

롤링 센서스 조사방법 연구를 위한  
해외출장 결과보고

2008. 5.

통계개발원

# 차 례

I 출장개요 .....	1
II 일정 .....	2
III 회의결과(요약) .....	4
IV 시사점 .....	13
[붙임 1] 네덜란드, 프랑스 관련 인사 인적사항 .....	14
[붙임 2] 네덜란드 가상센서스 관련 주요 질의응답 내용 ..	15
[붙임 3] 네덜란드 가상센서스 .....	17
[붙임 4] 프랑스 순환센서스 관련 주요 질의응답 내용 .....	35
[붙임 5] 프랑스 순환센서스 .....	39
[붙임 6] 네덜란드 통계청 제공 가상센서스 참고자료 .....	49
[붙임 7] 프랑스 통계청 제공 순환센서스 참고자료 .....	50

# 롤링센서스 조사방법 연구를 위한 해외출장 결과보고

## I. 출장개요

### ○ 출장목적

- 인구센서스 방법론 개선 연구(롤링센서스)와 관련하여
  - 본 연구의 핵심인 방법론 개선 검토를 위한 선진국의 선행 연구 결과 및 노하우 확보
  - 해당국의 실시 경험을 기초로 기획 및 실행단계에서의 시행착오, 문제점 및 해결방안 등 제도도입 시 선결사항 사전 검토

### ○ 출장자

- 사회통계실: 통계사무관 양경진
- 연구기획실: 통계주사 김황대

### ○ 방문기관:

- 네덜란드 통계청(CBS)
- 프랑스 통계청(INSEE)

### ○ 출장기간: 2008.4.21.~4.30.(10일간)

## II. 일정

### < 1일차: 08.4.21(월) >

- 인천 → 프랑크푸르트(Frankfurt) → 암스텔담(Amsterdam)
- 암스텔담에서 헤이그(Den Haag, 통계청 소재지)로 이동

### < 2일차: 08.4.22(화) >

- 네덜란드 통계청(CBS, Den Haag) 방문  
·회의주제: Registers and surveys for a Virtual Census
- 네덜란드 관계자: Mr. Eric Schulte Northolt,  
Mr. Frank Linder

### < 3일차: 08.4.23(수) >

- 네덜란드 통계청(CBS, Den Haag) 방문  
·회의주제: Registers and surveys for a Virtual Census,  
Repeated weighting in the context of the Virtual  
Census
- 네덜란드 관계자: Mr. Eric Schulte Northolt,  
Dr. Paul Knottnerus,  
Mr. Paul van der Laan

### < 4일차: 08.4.24(목) >

- 네덜란드 헤이그에서 프랑스 파리로 이동

### < 5일차: 08.4.25(금) >

- 프랑스 통계청(INSEE, Paris) 방문  
·회의주제: 프랑스 INSEE 소개 및 롤링센서스 개요
- 프랑스 관계자: Mr. Olivier Guillemin

< 6-7일차: 08.4.26(토)-27(일): 휴일 >

- 질의사항 점검
- 현지 탐방

< 8일차: 08.4.28(월) >

- 프랑스 통계청(INSEE, Paris) 방문  
·회의주제: French Rolling Census
- 프랑스 관계자: Mr. Olivier Lefebvre

< 9-10일차: 08.4.29(화)-30(수) >

- 귀국: 프랑스 파리 → 인천

### Ⅲ. 회의 결과(요약)

#### 1. 네덜란드: 가상 센서스(Virtual Census)

##### 가. 개요

- 1994년 전산화된 인구등록시스템이 도입되어 인구등록 자료와 표본조사 자료의 연결 가능성 확보
- 네덜란드 통계청에서는 인구등록시스템을 기반으로 하여 사회통계데이터베이스(Social Statistics Database; SSD) 구축
- 이렇게 구축된 사회통계데이터베이스와 기존의 표본조사 자료를 이용하여 2003년 네덜란드 통계청은 2001년 네덜란드 센서스 결과표를 작성하였으며 가상센서스(Virtual Census)라는 이름을 부여
- 전통적인 방식으로 센서스를 실시하였다면 약 3억 유로의 비용이 들었겠지만 가상 센서스의 비용은 3백만 유로에 그침
- 통계청은 반복가중기법(repeated weighting)이라는 새로운 추정 방법을 개발(자료원이 다른 자료를 이용 시 수치적으로 일관성 있는 결과표 생성)

## 나. 사용된 자료

- 행정등록자료: 사회보장, 고용보험, 주택등록, 연금 등
- 표본조사자료: 노동력조사, 고용소득조사, 가계소비조사 등
- 기타 자료: 지리주소 파일

## 다. 작성 방법

- 2001년 네덜란드 센서스의 결과표는 등록 자료와 기존의 표본조사 자료를 이용하여 작성되었으며, 수치적으로 일관성 있는 표의 작성을 위해 반복가중기법(repeated weighing)을 사용
- 네덜란드 통계청은 반복가중기법의 사용을 위해 VRD라는 통계 팩키지를 개발

## 라. 추정 및 자료처리: 반복가중기법(repeated weighting)

- 등록 자료와 표본조사 자료를 통합한 자료를 생성하기 위해 반복가중기법을 사용
- 반복가중기법은 각 표본조사에서 두개 이상의 가중치를 사용하여 각 추정치들이 서로 일치하도록 조정하고 회귀방정식을 반복하여 이용
- 반복가중기법 기본 절차
  - (1) 1단계: 등록부나 표본조사에서 관련정보 추출하여 data

block에 삽입

(2) 2단계: 각 data block에 일련의 가중치를 부여

(3) 3단계: 각 data block을 이용해 수치적으로 일관성 있는 결과표 생성

## 마. 자료 공표

□ 2001년 네덜란드 가상 센서스에서는 40개의 결과표를 생성

- 전국 단위: 28개
- 광역시 단위: 9개
- 기초자치단체 단위: 3개

□ 40개의 표를 분류별로 보면

- 8개 결과표: 주택관련 결과표
- 2개 결과표: 통근학관련 결과표
- 30개 결과표: 인구관련(직업, 교육수준 및 경제활동) 결과표

## 바. 전통적인 센서스와 가상센서스의 차이점

□ 전통적인 방식으로 센서스를 실시하였다면 약 3억 유로의 비용이 들었겠지만 가상 센서스의 비용은 3백만 유로에 그침

□ 전통적인 센서스가 의무적인 것이라고 할지라도 무응답(단위 무응답 및 항목 무응답)의 문제가 발생

- weighting 및 imputation을 통해 신뢰성 있는 결과를 공표하기는 부족
- 가상센서스에서는 소규모 표본조사에서만 무응답 문제 발생

□ 2001년 네덜란드의 가상 센서스는 전통적인 센서스를 실시한 다



른 나라들보다 늦게 시작했음에도 2000 라운드 센서스에 참여한 대부분의 나라들보다 더 일찍 40개의 결과표를 완성

- 일부 항목에 대해 표본조사 자료를 이용하기 때문에 상세한 자료 생산하는 것은 불충분

#### 사. 2001년 네덜란드 센서스 계획

- 센서스를 위한 별도의 표본조사 계획 없음
- 소지역통계 추정기법과 반복가중기법을 연결하여 센서스 결과표 작성 계획

## 2. 프랑스: 순환센서스(Rolling Census)

### 가. 개요

- 프랑스는 1999년 센서스 이후 전통적인 인구센서스를 포기하기로 결정하였는데 그 주된 이유는 센서스가 주기가 길어지고, 예산문제 및 지방자치제의 확산으로 인한 시의성 있는 소지역 통계에 대한 수요가 늘어났기 때문이다. 이러한 문제에 대응하기 위하여 프랑스의 순환센서스 탄생
- 순환센서스 연혁
  - ~1999년: 전통적인 센서스 실시(1939년 이후 부정기적으로 실시)
  - 2004~2007년: 1년 단위로 순환센서스 실시
  - 2008년 현재: 5년 추정치 작업을 진행 중

### 나. 표본 틀(Sampling frame)

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

## 다. 표본설계 및 표본선정

- 인구 만 명 미만의 모든 기초자치단체에 대해서는 센서스 실시 (해마다 기초자치단체 1/5에 대해서 센서스 실시)
- 인구 만 명 이상의 900개 기초자치단체에서도 매년 센서스 표본 조사 실시(매년 각 기초자치단체에서 8%의 거처 조사, 5년간 총 40%)
- 5개의 집단은 약 10가지 기준에 의해 균형을 맞추어 형성: 인구 (인구수, 성별, 연령계층), 기타 주거 관련[거처(dwelling) 수, 주거주지(principal residence) 수]

## 라. 조사 방법

- 센서스 표본조사는 매년 동일한 날에 관련된 모든 기초자치단체에서 시행(1월 3번째 목요일)
- 자료수집 방법(protocol)은 조사원이 질문지를 전달하고 수거하는 방식 (표집방법은 혁신적으로 변화했지만 자료수집의 경우는 일반적인 센서스에서 사용되는 방식을 그대로 따름)
  - 가구에 관한 질문: 가구원의 목록, 주거의 특징, 기준, 자동차 보유 대수 등 약 14개 문항
  - 가구원에 관한 질문: 나이, 성별, 출생지, 국적, 5년 전 거주지, 교육정도, 직업, 근무지 또는 통학지 등 25개 문항

- 조사가 불가능한 경우(부재, 응답거부)에는 주 거주지에 대해서 조사원이 미조사 거처조사표에 추정 가구원수를 기입하여 인구수 조정에 이용

#### 마. 자료처리

- 매년 조사된 9백만 개인 및 4백5십만 거처가 가중치 조정 후에 파일에 저장(세분된 파일은 모든 개인을 설명), 1년간 수집된 자료를 누적하여 당해 연도의 추정치를 계산
- 조사의 포괄률(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 의미
- 누적된 파일은 4천5백만 개인과 2천2백만 거처로 구성되어 있기 때문에 상세한 표 작성 가능

#### 바. 추정(Estimation)

- 대규모 기초자치단체의 경우에는 5년간 표본의 평균값을 계산한 후 중간연도의 RIL의 거처 수에 맞추어 판단
  - 5년간의 표본에 기초한 순환평균(rolling average) 이용, y-4년에서 y년까지의 5개 표본의 자료에서 거처당 평균 인구수 계산

- 이 평균 인구수가 중간기간(y-2년)을 대표(이 평균 인구수에 y-2년 1월 1일 현재의 거처수(RIL을 이용)를 곱하여 기초자치단체의 인구 산정)

## 사. 자료 공표

- 매년 각 기초자치단체에 대한 상세한 정보를 제공할 것이며 대규모 기초자치단체의 경우에는 통계 목적으로 합쳐진 대략 2천 명의 거주자로 구성된 IRIS(센서스 blocks)에 대한 상세한 정보를 제공
- 2004년 후반기에 약 100여 개의 대규모 기초자치단체에 대한 인구추정치도 제공(이러한 결과들은 인터넷을 통해 2005년 센서스 시작 이틀 전에 발표)
- 매우 효과적인 전국 및 지역적인 홍보를 통해 일반 국민들이 대규모 센서스에서보다 순환센서스에 대한 거부감 축소

## 아. 전통적인 센서스와 순환센서스의 차이점

- 표본 추출률: 순환표본조사의 5년간 누적표집률(40%)은 매우 높기 때문에 정확한 결과를 보증(이와는 대조적으로 전통적 센서스에서는 대개 약 15% 또는 25%의 표본을 사용)
- 연간 조사화: 표본의 사용 및 자료수집절차의 연간화는 INSSE, 각 기초자치단체, 조사원들에게 도움을 주어 양질의 자료수집 가능

- 비용의 분산: 순환 센서스는 또한 일을 분산함으로써 비용을 분산하고 안정적인 노동력 및 예산으로 운영하는 것이 가능
- 자료 시의성: 소지역에 대해서 매년 수정된 정보가 노후화된 정보보다 더 선호(순환 센서스는 10년 주기의 전통적 센서스보다 도시재개발 정책수행 등에 훨씬 더 유용)

## 자. 순환센서스의 성공요인

- (1) 기술적인 요인
  - 양질의 표본 틀 확보(기초 자치단체 하위 수준의 지리정보시스템의 사용가능성, 보완에 필요한 행정자료의 사용 가능성)
- (2) 방법론적인 요인
  - 표본설계 기법
- (3) 정치적인 요인
  - 센서스 동반자로서의 기초자치단체와의 광범위한 협의
  - 센서스 사용자들과의 협의

## IV 시사점

### 1. 네덜란드 가상 센서스

- 비용 절감
- 이용 가능한 모든 자료를 사용함으로써 기존 자료의 활용도 제고
- 조사환경의 악화에 대비한 대안
  - 조사거부 및 단독가구의 증가
  - 응답부담 경감, 통계청(지방청) 업무 부담 경감

### 2. 프랑스 순환센서스

- 1년 단위로 센서스 자료를 생산함으로써 시의성 제고
- 상세한 소지역 통계 작성 가능
- 센서스 비용의 분산

[붙임 1] 네덜란드, 프랑스 관련 인사 인적사항

국 가	성 명	인 적 사 항
네덜란드 (CBS)	Mr. Eric Schulte Northolt	Statistics Netherlands Division of Social and Spacial Statistics Development and support department Senior researcher and project leader <a href="mailto:else@cbs.nl">else@cbs.nl</a> Phone: 31(0)70 337 4931 Fax: 31(0)70 387 7429
	Mr. Frank Linder	
	Dr. Paul Knottnerus	Centraal Bureau voor de Statistiek Divisie Research en Ontwikkeling Sector Statistische methoden manager surveymethoden <a href="mailto:pkts@cbs.nl">pkts@cbs.nl</a> Phone: 31 (0)70 337 4957 Fax: 31 (0)70 337 5990
	Mr. Paul van der Laan	Statistics Netherlands Division of Social and Spacial Statistics Policy Staff Senior policy adviser <a href="mailto:plan@cbs.nl">plan@cbs.nl</a> Phone: 31 (0)70 337 5715 Fax: 31 (0)70 337 5982
프랑스 (INSEE)	Mr. Olivier Guillemain	INSEE, Director European and international affairs department <a href="mailto:DG75-Eu-Int@insee.fr">DG75-Eu-Int@insee.fr</a> Phone: 33(0)1 4117 5588
	Mr. Olivier Lefebvre	INSEE Chef du departement de la demographie <a href="mailto:olivier.lefebvre@insee.fr">olivier.lefebvre@insee.fr</a> Phone: 33(0)1 4117 5264 Fax: 33(0)1 4117 6741



## [붙임 2] 네덜란드 가상센서스 관련 주요 질의응답 내용

Q. 가상센서스란 무엇인가?

A. 행정등록 자료와 기존의 표본조사 자료를 이용하여 센서스를 실시하지 않고 EU 등의 국제기구가 요구하는 센서스 결과표를 통계적 기법을 사용하여 작성하는 것을 말한다.

Q. 가상 센서스의 방법론에 관한 연구기간은?

A. 1992~1995년

Q. 2001년 가상센서스에 투입된 인원 및 시간은?

A. 투입 인원: 10명  
투입 시간: 6,000시간

Q. 가상센서스를 위해 기존의 표본조사에 추가한 항목은 있는가?

A. 전혀 없다.

Q. 가상 센서스를 매년 실시하지 않은 이유는?

A. 1. 사회구조의 변화가 매년 심하지 않기 때문  
2. 매년 실시하면 비용이 많이 소요됨

Q. 2010 라운드 센서스에 참가 예정인 EU 국가의 수는?

A. 27개국

Q. 2010 라운드 센서스에 등록자료만 사용할 국가는?

A. 노르웨이, 핀란드, 스웨덴, 덴마크, 슬로비니아

Q. 등록부를 이용하여 네덜란드의 인구수 추정 시 특이점은?

A. EU 전체로 인구수를 파악하기 때문에 어느 나라에서 등록을 했는지에 따라 인구수가 달라짐

Q. hypercube의 셀 수는?

A. 40개의 테이블을 작성하는데 2백만 셀이 필요함  
110개의 테이블 작성 시 2천만 셀이 필요함

Q. 실제 가구와 등록상의 가구의 구분은 어떻게 하는가?

A. 등록상의 가구만 고려한다.

Q. 주소지와 실제 사는 곳이 다른 경우에 대비책은 무엇인가?

A. 주소지와 실제 사는 곳이 동일하도록 세금 혜택 등의 인센티브를 부여한다.

## [붙임 3] 네덜란드 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 1991년 센서스 라운드와 비해 상세한 정보를 요구하였으며, 센서스 결과표의 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 기반을 두고 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, 비경제 활동 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 연기금에 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 사회보장 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 정

보는 “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 때문 이리라. 이러한 자영업자는 2001년 센서스 결과표에 포함되지 않았다.

: 2000년 말 현재 7십9만 레코드

#### 다) 통합수혜파일

조기 은퇴한 사람들은 “통합수혜 파일”에 있는 생명보험이나 연금 자료를 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. 실제적으로 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 3개 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 대체법(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 사용되는 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 결과

표를 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 센서스에서는 충분히 추정할 수 없는 셀 값을 추정하기 위해서 소지역 추정법을 적용할 계획이다.

## 6) 반복가중기법(repeated weighting)

### 가) 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 결과표 생성한다.

### 나) 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.

## (1) 반복가중기법 순서

### (가) 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 수준의 클래스 변수는 계층적(hierarchical)이어야 한다. 즉 상위 단계에서의 표는 하위 단계의 결과표들을 합한 결과와 일치하여야 한다.

### (나) 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 구성요

소에 관한 기록이 있고, 1개의 표본조사에 기초를 둔 블록은 표본조사와 등록 자료를 연결한 블록이 된다. 2개 이상의 표본조사로 구성된 보다 복잡한 블록도 가능하다.

#### (다) 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 분산을 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 결과표 추정

둘째, 각 결과표의 모든 마진을 우선 추정. (각 마진은 결과표보다 더 큰 데이터 블록을 이용하여 추정 가능함)  
기 추정된 마진이 고려되어야 함

셋째, 가능하면 초기 가중치를 이용하여 결과표를 추정하여야 함

만약 초기 가중치를 이용한 추정치가 기 추정된 결과표와 일치하지 않는 경우에는 반복가중기법을 사용함

반복가중기법 사용 시 우선 결과의 어떤 마진이 기 추정된 결과표와 중복되는지를 검사하여야 한다. 이러한 공통 마진의 모수 합계치가 알려진 모수 합계치로 사용되며 회귀추정치에 조정에 대한 목표치가 된다. 실제로

조정(calibration)이라는 말은 초기 가중치를 조정하여 알려진 마진과 일치 되어야 한다는 제약조건을 충족시키는 것을 의미한다. 조정의 범위는 결과 표 및 기 추정된 결과표에 따라 달라진다. 원칙적으로 모든 결과표는 서로 다른 가중치를 사용하여 추정될 수 있음.

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 변수 간의 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 구체화해야 할 몇 가지 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

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

표1.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 사람에 관한 클래스 변수가 고려되고, 두 번째 부분은 가구에 관한 변수로 구성된다. 자연수인 가구수 변수 외에 가구의 크기라는 양적변수가 존재한다. 표2에 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 등록부이며 센서스의 핵심부분이며, LFS와 SEE는 표본조사이고 SSD는 고용인, 고용주 및 연금 등의 정보를 포함하는 등록부들의 결합이다. 일부 변수들은 등록부와 표본조사를 결합한 정보에 기초를 두었다. 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 1> 가상센서스에 사용된 변수

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’	양적 변수	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 경제 상태	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 나머지 한 세션은 개인 가구(private households)에 관한 것이다. 후자의 경우는 집단 가구에 관한 자료가 충분하지 못하기 때문에 필요하며, 결과표 작성에는 이 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.

## 라) 데이터 블록의 유도

위에서 설명하였듯이 반복가중기법은 직사각형의 완전히 채워진 마이크로 데이터블록을 사용한다. 여기서는 예로서 15-74세의 사람들에게 대한 데이터블록을 나타낸다. 이 VRD 세션은 두드러진 모든 데이터블록들을 사용하는 가장 광대한 것이다. 이 데이터블록들은 그림1에 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 구성하



였다. 이 데이터블록들을 크기의 순서에 따라 소개하겠다.

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. 표2에 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.

두 번째 데이터블록은 NACE<sup>1)</sup> 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 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.

---

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

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 레코드수가 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 신뢰할 만한 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.

#### 마) 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 때문이다. 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 사람들에 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  $w_{01,i}$ ,  $i=1, \dots, n_{01}$ 를 각각 2000년 LFS와 2001년 LFS의 공표가중치라고 하자. 2000년 LFS의 레코드에 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.

## [붙임 4] 프랑스 순환센서스 관련 주요 질의응답 내용

### < 건물등록부(RIL) 편집 >

Q. 1999년 센서스를 기초로 건물등록부(RIL; 표본틀)를 편집할 때 투입된 인원은?

A. 통계청 본부(센서스과) 직원 30명과 지역사무소 인원 15명이 투입되었다.

Q. 건물등록부의 편집을 완성하는데 걸린 기간은?

A. 3년간(2001~2003) 걸렸다. 2003년 중간에 표본조사를 위해 완료되었다.

### < 건물등록부 보완 및 점검 >

Q. 건물등록부 보완을 위한 조직이 있는가?

A. 24개의 지역사무소 직원이 건물등록부 보완 작업을 한다. 지역사무소의 담당 직원 수는 2~20명이고 전체로는 120명이다.

Q. 건물등록부 보완에 이용되는 행정자료는?

A. 지방세 파일, 건축허가, 우편국의 정보를 이용한다. 장래에는 사회보장파일과 건물철거명부도 이용할 계획이다.

Q. 건물등록부의 보완은 어떻게 하는가?

A. 담당 직원들이 사무실에서 GIS와 Register를 비교하여 수작업으로 한다.

Q. 건물등록부의 점검은 어떻게 하나?

A. 각 기초자치단체에 건물등록부를 송부하여 철거, 신규 등의 확인 작업을 한다. 각 기초자치단체에서 이 작업에 참가하는 인원 0~2명이다.

Q. 건물등록부의 오차율은 얼마나 되나?

A. 2003년에는 1.8~2% 차이가 있었지만, 2007년에는 1% 정도의 차이가 있다. 건물등록부의 오차율이 3~4%이면 괜찮다고 여겨진다.

Q. 건물등록부의 질은 어떤가?

A. 매년 80만 명이 출생하고 50만 명이 사망한다. 그러나 센서스 첫 해에는 60만 명 이상의 인구가 증가하다가 지금은 약 40만 명 정도의 인구가 증가하고 수치도 안정적이다. 아마 등록부의 질이 초기보다 개선되었기 때문이라고 생각된다.

### < 표집 및 추정 >

Q. 전수조사와 표본조사의 경계선인 10,000명 인구의 기준설정은?

A. 표본오차를 고려하여 최소 인구수를 10,000명으로 정하였다. 기초자치단체의 50%가 인구 만 명 미만이다.

Q. 무응답에 대해 대체(imputation)을 한다면 대체 방법은 무엇인가?

A. 조사 대상가구가 부재일 경우에는 특별한 조사표(미조사 거처 일람표)를 사용하여 이웃을 통해 사람 수만 조사하여 imputation한다. 핫덱(hot-deck) 방법을 사용한다.

Q. 프랑스에서의 공식통계는 1년 추정치, 3년 추정치, 5년 추정치 중 어떤 것을 사용하냐.

A. 5년 순환평균치를 공식통계로 사용한다.

Q. 표본추출단위가 건물이라면 가구와 관련된 모수의 추정은 어떻게 한다.

A. 조사 후에 거쳐 당 평균 가구 수를 구하여 추정한다.

## < 기타 >

Q. 순환센서스에 참여하는 본부 직원의 수는?

A. - 조사(자료 수집 및 행정) 관련 8명,  
- 자료처리, 추정, 데이터베이스관리, imputation(대체), checking, 표집 관련 9명,  
- 등록부, GIS 관련 8명,  
- 학생, 연구자 등의 고객의 교육 담당 전문가 1명  
- 2011년에 실시 예정인 특별조사(센서스 항목에 없는) 담당 전문가 1명  
- 그리고 비서가 2명 있다.

Q. 순환센서스의 조사인력은 얼마인가?

A. 전체 조사원의 규모는 18,000명이며 각 기초자치단체에서 채용하고 관리한다. 500명의 INSEE 직원이 supervisor의 역할을 하고 50명의 supervisor를 추가로 채용했다.

Q. 사회보장 파일에는 어떠한 내용이 들어있다.

A. 성별, 나이, 생일 및 주소가 들어있다. 이사할 경우에 신고가 의무 사항이 아니다.

Q. 조사항목을 변경할 때 고려 사항은 무엇인가?

A. 현재까지는 조사항목을 변경한 적이 없다. 2010년에 조사항목을 변경할 예정인데 매우 어려운 작업이다.

Q. 센서스에서 개인식별번호를 직접 또는 간접적으로 사용하는가?

A. 법적으로 사용 불가능하다.

Q. 전통적 센서스와 순환센서스의 비용은?

A. 전통적 센서스의 비용은 2억 유로가 들지만 순환센서의 비용은 연간 2천9백만 유로가 소요된다. 7년 주기의 센서스 비용과 비교하면 거의 동일하다.

Q. INSEE의 직원수는 몇 명인가?

A. 약 6천명이 근무하고 있다.



## [붙임 5] 프랑스 순환센서스

### 가. 프랑스 센서스의 역사

프랑스에서는 나폴레옹의 동생 루시안에 의해 1801년 처음으로 인구센서스가 실시되었다. 나폴레옹이 인구센서스를 실시한 주된 목적은 두 가지다. 첫째는 군사목적에 의한 것으로 전쟁을 위해 청년을 징병할 필요가 있어 각 지역의 청년수를 파악하는 것이고, 또 하나의 목적은 원로원 및 입법원의 의원 수는 나폴레옹헌법에 의해 지역의 인구수에 맞추어 정하도록 했기 때문에 각 지역별 인구수를 확정할 필요가 있었다. 인구센서스는 이 후 5년마다 실시되었다. 그러나 1939년의 2차 세계대전 발발 후 인구센서스는 부정기적으로 1946, 1954, 1962, 1975, 1982, 1990, 1999년에 실시되었다. 프랑스는 1999년 센서스 이후 전통적인 인구센서스를 포기하기로 결정하였는데 그 주된 이유는 센서스가 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 순환 센서스

#### 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. 현재까지는 프랑스가 전 세계에서 유일하게 순환센서스를 실시하는 나라이

다.

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.

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 되었다: 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 대규모<sup>2)</sup> 기초자치단체의 표본틀은 관련된 건물등록부(RIL; located building register)이며, 건물등록부는 주거용, 행정용, 산업용, 상업용 건물 등의 모든 건물의 목록이며 지리정보시스템을 이용하여 확인 가능하다.

---

2) 인구 만 명 이상의 자치단체를 말한다.

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 만 명 미만의 35,750개의 기초자치단체는 5개의 집단으로 나누어진다. 이 기초자치단체들은 전체 인구의 절반을 차지한다. 매년 한 개의 집단에 속하는 모든 기초자치단체들에 대해서 센서스를 실시한다. 첫 번째 집단에 속하는 기초자치단체들은 5년 후에 다시 조사된다.

5개의 집단은 약 10가지 기준에 의해 균형을 맞추어 형성된다. 인구(인구수, 성별, 연령계층), 기타 주거 관련[거처(dwelling) 수, 주 거주지(principal residence) 수]. 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.<sup>3)</sup> 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.

#### 4) 조사방법

센서스 표본조사는 매년 동일한 날에 관련된 모든 기초자치단체에서 시행된다(1월 3번째 목요일에 시작함).

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

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

다른 하나는 가구원에 관한 질문지이다. 질문지는 과거의 센서스에 사용된 질문지와 큰 차이가 나지 않는다. 가구에 관한 질문지는 가구원의 목록과 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 포괄률(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.

## 6) 추정

매년 수집된 정보는 5년 주기의 중간점에서의 특정한 날에 적용하기 위해서 조정된다. 대규모 기초자치단체의 경우에는 5년간 표본의 평균값을 계산한 후 중간년도의 RIL의 거처 수에 맞추어 판단한다. 소규모 기초자치

단체에 대해 사용되는 방법은 센서스 표본조사와 인구 기준 일간의 내삽 또는 외삽이다. 외삽은 주택세 등록자료를 이용하여 합쳐진다. 주택세 등록 자료는 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 거쳐당 평균 인구수가 계산된다. 이 평균 인구수가 중간기간( $y-2$ 년)을 대표한다. 이 평균 인구수에  $y-2$ 년 1월 1일 현재의 거쳐수(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 2천명의 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 하반기 이후로 약 100여 개의 대규모 기초자치단체에 대한 인구추정치도 제공할 수 있게 되었다. 이러한 결과들은 인터넷을 통해 2005년 센서스 시작 이틀 전에 발표되었다. 2005년 하반기에는 2006년에 조사된 인구 만 명 미만의 기초자치단체 및 또 다른 약 100여 개의 대규모 기초자치단체로 확대되었다. 그 결과는 2006년 1월 17일에 공표되었다. 이런 식으로 초기 성과물들을 발표함으로써 새로운 방식의 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. 인구 만 명 이상 기초자치단체의 조사원들은 조사해야 할 주소 목록을 가지고 있다. 이

것은 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 정책수행에 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 현장상태에 대한 지방의 이해가 훨씬 더 높기 때문에 이것은 매우 중요한 일이다. 행정자료가 없는 경우에는 프랑스의 해외에 있는 도(DOM; France's



overseas department)의 경우처럼 공무원들을 현장에 보내서 주택을 확인하는 것이 필요하기 때문에 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 통계적 확률표본(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 시스템

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

[붙임 6] 네덜란드 통계청 제공 가상센서스 관련 참고자료  
(별도 첨부)

1. 네덜란드 통계청 조직도
2. Power Point 설명자료
  - 가. Social Statistical Database: It's principles
  - 나. Educational attainment from administrative registers and surveys
  - 다. The use of registers and the alignment of tables estimated from different surveys(repeated weighing 기법 포함)
3. Statistical yearbook of the Netherlands 2007
4. Prins, C.J.M.(2000), "Dutch population statistics based on population register data"
5. Statistics Netherlands(2000), "Netherlands Official Statistics", Volume 15, Summer 2000
6. Statistics Netherlands, "The Dutch Virtual Census of 2001: A new approach by combining different sources"
7. Statistics Netherlands(2004), "The Dutch Virtual Census of 2001: Analysis and Methodology"
8. Nordholt, Eric Schulte, "The Dutch Virtual Census of 2001: A register-based approach combined with survey information"
9. Nordholt, Eric Schulte(2005), "The Dutch Virtual Census of 2001: A new approach by combining diferent sources"
10. Nordholt, Eric Schulte et. al.(2007), "Record matching for Census purposes in the Netherlands"

[붙임 7] 프랑스 통계청 제공 순환센서스 관련 참고자료  
(별도 첨부)

1. 순환센서스 조사표류: 4종
2. Power Point 설명자료
  - 가. The quality of the population census in France(1)
  - 나. The quality of the population census in France(2)
3. 프랑스 통계청 주요인사 전화번호: 내부자료
4. Dumais, Jean et al.(1999), "An alternative to traditional census taking: plans for France"
5. INSEE(1999), "Post 2000 Census in INSEE"
6. Desplanques(2000), "The new census in France"
7. Durr, Jean-Michel et al.(2001), "Redesign of the French census of population"
8. INSEE(2003), "The new French population census"
9. INSEE(2006), "The population Census in France: from general census to rolling census?"
10. INSEE(2007), "Quality management in the French census"
11. INSEE, "The Population Census from 2004 on"
12. Durr, Jean-Michel, "The French new rolling census"