

본 연구는 통계청의 학술연구
용역에 의해 수행되었음.

순환센서스 표본설계 및 추정방법 연구

2016. 12.

한국조사연구학회

본 연구는 통계청의 학술연구
용역에 의해 수행되었음.

순환센서스 표본설계 및 추정방법 연구

2016. 12.

연구수행기관 한국조사연구학회

책임 연구원 : 김영원 (숙명여자대학교 통계학과 교수)

공동 연구원 : 이기재 (방송통신대학교 정보통계학과 교수)

박인호 (부경대학교 통계학과 부교수)

제 출 문

통계청장 귀하

본 보고서를 “순환센서스 표본설계 및 추정방법 연구
용역사업”의 최종보고서로 제출합니다.

2016. 12.

수행기관 : 한국조사연구학회

연구책임자 : 김영원 (숙명여자대학교 통계학과 교수)
공동연구원 : 이기재 (방송통신대학교 정보통계학과 교수)
박인호 (부경대학교 통계학과 부교수)
곽은선 (숙명여자대학교 통계학과 박사과정)
연구보조원 : 이수빈 (숙명여자대학교 통계학과 석사과정)
장현주 (한국방송통신대학교 정보통계학과 조교)
김수진 (부경대학교 통계학과 석사과정)

■ 차례 ■

I. 서론	1
1. 연구목적 및 필요성	1
1) 연구목적	1
2) 연구 필요성	2
2. 연구내용 및 범위	3
1) 순환표본조사 표본설계 및 추정방법	3
2) 순환표본조사와 등록센서스 연계분석 방안	5
II. 2015년 인구주택총조사 표본조사 분석	6
1. 현행 표본설계	6
1) 조사구 현황	6
2) 표본배분	8
3) 표본추출	10
2. 2015년도 인구주택총조사 표본조사 데이터 분석	11
1) 표본크기 현황	11
2) 주요 조사항목 추정에 대한 분석: 시군구 단위	14
3) 주요 조사항목 추정에 대한 분석: 동읍면 단위	23
3. 순환표본조사 표본설계에 대한 시사점	29
1) 시군구 단위 분석	29
2) 동읍면 단위 분석	29
III. 순환표본조사 표본설계 방안	30
1. 표본설계 관련 기존 연구 결과	30
1) 통계청(2015) 한국인구학회 연구	30
2) 미국 ACS 표본설계	31
3) 프랑스 순환센서스 표본설계	32
2. 새로운 조사설계 방안	33
1) 기본원칙 및 통계공표 수준	33
2) 조사설계 방안 검토	34

3. 새로운 표본설계 방안	43
1) 조사설계 방안의 비교	43
2) 표본설계 방안: 2%(또는 4%) 조사	45
3) 표본설계 방안: 10% 조사(방안 2)	46
IV. 모집단 분할	49
1. 모집단 분할	49
1) 모집단 분할 개요	49
2) 조사구 기준 모집단 분할 방법	51
2. 도 단위 분할방법 비교	53
1) 주택 관련 변수 기준 도 단위 분할방법 비교	53
2) 가구 관련 변수 기준 도 단위 분할방법 비교	57
3) 인구 부문 기준 분할방법 비교	60
4) 도 단위 분할방법 비교 결과 종합	64
3. 시군구 단위 분할방법 비교	66
1) 주택 관련 변수 기준 시군구 단위 분할방법 비교	67
2) 가구 관련 변수 기준 시군구 단위 분할방법 비교	73
3) 인구 관련 변수 기준 시군구단위 분할방법 비교	78
V. 순환표본조사 추정 및 등록센서스 연계	84
1. 통계청 순환표본조사 및 추정 개요	84
2. 인구·가구 통합가중치 연구	86
1) 표본설계 및 설계가중치 산출	87
2) 칼리브레이션 추정	87
3) 일반화회귀추정	87
3. 소지역 추정을 포함한 효율적 추정방안	92
1) 소지역 추정을 위한 기본적 고려사항	92
2) 2차 제주도 시험조사에 기초한 소지역 추정 접근방식	93
4. 2차 시험조사 자료를 이용한 통합가중치 및 소지역 추정시산	97
1) 인구·가구 통합가중치 시산	97
2) 소지역 추정	117

5. 순환표본조사와 등록센서스 연계	123
1) 시점추정과 기간추정의 연계방안	123
2) 등록센서스와 인총 표본조사 자료 비교	123
참고문헌	125

■ 표 차례 ■

〈표 2-1〉 조사구 특성별 조사구 현황	7
〈표 2-2〉 조사구 규모별 동읍면 분포	8
〈표 2-3〉 시도별 표본규모	10
〈표 2-4〉 내재적 층화를 위한 분류지표	10
〈표 2-5〉 각 시군구별 동읍면, 표본조사구, 조사가구 수 현황	12
〈표 2-6〉 표본조사구 수의 시군구 및 동읍면별 분포 현황	14
〈표 2-7〉 각 시군구의 가족유형별 가구 분포 현황	15
〈표 2-8〉 각 시군구의 점유형태별 가구 분포 현황	16
〈표 2-9〉 각 시군구의 사용 방 수 분포 현황	18
〈표 2-10〉 각 시군구의 혼인상태 분포 현황	20
〈표 2-11〉 각 시군구의 경제활동상태 분포 현황	21
〈표 2-12〉 분석대상 동읍면의 표본 조사구, 응답가구, 응답자 현황	23
〈표 2-13〉 각 동읍면의 가족유형별 가구 분포 현황: 경기도	24
〈표 2-14〉 각 동읍면의 점유형태별 가구 분포 현황: 경기도	25
〈표 2-15〉 각 동읍면의 사용 방 수 분포 현황: 경기도	26
〈표 2-16〉 각 동읍면의 혼인상태 분포 현황: 경기도	27
〈표 2-17〉 각 동읍면의 경제활동상태 분포 현황: 경기도	28
〈표 3-1〉 각 방안별 주요 조사항목 추정에 대한 오차의 한계(예상)	36
〈표 3-2〉 시군구별 주요 조사항목 추정에 대한 오차의 한계(예상)	39
〈표 4-1〉 제주도와 대전광역시 등록센서스 조사구 매칭 현황	52
〈표 4-2〉 정렬방식에 따른 도단위 주택부분 추정오차 비교	55
〈표 4-3〉 정렬방식에 따른 도단위 가구부분 추정오차 비교	58
〈표 4-4〉 정렬방식에 따른 도단위 인구부분 추정오차 비교	61
〈표 4-5〉 정렬방식에 따른 시군구 단위 주택수 추정오차 비교	68
〈표 4-6〉 정렬방식에 따른 시군구단위 단독주택수 추정오차 비교	71
〈표 4-7〉 정렬방식에 따른 시군구단위 아파트수 추정오차 비교	72
〈표 4-8〉 정렬방식에 따른 시군구단위 가구수 추정오차 비교	74
〈표 4-9〉 정렬방식에 따른 시군구단위 가구규모별 가구수 추정오차 비교	77

<표 4-10> 정렬방식에 따른 시군구단위 인구수 추정오차 비교	79
<표 4-11> 정렬방식에 따른 시군구단위 10세 간격 인구 추정오차 비교	82
<표 4-12> 정렬방식에 따른 시군구단위 가구주 관계 추정오차 비교	83
<표 5-1> 제주도 조사구 모집단 현황	93
<표 5-2> 제주도 2차 시험조사 A·B 그룹 차수별 표본조사구 현황	94
<표 5-3> 제주도 2차 시험조사 A·B 그룹 일부 동읍면별 표본조사구 현황	94
<표 5-4> 2차 시험조사 월 표본규모 및 표본추출률	95
<표 5-5> 2차 시험조사 분기 표본규모 및 표본추출률	96
<표 5-6> 2차 시험조사 반기 표본규모 및 표본추출률	96
<표 5-7> 2차 시험조사 가구가중치 및 g-조정계수 분포	99
<표 5-8> 2차 시험조사 가구가중치 상관계수	100
<표 5-9> 2차 시험조사 인구가중치 및 g-조정계수 분포	102
<표 5-10> 2차 시험조사 인구가중치 상관계수	103
<표 5-11> 2차 시험조사 가구원수 총합추정	106
<표 5-12> 2차 시험조사 주택유형 총합추정	107
<표 5-13A> 2차 시험조사 남성·연령 총합추정	108
<표 5-13B> 2차 시험조사 여성·연령 총합추정	109
<표 5-14> 2차 시험조사 주거전용·영업겸용여부 총합추정	110
<표 5-15> 2차 시험조사 점유형태 총합추정	111
<표 5-16> 2차 시험조사 교육정도 총합추정	112
<표 5-17> 2차 시험조사 수학정도 총합추정	113
<표 5-18> 2차 시험조사 종교유무 총합추정	113
<표 5-19> 2차 시험조사 출생지 총합추정	114
<표 5-20> 2차 시험조사 1년전 거주지 총합추정	115
<표 5-21> 2차 시험조사 통근·통학여부 총합추정	116
<표 5-22> 2차 시험조사 가구원수 총합추정 분산비	118
<표 5-23> 2차 시험조사 주택유형 총합추정 분산비	118
<표 5-24> 2차 시험조사 주거전용·영업겸용여부 총합추정 분산비	118
<표 5-25> 2차 시험조사 성·연령 총합추정 분산비	119
<표 5-26> 2차 시험조사 점유형태 총합추정 분산비	119
<표 5-27> 2차 시험조사 교육정도 총합추정 분산비	120

〈표 5-28〉 2차 시험조사 수확정도 총합추정 분산비	120
〈표 5-29〉 2차 시험조사 종교유무 총합추정 분산비	121
〈표 5-30〉 2차 시험조사 출생지 총합추정 분산비	121
〈표 5-31〉 2차 시험조사 1년전 거주지 총합추정 분산비	121
〈표 5-32〉 2차 시험조사 통근·통학여부 총합추정 분산비	122

■ 그림 차례 ■

[그림 4-1] 정렬방식에 따른 5개 그룹별 도단위 주택부문 추정오차 산점도	56
[그림 4-2] 정렬방식에 따른 5개 그룹별 도단위 가구부문 상대오차 산점도	59
[그림 4-3] 정렬방식에 따른 5개 그룹별 도 단위 인구부문 추정오차 산점도	62
[그림 4-4] 제주도 정렬방식에 따른 평균상대오차 분석 결과	64
[그림 4-5] 대전광역시 정렬방식에 따른 평균상대오차 분석 결과	65
[그림 4-6] 정렬방식에 따른 5개 그룹별 시군구단위 주택수 추정오차 산점도	69
[그림 4-7] 정렬방식에 따른 시군구 주택수 평균상대오차 비교	70
[그림 4-8] 정렬방식에 따른 5개 그룹별 시군구단위 가구수 추정오차 산점도	75
[그림 4-9] 정렬방식에 따른 시군구 가구수 평균상대오차 비교	76
[그림 4-10] 정렬방식에 따른 5개 그룹별 시군구단위 인구수 추정오차 산점도	80
[그림 4-11] 정렬방식에 따른 시군구 인구수 평균상대오차 비교	81
[그림 5-1] 2차 시험조사 가구 설계가중치 및 GREG 가중치	100
[그림 5-2] 2차 시험조사 가구 설계가중치 및 g-조정계수	101
[그림 5-3] 2차 시험조사 인구 설계가중치 및 GREG 가중치	103
[그림 5-4] 2차 시험조사 인구 설계가중치 및 g-조정계수	104

I. 서론

1. 연구목적 및 필요성

1) 연구목적

본 연구는 통계청에서 순환표본조사(rolling sample survey)를 도입하기 위하여 중점적으로 검토해야 할 연간 표본배분 방안을 포함한 조사설계, 중복 조사를 배제하기 위한 모집단 분할 방법, 기간 추정을 포함한 순환표본조사 모수 추정과 등록센서스 연계 방안 등 3가지 사항에 대한 연구를 수행함을 목적으로 한다.

통계청에서는 2015년 등록센서스와 병행하여 20% 표본조사를 실시하였고, 앞으로는 매년 등록센서스와 병행하여 순환표본조사를 실시함으로써 등록센서스와 표본조사 자료를 기반으로 매년 주택, 인구, 가구 관련 주요 통계를 생산하는 것을 목표로 연구를 진행하여 왔다. 보다 효율적인 순환표본조사를 구현하기 위해 2014년과 2015년에 한국인구학회에서 수행한 『순환센서스 추정 및 분석 방안에 대한 연구』를 통하여 순환표본조사 도입에 필요한 기초적인 연구가 이루어졌지만, 주기별 통계 공표 수준에 따른 구체적인 조사설계를 비롯해 기간 추정과 등록센서스 연계 및 모집단 분할 등과 관련된 보다 심층적인 연구가 요구되고 있다.

이에 따라 이번 연구에서는 2014년, 2015년 연구 결과 및 관련 선행 연구들을 종합적으로 검토하고 보다 심화된 표본설계 모형 설정 및 추정 방안, 시점 추정과 기간 추정, 등록센서스 자료와 연계 등에 대한 심도 있는 연구를 수행하고자 한다.

2) 연구 필요성

2015년부터 인구주택총조사가 행정자료를 활용한 등록센서스 방식으로 전환됨에 따라 매년 인구, 가구 및 주택에 대한 지역별 총수가 산출되기 때문에 세부적인 항목에 대한 통계작성을 위해 기존의 인구주택총조사와 함께 시행되었던 표본조사 부문에서도 매년 또는 일정 주기로 광역시·도 및 시군구나 동읍면 등에 대한 지역통계를 산출하는 것이 바람직할 것이다.

순환표본조사를 도입해 매년 조사를 수행하는 경우 연간 표본조사에서 시군구 또는 동읍면 별 표본크기가 충분히 크기 않기 때문에 규모가 작은 시군구 또는 동읍면의 경우 2년 또는 3년 동안 조사된 자료를 누적해 통계를 작성하는 방안을 연구해야 하며, 표본이 중복 추출되는 것을 피할 수 있는 모집단 또는 표본추출틀 분할 방법도 연구해야 한다.

순환표본조사(rolling sample survey)는 기존 인구주택총조사와 병행하여 실시되던 표본조사를 일시에 수행하지 않고 일정기간, 일정주기로 나누어 순차적으로 표본을 조사하는 방식을 의미하며, 이를 통해 일정 주기를 갖고 시군구나 동읍면 등 소지역별 통계 생산이 가능할 수 있고, 동시에 조사원 관리 및 균등한 예산 배정 등에 있어서 유리한 방법이 될 수 있다.

특히 매년 작성이 가능할 것으로 예상되는 등록센서스 자료를 기반으로 한 기본적인 인구, 주택, 가구 통계와 함께 기존의 표본조사에서 작성되던 세부 항목별 통계를 시의성 있게 작성하기 위해서는 기존에 5년에 한번 수행하던 대규모 표본조사를 매년 균등하게 배분하여 조사하는 방식의 순환표본조사로 전환하는 것이 바람직할 것이다. 주어진 예산으로 적정 수준의 통계를 시의성 있게 생산할 수 있는 효율적인 순환표본조사를 도입하기 위해서는 연도별 작성 대상 통계 및 연간 표본 규모 등 다양한 주제에 관한 연구가 필요하다.

이런 점들을 고려해 통계청에서는 한국인구학회를 통해 2014년과 2015년에도 순환센서스 도입을 위한 기초연구를 수행하였다. 하지만 보다 구체적으로 조사주기, 표본 배분 등 표본설계 모형 설정 및 추정 방법을 개발하고, 특히 기간 추정값의 시점 추정값으로의 전환을 포함한 순환센서스와 등록센서스 연계 분석과 연간 표본의 중복을 피할 수 있는 효과적인 모집단 또는 표본추출틀 분할 방법 등에 대한 보다 심도 있는 연구가 수행될 필요가 있다. 본 연구에서는 순환표본조사 도입과 관련된 다양한 연구주제들 중 일부 내용을 다루고자 한다.

2. 연구내용 및 범위

1) 순환표본조사 표본설계 및 추정방법

본 연구에서는 우선 순환표본조사 관련 표본설계 모형 및 추정과 관련한 국내외 선행 연구 사례를 종합적으로 정리하고 시사점을 도출하고자 한다. 이를 위해 기존 국내외 순환센서스 혹은 순환표본조사의 사례들을 살펴보고 해당 조사들이 갖는 조사목적, 관련 예산 및 인프라 등 제반 여건에 따른 표본설계 모형 및 추정 방식의 선택 및 적용 결과 등을 검토하고자 한다.

대표적인 외국 사례로 프랑스 순환센서스의 경우에는 모집단 등록 자료의 부재, 센서스 관련 예산 압박, 지방자치정부의 지역 정보 요구 증가 등으로 인해 기존 부정기적으로 수행하던 센서스를 순환센서스 형태로 전환하였다. 프랑스의 경우 5년 주기의 순환센서스 조사를 위해 전국 가구를 균형을 이루도록 분할하여 5개 부차모집단을 구성하는 방식으로 조사를 수행하고 있다. 프랑스의 순환센서스는 기본적으로 5년 동안 모집단 전체(실제로는 인구의 70%)를 조사하는 것을 목표로 일정 시점에 모든 가구를 조사하는 전통적인 센서스 방식 대신 5등분 된 부차모집단을 매년 전수조사 하는 방식으로 5년에 걸쳐 한 주기의 센서스를 완료하는 것으로 볼 수 있다.

미국의 경우 미국지역사회조사(ACS: American Community Survey)를 통해 기존 센서스 틀은 유지하되 상세표본(long-form)에 대해서만 5년 주기의 순환조사로 전환하여 기존 10년의 긴 주기의 조사수행에서 오는 시의성 문제를 해결하고 있다. 미국지역사회조사의 경우 표본추출을 위해 주소파일(MAF: Mail Address File)을 행정 자료와 연계해 매년 갱신하고 있다. MAF를 기준으로 가구들은 균등하게 분할된 5개의 하위 추출틀(sub-frame) 중 하나에 속하며, 분할된 1개의 하위 추출틀에서 매년 전체 가구를 대표할 수 있는 표본 가구들을 추출하고 있다. 응답 부담을 고려해 한번 조사된 가구는 향후 5년간 조사대상에서 제외되도록 한다.

한편 질병관리본부에서 실시하는 국민건강영양조사는 우리나라의 전형적인 가구조사 형태를 취하며 3년 주기의 매년 조사를 수행하고 있으며, 최근 전국은 물론 광역 시도 수준의 추정이 가능하도록 순환표본조사로 전환하여 소지역 추정 가능성에 대한 연구를 진행하고 있다. 국민건강영양조사의 경우 연간 표본을 나누어 매월 지역

별로 순차적으로 조사함으로써 전문성을 갖는 조사원들이 지속적으로 조사에 참여하게 함으로써 조사의 품질을 제고하는 동시에 기간 추정 문제를 다루고 있다는 점에서 참고할 필요가 있다.

본 연구에서는 우선 이러한 국내외 선행 연구사례를 종합하여 우리나라 상황에 적합한 등록센서스 자료와 연계해 활용이 가능한 순환표본조사를 위한 효율적인 표본설계 모형 및 추정 방식의 선택에 있어서 참고할 시사점들을 도출하고자 한다.

아울러 순환표본조사를 위한 최적의 모집단 분할과 표본추출틀 유지 및 관리 방안을 검토하고자 한다. 순환표본조사를 위한 표본을 매년 안정적으로 추출하기 위해서는 적절한 표본추출틀을 구축하고 매년 등록센서스 자료를 기반으로 한 추출틀 업데이트가 작업이 필요하다.

예를 들어, 2015년 인구주택총조사 20% 표본조사에서는 표본추출틀로 조사구 리스트를 사용하였으며, 이는 2014년 가구주택기초조사와 2015년 조사구 정합성 확인 결과 등으로 설정한 거주 기반 정보를 활용한 것이다. 향후 기존의 조사구 명부 대신 등록센서스 정보를 바탕으로 한 등록센서스 자료를 기반으로 작성된 표본추출틀의 활용을 검토해 볼 필요도 있다.

표본추출틀 구축과 관련해 통계청(2015) 연구에서 검토했던 표본추출 후 표본을 기간별로 분할하는 방식 대신 모집단을 분할한 후 연간 또는 반기 표본을 추출하는 방안에 대한 장단점과 함께 도입 가능성을 검토해야 한다. 이를 위해 기존 조사구 정보에 등록센서스 정보를 연결하여 안정적인 표본추출틀을 구축하는 방안을 연구하고, 시간이 경과됨에 따라 발생하는 모집단 변동을 반영할 수 있는 표본추출틀 업데이트 및 관리 방안이 동시에 검토되어야 할 것이다. 장기적으로는 등록센서스의 가구 및 인구 자료를 이용해 실시간으로 조사구 정보를 업데이트 하는 시스템을 구축하는 것이 필요할 것이다.

또한 순환표본조사를 위한 조사설계 모형 중 최적의 조사주기, 기간별 표본배분 방안에 대한 타당성을 검토하고자 한다. 2014년과 2015년 한국인구학회 연구에서는 주로 월간 조사를 염두에 두고 연구가 진행된 것에 반해 이번 연구에서는 반기 또는 연간 조사를 수행하는 방안을 중점적으로 검토하고자 한다. 이와 관련해 순환표본조사에 포함된 문항별로 표집오차를 분석하고, 이를 토대로 시군구 및 동읍면 단위 통계작성 주기를 차별화하는 방안을 연구해 보고자 한다. 조사주기와 기간별 표본배분안의 결정은 앞서 논의한 표본추출틀의 결정, 지역단위 통계작성 문항 및 주기, 정도수준 평가, 조사업무 부담 등에 대한 종합적인 고려가 필요할 것으로 생각된다.

마지막으로 본 연구에서는 최적 순환표본조사 모형 하에서 등록센서스 자료를 벤

치마킹 하는 가구 또는 가구원 가중치 보정(calibration) 방법과 함께 효율적인 추정 방안을 검토하고자 한다. 이를 위해 최적의 조사주기, 기간별 표본 분할 및 배분 등에 따른 표본설계 모형에 적합한 가중치 산출을 포함한 효율적인 추정 방안을 연구한다. 또한 반기 또는 연간 조사를 수행하는 경우 문항별 상대표준오차 산출 방안과 함께 2년 또는 3년 자료를 누적하여 소지역통계를 작성하는 방안을 검토한다.

2) 순환표본조사와 등록센서스 연계분석 방안

순환표본조사 결과와 등록센서스 결과 연계 과정에서 발생하는 불일치 보정을 위한 가중치 조정 방안을 연구한다. 순환표본조사 결과와 등록센서스 결과의 불일치에 따른 문제를 해결할 수 있는 벤치마킹 가중치 조정 방안이 필요하며, 이 경우 가구 중심의 가중치 조정과 가구원 중심의 가중치 조정이 가능할 수 있으며, 궁극적으로 이들 두 가지 방법을 어떻게 병합해 사용할 것인지 연구가 필요하다.

기존 연구들을 참조하여 등록센서스에서 산출되는 보조정보를 기반으로 한 일반화회귀추정량(GREG)을 포함해 순환표본조사 결과와 등록센서스 결과를 연계하여 분석하는 방안을 검토하고자 한다. 또한 순환표본조사의 기간 추정값을 등록센서스의 시점 추정값으로의 전환 가능성에 대해 검토하고 구체적인 추정 방안을 제시하고자 한다.

기존의 인구주택총조사의 표본조사 부문을 순환표본조사 방식으로 전환함에 따라 동읍면이나 시군구의 경우 기존 한 시점의 특성만을 나타내는 시점 추정값(point-in-time estimate)와 다른 특정 기간의 평균적 특성을 나타내는 기간 추정값(period estimate)를 고려하게 되기 때문에 순환 표본조사를 통해 산출되는 기간 추정값과 매 5년 수행되는 등록센서스 결과를 연계해 활용하기 위해서는 기간 추정값을 시점 추정값으로 전환할 수 있는 연구가 필요하다. 관련 연구를 위해 기존 국내외 선행연구 결과를 살펴보는 동시에 순환표본조사와 등록센서스를 연계한 통계 작성을 위한 추정 방안을 개발하고 이론적인 타당성을 검토한다.

II. 2015년 인구주택총조사 표본조사 분석

1. 현행 표본설계

2015년 인구주택총조사는 행정자료를 활용한 등록센서스와 전체 인구주택총조사 조사구 중 20%를 추출하여 진행되는 표본조사로 실시되었다.

2015년 인구주택총조사의 표본조사에서 조사구는 2014년 가구주택기초조사 결과를 기초로 마련되었는데, 효율적인 모수 추정을 위하여 기존 60가구 내외의 조사구 크기를 30가구 내외로 줄였다. 가구주택기초조사는 인구주택총조사와 농업총조사의 조사구 설정을 위한 모집단 자료 제공을 목적으로 2014년 11월 1일을 조사기준시점으로 전국의 모든 내외국인과 이들이 살고 있는 거쳐(빈집 포함)를 조사대상으로 하였다.

2015년 인구주택총조사 표본조사는 기존 10% 표본조사에서 20% 표본조사로 표본크기를 크게 늘렸고, 이를 통해 시군구 수준에서 생산되는 통계의 추정 정확도를 높이고 동읍면 수준에서 주요 조사항목에 대한 공표를 추진하였다.

1) 조사구 현황

2015년 인구주택총조사 표본조사의 조사단위는 가구 및 개인이고, 표본설계에서 사용된 1차추출단위는 약 30가구를 기준으로 작성된 조사구(ED: Enumeration District)이다. 인구주택총조사의 조사구는 일반조사구와 특별조사구로 구분된다. 일반조사구는 주택의 종류에 따라 아파트조사구, 보통조사구, 섬조사구로 구분되고, 특

별조사구는 기숙시설이나 사회시설 조사구를 의미한다.

<표 2-1>은 조사구 특성별 조사구 현황이다. 조사구 특성별 조사구 분포를 살펴보면 총 693,510개 조사구 중 보통조사구가 52.1%, 아파트조사구가 46.0%, 섬조사구가 0.2%, 특별조사구가 1.7%이다. 아파트조사구(A), 보통조사구(1), 섬조사구(2)에 대해서만 표본조사를 실시하고, 나머지 특별조사구에 대해서는 전수조사를 실시하였다.

<표 2-1> 조사구 특성별 조사구 현황 (단위 : 개, %)

조사구 종류	조사구 특성	조사구 수	구성비	비 고
	전 체	693,510	100.0	
일반조사구	소 계	681,785	98.3	20% 표본조사
	아파트조사구(A)	318,944	46.0	
	보통조사구(1)	361,310	52.1	
	섬 조사구(2)	1,531	0.2	
특별조사구	기숙사회시설 조사구(3, 4)	11,725	1.7	전수조사

출처: 2015년 인구주택총조사 표본설계 결과, 통계청 표본과, 2015년

2015년 인구주택총조사 표본조사에서 고려한 통계공표의 최소 행정단위는 동읍면 구분이다. 동읍면별 통계작성 방안을 검토하기 위해서 동읍면 내 일반조사구 규모별 동읍면 수의 분포를 살펴보면 <표 2-2>와 같다.

우리나라는 3,492개의 동읍면으로 분할되어 있다. 이 중 5개 이하의 조사구가 있는 동읍면이 2곳, 6~20개 조사구가 있는 동읍면이 50곳, 21~50개 조사구가 있는 동읍면이 643곳 등으로 50개 이하의 조사구로 구성된 동읍면이 695개로 전체의 19.9%를 차지하고 있다. 반면 201~400개 조사구가 있는 동읍면이 1,031곳, 401개 이상의 조사구가 있는 동읍면도 388곳 등으로 200개 이상의 조사구로 구성된 동읍면이 1,419곳으로 전체의 40.6%이다.

현행 우리나라의 동읍면의 규모는 지역에 따라 대단히 큰 편차를 보이고 있다. 따라서 전체 동읍면을 일률적인 통계작성 단위로 고려한다면 규모가 작은 동읍면에 대해서는 전수조사를 해야 할 것이다.

<표 2-2> 조사구 규모별 동읍면 분포

(단위 : 개, %)

시도	전체 조사구	아파트 조사구	기타 조사구	동읍면 수	조사구 규모						
					~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

출처: 2015년 인구주택총조사 표본설계 결과, 통계청 표본과, 2015년

2) 표본배분

2015년 인구주택총조사 표본조사의 각 동읍면 및 시군구별 표본배분 방법은 Kish 배분, 제공근 배분, 역배분(power allocation) 등의 표본배분방법을 비교하여 결정하였다. 최종적으로 각 동읍면별 표본배분은 기본적으로 Kish 배분법을 사용하였으며, 조사구 수가 상대적으로 적은 동읍면은 조사구 추출률이 100%가 되는 상황이 발생하여 조사구 수가 5개 미만인 2개 동읍면은 전수조사하고, 나머지 동읍면 내 최대 조사구 추출률을 75%로 제한하여 조정하였다.

이후 각 동읍면별 표본배분 결과와 시군구별 예상 상대표준오차를 근거로 목표오차 달성 여부에 따라 이에 미달되는 시군구 지역에는 표본 조사구를 추가로 배분하였다. 이를 통해서 특·광역시외의 목표 상대표준오차는 12% 이하로, 도 지역의 시는

14% 이하, 군은 17% 이하가 되도록 추출률을 산정하였다.

2010년 인구주택총조사의 표본조사와 비교하여 실제 2015년 표본조사에서 추출되는 조사구 수 기준의 표본크기는 약 4배로 증가하였다. 이에 따라 시군구 통계의 예상 상대표준오차는 2010년 인구주택총조사 표본조사에 비해서 크게 감소하였다. 실제로 2010년 기준 상대표준오차가 25% 이상인 시군구가 61개로 전체 시군구의 24.3%였으나, 2015년 조사에서는 상대표준오차가 25% 이상인 시군구는 없을 것으로 예상되었다.

지역별 표본배분 후 ①외국인 수, ②1인 가구 수, ③대학생 수, ④다가구 주택, ⑤2010년 기준 인구주택총조사와 등록 자료의 가구 수 차이 등의 5개 변수의 크기와 비율에 따라 7,965개 조사구를 표본에 추가하였다. 이는 시군구 지역 이외의 동읍면별로 산출 예정인 통계의 안정성을 확보하기 위한 목적이며, 이를 통해 표본조사 결과의 활용목적에 부합하는 표본을 얻기 위함이다.

각 동읍면 내에서 조사구 특성(아파트, 보통+섬)별 표본배분은 비례배분법을 적용하였으며, 각 조사구 특성별 최소 1개 표본조사구를 배분하였다.

요약하면 2015년 조사에서 조사구 표본추출을 위하여 사용된 층화변수는 동읍면 구분과 조사구 특성(아파트, 보통+섬) 구분이다. 표본배분은 최대 추출률을 제한한 Kish 배분법을 적용하였고, 통계작성단위인 각 동읍면 및 시군구별 안정적인 통계생산이 가능하도록 추가적인 표본배분이 있었다.

각 시도별 표본 조사구 수는 <표 2-3>과 같다. 최종적으로 총 표본조사구 수는 134,529개로 전체 조사구 수 대비 추출률은 19.7%이다. 2010년 인구주택총조사와 비교하여 표본 조사구 수 기준으로 약 4.3배 증가하였다. 시도별 추출률은 서울이 15.9%로 가장 낮고, 전남이 31.4%로 가장 높다. 표본 조사구는 경기도가 25,989개로 가장 많고, 세종시가 814개로 가장 적다.

<표 2-3> 시도별 표본규모

(단위 : 개, %)

시도	동읍면	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

출처: 2015년 인구주택총조사 표본설계 결과, 통계청 표본과, 2015년

3) 표본추출

이 조사에서 1차추출단위는 조사구이고, 조사단위인 2차추출단위는 가구 및 가구원이다. 각 층에서 표본 조사구는 계통추출법을 적용하여 추출하였고, 조사구 추출 후 조사구 내 모든 가구가 조사하였다. 2015년 인구주택총조사 표본조사의 표본추출법은 층화집락계통추출법이라고 할 수 있다.

각 층 내에서 계통추출법에 따라 표본조사구를 추출할 때는 조사구를 <표 2-4>에 제시된 정렬변수를 기준으로 정렬하여 내재적 층화의 효과를 얻고자 하였다.

<표 2-4> 내재적 층화를 위한 분류지표

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

2. 2015년도 인구주택총조사 표본조사 데이터 분석

통계청은 2021년부터 인구주택총조사에 대한 표본조사를 5년에 한 번 시행하는 것이 아닌, 순환주기가 5년인 순환표본조사를 계획하고 있다. 특정 연도에서 조사방식은 매월, 분기, 반기, 연 1회 조사 등 다양한 방식에 대해서 조사의 현실적 측면과 장단점을 고려하여 최종 방식을 결정할 예정이다.

5년 주기의 순환표본조사를 위한 조사구 표본을 추출하는 경우, 층화, 표본크기 결정, 표본배분 등을 포함한 표본설계의 전반적인 절차는 현행 2015년 인구주택총조사의 표본조사를 위한 표본설계 방안을 따르는 것이 바람직할 것이다. 이는 현행 표본설계의 목적이 비용과 예산의 문제를 제외한다면 순환표본조사의 목적과 부합하기 때문이다 (통계청, 2015).

한국인구학회가 수행한 연구인 통계청(2015)에서는 2015년 인구주택총조사 결과를 분석하여 전체적인 통계 공표수준에 대한 결정하는 방안을 제안하였다. 본 절에서는 2015년 인구주택총조사 결과 중 경기도, 대전광역시, 제주도 등에 대한 조사 데이터를 분석하여 전체적인 통계공표 수준을 검토하고자 한다.

1) 표본크기 현황

2015년 인구주택총조사의 표본조사 중 경기도에 대한 결과를 분석한다. 본 분석에서는 조사구 특성코드(ED_PRPT_CD)가 1(보통조사구), 2(섬조사구), A(아파트조사구)인 경우¹⁾만을 분석한다. 조사구 특성코드가 3, 4인 경우는 인구주택총조사의 표본조사 대상이 아니다.

2015년 인구주택총조사 표본조사에서 대전광역시의 전체 표본 조사구는 3,411개이고, 조사 가구 수는 96,589호이다. 구별 표본 조사구 수의 최소값은 485개(대덕구)이고, 최대값은 1,020개(서구)이다.

제주도의 전체 표본 조사구는 1,833개이고, 조사 가구 수는 47,738호이다. 제주시의 표본 조사구 수는 1,313개이고, 서귀포시는 520개이다.

2015년 인구주택총조사 표본조사에서 경기도의 전체 표본 조사구는 25,901개이

1) 분석대상 표본조사구 중 조사 가구가 한 가구뿐인 17개 조사구는 입력오류로 간주하여 분석대상에서 제외하였음.

고, 조사 가구 수는 730,360호이다. 시군구 내 표본 조사구 수의 최소값은 192개 (연천군)이고, 최대값은 1,093개(남양주시)이다.

<표 2-5> 각 시군구별 동읍면, 표본조사구, 조사가구 수 현황

(1) 대전광역시

시군구	동읍면	표본조사구 수			조사가구 수		
		보통/섬	아파트	합계	보통/섬	아파트	합계
대전 동구	16	385	243	628	9,901	7,629	17,530
대전 중구	17	315	274	589	7,814	7,914	15,728
대전 서구	23	494	526	1,020	13,720	16,929	30,649
대전 유성구	11	271	418	689	6,743	12,298	19,041
대전 대덕구	12	256	229	485	6,345	7,296	13,641
전체	79	1,721	1,690	3,411	44,523	52,066	96,589

(2) 제주도

시군구	동읍면	표본조사구 수			조사가구 수		
		보통/섬	아파트	합계	보통/섬	아파트	합계
제주 제주시	26	953	360	1,313	25,132	9,234	34,366
제주 서귀포시	17	431	89	520	11,131	2,241	13,372
전체	43	1,384	449	1,833	36,263	11,475	47,738

(3) 경기도

시군구	동읍면	표본조사구 수			조사가구 수		
		보통/섬	아파트	합계	보통/섬	아파트	합계
수원 장안구	10	276	262	538	7,424	8,212	15,636
수원 권선구	11	345	316	661	9,277	9,054	18,331
수원 팔달구	10	340	159	499	9,173	4,657	13,830
수원 영통구	10	213	399	612	5,896	12,419	18,315
성남 수정구	16	577	108	685	15,192	3,062	18,254
성남 중원구	11	455	150	605	12,395	4,692	17,087
성남 분당구	22	296	593	889	7,447	18,776	26,223
의정부시	15	341	408	749	9,481	13,177	22,658
안양 만안구	14	293	215	508	7,938	6,356	14,294
안양 동안구	17	205	422	627	5,476	13,134	18,610
부천 원미구	20	449	401	850	12,928	11,847	24,775
부천 소사구	9	229	184	413	6,360	5,522	11,882
부천 오정구	7	277	118	395	7,600	3,344	10,944
광명시	18	303	338	641	8,218	10,839	19,057
평택시	22	571	456	1,027	13,379	13,530	26,909
동두천시	8	140	115	255	3,497	3,260	6,757
안산 상록구	13	590	201	791	16,013	6,073	22,086
안산 단원구	12	493	291	784	13,685	8,605	22,290
고양 덕양구	19	350	493	843	9,406	15,233	24,639
고양 일산동구	11	274	270	544	7,963	8,234	16,197
고양 일산서구	9	94	358	452	2,555	11,275	13,830
과천시	6	126	125	251	3,299	3,833	7,132

구리시	8	172	173	345	4,611	5,391	10,002
남양주시	16	414	679	1,093	10,411	21,090	31,501
오산시	6	245	334	579	6,079	10,172	16,251
시흥시	17	470	445	915	11,978	14,160	26,138
군포시	11	195	347	542	5,662	10,944	16,606
의왕시	6	124	219	343	3,498	6,963	10,461
하남시	12	212	160	372	5,454	4,884	10,338
용인 처인구	11	354	147	501	9,058	4,311	13,369
용인 기흥구	11	150	440	590	3,828	13,993	17,821
용인 수지구	9	92	420	512	2,718	13,860	16,578
파주시	17	359	458	817	8,807	13,743	22,550
이천시	14	312	166	478	7,777	5,046	12,823
안성시	15	354	186	540	8,748	5,783	14,531
김포시	13	231	421	652	5,786	12,781	18,567
화성시	24	531	560	1,091	12,955	16,845	29,800
광주시	10	479	187	666	13,165	5,564	18,729
양주시	11	194	248	442	4,710	7,353	12,063
포천시	14	426	128	554	9,886	3,657	13,543
여주군	12	360	99	459	7,931	2,738	10,669
연천군	10	164	28	192	4,363	703	5,066
가평군	6	191	43	234	4,148	1,098	5,246
양평군	12	323	42	365	6,809	1,163	7,972
전체	555	13,589	12,312	25,901	352,984	377,376	730,360

본 연구의 분석 대상 시도인 대전, 제주, 경기의 동읍면 내 표본 조사구 수에 따른 동읍면 수 현황은 <표 2-6>과 같다. 동읍면 내 표본 조사구가 20개 미만인 경우는 대전 11개, 제주도 1개, 경기도 9개 동읍면이다. 이들 동읍면에 대해서는 동읍면별 통계의 정확성이 떨어질 것으로 예상된다.

<표 2-6> 표본조사구 수의 시군구 및 동읍면별 분포 현황

구분 (표본 조사구)	대전광역시		제주도		경기도	
	동읍면 수	율%	동읍면 수	율%	동읍면 수	율%
2-9개	0	0.0%	0	0.0%	4	0.7%
10-19개	11	13.9%	1	2.3%	5	0.9%
20-29개	31	39.2%	15	34.9%	96	17.3%
30-39개	19	24.1%	13	30.2%	149	26.8%
40-49개	9	11.4%	7	16.3%	117	21.1%
50-59개	3	3.8%	2	4.7%	85	15.3%
60-69개	3	3.8%	0	0.0%	34	6.1%
70-79개	1	1.3%	1	2.3%	21	3.8%
80-89개	0	0.0%	1	2.3%	11	2.0%
90-99개	1	1.3%	0	0.0%	13	2.3%
100개 이상	1	1.3%	3	7.0%	20	3.6%
총합계	79	100.0%	43	100.0%	555	100.0%

2) 주요 조사항목 추정에 대한 분석: 시군구 단위

2015년 인구주택총조사의 20% 표본조사 데이터를 분석하여 전체적인 통계공표 수준을 검토하고자 한다. 시군구와 동읍면 구분별 주요 조사항목에 대한 추정값과 95% 신뢰수준 오차의 한계를 분석하여 통계공표 수준을 살펴본다. 본 연구에서 고려하는 주요 조사항목은 다음과 같다.

- 가구 부문(3): 가구유형, 점유형태, 사용 방 수(잠을 자는 방)
- 개인 부문(2): 혼인상태, 경제활동상태

앞서 제시한 주요 조사항목에 대한 추정은 응답항목별 모비율 추정에 해당한다. 모비율 추정에서 응답항목의 표본비율 추정값이 10% 미만인 경우는 추정의 상대표준오차가 대단히 크게 계산될 수 있어 추정의 정확도를 평가할 때 적합하지 않은 측면이 있으며, 이에 따라 오차의 한계를 살펴본 것이다.

(1) 가구 부문

<표 2-7>은 각 시군구의 가구유형별 분포와 오차의 한계 현황이다. 전체적으로 가족과 1인 가족 유형이 대부분을 차지하고 있으며, 95% 신뢰수준 오차의 한계는 전체적으로 2.5% 미만으로 대체로 안정적인 수준을 나타내고 있다.

<표 2-7> 각 시군구의 가족유형별 가구 분포 현황

(1) 대전광역시

시군구	각 응답항목별 비율 추정값					비율 추정에 대한 오차의 한계(95%)				
	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상
대전 동구	64.9	0.2	33.6	1.3	-	1.56	0.05	1.50	0.27	0.00
대전 중구	73.3	0.2	25.5	1.0	-	1.34	0.06	1.32	0.25	0.00
대전 서구	70.4	0.2	27.9	1.5	-	1.25	0.04	1.18	0.22	0.00
대전 유성구	68.7	0.1	29.3	1.9	-	1.56	0.05	1.51	0.33	0.00
대전 대덕구	71.6	0.1	27.1	1.3	-	1.99	0.04	1.92	0.29	0.00

(2) 제주도

시군구	각 응답항목별 비율 추정값					비율 추정에 대한 오차의 한계(95%)				
	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상
제주 제주시	71.1	0.2	26.2	2.5	-	0.96	0.04	0.86	0.28	0.00
제주 서귀포	70.2	0.2	26.8	2.8	0.0	1.46	0.07	1.27	0.58	0.05

(3) 경기도

시군구	각 응답항목별 비율 추정값					비율 추정에 대한 오차의 한계(95%)				
	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상
수원 장안구	74.8	0.2	23.8	1.2	-	1.32	0.06	1.29	0.26	0.00
수원 권선구	72.5	0.2	25.7	1.6	-	1.34	0.07	1.32	0.29	0.00
수원 팔달구	64.9	0.2	33.1	1.8	-	1.79	0.08	1.71	0.40	0.00
수원 영통구	72.0	0.2	26.3	1.5	-	1.52	0.06	1.47	0.28	0.00
성남 수정구	62.7	0.2	35.6	1.6	-	1.33	0.05	1.31	0.27	0.00
성남 중원구	69.2	0.2	29.1	1.5	-	1.47	0.06	1.38	0.40	0.00
성남 분당구	77.2	0.2	21.1	1.4	-	1.13	0.05	1.08	0.23	0.00
의정부시	75.3	0.2	23.4	1.1	-	1.38	0.06	1.30	0.22	0.00
안양 만안구	75.0	0.1	23.6	1.3	-	1.36	0.06	1.32	0.29	0.00
안양 동안구	80.8	0.2	18.0	1.1	-	1.17	0.06	1.09	0.23	0.00
부천 원미구	74.9	0.2	23.6	1.3	-	1.22	0.06	1.18	0.22	0.00
부천 소사구	75.9	0.2	23.0	1.0	-	1.58	0.07	1.52	0.25	0.00
부천 오정구	76.5	0.2	22.2	1.1	0.0	1.39	0.06	1.36	0.29	0.02
광명시	79.8	0.2	19.2	0.7	-	1.01	0.06	1.00	0.18	0.00
평택시	70.7	0.2	26.7	2.4	0.0	1.19	0.04	1.12	0.31	0.02
동두천시	72.1	0.2	26.5	1.2	-	1.88	0.08	1.84	0.39	0.00
안산 상록구	69.9	0.2	28.3	1.6	-	1.15	0.06	1.11	0.23	0.00
안산 단원구	70.3	0.2	27.8	1.7	-	1.36	0.05	1.29	0.31	0.00
고양 덕양구	76.3	0.2	22.1	1.4	0.0	1.13	0.05	1.09	0.22	0.00
고양 일산동	73.0	0.2	24.8	2.0	-	1.98	0.07	1.75	0.54	0.00
고양 일산서	82.4	0.3	16.4	1.0	0.0	1.28	0.07	1.23	0.24	0.01
과천시	79.6	0.3	19.1	1.0	-	1.50	0.12	1.51	0.29	0.00
구리시	77.4	0.2	21.4	1.0	-	1.54	0.08	1.47	0.30	0.00
남양주시	81.6	0.2	17.3	0.9	0.0	0.92	0.04	0.90	0.18	0.01

오산시	69.7	0.1	27.6	2.5	-	1.61	0.05	1.47	0.40	0.00
시흥시	70.8	0.2	26.3	2.7	0.0	1.25	0.05	1.20	0.31	0.01
군포시	79.3	0.2	19.5	1.0	-	1.40	0.06	1.35	0.22	0.00
의왕시	79.8	0.2	19.0	1.1	-	1.43	0.10	1.35	0.32	0.00
하남시	77.2	0.3	21.6	0.9	-	1.56	0.11	1.49	0.28	0.00
용인 처인구	72.1	0.3	25.1	2.4	0.0	1.86	0.09	1.67	0.47	0.02
용인 기흥구	79.9	0.2	18.8	1.1	-	1.36	0.06	1.35	0.22	0.00
용인 수지구	83.5	0.2	15.4	0.9	-	1.35	0.06	1.33	0.22	0.00
파주시	76.1	0.2	22.1	1.6	-	1.26	0.06	1.19	0.27	0.00
이천시	73.8	0.1	23.9	2.1	0.0	1.52	0.07	1.45	0.37	0.03
안성시	68.8	0.1	28.2	2.8	-	1.90	0.06	1.70	0.57	0.00
김포시	79.0	0.2	19.7	1.1	0.0	1.25	0.06	1.23	0.22	0.02
화성시	74.0	0.1	23.4	2.5	0.0	1.16	0.04	1.11	0.29	0.02
광주시	76.5	0.2	21.6	1.7	0.0	1.27	0.06	1.21	0.28	0.03
양주시	77.2	0.2	20.7	1.8	-	1.47	0.08	1.38	0.45	0.00
포천시	70.0	0.1	28.1	1.8	0.0	1.67	0.05	1.63	0.35	0.01
여주군	71.9	0.1	26.1	1.9	-	1.75	0.06	1.74	0.42	0.00
연천군	66.5	0.3	31.8	1.5	-	2.39	0.17	2.34	0.63	0.00
가평군	66.6	0.1	31.0	2.2	-	2.12	0.08	1.92	0.63	0.00
양평군	69.5	0.2	28.8	1.6	-	1.43	0.08	1.37	0.47	0.00

<표 2-8>은 각 시군구의 점유형태별 분포와 오차의 한계 현황이다. 각 시군구의 점유형태별 모비율의 추정에 대한 오차의 한계가 3%보다 큰 경우가 일부 있지만, 대체로 안정적인 추정이 가능한 것으로 판단된다.

<표 2-8> 각 시군구의 점유형태별 가구 분포 현황

(1) 대전광역시

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	자가	전세	월세1	월세2	사글세	무상	자가	전세	월세1	월세2	사글세	무상
대전 동구	51.0	12.8	27.6	3.9	1.9	2.8	2.13	0.83	2.00	0.48	0.50	0.44
대전 중구	61.8	13.5	19.1	2.8	0.5	2.3	1.86	0.88	1.63	0.54	0.12	0.26
대전 서구	52.8	16.7	26.0	1.7	0.4	2.4	1.41	0.70	1.42	0.19	0.10	0.31
대전 유성구	48.4	18.9	26.8	1.3	0.5	4.1	2.09	1.11	2.09	0.22	0.18	0.72
대전 대덕구	57.9	13.9	21.8	2.2	0.6	3.6	2.34	1.00	2.21	0.31	0.17	0.79

(2) 제주도

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	자가	전세	월세1	월세2	사글세	무상	자가	전세	월세1	월세2	사글세	무상
제주 제주시	56.5	4.0	17.5	3.5	11.6	6.9	1.32	0.45	1.24	0.31	0.58	0.43
제주 서귀포	63.0	1.7	11.4	3.0	10.7	10.2	2.02	0.30	1.85	0.46	0.98	0.93

(3) 경기도

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	자가	전세	월세1	월세2	사글세	무상	자가	전세	월세1	월세2	사글세	무상
수원 장안구	55.8	19.9	20.8	1.5	0.1	1.9	1.52	0.93	1.29	0.35	0.05	0.52
수원 권선구	49.3	19.8	26.7	1.8	0.3	2.2	1.99	1.02	1.97	0.39	0.09	0.54
수원 팔달구	43.0	22.0	30.0	2.7	0.3	2.0	1.90	1.15	1.99	0.58	0.13	0.27

수원 영통구	48.5	25.0	23.3	1.5	0.2	1.6	1.54	1.28	1.55	0.56	0.07	0.28
성남 수정구	29.8	34.0	28.9	3.2	0.4	3.7	1.25	1.14	1.31	0.71	0.13	0.31
성남 중원구	38.4	30.4	26.4	2.0	0.3	2.5	1.63	1.07	1.68	0.51	0.11	0.31
성남 분당구	46.8	23.8	26.5	0.8	0.1	2.1	1.51	0.92	1.67	0.39	0.04	0.52
의정부시	53.5	19.0	23.8	1.7	0.3	1.7	1.74	0.84	1.81	0.29	0.11	0.20
안양 만안구	53.3	20.4	20.8	2.5	0.3	2.6	1.71	1.08	1.35	0.67	0.13	0.56
안양 동안구	57.3	23.2	17.1	0.9	0.1	1.4	1.48	0.88	1.36	0.26	0.06	0.18
부천 원미구	54.4	19.7	22.8	1.6	0.2	1.2	1.50	0.95	1.36	0.59	0.07	0.15
부천 소사구	54.1	23.5	19.6	1.2	0.2	1.4	2.11	1.30	1.91	0.38	0.08	0.25
부천 오정구	56.7	20.1	20.0	1.3	0.3	1.7	2.01	1.04	1.71	0.24	0.12	0.31
광명시	53.6	22.9	20.0	1.1	0.2	2.2	1.64	0.89	1.66	0.22	0.09	0.37
평택시	53.9	16.0	22.0	2.7	0.4	5.0	1.71	0.80	1.66	0.35	0.11	0.73
동두천시	55.4	16.6	21.5	3.1	0.2	3.2	3.09	1.41	3.12	0.80	0.09	0.70
안산 상록구	44.1	20.1	31.2	2.2	0.4	2.0	1.53	0.92	1.43	0.49	0.22	0.36
안산 단원구	49.0	17.6	28.5	2.2	0.3	2.4	1.70	0.87	1.67	0.53	0.09	0.42
고양 덕양구	53.2	18.7	23.7	1.1	0.3	2.9	1.64	0.90	1.64	0.17	0.16	0.30
고양 일산동	49.3	19.2	27.0	2.3	0.2	2.0	2.25	0.99	2.34	0.97	0.13	0.35
고양 일산서	60.5	21.9	14.9	1.1	0.2	1.4	1.68	1.13	1.41	0.39	0.07	0.33
과천시	39.9	33.8	19.2	1.1	1.5	4.6	2.02	1.60	1.56	0.35	1.35	0.61
구리시	53.1	23.8	19.3	1.6	0.4	1.8	1.75	1.23	1.54	0.44	0.30	0.28
남양주시	57.0	19.1	20.4	1.1	0.2	2.2	1.46	0.73	1.64	0.19	0.06	0.29
오산시	46.6	15.2	31.5	4.2	0.2	2.2	2.40	0.99	2.49	0.63	0.10	0.48
시흥시	53.3	15.4	24.4	4.2	0.4	2.4	1.33	0.69	1.20	0.62	0.17	0.27
군포시	56.3	20.6	20.2	0.9	0.1	1.9	1.91	1.05	1.93	0.33	0.05	0.39
의왕시	53.7	22.9	20.4	0.9	0.2	2.0	2.14	1.39	2.21	0.35	0.07	0.46
하남시	47.8	24.2	23.2	1.3	0.2	3.3	2.08	1.48	2.31	0.52	0.09	0.56
용인 처인구	56.3	18.2	17.6	2.2	0.6	5.1	2.23	1.32	1.70	0.50	0.24	0.91
용인 기흥구	53.4	21.1	22.5	1.0	0.2	1.8	2.10	1.22	2.33	0.67	0.09	0.30
용인 수지구	59.8	21.8	15.7	1.2	0.2	1.3	1.56	0.95	1.58	0.52	0.12	0.17
파주시	51.2	17.2	25.5	1.9	0.4	3.9	2.00	0.95	2.21	0.48	0.11	0.53
이천시	62.3	13.9	14.3	2.4	0.7	6.3	2.45	1.03	2.14	0.39	0.28	1.43
안성시	59.8	12.5	17.3	4.5	1.1	4.9	2.45	1.32	1.91	1.03	0.31	0.72
김포시	53.1	18.5	23.8	1.2	0.3	3.2	1.90	1.32	1.97	0.33	0.12	0.61
화성시	55.0	18.2	20.1	2.2	0.2	4.3	1.76	0.94	1.70	0.48	0.07	0.50
광주시	59.5	20.6	15.0	1.4	0.2	3.3	1.92	1.32	1.22	0.23	0.07	0.42
양주시	55.8	15.4	21.0	1.3	0.3	6.3	2.48	1.05	2.79	0.24	0.10	1.03
포천시	59.4	10.1	18.9	2.3	0.4	8.8	2.37	0.92	2.23	0.33	0.13	1.01
여주군	69.8	11.8	10.2	2.7	0.3	5.2	1.90	1.05	1.44	0.43	0.14	0.85
연천군	69.5	8.5	11.3	2.7	0.6	7.4	3.87	1.84	3.65	0.74	0.25	2.53
가평군	66.6	10.9	12.3	2.7	0.5	7.0	3.26	1.61	2.61	0.60	0.19	1.73
양평군	70.4	11.5	9.0	2.1	0.3	6.6	2.03	1.14	1.22	0.51	0.12	1.35

<표 2-9>는 각 시군구의 사용 방 수 분포와 오차의 한계 현황이다. 전체적으로 오차의 한계가 2% 미만으로 대체로 안정적인 추정이 가능한 것으로 판단된다.

<표 2-9> 각 시군구의 사용 방 수 분포 현황

(1) 대전광역시

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	1개	2개	3개	4개	5개+	무응답	1개	2개	3개	4개	5개+	무응답
대전 동구	44.7	31.7	22.5	1.0	0.1	0.0	1.63	1.24	1.34	0.27	0.06	0.02
대전 중구	35.4	32.9	28.6	2.8	0.3		1.62	1.22	1.50	0.58	0.11	0.00
대전 서구	41.0	30.9	26.0	2.0	0.1	0.0	1.33	0.98	1.08	0.37	0.06	0.01
대전 유성구	39.4	29.1	28.5	2.9	0.1	0.0	1.64	1.15	1.44	0.51	0.07	0.01
대전 대덕구	36.0	34.9	27.5	1.5	0.1	0.0	1.93	1.56	1.83	0.49	0.05	0.03

(2) 제주도

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	1개	2개	3개	4개	5개+	무응답	1개	2개	3개	4개	5개+	무응답
제주 제주시	34.4	31.4	31.3	2.6	0.4	-	1.08	0.77	1.03	0.30	0.07	0.00
제주 서귀포	30.4	36.7	29.4	2.9	0.5	0.0	1.62	1.56	1.89	0.40	0.15	0.05

(3) 경기도

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	1개	2개	3개	4개	5개+	무응답	1개	2개	3개	4개	5개+	무응답
수원 장안구	31.7	33.5	32.3	2.4	0.1	0.0	0.75	0.67	0.78	0.36	0.02	0.01
수원 권선구	32.8	30.5	34.0	2.6	0.1	0.0	0.82	0.69	0.87	0.37	0.05	0.01
수원 팔달구	42.5	33.0	21.9	2.5	0.1	0.0	0.98	0.73	0.78	0.35	0.03	0.01
수원 영통구	36.9	27.8	31.5	3.6	0.1		0.83	0.61	0.79	0.34	0.09	
성남 수정구	36.3	43.3	19.5	0.9	0.1	0.0	0.74	0.64	0.60	0.20	0.01	0.01
성남 중원구	30.5	44.3	24.1	1.0	0.0	0.0	0.81	0.77	0.75	0.21	0.01	0.01
성남 분당구	33.7	30.7	29.2	6.0	0.4	0.0	0.75	0.54	0.66	0.34	0.07	0.00
의정부시	33.9	35.8	28.5	1.7	0.1		0.81	0.67	0.78	0.27	0.06	
안양 만안구	30.7	34.6	33.1	1.6	0.1		0.80	0.68	0.86	0.27	0.04	
안양 동안구	30.2	34.6	32.5	2.7	0.1	0.0	0.73	0.63	0.83	0.27	0.03	0.01
부천 원미구	33.5	32.1	31.9	2.3	0.2	0.0	0.68	0.53	0.68	0.24	0.04	0.01
부천 소사구	30.1	37.2	30.8	1.8	0.1		0.90	0.84	0.98	0.31	0.02	
부천 오정구	26.9	40.3	32.1	0.7	0.1		0.78	0.83	1.04	0.14	0.03	
광명시	31.2	39.7	27.3	1.8	0.1	0.0	0.62	0.66	0.79	0.27	0.02	0.01
평택시	33.9	33.2	30.5	2.2	0.2	0.0	0.69	0.60	0.74	0.23	0.03	0.00
동두천시	39.3	36.7	23.1	0.8	0.0	0.1	1.05	0.80	1.15	0.18	0.02	0.06
안산 상록구	33.1	33.4	31.4	1.9	0.2	0.0	0.68	0.63	0.72	0.24	0.07	0.01
안산 단원구	36.9	34.2	27.3	1.5	0.1	0.0	0.84	0.70	0.78	0.23	0.03	0.01
고양 덕양구	31.1	35.5	29.9	3.3	0.1	0.0	0.69	0.68	0.78	0.35	0.04	0.00
고양 일산동	38.1	28.4	28.4	4.8	0.3	0.0	1.15	0.71	0.85	0.38	0.06	0.01
고양 일산서	26.1	32.4	36.6	4.6	0.3	0.0	0.81	0.67	0.89	0.48	0.08	0.01
과천시	25.5	35.6	34.5	4.1	0.3		0.98	0.77	1.02	0.51	0.06	
구리시	26.5	30.6	39.5	3.3	0.2		0.94	0.96	1.21	0.69	0.04	
남양주시	26.1	34.6	35.4	3.8	0.1	0.0	0.55	0.55	0.65	0.29	0.02	0.00
오산시	44.0	31.5	23.2	1.2	0.1	0.0	0.89	0.72	0.77	0.18	0.03	0.01

시흥시	37.0	30.2	31.2	1.5	0.1		0.71	0.59	0.74	0.23	0.03	
군포시	34.7	36.8	25.6	2.8	0.1		0.86	0.73	0.91	0.32	0.02	
의왕시	33.6	35.9	28.5	1.9	0.1	0.0	0.87	0.74	0.91	0.26	0.03	0.01
하남시	27.7	37.8	31.5	2.7	0.2	0.0	1.00	1.09	1.18	0.43	0.05	0.02
용인 처인구	31.5	31.3	33.9	2.8	0.5	0.0	1.05	0.80	0.97	0.32	0.14	0.01
용인 기흥구	28.4	31.9	33.8	5.6	0.4		0.82	0.68	0.86	0.45	0.12	
용인 수지구	26.1	30.2	36.2	7.1	0.4		0.79	0.53	0.68	0.44	0.10	
파주시	33.4	34.5	29.2	2.7	0.2	0.0	0.74	0.58	0.70	0.25	0.03	0.00
이천시	32.7	34.7	30.5	1.9	0.2	0.0	0.95	0.88	0.93	0.19	0.04	0.01
안성시	32.8	33.0	31.4	2.4	0.3	0.0	1.07	0.91	1.11	0.26	0.06	0.02
김포시	32.2	33.1	29.8	4.8	0.2	0.0	0.75	0.66	0.72	0.44	0.03	0.01
화성시	33.5	29.7	34.5	2.2	0.1	0.0	0.65	0.55	0.73	0.21	0.02	0.00
광주시	33.8	31.6	32.2	2.1	0.2	0.0	0.79	0.48	0.69	0.21	0.04	0.01
양주시	34.8	35.3	27.9	1.8	0.2	0.0	0.81	0.81	0.92	0.26	0.06	0.01
포천시	43.3	34.0	21.0	1.5	0.2	0.0	0.88	0.68	0.76	0.15	0.04	0.02
여주군	32.8	34.2	30.2	2.4	0.4		1.02	0.79	0.89	0.26	0.07	
연천군	49.3	33.7	15.7	1.1	0.2		1.33	1.22	1.07	0.19	0.08	
가평군	24.8	34.9	37.2	2.8	0.3		1.43	1.35	1.49	0.28	0.07	
양평군	30.1	32.6	32.5	4.0	0.8		0.94	0.76	1.03	0.34	0.11	

(2) 개인 부문

<표 2-10>은 각 시군구의 혼인 상태별 분포와 오차의 한계 현황이다. 개인의 혼인상태는 ①미혼, ②유배우자(배우자 있음+동거), ③사별, ④이혼+별거, ⑤결측 등으로 구분하였다. 각 시군구에서 혼인 상태별 모비율 추정값의 95% 신뢰수준 오차의 한계는 전체적으로 1.5% 미만으로 안정적인 추정이 가능하다.

<표 2-10> 각 시군구의 혼인상태 분포 현황

(1) 대전광역시

시군구	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	미혼	유배우자	사별	이혼	무응답	미혼	유배우자	사별	이혼	무응답
대전 동구	32.8	51.4	7.9	6.6	1.3	0.74	0.92	0.37	0.39	0.11
대전 중구	29.5	55.2	8.0	5.8	1.6	0.54	0.71	0.34	0.33	0.12
대전 서구	32.3	55.5	5.4	5.1	1.7	0.53	0.62	0.22	0.24	0.10
대전 유성구	30.5	60.0	4.1	3.6	1.8	0.72	0.82	0.26	0.29	0.13
대전 대덕구	31.2	54.7	6.7	5.7	1.7	0.79	0.92	0.38	0.39	0.14

(2) 제주도

시군구	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	미혼	유배우자	사별	이혼	무응답	미혼	유배우자	사별	이혼	무응답
제주 제주시	29.2	55.1	7.4	6.4	1.9	0.42	0.65	0.24	0.30	0.10
제주 서귀포	25.6	56.1	10.3	6.3	1.6	0.69	1.10	0.49	0.53	0.16

(3) 경기도

시군구	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	미혼	유배우자	사별	이혼	무응답	미혼	유배우자	사별	이혼	무응답
수원 장안구	27.1	48.5	4.8	4.8	14.9	0.56	0.46	0.26	0.26	0.46
수원 권선구	27.5	48.3	4.5	4.6	15.2	0.59	0.47	0.24	0.24	0.50
수원 팔달구	28.8	47.7	6.1	5.8	11.5	0.76	0.64	0.33	0.32	0.56
수원 영통구	26.3	49.6	2.9	2.7	18.6	0.65	0.49	0.19	0.18	0.49
성남 수정구	29.9	45.9	6.4	7.0	10.8	0.58	0.57	0.26	0.29	0.39
성남 중원구	28.8	46.1	6.0	6.9	12.1	0.55	0.54	0.25	0.32	0.40
성남 분당구	25.2	52.4	3.8	3.0	15.6	0.45	0.45	0.20	0.19	0.39
의정부시	26.4	48.0	6.0	5.6	14.0	0.61	0.52	0.24	0.26	0.42
안양 만안구	28.0	49.4	5.7	5.2	11.8	0.55	0.55	0.25	0.28	0.45
안양 동안구	26.8	50.5	4.2	3.6	15.0	0.51	0.43	0.21	0.22	0.47
부천 원미구	28.5	48.6	4.8	4.7	13.4	0.48	0.43	0.22	0.21	0.36
부천 소사구	27.1	49.5	5.5	5.0	12.9	0.64	0.58	0.28	0.30	0.53
부천 오정구	29.0	46.1	5.9	6.2	12.8	0.60	0.61	0.31	0.36	0.49
광명시	25.0	49.8	5.2	4.4	15.6	0.47	0.45	0.26	0.24	0.42
평택시	24.7	49.1	5.7	4.6	15.9	0.56	0.47	0.24	0.23	0.43
동두천시	23.9	46.9	7.7	7.1	14.4	0.87	1.05	0.42	0.60	0.76
안산 상록구	28.6	46.5	4.7	6.0	14.2	0.47	0.44	0.21	0.26	0.37
안산 단원구	27.4	50.0	4.3	5.1	13.2	0.53	0.55	0.21	0.25	0.37
고양 덕양구	25.0	49.8	5.5	5.0	14.7	0.45	0.45	0.22	0.25	0.43
고양 일산동	27.4	49.6	4.5	4.5	14.0	0.63	0.61	0.24	0.31	0.49

고양 일산서	24.4	51.8	4.4	3.5	16.0	0.54	0.52	0.25	0.27	0.50
과천시	25.7	52.4	4.8	2.7	14.4	0.64	0.54	0.29	0.26	0.59
구리시	26.5	49.2	5.2	4.6	14.4	0.63	0.57	0.28	0.30	0.56
남양주시	21.9	51.6	5.0	4.3	17.2	0.35	0.32	0.17	0.19	0.37
오산시	24.1	49.6	3.7	3.9	18.7	0.61	0.52	0.21	0.25	0.54
시흥시	27.5	48.4	4.0	5.0	15.0	0.55	0.46	0.17	0.23	0.39
군포시	25.0	51.0	4.6	3.9	15.5	0.58	0.52	0.27	0.27	0.52
의왕시	25.9	51.1	4.7	3.8	14.5	0.67	0.59	0.26	0.31	0.56
하남시	23.7	51.9	5.5	4.2	14.7	0.59	0.58	0.34	0.32	0.55
용인 처인구	25.2	50.4	5.5	4.2	14.7	0.91	0.71	0.29	0.27	0.60
용인 기흥구	20.7	53.0	3.7	3.0	19.5	0.51	0.47	0.21	0.22	0.52
용인 수지구	21.7	54.1	3.4	2.3	18.5	0.60	0.44	0.18	0.16	0.54
파주시	21.6	51.0	5.7	4.2	17.4	0.56	0.50	0.25	0.25	0.48
이천시	21.8	51.7	6.2	3.8	16.6	0.70	0.65	0.32	0.31	0.72
안성시	24.1	50.6	6.5	4.0	14.8	0.92	0.71	0.31	0.28	0.62
김포시	22.2	51.6	4.7	3.8	17.7	0.84	0.60	0.23	0.24	0.54
화성시	22.2	50.7	3.8	3.2	20.1	0.61	0.42	0.18	0.19	0.47
광주시	23.9	51.7	4.4	4.4	15.6	0.65	0.54	0.20	0.22	0.49
양주시	23.2	50.1	5.7	4.8	16.2	0.99	0.76	0.33	0.35	0.70
포천시	25.7	50.5	7.0	4.5	12.2	1.15	0.84	0.37	0.34	0.65
여주군	21.4	52.5	8.3	3.8	14.0	0.74	0.69	0.41	0.31	0.67
연천군	20.4	52.6	10.6	4.2	12.2	1.20	1.24	0.75	0.49	1.35
가평군	19.7	53.0	10.2	5.1	12.1	1.08	1.12	0.73	0.52	0.96
양평군	18.5	55.8	9.3	4.0	12.4	0.74	0.82	0.44	0.34	0.69

<표 2-11>은 각 시군구의 경제활동상태별 비율 추정값과 오차의 한계 현황이다. 개인의 경제활동상태는 ①주로 일했음, ②가사, 학업(학교, 학원) 등 다른 활동을 하면서 틈틈이 도움, ③일자리가 있으나, 휴가 등으로 잠시 쉬, ④일하지 않았음, ⑤무응답 등으로 구분하였다.

각 시군구의 경제활동상태별 모비율 추정의 95% 신뢰수준 오차의 한계는 전체적으로 1.5% 미만으로 안정적인 추정이 가능하다.

<표 2-11> 각 시군구의 경제활동상태 분포 현황:

(1) 대전광역시

시군구	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	일했음	가사 등	잠시 쉬	일 않음	무응답	일했음	가사 등	잠시 쉬	일 않음	무응답
대전 동구	46.4	2.8	0.8	48.3	1.7	0.75	0.21	0.09	0.78	0.14
대전 중구	48.0	2.4	0.7	46.0	2.9	0.60	0.18	0.10	0.62	0.27
대전 서구	50.4	3.0	0.8	43.8	2.0	0.52	0.14	0.08	0.53	0.12
대전 유성구	50.4	3.8	1.1	40.6	4.1	0.65	0.23	0.11	0.63	0.27
대전 대덕구	49.8	2.3	0.8	44.4	2.8	0.92	0.19	0.10	0.88	0.24

(2) 제주도

시군구	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	일했음	가사 등	잠시 쉬	일 않음	무응답	일했음	가사 등	잠시 쉬	일 않음	무응답
제주 제주시	58.4	3.0	0.8	35.9	1.9	0.45	0.16	0.07	0.44	0.10
제주 서귀포	62.3	3.6	0.6	31.9	1.6	0.79	0.38	0.11	0.72	0.16

(3) 경기도

시군구	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	일했음	가사 등	잠시 쉬	일 않음	무응답	일했음	가사 등	잠시 쉬	일 않음	무응답
수원 장안구	50.7	3.1	1.1	42.3	2.9	0.59	0.18	0.10	0.59	0.23
수원 권선구	53.1	2.8	1.0	38.5	4.6	0.59	0.17	0.10	0.59	0.30
수원 팔달구	52.2	2.5	0.8	40.4	4.2	0.74	0.20	0.11	0.75	0.40
수원 영통구	53.6	3.6	1.3	38.0	3.6	0.61	0.18	0.11	0.60	0.21
성남 수정구	54.1	2.6	1.0	38.3	4.1	0.59	0.17	0.11	0.56	0.36
성남 중원구	55.2	2.7	1.0	38.2	2.9	0.56	0.18	0.10	0.54	0.22
성남 분당구	48.4	3.6	1.1	43.4	3.5	0.54	0.15	0.09	0.58	0.20
의정부시	49.4	2.5	0.7	43.0	4.4	0.56	0.15	0.07	0.63	0.29
안양 만안구	52.6	3.1	1.0	40.7	2.7	0.59	0.20	0.10	0.61	0.24
안양 동안구	51.9	3.5	1.0	40.3	3.2	0.49	0.18	0.09	0.49	0.21
부천 원미구	53.2	3.0	1.0	38.9	4.0	0.51	0.15	0.10	0.51	0.25
부천 소사구	52.2	2.7	0.8	40.3	3.9	0.62	0.21	0.11	0.62	0.32
부천 오정구	52.7	2.7	0.9	39.8	3.8	0.70	0.21	0.12	0.66	0.31
광명시	51.4	3.0	1.1	41.3	3.1	0.52	0.16	0.10	0.54	0.22
평택시	54.0	2.1	0.6	38.2	5.0	0.60	0.14	0.07	0.59	0.32
동두천시	48.3	2.7	0.7	42.2	6.0	1.01	0.28	0.14	0.99	0.60
안산 상록구	51.5	2.6	0.8	41.1	4.1	0.53	0.15	0.08	0.53	0.27
안산 단원구	56.1	2.3	0.9	35.6	5.1	0.63	0.14	0.17	0.58	0.40
고양 덕양구	50.2	2.9	1.0	41.7	4.2	0.51	0.14	0.09	0.52	0.26
고양 일산동	50.5	3.6	1.1	41.2	3.6	0.65	0.20	0.11	0.64	0.25
고양 일산서	49.0	3.2	0.9	44.0	3.0	0.64	0.20	0.10	0.63	0.23
과천시	47.0	4.3	1.2	44.2	3.3	0.72	0.29	0.15	0.77	0.30
구리시	51.2	2.9	0.8	41.1	4.0	0.80	0.24	0.11	0.79	0.34
남양주시	50.8	3.0	0.9	40.9	4.3	0.39	0.14	0.07	0.40	0.23
오산시	53.9	1.9	0.7	34.6	8.9	0.65	0.16	0.08	0.64	0.50
시흥시	56.2	2.5	0.8	36.2	4.3	0.54	0.14	0.08	0.51	0.25
군포시	52.9	3.0	0.9	38.9	4.4	0.59	0.18	0.09	0.58	0.29
의왕시	51.5	3.3	1.0	40.1	4.1	0.76	0.22	0.12	0.73	0.35
하남시	52.6	2.8	0.7	39.5	4.4	0.72	0.24	0.12	0.75	0.45
용인 처인구	54.8	2.9	0.7	36.8	4.8	0.86	0.22	0.09	0.82	0.45
용인 기흥구	49.5	3.2	1.1	41.5	4.7	0.62	0.19	0.10	0.61	0.28
용인 수지구	46.1	3.6	1.1	44.7	4.5	0.65	0.20	0.10	0.67	0.28
파주시	54.0	2.7	0.8	39.0	3.5	0.59	0.16	0.09	0.57	0.25
이천시	56.8	2.1	0.9	35.3	5.0	0.84	0.20	0.14	0.79	0.52
안성시	54.5	2.6	0.7	34.6	7.7	0.98	0.22	0.11	0.84	0.73
김포시	54.9	2.9	0.9	37.2	4.1	0.83	0.18	0.09	0.74	0.30
화성시	58.3	2.6	0.9	33.1	5.1	0.71	0.15	0.08	0.60	0.31
광주시	56.5	2.5	0.8	36.9	3.2	0.65	0.15	0.09	0.63	0.24
양주시	52.9	2.7	0.8	39.1	4.5	1.05	0.22	0.12	0.94	0.48

포천시	58.3	2.2	0.4	35.6	3.5	1.32	0.22	0.08	1.18	0.36
여주군	54.3	2.9	0.9	39.0	2.9	0.87	0.24	0.14	0.86	0.34
연천군	49.7	6.2	0.8	37.9	5.4	1.22	0.65	0.27	1.40	0.91
가평군	51.3	5.7	0.3	41.1	1.6	1.28	0.56	0.10	1.44	0.24
양평군	46.7	3.0	0.6	44.9	4.7	0.94	0.30	0.12	0.99	0.54

3) 주요 조사항목 추정에 대한 분석: 동읍면 단위

2015년 인구주택총조사의 20% 표본조사의 주요 조사항목에 대한 동읍면 구분별 추정값과 95% 신뢰수준 오차의 한계를 분석하여 통계공표 수준을 살펴본다. 본 연구에서 고려하는 주요 조사항목은 다음과 같다.

- 가구 부문(3): 가구유형, 점유형태, 사용 방 수(잠을 자는 방)
- 개인 부문(2): 혼인상태, 경제활동상태

수원시 장안구(10개 동), 의왕시(6개 동), 가평군(6개 읍면) 등을 대상으로 동읍면별 통계의 추정 정확도를 살펴본다. 각 동읍면별 표본조사구와 응답가구 현황은 <표 2-12>와 같다. 동읍면에서 조사되는 표본 조사구 수는 31~94개 범위이고, 응답가구 수와 응답자 수는 각각 673~2,701호와 1,321~6,809명 범위로 상당한 편차를 보이고 있다.

<표 2-12> 분석대상 동읍면의 표본 조사구, 응답가구, 응답자 현황

동읍면	표본 조사구 수	응답가구 수	응답자 수
3101154	45	1,295	3,403
3101155	94	2,701	6,809
3101156	49	1,580	4,698
3101157	52	1,516	4,199
3101158	49	1,334	2,898
3101159	38	1,061	2,633
3101160	55	1,544	4,030
3101161	44	1,175	2,718
3101162	79	2,437	7,900
3101163	33	993	3,224
3117051	36	1,167	3,384
3117052	60	1,720	4,701
3117053	76	2,364	7,085
3117054	51	1,626	4,441
3117055	70	2,092	5,920
3117056	50	1,492	3,994
3137011	49	1,276	3,034
3137031	36	718	1,758
3137032	42	1,085	2,511

3137033	37	684	1,509
3137034	40	811	1,720
3137035	31	673	1,321

(1) 가구 부문

<표 2-13>은 분석대상 동읍면의 가구유형별 분포와 오차의 한계 현황이다. 전체적으로 가족과 1인 가족 유형이 대부분을 차지하고 있으며, 95% 신뢰수준 오차의 한계는 대체로 안정적인 수준으로 판단된다.

<표 2-13> 각 동읍면의 가족유형별 가구 분포 현황: 경기도

시군구	각 응답항목별 비율 추정값					비율 추정에 대한 오차의 한계(95%)				
	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상	가족	가족+이외	1인 가족	남남 5인 이하	남남 6인 이상
3101154	74.5	0.1	24.6	0.8	-	3.56	0.16	3.59	0.60	-
3101155	68.0	0.2	29.8	2.0	-	4.01	0.16	3.75	0.83	-
3101156	82.9	0.1	16.2	0.8	-	4.07	0.16	4.05	0.55	-
3101157	77.8	0.3	21.2	0.8	-	3.31	0.26	3.39	0.60	-
3101158	62.3	0.2	35.9	1.6	-	4.30	0.24	4.38	0.95	-
3101159	71.7	0.1	27.1	1.2	-	4.78	0.18	4.71	0.87	-
3101160	72.8	0.2	26.3	0.8	-	5.59	0.16	5.61	0.77	-
3101161	65.8	0.2	30.9	3.2	-	5.02	0.30	4.51	1.63	-
3101162	83.4	0.0	15.7	0.9	-	2.25	0.06	2.19	0.66	-
3101163	92.2	-	7.7	0.1	-	2.23	0.00	2.23	0.01	-
3117051	81.1	0.3	17.3	1.3	-	3.16	0.34	3.20	0.99	-
3117052	77.5	0.2	21.3	1.0	-	3.76	0.15	3.58	0.67	-
3117053	85.8	0.3	13.1	0.8	-	2.00	0.26	1.96	0.69	-
3117054	76.0	0.1	22.1	1.8	-	3.71	0.14	3.00	1.17	-
3117055	80.0	0.3	18.5	1.1	-	2.94	0.29	2.78	0.55	-
3117056	74.7	0.1	24.5	0.7	-	5.63	0.11	5.48	0.53	-
3137011	70.8	0.2	27.4	1.6	-	3.63	0.21	3.58	0.77	-
3137031	65.0	0.1	32.8	2.1	-	4.28	0.10	4.13	1.47	-
3137032	64.6	0.1	32.3	3.0	-	5.25	0.13	4.64	1.69	-
3137033	63.1	0.1	34.5	2.3	-	4.85	0.16	3.87	1.76	-
3137034	68.8	0.1	28.5	2.6	-	6.83	0.22	5.68	2.19	-
3137035	58.9	0.1	38.8	2.2	-	3.68	0.09	3.73	1.18	-

<표 2-14>는 분석대상 동읍면의 점유형태별 분포와 오차의 한계 현황이다. 오차의 한계가 5%를 넘는 동읍면이 11곳이고, 4%-5%인 경우도 4곳이 있다. 가구유형별 분석에 비해서 좀 더 큰 오차의 한계를 나타내고 있다.

<표 2-14> 각 동읍면의 점유형태별 가구 분포 현황: 경기도

시군구	각 응답항목별 비율 추정값						비율 추정값 오차의 한계(95%)					
	자가	전세	월세1	월세2	사글세	무상	자가	전세	월세1	월세2	사글세	무상
3101154	52.0	23.2	18.8	3.7		2.2	4.46	2.94	3.95	1.46	0.00	0.72
3101155	48.4	19.6	29.6	1.7	0.1	0.7	4.21	2.49	4.34	1.09	0.14	0.32
3101156	66.7	19.0	11.0	1.8		1.5	4.74	2.99	3.36	2.13	0.00	0.55
3101157	59.8	19.0	19.2	1.0	0.1	0.9	3.43	2.93	2.96	0.58	0.18	0.55
3101158	36.8	25.3	34.1	1.1	0.1	2.7	5.82	3.82	5.25	0.57	0.11	0.86
3101159	47.8	25.0	24.4	1.2	0.1	1.5	5.84	3.18	5.21	0.66	0.18	0.68
3101160	52.0	18.1	23.6	0.9	0.2	5.2	6.64	2.64	4.76	0.54	0.28	4.32
3101161	42.3	24.6	27.4	2.6	0.1	3.0	5.39	4.09	4.13	1.01	0.21	1.08
3101162	70.4	16.2	11.9	0.5	0.2	0.8	2.49	2.04	2.12	0.47	0.17	0.49
3101163	83.2	11.2	4.6	0.2		0.9	2.15	1.81	1.18	0.19	0.00	0.71
3117051	65.7	18.8	11.2	1.7	0.4	2.1	4.35	2.81	3.26	0.77	0.39	1.11
3117052	58.0	20.4	18.1	1.0	0.1	2.4	6.14	3.00	5.67	0.84	0.14	1.15
3117053	69.3	17.7	10.2	0.6	0.0	2.2	3.25	2.12	2.36	0.44	0.08	1.27
3117054	48.3	20.2	28.1	1.3	0.0	2.0	6.00	2.80	7.22	1.76	0.09	1.39
3117055	44.0	34.9	18.7	0.7	0.4	1.3	2.96	3.17	2.72	0.35	0.26	0.47
3117056	34.3	22.3	40.8	0.7		2.0	7.97	5.99	9.97	0.55	0.00	0.84
3137011	64.6	9.3	18.8	2.7		4.6	6.93	2.14	6.90	1.15	0.00	2.90
3137031	71.6	10.3	10.0	3.6	0.4	4.2	6.77	5.68	3.33	1.77	0.43	1.36
3137032	59.0	15.5	14.9	3.1	0.2	7.2	7.08	3.22	4.79	1.46	0.25	4.01
3137033	81.9	4.1	2.6	2.3	1.0	8.1	2.81	1.67	1.26	1.36	0.88	1.97
3137034	62.4	16.6	6.0	1.7	1.8	11.5	9.17	6.45	3.35	0.81	1.04	7.36
3137035	80.4	1.3	2.3	1.6	1.2	13.2	4.97	0.79	1.30	0.84	0.82	4.98

<표 2-15>는 분석대상 각 동읍면의 사용 방 수 분포와 오차의 한계 현황이다. 오차의 한계가 5%를 넘는 동읍면이 10곳이고, 4%-5%인 경우도 6곳이 있다. 가구 유형별 분석에 비해서 좀 더 큰 오차의 한계를 나타내고 있다.

<표 2-15> 각 동읍면의 사용 방 수 분포 현황: 경기도

동읍면	각 응답항목별 비율 추정값						비율 추정값 오차의 한계 (95%)					
	1개	2개	3개	4개	5개+	무응답	1개	2개	3개	4개	5개+	무응답
3101154	36.1	35.7	26.9	1.0	0.1	0.2	3.56	2.97	4.59	1.02	0.12	0.22
3101155	36.1	30.1	32.0	1.6	0.1		3.91	3.82	3.99	1.32	0.15	0.00
3101156	24.2	32.3	38.0	5.6			4.66	5.18	6.64	3.94	0.00	0.00
3101157	31.9	33.6	32.0	2.3	0.2		3.93	3.36	3.67	2.66	0.24	0.00
3101158	45.1	33.1	21.1	0.7			5.20	3.68	4.49	0.45	0.00	0.00
3101159	37.6	37.0	25.0	0.5			6.38	4.84	4.84	0.34	0.00	0.00
3101160	27.9	39.0	32.7	0.4			5.79	4.67	5.57	0.34	0.00	0.00
3101161	33.6	42.9	22.3	1.1			5.63	5.82	4.60	0.82	0.00	0.00
3101162	25.2	26.1	44.4	4.3	0.0		2.86	2.27	3.19	2.46	0.08	0.00
3101163	18.9	32.3	41.8	6.8	0.2		3.99	4.30	5.71	3.81	0.28	0.00
3117051	30.0	36.9	28.4	4.5	0.1		4.32	3.57	4.52	2.19	0.12	0.00
3117052	34.5	41.0	23.2	1.3	0.1		3.79	4.19	4.62	1.05	0.14	0.00
3117053	26.8	35.7	36.4	1.0	0.1		3.28	2.70	4.02	0.89	0.09	0.00
3117054	34.6	32.4	30.5	2.4	0.1		3.50	4.27	4.57	1.43	0.17	0.00
3117055	34.5	34.7	29.2	1.5	0.1	0.1	3.81	2.39	3.39	1.25	0.13	0.12
3117056	44.0	35.2	18.2	2.5	0.1		6.24	4.39	4.79	1.39	0.12	0.00
3137011	18.2	35.9	43.7	2.0	0.1		5.72	6.08	6.59	0.91	0.22	0.00
3137031	43.8	33.3	21.1	1.6	0.2		4.89	3.89	3.74	0.85	0.28	0.00
3137032	22.6	31.8	42.2	3.3	0.2		6.76	5.67	6.24	1.41	0.21	0.00
3137033	26.2	41.0	29.5	2.7	0.6		3.86	3.52	3.75	1.05	0.47	0.00
3137034	22.3	37.3	35.9	3.7	0.7		7.61	6.78	8.41	1.56	0.55	0.00
3137035	25.4	32.3	36.1	5.0	1.2		4.18	2.89	4.12	1.77	0.84	0.00

(2) 개인 부문

<표 2-16>은 각 시군구의 혼인 상태별 분포와 오차의 한계 현황이다. 개인 부문 조사항목에 대한 오차의 한계는 다소 작게 나타나고 있다. 오차의 한계가 4%를 넘는 동읍면이 3곳이 있다.

<표 2-16> 각 동읍면의 혼인상태 분포 현황: 경기도

동읍면	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	미혼	유배우자	사별	이혼	무응답	미혼	유배우자	사별	이혼	무응답
3101154	27.2	47.9	5.5	6.8	12.6	1.70	1.67	0.89	1.04	1.38
3101155	30.2	47.1	4.0	3.3	15.3	1.88	1.46	0.61	0.56	1.09
3101156	22.8	51.1	4.6	3.3	18.2	1.64	1.26	0.75	0.69	1.82
3101157	26.3	49.0	4.5	5.1	15.0	1.51	1.09	0.65	0.77	1.53
3101158	29.2	46.2	6.5	9.8	8.4	1.97	2.07	1.01	1.15	1.18
3101159	28.6	45.9	5.8	7.1	12.4	2.17	1.65	0.92	1.07	1.73
3101160	28.3	46.1	5.5	6.3	13.7	1.63	1.89	1.44	1.11	1.43
3101161	32.7	42.3	8.8	7.2	9.0	2.55	2.10	1.03	1.16	1.04
3101162	23.7	52.1	2.3	1.5	20.5	1.30	0.69	0.35	0.30	1.14
3101163	24.9	54.1	3.6	1.9	15.5	1.50	1.14	0.56	0.56	1.57
3117051	27.0	52.0	5.3	4.1	11.6	2.55	1.69	0.81	0.89	1.49
3117052	26.1	49.1	6.0	5.0	13.8	1.48	1.42	0.71	0.75	1.36
3117053	24.6	53.3	4.0	3.1	15.0	1.18	1.17	0.50	0.58	1.13
3117054	27.8	51.2	3.7	3.0	14.3	2.10	1.73	0.61	0.73	1.45
3117055	26.0	50.7	4.7	3.8	14.8	1.35	0.98	0.56	0.55	1.21
3117056	25.1	49.0	5.0	4.8	16.2	1.98	2.07	0.81	1.26	1.74
3137011	20.5	50.3	11.0	5.2	12.9	1.83	1.89	1.43	0.90	1.80
3137031	16.7	56.6	9.9	3.9	13.0	2.24	2.63	1.82	1.02	3.11
3137032	20.5	52.9	9.1	6.2	11.4	1.85	2.31	1.69	1.41	1.70
3137033	18.1	58.3	11.4	4.4	7.7	4.40	3.25	1.72	1.15	2.45
3137034	22.6	50.1	8.0	4.9	14.4	4.36	3.92	1.39	1.29	2.34
3137035	15.9	57.8	13.6	3.8	8.8	2.41	3.12	1.79	0.88	3.28

<표 2-17>은 각 시군구의 경제활동상태별 비율 추정값과 오차의 한계 현황이다. 개인의 경제활동상태는 ①주로 일했음, ②가사, 학업(학교, 학원) 등 다른 활동을 하면서 틈틈이 도움, ③일자리가 있으나, 휴가 등으로 잠시 쉬, ④일하지 않았음, ⑤무응답 등으로 구분하였다. 표본의 크기가 상대적으로 큰 시 지역의 동에서는 전체적으로 오차의 한계가 2.5% 미만으로 안정적인 추정이 가능하지만, 표본의 크기가 작은 면 지역에서 오차의 한계가 4% 이상인 경우가 나타나고 있다.

<표 2-17> 각 동읍면의 경제활동상태 분포 현황: 경기도

동읍면	각 응답항목별 비율 추정값					비율 추정값 오차의 한계(95%)				
	일했음	가사 등	잠시 쉬	일 않음	무응답	일했음	가사 등	잠시 쉬	일 않음	무응답
3101154	51.0	2.7	1.2	41.9	3.2	2.01	0.60	0.46	1.97	0.91
3101155	48.2	3.5	1.2	44.1	3.0	1.78	0.44	0.23	1.67	0.59
3101156	52.1	2.0	0.8	42.0	3.1	1.49	0.44	0.25	1.57	0.77
3101157	51.9	3.2	1.0	41.6	2.3	1.96	0.57	0.37	1.93	0.58
3101158	56.1	2.7	1.5	38.3	1.3	2.04	0.66	0.41	1.95	0.43
3101159	53.8	3.1	1.2	39.0	2.9	2.06	0.88	0.43	2.23	0.93
3101160	49.6	2.5	1.0	43.8	3.1	2.19	0.55	0.29	2.45	0.76
3101161	48.1	3.5	0.9	45.2	2.3	2.09	0.80	0.37	1.91	0.69
3101162	49.1	3.6	1.2	42.1	4.0	1.22	0.43	0.26	1.18	0.62
3101163	49.1	4.1	0.7	43.8	2.3	1.44	0.74	0.28	1.77	0.77
3117051	56.9	2.6	1.2	36.0	3.4	4.22	0.55	0.34	3.57	0.91
3117052	52.5	3.1	1.0	40.1	3.4	1.62	0.54	0.35	1.59	0.75
3117053	52.2	2.9	1.0	39.7	4.2	1.43	0.42	0.24	1.42	0.74
3117054	48.9	3.9	0.9	40.9	5.4	1.76	0.57	0.26	1.65	1.08
3117055	49.9	3.2	1.1	41.2	4.5	1.44	0.54	0.28	1.49	0.85
3117056	50.8	3.8	1.1	41.0	3.2	1.95	0.67	0.29	1.97	0.66
3137011	50.8	4.1	0.2	43.0	1.8	2.40	0.95	0.16	2.64	0.44
3137031	48.7	7.3	0.3	42.2	1.5	3.52	2.10	0.26	4.68	0.51
3137032	52.2	3.5	0.3	42.3	1.7	2.68	1.02	0.21	2.83	0.48
3137033	51.2	7.1	0.3	40.7	0.8	3.86	1.18	0.22	4.06	0.36
3137034	51.5	5.6	0.5	40.3	2.1	3.38	1.28	0.32	3.58	0.95
3137035	55.7	15.3	0.1	28.0	1.0	2.68	1.90	0.14	3.33	0.47

3. 순환표본조사 표본설계에 대한 시사점

1) 시군구 단위 분석

2015년 인구주택총조사 표본조사(경기)를 분석한 결과에 의하면 시군구 단위의 추정의 정확도는 조사항목에 따라 약간의 차이는 있지만, 공표 가능한 수준으로 판단된다. 특히 2015년 인구주택총조사 표본조사에서 사용된 조사구의 규모가 30가구로 줄어들고 전체적인 추출률도 20%로 높아짐에 따라 추정의 정확도는 상대적으로 크게 높아졌다.

2010년 인구주택총조사 표본조사의 추출률은 10%였고, 이 경우에도 시군구 단위의 통계를 산출했다는 점을 고려하면 2015년 인구주택총조사 표본조사의 시도 및 시군구 단위 통계 산출은 별 문제가 없다고 할 수 있다.

2) 동읍면 단위 분석

2015년 인구주택총조사 표본조사(경기)는 전체 조사구 중에서 20%를 추출하였지만, 동읍면 단위의 추정 정확도가 다소 떨어지는 경우가 나타나고 있다. 이는 현행 동읍면 규모의 편차가 대단히 크고, 이에 따라 표본크기도 동읍면에 따라 차이가 크기 때문에 발생한 현상이다. 결과적으로 전체 동읍면에 대해서 같은 방식으로 통계를 공표하는 것은 현재까지의 분석 결과를 살펴볼 때는 현실적으로 어려울 것으로 판단되며, 규모가 작은 동읍면을 인근의 동읍면과 묶어서 일정 규모의 이상의 동읍면 단위로 통계를 공표하는 방안이 바람직할 것으로 판단된다.

다음 장에서는 순환표본조사를 위한 다양한 조사설계 방안에 대해서 동읍면 단위의 통계작성이 가능할 것인가에 대하여 검토하고자 한다. 20% 추출률로 조사구를 추출하여 조사한다고 할 때, 통계작성이 가능한 동읍면의 규모 기준에 대한 검토가 필요하다.

III. 순환표본조사 표본설계 방안

1. 표본설계 관련 기존 연구 결과

1) 통계청(2015) 한국인구학회 연구

순환표본조사의 목적은 각 조사시점별 안정적인 모집단 전체 또는 대규모 부모집단에 대한 추정결과를 산출함과 동시에 여러 시점에 걸쳐 조사된 자료를 통합하여 상대적으로 규모가 작은 부모집단에 대한 추정을 함께 하는 것으로 볼 수 있다.

인구주택총조사의 표본조사의 예를 들면 전국과 17개 광역 시도 통계의 추정의 매해 이루어지고, 시군구별 통계는 3년 내지 5년 동안 축적된 조사 자료를 토대로 안정적으로 생산할 수 있을 것이다.

순환표본조사를 위한 표본설계의 기본원칙은 다음 두 가지로 요약될 수 있다.

첫째, 각 순환표본 조사주기별로 전국 단위 또는 대규모 부모집단에 대한 안정적인 통계를 산출할 수 있도록 표본설계되어야 한다.

둘째, 여러 시점에서 조사된 자료를 누적하여 상대적으로 소규모인 부모집단들의 안정적인 통계를 산출할 수 있도록 해야 한다.

한국인구학회에서 수행한 연구인 통계청(2015)에서는 순환센서스(순환표본조사)를 도입하고자 할 때 중점적으로 점검해야 할 사항인 모집단 구축과 표본설계, 모수 작성과 추정, 결과 분석 등을 심층적으로 검토하는 것을 목적으로 하였다. 여기서는 통계청(2015) 연구의 주요 내용을 요약한다.

통계청 (2015)에서 제안한 표본설계의 기본방향을 정리하면 다음과 같다.

- 우리나라 순환표본조사의 추출단위는 주소가 아닌 조사구이며, 표집틀 갱신 주기가 5년이기 때문에 순환표본조사를 위한 조사구 표본추출은 매해 혹은 매월이 아닌 5년에 한 번 이루어지는 것이 효율적이다.
- 순환표본조사를 위한 표본추출 방안은 기본적으로 2015년 인구주택총조사 표본조사를 위한 표본설계 방안을 사용하는 것이 바람직할 것이다.
- 작성되는 통계의 공표 수준에 대한 결정은 2015년 인구주택총조사 결과를 분석하여 이루어져야 할 것이며, 현행 조사의 표본크기가 유지되어 적절한 방식으로 분할하여 매해 조사된다면 매해 전국 단위 및 광역시도 단위의 통계산출이 가능할 것으로 예상된다.
- 표본분할의 기본적인 목적은 매해 모집단 전체에 대한 안정적인 통계를 산출하기 위한 대표성이 높은 표본이 매해 조사되도록 하는 것이다. 이는 프랑스에서 사용되는 균형표본추출법과 유사하지만, 계통추출법 등의 확률추출법을 적용해서 표본을 분할할 것을 제안하였다.
- 배분된 표본조사구가 60개 미만인 소규모 동읍면의 비중이 높기 때문에 동읍면 단위에서 균형을 맞춘 표본을 매해 얻는 것은 가능하지 않다. 최소 층 크기를 60개 이상으로 만들기 위하여 단계적으로 층을 결합하는 방안과 시군구 내에서 추출된 조사구를 다시 정렬하고 계통추출법을 통해 표본을 분할하는 방식을 고려할 것을 제안하였다.

2) 미국 ACS 표본설계

미국 American Community Survey(이하 ACS)는 전통적인 센서스에서 시행되던 long-form 설문지 조사를 위한 표본조사보다 효율적인 대안으로 알려져 있다. 이와 같은 조사방식은 우리나라 인구주택총조사의 표본조사 방식에 대한 대체 방식으로 검토할 수 있는 최적의 선택으로 볼 수 있다 (통계청, 2015).

센서스가 대규모 인구집단에 대한 일시적인 스냅샷 정보를 제공하는 반면, 순환센서스는 표본을 분할하여 누적조사를 실시하기 때문에, 소지역에 대한 보다 신뢰성 있는 정보를 제공하기 위해서는 누적된 가구 자료를 결합할 필요가 있다. 센서스 long-form은 2005년에 ACS로 전면 이행하였으며, 2005년 조사결과를 2006년 최초로 공표하였다.

표본크기는 매월 평균 295,000가구를 조사하며, 2011년 6월 기준으로 1년 간 전국적으로 3.54백만 표본가구를 조사하였다. 각각의 카운티(County), 교구(Parish),

면(borough)의 총 3,142 지역단위에서 매월 ACS 조사를 수행하게 된다.

ACS의 표본틀로는 MAF(Master Address File)을 사용하는데, 매 6개월마다 우정국의 자료와 조사원들의 조사를 토대로 자료가 갱신된다. 전국적으로 MAF 갱신을 위해 약 100명의 조사원이 참여한다.

ACS는 인구 65,000명 이상인 지역에 대해서는 1년 추정값을 공표하며, 20,000명 이상 65,000명 미만 지역은 3년 추정값, 5년 추정값은 모든 센서스 블록(트랙?)이 공표 가능하다. 센서스 long-form이 ACS로 완전 이행된 2005년에 수집된 자료부터 다음 해에 자료가 공표되었는데, 2006년과 2007년에는 전년도 1년 추정값만 공표되었다. 2008년에는 '07년도 1년 추정값과 함께, '05-'07년 3년 누적치가 최초로 공표되었고, 2010년에는 '05-'09년의 5년 누적치가 최초로 공표되었다. '10년 이후에는 자료가 5년 이상 누적 되어, 매년 1년 추정값과 다년 추정값이 공표되었음. 가장 최근에 공표된 자료는 1년 추정값('13년), 3년 추정값('11년-'13년), 5년 추정값('09-'13년)로, 각각 '14.9.18, '14.10.23, '14.12.14일에 공표되었다.

참고로 ACS 켄터키 주의 추정값 작성 과정을 살펴본다. 켄터키 주는 6개의 Congressional district로 구성되는데, 센서스국은 2000년 ACS 자료에 기초하여 2001년에 Congressional district 수준의 자료를 공표하였다. 센서스국은 최소인구 100,000명을 기준으로 각각의 지역을 PUMAs(public use microdata areas)로 구분하며, 이 지역은 ACS 1년 추정값에 대한 공표가 가능하다(박시내·이주희, 2016).

센서스국은 2005년 ACS 자료에 기초하여 2006년 PUMAs 수준의 자료를 공표하였다. 켄터키 주의 120개의 카운티 중 12개 카운티는 인구수 65,000명 이상으로 1년 추정값 작성이 가능하며, 43개 카운티는 인구수 20,000이상으로 3년 추정값 작성이 가능하며, 65개 카운티는 인구수 20,000명 미만으로 5년 추정값만 작성이 가능하다. 센서스 트랙은 가장 작은 지역이며 평균 약 4,000명의 인구가 거주한다. ACS 자료는 센서스 트랙 수준의 자료 공표는 불가능하다. 단 예외적으로 34개의 ACS 시범 카운티는 신뢰할만한 자료를 생산한다.

3) 프랑스 순환센서스 표본설계

프랑스의 순환센서스는 미국의 ACS가 센서스의 일부인 표본조사를 대체하기 위하여 순환표본조사를 도입한 것과는 달리 전수조사 자체를 대체하고 있다. 프랑스의 순환센서스는 전통적인 센서스의 목적과 일치하며, 특정 시점에 프랑스 내 거주하는 개인 혹은 가구 정보를 취합하여 모집단의 인구 및 가구 분포를 추정하는 것을 주목적으로 한다.

프랑스의 순환센서스는 모집단 전체에 대한 전수조사가 아닌 표본조사를 통해서 센서스 결과를 산출하고 있다. 순환센서스의 주기는 5년이며, 조사는 매해 이루어지고 있다. 즉, 매해 정해진 특정 시점에 일종의 미니 센서스를 시행하는 것으로 볼 수 있다.

프랑스의 영토를 구성하는 가장 중요한 특징은 많은 수의 기초자치단체(municipalities)에 있는데, 36,680개로 대부분은 매우 작은 마을이지만 법률 적용을 위해서는 공식적인 인구수가 필요한 지역들이다. 단지 980개의 기초자치단체만이 인구 만 명 이상이고, 반면에 인구 만 명 미만인 지역은 27,000개에 달하고 있다. 프랑스 각 지역에서 만 명 미만의 기초자치단체들은 5개의 순환 그룹(rotation group)으로 나뉘는데, 이는 1999년 센서스의 통계 특성(statistical characteristics)에 기초하고 있다.

기초자치단체들은 5개의 집단으로 나누어 이들 지역의 거주자는 순환총조사 5년의 기간 중 1회만 조사된다. 반면 거주자 1만 명 이상의 기초자치단체는 매년 거주지의 8%의 거주만 조사된다. 따라서 5년 후에는 40%의 인구가 조사되어, 전체적으로 70%의 인구가 5년 주기 동안 조사된다.

2. 새로운 조사설계 방안

1) 기본원칙 및 통계공표 수준

2015년 인구주택총조사부터 전수 부문은 등록센서스 방식을 도입하여 매년 인구, 가구 및 주택에 대한 기본 사항에 대한 공표가 가능하도록 하였다. 등록센서스를 통해서 공표되지 않는 항목에 대해서는 2015년 인구주택총조사에서는 20% 표본조사를 실시하였으며, 2021년부터는 표본조사를 5년에 한번 실시하는 것이 아닌, 매월 또는 일정한 시점에 표본을 균등 분할하여 실시하는 순환센서스 방식 도입을 검토 중이다(박시내·정남수·전지혜·정진희, 2015).

2021년부터 적용 예정인 순환표본조사 조사설계의 쟁점은 5개년에 걸친 연도별 표본배분 방법과 일 년 동안 몇 차례로 분할하여 조사할 것인가를 결정하는 것이다. 제시한 두 가지 쟁점은 순환표본조사의 목적, 산출된 통계의 활용, 통계 공표 수준과 주기 등을 검토하여 결정할 사안이다.

통계청은 순환표본조사의 목적을 다음과 같이 잠정적으로 결정하였다. 매년 순환

표본조사를 통해서 등록센서스를 보완하는 시도단위 결과를 생산하고, 3개년 평균에 의한 시군구 자료(순환자료)를 생산하여 정책연구와 장기시계열 분석 등을 위한 자료로 제공하는 방안을 고려하고 있다.

참고로 2015년도 인구주택총조사 표본조사는 동읍면 단위의 통계생산을 위해 동읍면별로 표본크기를 산정하되, 안정적인 표본 수 확보가 곤란한 시군구의 경우 동읍면 수준에서 재조정하는 방안을 적용하였다. 앞선 II 장의 분석에서 살펴보았듯이 20% 추출률을 적용하여 표본조사를 진행하였지만, 동읍면의 규모가 큰 편차를 보이고 있는 상황에서 표본오차가 크게 나타나는 동읍면이 있어 일률적으로 전체 동읍면에 대해서 통계를 산출하는 것은 쉽지 않은 상황이다. 이에 따라 규모가 작은 동읍면 묶어서 새로운 통계작성 단위를 구성하는 방안을 제안하였으며, 전체 동읍면 단위의 통계 공표는 등록센서스 결과로 국한하는 방안이 바람직할 것이다.

5개년 순환표본조사 결과를 이용하여 동읍면 단위의 통계를 공표하는 방안은 20% 표본조사를 실시하여 동읍면 단위의 통계를 산출하는 경우와 유사한 통계적 정확도를 갖게 된다. 따라서 전체 행정 동읍면을 통계작성 단위로 고려한다면 추정의 정확도가 낮은 동읍면이 상당수 있을 것이다. 5개년 순환표본조사 결과를 이용하여 동읍면별 통계를 공표하는 방안으로는 지나치게 작은 규모의 동읍면을 묶어서 인근 동읍면과 함께 공표하는 방안이 현실적으로 검토할 수 있을 것이다.

2) 조사설계 방안 검토

(1) 조사 설계 방안

향후 매년 진행될 순환표본조사는 매년 조사를 통해서 등록센서스를 보완하는 시도단위 통계를 생산하고, 3개년 평균에 의한 시군구 자료(순환자료)를 생산하는 것을 기본원칙으로 하고자 한다. 이와 같은 기본원칙을 달성하기 위해서 본 연구에서 고려하고 있는 순환표본조사 방안은 다음과 같다. 제시하는 방안에서 5년 걸쳐 조사되는 표본의 추출률은 20%이다. 향후 각 방안의 장단점과 통계공표수준을 검토하여 최종 방안을 제시하고자 한다.

- 방안 1: 전체 조사구 중 매년 4%씩을 조사하는 방안
(4%, 4%, 4%, 4%, 4%)
- 방안 2: 매년 2%씩 조사, 등록센서스 연도에 10% 조사구를 추가조사 방안
(2%, 2%, 2%, 2%, 2%+10%)
- 방안 3: 매년 2.5%씩 조사, 등록센서스 연도에 7.5% 조사구를 추가조사 방안

(2.5%, 2.5%, 2.5%, 2.5%, 2.5%+7.5%)

방안 2와 3에서 매년 조사는 기본적인 조사항목에 대한 조사로 매년 공표되는 등록센서스 결과를 점검하고 보완하는 차원에서 이루어지는 조사이고, 등록센서스 해에 추가로 이루어지는 10% 또는 7.5% 추출률의 표본조사는 시도 및 시군구 단위의 통계 작성을 목표로 이루어지는 것으로 현행 인구주택총조사의 표본조사에 해당한다고 할 수 있다.

(2) 조사설계별 통계 작성 방안 비교

앞서 제시한 세 가지 조사설계 방안에 대해서 시도 및 시군구별 통계 작성의 통계적 정확도를 살펴보고자 한다. 2021년부터 시행될 예정인 순환표본조사는 시도별 통계는 매년 공표하고, 시군구별 통계는 3개년 순환표본조사 결과를 이용하여 산출하는 것을 기본원칙으로 하였다. 이와 같은 기본원칙에 입각하여 우선 세 가지 방안을 적용할 때 시도별 통계작성이 가능할 것인가에 대한 검토가 필요하다.

세 가지 방안을 적용하여 매년 순환표본조사를 실시한다고 할 때 산출되는 시도별 통계의 통계적 정확도를 살펴본다. 시도별 통계의 통계적 정확도는 앞선 II장에서 제시한 인구주택총조사 20% 표본조사 결과에 기초하여 산출한 것이다.

<표 3-1>은 방안 1, 2, 3의 추출률을 적용할 때 예상되는 주요 조사항목별 추정에 대한 오차의 한계이다. 방안 1, 2, 3을 적용할 때 예상되는 오차의 한계는 표본 조사구 수를 기준으로 다음 식에 따라 근사적으로 산출하였다.

$$\text{오차의 한계}_{new} = \text{오차의 한계}_{15} \times \sqrt{\frac{n_{15}}{n_{new}}}$$

여기서 n_{15} 와 오차의 한계₁₅는 각각 2015년 인구주택총조사 표본조사의 표본 조사구 수와 해당 조사항목 추정에 대한 오차의 한계이다.

방안 1이 상대적으로 다른 방안 2와 3에 비해서 시도별 추정의 정확도가 높게 나타나고 있다. 상대적으로 표본 조사구가 적은 제주도나 대전광역시에서 추정에 대한 오차의 한계가 다소 커지고 있다. 전반적으로 개인 조사항목에 대한 추정의 정확도는 대단히 높은 수준으로 볼 수 있으며, 이 경우는 추출률을 4%, 2%, 2.5% 등으로 낮추어도 시도별 추정의 문제는 없을 것으로 예상되었다. 반면, 가구 조사항목의 경우는 95% 신뢰수준 오차의 한계가 3%를 넘는 경우가 대전광역시와 제주도에 나타나고 있다. 이는 2015년도 인구주택총조사 표본설계를 그대로 두고 추출률만 낮추었을 때 해당 시도의 표본 조사구 수가 많지 않기 때문에 발생한 것이다. 이는 대전광역시나 제주도는 규모가 작은 시도이며, 이에 따라 2015년도 인구주택총조사 표본조사에서 배분되는 표본 조사구 수가 상대적으로 많지 않았기 때문이다.

지금까지의 분석결과를 종합해 볼 때 앞서 제시한 3가지 방안은 매년 가구나 인구 조사항목에 대한 시도별 추정은 가능할 것으로 판단된다. 다만, 순환표본조사를 위한 표본설계에서는 현행 표본설계의 표본 조사구를 단순 비례로 줄이는 것이 아닌 시도별 추정의 정확도를 일정 수준 이상을 유지할 수 있도록 시도별 표본 조사구 수의 최소 기준을 마련하는 방안을 검토할 수 있을 것이다.

〈표 3-1〉 각 방안별 주요 조사항목 추정에 대한 오차의 한계(예상)

(1) 대전광역시

• 가구 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 3,411)		방안1 (682개)	방안2 (341개)	방안3 (426개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
가구 유형	가족	69.7	0.96	2.16	3.05	2.73
	가족+이외	0.1	0.02	0.06	0.08	0.07
	1인 가족	28.7	0.92	2.05	2.90	2.60
	남남 5인 이하	1.5	0.14	0.31	0.44	0.40
	남남 6인 이상	-	-	-	-	-
점유 형태	자가	53.7	1.18	2.64	3.74	3.34
	전세	15.7	0.48	1.08	1.52	1.36
	월세_보증금	24.8	1.06	2.37	3.35	3.00
	월세_무보증금	2.2	0.18	0.40	0.57	0.51
	사글세	0.7	0.13	0.29	0.41	0.36
	무상	3.0	0.27	0.61	0.87	0.78
사용 방수	1	39.7	0.99	2.20	3.12	2.79
	2	31.5	0.62	1.39	1.96	1.76
	3	26.5	0.80	1.79	2.53	2.27
	4	2.1	0.23	0.51	0.72	0.65
	5+	0.1	0.04	0.08	0.11	0.10
	무응답	0.0	0.01	0.01	0.02	0.02

• 개인 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 3,411)		방안1 (682개)	방안2 (341개)	방안3 (426개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
혼인 상태	미혼	31.4	0.32	0.71	1.01	0.90
	유배우자	55.6	0.38	0.85	1.20	1.07
	사별	6.1	0.14	0.32	0.45	0.41
	이혼	5.2	0.15	0.34	0.48	0.43
	무응답	1.6	0.06	0.13	0.19	0.17
경제 활동 상태	일했음	49.3	0.32	0.72	1.01	0.91
	가사 등	2.9	0.09	0.21	0.29	0.26
	잠시 쉬	0.9	0.05	0.10	0.15	0.13
	일하지 않음	44.3	0.32	0.72	1.02	0.91
	무응답	2.7	0.10	0.22	0.32	0.28

(계속)

(2) 제주도

• 가구 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 1,833)		방안1 (367개)	방안2 (183개)	방안3 (229개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
가구 유형	가족	70.8	0.96	2.15	3.04	2.72
	가족+이외	0.2	0.04	0.08	0.11	0.10
	1인 가족	26.4	0.86	1.91	2.71	2.42
	남남 5인 이하	2.6	0.29	0.65	0.92	0.83
	남남 6인 이상	0.0	0.01	0.03	0.04	0.04
점유 형태	자가	58.2	1.37	3.07	4.34	3.88
	전세	3.4	0.44	0.99	1.40	1.25
	월세_보증금	15.9	1.27	2.83	4.01	3.59
	월세_무보증금	3.3	0.32	0.72	1.01	0.91
	사글세	11.4	0.65	1.45	2.05	1.83
	무상	7.8	0.47	1.06	1.49	1.34
사용 방수	1	33.4	1.10	2.47	3.49	3.12
	2	32.8	0.82	1.84	2.61	2.33
	3	30.8	1.09	2.44	3.45	3.08
	4	2.7	0.28	0.64	0.90	0.81
	5+	0.4	0.07	0.16	0.23	0.21
	무응답	0.0	0.02	0.03	0.05	0.04

• 개인 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 3,411)		방안1 (682개)	방안2 (341개)	방안3 (426개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
혼인 상태	미혼	28.2	0.40	0.89	1.26	1.13
	유배우자	55.3	0.54	1.20	1.71	1.52
	사별	8.2	0.24	0.54	0.77	0.69
	이혼	6.4	0.24	0.53	0.75	0.67
	무응답	1.8	0.09	0.21	0.29	0.26
경제 활동 상태	일했음	59.4	0.44	0.97	1.38	1.23
	가사 등	3.2	0.17	0.37	0.53	0.47
	잠시 쉬	0.8	0.07	0.15	0.21	0.19
	일하지 않음	34.9	0.42	0.94	1.33	1.19
	무응답	1.8	0.09	0.21	0.29	0.26

(계속)

(3) 경기도

• 가구 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 25,865)		방안1 (5,173개)	방안2 (2,587개)	방안3 (3,233개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
가구 유형	가족	74.9	0.30	0.67	0.95	0.85
	가족+이외	0.2	0.01	0.01	0.01	0.01
	1인 가족	23.4	0.28	0.35	0.38	0.37
	남남 5인 이하	1.5	0.05	0.05	0.05	0.05
	남남 6인 이상	0.0	0.00	0.00	0.00	0.00
점유 형태	자가	53.0	0.39	0.88	1.24	1.11
	전세	19.9	0.20	0.44	0.62	0.56
	월세_보증금	22.2	0.36	0.80	1.13	1.02
	월세_무보증금	1.8	0.09	0.20	0.28	0.25
	사글세	0.3	0.02	0.05	0.07	0.07
	무상	2.8	0.09	0.21	0.30	0.27
사용 방수	1	32.8	0.33	0.73	1.04	0.93
	2	33.7	0.24	0.54	0.76	0.68
	3	30.6	0.31	0.70	0.99	0.88
	4	2.8	0.11	0.25	0.35	0.32
	5+	0.2	0.02	0.04	0.06	0.05
	무응답	0.0	0.00	0.01	0.01	0.01

• 개인 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 3,411)		방안1 (682개)	방안2 (341개)	방안3 (426개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
혼인 상태	미혼	29.2	0.12	0.27	0.38	0.34
	유배우자	58.3	0.15	0.33	0.47	0.42
	사별	5.8	0.05	0.11	0.16	0.15
	이혼	5.1	0.06	0.13	0.18	0.16
	무응답	1.6	0.02	0.05	0.07	0.06
경제 활동 상태	일했음	52.6	0.12	0.28	0.39	0.35
	가사 등	2.9	0.03	0.07	0.10	0.09
	잠시 쉬	0.9	0.02	0.04	0.06	0.05
	일하지 않음	39.4	0.12	0.27	0.38	0.34
	무응답	4.2	0.06	0.13	0.18	0.16

앞서 제시한 세 가지 방안을 적용한다고 할 때 산출되는 시군구별 통계의 통계적 정확도를 살펴본다. <표 3-2>는 3년 동안의 조사데이터를 통합하여 통계를 작성하는 것을 가정할 때 예상되는 주요 조사항목별 추정에 대한 오차의 한계이다. <표 3-2>의 결과는 앞서 살펴본 수원시 장안구, 의왕시, 가평군에 대한 인구주택총조사 20% 표본조사 결과에 기초하여 산출한 것이다.

대체로 개인 조사항목에 대한 추정의 정확도는 높은 수준으로 볼 수 있으며, 매년

조사에서 추출률을 4%, 2%, 2.5% 등으로 낮추어도 3개년 조사데이터를 누적하여 통계를 산출할 때 추정의 정확도는 대체로 높은 수준으로 예상되었다. 반면, 가구 조사항목의 경우는 상대적으로 오차의 한계가 크게 나타나고 있다. 예를 들어 가평군의 점유형태별 비율 추정의 오차의 한계는 5%를 넘는다. 상대적으로 규모가 작은 시군구에 대해서는 추출률을 좀더 높이는 방안을 고려해야 할 것이다.

〈표 3-2〉 시군구별 주요 조사항목 추정에 대한 오차의 한계(예상)

(1) 수원시 장안구

• 가구 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 538)		방안1 (323개)	방안2 (161개)	방안3 (202개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
가구 유형	가족	74.8	1.32	1.70	2.41	2.15
	가족+이외	0.2	0.06	0.05	0.04	0.05
	1인 가족	23.8	1.29	1.37	1.50	1.46
	남남 5인 이하	1.2	0.26	0.25	0.24	0.25
	남남 6인 이상	0.0	0.00	0.00	0.00	0.00
점유 형태	자가	55.8	1.52	1.96	2.77	2.47
	전세	19.9	0.93	1.20	1.70	1.52
	월세_보증금	20.8	1.29	1.66	2.35	2.10
	월세_무보증금	1.5	0.35	0.45	0.64	0.57
	사글세	0.1	0.05	0.07	0.10	0.09
	무상	1.9	0.52	0.67	0.94	0.84
사용 방수	1	31.7	0.75	0.96	1.36	1.22
	2	33.5	0.67	0.87	1.23	1.10
	3	32.3	0.78	1.00	1.42	1.27
	4	2.4	0.36	0.46	0.66	0.59
	5+	0.1	0.02	0.03	0.04	0.03
	무응답	0.0	0.01	0.01	0.02	0.02

• 개인 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 538)		방안1 (323개)	방안2 (161개)	방안3 (202개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
혼인 상태	미혼	27.1	0.56	0.72	1.02	0.91
	유배우자	48.5	0.46	0.60	0.85	0.76
	사별	4.8	0.26	0.34	0.48	0.43
	이혼	4.8	0.26	0.34	0.48	0.43
	무응답	14.9	0.46	0.59	0.84	0.75
경제 활동 상태	일했음	50.7	0.59	0.76	1.07	0.96
	가사 등	3.1	0.18	0.24	0.33	0.30
	잠시 쉬	1.1	0.10	0.13	0.19	0.17
	일하지 않음	42.3	0.59	0.76	1.08	0.96
	무응답	2.9	0.23	0.29	0.41	0.37

(계속)

(2) 의왕시

• 가구 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 343)		방안1 (206개)	방안2 (103개)	방안3 (129개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
가구 유형	가족	79.8	1.43	1.85	2.62	2.34
	가족+이외	0.2	0.10	0.13	0.18	0.16
	1인 가족	19.0	1.35	1.75	2.47	2.21
	남남 5인 이하	1.1	0.32	0.41	0.58	0.51
	남남 6인 이상		0	0	0	0
점유 형태	자가	53.7	2.14	2.76	3.90	3.49
	전세	22.9	1.39	1.79	2.54	2.27
	월세_보증금	20.4	2.21	2.85	4.03	3.60
	월세_무보증금	0.9	0.35	0.45	0.63	0.56
	사글세	0.2	0.07	0.09	0.13	0.12
	무상	2.0	0.46	0.60	0.84	0.75
사용 방수	1	33.6	0.87	1.12	1.59	1.42
	2	35.9	0.74	0.96	1.36	1.21
	3	28.5	0.91	1.17	1.66	1.48
	4	1.9	0.26	0.34	0.48	0.43
	5+	0.1	0.03	0.04	0.05	0.04
	무응답	0.0	0.01	0.02	0.02	0.02

• 개인 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 3,411)		방안1 (682개)	방안2 (341개)	방안3 (426개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
혼인 상태	미혼	25.9	0.67	0.86	1.22	1.09
	유배우자	51.1	0.59	0.76	1.08	0.96
	사별	4.7	0.26	0.33	0.47	0.42
	이혼	3.8	0.31	0.40	0.56	0.50
	무응답	14.5	0.56	0.73	1.03	0.92
경제 활동 상태	일했음	51.5	0.76	0.98	1.39	1.24
	가사 등	3.3	0.22	0.29	0.41	0.36
	잠시 쉬	1.0	0.12	0.16	0.22	0.20
	일하지 않음	40.1	0.73	0.94	1.33	1.19
	무응답	4.1	0.35	0.46	0.65	0.58

(계속)

(3) 가평군

• 가구 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 234)		방안1 (140개)	방안2 (70개)	방안3 (88개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
가구 유형	가족	66.6	2.12	2.74	3.87	3.45
	가족+이외	0.1	0.08	0.11	0.15	0.13
	1인 가족	31.0	1.92	2.49	3.52	3.14
	남남 5인 이하	2.2	0.63	0.82	1.16	1.03
	남남 6인 이상		0.00	0.00	0.00	0.00
점유 형태	자가	66.6	3.26	4.22	5.96	5.32
	전세	10.9	1.61	2.08	2.94	2.62
	월세_보증금	12.3	2.61	3.37	4.77	4.26
	월세_무보증금	2.7	0.60	0.78	1.10	0.98
	사글세	0.5	0.19	0.24	0.35	0.31
	무상	7.0	1.73	2.23	3.16	2.82
사용 방수	1	24.8	1.43	1.85	2.62	2.34
	2	34.9	1.35	1.74	2.46	2.20
	3	37.2	1.49	1.93	2.72	2.43
	4	2.8	0.28	0.36	0.51	0.45
	5+	0.3	0.07	0.09	0.13	0.12
	무응답			0.00	0.00	0.00

• 개인 조사항목

구분		15년 인총 표본조사(20%) (표본 조사구: 3,411)		방안1 (682개)	방안2 (341개)	방안3 (426개)
		추정값	오차의 한계	오차의 한계	오차의 한계	오차의 한계
혼인 상태	미혼	19.7	1.08	1.40	1.98	1.77
	유배우자	53.0	1.12	1.45	2.04	1.82
	사별	10.2	0.73	0.94	1.33	1.19
	이혼	5.1	0.52	0.67	0.94	0.84
	무응답	12.1	0.96	1.24	1.76	1.57
경제 활동 상태	일했음	51.3	1.28	1.65	2.34	2.09
	가사 등	5.7	0.56	0.72	1.02	0.91
	잠시 쉬	0.3	0.10	0.12	0.17	0.16
	일하지 않음	41.1	1.44	1.87	2.64	2.36
	무응답	1.6	0.24	0.31	0.44	0.39

(3) 조사설계 방안 검토 결과

2021년부터 시행될 예정인 순환표본조사는 시도별 통계는 매년 공표하고, 시군구별 통계는 3개년 순환표본조사 결과를 이용하여 산출하는 것을 기본원칙으로 하였다. 앞서 살펴본 세 가지 방안을 적용하여 매년 순환표본조사를 실시한다고 할 때 예상되는 시도 및 시군구별 통계의 통계적 정확도를 요약하고 시사점을 정리한다.

앞서 제시한 3가지 방안은 매년 가구나 인구 조사항목에 대한 시도별 추정은 가능할 것으로 판단된다. 상대적으로 인구 조사항목에 대한 시도별 추정의 통계적 정확도는 상당히 높은 수준이다. 가구 조사항목에 대한 시도별 추정의 정확도는 부분적으로 다소 떨어지는 경우가 나타나고 있다. 따라서 순환표본조사 표본설계에서는 시도별 추정의 정확도를 일정 수준 이상을 유지할 수 있도록 시도별 표본 조사구수의 최소 기준을 마련하는 방안이 필요할 것이다.

앞서 제시한 3가지 방안에 적용할 때 3년 조사데이터를 누적하여 시군구별 통계를 산출하는 경우 예상되는 주요 조사항목별 추정의 정확도를 살펴보았다. 대체로 개인 조사항목에 대한 시군구별 추정의 정확도는 대체로 높은 수준으로 예상되었다. 반면, 가구 조사항목에 대한 추정의 정확도는 상대적으로 떨어지는 경우가 나타나고 있다. 3년 조사데이터를 누적하여 통계를 작성한다고 해도 상대적으로 규모가 작은 시군구에 대해서는 추출률을 좀 더 높이는 방안을 고려해야 할 것이다.

이를 위해 새로운 표본설계에서는 시군구의 규모에 따라 목표 오차를 만족할 수 있도록 표본크기를 결정하는 방안, 규모가 작은 시군구의 추출률을 상대적으로 높여서 설정하는 방안, Kish 배분법이나 맥배분법(power allocation) 등을 적용하여 규모가 작은 시군구에 상대적으로 더 많은 표본을 배분하는 방안 등을 고려할 수 있을 것이다.

3. 새로운 표본설계 방안

1) 조사설계 방안의 비교

통계청은 순환표본조사의 목적을 잠정적으로 다음과 같이 결정하였다. 매년 순환 표본조사를 통해서 등록센서스를 보완하는 시도 단위 결과를 생산하고, 3개년 평균에 의한 시군구 자료(순환자료)를 생산하여 정책연구와 장기시계열 분석 등을 위한 자료로 제공하는 방안을 고려하고 있다. 이와 같은 현실적인 제약조건 하에서 고려할 수 있는 아래 3가지 방안의 장단점과 통계 공표수준을 검토하여 최종 방안을 제시하고자 한다.

- 방안 1: 전체 조사구 중 매년 4%씩을 조사하는 방안
(4%, 4%, 4%, 4%, 4%)
- 방안 2: 매년 2%씩 조사하고, 등록센서스 연도에 10%를 추가 조사 방안
(2%, 2%, 2%, 2%, 2%+10%)
- 방안 3: 매년 2.5%씩 조사하고, 등록센서스 연도에 7.5%를 추가 조사 방안
(2.5%, 2.5%, 2.5%, 2.5%, 2.5%+7.5%)

방안 1은 순환표본조사의 당초 목적을 이상적으로 달성할 수 있는 방안이다. 매년 시도별 통계를 산출하고, 2년 또는 3년 조사데이터를 누적하여 시군구별 통계를 산출하는 방안이다. 이와 같은 조사설계 방안은 매년 일정한 조사업무가 부여되어 안정적인 예산과 인력 구조를 갖출 수 있을 때 적용 가능한 방안이다. 반면, 이 방안을 통해서 생산되는 시군구별 통계는 2년 또는 3년 조사데이터를 누적하여 생산되기 때문에 통계 활용의 측면에서 시의성의 문제가 지속적으로 제기될 수 있다는 점이 단점으로 지적될 수 있을 것이다.

방안 2와 3에서 매년 조사는 기본적인 조사항목에 대하여 조사하여 이를 기초로 시도별 통계는 매년 생산하고, 시군구별 통계는 3개년 자료를 통합하여 산출하는 것으로 목표로 한다. 또한 매년 공표되는 등록센서스 결과의 점검과 추정값의 보완을 목적으로 사용될 수 있을 것이다. 방안 2와 3에서 센서스 해에 추가로 이루어지는 10% 또는 7.5% 추출률의 표본조사는 시도 및 시군구 단위의 통계 작성을 목표로 이루어지는 현행 인구주택총조사의 표본조사에 해당한다고 할 수 있다. 따라서 방안

2와 3은 거의 같은 장단점을 갖게 된다.

방안 2와 3을 적용하게 되면 앞선 <표 3-2>에서 살펴본 바와 같이 매년 시도별 통계 산출과 3개년 조사데이터 누적을 통해서 시군구별 통계 산출이 가능할 것으로 예상된다. 앞서 <표 3-2>에서 산출한 주요 조사항목별 추정에 대한 오차의 한계는 3개년 조사 데이터를 미국 ACS 방식을 적용하여 합쳐(pooling) 분석하는 것을 가정한 것이다.

방안 2와 3은 매년 시도별 통계 산출 가능하고, 5년마다 시군구별 시의성이 있는 통계 산출이 가능하다는 장점이 있는 반면, 3개년 조사데이터를 누적하여 매년 시군구별 통계 산출이 가능하지만, 5년마다 산출되는 시군구별 통계와 조사기준시점의 차이로 인한 추정결과 괴리가 발생할 수 있다는 단점이 있다.

원칙적으로 보면 방안 2와 3에서 5년마다 상대적으로 대규모 표본조사를 진행하여 시군구별 통계를 공표하는 것과 3개년 조사데이터를 누적하여 시군구별 통계를 작성하는 것은 조사기준시점이 다른 두 통계가 함께 공표될 수 있기 때문에 양립하기 어려운 구조이다. 따라서 5년마다 대규모 표본조사(예를 들어 10% 조사)를 진행하고자 한다면 two-track으로 접근하는 것이 타당하다고 판단된다. 여기서 two-track이라는 의미는 매년 이루어지는 순환표본조사와 5년마다 이루어지는 표본조사의 조사목적과 조사항목을 달리하여 두 종류의 표본조사로 간주한다는 뜻이다. 참고로 다음 표는 순환표본조사는 매년 2%씩 진행하고, 기존 센서스 연도인 0이나 5로 끝나는 해에는 10% 표본조사를 진행하는 방안을 제시하고 있다.

구분	2021	2022	2023	2024	2025	2026	2027	...
순환표본조사	2%	2%	2%	2%	2%	2%	2%	...
표본조사					10%			...
시도 통계(1년)	○	○	○	○	○	○	○	
시군구 통계(3년)			○	○	○	○	○	
시군구 통계(1년)					○			

two-track으로 순환표본조사를 진행하는 방안은 인구주택총조사 조사항목을 세 유형으로 구분하여 공표하는 방안으로 볼 수 있다. 즉, 매년 공표되는 등록센서스 결과, 매년 2%(또는 2.5%) 표본조사를 통한 시도 단위의 통계, 5년마다 얻게 되는 시도 및 시군구 단위의 통계(10% 표본) 등이다. 여기서 매년 이루어지는 2%(또는 2.5%) 표본조사 결과는 등록센서스 결과를 점검하고 보정하기 위한 기초자

료로 활용하게 된다.

원칙적으로 매년 2% 표본조사 결과를 통해서 시도별 통계작성이 가능할 것으로 예상되지만, 이 경우도 연도별 시계열 흐름과 5년째 조사에서 얻은 추정결과의 일치성 여부가 중요한 쟁점이 될 수 있을 것이다. 이 점을 고려하면 매년 조사는 등록센서스 결과를 점검하고 보정하는 목적으로 이용하는 것이 바람직할 것으로 판단되며, 매 5년마다 10% 표본을 조사하여 시도 및 시군구 단위의 통계를 생산하는 방안이 현실적일 것으로 판단된다.

방안 2, 3을 적용하고자 할 때 표본추출 방안은 다음과 같다. 모집단을 5개 그룹으로 분할한 경우라면 매년 각 그룹에서 순차적으로 2%(또는 2.5%) 표본을 추출하여 조사하고, 5년째 되는 해에는 다섯 번째 그룹에서 4%(또는 5%) 표본을 추출하고, 나머지 각 그룹에서는 2%(2.5%) 표본을 추출하여 조사하는 방안을 고려할 수 있을 것이다. 다섯 번째 그룹에서 뽑힌 표본의 절반은 매년 실시되고 있는 표본조사 목적으로 이용하고, 나머지 절반은 5년 주기 표본조사에 활용하는 방안이다. 표본추출을 위한 다른 방안으로 0이나 5로 끝나지 않는 해에는 모집단을 4개로 분할하여 각 그룹에서 매년 2.5% 표본을 추출하여 시도별 추정과 등록센서스 결과 점검 및 보정의 목적으로 활용하고, 0이나 5로 끝나는 해에는 10% 표본을 추출하여 조사하는 방안을 적용할 수도 있을 것이다.

조사목적과 조사항목을 달리해서 two-track으로 접근하는 방식을 적용한다면 우선적으로 방안 2(2%, 2%, 2%, 2%, 2%+10%)를 고려할 수 있을 것이다. 여기서 매년 2%씩 이루어지는 순환표본조사는 등록센서스의 기준시점과 일치하여 조사하며, 조사 결과는 등록센서스 결과를 보정하고 점검하기 위한 목적으로 사용하는 방안이다. 5년마다 이루어지는 10% 표본조사를 통해서는 시도 및 시군구별 통계를 산출하는 하도록 하는 것이 바람직할 것이다.

앞서 제시한 3가지 방안 중에서 어느 방안을 선택할 것인가는 순환표본조사의 목적을 어디에 둘 것인가에 따라 결정될 사안이다. 순환표본조사의 기본원리에 충실하고자 한다면 매년 4%씩 조사하는 방안 1로 결정하는 것이 바람직할 것이다. 반면 방안 2나 3은 순환표본조사와 현행 인구주택총조사의 하이브리드 방식으로 조사목적은 달리하는 두 개의 조사로 진행한다는 의미이다. 이 중 순환표본조사는 매년 등록센서스와 기준시점을 동일하게 하여 진행함으로써 등록센서스 결과를 보정하고 점검하기 위한 목적으로 진행하며, 필요한 경우에는 2년 또는 3년 조사데이터를 누적하여 분석에 활용할 수도 있을 것이다. 5년마다 진행되는 표본조사는 현행 인구주택총조사의 표본조사와 마찬가지로 시도 및 시군구별 통계 작성을 목적으로 진행되는 것이다.

2) 월간 및 연간 조사의 장단점

통계청은 순환표본조사의 추출률이 결정되면 해당 연도의 조사방법으로는 해당 연도의 매월에 걸쳐 일정 비율을 조사하는 방안(월간 조사)과 등록센서스 조사기 준수점과 일치시켜 연 1회로 조사하는 방안(연간 조사)을 고려할 수 있을 것이다. 이 중 어느 방안을 선택할 것인가의 문제는 실사 인력의 운영, 생산되는 통계의 활용 목적에 따라 결정할 사안이다.

먼저 실사 인력의 운영적 측면을 살펴보면 월간 조사는 실사인력을 상시 고용하여 조사를 진행하게 되어 체계적인 조사가 가능하며, 이를 통해서 조사의 질을 안정적으로 유지하는 데 기여할 수 있다. 반면 연간 조사의 경우는 실사인력의 상시 고용이 어려운 상황에서도 적은 실사 예산으로 조사가 가능하게 된다는 장점이 있다.

다음으로 통계 활용 측면을 고려하면 월간 조사는 특정 조사시점이 아닌 1년 동안 연중 조사하게 되어 조사기간 동안의 기간 평균의 개념으로 해석할 수 있다는 장점이 있다. 또한 계절적 변동이 있는 조사항목의 경우는 월간 조사를 통해서 기간 평균의 의미가 좀더 분명하게 될 것이라는 장점이 있다. 반면 연간 조사는 특정 조사기준 시점에서 조사하는 방안으로 10% 표본조사의 조사시점과 동일하게 할 수 있어 표본조사를 통해서 생산되는 통계의 비교성이 높아지는 장점이 있다. 특히 매년 진행되는 표본조사 결과를 등록센서스 결과를 점검하고 확인할 목적으로 활용하고자 할 때는 등록센서스의 기준시점과 맞추어 연간조사로 진행하는 것이 바람직할 것이다.

3) 표본설계 방안: 2%(또는 2.5%, 4%) 조사

매년 표본 추출률을 2%(또는 2.5%, 4%)로 하는 순환표본조사 조사설계는 매년 시도단위 결과를 생산하고, 3개년 평균에 의한 시군구 자료(순환자료)를 생산하도록 하는 것을 목표로 한다.

하지만 우리나라의 시군구는 가구 수 또는 인구의 측면에서 대단히 큰 편차를 보이고 있다. 따라서 시군구의 규모(가구 또는 인구 기준)에 따라 추정의 정확도에 차등을 둘 필요도 있을 것이다. 전체 시군구를 가구 수나 인구 규모에 따라 3개 또는 4개 그룹으로 구분하고, 3년 조사데이터를 누적하여 분석한다고 가정할 때 목표 오차의 한계에 차등을 두어 표본크기를 결정할 수 있을 것이다.

다른 방안으로는 전국의 표본크기를 결정한 후 각 시군구에 대해서 Kish 배분이나 멱배분(power allocation)을 적용하여 시군구별 표본크기를 배분할 수 있을 것이다. 이후 시군구별 통계의 예상 오차의 한계를 산출하여 목표 오차의 한계를 만족할 수 있도록 조정한다. 이후 시도별 추정의 정확도(예상)를 계산하여 일정 수준 이상의 통계적 정확도를 만족할 수 있도록 표본크기를 조정하여 최종 시군구별 표본크기를 결정한다.

이와 같이 정해진 시군구의 표본크기에 따라 매년 해당 시군구의 5개 부모집단 중에서 선택된 하나의 부모집단에 대해서 아파트조사구와 보통조사구로 층화하여 비례배분한다. 이후 해당 부차모집단의 각 층에서 정해진 정렬기준에 따라 정렬한 후 계통추출법을 적용하여 표본조사구를 추출하는 방안을 고려할 수 있다.

4) 표본설계 방안: 10% 조사(방안 2 예시)

매 5년 간격의 인구주택총조사 연도에 10% 표본조사를 통해서 등록센서스 통계 생산과 함께 인구주택총조사가 시행되는 기준 연도의 시도 및 시군구 단위 세부 통계 생산을 목표로 한다. 이 방안은 기존의 2010년도 인구주택총조사 표본조사(10%)의 표본설계와 유사한 방법을 적용할 수 있을 것이다.

현행 행정구역상의 시군구는 가구 수 또는 인구의 측면에서 대단히 큰 편차를 보이고 있다. 따라서 앞서와 마찬가지로 시군구의 규모(가구 또는 인구 기준)에 따라 추정의 정확도에 차등을 둘 필요도 있을 것이다. 전체 시군구를 규모에 따라 3개 또는 4개 그룹으로 구분하고, 오차의 한계(목표)에 차이를 두어 접근할 수 있을 것이다. 이후 전체 추출률이 약 10%를 만족할 수 있도록 조정하여 시군구별 최종 표본크기를 결정한다.

다른 방안으로는 추출률 10% 기준으로 전국의 표본크기를 결정한 후 각 시군구에 대해서 Kish 배분이나 멱배분(power allocation)을 적용하여 시군구별 표본크기를 배분할 수 있을 것이다. 이후 시군구별 통계의 예상 오차의 한계를 산출하여 목표 오차의 한계를 만족할 수 있도록 조정함으로써 각 시군구별 최종 표본크기를 결정한다.

인구주택총조사 10% 표본조사는 전체 모집단을 기준으로 각 시군구에서 아파트조사구와 보통조사구로 층화하여 비례배분한다. 이후 각 층에서 표본조사를 정해진 정렬기준에 따라 정렬한 후 계통추출법을 적용하여 표본조사구를 추출한다.

5) 제3차 시험조사 방안

순환표본조사를 위한 제 3차 시험조사는 전체 순환표본조사 과정의 모의시험적인 성격을 갖게 되며, 2018년부터 3개년에 걸쳐 진행될 예정이다. 따라서 제3차 시험조사는 매년 2%(2.5%) 표본을 추출하여 3개년 조사 자료를 누적하여 생산하는 시군구별 통계의 추정 정확도를 평가하고, 매년 등록센서스 결과를 점검하고 보정하는 데 2% 표본조사 결과를 어떻게 활용할 것인가를 검토하기 위한 목적으로 활용해야 할 것이다. 앞선 <표 3-1>의 결과는 2%(2.5%) 표본조사를 통해서 매년 시도별 통계 작성은 별 문제가 없다는 점을 제시하고 있다. 따라서 제3차 시험조사는 순환표본조사의 실사 과정에 대한 점검과 더불어 3개년 누적자료를 이용하여 시군구별 안정적인 통계 산출이 가능할 것인가에 주안점을 맞추어야 할 것이다.

앞서 기술한 바와 같이 우리나라의 시군구는 가구 수 또는 인구의 측면에서 대단히 큰 편차를 보이고 있다. 실사 과정에 대한 점검과 3개년 누적자료를 이용한 통계 작성의 가능성 등을 검토하기 위해서는 다양한 규모의 시군구를 대상으로 조사를 진행하는 것이 바람직할 것이다. 전체 조사대상 시군구의 개수는 시험조사의 총예산에 따라 결정되겠지만, 가능하면 전체 시군구를 규모에 따라 3개 또는 4개 그룹으로 구분하여 각 그룹에서 1-2개의 시군구를 조사대상 시군구로 정하는 것이 바람직할 것이다.

제3차 시험조사는 실사 과정과 시군구별 통계 산출에 대한 점검뿐만 아니라 표본설계 방법에 대한 검토도 함께 이루어지도록 하는 것이 바람직하다. 따라서 앞서 제시한 2%(또는 2.5%, 4%) 표본조사에 대한 표본설계 기본원칙을 적용하여 전국 시군구에 대한 전체적인 표본설계를 진행한 후 각 규모 그룹에서 1-2개의 시군구를 랜덤추출하거나 임의로 지정하여 조사대상 시군구로 정하여 제3차 시험조사를 진행하는 것이 필요할 것이다.

IV. 모집단 분할

1. 모집단 분할

1) 모집단 분할 개요

순환표본조사를 도입하는 경우 5년을 하나의 순환주기로 설정하며, 연간 표본추출을 위해 전체 가구 모집단을 5개의 균형을 이루는 그룹(balanced group)으로 분할하여 표본추출틀(sampling frame)을 갱신하고 관리하는 동시에 매년 하나의 분할 표본추출틀을 사용함으로써 5년 한 주기 동안 표본의 중복을 피할 수 있다. 순환표본조사에서 순환주기는 통계 작성 대상 지역이나 작성 주기 등에 따라 2년이나 5년 또는 10년 주기로 가져갈 수도 있지만, 기존 우리나라 인구주택총조사가 5년 주기로 실시되었던 점을 감안해 5년을 주기로 순환표본조사를 시행하는 것을 전제로 모집단 분할 문제를 다루고자 한다.

순환센서스 또는 순환표본조사를 수행하는 대표적인 외국 사례를 보면 프랑스의 경우 전수조사 자체를 대체하는 목적으로 순환센서스(rolling census)를 도입하고 있으며, 전체 지역을 5년 동안 한번은 조사하는 것을 원칙으로 자치구역(municipality)을 기준으로 모집단을 5개 그룹으로 나누어 조사하고 있다.

대표적인 순환표본조사(rolling sample survey)에 해당하는 미국 지역사회조사(ACS; American Community Survey)의 경우, 가구들은 균등하게 나뉜 5개의 하위 프레임(sub-frame) 중 하나에 속하며 5년 마다 돌아가면서 조사되기 때문에, 매년 전체를 대표할 수 있는 표집을 할 수 있으면서 동시에 한번 조사된 가구는 향후 5년간 표집 대상에서 제외되는 서베이 휴식 프로그램(Survey Holiday Program)을 적용하고 있다(통계청, 2015). 미국 지역사회조사(ACS)에서는 모든 일반 주거단위(Housing Unit)와 요양원, 기숙사 등을 포함하는 집단거주시설(Group Quarters)에 대한 주소 정보를 포함하는 마스터주소파일(MAF: Master Address File)을 표본추출틀로 사용해 직접 가구를 표본으로 추출한다. 자료수집과정은 우선 1개월 동안 표본 가구가 온라인과 메일을 통해 조사에 참여하도록 유도한 후, 응답하지 않은 가구에 대해서는 순차적으로 전화조사와 방문조사를 통해 자료를 수집하는 방식으로 조사가 이루어진다.

스위스의 경우 등록 자료의 활용과 함께 센서스 표본조사를 실시하고 있으며 표본조사 부문은 순환조사 방식으로 등록조사, 구조조사, 주제중심조사, 유니버스조사 등 4개 조사로 구성되어 있다. 이 밖에도 등록센서스와 함께 표본조사를 실시하는 국가는 이스라엘, 대만, 독일, 스페인, 폴란드, 싱가포르, 터키, 리히텐슈타인 등이 있다(박은영·김의영, 2014).

이론적으로 보면 순환주기 5년 동안 동일 가구가 중복해 표본으로 추출되지 않도록 하는 방안으로는 모집단을 5개 그룹으로 분할해 관리하는 방식, 5년 한 주기 전체 표본을 추출한 후 표본을 균등 분할해서 연간 표본을 구성하는 방식, 매년 전체 표본추출틀에서 표본을 추출하데 기 추출된 가구(또는 조사구)는 다른 년도 표본에 중복 포함되지 않도록 제어하는 기법(영구번호 사용 등)을 통해 표본을 추출하는 방식 등이 있을 수 있다. 본 연구에서는 표본의 대표성 확보에 문제가 없도록 전국 모집단 표본추출틀을 5개 그룹으로 균형을 이루도록 분할하는 방안을 구현하는 데 초점을 두기로 한다.

전국을 중복되지 않는 5개의 그룹(또는 부차모집단)으로 분할하고 각 연도에는 각 분할 부차모집단에서 표본을 추출하는 방식으로 연간 표본을 추출하는 경우, 미국지역사회조사(ACS)처럼 마스터주소파일(MAF: Master Address File)을 표본추출틀로 사용해 모든 가구를 5개의 부차 추출틀(sub-frame)으로 나누는 방법이 있다. 이런 방식의 모집단 분할을 위해서는 모든 가구에 대한 주소가 누락이나 중복 없이 파악될 수 있는 주소 명부 DB를 구축하는 것이 필수적이며, 또한 우편이나 온라인 조사를 먼저 수행한 후에 방문조사를 함으로써 방문조사 실시에 따른 비용을 절감하는 자료수집 과정이 전제되어야 한다.

우리나라의 경우 아직 미국의 마스터주소파일과 동일한 수준의 포괄성(coverage)을 만족하는 주소 명부를 구축하지 못하고 있다는 점과 주소 명부를 이용해 가구를 직접 추출단위로 사용하는 동시에 우편조사와 방문조사를 단계별로 병합 조사하는 방식의 도입 가능성에 대한 검토가 충분히 이루어지지 않았다는 점 등을 고려할 때, 순환표본조사를 위해 주소를 기준으로 모집단을 분할하는 것은 현재 우리나라 상황에서는 불가능할 것으로 판단된다.

가구 기준으로 전체 가구 모집단을 분할하는 경우 각각의 부차 모집단(sub-population)이 균형을 이루게 한다는 측면에서 가장 효율적일 수 있지만, 현실적으로 가구를 최소단위로 사용하는 분할이 불가능하다면 대안으로는 조사구, 기초단위구, 국가기초구역 등을 기준으로 모집단을 분할하는 방안을 고려해야 할 것이다. 참고로 2015년 인구주택총조사 기준으로 조사구는 693,510개, 기초단위구는 350,000여개, 국가기초구역 35,000여개인 것으로 개략적으로 파악된다.

순환표본조사를 위한 모집단 분할에서는 두 가지 측면이 고려되어야 할 것이다.

첫째, 연도별 통계 산출이 안정적으로 이루어지기 위해서는 모집단을 5개 그룹으로 분할한 후 각 그룹이 균형을 이루는 것이 중요하다. 여기서 균형을 이룬다는 개념은 주택, 가구, 인구 등과 관련된 다양한 보조변수 관점에서 각 분할에서 얻어지는 통계가 차이가 나지 않는다는 것을 의미한다.

둘째, 매년 등록센서스 자료 연계 및 현장 확인을 통해 해당 연도의 부차 표본추출틀을 업데이트하고 유지 관리하는 작업이 원활하게 이루어질 수 있어야 한다. 이를 위해서는 향후 신규 주택이나 아파트 등이 건설되는 경우 해당 거처를 어떤 방식으로 5개로 분할된 추출틀에 균등 배분할 것인지, 또한 재개발 등으로 인해 기존 조사구에서 대규모 변동이 생기는 경우 이를 어떤 방식으로 5개 부차추출틀에 반영할 것인지 등에 대한 처리방안이 세밀하게 수립될 필요가 있다.

통계청이 확보하고 있는 표집틀 현황과 가급적 균형을 이루는 모집단 분할이라는 목적을 달성하기 위해서는 현재 통계청에서 관리하고 있는 인구주택총조사 조사구를 최소 분할 단위로 사용하는 것이 현실적으로 가능한 선택이라고 판단된다. 따라서 여기서는 조사구를 최소 분할 단위로 사용할 때 전체 가구 모집단을 5개의 균형을 이루는 그룹(balanced groups)으로 분할하는 방법을 검토해 보기로 한다.

2) 조사구 기준 모집단 분할 방법

전체 가구 모집단을 조사구를 기준으로 5개 그룹으로 균형이 되도록 분할하기 위해서는 먼저 동읍면 내의 조사구를 5개 그룹으로 나누고, 각 동읍면에서 구성된 그룹들을 묶어서 시군구 내 5개 그룹을 구성하는 방식으로 모집단을 분할하는 것이 효과적이다. 이를 위해서는 1차적으로 시군구 내 동읍면을 적절한 보조정보를 기준으로 정렬하고, 2차적으로 동읍면 내 조사구를 다시 적절한 보조변수를 사용해 정렬한 후 추출간격을 5로 하는 계통추출 과정을 통해 조사구를 5개로 분할하면 된다.

여기서는 2015년 인구주택총조사 조사구에 등록센서스 자료를 매칭시켜 작성한 조사구별 자료를 활용해 동읍면 내 조사구를 특정 기준으로 정렬한 후 계통추출하는 방식으로 동읍면 내 조사구를 5개 그룹으로 분할하는 방법을 검토하였다.

순환센서스에서는 기본적으로 주택, 가구 및 인구 관련 통계를 작성하는 것을 목적으로 하기 때문에 정렬방식의 선택에 있어서도 이런 핵심 변수를 사용하는 것이 바람직할 것이다. 따라서 본 연구에서는 우선 동읍면 정렬에 있어서는 지리적인 인접성을 고려한다는 의미에서 동읍면 코드를 이용한 정렬방식과 해당 지역의 규모를 나타내는 인구수를 정렬기준으로 사용하는 방법을 살펴보기로 한다. 조사구 정렬의

경우에는 조사구 코드, 해당 조사구 인구수와 함께 지역 특성을 나타내는 변수로 볼 수 있는 전체 인구 중 50대 이상 고령자 인구 비율을 사용하는 방안을 검토하기로 한다. 이런 관점에서 다음과 같은 6가지 정렬방식을 비교하였다.

- [정렬 A] : 동읍면 코드 정렬 후 조사구코드 정렬
- [정렬 B] : 동읍면 코드 정렬 후 인구수 정렬
- [정렬 C] : 동읍면 코드 정렬 후 고령비율(50세 이상 비율) 정렬
- [정렬 D] : 동읍면 인구수 정렬 후 조사구코드 정렬
- [정렬 E] : 동읍면 인구수 정렬 후 인구수 정렬
- [정렬 F] : 동읍면 인구수 정렬 후 고령비율 정렬

다양한 방법으로 구축된 모집단 분할이 어느 정도 수준의 균형을 만족하는지 평가하기 위해서는 2015년 등록센서스 자료와 전체 조사구 자료를 연계해 새로운 DB를 구축하는 작업이 선행되어야 한다. 하지만 아직 2015년 인구주택총조사의 등록센서스 자료와 20% 표본조사 조사구 자료 정리가 완료된 상태가 아니기 때문에 여기서는 제주도와 대전광역시에 대한 분석 결과만을 살펴보기로 한다. 향후 2015년 인구주택총조사 자료 정리가 완료되고 등록센서스 자료와 조사구 자료를 가구단위로 연계하는 작업이 완료되면 전체 17개 광역시·도에 대한 분석이 필요할 것이다.

본 연구에서 분할방법 비교 분석을 위해 사용한 등록센서스와 인구주택총조사 조사구 연계자료의 경우 매칭작업이 완벽하게 수행되지 않은 자료를 불가피하게 사용했기 때문에 등록센서스와 인총자료가 매칭되지 않은 일부 조사구를 제외하고 분석하였다. 분할방법 비교 분석을 위해 사용한 조사구수 현황은 <표 4-1>과 같다. 제주도의 경우 2015년 인구주택총조사의 8,205개 조사구 중 자료 매칭이 되지 않은 조사구 111개를 제외하고 8,094개 조사구를 분석에 사용하였고, 대전광역시는 20,886개 조사구 중 매칭이 되지 않은 조사구 401개를 제외한 20,485개 조사구를 분석에 사용하였다.

<표 4-1> 제주도와 대전광역시 등록센서스 조사구 매칭 현황

	지역	인총 조사구수	등록센서스 매칭 조사구수
제주특별자치도	전체	8,205	8,094
	서귀포시	2,042	1,986
	제주시	6,163	6,108
대전광역시	전체	20,886	20,485
	대덕구	2,609	2,586
	동구	3,653	3,615
	서구	6,313	6,251
	유성구	4,749	4,539
	중구	3,562	3,494

2. 도 단위 분할방법 비교

제주도와 대전광역시의 등록센서스 자료를 이용해 인구주택총조사 조사구 단위로 집계해 작성한 자료를 토대로 6가지 동읍면 및 조사구 정렬방식으로 구성된 분할이 얼마나 균형을 이루도록 구성되는지 확인해 보기로 한다. 여기서는 등록센서스 주요 변수를 기준으로 전체 자료에서 얻어진 총계와 5개의 분할된 그룹의 자료에서 산출되는 추정값이 얼마나 일치하는지를 살펴봄으로써 해당 분할 방법을 통해 구성된 그룹이 얼마나 균형을 이루고 있는지를 판단하기로 한다.

분할 방식(실제로는 조사구 정렬방식)의 효율성을 비교하기 위해 각 정렬방식을 통해 구축된 그룹별로 인구, 가구 및 주택 부문 주요 변수에 대한 광역시·도별 또는 시군구별 통계를 구한 후 전체 모집단에서 작성된 총계와 얼마나 차이가 발생하는지 살펴봄으로써 균등 분할 여부를 판단하였다.

모집단 분할 방식에 따른 차이를 요약된 형태로 살펴보기 위해 각 분할 방식을 통해 구축된 5개 그룹에서 얻은 모집단 총계 추정값들과 전체 모집단에서의 총계의 차이의 절대값을 기준으로 산출된 평균상대오차와 분할 중 오차가 가장 크게 발생한 경우의 상대오차를 나타내는 최대상대오차를 살펴보기로 한다. 각 변수별로 평균상대오차와 최대상대오차는 각 정렬방법으로 구성된 5개 분할($i = 1, 2, 3, 4, 5$)에서 얻어진 추정값(\hat{Y}_i)과 전체 모집단에서 계산된 총계(Y)와의 절대오차를 기준으로 다음과 같이 계산된 것이다.

- 평균상대오차 : $\frac{1}{5} \sum_{i=1}^5 \frac{|\hat{Y}_i - Y|}{Y} \times 100\%$
- 최대상대오차 : $Max\left[\frac{|\hat{Y}_i - Y|}{Y} \times 100\%, i = 1, 2, 3, 4, 5\right]$

1) 주택 관련 변수 기준 도 단위 분할방법 비교

주택부문의 경우 각 정렬방법으로 구성된 5개 분할에서 광역시·도별 전체 주택수, 단독 주택수 및 아파트 수 추정에 대한 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-2>과 같다.

전체 주택수의 경우 제주도에서는 [정렬 A]의 평균상대오차가 가장 작은 것으로

나타났으며 최대상대오차가 가장 작은 [정렬 F]에 비해 효율이 크게 떨어지지 않는 것으로 나타났다. 대전광역시에서는 최대상대오차와 평균상대오차 모두 주택수 추정에서 [정렬 A]가 가장 효과적인 것으로 나타났다. 결과적으로 제주도와 대전광역시 모두 전체 주택수를 기준으로 하는 경우 동읍명과 조사구 모두 지역적 인접성을 고려한 동읍면과 조사구 코드를 기준으로 한 [정렬 A] 방식을 이용한 분할 방법이 대체적으로 효과적인 것으로 나타났다.

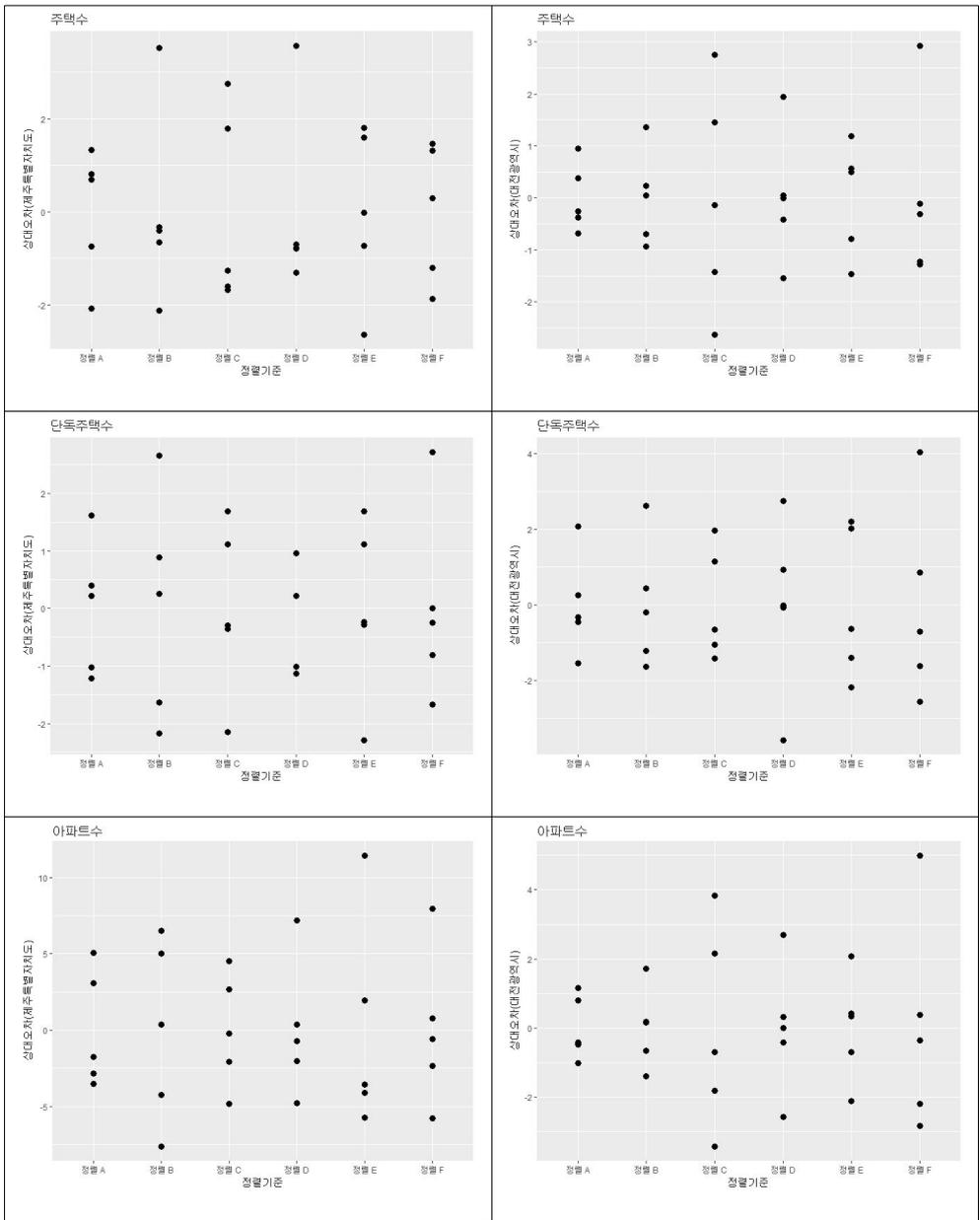
한편 단독주택의 경우, 제주도에서는 [정렬 D]가 대전광역시에서는 [정렬 A]가 대체적으로 효과적인 것으로 나타났다. 아파트 수의 제주도에서는 [정렬 C]가 대전광역시의 경우 [정렬 A]가 효과적인 것으로 보인다.

도 단위에서 주택부문 통계를 기준으로 정렬방법에 따른 비교 결과를 종합해 보면, 우선 대전광역시의 경우 전체 주택수, 단독주택, 아파트 모든 변수에 대해서 [정렬 A]가 효과적인 것으로 나타난 반면 제주도의 경우 변수에 따라 효과적인 정렬방법이 다른 것으로 나타났고, 특히 제주도의 경우 아파트 수에 대한 오차가 상대적으로 상당히 크게 발생한다는 특징을 보이고 있다.

한편 <그림 4-1>는 정렬방법에 따라 구성된 5개 분할에서 발생한 상대오차를 산점도 형식으로 나타낸 것이다. <그림 4-1> 맨 위 오른쪽 산점도를 보면 대전광역시 주택수의 경우 [정렬 A]를 기준으로 한 분할의 경우 5개 모든 그룹에서 상당히 안정적으로 전체 모집단 통계를 산출할 수 있다는 것을 확인할 수 있다.

〈표 4-2〉 정렬방식에 따른 도단위 주택부분 추정오차 비교

변수		총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]	
			최대 상대오차	평균 상대오차										
주택수	제주	195176	2.08%	1.13%	3.52%	1.41%	2.75%	1.82%	3.57%	1.43%	2.64%	1.36%	1.87%	1.23%
	대전	468492	0.95%	0.53%	1.36%	0.65%	2.75%	1.68%	1.94%	0.79%	1.46%	0.90%	2.93%	1.17%
단독 주택수	제주	83584	1.62%	0.90%	2.66%	1.53%	2.15%	1.12%	1.14%	0.86%	2.30%	1.13%	2.72%	1.09%
	대전	81285	2.07%	0.93%	2.61%	1.22%	1.96%	1.24%	3.58%	1.47%	2.20%	1.69%	4.04%	1.95%
아파트수	제주	62532	5.05%	3.25%	7.65%	4.76%	4.86%	2.87%	7.20%	3.03%	11.45%	5.36%	7.97%	3.49%
	대전	337890	1.15%	0.78%	1.71%	0.82%	3.82%	2.39%	2.69%	1.20%	2.12%	1.13%	4.99%	2.15%



[그림 4-1] 정렬방식에 따른 5개 그룹별 도단위 주택부문 추정오차 산점도

2) 가구 관련 변수 기준 도 단위 분할방법 비교

가구부문의 경우 각 정렬방법으로 구성된 5개 분할에서 제주도와 대전광역시에 대한 전체 가구수 및 1인, 2인, 3인 가구수 추정에 대한 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-3>과 같다.

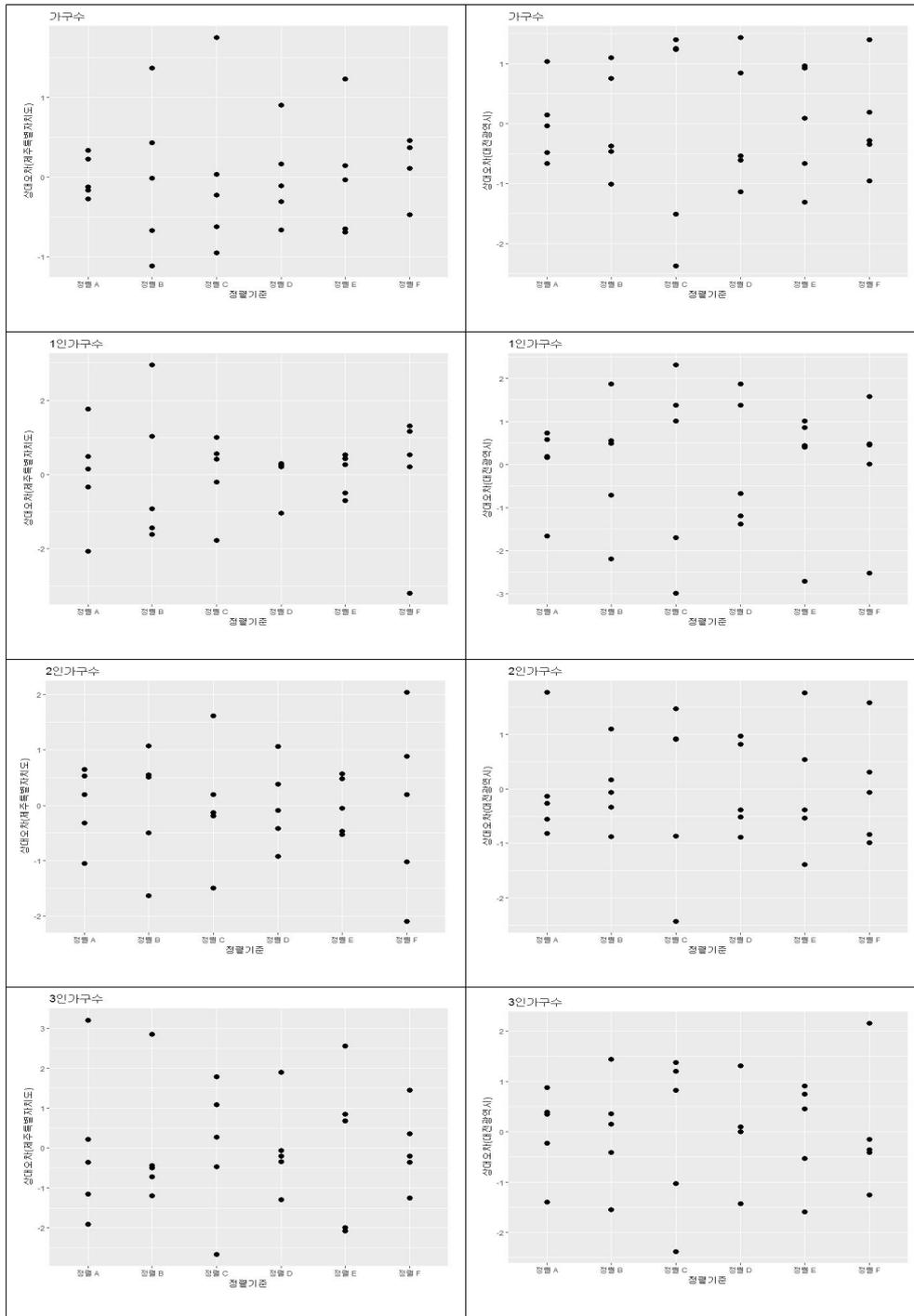
먼저 가구 부문 변수들의 경우 <표 4-3>의 주택부문에 비해 전반적으로 상대오차가 적게 발생하고 있다는 것을 알 수 있다. 대전광역시의 경우 전체 가구수 및 1인, 2인, 3인 가구수 추정에 있어서 모두 [정렬 A] 방식에 의한 분할이 가장 효과적인 것으로 나타났다.

한편 제주도의 경우 전체 가구수에 있어서는 [정렬 A]가 가장 오차가 적지만 1인 가구수의 경우 동읍면을 코드 대신 인구수로 정렬한 [정렬 D] 또는 [정렬 E]가 효과적인 것으로 나타났다. 2인 가구수의 경우에는 [정렬 E]가 3인 가구수의 경우에는 [정렬 F]가 가장 오차가 적어지는 분할방법인 것으로 나타났다. 결국 가구 부문의 경우에도 대전광역시의 경우에 모든 변수에 대해 [정렬 A]가 효과적인 분할방법인 것으로 나타났지만, 제주의 경우에는 어떤 변수를 기준으로 하는지에 따라 효과적인 분할방법 다시 말해 효과적인 정렬방식이 달라짐을 볼 수 있다.

가구 부문이 경우 전체 가구수, 1인 가구수, 2인 가구수 및 3인 가구수를 정렬방식에 따라 구성된 분할에 따라 오차가 어느 정도 발생하는지 살펴보기 위해 5개 분할별로 발생한 상대오차를 정리하면 <그림 4-2>와 같다. 첫 번째 행 첫 번째 열에 있는 제주도 가구수에 대한 상대오차 산점도를 보면, 다른 정렬방식에 비해 [정렬 A]와 [정렬 F] 방식을 통해 상대적으로 균형을 이루는 5개 그룹으로 분할하는 것이 가능하고, 이 경우 결과적으로 각 분할에서 연간 표본을 통해 상당히 안정적인 추정결과를 얻을 수 있다는 것을 확인할 수 있다. <그림 4-2>에 제시된 각 변수별 제주도와 대전광역시에 대한 4개 변수에 대한 정렬방식별 산점도를 보면 분할에 따라 오차가 평균적으로 어떤 수준으로 발생하는지와 함께 이상치에 해당하는 분할이 발생하는지 여부를 시각적으로 쉽게 확인할 수 있다.

<표 4-3> 정렬방식에 따른 도단위 가구부분 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
가구수	제주	224629	0.34%	0.23%	1.37%	0.72%	1.76%	0.72%	0.91%	0.43%	1.24%	0.55%	0.47%	0.38%
	대전	588159	1.03%	0.47%	1.10%	0.74%	2.38%	1.55%	1.44%	0.91%	1.31%	0.79%	1.40%	0.64%
1인 가구	제주	60274	2.07%	0.96%	2.95%	1.59%	1.77%	0.79%	1.04%	0.42%	0.70%	0.48%	3.20%	1.28%
	대전	172771	1.66%	0.66%	2.19%	1.16%	2.99%	1.88%	1.86%	1.30%	2.71%	1.08%	2.52%	1.01%
2인 가구	제주	59517	1.05%	0.55%	1.63%	0.85%	1.62%	0.73%	1.06%	0.58%	0.57%	0.42%	2.10%	1.25%
	대전	141941	1.78%	0.71%	1.10%	0.51%	2.43%	1.32%	0.97%	0.72%	1.76%	0.92%	1.58%	0.76%
3인 가구	제주	44277	3.20%	1.37%	2.85%	1.14%	2.67%	1.25%	1.89%	0.76%	2.55%	1.63%	1.44%	0.72%
	대전	122393	1.39%	0.65%	1.54%	0.78%	2.38%	1.36%	1.42%	0.57%	1.59%	0.85%	2.16%	0.86%



[그림 4-2] 정렬방식에 따른 5개 그룹별 도단위 가구부문 상대오차 산점도

3) 인구 부문 기준 분할방법 비교

인구 부문의 경우 각 정렬방법으로 구성된 5개 분할에서 제주도와 대전광역시에 대한 전체 인구수를 포함해 20대, 40대, 60대 인구수, 가구주와의 관계 중 자녀와 부모에 해당하는 인구수 추정에 대한 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-4>와 같다.

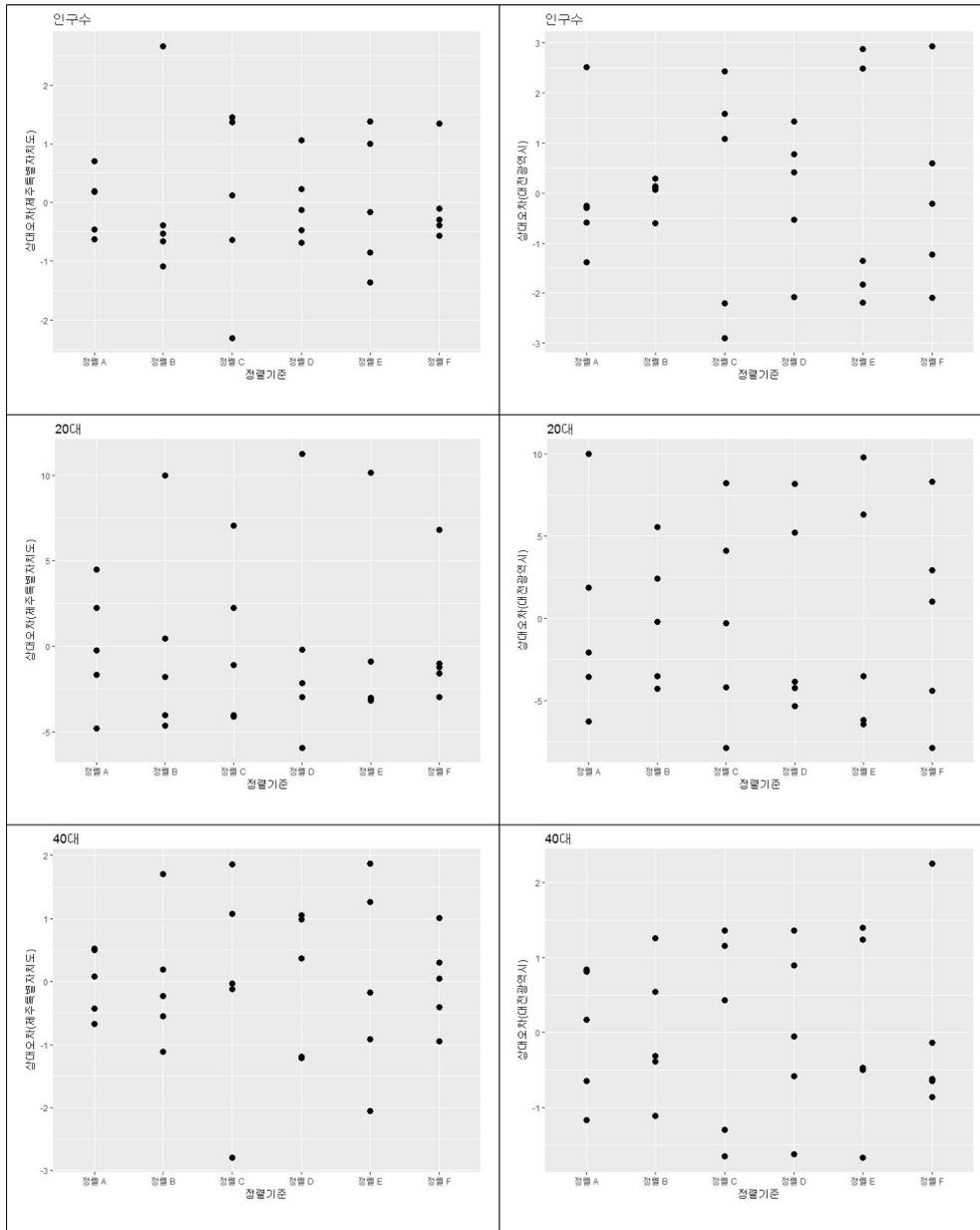
전체 인구수 추정 관점에서 보면 제주도에서는 [정렬 A], 대전광역시에서는 [정렬 B]가 가장 효과적인 것으로 나타났다. 연령대별 인구수 추정에 있어서는 20대 인구수의 경우 40대와 60대에 비해 상대오차가 상당히 크게 나타난다는 특징을 볼 수 있으며, 제주도에서는 [정렬 A] 대전광역시에서는 [정렬 B]가 상대오차가 작아지는 정렬방식인 것으로 나타났다. 40대 인구수의 경우 제주도와 대전광역시 모두 [정렬 A]가 대체적으로 효과적인 정렬방법인 것으로 나타났으며, 60대 인구수의 경우 제주도에서는 [정렬 E] 또는 [정렬 F], 대전광역시에서는 [정렬 F]가 분할에 따른 오차를 줄이는 정렬방식인 것으로 나타났다. 여기서 [정렬 F]는 조사구 정렬 기준으로 50대 이상 인구비율을 사용한 것이기 때문에 이런 결과는 예상했던 것과 일치하는 결과이다.

한편 제주도에서는 가구주 관계에서 자녀에 해당하는 인구수의 경우 [정렬 F], 가구주 관계가 부모인 경우 [정렬 B]가 오차가 적은 것으로 나타났고, 대전광역시에서는 자녀의 경우 [정렬 A], 부모의 경우 [정렬 F]가 효과적인 것으로 나타났다.

한편 인구 부문이 경우 각 변수별로 정렬방식에 따라 구성된 분할에 따라 오차가 어느 정도 발생하는지 살펴보기 위해 5개 분할별로 발생한 상대오차를 산점도로 정리하면 <그림 4-3>과 같다. 첫 번째 행 첫 번째 열에 있는 제주도 인구수 추정에 대한 상대오차 산점도를 보면, 다른 정렬방식에 비해 [정렬 A]가 상대오차가 가장 적게 발생하는 것을 볼 수 있고, [정렬 F]의 경우 이상점에 해당하는 그룹이 1개 발생하지만 이 것만 제외하면 오히려 [정렬 A]에 비해 상대적으로 안정적인 결과를 가져다 줄 수 있다는 것을 알 수 있다. 아울러 2번째 행에 있는 20대에 대한 산점도를 보면 대부분의 정렬방식에서 이상점에 해당하는 매우 큰 오차가 발생하는 분할이 구성될 가능성과 함께 상대오차가 다른 연령대에 비해 매우 크게 발생할 가능성이 높다는 것을 볼 수 있다.

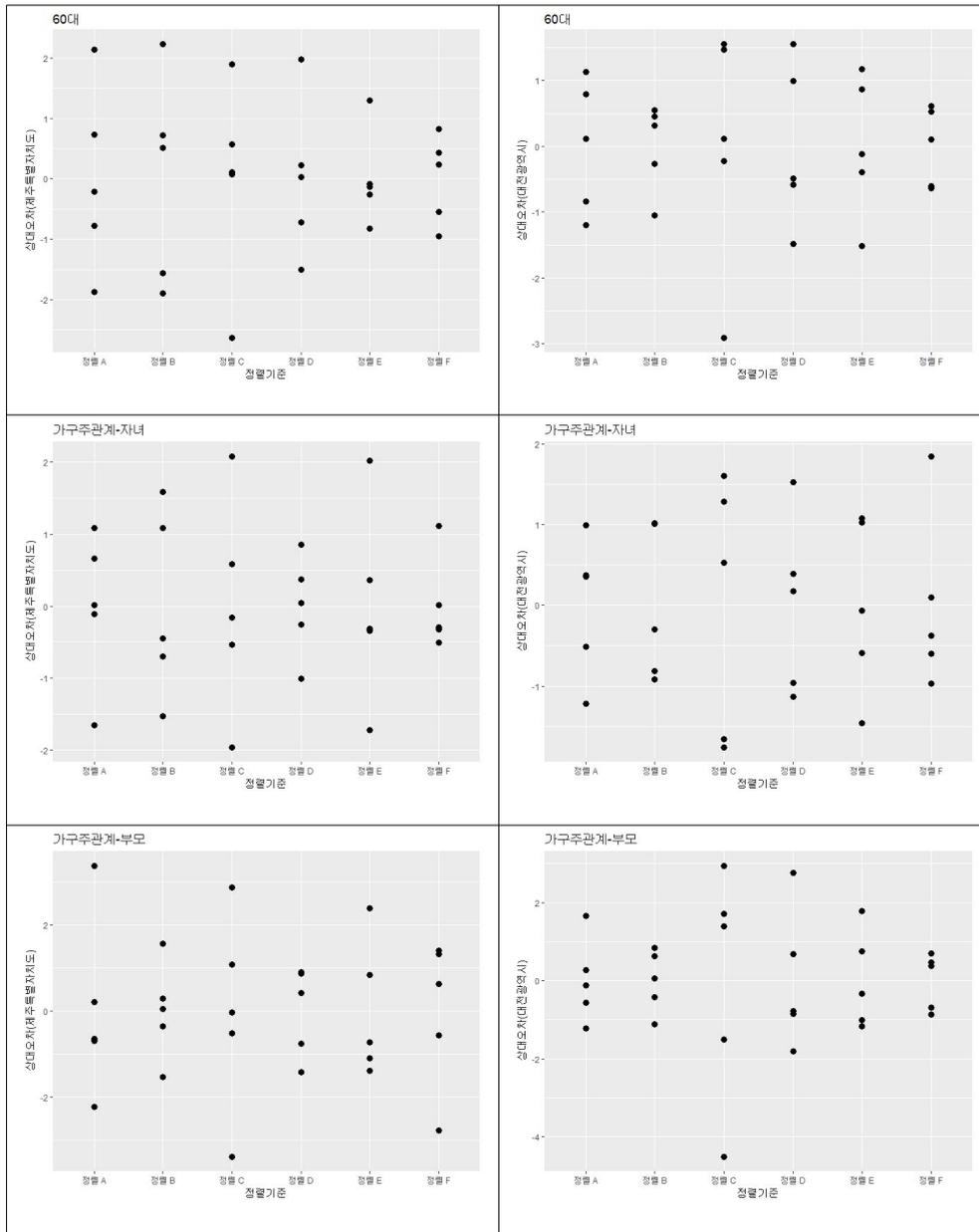
〈표 4-4〉 정렬방식에 따른 도단위 인구부분 추정오차 비교

변수		총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]	
			최대 상대오차	평균 상대오차	최대상대 오차	평균 상대오차	최대 상대오차	평균 상대오차						
인구수	제주	595573	0.71%	0.43%	2.67%	1.07%	2.32%	1.18%	1.06%	0.51%	1.39%	0.95%	1.35%	0.54%
	대전	1515211	2.52%	1.01%	0.60%	0.24%	2.90%	2.04%	2.08%	1.05%	2.87%	2.15%	2.94%	1.41%
20대	제주	65317	4.83%	2.70%	9.99%	4.18%	7.04%	3.71%	11.27%	4.51%	10.18%	4.07%	6.82%	2.73%
	대전	218049	10.02%	4.76%	5.57%	3.20%	8.24%	4.95%	8.18%	5.36%	9.80%	6.44%	8.31%	4.90%
40대	제주	105153	0.67%	0.44%	1.70%	0.76%	2.80%	1.18%	1.22%	0.96%	2.05%	1.26%	1.01%	0.55%
	대전	261420	1.16%	0.73%	1.26%	0.72%	1.65%	1.18%	1.62%	0.90%	1.67%	1.06%	2.26%	0.90%
60대	제주	113519	2.14%	1.15%	2.23%	1.39%	2.64%	1.05%	1.98%	0.89%	1.30%	0.52%	0.95%	0.60%
	대전	240527	1.20%	0.81%	1.05%	0.53%	2.91%	1.25%	1.56%	1.02%	1.52%	0.81%	0.63%	0.49%
가구주관계-자녀	제주	191690	1.65%	0.70%	1.59%	1.07%	2.08%	1.07%	1.01%	0.51%	2.02%	0.95%	1.11%	0.45%
	대전	477962	1.21%	0.69%	1.02%	0.81%	1.76%	1.37%	1.53%	0.84%	1.46%	0.84%	1.85%	0.78%
가구주관계-부모	제주	18782	3.37%	1.43%	1.56%	0.75%	3.39%	1.58%	1.42%	0.87%	2.39%	1.29%	2.78%	1.34%
	대전	41086	1.65%	0.77%	1.11%	0.61%	4.52%	2.41%	2.76%	1.38%	1.77%	1.01%	0.87%	0.62%



[그림 4-3] 정렬방식에 따른 5개 그룹별 도 단위 인구부문 추정오차 산점도

(그림 4-3 계속)

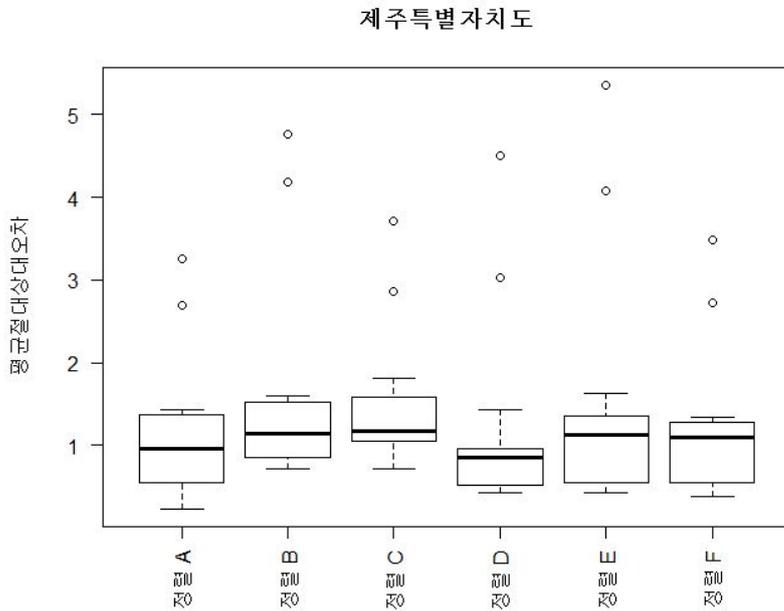


[그림 4-3] 정렬방식에 따른 5개 그룹별 도 단위 인구부문 추정오차 산점도

4) 도 단위 분할방법 비교 결과 종합

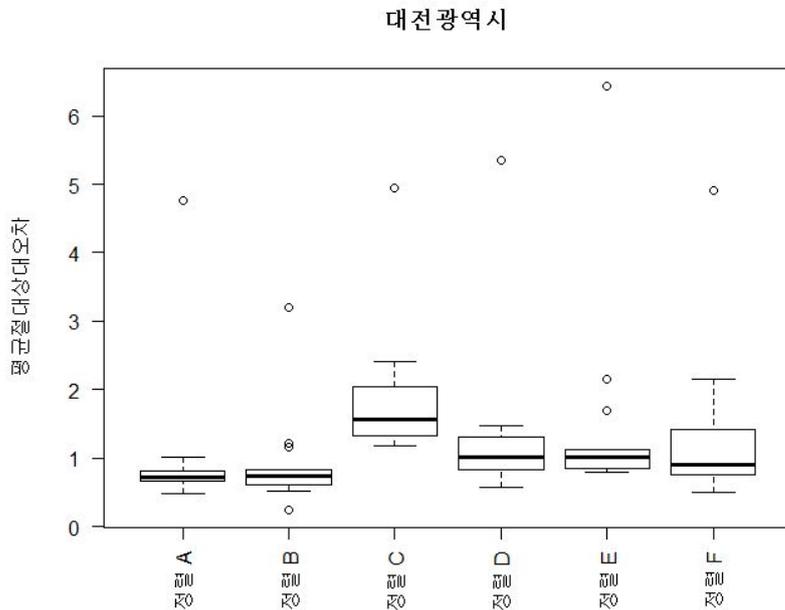
앞에서 살펴본 대로 주택, 가구, 인구 부문별로 비교한 결과를 보면, 변수나 지역별로 가장 효과적인 분할 결과를 도출해 주는 정렬방식이 동일하지 않다는 것을 볼 수 있다. 실제 광역시·도별로는 이런 차이가 있을 수 있고, 변수별로도 차이가 발생하게 된다. 앞에서 고려한 주택 부문 3개 변수, 가구 부문 4개 변수, 인구 부문 6개 변수에 대한 평균상대오차를 정렬방식에 따라 상자그림으로 정리해 보면 <그림 4-4>와 같다.

본 연구에서 고려한 12개 변수를 종합적으로 고려해 보면 대체적으로 [정렬 A]와 [정렬 D]가 안정적인 분할 결과를 주는 것으로 보인다. 여기서 이상점으로 표시된 것들은 앞에서 지적했던 것처럼 제주도에서 다른 변수들에 비해 평균상대오차가 크게 발생한 아파트 수와 20대 인구수에 대한 추정에서 발생한 것들이다. 실제 이런 이상점을 제외한다면 제주도의 경우 동읍면은 인구수로 정렬하고 동읍면 내에서 조사구는 조사구 코드로 정렬하는 [정렬 D]가 가장 안정적인 분할을 가져오는 것으로 보인다.



[그림 4-4] 제주도 정렬방식에 따른 평균상대오차 분석 결과

대전광역시의 경우 12개 변수를 종합적으로 고려해 보면 [정렬 A]가 가장 안정적인 분할 결과를 가져오는 것으로 보인다. 여기서 이상점으로 표시된 변수는 20대 인구수에 추정에 대한 것으로 이런 특수한 경우만을 제외하게 되면 대전광역시의 경우 동읍면코드와 조사구코드를 기준으로 정렬한 [정렬 A]를 기준으로 모집단을 분할하는 경우 대부분의 변수에 대해서 안정적인 통계작성이 가능한 모집단 분할이 가능한 것으로 나타났다.



[그림 4-5] 대전광역시 정렬방식에 따른 평균상대오차 분석 결과

제주도와 대전광역시를 대상으로 한 도 단위 모집단 분할 분석 결과를 참고로 하면 우리나라의 경우 외국사례와는 달리 모집단 분할 최소단위로 가구 대신 조사구를 사용하기 때문에 향후 순환표본조사를 성공적으로 수행하기 위해서는 다음 사항들에 대한 검토가 필요할 것으로 생각된다.

첫째, 미국이나 프랑스의 경우와는 달리 우리나라의 경우 가구 대신에 조사구를 최소단위로 모집단을 분할하게 되는 경우, 이는 이론적으로 2상 표본추출 (two-phase sampling)에 해당한다. 1단계로 전체 조사구 중 계통추출로 1/5 표본을 추출한 후(이 과정은 표본추출틀 분할에 해당 함), 추출된 조사구 중 일부를 다시 표본으로 추출하는 2단계 추출과정을 거치기 때문에 1단계 표본추출, 다시 말해

모집단을 분할에 따른 추가적인 표본추출오차(sampling error)를 분석에 반영하는 것이 필요할 것으로 보인다.

참고로 미국 ACS의 경우 가구를 최소단위로 모집단 분할을 하기 때문에 실제 1단계 추출에서의 표본크기가 매우 크고, 그룹들 간의 차이가 적은 모집단 분할이 가능하기 때문에 1단계 추출에 따른 표본추출오차는 무시해도 될 정도로 작을 것으로 보이지만, 조사구를 기준으로 분할을 하는 경우에는 각 광역시·도의 조사구 수가 크지 않을 수 있기 때문에 1단계 표본추출 과정 즉, 1/5 조사구 계통표본추출에 의한 5개 그룹으로의 모집단 분할에서 발생하는 표본추출오차를 분석과정에 반영하는 것이 필요할 것으로 보인다. 예를 들어, 제주도의 경우 모집단 분할 단계는 전체 조사구 8,205개 중에서 1,641개 조사구를 1단계로 1/5 계통표본추출하는 것에 해당하고, 조사구를 기준으로 보면 표본추출오차를 무시할 수준으로 표본크기가 크다고 볼 수 없다는 점에 유의할 필요가 있다.

둘째, 본 연구에서는 자료 활용상의 한계로 제주도와 대전광역시에 대해서만 살펴 보았지만, 가구 대신 조사구를 최소단위로 모집단을 분할하는 경우 20대 인구수 등과 같은 일부 변수에 대해서는 균형이 잡힌 모집단 분할이 불가능할 수 있다. 이렇게 분할된 그룹에서 전체 모집단을 대표하는 표본을 추출하기 위해서는 연간 표본추출단계에서 당 해 년도 부차 추출틀(sub-frame)을 대상으로 주요 변수에 대한 균형 여부를 사전에 충분히 검토하고, 이를 토대로 균형이 깨진 변수들에 대해서는 표본추출 확률을 조정하는 작업이 필요할 것으로 보인다. 이와 같이 조사구 특성에 따라 표본추출률을 조정하는 작업은 해당 연도에 사용되는 부차 추출틀을 기준으로 매년 독립적으로 이루어질 필요가 있다.

3. 시군구 단위 분할방법 비교

시군구 통계 작성을 목적으로 하는 경우 본 연구에서 제시한 정렬방식에 따라 얼마나 균형이 잡힌 모집단 분할이 가능한지 확인해 보기로 한다. 현재 도입을 추진 중인 순환표본조사 조사설계에 따르면 시군구 통계의 경우 매년 통계를 작성하는 것이 아니라 2년 혹은 3년 동안 조사된 자료를 통합해서 통계를 작성하게 될 가능성이 높기 때문에 시군구 단위에서는 분할 후 5개 그룹에서 발생하는 개별 상대오차는 크게 의미가 없을 수 있다.

하지만 인구주택총조사에서 핵심적인 변수에 해당하는 주택수, 가구수 및 인구수 등 3개 변수를 기준으로 각 정렬방식에 따른 시군구 단위에서의 상대오차를 살펴봄

으로써 본 연구에서 적용한 정렬방식을 통해 시군구 단위에서 어느 정도 균형이 잡힌 모집단 분할이 가능한지를 살펴본다는 측면에서 의미를 찾을 수 있을 것이다.

1) 주택 관련 변수 기준 시군구 단위 분할방법 비교

각 시군구의 주택수를 6가지 정렬방식을 통해 구축된 5개 분할에서 추정하는 경우 발생하는 추정오차를 기준으로 제주도와 대전광역시 내의 모든 시군구에 대한 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-5>과 같다.

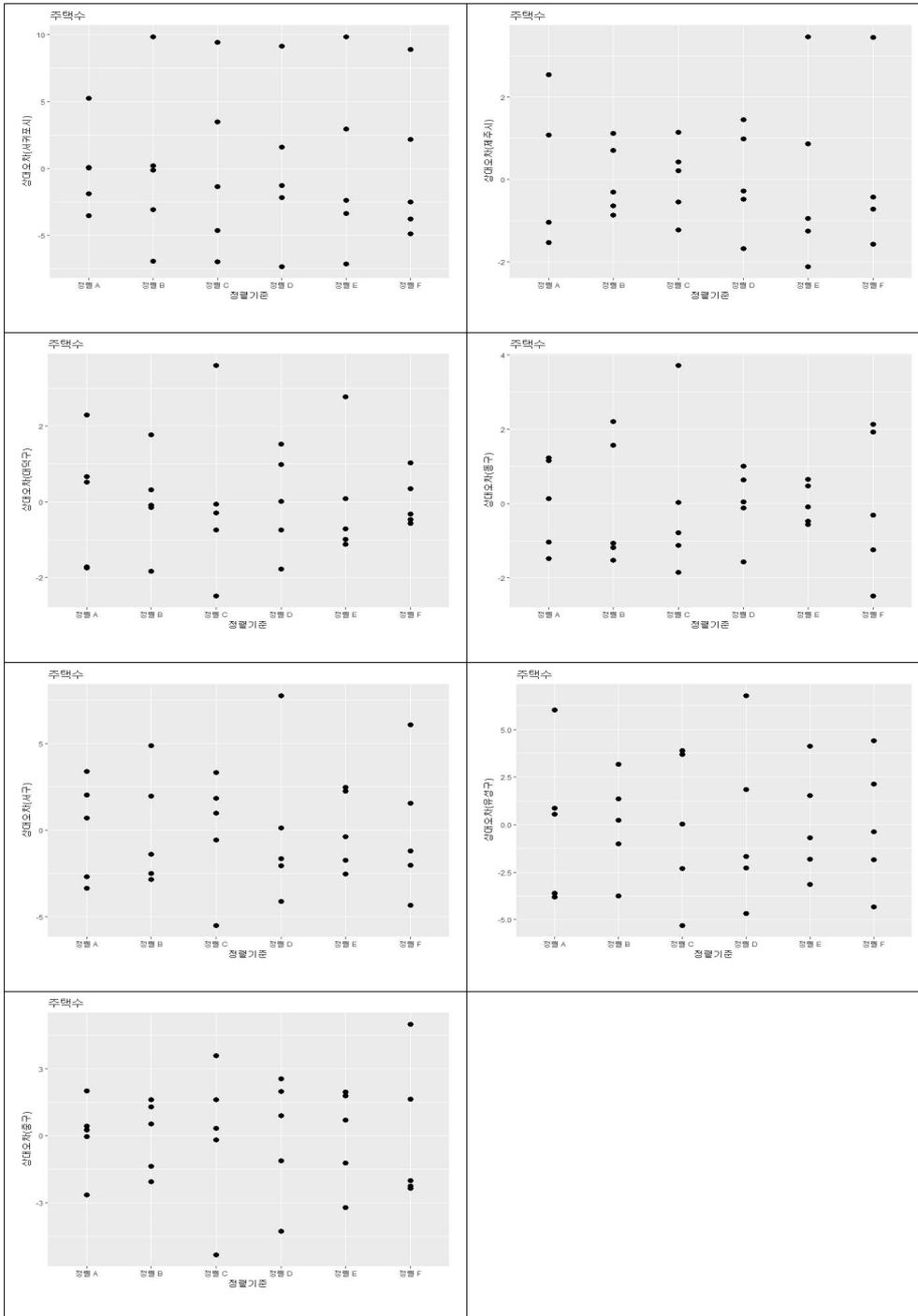
먼저 <표 4-2>에 제시된 도 단위에서의 분석결과와 비교해 보면 시군구별로 차이는 있지만 대체적으로 도 단위 통계작성을 고려했을 때에 비해 추정오차가 크게 발생하는 것을 볼 수 있다. 시군구별로 살펴보면 서귀포시의 경우 상대추정오차가 다른 시군구에 비해 가장 크게 나타났으며, 정렬방식 중 [정렬 A]의 평균상대오차가 가장 작은 것으로 나타났다. 제주시의 경우 서귀포시에 비해 상대적으로 추정오차가 작았으며 [정렬 B]와 [정렬 C]가 모집단 분할 관점에서 효과적인 정렬방식인 것으로 나타났다.

대전광역시의 경우 유성구의 경우 추정오차가 가장 크게 나타났으며, 동구의 상대적으로 추정오차가 작은 것으로 나타났다. 대덕구의 경우 [정렬 F], 동구의 경우 [정렬 E], 서구의 경우 [정렬 A], 유성구의 경우 [정렬 B], 중구의 경우 [정렬 A]가 가장 효과적인 것으로 나타났다. 시군구에 따라 조사구 주택 구성에 차이가 있기 때문에 모든 구에서 안정적인 분할을 구축해 주는 모집단 분할 방안은 존재하지 않을 것으로 생각된다.

한편 <그림 4-6>은 시군구별로 각 정렬방식에 따라 구축된 5개 분할에서 발생한 상대오차를 산점도 형식으로 나타낸 것이다. <그림 4-6>의 1행 1열에 해당하는 서귀포시에 대한 주택수 추정오차 산점도를 보면 [정렬 A]를 기준으로 한 분할의 경우 5개 모든 그룹에서 다른 정렬방식에 비해 안정적으로 모집단 주택수 산출이 가능하다는 것을 볼 수 있으며, 이는 주택수를 목적 변수로 생각하는 경우 [정렬 A] 방식으로 통해 상대적으로 다른 정렬방식에 비해 균형이 잡힌 분할을 구축할 수 있다는 것을 보여주고 있다. 마찬가지로 2행 2열에 있는 대전광역시 동구의 경우 다른 정렬방식에 비해 [정렬 E]가 가장 안정적인 모집단 분할을 구축해 주는 정렬방식이라는 것을 알 수 있다. 다른 시군구에 대해서도 <그림 4-6>을 통해 정렬방식에 따라 구축된 5개 분할에서 주택수 추정에 있어서 어떤 양상을 보여주고 있는지 파악할 수 있다.

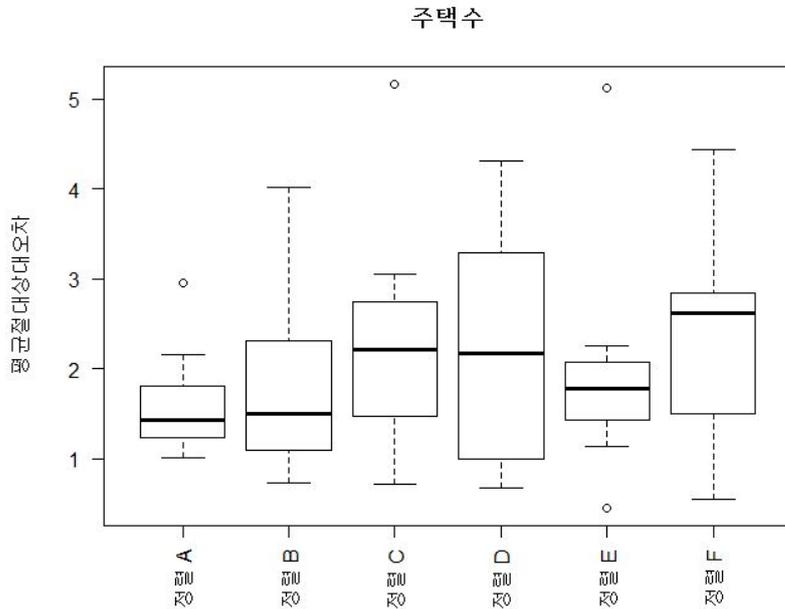
<표 4-5> 정렬방식에 따른 시군구 단위 주택수 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
주택수	서귀포시	53725	5.27%	2.16%	9.85%	4.02%	9.45%	5.17%	9.14%	4.31%	9.87%	5.13%	8.93%	4.44%
	제주시	141451	2.54%	1.45%	1.12%	0.73%	1.23%	0.71%	1.68%	0.98%	3.47%	1.73%	3.45%	1.38%
	대덕구	61560	2.29%	1.39%	1.84%	0.83%	3.61%	1.44%	1.77%	1.00%	2.77%	1.14%	1.03%	0.55%
	동구	77218	1.47%	1.01%	2.20%	1.51%	3.72%	1.50%	1.56%	0.67%	0.65%	0.45%	2.49%	1.62%
	서구	143555	3.38%	1.42%	4.85%	2.72%	5.53%	2.44%	7.76%	3.14%	2.53%	1.87%	6.05%	3.03%
	유성구	105809	6.01%	2.96%	3.76%	1.90%	5.32%	3.05%	6.79%	3.45%	4.13%	2.26%	4.42%	2.62%
	중구	80350	2.66%	1.08%	2.05%	1.37%	5.35%	2.22%	4.28%	2.17%	3.22%	1.78%	5.00%	2.66%



[그림 4-6] 정렬방식에 따른 5개 그룹별 시군구단위 주택수 추정오차 산점도

한편 각 정렬방식별로 제주도와 대전광역시의 7개 시군구에 대한 평균상대오차를 상자그림으로 정리한 결과는 [그림 4-7]과 같다. 7개 시군구의 평균상대오차를 종합적으로 보면 대체적으로 [정렬 A]와 [정렬 E]가 모든 시군구에서 안정적인 분할 결과를 주는 것으로 보이며, [정렬 E]의 경우는 이상점에 해당하는 서귀포시만 제외하면 가장 효과적인 분할 결과를 가져다주는 것으로 보인다.



[그림 4-7] 정렬방식에 따른 시군구 주택수 평균상대오차 비교

한편 7개 시군구를 대상으로 주택 관련 변수 중 단독주택 수와 아파트 수를 6가지 정렬방식을 통해 구축된 5개 분할에서 추정하는 경우 발생하는 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-6>와 <표 4-7>에 제시되어 있다.

〈표 4-6〉 정렬방식에 따른 시군구단위 단독주택수 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
단 독 주 택 수	서귀포시	30421	3.83%	2.14%	2.07%	1.50%	5.23%	2.21%	3.67%	2.20%	3.00%	1.67%	3.71%	1.59%
	제주시	53163	0.82%	0.47%	3.77%	2.06%	3.60%	2.14%	2.82%	1.14%	2.46%	1.86%	2.77%	1.22%
	대덕구	11052	5.95%	2.88%	7.21%	2.88%	7.35%	4.09%	11.61%	4.64%	2.60%	1.99%	5.81%	2.32%
	동구	20052	3.35%	1.42%	7.42%	3.05%	2.53%	1.01%	7.74%	3.10%	2.25%	1.16%	6.25%	2.67%
	서구	20617	3.16%	2.43%	2.87%	1.74%	4.48%	1.79%	5.73%	2.80%	4.62%	3.09%	3.62%	2.19%
	유성구	10510	8.18%	4.59%	11.99%	4.80%	6.66%	2.66%	7.18%	3.14%	9.51%	5.02%	7.71%	3.08%
	중구	19054	4.38%	1.75%	7.20%	3.41%	2.39%	1.88%	6.51%	2.69%	8.06%	4.75%	5.04%	2.31%

〈표 4-7〉 정렬방식에 따른 시군구단위 아파트수 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
아 파 트 수	서귀포시	11851	6.42%	2.99%	14.23%	9.85%	22.62%	11.65%	18.34%	9.78%	26.57%	12.50%	9.53%	5.36%
	제주시	50681	5.36%	3.49%	6.40%	5.01%	7.10%	2.84%	7.46%	2.98%	14.15%	5.66%	10.37%	4.16%
	대덕구	39534	2.08%	1.15%	5.04%	2.63%	5.00%	2.32%	1.98%	1.09%	5.23%	2.98%	3.44%	2.74%
	등구	46817	4.45%	2.08%	7.60%	4.80%	6.72%	2.69%	2.53%	1.04%	2.74%	1.62%	4.85%	2.53%
	서구	112731	4.13%	2.53%	6.92%	3.71%	5.31%	2.14%	9.81%	4.05%	4.09%	2.54%	8.42%	3.68%
	유성구	89204	6.70%	3.47%	5.16%	2.27%	7.08%	3.94%	8.80%	4.70%	3.06%	2.00%	5.19%	3.02%
	중구	49604	4.98%	2.56%	6.35%	4.03%	8.64%	4.23%	6.92%	3.55%	4.77%	2.62%	9.11%	4.04%

2) 가구 관련 변수 기준 시군구 단위 분할방법 비교

각 시군구의 가구수를 6가지 정렬방식으로 구축한 5개 분할에서 발생하는 추정오차를 기준으로 제주도와 대전광역시 내의 모든 시군구에 대한 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-8>과 같다.

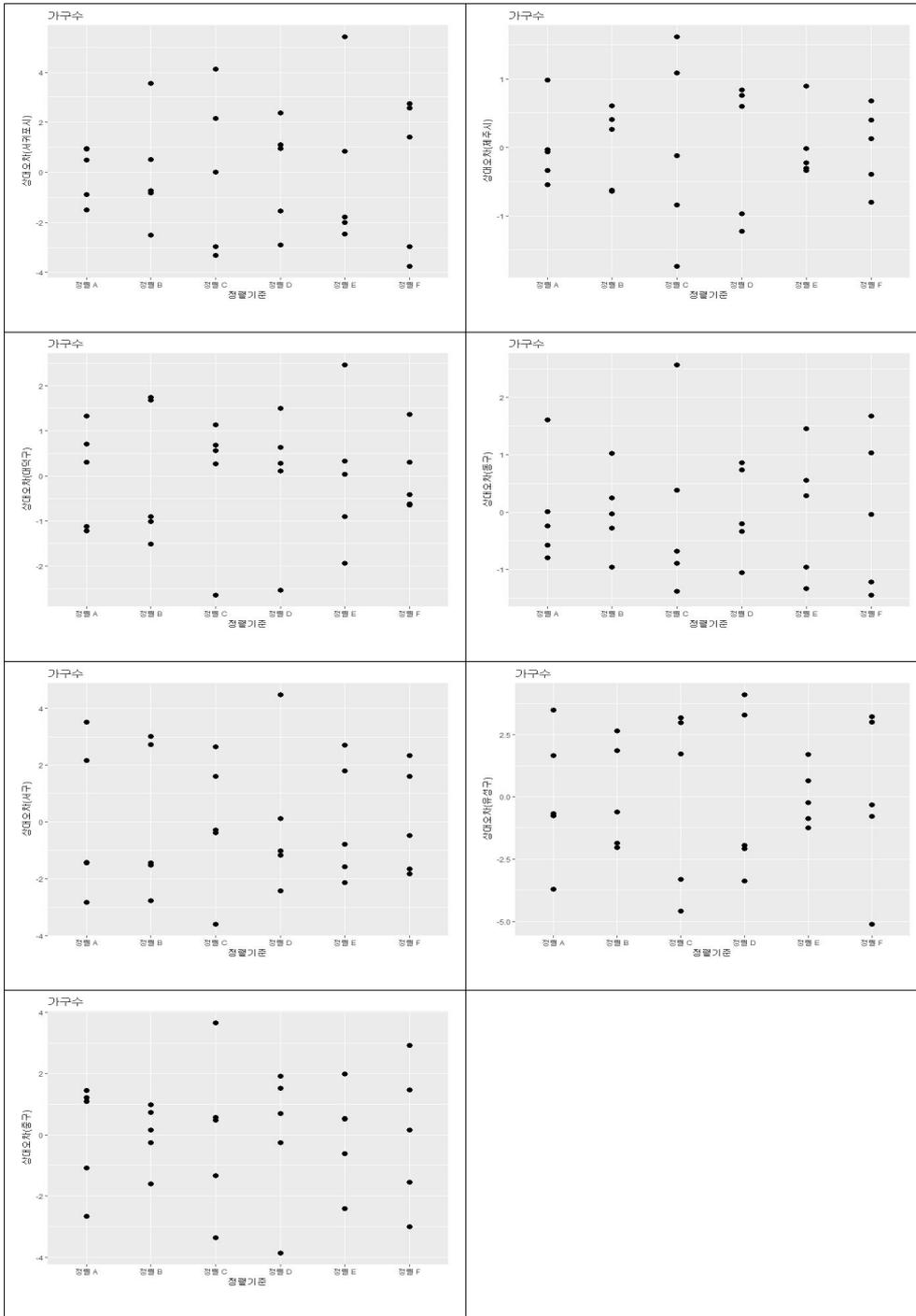
시군구별로 살펴보면 서귀포시의 경우 [정렬 A]의 평균상대오차가 가장 작은 것으로 나타났다. 제주시의 경우 조사구 코드를 기준으로 정리한 [정렬 A]와 조사구를 인구수로 정렬한 [정렬 E]가 모집단 분할 관점에서 효과적인 정렬방식인 것으로 나타났다.

대전광역시 대덕구의 경우 [정렬 F], 동구의 경우 [정렬 B], 서구의 경우 [정렬 F], 유성구의 경우 [정렬 E], 중구의 경우 [정렬 B]로 모집단을 분할하는 경우 가구수에 대한 평균상대오차가 가장 작은 것으로 나타났다. 주택수와 마찬가지로 가구수 관점에서도 시군구에 따라 조사구 가구 구성에 차이가 있기 때문에 모든 시군구에서 항상 안정적인 분할을 구축해 주는 분할 방안은 존재하지 않는 것으로 보인다. 아울러 앞에서도 언급했듯이 순환표본조사에서 시군구 통계는 2년 혹은 3년 자료를 통합해 통계를 작성하기 때문에 각 분할에서 발생하는 추정오차가 크더라도 순환표본조사에서의 해당 시군구에 대한 통계작성으로 그런 오차가 그대로 연결되는 것은 아니라는 점에 유의할 필요가 있다.

한편 [그림 4-8]은 시군구별로 각 정렬방식에 따라 구축된 5개 분할에서 발생한 상대오차를 산점도 형식으로 나타낸 것이다. [그림 4-8]의 1행 1열에 해당하는 서귀포시에 대한 가구수 추정오차 산점도를 보면 [정렬 A]를 기준으로 한 분할의 경우 5개 모든 그룹에서 다른 정렬방식에 비해 안정적으로 모집단 가구수 산출이 가능하다는 것을 볼 수 있다. 이는 가구수를 목적 변수로 하는 경우 [정렬 A] 방식을 통해 다른 정렬방식에 비해 상대적으로 균형이 잡힌 분할을 구축할 수 있다는 것을 보여주고 있다. 마찬가지로 1행 2열에 있는 제주시의 경우 다른 정렬방식에 비해 [정렬 E]가 가장 안정적인 모집단 분할을 구축해 주는 정렬방식이라는 것을 알 수 있다. 다른 시군구의 경우에도 [그림 4-8]의 산점도를 통해 정렬방식에 따라 구축된 5개의 분할에서 각각 가구수라는 변수를 기준으로 보면 어느 정도의 균형을 확보할 수 있는지 파악할 수 있다.

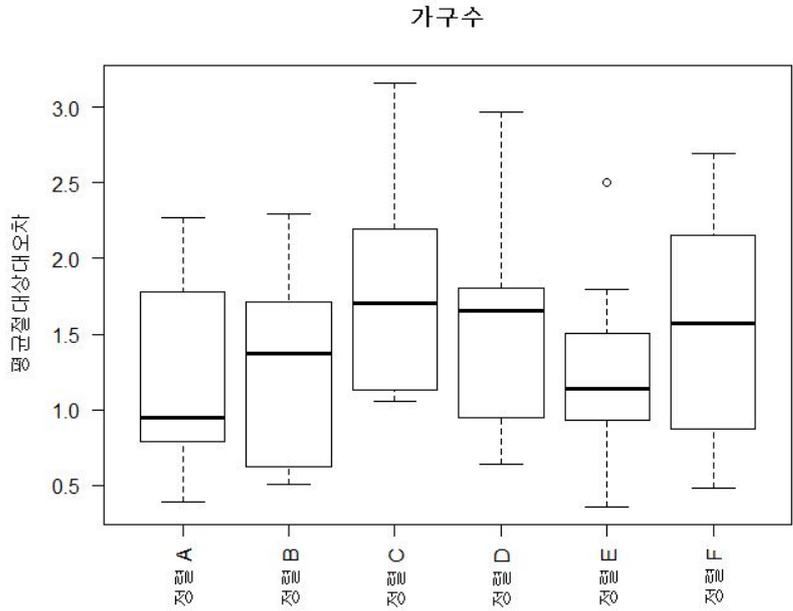
<표 4-8> 정렬방식에 따른 시군구단위 가구수 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
가구수	서귀포시	58191	1.50%	0.95%	3.55%	1.62%	4.12%	2.51%	2.89%	1.77%	5.42%	2.50%	3.76%	2.69%
	제주시	166438	0.98%	0.39%	0.64%	0.51%	1.74%	1.08%	1.22%	0.88%	0.89%	0.36%	0.80%	0.48%
	대덕구	75452	1.34%	0.94%	1.75%	1.37%	2.65%	1.06%	2.53%	1.01%	2.47%	1.14%	1.37%	0.67%
	동구	101020	1.60%	0.64%	1.02%	0.51%	2.57%	1.18%	1.05%	0.64%	1.45%	0.92%	1.67%	1.08%
	서구	187672	3.51%	2.27%	3.01%	2.29%	3.60%	1.70%	4.48%	1.84%	2.70%	1.80%	2.34%	1.58%
	유성구	128442	3.70%	2.05%	2.66%	1.81%	4.59%	3.16%	4.11%	2.97%	1.71%	0.95%	5.12%	2.49%
	중구	95573	2.67%	1.50%	1.61%	0.75%	3.66%	1.88%	3.87%	1.65%	2.41%	1.21%	3.01%	1.82%



[그림 4-8] 정렬방식에 따른 5개 그룹별 시군구단위 가구수 추정오차 산점도

한편 가구수를 기준으로 각 정렬방식별로 제주도와 대전광역시의 7개 시군구에 대한 평균상대오차를 상자그림으로 정리한 결과는 [그림 4-9]와 같다. 7개 시군구의 평균상대오차를 종합적으로 보면 대체적으로 [정렬 A]와 [정렬 E]가 모든 시군구에서 안정적인 분할 결과를 주는 것으로 나타났으며, [정렬 E]의 경우는 이상점에 해당하는 서귀포시만 제외하면 가장 효과적인 분할 결과를 가져다주는 것으로 보인다. 이런 정렬방식에 따른 평균상대오차 비교 결과는 [그림 4-7]의 주택수를 기준으로 한 분석결과와 일치한다.



[그림 4-9] 정렬방식에 따른 시군구 가구수 평균상대오차 비교

한편 제주도와 대전광역시의 7개 시군구를 대상으로 가구 관련 변수 중 1인, 2인 3인 가구원 가구의 수를 목적 변수로 설정한 경우 정렬방식에 따른 상대오차를 비교한 결과는 <표 4-9>에 정리되어 있다.

<표 4-9> 정렬방식에 따른 시군구단위 가구규모별 가구수 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
1인 가구	서귀포시	15761	3.65%	2.73%	3.86%	2.69%	5.01%	2.00%	2.34%	0.94%	6.15%	3.70%	7.02%	4.47%
	제주시	44513	2.23%	1.68%	2.99%	1.20%	2.21%	1.70%	2.24%	0.90%	2.50%	1.07%	2.86%	1.56%
	대덕구	20906	4.48%	1.79%	7.03%	4.88%	3.28%	1.98%	4.29%	1.85%	7.05%	3.11%	3.80%	2.10%
	동구	35167	3.78%	1.79%	3.61%	1.98%	2.23%	0.91%	2.45%	1.31%	4.29%	2.49%	1.26%	0.54%
	서구	53375	1.31%	0.68%	5.86%	2.72%	3.04%	1.58%	5.11%	2.04%	3.86%	1.68%	0.67%	0.36%
	유성구	38598	6.06%	2.55%	4.67%	2.90%	8.75%	5.90%	7.24%	4.19%	2.14%	1.09%	11.95%	5.07%
	중구	24725	4.04%	2.36%	4.89%	1.99%	5.99%	2.93%	4.37%	2.08%	1.92%	0.99%	1.60%	0.64%
2인 가구	서귀포시	17162	1.38%	0.99%	2.23%	1.01%	3.94%	2.35%	3.10%	2.19%	2.55%	1.20%	4.50%	2.84%
	제주시	42355	1.87%	0.75%	1.39%	0.93%	2.90%	1.50%	0.83%	0.53%	0.87%	0.53%	2.74%	1.10%
	대덕구	18730	1.87%	0.90%	2.38%	1.46%	2.08%	0.86%	1.87%	1.14%	3.60%	2.15%	2.48%	1.12%
	동구	24847	3.53%	1.44%	4.66%	1.86%	1.70%	1.08%	2.41%	1.40%	1.20%	0.49%	3.31%	1.38%
	서구	44443	4.46%	2.84%	3.44%	1.81%	2.91%	1.96%	3.63%	1.63%	2.82%	1.63%	2.97%	1.53%
	유성구	27967	4.12%	2.19%	4.05%	2.14%	5.26%	2.83%	4.10%	2.69%	2.06%	1.43%	5.39%	2.76%
	중구	25954	2.50%	1.39%	0.94%	0.55%	3.41%	1.85%	2.62%	1.05%	2.60%	1.58%	2.84%	1.18%
3인 가구	서귀포시	11234	8.24%	3.59%	5.17%	2.45%	7.84%	3.91%	3.82%	2.84%	9.93%	4.43%	3.24%	2.02%
	제주시	33043	1.54%	1.13%	2.06%	0.97%	1.73%	1.12%	1.75%	1.02%	1.02%	0.67%	3.03%	1.54%
	대덕구	16143	2.30%	1.61%	3.51%	1.91%	3.39%	1.69%	2.24%	1.30%	0.95%	0.45%	2.43%	1.07%
	동구	18892	2.13%	1.51%	1.74%	0.79%	3.30%	1.94%	3.48%	2.05%	1.68%	1.13%	2.58%	1.10%
	서구	40157	4.31%	2.99%	4.34%	2.20%	4.21%	2.18%	4.46%	1.79%	3.49%	2.11%	2.31%	1.58%
	유성구	26808	4.23%	2.39%	3.96%	2.07%	3.85%	2.13%	3.46%	1.87%	3.51%	2.37%	3.50%	1.48%
	중구	20393	2.22%	1.34%	4.92%	1.97%	3.32%	2.39%	3.25%	1.58%	4.47%	2.68%	4.50%	3.55%

3) 인구 관련 변수 기준 시군구단위 분할방법 비교

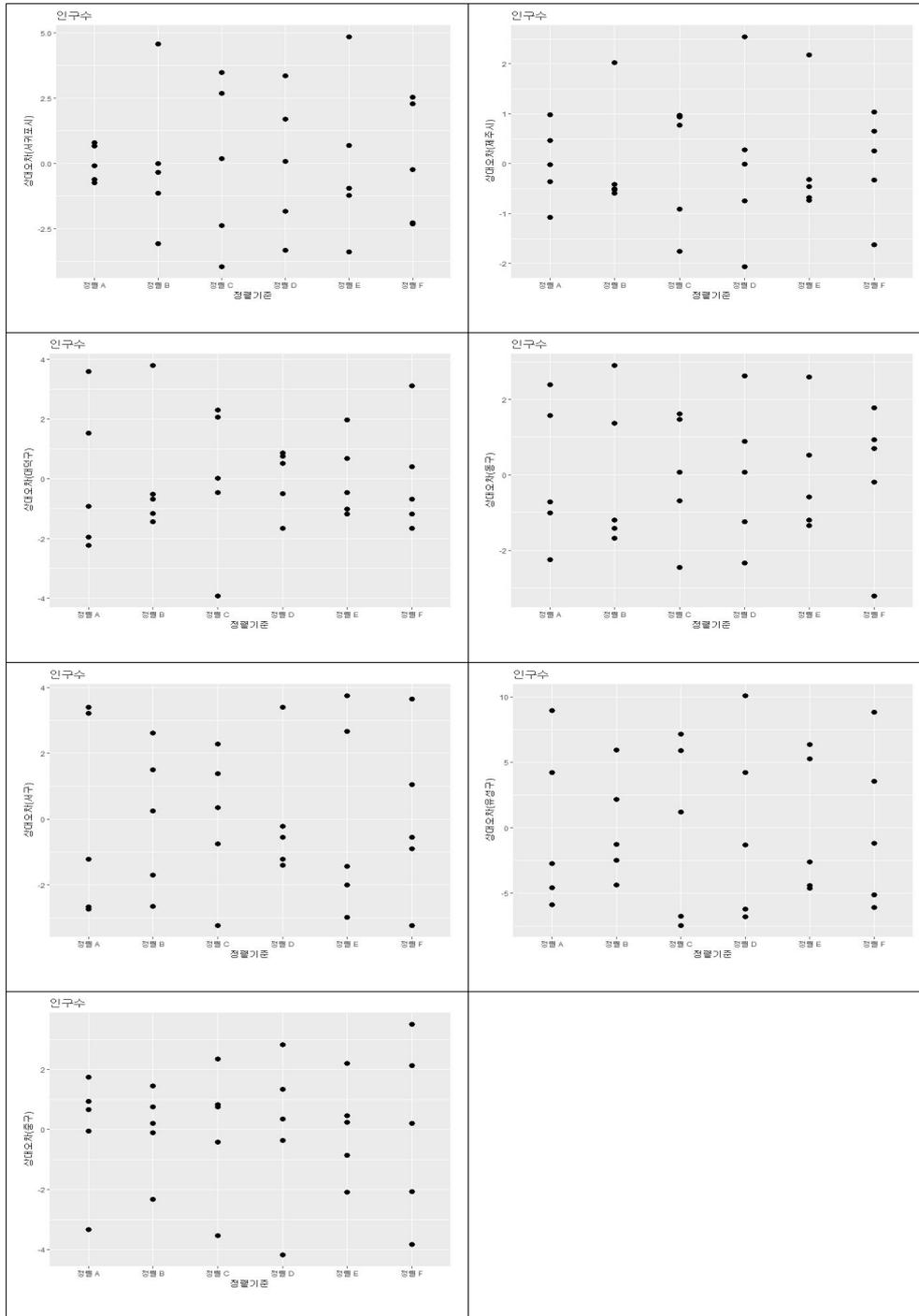
각 시군구의 인구수를 6가지 정렬방식으로 구축한 5개 분할에서 발생하는 추정오차를 기준으로 제주도와 대전광역시 내의 모든 시군구에 대한 최대상대오차와 평균상대오차를 정리한 결과는 <표 4-10>과 같다.

제주도 서귀포시와 제주시의 경우 모두 [정렬 A]의 평균상대오차가 가장 작은 것으로 나타났다. 대전광역시 대덕구의 경우 [정렬 D], 동구의 경우 [정렬 E], 서구의 경우 [정렬 D], 유성구의 경우 [정렬 B], 중구의 경우 [정렬 B]로 모집단을 분할하는 경우 가구수에 대한 평균상대오차가 가장 작은 것으로 나타났다. 인구수 관점에서 시군구에 따라 조사구 특성에 차이가 크기 때문에 모든 시군구에서 항상 안정적인 분할을 구축해 주는 분할 방안을 찾는 것은 어려울 것으로 보이고, 결국 각 시군구별로 별도의 분할 방법을 적용하는 것이 하나의 방안이 될 수 있을 것이다. 또한 순환표본조사에서 시군구 통계는 2년 혹은 3년 자료를 통합해 통계를 작성하기 때문에 각 분할에서 발생하는 추정오차가 순환표본조사에서 해당 시군구에 대한 통계작성으로 그대로 영향을 주는 것은 아니라는 점에 유의할 필요가 있다.

한편 [그림 4-10]은 시군구별로 각 정렬방식에 따라 구축된 5개 분할에서 발생한 인구수에 대한 상대오차를 산점도 형식으로 나타낸 것이다. [그림 4-10]의 1행 1열에 해당하는 서귀포시에 대한 인구수 추정오차 산점도를 보면 [정렬 A]를 기준으로 한 분할의 경우 5개 모든 그룹에서 다른 정렬방식에 비해 매우 안정적으로 모집단 가구수 산출이 가능하다는 것을 볼 수 있다. 이는 인구수를 목적 변수로 하는 경우 [정렬 A] 방식을 통해 다른 정렬방식에 비해 상대적으로 균형이 잡힌 추출틀 분할이 가능하다는 것을 보여주고 있다. 마찬가지로 2행 1열에 있는 대전광역시 대덕구의 경우 [정렬 D]가 가장 안정적인 모집단 분할을 구축해 주는 정렬방식이라는 것을 알 수 있다. 다른 시군구의 경우에도 [그림 4-10]의 추정오차 산점도를 통해 정렬방식에 따라 구축된 5개의 분할에서 각각 인구수라는 변수를 기준으로 보면 어느 정도의 균형을 확보할 수 있는지 살펴 볼 수 있다.

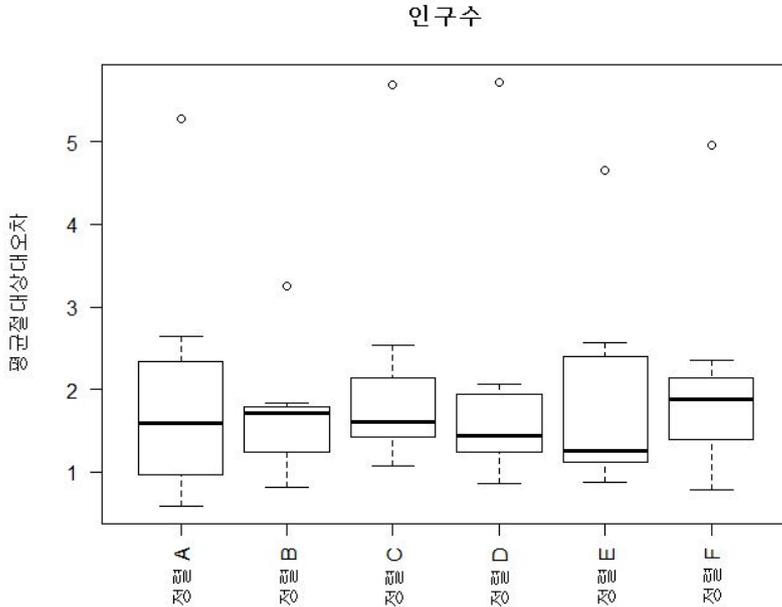
〈표 4-10〉 정렬방식에 따른 시군구단위 인구수 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
인구수	서귀포시	151204	0.79%	0.58%	4.57%	1.83%	3.96%	2.53%	3.36%	2.06%	4.86%	2.22%	2.53%	1.93%
	제주시	444369	1.07%	0.58%	2.03%	0.81%	1.76%	1.07%	2.55%	1.13%	2.18%	0.87%	1.62%	0.78%
	대덕구	195283	3.59%	2.04%	3.80%	1.52%	3.93%	1.75%	1.65%	0.86%	1.97%	1.06%	3.10%	1.40%
	동구	244425	2.39%	1.59%	2.91%	1.72%	2.46%	1.26%	2.63%	1.44%	2.60%	1.25%	3.22%	1.37%
	서구	483692	3.40%	2.64%	2.65%	1.74%	3.24%	1.60%	3.39%	1.36%	3.75%	2.57%	3.64%	1.88%
	유성구	344852	8.96%	5.28%	5.95%	3.25%	7.48%	5.70%	10.11%	5.73%	6.36%	4.66%	8.82%	4.96%
	중구	246959	3.33%	1.35%	2.31%	0.97%	3.53%	1.58%	4.18%	1.82%	2.20%	1.17%	3.82%	2.35%



[그림 4-10] 정렬방식에 따른 5개 그룹별 시군구단위 인구수 추정오차 산점도

한편 인구수를 기준으로 각 정렬방식별로 제주도와 대전광역시의 7개 시군구에 대한 평균상대오차를 상자그림으로 정리한 결과는 [그림 4-11]과 같다. 7개 시군구의 평균상대오차를 종합해 보면 대체적으로 [정렬 B]와 [정렬 D]가 모든 시군구에서 안정적인 분할 결과를 주는 것으로 나타났다. 여기서 이상점에 해당하는 지역은 대전광역시 유성구이다.



[그림 4-11] 정렬방식에 따른 시군구 인구수 평균상대오차 비교

한편 제주도와 대전광역시의 7개 시군구를 대상으로 인구 관련 변수 중 20대 인구, 40대 인구, 60대 인구 수를 목적 변수로 설정한 경우 정렬방식에 따른 상대오차를 비교한 결과는 <표 4-11>에 정리되어 있다. 또한 7개 시군구에서 가구주와의 관계가 자녀로 조사된 인구수와 가구주와의 관계가 부모로 조사된 인수를 기준으로 하는 경우 6가지 정렬방식에 따른 분할에서 발생하는 평균상대오차와 최대상대오차를 정리한 결과는 <표 4-12>에 제시되어 있다.

<표 4-11> 정렬방식에 따른 시군구단위 10세 간격 인구 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
20대	서귀포시	14580	11.04%	4.42%	7.00%	5.51%	5.21%	3.14%	7.13%	4.13%	1.92%	1.29%	3.22%	2.09%
	제주시	50737	7.01%	2.81%	10.86%	4.34%	8.97%	4.29%	13.60%	5.44%	12.73%	5.09%	8.21%	3.28%
	대덕구	26908	10.62%	4.60%	12.66%	5.06%	7.63%	3.87%	7.50%	3.23%	5.30%	2.76%	9.45%	4.09%
	동구	37587	7.66%	4.83%	9.37%	4.98%	6.66%	4.31%	8.24%	5.64%	8.79%	5.01%	8.44%	4.52%
	서구	69742	5.35%	3.69%	6.38%	3.28%	3.67%	1.56%	5.25%	2.30%	7.57%	3.03%	5.32%	3.52%
	유성구	54984	28.66%	17.92%	32.11%	18.27%	28.95%	17.77%	26.21%	20.91%	38.51%	21.93%	28.88%	17.73%
	중구	28828	5.94%	2.85%	5.07%	2.06%	3.39%	1.64%	7.17%	3.62%	2.92%	1.91%	6.36%	2.67%
40대	서귀포시	25266	3.40%	1.66%	7.42%	2.97%	6.45%	4.02%	5.12%	3.29%	8.19%	3.43%	4.65%	3.28%
	제주시	79887	1.20%	0.79%	0.77%	0.42%	2.19%	1.62%	1.19%	0.48%	1.87%	1.00%	1.59%	0.85%
	대덕구	34849	2.08%	1.02%	2.82%	1.53%	3.32%	2.02%	3.63%	1.60%	1.81%	0.97%	2.36%	1.18%
	동구	37033	1.57%	1.21%	2.46%	1.11%	2.65%	1.94%	1.25%	0.52%	0.31%	0.13%	4.02%	2.06%
	서구	86540	4.50%	2.69%	3.68%	2.40%	2.93%	1.32%	4.17%	1.98%	3.84%	2.69%	2.41%	1.70%
	유성구	61456	4.57%	2.04%	2.49%	1.29%	4.22%	1.87%	5.82%	3.38%	2.03%	1.04%	5.42%	2.17%
	중구	41542	1.21%	0.95%	2.72%	1.60%	4.29%	1.72%	4.63%	1.99%	2.36%	1.65%	3.76%	2.84%
60대	서귀포시	36784	5.82%	2.33%	2.42%	1.33%	3.46%	2.27%	4.03%	2.21%	3.57%	2.31%	1.48%	1.06%
	제주시	76735	4.07%	1.82%	3.43%	1.83%	2.24%	1.34%	2.03%	1.01%	1.52%	0.61%	1.38%	0.55%
	대덕구	32298	3.29%	2.13%	1.02%	0.79%	6.40%	2.56%	3.82%	1.56%	3.02%	1.74%	1.54%	0.62%
	동구	48689	1.68%	0.76%	3.24%	1.39%	2.98%	2.02%	1.67%	1.16%	0.95%	0.70%	2.93%	1.77%
	서구	70191	2.81%	1.91%	2.95%	2.17%	2.36%	1.19%	2.61%	1.19%	3.18%	1.97%	1.88%	1.01%
	유성구	36689	2.05%	1.51%	2.09%	1.05%	6.42%	2.57%	4.03%	1.88%	3.02%	2.28%	4.55%	1.82%
	중구	52660	6.60%	2.70%	4.22%	1.94%	4.31%	2.00%	5.36%	2.16%	2.10%	1.44%	4.35%	2.64%

〈표 4-12〉 정렬방식에 따른 시군구단위 가구주 관계 추정오차 비교

변수	총계	[정렬 A]		[정렬 B]		[정렬 C]		[정렬 D]		[정렬 E]		[정렬 F]		
		최대 상대오차	평균 상대오차											
가구주 관계 — 자녀	서귀포시	44633	3.29%	1.72%	5.99%	2.39%	5.06%	3.40%	5.08%	2.47%	7.35%	3.42%	3.08%	1.79%
	제주시	147057	1.59%	0.75%	1.64%	0.75%	1.67%	0.91%	1.49%	1.09%	0.68%	0.36%	2.39%	1.09%
	대덕구	63273	2.38%	1.24%	2.29%	0.92%	3.33%	1.64%	2.63%	1.06%	0.55%	0.36%	2.65%	1.19%
	동구	72054	1.83%	1.02%	1.38%	0.73%	4.12%	1.65%	1.22%	0.68%	1.01%	0.40%	2.89%	1.79%
	서구	157228	4.79%	2.93%	4.78%	2.65%	3.76%	2.58%	4.55%	2.01%	3.51%	2.33%	3.76%	2.50%
	유성구	107013	4.45%	2.26%	2.61%	1.48%	3.29%	1.44%	3.71%	2.24%	2.25%	1.09%	2.66%	1.06%
	중구	78394	3.09%	1.23%	2.61%	1.81%	4.12%	1.65%	4.48%	1.95%	3.21%	1.59%	3.83%	2.28%
가구주 관계 — 부모	서귀포시	5465	5.40%	2.82%	3.57%	1.68%	5.76%	3.48%	4.67%	2.09%	3.75%	2.96%	3.20%	1.35%
	제주시	13317	2.54%	1.16%	3.14%	1.47%	2.72%	1.86%	3.18%	1.28%	4.00%	1.60%	3.84%	1.54%
	대덕구	5085	8.36%	3.50%	4.13%	2.12%	4.33%	2.48%	3.24%	2.36%	4.03%	1.85%	4.52%	2.75%
	동구	7049	7.36%	3.73%	1.65%	1.15%	8.10%	3.70%	4.62%	2.68%	2.68%	1.07%	5.33%	2.54%
	서구	12760	4.62%	2.30%	2.35%	1.07%	7.48%	4.01%	4.43%	1.82%	5.17%	2.35%	6.70%	2.68%
	유성구	8312	2.43%	1.33%	6.58%	2.69%	6.40%	3.00%	3.63%	2.09%	3.27%	2.10%	5.33%	3.80%
	중구	7880	3.68%	1.93%	5.71%	2.28%	6.54%	2.61%	8.12%	4.29%	4.57%	2.08%	4.31%	2.92%

V. 순환표본조사 추정 및 등록센서스 연계

1. 통계청 순환표본조사 및 추정 개요

인구주택총조사는 2010년까지 전수조사와 표본조사의 두 가지 방식이 혼합된 형태로 진행되어 왔다. 전수조사는 전체 가구와 인구를 대상으로 기본적 항목을 조사하는 반면, 표본조사는 전체 가구의 약 10%에 대해 추가적인 다양한 세부 항목에 걸쳐 조사를 수행하였다.

통계청은 2015년에 전수조사를 대신하여 행정자료를 활용한 인구, 가구 및 주택과 관련한 통계를 작성하는 ‘등록센서스’를 실시하였다. 최신의 주택공시자료, 건축물대장, 주민등록자료 등을 연계한 등록(register) 상의 주택 및 가구 단위명부에 2014년 11월 가구주택기초조사를 통해 등록되어 있지 않은 주택의 거처와 가구를 보완한 가구 단위의 표본추출틀을 구축하였다. 2016년 이후, 각종 행정자료를 입수하여 매년 구축하는 주택 DB, 가구 DB, 가구원 DB에 있는 주택 ID, 주소 ID, 세대주 보안식별번호를 이용하여 이미 구축된 표본틀과 연계하여 변동분을 반영하고 기본적 사항에 대해 공표할 수 있도록 하였다(박은영·김의영, 2014).

또한 표본조사를 통해 과거 전수조사에 포함되었던 항목 중 행정자료로 대체하기 곤란한 항목을 포함한 심층 항목을 조사하였다. 표본조사임에도 불구하고 기존 인구주택총조사의 자료수집을 위한 역할을 반영하여 등록센서스와 구분하여 현재 ‘2015년 인구주택총조사 표본조사’라고 지칭하는 듯하다. 특히 동읍면 단위의 행정자료를 보완하기 위해 기존 10% 표본조사 부분을 20%로 확대하여 조사를 실시하였고, 더불어 효율적인 모수 추정을 목적으로 기존 60가구 내외의 조사구 크기를 30가구 내

외로 축소 조정하였다.

이러한 전통적 인구주택총조사의 패러다임 전환과 맞물려 2021년부터는 표본조사에 대해 순환방식으로 수행할 것을 검토 중에 있다. 20% 표본가구 전체에 대하여 2015 인구주택총조사에는 특성시점(2015년 11월1일)을 조사기준으로 조사를 수행하였으나, 2021년부터는 표본가구를 매년 서로 겹치지 않도록 균등하게 분할하되 5년 주기로 일부를 추가하는 순환방식에 의해 조사를 수행하는 것을 고려하고 있다. 이는 조사수행의 효율성 제고는 물론 표본조사 항목의 매년 통계 생산에 대한 수요에 대응하기 위한 것으로 판단된다.

이에 따라 기존 5년을 주기로 한 시점추정은 물론 소지역 추정과 관련해서 필요한 경우 기간추정이라는 새로운 개념의 재정립 및 이에 따른 적절한 추정방법의 고려도 필요할 것이다. 가구와 가구원을 표본조사에 모두 포함함에 따라 개별 조사단위를 함께 아우르는 시점별 인구·가구 통합가중치의 산출을 고려할 수 있다. 특정시점의 조사에서 충분한 정도수준을 갖출 수 없는 소지역들에 대하여는 일정기간 동안의 조사결과를 누적하여 추정하는 기간추정(period estimation)의 의미도 검토될 수 있을 것이다. 더불어 기간추정 하의 모집단, 모수, 가중치 산출, 결과분석 등의 다양한 요소 등의 많은 이슈들도 포함될 수 있을 것이다.

통계청에서는 조사체계의 개선, 모수 추정방법 및 조사결과에의 정확성을 검토하기 위한 노력의 일환으로 순환표본방식에 의한 일련의 시험조사를 수행하고 있으며 향후 지속적인 후속 연구가 진행될 예정이다. 기존 1~2차 시험조사에서는 매해 등록센서스를 갱신하여 매년 모수를 새롭게 작성함에 따라 월주기로 수행하는 순환표본조사를 통해 자료의 시의성을 확보하는 방안을 검토하였다. 예로, 2차 시험조사는 '15년10월부터 17년9월까지 총 24개월에 걸쳐 진행되고 있다.

2016년 12월 현재 통계청은 20% 표본에 대해 매월조사가 아닌 매년조사로 변경하고 조사시점별 표본분할 방식으로 다음의 3가지 방안을 고려하고 있다.

방안1: 전체 조사구 중 매년 4%씩 조사

방안2: 매년 2%씩 조사하되, 등록센서스 연도에 10%를 추가로 조사

방안3: 매년 2.5%씩 조사하되, 등록센서스 연도에 7.5%를 추가로 조사

본 장에서는 기본적인 순환표본조사 추정방법들에 대해 통계청의 2차 시험조사를 기준으로 살펴보고자 한다. 2절에서는 시점별 인구·가구의 개별가중치와 통합가중치의 산출방안을 비교·검토한다. 3절에서는 소지역 추정을 위한 일반화추정방안을 살펴보고 이에 필요한 가중치 작성 방안에 대해 미국 지역사회조사(ACS) 방식에

견주어 검토하고자 한다. 4절에서는 2차 시험조사 자료를 이용한 통합가중치 산출방안과 소지역 추정을 수행하고 그 결과를 살펴본다. 마지막으로 5절에서는 순환표본조사 결과와 등록센서스 결과를 연계하는 방안에 대해 간략히 논의한다.

2. 인구·가구 통합가중치 연구

가구조사에서는 종종 표본가구를 추출한 후 가구와 가구 내 모든 적격 가구원을 조사하게 된다. 조사 후 응답한 가구와 가구원(즉, 인구)은 일련의 가중치 산출과정을 통해 해당 개체들이 모집단내에서 대표하는 개체수를 추정하게 된다. 만약 단위별로 모집단 보조정보, 예를 들면 지리적 혹은 인구사회학적 분류별 총 인구수 등의 보조정보가 존재한다면 이를 이용하여 해당 모집단 총합과 표본추정 총합이 같아지도록 가중치를 조정하는 칼리브레이션을 고려할 수 있다.

칼리브레이션 조정에는 사후층화(post-stratification), 레이킹 비 조정(raking ratio adjustment) 등과 같은 모형보조추정(model-assisted estimation)이나 조정 전 가중치와의 차이를 최소화하여 칼리브레이션 조건을 만족시키도록 조정하는 조정(calibration adjustment)을 고려할 수 있다. 이러한 가중치 조정은 가구와 인구에 대해 별도로 실시하여 개별적인 가중치로도 산출할 수도 있지만, 가구와 인구 정보를 함께 수집하는 조사에서 가구내 가구원들에게 동일한 값을 부여하는 통합가중(integrated weighting) 방식을 고려한다면 두 가지 단위 특성을 함께 고려할 수 있어 매우 편리하게 된다. 예를 들면, 개인소득 총합은 가구소득이므로 통합가중치를 사용하면 단위별 변수 간 구조가 유지될 수 있다.

가구와 인구 설계가중치는 표본추출에 의해 정해지고 통합가중치는 가중치 조정 단계에서 단위별 특성을 통합적으로 고려하여 얻게 된다. 통계청(2015)은 개별적인 사후층화방법과 Lemaitre and Dufour(1987)의 통합가중방법을 모의실험을 통해 비교 검토하였다. 개별적으로 가중치를 산출한다면 가구와 인구 특성과의 관계가 적절히 반영되지 못할 수 있을 것이다(Steel and Clark, 2007).

본 절에서는 조사단위 구조에 기초한 통합가중치 접근을 반영한 인구·가구 통합가중치 산출법에 대해 살펴본다.

1) 표본설계 및 설계가중치 산출

가구 j 의 특성을 $(y_{jc}, \mathbf{x}_{jc})$ 으로 표기하고, 가구내 가구원 k 의 특성은 $(y_{jk}, \mathbf{x}_{jk})$ 으로 각각 표기하자. 여기서 y 와 \mathbf{x} 는 조사변수와 보조변수를 나타내며 첨자 c 는 집락 (cluster)을 뜻한다. 모집단은 N 개 가구로 구성되며 가구내 구성원은 N_j 명이라고 가정한다면 모집단은 가구모집단 $U_I = \{j: j=1, \dots, N\}$ 와 가구내 가구원모집단 $U_j = \{k: k=1, \dots, N_j\}$ 의 2단 집락구조인 $U = \cup_{j \in U_I} U_j$ 으로 표기할 수 있다. 가구 모집단에서 n 개 가구를 표본설계 $p(s_I)$ 에 따라 확률추출하고 가구내 가구원을 전수 조사한다면 가구표본은 $s_I = \{j: j=1, \dots, n\}$ 이 되며 인구표본은 다음과 같이 표기될 수 있다.

$$s = \cup_{j \in s_I} U_j = \{(jk): k=1, \dots, N_j, j=1, \dots, n\}$$

가구의 추출확률이 π_j 이라면 해당 설계가중치는 $d_j = 1/\pi_j$ 으로 정의되며, 가구내 개인 추출확률은 모두 1 (즉, 가구내 개인 조건부 설계가중치는 $d_{klj} \equiv 1$)이므로 인구 설계가중치는 다음과 같이 정의된다.

$$d_{jk} = d_j, d_{klj} = d_j \quad (5-1)$$

2) 칼리브레이션 추정

설계가중치에 대해 주로 (단위)무응답 조정을 포함하여 보조적인 모집단 정보에 대해 벤치마킹하여 최종가중치가 산출된다. 하지만 논의를 단순화하기 위해 본 절에서는 무응답 조정은 생략하고 칼리브레이션 만을 고려하고자 한다.

가구와 인구별로 모집단 보조정보에 칼리브레이션 한다면 단위별 최종가중치는 각각 다음과 같은 가구와 인구 수준의 칼리브레이션 조건을 만족하게 된다.

$$\sum_{j \in s_t} w_j \mathbf{x}_{jc} = \sum_{j \in U_t} \mathbf{x}_{jc} \quad (5-2)$$

$$\sum_{(jk) \in s} w_{jk} \mathbf{x}_{jk} = \sum_{(jk) \in U} \mathbf{x}_{jk} \quad (5-3)$$

단위별 가중치에 대한 칼리브레이션을 단순히 칼리브레이션 혹은 비통합 칼리브레이션이라고 칭하고자 한다. 칼리브레이션 가중치 산출과정을 단계별로 정리하면 다음과 같다.

- 단계1: 가구 설계가중치 d_j 에 대해 가구수준 보조정보 \mathbf{x}_{jc} 를 사용하여 (5-2)식을 만족하도록 칼리브레이션 가중치 산출
- 단계2: 단계1과는 독립적으로 개인 설계가중치 $d_{jk} (= d_j d_{k|j})$ 에 대해 인구수준 보조정보 \mathbf{x}_{jk} 를 사용하여 식 (5-3)을 만족할 수 있도록 칼리브레이션 가중치 산출

반면, 가구와 가구원간의 관계는 물론 확률추출의 구조를 고려하여 단위별 보조정보에 대한 반영은 물론 설계가중치의 구조를 함께 통합하여 고려하는 가중치 조정을 통합 칼리브레이션이라 칭한다. 통합 칼리브레이션의 가중치 산출과정을 단계별로 기술하면 다음과 같다.

- 단계1: 식 (5-3)의 최종가중치가 $w_{jk} = d_{k|j} w_j$ 의 관계를 갖는 것으로 정의하고 인구수준 칼리브레이션 등식을 가구 칼리브레이션 가중치의 함수로 표현
- 단계2: 가구 및 인구 보조변수벡터를 다음과 같이 새롭게 정의함

$$\mathbf{x}_{jc}^* = \begin{pmatrix} \mathbf{x}_{jc} \\ \tilde{\mathbf{x}}_{j\cdot} \end{pmatrix} \quad (5-4)$$

여기서 $\tilde{\mathbf{x}}_{j\cdot}$ 는 j 번째 가구내 인구 특성의 총합임(예, 전수조사에서는 $\sum_{U_j} \mathbf{x}_{jk}$ 이고 표본조사에서는 $\sum_{s_j} d_{k|j} \mathbf{x}_{jk}$ 으로 정의)

- 단계3: 가구 설계가중치 d_j 에 대해 단계 2에서 새롭게 정의된 가구수준 보조정보 \mathbf{x}_{jc}^* 를 사용하여 다음의 칼리브레이션 조건식 (5-5)을 만족하도록 최종 가중치 산출

$$\sum_{j \in s_I} w_j \mathbf{x}_{jc}^* = \sum_{j \in U_I} \mathbf{x}_{jc}^* \quad (5-5)$$

여기서 식 (5-5)을 다시 표기하면 다음과 같음

$$\left(\begin{array}{c} \sum_{s_I} w_j \mathbf{x}_{jc} \\ \sum_{s_I} w_j \left(\sum_{s_j} d_{k|j} \mathbf{x}_{jk} \right) \end{array} \right) = \left(\begin{array}{c} \sum_{U_I} \mathbf{x}_{jc} \\ \sum_{U_I} \left(\sum_{U_i} \mathbf{x}_{jk} \right) \end{array} \right) \quad (5-5a)$$

단계4: 단계 3에서 산출된 가구 칼리브레이션 가중치로 단계 1에서 가정한 대로 인구 칼리브레이션 가중치를 다음과 같이 산출

$$w_{jk} = d_{k|j} w_j \quad (5-6)$$

통계청(2015)에서는 통합가중치 산출방안으로 Lemaître & Dufour (1987)이 제안한 방법론을 고려하였다. 이는 “동일분담(equal share) 보조정보” 방식이라고 불리우는데 개별 가구원 보조변수 \mathbf{x}_{jk} 를 직접 사용하는 대신에 다음과 같이 정의되는 가구수준 평균값으로 인구 보조변수로 취하는, 즉 $\mathbf{x}_{jk}^* := \bar{\mathbf{x}}_j$. 으로 인구수준 칼리브레이션을 실행한다. 여기서 가구내 평균적 인구특성은 다음과 같이 표기될 수 있다.

$$\bar{\mathbf{x}}_j = \frac{1}{N_j} \sum_{U_j} \mathbf{x}_{jk} \quad (5-7)$$

따라서 앞서의 통합가중치는 인구정보를 가구수준으로 통합하여 가구기준으로 통합하였다면 후자는 인구수준 보조정보를 해당 가구평균을 취하는 동일분담 보조정보 통합방식을 통해 개인수준의 칼리브레이션을 수행하는 것이다.

3) 일반화회귀추정

일반화회귀추정은 관심변수 y 와 보조변수 $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ 의 선형적 연관성을 이용한 추정방법으로 추정식에서 선형형태로 정의되는 회귀가중치(regression

weight)는 사후층화 가중치나 레이킹 비 가중치를 특수한 형태로 포함하게 되며 또한 보조변수에 대한 식 (5-2)나 식 (5-3)과 같은 칼리브레이션 조건을 만족하게 한다.

먼저 모총합 $Y = \sum_U y_i$ 에 대한 일반회귀추정량은 다음의 형태로 정의된다.

$$\hat{Y}_{gr} = \sum_{i \in s} w_i y_i = \sum_{i \in s} d_i g_i y_i = \hat{Y}_\pi + (\mathbf{X} - \hat{\mathbf{X}}_\pi)' \hat{\mathbf{B}} \quad (5-8)$$

여기서 회귀가중치는 다음과 같이 정의되는데

$$w_i = d_i g_i$$

이는 추출확률 π_i 의 역수로 정의되는 설계가중치 $d_i = 1/\pi_i$ 에 다음의 g -조정계수가 곱하여져 정의된다.

$$g_i = 1 + (\mathbf{X} - \hat{\mathbf{X}}_\pi) \left(\sum_s d_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_i \quad (5-9)$$

또한 $\hat{Y}_\pi = \sum_s d_i y_i$ 와 $\hat{\mathbf{X}}_\pi = \sum_s d_i \mathbf{x}_i$ 는 각각 모총합 Y 와 \mathbf{X} 에 대한 설계가중치를 기준으로 계산되는 HT-추정량(Horvitz-Thompson estimator)이고, 추정된 회귀 계수 $\hat{\mathbf{B}}$ 는 다음과 같이 정의된다.

$$\hat{\mathbf{B}} = \left(\sum_s d_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_s d_i \mathbf{x}_i y_i \quad (5-10)$$

식 (5-8)을 사용하면 보조변수 \mathbf{x} 에 대한 표본총합추정량 $\hat{\mathbf{X}}_{gr}$ 과 모수 \mathbf{X} 는 다음의 칼리브레이션 성질을 갖게 된다.

$$\sum_s w_i \mathbf{x}_i = \mathbf{X} \quad (5-11)$$

모총합의 회귀추정량의 근사 분산추정량은 다음과 같이 정의된다(Särndal et al., 1992).

$$v(\hat{Y}_{gr}) = \sum_{i \in s} \sum_{j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \left(\frac{g_i e_i}{\pi_i} \right) \left(\frac{g_j e_j}{\pi_j} \right) \quad (5-12)$$

여기서 π_{ij} 는 두 개체 i 와 j 의 결합포함확률(joint inclusion probability)이고, $e_i = y_i - \hat{y}_i$ 은 잔차이며, $\hat{y}_i = \mathbf{x}_i \hat{\mathbf{B}}$ 는 개체 i 의 회귀추정값이다. 모총합의 HT-추정량의 분산추정식 $v(\hat{Y}_\pi)$ 은 식 (5-12)에서 $g_i e_i$ 대신에 y_i 이 대체되어 정의된다. 따라서 보조변수 \mathbf{x} 와 조사변수 y 간의 선형적 연관성이 높다면 $v(\hat{Y}_\pi)$ 에 비해 $v(\hat{Y}_{gr})$ 이 작은 값을 갖게 됨을 알 수 있다.

가구와 인구 회귀가중치는 해당 보조변수 \mathbf{x}_{jc} 와 \mathbf{x}_{jk} 을 각각 사용하여 회귀조정계수를 산출하면 된다. 먼저, 가구수준 보조변수 모총합과 표본추정량을 $\mathbf{X}_c = \sum_{U_I} \mathbf{x}_{jc}$ 과 $\hat{\mathbf{X}}_{c\pi} = \sum_{s_I} d_j \mathbf{x}_{jc}$ 으로 정의하고, 인구수준 보조변수 모총합과 표본추정량을 $\mathbf{X} = \sum_U \mathbf{x}_{jk}$ 과 $\hat{\mathbf{X}}_\pi = \sum_s d_{jk} \mathbf{x}_{jk}$ 으로 정의하자. 단위별로 식 (5-9)를 적용한 g-조정계수를 산출하면 다음과 같다.

$$g_j = 1 + (\mathbf{X}_c - \hat{\mathbf{X}}_{c\pi})' \left(\sum_{j \in s_I} d_j \mathbf{x}_{jc} \mathbf{x}_{jc}' \right)^{-1} \mathbf{x}_{jc} \quad (5-13)$$

$$g_{jk} = 1 + (\mathbf{X} - \hat{\mathbf{X}}_\pi)' \left(\sum_{j \in s} d_{jk} \mathbf{x}_{jk} \mathbf{x}_{jk}' \right)^{-1} \mathbf{x}_{jk} \quad (5-14)$$

또한 이들을 설계가중치에 곱하면 가구 회귀가중치 $w_j^{NI} = d_j g_{jc}$ 와 인구 회귀가중치 $w_{jk}^{NI} = d_{jk} g_{jk}$ 가 각각 산출된다.

반면 인구·가구 통합 회귀가중치는 식 (5-4)의 가구수준 인구·가구 통합특성 보조변수 \mathbf{x}_{jc}^* 를 사용하여 가구수준 g-조정계수를 산출하고, 이를 가구 설계가중치에 곱하여 가구와 인구 각각의 통합 회귀가중치를 계산하면 된다. 먼저, 가구와 인구의 통합특성에 대한 보조변수 모총합과 표본추정량을 각각 $\mathbf{X}_c^* = \sum_U \mathbf{x}_{jc}^*$ 과 $\hat{\mathbf{X}}_{c\pi}^* = \sum_{s_I} d_j \mathbf{x}_{jc}^*$ 으로 정의하고, 가구통합수준 g-조정계수를 다음과 같이 산출한다.

$$g_{jc}^{IC} = 1 + (\mathbf{X}_c^* - \hat{\mathbf{X}}_{c\pi}^*)' \left(\sum_{j \in s_I} d_j \mathbf{x}_{jc}^* \mathbf{x}_{jc}^{*'} \right)^{-1} \mathbf{x}_{jc}^* \quad (5-15)$$

가구 설계가중치에 가구 통합수준 g -조정계수를 곱하여 가구통합 회귀가중치와 인구 통합회귀가중치를 $w_j^{IC} = d_j g_{jc}^{IC}$ 와 $w_{jk}^{IC} = d_j g_{jc}^{IC} (= w_{jc}^{IC})$ 으로 각각 산출한다.

3. 소지역 추정을 포함한 효율적 추정방안

1) 소지역 추정을 위한 기본적 고려사항

2015년 인구주택총조사 표본조사의 목적은 행정자료에 기반한 ‘등록센서스’에 포함되지 못한 항목들을 보완하고 동읍면 단위의 통계산출을 가능하게 하는데 있다. 2021년 이후부터 5년에 걸친 단위 기간으로 균등하게 나누어 조사하는 순환표본으로의 전환을 고려함에 따라 2015년과는 달리 동읍면은 물론 시군구와 같은 상위 지역단위의 추정도 소지역 추정(small area estimation)로 접근해야 할 수도 있다. 따라서 특정시점에서 충분한 표본을 확보하지 못하게 되면 지역수준별 추정은 소지역 추정을 고려하여 추정치의 안정성을 확보하여야 될 것이다. 소지역 추정을 위해 5장 2절에서 살펴본 일반화회귀추정이 고려될 수 있지만 이 역시 적절한 표본규모와 해당 지역의 적절한 보조정보가 있어야 함을 전제로 한다.

미국 ACS조사에서는 누적시점 자료를 통합하여 표본수를 확보하는 기간추정 개념을 도입함으로 소지역 추정을 고려하고 있다. 예로, ACS조사에서는 주(state), 카운티(county), 인구 65,000명 이상 지역에 대해서 1년 표본을 통합하는 1년 추정을 적용하고 있다. 10년 주기 센서스의 전통적 상세(long-form, LF) 설문조사를 위한 가구추출률은 1/6(16.7%)이고, ACS조사의 가구추출률은 월조사 1/480(0.2%), 연간조사 1/40(2.5%), 5년조사 1/8(12.5%) 정도이다. 따라서 long-form 조사를 대체하고자 수행하는 ACS조사에서 long-form조사를 대체할 수 있는 적절한 표본크기는 5년 추정치로서 가능할 수 있다 (Fay, 2006).

ACS조사의 5년 추정치를 위한 가중치 산출은 1년 추정치의 기초가중치를 5로 나눈 후, 카운티를 가중지역(weighting area)으로 무응답 조정을 수행하고 카운티 수준의 표본추출률 총합에 대한 레이킹-비 조정과 같은 칼리브레이션을 적용하였으나, 10년의 긴 주기에서 오는 센서스 표본추출률과 ACS 표본 사이의 괴리로 말미암아 포함오차는 물론 분산의 증가를 초래하였다. 이에 따라 ACS에서도 행정정보의

이용을 검토하고, 센서스트랙(census track)이나 블록그룹(block group)과 같은 소지역에 대한 추정의 정도수준 향상을 위해 일반화회귀추정방법을 적용하였다. ACS 조사의 일반화회귀추정방법의 사용은 5년 추정을 위한 가중치 산출에만 적용하고 있으며 센서스트랙 수준 통계량의 분산감소만을 목적으로 한다(Fay, 2005, 2006).

따라서 ACS의 경험을 비추어 볼 때, 소지역 통계추정을 위해서는 (i) 충분한 표본수를 확보하여야 하며, (ii) 해당지역의 모총합 정보 자료가 존재하고 적절한 칼리브레이션 방법을 선택하여 적용하는 것이 필요하다고 할 수 있다.

2) 2차 제주도 시험조사에 기초한 소지역 추정 접근방식

순환센서스 제2차 시험조사의 대상은 제주특별자치도이며, '15년 10월부터 '17년 9월까지 총 24개월간에 걸쳐 조사가 진행 중이다. 매월 조사하는 순환조사를 고려하여 특정지역을 여러 시점에 걸쳐 중복되지 않게 반복하여 조사하고, 그 결과를 일정기간(또는 규모)을 누적하여 분석할 수 있도록 하며, 6개월을 1년으로 간주하여 4년을 실시하는 일종의 가상적 개념을 가정하였다.

시험조사를 위한 표본추출틀로는 2015년 인구주택총조사를 위해 설정한 조사구 중 특성이 아파트(A), 일반(1), 섬(2) 형태의 8,027개 조사구 명부를 사용하였는데 <표 5-1>은 제주도 조사구 모집단 현황을 정리하고 있다.

<표 5-1> 제주도 조사구 모집단 현황

구분	조사구특성							
	계	일반조사구				시설조사구		
		소계	A	1	2	소계	3	4
조사구수	8,207	8,027	2,230	5,733	64	178	77	101
가구수	229,761	229,583	60,089	167,658	1,836	178	77	101

출처: 통계청 표본과(2015.9.23.) 제2차 시험조사 표본추출 결과 보고서

표본규모로 24개월간 2015년 인구주택총조사 표본조사와 동일한 규모의 표본이 되도록 약 20%에 가까운 1,584개 조사구를 추출하였다. 동읍면별로 두 가지 조사구 특성인 아파트(A)와 기타(1,2)의 구분에 따라 층화한 후, 2015년 인구주택총조

사 20% 표본조사와 마찬가지로 층별로 조사구 세부특성, 단독비율, 조사구번호 순으로 정렬하여 계통추출하는 내재적 층화를 적용하였다. <표 5-2>는 6개월 단위 중 처음 2개 그룹인 A·B에 대해 월 단위 차수별 조사구 특성과 시 구분에 따른 표본조사구 현황을 정리하고 있다. <표 5-3>은 6개월 단위 2개 그룹 A·B에 대해 일부 읍면동 및 차수별 표본조사구 현황을 보여주고 있다.

<표 5-2> 제주도 2차 시험조사 A·B 그룹 차수별 표본조사구 현황

그룹 (6월단위)	차수 (월단위)	연월	전체	특성별		시군구별	
				아파트	기타	제주시	서귀포시
합계			1,584	386	1,198	1,060	524
A	1	'15.10	66	16	50	45	21
	2	11	66	17	49	44	22
	3	12	66	16	50	44	22
	4	'16. 1	66	16	50	44	22
	5	2	65	16	49	44	21
	6	3	66	16	50	44	22
B	7	4	66	16	50	44	22
	8	5	66	16	50	44	22
	9	6	66	16	50	44	22
	10	7	66	16	50	44	22
	11	8	66	16	50	44	22
	12	9	66	16	50	45	21

출처: 통계청 표본과(2015.9.23.) 제2차 시험조사 표본추출 결과 보고서

<표 5-3> 제주도 2차 시험조사 A·B 그룹 일부 읍면동별 표본조사구

구분	합계	A						B					
		1	2	3	4	5	6	1	2	3	4	5	6
제주	1,594	66	66	66	66	66	66	66	66	66	66	66	66
제주시	1,060	45	44	44	44	44	44	44	44	44	44	44	45
한림읍	40	2	1	2	1	2	2	2	1	2	1	2	2
애월읍	52	2	3	2	2	2	2	2	3	2	2	2	2
구좌읍	34	2	1	2	1	1	2	1	1	2	1	1	2
서귀포시	524	21	22	22	22	22	22	22	22	22	22	22	21
대정읍	36	2	1	2	2	1	1	2	1	2	2	1	1
남원읍	36	1	2	1	1	2	2	1	2	1	1	2	2
성산읍	34	1	1	2	1	1	2	2	1	2	1	1	2

출처: 통계청 표본과(2015.9.23.) 제2차 시험조사 표본추출 결과 보고서

월 단위 차수별 평균적 표본크기는 제주도 전체는 66개 조사구의 약 1,980가구로 이루어져 있고 조사구 특성별로는 아파트 조사구 16개의 약 480개 가구와 기타 조사구의 1,470개 가구와 시군구별로 제주시 조사구 44개의 1,320개 가구와 서귀포시 조사구 22개의 550개 가구로 구성된다. 또한 동읍면별 표본크기는 차수별로 1~2개 혹은 2~3개 조사구의 30~60개 가구 혹은 60~90개 가구로 구성된다.

2차 시험조사의 월 단위 표본에 대한 시군구별 표본추출률을 살펴보면 <표 5-4>와 같다. 제주도, 제주시와 서귀포시는 각각 표본추출률이 0.9%, 0.7%, 1.1%로 미국 ACS조사의 월 단위 표본의 0.2%보다는 많이 높음을 알 수 있다. 하지만 조사구내 모든 가구를 조사하는 일단집락추출을 적용함에 따라 설계효과가 존재하게 된다. 조사구내 가구와 인구 동질성(rate of homogeneity, roh)을 각각 0.02로 가정하면 Kish(1965)의 집락효과모형 $1 + (M-1)roh$ 을 적용하면 설계효과는 1.6 정도가 되며 이를 감안한 유효표본수(effective sample size)는 제주도 기준 표본가구수는 1,980개에서 유효표본수는 1,238개로 작아진다. 유효표본수를 적용한 제주도, 제주시, 서귀포시 월 표본의 (실질적) 표본추출률은 각각 0.5%, 0.5%, 0.7%로 작다. 미국 ACS조사의 표본설계에 집락추출을 적용하지 않고 있으므로 우리나라 2차 시험조사와 같은 집락효과는 존재하지 않는다.

<표 5-4> 시험조사 월 표본규모 및 표본추출률

지역	모집단		표본			유효표본	
	조사구	가구	조사구	가구	추출률	가구	추출률
제주도	8,027	240,810	66	1,980	0.9%	1,238	0.5%
제주시	6,044	181,320	44	1,320	0.7%	825	0.5%
서귀포시	1,983	59,490	22	660	1.1%	413	0.7%

통계청에서는 현재 2차 시험조사 표본설계 당시의 표본설계방침과는 달리 순환조사의 주기를 연 단위를 변경할 것을 고려중이며 구체적으로는 5장 1절에 언급한 3가지 방안 중에서 하나를 채택할 것을 고려하고 있다.

변경된 통계청의 순환조사 주기를 기준으로 볼 때 앞서 살펴본 월 표본의 표본추출률은 매우 작음을 알 수 있다. 반면 분기와 반기 표본을 통합하게 되면 제주도 전

체를 기준으로 표본추출률은 각각 2.5%와 5.0%이 된다. <표 5-5>와 <표 5-6>는 2차 시험조사에 대해 분기와 반기 단위로 표본을 통합하였을 때의 표본수와 표본추출률을 정리해 주고 있다. 따라서 반기는 통계청에서 현재 고려중인 방안 방안 (i)에 견줄 수 있는 표본규모가 되며, 분기는 등록센서스 연도에 추가되는 표본을 제외한 균등순환의 표본규모는 방안 (iii)과 유사하다고 할 수 있다.

<표 5-5> 시험조사 분기 표본규모 및 표본추출률

지역	모집단		표본			유효표본	
	조사구	가구	조사구	가구	추출률	가구	추출률
제주도	8,027	240,810	198	5,940	2.5%	3,713	1.5%
제주시	6,044	181,320	132	3,960	2.2%	2,475	1.4%
서귀포시	1,983	59,490	66	1,980	3.3%	1,238	2.1%

<표 5-6> 시험조사 반기 표본규모 및 표본추출률

지역	모집단		표본			유효표본	
	조사구	가구	조사구	가구	추출률	가구	추출률
제주도	8,027	240,810	396	11,880	5.0%	7,425	3.1%
제주시	6,044	181,320	264	7,920	4.4%	4,950	2.7%
서귀포시	1,983	59,490	132	3,960	6.7%	2,475	4.2%

이러한 점들을 볼 때, 현재 통계청에서 고려하는 연 단위 주기의 순환조사는 분기 혹은 반기 수준 규모의 표본통합을 고려하는 것이 바람직할 것으로 보인다. 다음 절에서는 논의의 단순화를 위해 분기 통합에 해당하는 2.5% 표본만을 고려하여 2차 시험조사 표본을 기준으로 통합가중치 산출과 소지역 추정에 대해 시산한 결과를 논하고자 한다.

4. 2차 시험조사 자료를 이용한 통합가중치 및 소지역 추정시산

1) 인구·가구 통합가중치 시산

본 소절에서는 2차 시험조사의 첫 3개월 표본(A-그룹내 1, 3, 4차의 총 3개월)에 대해서 적용하였다. 여기서 A-그룹내 2차는 2015년 11월에 해당하는 표본으로 2015년 인구주택총조사와 중복되는 표본조사구로 대체하였으므로 본 연구의 분석에서는 제외하였다. 칼리브레이션 조정을 위해 고려한 단위지역은 시군구인 제주시와 서귀포시이다.

먼저 2차 시험조사의 표본설계 내역을 간단히 살펴보고자 한다. 조사대상지역은 제주특별자치도내 제주시와 서귀포시의 모든 동읍면이고, 표본추출틀은 2015년 인구주택총조사를 위해 구축한 조사구 명부를 사용하였다. 동읍면별로 조사구 특성에 따라 아파트와 일반 조사구를 두 개 층으로 구분한 뒤, 층별로 할당된 표본조사구를 계통추출로 선택하고 조사구내 모든 가구와 가구내 모든 가구원을 조사하였다. 분석을 단순화하기 위해 가구내 모든 가구원이 응답한 가구와 가구원만을 시산에 포함하였다.

설계가중치는 추출확률의 역수로 정의된다. 특정 동읍면의 h 번째 층내 총 조사구와 표본조사구 수가 각각 N_h 와 n_h 이라면 i 번째 조사구 설계가중치는 다음과 같다.

$$d_{hi} = \frac{N_h}{n_h}$$

추출확률의 구조상 조사구내 가구, 가구내 가구원이 모두 조사에 포함되어 조건부 추출확률은 각각 1이 되며 단위별 가중치 값은 $d_{hi} = d_{hij} = d_{hijk}$ 으로 서로 같은 값을 갖게 된다. 만약 조사수행에 따른 무응답을 고려한다면, 조사구내 총가구 대비 응답가구와 가구내 총가구원 대비 응답가구원 수를 반영하는 가장 단순한 형태의 무응답 조정 가중치를 고려할 도 있을 것이다. 단위별 설계가중치에 대한 무응답 조정과 관련한 방법론은 기존 연구인 통계청(2015, 2장)에서 다루고 있어서 본 연구에서는 따로 고려하지 않았다.

2015 인구주택총조사 표본추출틀에 제공되는 가구부분의 거처종류와 가구원수, 그리고 인구부분의 성·연령 정보를 활용하여 가구와 인구 설계가중치 d_{hij} 과 d_{hijk} 에 대해 단위별로 g-조정계수 및 회귀가중치를 산출하였고, 더불어 인구·가구 통합 기준의 g-조정계수와 회귀가중치를 각각 산출하였다. 이때 가구와 인구 수준의 보조변수의 상세한 구분은 다음과 같다.

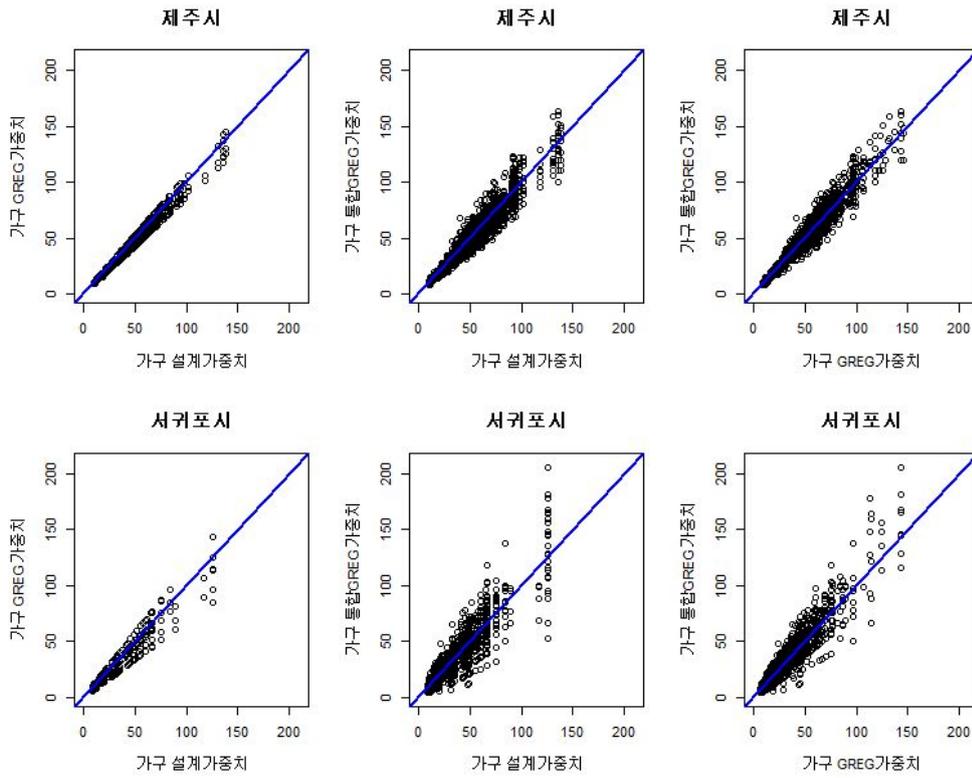
가구부분: 거처종류 (단독주택 거주 및 비단독주택 거주 여부)
 가구원수 (1인, 2인, 3인 이상 가구)
 인구부분: 성·연령 (남·여 × 0-19세, 20-39세, 40-59세, 60세 이상)

<표 5-7>은 2차 시험조사 2.5% (혹은 분기) 표본 가구들의 설계가중치, GREG 가중치, 통합 GREG 가중치, g-조정계수, 통합 g-조정계수의 기술통계량을 정리해주고 있다. 제주도 가구 설계가중치의 총합 171,897, 평균값 52.3, 표준편차 23.1이며, 최소 10.3에서 최대 137.9의 범위를 갖는다. 또한 균등가중치를 사용하지 않고 불균등가중치를 사용에 따른 평균추정량의 분산 증가분($L_w = 1 + cv_w^2$)은 1.195이다. GREG 가중치와 통합 GREG 가중치는 각각 총합이 164,617의 일치된 값을 보이는데 이는 설계가중치 합이 칼리브레이션 조건을 갖게 됨을 보여준다. GREG 가중치와 통합 GREG 가중치는 각각 설계가중치에 비해 평균값이 조금 낮아졌지만 최소값과 최대값의 범위는 늘어나는 방향으로 조정되었음을 알 수 있다. 하지만 GREG 가중치에 비해 통합 GREG 가중치는 그 범위가 조금 더 넓어졌고 분산 증가도 다소 커졌음을 알 수 있다. 또한 g-조정계수는 비통합과 통합 모두 평균이나 중위수 모두 1보다 작게 나타나고 있다. 서귀포시 가구 가중치와 g-조정계수는 제주도 가중치와 g-조정계수들과 대체적으로 유사한 분포를 보이지만 제주시에 비해 더 넓은 범위를 갖는다.

[그림 5-1]과 <표 5-8>는 각각 도시별 가구들의 설계, GREG, 통합 GREG 가중치들의 산점도와 피어슨 선형상관계수를 나타내고 있다. 앞서 살펴본 가중치별 분포와 비슷하게 GREG 가중치보다 통합 GREG 가중치가 설계가중치와의 상관관계가 낮음을 볼 수 있다. 특히 제주시에 비해 서귀포시의 가중치간 상관계수가 더 작음을 알 수 있다. 이러한 결과는 설계가중치에 대한 g-조정계수가 GREG보다 통합 GREG에서 더 다양한 값을 갖기 때문인 듯하다. [그림 5-2]는 설계가중치 대비 g-조정계수와 통합 g-조정계수가 갖는 분포를 도시별로 보여주고 있다. 제주시에 비해 서귀포시의 g-조정계수의 범위가 더 넓게 나타나고 있다.

<표 5-7> 2차 시험조사 가구가중치 및 g-조정계수 분포

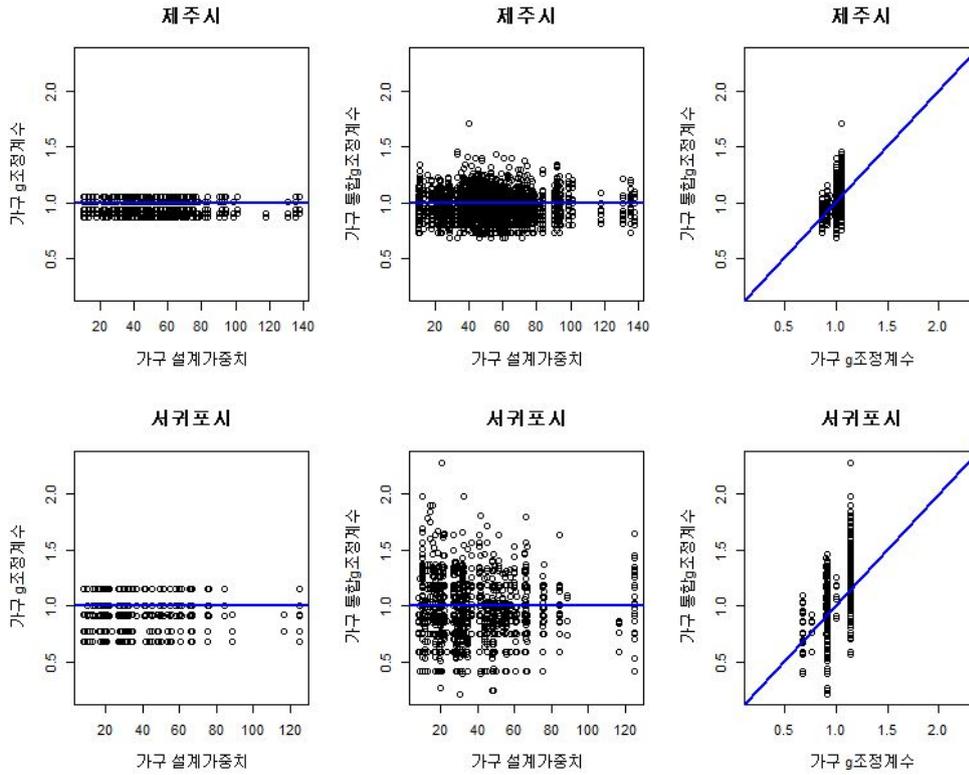
시군구	가구가중치	통계량								
		총합	평균	표준편차	최솟값	1분위수	중위수	3분위수	최댓값	L_w
제주시	설계가중치	171,897	52.3	23.1	10.3	36.9	50.8	63.9	137.9	1.195
	GREG가중치	164,617	50.1	22.2	9.1	35.2	49.9	61.8	145.1	1.196
	통합GREG가중치	164,617	50.1	23.2	7.7	35.1	48.3	61.9	163.5	1.215
	g-조정계수	-	0.959	0.064	0.864	0.909	0.942	1.008	1.052	-
	통합g-조정계수	-	0.956	0.119	0.685	0.882	0.950	1.034	1.704	-
서귀포시	설계가중치	62,053	38.1	23.7	8.3	19.7	32.2	51.6	124.9	1.387
	GREG가중치	57,965	35.6	22.9	5.6	18.3	30.8	46.9	143.3	1.414
	통합GREG가중치	57,965	35.6	24.8	3.6	17.1	29.8	47.8	205.9	1.487
	g-조정계수	-	0.939	0.140	0.678	0.908	0.917	0.999	1.147	-
	통합g-조정계수	-	0.942	0.263	0.211	0.743	0.910	1.159	2.281	-



[그림 5-1] 2차 시험조사 가구 설계가중치 및 GREG 가중치

<표 5-8> 2차 시험조사 가구가중치 상관계수

도시	상관계수	설계가중치	GREG가중치	통합GREG가중치
제주시	설계가중치	1.0000	0.9856	0.9561
	GREG가중치	0.9856	1.0000	0.9695
	통합GREG가중치	0.9561	0.9695	1.0000
서귀포시	설계가중치	1.0000	0.9598	0.8922
	GREG가중치	0.9598	1.0000	0.9270
	통합GREG가중치	0.8922	0.9270	1.0000

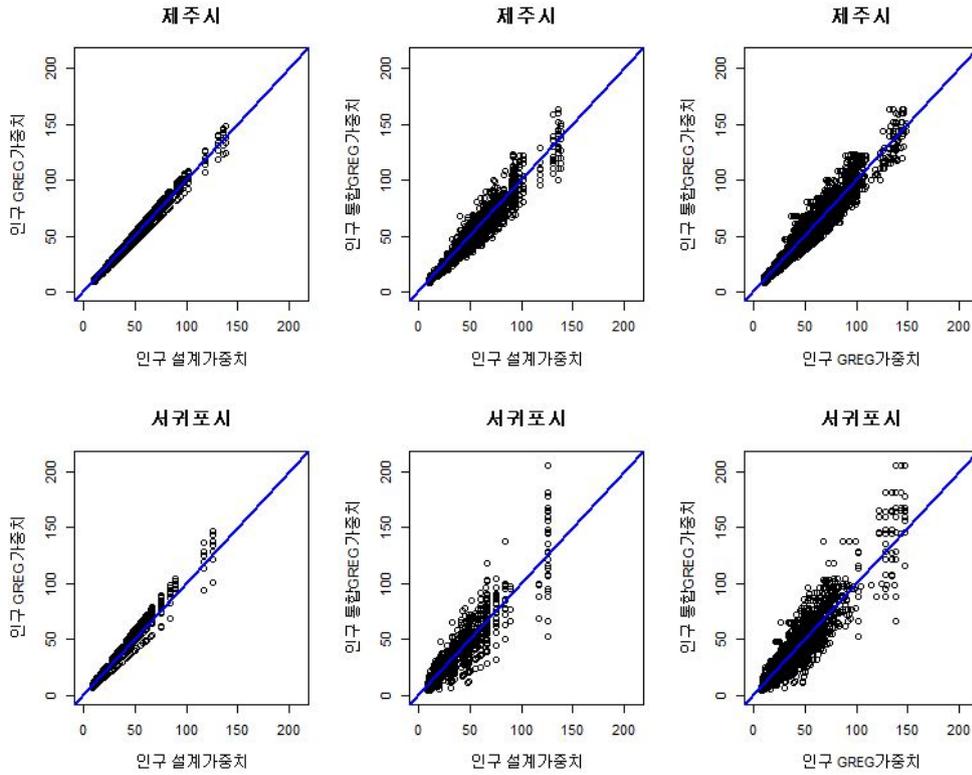


[그림 5-2] 2차 시험조사 가구 설계가중치 및 g-조정계수

<표 5-9>은 2차 시험조사 도시별 인구 설계가중치, GREG 가중치, 통합 GREG 가중치, g-조정계수, 통합 g-조정계수에 대한 기술통계량을 정리하고 있다. 가구내 모든 가구원을 조사하는 집락추출임으로 인구가중치 분포는 가구가중치의 분포와 거의 유사한 경향을 보인다. 가구가중치에서와 마찬가지로 제주시에 비해 서귀포시 인구가중치의 범위와 불균등성에 따른 분산증가분도 더 큼을 알 수 있다. [그림 5-3], [그림 5-4], <표 5-10>은 도시별 가중치간 산점도, 설계가중치와 g-조정계수간 산점도, 가중치간 상관계수를 각각 나타내고 있는데, 이 역시 가구가중치와 유사한 경향을 보여준다.

<표 5-9> 2차 시험조사 인구가중치 및 g-조정계수 분포

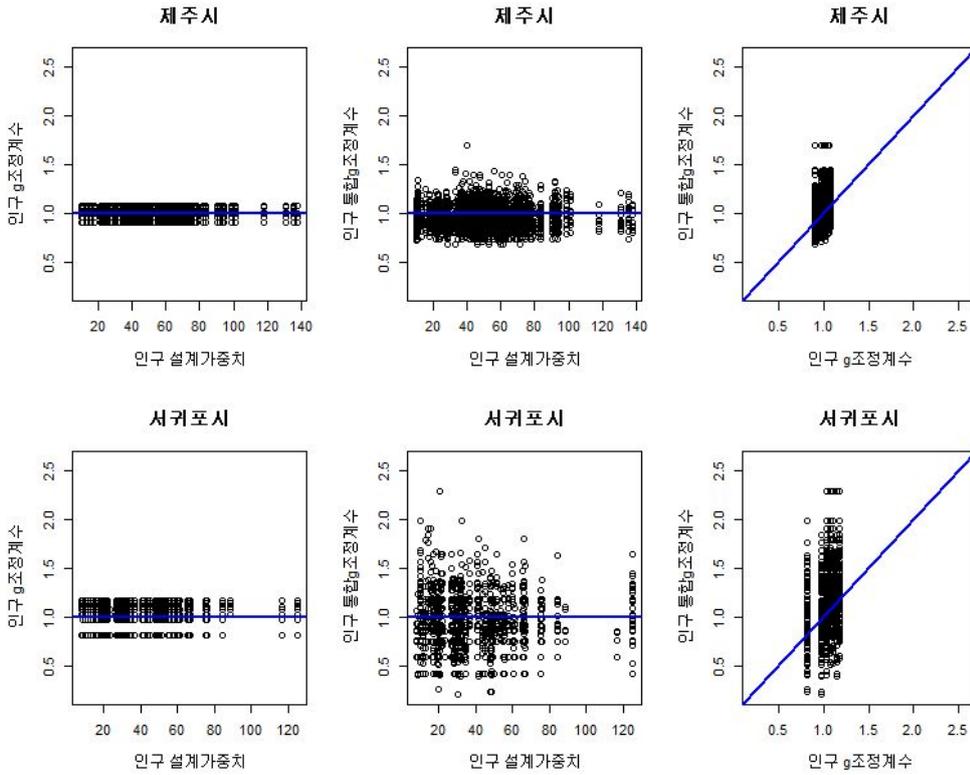
시군구	개인가중치	통계량								
		총합	평균	표준편차	최솟값	1분위수	중위수	3분위수	최댓값	L_w
제주시	설계가중치	434,800	52.2	22.3	10.3	39.5	52.2	63.1	137.9	1.182
	GREG가중치	436,034	52.3	22.7	9.4	39.0	51.7	64.2	148.5	1.188
	통합GREG가중치	431,674	51.8	23.4	7.7	38.0	50.7	63.6	163.5	1.203
	g-조정계수	-	1.001	0.059	0.907	0.967	1.001	1.051	1.077	-
	통합g-조정계수	-	0.992	0.132	0.685	0.895	0.989	1.077	1.704	-
서귀포시	설계가중치	149,328	38.1	23.0	8.3	19.7	32.9	49.9	124.9	1.366
	GREG가중치	148,684	37.9	23.7	6.7	18.9	33.8	52.2	146.8	1.389
	통합GREG가중치	147,304	37.5	25.9	3.6	18.9	31.6	49.7	205.9	1.477
	g-조정계수	-	0.996	0.132	0.809	0.809	1.027	1.106	1.175	-
	통합g-조정계수	-	0.996	0.294	0.211	0.779	0.975	1.182	2.281	-



[그림 5-3] 2차 시험조사 인구 설계가중치 및 GREG 가중치

<표 5-10> 2차 시험조사 인구가중치 상관계수

도시	상관계수	설계가중치	GREG가중치	통합GREG가중치
제주시	설계가중치	1.0000	0.9893	0.9488
	GREG가중치	0.9893	1.0000	0.9652
	통합GREG가중치	0.9488	0.9652	1.0000
서귀포시	설계가중치	1.0000	0.9680	0.8794
	GREG가중치	0.9598	1.0000	0.9151
	통합GREG가중치	0.8794	0.9151	1.0000



[그림 5-4] 2차 시험조사 인구 설계가중치 및 g-조정계수

앞서 선택한 보조변수를 활용한 GREG 가중치와 통합 GREG 가중치에 기초한 몇 가지 가구와 인구특성의 총합추정 결과에 대해 살펴보고자 한다. <표 5-11>는 가구원수 총합추정결과를 나타내며 <표 5-12>은 주택유형, <표 5-13A>와 <표 5-13B>는 남성과 여성의 연령별 인구 총합추정의 결과를 정리해 주고 있다. 이 3가지 특성은 앞서 나열한 범주로 칼리브레이션 조정에 고려한 것들이다.

먼저 가구원수를 살펴보면 1인과 2인은 칼리브레이션 조정에 사용한 범주와 일치하므로 GREG과 통합 GREG의 총합추정치 모두 같은 값을 보여준다. 하지만 3인, 4인, 5인 이상의 경우에는 칼리브레이션 조정에서 '3인 이상'이라는 하나의 큰 범주에서 조정되어 GREG과 통합 GREG 총합추정치의 결과는 다소 다르게 나타난다. 3인 이상의 범주로 묶어서 고려하였던 이유는 표본크기가 작아 보조변수가 적절히 정의되지 않거나 g-조정계수내 역행렬이 존재하지 않는 경우가 발생하기 때문이었다.

제주시와 서귀포시 모두의 GREG 총합추정치는 3인 가구가 많게 조정되어졌다면 통합 GREG 총합추정치는 4인과 5인 이상 가구가 더 많도록 조정되었음을 알 수 있다.

주택유형의 경우에는 칼리브레이션 조정에서 단독과 비단독의 두 범주가 고려되었다. GREG 총합추정치는 제주시의 아파트와 연립주택 규모가 미세하나마 서로 반대방향으로 조정됨을 볼 수 있다.

성·연령의 칼리브레이션 조정에서 대략 20세 간격으로 묶어서 고려되었다. 서귀포시는 GREG 총합추정치는 40대와 50대 크기를 유사하게 조정한 반면에, 통합 GREG 추정치는 40대에 비해 50대가 더 많은 것으로 조정되었다. 이러한 미세한 조정차이는 고려된 보조변수의 다변량적 관계에 따라 가구와 인구 특성이 반영되었으며 관심변수와의 관계에 따라 GREG 가중치와 통합 GREG 가중치에 기반한 총합 추정치는 달라질 수 있음을 알 수 있다. <표 5-45>-<표 5-21>은 주거전용·영업겸용여부, 점유형태, 교육정도, 수학정도, 종교유무, 출생지 등 몇 가지 조사변수에 대한 GREG 추정과 통합 GREG 추정결과를 포함하였다.

두 개의 단위특성을 함께 통합적으로 고려한 통합 GREG 가중치는 그렇지 않은 개별 GREG 가중치보다 다변량적 특성을 잘 보전할 것으로 예상되지만 이에 대한 연구는 좀 더 필요할 것으로 판단된다.

<표 5-11> 2차 시험조사 가구원수 종합추정

추정 방법	가구원수	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	1인	47,552	17,163	64,715	7.08	8.34	5.66
	2인	47,211	20,567	67,777	3.53	3.76	2.71
	3인	31,279	11,322	42,601	5.07	5.83	4.03
	4인	33,311	8,773	42,084	5.24	10.28	4.67
	5인이상	12,544	4,229	16,773	7.68	11.36	6.42
GREG 비통합	1인	43,792	15,619	59,411	0.00	0.00	0.00
	2인	41,844	17,142	58,986	0.00	0.00	0.00
	3인	32,070	11,759	43,829	3.53	5.02	2.91
	4인	34,043	9,041	43,084	3.16	6.50	2.82
	5인이상	12,868	4,403	17,272	5.96	9.70	5.08
GREG 통합	1인	43,792	15,619	59,411	0.00	0.00	0.00
	2인	41,844	17,142	58,986	0.00	0.00	0.00
	3인	30,171	10,183	40,354	2.73	4.17	2.32
	4인	34,663	9,866	44,529	2.88	5.98	2.58
	5인이상	14,147	5,155	19,302	4.90	8.16	4.18

<표 5-12> 2차 시험조사 주택유형 종합추정

추정 방법	주택유형	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	단독주택	75,990	37,096	113,086	6.9	6.7	5.1
	아파트	47,418	11,638	59,056	7.0	14.5	6.3
	연립 다세대	32,466	10,827	43,293	13.3	16.9	10.8
	기타	15,003	1,979	16,982	21.9	51.7	20.3
GREG 비통합	단독주택	73,891	37,658	111,549	0.0	0.0	0.0
	아파트	45,484	9,498	54,982	6.6	11.3	5.8
	연립 다세대	30,503	8,840	39,343	10.7	12.5	8.7
	기타	13,786	1,547	15,334	19.6	48.0	18.4
GREG 통합	단독주택	73,891	37,658	111,549	0.0	0.0	0.0
	아파트	45,214	9,696	54,910	6.6	11.2	5.8
	연립 다세대	30,633	8,715	39,348	10.7	13.0	8.8
	기타	14,093	1,591	15,684	19.4	51.1	18.4

<표 5-13A> 2차 시험조사 남성 · 연령 총합추정

추정 방법	성연령그룹	총합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	0-14세	35,461	10,970	46,432	5.7	7.6	4.7
	15-19세	15,741	4,792	20,533	7.5	12.4	6.5
	20-29세	20,952	6,258	27,210	6.6	10.7	5.7
	30-39세	32,196	9,266	41,463	5.9	9.8	5.1
	40-49세	38,304	12,188	50,491	4.4	6.8	3.7
	50-59세	34,903	12,609	47,512	5.1	7.6	4.3
	60세이상	36,751	20,289	57,040	5.1	5.4	3.8
GREG 비통합	0-14세	37,272	10,717	47,989	2.5	3.8	2.1
	15-19세	16,545	4,681	21,226	5.5	8.6	4.7
	20-29세	22,557	6,920	29,476	5.5	8.1	4.6
	30-39세	34,662	10,246	44,909	3.6	5.4	3.0
	40-49세	38,329	13,046	51,375	3.0	4.5	2.5
	50-59세	34,927	13,497	48,424	3.3	4.3	2.7
	60세이상	33,324	16,412	49,736	2.2	2.4	1.7
GREG 통합	0-14세	37,865	10,601	48,467	2.3	5.1	2.2
	15-19세	15,952	4,797	20,748	4.9	8.6	4.2
	20-29세	22,852	7,474	30,326	4.7	5.8	3.9
	30-39세	34,367	9,692	44,059	3.3	4.7	2.8
	40-49세	38,041	12,397	50,438	3.0	4.5	2.5
	50-59세	35,215	14,146	49,361	3.5	5.1	3.0
	60세이상	31,374	15,590	46,964	4.3	5.2	3.4

<표 5-13B> 2차 시험조사 여성·연령 종합추정

추정 방법	성연령그룹	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	0-14세	35,549	8,768	44,317	6.7	11.4	5.8
	15-19세	13,222	3,886	17,109	8.0	16.6	7.3
	20-29세	23,436	4,896	28,332	6.3	10.8	5.5
	30-39세	30,461	8,397	38,858	5.7	11.1	5.1
	40-49세	37,511	10,661	48,172	5.2	9.1	4.5
	50-59세	35,877	12,242	48,119	4.6	6.3	3.8
	60세이상	43,166	24,076	67,241	4.9	6.0	3.8
GREG 비통합	0-14세	36,286	10,064	46,351	2.7	5.6	2.4
	15-19세	13,497	4,461	17,957	7.2	12.7	6.3
	20-29세	24,962	5,753	30,715	4.8	9.6	4.4
	30-39세	32,446	9,866	42,312	3.7	5.6	3.2
	40-49세	36,258	10,949	47,207	3.3	5.0	2.8
	50-59세	34,679	12,572	47,251	3.4	4.4	2.8
	60세이상	39,140	19,476	58,615	1.9	2.0	1.5
GREG 통합	0-14세	36,889	10,119	47,008	2.6	5.8	2.4
	15-19세	12,894	4,406	17,300	7.2	10.9	6.1
	20-29세	24,990	6,199	31,189	4.6	7.4	4.0
	30-39세	32,418	9,420	41,838	3.3	5.4	2.9
	40-49세	36,284	10,769	47,053	3.1	5.2	2.7
	50-59세	34,653	12,752	47,405	3.5	5.2	2.9
	60세이상	36,965	18,926	55,891	4.8	4.9	3.7

<표 5-14> 2차 시험조사 주거전용·영업겸용여부 종합추정

추정방법	주거전용 영업겸용 여부	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	있음	6,536	2,946	9,481	19.9	19.0	14.9
	없음	164,348	59,034	223,381	2.5	3.2	2.0
GREG 비통합	있음	6,251	2,959	9,210	19.4	19.5	14.6
	없음	157,423	54,943	212,366	0.8	1.1	0.6
GREG 통합	있음	6,152	3,041	9,193	19.6	18.7	14.6
	없음	157,653	54,879	212,532	0.8	1.0	0.6

<표 5-15> 2차 시험조사 점유형태 종합추정

추정 방법	점유형태	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	자가	90,242	35,393	125,636	4.6	5.5	3.7
	전세(월세없음)	7,184	1,654	8,838	8.9	29.6	9.1
	보증있는월세	40,901	10,481	51,382	11.8	21.1	10.4
	보증없는월세	6,194	2,132	8,325	20.8	29.5	17.2
	사글세	11,445	5,204	16,649	12.8	15.6	10.0
	무상(관사, 사택)	14,824	7,116	21,940	9.3	8.8	6.9
GREG 비통 합	자가	87,516	33,653	121,169	3.5	4.4	2.9
	전세(월세없음)	6,935	1,410	8,345	8.6	30.1	8.8
	보증있는월세	38,126	9,063	47,190	9.3	16.9	8.5
	보증없는월세	5,800	1,965	7,765	20.7	27.1	16.9
	사글세	11,007	5,147	16,154	12.6	14.8	9.9
	무상(관사, 사택)	14,198	6,664	20,862	9.5	9.3	7.1
GREG 통합	자가	85,877	32,539	118,417	3.4	4.5	2.8
	전세(월세없음)	6,996	1,388	8,384	8.7	30.8	8.9
	보증있는월세	39,409	9,770	49,179	8.7	16.0	8.0
	보증없는월세	5,783	2,073	7,857	20.5	25.8	16.6
	사글세	11,126	5,297	16,423	12.6	14.4	9.8
	무상(관사, 사택)	14,523	6,851	21,374	9.2	9.3	6.9

<표 5-16> 2차 시험조사 교육정도 종합추정

추정 방법	교육정도	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포 시	제주도
HT	받지않았음	48,143	20,386	68,529	5.5	9.8	4.9
	초등학교	54,174	22,514	76,688	5.3	5.8	4.1
	중학교	40,789	16,999	57,787	4.7	5.8	3.8
	고등학교	114,786	44,365	159,151	4.1	5.5	3.4
	대학교(2,3년제)	59,166	16,970	76,136	4.9	6.7	4.1
	대학교(4년제이상)	100,881	24,616	125,497	4.4	7.9	3.9
	대학원석사과정	12,853	3,063	15,916	10.0	14.1	8.5
	대학원박사과정	2,097	385	2,482	15.6	36.4	14.3
GREG 비통 합	받지않았음	48,107	19,094	67,201	4.3	7.6	3.8
	초등학교	53,617	20,548	74,165	3.4	4.7	2.8
	중학교	39,913	16,103	56,015	4.0	4.2	3.1
	고등학교	113,582	44,667	158,249	3.0	3.1	2.4
	대학교(2,3년제)	61,027	18,519	79,547	3.5	3.8	2.8
	대학교(4년제이상)	103,038	26,228	129,265	3.3	6.2	2.9
	대학원석사과정	12,857	3,113	15,970	9.5	14.3	8.2
	대학원박사과정	2,082	388	2,471	15.5	37.5	14.3
GREG 통합	받지않았음	49,390	19,867	69,257	3.7	5.9	3.2
	초등학교	52,473	19,628	72,100	3.3	5.4	2.8
	중학교	38,337	16,249	54,586	4.0	4.1	3.1
	고등학교	112,800	44,790	157,589	2.9	2.3	2.2
	대학교(2,3년제)	60,971	18,253	79,224	3.4	3.6	2.8
	대학교(4년제이상)	101,514	25,187	126,701	3.2	6.5	2.9
	대학원석사과정	12,639	2,934	15,573	9.4	13.4	8.1
	대학원박사과정	1,988	380	2,368	15.2	39.1	14.3

<표 5-17> 2차 시험조사 수확정도 종합추정

추정 방법	수확 정도	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	졸업	286,119	102,453	388,572	2.8	3.0	2.2
	재학	80,589	20,477	101,065	4.7	9.9	4.3
	수료	1,366	716	2,083	29.8	39.0	23.7
	휴학	4,582	1,250	5,831	16.0	19.7	13.3
	중퇴	12,048	4,015	16,063	9.2	11.6	7.5
GREG 비통합	졸업	283,901	101,833	385,735	0.8	1.4	0.7
	재학	84,009	21,799	105,808	2.8	6.4	2.6
	수료	1,386	706	2,092	29.6	38.3	23.5
	휴학	4,820	1,360	6,181	15.4	20.1	12.8
	중퇴	11,957	3,867	15,824	8.8	11.5	7.2
GREG 통합	졸업	280,073	99,872	379,945	0.8	1.3	0.6
	재학	82,675	21,748	104,424	2.3	4.2	2.0
	수료	1,340	741	2,081	29.4	37.1	23.2
	휴학	4,794	1,316	6,110	15.0	18.1	12.5
	중퇴	11,792	3,743	15,535	8.8	11.8	7.3

<표 5-18> 2차 시험조사 종교유무 종합추정

추정 방법	종교여부	종합추정			상대표준오차(%)		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	있음	186,223	64,913	251,136	4.6	5.8	3.7
	없음	246,845	84,385	331,230	3.1	5.5	2.7
GREG 비통합	있음	184,560	63,024	247,584	2.9	5.0	2.5
	없음	249,849	85,636	335,485	2.1	3.7	1.8
GREG 통합	있음	182,550	61,411	243,962	3.0	4.9	2.6
	없음	247,744	85,876	333,621	2.2	3.5	1.9

<표5-19> 2차 시험조사 출생지 총합추정

추정 방법	출생지	총합추정			상대표준오차		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	현재 살고 있는 집	30,881	13,485	44,366	8.8	10.8	7.0
	같은시군구내 다른집	250,384	82,315	332,699	3.8	4.7	3.1
	다른시군구	142,748	50,714	193,462	4.0	5.9	3.3
	북한또는외국	8,776	2,784	11,560	12.1	12.9	9.7
GREG 비통합	현재 살고 있는 집	31,870	13,823	45,693	7.9	11.2	6.5
	같은시군구내 다른집	250,881	81,064	331,945	2.0	2.7	1.7
	다른시군구	142,496	50,958	193,454	3.2	4.0	2.6
	북한또는외국	8,879	2,814	11,693	12.2	13.6	9.8
GREG 통합	현재 살고 있는 집	33,112	14,722	47,834	7.3	10.6	6.0
	같은시군구내 다른집	249,228	81,714	330,942	1.9	2.8	1.6
	다른시군구	138,801	48,158	186,959	3.1	4.4	2.5
	북한또는외국	8,875	2,693	11,568	12.4	13.8	10.1

<표5-20> 2차 시험조사 1년전 거주지 종합추정

추정 방법	출생지	종합추정			상대표준오차		
		제주시	서귀포 시	제주도	제주시	서귀포 시	제주도
HT	태어나지않았음	5,170	2,095	7,265	16.3	14.0	12.3
	현재살고있는집	374,333	129,265	503,597	3.4	4.7	2.8
	같은시군구내다른 집	32,943	7,167	40,110	10.1	15.4	8.7
	다른시군구	18,308	10,542	28,850	12.1	14.7	9.4
	북한또는외국	2,120	220	2,340	24.4	49.1	22.6
GREG 비통합	태어나지않았음	5,329	2,242	7,571	16.1	13.3	12.0
	현재살고있는집	374,140	127,404	501,544	1.3	1.8	1.0
	같은시군구내다른 집	33,757	7,559	41,316	10.3	16.3	8.9
	다른시군구	18,783	11,204	29,986	11.6	14.9	9.1
	북한또는외국	2,202	240	2,442	24.1	49.5	22.3
GREG 통합	태어나지않았음	5,673	2,491	8,164	16.3	12.2	11.9
	현재살고있는집	370,828	126,596	497,424	1.2	1.6	1.0
	같은시군구내다른 집	33,248	7,280	40,528	10.1	16.5	8.9
	다른시군구	18,084	10,614	28,698	11.0	13.1	8.4
	북한또는외국	2,272	295	2,567	24.2	49.7	22.3

<표5-21> 2차 시험조사 통근·통학여부 종합추정

추정 방법	통근· 통학여부	종합추정			상대표준오차		
		제주시	서귀포시	제주도	제주시	서귀포시	제주도
HT	통근함	215,088	81,336	296,424	3.1	4.1	2.5
	통학함	55,301	13,849	69,149	5.2	12.4	4.9
	안함	102,841	37,376	140,217	3.8	5.9	3.2
GREG 비통합	통근함	215,142	81,272	296,415	1.4	2.2	1.2
	통학함	57,872	14,748	72,620	4.1	9.1	3.8
	안함	99,432	35,022	134,454	2.8	4.4	2.4
GREG 통합	통근함	213,383	80,281	293,664	1.3	2.5	1.2
	통학함	56,724	14,842	71,566	3.3	7.0	3.0
	안함	96,651	34,266	130,917	3.0	5.4	2.7

2) 소지역 추정

본 소절에서는 전 소절의 2.5% (혹은 분기) 표본자료를 이용할 때 시군구 수준의 (소)지역 총합추정에 있어서 HT-추정치에 비해 GREG 추정치와 통합 GREG 추정치의 분산이 얼마나 감소할 수 있었는지 살펴보았다.

<표 5-22>는 가구원수별 총 가구수의 GREG 추정치와 통합 GREG 추정치의 분산과 HT-추정치의 분산 대비 얼마나 작거나 큰지를 나타내고 있다. 1인 가구와 2인 가구는 칼리브레이션 조정범주와 같았으므로 두 GREG 총합추정치의 분산이 0이 되어 분산비는 0이 된다. 반면 3인, 4인, 5인 이상의 세부 범주에 대해서는 GREG 총합추정치에 비해 통합 GREG 추정치의 분산비가 전반적으로 작았는데 작게는 0.268에서 크기는 0.767로 나타났다.

<표 5-23>-<표 5-32>은 각각 주택유형, 주거전용·영업겸용여부, 점유형태, 성·연령, 교육정도, 수확정도, 종교유무, 출생지, 1년전 거주지, 통근·통학여부에 대한 HT-추정치 분산 대비 GREG 추정치와 통합 GREG 추정치의 분산비를 각각 나타내고 있다. 대부분의 변수 및 범주에서 분산비가 1보다 작은 값을 보이는데, 그 중 출생지와 1년전 거주지 특성에 대해서는 다소 많은 범주에서 분산비가 1보다 큰 값을 보여준다. 이는 GREG와 통합 GREG 총합추정 모두 HT-추정에 비해 작은 분산을 주고 있음을 나타낸다. 따라서 GREG에 의한 칼리브레이션 조정은 제주시와 서귀포 (소)지역 추정에서 매우 효율적인 방법임을 보여준다.

또한 앞 소절의 <표 5-11>-<표 5-21>에는 각 변수의 범주별 총합추정치의 상대표준오차를 정리하고 있다. 상당히 많은 범주에서 2015 인구주택총조사에서 고려한 각 시군구별 통계의 목표 CV인 구·시 14% (군 17% 이하)를 만족하지 못하는 경우가 상당수 존재한다. 따라서 표본수가 작거나 목표 CV를 상회하는 몇 개의 범주에 대해서는 다른 범주와 통합하여 통계작성을 고려해야 할 것으로 판단된다.

<표 5-22> 2차 시험조사 가구원수 총합추정 분산비

가구원수	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
1인	0.000	0.000	0.000	0.000	0.000	0.000
2인	0.000	0.000	0.000	0.000	0.000	0.000
3인	0.510	0.799	0.549	0.268	0.415	0.296
4인	0.378	0.424	0.382	0.328	0.429	0.342
5인이상	0.632	0.790	0.663	0.518	0.767	0.560

<표 5-23> 2차 시험조사 주택유형 총합추정 분산비

가구원수	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
1인	0.000	0.000	0.000	0.000	0.000	0.000
2인	0.819	0.400	0.722	0.799	0.416	0.718
3인	0.569	0.368	0.536	0.580	0.382	0.548
5인이상	0.676	0.525	0.676	0.691	0.630	0.706

<표 5-24> 2차 시험조사 주거전용·영업겸용여부 총합추정 분산비

주거전용, 영업겸용여부	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
있음	0.877	1.065	0.902	0.862	1.038	0.896
없음	0.091	0.092	0.091	0.088	0.090	0.090

<표 5-25> 2차 시험조사 점유형태 총합추정 분산비

점유형태	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
자가	0.543	0.563	0.571	0.497	0.565	0.531
전세(월세없음)	0.869	0.750	0.826	0.898	0.761	0.852
보증있는월세	0.530	0.479	0.568	0.502	0.501	0.548
보증없는월세	0.871	0.714	0.841	0.849	0.722	0.829
사글세	0.903	0.884	0.912	0.920	0.889	0.931
무상(관사, 사택)	0.953	0.971	0.960	0.936	1.027	0.963

<표 5-26> 2차 시험조사 성·연령 총합추정 분산비

성·연령		분산비(GREG/HT)			분산비(통합GREG/HT)		
성별	연령	제주시	서귀포시	제주도	제주시	서귀포시	제주도
남성	0-14세	0.206	0.234	0.210	0.191	0.412	0.244
	15-19세	0.594	0.465	0.568	0.429	0.478	0.436
	20-29세	0.783	0.693	0.766	0.606	0.426	0.581
	30-39세	0.422	0.378	0.414	0.368	0.253	0.355
	40-49세	0.474	0.499	0.480	0.459	0.449	0.469
	50-59세	0.414	0.374	0.406	0.485	0.574	0.522
	60세이상	0.152	0.126	0.146	0.532	0.545	0.550
여성	0-14세	0.169	0.318	0.193	0.163	0.345	0.193
	15-19세	0.844	0.769	0.832	0.767	0.556	0.723
	20-29세	0.673	1.079	0.729	0.597	0.753	0.634
	30-39세	0.477	0.349	0.454	0.378	0.298	0.366
	40-49세	0.374	0.317	0.361	0.339	0.327	0.342
	50-59세	0.513	0.504	0.510	0.544	0.744	0.581
	60세이상	0.129	0.073	0.112	0.720	0.415	0.657

<표 5-27> 2차 시험조사 교육정도 총합추정 분산비

교육정도	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
받지않았음	0.61246	0.53809	0.5925	0.46888	0.3462	0.44415
초등학교	0.40364	0.55294	0.4399	0.36312	0.65605	0.4104
중학교	0.66774	0.45691	0.6326	0.64263	0.45195	0.61078
고등학교	0.52454	0.31412	0.4912	0.48092	0.17155	0.43325
대학교(2,3년제)	0.54212	0.39151	0.5216	0.53121	0.32594	0.51793
대학교(4년제이상)	0.57582	0.70345	0.6096	0.53436	0.71852	0.58621
대학원석사과정	0.90405	1.05459	0.9348	0.86395	0.8218	0.86507
대학원박사과정	0.97386	1.08253	0.99096	0.85926	1.12551	0.9084

<표 5-28> 2차 시험조사 수학정도 총합추정 분산비

수학정도	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
졸업	0.074	0.224	0.092	0.068	0.181	0.081
재학	0.397	0.472	0.416	0.246	0.197	0.239
수료	1.020	0.936	0.995	0.935	0.963	0.958
휴학	1.024	1.226	1.042	0.960	0.935	0.970
중퇴	0.904	0.912	0.902	0.887	0.888	0.882

<표 5-29> 2차 시험조사 종교유무 총합추정 분산비

종교여부	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
있음	0.387	0.696	0.440	0.398	0.645	0.441
없음	0.479	0.451	0.476	0.491	0.417	0.476

<표 5-30> 2차 시험조사 출생지 총합추정 분산비

출생지	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
현재살고있는집	0.862	1.132	0.913	0.776	1.156	0.853
같은시군구내다른집	0.288	0.326	0.294	0.241	0.354	0.258
다른시군구	0.667	0.449	0.621	0.562	0.490	0.548
북한또는외국	1.043	1.135	1.051	1.079	1.056	1.084

<표 5-31> 2차 시험조사 1년전 거주지 총합추정 분산비

출생지	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
태어나지않았음	1.031	1.027	1.031	1.197	1.073	1.180
현재살고있는집	0.137	0.137	0.137	0.116	0.110	0.115
같은시군구내다른집	1.079	1.238	1.096	1.015	1.178	1.045
다른시군구	0.963	1.157	1.021	0.799	0.806	0.793
북한또는외국	1.048	1.206	1.060	1.128	1.841	1.173

<표 5-32> 2차 시험조사 통근·통학여부 종합추정 분산비

통근·통학여부	분산비(GREG/HT)			분산비(통합GREG/HT)		
	제주시	서귀포시	제주도	제주시	서귀포시	제주도
통근함	0.194	0.300	0.214	0.185	0.356	0.220
통학함	0.672	0.618	0.663	0.413	0.369	0.409
안함	0.518	0.490	0.507	0.579	0.701	0.610

5. 순환표본조사와 등록센서스 연계

통계청은 기존연구(통계청, 2015)에서는 ACS 조사와 같이 월 단위 순환조사를 기준으로 일정 기간 동안 여러 시점 기준의 자료를 누적하여 추정량을 산출하는 “기간추정(period estimate)”에 대하여 고려하였다. 더불어 1-2차 시험조사의 표본설계도 매월 수행하는 순환조사로 진행되었다.

하지만 통계청에서는 현재 매년 수행하는 순환조사를 검토하고 있으며 5장 3절에서 논의한 바와 같이 통계생산을 위한 최소의 지역단위로 시군구를 고려한다면 대부분의 시군구들에 대해서는 매년 조사자료를 이용하여 적절한 정도수준을 갖춘 통계생산이 가능하므로 굳이 기간추정을 고려할 이유는 없어 보인다. 일부 규모가 작은 시군구들에 대해서 기간추정을 고려할 수는 있으나 이보다는 인근 시군구와 통합하여 표본은 물론 모집단의 크기를 확보하는 것이 바람직할 것이다.

만일 일정기간 동안에 걸친 누적된 시점별 조사자료를 이용한 추정량을 고려하게 된다면 현재 고려 중인 연단위 순환조사 하에서는 3년 혹은 5년과 같은 다년추정을 고려하여야 할 것이다. 기간추정에 대한 모수 및 추정에 대한 논의는 기존연구(통계청, 2015)에도 포함되어 있지만, 본 절에서도 간단히 정리하고자 한다.

특정시점 t 의 모수가 θ_t 이면 총 T 시점 동안의 기간 $\mathcal{J} = \{t: t=1, \dots, T\}$ 에 대한 기간모수(parameter for period)는 다음과 같이 시점별 상대적 모집단 크기 γ_t 에 의한 시점모수(parameter for point-in-time)의 가중합으로 정의될 수 있다.

$$\theta_{\mathcal{J}} = \sum_{t \in \mathcal{J}} \gamma_t \theta_t \quad (5-16)$$

여기서 $\gamma_t = N_t/N_{\mathcal{J}}$ 는 시점별 상대적 모집단 크기이고 $N_{\mathcal{J}} = \sum_{t \in \mathcal{J}} N_t$ 는 총 T 기간의 모집단 크기의 합계이다. 만약, 기간 내 모집단 크기의 변동이 미미하다면 식 (5-16)은 다음과 같이 근사적 표현을 가진다.

$$\theta_{\mathcal{J}} \simeq \sum_{t \in \mathcal{J}} \gamma_t \theta_t := \theta_{\mathcal{J}}^* \quad (5-17)$$

여기서 $\gamma_{\mathcal{J}} = 1/T$ 는 시점모수 가중치이다. 앞서 정의한 기간모수는 조사 및 기준 시점이 상이한 모수를 통합하므로 직관적 해석이 용이하지 않을 수 있다. 다만 식

(5-16) 혹은 식 (5-17)의 모수 형태는 기간 동안 조사별 기준 시점 모수들의 가중 혹은 단순평균으로 각각 이루어진다.

시점모수의 (근사적) 불편추정량은 개별 시점의 시점모수 추정량 $\hat{\theta}_t$ 을 기간모수식에 대입한 추정량으로 기간추정량을 정의하면 다음과 같이 표기할 수 있다.

$$\hat{\theta}_J = \sum_{t \in J} \gamma_t \hat{\theta}_t \quad (5-18)$$

$$\hat{\theta}_J^* = \sum_{t \in J} \gamma_J \hat{\theta}_t \quad (5-19)$$

시점별 추정량은 (근사적으로) 독립이라는 가정 하에서는 시점별 분산추정량 $\hat{V}(\hat{\theta}_t)$ 의 가중합으로 표현된다. 예로, 식 (5-18)의 기간모수에 대한 분산추정량은 $\hat{V}(\hat{\theta}_J) = \sum_{t \in J} \gamma_t^2 \hat{V}(\hat{\theta}_t)$ 이 된다.

기간추정에 있어서도 보조정보를 이용한 일반화회귀추정으로 기간추정이 가능한데 이를 위해서는 해당 기간 내 시점별 보조변수 총합 또한 관련된 모수형태로 취합되어야 할 것이다. 예로, 식 (5-16)와 식 (5-17)의 기간모수를 사용한다면 이들에 대한 보조변수 총합은 각각 다음과 같이 정의되고 이에 따라 5장 2절에 논의한 바와 같이 회귀가중치를 산출하면 된다.

$$\bar{X}_J = \sum_J \gamma_t X_t \quad (5-20)$$

$$\bar{X}_J^* = \sum_J \gamma_J X_t \quad (5-21)$$

참고문헌

- 박시내 · 이주희. 2016. 순환센서스 도입을 위한 미국, 프랑스 순환센서스 사례와 시사점. 통계개발원 2016년 하반기 연구보고서.
- 박시내 · 정남수 · 전지혜 · 정진희. 2015. 순환센서스 시험조사 결과에 대한 인구·사회적 함의 분석. 통계개발원 2015년 상반기 연구보고서 제 I 권.
- 박은영 · 김의영. 2014. 순환센서스 적용을 위한 표본 틀 구축 및 갱신방안 연구. 통계개발원 2014년 상반기 연구보고서 제 II 권.
- 최재혁 · 김경미. 2014. “순환표본조사 표본설계 및 가중치 작성방법 연구”, 통계개발원 2014년 하반기 연구보고서 제 VI 권.
- 통계청. 2014. 순환센서스 추정 및 분석방안에 대한 연구, 한국인구학회 연구용역 보고서.
- 통계청. 2015. 2015년 순환센서스 추정 및 분석방안에 대한 연구, 한국인구학회 연구용역 보고서.
- Fay, R. (2006). Using administrative records with model-assisted estimation for the American Community Survey. In *Proceedings of the 2006 Joint Statistical Meetings on CD-ROM 2995-3001*. American Statistical Association, Alexandria, VA.
- Fay, R. (2005). Model-assisted estimation for the American Community Survey. In *Proceedings of the 2005 Joint Statistical Meetings on CD-ROM 3016-3023*. American Statistical Association, Alexandria, VA.
- Kish, L. (1965). *Survey Sampling*. Wiley, New York, NY.
- Lemaître, G., and Dufour, J. (1987). “An integrated method for weighting persons and families,” *Survey Methodology*, 13: 199-207.

- Nieuwenbroek, J. (1993). *“An integrated method for weighting characteristics of persons and households using the linear regression estimator.”* Netherlands Central Bureau of Statistics.
- Sarndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York, NY.
- Steel, D. G., and Clark, R. G. 2007. “Person-level household-level regression estimation in household surveys,” *Survey Methodology*, 33, 51-60.