

인구주택총조사 표본조사방법 개선을 위한 사례 연구

양경진 · 김황대

목 차

제1절 서 론	1
1. 연구 배경	1
2. 연구 목표	2
3. 연구 내용	2
제2절 센서스 해외 사례 연구	3
1. 센서스 실시 사례의 분류	3
2. 미국의 사례(미국지역사회조사)	10
3. 프랑스의 사례(순환센서스)	22
4. 네덜란드 사례(가상 센서스)	31
5. 해외사례 비교	37
제3절 우리나라의 센서스 표본조사 개선 방향	39
1. 조사 체계	41
2. 조사 방법	46
3. 표본추출 틀(Sampling frame)	47
4. 추정 및 공표	55
5. 조사 비용	57
6. 또 하나의 대안: ‘가상 센서스’	58
제4절 결론	62
<부록> 반복가중기법(Repeated weighting)	67

표 목 차

<표 1> 주요 공표 단위 및 공표 주기	16
<표 2> 최소 공표단위별 연간 표집률	17
<표 3> 2001년 ACS의 응답률	19
<표 4> ACS와 2000년 센서스의 대체율 비교	20
<표 5> 90% 신뢰구간 사용 예	21
<표 6> 전통적인 센서스와 ACS의 차이점	22
<표 7> 주민등록자료와 인구주택총조사 결과 비교	50
<표 8> 조사구 이용방식과 세대명부 이용방식 간의 고유번호 체계 비교	53
<표 9> ‘인구주택총조사’ 조사구와 ‘반’의 규모 비교	55
<표10> 순환표본조사 추정예산 규모	58
<표11> 가상센서스에 사용된 변수	71

그림 목차

[그림 1] 조사체계	42
[그림 2] 표본추출틀로서의 주민등록자료와 인구주택총조사 모집단과의 관계 ..	48
[그림 3] 주민등록 세대명부 양식	50
[그림 4] Data blocks for persons of 15-74 years of age	73

인구주택총조사 표본조사방법 개선을 위한 사례 연구 -미국, 프랑스, 네덜란드 사례를 중심으로-

제1절 서론

1. 연구 배경

통계조사 환경을 비롯한 사회 전반의 상황 변화는 대중에게서 정보를 얻고 또 대중에게 정보를 서비스해야 하는 통계작성 시스템의 변화에 필연적으로 영향을 미치게 되는데, 이러한 통계작성 시스템 변화의 조류에는 인구센서스도 예외일 수가 없게 되었다. 2000년 라운드 센서스에서는 세계 190개국 이상의 나라에서 인구센서스를 실시하였고 이들 센서스의 대부분은 전통적인 방식을 따랐으나, 2010년 라운드 센서스에서는 전통적 방식에 의한 자료수집, 처리, 공표의 기술의 대안을 만들어 시험, 이행하고 있는 나라들이 늘어나고 있다(UNSD, 2007).

이렇게 각 나라별로 환경변화에 적응하기 위하여 각자가 처해진 상황에 맞춰 센서스 시스템의 개선에 힘을 기울이고 있음은 경제적인 급속한 발전뿐만 아니라 각종 제도, 국민의식, 가치관 등 사회전반의 시스템들의 급속한 변화를 통해 역동적으로 변화해 가고 있는 우리나라에게 시사하는 바가 크다 하겠다. 실제로 5년을 주기로 하는 ‘인구주택 총조사’를 준비하고 조사해서 결과를 공표하는 단계 단계마다 이전의 조사에서보다 확연히 달라진 조사환경과 직면하게 되었는데, 이에 반해 통

계 이용자들의 수요는 더욱 전문화, 세분화되어 가고 있다는 것에 우리가 센서스 조사방법 개선이라는 주제로 고민해야 하는 근본적인 이유가 있다.

가장 최근의 우리나라 센서스인 ‘2005 인구주택총조사’는 전통적인 방식으로 실시되었으며 총 44개(전수 21개, 표본 23개)에 걸친 항목에 대해 전국의 상주 내·외국인을 대상으로 하였는데 조사결과는 자료처리의 단축을 통해 과거의 조사보다 신속하게 발표하여 시의성 측면에서 성공적인 조사였다고 평가받고 있다. 이렇게 통계작성에 있어 시의성이란 정확성과 함께 중요한 척도가 되고 있으나, 전통적 방식의 센서스에서는 조사 다음 연도에 모든 결과를 발표하는 등의 시의성 제고를 위한 노력을 아끼지 않는다고 해도 5년 주기조사라는 근본적인 한계는 수요자를 충족시킬 수 없었다.

이에 따라 순환표본에 의한 조사 시스템은 1년 단위로 조사결과 공표가 가능하고 조사비용을 연도별로 분산시킬 수 있다는 점에서 센서스 이용자들의 욕구 충족과 함께 조사비용 집행의 효율성까지도 확보할 수 있는 대안으로서 관심의 대상이 되고 있다.

2. 연구 목표

본 연구는 우리나라의 인구센서스 표본조사 개선에 관해 미국, 프랑스, 네덜란드 등 센서스 조사방법 개선의 선형국 실시사례를 벤치마킹하는 것으로 시작하여 ‘순환센서스(Rolling Census)’방식을 포함한 ‘순환표본조사(Rotation sampling survey)’를 중심으로 조사환경의 변화, 제도도입에 따른 문제사항 및 해결방안 등에 대한 검토를 통해 우리나라 특성에 맞는 실용적인 인구센서스 표본조사 개선 방향 제시를 목표로 한다.

3. 연구 내용

본 연구에서는 기본적으로 인구센서스 표본조사 방식의 개선 검토방향을 미국의 경우와 같이 기존의 전통적 센서스와 비교하여 보다 짧은 주기를 가지고 실시되는 순환표본 방식의 센서스 개념으로 접근해 가고

자 하였으며, 외국의 실시보고서를 중심으로 하는 선행 연구결과 검토를 통하여 통계수요에 부응하는 인구센서스 표본조사 방식의 개선방향을 제시하기 위해 노력하였다.

연구의 내용은 첫째, 선협국의 사례검토로서 미국의 ‘미국지역사회조사(ACS)’ 방식, 프랑스의 ‘순환센서스(Rolling Census)’ 방식, 네덜란드의 ‘가상센서스(Virtual Census)’ 방식을 중심으로 각 국의 실시 경험의 수록되어 있는 실무보고서를 중심으로 실시배경과 시행과정 및 결과 공표과정 등에 대해 살펴보았고, 두 번째로 검토 사례국간의 비교를 통해 우리나라의 상황과 부합하는 시스템의 개선과 관련된 시사점을 도출하고자 하였으며, 마지막으로 첫 번째와 두 번째 검토를 배경으로 한 우리나라의 센서스 표본조사 개선방향을 모색하는 부분으로 구성하였다.

제2절 센서스¹⁾ 해외 사례 연구

1. 센서스 실시 사례의 분류

센서스를 실시하는 주된 이유는 상세한 정보, 특히 소지역에 대한 상세한 정보가 필요하기 때문이다. 그러나 기존 센서스의 단점은 정보의 노후화와, 비용이 많이 들어 센서스 주기를 줄일 수 없다는 점이다. 또한 개인비밀보호의식의 증가 및 단독가구의 급증 등으로 인해 조사환경이 악화되어 무응답 문제가 증가하고 있다. 이러한 문제점들을 해결하기 위해서 전통적인 센서스를 포기하거나 변형하여 실시하는 나라들이 점점 늘어나고 있다. 무응답 문제를 해결하기 위해서 행정자료를 이용하고, 센서스의 막대한 비용을 여러 해에 걸쳐 분산하기 위해서 순환센서스를 실시하는 등 여러 유형의 센서스가 등장하고 있다.

UN은 이러한 여러 가지 유형을 정리하여 세계 각국의 센서스를 ① 전통적 센서스, ② 행정등록 자료에 기반을 둔 접근법, ③ 전통적 센서

1) 센서스는 라틴어로 세금징수를 의미한다. 센서스에 관한 기록은 B.C. 3000년 이집트에 까지 거슬러 올라간다.

스와 행정등록 센서스의 혼합형, ④ 행정등록 센서스와 표본조사의 혼합형, ⑤ 순환 센서스, ⑥ 전통적인 센서스와 특성치의 연간 갱신, ⑦ 모형 기반 센서스로 분류하였다(UN, 2007). 미국, 프랑스, 네덜란드 등 소수 관심 국에 대한 사례를 소개하기에 앞서 UN의 분류에 따른 해외 각국의 개괄적인 사례설명을 통해 센서스 실시방식에 대한 해외의 전반적인 흐름을 조망하고자 한다.

가. 전통적 센서스

2000 라운드²⁾ 센서스에서는 190개국 이상의 나라들이 인구센서스를 실시하였고 대부분의 나라들이 전통적인 방식을 따랐다. 인구센서스는 특정 시점에서 일련의 주제에 관해 인구 및 가구에 관한 정보를 수집하는 복잡한 작업이다. 수집된 자료는 편집, 평가, 분석 및 공표과정을 통해 국가 및 지역에 관련된 인구, 경제, 사회통계를 작성하는데 이용된다. 국민들이 센서스 질문지를 직접 작성할 경우도 있고 조사원을 고용하여 정보를 얻을 수도 있다. 조사원을 고용하여 센서스를 실시할 경우, 각 조사원은 비교적 짧은 기간인 특정의 시점에서 배정된 조사구내의 모든 가구 및 개인들을 조사함으로써 보편성과 동시성이라는 센서스의 요건을 충족하게 된다. 전수조사표와 표본조사표가 모두 사용된다. 전수조사표는 모집단 전체를 대상으로 하고 표본조사표는 표본으로 선정된 가구 및 인구를 대상으로 한다. 표본조사표에는 특정한 주제에 관한 상세한 질문들이 포함된다. 표본조사 추정치는 전체 포괄범위에 기초를 둔 것은 아니지만 센서스의 결과물로 간주된다.

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

전통적인 센서스는 특정 시점에서의 인구에 대한 스냅샷(snap shot)을 제공하고 상대적으로 소규모 지역에 대한 자료를 제공한다는 독보적인 장점이 있다. 전통적 센서스는 연방구조를 채택하고 있고 예산수립 및 배분에 필요한 모든 지역단위에 대해 다양한 사회경제 특성치를 한

2) 2000 라운드는 1995년 ~ 2004년을 말한다.

시점에서 동시에 생산할 것이 요구되고 있는 나라에 특히 적합하다. 선거구 경계선 확정작업 또한 동시성을 요구하고, 이 이유 때문에 전통적 센서스가 더 나을 것이다. 그러나 한편으로 전통적인 센서스는 센서스를 담당하는 부서가 행하는 업무 중 가장 많은 노력이 요구되며, 복잡하고, 비용이 많이 드는 자료수집활동으로 분류된다. 비용문제 외에 이 복잡한 작업에는 국민들이 참여해야한다는 인식과 동의가 요구된다. 복잡성과 비용문제 때문에 이러한 센서스는 보통 매 5년 또는 10년마다 한번씩 실시되고 따라서 센서스 자료는 여러 해가 지난 시의성이 떨어지는 자료가 된다.

전통적인 센서스를 실시하는 국가: 캐나다, 콜롬비아

나. 행정등록 자료에 기반을 둔 접근법

행정등록명부에 기초를 두고 센서스와 비슷한 결과를 생산한다는 개념이 2000 라운드 센서스에 등장했다. 이 방식은 1970년 대 이후로 논쟁이 되었고 시험되어 왔으며 몇몇 나라에서는 이 방식을 성공적으로 수행하여 1990 라운드 센서스 자료를 생산했다. 이 개념의 배경 철학은 현존하는 행정자료를 이용하는 것이다. 즉 가구, 주택 및 개인 등록부를 포함한 서로 다른 등록부를 이용하는 것이다. 다음 단계에서는 이러한 것들을 개인 수준에서 기업체, 세금, 교육, 고용 및 기타 관련 등록부와 연결시키는 것이다. 이론적으로는 개인의 이름을 이용하여 기록들을 연결하는 것이 가능하지만, 각 개인, 가구, 거처에 대한 고유번호의 존재가 가장 중요하다. 왜냐하면 서로 다른 등록부의 기록들을 효율적이고 신뢰할만한 연결을 가능하게 하기 때문이다.

이 접근법의 필수 조건은 국가가 고품질의 포괄범위가 좋은 인구등록부를 보유하고 있어야 하며 계속적으로 보완할 수 있는 시스템을 보유해야 된다는 점이다. 지역 등록부의 경우에 있어서는 계속적인 보완 및 다른 지역의 등록시스템과의 협조 관계가 잘 되어 있어야 한다. 품질 평가 또한 필요하다. 만약 이러한 조건들이 충족되지 않는다면 그 나라는 인구센서스를 기존 인구통계의 자료원으로 사용해야 한다.

등록 접근법의 주요 이점은 센서스 절차의 비용절감 및 자료 빈도의

확대이다. 그러나 행정등록부를 설립하고 실시하는 데에 센서스 자체보다 더 많은 비용이 든다. 등록의 필요성을 입증해야 하는 것은 보다 유용하고 효과적인 행정이지 통계만이 아니다. 행정자료를 사용하는 데에는 고려해야 할 단점이 있다. 한 가지 단점은 등록에 기초를 둔 설명은 이용할 수 있는 명부에 기초를 두고 형성할 수 있는 정보에 전적으로 의지해야 한다는 사실이다. 게다가 적지 않은 나라에서 통계를 만드는 등의 기타 목적으로 사용하는 것을 법적으로 제한하고 있다. 이 점이 설명할 수 있는 특성에 관해 제약을 가하고 국제 비교를 훼손하게 만들 수도 있다.

행정등록센서스를 실시하는 국가: 덴마크, 노르웨이, 스웨덴, 핀란드

다. 전통적 센서스와 행정등록 센서스의 혼합형

서로 다른 주제에 관한 등록부들이 동등하게 개발되어 있지 않다고 여겨지거나 요구되는 모든 센서스 결과를 생산할 만큼 신뢰할 수 있는 등록부를 보유하고 있지 않은 나라들은 등록부에서 이용할 수 있는 정보를 센서스의 기초자료로서 사용하는 방안을 고려할 수 있다. 2000 라운드 센서스에서 일부 국가는 인구등록부를 이용하여 거처의 주소, 거주자의 이름 및 기타 정보를 미리 기입한 조사표를 사용하였다. 이러한 정보는 면접 시에 확인받거나 수정되었다. 필요하지만 미리 기입되지 않은 정보는 면접 시에 수집되었다.

이러한 식으로 평균 면접 시간은 상당히 줄어들었고 비용의 감소로 나타났다. 응답자의 부담은 최소화되었으며 자료의 질이 개선되었다. 또한 포괄범위에 대한 평가가 인구등록과 현장조사 결과와의 비교를 통해 이루어졌다.

더욱이 등록부의 질에 대한 평가가 유사한 방식으로 이루어질 수 있다. 또한 일부 항목에 대해 현장 조사에서 수집된 정보에 기초를 두고 보완함으로써 개선될 수 있다. 이렇게 함으로써 양질의 등록부만 사용하여 센서스 중간연도에 대한 자료 생산의 빈도를 높이는 것이 가능할 것이다. 이러한 형태의 주기적인 센서스 자료 수집은 등록부에서 직접 얻을 수 있는 인구수의 수정을 가능하게 하고 이주 등과 관련된 등록되지 않는 사건이 일어남으로써 필요한 재고조사 활동으로 간주될

수 있다.

마지막으로, 이 방법은 전체 포괄범위에 기초를 두었기 때문에 전통적인 방식의 이점을 그대로 간직한다. 특정 시점에서의 전체 인구에 대한 스냅샷을 제공하고 상대적으로 소규모 행정구역에 대한 자료의 생산이 가능하다. 이러한 형태의 센서스를 시행한 경험이 있는 나라들은 전통적인 센서스에서의 결점이 완화된다고 주장한다.

전통적 센서스와 행정등록 센서스를 혼합하여 실시하는 국가: 스페인

라. 행정등록 센서스와 표본조사의 혼합형

일부 국가에서는 행정등록 센서스를 표본조사와 병행하여 실시한다. 이것은 등록센서스를 통해서 필요한 모든 정보를 얻을 수 없는 경우에 해당한다. 이 방식은 등록부의 기록과 표본 조사를 통해 얻은 자료를 결합하여 마이크로 수준 즉 개별 단위에서 연결한다. 다음 단계에서는 결과에 따라 자료를 이용해 표를 생성하든지 외삽(extrapolation)한다. 이 방식은 인구센서스의 전수/표본 형태와 유사한 것처럼 보인다. 기본 주제에 대해서는 각 개인 및 가구별로 자료를 수집하고 보다 상세한 주제에 대해서는 개인이나 가구의 표본조사를 이용하여 결과를 외삽하는 방식이다. 따라서 “인구주택센서스의 원칙 및 권고”에 기술되어 있는 표본의 사용에 대한 권고가 이 모형에 적용되지만 가중치 부여 절차에 대한 정교한 작업이 필요하기 때문에 약간의 수정이 필요하다.

이 방식의 주요 장단점은 순수한 행정등록센서스와 유사하다. 가상 센서스에 대한 반대는 거의 없고 무응답의 문제만이 표본조사의 자료가 사용되는 데 있어서 발생한다. 만약 무응답 문제가 표본조사에서 정정될 수 있다면 센서스에서 표본조사의 선택성 문제가 정정되는 것이 가능할 것이다. 단점은 마이크로 데이터에서 표를 생성하는데 더 많은 수고가 필요하다는 점이다. 왜냐하면 가중치 부여 문제가 발생하기 때문이다. 게다가 국민들을 대상으로 면접조사가 없어졌을 때 센서스 결과에 대한 관심을 얻는 것이 더 어려워 질 수 있다.

등록센서스와 표본조사를 병행하는 국가: 이스라엘, 네덜란드

마. 순환센서스(rolling census)

순환센서스는 전통적인 센서스의 대안으로서 연속적인 표본조사를 누적하는 방법이다. 이 방식은 전국을 특정 시점 또는 짧은 기간에 포괄하지 않고 장기간(보통 몇 년) 동안에 포괄하는 방식이다. 순환센서스의 두 가지 주요 요소는 조사기간과 표본 추출률이다. 조사기간은 자료의 보완 빈도와 관계가 있고 표본 추출률은 예산 및 공표에 필요한 지역단위와 관련이 있다. 예를 들면, 한 번의 연간 표본조사를 통해 국가단위의 자료를 생산하고, 5년을 누적하여 소지역 통계를 생산할 목적으로 표본틀을 생성하는 것이 가능하다. 연간 표본조사는 1년간에 걸쳐서 시행될 수도 있고 특정한 달 또는 좀 더 짧은 기간에 시행될 수도 있다.

이러한 방식을 이행하기 위해서는 매우 복잡한 표집 및 모형 기술, 소지역 단위에서의 표집을 허용할 수 있는 우량 질의 표본틀(매년 갱신되는 마스터 주소 파일이 필수적임), 그리고 이 방식에 대해 중앙, 지방 정부 및 사용자 단체 등 주요 이해 당사자들과의 성공적인 대화가 필요하다. 이 방식의 가장 큰 장점은 자료 갱신의 주기가 짧다는 점이다. 전통적인 센서스에서는 자료의 갱신이 5년 또는 10년 만에 이루어지지만 순환센서스에서는 매년 이루어진다. 또 다른 장점은 전통적인 센서스에서의 일시적인 큰 비용 및 인력의 부담을 완화한다는 점이다. 게다가 매년 작업의 개선을 이룰 수 있고 새로운 기술을 시험할 수 있다. 가장 큰 단점은 전체 모집단에 대한 동일 시점에서의 스냅샷을 더 이상 제공하지 않고, 조사 시점의 차이 때문에 지역 간의 비교를 복잡하게 만든다는 점이다. 참고로 서로 다른 시점에서 수집된 자료는 동일한 기준시점에 맞추어 조정된다.

순환센서스를 실시하는 나라: 프랑스

바. 전통적 센서스와 특성치의 연간 갱신

이 방식은 전통적인 센서스의 변이형으로 인구수의 조사에 초점을 두고 센서스 실시 해에는 기본 인구 자료만 수집하는 방식이다. 10년

간에 걸쳐 매년 대규모의 가구 표본조사를 통해 상세한 인구, 사회, 경제 및 주택 자료를 수집하여 센서스 표본조사표를 대신하는 방식이다. 매년 모든 주제에 대한 자료를 수집할 필요는 없다. 왜냐하면 나라마다 필요한 자료가 다를 수 있기 때문이다. 이 표본조사에서는 매년 일정 분량의 주소지를 표본으로 선정하여 센서스 주기의 일정 기간에 걸쳐 표본조사표의 표집률과 비슷하게 만든다. 소규모 행정단위에 대한 추정치의 신뢰도를 높이기 위해서 더 많은 비율의 주소지가 표본으로 추출된다. 표본은 누적되어 전통적인 센서스 표본조사의 표본과 유사한 세부지역단위에서 자료를 생성하게 된다. 표본설계에 따라 표본자료들은 가중치가 부여되고, 무응답 효과를 조정하고 과소포함 또는 과대포함을 수정하게 된다. 이 마지막 가중치 조정은 특성치에 대한 추정치가 주기적인 센서스의 표준치와 비교될 수 있도록 해준다. 마지막 가중치 작업이 완료되면 인구추정치, 비율, 평균, 중간값 등의 통계가 생성된다.

이 방식을 추진하게 하는 주요 동기는 두 가지다. 10년에 한번 센서스를 실시했을 때 얻을 수 있는 것보다 더 시의성 있고 관련된 자료를 제공한다는 점과 센서스와 관련된 작업의 위험을 줄인다는 점이다. 그러나 이러한 계획은 비용이 더 많이 들고 시작하는데 있어 기술적인 어려움이 있고, 포괄적인 계획, 개발 및 시험에 있어 장기 계획을 필요로 한다. 특히 주기적으로 전수조사를 하여 인구수를 구할 것을 요하는 법률이 있는 나라에서는 전수조사가 필수적이다.

전통적인 센서스와 특성치의 연간 갱신을 하는 나라: 미국, 페루

사. 모형 기반 방식 - 개발 중

제안된 센서스의 대안들이 성공의 정도의 차이는 있지만 정보 생산에 관한 다음의 관심사들의 일부를 해결하는 것은 분명하다. a) 소지역 b) 보다 넓은 범위의 주제 c) 보다 세부적이고 서로 다른 견지에서 연구됨 d) 보다 짧은 주기로 e) 생산 비용의 절감. 당분간 자동화된 완전한 등록부를 개발하는 것이 불가능하다면 각 국 통계청은 모형과 추정기법의 사용을 고려하는 것이 유용할 것이다.

비교적 짧은 기간 동안에, 선정된 질문에 관한 정보가 각 개인, 가구 및 거처에 대해서 수집된다. 센서스 중간연도 또는 센서스와 동시에 실시되는 표본조사에서는 과거에 조사된 항목 및 추가항목을 충분히 특성화하는데 필요한 추가 자료를 얻는다. 표본조사의 결과는 전수조사에서 구한 정보와 표본조사에서 얻은 정보와의 관계를 모형화하는 데 사용된다. 이렇게 개발된 모형은 특히 소지역 수준에서 센서스 자료를 설명 변수로 이용하여 표본조사에서 발생하는 정보의 공백을 채우는데 사용된다. 이 제안과 관련된 경험은 빈곤연구 분야에서 발견할 수 있다.

이 방식의 단점 중의 하나로는 동시에 조사를 할 경우에 보다 현장 조사가 더 복잡해진다는 것이다.

2. 미국의 사례(미국지역사회조사)

가. 미국 센서스의 역사

미국은 1787년 헌법에 세계 최초로 센서스 실시를 의무조항으로 두었으며, 센서스 실시 시기와 방법을 구체적으로 명시하였다. 이 헌법³⁾에 따라 1790년 미국 최초의 인구센서스 이래 매 10년 주기로 센서스가 실시되고 있다. 이 최초의 센서스에서는 6개의 항목⁴⁾을 조사하였으며 조사결과를 질병 가능한 청년의 수 등을 파악할 수 있게 하였다. 하원의원 수의 주별 배분 외에도 센서스 자료는 다양한 용도로 사용된다. 1975년 이후로 미국 센서스 국은 주 행정구역 및 선거구 설정에 필요한 소지역 인구자료 생산에 대한 책임을 맡게 되었다. 센서스 자료는 또한 의료보험 등의 정부 프로그램에 대한 예산 배분; 학교, 도로 및 기타 사회시설의 건설 계획 수립; 부동산 중개업체에 필요한 자료제공 및 이주를 원하는 사람들에게 주변 환경에 관한 정보를 제공하고; 미래의 변화를 예측하는 데 이용되는 추세를 알아내는 데 이용된다. 대부분의 센서스 자료는 다양한 지역단위까지 생산된다.

3) 헌법 제1조 제2항에는 하원의원수와 직접세의 주별 배분은 인구수에 비례하여 배분하고 인구수를 언제, 어떠한 방법으로 셀지를 구체적으로 명시하고 있다.

4) 이 6개의 조사항목은 가구주 성명, 가구원 수, 16세 이상의 백인자유인 수, 16세 미만의 백인 자유인 수, 기타 가구원의 성별, 기타 가구원의 피부색이다.

나. 미국지역사회조사(ACS: American Community Survey)

1) 개관

미국지역사회조사는 과거에 10년 단위로 실시한 센서스의 표본조사로 작성한 소지역(센서스 트랙트⁵⁾ 및 블록그룹⁶⁾통계의 연간자료를 생산하기 위하여 개발되었으며, 일련의 월간조사 표본을 사용한다. 소지역 통계를 생산하기 위해서 처음에는 5년간의 표본이 필요한데 5년간의 자료가 축적되고 나면 매년 소지역에 대한 통계가 생산된다. 좀 더 규모가 큰 지역을 대상으로는 3년 자료와 연간 자료를 생산하게 되며 과거의 센서스와 마찬가지로 ACS의 조사대상은 일반거처 및 집단거주시설에 거주하는 사람 모두를 포함하게 된다. ACS는 미국 및 푸에르토리코에서 실시되고 있으며, 푸에르토리코에서는 푸에르토리코 지역사회조사라고 불리어진다(US Census Bureau, 2006).

인구 및 주택의 특성에 관한 상세한 정보를 수집하는 대안으로서 연속적으로 자료를 수집하는 방법이 계속 연구되어 왔었지만 10년 주기의 센서스의 실질적인 대안으로는 생각되지 않았다. 그러나 1990년경에 연방정부, 주정부, 지방정부 및 민간부문의 현 시점에서 국가적으로 일관성 있는 자료에 대한 수요로 인해 정부에서는 사회통계, 경제통계 자료를 10년마다 한 번씩 수집하는 대신에 10년에 걸쳐 연속적으로 수집하는 가능성에 대해 검토하기 시작했다. 현 시점의 자료를 얻을 수 있다는 장점 및 비용절감, 계획수립, 센서스 포괄범위의 개선, 효율적인 수행 등의 예상되는 장점 때문에 센서스 국에서는 ACS를 2000년에 시작하기로 하고 준비하였으나, 신 계획의 표본설계, 조사방법, 데이터 생산품(data products)의 미묘한 점(nuances)을 이해할 필요성 때문에 ACS의 시행은 2000년 센서스 이후로 연기되었다. 이후 추가의 테스트, 이해관련 자들에 대한 접촉, 통계학회, 인구학회 등의 주요 이용자들과의 계속적

5) 센서스 트랙트는 카운티를 분할할 것으로 일반적으로 센서스 트랙트의 인구는 2,500~8,000명이 다.

6) 센서스블록은 자료집계 및 공표의 최소 지역 단위이다. 1990년 센서스에서는 7,020,924개의 센서스블록이 있었다. 블록그룹은 블록보다 한 단계 위의 지역단위로 여러 개의 센서스블록을 묶은 것으로 센서스 표본조사 결과를 공표하는 최소 지역단위이다.

인 접촉 후에 센서스국은 가구단위에 대해서는 2005년에, 집단구역에 대해서는 2006년에 ACS 본 조사를 시행하기에 이르렀다.

ACS의 역사는 4단계로 나눌 수 있다. 연속적인 자료수집의 개념이 발생한 1990-1993년까지의 계획 및 제의 단계, 몇 개의 지역을 대상으로 초기 프로토타입을 시험한 1994-1999년까지의 개발단계, 대규모 전국을 대상으로 조사를 실시하고 전국, 주 및 대규모 지역들에 대한 보고서를 작성한 2000-2004년까지의 적용단계(demonstration stage), 연간 약 300만 주소(미국) 및 약 3만 6천 주소(푸에르토리코)의 거처단위를 대상으로 2005년 1월에 시작한 완전시행 단계로 나눌 수 있다.

가) 계획의 기초 및 초기의 제안들

레슬리 키시(Leslie Kish)는 1981년에 센서스의 개념으로 순환표본설계(rolling sample design)라는 개념을 소개했다. 키시가 연구를 할 무렵에 센서스 국에서도 보다 자주 update된 자료의 필요성을 느끼기 시작했다. 1985년 하원에서는 중간연도 센서스를 허락했지만 예산은 배정되지 않았다. 1990년대 초 하원은 10년 주기의 센서스에 대한 대안에 대해 다시 흥미를 가지기 시작했으며 센서스국은 키시의 연구에 기초를 두고 1990년대 중반에 연속 측정 방법의 개발을 시작했다.

센서스국은 세분화된 표본자료 수집의 대안으로 연속 측정에 대한 연구계획을 마련하였고 연속측정을 위해 세가지 프로토타입을 개발했다. 실행 가능성 및 기술적인 가능성, 정책문제, 비용, 장점에 대한 직원들의 평가에 따라 센서스국은 한 개의 프로토타입을 선택하였고 개발을 계속하였다.

프로토타입 개발에 대해 여러 가지 결정이 이루어졌는데 비용 절감을 위해서는 우편조사를 해야 하고, 전화조사와 면접조사가 뒤따라야 한다는 결론을 내리게 되었다. 프로토타입의 지리적 구성, 표본추출율, 추계인구사용(use of population controls)에 관한 결정이 이루어졌다.

센서스 표본 조사와 같은 신뢰수준을 얻을 수 있는 소지역에 대한 5년치의 자료를 생산하기 위해서 매월 5십만 크기의 거처단위 표본이 제안되었다. 그러나 이 표본의 크기를 감당하기에는 비용이 너무 많이 들게 되었다. 하지만 비표본 오차를 줄임으로써 25만개의 표본으로 수용

할만한 신뢰수준을 얻을 수 있을 것으로 결정되었다.

나) 개발

1994년에 전담 부서의 창설이 이루어져 개발단계가 시작되었다. 전담 부서는 조사 프로토타입의 개발을 계속했고 아래의 설계요소들이 이 조사의 기초로서 검토되었다.

- ① 서로 독립된 월간 표본을 사용하여 1년에 걸쳐 연속적으로 자료를 수집해야한다.
- ② 세 가지의 자료수집방법이 사용되어야 한다(우편조사, 무응답에 대한 전화조사, 무응답에 대한 면접조사)
- ③ 거처 점유 상태 및 다른 많은 특성을 결정하는 조사기준일은 자료가 수집되는 바로 그날이어야 한다. 어떤 항목들에 대한 조사 대상 기간은 더 길어질 수 있다(예: 지난주, 지난 12개월)
- ④ ACS 추정치들은 센서스 중간연도의 인구 및 거처 추정치에 조정되어야 한다.
- ⑤ 모든 추정치들은 일정기간(예: 1년)에 걸쳐 조사된 월간 표본조사에서 수집된 자료를 합쳐서 생산되어야 한다.

초기의 개발보고서에는 여러 가지 형태가 있다. 연속추정계열로 알려진 20개의 보고서가 개발에 있어 아주 중요한 것들이다. 이 보고서들은 1993년에 시작되었고 최종 프로토타입에 반영되었다. 연속추정 계획은 1995년에 미국통계학회가 주관하는 JSA(Joint Statistical Meetings)에 공식적으로 소개되었다. Love, Dalzell과 Alexander(1995)가 이런 조사에 필요한 전제조건에 대해서 소개하였고 Dawson, Sebold, Love과 Weidman(1995)이 전화조사에 의한 자료수집 가능성에 관해 보고서를 작성했다. 당시에 연속추정자료로 평가된 가능한 수정안들에 대해서도 토의가 있었다.

1995년 11월에 4개의 지역(Rockland County, NY; Brevard County, FL; Multnomah County, Or; and Fulton County, PA)에 ACS의 운영시험이 시작되었다. 1996년 11월 시험은 확장되어 다양한 지리인구특성을 가진 지역을 포함하였다. 이러한 지역에는 Harris County, TX; Fort Bend County, TX; Douglas County, NE; Franklin County, OH; Oteto County,

NM이 있다. 이 시험은 방법과 절차의 정당성 평가와 장래 시행을 위해 비용모형을 개발하기 위해 실시되었다. 시험 결과에 따라 프로토타입 설계를 수정하였고 더 필요한 연구 영역을 알아내었다. 소지역추정, 추정방법, 무응답 추적, ACS 시험에서 가중치 부여, 항목무응답, 무응답률 및 농촌지역 자료의 질 등 여러 분야에 걸친 더 많은 연구가 있었다.

운영시험은 계속되었고 1998년에는 남캐롤리나의 두개 카운티 및 플로리다의 1개 카운티가 추가되었다. 남캐롤리나의 두개 카운티는 1998년 센서스 리허설 결과와 비교하기 위해서 1998년에만 포함되었다. 1999년에는 26개 주의 36개 카운티로 시험조사가 확장되었는데, 이 지역은 전국 대표 표본으로 구성되지 않았고 카운티의 인구 규모, 조사의 난이도 및 1990-1995년간의 인구성장을 고려하여 선정되었다. 지리적인 다양성 또한 고려되었고, 인종, 민족, 계절인구, 이동 취업자, 인디언 보호구역, 경제상태의 변동, 주 산업 및 주 직업이 고려되었다. ACS 프로그램을 평가하고 개선하는데 적극적으로 참여할 수 있는 사용자들이 있는 지역을 추가로 선정하였다. 운영시험의 결과에 근거하여 프로토타입의 개선이 이루어졌고 추가적인 연구영역이 확인되었다.

2) 표본틀 개발

가) 개요

미국 센서스국(U.S. Census Bureau)은 ACS의 주소, 기타 인구조사의 표본틀 및 10년 주기의 인구센서스의 주소로 사용하는 국가 마스터 주소 파일(national Master Address File; MAF)을 관리한다. MAF에는 미국 및 푸에르토리코의 일반거처, 집단거주시설 및 선택된 비거주시설(공공, 개인 및 상업용)에 대한 우편주소, 위치주소, 지리코드 및 거주구역에 관한 기타 정보가 담겨 있다. MAF는 TIGER(Topological Integrated Geographic Encoding and Referencing) 파일과 연계되어 있으며 TIGER는 디지털 센서스 지도 및 관련 특성을 표현할 수 있는 데이터베이스로서 자동화된 주소를 특정 지역에 부여하는 기능(geocoding)을 보유하고 있다.

초기 MAF는 2000년 센서스를 위해 개발되었는데 1990년 주소관리

파일, 미국 우편공사의 배달경로파일(Delivery Sequence File; DSF), 현장 조사 확인 및 지방정부가 제공한 주소로 구성되어 있다. MAF 보완은 DSF를 이용하고, Community Address Updating System(CAUS) 등을 이용하여 자동화, 수작업, 현장 확인을 통해 이루어진다.

나) MAF의 내용

MAF의 각 레코드에는 다음과 같은 정보의 전부 또는 일부를 담고 있다.

- 거처의 상태: 현존, 철거, 건축중 등
- 거주용 거처 또는 비거주용 건물
- 위도 및 경도
- 각종 지리정보(주, 카운티, 트랙트, 블록 등)

MAF 주소정보에는 도시형 주소 및 비도시형 주소를 모두 담고 있다.

- 도시형 주소: 예) 201 Main Street, Anytown, MA 01977
 - 전체 주소의 약 94.4% 차지
- 비도시형 주소: 예) 우편함 주소(PO box format) 등

3) 표본설계 및 표본선정

가) 개요

ACS의 표본은 D.C.를 포함한 미국 내의 3,141개 카운티와 푸에르토리코의 78개 기초자치단체에서 각각 선정된다. 일반거처와 집단거주시설에 대한 표본은 분리되어 선정된다. 일반거처에 대한 첫 번째 완전이행 표본은 2005년에 선정되었다. 매년 미국 내 약 3백만 일반거처와 푸에르토리코 내 약 3만6천 일반거처가 표본으로 선정된다. 집단거주시설에 대한 첫 번째 완전이행표본은 2006년에 선정되었으며 연간 집단거주시설 내의 약 2.5%의 사람들이 ACS표본에 포함된다. 표본의 신뢰도를 유사하게 하기 위해서 인구 규모가 큰 지역의 추출률은 인구 규모가 작은 지역보다 상대적으로 작으며 200가구(약 500명) 미만의 지역은 연간 10%를 표본으로 선정한다. 또한 우편 응답률이 낮은 지역의 추출률은 상대적으로 더 크게 한다. 5년 추정치의 표본 오차는 전통적인 센서스 표본조사 보다 1/3 더 클 것으로 예상이 되지만 전문적인 조사원의 채용

및 자료의 시의성으로 인한 비표본 오차의 감소로 인해 상쇄될 것으로 판단된다.

나) 일반거처에 대한 표집

일반거처에 대한 표본은 주표본과 추가표본으로 나누어진다. 주표본은 조사 전년도 8/9월에 선정되며 추가표본은 조사연도의 1/2월에 선정된다. 주표본의 표본들은 최근의 MAF가 이용되고 그 이후에 신축되는 거처에 대해서는 추가표본을 선정한다. 주표본으로 선정된 거처에 대해서는 12개월에 걸쳐 배분되고, 추가표본은 4월부터 12월까지 배분된다.

각 거처가 5년 동안 2번 이상 표본으로 선정되는 것을 방지하기 위해서 표본들은 5개의 부표본들로 나뉘어 지며 각 거처는 이 5개의 부표본들에 랜덤하게 할당된다. 이 5개의 부표본들 중 1개의 표본들이 1년간 사용된다.

선택된 1개의 부표본들을 이용하여 각 카운티별로 여러 가지 특성에 따라 주소지별로 정렬한 다음 1년 동안에 사용할 표본 거처를 계통 추출한다. 1년간 사용될 표본거처들은 다시 12등분하여 조사할 월에 할당된다.

〈표 1〉 주요 공표 단위 및 공표 주기

공표 단위	1년 추정치	3년 추정치	5년 추정치
	인구 6만5천 이상	인구 2만 이상	지리 구역 합계
United States	1	1	1
Census Regions ⁷⁾	4	4	4
Census Divisions ⁸⁾	9	9	9
States	51	51	51
Counties	761	1,811	3,141
Minor Civil Divisions	97	592	16,536
Places	476	1,983	25,161
American Indian and Alaska Native Areas	15	41	768

Metropolitan, Micropolitan, and Consolidated Statistical Areas	561	905	923
Congressional Districts	436	436	436
School Districts	879	3,290	14,505
Census Tracts	-	-	65,443
Block Groups	-	-	208,790
Puerto Rico	1	1	1
Municipios	12	65	78
Zonas Urbanas, Comunidades	9	20	225
Census Tracts	-	-	861
Block Groups	-	-	2,477

표집은 블록 단위로 하는데 블록을 포함하는 최소 공표단위(<표 1> 참조)의 크기에 따라 다르며 표집률은 <표 2>와 같다.

<표 2> 최소 공표단위별 연간 표집률

최소 공표 단위	표집률	
	미국	푸에르토리코
200가구 미만	10.0%	10.0%
200 ~ 800 가구 미만	6.9%	8.1%
800 ~ 1200 가구	3.6%	4.1%
1201가구 이상	2.3%	2.7%

4) 조사방법

표본으로 선정된 각 거처단위 주소는 특정 월에 배정되어 우편질문지를 받게 된다. 우편 발송에 적합하지 않는 주소는 발송불가주소로 분류된다. 거처와 거처내의 거주자들에 대한 조사는 배정된 월 및 그 다음

7) census region은 여러 개의 주를 합한 것으로 west, midwest, northeast, south로 구분된다.

8) 각 census region은 2~3개의 census division으로 구성된다.

두 달간에 걸쳐 작성가능하다. 우편 발송된 주소에서 응답이 없고 전화 번호가 있으면 CATI 담당직원을 통해 다음 달에 전화조사를 하게 된다. CATI조사에서도 응답이 없는 주소에 대해서는 이 중 표본을 선정하여 세 번째 달에 CAPI조사를 행하게 된다. 각 집단거주시설 표본은 특정 월에 배정되고 면접조사를 통해서 6주간에 걸쳐 조사를 하게 된다.

가) 우편조사

우편 발송이 가능한 약 95%의 가구에 대해서 매월 조사표류를 우편으로 발송하는데 ① 조사협조 공문 발송 ② 조사표류 발송 ③ 무응답 가구에 대해 조사독촉 공문 발송 ④ 조사표류 재 발송의 과정을 거친다.

우편조사에 대한 응답률은 약 50%로서 센서스 표본조사 응답률과 비슷하다.

나) CATI 조사

무응답 가구에 대해서는 우편조사 다음달에 3개의 콜 센터를 통해서 전화조사를 약 4주간에 걸쳐서 실시하며 CATI 조사의 응답률은 약 60%이다.

다) CAPI 조사

우편조사와 CATI 조사에서 무응답한 가구 및 우편발송불가 주소지에 대해서는 CATI 조사 다음 달에 면접조사를 하게 된다. 면접조사 대상 가구는 약 1/3을 표본으로 선정하여 조사한다. 무응답 가구에 대해 표본을 추출하여 면접조사를 실시하는 이유는 면접조사의 비용이 우편조사 비용의 약 10배가 들기 때문이다. 면접조사는 12개 지역사무소를 통해서 실시하며 조사원은 임시직이 아닌 정규직원이다. 또한 3년 이상의 조사 경험자를 슈퍼바이저로 이용하며 면접조사의 응답률은 93% 이상이다.

라) 전체 응답률

전체 응답률은 95% 이상이며, 표본설계의 복잡성 때문에 전체 응답률 계산에는 가중치가 이용된다.

〈표 3〉 2001년 ACS의 응답률

구 분	응 답 률
우편조사	51.1%
CATI	9.2%
면접조사	36.4%
무응답	3.3%

5) 자료처리

1년간 수집된 자료를 누적하여 당해 연도의 추정치를 계산한다. 자료처리에는 부호기입(coding), 편집(editing) 및 대체(imputation)의 3가지 주요 과정이 있다.

가) 부호 기입

부호기입에는 자동화 방식의 이용 및 수기로 진행하는데 인종, 사용 언어, 일하는 장소, 산업, 직업 등의 부호기입이 있다.

나) 편집

편집 과정에는 일차적으로 응답 조사표와 무응답 조사표로 분리하고 무응답으로 분류된 조사표는 버리고 나중에 가중치로 조정한다. 응답 조사표와 무응답 조사표는 조사표의 완성도에 따라 구분한다. 응답 조사표로 분류된 조사표에 대해서는 항목간의 논리가 맞지 않거나 무응답 항목이 있는 경우, 자동화된 시스템을 이용하여 찾아내어 대체한다.

다) 대체

대체 방법은 크게 부여(assignment)와 할당(allocation)으로 나눌 수 있다. 부여는 규칙에 근거한 것으로 해당 조사표 내의 다른 항목을 이용한다. 예를 들면 사람의 이름을 사용하여 성별을 부여하는 것이다. 할당은 다른 조사표의 내용을 사용하여 대체하는 것으로 핫덱(hot-deck) 방법을 사용하여 이웃의 정보를 사용한다. ACS의 대체율(imputation rate)을 2000년 센서스의 대체율과 비교해 보면 매우 낮은 것을 알 수 있다(<표 4> 참조).

〈표 4〉 ACS와 2000년 센서스의 대체율 비교

항 목	대체율(Imputation Rates)				
	2003 ACS	2002 ACS	2001 ACS	2000(C2SS) ⁹⁾	2000 Census
Number of vehicles available	1.0	1.1	1.3	1.6	6.2
Place of birth	6.2	4.4	4.6	6.4	10.1
Citizenship	0.4	0.4	0.4	0.5	0.8
Previous residence					
Mobility Status	2.2	2.5	2.6	4.0	6.9
Previous residence geography(one of more parts)	5.9	6.0	7.3	14.9	11.0
Employment status recode	3.4	3.5	3.8	6.0	10.9
Place of work geography (one of more parts)	5.2	4.9	5.3	9.9	10.7
Means of transportation to work	3.1	3.0	3.1	4.6	7.6
Private vehicle occupancy(car pooling)	4.1	3.9	4.1	5.8	10.0
Time leaving home to go to work	9.6	9.2	9.9	11.3	15.0
Travel time to work	7.0	6.9	7.2	8.7	11.8

6) 가중치 부여 및 추정(Weighting and Estimation)

가중치 부여는 세단계로 나눌 수 있다. 첫 번째는 표본설계시의 추출률이며, 이 추출율은 무응답률을 감안하여 조정되고, 추계인구를 이용하여 마지막으로 조정된다. 추계인구는 과거의 센서스 자료를 출생, 사망 등의 행정자료를 사용하여 보정한다. 추정은 연간 추정치, 3년 추정치

9) Census 2000 Supplementary Survey의 약어로 2000년 센서스 당시에 실시한 ACS의 시험조사임.

및 5년 추정치로 분류된다. 월간 표본조사를 누적하여 연간 추정치를 계산한다. 연간 추정치는 현재의 공표 지역 단위로 추정되며, 지역의 경계가 바뀌면 바뀐 지역의 추정치를 계산한다. 3년 추정치 및 5년 추정치는 연간 추정치들의 평균으로 계산된다. 중간 값의 계산은 연간 평균치의 중간 값을 사용하지 않고 전체 데이터의 중간 값으로 한다.

7) 자료 공표

연간 자료의 공표는 인구 6만5천 이상의 지역을 대상으로 하며 조사 다음 해 8월에 공표한다. 인구 6만5천명 이상의 카운티는 전체 인구의 82.5%를 차지한다.

3년 추정치의 공표는 인구 2만 이상의 지역을 대상으로 하며(2005~2007년의 자료는 2008년 8월에 공표), 인구 2만 이상인 카운티는 전체 인구의 95.5%를 차지한다.

5년 추정치의 공표는 인구 2만 미만의 지역(2005~2009년의 자료는 2010년 8월에 공표)을 대상으로 한다.

자료 공표 지역은 우편번호(Zip Code Areas), 학군(School Districts), 주 행정구역(State Legislative Districts), 인디언 보호구역(American Indian Reservations), - PUMAs(2000년 센서스 구역 중 인구 10만 이상), 센서스 트랙트(Census Tract) 등으로 분류할 수 있다.

ACS는 표본조사이기 때문에 연간 자료에 대해서는 점 추정치 및 신뢰구간을 발표하며 90% 신뢰구간을 사용한다.

〈표 5〉 90% 신뢰구간 사용 예

	추정치	신뢰구간	
		하한	상한
총 인구	373,815	369,411	378,219
남자	181,932	179,413	184,451
여자	191,883	189,383	194,383
5세 미만	28,015	26,762	29,268
5세 ~ 9세	26,222	24,323	28,121

8) 기타

ACS는 센서스 표본조사의 대체용이므로 헌법에 의해 의무적인 조사이다. 불응에 대해서는 \$100 ~ \$5,000까지의 벌금을 부여할 수 있다.

ACS의 조사 항목 수는 60개로 센서스 표본조사의 조사항목 수와 같으며 조사표 작성 시간은 약 38분이 소요된다. ACS의 총 비용은 10년간 누적하면 전통적인 센서스 비용과 거의 비슷하며 2005년 ACS의 총비용은 1억4천6백만 달러이다. 전통적인 센서스와 ACS의 차이점을 살펴보면 아래 표와 같다.

〈표 6〉 전통적인 센서스와 ACS의 차이점

	전통적인 센서스	ACS
조사 시점	0년 4월 1일	매월 조사
표본조사율	6가구 중 1가구 표본조사	5년간 8가구 중 1가구 조사
조사항목	5년전 거주지	1년전 거주지
결과의 빈도	10년 주기	연간 자료(대규모 지역), 3년 자료(중규모 지역), 5년 자료(소규모 지역)
소득	지난 1년간의 수입 조사 (2000년 센서스의 경우 1999년 1년간의 수입)	조사 시점에서 지난 1년간의 수입 조사

3. 프랑스의 사례(순환센서스)

가. 프랑스 센서스의 역사

프랑스에서는 나폴레옹의 동생 루시안에 의해 1801년 처음으로 인구 센서스가 실시되었다. 나폴레옹이 인구센서스를 실시한 주된 목적은 두 가지다. 첫째는 군사목적에 의한 것으로 전쟁을 위해 청년을 징병할 필요가 있어 각 지역의 청년수를 파악하는 것이고, 또 하나의 목적은 원로원 및 입법원의 의원 수는 나폴레옹헌법에 의해 지역의 인구수에 맞추

어 정하도록 했기 때문에 각 지역별 인구수를 확정할 필요가 있었다. 인구센서스는 이 후 5년마다 실시되었다. 그러나 1939년의 2차 세계대전 발발 후 인구센서스는 부정기적으로 1946, 1954, 1962, 1975, 1982, 1990, 1999년에 실시되었다. 프랑스는 1999년 센서스 이후 전통적인 인구센서스를 포기하기로 결정하였는데 그 주된 이유는 센서스의 주기가 길어지고, 예산문제 및 지방자치제의 확산으로 인한 시의성 있는 소지역 통계에 대한 수요가 늘어났기 때문이다. 이러한 문제에 대응하기 위하여 프랑스의 순환센서스는 탄생하게 되었다.

나. 프랑스 순환센서스(Rolling Census)

1) 개관

2004년에 프랑스는 순환센서스라는 새로운 센서스 방법을 도입하였다. 전통적인 센서스의 대안으로서 특정 시점에서 전국을 포괄하는 센서스를 실시하는 대신에 연속적인 표본조사를 실시하여 장기간에 걸쳐 전국을 포괄하는 방법을 순환센서스라고 한다. 표본조사를 하는 방식은 전국을 지역별로 나누어 돌아가면서 각 지역을 전수 조사하는 방식과 모든 지역에서 표본을 뽑아 조사하는 방법 등 여러 가지가 있을 수 있다. 순환센서스는 이런 식으로 매년 표본 조사된 자료를 누적하여 추정하는 방법을 사용한다(Kish, 1990). 현재까지는 프랑스가 전 세계에서 유일하게 순환센서스를 실시하는 나라이다.

프랑스에서 순환센서스를 도입한 이유는 두 가지가 있다. 첫 번째는 보다 최근의, 보다 정기적인 정보에 대한 수요의 증가이며, 두 번째는 전통적인 센서스에 관련되어 왔던 예산 및 인력 부담 문제를 해결하기 위해서이다(INSEE, 2006).

신선한 자료의 필요성은 프랑스 통계청(INSEE)의 고객들에 의해서 더 많이 노골적으로 표명되어왔다. 센서스의 주기가 점점 더 길어짐(지난 두 센서스 간 9년)에 따라 사회나 가족의 변화든 도시주변에 정착하는 경향이든, 도시 개발이든 간에 프랑스 인구 변화를 추적하는 것이 점점 더 어렵게 되어가고 있으며 지방 분권화에 의해 권력의 점진적인 위임 때문에 센서스 자료에 대해 지역사회들은 더욱 관심을 기울이게 되

었다. 예를 들면, 기초자치단체 또는 기초자치단체 연합체들은 도시 교통에 대해 새로이 책임을 맡게 되었으며, 사회복지관련 부서들은 중·고등학교에 대한 투자 및 운영, 직업훈련, 지역개발계획 등의 분야에서 더욱 많은 부분을 담당하게 되었다.

프랑스는 1999년 이전까지는 전통적인 센서스를 실시하였지만 센서스 실시에 대해 헌법상에 명문 규정이 없어 센서스를 실시할 때마다 법률을 제정하여야 했다. 또한 프랑스 선거를 고려해야 했기 때문에 보통 7~9년마다 센서스를 실시하였다. 조사주기가 긴 것 외에도 예산이나 인력 문제를 해결하기가 쉽지 않았다. 1997년에 계획된 센서스가 1999년으로 연기된 이유도 예산문제 때문이었다.

2) 표본틀 개발

각 대규모¹⁰⁾ 기초자치단체의 표본틀은 관련된 건물등록부(RIL; located building register)이며, 건물등록부는 주거용, 행정용, 산업용, 상업용 건물 등의 모든 건물의 목록이며 지리정보시스템을 이용하여 확인 가능하다. 건물등록부는 1999년 센서스를 기초로 편집되었고 건축허가, 지방세, 우편주소 등의 행정자료를 사용하여 계속 보완된다. 건물등록부는 매년 기초자치단체가 점검하며 INSEE가 마지막으로 확인한다.

3) 표본설계

프랑스의 기초자치단체는 그 수도 많고 매우 다양하기 때문에 특별한 표집 방법이 개발되었다. 약 3만 7천개의 기초자치단체가 있고 그 중 절반은 인구가 400명 미만이다.

인구 만명 미만의 모든 기초자치단체에 대해서는 센서스가 실시된다. 해마다 기초자치단체 1/5에 대해서 센서스가 실시된다. 인구 만명 미만의 35,750개의 기초자치단체는 5개의 집단으로 나누어진다. 이 기초자치단체들은 전체 인구의 절반을 차지한다. 매년 한 개의 집단에 속하는 모든 기초자치단체들에 대해서 센서스를 실시한다. 첫 번째 집단에 속하는 기초자치단체들은 5년 후에 다시 조사된다.

10) 인구 만 명 이상의 자치단체를 말한다.

5개의 집단은 인구(인구수, 성별, 연령계층), 기타 주거 관련[거처(dwelling) 수, 주 거주지(principal residence) 수] 변수 등 약 10가지 기준에 의해 균형을 맞추어 형성된다. 균형은 전국단위 및 프랑스의 26개 광역자치단체 단위에서도 이루어진다.

인구 만 명 이상의 900개 기초자치단체에서도 매년 센서스 표본조사가 이루어지지만 일부분만 조사된다. 매년 각 기초자치단체에서 8%의 거처가 조사된다. 따라서 5년 후에는 40%의 인구가 조사된다. 이것은 기초자치단체 및 기초자치단체의 각 구역에 대해 안정적인(robust) 정보를 보장하기에 충분한 표본의 크기이다.

각 기초자치단체의 센서스 표본은 다음의 표집 방법을 사용한 주거지의 부분집합에 기초를 둔다.

- (a) 집락 효과를 피하기 위해서 기초자치단체에서 가장 큰 주소들은 전부 조사한다. 이 주소들은 5개의 연간 집단으로 나뉜다.
- (b) 새로운 주소들은 전부 조사된다. 표집에 필요한 정보가 없기 때문이다. 이 새로운 주소들도 5개의 집단으로 나뉜다.
- (c) 기타 주소들은 인구 또는 주거-저량(stock) 기준에 따라 5개의 균형을 이루는 집단으로 나뉜다. 각 집단은 기초자치단체에서 골고루 분포된다.¹¹⁾ 매년 조사될 표본주소들은 올해의 집단에서 뽑히게 되며 조사될 모든 주소들은 기초자치단체의 전체 거처의 약 8%이다.

이런 식으로, 포괄적인 센서스가 매년 인구 만 명 미만의 기초자치단체의 1/5, 인구 만 명 이상의 기초자치단체 인구의 8%에 대해서 실시된다. 즉 4백5십만 거처와 9백만 인구가 매년 조사되는 셈이다. 5년 후에는 소규모 기초자치단체 전부와 대규모 기초자치단체에 거주하는 인구의 40%가 조사되게 되며, 전체적으로 70%의 인구가 5년 주기 동안 조사된다.

마지막으로 집단가구에 거주하는 집단, 이동주택에 사는 사람들, 그리고 노숙자들은 대규모로 조사된다. 이런 식으로 프랑스 영역에 살고 있는 모든 개인들을 조사하게 된다.

11) 따라서 동일한 거리(street)에 있는 주소들이 서로 다른 집단에 속할 수 있다.

4) 조사방법

센서스 표본조사는 매년 동일한 날에 관련된 모든 기초자치단체에서 시행된다¹²⁾. 자료수집 방법(protocol)은 조사원이 질문지를 전달하고 수거하는 방식이다. 표집방법은 혁신적으로 변화했지만 자료수집의 경우는 일반적인 센서스에서 사용되는 방식을 그대로 따른다. 즉 조사원이 전달하고 수거한다. 각 가구는 2가지 질문지를 받게 되는데, 그중 하나는 가구에 관한 것이고 다른 하나는 가구원에 관한 질문지이다. 질문지는 과거의 센서스에 사용된 질문지와 큰 차이가 나지 않는다. 가구에 관한 질문지는 가구원의 목록 및 주거의 특징과 기준에 관한 약 14개의 질문과 자동차 보유 대수로 구성되어 있다. 가구원에 관한 질문지는 나이, 성별, 출생지, 국적, 5년 전 거주지, 교육정도, 직업, 근무지 또는 통학지에 관한 25개의 문항으로 구성되어 있다.

조사가 불가능한 경우(부재, 장기간 출타, 응답거부)에는 주 거주지에 대해서 조사원이 미조사 거처 일람표(non-surveyed dwelling form)에 추정 가구원수를 기입하도록 하여 인구수를 조정하는데 이용된다.

5) 자료처리

2008년에 시작되는 세분된 파일(조사된 개인이나 가구당 기록)이 5년간의 조사된 파일을 합침으로써 매년 생산될 것이다. 따라서 일반적인 센서스처럼 교차표 생성(cross-tabulation)이 가능하다. 매년 조사된 9백만 개인 및 4백5십만 거처가 가중치 조정 후에 이 파일에 숫자로 나타난다. 따라서 세분된 파일은 모든 개인을 설명한다.

조사의 포괄율(만 명 미만의 기초자치단체의 경우 100%, 만 명 이상의 기초자치단체의 경우 40%) 및 정보가 개인별로 가중치가 부여된 자료들의 파일로 정리되어 있다는 사실은 아주 세분된 정보가 지리적으로 세분되고 소규모 집단에 대해서 추출될 수 있다는 것을 의미한다. 누적된 파일은 4천5백만 개인과 2천2백만 거처로 구성되어 있기 때문에 상세한 표를 만들 수 있다.

12) 1월 3번째 목요일에 시작함

6) 추정

매년 수집된 정보는 5년 주기의 중간점에서의 특정한 날에 적용하기 위해서 조정된다. 대규모 기초자치단체의 경우에는 5년간 표본의 평균 값을 계산한 후 중간년도의 RIL의 거처 수에 맞추어 판단한다. 소규모 기초자치단체에 대해 사용되는 방법은 센서스 표본조사와 인구 기준 일간의 내삽 또는 외삽이다. 외삽은 주택세 등록자료를 이용하여 합쳐진다. 주택세 등록자료는 연간 주택 저량의 변화를 나타내는 지표의 역할을 한다. 5년 순환을 이용하면, 2년이 초과할 경우에는 내삽/외삽은 불필요하다. 따라서 통계숫자는 매우 믿을만하다.

인구 만 명 이상의 기초자치단체에 사용되는 방법은 5년간의 표본에 기초한 순환평균(rolling average)이다. $y-4$ 년에서 y 년의 5개의 표본의 총합에서 거처당 평균 인구수가 계산된다. 이 평균 인구수가 중간기간($y-2$ 년)을 대표한다. 이 평균 인구수에 $y-2$ 년 1월 1일 현재의 거처수(RIL을 이용)를 곱하여 기초자치단체의 인구를 구한다.

인구 만 명 미만의 기초자치단체에 대해서는 y 년 말에 $y-2$ 년 초의 인구를 구하는 것이 필요하다. 인구 만 명 이상의 기초자치단체와 동일한 기준일을 사용한 정보를 제공하는 것이 중요하기 때문이다. 인구 만 명 미만의 기초자치단체들은 순환기간 전체에 걸쳐서 조사된다. 즉 $y-4$ 년에 1/5, $y-3$ 년에 1/5, 등으로 조사된다. $y-2$ 년에 조사된 기초자치단체들의 센서스 결과는 그대로 저장된다. 반면에 $y-1$ 년, y 년에 조사된 기초자치단체들의 $y-2$ 년 인구는 센서스 표본조사와 직전에 공표된 자료를 내삽(interpolation)하여 구한다. $y-4$ 년과 $y-3$ 년에 조사된 기초자치단체들에 대해서는 센서스 표본조사와 $y-2$ 년의 결과를 외삽(extrapolation)하여 구한다. 이것은 지방 주택세(local housing-tax) 자료에 기초를 둔다. 지방 주택세는 기초자치단체당 주택수의 변화를 나타내고 주택수의 변화와 거주자수의 변화간의 차이를 고려하여 조정된다. 지난 두 센서스 간의 차이는 인구의 변화를 주기위해서 주택세자료에 의해서 측정되는 변화에 적용된다.

7) 자료 공표

이 센서스는 매년 각 기초자치단체에 대한 상세한 정보를 제공할 것이며 대규모 기초자치단체의 경우에는 통계 목적으로 합쳐진 대략 2천 명의 거주자로 구성된 IRIS로 알려진 구역(blocks)에 대한 상세한 정보를 제공할 것이다.

3년간의 자료수집이 끝난 시점에서 신개념 센서스 실시에 관한 첫 번째 경과 보고서를 제공하는 것이 가능해졌으며 이것은 전반적으로 긍정적이었다. 새로운 방법에 대한 기초자치단체 및 전 국민의 폭넓은 지지가 있었다. 2005년과 2006년 센서스에는 2004년 센서스에 반대한 2개의 기초자치단체들을 포함한 모든 기초자치단체들이 참여하였다. 일반 국민 쪽에서는 대규모 센서스보다 관심이 적어져 표본조사 방법에 대한 두려움이 나타나지 않았다. 이는 매우 효과적인 전국 및 지역적인 홍보 때문이었다. 마지막으로 INSEE는 인구 만 명 미만의 모든 기초자치단체에 대한 인구추정치를 처음으로 제공할 수 있게 되었다. 그리고 2004년 후반기 이후로 약 100여 개의 대규모 기초자치단체에 대한 인구추정치도 제공할 수 있게 되었다. 이러한 결과들은 인터넷을 통해 2005년 센서스 시작 이틀 전에 발표되었다. 2005년 후반기에는 2006년에 조사된 인구 만 명 미만의 기초자치단체 및 또 다른 약 100여 개의 대규모 기초자치단체로 확대되었다. 그 결과는 2006년 1월 17일에 공표되었다. 이런 식으로 초기 성과물들을 발표함으로써 새로운 방식의 가치에 대해 사용자들을 확신시키게 되었다.

8) 프랑스 순환센서스와 전통적 센서스의 비교

순환 센서스는 INSEE, 기초자치단체 그리고 응답자들에 부여하는 부담을 완화하는 수단으로서 도입되었다. 5년간의 누적표집률은 매우 높기 때문에(대규모 기초자치단체의 경우 40%) 정확한 결과를 보증한다. 이와는 대조적으로 전통적 센서스에서는 대개 약 15% 또는 25%의 표본을 사용한다.

표본의 사용 및 자료수집절차의 연간화는 INSSE, 각 기초자치단체의 지역 센서스에 대한 책임이 있는 조정자들 및 자신의 임무에 집중하는

조사원들에게 도움을 주며 더 양질의 자료수집을 가능하게 한다. 인구만 명 이상 기초자치단체의 조사원들은 조사해야 할 주소 목록을 가지고 있다. 이것은 매우 중요한 점이다. 왜냐하면 이는 조사원들이 어디로 가야할지 정확히 알고 있다는 것을 의미하기 때문이다. 이 방법은 일반 센서스에 사용되는 지역내 모든 대상처를 모두 찾아 조사하는 방식보다 더 효율적이며, 미조사거처일람표와 결합할 경우 무응답의 문제를 줄이거나 보완하는데 도움이 된다.

순환 센서스는 또한 일을 분산함으로써 비용을 분산하고 안정적인 노동력 및 예산으로 운영하는 것을 가능하게 한다. 반면에 조사규모를 줄이기 되면서 충실한 준비를 할 수 있게 되었고 관리를 더욱 잘 할 수 있게 되었다.

표본조사의 무작위성이 다소 부정확함을 수반할지라도 소지역에 대해서는 매년 수정된 정보가 노후화된 정보보다 더 선호된다는 점은 분명하다. 순환 센서스는 10년 주기의 일반 센서스보다 도시재개발 정책 수행에 훨씬 더 유용하다.

9) 순환 센서스의 성공 요인

순환 센서스의 성공여부는 여러 가지 요인에 의해 결정되는 바, 기술적인 요인(기초 자치단체 하위 수준의 지리정보시스템의 사용가능성 및 보완에 필요한 행정자료의 사용가능성), 방법론적인 요인(특히 표본설계기법), 정치적인 요인(센서스 동반자로서의 기초자치단체와의 광범위한 협의 및 센서스 사용자들과의 협의)이 있다.

기술적으로 양질의 표본 틀이 중요하다. 표본 틀을 보완하기 위한 행정자료의 이용이 가능해야 하고 행정자료의 사용에 관한 지침을 세워야 한다.

기초자치단체와의 협동이 가장 중요하다. 왜냐하면 이러한 파일들의 내용이 지방 단위에서 정확히 점검되어야 하기 때문이다. 그리고 현장 상태에 대한 지방의 이해가 훨씬 더 높기 때문에 이것은 매우 중요한 일이다. 행정자료가 없는 경우에는 프랑스의 해외에 있는 도(DOM; France's overseas department)의 경우처럼 공무원들을 현장에 보내서 주택을 확인하는 것이 필요하기 때문에 표본 틀의 보완은 느리고 성가신

작업이 된다.

성공을 위한 또 다른 중요한 요인은 행정자료원의 사용가능성이다. 행정자료를 사용하면 표본틀(sampling base)을 보완할 수 있으며 필요할 경우에 조사 결과를 평가할 수 있을 것이다.

표본을 최적화하고, 조사결과를 외삽하고, 자료수집 시기와 관계없이 동시성을 가지게 하기 위한 조사와 모형화에 있어 방법론적인 전문성은 필수적이다. 균형표본기법(balanced-sample technique)과 기능에 대한 지식은 표본최적화-즉, 예산수립 및 추정치의 질-에 있어 극히 중요하다.

충분하고 일정한 예산이 있어야 한다. 순환 센서스는 센서스 비용을 줄이지 않고 다만 예산을 장기간에 걸쳐 분산하고 정보가 매년 생산되기 때문에 더 좋은 산출물을 보장한다. 매년 프랑스의 센서스 비용은 약 일반 센서스의 1/7이다. 한 번 시작된 후에는 계획은 방해받아서 안 된다. 잃어버린 1년을 복구하는 비용은 평상시의 1년간의 예산보다 훨씬 높기 때문이다.

따라서 전통적인 센서스에서 순환 센서스로의 전환은 많은 중요한 전제조건들과 철저한 준비(표본틀의 작성, 표집, 예산 협상 등)가 필요하고, 그렇기 때문에 상당한 협동적인 노력이 필요하다.

변화를 위해 준비하는데 있어 중요한 단계는 센서스 결과 이용자와의 협의이다. 10년 주기 자료의 snapshot 시스템에서 일정기간의 통계적확률표본(statistical probability representative)인 연간생산자료 시스템으로의 전환은 하룻밤사이에 이루어질 수 있는 것은 아니다. 프랑스에서는 또한 센서스의 오랜 참여자이며 신 센서스에서 동반자인 기초자치단체들과의 긴밀한 협의가 있었다. 상담에는 센서스 숫자들에 의존하는 국가적, 지역적 수준에서의 정치의사결정들도 포함되어야 한다. 만약 그들이 새로운 시스템을 승낙한다면 그들은 통계 서비스를 완전히 신뢰하여야 한다.

자료 수집 및 공표에 있어서는 집단적인 노력이 현재 진행 중이다. 국가센서스평가위원회가 국가통계정보위원회(CNIS; National Council on Statistical Information)에 설립되었다. 상원의원이 위원장을 맡고 있고 INSEE, 기초자치단체들과 센서스 사용자들을 끌어들이므로써 국가센서

스평가위원회는 자료수집 및 감독과정을 감시하고 여러 가지 프로토콜에 관해 수정안을 제시하고 센서스와 관련된 규제에 관한 수정 의견을 줄 것으로 기대된다. 자료 공표에 있어서는 INSEE와 기초자치단체들의 워킹그룹은 하위-기초자치단체 단계 및 그 위의 단계에서의 자료 배포 및 서비스에 관한 포럼을 제공할 것이다.

4. 네덜란드 사례(가상 센서스)

가. 네덜란드 가상 센서스(Virtual Census)

1) 개관

네덜란드는 1971년 센서스를 실시한 이후 국민들의 사생활보호의식의 확산 및 이로 인해 예상되는 무응답 증가 때문에 전통적인 센서스를 포기하게 되었다. 그러나 센서스 형태의 자료는 1981년과 1991년에도 여전히 필요하였기 때문에 네덜란드 통계청은 행정등록 자료와 기존의 표본조사 자료를 이용하여 결과표를 만들었다. 그러나 당시에는 행정자료를 개인 수준에서 이용할 수 없었기 때문에 표본조사 자료와 연결하여 사용하는 것은 불가능하였다(Statistics Netherlands, 2004).

1994년 전산화된 인구등록시스템이 도입되어 인구등록 자료와 표본조사 자료의 연결 가능성이 생겼다. 네덜란드 통계청에서는 인구등록시스템을 기반으로 하여 사회통계데이터베이스(Social Statistics Database; SSD) 구축을 시작했다. 이렇게 구축된 사회통계데이터베이스와 기존의 표본조사 자료를 이용하여 2003년 네덜란드 통계청은 2001년 네덜란드 센서스 결과표를 작성하였으며 가상센서스(Virtual Census)라는 이름을 부여하였다.

전통적인 방식으로 센서스를 실시하였다면 약 3억 유로의 비용이 들었겠지만 가상 센서스의 비용은 3백만 유로에 그쳤다. 3백만 유로에는 새로운 방법 개발 비용 및 관련 소프트웨어 비용, 결과 분석 비용이 포함되었지만 등록부의 비용은 포함되지 않았다.

또한 유럽통계국 및 기타 국제기구는 2001년 센서스 라운드에서 1991 센서스 라운드와 비해 상세한 정보를 요구하였으며, 센서스 결과

표의 전반적인 수치의 일관성이라는 요구사항을 충족하기 위해서 네덜란드 통계청은 반복가중기법(repeated weighting)이라는 새로운 추정방법을 개발하였다. 반복가중기법을 이용하면 자료원이 서로 다른 자료를 이용한다 할지라도 수치적으로 일관성 있는 결과표들을 생성할 수 있다. 반복가중기법은 회귀방법을 반복 적용하여 서로 다른 자료원을 사용한 추정치들 간의 수치적 불일치를 제거하는 방법이다.

2) 사회통계데이터베이스(Social Statistics Database; SSD)

1996년 네덜란드 통계청은 개인에 관해 이용 가능한 모든 정보를 보관하는 사회통계데이터베이스의 개발을 시작한다. SSD는 기초자치단체의 인구등록부에 기반으로 하여 사회보장, 고용보험 등의 행정자료 및 노동력 조사, 국민건강조사 등의 표본 조사 자료를 인구등록자료와 연결(match)하여 작성된다. 이 자료들은 성별, 생년월일, 주소 등의 변수를 이용하여 연결한다. 주민등록번호(social-fiscal number)는 1997년에 도입되었으며 많은 행정자료에 이용된다. 가구 단위에서는 지리주소파일, 주택등록, 가계소비조사(household budget survey) 등이 연결 고리로 이용된다. 목표 매칭률은 90%(오매칭률 5% 이하) 이상이다.

3) 2001년 가상센서스에 사용된 자료원

2001년 센서스 결과표의 생성을 위해 많은 자료들이 사회통계데이터베이스에서 구해진다. 사회통계데이터베이스는 인구 및 사회경제 자료를 포함한 데이터 파일을 마이크로 링크하거나 마이크로 통합한 것이다. 센서스의 결과표를 집계하기 위해서는 사회통계데이터베이스의 모든 통합된 파일이 필요한 것은 아니다. 필요로 하는 대부분의 정보는 인구적인 측면에 대한 정보이며 사회통계데이터베이스의 핵심인 인구등록부에서 구할 수 있다. 취업자에 대한 정보(고용인, 고용주 및 자영업자)는 사회통계데이터베이스 내의 통합직업파일(Integrated jobs file)에서 추출된다. 은퇴인구에 관한 정보(연금 및 생명보험)는 사회통계데이터베이스 내의 통합수혜파일(Integrated file of benefits)에서 구한다. 교육정도, 직업 및 실업, 비경제활동 등에 관한 등록부에서 구할 수 없는 정보는 노동력 조사가 주 자료원이다.

가) 인구등록자료

네덜란드 통계청에서 작성하는 인구 및 가구통계는 전산화된 지방자치단체 인구 등록부들(automated municipal population registers)에 기초를 두고 있다. 이 등록 체계는 GBA(Gemeentelijke Basis Administratie persoonsgegevens) 시스템으로 알려지고 있으며 지방자치단체의 인구자료에 대한 기본 등록이다. ‘기본’이라는 말은 GBA가 지역 등록부 시스템 내에서 인구자료에 관한 기본 등록부의 역할을 한다는 사실을 말한다. 이러한 등록부에는 “사회보장에 관한 지역 등록부”, “전기와 수도에 관한 지역 등록부”, 네덜란드에 거주하는 “외국인들을 다루는 경찰서의 지역 등록부” 및 “노인 연금에 관한 (국가)등록부”들이 포함된다. GBA 시스템은 1994년 10월 1일에 도입되었으며 분산형의 포괄적이고 결합력(cohesive) 있는 인구등록 시스템이다. 법적 조항 때문에 이러한 자치단체의 등록부에 필적하는 중앙 등록부는 없다. 이러한 면에서 이 시스템은 세계에서 유일한 것이다.

네덜란드의 각 지방자치단체는 그 지방자치단체의 모든 주민들에 관한 정보를 가지고 있는 인구 등록부를 소유하고 있다. 이러한 정보는 개인명부(personal list; PL)에 개별 거주자별로 기록되어 있다. 이 등록시스템 내에서 각 거주자들은 유일한 개인고유번호(personal identification number; PIN)가 주어지고, 이로 인해 지방자치단체는 개인의 자료를 배우자, 부모, 자녀들의 자료와 연결할 수 있다. 이러한 이유 때문에 각 개인명부에 각 주민의 개인고유번호뿐만 아니라 부모, 배우자 및 자녀의 개인번호도 저장되어 있다.

나) 통합직업파일

사회통계데이터베이스에는 ‘고용인 통합직업파일(integrated jobs file of employees)’ 및 ‘고용주와 고용인 없는 자영업자 통합직업파일(integrated jobs file of employers and self-employed persons without personnel)’이 있다.

(1) 고용인 통합 직업 파일(integrated jobs file of employees)

고용인 통합 직업 파일은 다음의 자료원들을 사용하여 마이크로-통합과정을 거쳐 생성된다.

(가) 고용인을 위한 고용인보험 계획 등록 시스템이라 불리는 직업등록부(2000년 말 현재 6백5십만 레코드).

(나) 고용 및 소득 조사(Survey on Employment and Earnings; SEE)

SEE는 대규모 기업 조사로서 임금청들(payroll administrations)로부터 전자데이터교환(Electronic data interchange)에 의해 자료가 얻어진다. 이 조사에서는 소득 및 근무시간뿐만 아니라 직업에 관한 특성치에 관한 정보가 얻어진다. 대부분의 대기업 자료는 등록부에 기초를 두고 구할 수 있지만 소규모 기업에 대해서는 표본조사를 하기 때문에 SEE의 표본설계는 복잡하다. 직업등록부에는 센서스 프로그램에 필요한 두 개의 변수 즉 ‘통상 일하는 시간’ 및 ‘근무처’가 없기 때문에 SEE는 필요하다(2000년 말 현재 3백만 레코드).

(다) FIBASE 등록부

FIBASE 등록부는 회계행정(fiscal administration)으로서 노동 및 사회 보장 수입에 대한 자료가 있으며 세금부과자료로 사용된다. FIBASE 등록부는 또한 직업에 관한 누락 정보를 완성하는데 사용된다(2000년 말 현재 720만 레코드).

(2) 자영업자에 관한 통합직업 파일

고용주 및 고용인 없는 자영업자(이후로 자영업자로 칭함)에 관한 정보는 “자영업자에 관한 통합직업 파일”에 저장된다. 이 정보는 “자영업자의 소득(profits)에 관한 최종소득세 평가 등록부(FITAP; final income tax assessments on profits)”에서 얻어진다. 불행히도 이 등록부에는 정확한 소득 기간에 관한 정보가 없다. 따라서 2000년도에 등록한 사람들은 센서스 기준 시점인 2001년 1월 1일 현재에 고용주 또는 고용인 없는 자영업자로 간주한다. 그러나 이러한 가정이 고용주나 고용인 없는 자영업자의 수에 대한 과대 추정을 배제할 수 없다. 2001년 센서스 결과표 프로그램을 편집하는 동안 약 4만 명(5%)의 자영업자에 관한 정보가 누락되었다. 이 사람들에게 대한 세금이 확정되지 않았는데 아마도 징세 당국과의 분쟁 때문이리라. 이러한 자영업자는 2001년 센서스 결과표에 포함되지 않았다(2000년 말 현재 7십9만 레코드).

다) 통합수혜파일

조기 은퇴한 사람들은 “통합수혜 파일”에 있는 생명보험이나 연금자료를 통해 추적한다. 이러한 종류의 정보는 FIBASE 등록부에서 마이크로-통합과정으로 구한다(2000년 말 현재 2백7십만 레코드).

라) 노동력 조사(Labor Force Survey; LFS)

노동력 조사는 가구표본조사로서 등록부에서 구할 수 없는 센서스 정보를 얻는 데 필요하다. 이 조사는 직업이나 교육정도 등의 센서스 변수와 관련이 있다. 노동력 조사는 또한 실업자인 경제활동인구를 정의하거나, 교육기관의 정규학생(full-time attendant) 또는 가족을 돌보는 것이 주활동인 비경제활동인구를 정의하는데 사용된다. 노동력조사는 일반가구를 대상으로 하는 표본조사로서 15세 이상을 대상으로 한다. 이 조사는 표집과 조사가 1년에 걸쳐 분산되는 연속 조사이다. 표본의 크기는 상대적으로 작아 약 10만명이 표본으로 선정되며 15세 이상 인구의 약 1%이다. 결과적으로 센서스 결과표 작성에 필요한 세분화된 단계에서 소규모 집단에 대한 추정치의 신뢰도에 문제가 있거나 추정 자체가 불가능한 현상을 야기한다. 이러한 이유 때문에 표본수를 늘리기 위해 2000년과 2001년의 노동력 조사를 합쳤다. 실제로 센서스 기준 시점 1년 전의 자료와 기준 시점 1년 후의 자료가 이런 식으로 수집되었다. 위에 언급한 노동력 조사의 센서스 변수들은 1년 동안 상대적으로 안정적이라고 가정하였으며, 따라서 오차의 크기가 크지 않으면서 조사 기준일의 상황을 대표하는 것으로 간주할 수 있다. 실제로 직업 및 실업 변수는 가정한 것 보다 더 큰 변화를 일으킬 수도 있다(2000-2001년 LFS 2십3만 레코드).

4) 추정방법

2001년 네덜란드 센서스의 결과표는 등록 자료와 기존의 표본조사 자료를 이용하여 작성되었으며, 수치적으로 일관성 있는 표의 작성을 위해 반복가중기법(repeated weighing)을 사용하였다. 네덜란드 통계청은 반복가중기법의 사용을 위해 VRD(Filling['Vullen' in Dutch] Reference Database)라는 통계팩키지를 개발하였다.

2001년 네덜란드 가상 센서스에서는 40개의 결과표를 생성했다. 이 중 28개는 전국단위의 결과표이고, 9개는 광역시 단위이며, 나머지 3개는 기초 자치단체 단위에서 작성되었다. 40개의 표를 분류별로 보면, 8개의 주택에 관한 표, 2개의 통근에 관련된 표, 나머지 30개의 표는 인구에 관한 것으로 직업, 교육수준 및 경제활동과 관련된 것이다.

등록부에 기초를 둔 표는 등록 자료를 합침으로써 계산된다. 표본조사에 기초를 둔 표의 작성을 위해서는 표본조사 자료가 이용되기 때문에 추정이 필요하다. 등록 자료와 표본조사 자료를 통합한 자료를 생성하기 위해 반복가중기법을 사용하였다.

5) 가상센서스와 전통적 센서스의 비교

가상 센서스는 비용이 적게 들며 과거의 센서스와 비교 가능하고 사회적 수용성이 더 크다. 전통적인 방식으로 센서스를 실시하였다면 약 3억 유로의 비용이 들었겠지만 가상 센서스의 비용은 3백만 유로에 그쳤다.

센서스 비용의 절감은 충분한 등록 정보를 보유하고 있는 나라에서만 가능하다. 전통적인 방식으로 실시한 2001년 캐나다의 센서스 비용은 약 4억5천만 유로가 들었고, 인구는 약 3천백6십만으로 네덜란드 인구의 약 2배이다. 캐나다에서는 충분한 등록 자료가 없기 때문에 가상 센서스의 실시는 불가능하다.

비용 측면 외에도 전통적인 센서스와 가상 센서스의 차이점이 존재한다. 전통적인 센서스가 의무적인 것이라고 할지라도 무응답(단위 무응답 및 항목 무응답)의 문제가 발생한다. 무응답을 가중치 부여(weighting) 및 대치법(imputation)을 통해 해결하는 것은 가치가 있는 일이지만, 잘 알려진 전통적인 센서스에서의 문제점은 센서스에 대한 참여가 제한적(limited)이고 선택적(selective)이라는 사실이다. 따라서 전통적인 무응답 보정 방법의 사용으로는 신뢰성 있는 결과를 공표하기에는 부족하다. 지난 1971년의 네덜란드 센서스에서는 국민들의 개인정보 보호의식으로 인해 센서스 자료 수집에 대한 반대에 직면했고, 다음 센서스가 실시된다면 무응답률은 점점 더 커질 것으로 예측되었다. 가상 센서스의 실시에 대한 반대는 거의 없으며 무응답의 문제는 센서스에 사

용되는 표본조사에 대해서만 일어난다. 표본조사에서 무응답 문제가 해결된다면 센서스에서 선택성(selectivity)의 문제가 해결된다는 점은 분명하다.

2001년 네덜란드의 가상 센서스는 전통적인 센서스를 실시한 다른 나라들보다 늦게 시작했다. 일부 등록부들은 비교적 나중에 이용 가능했음에도 불구하고 2000 라운드 센서스에 참여한 대부분의 나라들보다 더 일찍 40개의 결과표를 완성할 수 있었다. 사실 네덜란드는 완성된 40개의 결과표를 유럽통계청(Eurostat)에 가장 먼저 제출한 나라에 속한다. 네덜란드는 센서스 질문지를 검토하고 수정할 필요는 없지만 일부 항목에 대해서는 표본조사 자료를 이용해야했기 때문에 네덜란드 결과표에 요구되는 상세한 자료를 생산하는 것은 불가능했다. 다음의 센서스에서 충분히 추정할 수 없는 셀 값을 추정하기 위해서 소지역 추정법을 적용할 계획이다.

5. 해외사례 비교

센서스 응답률의 감소, 응답자의 센서스에 대한 부담, 센서스 비용 및 센서스를 실시하지 않는 센서스 중간연도에 대한 정확한 정보의 수요 등으로 인해 전통적인 센서스의 매력은 점점 줄어들고 있는 것이 세계적인 추세이다. 이 절에서는 전통적인 센서스를 포기하고 대안의 센서스 방법을 채택하고 있는 미국, 프랑스, 네덜란드의 센서스 방법들을 비교하고자 한다.

미국과 프랑스에서의 표본조사는 키시의 순환센서스의 개념을 사용하기 때문에 유사한 점이 많다. 네덜란드의 경우는 미국과 프랑스의 경우와 근본적으로 다르다. 네덜란드에서는 국민들의 사생활보호의식의 확산 및 이로 인해 예상되는 무응답 증가 때문에 전통적인 센서스를 포기하고 행정등록 자료와 기존의 표본조사 자료를 이용하여 통계적 기법만을 사용하여 가상으로 센서스를 실시한다.

순환센서스의 아이디어는 키시가 처음 주장하였으며 프랑스의 순환 센서는 키시의 아이디어를 그대로 따르고 있다. 미국의 지역사회조사도 키시의 아이디어 중 많은 부분을 따르고 있다. 두 나라 센서스의 가장

큰 차이점은 프랑스가 전통적인 센서스를 완전히 포기하고 순환센서스로 전환했던 반면에 미국은 헌법의 규정 때문에 10년 주기의 전통적 센서스를 여전히 실시해야 한다는 부분이다.

이러한 차이점 때문에 프랑스의 순환센서스에서는 조사 내용 보다는 인구수의 파악에 더 중점을 두고 있으며 미국의 지역사회조사에서는 인구수보다는 조사내용에 더 중점을 두고 있다. ACS는 10년 주기의 센서스표본조사를 대체하기 위해 고안되었다. 10년 주기의 전수조사는 여전히 실시된다.

방법적인 측면에서 보면 프랑스는 통계목적이 아닌 행정목적으로 작성/유지되는 주소록에 기반을 둔 반면에, ACS에서는 표본틀의 구축 및 주소록의 유지에 많은 노력을 기울인다. 또한 프랑스에서는 양질의 기초자치단체의 행정자료를 이용할 수 있지만, 미국의 경우에는 10년 주기의 전수조사 결과와 출생, 사망 등의 행정자료를 이용하여 추정된 추계인구에 맞춰 보정한다. 그러나 이러한 행정자료는 지역 단위의 총량치로서 프랑스의 경우처럼 소지역으로는 작성이 되지 않는다.

이 두 나라 모두 대규모 지역의 자료를 매년 작성하여 공표하지만, 프랑스의 경우에는 인구수에 초점을 두어 추정치의 계산 방식이 매우 간단하다. 그러나 미국의 추정방법은 매우 복잡하다.

소지역에 대한 개념은 두 나라 간에 약간 차이가 난다. 프랑스의 경우에는 인구 만 명 미만의 지역을, 미국의 경우에는 인구 이만 명 미만의 지역을 소지역으로 구분한다. 이러한 소지역에 대해 프랑스에서는 1년에 한번 조사하고 모델을 이용하여 센서스 중간연도와 센서스 이후연도에 대해 추정을 한다. 미국에서는 이러한 소지역을 매년 조사하고 연간 자료를 합쳐서 순환평균을 이용한 추정을 한다. 반면에 네덜란드 경우는 기존의 표본조사 자료를 이용하여 소지역에 대한 추정치를 생산하며, 기존의 표본조사의 표본수가 충분하지 못하기 때문에 소지역에 대한 자료가 빈약하다.

센서스 비용 측면에서 살펴보면 미국과 프랑스에서의 센서스 비용은 10년 주기의 전통적인 센서스 비용과 거의 비슷하다. 네덜란드의 경우에는 실제로 조사하는 부분이 없어 전통적인 센서스 비용의 1%만이 소요된다.

제3절 우리나라의 센서스 표본조사 개선 방향

인구센서스 표본조사 개선을 위한 해외사례 연구를 통한 벤치마킹 대상의 선정은 현 시점에서 우리나라의 객관적인 상황인식을 토대로 할 때 비로소 의미가 있다 할 것이다. 우리나라는 최근까지도 전통적 센서스를 실시하여 오면서 커다란 문제를 겪지 않았다고 볼 수 있으며 조사 환경 측면에서도 일부 선진국에서 발생되었던 집단적인 거부 움직임도 없었다고 볼 수 있다. 다만, 향후에 동일한 방식으로 센서스가 실시된다고 가정할 경우, 이에 소요될 조사비용의 압박은 우리나라가 캐나다와 일본 등과 같이 전통적인 센서스 방식을 고수해 가는 데에 있어 현실적인 장애요인이 되고 있음이 문제일 것이다.

이러한 현실적 상황 인식 하에서 볼 때, 과연 비용의 축소라는 문제에 관해 표본조사의 개편이 얼마나 영향을 끼칠 수 있는가?에 대한 고민의 해답은 자명해진다. 즉, 비용만을 고려한다면 센서스 전수조사 방식의 개편에 초점이 맞추어져야 할 것이지 굳이 표본조사의 개편을 논의할 필요가 없다는 것이다. 오히려 표본조사의 개편만을 놓고 본다면 자료제공의 시의성과 소지역 자료 확보를 위한 표본규모의 증가로 인해 오히려 조사비용은 증가하게 될 것이다. 이는 곧 센서스 표본조사방식의 개편이 우리나라의 센서스 실시환경에서 가장 중요한 부분이라 할 수 있는 비용의 문제를 해결할 수 있는 대안이 아닌 통계자료 제공의 시의성 제고 등 통계서비스의 질적인 향상을 위해 추진되어야 할 사항이라는 것이다. 이와 같은 표본조사 개편의 의미를 염두에 두면서 기존 시스템의 개선을 위한 도입에 적합한 해외 모델을 앞서 살펴본 해외 사례를 대상으로 선정하고 선택된 모델의 도입방향에 대해 고찰해 보고자 한다.

적합한 인구센서스 표본조사 모델을 선정하기 위한 해외 사례의 평가에 있어서 미국, 프랑스, 네덜란드의 인구센서스 시스템을 대상으로 하여 이들 나라에서 시행되고 있는 시스템의 신뢰성에 관한 측면과 새로운 시스템의 적용에 따른 실시 용이성 측면을 중심으로 살펴보았다.

- ① 시스템의 신뢰성: 네덜란드의 가상센서스 시스템과 프랑스 및 미국의 순환표본 시스템은 크게 센서스를 위한 별도의 표본조

사의 실시 여부에 의해 구분될 수 있으며, 프랑스와 미국은 전수조사의 실시방식에 따라 구분된다. 우선, 가상센서의 경우는 1971년의 센서스를 마지막으로 조사환경의 악화로 인해 부득이 센서스 실시를 포기하면서 별도의 조사 없이 센서스관련 결과표를 산출해야만 했던 상황에서 만들어진 체계로서 비록 1981년과 1991년을 거쳐 2001년까지 운영되어 오면서 나름대로 기틀을 갖췄다고는 하나 현재까지도 세부자료의 불충분성을 내포하고 있으므로 결과의 신뢰성에 대한 문제를 지적하지 않을 수 없다.

한편, 순환센서스를 세계에서 유일하게 실시하고 있는 프랑스의 경우도 통계조직 내부의 필요성이 아닌 외부적인 요인, 특히 예산의 압박으로 인한 조사비용 분산의 수단으로서 순환센서스를 기획하게 되었으며 2004년 시작 이후 2008년에 비로소 5년 순환주기의 첫 번째 5년 추정치 산출작업을 수행하고 있는 상황이다. 이들 나라에서는 제각기 새로운 시스템 시행이 성공적이었다고 자평하고 있음에도 불구하고 이러한 제도들이 아직은 시행 초기단계라는 점과 객관적인 평가가 이루어지지 않은 상황이라는 점에서 이들의 성공 정도를 객관적으로 판단하기 어렵다고 볼 수 있다. 따라서 이들 시스템의 신뢰성에 대한 평가는 태동배경에 대한 의구심 쪽에 더욱 무게가 실릴 수밖에 없게 되며, 장기적인 관점에서도 향후 이러한 시스템이 안정적으로 시계열 유지가 가능할지에 대한 판단도 어려운 상태이다.

이에 반해 미국의 경우는 새로운 시스템에 대한 통계작성기관의 자발적 수요(시의성 및 소지역 통계 제공 등)가 제기되어 각종 준비사항 및 적정한 전환 포인트 설정 등 시스템 도입배경에 대한 긍정적 평가가 가능하며 기본적인 인구 및 가구 지표를 생산할 수 있는 전수조사를 그대로 유지하므로 자료의 시계열적 연속성을 담보하고 있다는 점에서도 좋은 평가를 내릴 수 있다. 또한, 프랑스 순환센서스의 단점으로 지적되고 있는 횡단면 분석 가능성의 축소 문제와 관련해서도 전수조사가 유지되고 있어 기본적 인구·주택 지표에 대한 횡단면 분석자료 제공이 가

능하므로 수요자의 요구와도 부합할 수 있을 것으로 판단된다. 적어도 비용 등 외부적인 압박 요인에 의한 불가피한 선택이 아닌 통계기관 내부의 자발적인 의지를 통해 개발되어 왔다는 점은 시스템의 개념 설계, 각종 테스트 및 자료 제공 등 통계조사 전반적 개발과정에 걸쳐 신뢰성에 긍정적인 평가가 가능하다.

- ② 실시 용이성: 기존의 체계를 변화시키는 새로운 제도의 도입은 항상 혼란을 야기할 수밖에 없는데 특히 변화의 규모가 크고, 새로운 변화에 대한 준비기간이 충분하지 않을 경우, 이러한 혼란은 더욱 가중될 것이다. 인구센서스에 대한 새로운 변화에 있어서도 일시에 대규모의 급격한 변화보다는 큰 틀에서 기존의 체계를 유지하면서 개선의 범위를 단계적으로 확장해 나가는 것이 바람직할 것이다. 전수조사와 표본조사 모두에서 대규모 변화가 수반되는 프랑스와 네덜란드식의 개선은 센서스에 관계하는 제 주체들의 혼란으로 인해 새로운 제도의 좋은 취지와 상관없이 부정적 결과를 초래할 수 있으므로 전통적 센서스를 유지해 오던 우리나라가 변화를 수용하여 새로운 제도를 성공적으로 정착시키기 위해서는 센서스의 근간이라 할 수 있는 전수조사를 유지하면서 표본조사만을 순환 조사하는 미국의 형태가 받아들이기 용이하다는 것이다.

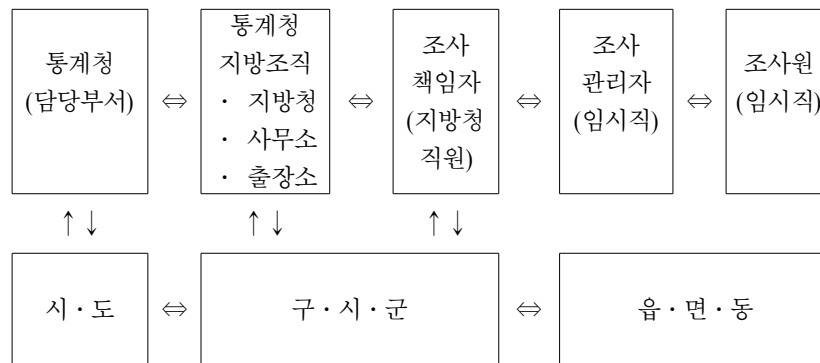
위와 같은 맥락에서 현재 우리나라의 표본조사방법 개선의 기본 모델을 미국의 ‘순환표본조사(Rotation sampling survey)방식으로 설정하고 앞서 살펴본 미국의 실시사례를 토대로 하여 우리나라 특성에 맞는 실용적인 인구센서스 표본조사의 개선 방향을 모색해 보고자 한다.

1. 조사 체계

가. 기본 체계

기본적으로 조사체계의 근간은 전통적인 센서스에서와 같이 ‘통계청

지자체-조사원'으로 이루어져야 함에는 틀림이 없으나 조사주기의 단축으로 인한 지자체 통계조직의 확충 유도에 관한 문제와 통계청 지방조직을 통한 센서스 조사원 현장관리시스템의 구축 및 1년 마다 동원되어야 할 수 천명의 현장조사인력에 대한 유지·관리 문제가 주요 과제로 남게 된다. 여기에서는 각 조사주체별 역할을 중심으로 전통적인 센서스에서 순환표본방식으로의 이전에 필요한 개선 사항을 조사를 담당하는 주체별로 살펴보기로 한다.



[그림 1] 조사체계

1) 통계청

가) 통계청 본청의 역할

인구센서스 전수조사방식의 변경과 관련된 문제는 논외로 하고 인구센서스 표본조사방법의 변화 부분만을 고려할 경우, 통계청 본청, 즉 담당부서인 '인구조사과'의 적절한 역할 재정립이 필연적으로 수반되는바, 인력투입 측면에서는 '전문화', '경량화'라는 특성과 자료제공 측면에서는 '신속화'라는 특성으로 대표되게 된다.

현재 실시되고 있는 '인구주택총조사'는 5년마다 한번 씩 실시되는 통계청의 큰 대사로서 전청 차원에서 실사지도 및 교육 인력의 지원이 수반되었으나 '순환표본조사'방식 경우 조사주기가 1년이고 전수조사를 배제하므로 현재와 같이 일시에 대규모 인력 지원은 필요가 없는 '경량화'와 표본 틀 유지관리 및 추정 등 보다 전문적인 업무를 수행할 수 있

도록 본부 인력의 ‘전문화’가 필요하게 된다.

또한 ‘신속화’라는 의미는 5년주기 조사에서 연간단위 성격으로 인구센서스가 변경되면서 자료 공표까지 일련의 업무 프로세스가 경상 작업화됨에 따라 현재 조사체계에서의 조사기획 분야의 중요성보다는 조사결과를 1년 단위로 신속하게 공표하여야 하는 부분과 관련되어 있는데, 이렇게 신속한 결과 공표를 위해서는 자료처리와 관련된 인력의 보강과 함께 프로그램 작성 등 전산처리 전반을 지원해줄 수 있는 외부의 전산전문 용역업체의 동원도 필요할 것이다.

나) 통계청 지방조직의 역할

프랑스는 전수조사를 망라하는 전체 인구센서스로서의 순환총조사를 지자체 인력 중심으로 수행하고 있고, 미국은 3개월 단위로 우편, 전화, 면접조사 등 연중 계속 조사하는 시스템으로 ACS 전담인력을 상시 고용하여 업무를 수행하고 있다. 미국의 경우 조사를 전담하는 인력을 상시 고용하는 것은 조사의 전문성 확보와 안정된 업무추진을 위해 바람직하다고 볼 수도 있겠으나, 우리나라의 현실적인 행정여건과 예산사정을 감안할 때 1년 중 어느 한 시점을 기준으로 일시 조사가 불가피하다. 또한 연중 계속조사를 할 경우에는 동일한 항목이 조사시점이 다른 각 지역별 조사결과가 상이할 경우 지역별 차이인지, 시점별 차이인지 불분명하게 되는 문제가 있을 수 있으며 이전의 전통적 센서스에서의 표본조사 장점인 횡단면분석의 가능성을 더욱 훼손시킬 수 있다는 단점이 있을 수 있다.

이러한 이유로 조사원은 임시적으로 고용하는 형태를 상정하고 통계청 지방조직의 역할을 고려한다면 첫째, 1년 주기 경상조사 개념의 실사관리자로서의 역할을 들 수 있으며, 두 번째로는 임시조사원의 전문성 유지를 위한 임시조사원 인력채용 DB 운영자로서의 역할을 꼽을 수 있겠다.

첫 번째 지방조직의 실사관리자(supervisor)로서의 역할의 경우, 기존의 센서스에서도 지방조직의 인력동원이 없었던 것은 아니나 여기서 언급하고자 하는 역할은 조사원의 교육, 실무지도 및 동별 임시관리자에 대한 총괄 관리 등의 업무를 망라하는 지역별 총괄관리자로서의 역할을

말한다. 상대적으로 소규모 표본을 1년 단위로 측정하는 시스템으로 바뀌게 되어 기존의 전통적 센서스에서 누려왔던 전폭적 인력지원이 불가능해지는데도 불구하고 표본가구의 분포는 여전히 소지역 추정을 위해 전국에 고루 분포하게 되므로 인구조사를 담당하는 부서의 인원만으로 충당하기에는 현실적으로 불가능하기 때문이다.

두 번째 임시조사원 인력채용 DB의 운영자로서의 역할은 임시조사원 동원에 따른 조사원의 전문성 결여 문제를 해결하고자 조사경험자를 재채용할 수 있도록 하는 임시조사원 DB를 구축, 유지관리하고 조사원 채용까지도 담당하는 운영자로서의 역할을 말한다. 미국의 경우 조사원 상시 고용제도를 통해 조사의 전문성을 확보하여 비표본 오차를 최대한 줄임으로써 불충분한 표본의 규모를 보완하고 있다. 정규고용 방식이 아닌 우리나라의 시스템 하에서 조사전문성을 확보하기 위한 최소한의 대안으로서 임시조사원 인력채용 DB의 운영과 채용 담당자로서의 역할은 반드시 필요하다.

2) 지방자치단체

기존의 전통적인 센서스 체제 하에서도 지자체의 역할은 중요했었고 순환표본조사 체제로 전환할 경우에도 여전히 지방자치단체의 참여는 절실하므로 지자체와의 원활한 협조체제 유지는 필수적이다. 원활한 협조관계 유지를 위해선 조사결과 집계 시 지자체가 필요로 하는 소지역 자료 생산 및 공표주기 단축을 통한 지자체의 센서스 조사 결과 활용도 제고 등으로 자발적인 참여를 유도하는 것이 가장 바람직 할 것이다. 프랑스는 순환센서스를 실시하면서 최초 비협조적이었던 지자체들이 정책에 필요한 소지역통계 생산으로 인해 협조적으로 바뀐 바 있는 좋은 예라 할 수 있다. 이렇게 관심유도가 성공적으로 이행될 경우 부수적으로 지자체내 통계조직의 활성화 및 확충도 기대할 수 있겠다.

지자체의 역할을 단계별로 살펴보면 읍·면·동의 경우 임시조사원을 수용할 수 있는 공간 제공과 사무기기 등 시설물 제공이 우선적으로 고려되어야 한다. 순환표본조사이므로 기존 센서스와 달리 1년에 한번 씩은 읍·면·동을 중심으로 조사가 이루어져야 하기 때문에 그러하다.

읍·면·동 기능전환에 따라 비록 조사의 진행에 직접적으로 참여하는 것까진 어렵더라도 사무실 및 사무기기 제공 등 부수적인 지원을 이끌어내는 것은 예산의 효율적 절감을 위해서도 중요하다.

구·시·군의 경우 현실적으로 담당자 지정이 가능한 지자체 최하위 기관이라 할 수 있다. 따라서 조사에 직·간접적으로 중요 역할을 담당하게 하여야 하는데 특히 지역내 홍보활동 분야에서 더욱 그러하다. 조사를 진행할 때 아파트와 같은 공동주택을 대상으로 안내방송 요청, 홍보물 설치 등 간접적 지원도 중요하며 조사 진행시 발생할 수 있는 응답자와의 마찰 해결에서도 효과적인 업무수행이 가능하므로 일선 조사관리체계에 있어 통계청 지방조직과 함께 중요한 부분을 담당하게 된다.

마지막으로 광역단위 지자체인 시·도의 경우는 조사를 진행함에 있어 실질적인 관여의 범위가 작을 수 있지만 산하 지자체와의 원활한 협조관계 도출을 위한 역할이 중요하다 하겠다. 통계청 본청 직원이 조사기간내 상주하며 시·도 조사본부의 역할을 담당해야 할 단계에 위치하므로 시·도 전체의 조사진행 상황 점검과 조사시 발생하는 문제사항 해결 등에 있어 통계청 본청과 직접적인 연락창구 역할도 담당해야 하는 단계이다.

3) 조 사 원

‘미국지역사회조사(ACS)’의 경우 연중 계속조사형태로서 전담인력을 상시 고용할 수 있는 체제로서 임시직원에 의존했던 10년 주기의 미국 센서스보다 높은 품질의 조사결과 도출이 가능하여 표본규모를 더 줄일 수 있는 밑거름이 되었다고 한다. 비단 이러한 사례를 들지 않더라도 조사원의 전문성은 조사의 밑바탕을 이루는 중요한 부분임은 명백하다. 우리나라의 형편상 연중 계속조사가 아닌 1년 1회 조사를 실시해야 하기 때문에 현실적으로 임시조사원을 활용할 수밖에 없는 상황에서 조사원의 전문성 확보를 통한 조사오류 축소는 표본규모의 합리적 결정을 위해서도 비중 있게 고려되어야 한다. 따라서 ‘인구주택총조사’ 및 각종 통계조사의 유경험자를 확보하여 관리하는 것이 중요한데, 이는 기존의 센서스에서는 90% 전수조사, 10% 표본조사라는 체제 하에서 비교적 단

순한 조사표를 통한 전수조사에 90%의 조사원이 동원되었지만 순환표본조사의 경우는 100% 표본조사표에 의한 조사이기 때문에 더욱 그러하다.

장기적으로는 조사원 채용과 관련하여 외주업체를 활용하는 방안도 고려해 볼만한데 이는 경상적으로 실시되는 가구대상의 조사들을 통합하여 우수한 조사인력을 보유하고 있는 업체에 외주를 주는 방법이다. 현실적으로 실행되기 위해서는 행정적 절차와 관련 규정의 검토가 선행되어야 하겠지만 1년 단위의 경상조사 성격으로 변화되는 센서스 표본조사에 있어서는 양질의 조사원 확보를 위한 업무를 대행해 줄 수 있는 수단으로 고려해 볼 수 있다.

2. 조사 방법

조사방법은 기존의 센서스 표본조사 방식과 순환표본조사 방식과의 차이가 있을 수 없으므로 면접조사를 비롯하여 우편조사, 인터넷조사 등 효율성과 필요성에 따라 맞는 방식을 선택할 수 있을 것이다. 다만, 기존의 표본조사는 조사구를 단위로 표본이 추출되어 정해진 조사구 전체를 조사하면 되었기 때문에 조사원이 표본대상 가구를 방문을 위한 대상가구 위치 파악 문제나 다음 대상가구로의 이동시간 등에 대해서 조사원 업무수행에 큰 지장이 없었다. 하지만 순환표본 방식에 있어 거처(주택)단위로 표본이 선정될 경우에는 대상거처를 찾는 문제와 표본이 보다 넓게 분포하게 됨에 따라 이동시간도 고려되어야 할 것이다. 이러한 문제의 해결을 위해서는 준비조사기간동안의 대상거처 위치 파악 및 적절한 동선 확보가 이전보다 중요하게 다루어져야 할 것이다. 따라서 조사지역 안내를 목적으로 하는 조사용 지도의 정밀성 제고에 대한 고려와 함께 상가, 관공서, 주요 건물 등 랜드마크로 활용될 수 있는 지형지물에 대한 정보를 최근기준 자료로 유지하는 것 등에도 좀더 세심한 주의가 필요하다.

한편, 조사를 실시함에 있어 현실적으로 부딪히게 될 가장 큰 문제는 5년마다 대규모로 실시되던 기존의 총조사에서와 같은 전국적 대규모 홍보가 상대적으로 소규모화된 순환표본조사에서는 기대할 수 없게 됨

으로 해서 조사원들의 원활한 조사 진행에 차질이 생길 것이 우려되는데 이에 대한 적절한 홍보대책도 세워야 할 것이다.

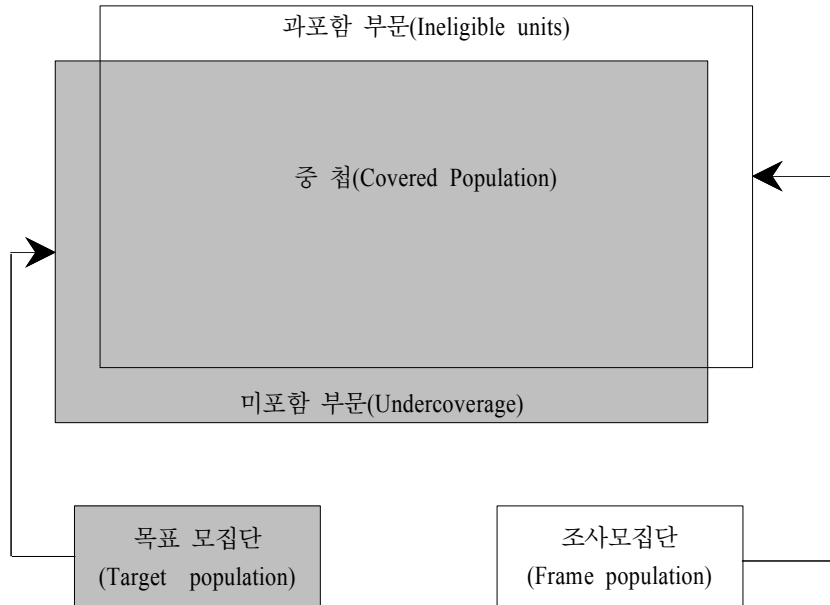
또한, 실제 조사수행시 표본가구의 부재·불응 등에 대한 대책이 필요한 바, 표본의 대체나 통계적 무응답 처리기법(Imputation) 적용을 고려해 볼 수 있는데, 통계적인 무응답 처리의 경우는 반드시 무응답 처리의 정확도를 제고하기 위하여 필요한 성별, 가구원 수, 주택의 종류 등 기본변수를 조사하여야 한다. 표본을 대체하는 방법의 경우, 순환되는 표본의 중복성 배제를 위해 향후 순환하여 대상이 될 표본과 대체로 이용된 표본에 대한 철저한 이력관리 등이 필요할 것이다.

3. 표본추출 틀(Sampling frame)

가. 표본추출 틀 결정 기준

특성을 알고자 하는 목표 모집단(Target population)의 개별 요소를 규명하고자 조사를 기획할 때 필요한 목록을 표본추출 틀로 규정할 수 있으며 표본조사의 근간이 되는 부분이라 할 수 있다. 표본의 중복성을 배제하고 1년 단위의 계속적인 조사를 통해 조사결과를 누적해야 하는 순환표본 방식의 표본조사에 있어서는 더욱 정교한 프레임이 요구된다 하겠다.

미국과 프랑스 모두 순환표본 추출을 위한 프레임으로 주소 파일과 지리정보 시스템(GIS)을 연결을 통해 데이터베이스를 구축하여 이용하고 있으며 매년 새로운 정보를 갱신해 오고 있다. 이러한 신규 데이터 갱신에는 비단 통계청뿐만 아니라 지자체까지를 망라하는 대규모 작업이 수반되고 있다. 이러한 GIS기반 표본추출 체제에서 결국 표본추출의 단위가 주소, 즉 거처단위가 되는 것이다. 이들 국가의 경우는 한국과 같은 주민등록제도가 없기 때문에 전국을 대상으로 하는 조사의 기본틀을 이러한 시스템으로 운영할 수밖에 없었을 것이다.



- * 목표 모집단(인구주택총조사 모집단): 일정 시점 현재 대한민국내 거주하는 모든 사람과 이들이 살고 있는 주택
- * 조사 모집단(주민등록명부): 시장·군수 또는 구청장에게 30일이상 거주할 목적으로 그 관할구역 안에 등록한 주소 또는 거소("거주지")를 가진 자("주민")

[그림 2] 표본추출틀로서의 주민등록자료와 인구주택총조사 모집단과의 관계

한국의 주민등록제도는 이러한 외국의 사례에서보다 우월한 표본추출 프레임 구축을 위한 여건을 제공하게 되는데, 몇 가지 보완이 필요한 문제들을 포함하여 표본추출 프레임으로서의 품질을 평가한 결과는 아래와 같다.

- ① 표본추출 틀로서의 정확성: 총조사가 목표로 하고 있는 모집단과 비교하여 과포함과 미포함 정도를 판단해 보는 항목으로서 서로 중첩되는 정보가 많고 과포함 및 미포함된 정보가 적어야 적절한 표본추출 틀이라 할 수 있다. 이러한 부분에 대한 정확한 측정은 방대하고 복잡한 작업으로서 단지, 2005년 총조사

결과와 주민등록자료간의 인구 및 가구(세대)를 비교해 봄으로써 정확성을 판단해 본다면 우선, 인구의 경우, 외국인이 포함되어 있지 않음에도 불구하고 주민등록자료가 3.2%정도로 높게 나오고 있는데 이를 모두 과포함으로 생각하기에는 곤란한 측면이 있다. 이러한 이유는 사망자에 대한 신고지체 등으로 인한 주민등록의 과포함 사유도 있다 할 수 있겠으나, '2005 인구주택총조사'의 사후조사결과에서 가구원 조사 누락률이 2.4%였다는 점을 감안한다면 센서스에서의 조사누락에 기인한다고도 볼 수 있기 때문이다. 또한, 가구(세대)의 경우도 11.7%의 차이를 보이며 주민등록자료가 높게 나타나는데 이는 인구와는 달리 주로 과포함의 범주에 속한다고 볼 수 있다. 이는 현실적으로 학군 배정, 주택 청약 등의 사유로 인해 행해지고 있는 허위 세대분리 문제가 실생활에서 공공연히 발생되고 있기 때문이다.

한편, 미포함 영역은 대표적으로 외국인과 주민등록 미등록자의 예를 들어 평가할 수 있을 것이다. 결혼 등을 통해 국적을 취득한 외국 출생자의 경우는 문제가 없지만 순수 외국인에 대해서는 주민등록상으로 파악할 수 없는 부분이며 일정한 거주지가 없어 주민등록을 할 수 없는 경우와 같이 틀 자체에서 근본적으로 배제되어 있는 부분은 미포함의 영역에 속한다 하겠다.

- ② 표본추출 틀로서의 타당성: 실질적인 표본추출의 틀이라 할 수 있는 주민등록 '세대명부'를 살펴보면 주소를 비롯하여 세대주의 성별 및 나이, 세대원 수 등 표본설계(sampling design)에 필요한 기본정보를 손쉽게 획득할 수 있어 표본추출 틀로서의 타당성이 입증된다. 목표로 하는 모집단을 대표할 수 있는가에 대한 타당성에서도 외국인 부문, 불성실 등록자, 가구와 세대간의 개념의 차이, 자료시점 등으로 인한 두 자료간의 근본적인 차이는 존재한다 하더라도 <표 7>에서와 같이 '05년말 기준 주민등록 인구(48,782천명)와 세대(17,857천 세대)가 '05년 11월 1

일기준 총조사 인구(47,279천명)와 가구(15,988천 가구)보다 각각 3.2%, 11.7% 더 높게 집계되고 있는 상황으로 볼 때, 총조사에서 추구하고 있는 목표 모집단을 충분히 커버하고 있다고 볼 수 있다.

〈표 7〉 '05년 기준 주민등록자료와 인구주택총조사 결과 비교

	주민등록자료 (A)	인구주택총조사 (B)	차 이 (A-B)	비 율 (A/B, %)
인 구	48,782천명	47,279천명	1,503,323	3.2
가구(세대)	17,858천세대	15,988천가구	1,869,237	11.7

* 자료출처: 행정안전부('05.12.31.), 통계청('05.11.1.)

[별지 제3호서식]						
세 대 명 부						
행정기관 출력범위				출력일자: 폐 이 자:		
일련 번호	세대주 성명 생년월일 (남/여)	주 소	변동일자 사 유	세대원 수 가족/동거 (남/여)	세대주(원) 확 인 (서명/인)	통장·이 (반)장확인 (서명/인)

210mm × 297mm[인쇄용지(특급) 70g/m²]

※ 색인부 내용 중 전입사항은 별지 제5호서식의 “주민등록 전입자명부”를 활용하고, 전출사항은 별지 제4호서식의 “주민등록 전출자명부”를 활용합니다. 이 난은 서식에 포함되지 않습니다.

[그림 3] 주민등록 세대명부 양식

- ③ 표본추출 틀로서의 시의성: 추출 틀이 최신의 상태로 유지되는지에 대한 평가로서 조사 이전 표본 추출작업을 진행하기 위해 틀이 되는 자료를 입수한 시점과 실제 조사시점간의 차이는 통상 반드시 필요한 기간이므로 논외로 하고, 순수하게 주민등록자료만의 갱신시차를 고려한다면 주민등록법에 존치하고 있는 신고유예기간 이외에는 갱신에 관한 시차를 발견할 수 있는 부분은 없다. 이러한 점은 기존의 조사구설정을 통한 표본추출 틀 구축작업과 비교해 볼 때, 월등하게 효율성을 평가받을 수 있는 부분으로서 표본 추출을 위한 표본설계 프로그래밍 작업 등이 완비되었다고 할 경우에는 주민등록 입수 시점에서부터 본조사 적용까지 단시간 내에 표본추출 작업이 가능할 수 있게 된다.
- ④ 표본추출 틀로서의 경제성: 주민등록이 표본추출의 틀로서 이용될 수 있는지에 대한 평가 중에서 크게 우수성을 인정받을 수 있는 항목으로서 기존의 전통적 센서스 방식에서의 인적, 물적 투입에 따른 비용측면의 경제성뿐만 아니라 표본 틀 구축에 투입되는 시간을 절감함에 따른 업무효율 측면의 경제성까지 만족시킬 수 있다 하겠다. 따라서 표본 틀로서의 품질에 대한 정확성 평가에서 나오는 ‘과포함’ 및 ‘미포함’ 문제를 충분히 상쇄하고도 남을 만큼의 충분한 경제성을 갖추고 있는 자료로서 평가할 수 있다.

나. 프레임 균등분할 문제

순환표본조사의 경우, 연속적인 측정을 통한 표본의 누적을 전제로 하므로 표본추출 틀에 대한 균등한 분할 문제를 반드시 고려해야 하는바, 매년 조사결과 발표를 위해서는 각 연도별로 대표성 있는 표본이 선정되어야 한다. 또한, 5년간 누적된 표본을 통한 최종적 추정에 있어서도 각 연도별 평균치의 동질성 확보 측면에서도 고려되어야 할 부분이다.

우리나라의 경우, 표본추출 틀로서 주민등록상의 세대명부를 이용한다면 이미 세대명부 자체가 읍·면·동 단위뿐만 아니라 ‘통’과 ‘반’ 그

리고 주소와 세대별로 정리되어 있고 ‘성별’, ‘연령’, ‘가구원 수’ 등 층화에 이용 가능한 변수들을 보유하고 있어서 읍·면·동내 계층별, 연도별로 주택 단위의 계통추출(systematic sampling)만으로도 균등분할에 대한 문제가 해결될 수 있을 것이다.

다. 표본추출 틀의 갱신

미국과 프랑스의 경우, 표본추출 틀이 우편주소를 기반으로 주소록이 작성되기 때문에 거처단위로 주소의 생성과 소멸에 따른 갱신 부분과 실제 표본 추출작업의 대상이 되는 GIS 데이터의 갱신부분으로 크게 나뉘어 진행되고 있다. 기본적으로 우리나라에서도 주소록과 GIS를 중심으로 유사한 갱신이 이루어져야 하겠지만, 주소록의 경우 주택별 주소뿐만 아니라 세대정보까지도 포괄하고 있는 주민등록 세대명부가 지자체를 통해 갱신되고 있으므로 통계청의 입장에서는 효과적으로 관리되고 있는 표본 틀을 적은 노력으로도 확보할 수 있다는 장점을 가질 수 있다.

이렇게 우리나라에서는 좋은 장점을 지닌 정교한 주소록(세대명부)을 보유하고 있으므로 GIS 정보의 이용은 미국, 프랑스와는 달리 주택 단위의 표본추출 틀로서 직접 이용할 것이 아니라, 조사원들의 실제조사시 표본가구의 방문을 위한 지도 작성을 위한 목적으로 한정해서 이용하게 되면 표본추출에 필요한 별도 속성정보의 연계가 필요치 않아 표본추출 작업의 시간 단축과 업무 효율을 기대할 수 있다. 조사원을 위한 GIS 이용이란 ‘새주소’시스템에 포함되어 있는 주소를 세대명부에서 추출된 주소와 연결한 지도를 담당 조사원에게 지급하여 조사대상 가구를 방문할 수 있도록 한다는 의미이다. 향후 GIS의 발전이 거듭되어 건물단위가 아닌 층별, 거처별 데이터에 행정자료에서 얻어지는 속성정보(성별, 연령, 세대주 정보 등)가 완벽히 연계되어 GIS 정보를 표본추출의 기본자료로서 활용할 수 있을 경우에는 표본 틀로서의 GIS 갱신 문제를 좀더 심도있게 검토해야 하겠지만, 현재의 세대명부만으로도 유용한 추출 틀이 될 수 있기 때문에 표본 틀의 갱신에 관한 검토에 있어 GIS 정보의 갱신은 고려하지 않았다.

중복되지 않는 표본의 연속적 누적을 위해서는 각 표본에 대한 유일한 ID관리가 필수적이기 때문에 세대명부의 갱신에 있어 가장 중요한

사항은 세대별로 고유번호를 관리하는 것이다. 현실적으로는 세대명부 데이터에 세대별로 고유번호를 붙여 파일을 관리하여야 한다는 의미인데, 이는 각 연도별 표본추출시 동일한 ‘세대’의 반복 추출을 방지하여 다양한 표본을 누적하기 위한 방법의 하나로서 기존의 조사구단위 추출에서 조사구번호를 부여했던 것과 같은 맥락이다.

한편, 기존의 조사구단위 표본추출 방식과 유사한 방법으로 ‘반’을 단위로 추출할 경우에는 이러한 ID부여과정이 조사구번호 체계와 유사한 방식으로 단순해질 수 있으나 ‘세대’를 단위로 할 경우에는 ‘반’ 이하 단계의 주택번호와 세대의 번호를 추가하여야 할 것이다. 특히 주택번호의 경우는 단순한 일련번호 형식만으로 컨트롤하기에는 유일성이 떨어져 불가피하게 ‘지번’을 사용하는 방식이 될 것이다. 아래의 내용은 기존 조사구번호 부여방식과 세대명부를 이용한 방식에서 새롭게 고려해야 할 부여체계를 비교하였다.

〈표 8〉 조사구 이용방식과 세대명부 이용방식 간의 고유번호 체계 비교

코드 체계	조사구 이용	세대명부 이용	
		‘반’ 단위	‘세대’ 단위
조사구 번호	· 동일 읍면동내에서 일련번호 세자리와 조사구 특성번호 한자리(A, 1~5)를 부여	-	-
‘반’ 번호	-	· 지자체 사용번호 그대로 이용	좌 동
거처 (주택) 번호	· 한 조사구내에서 조사순서대로 일련번호 부여 · 표본추출과는 상관없으며 조사시 식별번호로 이용	좌 동	· 한 반내에서 지번의본번과 부번 8자리 부여
세대 (가구) 번호	· 한 거처내에서 주인(대표)가 구부터 일련번호 부여 · 표본추출과는 상관없으며 조사시 식별번호로 이용	좌 동	· 한 거처내에서 주인(대표)가구부터 일련번호 부여 · 표본추출과는 상관없으며 조사시 식별번호로 이용

라. 표본추출 단위

미국과 프랑스의 경우 공통적으로 조사단위는 가구이지만 추출의 단위는 주소를 중심으로 한 거처단위가 되고 있다. 이는 표본 프레임이 우편주소를 중심으로 한 주소마스터 파일 체계이기 때문에 선택할 수 있는 최선의 방법이었다.

우리나라의 경우는 우선 ‘가구’와 유사한 ‘세대’라는 개념을 중심으로 구축되어 있는 행정등록 시스템이 있으므로 ‘세대’를 단위로 하는 표본추출이 가능하게 된다. 세대단위의 추출은 센서스 표본조사를 통한 소지역자료 등의 생산에 있어 표본의 대표성을 고려할 때 현실적으로 최선의 대안이 될 수 있음은 자명하다. 하지만 순환표본조사의 경우 표본의 중복을 배제하고 각기 다른 표본에 대한 조사결과를 누적해야 한다는 문제 때문에 표본 틀의 개별 세대에 대한 이력을 관리하는 데에 있어 현실적으로 한계가 따른다. 이러한 이력관리에는 기본적으로 고유한 ID 부여가 필수적인데 현실적으로 ‘세대’를 유일하게 구분할 수 있는 식별번호를 부여하는 것이 곤란하기 때문이다. 예를 들어 한 거처 내에 2가구 이상이 거주하는 경우에 이들 가구를 별도로 구분해낼 수 있는 방법은 가구주나 가구원의 성명 정도일 것이다. 그런데 이러한 성명을 코드화하여 ID로 사용하기에는 현실적으로 불가능하다. 따라서, 표본추출의 단위는 불가피하게 유일한 코드로 관리가 가능한 최하의 단위인 거처(주택)가 될 수밖에 없게 된다.

또 다른 대안으로서는 기존의 총조사에서 ‘조사구’를 단위로 하는 집락추출과 유사한 방식을 이용하는 것인데, 이는 지방자치 편제로 존재하는 ‘반’을 ‘조사구’와 같은 개념으로 이용하여 추출하는 방식이다. 기존 총조사의 ‘조사구’는 평균 60가구를 기준으로 묶게 되는데 반해, 지방자치법에 의한 ‘반’의 구성은 대개 20~30가구를 단위로 묶되 50호를 초과하지 않는 범위 내로 한정되어 있다. 2005년 ‘인구주택총조사’에서 최종 확정된 조사구가 273,188개인 반면, ‘반’의 수는 480,461개(2008.1.1. 기준)로 ‘반’의 구성이 세분화되어 있어 ‘조사구’를 단위로 했을 때보다 더욱 정교한 표본추출이 가능할 것이다. 우리나라의 통·리·반의 설치현황은 아래 <표 9>와 같다

〈표 9〉 ‘인구주택총조사’ 조사구와 ‘반’의 규모 비교

지 역	인.총 조사구 수 (A)	‘반’의 수			비 율 (B/A, %)
		계 (B)	도시	농촌	
전 국	273,188	480,461	367,364	113,097	75.9
서 울	56,424	101,948	101,948	-	80.7
부 산	20,348	28,121	27,271	850	38.2
대 구	13,571	21,921	20,383	1,538	61.5
인 천	14,367	21,288	19,777	1,511	48.2
광 주	7,772	11,097	11,097	-	42.8
대 전	8,195	13,674	13,674	-	66.9
울 산	5,663	9,405	7,655	1,750	66.1
경 기	56,575	86,610	70,600	16,010	53.1
강 원	9,188	20,683	13,706	6,977	125.1
충 북	8,767	17,985	8,170	9,815	105.1
충 남	11,661	24,812	14,555	10,257	112.8
전 북	11,001	23,960	13,314	10,646	117.8
전 남	11,684	22,984	7,908	15,076	96.7
경 북	16,606	38,661	15,495	23,166	132.8
경 남	18,287	32,108	18,772	13,336	75.6
제 주	3,079	5,204	3,039	2,165	69.0

*자료출처: 통계청(2005.11.1), 행정안전부(2008.1.1.)

4. 추정 및 공표

가. 추 정

추정과 관련해서는 선협국의 사례에서도 조사방법 개선을 위한 사전 작업시 많은 시간과 전문인력을 투입하여 추진한 바 있으며, 아직 우리나라에서도 본격적으로 연구가 진행되지 않아 더 많은 검토와 연구노력이 필요한 분야이다. 향후 순환표본 센서스 시행을 염두에 둔다면 사전

연구에서 반드시 거쳐야 하는 필수적인 분야로서 통계전문가의 고유영역이라 하겠다. 추정의 기본적인 접근법은 연차별로 조사된 데이터의 누적에 대한 이동평균법을 이용하게 되는데, 이는 5년을 단위로 새로운 연도의 결과가 추가되는 만큼 기존 조사결과 중 신규 데이터 선택을 위해 과거 데이터(6년 전)를 한 계열씩 제외해 가면서 5년간 자료의 평균을 이용하게 된다.

여기서 주목해야 할 부분은 추정에 동원된 자료에 따라 연도별 추정결과와 다년 추정치들 간의 상이성에 관한 것이다. 조사결과가 1년 단위로도 산출되므로 이를 기초로 한 1년 단위의 추정과 2개년 이상의 자료를 이용한 다년 추정이 있을 수 있게 되는데, 다년 추정의 경우, 3개년 자료이용 추정과 5개년 자료이용 추정결과가 주로 이용된다. 이때 어느 항목이 매년 증가 추세를 보인다거나, 등락을 거듭한다거나 하는 경우에 1년 추정과 다년 추정 간의 결과가 다르게 나타날 수 있다는 것이다. 결국, 조사결과 이용자들의 혼란유발과 이로 인한 신뢰성 저하가 문제로 될 수 있어서 프랑스의 경우도 공식적인 추정결과로서 5년 추정 자료만을 인정하고 있는 상황이다.

또한, 연속추정에 따른 자료의 누적이 가능하기 위해서는 조사항목에 대한 일관성의 유지가 필요하다. 연간 표본의 규모에 따라 달라지는 하겠지만 보통 소지역(읍·면·동) 자료를 생산하기 위해서는 5년간의 표본 누적이 요구되는데, 만약 정책기초자료서의 필요성 등으로 인해 신규항목이 1회용으로 추가될 경우, 기존의 총조사와는 달리 횡단면 분석자료로서의 성격이 약화된 순환표본 조사체계에서는 전국 또는 시·도별 자료의 제공으로 밖에 산출되지 않음을 분명히 인식해야 한다.

나. 결과 공표

센서스 결과산출 단계에서 순환표본의 가장 큰 장점인 연도별 자료 제공이라는 큰 장점이 부각되는데, 이는 순환표본조사 개발 연구의 태동 배경이라고도 할 수 있다. 여기서 유의해야 할 점은 순환표본을 적용하고도 최초 주기 5년 동안에는 지역별 공표 수준 측면에 있어서는 다소 불충분한 자료가 산출될 수밖에 없다는 것인데, 이는 소지역 추정을

위한 5년간 자료의 축적이 충분히 이루어지지 않았기 때문이다.

또한, 순환표본 방식의 도입 초기 5년간은 연도별로 연간 추정자료와 다년 추정자료를 동시에 발표해야 하겠지만, 소지역 추정에 필요한 표본의 축적이 완료된 이후, 즉 5년간의 순환조사 후에는 공식적인 인구센서스의 결과로서 5년 추정치만을 발표하는 것이 바람직할 것이다. 이는 이용자의 혼란을 방지한다는 측면뿐만 아니라, 표본의 충분한 누적에 완료된 이후에야 비로소 정도 높은 조사결과를 산출할 수 있기 때문이다.

이와 더불어, 결과공표에서 참고로 해야 할 것은 순환표본조사의 5년 추정치는 5년의 중간년도를 기준으로 한다는 점이다. 미국에서도 본격적인 순환표본 적용에 앞서 2004년에 2000년 센서스 조사결과를 ACS 결과와 비교할 때 대상이 되었던 자료는 1999 ~ 2001년의 3개년 ASC 추정결과로서 2000년을 기준으로 한 자료이었는데 이는 항목별 특성값을 5개년의 평균치로 이용하고 기준이 되는 인구와 가구, 주택 자료는 그 중간년도를 기준으로 하였다. 이렇게 기존의 전통적 센서스 방식에서의 결과는 해당 연도를 기준으로 전수결과와 표본 결과가 산출되었지만 순환표본조사를 도입하게 되면 전수결과 기준 연도와 표본조사 기준 연도가 달라진다는 점도 고려해야 한다.

5. 조 사 비 용

영국 통계청(ONS)에서 2003년에 검토한 순환센서스 도입 검토 보고서에 의하면 기존의 전통적 센서스에서 보다 순환센서스로 전환시 약 10%의 비용이 더 들 것이라고 밝히고 있다. 이는 프랑스와 같이 완전한 순환센서스 체제로 전환할 경우를 상정하여 평가한 것으로서 미국과 같이 전수조사를 그대로 유지하며 표본조사만을 순환표본 방식으로 대체하는 형태에서의 비용에 대한 평가와는 차이가 있을 수 있으며, 미국의 경우도 ACS관련 조사인력을 정규화하여 운영하고 있는 관계로 임시조사원을 동원해야 하는 우리나라 여건상 조사비용의 추산에 있어 외국의 사례를 직접적으로 참고한다는 것이 곤란함을 미리 밝혀둔다.

기본적으로 비용을 추정하기 위해서는 표본의 규모가 전제되어야 하

는데, 이러한 적정 표본규모의 산정에는 각 결과표별 표본오차 고려 등 장시간의 작업이 수반되므로 본 연구에서는 참고적인 자료로서 대략의 비용만을 추정해 보기위해 미국의 사례를 기준으로 하였다. 이는 미국 ACS에서의 각 센서스 몇 개의 block을 묶은 소지역 단위 결과 산출을 위한 추출률을 우리나라의 소지역(읍·면·동) 단위를 기준으로 적용할 경우로 별도의 센서스 block이 없는 우리나라의 상황과 정확히 일치한다고 보기 어려우나 빠른 시간 내에 대략의 표본규모를 산정해 볼 수 있게 해준다. 미국은 5년간 총 12.5%의 표본으로 조사를 실시하고 있다고 밝히고 있는데 이렇게 산출된 규모는 숙련된 정규직원을 동원함으로써 비표본오차를 철저히 관리해서 총오차를 줄인다는 취지에서 표본오차를 고려한 수준보다 적은 규모로 산출된 것이다.

이러한 ACS의 표본규모를 기준으로 하고 임시조사원을 동원해야 하는 우리나라의 현실을 감안하여, 대략 5년간 총 15%의 표본규모(1년간 3%)를 가정한다면 2005년 인구주택총조사 당시의 예산을 기준으로 5년간 약 369.5억원(1년 38.7억원) 정도의 비용이 필요한 것으로 추정된다. 이러한 추정에는 현실적으로 참고 가능한 자료가 2005년 예산자료로 한정되어 물가수준이 2005년을 기준으로 적용된 점과 ‘1가구당 비용’ 산출 시, 전수조사와 표본조사의 비용을 구분할 수 없어 동일하게 적용된 것 때문에 현실적으로 소요될 비용과 다소의 차이가 발생할 여지를 포함하고 있다 하겠다.

〈표 10〉 순환표본조사 추정예산 규모

2005년 예산 (단위: 억원)	1가구당 비용 (단위: 원)	1년 표본규모 (3%, 단위: 가구)	1년 추정예산 (단위: 억원)	5년 추정예산 (단위: 억원)
1,290	8,068.4	479,648	38.7	369.5

6. 또 하나의 대안‘가상 센서스’

가. 기본 체제

2001년 네덜란드 센서스 결과는 행정등록자료와 기존의 각종 표본조사

결과를 이용하여 센서스결과를 산출하였다. 이른바 ‘가상 센서스(Virtual census)’라는 개념으로 실시된 이러한 새로운 시도는 비록 자료의 불충분성에 관한 문제는 존재하지만 전통적 방식의 센서스와 비교하여 비용과 불응 문제가 축소되면서 효과적으로 센서스 결과 생산이 가능했다고 한다. 이 방식은 외국과의 센서스 결과 비교 및 과거 센서스 결과와의 시계열 비교에 필요한 결과표 작성에 목적을 두고 있으며, 이러한 방식도 센서스 실시방법에 따라 크게 나눌 경우, ‘행정등록센서스 + 표본조사’의 부류에 속하는 형태라 볼 수 있다.

이러한 ‘가상센서스’는 순환표본조사 방식과 비교해 볼 때, 크게 두 가지 면에서 차이점을 발견할 수 있게 되는데, 첫 번째는 인구센서스 표본조사 만을 위한 별도 조사를 실시하지 않는다는 것으로 이는 통계작성의 경제성 측면에서 가장 우수한 장점이라고 볼 수 있으나 한편으로는 센서스 결과 작성에 있어 가용 자료의 한계성은 내포하고 있다 하겠다. 두 번째로는 표본의 누적방식이 아니라는 점인데, 순환표본방식의 경우, 초기 자료 누적에 필요한 5년 정도의 기간이 지나면 비교적 충분한 데이터를 통해 1년 단위로 결과 산출이 가능해지고 표본도 중복되지 않아 대표성이 제고될 수 있다는 장점이 부각될 수 있는 부분이다. 네덜란드는 필요에 따라서 1년을 주기로 하는 집계 및 결과공표가 가능하다고 하지만, 지금도 과거와 같이 센서스 결과 공표주기를 10년으로 운영하고 있어 현실적으로 1년 단위의 공표가 가능한지에 대한 실질적인 입증은 되지 않고 있는 상태이다.

나. 기초 자료

가상 센서스의 근간을 이루고 있는 기초자료로는 지자체가 관리하고 있는 행정등록자료와 통계청에서 별도의 목적으로 실시되고 있는 몇 가지의 표본조사 결과이다. 우리나라의 상황과 비교할 경우, 유사한 행정등록자료와 표본조사가 실시되고 있어 인구센서스 결과 산출에 이용하기 위한 세부 검토가 전제된다면 기초자료로서 활용이 가능할 것이다.

가상 센서스에 있어서 가장 기본적으로 사용하고 있는 자료원은 지자체가 운영하고 있는 ‘인구등록부’(automated municipal population registers)

로서 크게는 우리나라 주민등록 시스템과 유사하다고 볼 수 있는데, 세부적으로는 사회보장관련 사항, 전기와 수도관련 사항, 외국인관련 경찰서 등록사항, 노인연기금관련 사항이 포함되어 있다는 점에서 다르다고 할 수 있다. 우리나라가 이와 비슷한 자료를 구축하기 위해서는 주민등록과 각 기관별로 흩어져 있는 기타 등록자료를 연계 구축하여야 할 것이며, 이러한 연계 구축에는 필수적으로 ‘개인식별번호(PIN: personal identification number)’를 기본 연계 key로 하여 개인별로 연계 되어야 함은 물론이다.

또한, 각종 인구·사회적 특성을 파악하기 위한 별도의 표본조사 결과 이용에 관해서는 크게 3가지의 자료를 동원하게 되는데, 직업 및 교육수준 파악을 위한 자료로서 ‘노동력 조사’(LFS: the Labour Force Survey) 결과와 경제활동관련 사항에 대규모 서베이인 ‘고용 및 소득 조사’(SEE: the Survey Employment & Earnings) 결과, 주거관련 사항과 관련된 ‘주거관련 등록자료 및 주거환경조사’(SHC: the Survey Housing Conditions) 결과가 그 것이다. 우리나라의 경우는 ‘경제활동인구조사’, ‘인력실태조사’, ‘가계조사’, ‘사회통계조사’, ‘가계자산조사’ 등이 이와 유사한 개념의 표본조사들로서 각 조사들 간의 항목에 대한 개념 일치 작업 및 각 조사의 표본설계 과정 검토 등을 거치게 되면 센서스 결과 산출을 위한 기초자료로서 이용이 가능한 자료들이다.

다. 작성 방법

네덜란드 통계청은 2001년 센서스에서 결과표 전반의 일관성을 유지하기 위하여 여러 자료에서 얻어진 데이터들의 수리적 일관성을 담보하는 새로운 추정방법을 개발하였다. 이 방법은 ‘repeated weighting’ 이라고 하며 이는 다른 소스에서 오는 각 테이블의 불일치성을 제거하기 위한 회귀적 방법을 반복 적용하는 것으로 결과표별 일관성 문제를 해결하였다. 하지만, 각 셀별 결측 값의 문제는 해결되기 어려워 셀을 통합하는 방법 등을 적용하게 되는데, 이러한 셀들이 많아진다면 전체적인 추정의 정도에 의문이 제기 될 수 있다.

이러한 결과표별 결측 셀 문제는 표본자료가 불충분할 경우 더욱 심

화될 것이고, 앞서 살펴 본 바와 같이 센서스 결과를 작성하기 위해 별도 목적의 여러 표본조사를 동원함에 있어 표본조사별로 승수가 달라 이에 따른 결과표별 일관성 조정 단계에서 이론적 합리성이 결여될 경우 문제될 소지가 있을 수 있다.

따라서 우리나라에서 가상 센서스 개념과 같은 방식으로 센서스 결과를 산출하고자 한다면 관련되는 기존 표본조사들의 표본규모 재조정, 조사항목들 간의 일관성 검토와 함께 보다 정교한 ‘repeated weighting’ 기법의 개발과 테스트 등에 대한 추가 연구가 진행되어야 할 것이다.

라. 도입에 따른 고려사항

크게 증가하고 있는 센서스 실시비용 문제는 조사환경의 악화라는 또 다른 이유와 함께 센서스 조사방법에 대한 변화를 유인하는 주요한 이유로서 작용해 온 것이 사실이다. 대표적으로 프랑스가 순환센서스 방식으로 전환한 배경에는 전통적 센서스에서 일시적으로 대규모 비용이 발생하는 문제 때문에 의회와의 마찰을 피해 센서스 예산을 나누어 집행하고자 하는 의도가 있었다는 것이 주지의 사실이다.

이러한 맥락에서 볼 때 ‘가상 센서스’의 경제성은 가장 큰 장점으로 부각될 수 밖에 없으나 자료의 불충분성과 관련된 단점과 함께 자료의 가용성 측면에 대한 고려, 특히 센서스 결과가 다른 표본조사의 표본 틀을 제공한다는 특성을 고려해 볼 필요가 있다. 센서스는 그 자체의 결과로서 여러 정책의 기초자료, 인구·사회 특성 분석을 통한 학술적 가치 등으로 충분히 중요한 가치가 있을 뿐만 아니라 부수적 기능으로서 가구를 대상으로 하는 통계조사의 설계에 있어서도 산업, 직업, 주택형태 등 기본정보를 제공하는 기본 틀이 되고 있기 때문인데, 네덜란드와 같이 성별, 나이, 주소 등 기본적인 정보뿐만 아니라 산업, 직업 등 표본설계에 필요한 추가적인 특성이 행정등록자료에 구축되어 있는 경우에는 문제가 없을 수 있었을 것이나 우리나라와 같이 전통적 센서스 방식을 유지해오던 경우, 행정등록자료의 연계 구축도 미비된 상태에서는 이에 대한 추가적인 고려가 필요할 것이다.

또한, 순환표본 시스템에서의 가장 큰 장점인 1년 단위의 결과생산

가능성과 관련하여 ‘버추얼 센서스’ 시스템에서도 그와 같은 목적이 달성될 수 있을 것인지에 대해서도 생각해 볼 필요가 있는데, 이론적으로는 가능할 수도 있다고 하지만, 실질적으로 적은 규모의 표본으로 여러 지역기준으로 결과를 매년 생산할 경우에 시계열 유지가 가능할 것인가에 대한 의문이 남게 된다. 현재 네덜란드는 필요성을 못 느껴 10년 단위로만 센서스 결과를 산출하고 있어서 이에 대한 실증은 아직 이루어지지 않았다고 볼 수 있다.

제4절 결 론

통계조사를 둘러싼 환경의 급격한 변화는 양적, 질적인 통계수요의 증가와 더불어 개인정보 보호의식 강화에 기인한 조사환경의 악화라는 양면적 성격을 띠고 있다. 통계생산을 담당해야 하는 중앙통계기관들의 고민은 이러한 부분에서 시작된다. 세계 각국의 센서스방법 개선의 노력도 이와 같은 부분에 초점이 맞춰지면서 조사응답 부담의 경감을 위해 행정등록센서스를, 통계 수요자의 욕구 충족을 위해 순환센서스를 개발하게 되는데, 여기서 순환센서스를 수요자 중심이라고 얘기할 수 있는 근거는 연도별 자료생산이 가능하다는 점과 소지역 통계 작성에 보다 유리하다는 점을 들 수 있다.

본 연구에서는 우리나라의 본격적인 센서스 개선 노력의 일환으로 외국의 개선사례 중 보다 수요자 중심적이라 할 수 있는 순환센서스를 대상으로 전반적인 실시현황을 살펴보고자 하였으며 이를 토대로 우리나라의 도입방향을 제시하고자 하였다. 순환센서스의 경우는 전수조사와 표본조사를 모두 아울러서 순환되는 표본의 누적을 통해 조사하는 개념으로서 전수조사를 행정등록자료로 대체할 수 있는 가능성이 높은 우리나라의 현실을 고려하여 표본조사만을 순환 조사하는 방식을 중심으로 검토하였다. 이른바 ‘순환표본조사(Rotation sampling survey)’라고 하는 방식 검토는 미국의 ‘미국지역사회조사’(ACS: American Community Survey)를 중심 모델로 하여 이루어 졌으며 전수조사 실시방식에 있어서는 우리나라의 개선방향과 다를 수 있다는 점을 감안하였다.

프랑스와 같은 ‘순환센서스(Rolling census)’이든 미국 ACS와 같은 ‘순환표본조사(Rotation sampling survey)’이든 순환조사에 관해서는 비판 목적 연차 표본의 누적이 전제되어야 하기 때문에 정밀한 표본추출 틀(sampling frame)의 구축과 유지는 순환표본 설계에 있어서 기본적인 분야로서 행정자료를 통한 표본추출 틀 제공에 대해 비중을 두고 검토하였다.

한편, 우리나라가 순환표본방식에 의한 조사를 본격적으로 도입하기 위해서는 추정분야 전문가에 의한 심도있는 추가 연구의 진행이 반드시 필요한데, 이는 순환표본의 특성상 정밀한 표본설계와 이에 따른 추정 과정이 매우 복잡하기 때문이다. 선형국의 사례에서도 표본추출 틀 구축분야와 함께 추정분야에도 사전 검토 및 연구에 충분한 투자가 이루어졌던 바와 같이 순환조사의 추정에 대하여 전문성과 충분한 시간을 가지고 별도의 주제로 연구해야 할 필요가 있다.

‘가상 센서스(Virtual census)’의 경우도 센서스방식 개선에 있어 고려해볼 만한 가치가 있는 시스템으로서 무엇보다도 적은 비용으로도 센서스 결과 산출이 가능하다는 장점을 들어 소개하였으며, 자료의 불충분성과 관련된 문제점에 대해서도 언급하였다. 그러나 여기서 한번 더 짚고 넘어가야 할 사항은 가상센서스로 전환할 수 밖에 없었던 1990년 즈음의 네덜란드 국내상황이다. 이는 당시 국민들의 센서스에 대한 저항과 그로 인해 센서스 실시를 포기할 수밖에 없었던 특수한 상황을 말하는 것으로서 그 후로 10년간의 네덜란드 통계청의 노력으로 겨우 자리를 잡았다고는 하나, 결국 가상센서스라는 것이 어쩔 수 없었던 현실적인 상황에서 잉태된 교육지책의 산물이었다는 점은 우리가 반드시 유념해야 할 부분이다. 이는 곧 현재 우리나라의 현실에 비추어 벤치마킹의 대상을 비교 평가해야 할 때, 가상센서스라는 시스템의 태생적 한계성과 역사적 특수성이 충분히 감안되어야 할 부분이라는 것이다.

우리나라와 비교하여 상대적으로 좀더 오랜 현대적 통계작성 역사를 가진 유럽과 미국과 같은 국가들이 전통적인 센서스의 한계상황을 좀더 먼저 맞이하여, 그 방식 전환에 대한 많은 고민과 준비를 통해 스스로의 자구책을 강구해 온 결과가 바로 ‘순환 센서스’와 ‘버추얼 센서스’ 등의 형태로 나타난 것으로 보인다. 우리나라도 센서스 방식 전환

에 대한 인식이 시작되고 있는 현재 시점에서 외부의 압박이 아닌 내부의 자발적 전환 의지에 의한 체계적 준비를 통해 적절한 전환 시기의 확보에 실기하지 않도록 선협국의 성공과 실패를 거울삼아 우리나라 통계현실과 부합하는 새로운 토착 모델 개발에 시간과 노력을 아끼지 말아야 할 것이다.

참고문헌

- Alexander, C.H. et al.,(1997), "Making estimates from the American Community Survey"
- Desplanques(2000), "The new census in France"
- Dumais, Jean et al.(1999), "An alternative to traditional census taking: plans for France"
- Durr, Jean-Michel et al.(2001), "Redesign of the French census of population"
- Durr, Jean-Michel(_), "The French new rolling census"
- GAO(2004), "American community survey: Key unsolved Issues"
- GAO(2007), "Census Bureau has improved the local update of Census Address Program, but challenges remain"
- Groves, Robert M. et al.(2004), 「Survey Methodology」, John Willy & Sons
- INSEE(1999), "Post 2000 Census in INSEE"
- INSEE(2003), "The new Franch population census"
- INSEE(2006), "The population Census in France: from general census to rolling census?"
- INSEE(2007), "Quality management in the French census"
- INSEE, "The Population Census from 2004 on"
- Kish, Leslie(1990), "Rolling Samples and Censuses", Survey Methodology, June 1990, Vol. 16, No. 1, pp 63-79, Statistics Canada
- Nordholt, Eric Schulte, "The Dutch Virtual Census of 2001: A register-based approach combined with survey information"
- Nordholt, Eric Schulte(2005), "The Dutch Virtual Census of 2001: A new approach by combining diferent sources"
- Nordholt, Eric Schulte et. al.(2007), "Record matching for Census purposes in the Netherlands"
- ONS(2003), "Alternatives to a Census: Rolling Census"
- ONS(2003), "Alternatives to a Census: Review of international approaches"
- Prins, C.J.M.(2000), "Dutch population statistics based on population register"

data"

Statistics Netherlands(2000), "Netherlands Official Statistics", Volume 15,
Summer 2000

Statistics Netherlands, "The Dutch Virtual Census of 2001"

Statistics Netherlands(2004), "The Dutch Virtual Census of 2001: Analysis
and Methodology"

Statistics Canada, 「Survey Skills Development Course 교재」, Statistics Canada
UN(2007), "Population and housing censuses",

<http://unstats.un.org/unsd/demographic/sources/census/alternativecensusedesigns.htm>

US Census Bureau(2000), "A better strategy is needed for managing the
nation's Master Address File"

US Census Bureau(2004), "Reengineering the census of population and
housing"

US Census Bureau(2006), "Design and Methodology: American Community
Survey", Technical paper 67

은기수 등(2007), 「2005년 인구주택총조사 종합분석 보고서」, 서울대학교
산학협력단

이건 등(2007), 2010년 인구주택총조사 방법론 연구」, 한국조사연구학회

통계청(2005), 「2005 인구주택총조사 지침서」

통계청(2006), 「2005 인구주택총조사 보고서」

통계청(2006), 「2005 인구주택총조사 종합평가보고서(I)」, 내부자료

<부록>

반복가중기법(Repeated weighting)

1. 개요

2001년 네덜란드 센서스의 결과표는 등록 자료와 기존의 표본조사 자료를 이용하여 작성되었으며, 수치적으로 일관성 있는 표의 작성을 위해 반복가중기법(repeated weighing)을 사용하였다.

표본조사 결과를 이용하여 결과표를 생성하면 ① 서로 다른 표본조사 결과를 이용하여 구한 추정치들이 수치적으로 불일치하고 ② 표본 오차로 인해 모수와 추정치 간의 차이가 항상 존재한다는 문제점이 발생한다. 반복가중기법은 이러한 통계적 불일치를 해결하기 위해 개발된 방법으로 각 표본조사에서 두개 이상의 가중치를 사용하여 각 추정치들이 서로 일치하도록 조정하고 회귀방정식을 반복하여 이용한다.

반복가중기법은 1단계로 등록부나 표본조사에서 관련 정보를 추출하여 직사각형의 data block에 삽입한 다음, 2단계로 각 data block에 일련의 가중치를 부여하고, 마지막 3단계에서는 각 data block을 이용하여 수치적으로 일관성 있는 결과표를 생성한다.

2. 반복가중기법

2001년 센서스 결과표 작성에 필요한 자료에는 인구 관련 변수 및 취업상태 등의 등록부에서 구할 수 있는 자료와 교육정도, 직업 등의 등록부에서 구할 수 없는 자료가 있다. 이 중 등록부에서 구할 수 있는 자료는 등록되어 있는 모든 인구를 포괄하며, 등록부에서 구할 수 없는 자료는 표본으로 선정된 사람에 한하여 정보를 얻을 수 있다. 표본조사 자료를 이용하는 경우에는 추정이 필요하다.

가. 반복가중기법 순서

1) 추정할 결과표 고려

각 결과표에 1개의 응답변수(자연수) 및 1개 이상의 클래스 변수가 관련된다. 응답 변수는 보통 인구수 및 가구수이며 클래스 변수에는 성별 및 나이 등이 있다. 클래스 변수는 1개 이상의 단계(level)가 존재할 수 있다. 예를 들면 한살 단위 계급, 5살 단위 계급 등이 있다. 서로 다른 단계의 클래스 변수는 계층적(hierarchical)이어야 한다. 즉 상위 단계에서의 표는 하위 단계의 결과표들을 합한 결과와 일치하여야 한다.

2) 사용 가능한 자료 고려(행정등록 자료 및 표본조사 자료)

이러한 자료들은 마이크로 레벨에서 연결(link)되며, 연결된 자료는 모집단 각 구성요소에 대해 1개의 레코드를 가지는 데이터베이스가 된다. 등록부에서 구한 자료는 모집단의 각 구성요소들에 대한 정보를 가지고 있으며 일부 구성요소들에 관한 추가 정보는 한개 이상의 표본조사에서 얻을 수 있다.

이 데이터베이스에서 직사각형의 완전히 채워진 데이터블록이 추출되고, 각 데이터블록은 특정의 공통 변수들을 가지는 모든 레코드로 구성된다. 예를 들면 등록블록에는 등록부에서 얻을 수 있는 모든 변수 및 구성요소에 관한 기록이 있고, 1개의 표본조사에 기초를 둔 블록은 표본조사와 등록 자료를 연결한 블록이 된다. 2개 이상의 표본조사로 구성된 보다 복잡한 블록도 가능하다.

3) 결과표 추정

이러한 직사각형의 데이터블록을 이용하여 필요한 결과표가 계산되거나 추정된다. 우선 각 데이터 블록에는 반복가중기법 적용을 위해 초기 가중치가 부여된다. 등록 블록내의 각 요소에 대해서는 초기 가중치는 1로 주어지고, 표본조사의 정보를 가지고 있는 요소에 대한 초기 가중치는 표본조사 결과의 공표 시에 사용된 가중치를 적용한다. 초기 가중치는 무응답 조정이 고려되며, 보조 정보를 이용하여 추정치의 분산

을 최소화하는데 이러한 정보는 공표 가중치에 포함되어 있다.

초기 가중치 결정 후 추정 방법은 첫째, 각 결과표의 작성에 필요한 모든 변수를 포함하는 가장 큰 데이터 블록을 이용하여 결과표를 추정하고

둘째, 각 결과표의 모든 마진을 우선 추정한다. 이때 이미 추정된 마진이 고려되어야 함

셋째, 가능하면 초기 가중치를 이용하여 결과표를 추정하여야 한다.

만약 초기 가중치를 이용한 추정치가 기 추정된 결과표와 일치하지 않은 경우에는 반복가중기법을 사용하여야 하며 반복가중기법 사용 시 결과의 어떤 마진이 기 추정된 결과표와 중복되는지를 우선적으로 검사하여야 한다. 이러한 공통 마진의 모수 합계치가 알려진 모수 합계치로 사용되며 회귀추정치의 조정에 대한 목표치가 된다. 실제로 조정(calibration)이라는 말은 초기 가중치를 조정하여 알려진 마진과 일치되어야 한다는 제약조건을 충족시키는 것을 의미한다. 조정의 범위는 결과표 및 기 추정된 결과표에 따라 달라진다. 원칙적으로 모든 결과표는 서로 다른 가중치를 사용하여 추정될 수 있다.

이론적으로는 반복가중기법을 사용하여 완전히 일치하는 결과표의 추정이 가능하지만, 실제적으로는 몇 가지 추정문제가 발생한다. 자주 발생하는 문제는 표본수가 0인 경우로, 모집단의 특정 범주가 표본에 선정되지 않은 경우로서 조정이 불가능 하다. 두 번째 문제로는 서로 다른 변수간의 상관관계이다. 이러한 상관관계는 서로 다른 두 가지 변수간의 편집규칙의 결과로서 발생할 수 있다. 예를 들면, 경제활동 상태와 직업 간의 관계를 들 수 있다. 취업자만이 직업을 가질 수 있다. 변수나 결과표를 만들 때, 변수들 간의 이러한 관계가 고려되어야 한다. 세 번째는 상호 불일치하는 마진들이다. 자료와 원하는 마진이 주어졌을 때, 결과표의 항목이 채워지지 않을 수가 있다.

반복가중기법을 사용하여 결과를 추정하기 위해, 네덜란드 통계청은 VRD라는 패키지를 개발하였다. 변수, 단계, 단계간의 계층관계, 데이터 블록이 VRD에 입력되면, 원하는 결과표가 생성(등록 정보만이 사용될 경우)되거나, 추정(표본조사 정보가 사용될 경우)된다.

반복가중기법을 사용하여 원하는 결과표의 추정을 위해 VRD를 사용하기 전에 아래의 내용과 같이 4가지의 구체화해야 할 사항이 있다.

- ① 모든 범주 및 모든 단계에 대한 변수들
- ② 각 클래스 변수의 서로 다른 단계들 간의 계층적 관계
- ③ 추정될 결과표
- ④ 직사각형 데이터블록의 구성

나. 결과표 및 사용된 변수

여기서는 반복가중기법을 사용하여 추정될 30개의 결과표에 대해서 설명한다. 이 결과표 중 일부는 1개 이상의 하위표로 구성된다. 따라서 추정될 결과표의 숫자는 30보다 조금 많다. 일부 결과표는 등록정보를 이용하며 이 경우에는 추정이 불필요하다. 표본조사 정보도 이용하는 결과표의 작성에는 반복가중기법을 사용한 추정이 필요하다.

결과표에 관한 검토가 완료된 후, 사용될 클래스 변수 및 범주, 단계 및 계층적관계가 구체화되어야 한다. 일부 세분된 결과표의 개체 수가 적은 셀에서 발생하는 예기되는 문제 때문에, 일부 클래스 변수는 실제로 필요한 단계수보다 더 많은 단계를 부여한다. 이로 인해 추정과정은 보다 더 유연해 진다. 추정과정에서 관측치가 없거나 그 수가 매우 적은 범주가 생길 경우에는 덜 세분화된 단계에서 재추정을 하는 일이 더 쉬워진다.

표7.에는 2001년 네덜란드 센서스의 결과표를 구성하는 모든 변수들이 나열되어 있으며, 각 클래스 변수들의 단계 수 및 자료 원이 표시되어 있다. 변수들은 두 부분으로 분류되어 있다. 첫 부분에는 인구에 관한 클래스 변수가 고려되고, 두 번째 부분은 가구에 관한 변수로 구성된다. 자연수인 가구수 변수 외에 가구의 크기라는 양적변수가 존재한다. 표 14.1에 사용된 자료원은 중앙인구등록(PR; central Population Register), 노동력조사(LFS; Labour Force Survey), 고용 및 소득 조사(SEE; Survey on Employment and Earnings) 및 사회통계데이터베이스(SSD; Social Statistical Database)이다. PR은 센서스의 핵심부분이며, LFS와 SEE는 표본조사이고 SSD는 고용인, 고용주 및 연금 등의 정보를 포함하는 등록부들의 결합이다. 일부 변수들은 등록부와 표본조사를 결합한 정보에

기초를 두었다. 따라서 특정 변수의 첫 번째 단계는 등록 정보에 의해서만 결정(따라서 모집단에 속하는 모든 요소에 대해 사용 가능)된다. 반면에 보다 상세한 단계들은 표본조사의 정보에 의해서도 결정되기 때문에 표본조사와 관련된 요소에 대해서만 사용 가능하다.

〈표 11〉 가상센서스에 사용된 변수

변수 설명	단계 수	자료원
인구에 관한 변수		
성	1	PR
연령	7	PR
가구 위치	3	PR
가구 형태	5	PR
가구 크기	3	PR
센서스 1년 전 거주지	2	PR
국적	5	PR
출생국	1	PR
지역(NUTS)	5	PR
교육정도	5	LFS+PR
경제활동	3	LFS+SSD
경제활동 상태	7	LFS+SSD
직업	4	LFS+SSD
직업 크기	3	SEE+SSD
산업	6	SSD
변수 ‘사람 수’	양적변수	PR
가구와 관련된 변수		
가구의 핵	2	PR
가구 형태	3	PR
자녀 수	6	PR
25세 미만 자녀수	6	PR
18세 미만 자녀수	6	PR
6살 미만 자녀수	6	PR
가구 형태/파트너의 경제상태	3	LFS+SSD
경제활동인구 수	2	LFS+SSD
가구 크기	양적 변수	PR
변수 ‘가구 수’	양적 변수	PR

변수들이 구체화된 후의 다음 단계는 추정 또는 구성할 결과표를 구체화하는 일이다. 두 가지의 서로 다른 형태의 대상(인구 및 가구)이 있기 때문에 적어도 두개의 VRD-session을 돌려야 한다.

관측치가 아주 적은 범주에서 예기되는 문제점과 일부 결과표는 전체 모집단을 목표로 하지 않기 때문에 인구를 대상으로 하는 결과표에 대해서는 여러 번의 session을 돌리기로 결정하였다. 각 세션은 고유의 목표 모집단을 가지고 있다. 첫 번째 세션에는 15살 미만의 어린이에 대한 하위표가 포함된다. 이 그룹에 대해서 변수들에 대한 값은 등록부의 정보만으로 구해진다. 그리고 이 그룹은 많은 다른 결과표의 모집단에서 제외된다. 유사한 이유로 75살 이상의 노인들에 대해서도 분리된 세션을 운영하는 것이 현명하다. 마지막으로 15-74세의 사람들에게 대해서는 두개의 세션이 존재한다. 한 세션은 합계 치에 관한 것이고 나머지는 세션은 일반가구(private households)에 관한 것이다. 후자의 경우는 집단 가구에 관한 자료가 충분하지 못하기 때문에 필요하며, 결과표 작성에는 집단가구에 대한 정보가 불필요하다.

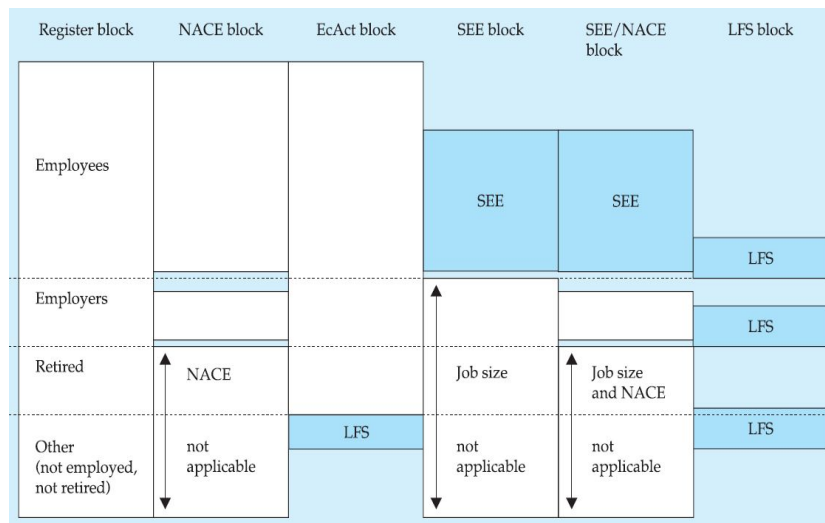
요약하여 말하면, 5개의 VRD 세션이 운영된다. 가구에 관한 세션 1개와 개인에 관한 4개의 세션이다. 네덜란드 전체를 목표 모집단으로 하는 결과표(하위 결과표)들은 3개의 부분(15세 미만, 15-74세, 75세 이상)으로 추정하기 때문이다. Eurostat이 요구하는 결과표를 얻기 위해서는 이 부분들을 합쳐야 한다.

다. 데이터 블록의 유도

위에서 설명하였듯이 반복가중기법은 직사각형의 완전히 채워진 마이크로 데이터블록을 사용한다. 여기서는 15-74세의 사람들에게 대한 데이터블록을 예시한다. 이 VRD 세션은 두드러진 모든 데이터블록들을 사용하는 가장 광대한 것이다. 이 데이터블록들은 [그림 4]에 그래프로 표시되어 있다. 각 데이터블록은 특정의 변수들을 공통적으로 가지고 있는 레코드로 구성되어 있고 각 레코드는 이 공통의 변수 모두를 가지고 있다. 15-74세의 사람들에게 관한 VRD 세션의 사용을 위해서 6개의 데이터블록을 구성하였다. 이 데이터블록들을 크기의 순서에 따라 소개하

겠다.

첫 번째 데이터블록은 등록블록이다. 이 등록블록은 PR과 SSD의 모든 등록정보를 포함하며 15-74세의 사람들(12,036,171)에 대한 목표 모집단의 모든 레코드를 포함하고 있다. 표 14.1에 대해서 설명하면, PR과 관련된 모든 변수들 및 등록부와 표본조사의 정보를 결합한 변수들의 첫 번째 단계가 첫 번째 데이터블록에 포함되어 있다. 이 블록으로부터 등록정보를 이용한 결과표만이 생성될 수 있으며 추정할 필요는 없다.



[그림 4] Data blocks for persons of 15-74 years of age

두 번째 데이터블록은 NACE¹³⁾ 블록(11,985,413 레코드)으로 불리어진다. 이 블록에는 취업하지 않은 15-74세의 개인에 관한 모든 기록 및 NACE 코드가 알려진 취업자가 포함된다. 다시 말하면 NACE 블록은 NACE 코드가 알려지지 않은 레코드를 제외한 목표 모집단의 모든 레코드를 포함한다. NACE 블록은 사실상 등록정보를 포함하고 있지만 표본조사 블록으로 취급된다. NACE 블록은 알려지지 않은 NACE 코드를 알려진 NACE 코드에 고르게 나눠주기 위해서 생성되었다. NACE

13) NACE 코드: 유럽 산업분류 기준에 기초를 두고 1970년에 도입되었으며 1990년에 개정되었음

블록은 등록블록에 나타난 모든 변수들 및 산업분류(NACE) 변수를 포함하며, 산업과 등록정보 만이 결합된 모든 결과표를 추정하는데 사용된다.

세 번째 블록은 EcAct 블록(8,793,530 레코드)이라 불린다. 이 블록은 등록명부에 있는 모든 취업자 및 은퇴한 사람들에 관한 레코드를 포함하고 LFS 상의 모든 비취업자 및 은퇴하지 않은 사람들을 추가한 것이다. 이 블록은 경제활동상태와 등록정보 만을 결합한 모든 결과표를 추정하기 위해서 생성된 것이다. EcAct 블록을 사용하여 모집단에 속하는 특정 그룹의 경제활동 상태를 추정할 때, 취업자 수, 고용자 수 및 은퇴한 사람의 수는 항상 등록명부를 사용하여 계산하며 따라서 추정이 불필요하다(만약 LFS 블록을 사용하면 동일한 결과표의 추정이 필요하다). EcAct 블록에는 등록부 상의 모든 변수들, 경제활동 상태 변수 및 경제활동 변수가 포함된다.

네 번째 블록은 SEE 블록(8,144,956 레코드)으로 불린다. 이 블록은 표본조사 블록 중 가장 규모가 큰 것이다. 이 블록은 SEE 표본조사의 15-74세 모든 취업자, 고용자 및 비취업자의 레코드를 포함한다. 이 블록에 포함되는 변수에는 SEE 변수 인 일자리 수(job size) 및 등록블록의 모든 변수가 있다. SEE 블록은 일자리 수와 등록 정보만이 결합된 결과표를 추정하는데 사용된다.

다섯 번째 블록은 SEE/NACE 블록(8,109,261 레코드)이라 불린다. 이 블록은 NACE와 SEE 블록의 겹치는 부분의 모든 레코드를 포함한다. 따라서 이 블록은 NACE 코드가 알려지지 않은 레코드를 제외한 모든 SEE 블록의 레코드를 포함한다. 이 블록에 포함되는 변수에는 등록블록의 모든 변수, 산업 및 일자리 수가 있다. 이 블록은 산업, 일자리 수, 등록 정보가 결합된 결과표를 추정하는데 사용된다.

여섯 번째 블록은 LFS 블록(163,741 레코드)이다. 이 블록은 레코드 수가 가장 적지만 변수의 수는 가장 많은 블록이다. LFS의 표본의 크기는 아주 작지만(일반 가구의 전체 모집단의 1%) 2000년, 2001년의 2년간의 자료를 합친 것이다. 이로 인해 소규모 부모집단에 관해 보다 더 신뢰할만한 추정이 가능해 진다. LFS 블록은 등록부에 있는 모든 변수 및 경제활동상태, 경제활동, 교육정도, 직업 변수를 포함하며 교육정도,

직업, 등록정보, 경제활동 상태, 경제활동과 관련된 결과표를 생성하는데 사용된다.

LFS 블록과 SEE 블록은 서로 겹치는 부분이 없으며, LFS 블록과 NACE 블록도 서로 겹치는 부분이 없다. 이는 불필요한 부분이다. 왜냐하면 LFS(교육정도, 직업) 변수와 NACE(산업) 블록의 변수 또는 LFS 및 SEE(일자리 수) 블록의 변수들을 결합한 결과표는 없기 때문이다.

라. 초기 가중치의 결정

위에서 설명된 각 데이터블록에 대해서, 반복가중기법 적용을 위한 일련의 초기 가중치들이 결정되어야 한다. 이것은 각 블록에 대해서 각각 분리되어 적용되어야 한다. 초기 가중치들은 세가지 단계를 통해서 결정된다.

우선 각 데이터 블록에 대해서 일련의 초기 가중치들이 결정되어야 한다. 센서스에서는 두가지 형태의 블록이 있다. 첫 번째 형태는 위에서 설명된 LFS 블록 등의 표본조사의 표본 레코드를 포함한다. 이 경우에 표본조사의 공표 가중치가 초기 가중치로 사용된다. 두 번째 형태의 블록은 등록과 표본조사를 합친 블록이다. 이러한 블록의 예로서 EcAct 블록을 들 수 있다. EcAct 블록은 등록부 상의 모든 취업자 및 은퇴자들을 포함하며 LFS 상의 모든 비취업자 및 비은퇴자도 포함한다. 두 번째 형태의 블록에 대해서, 표본조사와 관련된 모든 레코드는 초기 가중치로서 공표가중치를 가지며 나머지 다른 레코드들은 가중치 1을 가진다. 두가지 형태의 블록 모두에 있어서 공표 가중치를 사용함으로써 표본설계와 무응답에 대한 보정이 이루어진 것으로 가정한다. 더욱이 분산을 줄이기 위해 이러한 공표가중치의 결정에 있어 보조 정보가 사용되었다고 가정한다.

두 번째 단계로, 블록 가중치를 부여하기 위해 초기가중치들이 조정된다. 이는 2001년 1월 1일 현재의 인구수 합계와 일치하도록 가중치를 조정한다는 것을 의미한다. 이것이 자동적으로 이루어지는 것은 아니다. 왜냐하면 블록이 두개 이상의 표본조사로 구성될 수 있기 때문이

다. 예를 들자면, LFS 블록은 2000년과 2001년의 LFS로 구성되어 있으며, 따라서 가중치의 합은 약 2001년 1월 1일 현재의 인구수의 약 2배가 된다. 블록이 한 개의 표본조사의 모든 레코드로 구성된다 하더라도 조사기준 시점의 차이로 인해 공표 가중치의 합은 전체 인구수와 다를 수 있다.

마지막 단계로, 블록 가중치는 몇 개의 모집단 합계치들에 조정될 수 있다. 이 마지막 조정의 결과인 가중치가 반복가중기법의 초기 가중치가 된다. 이 마지막 조정에 있어서 반복가중기법을 사용하여 추정할 결과표가 고려된다. 초기가중치는 가능한 한 많은 결과표의 추정에 있어서 더 이상의 조정이 없이 사용할 수 있도록 결정된다.

예시의 목적으로 15-74세의 사람들에게 대해 VRD 세션의 LFS 블록에 대해 보다 자세하게 이러한 단계들은 설명하겠다. 이 블록은 2000년과 2001년 LFS의 모든 레코드를 포함하고 있다. 각 레코드는 2000년과 2001년의 LFS의 공표 가중치를 가지고 있으며 이 공표가중치가 초기 블록 가중치로서 사용된다.

다음으로 초기 가중치들은 블록 가중치로 바뀌어진다. LFS 가중치의 합은 인구수의 2배가 된다. N 을 2001년 1월 1일 현재 전체 인구수라고 하자. n_{00} 와 n_{01} 을 각각 2000년 LFS와 2001년 LFS의 표본크기(보다 정확히는 응답수)라고 하고, $w_{00,i}$ $i=1, \dots, n_{00}$ 와 $w_{01,i}$ $i=1, \dots, n_{01}$ 을 각각 2000년 LFS와 2001년 LFS의 공표가중치라고 하자. 2000년 LFS의 레

코드에 대한 가중치는 $\frac{N}{\sum_{i=1}^{n_{00}} w_{00,i}} \times \frac{n_{00}}{n_{00} + n_{01}}$ 를 곱하여 구할 수 있으며, 같은 방식으로 2001년 LFS에 대해서는 $\frac{N}{\sum_{i=1}^{n_{01}} w_{01,i}} \times \frac{n_{01}}{n_{00} + n_{01}}$ 를 구할 수 있다.

각 해의 첫 번째 요소는 각해의 LFS의 가중치의 합이 2001년 1월 1일 현재의 인구수와 같게 되는 것을 보장한다는 점을 주목하기 바란다. 두 번째 요소는 각 해의 표본크기(응답수)를 고려하여 2000년과 2001년의 가중치의 평균을 구한다. 15-74세 인구에 대한 VRD세션에서 이것은 2000년의 LFS 가중치에 0.56을 곱하고, 2001년의 LFS 가중치에 0.52를

곱한다는 것을 의미한다.

마지막으로 LFS의 블록가중치는 반복가중기법의 초기가중치를 구축하기 위해 인구변수에 조정될 수 있다.