

2015년 순환센서스 추정 및 분석방안에 대한 연구

2015년 12월 24일

통 계 청

제 출 문

제 출 문

통 계 청 장 귀 하

본 보고서를 “2015년 순환센서스 추정 및 분석방안에 대한 연구” 과제의 연구결과보고서로 제출합니다.

2015년 12월 24일

한국인구학회 김 석 호

연 구 진

연구책임자 김석호(서울대학교 사회학과 교수)
공동연구자 강정한(연세대학교 사회학과 교수)
강현철(호서대학교 응용통계학과 교수)
박민규(고려대학교 통계학과 교수)
송순관(뉴질랜드 통계청 박사)
연구보조원 문영민(호서대학교 응용통계학과 석사과정)
박서연(고려대학교 통계학과 석사과정)
신인철(서울대학교 사회발전연구소 연구조교수)
조진만(연세대학교 사회학과 석사과정)
하헌주(성균관대학교 사회학과 박사수료)

연구결과보고서 요약문

연구과제명	2015년 순환센서스 추정 및 분석방안에 대한 연구		
중심단어	총조사, 순환센서스, 조사방법론, 순환표본조사		
연구기관	한국인구학회	연구책임자	김석호
연구기간	2015. 8. 18 ~ 2015. 12. 24		

제1장 연구배경 및 구성

1. 연구의 배경

인구주택총조사는 1925년 처음 도입된 이후로 매5년마다 실시되는 조사로, 인구·가구·주택부문에 있어서 국가통계의 기반이 되어 왔다. 인구주택총조사는 전수조사(short-form)와 표본조사(long-form)라는 두 개의 방식으로 진행되어 왔는데, 조사환경이 변화됨에 따라 새로운 대안의 모색이 요구되고 있다. 이에 통계청은 인구주택총조사 패러다임의 전환을 모색하면서 다양한 방식의 연구를 수행하게 되었다. 먼저, 5년마다 대규모로 진행되던 인구주택총조사는 2016년부터 전수 부문은 등록센서스 방식을 도입하였으며, 등록센서스가 공표할 수 없는 항목을 포함한 기존 10% 표본조사 대신 20% 표본조사가 실시하고 있다. 또한 2021년 이후에는 표본조사를 5년에 한번 실시하는 것이 아닌, 매월 표본을 균등 분할하여 실시하는 순환센서스 방식 도입을 검토 중이다. 그러나 순환센서스 방식을 이행하기 위해서는 복잡한 표집 및 모형 기술, 우량의 표본틀이 필수적이며, 센서스가 제공해온 전체 모집단에 대한 일정 시점의 스냅샷이 더 이상 존재하지 않는

다는 점이 고려되어야 한다. 이러한 문제의식을 바탕으로 통계청은 2012년부터 순환센서스 도입을 위한 다양한 연구를 수행하여 왔다. 특히, 한국인구학회에 의뢰하여 2014년 5월부터 2014년 12월까지 약 8개월 간 ‘순환센서스 추정 및 분석 방안’에 대한 연구를 실시한 바 있다. 그러나 이 연구는 순환센서스와 기존 센서스 및 등록센서스와의 관계, 순환센서스 디자인, 모수추정 방법론, 순환센서스 활용방안 등에 대한 기초 연구에 그쳤다. 이에 본 연구에서는 2014년도의 연구 성과를 토대로 보다 심도 있는 연구를 수행하는데 목적을 두고 있다. 본 연구는 순환센서스를 도입하기 위해서는 모집단 구축과 표본설계, 모수 작성과 추정, 결과 분석 등을 중점적으로 검토할 필요가 있음을 강조한다. 이에 본 연구는 2014년 연구 결과를 바탕으로 실제적인 표본설계와 추정 및 결과 분석을 수행하고 있다.

제2절 연구의 구성

본 연구는 ‘인구주택총조사 표본규모할당 및 추정방법’, ‘순환표본조사 시행을 위한 표본설계 및 추정방안 연구’, ‘시험조사 자료의 심층 분석’, 그리고 ‘한국에서의 순환센서스 기능’ 등 크게 4개 부문으로 구성되어 있으며, 각 부문별 주요 연구내용은 다음과 같다.

제2장의 ‘인구주택총조사 표본규모할당 및 추정방법’은 크게 3개의 연구내용으로 구성되어 있다. 먼저, 제1절에서는 표본규모 할당방법에 대한 연구를 수행하였는데, 조사방법의 변화에 부응하기 위하여 정확도 높은 추정을 위한 표본설계의 기본방향을 제시한다. 제2절에서는 추정가중치 산정을 위한 시험조사 자료를 분석하였다. 제3절은 가중치 산정 절차를 제시하였다. 인구주택총조사의 승수 산정 방법은 읍면동을 부차모집단(sub-population)으로 하고, 부차모집단을 조사구 특성에 따라 층화(아파트조사구 및 일반조사구)하여, 가구인구 및 주택부문으로 나누었다. 그런 후에, 가중치산정 방법들을 검토하여 새로운 조사방법을 도입하는 인구주택총조사의 추정에 있어서 정확도를 높일 수 있는 방법을 제안하였다.

제3장은 ‘순환표본조사 시행을 위한 표본설계 및 추정방안 연구’로서, 제1절에서는 연구의 전체적인 연구내용을 개괄하고 있다. 그리고 제2절에서는 순환표본조사를 위한 표본설계 방안을 제시하였다. 제3절에서는 모수추정방안을 제시하였다. 먼저, 기간추정(Multi-year estimation or estimation for period) 개념에 대한 정의를 하고, 정의된 기간추정의 개념 하에서 모수 추정을 위한 가중치 작성내용을 우선적으로 다루었다. 추가적으로 추정 시 필요한 내용들도 검토하였다. 제4절에서는 기간 추정을 위한 기타 고려사항들을 제안하였다.

제4장은 ‘시험조사 자료의 심층 분석’을 실시하였다. 인천광역시 연수구와 경상남도 거창군의 시험조사 자료를 심층 분석함으로써 주요 항목별 변화 추이 및 시험조사에서의 무응답 현황 등 문제점을 살펴보았다.

제5장은 순환센서스의 인구사회학적, 정책적, 학문적 기능을 어떻게 극대화시

킬 수 있는지 모색하였다. 이를 위해 우선 이러한 기능을 성공적으로 수행하고 있는 미국의 순환센서스, ACS부터 검토하였다. 이와 더불어 다양한 사회조사를 순환센서스에 통합시킬 경우 어떻게 통합조사할 수 있는지에 대한 예시를 제시하였다. 이를 위해 통계청의 ‘사회조사’와 ‘지역별 고용조사’, ‘한국종합사회조사’ 및 OECD의 각종 사회지표를 검토하였다.

제6장은 ‘향후 순환센서스 연구를 위한 제언’에 해당되며, 각 장별 연구를 통해 도출된 함의들을 정리하고 2016년 이후의 순환센서스 연구를 위한 정책적 제언과 향후 과제들을 제시하였다.

제2장 인구주택총조사 표본규모할당 및 추정방법

제1절 표본규모 할당방법

과거 인구주택총조사는 조사방법이 전수와 표본을 혼합했었는데, 2015년 인구주택총조사는 일부 조사항목은 행정센서스 자료를 이용하고 나머지 조사항목은 표본조사를 하는 혼합형 조사방법으로 전환된다. 이 장은 정확도 높은 추정을 위한 표본규모 할당방법들을 검토하여 표본설계 기본방향을 제시하고자 하는데 그 목적이 있다.

표본설계자는 모집단의 특성, 표본들의 구성여부, 조사예산 및 추정의 정확도를 고려하여 표본설계 방법 및 표본규모를 결정하는 것이 일반적이다. 인구주택총조사는 전체 표본조사구가 사전에 설정되어 있으므로, 부차모집단으로 정의한 시도/시군구/읍면동으로 표본을 가장 합리적으로 할당하여 부차모집단의 추정치의 정확도를 향상시킬 수 있는 방법을 모색해야 한다.

표본조사가 생산하는 주요 통계치는 읍면동 부차모집단의 추정의 정확도를 고려한 설계를 요구한다. 읍면동별 높은 정확도가 요구되는 주요 통계치는 읍면동별 교육정도별 인구수, 혼인상태별 인구수, 가구원수별 가구수 및 거처종류별 가구수 등이다. 과거 인구주택총조사가 공표한 통계치는 시군구를 기준으로 하였는데, 2015년 조사는 읍면동 수준으로 통계치를 작성하는 것을 기본목표로 삼고 있다.

전체표본 조사구수를 행정구역별로 할당하는 방법은 크게 읍면동 직접 할당법 및 간접 할당법으로 구분된다. 읍면동 직접 할당법은 시도 및 시군구의 지역을 고려하지 않고 읍면동을 부차모집단(subpopulation)으로 지역층화하여 전체 표본조사구수를 읍면동에 직접 할당하고, 시군구 및 시도는 할당된 읍면동 표본조사구수를 누적하는 방법으로 간접적으로 표본규모를 계산한다. 읍면동 간접 할당법은 전국을 시도의 일차 부차모집단(first subpopulation)으로 지역층화하여 전체 표본조사구수를 먼저 시도에 할당하고, 할당된 시도의 표본 조사구수를 기초로 각 시도내의 시군구를 이차 부차모집단(second subpopulation)으로 지역층화한다. 이를 통해 시군구의 표본규모를 할당한 다음, 각 시군구내의 읍면동을 삼차 부차모집단(third subpopulation)을 지역층화하여 읍면동에 표본조사구수를 할당

한다.

이 장은 부차모집단의 분산 정도를 이용하여 표본규모를 부차모집단에 할당할 수 있는 네이만 배분법과, 부차모집단의 분산에 관계없이 사용할 수 있는 동등배분법(Equal allocation), 비례배분법(Proportional allocation), 제곱근배분법(Square root allocation), 거듭제곱배분법(Power allocation), 키쉬배분법(Kish allocation) 등의 적합성 여부를 검토하였다.

가령, 네이만 배분법은 특정 조사 항목을 설계변수로 선정하고, 과거 조사 자료를 이용하여 분산을 구한다. 이 방법의 적용을 위해서는 인구주택총조사에서 어떠한 항목들을 설계변수로 선정하고 변수의 분산에 따라 표본규모를 배분하는 문제가 중요하기 때문에 조사항목수 및 설계항목의 분산에 구애받지 않고 사용할 수 있는 표본배분법이 중점적으로 검토되어야 한다. 동등배분법, 비례배분법, 제곱근배분법, 키쉬배분법의 특성을 살펴보면, 전국/시도/시군구 추정에 중점을 두면 비례배분법이 적합하고, 읍면동만 추정에 초점을 맞추면 동등배분법이 적합함을 알 수 있었다.

가상적인 자료를 이용한 분석에서는, 상대적으로 규모가 작은 읍면동은 동등배분법이 효율적인 방법인 반면, 비례배분법은 효율성이 상대적으로 떨어지는 방법으로 판단된다. 또한 상대적으로 규모가 큰 동읍면은 이와 반대의 결론이 나타났다. 이러한 규모의 양극화 현상을 중립적인 방향으로 효율성을 조정할 수 있는 방법을 강구하는 것이 필요해 보인다.

시군구의 표본규모는 상대적으로 크기 때문에, 그 추정치가 일정한 오차의 범위 이내에서 정확도를 확보하는 것에 큰 문제는 없어 보인다. 읍면동 추정정도에 설계 초점을 맞추는 표본규모 할당 방법인 읍면동 직접 할당법이 설계의 기본 방향으로 설정되어야 한다. 규모가 큰 동읍면의 추정 정확도를 중시해야 하는지 아니면 규모가 작은 동읍면의 추정 정확도를 중시해야 하는지는 쉽게 결정하기 어려운 문제이다.

추정의 정확도는 표본크기가 클수록 높아지기 때문에, 표본 크기로 인한 정확도의 양극화 현상을 표본설계를 통해 일정 정도 완화할 수 있다. 본 절에서는 부차 모집단인 읍면동의 추정치와 시군구의 추정치의 상대적 중요성에 대한 고려가 이론적으로 견고하며 조사항목이 많은 방법에 적합한 키쉬방법이 표본규모 할당방법으로 제안되었다.

제2절 추정가중치 산정을 위한 시험조사 자료분석

본 절은 항목별 응답상태 및 응답특성을 고려하여 가중치 계산절차에 적용할 가구의 무응답 변수 도출방법을 결정하고, 가구수 추정을 자료의 출처별로 비교하여 행정센서스의 벤치마킹자료의 조정여부에 판단할 자료를 제공하고자한다. 이를 위해 시험조사 자료를 분석하였다.

시험조사는 5개 읍면동의 전체 1,063 조사구중 220개 조사구를 표본으로 선정하여 7,380가구를 대상으로 수행되었다. 해당 자료는 가구명부, 가구부분, 인구

부분, 주택부분 등으로 구성되어 있다. 원래 선정된 표본과 시험조사 간 차이가 나는 가구수는 309(4.2%)개인데, 이는 시험조사 과정에서 가구가 추가로 작성되어 조사 되었기 때문이다. 면 지역은 그 차이가 근소하고(1%미만), 동/읍 지역은 3%이상의 차이가 나는 것으로 나타났다. 이는 면 지역 조사와 동/읍 지역 조사의 시점이 다른 관계로 나타나는 현상이다.

거처종류를 무응답으로 처리한 가구는 551가구로 전체의 7.2%를 차지하는데, 거처종류는 가구 혹은 주택의 승수산정에 층화 변수로 사용이 가능한 변수이므로, 무응답부분은 반드시 대체값(imputation)으로 처리해야 층화 변수로 이용이 가능하다. 조사항목의 무응답의 형태를 분석하여 조사가구의 응답상태를 도출할 수 있었는데, 빈집여부 및 조사상태의 보고를 이용하여 결정 하도록 정리하였다.

시험조사를 시행한 읍면동을 등록센서스의 가구수와 표본으로 추정된 가구수와 2010년 인구총조사의 가구수와 비교하였다. 용운동, 면목동, 금왕읍은 표본추정값에 근접하고, 청하면은 등록센서스 값에 근접하며, 오천면은 등록센서스와 표본추정 중간값에 접근하는 것을 발견할 수 있었다. 용운동을 제외하고, 등록센서스가 표본추정 가구수보다 크게 집계 되었음이 자료분석 결과 드러났다. 일반적으로 면지역에서는 등록센서스가 표본추정보다 더 많이 집계되었다. 이는 주민등록은 있고 실체는 거주하지 않은 인구가 면지역에 더 많기 때문으로 판단된다. 동지역은 면지역과 반대 현상으로 표본추정이 더 많이 집계되었다. 이는 주민등록은 없고 실체는 거주하는 인구가 동지역에 많기 때문으로 추정된다.

등록센서스의 일부 가구는 부적격 가구가 있을 것이라는 가정 하에, 표본으로 추정된 적격가구 비율을 이용하여 등록센서스의 적격 가구수를 추정해 보면, 오천면은 여전히 등록센서스와 표본추정 간 큰 차이가 있고, 청하면은 부적격가구의 비율이 높아 표본추정 가구 수가 조정된 등록센서스 가구 수보다 큰 것으로 나타났다. 면목동과 금왕읍은 표본추정치 등록센서스 및 인구총조사의 추정가구 수가 유사한 것으로 집계되었다.

표본틀 작성을 위한 기초조사는 행정자료를 기반으로 시행한 조사이므로 등록센서스와 기초조사의 결과는 상관관계가 있을 것이라는 가정 하에, 적격가구 비율을 표본으로 추정함과 동시에 이를 근거로 등록센서스의 자료를 조정하여, 등록센서스 자료를 벤치마킹 자료로 이용할 수 있는지 여부를 살펴보았으나, 명확한 결론에 도달할 수 있는 근거를 시험조사 분석으로 도출하지 못하였다. 벤치마킹을 위한 등록센서스자료의 조정 여부는 좀 더 신중한 연구가 요구된다.

제3절 가중치 산정 절차

이 절의 목적은 인구주택총조사의 읍면동을 부차모집단으로 하여, 가구, 인구, 주택 통계를 반영한 가중치 산정 방법들을 검토함으로써 인구주택총조사의 추정 정확성을 높일 수 있는 방법을 제안하는 것이다.

표본추출 가중치는 표본 설계 가중치라 하는데, 이 가중치의 산정은 표본 조사구 추출 확률의 역수로 표시하기도 한다. 인구주택총조사의 표본틀은 조사구별

로 가구 수 정보가 제공되는데, 여기에서는 모집단 조사구의 가구 수와 추출된 표본 조사구의 추정 가구 수 정보를 이용하여, 표본 조사구의 추출 가중치를 조정함으로써 추정 정확도를 향상 시킬 수 있는 방안을 살펴보았다.

조사가 시행된 후에는 무응답 가구에 대한 처리가 불가피한데, 이 절에서는 조사로 도출된 가구응답 상태를 이용하여 가구 무응답에 대한 가중치를 적용해 조정을 할 수 있도록 하였다. 2010년 인구주택총조사는 가구 및 주택 분야는 전수조사로 시행되었으므로, 인구분야의 표본조사부문인 인구의 가중치는 조사된 총인구수를 벤치마크로 하여 사후층화 방법으로 승수를 산정하였다. 하지만 2015년 인구주택총조사는 가구 인구 및 주택 모두가 표본조사로 자료를 수집하는 관계로 벤치마크는 행정센서스 자료가 사용 가능하다. 이에 이 절에서는 가구, 개인, 주택으로 분리하여 각각 사후층화 방법으로 승수를 산정할 수 있는 방법을 제시하였다.

사후층화로 추정의 정확도를 향상시키기 위해서는 층화변수의 선택과 더불어 층화변수의 카테고리 결정도 가중치 산정의 중요한 사항이다. 층화변수는 추정에 중요한 역할을 할 수 있는 변수를 선택하여야 하고, 변수의 카테고리는 적당한 크기의 표본단위가 할당할 정도의 카테고리를 결정하여야 한다.

인구는 성별 및 5세 연령그룹, 가구는 가구원수, 주택은 거처의 종류를 각각 사후 층화변수로 제안하였다. 표본으로 조사된 층화변수의 카테고리과 등록센서스의 층화변수의 카테고리가 일치하지 않는 경우는 두 자료 간 카테고리의 매칭을 위하여 카테고리를 통합하는 방식으로 진행하였다.

인구와 가구의 추정 정확도를 향상시킬 수 있는 한 방법으로는, 동일가구에 속하는 모든 가구원들은 가구단위의 추정가중치와 인구단위의 추정가중치를 동일하게 부여하는 방법인, 가구·인구 통합 가중치방법(Lemaitre and Dufour: Integrated person-household weighting method)을 논의하였다. 인구의 사후 층화방법과 통합 가중치 방법의 비교를 위해 모의실험을 통한 각 방법의 추정 결과를 비교 검토하였다.

모의실험방법은 2010년 인구주택총조사 충청북도 원시자료를 모집단으로 하고 정의된 모집단으로부터 추정 항목들의 모수를 계산하고, 정의된 모집단에서 20% 조사구를 30회에 걸친 표본 조사구 무작위 추출에 의해 진행하였으며, 각 표본으로부터 사후층화 및 통합가중치 방법의 추정치를 모집단 모수와 비교하여 두 방법들 간 효율성을 비교하였다.

모의실험 추정항목으로는 전수조사 항목에 포함된 교육수준, 혼인상태, 가구주와의 관계를 선정하였고, 표본조사 항목으로는 사례수가 충분한 취업자, 실업자, 아동보육(부모), 활동제약(없음), 고령자생활비원천(부부직업) 등을 선정하였다. 방법 간 효율성 비교는 전수 항목에서는 읍면동별 상대 표본오차를 표본 항목에서는 시군구별 상대 표본오차를 계산하여 비교하였다.

분석결과, 전체적으로 통합가중치 방법이 사후 가중치 방법보다는 보다 효과적인 방법으로 드러났다. 그러나 분석 과정에서 가중치 분포를 살펴보면, 통합 가중치 방법으로 산출된 가중치가 일부 표본가구에1 작은 가중치가 할당되는 경우

도 나타났다. 이는 표본이론으로 설명이 불가능한 부분으로 이러한 가중치는 사전에 계산 과정에서 방지하여야 한다. 이를 방지하기 위해서는 정형화된 프로그램 도입이 전제 조건이므로 이 분야는 향후 연구 분야로 제안하였다.

본 장은 실제적인 시험조사 자료의 변수를 이용하여 제안한 사후층화 가중치 계산 절차 및 사후층화 변수 카테고리 자동통합 방법을 위해 알고리즘을 적용하여 정형화된 SAS 매크로 프로그램을 작성하였다. 이러한 정형화된 프로그램을 2015년 인구주택총조사에서 직접 이용 할 수 있도록 제시하였다.

제3장 순환표본조사 시행을 위한 표본설계 및 추정방안 연구

제1절 서론

본 장에서는 순환표본조사를 위한 표본설계 그리고 추정의 방안에 대한 내용을 다루었다. 순환표본조사를 위한 구체적인 표본설계 및 추정 방안에 대한 주요 이슈들을 정리하고 시험조사 결과에 적용하여 그 결과들을 살펴보았다. 본 연구 내용의 결과 및 필요한 추가 연구사항들 그리고 현재의 문제점 등을 정리하면 다음과 같다.

제2절 순환표본조사를 위한 표본설계

순환센서스에서는 모집단의 대표성을 유지할 수 있는 표본을 추출함과 동시에 매 공표시점 기준 실제 공표가 가능한 공표 수준의 결정을 위한 표본설계가 이루어져야 한다.

통계청에서 현재 고려하고 있는 순환센서스는 프랑스의 순환센서스와는 다른 구조를 갖기 때문에 각 시점 별 조사를 위한 표본의 구성을 위하여 확률적 표본설계와 비확률적 표본설계를 중복적으로 사용하는 것은 바람직하지 않을 것으로 판단된다.

각 시점 별 조사 대상 표본을 구성하기 위한 가능한 안으로는 읍면동 단위의 통계 생산 및 등록센서스의 보안을 목적으로 사전에 충분한 연구가 진행된 2015년 인구주택총조사의 표본조사를 위한 표본설계 방안을 이용하는 것이 바람직하다. 가령, 5년 주기로 조사가 진행될 표본을 추출하고 이를 60개월로 균등하게 분할하는 방안을 고려할 수 있을 것이다.

균등분할의 기준에 대한 연구는 추후에 추가적으로 이루어져야 하나 기본적으로 조사구 추출의 경우, 5년 동안 조사된 조사구를 추출하기 위하여 정렬된 조사구 리스트로부터 60개의 가능한 계통추출표본을 구축하고 이를 이용하는 방안을 고려할 수 있다. 이 때 균형 표본의 정의를 위하여 각 조사구의 추가적인 정보를 고려한다면 정렬기준에 이를 추가한 후 서로 겹치지 않는 계통추출을 통한 표본을 얻을 수 있을 것이다.

이에 대한 이론적 근거는 반복분산추출법 중 비복원 랜덤그룹화 방안에 있으

며 공표수준의 결정을 위해서는 각 조사시점 별 읍면동 포함 비율 등의 내용들을 함께 검토해야 한다.

현 시험조사 결과를 이용한 표본설계의 타당성 연구는 제한적일 수밖에 없으며 2015년 인구주택총조사의 표본조사 결과를 이용한 분석이 추 후 이루어져야 할 것으로 판단된다.

제3절 모수추정방안

본 절에서는 순환표본조사 결과를 이용하여 산출되는 추정량의 대응 모수에 대한 정의를 하였으며 이에 근거하여 기간추정량의 개념을 정립하였다. 기간추정량은 특정시점 기준의 추정량으로 이해할 수 없으며, 모수의 정의를 통해서 판단한다면 기간의 중간 시점에 대한 추정량으로도 언급할 수 있으나 이 역시 모순이 있는 해석이다. 따라서 공표 시 특정 기준 시점을 정의하는 것은 이론적으로 타당하지 않다.

기간추정에 맞는 조사항목을 새롭게 구성할 필요가 있으며 인플레이션과 같이 타당한 기준이 있을 경우 이에 근거하여 조사된 값의 수정이 추정보다 우선적으로 이루어져야 한다. 추정량의 계산은 다른 조사자료의 분석에서와 마찬가지로 가중치 산출을 통해 이루어지며, 이 과정은 설계가중치, 무응답 조정 그리고 벤치마킹을 통한 최종 가중치 산출의 단계를 거친다.

본 절에서는 시험조사 자료를 이용하여 각 가구원에 부여되는 설계가중치 및 사후층화를 통한 최종가중치를 산출하고 그 결과를 비교하였다. 시험조사의 결과를 볼 때, 20%의 조사구가 조사된 12개월 통합자료를 사용할 경우, 적어도 시군구 통계작성은 가능한 것으로 판단되나 시험조사를 위한 표집틀은 평균 60가구로 구성된 조사구의 리스트로 조사구 크기가 일정하지 않고 조사구 내 가구추출로 인한 가중치의 변동이 있어, 실제 30가구 기준의 조사구 추출이 이루어질 추 후 조사들에 대한 추론을 하는 것에는 한계가 있을 것으로 판단된다.

제4장 시험조사 자료의 심층 분석

제1절 서론

통계청은 2012년 8월 순환센서스 도입을 검토하기 위해 경상남도 거창군 시험조사 계획을 수립하여 2012년 10월부터 시험조사를 실시하였고, 2013년 3월부터 인천광역시 연수구의 5개동에 대해 시험조사를 실시하였다.

이 장에서는 인천광역시 연수구와 경상남도 거창군의 시험조사 자료를 심층 분석함으로써 주요 항목별 변화 추이 및 시험조사에서의 무응답 현황 등 문제점을 살펴보고자 한다. 또한 시험조사 자료를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 시군구 및 읍면동 단위의 시기별 공표 가능한 수준을 파악하고자 한다.

제2절 2010년 센서스 표본조사와 시험조사의 결과 비교

이 절에서는 2010년 인구주택총조사 10% 표본조사와 인천 연수구 및 경남 거창군의 시험조사의 주요 결과를 비교함으로써 항목별 변화 추이와 시험조사에서의 무응답 현황 등을 살펴보고자 한다.

1. 인구 부문

- 성별: 시험조사의 경우 추계인구(가중치 적용 후)에 비하여 여자의 비율이 약간 높고 남자의 비율이 약간 낮다.
- 나이: 통계청 인구총조사과의 분석에 의하면 인천광역시 연수구와 경상남도 거창군의 경우 2010년 인구주택총조사와 시험조사 자료의 5세 구간별 인구구조에 있어서 시차에 따라 고령화 되는 모습을 보이고 있다.
- 교육정도: 시지역인 연수구가 군지역인 거창군에 비해서는 '대학원 박사 이상'의 빈도수가 약간 많기는 하지만, 전체적으로 '대학원 이상' 특히 '박사 이상'의 빈도수가 상대적으로 매우 적다.
- 1년 전 거주지: '현 거주지'인 경우가 대부분을 차지하고 있으며, 인구이동 지표는 2010년 대비 감소한 것으로 나타났다.
- 통근·통학 여부: '안함'이 전체의 50% 정도를 차지하고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 있다. 1차년도 무응답은 연수구 614개, 거창군 209개로 높은 수준이다.
- 경제활동상태: '주로 일하였음'과 '일하지 않음'이 대부분을 차지하고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 있다. 1차년도 무응답은 연수구 140개, 거창군 468개로 높은 수준이다.
- 혼인상태: '미혼'과 '배우자 있음'이 대부분을 차지하고 있으며, 미혼율은 2010년 대비 약간 감소하였다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 있다. 1차년도 무응답은 연수구 1,474개, 거창군 2,528개로 매우 높은 수준이다.
- 총 출생아수: 4인 이하가 대부분을 차지하고 있으며, 평균 출생아수는 2010년 대비 약간 감소하였다.

2. 가구 부문

- 가구 구분: '가족'과 '1인 가구'가 대부분을 차지하고 있다. 연수구와 거창군 시험조사의 모두 가중치 적용 전과 후의 비율에 약간의 차이가 있다. 1인 가구의 비율은 2010년 대비 약간 높아졌다.
- 거주 기간: 시지역인 연수구의 경우 단기 거주 비율이 높고 장기 거주 비율은 낮다. 반면, 군지역인 거창군의 경우는 25년 이상 장기 거주 비율이 매우 높으며, 25년 이상 장기 거주 비율은 2010년 대비 상당히 감소하였다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이

가 있다.

- 사용 방 수: 잠을 자는 방의 수는 1~3개의 경우가 대부분을 차지하고 있으며, 2010년에 비하여 잠을 자는 방의 수는 상당히 감소하였다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 있다.
- 난방 시설: 시지역인 연수구의 경우 도시가스의 비율이 대부분을 차지하고 있는 반면, 군지역인 거창군의 경우는 기름보일러의 비율이 매우 높다. 거창군의 경우 2010년과 비교하여 난방 시설에 대한 변화가 큰 편이며, 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 있다.
- 점유 형태: 시지역인 연수구의 경우 자가의 비율이 가장 높으나 전세와 월세의 비율도 상당히 높은 편인 반면, 군지역인 거창군의 경우는 자가가 대부분을 차지하고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 있다.

제3절 시·군 단위의 시험조사 결과 분석

이 절에서는 인천광역시 연수구의 시험조사를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 시군구 단위의 시기별 공표 가능한 수준을 살펴보고자 한다. 참고로 경상남도 거창군의 경우는 20% 표본 자료에 대한 결과를 제시하였다.

< CV 기준 요약: 시·군 단위의 시험조사 결과 분석 >

구분	항목	누적				비고
		5% 표본	10% 표본	15% 표본	20% 표본	
인구	성별	○	○	○	○	
	나이	○	○	○	○	
	교육정도	○	○	○	○	'대학원 이상'으로 병합하는 경우 가능
	1년 전 거주지	○	○	○	○	'다른 지역'으로 병합하는 경우 가능
	통근·통학 여부	○	○	○	○	
	경제활동상태	○	○	○	○	'틈틈이/일시휴직'으로 병합하는 경우 가능
	직업: 종사상 지위	○	○	○	○	
	혼인상태	○	○	○	○	
가구	총 출생아수	△	○	○	○	'5명 이상'을 병합하는 경우 가능
	가구 구분	△	○	○	○	몇 개 범주를 병합하는 경우 가능
	거주 기간	○	○	○	○	'25년 이상'으로 병합하는 경우 가능
	잠을 자는 방	△	○	○	○	'4개 이상'으로 병합하는 경우 가능
	난방 시설	X	X	X	X	대폭적인 재분류시 가능
	점유 형태	△	○	○	○	'월세/사글세'로 병합하는 경우 가능

제4절 동·읍 단위의 시험조사 결과 분석

이 절에서는 인천광역시의 시험조사 대상 동 중 가장 작은 규모인 연수3동의 시험조사를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 읍면동 단위의 시기별 공표 가능한 수준을 살펴보고자 한다.

< CV 기준 요약: 동·읍 단위의 시험조사 결과 분석 >

구 분	항목	누적				비고
		5% 표본	10% 표본	15% 표본	20% 표본	
인 구	성별	○	○	○	○	
	나이	△	○	○	○	고연령 층의 일부 병합하는 경우 가능
	교육정도	△	○	○	○	'대학원 이상'으로 병합하는 경우 가능
	1년 전 거주지	△	○	○	○	'다른 지역'으로 병합하는 경우 가능
	통근·통학 여부	○	○	○	○	
	경제활동상태	△	△	△	○	'틈틈이/일시휴직'으로 병합하는 경우 가능
	직업: 종사상 지위	△	○	○	○	'무급 가족'의 경우 주의가 필요
	혼인상태	○	○	○	○	
총 출생아수	△	○	○	○	'4명 이상'을 병합하는 경우 가능	
가 구	가구 구분	△	△	△	△	몇 개 범주를 병합하는 경우 가능
	거주 기간	○	○	○	○	'20년 이상'으로 병합하는 경우 가능
	잠을 자는 방	△	△	△	○	'3개 이상'으로 병합하는 경우 가능
	난방 시설	X	X	X	X	대폭적인 재분류시 가능
	점유 형태	△	△	△	○	'월세/사글세'로 병합하는 경우 가능

제5장 한국에서의 순환센서스 기능

제1절 순환센서스의 사회적 기능

순환센서스는 각 지역의 변화를 정기적으로 측정하고 비교하는 기록으로서의 의미를 가진다. 순환센서스의 주요한 목적은 각 지역의 특성을 파악하고 비교함으로써 더 나은 공동체를 구축할 수 있는 토대를 만드는 것이라고 할 수 있다. 각 지방 정부가 정책을 수립하고 또한 시행하고 있는 정책의 효과를 평가하고 합의를 도출하는 데에 순환센서스는 보다 빠르고 정확한 자료를 통합적으로 제공하는 역할을 할 수 있다. 순환센서스를 조사하는 과정에서 시간에 따른 진출입에 의한 모집단의 변화를 파악하는 등 주기적인 조사에 지자체와의 협조가 매우 중요하기 때문에, 지자체가 정책을 평가하고 새로운 정책을 수립하는 데에 원활히 반영할 수 있도록 자료를 통합하고 가공해서 제공하는 것이 필요하다.

오늘날 세계적인 수준의 사회과학 저널들에서는 기존의 개인 위주의 서베이

자료보다는 실험의 기법이나 행정 자료를 활용한 연구가 증가하고 있다. 순환센서스는 기존의 행정 자료가 제공하던 정보의 질과 폭을 크게 넓힘으로써 학문적인 발전에 기여할 뿐만 아니라 정책적 함의가 높은 연구들을 양산해 낼 수 있을 것으로 기대된다.

제2절 American Community Survey(ACS) 적용 가능성 검토

1. ACS 조사의 특징

ACS는 2005년부터 미국 전역에 약 290만 가구를 표집하여 조사하는 것으로 시작하여, 2013년에는 약 350만 가구를 표집하는 등 지속적으로 확대되어 왔다. 응답률이 낮은 지역이나 에스키모, 인디언 등 소수 민족이 거주하는 지역들은 표집률을 다른 지역보다 높게 설정함으로써 이들의 대표성을 확보할 수 있도록 하고 있다. 한국에서도 현재 인구주택총조사 표집틀에 외국인 비율을 고려하여 취약지역에 추가 조사구를 배분하고 있는데, 다문화 가정이나 외국인 노동자 2세 조사 등을 위해 이러한 표집 방식을 적용해야 할 것이다. ACS 2015년의 설문지는 24개의 주거 관련 문항과 48개의 인구 관련 문항으로 구성되어 있다.

ACS 개인 원자료는 2000년까지 진행된 센서스와 동일한 방식으로 표집한 PUMS(Public Use Microdata Sample)을 로그인 없이 누구에게나 공개하는 방식으로 제공되고 있다. 2015년 현재에는 2000년부터 2013년까지의 마이크로데이터를 주 별로 1년 자료는 물론, 3년치와 5년치의 누적 자료까지 가구 자료와 개인 자료로 나누어서 업로드하였으며, 이 가구 자료와 개인 자료는 별도의 가구별 ID를 통해 병합하는 것도 가능하다. 지역 변수를 PUMA(Public Use Microdata Areas, 총 2,378개)로만 제공하고 각각을 인구 100,000명 이상이 되도록 하고 있다. 한국의 순환센서스도 개인과 가구 단위의 원자료를 사생활을 침해하지 않는 수준에서 공개할 지역 단위를 결정할 필요가 있다.

지역별 집계 자료는 웹사이트(Americana Fact Finder, factfinder.census.org)를 통해서 주제별, 지역별, 인종 및 민족 집단별, 산업별, 직업별로 나누어서 제공된다. 각 지역의 대표성을 확보하기 위해 1년치와 3년치의 추정치는 각각 인구 65,000명 이상과 20,000명 이상인 지역 단위에서만 공개한다. 비슷한 원칙을 한국의 순환센서스에 적용한다면 군과 구 단위에는 1년치와 3년치의 자료를, 읍과 면 단위에는 5년치의 자료를 공개하는 것과 유사할 것이다.

2. ACS 자료의 학문적, 정책적 활용 및 홍보

ACS 자료를 활용한 학문적 연구들은 크게 (1) ACS를 타 자료의 타당성 검토를 위한 준거 자료로 활용하는 경우, (2) ACS를 추정을 위한 주 자료로 사용하는 경우, (3) ACS를 지역 수준의 변수를 생성하는 데에 사용하는 경우, (4) 개인 수준의 자료를 활용하는 경우 등으로 구분된다. ACS를 보다 적극적으로 활용할수록 연구가 심화되기 때문에, (1)에서 (4)로 갈수록, 즉 분석적인 연구로 갈수록 수준이 높다. 따라서 한국의 순환센서스가 지역 집계자료나 개인 자료를 손쉽게

게 쓸 수 있도록 제공하는 것이 양질의 풍부한 분석을 활성화시키는 데에 바람직할 것으로 보인다.

ACS의 통계 자료는 특히 뉴욕타임즈(New York Times)가 지금까지 ACS를 인용한 200여 건의 기사에서 다루고 있는 것처럼 여러 언론에서 심도 있게 다루어지고 있다. 기사의 양과 질로 보았을 때 점점 그 중요성이 증가하고 있는데, ACS의 심층적인 언론보도에서 주목할 점은 다양한 주체 간의 연계활동이다. 통계청이 자료를 모으면 미네소타 대학이 자료를 가공하고, 뉴욕시립대학(CUNY)이 시각적 기법 등을 포함한 분석을 하고 뉴욕타임즈 같은 언론이 보도를 하는 것이다. 이러한 연계활동은 한국 순환센서스의 언론보도에도 활용될 수 있을 것이다.

미국의 연방 정부는 연간 총 \$40조의 예산을 ACS에 근거해 하위 기관이나 주 정부에 배분하며, 새로운 정책을 수립하거나 기존 정책을 보다 효율적, 효과적으로 집행하기 위해 ACS의 통계를 활용한다. 한국의 순환센서스의 정책 활용도를 높이기 위해서는 자료 제공을 바탕으로 다양하고 심층적인 자료 분석이 선행되어야 할 것이며, 더불어 순환센서스의 언론보도 및 정책적 활용을 추적하여 순환센서스가 사회에 주는 영향력을 누적적으로 파악되고 홍보할 수 있어야 할 것이다. 지금까지 논의된 ACS의 특성을 바탕으로 한국의 순환센서스에 적용할 만한 요소들을 정리하면 아래 표와 같다

< ACS의 특성 및 한국 순환센서스 적용 가능성 판별 조사표 >

구분	ACS 특성	정책적 적용 가능성	학술적 적용 가능성
조사 문항	현 거주지에 대한 각종 세금 및 비용	가구의 조세부담 관련 정책, 빈곤 가구에 대한 세제 혜택 등 지원 방안 마련	자산 규모 파악 및 소득을 넘어서 부의 불평등 연구
	국적 및 출신국	외국인 노동자, 다문화가정 현황 및 언어 사용 실태 파악	이주민의 정착과 2세들에 대한 연구
	결혼 상태/가족 구성원/조부모의 손자녀 돌봄	한부모가정 현황 파악 및 육아 관련 정책 수립	일-가정 양립과 육아에 대한 조부모 의존 현상
	교육 수준	학위 및 전공별 노동시장 진입 및 이직 실태 파악	지역 내, 지역 간 교육 불평등 연구
	복지 혜택 내역	복지 수요 파악 및 복지정책별 이용도 파악	빈곤 연구, 복지 정책의 효과 비교
조사 방법	매 월 대표성 있는 자료 수집	지역 단위별 주기적인 통계 생산	지역 규모에 따른 지역 단위 추정 신뢰성 연구
	가구 단위와 집단 거주, 쿼터의 구분	기숙사 등의 시설 거주, 1인 가족 등의 현황 파악	1인가구 증가요인과 결과에 대한 연구
자료 제공	American Fact Finder 웹을 통한 상세한 직업별, 산업별, 인증별, 지역별로 누적적 통계표 제공	집계자료를 제공할 수 있는 웹사이트 시스템 구축	기타 행정지역 단위 자료 혹은 기타 개인자료와의 병합을 통한 다층 분석
	개인자료 샘플 공개	개인자료 제공을 위한 심사 및 개인자료 접근 시스템 개선	전문적이고 참신한 시각에 근거한 연구 자료 제공
	심층분석의 주기적 언론 보도	통계청 수집, 미네소타대 가공, CUNY 분석, NYC 보도로 이어지는 연계활동	통계청 수집, KAIST 등에서 가공, 서울시립대 등 주요 거점 대학의 분석, 주요 일간지 보도로 이어지는 장기적 협력관계 형성

제4절 기타 조사의 활용 가능성 검토

순환센서스는 5년 주기로 총 인구의 20%만을 조사하는 대신 면접원이 방문 조사를 하며, 좀 더 많은 조사문항을 포함시켜 다양한 사회조사의 기능을 겸할 수 있다. 이에 본 절에서는 통계청이 실시하고 있는 조사들 중 ‘사회조사’와 ‘지역별 고용조사’를 주로 참조하여 가능한 문항들을 고려해보았다.

1. 사회조사 문항의 활용 가능성

현재의 사회조사는 2008년부터 매년 5개 부문으로 정착되어 실시되어 왔고, 18,000여 가구에서 만 13세 이상의 가구원을 대상으로 하기에 약 37,000여 명이 응답해 왔다. 순환센서스에 활용될만한 문항의 성격은 따라 크게 두 가지로 나눌 수 있다. 첫째는 지역별 특성을 보여줄 수 있는 문항으로, 복지, 환경, 안전, 나눔 등에 관한 문항들이다. 이들은 시군구(매년 혹은 3년)나 읍면동(5년) 등의 지역 단위별로 이들을 대표할 수 있는 지표를 생산하고 비교하는 데에 주로 활용될 수 있다. 이러한 문항을 바탕으로 지표를 산출함으로써 각 지방 정부에서의 정책을 위한 기초 자료로서의 역할을 할 수 있을 것이다. 두 번째 범주는 응답자의 태도나 상태에 관한 문항들이다. 주관적 계층 인식이나 계층 이동에 대한 전망을 묻는 문항들은 언론에 기획기사로 보도되기에 알맞은 주제이다. 예를 들어 2015년 11월 27일 한겨레 신문은 사회조사 계층이동 문항을 분석하여 “국민 10명 중 2명만 ‘계층상승 가능’”이라는 제목(부제: “2009년 36%서 올해 22%로 떨어져”)으로 1면에 기사를 게재했다. 이러한 기사는 2009년부터 격년으로 꾸준히 동일한 태도를 물었기에 가능한 기사이며, 향후 순환센서스에 통합하여 질문할 경우 그 누적적 가치가 지속될 수 있다.

2. 지역별 고용조사 문항의 활용 가능성

지역별 고용조사는 2008년부터 실시되어 시군 단위로 고용 현황과 산업별, 지역별 특성을 조사하는 것을 목표로 하는 대규모 조사이며, 연간 조사, 분기 조사를 거쳐 현재는 반기로 조사되어 실업률이나 취업률, 비경제활동인구 등의 고용지표를 산출하는 데에 활용된다. 10,000여 개의 조사구에서 총 190,000 가구를 표집하여 30개 내외의 문항들로 구성된 간단한 설문을 실시하기 때문에 비교적 작은 단위의 지역 범위인 시와 군 수준에서까지 통계를 제시할 수 있다. 순환센서스의 경우 약 두 달치 조사에 해당하는 규모의 조사이다.

3. OECD 지표 모듈 가능성

OECD에서 발표하는 지표들은 다른 국가들과 비교했을 때 한국 사회의 여러 단면들을 볼 수 있는 기회를 제공한다. 주요 결과들은 공공성 순위 꼴지, 안전의식 30위, 자살률 1위, 노인빈곤률 1위, 아동 삶의 만족도 꼴지, 스마트폰 보급량 1위, 단기 근로자 비중 OECD 1위, 노동생산성 낮음, 노동시간 2위, 1인당 독서량 꼴지, 중장년 실질 문맹률 최고, 성인 학습의지 꼴지 등으로 나타난다. 전국 수준

혹은 지역 수준에서 국제적으로 비교할 수 있는 지표를 생산하고 장기적으로 이를 개선하는 것은 정책적인 차원에서 중요할 것이며, 자살률 감소를 위한 우울감이나 자살생각 문항, 낮은 노동생산성과 학습의지 등을 개선하기 위한 직장만족도 등에 대한 조사가 가능할 것이다. 더불어 사회조사의 사회 지지망 문항, 사회참여 문항, 한국종합사회조사의 대인신뢰 및 기관에 대한 신뢰 등이 연관성이 높은 문항들이다.

4. 순환센서스 조사 모듈 설계의 예

모듈 항목들 중 고용조사는 동일한 항목을 3달 간격으로 4회까지 조사하며 두 회를 합하여 반기별(즉 6개월 단위)로 추정치를 공개한다. 사회조사 항목들은 환경, 나눔, 사회자본 및 계층의 총 세 모듈로 나누어 6개월 간격으로 총 2회 조사하여 연도별로 추정치를 제공한다. OECD 모듈은 지역단위보다 국가단위 추정치가 더 유용할 것이므로 1년에 1회만 조사하는 것으로 조사계획을 세웠다. 변동모듈은 주로 연말에 실시하여 국가적으로 혹은 지역적으로 중요한 이슈나 수요를 반영할 수 있도록 한다. 혹은 타 부처나 연구자들로부터 공모나 예산지원을 받아 문항을 구성하는 등의 방안이 가능할 것이다. 국가적인 현황파악의 경우는 향후 정규모듈에 편입시키려는 문항이나 모듈을 시범적으로 조사하는데 활용할 수 있을 것으로 보이며, 지자체의 정책 평가의 경우는 향후 순환센서스의 원활한 정착을 위해 지역행정의 협조를 얻는데 유용할 수 있다. 외부의 공모나 예산지원을 받는 경우는 학문적 발전과 센서스 예산 확보에 도움이 될 수 있을 것이다.

< 순환센서스 조사 설계의 예 >

조사가구: 매달 최대 9만 가구, 연 80만 가구, 5년 총 4백만 가구 = 모집단 20%				
조사 개 월	공통문항		모듈문항	
	내용 및 문항 수	내용	문항 수	시군구 단위 공개주기
1	*내용: 현재 인구주택총조사 기초 *문항수: 약 45	고용조사 1차	20~25	2차 조사와 통합
2		사회조사(환경) 1차		2차 조사와 통합
3		사회조사(나눔) 1차		2차 조사와 통합
4		고용조사 2차		반기별
5		사회자본, 계층 1차		2차 조사와 통합
6		OECD		연도별
7		고용조사 3차		4차 조사와 통합
8		사회조사(환경) 2차		연도별
9		사회조사(나눔) 2차		연도별
10		고용조사 4차		반기별
11		사회자본, 계층 2차		연도별
12		변동모듈		연도별

제6장 향후 순환센서스 연구를 위한 정책적 제언

제6장에서는 각 장별 연구를 통해 도출된 결과를 토대로 향후 순환센서스 연구를 위한 정책적 제언을 제시하였다. 먼저, 제2장의 ‘인구주택총조사 표본규모할당 및 추정방법’에 대한 연구를 통해 표본규모 할당방법, 추정가중치 산정을 위한 시험조사 자료분석, 그리고 가중치 산정 절차 등에 대한 과제를 제안하였다.

다음으로, 제3장의 ‘순환표본조사 시행을 위한 표본설계 및 추정방안 연구’에서는 시험조사의 목적이 불분명함을 지적함과 동시에 본 연구를 위한 다양한 분석 과정에서 발생한 문제점과 개선사항들을 제시하였다. 또한 순환표본조사를 위한 표본설계 시 유의사항과 순환표본조사 자료를 이용한 모수 추정을 위한 향후 과제를 제안하였다. 제4장에서는 ‘시험조사 자료의 심층분석’ 결과를 토대로 경상남도 거창군, 인천 연수구, 그리고 현재 제주도에서 진행 중인 시험조사의 결과분석을 어떻게 수행하여야 하는가, 그리고 시험조사 자료의 철저한 정제와 질 관리의 필요성을 언급하였다. 마지막으로 제5장에서의 ‘한국에서의 순환센서스 기능’과 관련된 연구를 통해 향후 순환센서스를 공동체 연구 활성화를 위한 기초자료로 어떻게 이용할 수 있을 것인지를 제시하였다. 특히, 지역 수준 지표 및 개인 자료 활용과 자료를 활용한 정책적·학문적 연구 활성화 가능성을 중심으로 논의하였다.

목 차

제1장 연구의 배경 및 구성	1
제1절 연구의 배경	1
제2절 연구의 구성	3
제2장 인구주택총조사 표본규모할당 및 추정방법	5
제1절 표본규모 할당방법	5
제2절 추정가중치 산정을 위한 시험조사 자료분석	16
제3절 가중치 산정 절차(Weighting steps)	25
제3장 순환표본조사 시행을 위한 표본설계 및 추정방안 연구	44
제1절 서론	44
제2절 순환표본조사를 위한 표본설계	48
제3절 모수추정방안	60
제4장 시험조사 자료의 심층 분석	95
제1절 서론	95
제2절 2010년 센서스 표본조사와 시험조사의 결과 비교	98
제3절 시·군 단위의 시험조사 결과 분석	106
제4절 동·읍 단위의 시험조사 결과 분석	117
제5장 한국에서의 순환센서스 기능	130
제1절 순환센서스의 사회적 기능	130
제2절 American Community Survey(ACS) 적용 가능성 검토	131
제3절 프랑스 롤링 센서스	147
제4절 기타 조사의 활용 가능성 검토	147
제6장 향후 순환센서스 연구를 위한 제언	163
제1절 인구주택총조사 표본규모할당 및 추정방법	163
제2절 순환표본조사 시행을 위한 표본설계 및 추정방안 연구	165
제3절 시험조사 자료의 심층 분석	167
제4절 한국에서의 순환센서스 기능: 공동체 연구 활성화를 위한 순환센서스 활용 ..	168

[표 차례]

<표 2-1> 2015년 조사구 종류 및 특성별 조사구수(잠정)	5
<표 2-2> 시도 및 조사구 규모별 읍면동수 및 조사구수	6
<표 2-3> 주요 통계생산에 필요한 항목 및 분류 (2010 조사기준)	7
<표 2-4> 표본규모 배분법에 사용한 공식들	10
<표 2-5> 가상적인 모집단 가구수, 4인가구원 특성치의 가구수 및 요구된 20% 표본 가구수	12
<표 2-6> Equal allocation	13
<표 2-7> Proportional allocation	13
<표 2-8> Square Root allocation	13
<표 2-9> Kish allocation	14
<표 2-10> 읍면동별, 표본할당방법별 상대표본오차 요약(단위:%)	14
<표 2-11> 시험조사 읍면동 현황	17
<표 2-12> 조사구 특성별 추출 승수	17
<표 2-13> 기초조사 및 시험조사의 가구수	18
<표 2-14> 거처의 종류별, 빈집 여부별, 조사상태별 및 가구원수별 가구수 ...	19
<표 2-15> 응답상태 분류별 조사된 표본가구수	20
<표 2-16> 응답상태 분류별 추정 모집단가구수	20
<표 2-17> 거처종류의 재분류	21
<표 2-18> 행정구역 및 거처종류별 추정가구수	22
<표 2-19> 출처별 가구수	23
<표 2-20> 등록센서스의 적격 가구수의 추정 요약	24
<표 2-21> 등록센서스의 적격 가구수의 거처종류별 추정	25
<표 2-22> 거처의 종류 카테고리 통합	29
<표 2-23> 가구규모의 카테고리 통합	30
<표 2-24> 연령 그룹의 카테고리 통합	30
<표 2-25> 교육정도의 방법별 결과비교	38
<표 2-26> 혼인상태의 방법별 결과비교	38
<표 2-27> 가구주관계의 방법별 결과비교	39
<표 2-28> 표본항목의 방법별 결과비교	40
<표 3-1> 2000-2020년 센서스 라운드의 국가 별 센서스 방안	45
<표 3-2> 조사구 규모별 읍면동 분포	53
<표 3-3> 시도별 표본규모	54
<표 3-4> 내재적 층화를 위한 정렬 변수	55
<표 3-5> 2015년 읍면동의 표본 조사구 수 분포	57
<표 3-6> 2015년 인구주택총조사 조사구 표본의 17개 광역시도별 분포	58
<표 3-7> 시범지역 조사구 모집단 분포	61
<표 3-8> 시범지역 표본 조사구 및 가구 분포	61

<표 3-9> 시범지역 설계가중치 분포	68
<표 3-10> 벤치마킹을 위한 모집단 정보: 인천시 연수구 남성	75
<표 3-11> 벤치마킹을 위한 모집단 정보: 인천시 연수구 여성	76
<표 3-12> 벤치마킹을 위한 모집단 정보: 거창군 남성	77
<표 3-13> 벤치마킹을 위한 모집단 정보: 거창군 여성	78
<표 3-14> 최종가중치 분포	80
<표 4-1> 경상남도 거창군 시험조사의 표본 가구수 및 인구수	96
<표 4-2> 인천광역시 연수구 시험조사의 표본 가구수 및 인구수	98
<표 4-3> 표본조사 분포 - 성별	98
<표 4-4> 표본조사 분포 - 나이	99
<표 4-5> 표본조사 분포 - 교육정도	100
<표 4-6> 표본조사 분포 - 1년 전 거주지	100
<표 4-7> 표본조사 분포 - 통근·통학 여부	101
<표 4-8> 표본조사 분포 - 경제활동상태	101
<표 4-9> 표본조사 분포 - 혼인상태	102
<표 4-10> 표본조사 분포 - 총 출생아수: 남자	102
<표 4-11> 표본조사 분포 - 총 출생아수: 여자	103
<표 4-12> 표본조사 분포 - 가구 구분	103
<표 4-13> 표본조사 분포 - 거주 기간	104
<표 4-14> 표본조사 분포 - 사용 방 수: 잠을 자는 방	104
<표 4-15> 표본조사 분포 - 난방 시설	105
<표 4-16> 표본조사 분포 - 점유 형태	105
<표 4-17> 시험조사 결과 - 성별	106
<표 4-18> 시험조사 결과 - 나이	107
<표 4-19> 시험조사 결과 - 교육정도	108
<표 4-20> 시험조사 결과 - 1년 전 거주지	109
<표 4-21> 시험조사 결과 - 통근·통학 여부	109
<표 4-22> 시험조사 결과 - 경제활동상태	110
<표 4-23> 시험조사 결과 - 종사상 지위	110
<표 4-24> 시험조사 결과 - 혼인상태	111
<표 4-25> 시험조사 결과 - 총 출생아수: 남자	112
<표 4-26> 시험조사 결과 - 총 출생아수: 여자	113
<표 4-27> 시험조사 결과 - 가구 구분	114
<표 4-28> 시험조사 결과 - 거주 기간	114
<표 4-29> 시험조사 결과 - 사용 방 수: 잠을 자는 방	115
<표 4-30> 시험조사 결과 - 난방 시설	116
<표 4-31> 시험조사 결과 - 점유 형태	117
<표 4-32> 연수구 연수3동 시험조사의 표본 가구수 및 인구수	117
<표 4-33> 거창군 거창읍 시험조사의 표본 가구수 및 인구수	118
<표 4-34> 시험조사 결과 - 성별	119

<표 4-35> 시험조사 결과 - 나이	119
<표 4-36> 시험조사 결과 - 교육정도	120
<표 4-37> 시험조사 결과 - 1년 전 거주지	121
<표 4-38> 시험조사 결과 - 통근·통학 여부	122
<표 4-39> 시험조사 결과 - 경제활동상태	122
<표 4-40> 시험조사 결과 - 종사상 지위	123
<표 4-41> 시험조사 결과 - 혼인상태	123
<표 4-42> 시험조사 결과 - 총 출생아수: 남자	124
<표 4-43> 시험조사 결과 - 총 출생아수: 여자	125
<표 4-44> 시험조사 결과 - 가구 구분	126
<표 4-45> 시험조사 결과 - 거주 기간	126
<표 4-46> 시험조사 결과 - 사용 방 수: 잠을 자는 방	127
<표 4-47> 시험조사 결과 - 난방 시설	128
<표 4-48> 시험조사 결과 - 점유 형태	129
<표 5-1> 월별 ACS 자료 수집 과정	132
<표 5-2> 2015년 ACS 설문지	133
<표 5-3> 2015년 ACS 설문지	134
<표 5-4> ACS를 활용한 기술적 추정 연구의 예	137
<표 5-5> ACS를 활용한 분석적 연구의 예	138
<표 5-6> 연방 정부와 주정부 및 하위기관의 ACS 정책적 활용 사례	145
<표 5-7> ACS의 특성 및 한국 순환센서스 적용 가능성 판별 조사표	146
<표 5-8> 순환센서스 조사 설계의 예	161

[그림 차례]

[그림 2-1] 할당방법 및 지역층별 상대표본오차의 비교	15
[그림 2-2] 가상적인 모집단 분석	16
[그림 2-3] 교육정도의 카테고리별 RSE(%)의 분포	38
[그림 2-4] 혼인상태의 카테고리별 RSE(%)의 분포	39
[그림 2-5] 가구주관계의 카테고리별 RSE(%)의 분포	39
[그림 2-6] 표본항목의 카테고리별 RSE(%)의 분포	40
[그림 2-7] Module processing flow	43
[그림 3-1] 인천 3개월 그룹 설계가중치	69
[그림 3-2] 인천 6개월 그룹 설계가중치	69
[그림 3-3] 인천 12개월 그룹 설계가중치	70
[그림 3-4] 인천 18개월 그룹 설계가중치	70
[그림 3-5] 인천 18개월 그룹 설계가중치	71
[그림 3-6] 거창 3개월 그룹 설계가중치	71
[그림 3-7] 거창 6개월 그룹 설계가중치	72
[그림 3-8] 거창 12개월 그룹 설계가중치	72
[그림 3-9] 거창 18개월 그룹 설계가중치	73
[그림 3-10] 거창 24개월 그룹 설계가중치	73
[그림 3-11] 조사시점 별 모집단 사후층 비율 차이 분포	79
[그림 3-12] 인천 3개월 그룹 최종 가중치	81
[그림 3-13] 인천 6개월 그룹 최종 가중치	81
[그림 3-14] 인천 12개월 그룹 최종 가중치	82
[그림 3-15] 인천 18개월 그룹 최종 가중치	82
[그림 3-16] 인천 24개월 그룹 최종 가중치	83
[그림 3-17] 거창 3개월 그룹 최종 가중치	83
[그림 3-18] 거창 6개월 그룹 최종 가중치	84
[그림 3-19] 거창 12개월 그룹 최종 가중치	84
[그림 3-20] 거창 18개월 그룹 최종 가중치	85
[그림 3-21] 거창 24개월 그룹 최종 가중치	85
[그림 3-22] 인천 3개월 및 6개월 최종가중치 대 설계가중치 비율	87
[그림 3-23] 인천 12개월 최종가중치 대 설계가중치 비율	88
[그림 3-24] 인천 18개월 및 24개월 최종가중치 대 설계가중치 비율	89
[그림 3-25] 가창 3개월 및 6개월 최종가중치 대 설계가중치 비율	90
[그림 3-26] 가창 12개월 최종가중치 대 설계가중치 비율	91
[그림 3-27] 가창 18개월 및 24개월 최종가중치 대 설계가중치 비율	92
[그림 5-1] ACS 개인자료 공개	135
[그림 5-2] ACS 집계자료 공개	136
[그림 5-3] New York 시의 민족별 인구 변화 추이(Ford Fessenden & Sam Roberts,	

2011)	142
[그림 5-4] Virginia의 민주당 성향 유권자들의 투표에 대한 분석(Cohn, 2014)	143
[그림 5-5] 남성 비노동자의 분포(Aisch, Katz, & Leonhardt, 2014)	143
[그림 5-6][예시 문항] 사회 조사 2015년 향후 늘려야 할 공공시설과 복지서비스	148
[그림 5-7][예시 문항] 사회 조사 2015년 현재 체감 환경	149
[그림 5-8][예시 문항] 사회 조사 2015년 기부 여부	150
[그림 5-9][예시 문항] 사회 조사 2015년 기부 내용	150
[그림 5-10][예시 문항] 사회 조사 2015년 기부의 정기성	151
[그림 5-11][예시 문항] 사회 조사 2015년 자원봉사 활동	151
[그림 5-12][예시 문항] 사회 조사 2015년 계층 의식 및 전망	152
[그림 5-13] 한겨레 기사에 실린 계층이동에 대한 태도 그래프	153
[그림 5-14][예시 문항] 사회 조사 2015년 여성 취업	154
[그림 5-15][예시 문항] 사회 조사 2015년 일-가정 양립에 대한 인지	154
[그림 5-16][예시 문항] 사회 조사 2015년 노인 생활비 수급원	155
[그림 5-17][예시 문항] 사회 조사 2015년 대학생 등록금 수급원	155
[그림 5-18][예시 문항] 지역별 고용조사 2015년 상반기 근로 시간 및 사업체 정보	156
[그림 5-19][예시 문항] 지역별 고용조사 2015년 상반기 근로 시간 및 사업체 정보	157
[그림 5-20][예시 문항] 지역별 고용조사 2015년 상반기 구직자의 구직 경로와 퇴직자의 퇴직 사유	157
[그림 5-21][예시 문항] 사회전반적 신뢰(General Trust): KGSS의 예	158
[그림 5-22][예시 문항] 제도에 대한 신뢰 (Institutional confidence): OECD는 군대, 사법부, 중앙정부에 대한 신뢰를 주로 고려, KGSS 문항의 예	159
[그림 5-23][예시 문항] OECD의 사회적 보호와 웰빙에 관한 비교 지표	159
[그림 5-24] 사회전반적 신뢰(General Trust): 사회조사의 응용	160

제1장 연구의 배경 및 구성

제1절 연구의 배경

인구주택총조사는 1925년 처음 도입된 이후로 매 5년마다 실시되는 조사로, 인구·가구·주택부문에 있어서 국가통계의 기반이 되는 매우 중요한 조사이다. 즉, 영토 내의 사람과 거쳐 전체(완전성)를 대상으로, 일시(동시성)에, 각각 개별적(개별성)으로, 일정한 주기(주기성)를 갖고 실시하는 국가기본 통계조사이다. 또한, 인구주택총조사는 국가기관, 개인기업, 학술단체 및 일반 대중의 정책입안, 계획수립, 연구 및 평가 등을 포함한 각종 분야의 기초자료로 활용되고 있으며, 모든 사회분야 통계의 기준(benchmark) 통계로서의 역할을 하고 있으며, 2015년 3월 현재, 가구대상 조사통계 143종 중 58종(40.6%)이 인구주택총조사 자료를 표본틀로 활용하고 있다(통계청, 2015).

인구주택총조사는 2010년까지 전수조사(short-form)와 표본조사(long-form)라는 두 개의 방식으로 진행되어 왔다. 전수조사는 전체 가구와 가구원을 대상으로 기본적인 항목을 조사한다. 반면 표본으로 선정된 가구와 가구원은 상대적으로 길고 다양한 항목의 조사에 응답하게 되는데, 이를 표본조사로 정의한다(박시내·정남수, 2015).

그러나 국가기본 통계조사로서 중요한 함의를 갖고 있는 인구주택총조사는 국내의 조사환경이 변화됨에 따라 몇 가지 도전에 직면해 있으며, 이에 대한 대응방안의 모색을 요구받고 있다. 먼저 위협요인이 되는 환경변화로는, 1인 가구 및 맞벌이 가구의 급격한 증가로 면접원의 야간 방문이 불가피하게 되었고, 인구고령화, 핵가족화 영향으로 고령자 부부가구 및 혼자 사는 고령자 가구가 증가하여 가구의 응답능력 약화되고 있음을 들 수 있다. 뿐만 아니라 출입을 통제해 조사원 접근이 곤란한 아파트, 다가구주택이 증가하는 추세이며, 사생활보호 의식이 확산되어 외부인의 가구방문을 거부하거나 개인과 가족의 정보 노출을 거부하는 경향이 증대되고 있다. 이로 인해 응답을 얻기가 날로 힘들어져 가고 있으며, 그 비용 또한 빠르게 가중되고 있다. 총조사 비용은 매 주기별로 약 50% 정도씩 급격히 증가[834억원('00년)→1,290억원('05년)→1,808억원('10년)]하여 국회, 예산 당국 등 정책 입안자들이 지속적으로 절감할 것을 요구해왔다.

반면, 기회요인이 되는 환경변화로는 주민등록부, 건축물대장, 가족관계등록부 등 각 부처의 행정자료가 전산화되고 정비되어 총조사에 활용 가능한 자료가 다양해지고 풍성해졌다는 사실을 들 수 있다. 또한, 광통신 및 인터넷 등 정보통신기술(ICT)의 발달로 인터넷조사의 내실화와 조사 관리의 시스템화가 용이하게 되었고,

ICR(Intelligent Character Recognition)의 문자 인식기술 향상으로 종이조사표의 효과적인 자료처리가 가능하게 되었다(통계청, 2015).

기존 인구주택총조사를 둘러싼 위의 위협요인과 기회요인에 대응하기 위해 통계청은 인구주택총조사 패러다임의 전환을 모색해왔다. 먼저, 5년마다 대규모로 진행되던 인구주택총조사는 2016년부터 전수 부문은 등록센서스 방식을 도입하여 매년 인구, 가구 및 주택에 대한 기본 사항에 대한 공표가 가능하도록 하였다. 더 나아가 통계청은 등록센서스에서 공표되지 않은 항목에 대해서는 2015년 20% 표본조사를 실시하였으며, 2021년부터는 표본조사를 5년에 한번 실시하는 것이 아닌, 매일 표본을 균등 분할하여 실시하는 순환센서스 방식 도입을 검토 중이다. 순환센서스는 전통적인 센서스방식의 대안으로서, 5년 주기로 일시에 실시되는 기존의 센서스 방식과는 달리 비반복적인 표본에 대해 매일 정기적으로 조사한 결과를 누적시켜 통계를 작성하는 방식이다. 이 방식은 통계의 시의성 및 정확성을 향상시키고, 인구·사회 관련 소지역 통계 수요에 부응하며, 일정 시점에 요구되는 비용과 인력의 부담을 완화할 수 있는 장점이 있다(박시내·정남수, 2015).

그러나 순환센서스 방식으로 이행하기 위해서는 복잡한 표집 및 모형 기술, 우량의 표본틀이 필수적이며, 센서스에서 보여줬던 전체 모집단에 대한 일정 시점의 스냅샷이 더 이상 제공되지 않는다는 점이 고려되어야 한다. 이를 위해 통계청은 2012년부터 순환센서스 도입을 위한 다양한 연구를 수행하여 왔다. 2012년 통계청의 인구총조사과와 동향분석실은 공동으로 “인구센서스표본조사방법연구회”를 조직하여 미국의 센서스와 지역사회조사(ACS)에 대한 연구를 시작하였다. 한편 통계청의 인구총조사과는 표본조사의 순환센서스도입을 위한 사전검토를 위한 시험조사 실시 계획안을 수립하고, 2010년 표본조사항목과 동일하게 순환센서스 시험조사를 2013년 1월~2014년 12월에 걸쳐 경남거창군과 인천 연수구에서 실시한 바 있다(박시내·정남수·전진혜·정진희, 2015). 그리고 제주특별자치도를 대상으로 2015년 10월부터 2017년 9월(24개월)까지 총 1,584조사구, 47,520가구를 대상으로 제2차 시험조사를 실시 중에 있다. 이와 함께 통계청은 순환센서스와 관련된 다양한 연구를 수행해왔다. 즉, ‘순환센서스 도입 검토를 위한 시험조사 자료 분석(2013)’, ‘2010년 센서스 자료를 활용한 순환센서스 다년 추정치 검토(2013)’, ‘순환센서스 적용을 위한 표본틀 작성 및 갱신 방안(2014)’, ‘거창군 및 인천시 순환센서스 시험조사 자료 분석(2014)’ 등의 연구를 수행해 왔다. 이러한 연구들은 주로 통계개발원을 중심으로 수행되어 왔는데, 통계청은 객관적이고 중립적인 외부 전문가들의 의견을 수렴하기 위해 한국인구학회에 의뢰하여 2014년 5월부터 2014년 12월까지 약 8개월 간에 걸쳐 ‘순환센서스 추정 및 분석 방안에 대한 연구’를 실시하기도 하였다. 한국인구학회의 2014년도 연구는 순환센서스의 도입을 위한 기본 방향과 기본 틀을 설정하는데 일정 부분 기여했지만, 순환센서스와 기존 센서스 및 등록센서스와의 관계, 순환센서스 디자인, 모수추정 방법론, 순환센서스 활용방안 등에 대한 기초 연구에 그쳤다는 한계를 남겼다. 이에 본 연구는 2014년의 연구 성과를 토대로 보다 심도 있는

연구를 수행하는데 목적을 두고 있다. 특히, 순환센서스를 도입하기 위해서는 모집단 구축과 표본설계, 모수 작성과 추정, 결과 분석 등을 중점적으로 검토할 필요가 있다. 아울러 순환센서스를 정책적 및 학술적 요구에 부합하게 효과적으로 활용할 수 있는 방안을 고민할 필요가 있다. 따라서 본 연구에서는 2014년 연구 등 기존 연구 결과를 참고하여 실제적인 표본설계와 추정 및 결과 분석, 그리고 활용방안에 대한 연구를 수행하고자 하였다.

제2절 연구의 구성

본 연구는 ‘인구주택총조사 표본규모할당 및 추정방법’, ‘순환표본조사 시행을 위한 표본설계 및 추정방안 연구’, ‘시험조사 자료의 심층 분석’, 그리고 ‘한국에서의 순환센서스 기능과 활용’ 등 크게 4개 부문으로 구성되어 있으며, 각 부문별 주요 연구내용은 다음과 같다.

제2장의 ‘인구주택총조사 표본규모할당 및 추정방법’은 크게 3개의 연구내용으로 구성되어 있다. 먼저, 제1절에서는 표본규모 할당방법에 대한 연구를 수행하였는데, 조사방법의 변화에 부응하기 위하여 정도 높은 추정을 위한 표본규모 할당방법들을 검토하여 바람직한 표본설계 기본방향을 제시한다. 제2절에서는 추정가중치 산정을 위한 시험조사 자료를 분석하였다. 시험조사의 결과를 분석하고자 하는 목적은 항목별 응답상태 및 응답특성을 고려하여 추정에 필요한 가중치 계산절차에 필요한 가구의 무응답 변수 도출방법을 결정하고, 가구수 추정을 자료의 출처별로 비교하여 벤치마킹자료의 조정여부에 판단할 자료를 제공하고 가중치산정의 처리절차를 제공한다. 제3절은 가중치 산정 절차를 제시하였다. 인구주택총조사의 승수 산정 방법은 읍면동을 부차모집단(sub-population)으로하고, 부차모집단을 조사구 특성에 따라 층화(아파트조사구 및 일반조사구)하여, 가구인구 및 주택부문으로 나누고, 가중치산정 방법들을 검토하여 새로운 조사방법을 도입하는 인구주택총조사의 추정에 정도를 높일 수 있는 방법을 제안한다.

제3장은 ‘순환표본조사 시행을 위한 표본설계 및 추정방안 연구’로서, 제1절에서는 연구의 전체적인 연구내용을 개괄하고 있다. 그리고 제2절에서는 순환표본조사를 위한 표본설계 방안을 제시하였다. 이를 위해 프랑스와 미국에서 사용되고 있는 표본설계 방안의 특징을 살펴보고 이의 적용 가능성을 검토한다. 그리고 기존 인구주택총조사의 표본설계 틀을 유지하면서 적절한 수준에서의 통계를 매 해 생산할 수 있는 방안을 고려한다. 제3절에서는 모수추정방안을 제시하였다. 먼저, 기간추정(Multi-year estimation or estimation for period)에 대한 정의하고, 정의된 기간추정의 개념 하에서 모수 추정을 위한 가중치 작성에 대한 내용을 ACS방안을 기초로 하여 우선적으로 다루고 이어서 추가적으로 추정 시 필요한 내용들을 검토한다. 제

4절에서는 기간 추정을 위한 기타 고려사항을 제안하였다. 기간추정을 위해서 고려한 가중치 작성 방안 이외에 가중치와 무관하게 고려해야 하는 사항들로 변수들의 표준화, 지역 혹은 구역 변화에 따른 지역 코드 갱신 그리고 인플레이션과 같은 경기 변동으로 인한 원화가치의 변화의 반영 등을 들 수 있는데, 이와 관련된 내용들을 간략히 제시하고자 한다.

제4장은 ‘시험조사 자료의 심층 분석’을 실시하였다. 인천광역시 연수구 및 경상남도 거창군의 시험조사 자료를 심층 분석함으로써 주요 항목별 변화 추이 및 시험조사에서의 무응답 현황 등 문제점을 살펴보고자 한다. 이를 위해 2010년 인구주택총조사 10% 표본조사와 인천 연수구 및 경남 거창군의 시험조사의 주요 결과를 비교함으로써 항목별 변화 추이 및 시험조사에서의 무응답 현황 등을 살펴보고자 한다. 그리고 인천광역시 연수구의 시험조사를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 시군구 단위의 시기별 공표 가능한 수준을 살펴보고자 한다. 참고로 경상남도 거창군의 경우는 20% 표본 자료에 대한 결과를 제시하였다.

제5장은 순환센서스의 인구사회학적, 정책적, 학문적 기능을 어떻게 극대화시킬 수 있는지 모색하였다. 이를 위해 우선 이러한 기능을 성공적으로 수행하고 있는 미국의 순환센서스, ACS부터 검토하였다. 다만, 지금까지의 연구에서 ACS의 설계 및 조사 특징들은 다루어온만큼, 본 연구에서는 ACS의 특징은 개괄적으로 살피고, 실제로 어떻게 자료가 제공되고 학문적, 정책적 활용이 어떻게 이루어지고 있는지를 좀 더 자세히 살펴보았다. 이와 더불어 다양한 사회조사를 순환센서스에 통합시킬 경우 어떻게 통합조사할 수 있는지에 대한 예시를 제시하였다. 이를 위해 통계청의 ‘사회조사’와 ‘지역별 고용조사’, 그리고 기타 조사로는 ‘한국종합사회조사’ 및 OECD의 각종 사회지표를 고려하였다.

제6장은 ‘향후 순환센서스 연구를 위한 정책적 제언’에 해당되며, 각 장별 연구를 통해 도출된 함의들을 정리하고 2016년 이후의 순환센서스 연구를 위한 정책적 제언과 향후 과제들을 제시하였다.

제2장 인구주택총조사 표본규모할당 및 추정방법

제1절 표본규모 할당방법

1. 연구목적

과거 인구주택총조사는 조사방법이 전수와 표본을 혼합했었는데, 2015년 인구주택총조사는 일부 조사항목은 행정센서스 자료를 이용하고 나머지 조사항목은 표본 조사를 하는 혼합형 조사방법으로 전환된다. 이 장은 정확도 높은 추정을 위한 표본규모 할당방법들을 검토하여 표본설계 기본방향을 제시하고자 하는데 그 목적이 있다.

2. 기본사항

가. 조사구정의

조사구(ED: EnumerationDistrict)는 평균 30가구를 기준으로 작성된 표본추출 단위로 정의된다. 표본추정의 효율성을 극대화하기 위하여 과거 인구주택 총 조사 기준인 평균 60가구에서 조사구 설정 기준 가구 수를 평균 30가구로 조정하였다.

나. 조사구 종류

조사구는 크게 일반조사구와 특별조사구로 구분하고, 일반조사구는 주택의 종류에 따라서 보통조사구와 아파트조사구로 나누었다. 조사구 특성별 조사구 분포를 살펴보면 총 681,982 조사구 중 아파트조사구는 52%, 보통조사구는 46%, 특별조사구는 2%로 구성되어 있다. 조사방법이 일반조사구에는 표본조사를, 특별조사구에는 전수조사를 적용하므로 본 장에서는 일반조사구에 해당하는 아파트조사구와 보통조사구에 표본설계에 대하여 검토할 예정이다.

<표 2-1> 2015년 조사구 종류 및 특성별 조사구수(잠정)

조사구 종류	조사구 특성	조사구수	구성비(%)	전수.표본
일반조사구	보통조사구	354,418	46	표본조사
	아파트조사구	313,502	52	
	소계	667,920	98	
특별조사구	기숙.사회시설조사구	14,062	2	전수조사
	총계	681,982	100	

일반조사구에서 시도 및 조사구 규모별 읍면동수의 분포를 살펴보면, 전국은 3,488개의 읍면동 행정구역으로 분할되어 있다. 조사구가 1-20개로 구성된 읍면동은 56개, 21-50개의 조사구로 구성된 읍면동은 661개, 51-100개의 조사구로 구성된 읍면동은 740개, 101-200개의 조사구로 구성된 읍면동은 740개, 201-300개의 조사구로 구성된 읍면동은 613개이며 300개 이상으로 구성된 읍면동은 769개로 다양한 조사구 크기의 읍면동으로 분포되어 있다.

<표 2-2> 시도 및 조사구 규모별 읍면동수 및 조사구수

시도	일반 조사구수	읍면동수						
		총계	조사구 규모					
			20.1월	21-50	51-100	101-200	201-300	300+
전국	667,920	3,488	56	661	649	740	613	769
11	128,703	423	1	4	11	60	148	199
21	46,344	210	2	5	22	89	44	48
22	32,056	139	-	1	13	56	37	32
23	36,109	147	2	11	21	18	46	49
24	19,090	95	-	5	19	34	21	16
25	20,463	78	-	1	5	21	27	24
26	14,687	56	-	2	2	19	14	19
29	2,402	12	-	2	3	5	-	2
31	150,312	552	10	24	62	113	132	211
32	21,717	188	8	44	65	41	18	12
33	21,964	153	4	49	34	27	19	20
34	29,296	207	-	56	68	34	23	26
35	26,137	241	7	104	56	34	21	19
36	26,792	297	10	124	96	38	15	14
37	39,261	332	6	123	92	56	19	36
38	45,056	315	6	98	74	79	21	37
39	7,531	43	-	8	6	16	8	5

다. 표본규모정의

인구주택 총조사의 20% 표본 추출은 일반조사구의 20% 표본이며(특별조사구는 전수조사이므로 20%의 표본규모에서 제외됨), 시도, 시군구 및 읍면동의 표본규모 할당의 검토를 위한 표본추출의 모집단 조사구에 표본들은 667,920조사구이다.

라. 표본설계 기본방향

표본설계자는 모집단의 특성, 표본들의 구성여부, 조사예산 및 추정의 정확도를 고려하여 표본설계 방법 및 표본규모를 결정하는 것이 일반적이다. 인구주택총조사는 전체 표본조사구가 사전에 설정되어 있으므로, 부차모집단으로 정의한 시도/시군구/읍면동으로 표본을 가장 합리적으로 할당하여 부차모집단 추정치의 정확도를 향상시킬 수 있는 방법을 모색해야 한다.

마. 표본설계에 제안된 고려사항

표본설계에 요구되는 기본적인 사항은 다음과 같다.

- 표본조사는 전체 일반 조사구 중 20% 조사구를 대상으로 한다.
- 표본조사가 생산하는 주요 통계치는 읍면동 부차모집단의 추정의 정확도를 고려한 설계를 요구한다. 읍면동별 높은 정확도가 요구되는 주요 통계치는 읍면동별 교육정도별 인구수, 혼인상태별 인구수, 가구원수별 가구수 및 거주종류별 가구수 등이다. 과거 인구주택총조사가 공표한 통계치는 시군구를 기준으로 하였는데, 2015년 조사는 읍면동 수준으로 통계치를 작성하는 것을 기본목표로 삼고 있다(통계생산에 필요한 주요 항목 및 분류 <표 2-3> 참조).
- 조사구수가 많은 읍면동은 추출율을 적게 하고, 조사구수가 적은 읍면동은 추출율을 크게 하여 조사의 정확도를 향상할 수 있는 방향으로 표본조사구수를 할당하였다.
- 벤치마킹자료는 인구 및 주택의 행정자료(행정센서스)를 이용하여 추정가중치를 산정하였다.
- 2010년 기준으로 조사자료와 등록자료를 읍면동별로 비교하여 양 자료의 차이가 큰 지역과 1인 가구 및 외국인 가구가 많은 지역은 취약지구로 분류하여 표본 규모를 할당된 규모보다 상향 조정하였다.
- 조사구설정 및 표본틀은 2014 가구주택 기초조사의 잠정결과를 이용하였다.

<표 2-3> 주요 통계생산에 필요한 항목 및 분류 (2010 조사기준)

	인구수		가구수	
	교육정도	혼인상태		거처종류
항목	초등학교 (졸업, 재학, 중퇴) 중학교 (졸업, 재학, 중퇴) 고등학교 (졸업, 재학, 중퇴) 대학교4년제미만(졸업, 재학, 수료, 휴학, 중퇴) 대학교4년제이상(졸업, 재학, 수료, 휴학, 중퇴) 대학원석사과정(졸업, 재학, 수료, 휴학, 중퇴) 대학원박사과정(졸업, 재학, 수료, 휴학, 중퇴) 안받았음(미취학 포함)	미혼 배우자있음 사별 이혼		주택 단독주택 아파트 연립주택 다세대주택 비거주용 건물내주택 주택이외의 거처
분류	-	성별 연령별(5세)	1인 2인 ... 7인이상	
2010공표수준	시군구	시군구	시군구	시군구

3. 표본설계

가. 지역별 표본규모 배분 방법

표본규모는 추정통계치의 생산 및 조사집행 편의상 행정구역별(시도, 시군구 및 읍면동) 표본이 요구된다. 전체표본 조사구수를 행정구역별로 할당하는 방법은 크게 읍면동 직접 할당법 및 간접 할당법으로 구분된다.

- **읍면동직접 할당법(Bottom-up):** 시도 및 시군구의 지역을 고려하지 않고 읍면동을 부차모집단(subpopulation)으로 지역층화하여 전체 표본조사구수를 읍면동에 직접 할당하고, 시군구 및 시도는 할당된 읍면동 표본조사구수를 누적하는 방법으로 간접적으로 표본규모를 계산한다.
- **읍면동간접 할당법(Top-down):** 전국을 시도의 일차 부차모집단 (1st subpopulation)으로 지역층화하여 전체 표본조사구수를 먼저 시도에 할당하고, 할당된 시도의 표본 조사구수를 기초로 각 시도내의 시군구를 이차 부차모집단 (2nd subpopulation) 지역층화한다. 이를 통해 시군구의 표본규모를 할당한 다음, 각 시군구내의 읍면동을 삼차 부차모집단(3rd subpopulation)으로 지역층화하여 읍면동에 표본조사구수를 할당한다.

나. 표본규모 배분법

표본규모를 배분하는 방법은 부차모집단의 분산 정도를 이용하여 표본규모를 부차모집단에 할당할 수 있는 네이만 배분법과, 부차모집단의 분산에 관계없이 사용할 수 있는 동등배분법(Equal allocation), 비례배분법(Proportional allocation), 제곱근배분법(Squareroot allocation), 거듭제곱배분법(Power allocation) 및 키쉬배분법(Kish allocation) 등의 적합성 여부를 검토하였다(<표 2-4> 참조).

(1) 동등배분법(Equal allocation method):

이 방법은 층별 분산의 값을 알지 못 할 경우에 사용이 가능한 가장 간편한 방법으로 각층의 표본수를 동일하게 배분한다.

- 층의 크기가 비슷한 경우에 유용한데, 층의 크기가 다양한 경우 작은 층은 추정의 정도가 향상되는 반면 큰 층은 추정의 정도가 감소되므로 층을 고려하지 않은 전체 추정은 정확도가 감소된다.
- 이 방법은 층별 추정이 필요하고, 층별 비용과 분산에 큰 차이가 없을 때 사용된다.
- 지역(읍면동)별 추정에 정확도를 향상시키는데 적합한 방법이다(Kish:1988).

(2) 비례배분법(Proportional allocation method):

이 방법은 층별 분산의 값을 알지 못 할 경우에 사용 가능한 방법으로 조사항목이 많은 표본조사에서 유용하다.

- 단점으로는 층의 크기가 작은 층은 표본수가 적게 할당되므로 층별 추정치를 비교하는데 문제가 발생할 수도 있다.
- 전체(시군구/시도/전국) 추정에 정확도를 향상시키는데 적합하다(Kish:1988).

(3) 네이만 배분법(Neyman allocation method):

이 방법은 층별 분산을 구할 수 있을 경우에 추출분산이 최소가 되게 층별 표본수를 할당하는 방법으로 표본설계에서 가장 효율이 높은 방법이다.

- 각 층별 분산의 추정은 과거의 조사자료나 예비조사로부터 추정되어 이용될 수 있는데 층별 분산의 근사치를 사용할 경우에는 그 효과가 기대할 만큼 좋지 못하다.
- 이 방법은 각 층별 분산이 상이한 경우 추출분산을 최소로 할 수 있고 추정의 정확도에 큰 효과를 기대 할 수 있는데, 층별 분산이 유사한 경우는 비례할당으로 접근한다.
- 여러가지 조사항목을 조사하는 표본조사에서는 어떤 항목을 설계변수로 지정하는가에 따라서 표본수가 다르게 배분될 수 있어 추정효율이 떨어질 수 있다.

(4) 제곱근 배분법(Square root allocation method):

제곱근 배분법은 거듭제곱근의 한 방법으로($K=0.5$) 층간의 표본오차를 평균화시키고자 하는 일반적인 방법이다.

- 이 방법은 층의 분산을 알지 못하는 경우에 사용 가능하며, 부차모집단의 다양하고 조사항목이 많은 조사에서는 멱배분(power) 결정에 어려움이 있는 경우 주로 사용되고 있다.

(5) 거듭제곱배분법(Power allocation method):

거듭제곱배분법은 비례배분법의 응용으로 층의 크기가 작은 층은 표본의 크기를 상대적으로 크게 할당하고, 층의 크기가 큰 층은 표본의 크기를 상대적으로 작게 할당하는 방법으로 층간의 표본오차를 평준화 시키고자 하는 방법이다.

- 이 방법은 층의 분산을 알지 못하는 경우에 사용이 가능하다.
- 부차모집단의 다양하고 조사항목이 많은 조사에서는 멱배분(power)의 결정에 어

려움이 있다. 이러한 경우, 평균적인 역배분(power)방법으로 제곱근배분법을 사용하는 것이 일반적인 추세이다.

- Bankier(1988)는 표본규모결정을 하는데 역배분(power)의 특성을 1981 캐나다 센서스 자료를 이용하여 설명하였다.

(6) 키쉬배분법(Kish optimal allocation method):

Kish(1976)가 제안한 방법으로 표본의 크기는 층별 및 전체의 중요 정도를 상대적으로 고려하여 할당하는 방법으로 층간의 표본오차를 평준화 시킬 수 있는 방법이다.

- 이 방법은 층의 분산을 알지 못하는 경우에 사용이 가능한 방법이며 조사항목이 많은 표본조사에서 사용한다.

<표 2-4> 표본규모 배분법에 사용한 공식들

Allocation method	Formula	Where
Equal	$n_h = n \frac{1}{H}$	N: population size (모집단: 총 조사구수) n: sample size(표본:표본조사구수)
Proportional	$n_h = n \frac{N_h}{N} = n W_h$	W: proportion of sub population S: standard deviation (표준편차)
Neyman	$n_h = n \frac{N_h S_h}{\sum N_h S_h} = n \frac{W_h S_h}{\sum W_h S_h}$ $n_h = n \frac{CV_h^2}{\sum CV_h^2} \text{ p167 박재수}$ $n_h = n \frac{N_h \sqrt{P_h Q_h}}{\sum N_h \sqrt{P_h Q_h}}, \text{ p 109 William G. Cochran}$ <p>If the maximum variance of proportion which is P=0.5 and Q=0.5, then proportional allocation</p>	CV: coefficient of variation(변동계수) P: Proportion(구성비) Q: (1-P) h: subscript for area(읍면동첨자) H: total number of areas(총읍면동수) K: power I _c : relative importance for combined areas(시도/시군구 중요정도) I _s : relative importance for individual areas(읍면동 중요정도)
Square root	$n_h = n \frac{\sqrt{N_h}}{\sum \sqrt{N_h}}$	
Power	$n_h = n \frac{N_h^k}{\sum N_h^k}, \text{ where } 0 < k < 1$ <p>If k=0.5, then square root allocation If k=0, then equal allocation</p>	
Kish optimal	$n_h = n \frac{\sqrt{H^{-2} + I W_h^2}}{\sum_{h=1} \sqrt{H^{-2} + I W_h^2}} \text{ where,}$ $I_c + I_s = 1 \quad I = \frac{I_c}{I_s}$ <p>With I = 0 , the Kish allocation reduces to the equal allocation. With I → ∞ , the Kish allocation tends to the proportional allocation</p>	

다. 배분법 사용의 현실성 검토

- 네이만 배분법은 특정 조사 항목을 설계변수로 선정하고, 과거 조사 자료를 이용하여 분산을 구한다. 이 방법의 적용을 위해서는 인구주택총조사에서 어떠한 항목들을 설계변수로 선정하고 변수의 분산에 따라 표본규모를 배분하는 문제가 중요하기 때문에 조사항목수 및 설계항목의 분산에 구애받지 않고 사용할 수 있는 표본배분법이 중점적으로 검토되어야 한다. 네이만 배분법에서 임의로 한 조사항목을 설계변수로 하여 분산을 최대치로 가정한다면($p=0.5$) 네이만 배분법은 비례배분법으로 접근한다.
- 거듭제곱 배분법은 부차모집단 및 조사항목이 많은 경우 역배분 결정에 어려움이 있어 인구주택총조사에는 적합하지 않다. 거듭제곱 배분법에서 $k=0.5$ 이면 제곱근 배분법과 동일하고, $k=0$ 이면 비례배분법과 동일하다.
- 키쉬배분법에서 I 의 값이 0에 근접하면 동등배분법과 동일하고, I 의 값이 무한대로 접근하면 비례배분법과 동일하게 되는 관계가 있다.
- 조사항목수 및 설계항목의 분산을 구하는 절차에 상관 없이 사용할 수 있는 표본배분법 방법들을 중점적으로 검토하여 배분법을 결정하는 것이 바람직하다.
- 키쉬(1987:section 7.3D)는 동배분법, 비례배분법, 제곱급배분법 및 키쉬 최적배분법을 실제 예를 들어 방법별 특성을 설명하였다.
- 키쉬(1988)의 토의에 따르면 전국/시도/시군구(cross domains) 추정에 중점을 두면 비례배분법이 적합하고, 읍면동(domain)만 추정에 초점을 맞추면 등배분법이 적합하다.

4. 표본규모할당 예시

표본규모 할당방법들을 설명하기 위하여 가상적으로 6개 읍면동으로 구성된 한 시군구의 가구수의 모집단을 다음 <표 2-5>와 같이 구성하였다. 가상적 표본설계에 주요사항은 다음과 같다.

- 가구원 규모별 가구수 중 4인 가구원으로 구성된 가구수를 주요항목 통계치로 선정하였고,
- 요구된 표본규모는 시군구 전체 가구수의 20%에 해당하는 가구수를 조사하고자 한다.

이 가상적인 표본설계의 목적은 각 표본규모 배분방법으로 산출하는 방법을 설명하고, 각 방법별로 할당된 표본규모의 표본추출 분산을 추정하여 각 표본규모할당 방법들의 효율성을 비교하고자 하였다.

<표 2-5> 가상적인 모집단 가구수, 4인가구원 특성치의 가구수 및 요구된 20% 표본가구수

Area (읍 면 동) (h)	모집단가구		4인가구원 특성치		요구된 20% 표본가구수
	가구수	구성비	가구수	구성비	
	Nh	Wh=Nh/N	Xh	Ph=Xh/X	n
A	500	0.07	255	0.51	n=7,000*0.2 =1400
B	600	0.09	330	0.55	
C	800	0.11	320	0.40	
D	1,000	0.14	430	0.43	
E	1,600	0.23	832	0.52	
F	2,500	0.36	1,000	0.40	
시군구	7,000(N)	1.00	3,167(X)	0.45	

가. 표본규모 배분법

표본설계 예시에서 사용한 표본규모 배분법은 앞서 사용의 현실성 검토 및 토의에서 제시된 네이만배분법과 거듭제곱배분법을 제외한 등배분법, 비례배분법, 제곱근배분법 및 키쉬배분법을 어떻게 적용하는지 가상모집단 자료를 이용하여 표본 배분을 설명하고 각 방법의 표본오차를 추정하여 방법의 특성들을 비교 검토하였다.

나. 표본추출분산추정방법

읍면동(Area) 4인 가구 규모의 가구수(X)의 표본추출분산은 모집단 가구수(N), 모집단의 구성비(P), 표본수(n) 및 표본 추출율(f)를 이용하여 다음의 추정공식을 이용하였고, 절대표본오차와 상대표본오차는 표본추출분산과 모집단 4인 가구원 규모의 가구수(x)를 이용하여 계산하였다.

$$\text{표본추출분산: } \hat{V}_h(x) = N_h^2(1-f_h)P_h(1-P_h)/n_h$$

- 절대표본오차: $SE_h = \sqrt{\hat{V}_h}$
- 상대표본오차: $RSE_h = 100 * SE_h / X_h$

각 할당방법의 표본추출분산의 크기는 지역 층의 구성비(P), 모집단 가구수(N), 4인 가구수(X)는 상수역할을 함으로써 분산추정에 영향이 없고 지역 층의 표본 수(n)로 할당된 추출율(f)만이 표본추출분산을 결정하는 유일한 변수로 작용한다. 다음 <표 2-6>, <표 2-7>, <표 2-8>, <표 2-9>에서 각 할당방법과 표본추출분산추정 계산 절차를 확인할 수 있다.

<표 2-6> Equal allocation

Area	표본			표본추출분산추정		
읍 면 동 (h)	할당비율	표본가구	추출율	표본추출분산	절대표본오차	상대표본오차
	Ah	nh	fh	\hat{V}_h	SE_h	RSE_h
	6/1	n*Ah	nh/Nh	가구수	가구수	%
A	0.17	233	0.47	143.800	11.99	4.70
B	0.17	233	0.39	234.912	15.33	4.64
C	0.17	233	0.29	469.241	21.66	6.77
D	0.17	233	0.23	810.309	28.47	6.62
E	0.17	233	0.15	2353.126	48.51	5.83
F	0.17	233	0.09	5862.931	76.57	7.66
시군구		1,398		9874.319	99.37	3.14

<표 2-7> Proportional allocation

Area	표본			표본추출분산추정		
읍 면 동 (h)	할당비율	표본가구	추출율	표본추출분산	절대표본오차	상대표본오차
	Ah	nh	fh	\hat{V}_h	SE_h	RSE_h
	Wh	n*Ah	nh/Nh	가구수	가구수	%
A	0.07	100	0.2	504.848	22.47	8.81
B	0.09	120	0.2	598.992	24.47	7.42
C	0.11	160	0.2	772.830	27.80	8.69
D	0.14	200	0.2	985.327	31.39	7.30
E	0.23	320	0.2	1602.448	40.03	4.81
F	0.36	500	0.2	2404.810	49.04	4.90
시군구	1	1400		6869.254	82.881	2.62

<표 2-8> Square Root allocation

Area	표본			표본추출분산추정		
읍 면 동 (h)	할당비율	표본가구	추출율	표본추출분산	절대표본오차	상대표본오차
	Ah	nh	fh	\hat{V}_h	SE_h	RSE_h
	$\sqrt{N_h}$	n*Ah/T	nh/Nh	가구수	가구수	%
A	22.36	159	0.32	269.671	16.42	6.44
B	24.49	174	0.29	365.671	19.12	5.79
C	28.28	201	0.25	575.040	23.98	7.49
D	31.62	225	0.23	848.002	29.12	6.77
E	40.00	285	0.18	1849.149	43.00	5.17
F	50.00	356	0.14	3623.662	60.20	6.02
시군구	196.76(T)	1,400		7531.195	86.782	2.74

<표 2-9> Kish allocation

읍 면 동 (h)	표본			표본추출분산추정		
	할당비율	표본가구	추출율	표본추출분산	절대표본오차	상대표본오차
	Ah	nh	fh	\hat{V}_h	SE_h	RSE_h
	$\sqrt{H^{-2} + IW_h^2}$ *	n*Ah/T	nh/Nh	가구수	가구수	%
A	0.18	173	0.35	237.550	15.41	6.04
B	0.19	179	0.30	351.228	18.74	5.68
C	0.20	193	0.24	607.000	24.64	7.70
D	0.22	209	0.21	932.087	30.53	7.10
E	0.28	270	0.17	1974.531	44.44	5.34
F	0.39	376	0.15	3398.400	58.30	5.83
시군구	1.467(T)	1,400		7500.796	86.607	2.73

*: 키쉬방법에서 l=1로 하여 표본규모를 결정하였다.

다. 추정된 표본오차의 분석

각 할당 방법의 추정된 상대표본오차를 <표 2-10>에 요약하였다.

- 6개 읍면동의 평균상대표본오차를 살펴보면 동등배분법이 6.04%로 가장 적고, 비례배분법이 6.99%로 가장 크고, 제곱근배분법 및 키쉬배분법이 각각 6.28%로 추정되었다.
- 시군구의 상대표본오차는 비례배분법이 2.62%로 가장 적고, 동등배분법이 3.14%로 가장 크고, 제곱근배분법 및 키쉬배분법은 각각 2.74% 및 2.73%로 추정되었다.
- 시군구의 상대표본오차는 읍면동의 평균상대표본오차와 비교하면 절반수준으로 정도가 향상되었다. 예를 들면 비례배분법은 읍면동 평균 6.99%에서 시군구의 2.62%로 추정정도가 향상될 것으로 기대되어진다.

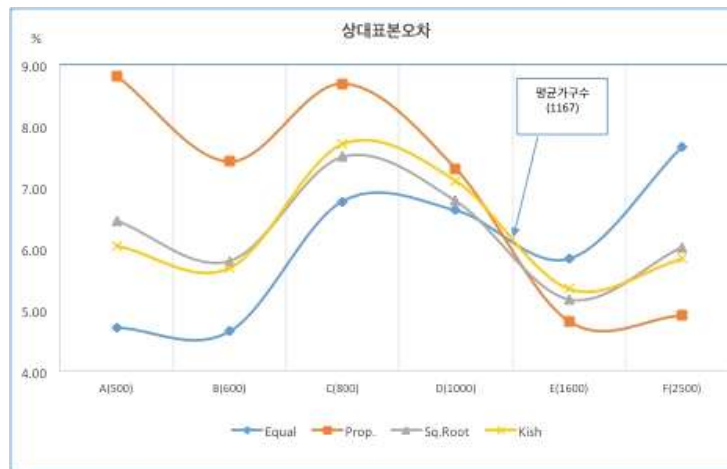
<표 2-10> 읍면동별, 표본할당방법별 상대표본오차 요약(단위:%)

Area 읍면동(가구수)	동등배분법 (Equal.)	비례배분법 (Prop.)	제곱근배분법 (Sq.Root)	키쉬배분법 (Kish)
A(500)	4.70	8.81	6.44	6.04
B(600)	4.64	7.42	5.79	5.68
C(800)	6.77	8.69	7.49	7.70
D(1000)	6.62	7.30	6.77	7.10
E(1600)	5.83	4.81	5.17	5.34
F(2500)	7.66	4.90	6.02	5.83
Ave.(1167)	6.04	6.99	6.28	6.28
시군구	3.14	2.62	2.74	2.73

각 할당방법의 추정된 상대표본오차를 읍면동의 가구수 규모별로 나열하여 비교하여 보면 다음 [그림 2-1]과 같다.

- 읍면동의 가구수가 가장 적은 A동부터 D동까지는 비례배분법이 상대표본오차가 다른 배분방법에 비하여 상대적으로 크게 추정되었고, 반면에 동등배분법은 상대적으로 작게 추정되었다.
- 읍면동의 가구수가 큰 E와 F동은 비례배분법이 상대표본오차가 다른 배분방법에 비하여 상대적으로 작게 추정되었고, 반면에 동등배분법은 상대적으로 크게 추정되었다.
- 제곱근배분법 및 키쉬배분법은 가구수가 크고 작은 그룹에 관계없이 비례배분법과 동등배분법의 평균수준정도로 상대표본오차가 추정되었다.
- 평균가구수(1167) 주변의 상대표본오차는 네 가지 방법 모두가 거의 유사하게 추정되어지는 경향이 있다.

[그림 2-1] 할당방법 및 지역층별 상대표본오차의 비교



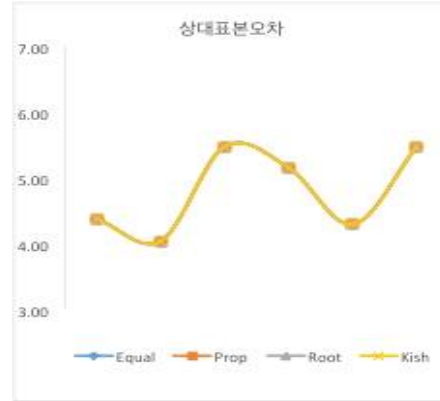
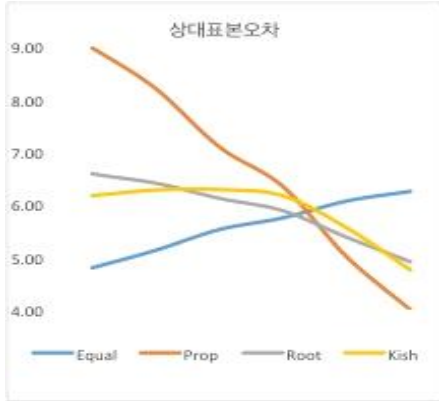
가구수 및 특성치의 영향

- 모집단 가구수를 변수로 하고 특성치를 상수로 가정한다면, [그림 2-2] 좌측 도표와 같이 읍면동 및 할당 방법별로 상대표본 오차의 추정 정도가 다르고,
- 모집단 가구수를 상수로 하고 특성치를 변수로 가정한다면, [그림 2-2] 우측 도표와 같이 읍면동별로 추정의 정도는 차이가 있지만 할당방법별로는 추정의 정도는 차이가 없다.

[그림 2-2] 가상적인 모집단 분석

Assuming that different N_h and same $Ph=0.5$

Assuming that same $N_h=2000$ and different Ph



제2절 추정가중치 산정을 위한 시험조사 자료분석

1. 자료 분석 목적

본 절은 항목별 응답상태 및 응답특성을 고려하여 가중치 계산절차에 적용할 가구의 무응답 변수 도출방법을 결정하고, 가구수 추정을 자료의 출처별로 비교하여 행정센서스의 벤치마킹자료의 조정여부에 판달할 자료를 제공하고자 한다. 이를 위해 시험조사 자료를 분석하였다.

2. 시험조사가 시행된 읍면동

시험조사는 2차(2013.9) 및 3차(2014.4)로 나누어 5개 읍면동을 대상으로 시행하였다. 조사된 표본 조사는 전체 1,063 조사구 중 220(21%)개 조사구를 표본으로 선정하여 조사하였고, 조사된 결과는 가구명부, 가구부분, 인구부분 및 주택부분으로 구분하여 구성하였다. 시험조사의 읍면동 조사구수는 <표 2-11>과 같다.

<표 2-11> 시험조사 읍면동 현황

시험조사	행정구역		조사구수		조사된 건수			
	코드	명칭	모집단	표본	명부	가구	인구	주택
2차 (2013.9)	2501057	대전 동구 용운동	376	68	2,025	1,730	3,648	1,075
	3403033	충남 보령시 오천면	59	15	589	362	748	356
	3701232	경북 포항시북구 청하면	84	23	849	457	941	425
3차 (2014.4)	1107052	서울 중랑구 면목2동	299	62	2,267	2,112	5,132	2,041
	3337012	충북 음성군 금왕읍	245	52	1,959	1,645	4,030	1,638
합계			1,063	220	7,689	6,306	14,499	5,535

표본조사구 추출 방법은 읍면동을 부차모집단으로 하고, 각 부차 모집단을 조사구 특성별(일반조사구 및 아파트조사구)로 층화하여 20%의 조사구를 표본으로 추출하였으므로 가중치는 각 층의 추출확률의 역수로 표시할 수 있는데, 이는 각 층의 총조사구를 표본조사구로 나누어 계산할 수 있다. 기초조사의 표본틀로부터 계산된 추출 가중치는 <표 2-12>과 같다.

<표 2-12> 조사구 특성별 추출 승수

행정구역	조사구* 특성	조사구			가구		
		총 조사구수	표본 조사구수	조사구 추출승수 (SelWgt)	총 가구수	표본 조사구수	가구 추출승수
1107052	1	241	49	4.9184	8,230	1,710	4.8129
1107052	A	58	13	4.4615	2,186	478	4.5732
2501057	1	216	40	5.4000	5,972	1,090	5.4789
2501057	A	160	28	5.7143	4,757	826	5.7591
3337012	1	130	28	4.6429	4,361	971	4.4912
3337012	A	115	24	4.7917	4,130	872	4.7362
3403033	1	59	15	3.9333	2,141	587	3.6474
3701232	1	78	17	4.5882	2,764	636	4.3459
3701232	A	6	6	1.0000	210	210	1.0000
합계		1,063	220		34,751	7,380	

*: 1=일반조사구, A=아파트조사구

3. 가구명부 분석

가. 가구명부 작성 목적

가구명부는 본 조사를 효과적으로 수행하고, 조사의 누락 및 중복방지를 하고자 하는 목적으로 작성되며, 작성된 결과는 가구부문 추정가중치 보정의 기초자료로 이용되기도 한다.

나. 가구명부의 주요항목

가구명부에 조사된 주요항목은 다음과 같다.

- 거처종류(HU_KIND_CD): 1(단독주택), 2(아파트), 3(연립주택), 4(다세대주택), 5(비거주용건물내주택), 6(주택이외의 거처)
- 빈집여부(UH_YN): 1(빈집), 2(빈방), 3(건물없음), 4(거처 가구가 아님)
- 조사상태(UN_SURV_RSN_CD): 1(조사완료), 2(부재), 3(불응)
- 남자가구원수(MAN_HM_CNT): 0, 1, 2,...
- 여자가구원수(WMAN_HM_CNT): 0, 1, 2,...

기초조사에서 작성된 가구는 시험조사 지역의 총 7,380가구이다. 원래 선정된 표본과 시험조사 간 차이가 나는 가구수는 309개(4.2%)인데, 이는 시험조사 과정에서 가구가 추가로 작성되어 조사되었기 때문이다. 면 지역은 차이가 근소하고(1% 미만) 동/읍 지역은 3%이상의 차이가 나는 것으로 나타났다. 이는 면 지역 조사와 동/읍 지역 조사의 시점이 다른 관계로 나타나는 현상이다.

<표 2-13> 기초조사 및 시험조사의 가구수

Hang	표본 조사구수	기초조사	시험조사	차이	
		가구수 (A)	가구수 (B)	B-A	%
1107052(서 울면목동)	62	2,188	2,267	79	3.6
2501057(대전 용운동)	68	1,916	2,025	109	5.7
3337012(충북 금왕읍)	52	1,843	1,959	116	6.3
3403033(충남 오천면)	15	587	589	2	0.3
3701232(경북 청하면)	23	846	849	3	0.4
Total	220	7,380	7,689	309	4.2

다. 주요항목별 응답분포

주요항목별 응답분포는 다음의 <표 2-14>와 같다. 항목별 무응답 여부에 대한 분석을 통해 무응답 가구와 부적격 가구를 판단하는 항목간 연관성을 도출할 수 있다.

- 거처종류를 무응답으로 처리한 가구는 551가구로 전체의 7.2%를 차지하는데, 거처종류는 가구 혹은 주택의 승수산정에 층화 변수로 사용이 가능한 변수이므로 무응답부분은 반드시 대체값(imputation)으로 처리해야 층화 변수로 이용이 가능하다.
- 빈집 여부 항목에 6,451(83.9%)가구는 응답가구로 간주되며, 나머지는 부적격 가구로 응답상태를 구분할 수 있다.
- 조사상태의 부재 및 불응인 305(4%)가구는 무응답으로 간주된다.
- 가구수 중 471(6.1%)가구에서 가구원수가 무응답으로 조사되었고, 0명으로 조사

된 가구는 918(11.9%)가구이다. 이는 조사표 검토 과정에서 확인 되어야 할 부분으로 보인다. 가구원수는 인구 혹은 가구부문의 총화변수로 사용 가능한 변수이므로 무응답 부문은 인구부문 조사결과를 이용하여 반드시 대체값(imputation)으로 처리하는 것이 바람직하다.

<표 2-14> 거처의 종류별, 빈집 여부별, 조사상태별 및 가구원수별 가구수

[1] 거처의종류별 가구수

HU_KIND_CD	COUNT	PERCENT
Missing(무응답)	551	7.2
1(단독주택)	4,059	52.8
2(아파트)	2,292	29.8
3(연립주택)	262	3.4
4(다세대주택)	428	5.6
5(비거주용건물내주택)	92	1.2
6(주택이외의 거처)	5	0.1
합계	7,689	100.0

[2] 빈집 여부별 가구수

UH_YN	COUNT	PERCENT
공란	6,451	83.9
1(빈집)	687	8.9
2(빈방)	246	3.2
3(건물없음)	15	0.2
4(거처가아님)	290	3.8
합계	7,689	100.0

[3] 조사상태별 가구수

UN_SURV_RSN_CD	COUNT	PERCENT
공란	1,241	16.1
1(조사완료)	6,143	79.9
2(부재)	155	2.0
3(불응)	150	2.0
합계	7,689	100.0

[4] 가구원수별 가구수

Persons	COUNT	PERCENT
Missing(무응답)	471	6.1
0	918	11.9
1	2,115	27.5
2	1,812	23.6
3	1,154	15.0
4	929	12.1
5+	290	3.8
합계	7,689	100.0

4. 가구의 응답상태 분류

조사항목의 무응답의 형태를 분석하여 조사가구의 응답상태를 도출할 수 있는데, 판단방법은 빈집여부 및 조사상태의 결과를 이용한다(절차 <표 2-15>참조). 가구의 응답상태(responsestatus) 구분은 다음과 같이 정의하여 추정 가중치 계산에 사용 할 수 있다.

- A: 응답(Response): 조사표가 작성된 조사가구
- B: 무응답(Nonresponse): 조사가 이루어지지 않은 가구
- C: 부적격(Ineligible): 조사대상이 부적합한 가구

표본가구(7,689)에 대한 시험조사의 응답상태를 보면 응답가구는 6,143(79.9%)가구이고, 무응답 가구는 308(4.0%)가구이며, 부적격 가구는 1,238(16.1%)가구로 집계되었다.

<표 2-15> 응답상태 분류별 조사된 표본가구수

가구명부		가구 응답상태	시험조사	
빈집여부 (UH_YN)	조사상태 (UN_SURV_RSN_CD)		가구수	%
공란	공란	B:무응답	3	0.04
	조사완료(1)	A:응답	6,143	79.89
	부재(2)	B:무응답	155	2.02
	불응(3)	B:무응답	150	1.95
빈집(1)	공란,조사완료(1),부재(2),불응(3)	C:부적격	687	8.93
빈방(2)	공란,조사완료(1),부재(2),불응(3)	C:부적격	246	3.20
건물없음(3)	공란,조사완료(1),부재(2),불응(3)	C:부적격	15	0.20
거처(가구)아님(4)	공란,조사완료(1),부재(2),불응(3)	C:부적격	290	3.77
Total			7,689	100.00

추출가중치(<표 2-12>: 조사구 추출 승수)를 사용한 추정 가구수를 살펴보면 (<표 2-16>), 추정된 총 가구수는 36,810가구로 적격 가구는 31,376(85.2%), 부적격 가구는 5,434(14.8%)가구이고 무응답 가구는 1,382(3.8%)가구이다.

<표 2-16> 응답상태 분류별 추정 모집단가구수

조사	행정구역	표본가구	추정된 총 가구수				
			응답가구 (A)	무응답가구 (B)	부적격가구 (C)	총 가구 (A+B+C)	적격가구 (A+B)
2차	3403033(충남 오천면)	589	1,420	43	854	2,317	1,463
	3701232(경북 청하면)	849	1,861	90	1,191	3,142	1,951
	2501057(대전 용운동)	2,025	9,583	448	1,163	11,194	10,031
3차	1107052(서울 면목동)	2,267	9,969	204	758	10,931	10,173
	3337012(충북 금왕읍)	1,959	7,161	597	1,468	9,226	7,758
	합계	7,689	29,994	1,382	5,434	36,810	31,376

5. 거처종류별 가구수 추정

거처종류는 가구명부에서 조사된 항목으로 가구 및 주택의 가중치를 부여하기 위한 층화 변수로 사용할 수 있다. 등록 센서스를 벤치마킹 자료로 사용한다면, 조사된 거처 종류 가운데 연립주택 및 다세대주택은 한 분류로 통합하고, 비거주용 건물 내 주택 및 주택 이외의 거처는 기타로 통합해야 등록 센서스를 이용할 수 있다. 분류의 연관관계는 <표 2-17>과 같다.

<표 2-17> 거처종류의 재분류

가구명부 (HU_KIND_CD)		추정에 이용할 재 분류 (TYPE)	
거처의 종류	코드	거처의 종류	코드
단독주택	1	단독주택	1
아파트	2	아파트	2
연립주택	3	연립, 다세대	3
다세대주택	4	기타	4
비거주용건물내 주택	5		
주택이외의 거처	6		
무응답	9	무응답	9

조사된 533(6.9%)가구는 거처종류가 무응답이므로 이러한 가구는 무응답 처리가 선행되어야 승수산정을 위한 사후층화가 가능하다. 거처종류별응답, 무응답 및 부적격 가구수를 행정구역별로 추정한 적격가구 비율을 <표 2-18>에 정리하였다.

<표 2-18> 행정구역 및 거처종류별 추정가구수

행정구역	거처종류 (TYPE)	표본 가구	표본으로 추정된 모집단가구수					
			응답 (A)	무응답 (B)	부적격 (C)	합계 (A+B+C)	적격가구 비율	적격가구 (A+B)
충남 보령시 오천면 (3403033)	합계	589	1,420	43	854	2,317		1,463
	1(단독주택)	373	1,416	43	8	1,467	0.9945	1,459
	2(아파트)	-	-	-	-	-	-	-
	3(연립. 다세대)	-	-	-	-	-	-	-
	4(기타)	1	4	-	-	4	1.0000	4
	9(무응답)	215	-	-	846	846	-	-
경북 포항시 북구 청하면 (3701232)	합계	849	1,861	90	1,191	3,142	-	1,951
	1(단독주택)	516	1,597	41	730	2,368	0.6917	1,638
	2(아파트)	120	47	33	94	174	0.4598	80
	3(연립. 다세대)	188	121	16	349	486	0.2819	137
	4(기타)	23	96	-	9	105	0.9143	96
	9(무응답)	2	-	-	9	9	-	-
대전 동구 용운동 (2501057)	합계	2,025	9,583	448	1,163	11,194		10,031
	1(단독주택)	1,092	5,001	233	664	5,898	0.8874	5,234
	2(아파트)	824	4,377	204	126	4,707	0.9732	4,581
	3(연립. 다세대)	39	194	11	5	210	0.9762	205
	4(기타)	2	11	-	-	11	1.0000	11
	9(무응답)	68	-	-	368	368	-	-
서울 중랑구 면목2동 (1107052)	합계	2,267	9,969	204	758	10,931	-	10,173
	1(단독주택)	1,363	6,246	148	310	6,704	0.9538	6,394
	2(아파트)	476	2,070	36	18	2,124	0.9915	2,106
	3(연립. 다세대)	322	1,505	20	59	1,584	0.9628	1,525
	4(기타)	30	148	-	-	148	1.0000	148
	9(무응답)	76	-	-	371	371	-	-
충북 음성군 금왕읍 (3337012)	합계	1,959	7,161	597	1,468	9,226		7,758
	1(단독주택)	729	2,665	241	478	3,384	0.8587	2,906
	2(아파트)	876	3,781	268	149	4,198	0.9645	4,049
	3(연립. 다세대)	141	525	88	42	655	0.9359	613
	4(기타)	41	190	-	-	190	1.0000	190
	9(무응답)	172	-	-	799	799	-	-

6. 자료 출처별가구수 비교

시험조사를 시행한 읍면동을 등록센서스의 가구수와 표본으로 추정된 가구수와 2010년 인구총조사의 가구수와 비교하였다(<표 2-19>참조). 먼저 등록센서스와 표본으로 추정된 가구수를 2차, 3차 시험조사별로 비교하여 보면 다음과 같다.

2차 시험조사 비교:

- 자료수집 시점은 2개월 차이로 변화가 근소하나, 면지역은 모두 등록센서스가 표본추정보다 크게 집계되었다. 이는 주민등록은 있지만 실제로는 거주하지 않은 인구가 많은 것으로 추정된다.
- 동지역은 면지역과 반대로 표본추정이 크게 집계되었다. 이는 주민등록은 없지만 실제로는 거주하는 인구가 많은 것으로 추정된다.

3차 시험조사 비교:

- 자료작성 시점은 5개월 차이로 변화가 근소하다. 동지역은 근소한 차이지만 등록센서스가 크게 집계되었고, 읍지역은 2차 시험조사와 유사한 추세로 등록센서스가 크게 집계되었다.

등록센서스 표본추정과 2010년 인구총조사를 비교하여보면 다음과 같다.

- 자료작성 시점은 3년 정도의 차이로 시점으로 인한 차이는 추정에 영향이 있을 것으로 보인다.
- 용운동, 면목동, 금왕읍은 표본추정값에 근접하고, 청하면은 등록센서스 값에 근접하며, 오천면은 등록센서스와 표본추정 중간값에 접근하는 것으로 나타났다.

용운동을 제외하고, 등록센서스가 표본추정 가구수보다 크게 집계 되었음이 자료분석 결과 드러났다.

<표 2-19> 출처별 가구수

행정구역	표본수 (2013.9)	기초조사 (2013.9)	등록센서스 (2013.11)	표본추정 (2013.9)	인구총조사 (2010.11)
2차조사 기준시점					
3403033(충남 오천면)	589	2,141	2,374	1,463	1,872
3701232(경북 청하면)	849	2,974	2,472	1,951	2,383
2501057(대전 용운동)	2,025	10,729	7,902	10,031	10,522
3차조사 기준시점					
1107052(서울 면목동)	2,267	10,416	10,715	10,173	10,206
3337012(충북 금왕읍)	1,959	8,491	8,385	7,758	7,514
합계	7,689	34,751	31,848	31,376	32,497

등록센서스에 부적격 가구가 있을 것이라는 가정하에, 표본으로 추정된 적격가구 비율을 이용하여 등록센서스의 적격 가구수를 추정하여 비교하여 보면 다음 <표 2-20>과 같다.

- 오천면은 여전히 등록센서스와 표본추정 간 큰 차이가 있고, 청하면은 부적격가구의 비율이 높아 표본추정 가구수가 조정된 등록센서스 가구 수보다 큰 것으로 나타났다.
- 면목동과 금왕읍은 표본추정치 등록센서스 및 인구총조사의 추정가구수가 유사한 것으로 집계되었다.

<표 2-20> 등록센서스의 적격 가구수의 추정 요약

행정구역	표본수	등록센서스 (조정전)	등록센서스 (조정후)	표본조사 추정	인구총조사
2차조사 기준시점	(2013.9)	(2013.11)		(2013.9)	(2010.11)
3403033(충남 오천면)	589	2,374	2,363	1,463	1,872
3701232(경북 청하면)	849	2,472	1,769	1,951	2,383
2501057(대전 용운동)	2,025	7,902	7,445	10,031	10,522
3차조사 기준시점	(2014.4)	(2013.11)		(2014.4)	(2010.11)
1107052(서울 면목동)	2,267	10,715	10,368	10,173	10,206
3337012(충북 금왕읍)	1,959	8,385	7,806	7,758	7,514
합계	7,689	31,848	29,751	31,376	32,497

적격가구수를 거처의 종류별로 나누어 분석하면 <표 2-21>과 같다.

- 충남 보령시 오천면은 대다수가 단독주택인 높은 적격가구 비율로 조정 후의 차이가 거의 없다.
- 경북 포항시 청하면은 기타를 제외한 모든 분류가 낮은 적격가구 비율로 조정 후 추정가구수가 큰 차이로 조정되어 표본으로 추정된 가구수에 근접하였다.
- 대전 용운동의 단독주택은 다소 낮은 적격가구 비율이고, 아파트는 높은 적격가구 비율이지만 전체적으로 높은 적격가구율로 추정 전후의 차이가 거의 없다.
- 서울 면목2동은 모든 거처종류가 높은 적격가구 비율로 조정 전후의 차이가 거의 없다.
- 충북 금왕읍은 단독주택의 적격가구율이 낮지만 다른 거처종류는 적격가구비율이 높아 조정 전후의 가구수에 다소 차이가 있다.

<표 2-21> 등록센서스의 적격 가구수의 거처종류별 추정

행정구역	거처의종류 (Type)	표본으로 추정된 가구수			등록센서스	
		합계	적격가구비율	적격가구	조정전	조정후
	합계	2,317	-	1,463	2,374	2,363
충남 보령시 오천면 (3403033)	1(단독주택)	1,467	0.9945	1,459	1,929	1,929
	2(아파트)	-	-	-	-	-
	3(연립.다세대)	-	-	-	11	-
	4(기타)	4	1.0000	4	434	434
	9(무응답)	846	-	-	-	-
	합계	3,142	-	1,951	2,472	1,769
경북 포항시 북구 청하면 (3701232)	1(단독주택)	2,368	0.6917	1,638	1,944	1,345
	2(아파트)	174	0.4598	80	31	14
	3(연립.다세대)	486	0.2819	137	70	20
	4(기타)	105	0.9143	96	427	390
	9(무응답)	9	-	-	-	-
	합계	11,194	-	10,031	7,902	7,445
대전 동구 용운동 (2501057)	1(단독주택)	5,898	0.8874	5,234	2,957	2,624
	2(아파트)	4,707	0.9732	4,581	4,318	4,202
	3(연립.다세대)	210	0.9762	205	340	332
	4(기타)	11	1.0000	11	287	287
	9(무응답)	368	-	-	-	-
	합계	10,931	-	10,173	10,715	10,368
서울 종랑구 면목2동 (1107052)	1(단독주택)	6,704	0.9538	6,394	5,930	5,656
	2(아파트)	2,124	0.9915	2,106	2,339	2,319
	3(연립.다세대)	1,584	0.9628	1,525	1,416	1,363
	4(기타)	148	1.0000	148	1,030	1,030
	9(무응답)	371	-	-	-	-
	합계	9,226	-	7,758	8,385	7,806
충북 음성군 금왕읍 (3337012)	1(단독주택)	3,384	0.8587	2,906	2,938	2,523
	2(아파트)	4,198	0.9645	4,049	3,811	3,676
	3(연립.다세대)	655	0.9359	613	454	425
	4(기타)	190	1.0000	190	1,182	1,182
	9(무응답)	799	-	-	-	-

제3절 가중치 산정 절차(Weighting steps)

1. 가중치 연구목적

이 절의 목적은 인구주택총조사의 읍면동을 부차모집단(sub-population)으로 하여, 가구, 인구, 주택 통계를 반영한 가중치 산정 방법들을 검토함으로써 인구주택총조사의 추정 정확성을 높일 수 있는 방법을 제안하는 것이다.

2. 표본 추출 가중치(selection weight)

표본추출 가중치는 표본들을 이용하여 추출 가중치 및 추출 가중치 조정을 위한 조중인자(adjusted factor)를 구할 수 있다. 이 절에서의 가중치 계산절차는 읍면동을 독립적으로 처리하도록 하여 공식에 나타나는 첨자를 간소화하였다.

가. 조사구추출 가중치(ED selection weight)

이 방법은 표본 설계 가중치(design-based weight), 혹은 Horvitz-Thompson (HT, 1952) 가중치라고도 하는데, 이 가중치의 산정은 표본 조사구 추출확률(probability of selection for sample unit)의 역수로 표시하기도 한다. 즉, 읍면동을 부차모집단으로 하고 부차모집단을 두 개의 층(일반조사구 및 아파트조사구: h=1,2)으로 층화하여 추출 가중치($EDSelWgt_h$)를 다음과 같이 산정할 수 있다.

$$EDSelWgt_h = \pi_h^{-1} = \frac{N_h}{n_h} \quad \text{-----}(1)$$

여기서 π_h =조사구 추출확률, N=총 조사구수, n=표본조사구수, h=조사구특성(1=일반조사구, 2=아파트조사구)

나. 조사구추출 가중치 조정(ED selection adjustment)

인구주택총조사의 표본틀(samplingframe)은 조사구별로 가구수 정보가 제공되는데, 여기에서는 모집단 조사구의 가구수와 추출된 표본 조사구의 추정 가구수 정보를 이용하여, 표본 조사구의 추출 가중치를 조정함으로써 추정 정확도를 향상시킬 수 있는 방안을 살펴보았다. 표본조사구의 조정된 추출가중치($AdjEDSelFact$)와 추출가중치($SelWgt$)는 다음과 같이 계산 할 수 있다.

$$AdjEDSelFact = \frac{\sum_h M_h}{\sum_h (EDSelWgt_h * m_h)} \quad \text{-----}(2)$$

M_h =모집단 조사구의 총가구수, m_h =추출된 표본조사구의 총 가구수

이 조정된 추출가중치는 가구명부에 있는 모든 조사 적격가구(j)에 할당한다. 즉,

$$SelWgt_{h,j} = EDSelWgt_h * AdjEDSelFact \quad \text{-----}(3)$$

3. 무응답 가구조정 (Non-responding household adjustment)

조사가 시행된 후, 조사로 도출된 가구응답 상태(response status: <표 2-15> 참조)를 이용하여 가구무응답에 대한 가중치를 적용해 조정을 할 수 있는데, 읍면동 응답 조정계수($AdjrespFact$)는:

$$AdjrespFact = \frac{\sum_{All\ eligible\ households \in k} Sel\ Wgt_k}{\sum_{Responding\ households \in k} Sel\ Wgt_k} \text{ -----(4)}$$

All eligible households=표본조사구의 총 적격가구(response status A+B), Responding households=총 응답가구(response status A), k =응답조정계수그룹(읍면동)

조사 시행 후 응답된 모든 표본가구(j)는 조정된 표본가구가중치(AdjSelHHWgt_{h,j})를 부여 받게 된다.

$$AdjSelHHWgt_{h,j} = Sel\ Wgt_h * AdjrespFact \text{ ----- (5)}$$

조사된 가구 개인 및 주택은 동일하게 조정된 표본가구 가중치를 부여하여, 가구 개인 및 주택관련 통계치 추정에 사용할 수 있는 것이 전통적인 표본설계 가중치 산정방법이다. 벤치마크자료 이용이 가능한 경우에는 사후층화 방법으로 추정의 정확도를 향상시킬 수 있으므로 사후층화 방법으로 추정가중치조정이 필요하다.

4. 사후층화를 이용한 추정가중치 조정(Post-stratification weight)

2010년 인구주택총조사는 가구 및 주택 분야는 전수조사로 시행되었으므로, 인구분야의 표본조사부문인 인구의 가중치는 조사된 총인구수를 벤치마크로 하여 사후층화 방법으로 승수를 산정하였다. 하지만 2015년 인구주택총조사는 가구 인구 및 주택 모두가 표본조사로 자료를 수집하는 관계로 벤치마크는 행정센서스 자료가 사용 가능하다. 이에 가구, 개인, 주택으로 분리하여 각각 사후층화방법으로 승수를 산정할 수 있는 방법을 제시하였다.

가. 가구부문 사후층화 (Post-stratificationfor household)

가구의 사후층화 조정계수는 다음과 같이 계산할 수 있다.

$$PostHHfact_l = \frac{PopHH_l}{\sum_j AdjSelHHWgt_{l,j}} \text{ -----(6)}$$

여기서, PopHH_l =행정센서스의 총 가구수 l =가구규모

가구분야의 사후층화 최종 가중치는 다음과 같이 각 응답 표본가구(j)에 할당된다.

$$HHWgt_j = AdjSelHHWgt_{l,j} * PostHHfact_l \text{ -----(7)}$$

나. 인구부문 사후층화(Post-stratification for person)

인구분야의 사후층화 가중치를 계산하기 위하여 가구분야의 사후층화 최종가중치(HHWgt)를 각 가구에 해당하는 모든 개인에 할당 한 다음, 인구분야의 사후층화

조정계수를 다음과 같이 계산할 수 있다.

$$PopPSfact_l = \frac{PopPS_l}{\sum_j HHWgt_{l,j}} \quad \text{-----}(8)$$

여기서, $PopPS_l$ = 행정센서스의 총 인구수, l = 성별 X 연령그룹.

인구분야의 사후층화 최종 가중치는 다음과 같이 응답한 개인(j)에 할당된다.

$$PSWgt_j = HHWgt_{l,j} * PostPSfact_l \quad \text{-----}(9)$$

다. 주택부문 사후층화(Post-stratification for house)

주택분야의 사후층화 조정계수는 다음과 같이 계산할 수 있다.

$$PostDWfact_l = \frac{PopDW_l}{\sum_j EDSel Wgt_{l,j}} \quad \text{-----}(10)$$

여기서, $PopDW_l$ = 행정센서스의 총 주택수, l = 거처의 종류.

주택분야의 사후층화 최종 가중치는 다음과 같이 각 응답가구의 표본주택(j)에 할당된다.

$$DWWgt_j = EDSel Wgt_{l,j} * PostDWfact_l \quad \text{-----}(11)$$

5. 사후층화 변수(Post-stratified variables)

사후층화로 추정의 정확도를 향상시키기 위해서는 층화변수의 선택과 더불어 층화변수의 카테고리 결정도 가중치 산정의 중요한 사항이다. 층화변수는 추정에 중요한 역할을 할 수 있는 변수를 선택하여야 하고, 변수의 카테고리는 적당한 크기의 표본단위가 할당할 정도의 카테고리를 결정하여야 한다.

가. 제안된 층화변수들

행정센서스의 자료로 가능한 층화변수 및 각각의 카테고리는 다음과 같다.

- 인구부문: 성별 X 5세 연령별(0-4, 5, 6-9, 10-11, 12, 13-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85세 이상)
- 가구부문: 거처의 종류(단독주택, 아파트, 연립·다세대 및 기타) 및 가구원수(1, 2, 3, 4, 5인 이상)
- 주택부문: 거처의 종류(단독주택, 아파트, 연립, 다세대, 비거주용 건물내 주택 및 주택 이외의 거처) 혹은 주거용 연면적(0-20, 20-40, 40-60, 60-85, 85-100, 100-130, 130-165, 165-230, 230m²이상 및 미상)

인구부문은 성별 및 5세 연령별 층화카테고리는 추정에 중요한 역할을 할 것으로 보이며, 각 카테고리별로 어느 정도의 표본수가 할당될 것이므로, 두 변수 모두 층화변수로 하여 초기 카테고리 사용 하는 것이 타당하다.

가구부문은 거처의종류 및 가구원수의 층화카테고리의 경우, 일부 분류는 적당한 크기의 표본수가 할당되기 어려운 세밀한 분류로, 거처의 종류 혹은 가구원수 중 한 변수만을 층화변수로 선택해야 하는데 이 중 가구원수가 추정에 중요한 변수로 이를 층화변수로 제안한다.

주택부문은 거처의 종류와 주거용 연면적 중 한 변수를 선택해야 하는데, 거처의 종류가 추정치 분류에 중요한 역할을 하므로 이를 층화변수로 제안한다.

나. 층화변수의 카테고리통합(Category collapsing for stratified variable)

표본으로 조사된 층화변수의 카테고리과 등록센서스의 층화변수의 카테고리가 일치하지 않는 경우는 두 자료 간 카테고리의 매칭을 위하여 카테고리를 통합하는 방식으로 진행하였다. 여기서 거처의 종류는 주택분야의 사후층화변수, 가구규모는 가구분야의 사후층화변수, 성별 및 연령 그룹은 인구분야의 사후층화변수로 사용하였다. 통합의 순서와 결정 방법은 다음과 같다.

□ 거처의종류

<표 2-22> 거처의 종류 카테고리 통합

거처의 종류	Type	Type1	Type2
1 단독	1	1	1
2 아파트	2	2	
3 연립주택	3		
4 다세대주택			
5 비거주용건물	4	3	
6 오피스텔			
7 호텔.여관			
8 기숙사			
9 판잣집.비닐하우스			
10 기타			

□ 가구규모(HHsize)

<표 2-23> 가구규모의 카테고리 통합

가구원수	HHsize	HHsize1	HHsize2	HHsize3
1명	1	1	1	1
2명	2	2	2	
3명	3			
4명	4			
5명이상	5	3		

□ 연령그룹(Age group)

<표 2-24> 연령 그룹의 카테고리 통합

Ageg	Age5g	Age10g	Age15g	Age50g	Age99g
0-4	0-4	0-9	0-14	0-14	0+
5	5-9				
6-9					
10-11	10-14	10-19	15-39	15-64	
12					
13-14					
15-19	15-19				
20-24	20-24	20-29			
25-29	25-29				
30-34	30-34	30-39			
35-39	35-39				
40-44	40-44	40-49	40-64		
45-49	45-49				
50-54	50-54	50-59			
55-59	55-59				
60-64	60-64	60-69	65+	65+	
65-69	65-69				
70-74	70-74	70-79			
75-79	75-79				
80-84	80-84	80+			
85+	85+				

6. 인구 및 가구의 통합 가중치 방법(Integrated person-household weighting)

가구·인구분야의 추정 정확도를 향상시킬 수 있는 한 방법으로는, 동일가구에 속하는 모든 가구원들은 가구단위의 추정가중치와 인구단위의 추정가중치를 동일하게 부여하는 방법인 가구·인구 통합 가중치방법(Lemaitre and Dufour(1987: The "integrated person-household weighting" method))을 논의하였다(앞서, 소개한 인구부문 단순 사후층화 방법은 동일가구내의 가구원은 성별 및 연령 그룹이 다르다면 각기 다른 추정가중치를 부여받는다).

모집단 크기를 N이라 하고, 추정하고자 하는 변수 Y가 값을 Y1, Y2, ... YN의

로 취한다고 하면, N-vector는 다음과 표현할 수 있고

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}$$

X를 층화변수(예: 성별 및 연령그룹)로 정의하고 이 층화변수의 표시변수(indicator variable)인 행렬 $N \times p$ 은 다음과 같이 표기할 수 있다.

$$X = \begin{bmatrix} X_{1,1} & \dots & X_{1,p} \\ \vdots & \ddots & \vdots \\ X_{N,1} & \dots & X_{N,p} \end{bmatrix}$$

이를 보조변수(auxiliary variables)로 정의하여 층화변수의 각 카테고리의 합계를 구하여 행렬 $p \times 1$ 로 표기하면 다음과 같다.

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, x_j = \sum_{i=1}^N X_{i,j}$$

예를 들면, 성별 및 연령그룹으로 사후층화하여 행렬변수의 카테고리의 합계를 남자연령그룹1($x_1=20$), 남자연령그룹2($x_2=10$), 여자연령그룹1($x_3=15$), 여자연령그룹2($x_4=20$)로 정의한 보조변수에 대한 벤치마킹 행렬은 다음과 같다.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \\ 15 \\ 20 \end{bmatrix}$$

다음과 같은 가상자료를 가지고 추정하고자 하는 가중치의 계산절차를 설명하고자 한다.

Area	ID	sex	age	weight
A	1	1	1	5.1
A	1	1	2	5.1
A	2	2	1	5.1
A	2	2	1	5.1
A	3	2	2	5.1
A	4	1	1	4.4
A	5	1	1	4.4
A	5	1	2	4.4
A	5	2	1	4.4
A	5	2	2	4.4
A	5	2	2	4.4

층화변수인 성별 및 연령그룹변수(sex age)를 가지고 표시변수로 변환하여 행렬을 작성하면 상기 예제 자료는 다음과 같은 11 x 4인 X행렬로 표시할 수 있다.

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad Z = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 2/2 & 0 \\ 0 & 0 & 2/2 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1/1 \\ \dots & \dots & \dots & \dots \\ 1/1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 2/5 \end{bmatrix}$$

X행렬에서 가구h의 사후층 j(=1, 2, 3, 4)에 모든 구성원 k에 대해서 합계를 계산할 수 있다.

$$U_{h,j} = \sum_k X_{k,j}$$

가구첨자를 이용한 사후층의 행렬은 $U_{h,j} = (U_{h,1}, U_{h,2}, U_{h,3}, U_{h,4})$ 이며, 가구번호 ID(h)=1, 2, 3, 4, 5의 행렬을 다음과 같이 요약 할 수 있다.

$$U_{h,1} = (1,1,0,0) \quad U_{h,2} = (0,0,2,0) \quad U_{h,3} = (0,0,0,1) \quad U_{h,4} = (1,0,0,0) \quad U_{h,5} = (1,1,1,2)$$

가구의 표본단위를 표기하는 첨자를 $i = 1, 2, \dots, n$ 으로 표시하고 단위행렬을 구하면 $Z_{i,j} = \frac{U_{h,j}}{l_h}$ 이며, 여기서 l_h 는 단위 가구의 총가구원수,

$$\text{즉,} \quad l_1 = 2, l_2 = 2, l_3 = 1, l_4 = 1, l_5 = 5 \text{이고}$$

$Z_{1,j}, Z_{2,j} = (1/2, 1/2, 0, 0) \quad Z_{3,j}, Z_{4,j} = (0, 0, 0, 2/2, 2/2) \\ Z_{5,j} = (0, 0, 0, 0, 1/1) \quad Z_{6,j} = (1/1, 0, 0, 0, 0) \quad Z_{7,j}, Z_{8,j}, Z_{9,j}, Z_{10,j}, Z_{11,j} = (1/5, 1/5, 1/5, 2/5)$ 인 행렬Z로 표기 할 수 있다.

선형 모델(linear model)의 방정식을 적용하여 추정 하고자 하는 변수(Y)는 $Y=ZC+E$ 이고 이 모델 방정식으로부터 계수를 추정하면 $\hat{C}=(Z^t\Pi^{-1}TZ)^{-1}Z^t\Pi^{-1}TY$, 이로부터 모수 y의 추정치는 $\hat{y} = \hat{C}x = Y^t\Pi^{-1}TZ(Z^t\Pi^{-1}TZ)^{-1}x = Y^tW$, 여기서 $W=\Pi^{-1}TZ(Z^t\Pi^{-1}TZ)^{-1}x$ $\Pi = \text{diag}(\pi_i) \Pi^{-1} = \text{diag}(\pi_i^{-1}) \quad \pi_i = i$ 가구단위가 표본에 포함될 확률

$$T = \text{diag}(t_i) \quad \text{여기서} \quad t_i = \begin{cases} 1: & \text{if unit } i \text{ is } \in \text{ the sample} \\ 0: & \text{Otherwise} \end{cases}$$

앞에서 소개한 예제자료를 이용하여 가중치($W=\Pi^{-1}TZ(Z^t\Pi^{-1}TZ)^{-1}x$) 계산절차를 살펴보면

$$\Pi^{-1}TZ = \begin{bmatrix} 1/2*5.1 & 1/2*5.1 & 0 & 0 \\ 1/2*5.1 & 1/2*5.1 & 0 & 0 \\ \hline 0 & 0 & 2/2*5.1 & 0 \\ 0 & 0 & 2/2*5.1 & 0 \\ \hline 0 & 0 & 0 & 1/1*5.1 \\ \hline 1/1*4.4 & 0 & 0 & 0 \\ \hline 1/5*4.4 & 1/5*4.4 & 1/5*4.4 & 2/5*4.4 \\ 1/5*4.4 & 1/5*4.4 & 1/5*4.4 & 2/5*4.4 \\ 1/5*4.4 & 1/5*4.4 & 1/5*4.4 & 2/5*4.4 \\ 1/5*4.4 & 1/5*4.4 & 1/5*4.4 & 2/5*4.4 \\ 1/5*4.4 & 1/5*4.4 & 1/5*4.4 & 2/5*4.4 \end{bmatrix} = \begin{bmatrix} 2.55 & 2.55 & 0.00 & 0.00 \\ 2.55 & 2.55 & 0.00 & 0.00 \\ 0.00 & 5.10 & 0.00 & 0.00 \\ 0.00 & 5.10 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 5.10 \\ 4.40 & 0.00 & 0.00 & 0.00 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \end{bmatrix}$$

$$Z' \Pi^{-1}TZ = \begin{bmatrix} 7.83 & 3.43 & 0.88 & 1.76 \\ 3.43 & 3.43 & 0.88 & 1.76 \\ 0.88 & 0.88 & 11.08 & 1.76 \\ 1.76 & 1.76 & 1.76 & 8.62 \end{bmatrix}$$

$$\Pi^{-1}TZ(Z' \Pi^{-1}TZ)^{-1} = \begin{bmatrix} 2.55 & 2.55 & 0.00 & 0.00 \\ 2.55 & 2.55 & 0.00 & 0.00 \\ 0.00 & 5.10 & 0.00 & 0.00 \\ 0.00 & 5.10 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 5.10 \\ 4.40 & 0.00 & 0.00 & 0.00 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \\ 0.88 & 0.88 & 0.88 & 1.76 \end{bmatrix} \begin{bmatrix} 0.22727 & -0.22727 & -0.000000 & -0.000000 \\ -0.22727 & 0.55564 & -0.015947 & -0.06379 \\ 0.00000 & -0.01595 & 0.094052 & -0.01595 \\ 0.00000 & -0.06379 & -0.015947 & 0.13229 \end{bmatrix}$$

$$W = \Pi^{-1}TZ(Z' \Pi^{-1}TZ)^{-1}x = \begin{bmatrix} 0 & 0.83734 & -0.04067 & -0.16266 \\ 0 & 0.83734 & -0.04067 & -0.16266 \\ 0 & -0.08133 & 0.47967 & -0.08133 \\ 0 & -0.08133 & 0.47967 & -0.08133 \\ 0 & -0.32532 & -0.08133 & 0.67468 \\ 1 & -1.00000 & -0.00000 & -0.00000 \\ 0 & 0.16266 & 0.04067 & 0.16266 \\ 0 & 0.16266 & 0.04067 & 0.16266 \\ 0 & 0.16266 & 0.04067 & 0.16266 \\ 0 & 0.16266 & 0.04067 & 0.16266 \\ 0 & 0.16266 & 0.04067 & 0.16266 \\ 0 & 0.16266 & 0.04067 & 0.16266 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \\ 15 \\ 20 \end{bmatrix} = \begin{bmatrix} 4.5102 \\ 4.5102 \\ 4.7551 \\ 4.7551 \\ 9.0203 \\ 10.0000 \\ 5.4898 \\ 5.4898 \\ 5.4898 \\ 5.4898 \\ 5.4898 \\ 5.4898 \end{bmatrix}$$

따라서 동일가구 내 모든 가구원들은 동일한 최종 가중치 $w_{h,j} = ID$ 를 부여 받게 된다.

$$w_{k,l} = \begin{pmatrix} w_{1,2} = 4.5102 \\ w_{1,2} = 4.5102 \\ w_{2,3} = 4.7551 \\ w_{2,4} = 4.7551 \\ w_{3,5} = 9.0203 \\ w_{4,6} = 10.0000 \\ w_{5,7} = 5.4898 \\ w_{5,8} = 5.4898 \\ w_{5,9} = 5.4898 \\ w_{5,10} = 5.4898 \\ w_{5,11} = 5.4898 \end{pmatrix}$$

따새로운 최종 가중치 벡터 W 는 다음 조건을 만족한다. $X^t W = x$

$$X^t W = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4.5102 \\ 4.5102 \\ 4.7551 \\ 4.7551 \\ 9.0203 \\ 10.0000 \\ 5.4898 \\ 5.4898 \\ 5.4898 \\ 5.4898 \\ 5.4898 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \\ 15 \\ 20 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x$$

이는 성별 및 연령그룹의 추정치는 벤치마크와 일치함을 증명한다.

앞에서 기술한 예제의 통합가중치 계산을 위한 SAS 프로그램을 소개하면 다음과 같다.

```

data
popn;

input hang $ sex $ageg
popn;
cards;

A 1 1 20
A 1 2 10
A 2 1 15
A 2 2 20
;
run;

data sam;

```

```

input Hang $ ID $sex $ ageg hhwgt;
cards;

A 1 1 1 5.1
A 1 1 2 5.1
A 2 2 1 5.1
A 2 2 1 5.1
A 3 2 2 5.1
A 4 1 1 4.4
A 5 1 1 4.4
A 5 1 2 4.4
A 5 2 1 4.4
A 5 2 2 4.4
A 5 2 2 4.4
;
proc sort data=sam;by hang sex ageg;
proc summarydata=sam nway;
  class hang sex ageg; var hhwgt;
  output out=a(rename=( _freq_=samsize)drop=_type_) sum=sumhhwgt;
data postfct; mergepopn a; by hang sex ageg;
  postfact=popn/sumhhwgt;
  dirwgt=popn/samsize;
data sam; merge sam postfct; by hang sex ageg;
  postwgt=hhwgt*postfact;
run;

%let agecat=2;
%let COLS=%eval( 2* &agecat );

proc sort data=sam;by hang id sex ageg;run;

data XMATRIX (drop= | sex ageg COLNUM);
set sam (keep = ID hang ageg sex hhwgt);
array COL(&cols);
if hhwgt>0 then do;
do I = 1 to &cols;
  col(I) = 0;
end;
select(SEX);
  when( '1' ) colnum = AGEg;
  when( '2' ) colnum = (&agecat + AGEg);
  otherwise;
end;
  COL(colnum) = 1;
output;
end;
run;
proc print noobs data=xmatrix;
run;

proc summary nway data = XMATRIX;
  class hang ID;
  var COL1-COL&cols;
  output out = Ujs (rename =( _FREQ_= Persons)drop=_TYPE_) sum = U1-U&cols;

```

```

run;
proc print data=ujs noobs;
run;

data PZMATRIX (drop= U1-U&cols I);
  merge Ujs XMATRIX;
  by hang ID;
  array U(&cols);
  array COL(&cols);
  array PCOL(&cols);
  length persons 3;
  do I = 1 to &cols;
    COL(I) = U(I) / persons;
    PCOL(I) = COL(I) * hhwgt;
  end;
run;

dataZ(keep=col1-col&cols)
  PZ(keep=pcol1-pcol&cols);
  set pzmatrix;
run;
data pop; set popn(keep=popn);
run;
proc print data=z noobs;
proc print data=pz noobs;
proc print data=pop noobs;
run;

PROC IML;
  use Z var _ALL_;
  read ALL into bigZ;
  use PZ var _ALL_;
  read ALL into PbigZ;
  use pop var _ALL_;
  read all into x;
  Wiw=PbigZ*inv(t(bigZ)*PbigZ)*x;
  create Wiw from Wiw[colname='Wiw'];
  append from Wiw;
  close Wiw;
quit;

datagregwgt(rename=(wiw=gregwgt));
  merge sam wiw;
run;
proc print data=gregwgt;

** check;
proc summary data=gregwgt nway;
  class hang sex ageg;
  var dirwgt postwgt gregwgt;
  id popn;
  output out=b(drop=_type__freq_)sum=;
proc print data=b noobs;
run;

```

7. 사후층화 방법과 통합가중치방법의 비교검토

가. 모의실험

모의실험자료: 2010년 인구주택총조사자료를 2015년 조사구 설정 개념에 일치하도록 평균 가구수를 60가구에서 30가구로 조정하는 조사구 분할을 실시한 충청북도 자료를 이용하였다.

모의실험방법: 2010년 인구주택총조사 충청북도 원시자료를 모집단으로 하고 정의된 모집단으로부터 추정 항목들의 모수를 계산하고, 정의된 모집단에서 20% 조사구를 30회에 걸친 표본 조사구를 무작위 추출에 의해 진행하였으며, 각 표본으로부터 사후층화 및 통합가중치 방법의 추정치를 모집단 모수와 비교하여 두 방법들 간의 효율성을 비교하였다. 여기서 가중치 산정방법은 무응답 조정을 제외한 모든 절차를 수행하여 조사결과를 추정하는 방법으로 진행하였다.

모집단 정의: 전수항목 비교를 위한 모집단은 읍면동으로 하고, 표본항목은 2010년 인구주택총조사에서 10%의 표본으로 조사되었으므로 읍면동을 모집단으로 하여 다시 20%의 표본을 추출한다면 표본의 대표성 문제가 제기된다. 따라서 표본항목 비교를 위한 모집단은 시군구로 확대하여 모집단으로 정의하여 표본 조사구를 추출하도록 하였다.

추정항목: 추정항목으로는 전수조사 항목에 포함된 교육수준, 혼인상태, 가구주와의 관계를 선정하였고, 표본조사 항목으로는 사례수가 충분한 취업자, 실업자, 아동보육(부모), 활동제약(없음), 고령자생활비원천(부부직업) 등을 선정하였다.

나. 모의실험 분석결과

방법간 효율성 비교는 전수항목에서는 읍면동별 상대 표본오차를 계산하여, 각 모집단별로 RSE_1 와 RSE_2 를 비교하여 효율성을 비교하였다.

$$\text{즉, } RSE_i = 100 * \frac{\sqrt{\sum_{j=1}^{30} (\theta - \hat{\theta}_j)^2 / 30}}{\theta}$$

여기서 $i = (1=\text{사후층화방법}, 2=\text{통합가중치방법})$ $\theta = \text{모수}$ $\hat{\theta} = \text{추정치}$

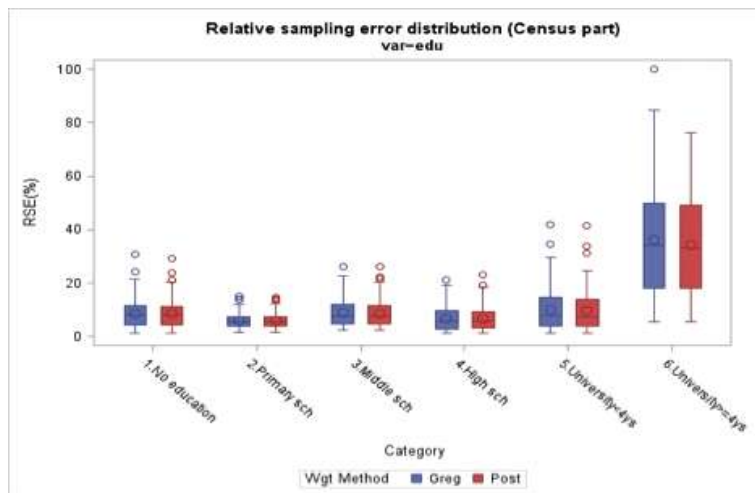
분석결과 교육정도(<표 2-25>, [그림 2-3])는 모든 카테고리가 사후 층화방법이 통합가중치 방법보다 효과적인 것으로 분석되었다.

<표 2-25> 교육정도의 방법별 결과비교

교육정도	효과적인 읍면동수			총 읍면동수
	사후층화 ($RSE_1 < RSE_2$)	통합가중치 ($RSE_1 > RSE_2$)	동일결과 ($RSE_1 = RSE_2$)	
안받았음(미취학포함)	87	67	.	154
초등학교 (졸업, 재학, 중퇴)	93	60	1	154
중학교 (졸업, 재학, 중퇴)	93	61	.	154
고등학교 (졸업, 재학, 중퇴)	87	67	.	154
대학교4년제미만(졸업, 재학, 수료, 휴학, 중퇴)	86	68	.	154
대학교4년제이상*(졸업, 재학, 수료, 휴학, 중퇴)	102	51	1	154

*대학원 포함

[그림 2-3] 교육정도의 카테고리별 RSE(%)의 분포

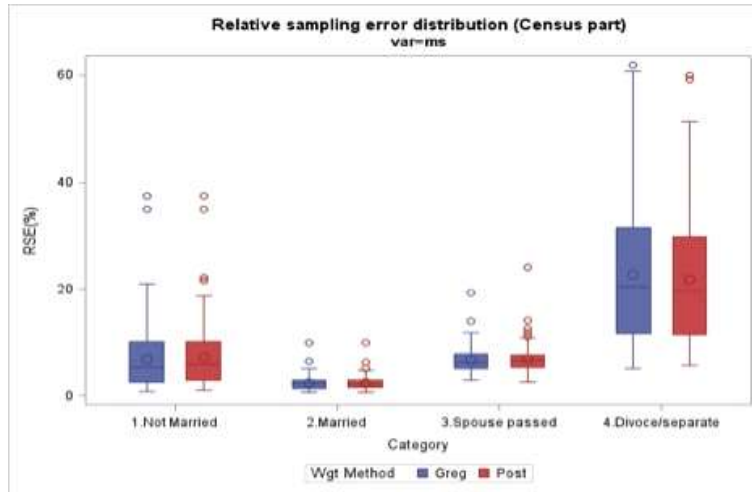


혼인상태(<표 2-26>, [그림 2-4])는 미혼, 배우자 있음, 사별의 카테고리는 통합가중치방법이 더 효과적이며, 이혼, 별거의 카테고리는 사후층화 방법이 더 효과적인 것으로 분석 되었다.

<표 2-26> 혼인상태의 방법별 결과비교

혼인상태	효과적인 읍면동수			총 읍면동수
	사후층화 ($RSE_1 < RSE_2$)	통합가중치 ($RSE_1 > RSE_2$)	동일결과 ($RSE_1 = RSE_2$)	
미혼	37	117	.	154
배우자 있음	71	83	.	154
사별	58	95	1	154
이혼, 별거	112	42	.	154

[그림 2-4] 혼인상태의카테고리별 RSE(%)의 분포

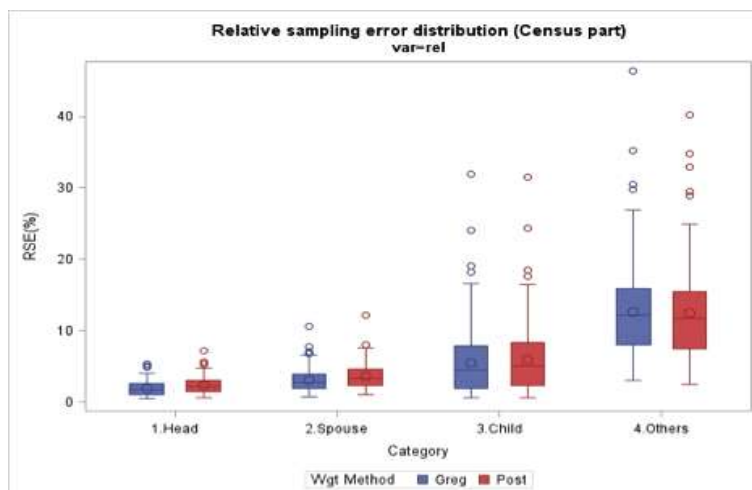


가구주의 관계(<표 2-27>, [그림 2-5])는 가구주, 배우자, 자녀의 카테고리는 통합 가중치 방법이 효과적인 방법으로 분석되었고, 기타 카테고리는 사후층화 방법이 효과적인 것으로 분석되었다.

<표 2-27> 가구주관계의 방법별 결과비교

가구주와의 관계	효과적인 읍면동수			총 읍면동수
	사후층화 ($RSE_1 < RSE_2$)	통합가중치 ($RSE_1 > RSE_2$)	동일결과 ($RSE_1 = RSE_2$)	
가구주	19	135	.	154
배우자	18	136	.	154
자녀	33	121	.	154
기타	88	66	.	154

[그림 2-5] 가구주관계의 카테고리별 RSE(%)의 분포

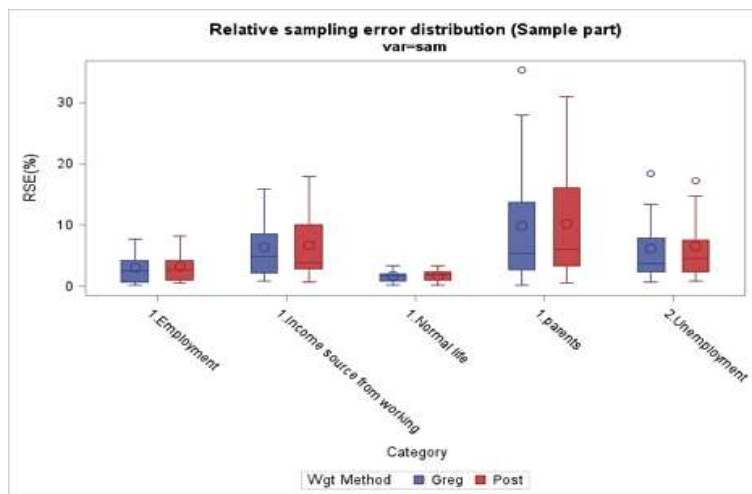


표본항목(<표 2-28>, [그림 2-7])은 모든 카테고리에서 통합가중치 방법이 사후
 층화 방법보다 효과적인 것으로 드러났다.

<표 2-28> 표본항목의 방법별 결과비교

교육정도	효과적인 읍면동수			총 읍면동수
	사후층화 ($RSE_1 < RSE_2$)	통합가중치 ($RSE_1 > RSE_2$)	동일결과 ($RSE_1 = RSE_2$)	
경제활동(취업)	3	10	.	13
경제활동(실업)	3	10	.	13
고령자생활비원천 (본인.배우자의 일, 직업)	6	7	.	13
활동제약(없음)	2	11	.	13
아동보육(부모)	6	7	.	13

[그림 2-6] 표본항목의카테고리별 RSE(%)의 분포



8. 가중치 계산 프로그램

본 장은 실제적인 시험조사 자료의 변수를 이용하여 제안한 사후층화 가중치
 계산 절차 및 사후층화 변수 카테고리 자동통합방법을 위해 알고리즘을 적용하여
 정형화된 SAS 매크로 프로그램을 작성하였다. 이러한 정형화된 프로그램을 2015
 인구주택총조사에서 직접 이용할 수 있도록 제시하였다.

SAS dataset and variables

다음은 반드시 준비하여야 할 SAS dataset와 필요한 변수들(변수명은 시험조사
 의 변수명을 그대로 이용하여 작성하였다)은 다음과 같다.

SAS datasets	Variables
PopED(표본틀)	Hang(행정구역) ED_NUM(조사구) ED_PRPT_CD(조사구특성) HH(가구수)
HHlist(가구명부)	Hang(행정구역) ID* UN_SURV_RSN_CD(조사상태) UH_YN(빈집여부)
Household(총조사 가구부문)	Hang(행정구역) ID*
HHmember(총조사 인구부문)	Hang(행정구역) ID* HM_SN(가구원번호) c2(성별) age(나이)
Dwelling(총조사 주택부문)	Hang(행정구역) ED_NUM(조사구) ED_PRPT_CD(조사구특성) HU_NUM(거처번호) c1(type:거처종류)
PopHousehold(벤치마킹 가구부문)	Hang(행정구역) HHsize(가구규모)** PopHH(가구수)
PopPerson(벤치마킹 인구부문)	Hang(행정구역) Sex(성별) Ageg(년령규모)*** PopPS(인구수)
PopDwelling(벤치마킹 주택부문)	Hang(행정구역) Type (거처종류) PopDW(주택수)

*)ID=HHD_LTNM_ID(가구명부ID) ED_NUM(조사구) ED_PRPT_CD(조사구특성) HU_NUM(거처번호) HHD_NUM(가구번호). **)인구부문으로부터 도출된 변수. ***)

SAS macro(SoCenWgt*) module parameters

```
%SoCenWgt(AdjNonR=N,
  PopED=data.poped,
  HHlist=data.hhlist,
  /* survey data */
  Household=data.householdimp,
  HHmember=data.hhmemberimp,
  Dwelling=data.dwellingimp,
  /* Benchmark */
  PopHousehold=data.benchmarkhousehold,
  PopPerson=data.benchmarkperson,
  PopDwelling=data.benchmarkdwelling,
  /* strat collapsing cut point */
  cuthhsize=5,
  cutageg=5,
  cuttype=5
);
```

Note: * SoCenWgt (System of CensusWeights)

Parameterspecification

- AdjNonR: 무응답을 가중치 계산절차에 포함여부 (Y/N)
- PopED: 표본 조사구 추출에 사용한 표본틀

- HHlist: 가구명부 파일
- Household: 가구분야 조사자료
- HHmember: 인구분야 조사자료
- Dwelling: 주택분야 조사자료

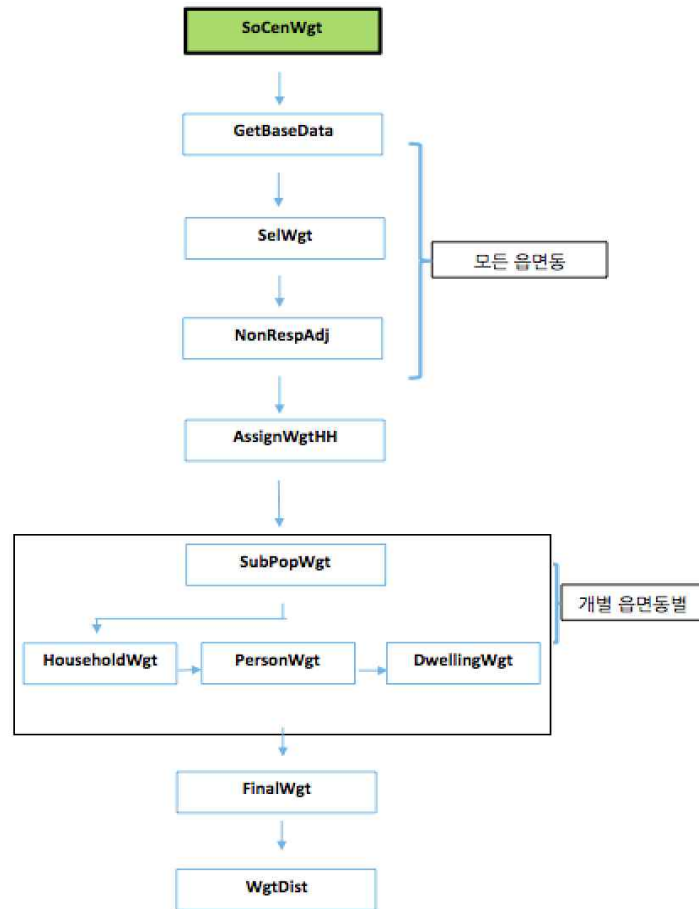
- PopHousehold: 가구 벤치마크 자료
- PopPerson: 인구부문 벤치마크 자료
- PopDwelling: 주택부문 벤치마크 자료

- Cuthhsize: 가구부문 층화변수 최저 추출률
- Cutageg: 인구부문 층화변수 최저 추출률
- Cuttype: 주택부문 층화변수 최저 추출률

%SoCenWgt submodules and processing flow

- %GetBaseData: Dataset preparation
- %SelWgt: Selection weight calculation
- %NonRespAdj: Non response adjustment
- %AssignWgtHH: Allocation of adjusted weight to household dataset
- %SubPopWgt: Eupmyeon dong processing
- %HouseholdWgt: Household weight calculation
- %PersonWgt: Person weight calculation
- %DwellingWgt: Dwelling weight calculation
- %FinalWgt: Final weights including integer weight
- %WgtDist: Weight distribution

[그림 2-7] Module processing flow



Outputs

All weight calculation step outputs

- This can be used for understanding of weight calculation steps

- CumHHWgt: Household weight
- CumPSWgt: Person weight
- CumDWWgt: Dwelling weight

Final weights including integer weight

- This can be used for estimation

- FinalCumHHWgt: Household weight
- FinalCumPSWgt: Person weight
- FinalCumDWWgt: Dwelling weight

SAS programs

- See appendix

제3장 순환표본조사 시행을 위한 표본설계 및 추정방안 연구

제1절 서론

센서스 혹은 인구주택총조사는 기준 시점의 인구, 가구 그리고 주택과 관련된 변수들을 이용하여 한 국가의 구조에 대한 일종의 스냅샷(snapshot)을 찍는다. 즉 센서스의 정의에 있어 중요한 요소는 “시점”과 “거주”기준이다. 국내에서 매 5년 마다 전국의 모든 가구 및 가구원의 분포를 거주지를 기준으로 파악하기 위하여 수행되었던 전통적인 인구주택총조사는 2015년부터 주민등록자료와 건축물 대장과 같은 행정자료를 활용한 등록센서스로 전환되었다. 등록센서스로의 전환은 기본적으로 거주 기준이 실제 거주지에서 행정 자료상의 거주지로 바뀌었음을 의미한다. 또한 조사 시점의 측면에서 보면 기존 인구주택총조사가 매 5년 마다 특정 시점의 인구, 가구 그리고 주택의 분포가 파악되었다면 등록센서스를 통해서는 행정자료를 사용함으로써 등록 기준으로 시의성을 갖춘 자료를 매 해 혹은 더 짧은 주기로 제공할 수 있게 되었다.

전통적인 센서스는 일반적으로 두 부분으로 구성되어 있다. 그 첫 번째는 국가 내 거주하는 모든 거주자 혹은 이들의 거주지 분포를 파악하기 위하여 모집단 전체를 조사하는 전수조사이며, 두 번째 부분은 이 중 일부를 추출하여 조사하는 표본조사이다. 실제로 센서스의 의미는 전자의 경우에 해당한다. 전수조사를 위해서는 상대적으로 짧은 설문문항으로 구성된 설문지를 사용하며 표본조사를 위한 설문지는 상대적으로 많은 설문항목을 포함하고 있다. 국내에서는 전수조사를 위한 설문항목을 전수항목(short-form survey item)으로, 표본조사를 위한 항목을 표본항목(long-form survey item)으로 구분하여 명명하고 있다. 국내에서 표본조사 대상자와 그 결과는 통계청에 의하여 별도로 관리되고 있으며 이들로부터 작성된 리스트와 정보는 경제활동인구조사와 같은 통계청 주관 조사들을 위한 표집틀로도 사용되고 있다.

조사비용의 증가 및 도시화에 따른 급격한 유동인구 증가로 인한 지역 인구, 가구 및 주택 분포의 변화 등으로 인하여 대부분의 센서스를 수행하던 국가에서는 전통적인 방식을 탈피하여 새로운 센서스를 지속적으로 도입하는 노력을 수행해왔고 2000년대에 들어와서는 점차적으로 새로운 센서스를 도입하여 실행하는 국가들이

증가하고 있다. 다음의 <표 3-1>은 2000년에서 2020년 사이에 센서스를 실시한 혹은 실시할 국가들이 사용했거나 사용할 예정인 센서스 방식을 나타내고 있다. 여전히 전통적인 센서스 방식을 유지하고 있는 국가들도 있으나 많은 국가들이 새로운 형태, 특별히 등록 자료를 이용한 센서스 방식으로의 전환을 하고 있는 것을 확인할 수 있다. 기본적으로 이러한 변화를 위해서는 안정적인 등록 자료가 존재하고 사용 가능해야 한다. 프랑스나 영국과 같이 이러한 등록 자료가 없는 국가에서는 다른 형태의 센서스 방법을 모색하여 시행하고 있다.

<표 3-1> 2000-2020년 센서스 라운드의 국가 별 센서스 방안

센서스 방안	국가 및 센서스 라운드
전수집계 (+ 표본조사)	한국(2010), 영국(2011), 캐나다(2011), 스위스(2000)
등록부 자료 + 전수집계	스페인(2001), 포르투갈(2001, 2011), 체코(2010), 발트 3국(2011)
등록부 자료(인구) + 전수집계(주택) + 표본조사	독일(2011)
등록부 자료 + 표본조사	한국(2015), 스페인(2011), 폴란드(2011), 슬로베니아(2001), 이스라엘(2008), 스위스(2010), 싱가포르(2000, 2010), 터키(2011), 네덜란드(2001, 2011), 벨기에(2011), 아이슬란드(2011)
완전 등록센서스	덴마크(2001), 핀란드(2001), 안도라(2001), 덴마크(2011), 노르웨이(2011), 슬로베니아(2011), 스웨덴(2011)

국내에서 진행되고 있는 등록센서스로의 전환은 기존 인구주택총조사와 함께 동일한 시점에 수행되던 기존 10% 표본조사의 변화도 요구하게 되었고, 2015년에는 자료의 안정성 검토 및 이와 관련한 추가 연구를 위하여 그 규모를 20%로 확대하였다. 기존의 10% 표본조사를 포함한 인구주택총조사는 특정 시점, 예를 들어, 2015년 11월 1일 0시 기준 대한민국 영토 내 거주 혹은 존재하는 인구, 가구 그리고 주택의 현황을 파악하기 위한 조사이며, 다른 가구조사의 표본설계를 위한 조사구 표집틀(sampling frame)로 활용된다. 통계청은 새로운 인구주택표본조사의 수행을 위해서 기존 평균 60가구로 구성된 조사구를 평균 30가구 규모로 정의하였고 새

롭게 정의된 조사구로 구축된 표집틀로부터 2015년 20% 표본조사를 위한 조사구 표본을 추출하였다. 조사구 내 가구 수의 감소로 인하여 조사구 내 가구 추출은 고려하지 않고 추출된 조사구 내 가구는 모두 조사대상 가구로 표본에 포함시켰다.

행정자료를 이용한 등록센서스로의 전환은 기본적인 인구, 가구 그리고 주택 분야 통계의 각 년도 별 산출을 가능하게 할 것이나 현행의 표본조사 주기 및 구조를 유지할 경우 표본조사 항목에 대한 통계는 여전히 5년의 주기로 작성될 것이다. 또한 20% 규모의 표본을 일시에 모두 조사할 경우에 발생하는 조사비용 대비 효율성은 일정 규모 이상의 조사를 매 해 수행하는 것에 비하여 낮을 수 있다는 문제가 제기되고 있다. 이에 통계청은 조사 상의 편의성, 조사비용의 효율적인 배분, 조사 인력의 상시 채용 그리고 시의성을 갖는 통계 생산 및 적절한 부가 항목의 추가 가능성을 고려하여 인구주택총조사의 중요한 개념인 “시점”에 대한 개념을 포기하고 대신 순환표본조사(rolling sample survey)를 실행하는 방안을 지난 몇 해 동안 지속적으로 검토해 왔다.

순환표본조사는 기본적으로 추출된 표본을 서로 겹치지 않은 부표본으로 분할한 후 분할된 각 부표본을 일정 주기를 두고 조사하는 방식의 조사를 의미한다. 혹은 순환 주기가 정해진 경우에는 매 조사 시점 별로 독립적인 표본을 추출하는 방안 역시 고려될 수 있다. 만약 분할 대상이 표본이 아닌 모집단인 경우에는 이를 흔히 순환센서스(rolling census)로 명명하기도 한다. 현재 국내에서 고려하고 있는 방안은 5년 단위로 20%의 표본을 추출하고 이를 60개월(5년)동안 매 월 조사하는 순환표본조사이다. 순환표본조사가 적용되는 대표적인 국내의 기존 조사로는 국민건강영양조사가 있다.¹⁾ 국민건강영양조사를 위한 표본은 매 3년마다 추출되며 이를 이용하여 매 년 전국단위 통계가 제공되고 있으며 3년 혹은 그 이상의 기간 자료를 통합하여 광역시도별 통계를 생산하고자 한다.

순환센서스 혹은 순환표본조사가 수행되고 있는 대표적인 국가로는 프랑스와 미국이 있다. 프랑스는 순환센서스를, 그리고 미국은 순환표본조사를 각각 사용한다. 프랑스의 경우에는 전통적인 센서스의 전수조사 부분을 대체하기 위한 방안으로 순환조사를 사용하고 있고 미국의 경우에는 기존의 표본조사 부분을 대체하기 위한 방안으로 순환조사를 도입하여 사용한다.

프랑스는 1801년부터 1999년까지 7년 단위로 전통적인 센서스를 수행해 왔으나 1997년에 예산상의 문제로 센서스를 수행하지 못함에 따라 전통적인 센서스의 대체 방안에 대한 연구를 시작했다. 프랑스의 경우, 모집단에 대한 등록 자료가 없기 때

1) 국민건강영양조사는 설문조사와 더불어 검진조사가 함께 이루어지고 검진조사를 위한 장비의 한계가 순환표본조사 도입의 하나의 이유이다.

문에 등록 자료기반 센서스를 수행할 수 없으며 1년의 단기간 예산 역시 제한되어 있어 순환센서스로 전환하였다. 프랑스의 순환센서스는 기본적으로 5년 동안 모집단 전체(실제로는 인구의 70%)를 포함하는 방안을 취하고 있으며 매 해 중·대 규모 지역의 인구 분포를 제공하기 위한 목적을 가진다. 이를 위하여 5년 동안 조사되는 지역 및 조사 단위의 균형을 맞추기 위한 표본설계 방안(실제로는 모집단 분할)이 사용되고 있고, 또한 매 년 수행되는 조사는 전통적인 센서스 방식의 축소판이라 할 수 있다. 즉, 매 해 조사는 특정 시점 기준으로 단기간 동안 한 번 이루어지며, 따라서 그 목적은 기준 시점 프랑스 인구 분포에 대한 스냅샷(snapshot)을 제공하는 것이다.

미국은 전통적인 센서스의 주기가 10년으로 다른 국가보다 그 주기가 길다. 따라서 미국의 센서스는 작성 통계의 시의성 문제에 직면해 왔다. 이를 극복하기 위한 대안으로 1994년 Census Bureau는 가구 및 인구 특성의 지속적인 측정을 위한 American Community Survey(ACS)를 활용하는 방안에 대한 연구를 시작했다. 2004년까지의 연구 및 시범 조사 등을 통해 순환표본조사를 통한 ACS의 기존 센서스 표본조사 대체 가능성을 확인하고 2005년 1월부터 매월 25만 가구(년 300만 가구, 약 2.5%)를 조사하고 있다. 표본추출은 기본적으로 매 월 독립적으로 주소 기반 표집틀로부터 이루어지며 이를 기반으로 각 년도 별로 미국 전체에 대한 표본조사 항목의 통계를 생산하고 있다.

한국의 2015년 인구주택총조사의 기존 전수조사 부분은 등록센서스로 대체되어 수행되고 있으며 표본조사 부분은 평균 30가구 기준으로 작성된 조사구 표집틀을 이용하여 20% 규모의 표본조사로 수행되고 있다. 즉 전수부분의 경우 새로운 센서스 방안이 적용되어 진행되고 있으나 표본조사는 예전과 동일하게 특정 시점 기준 그리고 등록 자료 기준이 아닌 거주 기준의 조사가 이루어지고 있는 것이다. 통계청에서는 전통적인 센서스의 틀 안에서 이루어지던 표본조사의 틀 역시 연구를 통해서 수정하려 계획하고 있고 그 방향은 미국에서 시행하고 있는 순환표본조사로 잡고 있다(한국인구학회, 2014).

본 장에서는 한국에서의 순환표본조사 실시를 위한 표본설계, 이를 이용하여 산출할 수 있는 가능한 추정량의 정의, 추정 방식 등을 살펴보기로 한다. 이를 위해 2절에서는 순환표본조사를 위한 표본설계를 우선 살펴본다. 프랑스와 미국에서 사용되고 있는 표본설계 방안의 특징을 살펴보고 이의 적용 가능성을 검토한다. 또한 기존 인구주택총조사의 표본설계 틀을 유지하면서 적절한 수준에서의 통계를 매 해 생산할 수 있는 방안을 모색한다. 이어 3절에서는 기간추정(Multi-year estimation or estimation for period)에 대한 정의를 하고 이를 위한 추정 및 가중치 산출 방안

을 검토한다. 마지막으로 고려된 표본설계 및 추정방안의 장단점과 필요한 추가 연구에 대한 내용을 소결에서 기술한다.

제2절 순환표본조사를 위한 표본설계

1. 표본설계의 기본원칙

순환표본조사와 일반적인 표본조사와의 가장 큰 차이는 그 목적에 있다. 일반적인 국가승인통계 작성을 위한 표본조사의 목적은 특정 조사 시점 및 설계변수(design variables) 혹은 관심변수(variables of interest)를 기준으로 모집단을 대표하는 표본을 추출하고 이에 근거하여 통계적으로 타당한 추정량을 산출하는 것에 있다. 따라서 표본설계의 중요한 요소인 층화, 집락화 그리고 표본크기의 결정이 이러한 목적에 맞게 결정된다.

순환표본조사를 위한 표본설계 역시 기본적으로는 동일한 기준으로 표본설계가 이루어지나 이에 추가하여 순환표본조사에서는 여러 시점동안 축적된 자료를 이용한 추정 방안 역시 고려되어야 한다. 즉 순환표본조사의 목적은 각 조사 시점 별 안정적인 모집단 전체 혹은 대규모의 부모집단 추정량을 산출함과 동시에 여러 시점에 걸쳐 조사된 자료를 통합하여 상대적으로 규모가 작은 부모집단에 대한 추정을 하는 것이다. 예를 들어, 순환표본조사를 국내 인구주택총조사의 표본조사를 위하여 사용할 경우를 고려해 보자. 이 때 전국 혹은 16(7)개 광역시도 통계의 추정은 매 해 이루지고 시군구 통계는 3년 내지 5년 동안 축적된 조사 자료를 이용하여 안정적으로 생산할 수 있을 것이다.

따라서 순환표본조사를 위한 표본설계의 기본원칙은 크게 두 가지로 요약할 수 있다. 첫째는 각 순환표본조사주기 별로 전국 단위 혹은 대규모 부모집단에 대한 안정적인 통계를 산출할 수 있는 설계이어야 한다는 것이다. 둘째는 여러 시점에 조사된 자료를 누적하여 상대적으로 소규모인 부모집단들의 안정적인 통계를 산출할 수 있는 설계이어야 한다는 것이다.

2. 해외 표본설계 사례 연구

본 장에서는 순환표본조사 혹은 순환센서스가 수행되고 있는 국가들에서 적용되고 있는 표본설계 방안들을 검토하고 국내 순환표본조사에 적용할 수 있는가를 살펴본다.

가. 프랑스의 순환센서스

서론에서 언급한 바와 같이 프랑스의 순환센서스는 미국의 ACS가 센서스의 일부인 표본조사를 대체하기 위하여 순환표본조사를 도입한 것과는 달리 전수조사 자체를 대체하고 있다. 따라서 프랑스 순환센서스의 목적은 전통적인 센서스의 목적과 일치한다. 즉, 특정 시점에 프랑스 내 거주하는 개인 혹은 가구 정보를 취합하여 모집단의 인구 및 가구 분포를 추정하는 것이 주요 목적이다. 다만 전통적인 방법이 모집단 전체를 관측하여 그 목적을 달성하는 것과는 달리 프랑스는 이를 표본조사를 통해 달성하는 방안을 취하고 있다. 프랑스 순환표본조사의 주기는 5년이며 조사는 매 해 이루어진다. 즉 매 해 정해진 특정 시점에 일종의 미니 센서스(mini Census)로 진행된다.

프랑스 행정구역은 총 36,682개의 기초 행정단위인 코뮌(communes)으로 구성되어 있다. 각 코뮌은 그 인구 규모가 매우 다양하다. 1999년 센서스에 의하면, 반 이상의 인구가 크기 10,000명 이하의 36,000 여 개의 코뮌에 거주하고 있으며 나머지 반은 그 크기가 100,000 이상인 1,000 여 개의 코뮌에 거주하고 있다(Durr and Clanché, 2013). 따라서 5년 순환주기의 각 년도 별 표본조사 대상을 결정하기 위해서는 추출단위인 코뮌의 크기(인구 수 기준)가 고려된다. 프랑스 통계청은 각 년도 별 조사 대상 코뮌을 결정하기 위하여 일종의 균형표본추출법(balanced sampling)을 사용하였다.

균형표본추출법은 모형 기반 서베이 분석가들 중 대표적으로 Royall(1992), Royall and Herson(1973) 그리고 Valliant, Dorfman and Royall(2000)에 의하여 제시된 방안으로 주어진 선형모형 하에서 최량선형불편예측치(Best Linear Unbiased Predictor: BLUP)의 분산을 최소화하는 표본추출방안으로 제시되었다. 또한, 이를 만족하는 표본을 추출하기 위한 추출방안으로는 제한적 랜덤추출법(restrictive random sampling)이 제안되었다. Park(2006)은 제한적 랜덤추출법에 의하여 구현된 균형표본의 확률적 구조를 유도하고 이를 바탕으로 BLUP이 단순임의추출 하에서 정의되는 회귀추정량과 근사적으로 일치(asymptotic equivalent)함을 증명하였다.

프랑스 통계청은 순환표본조사를 위한 각 년도별 조사 대상 표본을 선정하고자

전체 코뮌을 균형표본추출법을 활용하여 5개의 그룹으로 나누었다. 균형표본추출법의 적용을 위해서 전체 모집단을 먼저 규모가 큰 코뮌으로 구성된 코뮌 부모집단(U_L)과 상대적으로 규모가 작은 코뮌으로 구성된 코뮌 부모집단(U_S)으로 나누었다. U_S 는 지역(22개 지역), 연령 그룹(5개), 성별 그리고 주거 형태를 기준으로 서로 동일한 5개의 그룹으로 균형표본추출법을 통해 분할되었다.

U_L 은 균형표본추출법이 적용되기에 앞서 먼저 주소(address) 정보를 기반으로 총 4개의 층으로 재분류 되었다. 그 첫 번째 층은 코뮌 기관(communal establishment)으로 구성되었으며, 두 번째 층은 10가구 이상이 거주할 수 있는 공동주택으로 구성되어 있으며, 세 번째 층은 일반 단독 가구로 구성되어 있다. 마지막 네 번째 층은 신규 주소로 구성되어 있다. 각 층 별로 균형표본추출법에 의하여 층 내 주택 혹은 가구들은 5개의 그룹으로 분할되었고 4개의 각 층으로부터 분할된 5개의 그룹 및 U_S 로부터 주어지는 5개 그룹을 이용하여 매 년 조사될 5개의 독립적이 그룹이 생성된다. 기본적으로 5년의 주기 동안 전체 모집단이 포함되는 것이 원칙이나 U_L 의 단독가구 층으로부터는 일부만을 추출하여 약 70%의 주소지가 5년 동안 조사되도록 설계된 것이다.

프랑스의 순환센서스 표본설계는 전통적인 표본설계와는 다르다. 오히려 이는 5년 동안 매 해 균형적인 조사를 수행하기 위하여 조사 대상(주소지 혹은 코뮌)을 5개 그룹으로 나누는(splitting) 방안으로 이해된다. 프랑스 통계청은 매 해 36,682개 코뮌 및 U_L 에 속한 코뮌의 일부 부모집단 통계를 작성하여 공표하고 있다.

나. 미국 순환표본조사: ACS

미국 ACS는 전통적인 센서스에서 시행되던 표본조사보다 효율적인 대안으로 알려져 있다. 따라서 국내의 인구주택총조사의 일부로 수행되던 표본조사의 대체 방안으로 검토할 수 있는 최적의 선택이라고 판단된다.

미국의 ACS의 조사는 매 월 이루어지고 있으며 표본 추출 역시 매 월 독립적으로 이루어진다. ACS를 위하여 사용되는 표집틀은 Census Bureau의 대부분의 조사에서 주로 사용되는 Master Address File(MAF)로써, MAF는 미국과 푸에르토리코 내의 주택(housing unit), 그룹쿼터(group quarter: GQ²), 그리고 일부 비거주용 건물의 목록을 의미한다. MAF는 상시 갱신되는 자료이며 미국 내의 지리정보 시스템인 Topologically Integrated Graphic Encoding and Referencing(TIGER³)과 연동

2) Group Quarters (GQ) are places where people live or stay, in a group living arrangement, which is owned or managed by an entity or organization providing housing and/or services for the residents.

되어 사용된다. ACS의 조사방식은 기본적으로 우편조사이며 추가적으로 인터넷 및 CATI(computer assisted telephone interview)가 사용된다. 우편이 도달하지 못하는 주소지나 무응답자 중 추출된 가구 조사에 대해서는 CAPI(computer assisted telephone interview)가 사용된다.

본 장에서는 MAF 주택부분의 표본추출 방안을 간략하게 서술한다.⁴⁾ 기본적으로 ACS는 그 주기가 5년이며 한 주소가 5년 동안 1번을 초과하여 조사될 수 없도록 구성되어 있다. 따라서 표본추출을 위해서는 우선 표집틀인 MAF를 서로 겹치지 않는 5개의 부표집틀로 분할하게 된다. 각 부표집틀은 전체 주소의 약 20%를 차지하며, 각 표집틀은 모집단 대표성을 갖도록 작성된다. 이 과정에서 생성되는 5개의 부표집틀은 프랑스의 순환센서스의 각 순환그룹과 비슷한 성격을 갖는다.

각 부표집틀로부터 당 해 조사해야 하는 표본 주소들이 추출된다. 사전에 산출되는 추출률에 따라 조사 대상 주소는 각 카운티(county) 별로 추출되며 이는 다시 실제 조사가 이루어질 12개월에 배분된다. 추출률의 산정은 공표 수준을 고려하여 이루어진다. 만약 공표 수준에 해당하는 지역 간의 인구나 가구 기준 규모가 크게 차이가 나는 경우에는 규모가 작은 지역에 상대적으로 큰 추출률을 배정하는 방안을 적용하여 지역 간 표본 크기의 차이를 최소화하며 공표 수준에 합당한 안정적인 통계가 산출되도록 하였다. 표본 추출은 기본적으로 계통추출법이 적용되었다. 그룹 쿼터나 비거주용 건물 주소에 대한 추출 역시 비슷한 절차에 따라 이루어졌다.

월 별 표본주소지 수는 약 250,000개이며 연간 표본 수는 약 3,000,000개의 주소이다. 조사 결과의 공표는 지역 규모에 따라 1년, 3년 혹은 5년 간 수집된 자료를 활용하여 이루어진다. 그 규모가 65,000명 이상인 지역에 대해서는 매 해 추정량이 산출되며, 20,000명 이상 그리고 65,000명 미만 규모의 지역은 3년 단위로, 마지막으로 전체 지역 단위(시 단위까지) 별 통계는 5년 단위로 작성하여 공표된다. 추정량 산출을 위해서는 기본적으로 각 개인 혹은 가구에 부여되는 가중치를 산출하고 이를 각 관심변수에 적용하는 선형추정법을 사용한다. 기본적으로 가중치 산출 방안은 가구 조사 분석을 위해 사용되는 일반적인 방법, 즉 표본설계 가중치 작성, 무응답 조정 그리고 벤치마킹의 3단계에 걸쳐 이루어진다. 다만 우편조사 및 1차 무응답자들에 대한 재접촉 후 측정되는 자료는 조사시점과는 차이가 발생할 수 있어 이를 고려한 보정 방안이 가중치 산출을 위해 사용된다.⁵⁾

3) digital database of geographic features, such as roads, railroads, rivers, lakes, legal boundaries, census statistical boundaries, etc. covering the entire United States.

4) 표본추출에 대한 기술적인 자세한 내용은 Census Bureau(2014)를 참조하면 된다.

5) 이에 대한 보다 자세한 내용은 Census Bureau(2014)의 11장에 있다.

3. 2015년 인구주택총조사 표본조사를 위한 표본설계

2015년 인구주택총조사는 전통적인 방식에서 그 형식을 등록센서스와 표본조사의 두 트랙으로 전환한 후 첫 번째 수행되는 조사로써 그 효율성에 대한 검토가 이루어져야 하는 중요한 조사이다. 등록센서스는 기본적으로 등록 기반 정보를 이용한 전수 조사이며 이와 함께 이루어지는 표본조사는 거주 기반 정보를 수집하는 조사로서 두 결과상의 불일치는 일정 수준 존재하게 된다. 따라서 이러한 차이에 대한 바른 해석과 추후 가능한 보정 방안에 대한 연구는 기존 인구주택총조사의 목적인 국가 주요정책 수립을 위한 기초자료 제공과 표본추출틀로써의 역할을 동시에 고려해야 한다. 이러한 복합적인 목적을 달성하기 위하여 통계청은 표본조사의 규모를 2배로 늘려 20% 표본조사를 실시하고 있다. 조사구 크기가 평균 60가구에서 30 가구로 조정된 것을 감안하면 실제 표본조사구 수는 4배 증가한 효과를 갖는 것으로 간주될 수 있다.

언급된 목적 이외에 2015년 표본조사의 표본크기가 대폭 증가했기 때문에 지역별 통계 생산 즉 읍면동 단위 통계의 생산이 요구되고 있다. 이를 위하여 2015년 인구주택총조사 표본조사 설계를 위한 연구가 꾸준히 진행되어 왔으며 본 장에서는 기존 연구결과를 바탕으로 실제 인구주택총조사 표본조사에 사용된 표본설계 방안을 간단히 설명한다.

통계청이 실시한 2015년 20% 표본조사의 표집틀은 2014년 가구주택기초조사와 2015년 조사구 적합성 확인 결과 등으로 설정한 조사구 리스트를 사용하고 있다. 이러한 표본조사는 조사구 특성을 아파트(A), 보통(1), 섬(2)인 조사구로 제한하였고 나머지 특성을 갖는 조사구는 전수조사 대상 조사구로 정의하였다. 표집틀을 구성하고 있는 전체 조사구 수는 681,785개로 이의 읍면동 별 분포는 다음의 <표 3-2>와 같다. 전체 읍면동 중 695개 읍면동(19.9%)이 읍면동 내 조사구 수를 50개 이하이다.

<표 3-2> 조사구 규모별 읍면동 분포

(단위 : 개, %)

시도	전체 조사구	아파트 조사구	기타 조사구	읍면동 수	조사구 규모별 읍면동 수						
					~5	6~20	21~50	51~100	101~200	201~400	401~
전국	681,785	318,944	362,841	3,492 (100.0)	2 (0.1)	50 (1.4)	643 (18.4)	636 (18.2)	742 (21.2)	1,031 (29.5)	388 (11.1)
서울	134,654	56,280	78,374	423		1	3	9	53	257	100
부산	47,514	24,435	23,079	208		2	4	21	85	76	20
대구	31,840	15,550	16,290	139				14	56	59	10
인천	36,659	19,076	17,583	149		2	12	20	19	72	24
광주	19,264	11,990	7,274	95			5	19	35	28	8
대전	20,489	10,471	10,018	78			1	3	20	45	9
울산	15,314	8,532	6,782	56			1	2	19	26	8
세종	3,053	1,928	1,125	13			2	2	5	1	3
경기	150,360	78,839	71,521	555	1	8	22	59	120	242	103
강원	22,393	9,752	12,641	188	1	5	39	66	47	23	7
충북	22,548	10,095	12,453	153		2	52	31	29	28	11
충남	29,170	12,215	16,955	207			55	69	34	34	15
전북	26,356	11,746	14,610	241		8	103	56	32	33	9
전남	26,531	9,703	16,828	297		13	123	93	39	22	7
경북	40,332	14,664	25,668	332		4	122	92	58	37	19
경남	47,267	21,432	25,835	315		5	92	73	78	36	31
제주	8,041	2,236	5,805	43			7	7	13	12	4

각 읍면동 및 시군구 별 표본 배분을 위해서는 우선적으로 각 읍면동으로의 가능한 조사구 표본 배분 방안을 검토한 후 결정하였다. 읍면동 배분을 위해서는 기본적으로 Kish 배분법을 사용하였으며 조사구 수가 상대적으로 적은 읍면동은 조사구 추출률이 100%가 되는 문제가 발생하여 조사구 수가 5개 미만인 4개의 읍면동을 제외하고 나머지 읍면동 내 최대 조사구 추출률을 75% 수준으로 조정하였다. 각 읍면동 내 조사구 배분 결과와 2010년 인구주택총조사 결과를 활용하여 예상 CV를 산출하고 각 시군구 통계의 목표 CV의 달성 여부에 따라 이에 미달되는 시군구 지역에는 조사구를 추가적으로 배분하였다. 인구주택총조사 표본조사를 위해 산정된 각 시군구별 통계의 목표 CV는 다음과 같이 정의되었다.

- ▶ 서울 : 10% 이하
- ▶ 특·광역시 : 구 12% 이하, 군 14% 이하
- ▶ 경기 : 구·시 12% 이하, 군 15% 이하
- ▶ 도지역 : 구·시 14% 이하, 군 17% 이하

2010년에 비하여 실제 추출되는 조사구 수 기준 표본크기의 증가는 약 4배에 해당하기 때문에 시군구 통계의 예상 상대표준오차는 2010년에 비하여 매우 감소하였다. 실제로 2010년 기준 상대표준오차가 25% 이상인 시군구가 61개로 전체 시군

구의 24.3% 수준이었으나 2015년의 경우 상대표준오차가 25% 이상인 시군구는 없을 것으로 예상되고 있다.

지역 별 표본배분 이후에는 ① 외국인, ② 1인 가구, ③ 대학생, ④ 다가구 주택, ⑤ 2010년 기준 인총결과와 등록 자료의 가구 수 차이를 기준으로 취약지역을 선정하고 이 지역에 추가적으로 조사구를 배분하였다. 이는 표본조사 결과의 활용 목적에 부합하는 표본을 얻기 위함이며 또한 지역 이외의 관심 도메인별로 산출될 예정인 통계의 안정성을 확보하기 위함이다. 이 후 각 읍면동 내에서 조사구 특성에 따라 비례 배분을 함으로 최종 배분이 마무리 된다. 결론적으로 조사구 표본추출을 위하여 사용된 층은 읍면동*조사구특성이며 표본배분은 최대 추출률을 제한한 Kish 배분법을 사용하였고, 각 읍면동 및 시군구 별 안정적인 통계 생산이 이루어지도록 추가적인 강제 배분을 하였다. 조사구 재설정 반영 후 계산된 전체 681,909개 조사구의 19.7%에 해당되는 134,529개의 조사구가 표본조사구로 선정되었다. 참고로 2010년 전체 조사구 수는 304,352개이며 표본조사구 수는 31,156개로 추출률은 10.2%이다. 전체 조사구 수의 증가는 새롭게 정의된 조사구의 규모 변동에 기인한다. 2010년 조사구 크기는 평균 60가구였으나 2015년에 새롭게 정의된 조사구의 평균 크기는 30가구이다. <표 3-3>은 시도별 표본배분 결과를 보여주고 있다.

<표 3-3> 시도별 표본규모

(단위 : 개, %)

시도	읍면동 수	전체 조사구 수			표본 조사구 수			추출률		
		전체	아파트	기타	전체	아파트	기타	전체	아파트	기타
전국	3,494	681,909	319,227	362,682	134,529	52,573	81,956	19.7	16.5	22.6
서울	424	134,616	56,347	78,269	21,428	8,709	12,719	15.9	15.5	16.3
부산	207	47,512	24,447	23,065	8,225	3,916	4,309	17.3	16.0	18.7
대구	139	31,882	15,596	16,286	5,682	2,614	3,068	17.8	16.8	18.8
인천	149	36,692	19,090	17,602	6,246	2,966	3,280	17.0	15.5	18.6
광주	95	19,258	11,984	7,274	3,569	1,970	1,599	18.5	16.4	22.0
대전	79	20,500	10,488	10,012	3,408	1,681	1,727	16.6	16.0	17.2
울산	56	15,316	8,533	6,783	2,642	1,410	1,232	17.2	16.5	18.2
세종	13	3,051	1,928	1,123	814	424	390	26.7	22.0	34.7
경기	556	150,499	78,914	71,585	25,989	12,478	13,511	17.3	15.8	18.9
강원	188	22,395	9,782	12,613	5,995	2,001	3,994	26.8	20.5	31.7
충북	153	22,473	10,082	12,391	5,268	1,766	3,502	23.4	17.5	28.3
충남	207	29,177	12,212	16,965	6,982	2,195	4,787	23.9	18.0	28.2
전북	241	26,368	11,758	14,610	6,975	2,025	4,950	26.5	17.2	33.9
전남	297	26,573	9,730	16,843	8,332	1,764	6,568	31.4	18.1	39.0
경북	332	40,324	14,667	25,657	10,476	2,576	7,900	26.0	17.6	30.8
경남	315	47,246	21,439	25,807	10,658	3,624	7,034	22.6	16.9	27.3
제주	43	8,027	2,230	5,797	1,840	454	1,386	22.9	20.4	23.9

각 층 내에서의 표본추출은 계통추출법을 통해 이루어졌다. 계통추출법을 적용한 조사구 추출을 위해서는 우선 조사구를 <표 3-4>의 기준으로 정렬하여 내재적 층화의 효과를 얻고자 하였다. 즉, 조사구 추출을 위하여 사용된 표본추출 방안은 층화계통추출법이며 조사구 추출 후 조사구 내 모든 가구가 조사되기 때문에 전체 표본추출법은 층화 일단계 집락 추출법으로 정의할 수 있으며, 이 때 일차추출단위는 조사구이고 추출방법은 층화계통추출법이 사용되었다. 조사 단위는 조사구 내 가구 및 가구원으로 정의할 수 있다.

<표 3-4> 내재적 층화를 위한 정렬 변수

아파트 조사구	① 조사구 번호
보통 조사구	① 조사구특성(1/2) ② 단독비율 ③ 조사구 번호

4. 순환표본조사를 위한 표본설계 전략

현재 통계청에서 계획하고 있는 순환표본조사는 기본적으로 순환주기(rolling cycle)는 5년, 조사 시점은 매 월을 고려하고 있다. 즉, 총 60개월 동안 조사가 매 월 지속적으로 이루어지는 구조이다. 이러한 방식의 순환표본조사는 각 년도 별로 일정 규모 이상의 크기(인구 기준)를 갖는 지역 통계를 산출하는 것과 각 년도 별로 조사된 자료를 통합하여 보다 규모가 작은 지역 단위의 통계를 작성하는 것을 목적으로 한다. 순환표본조사를 위한 표본설계에 대한 전략을 세우기 위해서는 기본적으로 아래의 사항들을 고려해야 한다.

- 조사방법: 인터넷 조사, 면접조사
- 조사구 표집틀 갱신 주기: 5년
- 각 조사의 기준 시점: 월 단위 변경(예: 매 월 1일)

미국의 ACS와는 달리 국내에서는 인터넷 조사 혹은 면접조사가 이루어질 예정이기 때문에 자료수집 과정에서 발생하는 시간적 차이는 무시될 수 있다. 더불어 추출단위가 주소가 아닌 조사구이며 조사구 표집틀이 5년 주기로 갱신되기 때문에 순환표본조사를 위한 조사구 표본추출은 매 해 혹은 매 월이 아닌 5년에 한 번 이

루어지는 것이 효율적이다. 이러한 가정 하에서 국내의 순환표본조사를 위한 표본 설계의 중요한 요소는 추출된 조사구들을 조사시점 별로 배분하는 문제이다. 또한 각 조사의 기준시점이 다르기 때문에 전통적 센서스가 제공하는 통계, 즉 특정 시점 기준 모집단에 대한 스냅샷의 의미와는 본질적으로 다른 통계가 생산될 것이다.

5년 주기의 순환표본조사를 위한 조사구 표본을 추출할 경우, 층화, 표본크기 결정, 표본 배분 등을 포함한 표본설계의 전반적인 절차는 현행 2015년 인구주택총조사의 표본조사를 위한 표본설계 방안을 따르는 것이 바람직하다. 이는 현행 표본설계의 목적이 비용과 예산의 문제를 제외한다면 순환표본조사의 목적과 부합하기 때문이다. 순환표본조사의 수행은 비용, 예산 그리고 시의성의 이점을 취하고, 대신 전통적인 센서스 통계의 사전(辭典)적 의미를 포기하는 것을 전제하고 있다. 따라서 순환표본조사를 위한 표본설계의 목적은 2015년 인구주택총조사 표본설계의 목적과 일치한다. 따라서 2015년 인구주택총조사 표본설계의 구조를 그대로 반영하는 것이 바람직하다.

동일 시점에 전체 조사구의 20%를 모두 조사하는 현행 방식의 조사를 위해 추출된 표본을 순환표본조사에 맞게 구성하기 위하여 필요한 추가적인 절차는 이를 5년(60개월)으로 분할하는 것이다. 프랑스의 순환센서스에서는 일종의 균형 추출법(balanced sampling)을 적용하여 각 해에 조사되어야 하는 표본을 결정하고 있다. 이는 각 해에 조사된 표본이 모집단을 대표할 수 있도록 구성하여 매 해 시의성을 갖춘 그리고 통계적으로 타당한 수치를 제공하기 위함이다. 마찬가지로 국내에서도 순환표본조사를 통해 기본적으로는 매 해 통계적으로 타당한 표본조사 항목에 대한 추정량을 제시해야하기 때문에 5년(60개월)에 걸쳐 배분되는 각 년도 별 표본은 모집단 대표성을 유지해야 한다. 그러나 이를 위해서 프랑스와 같이 비확률표본추출법인 균형 추출법을 사용하는 것에는 주의가 필요하다. 전체 모집단을 분할하는 프랑스의 경우와는 달리 확률표본추출법을 통해 이미 추출된 표본을 분할하는 국내의 순환표본조사 하에서는 분할 과정에서 사용되는 비확률적 요소가 전체적인 표본설계를 비확률적 표본설계로 변형시킬 수 있는 위험이 존재하며 이로 인한 추정량의 통계적 타당성이 상실될 수 있다.

이와 같은 이유로 본 연구는 표본의 분할을 위해서 반복분산추정(replicate variance estimation)을 위해 사용되는 분할 방안을 제안한다(Sarndal, Swensson and Wretman(1992)). 이 방안의 핵심 요소는 분할 과정에서 비복원 추출법을 고려하되 원표본의 추출을 위해 사용한 표본추출법을 통해 생성된 확률구조를 유지하는 것이다. 2015년 인구주택총조사의 표본조사를 위한 표본추출을 위해서는 층화 일단계에서 집락추출법이 사용되었다. 이를 표본분할을 위해 사용한다면 각 층 내에서

순환 조사 횟수에 해당하는 조사구 표본을 계통추출법을 통해 추출하는 것을 의미한다. 예를 들어, 만약 전체 표본을 5개의 그룹으로 분할한다면 각 층 내에서 정렬된 조사구 리스트로부터 가능한 5개의 계통추출표본을 구성하고 이를 5개의 조사년도에 무작위로 부여함으로써 모집단 대표성을 가질 뿐 아니라 확률표본설계 조건을 만족하는 각 년도 별 표본을 얻을 수 있게 된다.

그런데 5년 동안 조사가 수행될 조사구 추출 후 이를 분할하는 과정에 앞서 우선적으로 고려할 사항이 있다. 각 년도 별 조사되는 분할된 표본이 모집단을 대표하기 위해서는 표본추출을 위해 사용된 각 층의 조사구 수가 최소 기준을 만족해야 한다. 예를 들어, 매월 조사를 5년 동안 수행하는 경우, 각 층 별 조사구 수는 최소 60개 이상이어야 한다. <표 3-5>는 2015년 인구주택총조사의 표본조사를 위한 표본의 각 읍면동별 조사구 수의 분포를 나타낸다. 각 읍면동에 배정된 조사구 수의 분포를 보면 전체 읍면동 중 약 89%에 해당하는 3,115 읍면동이 60개 미만의 조사구를 배정받았다. 만약 읍면동 외에 층을 구성하기 위하여 사용한 조사구 특성(일반, 아파트)을 고려하여 층 별 표본 조사구를 고려하면 60개 미만의 조사구를 포함한 층의 비율은 더욱 증가할 것이다.

<표 3-5> 2015년 읍면동의 표본 조사구 수 분포

조사구 수	빈도	백분율	누적 빈도	누적 백분율
- 5	15	0.43	15	0.43
6 - 29	1,510	43.07	1,525	43.50
30 - 59	1,590	45.35	3,115	88.85
60 - 119	361	10.30	3,476	99.14
120 -	30	0.86	3,506	100.00

즉, 매 월 조사가 이루어질 경우에는 각 월에 조사되는 조사구가 전국의 모든 읍면동을 포함할 수 없는 문제가 발생할 수 있다. 가장 적은 수의 조사구가 배분된 읍면동이 2개의 조사구를 배분받은 점을 고려하면 이러한 읍면동은, 통계의 공표여부는 논의하지 않더라도, 실제 5년의 조사가 모두 마무리 된 후에야 통계 생산이 가능하리라 판단된다. 이러한 관점에서 볼 때 통계 공표 주기를 만약 ACS와 같이 1년, 3년 그리고 5년으로 고려한다면 적어도 일정 수준 이상의 규모가 확보되는 지역 단위에서 각 공표 기준에 맞는 통계를 시점 별로 산출해야 할 것이다.

2015년 인구주택총조사 표본의 각 시군구 별 분포를 보면 경상북도 울릉군이 85개의 조사구로 가장 적은 표본 조사구를 부여 받았다. 참고로 각 시군구에 배분

된 표본 조사구 수의 평균은 535개이다. 이를 5개년에 배분한다면 매 해 경상북도 울릉군의 경우 17개, 평균적으로는 107개의 조사구가 각 시군구 수준에서 조사된다. 평균적으로 추정량의 통계적 정도를 나타내는 가장 중요한 요소인 일차추출단위의 크기가 각 시군구 별로 평균 100여개인 1년 자료를 통해서 시군구 통계를 산출하는 것은 매우 어려울 것이다. <표 3-6>은 2015년 인구주택총조사 조사구 표본의 17개 광역시도별 분포를 나타내고 있다. 5년 동안 순환표본조사가 이루어질 경우 각 광역시도 단위로 매 해 조사되는 최소 조사구수는 163개로 세종특별자치시가 이에 해당하며 평균적으로는 매 해 814개의 조사구가 각 광역시도 단위에서 측정될 것이다. 2015년 인구주택총조사의 표본조사결과를 바탕으로 추후 상세한 분석을 실시해야 자세한 공표수준을 결정할 수 있으리라 판단되나, 대략적으로 볼 때 대부분의 광역시도별 통계는 매 해 산출할 수 있을 것으로 여겨진다.

<표 3-6> 2015년 인구주택총조사 조사구 표본의 17개 광역시도별 분포

시도	읍면동수	전체 조사구			표본 조사구		
		합계	아파트	기타	합계	아파트	기타
전 국	3,494	681,909	319,227	362,682	134,529	52,573	81,956
서울특별시	424	134,616	56,347	78,269	21,428	8,709	12,719
부산광역시	207	47,512	24,447	23,065	8,225	3,916	4,309
대구광역시	139	31,882	15,596	16,286	5,682	2,614	3,068
인천광역시	149	36,692	19,090	17,602	6,246	2,966	3,280
광주광역시	95	19,258	11,984	7,274	3,569	1,970	1,599
대전광역시	79	20,500	10,488	10,012	3,408	1,681	1,727
울산광역시	56	15,316	8,533	6,783	2,642	1,410	1,232
세종특별자치시	13	3,051	1,928	1,123	814	424	390
경기도	556	150,499	78,914	71,585	25,989	12,478	13,511
강원도	188	22,395	9,782	12,613	5,995	2,001	3,994
충청북도	153	22,473	10,082	12,391	5,268	1,766	3,502
충청남도	207	29,177	12,212	16,965	6,982	2,195	4,787
전라북도	241	26,368	11,758	14,610	6,975	2,025	4,950
전라남도	297	26,573	9,730	16,843	8,332	1,764	6,568
경상북도	332	40,324	14,667	25,657	10,476	2,576	7,900
경상남도	315	47,246	21,439	25,807	10,658	3,624	7,034
제주특별자치도	43	8,027	2,230	5,797	1,840	454	1,386

각 년도 별 전국 대표성을 갖출 뿐 아니라 원하는 공표수준 별 통계 생산을 위한 표본분할의 방안으로는 두 가지 가능한 안을 고려할 수 있을 것이다. 그 첫 번째는 분할을 위해 표본설계 단계에서 사용된 층을 결합하여 표본분할을 위한 새로운 층을 구성하는 것이다. 이 때 새롭게 구성된 층은 기존 표본설계 층의 결합(stratum collapsing)을 통해 이루어지며 그 최소크기가 60개 조사구 이상이 되도록 해야 한다. 표본분할을 위한 층 결합의 우선순위는 추후 연구가 추가적으로 이루어져야 할 것이나 ① 조사구 특성과 ② 시군구의 순서를 고려할 수 있다. 2015년 인구주택총조사 표본에서 표본 조사구 수가 60개 미만인 읍면동을 하나 이상 포함하고 있는 시군구는 229개이다. 즉 229개 시군구 내에서는 배분된 조사구 수가 60개 미만인 읍면동은 서로 결합하여 새로운 층을 구성하여 표본을 분할하는 방안을 고려해야 할 것이다. 표본 분할을 위한 층화가 이루어진 후에는 각 층으로부터 60개의 서로 겹치지 않는 계통 추출 방안을 이용한 조사구 표본을 구성하고 이를 매 월 무작위로 배정하여 월 별 조사가 이루어질 조사구를 결정할 수 있다.

두 번째 방안은 표본설계 층의 결합을 위하여 지역 별로 별도의 작업을 수행하는 대신 5년 동안 조사되어야 하는 표본조사구 리스트를 시군구 내에서 분할하는 방안이다. 이는 표본설계 층을 모두 무시하고 전체 표본 분할을 위해서 시군구만을 층으로 간주하고 각 층 내에서 조사구를 ① 읍면동 코드, ② 조사구 특성(아파트, 기타), 그리고 ③ 각 조사구 특성 내 정렬 변수 순으로 정렬하여 계통추출법을 통해 상호배반적인 60개의 그룹을 구성하고 이를 각 월에 무작위로 배정하는 방안이다. 내재적 층화에 의하여 각 월 별 배분된 표본은 주어진 사용된 정렬 변수 기준으로서 유사한 분포를 갖게 되며, 따라서 1년간 누적된 자료는 17개 광역 시도 통계를 산출하기 위하여 사용될 수 있을 것이다.

본 절의 내용을 요약하면 다음과 같다.

- 표집틀 갱신 주기(5년)에 맞추어서 순환조사 표본을 한 시점에 추출
- 표본추출방안은 기본적으로 2015년 인구주택총조사의 표본조사를 위한 설계 방안을 사용
- 공표 수준에 대한 결정은 2015년 인구주택총조사 결과를 분석하여 이루어져야하나 현재의 표본크기가 유지되며 적절한 표본 분할 방안이 적용된다면 매 해 전국 단위 및 광역시도 단위의 통계 산출은 가능할 수 있을 것으로 판단됨
- 표본분할의 기본적인 목적은 매 해 모집단 전체에 대한 안정적인 통계를 산

- 출하기 위한 대표성 높은 표본이 매 해 조사되도록 하기 위한 것으로 프랑스에서 사용되는 균형표본추출법과 그 맥락을 같이함
- 다만, 센서스를 고려한 프랑스와는 달리 20% 표본의 분할을 목적으로 하는 국내 순환표본조사에서는 비확률표본추출법을 적용하기 보다는 전체 표본추출법과 동일한 방법인 계통추출방안을 통해 표본을 균형적으로 분할하는 것이 바람직함
 - 배분된 표본조사구가 60개 미만인 읍면동의 비중이 높기 때문에 읍면동 단위에서 균형을 맞춘 표본을 매 해 얻는 것은 가능하지 않으며, 따라서 시군구 단위에서 매 해 균형을 맞출 수 있는 방안을 모색해야 함. 이를 위해서는 최소 층 크기를 60개 이상으로 만들기 위하여 단계적으로 층을 결합하는 방안과 시군구 내에서 추출된 조사구를 다시 정렬하고 계통추출법을 통해 표본을 분할하는 방안을 고려할 수 있음
 - 언급된 방안들 역시 실제 2015년 인구주택총조사의 표본조사 결과를 이용한 실증적 연구가 추가적으로 이루어져야 할 것으로 판단됨

제3절 모수추정방안

1. 서론

순환표본조사로의 전환을 통해 발생하는 변화는 주로 추정부분이다. 추정과 관련된 문제를 논의하기 위해서는 기본적으로 순환표본조사의 장단점에 대한 이해가 필요하다. 직관적으로 파악할 수 있는 순환표본조사의 장점으로는 5년에 걸친 조사비용의 배분을 통한 단기 예산의 안정성 확보, 전문 조사원을 활용한 비표본 오류의 최소화 그리고 누적 자료를 이용한 시의성을 갖춘 기간 통계의 제공 및 공표 수준의 탄력성 확보 등이 있다. 단점으로는 매 월 새로운 가구 및 가구원을 조사함으로써 조사 현장에서의 업무 과중, 누적자료를 활용한 추정에 있어 모수 및 추정량 해석의 어려움, 조사시점의 대한민국 전체 모집단에 대한 스냅샷(snapshot) 정보를 제공하는 센서스의 목적을 포기해야 하는 문제, 그리고 일회의 표본조사와 비교할 때 복잡한 추정 방안이 사용되어야 하며 이의 해석 역시 용이하지 않은 점 등이 있다.

순환표본조사가 가지고 있는 장단점 중 직접적으로 추정문제와 관련된 문제들은 다음의 것들로 정리할 수 있다.

- ① 추정량의 공표 시점 및 공표 수준 결정
- ② 각 공표 시점 혹은 수준 별 모수에 대한 정의
- ③ 정의된 모수에 대응하는 추정량 및 분산 추정량 산출

본 절에서는 공표 시점과 공표 수준의 결정 문제는 고려하지 않고 이것들이 주어졌을 때 이에 맞는 모수와 이의 추정 방안에 대한 내용을 정리한다. 순환표본조사에서 흔히 사용되는 기간추정은 미국의 ACS에서 고려하는 “Multi-year estimation” 혹은 “estimation for period”을 의미한다. 즉 조사 시점 및 각 조사 별 기준 시점이 다르기 때문에 전통적인 센서스에서 고려하는 특정 시점 기준 추정량이 산출될 수 없고 일정 기간 동안 누적된 자료를 이용한 추정량이 사용된다는 의미로 “기간추정”이라는 용어가 사용된다.

추정량 혹은 추정 방안의 의미를 명확히 정의하기 위해서는 우선 추정하려 하는 모수의 정의가 필수적이다. 논의의 전개를 위하여 먼저 각 $t = 1, \dots, T$ 시점에 정의할 수 있는 모수와 이의 추정량을 각각 θ_t 와 $\hat{\theta}_t$ 로 정의한다. 또한 각 시점 별 모집단 크기를 N_t 로 표기하며 추정량 $\hat{\theta}_t$ 는 해당 모수에 대한 (근사) 비편향 추정량임을 가정한다. 통계청에서 수행하는 대부분의 표본조사를 위해서는 확률표본추출법이 사용되고 있고 이에 근거한 설계 가중치를 조정하여 추정이 이루어지고 있어 이러한 가정은 타당하다.

각 시점 별 모수 추정을 위한 표본수가 충분히 확보된 경우에는 대응되는 추정량을 사용할 수 있으나 순환표본조사의 목적은 조사 시점 별 표본 수가 충분하지 않은, 일종의 소지역에 대한 통계를, 여러 시점에 걸쳐 조사된 자료를 통합하여 산출함으로써 추정량의 통계적 정확성을 높이는 것에 있다. 다만 이 과정에서 언급한 바와 같이 추정량의 의미에 대한 해석은 조사 및 기준 시점이 다른 자료들의 통합으로 인하여 직관적인 해석이 어렵다. ACS에서 사용하고 있는 가중치 구축 방안을 통해 추정되는 모수는 다음과 같이 표현된다.

$$\theta_{EP} = \left(\sum_{t=1}^T N_t \right)^{-1} \sum_{t=1}^T N_t \theta_t \approx \frac{1}{T} \sum_{t=1}^T \theta_t \quad (3-1)$$

식 (3-1)의 근사표현은 각 시점 별 모집단 크기의 변동이 거의 발생하지 않는다는 가정 하에서 유도된 것으로 실제 조사 시점 간 차이가 1개월로 크지 않은 경우에는 두 모수는 거의 동일한 것으로 간주할 수 있다. 즉 ACS에서 정의하는 기간추

정이란 고려하는 기간 동안, 보다 정확히는 각 조사 시점에 대응하는 모수들의 평균을 추정하는 것으로 이해할 수 있다.

기간 추정에 대한 논의는 흔히 추정의 기준 시점에 대한 문제로 오해되어 논의되나 원칙적으로 기간추정의 개념은 언급한 바와 같이 여러 시점에 정의되는 모수의 평균 추정을 나타낸다. 굳이 해당 모수에 대한 기준 시점을 논의한다면 Brown and Beaujouan(2013)가 언급한 바와 같이 모수를 시점 t 로 정의하여 계산되는 $T^{-1} \sum_{t=1}^T t$ 로써 전체 기간의 중간 시점을 나타내는 것으로 이해할 수 있으나, 이 또한 정확한 해석은 아니며 이를 추정 단계에서 논의하는 것은 타당하지 않다. 본 절에서는 정의된 기간추정의 개념 하에서 모수 추정을 위한 가중치 작성에 대한 내용을 ACS방안을 기초로 하여 우선적으로 다루고, 추가적으로 추정 시 필요한 내용을 검토한다.

2. 기간추정을 위한 가중치 산출 방안

식 (3-1)에 주어진 모수를 추정하는 기간 추정 방안으로 고려할 수 있는 가장 기본적인 방안은 각 시점 별로 계산되는 추정량을 해당 모수에 대입하여 계산하는 것이다. 즉 첫 번째로 고려할 수 있는 추정량은 시점별 추정량의 (가중) 평균으로 다음과 같이 표현할 수 있다.

$$\hat{\theta}_{EP}^{(1)} = \left(\sum_{t=1}^T N_t \right)^{-1} \sum_{t=1}^T N_t \hat{\theta}_t \approx \frac{1}{T} \sum_{t=1}^T \hat{\theta}_t \quad (3-2)$$

송중호 외(2010)는 국민건강영양조사 자료를 이용하여 기간추정량의 정의를 위하여 식 (3-2)에 주어진 추정량 이외에 다양한 방안들을 검토하고 통계적으로 비교하여 (3-2)에 주어진 추정량이 다른 고려된 추정량보다 효율적임을 실증적으로 밝히고 있다. 이러한 연구결과를 바탕으로 질병관리본부에서는 식 (3-2)의 추정량을 산출할 수 있는 가중치를 각 통합 자료 제공 시 함께 제공하고 있다.

다만 순환표본조사 자료의 통계청 내부와 외부 사용자에게 각 년도 별 추정량 혹은 가중치를 제공하고 연구자 혹은 자료 사용자가 위에 주어진 공식을 적용하여 기간추정량을 산출하도록 하는 방안은 적절하지 않은 것으로 판단된다. 대신 기존 통계청 혹은 기타 국가승인통계 작성 기관들이 원자료를 제공하는 형식, 즉 관측된

값과 더불어 자료 분석을 위한 한 세트의 가중치를 제공하는 것이 바람직하며, 이것이 표준화된 절차에 따라 사용자가 분석을 수행할 수 있는 방안이 되어야 할 것이다. 실제 국내의 승인통계 작성 기관에서 수행하는 조사 뿐 아니라 통계청에서 순환표본조사의 도입을 위하여 벤치마킹을 고려하고 있는 미국의 ACS 역시 가중치 작성을 통한 추정방안을 사용하고 있다.⁶⁾

순환표본조사 자료를 이용한 기간추정을 위해 필요한 가중치 산출 작업은 크게 두 단계에 걸쳐 이루어진다. 첫 번째 단계는 단순히 해당 기간 내에 조사된 자료를 세로결합(set)하는 것이다. 즉 매 시점별로 동일한 변수가 서로 다른 관측단위로부터 측정되기 때문에 세로결합을 통해 자료를 통합하는 과정이다. 이 때 통합자료에는 각 시점별로 산출된 설계가중치(sampling weight) 내지 무응답 조정 가중치(non-response adjusted weight)가 함께 포함되어야 한다.

두 번째 단계는 벤치마킹 방안을 통해서 통합된 자료의 분석을 위한 최종가중치가 산출하는 것이다. 벤치마킹 기법을 이용한 최종가중치 산출을 위해서 벤치마킹을 위한 모집단 정보가 필요하다. 벤치마킹을 위한 모집단 정보의 산출은 기간추정의 정의에 의존한다. 언급한 바와 같이 기간추정을 해당 기간 모수의 평균으로 고려할 경우, 모집단 정보 역시 해당 기간의 모집단 정보의 평균을 사용하는 것이 바람직하다. 이를 나타내기 위하여 벤치마킹을 위한 t 시점의 보조정보를 $\mathbf{x}_{tU} = (x_{1,tU}, \dots, x_{k,tU})$ 로 표기하자. 예를 들어, 지역*성*연령을 사후층으로 정의하고 이의 모집단 분포를 벤치마킹을 위해 사용할 경우, $x_{i,tU}$ 는 t 시점에 i 번째 사후층의 모집단 비율 혹은 모집단 크기를 나타낸다. 각 시점 별 벤치마킹을 위한 모집단 정보가 주어졌을 때 통합표본자료의 분석을 위한 최종가중치 산출을 위해 사용되는 벤치마킹 정보는, 따라서 다음과 같이 정의된다.

$$\bar{\mathbf{x}}_{EP} = \frac{\sum_{t=1}^T \mathbf{x}_{tU}}{T} \quad (3-3)$$

최종가중치 산출을 위하여 주어진 벤치마킹 정보를 활용하는 방안으로는 회귀가중치, 레이킹 비 가중치, 그리고 사후층화 가중치 등을 고려할 수 있다. 실제 사후층화 가중치는 특별한 형태의 회귀가중치이며 이론적으로 레이킹 비 가중치는 회귀가중치와 근사적으로 일치(consistent)하기 때문에 간략히 회귀추정 방안을 본 절에서는 소개한다. 조사시점 t 에 조사된 개체 i 에 부여된 설계가중치 혹은 무응답 조

6) 이에 대한 자세한 내용은 Census Bureau(2014)의 11장에 기술되어 있다.

정 가중치를 d_{ti} 로 정의하자. 총 T 시점에 걸쳐 통합된 자료의 분석을 위하여 사용할 수 있는 회귀 가중치는 다음과 같다.

$$w_{ti} = \frac{d_{ti}}{T} + (\hat{\mathbf{x}}_d - \bar{\mathbf{x}}_{EP}) \left(\sum_{s=1}^T \sum_j d_{sj} \mathbf{x}_{sj}' \mathbf{x}_{sj} \right)^{-1} d_{ti} \mathbf{x}_{ti}' \quad (3-4)$$

$$\hat{\mathbf{x}}_d = \frac{1}{T} \sum_{t=1}^T \sum_i d_{ti} \mathbf{x}_{ti}, \quad \mathbf{x}_{ti} = (x_{1,ti}, \dots, x_{k,ti}) .$$

식 (3-4)에 주어진 가중치는 보조변수 \mathbf{x} 을 사후층을 나타내는 지시변수의 벡터로 표현할 경우 사후층화추정량을 산출하기 위한 가중치가 되며 만약 보조변수 \mathbf{x} 을 성, 연령그룹 등의 각 변수 별로 개체의 소속 여부를 나타내는 지시함수의 벡터로 표기할 경우에는 레이킹 비 추정량을 산출하기 위한 가중치와 근사적으로 일치한다. 실제 가중치 (3-4)는 ACS 자료를 이용한 기간추정을 위해 사용하는 가중치와 동일한 것으로 간주할 수 있다. 모수 (3-1)에 대한 기간추정량은 식 (3-4)의 가중치를 이용하여 다음과 같이 정의된다.

$$\hat{\theta}_{EP}^{(2)} = \sum_{t=1}^T \sum_i w_{ti} y_{ti} \quad (3-5)$$

주어진 가중치 (3-4)는 또한 벤치마킹 혹은 캘리브레이션(calibration) 성질을 만족한다. 즉, 주어진 가중치를 보조정보에 대한 기간추정을 위해 사용한다면 다음과 같이 벤치마킹 정보와 일치하게 된다.

$$\sum_{t=1}^T \sum_i w_{ti} \mathbf{x}_{ti} = \bar{\mathbf{x}}_{EP}$$

고려한 기간 동안 모집단의 변동이 적은 경우에는 식 (3-2)에 주어진 추정량은 식 (3-5)의 추정량과 거의 일치하나 회귀추정량과는 달리 캘리브레이션 성질은 만족하지 않는다. 각 시점 별 추정량 $\hat{\theta}_t$ 이 각 시점 기준으로 캘리브레이션 성질을 만족한다 하더라도 이의 가중 평균 혹은 단순평균으로 정의된 $\hat{\theta}_{EP}^{(1)}$ 는 캘리브레이션 성질을 항상 만족할 수는 없다.

고려한 두 추정량의 분산추정량은 각각 다음과 같이 계산할 수 있다. 먼저 $\hat{\theta}_{EP}^{(1)}$

의 분산은 각 시점 별 추정량이 근사적으로 독립이라는 가정 하에서 각 시점별 분산추정량의 가중합으로 아래와 같이 정의된다.

$$\hat{V}(\hat{\theta}_{EP}^{(1)}) = \frac{1}{T^2} \sum_{t=1}^T \hat{V}(\hat{\theta}_t) \quad (3-6)$$

회귀추정방안을 적용하여 산출되는 추정량 $\hat{\theta}_{EP}^{(2)}$ 추정량은 일반적인 회귀추정량의 분산을 추정하기 위한 공식을 통해 계산할 수 있으면 그 공식은 (3-7)과 같다. 공식의 정의를 위해서 표본설계 층을 h 로, 각 층 내에서의 모집단 조사구 수와 표본 조사구 수를 각각 N_h 그리고 n_h 로 정의한다. 비록 식 (3-5)는 시점에 대한 표기를 사용하였으나 분산추정의 식은 이의 간소화를 위하여 시점을 나타내는 첨자는 생략하였다. 즉, 총 T 시점에 걸쳐 조사된 전체 조사구 수를 n_h 로 표기하며 표본설계 시 사용한 층 정보를 이용하여 분산추정량의 식을 아래와 같이 나타낸다.

$$\hat{V}(\hat{\theta}_{EP}^{(2)}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (e_{hi\cdot} - \bar{e}_{h\cdot\cdot})^2 \quad (3-7)$$

$$\bar{e}_{h\cdot\cdot} = \frac{1}{n_h} \sum_i \sum_j w_{hij} e_{hij}$$

$$e_{hi\cdot} = \sum_j w_{hij} e_{hij}$$

$$e_{hij} = y_{hij} - \mathbf{x}_{hij} \hat{\beta}$$

식 (3-7)의 분산추정량은 서베이 자료 분석을 위한 모듈이 포함되어 있는 SAS 혹은 SPSS의 회귀분석 프로시저를 통해서 계산할 수 있다.

3. 시험조사 자료를 이용한 가중치 산출

본 절에서는 위에서 다룬 가중치 산출 방안을 2013년 1월부터 2014년 12월까지 진행된 순환표본조사를 위한 시험조사 결과에 적용하고 그 결과를 살펴본다. 시험조사는 경상남도 거창군과 인천광역시 연수구에서 소규모로 순환표본조사 수행의 가능성 및 문제점들을 파악하기 위하여 수행되었다. 거창군의 경우는 시험조사가 2012년 10월부터 수행되었으나 통계청의 요청으로 분석을 위해서는 인천광역시 연

수구와 동일한 기간에 조사된 24개월 자료를 활용하였다. 조사는 기본적으로 매월 이루어졌으나 인천광역시의 경우는 2013년에는 분기별로 조사가 이루어졌다.

시험조사가 수행된 지역의 모집단 조사구 및 표본 조사구 그리고 표본 가구 분포는 다음의 <표 3-7> 그리고 <표 3-8>과 같다. 본 연구를 위해서는 가구원 자료를 사용하였다. 수집된 자료는 3개월, 6개월, 12개월, 18개월 그리고 전체 24개월로 통합하여 분석하였다. 즉 <표 3-8>에서 3개월에 해당하는 8개의 그룹은 2013년 1월-3월, 2013년 4월-6월, 2013년 7월-9월, 2013년 10월-12월, 2014년 1월-3월, 2014년 4월-6월, 2014년 7월-9월 그리고 2014년 10월-12월에 해당한다. 6개월에 해당하는 4개의 그룹은 각각 2013년 1월-6월, 2013년 7월-12월, 2014년 1월-6월 그리고 2014년 7월-12월을 나타낸다. 12개월의 경우 중복되는 시기가 있어 3개의 그룹은 각각 2013년 1월-12월, 2013년 7월-2014년 6월 그리고 2014년 1월-12월에 해당하며 18개월은 각각 2013년 1월-2014년 6월 그리고 2013년 7월-2014년 12월을 나타낸다. 따라서 표본조사구 수는 각 배수의 관계가 있다. 12개월 동안 조사된 조사구 수가 전체 조사구 수의 약 20%에 해당하며 각 조사구로부터 평균 30가구를 조사하였다. 참고로 시험조사를 위해서는 기존 평균 60가구 기준의 조사구 리스트가 표본 추출을 위해 사용되었기 때문에 본 시험조사를 위해서는 추출된 조사구 내에서 가구 추출이 이루어졌고 각 조사구 내에서 30 가구가 추출된 것으로 자료상으로 파악된다.

표본조사구 분석을 통해 살펴볼 때 순환표본조사가 이루어질 경우, 각 읍면동 내에서 전체 조사구 중 평균 20%의 조사구가 추출될 예정이며 이는 시험조사의 12개월 분 조사자료에 해당한다. 표본 조사구 수와 추출률만을 기준으로 볼 때 20% 표본조사 결과를 이용한다면 시군구 통계는 안정적으로 산출할 수 있을 것으로 예측된다.

<표 3-7> 시범지역 조사구 모집단 분포

지역	추출그룹	모집단 조사구수	지역	추출그룹	모집단 조사구수
인천시	선학동	142	경상남도 거창군 ⁷⁾	A	245
	연수 1동	207		B	50
연수구	연수 2동	152		C	74
	연수 3동	118		D	56
	청학동	199			
	총	818		총	425

<표 3-8> 시범지역 표본 조사구 및 가구 분포

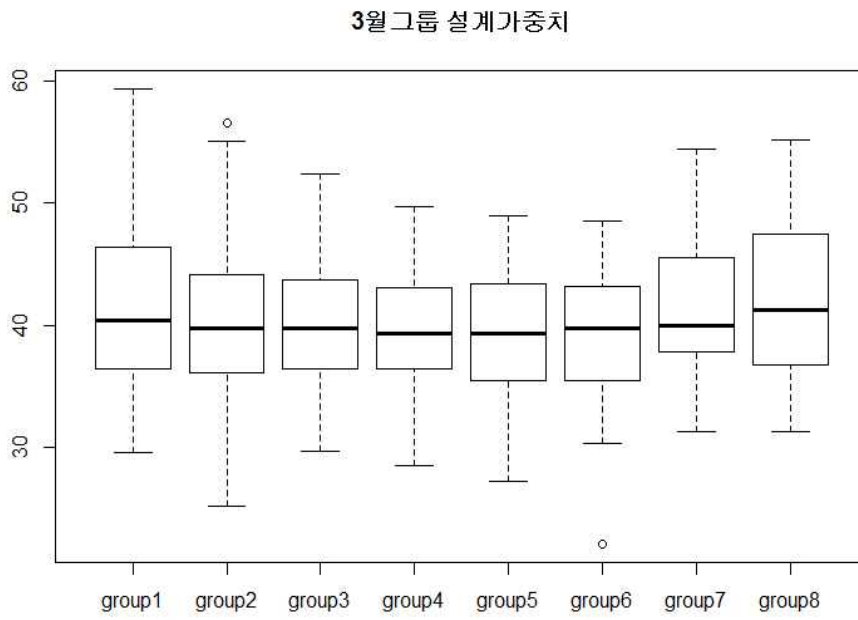
개월	그룹	인천시 연수구		거창	
		표본 조사구수	표본 가구수	표본 조사구수	표본 가구수
3	1	42	1,257	24	720
	2	42	1,260	24	720
	3	42	1,260	24	717
	4	42	1,260	24	719
	5	42	1,260	24	721
	6	42	1,260	24	723
	7	42	1,261	24	723
	8	42	1,260	24	722
6	1	84	2,571	48	1,440
	2	84	2,490	48	1,436
	3	84	2,550	48	1,444
	4	84	2,521	48	1,445
12	1	168	5,037	96	2,876
	2	168	5,040	96	2,880
	3	168	5,041	96	2,889
18	1	252	7,557	144	4,320
	2	252	7,561	144	4,325
24	1	336	10,078	192	5,765

7) A: 거창읍, B: 남상면, 남하면, 신원면, C: 웅양면, 가북면, 주상면, 가조면, D: 북상면, 고제면, 위천면, 마리면

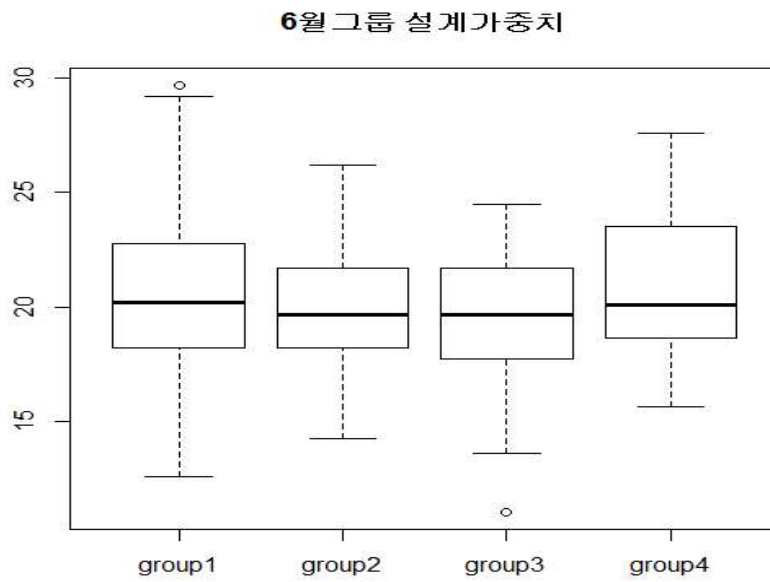
<표 3-9> 시범지역 설계가중치 분포

개월	그룹	인천시 연수구							거창군						
		기초통계량							기초통계량						
		mean	st.dev	min	Q1	median	Q3	max	mean	st.dev	min	Q1	median	Q3	max
3	1	41.740	6.655	29.580	36.480	40.460	46.430	59.340	34.420	11.223	17.270	23.150	35.830	34.420	54.940
	2	40.150	6.404	25.210	36.200	39.800	44.160	56.580	34.040	13.589	13.070	21.470	34.040	41.650	56.350
	3	40.460	5.193	29.770	36.480	39.800	43.700	52.400	32.760	10.903	11.110	29.490	34.030	42.190	57.870
	4	39.770	4.976	28.500	36.480	39.330	43.070	49.680	35.290	17.896	15.860	18.750	25.670	50.810	78.040
	5	39.660	5.051	27.230	35.500	39.330	43.470	43.470	35.280	12.852	15.830	22.920	32.670	44.550	59.390
	6	39.200	5.261	22.170	35.500	39.800	43.190	48.510	36.340	11.817	18.090	25.870	40.090	43.400	58.910
	7	41.490	5.497	31.360	37.810	40.020	45.600	54.390	44.090	14.927	15.810	35.000	46.030	54.940	74.240
	8	41.990	6.192	31.360	36.730	41.300	47.500	55.200	43.580	14.619	19.790	31.250	43.280	57.980	84.930
6	1	20.450	3.286	12.600	18.240	20.230	22.770	29.670	17.230	6.090	7.127	10.540	17.890	20.610	28.780
	2	20.060	2.549	14.250	18.240	19.670	21.740	26.200	16.930	5.904	6.667	12.000	16.530	21.780	33.440
	3	19.710	2.581	11.080	17.750	19.670	21.740	24.500	17.920	6.132	8.810	12.240	17.820	21.900	21.900
	4	20.870	2.930	15.680	18.640	20.120	23.550	27.600	21.840	7.097	8.781	17.240	21.830	27.610	40.440
12	1	10.130	1.470	6.302	9.121	9.121	11.090	14.840	8.536	2.830	3.941	6.108	8.167	10.310	16.720
	2	9.943	1.285	5.542	9.025	9.833	10.870	13.100	8.726	3.004	3.627	6.290	8.263	11.210	16.330
	3	10.150	1.413	5.542	9.142	9.950	11.090	13.800	9.958	3.377	3.854	6.741	10.260	11.780	19.750
18	1	6.690	0.946	3.694	6.017	6.017	7.297	9.890	5.784	1.949	2.627	4.079	5.570	7.273	10.970
	2	6.741	0.913	3.694	6.095	6.633	7.389	9.200	6.314	2.220	2.372	4.405	6.253	7.740	13.270
24	1	5.069	0.721	2.771	4.560	4.975	5.547	7.418	4.618	1.596	1.970	3.207	4.612	5.669	9.992

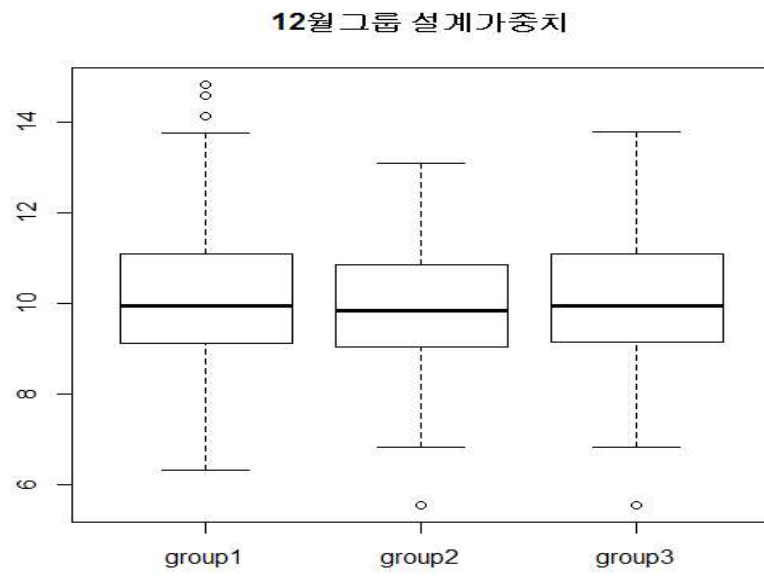
[그림 3-1] 인천 3개월 그룹 설계가중치



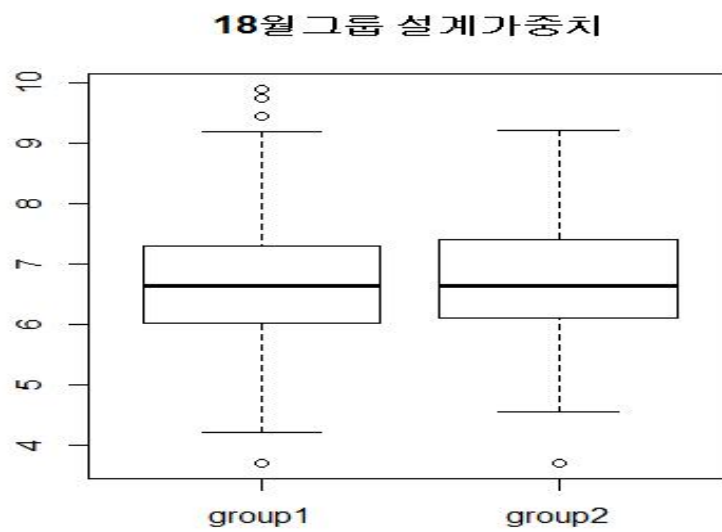
[그림 3-2] 인천 6개월 그룹 설계가중치



[그림 3-3] 인천 12개월 그룹 설계가중치

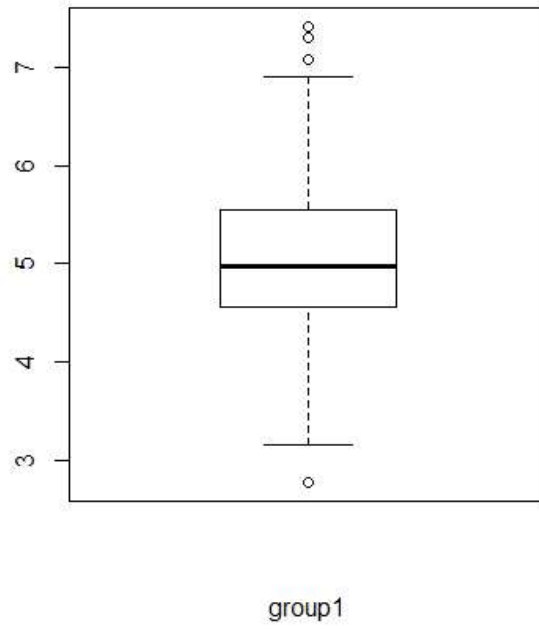


[그림 3-4] 인천 18개월 그룹 설계가중치



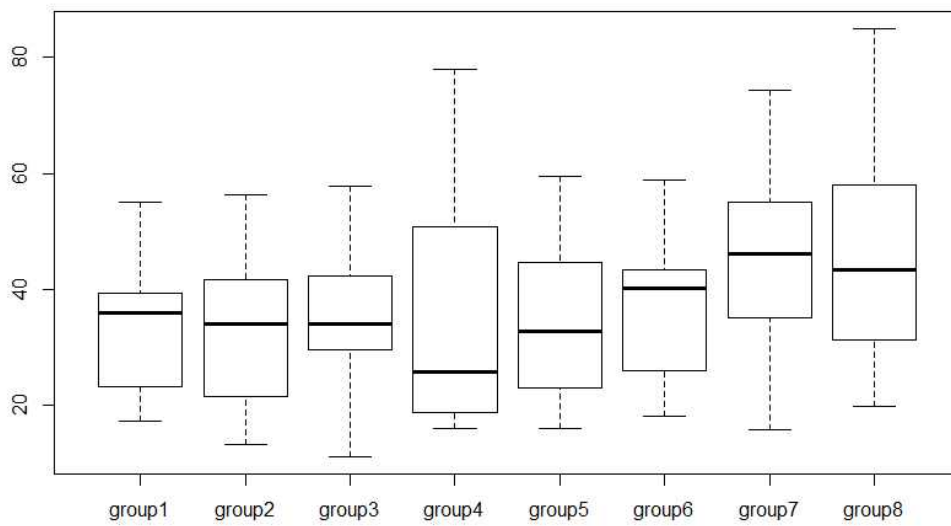
[그림 3-5] 인천 18개월 그룹 설계가중치

24월그룹 설계가중치



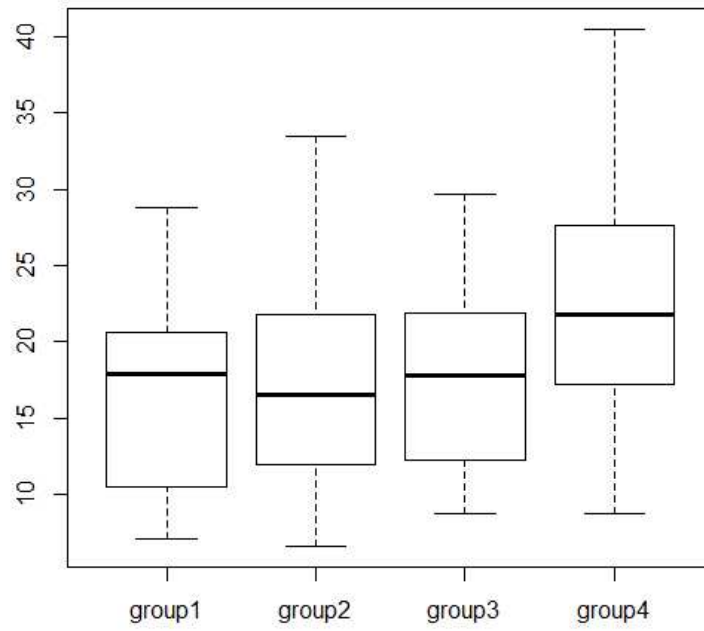
[그림 3-6] 거창 3개월 그룹 설계가중치

3월그룹 설계가중치



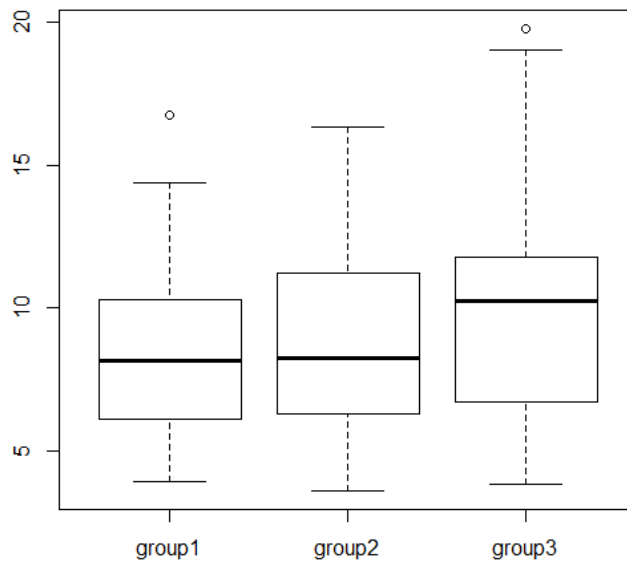
[그림 3-7] 거창 6개월 그룹 설계가중치

6월그룹 설계가중치

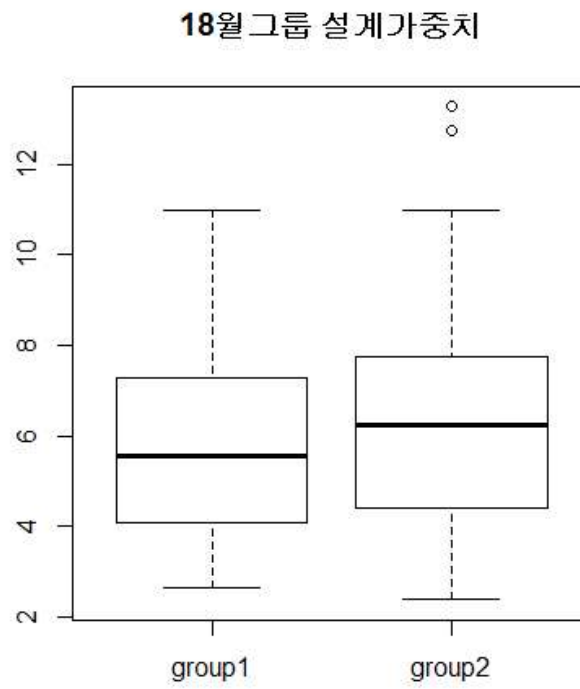


[그림 3-8] 거창 12개월 그룹 설계가중치

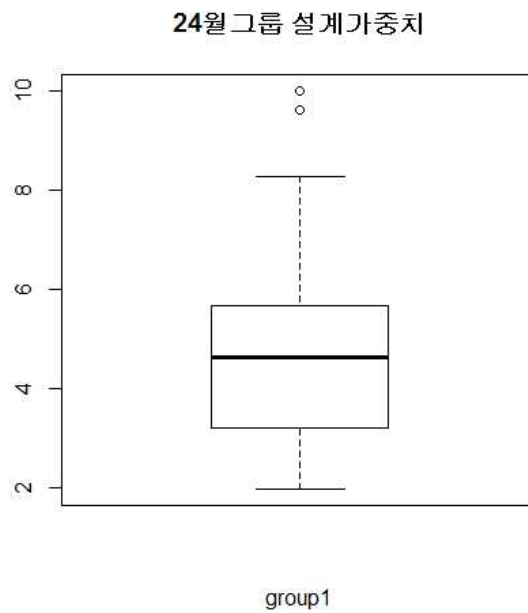
12월그룹 설계가중치



[그림 3-9] 거창 18개월 그룹 설계가중치



[그림 3-10] 거창 24개월 그룹 설계가중치



<표 3-9> 그리고 [그림 3-1]~[그림 3-10]은 각 그룹 별 설계가중치의 분포를 나타내고 있다. 설계가중치란 표본설계 상에서 부여되는 가중치를 의미하며 조사구 추출 및 가구원 추출이 균일 확률표본추출법(equal probability sampling design)인 단순임의추출법 혹은 계통추출법에 의하여 추출되었다는 가정 하에 다음과 같이 계산된다. 여기서 N_c 와 n_c 는 층 내 모집단 및 표본 조사구를, M_i 와 m_i 는 i 번째 조사구 내 모집단 및 표본가구 수를 나타낸다. 대부분의 경우, $m_i = 30$ 이나 일부 조사구에서는 1~2개의 가구가 적게 혹은 더 많이 추출되었다.

$$\text{설계가중치} = \frac{N_c}{n_c} \times \frac{M_i}{m_i}$$

설계가중치의 분포를 살펴보면, 우선 인천 연수구의 분포는 거창군에 비하여 안정적인 분포를 보이고 있음을 확인할 수 있다. 특별히 거창군의 3개월과 6개월 자료의 경우 평균 가중치 값 역시 그 변동이 심하게 발생하는 것을 알 수 있다. 이는 각 월 별 표본배분이 조사구 단위에서는 가능한 한 균등하게 이루어졌으나 각 조사구의 크기가 균일하지 않기 때문인 것으로 판단된다. 또한 거창군의 경우 인천과는 달리 각 읍면 단위의 조사구 수가 크지 않아 여러 읍면을 통합하여 층을 구성하였고 읍면 간 조사구 수 및 조사구 내 가구 수의 차이가 이러한 차이를 나타내고 있는 것으로 파악된다. 그러나 18개월 그리고 24개월 자료를 보면 상대적으로 안정적인 설계가중치 분포를 보이고 있으며 이를 통해 볼 때 고려한 거창군의 모든 읍면이 분석 대상 자료에 포함되어 안정적인 통계를 산출할 수 있을 것을 예상할 수 있다. 인천시 역시 동일한 구조를 나타내고 있으나 거창군 자료에 비하여 그 정도는 심하지 않은 것을 확인할 수 있다.

12개월 가중치, 약 20%의 조사구가 조사된 경우에 평균 가중치는 약 8-10으로 각 표본 가구 내 가구원이 전체 가구원 중 8-10명을 대표하는 것으로 판단할 수 있으며 이를 이용하여 시군구 통계를 산출하는 것이 가능하다. 다만, 군 지역의 경우는 각 순환표본조사의 시점에 따라 실제로 포함되는 읍면이 다를 수 있기 때문에 누적표본에 모든 읍면이 포함되었는지 확인하고 이의 포함 비율이 일정 수준 이상일 때 군 단위 통계를 산출하는 것이 바람직할 것이다.

<표 3-10> 벤치마킹을 위한 모집단 정보: 인천시 연수구 남성

연령그룹		0-4	5-9세	0-14	5-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85 -	
3	1	N	2,676	2,341	3,331	4,754	5,482	5,047	5,639	4,872	5,327	5,326	6,366	4,351	2,308	1,607	1,195	715	362	230
		%	4.32	3.78	5.38	7.68	8.85	8.15	9.11	7.87	8.60	8.60	10.28	7.03	3.73	2.59	1.93	1.15	0.58	0.37
	2	N	2,669	2,313	3,255	4,673	5,503	5,089	5,628	4,777	5,334	5,266	6,398	4,464	2,341	1,639	1,214	729	370	230
		%	4.31	3.74	5.26	7.55	8.89	8.22	9.09	7.72	8.62	8.51	10.34	7.21	3.78	2.65	1.96	1.18	0.60	0.37
	3	N	2,665	2,303	3,169	4,582	5,515	5,114	5,619	4,724	5,296	5,268	6,382	4,592	2,399	1,658	1,228	749	370	234
		%	4.31	3.72	5.12	7.41	8.91	8.27	9.08	7.64	8.56	8.52	10.32	7.42	3.88	2.68	1.99	1.21	0.60	0.38
	4	N	2,687	2,262	3,099	4,501	5,505	5,113	5,658	4,664	5,287	5,271	6,281	4,733	2,455	1,661	1,237	777	376	237
		%	4.35	3.66	5.01	7.28	8.91	8.27	9.15	7.55	8.56	8.53	10.16	7.66	3.97	2.69	2.00	1.26	0.61	0.38
	5	N	2,682	2,275	2,993	4,410	5,466	5,103	5,684	4,649	5,239	5,283	6,257	4,836	2,510	1,696	1,262	793	385	246
		%	4.34	3.68	4.85	7.14	8.85	8.26	9.20	7.53	8.48	8.55	10.13	7.83	4.06	2.75	2.04	1.28	0.62	0.40
	6	N	2,657	2,288	2,921	4,325	5,444	5,065	5,657	4,592	5,198	5,241	6,212	4,878	2,599	1,698	1,293	795	399	258
		%	4.32	3.72	4.75	7.03	8.85	8.23	9.20	7.46	8.45	8.52	10.10	7.93	4.23	2.76	2.10	1.29	0.65	0.42
	7	N	2,642	2,272	2,843	4,269	5,426	5,095	5,587	4,607	5,152	5,221	6,222	4,926	2,663	1,733	1,297	820	399	256
		%	4.30	3.70	4.63	6.95	8.83	8.29	9.09	7.50	8.39	8.50	10.13	8.02	4.34	2.82	2.11	1.33	0.65	0.42
	8	N	2,624	2,260	2,748	4,190	5,389	5,090	5,489	4,620	5,060	5,164	6,162	5,029	2,712	1,768	1,295	824	422	262
		%	4.29	3.70	4.50	6.86	8.82	8.33	8.98	7.56	8.28	8.45	10.08	8.23	4.44	2.89	2.12	1.35	0.69	0.43
6	1	N	2,673	2,327	3,293	4,714	5,493	5,068	5,634	4,825	5,330	5,296	6,382	4,407	2,325	1,623	1,205	722	366	230
		%	4.32	3.76	5.32	7.61	8.87	8.19	9.10	7.79	8.61	8.55	10.31	7.12	3.75	2.62	1.95	1.17	0.59	0.37
	2	N	2,676	2,283	3,134	4,542	5,510	5,114	5,638	4,694	5,292	5,270	6,331	4,663	2,427	1,660	1,233	763	373	235
		%	4.33	3.69	5.07	7.34	8.91	8.27	9.12	7.59	8.56	8.52	10.24	7.54	3.92	2.68	1.99	1.23	0.60	0.38
	3	N	2,670	2,282	2,957	4,368	5,455	5,084	5,671	4,620	5,219	5,262	6,234	4,857	2,555	1,697	1,277	794	392	252
		%	4.33	3.70	4.80	7.09	8.85	8.25	9.20	7.49	8.47	8.54	10.11	7.88	4.14	2.75	2.07	1.29	0.64	0.41
	4	N	2,633	2,266	2,796	4,230	5,408	5,092	5,538	4,614	5,106	5,193	6,192	4,978	2,688	1,751	1,296	822	411	259
		%	4.30	3.70	4.56	6.90	8.83	8.31	9.04	7.53	8.33	8.48	10.11	8.12	4.39	2.86	2.12	1.34	0.67	0.42
12	1	N	2,674	2,305	3,214	4,628	5,501	5,091	5,636	4,759	5,311	5,283	6,357	4,535	2,376	1,641	1,219	742	370	233
		%	4.32	3.72	5.19	7.48	8.89	8.23	9.11	7.69	8.58	8.54	10.27	7.33	3.84	2.65	1.97	1.20	0.60	0.38
	2	N	2,673	2,282	3,046	4,455	5,482	5,099	5,654	4,657	5,255	5,266	6,283	4,760	2,491	1,678	1,255	778	382	244
		%	4.33	3.70	4.93	7.22	8.88	8.26	9.16	7.54	8.51	8.53	10.18	7.71	4.03	2.72	2.03	1.26	0.62	0.39
	3	N	2,651	2,274	2,876	4,299	5,431	5,088	5,604	4,617	5,162	5,228	6,213	4,917	2,621	1,724	1,287	808	401	255
		%	4.31	3.70	4.68	6.99	8.84	8.28	9.12	7.51	8.40	8.51	10.11	8.00	4.27	2.81	2.09	1.31	0.65	0.42
18	1	N	2,673	2,297	3,128	4,541	5,486	5,089	5,648	4,713	5,280	5,276	6,316	4,642	2,435	1,660	1,238	760	377	239
		%	4.32	3.72	5.06	7.35	8.88	8.23	9.14	7.63	8.54	8.54	10.22	7.51	3.94	2.69	2.00	1.23	0.61	0.39
	2	N	2,660	2,277	2,962	4,380	5,458	5,097	5,616	4,643	5,205	5,242	6,253	4,833	2,556	1,703	1,269	793	392	249
		%	4.32	3.70	4.81	7.11	8.86	8.28	9.12	7.54	8.45	8.51	10.15	7.85	4.15	2.76	2.06	1.29	0.64	0.40
24	N	2,663	2,289	3,045	4,463	5,466	5,090	5,620	4,688	5,237	5,255	6,285	4,726	2,498	1,683	1,253	775	385	244	
	%	4.32	3.71	4.94	7.24	8.86	8.25	9.11	7.60	8.49	8.52	10.19	7.66	4.05	2.73	2.03	1.26	0.62	0.40	

<표 3-11> 벤치마킹을 위한 모집단 정보: 인천시 연수구 여성

연령그룹		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85 -	
3	1	N	2,483	2,261	2,994	4,327	5,204	4,353	4,677	4,386	5,236	6,050	5,984	3,838	2,172	1,703	1,751	1,402	988	637
		비율	4.11	3.74	4.95	7.16	8.61	7.20	7.74	7.26	8.66	10.01	9.90	6.35	3.59	2.82	2.90	2.32	1.64	1.64
	2	N	2,467	2,239	2,949	4,279	5,247	4,342	4,682	4,305	5,211	5,961	6,077	3,950	2,200	1,743	1,744	1,430	988	660
		%	4.08	3.70	4.88	7.08	8.68	7.18	7.74	7.12	8.62	9.86	10.05	6.53	3.64	2.88	2.88	2.36	1.63	1.63
	3	N	2,468	2,213	2,900	4,198	5,273	4,320	4,728	4,195	5,173	5,927	6,145	4,024	2,253	1,776	1,762	1,419	997	674
		%	4.08	3.66	4.80	6.94	8.72	7.15	7.82	6.94	8.56	9.81	10.17	6.66	3.73	2.94	2.91	2.35	1.65	1.65
	4	N	2,478	2,174	2,815	4,155	5,234	4,311	4,752	4,092	5,125	5,940	6,135	4,150	2,281	1,807	1,771	1,426	1,004	696
		%	4.11	3.60	4.66	6.88	8.67	7.14	7.87	6.78	8.49	9.84	10.17	6.88	3.78	2.99	2.93	2.36	1.66	1.66
	5	N	2,477	2,121	2,751	4,082	5,210	4,294	4,771	4,022	5,098	5,947	6,154	4,258	2,330	1,838	1,768	1,448	1,030	727
		%	4.11	3.52	4.56	6.77	8.64	7.12	7.91	6.67	8.45	9.86	10.20	7.06	3.86	3.05	2.93	2.40	1.71	1.71
	6	N	2,452	2,081	2,705	3,998	5,180	4,302	4,749	3,926	5,078	5,879	6,162	4,354	2,360	1,857	1,746	1,497	1,037	740
		%	4.08	3.46	4.50	6.65	8.62	7.16	7.90	6.53	8.45	9.78	10.25	7.24	3.93	3.09	2.90	2.49	1.73	1.73
	7	N	2,436	2,094	2,655	3,901	5,148	4,349	4,723	3,930	4,995	5,813	6,203	4,416	2,421	1,897	1,730	1,505	1,037	763
		%	4.06	3.49	4.42	6.50	8.58	7.25	7.87	6.55	8.32	9.68	10.34	7.36	4.03	3.16	2.88	2.51	1.73	1.73
	8	N	2,417	2,064	2,591	3,785	5,047	4,399	4,643	3,909	4,865	5,725	6,195	4,535	2,481	1,901	1,718	1,503	1,067	787
		%	4.05	3.46	4.34	6.35	8.46	7.38	7.79	6.56	8.16	9.60	10.39	7.60	4.16	3.19	2.88	2.52	1.79	1.79
6	1	N	2,475	2,250	2,972	4,303	5,226	4,347	4,680	4,346	5,224	6,006	6,031	3,894	2,186	1,723	1,747	1,416	988	649
		%	4.09	3.72	4.91	7.12	8.64	7.19	7.74	7.19	8.64	9.93	9.97	6.44	3.62	2.85	2.89	2.34	1.63	1.63
	2	N	2,473	2,194	2,857	4,176	5,254	4,316	4,740	4,144	5,149	5,933	6,140	4,087	2,267	1,792	1,766	1,423	1,000	685
		%	4.09	3.63	4.73	6.91	8.70	7.15	7.85	6.86	8.53	9.82	10.17	6.77	3.75	2.97	2.92	2.36	1.66	1.66
	3	N	2,465	2,101	2,728	4,040	5,195	4,298	4,760	3,974	5,088	5,913	6,158	4,306	2,345	1,848	1,757	1,472	1,033	734
		%	4.09	3.49	4.53	6.71	8.63	7.14	7.90	6.60	8.45	9.82	10.23	7.15	3.89	3.07	2.92	2.45	1.72	1.72
	4	N	2,427	2,079	2,623	3,843	5,098	4,374	4,683	3,920	4,930	5,769	6,199	4,476	2,451	1,899	1,724	1,504	1,052	775
		%	4.06	3.48	4.38	6.42	8.52	7.31	7.83	6.55	8.24	9.64	10.36	7.48	4.10	3.17	2.88	2.51	1.76	1.76
12	1	N	2,474	2,222	2,914	4,240	5,240	4,332	4,710	4,245	5,186	5,970	6,085	3,991	2,227	1,757	1,757	1,419	994	667
		%	4.09	3.68	4.82	7.02	8.67	7.17	7.79	7.02	8.58	9.88	10.07	6.60	3.68	2.91	2.91	2.35	1.65	1.65
	2	N	2,469	2,147	2,793	4,108	5,225	4,307	4,750	4,059	5,119	5,923	6,149	4,197	2,306	1,820	1,762	1,448	1,017	710
		%	4.09	3.56	4.63	6.81	8.66	7.14	7.88	6.73	8.49	9.82	10.20	6.96	3.82	3.02	2.92	2.40	1.69	1.69
	3	N	2,446	2,090	2,676	3,941	5,147	4,336	4,722	3,947	5,009	5,841	6,179	4,391	2,398	1,874	1,740	1,488	1,043	754
		%	4.07	3.48	4.46	6.57	8.57	7.22	7.87	6.58	8.35	9.73	10.29	7.32	4.00	3.12	2.90	2.48	1.74	1.74
18	1	N	2,471	2,182	2,852	4,173	5,225	4,320	4,727	4,154	5,154	5,951	6,110	4,096	2,266	1,787	1,757	1,437	1,007	689
		%	4.09	3.61	4.73	6.91	8.66	7.16	7.83	6.88	8.54	9.86	10.12	6.79	3.75	2.96	2.91	2.38	1.67	1.67
	2	N	2,455	2,125	2,736	4,020	5,182	4,329	4,728	4,013	5,056	5,872	6,166	4,290	2,354	1,846	1,749	1,466	1,029	731
		%	4.08	3.53	4.55	6.68	8.62	7.20	7.86	6.67	8.41	9.76	10.25	7.13	3.91	3.07	2.91	2.44	1.71	1.71
24	1	N	2,460	2,156	2,795	4,091	5,193	4,334	4,716	4,096	5,098	5,905	6,132	4,191	2,312	1,815	1,749	1,454	1,019	711
		%	4.08	3.58	4.64	6.79	8.62	7.20	7.83	6.80	8.46	9.81	10.18	6.96	3.84	3.01	2.90	2.41	1.69	1.69

<표 3-12> 벤치마킹을 위한 모집단 정보: 거창군 남성

연령그룹		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85 -	
3	1	N	1,145	1,304	1,841	2,303	1,954	1,507	1,582	1,813	2,410	2,390	2,741	2,276	1,790	1,466	1,710	1,207	649	346
		%	3.76	4.28	6.05	7.57	6.42	4.95	5.20	5.96	7.92	7.85	9.01	7.48	5.88	4.82	5.62	3.96	2.13	1.14
	2	N	1,131	1,292	1,814	2,293	1,968	1,488	1,561	1,804	2,406	2,396	2,766	2,293	1,792	1,507	1,701	1,228	659	352
		%	3.72	4.24	5.96	7.53	6.46	4.89	5.13	5.92	7.90	7.87	9.08	7.53	5.88	4.95	5.59	4.03	2.16	1.15
	3	N	1,143	1,277	1,778	2,283	1,983	1,473	1,565	1,771	2,392	2,401	2,780	2,324	1,787	1,506	1,699	1,244	668	355
		%	3.76	4.20	5.84	7.50	6.52	4.84	5.14	5.82	7.86	7.89	9.14	7.64	5.87	4.95	5.58	4.09	2.20	1.17
	4	N	1,150	1,260	1,763	2,288	1,999	1,460	1,581	1,743	2,364	2,453	2,762	2,364	1,809	1,495	1,677	1,260	690	359
		%	3.77	4.14	5.78	7.51	6.56	4.79	5.19	5.72	7.76	8.05	9.06	7.76	5.94	4.91	5.50	4.13	2.26	1.18
	5	N	1,141	1,258	1,716	2,318	1,965	1,445	1,596	1,704	2,365	2,523	2,696	2,437	1,806	1,518	1,642	1,295	702	371
		%	3.74	4.12	5.63	7.60	6.44	4.74	5.23	5.59	7.75	8.27	8.84	7.99	5.92	4.98	5.39	4.25	2.30	1.22
	6	N	1,127	1,261	1,689	2,304	1,966	1,413	1,580	1,683	2,363	2,537	2,689	2,479	1,815	1,546	1,617	1,322	710	388
		%	3.70	4.14	5.54	7.56	6.45	4.63	5.18	5.52	7.75	8.32	8.82	8.13	5.95	5.07	5.30	4.34	2.33	1.27
	7	N	1,132	1,255	1,667	2,287	1,981	1,410	1,542	1,698	2,329	2,538	2,709	2,488	1,823	1,582	1,600	1,325	708	397
		%	3.71	4.12	5.47	7.50	6.50	4.63	5.06	5.57	7.64	8.33	8.89	8.17	5.98	5.19	5.25	4.35	2.32	1.30
	8	N	1,128	1,250	1,628	2,279	2,011	1,389	1,511	1,669	2,333	2,534	2,716	2,524	1,839	1,601	1,560	1,346	740	391
		%	3.71	4.11	5.35	7.48	6.60	4.56	4.96	5.48	7.66	8.32	8.92	8.29	6.04	5.26	5.12	4.42	2.43	1.28
6	1	N	1,138	1,298	1,827	2,298	1,961	1,497	1,572	1,809	2,408	2,393	2,754	2,285	1,791	1,487	1,706	1,217	654	349
		%	3.74	4.26	6.00	7.55	6.44	4.92	5.16	5.94	7.91	7.86	9.05	7.50	5.88	4.88	5.60	4.00	2.15	1.15
	2	N	1,147	1,269	1,770	2,286	1,991	1,467	1,573	1,757	2,378	2,427	2,771	2,344	1,798	1,501	1,688	1,252	679	357
		%	3.77	4.17	5.81	7.50	6.54	4.82	5.16	5.77	7.81	7.97	9.10	7.70	5.90	4.93	5.54	4.11	2.23	1.17
	3	N	1,134	1,260	1,703	2,311	1,966	1,429	1,588	1,694	2,364	2,530	2,692	2,458	1,810	1,532	1,630	1,308	706	380
		%	3.72	4.13	5.58	7.58	6.45	4.69	5.21	5.55	7.75	8.30	8.83	8.06	5.94	5.02	5.34	4.29	2.32	1.24
	4	N	1,130	1,253	1,648	2,283	1,996	1,399	1,526	1,683	2,331	2,536	2,713	2,506	1,831	1,592	1,580	1,336	724	394
		%	3.71	4.11	5.41	7.49	6.55	4.59	5.01	5.53	7.65	8.32	8.91	8.23	6.01	5.23	5.19	4.39	2.38	1.29
12	1	N	1,142	1,283	1,799	2,292	1,976	1,482	1,572	1,783	2,393	2,410	2,762	2,314	1,795	1,494	1,697	1,235	667	353
		%	3.75	4.22	5.91	7.53	6.49	4.87	5.16	5.85	7.86	7.92	9.07	7.60	5.89	4.91	5.57	4.06	2.19	1.16
	2	N	1,140	1,264	1,737	2,298	1,978	1,448	1,580	1,725	2,371	2,479	2,732	2,401	1,804	1,517	1,659	1,280	693	368
		%	3.74	4.15	5.70	7.54	6.49	4.75	5.19	5.66	7.78	8.13	8.96	7.88	5.92	4.98	5.44	4.20	2.27	1.21
	3	N	1,132	1,256	1,675	2,297	1,981	1,414	1,557	1,688	2,347	2,533	2,703	2,482	1,821	1,562	1,605	1,322	715	387
		%	3.71	4.12	5.50	7.54	6.50	4.64	5.11	5.54	7.70	8.31	8.87	8.14	5.97	5.13	5.27	4.34	2.35	1.27
18	1	N	1,140	1,275	1,767	2,298	1,973	1,464	1,577	1,753	2,383	2,450	2,739	2,362	1,800	1,507	1,674	1,259	680	362
		%	3.74	4.19	5.80	7.54	6.48	4.81	5.18	5.75	7.82	8.04	8.99	7.75	5.91	4.95	5.50	4.13	2.23	1.19
	2	N	1,137	1,260	1,707	2,293	1,984	1,432	1,562	1,711	2,358	2,498	2,725	2,436	1,813	1,542	1,632	1,299	703	377
		%	3.73	4.14	5.60	7.53	6.51	4.70	5.13	5.62	7.74	8.20	8.95	8.00	5.95	5.06	5.36	4.26	2.31	1.24
24	1	N	1,137	1,270	1,737	2,294	1,978	1,448	1,565	1,736	2,370	2,472	2,732	2,398	1,808	1,528	1,651	1,278	691	370
		%	3.73	4.17	5.70	7.53	6.49	4.75	5.14	5.70	7.78	8.11	8.97	7.87	5.93	5.02	5.42	4.20	2.27	1.21

<표 3-13> 벤치마킹을 위한 모집단 정보: 거창군 여성

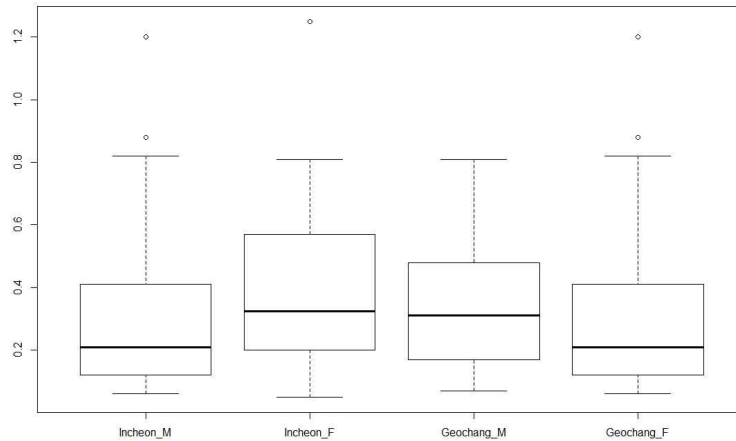
연령그룹		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85 -	
3	1	N	1,027	1,138	1,641	1,871	1,464	1,251	1,533	1,725	2,171	2,169	2,731	2,294	2,057	1,950	2,604	2,228	1,398	1,015
		%	3.18	3.53	5.08	5.80	4.54	3.88	4.75	5.35	6.73	6.72	8.46	7.11	6.37	6.04	8.07	6.90	4.33	3.15
	2	N	1,042	1,131	1,600	1,858	1,480	1,234	1,533	1,660	2,200	2,171	2,730	2,322	2,058	1,939	2,592	2,237	1,420	1,039
		%	3.23	3.51	4.96	5.76	4.59	3.83	4.76	5.15	6.82	6.73	8.47	7.20	6.38	6.01	8.04	6.94	4.40	3.22
	3	N	1,041	1,135	1,575	1,848	1,494	1,216	1,519	1,634	2,181	2,210	2,714	2,355	2,036	1,960	2,574	2,245	1,449	1,046
		%	3.23	3.52	4.89	5.73	4.64	3.77	4.71	5.07	6.77	6.86	8.42	7.31	6.32	6.08	7.99	6.97	4.50	3.25
	4	N	1,030	1,164	1,536	1,852	1,498	1,206	1,527	1,606	2,156	2,272	2,668	2,414	2,016	1,977	2,557	2,257	1,479	1,056
		%	3.19	3.61	4.76	5.74	4.64	3.74	4.73	4.98	6.68	7.04	8.27	7.48	6.25	6.13	7.92	6.99	4.58	3.27
	5	N	1,035	1,188	1,514	1,843	1,498	1,205	1,526	1,603	2,163	2,300	2,624	2,489	2,003	1,975	2,528	2,303	1,503	1,081
		%	3.20	3.67	4.68	5.69	4.63	3.72	4.71	4.95	6.68	7.10	8.10	7.69	6.18	6.10	7.81	7.11	4.64	3.34
	6	N	1,023	1,196	1,501	1,838	1,501	1,176	1,505	1,612	2,136	2,342	2,571	2,540	2,029	1,962	2,498	2,347	1,525	1,105
		%	3.16	3.69	4.63	5.67	4.63	3.63	4.64	4.97	6.59	7.23	7.93	7.84	6.26	6.05	7.71	7.24	4.71	3.41
	7	N	1,022	1,184	1,488	1,847	1,518	1,170	1,490	1,603	2,132	2,353	2,573	2,549	2,044	1,969	2,482	2,365	1,543	1,115
		%	3.15	3.65	4.59	5.69	4.68	3.61	4.59	4.94	6.57	7.25	7.93	7.86	6.30	6.07	7.65	7.29	4.76	3.44
	8	N	1,032	1,172	1,463	1,859	1,522	1,157	1,473	1,604	2,095	2,367	2,549	2,565	2,055	1,971	2,434	2,391	1,557	1,131
		%	3.18	3.62	4.51	5.74	4.70	3.57	4.55	4.95	6.47	7.30	7.87	7.92	6.34	6.08	7.51	7.38	4.81	3.49
6	1	N	1,034	1,135	1,620	1,865	1,472	1,242	1,533	1,692	2,186	2,170	2,731	2,308	2,058	1,945	2,598	2,232	1,409	1,027
		%	3.21	3.52	5.02	5.78	4.56	3.85	4.75	5.25	6.78	6.73	8.47	7.16	6.38	6.03	8.05	6.92	4.37	3.18
	2	N	1,036	1,149	1,556	1,850	1,496	1,211	1,523	1,620	2,169	2,241	2,691	2,384	2,026	1,969	2,565	2,251	1,464	1,051
		%	3.21	3.56	4.82	5.74	4.64	3.75	4.72	5.02	6.72	6.95	8.34	7.39	6.28	6.10	7.95	6.98	4.54	3.26
	3	N	1,029	1,192	1,507	1,840	1,500	1,190	1,516	1,607	2,149	2,321	2,598	2,514	2,016	1,969	2,513	2,325	1,514	1,093
		%	3.18	3.68	4.65	5.68	4.63	3.67	4.68	4.96	6.64	7.16	8.02	7.76	6.22	6.08	7.76	7.18	4.67	3.37
	4	N	1,027	1,178	1,475	1,853	1,520	1,164	1,482	1,604	2,114	2,360	2,561	2,557	2,049	1,970	2,458	2,378	1,550	1,123
		%	3.17	3.63	4.55	5.72	4.69	3.59	4.57	4.95	6.52	7.28	7.90	7.89	6.32	6.08	7.58	7.33	4.78	3.46
12	1	N	1,035	1,142	1,588	1,857	1,484	1,227	1,528	1,656	2,177	2,205	2,711	2,346	2,042	1,957	2,582	2,242	1,437	1,039
		%	3.21	3.54	4.92	5.76	4.60	3.80	4.74	5.13	6.75	6.84	8.40	7.27	6.33	6.07	8.00	6.95	4.45	3.22
	2	N	1,032	1,171	1,532	1,845	1,498	1,201	1,519	1,614	2,159	2,281	2,644	2,449	2,021	1,969	2,539	2,288	1,489	1,072
		%	3.19	3.62	4.74	5.71	4.63	3.71	4.70	4.99	6.68	7.06	8.18	7.58	6.25	6.09	7.86	7.08	4.61	3.32
	3	N	1,028	1,185	1,491	1,847	1,510	1,177	1,499	1,605	2,132	2,340	2,580	2,536	2,033	1,969	2,485	2,352	1,532	1,108
		%	3.17	3.66	4.60	5.70	4.66	3.63	4.62	4.95	6.58	7.22	7.96	7.82	6.27	6.08	7.67	7.26	4.73	3.42
18	1	N	1,033	1,159	1,561	1,852	1,489	1,215	1,524	1,640	2,168	2,244	2,673	2,402	2,033	1,961	2,559	2,269	1,462	1,057
		%	3.20	3.59	4.83	5.73	4.61	3.76	4.72	5.08	6.71	6.95	8.28	7.44	6.29	6.07	7.92	7.03	4.53	3.27
	2	N	1,030	1,173	1,513	1,848	1,505	1,188	1,507	1,610	2,144	2,307	2,617	2,485	2,030	1,969	2,512	2,318	1,509	1,089
		%	3.18	3.63	4.68	5.71	4.65	3.67	4.66	4.98	6.63	7.13	8.09	7.68	6.28	6.09	7.76	7.16	4.66	3.37
24	1	N	1,031	1,163	1,540	1,852	1,497	1,202	1,513	1,631	2,154	2,273	2,645	2,441	2,037	1,963	2,534	2,297	1,484	1,074
		%	3.19	3.60	4.76	5.73	4.63	3.72	4.68	5.04	6.66	7.03	8.18	7.55	6.30	6.07	7.84	7.10	4.59	3.32

<표 3-10>~<표 3-13>은 최종가중치 산출을 위하여 사용한 모집단 정보를 나타낸다. 시험조사 결과의 분석을 위해 필수적인 가중치 산출을 위하여 사용한 모집단 정보는 각 시범지역의 성*연령그룹으로 정의된 사후층 분포이다. 이는 식 (2.4)에 주어진 회귀가중치를 적용하기 위한 보조정보로 다음의 식을 고려한 경우이다.

$$\mathbf{x} = (x_1, x_2, \dots, x_k), \quad x_j = \begin{cases} 1, & \text{가구원이 사후층 } j \text{에 속한 경우} \\ 0, & \text{기타} \end{cases}$$

표에 주어진 값은 행정자치부의 주민등록자료로부터 조사 해당 월별 자료를 이용하여 계산된 값이다. 즉 2013년 1월- 3월까지의 통합자료의 벤치마킹 정보를 얻기 위해서는 각 월 별 성*연령그룹별 주민등록 인구를 파악하고 이의 평균을 구하여 벤치마킹을 위한 정보를 계산하였다.

[그림 3-11] 조사시점 별 모집단 사후층 비율 차이 분포

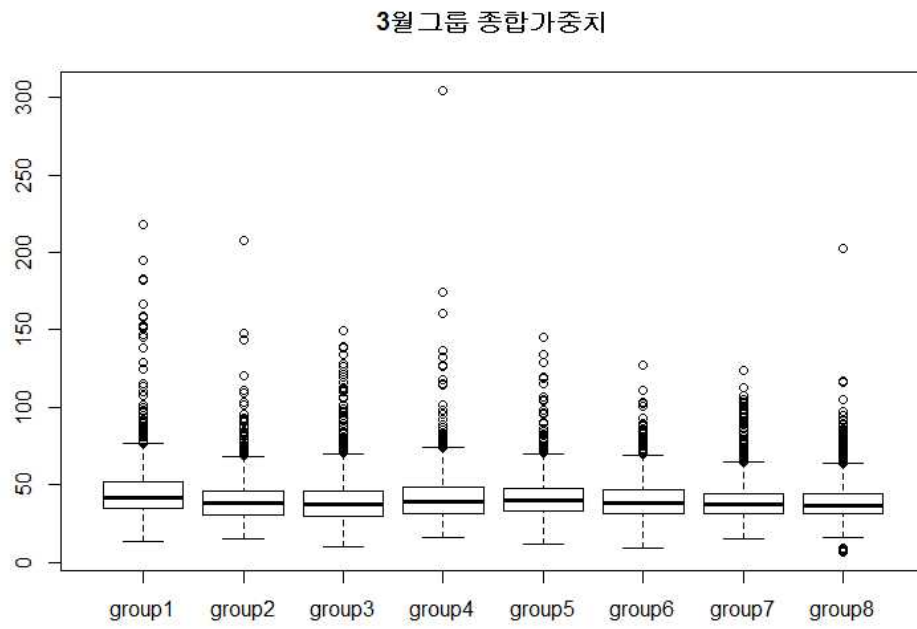


위 그림은 18개 사후층(연령)에서 계산된 18개 그룹 비율(%)의 최댓값에서 최솟값을 뺀 값의 요약 통계량을 나타낸다. 각 개월 그리고 그룹 별 분포를 비교하면 최대 약 1% 정도의 차이가 나는 사후층이 있으나 대부분의 경우는 0.1%-0.6% 혹은 그 미만의 변동을 나타내고 있다. 즉 모집단 분포의 월 간, 분기 간, 반기 간 혹은 년 간 변화는 매우 적은 것으로 판단할 수 있다. 다만, 각 연령그룹 및 지역 특성(도시 혹은 비도시 지역)에 따른 유동 인구의 비중은 상당한 차이가 있을 수 있으니 이 부분 역시 벤치마킹을 위한 모집단 정보 생성 시 고려하여야 할 것이다.

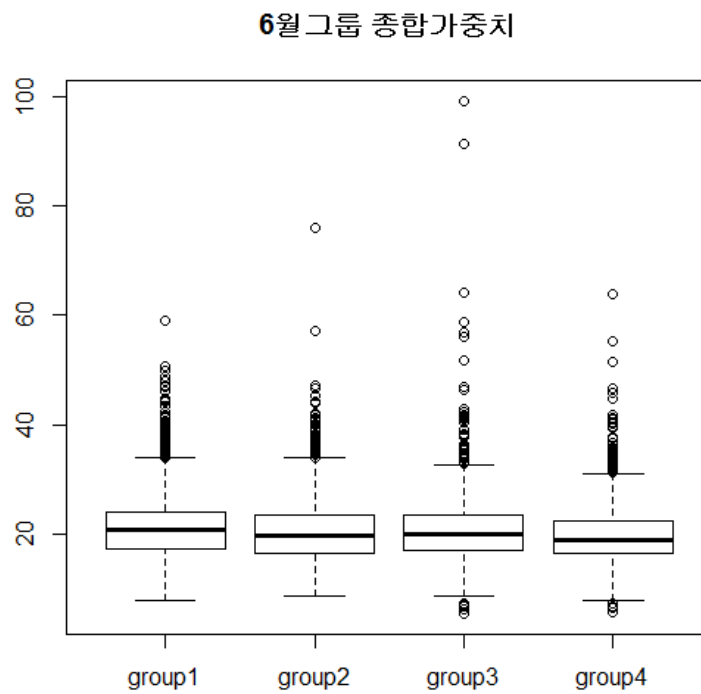
<표 3-14> 최종가중치 분포

개월	그룹	인천							거창						
		기초통계량							기초통계량						
		mean	st.dev	min	Q1	median	Q3	max	mean	st.dev	min	Q1	median	Q3	max
3	1	44.380	16.363	13.830	13.830	41.960	51.650	218.400	38.290	23.349	9.010	24.070	34.500	44.760	251.000
	2	40.080	13.802	14.760	30.900	37.970	45.880	207.700	40.670	22.376	7.538	23.770	38.000	52.160	52.160
	3	40.120	15.484	10.410	30.040	37.340	46.170	149.300	35.110	27.191	6.204	26.170	35.110	47.020	374.200
	4	41.160	14.975	16.400	31.330	39.380	48.380	304.300	40.430	26.790	10.850	20.480	33.560	51.800	286.200
	5	41.820	13.340	11.830	33.050	39.540	48.020	145.300	40.390	21.858	7.905	25.050	36.530	50.560	162.000
	6	40.350	13.187	9.164	31.220	38.280	46.470	127.200	42.370	23.524	11.420	25.830	38.550	52.280	225.000
	7	39.430	12.699	15.370	31.240	37.480	44.660	123.800	41.170	26.140	7.429	24.300	37.180	50.370	242.200
	8	39.060	12.545	6.857	31.280	36.630	44.280	203.000	39.030	24.592	7.096	22.770	36.240	47.630	266.200
6	1	21.090	5.571	7.929	17.290	20.630	23.960	59.020	19.830	10.569	4.676	12.490	18.740	24.270	130.200
	2	20.380	5.595	8.576	16.450	19.790	23.490	76.100	20.450	13.150	4.737	12.630	18.300	24.770	228.300
	3	20.560	5.402	5.462	17.020	19.960	23.310	99.190	20.890	10.748	6.214	13.060	18.670	26.160	172.000
	4	19.640	4.856	5.662	16.410	18.890	22.260	63.940	20.250	12.219	4.312	12.420	18.900	25.240	129.100
12	1	10.360	2.399	4.968	8.672	10.100	11.660	30.450	10.090	5.397	2.852	2.852	9.359	12.110	76.020
	2	10.230	2.297	3.441	8.690	10.020	11.520	26.430	10.360	5.427	2.872	2.872	9.549	12.590	91.790
	3	10.060	2.189	3.467	8.548	9.829	11.270	29.150	10.310	5.594	2.332	6.565	9.534	12.720	66.500
18	1	6.893	1.450	3.029	5.932	6.731	7.702	17.920	6.812	3.401	2.089	4.476	6.346	8.239	52.790
	2	6.730	1.450	2.519	5.776	6.624	7.518	15.550	6.852	3.698	1.533	4.424	6.382	8.367	45.560
24	1	5.103	1.010	2.104	4.438	5.009	5.103	12.130	5.097	2.661	1.301	3.280	4.752	6.217	32.000

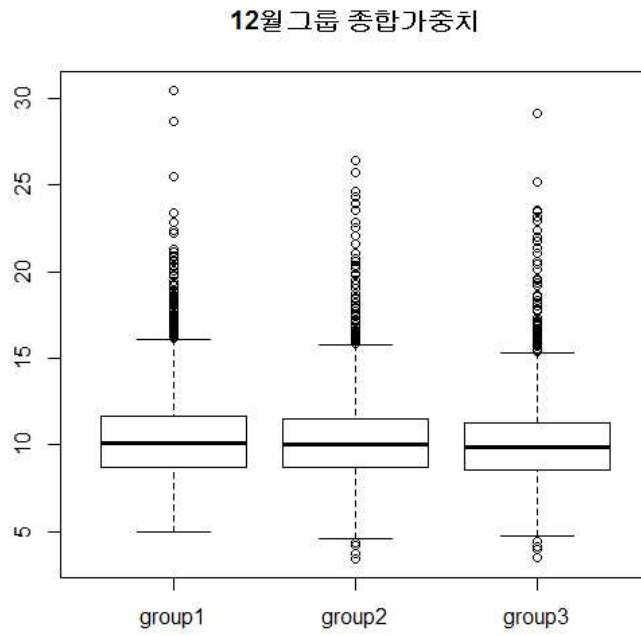
[그림 3-12] 인천 3개월 그룹 최종 가중치



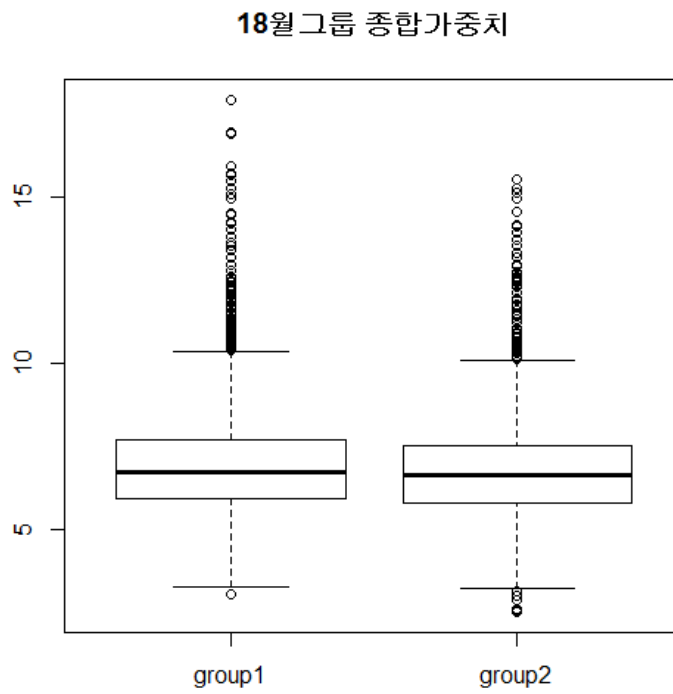
[그림 3-13] 인천 6개월 그룹 최종 가중치



[그림 3-14] 인천 12개월 그룹 최종 가중치

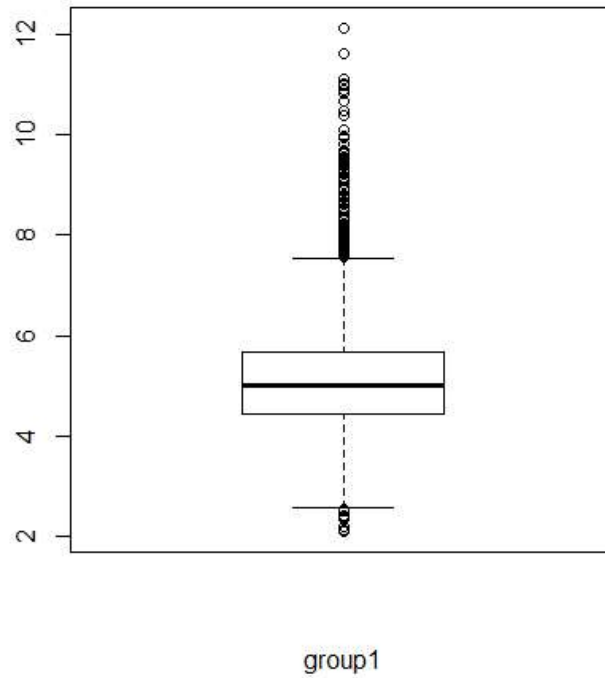


[그림 3-15] 인천 18개월 그룹 최종 가중치



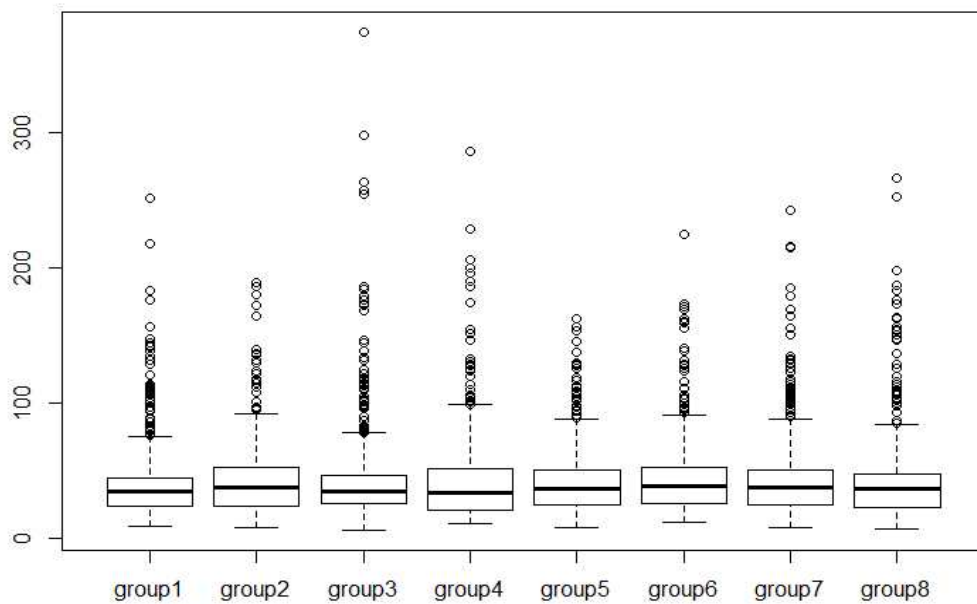
[그림 3-16] 인천 24개월 그룹 최종 가중치

24월그룹 종합가중치

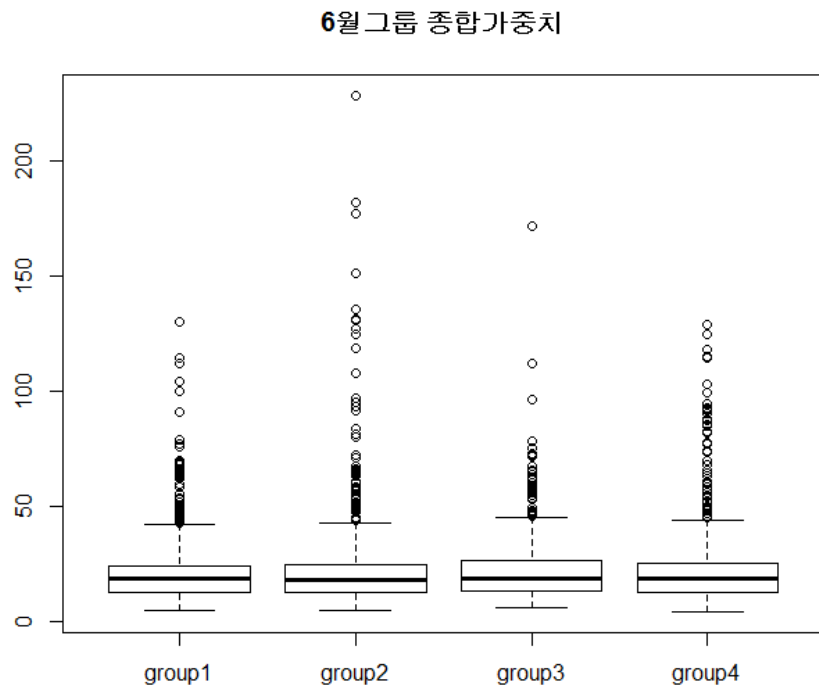


[그림 3-17] 거창 3개월 그룹 최종 가중치

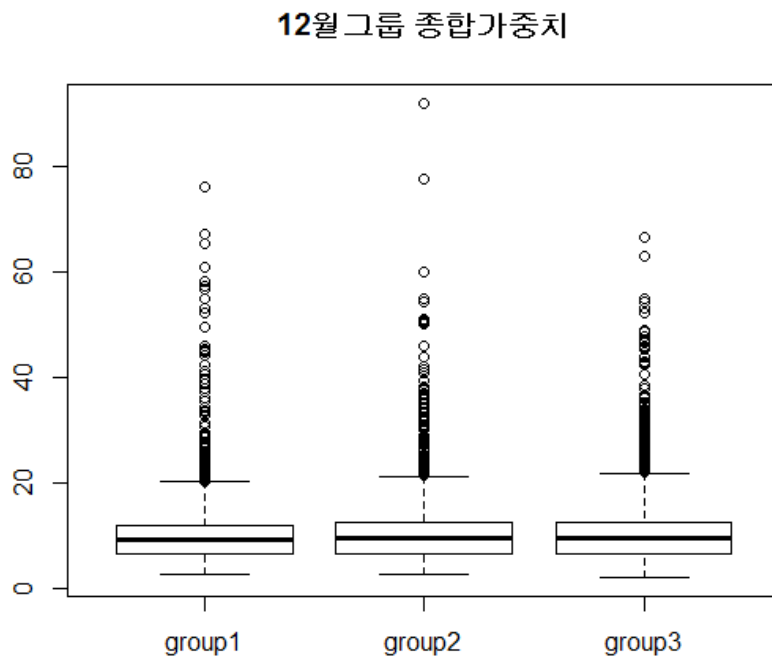
3월그룹 종합가중치



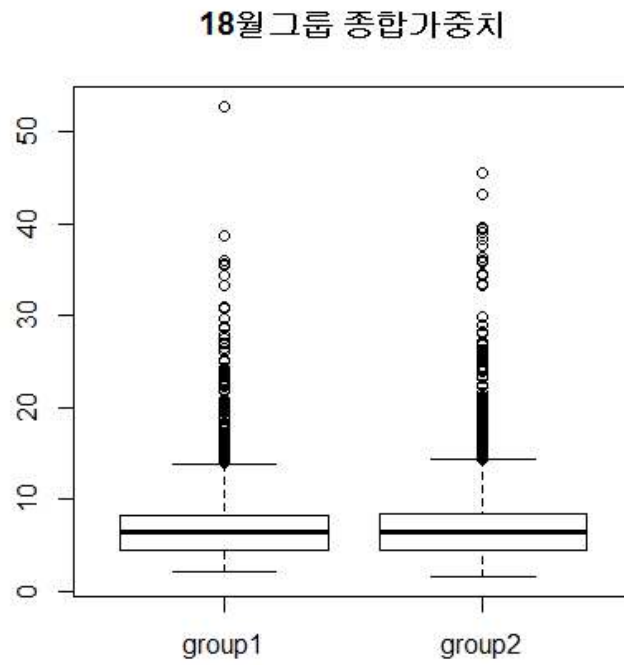
[그림 3-18] 거창 6개월 그룹 최종 가중치



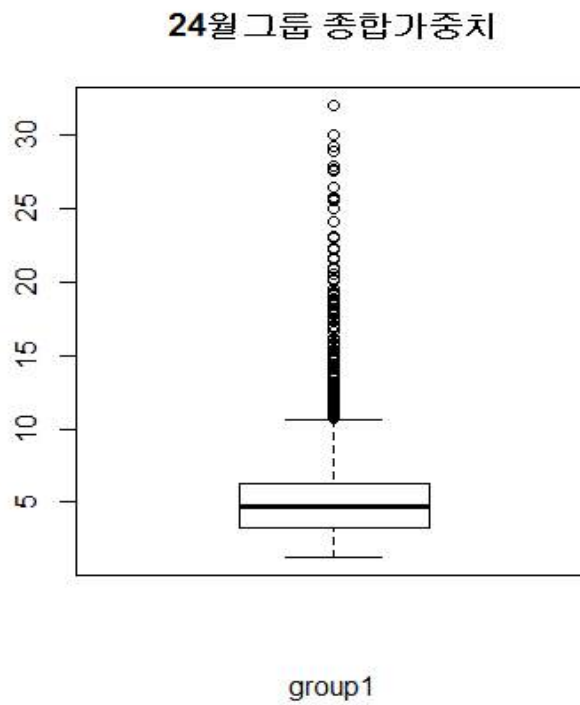
[그림 3-19] 거창 12개월 그룹 최종 가중치



[그림 3-20] 거창 18개월 그룹 최종 가중치



[그림 3-21] 거창 24개월 그룹 최종 가중치

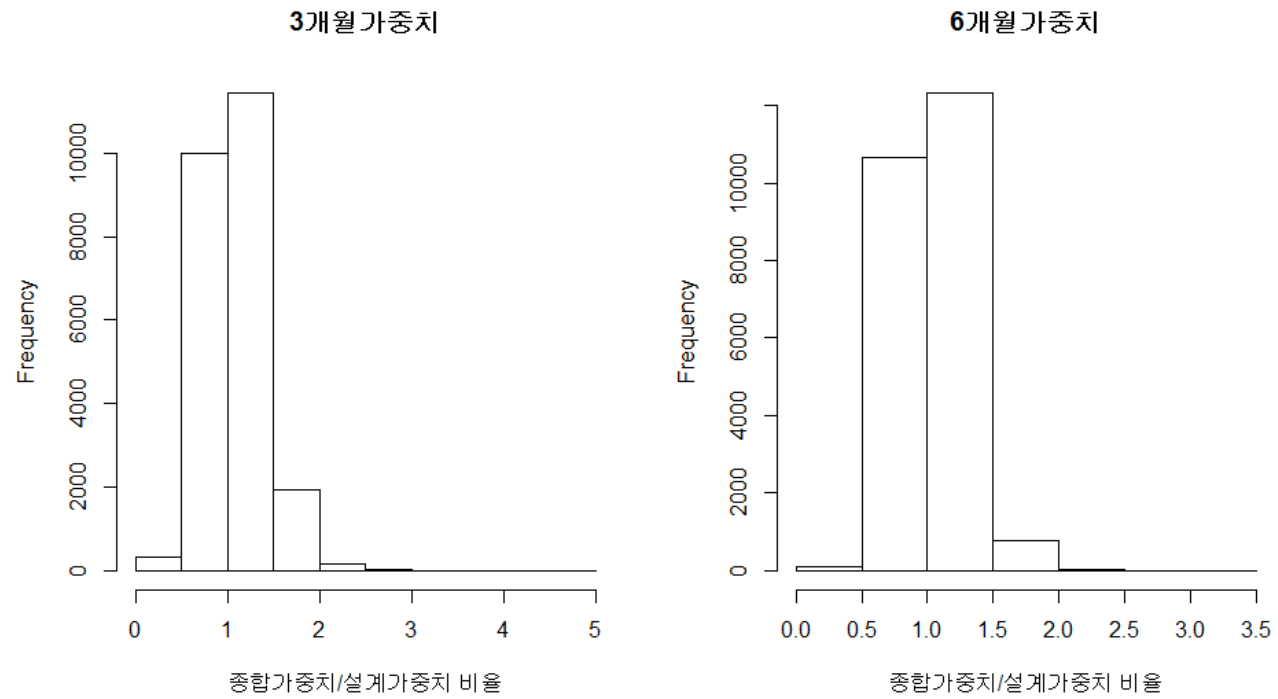


<표 3-14>와 [그림 3-12]~[그림 3-21]은 회귀추정방안을 통하여 산출되는 최종가중치의 분포를 나타내고 있다. 설계가중치의 분포에 비하여 전반적으로 가중치의 분산이 크게 나타나는 것을 확인할 수 있다. 이는 조사구 혹은 조사구 내 가구 표본을 추출할 때 각 조사구내 인구특성 변수를 고려하지 않았기 때문으로 여겨진다. 시험조사를 위해 사용된 표집틀은 언급한 바와 같이 각 조사구 내 평균 가구 수는 60가구이나 실제 조사구 별로 가구 수의 크기가 다른 경우가 발생하고, 그로 인하여 인구의 성별 그리고 연령별 분포가 모집단의 분포와 차이를 보이고 있는 것으로 판단된다. 회귀추정 혹은 사후층화 방안을 적용하여 산출되는 최종가중치의 분산을 가능한 한 최소화 하는 것이 최종 추정량의 분산을 최소화 하는 것임을 고려할 때 새롭게 구성된 조사구 표집틀로부터의 표본추출을 다시 실시할 경우에는 2015년 인구주택총조사 표본조사 결과를 통해 인구 특성 별 모집단 분포와 표본분포의 차이를 파악하고 이를 표본설계에 반영할 필요가 있다.

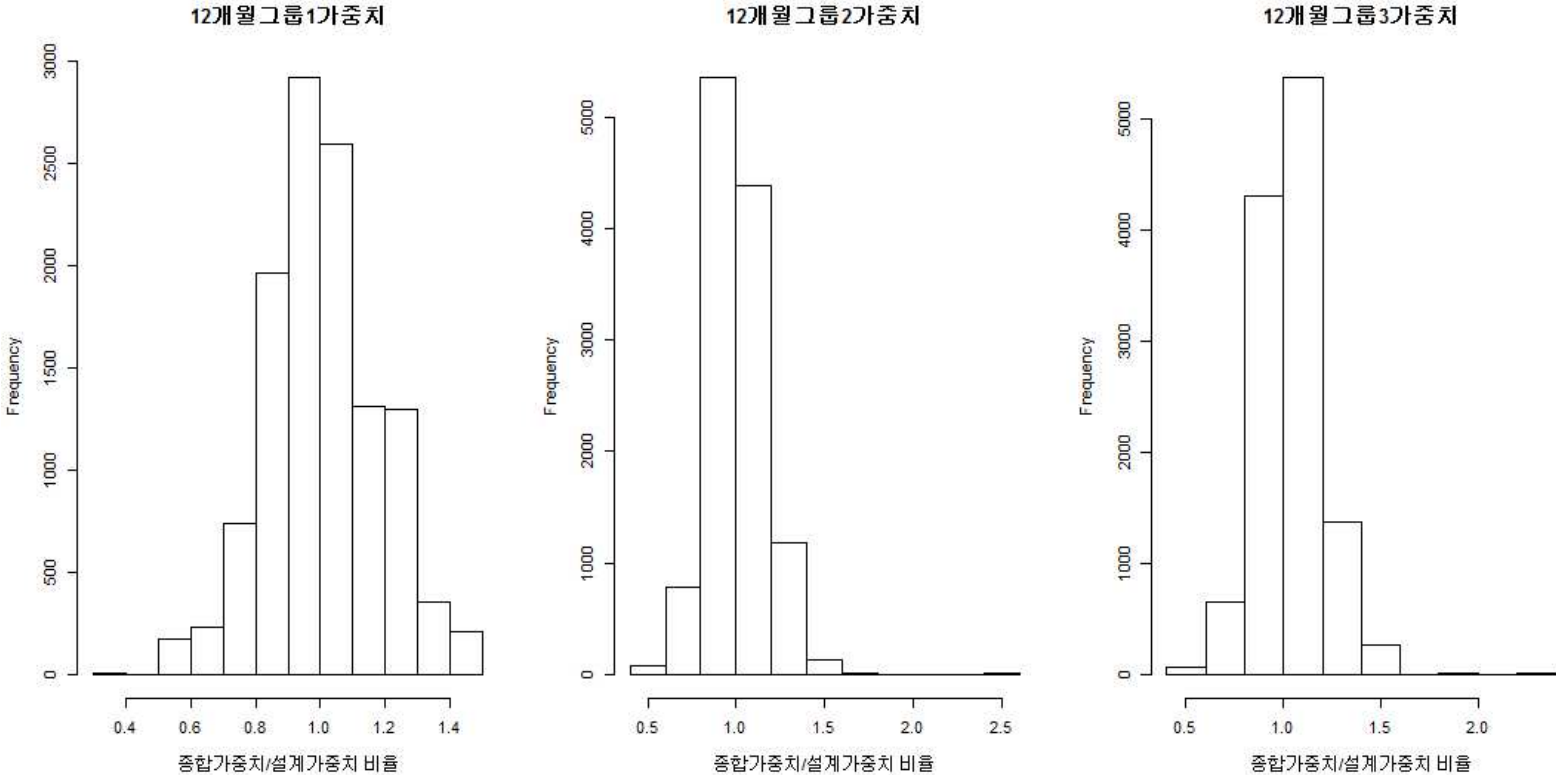
최종가중치 작성을 위해 사용한 가중치 조정 방안의 영향을 파악하기 위하여 설계가중치 대비 최종가중치의 비율을 산출하였다. [그림 3-22]~[그림 3-27]은 계산된 상대 비율의 히스토그램을 나타내고 있다. 통합을 위해 사용된 개월 수가 낮을수록 그 비율이 상대적으로 크고 분산 역시 커지고 있음을 확인할 수 있다. 전체 조사구 중 약 20%의 조사구가 분석을 위해 사용되는 12개월 통합 자료의 경우, 인천은 약 최대 3배정도 가중치가 증가하고 있고 거창군의 경우 최대 9배까지 가중치의 변동이 발생하는 것을 알 수 있다. 두 시범지역 모두에서 가중치 변동량의 평균은 1로써 평균적으로는 설계가중치의 변동이 크지 않게 나타나고 있으나, 특정 사후층에서는 표본 기반 추정치와 모집단의 차이가 다른 층에 비하여 크다.

조사구 간 가구 수 차이 그리고 이로 인한 인구학적 변인 기준의 차이가 최종가중치의 분산이 커지는 중요한 원인이나 이와 더불어 현재 표본추출은 기존 2010년 인구주택총조사를 이용하여 구축된 표집틀이 사용되었고 실제 벤치마킹을 위한 모집단 자료는 주민등록자료가 사용되었기 때문이 두 자료상의 차이 역시 본 결과에 영향을 준 것으로 판단된다. 실제 순환표본조사가 이루어진다면 벤치마킹 자료는 행정자료의 수정 후 작성되는 등록센서스 자료로부터 산출되어야 하며 이 경우에는 본 연구에서 나타난 가중치 변동의 폭은 상당히 감소할 것으로 예상할 수 있다.

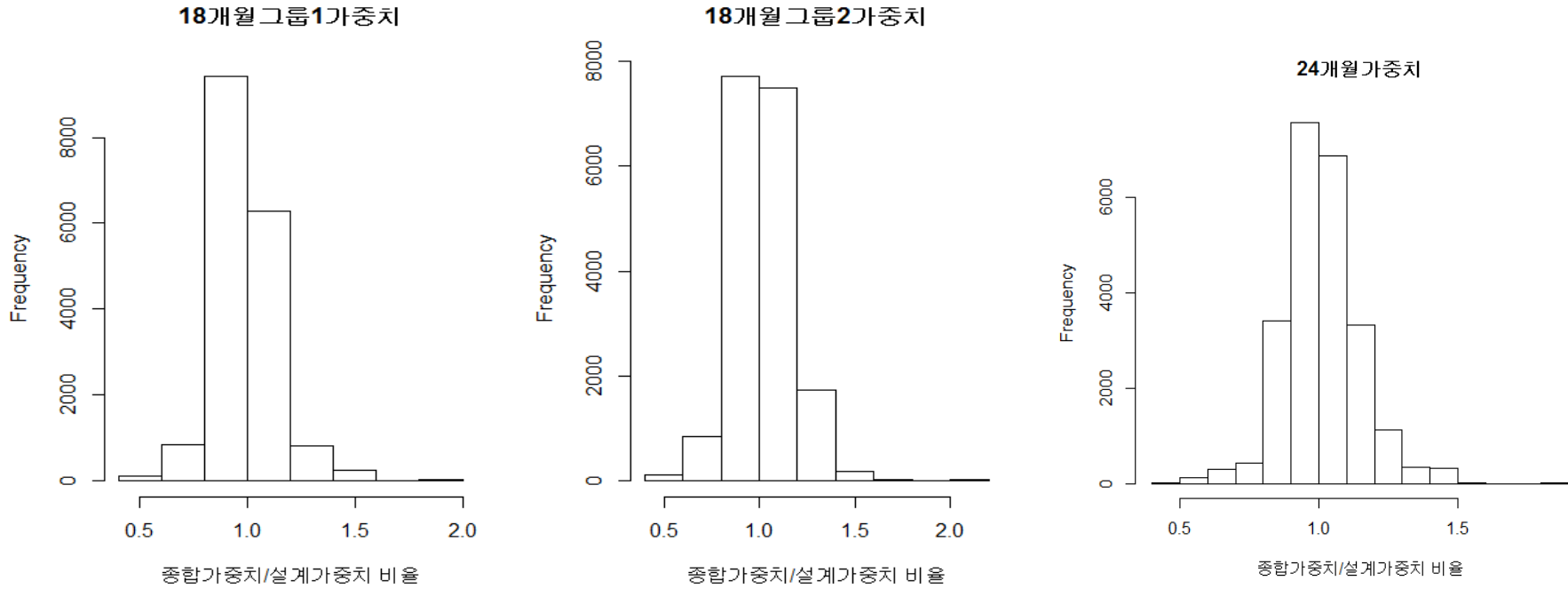
[그림 3-22] 인천 3개월 및 6개월 최종가중치 대 설계가중치 비율



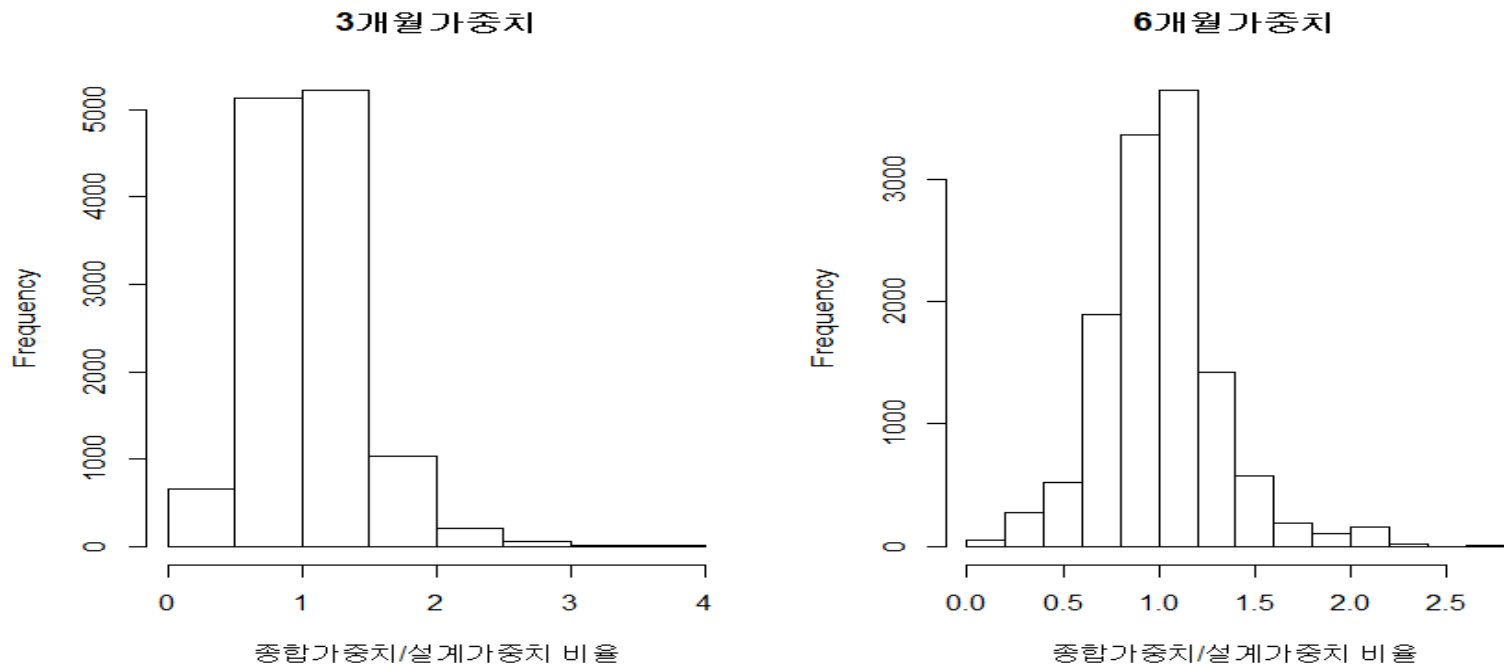
[그림 3-23] 인천 12개월 최종가중치 대 설계가중치 비율



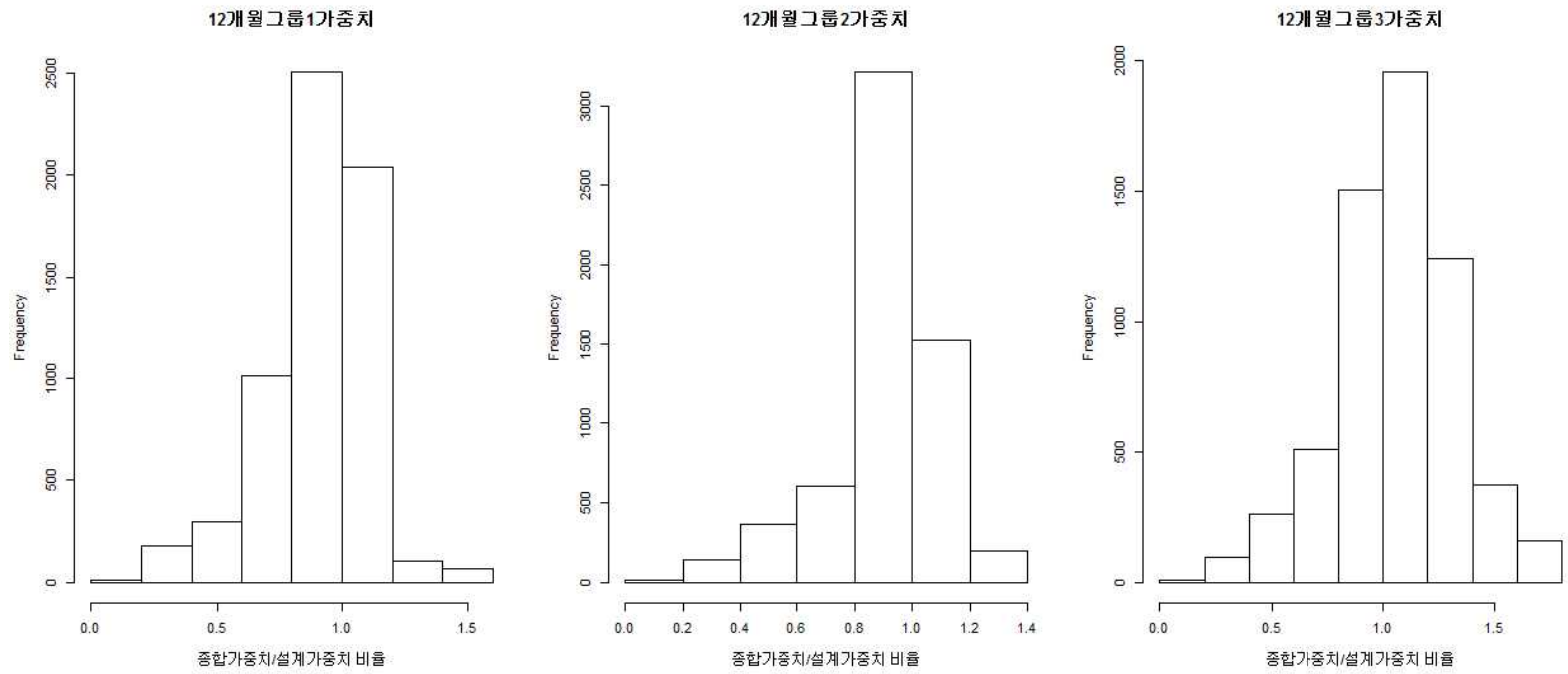
[그림 3-24] 인천 18개월 및 24개월 최종가중치 대 설계가중치 비율



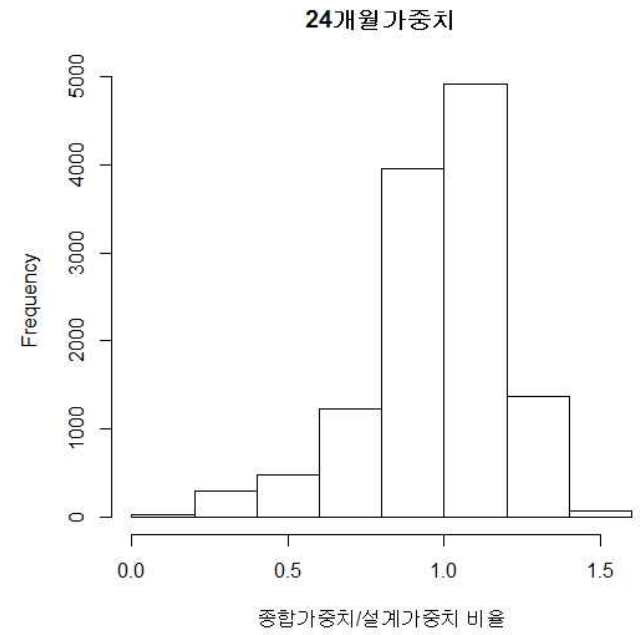
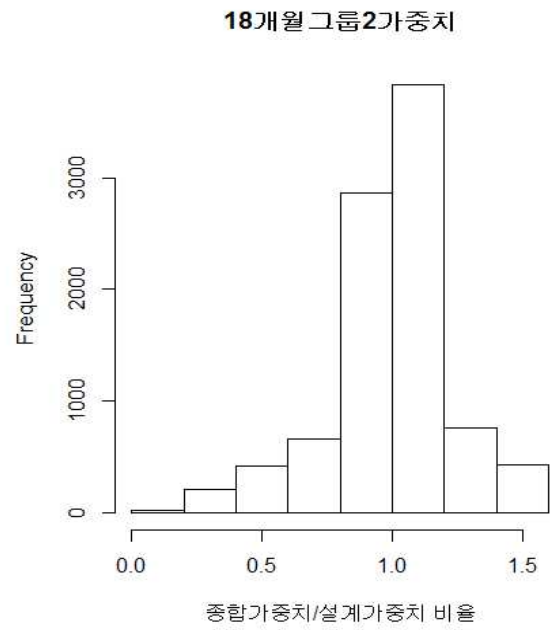
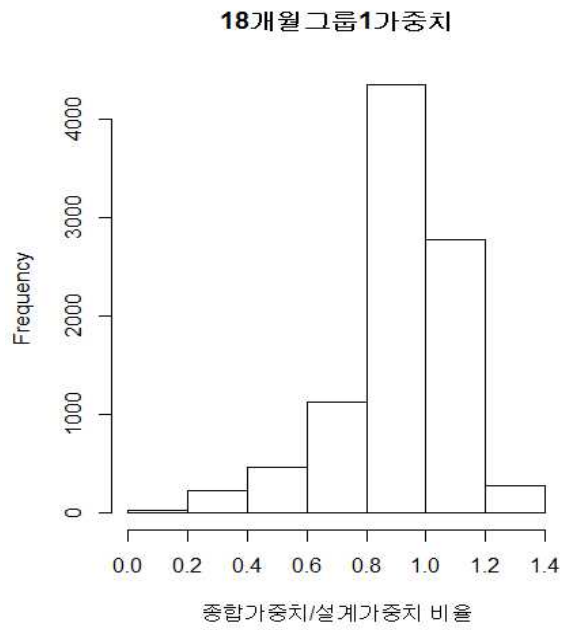
[그림 3-25] 가창 3개월 및 6개월 최종가중치 대 설계가중치 비율



[그림 3-26] 가창 12개월 최종가중치 대 설계가중치 비율



[그림 3-27] 가창 18개월 및 24개월 최종가중치 대 설계가중치 비율



4. 기간 추정을 위한 기타 고려사항

기간추정을 위한 가중치 작성 방안 이외에 가중치와 무관하게 고려해야 하는 사항들이 있다. 여기에는 변수들의 표준화, 지역 혹은 구역 변화에 따른 지역 코드 갱신, 그리고 인플레이션과 같은 경기 변동으로 인한 원화가치의 변화의 반영 등이 포함된다. 이에 대한 내용은 통계청(2014)의 기존 연구 3절에서 이미 세부적으로 연구된 바 있다. 본 절에서는 이러한 내용을 간략하게 정리한다.

기간추정을 위해 사전에 수정되어야 할 부분은 각 조사 문항들로 먼저는 누적 시점 조사 변수에 대한 문항 수정이 필요하다. 예를 들어, 2014년에 수행된 시험조사 설문지의 3번 나이 부분의 “집에서 세는 나이는 몇 살입니까?”와 26번 총 출생아수 부분의 “지금까지 낳은 자녀는 모두 몇 명입니까?”의 질문이 여기에 해당한다. 나이와 평생에 낳은 자녀의 수는 시간이 지남에 따라 그 값이 증가하기 때문에 기간추정 시에는 이러한 문제를 해결하기 위해 문항 자체를 수정해야 한다. 나이의 경우에는 실제 생일을 직접적으로 묻는 것이 바람직하며 기타 누적 시점 변수의 경우에는 해당 기간을 명시하는 것이 바람직하다. 누적 시점 변수에 대응하는 ACS의 설문문항들의 예를 살펴보면 다음과 같다.

- ACS Q10_A) At any time IN THE LAST 3 MONTHS, has this person attended school or college?
- ACS Q21. In the PAST 12 MONTHS did this person get married
- ACS Q24. Has this person given birth to any children in the PAST 12 MONTHS?
- ACS Q36. During the LAST 4 WEEKS, has this person been actively looking for work?
- ACS Q39_A) During the PAST 12 MONTHS (52 weeks), did this person work 50 or more weeks?
- ACS Q40. During the PAST 12 MONTHS, in the weeks worked, how many hours did this person usually work each week?

각 조사시점 별 수집된 자료의 병합을 통한 기간 추정을 위해서는 언급한 문항들, 특히 누적 시점 문항을 수정하여 조사를 수행하거나 관측된 값들을 분석시점에 맞도록 수정한 후에 자료를 통합하고 분석해야 한다. 이러한 설문 문항 상의 문제가 해결된 이후 자료의 통합과정에서 고려해야 하는 사항들은 다음과 같다.

- ① 거주지, 근무지 등의 지역 정보 정의를 위한 지리정보의 갱신
- ② 인플레이션 조정

지리정보는 통계청 내에서 행정구역 변동, 신규 개발지역 정보 등을 이용하여 전체 기간 중 공표 시점에 맞추어서 통합자료의 지리 정보를 갱신할 수 있을 것이다. 인플레이션 조정 역시 원화 가치의 변동에 따라 조사 기간 내 공표 시점에 맞추어 조정하여 기간 추정을 위한 데이터를 구성할 수 있다.

위의 내용은 인구주택총조사 표본조사 항목들에 국한하여 제안된 것이다. 즉 현재와 동일한 문항과 내용들이 순환표본조사를 통해 조사될 경우에 해당하며 추가적인 문항들이 포함된다면 추가 문항의 적절성 여부에 대한 검토가 우선적으로 이루어져야 할 것이다.

제4장 시험조사 자료의 심층 분석

제1절 서론

통계청은 2012년 8월 순환센서스 도입을 검토하기 위한 경상남도 거창군 시험조사 계획을 수립하여 2012년 10월부터 시험조사를 실시하였고, 도시지역의 특성을 파악하기 위해 2013년 3월부터 인천광역시 연수구의 5개동에 대해 시험조사를 실시하였다. 이 장에서는 인천광역시 연수구 및 경상남도 거창군의 시험조사 자료를 심층 분석함으로써 주요 항목별 변화 추이 및 시험조사에서의 무응답 현황 등 문제점을 살펴보고자 한다. 또한 시험조사 자료를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 시군구 및 읍면동 단위의 시기별 공표 가능한 수준을 파악하고자 한다. 먼저, 이절에서는 경상남도 거창군과 인천광역시 연수구(5개동)의 시험조사 현황에 대하여 간략히 서술한다.

1. 경상남도 거창군의 시험조사

통계청은 2012년 8월 순환센서스 도입을 검토하기 위한 시험조사 계획을 수립하였고, 2012년 10월부터 2014년 9월까지 경상남도 거창군에 대해 시험조사를 실시하였다. 2012년 10월~2012년 12월 조사대상 가구는 조사초기에 조사 내용이 안정적이지 않아 2014년 10월~2014년 12월에 다시 조사를 실시하였다(조사주기: 매월, 조사기준: 매월 1일 0시 현재).

가. 가창군 모집단 현황

거창군은 1개 읍과 11개 면으로 구성되어 있으며, 2010년 인구주택총조사 기준 전체 가구는 23,459가구이고 인구(가구원수)는 54,916명이다. 2014년 5월 현재 추계 인구는 60,962명으로 4년 전보다 약 6천 명 정도 증가했으며, 추계가구는 25,510가구로 약 2천 가구 정도 증가하였다. 2010년 기준 거창군 전체 조사구(454개) 중 시험조사 모집단인 일반조사구는 총 425개(93.6%)이며, 보통조사구(345개)가 아파트조사구(80개)보다 약 4배가량 많다.⁸⁾

나. 표본조사구 및 표본가구 추출방법

8) 아파트 조사구는 거창읍에만 존재한다.

추출대상은 일반조사구 중 30가구 이상인 421개 조사구이며(4개 조사구는 조사구 내 가구 수가 30가구 미만임), 면지역의 조사구수가 적어 3~4개의 인접 면끼리 묶어 총 60개 조사구 내외의 그룹을 설정하였다(A~D그룹). 표본조사구 추출은 조사구 특성, 조사구내 가구 수, 15-64세 인구 구성비, 행정구역(읍면동) 순으로 정렬하여 계통추출하였으며, 매월 8개 조사구씩 24개월분을 추출하였다(총 192개 조사구). 표본조사구 각 그룹별로는 매월 최소 1개의 조사구, 각 면별로는 6개월 동안 최소 2개의 조사구가 포함되도록 동일 특성 조사구내에서 조정하였다. 특히, 거창읍에만 있는 아파트조사구는 특정 월에 치우치지 않도록 매월 1~2개씩 골고루 배정하였다.

다. 시험조사 가구수 및 인구수

경상남도 거창군의 시험조사 결과 가구는 5,765가구(모집단 대비 25.4%), 인구는 12,319명(모집단 대비 21.4%)이 조사되었다. 시험조사 6개월 단위의 표본 가구수 및 인구수는 아래 표와 같다.

<표 4-1> 경상남도 거창군 시험조사의 표본 가구수 및 인구수

<가구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	22,381	1,440	6.4%	1,440	6.4%
12개월(2차)	22,545	1,436	6.4%	2,876	12.8%
18개월(3차)	22,666	1,444	6.4%	4,320	19.1%
24개월(4차)	22,709	1,445	6.4%	5,765	25.4%

<인구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	57,264	3,150	5.5%	3,150	5.5%
12개월(2차)	57,440	3,067	5.3%	6,217	10.8%
18개월(3차)	57,652	2,997	5.2%	9,214	16.0%
24개월(4차)	57,623	3,105	5.4%	12,319	21.4%

2. 인천광역시 연수구의 시험조사

통계청은 도시지역의 특성을 파악하기 위해 2013년 3월부터 2014년 12월까지 인천광역시 연수구의 5개동에 대해 시험조사를 실시하였다. 인천광역시 연수구 5개동은 최초 분기조사로 실시되었으나, 순환표본조사의 효율적 실시를 위해 2014년부터 매월조사로 전환되었다.

가. 연수구(5개동) 모집단 현황

인천광역시 연수구는 총 11개 동으로 구성되어 있고, 시험조사에는 선학동, 연수 1, 2, 3동, 청학동 등 5개동이 포함되었다. 인천광역시 연수구 5개동의 2010년 인구주택총조사 기준 전체 가구는 50,129가구이며, 인구(가구원수)는 128,545명이다. 2014년 6월 현재 추계인구는 119,384명으로 4년 전보다 약 9천명 정도 감소한 수준이고, 추계가구는 53,047가구로 약 3천가구 정도 증가하였다. 2010년 기준 인천광역시 연수구 5개동 조사구(829개) 중 시험조사 모집단인 일반조사구는 총 818개(98.7%)이며, 아파트조사구(448개; 54.8%)가 보통조사구(345개; 42.2%)보다 더 많다.

나. 표본조사구 및 표본가구 추출방법

추출대상은 전체 829개 조사 중 아파트 및 단독 등 일반조사구 818개를 대상으로 하였다. 표본조사구의 추출은 행정동, 조사구 특성, 조사구내 가구 수, 15-64세 인구 구성비 순으로 정렬하여 계통추출하였다. 2013년에는 매분기 42개 조사구로 4회 조사하여 168개 조사구를 조사하였으며, 2014년에는 매월 14개 조사구로 12회 조사하여 168개 조사구를 조사하였다. 5개 행정동별로 매분기 표본의 변동 폭 최소화 및 21명의 조사원이 동일 동내의 업무량 배정을 위해 8분기 동안 일정한 짝수의 표본조사구를 선정하였다. 표본가구 추출은 표본조사구별 1~60사이의 난수를 지정하여 해당난수에 해당하는 거처부터 연속하여 30가구를 표본가구로 선정하였다. 부재·불응 등으로 조사가 불가능할 경우에는 당초 선정된 마지막 표본가구 이후의 거처순서대로 대체가구로 선정하였다.

다. 시험조사 가구수 및 인구수

인천광역시 연수구(5개동)의 시험조사 결과 가구는 10,078가구(모집단 대비 20.6%), 인구는 23,884명(모집단 대비 19.8%)이 조사되었다. 시험조사 6개월 단위의 표본 가구수 및 인구수는 아래 표와 같다.

<표 4-2> 인천광역시 연수구 시험조사의 표본 가구수 및 인구수
<가구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	48,460	2,517	5.2%	2,517	5.2%
12개월(2차)	48,911	2,520	5.2%	5,037	10.3%
18개월(3차)	49,043	2,520	5.1%	7,557	15.4%
24개월(4차)	48,934	2,521	5.2%	10,078	20.6%

<인구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	122,180	5,796	4.7%	5,796	4.7%
12개월(2차)	121,966	5,935	4.9%	11,807	9.7%
18개월(3차)	121,535	5,991	4.9%	17,722	14.6%
24개월(4차)	120,502	6,162	5.1%	23,884	19.8%

제2절 2010년 센서스 표본조사와 시험조사의 결과 비교

이 절에서는 2010년 인구주택총조사 10% 표본조사와 인천 연수구 및 경남 거창군의 시험조사의 주요 결과를 비교함으로써 항목별 변화 추이 및 시험조사에서의 무응답 현황 등을 살펴보고자 한다.

1. 인구 부문

가. 성별

2010년 센서스 표본조사와 시험조사의 성별 분포는 <표 4-3>과 같다. 시험조사의 경우 추계인구(가중치 적용 후)에 비하여 여자의 비율이 약간 과다하고 남자의 비율이 약간 과소하게 조사된 것으로 나타났다.

<표 4-3> 표본조사 분포 - 성별

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
남자	50.4	49.8	50.9	46.8	45.8	47.8
여자	49.6	50.2	49.1	53.2	54.2	52.2

나. 나이

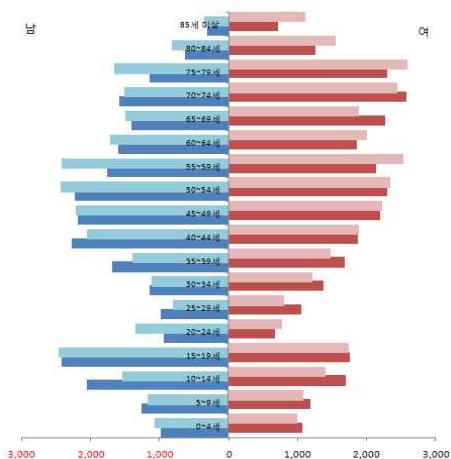
2010년 센서스 표본조사와 시험조사의 성별 분포는 <표 4-4>와 같다. 통계청 인구총조사과의 분석에 의하면 인천광역시 연수구와 경상남도 거창군의 경우 2010년 센서스와 시험조사 자료의 5세 구간별 인구구조에 있어서 시차에 따라 인구가 고령화 되는 모습을 보이고 있다(통계청 조사관리국 인구총조사과, 2015).

<표 4-4> 표본조사 분포 - 나이

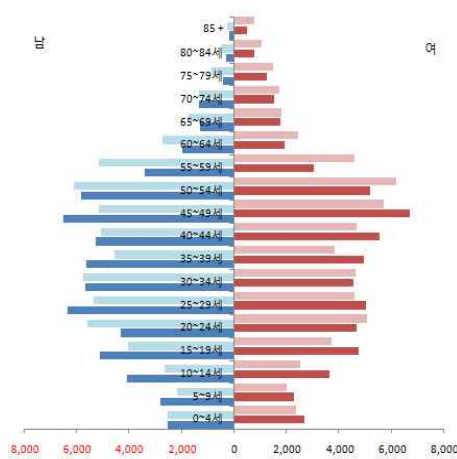
(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
0-4	4.0	4.6	4.1	2.5	3.2	3.6
5-9	4.2	3.6	3.5	2.7	3.9	3.9
10-14	6.2	4.4	4.3	4.0	5.4	5.1
15-19	7.5	6.5	6.4	13.6	4.4	7.3
20-24	6.9	7.4	8.8	3.3	1.6	3.7
25-29	9.4	8.4	8.2	2.8	2.5	2.8
30-34	7.8	9.2	8.6	2.7	3.6	4.1
35-39	8.6	7.4	7.0	3.9	4.5	5.0
40-44	8.4	8.5	8.1	4.9	6.2	6.8
45-49	10.6	8.9	9.0	5.3	6.8	7.7
50-54	9.0	10.1	10.2	6.3	8.2	8.3
55-59	5.3	7.6	8.1	7.2	9.1	8.6
60-64	3.3	4.0	4.3	6.4	7.8	6.5
65-69	2.8	3.1	2.9	7.8	7.0	5.9
70-74	2.3	2.6	2.5	10.2	9.3	6.9
75-79	1.4	1.8	2.0	8.3	8.4	7.4
80-84	1.3	1.2	1.3	5.1	5.2	4.1
85+	1.0	0.7	0.8	3.2	2.9	2.5

<경상남도 거창군>



<인천광역시 연수구>



주: 짙은 색이 '10년도 조사, 연한색이 순환 18개월 그룹임

출처: 인구주택총조사 개선을 위한 순환센서스 제1차 시험조사 결과보고(통계청 조사관리국 인구총조사과, 2015)

다. 교육정도

시지역인 연수구가 군지역인 거창군에 비해서는 ‘대학원 박사 이상’의 빈도수가 약간 많기는 하지만, 전체적으로 ‘대학원 이상’ 특히 ‘박사 이상’의 빈도수가 상대적으로 매우 적다.

<표 4-5> 표본조사 분포 - 교육정도

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
미취학	8.1	7.5	7.1	21.9	22.4	99.8
초등학교	11.7	9.9	9.9	28.3	24.7	3.3
중학교	9.9	9.5	9.5	10	13.7	2.9
고등학교	36.7	37.5	37.6	25	22.8	3.0
대학4년제미만	13.4	12.1	12.2	9.4	5.8	2.5
대학4년제이상	17.7	21.2	21.4	4.6	9.4	5.9
대학원 석사	2.0	2.0	2.0	0.6	0.9	6.3
대학원 박사	0.4	0.3	0.3	0.1	0.2	15.2
무응답		0.0	0.0		0.0	0.0
미상		0.0	0.0			

* 시험조사 무응답: 연수구 1개

* 시험조사 미상(입력 혹은 코드값 오류): 연수구 5개

라. 1년 전 거주지

1년 전 거주지는 ‘현 거주지’인 경우가 대부분을 차지하고 있으며, 인구인동 지표는 2010년 대비 감소한 것으로 나타났다.

<표 4-6> 표본조사 분포 - 1년 전 거주지

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
태어나지 않음	0.9	0.8	0.7	0.4	0.5	0.6
현거주지	77.7	80.8	81.0	86.8	95.5	94.7
이동:같은 시군	13.6	11.2	11.1	6.3	2.3	2.7
이동:다른 시군	7.3	6.5	6.6	6.4	1.6	1.9
북한또는 외국	0.4	0.6	0.6	0.1	0.1	0.1
무응답	0.0	0.0	0.0		0.0	0.0
미상	0.1					

* 시험조사 무응답: 연수구 9개, 거창군 2개

* 시험조사 미상(입력 혹은 코드값 오류): 연수구 1개

마. 통근·통학 여부

통근·통학 여부의 경우 ‘안함’이 전체의 50% 정도를 차지하고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-7> 표본조사 분포 - 통근·통학 여부

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
해당 없음	10.6	9.2	8.5	6.9	8.4	8.7
안함	46.3	47.9	48.0	50.6	53.9	50.8
통근	15.7	12.6	13.0	6.7	8.6	11.4
통학	27.4	30.2	30.5	35.8	29.1	29.1
미상	0.0					

* 해당 없음과 무응답(미상)이 분류되지 않음.

* 최재혁, 김미경(2014)에 의하면 1차년도 무응답은 연수구 614개, 거창군 209개로 높은 수준임.

바. 경제활동상태

경제활동상태의 경우 ‘주로 일하였음’과 ‘일하지 않음’이 대부분을 차지하고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-8> 표본조사 분포 - 경제활동상태

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
해당 없음	14.6	12.2	11.4	9.2	12.0	12.1
일함: 주로	43.5	48.7	48.7	44.3	53.3	50.9
일함: 틈틈히	2.6	1.6	1.7	6.3	1.2	1.0
일시 휴직	1.7	0.3	0.3	1.1	2.2	1.5
일하지 않음	37.6	37.3	37.9	39.2	31.3	34.5

* 해당 없음과 무응답(미상)이 분류되지 않음.

* 최재혁, 김미경(2014)에 의하면 1차년도 무응답은 연수구 140개, 거창군 468개로 높은 수준임.

사. 혼인상태

혼인상태의 경우 ‘미혼’과 ‘배우자 있음’이 대부분을 차지하고 있으며, 미혼율은 2010년 대비 약간 감소한 형태를 보이고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-9> 표본조사 분포 - 혼인상태

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
해당 없음	14.5	12.4	11.7	9.2	12.5	12.9
미혼	30.3	29.5	30.1	22.8	11.8	16.4
유배우	43.4	46.4	46.3	48.1	56.3	53.8
사별	6.1	5.4	5.6	17.9	16.8	14.1
이혼	5.3	6.3	6.4	1.9	2.5	2.7
미상	0				0.1	0.1

* 시험조사 미상(입력 또는 코드값 오류): 거창군 18개

* 해당 없음과 무응답(미상)이 분류되지 않음.

* 최재혁, 김미경(2014)에 의하면 1차년도 무응답은 연수구 1,474개, 거창군 2,528개로 매우 높은 수준임.

아. 총 출생아수

총 출생아수의 경우 4인 이하가 대부분을 차지하고 있으며, 평균 출생아수는 2010년 대비 약간 감소한 형태를 보이고 있다.

<표 4-10> 표본조사 분포 - 총 출생아수: 남자

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
해당 없음	69.7	69.5	69.9	59.6	56.0	59.8
0명	7.5	8.1	8.0	3.3	4.1	4.2
1명	14.0	14.5	14.3	12.2	15.8	15.6
2명	7.0	6.5	6.5	14.4	15.8	13.7
3명	1.3	1.0	1.0	6.7	5.5	4.4
4명	0.3	0.2	0.2	2.6	1.8	1.4
5명	0.1	0.0	0.1	0.8	0.6	0.4
6명		0.0	0.0	0.2	0.1	0.0
7명	0.0	0.0	0.0	0.1	0.0	0.0
8명		0.0	0.0	0.0		
9명		0.0	0.0			
19					0.3	0.4
20					0.0	0.0

* 거창군 시험조사의 코드값 19, 20은 조사 또는 입력 오류로 보임.

<표 4-11> 표본조사 분포 - 총 출생아수: 여자

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
해당 없음	69.7	69.8	70.2	59.6	56.6	60.3
0명	9.9	10.1	10.0	7.1	8.3	8.3
1명	12.6	13.1	12.8	12.4	15.5	14.7
2명	5.6	5.2	5.2	10.2	10.6	9.3
3명	1.4	1.3	1.3	5.8	5.3	4.3
4명	0.5	0.4	0.4	2.9	2.3	1.9
5명	0.1	0.1	0.1	1.4	0.9	0.8
6명	0.1	0.0	0.0	0.4	0.2	0.2
7명	0.0	0.0	0.0	0.1	0.1	0.0
8명		0.0	0.0	0.0		
10명				0.0	0.0	0.0
11명	0.0				0.0	0.0
12명					0.1	0.1

2. 가구 부문

가. 가구 구분

가구 구분의 경우 ‘가족’과 ‘1인 가구’가 대부분을 차지하고 있다. 연수구와 거창군 시험조사의 모두 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다. 1인 가구의 비율은 2010년 대비 약간 증가한 형태를 보이고 있다.

<표 4-12> 표본조사 분포 - 가구 구분

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
가족	63.9%	65.4	61.5	62.4	65.1	59.8
가족+가족이외	0.4%	0.6	0.6	0.2	0.2	0.3
1인 가구	32.9%	32.2	36.4	37.1	34.0	39.0
남남 5인이하	2.3%	1.8	1.5	0.3	0.3	0.3
남남 6인이상	0.0%	0.0	0.0	0.0	0.0	0.0
무응답		0.0	0.0		0.3	0.4
미상	0.2	0.0	0.0		0.0	0.0

* 시험조사 무응답: 연수구 3개, 거창군 20개

* 시험조사 미상(입력 혹은 코드값 오류): 연수구 1개, 거창군 2개

나. 거주 기간

시지역인 연수구의 경우 단기 거주 비율이 높고 장기 거주 비율은 낮은 형태를 보이고 있다. 반면, 군지역인 거창군의 경우는 25년 이상 장기 거주 비율이 매우 높은 것으로 나타났으며, 25년 이상 장기 거주 비율은 2010년 대비 상당히 감소한 형태를 보이고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-13> 표본조사 분포 - 거주 기간

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
1년 미만	28.0	22.6	23.1	6	3.5	4.6
1-2년 미만	12.4	13.8	13.9	3.3	5.4	6.6
2-3년 미만	9.1	9.8	10.0	4.3	4.0	4.8
3-5년 미만	10.3	9.9	9.9	3.8	6.0	7.1
5-10년 미만	17.9	18.3	18.3	12	14.2	17.1
10-15년 미만	10.5	9.9	9.7	6.2	11.0	12.5
15-20년 미만	10.0	8.3	8.0	4.5	6.0	6.0
20-25년 미만	1.0	6.5	6.2	6.7	7.8	7.5
25년 이상	0.7	0.9	0.8	53.2	41.5	33.1
무응답		0.0	0.0		0.5	0.6
미상		0.0	0.0		0.1	0.1

* 시험조사 무응답: 연수구 4개, 거창군 27개

* 시험조사 미상(입력 혹은 코드값 오류): 연수구 1개, 거창군 3개

다. 사용 방 수: 잠을 자는 방

잠을 자는 방의 수는 1-3개의 경우가 대부분을 차지하고 있으며, 2010년에 비하여 잠을 자는 방의 수는 상당히 감소한 형태를 보이기도 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-14> 표본조사 분포 - 사용 방 수: 잠을 자는 방

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
0개					0	0.0
1개	0.1	50.3	50.9		66.9	60.9
2개	47.3	31	29.0	54.2	23.2	25.0
3개	30.2	17.3	18.4	34.5	8.5	12.2
4개	21.1	1.3	1.6	10.8	0.7	1.0
5개	1.3	0.0	0.0	0.5	0.1	0.1
6개				0.0	0.0	0.0
무응답		0.0	0.0		0.6	0.8

* 시험조사 무응답: 연수구 3개, 거창군 37개

라. 난방 시설

시지역인 연수구의 경우 도시가스의 비율이 대부분을 차지하고 있는 반면, 군지역인 거창군의 경우는 기름보일러의 비율이 매우 높게 나타났다. 거창군의 경우 2010년과 비교하여 난방 시설에 대한 변화가 큰 편이며, 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-15> 표본조사 분포 - 난방 시설

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
무응답	0.5	0.0	0.0		0.7	0.5
중앙난방	9.1	10.2	9.6		3.0	4.0
지역난방	13.9	15.8	15.8		0.6	0.9
도시가스	75.1	73.2	73.8		12.6	17.9
기름보일러	0.1	0.1	0.1	68.5	52.8	46.3
프로판가스	0.0			10.9	11.2	14.9
전기보일러	1.3	0.4	0.4	7.3	6.2	5.6
연탄보일러		0.1	0.1	0.7	0.8	0.7
연탄아궁이				0.1	0.2	0.2
재래식아궁이				4.2	2.4	1.7
기타		0.1	0.1	8.3	9.6	7.2
미상		0.0	0.0			

* 시험조사 무응답: 연수구 3개, 거창군 40개

* 시험조사 미상(입력 혹은 코드값 오류): 연수구 1개

바. 점유 형태

시지역인 연수구의 경우 자가의 비율이 가장 높으나 전세와 월세의 비율도 상당한 반면, 군지역인 거창군의 경우는 자가가 대부분을 차지하고 있다. 거창군 시험조사의 경우 가중치 적용 전과 후의 비율에 약간의 차이가 나타나고 있다.

<표 4-16> 표본조사 분포 - 점유 형태

(단위: %)

구분	인천광역시 연수구			경상남도 거창군		
	2010 센서스 10% 표본조사	시험조사 20% 표본조사		2010 센서스 10% 표본조사	시험조사 20% 표본조사	
	가중치 적용 전	가중치 적용 전	가중치 적용 후	가중치 적용 전	가중치 적용 전	가중치 적용 후
자가	40.5	41.3	40.4	87.1	84.6	80.3
전세(월세없음)	19.3	18.1	18.3	2.8	2.7	3.8
보증있는 월세	19.2	26.6	26.8	4.4	7.1	10.1
보증없는 월세	15.2	10.9	11.4	1.6	1.4	1.6
사글세	0.9	0.0	0.0	0.9	0.9	0.9
무상(관사,사택)	2.7	3.0	3.0	3.2	2.5	2.5
무응답	2.3	0.1	0.1		0.8	0.8
미상	0.0					

* 시험조사 무응답: 연수구 13개, 거창군 44개

제3절 시·군 단위의 시험조사 결과 분석

이 절에서는 인천광역시 연수구의 시험조사를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 시군구 단위의 시기별 공표 가능한 수준을 살펴보고자 한다. 참고로 경상남도 거창군의 경우는 20% 표본 자료에 대한 결과를 제시하였다. 이 절에서 분석 대상으로 삼고 있는 항목은 다음과 같다.

<인구 부문>

성별, 나이(5세 단위), 교육정도, 1년 전 거주지, 통근·통학 여부, 경제활동상태, 직업: 종사상 지위, 혼인상태, 총 출생아수

<가구 부문>

가구 구분, 거주 기간, 사용 방 수: 잠을 자는 방, 난방 시설, 점유 형태

여기에서는 조사 자료를 6개월 단위로 구분하고 가중치를 부여하여 단순 및 누적 추정치와 CV(상대표준오차)를 계산하였으며 CV가 20% 이하이면 안정적인 값으로 판단하였다.

1. 인구 부문

가. 성별

성별의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다.

<표 4-17> 시험조사 결과 - 성별

<비율 추정치>								(단위: %)
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
남자	50.9	50.9	50.9	50.9	50.9	50.9	50.9	47.8
여자	49.1	49.1	49.1	49.1	49.1	49.1	49.1	52.2

<상대표준오차>								(단위: %)
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
남자	1.2	1.4	1.2	1.2	0.9	0.7	0.6	0.8
여자	1.3	1.5	1.3	1.2	1.0	0.8	0.6	0.8

나. 나이

나이의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다.

<표 4-18> 시험조사 결과 - 나이

<비율 추정치> (단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
0-4	4.1	4.1	4.1	4.1	4.1	4.1	4.1	3.6
5-9	3.6	3.5	3.5	3.5	3.5	3.5	3.5	3.9
10-14	4.9	4.7	4.5	4.3	4.7	4.5	4.3	5.1
15-19	7.1	6.9	6.6	6.4	6.9	6.6	6.4	7.3
20-24	9.0	9.0	9.0	8.8	9.0	9.0	8.8	3.7
25-29	8.0	8.0	8.0	8.2	8.0	8.0	8.2	2.8
30-34	8.7	8.8	8.8	8.6	8.8	8.8	8.6	4.1
35-39	7.3	7.0	6.9	7.0	7.0	6.9	7.0	5.0
40-44	8.5	8.4	8.3	8.1	8.4	8.3	8.1	6.8
45-49	9.2	9.2	9.1	9.0	9.2	9.1	9.0	7.7
50-54	10.2	10.1	10.2	10.2	10.1	10.2	10.2	8.3
55-59	7.0	7.4	7.7	8.1	7.4	7.7	8.1	8.6
60-64	3.7	3.9	4.1	4.3	3.9	4.1	4.3	6.5
65-69	2.7	2.7	2.8	2.9	2.7	2.8	2.9	5.9
70-74	2.5	2.5	2.5	2.5	2.5	2.5	2.5	6.9
75-79	1.7	1.9	1.9	2.0	1.9	1.9	2.0	7.4
80-84	1.1	1.1	1.2	1.3	1.1	1.2	1.3	4.1
85+	0.7	0.8	0.8	0.8	0.8	0.8	0.8	2.5

<상대표준오차> (단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
0-4	9.3	8.8	8.2	8.3	6.3	5.0	4.2	9.6
5-9	8.5	9.7	7.7	7.5	6.3	4.9	4.0	7.1
10-14	7.6	8.4	7.3	7.1	5.6	4.5	3.8	5.1
15-19	6.6	6.0	6.0	5.4	4.4	3.6	3.0	6.3
20-24	5.3	5.2	5.9	5.8	3.7	3.2	2.8	8.4
25-29	5.8	6.2	5.7	6.9	4.2	3.4	3.0	6.7
30-34	5.9	6.3	6.7	6.5	4.3	3.7	3.2	9.6
35-39	6.5	7.0	5.6	5.7	4.8	3.7	3.1	5.5
40-44	5.3	4.8	4.7	4.7	3.5	2.9	2.4	4.0
45-49	4.6	5.3	5.2	4.5	3.5	2.9	2.4	4.1
50-54	4.9	4.6	4.7	5.1	3.4	2.8	2.5	4.2
55-59	5.2	6.3	6.0	6.0	4.2	3.4	3.0	4.4
60-64	9.6	7.4	7.2	8.4	5.7	4.4	3.8	5.1
65-69	8.0	10.1	7.8	8.0	6.3	4.9	4.3	5.4
70-74	10.7	13.1	9.1	10.2	8.2	6.2	5.2	4.7
75-79	11.8	13.9	13.3	10.9	9.3	7.5	6.1	5.2
80-84	14.4	14.8	14.5	14.6	10.4	8.2	7.0	5.8
85+	19.7	19.3	23.7	22.2	13.5	11.5	10.0	7.2

다. 교육정도

연수구의 경우 전체적으로 모든 교육정도에서 CV가 안정적인 것으로 나타났다(대학원 이상의 경우에서 다소 높은 CV 값을 가지나 전체 차수에서 20% 이하의 값으로 안정적인 수준임). 교육정도별로는 대학원 이상으로 통합하여 모든 수준에서 5%, 10%, 15%, 20% 표본 누적결과가 공표 가능한 것으로 판단된다.

거창군의 경우 4년제 대학 이하의 교육정도에서는 모든 CV 값이 안정적이거나 대학원 이상의 교육정도는 통합을 하는 경우에도 CV 값이 안정적이지 않은 것으로 나타났다(최재혁, 김경미, 2014).

<표 4-19> 시험조사 결과 - 교육정도

<비율 추정치> (단위: %)								
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
미취학	7.1	7.0	6.7	7.1	7.1	7.0	7.1	18.7
초등학교	9.5	10.0	10.4	9.3	9.9	10.0	9.9	21.5
중학교	9.1	10.1	9.6	9.5	9.6	9.6	9.5	13.3
고등학교	38.2	37.2	38.5	36.9	37.7	37.9	37.6	26.1
대학4년제미만	12.5	12.4	12.2	12.1	12.4	12.3	12.2	7.8
대학4년제이상	21.5	21.0	20.6	22.4	21.2	21.1	21.4	11.2
대학원 석사	1.8	1.8	1.8	2.4	1.8	1.8	2.0	1.1
대학원 박사	0.2	0.4	0.3	0.1	0.3	0.3	0.3	0.2
무응답	0.0				0.0	0.0	0.0	0.0
미상	0.1	0.0			0.0	0.0	0.0	

<상대 표준오차> (단위: %)								
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
미취학	6.1	6.6	6.3	5.5	4.5	3.6	3.0	3.3
초등학교	4.9	5.4	6.6	6.2	3.7	3.4	3.0	2.9
중학교	5.5	5.6	5.2	4.7	3.9	3.1	2.6	3.0
고등학교	2.5	2.6	2.5	2.6	1.8	1.4	1.2	2.5
대학4년제미만	4.7	4.8	4.5	4.5	3.3	2.7	2.3	5.9
대학4년제이상	5.2	4.5	5.2	4.3	3.5	2.9	2.4	6.3
대학원 석사	10.9	12.4	11.3	12.6	8.3	6.8	6.1	15.2
대학원 박사	32.5	24.9	26.2	34.8	19.8	15.8	14.5	25.6
무응답	100				100	99.9	99.9	99.8
미상	99.5	100			86.9	85.7	85.0	

라. 1년 전 거주지

국내이동(1년전 거주지)의 경우 최근의 차수일수록 같은 시군구로의 이동은 적어지고, 다른 시군구로의 이동은 커지는 특성을 나타내고 있다. 국내이동(1년 전 거주지)별 인구수는 전체 항목에서 대체로 안정적인 CV 값을 가지는 것으로 나타났다. 북한 또는 외국 범주에서 다소 불안정하나, '다른 지역'으로 병합한다면 안정적

인 값을 가질 것으로 판단된다.

<표 4-20> 시험조사 결과 - 1년 전 거주지

<비율 추정치> (단위: %)								
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
태어나지 않음	0.7	0.8	0.7	0.7	0.8	0.7	0.7	0.6
현거주지	78.8	80.5	81.3	82.6	79.8	80.3	81.0	94.7
이동:같은 시군	14.5	13.9	8.7	7.9	14.1	12.3	11.1	2.7
이동:다른 시군	5.2	4.3	8.7	8.1	4.7	6.0	6.6	1.9
북한 또는외국	0.6	0.6	0.6	0.7	0.6	0.6	0.6	0.1
무응답	0.2				0.1	0.1	0.0	0.0

<상대표준오차> (단위: %)								
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
태어나지 않음	18.2	16.5	16.1	18.5	12.1	9.6	8.5	16.9
현거주지	1.9	1.7	1.6	1.7	1.2	1.0	0.9	0.6
이동:같은 시군	7.4	6.5	8.1	9.4	5.0	4.5	4.2	13.4
이동:다른 시군	13.3	15.1	9.9	9.6	9.8	7.2	5.8	14.2
북한 또는외국	24.2	35.3	27.9	23.3	21.1	16.7	13.5	48.0
무응답	42.0				42.7	40.6	40.9	71.1

마. 통근·통학 여부

통근·통학 여부의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다.

<표 4-21> 시험조사 결과 - 통근·통학 여부

<비율 추정치> (단위: %)								
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	8.1	9.0	8.7	8.9	8.5	8.5	8.5	8.7
안함	46.9	48.2	47.1	49.0	47.5	47.5	48.0	50.8
통근	14.4	13.9	13.0	13.0	13.9	13.3	13.0	11.4
통학	30.7	28.9	31.2	29.1	30.0	30.7	30.5	29.1

<상대표준오차> (단위: %)								
구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	7.6	6.9	6.0	5.7	5.1	3.9	3.2	6.9
안함	2.3	2.2	2.3	1.8	1.6	1.3	1.0	1.4
통근	5.4	4.6	5.2	4.4	3.6	2.9	2.4	4.9
통학	3.6	3.9	3.7	3.5	2.6	2.1	1.8	2.7

바. 경제활동상태

경제활동상태의 경우 대체로는 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다. 거창군의 경우 '일하였음'의 하위항목인 '틈틈이'와 '일시 휴직'은 응답자가 적어 추정치의 변동이 크며, CV 값도 안정적이지 않다. 따라서 공표 시 '일하였음'의 하위항목 중 '틈틈이'와 '일시 휴직'의 병합이 필요할 것으로 판단된다.

<표 4-22> 시험조사 결과 - 경제활동상태

<비율 추정치> (단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	11.8	12.1	11.3	11.5	11.9	11.5	11.4	12.1
일함: 주로	48.5	49.2	47.8	48.8	48.8	48.6	48.7	50.9
일함: 틈틈히	1.2	0.7	2.1	2.9	0.9	1.3	1.7	1.0
일시 휴직	0.3	0.4	0.2	0.1	0.3	0.3	0.3	1.5
일하지 않음	38.3	37.5	38.6	36.7	38.0	38.2	37.9	34.5

<상대 표준오차> (단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	6.2%	6.2%	5.1%	5.1%	4.3%	3.3%	2.8	5.4
일함: 주로	2.4%	2.1%	2.3%	1.9%	1.6%	1.3%	1.1	1.6
일함: 틈틈히	14.0%	27.2%	13.9%	10.0%	13.8%	10.1%	7.4	21.0
일시 휴직	26.5%	25.7%	26.4%	31.9%	18.3%	15.2%	13.8	24.2
일하지 않음	3.4%	3.2%	2.9%	2.8%	2.3%	1.8%	1.5	2.1

사. 직업: 종사상 지위

종사상 지위의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다.

<표 4-23> 시험조사 결과 - 종사상 지위

<비율 추정치> (단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	50.4	49.6	49.9	48.2	50.1	49.9	49.4	46.7
임금 근로자	39.6	40.4	40.5	41.8	39.9	40.2	40.6	19.8
고용원 없는 자영업	6.0	5.8	6.2	6.2	5.9	6.1	6.2	20.4
고용원 있는 자영업	2.6	3.0	2.4	2.7	2.8	2.7	2.7	1.8
무급 가족	1.4	1.1	1.0	1.0	1.3	1.2	1.2	11.3

<표 4-23> 시험조사 결과 - 종사상 지위 (계속)

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	2.4	2.2	2.3	2.0	1.6	1.3	1.1	1.6
임금 근로자	3.2	2.8	2.6	2.2	2.1	1.6	1.3	2.8
고용원 없는 자영업	6.1	5.5	6.5	5.2	4.1	3.5	2.9	2.7
고용원 있는 자영업	15.8	9.2	11.8	9.9	8.8	7.1	5.9	9.1
무급 가족	13.5	12.8	13.8	12.6	9.4	7.8	6.7	3.6

아. 혼인상태

혼인상태의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV 값을 가지는 것으로 나타났다.

<표 4-24> 시험조사 결과 - 혼인상태

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	13.0	12.1	11.3	11.5	12.5	11.9	11.7	12.9
미혼	30.7	30.3	31.5	29.8	30.3	30.4	30.1	16.4
유배우	44.7	45.8	45.1	47.5	45.5	45.7	46.3	53.8
사별	5.4	5.5	5.5	5.0	5.5	5.6	5.6	14.1
이혼	6.2	6.3	6.6	6.2	6.2	6.4	6.4	2.7
미상								0.1

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	6.7%	6.2%	5.0%	5.1%	4.5%	3.4%	2.9	5.8
미혼	3.6%	3.5%	2.8%	3.2%	2.5%	1.9%	1.6	3.8
유배우	2.6%	2.2%	2.6%	2.3%	1.7%	1.5%	1.2	1.2
사별	9.9%	8.8%	10.2%	8.9%	6.6%	5.5%	4.7	4.1
이혼	10.1%	9.3%	9.5%	9.4%	6.7%	5.6%	4.8	8.1
미상								100.0

자. 총 출생아수

총 출생아수의 경우 남자, 여자 모두 자녀수가 5명 이상의 경우 다소 불안정하며 5명 이하에서는 추정치 및 CV 값이 안정적인 것으로 나타났다. 자녀수 5명 이상을 병합한다면 안정적인 CV 값을 가질 것으로 판단된다.

<표 4-25> 시험조사 결과 - 총 출생아수: 남자

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	69.7	73.7	69.3	68.6	71.6	70.6	69.9	59.8
0	8.7	4.7	8.8	9.4	6.7	7.4	8.0	4.2
1	14.6	14.1	13.5	14.1	14.4	14.2	14.3	15.6
2	5.7	6.4	7.1	6.3	6.1	6.5	6.5	13.7
3	0.9	0.7	1.0	1.2	0.8	0.9	1.0	4.4
4	0.2	0.3	0.2	0.2	0.2	0.2	0.2	1.4
5	0.1	0.1	0.0	0.1	0.0	0.0	0.1	0.4
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7				0.0			0.0	0.0
8	0.0	0.0			0.0	0.0	0.0	
9	0.0	0.0			0.0	0.0	0.0	
19								0.4
20								0.0

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	0.8	1.0	0.8	0.8	0.7	0.5	0.5	1.1
0	4.8	13.1	4.9	4.3	6.0	4.2	3.2	5.9
1	3.3	3.0	3.5	2.9	2.2	1.8	1.5	2.2
2	5.1	6.9	5.1	5.2	4.3	3.3	2.7	2.5
3	14.0	16.2	15.8	14.4	10.4	8.6	7.3	5.6
4	28.9	27.9	33.9	31.4	19.4	17.1	15.0	8.1
5	59.2	73.0	57.7	69.6	44.5	35.6	31.8	13.8
6	99.9	73.4	100.1	100.3	58.9	51.5	45.0	36.8
7				71.4			71.7	45.6
8	100.4	99.7%			74.0%	76.7%	73.9	
9	100.4	100.0%			70.8%	71.8%	70.7	
19								100.1
20								100.1

<표 4-26> 시험조사 결과 - 총 출생아수: 여자

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	69.8	74.8	69.3	68.6	72.2	71.0	70.2	60.3
0	10.8	5.5	11.3	11.7	8.1	9.2	10.0	8.3
1	12.8	12.8	12.8	12.5	12.8	12.9	12.8	14.7
2	5.0	5.2	4.6	5.6	5.2	5.0	5.2	9.3
3	1.1	1.3	1.5	1.1	1.2	1.3	1.3	4.3
4	0.3	0.3	0.4	0.3	0.3	0.4	0.4	1.9
5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.8
6	0.0			0.0	0.0	0.0	0.0	0.2
7	0.0		0.0		0.0	0.0	0.0	0.0
8				0.0			0.0	
10								0.0
11								0.0
12								0.1

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 21.4%
	1차 4.7%	2차 4.9%	3차 4.9%	4차 5.1%	2차 9.7%	3차 14.6%	4차 19.8%	
해당 없음	0.8	1.1	0.8	0.8	0.8	0.6	0.5	1.1
0	4.1%	12.9%	4.2%	3.8%	5.7%	3.9%	3.0	4.5
1	3.4%	3.3%	3.1%	3.3%	2.3%	1.8%	1.6	2.0
2	6.7%	5.9%	5.6%	6.4%	4.3%	3.6%	3.2	3.0
3	12.4%	12.4%	11.8%	13.3%	8.8%	7.2%	6.3	5.1
4	23.5%	25.3%	22.4%	27.2%	16.9%	13.3%	11.6	7.5
5	50.0%	41.6%	38.3%	35.1%	31.6%	23.9%	19.4	12.1
6	100.8%			100.0%	100.3%	100.1%	71.4	20.4
7	99.8%		100.4%		99.9%	72.0%	72.2	42.9
8				100.7%			100.0	
10								100.1
11								100.1
12								100.1

2. 가구 부문

가. 가구 구분

가구 구분에 따른 가구 수는 ‘가족으로 구성된 가구’와 ‘1인 가구’ 및 ‘남남 5인 이하’의 경우는 추정치가 안정적이거나 나머지 가구 특성에 대해서는 추정치의 변동이 크다. 따라서 공표 시 ‘가족’과 ‘가족+가족 이외’의 병합, ‘남남 5인 이하’와 ‘남남 6인 이상’의 병합이 필요할 것으로 판단된다.

<표 4-27> 시험조사 결과 - 가구 구분

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
가족	61.6	62.1	61.8	61.6	61.8	61.7	61.5	59.8
가족+가족이외	0.8	0.9	0.7	0.1	0.8	0.8	0.6	0.3
1인 가구	35.9	35.9	36.0	36.6	36.1	36.2	36.4	39.0
남남 5인이하	1.5	1.1	1.4	1.7	1.3	1.4	1.5	0.3
남남 6인이상	0.0				0.0	0.0	0.0	0.0
무응답	0.1				0.1	0.0	0.0	0.4
미상		0.0			0.0	0.0	0.0	0.0

<상대표준오차>

(단위: %)

구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
가족	5.4%	5.1%	4.9%	5.2%	3.7%	2.9%	2.5	2.5
가족+가족이외	24.1%	26.5%	28.8%	57.1%	17.7%	15.0%	14.5	28.7
1인 가구	9.0%	8.6%	8.1%	8.4%	6.1%	4.8%	4.1	3.7
남남 5인이하	19.6%	22.5%	25.1%	22.3%	14.5%	12.5%	11.2	28.0
남남 6인이상	99.8%				99.7%	99.7%	99.6	99.8
무응답	58.0%				59.1%	58.9%	59.2	78.2
미상		100.0			100.0	100.0	100.0	100.0

나. 거주 기간

거주 기간의 경우 '25년 이상'을 제외하고는 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV 값을 가지는 것으로 나타났다. 거창군의 경우 '25년 이상'의 범주도 CV 값이 안정적이며, 연수구의 경우도 크게 나쁘지는 않은 것으로 판단된다.

<표 4-28> 시험조사 결과 - 거주 기간

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
1년 미만	24.7	22.7	21.8	22.8	23.8	23.2	23.1	4.6
1-2년 미만	14.2	14.6	13.3	13.4	14.5	14.1	13.9	6.6
2-3년 미만	9.0	10.3	10.3	10.6	9.6	9.8	10.0	4.8
3-5년 미만	7.9	10.6	10.3	11.0	9.1	9.5	9.9	7.1
5-10년 미만	18.5	18.9	17.9	18.0	18.7	18.3	18.3	17.1
10-15년 미만	10.2	9.0	10.2	9.6	9.6	9.7	9.7	12.5
15-20년 미만	8.7	8.2	7.7	7.6	8.4	8.2	8.0	6.0
20-25년 미만	5.8	4.7	8.1	6.2	5.3	6.2	6.2	7.5
25년 이상	0.7	1.0	0.6	0.8	0.9	0.8	0.8	33.1
무응답	0.2				0.1	0.1	0.0	0.6
미상	0.0				0.0	0.0	0.0	0.1

<표 4-28> 시험조사 결과 - 거주 기간 (계속)

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
1년 미만	7.6%	7.6%	7.4%	8.6%	5.4%	4.4%	3.9	11.3
1-2년 미만	6.4%	6.5%	6.6%	7.2%	4.5%	3.7%	3.3	10.7
2-3년 미만	7.6%	8.4%	7.4%	7.9%	5.7%	4.5%	3.9	8.8
3-5년 미만	8.7%	6.9%	7.6%	8.3%	5.7%	4.6%	4.0	8.8
5-10년 미만	6.0%	5.3%	5.4%	5.8%	4.0%	3.2%	2.8	7.8
10-15년 미만	9.0%	9.0%	7.7%	10.2%	6.5%	5.0%	4.5	8.0
15-20년 미만	11.5%	11.4%	10.9%	11.8%	8.0%	6.4%	5.7	12.3
20-25년 미만	14.0%	15.5%	17.0%	14.8%	10.3%	9.4%	7.9	9.7
25년 이상	30.2%	36.9%	32.0%	65.3%	24.9%	20.1%	23.3	4.8
무응답	48.4%				50.0%	50.0%	50.2	79.1
미상	100.0				100.0	100.0	100.0	100.0

다. 사용 방 수: 잠을 자는 방

잠을 자는 방의 경우 3개 이하에서 안정적인 추정치 및 CV 값을 가지고 있으며, 4개의 이상의 경우를 병합하면 안정적인 CV 값을 가질 것으로 판단된다.

<표 4-29> 시험조사 결과 - 사용 방 수: 잠을 자는 방

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
0개								0.0
1개	45.1	49.3	53.0	54.4	47.4	49.4	50.9	60.9
2개	29.8	29.8	29.9	26.9	29.8	29.8	29.0	25.0
3개	23.4	18.6	15.8	17.4	20.9	19.0	18.4	12.2
4개	1.4	2.3	1.4	1.3	1.9	1.7	1.6	1.0
5개	0.1				0.1	0.0	0.0	0.1
6개								0.0
무응답	0.1				0.1	0.0	0.0	0.8

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
0개								99.8
1개	8.0%	6.6%	5.9%	5.5%	5.1%	3.9%	3.1	2.6
2개	9.0%	7.9%	6.9%	5.9%	5.9%	4.6%	3.7	4.5
3개	12.1%	13.8%	12.9%	11.6%	9.1%	7.5%	6.4	8.5
4개	49.2%	36.0%	35.7%	38.3%	28.8%	23.2%	20.0	33.4
5개	73.4%				72.9%	73.0%	72.9	62.2
6개								99.6
무응답	56.5%				58.1%	58.1%	58.4	77.8

라. 난방 시설

시지역인 연수구의 경우 중앙난방, 지역난방, 도시가스 범주에서만 안정적인 CV 값을 가지며, 군지역인 거창군의 경우는 도시가스, 기름보일러, 프로판가스, 전기보일러, 연탄보일러 범주에서 안정적인 CV 값을 가지는 것으로 나타났다.

<표 4-30> 시험조사 결과 - 난방 시설

<비율 추정치> (단위: %)								
구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
무응답	0.1				0.1	0.0	0.0	0.5
중앙난방	11.0	8.4	11.6	7.0	9.9	10.5	9.6	4.0
지역난방	11.6	17.6	16.3	17.7	14.6	15.1	15.8	0.9
도시가스	77.1	73.5	72.1	72.8	75.2	74.2	73.8	17.9
기름보일러		0.1		0.4	0.1	0.0	0.1	46.3
프로판가스								14.9
전기보일러		0.1		1.6	0.0	0.0	0.4	5.6
연탄보일러		0.2		0.2	0.1	0.1	0.1	0.7
연탄아궁이								0.2
재래식아궁이								1.7
기타	0.1	0.1		0.3	0.1	0.0	0.1	7.2
미상	0.0				0.0	0.0	0.0	

<상대 표준오차> (단위: %)								
구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
무응답	58.2%				58.3%	58.6%	58.6	38.9
중앙난방	24.7%	31.6%	27.0%	35.6%	19.3%	15.6%	14.2	39.9
지역난방	28.6%	17.8%	22.2%	20.0%	15.5%	12.5%	10.7	94.6
도시가스	4.7%	4.8%	5.4%	4.4%	3.4%	2.8%	2.4	15.5
기름보일러		100.2%		100.2%	99.9%	99.8%	81.5	5.1
프로판가스								16.9
전기보일러		100.2%		90.4%	99.9%	99.8%	85.8	12.6
연탄보일러		100.2%		100.2%	99.9%	99.8%	70.6	17.8
연탄아궁이								38.9
재래식아궁이								14.5
기타	99.4%	100.2%		100.2%	70.5%	70.4%	73.9	9.2
미상	100.3%				100.0%	99.8%	99.7	

바. 점유 형태

주거 형태별 가구 수는 사글세의 경우를 제외하고 안정적인 CV 값을 가지고 있다. 보증금 없는 월세와 사글세 항목을 병합할 경우 공표가능한 수준이 될 것으로 판단된다.

<표 4-31> 시험조사 결과 - 점유 형태

<비율 추정치> (단위: %)								
구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
무응답	0.3	0.3			0.3	0.2	0.1	0.8
자가	38.4	40.3	41.5	42.5	39.3	39.9	40.4	80.3
전세(월세없음)	17.8	20.5	19.0	16.7	18.8	18.8	18.3	3.8
보증있는월세	30.4	27.2	24.5	23.3	29.1	27.8	26.8	10.1
보증없는월세	10.8	9.6	11.9	13.1	10.3	10.9	11.4	1.6
사글세	0.0	0.0			0.0	0.0	0.0	0.9
무상(관사,사택)	2.4	2.0	3.1	4.4	2.2	2.5	3.0	2.5

<상대표준오차> (단위: %)								
구분	단순				누적			거창군 25.4%
	1차 5.2%	2차 5.2%	3차 5.1%	4차 5.2%	2차 10.3%	3차 15.4%	4차 20.6%	
무응답	39.1%	45.0%			30.5%	30.6%	31.1	74.8
자가	9.8%	8.8%	8.7%	8.8%	6.5%	5.2%	4.5	2.3
전세(월세없음)	8.5%	11.1%	8.2%	8.8%	6.8%	5.3%	4.6	10.9
보증있는월세	11.2%	11.3%	12.7%	12.5%	7.9%	6.7%	5.9	16.9
보증없는월세	14.3%	17.4%	16.5%	15.6%	11.0%	9.2%	7.9	16.4
사글세	71.7%	100.2%			58.2%	57.8%	57.8	18.5
무상(관사,사택)	23.0%	20.5%	17.3%	12.9%	15.8%	11.8%	8.9	10.5

제4절 동·읍 단위의 시험조사 결과 분석

이 절에서는 인천광역시의 시험조사 대상 동 중 가장 작은 규모인 연수3동의 시험조사를 대상으로 표본 비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 읍면동 단위의 시기별 공표 가능한 수준을 살펴보고자 한다. 연수구 연수3동의 시험조사 결과의 가구는 1,800가구(모집단 대비 20.5%), 인구는 3,826명(모집단 대비 19.8%)이 조사되었다. 시험조사 6개월 단위의 표본 가구 수 및 인구수는 아래 표와 같다.

<표 4-32> 연수구 연수3동 시험조사의 표본 가구수 및 인구수

<가구수>					
	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	6,958	360	5.2%	360	5.2%
12개월(2차)	7,030	360	5.1%	720	10.2%
18개월(3차)	7,028	360	5.1%	1,080	15.4%
24개월(4차)	7,027	360	5.1%	1,800	20.5%

<표 4-32> 연수구 연수3동 시험조사의 표본 가구수 및 인구수 (계속)

<인구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	19,925	937	4.7%	937	4.7%
12개월(2차)	19,857	930	4.7%	1,867	9.4%
18개월(3차)	19,768	964	4.9%	2,831	14.3%
24개월(4차)	19,543	995	5.1%	3,826	19.6%

참고로 경상남도 거창군의 경우에는 거창읍에 대한 결과를 제시하였는데, 거창군의 경우 3-4개의 인접 면들을 묶어 그룹을 설정하고 조사하였기 때문에 한 개 면에 대한 분석이 가능하지 않기 때문이다.

거창군 거창읍의 시험조사 결과의 가구는 2,552가구(모집단 대비 18.7%), 인구는 6,311명(모집단 대비 4.1%)이 조사되었다. 시험조사 6개월 단위의 표본 가구수 및 인구수는 아래 표와 같다.

<표 4-33> 거창군 거창읍 시험조사의 표본 가구수 및 인구수

<가구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	13,288	630	4.7%	630	4.7%
12개월(2차)	13,490	627	4.6%	1,257	9.3%
18개월(3차)	13,593	662	4.9%	1,919	14.1%
24개월(4차)	13,668	633	4.6%	2,552	18.7%

<인구수>

	모집단 크기	단순		누적	
		표본크기	%	표본크기	%
6개월(1차)	37,443	1,587	4.2%	1,587	4.2%
12개월(2차)	37,701	1,596	4.2%	3,183	8.4%
18개월(3차)	37,855	1,576	4.2%	4,759	12.6%
24개월(4차)	37,980	1,552	4.1%	6,311	16.6%

1. 인구 부문

가. 성별

성별의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다.

<표 4-34> 시험조사 결과 - 성별

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
남자	48.0	48.0	48.0	48.3	48.0	48.0	48.3	49.1
여자	52.0	52.0	52.0	51.7	52.0	52.0	51.7	50.8

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
남자	2.5%	3.7%	2.8%	2.8%	2.2%	1.7%	1.5	1.1
여자	2.3%	3.4%	2.6%	2.6%	2.0%	1.5%	1.4	1.1

나. 나이

나이의 경우 15세-59세 구간에서는 20% 표본은 물론 5%, 10%, 15% 표본 누적 결과 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다. 또한 '85세 이상' 구간을 제외하고는 모든 5세 단위의 구간에서 대략 10% 표본부터는 안정적인 CV 값을 가지는 것으로 나타났다.

<표 4-35> 시험조사 결과 - 나이

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
0-4	5.2	5.3	5.2	5.3	5.3	5.2	5.3	4.8
5-9	4.1	4.0	4.1	4.1	4.0	4.1	4.1	5.0
10-14	5.2	5.0	5.0	4.7	5.0	5.0	4.7	6.5
15-19	6.5	6.3	6.0	6.0	6.3	6.0	6.0	9.5
20-24	8.4	8.2	8.0	7.6	8.2	8.0	7.6	4.6
25-29	7.1	7.2	7.1	7.2	7.2	7.1	7.2	3.4
30-34	8.6	8.7	9.0	8.8	8.7	9.0	8.8	5.4
35-39	7.2	7.1	6.9	6.9	7.1	6.9	6.9	6.5
40-44	7.8	7.7	7.7	7.8	7.7	7.7	7.8	8.6
45-49	8.8	8.7	8.5	8.3	8.7	8.5	8.3	8.8
50-54	10.4	10.3	10.2	10.1	10.3	10.2	10.1	8.5
55-59	6.9	7.4	7.6	8.0	7.4	7.6	8.0	7.7
60-64	3.5	3.7	4.1	4.3	3.7	4.1	4.3	5.2
65-69	2.8	2.8	2.8	2.8	2.8	2.8	2.8	4.4
70-74	2.6	2.7	2.6	2.5	2.7	2.6	2.5	4.1
75-79	2.3	2.4	2.5	2.5	2.4	2.5	2.5	3.6
80-84	1.4	1.4	1.5	1.6	1.4	1.5	1.6	2.1
85+	1.2	1.2	1.3	1.4	1.2	1.3	1.4	1.4

<표 4-35> 시험조사 결과 - 나이 (계속)

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
0-4	27.6%	24.4%	14.3%	16.0%	18.1%	12.6%	10.0	10.4
5-9	18.8%	18.8%	17.5%	12.4%	12.7%	10.3%	8.2	7.7
10-14	21.5%	20.0%	17.2%	7.2%	14.3%	11.1%	8.3	5.6
15-19	18.1%	11.4%	17.3%	13.9%	10.8%	9.1%	7.5	6.7
20-24	14.0%	10.7%	11.5%	13.3%	9.3%	7.3%	6.6	9.5
25-29	11.4%	12.9%	15.5%	21.9%	8.9%	7.9%	7.6	8.0
30-34	17.3%	18.0%	11.9%	15.3%	12.4%	9.7%	8.2	10.7
35-39	17.6%	27.1%	14.5%	16.4%	16.0%	11.7%	9.6	6.1
40-44	15.2%	8.1%	14.0%	12.6%	8.4%	7.3%	6.3	4.2
45-49	15.7%	16.6%	14.4%	10.6%	11.2%	8.8%	7.1	5.0
50-54	12.2%	11.2%	9.1%	19.4%	8.8%	6.9%	7.0	5.5
55-59	14.6%	17.7%	18.7%	17.6%	11.2%	9.5%	8.4	6.3
60-64	28.6%	20.3%	15.3%	22.8%	17.1%	12.3%	10.7	8.5
65-69	19.9%	19.5%	12.9%	20.3%	13.7%	10.0%	9.8	9.4
70-74	23.0%	29.9%	20.5%	21.9%	18.0%	14.5%	12.1	8.8
75-79	30.6%	40.3%	31.1%	19.7%	24.8%	19.3%	14.9	10.5
80-84	36.9%	35.0%	39.0%	34.6%	23.9%	19.8%	17.1	12.9
85+	48.4%	41.8%	70.2%	44.9%	32.6%	29.6%	24.6	13.9

다. 교육정도

교육정도의 경우 '대학원 이상'을 제외한 대부분의 범주에 대하여 전체 차수에서 CV가 안정적인 것으로 나타났다. 교육정도별로는 대학원 이상으로 통합하면 모든 수준에서 20% 표본 누적결과가 공표 가능할 것으로 판단된다.

<표 4-36> 시험조사 결과 - 교육정도

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
미취학	9.5	9.1	7.6	8.7	9.5	9.0	9.1	12.4
초등학교	10.5	11.6	11.6	10.1	11.0	11.3	11.0	17.1
중학교	8.2	10.2	8.9	9.9	9.2	9.1	9.2	13.5
고등학교	29.3	33.3	33.1	35.6	31.2	31.6	32.6	30.5
대학4년제미만	10.8	14.1	15.3	14.7	12.3	13.2	13.5	10.0
대학4년제이상	28.9	19.5	21.2	19.4	24.4	23.2	22.3	14.9
대학원 석사	2.5	2.0	2.1	1.5	2.3	2.2	2.1	1.3
대학원 박사	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.2

<표 4-36> 시험조사 결과 - 교육정도 (계속)

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
미취학	13.9%	18.1%	10.7%	12.6%	11.5%	8.1%	7.0	5.6
초등학교	13.0%	17.0%	15.2%	14.7%	10.0%	8.2%	7.1	4.3
중학교	14.4%	15.2%	9.3%	10.8%	10.3%	7.5%	6.0	4.0
고등학교	8.5%	7.1%	6.9%	6.4%	5.5%	4.2%	3.5	2.8
대학4년제미만	13.1%	10.0%	10.4%	7.8%	8.7%	6.9%	5.4	6.6
대학4년제이상	14.4%	12.0%	8.3%	10.0%	10.6%	8.0%	6.8	6.9
대학원 석사	27.5%	38.1%	26.9%	30.2%	22.3%	18.2%	16.1	17.7
대학원 박사	100.1%	69.0%	70.4%	101.1%	56.9%	44.0%	41.2	31.1

라. 1년 전 거주지

국내이동(1년전 거주지)별 인구수는 ‘현 거주지’와 ‘이동: 같은 시군’ 범주에서는 모든 차수에서 대체로 안정적인 CV 값을 가지는 것으로 나타났다. 북한 또는 외국 범주에서 다소 불안정하나, ‘다른 지역’으로 병합한다면 20% 표본 누적결과는 안정적인 값을 가질 것으로 판단된다.

<표 4-37> 시험조사 결과 - 1년 전 거주지

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
태어나지 않음	0.9	1.1	0.9	88.9	1.0	0.9	0.9	0.8
현거주지	88.5	88.6	85.8	5.4	88.4	87.7	88.0	93.1
이동:같은 시군	8.9	8.6	7.7	4.6	8.9	8.3	7.4	3.7
이동:다른 시군	1.3	1.6	5.7	0.3	1.5	3.0	3.4	2.4
북한 또는외국	0.5			0.8	0.2	0.1	0.1	0.1
무응답							0.1	

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
태어나지 않음	37.0%	28.2%	29.5%	2.4%	22.4%	18.1%	15.6	18.7
현거주지	1.6%	2.7%	2.2%	22.2%	1.6%	1.3%	1.1	0.8
이동:같은 시군	17.3%	23.1%	15.9%	27.9%	14.4%	10.9%	10.1	14.5
이동:다른 시군	61.6%	28.5%	19.1%	98.1%	29.4%	20.3%	16.5	16.4
북한 또는외국		99.5%			99.8%	99.9%	79.9	61.5
무응답	67.5%			33.5%	71.2%	72.3%	73.6	

마. 통근·통학 여부

통근·통학 여부의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다.

<표 4-38> 시험조사 결과 - 통근·통학 여부

<비율 추정치> (단위: %)								
구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	9.3	10.5	10.8	10.8	10.0	10.2	10.4	11.5
안함	41.1	43.2	41.9	46.2	42.2	42.0	43.1	44.7
통근	15.2	14.1	14.3	12.1	14.4	14.0	13.2	14.7
통학	34.4	32.3	33.0	30.9	33.4	33.8	33.3	29.1

<상대 표준오차> (단위: %)								
구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	17.5%	19.1%	10.8%	12.0%	13.1%	9.3%	7.6	7.4
안함	4.7%	6.2%	3.5%	3.9%	3.8%	2.9%	2.4	1.8
통근	13.9%	10.1%	11.8%	8.7%	8.6%	6.9%	5.6	5.3
통학	9.6%	12.6%	7.3%	9.1%	7.6%	5.7%	5.0	3.7

바. 경제활동상태

경제활동상태의 경우 ‘틈틈히 일함’과 ‘일시 휴직’ 범주를 제외하고는 대체로는 20% 표본은 물론 5%, 10%, 15% 표본 누적결과 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다. ‘일하였음’의 하위항목인 ‘틈틈이’와 ‘일시 휴직’은 응답자가 적어 추정치의 변동이 크며, CV 값도 안정적이지 않다. ‘일하였음’의 하위항목 중 ‘틈틈이’와 ‘일시 휴직’을 병합하면 20% 표본 누적결과는 안정적인 값을 가질 것으로 판단된다.

<표 4-39> 시험조사 결과 - 경제활동상태

<비율 추정치> (단위: %)								
구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	13.9	14.4	13.1	13.8	14.1	13.7	13.6	15.6
일함: 주로	42.9	42.2	42.0	44.6	42.7	42.4	43.1	46.0
일함: 틈틈히	1.2	0.7	1.7	2.7	0.9	1.1	1.5	0.6
일시 휴직	0.1	0.8	0.3	0.1	0.4	0.4	0.3	0.4
일하지 않음	41.9	42.0	42.9	38.8	41.8	42.3	41.5	37.3

<상대 표준오차> (단위: %)								
구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	12.3%	17.5%	9.5%	9.8%	10.7%	7.6%	6.4	5.8
일함: 주로	4.5%	6.1%	3.5%	4.0%	3.6%	2.8%	2.3	1.9
일함: 틈틈히	32.6%	37.9%	44.5%	23.0%	24.9%	23.8%	18.0	27.9
일시 휴직	102.6%	35.2%	54.2%	99.9%	37.3%	30.7%	30.2	26.9
일하지 않음	7.9%	9.7%	5.6%	6.3%	6.0%	4.4%	3.9	2.6

사. 직업: 종사상 지위

종사상 지위의 경우 ‘무급 가족’ 범주를 제외하고는 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 모두 안정적인 추정치와 CV를 가지는 것으로 나타났다. ‘무급 가족’을 포함한 모든 범주에서 20% 표본의 누적결과가 안정적인 값을 가질 것으로 판단된다.

<표 4-40> 시험조사 결과 - 종사상 지위

<비율 추정치> (단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	55.9	56.4	55.9	52.6	56.1	56.1	55.1	53.0
임금 근로자	34.0	36.9	37.0	39.5	35.4	35.9	36.8	26.3
고용원 없는 자영업	6.1	3.9	4.8	4.8	5.0	4.9	4.9	12.5
고용원 있는 자영업	2.7	2.1	1.7	2.3	2.5	2.3	2.3	2.4
무급 가족	1.3	0.7	0.6	0.7	1.0	0.9	0.8	5.7

<상대표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	3.5	4.8	3.0	3.5	2.9	2.2	1.9	1.7
임금 근로자	6.2	6.4	5.5	3.8	4.4	3.5	2.8	2.9
고용원 없는 자영업	9.4	16.7	20.1	15.9	9.8	9.2	7.8	4.9
고용원 있는 자영업	29.3	24.3	22.3	26.4	19.5	16.4	14.2	9.8
무급 가족	37.2	40.4	39.5	42.2	27.9	23.5	20.2	7.8

아. 혼인상태

혼인상태의 경우 모든 범주에서 20% 표본 누적결과는 안정적인 추정치와 CV 값을 가지는 것으로 나타났다.

<표 4-41> 시험조사 결과 - 혼인상태

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	15.7	14.3	13.1	13.8	15.0	14.3	14.1	16.9
미혼	25.6	28.6	25.2	22.6	26.9	25.9	24.8	19.6
유배우	47.8	42.1	51.7	53.4	45.2	47.6	49.2	51.3
사별	6.0	7.0	6.8	5.2	6.4	6.7	6.5	9.0
이혼	4.9	8.0	3.3	5.0	6.4	5.5	5.4	3.0
미상								0.1

<표 4-41> 시험조사 결과 - 혼인상태 (계속)

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	16.3%	17.8%	9.5%	9.8%	11.9%	8.4%	6.8	6.3
미혼	8.8%	9.8%	5.9%	8.1%	6.8%	4.7%	4.1	4.5
유배우	6.5%	8.2%	4.4%	4.8%	5.3%	3.8%	3.1	1.7
사별	29.3%	23.0%	22.5%	23.9%	18.3%	14.5%	12.6	7.9
이혼	34.0%	28.0%	36.5%	34.1%	21.6%	19.1%	16.4	10.4
미상								100.0

자. 총 출생아수

총 출생아수의 경우 남자, 여자 모두 자녀수가 4명 이상의 경우 다소 불안정하며 3명 이하에서는 추정치 및 CV 값이 안정적인 것으로 나타났다. 자녀수 4명 이상을 병합한다면 20% 표본 누적결과는 모두 안정적인 CV 값을 가질 것으로 판단된다.

<표 4-42> 시험조사 결과 - 총 출생아수: 남자

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	67.3	73.7	65.3	65.2	70.5	68.5	67.6	64.5
0명	9.1	5.0	10.6	11.2	6.9	8.2	8.9	4.6
1명	15.0	14.6	13.2	15.3	14.8	14.5	14.9	16.6
2명	7.1	5.3	8.9	6.6	6.4	7.1	7.0	10.5
3명	1.0	0.9	1.8	1.4	1.0	1.2	1.2	2.4
4명	0.3	0.5	0.1	0.4	0.4	0.3	0.4	0.6
5명								0.2
8명	0.3				0.1			
19명								0.5
20명								0.1

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	1.9	2.5	1.7	1.7	2.0	1.5	1.2	1.3
0명	16.7	30.8	8.9	12.2	15.7	10.2	8.0	7.4
1명	7.9	5.3	9.5	6.6	4.8	4.5	3.7	2.8
2명	11.0	15.8	12.9	14.6	9.8	7.5	6.4	3.9
3명	34.9	35.2	36.3	30.1	24.8	20.4	16.4	12.1
4명	56.4	66.1	103.4	66.7	43.4	41.2	34.7	17.7
5명								31.0
8명	102.6				101.1			
19명								100.3
20명								100.3

<표 4-43> 시험조사 결과 - 총 출생아수: 여자

<비율 추정치>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	68.0	74.1	65.3	65.2	70.8	68.8	67.8	65.0
0명	11.6	5.4	13.7	12.0	8.7	10.5	11.0	8.7
1명	13.3	12.7	14.9	14.3	12.9	13.4	13.6	14.3
2명	5.9	5.3	3.5	7.2	5.6	5.0	5.6	7.5
3명	0.8	2.1	1.7	0.9	1.5	1.6	1.5	2.6
4명	0.1	0.5	0.7	0.3	0.3	0.4	0.4	1.2
5명	0.2		0.2	0.1	0.1	0.1	0.1	0.4
6명								0.0
7명								0.0
10명								0.1
11명								0.1
12명								0.2

<상대 표준오차>

(단위: %)

구분	단순				누적			거창읍 16.6%
	1차 4.7%	2차 4.7%	3차 4.9%	4차 5.1%	2차 9.4%	3차 14.3%	4차 19.6%	
해당 없음	2.0	2.8	1.7	1.7	2.0	1.6	1.3	1.3
0명	8.2	33.3	7.6	12.6	13.6	9.1	7.4	6.1
1명	7.8	8.3	7.4	8.9	5.5	4.3	3.7	2.7
2명	10.3	16.2	14.3	14.3	8.8	8.6	7.8	4.6
3명	21.8	29.4	26.4	35.3	22.1	18.4	16.5	10.1
4명	105.3	43.6	56.7	55.7	45.0	35.8	30.0	14.2
5명	63.6		103.4	102.0	66.9	55.5	48.2	28.2
6명								100.1
7명								100.3
10명								100.3
11명								100.3
12명								100.3

2. 가구 부문

가. 가구 구분

가구 구분에 따른 가구 수는 ‘가족으로 구성된 가구’와 ‘1인 가구’의 경우는 20% 표본 누적결과 추정치가 안정적이거나 나머지 가구 특성에 대해서는 추정치의 변동이 크다. 따라서 공표 시 ‘가족’과 ‘가족+가족 이외’의 병합, ‘남남 5인 이하’와 ‘남남 6인 이상’의 병합 등이 필요할 것으로 판단된다.

<표 4-44> 시험조사 결과 - 가구 구분

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
가족	74.8	74.1	72.6	73.7	74.1	73.4	73.1	66.5
가족+가족이외	0.7	1.0	0.5		0.9	0.8	0.6	0.4
1인 가구	24.2	24.9	25.8	25.5	24.9	25.4	25.9	31.9
남남 5인이하	0.2		1.1	0.7	0.1	0.4	0.5	0.4
남남 6인이상								
무응답								0.7
미상								0.1

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
가족	9.6%	8.5%	8.7%	8.2%	6.3%	5.1%	4.4	3.5
가족+가족이외	72.7%	59.9%	100.2%		45.4%	41.7%	42.1	35.2
1인 가구	29.5%	25.8%	25.0%	23.7%	19.1%	15.1%	12.7	7.1
남남 5인이하	99.3%		81.5%	50.7%	99.4%	68.2%	46.9	33.6
남남 6인이상								
무응답								80.7
미상								100.0

나. 거주 기간

거주 기간의 경우 '25년 이상'을 제외하고는 20% 표본은 물론 5%, 10%, 15% 표본 누적결과 모두 안정적인 추정치와 CV 값을 가지는 것으로 나타났다. 20년 이상을 병합한다면 20% 표본 누적결과는 모두 안정적인 CV 값을 가질 것으로 판단된다.

<표 4-45> 시험조사 결과 - 거주 기간

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
1년 미만	8.5	10.4	13.3	13.1	9.5	10.8	11.3	6.9
1-2년 미만	7.0	9.5	10.1	11.6	8.3	8.7	9.4	8.9
2-3년 미만	6.3	9.5	9.6	10.6	7.8	8.5	9.0	6.6
3-5년 미만	4.9	10.4	13.0	9.8	7.5	9.3	9.4	9.5
5-10년 미만	24.1	22.2	20.7	26.0	23.1	22.2	23.3	22.8
10-15년 미만	21.1	12.9	12.3	12.8	17.1	15.5	14.8	15.7
15-20년 미만	10.7	13.3	11.2	6.7	12.0	11.7	10.5	6.9
20-25년 미만	15.8	11.5	9.4	9.3	13.8	12.6	11.8	8.3
25년 이상	1.2	0.3	0.4		0.7	0.6	0.5	13.4
무응답								0.9
미상	0.4				0.2	0.1	0.1	0.1

<표 4-45> 시험조사 결과 - 거주 기간 (계속)

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
1년 미만	13.3%	23.0%	12.5%	21.9%	13.9%	9.8%	9.3	12.2
1-2년 미만	17.9%	28.9%	24.3%	26.5%	18.2%	14.2%	12.4	12.7
2-3년 미만	14.2%	17.3%	18.3%	19.8%	12.4%	10.4%	9.2	10.2
3-5년 미만	29.2%	17.4%	18.7%	25.6%	16.7%	12.9%	11.3	10.5
5-10년 미만	9.7%	9.5%	12.9%	14.5%	6.7%	6.0%	5.9	9.4
10-15년 미만	13.5%	13.9%	9.6%	19.2%	11.4%	9.0%	7.9	10.3
15-20년 미만	13.2%	17.6%	22.4%	27.0%	11.4%	10.2%	10.1	17.2
20-25년 미만	18.0%	27.2%	19.3%	27.3%	15.7%	12.8%	11.5	13.2
25년 이상	42.3%	100.5%	100.0%		40.3%	37.8%	38.8	16.6
무응답								81.0
미상	98.3%				98.9%	99.2%	99.3	99.8

다. 사용 방 수: 잠을 자는 방

잠을 자는 방의 경우 3개 이하에서 20% 표본 누적결과가 안정적인 것으로 나타났으며, 3개의 이상의 경우를 병합하면 20% 표본 누적결과는 모두 안정적인 CV 값을 가질 것으로 판단된다.

<표 4-46> 시험조사 결과 - 사용 방 수: 잠을 자는 방

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
0개								0.0
1개	25.3	35.8	45.5	43.7	30.9	35.7	38.0	52.5
2개	42.0	38.8	42.4	39.1	39.8	40.7	40.5	28.9
3개	30.1	21.4	12.1	17.2	25.9	21.3	19.8	16.0
4개	2.7	4.1			3.3	2.3	1.7	1.1
5개								0.1
무응답								1.3

<상대 표준오차>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
0개								99.8
1개	28.3%	20.0%	14.8%	13.3%	16.8%	11.9%	9.3	4.5
2개	20.4%	16.7%	11.5%	8.0%	13.0%	9.4%	7.4	5.8
3개	32.5%	46.4%	48.7%	35.8%	26.7%	24.1%	20.4	10.2
4개	67.1%	100.0%			64.7%	64.0%	64.7	47.2
6개								99.9
무응답								79.1

라. 난방 시설

시지역인 연수3동의 경우 ‘지역난방’과 ‘도시가스’ 범주에서만 20% 표본 누적결과가 안정적인 CV 값을 가지며, 군지역인 거창읍의 경우에는 ‘도시가스’, ‘기름보일러’, ‘프로판가스’ 범주에서 안정적인 CV 값을 가지는 것으로 나타났다.

<표 4-47> 시험조사 결과 - 난방 시설

<비율 추정치> (단위: %)								
구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
중앙난방		0.3	8.8		0.2	2.8	2.2	6.6
지역난방	43.7	72.6	59.8	36.7	58.0	58.8	53.0	1.5
도시가스	56.1	27.1	31.3	63.3	41.7	38.3	44.8	29.7
기름보일러								30.2
프로판가스								24.4
전기보일러								4.3
연탄보일러								0.6
연탄아궁이								0.2
재래식아궁이								0.1
기타								1.9
무응답	0.2				0.1	0.1	0.1	0.4

<상대 표준오차> (단위: %)								
구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
중앙난방		100.0%	99.5%		100.0%	96.0%	96.2	40.0
지역난방	34.2%	18.2%	24.8%	40.0%	17.5%	14.0%	13.6	94.5
도시가스	26.5%	49.1%	44.2%	23.2%	24.3%	21.2%	16.0	15.3
기름보일러								12.1
프로판가스								17.1
전기보일러								25.5
연탄보일러								29.3
연탄아궁이								64.0
재래식아궁이								62.4
기타								30.3
무응답	101.3%				100.5%	100.1%	99.9	46.3

바. 점유 형태

주거 형태별 가구 수는 ‘자가’와 ‘전세’의 경우 20% 표본은 물론 5%, 10%, 15% 표본 누적결과가 안정적인 CV 값을 가지고 있다. ‘보증금 없는 월세’와 ‘사글세’, ‘무상’ 범주를 병합할 경우 20% 표본 누적결과는 공표가능한 수준이 될 것으로 판단된다.

<표 4-48> 시험조사 결과 - 점유 형태

<비율 추정치>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
자가	56.6	47.4	59.6	58.5	52.0	54.2	55.0	70.9
전세(월세없음)	10.4	14.7	20.4	17.1	12.5	14.9	15.4	6.1
보증있는월세	31.8	36.5	13.8	18.5	34.1	28.0	26.1	16.4
보증없는월세			0.4	0.4		0.1	0.2	
사글세								0.9
무상(관사,사택)	1.3	1.4	5.8	5.5	1.4	2.7	3.4	2.1
무응답								1.3

<상대표준오차>

(단위: %)

구분	단순				누적			거창군 18.7%
	1차 5.2%	2차 5.1%	3차 5.1%	4차 5.1%	2차 10.2%	3차 15.4%	4차 20.5%	
자가	20.3%	21.6%	11.5%	14.5%	14.6%	10.3%	8.5	4.2
전세(월세없음)	21.3%	29.0%	13.7%	17.2%	19.2%	12.8%	10.5	11.2
보증있는월세	42.1%	34.0%	60.7%	46.7%	26.2%	24.1%	21.4	17.1
보증없는월세			100.0%	98.6%		99.8%	69.8	
사글세								27.0
무상(관사,사택)	57.2%	69.4%	48.1%	43.5%	44.0%	35.4%	27.8	15.6
무응답								83.3

제5장 한국에서의 순환센서스 기능

제1절 순환센서스의 사회적 기능

1. 인구, 사회학적 의미

순환센서스는 각 지역의 변화를 정기적으로 측정하고 비교하는 기록으로서의 의미를 가진다. 지역 단위에서 종단적인 조사를 추적할 수 있기에 순환센서스는 지역 주민의 의식과 주관적 평가를 포함하여 세밀한 변수와 지표들을 매 년 혹은 최소 5년 간격으로 제공할 수 있다. 이 정보들은 중앙정부와 지방정부가 정책을 개발하고 평가하는 데에도 즉각적으로 활용이 가능하기 때문에, 활용성을 극대화할 수 있는 문항들을 개발하는 것이 필요하다. 나아가 개인과 지역을 연결할 수 있는 연구들의 특성을 검토함으로써 이를 촉진할 수 있는 지표들 위한 문항이 필요하다.

순환센서스를 통해 생산된 정보는 정부뿐만 아니라 민간에게까지 공개되어 지역 공동체 발전에 이바지할 수 있다. 순환센서스가 공공재로서의 역할을 할 수 있는 것이다. 소득 지니 계수와 같이 지역별 공개가 민감한 변수들의 경우에는 심사를 거쳐 제한적으로 공개하거나 지역의 크기에 따라 공개 변수의 범위를 위계적으로 제한하는 것도 고려할만 하다. 또한 순환센서스의 특성상 시군구와 같이 상대적으로 큰 단위는 연도별로 추정치를 제공하고, 작은 단위인 읍면동에는 3년이나 5년 간격의 추정치로 제공하면 자료의 신뢰성과 비교를 통한 활용성을 모두 확보할 수 있을 것이다.

한편 순환센서스는 기존 센서스와 달리 좀 더 다양한 설문이 가능하므로, 지역별 집계 지표들뿐만 아니라 각 개인의 설문을 대표성 있게 표집하여 마이크로데이터(microdata) 형태로 제공함으로써 분석적인 연구들을 포함하여 그 활용 범위를 크게 넓힐 수 있다. 뿐만 아니라 기타 서베이 자료의 응답자/응답가구의 주거 지역과 순환센서스의 지역별 정보를 결합하여 분석한다면 한국 사회의 공동체적 맥락의 중요성을 읽어낼 수 있는 방향으로 나아갈 수 있다는 점에서 잠재력이 매우 크다고 할 수 있다(노병일 & 곽현근, 2000).

2. 정책적 의미

순환센서스의 목적은 각 지역의 특성을 파악하고 비교함으로써 더 나은 공동체를 구축할 수 있는 토대를 만드는 것이라고 할 수 있다. 각 지방 정부가 정책을 수립하고 또한 시행하고 있는 정책의 효과를 평가하고 함의를 도출하는 데에 순환센

서스는 보다 빠르고 정확한 자료를 통합적으로 제공하는 역할을 할 수 있다. 순환 센서스를 조사하는 과정에서 시간에 따른 전출입에 의한 모집단의 변화를 파악하는 등 주기적인 조사에 지자체와의 협조가 매우 중요하기 때문에, 지자체가 정책을 평가하고 새로운 정책을 수립하는 데에 원활히 반영할 수 있도록 자료를 통합하고 가공해서 제공하는 것이 필요하다.

3. 학문적 기능

오늘날 전미 3대 경제학 저널을 포함한 세계적인 수준의 사회과학 저널들에서는 기존의 개인 위주의 서베이 자료보다는 실험의 기법이나 행정 자료를 활용한 연구가 날로 증가하고 있으며, 주변 분야와 하위 저널들에도 급격히 확산되고 있다. 순환센서스는 기존의 행정 자료가 제공하던 정보의 질과 폭을 크게 넓힘으로써 늘어나는 수요와 발맞추어 학문적인 발전에 기여할 뿐만 아니라 정책적 함의가 높은 연구들을 양산해 낼 수 있을 것으로 기대된다.

본 장에서는 이러한 순환센서스의 인구사회학적, 정책적, 학문적 기능을 어떻게 극대화시킬 수 있는지 모색해볼 것이며, 우선 이러한 기능을 성공적으로 수행하고 있는 미국의 순환센서스, ACS부터 검토해볼 것이다. 지금까지의 연구에서 ACS의 설계 및 조사 특징들은 다루어온만큼, 본 연구에서는 ACS의 특징은 개괄적으로 살펴보고, 실제로 어떻게 자료가 제공되고 학문적, 정책적 활용이 어떻게 이루어지고 있는지를 좀 더 자세히 살펴보고자 한다.

제2절 American Community Survey(ACS) 적용 가능성 검토

1. ACS 조사의 특징

가. ACS의 조사 설계 개관

ACS는 2005년부터 미국 전역에 약 290만 가구를 표집하여 조사하는 것으로 시작하여 2013년에는 약 350만 가구를 표집하는 등 지속적으로 확대되어 발전했다. 응답자가 서베이에 참여하는 것은 연방법(Title 13 U.S.C. Section 221)에 의해 응답 가능한 모든 문항에 의무적으로 응답하도록 강제하였으며, 응답을 거부하거나 고의적으로 불성실하게 응답한 경우 최대 \$100의 벌금을, 거짓으로 응답한 경우 최대 \$500의 벌금이 부과하는 등 양질의 응답을 확보하기 위한 제도적 절차가 마련되어 있으며, 서베이 참여에 대한 인센티브는 별도로 주어지지 않는다(박시내, 2012).

응답자를 표집하기 위해 주소 파일(MAF, Mail Address File)을 행정 자료와 위성에 의한 위치 정보 등을 종합하여 매년 갱신한다. 전체 주소에서 표집은 가구

단위(housing unit)과 공동으로 거주하는 집단 쿼터(group quarter)로 나누어 실시한다. 가구들은 균등하게 나뉜 5개의 하위 프레임(sub-frame) 중 하나에 속해 5년마다 돌아가면서 조사되기 때문에, 매 년 전체를 대표할 수 있는 표집을 할 수 있으면서 동시에 한 번 조사된 가구는 향후 5년간 표집 대상에서 제외되는 서베이 휴식 프로그램(Survey Holiday Program)을 적용한다. 향후 실시되는 한국의 순환서베이에서도 전체 모집단의 20% 중에서 추출하는 등 유사점이 많기 때문에 휴식 프로그램은 고려할 만한 방식이라고 할 수 있다. 매 년 1월에 새로 추가되는 주소들이 하위 프레임에 체계적으로 배분하고, 9-10월에는 전체의 주소를 16개의 층(stratum)으로 나누어 표집한다. 한편 저응답률 지역이나 에스키모, 인디언 등 소수민족이 거주하는 지역들은 표집률을 다른 지역보다 높게 설정함으로써 이들의 대표성을 확보할 수 있도록 하였다. 한국에서도 현재 인구주택총조사 표집틀에 외국인 비율을 고려하여 취약지역에 추가 조사구를 배분하고 있는데, 다문화 가정이나 외국인 노동자 2세 조사 등을 위해 이러한 표집 방식을 유지, 발전시켜야 할 것이다.

<표 5-1> 월별 ACS 자료 수집 과정

샘플 패널	자료 수집 월					
	2013년					
	1월	2월	3월	4월	5월	6월
2012년 11월	방문 조사					
2012년 12월	전화 조사	방문 조사				
2013년 1월	우편/인터넷	전화 조사	방문 조사			
2013년 2월		우편/인터넷	전화 조사	방문 조사		
2013년 3월			우편/인터넷	전화 조사	방문 조사	
2013년 4월				우편/인터넷	전화 조사	방문 조사
2013년 5월					우편/인터넷	전화 조사
2013년 6월						우편/인터넷

출처: Census Bureau of the United States. (2014). American Community Survey Design and Methodology.

나. ACS 조사 과정

우선 각각의 층에 맞게 표집률이 정해지면, 매 달 나누어 조사될 주소들을 표집하는 일차 표집(first sampling)을 실시한다. 조사는 총 3달에 나누어서 실시되며, 첫 번째 달에 조사되는 모든 주소에 설문지를 우편으로 보내고 인터넷 응답도 가능하게 한다. 첫 달 안에 응답이 이루어지지 않은 경우 두 번째 달에 전화 조사를 시도한다. 이렇게 두 달동안 응답이 완료되지 않은 가구의 경우에는, 주소가 불완전하거나 사서함 주소를 사용하여 이차 표집(second sampling)되는 대상들과 동일하게 면접원이 방문하는 면접 조사(CAPI: Computer Assisted Phone Interview)를 받게 된다.

한편 응답자가 집단 쿼터에 속한 경우, 즉 기숙사나 요양원, 군대, 노숙자 재활센터 등의 시설 혹은 기관에 거주할 때에는 별도의 표집이 이루어진다. 표집은 시

설이 아니라 거주자(resident) 단위로 이루어진다. 각 시설은 15인 이하의 작은 집단 쿼터와 그 이상인 집단 쿼터로 나뉘는데, 작은 집단 쿼터에 속하는 시설은 가구 조사와 마찬가지로 5개의 하위 프레임으로 우선 나뉘어, 시설을 먼저 표집하고 그 안에서 다시 응답자를 표집한다. 16인 이상의 시설들은 따로 하위 프레임을 나누지 않고 대신에 각 시설 내 거주민을 10인 집단 단위(groups of ten)로 묶은 후에 이 10인 단위를 표집한다. 면접원은 해당 시설에 방문한 후에 그 집단 내에서 1명을 무작위로 골라 응답을 받는다.

다. ACS 조사 내용

ACS 내용 협의회(ACS Content Council)를 구성하여 중앙 정부, 지방 정부, 대학 및 사기업들의 요구를 받아들이고 조정한다. ACS는 연방법에 따라 이루어지는 센서스 조사의 특성상 제약이 많고 내용이 엄격하게 제한되는데, 2003년에 OMB(Office of Management and Budget)에 의해 최종 승인된 설문지는 2000년 센서스의 long form과 거의 유사하게 25개의 주거 관련 설문과 42개의 인구 관련 설문으로 구성되어 있다. 그 내용은 다음의 <표 5-2>와 같다.

<표 5-2> 2015년 ACS 설문지

설문 대상	설문 내용
주거에 관한 설문	주거의 종류, 거주하는 건물이 지어진 년도, 언제부터 해당 건물에 살았는지, 집의 규모, 해당 지역에서 농업을 했다면 벌어들인 수입, 사업(business)으로 쓰고 있는지 여부, 방과 침실의 개수, 집에서 갖추고 있는 시설 유무(냉온수, 화장실, 욕조나 샤워실, 난방기구, 냉장고, 전화), 컴퓨터 유무(크기 기준으로 랩탑 이상 태블릿 이하-스마트폰 포함-로 구분) 및 인터넷 연결 여부와 서비스 종류, 자동차 수, 사용하는 난방 연료(가스, 전기, 기름, 석탄, 태양열, 기타 등등), 지난 달 전기비, 가스비, 수도비 연료비 각각의 액수 지난 1년간 식량 지원(food stamp) 혹은 영양보충 지원(supplemental nutrition assistance)을 받은 가구 구성원이 있는지 여부 집세 외의 관리비(condominium fee) 액수, 집이 자가인지 세(rent)인지, 그리고 자가라면 모기지(mortgage)나 론(loan)이 남아 있는지 아닌지 여부, 월세 액수, 자가일 경우 그 가치와 부동산세와 1년간 보험료, 다달이 갚는 모기지 액수
가구 구성원(최대 5명) 개인별 설문	성명, 가족관계, 성별, 연령, 인종 출생지, 시민권자 여부, 미국에 살기 시작한 시간, 교육수준, 대졸자면 학위를 받은 전공, 언어 사용, 1년 전에 살았던 장소, 가입한 의료보험 종류, 장애 여부(6가지 질문), 혼인 상태 및 과거 혼인력, 1년간 자녀 출생 여부, 같이 사는 손자나 손녀 유무 및 이들에 대한 생계 책임 여부, 군인 등의 공적 서비스 이력 유무 및 복무 시기 노동 여부, 직장의 위치, 직장명, 출근 시간 및 통근하는 데에 걸린 시간, 일시 해고(lay off) 여부 및 복직 가능성 여부, 노동 시간, 직장 종류(사기업, 정부, 자영업, NGO 등), 직장이 속한 산업, 직장의 종류(제조업, 도매업, 소매업, 기타 등등) 지난 1년간의 종류별 소득(월급, 자영업 소득, 투자에 따른 수익, 사회 보장(social security or railroad retirement), 소득 보전(supplemental security income), 공적 혜택이나 공적 지원, 각종 연금) 및 이를 총괄한 전체 수입
설문 대상 이외의 가구 구성원	성명, 성별, 연령

ACS의 설문 문항은 요청에 따라 연 단위로 수정될 수 있고, 이 경우 1년에 걸친 사전 조사(pre-testing)를 반드시 거치고 OMB의 승인을 받아야 한다. 이를테면 2003년에 개발된 설문지는 2007년까지 사용되었고, 2008년과 2009년에는 2006년에 진행된 내용 검사(content testing)에 따라 소폭 수정된 설문지가 사용되었다. 같은 과정을 거쳐 2010년부터는 컴퓨터 소유와 인터넷 사용과 관련된 문항이 추가되어 2015년까지 유지되었다. 그에 따라 2015년의 설문지는 24개의 주거 관련 문항과 48개의 인구 관련 문항으로 구성되어 있다.

라. ACS 자료 공개 방식: 개인 원자료(microdata) 공개방식

ACS 개인 원자료는 2000년까지 진행된 센서스와 동일한 방식으로 표집한 PUMS(Public Use Microdata Sample)을 로그인 없이 누구에게나 공개하는 방식으로 제공되고 있다. 2015년 현재에는 2000년부터 2013년까지의 마이크로데이터를 주별로 1년 자료는 물론, 3년치와 5년치의 누적 자료까지 가구 자료와 개인 자료로 나누어서 업로드하였으며, 이 가구 자료와 개인 자료는 별도의 가구별 ID를 통해 병합하는 것도 가능하다.

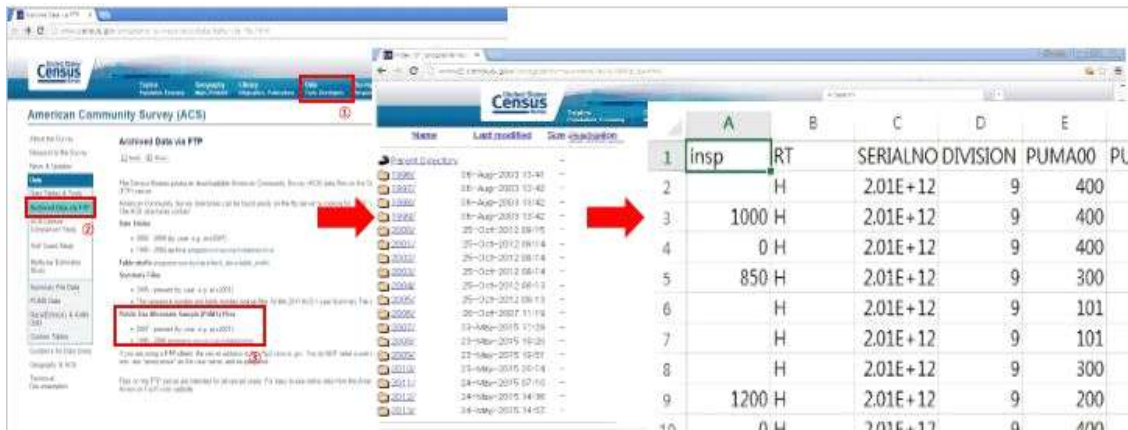
<표 5-3> 2015년 ACS 설문지

종류	기간	분할	총 변수	설문 변수	가중치	자료별 샘플	전체 샘플	
개인 자료	1년치	a	283	203	80	1,613,672	3,132,795	
		b	283	203	80	1,519,123		
	3년치	a	298	218	80	2,770,320	9,357,842	
		b	298	218	80	2,049,386		
		c	298	218	80	2,201,475		
		d	298	218	80	2,336,661		
	5년치	a	295	215	80	4,567,374	15,450,262	
		b	295	215	80	3,394,864		
		c	295	215	80	3,636,230		
		d	295	215	80	3,851,794		
	가구 자료	1년치	a	231	151	80	756,065	1,476,313
			b	231	151	80	720,248	
3년치		a	209	129	80	1,285,588	4,438,696	
		b	209	129	80	986,534		
		c	209	129	80	1,061,244		
		d	209	129	80	1,105,330		
5년치		a	205	125	80	2,090,075	7,219,000	
		b	205	125	80	1,608,820		
		c	205	125	80	1,725,672		
		d	205	125	80	1,794,433		

한편, ACS는 설문 내용을 통해 응답자의 신원이 특정되는 것을 방지하기 위해 (1) 자료에서 이름과 주소와 관련된 응답을 삭제하고, (2) 극단치에 의해 신원 파악이 가능한 개방형 질문, 예를 들어 연령이나 소득, 각종 세금액이나 비용, 시간, 주거 환경 등의 경우 탑코딩(top coding), 바텀 코딩(bottom coding) 방식으로 실제 값을 특정한 값으로 대체하는 방식으로 제공한다. (3) 또한 지역 변수를 지역(region, 총 4개), 구역(division, 총 9개), 주(state, 총 41개) 이외에는 PUMA(Public Use Microdata Areas, 총 2,378개)로만 제공하고 각각을 인구 100,000명 이상이 되도록 하고 있다. 한국의 순환센서스도 개인과 가구 단위의 원자료를 공개할 지역 단위를 결정할 필요가 있다.

2015년 현재 가장 최근의 ACS 자료로 제공되는 2013년의 PUMS의 경우, 원자료는 가구의 경우 1년치가 150만여 가구의 151개 변수를, 3년치는 450만여 가구의 129개 변수, 5년치는 750만여 가구의 125개 변수를 제공하고 있고, 개인자료의 경우 300만여 명의 203개 변수, 3년치는 900만여 명의 218개 변수, 5년치는 1,500만여 명의 215개 변수를 제공하고 있다. 한국의 순환센서스에서 조사할 적정 변수와 자료의 크기를 확정하는 것 또한 중요한 과제라고 할 수 있으며 아래에서 예시 조사방식을 제시해보려 한다.

[그림 5-1] ACS 개인자료 공개

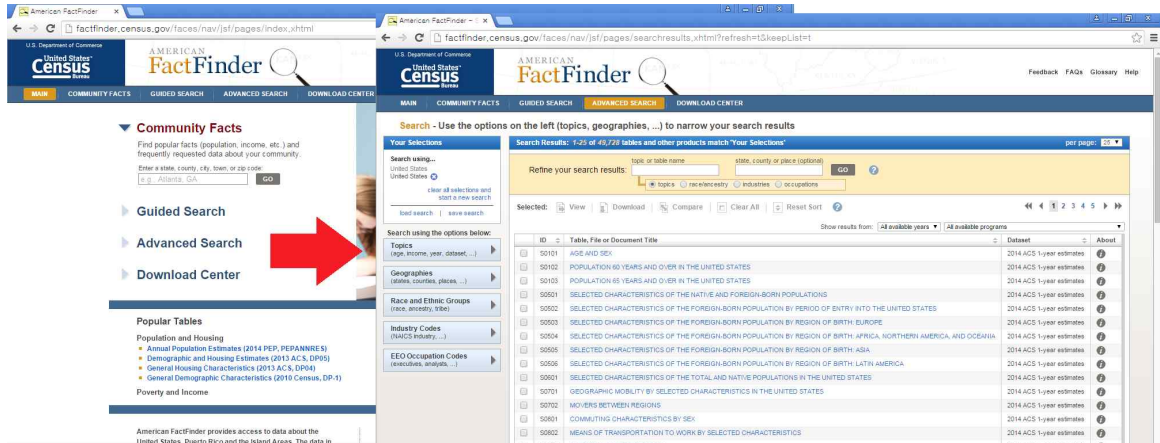


마. ACS 집계 자료(agggregated data) 공개 방식

지역별 집계 자료는 웹사이트(Americna Fact Finder, factfinder.census.org)를 통해서 주제별, 지역별, 인종 및 민족 집단별, 산업별, 직업별로 나누어서 제공하고 있다. 각 지역의 대표성을 확보하기 위해 1년치와 3년치의 추정치는 각각 인구 65,000명 이상과 20,000명 이상인 지역 단위에서만 공개한다. 결과적으로 1년치는

대략 800여 개의 카운티(county)와 500여 개의 메트로폴리탄(metropolitan)을, 3년치는 대략 1,800여 개의 카운티와 900여 개의 메트로폴리탄을 포괄하게 된다. 비슷한 원칙을 한국의 순환센서스에 적용한다면 군과 구 단위에는 1년치와 3년치의 자료를, 읍과 면 단위에는 5년치의 자료를 공개하는 것과 유사할 것이다.

[그림 5-2] ACS 집계자료 공개



2. ACS 자료의 학문적 활용

ACS 자료를 활용한 학문적 연구들은 크게 기술적(descriptive) 추정 연구와 변수 간 인과관계를 탐구하는 분석적(analytic) 연구로 나누어질 수 있다. 기술적 추정 연구는 다시 (1) ACS를 타 자료의 타당성 검토를 위한 준거 자료로 활용하는 경우와 (2) ACS를 추정을 위한 주 자료로 사용하는 경우로 나누어질 수 있고, 분석적 연구는 (3) ACS를 지역 수준의 변수를 생성하는 데에 사용하는 경우와 (4) 개인 수준의 자료(aggregate data)를 활용하는 경우로 구분된다.

기술적 추정 연구 중에서 (1) 준거 자료로 ACS를 활용하는 연구들의 경우 ACS에서의 성별, 인종별, 연령별 인구 분포를 해당 연구 자료의 인구 분포와 비교함으로써 그 자료의 대표성을 입증하고자 한다. 한편 (2) ACS를 주 자료로 추정하는 연구들은 지역별 실업률 변화 추이와 같이 ACS의 통계치들을 인용하여 현황을 파악하고자 한다.

(3) ACS를 지역 수준의 변수를 생성해 분석하는 연구들은 다양한 지역 단위들, 즉 동네(neighborhood), 지역 사회(community), 카운티, 주 등의 추정치들을 기반으로 다른 지역 변수들, 이를테면 지니 계수나 빈곤율, 히스패닉 비율 등과 병합하여 인과 관계를 입증하려고 한다. 마지막으로 (4) 개인 수준의 자료를 분석하는 경우는 개인 자료를 활용하여 특정한 사회 현상의 개인적 요인과 지역적 요인을 변인으로

하는 다양한 분석을 실시한다.

이처럼 ACS를 활용하는 연구는 그 주제와 분석 단위에 따라 매우 다양하지만, 그 빈도는 가장 간편한 방법인 준거 자료로서의 활용이 가장 많으며 반대로 개인 자료 분석이 가장 적다. 반면 연구의 수준이 ACS를 보다 적극적으로 활용할수록 보다 심화되는 경향이 있기 때문에, 분석적인 연구들이 보다 좋은 저널에 실리는 경우가 많다. 따라서 한국의 순환센서스가 지역 집계자료나 개인 자료를 손쉽게 쓸 수 있도록 제공하는 것이 양질의 풍부한 분석을 활성화시키는 데에 바람직할 것으로 보인다. 아래 표는 ACS를 이용한 네 가지 연구의 예들을 정리하고 있다.

<표 5-4> ACS를 활용한 기술적 추정 연구의 예

항목	연구	연구내용 요약
(1) 준거 자료로 활용	Burkhauser, R. V., A. J. Houtenville and J. R. Tennant (2014). "Capturing the Elusive Working-Age Population With Disabilities: Reconciling Conflicting Social Success Estimates From the Current Population Survey and American Community Survey." <i>Journal of Disability Policy Studies</i>	ACS와 CPS(Current Population Survey)에서 장애(disability)를 묻는 5-6가지 질문 문항이 장애인구의 수와 그들이 받는 복지 혜택의 비중을 과소추정(underestimate) 하고 그들의 실업률을 과대추정할 가능성이 있음을 주장하기 위해 해당 문항들에 의해 추정되는 장애인구의 특징을 비교
	Sweeney, S., J. Hastrup, A. Hyland and C. Rivard (2015). "A Follow-Up Community Survey of Knowledge and Beliefs About Cancer and Genetics."	지역 서베이의 인종과 교육수준의 분포를 2006년부터 2010년까지의 ACS와 비교함으로써 연구의 샘플링이 적절하게 이루어졌음을 주장
	Orrenius, P. and M. Zavodny (2009). "Do immigrants work in riskier jobs?"	미국인과 이민자들이 종사하는 산업과 직업을 조사해서, 이민자들이 더 위험한 직업에 일한다는 결과를 보여주기 위해 ACS의 개인 자료를 활용
(2) 추정	Fuller-Thomson, E., B. Yu, A. Nuru-Jeter, J. M. Guralnik and M. Minkler (2009). "Basic ADL Disability and Functional Limitation Rates Among Older Americans From 2000 - 2005: The End of the Decline?" <i>The Journals of Gerontology Series A: Biological Sciences and Medical Sciences</i>	ACS의 2000년부터 2005년까지 자료로 ADL 장애를 가진 사람들의 빈도가 9% 증가했음을 보임
	Kaye, H. S., C. Harrington and M. P. LaPlante (2010). "Long-Term Care: Who Gets It, Who Provides It, Who Pays, And How Much?" <i>Health Affairs</i>	장기간의 돌봄이 필요한 사람들이 나이에 따라, 주거 환경에 따라(자가, 시설, 공동거주 등) 어떻게 분포하는지를 살펴보기 위해 ACS의 2007년 자료를 활용하였고, 이를 바탕으로 누구에게 돌봄을 제공받는지, 누가 그 비용을 지불하는지를 다른 4가지의 대표성 있는 자료를 통해 보임
	Raphael, S. and E. Smolensky (2009). "Immigration and Poverty in the United States." <i>The American Economic Review</i>	이민이 증가할수록 미국의 빈곤율이 증가하는지에 대한 시뮬레이션 분석에서 ACS를 통해 획득한 국가 소득에서 노동 소득 이외에 추가로 발생하는 비중의 추정치를 사용하여, 이민의 효과가 실제로는 나타나지 않거나 미미한 수준임을 보임
	Saenz, V. B. and L. Ponjuan (2009). "The	히스패닉 남성들이 중고등학교에서 대학으로 진

Vanishing Latino Male in Higher Education." Journal of Hispanic Higher Education 8(1): 54-89.	학하는 비중이 점차 떨어지고 있는데, ACS의 인종별/성별 종사 직종을 인용하여 결과적으로 이들이 저숙련직이나 서비스직에서 차지하는 비중이 높아지고 있음을 보임
Goodman, A. C. and T. G. Thibodeau (2008). "Where are the speculative bubbles in US housing markets?" Journal of Housing Economics	2000년부터 2005년까지 집주인이 직접 거주하는 비율이 각 메트로폴리탄 지역에서 얼마나 증가했는지를 보이기 위해 ACS에서 해당 자료를 추출
Flores, S. M. and J. Chapa (2009). "Latino Immigrant Access to Higher Education in a Bipolar Context of Reception." Journal of Hispanic Higher Education	라틴계 시민 및 비시민(noncitizen)들의 거주지역을 파악하고 비교하기 위해 ACS의 2006년 자료를 사용

<표 5-5> ACS를 활용한 분석적 연구의 예

항목	연구	연구내용 요약
(3) 지역수준 변수 활용	Denton, E.-g. D., J. A. Shaffer, C. Alcantara, L. Clemow and E. Brondolo (2015). "Hispanic residential ethnic density and depression in post - acute coronary syndrome patients: Re-thinking the role of social support." International Journal of Social Psychiatry	지역수준에서 히스패닉의 인구 비중이 높을수록 서로에 대한 사회적 지지가 늘어나기 때문에 우울증이 감소할 것이라는 소수민족 밀도 가설 (ethnic density hypothesis)를 검증하기 위해 2005년부터 2009년까지의 ACS 자료를 활용
	Eckenrode, J., E. G. Smith, M. E. McCarthy and M. Dineen (2014). "Income Inequality and Child Maltreatment in the United States." Pediatrics	ACS를 이용해 country level에서 지니 계수를 추출하여, 소득 불평등이 증가할수록 아동 학대가 증가함을 보여주는 연구
	Edwards, K., M. Mattingly, K. Dixon and V. Banyard (2014). "Community Matters: Intimate Partner Violence Among Rural Young Adults." American Journal of Community Psychology	ACS를 이용해 지역수준에서의 빈곤율을 추출하여, 데이트 상대로부터 당하는 폭력이 인식된 지역 수준의 효능성 (community efficacy)가 낮을수록 높아진다는 연구의 통제변수로 사용
	Neville, L. (2012). "Do Economic Equality and Generalized Trust Inhibit Academic Dishonesty? Evidence From State-Level Search-Engine Queries." Psychological Science	주-수준의 사회경제적 지위, 지니 계수, 일반화된 신뢰 (generalized trust)를 변수화하고, 구글에서 대학 과제를 위해 페이퍼를 비윤리적으로 검색하는 것 (academically dishonest searches)에 불평등이 미치는 영향이 일반화된 신뢰에 의해 매개됨을 밝힘
	Gibson, J. and D. McKenzie (2011). "Eight Questions about Brain Drain." The Journal of Economic Perspectives	각 국가에서 미국으로의 두뇌자원 유출 (brain drain) 추세를 파악하기 위해 2008년 ACS를 활용해 외국인의 비중이 높은 직업군에 대한 정보를 활용
	Porter, M. (2010). "The Rent Gap at the Metropolitan Scale: New York City's Land-Value Valleys, 1990-2006." Urban Geography	젠트리피케이션이 발생하는 조건을 탐색하기 위해 1990년과 2000년의 센서스 자료와 더불어 2006년의 ACS에서 집의 시세 혹은 세를 추출
	Buehler, R. and J. Pucher (2012). "Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes." Transportation	ACS에서 가장 큰 100개 도시 중 90곳을 선정하고, 2006년부터 2008년까지 도시 별로 자전거 출퇴근하는 비중을 추출하여 도시 수준에서 분석
	Rentfrow, P. J., C. Mellander and R. Florida (2009). "Happy States of America:	2005년부터 2007년까지의 주-수준의 집의 가치를 ACS에 추출하여 사용했으며, 다른 경제 지표들

	A state-level analysis of psychological, economic, and social well-being." Journal of Research in Personality	(중위 소득, 지역 총생산)보다도 강하게 개인의 행복과 연관되는 것으로 나타남
	Brown, R. P., L. L. Osterman and C. D. Barnes (2009). ." Psychological Science	ACS의 3년 추정치(2005-2007)의 지니계수를 이용해 불평등을 추정하고, 고등학교에서 학생들이 총기를 가져오는 행위가 그 지역에서 명예를 중요하게 생각하는 주(state)에서 유의하게 두드러짐을 밝힌 연구
	Nagel, C. L., N. E. Carlson, M. Bosworth and Y. L. Michael (2008). "The Relation between Neighborhood Built Environment and Walking Activity among Older Adults." American Journal of Epidemiology	ACS의 1996년 자료를 이용해서 neighborhood 수준에서의 빈곤율, 인식된 문제들, 인식된 걷기의 안정성을 계산하였고, 이를 다른 2001년도의 서베이 자료와 병합하여 사용하여 평소에 걷기 운동을 하는 노인들이 얼마나 평균적으로 오래 운동하는지가 neighborhood 수준에서 자동차의 통행량이나 상업건물의 수와 관련이 있음을 밝힘
	Vigdor, J. (2008). "The Economic Aftermath of Hurricane Katrina." The Journal of Economic Perspectives	카트리나의 경제적 파급을 설명하기 위해 해당 지역의 2004년부터 2006년까지의 인종별 인구 구성비의 변화와 집값의 중위값 소폭 상승한 통계 인용
	Hess, D. B. and P. Ong (2002). "Traditional Neighborhoods and Automobile Ownership." Transportation Research Record: Journal of the Transportation Research Board	자동차 소유에 영향을 미치는 개인적, 지역사회적 요인들을 찾기 위해 ACS의 1996년 자료에서 중위 소득, 가구 밀도(household density), 백인 비율을 추출하여 사용하였으며, 이 중에서 중위 소득이 높을수록 차를 적게 소유하는 것으로 나타남
(4) microdata 분석	Fuller-Thomson, E. and T. Gadalla (2008). "Income inequality and limitations in activities of daily living: A multilevel analysis of the 2003 American Community Survey." Public Health	주(state) 수준의 임금 불평등이 가장 심한 세 주에서 개인 수준의 ADL 제약(ADL limitation)이 가장 높음을 증명하기 위해 2003년 ACS의 샘플 자료인 PUMS(public use microdata set)을 활용
	Sharma, A. (2012). "Exploratory spatial data analysis of older adult migration: A case study of North Carolina." Applied Geography	노스 캐롤리나의 노인들의 해안가 혹은 산지로의 이동이 패턴화됨을 보이기 위해 이동을 설명하기 위해 ACS 자료를 이용
	Fuller-Thomson, E., A. Nuru-Jeter, M. Minkler and J. M. Guralnik (2009). "Black-White Disparities in Disability Among Older Americans: Further Untangling the Role of Race and Socioeconomic Status." Journal of Aging and Health	흑인과 백인의 장애(disability) 불평등의 차이를 소득과 교육수준의 차이로 설명하는데, 2003년의 ACS를 바탕으로 소득과 교육 수준을 독립변수로, 그리고 장애의 2가지 측정인 FL과 ADL을 종속변수로 활용
	Percheski, C. (2008). "Opting Out? Cohort Differences in Professional Women's Employment Rates from 1960 to 2005." American Sociological Review	여성들의 전문직 고용이 보다 어린 코호트에서 증가하는지를 살펴보기 위해 1960년부터 2000년까지는 센서스 자료의 Integrated Public Use Microdata Series를, 2005년 자료는 ACS의 결과 이용해서 비교
	Fuller-Thomson, E. and M. Minkler (2007). "Central American Grandparents Raising Grandchildren." Hispanic Journal of Behavioral Sciences	중앙아메리카 출신의 손자를 돌보는 노인들이 상대적으로 젊고 혼인한 상태이며 고졸 이하의 학력을 가지고 이미 노동시장에서 은퇴한 특징을 지님을 밝힘
	Fuller-Thomson, E. and M. Minkler (2005). "American Indian/Alaskan Native Grandparents Raising Grandchildren:	2000년의 ACS의 부가자료를 이용해 아메리칸 인디언/알래스카 원주민들 중에서 손자녀를 돌보는 노인들이 빈곤하면서 건강이 나쁘고 자원도 제약

Findings from the Census 2000 Supplementary Survey." Social Work	되어 있음을 밝힘
Quigley, J. M. and S. Raphael (2004). "Is Housing Unaffordable? Why Isn't It More Affordable?" The Journal of Economic Perspectives	1960년부터 2000년까지 소득 수준에 따른 주택 감당 정도(housing affordability)의 증가를 살펴보면 2000년 자료는 ACS를 이용
Dávila, A. and M. Mora (2005). "Changes in the earnings of Arab men in the US between 2000 and 2002." Journal of Population Economics.	2000년부터 2002년 사이에 미국에 거주하는 아랍인들의 소득이 백인들과 비교했을 때에 유의하게 낮아졌으며, 그 원인이 임금구조의 변화나 아랍인 자체의 특성으로 설명되지 않았음을 분석하기 위해 개인 자료 활용
Latrice Sykes, L. (2005). "A home of her own: an analysis of asset ownership for non-married black and white women." The Social Science Journal	비혼인 흑인과 백인 사이에 주택 소유에 유의한 차이가 있음을 보이면서 ACS의 부가 자료를 사용
Roland G. Fryer, Jr. and S. D. Levitt (2004). "The Causes and Consequences of Distinctively Black Names." The Quarterly Journal of Economics	흑인들의 이름짓기가 1960년대에 백인과 구별되지 않는 방식이 도입되었고, 1970년대부터 백인과 동화되는 집단과 흑인의 정체성을 부여하는 집단으로 나뉘었는데, 그들의 사회경제적 배경이 실제로 달랐음을 보여주는 연구에서 다양한 자료들과 더불어, ACS를 기반으로 어머니와 자녀들간의 관계를 확인함
Lund, H. (2002). "Pedestrian Environments and Sense of Community." Journal of Planning Education and Research	ACS의 1996년 자료를 기반으로 지역사회(neighborhood)를 선정하여 전통적인 지역사회가 비교적 근래에 형성된 지역사회보다 그 지역에 대한 공동체 의식이 더 높음을 밝힘

ACS를 적극적으로 활용함으로써 깊이 있는 분석을 시도한 주목할 만한 연구로는 Sharma(2012)와 Brown et al(2009)를 들 수 있다. Sharma는 노스 캐롤라이나의 노인들이 해안가나 산지 등의 주변지로 이동하는 것이 패턴화되고 있음을 보여주기 위해 ACS의 지역 집계 자료를 활용하였다. 또한 Brown은 지역수준의 불평등을 ACS의 3년 추정치(2005-2007)의 지니계수를 이용해 추정하였으며, 고등학교에서 학생들이 충기를 가져오는 행위가 명예를 중요하게 생각하는 주(state)에서 유의하게 두드러짐을 밝힌 연구이며, 주 수준의 불평등은 학생들의 충기소지에 유의한 영향을 미치지 않음을 밝혔다.

한편 개인 자료를 활용한 대표적인 연구로는 Rolando & Levitt(2004)와 Dávila & Mora (2005)를 들 수 있다. Rolando & Levitt의 연구는 ACS의 사회경제적 변수를 다른 자료들과 결합하여 사용하였고 흑인들의 이름짓기가 1970년대부터 백인과 동화되는 집단과 흑인의 정체성을 부여하는 집단으로 나뉘었는데, 이 집단들의 사회경제적 배경이 실제로 달랐음을 보여주었다. 또한 Dávila & Mora의 연구는 2000년부터 2002년 사이에 미국에 거주하는 아랍인들의 소득이 백인들과 비교했을 때에 유의하게 낮아졌으며, 그 원인이 임금 구조나 아랍인 자체의 특성으로 설명될 수 없음을 ACS의 개인 자료를 이용해 분석하였다.

나아가 지역별 집계 자료와 개인 자료를 모두 활용한 연구도 있다. Fuller-Thompson & Gadalla(2008)는 개인 자료와 지역 자료를 모두 활용하여 분석하였는데, 주(state) 수준의 임금 불평등이 가장 심한 세 주에서 개인 수준의 건강상의 일상생활 제약(ADL limitation)이 가장 높았음을 보였다.

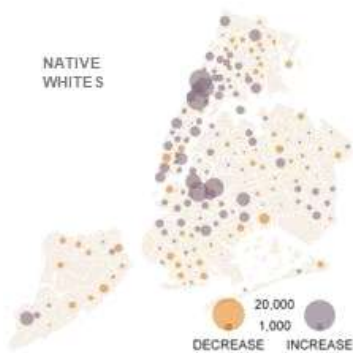
3. ACS 자료 홍보 및 정책적 활용

가. ACS 성과의 언론 보도

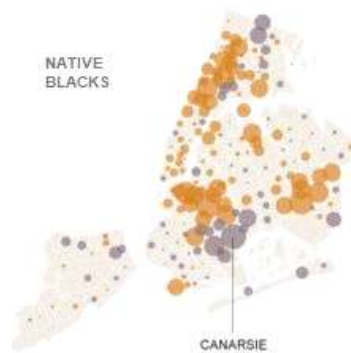
ACS의 통계 자료는 교육, 가족, 결혼, 인종, 이민, 노동 등의 전통적인 주제에서 컴퓨터와 인터넷 사용과 같이 최근에 주목하는 영역까지 다양한 분야에서 그 결과들을 3년에서 5년까지 누적해서 언론에 소개되는 경우가 많았다. 특히 뉴욕타임즈(New York Times)의 경우 지금까지 ACS를 인용한 200여 건의 기사에서 위의 주제들이 심도 있게 다루어지고 있다. 기사의 양과 질로 보았을 때 점점 그 중요성이 증가하고 있는데, ACS가 공식적으로 시작한 2005년부터 2010년까지는 47건의 기사가 나왔다면, 2011년부터 2015년까지 99건의 기사가 생산되었다. 또한 통계청(Census Bureau)의 공식 발표 자료나 외부 연구를 인용하는 것뿐만 아니라, 전문적인 필진들이 자체적으로 자료를 분석하여 심도 있는 분석을 발표하고 있다. Sam Roberts, Nate Cohen, Kirk Sample 등은 주기적으로 각자의 분야에서의 주목할 만한 분석 결과를 내 놓고 있으며, 내외부의 여러 필진들과의 공동 작업을 통해서도 양질의 기획기사 기사를 생산한다. 한국의 순환센서스를 보다 널리 활성화하기 위해서 통계청 내부의 분석진을 활성화하고 외부의 연구진들과의 협업도 전략적으로 활성화한다면 좋을 것이다. 아래 그림들은 이렇게 보도된 예들로서 대도시, 주(state), 전미 단위의 심층 분석을 시각화해서 제시하고 있다.

[그림 5-3] New York 시의 민족별 인구 변화 추이(Ford Fessenden & Sam Roberts, 2011)

Population Change in New York City Neighborhoods Since 2000



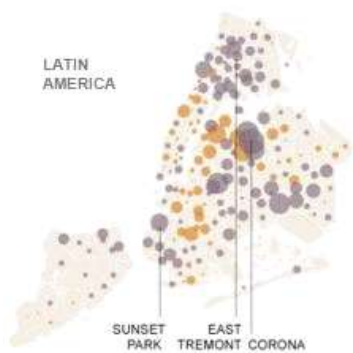
One of the principal demographic trends of recent years, U.S.-born whites moving into cities, is apparent in Manhattan and northern Brooklyn. White populations are declining in northeast Queens, where immigrant populations are increasing, and parts of Staten Island, which are attracting more blacks.



Black populations are declining in some traditionally black areas; in central Brooklyn and Harlem, where whites are moving in, but also in southeast Queens, which black families may be leaving for the suburbs. In Canarsie, the U.S.-born black population has grown 20 percent. In nearby Brownsville, it is down nine percent.



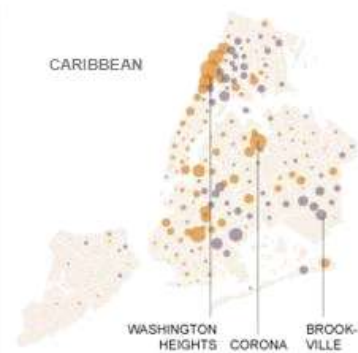
Chinese immigrants are declining in Lower Manhattan, and yielding to Indian and Bangladeshi growth in Jackson Heights and Woodside. But their numbers are rapidly growing in Flushing, Fresh Meadows, Sunset Park and Bensonhurst.



In what could be a classic ethnic replacement pattern, Mexican populations are growing in the Bronx and Corona, where native blacks are moving out, as well as in northern Manhattan neighborhoods, where older Dominican populations are declining.

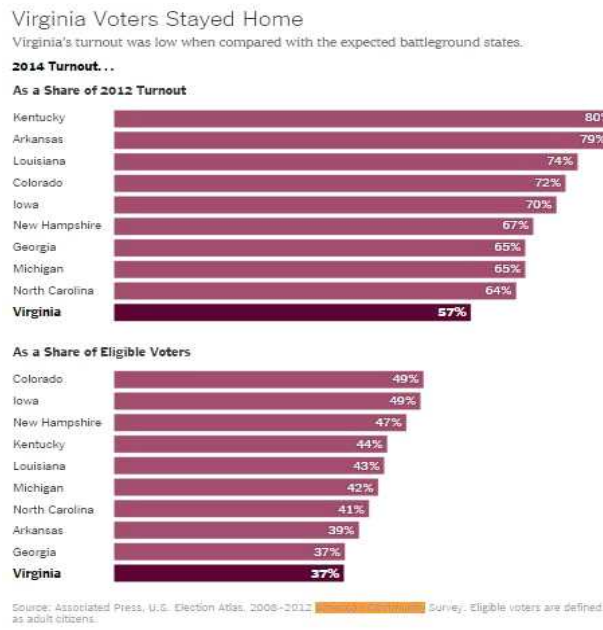


As Chinese move into Flushing and Queensboro Hill, Indians are moving out, and into Richmond Hill, Briarwood and Jamaica Hills. In Astoria, Pakistani and Bangladeshi populations are declining along with other immigrant groups.

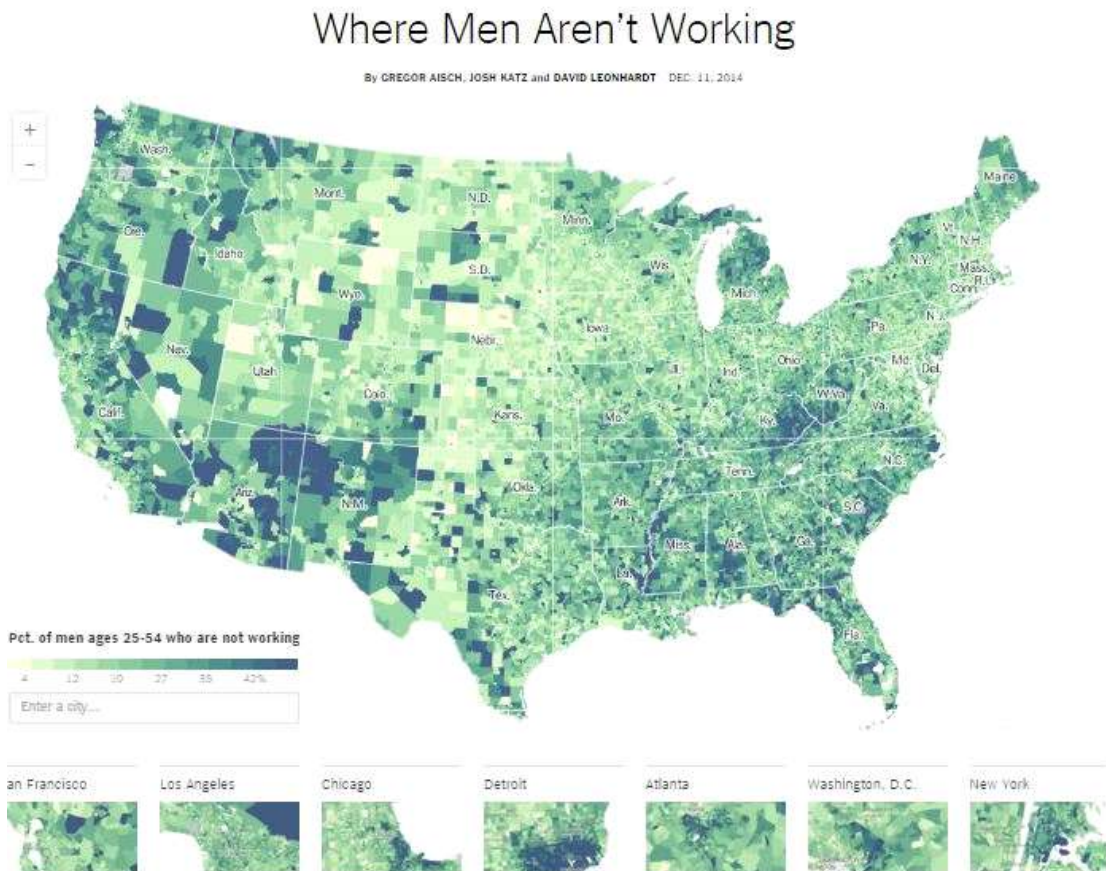


Whites are supplanting Dominicans in northern Manhattan neighborhoods. In parts of southeast Queens, Jamaicans, Haitians and Guyanese are moving in as U.S.-born black populations decline. In Corona, Caribbean immigrant populations have declined as Latin Americans have moved in.

[그림 5-4] Virginia의 민주당 성향 유권자들의 투표에 대한 분석(Cohn, 2014)



[그림 5-5] 남성 비노동자의 분포(Aisch, Katz, & Leonhardt, 2014)



이러한 ACS의 심층적인 언론보도에서 주목할 점은 다양한 주체 간의 연계활동이다. 통계청이 자료를 모으면 미네소타 대학이 자료를 가공하고, 뉴욕시립대학(CUNY)가 위에서 예를 든 시각적 기법 등을 포함한 분석을 하고 뉴욕타임즈 같은 언론이 보도를 하는 것이다. 이러한 연계활동은 한국 순환센서스의 언론보도에도 활용될 수 있기 때문에 본 보고서에서 좀 더 상술하려 한다.

미네소타 대학교(University of Minnesota)에서 학제적으로 운영하는 인구학 센터(Minnesota Population Center)의 IPUMS(Integrated Public Use Microdata Series) 프로젝트는 직접 센서스를 실행하고 자료를 축적해서 사용자들이 활용할 수 있도록 자료를 가공하고 제공하는 역할을 한다(<https://usa.ipums.org/usa/>). 홈페이지에 따르면 IPUMS-USA는 ACS를 포함한 총 16개의 전국(federal) 센서스 자료를 관할하고 있으며, 여러 센서스 자료들을 규격화(공통의 변수명 사용, 공통의 산업 및 직업 코드 사용 등)한다. 미네소타 대학의 이러한 표준적인 자료가공 활동은 미국연구재단(National Scientific Foundation)의 연구비를 수주하는 주요 경쟁력이다.

뉴욕시립대학(CUNY: City University of New York)의 도시연구센터(CUR: Center for Urban Research)는 ACS를 포함하여 이전의 센서스 자료들을 온라인을 분석할 수 있는 시스템을 구축하고 있다. 지속적으로 ACS를 이용해 대도시들, 특히 뉴욕 시의 지역적, 인구학적 변화들을 다각적으로 조명하는 분석을 발표한다(http://guides.newman.baruch.cuny.edu/nyc_data/nbhoods). 또한 지리 정보 시스템(GIS: Geography Information System)과 연동하여 센서스 자료를 분석하는 온라인 어플리케이션인 지도 서비스(Mapping Service)를 개발하여 누구나 센서스의 최소 트랙(census tract) 수준까지 다양한 자료를 시각화하여 이용할 수 있도록 하였다(Cork, 2013).⁹⁾ 이러한 CUR의 지도 서비스는 학문적인 활용뿐만 아니라 여러 정부, 비정부 기관, 기업 등의 사용자들이 각자의 목적에 맞게 활용할 수 있도록 하였다. New York Times에서 ACS를 활용한 기사들 중에는 CUNY의 시각화 자료를 적극적으로 인용할 만큼 이러한 서비스의 활용 가능성은 매우 높다고 할 수 있다(예: 「Then and Now - New York's Shifting Ethnic Mosaic」 by Ford Fessenden and Sam Roberts).

나. 정책 활용

미국의 연방 정부는 연간 총 \$40조의 예산을 ACS에 근거해 하위 기관이나 주 정부에 배분하며, 새로운 정책을 수립하거나 기존 정책을 보다 효율적, 효과적으로 집행하기 위해 ACS의 통계에 기반한다. 이렇게 집행된 주요한 정책들은 <표 5-6>과 같다. 한국의 순환센서스도 이처럼 정책 활용도를 높이기 위해서는 위에서 살펴

9) <http://www.longislandindexmaps.org/>에서는 Long Islands 지역에서의 ACS를 여러 지표들을 쉽게 시각화할 수 있다. 또한 특정한 주제에 대해 ACS와 다른 자료를 비교함으로써 시간에 따른 변화를 보여주기도 하는데, <http://www.urbanresearchmaps.org/plurality/>에서는 각 대도시별 2000년과 2010년의 자료의 인구학적 변화를 한눈에 알아보기 쉽도록 하였다.

본 것과 같이 자료 제공을 바탕으로 한 다양하고 심층적인 자료 분석이 선행되어야 할 것이며, 더불어 순환센서스의 언론보도 및 정책적 활용을 추적하여 순환센서스가 사회에 주는 영향력을 누적적으로 홍보할 수 있어야 할 것이다.

<표 5-6> 연방 정부와 주정부 및 하위기관의 ACS 정책적 활용 사례

기관	정책
연방 정부	<p>가족 관계, 소득, 직업, 주거: Community Services Block Grant Act 등에 의해 빈곤선을 결정하고 Temporary Assistance for Needy Families, Head Start Act에 따라 경제적, 교육적 지원이 필요한 가구의 특성을 파악하는 데에 활용</p> <p>연령, 성별, 이종, 민족성(ethnicity): 고용이나 주거 등에서 차별이 발생하는지 파악하고 Age Discrimination in Employment Act, The Rehabilitation Act, Equal Credit Opportunity Act 등의 정책적인 affirmative action을 통해서 격차를 지속적으로 줄이기 위해 노력</p> <p>가정에서 사용하는 언어: Voting Rights Act에 따라 투표하는 데에 언어 보조자(language assistance)의 수요를 지역별로 파악하여 예산 배정</p> <p>건강 상태, 1년 전의 주거지: AIDS나 결핵 등의 전염병 등의 확산을 측정</p> <p>장애 및 질병: Assistive Technology Act에 의해 장애를 가진 사람들을 지원할 수 있는 기술 개발을 지원, Carl D. Perkins Career and Technical Education, Elementary and Secondary Education Act 등을 통해 장애인의 취업 교육과 관련된 예산 배정에 활용</p> <p>군 전역자(veteran) 관련 문항: 군 전역자들의 건강, 고용, 주거를 보장하기 위한 정책들의 효과를 평가하고 부족한 지원의 수요를 파악</p> <p>컴퓨터 및 인터넷 사용 관련: Broadband Data Improvement Act과 연동하여 IT 관련 지역별 인프라를 파악하고 소외된 지역에 대한 망 구축 사업 예산에 반영</p> <p>난방 연료 및 난방비: Low Income Home Energy Assistance Program,</p> <p>푸드 스탬프: 정책 접근성을 파악하여 매 년 Program Access Index를 생산</p> <p>수도광열비: 항목별 Consumer Price Index를 생산</p> <p>배관, 주방, 통신 시설: National Center for Emerging and Zoonotic Infectious Diseases에 의한 전염병과 오염, 위험에 대한 노출 등의 현황을 파악하고, 보편적으로 해당 시설들이 보급되어 있는지 파악하는 데에 활용</p>
주 정부 및 하위 기관	<p>자동차 사용 및 출퇴근 수단 관련 문항: 도로와 대중교통의 지역별 수요 변화를 파악하고 Highway Safety Act, Moving Ahead for Progress in the 21st Century Act에 의해 카풀 등의 교통 관련 정보를 공유할 수 있도록 지원</p> <p>주거 형태 및 주거비: Community Development Block Grants, HOME Investment Partnership Program, Emergency Solutions Grant, Housing Opportunities for Persons with AIDS 등에 의거해 주택 수요가 높은 계층과 지역에 따른 보급 계획 수립</p> <p>건강 보험: State Children's Health Insurance Program 등에 의해 정부 지원 등으로 이루어지는 건강 보험의 사각지대를 파악하고 지원 예산 배정에 반영</p> <p>주거지 크기, 부동산 가격: 주별 GDP에서 부동산이 차지하는 비중 측정</p>

지금까지 논의된 ACS의 특성을 바탕으로 한국의 순환센서스에 적용할 만한 요소들을 정리하면 아래 <표 5-7>과 같다

<표 5-7> ACS의 특성 및 한국 순환센서스 적용 가능성 판별 조사표

	ACS 특성	정책적 적용 가능성	학술적 적용 가능성
조사 문항	현 거주지 특성 및 거주 기간, 모기지, 월세 액수	부동산/전·월세 정책 수립 및 평가 가구 및 증가에 대한 대책마련	1인가구, 공동 거주 등의 대안적 가족 형태에 대한 분석적 연구
	현 거주지에 대한 각종 세금 및 비용	가구의 조세부담 관련 정책 빈곤 가구에 대한 세제 혜택 등 지원 방안 마련	자산 규모 파악 및 소득을 넘어서 부의 불평등 연구
	고용 지위와 사업체 특성	비정규직 현황 및 정책 평가와 지원책 모색	정규직 가구와 비정규직 가구의 격차
	실업자 및 구직 방법	실업률, 고용률 등의 지표 생산, 시계열적 추적	실업자와 구직자의 유형 결정요인 연구
	인터넷 보급 및 연결 방식	정보 소수자(노인, 다문화) 문제 해결	스마트폰 시대 정보격차에 따른 사회적 불이익에 대한 분석
	국적 및 출신국	외국인 노동자, 다문화가정 현황 및 언어 사용 실태 파악	이주민의 정착과 2세들에 대한 연구
	결혼 상태/가족 구성원/조부모의 손자녀 돌봄	한부모가정 현황 파악 및 육아 관련 정책 수립	일-가정 양립과 육아에 대한 조부모 의존 현상
	장애와 건강 특성	지역별 장애편의 정책 수립	건강 및 활동 관련 연구
	교육 수준	학위 및 전공별 노동시장 진입 및 이직 실태 파악	지역 내, 지역 간 교육 불평등 연구
	복지 혜택 내역	복지 수요 파악 및 복지정책별 이용도 파악	빈곤 연구, 복지 정책의 효과 비교
	거주지 이동 및 통근	교통 수단(지하철, 도로 등) 정비를 위한 가수요 도출	지역별 젠트리피케이션 연구
조사 방법	우편/인터넷->전화->면접의 3단계 조사	지속적이고 체계적인 조사 방식 개선	각 조사 방법에 따른 응답의 질 차이점 분석
	매 월 대표성 있는 자료 수집	지역 단위별 주기적인 통계 생산	지역 규모에 따른 지역 단위 추정 신뢰성 연구
	가구 단위와 집단 거주 쿼터의 구분	기숙사 등의 시설 거주, 1인가족 등의 현황 파악	1인가구 증가요인과 결과에 대한 연구
자료 제공	American Fact Finder 웹을 통한 상세한 직업별, 산업별, 인종별, 지역별로 누적적 통계표 제공	집계자료를 제공할 수 있는 웹사이트 시스템 구축	기타 행정지역 단위 자료 혹은 기타 개인자료와의 병합을 통한 다층 분석
	개인자료 샘플 공개	개인자료 제공을 위한 심사 및 개인자료 접근 시스템 개선	전문적이고 참신한 시각에 근거한 연구 자료 제공
	인구에 대한 가장 폭넓고 정확한 자료 생산	순환 센서스와 등록 센서스 비교를 통한 상호 보완 모색	여타 서베이 자료의 타당성 검토를 위한 준거자료 활용
	심층분석의 주기적 언론 보도	통계청 수집, 미네소타대 가공, CUNY 분석, NYC 보도로 이어지는 연계활동	통계청 수집, KAIST 등에서 가공, 서울시립대 등 주요 거점 대학의 분석, 주요 일간지 보도로 이어지는 장기적 협력관계 형성
	통계청 내부 연구진 육성	각종 현안에 대한 시의성 있는 전문적 분석	비공개 자료를 이용하여 내부 분석자와 외부 연구진과의 공동작업을 통한 창의적 연구 환경 조성

제3절 프랑스 순환센서스(Rolling Census)

이 보고서의 3장에서 살펴본 것처럼, 프랑스의 순환센서스는 2차 세계대전 이후 예산 문제로 비정기적으로 실행되던 인구 총조사를 대체하기 위해 2004년부터 5년 단위로 실시되고 있다. 1997년부터 2008년까지의 준비 기간을 거쳐 매 해에 전체 인구의 1/7씩 5년간 전체 인구의 70%를 설문하도록 설계(Brown, 2013)되었고, 매년 €54,000,000 가량의 예산을 바탕으로 1월 15일부터 3월까지 조사를 실시한다.

총 36,682 개의 자치 구역(municipality)은 인구가 10,000명을 초과하는 980개의 자치구와 10,000명 이하인 나머지 자치구로 나누는데, 총 인구인 6,800만 명의 절반 정도씩 각각에 속한다. 10,000명 이상인 자치 구역은 주소의 층을 3단계(Communal Establishments-한 주소에 100개 이상의 거주지, Multi-Household Addresses-한 주소에 10개 이상의 거주지, Residential Addresses-한 주소에 10개 미만의 거주지)로 나누고, 각각을 5개의 균등한 순환 그룹으로 구성하여 10개 미만 거주지의 주소 그룹은 40%를 설문(전체의 8%), 그 이상은 전체 주소를 모두 조사한다. 또한 새롭게 생겨난 거주지는 당해에만 전체를 조사하고, 다음 해부터 기존의 층에 편입시켜 동일한 과정으로 설문을 진행한다.

10,000명 이하인 중소규모의 자치 구역은 지역, 연령, 성별, 거주 형태를 균등하게 대표하는 5개의 순환 그룹(rotation groups)으로 나뉘어 매 년 한 그룹씩 5년간 돌아가면서 조사한다. 코뮌(commune)이 선택되고 나면, 매년 1월-2월 사이에 각 주소별로 인구 전체를 선행으로 파악한다. 설문지는 우편을 통해 응답자에게 전달되고 최종적으로 응답을 확인하기 위한 전화(call-back)가 이루어지며, 인터넷으로도 응답이 가능하다.

설문 결과 생성되는 통계치는 INSEE의 웹사이트(www.recensement.insee.fr)를 통해 매년 1월에 공개되고 있으며, 원자료 또한 조사 시점에서 3년 후에 공개한다. 원자료는 가구별 자료와 개인별 자료를 나누어 공개하며, 가구 자료에서는 건물의 연식 등과 같은 거주지의 특성과 전기, 난방, 위생 시설 등의 내부 설비에 관해서, 개인 자료는 가족의 특성과 응답자 개인의 교육과 직업, 직장, 출신 지역 및 민족성(nationality)에 관한 내용 등을 포괄한다. 2012년의 원자료를 기준으로 가구 자료는 총 68개의 변수와 23,999,762개의 가구, 개인 자료는 도시 지역의 경우 89개 변수와 20,325,135명, 기타 지역의 경우 96개 변수와 20,325,135명의 자료를 공개하고 있다.

제4절 기타 조사의 활용 가능성 검토

순환센서스는 5년 주기로 총 인구의 20%만을 조사하는 대신 면접원이 방문조

[그림 5-7] [예시 문항] 사회 조사 2015년 현재 체감 환경

현재 체감 환경					
56 현재 살고 있는 지역의 환경은 어떻다고 생각하십니까? 다음 각 분야마다 해당되는 곳에 ○표를 하여 주십시오.					
	① 매우 좋다	② 약간 좋다	③ 보통이다	④ 약간 나쁘다	⑤ 매우 나쁘다
1. 대기					
2. 수질					
3. 토양(쓰레기, 폐기물로 인한 오염)					
4. 소음·진동					
5. 녹지 환경(공원, 가로수 등)					

다음으로는 사회적 활동에 대한 문항들인데, 우선 사회적 지지(social support) 문항과 시민 참여(civic engagement) 문항은 지역별 사회적 자본(social capital)을 측정하는 데에 유용하게 활용될 수 있으며, OECD 국가별 평가 항목에도 포함되어 있어 활용도가 높을 수 있다. 이러한 문항을 바탕으로 지표를 산출함으로써 각 지방 정부에서의 정책을 위한 기초 자료로서의 역할을 할 수 있을 것이다.

사회적 활동 문항들 중 나눔 혹은 기부 지수(giving index)로 활용할 수 있는 다양한 기부 및 자원봉사 활동을 격년 간격으로 묻고 있는 것이 사회조사의 고유한 특징이라고 할 수 있다. 관련 문항들은 통계개발원의 나눔실태 보고서로 꾸준히 분석되어 발표되고 있으며(박주연·이희길, 2013; 박주연·이희길, 2014), 사회복지공동모금회의 민간나눔자원 총량 추계 연구 등에서도 활용되었다. 이러한 사회적 활동에 대해 세부적으로 묻는 것은 다른 서베이에서 지금까지 상세히 다루지 않았기 때문에 사회조사만의 고유한 특징이라고 할 수 있다. 특히 나눔 지표는 시군구나 읍면동과 같은 지역 단위로 지표화하여 지역 공동체의 역량을 지표화할 수 있으며 World Giving Index, Giving UK, Giving Australia와 같이 국제적으로 비교가 가능한 변수들과의 호환을 통해 활용 가능성을 높일 수 있다. 사회조사의 기부와 자원봉사 관련 문항들은 비교적 자세하게 활동 횟수와 액수, 기관과 참여 경로 등을 묻고 있다.

[그림 5-10] [예시 문항] 사회 조사 2015년 기부의 정기성

기부의 정기성

27 귀하는 지난 1년 동안(2014. 5. 14. ~ 2015. 5. 13.) 주기적으로 기부하신 적이 있습니까?
 있다면 해당되는 현금 기부 또는 물품 기부의 주기를 하나 선택하여 ○표를 하여 주십시오.

※ 후원 경로가 여러 개일 경우, 가장 짧은 주기를 기준으로 응답하시기 바랍니다.

	주기적으로 기부한 적이 있다				⑤ 주기적으로 기부한 적이 없다
	① 월 단위	③ 분기(3개월)단위	③ 반기(6개월)단위	④ 기타()	
1. 현금					
2. 물품					

[그림 5-11] [예시 문항] 사회 조사 2015년 자원봉사 활동

자원봉사 활동

33 귀하는 지난 1년 동안(2014. 5. 14. ~ 2015. 5. 13.) 자원봉사 활동에 참여하신 적이 있습니까? 있다면 참여 횟수 및 1회 평균 활동 시간을 모두 써 주십시오.

※ 자원봉사 활동이란 개인 또는 단체가 지역사회·국가 및 인류사회를 위하여 대가 없이 자발적으로 시간과 노력을 제공하는 행위를 의미합니다.

① 있다

활동 분야	참여 횟수(연간)	1회 평균 시간
1. 아동, 청소년, 노인, 장애인, 재소자 등과 관련 (복지시설, 병원, 개인 가정 봉사, 방과후 교실 등)	□ 회	□ 시간
2. 환경 보전, 범죄 예방 등과 관련 (마을 청소, 방범 활동 등)	□ 회	□ 시간
3. 자녀 교육 등과 관련 (일일 교사, 교통 지도 등)	□ 회	□ 시간
4. 국가 및 지역 행사 등과 관련 (스포츠, 문화 행사 등에서의 안내, 통역 등)	□ 회	□ 시간
5. 재해 지역 주민 돕기 및 시설 복구 등과 관련 (해외 구호 포함)	□ 회	□ 시간
6. 기타 일반인을 위한 봉사 관련 (관공서 업무 지원, 무료 상담 등)	□ 회	□ 시간

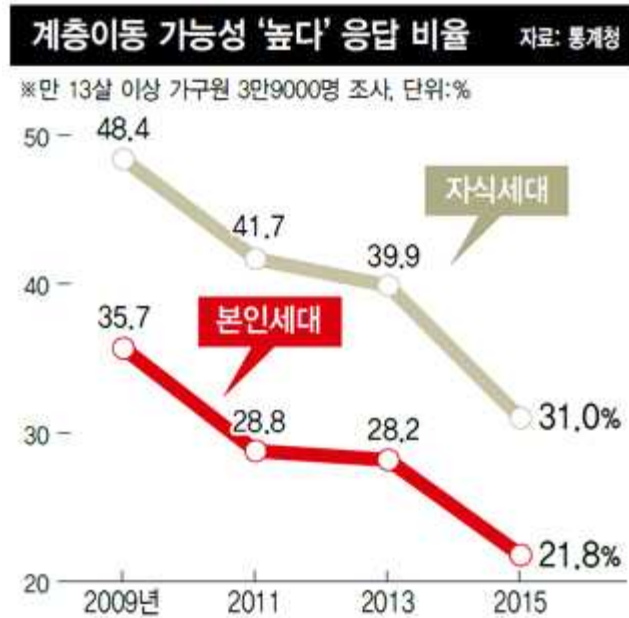
사회조사에서 순환센서스에 통합할 수 있는 세 번째 범주는 응답자의 태도나 상태에 관한 문항들이다. 주관적 계층 인식이나 계층 이동에 대한 전망을 묻는 문항들은 경제적 양극화, 경기 변동이 얼마나 체감되는지 알아볼 수 있는 문항으로 사회조사의 결과 보고서에서도 주요하게 인용되고 있다. 현재 한국 사회의 여러 지표들, 예를 들면 자살률이 OECD 국가 내에서 1위를 차지한다거나 성인의 학습 의지가 최하위권을 차지하는 등과 같은 현실은 사회의 안전장치로서 국민들의 희망 의식이 매우 낮음을 보여준다.

[그림 5-12] [예시 문항] 사회 조사 2015년 계층 의식 및 전망

계층의식	계층이동
38 귀하의 사회경제적 지위(소득, 직업, 교육, 재산 등을 고려)는 어디에 속한다고 생각하십니까?	39 우리 사회에서 노력한다면 개인의 사회경제적 지위가 높아질 가능성은 어느 정도라고 생각하십니까?
① 상 — [① 상 ② 하	① 매우 높다 ② 비교적 높다 ③ 비교적 낮다 ④ 매우 낮다 ⑤ 모르겠다
② 중 — [① 상 ② 하	39-1 우리 사회에서 현재 본인 세대 보다 자식 세대의 사회경제적 지위가 높아질 가능성은 어느 정도라고 생각하십니까?
③ 하 — [① 상 ② 하	① 매우 높다 ② 비교적 높다 ③ 비교적 낮다 ④ 매우 낮다 ⑤ 모르겠다

계층 의식이나 계층 이동 전망과 같은 문항들은 가장 중요한 차원의 희망을 나타내는 것이며, 언론에 기획기사로 보도되기에 알맞은 주제이다. 예를 들어 2015년 11월 27일 한겨레 신문은 사회조사 계층이동 문항을 분석하여 “국민 10명중 2명만 ”계층상승 가능“이라는 제목(부제: “2009년 36%서 올해 22%로 떨어져”)으로 1면에 기사를 게재했다. 이러한 기사는 2009년부터 격년으로 꾸준히 동일한 태도를 물었기에 가능한 기사이며, 향후 순환센서스에 통합하여 질문할 경우 그 누적적 가치가 지속될 수 있다. 그러나 이 기사는 ACS의 경우처럼 상세분석된 기획기사의 성격까지 다다르지는 못했으며, 향후 순환센서스에 통합되어 지역별 통계치까지 추정할 수 있다면 보도의 깊이와 질은 더욱 커질 것으로 생각한다. 순환센서스가 이러한 주관적 의식뿐만 아니라 부동산이나 금융 소득 등의 비근로 소득과 자산을 포괄할 경우 객관적인 사회적 위치와 결합하여 보다 풍부하고 참신한 해석을 가능하게 하는 변수로 활용될 수 있다.

[그림 5-13] 한겨레 기사에 실린 계층이동에 대한 태도 그래프



사회조사의 문항들 중에서 특정한 사회적 집단들에 관한 태도를 묻는 문항은 조사 시기의 사회적 현안에 따라 다양하게 대상을 달리하는 모듈로 시의성 있게 활용할 수 있다. 사회조사에 포함된 여성 취업에 대한 태도 문항들뿐만 아니라 다문화 가정이나 외국인 노동자, 자영업, 비정규직, 농어업 종사자에 대한 태도나 인지도 문항으로 확장되고 변형될 수 있다. 사회조사는 사회적 소수 집단에 대한 응답자들의 태도뿐만 아니라, 사회적으로 특정한 취약 집단에 속하는 이들의 현실적인 문제에 관한 문항도 포함한다. 아래 예시로 제시한 2015년의 노인들의 생활비와 대학생의 등록금 마련 관련 문항들은 응답자의 다양한 사회적 위치에 따라 변형되어 사용될 수 있고, 또한 조사 시점의 시의적인 현안에 맞게 모듈로 활용하는 것도 가능하다. 이를테면 다문화 가정의 2세의 사회적 적응이나 차별에 관한 문항, 외국인 노동자의 경제 활동에 관한 문항, 성적 소수자의 고용 등에 관한 문항 등은 취약 계층의 현실을 파악하고 이들에게 실질적으로 도움이 될 만한 정책을 개발하는 데에 유용할 것으로 보인다.

[그림 5-16] [예시 문항] 사회 조사 2015년 노인 생활비 수급원

생활비 마련 방법

20 귀하의 생활비는 주로 어떻게 마련하고 있습니까?

① 본인 및 배우자 부담 ② 자녀 또는 친척 지원 ③ 정부 및 사회단체 (정부 보조 등) ④ 기타 ()

① 근로소득 및 사업소득

② 재산소득(이자, 임대소득, 배당소득 등)

③ 연금(개인연금 포함), 퇴직급여

④ 예금(적금)

① 함께 살고 있음

② 함께 살고 있지 않음

[그림 5-17] [예시 문항] 사회 조사 2015년 대학생 등록금 수급원

대학생 등록금 마련 방법 ★

23-1 귀하의 2014년 1학기 대학(교) 총 등록금을 어떠한 방법으로 마련하셨습니까?
해당되는 곳 모두에 각각 기입하여 주십시오. 현재 휴학생은 최근 재학했던 학기를 기준으로 응답합니다.

1. 부모님(가족)의 도움				%
2. 대출(학자금 대출, 일반 대출, 마이너스 통장 등)				%
3. 스스로 벌어서 마련				%
4. 장학금				%
5. 기 타 ()				%
합 계	100%			

→ 26항으로

2. 지역별 고용조사 문항의 활용 가능성

지역별 고용조사는 2008년부터 실시되어 시군 단위로 고용 현황과 산업별, 지역별 특성을 조사하는 것을 목표로 하는 대규모 조사이며, 연간 조사, 분기 조사를 거쳐 현재는 반기로 조사되어 실업률이나 취업률, 비경제활동인구 등의 고용지표를 산출하는 데에 활용된다. 10,000여 개의 조사구에서 총 190,000 가구를 표집하여 30개 내외의 문항들로 구성된 간단한 설문문을 실시하기 때문에 비교적 작은 단위의 지역 범위인 시와 군 수준에서까지 통계를 제시할 수 있다.

순환센서스에 지역별 고용조사의 문항들을 확장에서 포함시킨다면 기존의 30여 개의 문항에 더해 고용 현안에 대한 보다 세밀하고 직접적인 주제를 다룰 수 있다. 예를 들어 사업체 내 복지 혜택, 승진이나 장기 근속 가능성 등에 대한 주관적인 기대, 직장생활에서의 어려움, 즉 과도한 근무 시간이나 승진에 대한 압박, 상사와

의 관계 등 일터에 대한 문항 등이 포함된다면 한국 사회의 고용 현안에 대한 주기적인 점검도 가능할 것이다. 물론 지금과 같은 문항 구성으로도 순환센서스에 통합되었을 경우 반기별 추정치 제시가 전국 단위뿐 아니라 시군구 수준에서도 가능할 것으로 보인다.

또한 구직자나 이직자, 퇴직자 등에 많은 맞춤형 문항을 제공한다면 구직을 희망하는 업종이나 준비 과정, 이직 사유와 이전 직장과의 주관적 비교, 퇴직자의 재취업 의사 등을 조사할 수 있다. 이에 따라 고용 불안정이 확대되어 가는 현실에 대한 연구와 정책 수립에 기여할 것으로 기대된다.

[그림 5-18] [예시 문항] 지역별 고용조사 2015년 상반기 근로 시간 및 사업체 정보

11 조사대상주간에 실제 몇 시간 일하셨습니다?

1 주된 일: 시간

2 다른 일: 시간

3 총 계: 시간

① 1~17시간 무급가족종사자 → 20 번으로
 ② 1~35시간 → 12 번으로
 ③ 36시간 이상

(※18시간 이상 무급가족종사자는 ② 또는 ③으로 표시)

12 조사대상주간에 어디에서 일하셨습니다?

● 사업체(직장)명 _____ **※ 조사관리자 기입**

● 사업체(직장)가 주로 하는 일 _____

● 사업체(직장) 소재지 _____ **※ 조사원 기입**

_____ (시, 도) _____ (시, 군, 구)

※ 조사대상주간의 직장(사업체)의 종사자수는 얼마나 됩니까?

1 1~4명 2 5~9명 3 10~29명 4 30~99명
 5 100~299명 6 300~499명 7 500명 이상

[그림 5-19] [예시 문항] 지역별 고용조사 2015년 상반기 근로 시간 및 사업체 정보

13 조사대상주간에 직장(사업체)에서 무슨 일을 하였습니다니까?

● 내가 한 일 _____

● 직명(직위) _____

● 부 서 명 _____

※ 조사관리자 기입

--	--	--	--

14 조사대상주간의 직장(일)에서의 자위는 무엇입니까?

● 임금근로자	● 비임금근로자
1 상용근로자	4 고용원이 있는 자영업자
2 임시근로자	5 고용원이 없는 자영업자
3 일용근로자	6 무급가족종사자

[그림 5-20] [예시 문항] 지역별 고용조사 2015년 상반기 구직자의 구직 경로와 퇴직자의 퇴직 사유

22 주로 어떤 방법과 경로로 직장(일)을 구해 보았습니까?
(주된 것 2가지까지 선택)

22-1 구직방법은 주로 무엇이었습니다() ()

1 시험 접수, 시험 응시	4 자영업 준비
2 구직 등록, 구직 응모	5 친척, 친구, 동료에게 소개 부탁
3 사업체에 문의, 방문, 원서제출	6 기 타 ()

22-2 주로 어떤 경로로 직장(일)을 알아보았습니까? () ()

1 공공 취업알선기관	4 학교, 학원
2 민간 취업알선기관	5 친척, 친구, 동료
3 대중매체(신문, 잡지, 광고 등)	6 기 타 ()

23 직장(일)을 구하기 위해 구직활동을 얼마 동안 계속 해왔습니까?

[] 개월

→ 번으로

24 이전 직장(일)을 그만둔 이유는 무엇입니까?

2 개인·가족적 이유

3 육 아

4 가 사

5 상산장애

6 정년퇴직·연로

7 작업여건(시간·보수 등) 불만족

8 직장의 후임·폐업

9 명예·초기퇴직, 정리해고

10 임시 또는 계절적 일의 완료

11 일거리가 없어서 또는 사업부진

12 기 타 ()

▶ **24-1** 개인·가족적 이유 중 구체적으로 어디에 해당됩니까?

1 결 혼

2 임신, 출산

3 자녀의 교육(초등학생)

4 가족돌봄(육아시외)

5 기 타 ()

▶ **28-1**번에서 "21년 이상 ⑩ 만 54세 이하"로 응답한 경우 질문 끝

3. 사회조사 및 지역별 고용조사 정리

사회조사와 고용별 고용조사 모두 조사 결과에서 추정된 비율(%) 인용은 활발하나, 자료를 적극적으로 분석한 연구는 많지 않다. 여타 서베이에서의 개인 수준 조사 자료와 병합할 수 있는 지역 변수들을 다양하게 생산함으로써 공동체에 관한 연구를 활성화하고, 또한 개인 자료(microdata)로서 한국사회의 현안들에 대한 신뢰도 높은 분석을 할 수 있다. 따라서 기존의 두 조사를 보다 확장하고 체계화하여 지역 특성 혹은 소수 인구집단에 대한 현황을 반영할 수 있는 다양한 조사 모듈을 개발하고, 암호화된 원자료 혹은 microdata 샘플을 공개함으로써 순환센서스가 보다 정책적 함의가 높은 분석적 연구를 이끌어내는 데에 중요한 기반으로 자리매김할 수 있을 것이다.

4. OECD 지표 모듈 가능성

OECD에서 발표하는 지표들은 다른 국가들과 비교했을 때 한국 사회의 여러 단면들을 볼 수 있는 기회를 제공한다¹⁰⁾. 주요 결과들은 공공성 순위 꼴지 (서울대학교 사회의 질 조사), 안전의식 30위, 자살률 1위, 노인빈곤률 1위, 아동 삶의 만족도 꼴지, 스마트폰 보급률 1위, 단기 근로자 비중 OECD 1위, 노동생산성 낮음, 노동시간 2위, 1인당 독서량 꼴지, 중장년 실질 문맹률 최고, 성인 학습의지 꼴지 (PIAAC 조사) 등으로 나타난다. 전국 수준 혹은 지역 수준에서 국제적으로 비교할 수 있는 지표를 생산하고 장기적으로 이를 개선하는 것은 정책적인 차원에서 중요할 것이며, 순환센서스를 활용해 조사해볼만한 항목을 아래 제시했다. 아래 문항들 외에도 자살률 대처를 위한 우울감이나 자살생각 문항, 낮은 노동생산성과 학습의지 등을 개선하기위한 직장만족도 요인 등에 대한 조사가 가능할 것이다.

사회자본 문항 조사

사회 전반적 신뢰, 대인 신뢰, 제도에 대한 신뢰 등과 같은 지표를 생산하는 것은 서베이 조사에서만 가능하며, 이러한 변수들은 국가간 비교에 중요하게 쓰이기 때문에 순환센서스에서 주기적으로 더 신뢰할 만한 지표를 생산하기 위한 모듈로 이용할 수 있다. 한국종합사회조사(Korean General Social Survey)에서 사용하는 일부 문항들은 전반적인 사회적 신뢰의 수준과 공적인 기관들에 대한 개별적인 신뢰의 정도를 측정한다.

[그림 5-21] [예시 문항] 사회전반적 신뢰(General Trust): KGSS의 예

귀하는 우리 사회가 어느 정도 믿을 수 있는 사회라고 생각하십니까? “매우 믿을 수 없다”에 0점을 주고, “매우 믿을 수 있다”에 10점을 준다면 몇 점을 주시겠습니까?

10)

http://live.media.daum.net/issue/OECD?post_id=686892541430499_687853271334426

[그림 5-22] [예시 문항] 제도에 대한 신뢰 (Institutional confidence): OECD는 군대, 사법부, 중앙정부에 대한 신뢰를 주로 고려, KGSS 문항의 예

다음은 우리나라의 주요 사회기관들입니다. 귀하는 이 기관들을 이끌어가는 사람들을 어느 정도 신뢰하는지 말씀해 주십시오.

보기 카드	매우 신뢰	다소 신뢰	거의 신뢰하지 않음	모르겠다
8) 중앙 정부 부처	①_	②_	③_	④_
10) 국회	①_	②_	③_	④_
11) 대법원	①_	②_	③_	④_
13) 군대	①_	②_	③_	④_

또한 사회적 지지망에 관한 OECD의 Better Life Index의 일부 문항들, 예를 들면 “귀하가 곤란한 상황에 있고 필요할 때마다 귀하를 도와줄 가족이나 친척, 친구 등이 있습니까?”와 같은 문항도 한국 사회의 사회자본을 측정하는 도구로 유용하게 활용될 수 있으면서 동시에 국제적인 비교도 가능해진다. 이러한 문항은 이미 사회조사에서 차용되고 있고, 순환센서스에서 충분히 포함되어 주기적으로 조사될 만한 문항이라 할 수 있다.

[그림 5-23] [예시 문항] OECD의 사회적 보호와 웰빙에 관한 비교 지표

Better Life Index - Edition 2015

Indicator	Housing		Income		Jobs			Personal earnings	Community		Education		Environment		Civic engagement		
	Dwellings without basic facilities	Housing expenditure	Rooms per person	Household net adjusted disposable income	Household net financial wealth	Employment rate	Job security		Long-term unemployment rate	Quality of support network	Educational attainment	Student skills	Years in education	Air pollution	Water quality	Consultation on rule-making	Voter turnout
Unit	Percentage	Percentage	Ratio	US Dollar	US Dollar	Percentage	Percentage	Percentage	US Dollar	Percentage	Percentage	Average score	Years	Micrograms per cubic metres	Percentage	Average score	Percentage
Australia	(E) 1.1	20	(E) 2.3	31 588	47 657	72	4.8	1.08	50 449	92	76	512	19.4	13	91	10.5	93
Austria	1	21	1.6	31 173	49 887	72	3.9	1.19	45 199	89	83	500	17	27	94	7.1	75
Belgium	2	21	2.2	28 307	83 876	62	5	3.88	48 062	94	72	509	18.9	21	87	4.5	89
Canada	0.2	21	2.5	29 365	67 913	72	6.4	0.9	46 911	92	89	522	17.2	15	91	10.5	61
Chile	9.4	18	1.2	(E) 14 533	17 733	62	(E) 4.4	(E) 1.59	(E) 22 101	86	57	436	16.5	46	73	2	49
Czech Republic	0.9	26	1.4	18 404	17 299	68	4.1	3.12	20 338	85	92	500	18.1	16	85	6.8	59
Denmark	0.9	24	1.9	26 491	44 488	73	5.6	1.78	48 347	95	78	498	19.4	15	94	7	88

본 예시 설계에서는 시군구 단위의 공개를 기준으로 삼았다.

<표 5-8> 순환센서스 조사 설계의 예

조사가구: 매달 최대 9만 가구, 연 80만 가구, 5년 총 4백만 가구 = 모집단 20%				
조사 개월	공통문항	모듈문항		
	내용 및 문항 수	내용	문항 수	시군구 단위 공개주기
1	* 내용: 현재 인구 주택총조사 기초 * 문항수: 약 45	고용조사 1차	20 ~ 25	2차 조사와 통합
2		사회조사(환경) 1차		2차 조사와 통합
3		사회조사(나눔) 1차		2차 조사와 통합
4		고용조사 2차		반기별
5		사회자본, 계층 1차		2차 조사와 통합
6		OECD		연도별
7		고용조사 3차		4차 조사와 통합
8		사회조사(환경) 2차		연도별
9		사회조사(나눔) 2차		연도별
10		고용조사 4차		반기별
11		사회자본, 계층 2차		연도별
12		변동모듈		연도별

모듈 항목들 중 고용조사는 동일한 항목을 3달 간격으로 4회까지 조사하며 두 회를 합하여 반기별(즉 6개월 단위)로 추정치를 공개한다. 현재 고용조사가 20만 가구 정도를 반기단위로 조사한 것을 고려하면 순환 센서스 2개월치, 약 18만 가구를 합하여 반기별로 추정하는 것이 합리적일 것으로 생각한다.

사회조사 항목들은 환경, 나눔, 사회자본 및 계층의 총 세 모듈로 나누어 6개월 간격으로 총 2회 조사하여 연도별로 추정치를 제공한다. 각 모듈에 포함될만한 문항들은 이미 앞에서 살펴보았다. 현재 사회조사 항목들이 격년으로 1만 8천 ~ 1만 9천 가구를 대상으로 13세 이상 구성원을 모두 조사하므로 응답자 수에서는 3만명 내외로 볼 수 있다. 순환센서스가 매년 2회 총 18만 가구를 대상으로 조사하는 것은 2년 동안 최대 36만 가구로서 조사대상자 수가 최소 10배 이상 증가하는 것을 뜻한다. 이처럼 조사대상자 수를 늘리게 되면 시군구 단위에서도 매년 추정치를 제공할 수 있을 것으로 생각한다.

OECD 모듈은 신뢰 및 삶의 질 문항들을 포함, OECD에서 국가간 비교를 위해 주로 사용할 수 있는 문항들로 조사한다. 예를 들어 자살률을 위한 우울증, 독서량 문항 등도 포함될 수 있을 것이다. 이러한 문항들은 지역단위보다 국가단위 추정이 더 유용할 것이므로 1년에 1회만 조사하는 것으로 예시 조사계획을 세웠다. 그러나

시군구 단위에서도 매년 추정이 가능할 것으로 생각하며, 최소한 2년 추정치는 시군구 단위에서도 안정적으로 제공할 수 있을 것이다.

변동모듈은 주로 연말에 실시하여 국가적으로 혹은 지역적으로 중요한 이슈나 수요를 반영할 수 있도록 한다. 국가적으로 중요한 현황을 전국적으로 파악해야 하거나, 지방자치정부가 실시한 정책을 연말에 평가하는 방식으로 지자체마다 문항을 구성하는 방식으로 운영될 수 있을 것이다. 혹은 타 부처나 연구자들로부터 공모나 예산지원을 받아 문항을 구성하는 등의 방안이 가능할 것이다. 국가적인 현황과 악의 경우는 향후 정규모듈에 편입시키려는 문항이나 모듈을 시범적으로 조사하는데 활용할 수 있을 것으로 보이며, 지자체의 정책 평가의 경우는 향후 순환센서스의 원활한 정착을 위해 지역행정의 협조를 얻는데 유용할 수 있다. 외부의 공모나 예산지원을 받는 경우는 학문적 발전과 센서스 예산 확보에 도움이 될 수 있을 것이다.

제6장 향후 순환센서스 연구를 위한 제언

인구주택총조사가 2015년부터 등록센서스로 대체되면서 통계청이 검토하고 있는 순환센서스(또는 순환표본조사)는 기존 ‘센서스 10% 표본조사’의 기능을 대신할 것으로 기대되고 있다. 만약 순환센서스가 표본조사의 기능을 대신할 뿐이라면 굳이 새로운 방식의 조사를 검토할 필요는 없을 것이다. 분명 순환센서스는 기존 표본조사가 한국사회에 제공해 온 기능 이상의 혜택들을 가져다 줄 것이다. 본 연구는 이러한 확신 하에 순환센서스를 실시하기 위해서 필요한 핵심적 사항들인, 순환센서스의 표본규모 할당, 표본설계 및 추정, 시험조사 자료 분석, 순환센서스 결과 활용 방안, 기타 주요 조사들과의 관계 및 통합가능성 타진 등의 내용들을 다루었다. 연구 결과, 순환센서스는 기존 표본조사의 기능을 충분히 대신할 뿐만 아니라 정보의 시의성, 추정의 정확성, 자료의 포괄성에 있어서 상당한 잠재력을 가지고 있음을 확인할 수 있었다.

우선, 제2장의 ‘인구주택총조사 표본규모할당 및 추정방법’은 크게 3개의 연구내용으로 구성되어 있다. 먼저, 제1절에서는 표본규모 할당방법에 대한 연구를 수행하였는데, 조사방법의 변화에 부응하기 위하여 정확도 높은 추정을 위한 표본설계의 기본방향을 제시하였다. 제2절에서는 추정가중치 산정을 위한 시험조사 자료를 분석하였다. 제3절은 가중치 산정 절차를 제시하였다. 인구주택총조사의 승수 산정 방법은 읍면동을 부차모집단(sub-population)으로 하고, 부차모집단을 조사구 특성에 따라 층화(아파트조사구 및 일반조사구)하여, 가구인구 및 주택부문으로 나누었다. 그런 후에, 가중치산정 방법들을 검토하여 새로운 조사방법을 도입하는 인구주택총조사의 추정에 있어서 정확도를 높일 수 있는 방법을 제안하였다.

제3장에서는 순환표본조사를 위한 표본설계 그리고 추정의 방안에 대한 내용을 다루었다. 순환표본조사를 위한 구체적인 표본설계 및 추정 방안에 대한 주요 이슈들을 정리하고 시험조사 결과에 적용하여 그 결과들을 살펴보았다. 제1절에서는 연구의 전체적인 연구내용을 개괄하고 있다. 그리고 제2절에서는 순환표본조사를 위한 표본설계 방안을 제시하였다. 제3절에서는 모수추정방안을 제시하였다. 먼저, 기간추정(Multi-year estimation or estimation for period) 개념에 대한 정의를 하고, 정의된 기간추정의 개념 하에서 모수 추정을 위한 가중치 작성내용을 우선적으로 다루었다. 추가적으로 추정 시 필요한 내용들도 검토하였다. 제4절에서는 기간 추정을 위한 기타 고려사항들을 제안하였다.

제4장은 ‘시험조사 자료의 심층 분석’을 실시하였다. 인천광역시 연수구와 경상남도 거창군의 시험조사 자료를 심층 분석함으로써 주요 항목별 변화 추이 및 시험조사에서의 무응답 현황 등 문제점을 살펴보았다.

제5장은 순환센서스의 인구사회학적, 정책적, 학문적 기능을 어떻게 극대화시킬

수 있는지 모색하였다. 이를 위해 우선 이러한 기능을 성공적으로 수행하고 있는 미국의 순환센서스, ACS부터 검토하였다. 이와 더불어 다양한 사회조사를 순환센서스에 통합시킬 경우 어떻게 통합조사할 수 있는지에 대한 예시를 제시하였다. 이를 위해 통계청의 ‘사회조사’와 ‘지역별 고용조사’, ‘한국종합사회조사’ 및 OECD의 각종 사회지표를 검토하였다.

본 장에서는 이제까지 검토한 순환센서스의 핵심적 내용들을 요약하고, 이로부터 도출된 순환센서스 시행 시 반드시 고려해야 하는 사항들을 정리한다. 그리고 순환센서스를 시행하기 위해서 필요한 향후 심층연구는 무엇인지에 대해 아래의 각절에서 논의한다.

제1절 인구주택총조사 표본규모 할당 및 추정 방법

표본규모 할당방법들을 이론적인 측면과 가상 자료로서 접근방법의 결과를 비교분석하여 살펴보았는데, 이론적인 측면에서도 방법들 각각의 다양한 장점 및 단점이 논의되었다. 표본규모 할당방법은 항목의 분산을 고려한 할당방법을 검토했는데, 이는 인구주택총조사는 조사항목이 많기 때문에 특정항목을 설계변수로 하여 그 분산에 초점을 맞추어 통계적인 표본규모를 할당하는 방식이다. 가상의 자료 분석에서는 상대적으로 규모가 작은 읍면동은 동등배분법이 효율적인 방법인 반면, 비례배분법은 효율성이 떨어지는 방법으로 판단된다. 또한 상대적으로 규모가 큰 동읍면은 이와 대조적인 결론이 나타났다. 이러한 규모의 양극화 현상을 중립적인 방향으로 조정하는 것이 필요해 보인다. 시군구의 경우에는 표본규모가 크기 때문에, 그 추정치를 일정한 오차의 범위 이내에서 산출하는데 큰 문제가 없을 것으로 보인다. 따라서 읍면동의 추정 정확도를 높이는 표본규모 할당 방법인 읍면동 직접 할당법(bottom-up method)이 설계의 기본 방향으로 고려되어야 할 것이다. 그러나 규모가 큰 동읍면에 초점을 맞출 것인가 또는 규모가 작은 동읍면에 초점을 맞출 것인가의 선택을 위해서는 보다 심도 있는 연구가 필요하다.

시험조사 가구명부 자료의 분석 가운데 항목 간 응답분포에서 빈집여부와 조사상태의 항목으로 도출된 가구의 응답상태는 무응답가구를 고려하여 표본추출승수를 조정할 수 있는 정보이다. 표본틀 작성을 위한 기초조사는 행정자료를 이용한 조사이므로 등록센서스와 기초조사의 결과는 밀접한 관계가 있을 것이라 가정 하에, 적격가구 비율을 표본으로 추정하고 등록센서스의 자료를 조정함으로써, 등록센서스를 벤치마킹 자료로 이용할 수 있는지 여부를 고려하였으나, 명확한 결론에 도달할 수 있는 근거를 시험조사 분석으로 도출하지 못하였다. 다만 면 지역에서는 표본으로 추정한 가구수가 등록센서스 가구수보다 낮게 집계되었을 것이며, 동 지역에서는 표본으로 추정한 가구수가 등록센서스 가구수보다 높게 집계되었을 것이다.

벤치마킹을 위한 등록센서스자료의 조정 여부는 좀더 신중한 연구가 요구되는 부분이다. 기초조사의 자료로 작성된 표본들의 변수들인 조사구, 가구수, 표본추출여부를 이용하여 표본 추출 승수를 작성하였다. 이 추출 승수를 실제로 조사에 사용된 가구명부에 부여하고, 이 가구명부로부터 응답상태변수를 도출하여 응답가구 및 무응답 가구수를 추정하였다. 이 과정을 거쳐 가구 무응답 조정가중치(조정된 표본가구 승수)를 도출하여, 이 승수를 실제조사에서 작성된 가구 인구 및 주택분야의 자료에 부여한 결과, 사후가중치 산정에 기초가중치 자료로 이용하도록 하는 것이 바람직한 가중치 추정절차로 판명되었다.

2010년 인구주택총조사 충청북도 자료를 이용하여 모의 실험한 결과는 항목별로 각각 다른 결과를 드러냈는데, 통합가중치 방법이 사후 가중치 방법보다는 더 효과적이다. 통합가중치 방법으로 산출된 가중치는, 때로는 특정 가구에 1보다 적은 가중치(음수 가중치 포함)가 할당되는 관계로(이는 표본이론으로 설명이 불가능함) 이를 사전에 방지 할 수 있는 정형화된 프로그램이 추가로 필요하다. 향후 통합가중치 방법을 사용하기 위해서는 정형화된 패키지를 도입하여 좀 더 심도 있는 연구를 제안한다.

제2절 순환표본조사 시행을 위한 표본설계 및 추정방안 연구

순환표본조사에 대한 원론적인 이해가 요구된다. 통계청에서 계획하고 명명하고 있는 순환센서스는 실제 센서스가 아닌 표본조사이다. 순환센서스를 고려하기 위해서는 프랑스와 마찬가지로 적어도 주소지 혹은 조사구 기반으로 전국 모든 지역을 포함해야 하며 기준 시점 역시 고정되어서 특정 시점 기준 한 국가의 인구, 가구 혹은 주택 분포의 스냅샷을 제공해야 한다. 순환표본조사를 수행할 경우, ACS 결과에서 드러난 바와 같이 센서스에서 제공하는 의미의 추정량을 생산할 수 없기 때문에 이 부분에 대한 사용자들의 이해가 우선 이루어져야 한다. 또한, 가능한 범위에서 순환표본조사로의 전환 시 설문문항에 대한 연구(questionnaire design)가 필요할 것으로 생각된다. 향후 순환센서스 기획 연구에서 가장 필요한 부분이 기존 센서스 표본조사에 포함된 항목별로 순환표본조사에서의 적합성을 따져보는 것이다.

앞서 살펴본 것처럼, 통계청은 순환센서스를 준비하면서 인천 연수구와 경남 거제군에서 순환표본조사 수행을 위한 시험조사를 수행했다. 현재에는 제주도에서 시험조사가 수행되고 있다. 그러나 연구진이 판단하기에 시험조사의 목적이 구체적으로 정해지지 않아 실제 연구를 통해 파악할 수 있는 부분들이 제한적으로 보인다. 이는 현재의 시험조사가 단순히 시험조사 지역에서 매 월 실사가 가능한지에 대한 검토를 위해 수행되었기 때문으로 판단된다. 조사 항목의 적절성, 표본분할 방안들의 비교, 다양한 추정 방안의 검토 등과 같은 구체적인 목적을 가지고서 시험조사의 설

계가 이루어지는 것이 바람직할 것이다. 따라서 향후 순환센서스 기획 연구는 순환 센서스 도입을 위한 시험조사의 목표를 재설정하고 그 목표를 달성할 수 있는 구체적인 방안과 내용에 대한 연구, 즉 시험조사 설계 연구가 되어야 할 것이다.¹¹⁾

모집단의 대표성을 유지할 수 있는 표본을 추출함과 동시에 매 공표시점 기준 실제 공표가 가능한 공표 수준의 결정을 위한 표본설계가 이루어져야 한다. 현재 통계청이 계획하고 있는 순환센서스는 프랑스의 순환센서스와는 다른 구조를 갖기 때문에 확률적 표본 설계와 비확률적 표본설계를 중복적으로 사용하는 것은 바람직하지 않다. 현재 가능한 방식으로는 읍면동 단위의 통계 생산 및 등록센서스의 보완을 목적으로 사전에 충분한 연구가 진행된 2015년 인구주택총조사의 표본조사를 위한 표본설계 방안을 이용하여 5년 주기로 조사가 진행될 표본을 추출하고 이를 60개월로 균등하게 분할하는 방안을 고려할 수 있을 것이다. 균등분할의 기준에 대한 연구는 추후에 추가적으로 이루어져야 하나 기본적으로 조사구 추출의 경우, 5년 동안 조사된 조사구를 추출하기 위하여 정렬된 조사구 리스트로부터 60개의 가능한 계통추출표본을 이용할 수 있다. 이 때 균형 표본의 정의를 위하여 각 조사구의 추가적인 정보를 고려한다면 정렬기준에 이를 추가한 후 서로 겹치지 않는 계통추출표본을 얻을 수 있을 것이다. 이에 대한 이론적 근거는 반복분산추출법 중 비복원 랜덤그룹화 방안에 있으며 공표수준의 결정을 위해서는 각 조사시점 별 읍면동 포함 비율 등의 내용들을 함께 검토해야 한다. 현 시험조사 결과를 이용한 표본설계의 타당성 연구는 제한적일 수밖에 없으며 2015년 인구주택총조사의 표본조사 결과를 이용한 분석이 추후 이루어져야 할 것으로 판단된다. 결론적으로 순환센서스 표본설계와 관련해 향후 순환센서스 기획연구에 포함되어야 하는 내용은 균등분할 기준에 대한 연구와 변수별 공표수준의 결정에 대한 연구가 되어야 할 것이다. 인구주택총조사 20% 표본조사를 분석하면 표본설계의 타당성 검토도 가능할 것이다.

본 연구에서는 순환표본조사 결과를 이용하여 산출되는 추정량의 대응 모수에 대한 정의를 하였으며 이에 근거하여 기간추정량의 개념을 정립하였다. 기간추정량은 특정시점 기준의 추정량으로 이해할 수 없으며 군이 모수의 정의를 통해서 판단한다면 기간의 중간 시점에 대한 추정량으로도 언급할 수 있으나 이 역시 모순이 있는 해석이며 따라서 공표 시 특정 기준 시점을 저하는 것은 이론적으로 타당하지 않다. 기간추정에 맞는 조사항목을 새롭게 구성할 필요가 있으며 인플레이션과 같이 타당한 기준이 있을 경우 이에 근거하여 조사된 값의 수정이 추정 단계보다 우선적으로 이루어져야 한다. 추정량의 계산은 다른 조사자료의 분석에서와 마찬가지로 가중치 산출을 통해 이루어지며 이 과정은 설계가중치, 무응답 조정 그리고 벤

11) 통계청에서 제공한 시험조사 자료에는 무응답 및 설문지의 선택 항목에 대응하지 않는 값들이 상당한 비율로 포함되어 있고 이러한 문제의 제기 및 자료의 수정 그리고 시험조사 설문지 요청을 지속적으로 요청했으나 이에 대한 처리가 이루어지지 않아 연구의 진행 상 그리고 그 결과의 활용에 제한적인 부분이 있다. 예를 들어, 시험조사를 위해 사용된 설문지는 2015년 11월 25일에 수령했으며 무응답과 같은 자료의 수정은 이루어지지 않아 오류로 판단되는 자료값을 그대로 사용하였다.

치마킹을 통한 최종 가중치 산출의 단계를 거쳐 이루어진다. 또한 본 연구는 시험조사 자료를 이용하여 각 가구원에 부여되는 설계가중치 및 사후층화를 통한 최종가중치를 산출하고 그 결과를 비교하였다. 시험조사의 결과를 볼 때 20%의 조사구가 조사된 12개월 통합자료를 사용할 경우, 적어도 시군구 통계작성은 가능한 것으로 판단되나 시험조사를 위한 표집틀은 평균 60가구로 구성된 조사구의 리스트로 조사구 크기가 일정하지 않고 조사구 내 가구추출로 인한 가중치의 변동이 있어 실제 30가구 기준의 조사구 추출이 이루어질 추 후 조사들에 대한 추론을 하는 것에는 한계가 있을 것으로 판단된다. 본 연구에서 고려된 사항들에 대한 실제 자료 기반 실증적 분석은 2015년 인구주택총조사 결과를 바탕으로 이루어져야 할 것이며 이 때 연구주제 역시 보다 구체적으로 제시될 필요가 있을 것이다. 조사문항의 문제점, 표본설계 특별히 표본분할의 효율성 그리고 모수 추정 및 이의 통계적 정확성에 기반 한 공표수준 결정 등의 내용이 추후 순환센서스 기획연구에서 다루어져야 할 것이다.

제3절 시험조사 자료의 심층 분석

제4장에서는 2012년~2014년에 걸쳐 시행된 인천광역시 연수구 및 경상남도 거창군의 시험조사 자료를 심층 분석함으로써 주요 항목별 변화 추이 및 시험조사에서의 무응답 현황 등 문제점을 파악하였다. 또한 시험조사 자료를 대상으로 표본비율(5%, 10%, 15%, 20%)별 주요 항목의 추정치 및 상대표준오차를 분석함으로써 시군구 및 읍면동 단위의 시기별 공표 가능한 수준을 파악하였다.

향후 추가적인 분석 및 연구를 위해 몇 가지 제언을 정리하면 다음과 같다. 첫째, 경상남도 거창군의 경우 3~4개의 인접 면들을 묶어 그룹을 설정하고 조사하였기 때문에 한 개 면에 대한 분석이 가능하지 않았다. 향후 분석 및 공표를 위해서는 적어도 읍면동 단위의 20% 표본 누적결과를 산출해 낼 수 있는 표본설계 혹은 가중치 부여 방안이 준비되어야 할 것이다. 둘째, 이미 시행된 거창군과 연수구의 시험조사와 현재 진행되고 있는 제주도의 시험조사 결과를 2015년 인구주택총조사의 등록센서스 및 표본조사 결과를 종합적으로 비교할 필요가 있다. 이러한 연구를 통해 순환센서스 결과의 정확도 수준, 공표 수준, 시기 등에 대한 결론을 도출할 수 있을 것이다. 셋째, 시험조사의 일부 항목에서 무응답이 상당히 존재하였고 많은 항목에서 적지 않은 무응답이 존재하였다. 향후 조사 단계에서 무응답을 줄이기 위한 방안 수립 및 노력이 필요하다. 또한 시험조사 자료에 있어서는 안 될 조사 오류 혹은 입력 오류로 보이는 값들이 존재하였다. 무응답 및 오류값은 자료의 신뢰도와 직결되는 문제이므로 자료 수집, 입력, 처리 단계에서 보다 체계적인 관리 방안이

수립되어야 한다. 이 또한 향후 순환센서스 기획 연구에서 다루어져야 할 내용이다.

제4절 한국에서의 순환센서스 기능: 공동체 연구 활성화를 위한 순환센서스 활용

순환센서스의 표집 계획에 따르면 최소한 매년 260여 개의 시군구 단위에서 지역 변수를 생성하는 것이 가능하다. 표집계획에 따르면 매달 읍면동 단위에서 층화 추출을 하므로 읍면동 단위에서도 매년 추정치를 제공하는 것이 원론적으로는 가능하다. 그러나 3500여 개의 모든 읍면동 단위에서 이러한 추정치를 제공하는 것이 신뢰성이 낮을 수 있음을 2장과 3장에서 살펴보았다. CV값 기준으로 군이나 구의 인구수가 많을수록 더 작은 CV값이 필요한데, 서울은 10%이하, 광역시의 구는 12%이하, 군은 14%이하, 도의 구/시는 14%이하, 군은 17%이하 등의 기준을 제시했다. 대략 CV값 15%정도 수준을 만족시킬 수 있는 시군구 단위에서 지역수준 추정치가 제공되어야 할 것이다. 그러나 주의할 점은 지역수준 추정치를 통계청이 공표하는 것과 지역수준 원자료를 제공하는 것은 다른 문제이다. 예를 들어 통계청이 큰 CV값 때문에 공표할 수 없는 수준의 자료(예: 1년 단위 읍면동 정보)도 응답자나 지역이 제대로 암호화된다면, 연구자의 책임과 재량에 의해 분석을 할 수 있도록 접근이 가능해야 할 것이다. 공표하는 추정치 이외의 자료에 대해서 통계청의 책임소재를 명확히 선 긋기 한다면, 이러한 자료 제공은 향후 CV 값을 낮출 수 있는 연구를 활성화하고 자료의 질을 높이는데 도움이 될 것으로 기대해볼 수 있다.

지역단위 각종 지표들을 제공할 수 있는 읍면동의 크기 및 변수를 선별하는 작업은 순환센서스 실시 이후에도 꾸준히 이루어져야 한다. 연구구와 거창군의 시험조사를 통해 판단해보자면, 조사구 수가 적인 일부 읍면동을 통합한다는 전제하여 5년간의 누적 추정치는 각종 변수들에 대해 비교적 안정적으로 제공할 수 있을 것으로 보인다. 좀 더 정확한 판단을 위해서는 2015년 20% 인구주택총조사 자료가 가용한대로 추가분석을 통해 보완해야 할 것이다. 어떤 변수들을 지역단위에서 제공할지, 개인 자료(microdata) 샘플 제공을 위한 암호화는 어떻게 할지 등의 문제는 순환 센서스 조사 항목들이 결정된 후 검토하고 결정할 사항이다. 특히 시군구 단위 보다는 읍면동 단위에서 보다 사회적으로 민감할 수 있는 변수들, 예를 들면 지역별 소득 지니 계수나 빈곤율, 실업률, 자살률 등을 공개할지 여부를 면밀히 검토할 필요가 있다. 지역단위 지표들은 국가통계포털(kosis.kr)을 통해 제공하면 무난할 것으로 보이나, 현재 국가통계포털의 지역통계는 시도별 수준에서 주로 제공하고 있다. 변수에 따라 시군구 단위나 읍면동 단위 제공이 필요하다. 심층분석을 위해서는 시군구단위까지 지역정보가 필요한 경우가 많으며 한국사회과학자료원은 이러한 연구수요를 반영하여 시군구 단위의 인구말도부터 상하수도 보급률, 노인 자살률까지

격년 간격으로 제공하고 있다. 이러한 자료를 활용하여 시군구 단위 연구가 최근 활발한 만큼(이민아·강정한, 2014; 한보영·강정한, 2015) 순환센서스 자료축적은 활발한 지역단위 연구를 촉발할 수 있을 것으로 본다.

더불어 개인 자료 샘플의 공개 방식에서 암호화와 공개 제한성 간의 균형 문제도 고려해야 한다. 가능하면 광범위한 암호화를 통해 좀 더 많은 범위에서 공개할 것인지, 아니면 최소한의 암호화를 한 후에 자세한 심사를 거쳐서 제한적으로 외부 연구자에게만 공개할 것인지 검토해야 한다. 두 방식을 제한 수준에 따라 혼용해서 사용할 수 있을텐데 현재 통계청의 마이크로데이터서비스시스템(MDSS, mdss.kostat.go.kr)을 통해 제공할 수 있다. 암호화에서 중요할 것으로 판단되는 것은 지역 정보인데, 응답가구의 지역정보를 통해 민감한 변수의 지역 추측이 가능해서는 안 될 것이다. 자료의 특성에 따라 공간적으로도 접근을 제한해야 한다면 MDSS 서비스 센터(통계청 본청(대전), 한국통계진흥원(성남), 한국개발연구원(KDI, 세종), 경인청 나라샘도서관(서울 강남), 서강대학교(서울 강북))에서만 접근할 수 있는 방안을 고려할 수 있을 것이다. 읍면동 단위에서 변수를 공개할 때, 민감한 변수들을 포함하되 지역 코드를 암호화하여 제공하는 것도 지역 간 비교 연구를 촉진시키기 위해 취할 수 있는 방안이다. 마치 microdata의 개인별 ID를 암호화하여 샘플로 제공하는 것과 같이, 3500여개 읍면동에 대한 상세한 변수를 암호화된 지역별 ID와 함께 심사를 통해서 외부 연구에 제공할 수 있다. 이 경우 개인 수준의 타 서베이 자료와의 변수 통합은 불가능하지만, 지역 수준의 정보를 풍부하게 활용할 수 있다는 장점을 가진다.

오늘날 영향력 있는 연구들은 서베이 자료 분석이나 실험실 실험(lab experiment)를 넘어서 현장 실험(field experiment)을 적극적으로 차용하여 현실의 삶을 개선시키려는 정책적 노력들을 활발히 하고 있다¹²⁾. 순환센서스에 의해 5년 주기로 지역의 변화를 파악하는 것이 가능하기 때문에, 260여 개의 시군구뿐만 아니라 3,500여개의 읍면동도 5년 단위로 다양한 지역 수준의 변수를 생성할 수 있다. 이를 정책 평가에 활용한다면 지역 정책이 조사와 조사 사이에 이루어질 때, 예시로 제시한 변동 모듈을 활용해 정책 시행 전과 후 차이를 파악할 수 있다. 혹은 외부 연구자들이 순환센서스 및 등록센서스의 행정자료를 중단적으로 활용해 정책 효과를 평가하거나 정책적 함의를 도출할 수도 있을 것이다. 즉, 본 장에서 제시한 예시 모듈들을 기반으로 필요에 따라 연구 시너지가 날 수 있는 소수 문항들을 유연하게 통합함으로써 자료의 활용가치를 높일 수 있을 것이다. 본 장에서 통합 가능한 예시로 제시한 사회조사나 지역별 고용조사의 경우도 분석적 연구를 위한 활용률은 높지 않다. 순환 센서스에 이러한 조사의 일부 주요 문항들이 포함된다 하더라도 이러한 용도의 활용률은 개선되지 않을 수도 있다. 따라서 각 모듈과 통합해서 조사할 수 있는 문항들을 공모(예를 들어 나눔 모듈과 함께 조사할 수 있는 문

12) 연구의 예로는 Banerjee et al. 2015, 그리고 대규모 연구 기관의 정책적 활동으로는 Poverty Action Lab(<http://www.povertyactionlab.org/>)을 들 수 있다.

항들 공모)해서 통합하고, 때로는 변동모듈 전체를 할애하여 조사할 수 있는 문항들의 집합을 공모하거나 설계하여 풍부한 사회조사 자료로서 순환센서스의 기능을 높일 수 있을 것이다.

공개한 자료 활용률을 높이기 위해서는 또한 순환센서스의 지역지표 및 개인 자료를 이용해 어떤 분석이 가능한지 예시가 잘 홍보되고 능력있는 외부 연구자들을 끌어들이는 필요가 있다. 이를 위해서는 통계청과 연구기관 및 대학 연구진이 잘 연계되어 언론에 심층 보도기사나 영향력있는 학술지에 논문을 게재하는 노력이 필요하다. ACS의 경우 살펴본 통계청과 미네소타 대학, 뉴욕시립대, 뉴욕 타임즈 간의 연계활동이 좋은 예이다. 또한 이러한 성과들이 정책적으로 활용된 영향력있는 사례들을 추적하고 잘 홍보하는 활동도 필요할 것이다. 이러한 방식으로 정책적, 학술적 활용 가능성이 홍보된다면, 창의적인 외부 연구진을 유도하고 성과를냄으로써 자료 활용도가 점점 높아지는 선순환을 이룰 수 있을 것이다. 특히 통계청 내부의 전문 연구진과 외부 연구자와의 협업으로 이러한 선순환을 강화시킬 수 있다면, 순환센서스가 지역공동체뿐 아니라 학문공동체에 미치는 순기능도 강화될 것이다. 이러한 선순환의 측면에서 ACS를 활용한 연구사례에서 주목할만한 연구주제는 경제적 불평등이다. ACS는 경제적 불평등에 관한 다양한 연구에 활용되어 왔고 순환센서스의 경우도 이러한 역할을 할 수 있을 것으로 본다. 그러나 이러한 연구를 위해서는 가구원의 소득에 관한 자세한 정보를 필요로 한다. 현재 인구주택총조사나 지역별 고용조사 등만으로는 조사가구 성인 가구원 각각에 대한 소득파악이 어려운데, 향후 소득조사의 정확성을 높일 수 있는 방안에 대한 고민과 더불어 상세한 소득조사를 한다면 유능한 연구자들을 유인할 수 있으리라 판단한다. 물론 이는 하나의 예시이며, 순환센서스가 연구자의 다양한 기획안을 선별하여 실험하는 장이 된다면 통계청, 지방 정부, 학계 간의 상호협조를 바탕으로 공동체의 발전에 기여할 수 있을 것이다.

참고문헌

- 박주연.이희길. 2014. 국내 나눔실태 2012-13. 통계개발원.
- 노병일 & 곽현근. (2005). 동네의 맥락적 특성이 주민의 정신건강에 미치는 영향. 보건과 사회과학, 17, 5-31.
- 박주연.이희길. 2013. 국내 나눔실태 2013. 통계개발원.
- 이민아.강정한. 2014. “한국 사회 자살률의 변동과 원인: 지역단위 지표를 이용한 패널 분석.” 『한국인구학』37(2): 1-19.
- 한보영.강정한. 2015. “한국사회의 고밀도 생활권과 이웃관계가 우울감에 미치는 영향: 한국종합사회조사(KGSS)와 시군구 단위 자료를 결합한 다수준 분석.” 『한국인구학』38(3): 75-111.
- Abhijit Banerjee, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, Christopher Udry. 2015. “A multifaceted program causes lasting progress for the very poor: Evidence from six countries” Science Vol. 348 no. 6236
- Aisch, G., J. Katz, & D. Leonhardt (2014) “Where Men Aren’t Working” New York Times
- Brown, J. J. and E. Beaujouan (2013). Review of the rolling census approach: and other survey-based options. 32.
- Census Bureau of the United States. 2014. American Community Survey (ACS) Design and Methodology
- Cohn, N. (2014) “What Happened in Virginia? Democrats Didn’t Show Up” New York Times
- Cork, D. L. (2013). Benefits, Burdens, and Prospects of the American Community Survey:: Summary of a Workshop, National Academies Press.
- Durr, J.-M. (2005). “The French new rolling census.” Statistical Journal of the United Nations Economic Commission for Europe 22(1): 3-12.
- Durr, J.-M. and F. Clanché(2013). “The French Rolling Census: a decade of experience ”.
- Fessenden F. and S. Roberts (2011) “Then and Now - New York's Shifting Ethnic Mosaic” New York Times
- Ha, Shang E., Soo Jin Cho, and Jeong-Han Kang*. 2015. “Group Cues and Public Opposition to Immigration: Evidence from a Survey Experiment in South Korea.” Journal of Ethnic and Migration Studies
- Kim, H. S. 2011. “Consequences of Parental Divorce for Child Development.” American Sociological Review 76: 487-511.