61	6,101	6,168	-1.1
	62	6,857	6,984	-1.9
	63	8,881	8,908	-0.3
	65	6,794	7,301	-7.5
	66	4,557	5,105	-12.0
	67	6,056	6,709	-10.8
	68	8,542	9,008	-5.5
	69	6,038	6,007	0.5
	70	5,750	6,543	-13.8
	71	6,505	7,115	-9.4
	72	9,437	10,711	-13.5
	73	9,028	9,762	-8.1
	74	6,582	8,087	-22.9
	75	7,106	7,792	-9.7
	76	10,574	10,542	0.3

시군구	읍면동	조사가구	행정가구	차이율
인천서구	51	9,307	9,577	-2.9
	53	15,955	15,513	2.8
	54	4,283	4,682	-9.3
	55	3,093	3,203	-3.6
	56	4,156	4,044	2.7
	58	8,837	9,094	-2.9
	59	5,894	6,390	-8.4
	60	6,237	6,316	-1.3
	62	4,999	5,297	-6.0
	63	7,156	7,251	-1.3
	64	7,115	7,220	-1.5
	65	4,474	4,711	-5.3
	68	12,684	13,524	-6.6
	69	7,089	7,680	-8.3
	70	6,963	6,982	-0.3
	71	10,604	10,451	1.4
	72	1,084	1,087	-0.3
	73	9,045	9,209	-1.8
광주동구	51	2,020	2,433	-20.4
	54	2,946	2,419	17.9
	56	3,944	3,750	4.9
	58	3,445	3,276	4.9
	59	3,758	3,719	1.0
	61	4,369	3,929	10.1
	62	2,714	2,140	21.1
	63	3,012	2,405	20.2
	64	2,768	2,096	24.3
	68	3,067	3,096	-0.9
	71	4,694	4,502	4.1
	72	3,217	3,132	2.6
	73	2,584	2,712	-5.0

시군구	읍면동	조사가구	행정가구	차이율
화성시	12	20,426	20,494	-0.3
	13	5,164	6,128	-18.7
	14	17,742	17,830	-0.5
	31	2,376	2,730	-14.9
	33	1,827	2,359	-29.1
	35	1,591	2,254	-41.7
	36	3,244	4,165	-28.4
	37	1,766	2,810	-59.1
	38	3,175	4,304	-35.6
	39	3,234	4,124	-27.5
	42	1,208	1,715	-42.0
	43	3,943	4,882	-23.8
	44	1,414	2,049	-44.9
	51	7,262	8,108	-11.6
	52	14,133	13,192	6.7
	53	8,401	8,358	0.5
	54	8,405	8,349	0.7
	55	5,545	5,603	-1.0
	56	4,284	4,443	-3.7
	57	7,516	7,764	-3.3
	60	9,697	9,615	0.8
	61	13,515	13,138	2.8
	62	12,555	11,992	4.5
양구군	11	3,975	4,607	-15.9
	31	1,244	1,480	-19.0
	32	713	890	-24.8
	33	500	613	-22.6
	34	487	523	-7.4
증평군	14	10,708	10,958	-2.3
	31	851	932	-9.5

시군구	읍면동	조사가구	행정가구	차이율
신안군	11	2,095	2,277	-8.7
	31	789	907	-15.0
	32	1,266	1,499	-18.4
	33	974	1,142	-17.2
	34	1,524	1,684	-10.5
	35	1,163	1,392	-19.7
	36	1,405	1,979	-40.9
	37	680	920	-35.3
	38	617	730	-18.3
	39	703	824	-17.2
	40	1,350	1,573	-16.5
	41	522	572	-9.6
	42	952	997	-4.7
	43	2,074	2,653	-27.9
김천시	11	2,984	3,191	-6.9
	31	1,095	1,217	-11.1
	32	1,145	1,369	-19.6
	33	1,048	1,095	-4.5
	34	1,592	1,610	-1.1
	35	1,772	1,905	-7.5
	36	1,221	1,349	-10.5
	37	1,387	1,541	-11.1
	38	767	854	-11.3
	39	915	1,011	-10.5
	40	1,422	1,499	-5.4
	41	800	856	-7.0
	42	590	615	-4.2
	43	990	1,053	-6.4
	44	485	540	-11.3
	55	2,092	2,128	-1.7
	56	8,695	8,434	3.0
	58	8,289	7,827	5.6
	59	3,748	3,552	5.2
	60	3,533	3,650	-3.3
울릉군	11	2,427	3,178	-30.9
	31	540	710	-31.5
	32	527	645	-22.4

시군구	읍면동	조사가구	행정가구	차이율
임실군	11	2,133	2,504	-17.4
	31	601	638	-6.2
	32	602	752	-24.9
	33	649	747	-15.1
	34	678	733	-8.1
	35	1,720	1,889	-9.8
	36	563	626	-11.2
	37	723	745	-3.0
	38	1,244	1,444	-16.1
	39	711	748	-5.2
	40	509	604	-18.7
나주시	41	560	616	-10.0
	11	2,670	3,243	-21.5
	31	1,237	1,498	-21.1
	32	1,347	1,609	-19.5
	33	716	856	-19.6
	34	1,138	1,305	-14.7
	35	1,329	1,542	-16.0
	36	1,699	1,996	-17.5
	37	975	1,155	-18.5
	38	1,918	2,357	-22.9
	39	1,706	2,143	-25.6
	40	1,255	1,745	-39.0
	41	840	970	-15.5
	42	1,892	2,314	-22.3
	51	2,693	2,528	6.1
	52	1,565	1,501	4.1
	54	2,445	2,832	-15.8
	55	3,454	3,527	-2.1
	58	1,663	1,905	-14.6
	60	1,253	1,652	-31.8

(3) 읍면동별 주택수

시군구	읍면동	조사주택	행정주택	차이율
금천구	51	4,736	4,131	12.8
	52	8,711	8,256	5.2
	53	3,510	3,558	-1.4
	54	5,289	5,122	3.2
	55	3,082	3,107	-0.8
	57	6,616	6,456	2.4
	58	7,804	7,249	7.1
	59	3,622	3,636	-0.4
	60	4,711	4,744	-0.7
	61	4,239	4,234	0.1
영등포구	54	11,412	10,869	4.8
	55	5,740	5,536	3.6
	56	10,647	9,894	7.1
	61	5,330	5,117	4.0
	62	6,119	6,169	-0.8
	63	4,239	3,952	6.8
	65	3,298	3,353	-1.7
	66	2,724	2,517	7.6
	67	2,993	2,974	0.6
	68	5,597	5,636	-0.7
	69	4,423	4,386	0.8
	70	2,780	2,802	-0.8
	71	3,475	3,419	1.6
	72	6,325	6,389	-1.0
	73	6,116	6,144	-0.5
	74	4,584	4,533	1.1
	75	4,320	4,297	0.5
	76	10,208	9,726	4.7

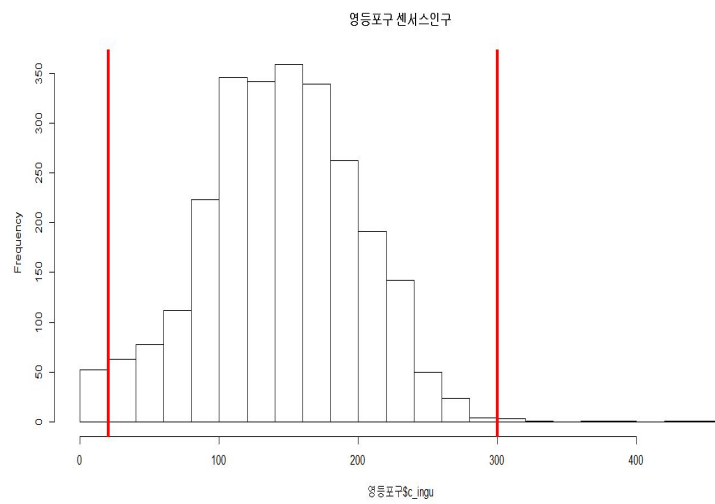
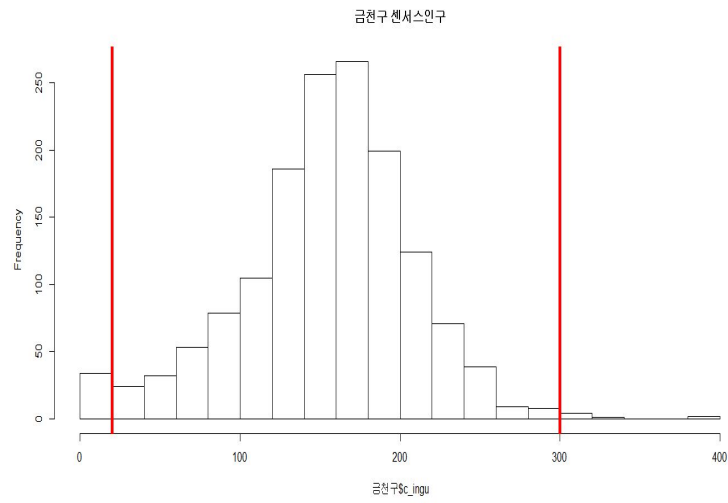
시군구	읍면동	조사주택	행정주택	차이율
화성시	12	19,751	20,058	-1.6
	13	5,174	5,218	-0.9
	14	16,359	16,512	-0.9
	31	2,357	2,464	-4.5
	33	1,866	1,974	-5.8
	35	1,685	1,824	-8.2
	36	3,604	4,012	-11.3
	37	1,913	2,174	-13.6
	38	3,226	3,518	-9.1
	39	3,526	3,426	2.8
	42	1,338	1,376	-2.8
	43	3,986	4,205	-5.5
	44	2,184	1,829	16.3
	51	7,465	7,717	-3.4
	52	9,273	9,444	-1.8
	53	8,226	8,267	-0.5
	54	8,126	8,156	-0.4
	55	5,575	5,640	-1.2
	56	4,350	4,386	-0.8
	57	7,322	7,248	1.0
	60	9,724	9,735	-0.1
	61	14,265	14,378	-0.8
	62	12,494	12,230	2.1
양구군	11	4,061	3,874	4.6
	31	1,336	1,314	1.6
	32	804	663	17.5
	33	587	508	13.5
	34	501	425	15.2
증평군	14	10,792	10,826	-0.3
	31	1,048	1,054	-0.6

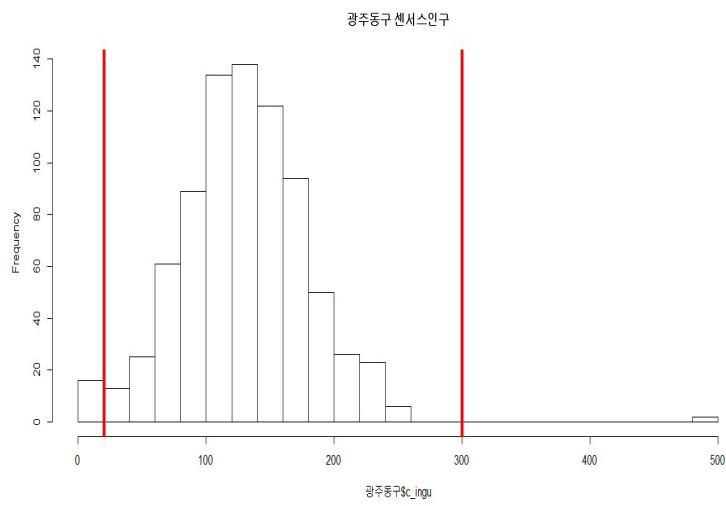
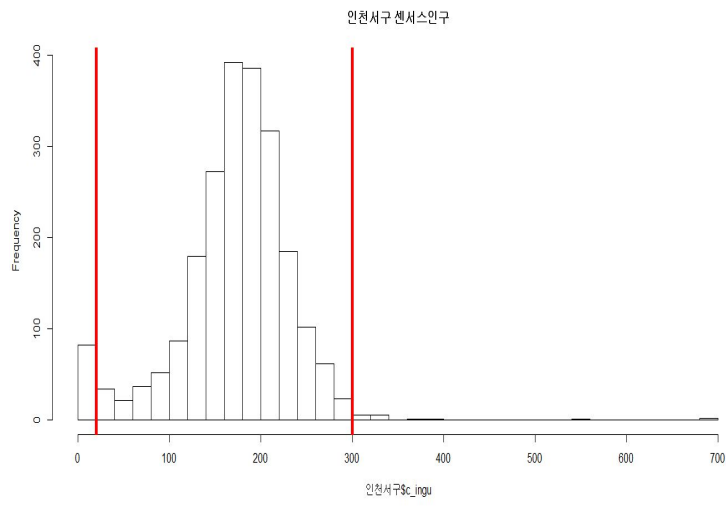
시군구	읍면동	조사주택	행정주택	차이율
인천서구	51	8,942	9,540	-6.7
	53	14,880	15,021	-0.9
	54	3,106	6,158	-98.3
	55	5,965	6,986	-17.1
	56	4,355	4,374	-0.4
	58	7,543	7,583	-0.5
	59	4,786	4,902	-2.4
	60	6,117	6,118	0.0
	62	4,790	5,084	-6.1
	63	7,200	7,227	-0.4
	64	6,688	6,834	-2.2
	65	3,561	3,517	1.2
	68	14,399	15,192	-5.5
	69	7,243	7,277	-0.5
	70	7,287	7,814	-7.2
	71	11,114	10,587	4.7
	72	2,232	2,565	-14.9
	73	8,507	8,537	-0.4
광주동구	51	1,082	1,594	-47.3
	54	1,551	1,592	-2.6
	56	1,793	1,879	-4.8
	58	2,919	2,941	-0.8
	59	2,851	2,920	-2.4
	61	3,071	3,107	-1.2
	62	1,511	1,591	-5.3
	63	2,195	2,174	1.0
	64	1,062	1,326	-24.9
	68	2,020	2,128	-5.3
	71	4,454	4,498	-1.0
	72	2,528	2,545	-0.7
	73	2,265	2,331	-2.9

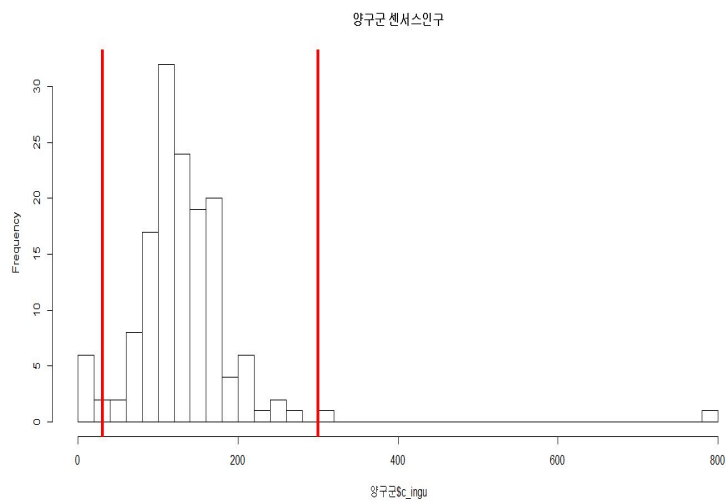
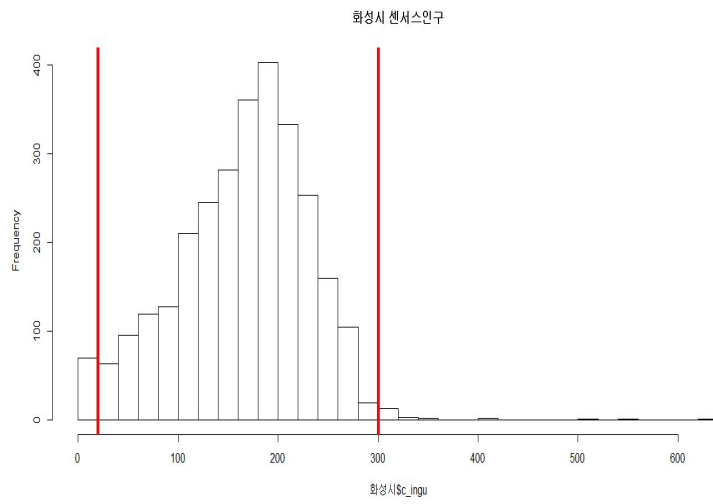
시군구	읍면동	조사주택	행정주택	차이율
신안군	11	2,280	2,062	9.6
	31	836	804	3.8
	32	1,357	1,272	6.3
	33	1,103	1,083	1.8
	34	1,631	1,671	-2.5
	35	1,281	1,347	-5.2
	36	1,593	1,415	11.2
	37	798	853	-6.9
	38	652	656	-0.6
	39	798	794	0.5
	40	1,527	1,502	1.6
	41	571	576	-0.9
	42	1,108	1,034	6.7
	43	2,227	2,369	-6.4
김천시	11	3,254	3,270	-0.5
	31	1,174	1,200	-2.2
	32	1,333	1,340	-0.5
	33	1,121	1,123	-0.2
	34	1,737	1,789	-3.0
	35	1,982	1,859	6.2
	36	1,348	1,302	3.4
	37	1,452	1,440	0.8
	38	842	879	-4.4
	39	1,054	990	6.1
	40	1,689	1,740	-3.0
	41	971	982	-1.1
	42	691	754	-9.1
	43	1,176	1,235	-5.0
	44	601	624	-3.8
	55	1,923	1,764	8.3
	56	8,142	8,245	-1.3
	58	7,837	7,872	-0.4
	59	3,797	3,650	3.9
	60	3,014	3,033	-0.6
울릉군	11	2,073	1,987	4.1
	31	688	767	-11.5
	32	595	660	-10.9

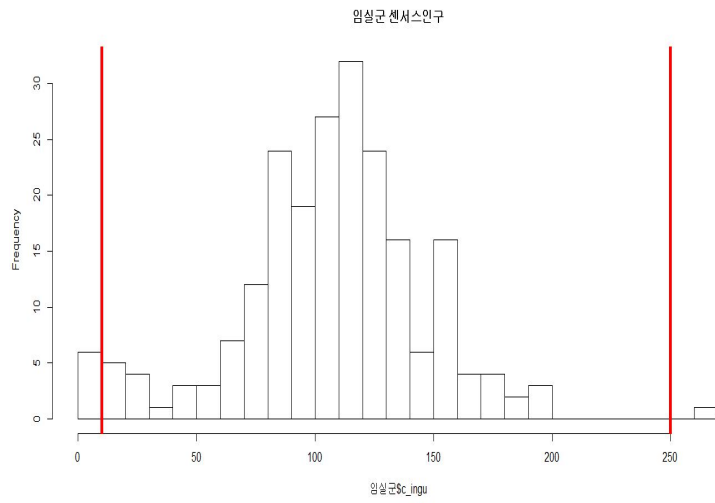
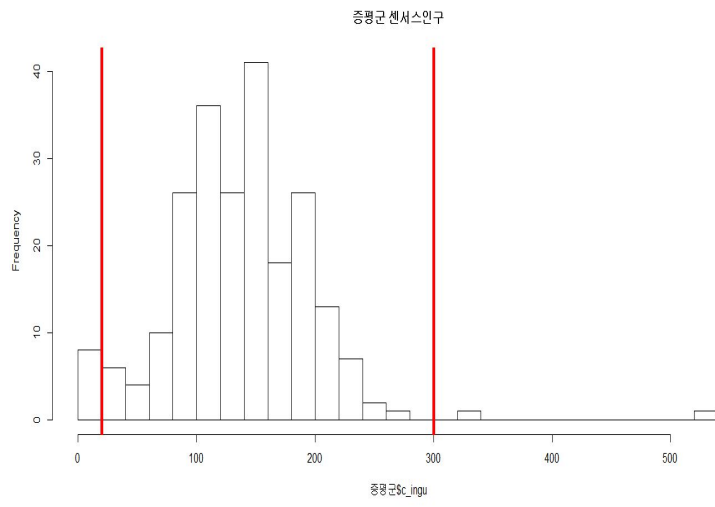
시군구	읍면동	조사주택	행정주택	차이율
임실군	11	2,473	2,721	-10.0
	31	638	713	-11.8
	32	898	981	-9.2
	33	724	725	-0.1
	34	747	879	-17.7
	35	2,031	2,139	-5.3
	36	648	692	-6.8
	37	886	953	-7.6
	38	1,370	1,446	-5.5
	39	838	857	-2.3
	40	587	680	-15.8
나주시	41	599	662	-10.5
	11	2,824	2,930	-3.8
	31	1,356	1,401	-3.3
	32	1,498	1,564	-4.4
	33	810	828	-2.2
	34	1,285	1,317	-2.5
	35	1,525	1,532	-0.5
	36	1,997	2,146	-7.5
	37	1,153	1,213	-5.2
	38	2,118	2,259	-6.7
	39	1,882	1,954	-3.8
	40	1,342	1,550	-15.5
	41	977	982	-0.5
	42	2,151	2,278	-5.9
	51	2,837	2,865	-1.0
	52	1,707	1,743	-2.1
	54	2,357	2,391	-1.4
	55	3,744	3,989	-6.5
	58	1,893	2,014	-6.4
	60	1,351	1,568	-16.1

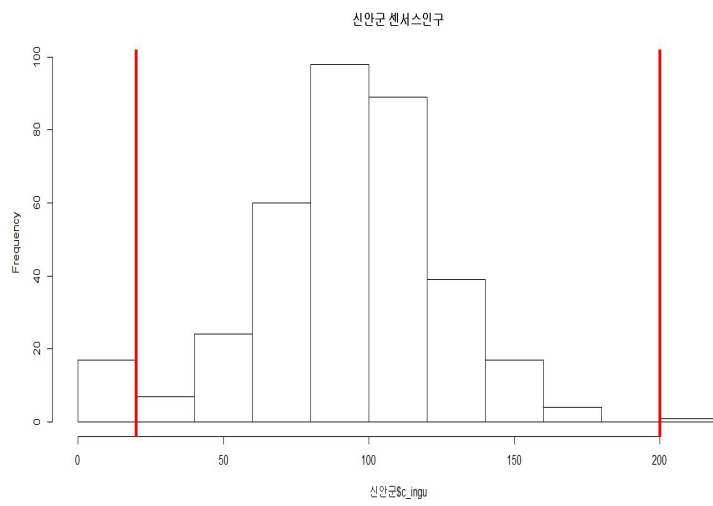
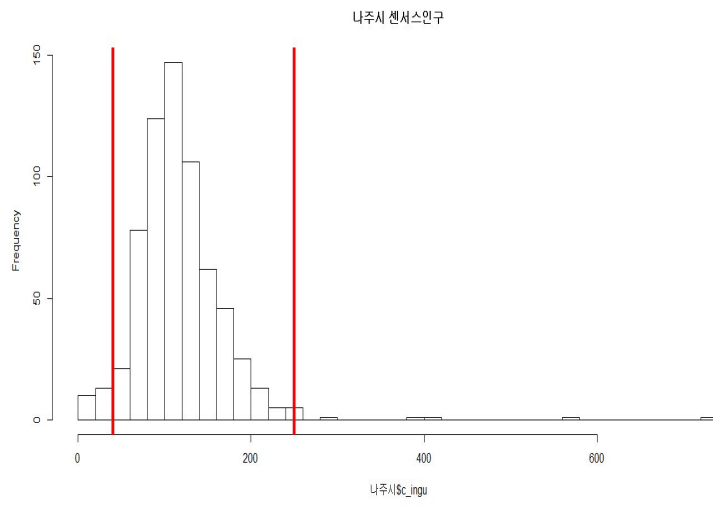
2-3. 지역별 조사구의 인구분포

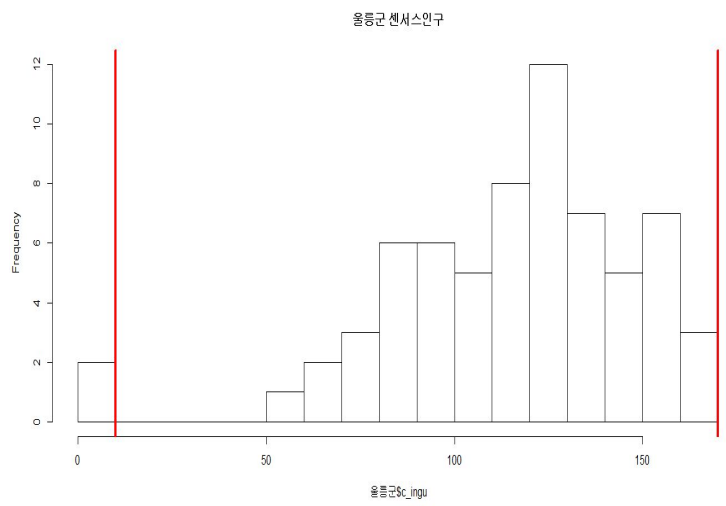
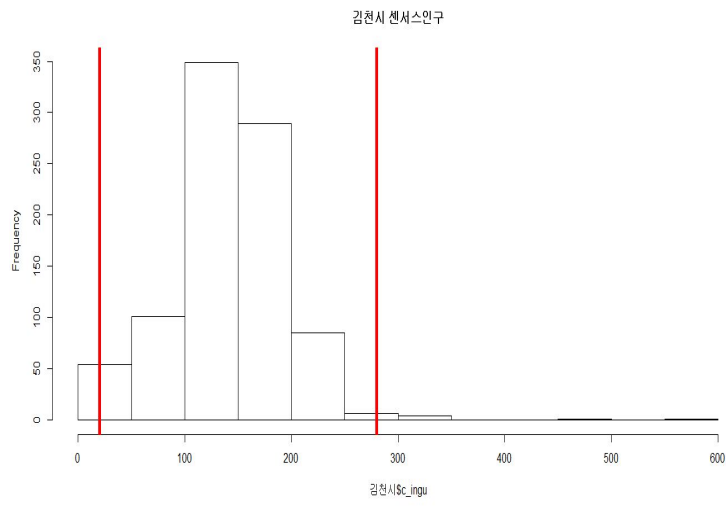












2-4. 읍면동별 추정 결과

(1) 인구 추정

지역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
금천구	51	157	32	20.4	18,133	18,414	-1.6	8.4
	52	212	43	20.3	31,061	30,424	2.0	6.1
	53	142	29	20.4	22,582	23,736	-5.1	4.6
	54	196	42	21.4	27,351	27,492	-0.5	4.7
	55	117	24	20.5	18,293	18,624	-1.8	4.9
	57	197	40	20.3	29,086	29,656	-2.0	5.0
	58	134	25	18.7	24,851	24,340	2.1	4.8
	59	71	14	19.7	12,357	12,266	0.7	6.6
	60	131	28	21.4	23,361	23,951	-2.5	3.5
	61	135	27	20.0	23,372	24,349	-4.2	4.2
	전체	1,492	304	20.4	230,447	233,252	-1.2	1.8
포천포구	54	211	32	15.2	31,264	30,240	3.3	8.3
	55	131	21	16.0	19,749	19,542	1.0	6.9
	56	200	33	16.5	32,240	31,668	1.8	5.8
	61	117	18	15.4	16,073	15,540	3.3	9.4
	62	120	20	16.7	20,531	20,748	-1.1	6.2
	63	169	27	16.0	20,835	20,394	2.1	6.5
	65	121	20	16.5	18,348	19,008	-3.6	5.9
	66	89	15	16.9	12,247	13,632	-11.3	7.6
	67	110	18	16.4	16,016	17,268	-7.8	6.5
	68	158	26	16.5	23,224	22,476	3.2	5.7
	69	114	19	16.7	17,748	15,996	9.9	8.5
	70	101	17	16.8	15,284	15,768	-3.2	6.5
	71	143	23	16.1	16,747	17,826	-6.4	7.0
	72	177	30	16.9	26,636	28,812	-8.2	5.7
	73	170	28	16.5	23,696	25,182	-6.3	6.9
	74	131	22	16.8	15,269	16,412	-7.5	12.2
	75	146	23	15.8	17,459	17,832	-2.1	7.1
	76	187	31	16.6	31,177	30,126	3.4	5.5
	전체	2,595	423	16.3	374,543	378,470	-1.0	1.8

지역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
인천 서구	51	152	20	13.2	29,074	29,588	-1.8	6.2
	53	263	31	11.8	47,399	44,845	5.4	6.1
	54	94	9	9.6	10,993	11,599	-5.5	17.8
	55	100	7	7.0	9,076	9,538	-5.1	34.4
	56	66	8	12.1	12,451	11,394	8.5	7.4
	58	142	18	12.7	24,381	26,443	-8.5	5.7
	59	98	12	12.2	16,566	17,219	-3.9	6.4
	60	106	12	11.3	17,933	17,739	1.1	6.8
	62	85	10	11.8	13,566	14,562	-7.3	7.1
	63	117	17	14.5	23,186	24,729	-6.7	5.2
	64	121	14	11.6	20,569	21,434	-4.2	5.6
	65	70	9	12.9	12,294	12,858	-4.6	8.2
	68	236	28	11.9	39,170	39,446	-0.7	6.9
	69	111	14	12.6	23,113	23,252	-0.6	6.1
	70	119	14	11.8	22,431	24,287	-8.3	6.0
	71	170	21	12.4	36,117	34,704	3.9	5.7
	72	40	3	7.5	3,475	3,619	-4.2	35.0
	73	156	20	12.8	27,263	27,677	-1.5	5.7
	전체	2,246	267	11.9	389,057	394,932	-1.5	2.1

지역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
광 주 동 구	51	37	10	27.0	4,625	4,420	4.4	6.8
	54	59	11	18.6	6,153	5,668	7.9	13.5
	56	63	16	25.4	8,050	7,101	11.8	5.4
	58	63	16	25.4	9,428	9,266	1.7	7.3
	59	71	19	26.8	8,877	8,825	0.6	5.5
	61	77	20	26.0	11,128	10,649	4.3	5.1
	62	61	13	21.3	5,941	5,334	10.2	11.4
	63	56	15	26.8	6,509	5,948	8.6	8.5
	64	51	12	23.5	5,639	4,036	28.4	18.7
	68	61	14	23.0	6,676	7,139	-6.9	8.6
	71	86	21	24.4	13,598	12,143	10.7	5.7
	72	58	15	25.9	9,246	8,510	8.0	5.9
	73	56	12	21.4	8,128	7,703	5.2	9.2
	전체	799	194	24.3	103,998	96,742	7.0	2.5
신 안 군	11	43	11	25.6	4,109	4,585	-11.6	12.0
	31	18	4	22.2	1,671	2,359	-41.2	16.1
	32	30	8	26.7	3,030	3,434	-13.3	12.4
	33	21	6	28.6	1,826	2,566	-40.5	10.0
	34	29	8	27.6	3,099	3,917	-26.4	9.0
	35	24	7	29.2	2,423	3,374	-39.2	7.7
	36	34	9	26.5	2,930	4,393	-49.9	8.9
	37	17	4	23.5	1,275	1,565	-22.7	14.1
	38	13	4	30.8	1,347	1,974	-46.5	13.9
	39	14	4	28.6	1,287	1,971	-53.1	12.5
	40	32	8	25.0	2,585	3,371	-30.4	10.2
	41	12	3	25.0	936	1,204	-28.6	23.3
	42	23	7	30.4	1,745	2,051	-17.5	10.7
	43	46	12	26.1	4,724	6,230	-31.9	8.3
	전체	356	95	26.7	32,987	42,991	-30.3	3.2

지 역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
화 성 시	12	367	53	14.4	63,694	62,618	1.7	4.9
	13	117	15	12.8	14,821	16,292	-9.9	10.8
	14	299	46	15.4	49,188	49,735	-1.1	5.4
	31	46	6	13.0	6,853	7,037	-2.7	10.4
	33	34	5	14.7	5,101	6,618	-29.7	14.5
	35	35	4	11.4	4,368	4,982	-14.1	19.7
	36	71	9	12.7	8,597	10,701	-24.5	11.8
	37	39	5	12.8	4,587	6,625	-44.4	15.3
	38	58	8	13.8	8,007	10,235	-27.8	11.8
	39	67	9	13.4	8,890	9,538	-7.3	14.1
	42	32	4	12.5	3,157	4,367	-38.3	19.8
	43	82	10	12.2	11,050	12,628	-14.3	15.5
	44	65	6	9.2	4,360	5,442	-24.8	36.9
	51	132	18	13.6	21,177	21,916	-3.5	8.0
	52	243	38	15.6	34,872	33,124	5.0	7.1
	53	142	20	14.1	27,098	26,655	1.6	4.5
	54	137	23	16.8	26,818	27,147	-1.2	5.4
	55	92	14	15.2	19,637	19,993	-1.8	5.9
	56	74	11	14.9	14,661	13,865	5.4	6.7
	57	131	18	13.7	22,817	23,099	-1.2	7.9
	60	166	25	15.1	32,109	31,726	1.2	4.3
	61	243	44	18.1	44,160	43,271	2.0	5.4
	62	197	30	15.2	38,138	38,036	0.3	4.7
	전체	2,869	421	14.7	474,160	485,647	-2.4	1.8

지역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
양 구 면	11	80	26	32.5	11,257	11,621	-3.2	10.7
	31	29	10	34.5	3,412	4,087	-19.8	12.3
	32	16	4	25.0	2,081	2,312	-11.1	12.9
	33	12	3	25.0	1,293	1,794	-38.7	16.6
	34	9	3	33.3	1,209	682	43.6	33.0
	전체	146	46	31.5	19,252	20,496	-6.5	7.1
중 평 군	14	206	54	26.2	29,169	30,814	-5.6	5.1
	31	20	4	20.0	2,035	2,124	-4.4	21.7
	전체	226	58	25.7	31,204	32,938	-5.6	5.1
임 실 군	11	44	10	22.7	5,352	6,172	-15.3	10.8
	31	11	2	18.2	1,190	1,194	-0.3	26.2
	32	13	3	23.1	1,263	1,371	-8.6	28.6
	33	15	3	20.0	1,346	1,326	1.5	29.8
	34	14	3	21.4	1,370	1,877	-37.0	17.9
	35	34	7	20.6	3,750	4,250	-13.3	8.1
	36	10	2	20.0	1,123	1,121	0.2	25.2
	37	16	3	18.8	1,425	1,440	-1.0	24.2
	38	25	5	20.0	3,021	4,123	-36.5	11.4
	39	15	3	20.0	1,504	1,813	-20.6	15.9
	40	11	2	18.2	1,016	1,321	-30.0	30.2
	41	11	3	27.3	1,130	1,239	-9.7	22.9
	전체	219	46	21.0	23,490	27,246	-16.0	5.1

지 역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
나 주 시	11	57	17	29.8	7,142	7,403	-3.7	12.2
	31	24	7	29.2	2,597	3,271	-26.0	9.8
	32	26	8	30.8	2,969	3,487	-17.4	6.3
	33	15	5	33.3	1,386	1,892	-36.5	13.9
	34	26	8	30.8	2,328	3,064	-31.6	9.7
	35	26	8	30.8	2,617	3,188	-21.8	7.5
	36	38	11	28.9	3,632	4,184	-15.2	8.1
	37	20	7	35.0	1,858	2,188	-17.8	8.8
	38	41	12	29.3	4,350	5,166	-18.8	6.5
	39	38	11	28.9	4,671	5,215	-11.6	12.5
	40	27	8	29.6	3,081	4,002	-29.9	11.9
	41	20	5	25.0	1,845	1,805	2.2	16.0
	42	39	13	33.3	3,862	5,326	-37.9	6.6
	51	47	15	31.9	7,546	7,236	4.1	4.8
	52	33	8	24.2	4,179	3,419	18.2	12.4
	54	51	16	31.4	6,411	7,023	-9.6	7.3
	55	70	23	32.9	10,152	8,892	12.4	10.7
	58	35	11	31.4	4,143	4,530	-9.3	11.3
	60	27	9	33.3	3,056	3,969	-29.9	9.9
	전체	660	202	30.6	77,825	85,261	-9.6	2.7

지역	동	N (개)	n (개)	f (%)	모집단 인구	추정 인구	차이율 (%)	RSE (%)
김천시	11	60	12	20.0	7,403	8,498	-14.8	10.4
	31	19	3	15.8	2,809	2,828	-0.7	24.2
	32	23	4	17.4	2,932	3,255	-11.0	18.5
	33	16	3	18.8	2,539	2,604	-2.6	11.3
	34	30	6	20.0	3,548	3,416	3.7	15.5
	35	43	6	14.0	5,105	4,679	8.3	21.9
	36	25	4	16.0	2,996	3,285	-9.6	23.5
	37	28	5	17.9	3,757	4,354	-15.9	15.9
	38	12	2	16.7	1,773	1,900	-7.2	18.4
	39	17	3	17.6	2,044	2,094	-2.5	19.5
	40	22	5	22.7	2,770	3,144	-13.5	13.1
	41	17	2	11.8	1,673	2,148	-28.4	23.7
	42	10	2	20.0	1,142	1,560	-36.6	25.5
	43	18	4	22.2	1,886	2,104	-11.6	19.6
	44	11	2	18.2	990	1,186	-19.8	25.0
	55	36	7	19.4	5,335	5,466	-2.5	8.2
	56	157	31	19.7	26,291	24,241	7.8	6.5
	58	145	31	21.4	22,971	22,452	2.3	4.7
	59	66	13	19.7	10,225	9,616	6.0	7.5
	60	62	12	19.4	8,661	9,752	-12.6	5.6
	61	73	15	20.5	9,973	10,165	-1.9	6.6
	전체	890	172	19.3	126,823	128,744	-1.5	2.6
영양군	11	45	18	40.0	5,579	7,163	-28.4	4.0
	31	11	4	36.4	1,078	1,393	-29.2	17.4
	32	11	4	36.4	1,080	1,228	-13.7	15.9
	전체	67	26	38.8	7,737	9,783	-26.4	4.6

(2) 가구수 추정

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
금천구	51	147	32	21.8	8,689	8,053	7.3	9.6
	52	200	43	21.5	12,299	12,175	1.0	4.4
	53	137	29	21.2	8,418	9,252	-9.9	3.1
	54	191	42	22.0	10,531	11,140	-5.8	3.1
	55	113	24	21.2	6,868	7,503	-9.2	3.1
	57	186	40	21.5	10,986	11,732	-6.8	3.2
	58	128	25	19.5	7,950	7,896	0.7	2.4
	59	67	14	20.9	4,177	4,403	-5.4	4.2
	60	129	28	21.7	8,211	8,969	-9.2	2.3
	61	133	27	20.3	8,404	9,276	-10.4	3.6
	전체	1,431	304	21.2	86,533	90,397	-4.5	1.5
영등포구	54	210	32	15.2	10,789	10,194	5.5	6.9
	55	128	21	16.4	7,411	7,794	-5.2	3.9
	56	190	33	17.4	11,629	10,854	6.7	4.1
	61	115	18	15.7	6,101	6,336	-3.9	6.6
	62	116	20	17.2	6,857	7,248	-5.7	3.5
	63	163	27	16.6	8,881	8,910	-0.3	4.2
	65	118	20	16.9	6,794	7,278	-7.1	4.6
	66	88	15	17.0	4,557	5,358	-17.6	5.4
	67	107	18	16.8	6,056	6,540	-8.0	4.4
	68	154	26	16.9	8,542	8,514	0.3	4.9
	69	113	19	16.8	6,038	6,456	-6.9	5.7
	70	100	17	17.0	5,750	6,282	-9.3	5.5
	71	142	23	16.2	6,505	7,236	-11.2	5.6
	72	173	30	17.3	9,437	10,788	-14.3	3.9
	73	168	28	16.7	9,028	9,138	-1.2	4.7
	74	119	22	18.5	6,582	7,620	-15.8	9.4
	75	141	23	16.3	7,106	7,566	-6.5	5.1
	76	184	31	16.8	10,574	10,476	0.9	3.6
	전체	2,529	423	16.7	138,637	144,588	-4.3	1.3

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
인 천 서 구	51	147	20	13.6	9,307	10,027	-7.7	5.7
	53	253	31	12.3	15,955	14,869	6.8	3.5
	54	93	9	9.7	4,283	4,811	-12.3	18.1
	55	99	7	7.1	3,093	2,929	5.3	33.1
	56	65	8	12.3	4,156	3,807	8.4	4.7
	58	139	18	12.9	8,837	9,993	-13.1	3.8
	59	96	12	12.5	5,894	6,140	-4.2	4.7
	60	103	12	11.7	6,237	6,455	-3.5	4.3
	62	84	10	11.9	4,999	5,382	-7.7	5.8
	63	116	17	14.7	7,156	7,504	-4.9	3.8
	64	120	14	11.7	7,115	7,111	0.1	4.0
	65	69	9	13.0	4,474	4,689	-4.8	5.6
	68	229	28	12.2	12,684	13,242	-4.4	4.3
	69	109	14	12.8	7,089	7,571	-6.8	4.6
	70	117	14	12.0	6,963	7,715	-10.8	4.3
	71	168	21	12.5	10,604	10,385	2.1	6.4
	72	39	3	7.7	1,084	1,124	-3.7	35.0
	73	155	20	12.9	9,045	9,044	0.0	5.3
	전체	2,201	267	12.1	128,975	132,799	-3.0	1.7

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
광 주 동 구	51	36	10	27.8	2,020	2,462	-21.9	5.4
	54	43	11	25.6	2,946	2,538	13.8	5.9
	56	60	16	26.7	3,944	3,579	9.3	4.6
	58	61	16	26.2	3,445	3,318	3.7	4.2
	59	69	19	27.5	3,758	3,637	3.2	3.9
	61	76	20	26.3	4,369	3,994	8.6	3.5
	62	52	13	25.0	2,714	2,185	19.5	6.4
	63	54	15	27.8	3,012	2,415	19.8	5.0
	64	45	12	26.7	2,768	2,097	24.3	6.3
	68	56	14	25.0	3,067	3,122	-1.8	5.2
	71	79	21	26.6	4,694	4,462	4.9	4.3
	72	56	15	26.8	3,217	2,953	8.2	5.0
	73	47	12	25.5	2,584	2,780	-7.6	5.7
	전체	734	194	26.4	42,538	39,543	7.0	1.4
신 안 군	11	39	11	28.2	2,095	2,181	-4.1	7.5
	31	15	4	26.7	789	1,043	-32.2	7.6
	32	28	8	28.6	1,266	1,407	-11.1	9.8
	33	20	6	30.0	974	1,229	-26.1	6.0
	34	27	8	29.6	1,524	1,764	-15.7	4.7
	35	22	7	31.8	1,163	1,498	-28.8	5.5
	36	33	9	27.3	1,405	1,747	-24.3	7.6
	37	15	4	26.7	680	728	-7.1	7.5
	38	12	4	33.3	617	837	-35.6	8.3
	39	13	4	30.8	703	987	-40.4	6.1
	40	30	8	26.7	1,350	1,554	-15.1	7.5
	41	11	3	27.3	522	613	-17.3	17.0
	42	22	7	31.8	952	914	4.0	7.1
	43	41	12	29.3	2,074	2,503	-20.7	4.9
	전체	328	95	29.0	16,114	19,002	-17.9	2.1

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
화성시	12	354	53	15.0	20,426	20,375	0.3	2.8
	13	103	15	14.6	5,164	6,348	-22.9	7.5
	14	294	46	15.6	17,742	17,354	2.2	4.5
	31	44	6	13.6	2,376	2,319	2.4	7.8
	33	32	5	15.6	1,827	2,576	-41.0	14.1
	35	34	4	11.8	1,591	2,001	-25.8	21.1
	36	64	9	14.1	3,244	4,171	-28.6	9.1
	37	34	5	14.7	1,766	2,846	-61.2	11.9
	38	53	8	15.1	3,175	4,644	-46.3	10.2
	39	59	9	15.3	3,234	3,941	-21.9	9.4
	42	28	4	14.3	1,208	1,927	-59.5	14.0
	43	70	10	14.3	3,943	4,806	-21.9	9.0
	44	59	6	10.2	1,414	2,021	-42.9	34.5
	51	124	18	14.5	7,262	8,119	-11.8	5.1
	52	239	38	15.9	14,133	12,921	8.6	3.5
	53	140	20	14.3	8,401	8,389	0.1	3.3
	54	135	23	17.0	8,405	8,538	-1.6	2.8
	55	87	14	16.1	5,545	6,059	-9.3	3.3
	56	72	11	15.3	4,284	4,313	-0.7	5.4
	57	125	18	14.4	7,516	8,038	-6.9	4.8
	60	163	25	15.3	9,697	9,711	-0.1	2.7
	61	240	44	18.3	13,515	13,272	1.8	4.6
	62	195	30	15.4	12,555	12,067	3.9	3.9
	전체	2,748	421	15.3	158,423	166,754	-5.3	1.3

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
양 구 민	11	74	26	35.1	3,975	4,328	-8.9	4.6
	31	26	10	38.5	1,244	1,547	-24.4	5.8
	32	15	4	26.7	713	892	-25.1	14.9
	33	10	3	30.0	500	740	-48.0	10.4
	34	8	3	37.5	487	241	50.6	20.0
	전체	133	46	34.6	6,919	7,748	-12.0	3.6
중 평 군	14	200	54	27.0	10,708	11,353	-6.0	3.0
	31	18	4	22.2	851	895	-5.1	14.7
	전체	218	58	26.6	11,559	12,247	-6.0	3.0
임 실 군	11	41	10	24.4	2,133	2,420	-13.5	7.5
	31	10	2	20.0	601	533	11.3	14.9
	32	12	3	25.0	602	510	15.2	22.0
	33	12	3	25.0	649	592	8.7	14.8
	34	13	3	23.1	678	793	-16.9	10.5
	35	33	7	21.2	1,720	1,790	-4.1	7.1
	36	9	2	22.2	563	524	6.9	15.6
	37	14	3	21.4	723	697	3.6	17.7
	38	24	5	20.8	1,244	1,535	-23.4	12.2
	39	14	3	21.4	711	747	-5.1	9.5
	40	9	2	22.2	509	601	-18.1	13.5
	41	10	3	30.0	560	556	0.8	8.5
	전체	201	46	22.9	10,693	11,299	-5.7	3.4

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
나 주 시	11	54	17	31.5	2,670	3,169	-18.7	6.1
	31	22	7	31.8	1,237	1,438	-16.2	5.8
	32	25	8	32.0	1,347	1,472	-9.3	4.2
	33	14	5	35.7	716	926	-29.3	7.6
	34	24	8	33.3	1,138	1,413	-24.2	6.3
	35	25	8	32.0	1,329	1,595	-20.0	4.6
	36	34	11	32.4	1,699	1,987	-17.0	5.0
	37	19	7	36.8	975	1,089	-11.7	5.1
	38	39	12	30.8	1,918	2,314	-20.7	6.3
	39	33	11	33.3	1,706	2,151	-26.1	5.3
	40	25	8	32.0	1,255	1,636	-30.3	6.7
	41	17	5	29.4	840	907	-8.0	12.3
	42	37	13	35.1	1,892	2,407	-27.2	5.3
	51	46	15	32.6	2,693	2,555	5.1	3.2
	52	29	8	27.6	1,565	1,262	19.4	10.5
	54	47	16	34.0	2,445	2,910	-19.0	5.2
	55	65	23	35.4	3,454	3,342	3.2	3.6
	58	33	11	33.3	1,663	1,984	-19.3	6.8
	60	26	9	34.6	1,253	1,565	-24.9	6.8
	전체	614	202	32.9	31,795	36,123	-13.6	1.4

지역	동	N (개)	n (개)	f (%)	모집단 가구	추정 가구	차이율 (%)	RSE (%)
김천시	11	56	12	21.4	2,984	3,255	-9.1	10.4
	31	16	3	18.8	1,095	1,161	-6.1	24.2
	32	19	4	21.1	1,145	1,428	-24.8	18.5
	33	15	3	20.0	1,048	1,020	2.6	11.3
	34	28	6	21.4	1,592	1,535	3.6	15.5
	35	37	6	16.2	1,772	2,060	-16.3	21.9
	36	22	4	18.2	1,221	1,278	-4.7	23.5
	37	22	5	22.7	1,387	1,599	-15.3	15.9
	38	11	2	18.2	767	748	2.4	18.4
	39	15	3	20.0	915	875	4.4	19.5
	40	21	5	23.8	1,422	1,399	1.6	13.1
	41	13	2	15.4	800	962	-20.3	23.7
	42	9	2	22.2	590	743	-26.0	25.5
	43	17	4	23.5	990	1,006	-1.6	19.6
	44	9	2	22.2	485	578	-19.2	25.0
	55	34	7	20.6	2,092	1,987	5.0	8.2
	56	147	31	21.1	8,695	8,629	0.8	6.5
	58	140	31	22.1	8,289	7,729	6.8	4.7
	59	63	13	20.6	3,748	3,367	10.2	7.5
	60	59	12	20.3	3,533	3,882	-9.9	5.6
	61	71	15	21.1	4,273	3,965	7.2	6.6
	전체	824	172	20.9	48,843	49,209	-0.7	2.6
영양군	11	44	18	40.9	2,427	3,103	-27.8	3.5
	31	10	4	40.0	540	638	-18.1	8.8
	32	10	4	40.0	527	548	-3.9	4.8
	전체	64	26	40.6	3,494	4,288	-22.7	2.9

(3) 주택수 추정

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
금천구	51	157	32	20.4	4,736	4,077	13.9	12.7
	52	211	43	20.4	8,711	8,313	4.6	9.2
	53	141	29	20.6	3,510	3,440	2.0	8.6
	54	195	42	21.5	5,289	5,088	3.8	8.6
	55	116	24	20.7	3,082	3,225	-4.6	9.8
	57	196	40	20.4	6,616	6,267	5.3	8.9
	58	133	25	18.8	7,804	7,335	6.0	5.3
	59	70	14	20.0	3,622	3,742	-3.3	9.2
	60	130	28	21.5	4,711	4,279	9.2	9.2
	61	134	27	20.1	4,239	3,948	6.9	9.5
	전체	1,483	304	20.5	52,320	49,713	5.0	3.3
영등포구	54	212	32	15.1	11,412	10,596	7.2	7.1
	55	130	21	16.2	5,740	6,234	-8.6	9.2
	56	199	33	16.6	10,647	9,840	7.6	6.2
	61	116	18	15.5	5,330	4,842	9.2	9.1
	62	119	20	16.8	6,119	6,018	1.7	7.1
	63	168	27	16.1	4,239	4,248	-0.2	11.8
	65	120	20	16.7	3,298	3,786	-14.8	13.7
	66	88	15	17.0	2,724	2,706	0.7	15.3
	67	109	18	16.5	2,993	3,228	-7.9	13.9
	68	157	26	16.6	5,597	5,562	0.6	10.6
	69	115	19	16.5	4,423	4,224	4.5	12.3
	70	101	17	16.8	2,780	2,772	0.3	14.1
	71	142	23	16.2	3,475	2,808	19.2	13.8
	72	176	30	17.0	6,325	6,492	-2.6	10.2
	73	169	28	16.6	6,116	6,606	-8.0	10.8
	74	130	22	16.9	4,584	4,604	-0.4	13.8
	75	145	23	15.9	4,320	4,422	-2.4	12.7
	76	186	31	16.7	10,208	9,522	6.7	5.0
	전체	2,582	423	16.4	100,330	98,510	1.8	2.6

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
인천 서구	51	151	20	13.2	8,942	9,978	-11.6	8.0
	53	262	31	11.8	14,880	14,256	4.2	5.9
	54	93	9	9.7	3,106	3,773	-21.5	20.6
	55	103	7	6.8	5,965	3,645	38.9	8.8
	56	65	8	12.3	4,355	4,207	3.4	4.8
	58	141	18	12.8	7,543	7,461	1.1	7.2
	59	97	12	12.4	4,786	4,820	-0.7	8.3
	60	105	12	11.4	6,117	5,893	3.7	6.7
	62	84	10	11.9	4,790	5,075	-6.0	6.7
	63	116	17	14.7	7,200	7,560	-5.0	4.1
	64	120	14	11.7	6,688	6,949	-3.9	6.5
	65	70	9	12.9	3,561	3,200	10.1	13.6
	68	249	28	11.2	14,399	12,867	10.6	5.5
	69	110	14	12.7	7,243	7,293	-0.7	5.3
	70	118	14	11.9	7,287	6,591	9.5	6.6
	71	169	21	12.4	11,114	10,960	1.4	9.4
	72	39	3	7.7	2,232	1,925	13.8	10.8
	73	155	20	12.9	8,507	8,550	-0.5	5.2
	전체	2,247	267	11.9	128,715	125,002	2.9	1.9

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
광 주 동 구	51	36	10	27.8	1,082	1,559	-44.1	13.4
	54	58	11	19.0	1,551	1,766	-13.9	17.7
	56	62	16	25.8	1,793	1,701	5.1	9.3
	58	62	16	25.8	2,919	3,345	-14.6	8.1
	59	70	19	27.1	2,851	2,830	0.7	7.0
	61	76	20	26.3	3,071	3,045	0.8	8.7
	62	60	13	21.7	1,511	1,525	-0.9	17.2
	63	55	15	27.3	2,195	2,139	2.6	10.2
	64	50	12	24.0	1,062	1,475	-38.9	14.1
	68	60	14	23.3	2,020	2,300	-13.9	11.1
	71	85	21	24.7	4,454	4,205	5.6	7.9
	72	57	15	26.3	2,528	2,561	-1.3	8.7
	73	55	12	21.8	2,265	2,523	-11.4	13.7
	전체	786	194	24.7	29,302	30,975	-5.7	3.3
신 안 군	11	42	11	26.2	2,280	1,883	17.4	10.7
	31	17	4	23.5	836	851	-1.7	16.6
	32	29	8	27.6	1,357	1,196	11.9	11.6
	33	20	6	30.0	1,103	1,166	-5.7	5.4
	34	28	8	28.6	1,631	1,684	-3.2	7.6
	35	24	7	29.2	1,281	1,278	0.3	9.9
	36	33	9	27.3	1,593	1,428	10.4	7.6
	37	16	4	25.0	798	718	10.1	13.7
	38	12	4	33.3	652	690	-5.8	7.1
	39	13	4	30.8	798	942	-18.0	5.9
	40	31	8	25.8	1,527	1,390	9.0	9.0
	41	11	3	27.3	571	595	-4.2	17.8
	42	22	7	31.8	1,108	893	19.4	5.3
	43	45	12	26.7	2,227	2,310	-3.7	9.6
	전체	343	95	27.7	17,762	17,020	4.2	2.8

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
화 성 시	12	367	53	14.4	19,751	19,874	-0.6	4.8
	13	116	15	12.9	5,174	5,651	-9.2	13.0
	14	298	46	15.4	16,359	17,246	-5.4	5.5
	31	45	6	13.3	2,357	2,427	-3.0	10.8
	33	33	5	15.2	1,866	2,096	-12.3	16.7
	35	34	4	11.8	1,685	1,785	-5.9	21.6
	36	70	9	12.9	3,604	3,792	-5.2	12.3
	37	38	5	13.2	1,913	2,339	-22.3	18.8
	38	57	8	14.0	3,226	3,840	-19.0	14.2
	39	66	9	13.6	3,526	3,265	7.4	14.0
	42	31	4	12.9	1,338	1,345	-0.5	21.1
	43	82	10	12.2	3,986	4,617	-15.8	17.6
	44	66	6	9.1	2,184	1,582	27.6	25.0
	51	131	18	13.7	7,465	7,848	-5.1	7.5
	52	242	38	15.7	9,273	9,350	-0.8	11.4
	53	141	20	14.2	8,226	8,423	-2.4	4.7
	54	136	23	16.9	8,126	8,440	-3.9	5.2
	55	91	14	15.4	5,575	5,870	-5.3	6.4
	56	73	11	15.1	4,350	4,259	2.1	6.1
	57	130	18	13.8	7,322	7,294	0.4	7.6
	60	165	25	15.2	9,724	9,442	2.9	3.8
	61	242	44	18.2	14,265	13,990	1.9	3.8
	62	196	30	15.3	12,494	12,181	2.5	3.4
	전체	2,850	421	14.8	153,789	156,956	-2.1	1.8

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
양 구 민	11	80	26	32.5	4,061	3,856	5.1	7.0
	31	28	10	35.7	1,336	1,313	1.8	9.4
	32	16	4	25.0	804	353	56.1	26.2
	33	11	3	27.3	587	405	31.0	20.5
	34	8	3	37.5	501	104	79.3	22.1
	전체	143	46	32.2	7,289	6,030	17.3	5.3
중 평 군	14	207	54	26.1	10,792	10,667	1.2	4.5
	31	20	4	20.0	1,048	853	18.6	12.8
	전체	227	58	25.6	11,840	11,520	2.7	4.2
임 실 군	11	43	10	23.3	2,473	2,754	-11.4	8.5
	31	10	2	20.0	638	574	10.0	13.9
	32	12	3	25.0	898	692	22.9	12.1
	33	14	3	21.4	724	638	11.9	32.0
	34	13	3	23.1	747	870	-16.5	11.4
	35	33	7	21.2	2,031	2,077	-2.3	8.4
	36	9	2	22.2	648	569	12.1	16.1
	37	15	3	20.0	886	875	1.3	24.9
	38	24	5	20.8	1,370	1,640	-19.7	13.0
	39	14	3	21.4	838	888	-6.0	10.0
	40	10	2	20.0	587	656	-11.8	25.6
	41	10	3	30.0	599	579	3.4	8.6
	전체	207	46	22.2	12,439	12,813	-3.0	4.2

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
나 주 시	11	56	17	30.4	2,824	3,021	-7.0	6.9
	31	23	7	30.4	1,356	1,367	-0.8	9.0
	32	25	8	32.0	1,498	1,580	-5.5	4.3
	33	14	5	35.7	810	913	-12.8	6.4
	34	25	8	32.0	1,285	1,432	-11.4	8.8
	35	25	8	32.0	1,525	1,620	-6.2	4.5
	36	37	11	29.7	1,997	2,080	-4.1	9.5
	37	19	7	36.8	1,153	1,188	-3.0	6.6
	38	40	12	30.0	2,118	2,231	-5.3	7.3
	39	37	11	29.7	1,882	2,049	-8.9	10.3
	40	26	8	30.8	1,342	1,543	-15.0	8.8
	41	19	5	26.3	977	910	6.8	18.4
	42	38	13	34.2	2,151	2,351	-9.3	6.4
	51	46	15	32.6	2,837	2,928	-3.2	3.1
	52	32	8	25.0	1,707	1,512	11.4	13.8
	54	50	16	32.0	2,357	2,509	-6.4	8.6
	55	69	23	33.3	3,744	4,107	-9.7	5.8
	58	34	11	32.4	1,893	2,200	-16.2	7.8
	60	26	9	34.6	1,351	1,558	-15.3	5.7
	전체	641	202	31.5	34,807	37,101	-6.6	1.9

지역	동	N (개)	n (개)	f (%)	모집단 주택수	추정 주택수	차이율 (%)	RSE (%)
김천시	11	59	12	20.3	3,254	3,319	-2.0	9.0
	31	18	3	16.7	1,174	1,093	6.9	21.5
	32	22	4	18.2	1,333	1,370	-2.8	22.8
	33	15	3	20.0	1,121	1,137	-1.4	9.7
	34	30	6	20.0	1,737	1,599	8.0	15.8
	35	42	6	14.3	1,982	1,978	0.2	21.9
	36	24	4	16.7	1,348	1,283	4.8	22.9
	37	27	5	18.5	1,452	1,394	4.0	20.3
	38	11	2	18.2	842	816	3.1	5.2
	39	16	3	18.8	1,054	894	15.2	17.1
	40	21	5	23.8	1,689	1,642	2.8	8.1
	41	16	2	12.5	971	1,098	-13.1	28.9
	42	9	2	22.2	691	947	-37.1	14.1
	43	17	4	23.5	1,176	1,011	14.1	16.7
	44	10	2	20.0	601	695	-15.6	25.4
	55	35	7	20.0	1,923	2,007	-4.4	9.1
	56	157	31	19.7	8,142	7,837	3.7	7.0
	58	144	31	21.5	7,837	8,170	-4.3	4.8
	59	65	13	20.0	3,797	3,668	3.4	6.1
	60	61	12	19.7	3,014	3,202	-6.2	8.1
	61	72	15	20.8	3,404	3,634	-6.8	8.4
	전체	871	172	19.7	48,542	48,795	-0.5	2.7
영양군	11	44	18	40.9	2,073	2,025	2.3	4.7
	31	10	4	40.0	688	803	-16.6	9.8
	32	10	4	40.0	595	475	20.2	5.6
	전체	64	26	40.6	3,356	3,303	1.6	4.4

제3장 행정자료와 표본조사자료를 이용한 인구, 가구 및 주택 통계 산출

제1절 머리말

- 행정자료를 기반으로 하여 특정 시점 기준으로 작성된 인구, 가구, 주택 관련 통계에는 일반적으로 오류가 있음이 알려져 있다. 이러한 행정자료의 오류를 보정하기 위한 방안으로 2010년 인구주택총조사에서 수행되던 표본조사(전체 조사구의 약 10% 규모)와 유사하지만, 추출률이 더 큰 규모의 표본조사를 실시하여 각 부모집단(副母集團: subpopulation)별 행정자료 기반 인구, 가구, 주택 관련 통계를 보정하는 방안이 요구되고 있다.
- 현재, 우리나라에서 통계청은 2015년 등록센서스를 실시할 경우, 그 결과를 효율적으로 보정하고, 이를 추출틀(sampling frame)로 이용하여 10% 표본조사가 아닌 20% 표본조사를 고려하고 있다. 또한 20% 표본추출을 위하여 행정구역인 읍, 면, 동과 인구주택총조사의 표본추출 단위인 조사구(ED: Enumeration District)의 개념을 보완하여 새로운 물리적인 구역개념인 추출단위의 정의 및 이의 리스트가 작성되고 있다. 현재 계획 중인 등록센서스와 대규모 표본조사가 실행 될 경우 각 관심 부모집단별 인구수, 가구수 및 주택수의 추정을 위하여 사용될 수 있는 정보는 각 부모집단별 행정자료 통계 그리고 20% 표본조사 결과로부터 산출되어지는 각 부모집단별 추정량이다.
- 본 연구의 목적은 주어진 두 세트의 각 부모집단 별 인구수, 가구수 및 주택수에 관한 정보를 이용하여 등록센서스 자료에 부여되어 실제 인구수, 가구수, 주택수의 산출을 위해 사용될 수 있는 각 개인, 가구 그리고 주택별 가중치(weight)를 산출하는 것이다. 즉 행정자료를 이용한 각 관심 부모집단 별 인구수, 가구수 주택수가 각 개인, 가구 그리고 주택의 가중치를 1로 정의하고

산출되어짐을 감안하면서, 20% 표본으로부터 얻어지는 각 부모집단 별 불편 추정량(unbiased estimator)을 이용하여 각 개인, 가구 그리고 주택 별 가중치를 보정하고 이를 인구수, 가구수 그리고 주택수 산출을 위하여 사용하는 방안을 제시하려 한다.

- 모집단에 대한 보조정보(auxiliary information)가 주어진 경우, 이를 이용한 효율적인 추정량으로는 흔히 회귀추정량(regression estimator)이 고려된다. 흔히 언급되는 사후층화추정량(post-stratification estimator)은 회귀추정량의 특별한 경우에 해당하며, 대부분의 조사자료 분석을 위해 사용된다. 회귀추정량의 성질 중 중요한 것은 보조변수의 모집단 총계(population totals)나 모평균(population means) 같은 모수 추정을 위해 이를 사용할 경우, 주어진 보조 변수에 대한 모집단 보조정보와 추정량이 정확하게 일치하는 calibration 성질이다. 즉 변수 y 의 모집단 총합에 대한 회귀추정량을

$$\hat{t}_{y,reg} = \sum_{i \in S} w_i y_i \quad (3.1)$$

로 정의할 때 모집단 총합 t_x 가 주어진 보조변수 x 의 회귀추정량은 다음의 식을 만족한다.

$$\hat{t}_{x,reg} = \sum_{i \in S} w_i x_i = t_x \quad (3.2)$$

- 본 연구에서는 행정자료 기반 인구, 가구, 주택의 총수를 보정하기 위하여 20% 표본조사 결과를 보조정보로 사용하는 회귀추정 방안을 고려한다. 이 경우, 주어진 보조정보가 일반적으로 회귀추정량의 정의 시에 사용되는 모집단의 “참값”(true population value)이 아닌 “표본에 근거한 추정량”(sample-based estimator)이기 때문에 이의 변동을 고려한 추정량을 사용하여야 하고, 따라서 능형회귀추정량(Ridge regression estimator)을 본 연구에서는 고려하였다.
- 본 연구에서는 2011년 통계청의 학술연구 용역사업에서 고려된 능형회귀추정

방안을 가구수 및 주택수의 산출을 위해서도 적용했으며, 인구수 산출을 위해서는 3가지 가능한 조정셀을 통계청 등록센서스 담당 주무부서와 논의하여 정의한 후 이 3가지 방안을 12개 시범지역으로 확대하여 비교·검토하는 작업을 진행하였다.

제2절 능형회귀추정량(Ridge regression estimator)

- 본 절에서 정의되는 능형회귀추정량은 인구수, 가구수, 주택수의 산출을 위하여 각각 적용되며 따라서 인구수, 가구수, 주택수의 산출을 위한 가중치는 각 통계별로 독립적으로 작성된다.
- 행정자료 기반 인구수, 가구수, 주택수에 체계적인 오류가 존재한다는 가정 하에서 불편추정량인 20% 표본조사 자료 기반 인구수, 가구수, 주택수 추정량을 벤치마킹을 위한 보조정보로 사용할 수 있다. 행정자료에 포함되어 있는 각 단위별로 p 개의 지시함수로 구성되어 있는 벡터 $\underline{x}_i$ 를 고려하자. 즉

$$\underline{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$$

이며, 여기서

$$x_{ij} = \begin{cases} 1, & i \in A_j \\ 0, & i \notin A_j \end{cases}$$

으로 $(A_1, \dots, A_p)$ 는 행정자료에 포함된 단위인 개인, 가구 그리고 주택들로 이루어진 A 의 분할(partition)로서 $\bigcup_{j=1}^p A_j = A$ 이며 모든 $j \neq j'$ 에 대하여 $A_j \cap A_{j'} = \phi$ 이다. 행정자료를 이용한 각 최소 공표 단위 부모집단, A_j , $j = 1, \dots, p$ 별 인구, 가구, 주택의 수는 $t_{Ax_j} = \sum_{i \in A} x_{ij} = \sum_{i \in A_j} 1$ 이며 벡터 $\underline{t}_A = \sum_{i \in A} \underline{x}_i = (t_{Ax_1}, \dots, t_{Ax_p})$ 는 각 최소 공표 단위 부모집단 별 행정자료 기반 통계를 그 원소로 갖는다. 즉 행정자료 기반 각 최소 공표 단위 부

모집단 별 통계는 A 에 포함되는 모든 단위에게 가중치 1을 부여한 후 각 최소 공표 단위 부모집단별 가중치 합을 통하여 산출된다. 용어의 단순화를 위하여 최소 공표 단위 부모집단을 부모집단으로 정의한다. 각 단위 별 p 개의 지시함수 중 오직 하나의 지시함수만이 1의 값을 갖으며 나머지 지시함수들의 값은 0으로 정의된다.

- 행정자료 기반으로 산출된 인구, 가구 및 주택수의 보정을 위해서는 20% 표본 조사 결과를 통해 산출된 각 부모집단별 불편추정량을 사용한다. 20% 표본조사자료를 이용한 추정량은

$$\hat{\underline{t}}_s = \sum_s \pi_i^{-1} \underline{x}_i = (\hat{t}_{sx_1}, \dots, \hat{t}_{sx_p}) \quad (3.3)$$

이며, 여기서 π_i 는 i 번째 개체가 표본에 포함될 확률을 의미한다. 기존의 회귀추정량 혹은 사후총화추정량은 모집단 정보를 보조정보로 이용하여 표본으로부터 얻어지는 추정량을 보정하기 위하여 사용되었으나 본 연구에서는 일종의 모집단 정보인 행정자료로부터 산출된 $\underline{t}_A$ 을 보정하기 위하여 불편추정량인 $\hat{\underline{t}}_s$ 을 보조정보로 사용한다.

- 모집단 총합 $t_y = \sum_U y_i$ 에 대한 능형회귀추정량 $\hat{t}_{y, rreg}$ 는

$$\hat{t}_{y, rreg} = \sum_{i \in A} w_{i, rreg} y_i = t_{Ay} + (\hat{\underline{t}}_s - \underline{t}_A) \hat{\underline{\beta}}_{rid} \quad (3.4)$$

로 정의되며, 여기서

$$w_{i, rreg} = 1 + (\hat{\underline{t}}_s - \underline{t}_A) (X_A' X_A + \Lambda)^{-1} X_A' \underline{x}_i \quad (3.5)$$

$$\hat{\underline{\beta}}_{rid} = (X_A' X_A + \Lambda)^{-1} X_A' \underline{y}_A \quad (3.6)$$

이고 $t_{Ay} = \sum_A y_i$, $X_A = (\underline{x}_1', \dots, \underline{x}_{N_A}')'$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, N_A 는 행정자료 기반 전체 인구수, 가구수 및 주택수를, 그리고 U 는 특정시점 기준 실제 한국 전체 인구, 가구 및 주택의 집합을 나타낸다. 대각 행렬 Λ 의 모든 원소가 0인 경우, $\hat{t}_{y, rreg}$ 는 $\hat{t}_s$ 을 보조정보로 사용하는 회귀추정량, 즉 사후총화추정량과 동일하다. 능형회귀추정량 (3.4)를 정의하기 위하여 사용된 가중치 (3.5)는 다음의 함수

$$\sum_A (w_{i, rreg} - 1)^2 + \left(\sum w_{i, rreg} \underline{x}_i - \hat{t}_s \right) \Lambda^{-1} \left(\sum w_{i, rreg} \underline{x}_i - \hat{t}_s \right)' \quad (3.7)$$

를 최소화하는 w_i 로 정의된다. $\hat{\beta}_{rid}$ 는 일반적으로 능형회귀계수 추정량으로 알려져 있다.

- 제시된 능형회귀추정량은 회귀 혹은 사후총화추정량과 달리 행정자료로부터 산출된 $\underline{x}_i$ 의 총합을 구하기 위하여 사용될 경우, 보조정보인 $\hat{t}_s$ 과 일치하지 않는다. 즉 일반적으로 알려진 calibration 성질을 만족하고 있지 않다. 본 연구에서 능형회귀추정량을 고려한 이유는 바로 보조정보인 $\hat{t}_s$ 이 표본조사를 통하여 산출된 값으로서 비록 불편추정량이지만 하나 변동을 갖는 통계량이고 따라서 $\hat{t}_s$ 에 직접적으로 벤치마킹하는 것이 바람직하지 않기 때문이다.
- 능형회귀추정량의 계산을 위해서는 가중치 (3.5)의 산출에 필수적인 대각행렬 Λ 를 정의해야 한다. 능형회귀추정량의 정의를 위한 기존 연구로 Park and Cho (2008)는 y 와 x 사이의 선형회귀 모형 하에서 능형회귀추정량의 MSE를 최소화하는 Λ 를 제안하였으며 Rao and Singh (1997) Λ 의 정의를 위하여 Ridge Shrinkage Method를 고려하였다.
- 본 연구에서는 Λ 의 정의를 위하여 보조정보로 사용되는 $\hat{t}_s$ 의 통계적 정도(精

度: precision)를 나타내는 분산을 사용하였다. 즉 $\hat{t}_s$ 의 원소 중 그 분산이 큰 경우에는 해당 λ 의 대각원소 값을 크게 함으로 행정자료 결과에 대한 보정의 폭을 줄이고 반대로 분산이 작은 경우에는 해당 λ 의 대각원소 값을 0에 가깝게 부여하여 행정자료 결과에 대한 보정의 폭을 늘리는 방안을 고려하였다. 즉 각 보조변수 x_j , $j = 1, \dots, p$ 별 주어진 확률 α_j 에 대하여

$$\Pr \left[\left| \hat{t}_{x_j, rreg} - \hat{t}_{sx_j} \right| < \delta_j \right] \geq 1 - \alpha_j \quad (3.8)$$

를 만족하는 한계 값 δ_j 를 정의하고 정의된 δ_j 를 이용하여 (3.5)의 계산 시 필요한 λ 값을 정의하였다. (3.4)의 $\hat{t}_{x_j, rreg}$ 를 부모집단 별 행정자료 기반 통계, t_{Ax_j} 와 표본자료를 이용한 통계, $\hat{t}_{sx_j}$ 를 이용하여 표현하면 다음과 같다.

$$\hat{t}_{x_j, rreg} = \left(1 - \frac{t_{Ax_j}}{t_{Ax_j} + \lambda_j} \right) t_{Ax_j} + \frac{t_{Ax_j}}{t_{Ax_j} + \lambda_j} \hat{t}_{sx_j} \quad (3.9)$$

즉 각 부모집단 별 통계의 능형회귀추정량은 행정자료 기반 통계와 표본자료 기반 통계의 가중 평균으로 표현되며 부여되는 가중치는 λ 의 함수로 나타난다. 즉 λ 의 값이 커짐에 따라 능형회귀추정량은 행정자료 기반 통계에 접근하며 λ 의 값이 0에 가까워짐에 따라 능형회귀추정량은 표본자료 기반 불편 추정량으로 수렴한다.

- 식 (3.9)를 식 (3.8)에 적용한 결과는 다음과 같다.

$$\Pr \left[\left| \left(1 - \frac{t_{Ax_j}}{t_{Ax_j} + \lambda_j} \right) (t_{Ax_j} - \hat{t}_{sx_j}) \right| < \delta_j \right] \geq 1 - \alpha_j \quad (3.10)$$

여기서 t_{Ax_j} 와 λ_j 는 상수로 간주할 수 있으며 표본자료 기반 불편 추정량 $\hat{t}_{sx_j}$ 은 확률변수로서 그 분포는 근사적으로 $N(t_{x_j}, Var(\hat{t}_{sx_j}))$ 을 따르게 된다. 주

어진 근사분포의 가정 하에서 식 (3.10)을 만족하는 δ_j 의 상한 값을 구하면 다음과 같다.

$$\begin{aligned}\delta_j &= z_{\alpha_j/2} \left(1 - \frac{t_{Ax_j}}{t_{Ax_j} + \lambda_j} \right) \sqrt{\text{Var}(\hat{t}_{sx_j})} \\ &\leq z_{\alpha_j/2} \sqrt{\text{Var}(\hat{t}_{sx_j})}.\end{aligned}\quad (3.11)$$

여기서 $z_{\alpha_j/2}$ 는 표준정규분포의 상위 $\alpha_j/2$ 분위수이다. (3.11)에 주어진 δ_j 의 상한 값을 이용하여 (3.8)을 만족하는 λ_j 는 다음과 같이 정의될 수 있다.

$$\lambda_j = \begin{cases} \frac{\delta_j t_{Ax_j}}{\sqrt{t_{Ax_j} - \hat{t}_{sx_j}} - \delta_j}, & \text{if } \sqrt{t_{Ax_j} - \hat{t}_{sx_j}} > \delta_j \\ 0, & \text{otherwise} \end{cases} \quad (3.12)$$

즉, 능형회귀추정량 $\hat{t}_{x_j, rreg}$ 의 가중치 (3.5)를 위한 λ 를 결정하기 위하여 고려된 (3.8)의 $\sqrt{t_{x_j, rreg} - \hat{t}_{sx_j}}$ 를 결국 (3.10)과 같이 $\sqrt{t_{Ax_j} - \hat{t}_{sx_j}}$ 로 표현된다. $\sqrt{t_{Ax_j} - \hat{t}_{sx_j}} \geq \delta_j$ 만족할 때, 즉 $\lambda_j = 0$ 인 경우에는 $\hat{t}_{x_j, rreg} = \hat{t}_{sx_j}$ 로 표현되며 λ_j 값이 클수록 $\hat{t}_{x_j, rreg} \approx t_{Ax_j}$ 는 것을 알 수 있다.

- 각 부모집단 별로 정의되는 λ_j 는 표본조사 자료 기반 부모집단 추정량의 분산에 비례하는 값으로 분산이 커질수록 정의되는 능형회귀추정량은 행정자료 결과에 근접하게 된다.
- 본 연구에서 제시되는 추정량은 기본적으로 표본조사 자료 기반 추정량이 불편 추정량이며 동시에 일치 추정량인 사실에 기인하여 그 분산이 작은 경우, 실제 참 모수의 값과 일치하게 되는 통계적 이론에 그 사용의 근거를 두고 있다.

제3절 12개 시범지역에 대한 인구, 가구 및 주택자료 분석 결과

- 본 절에서는 통계청의 2012년 학술연구 용역사업을 위한 12개 시범지역 자료를 이용하여 행정자료의 각 부모집단 별 인구, 가구, 주택 통계의 보정을 위해 제2절에서 제시된 능형회귀추정량을 2010년 인구주택총조사 결과와 가장 유사한 불편 추정량을 제공하는 표본(최량표본)을 통해 평가하고자 한다. 표본추출을 위해서는 12개 시범지역의 행정자료를 2010년 인구주택총조사 조사구에 매칭(연계)시켜 작성된 행정자료 기반 조사구의 리스트가 추출틀로 사용되었다. 추출된 조사구 별 통계는 2010년 인구주택총조사(이하 센서스로 표기) 결과를 참값으로 간주하여 사용하였다. 12개 시범지역에 대한 내용은 다음의 <표 3-1>과 같다.

<표 3-1> 12개 시범지역

시도코드	시도명	시군구코드	시군구명	동수
11	서울특별시	11180	금천구	10
		11190	영등포구	18
23	인천광역시	23080	서구	18
24	광주광역시	24010	동구	13
31	경기도	31240	화성시	23
32	강원도	32380	양구군	5
33	충청북도	33390	증평군	2
35	전라북도	35350	임실군	12
36	전라남도	36040	나주시	19
		36480	신안군	14
37	경상북도	37030	김천시	21
		37430	울릉군	3
총 합계				158

- 12개 시범지역에 대한 인구수, 가구수 및 주택수 통계를 위한 능형회귀추정량은 식 (3.9)와 같이 특정 부모집단에서 행정자료의 총합과 조사자료의 총합추정량의 가중평균(weighted mean)으로 정의되므로, 능형회귀추정량을 정의하

기 위한 부모집단의 결정이 선행되어야 한다. 여기서의 부모집단은 최소 공표단위 부모집단과는 다른 개념으로 능형회귀 추정방안의 적용을 위하여 구성하는 일종의 조정 셀로 정의하도록 한다. 식 (3.5)의 능형회귀 가중치는 다음과 같이 다시 표현할 수 있다.

$$w_{i,reg} = 1 + \frac{\hat{t}_{sx_j} - t_{Ax_j}}{t_{Ax_j} + \lambda_j} \quad (3.13)$$

여기서, j 는 i 번째 개체가 속한 부모집단으로 가중치 보정을 위해 미리 정의된 조정 셀을 의미한다. 이하 표기의 혼동을 방지하기 위하여, 부모집단은 최소 공표단위 부모집단으로 정의하며, 조정 셀은 능형회귀추정방안을 적용하기 위하여 정의된 부모집단으로 정의한다. $w_{i,reg}$ 는 각 조정 셀 j 안에서 동일한 능형회귀 가중치가 산출되므로 이는 각 조정 셀 별 인구수, 가구수, 주택수 산출 시 부여되는 개인, 가구, 주택에 부여되는 일종의 조정계수가 된다.

1. 인구

- 읍면동별*성별*연령5세별로 정의된 부모집단별 인구수 산출을 위한 능형회귀 추정방안은 통계청과 논의하여 가능한 세 가지 조정 셀을 정의한 후, 이 세 가지 방안을 12개 시범지역에서 대하여 비교하였다. 관심 부모집단별 인구수 산출을 위한 세 가지 조정 셀 구성방안은 다음과 같다.

- 1) 방안 1 - 읍면동별(158)*성별(2)*연령10세별(11) 조정 셀 구성
- 2) 방안 2 - 읍면동별(158)*성별(2)*연령3그룹(14세이하/15세~64세/65세 이상)별(3) 조정 셀 구성
- 3) 방안 3 - 동부/읍부/면부별(3)*성별(2)*연령10세별(11) 조정 셀 구성

- 능형회귀추정 방안 1, 2, 3을 통하여 각 조정 셀에서 보정된 개인 가중치의 분포는 다음의 <표 3-2>와 같다. 세 가지 방안으로 보정된 가중치의 평균은 모

두 ‘1’보다 작은 값으로 행정자료가 평균적으로 과다집계(overcount)됨을 보여주고 있다. 또한 세 가지 능형방안의 5%에서 95%까지의 분포는 0.8~1.0 사이의 값으로 비슷하게 나타나고 있으며, 하위 5%와 상위 5% 부분에서 차이를 보여주고 있다. 이러한 차이는 조정 셀을 구성함에 있어서 방안 1에서 가장 많은 셀을 그리고 방안 3에서 가장 적은 셀을 고려한 결과이다.

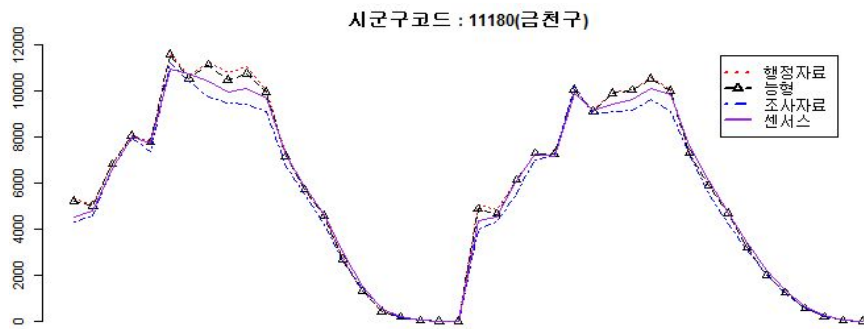
<표 3-2> 12개 시범지역의 인구 능형회귀 가중치의 분포

능형회귀추정	mean	sd	min	5%	10%	Q1	median	Q3	95%	max
방안1	0.976	0.0926	0.088	0.830	0.925	1.000	1.000	1.000	1.000	4.229
방안2	0.979	0.0792	0.294	0.840	0.929	1.000	1.000	1.000	1.000	3.310
방안3	0.939	0.0683	0.581	0.820	0.874	0.932	0.948	0.980	1.000	1.116

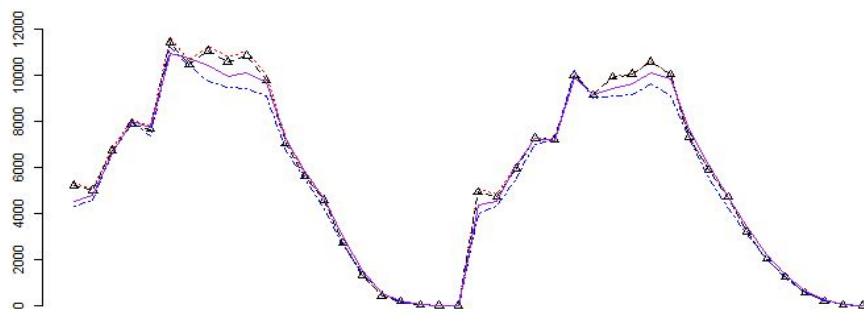
- 센서스와 행정자료의 총합 그리고 표본자료와 능형회귀추정 방안1, 2, 3의 총합 추정량의 성, 연령, 시군구, 성*연령, 성*시군구, 시군구*연령 그리고 시군구*성*연령 별 인구수 분포는 <부록 표 3-1>부터 <부록 표 3-7>을 참고하면 된다.
- 읍면동별(158)*성별(2)*연령5세별(21)로 구성된 총 6,636개의 관심 부모집단 별 센서스와 행정자료의 인구수 그리고 조사자료와 능형방안 1, 2, 3을 통하여 추정한 인구수는 다음의 <그림 3-1>부터 <그림 3-12>와 같다. 여기서, x 축은 각 시군구에 속하는 부모집단을 나타낸다. 총 6,636개의 부모집단 중 303개는 해당 부모집단의 인구수가 센서스, 행정자료, 조사자료 그리고 능형회귀추정 방안 1, 2, 3에서 모두 인구수가 ‘0’으로 나타나 다음의 그림에서는 제외되었다.
- 센서스 인구를 참값(true value)으로 간주하고 행정자료를 살펴보면 특정 부모집단에서 행정자료의 인구수와 센서스의 인구수의 차이가 크게 나는 것을 확인할 수 있으며, 따라서 행정자료의 보정이 필요함을 알 수 있다. 또한 조사자료의 추정 인구수 역시 특정 부모집단에서 센서스의 인구수와 그 차이가 크게 나타남을 확인할 수 있다. 따라서 행정자료의 인구수와 조사자료의 인구수의 차이가 큰 곳에서 능형회귀추정 방안을 통한 보정은 적절해 보이며, 평

균적으로 능형회귀추정 방안을 통한 인구수 추정 결과가 센서스 결과와 차이가 작음을 확인할 수 있다. 그리고 능형회귀추정량을 정의하기 위한 조정 셀과 관심 부모집단이 다르게 나타남에 따라 능형회귀추정 방안 1, 2, 3의 결과가 항상 행정자료 인구수와 조사자료 추정 인구수 사이에 존재하지는 않음을 확인할 수 있다.

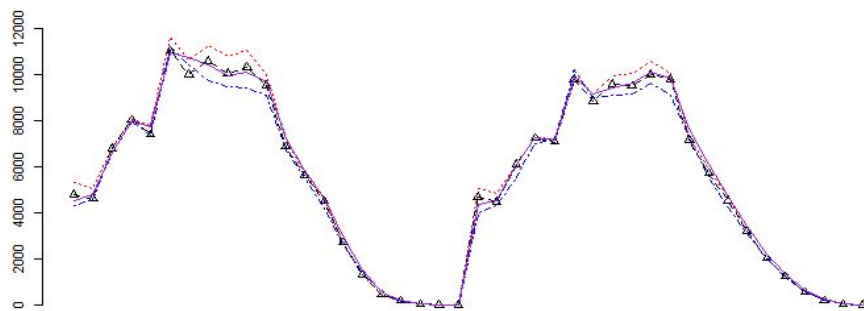
<그림 3-1> 금천구에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

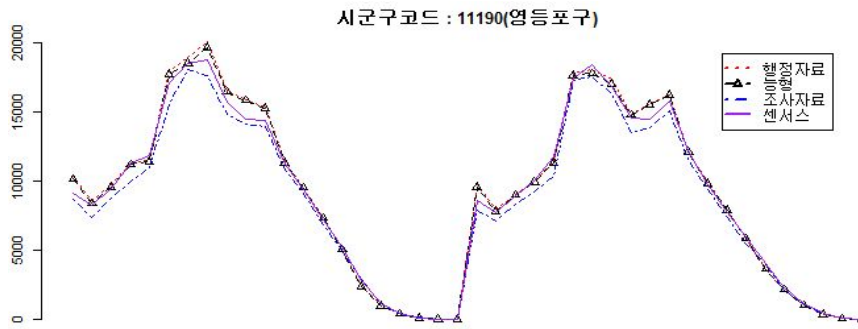


능형2총합추정량의 비교

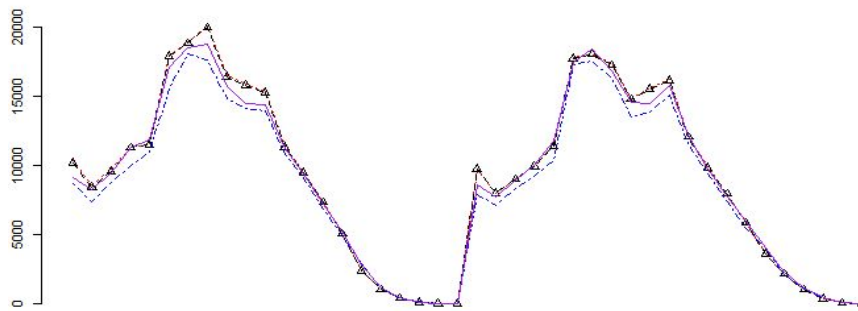


능형3총합추정량의 비교

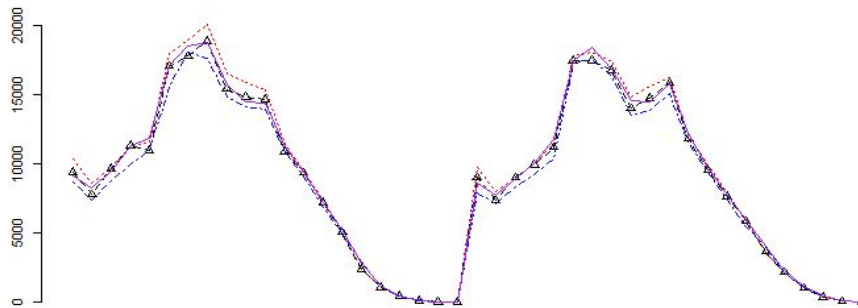
<그림 3-2> 영등포구에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

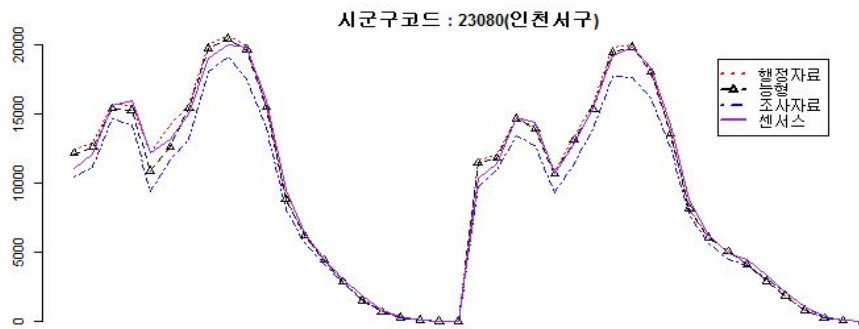


능형2총합추정량의 비교

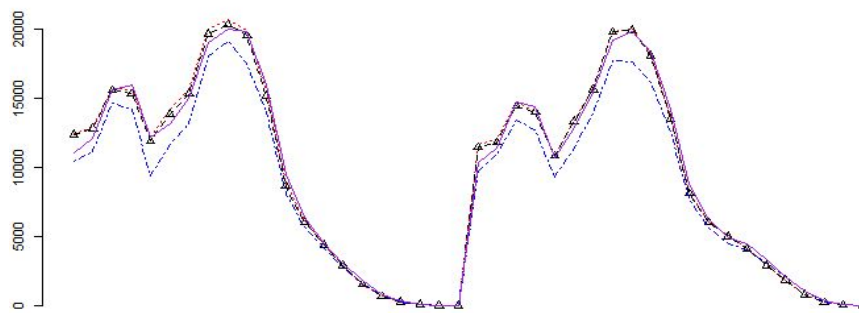


능형3총합추정량의 비교

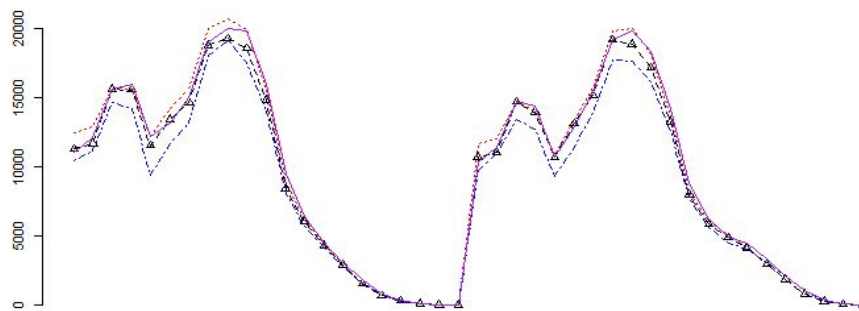
<그림 3-3> 인천서구에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

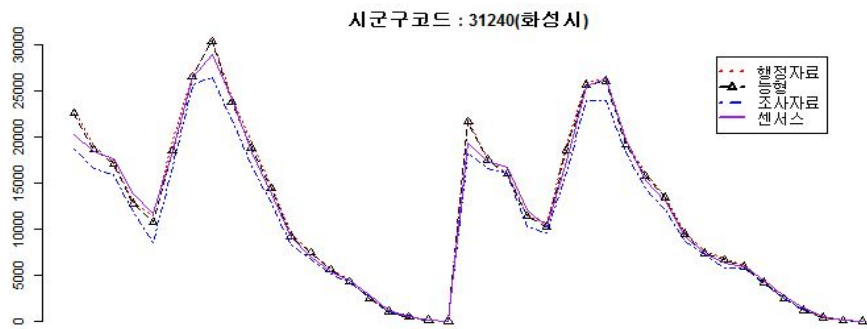


능형2총합추정량의 비교

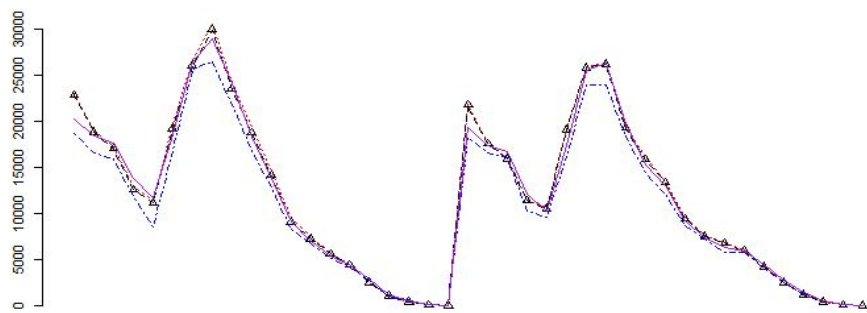


능형3총합추정량의 비교

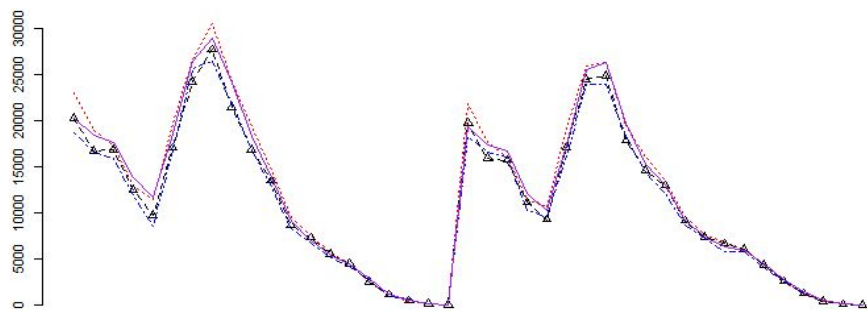
<그림 3-4> 화성시에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

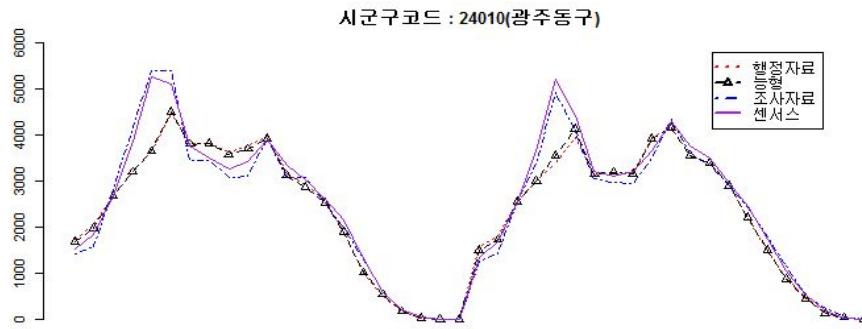


능형2총합추정량의 비교

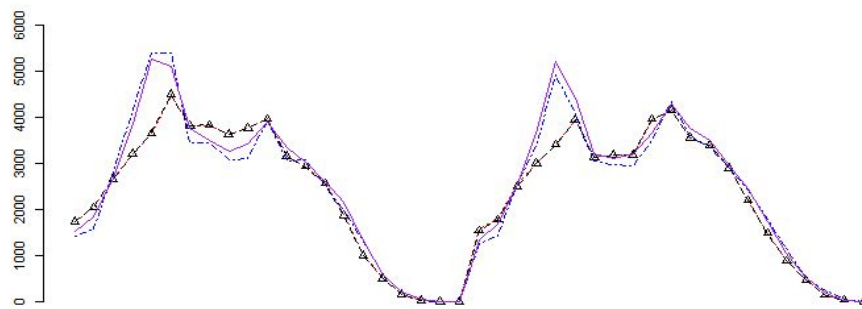


능형3총합추정량의 비교

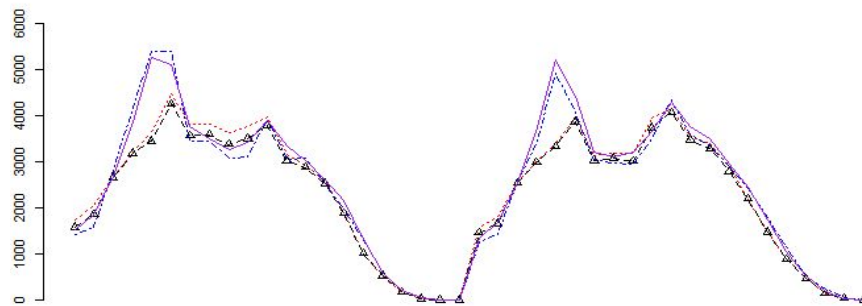
<그림 3-5> 광주동구에 포함된 부모집단별 인구수 비교



능형1출합추정량의 비교

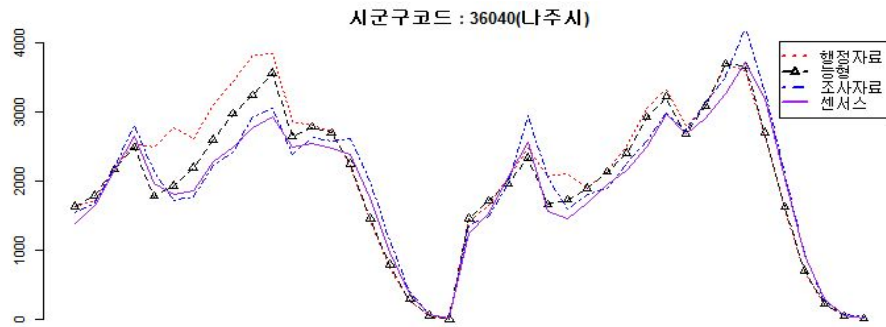


능형2출합추정량의 비교

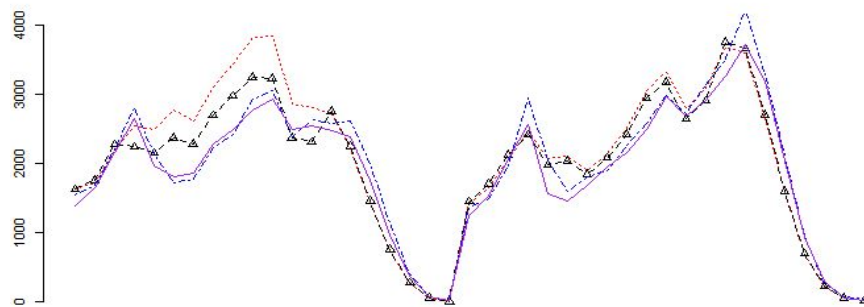


능형3출합추정량의 비교

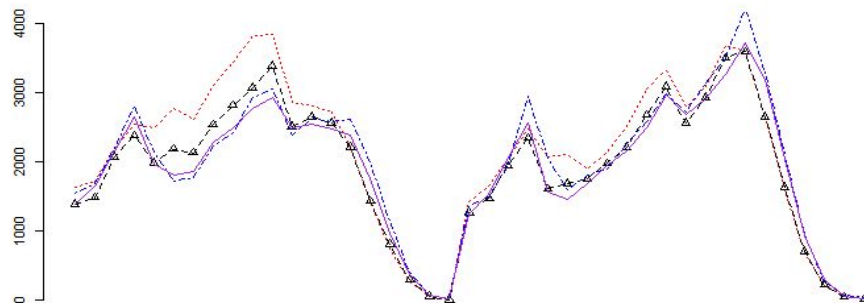
<그림 3-6> 나주시에 포함된 부모집단별 인구수 비교



능형1출합추정량의 비교

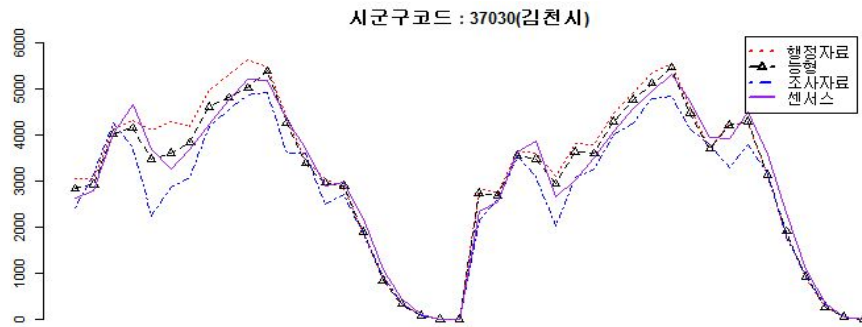


능형2출합추정량의 비교

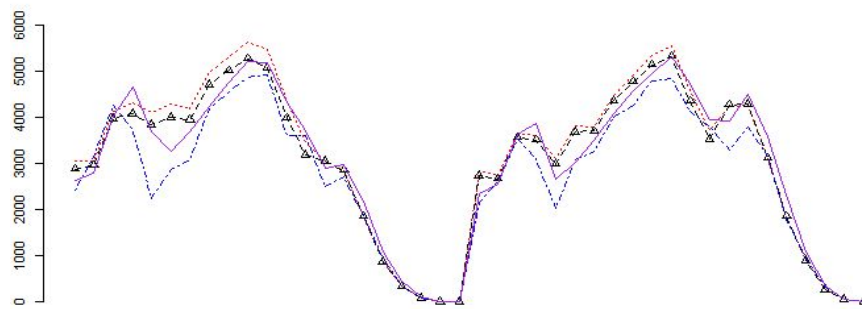


능형3출합추정량의 비교

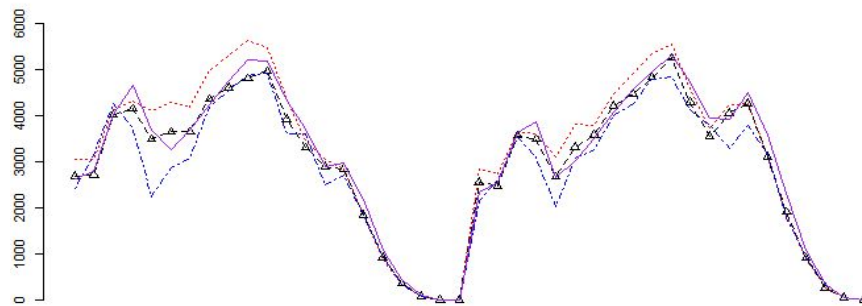
<그림 3-7> 김천시에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

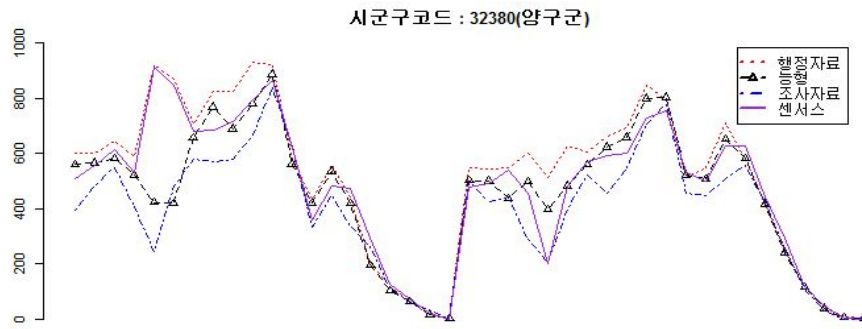


능형2총합추정량의 비교

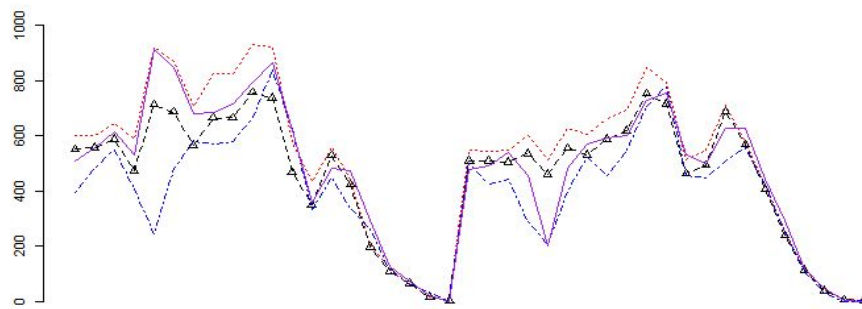


능형3총합추정량의 비교

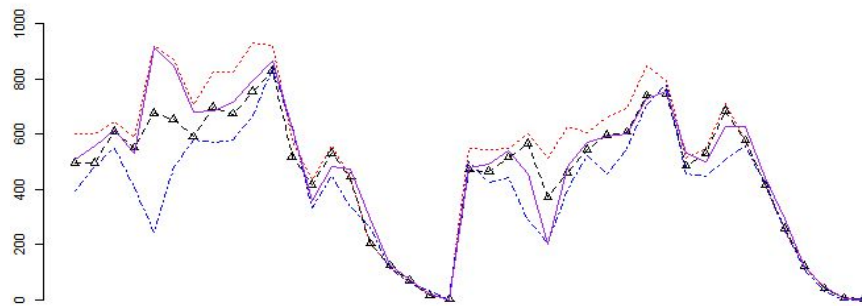
<그림 3-8> 양구군에 포함된 부모집단별 인구수 비교



능형1출합추정량의 비교

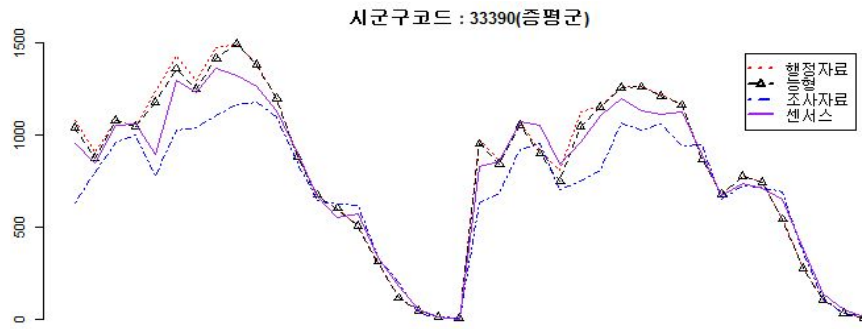


능형2출합추정량의 비교

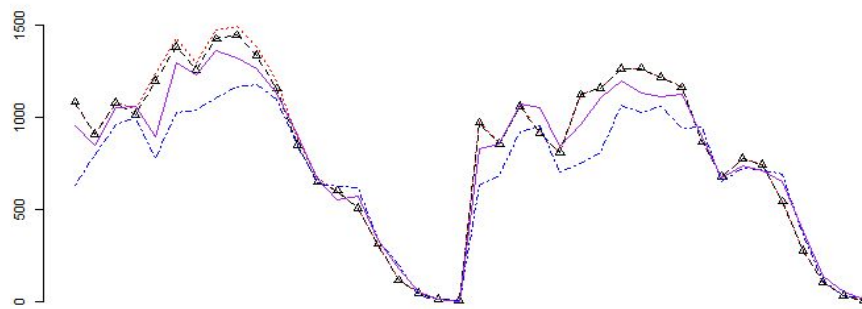


능형3출합추정량의 비교

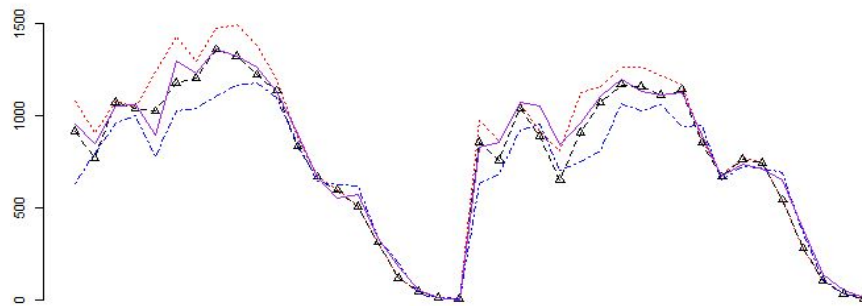
<그림 3-9> 증평군에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

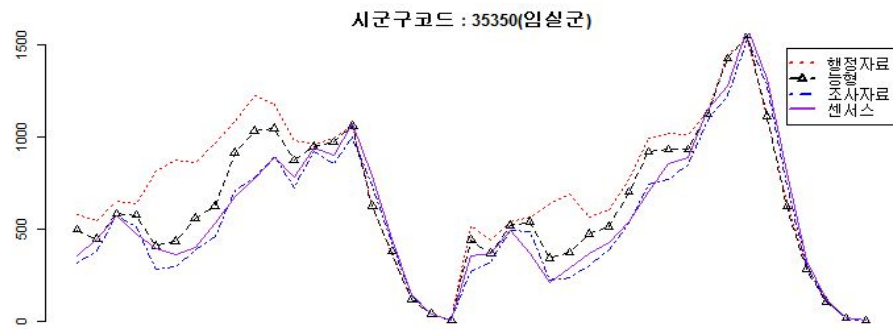


능형2총합추정량의 비교

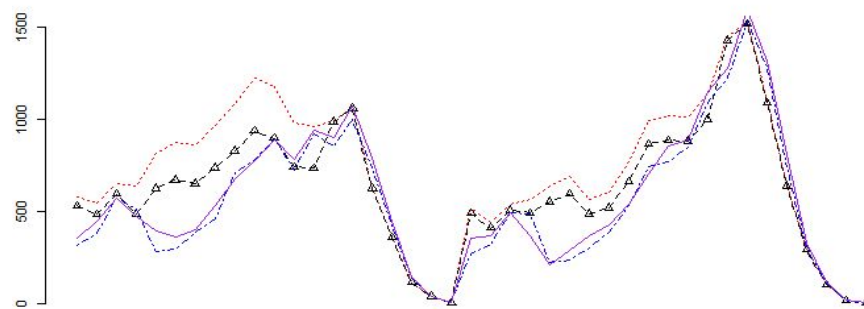


능형3총합추정량의 비교

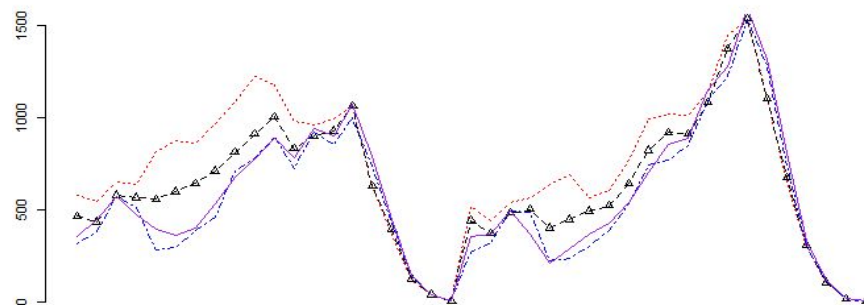
<그림 3-10> 임실군에 포함된 부모집단별 인구수 비교



능형1출합추정량의 비교

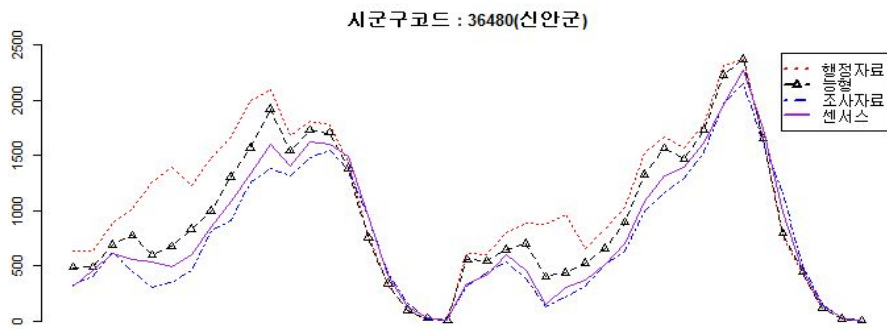


능형2출합추정량의 비교

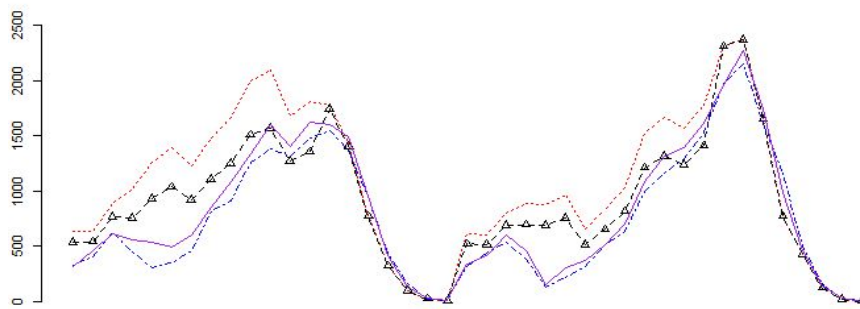


능형3출합추정량의 비교

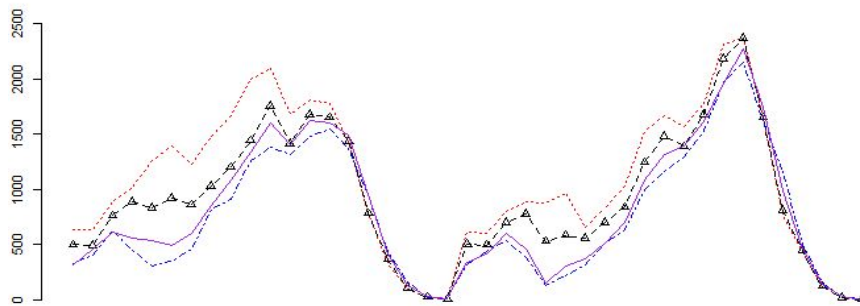
<그림 3-11> 신안군에 포함된 부모집단별 인구수 비교



능형1총합추정량의 비교

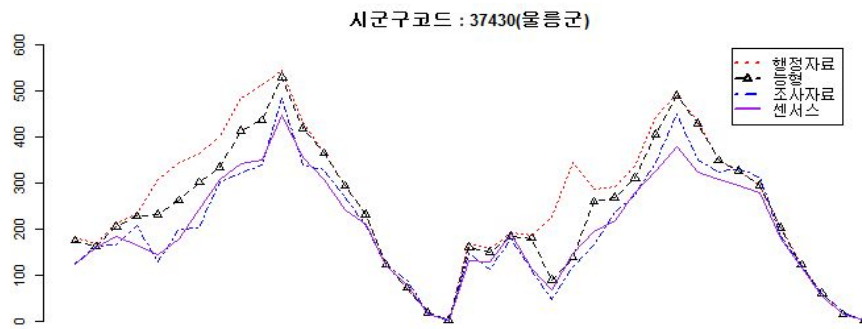


능형2총합추정량의 비교

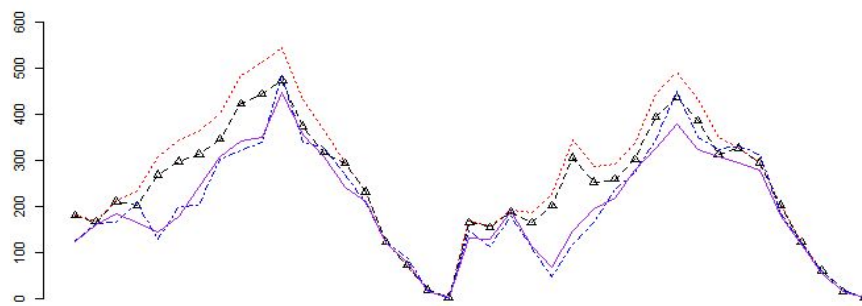


능형3총합추정량의 비교

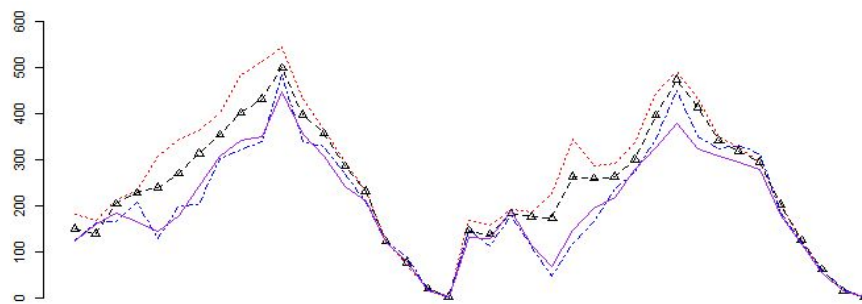
<그림 3-12> 울릉군에 포함된 부모집단별 인구수 비교



능형1출합추정량의 비교



능형2출합추정량의 비교



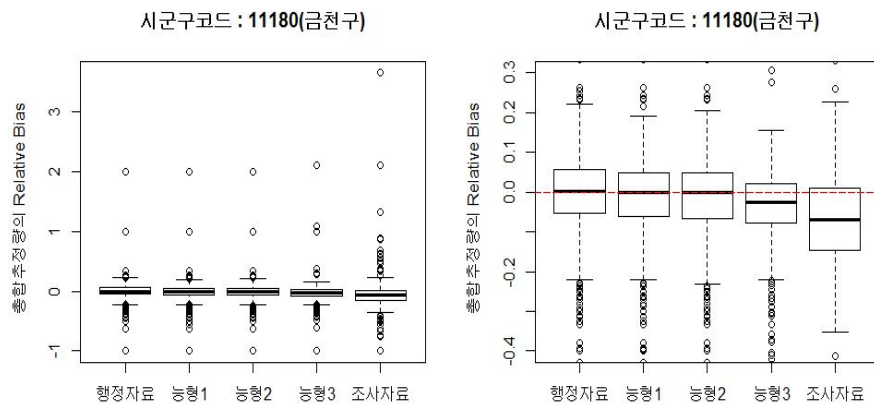
능형3출합추정량의 비교

- 능형회귀추정 방안을 사용한 각 부모집단별 인구수 추정 결과를 행정자료와 조사자료 결과와 비교하기 위하여 *Rel.Bias*(Relative Bias : 상대편의)를 다음과 같이 정의하였다.

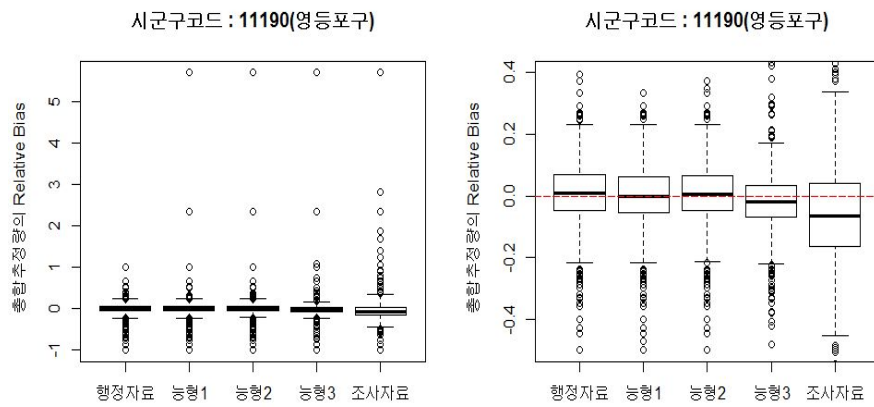
$$Rel.Bias = \frac{Bias}{True\ Value} = \frac{Estimator\ of\ Total - Census\ Total}{Census\ Total} \quad (3.14)$$

여기서, 센서스총합(Census Total)을 참 값으로 하여, 총합추정량(Estimator of Total)으로 행정자료와 조사자료 그리고 세 가지 능형회귀추정 방안 통하여 산출된 인구수 혹은 추정 인구수를 고려할 수 있다. 다섯 가지 총합추정량의 *Rel.Bias*는 센서스총합이 '0'으로 *Rel.Bias*가 정의되지 않는 부모집단을 제외한 총 6,266개의 부모집단에 대하여 계산되었다. 다음의 <그림 3-13>부터 <그림 3-24>는 각 시군구 안에 있는 관심 부모집단의 총합추정량의 *Rel.Bias*의 분포를 상자그림으로 표현한 것으로 좌측의 그림은 전체 분포이며, 우측의 그림은 분포의 중심을 중점으로 표현한 것이다. 특정 시군구에서 행정자료나 조사자료의 *Rel.Bias*는 평균적으로 '0'에 가깝게 나타나거나 '0'과 큰 차이를 보이는 시군구가 존재하는 반면, 능형회귀추정 방안1, 2, 3은 전체 시군구에 걸쳐 행정자료와 조사자료의 *Rel.Bias*보다 평균적으로 '0'으로 부터의 차이가 크게 나타나고 있지 않다.

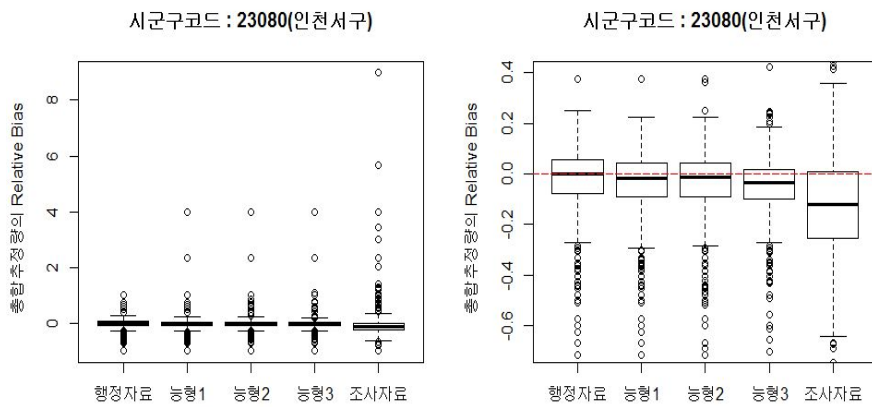
<그림 3-13> 금천구에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



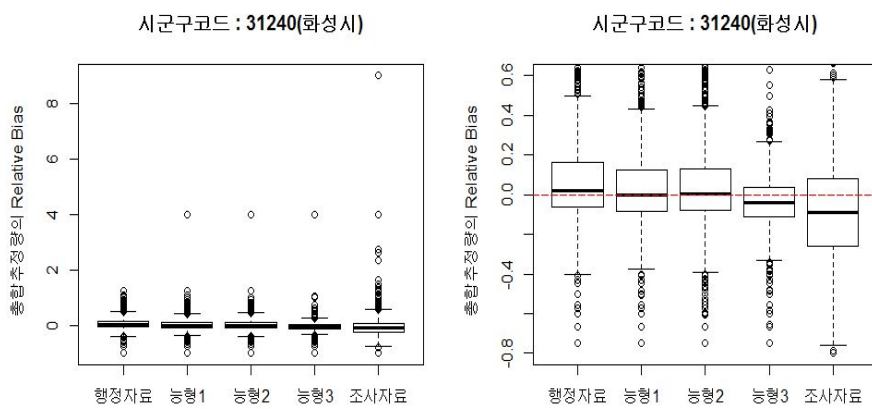
<그림 3-14> 영등포구에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



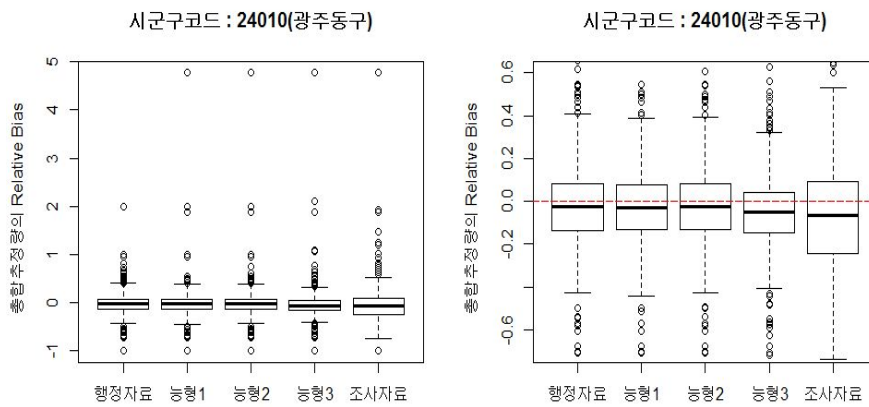
<그림 3-15> 인천서구에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



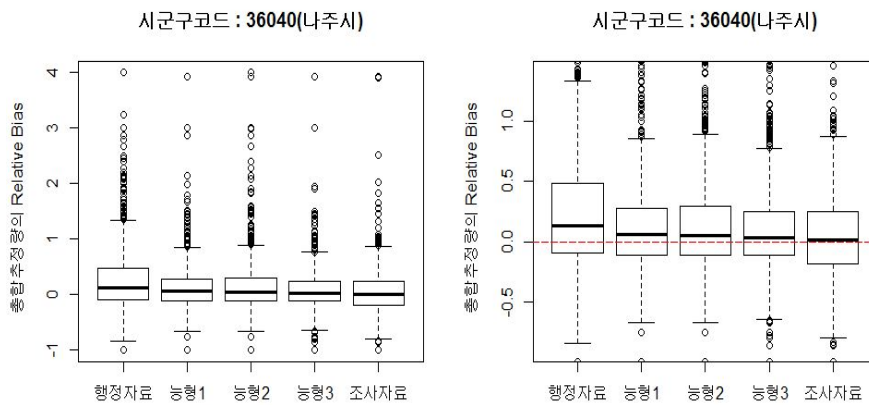
<그림 3-16> 화성시에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



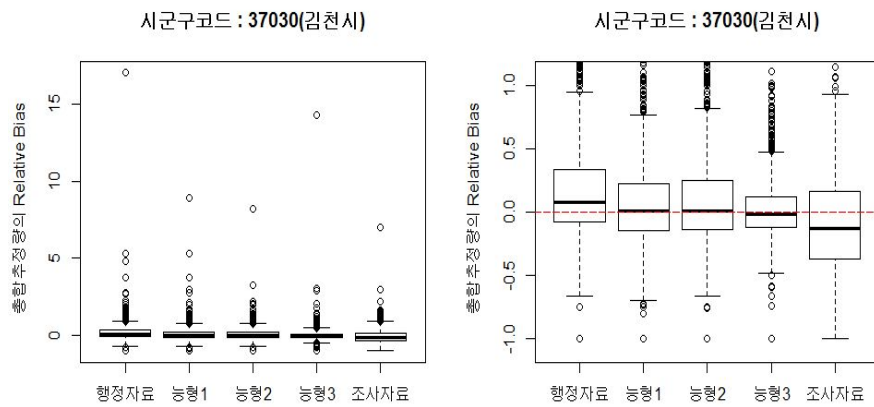
<그림 3-17> 광주동구에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



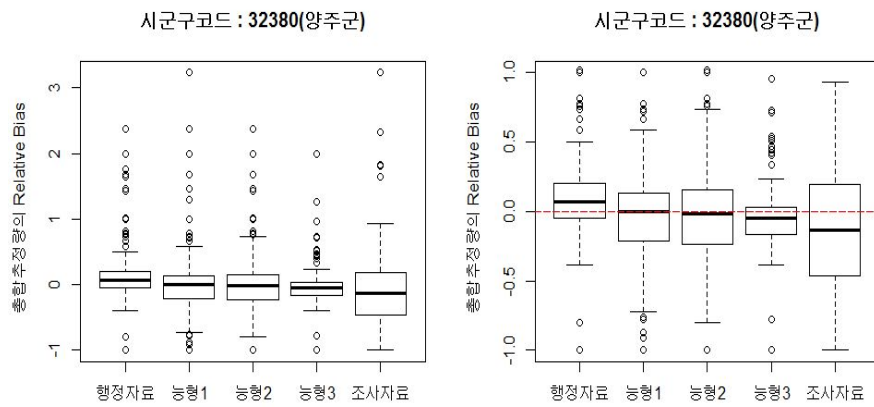
<그림 3-18> 나주시에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



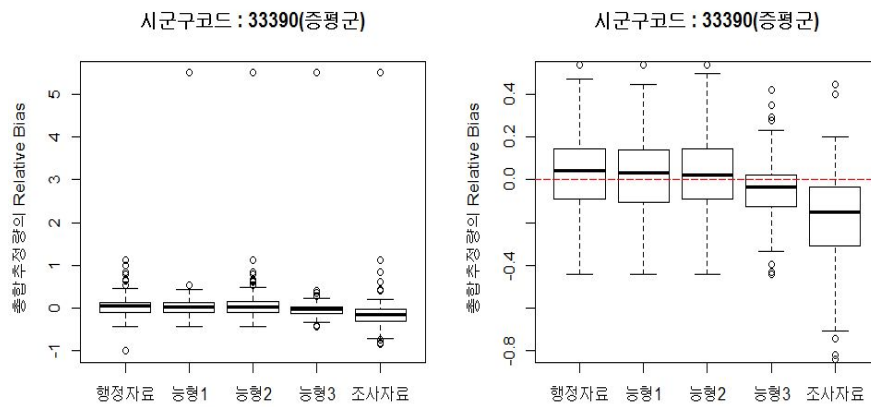
<그림 3-19> 김천시에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



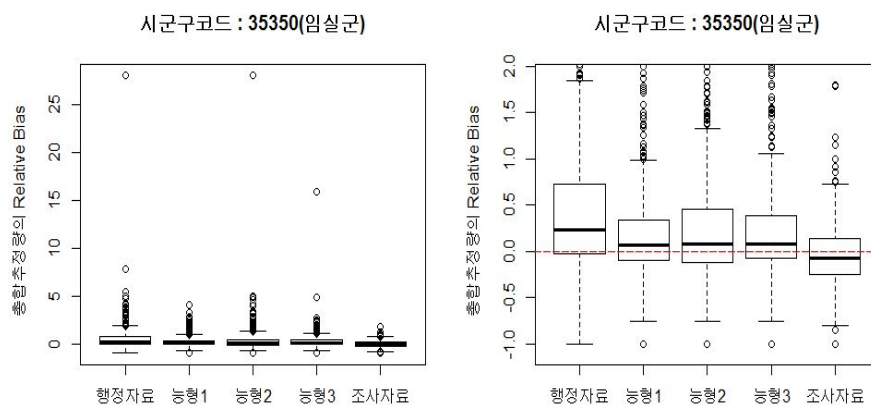
<그림 3-20> 양주군에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



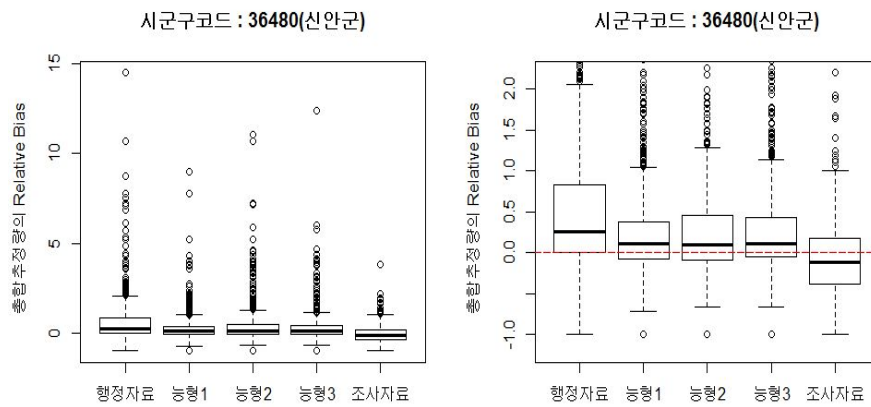
<그림 3-21> 증평군에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



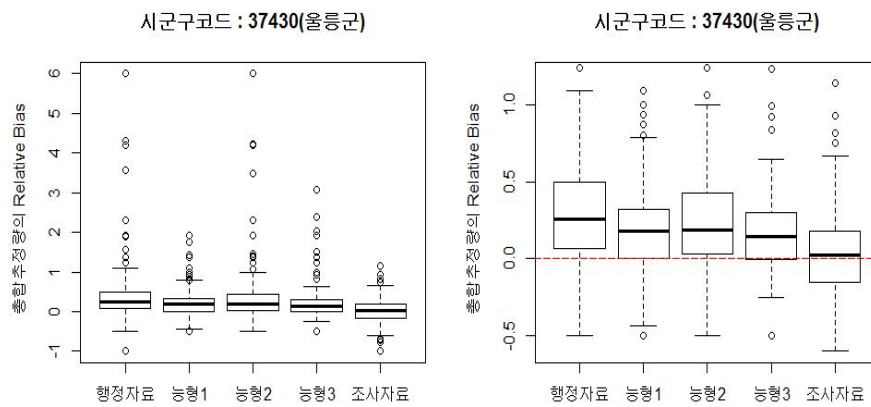
<그림 3-22> 금천구에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림



<그림 3-23> 신안군에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림

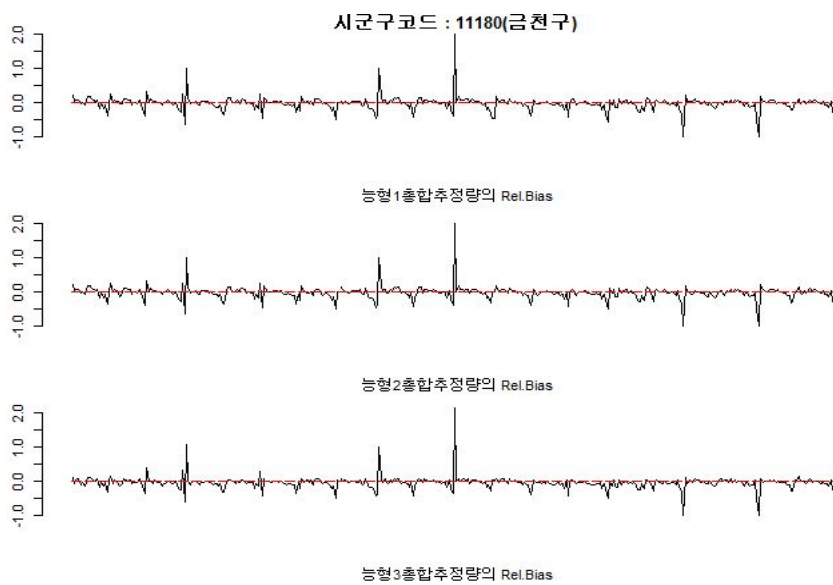


<그림 3-24> 울릉군에 포함된 부모집단의 인구추정량의 *Rel.Bias* 상자그림

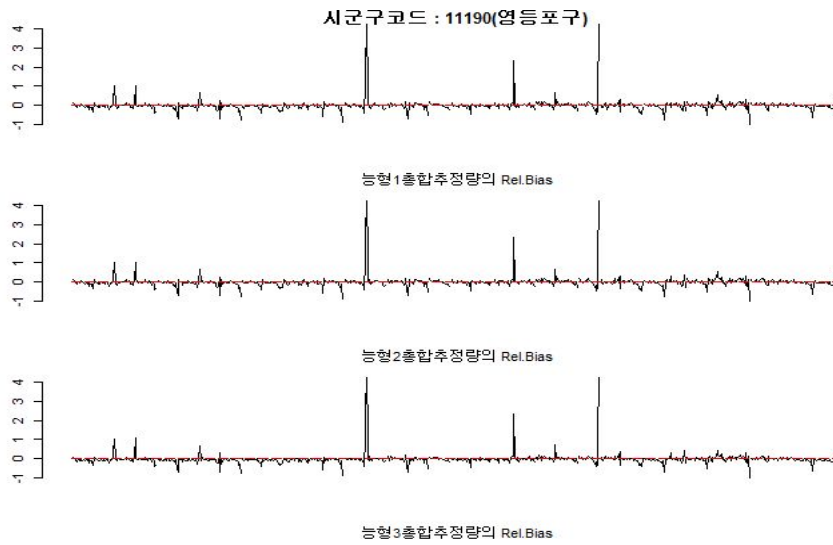


- 능형회귀추정 방안 3가지를 중심으로 비교하고자 다음의 <그림 3-25>부터 <그림 3-36>와 같이 능형회귀추정 방안1, 2, 3의 *Rel.Bias*를 각 시군구 안에 있는 관심 부모집단별로 표현하였다. 여기서 x 축은 각 시군구 안에 있는 관심 부모집단을 의미한다. 특정한 능형회귀추정 방안이 다른 방안에 비하여 *Rel.Bias*가 항상 작거나 크게 나타나지 않고 부모집단에 따라 다르게 나타나는 것을 확인 할 수 있다. 그러나 세 가지 방안의 *Rel.Bias*의 패턴은 비슷한 모양을 보여주고 있다. 그러나 방안 2의 경우 특정 부모집단에서 그 같이 다른 방안에 비하여 크게 나타나는 곳이 나타나고 있어, 방안1과 3에 비하여 불안정적인 양상을 보이고 있다.

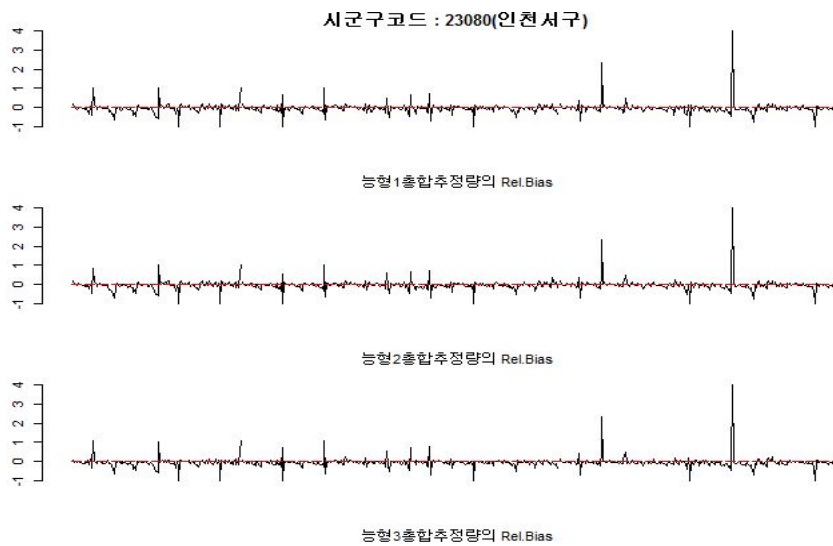
<그림 3-25> 금천구에 포함된 부모집단별 인구수 *Rel.Bias* 비교



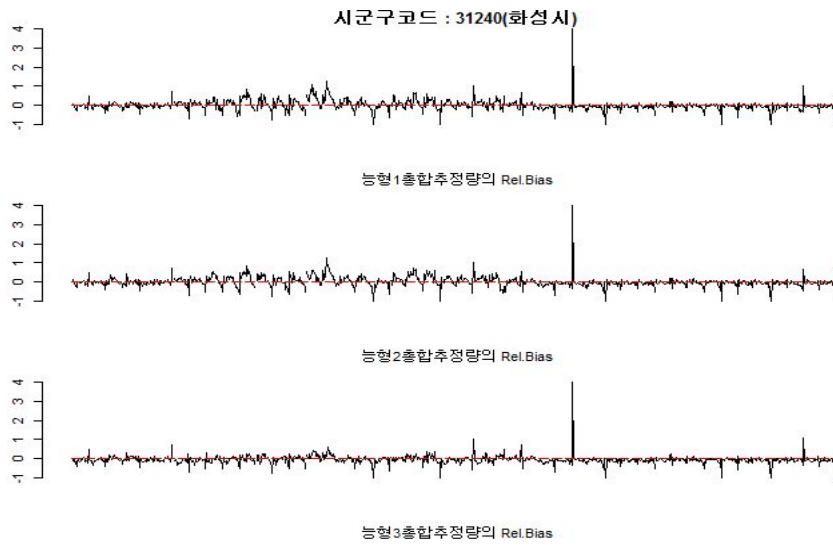
<그림 3-26> 영등포구에 포함된 부모집단별 인구수 *Rel.Bias* 비교



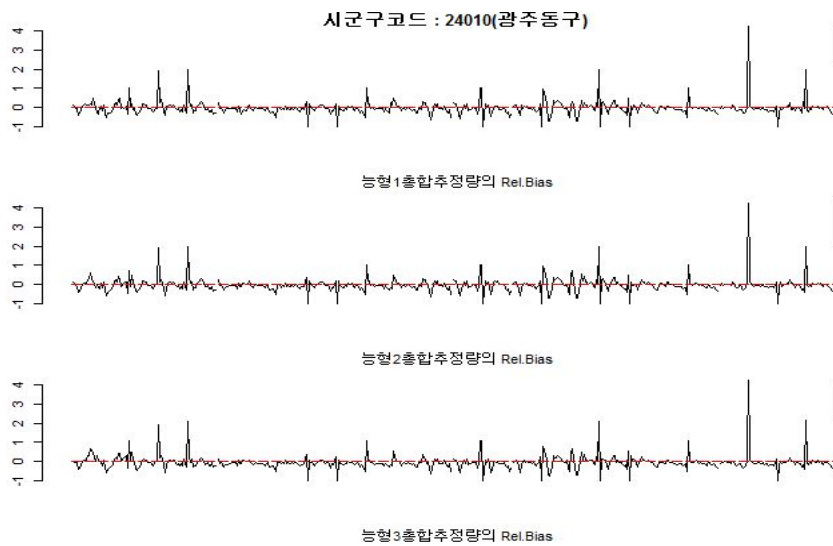
<그림 3-27> 인천서구에 포함된 부모집단별 인구수 *Rel.Bias* 비교



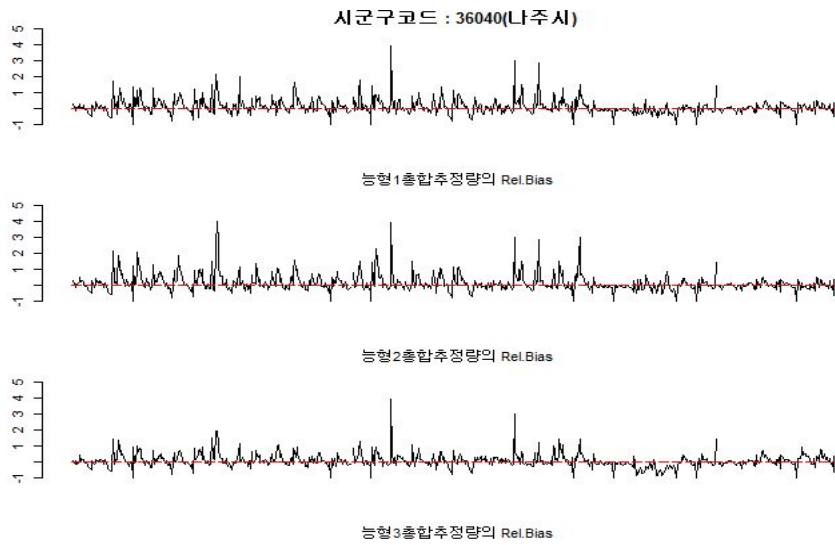
<그림 3-28> 화성시에 포함된 부모집단별 인구수 *Rel.Bias* 비교



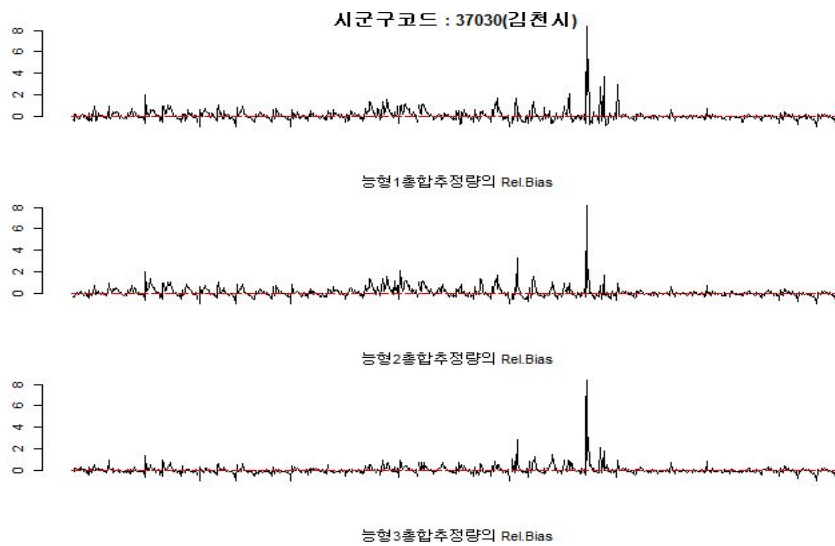
<그림 3-29> 광주서구에 포함된 부모집단별 인구수 *Rel.Bias* 비교



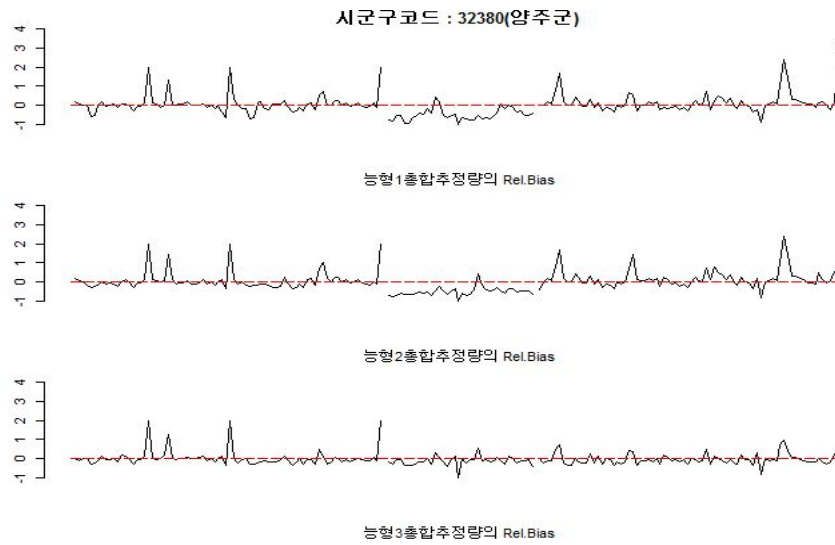
<그림 3-30> 나주시에 포함된 부모집단별 인구수 *Rel.Bias* 비교



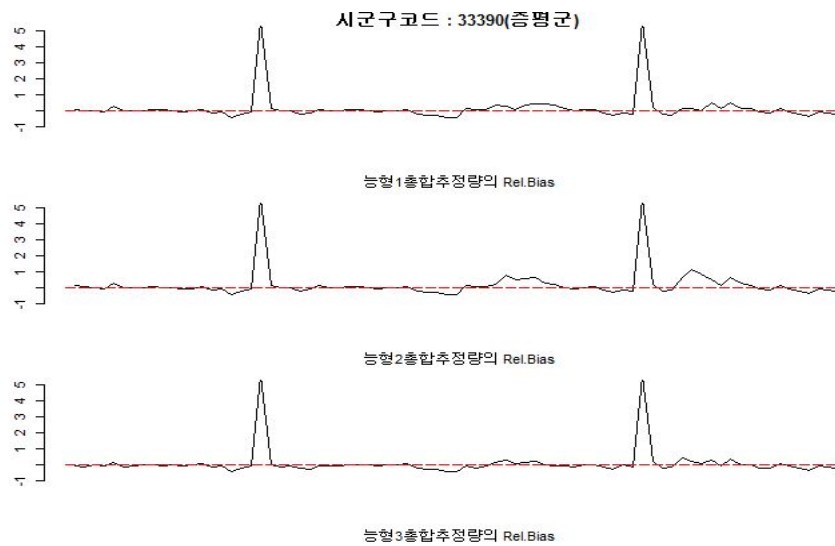
<그림 3-31> 김천시에 포함된 부모집단별 인구수 *Rel.Bias* 비교



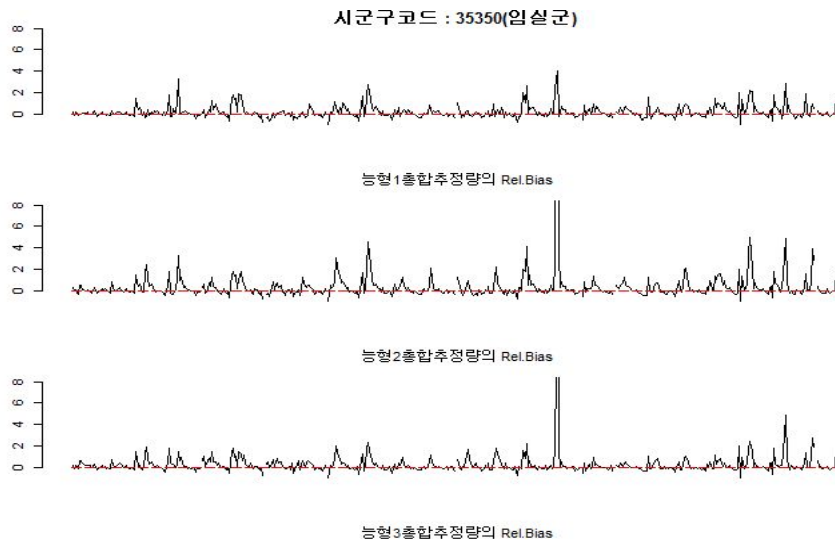
<그림 3-32> 양구군에 포함된 부모집단별 인구수 *Rel.Bias* 비교



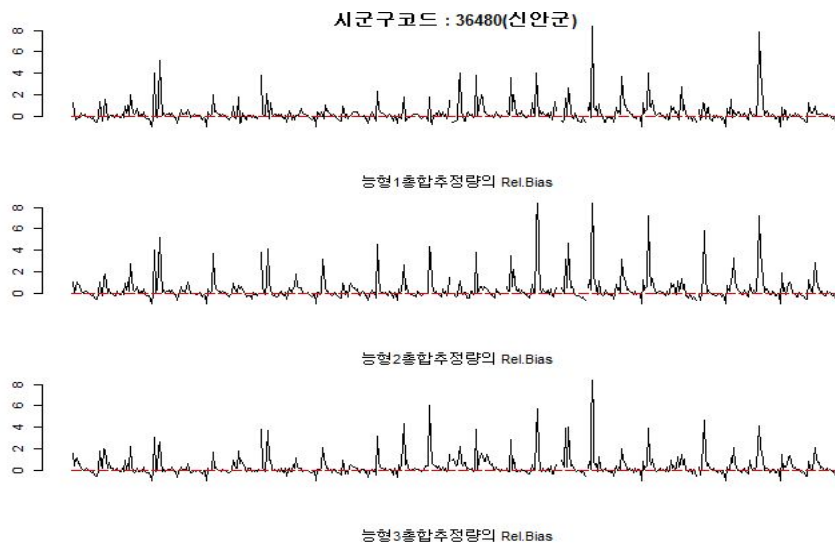
<그림 3-33> 증평군에 포함된 부모집단별 인구수 *Rel.Bias* 비교



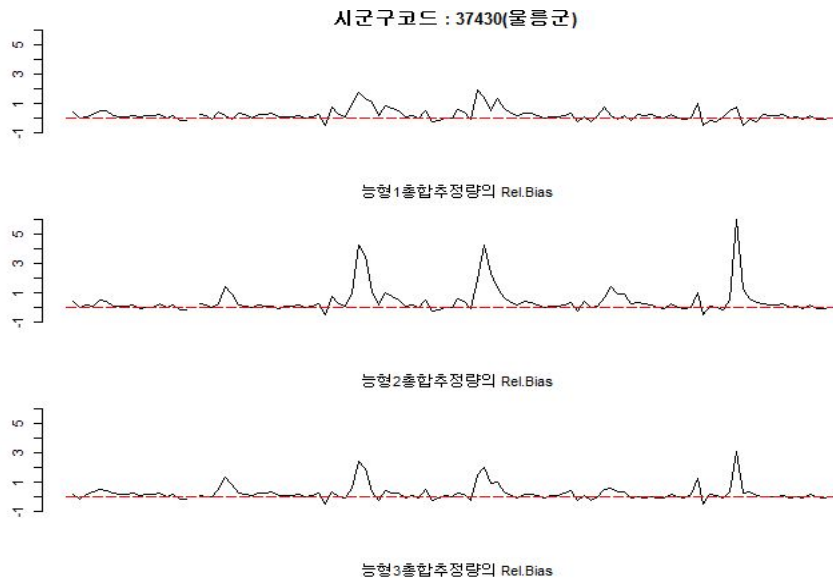
<그림 3-34> 임실군에 포함된 부모집단별 인구수 *Rel.Bias* 비교



<그림 3-35> 신안군에 포함된 부모집단별 인구수 *Rel.Bias* 비교

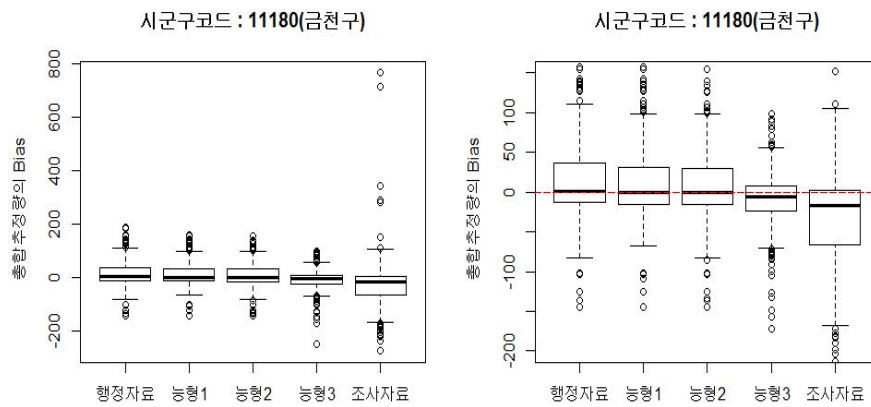


<그림 3-36> 울릉군에 포함된 부모집단별 인구수 *Rel.Bias* 비교

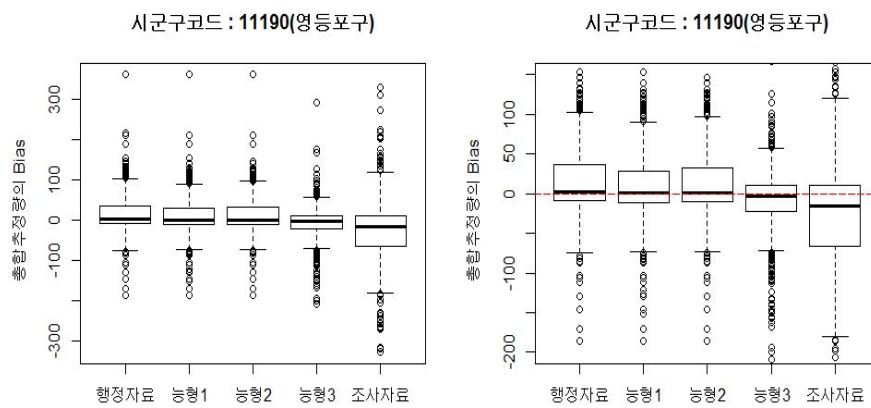


- *Rel.Bias*에서 고려한 다섯 가지 총합추정량의 *Bias* (*Rel.Bias*의 분자 부분)를 관심 부모집단별로 계산하였다. 단, 센서스, 행정자료 그리고 조사자료 인구수가 모두 '0'인 303개의 부모집단은 비교에서 제외하였다. 다음의 <그림 3-37>부터 <그림 3-48>은 각 시군구 안에 있는 관심 부모집단의 총합추정량의 *Bias*의 분포를 상자그림으로 표현한 것으로 *Rel.Bias*의 결과와 유사하게 나타나고 있다.

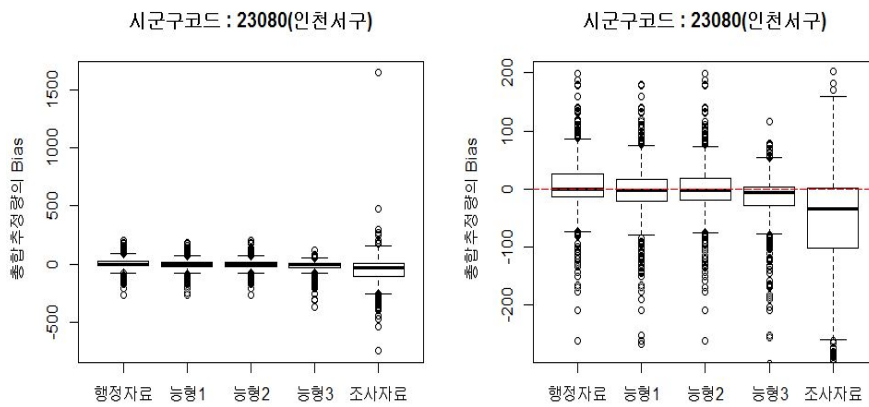
<그림 3-37> 금천구에 포함된 부모집단의 인구추정량의 Bias 상자그림



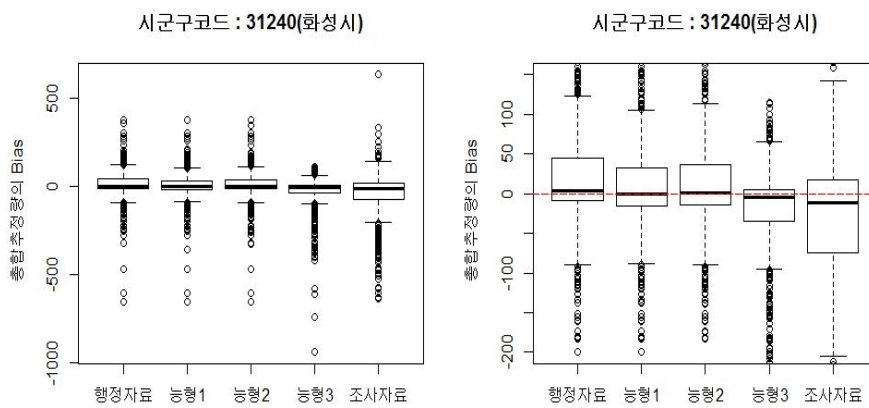
<그림 3-38> 영등포구에 포함된 부모집단의 인구추정량의 Bias 상자그림



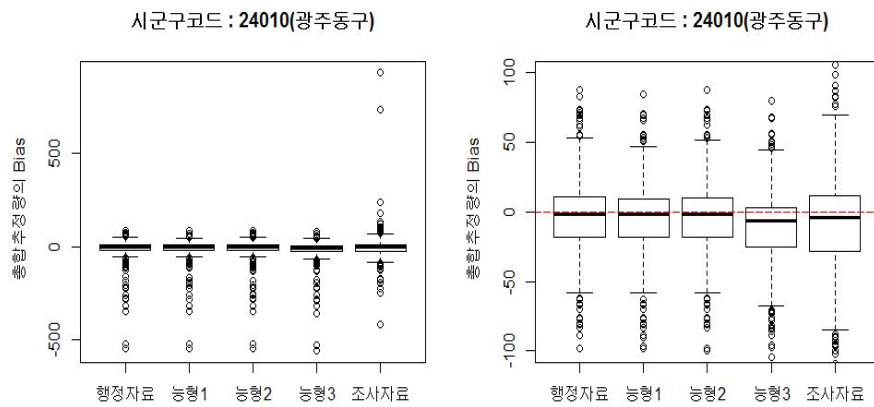
<그림 3-39> 인천서구에 포함된 부모집단의 인구추정량의 Bias 상자그림



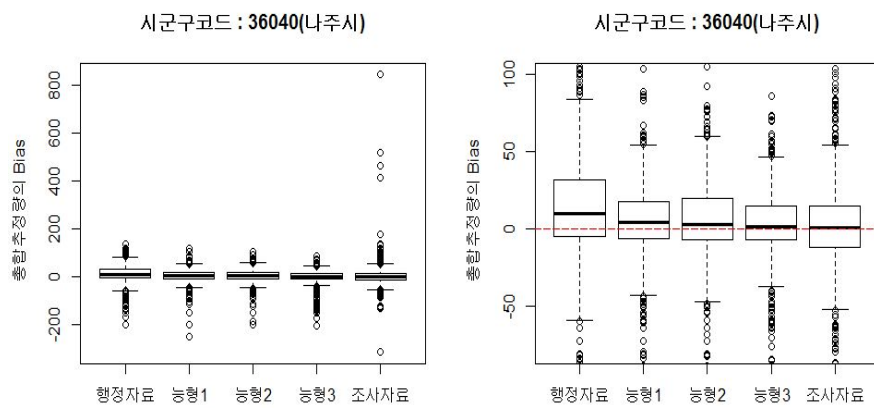
<그림 3-40> 화성시에 포함된 부모집단의 인구추정량의 Bias 상자그림



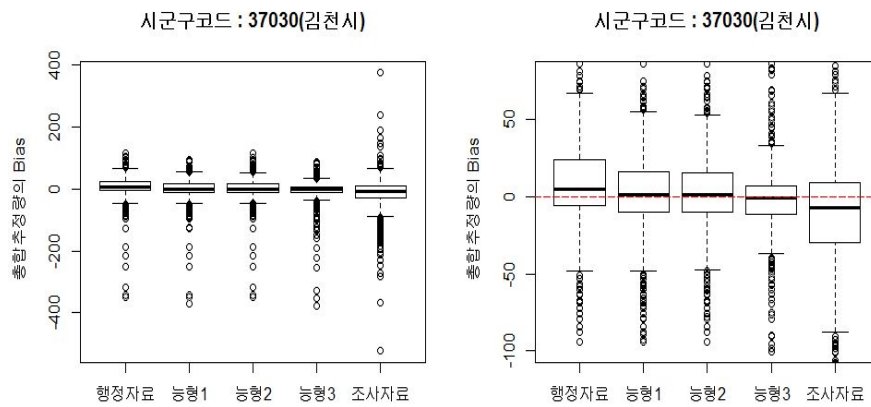
<그림 3-41> 광주동구에 포함된 부모집단의 인구추정량의 Bias 상자그림



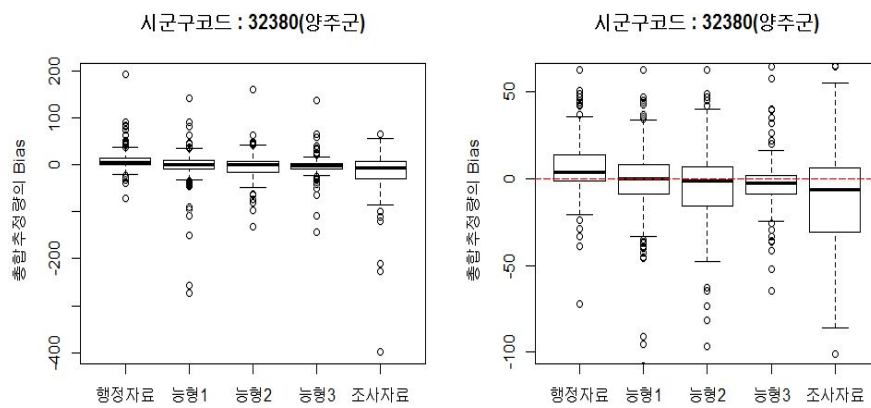
<그림 3-42> 나주시에 포함된 부모집단의 인구추정량의 Bias 상자그림



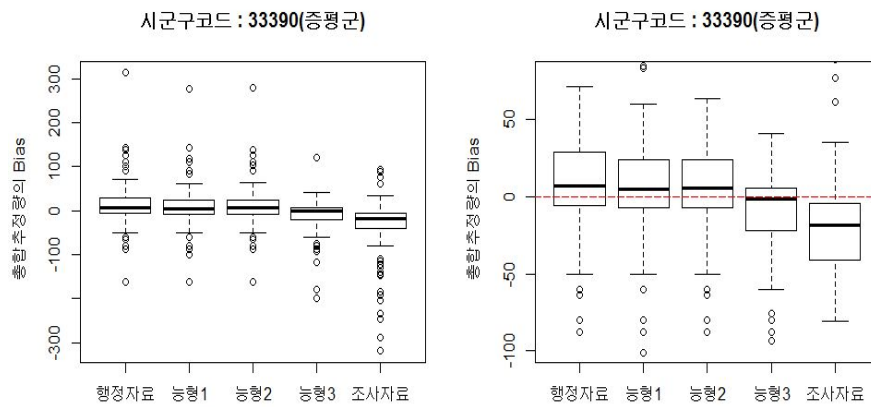
<그림 3-43> 김천시에 포함된 부모집단의 인구추정량의 Bias 상자그림



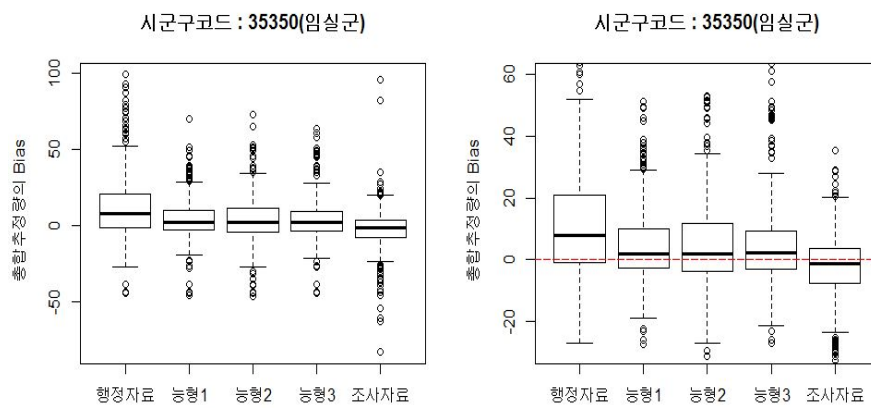
<그림 3-44> 양주군에 포함된 부모집단의 인구추정량의 Bias 상자그림



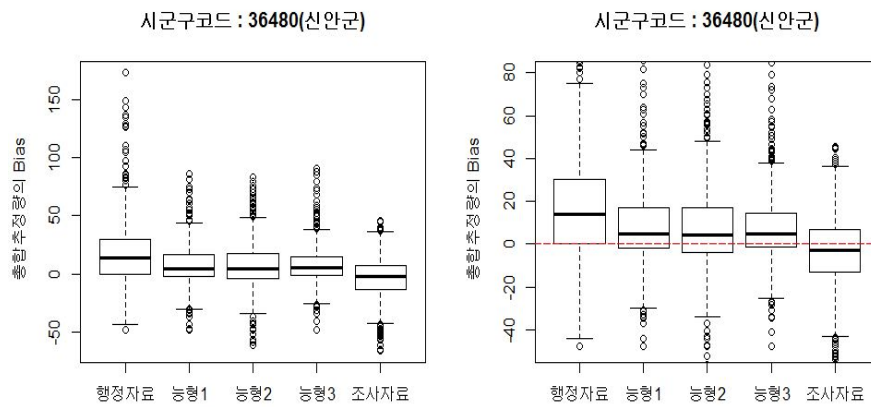
<그림 3-45> 증평군에 포함된 부모집단의 인구추정량의 Bias 상자그림



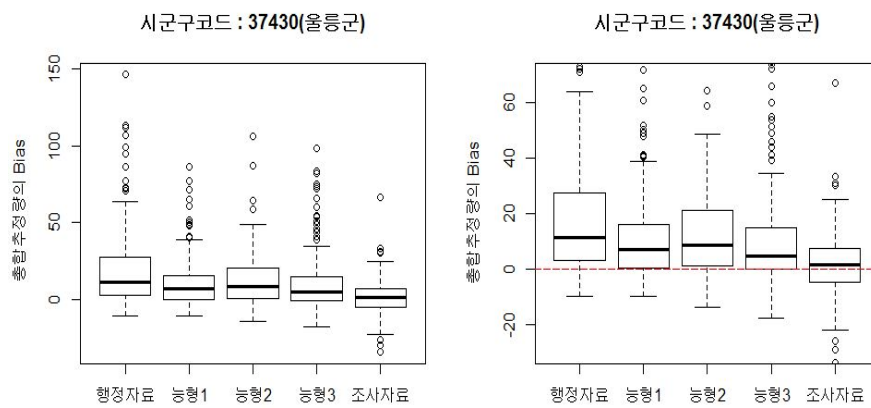
<그림 3-46> 임실군에 포함된 부모집단의 인구추정량의 Bias 상자그림



<그림 3-47> 신안군에 포함된 부모집단의 인구추정량의 Bias 상자그림

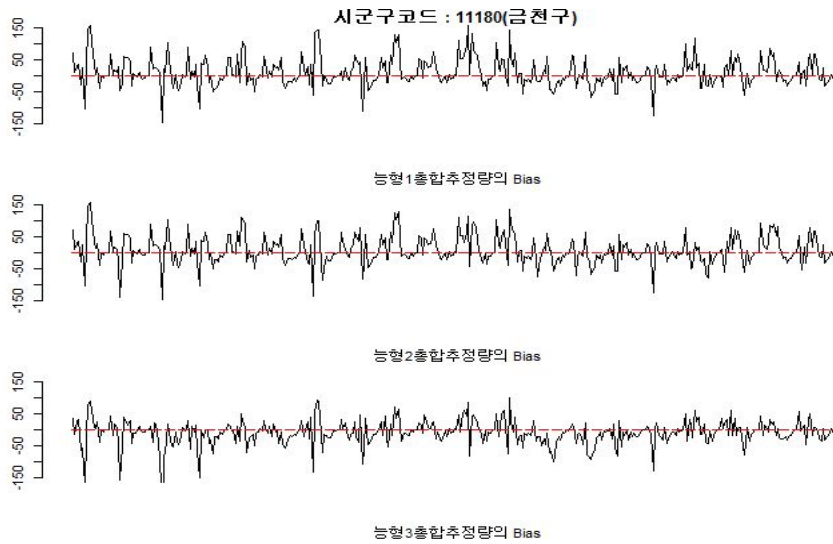


<그림 3-48> 울릉군에 포함된 부모집단의 인구추정량의 Bias 상자그림

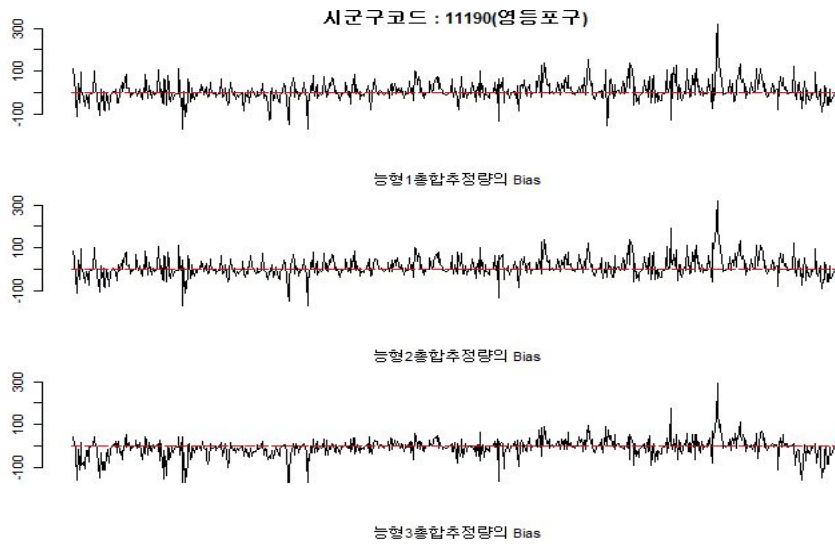


- 능형회귀추정 방안 3가지를 중심으로 $Bias$ 를 각 시군구 안에 있는 부모집단별로 표현한 결과는 다음의 <그림 3-49>부터 <그림 3-60>과 같다. 여기서 x 축은 각 시군구 안에 있는 관심 부모집단을 의미한다. $Bias$ 도 $Rel.Bias$ 와 유사하게 나타나고 있음을 확인할 수 있다.

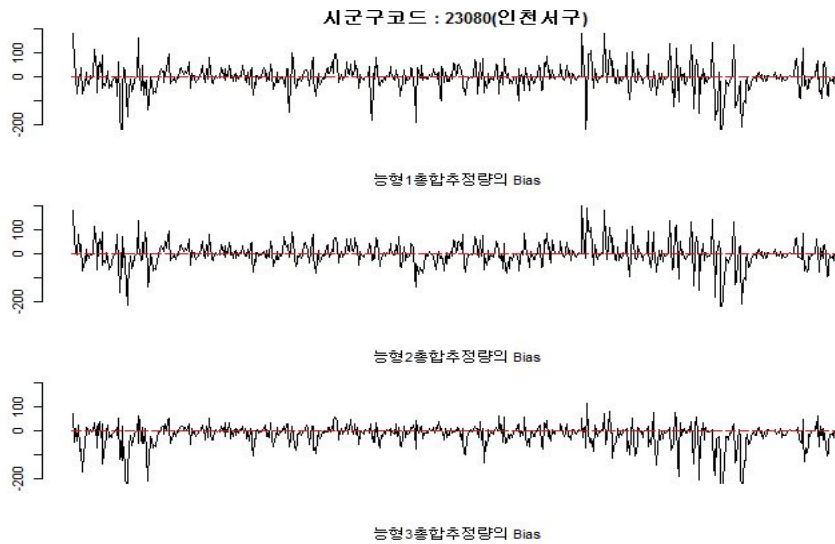
<그림 3-49> 금천구에 포함된 부모집단의 인구수 $Bias$ 비교



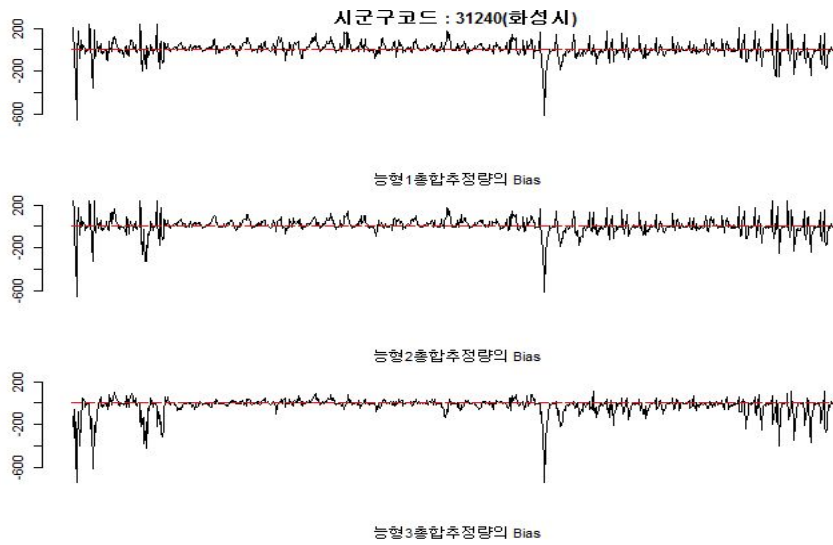
<그림 3-50> 영등포구에 포함된 부모집단의 인구수 *Bias* 비교



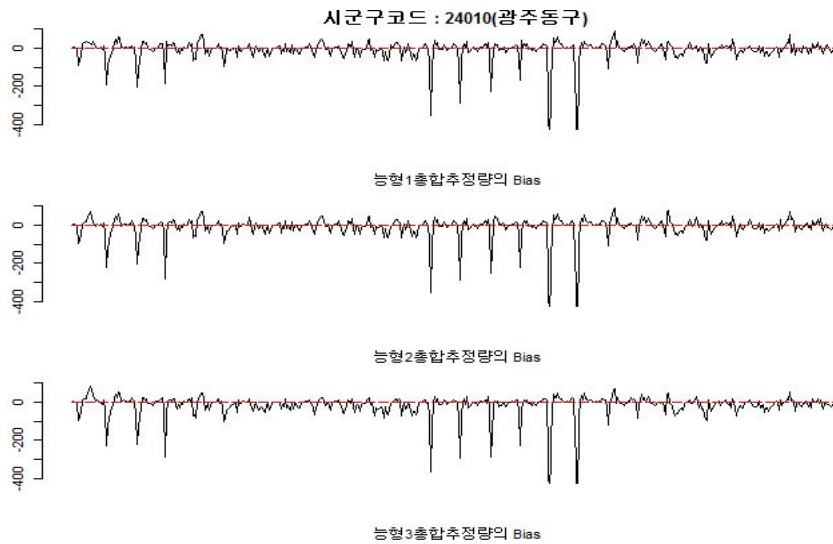
<그림 3-51> 인천서구에 포함된 부모집단의 인구수 *Bias* 비교



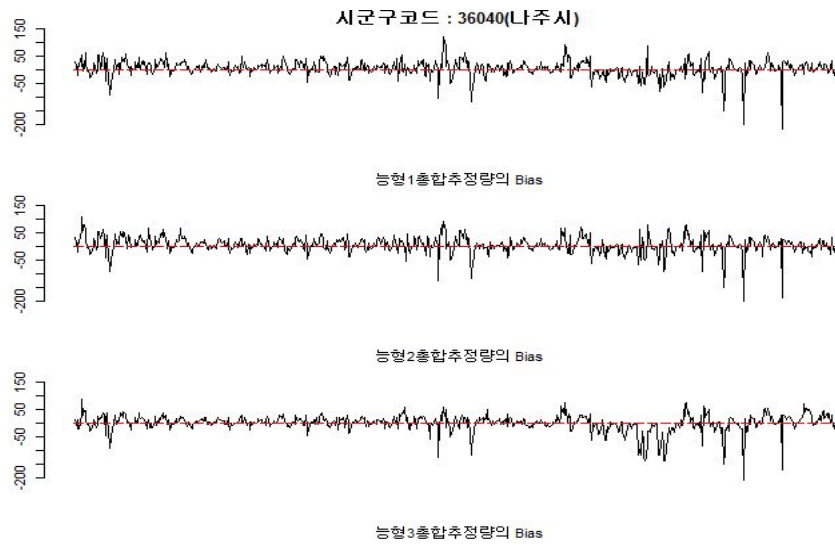
<그림 3-52> 화성시에 포함된 부모집단의 인구수 *Bias* 비교



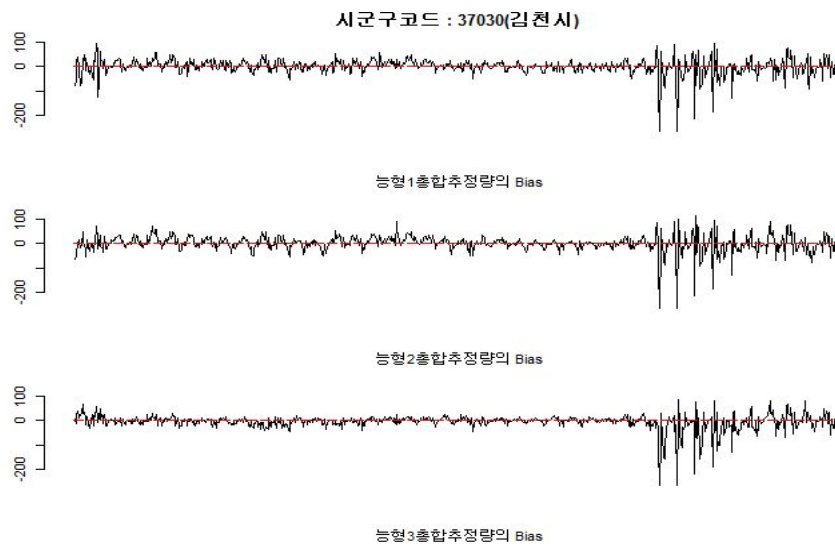
<그림 3-53> 광주동구에 포함된 부모집단의 인구수 *Bias* 비교



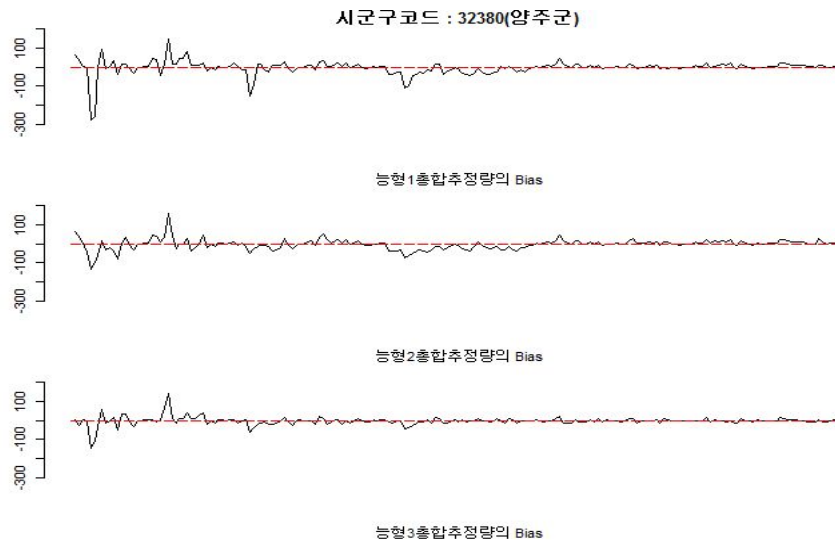
<그림 3-54> 나주시에 포함된 부모집단의 인구수 Bias 비교



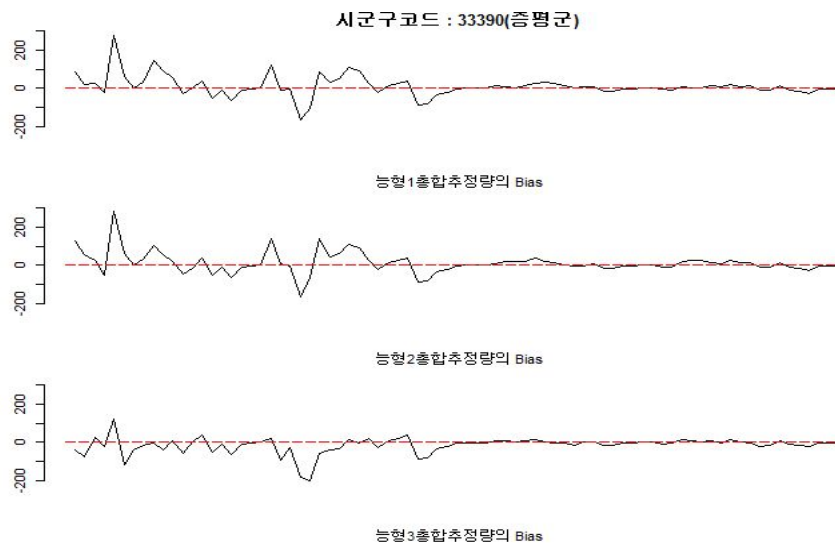
<그림 3-55> 김천시에 포함된 부모집단의 인구수 Bias 비교



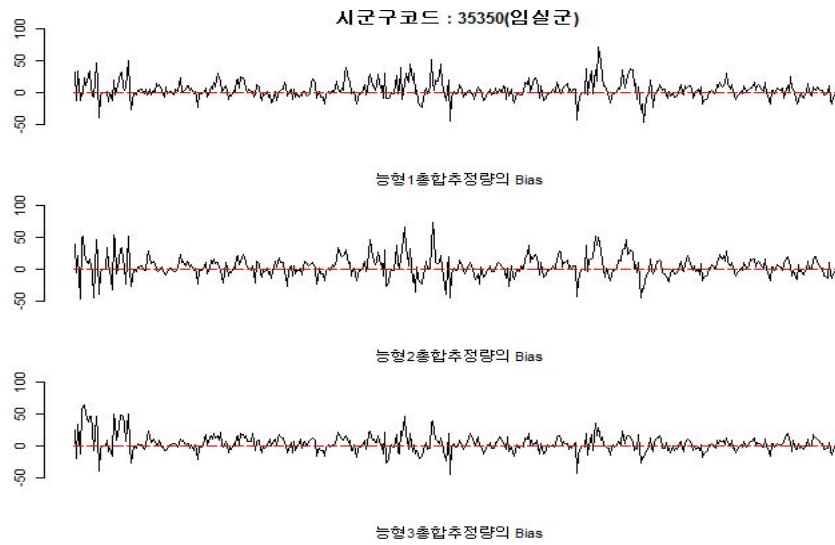
<그림 3-56> 양구군에 포함된 부모집단의 인구수 Bias 비교



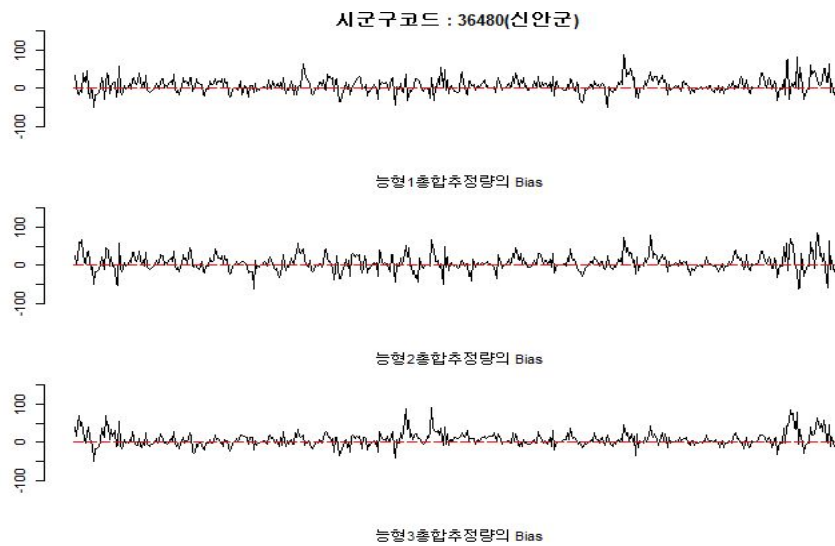
<그림 3-57> 증평군에 포함된 부모집단의 인구수 Bias 비교



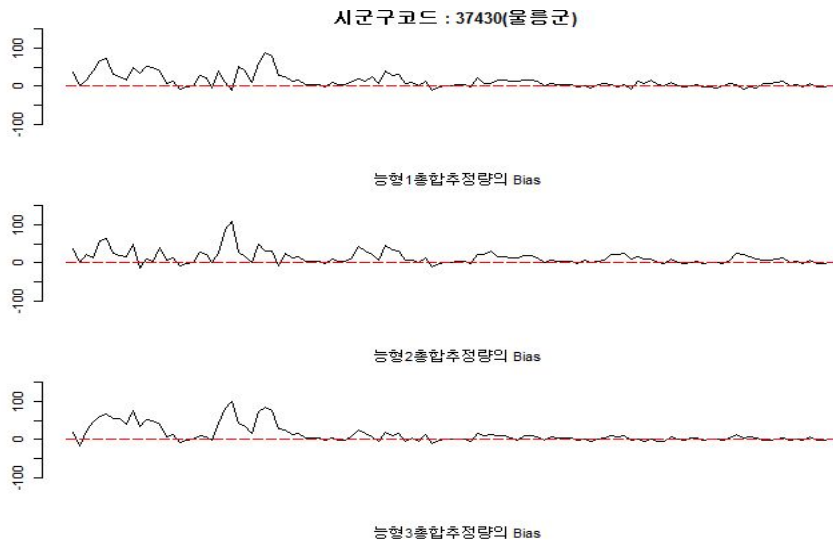
<그림 3-58> 임실군에 포함된 부모집단의 인구수 Bias 비교



<그림 3-59> 신안군에 포함된 부모집단의 인구수 Bias 비교



<그림 3-60> 울릉군에 포함된 부모집단의 인구수 Bias 비교



- 관심 부모집단에서 참 값(센서스총합)과 다섯 가지 총합추정량을 비교를 위하여 동질성 여부를 판단하는 χ^2 통계량을 적용하였다. 여기서의 χ^2 통계량은 단순히 참 값과의 이질성을 나타내는 척도로 사용하였으며 그 값이 클수록 참 값과의 차이가 크다고 볼 수 있다. 그 결과는 다음의 <표 3-3>과 같다. 능형 회귀추정 방안1, 2, 3의 χ^2 통계량 값 모두 행정자료와 표본자료의 값보다 작게 나타나고 있음을 확인할 수 있다.

<표 3-3> 부모집단의 인구추정량의 참 값과의 동질성 비교

시군구	행정자료	능형1	능형2	능형3	표본자료
χ^2 통계량 (df)	24,224 (6,332)	17,855 (6,332)	18,797 (6,332)	14,986 (6,332)	40,321 (6,265)

2. 가구

- 읍면동별*가구원수별로 정의된 부모집단에 대한 가구수 산출을 위한 능형회귀 추정방안을 위한 조정 셀은 통계청과의 논의 후 부모집단 읍면동별(158)*가구원수별(5)과 동일한 790개의 셀로 정의하여 12개 시범지역에서 대하여 비교하였다.
- 능형회귀추정 방안을 통하여 각 조정 셀에서 보정된 가구 가중치의 분포는 다음의 <표 3-4>와 같다. 능형회귀 가중치의 평균 역시 인구 부분과 마찬가지로 '1'보다 작은 값으로 행정자료가 평균적으로 과대 집계됨을 보여주고 있다. 능형회귀 가중치의 5%에서 95%까지의 분포는 0.74~1.03의 값으로 작은 변동을 보이는 반면 하위 5%와 상위 5% 부분에서 변동이 크게 나타나고 있다.

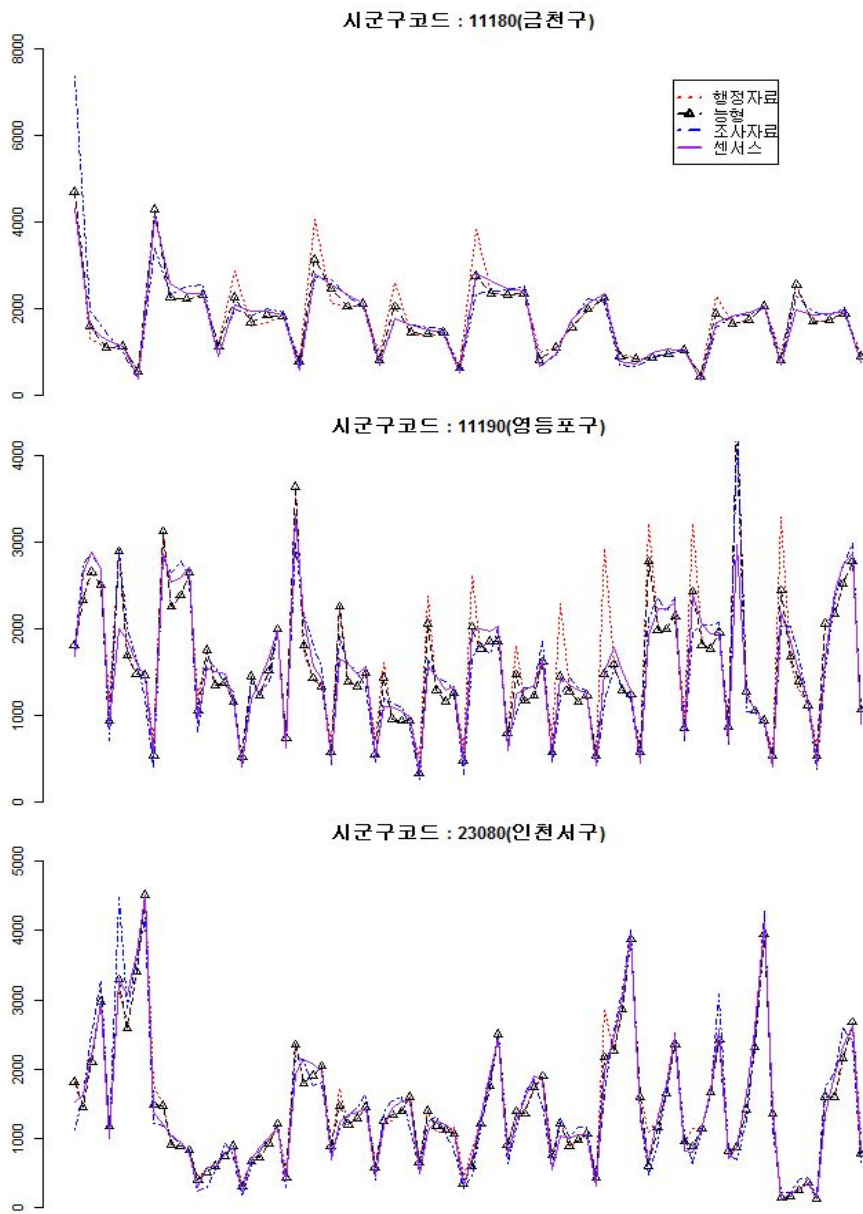
<표 3-4> 12개 시범지역의 가구 능형회귀 가중치의 분포

	mean	sd	min	5%	10%	Q1	median	Q3	95%	max
능형회귀가중치	0.966	0.1240	0.170	0.743	0.807	1.000	1.000	1.000	1.034	3.394

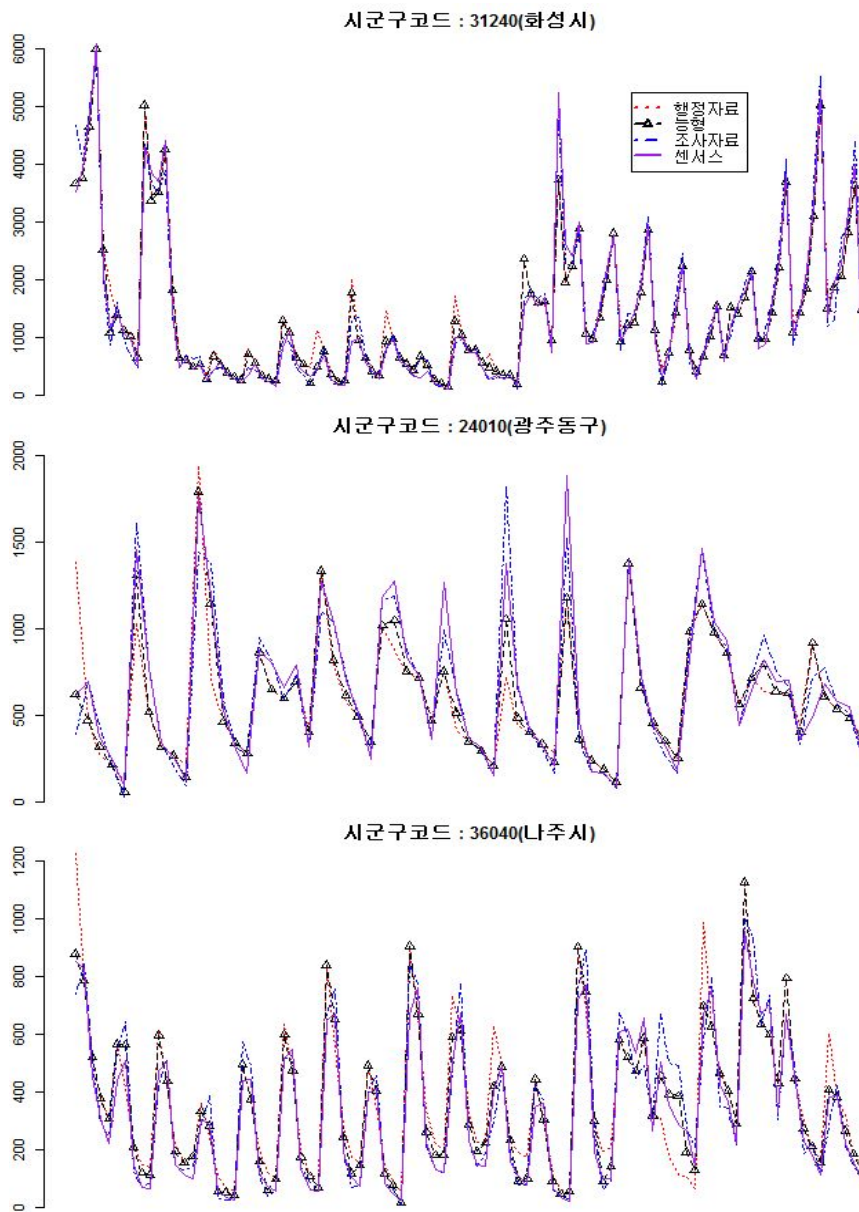
- 센서스와 행정자료의 총합 그리고 표본자료와 능형회귀추정 방안의 총합 추정량의 가구원수, 시군구, 가구원수*시군구 별 가구수 분포는 각 각 <부록 3-8>, <부록 3-9> 그리고 <부록 3-10>을 참고하면 된다.
- 읍면동별(158)*가구원수별(5)의 총 790개의 부모집단별 센서스와 행정자료의 가구수 그리고 조사자료와 능형회귀추정 방안의 추정 가구수는 다음의 <그림 3-61>부터 <그림 3-64>와 같다. 여기서, x 축은 각 시군구에 속하는 부모집단을 나타낸다. 인구부분과 마찬가지로 특정 부모집단에서 행정자료 가구수와 센서스 가구수의 차이가 크게 나는 것을 확인할 수 있으며, 또한 조사자료 추정 가구수 역시 특정 부모집단에서 센서스 가구수와 차이가 나타남을 확인할 수 있다. 따라서 가구부분에서도 행정자료 가구수와 조사자료 가구수의 차이가 큰 곳에서 능형회귀추정 방안을 통한 보정은 적절해 보이며, 평균적으로 능형회귀추정 방안을 통한 가구수 추정 결과가 센서스 결과와 차이가 작음을

확인할 수 있다. 그리고 능형회귀추정량을 정의하기 위한 조정 셀과 관심 부
모집단이 같게 나타남에 따라 능형회귀추정 방안의 결과는 항상 행정자료 가
구수와 조사자료 추정 가구수 사이에 존재함을 확인 할 수 있다.

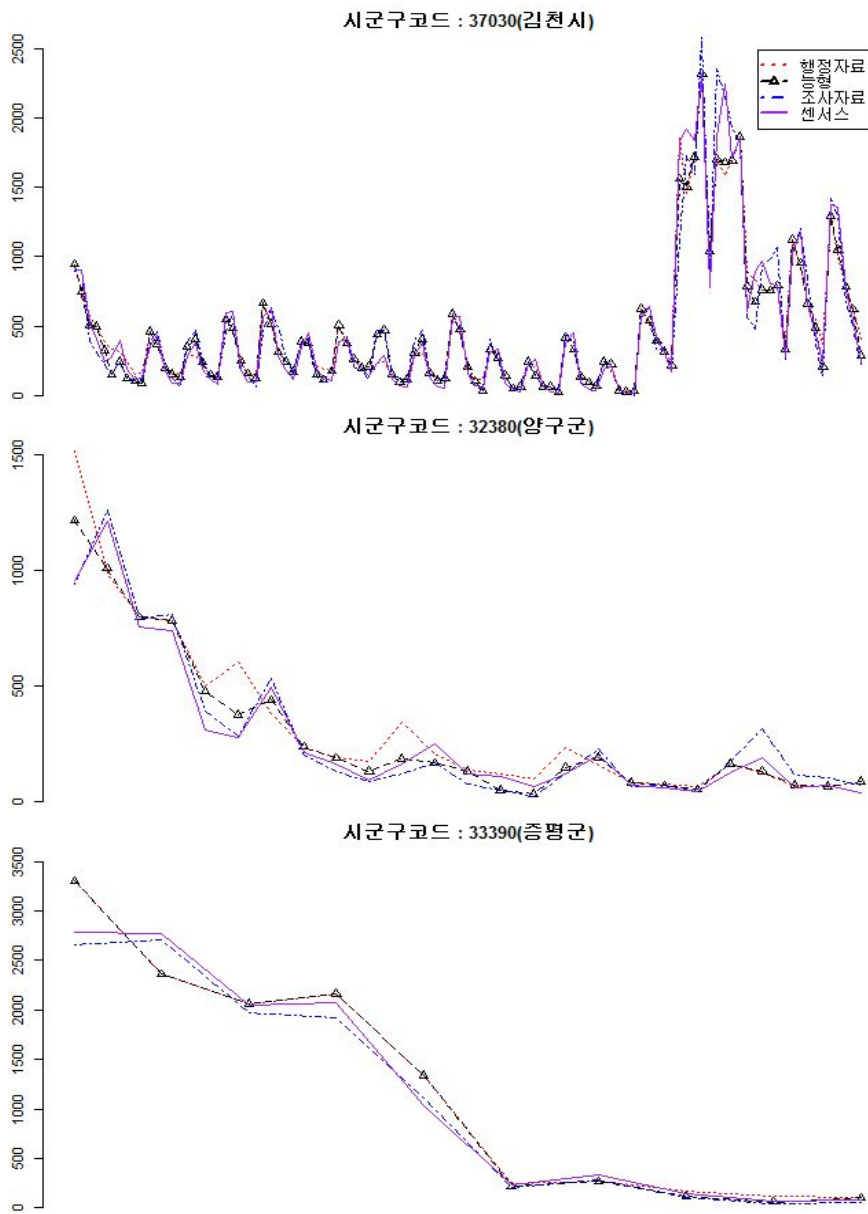
<그림 3-61> 금천구, 영등포구, 인천서구에 포함된 부모집단별 가구수 비교



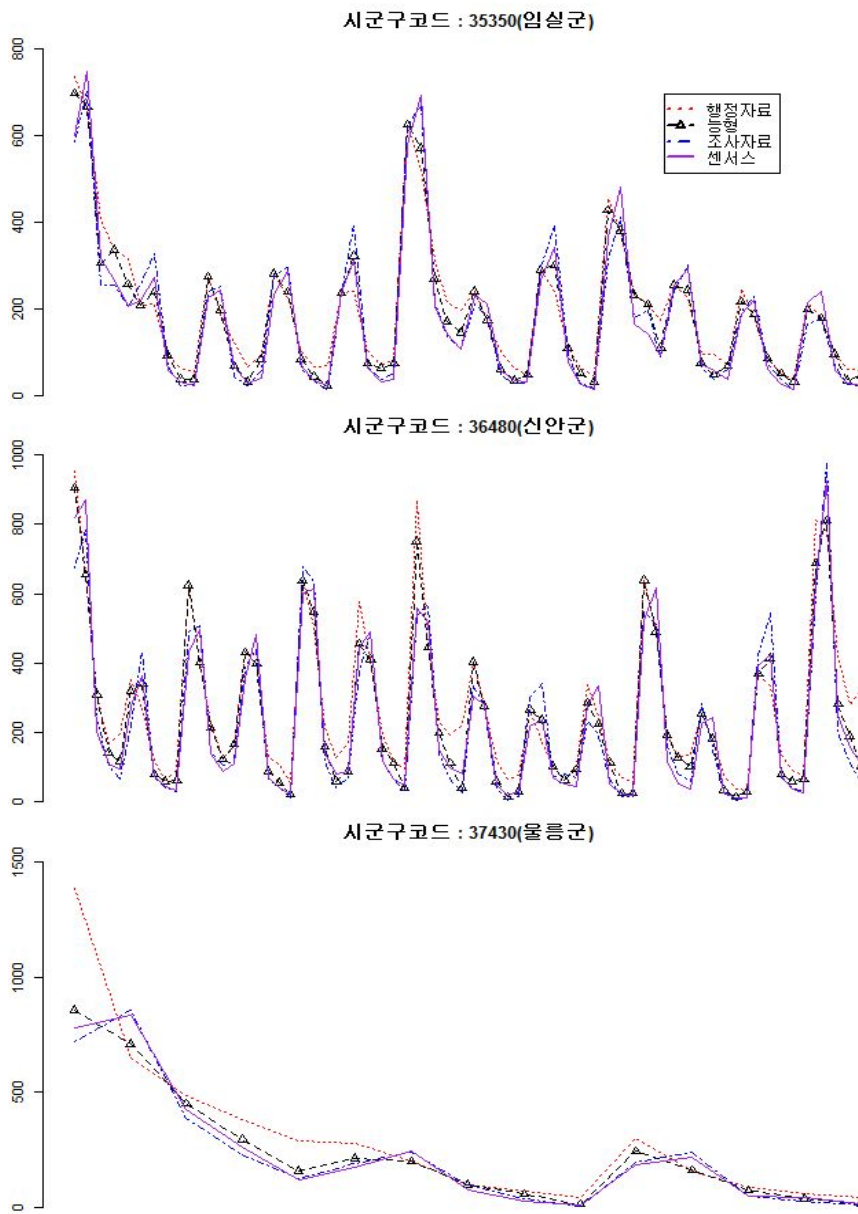
<그림 3-62> 화성시, 광주동구, 나주시에 포함된 부모집단별 가구수 비교



<그림 3-63> 김천시, 양구군, 증평군에 포함된 부모집단별 가구수 비교

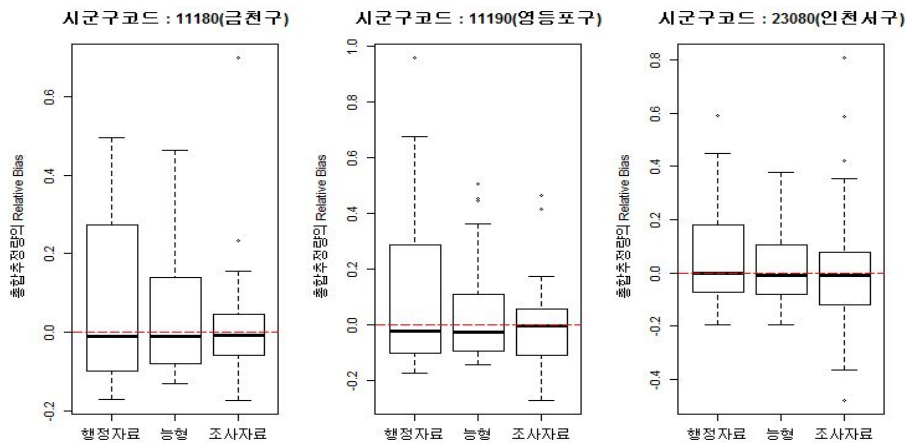


<그림 3-64> 임실군, 신안군, 울릉군에 포함된 부모집단별 가구수 비교

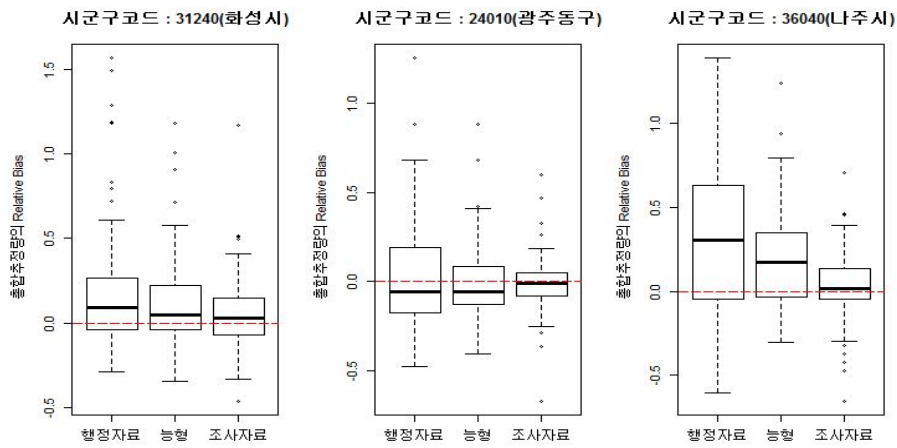


- 능형회귀추정 방안을 사용한 각 부모집단별 가구수 추정 결과를 행정자료와 조사자료 결과와 비교하기 인구부분에서 사용한 식 (3.14)의 *Rel.Bias*를 사용하였다. 다섯 가지 총합추정량의 *Rel.Bias*는 총 790개의 관심 부모집단에 대하여 계산되었다. 다음의 <그림 3-65>부터 <그림 3-68>은 각 시군구 안에 있는 관심 부모집단의 총합추정량의 *Rel.Bias*의 분포를 상자그림으로 표현한 것이다. 인구부분과 마찬가지로 특정 시군구에서 행정자료나 조사자료의 *Rel.Bias*는 평균적으로 '0'에 가깝게 나타나거나 '0'과 큰 차이를 보이는 반면, 능형회귀추정 방안은 전체 시군구에 걸쳐 행정자료와 조사자료의 *Rel.Bias*보다 평균적으로 '0'으로 부터의 차이가 크게 나타나고 있지 않다.

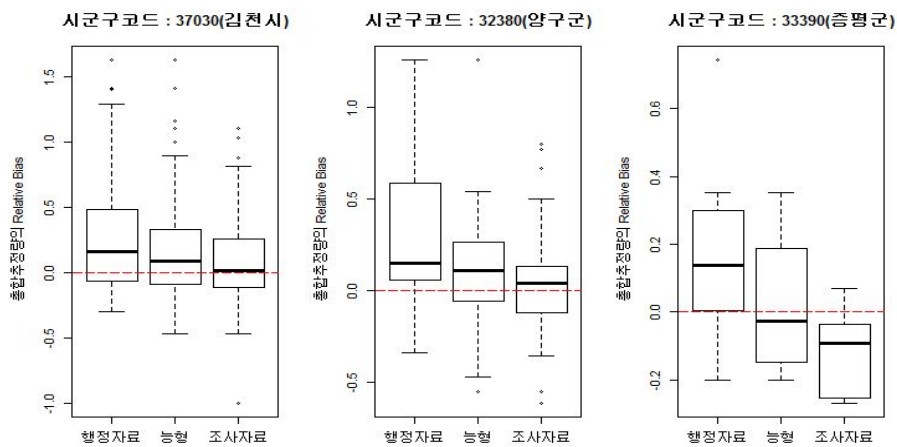
<그림 3-65> 금천구, 영등포구, 인천서구에 포함된 부모집단의 가구추정량의 *Rel.Bias* 상자그림



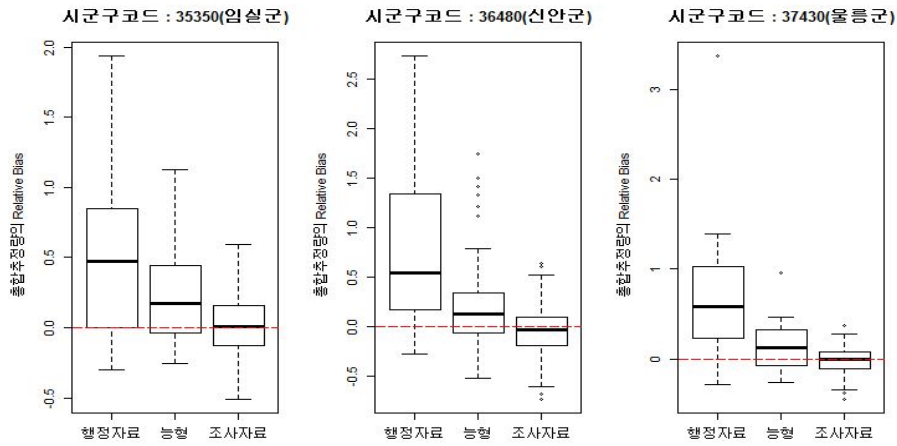
<그림 3-66> 화성시, 광주동구, 나주시에 포함된 부모집단의 가구추정량의 *Rel.Bias* 상자그림



<그림 3-67> 김천시, 양구군, 증평군에 포함된 부모집단의 가구추정량의 *Rel.Bias* 상자그림

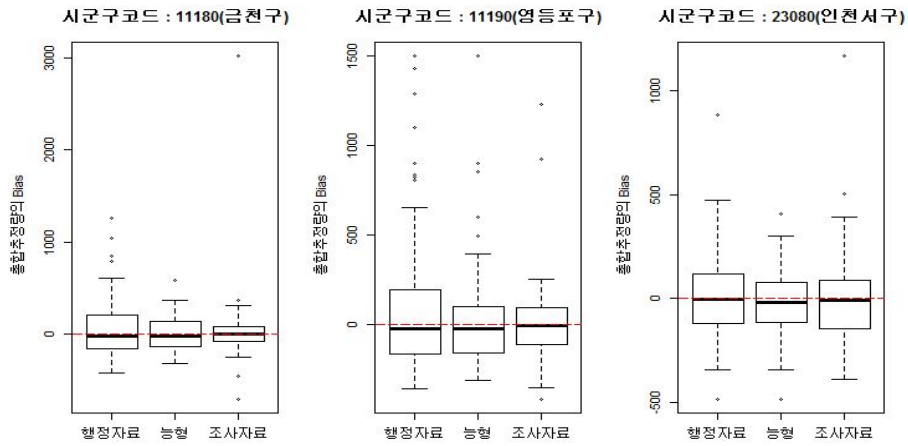


<그림 3-68> 임실군, 신안군, 울릉군에 포함된 부모집단의 가구추정량의 *Rel.Bias* 상자그림

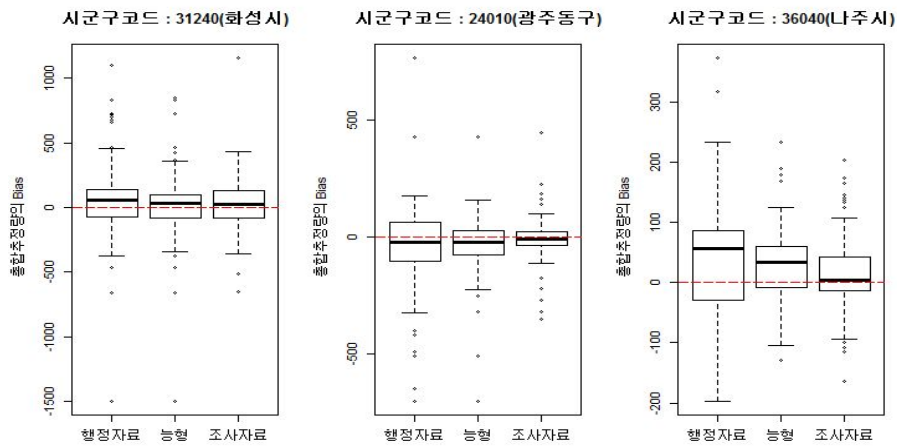


- *Rel.Bias*에서 고려한 다섯 가지 총합추정량의 *Bias*를 관심 부모집단별로 계산한 결과의 각 시군구별 분포는 다음의 <그림 3-69>부터 <그림 3-72>와 같다. 그 결과는 *Rel.Bias*와 유사하게 나타나고 있다.

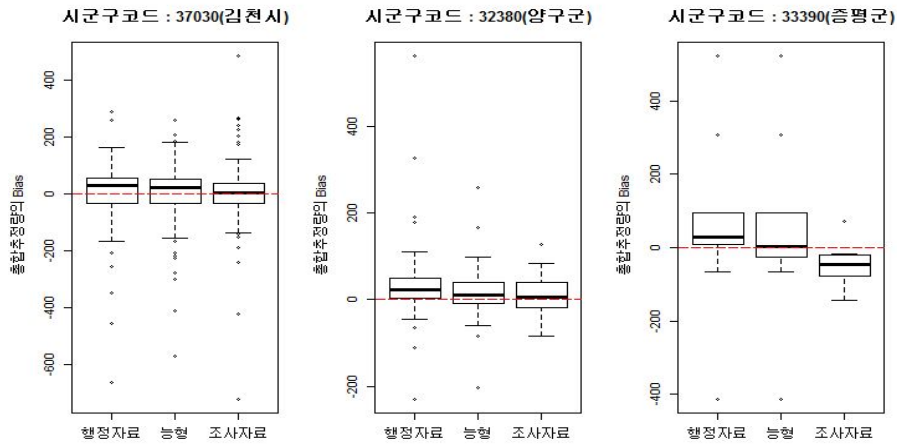
<그림 3-69> 금천구, 영등포구, 인천서구에 포함된 부모집단의 가구추정량의 Bias 상자그림



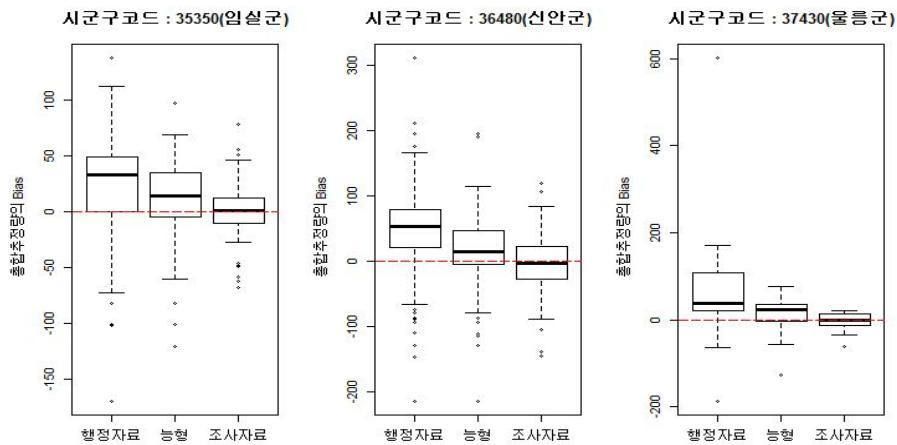
<그림 3-70> 화성시, 광주동구, 나주시에 포함된 부모집단의 가구추정량의 Bias 상자그림



<그림 3-71> 김천시, 양구군, 증평군에 포함된 부모집단의 가구추정량의
Bias 상자그림



<그림 3-72> 임실군, 신안군, 울릉군에 포함된 부모집단의 가구추정량의
Bias 상자그림



3. 주택

- 읍면동별*주택유형별로 정의된 부모집단에 대한 주택수 산출을 위한 능형회귀 추정방안을 위한 조정 셀은 통계청과의 논의 후 부모집단 읍면동별(158)*주택 유형별(3)과 동일한 474개의 셀로 정의하여 12개 시범지역에서 대하여 비교하였다.
- 능형회귀추정 방안을 통하여 각 조정 셀에서 보정된 주택 가중치의 분포는 다음의 <표 3-5>와 같다. 능형회귀 가중치의 평균 역시 인구, 가구 부분과 마찬가지로 ‘1’보다 작은 값으로 행정자료가 평균적으로 과대 집계됨을 보여주고 있다. 능형회귀 가중치의 5%에서 95%까지의 분포는 0.99~1.00의 값으로 작은 변동을 보이는 반면 하위 5%와 상위 5% 부분에서 변동이 크게 나타나고 있다.

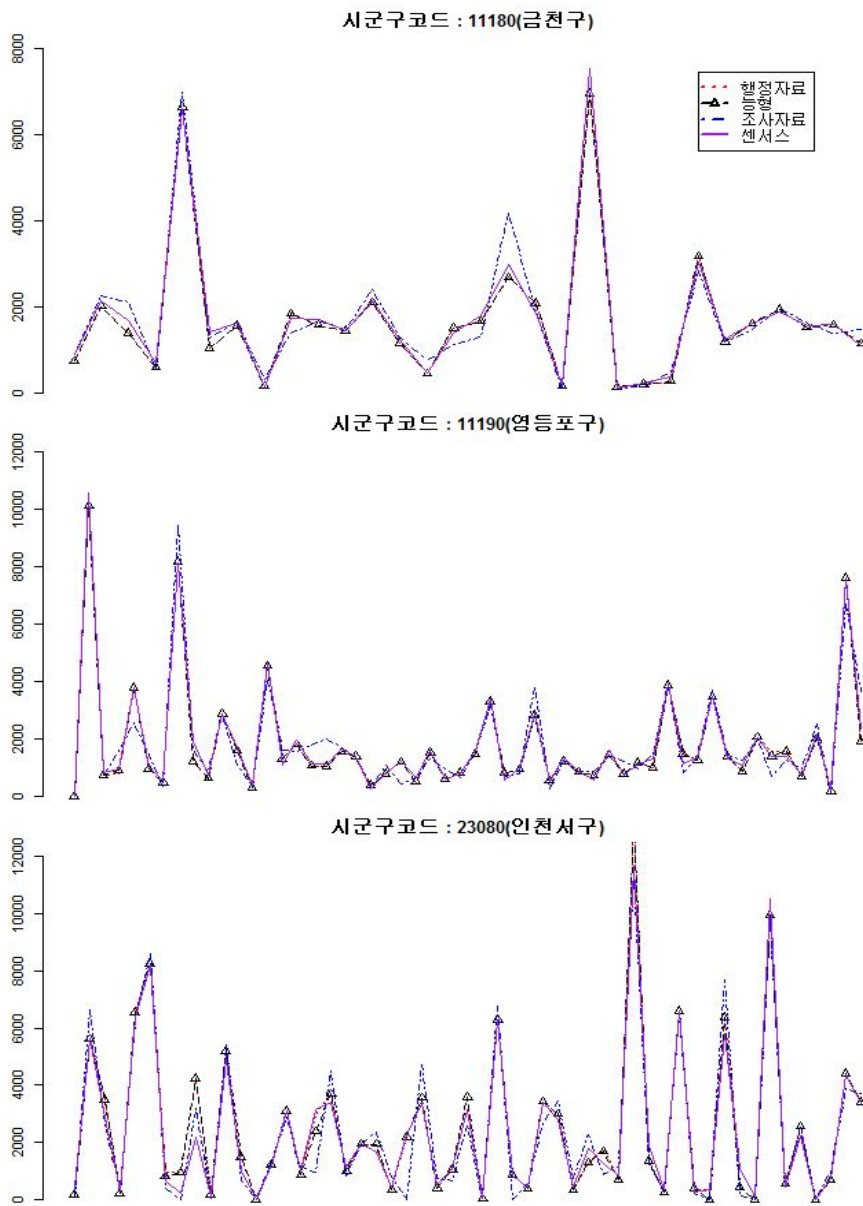
<표 3-5> 12개 시범지역의 개인 능형회귀가중치의 분포

	mean	sd	min	5%	10%	Q1	median	Q3	95%	max
능형회귀가중치	0.993	0.0875	0.026	0.986	1.000	1.000	1.000	1.000	1.000	1.939

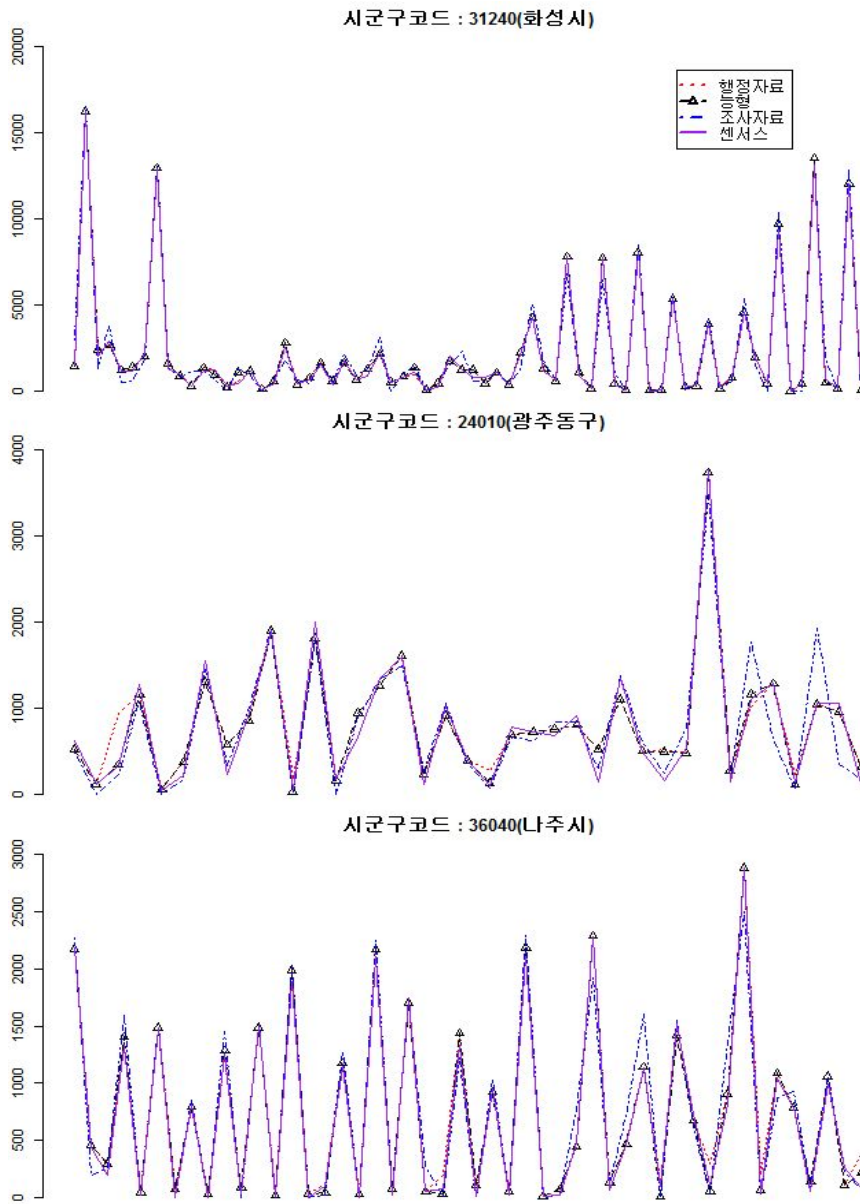
- 센서스와 행정자료의 총합 그리고 표본자료와 능형회귀추정 방안의 총합 추정량의 주택유형, 시군구, 주택유형*시군구 별 가구수 분포는 각 각 <부록 3-11>, <부록 3-12> 그리고 <부록 3-13>을 참고하면 된다.
- 읍면동별(158)*주택유형별(3)의 총 474개의 부모집단별 센서스와 행정자료의 주택수 그리고 조사자료와 능형회귀추정 방안의 추정 주택수는 다음의 <그림 3-73>부터 <그림 3-76>과 같다. 여기서, x 축은 각 시군구에 속하는 부모 집단을 나타낸다. 인구, 가구부분과 마찬가지로 행정자료 주택수와 조사자료 주택수의 차이가 큰 곳에서 능형회귀추정 방안을 통한 보정은 적절해 보이며, 평균적으로 능형회귀추정 방안을 통한 주택수 추정 결과가 센서스 결과와 차이가 작음을 확인할 수 있다. 그리고 능형회귀추정량을 정의하기 위한 조정 셀과 관심 부모집단이 같게 나타남에 따라 가구부분과 마찬가지로 능형회귀추

정 방안의 결과는 항상 행정자료 주택수와 조사자료 추정 주택수 사이에 존재
함을 확인 할 수 있다.

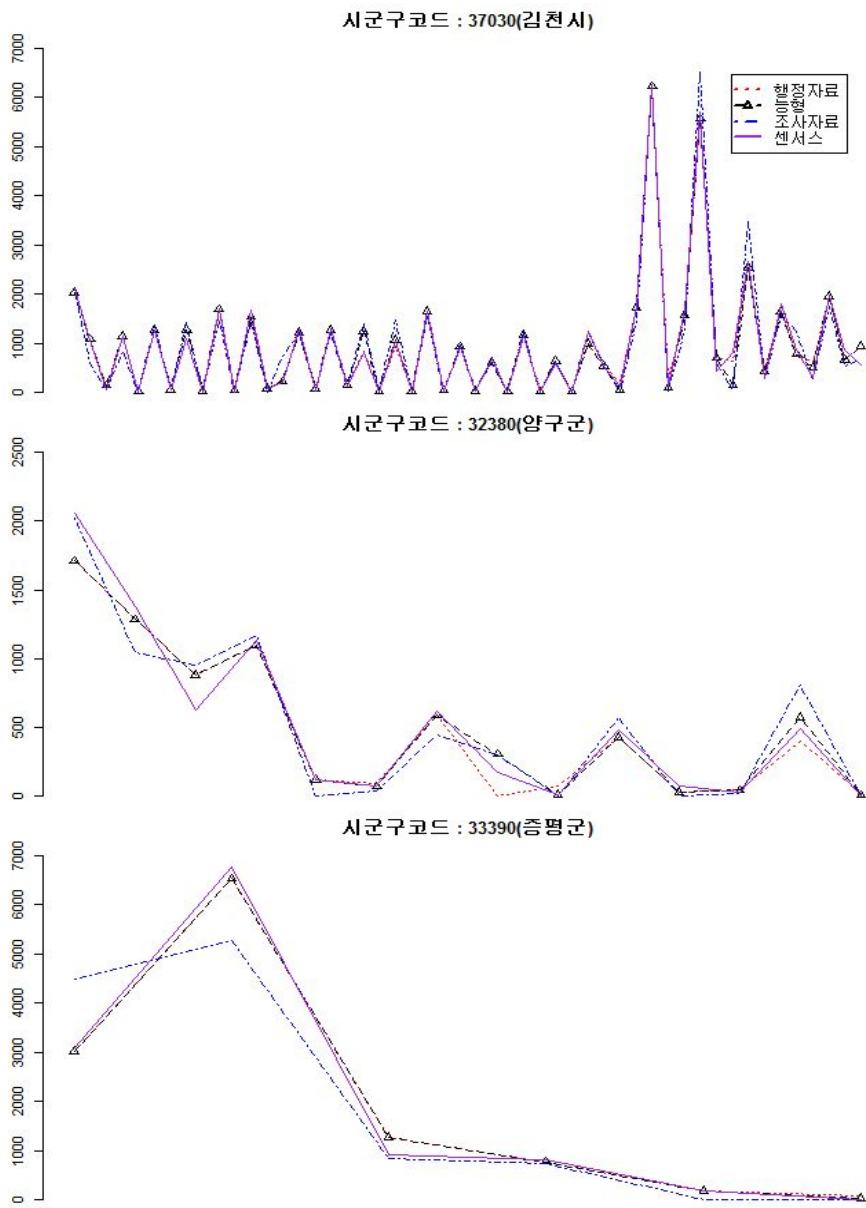
<그림 3-73> 금천구, 영등포구, 인천서구에 포함된 부모집단의 주택수 비교



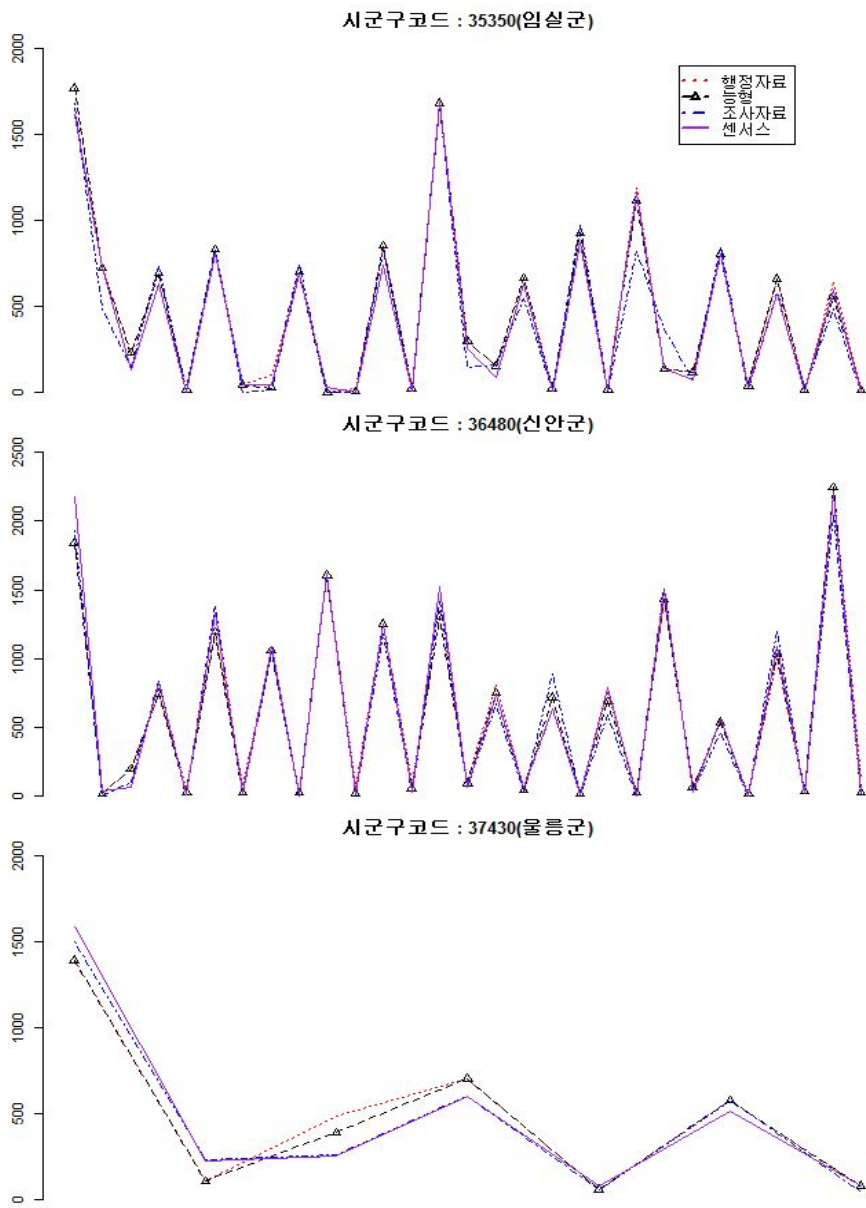
<그림 3-74> 화성시, 광주동구, 나주시에 포함된 부모집단의 주택수 비교



<그림 3-75> 김천시, 양구군, 증평군에 포함된 부모집단의 주택수 비교

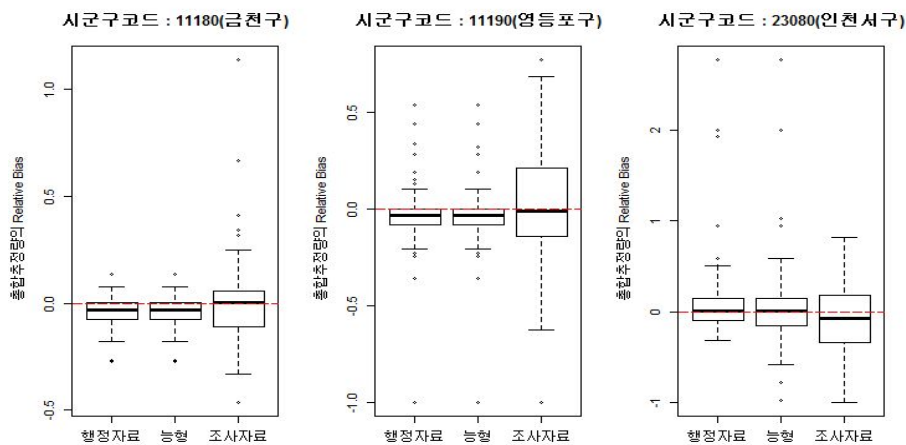


<그림 3-76> 임실군, 신안군, 울릉군에 포함된 부모집단의 주택수 비교

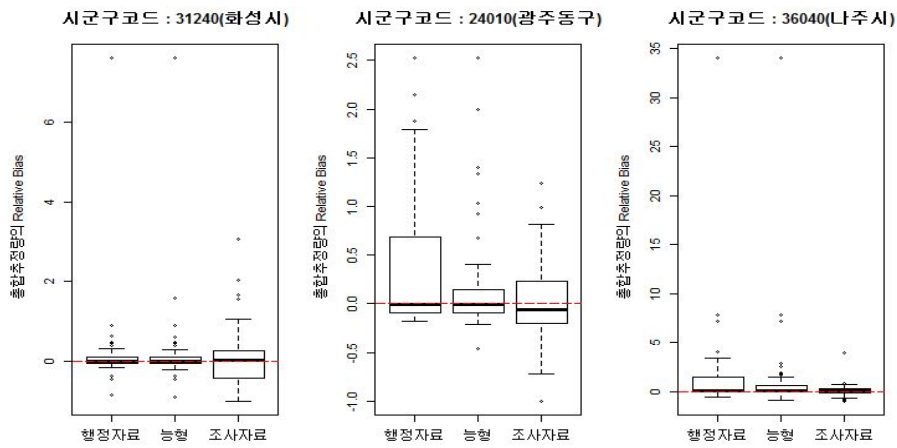


- 능형회귀추정 방안을 사용한 각 부모집단별 주택수 추정 결과를 행정자료와 조사자료 결과와 비교하기 인구, 가구부분에서 사용한 식 (3.14)의 $Rel.Bias$ 를 사용하였다. 다섯 가지 총합추정량의 $Rel.Bias$ 는 센서스총합이 '0'인 50개의 부모집단을 제외한 424개의 부모집단에 관심 부모집단에서 계산되었다. 다음의 <그림 3-77>부터 <그림 3-80>은 각 시군구 안에 있는 관심 부모집단의 총합추정량의 $Rel.Bias$ 의 분포를 상자그림으로 표현한 것이다. 인구, 가구 부분과 마찬가지로 특정 시군구에서 행정자료나 조사자료의 $Rel.Bias$ 는 평균적으로 '0'에 가깝게 나타나거나 '0'과 큰 차이를 보이는 반면, 능형회귀추정 방안은 전체 시군구에 걸쳐 행정자료와 조사자료의 $Rel.Bias$ 보다 평균적으로 '0'으로 부터의 차이가 크게 나타나고 있지 않다.

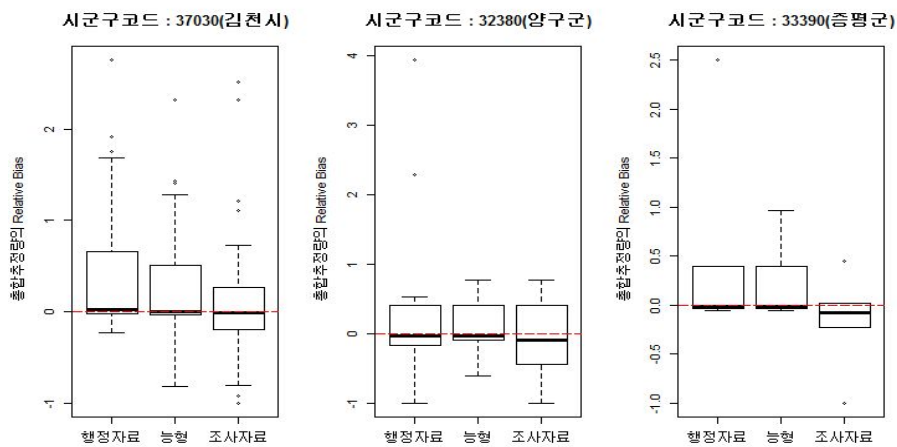
<그림 3-77> 금천구, 영등포구, 인천서구에 포함된 부모집단의 주택추정량의 $Rel.Bias$ 상자그림



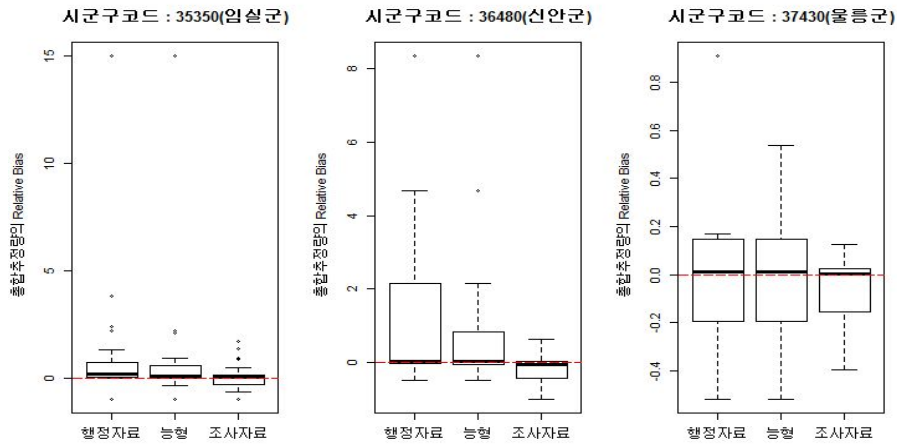
<그림 3-78> 화성시, 광주동구, 나주시에 포함된 부모집단의 주택추정량의 *Rel.Bias* 상자그림



<그림 3-79> 김천시, 양구군, 증평군에 포함된 부모집단의 주택추정량의 *Rel.Bias* 상자그림

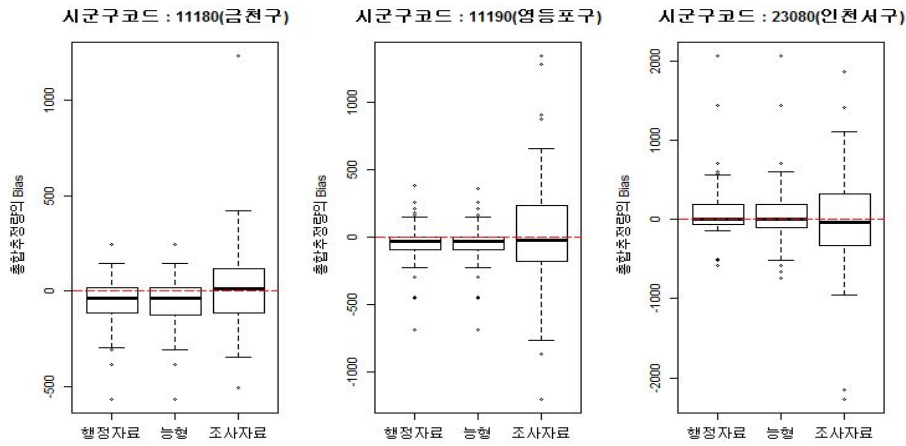


<그림 3-80> 임실군, 신안군, 울릉군에 포함된 부모집단의 주택추정량의 *Rel.Bias* 상자그림

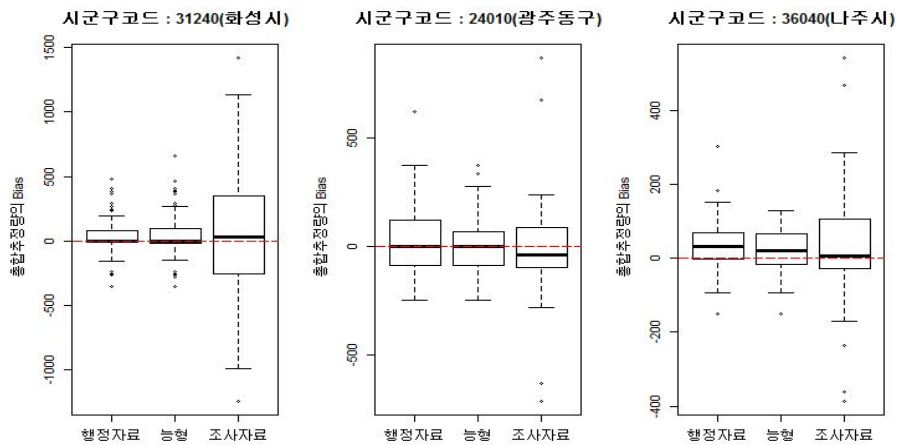


- *Rel.Bias*에서 고려한 다섯 가지 총합추정량의 *Bias*를 관심 부모집단별로 계산한 결과의 각 시군구별 분포는 다음의 <그림 3-81>부터 <그림 3-84>와 같다. 관심 부모집단 중 49개의 부모집단은 센서스총합, 행정자료총합, 능형회귀추정량, 조사자료의 총합추정량의 값이 모두 '0'으로 비교에서 제외되었다.

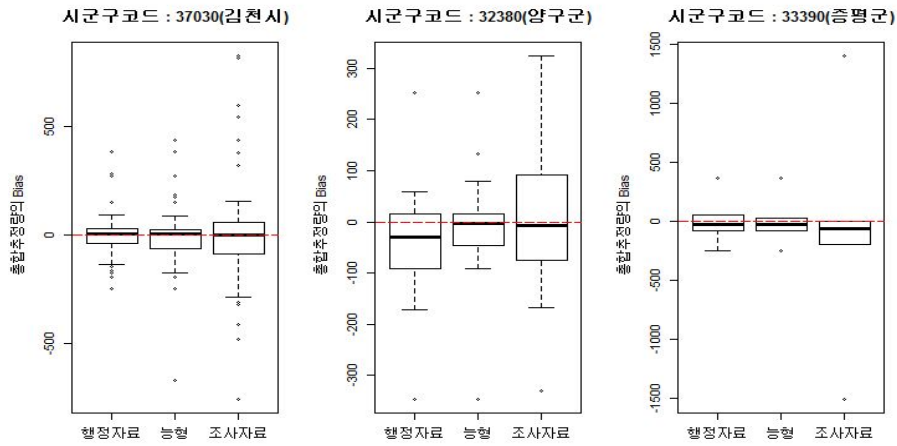
<그림 3-81> 금천구, 영등포구, 인천서구에 포함된 부모집단의 주택추정량의 Bias 상자그림



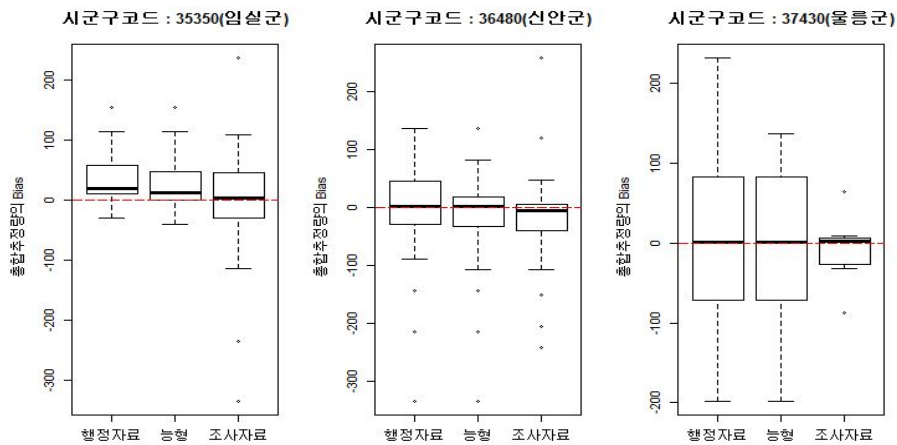
<그림 3-82> 화성시, 광주동구, 나주시에 포함된 부모집단의 주택추정량의 Bias 상자그림



<그림 3-83> 김천시, 양구군, 증평군에 포함된 부모집단의 주택추정량의
Bias 상자그림



<그림 3-84> 임실군, 신안군, 울릉군에 포함된 부모집단의 주택추정량의
Bias 상자그림



제4절 최종데이터를 통한 인구, 가구 그리고 주택자료 분석 결과

- 제3절에서 다룬 12개 시범지역에 대한 인구, 가구 그리고 주택에 대한 분석에서는 최종데이터의 이전의 중간데이터를 사용하여 인구, 가구 그리고 주택자료의 분석을 실시하였다. 본 절에서는 중간데이터 이후 작성된 최종데이터를 통한 인구, 가구 그리고 주택자료의 12개 시범지역에 대한 분석을 실시하였으며 이를 바탕으로 행정자료, 조사자료 그리고 능형회귀방안을 비교하였다.
- 본 절에서 인구, 가구 그리고 주택자료의 통계 산출을 위하여 고려된 행정자료, 조사자료 그리고 능형회귀추정 방안의 비교를 위하여 사용한 척도는 제3절에서 정의한 식 (3.14)의 $Rel.Bias$ 이다. 또한, 비교를 위하여 정의되는 관심 부모집단은 제3절에서 정의한 것과 동일하며, 인구수 추정을 위하여 고려한 능형회귀추정 방안 1, 2, 3에 방안 4를 추가하여 비교하였다.

1. 인구

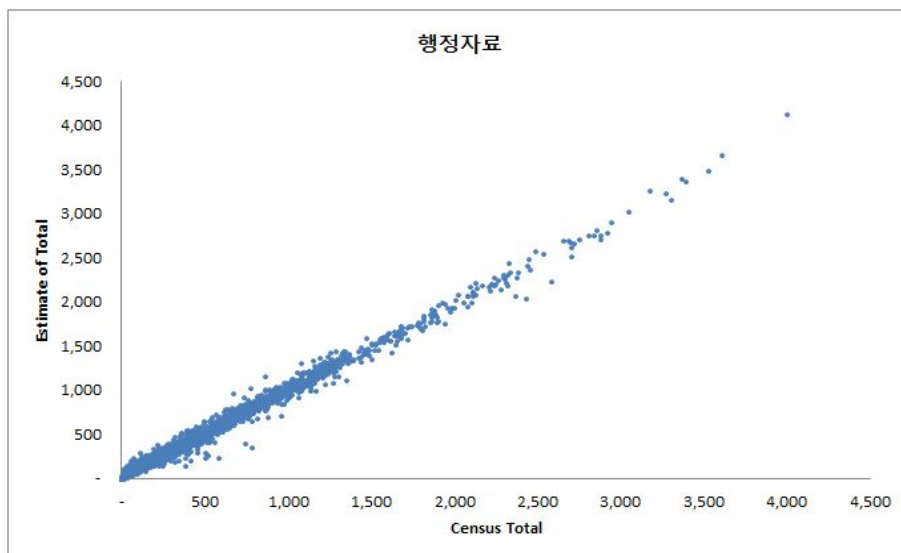
- 인구수 산출을 위한 능형회귀추정 방안은 제2절에서 정의한 방안1, 2, 3에 방안4를 추가하였으며, 방안4는 시군구별(12)*읍부/면부/동부별(3)*성별(2)*연령 10세별(11)로 조정 셀을 구성하였다. 12개 시범지역의 최종데이터를 사용하여 개인 가중치를 네 가지 방안을 통하여 보정한 능형회귀 가중치의 분포는 다음의 <표 3-6>과 같다. 제2절의 <표 3-2>의 중간데이터로부터 얻어진 분포와 비교하면 양 끝 쪽의 5%에 해당하는 분포가 더 안정적임을 알 수 있다. 이는 최종표본의 높은 모집단 대표성에 기인한다.

<표 3-6> 12개 시범지역의 인구 능형회귀 가중치의 분포 (최종데이터)

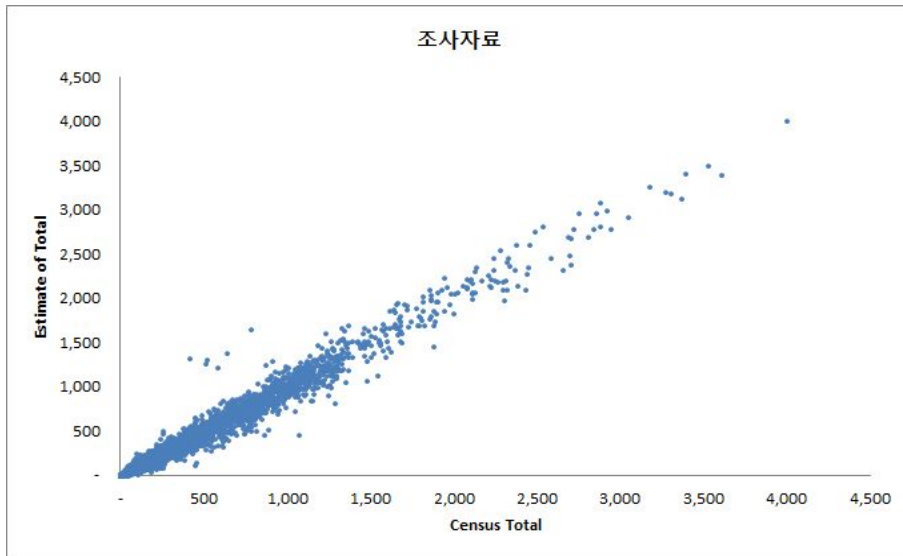
능형회귀가중치	mean	std.	min.	5%	Q1	median	Q3	95%	max.
방안1	0.983	0.074	0.092	0.885	1.000	1.000	1.000	1.000	1.944
방안2	0.983	0.061	0.331	0.856	1.000	1.000	1.000	1.000	1.215
방안3	0.974	0.074	0.539	0.824	0.979	1.000	1.000	1.009	1.009
방안4	0.978	0.074	0.291	0.840	1.000	1.000	1.000	1.000	1.150

- 관심 부모집단인 읍면동별(158)*성별(2)*연령5세별(21)의 참 값인 센서스 인구수를 x 축으로 행정자료, 조사자료 그리고 능형회귀추정 방안1, 2, 3, 4로 산출된 인구수를 y 축으로 한 산점도는 다음의 <그림 3-85>부터 <그림 3-90>이다. 그림을 전체적으로 살펴보면, 행정자료와 조사자료에 비하여 능형회귀추정 방안 1, 2, 3, 4의 결과가 조금 더 직선상에 가깝게 나타나고 있음을 확인할 수 있다.

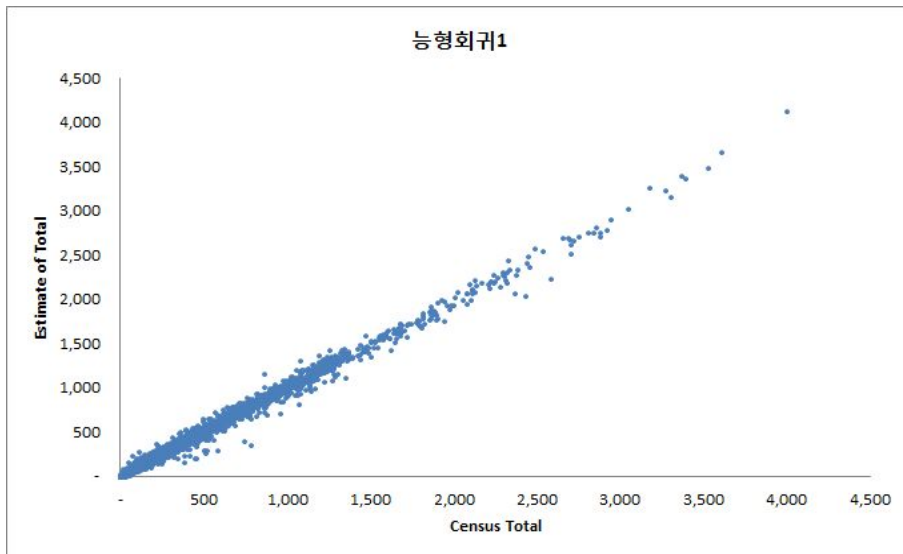
<그림 3-85> 행정자료와 센서스 부모집단별 인구수 산점도



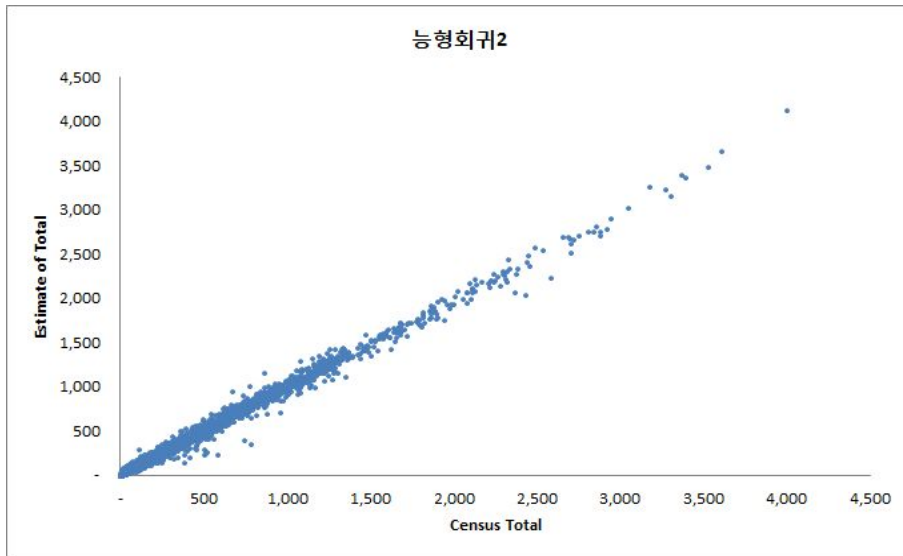
<그림 3-86> 조사자료와 센서스 부모집단별 인구수 산점도



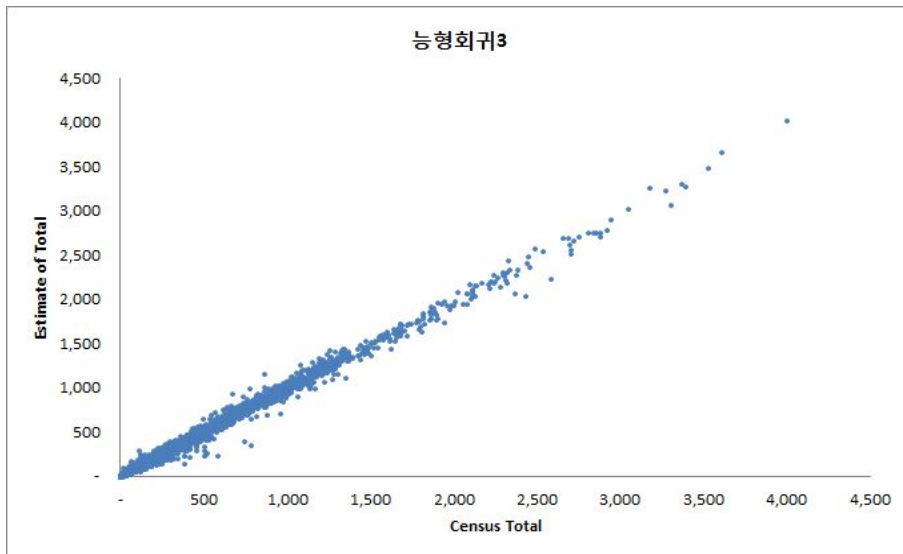
<그림 3-87> 능형회귀추정 방안1과 센서스 부모집단별 인구수 산점도



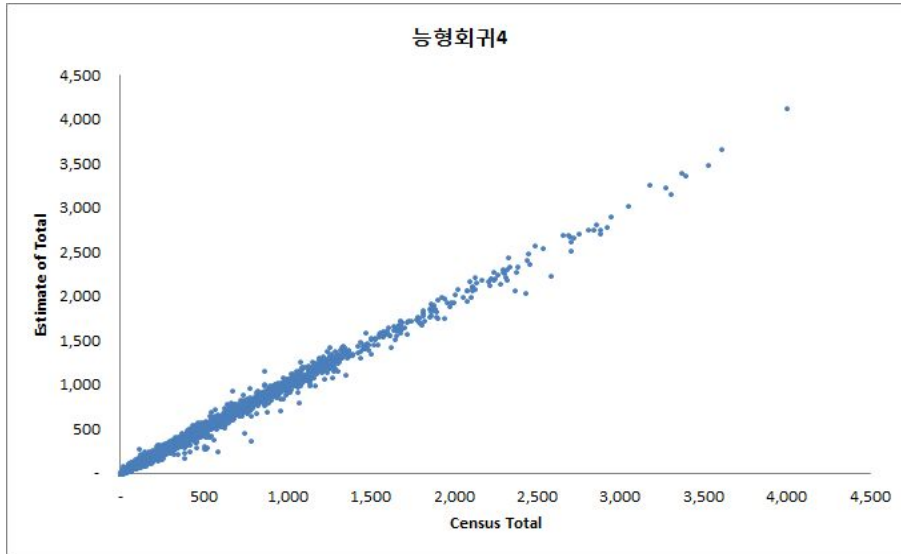
<그림 3-88> 능형회귀추정 방안2와 센서스 부모집단별 인구수 산점도



<그림 3-89> 능형회귀추정 방안2와 센서스 부모집단별 인구수 산점도



<그림 3-90> 능형회귀추정 방안2와 센서스 부모집단별 인구수 산점도



- 행정자료, 조사자료 그리고 능형회귀추정 방안1, 2, 3, 4를 통하여 관심 부모 집단 읍면동별(158)*성별(2)*연령5세별(21)의 지역별, 성별, 연령별 주변분포를 구성하여 그 *Rel.Bias*를 계산하였다. 여기서 지역별 주변분포는 시군구와 읍면동 수준에서 모두 살펴보았다. 성별, 연령별, 시군구별 그리고 읍면동별 주변분포의 *Rel.Bias*는 각각 다음의 <표 3-7>부터 <표 3-10>와 같다. <표 3-7>의 성별 수준의 분포에서는 전체적으로 *Rel.Bias*가 작은 값으로 센서스 총합에 비하여 크게 벗어나지 않음을 확인할 수 있다. <표 3-8>의 연령5세별 수준의 분포에서는 90세 이상의 범주에 대하여 능형회귀추정 방안3의 *Rel.Bias*가 -0.150, -0.187, -0.267로 크게 나타나고 있으나, 나머지 방안 1, 방안2, 방안4는 센서스총합에 비하여 크게 벗어나지 않고 있음을 확인할 수 있다. <표 3-9>의 시군구별 수준의 분포에서는 임실군(35350)과 신안군(36480) 그리고 울릉군(37430)의 *Rel.Bias*가 크게 나타나고 있음을 확인할 수 있다. <표 3-10>의 읍면동별 수준의 분포에서도 마찬가지로 규모가 작은 읍면동지역에서 행정자료의 *Rel.Bias*가 크게 나타나고 이로 인한 능형회귀추정의 결과 역시 약간의 보정은 하고 있으나 안정적이라고 할 수는 없다.

<표 3-7> 성별(2) 주변분포에서의 인구수 *Rel.Bias*의 비교

성별	센서스총합	<i>Rel.Bias</i>					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	표본조사
남자	949,345	0.045	0.021	0.020	0.009	0.014	-0.012
여자	942,178	0.025	0.013	0.015	0.008	0.011	-0.006

<표 3-8> 연령5세별(21) 주변분포에서의 인구수 *Rel.Bias*의 비교

연령	센서스총합	<i>Rel.Bias</i>					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	표본조사
0-4	102,512	0.012	0.007	0.008	0.009	0.007	0.030
5-9	101,231	0.011	0.006	0.007	0.008	0.006	0.003
10-14	118,871	0.007	-0.004	0.002	-0.006	-0.003	0.005
15-19	119,295	0.022	0.007	0.003	0.006	0.010	-0.060
20-24	107,991	0.010	-0.036	-0.011	-0.034	-0.035	-0.048
25-29	141,665	0.027	-0.013	0.007	-0.011	-0.011	-0.010
30-34	162,706	0.038	0.019	0.022	0.015	0.016	0.009
35-39	174,695	0.047	0.028	0.031	0.023	0.025	0.014
40-44	161,547	0.073	0.055	0.053	0.037	0.044	0.009
45-49	151,949	0.055	0.034	0.030	0.013	0.021	-0.013
50-54	140,624	0.074	0.059	0.044	0.030	0.036	-0.012
55-59	104,346	0.060	0.044	0.028	0.014	0.020	-0.003
60-64	85,945	0.039	0.027	0.000	0.014	0.031	-0.016
65-69	72,098	0.066	0.054	0.060	0.039	0.057	-0.006
70-74	63,165	-0.003	-0.004	-0.008	-0.003	-0.002	-0.018
75-79	43,677	-0.052	-0.053	-0.056	-0.052	-0.051	-0.032
80-84	24,254	-0.047	-0.056	-0.051	-0.055	-0.048	-0.074
85-89	10,824	-0.035	-0.044	-0.039	-0.043	-0.035	-0.141
90-94	3,403	0.022	-0.042	0.017	-0.150	-0.047	-0.264
95-99	642	-0.023	-0.086	-0.027	-0.187	-0.089	-0.339
100이상	83	0.048	0.048	0.046	-0.267	0.011	-0.399

<표 3-9> 시군구별(12) 주변분포에서의 인구수 $Rel.Bias$ 의 비교

시군구	센서스총합	$Rel.Bias$					
		등록자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	표본조사
11180	230447	0.023	0.018	0.019	0.017	0.013	-0.013
11190	374543	0.024	0.018	0.022	0.018	0.014	-0.001
23080	389057	0.009	0.006	0.007	0.003	0.008	0.008
24010	103998	-0.043	-0.047	-0.046	-0.049	-0.030	-0.022
31240	474160	0.027	0.014	0.012	-0.008	0.011	-0.008
32380	19252	0.085	0.059	0.077	-0.003	0.078	0.026
33390	31204	0.048	0.041	0.038	0.032	0.041	-0.012
35350	23490	0.242	0.119	0.120	0.089	0.101	-0.045
36040	77825	0.152	0.058	0.050	0.057	0.027	-0.039
36480	32987	0.317	0.135	0.123	0.125	0.069	-0.031
37030	126823	0.053	0.013	0.015	-0.004	-0.005	-0.050
37430	7737	0.307	0.223	0.202	0.236	0.221	0.061

<표 3-10> 읍면동별(158) 주변분포에서의 인구수 $Rel.Bias$ 의 비교

읍면동	센서스총합	$Rel.Bias$					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	조사자료
1118051	18,133	0.035	0.030	0.024	0.029	0.023	-0.051
1118052	31,061	-0.004	-0.005	-0.004	-0.010	-0.014	-0.016
1118053	22,582	0.023	0.023	0.023	0.016	0.012	0.020
1118054	27,351	0.028	0.017	0.024	0.021	0.016	-0.035
1118055	18,293	0.043	0.027	0.039	0.036	0.031	0.017
1118057	29,086	0.048	0.047	0.041	0.041	0.037	-0.039
1118058	24,851	-0.011	-0.020	-0.011	-0.017	-0.021	-0.011
1118059	12,357	0.005	0.003	0.005	-0.001	-0.005	-0.049
1118060	23,361	0.036	0.029	0.022	0.030	0.026	-0.029
1118061	23,372	0.034	0.034	0.034	0.028	0.024	0.060
1119054	31,264	-0.035	-0.036	-0.035	-0.041	-0.044	0.017
1119055	19,749	0.028	0.012	0.008	0.023	0.019	-0.039
1119056	32,240	-0.013	-0.013	-0.013	-0.018	-0.022	-0.006
1119061	16,073	-0.006	-0.006	-0.006	-0.011	-0.014	-0.033
1119062	20,531	-0.013	-0.013	-0.013	-0.019	-0.022	-0.005
1119063	20,835	-0.009	-0.010	-0.009	-0.015	-0.018	-0.042
1119065	18,348	0.027	0.026	0.027	0.021	0.017	0.003
1119066	12,247	0.050	0.050	0.050	0.043	0.039	0.045
1119067	16,016	0.052	0.052	0.052	0.045	0.041	0.041
1119068	23,224	0.039	0.032	0.037	0.034	0.030	-0.027
1119069	17,748	-0.004	-0.004	-0.004	-0.010	-0.014	-0.047
1119070	15,284	0.077	0.062	0.077	0.070	0.066	0.031
1119071	16,747	0.076	0.067	0.076	0.070	0.066	0.048
1119072	26,636	0.063	0.060	0.062	0.057	0.053	0.017
1119073	23,696	0.048	0.033	0.048	0.042	0.038	0.029
1119074	15,269	0.137	0.094	0.130	0.128	0.123	0.008
1119075	17,459	0.051	0.044	0.048	0.045	0.041	-0.021
1119076	31,177	-0.021	-0.021	-0.021	-0.026	-0.029	-0.017
2308051	29,074	0.009	0.006	0.009	0.004	0.009	-0.003
2308053	47,399	-0.011	-0.013	-0.011	-0.017	-0.011	-0.046
2308054	10,993	0.059	0.059	0.059	0.052	0.059	0.140
2308055	9,076	0.031	0.018	0.028	0.025	0.031	0.034

<표 3-10> 읍면동별(158) 주변분포에서의 인구수 $Rel.Bias$ 의 비교-(계속)

읍면동	센서스총합	$Rel.Bias$					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	조사자료
2308056	12,451	-0.009	-0.021	-0.017	-0.014	-0.009	-0.085
2308058	24,381	0.011	0.011	0.011	0.005	0.011	0.037
2308059	16,566	0.049	0.027	0.023	0.043	0.049	-0.016
2308060	17,933	0.010	0.010	0.010	0.004	0.010	-0.043
2308062	13,566	0.038	0.038	0.038	0.032	0.038	0.048
2308063	23,186	0.009	0.008	0.009	0.003	0.009	0.068
2308064	20,569	0.019	0.014	0.019	0.013	0.018	0.023
2308065	12,294	0.017	0.017	0.017	0.011	0.017	0.020
2308068	39,170	0.024	0.023	0.023	0.019	0.024	0.034
2308069	23,113	0.026	0.023	0.026	0.020	0.025	-0.005
2308070	22,431	-0.028	-0.023	-0.026	-0.032	-0.028	0.028
2308071	36,117	-0.023	-0.023	-0.023	-0.028	-0.023	-0.016
2308072	3,475	-0.024	-0.031	-0.024	-0.028	-0.024	0.127
2308073	27,263	0.007	0.007	0.007	0.001	0.007	0.005
2401051	4,625	0.000	-0.038	-0.005	-0.008	0.014	-0.053
2401054	6,153	-0.091	-0.084	-0.091	-0.096	-0.074	-0.043
2401056	8,050	-0.035	-0.062	-0.045	-0.041	-0.019	-0.075
2401058	9,428	-0.041	-0.040	-0.040	-0.047	-0.031	0.077
2401059	8,877	-0.013	-0.013	-0.013	-0.019	0.002	0.037
2401061	11,128	-0.053	-0.047	-0.053	-0.059	-0.041	0.013
2401062	5,941	-0.121	-0.080	-0.121	-0.126	-0.105	-0.005
2401063	6,509	-0.074	-0.065	-0.074	-0.080	-0.060	-0.012
2401064	5,639	-0.132	-0.162	-0.135	-0.136	-0.109	-0.188
2401068	6,676	0.030	0.040	0.030	0.023	0.044	0.038
2401071	13,598	-0.035	-0.053	-0.049	-0.041	-0.023	-0.099
2401072	9,246	-0.038	-0.039	-0.038	-0.044	-0.026	0.000
2401073	8,128	-0.001	-0.004	-0.001	-0.009	0.010	-0.018
3124012	63,694	0.005	0.005	0.005	0.000	0.005	-0.007
3124013	14,821	0.130	0.089	0.108	0.124	0.130	-0.015
3124014	49,188	-0.002	-0.002	-0.003	-0.007	-0.002	0.022
3124031	6,853	0.097	0.037	0.049	-0.118	-0.011	0.006
3124033	5,101	0.193	0.137	0.185	-0.027	0.079	0.048

<표 3-10> 읍면동별(158) 주변분포에서의 인구수 $Rel.Bias$ 의 비교-(계속)

읍면동	센서스총합	$Rel.Bias$					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	조사자료
3124035	4,368	0.288	0.031	-0.057	0.050	0.162	-0.160
3124036	8,597	0.241	0.210	0.231	0.018	0.124	-0.001
3124037	4,587	0.410	0.280	0.235	0.164	0.275	0.066
3124038	8,007	0.211	0.085	0.069	-0.039	0.079	-0.040
3124039	8,890	0.207	0.093	0.052	-0.016	0.092	-0.110
3124042	3,157	0.294	0.202	0.174	0.042	0.155	-0.034
3124043	11,050	0.148	0.132	0.127	-0.080	0.032	0.034
3124044	4,360	0.195	0.141	0.195	-0.058	0.072	-0.025
3124051	21,177	0.052	0.042	0.012	0.045	0.051	-0.091
3124052	34,872	-0.037	-0.037	-0.037	-0.040	-0.037	0.007
3124053	27,098	-0.008	-0.005	-0.009	-0.012	-0.009	-0.008
3124054	26,818	-0.009	-0.009	-0.009	-0.012	-0.009	0.006
3124055	19,637	-0.022	-0.022	-0.022	-0.026	-0.022	-0.008
3124056	14,661	0.016	-0.001	-0.003	0.012	0.016	-0.097
3124057	22,817	0.019	0.019	0.019	0.014	0.018	0.003
3124060	32,109	-0.009	-0.010	-0.009	-0.013	-0.009	-0.016
3124061	44,160	-0.018	-0.019	-0.019	-0.022	-0.019	0.005
3124062	38,138	-0.037	-0.037	-0.037	-0.040	-0.037	0.018
3238011	11,257	0.070	0.065	0.070	0.065	0.065	0.045
3238031	3,412	0.103	0.062	0.103	-0.105	0.094	0.013
3238032	2,081	0.078	0.025	0.019	-0.120	0.069	-0.159
3238033	1,293	0.131	0.069	0.127	-0.079	0.121	0.277
3238034	1,209	0.127	0.037	0.115	-0.070	0.119	-0.055
3339014	29,169	0.041	0.040	0.041	0.036	0.040	-0.008
3339031	2,035	0.159	0.050	-0.002	-0.035	0.050	-0.074
3535011	5,352	0.218	0.164	0.186	0.212	0.164	0.009
3535031	1,190	0.238	0.093	0.171	0.056	0.089	-0.116
3535032	1,263	0.466	0.317	0.232	0.220	0.246	-0.210
3535033	1,346	0.265	0.065	0.134	0.072	0.099	-0.178
3535034	1,370	0.280	0.135	0.113	0.086	0.115	0.001
3535035	3,750	0.220	0.105	0.035	0.020	0.045	-0.037
3535036	1,123	0.246	0.041	0.013	0.064	0.084	-0.201

<표 3-10> 읍면동별(158) 주변분포에서의 인구수 $Rel.Bias$ 의 비교-(계속)

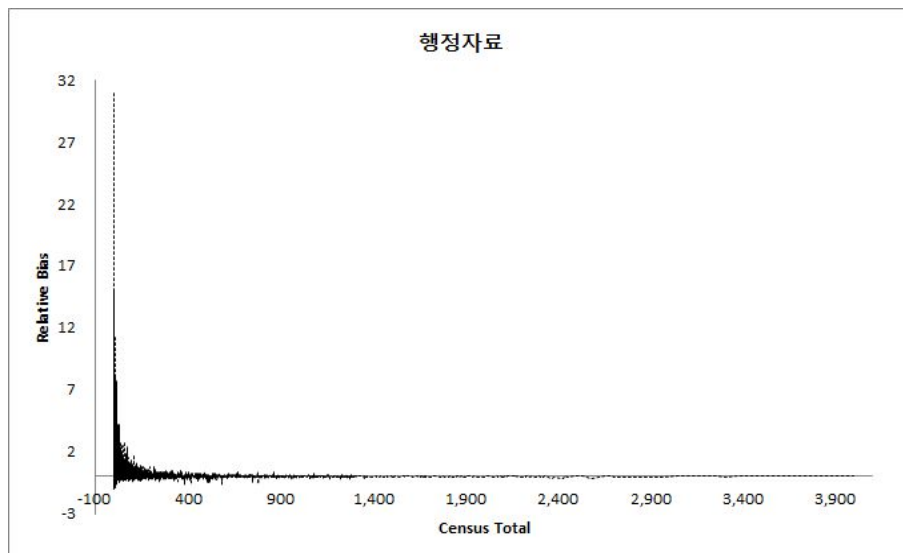
읍면동	센서스총합	$Rel.Bias$					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	조사자료
3535037	1,425	0.154	0.002	-0.013	0.001	0.031	-0.140
3535038	3,021	0.249	0.187	0.249	0.033	0.067	0.117
3535039	1,504	0.173	0.024	-0.009	-0.011	0.019	-0.028
3535040	1,016	0.271	0.093	0.046	0.088	0.123	-0.027
3535041	1,130	0.246	-0.017	0.046	0.056	0.089	-0.157
3604011	7,142	0.135	0.019	0.000	0.129	0.019	-0.140
3604031	2,597	0.319	0.160	0.122	0.111	0.075	-0.005
3604032	2,969	0.271	0.085	0.073	0.058	0.021	-0.042
3604033	1,386	0.294	0.267	0.334	0.093	0.063	0.207
3604034	2,328	0.252	0.121	0.106	0.061	0.030	-0.010
3604035	2,617	0.231	0.054	0.006	0.046	0.022	-0.060
3604036	3,632	0.211	0.073	0.056	0.020	-0.009	-0.096
3604037	1,858	0.255	0.115	0.102	0.068	0.039	-0.047
3604038	4,350	0.254	0.137	0.160	0.051	0.020	-0.003
3604039	4,671	0.221	0.086	0.064	0.024	-0.006	-0.080
3604040	3,081	0.390	0.138	0.054	0.149	0.105	-0.064
3604041	1,845	0.203	-0.002	-0.044	0.028	0.001	-0.229
3604042	3,862	0.338	0.208	0.189	0.131	0.099	0.085
3604051	7,546	-0.056	-0.058	-0.052	-0.061	-0.072	0.014
3604052	4,179	-0.049	-0.063	-0.067	-0.055	-0.064	-0.103
3604054	6,411	0.110	0.059	0.108	0.103	0.089	0.006
3604055	10,152	-0.016	-0.018	-0.016	-0.021	-0.037	-0.033
3604058	4,143	0.128	0.036	-0.009	0.121	0.107	-0.072
3604060	3,056	0.294	0.192	0.188	0.284	0.270	-0.001
3648011	4,109	0.201	-0.046	-0.073	0.194	-0.046	-0.150
3648031	1,671	0.250	0.192	0.238	0.047	0.027	0.056
3648032	3,030	0.169	0.123	0.154	-0.025	-0.053	0.010
3648033	1,826	0.281	0.168	0.244	0.096	0.078	0.087
3648034	3,099	0.225	0.076	0.078	0.034	0.008	-0.052
3648035	2,423	0.252	0.108	0.044	0.058	0.027	-0.077
3648036	2,930	0.544	0.331	0.279	0.257	0.208	0.088
3648037	1,275	0.576	0.285	0.274	0.327	0.295	-0.050

<표 3-10> 읍면동별(158) 주변분포에서의 인구수 $Rel.Bias$ 의 비교-(계속)

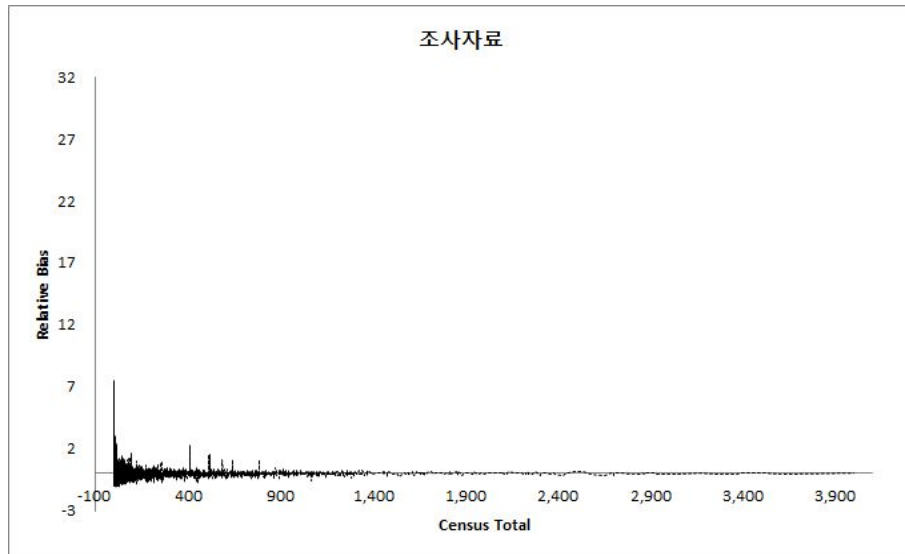
읍면동	센서스총합	$Rel.Bias$					
		행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	조사자료
3648038	1,347	0.310	0.241	0.310	0.080	0.046	0.200
3648039	1,287	0.354	0.245	0.211	0.147	0.120	0.061
3648040	2,585	0.350	0.008	-0.060	0.139	0.111	-0.185
3648041	936	0.263	0.072	0.158	0.076	0.056	0.017
3648042	1,745	0.245	0.155	0.184	0.055	0.035	-0.067
3648043	4,724	0.441	0.170	0.110	0.193	0.154	-0.063
3703011	7,403	0.122	0.117	0.122	0.116	0.117	0.069
3703031	2,809	0.147	0.037	-0.013	-0.045	-0.015	-0.206
3703032	2,932	0.205	0.030	-0.038	0.011	0.042	-0.219
3703033	2,539	0.124	0.085	0.023	-0.067	-0.036	-0.013
3703034	3,548	0.092	0.071	0.081	-0.075	-0.048	-0.058
3703035	5,105	0.062	-0.011	-0.043	-0.109	-0.079	-0.189
3703036	2,996	0.176	0.085	0.063	-0.033	0.000	-0.069
3703037	3,757	0.099	0.039	0.050	-0.089	-0.058	-0.027
3703038	1,773	0.262	0.031	0.003	0.051	0.083	-0.129
3703039	2,044	0.253	0.029	-0.021	0.050	0.082	-0.159
3703040	2,770	0.178	0.145	0.174	0.002	0.034	-0.030
3703041	1,673	0.124	0.042	0.059	-0.042	-0.015	0.048
3703042	1,142	0.143	-0.010	0.083	-0.018	0.010	0.110
3703043	1,886	0.189	0.012	-0.081	0.017	0.049	-0.147
3703044	990	0.186	0.117	0.148	0.018	0.046	0.065
3703055	5,335	0.029	0.028	0.019	0.022	0.007	0.050
3703056	26,291	-0.012	-0.043	-0.026	-0.017	-0.033	-0.128
3703058	22,971	-0.012	-0.014	-0.012	-0.018	-0.033	0.004
3703059	10,225	-0.019	-0.027	-0.025	-0.024	-0.039	-0.028
3703060	8,661	0.071	0.028	0.071	0.063	0.047	-0.012
3703061	9,973	0.055	0.041	0.055	0.048	0.030	-0.005
3743011	5,579	0.289	0.215	0.193	0.283	0.215	0.071
3743031	1,078	0.431	0.274	0.178	0.181	0.308	0.002
3743032	1,080	0.271	0.218	0.267	0.051	0.166	0.067

- *Rel.Bias*와 인구수와의 관계를 살펴보기 위하여 읍면동별(158)*성별(2)*연령 5세별(21)로 구성된 관심 부모집단의 통계를 행정자료, 조사자료, 능형회귀추정 방안1, 2, 3, 4에서 산출하여 *Rel.Bias*를 계산하였다. 다음의 <그림 3-91>부터 <그림 3-95>은 참 값인 센서스 인구수를 x 축으로 행정자료, 조사자료 그리고 능형회귀추정 방안1, 2, 3, 4로 산출된 인구수 *Rel.Bias*를 y 축으로 나타낸 플롯이다. 인구수 산출을 위한 다섯 가지 방안의 결과를 종합적으로 살펴보면, 센서스총합이 작은 곳에서 *Rel.Bias*의 변동이 매우 큼을 확인할 수 있다. 이것은 *Rel.Bias*에서 분모에 해당하는 센서스총합이 매우 작음으로써 발생하는 것으로, 즉 관심부모집단의 참 값이 작을수록 그 값을 추정함에 있어 어려움이 따르는 것을 확인할 수 있다. 이러한 문제는 다섯 가지 방안 모두에서 나타나는 현상이다.

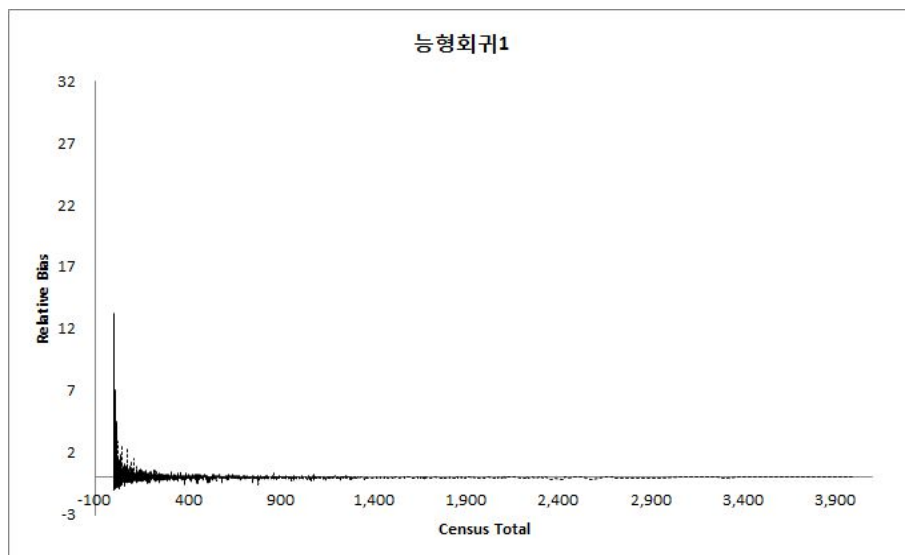
<그림 3-91> 부모집단별 행정자료 *Rel.Bias*와 센서스 인구수 플롯



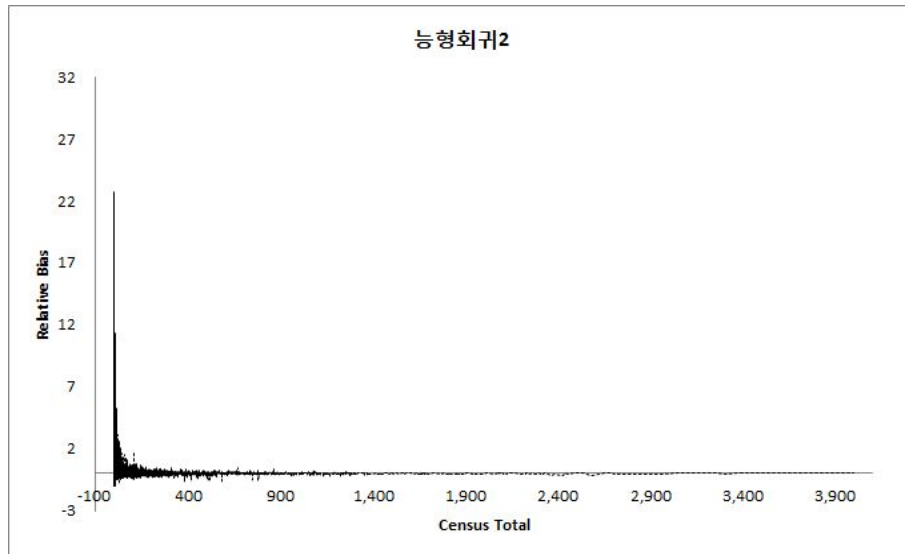
<그림 3-92> 부모집단별 조사자료 $Rel.Bias$ 와 센서스 인구수 플롯



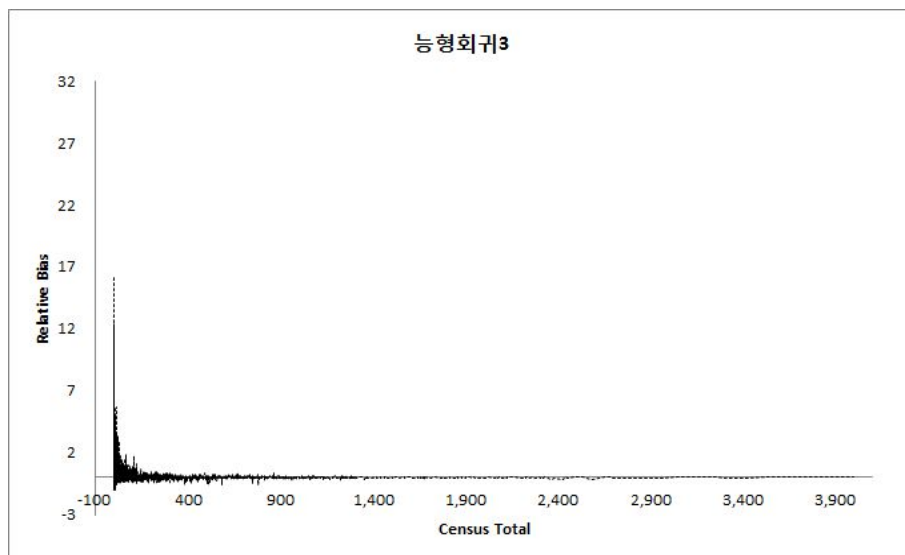
<그림 3-93> 부모집단별 능형회귀추정 방안1 $Rel.Bias$ 와 센서스 인구수 플롯



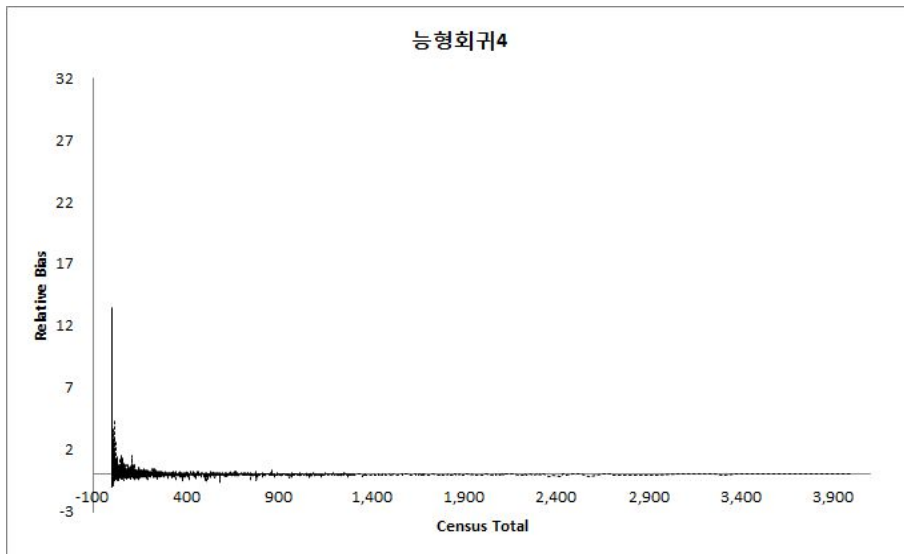
<그림 3-94> 부모집단별 능형회귀추정 방안2 $Rel.Bias$ 와 센서스 인구수 플롯



<그림 3-95> 부모집단별 능형회귀추정 방안3 $Rel.Bias$ 와 센서스 인구수 플롯



<그림 3-96> 부모집단별 능형회귀추정 방안4 $Rel.Bias$ 와 센서스 인구수 플롯



- 관심 부모집단별 다섯 가지 추정 방법의 $Rel.Bias$ 와 센서스 인구수의 관계를 살펴보기 위하여 센서스 인구수를 12개의 그룹으로 구성하여 그룹 내의 $Rel.Bias$ 의 평균과 표준편차를 다섯 가지 추정 방법에 대하여 계산한 결과는 다음의 <표 3-11>과 같다. 각 그룹별 $Rel.Bias$ 의 평균은 행정자료가 다른 방법에 비하여 대체로 크게 나타나고 있으며, ‘10명 미만’부터 ‘20-40명 미만’의 그룹에서는 능형회귀추정 방안1~방안4 모두 $Rel.Bias$ 가 0.144~0.497 사이의 값으로 크게 나타나고 있다. 따라서 능형회귀추정 방안을 통한 관심 부모집단의 규모가 작은 곳에서의 추정에 있어서는 주의를 필요로 한다.

<표 3-11> 센서스 인구수에 따른 *Rel.Bias*의 평균과 표준편차

센서스 인구수 (셀 수)	통계량	행정자료	능형회귀1	능형회귀2	능형회귀3	능형회귀4	조사자료
10명 미만 (554)	mean sd	0.524 2.033	0.245 1.136	0.444 1.728	0.169 1.274	0.201 1.044	-0.225 1.190
10-20명미만 (434)	mean sd	0.670 1.013	0.304 0.605	0.497 0.797	0.331 0.602	0.288 0.520	-0.105 0.543
20-40명 미만 (593)	mean sd	0.447 0.668	0.155 0.407	0.262 0.487	0.168 0.391	0.144 0.312	-0.084 0.412
40-60명 미만 (502)	mean sd	0.261 0.416	0.113 0.288	0.126 0.302	0.056 0.235	0.054 0.213	-0.036 0.347
60-80명 미만 (378)	mean sd	0.217 0.366	0.095 0.262	0.089 0.267	0.049 0.231	0.057 0.206	-0.042 0.293
80-100명 미만 (322)	mean sd	0.144 0.259	0.060 0.215	0.028 0.183	0.014 0.169	0.037 0.164	-0.046 0.309
100-150명 미만 (583)	mean sd	0.114 0.228	0.060 0.179	0.032 0.195	0.020 0.181	0.043 0.171	-0.028 0.236
150-200명 미만 (408)	mean sd	0.080 0.164	0.041 0.151	0.032 0.143	0.018 0.138	0.032 0.132	-0.023 0.207
200-300명 미만 (559)	mean sd	0.055 0.136	0.032 0.122	0.028 0.119	0.012 0.125	0.029 0.109	0.003 0.203
300-500명 미만 (674)	mean sd	0.029 0.109	0.020 0.099	0.018 0.097	0.003 0.092	0.016 0.084	-0.007 0.184
500-1,000명 미만 (805)	mean sd	0.019 0.082	0.013 0.077	0.015 0.080	0.012 0.080	0.012 0.076	-0.008 0.167
1,000-4,000명 미만 (454)	mean sd	0.001 0.048	-0.001 0.050	-0.001 0.048	-0.004 0.046	-0.004 0.046	-0.002 0.096
총합계 (6,266)	mean sd	0.203 0.747	0.090 0.429	0.127 0.615	0.068 0.456	0.073 0.381	-0.049 0.452

2. 가구

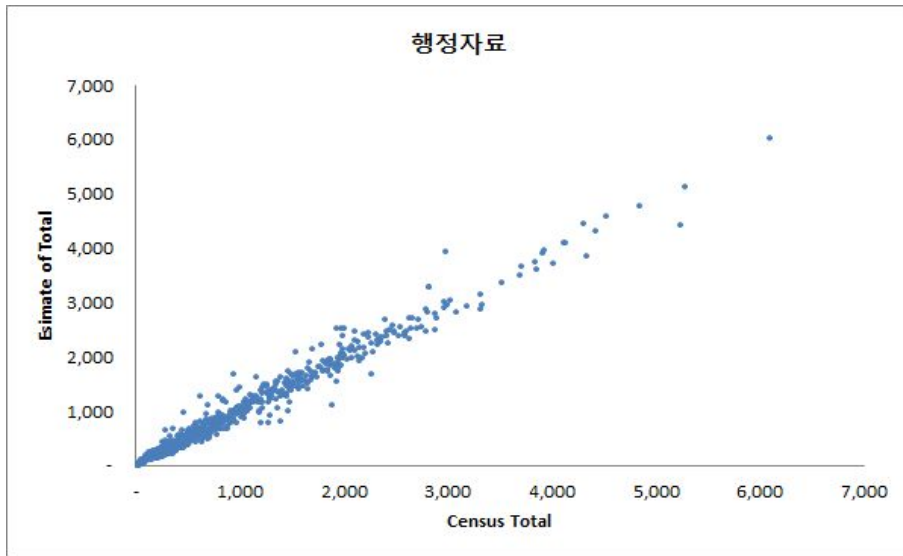
- 12개 시범지역의 최종데이터를 사용하여 가구 가중치를 능형회귀추정 방안을 통하여 보정한 능형회귀 가중치는 다음의 <표 3-12>와 같다. 제2절의 <표 3-4>의 중간데이터로부터 얻어진 분포와 비교하면 인구부분과 마찬가지로 양 끝 쪽의 5%에 해당하는 분포가 더 안정적임을 알 수 있다.

<표 3-12> 12개 시범지역의 가구 능형회귀 가중치의 분포 (최종데이터)

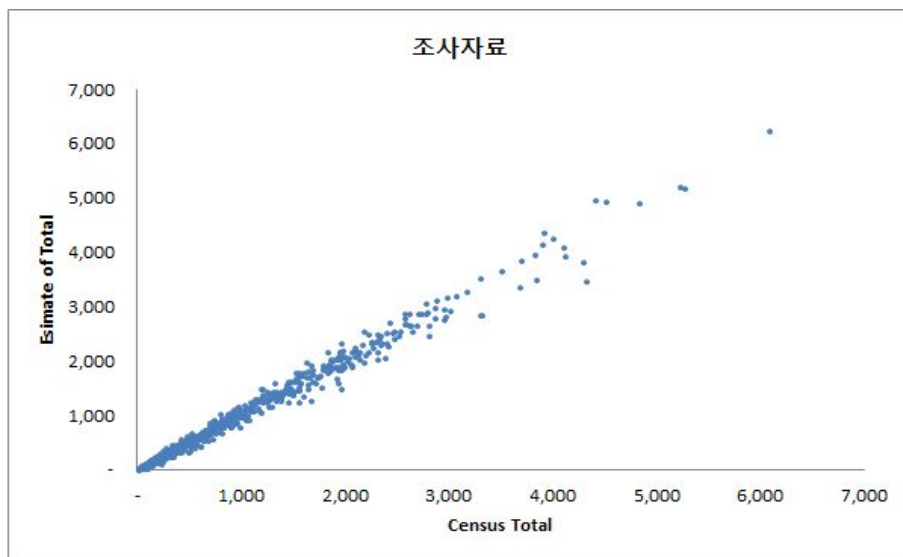
능형회귀가중치	mean	std.	min.	5%	Q1	median	Q3	95%	max.
가구	0.981	0.086	0.217	0.816	1.000	1.000	1.000	1.003	1.519

- 관심 부모집단인 읍면동별(158)*가구원수별(5) 참 값인 센서스 가구수를 x 축으로 행정자료, 조사자료 그리고 능형회귀추정 방안으로 산출된 가구수를 y 축으로 한 산점도는 다음의 <그림 3-97>부터 <그림 3-99>이다. 그림을 전체적으로 살펴보면, 능형회귀추정 방안을 적용한 것이 센서스 가구수가 적은 쪽에서는 행정자료보다 그리고 센서스 가구수가 많은 쪽에서는 조사자료보다 조금 더 직선상에 가깝게 나타나고 있음을 확인할 수 있다.

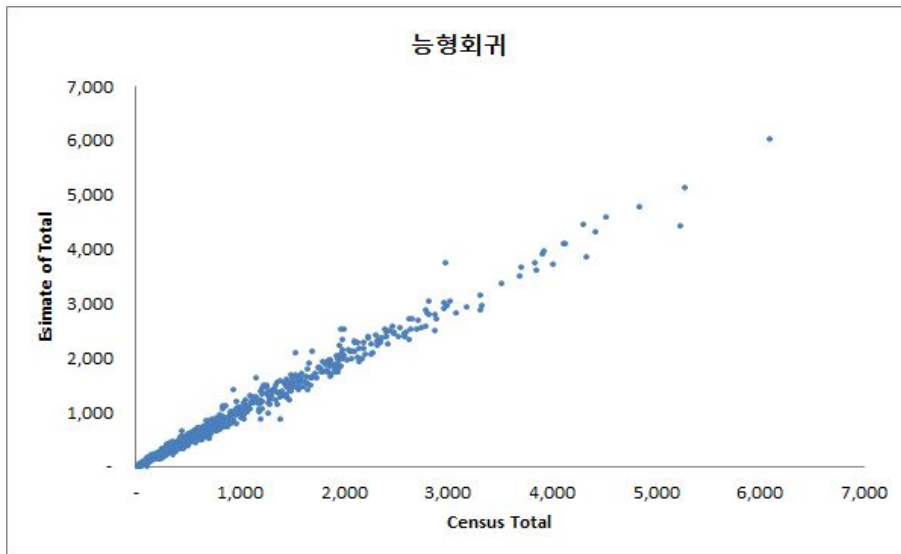
<그림 3-97> 행정자료와 센서스 부모집단별 가구수 산점도



<그림 3-98> 조사자료와 센서스 부모집단별 가구수 산점도



<그림 3-99> 능형회귀추정 방안과 센서스 부모집단별 가구수 산정도



- 행정자료, 조사자료 그리고 능형회귀추정 방안을 하여 관심 부모집단 읍면동 별(158)*가구원수별(5)의 지역과 가구원수 별 주변분포를 구성하여 그 *Rel.Bias*를 계산하였다. 여기서 지역별 주변분포는 시군구와 읍면동 수준에서 모두 살펴보았다. 가구원수별, 시군구별 그리고 읍면동별 주변분포의 *Rel.Bias*는 각각 다음의 <표 3-13>, <표 3-14> 그리고 <표 3-15>와 같다. <표 3-13>의 가구원 수준의 분포에서는 ‘5인 이상’가구를 제외한 나머지 수준에서 *Rel.Bias*가 작게 나타나고 있다. <표 3-14>의 시군구별 수준의 분포에서는 모든 시군구에서 *Rel.Bias*가 작게 나타나고 있으며, <표 3-15>의 읍면동 수준의 분포에서는 일부 읍면동을 제외한 대부분의 읍면동에서 *Rel.Bias*가 작게 나타나고 있음을 확인할 수 있다.

<표 3-13> 가구원별(5) 주변분포에서의 가구수 *Rel.Bias*의 비교

가구원수	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
1인	163,431	0.114	0.059	-0.019
2인	165,408	-0.045	-0.033	0.025
3인	143,913	0.032	0.016	0.024
4인	155,003	0.044	0.031	0.031
5인 이상	56,768	0.188	0.148	0.027

<표 3-14> 시군구별(12) 주변분포에서의 가구수 *Rel.Bias*의 비교

시군구	센서스총합	<i>Rel.Bias</i>		
		등록자료	능형회귀	조사자료
11180	86,533	0.055	0.036	-0.010
11190	138,637	0.057	0.042	0.008
23080	128,975	0.025	0.022	0.039
24010	42,538	-0.069	-0.049	0.008
31240	158,423	0.050	0.027	0.021
32380	6,919	0.173	0.108	0.062
33390	11,559	0.029	0.025	0.021
35350	10,693	0.127	0.057	0.012
36040	31,795	0.154	0.081	0.029
36480	16,114	0.188	0.053	0.004
37030	48,843	0.014	0.004	0.000
37430	3,494	0.297	0.087	0.023

<표 3-15> 읍면동별(158) 주변분포에서의 가구수 *Rel.Bias*의 비교

읍면동	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
1118051	8,689	-0.017	-0.021	-0.074
1118052	12,299	0.005	0.006	-0.002
1118053	8,418	0.078	0.058	0.025
1118054	10,531	0.080	0.034	-0.043
1118055	6,868	0.111	0.041	-0.044
1118057	10,986	0.092	0.062	-0.027
1118058	7,950	0.000	0.000	0.018
1118059	4,177	0.030	0.028	0.044
1118060	8,211	0.086	0.080	0.009
1118061	8,404	0.084	0.075	0.026
1119054	10,789	-0.022	-0.044	-0.002
1119055	7,411	0.092	0.092	0.004
1119056	11,629	-0.042	-0.042	0.023
1119061	6,101	0.011	0.009	0.030
1119062	6,857	0.019	0.028	0.041
1119063	8,881	0.003	0.003	-0.034
1119065	6,794	0.075	0.075	0.024
1119066	4,557	0.120	0.120	0.096
1119067	6,056	0.108	0.055	0.025
1119068	8,542	0.055	0.055	-0.001
1119069	6,038	-0.005	-0.005	0.014
1119070	5,750	0.138	0.058	-0.056
1119071	6,505	0.094	0.086	-0.011
1119072	9,437	0.135	0.078	0.013
1119073	9,028	0.081	0.064	0.000
1119074	6,582	0.229	0.200	0.014
1119075	7,106	0.097	0.075	-0.013
1119076	10,574	-0.003	-0.003	0.004
2308051	9,307	0.029	0.029	0.070
2308053	15,955	-0.028	-0.031	-0.021
2308054	4,283	0.093	0.093	0.065
2308055	3,093	0.036	0.036	0.015

<표 3-15> 읍면동별(158) 주변분포에서의 가구수 *Rel.Bias*의 비교-(계속)

읍면동	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
2308056	4,156	-0.027	-0.027	-0.023
2308058	8,837	0.029	0.029	0.070
2308059	5,894	0.084	0.084	0.018
2308060	6,237	0.013	0.013	0.048
2308062	4,999	0.060	0.063	0.079
2308063	7,156	0.013	0.013	0.073
2308064	7,115	0.015	0.015	0.033
2308065	4,474	0.053	0.053	0.024
2308068	12,684	0.066	0.066	0.073
2308069	7,089	0.083	0.025	0.010
2308070	6,963	0.003	0.003	0.092
2308071	10,604	-0.014	-0.013	0.030
2308072	1,084	0.003	0.018	0.125
2308073	9,045	0.018	0.018	0.018
2401051	2,020	0.204	-0.108	-0.093
2401054	2,946	-0.179	-0.068	0.028
2401056	3,944	-0.049	-0.031	-0.045
2401058	3,445	-0.049	-0.049	0.039
2401059	3,758	-0.010	-0.006	0.040
2401061	4,369	-0.101	-0.053	0.025
2401062	2,714	-0.211	-0.126	0.056
2401063	3,012	-0.202	-0.164	0.003
2401064	2,768	-0.243	-0.043	0.047
2401068	3,067	0.009	0.009	0.040
2401071	4,694	-0.041	-0.042	-0.036
2401072	3,217	-0.026	-0.029	-0.031
2401073	2,584	0.050	0.032	0.031
3124012	20,426	0.003	0.003	0.040
3124013	5,164	0.187	0.179	0.068
3124014	17,742	0.005	0.005	0.002
3124031	2,376	0.149	0.014	-0.035
3124033	1,827	0.291	0.247	0.112

<표 3-15> 읍면동별(158) 주변분포에서의 가구수 $Rel.Bias$ 의 비교-(계속)

읍면동	센서스총합	$Rel.Bias$		
		행정자료	능형회귀	조사자료
3124035	1,591	0.417	0.070	-0.103
3124036	3,244	0.284	0.210	0.032
3124037	1,766	0.591	0.217	0.131
3124038	3,175	0.356	0.229	0.078
3124039	3,234	0.275	0.230	-0.011
3124042	1,208	0.420	0.133	-0.097
3124043	3,943	0.238	0.116	0.037
3124044	1,414	0.449	0.190	0.000
3124051	7,262	0.116	0.112	-0.004
3124052	14,133	-0.067	-0.067	0.019
3124053	8,401	-0.005	-0.001	0.036
3124054	8,405	-0.007	-0.007	0.042
3124055	5,545	0.010	0.010	0.026
3124056	4,284	0.037	0.037	-0.045
3124057	7,516	0.033	0.030	0.019
3124060	9,697	-0.008	-0.011	0.019
3124061	13,515	-0.028	-0.028	0.033
3124062	12,555	-0.045	-0.045	-0.003
3238011	3,975	0.159	0.112	0.079
3238031	1,244	0.190	0.096	0.044
3238032	713	0.248	0.092	-0.089
3238033	500	0.226	0.226	0.336
3238034	487	0.074	0.004	-0.098
3339014	10,708	0.023	0.026	0.026
3339031	851	0.095	0.005	-0.038
3535011	2,133	0.174	0.166	0.097
3535031	601	0.062	0.007	-0.065
3535032	602	0.249	-0.153	-0.351
3535033	649	0.151	0.018	-0.120
3535034	678	0.081	0.081	0.089
3535035	1,720	0.098	0.009	0.027
3535036	563	0.112	0.009	-0.129

<표 3-15> 읍면동별(158) 주변분포에서의 가구수 *Rel.Bias*의 비교-(계속)

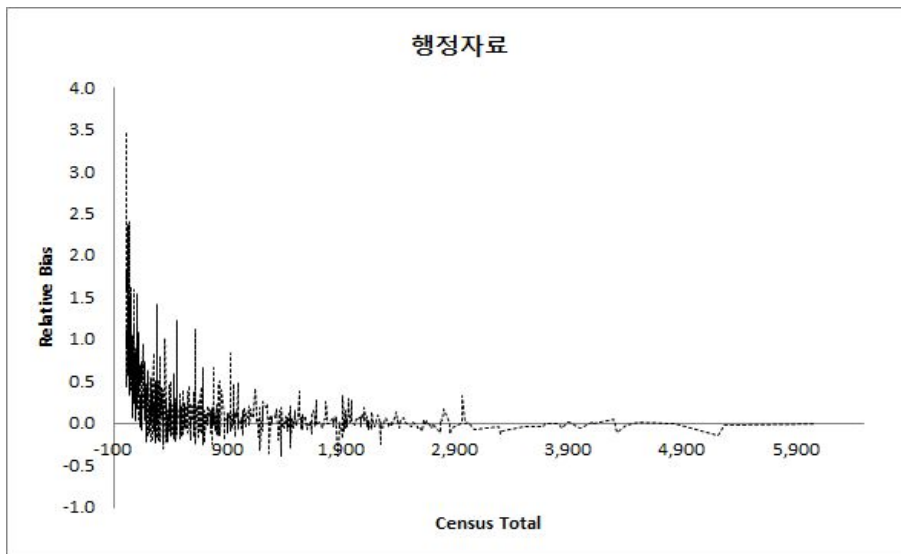
읍면동	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
3535037	723	0.030	-0.018	-0.025
3535038	1,244	0.161	0.223	0.156
3535039	711	0.052	0.004	0.031
3535040	509	0.187	0.098	0.085
3535041	560	0.100	-0.113	-0.048
3604011	2,670	0.215	0.184	0.033
3604031	1,237	0.211	0.098	-0.021
3604032	1,347	0.195	0.072	-0.006
3604033	716	0.196	0.216	0.193
3604034	1,138	0.147	0.086	0.078
3604035	1,329	0.160	0.055	0.007
3604036	1,699	0.175	0.072	-0.015
3604037	975	0.185	0.050	-0.009
3604038	1,918	0.229	0.083	-0.017
3604039	1,706	0.256	0.153	0.073
3604040	1,255	0.390	0.181	0.022
3604041	840	0.155	-0.033	-0.091
3604042	1,892	0.223	0.137	0.055
3604051	2,693	-0.061	-0.034	0.051
3604052	1,565	-0.041	-0.067	-0.109
3604054	2,445	0.158	0.099	0.054
3604055	3,454	0.021	0.022	0.085
3604058	1,663	0.146	0.146	0.057
3604060	1,253	0.318	0.112	0.012
3648011	2,095	0.087	-0.064	-0.086
3648031	789	0.150	0.087	0.116
3648032	1,266	0.184	0.080	-0.025
3648033	974	0.172	0.115	0.103
3648034	1,524	0.105	0.037	-0.021
3648035	1,163	0.197	0.011	-0.002
3648036	1,405	0.409	0.149	0.038
3648037	680	0.353	0.207	-0.038

<표 3-15> 읍면동별(158) 주변분포에서의 가구수 $Rel.Bias$ 의 비교-(계속)

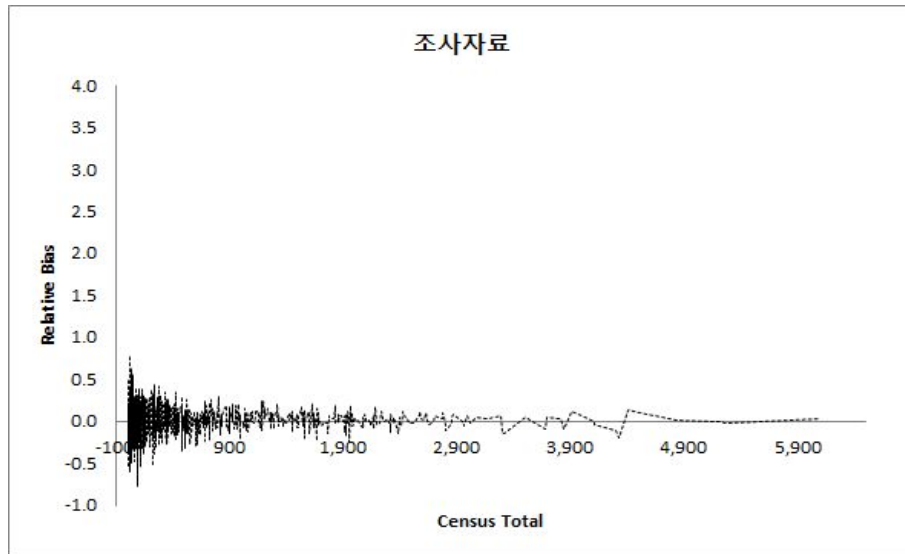
읍면동	센서스총합	$Rel.Bias$		
		행정자료	능형회귀	조사자료
3648038	617	0.183	0.251	0.281
3648039	703	0.172	0.111	0.140
3648040	1,350	0.165	-0.062	-0.114
3648041	522	0.096	0.023	0.095
3648042	952	0.047	-0.044	-0.037
3648043	2,074	0.279	0.082	-0.016
3703011	2,984	0.069	0.069	0.065
3703031	1,095	0.111	0.013	-0.087
3703032	1,145	0.196	0.061	-0.012
3703033	1,048	0.045	-0.026	0.019
3703034	1,592	0.011	-0.006	0.000
3703035	1,772	0.075	0.075	0.042
3703036	1,221	0.105	0.036	-0.057
3703037	1,387	0.111	0.086	0.042
3703038	767	0.113	-0.014	-0.131
3703039	915	0.105	-0.035	-0.079
3703040	1,422	0.054	0.016	-0.080
3703041	800	0.070	0.064	0.170
3703042	590	0.042	0.015	0.154
3703043	990	0.064	-0.037	-0.119
3703044	485	0.113	0.205	0.212
3703055	2,092	0.017	-0.018	-0.012
3703056	8,695	-0.030	-0.020	-0.025
3703058	8,289	-0.056	-0.023	0.020
3703059	3,748	-0.052	-0.059	-0.021
3703060	3,533	0.033	0.040	0.026
3703061	4,273	-0.009	0.008	-0.009
3743011	2,427	0.309	0.071	0.024
3743031	540	0.315	0.125	0.048
3743032	527	0.224	0.125	-0.009

- *Rel.Bias*와 가구수와의 관계를 살펴보기 위하여 읍면동별(158)*가구원수별(5)로 구성된 관심 부모집단의 통계를 행정자료, 조사자료 그리고 능형회귀추정 방안으로 산출하여 *Rel.Bias*를 계산하였다. 다음의 <그림 3-100>부터 <그림 3-102>는 참 값인 센서스 가구수를 x 축으로 행정자료, 조사자료 그리고 능형회귀추정 방안의 인구수 *Rel.Bias*를 y 축으로 나타낸 플롯이다. 인구 부분과 마찬가지로 센서스총합이 작은 곳에서 *Rel.Bias*의 변동이 매우 큼을 확인할 수 있다.

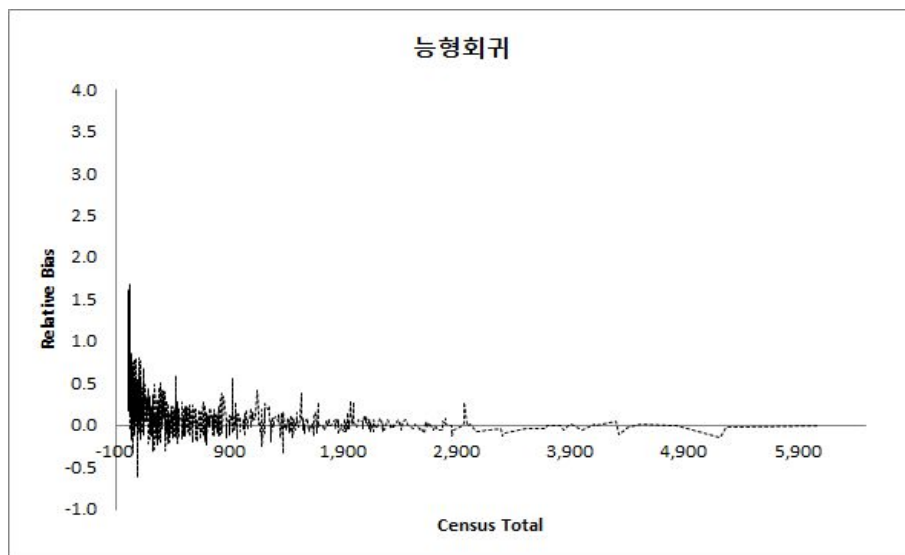
<그림 3-100> 부모집단별 행정자료 *Rel.Bias*와 센서스 가구수 플롯



<그림 3-101> 부모집단별 조사자료 $Rel.Bias$ 와 센서스 가구수 플롯



<그림 3-102> 부모집단별 능형회귀추정 방안 $Rel.Bias$ 와 센서스 가구수 플롯



- 관심 부모집단별 세 가지 추정 방법의 *Rel.Bias*와 센서스 가구수의 관계를 살펴보기 위하여 센서스 가구수를 13개의 그룹으로 구성하여 그룹 내의 *Rel.Bias*의 평균과 표준편차를 세 가지 추정 방법에 대하여 계산한 결과는 다음의 <표 3-16>과 같다. 각 그룹별 *Rel.Bias*의 평균은 행정자료가 다른 방법에 비하여 대체로 크게 나타나고 있으며, ‘50가구 미만’부터 ‘100-200명 미만’의 그룹에서는 능형회귀추정 방안의 *Rel.Bias*가 0.247~0.573 사이의 값으로 크게 나타나고 있다. 따라서 인구 부분과 마찬가지로 능형회귀추정 방안을 통한 관심 부모집단의 규모가 작은 곳에서의 추정에 있어서는 주의를 필요로 한다.

<표 3-16> 센서스 가구수에 따른 *Rel.Bias*의 평균과 표준편차

센서스 가구수 (셀 수)	통계량	행정자료	능형회귀	표본자료
50가구 미만 (54)	mean	1.210	0.573	-0.012
	sd	0.568	0.431	0.326
50-100가구 미만 (63)	mean	0.707	0.306	-0.004
	sd	0.323	0.263	0.243
100-200가구 미만 (80)	mean	0.450	0.247	0.021
	sd	0.286	0.216	0.207
200-300가구 미만 (79)	mean	0.152	0.073	0.010
	sd	0.285	0.184	0.184
300-400가구 미만 (56)	mean	0.129	0.085	0.024
	sd	0.243	0.168	0.139
400-600가구 미만 (86)	mean	0.125	0.077	0.015
	sd	0.210	0.143	0.128
600-800가구 미만 (74)	mean	0.063	0.021	0.018
	sd	0.220	0.120	0.107
800-1,000가구 미만 (49)	mean	0.085	0.061	0.039
	sd	0.199	0.142	0.106
1,000-1,300가구 미만 (55)	mean	0.043	0.050	0.040
	sd	0.137	0.117	0.084
1,300-1,600가구 미만 (52)	mean	0.023	0.018	0.020
	sd	0.123	0.104	0.077
1,600-2,000가구 미만 (55)	mean	0.029	0.024	0.012
	sd	0.121	0.091	0.095
2,000-2,500가구 미만 (38)	mean	0.016	0.012	0.003
	sd	0.075	0.051	0.065
2,500가구 이상 (49)	mean	-0.015	-0.020	0.008
	sd	0.080	0.064	0.073
총합계 (790)	mean	0.240	0.122	0.015
	sd	0.418	0.242	0.163

3. 주택

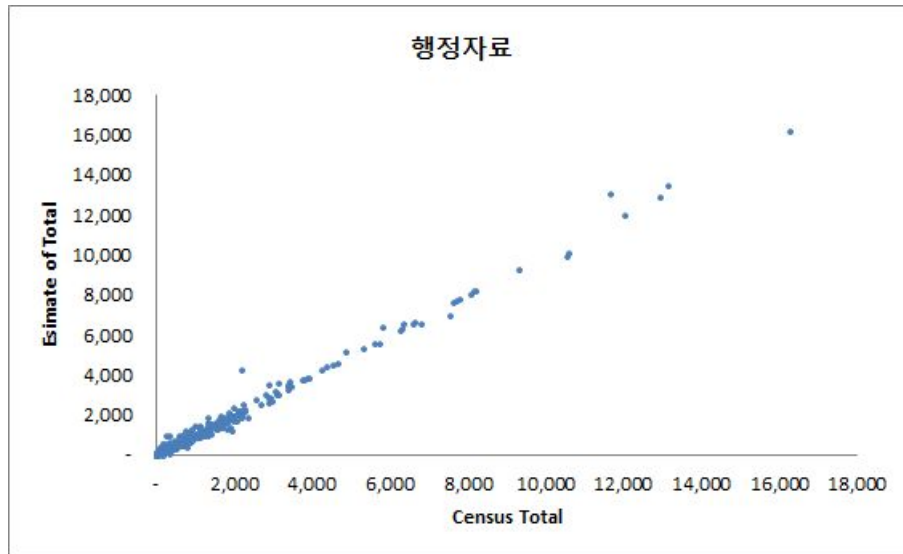
- 12개 시범지역의 최종데이터를 사용하여 주택 가중치를 능형회귀추정 방안을 통하여 보정한 능형회귀 가중치는 다음의 <표 3-17>과 같다. 제2절의 <표 3-5>의 중간데이터로부터 얻어진 분포와 비교하면 인구, 가구 부분과 마찬가지로 양 끝 쪽의 5%에 해당하는 분포가 더 안정적임을 알 수 있다.

<표 3-17> 12개 시범지역의 가구 능형회귀 가중치의 분포 (최종데이터)

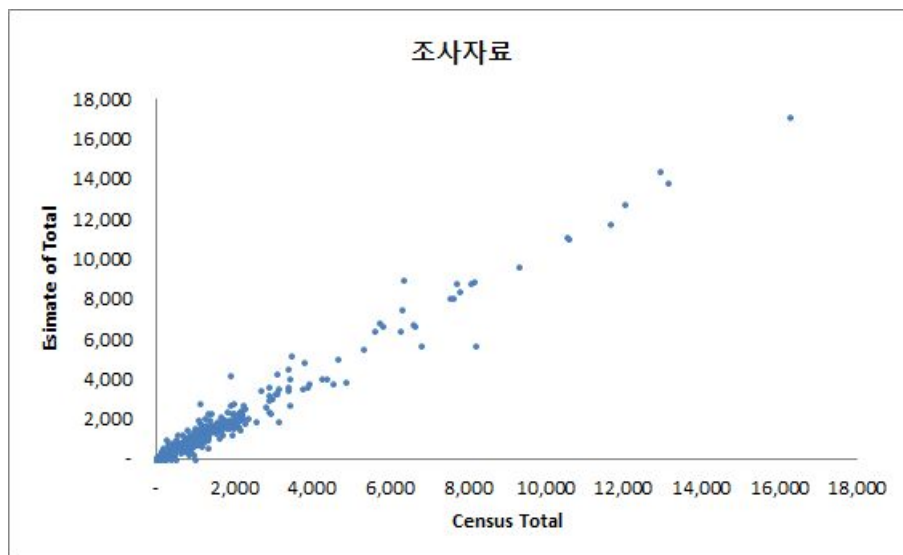
	mean	std.	min.	5%	Q1	median	Q3	95%	max.
능형회귀가중치	0.990	0.074	0.050	0.939	1.000	1.000	1.000	1.020	1.208

- 관심 부모집단인 읍면동별(158)*주택유형별(3) 참 값인 센서스 주택수를 x 축으로 행정자료, 조사자료 그리고 능형회귀추정 방안으로 산출된 주택수를 y 축으로 한 산점도는 다음의 <그림 3-103>부터 <그림 3-105>이다. 그림을 전체적으로 살펴보면, 행정자료와 조사자료에 비하여 능형회귀추정 방안의 결과가 조금 더 직선상에 가깝게 나타나고 있음을 확인할 수 있다.

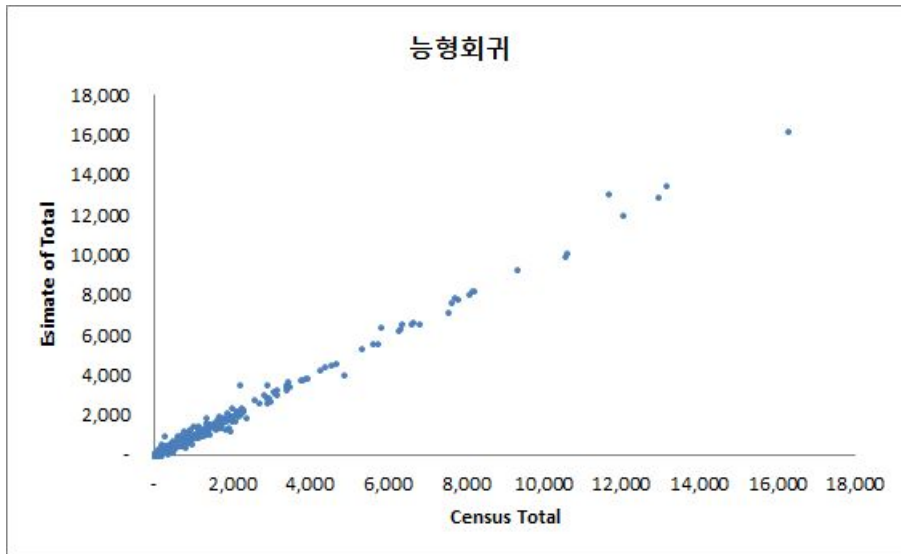
<그림 3-103> 행정자료와 센서스 부모집단별 주택수 산점도



<그림 3-104> 조사자료와 센서스 부모집단별 주택수 산점도



<그림 3-104> 능형회귀추정 방안과 센서스 부모집단별 주택수 산점도



- 행정자료, 조사자료 그리고 능형회귀추정 방안을 하여 관심 부모집단 읍면동 별(158)*주택유형별(3) 지역과 주택유형 별 주변분포를 구성하여 그 *Rel.Bias*를 계산하였다. 여기서 지역별 주변분포는 인구, 가구 부분과 마찬가지로 시군구와 읍면동 수준에서 살펴보았다. 주택유형별, 시군구별 그리고 읍면동별 주변분포의 *Rel.Bias*는 각각 다음의 <표 3-18>, <표 3-19> 그리고 <표 3-20>과 같다. <표 3-18>의 주택유형 수준의 분포에서는 ‘기타’부분을 제외한 나머지 수준에서 *Rel.Bias*가 작게 나타나고 있다. <표 3-19>의 시군구별 수준의 분포에서는 모든 시군구에서 *Rel.Bias*가 작게 나타나고 있으며, <표 3-20>의 읍면동 수준의 분포에서는 일부 읍면동을 제외한 대부분의 읍면동에서 *Rel.Bias*가 작게 나타나고 있음을 확인할 수 있다.

<표 3-18> 주택유형별(3) 주변분포에서의 주택수 *Rel.Bias*의 비교

주택유형	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
단독주택	172,707	-0.043	-0.039	0.021
아파트	323,846	-0.003	-0.006	0.076
기타	103,938	0.154	0.100	0.030

<표 3-19> 시군구별(12) 주변분포에서의 주택수 *Rel.Bias*의 비교

시군구	센서스총합	<i>Rel.Bias</i>		
		등록자료	능형회귀	조사자료
11180	52,320	-0.035	-0.033	0.018
11190	100,330	-0.026	-0.026	0.043
23080	128,715	0.051	0.031	0.043
24010	29,302	0.045	0.001	0.082
31240	153,789	0.013	0.009	0.062
32380	7,289	-0.069	-0.047	0.108
33390	11,840	0.003	0.007	0.010
35350	12,439	0.081	0.052	0.035
36040	34,807	0.049	0.026	0.072
36480	17,762	-0.018	-0.020	0.047
37030	48,542	0.002	-0.006	0.073
37430	3,356	0.017	-0.006	0.045

<표 3-20> 읍면동별(158) 주변분포에서의 주택수 *Rel.Bias*의 비교

읍면동	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
1118051	4,736	-0.128	-0.126	-0.057
1118052	8,711	-0.052	-0.052	0.044
1118053	3,510	0.014	0.020	0.012
1118054	5,289	-0.032	-0.032	0.041
1118055	3,082	0.008	0.008	0.074
1118057	6,616	-0.024	-0.024	-0.030
1118058	7,804	-0.071	-0.052	0.069
1118059	3,622	0.004	0.004	0.098
1118060	4,711	0.007	0.007	-0.056
1118061	4,239	-0.001	-0.025	-0.020
1119054	11,412	-0.048	-0.048	0.019
1119055	5,740	-0.036	-0.036	0.170
1119056	10,647	-0.071	-0.071	0.005
1119061	5,330	-0.040	-0.040	-0.008
1119062	6,119	0.008	0.008	0.032
1119063	4,239	-0.068	-0.068	0.117
1119065	3,298	0.017	0.015	0.175
1119066	2,724	-0.076	-0.076	0.096
1119067	2,993	-0.006	-0.006	0.151
1119068	5,597	0.007	0.007	0.036
1119069	4,423	-0.008	-0.008	-0.035
1119070	2,780	0.008	0.008	0.038
1119071	3,475	-0.016	0.019	-0.159
1119072	6,325	0.010	0.017	0.045
1119073	6,116	0.005	-0.016	0.106
1119074	4,584	-0.011	-0.011	0.034
1119075	4,320	-0.005	-0.005	0.073
1119076	10,208	-0.047	-0.047	0.014
2308051	8,942	0.067	0.059	0.065
2308053	14,880	0.009	0.009	0.005
2308054	3,106	0.983	0.750	0.022
2308055	5,965	0.171	-0.029	-0.360

<표 3-20> 읍면동별(158) 주변분포에서의 주택수 $Rel.Bias$ 의 비교-(계속)

읍면동	센서스총합	$Rel.Bias$		
		행정자료	능형회귀	조사자료
2308056	4,355	0.004	0.004	0.008
2308058	7,543	0.005	0.005	0.043
2308059	4,786	0.024	0.024	0.061
2308060	6,117	0.000	0.001	0.034
2308062	4,790	0.061	-0.004	0.119
2308063	7,200	0.004	0.004	0.094
2308064	6,688	0.022	0.022	0.078
2308065	3,561	-0.012	-0.012	-0.068
2308068	14,399	0.055	0.055	0.143
2308069	7,243	0.005	0.005	0.075
2308070	7,287	0.072	0.072	0.181
2308071	11,114	-0.047	-0.053	0.035
2308072	2,232	0.149	0.047	-0.097
2308073	8,507	0.004	0.004	0.063
2401051	1,082	0.473	-0.026	-0.095
2401054	1,551	0.026	0.028	0.183
2401056	1,793	0.048	-0.011	-0.025
2401058	2,919	0.008	-0.040	0.200
2401059	2,851	0.024	0.024	0.055
2401061	3,071	0.012	0.012	0.063
2401062	1,511	0.053	-0.020	0.059
2401063	2,195	-0.010	-0.034	0.054
2401064	1,062	0.249	0.249	0.271
2401068	2,020	0.053	-0.069	0.100
2401071	4,454	0.010	0.010	-0.003
2401072	2,528	0.007	-0.035	0.083
2401073	2,265	0.029	0.029	0.192
3124012	19,751	0.016	0.016	0.047
3124013	5,174	0.009	0.009	0.055
3124014	16,359	0.009	0.009	0.113
3124031	2,357	0.045	-0.099	0.069
3124033	1,866	0.058	0.048	0.198

<표 3-20> 읍면동별(158) 주변분포에서의 주택수 *Rel.Bias*의 비교-(계속)

읍면동	센서스총합	<i>Rel.Bias</i>		
		행정자료	능형회귀	조사자료
3124035	1,685	0.082	-0.017	-0.003
3124036	3,604	0.113	0.113	0.009
3124037	1,913	0.136	0.136	0.124
3124038	3,226	0.091	0.077	0.086
3124039	3,526	-0.028	-0.028	0.084
3124042	1,338	0.028	-0.129	-0.026
3124043	3,986	0.055	0.055	0.164
3124044	2,184	-0.163	-0.163	-0.218
3124051	7,465	0.034	0.034	0.066
3124052	9,273	0.018	0.018	0.083
3124053	8,226	0.005	0.023	0.077
3124054	8,126	0.004	0.004	0.091
3124055	5,575	0.012	0.012	0.084
3124056	4,350	0.008	0.008	0.000
3124057	7,322	-0.010	-0.010	0.050
3124060	9,724	0.001	0.001	0.019
3124061	14,265	0.008	0.008	0.056
3124062	12,494	-0.021	-0.021	0.046
3238011	4,061	-0.046	-0.046	0.088
3238031	1,336	-0.016	0.113	0.085
3238032	804	-0.175	-0.211	0.132
3238033	587	-0.135	-0.135	0.390
3238034	501	-0.152	-0.113	-0.033
3339014	10,792	0.003	0.012	0.024
3339031	1,048	0.006	-0.042	-0.141
3535011	2,473	0.100	0.064	0.131
3535031	638	0.118	0.118	-0.101
3535032	898	0.092	-0.052	-0.219
3535033	724	0.001	0.001	-0.085
3535034	747	0.177	0.158	0.123
3535035	2,031	0.053	0.004	0.034
3535036	648	0.068	0.068	-0.084

<표 3-20> 읍면동별(158) 주변분포에서의 주택수 $Rel.Bias$ 의 비교-(계속)

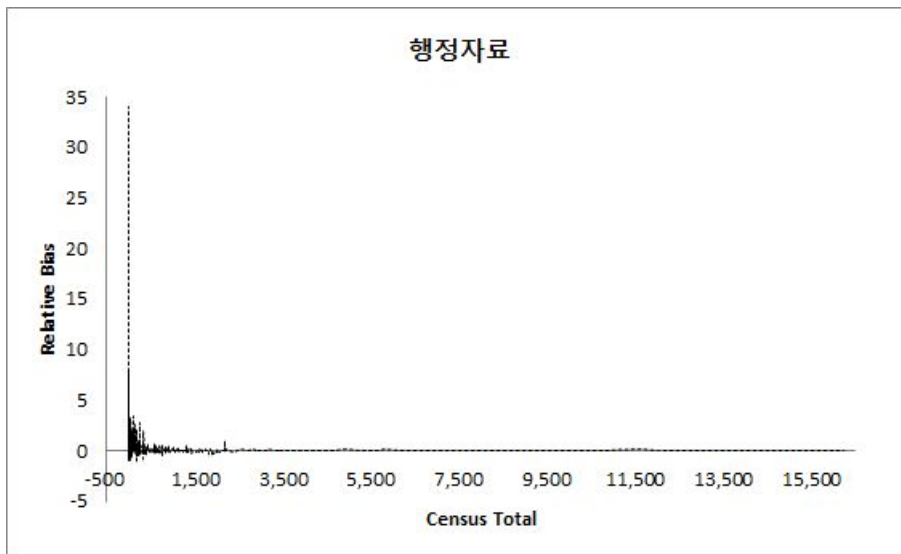
읍면동	센서스총합	$Rel.Bias$		
		행정자료	능형회귀	조사자료
3535037	886	0.076	0.076	-0.087
3535038	1,370	0.055	0.055	0.217
3535039	838	0.023	0.023	0.094
3535040	587	0.158	0.158	0.095
3535041	599	0.105	0.057	-0.042
3604011	2,824	0.038	0.023	0.073
3604031	1,356	0.033	-0.010	0.041
3604032	1,498	0.044	0.008	0.036
3604033	810	0.022	0.090	0.218
3604034	1,285	0.025	-0.025	0.121
3604035	1,525	0.005	-0.017	0.045
3604036	1,997	0.075	0.075	-0.007
3604037	1,153	0.052	0.035	0.048
3604038	2,118	0.067	0.030	0.016
3604039	1,882	0.038	0.038	0.121
3604040	1,342	0.155	0.086	0.093
3604041	977	0.005	-0.004	-0.040
3604042	2,151	0.059	0.048	0.109
3604051	2,837	0.010	0.010	0.089
3604052	1,707	0.021	0.021	-0.085
3604054	2,357	0.014	0.014	0.113
3604055	3,744	0.065	0.020	0.136
3604058	1,893	0.064	0.064	0.118
3604060	1,351	0.161	0.009	0.034
3648011	2,280	-0.096	-0.080	0.004
3648031	836	-0.038	-0.053	0.143
3648032	1,357	-0.063	-0.104	-0.023
3648033	1,103	-0.018	0.004	0.112
3648034	1,631	0.025	-0.001	0.058
3648035	1,281	0.052	0.052	0.042
3648036	1,593	-0.112	-0.045	0.069
3648037	798	0.069	0.069	-0.022

<표 3-20> 읍면동별(158) 주변분포에서의 주택수 $Rel.Bias$ 의 비교-(계속)

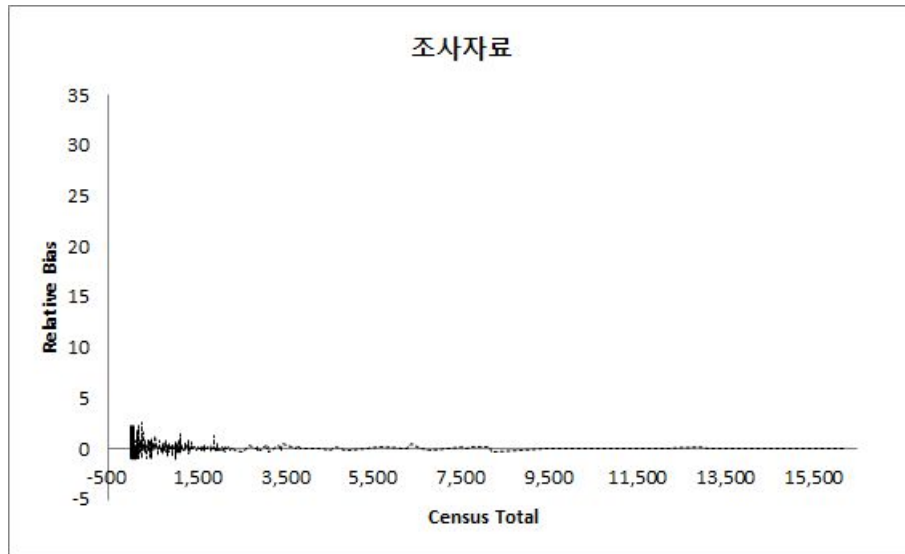
읍면동	센서스총합	$Rel.Bias$		
		행정자료	능형회귀	조사자료
3648038	652	0.006	0.006	0.214
3648039	798	-0.005	0.010	0.188
3648040	1,527	-0.016	-0.054	-0.048
3648041	571	0.009	0.033	0.126
3648042	1,108	-0.067	-0.067	-0.014
3648043	2,227	0.064	0.037	0.042
3703011	3,254	0.005	0.005	0.133
3703031	1,174	0.022	0.022	-0.056
3703032	1,333	0.005	0.005	0.096
3703033	1,121	0.002	-0.001	0.118
3703034	1,737	0.030	0.030	0.067
3703035	1,982	-0.062	-0.062	0.129
3703036	1,348	-0.034	-0.064	0.014
3703037	1,452	-0.008	-0.008	0.059
3703038	842	0.044	-0.023	-0.034
3703039	1,054	-0.061	-0.076	-0.032
3703040	1,689	0.030	0.030	-0.043
3703041	971	0.011	0.011	0.221
3703042	691	0.091	0.080	0.192
3703043	1,176	0.050	0.034	-0.097
3703044	601	0.038	0.137	0.198
3703055	1,923	-0.083	-0.102	0.158
3703056	8,142	0.013	-0.011	0.028
3703058	7,837	0.004	0.004	0.113
3703059	3,797	-0.039	-0.017	0.038
3703060	3,014	0.006	-0.063	0.101
3703061	3,404	0.049	0.049	0.100
3743011	2,073	-0.041	-0.079	0.053
3743031	688	0.115	0.115	0.075
3743032	595	0.109	0.109	-0.019

- *Rel.Bias*와 주택수와의 관계를 살펴보기 위하여 읍면동별(158)*주택유형별(3)로 구성된 관심 부모집단의 통계를 행정자료, 조사자료 그리고 능형회귀추정 방안으로 산출하여 *Rel.Bias*를 계산하였다. 다음의 <그림 3-106>부터 <그림 3-108>은 참 값인 센서스 가구수를 x 축으로 행정자료, 조사자료 그리고 능형회귀추정 방안의 인구수 *Rel.Bias*를 y 축으로 나타낸 플롯이다. 인구, 가구 부분과 마찬가지로 센서스총합이 작은 곳에서 *Rel.Bias*의 변동이 매우 큼을 확인할 수 있다.

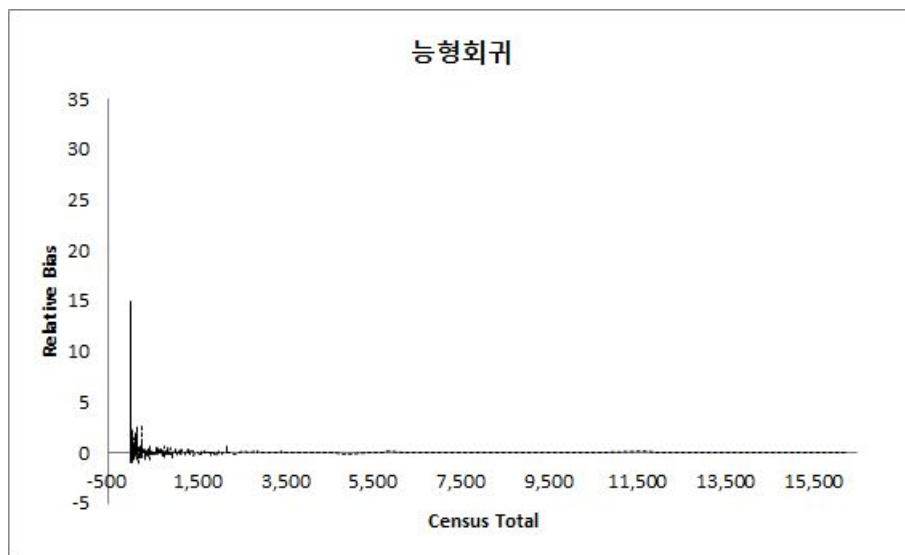
<그림 3-106> 부모집단별 행정자료 *Rel.Bias*와 센서스 주택수 플롯



<그림 3-107> 부모집단별 조사자료 *Rel.Bias*와 센서스 주택수 플롯



<그림 3-106> 부모집단별 능형회귀추정 방안 *Rel.Bias*와 센서스 주택수 플롯



- 관심 부모집단별 세 가지 추정 방법의 *Rel.Bias*와 센서스 주택수의 관계를 살펴보기 위하여 센서스 주택수를 9개의 그룹으로 구성하여 그룹 내의 *Rel.Bias*의 평균과 표준편차를 세 가지 추정 방법에 대하여 계산한 결과는 다음의 <표 3-21>과 같다. 각 그룹별 *Rel.Bias*의 평균은 행정자료가 다른 방법에 비하여 대체로 크게 나타나고 있으며, ‘50주택 미만’부터 ‘100-300주택 미만’의 그룹에서는 능형회귀추정 방안의 *Rel.Bias*가 0.265~1.356 사이의 값으로 크게 나타나고 있다. 따라서 인구, 가구 부분과 마찬가지로 능형회귀추정 방안을 통한 관심 부모집단의 규모가 작은 곳에서의 추정에 있어서는 주의를 필요로 한다.

<표 3-21> 센서스 주택수에 따른 *Rel.Bias*의 평균과 표준편차

센서스 주택수 (셀 수)	통계량	행정자료	능형회귀	조사자료
50주택 미만 (56)	mean	2.783	1.356	-0.109
	sd	5.035	2.658	0.859
50-100주택 미만 (24)	mean	0.637	0.265	0.053
	sd	0.981	0.739	0.854
100-300주택 미만 (45)	mean	0.487	0.306	0.225
	sd	0.787	0.716	0.857
300-600주택 미만 (45)	mean	0.070	-0.007	0.124
	sd	0.398	0.274	0.487
600-1,000주택 미만 (66)	mean	0.018	0.009	0.021
	sd	0.199	0.206	0.330
1,000-1,500주택 미만 (68)	mean	-0.008	-0.004	0.104
	sd	0.118	0.121	0.314
1,500-2,000주택 미만 (47)	mean	-0.042	-0.037	0.044
	sd	0.103	0.102	0.243
2,000-5,000주택 미만 (48)	mean	0.017	0.004	0.007
	sd	0.155	0.112	0.186
5,000주택 이상 (25)	mean	0.002	0.003	0.064
	sd	0.040	0.039	0.125
총합계 (424)	mean	0.462	0.223	0.055
	sd	2.074	1.109	0.537

4. 토의

- 최종데이터를 활용하여 산출된 인구수 추정량들을 비교한 결과 네 가지 능형회귀 추정량의 값들은 대략적으로 행정자료와 조사자료 기반 추정량의 사이에 있음을 알 수 있다.
- 인구수 산출을 위해 고려한 네 가지 능형회귀 추정방안들의 차이는 추정 셀의 정의에 따른 차이이며 네 가지 가능한 방안 중 방안1 혹은 방안4가 적절한 것으로 판단된다.
- 인구수 통계 공표 기준을 결정하기 위하여 살펴 본 여러 단위의 분석 결과 읍면동별(158)*성별(2)*연령5세별(21) 통계는 전체적으로 행정자료는 20%, 조사자료는 5% 그리고 능형회귀 방안1과 방안4는 각각 9%와 7% 수준의 상대편향을 나타내고 있고 표준편차 측면에서는 능형회귀방안이 가장 효율적으로 나타났다. 그러나 특정 셀 내의 추정 인구수가 매우 적은 경우(40명 미만) 상대편향이 크게 나타나는 셀이 많고 따라서 이러한 셀에 대한 통계의 공표에는 주의가 필요하다.
- 시군구별, 성별, 연령별 통계 산출에 있어서도 능형회귀추정 방안의 적용이 적절한 것으로 판단되나 언급한 바와 마찬가지로 인구수가 적은 읍면동 혹은 고연령대의 인구수 산출에는 주의가 필요하다.
- 가구 및 주택수 산출에 있어서도 능형회귀추정 방안이 편향과 표준오차 측면에서 적절한 통계적 정도를 나타내고 있으나 인구 추정에서와 마찬가지로 공표단위 셀에 대응되는 가구 혹은 주택수가 적은 경우 주의가 필요하다.
- 능형회귀 추정 방안의 효율성은 추정 셀의 정의와 더불어 능형회귀 계수를 결정하기 위하여 사용된 z 값에 의존한다. 조사자료의 통계적 효율성이 높은 공표단위 셀의 경우 신뢰계수의 값을 낮추어 행정자료 보정의 폭을 넓히는 방안을 고려할 수 있을 것이다.

[부록 3-1] 인구, 가구, 주택 부문 주변분포

<부록 표 3-1> 성별 인구수 주변분포

성별	센서스	행정자료	능형1	능형2	능형3	조사자료
남	949,345	991,180	959,686	962,277	919,498	881,006
여	942,178	962,772	946,635	950,348	915,840	882,673

<부록 표 3-2> 연령별 인구수 주변분포

연령	센서스	행정자료	능형1	능형2	능형3	조사자료
0-4	102,512	117,637	114,828	116,010	105,739	95,815
5-9	101,231	106,557	104,127	105,146	95,853	94,641
10-14	118,871	119,307	117,504	117,533	117,368	112,187
15-19	119,295	116,566	114,422	113,882	114,443	108,062
20-24	107,991	109,556	102,134	106,636	99,821	94,579
25-29	141,665	154,041	144,919	150,510	141,321	133,007
30-34	162,706	167,817	163,895	164,572	156,740	152,828
35-39	174,695	185,814	181,403	182,184	173,476	163,468
40-44	161,547	169,055	165,177	165,175	154,841	150,320
45-49	151,949	163,194	158,687	158,711	148,730	141,712
50-54	140,624	146,279	144,581	141,743	138,912	131,555
55-59	104,346	104,181	102,833	100,594	98,714	96,636
60-64	85,945	86,579	85,763	83,188	83,717	82,120
65-69	72,098	75,595	74,739	75,332	72,943	67,933
70-74	63,165	60,662	60,327	60,511	60,662	59,490
75-79	43,677	38,188	38,028	38,100	38,188	41,848
80-84	24,254	20,468	20,511	20,391	21,091	22,977
85-89	10,824	9,118	9,140	9,082	9,376	10,360
90-94	3,403	2,710	2,683	2,699	2,765	3,363
95-99	642	521	515	521	532	687
100이상	83	107	107	106	107	91

<부록 표 3-3> 시군구별 인구수 주변분포

시도	시군구	센서스	행정자료	능형1	능형2	능형3	조사자료
11	180	230,447	236,176	234,320	233,539	226,647	219,993
11	190	374,543	384,586	380,751	382,564	368,959	354,566
23	080	389,057	392,915	385,448	388,270	376,222	350,797
24	010	103,998	98,021	97,798	97,926	94,421	101,072
31	240	474,160	486,909	478,052	479,554	448,588	435,854
32	380	19,252	21,032	18,512	18,422	18,564	16,085
33	390	31,204	32,631	32,077	32,206	30,068	27,916
35	350	23,490	29,234	25,434	25,542	25,360	22,364
36	040	77,825	86,306	81,165	81,494	77,813	81,554
36	480	32,987	43,728	37,029	36,679	37,300	30,682
37	030	126,823	132,187	126,492	127,122	122,122	114,884
37	430	7,737	10,227	9,243	9,307	9,275	7,910

<부록 표 3-4> 성*연령별 인구수 주변분포

성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
남	0-4	52,814	60,624	58,977	59,758	53,921	49,279
	5-9	52,252	55,231	53,842	54,514	49,177	48,192
	10-14	61,231	61,909	60,857	61,045	60,906	58,103
	15-19	63,005	61,540	60,301	59,565	60,411	57,233
	20-24	57,228	57,741	52,530	55,561	51,883	47,758
	25-29	73,046	79,989	73,586	77,332	72,282	67,494
	30-34	83,149	86,940	84,372	84,424	79,555	78,220
	35-39	90,841	98,970	96,043	96,132	90,586	84,932
	40-44	84,513	90,090	87,250	87,138	81,364	79,060
	45-49	78,647	85,476	82,240	82,209	76,721	73,034
	50-54	70,736	74,693	73,568	71,489	69,885	66,415
	55-59	51,276	51,697	50,841	49,268	48,252	46,924
	60-64	41,804	42,386	42,013	40,174	41,191	40,025
	65-69	33,442	34,575	34,213	34,472	33,527	31,676
	70-74	27,175	25,793	25,564	25,720	25,793	25,408
	75-79	16,754	14,096	13,995	14,074	14,096	16,222
	80-84	7,626	6,295	6,348	6,281	6,664	7,359
	85-89	2,946	2,451	2,466	2,439	2,587	2,828
	90-94	742	580	578	579	593	714
	95-99	110	88	87	88	90	129
	100이상	8	16	16	16	16	0
여	0-4	49,698	57,013	55,851	56,251	51,819	46,536
	5-9	48,979	51,326	50,285	50,632	46,675	46,449
	10-14	57,640	57,398	56,646	56,488	56,462	54,084
	15-19	56,290	55,026	54,121	54,317	54,031	50,829
	20-24	50,763	51,815	49,604	51,075	47,938	46,821
	25-29	68,619	74,052	71,334	73,178	69,039	65,513
	30-34	79,557	80,877	79,522	80,147	77,186	74,608
	35-39	83,854	86,844	85,360	86,052	82,890	78,537
	40-44	77,034	78,965	77,927	78,036	73,477	71,260
	45-49	73,302	77,718	76,446	76,502	72,009	68,679
	50-54	69,888	71,586	71,013	70,254	69,027	65,139
	55-59	53,070	52,484	51,992	51,326	50,462	49,711
	60-64	44,141	44,193	43,751	43,014	42,527	42,095
	65-69	38,656	41,020	40,526	40,859	39,416	36,257
	70-74	35,990	34,869	34,763	34,791	34,869	34,082
	75-79	26,923	24,092	24,033	24,027	24,092	25,625
	80-84	16,628	14,173	14,163	14,110	14,427	15,618
	85-89	7,878	6,667	6,675	6,643	6,789	7,532
	90-94	2,661	2,130	2,105	2,121	2,172	2,649
	95-99	532	433	428	433	442	558
	100이상	75	91	91	90	91	91

<부록 표 3-5> 시군구*성별 인구수 주변분포

시도	시군구	성별	센서스	행정자료	능형1	능형2	능형3	조사자료
11	180	남	116,529	120,650	119,364	118,565	114,865	111,407
		여	113,918	115,526	114,956	114,974	111,782	108,586
11	190	남	186,507	193,517	191,301	192,237	184,115	176,323
		여	188,036	191,069	189,449	190,327	184,845	178,243
23	080	남	196,367	199,598	194,458	196,394	189,573	176,814
		여	192,690	193,317	190,990	191,876	186,649	173,984
24	010	남	51,436	49,037	48,686	49,064	46,881	50,624
		여	52,562	48,984	49,112	48,862	47,540	50,449
31	240	남	242,708	250,010	244,767	244,621	227,071	220,416
		여	231,452	236,899	233,285	234,933	221,517	215,438
32	380	남	10,163	10,858	9,165	9,115	9,354	8,009
		여	9,089	10,174	9,347	9,307	9,209	8,077
33	390	남	15,678	16,778	16,463	16,374	15,324	14,097
		여	15,526	15,853	15,614	15,833	14,744	13,819
35	350	남	11,012	14,577	12,138	12,108	12,198	10,443
		여	12,478	14,657	13,296	13,434	13,162	11,922
36	040	남	37,000	43,266	39,311	39,082	37,960	38,322
		여	40,825	43,040	41,855	42,411	39,852	43,232
36	480	남	16,115	22,249	17,934	17,947	18,174	14,603
		여	16,872	21,479	19,095	18,732	19,126	16,079
37	030	남	61,842	65,347	61,294	62,009	59,258	55,934
		여	64,981	66,840	65,198	65,113	62,864	58,951
37	430	남	3,988	5,293	4,805	4,761	4,726	4,015
		여	3,749	4,934	4,438	4,546	4,549	3,895

<부록 표 3-6> 시군구*연령별 인구수 주변분포

시도	시군구	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
11	180	0-4	8,880	10,369	10,094	10,151	9,479	8,296
		5-9	9,357	9,941	9,659	9,714	9,088	8,904
		10-14	12,707	12,953	12,953	12,680	12,896	12,159
		15-19	15,220	15,347	15,347	15,172	15,280	14,957
		20-24	14,875	15,031	15,013	14,871	14,477	14,578
		25-29	20,844	21,641	21,621	21,420	20,831	21,411
		30-34	19,896	19,790	19,645	19,584	18,834	19,418
		35-39	19,818	21,187	21,028	20,964	20,167	18,807
		40-44	19,602	20,843	20,484	20,623	19,554	18,664
		45-49	20,230	21,653	21,281	21,419	20,315	19,044
		50-54	19,539	20,017	19,938	19,806	19,337	18,232
		55-59	14,828	14,507	14,450	14,357	14,016	14,041
		60-64	11,971	11,636	11,615	11,517	11,328	11,081
		65-69	9,339	9,283	9,265	9,283	9,037	8,438
		70-74	6,427	5,936	5,898	5,936	5,936	5,747
		75-79	3,723	3,337	3,316	3,337	3,337	3,439
		80-84	1,946	1,677	1,688	1,677	1,692	1,718
		85-89	855	747	751	747	753	753
		90-94	327	226	222	226	234	261
11	190	95-99	55	46	45	46	48	48
		100이상	8	9	9	9	9	0
		0-4	17,673	20,140	19,701	19,921	18,410	16,620
		5-9	15,892	16,584	16,217	16,398	15,159	14,483
		10-14	18,436	18,763	18,572	18,544	18,681	17,160
		15-19	21,293	21,329	21,106	21,233	21,235	19,205
		20-24	23,550	22,977	22,726	22,870	22,142	21,388
		25-29	34,511	35,805	35,398	35,638	34,505	32,807
		30-34	36,960	37,012	36,305	36,847	35,250	35,639
		35-39	35,541	37,414	36,666	37,236	35,607	33,942
		40-44	30,264	31,347	31,207	31,196	29,404	28,379
		45-49	28,984	31,490	31,344	31,333	29,547	27,964
		50-54	30,173	31,560	31,493	31,409	30,497	29,004
		55-59	23,455	23,470	23,419	23,358	22,680	22,481
		60-64	19,109	19,390	19,345	19,288	18,876	18,488
		65-69	14,916	15,272	15,241	15,268	14,864	14,229
		70-74	11,228	10,910	10,910	10,906	10,910	10,417
		75-79	6,839	5,987	5,987	5,985	5,987	6,920
		80-84	3,482	3,159	3,143	3,158	3,193	3,279
		85-89	1,603	1,440	1,433	1,439	1,454	1,426
		90-94	516	426	426	426	442	592
		95-99	106	82	82	82	85	132
		100이상	12	29	29	29	29	13

<부록 표 3-6> 시군구*연령별 인구수 주변분포 (계속)

시도	시군구	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
23	080	0-4	21,378	24,081	23,608	23,823	22,012	20,128
		5-9	23,504	24,882	24,373	24,596	22,744	22,051
		10-14	30,346	30,439	30,027	30,018	30,306	28,111
		15-19	30,327	29,663	29,205	29,304	29,533	26,860
		20-24	22,896	23,053	21,535	22,738	22,198	18,747
		25-29	25,968	27,578	25,708	27,201	26,567	23,016
		30-34	30,430	31,245	30,714	30,907	29,768	27,030
		35-39	38,273	39,860	39,193	39,461	37,973	35,723
		40-44	39,901	40,638	40,277	40,209	38,128	36,804
		45-49	38,154	38,073	37,695	37,603	35,715	33,683
		50-54	30,144	29,115	29,074	28,715	28,105	26,602
		55-59	18,231	17,003	16,974	16,764	16,418	15,839
		60-64	12,664	12,228	12,242	12,061	11,907	11,354
		65-69	9,337	9,472	9,479	9,402	9,217	8,607
		70-74	7,540	7,061	6,936	7,019	7,061	6,818
		75-79	5,131	4,497	4,420	4,462	4,497	4,670
		80-84	2,957	2,529	2,514	2,505	2,551	2,907
		85-89	1,363	1,098	1,092	1,086	1,108	1,357
		90-94	418	328	314	324	340	381
24	010	95-99	83	58	56	58	60	110
		100이상	12	14	14	14	14	0
		0-4	2,880	3,325	3,185	3,284	3,039	2,693
		5-9	3,525	3,862	3,701	3,820	3,529	3,016
		10-14	5,232	5,224	5,224	5,167	5,201	5,384
		15-19	7,528	6,203	6,203	6,214	6,176	7,554
		20-24	10,490	7,041	7,206	7,054	6,782	10,313
		25-29	9,517	8,440	8,636	8,455	8,126	9,468
		30-34	6,975	6,936	6,951	6,945	6,599	6,520
		35-39	6,624	7,006	7,018	7,014	6,666	6,411
		40-44	6,470	6,810	6,729	6,821	6,387	6,045
		45-49	7,081	7,710	7,614	7,716	7,236	6,626
		50-54	8,187	8,128	8,091	8,134	7,854	8,274
		55-59	7,126	6,714	6,675	6,715	6,490	6,646
		60-64	6,507	6,328	6,266	6,330	6,157	6,410
		65-69	5,572	5,478	5,422	5,456	5,331	5,413
		70-74	4,619	4,080	4,116	4,074	4,080	4,379
		75-79	3,048	2,492	2,515	2,489	2,492	3,085
		80-84	1,632	1,394	1,403	1,391	1,411	1,750
		85-89	723	629	624	627	635	755
		90-94	212	174	172	174	180	271
		95-99	44	43	42	43	45	52
		100이상	6	4	4	4	4	6

<부록 표 3-6> 시군구*연령별 인구수 주변분포 (계속)

시도	시군구	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
31	240	0-4	39,708	44,851	44,209	44,598	39,964	36,944
		5-9	35,725	36,651	36,127	36,427	32,639	33,058
		10-14	34,323	33,258	33,044	32,990	32,549	31,958
		15-19	25,958	24,391	24,153	23,967	23,671	22,084
		20-24	21,959	22,095	20,982	21,697	19,008	18,064
		25-29	35,908	38,929	37,040	38,295	34,166	32,981
		30-34	51,950	52,578	52,185	51,848	48,649	49,505
		35-39	55,213	56,853	56,425	56,092	52,564	50,358
		40-44	43,914	43,581	42,904	42,857	39,297	40,217
		45-49	33,598	35,358	34,600	34,616	31,463	31,143
		50-54	26,820	28,189	27,839	27,513	26,443	24,868
		55-59	18,366	19,002	18,748	18,506	17,782	17,051
		60-64	14,451	15,196	14,832	14,774	14,680	13,847
		65-69	11,704	12,604	12,266	12,400	12,162	10,896
		70-74	10,489	10,575	10,225	10,403	10,575	9,974
		75-79	7,384	6,873	6,656	6,754	6,873	6,790
		80-84	4,009	3,612	3,537	3,545	3,730	3,438
		85-89	1,919	1,672	1,639	1,642	1,723	1,919
		90-94	614	524	524	515	530	660
		95-99	136	105	105	104	106	90
		100이상	12	12	12	12	12	10
32	380	0-4	987	1,152	1,062	1,061	966	892
		5-9	1,048	1,145	1,066	1,065	962	909
		10-14	1,156	1,193	1,019	1,091	1,125	998
		15-19	984	1,189	1,021	1,009	1,115	693
		20-24	1,113	1,436	821	1,174	1,046	452
		25-29	1,330	1,497	901	1,242	1,114	875
		30-34	1,249	1,312	1,216	1,099	1,133	1,104
		35-39	1,275	1,488	1,390	1,252	1,295	1,028
		40-44	1,319	1,518	1,345	1,283	1,280	1,129
		45-49	1,522	1,776	1,577	1,508	1,494	1,374
		50-54	1,616	1,716	1,688	1,451	1,576	1,620
		55-59	1,149	1,091	1,083	932	1,001	1,078
		60-64	862	985	926	843	948	782
		65-69	1,111	1,265	1,185	1,215	1,215	956
		70-74	1,102	1,022	1,001	992	1,022	896
		75-79	734	620	613	604	620	689
		80-84	423	367	346	350	383	368
		85-89	194	187	181	181	194	173
		90-94	70	58	57	56	58	63
		95-99	6	10	9	10	10	0
		100이상	2	5	5	5	5	4

<부록 표 3-6> 시군구*연령별 인구수 주변분포 (계속)

시도	시군구	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
33	390	0-4	1,781	2,057	1,990	2,052	1,769	1,261
		5-9	1,706	1,774	1,716	1,767	1,525	1,476
		10-14	2,121	2,146	2,133	2,138	2,112	1,885
		15-19	2,109	1,963	1,947	1,930	1,929	1,959
		20-24	1,739	2,044	1,927	2,005	1,674	1,483
		25-29	2,256	2,550	2,405	2,504	2,088	1,780
		30-34	2,335	2,458	2,401	2,417	2,273	1,847
		35-39	2,559	2,740	2,670	2,691	2,525	2,170
		40-44	2,451	2,760	2,755	2,711	2,479	2,193
		45-49	2,375	2,599	2,593	2,553	2,333	2,238
		50-54	2,252	2,360	2,360	2,318	2,275	2,041
		55-59	1,797	1,750	1,750	1,717	1,683	1,798
		60-64	1,333	1,351	1,351	1,326	1,336	1,297
		65-69	1,292	1,377	1,377	1,377	1,360	1,346
		70-74	1,280	1,250	1,250	1,250	1,250	1,337
		75-79	985	856	856	856	856	1,029
		80-84	561	394	394	394	399	561
		85-89	195	151	151	151	153	148
		90-94	65	44	44	44	44	45
		95-99	12	7	7	7	7	28
35	350	0-4	707	1,105	938	1,021	905	582
		5-9	815	986	818	895	805	705
		10-14	1,067	1,187	1,104	1,101	1,062	1,064
		15-19	841	1,205	1,114	979	1,070	998
		20-24	604	1,452	749	1,176	960	508
		25-29	652	1,567	805	1,267	1,048	534
		30-34	770	1,429	1,032	1,138	1,135	688
		35-39	958	1,572	1,137	1,256	1,236	846
		40-44	1,218	1,855	1,614	1,491	1,454	1,244
		45-49	1,486	2,220	1,952	1,804	1,735	1,527
		50-54	1,746	2,200	1,977	1,782	1,922	1,662
		55-59	1,676	1,994	1,804	1,621	1,742	1,569
		60-64	2,091	2,104	2,072	1,734	1,980	2,017
		65-69	2,176	2,441	2,397	2,415	2,299	2,076
		70-74	2,666	2,599	2,610	2,576	2,599	2,528
		75-79	2,103	1,731	1,738	1,715	1,731	2,025
		80-84	1,253	1,005	998	996	1,070	1,171
		85-89	474	412	404	408	437	437
		90-94	156	144	144	142	144	157
		95-99	24	22	22	22	22	20
		100이상	7	4	4	4	4	8

<부록 표 3-6> 시군구*연령별 인구수 주변분포 (계속)

시도	시군구	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
36	040	0-4	2,638	3,062	3,093	3,073	2,650	2,888
		5-9	3,170	3,379	3,502	3,460	2,948	3,169
		10-14	4,224	4,245	4,129	4,398	4,008	4,155
		15-19	5,218	5,020	4,813	4,655	4,727	5,754
		20-24	3,525	4,545	3,439	4,127	3,593	4,184
		25-29	3,276	4,880	3,653	4,404	3,867	3,309
		30-34	3,554	4,517	4,085	4,127	3,876	3,584
		35-39	4,223	5,240	4,726	4,773	4,512	4,120
		40-44	4,649	5,944	5,377	5,400	5,024	4,689
		45-49	5,258	6,875	6,160	6,189	5,745	5,509
		50-54	5,897	7,175	6,780	6,412	6,472	6,026
		55-59	5,187	5,639	5,331	5,007	5,074	5,094
		60-64	5,442	5,882	5,866	5,213	5,582	5,779
		65-69	5,742	6,404	6,391	6,508	6,063	6,085
		70-74	6,106	5,806	5,905	5,909	5,806	6,847
		75-79	4,951	4,085	4,159	4,166	4,085	5,301
		80-84	2,997	2,316	2,419	2,358	2,439	3,248
		85-89	1,315	951	988	968	998	1,350
		90-94	379	273	280	280	276	340
		95-99	59	54	54	55	55	89
		100이상	15	14	14	14	14	34
36	480	0-4	656	1,250	1,043	1,065	1,007	643
		5-9	877	1,236	1,034	1,056	996	846
		10-14	1,218	1,685	1,340	1,461	1,467	1,170
		15-19	1,016	1,907	1,477	1,454	1,666	832
		20-24	687	2,140	1,002	1,627	1,364	438
		25-29	794	2,358	1,112	1,794	1,509	571
		30-34	977	1,893	1,367	1,443	1,424	784
		35-39	1,369	2,309	1,660	1,763	1,739	1,329
		40-44	1,785	2,695	2,197	2,068	2,042	1,547
		45-49	2,422	3,515	2,896	2,719	2,685	2,260
		50-54	2,915	3,759	3,480	2,886	3,235	2,541
		55-59	2,799	3,248	3,004	2,510	2,804	2,613
		60-64	3,241	3,587	3,455	2,767	3,359	2,998
		65-69	3,565	4,086	3,932	4,054	3,835	3,515
		70-74	3,755	3,800	3,754	3,776	3,800	3,522
		75-79	2,697	2,437	2,408	2,422	2,437	2,576
		80-84	1,405	1,106	1,134	1,098	1,180	1,602
		85-89	594	529	548	527	562	665
		90-94	176	152	149	152	152	168
		95-99	36	31	30	31	31	48
		100이상	3	5	5	5	5	14

<부록 표 3-6> 시군구*연령별 인구수 주변분포 (계속)

시도	시군구	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
37	030	0-4	4,969	5,896	5,568	5,616	5,241	4,592
		5-9	5,323	5,790	5,600	5,625	5,182	5,748
		10-14	7,668	7,811	7,568	7,547	7,575	7,797
		15-19	8,522	7,928	7,628	7,597	7,636	6,847
		20-24	6,338	7,206	6,412	6,827	6,163	4,250
		25-29	6,282	8,107	7,240	7,687	6,967	5,937
		30-34	7,173	7,998	7,433	7,650	7,227	6,343
		35-39	8,315	9,450	8,889	9,078	8,574	8,198
		40-44	9,353	10,243	9,566	9,792	9,088	8,815
		45-49	10,163	10,969	10,134	10,414	9,633	9,660
		50-54	10,510	11,024	10,843	10,406	10,221	9,750
		55-59	9,053	8,900	8,748	8,346	8,212	7,736
		60-64	7,659	7,178	7,080	6,706	6,865	7,414
		65-69	6,806	7,292	7,162	7,332	6,955	5,774
		70-74	7,464	7,096	7,194	7,143	7,096	6,502
		75-79	5,781	4,947	5,034	4,984	4,947	5,021
		80-84	3,392	2,715	2,740	2,727	2,841	2,729
		85-89	1,517	1,222	1,248	1,227	1,277	1,299
		90-94	451	343	334	343	347	401
		95-99	78	61	61	63	62	72
		100이상	6	11	11	11	11	0
37	430	0-4	255	349	336	345	296	275
		5-9	289	327	312	322	276	277
		10-14	373	403	390	399	387	346
		15-19	279	421	407	367	404	318
		20-24	215	536	322	470	413	175
		25-29	327	689	400	603	534	319
		30-34	437	649	561	567	573	369
		35-39	527	695	602	605	619	539
		40-44	621	821	724	724	702	595
		45-49	676	956	841	837	828	685
		50-54	825	1,036	1,018	910	974	934
		55-59	679	863	848	760	811	689
		60-64	615	714	714	630	697	654
		65-69	538	621	621	621	604	599
		70-74	489	527	527	527	527	522
		75-79	301	326	326	326	326	304
		80-84	197	194	194	194	201	208
		85-89	72	80	80	80	82	80
		90-94	19	18	18	18	18	23
		95-99	3	2	2	2	2	0

<부록 표 3-7> 시군구*성*연령별 인구수 분포

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
11	180	남	0-4	4,524	5,297	5,225	5,222	4,805	4,312
			5-9	4,812	5,078	5,005	5,001	4,606	4,568
			10-14	6,609	6,824	6,824	6,736	6,793	6,609
			15-19	7,975	8,066	8,066	7,897	8,030	7,969
			20-24	7,738	7,810	7,760	7,656	7,403	7,344
			25-29	10,940	11,644	11,569	11,431	11,037	11,187
			30-34	10,744	10,650	10,526	10,453	10,002	10,398
			35-39	10,415	11,260	11,125	11,046	10,575	9,714
			40-44	9,967	10,786	10,466	10,574	10,053	9,492
			45-49	10,103	11,076	10,745	10,851	10,323	9,423
			50-54	9,672	9,979	9,936	9,777	9,534	9,117
			55-59	7,199	7,184	7,153	7,040	6,864	6,741
			60-64	5,860	5,723	5,715	5,609	5,624	5,537
			65-69	4,630	4,580	4,574	4,580	4,501	4,207
			70-74	3,004	2,716	2,700	2,716	2,716	2,635
			75-79	1,517	1,315	1,308	1,315	1,315	1,403
			80-84	551	433	439	433	448	522
			85-89	205	181	183	181	187	170
			90-94	55	37	35	37	38	53
			95-99	8	9	8	9	9	6
			100이상	1	2	2	2	2	0
		여	0-4	4,356	5,072	4,869	4,929	4,674	3,984
			5-9	4,545	4,863	4,654	4,713	4,482	4,335
			10-14	6,098	6,129	6,129	5,945	6,103	5,550
			15-19	7,245	7,281	7,281	7,275	7,250	6,988
			20-24	7,137	7,221	7,253	7,215	7,074	7,233
			25-29	9,904	9,997	10,052	9,989	9,794	10,224
			30-34	9,152	9,140	9,119	9,132	8,831	9,019
			35-39	9,403	9,927	9,903	9,918	9,592	9,093
			40-44	9,635	10,057	10,018	10,048	9,501	9,171
			45-49	10,127	10,577	10,535	10,568	9,992	9,621
			50-54	9,867	10,038	10,002	10,029	9,804	9,116
			55-59	7,629	7,323	7,297	7,317	7,152	7,300
			60-64	6,111	5,913	5,900	5,907	5,704	5,544
			65-69	4,709	4,703	4,691	4,703	4,537	4,230
			70-74	3,423	3,220	3,198	3,220	3,220	3,113
			75-79	2,206	2,022	2,008	2,022	2,022	2,035
			80-84	1,395	1,244	1,249	1,244	1,244	1,196
			85-89	650	566	568	566	566	583
			90-94	272	189	186	189	196	208
			95-99	47	37	37	37	38	42
			100이상	7	7	7	7	7	0

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
11	190	남	0-4	9,106	10,372	10,144	10,195	9,408	8,691
			5-9	8,212	8,575	8,382	8,426	7,778	7,337
			10-14	9,456	9,716	9,587	9,547	9,672	8,800
			15-19	11,293	11,352	11,202	11,295	11,301	9,970
			20-24	11,814	11,563	11,413	11,502	10,960	10,999
			25-29	17,063	17,990	17,746	17,895	17,052	15,518
			30-34	18,549	18,926	18,528	18,831	17,775	18,060
			35-39	18,742	20,084	19,654	19,978	18,863	17,665
			40-44	15,710	16,511	16,438	16,424	15,388	14,841
			45-49	14,479	15,905	15,831	15,817	14,824	14,106
			50-54	14,398	15,347	15,280	15,264	14,663	13,933
			55-59	11,265	11,368	11,317	11,308	10,861	10,874
			60-64	9,358	9,515	9,512	9,463	9,351	9,085
			65-69	7,236	7,334	7,332	7,334	7,207	6,878
			70-74	5,231	5,061	5,061	5,061	5,061	4,982
			75-79	2,835	2,358	2,358	2,358	2,358	2,854
			80-84	1,162	1,012	995	1,012	1,046	1,184
			85-89	457	415	408	415	429	385
			90-94	117	89	89	89	93	134
			95-99	20	17	17	17	18	28
			100이상	4	7	7	7	7	0
		여	0-4	8,567	9,768	9,556	9,726	9,002	7,929
			5-9	7,680	8,009	7,835	7,973	7,381	7,146
			10-14	8,980	9,047	8,986	8,997	9,008	8,360
			15-19	10,000	9,977	9,904	9,938	9,934	9,236
			20-24	11,736	11,414	11,313	11,368	11,182	10,389
			25-29	17,448	17,815	17,652	17,743	17,453	17,288
			30-34	18,411	18,086	17,777	18,016	17,475	17,579
			35-39	16,799	17,330	17,012	17,258	16,745	16,276
			40-44	14,554	14,836	14,770	14,772	14,016	13,538
			45-49	14,505	15,585	15,513	15,516	14,723	13,858
			50-54	15,775	16,213	16,213	16,145	15,834	15,071
			55-59	12,190	12,102	12,102	12,050	11,819	11,607
			60-64	9,751	9,875	9,833	9,825	9,526	9,404
			65-69	7,680	7,938	7,909	7,934	7,657	7,351
			70-74	5,997	5,849	5,849	5,845	5,849	5,435
			75-79	4,004	3,629	3,629	3,627	3,629	4,065
			80-84	2,320	2,147	2,148	2,146	2,147	2,095
			85-89	1,146	1,025	1,026	1,024	1,025	1,040
			90-94	399	337	337	337	349	458
			95-99	86	65	65	65	67	104
			100이상	8	22	22	22	22	13

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
23	080	남	0-4	11,068	12,446	12,143	12,369	11,289	10,394
			5-9	12,115	12,892	12,577	12,806	11,694	11,114
			10-14	15,595	15,694	15,383	15,570	15,623	14,667
			15-19	15,971	15,631	15,282	15,326	15,561	14,180
			20-24	12,164	12,163	10,889	11,896	11,529	9,427
			25-29	13,114	14,167	12,602	13,845	13,428	11,641
			30-34	15,024	15,585	15,385	15,299	14,637	13,159
			35-39	19,028	20,007	19,759	19,671	18,790	18,031
			40-44	20,036	20,671	20,449	20,314	19,266	19,133
			45-49	19,860	19,913	19,666	19,517	18,559	17,526
			50-54	15,933	15,538	15,506	15,201	14,845	14,048
			55-59	9,444	8,829	8,809	8,629	8,435	8,096
			60-64	6,406	6,171	6,199	6,031	6,064	5,707
			65-69	4,464	4,401	4,420	4,407	4,325	4,139
			70-74	3,056	2,908	2,843	2,920	2,908	2,788
			75-79	1,812	1,535	1,498	1,539	1,535	1,611
			80-84	815	662	662	667	684	763
			85-89	340	291	291	291	301	269
			90-94	104	82	82	82	85	90
			95-99	17	9	9	9	9	30
			100이상	1	3	3	3	3	0
		여	0-4	10,310	11,635	11,465	11,455	10,723	9,734
			5-9	11,389	11,990	11,796	11,790	11,050	10,937
			10-14	14,751	14,745	14,644	14,448	14,682	13,443
			15-19	14,356	14,032	13,923	13,978	13,972	12,681
			20-24	10,732	10,890	10,646	10,842	10,669	9,320
			25-29	12,854	13,411	13,105	13,355	13,138	11,375
			30-34	15,406	15,660	15,329	15,609	15,131	13,871
			35-39	19,245	19,853	19,434	19,790	19,183	17,692
			40-44	19,865	19,967	19,827	19,895	18,863	17,672
			45-49	18,294	18,160	18,029	18,085	17,156	16,157
			50-54	14,211	13,577	13,568	13,514	13,260	12,554
			55-59	8,787	8,174	8,165	8,135	7,983	7,743
			60-64	6,258	6,057	6,042	6,030	5,843	5,646
			65-69	4,873	5,071	5,059	4,995	4,892	4,468
			70-74	4,484	4,153	4,093	4,099	4,153	4,030
			75-79	3,319	2,962	2,922	2,922	2,962	3,058
			80-84	2,142	1,867	1,852	1,838	1,867	2,143
			85-89	1,023	807	801	795	807	1,088
			90-94	314	246	232	242	255	291
			95-99	66	49	47	48	51	80
			100이상	11	11	11	11	11	0

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
24	010	남	0-4	1,528	1,743	1,684	1,737	1,581	1,432
			5-9	1,845	2,053	1,979	2,046	1,862	1,592
			10-14	2,679	2,673	2,673	2,664	2,661	2,814
			15-19	3,874	3,205	3,205	3,216	3,191	4,194
			20-24	5,270	3,637	3,648	3,650	3,447	5,386
			25-29	5,102	4,484	4,505	4,499	4,250	5,402
			30-34	3,754	3,803	3,800	3,812	3,572	3,455
			35-39	3,506	3,824	3,821	3,832	3,591	3,449
			40-44	3,268	3,625	3,569	3,636	3,379	3,084
			45-49	3,411	3,753	3,695	3,759	3,498	3,114
			50-54	3,885	3,962	3,925	3,968	3,785	3,933
			55-59	3,352	3,157	3,118	3,158	3,016	3,044
			60-64	3,015	2,931	2,869	2,933	2,880	3,069
			65-69	2,616	2,580	2,524	2,562	2,535	2,517
			70-74	2,171	1,884	1,900	1,877	1,884	1,971
			75-79	1,314	1,011	1,019	1,007	1,011	1,307
			80-84	592	507	537	504	524	592
			85-89	209	166	175	164	172	215
			90-94	40	33	33	33	34	40
			95-99	5	5	5	5	5	12
			100이상	0	1	1	1	1	0
		여	0-4	1,352	1,582	1,501	1,546	1,458	1,260
			5-9	1,680	1,809	1,722	1,774	1,667	1,424
			10-14	2,553	2,551	2,551	2,503	2,540	2,570
			15-19	3,654	2,998	2,998	2,998	2,985	3,361
			20-24	5,220	3,404	3,557	3,404	3,335	4,927
			25-29	4,415	3,956	4,131	3,956	3,876	4,066
			30-34	3,221	3,133	3,152	3,133	3,027	3,065
			35-39	3,118	3,182	3,197	3,182	3,075	2,962
			40-44	3,202	3,185	3,159	3,185	3,009	2,960
			45-49	3,670	3,957	3,919	3,957	3,738	3,511
			50-54	4,302	4,166	4,166	4,166	4,069	4,341
			55-59	3,774	3,557	3,557	3,557	3,474	3,603
			60-64	3,492	3,397	3,397	3,397	3,277	3,341
			65-69	2,956	2,898	2,898	2,893	2,795	2,896
			70-74	2,448	2,196	2,216	2,198	2,196	2,409
			75-79	1,734	1,481	1,496	1,483	1,481	1,779
			80-84	1,040	887	867	887	887	1,158
			85-89	514	463	449	462	463	539
			90-94	172	141	139	142	146	231
			95-99	39	38	37	38	39	40
			100이상	6	3	3	3	3	6

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
31	240	남	0-4	20,325	22,991	22,556	22,841	20,240	18,685
			5-9	18,377	18,992	18,652	18,857	16,712	16,537
			10-14	17,667	17,191	17,065	17,031	16,845	15,870
			15-19	13,794	12,896	12,763	12,569	12,524	11,799
			20-24	11,707	11,438	10,728	11,129	9,739	8,491
			25-29	18,587	19,731	18,475	19,215	17,060	16,815
			30-34	26,384	26,695	26,506	26,087	24,183	25,611
			35-39	28,889	30,560	30,343	29,914	27,715	26,403
			40-44	24,140	24,173	23,735	23,587	21,448	21,846
			45-49	18,287	19,344	18,847	18,766	16,883	16,843
			50-54	13,810	14,638	14,395	14,136	13,484	12,723
			55-59	9,144	9,427	9,256	9,073	8,654	8,345
			60-64	7,019	7,541	7,427	7,244	7,293	6,663
			65-69	5,334	5,699	5,602	5,620	5,507	5,122
			70-74	4,539	4,495	4,314	4,420	4,495	4,172
			75-79	2,845	2,545	2,447	2,504	2,545	2,770
			80-84	1,205	1,077	1,077	1,060	1,140	959
			85-89	492	445	445	437	470	620
			90-94	141	110	110	108	111	130
			95-99	22	22	22	22	22	10
		여	0-4	19,383	21,860	21,652	21,757	19,724	18,259
			5-9	17,348	17,659	17,476	17,571	15,926	16,520
			10-14	16,656	16,067	15,978	15,959	15,704	16,088
			15-19	12,164	11,495	11,390	11,398	11,147	10,285
			20-24	10,252	10,657	10,253	10,568	9,269	9,572
			25-29	17,321	19,198	18,565	19,080	17,107	16,166
			30-34	25,566	25,883	25,679	25,761	24,467	23,893
			35-39	26,324	26,293	26,081	26,177	24,849	23,955
			40-44	19,774	19,408	19,169	19,270	17,849	18,371
			45-49	15,311	16,014	15,753	15,849	14,580	14,300
			50-54	13,010	13,551	13,443	13,378	12,959	12,145
			55-59	9,222	9,575	9,491	9,433	9,128	8,706
			60-64	7,432	7,655	7,405	7,530	7,387	7,184
			65-69	6,370	6,905	6,664	6,780	6,655	5,774
			70-74	5,950	6,080	5,911	5,983	6,080	5,802
			75-79	4,539	4,328	4,210	4,250	4,328	4,020
			80-84	2,804	2,535	2,460	2,484	2,590	2,478
			85-89	1,427	1,227	1,194	1,205	1,253	1,299
			90-94	473	414	414	407	418	530
			95-99	114	83	83	83	84	80
			100이상	12	12	12	12	12	10

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
32	380	남	0-4	509	603	560	551	496	393
			5-9	555	603	565	557	496	482
			10-14	616	645	582	587	609	554
			15-19	529	588	522	473	550	406
			20-24	912	921	423	712	676	245
			25-29	847	868	419	686	653	478
			30-34	679	708	656	567	591	579
			35-39	685	825	767	664	698	571
			40-44	716	825	688	665	673	579
			45-49	795	930	779	756	754	668
			50-54	862	921	886	736	829	837
			55-59	620	576	561	468	517	622
			60-64	361	436	419	349	418	334
			65-69	484	555	535	528	531	448
			70-74	473	445	419	424	445	338
			75-79	292	205	196	198	205	262
			80-84	128	117	104	110	125	114
			85-89	76	68	64	65	71	63
			90-94	23	16	16	16	16	34
			95-99	1	3	3	3	3	0
		여	0-4	478	549	502	510	471	499
			5-9	493	542	501	508	466	427
			10-14	540	548	437	504	515	444
			15-19	455	601	499	536	566	288
			20-24	201	515	398	462	371	207
			25-29	483	629	482	556	461	397
			30-34	570	604	560	532	543	524
			35-39	590	663	623	588	597	457
			40-44	603	693	657	618	607	550
			45-49	727	846	798	752	740	706
			50-54	754	795	802	715	747	782
			55-59	529	515	522	464	484	457
			60-64	501	549	507	493	531	448
			65-69	627	710	651	687	684	507
			70-74	629	577	582	568	577	558
			75-79	442	415	417	407	415	427
			80-84	295	250	241	240	258	254
			85-89	118	119	117	115	123	110
			90-94	47	42	41	40	42	30
			95-99	5	7	6	7	7	0
			100이상	2	5	5	5	5	4

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
33	390	남	0-4	952	1,082	1,039	1,082	914	631
			5-9	849	910	874	910	768	793
			10-14	1,052	1,080	1,080	1,080	1,074	962
			15-19	1,056	1,048	1,048	1,015	1,040	1,003
			20-24	894	1,235	1,179	1,196	1,023	778
			25-29	1,294	1,427	1,358	1,381	1,179	1,029
			30-34	1,232	1,299	1,249	1,258	1,203	1,039
			35-39	1,359	1,475	1,413	1,426	1,358	1,104
			40-44	1,322	1,494	1,494	1,445	1,322	1,166
			45-49	1,265	1,381	1,381	1,335	1,221	1,177
			50-54	1,129	1,198	1,198	1,156	1,134	1,098
			55-59	902	883	883	850	832	848
			60-64	663	674	674	649	667	645
			65-69	554	601	601	601	595	624
			70-74	570	506	506	506	506	618
			75-79	336	312	312	312	312	335
			80-84	179	116	116	116	119	197
			85-89	55	45	45	45	46	34
			90-94	14	12	12	12	12	13
			95-99	1	0	7	7	7	7
		여	0-4	829	975	951	970	855	631
			5-9	857	864	842	857	757	683
			10-14	1,069	1,066	1,053	1,058	1,038	923
			15-19	1,053	915	899	915	889	956
			20-24	845	809	748	809	651	705
			25-29	962	1,123	1,047	1,123	909	751
			30-34	1,103	1,159	1,151	1,159	1,070	808
			35-39	1,200	1,265	1,257	1,265	1,167	1,066
			40-44	1,129	1,266	1,261	1,266	1,157	1,027
			45-49	1,110	1,218	1,212	1,218	1,111	1,061
			50-54	1,123	1,162	1,162	1,162	1,141	943
			55-59	895	867	867	867	851	950
			60-64	670	677	677	677	669	652
			65-69	738	776	776	776	765	722
			70-74	710	744	744	744	744	719
			75-79	649	544	544	544	544	694
			80-84	382	278	278	278	280	364
			85-89	140	106	106	106	107	114
			90-94	51	32	32	32	32	32
			95-99	11	7	7	7	7	22

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
35	350	남	0-4	353	582	496	531	465	315
			5-9	449	543	449	484	433	384
			10-14	574	650	584	595	579	573
			15-19	474	640	575	488	568	514
			20-24	393	815	408	624	557	280
			25-29	364	875	433	672	599	298
			30-34	404	862	558	651	642	385
			35-39	532	970	624	737	712	460
			40-44	678	1,088	914	828	812	713
			45-49	779	1,224	1,033	937	911	782
			50-54	890	1,180	1,045	897	1,004	895
			55-59	786	978	873	741	832	722
			60-64	940	963	946	735	899	923
			65-69	903	994	972	988	927	854
			70-74	1,073	1,064	1,061	1,058	1,064	997
			75-79	787	629	627	626	629	738
			80-84	445	361	378	359	396	426
			85-89	144	116	120	115	127	135
			90-94	38	39	39	39	39	42
			95-99	6	4	4	4	4	6
		여	0-4	354	523	443	490	441	267
			5-9	366	443	369	411	372	321
			10-14	493	537	519	507	483	491
			15-19	367	565	539	491	503	484
			20-24	211	637	342	552	403	227
			25-29	288	692	372	595	449	236
			30-34	366	567	474	486	493	303
			35-39	426	602	512	519	523	386
			40-44	540	767	700	663	642	530
			45-49	707	996	919	867	824	745
			50-54	856	1,020	932	885	918	767
			55-59	890	1,016	932	881	911	846
			60-64	1,151	1,141	1,126	999	1,082	1,094
			65-69	1,273	1,447	1,425	1,427	1,372	1,222
			70-74	1,593	1,535	1,550	1,517	1,535	1,531
			75-79	1,316	1,102	1,111	1,090	1,102	1,286
			80-84	808	644	621	637	674	745
			85-89	330	296	284	292	310	302
			90-94	118	105	105	104	105	115
			95-99	18	18	18	18	18	14
			100이상	7	4	4	4	4	8

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
36	040	남	0-4	1,385	1,636	1,636	1,628	1,390	1,545
			5-9	1,642	1,722	1,787	1,760	1,481	1,678
			10-14	2,154	2,198	2,167	2,277	2,067	2,194
			15-19	2,653	2,540	2,483	2,240	2,382	2,805
			20-24	1,966	2,483	1,778	2,151	1,984	2,135
			25-29	1,814	2,772	1,927	2,371	2,184	1,711
			30-34	1,862	2,622	2,192	2,281	2,130	1,779
			35-39	2,272	3,101	2,596	2,695	2,539	2,229
			40-44	2,484	3,444	2,973	2,974	2,813	2,426
			45-49	2,770	3,819	3,241	3,246	3,070	2,930
			50-54	2,929	3,834	3,564	3,231	3,388	3,044
			55-59	2,495	2,852	2,645	2,365	2,510	2,381
			60-64	2,537	2,799	2,778	2,307	2,648	2,640
			65-69	2,480	2,716	2,697	2,753	2,560	2,578
			70-74	2,388	2,214	2,245	2,248	2,214	2,616
			75-79	1,751	1,435	1,455	1,460	1,435	2,008
			80-84	959	745	794	756	810	1,138
			85-89	377	271	289	275	294	400
			90-94	72	58	58	59	59	70
			95-99	10	5	5	5	5	15
		여	0-4	1,253	1,426	1,457	1,445	1,260	1,343
			5-9	1,528	1,657	1,715	1,700	1,467	1,491
			10-14	2,070	2,047	1,962	2,120	1,941	1,960
			15-19	2,565	2,480	2,330	2,415	2,345	2,949
			20-24	1,559	2,062	1,660	1,976	1,609	2,050
			25-29	1,462	2,108	1,726	2,033	1,683	1,598
			30-34	1,692	1,895	1,893	1,845	1,746	1,805
			35-39	1,951	2,139	2,131	2,078	1,974	1,891
			40-44	2,165	2,500	2,404	2,426	2,211	2,263
			45-49	2,488	3,056	2,919	2,943	2,675	2,579
			50-54	2,968	3,341	3,217	3,181	3,085	2,982
			55-59	2,692	2,787	2,686	2,642	2,563	2,713
			60-64	2,905	3,083	3,087	2,906	2,934	3,139
			65-69	3,262	3,688	3,694	3,755	3,503	3,507
			70-74	3,718	3,592	3,660	3,662	3,592	4,230
			75-79	3,200	2,650	2,704	2,705	2,650	3,293
			80-84	2,038	1,571	1,624	1,601	1,629	2,110
			85-89	938	680	700	693	704	949
			90-94	307	215	222	220	217	270
			95-99	49	49	49	50	49	74
			100이상	15	14	14	14	14	34

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
36	480	남	0-4	320	636	487	541	499	329
			5-9	458	634	489	542	497	404
			10-14	610	884	693	769	765	629
			15-19	555	1,024	774	758	889	447
			20-24	532	1,262	600	934	833	308
			25-29	491	1,392	676	1,038	923	348
			30-34	603	1,231	837	924	862	464
			35-39	856	1,481	1,001	1,107	1,034	817
			40-44	1,088	1,668	1,303	1,253	1,202	913
			45-49	1,341	1,994	1,570	1,509	1,441	1,266
			50-54	1,596	2,091	1,917	1,569	1,754	1,382
			55-59	1,403	1,682	1,540	1,270	1,414	1,315
			60-64	1,625	1,810	1,729	1,356	1,680	1,478
			65-69	1,604	1,778	1,704	1,746	1,652	1,546
			70-74	1,486	1,431	1,385	1,407	1,431	1,371
			75-79	950	787	758	772	787	953
			80-84	421	333	338	325	369	438
			85-89	137	99	101	97	110	164
			90-94	31	26	26	26	26	19
			95-99	8	6	6	6	6	10
		여	0-4	336	614	556	524	508	314
			5-9	419	602	546	515	498	442
			10-14	608	801	648	692	702	541
			15-19	461	883	703	696	777	385
			20-24	155	878	402	693	530	130
			25-29	303	966	437	756	586	222
			30-34	374	662	530	519	562	320
			35-39	513	828	659	655	705	511
			40-44	697	1,027	893	816	840	633
			45-49	1,081	1,521	1,326	1,210	1,245	994
			50-54	1,319	1,668	1,563	1,317	1,481	1,160
			55-59	1,396	1,566	1,464	1,240	1,390	1,298
			60-64	1,616	1,777	1,726	1,411	1,679	1,519
			65-69	1,961	2,308	2,228	2,308	2,183	1,969
			70-74	2,269	2,369	2,369	2,369	2,369	2,151
			75-79	1,747	1,650	1,650	1,650	1,650	1,623
			80-84	984	773	796	773	811	1,164
			85-89	457	430	447	430	452	501
			90-94	145	126	123	126	126	149
			95-99	28	25	24	25	25	38
			100이상	3	5	5	5	5	14

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
37	030	남	0-4	2,620	3,055	2,832	2,880	2,684	2,425
			5-9	2,777	3,061	2,921	2,958	2,710	3,137
			10-14	4,035	4,142	4,013	3,978	4,012	4,263
			15-19	4,664	4,315	4,152	4,087	4,148	3,740
			20-24	3,692	4,105	3,471	3,842	3,491	2,235
			25-29	3,251	4,295	3,613	4,002	3,647	2,866
			30-34	3,673	4,196	3,834	3,948	3,645	3,089
			35-39	4,248	4,980	4,605	4,716	4,356	4,188
			40-44	4,762	5,322	4,808	5,016	4,607	4,546
			45-49	5,206	5,625	5,016	5,272	4,805	4,858
			50-54	5,185	5,461	5,386	5,080	4,966	4,922
			55-59	4,311	4,330	4,267	3,992	3,920	3,597
			60-64	3,713	3,458	3,379	3,180	3,310	3,615
			65-69	2,895	3,043	2,957	3,059	2,901	2,495
			70-74	2,973	2,837	2,898	2,853	2,837	2,712
			75-79	2,193	1,841	1,894	1,859	1,841	1,861
			80-84	1,089	860	836	866	926	937
			85-89	437	335	326	334	361	353
			90-94	104	75	75	75	76	88
			95-99	12	8	8	8	8	8
			100이상	2	3	3	3	3	0
		여	0-4	2,349	2,841	2,737	2,736	2,557	2,168
			5-9	2,546	2,729	2,679	2,667	2,471	2,610
			10-14	3,633	3,669	3,555	3,569	3,563	3,534
			15-19	3,858	3,613	3,475	3,511	3,488	3,107
			20-24	2,646	3,101	2,941	2,985	2,672	2,014
			25-29	3,031	3,812	3,627	3,685	3,320	3,071
			30-34	3,500	3,802	3,598	3,702	3,581	3,254
			35-39	4,067	4,470	4,284	4,362	4,218	4,010
			40-44	4,591	4,921	4,758	4,775	4,482	4,270
			45-49	4,957	5,344	5,117	5,143	4,828	4,802
			50-54	5,325	5,563	5,457	5,326	5,255	4,829
			55-59	4,742	4,570	4,481	4,354	4,293	4,139
			60-64	3,946	3,720	3,701	3,526	3,555	3,799
			65-69	3,911	4,249	4,205	4,273	4,054	3,279
			70-74	4,491	4,259	4,296	4,290	4,259	3,790
			75-79	3,588	3,106	3,140	3,125	3,106	3,160
			80-84	2,303	1,855	1,904	1,860	1,914	1,792
			85-89	1,080	887	923	893	916	946
			90-94	347	268	259	268	271	313
			95-99	66	53	53	54	54	64
			100이상	4	8	8	8	8	0

<부록 표 3-7> 시군구*성*연령별 인구수 분포 (계속)

시도	시군구	성별	연령	센서스	행정자료	능형1	능형2	능형3	조사자료
37	430	남	0-4	124	181	175	180	150	127
			5-9	161	168	162	167	139	164
			10-14	184	212	206	211	204	167
			15-19	167	235	228	202	228	207
			20-24	146	309	232	269	240	129
			25-29	179	344	262	297	270	199
			30-34	241	363	301	314	314	202
			35-39	309	403	335	346	354	302
			40-44	342	483	414	423	402	321
			45-49	351	512	436	444	432	340
			50-54	447	544	529	474	499	484
			55-59	355	431	419	374	397	339
			60-64	307	365	365	317	357	329
			65-69	242	294	294	294	286	268
			70-74	211	232	232	232	232	208
			75-79	122	123	123	123	123	121
			80-84	80	72	72	72	76	89
			85-89	17	19	19	19	20	19
			90-94	3	3	3	3	3	1
		여	0-4	131	168	161	165	146	148
			5-9	128	159	150	154	138	113
			10-14	189	191	184	188	183	179
			15-19	112	186	180	165	176	111
			20-24	69	227	90	201	173	47
			25-29	148	345	138	305	264	119
			30-34	196	286	260	253	259	167
			35-39	218	292	268	259	264	237
			40-44	279	338	310	302	300	274
			45-49	325	444	406	394	396	345
			50-54	378	492	489	436	474	450
			55-59	324	432	429	386	414	350
			60-64	308	349	349	314	341	324
			65-69	296	327	327	327	319	331
			70-74	278	295	295	295	295	314
			75-79	179	203	203	203	203	184
			80-84	117	122	122	122	125	119
			85-89	55	61	61	61	62	61
			90-94	16	15	15	15	15	22
			95-99	3	2	2	2	2	0

<부록 표 3-8> 가구원수별 가구수 분포

가구원수	센서스	행정자료	능형	조사자료
1인	163,431	199,807	181,305	167,564
2인	165,408	142,579	146,610	165,590
3인	143,913	138,750	137,815	145,690
4인	155,003	155,919	154,033	155,006
5인이상	56,768	75,215	68,458	55,380

<부록 표 3-9> 시군구별 가구수 분포

시도	시군구	센서스	행정자료	능형	조사자료
11	180	86,533	90,065	86,813	88,606
11	190	138,637	146,970	139,908	138,672
23	080	128,975	130,689	127,826	128,023
24	010	42,538	38,868	39,373	42,153
31	240	158,423	164,961	160,591	161,309
32	380	6,919	8,201	7,297	7,192
33	390	11,559	12,136	11,997	11,062
35	350	10,693	11,949	11,244	10,740
36	040	31,795	35,774	34,465	33,191
36	480	16,114	19,078	17,038	15,871
37	030	48,843	49,033	48,051	48,976
37	430	3,494	4,546	3,617	3,436

<부록 표 3-10> 시군구*가구원수 별 가구수 분포

시군구	가구원수	센서스	행정자료	능형	조사자료
11180	1인	23,274	29,203	25,609	25,055
	2인	19,298	16,812	17,597	19,145
	3인	18,847	17,127	17,354	19,169
	4인	18,689	18,460	18,460	18,808
	5인이상	6,425	8,463	7,793	6,429
11190	1인	34,158	46,732	40,937	35,130
	2인	32,312	28,795	28,930	32,400
	3인	31,145	28,323	28,460	31,127
	4인	30,647	29,595	29,595	30,505
	5인이상	10,375	13,525	11,987	9,510
23080	1인	21,661	25,710	23,781	22,009
	2인	26,018	22,980	23,060	25,168
	3인	31,400	29,160	29,169	31,652
	4인	37,761	37,840	37,723	37,551
	5인이상	12,135	14,999	14,094	11,643
24010	1인	15,022	14,193	13,882	14,729
	2인	11,220	8,218	9,190	11,412
	3인	7,124	6,532	6,643	7,212
	4인	6,146	5,849	5,849	5,785
	5인이상	3,026	4,076	3,809	3,016
31240	1인	29,792	36,826	33,491	31,178
	2인	32,906	30,338	30,433	33,241
	3인	35,184	34,159	34,136	36,163
	4인	44,775	44,135	43,945	45,212
	5인이상	15,766	19,503	18,586	15,517
32380	1인	1,652	2,865	2,093	1,650
	2인	2,347	1,854	1,939	2,502
	3인	1,228	1,320	1,323	1,251
	4인	1,142	1,240	1,157	1,167
	5인이상	550	922	785	621
33390	1인	3,009	3,550	3,516	2,855
	2인	3,108	2,628	2,628	2,989
	3인	2,196	2,233	2,180	2,082
	4인	2,136	2,280	2,227	1,976
	5인이상	1,110	1,445	1,445	1,160

<부록 표 3-10> 시군구*가구원수 별 가구수 분포 (계속)

시군구	가구원수	센서스	행정자료	능형	조사자료
35350	1인	3,609	4,061	3,946	3,636
	2인	4,333	3,429	3,692	4,324
	3인	1,257	1,859	1,548	1,224
	4인	845	1,359	1,108	871
	5인이상	649	1,241	949	686
36040	1인	10,545	13,272	12,149	10,891
	2인	10,662	9,276	9,915	11,688
	3인	4,603	5,548	5,372	4,639
	4인	3,700	4,262	3,875	3,482
	5인이상	2,285	3,416	3,154	2,490
36480	1인	6,070	7,523	7,027	6,082
	2인	6,892	5,561	5,831	6,858
	3인	1,528	2,569	2,059	1,522
	4인	874	1,630	1,149	795
	5인이상	750	1,795	973	614
37030	1인	13,494	13,909	13,558	13,237
	2인	15,010	11,679	12,319	14,526
	3인	8,848	9,248	8,942	9,108
	4인	7,947	8,753	8,548	8,554
	5인이상	3,544	5,444	4,684	3,551
37430	1인	1,145	1,963	1,317	1,112
	2인	1,302	1,009	1,076	1,339
	3인	553	672	629	540
	4인	341	516	398	301
	5인이상	153	386	198	144

<부록 표 3-11> 시군구 별 주택수 분포

시도	시군구	센서스	행정자료	능형	조사자료
11	180	52,320	50,493	50,464	53,841
11	190	100,330	97,713	97,586	101,927
23	080	128,715	135,316	133,076	127,546
24	010	29,302	30,626	29,725	28,891
31	240	153,789	155,791	156,583	158,549
32	380	7,289	6,784	6,866	7,406
33	390	11,840	11,880	11,846	11,399
35	350	12,439	13,448	13,186	12,281
36	040	34,807	36,524	35,681	36,438
36	480	17,762	17,438	17,124	17,163
37	030	48,542	48,663	48,236	48,964
37	430	3,356	3,414	3,320	3,293

<부록 표 3-12> 주택유형 별 주택수 분포

주택분류별	센서스	행정자료	능형	조사자료
단독	172,707	165,217	164,986	174,363
아파트	323,846	322,952	322,741	323,559
기타	103,938	119,921	115,965	109,776

<부록 표 3-13> 시군구*주택유형 별 주택수 분포

시도	시군구	주택분류별	센서스	행정자료	능형	조사자료
11	180	단독	11,114	10,362	10,333	10,412
		아파트	24,873	23,794	23,794	26,862
		기타	16,333	16,337	16,337	16,567
11	190	단독	19,598	17,928	18,061	19,388
		아파트	60,146	59,345	59,345	60,524
		기타	20,586	20,440	20,180	22,015
23	080	단독	8,177	7,481	6,650	8,383
		아파트	79,499	81,774	81,165	76,708
		기타	41,039	46,061	45,261	42,456
24	010	단독	14,525	12,990	13,132	15,605
		아파트	11,602	11,461	11,460	9,490
		기타	3,175	6,175	5,132	3,796
31	240	단독	26,779	25,593	25,759	26,306
		아파트	110,831	111,106	111,505	114,586
		기타	16,179	19,092	19,319	17,657
32	380	단독	4,781	4,232	4,404	5,020
		아파트	1,749	1,439	1,439	1,351
		기타	759	1,113	1,023	1,036
33	390	단독	3,930	3,800	3,800	5,249
		아파트	6,976	6,727	6,727	5,271
		기타	934	1,353	1,319	879
35	350	단독	10,807	11,419	11,265	10,759
		아파트	1,195	1,213	1,213	1,010
		기타	437	816	708	513
36	040	단독	25,156	25,419	25,568	27,314
		아파트	8,596	8,450	8,450	8,271
		기타	1,055	2,655	1,663	852
36	480	단독	17,323	16,446	16,408	16,810
		아파트	40	20	20	0
		기타	399	972	696	353
37	030	단독	27,811	26,873	26,933	26,432
		아파트	18,110	17,513	17,513	19,255
		기타	2,621	4,277	3,790	3,276
37	430	단독	2,706	2,674	2,674	2,685
		아파트	229	110	110	232
		기타	421	630	536	376

[부록 3-2] 2013년 연구의 방향

- 2010년 전국자료를 이용한 인구·가구·주택의 총수 추정
 - * 공표단위 설정을 위한 능형회귀방안 적용 결과 분석
 - * 조사자료의 통계적 정확성을 고려한 신뢰계수 결정에 대한 연구
 - * 인구·가구·주택의 총수 산출 결과에 대한 다각적인 검토
 - 가능한 공표단위별 결합 및 주변분포 분석
- 레이킹 방안 검토
 - * 인구, 가구, 그리고 주택 자료의 레이킹(raking) 방안 연구
 - 레이킹 수준에 대한 검토
 - 세 자료의 consistency(일치성)를 고려한 레이킹 방안 검토
 - 통계청과의 협의가 매우 중요
 - 레이킹 후 산출된 최종 가중치를 이용한 인구·가구·주택수 통계 산출 및 검증

제4장 소지역 추정방법론

- 행정자료 인구수는 일반적으로 오류가 있음이 알려져 있으므로 이를 수정하기 위한 근거로 20% 표본조사가 고려되고 있다. 조사자료와 행정자료를 동시에 하나의 모형에 고려하여 행정자료를 수정하는 방법에 관한 연구가 필요하다. 특히 읍면동 수준에서 지역, 성, 연령과 같은 인구학적 특성별 인구수의 산출을 위해 소지역 추정(small-area estimation) 방법을 고려한다.
- 20% 표본조사 결과에서 얻은 각 부모집단(subpopulation)별 인구수 추정값과 행정자료를 사용하여 소지역 모형을 설정한다. 이 경우 읍면동 같은 소지역에서는 성별 연령별 부모집단에 할당되는 표본수가 상대적으로 작게 되어 각 부모집단별 인구수 산출시 사용되는 추정값은 신뢰할 수 없을 정도로 불안정하게 된다. 이를 극복하기 위한 방안으로써 이웃의 비슷한 부모집단으로부터 “정보를 빌려옴”(borrowing strength)으로 해서 부모집단별 추정의 효율성을 증가시키는 소지역 추정을 고려하고자 한다.
- 지역별 성별 연령(5세 단위)별 부모집단을 소영역(small domain)으로 설정하고 20% 표본조사에서 표본추출단위(조사구)를 고려하여 단위기반 소지역모형(unit-level small area model)을 설정한다. 지역으로는 시군구 수준과 읍면동 수준을 동시에 고려한다.
- 제4장에서는 지역·성·연령별 인구수를 산출하기 위해 네 가지 추정방법 - 직접 추정량(Direct estimator), 비 추정량(Ratio estimator), 경험적 최량선형불편 추정량(EBLUP: Empirical Best Linear Unbiased Predictor), 그리고 계층적 베이지 추정량(HB: Hierarchical Bayes estimator)을 사용하고자 한다. 이 때 직접 추정량과 비 추정량은 설계기반 추정방법(design-based estimation)에 해당되고 EBLUP과 HB 추정량은 모형기반 추정방법(model-based estimation)에 해당된다. 동일한 방법으로 지역·가구원수별 가구수와 지역·주택유형별 주택수를 산출한다.

- 센서스 값(gold standard)과 추정값들의 비교를 위한 그래프는 각 절의 마지막 부분에서 확인할 수 있으며, 실제 추정량들의 결과값은 부록의 형태로 별도 제출한다.

제1절 지역·성·연령별 인구수 산출 방안(1)

- 제1절에서는 지역·성·연령별 인구수를 추정하기 위해 직접 추정량, 비 추정량, EBLUP 추정량, HB 추정량을 계산하는 방법론에 대해 간단히 소개한 후, 이 네 가지 추정량의 비교를 위한 요약치 계산 또한 함께 소개하고자 한다.
- 네 가지 추정량들을 소개하기 전에 네 가지 추정량 계산에 공통으로 쓰이는 표기들을 먼저 정의하도록 하자. $\mathbf{y} = \{y_{ijsa}\}$ 는 인구수를 나타내는 벡터이며, 첨자 $i(i = 1, \dots, m)$ 는 소지역을, $j(j = 1, \dots, N_i)$ 는 해당 소지역에 속하는 조사구를, $s(s = 1, 2)$ 와 $a(a = 1, \dots, 21)$ 는 각각 성별과 연령을 나타낸다. 이 때, N_i 는 i 번째 지역의 전체 조사구 수를, n_i 는 i 번째 지역의 표본 조사구 수를 나타낸다. 또한, x_{ijsa} 는 행정자료, ν_s 는 성별, z_a 는 연령, 그리고 $\nu_s z_a$ 는 성별과 연령의 교호작용을 의미하며, 이러한 보조변수로 구성된 벡터를 $\mathbf{x}_{ijsa}$ 라 표기하기로 한다. 보조변수의 값은 다음과 같이 정의된다.

$$v_s = \begin{cases} -1 & \text{if } s = 1 \text{ (남)} \\ 1 & \text{if } s = 2 \text{ (여)} \end{cases}$$

$$z_a = \begin{cases} -10 & \text{if } a = 1 \text{ (age 0-4)} \\ -9 & \text{if } a = 2 \text{ (age 5-9)} \\ -8 & \text{if } a = 3 \text{ (age 10-14)} \\ -7 & \text{if } a = 4 \text{ (age 15-19)} \\ -6 & \text{if } a = 5 \text{ (age 20-24)} \\ -5 & \text{if } a = 6 \text{ (age 25-29)} \\ -4 & \text{if } a = 7 \text{ (age 30-34)} \\ -3 & \text{if } a = 8 \text{ (age 35-39)} \\ -2 & \text{if } a = 9 \text{ (age 40-44)} \\ -1 & \text{if } a = 10 \text{ (age 45-49)} \\ 0 & \text{if } a = 11 \text{ (age 50-54)} \\ 1 & \text{if } a = 12 \text{ (age 55-59)} \\ 2 & \text{if } a = 13 \text{ (age 60-64)} \\ 3 & \text{if } a = 14 \text{ (age 65-69)} \\ 4 & \text{if } a = 15 \text{ (age 70-74)} \\ 5 & \text{if } a = 16 \text{ (age 75-79)} \\ 6 & \text{if } a = 17 \text{ (age 80-84)} \\ 7 & \text{if } a = 18 \text{ (age 85-89)} \\ 8 & \text{if } a = 19 \text{ (age 90-94)} \\ 9 & \text{if } a = 20 \text{ (age 95-99)} \\ 10 & \text{if } a = 21 \text{ (age 100+)} \end{cases}$$

- 제1절에서는 지역 · 성 · 연령별 인구수가 관심 모수이며, 이를 Y_{isa} 로 표기하

도록 한다. 즉, $Y_{isa} = \sum_{j=1}^{N_i} y_{ijsa}$ 이며, 이는 지역이 i , 성별이 s , 연령이 a 인 그룹에 속하는 인구수를 의미한다.

1. 직접 추정량

- 관심 모수 Y_{isa} 의 직접 추정량이란 순전히 표본 조사값만을 이용하여 모수를 추정하는 것으로 표본의 합에 가중치를 곱함으로써 얻을 수 있다. 가중치에 관한 정보가 없을 경우에는 표본의 합에 i 번째 지역에 속하는 전체 조사구 수 N_i 와 표본 조사구 수 n_i 의 비를 곱하면 얻을 수 있다.
- 본 연구에서는 지역 i 에 대한 가중치 정보를 이용할 수 있으므로 이에 대한

직접추정량을 수식으로 표현하면 다음과 같다.

$$\hat{Y}_{isa}^D = w_i \sum_{j=1}^{n_i} y_{ijsa} \quad (4.1)$$

단, w_i 는 i 번째 지역에 대한 가중치이다.

2. 비 추정량

- 비 추정량은 표본 조사값만을 이용하는 직접추정량과는 달리, 보조정보를 이용하여 관심 모수 Y_{isa} 를 추정하는 것이다. 본 연구에서는 지역·성·연령별 인구수를 추정하기 위한 보조 정보로 행정자료인 x_{ijsa} 를 이용한다. 비 추정량을 수식으로 표현하면 다음과 같다.

$$\begin{aligned} \hat{Y}_{isa}^R &= \sum_{j=1}^{N_i} x_{ijsa} \frac{\bar{y}_{isa}^w}{\bar{x}_{isa}^w} \\ &= \sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} x_{ijsa}^* \frac{\bar{y}_{isa}^w}{\bar{x}_{isa}^w} \end{aligned} \quad (4.2)$$

단, $\bar{y}_{isa}^w = \sum_{j=1}^{n_i} \tilde{w}_{ijsa} y_{ijsa}$ 이고 $\bar{x}_{isa}^w = \sum_{j=1}^{n_i} \tilde{w}_{ijsa} x_{ijsa}$ 이며, $\tilde{w}_{ijsa}$ 는 $\sum_{j=1}^{n_i} \tilde{w}_{ijsa} = 1$

을 만족하는 정규화된 가중치(normalized weight)이다. 즉, $\bar{y}_{isa}^w$ 와 $\bar{x}_{isa}^w$ 는 각각 지역 i , 성별이 s , 연령이 a 인 그룹에 속하는 표본 조사값의 가중 평균과 표본에 대응되는 행정자료의 가중 평균이다.

- 그러나 본 연구에서 이용하는 $\tilde{w}_{ijsa}$ 의 값이 오로지 지역에만 영향을 받기 때문에 $\tilde{w}_{ijsa} = \tilde{w}_i$ 이므로, 식 (4.2)는 다음과 같이 가중치의 값이 상쇄됨을 알 수 있다.

$$\begin{aligned}
\hat{Y}_{isa}^R &= \sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} x_{ijsa}^* \frac{\bar{y}_{isa}^w}{\bar{x}_{isa}^w} \quad (4.3) \\
&= \sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} x_{ijsa}^* \frac{\tilde{w}_i \sum_{j=1}^{n_i} y_{ijsa}}{\tilde{w}_i \sum_{j=1}^{n_i} x_{ijsa}} \\
&= \sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} x_{ijsa}^* \frac{\sum_{j=1}^{n_i} y_{ijsa}}{\sum_{j=1}^{n_i} x_{ijsa}}
\end{aligned}$$

3. EBLUP 추정량

- 소지역에 해당하는 인구수를 추정하기 위해 다음과 같은 모형을 고려한다.

$$\begin{aligned}
y_{ijsa} &= \mathbf{x}_{ijsa}^T \boldsymbol{\beta} + u_i + \epsilon_{ijsa} \quad (4.4) \\
\mathbf{x}_{ijsa}^T \boldsymbol{\beta} &= \beta x_{ijsa} + \alpha \nu_s + \gamma z_a + \delta \nu_s z_a
\end{aligned}$$

단, $i = 1, \dots, m$, $j = 1, \dots, N_i$, $s = 1, 2$, $a = 1, \dots, 21$ 이며, u_i 는 지역 임의 (랜덤) 효과(random area effect)를 나타낸다.

이 때, u_i 와 ϵ_{ijsa} 는 서로 독립이며, $u_i (i = 1, \dots, m)$ 의 평균이 0이고, 분산이 σ_u^2 인 분포를, $\epsilon_{ijsa} (i = 1, \dots, m, j = 1, \dots, N_i, s = 1, 2, a = 1, \dots, 21)$ 는 평균이 0이고 분산이 σ^2 인 분포를 가정하되, 각각 독립이며 동일한 분포를 따른다(i.i.d)고 가정한다.

- 모형 식 (4.4)는 다음과 같이 행렬 구조로 다시 표현할 수 있다.

$$\mathbf{y}_{isa} = X_{isa} \boldsymbol{\beta} + v_i \mathbf{1}_{isa} + \boldsymbol{\epsilon}_{isa}, \quad i = 1, \dots, m \quad (4.5)$$

여기서 X_{isa} 는 $N_i \times p$ 행렬을, y_{isa} , $\mathbf{1}_{isa}$, 그리고 ϵ_{isa} 는 $N_i \times 1$ 벡터를 나타내며, $\mathbf{1}_i = (1, \dots, 1)^T$ 이고 p 는 보조변수의 개수를 나타낸다.

- Y_{isa} 는 본 연구에서의 첫 번째 관심 모수로, i 번째 지역에서 성별이 s 이고 연령이 a 인 그룹에 속하는 인구수를 나타낸다. 이를 우리는 다음과 같이 표본(sample)과 비표본(non-sample)으로 분해할 수 있다.

$$Y_{isa} = \sum_{j=1}^{N_i} y_{ijsa} = \sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} y_{ijsa}^* \quad (4.6)$$

단, 위첨자 *는 비표본임을 나타낸다.

- 위의 모형을 기반으로 하는 EBLUP 추정량을 구하는 절차를 소개하고자 한다. 본 연구에서는 추출률(sampling fraction), $f_i = n_i/N_i$ 를 무시할 수 없을 경우의 EBLUP 추정량을 계산하였다. 먼저, 관심 모수 Y_{isa} 는 다음과 같이 표현할 수 있다.

$$Y_{isa} = n_i \bar{y}_{isa} + (N_i - n_i) \bar{y}_{isa}^* \quad (4.7)$$

단, $\bar{y}_{isa}$ 와 $\bar{y}_{isa}^*$ 는 각각 지역이 i , 성별이 s , 그리고 연령이 a 인 그룹에 속하는 표본 조사구의 평균과 비표본 조사구의 평균이다.

- 모형 식 (4.6)에서 비표본 인구수 y_{ijsa}^* 를 추정치 $\mathbf{x}_{ijsa}^{*T} \tilde{\boldsymbol{\beta}} + \tilde{u}_i$ 로 대체하여 전개하면 다음과 같은 Y_{isa} 에 대한 BLUP 추정량을 얻을 수 있다.

$$\tilde{Y}_{isa}^{BLUP} = n_i \bar{y}_{isa} + (N_i - n_i) (\bar{\mathbf{x}}_{isa}^{*T} \tilde{\boldsymbol{\beta}} + \tilde{u}_i) = n_i \bar{y}_{isa} + (N_i - n_i) \tilde{y}_{isa}^{*H} \quad (4.8)$$

여기서 $\bar{\mathbf{x}}_{isa}^*$ 는 지역이 i , 성별이 s , 그리고 연령이 a 인 그룹에 속하는 비표본 조사구의 행정자료에 대한 평균을 나타내는 값이며,

$$\tilde{y}_{isa}^{*H} = \gamma_i [\bar{y}_{isa} + (\bar{\mathbf{x}}_{isa}^* - \bar{\mathbf{x}}_{isa})^T \tilde{\boldsymbol{\beta}}] + (1 - \gamma_i) \bar{\mathbf{x}}_{isa}^{*T} \tilde{\boldsymbol{\beta}} \quad (4.9)$$

이고, $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma^2 / n_i)$ 이다.

- 이 때, (4.9)식에서 σ_u^2 과 σ^2 대신 각각의 추정량 $\hat{\sigma}_u^2$ 과 $\hat{\sigma}^2$ 을 대입하면 다음과 같이 EBLUP 추정량을 구할 수 있다.

$$\hat{Y}_{isa}^{EBLUP} = n_i \bar{y}_{isa} + (N_i - n_i) \hat{y}_{isa}^{*H} \quad (4.10)$$

단, $\hat{y}_{isa}^{*H} = \hat{\gamma}_i [\bar{y}_{isa} + (\bar{\mathbf{x}}_{isa}^* - \bar{\mathbf{x}}_{isa}) \hat{\boldsymbol{\beta}}] + (1 - \hat{\gamma}_i) \bar{\mathbf{x}}_{isa}^{*T} \hat{\boldsymbol{\beta}}$ 이고, $\hat{\gamma}_i = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \hat{\sigma}^2 / n_i)$ 이다.

4. 계층적 베이즈 추정량

- 계층적 베이즈 추정량은 앞서 소개한 EBLUP 추정량과 모형 가정에서부터 차이가 있다. 계층적 베이즈 추정량을 구하기 위해 기반이 되는 모형은 (4.4)와 동일한 구조를 가지지만, u_i 와 ϵ_{ijsa} 에 ‘정규성’이라는 분포 가정을 추가한다. 다시 말해, u_i 와 ϵ_{ijsa} 는 서로 독립이며, $u_i (i = 1, \dots, m)$ 는 평균이 0이고, 분산이 σ_u^2 인 정규분포를, $\epsilon_{ijsa} (i = 1, \dots, m, j = 1, \dots, N_p, s = 1, 2, a = 1, \dots, 21)$ 는 평균이 0이고 분산이 σ^2 인 정규분포를 가정하되, 각각 독립이며 동일한 분포를 따른다(i.i.d)고 가정한다.
- $y_{ijsa} (j = 1, \dots, n_i)$ 는 표본으로서 현재 알고 있는 값이기 때문에, Y_{isa} 를 추정

하는 문제는 결국 식 (4.6)의 두 번째 항인 $\sum_{j=n_i+1}^{N_i} y_{ijsa}^*$ 를 추정하는 문제와 동일하게 된다. 따라서 본 연구에서는 모수 Y_{isa} 를 추정하기 위해, 마르코프 연쇄 몬테 칼로(Markov Chain Monte Carlo; MCMC)를 이용하여 비표본 $y_{ijsa}^* (j = n_i + 1, \dots, N_i)$ 의 예측 분포를 찾는다.

- 이를 위해 먼저 σ_u^2 과 σ^2 의 역수를 각각 R_u 와 R 로 표기하자. 그러면, (4.4) 모형 식은 다음과 같은 계층적 베이지안 형태로 표현할 수 있다.

$$\text{I. } y_{ijsa} | \mathbf{u}, \boldsymbol{\beta}, R_u = r_u, R = r \sim N(\mathbf{x}_{ijsa}^T \boldsymbol{\beta} + u_i, r^{-1}), \quad (j = 1, \dots, N_i)$$

$$\text{II. } u_i | R_u = r_u \sim N(0, r_u^{-1})$$

$$\text{III. } \boldsymbol{\beta} \sim \text{Uniform}(\mathbf{R}^p), R_u \sim \text{Gamma}(\frac{1}{2}a, \frac{1}{2}b), R \sim \text{Gamma}(\frac{1}{2}c, \frac{1}{2}d)$$

(단, $\boldsymbol{\beta}$, R_u , 그리고 R 은 서로 독립이다.)

- 모수 추정량의 대입을 이용하는 EBLUP 추정량과는 달리, 계층적 베이지안 추정량에서는 초모수(hyperparameter)에 초사전분포를 가정하여 MCMC에 근거한 사후평균을 이용한다는 점이 또 다른 큰 차이점이다.
- 위와 같이 주어진 모형에 대한 계층적 베이지안 분석을 수행하기 위해 본 연구에서는 깁스 샘플러(Gibbs sampler)를 사용하여 계산을 수행하였다. 깁스 샘플러를 사용하기 위해 표본 $\mathbf{y}_s$ 및 다른 모수들이 주어졌을 때의 조건부 확률분포를 각각 구하면 다음과 같다.

$$\text{i. } \boldsymbol{\beta} | u_i, r, r_u, \mathbf{y}_s \sim N((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_s, r^{-1} (\mathbf{X}^T \mathbf{X})^{-1})$$

$$\text{ii. } r_u | \mathbf{u}, \boldsymbol{\beta}, r, \mathbf{y}_s \sim \text{Gamma}(\frac{1}{2}(\sum_i u_i^2 + a), \frac{1}{2}(m + b))$$

$$\text{iii. } r | \mathbf{u}, \boldsymbol{\beta}, r_u, \mathbf{y}_s \sim \text{Gamma}(\frac{1}{2} \sum_i \sum_j \sum_s \sum_a (y_{ijsa} - \mathbf{x}_{ijsa}^T \boldsymbol{\beta} - u_i)^2, \frac{1}{2}(\sum_i n_i \times 2 \times 21 + d))$$

iv. $u_i | \beta, r, r_u, \mathbf{y}_s \sim$

$$N\left(\frac{r}{r(n_i \times 2 \times 21) + r_u} \sum_j \sum_s \sum_a (y_{ijsa} - \mathbf{x}_{ijsa}^T \beta), \{r(n_i \times 2 \times 21) + r_u\}^{-1}\right)$$

단, $X = (\mathbf{x}_{1111}^T, \dots, \mathbf{x}_{1n_1 221}^T, \dots, \mathbf{x}_{m111}^T, \dots, \mathbf{x}_{mn_m 221}^T)^T$ 이다.

- 구체적으로 2,000개의 반복을 사용하였으며 β , r_u , r , 그리고 $\mathbf{u}$ 의 난수를 각각 2,000개 생성한 후, 초기값의 영향을 제거하기 위해 반복의 앞부분 1,000개를 버리고(burn-in) 나머지 1,000개를 사용하여 사후평균을 계산하였다.
- 다음은 위와 같이 MCMC를 이용한 모수 추정치를 가지고 관심 모수인 Y_{isa} 의 추정량을 찾는 과정을 설명한 것이다.
먼저, 조건부 확률 공식을 이용하면 $\mathbf{y}_s$ 가 주어졌을 때 관심 모수 Y_{isa} 의 조건부 확률분포 및 평균을 다음과 같이 표현할 수 있다.

$$i. p(Y_{isa} | \mathbf{y}_s) = \int p(Y_{isa} | \beta, r, \mathbf{u}, \mathbf{y}_s) \pi(\beta, r, \mathbf{u} | \mathbf{y}_s) d\beta dr d\mathbf{u} \quad (4.11)$$

$$ii E[Y_{isa} | \mathbf{y}_s] = \int \sum_{j=1}^{N_i} y_{ijsa} p(Y_{isa} | \mathbf{y}_s) dY_{isa} \quad (4.12)$$

$$= \int \int Y_{isa} p(Y_{isa} | \beta, r, \mathbf{u}, \mathbf{y}_s) \pi(\beta, r, \mathbf{u} | \mathbf{y}_s) d\beta dr d\mathbf{u} dY_{isa}$$

- 이 때

$$Y_{isa} | \beta, r, \mathbf{u}, \mathbf{y}_s \sim N\left(\sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} \mathbf{x}_{ijsa}^{*T} \beta + (N_i - n_i)u_i, (N_i - n_i)r^{-1}\right)$$

이므로, $\mathbf{y}_s$ 가 주어졌을 때 Y_{isa} 의 평균은 다음과 같이 다시 표현할 수 있다.

$$ii *. E[Y_{isa} | \mathbf{y}_s] =$$

$$\int \left(\sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} \mathbf{x}_{ijsa}^{*T} \boldsymbol{\beta} + (N_i - n_i) u_i \right) \pi(\mathbf{u}, \boldsymbol{\beta}, r | \mathbf{y}_s) d\mathbf{u} d\boldsymbol{\beta} dr \quad (4.13)$$

- 따라서 계층적 베이지 추정량 $\hat{Y}_{isa}^{HB}$ 는 위의 MCMC의 결과로 계산된 모수 추정치를 이용하여 다음과 같이 구할 수 있다.

$$\hat{Y}_{isa}^{HB} \approx \frac{1}{T} \sum_{t=1}^T \left\{ \sum_{j=1}^{n_i} y_{ijsa} + \sum_{j=n_i+1}^{N_i} \mathbf{x}_{ijsa}^{*T} \boldsymbol{\beta}^{(t)} + (N_i - n_i) u_i^{(t)} \right\} \quad (4.14)$$

(단, $i = 1, \dots, m$, $j = n_i + 1, \dots, N_i$, $s = 1, 2$, $a = 1, 2, \dots, 21$, $t = 1, 2, \dots, T$ 이며, $\boldsymbol{\beta}^{(t)}$, $u_i^{(t)}$ 는 각각 $\boldsymbol{\beta}$ 와 u_i ($i = 1, \dots, m$)의 MCMC 알고리즘에서 t 번째 반복에서 계산된 모수 추정치이다.)

- T 는 MCMC 시뮬레이션의 반복 횟수 중 초기값의 영향을 제거하기 위한 burn-in 기간(여기서는 1,000)을 제외한 횟수, 즉, 1,000이 된다. Burn-in 후의 시뮬레이션 반복 횟수 마다 나오는 MCMC 모수 추정치를 대입하여 계산한 뒤 평균을 취하여 관심 모수를 추정하는 것이 바로 MCMC의 몬테 카를로 부분에 해당된다. 이렇게 추정한 결과 값은 부록 표에 제시되어있다.

5. 추정량 비교를 위한 요약치 계산

- 앞서 소개한 네 가지 추정량 - 직접 추정량, 비추정량, EBLUP 추정량, 계층적 베이지 추정량 중 어떤 추정량이 실제 인구 수와 어느 정도로 가깝게 추정이 되는지를 비교하는 것 또한 본 연구의 목표 중 하나이다. 따라서 우리는 추정량의 비교를 위한 수치적 정보를 제공하는 요약치를 이용하여 네 가지 추정량을 비교하고자 한다.
- 본 연구에서는 센서스 인구수를 실제 인구수로 보았으며, 앞서 소개한 네 가지 추정량의 비교를 위한 요약치로 다음의 네 가지를 고려한다.

i . Average Relative Bias :

$$(m \times 2 \times 21)^{-1} \sum_{i=1}^m \sum_{s=1}^2 \sum_{a=1}^{21} |y_{isa}^c - \hat{y}_{isa}| / y_{isa}^c$$

ii . Average Squared Relative Bias :

$$(m \times 2 \times 21)^{-1} \sum_{i=1}^m \sum_{s=1}^2 \sum_{a=1}^{21} |y_{isa}^c - \hat{y}_{isa}|^2 / (y_{isa}^c)^2$$

iii . Average Absolute Bias :

$$(m \times 2 \times 21)^{-1} \sum_{i=1}^m \sum_{s=1}^2 \sum_{a=1}^{21} |y_{isa}^c - \hat{y}_{isa}|$$

iv . Average Squared Deviation :

$$(m \times 2 \times 21)^{-1} \sum_{i=1}^m \sum_{s=1}^2 \sum_{a=1}^{21} (y_{isa}^c - \hat{y}_{isa})^2$$

단, y_{isa}^c 와 $\hat{y}_{isa}$ 는 각각 지역이 i , 성별이 s , 그리고 연령이 a 인 그룹에 속하는 센서스의 인구수(실제 값)와 그에 해당하는 추정값이다.

- 본 연구에서는 소지역 i 를 시군구, 읍, 면, 동으로 각각 정의하여 네 가지 추정량들을 계산하였으며, 이에 대응하는 요약치 결과를 다음과 같이 표로 나타내었다.

<표 4-1> ‘시군구’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.1002	0.0515	104.97	32972.66
Ratio	0.0864	0.0452	64.23	9485.10
EBLUP	0.1300	0.0821	67.45	9941.22
HB	0.7310	27.5751	206.28	139667.48

<표 4-2> '읍'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.2108	0.1214	44.88	5259.29
Ratio	0.1484	0.0659	27.08	1861.97
EBLUP	0.1818	0.1084	26.26	2027.87
HB	1.1622	87.1695	48.88	7348.72

<표 4-3> '면'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.3013	0.2177	13.65	479.45
Ratio	0.2717	0.2074	11.12	312.38
EBLUP	0.3979	1.6991	10.50	218.32
HB	0.6182	5.1740	11.49	245.78

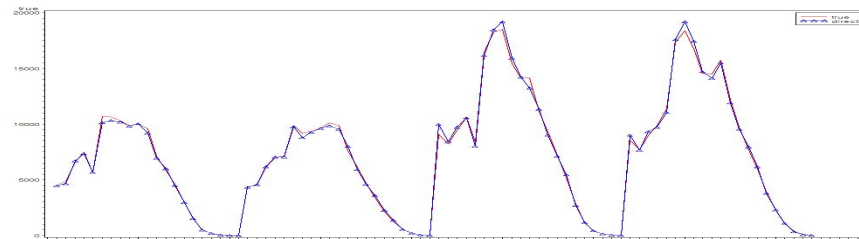
<표 4-4> '동'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.2026	0.1953	39.55	3971.68
Ratio	0.1380	0.0808	21.49	1113.90
EBLUP	0.1603	0.1014	21.56	1158.96
HB	0.3506	3.1275	21.84	1377.68

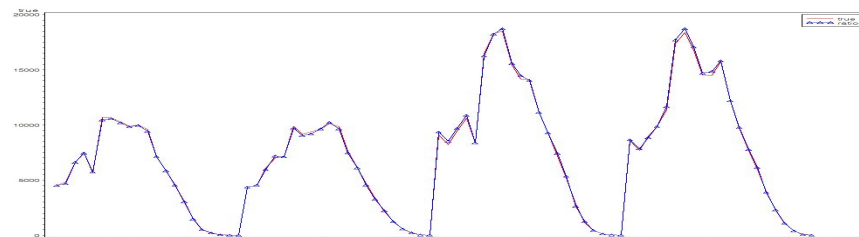
- 위의 표를 보면, 지역·성·연령별 인구수에 대해서는 비 추정량이 나머지 세 추정량에 비해 실제 값에 가깝게 추정되었음을 확인할 수 있다. 그 다음으로 EBLUP 추정량이 실제 값에 가깝게 추정이 되었음을 확인할 수 있으며, 특히 '면'의 경우에는 비 추정량보다 EBLUP 추정량의 AAB와 ASD 값이 더 작은 것을 확인할 수 있는데, 이는 비 추정량에 비해 EBLUP 추정량이 큰 값에서 좀 더 잘 추정이 되었음을 의미한다.

<그림 4-1> 센서스 값과 추정값 비교 (금천구, 영등포구)

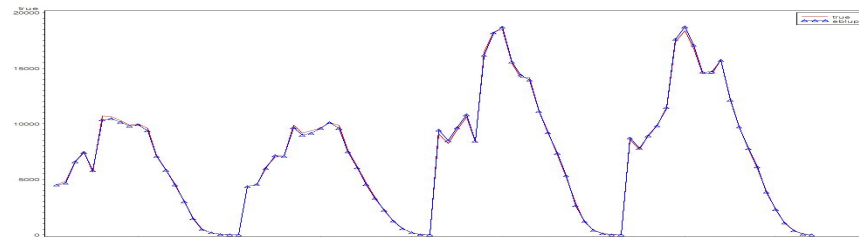
Direct



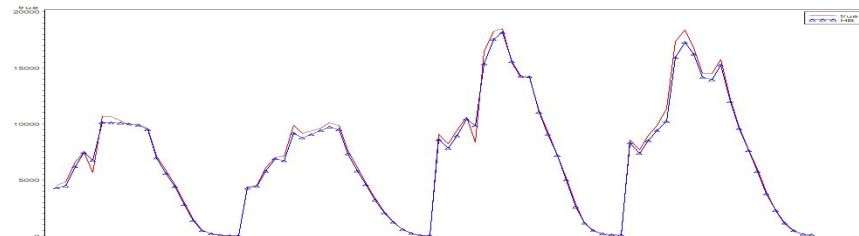
Ratio



EBLUP



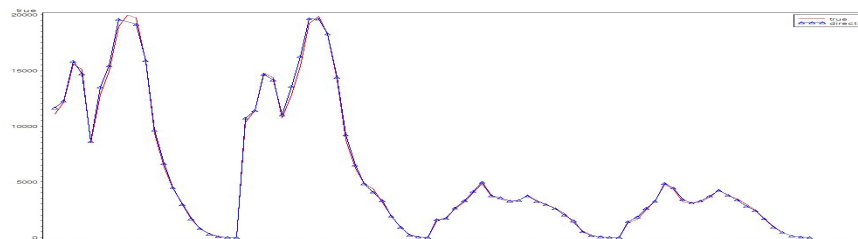
HB



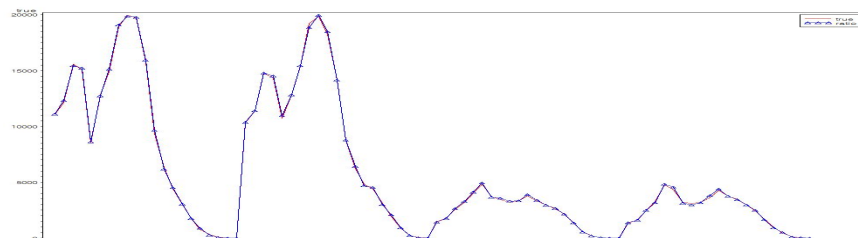
★★ 센서스 값 : — , 추정값 : —

<그림 4-2> 센서스 값과 추정값 비교 (인천서구, 광주동구)

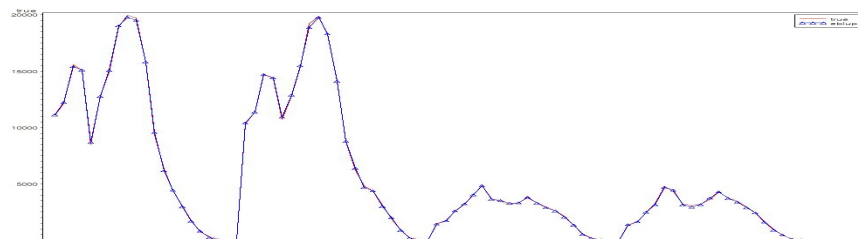
Direct



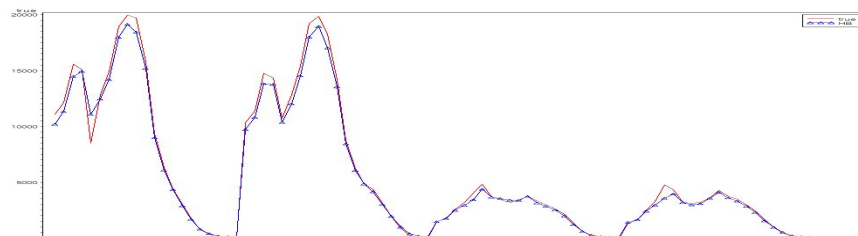
Ratio



EBLUP



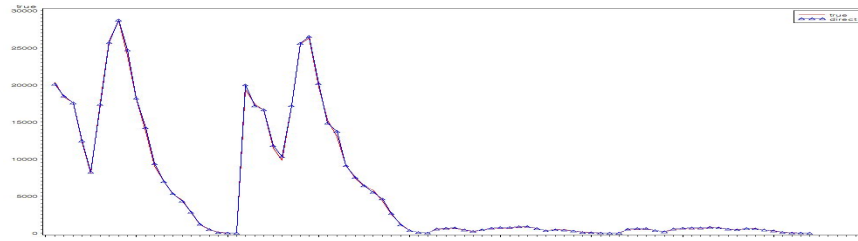
HB



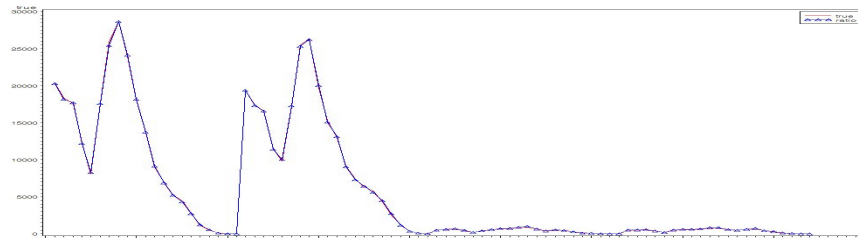
★★ 센서스 값 : — , 추정값 : —

<그림 4-3> 센서스 값과 추정값 비교 (화성시, 양구군)

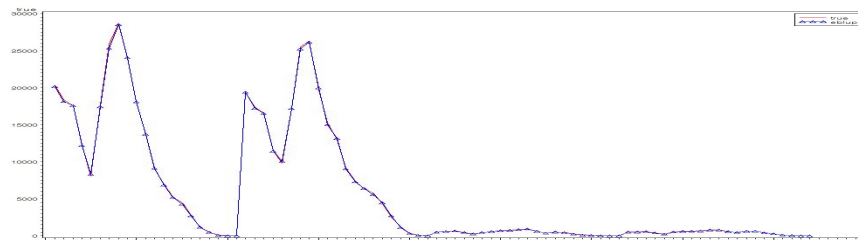
Direct



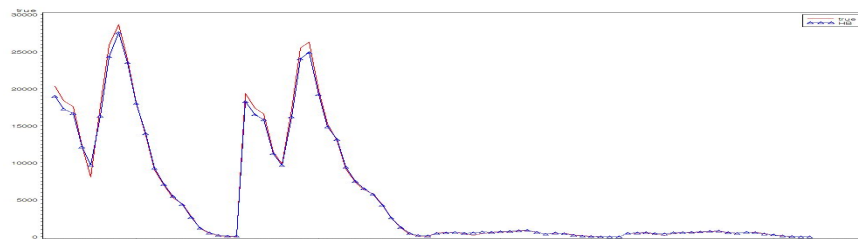
Ratio



EBLUP



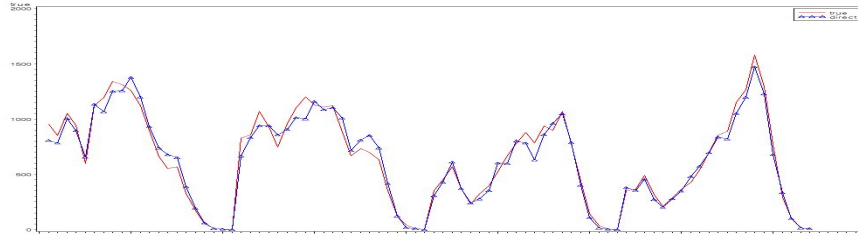
HB



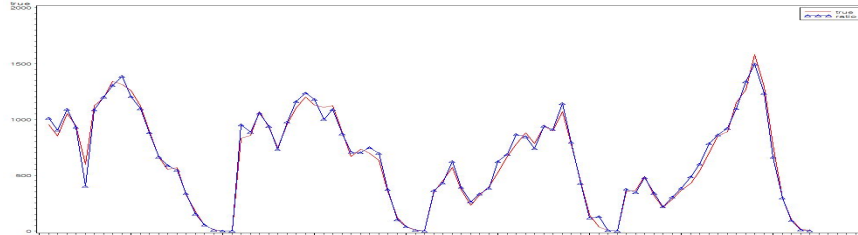
★★ 센서스 값 : — , 추정값 : —

<그림 4-4> 센서스 값과 추정값 비교 (증평균, 임실군)

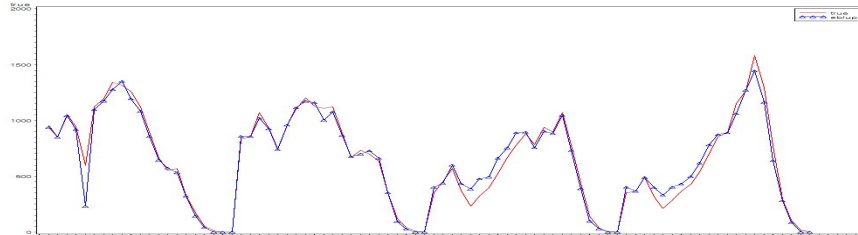
Direct



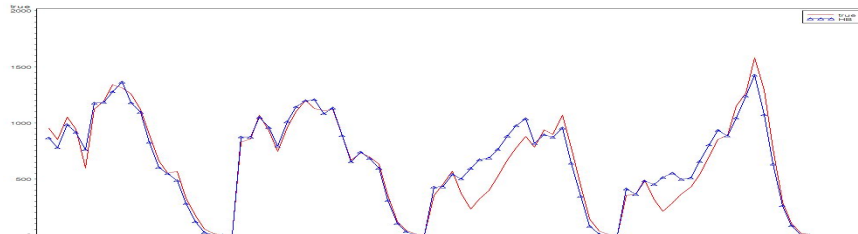
Ratio



EBLUP



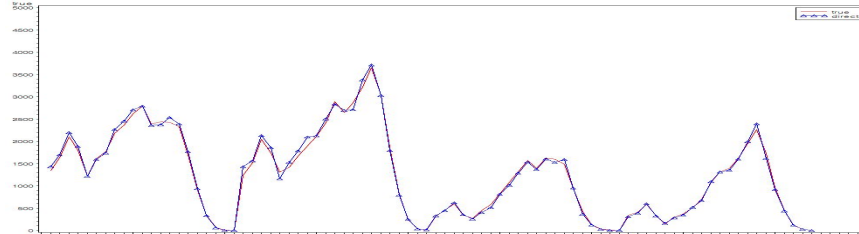
HB



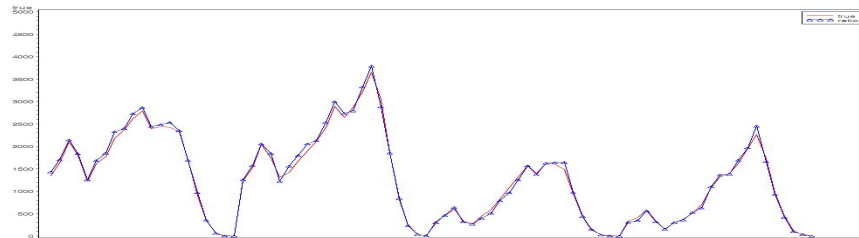
** 센서스 값 : — , 추정값 : —

<그림 4-5> 센서스 값과 추정값 비교 (나주시, 신안군)

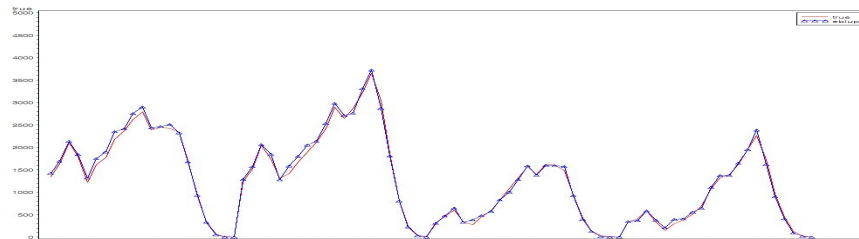
Direct



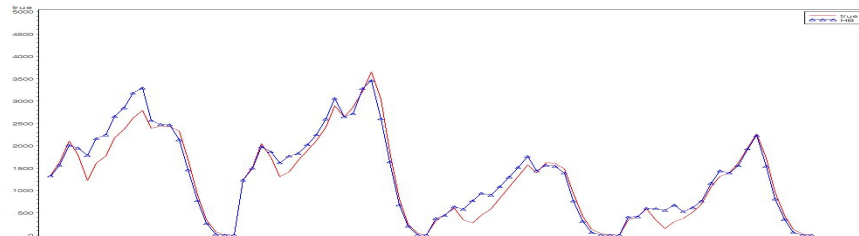
Ratio



EBLUP



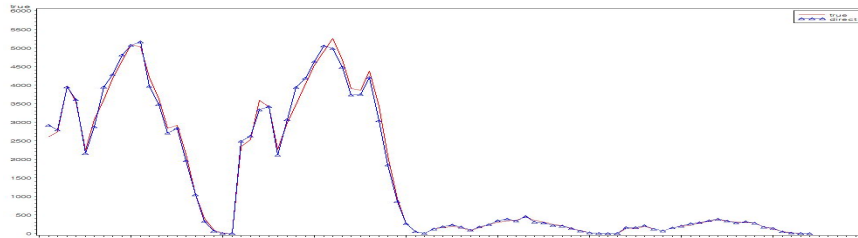
HB



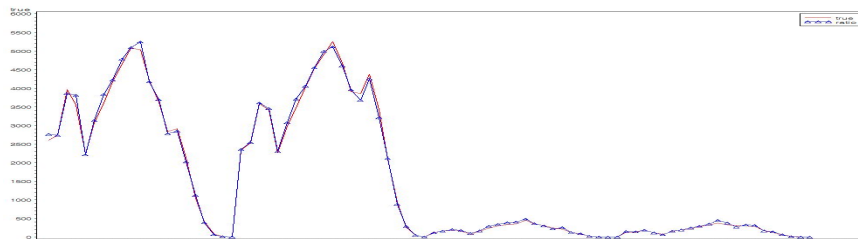
** 센서스 값 : — , 추정값 : —

<그림 4-6> 센서스 값과 추정값 비교 (김천시, 울릉군)

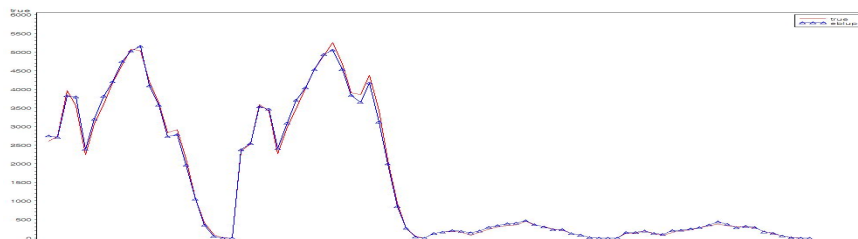
Direct



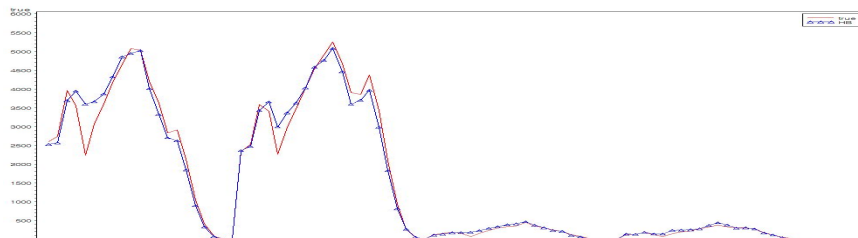
Ratio



EBLUP



HB



★★ 센서스 값 : — , 추정값 : —

제2절 지역·성·연령별 인구 수 산출 방안(2)

- 이 절에서는 지역·성·연령별 인구수를 산출하는 또 다른 방안을 제안하고자 한다. 이는 제1절에서 소개한 직접 추정량, 비추정량, EBLUP 추정량, 그리고 계층적 베이지 추정량을 그대로 사용하되, 보조변수인 연령을 3개의 범주로 구분하여 소영역 인구수를 추정한 다음, 행정자료의 정보를 이용하여 다시 21개의 연령 범주에 대한 인구수를 산출하는 방법이다.
- 계층적 베이지 추정량을 예로 들어 보자. 먼저 보조변수이자 분류변수인 연령에 대한 첨자 $a(a = 1, \dots, 21)$ 를 A 로 다음과 같이 다시 정의하자.

$$A = \begin{cases} 1 & \text{if } a = 1, 2, 3 \quad (\text{age } 0-14) \\ 2 & \text{if } a = 4, \dots, 13 \quad (\text{age } 15-64) \\ 3 & \text{if } a = 14, \dots, 21 \quad (\text{age } 65+) \end{cases}$$

첨자 a 는 연령을 21개의 범주로 나누었다면 위에서 정의한 첨자 A 는 21개의 범주를 다시 크게 3개의 범주로 구분하였다. 첨자 A 를 이용하여 보조변수 z_a 에 대응되는 z_A 를 다음과 같이 정의하자.

$$z_A = \begin{cases} -1 & \text{if } A = 1 \quad (\text{age } 0-14) \\ 0 & \text{if } A = 2 \quad (\text{age } 15-64) \\ 1 & \text{if } A = 3 \quad (\text{age } 65+) \end{cases}$$

- 계층적 베이지 추정량을 구하기 위해 기반이 되는 모형은 (4.4)와 동일한 구조를 가지되 변수 z_a 대신 z_A 를 대입하면 된다. 제1절-4. 계층적 베이지 추정량에서 소개한 절차를 그대로 수행하면 지역이 i , 성별이 s , 그리고 연령이 a' 인 그룹에 속하는 인구수에 대한 계층적 베이지 추정량을 얻을 수 있다. 이를 $\hat{Y}_{isA}^{HB(2)}$ 라 표기하자. 그러나 $\hat{Y}_{isA}^{HB(2)}$ 는 우리가 원하는 관심 모수 Y_{isa} 에 대한 추정량이 아니다. 따라서 본 연구에서는 관심 모수에 적합한 추정량을 계산하기 위해 행정자료의 정보를 한 번 더 이용하였으며, 수식으로 표현하면 다음과 같다.

$$\hat{Y}_{isa}^{HB(2)} = \hat{Y}_{isA}^{HB(2)} \frac{\sum_{j=1}^{N_i} x_{ijsa}}{\sum_{a \in A} \sum_{j=1}^{N_i} x_{ijsa}} \quad (4.15)$$

- 위의 방법은 EBLUP 추정량, Ratio 추정량에도 그대로 적용하여 $\hat{Y}_{isa}^{EBLUP(2)}$, $\hat{Y}_{isa}^{R(2)}$ 를 얻을 수 있으며, 세 가지 추정량 또한 제1절에서 소개한 요약치를 이용하여 실제 값과 어느 정도 차이가 있는지를 비교할 수 있다.
- 방안(2)를 이용하여 계산한 네 가지 추정량들을 비교하기 위한 요약치 결과를 다음과 같이 표로 나타내었다. 이 절에서는 소지역 i 를 시군구를 제외한 읍, 면, 동으로만 각각 정의하여 계산한다.

<표 4-5> ‘읍’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.1623	0.0928	34.78	3113.44
Ratio	0.1496	0.1009	27.51	2223.05
EBLUP	0.1431	0.0982	25.68	2088.59
HB	0.1582	0.1146	33.30	4443.22

<표 4-6> ‘면’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.3343	0.6451	12.65	365.89
Ratio	0.3301	0.6774	11.28	282.15
EBLUP	0.3146	0.5840	11.41	274.77
HB	1.9925	4.3886	66.57	1772.04

<표 4-7> ‘동’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.1099	0.0389	27.03	2053.09
Ratio	0.0924	0.0363	21.30	1553.49
EBLUP	0.0913	0.0347	21.14	1547.92
HB	5.3116	2.3320	1152.91	87463.97

- 위의 표를 보면, 지역·성·연령별 인구수에 대해서는 EBLUP 추정량이 나머지 세 추정량에 비해 실제 값에 가깝게 추정되었음을 알 수 있다. 그 다음으로 비 추정량이 실제 값에 가깝게 추정이 되었음을 확인할 수 있으며, 특히 ‘면’의 경우에는 EBLUP 추정량보다 비 추정량의 ARB와 AAB의 값이 더 작은 것을 확인할 수 있으며, 전반적으로 비 추정량과 EBLUP 추정량의 비교 요약치 결과가 다른 나머지 추정량에 비해 크게 차이가 나지 않는다고 해석할 수 있다.
- 또한, 방안 (1)과 방안 (2)의 요약치 결과 값을 비교해보면, 직접 추정량과 비 추정량의 경우 방안 (2)의 값이 대체로 더 작은 것을 확인할 수 있다. 반면, EBLUP 추정량의 경우, 방안 (2)에서 ARB와 ASRB의 값이 작아지는 반면, AAB와 ASD의 값은 커지는 경향이 있으며, 계층적 베이지스 추정량의 경우 ‘읍’을 제외한 ‘면’과 ‘동’에서 값이 크게 증가하는 것을 확인할 수 있다.

제3절 지역·가구원수별 가구수 산출 방안

- 이 절에서는 지역·가구원수별 가구수를 추정하기 위한 직접 추정량, 비 추정량, EBLUP 추정량, 계층적 베이지스 추정량을 사용하고, 이 추정량들을 비교하기 위한 네 가지 요약치도 함께 계산하고자 한다.
- 이 절에서 $y = \{y_{ig}\}$ 는 인구를 나타내는 모수벡터이며, 첨자 $i(i = 1, \dots, m)$ 는 소지역을, $j(j = 1, \dots, N_i)$ 는 해당 소지역에 속하는 조사구를, $g(g = 1, 2, \dots, 5)$ 는 가구원 수를 나타낸다. 이 때, N_i 는 제1절에서와 마찬가지로

로 i 번째 지역의 전체 조사구 수를, n_i 는 i 번째 지역의 표본 조사구 수를 나타낸다. 또한, x_{ijg} 는 행정자료, v_g 는 가구원 수를 나타내며 이러한 보조변수로 구성된 벡터를 $\mathbf{x}_{ijg}$ 라 표기하기로 한다. 보조변수 v_g 는 다음과 같이 정의된다. 보조변수의 값은 다음과 같이 정의된다.

$$v_g = \begin{cases} 1 & \text{if } g = 1 \text{인 가구} \\ 2 & \text{if } g = 2 \text{인 가구} \\ 3 & \text{if } g = 3 \text{인 가구} \\ 4 & \text{if } g = 4 \text{인 가구} \\ 5 & \text{if } g = 5 \text{인 이상 가구} \end{cases}$$

- 이 절에서는 지역 . 가구원수별 가구 수가 관심 모수이며, 이를 Y_{ig} 로 표기한다. 즉, $Y_{ig} = \sum_{j=1}^{N_i} y_{ijg}$ 이며, 이는 지역이 i , 가구원 수가 g 인 그룹에 속하는 가구 수를 의미한다.
- 관심 모수 Y_{ig} 의 직접 추정량, 비 추정량, EBLUP 추정량, 계층적 베이지 추정량은 앞 절에서 소개한 공식과 동일한 형태를 가진다. 마찬가지로 네 가지 추정량의 비교를 위한 요약치 또한 동일하다.
- 다음은 앞서 제안한 추정량들을 비교하기 위한 요약치 결과를 표로 나타낸 것이다.

<표 4-8> ‘시군구’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.0407	0.0032	254.29	208914.87
Ratio	0.0385	0.0036	211.70	92226.45
EBLUP	0.0364	0.0031	192.63	81500.67
HB	0.1752	0.0616	844.30	1328858.29

<표 4-9> ‘읍’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.1049	0.0246	97.72	21676.11
Ratio	0.0921	0.0224	84.07	17993.30
EBLUP	0.0694	0.0131	71.55	11751.46
HB	0.1640	0.0503	127.03	25265.67

<표 4-10> ‘면’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.1658	0.0449	29.74	1958.22
Ratio	0.1596	0.0569	24.95	1349.79
EBLUP	0.1611	0.0479	30.24	2087.81
HB	0.2475	0.1434	43.84	4523.83

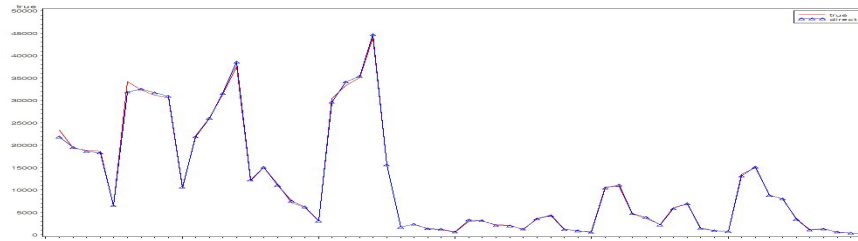
<표 4-11> ‘동’에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.0766	0.0099	85.92	15792.37
Ratio	0.0610	0.0076	63.57	8501.68
EBLUP	0.0648	0.0063	75.96	10648.09
HB	0.0987	0.0225	103.83	24779.70

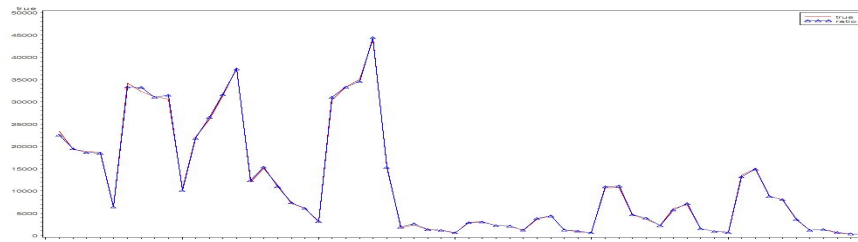
- 위의 표를 보면, 지역.가구원수별 가구 수에 대해서는 소지역을 ‘시군구’와 ‘읍’으로 정의했을 때, EBLUP 추정량이 실제 값에 가장 가까이 추정된 반면, ‘면’과 ‘동’에 대해서는 비 추정량이 대체로 실제 값에 가깝게 추정되었음을 확인할 수 있다.

<그림 4-7> 각 시군구별 센서스 값과 추정값 비교

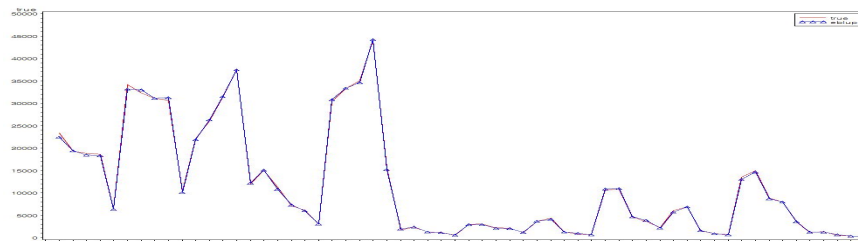
Direct



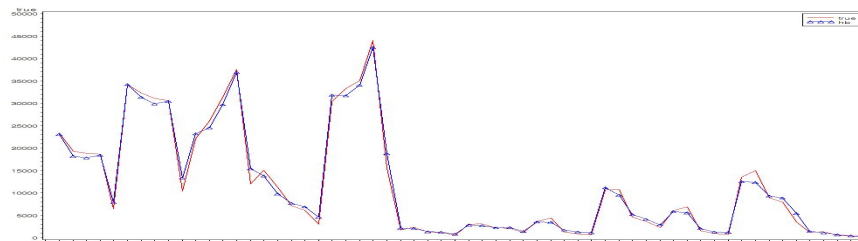
Ratio



EBLUP



HB



** 센서스 값 : — , 추정값 : —

제4절 지역·주택유형별 주택수 산출 방안

- 이 절에서도 지역·주택유형별 주택수를 추정하기 위해 직접 추정량, 비 추정량, EBLUP 추정량, HB 추정량을 사용하고, 이들 네 가지 추정량의 비교를 위한 요약치를 계산하고자 한다.
- 비슷하게, $y = \{y_{ih}\}$ 는 주택수를 나타내는 벡터이며, 첨자 $i (i = 1, \dots, m)$ 는 소지역을, $j (j = 1, \dots, N_i)$ 는 해당 소지역에 속하는 조사구를, $h (h = 1, 2, 3)$ 는 주택유형을 나타낸다. N_i 와 n_i 는 앞서와 마찬가지로 각각 i 번째 지역의 전체 조사구 수와 표본 조사구 수를 나타낸다. 또한, x_{ijh} 는 행정자료, v_h 는 주택유형을 나타내며 이러한 보조변수로 구성된 벡터를 x_{ijh} 로 표기하기로 한다. 보조변수 v_h 는 다음과 같이 정의하였다.

$$v_h = \begin{cases} 1 & \text{if } h = \text{단독주택} \\ 2 & \text{if } h = \text{아파트} \\ 3 & \text{if } h = \text{그 외} \end{cases}$$

- 이 절에서는 지역·주택유형별 주택수가 관심 모수이며, 이를 Y_{ih} 로 표기한다. 즉, $Y_{ih} = \sum_{j=1}^{N_i} y_{ijh}$ 이며, 이는 지역이 i , 주택유형이 h 인 그룹에 속하는 주택수를 의미한다.
- 관심 모수 Y_{ih} 의 직접 추정량, 비 추정량, EBLUP 추정량, 계층적 베이지 추정량은 앞 절에서 소개한 공식과 동일한 형태를 가진다. 마찬가지로 네 가지 추정량의 비교를 위한 요약치 또한 동일하다.
- 다음은 제안한 추정량들을 비교하기 위한 요약치 결과를 표로 나타낸 것이다.

<표 4-12> '시군구'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.1368	0.0602	841.30	1361756.52
Ratio	0.0837	0.0366	429.58	1282160.88
EBLUP	0.1428	0.0695	587.11	1109101.39
HB	0.1827	0.1148	697.06	1500925.94

<표 4-13> '읍'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.2987	0.1688	310.83	217863.01
Ratio	0.1721	0.1131	82.13	16791.25
EBLUP	0.1262	0.0527	94.34	15003.57
HB	0.2145	0.2274	101.26	16289.76

<표 4-14> '면'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.3440	0.3420	80.39	24589.12
Ratio	0.2637	0.3010	41.69	15917.86
EBLUP	0.2087	0.1512	22.62	1625.60
HB	0.5151	3.0361	32.85	3742.03

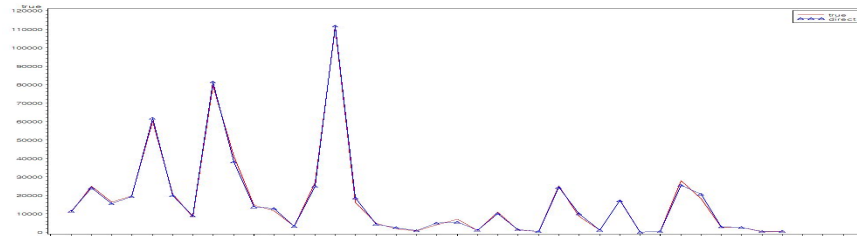
<표 4-15> '동'에 대한 추정량 비교 요약치 결과

	ARB	ASRB	AAB	ASD
Direct	0.2761	0.1646	261.12	190748.62
Ratio	0.1836	0.4801	138.37	549613.07
EBLUP	0.7230	39.4885	101.37	43564.44
HB	0.7653	61.3849	121.12	54832.53

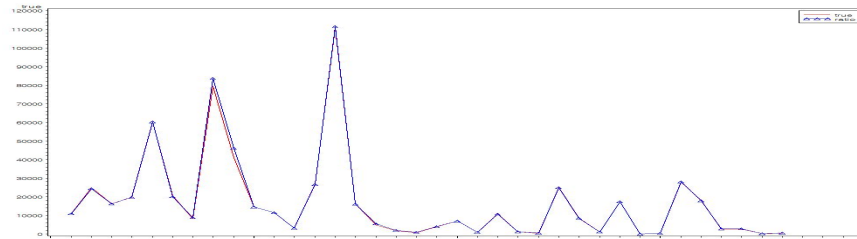
- 위의 표를 보면, 지역·주택유형별 주택 수에 대해서는 대체로 EBLUP 추정량이 실제 값에 가장 가까이 추정된 반면, '시군구'에 대해서는 비 추정량이 대체로 실제 값에 가장 가깝게 추정되었음을 확인할 수 있다.

<그림 4-8> 각 시군구별 센서스 값과 추정값 비교

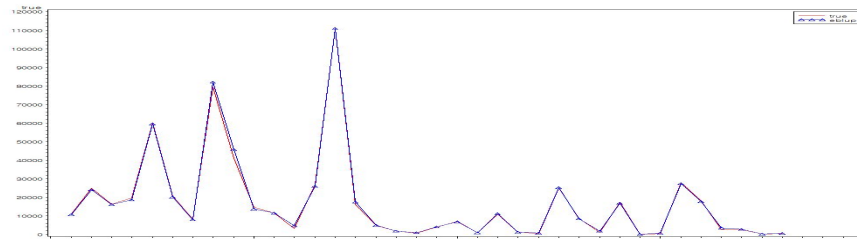
Direct



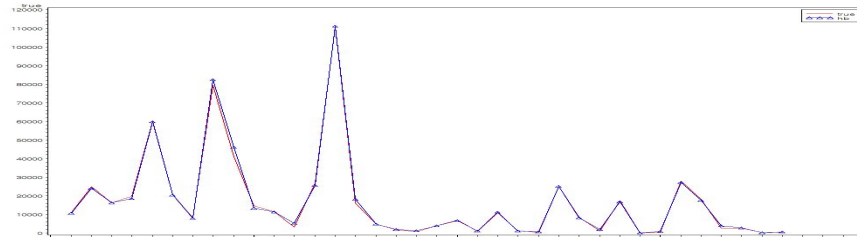
Ratio



EBLUP



HB



** 센서스 값 : — , 추정값 : —

제5절 향후 연구 방안

- 이 장에서 지역·성·연령별 인구수, 지역·가구원수별 가구수 및 지역·주택유형별 주택수를 산하기 위해 소지역 추정방법을 사용하였다. 특히 지역으로 읍면동을 고려할 경우 이들 소지역 추정값을 구하는 과정에서 간과해서는 안되는 점이 있다. 일반적으로 대지역 수준에서는 소지역과는 달리 직접 추정량의 신뢰도가 상당히 높다고 할 수 있다. 그러나 소지역 추정량을 구하여 대지역의 범위에 해당하는 소지역 추정값을 합한 값은 신뢰할 만한 직접 조사추정값과 다소 차이가 있을 수 있다.
- 따라서 대지역에 해당하는 소지역 추정값을 합했을 때 대지역의 직접 추정량과 같은 값을 갖도록 조정해줄 필요가 있는데 이를 벤치마킹 방법이라 한다.
- 능형회귀추정 방법은 회귀추정량의 벤치마킹 혹은 Calibration 성질에 의해 회귀추정량의 보조정보인 20% 표본조사 결과와 일치하게 된다. 즉, 능형회귀추정은 일종의 벤치마킹 추정이라 할 수 있다.
- 따라서 구한 역·성·연령별 인구수, 지역·가구원수별 가구수 및 지역·주택유형별 주택수에 대한 소지역 추정값을 벤치마킹 시킨 후 능형회귀추정값과 비교해 보는 것이 의미가 있을 것이다.

[부 록 4-1] 인구수 추정에 대한 주변분포 (시군구)

1. 인구수 추정에 대한 주변분포

1) 성별 인구 추정량

성별	센서스	행정자료	Direct	Ratio	EBLUP	HB
남	912,176	979,651	916,639	917,568	912,069	902,984
여	929,401	955,547	936,377	934,396	929,393	901,326

2) 연령별 인구 추정량

연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
0-4	102,307	103,537	105,594	103,172	103,227	97,525
5-9	101,002	102,071	101,523	101,447	101,038	96,126
10-14	118,273	119,050	119,441	118,489	117,989	111,963
15-19	109,294	117,033	109,220	110,458	110,150	108,820
20-24	88,075	106,032	88,491	89,288	89,868	96,471
25-29	137,747	143,720	138,381	138,154	138,146	133,505
30-34	161,046	167,763	163,605	161,489	161,295	155,833
35-39	173,479	181,989	176,763	174,053	173,638	168,808
40-44	160,297	172,345	162,673	161,533	161,001	159,201
45-49	150,503	159,264	149,681	152,185	150,939	147,660
50-54	139,204	149,814	138,561	139,948	139,063	138,498
55-59	103,395	109,770	103,872	103,780	103,082	101,989
60-64	85,194	88,531	84,548	85,208	84,283	82,702
65-69	71,433	76,235	71,627	71,888	71,126	71,009
70-74	62,243	62,298	62,005	62,920	61,668	59,054
75-79	42,388	40,455	42,287	41,637	40,873	38,769
80-84	22,700	21,963	22,467	23,311	22,446	21,368
85-89	9,635	9,602	9,298	9,711	9,150	9,564
90-94	2,802	3,034	2,506	2,768	2,265	3,335
95-99	498	568	424	481	217	1,231
100+	62	124	50	42	0	878

3) 지역별 인구 추정량

시도	시군구	센서스	행정자료	Direct	Ratio	EBLUP	HB
11	180	225,999	234,435	223,334	224,377	222,878	219,599
11	190	366,957	380,144	369,494	371,035	369,339	357,109
23	080	382,087	391,165	387,692	383,968	382,064	367,606
24	010	99,451	96,598	100,556	100,469	98,979	95,336
31	240	463,403	481,346	467,448	463,398	461,624	450,955
32	380	17,529	20,391	18,660	18,897	18,344	18,720
33	390	30,175	32,344	30,289	30,282	29,249	29,797
35	350	22,909	29,029	21,977	23,327	23,692	24,935
36	040	72,646	86,003	74,157	74,574	74,544	76,097
36	480	32,023	42,865	31,603	31,891	32,219	34,930
37	030	120,789	130,767	119,908	121,535	120,263	120,845
37	430	7,609	10,111	7,900	8,212	8,268	8,381

4) 성*연령(5세단위)별 인구 추정량

성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
남	0-4	52,678	53,328	54,189	53,139	53,047	49,626
	5-9	52,124	52,756	52,453	52,422	52,100	49,115
	10-14	60,882	61,526	61,765	61,303	60,923	57,349
	15-19	56,195	61,453	56,017	56,959	56,680	56,427
	20-24	39,382	55,778	39,502	39,805	40,338	48,945
	25-29	69,631	74,307	69,058	69,184	69,226	68,101
	30-34	81,728	86,562	82,143	81,638	81,488	79,624
	35-39	89,845	96,175	91,565	90,528	90,271	88,474
	40-44	83,521	91,801	85,142	84,032	83,860	84,000
	45-49	77,521	83,896	77,337	78,155	77,622	77,020
	50-54	69,652	76,477	69,036	70,120	69,646	70,104
	55-59	50,602	54,217	50,657	51,096	50,658	50,031
	60-64	41,274	43,303	41,181	41,206	40,772	40,134
	65-69	33,065	35,401	33,054	33,548	33,088	32,809
	70-74	26,794	26,766	26,761	27,101	26,513	25,244
	75-79	16,420	15,452	16,421	16,107	15,732	14,781
	80-84	7,332	6,940	7,232	7,684	7,248	6,782
	85-89	2,769	2,685	2,544	2,792	2,485	2,699
	90-94	659	700	507	667	372	874
	95-99	94	107	68	76	1	447
	100+	8	21	6	5	0	398

여	0-4	49,629	50,209	51,405	50,033	50,180	47,899
	5-9	48,878	49,315	49,071	49,025	48,939	47,011
	10-14	57,391	57,524	57,676	57,186	57,066	54,614
	15-19	53,099	55,580	53,203	53,499	53,470	52,393
	20-24	48,693	50,254	48,989	49,483	49,530	47,525
	25-29	68,116	69,413	69,323	68,970	68,920	65,403
	30-34	79,318	81,201	81,462	79,851	79,807	76,209
	35-39	83,634	85,814	85,198	83,525	83,366	80,334
	40-44	76,776	80,544	77,531	77,501	77,141	75,201
	45-49	72,982	75,368	72,344	74,030	73,317	70,640
	50-54	69,552	73,337	69,525	69,829	69,417	68,394
	55-59	52,793	55,553	53,215	52,684	52,424	51,958
	60-64	43,920	45,228	43,367	44,002	43,511	42,568
	65-69	38,368	40,834	38,573	38,340	38,038	38,201
	70-74	35,449	35,532	35,244	35,819	35,154	33,810
	75-79	25,968	25,003	25,866	25,529	25,141	23,988
	80-84	15,368	15,023	15,235	15,627	15,198	14,586
	85-89	6,866	6,917	6,753	6,919	6,665	6,865
	90-94	2,143	2,334	1,999	2,100	1,893	2,461
	95-99	404	461	357	406	216	784
	100+	54	103	44	38	0	479

5) 지역*성별 인구추정량

시 도	시군구	성별	센서스	행정자료	Direct	Ratio	EBLUP	HB
11	180	남	112,905	119,675	111,030	112,114	111,304	110,771
		여	113,094	114,760	112,304	112,263	111,573	108,828
11	190	남	179,978	190,781	181,127	181,726	180,945	177,666
		여	186,979	189,363	188,367	189,309	188,394	179,443
23	080	남	190,473	198,373	193,176	191,584	190,458	185,006
		여	191,614	192,792	194,516	192,384	191,606	182,600
24	010	남	48,444	48,164	48,890	49,176	48,386	46,981
		여	51,007	48,434	51,666	51,293	50,593	48,355
31	240	남	234,810	246,713	236,183	234,446	233,379	229,192
		여	228,593	234,633	231,265	228,952	228,244	221,762
32	380	남	8,679	10,495	9,228	9,305	9,028	9,270
		여	8,850	9,896	9,432	9,592	9,316	9,450
33	390	남	14,985	16,580	15,087	14,838	14,265	14,469
		여	15,190	15,764	15,201	15,443	14,984	15,328
35	350	남	10,635	14,420	10,194	11,007	11,263	12,121
		여	12,274	14,609	11,783	12,320	12,429	12,814
36	040	남	34,018	43,025	34,735	35,083	35,181	37,152
		여	38,628	42,978	39,422	39,490	39,363	38,945
36	480	남	15,445	21,712	15,136	15,315	15,485	17,332
		여	16,578	21,153	16,468	16,575	16,734	17,598
37	030	남	57,944	64,513	57,889	58,773	58,169	58,810
		여	62,845	66,254	62,019	62,762	62,093	62,035
37	430	남	3,860	5,200	3,965	4,200	4,206	4,213
		여	3,749	4,911	3,935	4,012	4,063	4,169

6) 지역*연령별 인구추정량 - (1)

시도	시군구	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
11	180	0-4	8,879	9,051	8,769	8,862	8,829	8,593
		5-9	9,353	9,394	9,188	9,221	9,188	8,913
		10-14	12,679	12,609	12,849	12,568	12,545	11,944
		15-19	14,514	15,375	14,377	14,627	14,538	14,369
		20-24	12,746	14,603	12,716	12,850	12,839	13,443
		25-29	20,560	20,574	19,887	20,157	20,002	19,302
		30-34	19,774	20,146	19,090	19,609	19,419	18,858
		35-39	19,691	20,600	19,445	19,355	19,271	19,131
		40-44	19,489	20,960	19,400	19,452	19,358	19,405
		45-49	20,111	21,011	19,862	20,203	20,020	19,587
		50-54	19,448	20,496	18,723	19,055	18,927	18,973
		55-59	14,765	15,322	14,943	14,600	14,547	14,288
		60-64	11,920	12,060	11,882	11,938	11,825	11,356
		65-69	9,284	9,592	9,079	9,090	9,020	8,967
		70-74	6,355	6,304	6,532	6,294	6,251	5,994
		75-79	3,612	3,548	3,749	3,676	3,630	3,443
		80-84	1,779	1,731	1,865	1,742	1,718	1,716
		85-89	741	761	705	772	722	802
		90-94	255	248	225	275	225	320
		95-99	36	38	43	28	4	112
		100+	8	12	5	5	0	84
11	190	0-4	17,662	17,829	18,965	18,014	18,144	16,922
		5-9	15,884	15,974	16,054	16,353	16,264	15,202
		10-14	18,419	18,403	19,045	18,514	18,551	17,426
		15-19	20,462	21,118	20,275	20,729	20,615	19,852
		20-24	19,722	21,525	19,078	20,026	19,891	20,040
		25-29	33,805	33,197	33,742	33,952	33,757	31,275
		30-34	36,645	37,152	37,642	36,939	36,900	34,802
		35-39	35,286	36,857	36,620	35,791	35,770	34,378
		40-44	29,961	31,977	30,628	30,261	30,200	29,716
		45-49	28,627	30,216	28,319	29,307	29,044	28,195
		50-54	29,889	31,745	28,724	29,843	29,635	29,456
		55-59	23,301	24,750	23,221	23,326	23,224	23,021
		60-64	19,023	19,962	18,561	19,056	18,911	18,644
		65-69	14,853	15,782	15,046	15,190	15,091	14,802
		70-74	11,146	11,308	11,607	11,511	11,440	10,768
		75-79	6,756	6,638	6,482	6,532	6,462	6,363
		80-84	3,387	3,451	3,447	3,550	3,469	3,424
		85-89	1,529	1,596	1,503	1,517	1,467	1,665
		90-94	488	520	457	513	453	670
		95-99	100	110	66	94	49	282
		100+	12	34	12	17	0	207

6) 지역*연령별 인구추정량 - (2)

시도	시군구	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
23	080	0-4	21,378	21,041	22,330	21,458	21,530	19,906
		5-9	23,499	23,497	23,635	23,728	23,614	22,142
		10-14	30,331	30,147	30,449	30,219	30,104	28,287
		15-19	29,413	30,744	28,814	29,738	29,505	28,701
		20-24	19,212	23,252	19,614	19,587	19,641	21,441
		25-29	25,677	26,122	27,079	25,553	25,648	24,473
		30-34	30,306	30,688	31,665	30,563	30,555	28,747
		35-39	38,164	38,540	39,219	37,940	37,899	35,962
		40-44	39,782	40,929	39,783	39,820	39,579	38,099
		45-49	37,971	37,922	37,416	38,266	37,848	35,470
		50-54	29,942	30,792	30,241	30,080	29,911	28,742
		55-59	18,092	18,517	18,829	18,431	18,349	17,430
		60-64	12,535	12,815	13,126	12,608	12,570	12,122
		65-69	9,223	9,711	9,298	9,202	9,134	9,198
		70-74	7,379	7,290	7,029	7,573	7,363	7,033
		75-79	4,902	4,791	4,963	4,736	4,684	4,679
		80-84	2,692	2,686	2,704	2,957	2,812	2,754
		85-89	1,189	1,238	1,187	1,175	1,103	1,370
		90-94	326	366	269	287	214	554
		95-99	68	67	44	49	2	277
		100+	6	10	0	0	0	219
24	010	0-4	2,785	2,873	3,015	2,790	2,820	2,873
		5-9	3,438	3,459	3,529	3,352	3,373	3,416
		10-14	5,064	5,001	5,227	5,125	5,097	4,920
		15-19	6,440	5,910	6,540	6,441	6,370	5,878
		20-24	8,803	6,623	8,982	8,948	8,721	7,012
		25-29	9,151	8,367	9,409	9,515	9,312	8,331
		30-34	6,859	7,208	7,127	6,780	6,772	6,839
		35-39	6,521	6,819	6,663	6,521	6,485	6,505
		40-44	6,362	6,799	6,548	6,480	6,439	6,480
		45-49	6,976	7,268	7,054	7,147	7,053	6,964
		50-54	8,059	8,119	7,965	8,294	8,124	7,832
		55-59	7,034	6,948	7,012	7,139	7,010	6,739
		60-64	6,456	6,377	6,321	6,377	6,282	6,155
		65-69	5,542	5,525	5,357	5,601	5,477	5,368
		70-74	4,579	4,304	4,409	4,631	4,493	4,283
		75-79	2,996	2,649	3,164	2,998	2,966	2,751
		80-84	1,541	1,489	1,417	1,454	1,413	1,551
		85-89	648	636	603	697	642	770
		90-94	165	178	173	136	108	313
		95-99	27	40	42	41	24	195
		100+	5	6	0	0	0	160

6) 지역*연령별 인구추정량 - (3)

시도	시군구	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
31	240	0-4	39,703	39,594	40,080	39,657	39,548	37,217
		5-9	35,704	35,940	35,621	35,439	35,321	33,727
		10-14	34,198	34,546	34,227	34,275	34,118	32,452
		15-19	23,715	24,976	24,231	23,610	23,627	23,373
		20-24	17,848	20,828	18,529	18,333	18,382	19,336
		25-29	34,822	34,529	34,572	34,863	34,619	32,481
		30-34	51,377	51,837	51,278	50,760	50,577	48,374
		35-39	54,942	56,262	55,254	54,915	54,685	52,449
		40-44	43,712	45,887	44,882	44,144	44,027	42,728
		45-49	33,409	35,382	33,029	33,255	33,094	32,877
		50-54	26,619	29,161	27,931	26,859	26,861	27,053
		55-59	18,197	20,097	18,465	18,209	18,168	18,678
		60-64	14,319	15,597	14,457	14,165	14,131	14,553
		65-69	11,550	12,787	11,727	11,638	11,583	11,972
		70-74	10,281	10,736	9,815	9,904	9,832	10,104
		75-79	7,114	7,137	7,419	7,210	7,161	6,885
		80-84	3,683	3,725	3,858	3,946	3,854	3,723
		85-89	1,622	1,682	1,577	1,668	1,589	1,780
		90-94	485	520	409	435	383	684
		95-99	96	110	81	107	62	304
		100+	7	13	7	6	0	205
32	380	0-4	986	1,003	1,102	1,017	1,003	967
		5-9	1,045	1,076	1,252	1,135	1,115	1,053
		10-14	1,145	1,194	1,358	1,286	1,247	1,172
		15-19	755	1,015	774	856	866	918
		20-24	381	1,294	405	376	492	902
		25-29	912	1,455	1,020	944	992	1,200
		30-34	1,134	1,298	1,331	1,186	1,174	1,192
		35-39	1,237	1,413	1,465	1,335	1,317	1,314
		40-44	1,280	1,523	1,462	1,408	1,385	1,403
		45-49	1,497	1,665	1,645	1,706	1,627	1,585
		50-54	1,605	1,741	1,602	1,857	1,723	1,670
		55-59	1,148	1,223	1,157	1,226	1,169	1,153
		60-64	861	919	774	850	821	842
		65-69	1,098	1,234	1,048	1,165	1,115	1,136
		70-74	1,091	1,063	984	1,186	1,069	1,026
		75-79	718	604	697	695	644	594
		80-84	399	394	381	389	365	363
		85-89	176	196	158	208	178	178
		90-94	55	68	43	69	44	51
		95-99	5	8	3	3	0	1
		100+	1	5	0	0	0	0

6) 지역*연령별 인구추정량 - (4)

시도	시군구	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
33	390	0-4	1,781	1,786	1,473	1,962	1,794	1,741
		5-9	1,706	1,703	1,618	1,788	1,706	1,652
		10-14	2,121	2,135	1,945	2,148	2,064	2,036
		15-19	1,873	2,004	1,835	1,867	1,843	1,874
		20-24	1,341	1,991	1,511	1,132	970	1,554
		25-29	2,082	2,405	2,041	2,063	2,056	2,187
		30-34	2,295	2,505	2,079	2,359	2,287	2,328
		35-39	2,541	2,663	2,254	2,540	2,447	2,476
		40-44	2,442	2,792	2,417	2,565	2,511	2,571
		45-49	2,366	2,483	2,463	2,205	2,194	2,262
		50-54	2,241	2,446	2,303	2,183	2,159	2,228
		55-59	1,791	1,857	1,942	1,752	1,723	1,715
		60-64	1,328	1,331	1,454	1,369	1,321	1,256
		65-69	1,286	1,418	1,489	1,289	1,273	1,288
		70-74	1,265	1,238	1,504	1,291	1,260	1,170
		75-79	957	913	1,127	1,033	982	876
		80-84	525	453	613	524	496	427
		85-89	171	162	183	157	136	126
		90-94	54	52	27	48	27	30
35	350	95-99	9	7	11	8	0	0
		100+	0	0	0	0	0	0
		0-4	707	998	687	739	805	839
		5-9	815	928	784	775	804	791
		10-14	1,060	1,155	1,066	1,102	1,091	1,024
		15-19	688	1,156	642	731	832	950
		20-24	443	1,448	442	478	717	1,107
		25-29	608	1,548	556	637	881	1,223
		30-34	763	1,467	702	768	924	1,182
		35-39	950	1,505	1,083	1,107	1,158	1,273
		40-44	1,208	1,815	1,171	1,285	1,370	1,540
		45-49	1,480	2,048	1,500	1,650	1,670	1,779
		50-54	1,733	2,298	1,622	1,705	1,768	1,970
		55-59	1,670	1,939	1,450	1,659	1,644	1,701
		60-64	2,088	2,164	1,911	2,037	1,964	1,939
		65-69	2,163	2,345	2,158	2,247	2,154	2,112
		70-74	2,651	2,606	2,522	2,636	2,492	2,382
		75-79	2,072	1,855	2,016	2,029	1,894	1,714
		80-84	1,209	1,118	1,080	1,089	1,029	978
		85-89	434	433	443	412	381	340
		90-94	141	172	114	224	116	92
		95-99	21	25	18	15	0	0
		100+	5	6	9	2	0	0

6) 지역*연령별 인구추정량 - (5)

시도	시군구	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
36	040	0-4	2,577	2,806	2,859	2,682	2,703	2,554
		5-9	3,147	3,326	3,258	3,273	3,261	3,064
		10-14	4,147	4,331	4,328	4,197	4,196	3,989
		15-19	3,536	4,335	3,742	3,674	3,687	3,814
		20-24	2,527	4,266	2,373	2,471	2,596	3,398
		25-29	3,026	4,729	3,123	3,244	3,332	3,931
		30-34	3,438	4,741	3,523	3,642	3,696	4,066
		35-39	4,078	5,364	4,362	4,378	4,404	4,682
		40-44	4,479	5,970	4,576	4,523	4,547	5,092
		45-49	5,029	6,686	5,214	5,262	5,287	5,781
		50-54	5,680	7,295	5,629	5,863	5,881	6,357
		55-59	5,039	5,824	5,051	5,172	5,146	5,213
		60-64	5,298	5,749	5,084	5,277	5,226	5,195
		65-69	5,648	6,341	5,917	5,858	5,820	5,748
		70-74	5,981	5,944	6,107	6,137	6,045	5,589
		75-79	4,738	4,277	4,798	4,571	4,540	4,059
		80-84	2,768	2,558	2,732	2,822	2,743	2,410
		85-89	1,144	1,037	1,127	1,190	1,136	937
		90-94	308	345	309	290	271	219
		95-99	47	57	34	40	28	0
		100+	11	22	12	9	0	0
36	480	0-4	654	1,092	638	605	656	775
		5-9	862	1,117	833	807	837	850
		10-14	1,183	1,532	1,225	1,201	1,229	1,236
		15-19	675	1,708	679	634	705	1,171
		20-24	415	2,011	417	417	587	1,322
		25-29	744	2,317	684	690	871	1,605
		30-34	955	1,959	859	863	974	1,418
		35-39	1,344	2,189	1,329	1,314	1,379	1,693
		40-44	1,775	2,650	1,691	1,590	1,643	2,062
		45-49	2,397	3,236	2,379	2,347	2,388	2,672
		50-54	2,887	3,782	2,847	2,924	2,959	3,201
		55-59	2,788	3,296	2,722	2,755	2,751	2,829
		60-64	3,233	3,497	3,211	3,312	3,243	3,116
		65-69	3,558	3,911	3,523	3,606	3,545	3,471
		70-74	3,742	3,942	3,984	4,094	3,960	3,626
		75-79	2,669	2,555	2,570	2,637	2,548	2,303
		80-84	1,382	1,300	1,284	1,369	1,297	1,115
		85-89	573	565	560	564	529	405
		90-94	156	172	144	122	100	61
		95-99	29	28	28	40	18	0
		100+	2	6	0	0	0	0

6) 지역*연령별 인구추정량 - (6)

시도	시군구	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
37	030	0-4	4,940	5,166	5,393	5,115	5,119	4,883
		5-9	5,260	5,350	5,413	5,270	5,250	5,043
		10-14	7,553	7,595	7,274	7,463	7,356	7,119
		15-19	6,955	8,282	7,021	7,265	7,256	7,601
		20-24	4,500	7,665	4,264	4,513	4,795	6,586
		25-29	6,051	7,815	5,938	6,212	6,288	7,027
		30-34	7,068	8,096	7,872	7,538	7,514	7,491
		35-39	8,205	9,081	8,464	8,276	8,237	8,354
		40-44	9,191	10,273	9,436	9,330	9,273	9,435
		45-49	9,965	10,437	10,121	10,082	9,958	9,707
		50-54	10,278	10,916	10,140	10,354	10,210	10,106
		55-59	8,892	9,125	8,440	8,776	8,618	8,471
		60-64	7,518	7,331	7,201	7,646	7,408	6,921
		65-69	6,690	6,949	6,449	6,455	6,376	6,405
		70-74	7,284	7,034	7,036	7,093	6,947	6,593
		75-79	5,553	5,148	4,995	5,229	5,079	4,829
		80-84	3,138	2,844	2,881	3,233	3,041	2,729
		85-89	1,336	1,210	1,186	1,269	1,197	1,142
		90-94	350	366	326	354	312	338
		95-99	57	74	53	56	28	60
		100+	5	10	5	4	0	5
37	430	0-4	255	298	283	269	275	253
		5-9	289	307	340	307	307	273
		10-14	373	402	448	391	391	359
		15-19	268	410	290	287	308	319
		20-24	137	526	160	158	237	329
		25-29	309	662	330	325	388	470
		30-34	432	666	438	482	503	536
		35-39	520	696	605	583	585	592
		40-44	616	770	680	676	669	669
		45-49	675	910	680	756	757	781
		50-54	823	1,023	835	931	904	910
		55-59	678	872	640	736	732	751
		60-64	615	729	568	574	580	603
		65-69	538	640	538	546	539	543
		70-74	489	529	478	571	516	485
		75-79	301	340	308	290	284	274
		80-84	197	214	205	233	208	180
		85-89	72	86	65	83	71	50
		90-94	19	27	13	14	12	4
		95-99	3	4	0	0	0	0
		100+	0	0	0	0	0	0

7) 지역*성별*연령별 인구 추정량 - (1)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
11	180	남	0-4	4,524	4,614	4,451	4,503	4,475	4,303
			5-9	4,812	4,791	4,638	4,680	4,650	4,469
			10-14	6,604	6,578	6,685	6,602	6,568	6,171
			15-19	7,488	8,114	7,349	7,444	7,409	7,460
			20-24	5,631	7,606	5,675	5,736	5,778	6,741
			25-29	10,692	10,928	10,135	10,449	10,349	10,138
			30-34	10,637	10,861	10,317	10,567	10,459	10,110
			35-39	10,300	10,929	10,192	10,178	10,126	10,079
			40-44	9,864	10,935	9,798	9,821	9,782	10,008
			45-49	9,995	10,700	10,028	9,962	9,913	9,868
			50-54	9,590	10,332	9,193	9,414	9,351	9,486
			55-59	7,144	7,512	6,941	7,142	7,067	6,965
			60-64	5,820	5,942	5,979	5,834	5,801	5,557
			65-69	4,595	4,724	4,486	4,540	4,491	4,402
			70-74	2,966	2,923	2,940	3,027	2,968	2,779
			75-79	1,478	1,450	1,509	1,455	1,435	1,386
			80-84	518	478	498	502	479	481
			85-89	190	201	177	212	184	220
			90-94	50	49	29	42	20	76
			95-99	6	5	10	5	0	37
			100+	1	3	0	0	0	35
		여	0-4	4,355	4,437	4,317	4,360	4,354	4,290
			5-9	4,541	4,603	4,550	4,541	4,538	4,443
			10-14	6,075	6,031	6,164	5,966	5,977	5,774
			15-19	7,026	7,261	7,029	7,183	7,129	6,909
			20-24	7,115	6,997	7,041	7,113	7,061	6,702
			25-29	9,868	9,646	9,752	9,708	9,653	9,164
			30-34	9,137	9,285	8,773	9,043	8,959	8,748
			35-39	9,391	9,671	9,253	9,177	9,145	9,052
			40-44	9,625	10,025	9,602	9,631	9,576	9,397
			45-49	10,116	10,311	9,834	10,241	10,107	9,719
			50-54	9,858	10,164	9,530	9,641	9,577	9,487
			55-59	7,621	7,810	8,002	7,457	7,480	7,323
			60-64	6,100	6,118	5,903	6,103	6,024	5,798
			65-69	4,689	4,868	4,593	4,550	4,529	4,565
			70-74	3,389	3,381	3,592	3,268	3,283	3,215
			75-79	2,134	2,098	2,240	2,222	2,195	2,057
			80-84	1,261	1,253	1,367	1,240	1,239	1,235
			85-89	551	560	528	560	538	582
			90-94	205	199	196	233	205	244
			95-99	30	33	34	23	4	75
			100+	7	9	5	5	0	48

7) 지역*성별*연령별 인구 추정량 - (2)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
11	190	남	0-4	9,098	9,135	9,964	9,355	9,428	8,616
			5-9	8,208	8,302	8,412	8,521	8,467	7,830
			10-14	9,449	9,489	9,727	9,674	9,642	8,933
			15-19	10,615	11,198	10,552	10,882	10,799	10,436
			20-24	8,348	10,919	8,020	8,370	8,429	9,820
			25-29	16,453	16,471	16,158	16,254	16,168	15,364
			30-34	18,268	18,878	18,452	18,220	18,178	17,534
			35-39	18,508	19,595	19,182	18,722	18,706	18,168
			40-44	15,430	16,861	15,918	15,564	15,549	15,551
			45-49	14,146	15,421	14,215	14,468	14,387	14,268
			50-54	14,140	15,426	13,226	14,032	13,924	14,202
			55-59	11,128	11,918	11,326	11,139	11,111	11,040
			60-64	9,287	9,732	9,009	9,281	9,199	9,059
			65-69	7,198	7,685	7,108	7,422	7,334	7,184
			70-74	5,188	5,284	5,451	5,361	5,329	5,011
			75-79	2,800	2,725	2,683	2,631	2,606	2,601
			80-84	1,146	1,126	1,165	1,255	1,200	1,148
			85-89	439	459	415	428	399	504
			90-94	108	122	120	122	92	196
			95-99	17	26	18	20	0	107
			100+	4	9	6	5	0	92
		여	0-4	8,564	8,694	9,001	8,659	8,716	8,305
			5-9	7,676	7,672	7,642	7,832	7,797	7,373
			10-14	8,970	8,914	9,318	8,840	8,910	8,493
			15-19	9,847	9,920	9,723	9,847	9,815	9,416
			20-24	11,374	10,606	11,058	11,656	11,463	10,220
			25-29	17,352	16,726	17,584	17,698	17,589	15,911
			30-34	18,377	18,274	19,190	18,719	18,723	17,268
			35-39	16,778	17,262	17,438	17,068	17,064	16,210
			40-44	14,531	15,116	14,710	14,696	14,652	14,165
			45-49	14,481	14,795	14,104	14,839	14,657	13,927
			50-54	15,749	16,319	15,498	15,811	15,712	15,254
			55-59	12,173	12,832	11,895	12,187	12,113	11,981
			60-64	9,736	10,230	9,552	9,776	9,712	9,585
			65-69	7,655	8,097	7,938	7,768	7,757	7,618
			70-74	5,958	6,024	6,156	6,150	6,111	5,757
			75-79	3,956	3,913	3,799	3,901	3,856	3,761
			80-84	2,241	2,325	2,282	2,295	2,269	2,276
			85-89	1,090	1,137	1,088	1,089	1,068	1,161
			90-94	380	398	337	391	361	473
			95-99	83	84	48	75	49	175
			100+	8	25	6	13	0	114

7) 지역*성별*연령별 인구 추정량 - (3)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
23	080	남	0-4	11,068	10,868	11,625	11,096	11,123	10,187
			5-9	12,113	12,124	12,250	12,330	12,239	11,344
			10-14	15,584	15,524	15,808	15,441	15,405	14,466
			15-19	15,062	16,140	14,684	15,222	15,099	14,936
			20-24	8,498	12,326	8,610	8,618	8,703	11,074
			25-29	12,841	13,419	13,513	12,737	12,765	12,450
			30-34	14,914	15,252	15,420	15,127	15,068	14,213
			35-39	18,937	19,324	19,579	19,092	19,032	17,978
			40-44	19,949	20,660	20,176	19,872	19,789	19,148
			45-49	19,724	19,813	19,128	19,778	19,524	18,447
			50-54	15,782	16,345	15,867	15,934	15,812	15,194
			55-59	9,350	9,591	9,621	9,651	9,557	8,991
			60-64	6,312	6,496	6,636	6,179	6,179	6,080
			65-69	4,399	4,607	4,483	4,513	4,455	4,353
			70-74	2,997	3,015	2,951	3,033	2,969	2,890
			75-79	1,762	1,690	1,667	1,743	1,683	1,667
			80-84	762	746	791	878	808	801
			85-89	310	319	287	264	225	386
			90-94	93	97	68	65	22	184
			95-99	15	13	10	13	1	111
			100+	1	4	0	0	0	104
		여	0-4	10,310	10,173	10,705	10,363	10,407	9,719
			5-9	11,386	11,373	11,386	11,399	11,375	10,799
			10-14	14,747	14,623	14,641	14,778	14,699	13,820
			15-19	14,351	14,604	14,130	14,516	14,406	13,765
			20-24	10,714	10,926	11,003	10,968	10,939	10,367
			25-29	12,836	12,703	13,566	12,816	12,883	12,023
			30-34	15,392	15,436	16,245	15,436	15,487	14,534
			35-39	19,227	19,216	19,639	18,848	18,867	17,984
			40-44	19,833	20,269	19,607	19,948	19,790	18,951
			45-49	18,247	18,109	18,288	18,488	18,323	17,024
			50-54	14,160	14,447	14,375	14,145	14,100	13,548
			55-59	8,742	8,926	9,208	8,780	8,791	8,440
			60-64	6,223	6,319	6,490	6,429	6,391	6,042
			65-69	4,824	5,104	4,815	4,689	4,678	4,845
			70-74	4,382	4,275	4,077	4,541	4,394	4,142
			75-79	3,140	3,101	3,296	2,993	3,000	3,012
			80-84	1,930	1,940	1,913	2,079	2,004	1,953
			85-89	879	919	900	910	878	983
			90-94	233	269	201	222	192	369
			95-99	53	54	34	36	1	165
			100+	5	6	0	0	0	114

7) 지역*성별*연령별 인구 추정량 - (4)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
24	010	남	0-4	1,465	1,506	1,598	1,427	1,442	1,463
			5-9	1,791	1,829	1,697	1,746	1,731	1,764
			10-14	2,587	2,572	2,634	2,637	2,606	2,501
			15-19	3,180	2,960	3,291	3,277	3,224	2,931
			20-24	4,031	3,374	4,132	4,121	4,029	3,432
			25-29	4,825	4,450	4,962	4,952	4,862	4,378
			30-34	3,670	3,888	3,737	3,655	3,633	3,660
			35-39	3,436	3,760	3,560	3,568	3,533	3,550
			40-44	3,189	3,632	3,268	3,309	3,283	3,396
			45-49	3,337	3,602	3,329	3,340	3,310	3,384
			50-54	3,785	3,912	3,729	3,900	3,821	3,737
			55-59	3,290	3,233	3,230	3,390	3,301	3,143
			60-64	2,979	2,946	2,945	2,942	2,899	2,837
			65-69	2,594	2,595	2,577	2,646	2,590	2,520
			70-74	2,155	1,991	2,008	2,138	2,065	1,976
			75-79	1,307	1,141	1,486	1,346	1,343	1,207
			80-84	580	550	499	546	520	589
			85-89	203	186	157	196	171	251
			90-94	35	30	46	36	23	106
			95-99	5	6	4	3	0	80
			100+	0	1	0	0	0	76
		여	0-4	1,320	1,367	1,417	1,363	1,379	1,410
			5-9	1,647	1,630	1,832	1,606	1,642	1,652
			10-14	2,477	2,429	2,592	2,488	2,491	2,420
			15-19	3,260	2,950	3,249	3,163	3,146	2,947
			20-24	4,772	3,249	4,850	4,828	4,693	3,580
			25-29	4,326	3,917	4,447	4,563	4,449	3,953
			30-34	3,189	3,320	3,391	3,125	3,139	3,179
			35-39	3,085	3,059	3,103	2,953	2,952	2,955
			40-44	3,173	3,167	3,280	3,171	3,156	3,083
			45-49	3,639	3,666	3,725	3,807	3,742	3,580
			50-54	4,274	4,207	4,236	4,394	4,303	4,095
			55-59	3,744	3,715	3,783	3,749	3,708	3,597
			60-64	3,477	3,431	3,376	3,435	3,383	3,318
			65-69	2,948	2,930	2,780	2,954	2,887	2,848
			70-74	2,424	2,313	2,400	2,492	2,428	2,307
			75-79	1,689	1,508	1,678	1,652	1,623	1,545
			80-84	961	939	918	909	893	962
			85-89	445	450	445	502	470	519
			90-94	130	148	127	100	85	208
			95-99	22	34	38	38	24	114
			100+	5	5	0	0	0	84

7) 지역*성별*연령별 인구 추정량 - (5)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
31	240	남	0-4	20,324	20,333	20,076	20,285	20,145	18,963
			5-9	18,365	18,531	18,476	18,146	18,107	17,247
			10-14	17,578	17,839	17,573	17,696	17,584	16,636
			15-19	12,293	13,126	12,429	12,204	12,188	12,135
			20-24	8,048	10,729	8,224	8,274	8,311	9,677
			25-29	17,643	17,430	17,374	17,581	17,435	16,292
			30-34	25,883	26,224	25,705	25,441	25,343	24,343
			35-39	28,663	29,724	28,730	28,665	28,526	27,561
			40-44	23,980	25,410	24,709	24,130	24,082	23,519
			45-49	18,155	19,526	18,193	18,175	18,099	18,031
			50-54	13,665	15,109	14,218	13,732	13,717	13,917
			55-59	9,043	10,014	9,366	9,149	9,124	9,257
			60-64	6,935	7,650	6,946	6,880	6,850	7,076
			65-69	5,253	5,918	5,327	5,211	5,188	5,475
			70-74	4,448	4,654	4,298	4,308	4,270	4,356
			75-79	2,785	2,724	2,814	2,717	2,691	2,614
			80-84	1,153	1,161	1,190	1,216	1,174	1,171
			85-89	457	461	468	530	478	522
			90-94	122	126	61	95	66	201
			95-99	17	23	7	12	0	109
			100+	0	1	0	0	0	91
		여	0-4	19,379	19,261	20,005	19,372	19,403	18,255
			5-9	17,339	17,409	17,145	17,293	17,213	16,480
			10-14	16,620	16,707	16,655	16,579	16,534	15,816
			15-19	11,422	11,850	11,803	11,406	11,438	11,239
			20-24	9,800	10,099	10,305	10,059	10,071	9,659
			25-29	17,179	17,099	17,199	17,282	17,185	16,189
			30-34	25,494	25,613	25,573	25,319	25,234	24,031
			35-39	26,279	26,538	26,524	26,250	26,159	24,888
			40-44	19,732	20,477	20,173	20,014	19,944	19,209
			45-49	15,254	15,856	14,836	15,079	14,995	14,846
			50-54	12,954	14,052	13,713	13,127	13,144	13,136
			55-59	9,154	10,083	9,099	9,060	9,044	9,421
			60-64	7,384	7,947	7,510	7,285	7,281	7,477
			65-69	6,297	6,869	6,400	6,428	6,396	6,497
			70-74	5,833	6,082	5,516	5,595	5,562	5,748
			75-79	4,329	4,413	4,604	4,492	4,470	4,271
			80-84	2,530	2,564	2,669	2,730	2,680	2,552
			85-89	1,165	1,221	1,109	1,138	1,111	1,258
			90-94	363	394	348	341	317	483
			95-99	79	87	74	96	62	196
			100+	7	12	7	6	0	114

7) 지역*성별*연령별 인구추정량 - (6)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
32	380	남	0-4	509	516	570	497	494	470
			5-9	555	569	655	612	592	541
			10-14	607	640	725	696	667	612
			15-19	395	525	426	468	462	466
			20-24	199	792	231	200	250	515
			25-29	449	857	460	417	463	641
			30-34	573	704	664	566	569	604
			35-39	653	772	764	707	696	697
			40-44	679	826	758	743	730	742
			45-49	773	870	865	889	847	817
			50-54	853	928	880	1,012	932	887
			55-59	619	644	612	650	616	599
			60-64	361	400	332	382	363	362
			65-69	478	551	448	551	512	508
			70-74	470	474	399	457	431	432
			75-79	288	209	268	267	240	208
			80-84	122	113	85	93	87	92
			85-89	75	73	67	80	66	62
			90-94	20	29	18	17	11	15
			95-99	1	2	0	0	0	0
			100+	0	1	0	0	0	0
		여	0-4	477	487	533	520	509	497
			5-9	490	507	597	523	523	511
			10-14	538	554	633	591	580	560
			15-19	360	490	347	388	404	453
			20-24	182	502	174	176	241	388
			25-29	463	598	560	526	529	559
			30-34	561	594	667	619	605	588
			35-39	584	641	700	627	620	617
			40-44	601	697	704	665	655	662
			45-49	724	795	780	817	780	768
			50-54	752	813	722	845	791	784
			55-59	529	579	545	576	552	553
			60-64	500	519	442	467	459	480
			65-69	620	683	600	614	603	628
			70-74	621	589	585	730	637	594
			75-79	430	395	429	428	404	386
			80-84	277	281	295	296	278	271
			85-89	101	123	91	127	112	115
			90-94	35	39	24	52	34	36
			95-99	4	6	3	3	0	1
			100+	1	4	0	0	0	0

7) 지역*성별*연령별 인구 추정량 - (7)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
33	390	남	0-4	952	949	807	1,011	937	866
			5-9	849	855	784	903	850	778
			10-14	1,052	1,082	1,005	1,090	1,044	984
			15-19	944	1,040	895	929	918	916
			20-24	597	1,185	655	400	230	761
			25-29	1,123	1,366	1,134	1,086	1,096	1,176
			30-34	1,194	1,329	1,066	1,200	1,174	1,185
			35-39	1,341	1,426	1,253	1,303	1,275	1,277
			40-44	1,313	1,520	1,256	1,386	1,351	1,363
			45-49	1,258	1,318	1,378	1,205	1,192	1,180
			50-54	1,120	1,237	1,199	1,095	1,083	1,096
			55-59	897	921	933	881	858	828
			60-64	661	670	739	663	645	603
			65-69	552	617	678	587	574	547
			70-74	568	541	651	544	531	487
			75-79	326	317	388	337	322	281
			80-84	172	149	194	152	142	118
			85-89	52	45	61	55	42	23
			90-94	13	13	8	9	0	0
			95-99	1	0	4	1	0	0
			100+	0	0	0	0	0	0
		여	0-4	829	837	666	951	857	875
			5-9	857	848	834	884	857	873
			10-14	1,069	1,053	940	1,057	1,020	1,052
			15-19	929	964	940	937	924	958
			20-24	744	806	857	731	740	793
			25-29	959	1,039	906	977	959	1,010
			30-34	1,101	1,176	1,013	1,159	1,114	1,144
			35-39	1,200	1,237	1,001	1,237	1,172	1,199
			40-44	1,129	1,272	1,161	1,179	1,160	1,208
			45-49	1,108	1,165	1,085	1,000	1,001	1,082
			50-54	1,121	1,209	1,104	1,089	1,077	1,133
			55-59	894	936	1,009	870	865	887
			60-64	667	661	716	706	676	653
			65-69	734	801	811	702	698	741
			70-74	697	697	853	747	729	683
			75-79	631	596	739	697	661	595
			80-84	353	304	419	372	353	309
			85-89	119	117	122	101	94	104
			90-94	41	39	19	39	27	30
			95-99	8	7	8	7	0	0
			100+	0	0	0	0	0	0

7) 지역*성별*연령별 인구 추정량 - (8)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
35	350	남	0-4	353	528	306	364	402	428
			5-9	449	517	429	430	438	429
			10-14	570	614	611	623	600	541
			15-19	371	624	371	393	434	500
			20-24	232	787	240	261	386	594
			25-29	322	862	277	332	478	672
			30-34	397	877	353	384	493	686
			35-39	524	922	603	621	661	764
			40-44	670	1,058	599	686	753	882
			45-49	778	1,141	803	866	889	975
			50-54	880	1,231	784	844	894	1,036
			55-59	782	958	631	739	755	818
			60-64	937	1,010	859	940	903	895
			65-69	898	987	961	909	884	872
			70-74	1,069	1,037	1,049	1,141	1,047	955
			75-79	781	699	789	796	731	641
			80-84	439	399	401	428	388	345
			85-89	142	121	109	116	98	79
			90-94	35	43	14	129	29	8
			95-99	6	5	5	5	0	0
			100+	0	0	0	0	0	0
		여	0-4	354	470	380	376	403	412
			5-9	366	411	355	345	366	361
			10-14	490	541	455	479	491	483
			15-19	317	532	271	338	398	450
			20-24	211	661	202	217	331	514
			25-29	286	686	279	305	403	551
			30-34	366	590	349	384	431	496
			35-39	426	583	481	486	497	509
			40-44	538	757	573	600	617	657
			45-49	702	907	696	784	781	804
			50-54	853	1,067	838	861	873	934
			55-59	888	981	819	920	889	883
			60-64	1,151	1,154	1,052	1,097	1,061	1,044
			65-69	1,265	1,358	1,197	1,338	1,270	1,240
			70-74	1,582	1,569	1,473	1,495	1,445	1,426
			75-79	1,291	1,156	1,227	1,234	1,162	1,072
			80-84	770	719	679	661	641	633
			85-89	292	312	334	296	282	261
			90-94	106	129	100	95	87	84
			95-99	15	20	14	10	0	0
			100+	5	6	9	2	0	0

7) 지역*성별*연령별 인구 추정량 - (9)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
36	040	남	0-4	1,341	1,469	1,430	1,426	1,420	1,323
			5-9	1,630	1,717	1,695	1,702	1,689	1,566
			10-14	2,105	2,199	2,196	2,142	2,134	2,010
			15-19	1,796	2,267	1,885	1,835	1,841	1,948
			20-24	1,222	2,284	1,214	1,254	1,319	1,781
			25-29	1,605	2,668	1,595	1,684	1,748	2,162
			30-34	1,771	2,707	1,737	1,847	1,898	2,240
			35-39	2,177	3,139	2,266	2,325	2,351	2,656
			40-44	2,370	3,429	2,452	2,393	2,412	2,847
			45-49	2,613	3,781	2,709	2,725	2,754	3,180
			50-54	2,788	3,884	2,793	2,867	2,897	3,294
			55-59	2,391	2,938	2,359	2,440	2,437	2,566
			60-64	2,438	2,769	2,372	2,487	2,463	2,475
			65-69	2,426	2,733	2,537	2,533	2,511	2,466
			70-74	2,331	2,285	2,390	2,349	2,320	2,130
			75-79	1,698	1,540	1,768	1,690	1,673	1,458
			80-84	902	818	938	968	932	770
			85-89	342	317	336	353	333	262
			90-94	64	72	59	62	51	17
			95-99	8	9	3	3	0	0
			100+	0	0	0	0	0	0
		여	0-4	1,236	1,337	1,429	1,256	1,283	1,231
			5-9	1,517	1,609	1,563	1,572	1,572	1,498
			10-14	2,042	2,132	2,132	2,055	2,062	1,979
			15-19	1,740	2,068	1,857	1,839	1,846	1,865
			20-24	1,305	1,982	1,159	1,217	1,277	1,618
			25-29	1,421	2,061	1,528	1,560	1,584	1,769
			30-34	1,667	2,034	1,785	1,795	1,798	1,826
			35-39	1,901	2,225	2,097	2,053	2,052	2,026
			40-44	2,109	2,541	2,124	2,129	2,136	2,245
			45-49	2,416	2,905	2,504	2,536	2,533	2,600
			50-54	2,892	3,411	2,836	2,997	2,983	3,063
			55-59	2,648	2,886	2,692	2,733	2,709	2,646
			60-64	2,860	2,980	2,712	2,790	2,763	2,721
			65-69	3,222	3,608	3,380	3,326	3,309	3,282
			70-74	3,650	3,659	3,717	3,788	3,725	3,459
			75-79	3,040	2,737	3,030	2,881	2,867	2,601
			80-84	1,866	1,740	1,794	1,855	1,811	1,639
			85-89	802	720	791	838	803	675
			90-94	244	273	250	228	220	201
			95-99	39	48	31	37	28	0
			100+	11	22	12	9	0	0

7) 지역*성별*연령별 인구 추정량 - (10)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
36	480	남	0-4	320	551	329	298	311	372
			5-9	453	575	446	459	468	441
			10-14	596	790	623	639	647	633
			15-19	347	868	354	316	332	575
			20-24	269	1,147	256	266	382	769
			25-29	447	1,369	401	392	478	937
			30-34	584	1,244	515	507	569	891
			35-39	833	1,416	814	787	826	1,083
			40-44	1,082	1,676	1,015	965	1,000	1,300
			45-49	1,317	1,869	1,286	1,250	1,280	1,513
			50-54	1,573	2,106	1,535	1,564	1,589	1,763
			55-59	1,394	1,694	1,369	1,377	1,377	1,438
			60-64	1,620	1,776	1,607	1,624	1,598	1,561
			65-69	1,598	1,739	1,531	1,638	1,596	1,540
			70-74	1,481	1,512	1,590	1,642	1,574	1,392
			75-79	944	843	953	976	926	764
			80-84	417	379	365	438	394	304
			85-89	133	124	123	150	125	57
			90-94	30	27	25	24	11	0
			95-99	7	7	4	4	0	0
			100+	0	0	0	0	0	0
		여	0-4	334	541	309	308	345	403
			5-9	409	542	387	347	368	409
			10-14	587	742	602	562	582	603
			15-19	328	840	326	318	374	595
			20-24	146	864	161	151	206	553
			25-29	297	948	284	298	394	669
			30-34	371	715	344	356	405	527
			35-39	511	773	516	526	552	611
			40-44	693	974	676	625	642	762
			45-49	1,080	1,367	1,093	1,097	1,108	1,159
			50-54	1,314	1,676	1,312	1,360	1,370	1,437
			55-59	1,394	1,602	1,353	1,378	1,374	1,391
			60-64	1,613	1,721	1,604	1,688	1,645	1,555
			65-69	1,960	2,172	1,992	1,968	1,949	1,930
			70-74	2,261	2,430	2,395	2,452	2,385	2,235
			75-79	1,725	1,712	1,617	1,662	1,622	1,538
			80-84	965	921	919	932	903	811
			85-89	440	441	438	414	404	348
			90-94	126	145	119	99	89	61
			95-99	22	21	25	37	18	0
			100+	2	6	0	0	0	0

7) 지역*성별*연령별 인구 추정량 - (11)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
37	030	남	0-4	2,600	2,705	2,910	2,760	2,747	2,522
			5-9	2,738	2,779	2,789	2,731	2,709	2,568
			10-14	3,966	4,001	3,945	3,858	3,826	3,690
			15-19	3,548	4,362	3,606	3,812	3,793	3,945
			20-24	2,239	4,323	2,152	2,211	2,385	3,596
			25-29	3,070	4,150	2,872	3,136	3,196	3,663
			30-34	3,601	4,248	3,941	3,832	3,812	3,866
			35-39	4,171	4,767	4,281	4,221	4,199	4,324
			40-44	4,658	5,334	4,805	4,778	4,745	4,853
			45-49	5,075	5,374	5,068	5,091	5,025	4,947
			50-54	5,031	5,436	5,160	5,243	5,158	5,023
			55-59	4,210	4,350	3,965	4,176	4,090	4,012
			60-64	3,617	3,546	3,474	3,691	3,571	3,324
			65-69	2,832	2,943	2,702	2,778	2,727	2,703
			70-74	2,910	2,816	2,833	2,845	2,782	2,625
			75-79	2,129	1,983	1,958	2,023	1,962	1,849
			80-84	1,041	936	1,040	1,119	1,048	899
			85-89	409	359	326	388	348	330
			90-94	86	88	58	66	47	70
			95-99	11	11	5	11	0	2
			100+	2	2	0	0	0	0
		여	0-4	2,340	2,461	2,483	2,355	2,372	2,362
			5-9	2,522	2,571	2,624	2,538	2,541	2,475
			10-14	3,587	3,594	3,328	3,605	3,530	3,429
			15-19	3,407	3,920	3,415	3,453	3,463	3,655
			20-24	2,261	3,342	2,112	2,302	2,409	2,990
			25-29	2,981	3,665	3,067	3,075	3,093	3,364
			30-34	3,467	3,848	3,932	3,706	3,701	3,625
			35-39	4,034	4,314	4,183	4,055	4,037	4,030
			40-44	4,533	4,939	4,630	4,552	4,528	4,583
			45-49	4,890	5,063	5,053	4,991	4,934	4,760
			50-54	5,247	5,480	4,980	5,111	5,052	5,083
			55-59	4,682	4,775	4,475	4,600	4,529	4,459
			60-64	3,901	3,785	3,727	3,955	3,837	3,597
			65-69	3,858	4,006	3,747	3,677	3,649	3,702
			70-74	4,374	4,218	4,203	4,247	4,165	3,968
			75-79	3,424	3,165	3,037	3,206	3,117	2,981
			80-84	2,097	1,908	1,841	2,114	1,993	1,830
			85-89	927	851	860	881	849	812
			90-94	264	278	267	288	265	268
			95-99	46	63	49	45	28	58
			100+	3	8	5	4	0	5

7) 지역*성별*연령별 인구 추정량 - (12)

시도	시군구	성별	연령	센서스	행정자료	Direct	Ratio	EBLUP	HB
37	430	남	0-4	124	154	123	118	123	112
			5-9	161	167	183	163	161	138
			10-14	184	198	233	205	200	173
			15-19	156	229	175	176	182	178
			20-24	68	306	93	93	138	187
			25-29	161	337	178	165	190	228
			30-34	236	350	238	292	292	293
			35-39	302	401	343	339	338	338
			40-44	337	460	388	385	384	389
			45-49	350	481	335	405	403	411
			50-54	445	531	453	483	469	467
			55-59	354	444	305	361	363	374
			60-64	307	366	283	302	301	305
			65-69	242	302	218	221	227	237
			70-74	211	234	200	256	227	210
			75-79	122	131	138	126	119	105
			80-84	80	85	65	88	76	64
			85-89	17	20	18	20	14	2
			90-94	3	4	3	1	0	0
			95-99	0	0	0	0	0	0
			100+	0	0	0	0	0	0
		여	0-4	131	144	160	151	152	140
			5-9	128	140	158	145	146	135
			10-14	189	204	215	187	191	186
			15-19	112	181	115	111	127	141
			20-24	69	220	68	65	99	142
			25-29	148	325	153	160	198	242
			30-34	196	316	200	190	211	243
			35-39	218	295	263	244	248	254
			40-44	279	310	293	290	285	281
			45-49	325	429	345	350	355	369
			50-54	378	492	383	448	436	442
			55-59	324	428	335	375	370	378
			60-64	308	363	285	272	280	298
			65-69	296	338	320	325	313	306
			70-74	278	295	278	315	289	275
			75-79	179	209	170	163	164	168
			80-84	117	129	140	144	132	116
			85-89	55	66	48	63	56	48
			90-94	16	23	10	13	12	4
			95-99	3	4	0	0	0	0
			100+	0	0	0	0	0	0

제5장 등록센서스(인구, 가구, 주택) 결과생산 방식에 대한 평가 연구

제1절 연구목적

- 이 연구는 등록센서스 실시목적으로 편집되는 행정자료 데이터베이스에 포함된 센서스 항목에 대한 통계생산 방식을 평가하는 연구이다. 이것은 사실상 등록센서스 통계를 생산하는데 필요한 행정자료의 품질도 평가되어야 한다는 것을 의미한다.
- 인구, 가구, 주택에 관한 가장 기본적인 통계는 지금까지 현지실사의 형태로 응답자로부터 직접 획득한 센서스 자료에서 얻을 수 있다. 그러나 우리나라에는 인구센서스 이외에 여러 형태의 행정자료가 존재한다. 인구, 가구, 주택 부문에 관해서는 주민등록자료, 가족관계부, 건축물대장 등 여러 형태의 행정자료가 있다. 우리나라 통계청이 2015년 센서스에서는 이러한 행정자료를 토대로 센서스 추정방법론을 적용하여 보완된 센서스 통계를 작성함으로써, 2010년까지 5년마다 실시해 온 전통적 센서스의 전수항목을 대체하는 센서스 방식의 패러다임 전환(paradigm shift)을 시도하고 있다
- 등록센서스 실시의 전제조건으로 행정자료의 품질을 이야기하지만, 행정자료는 품질 면에서 결코 완전하지 않다(UNECE, 2007; Wallgren and Wallgren, 2007; Statistics Sweden, 2011). 물론 어느 통계자료도 100% 완벽하지 않다. 그렇기 때문에 전통적 방식으로 실시되는 인구주택총조사 자료도 여러 인구학적 가정, 표본조사, 사후조사 등을 토대로 자료의 포괄범위나 응답내용을 점검하여 품질개선 작업을 추진하기도 하였다. 또 행정자료로 전통적 방식의 인구주택총조사 자료를 대체하는 경우, 행정자료의 포괄범위 조정(coverage

adjustment)은 물론 기재내용이 정확하지 않기 때문에 이를 점검하여 보정하고, 센서스 통계 작성에 이용하여야 한다. 그런데 행정자료의 기재내용은 센서스 모집단의 크기, 곧 인구, 가구, 주택의 총규모를 산출하는데 이용되는 성별, 연령별, 거주지별 특성을 제외하고, 행정자료 안에 포함된 센서스 항목들의 내용에는 오류가 없는지, 또는 오류가 있다면 어떻게 개선할 수 있는 방법이 있는가에 대해서 체계적으로 검토된 적이 거의 없다.

- 유럽이나 아시아의 일부 등록센서스 실시국가에서 직접추정(Direct Estimator), 비추정(Ratio Estimator), 이원회귀추정(DSE)) 등의 센서스 추정 방법론을 이용하여, 인구, 가구, 주택 부문 센서스 모집단의 크기를 확정하고 있다. 현재, 우리나라 등록센서스 추정방법론으로서, 가장 현실성이 있는 방안으로 능형회귀추정량(Ridge regression estimator)에 대한 연구가 집중적으로 이루어지고 있으며, 그 결과는 행정자료를 인구, 가구 및 주택분야에서 이용할 때 센서스 모집단인 인구, 가구와 주택의 크기를 산출한다(한국인구학회, 2011과 본 보고서 제3장 참고). 또, 능형회귀추정량과 비교목적으로 읍면동 수준에서 소지역추정 방법론의 다양한 추정방안을 검토되고 있는데, 직접추정량, 비추정량, EBLUP (Empirical Best Linear Unbiased Predictor), HB 추정량이 그 예이다(본 보고서의 제4장을 참고).
- 우리나라의 2015년 등록센서스의 추진과정에서 센서스 부모집단의 크기를 추정하는데 이용될 것으로 예상되는 능형회귀추정량은 행정자료 기반 통계와 표본자료 기반 통계의 “가중평균”(weighted mean)이며, 가중치는 λ 의 함수로 표시된다. 곧 λ 의 값이 커짐에 따라 능형회귀추정량은 행정자료 기반 통계에 접근하며, λ 의 값이 0에 가까워짐에 따라 능형회귀추정량은 표본자료 기반 불편 추정량으로 수렴한다. 물론, 부모집단별 능형회귀추정량에 의하여 산출된 통계와 행정자료 기반 통계를 비교하여, 조사구(ED)별로 “센서스 가중치(census weight)”를 산출하며, 이 가중치는 다시 행정자료 데이터베이스에 포함된 센서스 항목에 대한 통계를 작성하는데 사용된다. 여기서 능형회귀추정량에 의하여 도출될 수 있는 “센서스 가중치”는 추정목적에 의거하여 각각의 가정에 따라 몇 가지 세트가 작성될 수 있다. 곧, 이러한 세트의 센서스 가중치가 작성되면, 행정자료의 개체 레코드(individual record)에 가중치

가 설정되고, 그 가중치는 행정자료에 포함된 센서스 조사항목(예: 혼인상태, 가구주와의 관계, 가구원수, 주택연면적, 주택건축연도 등)에 대하여, 조사표를 토대로 하는 전통적 센서스의 전수집계 결과를 대체하는데 있어서도 그대로 사용된다.

- 올해의 등록센서스 학술연구 용역사업에서, 표본설계 적정화와 모집단(또는 부모집단) 크기의 추정 목적으로 인구, 가구, 주택 부문의 행정자료가 기존의 센서스 자료와 비교해서 인구수, 가구수, 주택수가 큰 차이가 없는 경우와 약간의 차이가 있는 경우, 차이가 많이 나는 경우 등으로 구분하고 있다. 예를 들어 센서스 자료의 인구 및 가구분포 등과 행정자료의 인구 및 가구분포를 비교하면 지역별로 차이가 거의 없는 지역, 약간 차이나는 지역, 차이가 많이 나지 않는 지역을 구분할 수 있다. 그런데 그뿐 아니라 차이가 어디서 왜 나타나는지도 파악해야 할 필요가 있다. 더 나아가 능형회귀추정량을 이용하여 행정자료의 개체 레코드에 적용되는 센서스 가중치를 계산하고, 이것으로 보완된 행정자료를 만드는 경우, 부모집단 또는 영역 정의(domain definition)에 고려되지 않은 다른 센서스 항목들의 분포(가령, 성별과 연령별 분포만을 보완하고, 혼인상태, 가구주와의 관계 등 새로운 항목을 추가하여 만들 수 있는 새로운 분포)는 원래의 행정자료와 보완된 행정자료를 비교하면 어떤 차이가 나는지는 아직 체계적으로 점검해 본 적이 없다.
- 제5장은 통계청이 2015년에 등록센서스를 본격적으로 실시함에 따라, 행정자료 데이터베이스에 포함된 센서스 항목으로 전통적 방식으로 실시되는 센서스의 조사항목을 대체하는 경우, 그 데이터베이스에 포함된 센서스 항목에 대한 통계를 어떻게 보완하여야 할 것인지, 그리고 보완한 자료에서 도출되는 통계는 어떤 특성을 갖게 되는지 분석하여 행정자료 기반 통계 작성방식을 평가하는 것이 목적이다. 구체적으로는 등록자료(행정자료)를 기준으로 놓고, 등록자료에 포함된 인구, 가구, 주택의 센서스 항목의 통계와 인구센서스의 해당 항목의 통계 사이에 얼마나 유사한지를 점검해본다. 또 다른 점검은 인구, 가구, 주택 부문의 등록자료를 능형회귀추정량을 이용해 보완했을 때 원 등록자료와 보완된 등록자료 사이에 얼마나 유사한지 아니면 상이한지를 검토한다.

- 이 연구를 위해 간단한 지표를 개발하여 사용한다. 인구학에서는 인구학적 분포를 포함하고 있는 자료의 질을 검토할 때 비유사성 지수(index of dissimilarity)를 구하여 둘 혹은 그 이상의 분포를 비교한다. 이 연구에서도 비유사성 지수는 등록자료 상의 분포와 인구센서스의 분포를 비교할 때 가장 기본적이고 중요한 분석 지표가 된다. 그러나 이 연구에서 다루는 12개 지역의 등록자료 상의 분포가 인구센서스 상의 분포와 어디서 얼마나 같고 다른지를 구체적으로 검토하기 위해서는 비유사성 지수를 이용한 분석 외의 다른 분석이 필요하다. 인구의 경우 구체적으로 지역에 따라 어느 성별, 어느 연령층에서 분포상의 차이가 나는지를 알아보기 위해서는 분포의 전체 차이뿐만 아니라 각 성별, 연령별로 분포의 차이를 검토해야 한다. 이를 위해서는 센서스 자료와 행정자료, 그리고 가중치 보정 이전과 이후의 행정자료들 간의 오차크기라는 지표를 개발하여, 그것을 중심으로 보완된 행정자료를 이용하여 작성이 가능하다고 생각되는 센서스 항목의 통계 품질을 점검하고자 한다. 자료들에서 도출되는 통계 간의 오차는 각 지역별로 주어진 센서스 항목의 오차를 구한 후, 그 절대값을 취하였다, 각 지역에서 센서스 항목에 따라 나타난 오차의 절대값을 합산하고 이 중에서 주어진 센서스 항목별 범주가 차지하는 비율을 통해 오차크기의 비율을 산정하였다. 이러한 오차크기의 비율은 지역과 센서스 항목에 따라 발생하는 전체 오차(절대값) 중에서 특정 지역과 항목에서 발생하는 오차의 상대적 크기를 제시한다. 이는 제1장에서 이상시군구를 탐색할 때 개발한 지표 중 RLAD와 같은 개념이다. 이상시군구를 탐색할 때는 이 지표가 실질적으로 사용되지는 않았다. 그러나 행정자료와 인구센서스자료의 인구분포를 비교할 때는 이 지표가 유용하기 때문에 이 장에서는 이 지표를 적극 활용한다.

제2절 연구방법

- 본 연구에 참여하고 있는 통계학자들은 행정자료의 완전성(completeness) 곧 포괄범위의 문제점을 해결하기 위하여 능형회귀추정방안을 고려하여 행정통계와 조사통계의 “가중평균”으로 센서스 모집단(또는 부모집단)의 크기를

확정한다. 두말할 필요도 없이, 이러한 센서스 모집단(또는 부모집단)의 크기를 해당 행정자료의 크기와 비교하여 “센서스 가중치”(census weight)를 산출하여, 새로운 인구, 가구, 주택분포를 구하는 작업은 행정자료 기반 센서스 통계 작성의 기초가 된다. 제5장의 연구는 능형회귀추정량을 적용하여 계산한 센서스 모집단(또는 부모집단)의 크기를 해당 행정자료의 크기와 비교하여 계산한 센서스 가중치를 행정자료에 다시 적용하여, 등록센서스에서 작성 가능하다고 판단되는 조사항목에 대하여 얻은 새로운 인구, 가구, 주택 자료가 갖는 특성을 점검하여 등록센서스 결과 생산방식을 평가하는 것이다. 이를 위해 다음과 같은 방법론적 절차를 밟았다.

- 통계청 등록센서스 담당부서는 본 보고서의 제1장에서 진행된 인구, 가구, 주택 등의 측면에서 행정자료와 센서스 자료와 비교한 후 차이의 정도를 검토하여, 다음과 같은 12개 연구 대상지역을 선정하였다: 서울 금천구, 서울 영등포구, 인천 서구, 광주 동구, 경기도 화성시, 강원도 양구군, 충청북도 증평군, 전라북도 임실군, 전라남도 나주시, 전라남도 신안군, 경상북도 김천시, 경상북도 울릉군. 기본적으로 이 연구는 이 12개 지역의 인구, 가구, 주택에 관한 행정자료를 이용한다.
- 이 연구에 참여한 연구진 중 통계학자는 인구 부문의 “센서스 가중치”를 산출하는데 필요한 능형회귀추정량을 다음과 같이 각기 다른 4개의 방안을 고려하였다(제3장 참고).
 - 1) 방안 1 (릿지 1) - 읍면동별 성별 연령(10세별) 가중치
 - 2) 방안 2 (릿지 2) - 읍면동별 성별 연령 3그룹(14세 이하/ 15-64세/ 65세 이상)별 가중치
 - 3) 방안 3 (릿지 3) - 동부/읍부/면부별 성별 연령(10세별) 가중치
 - 4) 방안 4 (릿지 4) - 시군구 내에서 동부/읍부/면부별 성별 연령(10세별) 가중치

이와 같이 지역의 소구분과 연령구분이 다르기 때문에, 그 기준의 차이에 따라 서로 다른 가중치의 세트가 작성될 수 있음은 물론이다.

- 가구와 주택 부문에서는 “센서스 가중치”를 계산하는데 필요한 능형회귀추정량은 하나의 방안만 고려되었다.
- 행정자료의 센서스 항목에 대하여 등록센서스 결과생산 방식을 평가하는 본 연구는 능형회귀추정량을 통해 확정된 센서스 모집단의 크기를 해당 부모집단 행정자료의 크기와 비교하여 “센서스 가중치”를 구하고 그것을 성별 및 연령별 변수 이외의 다른 센서스 항목을 대상으로 새로운 분포를 만들었다. 첫째, 인구 부문은 행정자료에 포함된 센서스 조사항목인 혼인상태와 가구주와의 관계라는 항목을 이용하였다. 즉 능형회귀추정량과 행정자료의 크기를 비교하여 만든 센서스 가중치는 성별, 연령별 이라는 두 센서스 조사항목만을 고려하여 산출되었다. 따라서 두 항목을 고려한 센서스 가중치에 의하여 만들어진 성별 연령별 인구분포는 센서스 가중치 적용 이전의 행정자료에서의 성별 연령별 인구분포와 비교할 때 큰 차이가 없을 수 있다. 그러나 성별, 연령별 이라는 두 센서스 항목만 이용하여 만든 가중치를 다시 이용하되, 이번에는 혼인상태와 가구주와의 관계라는 두 센서스 조사항목을 추가로 고려하여 분포를 만들 수 있다. 즉 “성별 연령별 구조”라는 인구분포가 원래의 분포라면 본 연구에서는 성별, 연령별 혼인상태 분포라는 새로운 분포를 만드는데, 성별, 연령별 부모집단 능형회귀추정량과 해당 행정자료의 크기를 비교하여 만든 센서스 가중치를 적용하는 것이다. 그러면 가중치가 전혀 부여되지 않은 행정자료 상의 성별, 연령별 혼인상태 분포와 가중치를 부여하여 만든 성별, 연령별 혼인상태의 분포가 얼마나 같은지 혹은 차이가 나는지는 미리 예측하기가 힘들다. 이 연구는 바로 성과 연령이라는 두 센서스 항목만을 가지고 산출한 가중치를 혼인상태나 가구주와의 관계라는 제3의 센서스 항목을 도입하여 새로운 분포를 만들게 되면, 원래의 행정자료 상의 분포와 얼마나 달라지는지 확인하여, 등록센서스 결과생산 방식에 대한 아이디어를 얻게 된다.
- 우리는 등록센서스 결과 생산방식의 성능을 평가하기 위하여 센서스 가중치를 능형회귀추정량 4개의 세트에 대하여 릿지 1, 릿지 2, 릿지 3, 릿지 4로 명명하고, 이에 따라 보정된 행정자료에서 작성된 센서스 항목의 통계들을 비교 검토한다. 결국 인구 부문의 경우 12개 연구지역을 대상으로 원래의 행정자

료, 릿지 1, 릿지 2, 릿지 3, 릿지 4에 의하여 보완된 행정자료에서 도출되는 지역별 성별 연령별 혼인상태 분포표와 지역별 성별 연령별 가구주와의 관계라는 분포표를 얻는다.

- 등록센서스 결과 생산방식의 품질을 평가하는 기본적인 방법은 행정자료에서 얻은 지역별 성별 연령별 혼인상태 분포를 릿지 1, 릿지 2, 릿지 3, 릿지 4 등 4종류의 센서스 가중치를 적용하여 만들어지는 지역별 성별 연령별 혼인상태 분포와 비교하는 것이다. 혼인상태를 이용한 자료의 비교 검토가 끝나면 다시 가구주와의 관계를 이용하여, 등록자료의 지역별 성별 연령별 가구주와의 관계 분포를 릿지 1, 릿지 2, 릿지 3, 릿지 4 등의 4종류의 가중치를 적용하여 만든 지역별 성별 연령별 가구주와의 관계 분포와 비교하게 된다.
- 분포의 비교는 어떻게 이루어지는가?: 우리는 이 연구에서 분포 사이의“차이”를 구하는 방식으로 분포를 비교하고자 하였다. 차이를 구하는 방법은 다음과 같다. 1) 단순한 차이를 구해 연령별 혼인상태 분포의 각 셀 값이 양(+)의 값이나 음(-)의 값을 모두 가질 수 있도록 허용하는 방식. 2) 양(+)의 값과 음(-)의 값을 합하면 값이 상쇄될 수 있으므로 차이를 구한 다음 절대값을 취하는 방식. 따라서 단순히 한 분포의 값에서 다른 분포의 값을 뺄 때 양과 음의 값을 허용하는 분포와 절대 값으로 나타나는 분포 등 두 가지 분포를 얻을 수 있다.
- 그리고 두 인구분포를 비교하는 또 하나의 방법은 3) 비(ratio)를 구하는 것이다. 비를 구하면 1 이상 혹은 1 미만의 값을 얻게 되고, 이 비 값은 두 분포의 차이를 나타내는 지표로 활용할 수 있다. 즉 1보다 크면 앞 분포의 값이 뒤 분포의 값보다 큰 것이고, 1보다 작으면 뒤 분포의 값이 앞 분포의 값보다 큰 것이다. 이를 통해 두 분포의 각 셀별로 값의 크고, 작음, 그리고 이런 값의 분포가 지역별, 성별, 연령별로 어떤 패턴을 보이는지, 일관성이 있는지, 아니면 사례마다 특정한 유형이라고 볼 수 있는지를 점검한다.
- 이와 같은 방식을 취하면 12개 지역에서 수많은 표가 만들어진다. 인구부문 등록센서스 결과 생산방식의 품질을 평가하는 경우 예를 들어 보자. 릿지 1의

경우, 인구부문 행정자료에서 얻은 분포, 릿지 1에서 얻은 분포, 릿지 1의 인구분포에서 인구 등록자료의 인구분포를 빼서 얻은 분포, 절대값을 취한 분포, 비를 구한 분포 등 다섯 개의 표가 만들어진다. 릿지값이 4개가 있으므로 릿지 1, 2, 3, 4에 대해서 20개의 표가 만들어진다. 그런데 우리가 다루는 지역은 12개 지역이므로 $12 \times 20 = 240$ 개의 표가 만들어진다.

- 우리는 인구부문 자료에서 혼인상태 항목 이외에 가구주와의 관계라는 항목의 분포도 등록센서스 결과생산 방식의 관점에서 검토해야 한다. 그럼 결국 인구부문 행정자료와 센서스 가중치를 적용해서 얻은 인구분포를 비교하기 위해 $240 \times 2 = 480$ 개의 표를 산출하도록 되어 있다.
- 인구부문에서 행정자료와 인구센서스자료를 비교할 때는 다음과 같은 지표를 활용한다. 이 지표는 이 연구보고서 제1장의 이상 시군구 탐색 연구에서 제시한 지표(RLAD)와 동일하다

$$\frac{\text{컨조사자료} - \text{행정자료컨}}{\sum \text{컨조사자료} - \text{행정자료컨}}$$

여기서, $\sum \text{컨조사자료} - \text{행정자료컨}$ 는 소속지역

- 인구, 가구, 주택의 행정자료와 인구센서스자료, 행정자료와 이를 능형회귀추정량으로 보완한 릿지1, 릿지2, 릿지3, 릿지4 등의 분포와 비교할 때, 각 분포의 각 셀마다 독특한 특성이 있기 때문에 이를 다 살펴볼 필요가 있지만, 현실적으로 이는 불가능한 작업이 될 수 있다. 그래서 이 연구는 인구학에서 두 분포를 비교하여 자료의 완전성을 검증할 때 사용하는 대표적인 방법을 활용하였다. 인구학에서는 비유사성지수(index of dissimilarity)를 구하여 두 분포에 얼마나 차이가 나는지를 검증한다(Kuha and Firth, 2011). 이 연구에서 등록센서스 결과 생산방식의 품질을 평가하는 것은 결국 지역별, 성별, 연령별, 혼인상태별 분포를 구해 비교하는 것이기 때문에 기본적으로 비유사성지수를 구해 두 분포의 차이를 보는 것이 가장 기본적이고 효과적인 방법 가운데 하나다. 그래서 이 연구는 기본적으로 여러 분포를 비교하기 위해 비유사성지수를

구하였다.

- 비유사성지수는 다음과 같은 공식으로 구한다.

$$\frac{1}{2} \sum_{i=1}^N \left| \frac{b_i}{B} - \frac{w_i}{W} \right|$$

이 공식에서 B는 기준이 되는 분포의 총수이다. b_i 는 i 번째 셀의 값이다. 즉 전체에서 해당 셀이 차지하는 비율이 된다. W는 비교하는 분포의 총수이다. w_i 는 i 번째 셀의 값이다. 즉 전체에서 해당 셀이 차지하는 비율이 된다. 두 분포의 차이를 보기 위해서는 양의 값과 음의 값이 상쇄되는 것을 막기 위해 절대값을 취한다. 그런데 절대값을 취해 합을 구하면 원래 값의 두 배가 되는 효과가 있기 때문에 다시 2로 나누어준다.

- 지금까지 우리는 행정자료와 인구센서스질, 그리고 행정자료에서 얻은 인구분포를 4종류의 센서스 가중치를 적용해서 얻은 인구분포와 비교하여 등록센서스 결과생산방식을 평가하는 방법을 설명하였다. 그런데 이런 작업은 기본적으로 행정자료를 기본으로 하고, 행정자료의 센서스 항목에 내용오류가 있는 경우 그것을 어떻게 보완할 것인지만을 고려하는 것이다. 센서스가 보편성(universality)의 원칙을 준수해야 하는 이상, 행정자료 자체의 완전성을 점검해야 하지만, 이것은 행정자료의 “개체 레코드”(individual records)가 과다집계되어 있는지, 아니면 과소집계되어 있는지 등에 관해서 평가하는 것이 아니다.
- 본 연구는 등록센서스 결과생산방식의 품질을 점검하는 연구이지만, 행정자료에 입력된 센서스 항목의 내용오류(content error)를 점검하는 것도 대단히 중요한 과제이다. 아마도, 현재 상태에서 더 중요한 것은 행정자료의 경우, 기본적으로 인구주택총조사와 비교해 보는 작업이 필수적이다. 물론 인구주택총조사는 2010년에 실시되었고, 행정자료도 센서스 실시시점인 2010년 11월 1일을 기준으로 공적 장부인 주민등록, 가족관계등록부, 건축물대장에서 추출

된 자료이기 때문에, 자료 간 비교의 동시성 원칙(simultaneity principle)이 지켜지고 있다고 볼 수 있다.

제3절 등록센서스 통계생산방식 평가: RLAD 지표를 활용한 행정자료 와 인구센서스자료 비교 평가

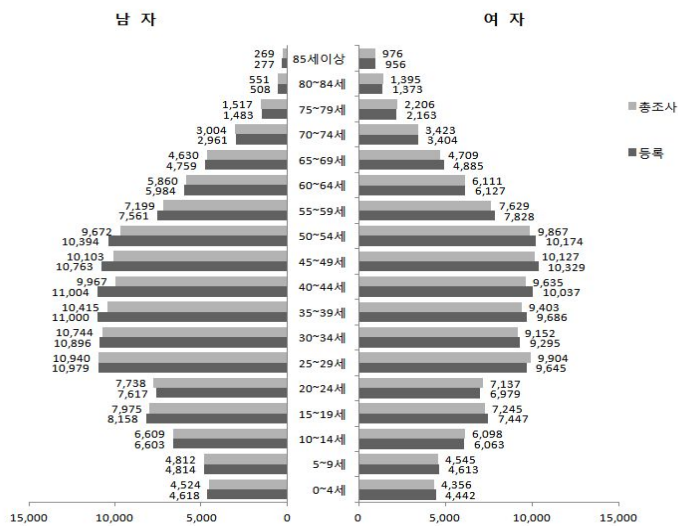
1. 2010년 센서스와 행정자료의 인구

- 인구부문의 등록센서스 결과 생산방식을 평가할 때, 주축이 되는 자료는 2010년 센서스 자료와 행정자료이다. 따라서 센서스 자료와 주민등록이나 가족관계등록부 등 공적장부의 형태로 주어지는 행정자료의 기본 성격을 이해할 필요가 있다.
- 이 연구에서는 먼저 12개 지역의 행정자료와 인구센서스자료의 성별 연령별 인구구조를 비교하면서 지역별 자료의 특성을 통해 행정자료의 품질을 평가해 본다.
- 이 절에서는 각 지역별로 성별 연령별 인구구조 분석을 위해 사용한 자료와 각 지역별 성별 연령별 인구구조를 통해 행정자료의 성격을 이해해 본다.

<표 5-1> 서울시 금천구의 등록자료와 인구센서스자료의 성별 연령별 구조

금천구	남자		여자	
	등록	총조사	등록	총조사
0~4세	4,618	4,524	4,442	4,356
5~9세	4,814	4,812	4,613	4,545
10~14세	6,603	6,609	6,063	6,098
15~19세	8,158	7,975	7,447	7,245
20~24세	7,617	7,738	6,979	7,137
25~29세	10,979	10,940	9,645	9,904
30~34세	10,896	10,744	9,295	9,152
35~39세	11,000	10,415	9,686	9,403
40~44세	11,004	9,967	10,037	9,635
45~49세	10,763	10,103	10,329	10,127
50~54세	10,394	9,672	10,174	9,867
55~59세	7,561	7,199	7,828	7,629
60~64세	5,984	5,860	6,127	6,111
65~69세	4,759	4,630	4,885	4,709
70~74세	2,961	3,004	3,404	3,423
75~79세	1,483	1,517	2,163	2,206
80~84세	508	551	1,373	1,395
85세이상	277	269	956	976

<그림 5-1> 서울시 금천구의 등록자료와 인구센서스자료의 성별 연령별 구조



- 금천구는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서는 행정자료와 인구센서스자료 사이에 차이가 가장 큰 지역의 하나로 선정된 지역이다(SWORD 및 이상 시군구 논의에 관해서는 본 보고서의 제4장을 참고). 그러나 인구구조의 모양으로는 현재 한국의 전체 인구구조와 유사한 모양을 하고 있다.
- 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

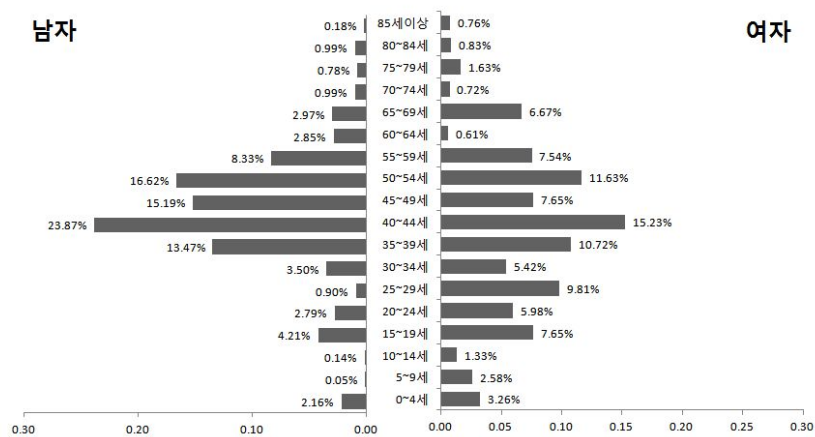
<표 5-2> 서울시 금천구의 등록자료와 인구센서스자료의 연령별 성비

	금천구		
	등록	총조사	성비의 차
0~4세	104.0	103.9	0.1
5~9세	104.4	105.9	-1.5
10~14세	108.9	108.4	0.5
15~19세	109.5	110.1	-0.5
20~24세	109.1	108.4	0.7
25~29세	113.8	110.5	3.4
30~34세	117.2	117.4	-0.2
35~39세	113.6	110.8	2.8
40~44세	109.6	103.4	6.2
45~49세	104.2	99.8	4.4
50~54세	102.2	98.0	4.1
55~59세	96.6	94.4	2.2
60~64세	97.7	95.9	1.8
65~69세	97.4	98.3	-0.9
70~74세	87.0	87.8	-0.8
75~79세	68.6	68.8	-0.2
80~84세	37.0	39.5	-2.5
85세이상	29.0	27.6	1.4

- 등록자료나 인구센서스 모두, 금천구는 30~34세, 35~39세, 25~29세 등 젊은 성인 연령층에서 남자가 여자보다 더 많다. 그런데 성비로 보면 등록자료가 인구센서스자료보다 동일한 연령층에서 훨씬 높아, 등록자료상으로는 남자가 훨씬 더 많이 기재되어 있는 것으로 나타난다.

- 금천구의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면 남녀 모두 40~44세에서의 차이가 등록자료와 인구센서스자료의 분포의 차이의 가장 많은 부분을 차지하고 있다. 금천구에서는 40~44세, 45~49세, 50~54세, 55~59세 등 40대, 50대 성인 연령층에서 등록자료와 인구센서스자료의 기록이 가장 차이가 많이 난다. 금천구에 행정자료상으로 등록되어 있는 인구와 실제로 상주하는 인구를 조사한 인구센서스조사상의 인구의 차이가 40, 50대 가장 및 그 배우자인 연령층에서 가장 많이 나타나고 있고 그 이유를 찾아낼 필요가 있다.

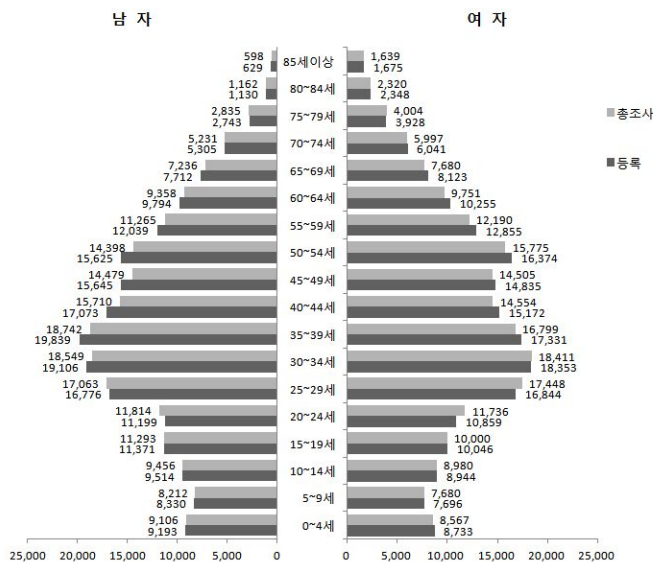
<그림 5-2> 금천구의 성 및 연령별 인구수의 오차비율분포



<표 5-3> 서울시 영등포구의 등록자료와 인구센서스자료의 성별 연령별 구조

영등포구	남자		여자	
	등록	총조사	등록	총조사
0~4세	9,193	9,106	8,733	8,567
5~9세	8,330	8,212	7,696	7,680
10~14세	9,514	9,456	8,944	8,980
15~19세	11,371	11,293	10,046	10,000
20~24세	11,199	11,814	10,859	11,736
25~29세	16,776	17,063	16,844	17,448
30~34세	19,106	18,549	18,353	18,411
35~39세	19,839	18,742	17,331	16,799
40~44세	17,073	15,710	15,172	14,554
45~49세	15,645	14,479	14,835	14,505
50~54세	15,625	14,398	16,374	15,775
55~59세	12,039	11,265	12,855	12,190
60~64세	9,794	9,358	10,255	9,751
65~69세	7,712	7,236	8,123	7,680
70~74세	5,305	5,231	6,041	5,997
75~79세	2,743	2,835	3,928	4,004
80~84세	1,130	1,162	2,348	2,320
85세이상	629	598	1,675	1,639

<그림 5-3> 서울시 영등포구의 등록자료와 인구센서스자료의 성별 연령별 구조



- 영등포구는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서는 행정자료와 인구센서스자료 사이에 차이가 가장 큰 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 최근 들어 0-4세등 유년층 인구가 늘어나고 있는 것처럼 보인다.
- 영등포구의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

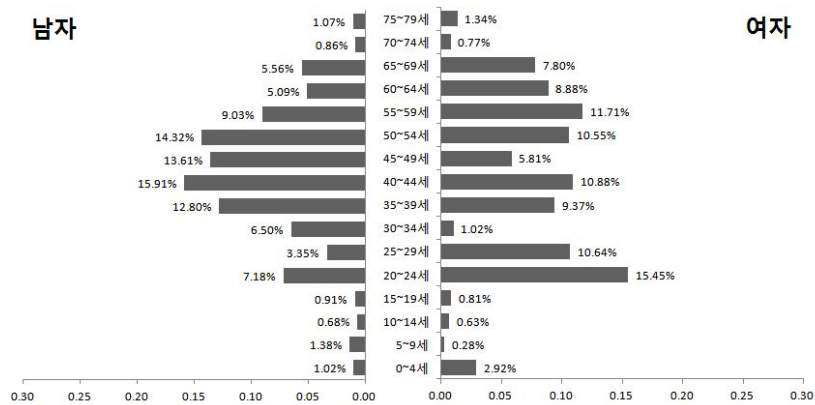
<표 5-4> 서울시 금천구의 등록자료와 인구센서스자료의 연령별 성비

	영등포구		
	등록	총조사	성비의 차
0~4세	105.3	106.3	-1.0
5~9세	108.2	106.9	1.3
10~14세	106.4	105.3	1.1
15~19세	113.2	112.9	0.3
20~24세	103.1	100.7	2.5
25~29세	99.6	97.8	1.8
30~34세	104.1	100.7	3.4
35~39세	114.5	111.6	2.9
40~44세	112.5	107.9	4.6
45~49세	105.5	99.8	5.6
50~54세	95.4	91.3	4.2
55~59세	93.7	92.4	1.2
60~64세	95.5	96.0	-0.5
65~69세	94.9	94.2	0.7
70~74세	87.8	87.2	0.6
75~79세	69.8	70.8	-1.0
80~84세	48.1	50.1	-2.0
85세이상	37.6	36.5	1.1

- 등록자료나 인구센서스 모두, 영등포구는 25-29세에서 성비가 100이하로 여성이 남성보다 더 많은 것으로 나타난다. 반면에 15-19세의 중학교 및 고등학교 연령층에서는 남성이 여성보다 훨씬 많다. 또한 35-39세, 40-44세 등 젊은 성인층에서 남성이 여성보다 훨씬 많다. 그러나 역시 나이가 들수록 남성의 수가 여성보다 급감한다.

- 영등포구의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구 수 기록의 오차비율의 분포를 보면 여성의 경우 20~24세의 연령층에서 등록 자료와 인구센서스자료의 기록이 두 자료의 기록의 차이를 낳는데 가장 크게 기여하고 있다. 여성 기록 가운데 등록 자료와 인구센서스자료의 차이의 총 오차의 15% 이상이 바로 이 연령층에서 비롯되고 있다. 여성의 경우, 반면에 30~34세층은 두 자료에 기록의 차이가 매우 작은 편이다. 남성의 경우 30~34세부터 55~59세까지 성인층의 기록에 차이가 있음을 알 수 있다. 어느 연령층 이 가장 문제라고 지적하기 어렵게, 남성의 등록자료와 인구센서스자료의 연령별 기록 사이에는 오차가 고르게 분포하고 있다.

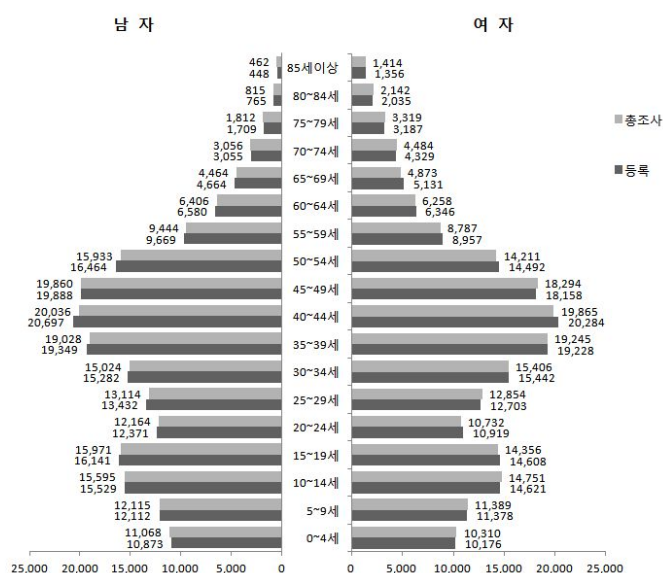
<그림 5-4> 영등포구의 성 및 연령별 인구수의 오차비율분포



<표 5-5> 인천 서구의 등록자료와 인구센서스자료의 성별 연령별 구조

인천서구	남자		여자	
	등록	총조사	등록	총조사
0~4세	10,873	11,068	10,176	10,310
5~9세	12,112	12,115	11,378	11,389
10~14세	15,529	15,595	14,621	14,751
15~19세	16,141	15,971	14,608	14,356
20~24세	12,371	12,164	10,919	10,732
25~29세	13,432	13,114	12,703	12,854
30~34세	15,282	15,024	15,442	15,406
35~39세	19,349	19,028	19,228	19,245
40~44세	20,697	20,036	20,284	19,865
45~49세	19,888	19,860	18,158	18,294
50~54세	16,464	15,933	14,492	14,211
55~59세	9,669	9,444	8,957	8,787
60~64세	6,580	6,406	6,346	6,258
65~69세	4,664	4,464	5,131	4,873
70~74세	3,055	3,056	4,329	4,484
75~79세	1,709	1,812	3,187	3,319
80~84세	765	815	2,035	2,142
85세이상	448	462	1,356	1,414

<그림 5-5> 인천 서구의 등록자료와 인구센서스자료의 성별 연령별 구조



- 인천 서구는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 역시 행정자료와 인구센서스자료 사이에 차이가 가장 큰 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 20세부터 34세까지의 젊은 성인층 인구가 적고, 상대적으로 10대의 인구가 많은 지역적 특성을 보이고 있다.
- 인천 서구의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

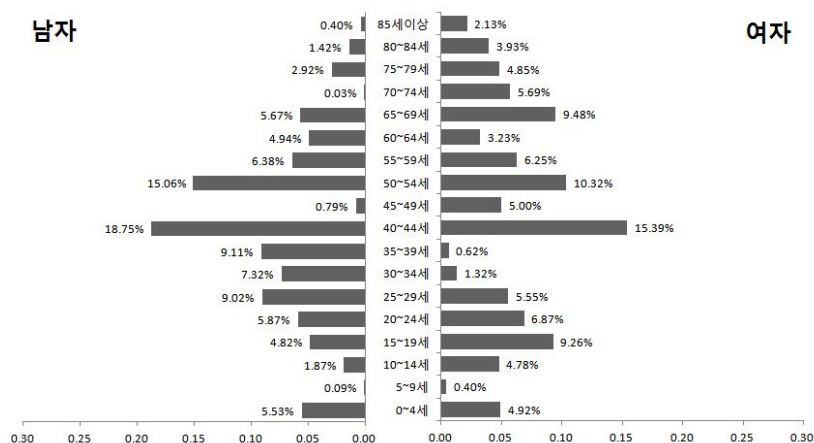
<표 5-6> 인천 서구의 등록자료와 인구센서스자료의 연령별 성비

	인천서구		
	등록	총조사	성비의 차
0~4세	106.8	107.4	-0.5
5~9세	106.5	106.4	0.1
10~14세	106.2	105.7	0.5
15~19세	110.5	111.2	-0.8
20~24세	113.3	113.3	0.0
25~29세	105.7	102.0	3.7
30~34세	99.0	97.5	1.4
35~39세	100.6	98.9	1.8
40~44세	102.0	100.9	1.2
45~49세	109.5	108.6	1.0
50~54세	113.6	112.1	1.5
55~59세	107.9	107.5	0.5
60~64세	103.7	102.4	1.3
65~69세	90.9	91.6	-0.7
70~74세	70.6	68.2	2.4
75~79세	53.6	54.6	-1.0
80~84세	37.6	38.0	-0.5
85세이상	33.0	32.7	0.4

- 등록자료나 인구센서스 모두, 인천 서구는 30~34세 젊은 성인층에서 성비가 100이하로 여성이 남성보다 더 많은 것으로 나타난다. 반면에 50~54세층에서는 남성이 여성보다 훨씬 더 많이 거주하고 있다. 60대 초반에도 남성이 여성보다 더 많은 도시적 속성을 보여주고 있다. 등록자료와 인구센서스자료 사이에 성비의 차는 별로 크지 않다.

- 인천 서구의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구 수 기록의 오차비율의 분포를 보면 남성과 여성 모두 40~44세에서 두 자료의 기록의 차이가 전체 차이에서 가장 많은 비율을 차지하고 있다. 반면에 인접한 45~49세층이나 35~39세층의 두 자료의 기록의 차이는 전체 오차에서 별로 큰 비중을 차지하지 않고 있다. 인천 서구에서는 0~4세층에서 등록자료와 인구센서스자료의 기록의 차이도 유의미한 차이를 보이고 있음을 알 수 있다.
- 인천 서구에서는 등록자료와 인구센서스자료의 차이가 남성과 여성에 따라 크게 다른 연령층이 보인다. 예를 들어 30대층에서는 남성의 경우 두 자료의 기록에 차이가 큰 편이고, 여성의 경우는 45~49세, 70~74세층의 기록에 두 자료의 기록의 차이가 큰 편이다.
- 전반적으로 인천 서구의 등록자료와 인구센서스자료의 성별, 연령별 인구기록은 어떤 특정한 유형을 찾기 힘들 정도로 부정형적인 차이가 나타난다.

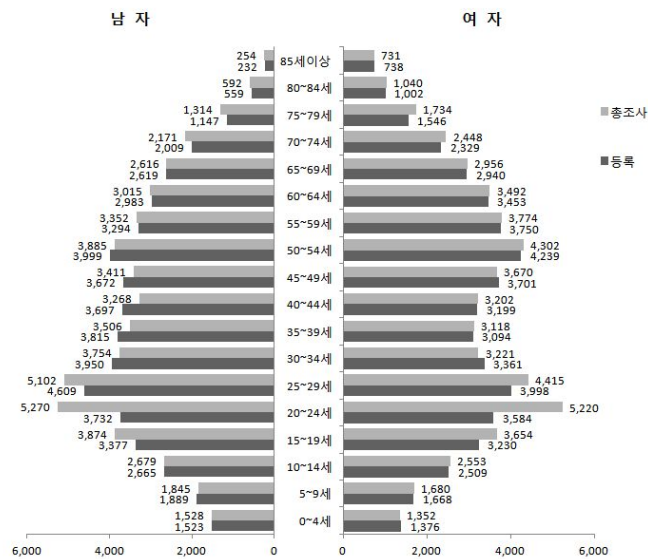
<그림 5-6> 인천 서구의 성 및 연령별 인구수의 오차비율분포



<표 5-7> 광주 동구의 등록자료와 인구센서스자료의 성별 연령별 구조

광주동구	남자		여자	
	등록	총조사	등록	총조사
0~4세	1,523	1,528	1,376	1,352
5~9세	1,889	1,845	1,668	1,680
10~14세	2,665	2,679	2,509	2,553
15~19세	3,377	3,874	3,230	3,654
20~24세	3,732	5,270	3,584	5,220
25~29세	4,609	5,102	3,998	4,415
30~34세	3,950	3,754	3,361	3,221
35~39세	3,815	3,506	3,094	3,118
40~44세	3,697	3,268	3,199	3,202
45~49세	3,672	3,411	3,701	3,670
50~54세	3,999	3,885	4,239	4,302
55~59세	3,294	3,352	3,750	3,774
60~64세	2,983	3,015	3,453	3,492
65~69세	2,619	2,616	2,940	2,956
70~74세	2,009	2,171	2,329	2,448
75~79세	1,147	1,314	1,546	1,734
80~84세	559	592	1,002	1,040
85세이상	232	254	738	731

<그림 5-7> 광주 동구의 등록자료와 인구센서스자료의 성별 연령별 구조



- 광주 동구는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 역시 행정자료와 인구센서스자료 사이에 차이가 중간 정도에 해당하는 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 20대 청년층의 인구가 많고, 상대적으로 30대와 40대의 젊은 성인층 인구가 적은 지역적 특성을 보이고 있다. 동시에 행정자료와 인구센서스자료 사이에 인구구조가 차이가 많이 나는 것으로 보인다.
- 광주 동구의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-8> 광주 동구의 등록자료와 인구센서스자료의 연령별 성비

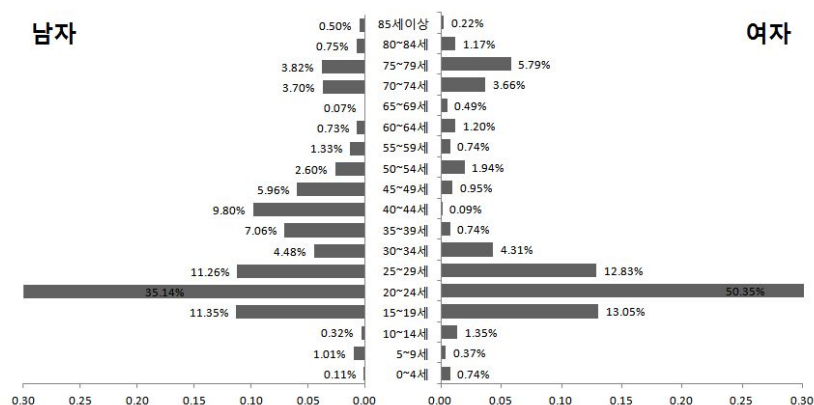
	광주동구		
	등록	총조사	성비의 차
0~4세	110.7	113.0	-2.3
5~9세	113.2	109.8	3.4
10~14세	106.2	104.9	1.3
15~19세	104.6	106.0	-1.5
20~24세	104.1	101.0	3.2
25~29세	115.3	115.6	-0.3
30~34세	117.5	116.5	1.0
35~39세	123.3	112.4	10.9
40~44세	115.6	102.1	13.5
45~49세	99.2	92.9	6.3
50~54세	94.3	90.3	4.0
55~59세	87.8	88.8	-1.0
60~64세	86.4	86.3	0.0
65~69세	89.1	88.5	0.6
70~74세	86.3	88.7	-2.4
75~79세	74.2	75.8	-1.6
80~84세	55.8	56.9	-1.1
85세이상	31.4	34.7	-3.3

- 등록자료나 인구센서스 모두, 광주 동구는 45세 이상 연령층에서는 남자가 여자보다 적은 지역으로 나타난다. 이 연구에서 다루는 다른 도시 지역과 달리 광주 동구는 45세가 넘어가면 남자의 수가 여자의 수보다 적어지기 시작하는

지역적 특성을 보이고 있다.

- 광주 동구의 등록자료와 인구센서스자료의 차이를 보기 위해 성비를 비교해보면, 35세부터 49세까지의 연령층에서 등록자료와 인구센서스 자료의 성비가 크게 차이난다. 남성은 등록자료상 숫자가 인구센서스자료상 숫자보다 더 많은 것으로 나타나고, 반면에 여성은 이 연령층에서 인구센서스자료상 숫자가 등록자료상 숫자보다 더 많은 것으로 나타나면서 두 자료 사이에 성비의 차이가 크다.
- 광주 동구의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면 이 지역에서 등록자료와 인구센서스자료의 차이를 가장 크게 만드는 요인은 20대초반의 인구기록에서 비롯된다. 20-24세 기록에서 두 자료 사이의 차이가 두 자료의 인구 기록의 전체 오차의 매우 큰 부분을 차지한다. 왜 이 연령층에서 등록자료와 인구센서스자료상 기록의 차이가 큰지 후속 정밀연구 대상이다.

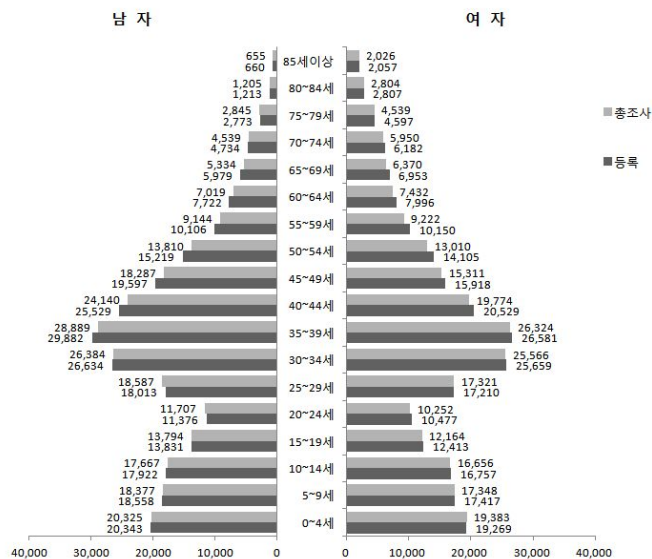
<그림 5-8> 광주 동구의 성 및 연령별 인구수의 오차비율분포



<표 5-9> 화성시의 등록자료와 인구센서스자료의 성별 연령별 구조

화성시	남자		여자	
	등록	총조사	등록	총조사
0~4세	20,343	20,325	19,269	19,383
5~9세	18,558	18,377	17,417	17,348
10~14세	17,922	17,667	16,757	16,656
15~19세	13,831	13,794	12,413	12,164
20~24세	11,376	11,707	10,477	10,252
25~29세	18,013	18,587	17,210	17,321
30~34세	26,634	26,384	25,659	25,566
35~39세	29,882	28,889	26,581	26,324
40~44세	25,529	24,140	20,529	19,774
45~49세	19,597	18,287	15,918	15,311
50~54세	15,219	13,810	14,105	13,010
55~59세	10,106	9,144	10,150	9,222
60~64세	7,722	7,019	7,996	7,432
65~69세	5,979	5,334	6,953	6,370
70~74세	4,734	4,539	6,182	5,950
75~79세	2,773	2,845	4,597	4,539
80~84세	1,213	1,205	2,807	2,804
85세이상	660	655	2,057	2,026

<그림 5-9> 화성시의 등록자료와 인구센서스자료의 성별 연령별 구조



- 화성시는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 역시 행정자료와 인구센서스자료 사이에 차이가 매우 큰 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 30대 후반과 40대의 젊은 성인층 인구가 많은데 반해, 10대와 20대 등 유소년 및 청년층 인구가 매우 적은 지역이다. 도시지역과 농촌지역의 성격이 혼재되어 있는 도시지역의 인구구조를 지니고 있다.
- 화성시의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-10> 화성시의 등록자료와 인구센서스자료의 연령별 성비

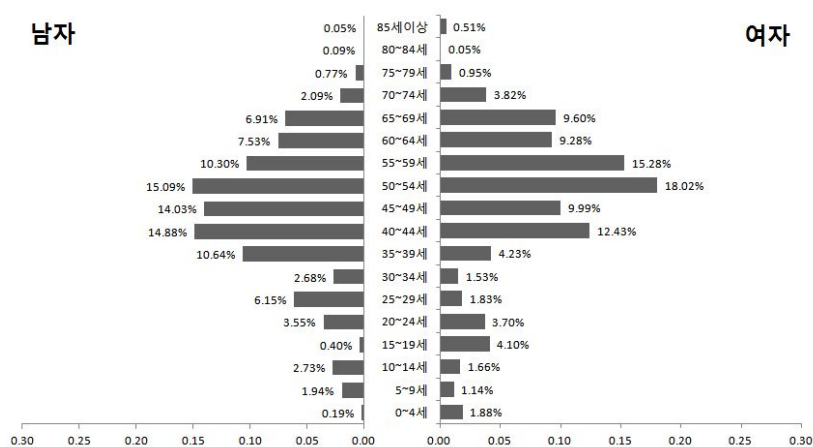
	화성시		
	등록	총조사	성비의 차
0~4세	105.6	104.9	0.7
5~9세	106.6	105.9	0.6
10~14세	107.0	106.1	0.9
15~19세	111.4	113.4	-2.0
20~24세	108.6	114.2	-5.6
25~29세	104.7	107.3	-2.6
30~34세	103.8	103.2	0.6
35~39세	112.4	109.7	2.7
40~44세	124.4	122.1	2.3
45~49세	123.1	119.4	3.7
50~54세	107.9	106.1	1.7
55~59세	99.6	99.2	0.4
60~64세	96.6	94.4	2.1
65~69세	86.0	83.7	2.3
70~74세	76.6	76.3	0.3
75~79세	60.3	62.7	-2.4
80~84세	43.2	43.0	0.2
85세이상	32.1	32.3	-0.2

- 등록자료나 인구센서스 모두, 화성시는 15~19세층에서 남자가 여자보다 많고, 그 다음 40대에서 남자가 여자보다 훨씬 더 많은 성별 구조를 지니고 있다. 그러나 등록자료나 인구센서스자료 모두 연령별 성비 구조에 큰 차이는 없다.

20대 전반에 등록자료보다 인구센서스의 성비가 더 높다. 즉 이 연령층에서 인구센서스상으로는 남자가 여자보다 더 높게 거주하고 있는 것으로 나타난다. 그 외의 연령에서는 등록자료와 인구센서스자료 사이에 연령별 성비 구조는 별 차이없이 양호하게 나타난다.

- 화성시의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면 이 지역에서 등록자료와 인구센서스자료의 차이를 가장 크게 만드는 요인은 40대, 50대, 60대의 중장년층에서의 기록 차이이다. 특히 여성의 경우 50대에서 등록자료와 인구센서스상의 인구기록의 차이가 컸다. 반면에 남자의 경우 40대와 50대 초반의 인구기록이 등록자료와 인구센서스자료 사이에 차이가 큰 것으로 나타난다. 왜 이처럼 남성과 여성의 인구기록에 약간 다른 연령대에서 등록자료와 인구센서스자료 사이에 차이가 나는지, 지역의 특성을 밝히는 후속 정밀연구가 이루어질 필요가 있다.

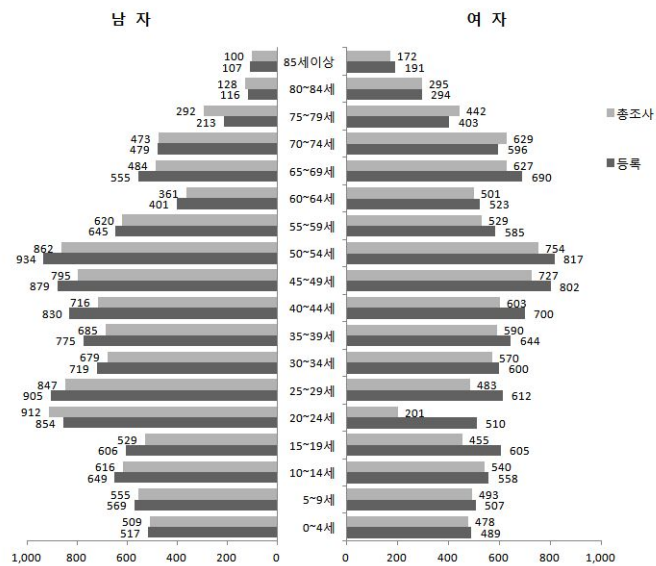
<그림 5-10> 화성시의 성 및 연령별 인구수의 오차비율분포



<표 5-11> 양구군의 등록자료와 인구센서스자료의 성별 연령별 구조

양구군	남자		여자	
	등록	총조사	등록	총조사
0~4세	517	509	489	478
5~9세	569	555	507	493
10~14세	649	616	558	540
15~19세	606	529	605	455
20~24세	854	912	510	201
25~29세	905	847	612	483
30~34세	719	679	600	570
35~39세	775	685	644	590
40~44세	830	716	700	603
45~49세	879	795	802	727
50~54세	934	862	817	754
55~59세	645	620	585	529
60~64세	401	361	523	501
65~69세	555	484	690	627
70~74세	479	473	596	629
75~79세	213	292	403	442
80~84세	116	128	294	295
85세이상	107	100	191	172

<그림 5-11> 양구군의 등록자료와 인구센서스자료의 성별 연령별 구조



- 양구군은 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 차이가 매우 작은 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 50대까지는 남성의 인구가 여성의 인구보다 많고, 60대부터는 여성인구가 많다. 군사지역이 포함되어 20대 남성이 20대 여성보다 월등히 많음이 쉽게 나타난다.
- 양구군의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-12> 양구군의 등록자료와 인구센서스자료의 연령별 성비

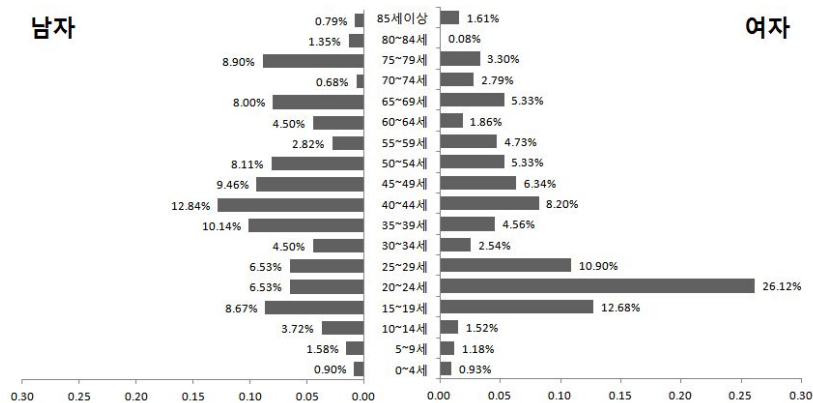
	양구군		
	등록	총조사	성비의 차
0~4세	105.7	106.5	-0.8
5~9세	112.2	112.6	-0.3
10~14세	116.3	114.1	2.2
15~19세	100.2	116.3	-16.1
20~24세	167.5	453.7	-286.3
25~29세	147.9	175.4	-27.5
30~34세	119.8	119.1	0.7
35~39세	120.3	116.1	4.2
40~44세	118.6	118.7	-0.2
45~49세	109.6	109.4	0.2
50~54세	114.3	114.3	0.0
55~59세	110.3	117.2	-6.9
60~64세	76.7	72.1	4.6
65~69세	80.4	77.2	3.2
70~74세	80.4	75.2	5.2
75~79세	52.9	66.1	-13.2
80~84세	39.5	43.4	-3.9
85세이상	56.0	58.1	-2.1

- 양구군처럼 군인이 많은 지역에서 등록자료와 인구센서스자료의 특성이 분명하게 드러난다. 20대는 등록자료의 성비가 남자가 매우 많음을 보여준다. 그러나 인구센서스자료는 남자가 그저 많은 정도가 아니라 20대초반에 성비가 453에 이를 정도로 남자의 인구가 여자 인구의 거의 다섯배에 육박하는 특성

을 지니고 있다. 20대를 제외하고는 70대 후반에 등록자료와 인구센서스자료 사이에 성비의 차이가 크게 나타난다.

- 양구군의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면 이 지역에서 등록자료와 인구센서스자료의 차이를 가장 크게 만드는 요인은 다시 20대 기록에서 발견된다. 20대 남성의 경우 등록자료의 기록과 인구센서스 기록 사이에 상대적으로 큰 차이는 없다. 그러나 20대 여성의 경우 등록자료에 기록된 여성의 기록에 비해 인구센서스의 여성의 기록이 무척 부족하다. 그래서 20대의 성비가 매우 왜곡되어 있는 현상도 보였지만, 등록자료와 인구센서스의 기록의 차이도 20대 여성의 경우 두 자료의 기록에 너무 큰 차이가 있어, 두 자료의 차이를 확대시키는 결과를 빚고 있다. 양구군에서는 10대 후반과 20대의 여성 기록이 등록자료와 인구센서스자료 사이에 차이가 크게 나타나는데 그 이유가 무엇인지 현재로서는 파악하기 어렵다.

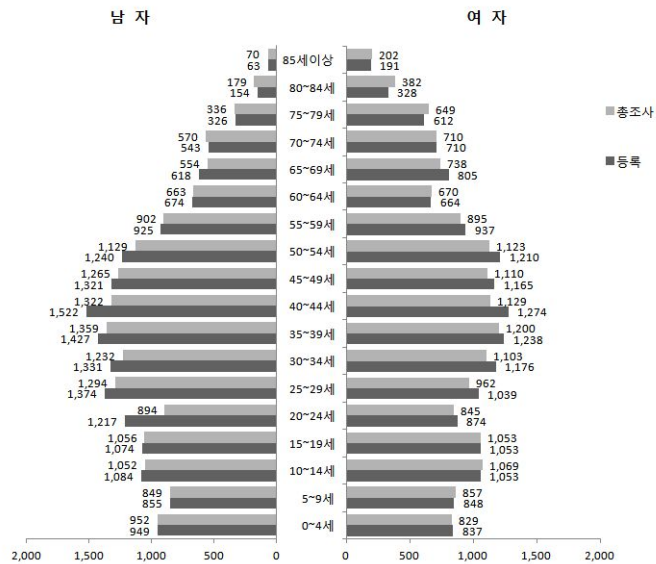
<그림 5-12> 양구군의 성 및 연령별 인구수의 오차비율분포



<표 5-13> 증평군의 등록자료와 인구센서스자료의 성별 연령별 구조

증평군	남자		여자	
	등록	총조사	등록	총조사
0~4세	949	952	837	829
5~9세	855	849	848	857
10~14세	1,084	1,052	1,053	1,069
15~19세	1,074	1,056	1,053	1,053
20~24세	1,217	894	874	845
25~29세	1,374	1,294	1,039	962
30~34세	1,331	1,232	1,176	1,103
35~39세	1,427	1,359	1,238	1,200
40~44세	1,522	1,322	1,274	1,129
45~49세	1,321	1,265	1,165	1,110
50~54세	1,240	1,129	1,210	1,123
55~59세	925	902	937	895
60~64세	674	663	664	670
65~69세	618	554	805	738
70~74세	543	570	710	710
75~79세	326	336	612	649
80~84세	154	179	328	382
85세이상	63	70	191	202

<그림 5-13> 증평군의 등록자료와 인구센서스자료의 성별 연령별 구조



- 증평군은 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 차이가 매우 작은 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 남성의 인구가 여성의 인구보다 많고, 남성
에 비해 성인층에서 여성의 인구가 더 적은 것으로 나타난다. 특히 20대 초반
의 여성의 수가 매우 적어 인구피라미드에서 뚜렷하게 드러나고 있다.
- 양구군의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위
해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록
자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-14> 증평군의 등록자료와 인구센서스자료의 연령별 성비

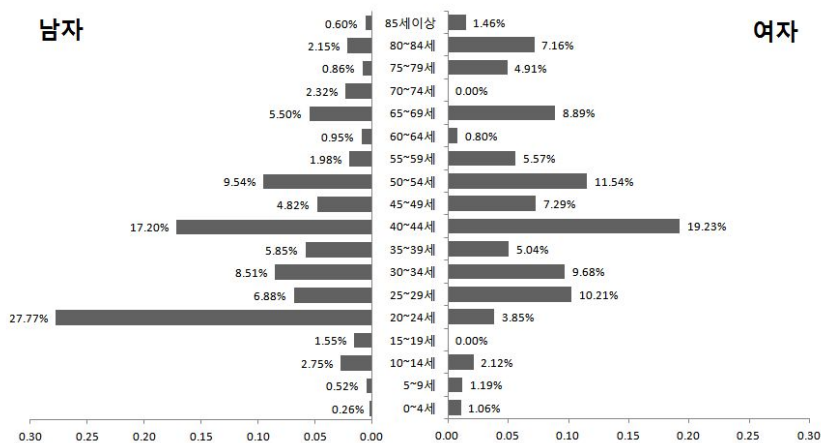
	증평군		
	등록	총조사	성비의 차
0~4세	113.4	114.8	-1.5
5~9세	100.8	99.1	1.8
10~14세	102.9	98.4	4.5
15~19세	102.0	100.3	1.7
20~24세	139.2	105.8	33.4
25~29세	132.2	134.5	-2.3
30~34세	113.2	111.7	1.5
35~39세	115.3	113.3	2.0
40~44세	119.5	117.1	2.4
45~49세	113.4	114.0	-0.6
50~54세	102.5	100.5	1.9
55~59세	98.7	100.8	-2.1
60~64세	101.5	99.0	2.6
65~69세	76.8	75.1	1.7
70~74세	76.5	80.3	-3.8
75~79세	53.3	51.8	1.5
80~84세	47.0	46.9	0.1
85세이상	33.0	34.7	-1.7

- 증평군의 인구피라미드에서 20~24세 여성의 수가 여성내에서 다른 연령, 또
남성과 비교해도 숫적으로 적은 현상이 관찰되었는데, 등록자료와 인구센서스
의 연령별 성비를 비교해보아도, 20대 초반의 두 자료의 성비차가 가장 크게
나타나고 있다. 이 연령대의 성비가 등록자료상으로는 139인데 반해, 인구센

서스상으로는 114이다. 등록자료상으로 왜 여성들이 과소 등록되어 있고, 상대적으로 남성이 과대등록되어 있는 것으로 나타나는지 현재 그 이유를 정확히 알 수는 없다.

- 증평균 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면, 예상과는 달리 20대 초반 남성의 인구기록이 등록자료와 인구센서스자료 사이에 가장 크게 차이가 나고 있음이 보인다. 남성의 이 연령층에서 두 자료사이의 차이가 남성의 두 자료의 전체 차이 가운데 차지하는 비중이 제일 높다. 이 연령층을 제외하면 40대 초반에 남성과 여성 모두 등록자료와 인구센서스자료 사이에 오차가 크게 나타난다. 모두 등록자료 상으로는 많은 인구가 기록되어 있는데, 인구센서스상으로는 거주하지 않는 인구가 많은 것으로 나타난다. 이 연령층에서 왜 이처럼 두 자료 사이에 차이가 나는지 현재로서는 정확히 말하기 어렵다. 50대 초반의 연령에서도 남성과 여성 모두 두 자료 사이의 기록에 차이가 있고, 전체 오차의 상당부분을 차지한다.

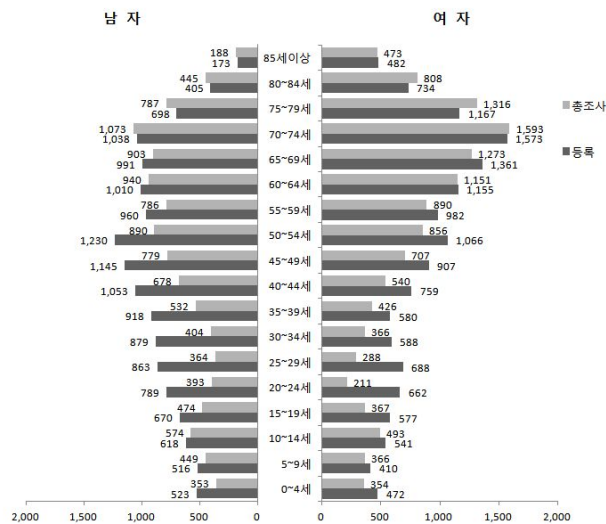
<그림 5-14> 증평균의 성 및 연령별 인구수의 오차비율분포



<표 5-15> 임실군의 등록자료와 인구센서스자료의 성별 연령별 구조

임실군	남자		여자	
	등록	총조사	등록	총조사
0~4세	523	353	472	354
5~9세	516	449	410	366
10~14세	618	574	541	493
15~19세	670	474	577	367
20~24세	789	393	662	211
25~29세	863	364	688	288
30~34세	879	404	588	366
35~39세	918	532	580	426
40~44세	1,053	678	759	540
45~49세	1,145	779	907	707
50~54세	1,230	890	1,066	856
55~59세	960	786	982	890
60~64세	1,010	940	1,155	1,151
65~69세	991	903	1,361	1,273
70~74세	1,038	1,073	1,573	1,593
75~79세	698	787	1,167	1,316
80~84세	405	445	734	808
85세이상	173	188	482	473

<그림 5-15> 임실군의 등록자료와 인구센서스자료의 성별 연령별 구조



- 임실군은 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 차이가 매우 작은 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 젊은 층과 여성의 인구가 적고 고령화가 심화된 인구구조를 보여준다. 특히 등록자료보다 인구센서스자료에 기록된 인구가 훨씬 적음이 한 눈에 보인다. 여성은 70대 인구가 높은 비중을 차지하고 있다.
- 임실군의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-16> 임실군의 등록자료와 인구센서스자료의 연령별 성비

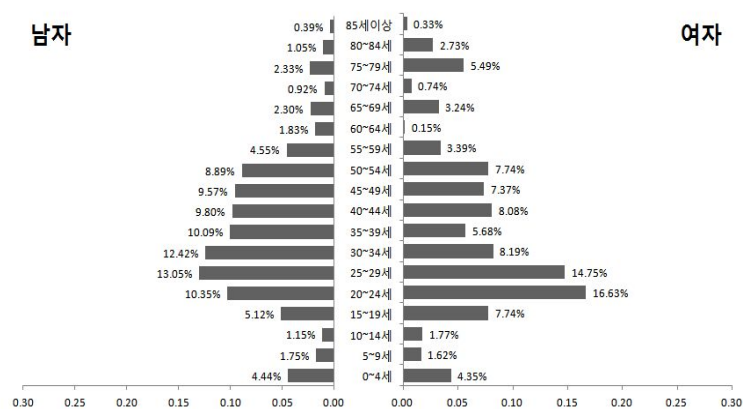
	임실군		
	등록	총조사	성비의 차
0~4세	110.8	99.7	11.1
5~9세	125.9	122.7	3.2
10~14세	114.2	116.4	-2.2
15~19세	116.1	129.2	-13.0
20~24세	119.2	186.3	-67.1
25~29세	125.4	126.4	-1.0
30~34세	149.5	110.4	39.1
35~39세	158.3	124.9	33.4
40~44세	138.7	125.6	13.2
45~49세	126.2	110.2	16.1
50~54세	115.4	104.0	11.4
55~59세	97.8	88.3	9.4
60~64세	87.4	81.7	5.8
65~69세	72.8	70.9	1.9
70~74세	66.0	67.4	-1.4
75~79세	59.8	59.8	0.0
80~84세	55.2	55.1	0.1
85세이상	35.9	39.7	-3.9

- 임실군에서 등록자료와 인구센서스의 연령별 성비 기록에 가장 큰 차이가 나는 연령대는 20대 초반이다. 등록자료상으로는 이 연령대 성비가 119로 성비가 약간 왜곡된 것처럼 보이는데, 인구센서스상으로는 이 연령대 성비가 무려

186이나 되어, 여성이 턱없이 부족한 상황임을 알 수 있다. 그러나 성비만으로 보면 등록자료상 성비가 인구센서스자료에 나타난 성비보다 대체로 훨씬 높음을 알 수 있다. 그래서 30대에도 등록자료상의 성비가 훨씬 높아, 등록자료의 성비와 인구센서스의 성비의 차이를 구해보면 이 연령대에서도 두 자료 사이에 성비의 차이가 크게 남을 알 수 있다..

- 임실군 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면, 20대부터 50대초반까지 등록자료와 인구센서스자료와의 기록차이가 크다. 그런데 여성의 경우 20대 전후반에 등록자료와 인구센서스자료 상의 기록차이가 커서 여성기록의 오차의 30% 이상이 20대 여성의 등록자료와 인구센서스자료 사이의 차이에서 비롯되고 있는 것으로 나타난다.

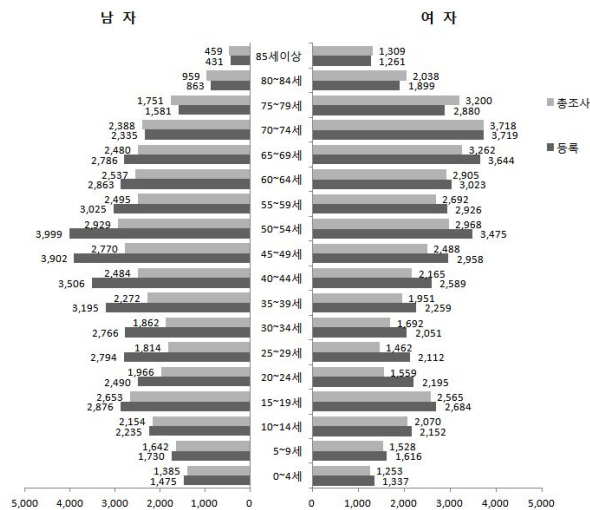
<그림 5-16> 임실군의 성 및 연령별 인구수의 오차비율분포



<표 5-17> 나주시의 등록자료와 인구센서스자료의 성별 연령별 구조

나주시	남자		여자	
	등록	총조사	등록	총조사
0~4세	1,475	1,385	1,337	1,253
5~9세	1,730	1,642	1,616	1,528
10~14세	2,235	2,154	2,152	2,070
15~19세	2,876	2,653	2,684	2,565
20~24세	2,490	1,966	2,195	1,559
25~29세	2,794	1,814	2,112	1,462
30~34세	2,766	1,862	2,051	1,692
35~39세	3,195	2,272	2,259	1,951
40~44세	3,506	2,484	2,589	2,165
45~49세	3,902	2,770	2,958	2,488
50~54세	3,999	2,929	3,475	2,968
55~59세	3,025	2,495	2,926	2,692
60~64세	2,863	2,537	3,023	2,905
65~69세	2,786	2,480	3,644	3,262
70~74세	2,335	2,388	3,719	3,718
75~79세	1,581	1,751	2,880	3,200
80~84세	863	959	1,899	2,038
85세이상	431	459	1,261	1,309

<그림 5-17> 나주시의 등록자료와 인구센서스자료의 성별 연령별 구조



- 나주시는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 차이가 중간 정도되는 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 50대 연령이 제일 많고, 나이가 적을수록 인구가 줄어든다, 15세부터 19세까지의 인구가 전체 인구 가운데 비교적 높은 비율을 차지하고 있다. 여성인구는 매우 고령화되어 있다.
- 나주시의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-18> 나주시의 등록자료와 인구센서스자료의 연령별 성비

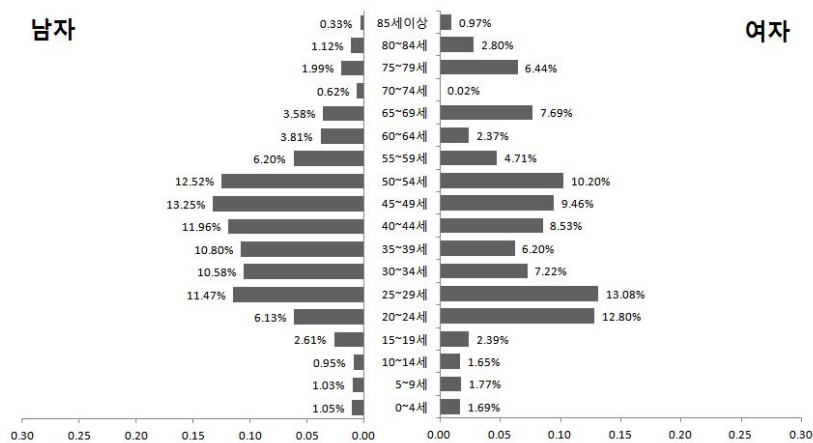
	나주시		
	등록	총조사	성비의 차
0~4세	110.3	110.5	-0.2
5~9세	107.1	107.5	-0.4
10~14세	103.9	104.1	-0.2
15~19세	107.2	103.4	3.7
20~24세	113.4	126.1	-12.7
25~29세	132.3	124.1	8.2
30~34세	134.9	110.0	24.8
35~39세	141.4	116.5	25.0
40~44세	135.4	114.7	20.7
45~49세	131.9	111.3	20.6
50~54세	115.1	98.7	16.4
55~59세	103.4	92.7	10.7
60~64세	94.7	87.3	7.4
65~69세	76.5	76.0	0.4
70~74세	62.8	64.2	-1.4
75~79세	54.9	54.7	0.2
80~84세	45.4	47.1	-1.6
85세이상	34.2	35.1	-0.9

- 나주시에서 등록자료와 인구센서스의 연령별 성비 기록에 가장 큰 차이가 나는 연령대는 30대, 40대, 50대이다. 이 연령대에서 인구센서스자료 상으로는 남성이 여성보다 물론 많지만, 등록자료상의 성별기록의 왜곡보다는 덜하다. 그 결과 등록자료의 성비와 인구센서스자료의 성비의 차이가 30대부터 50대

사이에 가장 높게 나타난다. 왜 이처럼 등록자료상으로 남녀의 기록이 크게 차이가 나고, 인구센서스자료와 비교해도 성비의 차이가 이 연령대에서 크게 나타나는지 앞으로 더 심층분석이 필요하다.

- 나주시 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면, 여성의 경우 20대의 기록에 등록자료와 인구센서스자료 사이의 오차가 크다. 그 다음은 40대 후반과 50대 초반의 기록에 두 자료의 차이가 크다. 남성의 경우에는 20대 후반부터 50대 초반까지 거의 모든 연령대에서 등록자료와 인구센서스자료 사이의 기록에 차이가 고르게 크게 나타나고 있다.

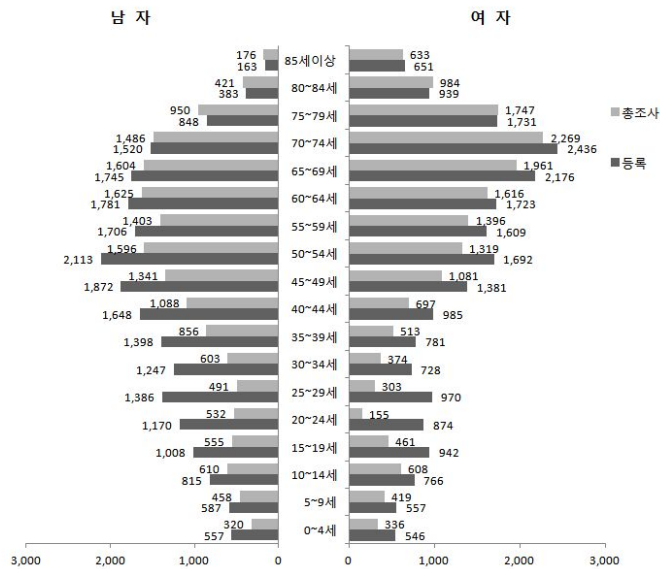
<그림 5-18> 나주시의 성 및 연령별 인구수의 오차비율분포



<표 5-19> 신안군의 등록자료와 인구센서스자료의 성별 연령별 구조

신안군	남자		여자	
	등록	총조사	등록	총조사
0~4세	557	320	546	336
5~9세	587	458	557	419
10~14세	815	610	766	608
15~19세	1,008	555	942	461
20~24세	1,170	532	874	155
25~29세	1,386	491	970	303
30~34세	1,247	603	728	374
35~39세	1,398	856	781	513
40~44세	1,648	1,088	985	697
45~49세	1,872	1,341	1,381	1,081
50~54세	2,113	1,596	1,692	1,319
55~59세	1,706	1,403	1,609	1,396
60~64세	1,781	1,625	1,723	1,616
65~69세	1,745	1,604	2,176	1,961
70~74세	1,520	1,486	2,436	2,269
75~79세	848	950	1,731	1,747
80~84세	383	421	939	984
85세이상	163	176	651	633

<그림 5-19> 신안군의 등록자료와 인구센서스자료의 성별 연령별 구조



- 신안군은 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 차이가 별로 나지 않는 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 젊은층의 인구는 매우 적고 대신 나이 들수록 인구규모가 커지는, 전형적인 고령화된 농촌의 인구구조를 지니고 있다. 등록자료상의 인구보다 인구센서스상의 인구가 적은 예를 보여준다.
- 신안군의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-20> 신안군의 등록자료와 인구센서스자료의 연령별 성비

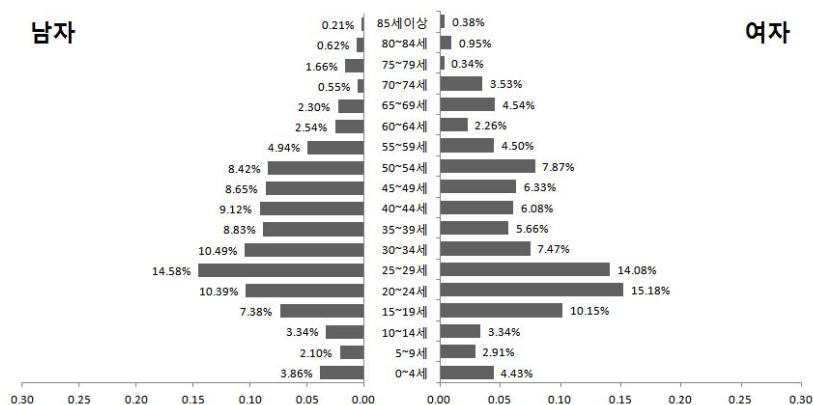
	신안군		
	등록	총조사	성비의 차
0~4세	102.0	95.2	6.8
5~9세	105.4	109.3	-3.9
10~14세	106.4	100.3	6.1
15~19세	107.0	120.4	-13.4
20~24세	133.9	343.2	-209.4
25~29세	142.9	162.0	-19.2
30~34세	171.3	161.2	10.1
35~39세	179.0	166.9	12.1
40~44세	167.3	156.1	11.2
45~49세	135.6	124.1	11.5
50~54세	124.9	121.0	3.9
55~59세	106.0	100.5	5.5
60~64세	103.4	100.6	2.8
65~69세	80.2	81.8	-1.6
70~74세	62.4	65.5	-3.1
75~79세	49.0	54.4	-5.4
80~84세	40.8	42.8	-2.0
85세이상	25.0	27.8	-2.8

- 신안군에서 등록자료와 인구센서스의 연령별 성비 기록에 가장 큰 차이가 나는 연령대는 20대 전반이다. 이 연령대에서 등록자료상으로는 성비가 133.9로 남성이 여성보다 많다. 그러나 실제 인구를 반영하는 인구센서스자료 상으로는 성비가 무려 343.2로 나타난다. 배를 타는 젊은 남성이 많은 것인지, 아니

면 염전 등의 산업에 종사하는 외지 남성들이 많아서 그런 것인지 확실하지 않다. 그 외의 연령에서도 등록자료의 성비와 인구센서스상의 성비가 차이가 많이 나고 있다.

- 신안군의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면, 여성의 경우 등록자료와 인구센서스자료의 기록에 20대에서 차이가 제일 크게 나타난다. 20대에서 나타나는 오차의 크기가 여성 기록에서 나타나는 전체 오차의 30%를 차지한다. 남성의 경우에도 20대 후반의 기록에서 등록자료와 인구센서스자료의 차이가 크다. 남성의 경우 20대 후반 이상의 성인 연령층에서도 등록자료와 인구센서스자료 사이에 오차가 거의 일정한 비율로 존재한다. 고령연령층과 유소년연령층을 제외한 성인 연령층에서 골고루 분포의 오차가 존재하는 셈이다. 여성의 경우에는 20대에 오차가 좀더 집중되어 있어, 신안군의 경우에도 성별로 등록자료와 인구센서스자료 사이에 오차가 다르게 나타나는 특성이 있음을 알 수 있다. 신안군에서 젊은 여성은 등록자료상으로는 존재하지만 실제로는 존재하지 않는 현실이 여기에 나타난다.

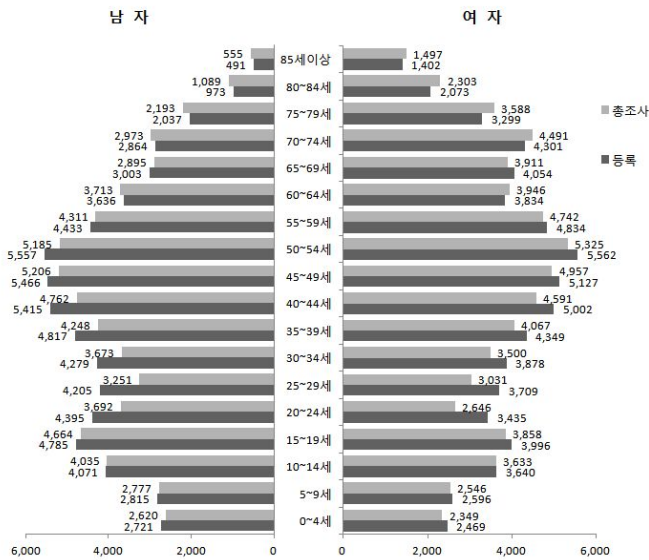
<그림 5-20> 신안군의 성 및 연령별 인구수의 오차비율분포



<표 5-21> 김천시의 등록자료와 인구센서스자료의 성별 연령별 구조

김천시	남자		여자	
	등록	총조사	등록	총조사
0~4세	2,721	2,620	2,469	2,349
5~9세	2,815	2,777	2,596	2,546
10~14세	4,071	4,035	3,640	3,633
15~19세	4,785	4,664	3,996	3,858
20~24세	4,395	3,692	3,435	2,646
25~29세	4,205	3,251	3,709	3,031
30~34세	4,279	3,673	3,878	3,500
35~39세	4,817	4,248	4,349	4,067
40~44세	5,415	4,762	5,002	4,591
45~49세	5,466	5,206	5,127	4,957
50~54세	5,557	5,185	5,562	5,325
55~59세	4,433	4,311	4,834	4,742
60~64세	3,636	3,713	3,834	3,946
65~69세	3,003	2,895	4,054	3,911
70~74세	2,864	2,973	4,301	4,491
75~79세	2,037	2,193	3,299	3,588
80~84세	973	1,089	2,073	2,303
85세이상	491	555	1,402	1,497

<그림 5-21> 김천시의 등록자료와 인구센서스자료의 성별 연령별 구조



- 김천시는 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 중간 정도의 차이가 나는 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 30대의 젊은 성인의 인구가 적고, 남성보다는 여성의 고령화가 더 심한 인구구조를 보여주고 있다. 일건 등록자료와 인구센서스자료 상에 큰 차이가 없는 것처럼 보인다.
- 김천시의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-22> 김천시의 등록자료와 인구센서스자료의 연령별 성비

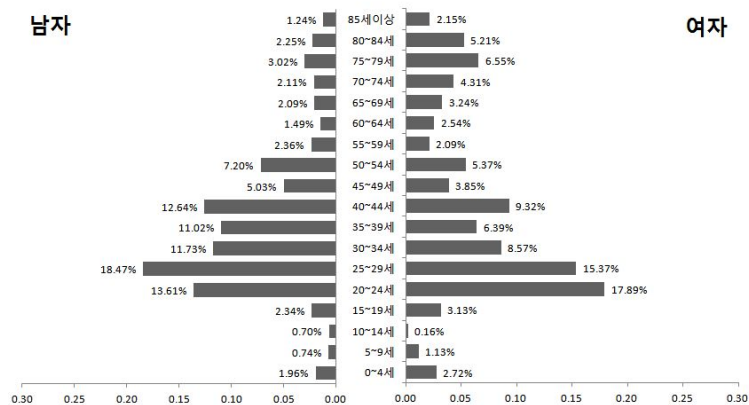
	김천시		
	등록	총조사	성비의 차
0~4세	110.2	111.5	-1.3
5~9세	108.4	109.1	-0.6
10~14세	111.8	111.1	0.8
15~19세	119.7	120.9	-1.1
20~24세	127.9	139.5	-11.6
25~29세	113.4	107.3	6.1
30~34세	110.3	104.9	5.4
35~39세	110.8	104.5	6.3
40~44세	108.3	103.7	4.5
45~49세	106.6	105.0	1.6
50~54세	99.9	97.4	2.5
55~59세	91.7	90.9	0.8
60~64세	94.8	94.1	0.7
65~69세	74.1	74.0	0.1
70~74세	66.6	66.2	0.4
75~79세	61.7	61.1	0.6
80~84세	46.9	47.3	-0.3
85세이상	35.0	37.1	-2.1

- 김천시는 위에서 검토한 농촌지역에 비해서 등록자료와 인구센서스자료의 성비 사이에 그리 큰 차이가 보이지 않는다. 20대 초반에 등록자료의 성비보다 인구센서스자료의 성비가 더 높고, 이 성비 차이는 어느 연령대보다 높은 것으로 나타난다. 그러나 다른 연령대에서는 두 자료의 성비 사이에 그렇게 큰

차이가 나타나지 않는다.

- 김천시의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면, 등록자료와 인구센서스자료 사이의 기록에 나타나는 차이는 비교적 젊은 연령의 기록에 기인한다. 남성과 여성 모두 20대 초반과 20대 후반 연령대에서 등록자료와 인구센서스자료의 인구기록에 차이가 크고, 이 차이는 성별 두 기록상의 전체 차이에서 가장 높은 비중을 차지한다. 남성의 경우에는 30대와 40대의 기록도 두 자료사이에 기록의 차이가 큰 편이고, 여성의 경우는 70대 이상의 고령층에서 두 자료의 기록의 차이가 큰 것으로 나타난다.

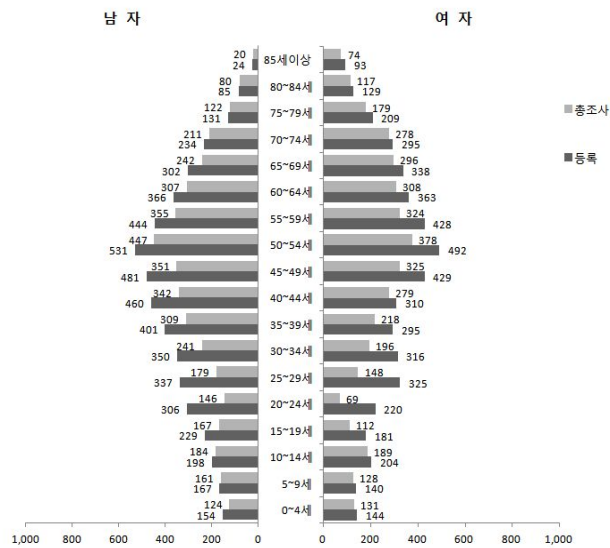
<그림 5-22> 김천시의 성 및 연령별 인구수의 오차비율분포



<표 5-23> 울릉군의 등록자료와 인구센서스자료의 성별 연령별 구조

울릉군	남자		여자	
	등록	총조사	등록	총조사
0~4세	154	124	144	131
5~9세	167	161	140	128
10~14세	198	184	204	189
15~19세	229	167	181	112
20~24세	306	146	220	69
25~29세	337	179	325	148
30~34세	350	241	316	196
35~39세	401	309	295	218
40~44세	460	342	310	279
45~49세	481	351	429	325
50~54세	531	447	492	378
55~59세	444	355	428	324
60~64세	366	307	363	308
65~69세	302	242	338	296
70~74세	234	211	295	278
75~79세	131	122	209	179
80~84세	85	80	129	117
85세이상	24	20	93	74

<그림 5-23> 울릉군의 등록자료와 인구센서스자료의 성별 연령별 구조



- 울릉군은 본 연구의 제1장 이상 시군구 탐색연구의 지표 중 SWORD 면에서 행정자료와 인구센서스자료 사이에 별로 차이가 나지 않는 지역의 하나로 선정된 지역이다. 인구피라미드 상으로는 남성보다 여성의 인구유출이 더 많은 것으로 보이는데, 중장년 연령층의 인구가 많고, 고령층과 유소년층의 인구는 상대적으로 적은 인구구조를 보인다.
- 울릉군의 행정자료와 인구센서스자료의 성별 연령별 기록 특성을 알아보기 위해 등록자료와 인구센서스자료의 연령별 성비를 구하였고, 각 연령별로 등록자료의 성비와 인구센서스자료의 성비의 차를 구하였다.

<표 5-24> 울릉군의 등록자료와 인구센서스자료의 연령별 성비

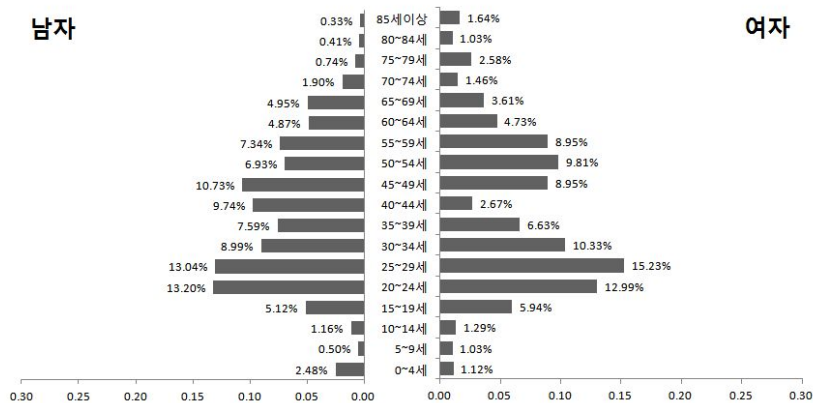
	울릉군		
	등록	총조사	성비의 차
0~4세	106.9	94.7	12.3
5~9세	119.3	125.8	-6.5
10~14세	97.1	97.4	-0.3
15~19세	126.5	149.1	-22.6
20~24세	139.1	211.6	-72.5
25~29세	103.7	120.9	-17.3
30~34세	110.8	123.0	-12.2
35~39세	135.9	141.7	-5.8
40~44세	148.4	122.6	25.8
45~49세	112.1	108.0	4.1
50~54세	107.9	118.3	-10.3
55~59세	103.7	109.6	-5.8
60~64세	100.8	99.7	1.2
65~69세	89.3	81.8	7.6
70~74세	79.3	75.9	3.4
75~79세	62.7	68.2	-5.5
80~84세	65.9	68.4	-2.5
85세이상	25.8	27.0	-1.2

- 울릉군은 등록자료의 성비와 인구센서스자료의 성비 사이에 차이가 많은 연령층이 다수 존재한다. 두 자료의 성비 사이에 가장 차이가 큰 연령은 20~24세 층이다. 이 연령층에서 등록자료의 성비는 139인데, 인구센서스자료의 성비는 211.6으로 성비의 차이가 72.5나 된다. 반면에 40~44세 연령층은 등록자료 상

으로는 성비가 148.4인데, 인구센서스자료 상으로는 성비가 122.6으로 나타난다. 젊은층에서는 등록자료상으로는 기록되어 있으나 실제로는 여성인구의 유출이 있는 것으로 보인다. 반면에 장년층에서는 실제로는 거주하나 등록은 다른 곳에 하고 있는 인구가 울릉군에 존재하는 것으로 보인다. 이 또한 성별, 연령별, 지역별 특성이 일관성없게 존재하는 한 예를 보여준다.

- 울릉군의 등록자료와 인구센서스자료를 비교하기 위해 성별로 연령별 인구수 기록의 오차비율의 분포를 보면, 여성들의 경우 20대 및 30대 초반에 두 자료 사이의 기록 차이가 큼을 알 수 있다. 이는 남성도 마찬가지이다. 남성은 20대의 젊은층에서 두 자료 사이의 기록에 차이가 크다. 남성은 40대에도 두 자료의 기록의 차이가 큰데, 여성은 40대 초반에는 등록자료와 인구센서스자료 사이의 기록에 상대적으로 차이가 적다. 울릉군도 성별로 등록자료와 인구센서스자료의 기록 사이에 차이가 있고, 연령별로도 젊은 연령층 및 중장년층에서도 기록상의 차이가 발생하고 있음을 알 수 있다. 그러나 이러한 기록의 차이가 일관성있는 모습으로 보이지 않아 이 현상을 어떻게 해석해야 하는지 추후의 분석과 연구가 필요하다.

<그림 5-24> 울릉군의 성 및 연령별 인구수의 오차비율분포



2. 2010년 센서스와 행정자료의 가구

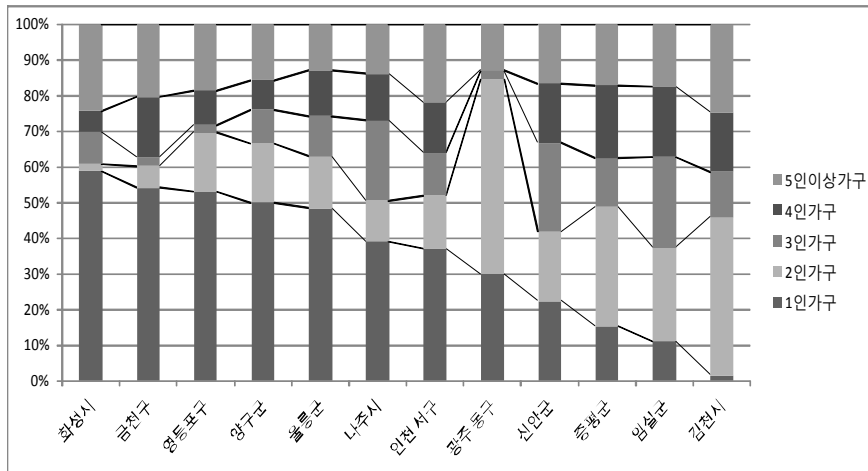
- 이 절에서는 행정자료와 인구센서스자료를 이용해 가구측면에서 행정자료의 성격을 이해하고자 한다. 인구자료에 비해 가구자료는 매우 단순하다. 인구자료는 지역별로 성, 연령구조를 지니고 있고, 성별로 성비를 구해 비교하고, 연령별 성비를 비교하며, 인구규모를 비교하는 등 여러 측면에서 비교해야 할 측면이 있다.
- 그러나 이에 반해 가구는 분포가 매우 단순하다. 가구의 분포에는 성 및 연령이라는 변수가 동원되지 않고, 단순히 지역별 가구원수에 따른 분포만을 볼 수 있을 뿐이다. 본 분석에서도 12개 지역별로 등록자료와 인구센서스자료 사이에 가구원수별 가구분포가 어떻게 같고 차이나는지를 검토하여 등록자료의 성격을 지역별로 파악하려고 한다.
- 다음 <표 5-25>는 지역별로 등록자료와 인구센서스자료의 가구원수별 분포의 차이가 어디서 발생하는지를 보여준다. 이 표는 먼저, 12개 지역별로 등록자료의 가구원수별 분포와 인구센서스자료의 가구원수별 분포를 이용한다. 그 다음 등록자료의 각 가구원수별 비율에서 인구센서스자료의 각 가구원수별 비율을 뺀다. 즉 비율의 차이를 구한다. 그 다음 모든 비율의 차이를 합한다. 그 후 각 가구원수별 가구범주에서 두 자료 분포 사이의 비율의 차이가 전체 비율의 차이에서 차지하는 비율을 구한다.
- 이런 방식을 통해 구한 수치는 등록자료의 가구원수별 가구분포가 인구센서스의 가구원수별 가구분포와 비교할 때 어느 가구원수별 가구 때문에 차이가 많이 나는지를 알려준다. 또한 1인가구 혹은 비 1인가구에 따른 차이인지를 확인하면 최근 증가하는 1인가구의 영향 및 도시와 농촌지역의 1인가구의 영향이 등록자료와 인구센서스자료의 성격을 다르게 만드는데 기여하고 있는지 알 수 있을 것이다.

<표 5-25> 지역별 등록자료와 인구센서스자료의 가구원수별 가구 분포의 차이

	1인가구	2인가구	3인가구	4인가구	5인이상가구
금천구	54.21%	6.14%	2.39%	16.96%	20.30%
영등포구	53.09%	16.50%	2.13%	9.74%	18.54%
인천서구	37.04%	15.06%	11.87%	14.13%	21.90%
광주동구	29.99%	54.67%	2.33%	0.08%	12.93%
화성시	58.95%	1.90%	8.96%	6.00%	24.19%
양구군	50.06%	16.57%	9.46%	8.29%	15.62%
증평군	15.30%	33.66%	13.52%	20.43%	17.08%
임실군	11.20%	26.17%	25.61%	19.51%	17.51%
나주시	39.23%	11.42%	22.47%	13.04%	13.84%
신안군	22.52%	19.30%	24.90%	16.73%	16.55%
김천시	1.67%	44.20%	12.70%	16.82%	24.60%
울릉군	48.47%	14.59%	11.32%	12.75%	12.88%

- 인구와 마찬가지로 가구수도 등록자료와 인구센서스자료에 따라 다르다. 게다가 가구원수별 가구수 및 그 분포 또한 등록자료와 인구센서스자료에 따라 다르다. 게다가 도시와 농촌이라는 지역을 고려하면 지역별로 등록자료의 가구수 및 가구분포와 인구센서스자료의 가구수 및 가구분포 사이에 차이가 있을 것은 명약관화하다. 이제 우리는 이 절에서 가구를 가구원수로 구분할 때 지역별로 등록자료와 인구센서스자료 사이에는 어떤 차이가 있는지를 밝혀보려고 한다.
- 1인가구의 수가 등록자료와 인구센서스자료 사이에 크게 달라 등록자료와 인구센서스자료의 가구수 및 가구분포의 차이에 1인가구가 미치는 영향이 전체 오차의 50%를 넘어서는 지역은 전체 12개 지역 가운데 서울 금천구, 서울 영등포구, 화성시, 양구군 등 네 지역이다. 양구군을 제외하면 모두 대도시 및 수도권에 있는 도시지역이다. 대도시 지역에서 증가하고 있는 1인가구는 거주지에 등록이 안되어 있으면서 실제로는 존재하고 있는 현실을 말해주고 있다. 따라서 등록자료를 이용해서 가구원수별 가구분포를 만들 때는 대도시 지역에서 1인가구를 어떻게 파악하고 대처할 것인지가 관건임을 보여준다.

<그림 5-25> 지역별 등록자료와 인구센서스자료의 가구원수별 가구 분포의 차이



- 반면에 김천시는 1인가구가 등록자료와 인구센서스자료의 가구수 및 가구원수 별 가구 분포의 차이에 미치는 영향이 극히 미미하다. 실제로는 거의 없다고 말할 수준이다. 김천시를 제외하면 심지어는 농촌지역이라 할지라도 1인가구의 존재가 등록자료와 인구센서스자료 사이의 분포 차이를 낳는 중요한 역할을 하는데, 김천시에서는 이런 현상이 전혀 보이지 않는다. 어떻게 이런 현상이 김천시에서 발생하는지 추후 방문연구 등을 통해 밝혀낼 필요가 있다.
- 광주 동구는 2인가구의 측면에서 등록자료와 인구센서스자료 사이에 큰 괴리가 있는 것으로 나타난다. 가구원수별 분포의 전체 오차 가운데 50% 이상이 광주 동구의 경우에는 2인가구에서 비롯되고 있다. 김천시와 증평군도 2인가구의 분포의 차이가 등록자료와 인구센서스자료의 가구 자료의 차이를 낳는 중요한 요소임을 보여주고 있다. 김천시는 1인가구는 자료의 차이를 낳는 중요한 요소가 아닌데, 2인가구는 두 자료의 분포 차이를 낳는 중요한 요소이다. 왜 이런 현상이 발생하는지 몹시 흥미롭다. 그러나 현재 이 현상을 적절히 설명할 수는 없고, 추후 연구가 꼭 필요함을 보여준다.

- 하나의 범주로 묶여 있는 5인 이상의 가구도 역시 두 자료의 차이를 만드는 주요한 요소이다. 그러나 현재 5인 이상의 가구원이면 하나의 범주로 묶어 분석하고 있기 때문에 구체적으로 어느 지역에서 어떤 이유로 등록자료와 인구센서스자료 사이에 차이가 나는지 말하기 힘들다.
- 요약하면, 전 지역, 특히 대도시 지역을 중심으로 등록자료의 가구원수별 가구와 인구센서스자료의 가구원수별 가구 사이에 차이가 존재한다. 최근 급속도로 증가하고 있는 1인가구는 등록되지 않고, 실제로 존재하는 가구일 가능성이 높다. 이 경우 행정적으로 어떻게 1인가구를 정확히 파악할 것인가가, 등록자료의 가구원수별 가구 기록을 신뢰하고 이용할 수 있는 관건이 될 전망이다.

3. 2010년 센서스와 행정자료의 주택

- 이 절에서는 행정자료와 인구센서스자료를 이용해 주택 측면에서 행정자료의 성격을 이해하고자 한다. 주택 자료 또한 인구자료에 비해 매우 단순하다. 인구자료는 등록자료와 인구센서스자료를 비교하면서 등록자료의 성격을 이해할 여러 독립변수가 많이 존재한다. 특히 성별, 연령별 인구구조는 매우 중요하다. 그러나 주택자료의 경우, 가구자료와 마찬가지로 성별, 연령별 변수도 없고, 자료의 품질을 평가할 기준이 비교적 단순하다. 이 절에서는 주택의 유형 및 주택의 건축연도 자료를 중심으로 12개 지역의 등록자료와 인구센서스자료의 주택자료의 품질을 평가하려고 한다.
- 다음 <표 5-26>은 지역별로 등록자료와 인구센서스자료의 주택유형별 분포의 차이가 어디서 발생하는지를 보여준다. 이 표는 먼저, 12개 지역별로 등록자료의 주택유형별 분포와 인구센서스자료의 주택유형별 분포를 이용한다. 그 다음 등록자료의 각 주택유형별 비율에서 인구센서스자료의 각 주택유형별 비율을 뺀다. 즉 비율의 차이를 구한다. 그 다음 모든 비율의 차이를 합한다. 그 후 각 주택유형별 주택범주에서 두 자료 분포 사이의 비율의 차이가 전체 비

율의 차이에서 차지하는 비율을 구한다.

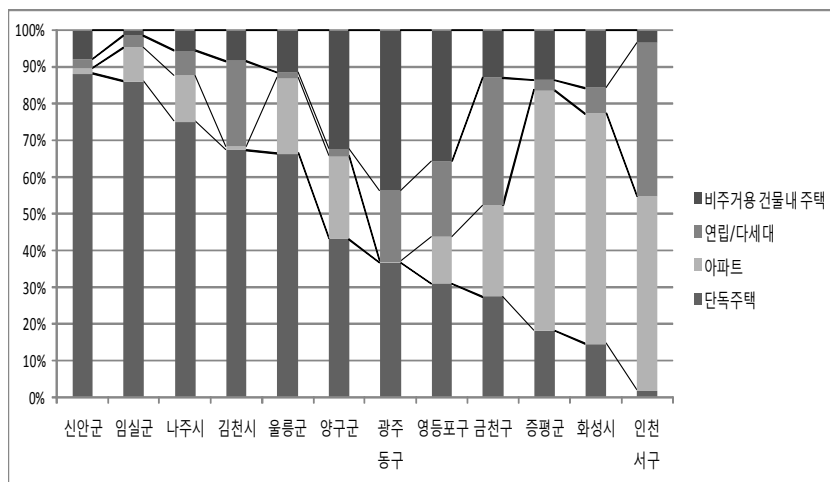
<표 5-26> 지역별 등록자료와 인구센서스자료의 주택유형별 주택 분포의 차이

	단독주택	아파트	연립/다세대	비주거용 건물내 주택
금천구	27.51%	24.82%	34.67%	12.99%
영등포구	31.06%	12.73%	20.65%	35.56%
인천서구	1.93%	52.81%	41.74%	3.52%
광주동구	36.71%	0.32%	19.14%	43.83%
화성시	14.57%	62.74%	7.04%	15.65%
양구군	43.16%	22.62%	1.71%	32.51%
증평군	18.26%	65.26%	3.01%	13.47%
임실군	85.95%	9.27%	3.40%	1.37%
나주시	75.04%	12.70%	6.78%	5.48%
신안군	88.11%	1.58%	2.38%	7.92%
김천시	67.40%	0.95%	23.23%	8.42%
울릉군	66.41%	20.46%	1.74%	11.39%

- <표 5-26>은 지역별로 등록자료와 인구센서스자료의 주택유형별 분포의 차이를 보여준다. 단독주택의 수와 분포가 등록자료와 인구센서스자료의 주택분포에 미치는 영향은 극과 극이다. 신안군의 경우에는 등록자료상 파악되는 단독주택과 인구센서스에서 파악되는 단독주택의 수가 매우 차이가 난다. 그래서 주택유형별 분포에서 두 자료상의 차이에 미치는 영향이 신안군의 경우 단독주택의 경우 88.1%나 된다. 신안군의 주택분포에서 등록자료와 인구센서스자료 사이에 차이가 있다면 그 대부분의 원인이 단독주택에 있다는 뜻이다. 임실군도 신안군 못지 않게 두 자료의 주택분포의 차이 가운데 86%가 단독주택에서 비롯된다. 나주시, 김천시, 울릉군 등도 단독주택의 영향이 크다.
- 반면에 인천 서구는 단독주택이 등록자료와 인구센서스자료의 주택분포의 차이에 미치는 영향은 극히 미미하다. 인천 서구의 경우 두 자료의 분포에서 파악되는 전체 오차의 1.9%만이 단독주택에 기인한다. 화성시, 금천구, 영등포구, 광주동고 등도 비교적 단독주택의 영향이 작은 지역이다.
- 다른 한 편, 아파트가 두 자료의 주택분포의 차이에 미치는 영향은 화성시에서 가장 크게 나타난다. 화성시의 경우 두 자료의 주택분포의 차이에 아파트가

62.7%를 차지하고 있다. 농촌지역인 증평군은 화성시보다 더 높은 비율로 아파트가 주택분포의 차이에 미치는 영향이 더 크다. 그 다음은 인천 서구이다. 12개 지역 중 이 세지역에서는 아파트가 등록자료와 인구센서스자료의 주택분포의 차이에 미치는 영향이 50% 이상이다. 그러나 놀랍게도 광주 동구, 김천시, 신안군 등에서는 등록자료상의 주택수나 인구센서스자료상의 주택수의 측면에서 아파트의 수에는 별다른 차이가 없다. 지역별로 이처럼 주택유형별로 오차의 크기가 어떤 일관성을 찾기 힘들게 변하고 있다.

<그림 5-26> 지역별 등록자료와 인구센서스자료의 주택유형별 주택 분포의 차이



- 광주 동구는 매우 특이하게도 비주거용 건물내 주택의 수가 등록자료와 인구센서스자료의 주택의 수와 매우 차이가 나고, 전체 차이 가운데 이 범주의 주택이 차지하는 차이가 44%나 된다. 왜 이처럼 광주 동구에서 비주거용 건물에 주택이 들어서 있고, 이런 주택의 파악이 등록자료와 인구센서스자료 사이에 것처럼 차이가 나는지 매우 흥미롭다. 지역별로 주택의 유형을 파악하는 인식이 다른 것인가. 아니면 지역에 따라 그런 유형의 주택을 다른 유형의 주택으로 기록하는 것인지 알 수 없다. 이 또한 지역별로 주택의 유형을 등록자료와 인구센서스자료에서 파악하는 방식이 다른 것인지 다시 추가 연구를 할

필요가 있다고 본다.

- 이 절의 내용을 간단히 요약하면 등록자료와 인구센서스자료를 이용해 주택을 파악할 때, 주택의 유형 측면에서 지역별로 큰 차이가 존재한다. 심지어 극과 극으로 차이가 난다고 할 정도로 지역에 따라 주택유형에 따른 자료의 차이가 크게 존재한다. 그런데 이런 차이가 지역별로 어떤 일관성이나 공통된 패턴을 파악하기 힘들다. 따라서 우리는 등록자료로 주택을 파악하고자 할 때는 인구보다도 더 많은 관심과 노력이 기울여야 할 필요가 있음을 절감한다.
- 다음에는 주택의 연면적에 따라 등록자료와 인구센서스자료 사이에 주택을 파악하는데 어떤 차이가 존재하는지 살펴본다. 다음 <표 5-27>은 주택연면적별 등록자료와 인구센서스자료 사이의 차이에 미치는 영향을 보여준다.

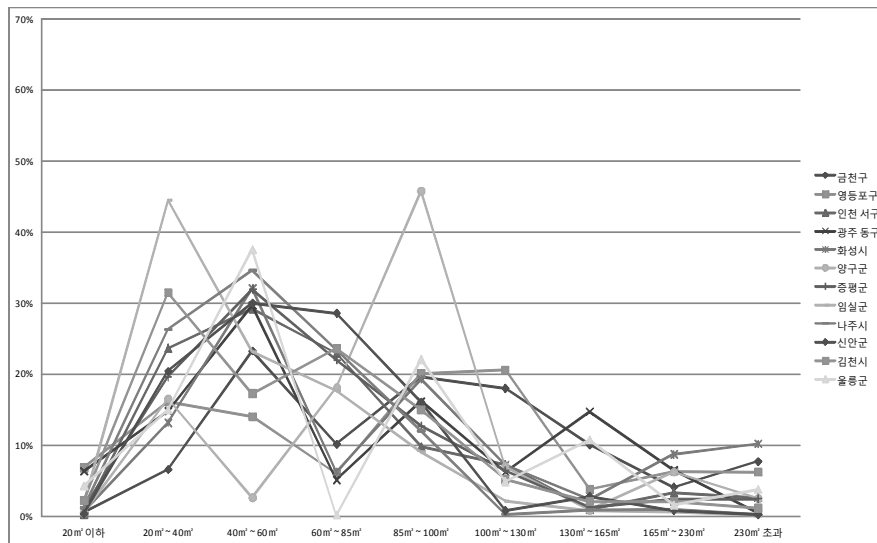
<표 5-27> 지역별 등록자료와 인구센서스자료의 주택유형별 주택 분포의 차이

	2 0 m ² 이하	2 0 m ² ~ 40 m ²	4 0 m ² ~ 60 m ²	6 0 m ² ~ 85 m ²	8 5 m ² ~ 100 m ²	1 0 0 m ² ~ 130 m ²	1 3 0 m ² ~ 165 m ²	1 6 5 m ² ~ 230 m ²	230 m ² 초과
금천구	0.54%	6.60%	23.26%	10.14%	19.63%	17.96%	10.03%	4.13%	7.70%
영등포구	6.83%	16.11%	14.05%	6.00%	20.07%	20.60%	3.79%	6.33%	6.22%
인천 서구	0.30%	23.66%	29.17%	22.92%	9.75%	7.25%	0.95%	3.34%	2.66%
광주 동구	6.32%	14.81%	29.64%	5.11%	16.18%	6.22%	14.72%	6.51%	0.49%
화성시	0.63%	13.16%	32.10%	6.27%	19.30%	7.21%	2.40%	8.74%	10.20%
양구군	0.44%	16.56%	2.61%	18.08%	45.75%	6.97%	0.87%	6.21%	2.51%
증평군	1.10%	19.61%	31.95%	21.97%	12.75%	6.46%	1.27%	2.42%	2.48%
임실군	1.81%	44.51%	23.17%	17.67%	9.04%	2.19%	0.80%	0.64%	0.18%
나주시	1.18%	26.28%	34.62%	23.44%	11.98%	0.32%	0.93%	1.00%	0.25%
신안군	0.36%	20.39%	30.01%	28.56%	16.00%	0.81%	2.77%	0.82%	0.29%
김천시	2.19%	31.45%	17.27%	23.62%	14.99%	5.19%	2.10%	2.07%	1.11%
울릉군	4.26%	14.98%	37.54%	0.16%	22.08%	4.89%	10.73%	1.58%	3.79%

- <표 5-27>은 주택의 연면적에 따라 등록자료와 인구센서스자료 사이에 주택의 분포에 어떤 차이가 존재하는지를 보여준다. 그런데 이 표를 보면 주택의 연면적에 따른 두 자료의 차이는 지역에 따라 전혀 일관성 혹은 일정한 패턴

을 찾기가 거의 불가능함을 보여준다. 연구자의 입장에서 이처럼 당혹스러울 수가 없다. 나름 지역적 특성에 따른 주택의 연면적별 분포가 등록자료와 인구센서스자료의 차이를 보여줄 것이라고 기대했지만, 실제로 오차가 어느 연면적별 주택유형에 따른 것인지 지역별로 살펴보면 어떤 일관성이나 공통적인 유형을 찾기 힘들다.

<그림 5-27> 지역별 등록자료와 인구센서스자료의 주택 연면적별 주택 분포의 차이



- 이런 사실은 등록자료로 주택의 연면적별 주택유형을 파악하는 것이 얼마나 어렵고, 얼마나 신뢰할 수 있으며, 시계열적인 파악은 가능할 것인지 의문을 제기하고 있음을 보여준다. 도시와 농촌, 오래 전에 개발된 도시와 새로 개발되고 있는 도시 등 여러 기준에서 바라보아도 주택연면적별 주택유형의 분포의 차이에는 어떤 일관성과 안정된 패턴을 찾기 힘들다. 따라서 등록자료를 인구센서스자료에 비교해서 어떤 장점이나 단점 혹은 특성이 있는지, 주택연면적의 측면에서는 말하기 힘들다.

4. 소결: 2010년 센서스와 행정자료

- 지금까지 이 연구는 2010년 인구센서스자료와 등록자료를 가지고 인구, 가구, 주택의 측면에서 12개 지역별로 어떤 공통점과 차이점이 있는지를 비교 분석하였다. 이러한 비교분석은 등록자료가 인구센서스자료에 비해 어느 지점에서 얼마나 차이나는지를 살펴보면서 등록자료를 앞으로 본격적으로 시행될 등록센서스의 가장 기초적이고 중요한 자료로 활용할 가능성을 타진하기 위함이다. 즉 등록자료의 품질을 평가하고, 앞으로 본격적으로 시행될 등록센서스의 주 자료로 활용될 수 있는 가능성을 탐색하기 위해 이 연구가 실시되었다.
- 인구의 측면에서는 본 보고서의 제1장 이상 시군구 탐색연구에서 제시된 RLAS라는 지표를 활용하여 등록자료와 인구센서스자료의 성별, 연령별 분포를 12개 지역에서 모두 비교 분석하였다. 이런 과정에서 지역별로 성별, 연령별 두 자료의 분포가 어떤 일관성이나 안정된 패턴을 찾기 힘들게 존재한다는 사실을 발견하였다. 등록자료를 등록센서스의 주요한 자료로 활용하기 위해서는 등록자료의 결함이 구체적으로 어떻게 나타나고, 그 결함의 균일성, 균제성이 파악되어야 한다. 그런데 인구의 측면에서 등록자료와 인구센서스자료의 분포가 불규칙하고 일관성이 없다는 사실은 등록자료를 이용할 때 무엇을 주의해야 할지 일반화할 수 있는 원칙을 만들기 힘들다는 것을 말해준다.
- 이는 가구원수를 기준으로 한 가구유형, 단독주택, 아파트 등으로 구분한 주택유형, 주택의 연면적으로 구분한 주택유형 등으로 등록자료와 인구센서스자료를 비교 분석할 때도 동일하게 발견된다. 지역별로 두 자료의 분포의 차이를 발생하는 요인도 일관성없이 달라진다. 따라서 등록자료가 옳은 것인지, 아니면 인구센서스자료가 옳은 것인지 이와 같은 단순한 비교로는 말할 수 없다. 어느 자료를 더 신뢰할 수 있고, 결함이 있는 자료를 어떻게 보정할 것인지 더 많은 노력과 연구가 필요함을 절실히 느낄 수 있는 것이다. 2013년에는 이런 목적으로 더 심층적인 연구가 필요하고, 추후 연구과제로 제안하려고 한다.

제4절 등록센서스 통계생산방식 평가: 비유사성 지수의 활용

1. 2010년 센서스와 행정자료

- 이 절에서는 행정자료와 인구센서스자료를 비교하는데, 인구의 측면에서 성별, 연령별 비교 분석에 더하여 새로운 변수를 하나 더 추가한 비교 분석을 실시한다. 인구의 경우, 혼인상태와 가구주와의 관계라는 두 변수의 자료가 생성되었다. 이미 등록자료에서 혼인상태의 정의와 인구센서스의 혼인상태의 정의가 다르기 때문에 등록자료상의 성별, 연령별, 혼인상태별 분포와 인구센서스상의 성별, 연령별, 혼인상태별 분포 사이에는 큰 차이가 있을 수밖에 없다는 것을 누구나 인정할 수 있다. 이 절에서는 정의가 다른 한계까지 포함하면 등록자료를 이용한 인구자료와 인구센서스를 이용한 인구자료에 큰 간격이 존재하는 예의 하나로 등록자료의 성별, 연령별, 혼인상태별 분포를 인구센서스자료의 성별, 연령별, 혼인상태별 분포를 비교한다. 다시 반복하면, 이러한 비교는 등록자료의 결함을 드러내기 위함이 아니라, 정의가 다른 두 자료가 존재하고, 한 자료를 다른 한 자료로 대체하는 것이 얼마나 엄청난 작업인지 확인하기 위함이다.
- 2010년 센서스의 혼인상태는 미혼, 기혼(유배우자), 이혼, 사별 등 네 범주로 되어 있다. 행정자료의 혼인상태도 기본적으로 이와 동일하다. 그런데 행정자료는 센서스의 혼인상태의 네 범주 외에 “미상”이 있다. 혼인상태가 명확하지 않은 사람에 대해 미상으로 처리하여 결국 행정자료의 혼인상태 범주는 다섯 범주이다.
- 인구부문의 행정자료를 센서스 가중치의 4개 세트를 적용하여 보완한 행정자료와 비교하는 경우 미상이라는 범주가 있어도 비교에 전혀 문제가 없다. 왜냐하면 능형회귀추정량의 센서스 모집단(또는 부모집단) 크기와 행정자료의 크기를 비교하여 구한 “센서스 가중치”의 4개 세트를 원래의 행정자료에 적용하여 다섯 범주가 있는 혼인상태별 분포표를 만들어 비교하면 되기 때문

이다. 그러나 센서스의 혼인상태별 인구분포를 행정자료 상의 분포 및 릿지회귀방정식에서 얻은 가중치를 행정자료에 적용하여 구한 분포와 비교할 때는 범주가 달라 문제가 생긴다. 즉 센서스 자료의 혼인상태는 4개 범주인데, 행정자료 및 센서스 가중치를 적용하여 보완된 행정자료의 혼인상태별 인구분포는 혼인상태의 범주가 다섯 범주이기 때문에 직접 비교에 문제가 생긴다.

- 행정자료의 혼인상태가 “미상”인 것은 아직 센서스 항목에 대한 행정자료 데이터베이스의 구축이 진행 상태이기 때문에, 현재 이 문제를 만족스럽게 해결하는 방법은 없다. 혼인상태의 범주가 자료마다 다르기 때문에 센서스의 인구분포를 기준으로 행정자료 및 가중치를 적용한 행정자료들의 인구분포를 비교할 때는 부득이하게 미상을 제외하고, 미혼, 기혼, 이별, 사별 등의 네 범주의 분포만을 비교할 수밖에 없다.

2. 2010 인구센서스와 행정자료의 혼인상태별 비교: 센서스 가중치 적용 전후 행정자료의 인구분포 비교 1

- 등록센서스 결과생산 방식의 평가에서 가중치를 적용하지 않은 행정자료가 기준이 되고, 센서스 가중치 4개의 세트를 당초의 행정자료에 적용하여 얻은 분포를 비교하기 위해 비유사성지수를 구한 결과는 <표 5-28>과 같다.
- 먼저, 가중치 미적용의 행정자료를 기준으로 해서 센서스 가중치의 세트를 적용하여 얻은 혼인상태별 분포를 비교했을 때 가중치 미적용의 행정자료의 혼인상태별 분포와 가장 유사한 분포는 릿지 2 가중치로 보완된 행정자료의 혼인상태 분포이다. 예를 들어 서울 금천구의 경우 행정자료의 분포와 릿지 2 보정 행정자료의 분포의 비유사성지수는 남자 분포의 경우 0.122로 릿지 1 보정 행정자료의 0.449, 릿지 3보정 행정자료의 0.397, 릿지 4 보정 행정자료의

<표 5-28> 행정자료(가중치 적용 이전)의 지역별, 성별, 연령별 혼인상태 분포와
 센서스 가중치(릿지 1, 2, 3, 4) 적용 행정자료 간의 지역별, 성별, 연령별
 혼인상태 분포 사이의 비유사성지수

지역		행정자료(가중치 적용이전) vs. 행정자료(가중치 적용 이후)			
		릿지1	릿지2	릿지3	릿지4
금천구	남자	0.449	0.122	0.397	0.999
	여자	0.285	0.184	0.195	0.602
영등포구	남자	0.487	0.061	0.423	0.599
	여자	0.120	0.126	0.213	0.129
인천서구	남자	0.269	0.052	0.460	
	여자	0.080	0.051	0.264	0.202
광주동구	남자	0.872	0.104	0.385	2.118
	여자	1.517	0.609	0.158	2.352
화성시	남자	0.667	0.615	1.135	0.977
	여자	0.264	0.237	0.776	0.584
양구군	남자	1.550	0.352	3.093	0.415
	여자	1.912	1.499	2.113	1.400
중평군	남자	0.646	0.538	0.663	0.646
	여자	0.514	0.294	0.866	0.514
임실군	남자	5.959	4.546	5.886	7.237
	여자	4.309	2.005	4.018	5.508
나주시	남자	5.049	4.579	3.352	5.415
	여자	2.947	2.366	2.094	2.739
신안군	남자	7.107	5.541	6.652	9.621
	여자	7.582	5.773	4.733	9.602
김천시	남자	2.042	1.224	1.977	3.006
	여자	0.815	0.697	1.197	1.667
울릉군	남자	2.270	1.926	1.938	2.062
	여자	5.249	5.108	1.490	5.374

0.999에 비해 가장 작은 것으로 나타난다. 비유사성 지수가 가장 작다는 의미는 두 분포 사이에 차이가 가장 작다는 뜻이다. 반면에 두 분포 사이에 가장 차이가 큰 경우는 릿지 4 보정 행정자료의 분포이다. 릿지 2 보정 행정자료의 분포는 0.122의 비유사성지수를 갖는 반면, 릿지 4 보정 행정자료의 분포는

비유사성지수가 0.999로 가중치를 적용하지 않은 행정자료의 분포와 릿지 1 보정 행정자료의 분포 차이에 비해, 가중치를 적용하지 않은 행정자료의 분포와 릿지 4 보정 행정자료의 분포의 차이가 큼을 의미하고 있다.

- 반면에 릿지 1 보정 행정자료의 혼인상태분포와 릿지 3 보정 행정자료의 혼인상태분포는 릿지 2 보정 행정자료와 릿지 4 보정 행정자료의 혼인상태분포보다 편차가 적다. 덜 극단적인 분포라는 의미이다. 이러한 결과는 각 능형회귀 추정량을 토대로 센서스 가중치를 구한 방법을 이해하면 쉽게 받아들일 수 있는 결과이다. 릿지 2 보정 행정자료의 센서스 가중치는 연령을 세범주(15세 미만/15-64세/65세 이상)로 나누어 구했기 때문에 원래의 분포인 행정자료의 분포에서 크게 일탈하지는 않는다. 반면에 릿지 4 보정 행정자료의 가중치는 일단 시군구별로 나눈 자료에서 각 시군구 내에서 다시 동부/읍무/면부로 지역을 한 번 더 고려했기 때문에 자연히 분포가 원래 행정자료의 분포와 차이가 크게 날 수밖에 없다. 그리고 이런 예상은 혼인상태분포 항목에 대한 등록 센서스 결과 생산방식의 품질에 반영된다. 릿지 1 보정 행정자료와 릿지 3 보정 행정자료는 비슷한 방식으로 센서스 가중치를 구했기 때문에 혼인상태분포 항목에 대한 릿지 1 보정 행정자료의 분포는 릿지 3 보정 행정자료의 분포와 유사하다.
- 12개 대상지역 가운데 비유사성지수가 가장 높은 지역은 전라남도 신안군이다. 신안군은 남녀 모두 다른 11개 대상지역에 비해 행정자료(가중치 미적용)와 센서스 가중치를 적용하여 보완된 행정자료들의 분포 사이에 괴리가 가장 크다. 분포 사이에 가장 차이가 적을 것으로 예상되는 릿지 2 보정 행정자료의 분포와의 차이도 커서 비유사성지수가 남자는 5.541, 여자는 5.773에 이른다. 행정자료(가중치 미적용)와 센서스 가중치를 적용하여 보완된 행정자료들 간의 차이가 가장 클 것으로 예상되는 릿지 4 보정 행정자료와의 비유사성 지수는 신안군 남자의 경우 9.621, 여자는 9.602를 나타낸다.
- 신안군 다음으로 행정자료(가중치 미적용)와 센서스 가중치를 적용하여 보완한 행정자료들 간의 차이가 큰 지역은 전라북도 임실군이다. 임실군은 남자의 경우 릿지 1 보정 행정자료를 기준으로 했을 때 비유사성지수가 5.959이고,

여자의 경우는 4.309이다. 임실군 다음으로는 전라남도 나주시가 가중치 미적용 행정자료와 센서스 가중치를 적용하여 보완된 행정자료 사이에 괴리가 크다.

- 울릉군의 경우에는 여자 분포의 괴리가 남자 분포의 괴리보다 크다. 릿지 3 보정 행정자료가 오히려 예외적인 경우로 보이는데, 릿지 3 보정 행정자료를 제외하면 남자분포보다 여자분포가, 즉 성별, 연령별 혼인상태 분포에서 가중치 미보정 행정자료와 가중치 보정 행정자료의 분포에 차이가 크게 나타나고 있다.
- 반면에 대도시지역, 특히 수도권과 광역도시권에 속하는 서울의 금천구, 영등포구, 인천 서구, 광주 동구, 화성시 등에서는 가중치 미보정 행정자료의 혼인상태 분포와 가중치 보정 행정자료의 혼인상태 분포 사이에 괴리가 상대적으로 작다.
- 동일한 지역이라 할지라도 성별에 따라서 가중치 미적용 행정자료와 가중치 보정 행정자료의 분포 사이에 차이가 있는 것이 확인되고 있다. 보통 인구학에서 비유사성지수를 구하는 경우는 둘 혹은 그 이상의 성별 연령별 인구구조가 서로 얼마나 같거나 다른지를 검토하기 위해 사용한다. 그러나 이 연구에서는 지역별, 성별로 분류한 후 각각의 경우에 다시 연령별 혼인상태 분포를 비교 검토하고 있기 때문에 남자의 인구분포와 여자의 인구분포를 따로 떼어내 비유사성지수를 구하고, 성별에 따른 분포의 차이가 나타나는지를 검증할 수 있다. 성별 분포의 차이를 좀더 살펴보자. <표 5-29>는 12개 대상지역별 비유사성지수의 성별 차이를 보여준다.

<표 5-29> 지역별 비유사성지수의 성별 차이

지역	비유사성지수의 성별 차이	지역	비유사성지수의 성별 차이
서울 금천구	0.062	증평군	-0.244
서울 영등포구	0.065	임실군	-2.541
인천 서구	-0.001	나주시	-2.213
광주 동구	0.505	신안군	0.232
화성시	-0.378	김천시	-0.527
양구군	1.147	울릉군	3.183

주: 가중치를 적용하지 않은 행정자료와 릿지 1로 조정한 행정자료를 비교하여, 센서스 가중치를 구하여 보완된 행정자료를 비교해서 얻은 성별 비유사성지수의 차이이다.

- <표 5-29>에서 양(+)의 값은 남자의 경우, 동일한 지역이라고 할지라도 가중치 미적용 행정자료와 릿지 1로 보정한 행정자료 간의 분포의 차이가 여성의 경우보다 더 크다는 의미이다. 반대로 음(-)의 값은 여성의 분포의 차이가 남성의 분포의 차이보다 더 크다는 의미이다. 서울 금천구, 서울 영등포구, 광주 동구, 강원도 양구군, 전라남도 신안군, 경상북도 울릉군은 남자가 여자보다 가중치 미적용 행정자료와 릿지 1 보정 행정자료 간의 분포의 괴리가 더 크다. 반면에 인천 서구, 경기 화성시, 충청북도 증평군, 전라북도 임실군, 전라남도 나주시, 경상북도 김천시에서는 남자의 행정자료 분포와 릿지 1의 분포 사이의 차이가 여성의 경우보다 작다. 흥미로운 것은, 이 연구에서 대상으로 삼은 12개 지역에서 절반은 남성의 가중치 미적용 행정자료 분포와 릿지 1 보정 행정자료 간의 분포의 차이가 여성의 가중치 미적용 행정자료 분포와 릿지 1 보정 행정자료 간의 분포의 차이보다 크고, 나머지 절반은 작게 나타나고 있다. 우연일 수도 있겠지만, 이 결과를 통해 성별에 따른 차이를 일반화하기 힘들음을 알 수 있다.
- 비유사성 지수의 성별 차이를 절대값으로 하여 규모의 차이를 생각하면 남성의 분포와 여성의 분포 사이에 가장 차이가 큰 지역은 경상북도 울릉군이고, 그 다음은 전라북도 임실군, 그 다음은 전라남도 나주시, 그 다음은 강원도 양구군으로 나타난다. 나주시를 제외하면 모두 도시생활에서 멀리 떨어진 농촌

지역이라고 말할 수 있다. 반면에 도시지역인 인천 서구, 서울 금천구, 영등포구는 성별에 따른 분포의 차이가 거의 없다. 충청북도 증평군, 전라남도 신안군에서도 성별에 따른 분포의 차이가 매우 적다. 전라남도 신안군의 경우 인구분포의 측면에서는 12개 지역에서 가중치 미적용 행정자료의 인구분포와 센서스 가중치를 적용한 행정자료 간의 인구분포에 가장 괴리가 큰 지역이었는데, 남자의 분포나 여자의 분포는 거의 비슷한 수준으로 행정자료와 센서스 가중치를 적용하여 보완한 행정자료 사이에 차이가 있는 것으로 나타나고 있다.

3. 2010 인구센서스와 행정자료의 혼인상태별 비교: 센서스 가중치 적용 전후 행정자료의 인구분포 비교 2

- 지금까지는 센서스 가중치를 적용하지 않은 행정자료를 기준으로 하고, 릿지 1, 2, 3, 4의 센서스 가중치를 적용하여 보완된 행정자료의 분포를 비교해 보았다. 그런데 센서스 가중치를 적용하지 않은 행정자료를 기준으로 삼기 전에 등록센서스 결과 생산방식의 품질에 당장에 영향을 미칠 수 있는 행정자료 품질의 평가 가능성을 더 엄밀히 파악해 볼 필요가 있다. 행정자료 데이터베이스에 포함된 센서스 항목에 대한 품질 점검은 우리는 조사표를 이용하여 현장 실사로 획득한 인구주택총조사 자료를 활용해 볼 필요가 있다.
- 현재의 행정자료는 2010년 센서스 시점을 기준으로 획득한 주민등록, 가족관계등록부를 기준으로 작성된 것이기 때문에, 이것을 비교하는데 2010년의 인구주택총조사 자료가 최적의 조사자료이다. 인구주택총조사에서 혼인상태는 15세부터 파악하고 있다. 통계청 웹사이트 국가통계포털(www.kosis.kr)에서 지역별, 성별, 연령별 혼인상태에 관한 인구주택총조사 자료는 15세 미만은 제공되지 않는다. 그러나 우리가 이 분석에서 활용하고 있는 가중치 미적용 행정자료나 가중치 적용 행정자료의 세트는 모두 0세부터 5세 간격으로 혼인상태 분포표를 만들어 분석에 이용하였다. 그런데 센서스 자료는 15세 이상부

터 제공되기 때문에 센서스 자료와 행정자료를 비교하고, 아울러 릿지 1, 2, 3, 4 등 4개의 센서스 가중치 세트에 의하여 보완된 행정자료와 비교하기 위해서는 모두 15세 이상의 연령분포만을 고려해야 한다.

- 또 하나의 문제점은 인구주택총조사의 혼인상태 분포는 미혼, 기혼, 이혼, 사별 등 네 범주로 되어 있지만, 행정자료 및 릿지 분포는 미혼, 기혼, 이혼, 사별 및 미상으로 되어 있는 것이다. 이 역시 인구주택총조사의 범주에 맞춰서, 가중치 미보정 행정자료 및 가중치 보정 행정자료의 경우에도 혼인상태는 미혼, 기혼, 이혼, 사별 등의 분포만 고려한다.
- 분석대상인 12개 대상지역의 센서스 자료의 지역별, 성별 연령별 혼인상태분포는 부록에 제시되어 있다.
- 통계청에서 이 연구를 위해 연구진에게 제공한 인구부문 행정자료는 12개 지역의 지역별, 성별, 연령별 혼인상태의 집계표(aggregate table)이다. 그리고 조사구(ED)별로 능형회귀추정량의 추정값이 제공되었다. 이 추정값과 해당 조사구의 행정자료 크기를 비교하여, 원래의 행정자료 보완에 적용되는 릿지 1, 2, 3, 4의 센서스 가중치(census weight)를 계산하였다.
- 2010년 인구주택총조사의 인구자료에서 본 연구의 12개 지역의 성별, 연령별 혼인상태 분포를 인구부문 가중치 미적용 행정자료 및 릿지 1, 2, 3, 4의 센서스 가중치에 의하여 보완된 행정자료의 분포와 비교해본다. 그 결과는 <표 5-30>에 제시되어 있다.
- 먼저 인구주택총조사의 성별, 연령별 혼인상태 분포와 12개 지역 행정자료의 성별, 연령별 혼인상태 분포를 비교하면, 가중치 미적용 행정자료와 릿지 1, 2, 3, 4의 센서스 가중치를 적용한 행정자료들을 비교해서 나온 차이와 비교가 되지 않을 정도로 큰 차이가 나는 것을 알 수 있다. <표 5-28>에서 가중치를 적용하지 않은 행정자료의 인구분포와 릿지 1, 2, 3, 4 등 센서스 가중치를 적용한 행정자료들의 인구분포를 비교한 비유사성지수는 제일 차이가 많이 난 경우 8을 넘지 못했다. 그러나 <표 5-30>에서 인구주택총조사의 인구분포

와 센서스 가중치 미적용의 행정자료의 인구분포를 비교한 비유사성지수는 가장 작은 수치가 화성시 남자의 비유사성 지수로 6보다는 크고 7보다는 작은 수치이다. 이 외에는 비유사성지수가 모두 7을 넘는다.

- 12개 연구지역에서 인구주택총조사의 인구분포와 센서스 가중치 미적용 행정자료의 인구분포 사이에 가장 차이가 적은 지역은 화성시로 나타난다. 화성시의 남자의 경우 두 분포의 차이를 알려주는 비유사성지수가 6.356이고 여자의 경우 11.904이다. 그 다음은 서울 영등포구로 남자는 비유사성지수가 7.225, 여자는 11.642로 나타난다. 그 다음은 인천 서구로 남자는 7.466, 여자는 12.226으로 나타난다.
- 반면에 인구주택총조사의 인구분포와 가중치 미적용 행정자료의 인구분포를 비교해서 비유사성지수를 구한 경우, 비유사성지수가 가장 크게 나타나는 지역은 전라남도 신안군이다. 신안군의 경우 남자의 비유사성지수는 17.230이고 여자의 비유사성지수는 39.744이다. 전라북도 임실군도 이와 비슷해서 남자의 비유사성지수는 17.143이고 여자는 38.554로 두 분포가 매우 차이가 남을 보여주고 있다.
- 인구주택총조사의 인구분포와 릿지1, 2, 3, 4의 센서스 가중치를 적용한 행정자료를 비교한 결과는 인구주택총조사와 행정자료의 인구분포를 비교한 것과 크게 다르지 않다. 이는 행정자료의 센서스 가중치인 릿지 1, 2, 3, 4가 능형 회귀추정량의 추정값과 관련 부모집단별 행정자료의 크기를 비교하여 얻어지는 것으로, 대부분의 센서스 가중치는 1.0보다 적지만, 이에 근접하는 값을 가지기 때문이다. 그래서 <표 5-30>에 나타난 비유사성지수는 지역별 성별에 따라서 별로 차이가 나지 않음을 보여주고 있다.

<표 5-30> 인구주택총조사 자료에 대한 가중치 미적용 행정자료, 랫지 1, 2, 3, 4 가중치 보정 행정자료의 분포 차이: 비유사성지수

지역	성별	행정자료				
		가중치 미적용	랫지1	랫지2	랫지3	랫지4
금천구	남자	7.869	7.836	7.851	7.815	7.708
	여자	13.334	13.364	13.334	13.365	13.477
영등포구	남자	7.225	7.235	7.189	7.214	7.222
	여자	11.642	11.623	11.635	11.664	11.662
인천서구	남자	7.466	7.372	7.447	7.537	7.466
	여자	12.226	12.192	12.225	12.245	12.163
광주동구	남자	10.26	9.741	10.272	10.325	10.008
	여자	17.617	16.770	17.637	17.605	16.930
화성시	남자	6.356	6.333	6.092	6.204	6.266
	여자	11.914	11.834	11.810	11.967	11.914
양구군	남자	10.108	10.070	10.119	10.742	10.120
	여자	23.653	23.068	23.688	23.427	23.059
중평군	남자	10.565	10.211	10.601	10.35	10.211
	여자	17.819	18.001	17.919	17.885	18.001
임실군	남자	17.143	12.379	15.109	12.389	12.917
	여자	38.554	36.939	38.481	37.297	36.364
나주시	남자	11.484	8.870	10.614	9.137	8.693
	여자	31.014	30.933	31.506	30.549	30.413
신안군	남자	17.230	11.608	14.937	12.920	10.320
	여자	39.744	36.750	40.341	38.613	36.658
김천시	남자	9.400	7.681	9.065	8.073	7.541
	여자	23.382	23.066	23.566	23.252	23.025
울릉군	남자	14.128	14.267	13.666	13.378	14.291
	여자	27.744	25.348	28.357	27.223	25.467

- 결론적으로 인구총조사의 인구분포와 가중치 미적용 행정자료의 인구분포 그리고 행정자료를 보정하기 위하여 릿지 1, 2, 3, 4의 센서스 가중치를 적용한 행정자료의 인구분포 사이에는 상당히 큰 차이가 존재한다. 따라서 행정자료의 인구분포를 기준으로 하여 등록센서스 결과생산의 방식에 대한 평가 연구도 이루어져야 하지만, 그와 함께 행정자료의 완전성을 검증하기 위해 인구주택총조사 자료 등 여러 다른 자료를 활용해서 행정자료의 품질, 특히 혼인상태에 대한 센서스 정의와 행정자료 정의를 비교하는 작업이 개체자료(individual records)의 수준에서 절대적으로 필요하다고 생각한다.

<표 5-31> 인구총조사의 인구분포와 가중치 미적용 행정자료, 센서스 가중치 릿지 1, 2, 3, 4를 적용한 행정자료를 비교해서 얻은 비유사성지수의 성별 차이

지역	센서스 vs 미적용	센서스 vs 릿지1	센서스 vs 릿지2	센서스 vs 릿지3	센서스 vs 릿지4
금천구	-5.465	-5.528	-5.483	-5.550	-5.769
영등포구	-4.417	-4.388	-4.446	-4.450	-4.440
인천서구	-4.760	-4.820	-4.778	-4.708	-4.697
광주동구	-7.357	-7.029	-7.365	-7.280	-6.922
화성시	-5.558	-5.501	-5.718	-5.763	-5.648
양구군	-13.545	-12.998	-13.569	-12.685	-12.939
중평군	-7.254	-7.790	-7.318	-7.535	-7.790
임실군	-21.411	-24.560	-23.372	-24.908	-23.447
나주시	-19.530	-22.063	-20.892	-21.412	-21.720
신안군	-22.514	-25.142	-25.404	-25.693	-26.338
김천시	-13.982	-15.385	-14.501	-15.179	-15.484
울릉군	-13.616	-11.081	-14.691	-13.845	-11.176

- 가중치 미적용의 행정자료의 인구분포를 릿지 1, 2, 3, 4 3 등 4개 세트의 센서스 가중치를 적용한 행정자료들의 인구분포와 비교할 때도 나타났듯이, 인구주택총조사의 인구분포를 가중치 미적용의 행정자료 및 릿지 1, 2, 3, 4 등 센서스 가중치를 적용한 행정자료들의 인구분포와 비교할 때도 성별에 따른 차이가 크게 부각되고 있다. 이를 검증해보기 위해 12개 각 지역별로 남자의 비유사성지수에서 여자의 비유사성지수를 뺀 값을 구해보았다. 그 결과가 위의 <표 5-31>에 제시되어 있다.
- <표 5-31>을 보면 서울 금천구, 영등포구, 인천 서구, 광주 동구, 화성시 등 도시지역에서는 상대적으로 남성의 인구총조사와 가중치 미적용 행정자료의 연령별 혼인상태 분포와 여성의 인구총조사와 가중치 미적용 행정자료의 연령별 혼인상태 분포의 차이에 상대적으로 큰 차이가 없는 것으로 드러났다. 예를 들어 서울 금천구의 경우 여자의 경우 인구총조사와 가중치 미적용 행정자료의 분포 사이에서 나타나는 비유사성지수와 남자의 비유사성지수의 차이가 -5.465로 여자의 비유사성지수가 남자의 비유사성지수보다도 높은 것이 보인다. 즉 여자의 경우 인구총조사의 인구분포와 가중치 미적용 행정자료의 인구분포 사이의 차이가 남자의 경우 인구총조사의 인구분포와 가중치 미적용 행정자료의 인구분포 사이의 차이보다 더 큰 것이다. 그러나 이와 같은 차이가 있음에도 불구하고 다른 지역에 비하면 상대적으로 그 차이가 작다는 뜻이다.
- 반면에 신안군에서는 여자의 경우 인구총조사의 연령별 혼인상태 분포가 가중치 미적용 행정자료의 연령별 혼인상태 분포와 크게 차이가 난다. 두 비유사성지수의 차이가 22.514나 된다. 인구총조사의 분포와 가중치 미적용 행정자료의 분포 사이에 차이가 많은 것이다. 신안군의 뒤를 이어, 임실군, 김천시, 나주시 등 시 지역에서도 역시 여자의 비유사성지수가 남자의 비유사성지수보다 매우 크다. 즉 남자에 비해 여자의 혼인상태별 인구자료에 행정자료와 조사자료 간에 괴리가 크다는 뜻이다. 이 뒤를 양구군과 울릉군이 잇고 있다.
- 가중치 미적용 행정자료의 인구분포를 릿지 1, 2, 3, 4 등의 센서스 가중치를 적용한 행정자료들의 인구분포와 비교할 때는 비유사성지수가 상대적으로 작음에도 불구하고 4개 세트의 센서스 가중치 가운데 어느 것을 선택하여 행정

자료를 보완하는 것이 제일 좋을 것인가에 대하여 살펴보았다. 그러나 이는 가중치 미적용의 행정자료를 기준으로 하고, 센서스 가중치 중에서 어느 것을 이용하여 행정자료를 보완하는 것이 제일 좋을 것인가를 보는 작업이었다. 즉 행정자료가 등록센서스 결과생산의 품질을 결정하는 기준처럼 간주된 것이다.

- 그런데 행정자료가 등록센서스 결과 생산방식의 기준이 아니고 다른 자료와 비교할 때 행정자료가 얼마나 유사한 자료인지 아니면 차이가 많은 자료인지를 살펴보기 위해 2010년 인구주택총조사의 지역별, 성별, 연령별 혼인상태 분포를 기준으로 하고, 행정자료의 동일한 분포, 릿지 1, 2, 3, 4 등 4개 세트의 가중치를 적용하여 얻은 동일한 분포를 비교하면, 가중치 미적용의 행정자료와 릿지 1, 2, 3, 4 등의 4개 세트 가중치를 적용한 행정자료의 분포 사이에는 차이가 별로 크지 않다고 볼 수 있음을 알 수 있다. 인구총조사의 인구분포와 센서스 가중치의 적용여부에 관계없이, 행정자료의 인구분포에 오히려 큰 차이가 존재하는 것이다. 이런 상황에서 가중치 미적용의 행정자료만을 기준으로 하여 릿지 1, 2, 3, 4 등 4개 세트의 센서스 가중치를 적용하여 보완된 행정자료에서 나오는 분포를 비교하는 것은 그리 중요한 작업이 아닐 수도 있음을 보여주고 있다.
- 올해의 연구 가운데 가중치를 적용하지 않은 행정자료를 중심으로 릿지 1, 2, 3, 4 등 4개 세트의 센서스 가중치를 적용하여 보완된 행정자료를 가지고 등록센서스 통계생산방식의 품질을 점검하는 작업은 행정자료의 품질점검까지 포함하기 때문에 등록센서스와 관련된 전반적 품질점검 작업의 시작이라고 볼 수 있다. 이 연구에서 본격적으로 시도하지 못하고 있지만, 시험적으로 인구주택총조사 자료에서 12개 지역을 골라 성별, 연령별 혼인상태 분포를 기준으로 행정자료의 동일한 분포와 비교하면 등록센서스 통계생산 방식의 품질진단의 영역에서 또 다른 차원의 논의가 전개될 수도 있을 것이다.
- 올해의 연구는 인구의 측면에서 지역별, 성별, 연령별 혼인상태분포와 지역별, 성별, 연령별 가구주와의 관계에 대해 가중치 미적용의 행정자료를 기준으로 릿지 1, 2, 3, 4 등의 센서스 가중치를 적용한 행정자료에서 얻어진 분포를 비교하기 위한 수많은 표를 산출하고, 이를 요약해주는 비유사성지수를 구해 분

포의 차이를 확인하였다. 가구주와의 관계도 매우 중요한데, 이 연구에서는 지역별, 성별, 연령별 가구주와의 관계에 따른 분포의 차이를 검증하기 위한 비 유사성지수는 산출하지 않았다.

- 또 행정자료 자체의 품질평가를 위해 2010년 인구주택총조사 자료를 이용하여 12개 지역별, 성별, 연령별 혼인상태 분포를 만들고, 이를 가중치 미적용 행정자료의 혼인상태 분포, 릿지 1, 2, 3, 4 등 4개의 가중치 세트를 적용하여 보완된 행정자료에서 얻어진 혼인상태의 분포와 시험적으로 비교하여 보았다. 이를 통해서도 행정자료 자체에 대해 여러 조사자료를 통해 다시 한 번 품질을 본격적으로 진단하는 것이 필수적임을 확인할 수 있었다.
- 행정자료를 센서스 가중치를 이용하여 보완하고, 행정자료의 다양한 자료원을 발굴하여 등록센서스를 실시한다는 전제 하에 2013년에는 행정자료 자체의 품질을 검증하는 새로운 연구가 필수적이다. 올해의 연구는 등록센서스 결과 생산 방식의 품질 점검에 초점이 맞추어졌기 때문에, 가중치를 적용하지 않은 행정자료를 기준으로 하여, 가중치를 적용하여 보완된 행정자료들과의 괴리가 어느 정도가 되는가를 검토한 것에 불과하다. 2013년 연구에서는 행정자료를 다른 여러 자료(조사자료, 행정자료)와 비교해서 행정자료의 성격을 더 잘 이해하고, 행정자료의 센서스 항목에 대한 품질을 체계적으로 점검할 수 있는 여러 방법론적 모색이 여전히 필요하다고 말할 수 있다.

4. 가중치 적용 전후 행정자료의 가구분포 비교

- 본 연구는 행정자료를 인구자료로서만 아니라 가구 및 주택의 측면에서도 품질을 평가하였다. 가구 및 주택에 관한 행정자료의 품질을 검증하기 위해 역시 능형회귀추정량에 의하여 산출된 가구 모집단(또는 부모집단)의 크기를 원래의 가구 모집단 크기와 비교하여, 센서스 가중치가 산출되었다. 이 가중치를 이용하여 새로운 가구 및 주택 분포가 도출되었다.

- 인구자료를 검증하는 것과 마찬가지로 방식으로 가구 및 주택의 가중치 미적용 행정자료 분포와 능형회귀추정량을 이용하여 계산한 센서스 가중치를 적용한 행정자료의 분포에 얼마나 차이가 있는지를 보기 위해, 1) 인구자료를 검증할 때와 마찬가지로 가중치 미적용 행정자료와 가중치 적용 행정자료를 이용하여 가구와 주택의 분포에 관해 산출한 자료의 각 셀의 차이를 구하고, 2) 그 다음에는 차이의 절대값을 구하였다. 3) 마지막으로 비(ratio)를 구해 분포의 차이를 검증하려고 하였다.
- 인구자료의 검증과 마찬가지로 위와 같은 방식으로 수많은 표가 산출되고 일목요연하게 등록센서스 결과 생산방식의 품질을 검증하는 것이 불가능하다. 그래서 가구 행정자료(가중치 미적용)의 분포와 센서스 가중치를 적용하여 보완한 행정자료로부터 얻은 가구 분포의 차이를 검증하기 위해 비유사성지수를 구하였다. 그 결과가 다음 <표 5-32>에 제시되어 있다.
- 12개 지역별 가구구분별 가구원수 분포에 관한 가중치 미적용의 행정자료와 릿지의 센서스 가중치를 적용한 가구구분별 가구원수 분포를 비교하여 얻은 비유사성지수를 보면 금천구, 영등포구, 인천 서구, 광주 동구, 화성시 등 도시 지역의 비유사성지수가 상대적으로 높은 편이다. 금천구의 비유사성지수가 6.436이고, 영등포구도 비유사성지수가 4.801에 달한다. 이것은 일부 지역에 대하여, 가중치를 적용하지 않은 행정자료의 품질이 가중치를 적용하여 등록 센서스 자료를 생산하는 방식에 의하여 통계의 품질에 영향을 미칠 수 있음을 의미한다.
- 반면에 다른 군 지역의 지방의 시지역에서는 가구에 관한 행정자료와 능형추정량으로 보완된 행정자료의 분포 사이에는 차이가 크지 않다. 임실군은 비유사성지수가 0.201, 증평군은 0.291 등으로 원 행정자료 상의 가구구분별 가구원수 분포나 센서스 가중치를 적용한 행정자료에서 얻은 가구구분별 가구원수 분포 사이에 별 다른 차이가 보이지 않는다.

<표 5-32> 행정자료(가중치 미적용)의 지역별 가구구분별 가구원수 분포와 행정자료(가중치 적용)에서 얻은 지역별 가구구분별 가구원수 분포를 비교해서 얻은 비유사성지수

지역	비유사성지수
금천구	6.436
영등포구	4.801
인천 서구	2.198
광주 동구	1.526
화성시	3.800
양구군	0.455
증평군	0.291
임실군	0.201
나주시	0.818
신안군	0.859
김천시	0.809
울릉군	0.344

- 결론적으로 12개 지역별 가구구분별 가구원수 분포에 관한 행정자료와 센서스 가중치를 부여해서 얻은 가구구분별 가구원수 분포를 비교하면 일단 서울 수도권의 대도시 지역에서 차이가 크고, 지방에서는 가중치 미적용의 행정자료의 가구분포와 센서스 가중치를 이용하여 보완한 행정자료의 가구분포 사이에 큰 차이가 없는 것으로 보인다. 현재, 이러한 차이는 많은 경우, 서울 등 수도권의 경우, 가구구분에서 인구주택총조사와 행정자료 간에 존재하는 1인가구의 차이 때문이라고 추측할 수는 있다. 그러나 이런 현상이 다른 지역에도 일관성있게 적용되는지는 다음 연구에서 검토해 보아야 한다.

5. 가중치 적용 전후 행정자료의 주택분포 비교

- 인구, 가구 부문에 이어 본 연구에서 검토한 것은 주택 부문에 관한 가중치 미 적용 행정자료의 품질을 점검해 보는 것이다. 이를 위해 행정자료에서 지역별 주택구분별 연면적 분포 및 지역별 주택구분별 건축연도 분포를 파악하고, 주택 부문의 센서스 모수를 산출하기 위하여 작성된 능형회귀추정량과 행정자료를 비교하여 센서스 가중치를 계산하고, 그 가중치를 행정자료에 적용하여 보완된 행정자료를 만들고, 이를 이용하여 다시 동일한 표를 산출하였다. 주택에 관한 등록센서스 결과생산 방식을 평가하는 것은 원래 주택에 관한 행정자료에서 확보한 표와 센서스 가중치를 적용하여 얻은 행정자료에서 얻은 표를 비교하는 것이다. 인구 및 가구에서와 마찬가지로 가중치 적용 이전의 행정자료의 분포가 있다면, 여기에 능형회귀추정량을 이용하여 센서스 모집단의 크기를 확정하고 이를 행정자료의 크기와 비교하여 계산한 센서스 가중치를 이용하여 보완된 행정자료의 분포를 비교하여 두 분포 사이에 어떤 차이가 나타나는지, 지역별로 어떤 차이가 나타나는지를 점검하는 방식을 택한다.
- 주택에 관한 분포를 가중치 미보정 행정자료와 센서스 가중치를 적용하여 보완된 행정자료의 주택분포를 비교하는 방식은 인구, 가구와 마찬가지로이다. 첫째, 두 분포의 차이를 구하기 위해 센서스 가중치에 의하여 보완된 행정자료에서 센서스 가중치를 적용하지 않은 분포의 각 셀의 값을 빼거나, 둘째, 절대값을 취하거나, 셋째, 비(ratio)를 구해 주택에 관한 등록센서스 결과생산 방식의 품질을 평가하려고 하였다.
- 그러나 현실적으로 두 분포의 차이를 알기 위해 비유사성지수를 구하여 가중치를 적용하지 않은 행정자료 상의 주택의 분포와 가중치를 적용하여 보완된 행정자료 상의 주택의 분포를 비교하는 것이 가장 기본적인 방식이다. 다음 <표 5-33>은 지역별 주택구분별 연면적 분포 및 지역별 주택구분별 건축연도 분포의 비유사성지수를 구한 결과를 보여주고 있다.

<표 5-33> 행정자료(가중치 미적용)의 지역별 주택구분별 연면적 분포, 주택건축연도 분포와 가중치 적용 행정자료에서 얻은 지역별 주택구분별 연면적 분포, 주택건축연도 분포를 비교해서 얻은 비유사성지수

지역	주택연면적	주택건축연도
금천구	0.777	0.278
영등포구	0.222	0.221
인천서구	0.558	0.920
광주동구	3.383	2.307
화성시	0.258	0.278
양구군	1.359	1.347
중평군	0.594	0.581
임실군	0.780	0.779
나주시	0.888	0.891
신안군	1.085	1.085
김천시	0.709	0.715
울릉군	1.452	1.372

- 주택 연면적의 측면에서는 서울 영등포구의 주택 행정자료가 센서스 가중치를 적용하여 보완된 행정자료를 만든 경우에 비추어보아도 별다른 차이가 나지 않음을 알 수 있다. 주택구분별 주택연면적 분포의 비유사성지수가 영등포구에서는 0.222에 불과하다. 그 다음은 화성시로 비유사성지수가 0.258로 금천구와 거의 비슷한 수치를 보인다. 인천 서구는 비유사성지수가 0.558로 이 두 지역과 마찬가지로 가중치 미적용 행정자료와 가중치를 적용한 행정자료에서 얻어진 분포와의 사이에 큰 차이가 없다.
- 반면에 광주 동구, 울릉군, 양구군, 신안군 등의 주택에 관한 행정자료는 센서스 가중치를 적용하여 만든 행정자료와 비교하면 분포 상 상대적으로 차이가 크다.

- 주택 건축연도 분포를 분석해 보아도, 서울의 영등포구, 금천구, 화성시 등에서는 행정자료에 센서스 가중치의 적용 여부에 관계없이 지역별 주택구분별 주택 건축연도 분포 사이에 별 다른 차이가 없다. 반면에 광주 동구, 울릉군, 양구군, 신안군 등에서는 주택 연면적의 분포와 마찬가지로 주택연도의 분포에서도 역시 마찬가지로 가중치를 적용하지 않은 행정자료와 가중치를 적용하여 만든 행정자료의 분포 사이에 그 차이가 상대적으로 크다. 그러나 행정자료를 센서스 가중치로 보정한 결과는 본질적으로 센서스 가중치를 적용한 행정자료와 매우 크게 다르지는 않아서 비유사성지수가 그렇게 크게 나타나지 않는다. 이것은 능형회귀추정량에 의한 등록센서스 결과 생산방식이 주택부문의 통계품질에 그리 큰 영향을 미치지 않는다는 것을 의미한다.

제5절 자료원 차이에 따른 품질평가 요소 및 제언³⁾

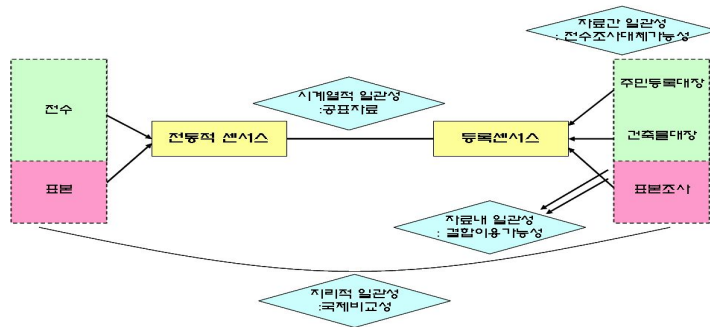
- 이 절은 인구센서스 주요 전수항목과 이에 상응하는 등록자료 항목들의 개념 및 분류방식 등을 중심으로 그 대체가능성과 주의사항을 점검하고 이를 향상시키기 위한 방안을 논의한다. 여기에서 고려된 항목들은 인구부문의 성, 연령, 가구주와의 관계, 혼인상태, (4항목), 가구부문의 가구구분, 가구원수(2항목), 주택부문의 주택연면적, 주택구분, 건축연도(3항목) 등이다.

1. 통계품질기준으로서의 일관성

- 일관성은 대체나 자료나 통계작성에 있어서 공통적인 개념, 정의, 분류 등을 적용함으로써 시공간적 비교가능성을 유지하는 것이라 할 수 있다. 김정석(2009)은 등록센서스의 시행에 따른 일관성의 이슈를 크게 다음의 <그림 1>과 같이 정리한 바 있는데, 이에 기대어 등록자료 생성, 가공 및 이용상의 유의사항을 정리해 보기로 하겠다.

3) 본 절은 2009년도 ‘등록센서스 기법개발 연구’보고서 제6장 등록센서스의 결과표 작성과 통계자료 시계열 변동 보완방법(김정석 집필)과 제8장 현행 센서스와 행정자료의 주요 개념 비교분석 및 등록센서스 제도적·법률적 기반개선의 과제(최진호 집필)에 기반하였음을 밝혀둔다.

<그림 5-28> 등록센서스로의 이행에 따른 일관성 이슈들



- <그림 5-28>에 나타난 바와 같이, 전통적인 센서스는 거대규모의 사회조사로 단일대상, 단일조사표에 근거한다. 물론 전수조사와 표본조사로 나뉘기는 하지만 그 대상은 동일하며, 조사문항도 전수문항의 추가에 불과하다. 이에 비해 등록센서스는 여러 개의 행정자료를 이용하는 동시에 표본조사를 병행함으로써, 다양한 자료원을 가지게 된다.
- 인구센서스와 등록자료간의 일관성: 전수조사항목을 대체하기 위한 등록자료를 구축할 때 갖는 부담은 기존 인구센서스자료와 개별 행정자료간의 일관성이다. 이를 위해서는 일차적으로 두 자료간에 대체가능한 항목들의 개념, 정의, 분류상의 차이가 파악하는 작업이 전제가 된다. 그 다음으로 두 자료간의 실질적인 수치비교와 보정작업이 요청된다.
- 등록자료내 (다양한 행정자료와 표본조사자료간) 일관성: 인구센서스와 등록자료간의 일관성이 대체로 전수조사항목 대체에 초점을 둔 것이라 할 수 있다. 등록자료내 일관성은 주민등록부, 가족관계등록부, 건축물대장 등의 다양한 행정자료와 표본조사자료를 서로 결합하여 사회적으로 유의미한 공표자료를 제공할 때 중요한 이슈가 될 수 있다. 각 자료원으로부터 나온 항목들을 결합하여 사용하려 할 때 자료원간의 항목개념, 정의, 분류 등의 모순이 없으며, 의미있는 결합이 가능해질 것인가가 주요 관심사가 될 것이다.
- 시계열적 일관성: 시계열적 일관성은 등록센서스를 구축하는 단계에서 이슈

가 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.

- **지리적 일관성:** 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 센서스에서 획득할 수 있는 국제비교 가능성보다는 기대되는 바가 적을 것으로 보인다.
- 등록센서스를 준비하고 그 자료를 구축하는 현 단계에서는 무엇보다도 인구센서스와 등록자료간의 일관성 확보가 일차적인 과제가 될 것이다. 그러나 센서스에서 사회적으로 정책적으로 의미있는 자료를 제공하기 위해, 여러 부문에 걸친 항목들을 서로 결합해 공표자료를 생산한다는 점을 감안한다면 등록자료내 일관성에 대해서도 많은 주의와 관심이 요구된다. 시계열적, 그리고 지리적 일관성은 앞서 언급한 두 가지 일관성이 얼마나 확보되느냐에 따라 결정될 것으로 보인다.

2. 등록자료와 인구센서스 자료간 일관성: 개념 및 자료성격 차이

- 다음의 <표 5-34>는 현재 통계청이 등록자료로 대체가능할 것이라 판단하고 점검중인 인구센서스의 주요 전수항목과 이에 상응하는 등록자료 자료원을 정리한 것이다. 인구부문에 해당하는 자료는 대체로 주민등록부를 이용하지만, 혼인상태는 가족관계등록부를 이용하고 있다. 한편 주택부문에 해당하는 자료는 건축물대장을 이용하고 있다.

<표 5-34> 주요 대체항목별 자료원

구분	인구센서스	행정자료	주자료명 <기관>
인구 (4)	성별, 연령, 가구주와 관계	성별, 연령, 세대주와 관계	주민등록자료 <행정안전부>
	혼인상태	혼인상태	가족관계등록부 <대법원>
가구 (2)	가구구분	가공(세대활용)	주민등록자료 <행정안전부>
	가구원수	가공	
주택 (3)	주택구분, 주택연면적 건축연도	건축물 주용도 주용도, 기타용도 면적 사용승인일자	건축물대장 <국토해양부>

- 연령-실제나이와 주민등록상 나이: 인구센서스에서는 주민등록상의 나이와 관계없이 가구원의 양력, 음력과 실제 생년월일을 조사한다. 한편, 주민등록부상의 연령은 본인들의 신고에 기초한다. 따라서 실제 연령에 비해 신뢰성이 떨어진다고 알려져 있다. 특히 과거에는 출생신고가 늦어진 경우가 많아, 실제 나이가 주민등록부상의 나이보다 많은 사람들이 적지 않은 것으로 판단된다. 그러나 최근으로 올수록 출생신고가 제때 이루어져 실제나이와 주민등록상의 나이차가 많은 경우가 급격히 줄어들었다. 인구센서스상의 실제나이와 주민등록부상의 나이 차이는 결국 출생코호트에 따라 다르다는 점에 유의할 필요가 있다.
- 가구주와의 관계와 세대주와의 관계: 인구센서스에서는 주민등록상의 세대주와는 관계없이 ‘가구를 실질적으로 대표하는 사람’으로 정의한다. 이에 비해, 주민등록부의 세대주는 ‘세대를 대표하는 자’로서 주민의 신고에 의하여 선정이 가능하다(다만 성년이 있는 세대에서 고의적으로 미성년자를 세대주로 선정하여 기재하는 것은 금지되어 있다). 주민등록법이나 시행령에 세대주에 대한 정의는 없지만, 업무지침에 세대란 ‘주거 및 생계를 같이하는 집단을 의미한다고 규정하고 있다. 대체로 가구주는 경제력이 있는 의사결정권자인 반면, 세대주는 연장자인 경우가 많은 것으로 알려져 있다. 가구주와의 관계와 세대주와의 관계가 어느 정도 일치하는가를 보기 위해서는 가구주와 세대주의

성별 연령별 특성을 서로 비교 검토해보는 것이 필요하다. 한편 가족이나 가구구조가 소규모화되고 핵가족화되는 사회적 추세를 감안한다면 가구주와 세대주간의 차이가 적을 가능성이 높아지지만. 그러나 이혼사례가 늘어나고 기러기가족이나 주말부부 등이 가족이 다양해지고 있다는 점을 고려한다면, 세대주와 가구주의 차이가 발생한 가능성도 적지 않다. 가령, 주말부부의 경우 주거처지에 남은 부인이 세대주(자녀의 학교 등으로 인해 이주의 신고의무를 수행)가 되고, 떨어져 살고 있는 남편이 가구주가 될 가능성이 있다.

- 가구주와의 관계/ 세대주와의 관계 분류: 인구센서스에서는 가구주, 가구주배우자 등을 비롯해 기타 동거인에 이르기까지 총14가지로 가구주와의 관계를 분류하고 있다. 한편, 주민등록상의 세대주와의 관계는 세대주 본인, 처, 남편, 자녀를 비롯해 총75가지로 그 관계를 분류하고 있다. 주민등록상의 세대주와의 관계가 훨씬 세분화되어 있어, 분류방식에 있어서 대체가능성은 별 문제가 없을 것으로 보인다.
- 혼인상태-사실혼과 법률혼: 인구센서스에서는 혼인 및 이혼신고와 관계없이 사실상의 혼인상태를 묻는다(이와 함께 인구센서스는 조사대상자의 응답에 의존한다). 이에 비해, 가족관계등록부상의 혼인상태는 법률상의 혼인상태를 의미한다. 혼전동거나 이혼의 전단계로서 별거 등이 증가하는 상황에서 둘 간의 차이가 커질 가능성이 존재한다.
- 가구구분: 인구센서스에서는 가구를 일반가구, 집단가구, 외국인가구로 나누며 이중에서 일반가구를 가족으로 이루어진 가구, 가족과 5인이하의 남남이 함께 사는 가구, 가족이 아닌 5인 이하의 남남이 함께 사는 가구, 1인가구 등으로 나누고 이를 조사한다. 한편, 세대개념을 사용하는 주민등록부상에는 이러한 구분이 없다. 다만 세대원들과 세대주와의 관계를 중심으로 이를 가공할 수 있다. 따라서 세대주와의 관계나 세대라는 개념이 가구주와의 관계나 가구라는 개념에 근사할수록 자료원에 따른 차이 또한 적을 것으로 기대된다.
- 가구원수-가구와 세대: 인구센서스에서 가구는 1인 또는 2인 이상이 모여서 취사, 취침 등 생계를 같이 하는 생활단위로 정의된다. 따라서 가사를 함께 하

는 가구원은 모두 해당 가구의 구성원으로 파악된다. 예컨대, 먼 친척이나 종업원 등 가족이 아니지만 가사를 함께 하는 사람은 가구원에 포함시켜서 조사된다. 반면, 군복무, 취업, 학업 등의 이유로 함께 살고 있지 않는 가족원들은 가구원에서 제외된다. 주민등록부에서 사용하는 세대 또한 주거지와 생활단위를 함께 하는 집단으로 친족을 중심으로 하지만 타인이 포함시키기도 한다는 점에서 가구와 큰 차이가 없다고 한다. 따라서 세대원수를 가구원수로 대체할 가능성도 높아진다.

- 주택종류-거처종류와 건축물용도: 인구센서스에서 거처란 사람이 살고 있는 모든 장소로, 구조적으로 분리되고 독립된 하나의 거주단위로 정의된다. 모든 거처는 크게 주택과 주택 이외의 거처로 구분한다. 주택은 가구가 독립적으로 살림을 할 수 있도록 지어진 집인데, 다시 단독주택, 아파트, 연립주택, 다세대주택, 비거주용건물 내 주택으로 분류된다. 단독주택은 일반단독, 다가구단독, 영업겸용단독을 다 포함한다. 건축물대장은 용도별로 건축물의 종류를 구분하는데 모두 28개로 분류하고 있다. 이 중에서 총조사의 거처에 해당하는 건축물로는 단독주택과 공동주택의 분류가 있다. 먼저 단독주택은 다시 세분해서 단독주택, 다중주택, 다가구주택, 공관으로 구분하고 있다. 다중주택은 여러 사람이 장기간 거주할 수 있는 구조로 되어있으나 독립된 주거의 형태를 갖추지 아니한 것으로 인구센서스에는 포함되지 않는다. 다음으로 공동주택은 아파트, 연립주택, 다세대주택, 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. 인구센서스와 건축물대장의 개념정의를 큰 차이가 나지 않으며, 건축물대장의 종류 구분이 더 세분화되어 있어 대체가능성이 높다고 판단된다.
- 건축연도: 인구센서스에서 건축연도는 신축시기를 원칙으로 하나 증축, 개축 면적이 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, 개축시

기와 증개축 면적이 별도로 표시된다. 따라서 총조사의 정의에 맞추기 위해서는 건축물 대장에서 증개축 여부와 그 면적을 따져서 건축시기를 조정해야 한다

- 대체항목들의 개념과 분류차이 이외에도 인구센서스자료와 등록자료간의 차이는 자료원의 성격에 기인하기도 한다. 무엇보다도 주민등록자료는 신고자료라는 점을 감안해야 한다. 인구센서스를 대체할 등록센서스의 자료원으로 가진 가장 큰 문제는 주민등록신고지와 실제 거주지와 불일치를 들 수 있다. 여기에는 세금과 주택청약, 부동산취득 등의 이익을 위한 허위등록으로부터 자녀를 인기학군의 학교로 보내기 위한 허위등록 등 그 원인이 다양하다.
- 이주신고를 하지 않아 등록이 계속 전거주지로 되어 있는 경우 등도 있는데 이는 지역별 수준에서 인구센서스자료와 등록자료간의 차이를 초래할 것이다. 전국적 수준에서 총인구수 차이에는 주민등록인구에서는 집계되지만 인구센서스에서는 제외된 해외장기거주자를 들 수 있다. 이외에도 출생신고나 사망신고의 지연이, 주민등록지권말소에 따른 과소과대 집계 가능성도 있다.
- 건축물 대장자료가 센서스의 주택항목을 대체하는 데 있어서 주택분류 방식 이외에도 건축물의 실제사용과 대장상의 기재 차이라는 문제점이 있다. 주택 면적에 관련된 연건평에 있어서도 두 자료간의 차이가 적지 않은데, 대체적으로 건축물대장보다 인구주택총조사에서 더 크게 조사된 것으로 보고된다. 특히 단독주택의 대지면적은 건축물대장에 누락된 사례가 매우 많은 것으로 알려져 있다. 현재 건축물대장이 갖고 있는 문제점으로는 건축물 대장 기재항목 가운데 주요 항목이 기재되어 있지 않은 누락이 항목별로 다르기는 하지만 상당한 정도에 이르고 있다는 것이다.

3. 등록자료내 일관성 및 시계열적 일관성 확보과제 및 제언

- 이상에서는 전수조사 주요 항목을 대체할 수 있는 등록자료의 항목개념과 자료원의 특성을 중심으로 인구센서스와 등록자료간의 일관성을 논의하였다. 현

재 통계청에서 제공된 행정자료(그리고 그에 대한 보정자료)를 기존의 인구센서스 일부와 비교하는 오차를 분석하는 작업은 이러한 자료간 일관성을 점검에 해당한다.

- 등록자료가 구축되는 과정에서 이용되는 행정자료 사이의 일관성에 대해서도 주목할 필요가 있다. 대체로 공표자료는 단일항목에 대한 정보보다는 항목의 결합을 통해 전달되는 경우가 많은데, 이는 사회적으로 유의미한 항목간의 상관성을 반영하기 때문이다. 항목간의 결합은 곧 자료원간의 연계를 의미한다. 가령 가구형태별 주택유형을 살피려면, 주민등록부에서 마련된 가구형태자료와 건축물대장에서 마련된 주택유형자료가 결합된다.
- 행정자료원간의 결합을 위해서는 몇 가지 문제가 발생한다. 우선 주민등록부의 세대와 건축물대장상의 주택점유자를 어떻게 연결시킬 것인가라는 점이다. 세대구성원의 고유번호와 주택점유자의 고유번호(이는 다시 개인고유번호와 세대고유번호로 구분가능)가 부여되고 서로 인식될 수 있어야 한다. 다음으로 각 행정자료원의 정확성과 신뢰성의 수준이 서로 견줄 수 있는 수준이어야 한다. 특히 건축물대장의 신뢰성이 상대적으로 낮은 상태라는 점을 감안한다면, 자료원간의 결합에 적지 않는 어려움이 생긴다.
- 행정자료원간의 결합을 추구하는 과정에서 발생하는 또 다른 문제는 행정자료의 보정을 위한 가중치를 어떻게 항목간 결합에 반영할 것이라는 문제이다. 현재의 보정치는 인구(개인), 가구, 주택으로 나누어져 있으며 각 부문에 반영되고 있다. 앞서 언급한 바와 같이 가구형태별 주택유형의 공표자료를 준비하게 될 경우에 가구와 주택 각각의 보정치를 동시에 사용해야 한다. 인구부문과 가구부문, 인구부문과 주택부문 항목의 결합에서도 이런 상황이 발생한다.
- 등록센서스의 주요 자료원이 되는 표본조사자료와 행정자료원간 일관성에도 주목할 필요가 있다. 특히 혼인상태 항목은 그 활용도가 매우 높기 때문에 표본조사항목에 포함될 가능성이 크다. 가령 혼인상태별 출산력, 혼인상태별 경제활동참가율 등을 산출하기 위해서는 가족관계등록부에서 도출된 혼인상태자료만으로는 불가하다. 이럴 경우, 혼인상태는 가족관계등록부상의 혼인상태

(전수항목대체용)과 표본조사상의 혼인상태 두 가지가 공존하게 된다. 그리고 이 둘은 법률혼과 사실혼이라는 개념 차이를 안고 있다.

- 행정자료를 이용한 등록센서스와 인구주택총조사는 그 자료원의 변화를 의미하기 때문에 기존의 인구주택총조사와 등록센서스간의 시계열적 일관성을 확보하는 데는 적지 않은 노력이 필요하다. 이러한 노력은 등록센서스를 기획하고 준비하는 단계에서부터 이루어질 필요가 있다.
- 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. 특히 자료내 일관성이 담보되어야 할 것이다.
- 결과활용 단계에서 공표자료형태와 구성 내용이 가능하면 기존의 센서스와 유사하도록 할 필요가 있다. 이를 위해서는 최소한의 수준에서 시계열적 일관성을 보여줄 수 있는 항목과 내용이 무엇인지를 고민하여야 할 것이다. 또한 공표자료에 각주와 비교란을 적극 활용하여 시계열적 이용에 따른 유의사항을 제시하여야 할 것이다. 다음의 <표 5-35>은 공표결과표를 중심으로 변동사항을 예측하고 유의사항을 표기하기 위한 예시이다.
- 이와 관련하여 생각해 볼 수 있는 것이 메타자료(Meta Data)의 생성이다. 이러한 메타자료에는 등록센서스로의 변동근거를 비롯하여, 자료원의 비교소개, 조사항목추정과 추정방식의 소개 등이 필수적이다. 또한 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 이

해가 높아질 것이다.

<표 5-35> 인구센서스 전수항목 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.	연령 및 성, 혼인상태별 인구(15세 이상)	전수▷가족관계등록부상의 혼인상태명시/유의점 제시
6표.	성, 연령 및 가구주와의 관계별 인구	전수▷가구주와 관계를 세대주와관계로 변동/유의점 제시
7표.	가구주와의 관계, 성 및 혼인상태별 인구	전수▷가족관계등록부상의 혼인상태명시/유의점 제시
8표.	성, 혼인상태, 연령 및 세대구성별 인구	전수▷가족관계등록부상의 혼인상태명시/유의점 제시
9표.	성, 혼인상태, 연령 및 교육정도별 인구 (15세 이상)	표본: 혼인상태/교육정도 표본조사의 혼인상태와 가족관계등록부상의 혼인상태 차이명시???

가구편(전수)		등록센서스예상표
1표.	행정구역, 가구형태별 가구 및 가구원	전수▷세대주관계를 가공/유의점 제시
2표.	행정구역, 세대구성별 가구 및 가구원	전수▷세대주관계를 가공/유의점 제시
3표.	가구주의 성, 연령 및 혼인상태별 가구 (일반가구)	전수▷가구주에서 세대주로의 변동/혼인상태측정방식변동/유의점제시

4표.	가구주의 성, 연령 및 세대구성별 가구 (일반가구)	전수▷세대주/세대주와 관계가공/유의점제시
5표.	세대구성 및 가구주의 교육정도별 가구 (일반가구)	표본??: 교육정도
6표.	세대구성 및 가구원수별 가구(일반가구)	전수▷세대주관계가공/세대구성원/유의점제시
7표.	가구의 크기 및 핵수별 가구(일반가구)	전수▷세대주관계를 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 소유여부파악가능
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

제6절 종합 및 결론

- 이 연구는 인구, 가구, 주택분야에서 행정자료를 이용하여 등록센서스 결과를 생산하는 방식에 대한 평가적 연구이다. 이 연구를 위해 이 연구에 함께 참여하고 있는 통계학자들이 능형회귀추정량(Ridge Regression Estimator)을 이용하여 센서스 모집단(곧, 인구, 가구, 주택)의 크기를 산출하고, 이를 이용하여, 개체자료(individual records)에 가중치를 설정하는 방안을 마련하였다. 이러한 개체자료에 대한 “센서스 가중치”(census weights)로 보완된 행정자료를 이용하여, 등록센서스 항목에 대하여, 네 가지의 분포표를 산출하였다.
- 가구나 주택 부문의 경우에도 능형회귀추정량을 가구수, 주택수를 산출하였는데, 이 자료를 이용하여 가구와 주택의 개체자료에 대하여 “센서스 가중치”를 설정하였으며, 이 가중치를 개체자료에 곱하여, 새로운 가구통계와 주택통계를 산출하였다.

- 먼저 인구부문에서는 12개 지역에 대하여 성별 연령별 혼인상태별 인구 분포 및 지역별 성별 연령별 가구주와의 관계 분포를 산출할 수 있었다. 이를 원래의 가중치 미적용 행정자료 및 센서스 가중치로 보완된 행정자료에서 분포표를 산출하게 되면 등록센서스 결과 생산방식에 대한 평가적 연구를 진행할 수 있다. 이 평가적 연구에는 두 자료 간에 1) 단순 차이를 구하는 방법, 2) 차이의 절대값을 구하는 방법, 3) 두 분포의 비(ratio)를 구하는 방법을 통해 수백 개 이상의 표를 산출하고, 이를 검증하는 작업을 수행해야 했다.
- 인구학에서는 두 분포의 차이를 평가하기 위해 비유사성지수 (index of dissimilarity)를 구해서 비교한다. 이 연구에서도 인구자료에서 가중치 미적용의 행정자료와 릿지 1, 2, 3, 4 등 4개 세트의 센서스 가중치를 적용하여 보완된 행정자료 간에 분포의 차이를 보기 위해 비유사성지수를 구해 비교해 보았다. 인구자료의 품질 평가에서는 지역별, 성별, 연령별 혼인상태의 분포만 평가하였다.
- 성별 분포를 보게 되면 동일한 지역에서도 남자의 가중치 미적용 행정자료 및 릿지 1, 2, 3, 4 등 센서스 가중치를 적용하여 보완된 행정자료의 분포 사이에서 얻어진 비유사성지수가 여자의 경우에 얻어진 비유사성지수와 차이가 있을 수 있다. 그래서 인구부문 등록센서스 결과 생산방식의 평가에서는 남자 자료와 여자 자료의 분포 차이도 비교하였다.
- 이 연구에서 인구 부문은 등록센서스 결과 생산방식을 평가할 때, 주안점은 가중치를 적용하지 않은 행정자료와 가중치를 적용한 행정자료의 센서스 항목별 분포를 비교하는데 두었지만, 등록센서스 결과 생산방식을 평가하는 또 다른 중요한 방법은 다른 자료원의 인구자료와 행정 자료원의 인구자료를 비교하는 것이다. 이를 위해 2010 인구주택총조사에서 해당 12개 지역의 성별, 연령별, 혼인상태 분포를 구하고, 인구주택총조사의 인구분포를 기준으로 가중치 적용 전후의 행정자료들의 분포를 비교하면서 행정자료의 센서스 항목별 품질도 평가하였다. 이 역시 남자의 혼인상태 분포와 여자의 혼인상태 분포에 차이가 있을 수 있으므로 성별에 따른 비유사성지수의 차이도 검토하였다.

- 검토 결과, 릿지 2 방식으로 센서스 가중치를 설정하여 보완된 행정자료의 분포가 원래 가중치 미적용의 행정자료와 분포가 비슷하고, 릿지 4 방식으로 보완된 행정자료의 분포가 원 행정자료의 분포와 가장 차이가 남을 확인하였다. 릿지 1, 3 방식의 분포는 그 중간에 위치하고 있었다. 이 결과가 릿지 1 혹은 릿지 3 방식으로 등록센서스 결과 생산하는 것이 가중치 보정 이전의 행정자료와의 괴리를 줄이는데 기여하지만, 그것이 행정자료의 센서스 항목에 “내용오류”(content error)가 없다는 것을 확인하는 것은 아니다.
- 가구와 주택은 센서스 가중치가 하나만 주어지기 때문에 비교가 상대적으로 단순하였다. 행정자료의 가구자료와 주택자료의 등록센서스 결과 생산방식을 평가하기 위해 역시 인구자료와 마찬가지로 비유사성지수를 구해 가중치 적용 이전과 이후의 행정자료 간 분포의 차이를 검토하였다.
- 비유사성 지수와 함께, 이 연구에서는 인구센서스와 행정자료 간 오차크기비율을 살펴보았다. 연령대별 오차크기는 대체로 15세 미만 연령층과 고령층에서 높게 나타났으며, 이러한 현상은 남녀간 큰 차이를 보이지 않았다. 혼인상태별 오차크기는 남녀에 따라 다소 차이를 보였다. 남자의 경우 지역에 따라 혼인상태별 오차크기가 심한 차이를 보였다. 이에 비해 여성들에게서 다소 안정적인 오차크기가 관찰되었다. 또한 남성에 비해 이혼과 사별자들의 오차크기가 매우 크다는 점도 부각되었다.
- 가구주와의 관계에 따른 인구센서스와 행정자료 간 오차크기 또한 남녀에 따라 차이를 보였다. 남자들이 경우는 대체로 가구주와 자녀에서, 여자들의 경우는 이외에 배우자에서 오차가 크게 나타났다. 또한 대도시 지역일수록 기타 친인척과 기타 동거인에서 발생한 오차비율이 상대적으로 높게 발견되었다. 이 경우에도 가구주 관계에 대한 행정자료의 품질을 개선하고, 필요에 따라 가구생성의 규칙을 마련하여야 할 필요가 있을 것이다.
- 가구원수별 인구센서스와 행정자료 간 오차크기비율은 대체로 1인가구에서 가장 높고, 2인가구가 그 뒤를 이었다. 그러나 지역별 차이가 일관적이지는 않은 것으로 보인다, 다만 광주 동구의 경우, 2인가구의 오차비율이 매우 크다는 점

이 부각되었다.

- 주택연면적별 인구센서스와 행정자료간 오차크기비율에서는 연면적이 아주 작거나 큰 주택에서 오차크기가 작은 것으로 나타났다. 이 또한 지역간 편차가 매우 큰 것으로 보인다. 주택구분별 오차크기에서는 농촌지역에서는 단독주택이 도시지역에서는 아파트의 오차크기가 대체로 높게 나타났다. 건축연도별 오차크기에서는 가장 최근에 지은 집이나 2004년 이전에 지은 집들의 오차크기가 대체로 높게 나타났다.
- 이 연구에서는 행정자료를 센서스 가중치로 보완하기 전과 보완한 후의 오차크기를 통해 추정치의 질을 검토하였다. 연령대별 오차크기에서는 앞서 살펴본 인구센서스와 행정자료간의 오차크기 유형과는 달리, 대체로 20대 연령층에서 오차가 큰 것으로 나타났다. 여자들의 경우는 20대 미만 층에서도 다소 오차비율이 높게 나타났다. 혼인상태별 오차크기에서는 남성들에 비해 여자들이 다소 지역 간 편차가 적게 나타났다. 앞서 살펴본 인구센서스와 행정자료간의 오차유형과는 달리, 여성들의 경우도 이혼이나 사별자들의 오차크기가 작게 나타났다.
- 가구주와의 관계별 가중치 미적용의 행정자료와 가중치를 적용한 행정자료 간 오차크기에서는 인구센서스와 행정자료간의 오차크기 유형과 유사한 모습이 관찰되었다. 남성들의 경우는 대체로 가구주나 자녀에게서, 여자들의 경우는 이 외에 배우자에게서도 오차가 큰 것으로 나타났다.
- 가구원수별 가중치 미적용의 행정자료와 가중치를 적용한 행정자료 간 오차크기비율도 대체로 1인가구와 2인가구에서 높게 나타나, 앞서 살펴본 인구센서스와 행정자료간의 오차유형과 유사한 모습을 보였다. 또한 광주 동구의 경우, 2인가구의 오차비율이 매우 크다는 점이 여기서도 재확인되었다. 가구유형별 오차크기 비율에서는 역시 가족가구와 1인가구의 오차비율이 가장 높다. 가족가구 오차가 높은 지역은 1인가구 오차비율이 매우 낮게 나타난다. 1인가구 오차비율이 높은 지역은 대체로 서울지역이라 할 수 있으나 지역별 일관성을 논하기는 힘들다.

- 주택연면적별 가중치 미적용의 행정자료와 가중치를 적용한 행정자료 간 오차크기비율에서도 연면적이 아주 작거나 큰 주택에서 오차크기가 작은 것으로 나타났다. 이러한 현상은 인구센서스와 행정자료간의 오차크기 유형과 유사하다. 주택구분별 오차크기분포는 인구센서스와 행정자료간의 오차크기 유형과는 차이를 보인다. 또한 건축연도별 오차크기 분포 또한 인구센서스와 행정자료간의 오차크기 유형과 차이를 보이고 있다. 건축연도별 오차크기가 가장 최근에 건축한 주택의 경우는 낮다.
- 이 연구는 기본적으로 가중치 미적용의 행정자료를 기준으로, 능형회귀추정량으로 산출된 인구, 가구, 주택의 총수를 비교하여 센서스 가중치를 계산하여 새롭게 보완된 행정자료를 만든 후 이들 행정자료를 비교하면서 가중치 미적용의 행정자료와 가중치를 적용한 행정자료의 품질을 진단하여 등록센서스 결과 생산방식을 종합적으로 평가하는 연구이다. 이 연구는 앞으로 행정자료를 잘 보완하여 등록센서스를 실시할 때 중요한 시사점을 던져 줄 것이다.
- 그러나 행정자료의 보완이 인구자료의 경우 성별, 연령별 총수의 보완에 중점을 두고 있었고, 이 결과물을 혼인상태 등 추가적인 변수를 대입하여 새로운 분포를 만들면 원래 생각하지 못했던 여러 문제점을 찾아볼 수 있었다. 더욱이 행정자료의 ‘자료’로서의 성격과 품질을 평가하기 위해 인구주택총조사 자료 등을 대입하는 경우 혼인상태의 정의가 행정자료와 조사자료에서 달라짐으로 해서 새로운 문제를 확인할 수 있었다. 이러한 연구는 행정자료의 품질을 포함해서 등록센서스 결과 생산방식을 진단하는데 매우 중요한 연구주제이다. 2012년의 연구에서는 이런 점이 충분히 이루어지지 못했다. 행정자료의 정리, 능형회귀추정량을 이용한 센서스 가중치 산출 및 이를 이용하여 보완된 행정자료에서의 새로운 분포의 산출에 오랜 기간이 소요되어 추가적인 작업을 진행하지 못했다. 2013년의 연구에서는 이런 부분이 대폭 보완될 것으로 기대한다.

제7절 제언: 2013년 과제

- 2012년의 연구는 행정자료를 기준으로 행정자료의 인구 부문에서는 성별, 연령별 총수를, 가구측면에서는 가구구분별 총수를 그리고 주택 부문에서는 역시 주택구분별 총수를 기준으로 능형회귀추정량을 행정자료의 크기와 비교하면서 센서스 가중치를 산출했다. 이 가중치를 이용하여 행정자료를 이용하여 등록센서스 결과 작성방식을 진단하기 위해 인구의 경우 혼인상태와 가구주와의 관계, 가구의 경우 가구원수, 주택의 경우 주택 연면적, 주택의 건축연도라는 새로운 변수를 도입하여 가중치 미적용 행정자료 상의 분포와 가중치를 적용하여 보완된 행정자료의 분포를 비교하였다.
- 경우에 따라 성별, 연령별 등의 총수만을 가지고 보정치를 산출했는데, 혼인상태 등 제 3의 변수를 도입하여 분포를 새로 만들고 이를 품질진단작업에 사용할 필요가 있는지 의문을 제기할 수도 있다. 그러나 바로 그 이유 때문에 즉 보정치를 산출하는데 여러 요인을 다 고려하기 힘들다. 따라서 대표적인 요인을 사용해 가중치를 산출한다. 그러나 실제로 통계청에서 나중에 여러 형태를 집계표를 공표할 때는 실로 다양한 항목의 변수들을 이용한다.
- 따라서 2013년 행정자료의 품질 검증 작업에는 보다 더 다양한 제3의 센서스 항목을 이용한 분포를 산출하고, 이를 행정자료, 행정자료를 센서스 가중치로 보완한 자료 등에 제 3의 변수를 대입하여 새롭게 만든 여러 분포표를 비교 검토하여, 어떤 변수를 새로 추가했을 때 문제가 큰지, 아니면 별로 문제가 없는지 등을 점검하여, 센서스 항목별 특성을 이해하여야 한다. 즉, 행정자료의 품질 검증에는 센서스 가중치 작성에 고려된 보정요인만 고려해서는 안 되고, 보정요인 아닌 제3의 변수를 가능한 더 많이 활용하여 여러 변수를 새로 도입해도 별 문제가 없는지, 아니면 전혀 예상하지 못한 새로운 변화가 생기는지를 검증해야 한다. 이 때문에, 2013년에는 인구의 경우 혼인상태 및 가구주와의 관계 이외의 변수들, 가구의 경우 가구원수 이외의 변수들, 주택의 경우 주택의 연면적과 건축연도 이외의 변수들을 활용한 분포표를 만들고 이 분포표를 통해 행정자료를 이용하는 등록센서스 결과 생산방식을 종합적으로 평가해

야 할 것이다.

- 보다 더 중요한 것은 행정자료의 품질을 평가하기 위해 행정자료 아닌 다른 자료, 가장 대표적으로는 센서스 자료를 이용하여 행정자료의 품질을 평가하는 작업을 체계적으로 진행하여야 한다. 센서스 자료도 완전한 자료가 아니다. 그러나 사후조사 등을 통해 센서스 자료는 보완되고 있다. 이런 센서스 자료와 비교해서 행정자료는 어떤 특성 혹은 현상을 보이고 있는가, 어디서 과다 혹은 과소집계가 이루어지고 있는가를 검증하는 것이 행정자료의 안정적 이용을 위해 필수적이다. 또한 시계열적 분석을 이어가기 위해서도 행정자료를 센서스 자료와 연계해서 품질을 평가하는 것이 필수적이다. 2013년에는 센서스 자료를 기준으로 행정자료, 그리고 행정자료를 가중치로 보완한 자료를 모두 함께 비교하면서 평가하는 작업이 진행되어야 한다.

<부표 5-1> 2010 인구주택총조사: 금천구의 성별 연령별 혼인상태 분포

금천구	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	7,949	21	4	1	7,975
	20~24세	7,653	82	3	0	7,738
	25~29세	9,825	1,083	28	4	10,940
	30~34세	6,755	3,870	117	2	10,744
	35~39세	3,911	6,109	379	16	10,415
	40~44세	2,225	6,901	784	57	9,967
	45~49세	1,306	7,653	1,029	115	10,103
	50~54세	598	7,871	1,026	177	9,672
	55~59세	203	6,135	641	220	7,199
	60~64세	79	5,110	394	277	5,860
	65~69세	40	4,045	234	311	4,630
	70~74세	20	2,552	126	306	3,004
	75~79세	6	1,216	43	252	1,517
	80~84세	2	390	8	151	551
	85세+	0	145	1	123	269
여자	15~19세	7,205	37	2	1	7,245
	20~24세	6,924	204	9	0	7,137
	25~29세	7,784	2,071	46	3	9,904
	30~34세	3,652	5,256	228	16	9,152
	35~39세	1,796	6,864	668	75	9,403
	40~44세	829	7,527	1,091	188	9,635
	45~49세	374	8,036	1,272	445	10,127
	50~54세	260	7,775	1,134	698	9,867
	55~59세	121	5,865	652	991	7,629
	60~64세	59	4,328	375	1,349	6,111
	65~69세	42	2,829	179	1,659	4,709
	70~74세	22	1,496	86	1,819	3,423
	75~79세	19	606	44	1,537	2,206
	80~84세	11	139	15	1,230	1,395
	85세+	9	47	5	915	976

<부표 5-2> 2010 인구주택총조사: 영등포구의 성별 연령별 혼인상태 분포

영등포구	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	11,259	30	3	1	11,293
	20~24세	11,723	88	2	1	11,814
	25~29세	15,393	1,642	27	1	17,063
	30~34세	10,241	8,147	153	8	18,549
	35~39세	5,842	12,417	462	21	18,742
	40~44세	2,985	11,828	844	53	15,710
	45~49세	1,491	11,736	1,115	137	14,479
	50~54세	763	12,298	1,154	183	14,398
	55~59세	358	9,866	751	290	11,265
	60~64세	141	8,331	506	380	9,358
	65~69세	57	6,424	296	459	7,236
	70~74세	43	4,591	132	465	5,231
	75~79세	12	2,385	45	393	2,835
	80~84세	3	889	20	250	1,162
	85세+	0	370	2	226	598
여자	15~19세	9,976	23	1	0	10,000
	20~24세	11,509	223	2	2	11,736
	25~29세	13,420	3,968	55	5	17,448
	30~34세	6,920	11,201	270	20	18,411
	35~39세	3,267	12,725	734	73	16,799
	40~44세	1,553	11,686	1,109	206	14,554
	45~49세	773	11,914	1,350	468	14,505
	50~54세	482	12,853	1,458	982	15,775
	55~59세	245	9,631	941	1,373	12,190
	60~64세	135	7,149	510	1,957	9,751
	65~69세	86	4,920	282	2,392	7,680
	70~74세	53	3,024	127	2,793	5,997
	75~79세	33	1,231	59	2,681	4,004
	80~84세	17	339	16	1,948	2,320
	85세+	5	116	3	1,515	1,639

<부표 5-3> 2010 인구주택총조사: 인천 서구의 성별 연령별 혼인상태 분포

인천서구	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	15,923	39	6	3	15,971
	20~24세	11,994	163	7	0	12,164
	25~29세	10,746	2,320	45	3	13,114
	30~34세	6,353	8,429	237	5	15,024
	35~39세	4,187	14,125	672	44	19,028
	40~44세	2,555	16,094	1,302	85	20,036
	45~49세	1,345	16,645	1,717	153	19,860
	50~54세	605	13,525	1,519	284	15,933
	55~59세	218	8,114	830	282	9,444
	60~64세	106	5,571	394	335	6,406
	65~69세	45	3,827	230	362	4,464
	70~74세	28	2,565	92	371	3,056
	75~79세	8	1,440	52	312	1,812
	80~84세	3	555	11	246	815
	85세+	5	259	7	191	462
여자	15~19세	14,312	37	2	5	14,356
	20~24세	10,212	501	19	0	10,732
	25~29세	7,925	4,800	121	8	12,854
	30~34세	3,113	11,848	414	31	15,406
	35~39세	1,490	16,536	1,092	127	19,245
	40~44세	746	16,920	1,857	342	19,865
	45~49세	404	15,199	2,005	686	18,294
	50~54세	245	11,418	1,481	1,067	14,211
	55~59세	132	6,666	757	1,232	8,787
	60~64세	69	4,253	377	1,559	6,258
	65~69세	38	2,736	181	1,918	4,873
	70~74세	20	1,815	103	2,546	4,484
	75~79세	23	809	39	2,448	3,319
	80~84세	10	231	18	1,883	2,142
	85세+	8	84	3	1,319	1,414

<부표 5-4> 2010 인구주택총조사: 광주 동구의 성별 연령별 혼인상태 분포

광주동구	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	3,868	6	0	0	3,874
	20~24세	5,203	62	3	2	5,270
	25~29세	4,750	344	7	1	5,102
	30~34세	2,607	1,090	54	3	3,754
	35~39세	1,493	1,858	144	11	3,506
	40~44세	685	2,302	257	24	3,268
	45~49세	395	2,620	338	58	3,411
	50~54세	247	3,101	434	103	3,885
	55~59세	105	2,773	310	164	3,352
	60~64세	57	2,634	146	178	3,015
	65~69세	29	2,284	76	227	2,616
	70~74세	19	1,878	28	246	2,171
	75~79세	8	1,087	10	209	1,314
	80~84세	3	458	4	127	592
	85세+	0	166	2	86	254
여자	15~19세	3,642	10	2	0	3,654
	20~24세	5,092	120	6	2	5,220
	25~29세	3,679	707	29	0	4,415
	30~34세	1,547	1,578	91	5	3,221
	35~39세	654	2,245	197	22	3,118
	40~44세	263	2,530	332	77	3,202
	45~49세	138	2,936	417	179	3,670
	50~54세	97	3,428	427	350	4,302
	55~59세	59	2,852	281	582	3,774
	60~64세	33	2,495	129	835	3,492
	65~69세	19	1,850	60	1,027	2,956
	70~74세	12	1,268	31	1,137	2,448
	75~79세	9	584	10	1,131	1,734
	80~84세	3	179	3	855	1,040
	85세+	2	34	0	695	731

<부표 5-5> 2010 인구주택총조사: 화성시의 성별 연령별 혼인상태 분포

화성시	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	13,746	46	2	0	13,794
	20~24세	11,521	182	3	1	11,707
	25~29세	14,177	4,351	54	5	18,587
	30~34세	9,101	17,056	216	11	26,384
	35~39세	4,775	23,471	607	36	28,889
	40~44세	2,352	20,771	938	79	24,140
	45~49세	1,047	16,060	1,047	133	18,287
	50~54세	474	12,273	863	200	13,810
	55~59세	182	8,198	536	228	9,144
	60~64세	67	6,347	308	297	7,019
	65~69세	32	4,731	207	364	5,334
	70~74세	17	3,981	90	451	4,539
	75~79세	6	2,334	33	472	2,845
	80~84세	6	881	8	310	1,205
	85세+	1	393	4	257	655
여자	15~19세	12,102	60	0	2	12,164
	20~24세	9,467	772	12	1	10,252
	25~29세	7,763	9,415	131	12	17,321
	30~34세	3,217	21,958	343	48	25,566
	35~39세	1,534	23,855	823	112	26,324
	40~44세	791	17,689	1,061	233	19,774
	45~49세	438	13,267	1,149	457	15,311
	50~54세	255	10,853	1,051	851	13,010
	55~59세	155	7,230	568	1,269	9,222
	60~64세	63	5,179	325	1,865	7,432
	65~69세	54	3,896	151	2,269	6,370
	70~74세	34	2,894	65	2,957	5,950
	75~79세	20	1,461	40	3,018	4,539
	80~84세	10	425	23	2,346	2,804
	85세+	6	132	7	1,881	2,026

<부표 5-6> 2010 인구주택총조사: 양구군의 성별 연령별 혼인상태 분포

양구군	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	529	0	0	0	529
	20~24세	862	50	0	0	912
	25~29세	507	339	0	1	847
	30~34세	218	454	7	0	679
	35~39세	136	518	29	2	685
	40~44세	113	552	47	4	716
	45~49세	81	655	54	5	795
	50~54세	42	740	68	12	862
	55~59세	17	562	32	9	620
	60~64세	4	322	20	15	361
	65~69세	1	433	14	36	484
	70~74세	1	424	3	45	473
	75~79세	1	254	3	34	292
	80~84세	0	101	1	26	128
	85세+	0	64	0	36	100
여자	15~19세	452	3	0	0	455
	20~24세	150	50	0	1	201
	25~29세	170	310	3	0	483
	30~34세	77	488	5	0	570
	35~39세	41	534	10	5	590
	40~44세	22	533	39	9	603
	45~49세	20	654	27	26	727
	50~54세	17	670	29	38	754
	55~59세	15	426	13	75	529
	60~64세	4	373	11	113	501
	65~69세	3	382	10	232	627
	70~74세	4	321	2	302	629
	75~79세	1	139	1	301	442
	80~84세	0	38	5	252	295
	85세+	2	12	0	158	172

<부표 5-7> 2010 인구주택총조사: 증평군의 성별 연령별 혼인상태 분포

증평균	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	1,054	2	0	0	1,056
	20~24세	864	30	0	0	894
	25~29세	937	349	8	0	1,294
	30~34세	512	699	19	2	1,232
	35~39세	337	958	60	4	1,359
	40~44세	208	1,015	94	5	1,322
	45~49세	121	1,019	109	16	1,265
	50~54세	54	968	87	20	1,129
	55~59세	25	778	74	25	902
	60~64세	9	589	36	29	663
	65~69세	6	487	12	49	554
	70~74세	1	491	10	68	570
	75~79세	2	277	2	55	336
	80~84세	1	135	1	42	179
	85세+	0	37	0	33	70
여자	15~19세	1,049	4	0	0	1,053
	20~24세	769	74	2	0	845
	25~29세	430	524	8	0	962
	30~34세	188	886	25	4	1,103
	35~39세	89	1,028	69	14	1,200
	40~44세	30	982	80	37	1,129
	45~49세	32	929	97	52	1,110
	50~54세	24	905	86	108	1,123
	55~59세	11	697	54	133	895
	60~64세	2	464	21	183	670
	65~69세	4	431	16	287	738
	70~74세	2	345	4	359	710
	75~79세	4	213	4	428	649
	80~84세	1	70	1	310	382
	85세+	0	16	0	186	202

<부표 5-8> 2010 인구주택총조사: 임실군의 성별 연령별 혼인상태 분포

임실군	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	467	5	1	1	474
	20~24세	378	15	0	0	393
	25~29세	308	55	0	1	364
	30~34세	225	176	2	1	404
	35~39세	174	326	30	2	532
	40~44세	145	491	39	3	678
	45~49세	96	621	55	7	779
	50~54세	61	744	72	13	890
	55~59세	29	688	39	30	786
	60~64세	15	850	30	45	940
	65~69세	5	828	13	57	903
	70~74세	3	975	5	90	1,073
	75~79세	0	683	4	100	787
	80~84세	0	374	1	70	445
	85세+	0	124	0	64	188
여자	15~19세	364	3	0	0	367
	20~24세	174	36	1	0	211
	25~29세	151	136	1	0	288
	30~34세	67	291	6	2	366
	35~39세	27	377	16	6	426
	40~44세	28	480	15	17	540
	45~49세	25	624	29	29	707
	50~54세	11	743	23	79	856
	55~59세	13	730	21	126	890
	60~64세	9	875	9	258	1,151
	65~69세	3	826	5	439	1,273
	70~74세	6	820	7	760	1,593
	75~79세	5	432	3	876	1,316
	80~84세	2	139	2	665	808
	85세+	0	26	0	447	473

<부표 5-9> 2010 인구주택총조사: 나주시의 성별 연령별 혼인상태 분포

나주시	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	2,645	8	0	0	2,653
	20~24세	1,934	32	0	0	1,966
	25~29세	1,560	248	6	0	1,814
	30~34세	1,029	794	38	1	1,862
	35~39세	762	1,394	104	12	2,272
	40~44세	555	1,752	160	17	2,484
	45~49세	398	2,057	254	61	2,770
	50~54세	228	2,402	212	87	2,929
	55~59세	126	2,104	153	112	2,495
	60~64세	75	2,170	97	195	2,537
	65~69세	37	2,171	65	207	2,480
	70~74세	15	2,092	19	262	2,388
	75~79세	4	1,435	9	303	1,751
	80~84세	4	712	3	240	959
	85세+	5	300	2	152	459
여자	15~19세	2,558	6	0	1	2,565
	20~24세	1,452	103	2	2	1,559
	25~29세	893	551	15	3	1,462
	30~34세	408	1,230	44	10	1,692
	35~39세	194	1,641	99	17	1,951
	40~44세	109	1,847	159	50	2,165
	45~49세	106	2,079	171	132	2,488
	50~54세	79	2,454	129	306	2,968
	55~59세	28	2,102	61	501	2,692
	60~64세	23	2,031	40	811	2,905
	65~69세	22	1,886	30	1,324	3,262
	70~74세	12	1,698	10	1,998	3,718
	75~79세	8	927	6	2,259	3,200
	80~84세	5	314	1	1,718	2,038
	85세+	8	87	1	1,213	1,309

<부표 5-10> 2010 인구주택총조사: 신안군의 성별 연령별 혼인상태 분포

신안군	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	553	2	0	0	555
	20~24세	516	15	1	0	532
	25~29세	380	107	4	0	491
	30~34세	322	260	20	1	603
	35~39세	313	506	31	6	856
	40~44세	294	718	67	9	1,088
	45~49세	212	1,011	101	17	1,341
	50~54세	157	1,299	103	37	1,596
	55~59세	64	1,214	60	65	1,403
	60~64세	35	1,465	38	87	1,625
	65~69세	7	1,424	23	150	1,604
	70~74세	10	1,331	6	139	1,486
	75~79세	2	810	2	136	950
	80~84세	0	318	1	102	421
	85세+	0	108	1	67	176
여자	15~19세	457	4	0	0	461
	20~24세	132	22	1	0	155
	25~29세	159	141	3	0	303
	30~34세	68	297	6	3	374
	35~39세	37	461	9	6	513
	40~44세	19	644	16	18	697
	45~49세	17	992	22	50	1,081
	50~54세	15	1,202	15	87	1,319
	55~59세	8	1,210	10	168	1,396
	60~64세	2	1,227	5	382	1,616
	65~69세	2	1,248	3	708	1,961
	70~74세	4	1,082	2	1,181	2,269
	75~79세	0	579	2	1,166	1,747
	80~84세	1	193	2	788	984
	85세+	5	54	1	573	633

<부표 5-11> 2010 인구주택총조사: 김천시의 성별 연령별 혼인상태 분포

김천시	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	4,649	10	2	3	4,664
	20~24세	3,644	45	3	0	3,692
	25~29세	2,711	527	10	3	3,251
	30~34세	1,764	1,866	40	3	3,673
	35~39세	1,129	2,965	134	20	4,248
	40~44세	716	3,746	262	38	4,762
	45~49세	404	4,365	369	68	5,206
	50~54세	226	4,496	353	110	5,185
	55~59세	92	3,866	196	157	4,311
	60~64세	47	3,303	153	210	3,713
	65~69세	21	2,579	69	226	2,895
	70~74세	19	2,596	31	327	2,973
	75~79세	9	1,857	16	311	2,193
	80~84세	2	824	2	261	1,089
	85세+	2	335	2	216	555
여자	15~19세	3,842	14	0	2	3,858
	20~24세	2,480	161	3	2	2,646
	25~29세	1,874	1,121	33	3	3,031
	30~34세	774	2,642	68	16	3,500
	35~39세	364	3,496	169	38	4,067
	40~44세	198	3,983	294	116	4,591
	45~49세	121	4,261	336	239	4,957
	50~54세	104	4,408	309	504	5,325
	55~59세	52	3,735	184	771	4,742
	60~64세	51	2,771	84	1,040	3,946
	65~69세	31	2,280	51	1,549	3,911
	70~74세	29	2,096	39	2,327	4,491
	75~79세	10	1,041	20	2,517	3,588
	80~84세	11	337	17	1,938	2,303
	85세+	6	89	4	1,398	1,497

<부표 5-12> 2010 인구주택총조사: 울릉군의 성별 연령별 혼인상태 분포

울릉군	연령별	미혼	유배우	이혼	사별	계
남자	15~19세	166	1	0	0	167
	20~24세	143	3	0	0	146
	25~29세	152	26	1	0	179
	30~34세	106	133	2	0	241
	35~39세	82	218	9	0	309
	40~44세	57	265	19	1	342
	45~49세	33	289	25	4	351
	50~54세	19	392	24	12	447
	55~59세	9	308	24	14	355
	60~64세	8	274	11	14	307
	65~69세	2	222	5	13	242
	70~74세	2	188	2	19	211
	75~79세	0	109	1	12	122
	80~84세	1	58	0	21	80
	85세+	0	15	0	5	20
여자	15~19세	112	0	0	0	112
	20~24세	63	6	0	0	69
	25~29세	81	67	0	0	148
	30~34세	41	154	1	0	196
	35~39세	13	193	10	2	218
	40~44세	12	251	10	6	279
	45~49세	11	285	17	12	325
	50~54세	6	329	15	28	378
	55~59세	5	267	10	42	324
	60~64세	3	221	5	79	308
	65~69세	0	188	2	106	296
	70~74세	2	141	2	133	278
	75~79세	0	55	0	124	179
	80~84세	0	16	0	101	117
	85세+	0	6	0	68	74

제6장 2010년과 2020년 라운드 등록센서스 관련 동향의 검토

- 본 장에서는 우리나라가 2015년 등록센서스를 준비하면서 2012년 학술용역 연구의 과제를 정리하면서, 세계 각국의 등록센서스 실시경과 및 준비상황을 중심으로 살펴보고, 그것이 우리나라 등록센서스 준비와 시행에 주는 함의를 요약하기로 한다,
- 우리나라는 2015년에 등록부 자료를 기준으로 하는 인구주택총조사를 실시할 예정으로 되어 있다. 이것은 종전 특히, 2010년 센서스에서 주택부문의 일부 항목에 대하여 건축물대장에 기재된 정보(예: 주택유형, 주택연면적)를 이용한 것을 제외하면, 2015년의 인구주택총조사는 행정(등록)자료와 조사자료¹⁾를 결합하여 센서스 모집단(인구, 가구, 주택)의 총수와 사회경제적 특성을 추정한다는 점에서 아시아 지역에서는 싱가포르, 이스라엘, 대만을 이어서 새로운 영역의 중대한 실험을 진행하고 있는 셈이다.
- 특히, 2015년의 인구주택총조사는 주민등록, 가족관계등록부, 건축물대장이 핵심 등록부가 되고 있음에도 불구하고, 현지실사의 결과를 주민등록 자료와 결합하는 것으로, 센서스 모집단(인구, 가구, 주택)의 크기는 물론 행정자료에 포함된 센서스 항목의 정확성을 체계적으로 검토한 경험이 사실상 전무(全無)한 상황에서, 전통센서스에서 등록센서스 방식으로 이행하는 것은 센서스 방법론의 패러다임 전환(paradigm shift)에서 또 하나의 새로운 이정표가 된다고 할 수 있다.

1) 본 장에서 행정자료는 주민등록(인구대장), 가족관계등록부, 건축물/주택대장 등의 행정 목적으로 도입된 등록부 자료를 가리키고, 조사자료는 센서스 또는 표본조사에서 획득된 자료를 가리키는 것으로 한다. 특히, 조사자료가 전수집계인 경우에는 '센서스'라고 할 수 있다.

제1절 등록부 이용 센서스 방식의 개관

- 인구주택총조사, 곧 센서스에는 등록부를 이용하는 센서스와 그렇지 않은 센서스의 두 종류가 존재한다. 또 등록부를 이용하는 센서스에도 등록부를 단순한 「보조정보」로 이용하는 센서스와 등록부를 「주축정보」로 이용하는 센서스가 있다.
- 단순히 등록부를 “보조정보”로 이용하는 센서스가 아니라, “주축정보”로 하는 센서스를 실시하여 통계를 작성하는 관행은 1970년대에 구상되어, 다년간 북유럽 국가에서 이용되어 왔다(Wallgren and Wallgren, 2007; UNECE, 2007). 현지실사 대신에 등록부의 포괄범위와 기재정보를 기준으로 하여 실시되는 센서스, 곧 등록센서스는 다음과 같은 특징을 공유한다.
 - a. 모든 나라의 경우에, 행정자료의 품질이 개선되면서 전통센서스가 행정자료로 대체되는 순차적 과정을 거친다. 이 과정은 15-20년에 걸쳐서 이루어지고, 100% 행정자료만을 이용하여 실시되는 센서스로 이행하는데 더 많은 시간이 요구된다.
 - b. 행정자료의 통계적 이용을 가능하게 하고, 개인 사생활을 보호하기 위하여 통계법이나 개인정보법과 같은 법령의 준비가 필요하다.
 - c. 국가통계의 작성에 대하여 일반대중의 행정자료 이용에 대한 수용(public approval)이 필요하다.
 - d. 개체자료의 매칭(또는 연계)(matching and linking)가 가능하도록, 개인 또는 개체 식별번호체계가 통합되어야 한다.
 - e. 등록부는 원칙적으로 포괄범위(coverage) 곧 과다범위나 과소범위의 문제가 없어야 하고, 등록부에 기재된 정보는 센서스 항목의 통계를 작성하는데 있어서 아무런 문제가 없을 정도로 신뢰도가 높아야 한다.

<그림 6-1> 세계 주요국 등록부 이용방식: 2000년 및 2010년 센서스 라운드

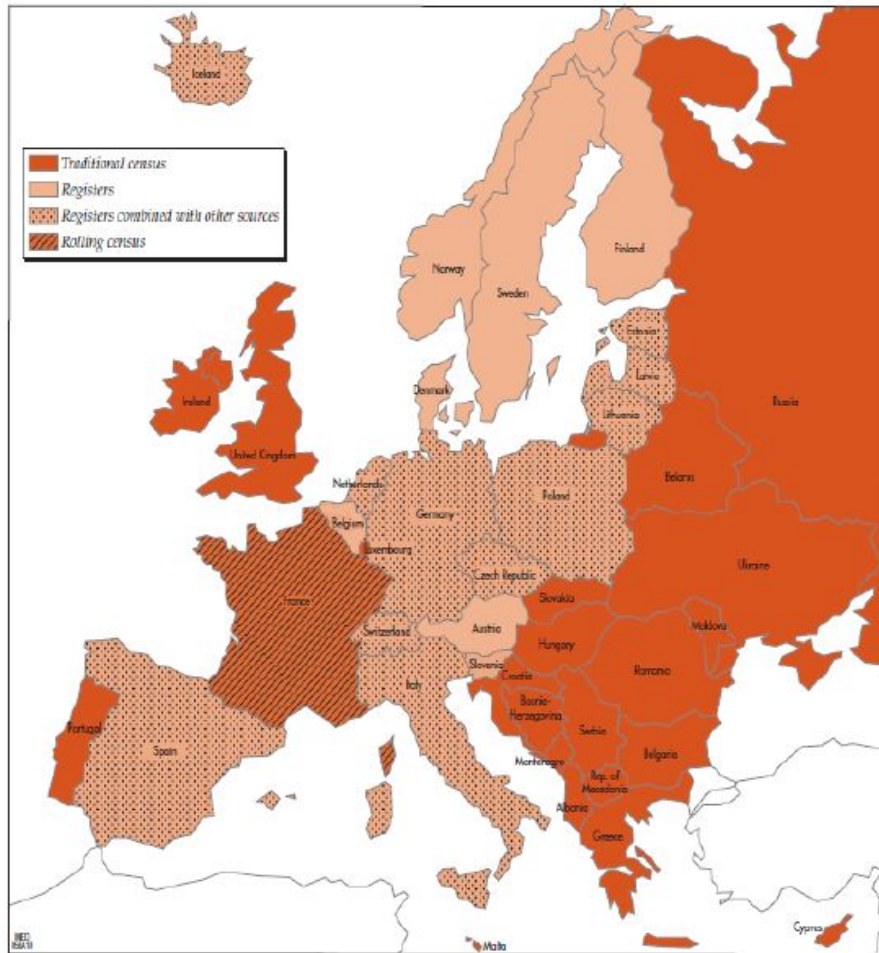
전수집계(전통센서스)	한국(2010), 영국(2011), 캐나다(2011), 스위스(2000) 등 유럽과 아시아 일부를 제외한 대부분
등록부 자료 + 전수집계(전통센서스)	스페인(2001), 포르투갈(2001, 2011), 이탈리아(2001, 2011) 체코(2010), 발트3국(에스토니아, 라트비아, 리투아니아) (2011)
등록부 자료(인구) + 전수집계(주택) + 특수목적의 표본조사	스페인(2011), 독일(2011)
등록부 자료 + 특수목적의 표본조사	폴란드(2011), 슬로베니아(2001), 이스라엘(2008) 스위스(2010), 싱가포르(2000, 2010), 대만(2010), 터키(2011), 한국(2015)
등록부 자료 + 기존의 표본조사	네덜란드(2001, 2011), 벨기에(2011), 아이슬란드(2011)
완전 등록센서스	덴마크(2001), 핀란드(2001), 안도라(2001), 덴마크(2011) 노르웨이(2011), 슬로베니아(2011), 스웨덴(2011)

- 행정 목적으로 도입된 등록부(register)를 이용하는 센서스는 (1) 등록부 자료 + 전수집계 (2) 등록부 자료(인구) + 전수집계 + 특수목적의 표본조사 (3) 등록부 자료 + 특수목적의 표본조사 (4) 등록부 자료 + 기존의 표본조사 (5) 완전 등록센서스 등으로 구분할 수 있다 (<그림 6-1> 참고). 위의 분류에서 (2)-(4)의 4가지 유형은 “통합센서스”(integrated census) 또는 “결합센서스”(combined census)로 명명할 수 있다. (1)의 경우는 등록부 자료가 센서스 통계를 작성하는데 이용되기는 하지만 주축자료가 아니라, 전통적 센서스의 현지실사를 지원하는 “보조정보”로 사용되고 있다.
- 영국 통계청은 2021년 센서스 모형을 준비하는 프로젝트에서 세계 각국에서 시도되는 센서스 모형을 (1) 행정자료를 이용하여 전통적 센서스를 보강하는 센서스 (2) 등록부 또는 행정자료를 기반으로 하는 접근방식(북유럽, 네덜란드, 스위스, 이스라엘, 이탈리아의 2021년 계획) (3) 전통적 센서스의 추가적 변형(프랑스의 순환센서스, long/short-form 센서스)으로 구분한다(UK Office for National Statistics, 2012a, 2012b). 첫 번째의 모형은 행정자료를 이용하는 센서스, 곧 본 연구의 “등록센서스”에 해당한다고 볼 수 있고.

세 번째의 모형의 경우 프랑스는 건축물대장²⁾을 이용하여 표본추출을 한다는 점에서, 행정자료가 표본의 추출틀(sampling frame)로 이용되기는 하지만, 등록센서스의 기본전제인 “등록부에서 센서스 항목의 변수를 직접 추출해야 한다는 조건에 부합되지 않는 행정자료의 이용형태이기 때문에, 본 연구에서는 논외(論外)의 대상으로 한다.

2) 프랑스의 건축물대장(RIL: Repertoire d'immeubles localises)은 당초 1999년 인구센서스 시점에 도입된 것으로, 주거용, 행정용, 산업용, 상업용 건물을 전부 기재하고 있는 리스트이다. 이 대장은 건축허가, 지방세 자료, 우체국 주소명부와 같은 행정자료를 이용하여 업데이트되며, 지방자치단체가 매년 점검하고, 프랑스 통계청(INSEE)의 최종 확인을 거치게 되어 있다.

<그림 6-2> 유럽지역의 센서스 방법론 지도: 2010년 라운드



Traditional Census: 전통적 센서스 Registers: 등록부
 Registers contributed with other sources: 등록부와 다른 자료의 결합
 Rolling Censuses: 순환센서스

자료: Paolo Valente (2010), "Census taking in Europe: how are populations counted in 2010?" (Population et Societes, No. 467, May 2010, INED)

1. 등록부 자료 + 전수집계

- 전수집계(전통적 센서스)와 등록부 자료를 동시에 이용하여 사실상, 두 종류의 센서스를 실시하는 것처럼 보이는 나라들이 있다. 이 때문에, 이러한 센서스는 “Register-supported Census” 또는 “Register-assisted Census”라고 불린다. 물론, 이것은 다음에 보는 “Register-based Census”와는 다른데, 그 차이는 등록부가 센서스 통계를 작성하는데 보조정보이나 아니면 주 측정정보이나에 따라 달라진다.
- 2000년대 라운드에 이러한 센서스를 실시했던 나라에는 이탈리아, 스페인, 포르투갈이 있었다. 2010년 센서스 라운드에서도 체코공화국, 에스토니아, 이탈리아, 라트비아, 리투아니아, 스페인 등 모두 6개 나라³⁾가 있다.
- 우리나라는 2010년까지 전통적 센서스를 실시하였으며, 전수항목은 long-form 조사표로 표본항목은 short-form 조사표로 하여, 조사 모집단으로부터 직접 자료를 수집하였다. 2010년 센서스에서, 우리나라도 전수집계 센서스 항목에서, 건축물 대장과 같은 행정자료를 일부 활용하기는 하였지만, 센서스의 핵심 정보인 인구규모를 정확하게 추정하기 위하여 등록센서스의 백본(backbone)이라고 할 수 있는 주민등록이나 가족관계등록부를 적극적으로 활용하여 인구나 가구를 집계하지는 않았다는 점에서, 등록부의 기재내역을 보조정보(auxiliary information)로 이용하는 센서스 실시국가로 분류하는 것도 무리일 것이다.
- 스페인은 2010년 라운드의 등록센서스를 실시하기 전에, 곧 2001년 센서스에서는 인구대장과 같은 등록부의 행정자료와 전통적 센서스의 전수집계를 결합하면서, 동시에 토지대장(Cadaster)⁴⁾, 국민신분증 및 주민증(내무부 주관), 인

3) 이 중에서 발트 3국(Three Baltic States)는 에스토니아, 라트비아, 리투아니아다. 이들은 북·중유럽에 위치하며, 발트 해의 동쪽 끝, 러시아연방과 벨라루스(Belarus)의 서쪽 경계선에 입지하고 있다. 이들 나라는 1991년 구(舊) 소련으로부터 독립했으며, 제2차 세계대전 이전에는 나치독일에 의하여 발트지역의 4번째 국가로 간주되었다.

4) 토지대장은 토지에 관련된 물리적 현황과 법적 권리관계 등을 등록·공시하여, 국가적

구동태통계, 사회보장자료, 국세청 자료를 보조정보로 이용하였으며 전기소비량, 최종교육연수, 실업 관련 자료도 센서스 모집단의 크기, 특히 인구규모를 확정하는 목적에 이용하고자 노력하였다.

- 등록부 자료와 전수집계(전통센서스) 자료를 동시에 이용하는 센서스는 등록부나 여타 행정자료에 센서스 목적에 필요한 특정 조사항목에 대한 자료가 누락되어 있고, 표본조사로는 불충분한 정도로 상세한 자료가 요구되는 경우에 실시한다. 또, 통계법에 센서스 실시의 목적을 인구대장의 정보내용을 보완하는 것이라고 규정하는 경우에, 전수집계 결과를 인구대장의 과다범위와 과소범위를 통계적으로 추정하여, 센서스가 마땅히 집계해야 할 목표인구(census target population)를 추정하기 위한 기초자료로 사용한다.
- 등록부에서 일부의 센서스 항목에 대하여 자료를 입수할 수 있지만, 센서스의 정의에 부합되는 통계단위(인구, 가구, 주택) 전부에 대하여, 모든 센서스 항목에 대한 정확한 자료를 얻기 위하여 전수집계(전통센서스)를 실시한다. 전수집계 결과는 포괄범위를 평가하고, 등록부 자료의 정확성을 점검하는데 사용한다. 등록부 자료의 핵심 항목(성명, 성별, 연령, 주소)은 조사표에 미리 인쇄하여 배포하여, 등록부에 기재된 통계단위가 전부 현지실사에 집계되는가를 확인하며, 전수집계는 등록부 행정자료의 정확성을 100% 전면적으로 점검하여, 센서스 목표인구를 집계 또는 추정하는데 사용된다.
- 이러한 방식으로 등록부의 행정자료를 이용하여 전수집계를 하는 전통적 센서스가 갖는 장점은 우선, 자료수집 자체가 장래 부분 등록센서스 또는 완전 등록센서스의 시행을 최종목표로 하여, 공적 등록부(official register)를 도입하고, 그 등록부를 업데이트(update)하거나 기재오류(registration error)를 개선하는데 사용할 수 있다는 것이다. 그러나 여기서, 센서스의 실시시점에 통계

인 측면에서는 영토의 넓이와 경계를 표시하며 개인적인 측면에서는 해당 토지에 대한 소유권의 보호 및 재산의 가치를 표현하고, 더불어 그 변경 사항을 영속적으로 관리하는 국가의 토지 행정자료이다. 처음에는 과세를 위한 수단으로 출발하였지만, 현재에는 토지 소유권 및 가격 정보, 필지 관계 정보를 기록하고 보관·제공 및 운영을 목적으로 하는 다목적 토지대장(Multipurpose Cadastre)으로 발전하고 있다

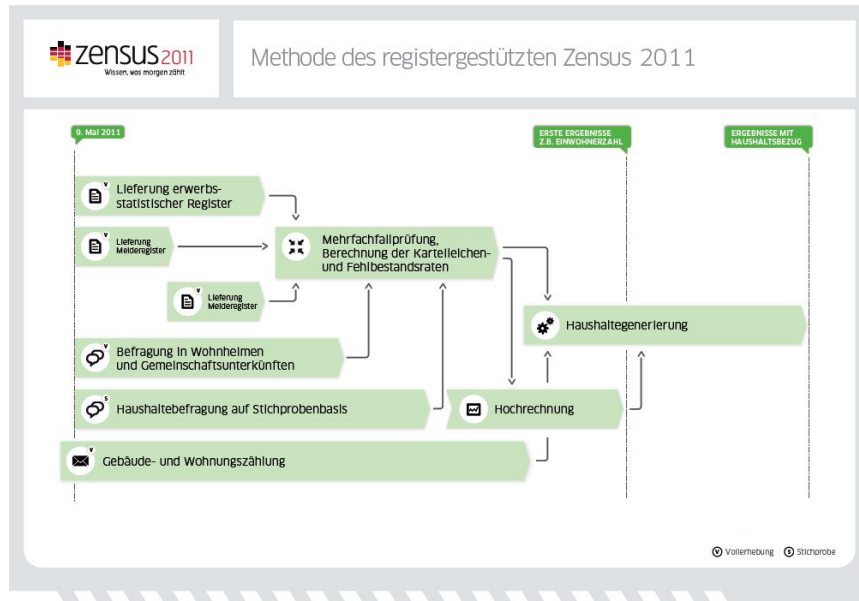
조사(전수집계=센서스)와 행정관리(등록부 업데이트) 간의 차이점을 응답자에게 명백하게 전달해야 하는 것이 무엇보다 중요하다. 또, 등록부의 행정자료 특히 주소대장(address register)을 표본추출의 프레임(sampling frame)으로 이용하기 때문에, 전통적 센서스보다 현지실사 운영이 더욱 더 효율적일 수 있다는 장점도 있다.

- 물론, 이러한 방식으로 등록부를 이용하는 전통적 센서스가 갖는 단점은 현지실사로 전수집계를 해야 하기 때문에 전통적 센서스와 같이 조사비용이 많이 들고, 응답자의 부담이 경감되기가 쉽지 않고, 등록부에서 추출된 주소대장의 완전성에 문제가 있는 관계로 지역프레임(areal frame)을 동시에 이용해야 하는 폐단이 있기 때문에 전국인구는 물론 지방자치단체의 인구나 여타 센서스 모집단(예: 인구, 가구, 주택)의 크기를 수리적으로 추정하는 방법론 자체도 현실적으로 적용하기가 간단하지 않다고 할 수 밖에 없다.

2. 등록부 자료(인구) + 전수집계 (주택) + 특수목적의 표본조사

- 우리나라의 경우, 2015년 등록센서스를 실시한다고 할 때, 주민등록, 가족관계 등록, 건축물/주택대장과 같은 핵심 등록부에서 행정자료의 포괄범위와 조사항목의 정확성에 아무런 문제가 없는 경우에는 이 등록부 자료를 이용하여 센서스 통계를 작성할 수 있다. 그러나 등록부들 중 어느 하나(예: 건축물/주택대장)가 포괄범위나 기재항목의 정확성에 있어서 문제가 있는 경우에는 관련 변수를 전수집계(complete enumeration)하고, 향후 그 결과를 이용하여 등록부를 업데이트하여 품질개선을 시도할 수 있다.
- 이러한 방식으로 센서스를 실시하는 나라에서는, 센서스 모집단 특히 인구부문에 대한 표본조사를 실시하여, 등록부의 포괄범위(과다범위, 과소범위)를 점검하고, 등록부에 없는 조사항목에 대한 자료를 별도로 수집하는 이유는 포괄범위 점검이나 조사항목의 통계표 작성에서 약간의 추출오차를 포함하여도, 국가통계인 센서스의 품질에 그리 큰 손상이 없다고 판단하기 때문이다.

<그림 6-3> 독일 2011년 등록센서스의 방법



주: 고용대장(Erwerbsstatistischer Register) 주민등록(Meldenregister)
중복, 이중기재(Mehrfachfallprüfung) 과다범위(Kartelleichenrate)
과소범위(Fehlbestandsrate) 추정(Hochrechnung)
표본 기반 가구조사(Haushaltenbefragung auf Stichprobenbasis)
가구생성(Haushaltegenerierung)

- 2010년 센서스 라운드에서 이러한 방식으로 센서스를 실시한 나라에는 독일과 스페인이다. 독일과 스페인의 2011년 등록센서스는 기존의 건축물대장이 포괄범위에 심각한 문제가 있었고, 기재내용의 정확성에도 문제가 있었기 때문에, 건축물/주택 센서스를 전통적 방식으로 실시하지 않으면 안 되었다.
- 독일은 2011년 등록센서스에서 주택부문에 대해서 전수집계를 실시하고 주택(주소)별로 거주자를 조사하여, 이들을 다시 주민등록과 결합하여 “가구생

성”(household generation)이라는, 대단히 특수한 방법론적 절차를 밟았다. 또, 독일 연방통계청은 사회경제통계 자료를 획득하기 위하여, 핵심등록부로 연방고용청 고용대장(Employment Register)을 이용하였다. 한편, 주민등록이나 고용대장에서 입수할 수 없는 센서스 항목에 대해서는 주소명부에서 10% 정도의 표본을 추출하여, 센서스 통계표를 작성하도록 되어 있다.

- 독일 연방통계청(Federal Statistical Office)은 2011년 센서스에서 주민등록을 주소명부/건축물대장을 구축하기 위한 일차적 자료원으로 사용하였다. 또, 주민등록은 표본조사를 실시하기 위한 프레임으로 사용되었으며, 인구통계, 지리통계 관련 센서스 항목을 포함하는데 사용되는 핵심자료로 활용되었다. 여기서, 독일의 2011년 센서스를 표본조사를 실시할 때, 등록부의 주소명부를 이용하였다는 점에서, 전통적인 센서스나 일부 등록센서스 실시국가가 보완적으로 실시하는 표본조사가 지역 프레임(areal frame), 특히 우리나라가 조사구를 설정하여 이른 기반으로 표본설계를 한 것과는 좋은 대조를 이룬다.
- 스페인은 2011년 등록센서스를 추진하면서, 글로벌 재정위기에 직면하여 센서스 비용 삭감을 목적으로, 인구부문은 주민등록 자료를 이용하고, 주택부문은 전수조사로 하고, 추가적으로 추출률 12.5%의 표본조사 자료를 획득하여, 이들을 결합하는 방식으로 전환했다. 스페인은 2011년 센서스에서 건축물/주택대장이 실제로 거주자가 있는 건축물/주택을 대상으로 하여 작성되었기 때문에, 주택부문에 대해서는 빈집(vacant estate)에 대한 추가 자료를 수집할 목적으로 전수집계를 하기로 결정하였다.
- 이탈리아는 2011년 센서스에서 등록부 자료 외에 센서스를 short/long form 조사표를 동시에 사용하여, 이들 자료를 결합하는 방식으로 센서스를 실시하면서, 2021년부터 순환 등록센서스(rolling register-based census)로 이행하기 위한 준비를 하였다.
- 이러한 준비과정에서, 2010년 센서스 라운드에서 이탈리아가 등록부로 자치단체 주민등록(Anagrafe)을 이용하기는 하지만, 현지실사를 전수집계와 표본집계로 구분하고, 표본조사에도 조사문항의 분량에 따라 두 종류로 구분하여, 이

탈리아의 2010년 라운드 센서스에서는 사실상 세 종류의 조사표(short-, middle-, and long-form)가 사용되었다.

<표 6-1> 등록부 자료(인구)+ 전수집계(주택)+ 특수목적의 표본조사
실시국가의 표본추출과 등록부 과다점검 방식

국가	실시연도	표본추출률	등록부 과다점검
독일	2011	9.6%(인구기준)	등록부 중복삭제로 과다점검, 사후조사(반복조사) 실시
스페인	2011	12.3%(인구기준) 11.9%(거처기준)	보조 행정자료 이용, recount factor 계산

- 이탈리아는 2020년 센서스 라운드에는 우리나라의 주민등록과 가족관계등록부를 합친 것에 해당하는 “Anagrafe”라 불리는 핵심 등록부를 주축정보로 이용하는 순환센서스로 전환할 계획이지만, 특수목적의 표본조사 자료를 자치단체 주민등록 자료와 결합하는 “통합센서스” 방식으로 이행한다고 볼 수 있다. 또, 스페인은 건축물/주택대장의 정확성이 확보된다면 2020년 센서스 라운드에서는 특수목적의 표본조사를 인구대장, 건축물/주택대장의 자료를 결합하는 등록센서스를 실시하게 될 것이다.
- 독일 통계청은 주민등록을 등록센서스의 백본으로 고용대장(Employment Register)을 같이 이용하고, 건축물/주택에 대하여 전수조사를 실시하고, 특수목적의 별도 표본조사를 실시하면서, 센서스 모집단 9.6%(인구기준)의 표본을 주민대장에서 선정하였으며, 등록부의 중복삭제로 과다범위를 점검하기로 하였다. 스페인은 2011년 센서스에서 인구기준 12.3%, 거처기준 11.9%의 표본을 인구대장과 현지실사에서 각각 선정하였으며, 등록부의 과다범위 점검을 위해서 행정자료를 보조정보를 이용하여 최종적으로 재검인자(recount factor)를 계산하였다 (<표 6-1> 참고).

3. 등록부 자료 + 특수목적의 표본조사

- 2000년 센서스 라운드에는 동아시아에서 싱가포르가 처음으로, 가구등록데이터베이스(Household Registration Database)와 주택/거처대장(Dwelling Register)을 핵심 등록부로 하고, 등록부에 포함되지 않은 센서스 항목을 조사하기 위한 특수목적의 표본조사를 실시하여, 등록부와 표본조사를 결합하는 방식으로 센서스를 실시하였다. 싱가포르는 인구와 주택에 대하여 등록부를 정비하는 작업을 1990년대부터 꾸준히 추진하였기 때문에, 등록부의 완전성 포괄범위에 대하여 전수조사를 실시할 필요는 없었고, 모집단의 일부에 대한 표본조사만 실시하여 이를 결합하여 사회경제통계를 작성하는 센서스 방식으로 전환하였다.
- 북유럽에도 완전 등록센서스를 실시하지 못했던 국가들, 가령 스웨덴, 노르웨이도 이러한 방식으로 센서스를 실시하였다. 그러나 북유럽의 부분 등록센서스는 행정자료를 거의 100% 이용하여 센서스 통계를 작성하기 때문에, 완전 등록센서스라고 해도 무방할 정도이다.
- 2010년 센서스 라운드에서 이러한 접근방식을 채택한 나라로는 북유럽을 제외하고, 유럽 지역에서는 이스라엘, 스위스, 터키가 있으며, 동아시아 지역에서는 싱가포르(Singapore), 대만(Taiwan-Republic of China)이 있다.
- 우리나라 통계청은 2010년 전통적 센서스를 마지막으로 하고, 2015년부터는 주민등록, 가족관계등록부, 건축물대장을 핵심 등록부 자료로 하여, 사실상 인구총조사, 가구총조사, 주택총조사를 동시에 실시하며, 특수목적의 표본조사로 이들 등록부의 행정자료를 보완하는 통합센서스 또는 결합센서스를 실시할 계획이다. 이러한 방식의 센서스 패러다임 전환은 조사자료와 행정자료의 매칭과 연계(matching and linking), 행정자료의 포괄범위 점검 및 기재내용의 오류점검, 그리고 센서스 모수의 추정방법론을 신규로 개발해야 한다는 점에서 결코 간단한 실험은 아니라고 할 수 있을 것이다.

- 등록부 자료와 표본조사 자료를 결합하는 센서스 방식으로, 전통적 센서스의 전수집계 항목인 short-form 조사표에 해당하는 것을 등록부의 행정자료로 대체하며, long-form 항목에 해당하는 것을 표본조사로 위하여, 노동력조사(LFS)와 같은 기존의 통계조사를 이용하는 것이 아니라 별도로 표본조사(*ad hoc survey*)를 실시하여, 그 조사에서 획득한 자료를 등록부 자료와 결합하여 센서스 통계를 작성한다.
- 특수 목적의 표본조사는 ① 등록부의 정확성, 특히 포괄범위와 기재내용을 점검하기 위하여, ② 등록부에서 가용하지 않은 조사항목에 대한 자료를 수집하기 위하여 실시될 수 있다. 물론, 이들 두 가지 목적의 표본조사는 분리될 수도 있고, 결합될 수도 있다. 가령, 이스라엘은 포괄범위와 조사항목 보완 목적의 표본조사를 동시에 실시하고, 영국은 이들 조사를 분리할 계획을 가지고 있는 것처럼 보인다. 등록부 자료와 특수목적의 표본조사 자료를 결합하는 센서스를 실시하기 위해서는 상이한 자료원에서 추출한 개체자료(individual data)를 연계할 수 있는 능력과 수리적 방법론의 개발을 필요로 한다.
- 특수목적의 표본조사 자료를 등록부 자료에 결합하는 센서스는 인구대장의 정확성을 점검하고, 인구집계를 현실성이 있도록 조정할 수 있으며, 표본조사를 실시하기는 하지만 등록부 자료를 이용한다는 점에서 응답자 부담이 상당히 경감될 수 있고, 조사비용도 상당 정도 절약될 수 있다는 장점이 있다.
- 2010년 라운드 센서스에서 특수목적의 표본조사는 대부분 센서스 모집단에서 20% 수준의 추출률로 실시되었지만, 12-13%대의 표본추출률로 표본조사를 실시했던 경우도 있다(<표 6-2> 참고). 싱가포르(Singapore)와 폴란드(Poland)의 표본조사는 추출률이 20% 수준이고, 이스라엘(Israel), 대만(Taiwan-Republic of China)의 표본조사가 추출률이 16-17%이며, 터키(Turkey)의 표본조사 추출률이 12-13%이다.
- 등록부에서 과다범위(OC: overcoverage)의 정도를 점검하기 위하여 표본조사를 실시하는 나라는 이스라엘(Israel)이 유일하며, 과소범위를 점검하는 표

본조사와 합하여 모두 20% 수준의 표본조사를 실시하였다. 폴란드(Poland)는 short-form의 형태로 등록부의 정확성을 점검하기, 등록부의 데이터베이스를 구축하고 이들의 정확성을 점검하기 위한 간단한 인터넷조사를 실시하였다. 한편, 싱가포르(Singapore)와 터키(Turkey)는 등록부의 정확성을 점검하는 품질관리를 별도로 하지 않고, 등록부를 이용하는 센서스를 실시하였다. 폴란드, 싱가포르, 대만, 터키는 모두 등록부의 과소범위를 점검하기 위한 조사를 사회경제 통계를 얻기 위한 표본조사와 동시에 실시하였다.

<표 6-2> 등록부자료 + 특수목적의 표본조사 실시국가의 표본추출과 등록부
과다점검 방식

국가	실시연도	표본추출률	등록부 과다점검
싱가포르	2000, 2010	20%(인구기준)	별도의 과다점검 없음
이스라엘	2008	16-17% (인구기준)	20% 과다점검 표본조사 실시
대만	2019	16-17% (인구기준)	없음
폴란드	2011	20% (인구기준)	Short form 등록부 확인 전 수조사(인터넷 방문) 실시
터키	2011	12-13%(인구기준)	별도의 과다점검 조사 실시하 지 않음

- 문제는 전통적 방식의 센서스 통계는 단순집계인데 반하여, 행정자료와 특수 목적의 표본조사의 결합으로 실시되는 결합 센서스 또는 통합센서스에서는 인구주택총조사의 모집단 특성을 인구부문, 가구부문, 주택부문에 대하여 별도로 표본자료에서 추정해야 하기 때문에 추정방법론 자체가 대단히 복잡할 수 있으며, 표본조사 자체에서 나온 정보도 지리적 상세성이 부족하고, 특히 전수조사가 아니기 때문에, 인구규모나 적은 자치단체나 하위 행정단위의 소지역 (small-area)에 대해서는, 아무리 세련된 간접추정(indirect estimation) 방법론을 적용한다고 하더라도 센서스 통계의 정확성에 의문이 제기될 수밖에 없다.

4. 등록부 자료 + 기존의 표본조사

- 등록부 자료와 “기존의 표본조사”(existing sample survey) 자료를 결합하는 센서스는 2000년 센서스 라운드에서 네덜란드에서 “가상센서스”(Virtual Census)라는 이름으로 실시되었고, 2010년 센서스 라운드에는 아이슬란드가 추가되었다. 위에서 벨기에는 2010년 센서스 라운드에서 완전 등록센서스 실시국가에 포함시켰지만, 벨기에 통계청에 의하면, 유럽연합 센서스 규정(EU Census Regulation)에 열거된 항목에 대한 통계표를 등록부 자료로 작성하는데 문제가 있을 때, 기존의 표본조사 자료를 활용할 계획으로 있었기 때문에, 네덜란드와 같은 종류의 가상센서스 실시국가로 분류할 수도 있을 것이다.
- 센서스 조사항목 중 일부가 등록부에서 입수 불가능한 경우에, 기존 표본조사(예: 경제활동인구조사, 사회통계조사, 생활상태조사 등)의 자료를 이용하는 방식으로, 자료는 개체수준에서 연계될 수 있으며, 전통적 센서스와 같은 현지 실사에 의한 자료수집을 하지 않는 것이 특징이다.
- 네덜란드의 가상센서스는 사회통계 데이터베이스(SSD: Social Science Database)를 자료원으로 사용한다. 이 데이터베이스는 다양한 등록부의 행정 자료와 표본조사를 마이크로 수준에서 연계(micro-linking)하여 구축되었으며, 사람, 가구, 일자리, 사회급부 등에 대한 정보를 포함한다. 센서스가 마땅히 추정해야 할 총인구나 특성별 인구의 수치는 표본조사와 등록부의 총계(total)를 반복가중(repeated weighting)과 캘리브레이션 절차(calibration procedure)를 통하여 확보하도록 되어 있다. 여기서 반복가중은 개개의 표본조사에 대하여 한 세트 이상의 표본가중치를 사용하는 것이 특징이다. 개별 통계표를 추정하기 위하여, 가중치는 해당 통계표의 항목소계가 이미 추정이 완료된 통계표의 항목 소계와 일치하도록 조정한다(Houbiers et al. 2003).
- 센서스 방식으로, 등록부 자료와 기존의 표본조사를 결합하는 센서스는 응답자의 부담을 경감하고, 현장실사를 실시하지 않기 때문에 조사비용이 경감될

수 있고, 센서스 결과와 여타 조사 결과의 정합성(consistency)이 크다는 장점이 있다. 그러나 이러한 센서스 방식의 단점으로는 서로 다른 등록부와 표본조사의 개체정보를 성공적으로 매칭/연계(matching and linking)하는 과정이 대단히 복잡하다는 것이고, 표본조사에서 추출된 정보도 표본조사의 한계를 극복할 수는 없으며, 아무리 정교한 통계적 추정방법론이 도입된다고 하더라도 이해당사자들의 수요를 정확하게 조준할 수 있는 상세한 센서스 통계를 작성하는데 불충분할 수 있다는 것이다.

5. 완전 등록센서스

- 완전 등록센서스는 다양한 등록부의 자료를 통합함으로써 실시되는 센서스이며, 장기간에 순차적인 과정을 밟아서 완결된 방식이다(등록부의 도입과정과 최초 이용연도에 대해서는 <표 6-3>을 참고할 것). 이 경우 인구주택총조사는 “등록정보시스템”(register information system)의 일부로 볼 수 있다(Wallgren and Wallgren, 2007).
- 완전 등록센서스를 실시하기 위한 요건으로는 (1) 주민등록부, 건축물/주택대장의 정비 및 가용성(availability) (2) 가구/가족을 확인하여, 인구와 주택을 연계할 수 있는 방법 (3) 주민등록부의 통계적 이용에 대한 여론과 법률적 지원 등을 포함한다. 그러나 이러한 완전 등록센서스의 실시는 이론적으로는 간단할 수 있지만, 현실적으로 대단히 어려운 점이 많기 때문에, 오랜 시간에 걸쳐 점진적으로 성취될 수 있는 과정이라는 것을 명심할 필요가 있을 것이다.
- 완전 등록센서스는 2000년 센서스 라운드에서 안도라, 덴마크, 핀란드에서 실시되었으며, 2010년 센서스 라운드에서는 상기의 3개 나라 외에, 오스트리아, 벨기에, 노르웨이, 슬로베니아, 스웨덴 등 5개국이 추가되었다. 완전 등록센서스는 대부분 실시국가의 인구규모가 작다는 특징이 있지만, 향후 완전 등록센서스의 실시는 인구규모와는 달리 등록부의 정비 및 가용성에 의하여 중대한 영향을 받게 될 것이다.

- 아마도, 2021년 센서스 라운드에서는 포르투갈이 완전 등록센서스 국가에 추가될 것이다. 포르투갈의 센서스 패러다임 변경 방법은 전통센서스에서 완전 등록센서스로의 이행이라는 점에서 대단히 급격한 방법론 전환이라고 볼 수 있다. 그러나 포르투갈 통계청은 완전 등록센서스를 실시하지 못하는 경우에는 표본조사를 실시하여 등록부의 기재내용이 불충분한 조사항목을 보완할 계획을 세우고 있다.

<표 6-3> 북유럽 국가의 부문별 등록부 도입 연도 및 센서스 최초 이용연도

	Denmark		Finland		Norway		Sweden	
Type of register	Established	First used in census	Established	First used in census	Established	First used in census	Established	First used in Census
Central Population Register	1968	1981	1969	1970	1964	1970	1967	1975
Business Register	1975	1981	1975	1980	1965	1980	1963	1975
Dwellings	1977	1981	1980	1985	2001	2011	2008?	2011?
Housing conditions	1977	1981	1980	1985	2001	2011	2008?	2011?
Education	1971	1981	1970	1975	1970	1980	1985	1990
Employment	1979	1981	1987	1990	1978	2001	1985	1985
Family	1968	1981	1978	1980	1964	1980	1960	1975
Household ^a	1968	1981	1970	1975	2001	2011	2011?	2011?
Income	1970	1981	1969	1970	1967	1980	1968	1975
Totally register-based census		1981		1990		2011		2011?

^a Household-dwelling unit, i.e. all the persons living in one dwelling

- 행정자료 또는 행정자료가 등록센서스에서 유용하게 사용되기 위해서는 다음과 같은 요건을 구비하여야 한다.
 - (1) 자료의 적절한 포괄성. 주제별 포괄성에 문제점이 있는 경우에는, 전통적 방법을 이용하여 보완적인 자료수집을 하도록 해야 한다.

(2) 인구주택총조사의 센서스 통계를 작성하는데 필요한 3개의 기본단위는 ① 사람 ② 기업/사업체 ③ 거처(주택)이다.

(3) 신뢰할 수 있는 자료 갱신주기

(4) 개념의 안정성이 중요하며, 정의가 바뀌면 문제가 생긴다.

(5) 자료 품질의 요건을 충족해야 한다.

(6) 통계작성기관인 통계청은 등록부 관리 기관과 충분히 협력해야 한다.

물론, 이러한 요건들은 부분 등록센서스를 실시하는 우리나라에도 그대로 적용되어야 할 것이다.

- 완전 등록센서스의 장점에는 (1) 응답자 부담의 경감 (2) 센서스 자료의 매년 이용가능성 (3) 고품질의 등록부가 도입되고 나면, 전수집계에 의한 전통적 센서스 실시보다 저비용으로 센서스 통계의 작성이 가능함 등을 포함한다. 그러나 단점으로는 (1) 등록부의 도입과 유지관리에 많은 자원이 요구되고 (2) 센서스 항목에 포함될 수 있는 것이 등록부에 포함된 변수항목에만 국한되고, (3) 센서스 통계의 품질이 궁극적으로 등록부의 포괄범위와 품질에 종속되며 (4) 등록부의 자료는 통상적으로, 비통계적, 행정관리를 목적으로 수집되며, (5) 통계작성기관이 등록부 관리기관에 종속된다는 것이다.

제2절 주요국의 등록센서스 추진 및 실시경과

1. 독일

- 독일은 1875년부터 1910년까지는 5년마다 전통적 센서스(Volkszählung)를 실시하였으며, 전후 서독지역은 1970년까지 5-9년의 간격으로 몇 차례의 인구센서스를 실시하였다. 1956년의 건축물/주택센서스에서는 거주인구를 마이크로센서스로 집계하기 시작하였다. 한편, 통일이전의 동독지역에서는 1949-1990년의 50여 년간 4차례의 센서스가 실시되었으며, 1964년에 실시된 센서스 결과만이 공표되었다. 동독지역의 마지막 센서스는 1981년 12월 31일 기준일로 하여 실시되었다.
- 독일의 등록센서스 도입과정에서, 센서스 자체가 “정치화”되었다는 사실을 무시할 수 없다. 1981년 5월 28년 당초에 계획되었던 인구센서스는 연방정부와 지방정부의 예산협력 문제가 해결되지 않았기 때문에, 1983년 4월 27일로 연기되었다. 그러나 이러한 당초의 센서스 실시계획은 독일연방헌법재판소(Bundesverfassungsgericht)의 센서스 위헌판결로 1987년 5월 27일로 재차 연기되었다.
- 독일은 1987년의 마지막 센서스 이후 1990년과 2000년 라운드의 센서스를 실시하지 못하고, 통계법상의 여러 제약 때문에, 전통센서스를 폐기하고 행정자료를 이용하는 센서스로 이행하기로 결정하였다. 1998년 행정자료 옵션을 이용하는 몇 가지 새로운 센서스 모델이 제시되었으며, 2001년부터 2003년까지 1% 표본추출에 의한 마이크로센서스 실시를 통해서 최종모델이 선택되었고, 그 후에 엄격한 독일 연방통계법⁵⁾(BStatG: Federal Statistics Act of 1987)

5) 독일의 연방통계법은 독일 국가통계에 관련된 기본법으로서, 품질개선, 중립성, 과학적 독립성, 통계청의 업무, 통계적 비밀보호에 관한 규정을 규정한다. “법 규정이 없는 통계조사는 실시할 수 없다”(no survey without legal basis)라는 원칙에 입각하여, 연방통계의 하나하나가 모두 개별법이나 규정에 의하여 실시되도록 하고 있다. 특히, 개별

의 요구사항과 행정자료의 이용 가능성 때문에 필요한 부분이 수정이 가해졌다.

<표 6-4> 독일의 2011년 등록센서스 추진 및 실시 스케줄

기획 및 준비	시점
1981년 센서스 1983년으로 연기	1981년
1983년 센서스 위헌판결	1983년 12월
1987년 최후의 전통센서스 실시	1987년
새로운 센서스 모델 제안	1998년
새로운 센서스 예비조사 관련 법률 제정	2001년 7월
2011년 센서스 모델 잠정 확정	2001년 9월 -2003년 9월 (마이크로 센서스 실시시점)
대연정 CDU/SPD 2010년 라운드 EU센서스 참가 결정	2006년 11월 11일
2011년 센서스 준비법 효력발생	2007년 12월 13일
2011년 센서스법 제정	2008년 12월 3일 제정/2009년 7월 8일 공포
등록센서스 표본설계 및 추정방법론 개발	2008년-2011년
건축물/주택센서스 우편조사 실시	2011년 4월말-2012년 7월(14개월)
센서스 표본조사 실시	2011년 5월 9일
센서스 반복조사(사후조사) 실시	표본조사 종료 후 수주일 내 실시
연방/주/자치단체 인구규모의 확정	2013년 상반기
센서스 상세분석 결과 공표	2014년 상반기

- 2005년 11월 11일, 독일 대연정 연립파트너인 기민당(Christian Democratic Union)과 사민당(Social Democratic Party)는 2010년 라운드의 유럽연합 센서스에 참가하기로 결정하고, 2006년 8월 29일 정책결정에서, 연방내각은 독일의 2011년 센서스를 등록센서스 방식으로 실시하기로 결정하였다. 유엔 통계처(UNSD)가 주관하는 인구주택총조사 세미나(Seminar on Population and Housing Censuses)에서, 독일 연방통계청은“독일은 2010/2011년 라운드

연방통계조사에 대한 법규는 조사의 특징, 조사의 유형, 기준기간, 기준일자, 주기성, 응답대상자를 철저히 정하고 있다.

범 유럽 센서스에 참가하고, 센서스 방식으로 등록센서스로 하겠다.”고 공식 발표하였다.

- 2007년 12월 13일, 센서스 준비법(Zensusvorbereitungsgesetz 2011 – ZensVorbG 2011)이 효력을 발생했다. 센서스 준비법은 독일의 2011년 센서스의 구체적 준비를 위한 법적 기반을 제공하였다. 2008년 12월 3일, 독일 연방의회는 2011년 센서스법(Zensusgesetz 2011 – ZensG 2011)을 통과시켜, 독일의 2011년 센서스 시행을 관리하도록 하였다.
- 독일 센서스 준비법은 1980년대에 전개되었던 센서스 위헌소송과 연방헌법재판소의 위헌판결을 염두에 두고, 적용영역, 주소/건축물대장과 출생지/출신국 대장의 구조, 건축물/주택 센서스의 준비를 위한 자료전송 시스템, 공통규정 및 작업준비 절차 등을 상세하게 규정하고 있다.
- 독일의 2011년 센서스법은 6개 부분으로 구성되며, (1) 일반방침 (2) 자료의 수집과 통합(가구생성), (3) 센서스 실시 조직 (4) 센서스 결과의 품질을 확보하기 위한 조치 (5) 정보제공 의무와 정보보호 (6) 최종규정으로 이루어져 있다.
- 독일 연방통계청의 공식발표에 의하면, 독일에서 2011년 센서스의 실시가 필요한 이유는 구서독의 1987년 최종 센서스와 구동독의 1981년 최종센서스 이후, 그 자료를 인구분석(DA: Demographic Analysis) 곧 인구방정식(Demographic Equation)으로 센서스 실시 마지막 시점 이후, 인구동태와 인구이동의 수치를 보완/조정한 수치만을 사용하여 왔기 때문이다. 이 때문에, 독일 연방통계청은 센서스에 의한 인구 규모는 인구동태와 국제인구이동 자료를 이용하여 인구추계를 한 것보다 상당히 늘어날 것이라고 가정하였다.
- 독일 센서스위원회(Census Commission)⁶⁾ 위원장은 2011년 센서스 준비법을

6) 독일 센서스위원회는 9명의 위원으로 구성되며, 연방내무장관이 지명한다. 위원은 사회학, 통계학, 인구학, 경제학, 정치학 등 센서스 자료의 과학적 이용과 관련 학술분야를 대표하는 학자들로 구성된다.

제정하기 위한 연방의회 청문회에서 “2011년 센서스는 엄격히 말하자면, 조사대상은 독일 시민이 아니라 그들에 대한 자료를 포함한 익명화된 컴퓨터 기록(anonymized computer records)이다.”라고 했다. 행정자료로는 주민대장(Population Register)과 연방고용등록부(Federal Employment Register)이고, 등록부 자료로서 당시 문제가 있다고 판단했던 건축물과 주택에 관한 자료는 건축물/주택 소유주(또는 관리자)로부터 우편조사(postal survey)로 수집하기로 결정하였다.

- 독일 연방통계청은 인구센서스와 함께 주택센서스의 실시가 필수적이라고 판단하고, 인구부문은 등록센서스로 하되, 주택/거처부문은 전수집계를 하여, 주택대장을 구축하기로 하였다. 독일에는 현재 어떤 기초자치단체에 거처가 많이 비어 있으며, 주택부족이 어디에서 얼마나 심각하며, 과열화되어 있는가를 알 수 있는 통계가 존재하지 않는다. 이 때문에, 연방통계청은 2011년 센서스의 일부로서 주택공간이 있는 모든 공간에 대한 조사를 실시하는 것이 필수적이라고 판단하여 왔다. 독일 국민이 어떻게 보호되고 있는가에 대한 신뢰성 있는 통계적 정보는 주택/거처 정보를 통해서만 확보될 수 있다.
- 건축물/주택 센서스는 2011년 4월 말부터 2012년 7월까지 장장 14개월에 걸쳐서, 주택과 아파트의 보유자와 관리자 1,750만 명에게 우편 조사표를 배포하여 해당 문항을 대한 응답을 획득하였다. 조사표는 건물유형, 건축연도, 거처총수, 난방형태 등의 질문을 포함하며, 단독주택에 대해서는 건평, 필요방수, 거주자 수 등에 대한 정보를 수집한다. 건축물/주택 센서스의 결과는 연방/주/자치단체 정부의 주택정책과 도시계획정책에 대한 결정을 내리기 위한 필요정보를 제공한다.

(표본조사)

- 독일 2011년 등록센서스의 표본조사 기준일자는 2011년 5월 9일이다. 표본조사의 조사표에는 “가구표본은 모집단에 대한 자료품질을 개선으로 관련정보를 추가적으로 획득하기 위한 것이며, 거쳐주소가 있는 개인에 대하여 실시되는 표본조사 목적에 사용된다. 표본 크기는 인구, 가구, 주택을 연계하는 주소명부 모집단의 9.6%이며, 조사 기준일은 2011년 5월 9일이다. 조사는 연방통계법에 의거하여, 2009년 7월 8일에 공표된 2011년 센서스법에 따라 실시된다.”고 하고 있다. 결국, 표본조사는 조사목적, 기준일자, 법적근거를 명시하고 있다고 볼 수 있다.
- 먼저, 가구 표본조사(Household Sample Survey)는 2011년 센서스의 구성요소로, 총인구의 대략 10%에 해당하는 응답자 790만 명을 대상으로 면접을 진행한다⁷⁾. 가구조사는 1만 명 이상의 거주자가 있는 지방자치단체의 인구규모를 결정하는데 사용한다. 행정자료는 인구이동이 빈번하게 일어나는 대도시 공동체의 경우는 인구규모를 확정하는데 어려움이 많으며, 행정자료가 인구규모를 확정하는 유일한 자료원이 되어서는 안 된다는 것이 독일 통계청의 판단이다. 이 때문에, 유럽연합 센서스 규정이 정한 대로, 가구조사 결과를 이용하여, 등록부의 과다범위와 과소범위를 통계적으로 보완하는 수치를 발표하도록 되어 있다.

7) 독일의 2011년 센서스 표본조사는 지방자치단체를 인구크기에 따라 (1) 2000명 미만 (2) 2000-10,000명 (3) 10,000-50,000명 (4) 50,000 명으로 구분하여, 추출되는 주소지의 개수(address size)를 차등화하고, 같은 자치단체 안에서도 행정자료의 포괄범위 오차(과다범위, 과소범위)는 전입자 유무와 가구구성에 따라 달라지기 때문에, 이들 주소지의 개수를 다시 5개의 층(strata)(층 1: 해당 주소지에 1년 전 전입자 유무, 성씨가 다른 사람의 수 1-2명, 층 2: 해당 주소지에 1년 내 전입자수 유무, 성씨가 다른 사람의 수 1-2명, 층 3: 해당 주소지에 1년 전 전입자수 유무, 성씨가 다른 사람의 수 3-12명, 층 4: 해당 주소지에 1년 내 전입자 유무, 성씨가 다른 사람 3-12명, 층 5: 성씨가 다른 사람 13명 이상)으로 층화(stratification)하였다. 이들 층 안에서 분산이 크기 때문에, 니이만 할당법(Neyman allocation)을 사용하여 성씨가 다른 사람이 많은 주소지나 전입자의 수가 상대적으로 많은 주소지를 많이 추출하는 방안을 채택하였다(Ralf Münnich, Siegfried Gabler u.a. 2012, Stichprobenoptimierung und Schätzung im Zensus 2011).

- 주소명부 모집단의 10%를 표본으로 선정하여 실시되는 가구 표본조사는 독일의 등록부 자료에 전혀 포함되어 있지 않거나 불충분한 자료들에 대하여 질문을 포함한다. 여기에는 교육/전문자격(최종학력, 재학여부 특히 경제활동/고용 상태, 자영업, 종교, 이민경험 등에 대한 조사항목이 들어 있다.
- 응답자는 수리적/통계적 확률법칙에 근거하여, 사전에 선정된 것이며, 2011년 5월 9일에 선정된 응답자를 대상으로 조사원이 가구방문을 통하여 응답자에게 보고의무를 부과한다. 표본조사를 모집단의 일부에 대해서 실시하기 때문에, 그 결과는 기초 자치단체는 독일 전체로 조사결과를 확대하기 위한 승수 요인(乘數要因)을 적용한다.
- 이와 관련된 표본설계 적정화와 추정방식은 연방통계청의 독자적인 연구에 의하여 세부화되었으며, 그 보고서는 2012년 8월에 간행되었다(Ralf Münnich, Siegfried Gabler u.a. 2012, Stichprobenoptimierung und Schätzung im Zensus 2011). 이 보고서는 독일의 2011년 등록센서스의 목표를 두 가지로 설정하였다. 하나는 전국수준과 지역수준의 인구규모를 결정하고, 다른 하나는 등록부에 포함되어 있지 않는 변수를 신뢰도 또는 통계적 정도(精度: precision) 수준에서 그리 큰 문제가 없는 방식으로 센서스 통계표를 작성하는 것이었다.
- 독일은 2011년 센서스에서 사후조사, 곧 반복조사를 실시하여 추정인구의 통계적 정도를 개선하고자 노력하였다. 가구조사 주소의 5%에 대해서, 1차 조사가 끝난 후 몇 주가 지나서, 응답자들은 제2차 면접의 대상이 된다. 이 반복조사는 1차 조사와는 다른 면접원에 의하여 실시되며, 간단한 조사표로 몇 가지 특성에 대한 질문으로 하도록 하였다. 이것은 인구 1만 명 이상의 기초 자치단체에 대하여 실시된다. 이러한 반복조사 곧 사후조사(PES)는 국제적으로 보편화되어 있는 절차로서, 인구규모의 확정에서 가구조사의 품질을 평가하는데 도움을 주기 때문에 실시되었다.

(결과 공표)

- 독일 통계청은 연방/주/자치단체 수준의 인구는 2013년 상반기 중 확정되어 공표될 수 있을 것으로 보고 있다. 또, 2014년 상반기에는 상세한 분석결과가 공표될 것이며, 인터넷에 제시될 수 있을 것이다. 전반적으로, 독일의 등록센서스는 1987년 센서스 위헌판결 이후에 대단히 조심스럽게 진행되어 왔으며, 개인식별번호(인구대장) 없이 자료연계 작업이 실시되기 때문에 시간이 많이 소요되는 절차라고 보면 될 것이다.
- 현재, 독일의 행정체제가 연방시스템이고, 연방주와 기초자치단체들이 복잡한 행정체제를 이루어지기 때문에, 등록센서스는 추정방법론이 직접추정은 물론 다양한 간접추정, 특히 모형기반 추정방법(model-based estimation)이 가능하도록 다양하게 연구되고 있고 있지만, 가구생성(household generation) 절차가 대단히 복잡하기 때문에, 결과공표가 정해진 일정표에 따라 진행될 것인지에 대하여 여전히 불확실성의 요인은 있다고 할 수 있다.

2. 스위스

- 스위스 최초의 센서스는 1850년 3월 상원의원 스테파노 프란스치니(Stefano Franscini)의 주도하에 실시되었다. 인구수에 대한 조사 외에, 성별, 연령, 혼인상태, 직업, 경제활동, 종교 등에 대한 질문을 포함하였다. 1860년부터 2000년까지 150년간에 걸쳐서, 10년마다 센서스를 실시하였다. 이 센서스 기간 중 주기성 원칙을 위반한 센서스는 선거법 개정으로 1888년 센서스, 1940년 군동원령에 의한 1941년 센서스 등 2 차례가 있었다.
- 스위스 통계청((BFS: Bundesamt für Statistik)은 2000년에 마지막으로 전통적 센서스를 실시하였다. 스위스 통계청은 2010년 이후에는 새로운 형태의 센서스를 실시하도록 되어 있다. 인구조사에 대해서는 우선적으로 주민등록 정보를 이용하고, 그것을 표본조사의 조사항목으로 보완하는 절차를 밟게 되어 있다. 전체 국민의 5%를 대상으로 방문면접 또는 전화조사로 조사를 하게

되며, 2010년 12월 31일은 등록센서스의 최초 시행 기준일로 되어 있었다.

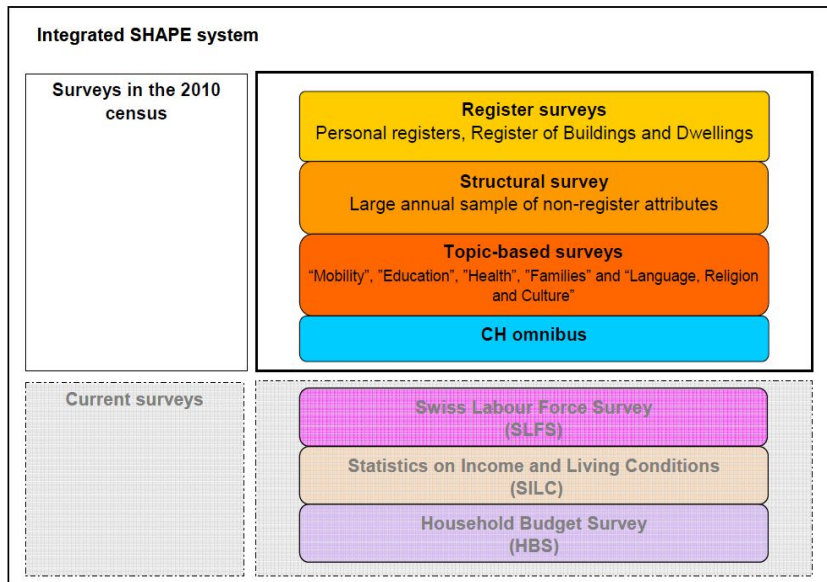
- 스위스 의회는 2006년에 내각의 요청으로, 센서스법을 심의하였다. 당시, 스위스 의회는 국가통계는 인구, 경제, 사회, 교육, 연구, 공간과 환경 등의 현황과 동향에 대하여 모집단의 특성을 대표하는 결과를 작성하여야 하며, 급변하는 사회경제적 정세에 비추어서, 주제의 폭 또는 다양성이나 정보의 시의성이 지리공간의 깊이에 못지않게 중요하다는 사실을 강조했다. 스위스 의회는 통계는 이용자의 요구에 부응해야 하며, 이제는 필요한 정보가 개별조사의 형태가 아니라 종합적이고 통합적인 자료수집, 자료평가, 분석시스템 등에 의하여 적정화되어야 한다고 보면서, 인구주택센서스는 국가통계영역의 상당한 영역에 대하여 필요한 정보를 제공해야 한다고 보고 있다.
- 스위스에서는 센서스 자료의 필요성은 헌법이나 각종 법률과 시행령은 물론 행정업무를 수행하는 연방정부, 지방자치단체 정부의 다양한 수요 때문이다. 또, 정책분야, 학술연구, 교육, 경제와 사회, 매스미디어, 일반대중들로 센서스 자료의 이용을 통하여 많은 이익을 얻고 있다.
- 스위스 의회는 내각은 통계작성의 시의성, 주제영역, 응답자 부담경감 등의 새로운 요구사항을 들면서, 자료수집 방법의 전면적 변화를 요구하였는데, 이것은 표준화된 등록대장의 이용과 보완적인 표본조사를 기반으로 한 것이었다. 그 후, 스위스 의회는 새로운 센서스 방식에 대해서, 광역자치단체와 이해당사자들을 대상으로 공청회를 개최하였다.
- 특히, 광역자치단체는 등록부 조사 외에, 등록부에 없는 센서스 조사항목을 조사하기 위하여, 인구부문 전수조사의 유지를 옹호하고 나섰다. 물론, 이 전수조사는 10년에 한 번씩 실시된다. 무엇보다도, 경제계는 내각의 새로운 센서스 방식을 옹호했는데, 그 이유는 이 방식은 많은 주제에 대하여 조사예산과 응답부담의 경감을 가져올 것이기 때문이었다.
- 스위스 내각은 2006년 4월 26일 공청회에서 새로운 센서스 방식에 대하여 국민의 합의를 유도하는데 실패하였다. 이 때문에, 당초의 센서스 개혁안에, 매

년 구조조사(Structural Survey)를 실시한다는 조건을 추가하여, 광역자치단체와 여타 정부부처의 요구를 수용하기로 결정하였다.

- 앞에서도 언급하였듯이, 센서스 통계를 얻기 위한 구조조사는 전통적 센서스의 조사항목 중에서, 주민등록대장, 건축물대장 등과 같은 행정자료에서 입수할 수 없는 센서스 기본항목을 대규모 표본에 대한 현지실사를 실시하는 것을 의미한다.
- 스위스 내각은 구미선진국의 센서스 경험을 통하여, 주민등록이나 건축물대장의 특성 중에서 센서스의 기본적 특성에 해당하는 것은 소지역 통계는 물론 소집단 통계에 이르기까지 표본조사로 생산이 가능할 것이라고 판단하였다.
- 추가적으로, 2006년 4월 24일 스위스 의회에 보내는 법률 권고안에 스위스 정부의 센서스 개혁안 외에 광역자치단체의 안을 첨부하였다. 스위스 내각의 새로운 센서스 개혁안은 “인구 및 가구를 종합하는 통계시스템에 등록부 조사와 표본조사를 통합하는 것”이라고 할 수 있다. 스위스 내각이 구상하고 있는 통계조사는 등록부조사, 구조조사, 주제조사를 포함하였다.
- 스위스의 새로운 센서스는 주로 등록부조사(Register Survey: Registererhebung)을 기반으로 하며, 이 조사에 기존의 행정자료를 이용한다. 스위스 통계청은 행정자료로, 광역자치단체와 기초자치단체의 주민등록(Einwohnerregister), 연방인구대장(Bundespersonenregister), 그리고 연방 건축물/주택대장(Gebäude- und Wohnungsregister)을 이용한다. 행정자료에 포함되지 않은 조사항목에 대해서는 보완적인 표본조사, 특히 구조조사(Structural Survey), 주제조사(Thematic Survey), 옴니버스조사(Omnibus Survey)를 통하여 수집하도록 되어 있다(<그림 6-4>을 참고할 것).
- 등록부조사(Register Survey: Registererhebung)은 총인구에 대한 기초정보를 제공한다(<그림 7-1> 참고). 이 조사는 인구와 가구의 구조, 거주와 주택의 특성에 대한 결과를 제공한다. 조사결과는 매년 8월 말 인구통계(population statistics), 건축물/주택통계(building/housing statistics)로 공표하고, 매년 초

에는 가구통계(household statistics), 주택상태(housing conditions)의 특성으로 공표하도록 되어 있다.

<그림 6-4> 스위스의 2010년 등록센서스 관련 통계시스템



- 구조조사(Structural Survey: Strukturerhebung)는 매년 20만 명을 표본으로 하여, 종이 조사표 조사를 실시한다. 광역자치단체와 대도시에 대해서는 표본을 확대할 수 있도록 되어 있다. 이 조사는 인구구조에 대한 추가적인 통계를 작성하며, 이를 통하여 등록부의 정보를 보완한다. 최초의 조사결과는 표본조사 실시 1년 후에 공표한다. 2010-2012년 기간에 대한 누적조사 결과는 2013년 말에 공표한다.
- 주제조사(Thematic Survey: Thematische Erhebungen)의 주제는 5개로, 1만 명에서 4만 명의 범위에서 표본을 선정하여, 조사표로 설문을 하도록 되어 있다. 이 조사의 통계로, 구조조사의 정보를 심화하는 것이 주목적이다. 최초

의 결과는 주제조사가 실시되고 1년 후 공표한다. 5개 주제조사의 실시시점을 보면, 교육 및 훈련 부문은 2011년, 건강 부문은 2012년, 가족/가구 부문은 2013년, 언어, 종교, 문화 부문은 2014년, 이동과 수송 부문은 2015년, 2016년에는 다시 교육 및 훈련 부문의 조사를 실시한다.

- 옴니버스 조사(Omnibus)는 특정의 정책이나 과학적 영역의 구체적인 주제에 대하여 3,000명 정도를 표본으로 하여 진행되는 조사로, 이 조사는 실시 이후 6개월 이후에 공표한다. 옴니버스 조사의 특징은 전화 표본조사이고, 매년 최소 1회 이상의 조사를 실시하고, 국가적 필요성에 따라 신축적으로 주제를 선정하며, 전국수준의 특정 주제에 대하여 정보를 획득하는 것이다. 가령, 2010년에는 인터넷이용, 2011년에는 생활에 관한 복수의 주제에 관한 조사를 실시하였다.
- 스위스 내각은 자신들 제안한 센서스 개혁안이 광역자치단체의 대안보다 총비용면에서 저렴하다는 결론에 도달하였다. 스위스 내각은 2008-2015년의 8년간에 걸쳐 6,920만 스위스 프랑(약 800억 원, 공정환율 1 Swiss Franc = 1,160원)에 해당한다.
- 현재 스위스 인구가 2010년 총인구 추정규모가 780만으로 우리나라 총인구는 스위스의 6배를 넘기 때문에, 단순계산을 하더라도 한화 4,800억 원에 해당하는 금액이 센서스 개혁안을 추진하는데 투입되는 비용이라고 할 수 있다. 사실상, 스위스의 등록센서스 실시비용은 우리나라에서 전통적 센서스로 실시되는 인구주택총조사 실시비용의 2.5배에 해당하는 비용으로, 우리나라가 등록 방식으로 센서스 방법론을 전환할 때, 고려해야 할 사항이다.
- 스위스 광역자치단체인 칸톤의 센서스 개혁안을 보면, 같은 기간에 전통적 센서스의 조사항목에는 들어있으나 등록부에는 포함되어 있지 않는 조사항목에 대해서 전수조사를 실시하여야 한다고 주장하였기 때문에, 표본 확대의 비용이 엄청나다는 것을 의미한다. 같은 기간에 광역자치단체의 센서스 개혁안은 총 소요예산이 1억 371만 스위스 프랑으로 연방정부에 9,960만 스위스 프랑, 광역자치단체에 3,750만 스위스 프랑이 할당되어야 할 것이다. 스위스 내각은

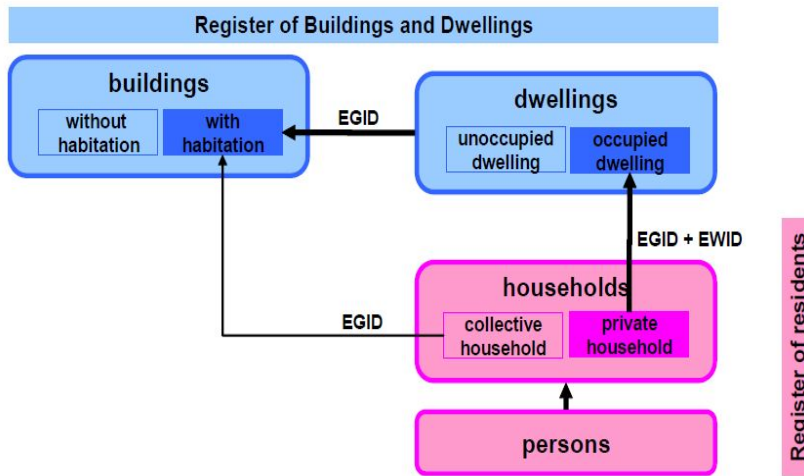
새로운 센서스법을 채택하면서, 2008-2015년 센서스 비용조달에 관한 연방시행령의 채택을 요구하였는데, 순수 센서스 관련 예산 요구액은 6,920만 스위스 프랑이었으며, 센서스 실시와 관련하여 외부전문가 기용에 소요되는 비용으로 2,150만 스위스 프랑을 요구하였다.

- 스위스 연방통계청은 매년 실시되는 스위스 노동력조사(SLFS, Swiss Labor Force Survey)와 가계조사(HBS, Household Budget Survey) 외에, 일련의 통계조사를 5년 주기로 실시하게 된다. 통계조사는 순환방식으로 진행되는 조사를 2010년의 여행활동 마이크로센서스(Microcensus on Travel Behavior)를 기점으로, 2011년에는 스위스 성인교육조사(Swiss Adult Education Survey), 2012년에는 스위스 건강조사(Swiss Health Survey), 2013년에는 가족 및 세대조사(Family and Generations Survey), 2014년에는 언어, 종교, 문화에 관한 조사(Survey on Language, Religion, and Culture)를 실시할 예정으로 되어 있다. 스위스 연방통계청은 또한 매년 소득 및 생활여건에 관한 조사(SILC: Survey on Income and Living Conditions)를 실시한다.

(주택센서스와 인구센서스의 관련성)

- 스위스 통계청은 2010년 센서스는 유엔 유럽경제위원회(UNECE, United Nations Economic Commission for Europe)와 유럽연합 통계청(Eurostat: Statistical Office of the European Union)의 공동 건의사항을 바탕으로, 주택총조사를 인구총조사에 통합하여 실시하기로 결정하였다.
- 인구총조사는 개인, 가구, 주택, 거처 등의 다양한 통계적 실체(statistical entities)에 대한 자료를 생산하고, 여기서 다양한 통계단위들은 건물식별번호(EGID)와 주택식별번호(EWID)를 이용하여 상호대조할 수 있도록 되어 있다(그림 6-5> 참고).

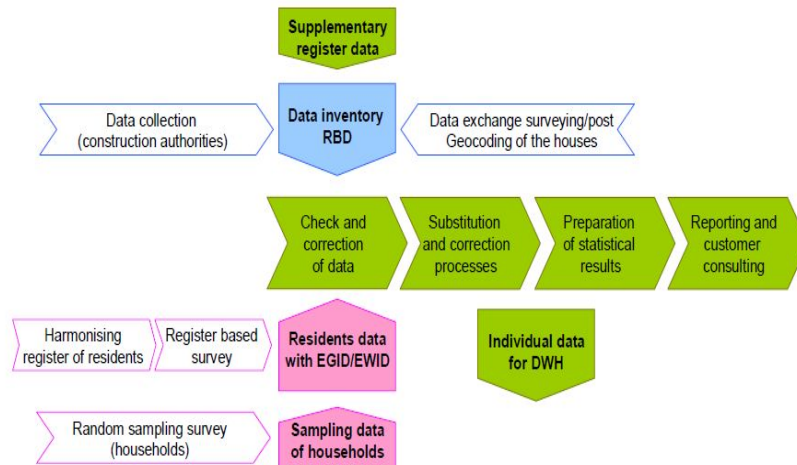
<그림 6-5> 스위스 인구주택총조사 및 관련 등록부의 기본구조



- ① 스위스 영내의 모든 상주인구는 일반가구 또는 집단가구의 구성원이다.
- ② 건물/주택에 별도로 부여된 고유식별번호(EGID-EWID)를 이용하여, 일반 가구는 특정 거처와 연계되어, 점유 중인 주택스톡(dwelling stock)의 비율을 계산할 수 있다.
- ③ 건물의 고유식별번호는 모든 거처에 대하여, 거처의 점유 여부에 관계없이 거처가 소속된 건축물과 연계한다.
- ④ 집단가구는 일반적으로 연방정부 건축물/주택대장 상의 거처를 점유하는 가구가 아니다. 따라서 집단가구는 건물의 고유식별번호를 이용하여 건물과 직접 연계하도록 하는 절차를 채택하고 있다.
- ⑤ 거처가 있는 모든 건물이나 집단가구와 연계되는 건물은 모두 사람이 거주하기 위하여 사용되는 건물로 간주한다. 이들은 앞에서 말한 총 주택스톡(total housing stock)의 비율을 구성하는 것으로 한다.

- 위에서 언급한 상호대조 방식을 통하여, 건물과 주택에 대한 정보는 물론 생활 여건과 관련된 주택총조사의 필요정보를 적출하는 것이 가능하다.

<그림 6-6> 2010년 주택센서스의 자료원과 주요절차



- <그림 6-6>는 2010년 주택센서스의 자료원(data source)과 주요절차를 보여 주고 있다.
- 건물/주택대장은 2002년 1월부터 서비스에 들어갔고, 관리기관은 광역자치체인 칸톤과 기초자치단체이다. 여러 광역자치단체에서는 스위스 통계청의 요구 사항에 따라 관리되는 독자적인 건물/주택대장을 가지거나, 향후에 그러한 공적 장부를 도입할 계획이다.8).

8) 스위스는 강력한 연방 국가이며, 행정은 3단계 구조로, 연방-광역자치단체(칸톤) - 기초자치단체(시)의 상하서열로 구성된다. 그러나 집행기능 대부분은 광역자치단체나 기초자치단체에 위임된다. 따라서 등록센서스에 이용되는 공적 장부의 특성은 가변적이고 통일성이 결여되어 있다. 가령, 26개 칸톤 중, 19개 칸톤은 건물/주택대장과 연결된 칸톤을 단위로 하는 주택보험이 있으며, 방식의 차이가 있기는 하지만 다양한 행정규제를 받도록 되어 있다. 곧 어떤 공식적 통계조사도 개별 광역자치단체의 특수 상황을 고려하여, 시행상의 절차적 검토를 받지 않으면 안 된다.

- 건물/주택대장은 전체 또는 일부가 거주(habitation) 목적으로 사용되는 모든 건물을 포함해야 하며, 거주목적에 전혀 사용되지 않은 주택의 경우에도 거주(dwelling place)의 상세내역은 관련 광역자치단체나 기초자치단체의 요구에 따라서 등록부에 포함하도록 하고 있다. 그러나 연방정부는 그러한 보완적 정보를 업데이트하는 의무를 규정하지는 않고 있다.
- 건물/주택대장에서 모든 건물과 주택은 예외없이, 식별번호를 할당받는다. 건축물 식별번호는 EGID (연방건축물식별번호, federal building identifier), 주택 식별번호는 EWID (연방주택식별번호, federal housing identifier)로 표시된다. 이들 식별번호, 건물/주택의 건축연도, 규모, 전력공급 등 기본정보가 건물/주택대장에 기록된다. 그러나 점유자, 소유권, 임대여부 등의 정보는 기록되지 않는다.
- 2010년의 주택센서스는 다양한 행정자료를 이용하여 실시되도록 되어 있다. 자료의 제일차 자료원은 연방정부의 건물/주택대장이다. 통계측량기관, 우체국 등과의 규칙적인 자료교환은 건물의 지리좌표를 제공하며, 모든 건물이 제대로 기재되었다는 것을 확인하는데 도움을 줄 것으로 기대하고 있다.
- 행정자료, 특히 등록부에서 입수 가능한 추가 자료는 연방정부의 건물/주택대장의 품질이나 최신 업데이트를 보정하기 위하여 사용되며, 또 통계목적으로 추가적인 건축물이나 주택정보를 획득하기 위하여서도 사용한다.
- 기초자치단체의 인구대장(주민등록)은 2010년 스위스 주택센서스의 중요한 정보원이다. 기초자치단체의 인구대장(주민등록)을 표준화하면서, 모든 개인은 예외 없이, 주택대장의 건물식별번호 및 주택식별번호를 배정받게 된다. 인구대장의 개인과 건물/주택대장의 거주자료에 연계하여, 행정자료를 기반으로 가구구조를 파악하고, 생활여건에 대한 핵심적인 정보를 획득한다.
- 스위스 통계청은 2010년 인구주택총조사에서 기존 등록부, 곧 행정자료에 기재되어 있지 않는 센서스 항목에 대해서는 추가적인 표본조사를 연례적으로 실시하여, 그 정보를 획득하도록 하였다. 물론, 스위스의 2010년 인구주택총

조사의 표본조사는 일반가구만을 대상으로 하는 조사이며, 건물 소유주나 재산관리 기관은 통계조사를 목적으로 접촉하는 것이 허용되지 않고 있다.

- 건물/주택대장의 건물/주택자료, 자치체 인구대장의 개인자료, 그리고 표본조사의 결과는 점검되고, 필요한 경우, 보정대상이 된다. 스위스의 2010년 인구주택총조사에서는 비용절감을 이유로 하여, 자료가 완전성을 결여하거나 부정확한 경우에는, 대체(imputation)와 보정(correction)의 절차를 사용하여, 통계적으로 허용되는 범위 안에서 보완된다. 건물/주택대장의 건축물은 모두 지리좌표(geo-code)를 부여하게 되며, 이 지리좌표는 소지역에 대해서도 자료를 제시하고 평가하는 것을 가능하게 만들어 준다. 확인된 자료는 스위스 통계청 내부 데이터 창고(CODDAM)에서 입수 가능하도록 할 방침이며, 기본 자료는 통계표, 다이어그램, 각종 통계분석을 통하여 간행물의 형태로 제시할 방침으로 되어 있다.

3. 스페인

- 스페인 통계청은 2001년 인구주택총조사를 등록부 자료와 전수집계(전통센서스)를 결합하는 방식으로 센서스를 실시하고, 같은 방식으로 2011년 센서스를 실시할 것으로 예상되었다. 그러나 스페인 통계청은 인구부문에 대해서는 전수집계를 폐기하고, 주택부문에 대해서만 전수집계를 실시하는 독일의 등록센서스와 유사한 방식의 부분 등록센서스(partial register-based census)를 실시하기로 결정하였다.
- 따라서 스페인의 2011년 인구주택총조사는 (1) 인구/가구부문의 행정자료 파일, (2) 주택부문의 전수집계 현지실사(센서스) (3) 대규모 표본조사로 이루어졌다. 2001년 센서스의 운영과 다음의 3개 요소로 구성되어 있다.

- ① 센서스 준비를 위한 등록부 파일
- ② 현지실사: 모든 건물의 지리좌표를 확정하는 광범한 건축물 센서스

③ 인구와 거처의 특성을 확인하기 위하여 인구의 일부를 대상으로 실시되는
대규모 표본조사

- 2011년 스페인 센서스는 당초에 센서스 목적에 부합되는 행정자료, 특히 공적 장부 자료를 이용하고, 전수집계로 그 자료를 보완하는 것이었다. 이러한 센서스의 목적은 인구집계의 정확성을 개선하고, 등록부의 결함에서 얻을 수 없는 센서스 항목에 대한 상세정보를 획득하기 위한 것이었다.
- 이러한 센서스는 등록부와 표본조사의 결합에 의하여 생산되는 센서스 통계와 중대한 차이가 있는 것은 두말할 필요도 없다. 그 차이점 중에서 중요한 것을 열거하면 다음과 같다.
 - ① 행정자료에서 얻을 수 없는 정보를 단순히 표본조사로 획득하는 것이 아니라, 전수집계에 의하여 확보한다.
 - ② 등록부에 의하여 획득한 인구집계는 포괄범위(과다, 과소)에 문제가 있기 때문에, 그것을 최고의 수치라고 단정하지 못하며, 전수집계(complete enumeration)에 의하여 점검하고 보완되어야 한다. 따라서 전통적 센서스 방식으로 실시되는 전수집계는 주민등록과 같은 등록부의 포괄범위에 대한 확실한 점검 장치가 된다. 인구센서스는 스페인 통계법이 정하는 바에 의거하여, 등록부의 과다집계와 과소집계에 대한 확실한 정보를 제공하여, 인구집계를 보완하는데 중요한 정보를 제공한다.
- 주택센서스를 전통적 집계방식대로 실시하게 된 것은 주택/거처명부가 행정자료 파일로 사용되는데 어려움이 많았기 때문이다. 인구집계도 전수조사로 이루어는 것으로 당초에 예정되어 있었지만, 최종적으로는 12-13%의 표본조사를 실시하여, 그것으로 행정자료 파일을 보완하고, 행정구역별로 소지역통계를 작성하는데 활용하기로 결정하였다. 상대적으로 행정자료 파일에 본격적으로 의존하는 센서스 방식으로의 전환이 황급하게 이루어졌다고 볼 수 있다.

- 스페인의 2011년 인구주택총조사는 2011년 5월에 제2차 시범조사를 실시하여 센서스 절차를 총괄점검하고, 본조사(main survey)는 2011년 9월에 시작하여, 2012년 3월 말까지 실시되는 것으로 되어 있다. 단계별 작업은 다음과 같다.
 - ① 2011년 9월: 표본으로 선정된 가구에 안내편지를 보내고, 그 후에 응답자들은 조사표에 온라인으로 응답하는 옵션을 선택할 수 있다. 인터넷조사의 자료수집은 2012년 5월에 완료된 것으로 보고되고 있다.
 - ② 2011년 11-12월: 스페인 통계청이 임시직으로 고용한 조사원이 건물/주택 센서스를 실시한다.
 - ③ 2012년 1-3월: 조사원들이 센서스 자료의 수집을 완료하였다.
- 현지실사 작업에서, 조사원 4000명, 집단관리자 800명 (조사원 5명당 1명의 집단관리자), 지역책임자 160명(지역관리자 5명당 1명의 지역책임자)을 고용하였다. 조사원의 선정은 통계청 내 고용센터 요원 중에서 일차적으로 선정하고, 필요한 경우, 공공고용센터 사무소에서 조사원을 선정하도록 하였으며, 센서스 조사원의 자격요건을 충족시킬 수 있도록 노력하였다. 어떤 경우에도, 조사원으로서의 적격성 여부를 평가하기 위하여, 센서스 매뉴얼에 입각하여 선발시험을 치르도록 하였다.
- 스페인 통계청(National Institute of Statistics)은 행정자료를 토대로 작성되는 인구규모(기초자치단체별, 연령, 성별, 출생지, 국적별)의 최종수치는 센서스 실시 이후 1년만인 2012년 말경에 간행될 것으로 예상하고 있다. 여타 센서스 항목의 변수는 2013년에는 이용가능하게 될 것으로 보고 예상하고 있다.

4. 싱가포르

(센서스의 준비과정)

- 싱가포르 통계국이 확인하고 있듯이, 센서스는 정부와 국민의 자료수요를 충족시키기 위하여 전국적으로 실시되는 최대 규모의 통계조사이다. 싱가포르 통계국은 1990년부터 등록부 자료와 전수자료를 결합하는 방식으로 센서스를 실시하다가, 2000년에는 제1차 등록센서스를 실시하고, 2010년 제2차 등록센서스를 실시하면서, 다양한 정부부처가 자문에 응하고, 기획 작업에 참가하였다.
- 싱가포르의 등록센서스는 2000년 라운드와 2010년 라운드에 걸쳐 두 차례에 실시되었지만, 2000년 센서스도 1990년의 준비작업을 거쳐서 실시된 것이라고 보아야 할 것이다. 1990년 센서스는 싱가포르 시민과 영주권자의 신분번호(UIN: Unique Identification Number)와 외국인신분번호(FIN: Foreign Identification Number)를 이용하여 싱가포르 정부의 데이터베이스와 연계하는 작업을 실시하였다. 실제로, 응답자의 취학여부와 최종학력에 대한 정보를 인구통계 기본정보와 연계하여, 센서스 실시이전에 데이터베이스를 구축하였다. 특히, 싱가포르는 이탈리아나 스페인이 2000년 라운드에 실시하였던 등록부의 지원을 받는 센서스, 곧 “register-assisted census”를 실시하였다고 할 수 있으며, 센서스 조사표는 등록부 데이터베이스(Register database)의 자료를 인쇄하여 응답자들이 현지실사에서 확인하는 작업을 거치도록 하였다.
- 싱가포르는 2000년 라운드에 이어서, 2010년 라운드에도 등록센서스를 실시하기 위하여, 통산부 사무차관(Permanent Secretary)을 위원장으로 하는 센서스기획위원회(Census Planning Committee)가 센서스 기준일자인 2010년 8월 1일로 기준으로, 만 2년 전인 2008년 6월에 발족하여, 센서스에 대한 전반적 지침이나 방향을 제공하는 역할을 수행하였다.

<표 6-5> 싱가포르 2010년 센서스의 기획/준비 일정표

기획 및 준비	시점
2010년 센서스 방식 및 예산 검토 작업	2008년 1월-6월
센서스기획위원회 출범	2008년 6월
센서스 자문 및 조사항목 검토작업	2008년 6월-11월
정보기술(IT) 입찰준비 및 계약완료	2008년 3월-2009년 1월
2010년 센서스 IT 시스템 개발	2009년 1월-2010년 2월
센서스 본부 사무소 정비작업	2009년 2월-6월
2010년 센서스 웹사이트 디자인 및 활동개시	2009년 4월-2010년 3월
지역사무소 조달 및 설비 업무 개시	2009년 6월-2010년 3월
2010년 센서스 홍보사업	2009년 9월-2010년 9월
홍보사업 제1호: 싱가포르 통계청의 2010년 센서스 관련 뉴스레터 논문게재	2009년 9월
예비조사	2009년 11월
표본설계 및 선택	2009년 11월-12월
센서스 운영	
전화조사원, 현장조사원, 코딩담당 및 편집담 당 임시직원 모집	2010년 2월-4월
대규모 가구 조사	2010년 2월
현장요원 훈련	2010년 3월-4월
2010년 센서스 공식출범 및 기자회견	2010년 3월 12일
자료수집	2010년 3월-8월
자료처리	2010년 3월-8월
자료배포	
속보결과 공표	2010년 8월
상세보고서 공표	2011년 1월-2월

- 센서스의 중요성을 감안하여, 정책기획이나 정책평가 목적으로 센서스 자료를 이용하는 핵심부서가 센서스기획위원회에 참가하였는데, 가령 지역개발부, 청년 . 체육부, 교육부, 재무부, 내무부, 고용부, 국가개발부, 통산부, 국무총리실 등이 이에 해당한다.
- 2008년 1월에 2010년 센서스 기획작업이 개시되었다. 센서스 조사표에 포함될 자료항목에 대한 의견을 받기 위하여, 이용자 자문이 실시되었다. 몇 단계

의 검토작업이 실시되어 전국수준에서 제안된 조사항목의 적합성이나 센서스에 포함되는 경우, 적합성에 대한 연구가 진행되었다. 2000년 센서스의 개선/보완사항을 확인하고, 센서스의 문항구성을 다듬거나 운영의 흐름을 세련되게 하는 일에 착수하였다.

- 싱가포르 통계국은 1998년 5월에 2000년 인구센서스 조사항목을 기획하는 작업에 착수하였다. 다수의 정부부처와 법정기관들이 센서스 통계자료의 주 이용자이기 때문에, 이들 기관에 센서스 실시기관인 싱가포르 통계국은 기존 조사항목들에 대하여 자문을 구하였다. 이들 기관은 싱가포르 통계국에 추가적으로 필요하다고 판단하는 조사항목에 대하여 답신(答申)하였다. 이러한 부처간 대화과정에서 2000년 행정자료 기반 센서스의 20% 표본조사에서 획득해야 할 조사항목에 대한 리스트를 작성할 수 있었다. 싱가포르 통계청은 해당 리스트를 상세히 검토하여, 조사항목 최종리스트를 작성하여 싱가포르 센서스 기획위원회(Census Planning Committee)에 승인을 요청하였다.
- 2010년 센서스는 공식적으로 2010년 3월 12일에 개시되었으며, 3월부터 8월에 걸쳐 6개월간 자료수집이 이루어졌다. 2010년 6월 기준 싱가포르 인구의 인구통계 프로필을 제공하는 속보자료(Advance Census Release)가 2010년 8월에 공표되었다. 센서스 결과의 상세한 통계표를 담은 자료집은 2011년 2월에 공표되었다. 주요활동의 핵심 스케줄은 <표 6-5>와 같다.

(표본설계 및 선택)

- 싱가포르의 2010년 표본 추출틀은 싱가포르 통계국이 관리하는 전국주택데이터베이스(NDD: National Database of Dwellings)를 토대로 하였다. 전국주택데이터베이스는 싱가포르의 상주목적의 단위주택 명부이다. 싱가포르 2010년 센서스의 표본조사는 상주주택에 거주하는 가구만을 대상으로 하기 때문에, 군 대막사, 호스텔, 호텔 등 시설가구는 추출틀에서 제외하는 것으로 하였다.
- 싱가포르 2010년 센서스 표본은 비례할당 층화추출로 선정하였다. 층화는 싱

가포르 도시재개발청(Urban Redevelopment Authority)에 의하여 구획된 계획지역(planning area)을 바탕으로 정의되었다. 추출틀에 있는 모든 단위거처는 우선 계획지역에 따라 층(層: strata)으로 나누어졌다, 각 층 안에서, 단위거처는 거처유형 별로 구분하였다. 표본은 광범위한 거처유형별로, 비(非)복원 단순무작위추출방법으로 추출하였다. 층별로 선정된 표본은 모두 합하여 20만 단위거처이며, 이것은 센서스 추정방법론에 요구되는 적정 표본요건을 충족시키는 크기였다.

(조사항목)

- 싱가포르 통계청은 2008년 7월부터 12월까지 정부부처와 법정 위원회와 협의하여, 2010년 센서스에서 수집해야 할 문항들을 수집하기 시작하였다. 센서스 표본조사에서 집계되어야 할 항목은 다양한 요인을 고려하였다. 그 이유는 조사항목은 조사표의 포맷, 응답내용의 정확성과 품질, 자료처리단계, 센서스 결과 공표의 일정 등에 중대한 영향을 미치기 때문이다.
- 2010년 센서스 조사표에 포함되어야 할 항목을 평가하면서, 싱가포르 통계청은 유엔의 인구주택총조사 원칙 및 권고사항(United Nations, 2008)에 나오는 (1) 국가적 우선순위 (2) 항목의 적합성 (2) 자원의 이용가능성 (4) 국제비교의 가능성 등을 고려하여 최종안을 마련하였다.

(예비조사)

- 싱가포르 통계청은 2010년 센서스는 2000년 센서스를 반복하는 것이었기 때문에, 센서스 운영의 모든 면을 포함하는 예비조사를 광범하게 실시하지는 않았다. 대신, 싱가포르 통계청은 울트라 MPC (미니 노트북보다 작은 컴퓨터로, “오리가미”로 불림)를 사용하여 현지실사에서 자료수집의 효율성을 증대시키는 절차와 관련되는 사항들을 점검하였다.

- 예비조사는 2009년 11월에 2주간에 걸쳐서 실시되었다. 싱가포르 중부지역의 500여개 단위주택을 선정하여, 울트라 MPC 어플리케이션들이 점검되었다. 현장요원들에게 작업절차와 수집자료의 개념과 정의에 대한 브리핑을 실시하였다. 2010년 센서스의 자료수집방식을 시뮬레이션하기 위하여 현지실사 작업은 인터넷조사, 전화조사, 방문조사 등 3가지 종류로 실시되었다. 예비조사에서, 울트라 MPC와 현장관리시스템 간의 자료전송 문제와 데이터베이스와의 동기화(synchronization)에 대해서도 점검대상이 되었다.

(표본조사)

- 2010년 표본조사는 센서스 기준일 6월 30일을 중간에 두고, 2010년 3월 중순부터 8월 말까지 실시되었다. 인터넷, CATI (Computer-Assisted Telephone Interviewing), 방문조사 등 3가지 자료수집 방식을 사용하였으며, 수집된 자료는 오류와 누락을 제거하기 위하여 자료를 편집하기 위한 자료처리센터로 이첩되었으며, 그 곳에서 제표작업과 분석작업이 가능하도록 코딩을 하였다.
- 2010년 표본조사를 실시하기 위하여, 선정된 거처는 20개 묶음(batch)로 구분하였다. 묶음들은 최소 7,200개 거처 최대 8,600개 거처의 크기로 이루어졌다. 20개 묶음의 각각에 대하여 3-4일 간격을 두고, 엇갈리게 조사를 개시하도록 하여, 3가지 자료수집방식에 대하여 작업량이 분산되도록 하였다.
- 현지실사가 종료될 때는 현장조사 가동 인원이 줄어드는 경향이 있기 때문에, 마지막 묶음에 대해서는 처음의 묶음보다 더 많은 주택수를 할당하였다. 처음의 몇 개 묶음에서는 주택수를 할당하여 현장조사원이 짧은 기간 안에 운영절차와 면접기술을 익히는데 도움이 되도록 하였다.

(자료처리)

- 인터넷, CATI, 방문조사 등의 방식으로 수거된 센서스 조사표는 광학적으로 캡처(optical capture)하여, 센서스 데이터베이스에 업데이트되었다. 훈련된 코딩요원이 직업, 산업, 교육 자료에 대해서는 컴퓨터 지원을 받아 온라인으로 코딩작업을 했으며, 편집요원은 조사표의 내적 일관성과 완전성을 점검하였다. 약 100명의 코딩요원과 에디팅 요원이 1일 2교대로 자료의 일관성이나 품질 개선을 위하여 개별기록의 편집과 확인을 진행하였다.

(결과 배포)

- 3가지 자료수집방식으로 이루어진 등록기반 센서스는 싱가포르의 2010년 센서스 결과를 적시에 포함하는데 기여하였다. 통계청내 프로젝트를 통하여, 2010년 센서스 결과를 신뢰성 있고, 이용하기 편한 제표화 시스템을 개발하였다. 제표화 시스템은 Microsoft Excel 마크로 소프트웨어를 Excel 스프레드시트로 내보기를 하기 전에 SAS 서버에서 집계된 자료를 컴파일하기 위한 인터페이스로 이용하였다. 제표화 시스템은 SAS 엔진에 탑재되어 있기 때문에, 개체자료 처리와 제표화 작업을 신속하게 진행할 수 있기 있었다.
- 싱가포르 통계청은 2010년 등록센서스에서 2010년 8월 말에 “속보 보고서”(Advance Census Release)를 간행하였다. 여기에는 인구규모, 인구성장률, 싱가포르 인구의 지리적 분포를 센서스 기준일자인 6월 30일 현재로 정리한 통계표를 수록하였다. 표본조사 자료 처리를 완결한 후, 통계청은 2011년 1-2월 중 상세한 집계결과를 시리즈 형태로 출판하였다. 이 보고서들은 다음과 같다.

시리즈 1: 인구특성, 교육, 언어, 종교. 2011년 1월 12일

시리즈 2: 가구와 주택. 2011년 2월 15일

시리즈 3: 지리적 분포와 수송 2011년 2월 24일

5. 대만

- 대만 통계청은 2010년 등록센서스로 이행하기 위한 준비로서, 2000년 센서스에서 호적등기부(戶籍登記簿) 자료와 전통센서스의 전수집계 자료를 결합하여, 이른바 “등록부를 이용하는 센서스”를 실시하였다고 할 수 있다.
- 대만 통계청은 2005년 이후 해외사례 연구를 진행하고, 지속적으로 행정자료를 기준으로 하는 등록센서스를 기획하면서, 행정자료와 조사자료의 연계 가능성에 약간의 의문이 존재하는 상황에서 2010년에는 부분 등록센서스, 곧 “등기부를 표본조사로 보완하는 등록센서스”(公務登記輔以抽樣調查方式)를 실시하기로 결정하였다.
- 센서스 기준일은 2010년 12월 26일이었다. 직전 센서스는 2000년 12월 16일이 기준일이어서, 센서스의 주기성(periodicity) 원칙이 정확하게 지켜지지는 않는다. 조사기간은 2010년 12월 26일부터 2011년 1월 22일까지, 4주간 계속하도록 되었다.
- 대만 주계처 통계국은 2010년 센서스 시험조사 자료와 호적등기부 자료의 매칭률이 75% 수준으로, 현재의 우리나라보다 행정자료와 센서스자료의 매칭/연계율이 저조한 상황에서, 2010년부터는 전통적 센서스 방식을 포기하고 호적등기부의 인구통계 정보를 활용하고 이것을 추출틀(sampling frame)로 하는 표본조사에 의하여, 인구, 가구와 주택에 대한 상세정보를 획득하도록 하였다.
- 센서스 지구 범위는 대만지구를 센서스 지구 범위로 하며, 여기에는 대만지구 각 직할시, 각 현과 복건성(福建省)의 금문현(金門縣)과 연강현(連江縣)을 포함하였다.
- 등록부 등 행정자료(公務登記)를 표본조사로 보완하는 방식으로 실시하며, 행

정자료 파일을 연계/통합하여, 대만지구 상주인구 총수와 기본 자료를 생산하고, 다시 16% 표본을 조사구에서 선정하여 표본조사를 실시하고, 각 지역(시, 진, 구)의 상주인구와 관련 상세자료를 추계하였다.

- 센서스 대상: 센서스 표준시각, 대만지구 경내의 표본조사구 범위 안에 있는 상주인구(대만정부가 파견한 해외근무자 및 그 가족, 외국인 노동자, 외국인 교포는 포함하고, 각국의 대만주재 문무관 근무자 및 그 가족들은 제외한다), 그리고 표본조사구의 범위 안의 소유주택을 2010년 대만 센서스의 실지조사 대상으로 하였다.
- 센서스 방법: 방문자조사(interview survey), 조사표 유치(mailed questionnaire), 인터넷 조사 등의 3가지 방법으로 현지실사를 실시하며, 지역별 도시화의 정도에 따라 탄력적으로 운용하였다.
- 센서스 자료의 비밀보호: 센서스에서 획득한 개별자료는 엄격하게 비밀을 유지하며, 자료관리 및 서비스는 통계법과 관련 규정에 의하여 실시하였다.
- 센서스 조사항목
 - (1) 주택상황: 거처의 구분, 거처의 빈집 여부 파악, 주택용도, 주택 기본특성 등의 문항
 - (2) 가구상황: 가구별, 주택소유자 및 해당 가구에 거주한 시점 등의 문항
 - (3) 인구상황: 인구의 기본특성, 상주거주 및 이동현황, 경제활동 현황, 취학지/근무지 위치, 통학/통근 상황, 내국인 건강상태 등의 문항
- 행정자료 파일 이용 항목
 - (1) 내국인 인구: 전체 가구 호적자료 파일, 내국인 출입국 자료파일, 시설수용 인원 파일, 학적자료 파일, 건강보험자료 파일, 심신장애자 관리파일 등 대만 내국인 관련 등록부 등 행정자료 파일
 - (2) 외국인 인구: 외국인 인구 파일, 외국인 국적 노동자 파일, 홍콩, 마카오 및 대륙지구 대만입국 인사 파일 등 외국인 인구 관련 행정자료 파일
 - (3) 건물: 촌(리) 주소대장, 재산세 중 주택관련 파일, 각 직할시 및 현(시)의

주택세 파일 등 건물관련 행정자료 등록부 파일

- 주관기관: 행정원 주계처(主計處)를 주관기관으로 하며, 센서스 추진위원회를 설치하여, 센서스 업무를 책임지고, 관리하였다.
- <표 6-6>은 대만 인구주택총조사의 상세 실시(예정)일정을 (1) 센서스 계획 및 설계 단계 (2) 조사 사전준비 단계 (3) 현지실사단계 (4) 자료처리보고 단계 (5) 후속작업 단계 등의 순서대로 보여주고 있다.
- 센서스 계획 및 설계 단계에는 모두 31개의 세부작업이 있으며, 이들은 모두 2007년 1월부터 시작하여 2011년 12월에 종료되는 것으로 되어 있다. 구체적으로 2010년 12월 인구주택총조사 현지실사 직전까지 해외 센서스 제도 및 실시방법 연구, 수치화된 조사구의 설정 및 타당성 검토 작업, 센서스 목적에 이용되는 공적 등록부의 평가 및 매칭/연계 작업, 표본조사의 추출방식 설계 및 추정방법론의 개발, 시험조사(제1차, 제2차, 제3차)의 실시 및 평가작업, 전수조사 관리방식의 확정, 센서스 조사표의 작성 및 조사항목 용어 정의, 센서스 실시관련 규칙(안)의 준비안 마련, 센서스 통계 결과표 작성방식 확정, 센서스 조직 설치, 인원파견, 훈련방식 결정, 센서스 경비지출 기준 확정, 센서스 주소명부 운영방식 및 조사표 배포방식 결정, 자료처리방식 및 작업방법 연구 등의 다양한 세부사업을 포함한다.
- 등록센서스의 경우에도, 표본조사를 실시하고, 이것을 이용하여 센서스 통계를 작성하기 때문에 조사 사전준비 단계, 현지실사단계, 자료처리보고 단계, 후속작업 단계 등에서는 전통센서스 또는 전수집계와 행정자료를 동시에 이용하는 센서스의 경우와 그리 다르지 않은 것으로 나타난다. 이것은 전통센서스에서 등록센서스로 센서스 방법론의 패러다임 전환(paradigm shift)을 준비하는 나라에서, 센서스 계획 및 설계 단계의 세부작업이 대단히 중요함을 말해주는 것이라고 할 수 있다.

<표 6-6> 2010년 대만 센서스 프로젝트 상세 실시(예정)일정

프로젝트 항목		실시일정
센서스 계획 및 설계 작업 단계	1. 해외 센서스 제도 및 실시방식 연구	2007년 1월-2010년 11월
	2. 수치화된 조사구(geo-coded EA)의 설정	2007년 1월-2010년 3월
	3. GIS 수치화된 조사구 구획의 타당성 검토작업	2009년 1월-2010년 9월
	4. 센서스에 이용될 공적 등록부 파일의 평가작업	2007년 1월-2008년 12월
	5. 공적 등록부 파일의 연계 및 보완 작업	2009년 1월-2011년 12월
	6. 인구주택총조사 규칙(안)의 준비작업	2007년 1월-2008년 12월
	7. 표본추출 방식의 설계 및 추정방식의 개발	2007년 1월-2009년 12월
	8. 센서스 조사표(안) 작성, 조사항목 및 용어정의 연구	2007년 7월-2010년 4월
	9. 센서스 방식(초안)의 확정	2009년 1월-7월
	10. 센서스 방식(초안)의 행정원 보고 및 실시 승인요청	2009년 6월-7월
	11. 제1차 시험조사 실시 및 개선사항 점검	2007년 9월-2008년 8월
	12. 제2차 시험조사 실시 및 개선사항 점검	2008년 10월-2009년 8월
	13. 제3차 시험조사 실시 및 개선사항 점검	2009년 8월-2010년 4월
	14. 센서스 방식(안)의 관련기관 회람 및 검토의견 요청	2009년 2월-2010년 4월
	15. 전수조사 작업방식의 확정	2008년 7월-2010년 4월

	16. 센서스통계 결과표 작성방식(안)의 확정	2009년 2월-2010년 6월
	17. 조사표 설명 및 검토기준 확정	2009년 2월-12월
	18. 각급 센서스 조직 설치, 인원파견, 훈련방식의 확정	2009년 1월-2010년 5월
	19. 각급 센서스조직 검토/조사요원 필수공지사항 확정	2009년 7월-2010년 6월
	20. 센서스 경비수지 기준 및 처리작업의 확정	2009년 7월-2010년 6월
	21. 센서스 조직 및 업무인원 선정 고려사항의 확정	2009년 7월-2010년 6월
	22. 센서스 홍보작업 실시핵심 및 홍보자료 작성의 확정	2008년 1월-2010년 8월
	23. 조사표 배포방식의 확정	2009년 7월-2010년 6월
	24. 센서스 감독작업 방법의 확정	2009년 7월-2010년 6월
	25. 센서스 주소명부 및 기타 운영관련 작업방식의 확정	2008년 1월-2010년 6월
	26. 자료처리 및 작업방식의 확정	2008년 1월-2009년 12월
	27. OCR 판독장치 연구개발작업	2008년 7월-2010년 7월
	28. 센서스 행정관리시스템 개발	2009년 6월-2010년 1월
	29. 센서스 자료 기기점검 및 오류상태 확인	2009년 1월-2010년 7월
	30. 인터넷조사 관리시스템의 개발	2008년 7월-2011년 2월
	31. 시스템분석 실시 및 프로그램 설계/테스트작업	2008년 7월-2011년 2월
조 사	32. 센서스 표본추출 작업 실시	2010년 1월-6월
	33. 홍보용 단편, 단극, 표어 및 각항 홍보용품 제작	2010년 1월-8월
	34. 홍보작업 추진	2009년 7월-2011년 1월

사전준비단계	35. 센서스작업 매뉴얼, 참고자료 및 각종 문건 편제작업	2010년 5월-9월
	36. 센서스 조사구지도 제작 및 센서스 명부 편제작업	2010년 5월-9월
	37. 센서스 관련 각종 질의응답(FQA) 소책자 편제작업	2010년 1월-8월
	38. 자료처리 테스트 및 조사원 훈련작업	2010년 2월-2011년 3월
	39. 광역자치단체(직할시, 현시) 조사본부 설치	2010년 9월-2011년 2월
	40. 기초자치단체 조사지부 설치	2010년 9월-2011년 2월
	41. 각급 센서스 요원 선정	2010년 9월-10월
	42. 중앙간부요원 대상 설명회 및 연찬회 개최	2010년 7월-9월
	43. 감독요원 및 강사 대상 설명회 개최	2010년 8월-9월
	44. 각급 센서스 요원 대상 강습회 개최	2010년 11월-12월
	45. 기자회견 개최	2010년 12월
	46. 센서스 조사표 및 각종 물품 현지 조사본부 및 지부로 배포 및 전달	2010년 10월-11월
	47. 조사경비의 배정	2010년 9월-11월
현지실사단	48. 현지실사작업의 실시	2010년 9월-12월
	49. 업무관련 작업회보 관리	2010년 12월-2011년 1월
	50. 감독원 회의 및 순회감독 실시	2011년 12월-2011년 2월
	51. 센서스 조사표 작성결과 편집작업 실시	2011년 1월-2월
	52. 센서스 조사표 및 관련문건 수거 및 조사경비의 확인 및 집행	2011년 2월
	53. 사후조사	2011년 2월-3월

계		
자료처리보고	54. 자료수집 및 오류검정 작업	2011년 3월-5월
	55. 각급 센서스 조직 및 작업요원 업무성과 평가작업 실시	2011년 2월-7월
	56. 자료오류 개정, 보완 및 추계작업	2011년 4월-6월
	57. 행정자료와 조사표 자료를 이용하여 결과표 작성 준비	2011년 6월-9월
	58. 센서스 속보자료 작성 및 보고서 제책작업	2011년 5월-7월
	59. 센서스 속보자료 보고서의 행정원 보고 및 인쇄 발행	2011년 9월-10월
	60. 센서스 총보고서 결과분석 및 보고서 편집작업	2011년 11월-2012년 7월
	61. 센서스 총보고서 행정원 보고 및 인쇄발행	2012년 11월-12월
후속작업	62. 센서스 기획 및 설계요원 업무성과 평가작업 실시	2012년 11월-12월
	63. 센서스 사업의 평가 및 개선사항 마련	2012년 7월-12월
	64. 센서스가 실시되지 않은 연도에 대한 기초자치단체 상주인구 추계를 위한 연구사업 실시	2013년 이후
	65. 행정 등록부 자료 파일 및 인구센서스 데이터베이스 작성에 관한 연구	2013년 이후
	66 보완목적의 특별조사 실시	2013년 이후

6. 폴란드

(센서스 준비과정)

- 폴란드는 2002년 5월에서 6월에 걸쳐서 전통적 방법으로 센서스를 실시하였으며, 2011년에는 등록센서스로 센서스 방식을 전환하였다. 폴란드의 등록센서스는 유럽지역 센서스 사업의 일환으로 실시되었으며, 유럽연합 통계청(Eurostat: Statistical Office of the European Union)이 정한 센서스 규정을 준수하여, 등록부와 이를 보완하는 표본조사의 문항을 결합하여 센서스 통계를 작성하기로 하였다.
- 폴란드 통계청은 2011년 등록센서스 전환을 준비하면서, 몇 차례의 시험조사를 시행하고, 2011년에 등록센서스 본조사를 시행하고, 그 결과를 2013년에 공표함으로써, 센서스 전환 프로젝트를 완료하는 것으로 되어 있다.
- 폴란드의 2011년 센서스는 결합센서스 모형으로 등록자료(인구통계 기본변수는 전수조사)와 20% 특별목적의 표본조사 자료를 결합하는 것이 특징이라고 할 수 있다. 가령, 폴란드의 2011년 등록부 자료 중 인구통계 변수는 인터넷 조사에 의하여 전면적으로 기재내용의 정확성을 점검한다는 점에서, 일반적으로 실시되는 부분 등록센서스 국가의 센서스와는 차이가 있다고 할 수 있을 것이다.
- 폴란드는 표본조사 실시와 관련하여, 전화인터뷰는 2011년 4월 8일부터 6월 30일까지 3개월 동안 실시하였다. 컴퓨터 전화면접(CATI)에 선정된 형식은 행정자료나 인터넷 자기기입식으로 획득한 자료로 채워진 것들로, 응답자로부터 얻을 수 있는 답은 별로 없다.

- 폴란드 통계청은 행정자료를 표본조사를 위한 자료원, 센서스 표본추출단위를 구체화하기 위한 정보원(주소 및 주택 추출틀, 농장), 그리고 임퓨테이션, 자료 추정, 센서스의 품질을 비교하고 최종확정하기 위한 자료로 사용할 것이다. 폴란드 통계청은 300여개의 행정자료를 확보하였으며, 그 중에서 3개는 센서스 정보시스템에 입력될 수 있는 유용한 정보를 포함하고 기본대장이 될 수 있고, 또 30여개 행정자료가 보조정보를 획득하는데 이용될 수 있다고 밝히고 있다.

<표 7-7> 폴란드 2011년 등록센서스의 기본대장

기본대장의 이름	대장 관리기관
주민등록시스템(Population Identification Register System): - 중앙전산주민등록시스템(PESEL) - 지방자치단체 등록파일	내무부(Ministry of Interior Affairs and Administration) 자치단체 등록사무소
사업체대장(REGON)	폴란드 통계청
행정구역대장/건축물 대장	폴란드 통계청

(전수조사)

- 폴란드 통계청은 2010년 4월 1일부터 5월 31일까지, 2개월간에 걸쳐서 소규모의 센서스 테스트를 16개 단위지역에서 실시하여, 등록센서스로의 이행에 따른 필수적인 절차에 대하여 점검작업을 진행하였다.
- 폴란드의 2011년 전수조사는 행정자료와 모든 응답자가 정보를 제공하는 간단한 조사표로 보완되는 방식으로 진행하였다. 전화면접과 개인면접을 사용하는 자료보완은 자료획득을 위한 채널이라기보다는 보완적 채널에 해당하는 것이었다. 자료를 획득하는 기본방식은 전수조사로서, 컴퓨터를 이용한 면접

(CATI)과 마스터 파일을 포함하였다.

- 센서스의 완전성 곧 포괄범위를 점검하기 위한 두 개의 기록을 사용하는데, 그것은 인구자료와 주택자료이다. 등록부에서 추출된 변수로 이루어지는 마스터 기록은 자료의 수집을 지원하는 4번째의 채널로서, 인터넷조사, 전화조사, 직접면접과는 별도로 운영된다.
- 간단한 조사의 출발점은 센서스 실사 마이크로데이터베이스(Micro Database)의 마스터 레코드로서, 마스터 레코드의 자료와 함께 조사의 포괄범위를 계산하기 위하여 사용한다.
- 등록부에서 획득한 자료로 완전히 포괄되는 조사는 센서스 운영 마이크로 데이터에서 자동적으로 보관되고, 다른 채널로 이관되지 않는다.

(표본조사)

- 표본조사는 폴란드 영토 안에 영구적으로 또는 일시적으로 거주하는 사람들을 대상으로 실시되며, 표본추출은 가구를 대상으로 하여 실시하였다.
- 표본조사는 CAI 방법과 CAPI 방식으로 실시되고, 자료는 CATI 방식으로 보완하였다. 거처의 20%를 표본으로 추출하여 표본조사를 실시하였다. 조사의 완전성을 관리하기 위하여, 표본추출률에 대한 점검을 실시하며, 센서스 조사원 거처를 대상으로 조사하였다.
- 폴란드 통계청은 행정자료를 연계하는 방법으로 주소를 사용하였으며, 단위통계자료에 대한 주소지의 (x,y) 좌표를 이용하여, 이들 자료에 대한 종전의 공간확인 시스템(system of spatial identification)을 변경시켜, 지역분류(센서스 구역)에서 주소지 분류로 이행하였다. 이것은 2011년 폴란드 센서스의 획기적 혁신이라고 생각되며, 센서스 자료의 소지역 수준 그룹화 등 지리통계의

방향을 새로이 제시할 것으로 기대하고 있다. 또한, 이 좌표시스템은 공간적 성격의 마이크로 데이터베이스를 구축하는데 도움을 줄 것이며, 다양한 공간 분석을 하는데도 도움을 줄 것이라고 판단하고 있다.

- ① 인구통계(자녀 거주지와 부모 거주지의 평균거리, 취업지, 학교 간의 통근, 병원까지의 거리)
- ② 도시화와 공간기획(도시집적체, 대도시의 경계를 결정하고, 토지개발계획의 작성에 유용함)
- ③ 농업과 환경(작물구조, 환경오염의 분석)
- ④ 경제(각종 도로 및 산업투자계획의 효과 분석)

(자료처리와 분석작업)

- 센서스 자료의 처리과정은 우선, 관련 법규에 규정된 행정자료 특히 등록부에서 센서스 조사항목을 추출하는데서 출발한다. 폴란드 통계법은 국가통계의 작성기관인 통계청이 센서스 대상에 접근할 수 있는 권한을 갖도록 규정하고 있다. 행정자료는 단위자료를 결합하는데 필요한 개인자료와 신분번호를 이용하였다.
- 행정자료에서 획득한 모든 단위자료는 통계등록부(statistical register)로 변환하는데, 여기에서 정선(精選), 이중등록 삭제, 자료 표준화 등의 절차를 진행한다. 이 절차는 DOS SAS 환경에서 진행되도록 하였다. 동시에, 등록부에서 획득한 입력자료, 정선절차, DOS 절차를 적용한 이후의 최종 품질 등에 대한 메타자료를 수집하였다.
- 정선된 자료는 운영 마이크로 데이터베이스(Operational Microdata Base)에 획득된 등록부에 일치하는 논리적 레이어(logical layer)의 웨이브로 적재하도

록 하였다.

- 폴란드 통계청은 객관적 이론적 통계적 지식을 기반으로, 자료를 요약하고, 연계, 구조, 시스템, 요인, 집중 그리고 통계적 가설 검정이나 결과를 일반화하기 위한 다양한 분석을 시도할 계획으로 되어 있다. 현재 시도되고 있는 절차는 통계적 기술, 상관분석, 편차분석, 회귀분석, 모델분석, 차원축소, 분류, 비모수검정, 통계적 추론, 공간분석 등을 포함한다.
- 센서스 본조사가 끝나고, 센서스 대상 모집단의 2%에 대하여 사후조사(PES: Post-Enumeration Survey)를 실시하도록 되어 있다. 목적은 센서스의 완전성 곧 포괄범위를 점검하고, 센서스에서 획득한 변수의 신뢰도를 결정하고, 다양한 변수(행정자료의 출처, 조사원, 등록부, 응답자 등의 자료관련 변수)가 자료의 품질에 어떤 영향을 미치는가를 연구할 예정으로 되어 있다.
- 센서스 결과보고서는 유럽연합 통계청(Eurostat: European Union Statistical Office))이 규정하고 있는 방법론에 의거하여, 2013년까지 작성을 완료할 예정으로 되어 있다. 또 국내이용자를 위하여 다양한 통계표를 마련하여, 그것을 종이 보고서와 인터넷으로 배포할 예정으로 되어 있다.

7. 터키

(센서스 방법론 전환의 계기)

- 터키는 2011년 등록센서스를 추진하기 전에는 전통적 센서스를 실시하였고, 조사표와 현지실사에 의한 전통적 센서스는 인구규모와 인구특성을 파악하기 위한 주요 자료원이 되었다.

- 터키가 센서스 방법론의 전환에 당면한 장애요인이 한두 가지가 아니었지만, 유럽연합(EU)의 정식 회원국이 되기를 원하는 상태에서, 유럽에서 실시되는 센서스와 같이 상주인구를 집계하기 위하여, 전통적 센서스에서 사용된 종전의 현주인구 집계방식을 폐기하고자 하였다. 그러나 문제는 행정자료에는 인구규모나 인구특성에 대한 센서스 변수가 찾아보기가 힘든데, 그것은 이들 자료를 5년 또는 10년마다 실시되는 전통적 인구센서스에서 획득하였기 때문이다. 더욱이, 전통적 센서스의 자료처리에는 3년 반의 기간이 소요되었기 때문에, 인구통계에 대한 업데이트한 정보가 존재하지 않았다.
- 2011년 인구센서스 이전에는 센서스 조사표를 이용하는 전통적 센서스에서, 전국적으로 통금(通禁)을 선포하여, 하루 만에 현지실사 작업을 완료하여야 했기 때문에, 극히 제한된 문항만을 조사항목으로 선정할 수 있었다. 센서스 조사항목에 포함된 주요 문항은 인구특성, 사회경제특성, 그리고 단위주택에 대한 몇 개의 문항뿐이었다. 특히, 인구이동, 장애, 교육, 가구상태, 단위주택의 특성에 대한 상세한 조사항목은 센서스 조사표에 포함될 수가 없었다.
- 터키 통계청은 2011년 이전의 전통적 센서스 실시에서 발생하는 문제를 해결하고, 지역별 인구에 대하여 업데이트한 정보를 획득하고, 인구이동을 추적하기 위하여, 주소 기반 인구등록시스템(ABPRS: Address-Based Population Registration System)을 도입하기로 결정하였다. 이 등록시스템의 일차적 목적은 상주지 주소에 입각하여 기존의 행정자료를 체계화하였다는 점에서, 2011년 등록센서스 전환을 위한 획기적 조치라고 볼 수 있다.
- 2006년 4월 25일자 인구등록법(Population Registration Law, No 5400)에 의거하여, 터키 통계청(TurkStat, Turkish Statistical Institute)은 인구등록시스템의 수립이라는 중책을 맡게 되었으며, 터키 내무부 인구 및 국적 담당국(GDPCA, General Directorate of Population and Citizenship Affairs, Interior Ministry)이 수립된 인구등록시스템을 개발하고 유지.관리하는 업무를 맡도록 하였다.
- 인구등록법에 의거하여, 전국주소데이터베이스(NAD, National Address

Database)를 수립하여, 터키 내의 모든 주소정보를 포함하는 명부를 작성하였다. 다음에는 터키 영내에 거주하는 내국인과 외국인의 상주지 주소를 획득하고 이들 주소정보와 중앙호적시스템(MERNIS, Central Civil Registration System)하여, 주소 기반 인구등록시스템(ABPRS, Address-Based Population Registration System)을 구축하였다. 더욱이, 터키의 모든 정부기관이나 관련 사업체는 그들의 행정업무에서 인구등록시스템을 이용할 수 있도록 하였다.

- 2007년 주소기반 인구등록시스템(ABPRS)의 수립 이후, 등록시스템의 자료를 바탕으로 매년 1월에 전년도 12월 31일 기준으로 행정구역(주, 구역, 읍, 마을, (도시의) 지구)별로 인구수와 기본특성(성/연령별 구조, 등록지, 국적, 문자해독 및 최종학력, 법적 혼인상태, 주 수준의 국내인구이동)에 대한 정보를 공표하도록 하였다.
- 터키 통계청의 2007-2011년 국가통계 프로그램에 의하면, 2011년 터키 인구주택총조사는 유엔의 2010년 라운드 인구주택센서스 원칙 및 권고사항이나 유럽연합의 2010년 라운드 인구주택센서스 요구사항 존중하여 실시하기로 하였다. 이에 따라, 2011년 인구주택총조사를 행정자료와 표본조사를 조합하는 방식으로 진행하기로 하였다.
- 터키 통계청은 전통적 센서스에서 행정자료를 이용하는 센서스로 전환하기 위하여, 2009년 11월 1일 자문위원회 제1차 회의를 개최하였다. 여기서, 실무위원회를 설치하기로 하는 안이 승인되었다. 당시의 회의는 행정자료 기반 센서스의 방법론을 연구하고, 기술 관련 문헌을 준비하고, 센서스가 조사해야 할 항목을 제시하기 위하여 6개 실무위원회를 설치하기로 하였다. 실무위원회가 담당하는 주제영역은 다음 사항을 포함하였다.

- (1) 센서스 방법론
- (2) 센서스 응용
- (3) 자료처리
- (3) 센서스 실시에 필요한 예산
- (4) 행정자료 기반 센서스의 홍보작업

(5) 센서스 결과의 배포문제

- 터키 통계청은 무엇보다 먼저, 행정자료 기반 센서스를 추진하기 위하여 두 종류의 조사표를 작성하였다. 하나는 short-form 조사표로 이것은 전수조사에 사용하며, 인구등록시스템을 점검하기 위한 것이다. 다른 하나는 long-form 으로, 표본조사에 적용하여, 인구특성을 상세하게 파악하기 위한 정보를 획득 하는데 사용한다. 조사표에 대한 사전조사(Pre-test)가 두 차례에 걸쳐, 2009년 1월과 2월에 수도 앙카라(Ankara)의 몇 개 지역과 주변 농촌 지역에서 실시되었다. 당시에 실시된 사전조사의 목적은 다음과 같았다.

① 조사항목의 질문이 명쾌한가, 의미가 있는가, 모든 사람들이 이해할 수 있는가, 그리고 질문에 대한 선택지 수가 충분한가에 관련되는 사항을 점검한다.

② 조사표의 설계, 조사표의 각 파트별 응용의 유용성을 파트와 문항의 순서, 조사항목에 응답하는데 필요한 공간이나 박스의 크기 등에 관련되는 사항을 점검한다.

③ 조사표의 모든 조사항목에 대한 응답시간을 추정한다.

(시험조사)

- 터키 통계청은 2010년 3월-4월에 걸쳐, 센서스 표본조사의 조사방식을 테스트하기 위하여, 수도 앙카라의 아큐르트구(Akyurt District)에서 시험조사를 실시하였다.

① 조사표를 학교교사가 학생들에게 전달하고, 가구별로 조사표를 완성하여 학교로 반송한다. (교육부의 협력을 필요로 하였음)

② 조사원 방법(가구방문을 통한 직접조사)

③ 가구주/자계식(조사원이 센서스 조사표를 직접 전달하고, 가구별로 조사표를 완성하여 조사원이 직접 수거하는 방식)

- 앙카라 아큐르트 사전조사에서 집계지역, 자료입력, 응용프로그램이 또한 점검 대상이 되었다. 2010년 7월에는 센서스 방법론이 사전조사와 시험조사에서 변경되었다. 터키 통계청의 청장과 지역사무소의 행정책임자가 참석한 가운데, 현지실사는 전수조사를 하면서, 독특한 조사표를 적용하기로 결정하였다.
- 2010년 9-10월, 터키 통계청은 센서스의 모든 단계를 테스트하기 위하여, 시험조사를 실시하였다. 시험조사의 기준일자는 2010년 10월 3일이었고, 기준 주일은 2010년 9월 27일에서 10월 3일까지로 하였다. 시험조사를 위한 훈련은 터키 통계청 지역사무소의 직원들이 담당하였다. 시험조사의 현지실사는 2010년 10월 4일에 시작하였다. 터키 통계청의 지역사무소가 우선적으로, 조사구와 전국주소데이터베이스를 관리하였다. 그리고 이들 지역사무소는 조사표를 가지고 현지실사를 관리하였다. 가구조사표, 집단시설조사표 (앙카라 지역만 조사하였음), 노숙자 인구에 대한 조사표(주별 모든 노숙자 시설을 대상으로 함)가 점검대상이 되었다.
- 시험조사는 214개 지역(81개 주, 81 구역 센터, 52개 마을)과 257개 조사구에서 실시되었다. 대략 5만 가구가 집계되었다. 시험조사의 현지실사는 10월 말에 끝이 났다. 시험조사의 평가회의는 지역사무소 행정책임자가 참석한 가운데, 2010년 11월 4-6일에 개최되었다. 시험조사의 결과를 분석하기 위한 연구가 2010년 10월-12월의 3개월 동안 진행되었다. 2010년 12월 23-24일 7개 지역사무소 행정책임자(앙카라, 이스탄불, 이즈미르, 트라브존, 하타이, 네브세히르, 안탈야 지역)가 참석한 가운데, 센서스 현장조직을 평가하고, 센서스 조사표를 다시 토론하기 위하여 회의를 개최하였다.
- 센서스 방법론의 논의하고, 당시 진행 중이던 연구를 평가하기 위한 임시회의가 2010년 12월 27일 관련 부처의 장관과 행정책임자들이 참석한 가운데 개최되었다. 당시 참석한 부처는 다음과 같다.

- 국가기획위원회(SPO)
 - 내무부 인구 및 국적 담당국(GDPCA)
 - 보건부
 - 교육부
 - 고등교육평의회
- 임시회의 결과에 따라, 전수조사가 아니라, 등록시스템과 표본조사를 결합하는 행정자료 기반 센서스를 실시하기로 결정하였다. 2011년 인구주택총조사의 추정수준과 표본추출, 국제기구의 자료요청을 논의하기 위한 회의가 2010년 12월 30일에 터키 통계청장의 주재 아래 관련부서의 참여로 개최되었다. 센서스의 현지실사를 위한 표본추출의 방법론은 당시 진행 상태였다. 조사표에 포함될 변수를 결정하고, 국제기구의 자료요구를 평가하기 위하여, 2010년 12월 31일에 관련기관의 최고 책임자들이 참석한 가운데 회의를 개최하였다.
 - 가구 조사표의 마지막 사전조사는 터키 통계청 지역사무소 주관으로 실시되었으며, 구체적으로, 응답자들이 문항을 쉽게 이해할 수 있는지, 선택지의 개수는 충분한지 등을 점검하였다. 가구 조사표는 2011년 5월 13일에 완성되었다.
 - 터키의 2011년 행정자료 기반 센서스의 운영, 준비작업, 응용단계, 평가업무 등 센서스 운영을 관리하기 위하여, 터키 통계청은 이해당사자들로 구성된 자문위원회를 구성하였다. 자문위원회의 구성원은 다음과 같다.
 - 내무부, 인구 및 국적 담당국(GDPCA)
 - 국가기획위원회(SPO, State Planning Organization)
 - 하세테프 대학교 인구문제연구소(Hacettepe University, Institute of Population Studies)
 - 교육부(National Education Ministry)
 - 보건부(Health Ministry)
 - 법무부(Justice Ministry)

(인구등록시스템의 개발)

- 터키의 2011년 인구주택총조사는 2011년 10-11월에 인구등록시스템이라는 행정자료를 가지고, 이것을 이용하여 대규모 표본조사를 실시하는 방식으로 실시하였다. 센서스의 기준일자는 2011년 10월 2일이고, 현지실사는 2011년 10월 3일에 시작하여, 자료 수집은 2개월 동안 계속되었다. 2011년 총조사의 실사작업 조사원이 CAPI 방식으로 1대1 면접조사로 진행하였다. 센서스 실시를 위하여 정부예산이 배정될 것이다. 표본규모는 250만 가구로, 전체가구의 13.5%에 해당하는 규모로 실시하였다.
- 2011년 인구주택총조사의 모집단(인구, 가구, 주택)에 대한 추정작업은 주 수준과 지구 수준(구체적으로, 모든 주에 대해서, 모든 주도에 대해서, 그리고 2만 명이상의 지구 중심도시에 대해서, 모든 지구의 마을 합계에 대해서), 집단 시설이 있는 장소에 대해서는 전수집계를 하도록 하였다.
- 조사구는 2011년 8월에, 업데이트된 전국주소데이터베이스(NAD)를 이용하여 터키 영내의 모든 주소들을 고려하여, 터키 통계청 방법론 부서(Methodology Department)가 결정을 하게 될 것이다. 전체 조사구(EA: Enumeration Areas)를 확정한 후에, 표본조사에 선정될 조사구를 정하도록 하였다.
- 조사표에 대해서는 방법론 실무위원회에 의하여 연구가 진행되고 있다. 최종 가구조사표는 현재 완성되어 있으며, 2011년 말에 실시될 센서스 현지실사를 위하여 인쇄하였으며, 집단시설이 있는 장소에 대한 조사표도 별도로 만들었다.
- 주소 기반 인구통계시스템(ABPRS)를 구축한 후, 인구규모, 성/연령별 인구구조, 법적 혼인상태, 출생지, 문자해득 및 최종교육연수, 가구의 크기, 가구주와의 관계 등은 마을 수준에서 등록시스템에서 획득 가능한 통계이다. 터키 통계청은 주소 기반 인구등록시스템이나 다른 행정자료(등록부 포함)와 정례적으로 실시되는 가구조사에서 획득할 수 없는 센서스 변수에 대한 정보는 2011

년 인구주택총조사의 표본조사로 입수하도록 하였다. 터키 통계청은 2011년 등록센서스 조사표에 포함한 항목을 논의하는 과정에서, 유럽연합 통계청(Eurostat)의 핵심 조사항목과 터키 정부의 국가적 자료수요를 고려하여 센서스 항목의 최종 리스트를 결정하였다.

(센서스 조사항목)

- 현재, 터키 통계청이 2011년 인구주택총조사 표본조사에 수집하게 될 정보의 전체적 윤곽은 다음과 같은 것을 포함하였다.
 - 건물 및 주택 특성
 - 혼인상태
 - 상주지 및 5년전 거주지
 - 인구이동의 이유
 - 해외 거주 여부, 터키 입국연도
 - 장애 여부
 - 문자해독 및 최종교육연수
 - 노동력
 - 출산력
 - 사망(영유아, 아동, 성인)
- 행정구역별 인구는 센서스 기준일자인 2011년 10월 2일을 기준으로, 인구등록시스템에서 집계하도록 하였다. 그 후, 인구수, 성/연령별 구조, 가구규모는 등록부에서 직접 추출하였으며, 표본조사에서 획득되는 다른 특성들은 센서스 추정방법론과 캘리브레이션(calibration) 방법을 이용하여, 센서스 모집단 특성을 추정하도록 하였다.
- 터키 통계청은 표본조사의 현지실사는 3단계로 진행하였다.

1. 표본조사구에서 주소를 관리하여 (2011년 9월 전반부), 새로운 주소를 데이터베이스에 입력한다.
 2. 주소 기반 인구등록시스템에 등록된 사람을 등록시스템에서 인쇄한 가구 리스트별로 (Form B) 표본조사구 주소의 실제상황과 비교하여 점검작업을 실하였다(조사표 적용의 제1단계, 2011년 10월-12월).
 3. 표본조사구의 모든 주소(일반적인 거처나 단위주택으로 사람이 거주하도록 설계되지 않은 건물 가령, 대피소, 축사, 천막까지 포함한다)를 방문하여, 해당 주소에 가구가 거주하는지 아닌지를 점검하고, 거주하는 경우에는 해당 가구를 면접하여 조사표를 작성하도록 하였다. (2011년 10월-11월).
- 현지실사에서, 각 지역의 조사원, 관리자, 책임자는 모두 수당을 지급한다. 현지실사에서, 1일 15개의 면접이 가능할 것으로 가정하였다. 현지실사에서 다양한 방법을 이용하여 점검 작업을 실시하였다.
 - 각 지역의 관리자, 책임자에 대하여 중앙관리시스템에서 추출되는 일일 및 주일 보고서
 - 관리자에 의한 현지 조사원의 업무량 관리 및 책임자에 의한 관리자의 업무량 관리
 - 각 지역에 대한 3단계 자료분석(관리자, 책임자, 터키 통계청 센서스 실시본부)
 - 터키 통계청은 SAS 통계분석을 사용하여, 터키 통계청과 지역사무소는 자료분석을 진행할 예정이다. 터키 통계청은 현지실사 과정을 관리하기 위하여, 다양한 보고용 프로그램을 개발하고 있으며, 자료분석을 위한 보고프로그램의 연구는 진행되고 있다. 표본조사가 끝나고 나면, 품질점검과 품질관리 목적으로 등록부에 따른 조사자료의 상세한 분석이 실시하도록 되어 있다. 그러나 최종결과표의 간행에 대한 예상 일정은 아직도 공표된 자료가 없다.

제3절 주요국의 2020년 라운드 센서스 계획과 준비상황

1. 영국의 “Beyond 2011 Programme”

- 영국은 유럽 대륙의 등록센서스 실시 국가들과는 달리, 2011년까지 현지실사에 의한 전통적 센서스를 실시하여 왔다. 그 이유는 유럽지역에서 흔히 볼 수 있는 인구대장, 곧 우리나라의 주민등록과 같은 시스템이 존재하기 않기 때문이다. 영국왕립통계학회가 70여 년 전부터 주민등록시스템의 구축을 주장하고, 2004년에도 영국통계청이 직접 주민등록시스템 구축의 타당성 연구를 진행하였으나, 그것은 현실화되지 않고 있다.
- 영국에는 주민등록시스템이 없는 대신, 선거명부(Election Register)가 존재한다. 가구주는 가구 내 유권자를 등록하도록 되어 있으며, 2009년 정당 및 선거법(Political Parties and Elections Act 2009)에 의거하여, 유권자 등록은 법정 의무가 되었다.
- 2008년 노동당 정부는 신분증법(Identity Cards Act 2008)을 제정하여, 자발적으로 신분등록을 시작하였으나, 2010년 총선 후, 새로이 집권한 보수당-자민당 연립정부는 국민들의 불신대상이 되고 있던 신분증법의 시행을 폐지하고, 신분등록제도의 계속 추진을 중단하였다.
- 영국 통계청(Office for National Statistics)는 잉글랜드와 웨일즈 지방의 국민건강보험 환자명부(NHS Patient Register)를 행정자료 옵션(administrative data option)의 자료원으로 검토작업을 진행하였다. 이 환자명부는 국민건강보험 일반개업의(General Practitioners)들에게 등록된 모든 사람들의 기록으로, 이것은 인구현상을 측정하기 위한 명부는 아니지만, 잉글랜드와 웨일즈 지방에 거주하는 모든 사람들에 대한 정보를 포함한다고 볼 수 있는 가장 방대한 자료이다.

- 영국에서는 통계청은 상부 관리기구인 통계기구(UK Statistics Authority)의 지침을 바탕으로, 「2011년 이후 센서스 실시계획」(Beyond 2011 Programme)을 확정하고, 센서스 옵션에 대한 연구결과를 2014년까지 영국의 회에 보고하도록 되어 있다.
- 영국의 「2011년 이후 센서스 실시계획」(Beyond 2011 Programme)의 관리 방식은 현재 개발 중이며, 관리체제의 핵심집단은 다음과 같다.
 - ① 영국통계기구(UK Statistics Authority)는 내각 특정부처의 감독 하에 놓이지 않고, 의회에 직접 보고한다. 또한 영국통계기구는 영국통계청의 운영에 대한 감독권을 갖는다.
 - ② 영국통계청 실무위원회(Executive Committee)는 영국통계기구에 직접 보고하며, 전반적 전략과 정책에 대한 책임을 지도록 되어 있다. 실무위원회는 전체 예산은 물론, 「2011년 이후 센서스 실시계획」 개별 프로젝트에 대한 각종 업무를 승인한다. 실무위원회는 프로젝트의 착수와 실행을 선임담당자(Senior Responsible Officer)에게 위임하며, 쟁점사항은 방향과 해결방안을 제시한다.
 - ③ 「2011년 이후 센서스 실시계획」 프로젝트 위원회는 선임담당자를 위원장으로 하며, 프로젝트 관리팀에게 파생 업무를 최대한 위임하여 업무가 순조롭게 이루어질 수 있도록 한다. 위원회는 선임담당자를 지원하여, 의사결정이나 승인과정에서 본연의 책임을 다하는 것을 돕는다.
 - ④ 영국인구위원회(UK Population Committee)는 잉글랜드/웨일즈, 스코틀랜드, 북아일랜드 통계담당기관으로 구성되며, 이들은 영국 전체의 인구통계와 인구분석 프로그램에 대한 합의를 형성하고, 그것을 감독한다. 위원회는 「2011년 이후 센서스 실시계획」 프로젝트의 장래 방향에 대하여 수시로 전략적인 조언을 하는 것을 업무로 한다.
 - ⑤ 부처간 전략집단(Interdepartmental Strategy Group)은 범정부적 차원에

서 센서스 통계 이용자들을 센서스 방법론 개혁에 참여시켜서, 그들의 자문을 받도록 한다.

- ⑥ 「2011년 이후 센서스 실시계획」 프로젝트 설계 및 세부추진 위원회 (Beyond 2011 Design and Delivery Board)는 프로젝트 결과의 표준화를 확실히 하고, 프로젝트의 세부추진 사항을 조정하고 관리하는 책임을 맡으며, 프로젝트의 결과와 편익이 프로그램 업무추진계획에 정해진 대로 이행될 수 있도록 한다.

- 영국 통계청(ONS)가 고려하는 2011년 이후의 센서스 옵션은 모두 8개로, 이들은 모두 3개 그룹으로 대별할 수 있다(<그림 6-8> 참고).

(a) “전통적 센서스” 옵션(traditional census options)

- ① 전통적 센서스 (Long Form 전수조사)
- ② Long Form (표본조사) + Short Form (전수조사)
- ③ 미국 센서스 모델: Short Form 센서스 (전수조사) + 연속 표본조사 (“Rolling sample”)
- ④ 프랑스 센서스 모델: 순환센서스 (주기는 5년 또는 10년)

(c) 행정자료 - 표본조사 옵션(Administrative data - survey options)

- ⑤ 행정자료 - 집계수준 (aggregate level)
- ⑥ 행정자료 - 개체수준 (record level)
- ⑦ 행정자료 - 중간수준 (intermediate level)

(d) 표본조사 옵션("survey" option)

- ⑧ 주소명부 + 표본조사

<그림 6-8> 2011년 이후 영국 센서스의 옵션 개괄



자료: UK Office for National Statistics (2012), Beyond 2011 'The Options Explained' Methods & Policies Report' (영국 통계청 웹사이트 <http://www.ons.gov.uk/>에서 입수가능)

• 옵션 5 (행정자료-집계수준)

- 이 옵션에서는 영국 정부가 관리하는 다양한 등록부, 가령 선거인 명부(Election Register)와 국민건강보험 일반개업의 행정자료(NHS GP Register Data)를 자료원(data source)으로 하여, 이러한 공적 자료에서 생성한 집계자료(aggregate data)를 이용하여, 센서스 목적에 부합하는 인구규모를 잠정적으로 추정한다.
- 잠정 추정인구(provisional population estimates)는 다양한 등록부의 자료를 연령, 성별, 지역 등의 변수로 총화하여, 그것을 가중평균(weighted

average)하여 작성한다.

- 포괄범위 점검조사 자료와 같은 추가정보를 이용하여, 총화된 잠정 추정인구(provisional population estimates)에 적용될 수 있는 가중치(weights)를 적용하여, 추정인구를 확정할 수 있다. 이 때 가중치 계산을 위해서는 비추정법(Ratio Estimation)을 사용한다. 이 가중치를 규칙적으로 업데이트한다면, 추정인구를 매년 작성할 수 있고, 센서스를 매년 실시하는 효과를 얻을 수 있다.
- 물론, 경제사회통계는 포괄범위 점검과는 별도의 표본조사를 실시하여 수집하여야 하며, 가까운 장래에는 공공부문과 민간부문의 행정자료 정보로 표본조사의 정보를 일부는 대체할 수 있다.

• 옵션 6 (행정자료-개체수준)

- 이 옵션에서는, 개체수준의 행정자료(주소명부를 포함)를 전부 매칭/연계하여, 잠정 추정인구를 작성하는 것이 그 목적이다.
- 정례적인 포괄범위 점검조사를 실시하여, 잠정 추정인구의 정확성을 평가할 수 있으며, 여기서는 행정자료의 포괄범위, 바로 과소범위(undercoverage)와 과다범위(overcoverage)를 측정하여야 한다.
- 추정방법(이원시스템추정, 비추정)은 영국의 2001년과 2011년의 One Number Census 프로젝트에서 시험되었으며, 이러한 추정방법론은 잠정 추정인구에 적용되는 가중치를 도출하여, 최종적으로 추정인구를 확정하기 위하여 사용할 수 있을 것으로 영국 통계청(ONS: Office for National Statistics)은 판단하고 있다.
- 행정자료의 경우에는 추가적인 통계적 보완, 가령 편집이나 임putation을 통한 보완이 가능할 수도 있다.
- 센서스의 사회경제통계를 작성하기 위한 자료는 표본조사를 실시하여 별도로

수집되어야 하며, 장래 행정자료의 정보로 표본조사의 일부항목을 대체할 수 있다.

- 옵션 7 (행정자료-중간수준)

- 이 옵션은 행정자료 접근방식을 집계수준과 개체수준에서 모두 이용한다. 집계행정자료는 잠정추정인구(provisional population estimates)를 작성하기 위하여 사용한다.

- 과소범위와 과다범위를 측정하는 포괄범위 점검조사 자료는 포괄범위를 점검하는 지역에서 개체 행정자료(record-level administrative data)와 주소명부(address register)를 매칭하고 연계하기 위하여 사용될 수 있다.

- 추정방법론(이원시스템추정, 비추정)을 사용하여 잠정추정인구에 가중치를 적용하여, 최종적으로 추정인구를 확정하는데 이용할 수 있다.

- 이 경우에도 사회·인구학적 통계를 작성하기 위한 자료는 포괄범위 점검 목적의 표본조사와는 별도의 표본조사에 의하여 수집되어야 하며, 장래 행정자료의 정보로 표본조사의 일부항목을 대체할 수 있다.

- 옵션 8 (주소명부 + 표본조사)

- 주소명부(address register)와 표본조사를 결합하여, 센서스 모집단의 인구와 관련 특성을 추정한다.

- 가령, 표본조사에 포함된 개별 주소에 거주하는 평균 인구를 해당 지역의 주소 개수로 곱하여, 해당 지역의 인구를 추정하는 비추정법(ratio estimation)을 이용한다.

- 추정치는 매년 업데이트 가능하지만, 소지역 추정(small-area estimation)에 많은 시간과 노력이 요구될 수 있다는 것이 현재의 관심영역이다. 세심한 총

화작업(stratification)을 진행하여, 추정인구의 품질을 개선할 수 있지만, 이 접근방식은 주소명부의 품질이나 내적 정합성(internal consistency), 가구규모의 지역편차에 대단히 민감하다는 문제점이 있다.

– 행정자료나 추가적인 포괄범위 점검조사(coverage survey)를 이용하여, 주소명부의 품질을 개선하고, 표본조사 자료를 소지역 추정 목적에 이용하여, 추정인구의 품질을 개선하는 것이 이 옵션의 일차적 목표이다..

- 영국은 전통적 센서스의 대안을 마련하기 위하여, 다음과 같은 세트기준(set criteria)을 사용하여, 반복적 평가를 실시한다.

(a) 통계이용자의 요구사항을 충족시킬 수 있는 능력

(i) 작성된 통계의 정확성

(ii) 통계작성의 빈도 (매년 업데이트하는가, 아니면 5년, 10년마다
업데이트하는가의 문제)

(iii) 통계가 작성되는 지리적 상세도(detail)

(b) 작성비용(총 소요비용과 시점별 소요비용의 변동폭)

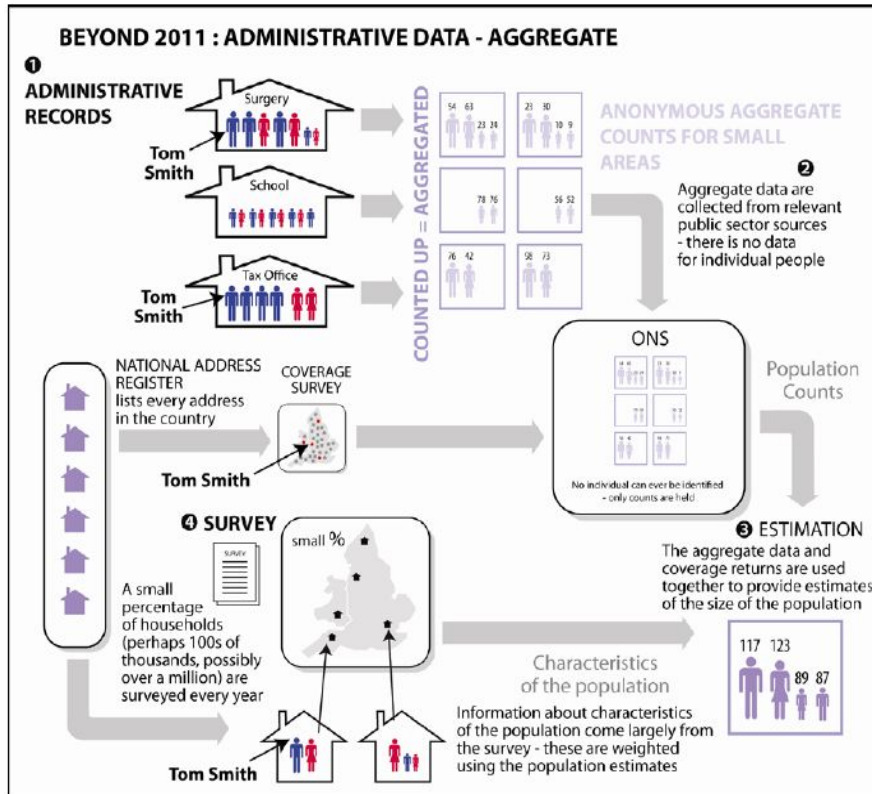
(c) 기술적, 법률적 실행가능성

(d) 위험

(e) 공적부담과 공적 수용도

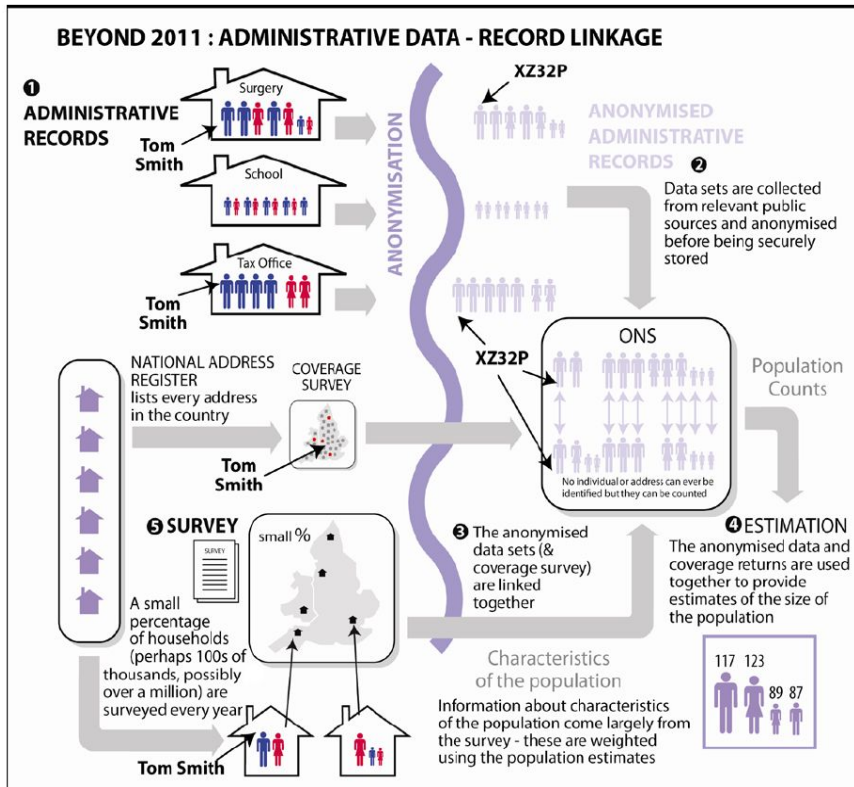
참고: 행정자료 센서스 옵션 5, 6, 7, 8

(1) 옵션 5: 행정자료-집계수준



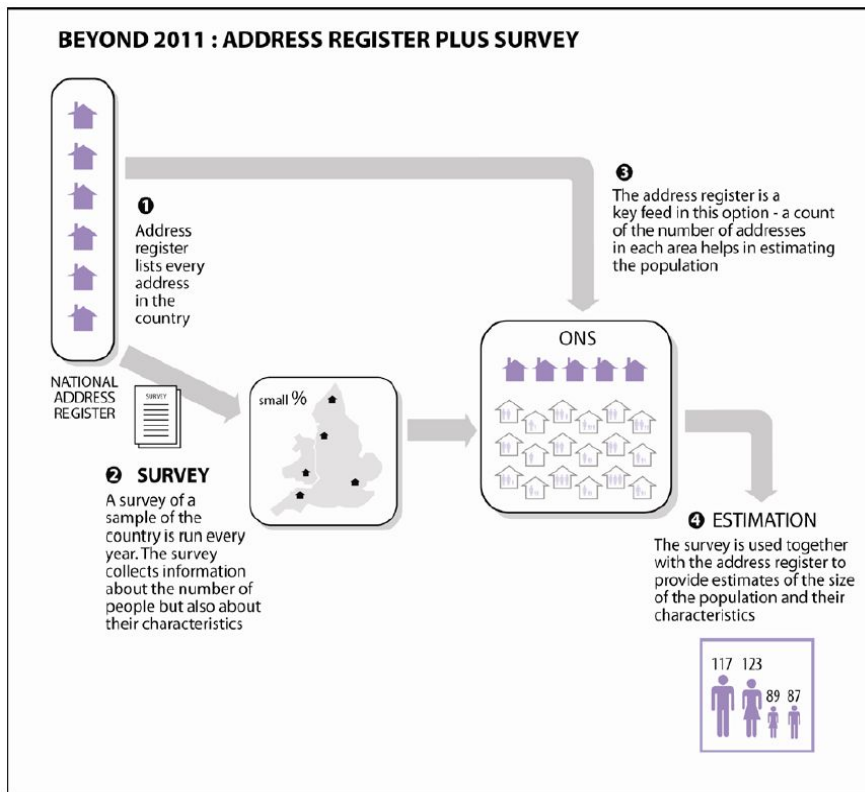
이 옵션은 다양한 행정자료원(예: 선거인등록부, 건강보험 환자등록부)에서 추출한 집계자료를 이용하여, 잠정적인 추정인구를 결정한다. 포괄범위 점검을 위한 표본조사를 실시하여, 잠정 추정인구의 품질을 개선한다. 센서스 모집단의 일부에 대하여 포괄범위 점검 목적과 다른 별도의 표본조사를 실시하여, 인구의 사회경제적 특성에 대한 자료를 수집한다.

(2) 옵션 6과 7: 행정자료-개체수준 및 중간수준



이 옵션에서는 다양한 행정자료원에서 개체별로 나온 상세자료를 매칭/연계하여, 잠정적인 추정인구를 결정한다. 여기에는 개체자료의 전부를 연계할 수도 있고, 일부를 연계할 수도 있다. 여기에서도 포괄범위 점검 조사를 실시하여, 행정자료로 집계된 잠정 추정인구의 과다범위와 과소범위를 보완하여, 추정인구의 품질을 개선할 수 있다. 또, 집계수준 행정자료 옵션과 같이, 인구의 사회경제적 특성에 관한 통계를 수집하기 위하여, 포괄범위 점검 목적과는 별도의 표본조사를 실시할 수 있다.

(3) 옵션 8: 주소명부+표본조사



이 옵션에서는 인구규모를 추정하기 위한 정확한 주소명부(address register)가 필요하며, 이것을 기반으로 표본조사를 연차적으로 실시하여, 센서스 통계를 작성한다.

2. 이스라엘 2020년 통합 순환센서스 운영계획

- 이스라엘 통계청(Central Bureau of Statistics)은 1949년 국가창설과 함께, 센서스를 실시하고 최초의 센서스 실시시점에 구축된 주민등록 시스템을 10여년의 주기로 실시된 센서스에서 적극적으로 이용하여 왔다. 그러나 이들 센서스는 주민등록과 센서스의 전수집계를 결합하는 센서스에 근접하며, 등록부의 포괄범위나 기재내용을 보조정보로 이용하는 성격이 강한 센서스였다고 할 수 있다.
- 이스라엘 통계청은 경비절감과 응답부담의 경감을 목적으로, 등록부를 보조정보로 이용하는 센서스에서 등록부를 주축정보로 이용하여, 등록부의 포괄범위를 보완하고, 기재내용을 적극적으로 이용하여 센서스를 실시하는 방법, 곧 「통합센서스」(integrated census)로 전환을 추진하였다. 이스라엘의 2008년 통합센서스는 등록부와 표본조사를 결합하는 방식으로 실시된 센서스로서, 아시아 지역에서 싱가포르의 2000년 등록센서스와 함께 센서스 방법론의 전환에 있어서 선도적 역할을 하고 있다고 할 수 있다.
- 이스라엘 통계청은 이러한 2008년 센서스를 토대로 또 하나의 도약을 준비하고 있다(Statistics Israel, 2012). 그 도약은 2020년 라운드의 센서스에서 통합 순환센서스(IRC: Integrated "Rolling" Census)를 실시하는 것으로 나타나고 있다. 구체적으로, 2020년 라운드 센서스에서는 2008년 통합센서스의 기본원칙이 그대로 적용되지만, 10년 기간에 걸쳐서 순환표본조사를 실시하는 센서스를 운영하기로 방침을 확정하였다
- 이스라엘 통계청은 당초의 통합센서스에서 순환센서스 방식을 혼합하기로 한 이유로 한정된 센서스 예산과 응답자의 부담경감을 들고 있다. 이스라엘 재무부가 내린 결정은 2021년 이후의 센서스는 2008년 센서스에 배정된 총예산의 범위를 초과해서는 안 된다는 것이고, 남은 예산은 자료원이나 통계목적 등록부의 개발을 지원하는 센서스를 개발하는데 사용할 수 있다는 것이다. 예산 가용성에 대한 기대는 센서스 사업에서 예산 정점화(頂點化)를 피함으로 해서

절약되는 비용의 평가를 근거로 한 것이다.

- 2008년 센서스 데이터에 대한 통계적 검정의 결과, 과소범위(UC) 점검조사의 17% 표본이나 과다범위(OC) 점검조사의 20% 표본과 비교해서 각각의 점검 조사에 10%의 표본을 선정한다고 하더라도 센서스 모수(母數)인 인구 추정치의 분산(dispersion)이 그리 크게 증가하지 않다는 것이 확인되었다. 이 때문에, 이스라엘 통계청은 센서스 모집단인 응답자에 주어지는 응답부담을 경감하기 위하여, 과소범위와 과다범위의 점검에 이용되는 표본조사의 표본규모를 1/2로 축소하기로 결정하였다.
- 이스라엘 통계청은 시간과 비용의 절감을 통하여 남은 예산으로 부분 등록센서스에서 완전 등록센서스를 목표로 하여, 건축물/주거대장, 교육대장, 소득대장 등 3개 주요대장을 개발하는데 관심을 기울이고 있다.

(a) 건축물/주거대장(BDR: Building/Dwelling Register)은 지방세 대장(local tax register)을 기반으로 하며, 이스라엘 통계청이 주관하는 모든 가구조사의 표본프레임으로 사용할 예정으로 되어 있다. 현재, 건축물/주거대장을 도입하는데 필요한 통계단위 명부의 품질은 차이가 있으며, 어떤 자료는 품질 면에서 불충분하며, 등록부 대장에 포함되기 위해서는 추가적인 개선작업이 필요하다.

(b) 교육대장(Education Register)은 이스라엘 영토 안에서 인·허가된 교육기관 전부에서 매년 접수되는 파일을 기초로 한다. 교육대장은 통합 순환센서스의 교육 자료의 품질을 개선하여, 센서스 조사항목의 최종적인 완성도를 높이는데 사용할 예정으로 되어 있다.

(c) 소득대장(Income Register)은 소득세 파일을 기초로 작성되며, 사회보장급여(social security benefits)로 수령된 소득 자료를 포함하게 된다. 2008년 통합센서스에서 센서스 파일의 소득정보의 흠결을 보완하기 위하여 사용한 적이 있으며, 현재 작성되고 있는 소득대장은 2020년 통합 순환센서스에서 핵심자료로 사용하게 될 예정이다.

- 이스라엘 통계청은 2020년 통합 순환센서스의 실시와 관련된 시설 센서스(시설에 입소한 인구의 완전한 센서스)를 2년을 주기로 실시할 예정으로 되어 있다. 각 주기의 1차 연도에는 시설대장(Institutions Register)을 업데이트하는데 시간을 할애한다. 각 주기의 2차 연도에는 (1) 등록부에 기재된 모든 시설의 인구의 100%에 대한 인구통계조사 (2) 시설인구의 일부에 대한 사회경제통계 표본조사로 두 가지 조사를 실시한다. 시설센서스의 표본조사는 2단계 표본을 선정하여 실시된다. 제1 단계는 시설의 10%를 추출하고, 제2단계는 추출된 시설 안의 20% 입소자를 포함한다. 이스라엘 통계청은 통합 순환센서스의 취지에 맞추어 10년을 센서스 주기로 하여, 시설의 50%를 추출하고, 사회경제통계 표본조사에서 시설인구의 10%를 면접하도록 한다.
- 표본추출 방법론은 원칙적으로 2008년 통합센서스를 위하여 개발된 방식을 준용하고, 필요한 경우에 한하여 약간의 수정을 한다.
 - (a) 과소범위(UC) 점검 조사 - 등급별로 기초자치단체를 매년 3단계 추출하며, 기초자치단체 내 통계지역(SA: statistical areas)의 10%, 추출된 통계지역 내(가능한 모든 경우)의 주거단위 또는 건물의 10%를 추출한다.
 - (b) 과다범위(OC) 점검 조사 - 표본조사의 추출틀(sampling frame)로, 보관된 지리좌표가 부여된 중앙인구대장(CPR: Central Population Register)을 이용한다. 추출된 통계지역에 등록된 행정가구(administrative households)의 10%를 추출하는데, 이것은 매년 총가구수의 1%에 해당한다.

표본으로 추출되지 않은 통계지역에서 추가적으로 1%를 매년 표본으로 추출하여 과다집계 점검 조사를 실시하고, 이것을 기초자치단체 수준의 과다범위를 추정하는데 사용한다.
- 이스라엘 통계청은 통계조사 중 표본규모가 상대적으로 큰 노동력조사(LFS: Labor Force Survey)를 센서스 포괄범위와 신뢰도를 지원하기 위하여 하기 위하여 사용하기로 하였다. 표본크기의 확충, 지리적 포괄범위의 개선, 그리고 노

동력조사 문항에 센서스 문항의 도입을 통하여, 노동력조사를 기초자치단체 수준에서 과소집계 점검의 목적으로 이용할 예정이다.

- 2008년 통합센서스의 운영을 목적으로 개발된 기술적 수단과 애플리케이션(application)들을 약간 변용하여 2010년 이후에 계속해서 사용하기로 하였다. 2008년 센서스에서 발견된 흠결사항은 수정되었다. 이러한 약간의 변화는 개발에 소요되는 시간과 인원을 감축시켰으며, 2010년 시설센서스 제1라운드와 2012년 통합 순환센서스의 제1라운드의 착수를 가능하게 하였다.
- 2008년 통합센서스와 향후의 통합 순환센서스 목적으로 개발된 테크놀로지와 운영절차는 이스라엘 통계청이 실시하는 다른 가구조사에도 적합한 것으로 판명되었으며, 적은 시간과 직원을 투입하여 실시되어야 하는 특수목적의 통계조사를 지원할 수 있도록 수정되었다.
- 2008년 통합센서스를 개발한 대부분의 숙련된 직원들은 이스라엘 통계국에 그대로 근무하고 있으며, 통합 순환센서스의 개발에 참여하였다. 관리담당 직원도 또한 센서스 운영에 참여하고 있다.
- 현재, 이스라엘 통계청은 2012년 2월에 시작하여, 2012년 11월에는 끝내기로 되어 있는 2011년 통합 순환센서스에서 자료수집단계에 와 있다. 순환표본조사의 기준일자는 2011년 11월이다. 중앙인구대장과 건물/거처대장이 표본조사의 추출틀이다. 시설센서스의 제1라운드는 완성되었으며, 이들 자료는 2011년 추정인구 산정에 포함될 것이다.
- 통계지역에 대한 연차별 추정인구는 업데이트된 센서스 결과를 기준으로 수정될 것이며, 기초자치단체에 조정된 추정치로 캘리브레이션(calibration)될 것이다. 조정수치는 통합 순환센서스 결과와 추정 총인구(estimated total population)를 기준으로 할 것이며, 보완된 중앙인구대장의 인구동태/인구이동 등록을 고려하게 될 것이다. 시설인구의 추정치는 격년으로 작성될 것인데, 그것은 가구조사의 추정이나 캘리브레이션 목적(calibration purpose)으로 사용되는 비 시설인구(non-institutional population)의 추정치를 작성하기 위한

것이다.

- 사회경제통계 자료는 센서스 주기의 중간시점(10년 주기로 6년째에 해당하는 연도)부터 생산할 예정으로 되어 있다. 자료수집의 제1차 10년 주기는 2022년에 완결될 수 있다. 그럼에도 불구하고, 이스라엘 통계청은 향후, 통합 순환센서스의 사회경제통계에 대한 잠정결과를 공표할 예정으로 되어 있다.

3. 이탈리아의 순환 등록센서스 모델

(이탈리아 2011년과 2021년 센서스 대전환)

- 이탈리아 인구주택총조사는 1861년부터 10년 주기로 계속 실시되었으며, 센서스 방법론의 주요특징은 조사표를 이용하여 이탈리아에 거주하는 내국인과 외국인을 전수 조사하는 것이었다. 센서스 조사표는 조사원이 배포하여 수거하지만, 응답자가 조사항목마다 직접 정보를 제공하는 자계식(自計式)로 진행된다. 이탈리아 센서스는 상세조사표(long-form questionnaire)만을 사용했으며, 따라서 조사항목의 모든 자료는 표본조사가 아니라, 전수조사로만 수집되었다는 것이 특징이다.
- 이탈리아에서, 인구주택총조사의 실시 목적으로 “법정인구”를 정하고, 상주 인구의 인구통계 및 사회경제적 특성에 대한 정보를 수집하는 것 외에도, 센서스 현지실사를 통하여 기초 자치단체체 인구대장(Anagrafi)을 업데이트하기 위하여 자료를 획득한다. 실제로, 인구대장법(Regolamento Anagrafico)은 센서스 결과를 이용하여, 지방자치단체의 주민등록시스템을 업데이트해야 한다고 규정하고 있다.
- 이탈리아 통계청(ISTAT, Italian National Institute of Statistics)는 센서스의 집계에 책임을 지고, 센서스 설계와 조정업무를 직접 수행하며, 현지실사는 대

부분, 중앙정부에 의하여 할당 된 예산 외에 독자적인 예산과 자원을 이용하는 기초자치체에 의하여 실시된다. 이탈리아 통계청은 또한 현지실사 자료와 인구대장자료를 비교하여 포괄범위를 점검하고, 조사표의 응답내용에 대한 편집 작업을 일차적으로 실시하고, 최종적으로는 센서스 결과를 바탕으로 인구대장의 오류를 수정한다.

- 이탈리아는 2001년과 2011년 센서스의 센서스를 등록부와 전수집계를 결합하는 방식으로 진행하여 왔지만, 향후의 센서스 실시와 관련해서, 다른 나라와 마찬가지로 센서스 자료의 시의성(時宜性)이라는 품질기준을 고려해야 하는 압력에 직면하였다. 이 때문에 연차적 자료수집에 의한 센서스 실시를 계획하고 있으며, 기초자치단체의 인구대장을 이용하는 복수시점 센서스(MTP Census: Multiple Time Population Census)를 염두에 두고 있다. 이 계획에 의하면, 이탈리아 인구 모집단의 일부를 포괄하는 표본조사를 연속적이고 누적적으로 실시함으로써 자료를 수집하는 센서스가 이탈리아 통계청이 의도하고 있는 최종목표임을 알 수 있다.
- 이탈리아의 지방자치단체는 앞서서도 설명했지만, Anagrafe 제도를 통하여 주민, 가구원관계, 상주지 주소, 출생, 혼인, 인구이동 등의 인구동태사건을 기록하는 주민등록시스템을 두고 있다. 이러한 인구대장(population register)에는 센서스 조사구의 경계나 코드, 그리고 주소명부를 포함하고 있다. 그러나 심각한 문제점은 지방자치단체별로, 주소명부의 품질이 차이가 크다는 것이다. 이들 자료의 표준화(harmonization)와 규격화는 향후에도 여전히 센서스 실시의 심각한 문제점으로 등장하고 있다.
- 2011년의 이탈리아 센서스는 인구대장과 같은 행정자료를 이용하기는 하지만, 기본적으로 전통적 센서스 방식에 의거하여 실시되었지만, 다양한 혁신을 포함하고 있었다. 가령, (1) 지방자치단체 인구대장에서 가구와 주소의 명부를 획득하고, (2) 이 리스트에 포함된 모두 가구에 조사표를 송달하고 (3) 웹, 우편수거 또는 자치단체별 수거센터 등 조사표 수거방식의 복수채널화 (5) 인터넷 조사 관리시스템의 도입 (6) 조사원별 직접 무응답과 과소범위의 정도를

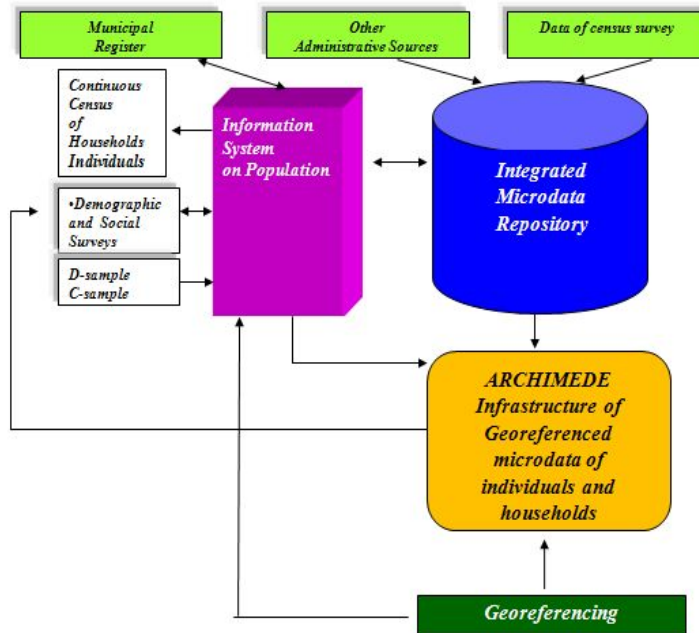
확인하는 절차의 도입 등을 정하는 것이었다.

- 2011년의 등록부를 보조정보(auxiliary information)로 이용하는 전수집계 센서스에서 주도(州都)와 인구규모가 2만 명 이상인 지방자치단체에 대해서는 조사방식의 개선이 있었다. 구체적으로 (1) 센서스 사전조사(事前調査)를 실시하여, 주소명부를 현장점검하고, 점검된 주소에 대해서, 거처(주택)단위의 수에 대한 정보를 작성하며, (2) 1만 5,000명 인구규모의 센서스지역(census area)을 구획하며, (3) 가구표본에 대해서 long-form 조사표로 사회경제통계 자료를 수집하는 것으로 이루어졌다.
- 이러한 2011년 센서스 방법론 개선은 응답부담과 운영비용을 축소하는데 필요한 조치로 기획되어진 것이었다. 센서스 목적에 행정자료의 적합성을 강화하기 위하여 추가적인 변화를 도입하는 것이 필요하다는 것이 이탈리아 통계청(ISTAT)의 판단이다. 우선 센서스 비용이 너무 방대하고, 조사 자체가 단일 시점에 집중되어 있기 때문에 통계적 효율성(statistical efficiency)도 낮은 편이었다. 또, 행정자료의 이용도 최적 수준에 이르지 못하고 있으며, 지리적/시간적 상세성도 제한된 수준에 머물고 있었다.
- 이탈리아 통계청의 판단을 종합적으로 평가하면, 현재, 2011년 센서스의 다양한 방법론 개혁에도 불구하고, 그것만으로는 센서스의 비용-편익 간 균형을 안정적이고 지속적으로 달성하는데 충분하지 않다. 실제로, 센서스가 여전히 고비용이고, 그 비용이 특정 시기에 집중되어 있으며, 행정자료의 이용이 센서스 통계의 작성에 최적조건을 충족시키지 못하고 있다.
- 주지하다시피, 이탈리아에서도 센서스 자료의 시의성 상실 속도는 빠르며, 지리적으로 상세한 소지역 자료(small-area data)는 10년을 주기로 공급되고 있다. 이 때문에, 이탈리아 통계청은 센서스 방법론을 전면적 개편이 필요가 있다고 판단하고 있다.

- 미시간대학 통계학 교수였던 Leslie Kish(1910-2000)는 “인구의 집계만이 아니라, 다양한 사회경제변수에 대하여 지리적으로 상세한 연차통계를 제공하는 것은 많은 나라에서 순환표본의 주목적이 될 것이다”라고 했다(Kish, 1990, 1998, 1999).
- 이러한 제안을 바탕으로, 2000년대부터 미국 센서스국(Census Bureau)은 10년 주기의 센서스를 업데이트하기 위하여, 순환표본 방식을 사용하는 American Community Survey(ACS)를 발족하였다(US Census Bureau, 2004). 이 조사는 대규모의 표본조사로, 인접연도의 자료를 통합(pooling)하여, 상이한 지리적 상세수준에서 이동평균법(moving average methods)에 의해서 추정치를 산출하는데 사용한다. 현재, 프랑스도 미국과는 아주 다른 방식이기는 하지만, 순환센서스 방식을 채택하는 나라이다. 프랑스 통계청(INSEE: Institut National de la Statistique et des Études Économiques)은 2004년 이후 5년을 주기로, 60%의 인구를 조사하고, 이동평균 추정치를 이용하여 인구집계를 업데이트한다.
- 센서스 방식으로 순환조사의 주된 장점은 인접한 시점의 자료를 누적하여 추정치의 효율성을 증대시킬 수 있다는 것이다. 이러한 이점은 주기적으로 지리적으로 사전 정의된 지역(pre-defined area)에 대하여 추정치를 산출할 수 있다는 것이다.
- 이탈리아는 2021년 센서스에서, 행정자료의 이용과 함께, 다년간 순환 표본조사를 운용하는 “순환센서스”(rolling census) 방식을 구상하고 있다. 이탈리아 통계청은 매몰비용(sunk cost)이 너무나 큰 1회성 대규모 통계조사인 전통적 센서스의 문제점을 극복하기 위하여, 전통적 센서스에서 1회성 통계조사가 아닌, 상이한 시점에서 특정의 센서스 통계를 작성하는 연속적 조사로 이행하고자 한다. 이것은 자료원과 현장운영의 기존 네트워크를 원용하는 새로운 센서스 디자인을 의미하게 될 것이다.
- 주 행정자료원은 자치단체 주민등록(LAC: Liste Anagrafiche Comunale)이

다. 주민등록 관리규정(Regolamento Anagrafico)에 의하면, 이탈리아의 모든 자치단체는 “Anagrafe”라 불리는 자치단체 주민등록을 관리하도록 되어 있다. 자치단체에 상주하는 모든 사람들은 이 등록부에, 가구원이나 상주지 주소에 대한 정보와 함께 기재하도록 되어 있다. 더욱이 출생, 혼인, 국내이동, 해외이동 등의 생애사건을 모두 “Anagrafe”에 기재하도록 되어 있다.

<그림 6-10> 이탈리아 인구정보시스템 내 입력자료와 산출자료의 관계



- 센서스 지도작성을 업데이트하기 위한 순환전략이 연속센서스를 위하여 채택 될 것이다. 전국은 5개의 지역으로 분할되고, 이들 지역은 인구 2만 명 이하의 지방자치단체 20%, 인구 2만 명 이상의 지방자치단체에 속한 센서스 지역의 20%를 포함하게 될 것이다. 이들은 크기, 지리적 분포, 자치단체의 통계적 대표성 등의 제반기준에서 균형을 유지할 수 있도록 할 것이다. 각 지역의 센서스 지도작성, 집계지역, 센서스지역, 그리고 여타 하위지역의 분할을 5년마다

순차적으로 업데이트하도록 할 것이다.

- <그림 6-10>은 행정자료와 표본조사를 결합하여 구축하는 이탈리아 정보시스템에서 입력자료와 산출자료의 관계를 보여준다. 정보시스템의 구축이 성공하기 위해서는 통계적 단위가 다르고 다양한 항목이 포함되어 있는 자료를 효과적으로 통합하는데서 시작하기 때문에, 행정자료원에서 수집된 모든 자료는 "Integrated Micro-data Repository"라 불리는 공통환경 안으로 직접 또는 간접으로 유입된다. 자료원 간의 자료연계(record linkage)와 개념통합(conceptual integration)은 이 프레임 안에서 이루어지도록 되어 있다.
- "Integrated Micro-data Repository"의 입력자료들 중 대표적 역할을 하는 것은 우리나라의 주민등록부에 해당하는 자치단체 등록부 자료이다. 이 자료는 다른 자료원보다 행정목적이나 통계목적에 적합성이 큰 자료이다. 이 자료는 자치단체에 중요한 법적 효력을 갖는 인구대장에 대한 특별 규정에 의하여 관리된다. 이 때문에, 인구대장 자료는 "Integrated Micro-data Repository"에 통합되기 전에 인구정보시스템을 통하여 일차적으로 처리되도록 하고 있다. 이러한 일반적 프레임웍 안에서, 통계조사는 품질관리 조사로 활용하도록 되어 있다. 곧 C-sample은 인구등록부 포괄범위 오류(coverage error)를 점검하며, D-sample은 항목별 정보를 개선하는 도구로 사용하도록 할 예정이다.
- 인구정보시스템에서 핵심 역할을 하는 또 다른 등록부는 전국도로/주소명부로, 여기에는 지리좌표가 부여된다. 이 명부는 인구정보시스템에 포함된 통계단위를 현장조사에서 파악하기 위한 핵심자료이다. 전국도로/주소는 통합시스템을 개발하고, 지도 작성을 통하여 현재 계획되고 센서스를 관리하는데 없어서는 안 되는 핵심적 요소이다.
- 이탈리아 통계청은 센서스 지도 작성을 위하여, 전국도로/주소명부에서, 장소, 센서스구역, 조사구역 정보를 업데이트하여, 순환센서스의 기본요건을 충족시키도록 노력할 예저으로 되어 있다. 초음파 장치도구와 같이, 이탈리아 통계청

은 지속적으로 전국도로/주민번호(National Roads/Civic Numbers) 자료를 업데이트할 것이며, 이 문제를 해결하는데 2011년 인구주택총조사 지원기관들과 긴밀하게 협조할 예정으로 되어 있다. 통계목적에 위한 인프라로 구축된 토지이용정보는 이탈리아의 지리공간을 파악하는데 대단히 중요하기 때문에, 정보의 집합창고로 보완되어 필요시 접근할 수 있도록 할 예정이다.

(C-sample과 D-sample)

- 두 개의 기본적인 주제 시스템(thematic system)은 행정자료원(administrative data source)을 통합하고, 관리목적의 조사와 정보개선을 위한 조사를 이용하여, 정합성이 있는 통계를 작성하는데 이용하도록 되어 있다. 이 모델에서, C-sample과 D-sample로 언급되는 표본조사에서 특정의 역할이 부여된다. C-sample은 행정자료원의 포괄범위를 통계적으로 보완하는 것이 주목적이고, D-sample은 센서스 조사항목으로 행정자료원에 없는 정보를 제공하기 위한 것이다.
- 이탈리아 통계청(ISTAT)은 2013-2015년 전략수립계획(Strategic Plan)에서, 2017년 5년 주기의 제1차 순환센서스(rolling census)를 착수하여, 유럽연합 규정(EU regulation)에 의거하여 센서스 통계를 작성해야하는 2021년 센서스 실시예에 이르기까지 통합공간자료(integrated spatial data)의 실행을 가능하게 하는데 필요한 자원을 충분히 보유하고 있다고 판단한다.
- 이탈리아 2021년 센서스에서, 센서스 운영의 2가지 목표를 별도로 달성하기 위하여 C-sample과 D-sample을 정하여, 두 종류의 서로 다른 표본조사를 실시하게 될 것이다. C-sample은 인구/가구 통계를 작성하고, D-sample은 사회경제 통계를 작성하는데 사용할 예정이다.
- 두 개의 표본조사는 범위가 특징에 있어서 서로 차이가 있도록 설계할 것이다. C-sample 조사는 등록부에 포함된 인구통계 항목(성별, 연령, 혼인상태, 국

적) 등을 포함하며, 매년 인구구조를 추정하고, 보완하는데 사용한다. D-sample 조사는 국내정책의 우선순위와 국제 비교목적에 충실하도록, 개인/가구의 사회경제통계를 작성하도록 설계된다. 곧, D-sample 조사는 유럽연합 규정에 의거하여, 인구대장에 포함되지 않은 핵심항목 자료를 수집하고, 나아가 통계이용자의 요구를 충족시켜, 10년 주기로 실시되는 전통적 센서스의 정확성이 빠른 속도로 줄어드는 폐단을 제거하고, 자료업데이트는 빈번하게 하기 위한 것이다.

- C-sample 조사는 지역(area)을 추출틀로 사용하며, 개별 자치단체 인구대장의 과소범위(undercoverage)와 과다범위(overcoverage)를 측정하여, 인구통계의 보정요인(correction factor)을 계산하는데 사용한다. 등록부에는 해당 자치단체에서 통상적으로 거주하지 않는 사람을 포함하며(“과다범위”에 해당함), 때로는 해당 자치단체에 통상적으로 거주하지만 등록되지 않은 사람을 포함한다(“과소범위”에 해당함). 특히, 이탈리아에서는 외국인의 유입이 증가하고 있기 때문에, 이로 인하여 발생하는 오류를 추정하는데 필요한 통계적 방법을 개발할 필요가 있는 것으로 파악되고 있다. 이들 추정치는 이원시스템 추정법(또는 “capture-recapture” method라고 불림)으로 작성될 것이다.
- 표본으로 선정된 조사구 또는 주소명부에 대하여, 현장조사에 short form 조사표를 이용하여 전수조사를 실시한다. 조사는 어떠한 행정적 활동과 병행하여 실시되어서는 안 되도록 할 것이며, 해당 자치단체에 거주하지 않는 조사원들에 의하여 현장조사가 실시되도록 해야 마땅하다고 본다.
- C-sample 조사는 매년 자치단체(LAU2)와 자치단체 하위지역(센서스구)에 대하여 인구통계 추계치를 정확하게 작성할 수 있도록 설계할 것이며, 2016년에 그 결과가 최초로 공표될 것이다.
- 센서스 연도(가령, 2021년, 2031년 등)에는 C-sample 조사의 special wave를 이용하여, 해당 자치단체의 법정인구(de jure population)를 작성하도록 할 예정이다. 따라서 2021년에는 C-sample 조사에 의하여 보완된 지방자치단체의 법정인구를 처음으로 작성하여야 할 것이다.

- 인접시점의 통계자료를 누적하여 얻게 되는 각종 추정치에서 효율성이 증대됨으로 해서, 상이한 지리적 단위의 소지역에 대해서 주기적으로 품질이 양호한 추정치 작성이 가능해질 것이다. 이탈리아 통계청은 D-sample 조사에는 long-form 조사표를 이용하여, 사회경제통계를 작성할 계획으로 되어 있다.
- D-sample 조사에서 수집될 항목을 확인하기 위하여, 이탈리아 통계청(ISTAT)은 등록부에 포함된 사회경제자료에 관한 예비조사를 실시할 계획으로 되어 있다. 예비조사의 목적은 통계조사에 의하여 수집될 수 있는 자료를 입수 가능한 행정자료 정보로 대체하기 위한 것이다. 이것은 2021년 센서스를 위하여 이미 시작된 행정자료원에 대한 심층적 연구를 강화할 필요가 있음을 말해주는 것이라고 할 수 있을 것이다.
- D-sample 조사는 매년 추정치를 정확하게 작성하도록 설계할 예정이다. 전국 수준과 NUTS1 지리공간 수준에서, 2017년부터 매년 수집되는 표본에 의하여 추정치를 작성할 수 있을 것이다. 한편, 소지역 통계를 작성하기 위해서는 인접연도의 표본을 통합해야 한다. 생활권지역(NUTS2)와 광역자치단체(NUTS3)에 대해서는 3개의 인접연도 표본을 통합하여 추정치를 작성할 수 있으며, 2018년부터 해당 통계를 작성하여 2019년에는 공표할 예정으로 되어 있다. 자치단체(LAU2)와 자치단체 하위지역(센서스구)에 대해서는 5개의 인접연도 표본을 통합하여 추정치를 작성할 수 있으며, 2019년부터 작성하여 2021년에 처음으로 공표할 예정으로 되어 있다.
- D-sample 조사는 C-sample 조사의 인구집계와 인구구조 결과를 바탕으로 하여 실시될 것이다.
- 2단계 표본추출(two-stage sampling)을 실시하며, 자치단체를 1단계 추출단위로 하고, 가구를 2단계 추출단위로 한다. 2만 명 이하의 자치단체는 매년 추출대상이 되는 것이 아니라, 5개의 그룹으로 구분하여, 5년마다 한 번씩 추출될 수 있도록 하는 “wave”로 편성하게 될 것이다. 2만 명 이상의 자치단체는 509개인데, 매년 추출대상이 되도록 할 예정이며, 추출된 자치단체에 대해

서는 매년 가구표본이 추출될 것이다.

- 제1차 웨이브는 2017년에 시작될 것이며, 각각의 웨이브는 최저 60만 가구, 최고 90만 가구를 조사하는 시스템이 될 것이다. 현재 고려되는 또 하나의 옵션은 2017년과 2021년의 두 웨이브에서만 집중적으로 자료를 수집하고, 이 경우 각각 최저 150만 가구와 최대 220만 가구를 조사대상에 포함시키도록 하는 것이다.
- 이탈리아 통계청은 소지역 단위(자치단체와 자치단체 하위지역)에 대하여 표본추출률 때문에 발생하는 오차의 문제는 빈번한 주기로 통계정보를 작성함으로써 어느 정도 상쇄될 수 있다고 판단하고 있다.
- 이탈리아 통계청은, 자치단체 인구대장(LAC)에서 조사대상 가구를 추출하여 별도 명부를 작성하면, 운영상의 부담을 경감하기 위하여 2011년 라운드에 채택된 몇가지 옵션을 이용할 수 있을 것이라고 본다. 가령 (1) 우편방식의 조사표 배포 (2) 인터넷 반송, 우편반송, 자치단체 행정사무소 조사표 반송부서 설치 등 다채널 방식의 자료수집 (3) 무응답자에 대한 현장조사원 파견 (4) 인터넷 조사 관리 시스템(SGR: Web survey management system)의 이용 등은 2011년 이탈리아 통계청이 2011년 센서스에 도입된 조사방식의 혁신사례에 해당한다.

제4절 최근 동향의 우리나라 2015년 등록센서스에 대한 합의

- 우리나라의 경우 다른 나라와 비교할 때, 2015년에 행정자료와 전수집계를 동시에 이용하는, 이른바 “Register-Assisted Census”를 실시하지 않고, 비교적 짧은 시간에 전통적 센서스로 등록부를 주축정보로 하고, 등록부에 없는 조사항목을 표본조사를 실시하는 새로운 “결합센서스”방식으로 전환에 성공한다면, 그것은 대단히 이례적인 성공사례에 속하게 될 것이다.
- 2010년 라운드와 2020년 라운드의 등록센서스 실시(예정) 동향을 검토하여 보면, 전통적 센서스에서 등록부를 기반으로 하는 센서스로 전환하는 과정에서 다양한 어려움에 직면할 수 있음을 알 수 있다. 이 때문에, 스페인이나 독일과 같은 경우, 등록부 중에서 건물/주택대장의 포괄범위나 기재내용의 불비(不備)로 주택부문에 대해서 전수조사(스페인: 현장실사 독일: 우편조사)를 실시하고, 이를 기반으로 새로운 등록부를 구축하는 방식을 채택하기도 하였다. 한편, 독일의 경우는 통계법이나 등록센서스법의 규정 때문에, 센서스 목적으로 구축된 건물/주택대장도 영구적인 등록자료로 이용하지 못하고, 센서스 실시 이후 일정기간 이내에 폐기처분을 해야 하는 어려운 상태에 있음을 직시할 필요도 있을 것이다.
- 대부분의 등록센서스 실시국가에서는 등록부를 이용하여, Pre-Census Database 곧 통계목적의 명부를 작성하여, 이것은 전반적으로 전통적 센서스와 마찬가지로, 개별성(individuality)의 원칙과 보편성(universality)의 원칙을 준수하는 방식으로 구축된다. 또 이 등록자료는 센서스 모수를 추정하기 위한 기초자료로서는 물론, 표본조사의 추출틀(sampling frame)로 이용되고 있다.
- 등록센서스를 준비하는 국가에서, 등록부의 포괄범위나 기재내역에 대한 확신이 서지 않는 경우에는, 등록부의 포괄범위나 기재내역의 정확성을 점검하기 위하여 전수조사를 실시하는 경우도 있다. 물론, 이것은 빠른 시간 안에, 등록부가 품질요건을 충족한다면, 부분 등록센서스로 이행하고 나아가 순환 등록센서스 또는 완전 등록센서스로 이행하기 위한 절차라고 볼 수 있다.

- 영국은 행정자료 기반 센서스 준비상황은 대단히 특이하다. 주민등록이나 건물/주택대장이 없는 상태에서, 2011년 이후의 센서스에서 전통적 센서스를 폐기하는 방향으로 나아가고 있으며, 이것은 우리에게도 대단히 중요한 교훈을 주고 있다. 그 이유는 행정자료가 대단히 부족한 상황에서, 국민건강보험 환자 행정자료나 선거인 명부를 인구부문 자료를 추출하기 위한 행정자료로 사용하고 있으며, 이를 위하여 행정자료 옵션으로 비교적 간단한 집계자료 옵션 (aggregate data option)으로 전환할 가능성도 제시하고 있기 때문이다. 물론, 이러한 경우에도 영국 통계청은 포괄범위 점검조사를 과다범위와 과소범위 두 가지 면에서 실시할 계획으로 되어 있다. 또 이스라엘이나 대부분의 등록센서스 실시국가들과는 달리 포괄범위 목적이 아닌 별도의 통계조사를 실시하여, 종전의 전통적 센서스에서 획득하였던 사회경제통계 항목에 대한 정보를 획득할 계획으로 되어 있다.
- 이스라엘은 2008년 통합센서스의 성공을 계기로 하여, 빠른 속도로 순환 통합센서스로 나아가 있다. 특히, 주택대장을 개발하고, 교육대장, 소득대장에 대한 별도의 대장을 추가적으로 구축하여, 궁극적으로 완전 등록센서스를 목표로 하는 방향으로 나아가고 있다. 당초에 우리나라의 2015년 등록센서스가 이스라엘의 센서스 추정방법론을 벤치 마크할 것으로 예상되었지만, 과다범위와 과소범위를 별도로 추정하지 않고, 행정자료와 조사자료의 차이에 초점을 맞추어, 표본설계를 적정화하고, 릿지회귀추정이나 소지역추정 방법론을 적용하고자 노력하고 있지만, 여전히 우리나라의 등록센서스도 매년 추정인구를 산출하는 것을 목표로 하기 때문에, 이스라엘로부터 여전히 배워야 할 것이 많다고 생각한다.
- 이탈리아의 센서스 방법론도 대단히 빠른 속도로 진전하고 있다. 2001년 센서스에서는 행정자료와 전수집계를 조합하는 방식으로 센서스를 실시하였지만, 2011년 센서스에서는 조사표를 short-form과 long-form으로 구분하여 실시하여, 2021년 센서스에서 순환 등록센서스로 이행할 준비를 하고 있다. 2021년 순환센서스에서 등록부 기준 과다와 과소를 점검하는 C-sample을 이

용한 조사와, 사회경제통계를 얻기 위한 D-sample 조사를 구분할 계획이다. 이것도 중장기적으로, 우리나라의 2015년 등록센서스가 연차적으로 인구통계를 생산하고, 필요한 경우도 동시에 사회경제통계를 작성하기를 원한다면, 반드시 벤치 마크해야 할 바람직한 전환방향이라고 판단된다.

제7장 주요국의 추정방법론과 우리나라 2015년 등록센서스의 당면과제

- 센서스 방법론의 전환에는 행정자료의 완전성 문제를 포함하여 과학적 문제, 특히 표본추출과 통계적 추정문제는 물론, 개인의 사생활 보호에 따른 센서스의 정치화 현상 등 무수히 극복해야 할 과제들이 있다. 2012년 통계청 등록센서스 학술연구 용역보고서의 마지막 장에서는, 등록부의 기재내용을 주축정보로 하여 센서스를 실시한 나라들이 참고해야 할 유럽연합(EU)이 정한 센서스 관련 규정, 특히 센서스 목표모집단 관련 사항과 센서스 추정방법론에 대한 다양한 접근방식을 소개하는 것이 목적이다.
- 본 장에서 센서스 추정방법론에 대해서는 해당 국가의 통계청 웹사이트나 관련문서에 보고된 내용을 기초로 할 것이지만, 센서스 추정방법론의 세부사항은 사실상 등록센서스 실시국가 간에 큰 차이가 없고, 해당 국가의 행정구역이 얼마나 복잡하며, 센서스 통계작성의 상세성과 깊은 관련이 있다. 검토결과는 우리나라 등록센서스 추정방법론이 등록센서스 실시국가는 물론 유럽연합의 국제적 기준과 비교하여 어떠한 특성을 갖는가에 대한 일반적 아이디어를 제공하는데 이용될 것이다.

제1절 EU 센서스 규정의 목표모집단 관련사항

- 유럽연합 통계청(Statistical Office of the European Community)은 센서스 관련 규정에서 등록센서스를 실시하는 경우, “등록부 자료원에 대한 보고서는 목표모집단(target population)에 대하여 기대되는 정보량과 등록부 자료

원에서 실제로 이용 가능한 정보량 간 매칭/연계(matching/linking)의 정도를 포함하여야 마땅하다”고 정하고 있다.

- 요약하면, 센서스 보고서는 (1) 공표된 자료가 어느 정도 센서스 모집단의 실제크기(true size)를 반영하는가? (2) 공표된 자료가 실제관찰을 기준으로 작성한 것이 어느 정도인가? (3) 행정자료원의 기록들 중 어느 정도가 센서스 목적에 이용되었는가? 등의 질문에 대한 해답을 포함해야 한다고 본다. 곧 이 질문에 대한 해답은 행정자료의 포괄범위와 임putation 자료와 삭제된 자료에 대한 전반적 평가를 통하여 주어질 수 있다.
- 유럽연합 센서스 규정 No. 1151/2010(EU Regulation No 1151/2010) 부칙 II (Annex II)은 행정자료의 포괄범위와 임putation 자료와 삭제된 자료에 대한 전반적 평가를 토대로 하여, 유의미한 방식으로 해석될 수 있는 일련의 지표를 요구하고 있다. 이 지표의 세트는 (1) 추정된 목표모집단(estimated target population)과 센서스 모집단(곧, 포괄범위 오차에 의하여 영향을 받은 모집단)의 정합성 정도를 보여주고, (2) 레코드 임putation(imputation)과 삭제(deletion)의 영향을 보여주고, (3) 센서스 모집단을 목표모집단에 근접시키기 위하여 자료 레코드에 가중치를 조정하는 기법 또는 방법론의 영향을 보여주고, (4) 센서스 모집단의 표본을 추출틀로 사용한 경우, 추출틀의 영향을 보여주는 것이라야 한다.

1. 용어의 정의

- 목표모집단(target population)은 통계단위(예: 사람, 가구, 주택/거처)의 이상적인 값 또는 참값에 해당하는 모집단이다. 이것은 이론적 개념(theoretical concept)로서, 센서스가 통계단위를 누락이나 중복없이 집계하여 100% 포괄범위를 달성하여 보고해야 할 것을 말한다. 목표모집단은 하나 이상의 특정 항목에 대하여 통계표를 작성하는데 사용되어야 할 모든 통계단위의 전체집합을 의미하며, 이들 통계단위는 특정지역의 범위 안에 있는 모든 단위를 포함

해야 하기 때문에, 센서스의 기본원칙인 보편성(universality)의 원칙이 적용된다. 또 목표인구는 누락(omission)이나 중복(duplication)의 오류를 범하지 않고, 유효한 모든 통계단위를 오직 한 번씩만 집계하여야 한다. 센서스 집계 결과는 목표인구의 모든 통계단위에 대한 정보를 바탕으로 집계하는 것이 이상적이다.

- 목표모집단 추정치(estimated target population)는 (이상적인) 목표모집단에 대한 최고의 입수 가능한 추정치이며, 이 추정치는 포괄범위 평가 작업에 의하여 확인된 오류를 보완하여 산정한 센서스 모집단(census population)을 기준으로 하여야 한다. 목표모집단 추정치에 대해서는 대개 핵심수치만 입수 가능한데, 그 예로는 성.연령(5세 간격)별 인구가 있다. 유럽연합 규정 번호 1151/2010의 요건을 충족시키기 위해서, NUTS2 수준에서는 목표모집단의 총 규모에 대한 추정치는 입수 가능해야 한다.
- 유럽연합(European Union)의 지역통계단위(NUTS: Nomenclature of Units for Territorial Statistics)는 기존의 국가별 행정구역분류 체계를 기반으로 한다(<표 7-1> 참고). 지역수준의 행정구역이 2개 이하이거나, 기존 행정구역의 크기가 너무 작은 경우에는, 제2수준 또는 제3수준이 될 수 있다. 이것은 제1수준(예: 프랑스, 이탈리아, 그리스, 스페인), 제2수준(예: 독일) 또는 제3수준(예: 벨기에)이 될 수도 있다. 소국의 경우(예: 룩셈부르크, 사이프러스, 아일랜드)는 국가 전체가 NUTS2 또는 NUTS 3 수준까지 하나만 존재하여, 수준 1, 2, 3이 전국이 똑 같은 경우가 있다.
- 유럽연합은 통계단위지역(NUTS)를 설정하면서 제1수준은 3백만~7백만 명, 제2수준은 80만~3백만 명, 제3수준은 15만~80만 명을 기준으로 하고 있다. 물론 이러한 인구규모를 이용하여, 단위통계지역을 설정하기 위한 기준으로 사용하지만, 이들은 엄격하게 적용되지 않는다. 가령, 프랑스에는 인구수가 엄청난 차이가 (가령 인구가 3만 명인 지역과 인구가 120만 명인 지역이 모두 NUTS 2 지역으로 설정되는 경우가 있다).

<표 7-1> 유럽지역 NUTS(Nomenclatures of Units for Territorial Statistics) 구분

Countries		NUTS1		NUTS 2		NUTS 3	
유럽연합 회원국	27		97		271		1303
오스트리아	AT	Groups of states	3	States	9	Groups of districts	35
벨기에	BE	Regions	3	Provinces (+ Brussels)	11	Arrondissements (Verviers split into two)	44
불가리아	BG	Regions	2	Planning regions	6	Oblasts	28
사이프러스	CY	—	1	—	1	—	1
체코	CZ	—	1	Oblasts	8	Regions	14
독일	DE	States	16	Government regions (or equivalent)	39	Districts	429
덴마크	DK	—	1	Regions	5	Lands	11
에스토니아	EE	—	1	—	1	Groups of counties	5
스페인	ES	Groups of autonomous communities	7	17 Autonomous communities and autonomous cities	19	Provinces + Islands + Ceuta and Melilla	59
핀란드	FI	Mainland Finland, Åland	2	Large areas	5	Regions	20
프랑스	FR	Z.E.A.T. + DOM	9	Regions+ DOM	26	Departments+ DOM	100
그리스	GR	Groups of development regions	4	Peripheries	13	Prefectures	51
헝가리	HU	Statistical large	3	Planning and statistical	7	Counties + Budapest	20

		regions		regions			
아일랜드	IE	—	1	Regional Assemblies	2	Regional Authorities	8
이탈리아	IT	Groups of regions	5	Regions (Trentino-Alto Adige/Südtirol split into two)	21	Provinces	107
리투아니아	LT	—	1	—	1	Counties	10
룩셈부르크	LU	—	1	—	1	—	1
라트비아	LV	—	1	—	1	Statistical regions	6
몰타	MT	—	1	—	1	Islands	2
네덜란드	NL	Groups of Provinces	4	Provinces	12	COROPregions	40
폴란드	PL	Regions	6	Voivodeships	16	Subregions	66
포르투갈	PT	Continental (+ Azores and Madeira)	3	Regional Coordination Commissions + Autonomous regions	7	Groups of municipalities	30
루마니아	RO	Macro regions	4	Regions	8	Counties + Bucharest	42
스웨덴	SE	Regions	3	National areas	8	Counties	21
슬로베니아	SI	—	1	Macroregions	2	Statistical regions	12
슬로바키아	SK	—	1	Oblasts	4	Regions	8
영국	UK	Regions of England	9	Counties (some grouped); Inner and Outer London	30	Upper tier authorities or groups of lower tier authorities (unitary authorities or districts)	93
		Wales	1	Groups of	2	Groups of	12

				u n i t a r y authorities		u n i t a r y authorities	
		Scotland	1	Groups of u n i t a r y authorities or districts	4	Groups of u n i t a r y authorities or districts	23
		Northern Ireland	1	—	1	Groups of districts	5
비준국	1		1		3		21
크로아티아	HR	—	1	Regions	3	Counties	21
후보국	5		14		34		121
아이슬란드	IS	—	1	—	1	C a p i t a l area/Rest of country	2
마케도니아	MK	—	1	—	1	Statistical regions	8
몬테네그로	ME	—	1	—	1	—	1
세르비아	RS	Groups of Regions	2	Regions	5	Districts	29
터키	TR	Regions	12	Sub-regions	26	Provinces	81
EFTA 회원 국	3		3		15		46
리히텐슈타인	LI	—	1	—	1	—	1
노르웨이	NO	—	1	Regions	7	Counties	19
스위스	CH	—	1	Regions	7	Cantons	26

- 목표모집단 추정치(estimated target population)는 목표모집단에 대하여 입
수 가능한 최고의 근사치(approximation)를 의미한다. 목표모집단의 크기는
센서스 모집단의 크기에 과소범위(under-coverage)의 크기를 보태고, 과다범
위(over-coverage)의 크기를 제거하여 추정한다.
- 센서스 모집단은 센서스 자료, 등록센서스의 경우에는 행정자료가 실제로 반
영하는 모집단을 가리킨다. 센서스 모집단은 개별국가의 통계작성기관이 특정
항목에서 대하여 최고품질이라고 판단하는 센서스 자료(census data)의 원천
(source)에 반영되어야 한다. 센서스 모집단에 포함되는 데이터 레코드(data

record)는 임putation으로 생성된 레코드("imputed records")를 포함하고 삭제된 레코드("deleted records")를 제외해야 한다.

- 센서스 모집단은 특정의 목표모집단 대신에 1개 이상의 항목에 대하여 센서스 결과로 실제로 대표되는 통계단위(statistical units)의 집합이다. 센서스 모집단에 대한 데이터 레코드(data records)는 특정의 목표모집단의 자료원(행정 등록부, 가령 주민등록, 주택대장)에 들어있는 데이터 레코드를 가리킨다.
- 사람들(persons)에 관한 모든 자료원은 정의상, 목표모집단으로서 그 크기를 가지며, 실제로 목표모집단의 크기에 대한 추정치가 동일하여야 한다. 또, 센서스 모집단의 크기는 동일하여야 마땅할 것이다. 그러나 실제로 센서스 모집단의 크기는 조사항목별로 달라지게 마련인데, 그 이유는 조사항목별로 자료의 품질이 완전성에 있어서 차이를 보이기 때문이다.

2. 데이터 레코드의 임putation과 삭제

- 인구주택총조사의 기본적 특성에 들어 있는 “개별집계”(individual enumeration)와 “정해진 영토 내의 보편성”(universality within a defined territory)의 두 가지 원칙은 행정자료, 곧 등록부(예: 주민등록부, 건축물/주택대장)는 목표모집단의 모든 통계단위에 대하여 정확하게 하나의 데이터 레코드(data record)를 포함하는 것이 이상적이라는 것을 의미한다. 그러나 실제로 이것은 불가능한 경우가 많으며, 행정목적의 등록부에는 데이터 레코드가 누락되어 있거나 너무 많은 레코드가 들어 있는 경우가 있다.
- 이 문제를 해결하기 위하여, 해당국가의 통계작성기관은 대체 레코드를 임putation을 하거나 부적격인 레코드를 삭제하는 조치를 취할 수 있다. 다만, 유럽연합 규정(EC Regulation) No. 763/2008는 회원국들이 센서스 결과표를 작성할 때, “등록센서스의 등록부와 방법론이 센서스의 기본적 특성을 충족시키는 정도”(extent to which the chosen data sources and methodology meet the essential features) (제6조 제2항)에 대한 보고서를 제출할 것을

의무화하고 있다. 여기서 “정도”(extent)는 데이터 레코드의 임퓨테이션 건수와 삭제 건수에 따라 달라질 것이다.

3. 포괄범위 평가작업

- 센서스 모집단은 아무리 엄격한 방법으로 임퓨테이션에 의하여 레코드 수를 늘리고, 삭제에 의하여 레코드를 줄인다고 하더라도, 이상적인 목표모집단을 대부분 적중(target)시키지 못한다. 유럽연합 회원국들은 포괄범위(coverage)를 평가하여, 전체 모집단(사람)의 과소범위와 과다범위를 추정할 것을 의무화하고 있지만, 그렇다고 해서 평가방법을 법제화하거나 평가결과의 합격점을 특별히 정하고 있지는 않다.
- 물론, 센서스에서 준수되어야 할 보편성의 원칙에서, 완전성(completeness) 곧 포괄범위에 심각한 문제가 생기게 되면, 그것은 인구주택총조사의 품질은 심각한 손상을 입게 된다. 이 때문에, 유럽연합 회원국은 등록부 자료의 과다범위와 과소범위를 별도로 추정하여 보고할 필요가 있다.
- 과소범위가 높은 수준인 경우에도, 그것이 과다범위가 높은 수준에 있는 것에 의하여 보상될 수 있다. 그러나 높은 수준의 과다범위는 공표되는 자료가 현실적인 관찰, 곧 이용 가능한 관찰을 반영하지 못하는 상황을 의미한다. 가령, 과다범위가 100%이고 과소범위가 100%인 경우에는 센서스 모집단에 내재하는 엄청난 품질문제를 은폐하게 될 가능성이 크다.
- 과소집계와 과다집계의 정도에 대한 정보는 추정치에 근거한 것으로, 각종 오류의 영향을 배제할 수 없다. 이 때문에, 센서스 모집단과 목표모집단의 추정정치 간에 정합성에 대한 완전한 해답을 얻을 수는 없다. 이 때문에, 유럽연합 규정 No. 1151/2010은 과소/과다범위 추정치의 품질정보는 물론, 과소/과다범위를 평가하는데 사용된 각종 조사와 방법론에 관한 메타데이터

(metadata)를 제공할 것을 요구하고 있다.

- 등록센서스는 물론 여타 센서스에 적용되는 목표모집단, 센서스 모집단, 추정 목표모집단의 개념을 정리하면 다음과 같다.

- 목표모집단(Target population): 기준시점 (reference date)에서, 정해진 지리적 지역 내에서 센서스 조사항목에 해당하는 모든 통계단위(statistical units)(예: 인구, 가구, 주택)의 집합. 목표모집단은 유효한 통계단위를 누락(omission), 중복(duplication) 없이, 오직 한번만을 포함하는 것이어야 한다. 이들 통계단위는 특정 지역의 범위 안에 있는 모든 단위를 포함해야 하기 때문에, 센서스의 기본원칙인 보편성(universality) 원칙이 적용된다.

- 센서스 모집단(Census population): 특정 목표 모집단에 대하여 1개 이상의 항목에 대한 센서스 결과에 의하여 실제 반영되어야 할 통계단위의 집합. 센서스 모집단의 데이터 레코드(data record)는 특정 목표모집단의 데이터 레코드에 임putation(임putation)에 의한 데이터 레코드를 포함하고, 삭제된 데이터레코드를 제외하여야 한다. 센서스 모집단은 센서스의 기본원칙인 개별성(individual enumeration)과 보편성(universal in a given territory)의 원칙을 준수하여야 한다.

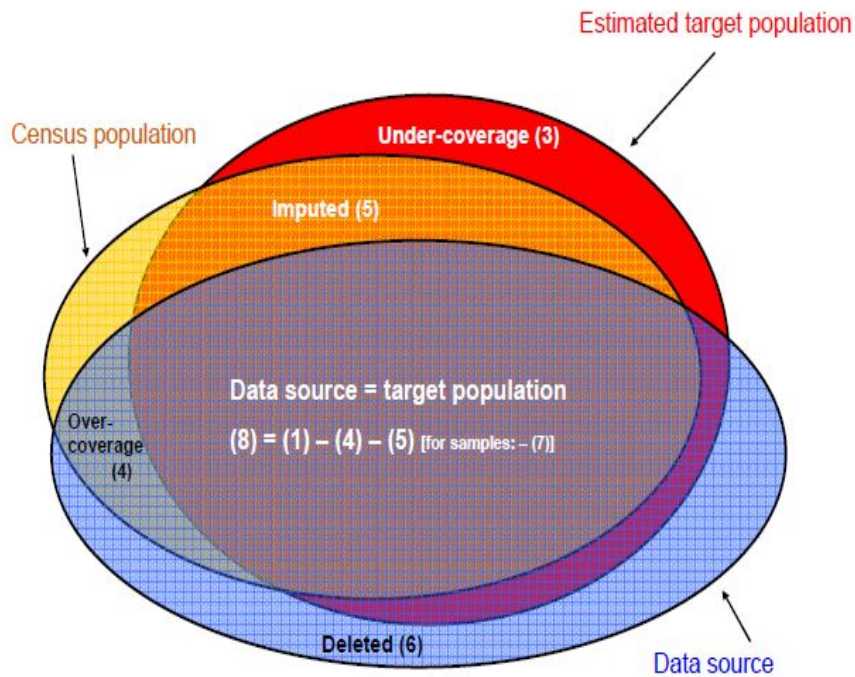
- 추정 목표모집단(Estimated target population): 목표모집단에 근접하는 최고품질의 입수 가능한 수치 추정 목표모집단은 포괄범위 평가작업(coverage assessment)을 통하여, 센서스 모집단(census population)의 크기에 과소집계(UC)의 크기를 보태고, 과다집계(OC)의 크기를 빼서 구한다. 목표모집단은 센서스의 조사항목 모두에 대하여 추정되는 것이 아니라, 일부 항목 가령 성/연령별 인구에 대하여 추정되는 경우가 대부분이다.

- 등록센서스는 물론 여타 방식의 센서스에서 센서스 결과를 평가하는 지표를 유럽연합 센서스 규정은 다음과 같이 정리하고 있다.

(1) 추정 목표모집단과 센서스 모집단(곧, 포괄범위 오차에 의하여 영향을 받

- 은 모집단)의 정합성 정도를 보여주고,
(2) 레코드 임putation(impudation)과 삭제(deletion)의 영향을 보여주고,
(3) 센서스 모집단을 목표모집단에 근접시키기 위하여 자료 레코드에 가중치를 조정하는 기법 또는 방법론의 영향을 보여주고,
(4) 센서스 모집단의 표본을 프레임으로 사용한 경우, 프레임의 영향을 보여주는 것이라야 한다.

<그림 7-1> 목표모집단에 대한 질적 정보



참고: Data Source = target population: 자료원 = 목표모집단

Census Population = 센서스 모집단

Estimated Target Population = 추정 목표모집단

Imputed: 임퓨테이션 Deleted: 삭제

Undercoverage: 과소범위 Overcoverage:

<그림 7-2> 포괄범위의 3가지 경우: 과소문제가 심각한 경우, 과다문제가 심각한 경우, 과소/과다문제가 모두 심각한 경우

		High under-coverage	High over-coverage	High under- and over-coverage
(1)	Census population	700	1300	1000
Q 1	% of estimated target population	70%	130%	100%
(2)	Estimated target population	1000	1000	1000
(3)	Under-coverage	300	0	300
	% of census population	43%	0%	30%
(4)	Over-coverage	0	300	300
	% of census population	0%	23%	30%
(5)	Imputed records	0	0	0
	% of census population	0%	0%	0%
(6)	Deleted records	0	0	0
	% of census population	0%	0%	0%
(7)	Complementary set of units	0	0	0
(8)	Data source = target population	700	1000	700
Q 2	% of census population	100%	77%	70%
Q 2 * Q 1	% of estimated target population	70%	100%	70%
Q 3	% of records in the data source	100%	77%	70%
	Net coverage	70%	130%	100%

- <그림 7-2>는 포괄범위의 3가지 경우, 곧 (1) 과소문제가 심한 경우(high under-coverage) (2) 과다문제가 심한 경우(high over-coverage), (3) 과소문제와 과다문제가 모두 심한 경우(high under- and over-coverage)를 보여준다.
- (1)의 경우는 300명이 과소집계되었고 (2)의 경우는 300명이 과다집계되었으며, (3)의 경우는 300명이 과소집계되었을 뿐만 아니라 300명이 과다집계되었음을 말해준다.

- 추정목표인구와 센서스 인구와의 정합성(Q1): 추정 목표인구의 크기는 3가지 경우 모두 1000명이지만, 센서스 인구는 (1) 700명 (2) 1300명 (3) 1000명으로, 센서스 인구는 추정 목표인구의 70%, 130%, 100%를 집계하고 있음을 말해준다.
- 삭제와 임putation의 영향(Q2): 자료원에는 목표인구의 크기가 (1) 700명 (2) 1000명 (3) 700명이다. 이들 목표인구는 센서스 인구의 100%, 77%, 70%에 각각 해당하고, 추정 목표인구의 70%, 100%, 70%에 각각 해당한다(Q1+Q2). 또 자료원에는 개체 레코드가 센서스 인구의 100%, 77%, 70%가 포함되어 있다(Q3). 순포괄범위(net coverage)는 (1) 70% (2) 130% (3) 100%이다.

<그림 7-3> 과소범위 문제가 심각한 경우와 임퓨테이션 문제가 심각한 경우

		High under-coverage	High imputation ("number-1-census")
(1)	Census population	700	1000
Q 1	% of estimated target population	70%	100%
(2)	Estimated target population	1000	1000
(3)	Under-coverage	300	0
	% of census population	43%	0%
(4)	Over-coverage	0	0
	% of census population	0%	0%
(5)	Imputed records	0	300
	% of census population	0%	30%
(6)	Deleted records	0	0
	% of census population	0%	0%
(7)	Complementary set of units	0	0
(8)	Data source = target population	700	700
Q 2	% of census population	100%	70%
Q 2 * Q 1	% of estimated target population	70%	70%
Q 3	% of records in the data source	100%	100%
	Net coverage	70%	100%

- <그림 7-3>은 (1) 포괄범위에서 과소문제(undercoverage problem)가 심각한 경우와 (2) 임퓨테이션(imputation problem) 문제가 심각한 경우로 구분한 결과를 보여주고 있다.
- (1)과 (2)의 경우, 모두 추정 목표인구는 1000명다. 그러나 (1)의 경우는 300명이 과소집계되었고, (2)의 경우는 과소나 과다로 집계되지 않았지만, 300개의 임퓨테이션 레코드가 있다. 센서스 인구는 (1)의 경우 700명, (2)의 경우 1000명으로, 각각 추정 목표인구의 70%와 100%에 해당한다(Q1).
- 자료원에 포함된 목표인구는 (1)과 (2) 모두 700명으로, 센서스 인구의 100%와 70%에 해당한다(Q1). 또, 자료원에 포함된 목표인구는 추정 목표인구의 70%와 70%이며(Q1+Q2), 자료원에 들어있는 개체 레코드는 센서스 인구의 100%와 100%가 각각 포함되어 있다(Q3). 순포괄범위(net coverage)는 (1)의 경우는 70%이고 (2)의 경우는 100%이다.

<그림 7-4> 포괄범위는 문제없지만, 삭제된 레코드가 많은 경우



- 포괄범위에는 문제가 없지만, 삭제된 레코드가 많은 경우("many invalid records")를 <그림 7-4>가 보여준다. 이 경우는 센서스 인구, 추정된 목표인구는 같고, 과다범위나 과소범위의 문제는 없지만, 삭제된 기록이 센서스 인구의 30%에 이른다.

<그림 7-5> 자료원이 표본인 경우

		Source is sample
(1)	Census population	1000
Q 1	% of estimated target population	100%
(2)	Estimated target population	1000
(3)	Under-coverage	0
	% of census population	0%
(4)	Over-coverage	0
	% of census population	0%
(5)	Imputed records	0
	% of census population	0%
(6)	Deleted records	0
	% of census population	0%
(7)	Complementary set of units	666
(8)	Data source = target population	334
Q 2	% of census population	33%
Q 2 * Q 1	% of estimated target population	33%
Q 3	% of records in the data source	100%
	Net coverage	100%
	Sample size	33%
	Stat. units used for extrapolation	334

- 자료원(data source)이 표본(sample)인 경우의 모습을 <그림 7-5>가 보여준다. 위의 자료에서는 표본자료가 센서스 모집단 크기의 33%에 해당한다. 곧 표본자료를 이용하여 센서스 인구(census population)와 추정 목표인구(estimated target population)를 계산한 것이라고 볼 수 있다.

<그림 7-6> 자료원에 문제가 없는 경우, 레코드 임퓨테이션 문제가 심각한 경우,
항목 임퓨테이션 문제가 심각한 경우

	No sizeable problems	High record imputation	High item imputation
(1) Census population	1000	1000	1000
(2) Number of records with information on the topic % of census population	995 100%	995 100%	995 100%
(3) Number of imputed records with information on the topic % of census population	0 0%	300 30%	0 0%
(4) Item imputation for the topic % of census population	0 0%	0 0%	300 30%
(5) Item no-information for the topic % of census population	0 0%	0 0%	0 0%
(6) Non-imputed observations on the topic % of census population	995 100%	695 70%	695 70%

- (1) 자료원에 문제가 없는 경우, (2) 레코드 임퓨테이션(record imputation)의 문제가 심각한 경우, (3) 항목 임퓨테이션(item imputation)의 문제가 심각한 경우 등 3가지의 센서스 품질과 관련된 상황을 보여주고 있다.

<그림 7-7> 센서스 자료원에 문제가 없는 경우, 해당 항목정보가 없는 레코드가 많은 경우, 누락된 정보를 임퓨테이션한 경우

	No sizeable problems	Many records without information on the item	Some missing information is imputed
(1) Census population	1000	1000	1000
(2) Number of records with information on the topic % of census population	995 100%	300 30%	500 50%
(3) Number of imputed records with information on the topic % of census population	0 0%	0 0%	0 0%
(4) Item imputation for the topic % of census population	0 0%	0 0%	200 20%
(5) Item no-information for the topic % of census population	0 0%	0 0%	0 0%
(6) Non-imputed observations on the topic % of census population	995 100%	300 30%	300 30%

- <그림 7-7>은 (1) 센서스 자료에 문제가 없는 경우(no sizeable problems), (2) 해당 항목 정보가 없는 레코드가 많은 경우(many records without information on the item), (3) 누락된 정보를 임퓨테이션한 경우(some missing information is imputed) 등 3가지의 경우에 대하여 센서스의 품질에 관련된 상황을 보여주고 있다.

4. 센서스 가중치 설정

- 유럽연합 센서스 규정은 개체자료(예: 가구원, 가구, 주택)에 1이 아닌 가중치를 설정할 수 있다고 규정하고 있다. 이것은 (1) 개체자료를 가상적, 비실체적으로 임퓨테이션을 하는 경우 (2) 센서스 모집단의 일부에 대한 가중치를 체계적으로 늘리거나 줄임으로 해서 자료의 크기를 확대하거나 축소하기를 원하는 경우 (3) 자료의 설득력이나 정합성을 개선하기 위하여 자료 가중치를 조정하는 방법을 이용하는 경우들이다.
- 이러한 관행은 센서스의 개별성 원칙이나 보편성의 원칙을 위반할 수 있다. 이 때문에, 이러한 가중치 설정작업은 개체자료의 임퓨테이션이나 삭제가 자료원의 품질측정에 있어서 동일한 효과를 가진다는 사실을 확인해야만 한다.
- 센서스 통계작성에서 개체자료가 1보다 큰 가중치를 설정하여 이용된다면, 그 가중치에서 1을 뺀 값은 임퓨테이션 레코드(imputed records)로 간주되어야 하며, 1보다 작은 가중치를 설정하여 이용된다면, 1에서 그 가중치를 뺀 값은 삭제된 레코드(deleted records)의 건수로 간주되어야 할 것이다.
- 올해의 등록센서스 통계작성을 위한 가중치 설정(본 보고서의 제3장의 연구결과)에서, 대부분의 센서스 가중치는 1보다 작은 것으로 설정되었는데, 1에서 그 가중치를 뺀 값은 주민등록자료나 건축물대장과 같은 목표모집단(target population)에서 삭제되어야 할 레코드의 건수를 의미하는 것으로 해석되어야 마땅할 것이다.

제2절 등록센서스 실시국가의 표본설계와 추정방법론

- 2012년 통계청 등록센서스 연구용역 사업이 표본설계와 추정방법론을 중심으로 하고 있기 때문에, 목표모집단(예: 가구원, 가구, 주택)에 대한 개체자료의 삭제나 임putation 문제를 제외하고, 주로 센서스 추정방법론을 중심으로 해서 살펴본다.
- 등록센서스 실시(예정)국가는 대부분 등록부를 주축정보로 하여 센서스 통계를 작성하기 때문에 등록부의 정확성 곧 포괄범위를 점검하기 위하여, 센서스 모집단의 일부에 대하여 표본조사를 실시하기 때문에, 표본설계와 그에 따른 추정방법론은 등록센서스 운영의 핵심적 과제의 하나이다.
- 대부분의 통계청이 실시하는 통계조사는 조사가중치(survey weights)을 이용하여, 목표모집단의 값을 추정하는 이른바 “설계기반 추정방법”을 선택하기 때문에, 표본설계는 대단히 중요한 절차이다. 센서스에서 예외는 아니다. 그러나 인구학적 센서스 추정방법으로는 전통적으로 인구통계분석(DA: Demographic Analysis) 방식이 사용되거나, 이원시스템 추정법과 같은 “모델 기반 추정방식”(model-based estimation)이 채택되고 있기 때문에 이원시스템 추정의 응용방식을 살펴보는 것은 중요하다(Rao, 2003).
- 통계적 모형은 종속변수(예: 총수, 평균, 비율)로 알려진 일련의 변수를 예측하는데 있어서, 우리가 관심을 갖는 변수와 관련 변수(보조변수, 독립변수)의 관계를 수학적으로 표현한 것이다. 이들 모형은 개별단위에 대해서 종속변수와 보조변수에 대해서 관측된 값을 포함하는 자료에 적합시키도록 되어 있다. 적합과정에서 절편(intercept)과 기울기(slope) 같은 모형 파라미터의 추정치를 계산한다. 여기서 물론, 단위는 개체(개인)가 될 수 있고, 가구가 될 수도 있고, 조사구(ED: Enumeration District) 또는 읍면동 아니면 시군구 등의 소지역 자체도 될 수 있다. 모형은 오차항을 포함하여, 종속변수의 예측값이 실제 관측값과 얼마나 차이가 나는가를 기술하도록 할 수 있다.

- 등록부를 이용하여 센서스를 실시하는 나라들의 경우, 등록부의 자료를 보완하고, 추가적으로 사회경제 통계자료를 획득하기 위하여 표본조사를 실시하는 나라의 경우는 표본설계와 추정방법론의 문제가 중요하다. 물론 통계학적 추정방법에서는 표본설계가 중요하지만, 인구학적 추정방법에서는 그러지 아니한다. 인구학적 추정방법에는 인구균형방정식(Population Equation), 이원시스템 추정방법(DSE) 등이 포함되고, 통계학적 추정방법으로는 직접추정방법(Direct Estimation) (예: HT 추정량, Generalized Regression Estimator 등) 비추정방법(Ratio Estimation), 표준회귀추정방법(Standard Regression Estimation) 등이 있는데, 이들 방법이 혼용하여 사용하는 경우가 많다.

<표 7-2> 등록센서스 표본조사의 설계와 추정방법론

국가	행정자료 품질	표본추출	추정방법론
영국	전수집계	2% 수준(포괄범위조사)	이원시스템추정, 비추정
독일	중간급	10% 수준(센서스 표본조사), 표본조사 표본의 5% 수준(반복조사)	GREG(일반회귀추정), 변형된 이원시스템추정(사실상 비추정)
스페인	중간급	13% 수준(센서스 표본조사)	비추정
이스라엘	중간급	16-17% (센서스 표본조사-과소범위점검조사), 20% (과다범위조사)	이원시스템추정
싱가포르	고급	20% (센서스표본조사)	행정자료를 그대로 이용
대만	중간급	16-17% (센서스 표본조사)	비추정
터키	중간급	12-13% (센서스 표본조사)	비추정
이탈리아	중간급	4-5% (순환 표본조사)	비추정, 이원추정방법
네덜란드	고급	없음	매크로통합 및 반복가중
벨기에	고급	없음	매크로통합 및 반복가중

- 등록센서스를 실시하는 나라에서는 센서스 추정방법론으로, 인구 부문에서 이원시스템 추정방법을 이용하는 경우가 상당수에 이른다. 이원시스템 추정방법을 정확하게 적용하게 위해서는 과다범위와 과소범위를 정확하게 확인할 수 있는 포괄범위 점검조사가 필요하다. 이러한 방식으로 포괄범위를 점검하는

나라에는 2000년 라운드에 전통적 센서스를 실시하였던 미국, 스위스가 있었고, 등록센서스를 실시하는 국가 중에서는 이스라엘이 유일하게 2008년 통합 센서스에서 과다범위와 과소범위를 점검하는 표본조사를 실시하였다.

- 인구통계의 관점에서, 이스라엘의 등록센서스나 미국 센서스 본조사의 사후 조사에서 사용되는 이원시스템 추정방식이 적합하다는 것에 대해서는 두말할 필요가 없다. 두 차례의 조사에서 제1차 조사인 과소범위를 추정하는 경우에는 문제가 적겠지만, 제2차 조사인 과다범위를 추정하기 위한 조사의 경우, 그것이 전화조사로 실시된다면 우리나라의 상황에서 심각한 문제를 야기할 수도 있다. 그럼에도 불구하고, 본 연구는 이스라엘식의 과다범위를 추정하는 문제가 해결되지 않는다면, 해당지역에서 과다범위에 해당되는 주민들에 대한 별도의 조사방법을 현지실사 방식으로 해서라도 개발하는 것이 등록센서스 통계 작성방식에 대한 신뢰도를 올릴 수 있는 방안이라고 판단한다.
- 물론, 이스라엘 방식에 대한 대안적 추정방식은 독일, 스페인, 대만의 2010년 라운드 센서스에서 사용된 비추정법(ratio estimation)을 활용하는 것이다. 이러한 비추정법을 외견상 이스라엘 방식의 대안이라고 할 수 있겠지만, 독일의 방식은 주민대장과 우편조사에 의하여 확인된 주소명부(Address register)를 이용하여 표본을 추출하였고, 대만의 방식은 지역프레임(area frame)을 사용하여 표본을 추출하였다. 독일의 등록센서스에서, 직접추정 특히 Generalized Regression Estimator에 의하여 과다범위와 과소범위의 크기를 추정하기는 하지만 이원시스템 추정법의 경우는 사실상 행정인구와 조사인구를 비교하는 비추정법이었고, 대만의 등록센서스도 독일과 같이, 행정인구와 조사인구를 비교하는 비추정법이었다고 할 수 있다.
- 등록센서스의 추진과정에서 등록부 자체의 과다범위를 조사할 수 있고, 과소범위를 조사할 수 있다. 과다범위는 원칙적으로 등록부의 주소명부를 이용하고, 과소범위는 해당 지역의 지역프레임(area frame)을 사용한다. 행정자료의 집계범위, 특히 과소집계를 추정하기 위하여, 표본조사를 실시하는데, 이것은 현행 인구주택조사 본조사의 실시 후에 본조사의 집계범위(과소범위)를 추정하기 위하여 실시하는 사후조사와 사실상 같다. 행정자료의 집계범위를 추정

하기 위하여, 행정자료에 포함된 변수와 집계범위를 추정하기 위하여 실시하는 표본조사에 포함된 변수를 매칭/연계하여, 그것을 토대로 하여, 행정자료의 포괄범위를 추정하고, 행정자료의 문제점을 보완하여 센서스통계의 정확성을 개선하고자 노력한다.

1. 영국

- 영국은 현재, 보수당-자민당 연립정권이 전통적 센서스를 폐기하기로 잠정적 결정을 내린 이상, 2021년의 센서스에서 주소명부(address register option) 또는 행정자료 옵션(administrative data option)을 사용할 가능성이 크다고 판단된다. 그러나 물론, 집계수준(aggregate level)에서 행정자료를 사용하게 되느냐 아니면 개체수준(record-level) 또는 중간수준(intermediate level)에서 사용하게 되느냐에 따라 센서스 모집단, 바로 센서스 인구 크기를 추정하는 방법론에서 차이가 나게 될 것이다 (본 보고서의 제 6장 영국의 “Beyond 2011 Programme”을 참고할 것).
- 영국은 개체수준에서 행정자료를 사용하게 된다면, 아마도 전통적인 인구추정 방법론인 이원시스템추정법을 이용하여, 센서스 모집단의 인구 부문을 추정하고 이를 통하여 조사자료를 보완하게 될 가능성이 크다. 영국의 행정자료 옵션은 추정인구만을 작성하면 되기 때문에, 우리나라의 등록센서스와 비교할 때, 영국의 센서스 추정방법론이 그리 복잡하게 전개되지는 않을 것이다.
- 영국의 2011년 센서스에서 이원시스템추정법(DSE)의 일차적 목표는 “One Number Census” 프로젝트의 일부로서, 2011년 센서스 본조사에서 조사되지 않은 인구와 가구의 수를 추정하여 센서스 본조사의 자료를 보완하는 것이다. 또 하나의 목적은 2011년 센서스에서 1회 이상, 잘못된 장소에서 조사된 인구와 가구의 수를 추정하는 것이다. 결국, 인구와 가구의 누락 및 중복을 보완하기 위하여 센서스 본조사의 포괄범위를 조정하는 방법론이라고 볼 수 있다.

- 영국의 “One Number Census”를 달성하기 위한 포괄범위 평가 및 조정방법론은 6단계의 절차를 포함하는 것이었다.

① 영국 통계청은 센서스 포괄범위조사(CCS: Census Coverage Survey)를 본조사와는 별도로 실시한다. 포괄범위조사는 센서스 본조사의 포괄범위를 조사할 목적으로 설계되었으며, 표본은 모든 지방자치단체에서 선정되었다.

② 포괄범위조사의 결과는 2011년 센서스 본조사의 결과와 개체별로 1 : 1로 연계작업을 실시하여, 정밀대조를 하였다.

③ 센서스 본조사 데이터베이스를 점검하여, 포괄범위조사 자료와 대조하여, 그 결과는 센서스 본조사의 과다집계 정도를 추정하는데 사용하였다.

④ 표본지역의 인구는 센서스 본조사와 포괄범위조사 모두에서 자료매칭의 결과를 토대로 누락된 사람들의 추정을 가능하게 하는 이원시스템추정방법을 이용하여 추정되었다.

⑤ 성별, 연령별 인구는 지방자치단체(LAD: Local Authority Districts) 별로 표준회귀추정법과 소지역추정방법을 결합하여 추정되었다.

⑥ 추정결과, 센서스 본조사에 누락된 것으로 판명된 가구와 인구는, 과다집계의 정도를 고려하여, 2011년 센서스 데이터베이스에 임putation) 절차를 통하여 보완되었다.

물론, 이러한 인구규모 조정방식은 200여개의 지방자치단체(LADs)¹⁾ 수준의 인구규모 범위 안에서만 이루어지도록 하였다. 일반적으로, 지방자치단체의 규모가 상대적으로 큰 영국의 경우, 2011년 센서스에도 인구규모가 대단히 적은 지역에 대한 추정방법론 문제로 그리 큰 어려움에 봉착하지 않아도 되었다.

- 영국의 2011년 센서스 포괄범위 점검절차는 (1) 전체인구의 완전집계인 센서스 본조사(census main survey) (2) 사후조사(PES: Post-Enumeration Survey)에 의한 인구의 정교한 재집계 (3) 추정지역 수준에서 과소집계의 정도를 추정하기 위한 센서스 본조사와 사후조사 표본의 매칭작업(matching procedure) (4) 지방자치단체별로, 과소집계를 조정한 추정인구를 산출하는 추정 및 보정작업으로 구성된다.

- 영국의 2011년 센서스는 전통적 센서스이기 때문에, 이것을 2011년 이후의

1) 우리나라의 지방자치단체는 전국이 특별자치시도를 제외하고 광역수준에서는 시도 기초 수준에서는 시군구로 2층 체제로 구성되어 있다. 그러나 영국은 지역에 따라 다르다. 잉글랜드에서는 2층과 단층이 혼재하며, 웨일즈, 스코틀랜드, 북아일랜드에서는 단층체로 통일되어 있다. 2층 체제는 County Council과 District Council로 구성되어 있다. County는 우리나라의 시도에 해당하는 광역자치단체이고, District는 우리나라의 시군구에 해당하는 기초자치단체이다. 잉글랜드의 1층 체제 자치단체로는 대도시권에 존재하는 “대도시권 구역(Metropolitan District)”, “비대도시권 통합구역(Unitary Council)”을 들 수 있다. 이들은 광역/기초 자치단체의 기능을 공유하는 자치단체이다. 런던은 GLA (Greater London Authority) (32개의 London Borough Council)와 City of London Corporation (통상 City라고 불림)으로 구성된다. 또, 웨일즈와 스코틀랜드의 단층 자치단체가 “Unitary”, 북아일랜드에서는 “District”라고 불린다.

통계목적으로 설정되는 지방자치단체(LAD: Local Authority District)는 잉글랜드의 경우 2012년 현재, 런던의 33개(곧 32개의 Borough와 1개의 “City”), 대도시권의 단층체제로 구성된 32개의 “대도시권 구역(Metropolitan District)”, 비대도시권의 단층체제로 구성된 56개의 통합구역(Unitary Council)과 2층 체제로 구성된 자치제도의 하위 자치단체에 해당하는 201개의 구역(District Council)을 합하면 311개가 있다. 웨일즈, 스코틀랜드, 북아일랜드는 모두 단층체제이기 때문에, 통계목적의 지방자치단체는 32개, 22개, 26개가 설정되어, 영국 전체에는 391개의 지방자치단체 통계작성지역이 존재하는 셈이 된다.

행정자료 옵션에 적용한다면, 행정자료 또는 등록부 자료의 집계결과라고 생각하면 되겠고, 이 등록부 자료는 과다집계(OC) 곧 사후조사 표본으로 부적격한 표본의 구성요소들이 없어야 하는 전제조건이 포함된다고 보면 될 것이다.

- 영국의 센서스에서 인구총수를 추정하는데 이용되는 이원시스템 추정방법은 설계기반추정이 아니라 모형 기반 추정에 해당한다. 모형 기반 추정방법을 적용하기 위해서는 성/연령 또는 다른 핵심 인구통계 변수에 따라 총화된 정확한 추정인구가 있어야 한다. 그러나 영국의 사후 표본조사는 2% 추출률의 표본을 선정하여 실시된 것이기 때문에, 200여개의 지방자치단체보다 적은 지역에 대하여 만족할 만한 수준의 추정인구가 존재하지 않는다.
- 영국 통계청(ONS)은 이러한 어려움을 해결하기 위하여, 회귀모형에 기반을 둔 추정방법을 다양하게 개발한다.
- 직접추정(곧 비추정)은 특정 지방자치단체의 조사구(postcode)에서 나온 자료만을 이용하여, 그 자치단체 내의 유사성을 이용하고, 센서스 집계결과와 이원시스템추정값에 기반을 둔 비(ratio)를 승수(multiplier)로 활용한다. 이러한 추정방식은 원래의 자료에서 나온 정보를 명시적으로 이용하기 때문에, 편향성이 적다고 볼 수 있다.
- 합성추정량(synthetic estimator)은 특정 추정지역(Estimation Area) 내의 모든 지방자치단체 자료를 이용하고, 그 추정지역 내에서 추출된 모두 조사구의 유사성을 고려하여, 연령/성별 집단효과를 고려하는 회귀모형을 적합시킨다.
- 소지역 고정효과 모형(small area fixed effect model)은 특정 추정지역 내의 모든 지방자치단체의 조사구에서 나온 자료의 강점(強點)을 차용(“borrowing strength”)하고, 추정지역에서 표본으로 추출된 조사구의 유사성을 고려하여, 통계적 정도(精度: precision) 수준을 추가적으로 개선하기 위하여, 성/연령별 집단효과를 구분한다. 이 모형은 합성추정과 유사하지만, 지방자치단체별 효과를 추정 모형에 별도로 포함한다는 것이다.

- 영국 통계청은 추정에 필요한 표본크기를 늘리기 위하여, 인구 50만 명을 기준으로 책정된 추정지역, 곧 지방자치단체(LAD)보다 더 규모가 큰 광역지역을 상정할 수 있는가를 검토하고, 그 광역지역 내의 지방자치단체 집단들 간의 유사성을 탐색하고, 광역지역 내의 모든 지방자치단체 내의 모든 조사구(EA)를 이용하여, 좀 더 복잡한 모형, 곧 임의(랜덤)효과가 있는 모형들을 생각하는 것이었다. 이들 모형은 다음과 같다.

① 광역지역 Fixed effects model

- 지방자치단체의 효과를 고정시킨 모형
(Fixed LAD effects model)
- 연령/성별 효과를 고정시킨 모형
(Fixed age-sex effects model)

② 광역지역 Random effects model

- 지방자치단체의 효과를 랜덤한 것으로 보는 모형
(Random LAD effects)
- 연령/성별 효과를 랜덤한 것으로 보는 모형
Random age-sex effects

- 모든 센서스 추정결과는 산출 이후, 인구균형방정식, 통계조사, 질적 정보, 행정자료 등을 이용하여 품질확인절차(quality assurance procedure)를 거쳤으며, 추정결과가 설득력이 있는 것임을 이해당사자들에게 확인시키고자 노력을 하였다.

2. 스페인

- 스페인은 앞에서 설명한 대로, 인구대장(주민등록)을 핵심 등록부로 하고, 건축물/주택부문은 빈집에 대한 정보가 부족하기 때문에 전수 현장조사로 진행되며, 핵심등록부의 일부에서 표본을 추출하는 방식으로 표본조사를 진행하고 이를 기반으로 센서스 모집단의 크기와 여타 주요 모수를 추정하는 것이, 2011년에 실시된 등록센서스의 기본적 골격이었다.
- 주민대장의 정보를 보완하기 위하여 실시되는 표본조사의 정보 수집은 다음에서 설명하는 대로 실시하였다.

A) 이론적으로 주거처(主居處: main residence)들에 해당하는 프레임 A (행정자료를 이용하여 만든 센서스 파일, 기본적으로 통계목적의 등록부 곧 “statistical register”에 해당하는 것임)에서 선정되는 통계단위에 대해서는 다음의 2단계로 현지실사 작업을 진행한다.

- 제1 단계의 표본자료 수집: 이 단계에서는 일반우편과 인터넷, 두 가지 방식을 사용하여, 응답자들이 조사표를 완성한다. 인터넷 응답을 원하는 가구를 위하여, 조사표에 접속하는데 필요한 패스워드를 제공한다.

일정 시간이 경과하고 나면, 인터넷을 이용하여 조사표에 응답하지 않은 사람들은 스페인 통계청이 전달한 조사표를 사용하여 응답하고, 그 조사표를 일반우편을 이용하여 반송한다.

우편이나 인터넷으로 센서스 조사표의 조사항목에 응답을 제공하지 않은 거처에 대해서는 일정시간이 경과하고 나면, 독촉장을 보낸다.

- 제2단계의 표본조사: 제1단계 표본조사에서 인터넷이나 일반우편으로 응답하지 않은 거처로서, 조사원이 직접 방문할 필요가 있는 거처들의 일정 비율을 제1단계 조사의 표본에서 선정한다. 거처별로, 집단별로 이러한 무응

답 단위의 수를 계산하고, 컴퓨터 개인면접(CAPI)을 이용하여, 센서스 조사 표에 대한 정보를 수집하도록 한다. 개인접속(CAPI) 방식은 건물/주택 센서스에서 정보 수집을 목적으로 휴대용 컴퓨터를 이용하여 실시되었기 때문에, 센서스 표본조사에서 이용하는데 아무런 어려움이 없을 것으로 예상하였다.

B) 이론적으로 주거처가 아닌 거처, 곧 부거처(副居處: secondary residence)에 해당하는 것으로, 프레임 A (통계목적의 등록부)에서 선정된 통계단위에 대해서는, 이들은 포괄적인 검토 작업을 위한 현지실사에서 검토된다. 이들이 실제로 주거처인 경우에는, 아래에서 말하는 신규 등록을 하는 방식으로 하고, 만약에 주거처가 아닌 거처인 경우에는, 그에 해당하는 상태를 할당한다.

C) 건물/주택 센서스를 하는 과정에서 신규로 등록된 거처에서 선정된 통계 단위(프레임 B)에 대해서는, 등록부에 기재되지 않은 센서스 조사항목의 정보를 다음과 같은 방식으로 수집한다.

- 건물/주택 센서스를 실시하면서, 신규 등록이 발생하는 경우에, 휴대용 컴퓨터에 설치된 응용프로그램은 신규로 등록된 건축물을 확률적으로 선정하는 알고리즘을 가지고 있으며, 그것이 표본의 일부가 될 것인지를 결정한다. 여기서, 해당 부지가 단순한 장소에 불과한지, 아니면 주거처인지, 아니면 주거처가 아닌 부거처인지 아니면 빈집인지를 확인한다. 만약에, 그 주택이 주거처인 경우에는, 조사원은 해당 가구가 조사표를 온라인으로 반송할 수 있도록 하거나, 일반우편으로 반송할 수 있도록 종이 조사표를 유치하고 떠난다.
- 온라인이나 일반우편으로 모두 응답을 하지 않은 나머지 거처에 대해서는 CAPI 방식으로 자료 수집하여, 위에서 설명한 제2단계의 표본조사에 포함하도록 한다.

- 앞에서 언급하였듯이, 2011년의 스페인 인구주택총조사는 사전 센서스 파일

(pre-census file), 건물/주택 센서스, 표본조사 등 3개의 구성요소로 이루어졌다. 이들 3개의 핵심요소는 인구와 가구의 주택특성을 확인하고, 유럽연합 통계청(Eurostat: Statistical Office of the European Union)이 설정한 포괄 범위를 정확하게 점검하는데 필요한 표본조사를 이용하여, 센서스 최종 결과표를 작성하는데 사용하도록 하였다(본 장의 제1절을 참고할 것).

(표본조사의 목적)

- 표본조사의 목적은 주어진 특성집단별로 행정인구 수치를 조정하여 전체인구를 추정하는 것이다. 표본조사 결과를 바탕으로, 스페인 통계청은 행정인구가 표본조사의 인구에 속할 확률을 계산하여, 계수조정요인(recounting factors)으로 이용한다.
- 상이한 지리적 기준의 상세수준에서, 인구와 거처의 특성을 추정하기 위하여, 표본설계를 기반으로 하여 승수요인을 추정하여, 기초자치단체 인구에 캘리브레이션(calibration) 하여, 여기에서 계산된 이른바 "캘리브레이션 가중치"(calibration weights)를 사용한다.
- 위의 두 가지 표본조사 목적에 부합되게, 인구주택총조사가 표본추출의 1차 단위인 센서스 구역별로 구분된 정보를 획득하는데 필요한 유일한 통계조사라는 것을 염두에 두고, 모든 센서스 구역에서 표본을 추출하도록 하는 것이 2011년 스페인 센서스 표본조사의 특징이다.

(표본조사의 모집단 프레임)

- 스페인 등록센서스의 표본 추출틀은 행정자료파일, 바로 통계목적의 등록부(PCF: pre-census file)다. 이 파일은 표본조사를 실시하기 전에, 인구대장, 토지대장 등을 연계/통합(linkage-integration)하여 만든 것이다.

- 행정자료파일(PCF)은 개별 기초자치단체에 존재하는 부동산 부지와 그 곳에 살고 있는 사람들의 명부이며, 여기에는 이들 두 통계단위에 대하여 입수 가능한 정보를 추가하여 작성되는 것이다. 이 파일은 바로 제안된 목표를 달성하기 위하여, 표본설계의 효율화 목적에 이용될 수 있을 것이다.
- 다양한 방식으로 대조작업을 진행하고 나면, 거처는 행정자료파일(PCF)에서, (1) 주소가 확인되는 것과 (2) 주소가 확인되지 않는 것의 두 종류로 구분된다. (1)의 경우는 우편주소를 통하여, 현지실사에 주소를 찾을 수 있는 통계단위들이다. 후자는 완전한 주소가 없는 통계단위로서 현지실사에서 어디에 입지하고 있는가를 확인할 수 없다.
- 표본선정을 목적으로, 자치단체별로 모든 거처는 두개의 프레임으로 구분한다, 프레임 A는 행정자료파일에서 주소가 확인되는 통계단위이고, 프레임 B는 표본조사의 현지실사 작업에서 광범위한 현장방문을 통하여 신규 등록된 거처들의 세트로 구성되었다.

(표본의 구성)

- 센서스의 목적을 달성하기 위하여, 거처들의 표본을 프레임 A을 기준으로 하느냐, 아니면 프레임 B를 기준으로 하느냐에 따라서 달라지는 표본을 선정한다. 표본조사의 표본은 모든 센서스 구역에서 선정된다.

(1) 프레임 A

- 프레임 A에서 표본추출을 한다고 했을 때, 프레임 A가 바로 거처용도인 건축물이라는 것을 유념할 필요가 있고, 이것은 현지실사를 하기 전에 확인된다. 스페인 통계청은 “행정자료파일”이 주(主)거처(main dwelling)라고 판단하는 그룹에 속하는 거처들에 대하여 일반우편으로 접촉을 하도록 했다. 주거처의 가구들은 온라인으로 센서스 조사표에 응답을 하든가, 아니면 완성된 조사표를 일반우편으로 반송하는 절차를 완료함으로 해서, 응답자로서의 의무를

종결한다.

- 일정 시간이 경과하고 나서, 우편조사에 응답하지 않은 나머지 거처들의 세트에서, 센서스 조사표를 컴퓨터 개인면접(CAPI)으로 완성하게 될 하위표본(sub-sample)을 선정한다.
- 주거처가 아닌 경우 건물/주택 센서스(Building Census)를 실시하는 동안에 검토대상이 되도록 하고 있으며, 그 과정에서 해결될 수 없는 주거처가 아닌 건축물 부지에 대해서는 컴퓨터 개인면접(CAPI)으로 센서스 자료를 수집하도록 하고 있다.
- 그렇기 때문에, 프레임 A를 기준으로 실시되는 표본조사는 두 종류의 서로 다른 거처를 조사 대상으로 하는데, 하나는 주거처를 포함하는 것이고, 다른 하나는 주거처가 아닌 거처를 포함하는 것이다.

(2) 프레임 B

- 프레임 B의 표본은 현지실사를 전수조사 형식으로 진행할 때, 선정되는 것으로, 여기에서는 무작위 확률추출 베르누이 절차(random-selection Bernouilli procedure)를 사용한다.
- 프레임 B의 표본은 앞에서 본 추출방식 때문에, 3종류의 거처로 구성된다.
 - ① 인터넷이나 일반우편으로 조사표를 반송한 주(主)거처와 주거처가 아닌 거처로 이들은 모두 프레임 A에서 선정된 것이다.
 - ② 프레임 A에서 선정된 주(主)거처로서, 이것은 프레임 A의 하위표본에서 선정된다.
 - ③ 프레임 B에서 선정된 거처들

(표본의 크기)

- 스페인 통계청은 기초자치단체 수준의 정보를 획득해야 하는 표본조사의 일차적 목표를 염두에 두고, 표본조사에 가용한 예산까지 고려하여, 표본의 크기는 3백만 거처 정도를 예상하였으며, 이것은 전체 거처의 11.9%에 해당하며, 인구규모로 보면, 추출률은 12.3%에 해당한다.
- 표본조사의 목적을 달성하는데 적합하다고 판단하는 이 표본규모가 기초자치단체별로 어떻게 분포되는가는 추정치의 추출오차를 어느 정도 수준으로 하는 것과 관련되며, 이것은 다시 기초자치단체의 크기와 직결된다고 볼 수 있다.
- 결국, 인구규모가 대단히 작은 기초자치단체와 대단히 큰 기초자치단체 간에는 추출률이 달라졌다. 한마디로, 스페인의 2011년 센서스는 모든 기초자치단체에 대하여 추출률이 동등한 단순확률추출과는 달리, 자치단체의 크기(인구, 거처)에 따라 표본 추출률이 달라지는, 이른바 “복합적 표본설계”(complex sample design)이다.
- 표본설계의 효율성을 극대화하기 위하여, 스페인 통계청은 (1) 프레임 A에서 추출되는 주거처가 아닌 건물의 비율을 주거처인 건물의 40%로 하고, (2) 프레임 B에서 추출되는 거처에 대해서는 주거처가 아닌 건물의 비율을 주거처인 건물의 60%로 하였다. 다만, 전수조사를 하는 하위표본에 대해서는 프레임 A와 프레임 V의 주거처가 아닌 건물을 전부 전수 조사하는 것으로 하였다.
- 스페인 통계청은 이 기준을 사용하면, 인구가 2만 명 정도 되는 기초자치단체의 특성치가 10%인 변수의 추정치는 대체로 변이계수(CV)가 7.7%에 이르게 될 것이라고 보았다.

(표본의 선정)

- 표본은 모든 기초자치단체에서 선정되며, 주거처인 건물이나 주거처가 아닌 건물에 대해서는 그 건축물의 모집단 안에서는 동등 확률의 기준을 충족시켜서 표본을 선정하였다.
- 프레임 A에서 추출하는 표본은 출발점은 무작위로 선정하는 계통추출방식(系統抽出方式)으로 하였다. 또, 프레임 A의 표본에서, 일반우편이나 인터넷으로 응답을 하지 않은 거처들의 하위표본을 선정하는 작업도 출발점을 무작위로 하는 계통추출방식을 이용하였다.
- 프레임 B에서 추출하는 표본은 전수조사에서 신규로 등록된 거처의 세트에서 선정하며, 이 경우에는 베르누이 절차를 이용하여 추출단위에 동등 확률의 원칙을 적용하였다.

(추정량의 계산식)

- 개별 기초자치단체에서 거처특성과 인구특성의 추정량은 아래에 설명하는 캘리브레이션 기법을 적용하여, 무응답 보정을 거친 승수추정량이 된다. 초기 승수추정량은 선택확률의 역수로 한다.

(표본 포함확률의 계산)

- 표본추출 절차에 의거하여, 개별 기초자치단체 내에 3종류의 거처가 있으며,

거기에 각각의 표본 포함확률의 역수를 승수요인으로 하여, 그 승수요인을 적용하였다.

1. 프레임 A에서 추출되는 거처들 (하위표본 추출은 없음)

1.1. 주(主)거처

- 프레임 A에 들어 있는 주(主)거처가 표본에 선택될 확률은 다음 식으로 계산하였다.

$$\Pr(V_i^{AP}) = f_r^{AP} = \frac{V_{AP}^t}{V_{FPC}^{AP}}$$

여기서 f_r^{AP} = 주(主)거처 그룹에 적용되는 표본 추출률

V_{AP}^t = 프레임 A를 구성하는 사전 센서스 파일(PCF)에 들어 있는 주거처들의 총수

V_{FPC}^{AP} = 프레임 A에 추출된 이론상의 주(主)거처들의 총수

1.2. 주(主)거처가 아닌 거처들

- f_r^{ANP} 는 주(主)거처가 아닌 거처의 경우에 대하여 사용되는 표본 추출률로서, 이것은 주거처가 아닌 거처가 표본에 포함될 확률의 의미하는데, 그것은 아래 계산식으로 표시된다.

$$\Pr(V_i^{ANP}) = f_r^{ANP} = \frac{V_{ANP}^t}{V_{FPC}^{ANP}}$$

위의 식에서 보는 각 항의 의미는 각 항이 주거처가 아닌 거처들에 적용된다는 것을 제외하고는 주(主)거처의 계산식에 나타나는 항들과 그 의미가 유사

하다.

2. 프레임 A에서 추출되는 거처들 (하위표본 추출의 경우)

- 위와 비슷한 절차를 거치면서, f_{rs}^{AP} 를 프레임 A에서 추출된 주거처 중에서, 일반우편이나 인터넷으로 조사표를 반송하지 않은 하위표본 중에서의 추출률이라고 하고, 거처 i 가 표본에 포함될 확률을 계산한다면, 그 계산식으로 다음과 같이 표시된다.

$$\Pr(V_{is}^{AP}) = f_r^{AP} \cdot f_{rs}^{AP} = \frac{V_{AP}^t}{V_{FPC}^{ANP}} \cdot \frac{V_{APs}^t}{V_{AP}^t - V_{AP}^e}$$

여기서 V_{APs}^t = 하위표본에서 선정된 주거처들의 이론적 총수가 되고 V_{AP}^e = 일반우편이나 인터넷으로 응답한 표본에 추출된 가구의 그룹이 된다.

3. 프레임 B의 거처들

- 이 프레임에서 거처들이 추출된 확률은 동일하다. f_r^B 를 프레임 B에서 이용하는 표본 추출률이라고 한다면, 거처 i 가 표본에 속할 확률은

$$\Pr(V_i^B) = f_r^B$$

프레임 B에서 추출되는 거처의 총수는 표본에서 용도가 주택인 건축물에서 추정한다. 따라서 현장조사에 의한 건물/주택 센서스가 끝나고 나면, $\widehat{V}_B$ 를 계산할 수 있다.

- 초기 승수요인은 표본에 포함될 확률의 역수로 구하지만, 이 수치는 각 경우에 대하여 입수 가능한 무응답의 비율을 이용하여 필요한 보정절차를 거쳐야 한

다.

(표본에 따른 상이한 거처유형의 추정)

- 개별 기초자치단체의 유형별로 이루어지는 총거처수의 추정에 대해서는, 비추정법(ratio estimation)을 사용한다. 여기서, 사전 센서스 파일에서 계산한 총거처수와 광범위한 현지실사에 획득한 정보에서 획득한 총거처수를 보조정보(補助情報)로 사용한다.
- 주어진 유형의 거처 t 에 대한 승수추정량(V_t)을 계산하기 위한 식은 다음과 같다.

$$\widehat{V}_t = \sum_i f_i \cdot t_i$$

여기서 t_i = 거처 i 가 유형 t 에 해당하느냐 하지 않느냐에 따라, 값을 1 또는 0를 취하는 변수이고, f_i = 그 주택의 승수요인에 해당한다.

- 그 합은 해당 기초자치단체의 모든 주택으로 확대된다.

$$\widehat{\widehat{V}}_t = \widehat{V}_t \cdot \frac{V}{\widehat{V}}$$

여기서 V 의 값은 포괄적인 현지실사에서 삭제되지 않는 사전 센서스 파일(PCF, 곧 프레임 A)의 총거처수와 신규 등록 거처들의 표본에서 추정된 거처의 수를 합한 것이다.

최종적으로 수정된 승수요인, 바로 가중치는 다음과 같이 계산한다.

$$w_i = f_i \cdot \frac{V}{\sum_i f_i}$$

(인구의 계산)

- 사전센서스파일(PCF)은 기초자치단체 주민대장을 토대로 하며, 이것은 특정 시점에 기초자치단체에 거주하는 주민명부이다. 이 주민대장과 여타 행정자료(사회보장파일, 국세청 조세자료 파일)을 비교 대조하여, 이 파일에 나타나는 사람들의 대다수는 스페인 영내 거주 여부를 확인할 수 있다. 그러나 행정자료파일(PCF)에서, 스페인 영내 거주 여부가 확인되지 않는 경우도 있는데, 그 이유는 주민대장에 기재된 사람들이 행정자료 파일에는 등록되어 있지 않기 때문이다.
- 주민대장에 기재된 사람으로, 행정자료(사회보장파일, 국세청 조세파일 등)에서 스페인 영내 거주사실이 확인되는 사람은 계수조정요인(recount factor)을 1 (FR =1)을 할당하고, 행정자료에서 확인이 불가능한 사람은 표본자료 정보를 이용하여 계수조정요인을 할당한다.

(계수조정요인)

- 계수조정요인(FR: recount factor)은 행정자료파일(PCF) 기록과 표본조사기록을 통합하여 계산하여 조정하는 계수이다.
- 스페인 영내에 거주가 확인되지 않은 사람들의 계수조정요인(FR)의 추정에 대해서는, 이들을 포함하는 인구집단을 사용하도록 할 것이며, 이들은 센서스 조사표에 포함된 변수들의 파생특성을 통하여 정의되기 때문에, 표본정보를 이용하여 추정할 수 있을 것이다. 결국, 스페인 영내 거주가 확인되지 않은 사람들에게 대해서는 행정자료파일(PCF)과 여타 보조파일을 비교 대조하는 작업을

해도 그 결과는 완전하지 않기 때문에, 이 계수조정요인을 산출하는 절차는 스페인 센서스의 필수적인 방법론적 절차이다.

- 위의 결과에서, 우리가 G 라고 부르기로 하는, 상기의 스페인 영내 거주하는 사람들의 집단 중 하나인 집단 P_G 는 스페인 영내 거주가 확인되고, 등록부에서 집계되는 사람들(S_G)과 표본조사 결과에서 추정되는 사람들을 포함하게 된다. 이 집단에 대해서, 추정인구는 $\widehat{D}_G$ 로 하고, 사전센서스파일에 포함되어 있는 사람은 D_G 로 한다.
- 표본조사 결과에서 추정되는 집단에 대하여, $P_G = S_G + D_G$ 이다. 거처표본에서 집단 G 의 인구 추정은 다음과 같은 방식으로 실시된다.

$$\widehat{P}_G = \sum_i w_i \cdot p_{iG}$$

여기서 합의 공식은 고려대상이 되는 지리적 범위에 있는 표본의 모든 거처 i 에 그대로 확장된다. w_i 는 거처의 가중치이고, P_{iG} 는 그룹 G 에 속한 표본의 거처 i 의 모든 사람들을 가리킨다.

- 결과적으로, 그룹 G 의 추정식 D_G 은 $\widehat{D}_G = \widehat{P}_G - \widehat{S}_G$ 이고, 따라서 그룹 G 의 계수조정요인은 다음과 같은 수식으로 표현할 수 있다.

$$FR_G = \frac{\widehat{D}_G}{D_G} = \frac{\widehat{P}_G - \widehat{S}_G}{D_G} = \frac{\sum_i w_i \cdot P_{iG} - S_G}{D_G}$$

여기서, 분모 D_G 는 사전 센서스 파일에 기재되어 있지만, 스페인 영내에서 거주사실이 확인되지 않은 그룹 G 의 총인구라고 할 수 있다.

- 이 절차가 종료되면, 모든 사람들의 계수조정요인(FR, recount factor)이 사전 센서스 파일에 배정되며, 그 요인은 어떤 사람은 1이 될 것이고 어떤 사람

은 위의 절차에 의하여 계산된 계수조정요인을 갖게 될 것이다.

- 주어진 지역(geographical area)의 인구는 그 지역의 사전 센서스 파일(PCF: pre-census file)에 기재된 인구의 계수조정요인(recount factor)를 합산한 것이다.

(거처특성의 계산)

- 센서스에서 이용되는 거처의 특성은 상주목적의 주거처만을 상대로 하여 통계를 작성한다. 상이한 주택유형의 추정에서, 사용하는 계산식은 다음과 같다.

$$V = \widehat{V}_{pr} + \widehat{V}_s + \widehat{V}_v + \widehat{V}_{of}$$

각각의 합은 주거처(pr, main dwelling), 주거처가 아닌 거처(s, secondary dwelling), 빈집(v, empty dwelling), 다른 용도의 주택(of, dwellings for other purpose)에 대하여 추정된 총수이다.

- 주거처 총수의 보완된 추정량은 다음과 같은 수식으로 표현한다.

$$\widehat{V}^* = \frac{P}{\sum \widehat{V}_{pr}} = \widehat{V}_{pr} \frac{P}{\widehat{P}}$$

여기서 P = 계수조정요인(FR)과 $\widehat{V}_{pr}$ 을 계산하여 구한, 행정자료파일(PCF)의 인구, $\widehat{V}_{pr}$ 과 $\widehat{P}$ = 표본조사 결과에서 추정된 가구와 인구의 총수로
서, 위에서 언급된 절차에 따라 계산된다.

(인구특성의 추정)

- 인구특성 Y 의 추정은 비추정법 사용하며, 보조변수는 캘리브레이션 그룹 T 에서 구한 기초자치단체의 인구이다. 가구특성의 경우와 마찬가지로, 다음의 비추정식이 사용된다.

$$\widehat{P}_M^T = \widehat{P}_M \cdot \frac{P_{MT}}{\widehat{P}_{MT}} = \sum_{i \in S_M} w_i^f \frac{P_{MT}}{\sum_{i \in S_M} w_i^f p_i^T} \cdot y_i$$

합산의 공식은 기초자치단체 표본(S_M)의 모든 거처에 적용된다. w_i^f 는 거처의 가중치이다. p_i^T 는 캘리브레이션 그룹 T 에 속한 거처 i 에 거주하는 모든 사람의 수이다. y_i 는 특성 Y_i 를 갖는, 그룹 T 에 속하는 거처 i 의 총인구를 말한다. 또 P_{MT} 와 $\widehat{P}_{MT}$ 는 기초자치단체의 캘리브레이션 그룹 T 에 대한 행정인구와 추정인구를 가리킨다.

- 스페인 통계청은 2011년 센서스에서 캘리브레이션 그룹은 기초자치단체의 크기(인구규모)에 따라서, 상이한 변수들을 비교 검토하면서 정의하도록 할 계획이라고 밝히고 있다.

(추출오차)

- 센서스 통계 작성에서, 거처(주택) 표본조사를 이용한다는 사실은 센서스 통계가 표본추출오차의 영향을 받는 것을 의미하며, 표본추출오차는 표본의 크기뿐만 아니라, 추정되어야 할 모집단의 특성이 무엇이냐에 따라서도 영향을 받는다는 것을 의미한다.
- 주, 자치단체, 전국 수준의 센서스 통계표 작성에서, 추출률이 자치단체의 크기에 따라 차이가 나기 때문에, 잭크나이프 방법(Jackknife method)를 사용한

다. 이 방법에 근거하여, 스페인 통계청은 다음 방식으로, 모집단 특성 Y 의 추정량의 분산을 추정할 계획이다.

$$\widehat{V}(\widehat{Y}) = \sum_{a \in A} \frac{n_a - 1}{n_a} \sum_{i \in a} (\widehat{Y}_{(ia)} - \widehat{Y})^2$$

여기서 A = 추정이 이루어지는 지역을 말한다.

$a \in A$, $a = 1, 2, 3, \dots, N_a$, 는 지역 A 안에서 동일한 추출률을 갖는 기초자치 단체의 동질적인 집단을 의미한다. (ia) 는 그룹 a 안에 있는 거처들의 세트 i 를 가리킨다. n_a = 그룹 a 안에 있는 세트 (ia) 의 수를 가리킨다. $\widehat{Y}_{(ia)}$ = 모 집단의 특성, 바로 모수 Y 의 추정치로서, 표본의 거처들 (ia) 의 추정집단에서 구한다. $\widehat{Y}$ = 모집단의 특성인 모수 Y 을 전체 표본에 대하여 추정한 수치이다.

- 추정량을 구하기 위하여, 승수요인을 재계산하기보다는 거처들의 그룹 i 가 제거되는 그룹의 $\frac{v_a}{v_a - v(ia)}$ 라는 승수요인을 곱하여 계산한다. 위의 공식에 따라서

$$\widehat{Y}_{(ia)} = \sum_{j \notin a} F_j y_j + \sum_{\substack{j \in A \\ j \notin ia}} F_i \frac{v_a}{v_a - v(ia)} y_i$$

여기서 ia = 지역 A 에 있는 거처들의 그룹, v_a = 그룹 a 에 있는 거처의 총수, $v(ia)$ 는 집단 (ia) 의 주택 총수를 가리킨다.

- 스페인 통계청은 2011년 등록센서스의 결과표를 제시할 때, 추정결과는 상대 표본추출오차(RSE: Relative Sampling Error)(%로 표시), 변이계수(CV: coefficient of variation)를 함께 제시할 예정이다. 변이계수는 다음과 같은

식으로 표시한다.

$$\widehat{CV}(\hat{Y}) = \frac{\sqrt{\widehat{V}(\hat{Y})}}{\hat{Y}} \cdot 100$$

표본추출오차는 특정 확률(통상 $1-\alpha=0.95$)을 기준으로, 센서스 항목의 특성별 참값이 존재할 수 있는 신뢰구간을 계산하는데 도움을 준다.

- 스페인 통계청은 기초자치단체 통계표에 대하여, 표본선정에 고려된 추출방법을 염두에 두고, 총별 합계(class totals)가 다르거나 추출률이 다른 경우에 센서스 통계표에 변이계수(CV: Coefficient of Variation), 상대표본추출오차(RSE: Relative Sampling Error)를 동시에 계산하도록 되어 있다.

3. 독일

- 독일의 등록센서스는 표본설계와 추정모형에 관한 논의는 Robert Munnich, Siegfried Gabler 등 트리에르 대학 통계학 교수들이 집필하여 연방통계청이 2012년에 출판한 보고서 「독일 2011년 센서스의 표본적정화와 추정방법론」(Stichprobenoptimierung und Schätzung im Zensus 2011)에 상세히 소개되어 있다(<그림 7-8> 참고). 이 보고서는 추정방법론을 표본추출과 관련하여 전반적인 소지역모형을 검토하면서, 설계 기반 추정방법으로 HT 추정방법(Horovitz-Thompson estimator), 일반화 회귀추정방법(GREG: Generalized Regression Estimator)과 모형 기반 추정으로 본 EBLUP(Empirical Best Linear Unbiased Predictor), HB 추정(Hierarchical Bayesian estimator), 공간모형(Spatial Model)을 소개한다. 마지막으로 전통적 인구추정방법에 해당하는 이원시스템 추정방법(DSE)의 응용방안으로 SPREE (Structure Preserving Estimator)도 고려한다.
- 독일의 센서스 추정방법론은 일차적으로 설계 기반 추정량에 관심을 기울이고,

계속해서 모형 기반 추정량과 인구학적 추정방법인 이원시스템추정량에 대하여 언급한다. 또, 독일은 연방국가로서 행정체계가 대단히 복잡하기 때문에, 소지역 추정에 대해서 많은 관심을 기울이고 있다. 설계 기반 추정량은 조사 가중치를 승수요인으로 하여, 표본 내 관측치를 가중하는 방법이다. 조사가중치의 계산은 표본설계와 목표모집단에 대한 보조변수를 이용한다. 표본 관측치를 가중화한 결과를 미지의 모집단 수치의 설계 비편향 추정량 (design-unbiased estimator)으로 한다. 이와 관련된 추론은 모수의 추정값이 고정되었다고 가정하고, 표본설계에 의하여 도출된 확률분포를 이용하여 모수를 추정한다.

- 여기서는 센서스 추정방법론 중에서 인구학자들에게는 인구균형방정식과 더불어, 센서스 모집단의 크기를 추정하는데 흔히 이용되는 방법론이기는 하지만, 우리나라 센서스 추정방법론의 개발에서는 통계청의 2010년 학술연구 용역사업에서 폐기되었던 이원시스템 추정방법(DSE)의 독일식 응용방식에 대하여 살펴보기로 한다.

(이원 시스템 추정모형의 골격)

- 독일의 등록센서스는 설계 기반 추정방법을 중심으로 센서스 추정방법론을 전개하지만, 인구통계분석에서 일반적으로 사용되는 이원 시스템 추정모형의 기본원리가 그대로 사용되고 있다. 그러나 독일 연방통계청의 추정방법론 보고서는 연방주별로 지방자치단체의 규모가 상대적으로 작은 지역에 대해서는 이원시스템 추정모형을 변형하여야 할 필요가 있다고 주장한다(Munnich, Gabler et al, 2012).
- 고전적 이원시스템 추정모형은 인구의 규모가 N_p (미지수)인 모집단 U 는 폐쇄인구(closed population)으로, 출생, 사망, 인구이동이 발생하지 않는다고 가정한다. 물론, 이러한 가정은 동물인구의 이원시스템 추정모형에서 사용되고 있지만, 센서스 모집단의 총수를 추정하는데 그대로 적용되지는 않을 수 있다.

<그림 7-8> 독일 센서스의 연방통계청 표본설계와 추정방법론 보고서



- 자료생성은 다항분포(multinomial model)로 기술될 수 있으며, 모집단 U 의 요소 i (곧 $i \in U$)는 (1) List C 에 들어있지만, List S에는 빠져 있고 ($i \in List C, i \notin List S$) (2) List C에 빠져있지만, List S에는 들어있고 ($i \notin List C, i \in List S$) , (3) List C와 List S에 모두 들어 있고 ($i \in List C, i \in List S$) (4) List C와 List S에 모두 빠져 있는 경우 ($i \notin List C, i \notin List S$) 로 구분할 수 있다는 가정에서 출발한다.
- 요소 i 가 4개의 셀(cell) 중 어느 하나에 속할 확률이 동등하다고 하고 그것을 p_A 라고 하면, 다항분포의 파라미터 p_A 의 첨자 $A = UC1, UC2, 00, 11$ 로 표기할 수 있을 것이다. 구체적으로

p_{uc1} = List C 에 들어있지만, List S에는 빠져 있을 확률

p_{uc2} = List C에 빠져있지만, List S에는 들어있을 확률

p_{00} = List C와 List S에 모두 들어 있을 확률

p_{11} = List C와 List S에 모두 빠져 있을 확률

<표 7-3> 2x2 횡단제표

		List S		
		IN	OUT	Total
List C	IN	N_{00}	N_{uc1}	N_c
	OUT	N_{UC2}	N_{11}	?
	Total	N_s	?	N_p

- 이 모형에 따라서, N_p 개의 실험을 독립적으로 시행하고 난 결과는 2x2 횡단 제표 (<표 7-3>)의 셀 빈도(cell frequency)로 표시되어 있다.

- 위의 두 List C, D는 범위오차(coverage error)를 포함한다. 이 범위오차를 추정하기 위하여, 포획/재포획 모형(capture-recapture model)을 이용하여 폐쇄인구 U 의 크기를 추정하는 방법을 채택한다. 이 방법은 흔히들 피터센 추정량(Petersen estimator)라 불린다.

$$\hat{N}_p = N_c \frac{N_s}{N_{00}}$$

- 이 방식에는 포획/재포획 모형의 가정, 바로 List C와 List S는 상호 독립적으로 선정된다는 가정이 들어 있다. 두 개의 List를 모두 전수조사로 하는 것은 비용 문제 때문에 힘들다. 이 때문에, List C만 센서스의 전수조사 결과를 이용하고, 두 번 째의 List S는 표본조사를 이용하여, 이를 List C와 연계하여, N_s 와 N_{00} 를 추정하고, $\hat{N}_p$ 를 아래와 같이 추정한다.

$$\hat{N}_p = N_c \frac{\hat{N}_s}{\widehat{N_{00}}}$$

- 이 추정방법을 이원시스템추정량(DSE)이라고 부른다. 물론, 이 추정절차는 2개 이상의 List에 모두 적용될 수 있다(Zalavsky, 1989).
- 이원시스템 추정량(DSE)은 List C와 List S의 모두에 빠져 있는 요소의 개수 (N_{11})가 무시될 수 있다고 가정하기 때문에 추정이 가능하다. 이원시스템추정량(DSE)은 모집단 U 의 요소 i 는 List C의 표본에서 모두 똑 같은 포획확률(capture probability) $p_{00} + p_{uc1}$ 을 갖고, List S의 표본에서 모두 똑 같은 재포획확률(recapture probability) $p_{00} + p_{uc2}$ 을 갖는다고 가정한다.

- 사실, 이 가정은 현실성이 없는 가정임은 두 말할 필요도 없다. 인구통계분석(demographic analysis)에 의하면, 성별, 연령별 집단이 달라지면 모집단 U 의 요소 i 는 집단별 확률이 달라진다. 곧 개별 집단 간에 존재해야 하는 동질성 가정(homogeneity assumption)을 위반하게 되면, 모집단 U 를 몇 개의 부모집단별로 이원시스템 추정량(DSE)을 적용하고, 최종적으로 부모집단 추정결과를 더해야 한다.

(미국과 스위스의 이원시스템추정량)

- 독일의 등록센서스 표본적정화 및 추정방법론 연구(Munnich et al, 2012)는 미국과 스위스의 이원시스템 추정방식을 고전적 방식으로 보고 있다. 미국은 2000년 센서스에서 사후조사, 특히 정확도/포괄범위 평가조사(ACE: Accuracy and Coverage Evaluation Survey)라 불리는 프로젝트를 수행했다. 이 조사는 이원시스템 추정량(DSE)으로 N_s/N_{00} 를 추정하기 위하여, 포괄범위를 평가하는 것이 주목적이었다. 이것은 앞에서 사용한 용어를 기준으로 하면, List S에 해당하고, 미국 센서스 사후조사(PES) 일반적 용법으로는 P-Sample로 알려져 있다.
- 미국 2000년 센서스에서 이원시스템 추정량(DSE)은 다음과 같은 형태로 표시된다.

$$\hat{N}_p^{USA, 2000} = \hat{N}_c \frac{\hat{N}_s}{\hat{N}_{00}}$$

이 절차를 확장하여 변형한 사례들은 Hogan (2003)의 연구에서 찾아 볼 수 있다.

- 미국에서 개발된 추정방법은 2000년 스위스 센서스의 포괄범위 추정의 기초

가 되기도 했다. 과소범위(UC)의 추정엔 센서스 본조사 실시 1개월 후에 실시되는 P-Sample를 이용한다. 과다범위(OC)는 실제 센서스 본조사에 응답한 사람들을 모집단으로 해서, 표본(보통, E-Sample이라고 부름)을 추출하여 추정한다. 두 개의 표본은 2000년 12월 15일에 실시된 스위스 포괄범위조사(Swiss Coverage Survey)의 일부이다.

- 이 방법으로 스위스 통계청은 2000년 센서스의 과다범위(OC)는 목표모집단의 0.4%, 과소범위는 1.6%라고 추정하였다. 이 범위오차를 이원시스템추정량(DSE)에 조합하여 계산한 결과, 2000년 센서스의 범위오차(coverage error)는 1.4%였다. 하위모집단이나 표본추출 방식에 대한 설명은 Renault (2001)을 보면 알 수 있다.

(독일의 이원시스템추정량)

- 독일의 2011년 등록센서스는 우리나라의 2015년 등록센서스와 마찬가지로, 주민등록대장을 이용하였다. 자치단체의 행정인구(τ_R)는 그 지역에 거주하지 않는 인구, 곧 등록부 기준 과다 인구(τ_{oc})를 포함한다. 또 그 지역에 거주하고 있지만, 인구대장에 등록되어 있지 않은 인구, 곧 등록부 기준 과소 인구(τ_{uc})를 포함한다.
- 독일에서는 건축물대장을 구축하기 위한 별도의 우편조사가 실시되었으며, 주민등록과 주소명부/건축물대장 동시적으로 센서스에 이용한다. 따라서 N 은 해당지역의 센서스 모집단 U 에 과다 또는 과소의 형태로 존재하는 모든 주소의 개수이며, N_p 는 이들 주소에 거주하는 모든 사람들의 합계이다. 또 주민등록에 기재되어 있을 뿐만 아니라, 자치단체의 실제거주자 수를 N_{00} 라고 하면, $N_p = \tau_{oc} + \tau_{00} + \tau_{uc} = \tau_c + \tau_{uc}$ 이다.
- 2011년 독일 센서스 추정방법론의 중대과제는 인구수($\tau_c = \tau_{00} + \tau_{uc}$)의 참값(true value)은 미지수이기 때문에, 그것을 추정하는 방법을 합리적으로 모색

하는 것이었다. 우리는 이 센서스 인구의 모집단을 C 로 표기한다. 한편, 이미 언급한 것처럼, 행정인구 ($\tau_r = \tau_{00} + \tau_{oc}$)의 크기는 기지수(既知數)이다.

- 2011년 독일 센서스에서 연방통계청은 통계학에서 흔히 사용되는 설계기반 추정법(design-based estimation) 외에 전통적인 인구통계 추정법인 이원시스템추정량(DSE)의 이론을 일부 수정하여, 자치단체별 센서스 인구(τ_c)의 참값을 추정한다.
- 앞에서 설명한 것과 같은 방식으로, 우리는 폐쇄인구 U 에서 각각의 확률 $p_{iuc}, p_{i00}, p_{ioc}$ 로 이루어진 무작위실험(random experiment)이 U 의 요소(사람) i 가 3개의 별도 범주, 곧 (1) 등록부 기준 과소 인구에 속하거나 (2) 등록부에 기재되고, 센서스 기준에도 부합되는 인구에 속하거나 (3) 등록부 기준 과다인구에 속하는 3개의 별도 범주를 결정한다고 보고 시작한다. 곧 $p_{iuc} = P(i \notin R, i \in C)$, $p_{i00} = P(i \in R, i \in U)$, $p_{ioc} = P(i \in R, i \notin C)$ 이다. 동시에 U 의 요소(사람) i 가 등록부에 소속될 확률 ($p_{iR} = P(i \in R)$)과 센서스에 속할 확률($p_{iC} = P(i \in C)$)은 다음과 같은 관계를 갖는다고 가정한다.

$$p_{iC} = p_{i00} + p_{iuc}$$

$$p_{iR} = p_{i00} + p_{ioc}$$

<표 7-4> 포괄범위의 3개 범주와 빈도

		범위종류	빈도1	빈도2	빈도3
등록부 R	IN	과다범위(OC)	τ_{oc}	τ_R	τ_{oc}
		00	τ_{00}		τ_C
	OUT	과소범위(UC)	τ_{uc}	τ_{uc}	
			N_p	N_p	N_p

- <표 7-4>는 모집단 U 의 N_p 에 대해서 상호독립적인 반복실험 후에, 세 범주 $p_{ioc}, p_{i00}, p_{iuc}$ 와 그것의 빈도 $\tau_{oc}, \tau_{00}, \tau_{uc}$ 간의 관계를 보여준다. 여기서, 중요한 것은 등록센서스 실시국에서는 등록부 R 이 미국이나 여타 전통적 센서스를 실시하는 국가의 센서스 본조사(census main survey) 역할을 한다는 것을 명심할 필요가 있다.

<표 7-5> 표본 S 의 빈도분포

		범위종류	빈도 1 (S)	빈도 2 ($\bar{S}$)	추정1	추정2	추정3
등 록 부 R	IN	과다범위 (OC)	$\tau_{oc, S}$	$r_{oc, \bar{S}}$	$\hat{\tau}_{oc}$	$\hat{\tau}_R$	$\hat{\tau}_{oc}$
		00	$\tau_{00, S}$	$r_{00, \bar{S}}$	$\hat{\tau}_{00}$		$\hat{\tau}_C$
	OUT	과소범위 (UC)	$\tau_{uc, S}$	$r_{uc, \bar{S}}$	$\hat{\tau}_{uc}$	$\hat{\tau}_{uc}$	
			n_p	$N_p - n_p$	$\hat{N}_p$	$\hat{N}_p$	$\hat{N}_p$

- 향후 2015년 한국의 등록센서스는 물론 독일의 2011년 등록센서스와 같이 센서스 예산 경감과 응답자의 부담 경감이 등록센서스 실시의 일차적 이유이다. 따라서 응답자 전체가 아니라 그 일부에 대하여 표본조사를 실시한다. 독일의 경우 표본 S 는 자치단체의 주소명부를 이용한다. 선정된 주소에 거주하는 모든 사람은 표본조사에 포함된다. n 개의 주소지에서 실제로 거주하거나 등록되어 있는 n_p 명에 대한 표본 S 에 대해서, 무작위 실험을 할 때, $\forall i \in U$ 라고 할 때, 요소(사람) i 가 6개의 범주에 속할 확률은 $p_{S, ioc}, p_{S, i00}, p_{S, iuc}, p_{\bar{S}, ioc}, p_{\bar{S}, i00}, p_{\bar{S}, iuc}$ 로 표기할 수 있다. 여기서, $\bar{S}$ 는 모집단 U 에서 S 의 보집합(補集合)이다. 또 $\forall i \in U$ 에 대하여 동등 확률을 $p_{ioc} = p_{oc}, p_{i00} = p_{00}, p_{iuc} = p_{uc}$ 로 표기한다. <표 7-5>은 이 무작위실험의 빈도수를

보여주며, 여기서 $i \in \bar{S}$ 이면, 요소(사람) i 가 표본 S 의 주소지에 거주하지 않는다는 것을 의미한다. 각 행의 최종 열은 앞에 있는 두 행의 합이 아니라, 세 개 범주 $OC, 00, UC$ 에 대하여 추정된 인구의 크기이다.

- 표본 S 에는 모집단 U 에서 선정된 모든 주소지에 거주하는 사람(N_p)이 모두 포함된다. 이 때문에, 크기 n_p 의 표본 S 는 $\frac{n_p}{N_p} \times 100\% = \theta \times 100\%$, 곧 N_p 의 일부에 해당한다. 표본에 선정된 모든 주소지의 모든 사람들이 세 개의 범주 $OC, 00, UC$ 중의 하나에 배치될 수 있다면, 이 모형에서 p_j 의 최대우도 추정량(MLE: Maximum Likelihood Estimator)는 다음과 같다.

$$p_j = \frac{\tau_{j,S}}{n_p} = \frac{\tau_j}{N_p}, j \in \{OC, 00, UC\}$$

- 앞에서 설명한 것처럼, 행정인구 τ_R 은 기지수(既知數), 센서스인구 τ_C 의 참값은 미지수(未知數)이기 때문에, 추정하지 않으면 안 된다. 이미 설명한 것처럼, $p_{iC} = p_{i00} + p_{iuc}$ 이고 $p_{iR} = p_{i00} + p_{ioc}$ 이기 때문에, 독일의 자치단체에서 이원시스템추정량(DSE)에 의한 τ_C 는 다음과 같이 추정한다.

$$\hat{\tau}_C^{DSE-DEU} = \tau_R \frac{\tau_{00,S} + \tau_{uc,S}}{\tau_{00,S} + \tau_{oc,S}} = \tau_R \frac{\tau_{S,C}}{\tau_{S,R}}$$

곧, $\hat{\tau}_C^{DSE-DEU}$ 는 비추정량(ratio estimator)이다.

- 사람들은 주소지에 같이 살고 있기 때문에, 개인수준에서는 주소지 표본은 개인의 집락표본이라고 할 수 있다. 우리가 주소지를 층화하고, 주소지와 그에 속한 사람들을 상이한 확률로 선정하게 된다면, 표본 S 의 주소지 i 는 선정된

주소지의 포함확률(inclusion probability)의 역수(inverse)에 해당하는 설계가 중치(survey weights)를 부여해야 한다. 이 경우에, 이원시스템추정량(DSE)는 다음과 같다.

$$\begin{aligned} \hat{\tau}_C^{DSE-DEU} \\ = \tau_R \frac{\sum_{i \in S} w_i \tau_{i,C}}{\sum_{i \in S} w_i \tau_{i,R}} = \tau_R \frac{\hat{\tau}_C}{\hat{\tau}_R} \end{aligned}$$

- 독일의 이원시스템추정량(DSE-DEU)은 고전적 이원시스템추정량(classical DSE)이 4개의 범주를 갖는데 대하여, 3개의 범주로만 이루어져 있다. 결국, 독일의 이원시스템 추정방식은 등록부 R에 기재되지 않고, 해당 자치단체에 거주하지도 않는 범주는 존재하지 않는다는 가정에서 출발하고 있다.
- 독일의 등록센서스에서는 τ_{00} 를 추정하는 것이 아니라, τ_C 를 추정하는 것이 일차적 목적이 된다. 이 파라미터는 표본 S에 의하여 추정되는데, 표본 S는 미국의 고전적 이원시스템추정량에서 P-sample에 해당하는 것이다. 미국에서와 같이 E-sample, 곧 등록부에 대한 추가적인 표본조사를 실시하지는 않는다. 등록센서스에서 E-sample식의 추가적인 등록부 표본을 뽑아서 점검을 하는 나라는 이스라엘인데, 본 연구의 다른 부분에서 검토한다.

(Chapman 추정량)

- DSE 모형의 동질성(同質性) 가정을 충족시키기 위해서는, 연령별, 성별 또는 다른 특성별 동질적인 범주 $k \in \{1, 2, \dots, K\}$ 을 나누어서, 이들 하위모집

단 각각에 대하여, DSE 모형을 적용하여야 한다. 이 경우, 두 가지 상황이 발생한다.

- (1) 등록부 R의 범주 k에 해당하는 사람이 표본에 없는 경우이다

$$(\tau_{S,R,k} = 0).$$

- (2) 표본으로 선정된 주소지에 범주 k에 해당하는 사람이 거주하지 않는 경우이다($\tau_{S,C,k} = 0$).

첫째의 경우에는 독일의 등록센서스에서 제안된 이원추정량을 적용할 수가 없다. 둘째의 경우, 그것이 표본의 크기가 작아서 생기는 제로 셀(표본 제로-sample zero), 표본의 크기와는 관계없이 생기는 제로 셀인지(구조적 제로-structural zero)인지 분명하지 않다. 구조적 제로의 경우, 모집단에 관찰치가 없어서 제로 값을 부여해도 무방한 셀을 의미한다. 한 셀의 값이 표본 제로인 경우, 구조적 제로가 아니라고 하더라도 표본에는 관찰치가 없고, 이 경우에도 셀 값은 제로를 부여하게 된다.

- DSE 모형의 이러한 한계점 때문에, Chapman(1951)은 수정된 추정방법을 개발하였다. 수정방식은 $\tau_{00,S}$ 에 1을 더하여, 새로운 $\tau_{00,S}^* = \tau_{00,S} + 1$ 로 한다. 또 주변빈도(marginal frequency)도 $\tau_{00}^* = \tau_{00} + 1$, $n_p^* = n_p + 1$, $N_p^* = N_p + 1$ 도 변경된다. 따라서, $\theta^* = (n_p + 1)/(N_p + 1) = n_p^*/N_p^*$ 이다.

- 독일의 자치단체별 특성 범주 k에 대한 τ_C 에 대한 Chapman 추정량은 다음과 같다.

$$\widehat{\tau_{C,k}^{CHAP-DEU}} = (\tau_{R,k} + 1) \frac{\widehat{\tau_{C,k}} + 1}{\widehat{\tau_{R,k}} + 1} - 1$$

위의 추정량에서는 분모나 분자가 제로(0)가 될 수 없기 때문에, 앞에 말한 구

조적 제로가 표본 제로에 의하여 추정이 불가능해지는 것을 피할 수 있다.

3. 대만

- 대만의 등록센서스 전환이 2010년 센서스 라운드에서 이루어졌지만, 2015년 우리나라의 등록센서스 전환이 이루어지는 현재의 상황과 상태가 유사하다. 전반적으로, 우리나라보다 행정자료는 풍부한 것처럼 보이지만, 그 자료의 품질은 그리 우수하지 않은 것으로 판명되었다. 대만의 행정자료 품질은 실제로, 이스라엘의 등록부 품질과 유사한 것으로 알려졌다.
- 대만은 2010년 센서스는 등록센서스로 전환하면서, 행정자료 상의 등록주소(registered address)와 상주주소(usual-residence address)의 불일치 문제를 효율적으로 해결할 수 없으며, 전통적 센서스의 사회경제 분야 조사항목에 대한 상세정보를 포함하지 못하고 있다. 따라서 센서스 목적의 표본조사를 실시할 필요가 있었다. 지리적 관점에서, 센서스는 공간적 의미를 가지며 센서스 기준을 충족하는 모든 통계단위를 면접하는 전수집계다. 한편, 표본조사는 지정된 조사대상자를 면접하기 위한 것이다. 여기서, 조사구는 표본설계에서 추출단위로 사용한다. 추가적으로, 조사구는 동질적이고, 안정적이며, 공간적 의미를 지니기 때문에 상주인구를 추정하는데 도움이 되는 단위이다.

<표 7-6> 대만 2010년 센서스의 향진시구(행정구역)별 총별 할당 방식

항목별	배치총수	향진 수	구성비
총계	-	368	100.00
행정인구수 < 5,000	1	24	6.52
5,000 □ 행정인구수 < 10,000	2	38	10.33
10,000 □ 행정인구수 < 25,000	3	83	22.55
25,000 □ 행정인구수 < 50,000	4	91	24.73
50,000 □ 행정인구수 < 100,000	5	63	17.12
100,000 □ 행정인구수 < 200,000	6	42	11.41

행정인구수 □ 200,000	7	27	7.34

- 표본설계의 타당성을 점검하고, 신뢰할 수 있는 상주인구 통계를 추정하기 위하여, 2000년 센서스와 호구등록파일을 이용하여 다양한 모듈을 개발하였다. 신뢰도가 큰 상주인구를 추정하기 위하여, 모든 일반가구의 구성원은 총화집락추출(stratified cluster sampling) 방법을 사용하여, 표본추출을 했다. 집단시설의 구성원이나 특수시설의 인구는 해당업무를 관장하는 부서가 직접 면접하도록 하였다.

<표 7-7> 행정구역의 표본추출률

조사구별 추출률 (%)	행정구역의 수	퍼센트 (%)
합계	368	100.00
7 ~ 10	54	14.67
10 ~ 20	117	31.79
20 ~ 30	57	15.49
30 ~ 40	29	7.88
40 ~ 50	24	6.52
50 ~ 60	20	5.43
60 ~ 75	23	6.25
100	44	11.96

- 대만 영토 내에는 25개의 현시(농촌/도시지역, 우리나라의 광역자치단체에 해당)가 있으며, 368개의 향진(우리나라의 기초자치단체에 해당함)이 존재한다. 센서스 통계는 향진별로 분류된다. 따라서 향진은 표본설계의 부모집단(副母集團: subpopulation)으로 사용된다. 각 하위모집단의 총수는 행정인구의 총수에 따라 달라진다. 배치된 층(stratum)의 최저치는 1이고 최고치는 7이다(<표 7-6> 참고). 층화 기준은 DH (Dalenius-Hodges) 방법을 사용하고, 각 층의 표본규모는 니만 할당법(Neyman allocation)으로 결정하며, 각 층의 표본 조사구(sample EA)는 계통추출법(systematic sampling)으로 선택하도록 하였다. 표본 조사구 내에서는 모든 가구의 가구원과 주택을 조사대상으로 선

정하였다.

- 부모집단(subpopulation)에 대한 표본조사구의 개수는 신뢰도 95%와 추정오차 3%를 기준으로 하여 결정하였다. 그러나 추출률이 대도시 지역에서 상대적으로 낮았기 때문에, 표본조사구의 수는 상세 통계표를 작성할 때, 추정의 정확성을 고려하여, 표본 조사구의 개수를 추가하였다.

<표 7-8> 대만 2010년 센서스 조사구 통계

		조사구			표본조사구	
		개수	평균 등록 가구수	평균 등록 인 구수	개수	표본추출률 (%)
합계	總計	70,924	110	325	10,747	15.2
타이페이현	臺北縣	12,130	113	318	1,403	11.6
이란현	宜蘭縣	1,322	115	348	245	18.5
타오위안현	桃園縣	6,122	109	328	711	11.6
신주현	新竹縣	1,511	105	336	361	23.9
마오리현	苗栗縣	1,677	102	333	348	20.8
타이중현	臺中縣	4,320	107	361	719	16.6
장화현	彰化縣	3,589	100	364	623	17.4
난타우현	南投縣	1,649	103	320	387	23.5
원린현	雲林縣	2,133	107	337	451	21.1
자이현	嘉義縣	1,549	113	351	332	21.4
타이난현	臺南縣	3,157	116	348	599	19.0
카오슝현	高雄縣	4,136	103	299	679	16.4
핑둥현	屏東縣	2,661	103	329	549	20.6
타이둥현	臺東縣	780	102	297	275	35.3
화롄현	花蓮縣	1,150	104	295	280	24.3
핑후현	澎湖縣	280	119	344	86	30.7
지룽시	基隆市	1,293	113	298	202	15.6
신주현	新竹縣	1,256	111	327	137	10.9
타이중시	臺中市	3,435	110	312	505	14.7
자이시	嘉義市	855	110	319	97	11.3
타이난시	臺南市	2,415	110	318	279	11.6
타이페이시	臺北市	8,174	118	317	885	10.8
카오슝시	高雄市	5,024	115	303	498	9.9
롄장현	連江縣	21	99	474	21	100
진먼현	金門縣	285	112	333	75	26.3

- 한편, 조사구의 표본추출률이 75%를 넘는 경우에는, 해당 향진은 100% 전수 조사를 실시하도록 하였다. 현 1개와 향진시구 44개가 이 범주에 속했다(<표 7-7> 참고). 표본추출률은 모든 향진시구에 대하여 최저치는 7.09%였다. 향시의 조사구의 추출률은 9.9%가 최소치이고, 35.3%가 최대치였다 (<표 7-8> 참고). 조사구의 총수는 7만 924개였으며, 표본으로 추출된 조사구는 1만 747개로, 추출률은 15.2%였다. 실제로, 표본조사의 대상이 된 가구는 118만이었 고, 이것은 인구기준으로 356만 명에 해당하는 수치였다.
- 표본조사가 실시된 향진시구에 대해서는 통계단위(인구, 가구, 주택)의 추정치 를 계산하기 위하여, 별도의 비추정법(ratio estimation)을 사용하는데, 이 추 정법은 일종의 전통적 인구통계기법으로, 통계학의 “고정효과 모형”(fixed effect model)에 해당한다고 할 수 있다. 곧, 표본 조사구(sample EA)의 행 정인구(또는 행정가구)의 총수를 보조변수로 사용하여, 표본조사가 실시되지 않은 향진시구의 경우에 대해서는 현지실사 중 획득한 면접 자료를 토대로 작 성하였다. 시설가구 등의 특수인구(special population)에 대한 통계는 상주지 정보를 이용하여 작성되며, 특수인구의 기록은 사전에 호구등록파일에 삭제하 는 조치를 취하였다.

$$r_{ih} = \frac{\sum_{j=1}^{n_{ih}} x_{ihj}}{\sum_{j=1}^{n_{ih}} y_{ihj}} = \frac{x_{ih}}{y_{ih}}$$

$$\widehat{X_{ih}'} = r_{ih} \times Y_{ih}$$

$$\widehat{X_i} = \sum_{h=1}^{L_i} \widehat{X_{ih}} = \sum_{h=1}^{L_i} r_{ih} \times Y_{ih}$$

$$V(\widehat{X_i}) = \sum_{h=1}^{L_i} N_{ih}^2 \frac{N_{ih} - n_{ih}}{N_{ih}} \times \frac{1}{n_{ih}} S_{s_{ih}}^2$$

$$\begin{aligned} S_{s_{ih}}^2 &= \frac{1}{N} \sum_{j=1}^{N_{ih}} (x_{ijh} - r_{jh} y_{ihj})^2 \\ &= (S_{s_{ih}}^2 + r_{ih}^2 S_{y_{ih}}^2 - 2r_{ih} \rho_{ih} S_{\xi h} S_{y_{ih}}) \end{aligned}$$

$$\rho_{ih} = \frac{S_{xy_{ih}}}{S_{\xi h} S_{y_{ih}}}$$

기호설명:

x_{ihj} : i 번째 하위모집단, h 번째 층, j 번째 센서스 조사구 모항(某項)의 상주인구, 상주가구, 또는 상주주택 특성치의 총수 (특별조사가구는 포함하지 않음)

y_{ihj} : i 번째 하위모집단, h 번째 층, j 번째 센서스 조사구 모항(某項)의 행정인구수 혹은 행정가구수 또는 행정주택수

$\widehat{X_j}$: 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

- 추정과정에서, 내국인과 외국인 통계, 그리고 향진시구(행정구역)별 주택통계는 공적등록파일의 연계로 획득한다. 공적등록파일은 표본조사 자료를 점검하고, 추정 가중치(estimation weights)를 설정하기 위하여 사용한다. 공적등록 파일에 등록되어 있지만, 실제로 상주하지 않는 인구에 대해서는, 표본조사에서 획득한 상주지 정보로 지역 간 전출입 행렬(interregional migration matrix)을 계산하며, 이것을 다시 상주인구를 추정하는데 사용한다.

4. 이스라엘

- 이스라엘 중앙통계국(Central Bureau of Statistics)는 2009년 7월 2일자로, 웹사이트에 2008년 인구센서스 표본조사의 제2차 조사에 해당하는 “전화확인조사”를 통한 자료수집이 종결되었다고 발표했다. 해당 표본조사는 두 단계로 진행되었다.
 - ① 제1단계는, 현장 실지조사로서, 일대일 면접조사가 실시되었다. 이 단계는 2008년 12월 28일부터 3월 8일까지 100일 동안 계속되었다. 이 기간 중, 대략 2,000명의 현장조사원을 동원하여 33만 가구의 130만 명에 해당하는 인구를 대상으로 현장 실지조사를 실시하였다.
 - ② 제2단계는 주민등록의 과다범위를 점검하는데 필요한 정보를 확인하기 위하여 실시되는 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 시작하여, 7월 2일에 종료하였다.

- 이스라엘 중앙통계국은 센서스 표본조사에서 수집한 정보를 이용하여, 이스라엘의 인구 총수(total)와 인구통계, 사회경제적 특성에 대하여 신뢰도가 높은 통계를 작성하기 위하여 노력하고 있다. 2009년 7월 제2단계 조사인 과다범위 점검조사가 끝나고 나서, 이스라엘 중앙통계국은 현장 조사원들에 협력하여 제1단계의 과소범위 점검조사와 제2단계의 과다범위 점검조사가 성공적으로 끝나도록 도와준 이스라엘 국민들에게 감사를 드린다는 이야기를 하고 있다.
- 이스라엘 센서스에서 통합센서스 파일(ICF)은 통합행정파일에 등재된 개인의 인구학적 특성으로 구성된다. 또 통합행정파일에 등재된 개인은 추정 목표모집단에서 실제로 몇 명에 해당하는가를 명시적으로 보여주는 센서스 가중치를 포함한다. 통합행정파일에 등재된 개인의 센서스 가중치가 2.0이면, 그는 추정 목표모집단에서 실제로 2명의 특성을 대표하게 된다. 곧, 해당 개인은 통합행정파일 주소지에 실제로 거주하는 사람 1명과 통합행정파일 주소에 누락되어 다른 주소지에 거주하는 사람 1명을 대표한다.
- 사회경제파일(SEF) 통합센서스 파일은 표본조사에서 집계된 가구와 개인에 대한 자료이다. 이 파일은 인구통계 속성과 사회경제적 속성의 양자를 모두 포함하는 파일이다. 표본 조사구에서 집계된 모든 사람들은 추정 목표모집단의 특성을 대표할 수 있도록 센서스 가중치를 부여받게 된다. 그 결과, 우리가 이스라엘의 수도권(텔아비브)에서 조사된 모든 개인들에 대한 모든 센서스 가중치를 합산하게 되면, 통합센서스 파일 인구 총수(IAF population total)와 사회경제 파일 인구 총수(SEF population total)는 동일하게 된다.
- 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 해당한다. 또, 주소불명으로 인하여 통계처리에 어려움이 있는 사람들의 파일도 작성되어야 한다.

(추정방법론: 개관)

- 이스라엘의 인구대장, 곧 주민등록(Population Registry)은 행정인구의 과다범위(OC)와 과소범위(UC) 때문에, 센서스가 필요로 하는 모집단 인구의 정의에 부합되지 않는다. 전국수준의 과다는 출국이민의 기록으로, 이들은 이스라엘에 더 이상 거주하지 않지만 주민등록에 여전히 기재되어 있는 사람(주민등록인구의 7% 수준)을 포함한다(Statistics Israel, 2005, 2007).
- 이스라엘의 센서스 표본조사는 두가지 기능을 가지는데, 그 중에 하나가 포괄범위를 점검하는 것이고, 다른 하나는 행정자료에 없는 센서스 항목에 대한 통계를 작성하는 것이다. 표본조사의 포괄범위 점검은 과다범위 인구에 대한 정보를 포함하지 않기 때문에, 추가적인 현지실사가 필요하다. 이 과다범위를 점검하는 조사는 영국의 2021년 행정자료 센서스 옵션에서와 마찬가지로, 별도의 조사로 실시된다. 전국수준의 과소범위는 주민등록번호가 없이, 이스라엘에 거주하는 사람들로 구성된다. 이스라엘 통계청은 전국수준의 과소범위를 추정하거나 조정하지는 않는다.
- 이스라엘의 통합센서스에서 포괄범위 오차는 대부분 지역수준에서 발생한다고 가정한다. 사실, 총인구의 25% 정도가 실제 거주지가 아닌 다른 주소에 있는 주민등록부에 기재되어 있다. 이러한 잘못된 주소는 한 지역에서는 과소범위의 문제를 야기하고, 다른 지역에서는 과다범위의 문제를 야기한다. 주민등록부에 남아있는 출국이민이 야기하는 잘못된 기록과는 달리, 이러한 유형의 과다범위는 포괄범위 조사를 실시하여 확인되고, 조정될 수 있을 것이라고 본다.
- 우선, 주민등록부의 주소정보는 다른 행정자료 파일의 주소정보를 바탕으로 갱신한다. 이러한 절차를 거쳐서 낡았거나 잘못된 기록의 문제를 해소하고 나면, 최종 결과물은 통합행정파일이 작성되는, 이 파일은 추정목적의 등록부로 이용될 수 있다.
- 이스라엘은 총인구의 20%를 표본으로 하는 센서스 표본조사를 실시하였다. 이 조사의 목적은 누누이 강조하지만, 이중기능을 가진다. 하나는 주민등록부

를 근간으로 하는 추정치를 보완하는데 필요한 정보를 획득한다. 다른 하나는 역사적으로 long-form 조사표에 해당하는 것으로, 상세한 인구통계, 사회경제통계 정보를 획득하기 위한 것이다²⁾.

- 이스라엘의 2008년 센서스와 마찬가지로, 20% 표본조사를 실시하여, 포괄범위의 점검과 소지역 사회경제통계의 획득은 비용과 효율성을 근거로 고려될 만한 필요가 있는 조치이다. 그러나 몇 가지 현실적인 문제를 고려하여야 할 것이다.
- 우선, “포괄범위 점검”과 “사회경제통계 획득”이라는 두 가지 목적이 상이한 것으로 서로 충돌할 수 있기 때문에, 이를 개별적으로 달성하는데 필요한 조화된 추출틀(harmonized sampling frame)을 최적화하는데 상당한 어려움이 생길 수 있다. 둘째로, 응답부담이란 면에서 상당한 문제가 생기면서, 포괄범위의 점검에 부정적 영향을 미칠 수 있다. 그러나 이들 문제를 극복하기 위하여, 모듈식 설계(modular design)를 실행하는 것이 필요하며, 그것을 가능하게 한 것이 이스라엘의 2008년 센서스라고 할 수 있다.
- 통합행정파일의 포괄범위 오차는 통계지역(SA: statistical area)별로 정의되었다. 가령, 통합행정파일의 과소범위(UC)는 특정의 통계지역에 거주하지만, 통합행정파일에서 다른 통계지역에 기재되어 있는 사람을 의미한다. 또 통합행정파일의 과다범위(OC)는 통합행정파일에서 특정의 통계지역에 기재되어 있지만, 다른 통계지역에 거주하고 있거나(이 경우, 지역수준의 과다범위에 해당함) 또는 아예 해외로 출국하여 이스라엘 영내에 거주하지 않는 사람(이 경우는 전국수준의 과다범위에 해당함)을 가리키는 것이 된다.
- 이스라엘 센서스의 추정절차는 지역수준의 과소범위와 과다범위를 추정하기 위한 일종의 이원시스템 추정모형(DSE)을 이용한다. 추정대상 인구는 주민등록부에 기재되지 않은 사람은 집계되어서는 안 된다고 가정하기 때문에, 전국수준의 과소범위는 존재하지 않는다. 전국수준의 과다범위는 존재하지 않도록

2) 이스라엘의 전통센서스에서 short-form 조사표는 연령, 성별, 주소지, 출생지, 입국연도, 인종/종족, 혼인상태, 종교, 가족/친족관계 등에 대한 문항을 포함하였다.

통합행정파일을 작성하기 때문에, 지역수준의 과다범위와 과소범위를 모두 합산하게 되면 그것은 전국수준에서 제로(0)가 되기 때문에, 각 지역의 과다범위와 과소범위는 이원시스템 추정모형을 사용하여 추정이 가능하다고 볼 수 있다.

- 이스라엘 통합센서스에서 추정인구의 정확성은 (1) 통합행정파일과 현지실사 기록의 독립성 (2) 모집단의 모든 사람들이 통계지역(SA) 통합행정파일에 기재될 확률뿐만 아니라 현지실사에서 집계될 확률이 동일하고, (3) 통계지역 내 조사구별로 통합행정파일 과다집계(IAF OC)의 분포가 “참” 모집단 인구의 분포와 동일하다는 가정이 충족되느냐 여부에 따라 달라진다.
- 현장조사원이 통합행정파일 리스트에 대하여 사전지식이 없기 때문에, 독립성 가정은 충족된다고 볼 수 있다. 나머지 2개 가정은 과다범위나 과소범위의 오류가 비슷하도록, 모집단을 동질적 집단으로 구분함으로 해서 충족된다고 볼 수 있다.
- 이원시스템 추정방법은 개별조사구별로 두 개의 리스트를 필요로 한다. 하나는 조사구(EA)에 포함된 주소지의 통합행정파일에 기재된 사람들의 리스트이고 다른 하나는 조사구의 현장조사에서 실제로 집계된 사람들의 리스트이다. 현장조사원에게 조사대상이 되는 주소지에 거주하는 사람이나 가구에 관한 통합행정파일 명부를 제공하지 않기 때문에, 통합행정파일 리스트와 현장조사 집계 리스트는 상호독립이라고 가정할 수 있다.
- 표본조사가 실시되고 나면, 통합행정파일 리스트와 표본조사 리스트는 매칭작업을 실시한다. 매칭작업은 주민등록번호(ID number)와 성/연령 등의 핵심변수를 이용하여 전국수준에서 실시된다. 이것은 두 개 리스트에 동시에 포함되어 있는 사람들과 하나의 리스트에는 포함되어 있지만, 다른 리스트에는 포함되어 있지 않은 사람들을 구별한다.
- 이스라엘 통계청은 통합행정파일 리스트에 기재되어 있지만 표본조사에서 집계되지 않은 사람들을 찾아내는 절차를 진행하였다. 이 조사는 등록부 과다범

위 점검조사(OC check survey)³⁾라 불리는 것이다. 이 조사에 의하여 (1) 통합행정파일에 기재된 사람이 자신의 통합행정파일 통계지역에 실제로 거주하지 않고, 다른 통계지역 또는 해외에 거주함으로써 발생하는 “통합행정파일 과다범위”(IAF overcoverage)와 (2) 표본 조사구(EA)의 통합행정파일에 기재되었으나 점검조사에서 조사되지 않은 “통합행정파일 과다”(IAF excess)로 구분하였다.

- 추출된 조사구(EA)의 통합행정파일에 기재되었으나 점검조사에서 조사되지 않은 사람을 찾아내는 점검조사는 (1) 통합행정파일에 잘못된 주소지, 잘못된 장소에 기재된 사람을 정정하고, (2) 이스라엘 영내에 거주하지 않는 사람들을 정정하는 작업((1)과 (2)는 과다범위로 분류된 것들임)을 포함하게 된다. 이들의 주소지 정보는 몇 단계를 거쳐서 획득하도록 되어 있다. 먼저, 현장조사원은 통합행정파일에는 기재되어 있지만 집계되지 않은 사람들의 리스트를 가지고 현장으로 가서, 그들을 찾거나 이웃사람들로부터 그들의 주소지 정보를 획득한다. 다음으로 이들 중 주소지를 확인할 수 없는 사람들에 대해서는 전화로 주소지를 확인하고자 노력한다.
- 이스라엘 통계청은 주민등록파일을 다른 행정자료 파일(자동차등록, 전력이용료 계정, 초중고교 학생파일)과 비교하여, 주소정보를 정정하려고 노력한다. 국내체류 여부는 출입국 파일의 정보를 이용하여 정정한다. 주민등록파일 업데이트를 이용하여, 당초 파일추출 이후 발생한 출생건수와 사망건수를 확인한다. 다른 통계지역 전출자나 해외로 출국한 사람은 통합행정파일 기재 오류(IAF entry error)로 분류한다.

<표 7-9> 이스라엘의 이원시스템추정법(DSE)

	통합행정파일(IAF)	
	Yes (집계)	No (누락)

3) 이것은 센서스 사후조사에서 흔히들 E 표본(Enumeration Sample)이라고 불리는 것이다. 한편, 과소범위를 조사하는 센서스 표본조사는 P 표본(Population Sample)이라 불리는 것이다.

표본조사	Yes (집계)	n_{11}	n_{01}
	No (누락)	n_{10}	n_{00}

주: 이스라엘의 이원시스템 추정방법은 통계지역(SA)별로 적용되며, 통합행정파일과 표본조사의 리스트는 통계지역(SA: statistical area)별로 비교된다.

- 이스라엘의 2008년 센서스에서, 특정 주소지에서 추출된 모든 사람은 (1) 통합행정파일 주소지와 현지실사 주소지가 일치하는 경우 (2) 통합행정파일에는 기재되어 있으나 현지실사에서는 집계되지 않은 경우 (3) 현지실사에서는 집계되었으나 통합행정파일에는 기재되지 않은 경우로 구분된다.
- 이원시스템 추정방법 중에서도 고전적 방식을 사용하면, 이들 범주 중 처음의 경우는 n_{11} , 마지막 경우는 n_{10} 이다. 그러나 두 번째 경우는 과다범위(X), 곧 여타 통계지역 (SA)에 거주하거나 해외에 거주하는 사람들을 조정하여야 한다. 따라서 n_{01} 은 통합행정파일 집계결과에서 과다범위를 뺀 수치와 동일하여야 한다($n_{01} = IAF - X$).
- 여기서 중요한 것은 현지실사에는 오류가 없다고 하며, 가능하면, 가구방문에 의한 직접면접을 실시하는 것이 이스라엘 통계청면접조사 지침이다. 전국수준에서는 X는 통합행정파일 기재오류, 곧 해외출국자를 센서스 모집단에 포함시킨 오류만을 포함하도록 되어 있다. 한편, 지역수준에서는 X는 전국수준의 통합행정파일 기재오류와 지역수준의 과다범위(local OC), 곧 사람들이 실제로는 다른 조사구(EA)에 거주함으로 해서 발생하는 오류를 모두 포함하도록 되어 있다.
- 이스라엘의 2008년 센서스에서, 주어진 통계지역에 대한 통합행정파일의 과소범위(IAF UC)는 해당 통계지역에 거주하지만, 통합행정파일에는 다른 주소지에 기재되어 있는 사람을 가리킨다. <표 6-9>의 분류표를 이용하여, 과소범위는 다음과 같이 추정한다(Kamen, 2005).

$$\begin{aligned}\hat{p}_{1+} &= \frac{\text{통합행정파일과 현지실사가 일치하는 사람의 수}}{\text{현지실사에서 집계된 모든 사람의 수}} \\ &= \frac{n_{11}}{n_{11} + n_{10}}\end{aligned}$$

- 통합행정파일 과다범위(IAF OC)은 주어진 통계지역에 대하여, 통합행정파일에 기재되었지만, 그 지역에 거주하지 않는 모든 사람들로 구성된다. 이것은 통합행정파일 집계오류(erroneous enumeration)(이것은 통합행정파일 과다, “IAF excess”를 뺀 수치)와 <표 6-6>의 분류표를 결합하여 추정한다. 분모는 추출된 셀의 크기에 대한 추정량(estimator of the sampled cell size)이다(Kamen, 2005).

]

$$\begin{aligned}\hat{\lambda}_{1+} &= (\text{통합행정파일의 집계오류}) \div \\ &\quad \{(\text{통합행정파일과 현지실사가 일치하는 사람의 수}) \\ &\quad + (\text{통합행정파일의 순과다})\} / \text{과다범위} \\ &= \frac{X}{(n_{11} + n_{01}) / \hat{p}_{1+}}\end{aligned}$$

- 과다범위 조정 파라미터($\hat{\lambda}_{1+}$)는 과다범위 기록(X , 전국수준의 잘못된 과다범위와 지역수준의 과다범위)을 이미 과소범위($\hat{p}_{1+}$)를 조정한 통합행정파일 집계결과($n_{11} + n_{01}$)로 나누어 계산한다. 이것은 등록부의 과다범위가 문제가 되지 않은 북유럽이나 여타 지역과 달리, 이스라엘 통계청은 전국수준의 과다범위는 무시할 수 있다고 가정할 수가 없었기 때문이다. 이 문제를 다루기 위해서, 전국수준의 과다범위를 추정하여, 이원시스템 추정법에서 이용하는 수치에서 그들을 제(除)하는 방식을 채택하였다. 더욱이, 지역수준의 과다범위도 이원시스템 추정 수치에서 제(除)하였고, 이 때문에 과다범위에 대한 전

반적 조정이 이루어졌다고 할 수 있다.

- 통계지역별로 지역수준의 과소범위와 과다범위의 추정치는 통합행정파일의 주소패턴에서 차이가 있는 인구통계집단별로 계산하도록 되어 있다. 과다범위와 과소범위의 추정치는 통합행정파일에 기재된 사람의 수와 결합되어, 인구규모를 추정하게 된다. 센서스 가중치(census weight)는 통합행정파일의 개별기록에 대해서 모두 작성되고, 그 가중치는 개별기록이 몇 명의 실제 사람을 가리키는가를 의미하는 수치로서 센서스 통계에서 대단히 중요한 역할을 하게 된다. 어떤 인구집단에 대한 센서스 추정치는 그 구성원에 할당된 통합행정파일 “센서스 가중치의 총합”(the sum of the IAF census weights)이다.

제3절 센서스의 소지역 추정문제

- 등록센서스에서 표본조사를 이용하여, 인구, 가구, 주택 총수와 같은 센서스 모집단의 크기를 추정할 때, 행정구역의 크기가 큰 지역의 경우에는 설계 기반 직접추정을 이용하거나, 앞에서 언급한 이원시스템 추정이나 비추정이 모형 기반 추정방법으로 이용된다. 소지역 추정(small-area estimation)은 대규모의 표본조사에 관심 부모집단으로 표본의 크기가 작은 소지역이 포함되어 있을 때, 그 부모집단의 파라미터의 추정을 포함하는 다양한 통계적 방법들을 총괄하는 것이다.
- “소지역”이란 용어는 미국의 경우, 군단위 지리적 지역(geographic area)를 가리키거나, 때로는 특정 지역 내 인구특성 이 경우는 소집단(small domain)을 의미하기도 한다. 그러나 표본 자체로서는 통계적으로 유의미한 추정이 가능하지 않은 소규모의 지역을 의미하는 것으로 사용되기도 한다. 곧 모집단 전체를 대상으로 하는 표본조사를 실시하더라도, 특정의 지역 안에서의 표본은 크기가 너무 작아서, 그 자료에서 통계적으로 유의미한, 곧 정도(精度) 상 문제가 없는 추정치를 산출하는 것이 쉽지 않은 경우를 말한다. 이 때문에, 그러한 지역에 대하여 파라미터를 추정하고자 하는 경우에는 존재하는

등록부 자료나 전통적 센서스에서 도출된 추가정보를 이용하여 추정치를 작성하지 않으면 안 된다.

1. 인구학적 추정과 통계학적 추정

- 인구주택총조사의 소지역 추정과 관련하여, 인구학에는 두개의 서로 다른 흐름이 있음을 주지할 필요가 있다. 하나는 출생, 사망, 이동(전입, 전출)과 같은 인구변동요인을 고려하여, 총인구를 추정하는 직접추정으로 인구균형방정식(demographic balancing equation)과 같은 모형에 의존하는 경우가 많다. 다른 하나는 간접추정으로 사회경제적 변동요인을 고려하여 총인구를 추정하는 방식으로, 표본조사에 의존하는 경우가 많다.
 - ① 직접추정: 주어진 인구의 귀속적 특성(성별, 연령, 인구 등)은 물론 총인구를 추정하기 위하여 개발되는 방법
 - ② 간접추정: 인구의 성취적 특성(교육수준, 고용상태, 소득, 혼인상태)을 추정하기 위하여 사용되는 방법
- 센서스 추정방법론에서는 통계학적 추정 정의(estimation definition)를 따르는 것이 보통이다. 통계학 특히 표본설계(sample design)의 입장에서 보는 직접추정은 인구학자들의 추정 정의로는 간접추정에 해당한다고 볼 수 있다. 그 이유는 인구추정 목적에 이용되는 자료가 표본조사에서 나온 자료로서, 인구변동의 징후군에 해당하는 변수(예: 주택가격, 초등학교 재학률)들을 사용하며, 관심을 기울이는 현상을 직접적으로 대변하지 않기 때문이다.

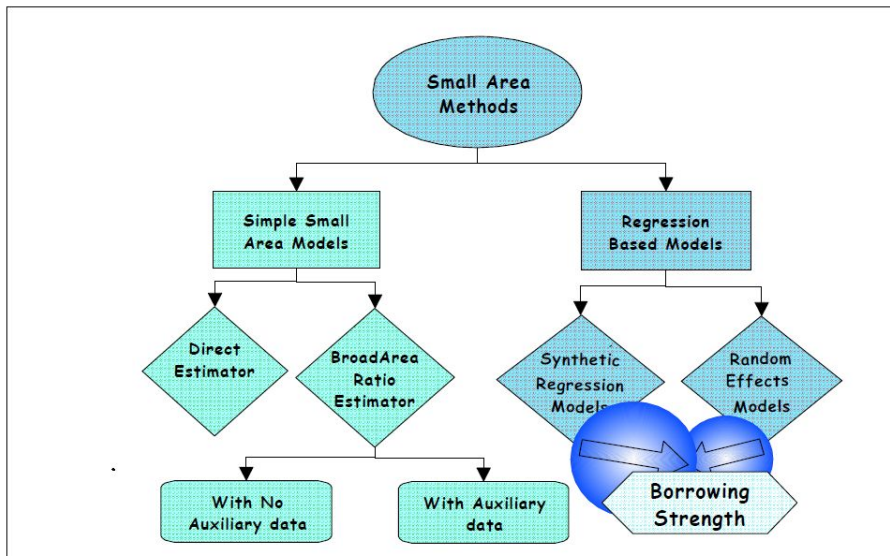
2. 소집단 추정의 문제

- 보통 사람들은 인구주택총조사, 특히 전통적 센서스의 경우에는 집계(counting, enumeration)의 연습으로 생각하는 경향이 있다. 그러나 센서스의 개념을 행정자료와 표본조사를 결합하여, 센서스 모집단의 크기를 확정할 때에는, 이들 수치는 좀 더 적절하게 추정으로 보아야 하며, 그렇게 되면, 편향-분산 트레이드오프(bias-variance tradeoff)를 포함하여, 무수히 많은 통계적 쟁점이 전면에 부각된다.
 - 등록센서스 특히 행정자료와 조사조사를 결합하는 경우, 소지역 추정(small-area estimation)의 목적은 센서스에서 표본조사로 실시된 조사항목에 대하여 통계작성단위가 소지역일 때, 그 지역에 대한 신뢰성 있는 추정치를 작성하는 것이다. 여기서 주요 과제는 소지역의 경우 표본수가 작다는 문제를 극복하고, 추정치의 품질을 개선하기 위하여 혁신적인 접근방식을 사용하는 것이다. 소지역 추정은 조사자료//다른 자료들 간에 내재하는 관계를 설정할 수 있을 때 그 방법론이 효율적일 수 있다. 이 때, 소집단 추정목적으로, 그 관계의 장점을 차용하는 것이라고 볼 수 있다. 이들 관계는 <그림 7-5>에 도식화되어 있는데, 이를 간단히 설명하면 다음과 같다(Australian Bureau of Statistics, 2003; Rao, 2003).
- (1) 조사기반 직접추정치와 행정자료, 센서스 또는 여타 자료원에서 나온 보조 정보와 존재하는 관계를 이용한다.
 - (2) 시계열적으로 관찰되는 직접추정치 간의 상관관계를 이용한다.
 - (3) 이웃하는 소지역(neighboring areas)들 간의 공간적 관계를 이용한다.
 - (4) 상이한 소지역이기는 하지만 좀 더 넓은 지역에서 관찰되는 유사특성을 가진 통계단위들 간의 횡단적 관계를 이용한다.

(5) 위의 어떠한 조합이라도 활용한다.

- 모든 소지역 추정 문제는 접근방식이나 기법 자체가 등록센서스의 추정문제와

<그림 7-9> 소집단 추정모형 구축의 프레임워크



관련되는 경우에는 당면과제에 적합한가를 확인하기 위하여 세심한 평가를 거쳐야 할 필요가 있다.

- 소지역 방법은 소지역 수준에서 보조정보의 이용가능성이나 보조정보나 추정대상이 되는 변수 간의 관계에 의존하게 된다.

주: Small Area Methods = 소지역 방법
 Simple Small Area Models = 간단한 소지역 모형
 Regression Based Models = 회귀 기반 모형
 Direct Estimator = 직접추정방식
 Broad Area Ratio Estimator = 광범한 지역 비 추정방식
 Auxiliary data = 보조자료(정보)

Synthetic Regression Models = 합성 회귀모형
Random Effects Models = 임의효과 모형
Borrowing Strength = 장점의 차용(원용)

- 본질적으로, 소지역 수준에서 추정결과의 정확성을 개선하기 위해서는, 보조정보에서 “borrowing strength”의 가능성을 탐색하는 것이다.
- 물론, 소지역 모형은 단순한 모형에서 대단히 복잡한 모형에 이르기까지 다양하며, 복잡성이 증가할수록 많은 시간, 노력, 기술, 정보의 요구량이 증가한다.
- 양적 방법과 질적 방법이 주어진 방법에 대한 최고의 모형을 선택하기 위하여 선택되어야 한다. 모형구축의 기본적 프레임워크는 다음과 같이 요약할 수 있다.
- 직접추정량(direct estimator)은 센서스 추정방법론에서는 모든 소지역에서 선택된 표본단위에 조사가중치(survey weights)를 적용하여 추정치를 획득하는 방법을 가리킨다.
 - 우리나라에서도 대부분의 통계조사는 상당수가 전국수준 또는 시도별(광역자치단체) 수준에서만 신뢰도가 높은 추정치를 제공할 수 있도록 표본설계가 되어 있기 때문에, 표본의 크기가 작은 경우에는 신뢰도가 높은 추정치를 시군구 또는 읍면동 수준에서 작성한다는 것은 대단히 어려운 일이다.
 - 더욱이, 특정 수준에서 표본이 아예 없는 경우도 생겨날 수 있다. 이러한 경우, 소지역 모형을 구축하는 작업을 행하는 경우, 여타 추정치와 비교목적으로, 직접추정치를 작성하는 경우가 있다.

- 이것은 최종적으로 편향이나, 가법성(additivity)에 관련된 품질진단을 하는데 도움이 될 수도 있다.
- 상위지역의 비를 하위지역에 적용하는 비추정량(broad area ratio estimator): 이것은 사실상 가장 간단한 소지역 추정모형의 하나라고 할 수 있다.
 - 이것은 통계조사에서 획득한 비율을 단순히 행정자료에서 획득한 소지역 인구에 적용하여 계산할 수 있다. 물론, 이 비추정량의 성공여부는 상위지역을 어떻게 선택하는가에 의존하게 될 것이다.
 - 상위지역은 신뢰성이 있는 직접추정이 가능할 정도로 범위가 넓어야 하고, 하위지역은 상위지역 안에서 관심있는 항목과 관련하여 동질성을 유지할 수 있을 정도로 작아야 한다.
 - 사실, 이것은 현실과 거리가 있는 가정이 될 수도 있고, 이 때문에 통계이용자는 이 문제를 잘 인식하여야 할 것이다.
- 센서스 추정방법론에서 소지역 추정을 성공적으로 진행하는데 필요한 조건은 반응변수(예: 총인구, 과다집계 또는 과소집계의 정도)와 높은 상관관계가 있는 정확한 보조정보(auxiliary information)의 입수여부이다. 여기서 보조정보는 추정모형에 포함되어야 할 하나 또는 둘 이상의 변수로서, 등록센서스에서는 일차적으로 등록부에서 도출된 변수라야 할 것이다. 보조정보의 요건으로서 고려되어야 할 사항은 다음과 같다.
 - (1) 센서스 추정방법론 특히 추정지역이 나라 전체나 광역자치단체가 아니라, 기초자치단체 또는 기초자치단체의 일부인 경우에 소지역추정방법이 적용되어야 하는 전체 모집단을 포괄하여야 한다. 모집단의 특정 지역에 보조정보가 입수 불가능한 경우에는, 그 지역에 대해서 소지역 예측이 불가능하며, 그런 보조정보가 없는 지역은 예측모형에서 사실상 제외되어야 마땅할 것이다.

(2) 특정의 소지역에 소속한 모든 단위가 정확하게 파악될 수 있도록 하기 위하여, 신뢰도가 높은 지리정보(geographic information)를 포함하여야 한다.

(3) 반응변수와 모형에 사용되는 보조정보는 동일시점의 자료여야 한다. 이 점은 등록부 자료의 주기별로 항상 업데이트되어야 할 필요가 있음을 말해준다고 할 수 있다.

- 보조정보를 이용하는 비추정량(ratio estimator with auxiliary data): 이것은 관련 변수와 상관관계가 있는 정보를 이용하는 것으로, 그 정보는 소지역 간의 구성상 차이점을 조정하는 추정치를 도출하기 위하여 소지역 수준에서 이용한다.

- 가령, 연령이나 성별 특성이 특정지역에 거주하는 인구의 크기와 관련된다 고 하는 것이 해당지역의 포괄범위 오차(coverage error)와 관련된다 고 하는 사실이 있을 수 있다. 가령, 젊은 연령층의 인구가 많으면 많을수록 포괄 범위의 오차가 클 수 있다.

- 이 경우에는 연령별 범위오차율이 연령별로 분류된 소지역인구에 그대로 적용될 수 있다. 이것은 흔히들 말하는 결정론적 모형(deterministic model)로서, 포괄범위는 연령변수에 의해서만 결정되고 다른 변수의 영향은 없다고 가정한다.

- 이것은 상위지역의 비를 하위지역에 적용하는 비추정량과 관련하여 사용될 수 있다. 이 경우에도, 기본가정은 대단히 강한 것으로, 그 가정을 그대로 적용하는데 세심한 현실적 판단이 필요할 것이다.

- 회귀모형(regression based model): 이것은 다양한 변수를 기초로 해서, 소지역 추정을 가능하게 한다. 이 다변량 분석은 지역수준이나 또는 분석단위(개인/가구/주택) 수준 또는 이들을 결합한 수준에서 적용될 수 있다.

- 이 모형형태는 자료의 특성(연속변수, 계수자료, 범주자료 등)을 반영할 필

요가 있을 것이다. 모형들은 “random effects”를 포함하도록 할 수 있다.

- 이 방법은 해당 소지역의 표본이 작지만, 지역효과를 추정하는데 이용할 수 있는 충분하지는 않지만, 어느 정도의 표본이 있는 경우에 적용이 가능하다. 지역자체의 특성을 허용하지 않는 모형은 흔히들 “합성모형”(synthetic model)이라고 불린다.
- 합성 회귀모형(synthetic regression model)은 보조변수와 반응변수 간에 결정론적 관계(deterministic relationship)를 수리적으로 표현하기 위하여, 이용 가능한 모든 보조정보를 이용한다.
 - 이 추정모형은 반응변수의 체계적 편차(systematic variation)는 모두 보조 정보가 취하는 값에 차이에 의하여 설명된다고 가정한다.
 - 곧 “Random noise” 또는 “stochastic variation”으로 언급되는 나머지 차이는 모형에 의하여 예측된 수치와 자료의 실제값의 차이에 의하여 설명된다고 가정한다. 이 차이는 흔히들 “random error” 또는 잔차(residuals) 또는 교란요인(disturbance error)이라고 한다.
- “임의효과” 모형은 소지역 추정에 “임의효과”(random effect)를 포함하여, 합성효과(synthetic effect) 모형을 개선한다. 이것은 보통 상수항과 오차항을 포함시켜서 추정을 하는 것을 말한다. 이것의 해석은 모든 소지역에 통용되는 상수항을 포함하기 때문에, 그것을 기준으로 해서 보조변수가 영향을 미치는 것이다. 이것은 지역 간 편차를 고려하는데 충분하지만, 어떤 경우에는 상수항보다는 파라미터 계수에 “임의효과”(임의효과)를 포함시키기도 한다. 물론 이러한 모형은 복잡성을 증가시키며, 센서스의 소지역 추정모형에서 사용되지 않은 수도 있다.
- 소지역 추정 모형에서, “임의효과”의 추가는 추정결과가 더욱 더 정교해지고, 강건함(robustness)을 보일 수 있다.

- 선형모형의 경우는 “임의효과” 모형은 이론적으로 직접추정치의 정확성을 개선하고, “합성효과” 모형과 관련된 불확실성을 줄이는 소집단 추정 결과를 제공한다는 점에서 장점이 있다.
- 총오차(표본추출 오차와 비표본추출 오차의 합계) 대비 추출오차가 상대적으로 작은(곧, 표본의 크기가 상대적으로 큰) 소지역의 경우, random effect model은 직접추정치(direct estimates)에 더 많은 가중치를 부여한다.
- 한편, 표본추출 오차가 큰 지역의 경우에는 모형 기반 추정치(model based estimates)에 더 많은 가중치를 부여하는 것으로 알려져 있다.
- 캘리브레이션 추정량(calibration estimator): 캘리브레이션 추정량을 구하기 위해서는, 원래의 조사가중치(보통, 표본 포함확률의 역수)는 새로운 “캘리브레이션” 가중치로 대체하는데, 이 가중치는 원래의 가중치와 그리 큰 차이가 나지 않도록 하고, 모집단에 대하여 입수가 가능한 보조변수를 이용하여 계산한다.
 - 이 보조변수에 대한 소집단 추정치는 캘리브레이션된 가중치를 이용하여 계산하면, 기지수인 모집단 합계(population total)와 동일하다. 캘리브레이션의 간단한 예는 소지역별로 추정된 성/연령별 인구이다.
 - 이 경우, 조사가중치는 캘리브레이션하여, 성/연령별 인구규모가 기지(既知)의 인구수와 일치하도록 조정한다.
- 올해의 통계청 등록센서스 용역연구사업의 핵심과제 중 하나는 센서스 추정방법론, 특히 소집단 추정방법의 성공여부를 결정하는데 있어서 다양한 요인들이 작용하게 될 것이고, 이들 요인들을 고려하여 등록센서스에서 이용될 소지역 추정방법을 개선하여 나가지 않으면 안 될 것이다.

(1) 통계청과 이용자의 상호작용(interaction between NSO and statistics users): 통계청은 소지역 추정결과를 어떻게 사용할 것인가를 알고 있는 이용자와 같이 일할 수 있는 능력을 키워야 하며, 그것은 궁극적으로 이용자의 요구사항을 확인하고, 추정결과의 타당성을 확인하는데 적합할 것이다.

(2) 관심변수(variable of interest): 이들 변수는 일반적으로 흔한 인구특성이어야 한다. 가령, 모집단의 전체 비중 중에서 10%가 안 되는 희귀한 특성은 모형화하기가 쉽지 않으며, 결국은 신뢰도(reliability), 또는 정도(精度 precision)의 수준이 낮은 소집단 추정치로 끝나버리기 마련이다. 가령, 이들 변수의 특성을 분해하는데 가능하면 최소한으로 축소하여야 한다. 또 소지역 수준의 추정결과가 일관성이 있어야 하며, 비교적 간단한 “합성” 모형으로 추정하는 것이 쉬워진다.

(3) 소지역의 크기(size of small area): 소지역의 표본수가 직접추정(direct estimation)에는 적합하지 않더라도 어느 정도가 확보되어야지, 추정결과의 신뢰도는 높아지기 마련이다. 그러나 실제로 소지역 자료가 이용자에게 의미 있을 필요가 있기 때문에, 이러한 이상적 목표는 달성되지 못하는 경우가 많다. 일반적으로 소지역의 크기가 작아지면 그 지역의 추정치는 신뢰도가 떨어지기 마련이다.

일반적으로, 연구자들은 인구규모가 최소 20,000명 이상의 지역에 대해서만 소지역추정의 결과를 공표하도록 하고 있다. 그러나 많은 소지역 추정에서는 이 보다 작은 지역에 대해서도 추정결과를 작성해야 할 필요가 있을 것이다. 이 경우에는 보조정보와 상관관계(correlation)가 대단히 크다고 가정하지 않으면 안 될 것이다. 어떤 지역은 크기가 너무 작기 때문에, 추정결과에 대한 신뢰도에 확신을 가지는 못하는 경우, 최종 통계작성을 유보해야 할 필요가 있을 것이다.

(4) 보조정보(auxiliary data): 등록센서스에서는 등록부나에 포함된 기재사항을 보조정보로 하고, 센서스 모집단의 반응변수(예: 총인구, 과다집계, 과소집계)를 추정한다. 우선, 지역별 인구규모는 표본조사 자료에서 얻은 인구규모와

의미있다고 가정하는 행정자료의 인구규모가 있어야 한다. 반응변수는 행정자료나 표본자료는 모든 소지역에 존재해야 하며, 원칙적으로 정확한 방법으로 수집되어야 할 필요가 있을 것이다. 또, 행정자료와 조사자료에서 마련된 관계가 조사자료가 없는 지역에 적용되는 경우에는, 비슷한 포괄범위와 개념정의가 필수적이라고 할 수 있을 것이다.

- 등록센서스에서는 소지역 추정치는 통계모형을 표본조사가 실시된 지역 곧 행정자료와 표본자료를 조합하여 새로운 반응변수를 만들 수 있는 지역에 적합시키고, 이 모형을 표본조사가 실시되지 않은 소지역에 대해서는 소지역 수준에서 행정자료 수준에 입수 가능한 보조정보를 적용함으로써 얻을 수 있다. 물론, 이 경우에 보조정보의 다양한 조합을 고려하여 몇 가지 후보모형(candidate model)을 생각할 수 있을 것이다. 이 후보모형 중에서 몇 가지 기준을 정하여 가장 신뢰할만한 것을 최종 모형으로 선택하는데, 그 기준은 다음과 같다.

(1) 종전의 연구나 일반적 상식에 비추어보아 어느 정도 후보모형이 타당성이 있는가?

(2) 각 모형이 관찰된 자료를 어느 정도 잘 적합시키는가?

(3) 각 후보모형에서 예측된 소지역추정치의 정확성은 어느 정도인가?

- 후보모형 중에서 어느 것을 사용할 것인가에 대해서는 위의 다양한 기준을 사용하여 평가를 해야 할 것이다. 후보모형 중에서 최고의 모형을 선택하면, 통계이용자에게 모형자체의 평가는 물론, 그 모형에서 얻은 소지역 추정치의 품질을 평가하도록 하는 것이 중요하다. 센서스 통계품질의 측정기준은 소지역 추정의 정확성, 타당성, 일관성 등이 있을 수 있다.

(1) 소지역 추정결과를 직접추정 결과와 비교하는 편향성 테스트(bias test)

(2) 모형 가정과 적합도의 정도에 대한 검정

(3) 소지역 추정결과를 합산하여, 그것이 상위 소지역 결과와 동일해지는가에 대한 점검

(4) 소지역 추정결과와 다양성에 대한 현지 전문가들의 의견청취

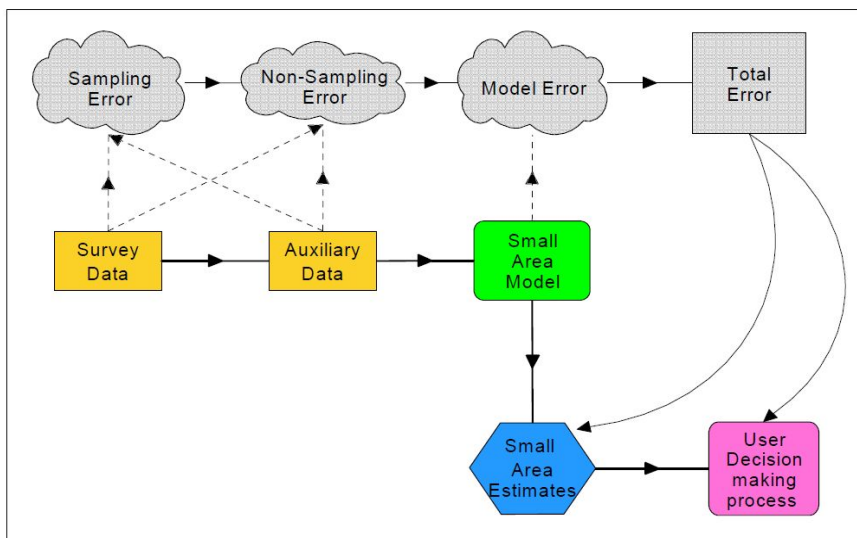
(5) 평균제곱오차의 제곱근(RMSE: root mean squared error)을 모델추정치로 나누어서 그것을 평가하는 작업: 여기서 평균제곱오차의 제곱근(RMSE)은 조사추정치에 대해서 계산한 표본추출오차와 유사하다.

- 센서스 추정결과와 품질측정은 소지역 후보모형의 상대적 성능을 비교하는데 있어서 대단히 중요하지만, 소지역 추정결과를 최종적으로 선택하는데 있어서, 현명한 판단이나 전문가의 의견을 듣는 것이 대단히 중요할 것이다.
- 등록센서스에서 전국수준의 자료가 아닌 소지역에 대하여 다양한 추정방법을 사용하는 것이 불가피하다. 그러나 이들 추정결과는 다른 국가통계와는 별도로 취급되어야 하며, (1) 그 방법과 가정들을 최종 발표에서 상세하게 기술하지 않으면 안 될 것이다. (2) 그들의 타당성은 이용자가 통계작성 결과에 대하여 신뢰성을 가질 수 있도록 정확하게 평가되어야 한다. (3) 이들의 기본모형은 이용자가 이해할 수 있도록 기술되어야 하며, 기본가정의 타당성은 핵심 이용자와 토론의 대상이 되어야 한다. (4) 작성결과의 품질은 양적인 방법으로 기술되어야 하며, 모형이 대단히 복잡하고, 방법론의 선택이 상당한 경우에 모형에 대한 심사평가(peer review)가 있어야 한다.
- 등록센서스의 추정방법론 개발에서 소지역 추정은 그 자료를 생산하는 방식에 따라 다양한 오류를 포함할 수 있음을 주지할 필요가 있다. 물론, 그 오류는 추출오차(sampling error), 비(非)추출오차(non-sampling error), 모형오차(model specification error) 등 적어도 추정결과에 영향을 줄 수 있는 3 종류의 오류를 말한다(<그림 7-10> 참고). 이들 오류가 특정 소지역 추정에 영향을 미치는 정도는 소지역 추정방식이나 기초자료의 종류에 의존한다.
- 센서스 통계표 작성결과를 소지역 추정의 형태로 이용자에게 공표할 때, 보고

서는 부록에 이용된 개념을 추정방식을 명확하게 설명해야 한다. 보고서에 수록해야 하는 내용은

- (1) 기본문제, 추정치의 포괄범위와 적용가능성에 대한 소개
- (2) 소지역 추정의 다양한 절차에 대한 개관(특정 모형, 관련 변수, 기본가정)
- (3) 다양한 세트의 소지역추정치에 관련된 품질문제를 확인하여 주는 내용
- (4) 소지역 추정결과를 사용하는 방법에 관한 필요한 지침(가령, 소지역 추정의 한계) 주요 쟁점과 권고사항(예: 소지역 추정결과의 집계, 추정지역의 특징)

<그림 7-10> 소지역 추정에서의 오류의 원천



주: Sampling error = 표본추출 오차
 Non-sampling error = 비 표본추출 오차
 Model error = 모형설정으로 인한 오차

Total Error = 총오차, 곧 표본추출과 관련된 오차, 관련되지 않은 오차,
 모형설정 문제로 인한 오차의 총합계
 Survey Data: 조사자료
 Auxiliary Data: 보조정보
 Small Area Model 소지역 모형
 Small Area Estimates 소지역 추정치
 User Decision making process: 통계이용자가 추정결과에 대하여
 결정을 내리는 과정

3. 추정결과의 품질진단

- 센서스 추정결과의 품질진단(quality diagnostics)은 다음과 같은 사항들을 포함하여야 마땅할 것이다.

- (1) 소지역 추정량의 값을 작성하는데 사용한 모형의 종류, 이들의 설득력, 타당성, 적합성의 정도
- (2) 다양한 품질지표를 이용하여 작성한 소지역 추정치의 성능 평가
- (3) 소지역 자료를 적용하거나 해석할 때 이용자들이 의식해야 하는 여타 품질상의 쟁점

물론, 상세한 기술보고서(technical report)가 작성되어, 행정자료를 기반으로 하는 센서스 통계품질에 대한 견해를 담아야 할 필요가 있음은 두말할 필요가 없을 것이다.

- <그림 7-7>에서 보듯이, 양질의 보조정보는 양질의 센서스 추정치를 획득하는 전제조건이다. 물론, 센서스 통계의 품질은 상대적 용어이며, 결국은 정부나 기업, 또는 일반 통계이용자가 어떠한 결정을 내리는가에 의존하는바가 크다

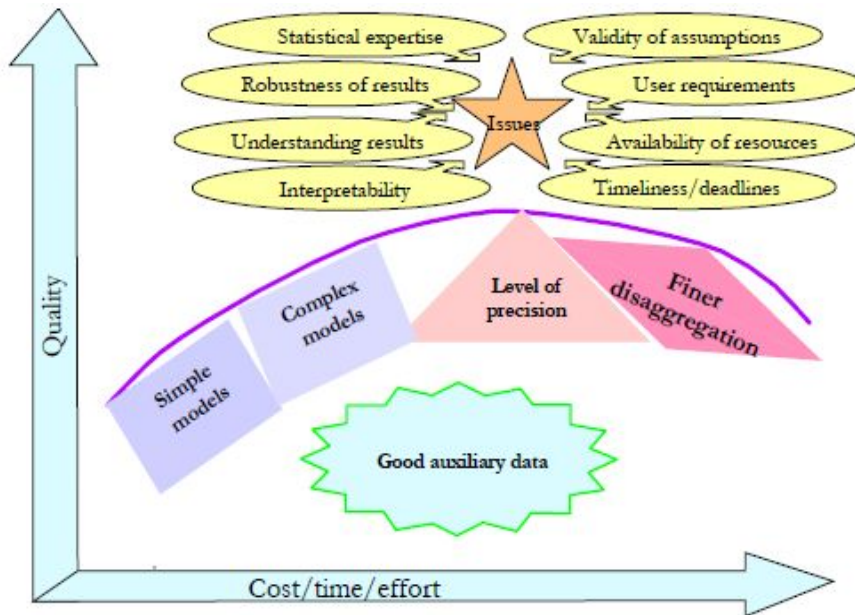
고 할 수 있다.

- 통계청이 등록센서스에서 양질의 보조자료를 보유하고 있거나 그러한 자료에 대한 접근이 가능하다고 가정하면, 우리는 고품질의 센서스 통계를 작성하는데 좀 더 정교한 방법을 개발하는데 도움을 얻을 수 있을 것이다. 곧 이것은 <그림 7-7>에서 품질과 비용의 관계가 양(+)의 관계라는 것을 보여준다. 같은 곡선은 추가적인 비용/시간/노력이 센서스 통계의 품질을 개선시키는 효과가 미미하거나 오히려 줄어드는 시점이 있으며, 결국 통계적 정도(精度: precision)에는 최적점(最適点: optimal point)이 존재한다는 것을 확인하여 준다. 또, 대단히 정교한 기법을 활용한다고 하더라도, 그것이 통계적 정도를 개선하는데 한계가 있으며 오히려, 품질측정의 한 차원인 시의성(timeliness)을 감소시키는 부정적 결과를 가져다 줄 수 있다는 것도 기억할 필요가 있다. 특히, 읍면동과 같은 너무나 상세한 수준에서 센서스 통계를 작성하거나, 센서스 조사항목을 읍면동 수준에서 5세 간격으로 통계를 작성하게 되면, 저출산으로 어려움을 겪는 소지역에서는 추정결과의 품질이 대단히 낮은 수준에 이르게 되어 센서스의 전반적 품질이 크게 손상될 수도 있다.

쟁점: Statistical expertise (통계적 전문성)
Robustness of results (결과의 강건성)
Understanding results (결과의 이해)
Validity of assumptions (가정의 타당성)
User requirement (이용자 요구사항)
Availability of resources (자원의 이용가능성)
Timeliness/deadline (시의성/공표일자)
Level of precision (정도(精度)의 수준)
Simple models (간단한 모형)
Complex models (복잡한 모형)
Finer disaggregation (범주별 통계작성의 정교화)

- <그림 7-7>의 비용곡선에서 보는 것처럼, 소지역 추정의 품질과 비용에 관련되는 다른 쟁점들이 있을 수 있다. 가령, 상대적으로 복잡한 모형을 사용하게 되면, 모형결과를 이해하고 해석하는데 도움을 주는 주제영역의 지식이 상당히 많이 필요하다. 이러한 지식은 모형에 내재하는 가정의 타당성을 검정하고

<그림 7-7> 센서스 소지역 통계의 품질과 비용/시간/노력



모형결과의 강건성(robustness)이나 민감도(sensitivity)를 점검하는데 대단히 중요하다.

- 최종적으로 등록센서스에서, 위에서 논의된 품질과 비용과 관련하여 언급해야 할 중요한 점이 있다. 우선, 단순성(simplicity)은 소지역 추정결과의 해석가능성에 도움을 준다는 점에서 대단히 중요한 차원이다. 우리는 분명히 말하자면, 방법론이 간단하다고 해서, 그것이 품질이 낮은 추정결과를 가져다주지 않는 것을 강조하고자 한다. 상대적으로 복잡한 방법론은 센서스 추정, 특히 소지역 추정에서 확인 가능한 개선점이 있을 때 시도되는 것이 바람직할 것이다. 둘째로, 세련된 방법론을 사용한다고, 비용이 커진다는 결론에 도달할 필요는 없다. 가령, 소지역 추정을 행할 수 있는 분석적 능력이 확고하게 정착되기 시작하면, 그러한 복잡한 추정을 행하는데 드는 비용/시간/노력이 추가적으로 많이 소요되지는 않는다고 할 수 있을 것이다.

제3절 우리나라 2015년 등록센서스의 당면과제와 일정점검

1. 한국인구학회 주관 등록센서스 연구의 진행과정

(2010년 연구)

- 한국인구학회가 주관한 통계청의 2010년 연구용역 과제는 2010년 인구주택총조사 시범예행조사 지역(대전 유성구, 충남 홍성군)의 조사자료와 행정자료를 이용하였으며, 조사자료는 각 읍면동에서 20%의 통을 계통추출(systematic sampling)하였다. 유성구는 통 인구 100명 이하, 홍성군은 통 인구 50명 이하의 다른 통과 합하여 표본으로 추출되지 않도록 조정하였다.
- (추정방법) 2010년의 인구학회 학술연구 용역과제에서 센서스 추정방법론을 개발하기 위하여, 다양한 방법이 시도되었다. 구체적으로 이원시스템 추정(DSE), 단순회귀추정, 원점회귀추정, 비추정 방식으로 추정작업을 진행하였다. 또 5가지 모형의 성능은 RSE(상대표준오차), MSEr(평균제곱오차비)를 기준으로 하여 비교 검토되었다. 일반적으로, RSE 기준으로 이원시스템 모형이 전반적으로 우수하였으며, MSEr 기준으로는 단순회귀모형이 더 나은 것으로 나타났다.

- 추정방법의 가정을 보면, 단순회귀모형은 특정집단에서 집계된 결과를 그 지역에 사는 해당 집단의 상주인구라고 가정하고, 등록부의 행정자료를 보조정보로 사용하여 추정하였다. 사실상, 이 경우에는 등록부의 행정자료에는 추정단위 안(within estimation units)에 과다범위로 조사되는 상주인구는 존재하지 않거나, 추정단위 간에(across estimation units) 과다범위로 인한 과다집계율(overcount rate) (곧, 등록부 내의 행정인구 대비 비상주 인구의 비)은 동일하다고 가정하는 것이라 볼 수 있다.
- 이원시스템 추정모형(DSE: Dual System Estimation) 또는 포획/재포획 추정모형(C-RCE: Capture-Recapture Estimation)은 인구부문의 경우, 등록부의 행정자료와 현장실사의 조사자료가 모두 상주인구(usual-resident population)를 적중하는데 실패하고 있다고 가정한다. 그 이유는 행정자료는 등록부 자체에는 과다범위로 인한 과다집계가 있는데, 이 과다집계는 등록부에 포함된 인구가 해외출국, 사망, 국내전출 등으로 생겨난 과다(excess)가 있고, 실제로 거주하고 있는데 일시적인 출타(temporary home-leaving)와 최근에 신규로 전입하여, 등록부에 기재되지 않아서 현장실사의 조사자료에서 누락된 사람들이 있다고 본다. 또, 등록부의 행정자료나 현장실사의 조사자료 어느 곳에도 포함되지 못한 상주인구가 있다고 가정하고 이들을 추정하는 모형이다.
- 2010년 인구학회의 연구결과는 RSE(상대표준오차)가 큰 연령대를 구, 군별로 통합하고, RSE가 큰 지역을 통합하여 단순회귀 추정모형과 이원시스템 추정모형을 이용하는 것이 바람직하다는 결론에 도달하였다. 실제로, 이러한 결론은 등록센서스를 실시하는 거의 모든 국가에서 채택된 센서스 추정방법론이라고 볼 수 있으며, 비교적 추정지역이 크면 클수록, 일반화된 회귀모형(GREG: Generalized Regression)이나 이원시스템 추정모형을 이용하면 센서스 모집단의 크기를 추정하는데 그리 큰 문제가 없다는 결론을 내렸다.
- 통계청의 등록센서스 주무부서는 이원시스템 추정모형과 단순회귀 추정모형은 모두 문제점을 가지고 있는 것으로 파악하였다. 구체적으로, 20% 표본조사를 실시하는 경우, 과소범위 파라미터를 추정하는 것은 그리 어렵지 않다고 할

수 있지만, 등록부의 과다범위 파라미터는 통상적인 조사방법이 아닌 전화조사 등과 같은 방식에 의존해야 하기 때문에, 이원시스템 추정모형에 의하여 센서스 모집단을 추정하는 방식을 포기하기에 이르렀다. 또, 단순회귀추정 모형은 추정결과가 행정자료와 조사자료의 인구범위를 벗어나는 값을 지니게 됨으로 해서, 이 방식을 포기하고 다른 대안을 찾기로 하였다.

(2011년 연구)

- 2011년 한국인구학회 학술용역 과제는 전년도 연구의 후속으로, 2010 인구주택총조사 시범예행조사 지역(대전 유성구, 충남 홍성군)의 조사자료와 행정자료를 다시 사용하면서, 센서스 추정방법론을 개발하는데 관심을 기울였다. 여기에서 고려되기 시작한 방법론은 능형회귀 추정방법(Ridge Regression Estimation)이다.
- 능형회귀 추정방법은 현장실사의 조사자료와 등록부의 행정자료에서 각 관심 하위모집단별 인구 예측치가 기지수(known number)으로 주어졌다고 가정하고, 최종 관심 부모집단별 인수로 이 두 수치의 가중평균을 고려하는 것이다. 여기서는 20% 표본에 대한 현지실사의 결과 주어진 보조정보가 일반적으로 회귀추정량을 정의할 때 사용하는 모집단의 참값("true" population value)이 아닌 표본에 근거한 추정량이기 때문에 이의 변동을 고려한 능형회귀추정량을 사용하는 것이 이 방식의 핵심이라고 할 수 있다.
- 현지실사를 위한 표본 추출을 할 때 이상 조사구는 전수 조사 대상으로 분류하여 통계적 안정성을 높이고, 표본은 층화집락추출로 하고 동지역은 CV가 6~8%이내, 면지역은 9% 이내인 지역을 설정하였다.
- 시뮬레이션의 결과, 표본조사 기반 불편추정량의 경우, 표본 간 분산이 능형회귀추정량보다 크게 나타나 정확성 및 효율성은 낮게 나타난다. 능형회귀추정

량의 경우 표본조사 결과인 불편추정량의 분산과 조합하여 사용함으로 오차가 행정자료 결과보다 적으며 변동은 불편추정량보다 적게 나타난다.

- 20% 표본으로부터 산출된 추정량의 분산이 커질수록 능형회귀추정량은 행정자료 부모집단 인구에 근접하며, 반대로 추정량의 분산이 적을수록 표본조사 기반 불편추정량으로 수렴한다고 가정한다.
- 보통, 사람들은 인구주택총조사라고 하면, 단순집계를 생각하지만, 그것을 추정 문제라고 보면, 분산(variance)과 편향(bias) 간의 트레이드오프를 포함해서 온갖 통계적 문제가 발생한다. 이 때문에, 2011년의 추정방법론 연구는 능형회귀추정량에 관심이 집중되었으며, 이 방법은 추정결과의 결과가 편향성이 있지만 분산이 적은 추정량을 선택하는 것이라고 볼 수 있다.

(금년도의 연구)

- 2012년 한국인구학회 학술용역 과제는 전년도 연구의 후속으로, 2010 인구주택총조사와 2010년 행정자료 기준 비교를 진행하는 것으로, 12개 시군구에 대하여 인구, 가구, 주택 등 3개 부문의 총수를 시군구와 읍면동 별로 비교하여, 이상지역(outlier area)을 색출하고, 인구, 가구, 주택 부문을 동시에 고려하여 표본설계를 적정화하는 방안을 연구하였다.
- 센서스 추정방법론과 관련하여, 2011년 학술용역 과제에서와 같이 능형회귀추정방법론을 세련화하고자 하는 노력이 12개 시군구 지역을 중심으로 해서 이루어졌다. 행정자료를 보완하여, 인구부문의 센서스 통계를 작성하기 위하여 4개의 능형파라미터(Ridge parameters)를 개발하였으며, 가구부문과 주택부문에 대해서는 1개의 능형 파라미터가 개발되었다.
- 소지역추정방법론은 지역별 성별 연령(5세단위)별 부모집단을 소영역(small domain)으로 설정하고 20% 표본조사에서 표본추출단위(조사구)를 고려하여 단위기반 소지역모형(unit-level small area model)을 설정한다. 지역으로는

시군구 수준과 읍면동 수준을 동시에 고려한다. 지역·성·연령별 인구수를 산출하기 위해 네 가지 추정방법 - 직접 추정량(Direct estimator), 비 추정량(Ratio estimator), 경험적 최량선형불편 추정량(Empirical Best Linear Unbiased Predictor; EBLUP), 그리고 계층적 베이즈 추정량(Hierarchical Bayes estimator; HB)을 사용하고자 한다. 이 때 직접 추정량과 비 추정량은 설계기반 추정방법에 해당되고 EBLUP과 HB 추정량은 모형기반 추정방법에 해당된다. 동일한 방법으로 지역·가구원수별 가구수와 지역·주택유형별 주택수를 산출한다.

- 등록센서스 준비과정에서 행정자료에 대한 품질문제는 추정문제와 별도 대단히 중요한 과제임은 틀림없다. 실제로, 품질은 포괄범위뿐만 아니라 행정자료의 기재내용이 전통적 센서스의 조사항목을 대체하는 아무런 문제가 없는가를 철저히 점검하여야 할 필요가 있다. 올해의 연구에서 행정자료의 혼인상태, 가구주와의 관계, 가구원수별 가구, 연면적별 주택수, 건축연도별 주택수 등에 대한 등록센서스 결과 작성방식을 평가했지만, 전반적으로 등록센서스 통계의 문제는 포괄범위를 조정하는 가중치 설정의 문제가 아니라, 행정자료 자체의 품질문제가 중요했다. 주택부문 통계를 제외하고는 혼인상태별 인구, 가구주와의 관계별 인구, 가구원수별 가구는 현재로서는 등록센서스 통계작성의 품질요건을 충족시킨다고 보기는 힘들다. 가구생성을 통하여, 행정자료의 지역별 과다범위와 과소범위를 검토하고, 필요한 경우에 그것을 보완할 수 있는 방법론을 개발할 필요가 있다.

2. 2015년 등록센서스의 당면과제과 일정점검

- 1925년 일제강점기 시대에 간이국세조사란 이름으로 우리나라의 인구주택총조사가 처음으로 실시된 이후, 거의 예외 없이 5년마다 실시되어 왔다. 조사방식은 초기에는 전수조사 방식으로 진행되다가, 해방 이후, 1970년대 들어서 본격적으로 표본조사를 센서스에 도입하여, 전수항목과 표본항목을 구분하기

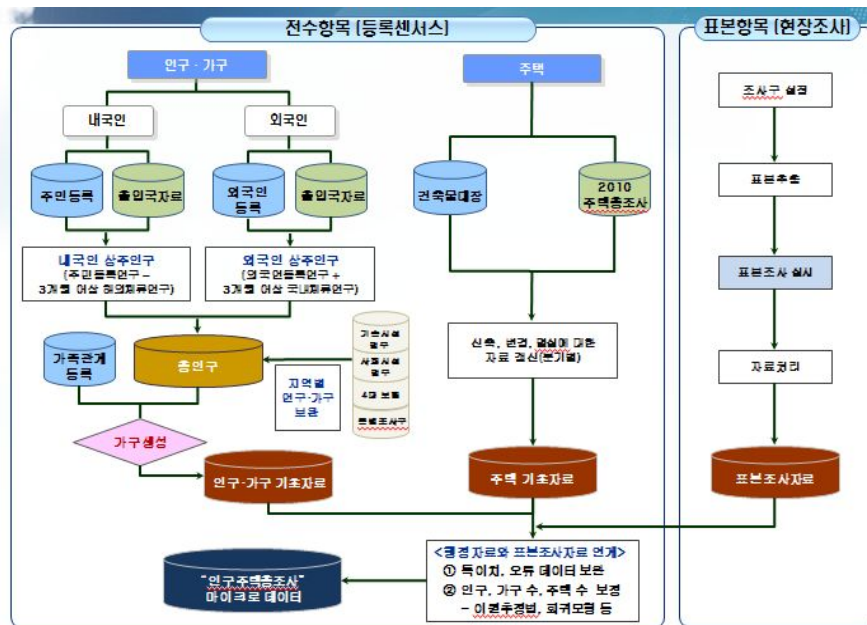
에 이르렀다.

- 한편 통계청은 2000년대에 들어서 행정자료 특히 건축물대장의 자료를 센서스에 활용하기 시작하였으며, 이러한 경험을 바탕으로 2015년부터, 주민등록과 건축물대장을 근간으로 하는 행정자료를 확보하고, 그것을 보완하는 표본조사를 실시하여, 그것을 결합하는 부분 등록센서스(partially register-based census) 또는 결합센서스(combined census)로 이행하기로 하였다. 통계청은 또한 2015년 등록센서스의 성공적 추진을 위하여, 2011년에 통계청 내 조사관리국 소속으로 등록센서스 전담부서를 설치하고, 센서스 목적의 행정자료 데이터베이스를 구축하고, 이를 표본조사 자료와 결합하여 센서스 모집단의 파라미터를 추정하는 연구를 집중적으로 진행하고 있다.

(등록센서스 추진체계도 관련)

- <그림 7-12>는 우리나라가 2015년에 실시하게 될 등록센서스의 흐름도를 제시한다. 등록센서스에서 전수항목은 전적으로 등록부에 의존하는 것으로서, 인구, 가구는 주민등록부, 가족관계등록부, 주택은 건축물대장을 자료원으로 하여 실시된다.
- 2015년에 실시되는 인구주택총조사는 엄격히 말하자면, 인구부문, 가구부문, 주택부문에 대하여 각각 행정자료를 이용하여 집계되어야 할 전수조사, 곧 센서스이기는 하지만, 그것을 편의상 동일시점에서 실시하는 것이라고 볼 수 있다. 특히, 모든 개인은 가구의 구성원이 되고, 가구의 구성원은 특정의 주택에 거주한다고 생각한다면, 인구-가구, 가구-주택, 인구-주택과 같이 3방향 매칭/연계(three-way matching and linking)에 의하여 생산될 수 있는 등록기반 센서스 통계(register-based census statistics)는 응답부담과 통계비용을 경감하면서 작성되는 통계로서, 센서스 방법론의 전환이라는 점에서 엄청난 도전이며, 전통센서스 방식으로 조사표를 단순히 집계해온 통계청으로 대단히

<그림 7-12> 2015년 등록센서스 추진체계도



어려운 대과제를 수행하고 있다고 볼 수 있을 것이다.

- 2012년의 센서스 추정방법론의 연구에서 특징적인 것은 등록센서스 담당 주무부서가 연구자들에게 조사구 수준에서 집계된 자료(aggregate data)를 제공하였다는 것이다. 실제로, 등록센서스 통계데이터베이스를 담당하는 주무부서가 개체수준의 행정자료와 조사자료를 2010년 인구주택총조사 시점에서 매칭하고 연계하고 있다고 판단되지만, 이것이 표본설계나 추정방법론의 개발에 있어서 세련된 방식으로 센서스 모집단을 확정하고 행정자료 기반 통계표를 작성하는데 있어서 품질점검에 필요한 지표를 개발하는데 도움이 될 수 있도록, 관련 통계데이터베이스를 적극 활용하였으면 하는 기대가 있다.
- 2015년 등록센서스에서 인구부문의 센서스 모집단은 내국인의 경우는 주민등록을 이용하고, 외국인의 경우에는 외국인등록을 이용하여 인구자료를 추출하

도록 되어 있다. 가족관계등록부는 가구생성(household generation) 목적에 이용되며, 대학생 기숙사 명부, 요양시설, 교도소 수감인원 명부, 4대 보험(국민연금, 건강보험, 고용보험, 산재보험) 등의 행정자료도 인구, 가구 등의 자료를 보완하는데 이용하도록 되어 있다.

- 우리나라의 2015년 등록센서스의 추정방법론 개발과정에서 행정자료는 개체 수준(record level)에서 삭제되거나 임퓨테이션된 것이 바람직하다. 또, 그것이 센서스의 기본원칙인 개체성(individuality)과 보편성(universality)의 원칙을 제대로 충족시킬 수 있고, 등록센서스의 데이터베이스를 구축하는 근본적 취지에도 부합된다고 판단하고 있다. 그러나 현재로서는 등록센서스의 데이터베이스를 종전의 전통적 센서스에서 사용되는 조사항목의 정의(定義)와 일치하는 방식으로 구축하는 작업은 그리 간단하지는 않을 것이라고 판단한다. 이 때문에 등록센서스 담당 주무부서는 등록센서스의 데이터베이스 구축에 필요한 기초정보를 표본조사의 수준에서 확보할 것인지 아니면 다른 방식으로 확보할 것인지에 대하여 조만간 결정을 내려야 할 것으로 판단한다. 현재, 행정자료에서 성별, 연령, 거주지 등을 조정하는 방법론을 개발하고 있지만, 인구 부문(혼인상태, 가구주와의 관계), 가구 부문(가구원수별 가구), 주택 부문(연면적별 주택수, 건축연도별 주택수) 중에서, 특히 인구/가구 부문의 행정자료에 대한 품질을 개선하는 방안에 대한 연구를 집중적으로 진행할 필요가 있다.
- 2015년 등록센서스에서 20% 표본자료는 행정자료를 보완하고, 나아가 우리나라의 인구를 성, 연령별로 추정하여 행정자료는 물론 표본자료의 가중치를 부여하는데 사용하게 될 것이다. 특히, 센서스 모집단을 보완하기 위하여 상당부분 표본조사 자료에 의존하는 정도가 큰 부분 등록센서스가 2015년 우리나라의 센서스에 있어서의 중요한 특징이라고 할 수 있다.
- 본 연구가 한국인구학회의 주관으로 지난 3년 동안 다양한 실험을 통하여, 모집단의 추정과 가중치를 부여하는 방법에 대하여 상세한 검토를 실시하고 있지만, 비추정법과 이원시스템추정법, 그리고 최근에 본 연구에서 개발되고 있는 능형회귀모형(Ridge regression estimation)와 소지역 추정방법론의 장단

점을 비교하여, 최종적으로 어떤 방법을 택하여야 할 것인가는 우리나라 통계법이나 행정자료 이용관련 법률의 제정상황, 개인정보보호법의 제한방식 등을 충분히 고려하여 이들 중에 하나를 선택할 수 있을 것이라고 생각된다.

(등록센서스 연도별 주요업무 관련 일정)

- <그림 7-13>는 통계청의 등록센서스 담당 주무부서가 연도별로 추진해야 하는 주요업무의 일정표를 보여준다. 이 추진계획표는 대만 주계처가 2010년 등록센서스를 추진하면서, 마련한 일정표 이상으로 대단히 상세한 일정표를 마련하고 있어서, 등록센서스를 체계적으로 일정의 차질없이 추진하고 있다는 인상을 대외적으로 보여주는 것이라고 할 수 있다.
- 이 일정표에 의하면, 2012년도에는 (1) 행정자료와 2010년 총조사 자료 간에 개별 연계분석을 실시하는 것으로 되어 있으며 (2) 능형회귀추정모형(Ridge regression estimator)을 230개 시군구(기초자치단체)에 일반화 가능성을 검토하며, (3) 기초자치단체의 하위지역인 읍면동에 능형회귀추정방법을 적용하거나, 그것이 적합하지 않은 경우에는 별도의 소지역추정방법론을 개발하는 것으로 되어 있다.

<그림 7-13> 등록센서스과의 연도별 핵심추진업무

연도	핵심 추진업무
2012년	▶ 행정자료와 2010 총조사 자료 간 개별 연계 분석(~12월) ▶ 능형회귀추정모형을 230개 시군구에 일반화 가능성 검토 ▶ 읍면동 자료에 대한 능형회귀 또는 소지역 적용 방법 검토
2013년	▶ 2015 인구주택총조사 방식 전환에 대한 대국민 공표 : 7월 ▶ 2015 인구주택총조사 등록센서스 항목 확정 : 12월 ▶ 인구, 가구수, 주택수 보정방법 확정 : 12월 ▶ 표본조사 규모 및 표본추출률 확정 : 12월[* 인구총조사과 주관]
2014년	▶ 2015 등록센서스 실시계획 수립 : 6월 ▶ 2013년 기준 등록센서스 시산 : 6월 ▶ 2015 인구주택총조사 시범예행조사 : 11월[* 조사 기획·실시는 인구총조사과 주관]
2015년	▶ 2015 등록센서스 세부시행계획 수립 : 3월 ▶ 2015.11. 1기준 인구주택총조사 실시 : 11월[* 조사 기획·실시는 인구총조사과 주관]
2016년	▶ 2015 등록센서스(전수부문) 결과 공표 : 9월

- 한편, 2013년에는 등록센서스로 작성될 조사항목을 확정하고, 센서스 전환방식에 대한 대국민 공표 및 홍보전략 마련, 모델적용을 위한 시험조사를 실시하고, 인구수, 가구수, 주택수에 대한 보정방법을 마련하고, 표본조사 규모 및 추출률을 확정하는 것으로 되어 있다.

- 2013년부터는 등록센서스 실시와 관련하여, 통계청의 등록센서스 담당 주무부서만이 아니라 다른 부서들의 도움을 얻어서라도 행정자료의 포괄범위를 점검하는 표본조사를 실시하는 것이 바람직하다고 생각한다.

- 본 연구는 전통센서스 사후조사의 P-sample 개념으로 실시되는 표본조사로 과소범위를 조정할 수는 있지만, 등록부 자체를 가지고 표본(곧 E-sample에 해당함)을 선정하여 별도의 표본조사가 실시되지 않으면 과다범위를 점검하기는 쉽지 않을 것이라는 사실은 강조하였다.

- 사실, 센서스 모집단의 크기를 세련된 추정방법론을 적용하여 확정한다고

하더라도, 과다범위를 조정한 행정인구, 행정가구, 행정주택에 대한 자료를 가지고 있다면, 추가적으로 표본조사를 실시하여 과소범위를 조정하는 것이 추정결과를 더욱 더 신뢰성있는 것으로 만들 수 있다고 판단한다. 이 문제에 대하여 빠른 결정이 내려져야 한다.

- 전반적으로, 현재 등록센서스 목적으로 등록부 자료에서 이용가능한 센서스 조사항목의 개수가 그리 많지는 않겠지만, 좀 더 세심한 방법으로 이용가능한 항목을 확정해야 될 필요가 있을 것이다.
- 또, 인구수, 가구수, 주택수 등 센서스 모집단의 크기를 추정하기 위한 방법론은 조사구(ED) 수준에서 행정자료와 20% 수준의 조사자료를 비교하여 센서스 목표인구의 크기를 확정하는 능형회귀추정모형으로 사실상 정해져 있는데, 이것을 통해 센서스 모집단의 크기를 추정하고 그것의 논리적 타당성을 시험조사 절차에서 검정을 해보아야 할 것이다.
- 또, 시험조사에서 포괄범위 점검을 위한 표본조사와 사회경제통계를 얻기 위한 표본조사를 분리할 것인지 통합할 것인가에 대하여도 심사숙고해볼 필요가 있을 것이다. 만약에 이스라엘식으로 포괄범위(과소범위) 점검조사와 사회경제통계 표본조사를 병행하여 실시한다면, 응답자 부담을 줄이기 위하여 어떠한 조치를 해야 될 것인가에 대해서도 생각해 볼 필요가 있을 것이다.
- 현재, 우리나라의 센서스 모집단 추정방식은 능형회귀 추정량(Ridge Regression Estimator)에 근거하고 있다. 이 방법은 시군구의 기초자치단체에 적용되고, 소지역 추정방법론의 추정결과와 비교하여 그것이 크게 무리가 없는 한, 시군구 기초지방자치단체의 하위지역인 읍면동 수준에 대하여 적용가능성이 검토되고 있다.
- 현재로서는, 2013년에도 이 방법을 계속 연구하고 있기 때문에, 2015년 등록센서스 추진에서, 인구수, 가구수, 주택수 등 모수 추정방법론으로 채택될 가능성이 크다.

- 그러나 유럽연합 센서스 규정에 의하면, 모든 등록센서스에서는 등록부의 자료품질을 점검하여야 할 뿐만 아니라, 등록부 자료의 과다범위와 과소범위를 전국 수준에서는 물론 기초자치단체에 대해서 추정결과를 제시하도록 되어 있다.
- 이 때문에, 우리나라도 등록센서스 실시와 관련된 국제적 표준관례를 중시한다면, 등록센서스 데이터베이스의 정보와 20% 표본조사의 결과를 결합한 정보를 개체수준에 이용하는 직접추정량(direct estimator)을 포함하여 다양한 해결방법을 찾는 것은 대단히 중요할 것이라고 판단한다.
- 인구주택총조사 통계의 중요한 특성은 시계열로 인구특성은 물론 사회경제적 특성의 변동 과정을 기술하는 것이다. 곧, 특정의 사건이 월차별로, 연차별로 어떻게 발전하는가를 기술하는 것이다. 따라서 행정자료의 개념은 장기간에 걸쳐서 불변의 위치에 있을 필요가 있다. 그렇지 않으면, 시점마다 비교 가능한 수치를 확보하는데 문제가 생기게 된다. 어떻게 경우에는 정확성에 문제가 있을 수 있지만, 변동 상황에 대하여 약간의 조정이 필요한 경우가 있을 것이다.
- 인구주택총조사 목적으로, 행정자료를 사용할 때는, 목표 모집단에 대하여 행정자료의 과다집계(overcoverage)와 과소집계(undercoverage)의 크기를 확실하게 파악하는 것이 대단히 중요하다. 이러한 포괄범위를 조사통계의 방식으로 규명하고자 노력한 경우는 이스라엘의 2008년 통합센서스(integrated census)가 있고, 이탈리아가 2011년에 계획하고 있는 순환센서스에서도 포괄범위를 측정하는 문제를 포함하고 있다고 볼 수 있다. 현재, 우리나라가 시도하고 있는 추계방식은 순과다범위(net coverage)를 조정하려는 노력이라고 볼 수 있겠지만, 그것이 인구추정모형(population estimation model)의 정교한 논리에 근거하지 않고 있기 때문에 개선이 가능한 범위 안에서, 추정방식을 개선할 필요성은 있다고 판단하고 있다.
- 통계청의 2015년 센서스 실시와 관련하여, 주무부서인 등록센서스과에서는

2014년에는 등록센서스 실시계획을 수립하고, 2015년 11월 1일을 기준으로 하여 등록센서스를 실시하기 위하여, 20% 표본조사를 인구총조사과에 위탁하여 실시하도록 되어 있다. 앞에서도 언급했지만, 표본조사는 등록부 자료의 품질을 점검하고 각종 추정과 가중치 설정에 이용하며, 등록부에 포함되어 있지 않은 센서스 항목인 사회경제통계 자료를 얻기 위한 도구로 활용될 예정이다. 또 센서스 실시 1년 후에는 2016년에는 전수항목에 해당하는 등록센서스 결과, 곧 등록기반 센서스 통계표(register-based census tables)를 상반기에는 발표하는 것으로 되어 있다.

(등록센서스 준비 세부과제 관련 사항)

- 통계청이 2015년 등록센서스 실시를 위하여 준비하고 있는 구체적인 과제로는 (1) 등록센서스 인프라 구축 (2) 저비용/고효율의 등록센서스 기법 개발로 구분되는데, 처음의 과제로는 행정자료 확보, 추가 입수를 위한 관계기관 간의 협의, 자료입수 인프라 등 행정자료의 안정적 확보, 행정자료의 표준화, 연계키 확보, 품질점검, DB 구축 및 컴퓨터 소프트웨어 개발 등 행정자료 표준화 통계 데이터베이스 구축이 포함되어 있다.
- 두 번째의 과제에는 행정자료의 특성분석, 전수항목 확정 등 행정자료 분석을 통한 항목확정, 그리고 행정자료간 연계분석, 2010년 총조사 자료와의 연계분석, 인구수, 가구수, 주택수 보정방안 확정, 등록센서스 실시국가 또는 실시예정국가의 해외사례 연구, 전문가와의 공동연구, 등록센서스 시산, 표본조사 실시, 전수결과 공표 등 행정자료 연계기법 및 보정기법 연구로 구성되어 있다.

<그림 7-14> 통계청의 등록센서스 추진을 위한 연차별 세부과제 (1)

전략	주요과제	주요 내용	일정	2012				2013				2014				2015				2016			
				1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
등록센서스 인프라 구축	행정자료의 안정적인 확보	기 입수된 행정자료 확보	매년 11.1																				
		주거 입수율 위한 협의	'12. 1~'12																				
		자료입수 인프라 구축	'12.1~'14.12																				
	행정자료 표준화 및 체계화 구축	행정자료 표준화	'12. 1~'15.12																				
		연계기 확보	'12. 1~'12.12																				
		품질점검	'12. 1~'15.12																				
		DB 구축 및 프로그램 개발	'12. 1~'15.12																				
저비용·고효율의 등록센서스 기법 개발	행정자료 분석을 통한 항목 선정	행정자료 특성 분석	'12.1~'12.12																				
		전수항목 확정*	'13. 6																				
	행정자료 연계 및 보정기법 연구	행정자료간 연계 분석	'11. 5~'12																				
		2010총조사와 연계분석	'12. 1~'12																				
		인구, 가구수, 주택수	'15. 12																				
		보정 방안 확정*																					
		해외사례 연구	'12. 10																				
		전문가와 공동연구	'12 ~ '15																				
		등록센서스 시산	'15 ~ '14																				
		표본조사 실시	'15. 11. 1																				
		전수결과 공표*	'16. 9																				

- 올해의 연구에서도 행정자료의 품질에 대한 연구가 부분적으로 진행되었지만, 입력자료나 산출자료로서의 행정자료의 품질에 대하여 끊임없이 관심을 기울여야 할 것이라고 생각된다.
 - 현재 등록센서스 담당부서는 능형회귀추정량과 소집단추정방법론에서 얻은 센서스 가중치를 기반으로 센서스 모집단의 규모를 확정할 것이라고 생각되지만, 센서스 통계작성을 위한 입력자료로서, 성별, 연령별, 거주지별 정보에 대한 품질이 최종적으로 확인되지 않은 이상, 이들을 추정목적으로 사용하기 전에 어느 정도 개선할 수 있는가에 대한 품질개선 작업이 선행되어야 할 것이다.
 - 한마디로 자료의 삭제나 임putation으로 행정자료의 품질을 전면적으로 개선할 수 없듯이, 추정방법론으로서 전수조사를 대체하는 행정자료의 품질을 전면적으로 개선할 수는 없다. 혼인상태와 같이 법적 혼인상태와 사실

혼인상태 간의 괴리가 있거나, 1인가구와 같은 경우에는 주민등록, 가족관계등록부 등의 자료는 물론, 여타 보조적인 행정자료를 이용하여 행정자료의 품질을 개선하는 방안을 꾸준히 진전시켜나가야 할 것으로 생각한다.

- 통계청 등록센서스 담당부서는 등록센서스의 센서스 조사항목에 대해서 산출되는 통계표를 빠른 시간 안에 공표해야 한다는 강박관념에서 해방될 필요가 있다. 현재 진행되는 연구 자체에서도 입증되었듯이, 인구, 가구, 주택 부문에서 센서스 모집단(또는 부모집단)의 크기를 확정하는 것은 결코 간단한 문제는 아니다. 센서스 비용의 절약이나 사생활 보호차원에서 실시되는 등록센서스의 실시가 통계적으로는 대단히 복잡한 문제를 포함하고 있다는 점을 항상 유념할 필요가 있을 것이다. 이 때문에, 등록센서스 결과를 공표하는 일정에는 약간의 신축성을 부여하는 방안을 마련하지 않으면 안 될 것이라고 판단하고 있다.

<그림 7-15> 통계청의 등록센서스 추진을 위한 연차별 세부과제 (2)

전략	주요과제	주요 내용	일정	2012				2013				2014				2015				2016			
				1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
결과자료의 활용가치 증대	매년 표본추출 및 총조사 자료 제공	표본설계 연구	'12.1~'15.12																				
		행정자료 이용 기초단위구분 안	'12.1~'15.9																				
		표본설계 확정*	'15. 12																				
	기준인구의 일원화 등 활용도 제고	기준인구 작성 일원화 연구	'12. 1~'15.12																				
		총조사 결과 활용에 대한 법제화	'12 ~ '15																				
국민적 공감대 형성 및 개인정보보호 강화	대국민 홍보 및 전문가와 협력관계 구축	방식 전환에 대한 대국민 공표*	'15. 7																				
		등록센서스 효용성 홍보	'15.8~'15.11																				
		관계기관 워크숍	'12 ~ '15																				
		전문가회의(연1~2회)	'12 ~ '15																				
	법령 제정 및 개인정보보호 강화	행정자료 활용 기본법 마련	'12 ~ '15																				
		관계기관 의견조회 및 협의	'15. 5																				
		행정자료 활용 기본법 공표	'14. 1																				

- 통계청이 2015년 등록센서스 실시를 위하여 준비하고 있는 세번째 추가 과제로 설정되어 있는 것은 (1) 결과자료의 활용가치를 증대하고 (2) 국민적 공감대 형성 및 개인정보보호를 강화하기 위한 조치를 마련하는 것이 포함되어 있다(<그림 7-15> 참고).
- 결과자료의 활용가치를 증대시키기 위하여, 매년 표본추출틀 및 총조사 자료를 제공하고, 기준인구의 일원화 및 활용도를 개선하며, 이와 관련된 세부과제로는 표본설계 연구, 행정자료 이용 기초단위구 보완, 표본설계 확정, 기준인구 작성 일원화 연구, 총조사 결과 활용에 대한 법제화를 포함한다.
 - 등록센서스를 2015년에 실시하고 나면, 곧장 등장하게 되는 문제가 5년단위로 이러한 센서스를 실시하는 센서스로 머물 것인가, 아니면 연단위로 등록센서스를 실시하여 인구, 가구, 주택에 관한 센서스를 실시할 것인가에 대한 문제가 될 것이다. 이 때문에, 본 연구는 빠른 시간 안에 어떠한 방식으로 연단위로 등록센서스 곧 등록 기반 통계(Register-based statistics)를 어떤 영역에 걸쳐서 생산할 것이며, 사회경제통계는 연 단위 주기로 아니면 5년 주기로 자료를 생산할 것인가에 대한 문제에 직면하게 될 것이다.
 - 본 연구는 등록센서스 실시국가의 경험을 근거하여 판단할 때, 매년 인구통계 특히 성별, 연령별 인구구조에 대한 자료를 제공하려고 하는 경우에는, 현재 기획되고 있는 표본조사에서 과소범위를 추정하는 기능과 별도의 과다범위를 추정하는 기능을 동시에 고려하는 표본조사를 실시할 필요가 있다.
 - 이 때문에, 2015년에 실시하는 등록센서스에서도 과소범위를 점검한 표본조사와 사회경제통계 자료를 획득하는 표본조사를 하나의 표본조사로 통합할 것인가에 대하여 심사숙고해야 하며, 궁극적으로는 2010년 인구주택총조사 표본항목 조사표는 물론 대만이나 독일 등의 등록센서스 실시국가에 의하여 사용되는 표본조사 조사표를 참고하여, 등록센서스 실시에 부합되는 표본조사표를 개발하여야 할 것이다.

- 본 연구는 또한 우리나라의 등록센서스가 순환 등록센서스(Rolling Register-based Census)로 중장기적으로 진화하게 된다면, 포괄범위를 점검하는 표본조사와 사회경제통계를 점검하는 조사의 분리를 의미하게 되는 것이라고 본다. 이것은 아마도, 이탈리아식의 C-sample과 D-sample 표본 조사를 별도 설정하는 것이 될 것이다.
- 현재의 규모가 작은 표본조사를 매년 실시하여 포괄범위나 사회경제통계를 획득하는 이스라엘식의 통합 순환센서스(IRC: Integrated Rolling Census)로 진화할 수도 있을 것이다. 물론, 또 다른 옵션으로 현재 기획되고 있는 2015년의 센서스를 2010년의 스위스식 순환 등록센서스의 방식으로 진전시키면서, 센서스를 등록부조사, 구조조사, 주제조사, 오피버스조사로 구분하고 주제조사는 현재 통계청이 사회통계기획과 주무부서가 되어 실시하는 사회조사와 경제활동인구조사를 결합하여 센서스의 다양한 모듈을 개발할 수도 있을 것이다.
- 등록센서스 담당 주무부서는 네 번째의 세부과제로서 국민적 공감대 형성과 개인정보보호 강화를 위해서는 대국민 홍보 및 전문가와 협력관계를 구축하고, 법령제정 및 개인정보보호 강화를 위한 전략을 마련하고 있다. 이를 위한 구체적 과제로는 센서스 방식 전환에 대한 대국민 공표, 등록센서스의 효율성 홍보, 관계기관 워크숍, 전문가회의, 행정자료 활용 기본법 마련, 관계기관 의견조회 및 협의, 행정자료 활용 기본법 공표 등을 포함하고 있다.
- 2015년 등록센서스를 실시하기 전에, 주무부서는 끊임없이 등록센서스의 결과에 의하여 영향을 받을 수 있는 이해당사자들과 접촉하여야 할 것이라고 판단한다. 등록센서스가 행정자료를 이용하는 것인 만큼, 이에 필요한 우호적인 여론형성이나 개인정보보호법과 통계법의 통계비밀보호(statistical confidentiality) 조항에 관한 사항도 충분히 홍보되어야 할 것이라고 판단한다.
- 또 법령제정 특히 행정자료의 이용에 관한 법률의 제정을 빠른 시간 안에 진행되고, 이 법률을 근거로 하여 행정자료를 센서스 목적에 이용할 수 있

도록 노력을 하여야 할 것이라고 생각한다. 또, 행정자료의 활용과 관련하여, 국가통계로서의 품질기준을 준수하기 위하여, 행정자료를 구성하는 “통계단위”(statistical unit)의 일부를 표본으로 선정하여 포괄범위나 기재내용의 정확성을 점검하는데 따른 안전장치를 두는 방안도 법조항에 삽입하도록 하여야 할 것이다.

- 마지막으로, 우리나라도 2015년 등록센서스를 본격적으로 추진하게 되는 이상, 대만의 2010년 센서스 실시에 따른 시행규칙을 참고하여(<참고 자료 7-2>을 볼 것), 등록센서스 실시계획의 초안을 마련하여야 할 것이라고 생각된다. 이 실시계획에는 총칙은 물론, 센서스 실시의 범위와 대상, 센서스 방식, 표본추출방법, 조사항목, 센서스 목적에 이용되는 등록부의 종류, 센서스 명부, 주관기관, 협력기관과 업무분장, 임시 센서스 조직, 조사구 구획, 센서스 인원, 홍보 및 훈련, 현지실사, 사후조사, 센서스 결과의 심의 및 배포 등으로 구성되는 것이 바람직할 것이다.

<참고 자료 7-1-1> 대만의 등록센서스 보완 센서스 조사표

報名日期：109年12月15日(星期日)起至12月26日(星期五)止
報名時間：上午9時至下午5時(12月26日除外)

注意事項：
1.請使用藍色或黑色鋼筆，以正楷填寫各項資料。
2.文字填寫請注意：數字請書寫正確，阿拉伯數字請用半角符號「.」填寫。
3.如有填錯(或)填漏，應重新填寫，填錯者由申請者自行負責。
4.不可填寫虛假、重複資料，違者將取消其申請資格。

99年人口及住宅普查表

普查標準時點：民國99年12月26日零時正

1.本普查表分為「基本戶」及「非基本戶」兩種，由戶長或代理人，持本普查表，向戶籍所在地之戶政事務所領取。
2.本普查表之填寫，應以戶籍登記資料為準，如有不一致者，應以戶籍登記資料為準。
3.本普查表之填寫，應以戶籍登記資料為準，如有不一致者，應以戶籍登記資料為準。

三、人口資料 (包括戶籍所在地及所有經常居住在本戶之人口) (如有同一姓名者，請填寫同一位)

(1)姓名及性別：
☐ 男 ☐ 女

(2)出生年月日：
 民國 年 月 日 西曆 年 月 日

(3)是否為本籍籍：
☐ 是 ☐ 否

(4)婚姻狀況：
☐ 未婚 ☐ 有配偶(同居人) ☐ 已離婚或分居 ☐ 已離婚死亡

(5)與戶長關係：
☐ 戶長 ☐ 配偶 ☐ 父母 ☐ 子女 ☐ 兄弟姊妹 ☐ 其他共同生活者
☐ 配偶(同居人) ☐ 前配偶 ☐ 其他共同生活者
☐ 父母(父) ☐ 父母(母) ☐ 兄弟姊妹 ☐ 其他共同生活者
☐ 配偶(父) ☐ 配偶(母) ☐ 兄弟姊妹 ☐ 其他共同生活者

(6)教育程度：
☐ 小學 ☐ 國小 ☐ 國中 ☐ 高中 ☐ 大學 ☐ 研究所以上
☐ 小學 ☐ 國小 ☐ 國中 ☐ 高中 ☐ 大學 ☐ 研究所以上
☐ 小學 ☐ 國小 ☐ 國中 ☐ 高中 ☐ 大學 ☐ 研究所以上

(7)現任職業類別：
 1.職業類別：(請勾選)
☐ 1.農、林、漁、牧業 ☐ 2.製造業 ☐ 3.服務業 ☐ 4.其他
 2.職業類別：(請勾選)
☐ 1.農、林、漁、牧業 ☐ 2.製造業 ☐ 3.服務業 ☐ 4.其他

(8)現任職業類別：
 1.職業類別：(請勾選)
☐ 1.農、林、漁、牧業 ☐ 2.製造業 ☐ 3.服務業 ☐ 4.其他
 2.職業類別：(請勾選)
☐ 1.農、林、漁、牧業 ☐ 2.製造業 ☐ 3.服務業 ☐ 4.其他

(9)現任職業類別：
 1.職業類別：(請勾選)
☐ 1.農、林、漁、牧業 ☐ 2.製造業 ☐ 3.服務業 ☐ 4.其他
 2.職業類別：(請勾選)
☐ 1.農、林、漁、牧業 ☐ 2.製造業 ☐ 3.服務業 ☐ 4.其他

四、住宅資料 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(9)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

五、住宅狀況 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(9)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

六、住宅狀況 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(9)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

七、住宅狀況 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(9)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

八、住宅狀況 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(9)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

九、住宅狀況 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(9)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

十、住宅狀況 (包括本戶所有住宅及所有經常居住在本戶之住宅)

(1)住宅地址：
 縣市代碼： 鄉鎮市區代碼： 里鄰代碼： 門牌： 戶號：

(2)住宅用途：
☐ 住宅 ☐ 商業 ☐ 工業 ☐ 其他

(3)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(4)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(5)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(6)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(7)住宅所有權：
☐ 自有 ☐ 共有 ☐ 其他

(8)住宅所有權：

199012

三、人口状况(续) (在沪户籍出生的人口; 在沪户籍不在沪出生的人口; 在沪户籍不在沪出生的人口)

(1) 户籍在沪出生有经常居住在本市; 已死亡; 已不在本市居住	(2) 户籍在沪出生有经常居住在本市; 已死亡; 已不在本市居住	(3) 户籍在沪出生有经常居住在本市; 已死亡; 已不在本市居住
姓名: 性别: ☐ 男 ☐ 女	姓名: 性别: ☐ 男 ☐ 女	姓名: 性别: ☐ 男 ☐ 女
民族: 出生年月日:	民族: 出生年月日:	民族: 出生年月日:
婚姻状况: ☐ 未婚 ☐ 已婚 ☐ 离异 ☐ 丧偶	婚姻状况: ☐ 未婚 ☐ 已婚 ☐ 离异 ☐ 丧偶	婚姻状况: ☐ 未婚 ☐ 已婚 ☐ 离异 ☐ 丧偶
配偶姓名:	配偶姓名:	配偶姓名:
子女姓名:	子女姓名:	子女姓名:
父母姓名:	父母姓名:	父母姓名:
教育程度: ☐ 小学 ☐ 初中 ☐ 高中 ☐ 大学 ☐ 研究生	教育程度: ☐ 小学 ☐ 初中 ☐ 高中 ☐ 大学 ☐ 研究生	教育程度: ☐ 小学 ☐ 初中 ☐ 高中 ☐ 大学 ☐ 研究生
职业: ☐ 工人 ☐ 农民 ☐ 干部 ☐ 其他	职业: ☐ 工人 ☐ 农民 ☐ 干部 ☐ 其他	职业: ☐ 工人 ☐ 农民 ☐ 干部 ☐ 其他
住址:	住址:	住址:
联系电话:	联系电话:	联系电话:
健康状况: ☐ 良好 ☐ 一般 ☐ 较差	健康状况: ☐ 良好 ☐ 一般 ☐ 较差	健康状况: ☐ 良好 ☐ 一般 ☐ 较差
其他:	其他:	其他:

Please mark your answers in the official 2011 Census questionnaire.



Census of Buildings and Housing on 9 May 2011
 Please return the questionnaire within 2 weeks of receipt in the return envelope or complete it online.

online Save postage! You may also complete the questionnaire on the internet. Everything is prepared for you at www.zensus2011.de.

Your questionnaire number: Your activation code:

Purpose of the survey
 This survey collects data on the structure of the housing stock. To this end, the owners, managers or other parties entitled to use or dispose of a building or dwelling are questioned in writing. The survey is part of the 2011 Census.

General information

Property address

Building ID:

All questions in this questionnaire refer to this address.

- 1 Is there at least one building with residential space or an inhabited accommodation at the property address?
 Yes ... ☐ No ... ☐
- 2 Are you obliged to provide information on the building(s) or dwelling(s) at the property address because you are, for example, the owner?
 Yes ... ☐ No ... ☐
- 3 Data on the new owner (or manager):
 First name:
 Surname / company:
 Street, House number:
 Postcode, place name, country:
- 4 Do you own other buildings with residential space at this property address (e.g. rear building) for which you have not received a questionnaire?
 Yes ... ☐ No ... ☐

Please give a daytime telephone number where we can contact you to resolve any queries!

For legal information please refer to the annex, which forms part of the questionnaire.
 Your direct Census contact for any enquiries:
www.zensus2011.de

Yes ... ☐ Go to question 2.
 No ... ☐ End of survey. Please return the questionnaire.

Yes ... ☐ Go to question 4.
 No ... ☐ Go to question 3 and return the questionnaire afterwards.

Please request a questionnaire for each of the other buildings. Now go to question G1 on page 2.
 Go to question G1 on page 2.

Please mark your answers in the official 2011 Census questionnaire.

Please proceed as follows:

1. Answer the questions in the order given. Please mark only one answer for each question.
2. Please use a black or blue ballpoint pen to complete the questionnaire.
3. When entering text, please use block capitals.

Example: Surname:
Street:

4. If you wish to correct an answer, please make sure that the correction is clear so that there can be no doubt about your answer.

Example: Use of the dwelling on 9 May 2011

Occupied by owner ☐

Let for residential use ☐

5. Instructions on the questions are highlighted like this: **11**. When answering the questions, please observe instructions **11** to **14** in the document attached.

Questions about the building

G1 What type of building is it? **11**

In residential buildings at least half of the floor space is used for residential purposes. In other buildings with residential space less than half of the floor space is used for residential purposes.

Residential building (no residential home) ☐

Other building with residential space, e.g. commercial building containing a dwelling ☐

Residential home where occupants have their own household ☐

Inhabited accommodation, e.g. summer-house, portacabin ☐

G2 How many dwellings are in the building? **11**

Please indicate "1" for a one-family house. Include dwellings used for commercial purposes.

Number of dwellings

G3 What type of building is it? **11**

Detached one-family or multi-family house ☐

Semi-detached one-family or multi-family house ☐

Terraced one-family or multi-family house ☐

Other type of building ☐

G4 In what year was the building completed? **11**

If you do not know the exact year of construction, please give an estimate.

G5 Who owns the building? **11**

For buildings containing freehold dwellings please always mark **commonhold owners** even if all the freehold units have the same owner.

Commonhold owners ☐

Private individual(s) ☐

Housing co-operative ☐

Private sector housing company ☐

Other private sector company ☐

Municipality or municipal housing company ☐

Federation, State ☐

Non-profit organisation, e.g. church ☐

G6 What is the main type of heating in the building? **11**

District heating (long-distance heating) ☐

Block heating ☐

Central heating ☐

Single-storey heating ☐

Single-room or multi-room stoves (incl. night storage heating) ☐

No heating in the building or in the dwellings ☐

Please mark your answers in the official 2011 Census questionnaire.

Questions about dwellings

You can enter data for up to 6 dwellings on the following pages.

If you are the owner of a one-family house or of a single freehold dwelling, please enter the data under "Dwelling 1".

Dwelling 1

W1 Please enter the first names and surnames of up to two persons living in the dwelling on 9 May 2011.

Start with the adults first. ☐

Please enter your own name even if you are the owner-occupier of the house/dwelling. Leave these boxes blank if it is a diplomat's dwelling, a dwelling of foreign armed forces, if the dwelling is vacant or used exclusively for commercial purposes.

Occupant 1:	First name	
	Surname	
Occupant 2:	First name	
	Surname	

Is it a freehold dwelling? (let or owner-occupied) Yes ... ☐ No ... ☐

Go to question W2. Go to question W3.

W2 Who owns the dwelling? ☐

Private individual(s) ☐

Private sector company ☐

Public sector company, church or the like ☐

Housing co-operative ☐

W3 How is the dwelling being used on 9 May 2011? ☐

Occupied by owner ☐

Let for residential use ☐

Vacant ☐

Whole dwelling used exclusively for commercial purposes. If the whole dwelling is used for commercial purposes, there are no more questions to answer on this dwelling. ☐

W4 What is the floor space of the dwelling? ☐

Floor space in m²

W5 How many rooms are in the dwelling? ☐

Please count in the kitchen as well as every room of 6m² or more (except bathroom, toilet, corridor).

Number of rooms

W6 Is there a toilet in the dwelling? ☐

Yes ... ☐ No ... ☐

W7 Is there a bathtub or shower in the dwelling? ☐

Yes ... ☐ No ... ☐

W8 How many persons are living in the dwelling on 9 May 2011? ☐

Number of persons

W9 This dwelling is ... ☐

... a holiday or leisure dwelling (not continuously used for commercial/hotel purposes) ☐

... a diplomat's dwelling/dwelling of foreign armed forces ☐

... none of the two ☐

Please continue here

Are you obliged to provide information on other dwellings in the building? Yes ... ☐ No ... ☐

Go to dwelling 2. End of survey. Please return the questionnaire.

Guideline for the language support in ENGLISH, DO NOT give answers in this English version, which just provides language support.

GWZ 2011

Please mark your answers in the official 2011 Census questionnaire.

Dwelling 6

W1 Please indicate up to two persons who are living in the dwelling on 9 May 2011.

Occupant 1: First name

Surname

Occupant 2: First name

Surname

W2 Only for freehold dwellings:
Who owns the dwelling?

Private individual(s) ☐

Private sector company ☐

Public sector company, church or the like ☐

Housing co-operative ☐

W3 Use of the dwelling on 9 May 2011?

Occupied by owner ☐

Let for residential use ☐

Vacant ☐

Whole dwelling used exclusively for commercial purposes ☐

End of survey for this dwelling ☐

How many rooms are
W5 in the dwelling?

W6 Toilet in the dwelling

Yes ☐ No ☐

W7 Bathtub or shower in the dwelling

Yes ☐ No ☐

W8 How many persons are living in the
dwelling on 9 May 2011?

W9 This dwelling is ...

... a holiday or leisure dwelling ☐

... a diplomat's dwelling/dwelling of foreign
armed forces ☐

... none of the two ☐

W4 Floor space of the dwelling in m²

If you are responsible for providing information on more than 6 dwellings in the building, please use the separate continuation questionnaire. If you have not received a continuation questionnaire, please contact us to request one.

Comments

If you wish to make comments or add supplementary data on the building or the dwellings therein, please enter that information here.

Thank you very much for your support.

Census of Buildings and Housing on 9 May 2011

Instructions for page one of the questionnaire

A separate building and housing questionnaire has to be completed for each building with residential space and for each inhabited accommodation.

A building with residential space is defined as:

- every detached building or
- every building in a group of semi-detached buildings separated by a firewall. As a rule, semi-detached buildings can be recognised by their separate entrance and separate staircase which lead to at least one dwelling, irrespective of whether the dwellings are occupied or vacant.

Inhabited accommodations include inhabited summerhouses, inhabited construction site trailers, inhabited portacabins and the like.

If you are the owner of several buildings with residential space, please request a separate building and housing questionnaire for each building, even if these buildings have the same address.

Instructions on the questions about the building

- 1 Please mark residential building if at least half of the total useful floor space in the building is used for residential purposes and it is not a residential home. If less than half of the total useful floor space is used for residential purposes, because the building mainly contains, for example, shops or offices, please mark other building with residential space.

Residential homes are residential buildings that are mainly used to meet the housing needs of specific population groups (e.g. student hall of residence, home for elderly people).

Residential homes contain common rooms. The occupants of residential homes have their own household. Please mark residential home even if the building includes, for example, a long-term care unit.

Inhabited accommodations are temporary structures. They include, for instance, summerhouses, construction site trailers, shacks, mobile homes (e.g. on camping sites), portacabins, huts in allotment gardens, hunting lodges, vineyard huts, Alpine cabins and permanently anchored houseboats provided that the latter are continuously inhabited.

- 2 Please indicate the number of all dwellings in the building, irrespective of whether they are currently occupied or not. Accessory apartments in one-family houses have to be counted as separate dwellings. Please count in dwellings that were originally intended for residential purposes but are now used exclusively for commercial purposes (e.g. because they accommodate medical practices or offices). Do not include salesrooms. In the case of residential homes (where occupants have their own household), each self-contained unit is counted as a dwelling if it can be entered directly from a staircase, corridor or the outside.

- 3 Detached one-family or multi-family house: This may be, for instance, a detached one-family house or a detached high-rise building.

Semi-detached house: A building erected side-by-side with one other building, irrespective of whether the buildings are one-family or multi-family houses.

Examples of semi-detached houses:



Terraced house: A building erected side-by-side with at least two other buildings, irrespective of whether they are one-family or multi-family houses. The buildings need not have the same design and may be set in a staggered line

or at different levels. This category includes end of terrace houses.

Examples of terraced houses:



Other type of building: All types of buildings that are not detached, semi-detached or terraced houses and all types of inhabited accommodations.

- 4 Indicate the year in which the building was originally completed. For buildings that were rebuilt after complete destruction, the year of construction is the year in which the building was rebuilt. If you do not know the exact year of construction, please give an estimate.

- 5 Commonhold owners: This category applies always and exclusively where a building (multi-family house) has been split into freehold dwellings in accordance with the Housing Property Act. Commonhold owners must not be marked in the case of adjacent one-family houses (e.g. semi-detached or terraced houses) that have been split into separate units under the Housing Property Act. In that case, one of the other categories must be chosen (usually: private individual(s)). Private individual(s): All natural persons. This refers, for instance, to individuals, couples or other groups of persons, such as communities of heirs.

Housing co-operative: All housing companies with the legal form of a co-operative.

Private sector housing company: All housing companies under private law, excluding co-operatives.

Other private sector company: All companies under private law owning dwellings that were not primarily purchased to be rented out (e.g. banks, insurance companies, funds).

Municipality or municipal housing company: Company or institution in which the municipality holds more than 50% of the nominal capital or voting rights.

Federation, State: Company or institution in which the Federation or the State holds more than 50% of the nominal capital or voting rights.

Non-profit organisation: e.g. churches.

- 86 District heating: District heating is a system where the heat for entire residential areas is produced centrally in a district heating plant (also called long-distance heating).

Block heating: This applies where the heat for an entire block of buildings is produced by a central heating system whose heat source is inside, attached to, or very close to one of the buildings (also called short-distance heating).

Central heating: A central heating system supplies heat to all dwelling units in a building from one central source of heat that is located inside the building (usually in the cellar).

Single-storey heating: Single-storey heating is a system where all the rooms of a self-contained dwelling are heated from a central source of heat (e.g. gas furnace) that is usually located inside the dwelling.

Single-room or multi-room stoves: Single-room stoves (e.g. coal stoves or night storage heaters) only heat the room in which they are placed; usually, they are firmly installed. A multi-room stove (e.g. tile stove) supplies heat to several rooms at once (sometimes by air ducts).

Heat recovery systems/controlled ventilation systems in passive houses are considered here as "heating" and should be allocated to the relevant category (usually: central heating).

If there are several heating systems in a building, please mark the one that supplies heat for the largest proportion of the floor space.

Instructions on the questions about dwellings

A dwelling is defined as a self-contained unit for residential use that usually consists of adjoining rooms and enables the occupants to have their own household.

Dwellings have a separate entrance with direct access from the outside, a staircase or vestibule. The dwelling may however include cellar or attic rooms that are outside the actual dwelling but have been converted for residential use (e.g. mansards).

- 87 The first names and surnames of up to two persons who permanently or regularly live in the dwelling even if they are not staying in the dwelling on 9 May. In the case of let dwellings, please look up the names in the rental contract, if necessary. If you do not know the names of the persons actually living in the dwelling, please enter the names of the persons with whom the contract was made. For holiday dwellings (see 88), please enter the names of the persons who normally use the dwelling.

- 88 This question must only be answered for freehold dwellings in multi-family houses, that is, whenever the category "commonhold owners" was marked in question G5.

Private sector company: Including private sector housing companies.

Public sector companies, churches or the like: Companies or institutions in which municipalities, the Federation or the States hold more than 50% of the nominal capital or voting rights. Non-profit organisations such as churches also come under this category.

- 89 Occupied by owner: If at least one of the occupants owns the dwelling, even if the dwelling is occupied by sub-tenants besides the owner.

Please mark let for residential use if the dwelling has been rented out for residential use or if someone is allowed to use it rent-free.

A dwelling is considered as vacant if it is neither rented out on the reference day nor occupied by the owner and is not a holiday or leisure dwelling (see 90). If the dwelling cannot be used for some time because of alterations or renovations – while the tenancy continues – the dwelling is not considered as vacant.

The category exclusive commercial use applies if the dwelling is not used for residential purposes but only and exclusively for commercial purposes (e.g. as office, medical practice). Dwellings that are continuously let for commercial/hotel purposes as holiday dwellings are also used exclusively for commercial purposes.

- 90 Floor space: Total useful floor space of the dwelling. Rooms outside the self-contained dwelling (e.g. mansards) and cellar and attic rooms converted for residential use also form part of the dwelling.

To calculate the useful floor space (in accordance with the Useful Floor Space Ordinance), include:

- the full floor space of rooms/parts of rooms with a clear height of at least 2 metres;
- half the floor space of rooms/parts of rooms with a clear height of at least 1 metre but less than 2 metres; conservatories without heating, swimming pools and similar fully enclosed rooms;
- usually a quarter, but no more than half, of the floor space of balconies, loggias, roof gardens, terraces.

If you do not know the exact floor space, please give an estimate.

- 91 The number of rooms includes all living, dining and bedrooms and other separate rooms (e.g. habitable cellar and attic rooms) of at least 6m² as well as self-contained kitchens irrespective of their size. The bathroom, toilet, corridor and utility rooms are in principle not included. A living room with a dining area, bedroom area or kitchenette should be counted as one room. Accordingly, open-plan dwellings in which the living spaces are not structurally separate from each other (e.g. loft apartments) consist of only one room.

- 92 Bathtub or shower: Mark no if there are only portable shower units or bathtubs that are not permanently connected to the water supply and drain pipes.

- 93 The number of persons living permanently or regularly in the dwelling even if they are not staying in the dwelling on 9 May. If the exact number is not known, please indicate the number of persons living in the dwelling as known to you. For private holiday and leisure dwellings (see 94) please indicate the number of persons normally using the dwelling.

- 94 Holiday or leisure dwellings are dwellings where people only spend private leisure time (e.g. weekends, holidays, vacations, etc.). Holiday and leisure dwellings can be found in all kinds of buildings (e.g. weekend and holiday home, multi-family house). They may be used by the owner or be rented out permanently (or let free of charge) to a third party for leisure-time use.

This does not include holiday dwellings that are continuously let for commercial/hotel purposes.

The dwelling is a diplomat's dwelling or a dwelling of foreign armed forces if at least one of the occupants is a member of a foreign diplomatic mission, of foreign armed forces or the like, irrespective of whether they are owners or tenants.

Census of Buildings and Housing on 9 May 2011

Notification in accordance with Article 17 of the Federal Statistics Law

Purpose, type and scope of the survey, legal bases

Current official statistical data on buildings and dwellings are based on the results of the population census conducted in the old States in 1987 and of the census of buildings and housing conducted in the new States in 1995 as well as on intercensal estimates produced since that time. Updating those data and putting them on a new basis requires the collection of data on buildings and dwellings as part of the 2011 Census. Register data of the public administration cannot be used because there are no relevant registers. The survey covers all buildings with residential space, inhabited accommodations and dwellings.

To control the process of conducting the census of buildings and housing, the Federal Statistical Office has set up a register of addresses and buildings on the basis of the Act on Preparing a Register-Based Census of 8 December 2007 (Federal Law Gazette I, p. 2808), last amended by Article 3 of the Law of 8 July 2009 (Federal Law Gazette I, p. 1781). In that register, individual data on dwellings and owners are stored for every housing address.

The reference date of the census of buildings and housing is 9 May 2011.

The survey is conducted on the basis of the 2011 Census Act of 8 July 2009 (Federal Law Gazette I, p. 1781) in conjunction with the Federal Statistics Law of 22 January 1987 (Federal Law Gazette I, pp. 462, 565), last amended by Article 3 of the Law of 7 September 2007 (Federal Law Gazette I, p. 2246).

Data are collected in accordance with Article 6, paras. 2 and 3 of the 2011 Census Act.

The survey involves an obligation to provide information.

The obligation to provide information is laid down in Article 18, para. 2 of the 2011 Census Act in conjunction with Article 15, para. 1 of the Federal Statistics Law. Pursuant to those provisions, the owners, managers and other parties entitled to use or dispose of the buildings or dwellings are obliged to provide information.

Persons considered as owners include the persons to whom the buildings and dwellings are to be allocated in economic terms in accordance with Article 39, para. 2 of the Taxation Code.

To provide the information, an online questionnaire may be used instead of the paper questionnaire.

Managers who are not in a position to answer the questions are obliged to indicate the names and addresses of the owners so that the latter can be questioned. If you are no longer responsible for providing information on 9 May 2011, you are obliged to provide the names and addresses of the purchasers to the relevant statistical office. If you do not have that information, you must indicate the names and addresses of the persons (owners, managers and other parties entitled to use or dispose of the building or dwelling) who are able to provide the information.

Pursuant to Article 15, para. 6 of the Federal Statistics Law, objections and actions for rescission brought against the summons to provide information will not have a suspensive effect.

Statistical confidentiality

In accordance with Article 16 of the Federal Statistics Law, the individual data collected are kept secret. Only in exceptional cases explicitly provided for by law may individual data be passed on.

For transmission to the legislative bodies and for planning purposes, but not for the handling of individual cases, tables containing statistical results – even where table cells contain only one case – may be transmitted by the statistical offices of the Federation and the States to Supreme Federal or State authorities in accordance with Article 22, para. 1 of the 2011 Census Act.

Under Article 22, para. 2 of the 2011 Census Act, the Statistical Offices of the Federation and the States may, exclusively for purposes of municipal statistics, transmit individual data on the survey variables and on the auxiliary variables of "street" and "house number" or individual data grouped by block sides to the municipal statistical units for their sphere of competence. However, such transmission is permitted only if statistical confidentiality is ensured by legally required measures, especially with regard to separating the municipal statistical units in terms of space, organisation and staff from those units that are responsible for non-statistical tasks. The auxiliary variables must be deleted there at the earliest possible time, but not later than two years after transmission.

This means that the data must not be transmitted to the administration for administrative purposes. That prohibition to transfer the data back means, for example, that the population registers must not be corrected by means of the data collected by the Census.

Pursuant to Article 16, para. 6 of the Federal Statistics Law, only effectively anonymised individual data may be made available to institutions of higher education or other institutions entrusted with independent scientific research for the purpose of carrying out scientific projects.

The obligation to keep data secret applies to persons who have received individual data in accordance with the above provisions to the same extent as to staff members of the statistical offices.

Rights and duties of the interviewers

The census of buildings and housing is conducted as a written survey. In accordance with Article 11, para. 5 of the 2011 Census Act, however, interviewers may be employed in exceptional cases.

The interviewers must show their identification and have specially been sworn to secrecy. They must not use the information obtained from their activity in other procedures or for other purposes. This obligation will continue to apply after they have ended their activity as interviewers.

Auxiliary variables, separation and deletion, reference numbers

The family name, former names, first names and address of the persons obliged to provide information, telecommunication numbers of the persons responsible for providing information or of another person available for queries, the name and first name of up to two persons using the dwelling, the number of occupants per dwelling, the street, house number and address supplements of the dwelling are auxiliary variables which will only be used for the technical conduct of the survey. They will be separated from the survey variables at the earliest possible time, be stored separately and deleted as soon as the statistical offices have finished checking the survey variables and auxiliary variables for consistency and completeness, but not later than four years after 9 May 2011. The questionnaires will be destroyed after census processing has been finished, but not later than four years after 9 May 2011.

The data from the register of addresses and buildings on the family name and first name or on the name and address of the relevant owners, holders of a heritable building right, managers or other parties entitled to dispose of the buildings or dwellings, always in connection with the selected housing addresses, may be used to prepare and conduct sample surveys.

Those data as well as any other data contained in the register of addresses and buildings will be deleted at the earliest possible time, but not later than six years after 9 May 2011.

The questionnaire number contains a sequence of random digits and allows allocating the questionnaire to the person concerned. It also includes a check digit. However, it does not contain any information on the relevant person.

The activation code is a sequence of random characters which, together with the questionnaire number, is used to identify the respondents if they participate in the online survey.

The bar code printed on the bottom of the questionnaire represents the document code shown directly above (sequence of 38 digits). The document code is a reference number used only for organising the survey and processing procedure. It contains the questionnaire number, the information that this is a questionnaire of the census of buildings and housing as part of the 2011 Census referring to the reference date of 9 May 2011, and indicates which State is responsible.

The building ID is used to distinguish between the buildings included in the survey and to efficiently process the survey. It consists of a random serial number.

[참고자료 7-2]

대만 2010년 인구주택총조사 실시계획

(1) 총칙

1. 법적 근거: 행정원 주계처(이하 본처)는 통계법 제3조, 제4조, 제10조와 동법 시행세칙 제7조, 제25-제32조, 그리고 대만 2010년 인구주택총조사 방안 제15항 규정, 개정된 2010년 인구주택총조사(이하 센서스라고 칭함) 실시계획에 의거하여 센서스 사업 실시의 기준을 마련한다.
2. 센서스 기준시점: 본 센서스는 2010년 12월 26일을 센서스 기준일로 하며, 해당 일자의 자정을 기준 시각으로 한다. 또 기준일 전의 1주(12월 19일-12월 25일)을 센서스 기준주로 한다.
3. 센서스 실시기간: 본 센서스의 현지실사는 2010년 12월 26일부터 시작하여, 2011년 1월 22일에 종료한다.

(2) 센서스의 범위와 대상

4. 센서스 조사구역의 범위: 대만지구의 직할시(直轄市) 현(市) 및 복건성(福建省) 금문현(金門縣), 연강현(連江縣)을 포함하는 지역을 조사구역으로 한다.
5. 센서스 실시대상:
 - (1) 주택부문: 센서스 기준시점에 대만-복건지구의 표본조사구(sample census area) 범위에 있는 주택을 모두 센서스 대상으로 한다. 그 범위는 다음과 같다.

1. 사람들이 상주하는 주택, 건물 및 거처 (다만, 대만에 주재하는 외국인 공무원과 그의 가족들이 거주하는 주택, 건물 및 거처는 조사대상이 아님)

2. 비어있는 주택

(2) 가구 및 인구 부문: 센서스 기준시점에, 대만/북건지구의 표본조사구 범위에서 정해진 가구에 6개월 이상 상주하고 있거나, 6개월 이상 상주할 계획으로 되어 있는 인구를 센서스 대상 가구 및 인구로 하며, 그 범위는 다음과 같다.

1. 대만 국적을 가진 국민을 모두 포함하며, 정부가 해외에 파견한 공무원이나 그들의 가족도 여기에 포함한다. (다만, 해외에 유학, 사업 상 출국하여 6개월 이상 거주하지 않는 사람들은 제외한다.)

2. 국적이 외국이지만 중국본토, 홍콩 마카오에 적을 두고 있는 사람은 조사하고, 외국인 노동자나 외국인 교포는 포함한다. (다만, 대만에 주재하는 외국인 공무원이나 그들의 가족은 조사하지 아니한다.)

(3) 센서스 방식

6. 센서스 실시방법: 조사원이 표본조사구의 범위 내에 있는 가구를 직접 방문하여 조사하는 방식을 주로 하고, 인터넷조사나 우편조사의 방법으로 직접방문을 대신할 수 있다. 또, 행정자료 특히 등록부 자료를 연계하고, 센서스의 현실실사를 지원한다.

7. 자료제공 의무자: 본 센서스에서 가구주, 가구주의 대리인, 또는 가구내 여타 구성원 혹은 기타 관리인이 자료제공의 책임을 진다.

8. 센서스 자료의 비밀유지: 본 센서스의 개별자료는 엄격히 비밀을 유지하여야 하며, 통계분석에 이용되는 되는 것을 제외하고, 다른 용도로는 사용할 수 없

다.

(4) 표본추출방법

9. 표본추출 모집단: 대만지구에 대하여 지리정보시스템(GIS)를 이용하여 주소명부와 호적행정자료를 연계하여 수치화된 조사구(매 조사구당 110 ± 30 호로 책정함)를 설정하고, 복건성 금문현, 연강현 2개현에 대해서는 인위적으로 조사구를 구획하여, 표본추출 모집단으로 한다.
10. 추출방법: 일반조사가구는 대만/복건지구 기초자치단체(진, 시, 구)를 하위모집단으로 하여, 조사구를 최소추출단위로 하여 층화집락추출법(stratified cluster sampling)으로 추출하며, 모두 모집단 조사구의 16%를 표본조사 목적으로 선정한다. 한편, 조사구에서 집단가구(非普通住戶) 거주자가 100인을 넘는 경우에는, 원칙적으로, 전수조사 방식으로 현지실사를 관리한다.
11. 추계방법: 비추정법(ratio estimation method)으로 추정한다. 같은 기초자치단체(진, 시, 구)의 같은 층의 총수와 같은 층의 호적인구(戶籍家口) 총수를 이용하여 비추정(비推定)을 해서 각 하위 모집단의 특칭치(총수) 자료로 한다.

(5) 조사항목

12. 조사항목
- (1) 주택부문
1. 거주상태 2. 주택용도 3. 주택 총방수 및 부지면적
- (2) 가구부문
1. 가구구분 2. 주택(거처) 소유형태 3. 타지 주택소유 여부
2. 주택(거처) 거주기간

(3) 인구부문

1. 성명 및 성별 2. 생년월일 3. 국적소유 여부
4. 혼인상태 4. 가구주와의 관계
5. 상주상태 7. 교육수준 8. 사용 언어의 종류
9. 5년전 거주지 10. 가계운영의 책임 여부
11. 조사기준 주 취업상태 12. 취업 또는 취학 장소
13. 통근 또는 통학 상태 14. 현존 자녀수 및 거주지
15. 건강 상태

(6) 센서스 목적에 이용되는 등록부 자료

13. 본 센서스에 이용되는 공적 등록부 자료는 다음과 같다.

(1) 본국인구

1. 전가구 호적자료 파일: 내정부 호적담당부서가 책임지고 제공한다.
2. 내국인 출입국자료 파일: 내정부 출입국 및 이민담당부서가 책임지고 제공한다.
3. 수용자 및 수감자 파일: 법무부가 책임지고 제공한다.
4. 학적자료 파일: 내정부 호적담당부서가 책임지고 제공한다.
5. 건강보험자료: 행정원 위생담당부서가 책임지고 제공한다.
6. 장기보호, 요양 등의 파일, 기초생활수급자, 심신장애자 파일: 내정부 사회담당부서가 책임지고 제공한다.

(2) 외국인구

1. 외국인 교포인구 파일: 내정부 출입국/이민담당부서가 책임지고 제공한다.
2. 외국인 노동자 파일: 행정원 외국인노동자위원회가 책임지고 제공한다.
3. 홍콩, 마타오, 중국대륙에서 대만으로 온 자의 행정자료는

행정부 출입국/이민담당부서가 책임지고 제공한다.

(3) 건물

1. 村(里) 주소명부 파일: 내정부 호적담당부서가 책임지고 제공한다.
2. 주택 등 재산세 파일: 재정부 재산세 자료를 중심으로 책임지고 제공한다.
3. 縣(市) 주택세 파일: 재정부 재산세 자료를 중심으로 책임지고 제공한다.

(7) 센서스 명부

14. (센서스 명부 자료 출처) 센서스 명부를 작성할 때는, 가구번호, 주소, 호적상 호주 기본자료 등을 포함하여야 하며, 그 자료의 출처는 다음과 같다.

- (1) 내정부 2010년 6월 기준 호적자료 파일 및 村(里) 주소명부 파일
- (2) 각종 조사대상은 주관기관이 2010년 10월 하순 이전에 최신자료를 제공한다.
 1. 병영의 숙소, 군사학교, 훈련소, 군병원, 군교도소 등에 통상적으로 거주하는 장교, 부사관, 사병, 훈련병 등 군사요원이나 문관요원, 그리고 각 군 시설에 고용된 공원들의 자료는 국방부가 제공한다.
 2. 병영의 숙소에서 통상적으로 거주하는 해안경비 군인, 경찰, 문관 및 관련 인원의 자료는 행정원 해안경비부서(海岸巡防署)가 제공한다.
 3. 중앙경찰대학, 대만경찰전문학교에서 기숙생활을 하는 학생들의 자료는 내정부 경찰행정부서(警政署)가 제공한다.
 4. 수용기관에서 집중적으로 또는 개별적으로 기숙생활을 하는 일반 대체복무요원의 자료는 내정부 대체복무담당부서(役政署)가 제공한다.
 5. 행정원 재향군인회 소속 재향군인의 집이나 자비로 호스피스센터에 거주하는 재향군인과 그의 가족들의 자료는 재향군인회가 제공한다.
 6. 공사립 병의원 소속 요양원이나 휴양소에 거주기간이 6개월 이상

된 장기입소자는 행정원 위생담당부서(衛生署)가 관련자료를 제공한다.

7. 각 정부기관의 해외요원이나 해외에서 공동생활을 하는 가족, 국내에서 사적 고용계약에 따라 외국에 파견된 인원은 외무부가 관련자료를 제공한다.
8. 전문학교 이상 공사립 학교의 기숙사 거주 학생은 교육부가 관련자료를 제공한다.
9. 교도소, 교도요원 각종 교정기관의 수용인은 법무부가 관련자료를 제공한다.
10. 산업부문 외국인 노동자, 요양기관 외국인 요양노동자는 노동위원회 관련자료를 제공한다.

(3) 현장실사 과정에서 발견된 센서스 대상

15. 센서스 명부의 편제 및 응용

- (1) 내정부가 제공한 호적자료 파일과 촌(리) 주소파일을 컴퓨터에 입력하여 정리하고, 통계청에서 센서스 명부를 편제하고, 센서스 실시 이전에 지역본부에 명부를 전달한다.

- (2) 각종 센서스 명부는 센서스 작업에 이용한 후, 조사표와 함께 통계청으로 환송한다.

(8) 주관기관

16. (주관기관) 대만 통계청이 센서스 주관기관이 되고, 센서스 운영의 설계와 추진에 있어서 기획과 조정의 책임을 수행하도록 한다.

(9) 협력기관 및 업무분장

17. (협력기관) 본 센서스 대상 중에서, 각각의 유관 주관기관을 전수조사 책임자로 하며, 그 조사업무의 분장은 중앙은 각 유관 주관기관이 책임자가 되고, 나머지는 행정구역에 따라, 지방정부가 설비한 임시 센서스 조직이 책임자가 된다. 본 센서스 목적에 운용되는 공적 등록부 자료 파일은 중앙의 각각 유관기관이 책임을 지고 제공하여야 한다.

18. (전수조사범위) 각각의 유관 주관기관가 책임을 지는 전수조사의 범위는 다음과 같다.

- (1) 국방부: 병영의 숙소, 군사학교, 훈련소, 군병원, 군교도소 등에 통상적으로 거주하는 장교, 부사관, 사병, 훈련병 등 군사요원이나 문관요원, 각 군 시설에 고용된 공원
- (2) 행정원 해안경비부서(海岸巡防署): 병영의 숙소에서 통상적으로 거주하는 해안경비 군인, 경찰, 문관 및 관련 인원
- (3) 내정부 경찰행정부서(警政署): 중앙경찰대학, 대만경찰전문학교에서 기숙생활을 하는 학생
- (4) 내정부 대체복무담당부서(役政署): 수용기관에서 집중적으로 또는 개별적으로 기숙생활을 하는 일반 대체복무요원.
- (5) 행정원 재향군인회: 행정원 재향군인회 소속 재향군인의 집이나 자비로 호스피스센터에 거주하는 재향군인과 그의 가족들
- (6) 행정원 위생담당부서(衛生署): 공사립 병원 소속 요양원이나 휴양소에 거주기간이 6개월 이상 장기입소자
- (7) 외무부: 각 정부기관의 해외요원이나 해외에서 공동생활 하는 가족, 국내에서 사적 고용계약에 따라 외국에 파견된 인원
- (8) 교육부: 전문학교 이상 공사립 학교의 기숙사 거주 학생
- (9) 법무부: 교도소, 교도요원 각종 교정기관의 수용인
- (10) 노동위원회: 산업부문 외국인 노동자, 요양기관 외국인 요양노동자

19. (전수조사 센서스 인원): 이상의 전수조사에 필요한 각급 조사인원은 각 해

당 주관기관의 관련 직원을 위촉하거나 파견한다.

(10) 임시 센서스 조직

20. (전수조사 조직) 국방부 전수조사는 국방부에 인구주택총조사 본부를 설치하고, 나머지 전수조사는 각 상관 주관기관에 지도원을 두고, 대만 통계청의 지휘감독을 받으며, 해당 전수조사 범위의 인구주택총조사 업무를 관리한다.

21. (직할시, 현(시) 조직) 대만지구 각 직할시, 대만성 각현(시), 복건성 금문현, 연강현 정부는 별도로 인구주택총조사 지역본부를 두고, 대만 통계청의 지휘감독을 받아서, 해당지구의 인구주택총조사 업무를 관리한다.

22. 향(진, 시, 구) 조직: 각 향(진, 시, 구)의 관청에는 인구주택총조사 지역지부를, 대만 통계청 및 해당 상급 지역본부의 지위감독을 받아, 해당지역의 인구주택총조사 업무를 관리한다.

23. 센서스 조직 설치기간

- (1) 전수조사 조직: 2010년 10월 1일(기준일 3개월전)에 설치하고, 2011년 2월 28일에 해산한다. (국방부 인구주택총조사본부는 2011년 3월 15일에 해산한다.)
- (2) 직할시, 현(시) 조직: 2010년 9월 1일에 설치하여, 2011년 2월 28일에 해산한다.
- (3) 향(진, 시, 구) 조직: 2010년 9월 21일에 설치하여, 2011년 2월 20일에 해산한다.

(11) 조사구 구획

24. (조사구 및 지도구) 지리정보시스템을 운용하여, 대만지구 각 직할시, 현(시)는 수치화조사구(geo-coded enumeration area)를 설치하고, 복건성 금문현, 연강현 2개 현은 현장실사 방식으로 적당한 호수나 인구규모를 기준으로 조사구를 구획하며, 주변에 위치하는 약간의 조사구를 하나의 지도구로 하여, 현장실사작업 실시와 센서스 인력 배치를: 추진하는데 도움이 되도록 한다.
25. (선도구) 직할시, 현(시) 센서스 지역본부는 관할 향(진, 시, 구) 수, 인구수 및 선정된 조사구수를 기준으로 적당히 선도구를 구획하며, 원칙적으로 직할시, 현(시)에 1개의 선도구를 설치하도록 하며, 선정된 조사가 많은 경우는 선도구를 추가적으로 설치하여도 무방하다. 전수조사 기관도 선도구를 설치하여, 작업방식을 통일하여, 센서스 현장실사를 순조롭게 진행하는데 도움이 되도록 한다.
26. (조사구 구획 책임자)
- (1) 대만지구: 조사구 및 선도구 구획은 대만 통계청이 관리하며, 지도구 구획은 각 직할시, 현(시) 센서스 지역본부와 소속 향(진, 시, 구) 지역지부가 함께 관리한다.
 - (2) 복건성: 조사구 및 지도구는 금문현, 연강현 지역본부와 소속 향(진, 시, 구) 지역지부가 관리한다.

(12) 센서스 인원

27. 각급 센서스 인원배치는 본 센서스 사업을 추진하고 집행하기 위하여, 각 향의 작업수요에 의하여, 아래와 같은 센서스 인원을 배치한다.
- (1) 센서스 조직 행정인원: 각급 센서스 조직 책임자, 부책임자, 집행위원, 총간사, 부총간사, 팀장, 부팀장 및 간사 등 겸임인원은 각급 센서스 조직 설치 요강에 의하여 위촉하거나 파견한다. 센서스 조직의 책임자와 부책임자는 조직 설치 10일 전에 위촉/파견하며, 여타 공작인원은 해당 조직 설치

후 1주일 이내에 위촉/파견한다.

(2) 각급 조사인원:

1. 조사요원: 조사구마다 조사원 1인을 배치하며, 각 직할시, 현(시) 센서스 지역본부와 소속 향(진, 시, 구) 지역지부와 지도원이 함께, 현직의 민정, 호정, 복지, 교육, 재정 관련 이원, 촌(리)장 등이나 미취업의 전문대학 이상의 학력 소유자, 또는 민간의 관심있는 자를 선정하여 업무를 수행하도록 한다.
2. 지도요원: 지도구마다 지도원 1인을 배치하며, 각 직할시, 현(시) 센서스 지역본부와 소속 향(진, 시, 구) 지역지부가 현직의 민정, 호정, 재정, 교육 인원 등 기타 적당한 인원을 선정하여 위촉/파견하도록 한다.
3. 검토요원(審核員): 지도구마다 심핵원 1인을 배치하며, 직할시, 현(시) 센서스 지역본부는 현직의 민정, 호정, 재정, 지역통계조사 종사 인원 또는 공무원 중 우수한 지도원, 조사원 또는 적당한 교사인원을 위촉/파견한다.
4. 선도요원(督導員): 선도구마다 선도원 1인을 배치하며, 대만 통계청이 유관기관이나 통계조사 경험이 있는 자를 위촉/파견한다.

28. (조사인원의 보수규정) 조사원, 지도원, 심핵원은 센서스 기간 중 센서스 임무를 수행함에 있어서, 그 업무량을 고려하여 보수를 계상하도록 한다.

(13) 훈련 및 홍보

29. (훈련방식) 각급 센서스 인원의 훈련은 아래 열거한 방식에 따라 관리하도록 한다.

- (1) 센서스업무 세미나: 2010년 8월 중 대만 통계청이 주관하며, 참가인원은 직할시, 현(시) 정부 민정업무주관 국장, 주계처 처장 및 부처장(주계실 주임 및 부주임), 전수조사 담당기관 주관인원 및 대만 통계청이 지정하는 인원

- (2) 센서스 행정업무 관리체계 워크숍: 2010년 8월부터 2010년 9월까지, 참가인원은 직할시, 현(시) 정부가 주관하며, 참가인원은 대만 통계청 및 각 진할시, 현(시) 정부 및 향(진, 시, 구) 관청이 지정하는 인원
- (3) 지방간부인원 세미나: 2010년 9월부터 2010년 10월까지 직할시, 현(시) 센서스 지역본부가 주관하며, 참가인원은 지역본부장, 부본장, 집행위원, 정부총간사, 각조 팀장, 부팀장, 간사, 각 향(진, 시, 구) 센서스 지역지부 주임, 총간사 등
- (4) 강사 및 선도원 워크숍: 2010년 11월까지 대만 통계청이 주관하며, 참가인원은 센서스 교육강사, 선도원, 직할시, 향(시) 센서스 지역본부 및 국방부 센서스 본부 부장, 부조장 기타 전수조사 관리기관 간사 등의 인원
- (5) 센서스 실무교육회의: 2010년 12월 중, 직할시, 현(시) 센서스 지역본부와 소속 향(진, 시, 구) 지역 지부는 교통이 편리한 장소를 선정하고 구역을 나누어 관리하며, 참가인원은 행정인원, 검토원, 지도원, 조사원 및 기타 지정인원 등
- (6) 전수조사 실무교육회의: 2010년 12월 사이, 전수조사 관리기관이 자체적으로 실시하며, 참가원인 팀장, 부팀장, 간사, 및 각급 조사인원을 포함한다.
30. (강사 및 훈련교재): 본 센서스 실무회의 강사의 위촉/파견 및 훈련교재의 편제는 대만 통계청이 총괄적으로 관리한다.
31. (홍보사업): 본 센서스는 지방 센서스 조직이 성립한 후, 홍보사업을 전개하고, 현장실사 전 2주간에 걸쳐 집중적으로 홍보하며, 현장실사가 끝나면 종료한다. 지방 각급 센서스 조직은 통계청의 홍보작업계획에 의거하여, 적극적으로 지역특성에 맞는 홍보사업을 추진한다.

(14) 현지실사

32. (센서스 공고) 본 센서스는 현지실사 1개월 전에 행정원이 본 센서스의 법령에 의거하여, 목적, 기준시간, 실시시간, 구역범위 및 대상, 조사항목 등의 사항을 공고한다.
33. (조사표의 배포) 실시조사에 소용되는 각항의 조사표는 대만 통계청이 2010년 11월 말에는 각급 임시 센서스 조직에 배포하고, 해당 센서스 조직은 실무 교육회의 전에 개최하여, 지체 없이 관련 조사표를 각급 센서스 인원에게 배포하도록 한다.
34. (현지실사의 실시)
- (1) 방문조사: 센서스 실시기간, 조사원은 조사원증을 패용하고, 필요한 조사표건을 휴대하여야 하며, 직접 해당가구를 방문하여 조사를 관리한다. 실제 상황을 보고, 현장에서 조사표를 기입할 수 있거나, 해당가구의 대표자에게 조사표 기입을 요청하거나 인터넷으로 응답하는 것을 요구할 수 있다. 규정에 의거하여, 검토원은 조사표와 관련 표를 점검하고, 이것을 지도원에게 전달한다.
 - (2) 지도업무: 센서스 실시 기간, 지도원은 조사원의 방문조사, 특히 면접기술, 면접진행을 포함하여 어려운 문제를 협조 해결하는 방식으로 지도한다. 조사원이 조사표를 완성하면, 추가적으로 검토하고 수합하여, 해당 지역지부에 반송하여야 한다.
 - (3) 검토사업: 검토요원은 지도요원이 반송한 조사표를 접수하고, 필요한 경우에 편집하여, 규정에 의거하여 엄격히 재검하고, 누락이나 모순이 있으면 지도요원이나 조사용원이 재조사를 통하여 정정하도록 한다.
 - (4) 선도사업: 센서스 실시기간, 선도요원은 법규정에 따라 해당 센서스 조직과 더불어 현지실사 작업을 확실히 하고, 조사 진행을 파악하고, 각종의

어려운 문제를 해결한다. 수시로 조사표를 뽑아서, 유실 여부를 파악하고, 필요한 경우 철저한 감독을 통하여 센서스 자료의 품질을 확보하도록 한다.

35. (조사표 반송) 지방 각급 센서스 조직은 조사표를 재검하여, 규정에 의거하여 기한 전에 대만 통계청으로 반송하여야 한다.

(15) 사후조사

36. (조사목적) 본 센서스의 비표본오차(non-sampling error)의 수준을 측정하고, 자료품질을 개선함으로 해서, 각계에서 센서스 자료를 이용하는데 참고가 되도록 한다.

37. (사후조사 실시기간): 사후조사는 현장방문의 형태로 실시되고, 2011년 3월 1일부터 3월 15일까지 15일에 걸쳐서 실시한다.

38. (사후조사 방식): 적당한 표본을 선택하여, 조사원이 직접 면접하는 방식으로 관리한다.

(16) 자료처리

39. (자료처리방식): 본 센서스의 자료는 컴퓨터 처리를 주로 하고, 수작업으로 보완하여 상호 배합하여 진행한다.

40. (컴퓨터 소프트웨어 설계 및 처리) 본 센서스의 광학문자판독시스템(OCR: Optical Character Recognition) 설계는 민간회사에 위탁할 것이며, 기타 소프트웨어 설계 및 컴퓨터 작업은 대만 통계청 컴퓨터 자료처리센터가 책임을 지고 처리한다. 컴퓨터 소프트웨어 설계는 2010년 12월 말 이전에 완료한다.

41. (자료등록)

(1) 종이 조사표: 조사표는 광학문자판독시스템(OCR)을 채택하여, 자료를 등록하며, 내검 작업(internal consistency check)을 진행하여, 자료의 정확도를 개선한다.

(2) 인터넷조사표: 인터넷상으로 응답한 자료는 네트워크 상 점검을 완료하여, 센서스 자료파일에 등록한다. 인터넷조사 시스템 설계는 민간회사에 위탁한다.

42. (결과표식) 본 센서스의 통계결과 표식은 센서스 조사항목을 교차분류하고, 통계이용자의 요구를 존중한다.

43. (결과표의 인쇄): 본 센서스의 통계자료는 컴퓨터 점검과 수작업 점검을 거친 후, 조사표 결과표식을 정하고, 분기별로 아래의 각 통계표를 작성한다.

(1) 속보 센서스 결과표: 2011년 6월말까지 완성

(2) 센서스 종합보고 결과표: 2012년 6월 말까지 완성

(3) 센서스 보충보고 결과표: 2012년 11월 말까지 완성

(17) 센서스 보고서 심의 및 배포

44. (결과분석 및 심의) 본 센서스 자료의 통계결과는 주요 조사항목에 대하여 실시되고, 각종 상관자료를 비교분석하여, 분석 의견과 함께 첨부되어야 하며, 그 결과는 대만통계청에 보고되면, 센서스 위원회가 심의한다.

45. (행정원 결과요약 보고) 본 센서스 결과의 속보자료와 총괄보고 자료는 센서스 위원회가 심의한 이후, 분석결과를 요약설명하여, 행정원에 보고하여 최종 확정한다.

46. (센서스 보고서 배포) 본 센서스 자료의 통계결과는 체계적 분석을 거쳐서, 보고서를 작성 배포하여 센서스 자료의 활용을 증진시키는 것이다. 센서스 보고서의 제책작업은 대만 통계청이 총괄적으로 관리하며, 배포작업은 다음과 같이 실시된다.

- (1) 센서스 속보자료 보고서: 2011년 7월 22일 전 완성
- (2) 센서스 총괄보고서: 2012년 7월 이전 완성
- (3) 센서스 보충보고서: 2012년 12월 전 완성

47. (센서스 결과 파일작업): 상세결과 통계표는 파일로 작성하여 장래사용에 대비하여야 하며, 센서스의 주요항목은 표준분류방식에 따라 분류되어야 하며, 파일의 내용, 포맷, 응용방법 등에 관한 사용설명서를 첨부하여 관련 정부기관이나 유관조직에 배포하여야 한다.

부 칙

- 48. (사업의 진행) 센서스 운영의 모든 단계는 상세한 운영방법이나 작업 스케줄에 대한 실질적 필요성에 의거하여 점진적으로 추진되어야 한다.
- 49. (응답의무 및 권리) 통계법 제20조 및 제21조의 규정에 의거하여, 본 센서스에 대한 응답자는 상세한 보고를 할 의무가 있으며, 각급 보사인원은 그 직책을 이용하여 응답자의 권리를 침해할 수 없다.
- 50. (센서스조직 직인) 각급 센서스 조직은 대외적으로 공문서를 발행할 때에는 직인과 서명인을 사용하여야 한다. 겸임 또는 본직이 되는 기관의 책임자는 해당기관의 직인을 대신 사용할 수 있다.
- 51. (포상 및 징계) 센서스 운영이 종결된 후, 센서스 사업의 성과를 평가한다. 각급 센서스 조직 및 센서스 인원을 센서스 작업을 처리하고, 그 조사의 자료 품질이 우수한 자는 포상을 하고, 센서스 운영에 태만이나 오류가 있었던 사람은 적정한 처벌을 한다. 그 중 센서스 인원의 업무성과법에 의거하여 각 센서스 조직을 포상하거나 징계할 수 있다.

52. (센서스 경비) 본 센서스의 소요경비는 대만통계청의 사업계획에 의거하여, 연도별 나누어 편성하고, 규정에 의거하여 지출용도를 명시하며, 각 전수조사의 주관기관이나 각급 센서스 조직은 상급 정부기관은 실제 소요비용을 부분 예산으로 편성하여, 배합하여 운용한다.
53. 검토 및 평가: 본 센서스 업무를 완료한 이후, 센서스의 자표품질과 관련하여, 응용, 작업규칙, 집행과정을 분석하고, 검토 및 평가를 진행하고, 후속연구의 진행을 관리한다.

참고문헌

1. 한국어

- 건축법(2009), 건축법 시행령, 건축물대장의 기재 및 관리 등에 관한 규칙, 법제처 홈페이지
- 김근영.김순관.이신해(2003), "Comparing Census Data with Registration Data with Respect to Population and Household in the Seoul Metropolitan Region." 『국토계획』 38(1): 237-249.
- 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 보장”, 『통계연구』 13(2): 1-40. 통계청.
- 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.
- 이내성(2006), “주민등록 행정자료를 활용한 인구통계 조사방법 연구.” 미발간 통계연구회 발표자료.
- _____(2008), “총조사에 주민등록 행정자료의 활용을 위한 자료매칭 연구”. 조사연구 9(2): 119-149
- 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.
- 이지연(2007a). “2010년 인구주택총조사 방법론 연구 기초자료.” 미발간 통계

청 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.
- 주민등록법(2009), 주민등록법 시행령, 법제처 홈페이지
- 통계청(2004a) 『2005 인구주택총조사 기본계획안(요약)』.
- _____(2004b), 『지역통계 생산을 위한 도시화지역 설정 연구』.
- _____(2005). 『2005 인구주택총조사 조사지침서』.
- _____(2006a). 『2005 인구주택총조사 종합평가보고서』.
- _____(2006b). 『도시화지역 획정을 위한 시스템 개발』.
- _____(2006c). 『통계지리정보시스템 구축을 위한 ISP/BPR 사업』.
- _____(2007). 『2010 인구주택총조사 추진기본계획 (잠정안)』.
- _____(2007), 건축물대장 활용방법 연구, 2007
- _____(2007), 2010년 인구주택총조사방법론 연구
- _____(2008), 등록센서스 실시 기본계획, 2008
- _____(2012), 국가통계포털(Korea Statistical Information Service).
(<http://www.kosis.kr/>)
- 통계교육원 (2008), 인구통계 기초와 응용, 2008
- 통계개발원 사회통계실(편) (2008a), 『행정자료의 통계적 활용』, 통계청 통계개발원.
- _____(2008b) 『센서스 방법론 연구』, 통계청 통계개발원.
- 행정안전부(2008), "주민등록법 일부개정법률(안) 입법예고." 행정안전부 공고 제2008-51호(2008년 5월 28일).
- _____(2011), 『행정구역 실무편람』. 행정안전부

한국인구학회(2010), 「행정자료를 활용한 등록센서스 시산에 관한 연구」, 통계청 인구총조사과 등록센서스 최종보고서, 한국인구학회.

_____ (2011), 「행정자료 기반 센서스 실시국가의 심층연구를 통한 방법론 개발 연구」, 통계청 등록센서스과 등록센서스 최종보고서, 한국인구학회.

2. 외국어문헌

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

中華民國 行政院(2009b), 九十九年人口及住宅普查-公務登記資料連結備對 作業

中華民國 行政院(2009c), 九十九年人口及住宅普查抽樣設計

江桂英、謝明憲(2010), 普查母體建置採用資料庫處理之研究－以人口及住宅普查為例, 中國 行政院 主計處 電子中心分析設計組

顏貝珊(2009) "2010년 各國戶口普查制度之研究"(An International Study of 2010 Population Census Methods), 臺灣 國立政治大學 統計學科 碩士學位論文

劉訓蓉、劉惠玲、周元暉、程郭義、王曉雲(2007), 2010年人口及住宅普查構想與發展之研究, 中國 行政院 主計處 第4局 秘書室

Abbott, Own. (2009) "2011 UK Census Coverage Assessment and Adjustment Methodology", Population Trends 137: 25-32.

Australian Bureau of Statistics (2006), A Guide to Small Area Estimation - Version 1.1, Canberra, Australia.

Ballano, Carlos (2008), "A Census of Population Based on an Administrative Register", Proceedings of Statistics Canada Symposium 2008, Data Collection: Challenges, Achievements and New Directions.

Blum, Olivia (2005), "Wholistic Approach to Multifaceted Integrated Census", Central Bureau of Statistics, Jerusalem, Israel. (<http://www.fcs.gov/03papers/Blum.pdf>)

Brown J, Abott O and Smith P A (2011) Design of the 2001 and 2011 Census Coverage Surveys for England and Wales. Journal of the Royal

- Statistical Society: Series A (Statistics in Society), Volume 174:4, 881-906.
- Chapman, D. (1951). "Some properties of the hypergeometric distribution with applications to zoological censuses". University of California Publications in Statistics 1: 131-160.
- Duncan, Otis and Beverly Duncan (1955), A Methodological Analysis of Segregation Indexes. American Sociological Review 20: 210-217.
- 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).
- Elkin M, Dent P and Rahman N (2012) 'Review of International Approaches to estimating and adjusting for over and under coverage': Beyond 2011 Research Report R2 , to be made available on the ONS website at <http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/news-and-events/reports-and-publications/index.html>
- 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.

- Fellegi, I. P. and A. B. Sunter (1969), "A theory of Record Linkage", *Journal of American Statistical Association* 64: 183-1210.
- Fienberg, S. E. (1992). "Bibliography on capture-recapture modelling with application to census undercount adjustmen", *Survey Methodology* 18: 143-154.
- Gerull, Susanne and Paul-Linke-Ufer (2008) "Census 2011 in Germany: The use of population register data for quantifying certain subgroups of homeless people". Berlin, Germany.
- 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.
- Golata, Elzbieta and Jan Paradysz (2007), "Indirect Estimation in Population Census 2011 in Poland", Department of Statistics, Poznan University of Economics, Poland, *Studia Demograficzne (Demographical Studies)* Vol: 1: 40-60.
- Herrador, Montserrat and Juana Porras (2008), "The use of administrative register to improve sampling frame. Experiences and possibilities in Spain", Seminar on Registers in Statistics - methodology and quality 21 - 23 May, 2007 Helsinki.
- Hogan, H. (2003) "The accuracy and coverage evaluation: Theory and design", *Survey Methodology* 29(2): 129-138.
- Houbiers, M., P. Knottnerus, A. H. Kkoese, R.H. Renssen, and V. Snijders. (2003). "Estimating Consistent Table Sets: Position Paper on Repeated Weighting." Discussion paper 03005, Statistics Netherlands: Voorburg / Heerlen.
- Kamen, Charles S. (2005) "The 2008 Israel Integrated Census of Population and Housing: Basic conception and procedure". Central Bureau of Statistics, Israel (http://www1.cbs.gov.il/www/mifkad/census2008_e.pdf)
- 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.
- Luo, Kuo-Hua (2011), "2010 Population and Housing Censuses in Taiwan", A paper presented at the International Workshop on Internet Survey 2011, September 7-8, Daejeon, Korea
- Kuha, Jouni and Firth, David (2011) On the index of dissimilarity for lack of fit in loglinear and log-multiplicative models. *Computational statistics & data analysis* 55 (1): 375-388.
- 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.
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press.
- Mule Thomas (2008). 2010 Census Coverage measurement Estimation methodology. U.S. Bureau of the Census, Washington, DC.
- Munnich, R. (2008). "Varianzschatzung in komplexen Erhebungen", (Variance Estimation in Complex Surveys). *Austrian Journal of Statistics* 37:319.334.
- _____, Bihler, W., Bjørnstad, J., Davison, A., Sardy, S., Haslinger, A., Kottner, P., Laaksonen, S., Ohly, D., Schurle, J., Wiegert, R.,

Oetliker, U., Renfer, J.-P., Quatember, A., Skinner, C., and Berger, Y. (2003). Data quality in complex surveys. DACSEIS Deliverable 1.1. www.dacseis.de.

_____, Gabler, S., Ganninger, M., and Thees, N. (2008). Zensus Stichproben-Projekt, Technical Annex. Version: 5. Februar 2008. 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.

_____, Gabler, S., Ganninger, M., Burgard, J. P., and Kolb, J.-P. (2012). Stichprobenoptimierung und Schätzung im Zensus 2011 (Sample Optimization and Estimation in Census 2011). Wiesbaden: Statistisches Bundesamt.

Myrskylä, Pekka (2000), Effects of Changes to Registers: Experiences from Finnish Register-based Censuses." INSEE-Eurostat Seminar on Census after 2001, Paris.

National Research Council (2004), *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.

_____. (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.

_____. (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.

-
- _____ (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.
- Park, M. and Cho, H. (2008). Minimum MSE regression estimator with estimated population quantities of auxiliary variables. *Computational Statistics and Data Analysis* 53: 394-404.
- Pfefferman, Danny (2002). "Small area estimation-New developments and directions", *International Statistical Review* 70(1): 125-143.
- Paolo Valente (2011), "Census taking in Europe: how are populations counted in 2010?", *Population and Societies*, no. 467, May 2010
- Park, M. and Cho, H. (2008), Minimum MSE regression estimator with estimated population quantities of auxiliary variables. *Computational Statistics and Data Analysis* 53: 394-404.
- Rao, J.N.K. (2003), *Small area estimation*, New York: Wiley.
- _____ and Singh, A. (1997), "A ridge shrinkage method for range restricted weight calibration in survey sampling", In *Proceedings of the Section on Survey Research Methods*, American Statist Association, 57-64.
- Renaud, A. (2007) "Estimation of the coverage of the 2000 census of population in Switzerland: Methods and results", *Survey Methodology* 33:199-210.
- Saei, A., Zhang, L.-C., and Chambers, R. (2005). Generalised structure

- preserving estimation for small areas. *Statistics in Transition* 7(3): 685-696.
- Sarndal, C. E., Swemsson, B and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.
- Sekar, G. A. F. and Deming, W. E. (1949). "On a method of estimating birth and death rates and the extent of registration". *Journal of the American Statistical Association* 44:101-115.
- Siegel, J. and David A. Swanson (eds.) (2004), *The Methods and Materials of Demography*, New York: Academic Press (Elsevier)
- Smith, Paul (2007), "Targetting 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.
- Statistics Israel (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.
- Swiss Federal Statistical Office (2007), Die Harmonisierung amtlicher Personenregister-Kantonale und kommunale Einwohnerregister Amtlicher Katalog der Merkmale., Version 1.0, Neuchâtel
- (스위스 연방통계청이 등록부조정법이 통과되고 나서, 광역자치체와 기초자치체의 공적 인구대장의 특징과 표준화를 위한 방안을 제시한 매뉴얼)
-
- (2009). Merkmalskatalog. Eidgenössisches Gebäude- und Wohnungsregister. Version 3.4, Neuchâtel.
- Statistics 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.
-
- (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

_____ (2012), "The Integrated Rolling Census in Israel", Conference of European Statisticians, Economic Commission for Europe. Paris, 6-8 June, 2012.

Statistics Singapore (Department of Statistics) (2002), Census of Population, 2000 Administrative Report. Department of Statistics, Ministry of Trade and Industry, Singapore.

_____ (2011), Census of Population, 2010 Administrative Report, Department of Statistics, Ministry of Trade and Industry, Singapore. 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.

Statistics Spain (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.

_____ (2005): "Outline of the census type planned for 2011 in Spain". Submission to the United Nations Statistics Division Website on 2010 World Population and Housing Censuses., National Institute of Statistics, Madrid, Spain,
<http://unstats.un.org/unsd/Demographic/sources/census/Spainpdf.pdf>

_____ (2011), "Demographic Census Project 2011", Subdirectorato-General for Statistics on the Population,, National Institute of Statistics of Spain.

Statistics Turkey (Turkish Statistical Institute) (2008a), "Address Based

Population Registration System 2007 Population Census Results”, January 21, 2008, Turkey.

(2008b), "New method for 2010 population and housing census of Turkey: Considerations about data quality and coverage" Joint UNECE/Eurostat Meeting on Population and Housing Censuses, at the Eleventh Meeting Geneva, 13-15 May 2008

Statistics Sweden (2011), "Quality Assessment of Administrative Data", Methodology reports from Statistics Sweden.

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) (2003), "Alternative to a Census: Review of Interantional Approaches." Census Strategic Development Review. (<http://www.statistics.gov.uk>.)

(2012a), "Beyond 2011: Administrative Data Sources Report: NHS Patient Register." United Kingdom Office for National Statistics.

(2012b), "Beyond 2011: The Options Explained." United Kingdom Office for National Statistics.

(2012c), "Beyond 2011: Exploring the Challenges of Using Administrative Data." United Kingdom Office for National Statistics.

(2012d), "A review of international approaches to estimating and adjusting for under- and over-coverage." a paper prepared by Meghan Elkin, Pamela Dent, and Nargis Rahman, United Kingdom Office for National Statistics

United Nations (1969), *Methodology and Evaluation of Population Registers and Similar Systems*. Department of Economic and Social Affairs, New York: United Nations .

-
- _____ (2001), *Principles and Recommendations for a Vital Statistics System. Revision 2*, Department of Economic and Social Affairs, New York: United Nations.
-
- _____ (2008), *Principles and Recommendations for Population and Housing Censuses, Revision 2*, Department of Economic and Social Affairs, New York: United Nations.
-
- _____ (2009), *Fundamental Principles of Official Statistics. United Nations*. Department of Economic and Social Affairs, New York: United Nations. (unstats.un.org/unsd/dnss/gp/fundprinciples.aspx)
- UNECE (2006), *Recommendations for 2010 Censuses of Population and Housing*, Conference of European Statisticians, United Nations Economic Commission for Europe, New York: United Nations.
-
- _____ (2007), *Register-based Statistics in the Nordinc Countries: Review of Best Practices with Focus on Population and Social Statistics*. New York: United Nations.
-
- _____ (2012), "Overview of the 2010 round of population and housing censuses in the United Nations Economic Commission for Europe region", Economic Commission for Europe Conference of European Statisticians, sixtieth plenary session Paris, 6-8 June 2012 Item 3 (a) of the provisional agenda 2010 round of censuses – innovations and lessons learned
- 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>.
- Wagstaff, Heather and Steven Rogers (2007), "Optimising 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, Anders and Britt Wallgren (2007), *Register-based Statistics: Administrative Data for Statistical Purposes*, Wiley Series in Survey Methodology. New York: John Wiley and Sons.

Zaslavsky, A. M. (1989) "Multiple-system methods for census coverage evaluation", in Proceedings of the Survey Research Methods Section American Statistical Association.