

시계열자료 구축을 위한 인구센서스의 일관성 검토

김 가 은, 전 광 희†

충남대학교 사회학과

우리나라의 인구센서스는 인구의 규모, 구조, 분포와 관련 특성을 파악하는 국가 기본통계조사로서, 통계청은 1960년부터 2015년까지의 센서스 자료를 2% 마이크로데이터로 제공하여, 모든 이용자에게 서비스를 한다. 이 연구는 이용자 입장에서 유엔의 공식통계 원칙 중 하나인 일관성 측면에서 센서스 자료들을 검토하여 자료의 활용성을 개선하고자 한다. 평가의 관점은 마이크로데이터의 일관성, 항목의 반복성, 개념과 조사방법, 접근법, 분류, 파생변수, 인구 및 가구주택 항목의 연계방법 등이다. 검토의 결과, 통계청이 제공하는 2% 마이크로데이터의 일부(1975년 및 1980년)는 상세한 항목이 있는 표본조사가 아니라 전수조사 자료를 기반으로 제작되었으며, 항목별 조사주기의 전략 강화, 등록센서스 전환에 따른 시계열 불일치의 문제, 경제활동 항목의 불완전성, 항목 응답 분류의 잦은 변경, 파생변수의 관리 미흡, 일부 연도의 인구와 가구·주택 간 자료세트의 통합 불편 등으로 자료의 일관성/정합성에 문제가 발생하여 활용성을 떨어뜨리는 것으로 파악되어 지속적인 개선이 필요하다고 본다.

주요어: 인구센서스 2% 개별자료, 인구센서스 일관성, 인구센서스 정합성

† 교신저자(Corresponding Author) : 전광희, 충남대학교 사회학과 교수, 대전광역시 유성구 대학로 99, E-mail : jkh96@cnu.ac.kr

김가은, 충남대학교 사회학과 박사수료(제1저자)

■ 최초투고일 : 2020년 5월 11일 ■ 심사마감일 : 2020년 7월 7일 ■ 게재확정일 : 2020년 7월 16일

1. 서론

우리나라의 인구주택총조사, 곧 센서스는 1925년부터 2015년까지 5년마다 주권이 공식적으로 행사되는 전국의 인구와 가구 및 주택을 파악하여 왔다. 특히 2015년에는 20%의 표본조사를 실시하고, 이를 행정자료와 결합하여 센서스 통계를 작성하는 부분적 등록센서스를 도입하고, 2016년부터 2019년에 대해서는 별도의 표본조사를 실시하지 않고, 행정자료를 편집하여 매년 인구, 가구, 주택의 일부 항목에 대한 자료를 제공하고 있다.

인구주택총조사는 한 시점을 기준으로, 전국을 대상으로 누락과 중복이 발생하지 않도록 조사대상을 파악해가기 때문에, 방대한 인력과 예산이 투입되는 국가적인 대규모 프로젝트이다. 유엔의 원칙과 권고에 의하면, 인구센서스는 국가 기본통계조사임을 염두에 두고, 개별성(individual enumeration), 보편성(universality), 동시성(simultaneity), 주기성(periodicity)이라는 기본적 특성을 갖출 것을 요구하고 있다(United Nations, 2017).

본질적으로 인구주택총조사는 전국의 모든 대상을 파악하기 때문에, 시군구는 물론 읍면동보다 작은 소지역에 이르기까지 통계 생산이 가능할 뿐만 아니라, 한 나라에서 공식통계를 작성하기 위한 인구 및 사회조사의 표본 틀을 제공하는 역할도 수행한다. 이러한 센서스의 여러 가지 특성 중에서도 아마 가장 중요하게는 오랜 기간 주기적으로 조사를 실시해 왔기 때문에 우리나라의 근현대 생활의 변화 양상을 살펴보는 데 가장 유용한 수단이라고 할 수 있다. 특히 센서스는 임의적이지 않고 편향이 없이 모든 조사대상을 전부 포함하고 있다는 점에서 하나의 수치로 표현된 객관적인 역사기록으로 간주될 수 있을 것이다.

센서스 자료가 이렇게 가치가 있는 사회변화 측정수단으로 제대로 활용되기 위해서는, 측정하려

는 내용이 당시의 시대 변화를 반영할 수 있도록 유연해야 하는 부분도 있지만, 다른 한편으로는 비교도 가능하도록 내용적으로 일관성을 유지하는 것도 대단히 중요할 것이다. 동일한 내용이라도 개념이나 범위가 바뀌거나 분류 방식이 다르다면 비교가 불가능해진다. 이러한 이유로 유엔에서도 국가통계는 원칙적으로 일관성/정합성(consistency/coherence)의 기준을 충족시켜야 한다고 강조하고 있다(United Nations, 2017).

이 연구에서는 우리나라의 인구센서스 중 최초의 현대적인 센서스로 평가를 받고 있는 1960년 인구센서스 이후 2015년 센서스까지 지난 60여년간 인구센서스의 주요 항목들을 시계열 자료로 구축하면서, 우리나라 인구센서스 자료들이 갖는 비교성과 일관성의 문제를 평가하고 이들 자료가 사회변동의 양상을 연구하는데 어떠한 사회학적 내지는 통계학적인 시사점을 제시하는가를 검토하고자 한다.

2. 연구방법-일관성의 정의를 중심으로

인구주택총조사는 일차적으로 가구단위에서 출발하기 때문에, 이를 관찰단위(observation unit) 또는 조사단위(survey unit)라고 한다. 그런데 인구센서스의 결과를 분석하거나 집계할 때는 개인단위로 이루어진다. 실제로 연령별 인구, 직업별 인구 등은 가구단위가 아닌 가구원단위 또는 개인단위로 집계가 되는데, 이러한 단위를 분석단위 또는 집계단위(tabulation unit)라고 한다. 가장 기본적인 단위는 집계단위로 이것을 기본단위라고 하는데 조사목적에서 정해진다. 특히 조사대상을 분류 집계할 필요성이 있을 때에는 어떤 단위로 분류할 것인지 미리 결정해야 하며, 집계단위

는 이 단위에 맞추어야 한다(박재수, 1989). 인구 센서스의 집계단위는 기본적으로 인구 또는 가구(이들이 함께 거주하는 거처)이기 때문에 조사항목은 크게 인구항목과 가구 및 주택항목으로 구분된다.

인구센서스는 국가기본통계조사로 전수조사이기 때문에 어떤 항목을 조사하는지 그리고 어떤 용어를 사용하며, 그 개념과 정의는 무엇인지가 인구 및 사회학적으로 매우 중요하다. 예를 들어 혼인상태 항목을 질문할 때 동거나 별거를 응답 항목에 포함하여 조사하거나, 미혼자에게도 출생아 수를 질문하여 미혼모 통계를 내거나, 가구주와의 관계에서 편부나 편모, 조손가정 등을 파악하여 통계를 작성하는 것과 같은 행위는 전 국민을 대상으로 조사하는 과정에서 사회 관념(social norm)으로 수용성을 높이는 방향으로 영향을 미칠 개연성이 높을 것으로 저자는 판단하고 있다. 아울러 관행적으로 5년마다 실시하는 인구센서스의 조사항목을 선정할 때는 항상 전문가나 이용자의 의견을 듣고 반영하기 때문에 조사항목을 신설하거나 폐지하거나 항목별 조사주기를 정하는 문제는 어느 정도는 시대별로 그 항목에 대한 수요 즉, 중요성의 변화를 반영하는 것으로 보아도 무방할 것이다.

물론 이 연구에서는 그러한 인구센서스 항목 선정과 용어, 항목별 조사주기에 대해 인구학이나 사회학적 맥락에서의 함의를 직접 다루지는 않을 것이다. 이 연구의 주요 목적은 이용자의 관점에서 우리나라 인구센서스 조사항목의 일관성 및 실용성을 검토하는 것이다. 다만 이러한 시도가 인구학이나 사회학 연구자들의 지식사회학적 탐구를 지원하고 도움을 줄 것으로 기대한다.

이에 따라 이 연구의 대상이 되는 자료는 원칙적

으로, 통계청 마이크로데이터 시스템(MDIS: Micro Data Integrated System)에서 통계이용자에게 무료로 제공하는 1960년부터 2015년까지 12개 인구주택총조사의 2% 자료 세트)로 한정한다.

이 연구에서 일관성(一貫性)은 이 12개의 데이터 세트에 있는 센서스 조사항목이 동일하게 조사되었는지를 의미하며, 이를 다음과 같이 6가지 범주로 구분하여 점검한다.

- 1) 마이크로데이터의 일관성
- 2) 항목의 주기성 또는 반복성
- 3) 개념/정의 및 조사방법, 접근법
- 4) 분류
- 5) 파생변수
- 6) 인구 및 가구주택 항목의 연계방법

먼저 개념/정의에서는 해당 조사항목을 어떠한 접근법으로 작성했는지 검토할 것이다. 예를 들어 인구를 파악하는 방법은 조사원이 대상자를 발견한 장소에서 파악하는 현주지 방식(*de facto* approach)과 대상자를 통상적으로 거주하는 지역에서 파악하는 상주지 방식(*de jure* approach) 방법, 또 등록센서스 전환으로 대상자를 주민등록과 같은 등록지를 기준으로 파악하는 등록지 방식(registration-place approach) 또는 법정거주지 방식이 있다. 접근방식에 따라 인구가 다르게 파악되어 인구분포가 행정구역별로 달라질 수 있다. 경제활동동인구나 집단가구의 집계방법에 대해서도 개념을 검토한다. 포괄범위(coverage)는 조사대상 연령을 위주로 검토한다. 예를 들어 경제활동인구조사는 1960년 처음 도입된 이후 14세를 경제활동 개시연령으로 조사를 하다가 1985년부터는 개시연령을 15세로 변경하였다. 출생아 수를 조사시점 혼인상태에 있는 여성들에게만 질문하였

1) 1960년, 1966년, 1970년, 1975년, 1980년, 1985년, 1990년, 1995년, 2000년, 2005년, 2010년, 2015년

는지 아니면 15세 이상 전체여성을 대상으로 질문을 하였는가에 따라라도 결과가 달라질 것이다.

분류(classification)는 행정구역분류, 직업분류와 산업분류는 물론이고, 학교 급별 분류 등 센서스 간 분류의 차이가 해당된다.

응답(response category)범주는 개방형(사후 코딩방법, post-coding)인지 폐쇄형(사전 코드방법, pre-coding)인지 구분하고, 응답범주의 개수나 종류의 변화는 물론 순서가 일관성이 있는지도 함께 분석한다.

파생변수(derived variable)는 센서스 조사항목에 포함되지 않아서 직접 조사하지는 않았지만, 조사항목과 응답범주를 재구성한 후 집계하여 새로이 생성하는 항목이다. 예를 들면 가구원 수, 세대유형(living arrangement), 세대(世代 generation) 수, 핵(核, nucleus) 수 등이다. 이 연구에서는 센서스 간 파생변수의 종류와 집계방법의 차이가 있는지도 검토한다.

마지막으로 인구 및 가구/주택 항목의 연계방법(linkage method)이다. 인구주택총조사는 가구와 거기에 거주하는 인구를 조사하기 때문에, 한 데이터가 다른 데이터를 포함하는 구조적 자료(structured data)의 형태를 갖고 있다. 따라서 개인의 가구특성을 분석하려면 자료를 다루는 일정 수준의 자료처리 기술이 필요하다. 그런데 이러한 접근법도 센서스마다 다소 다르기 때문에 통계 이용자들에게 어려움을 주고 있다. 이 또한 자료이용의 편의성 및 일관성 측면에서 검토할 필요가 있다고 보았다.

1) 인구주택총조사의 탐색

(1) 1960년 이전

우리나라의 인구주택총조사는 1925년 처음으로

실시되었으며, 이후 예외가 있기는 하지만 5년마다 거의 한번 씩 실시되었다. 일제는 당초 1920년 제국주의 수탈목적으로 제국판도(帝國版圖) 전체에 대한 인구센서스를 기획하고 조선, 대만, 일본에서 최초로 근대적인 센서스를 실시하려고 계획했으나, 한반도에서는 3.1 독립운동의 여파로 국민들이 조사에 비협조적이어서 조사원을 구하지 못해 실시하지 못하고, 5년 후인 1925년 최초로 제1회 간이국세조사를 실시하였다. 이 조사는 근대적인 의미의 최초의 인구센서스로 성명, 성별, 생년월일, 배우관계, 본적(또는 국적) 등 5개의 항목에 대하여 조사를 실시하였다(박명규 외, 2003).

이후 매 5년마다 정기적으로 인구센서스가 실시되었다. 이 조사들의 조사항목 변화를 보면, 1930년에는 1925년의 항목에 4개 항목(가구주와의 관계, 직업, 출생지, 문맹여부)이 추가되었으며, 1935년에는 총 6개로 1925년 항목에 상주지 1개 항목이 추가되었다. 1940년에는 전쟁을 위한 인력수탈 목적의 항목들이 추가되었는데, 지정기능(126종)과 병역관계가 포함되었으며, 결과는 보안상 이유로 공표되지 않았다. 1944년에는 전사인력동원을 위한 자원조사법에 근거해 조사를 실시했는데, 1940년에 조사하던 가구주와의 관계, 출생지를 제외하고, 특수기능과 중사상지위를 신규로 포함하였다.

광복 후에는 1949년에 정부수립 및 1950년 민의원 선거일정(5월)에 따라 1년 앞당겨 실시하였으나, 한국전쟁으로 속보자료를 제외한 모든 자료가 불타버렸다.

(2) 1960년 이후

우리나라의 인구센서스는 1960년을 계기로 현재와 같은 모습을 갖추었기 때문에, 보통 1960년 센서스를 최초의 현대적인 인구센서스로 보고 있

다. 유엔 통계고문관의 기술지도 하에 사후조사를 도입하고, 표본집계(20%) 방식을 최초로 도입하였다. 1949년 및 1955년에 적용한 현주(de facto) 개념 대신 상주(de jure) 개념을 처음으로 도입하였고, 시험조사도 처음으로 실시하였다. 그동안의 센서스가 인구만 조사하였으나 처음으로 주택센서스를 도입하여 함께 조사하였다. 하지만 세는 나이만 조사하여 만 나이 0세가 없는 특징이 있기도 하다. 한편 경제활동과 출산력에 관한 문항이 처음으로 센서스 조사항목에 포함되었다.

이후 우리나라의 인구센서스는 지속적으로 발전을 거듭하였다. 1966년 센서스는 1962년부터 시작된 경제개발5개년계획의 재원 조달로 1년 후인 1966년에 주택은 제외하고 인구만 조사를 했으나, 최초로 1/10 표본조사를 도입하였고, 센서스 자료처리를 위해 1967년 우리나라 최초의 주전산기인 IBM1401을 도입하였다. 연령 파악의 정확성을 개선하기 위해 세는 나이, 생년월일, 양/음력, 띠(12간지)를 처음으로 조사하기도 했다. 1970년에는 주택센서스를 다시 통합하였는데, 대도시 인구집중의 문제로 인구이동 항목이 최초로 추가되었다. 1975년에는 표본조사의 규모를 5%로 줄여서 조사를 하였다. 1985년 센서스는 표본조사는 하지 않고 전수조사만 실시하였는데, 성씨 및 본관 조사와 종교 항목을 처음으로 조사하였다. 1990년 센서스는 OMR(광학마크판독, Optical Marker Recognition) 조사표를 사용하였으며, 처음으로 빈집조사표를 사용하였다. 1995년에는 교통문제의 중요성을 감안하여 통근통학 항목을 처음으로 조사하였다. 2015년에는 표본조사는 기존에 10%에서 20%로 규모를 확대하는 대신 전수조사는 실시하지 않고 가족관계등록부, 주민등록부 등 행정자료를 이용하여 센서스 통계를 작성하는 부분적 등록센서스를 처음으로 도입하였다. 이와 같은 부분적인 등록센서스 도입으로 2016년

부터는 매년 별도의 20% 표본조사 없이 행정자료를 이용하는 센서스 통계가 인구, 가구, 주택의 10여개 항목에 대해서는 매년 기본통계의 형태로 제공되고 있다.

3. 항목의 일관성 검토

1) 마이크로 자료의 일관성

우리나라의 인구센서스는 1925년부터 몇 번의 예외를 제외하고 5년마다 규칙적으로 실시되어 왔다. 1960년에는 조사를 한 이후에 신속하게 집계를 위하여 20%를 추출한 자료를 표본이라고 했기 때문에 전수자료와 표본자료가 사실상 동일하다. 실제로 엄격한 의미에서 센서스 조사과정에 표본조사의 기법을 도입한 것은 1966년이 처음이다. 간단하고 응답부담이 적은 기본 항목은 모든 응답자가 조사항목에 정보를 제공하는 전수조사 방식을 적용하고, 심층적이며 사생활을 침해하는 정도가 큰 조사항목은 일부 응답자들만을 대상으로 표본(1975년은 5%, 1980년은 15%, 1985년 0%, 2010년은 20%이며 1966년, 1970년, 1990년부터 2010년은 10%)으로 선정하여 조사를 했다. 따라서 전수조사는 항목이 적은 반면 표본조사는 항목의 범위가 다양하고 항목 수가 많다.

1980년 이전에만 해도 마이크로데이터는 거의 찾은 사람이 없었다. 자료처리의 최종 목표는 프린트된 보고서를 만드는 것이었기 때문에 마이크로데이터를 중요하게 생각하고 관리하는 사람이 거의 없었던 것으로 알려지고 있다. 따라서 1960년 및 1970년 마이크로 자료(1966년 센서스는 원래부터 인구만 조사)는 가구주택 자료는 없고, 인구자료만 이용자에게 제공되고 있다. 가구주택 자료가 보관되어 있지 않을 것으로 짐작이 된다. 대부

분의 연도는 항목이 많은 표본자료에서 1%나 2% 자료를 만들어 제공한다. 그런데 1975년, 1980년 및 1985년은 전수조사 결과에서 2% 자료를 추출하여 제공하고 있다. 1985년은 표본조사를 실시하지 않은 해이기 때문에 전수자료에서 2%를 추출했지만, 1975년 및 1980년은 각각 5%, 15%의 표본조사를 실시했는데도 이유 없이 항목이 적은 전수조사에서 2%를 추출하여 제공하고 있다.

<표 1> 통계청의 인구주택총조사 마이크로 자료 종류

	전수자료		표본자료	
	인구	가구주택	인구	가구주택
1960			0	
1966			0	
1970			0	
1975	0	0		
1980	0	0		
1985	0	0		
1990			0	0
1995			0	0
2000			0	0
2005			0	0
2010			0	0
2015			0	0

2) 항목의 주기성 또는 반복성

이 연구에서는 인구항목에 국한하여 인구센서스 조사항목의 주기성 또는 반복성을 검토하기로 한다.

(1) 매년 반복조사되는 기본항목

인구센서스는 거대한 국가적 프로젝트로서 기

본통계조사이기 때문에 인구의 규모, 구조, 분포 및 인구·사회·경제적인 특성에 관한 기본적인 사항들은 매년 반복적으로 조사가 되었다.

성명, 가구주와의 관계, 성별, 생년월일, 교육정도, 혼인상태, 경제활동, 산업, 직업, 종사상 지위 등은 1966년부터 2015년까지 한 번도 빠지지 않고 조사가 되었다.

그 밖에 기본항목에 버금가도록 거의 매년 조사하는 항목은 총 출생아 수로, 1966년부터 2015년 중 1995년을 제외하고 계속 조사하고 있으며, 5년전 거주지도 1970년 이후 매회 센서스에서 조사하고 있다.

다소 반복성이 떨어지지만 출생지는 1966년부터 2015년까지 중, 1966년, 1975년, 2005년을 제외하고는 매회 조사가 되었다. 1년 전 거주지는 1980년 이후 1995년과 2005년을 제외하고 계속 포함되었다. 아울러 통근/통학 여부 및 장소, 출발 도착시간 및 이용교통수단 등 통근/통학 항목도 1980년 처음 포함된 이후, 1990년부터 2015년까지 매회 조사가 되고 있다.

이에 따라 우리나라 인구센서스에서 중요하게 취급되는 인구항목은 성, 연령, 가구주와의 관계, 교육정도, 혼인상태 등 기본적인 인구통계 항목과 경제활동, 출산력(총 출생아 수), 인구가동(출생지, 1년 전 거주지, 5년 전 거주지), 통근/통학 항목임을 알 수 있다.

(2) 일정한 주기로 조사하는 항목

매회 조사할 정도로 중요한 항목은 아니지만 주기적으로 조사하는 항목 중, 종교는 10년마다(1985년, 1995년, 2005년, 2015년) 조사하며, 성씨 및 본관은 15년(1985년, 2000년, 2015년²⁾) 마다 작

2) 2015년 성씨 및 본관은 현장조사가 아닌 등록센서스 시스템으로 집계가 되었다.

성되고 있다.

(3) 최근에 추가되어 계속 조사되는 항목

저출산/고령화, 돌봄, 노동, 소외 등 사회현안과 관련되어 개발된 항목들은 개발된 이후 지속적으로 매회 조사가 되고 있다. 아동보육 상태는 2000년 처음 도입된 후 2015년까지 매회 조사가 되고 있으며, 활동제약, 추가계획자녀수도 2005년 처음 조사한 이후 2015년까지 3번 연속 조사가 되고 있다. 비공식분야 경제활동과 관련이 있는 근로장소에 관한 항목도 2005년에 처음 만들어진 이후 2015년까지 계속 포함되었다. 사회활동 항목은 2010년에 처음 개발된 후 2015년에도 조사가 되어 주요항목으로 자리매김을 하고 있다.

(4) 1회성으로 조사하고 없어진 항목

인구센서스 조사항목에는 1회에 그친 조사항목들도 적지 않다. 반복해서 관찰할 필요가 없다는 것은 사실 센서스 항목으로서 거의 가치가 없다고 보아야 한다.

컴퓨터 활용상태, 인터넷 활용상태, 개인휴대용 통신기기는 2000년 센서스에서 한번만 조사가 있었는데, 센서스 보다는 정보통신 관련 행정자료나 표본조사가 더 적합한 항목들이다. 자녀 거주장소, 주부양자, 생계수단 등 고령자 관련 항목들도 2000년에만 조사가 되고 이후 없어진 항목들이다.

한편, 2015년에 처음 도입된 항목으로, 자녀출산시기, 결혼 전 취업여부, 경력단절 등 항목은 2020년 센서스 시범예행조사 조사항목에 포함된 반면 전공계열, 현재 직업 근무년수 등은 포함되지 않아 향후 더 이상 조사항목에 포함되지 않을 가능성이 큰 것으로 전망이 되고 있다.

3) 개념 및 조사방법, 접근법

우리나라 인구센서스에서 인구를 파악하는 방식은 크게 세 차례의 변화가 있었다. 첫 번째는 1925년 센서스에서 처음 적용하여 1955년 센서스까지 적용한 현주지 방식(de facto approach) 개념에 의한 인구집계인데, 이 개념은 센서스 기준일(census day) 자정에 가구 또는 거처에서 발견된 모든 사람을 그 주소의 인구로 파악하는 접근법이다. 이 방식에 의하면 출장, 여행 등으로 잠시 방문한 사람도 발견된 장소의 인구로 포함된다.

두 번째는 상주지 방식(de jure approach) 개념이다. 사람들이 평소 거주하는 장소에서 인구를 파악하는 접근법이며, 이 방식에 따르면 출장, 여행 등으로 잠시 집을 떠나 있는 사람도 평소 거주하는 장소에서 파악된다. 여기서 평소는 3개월 이상 거주하여 왔거나 앞으로 거주할 의향이 있는 장소를 의미한다. 이 접근법은 유엔의 센서스 원칙이 권고한 6개월 또는 12개월 기준보다 느슨한 거주기간으로서, 1960년에 처음 도입되어 2010년까지 적용되었다.

세 번째는 등록인구 개념이다. 2015년부터 도입된 등록센서스에서는 거주지는 기본적으로 주민등록의 등록지를 근거로 한다. 주민등록법 제16조(거주지의 이동)에 따르면 국민들이 거주지를 이동하면 전입한 날부터 14일 이내에 전입신고를 하도록 되어 있는데, 이때 거주지는 30일 이상 거주 의사가 있어야 한다. 인구센서스와 기간에서 차이가 나는데 주목할 필요가 있다고 본다.

주민등록은 주택착약이나 취업, 취학 등의 요인으로 실제 거주하지 않는 지역에 등록이 되는 경우가 적지 않고, 상주지 개념의 조사통계는 조사누락 등 오차가 적지 않은 상황에서, 상주에서 등록센서스로의 조사방식의 변경으로 지역별 인구

의 규모와 구조에 상당한 불일치를 야기할 것으로 보인다(전광희, 2014). 특히, 대학가 지역, 공단, 원룸 등 인구이동이 빈번한 지역에서 시계열 단절이 우려된다.

<표 2>는 지난 2000년 이후 인구센서스에 나타난 서울시 일부 동의 인구추이이다. 2000년~2010년 현장조사 기반의 인구센서스와 2015년 등록센서스의 결과가 크게 차이가 나서 추이가 맞지 않는 동의 예시이다. 주택재개발, 신규 아파트 분양 등으로 실제 인구의 급격한 증가나 감소가 이루어졌을 수도 있지만, 그러한 요인이 없다면 인구센서스 방법론의 차이에서 오는 문제에 기인한다고 보아야 할 것이다. 아울러 2016년부터 2018년까지 등록센서스는 시군구 인구만 제공하고 있어, 시계열 자료로 사용하는데 제약이 있다.

<표 2> 서울 주요 동의 센서스 인구 추이

행정구역	2000	2005	2010	2015
신촌동	-	-	17,748	25,688
가락1동	21,271	19,780	14,850	746
가리봉동	-	-	12,531	19,330
고덕2동	21,797	20,046	17,747	9,029
대림2동	23,273	20,427	16,747	26,088
명동	1,792	2,855	2,140	3,328
방화1동	27,243	31,870	29,226	41,615
북가좌1동	20,157	18,308	9,960	19,059
북아현동	-	-	14,477	9,710
서빙고동	12,877	9,765	10,036	13,068
소공동	1,093	689	736	1,939
시흥1동	17,465	16,414	29,086	37,152
신내1동	28,975	28,089	30,904	40,539
신정1동	25,908	23,664	22,522	17,907
신정3동	33,746	46,538	45,918	53,880
아현동	-	-	16,021	26,928
양재1동	25,278	25,539	23,840	44,964
오류2동	24,970	28,603	27,494	44,562
원효로1동	6,897	5,877	9,983	13,472
장위3동	21,262	15,841	18,117	13,796

행정구역	2000	2005	2010	2015
종로1, 2, 3, 4가동	5,928	4,656	5,900	7,598
중림동	10,334	12,759	11,565	8,392
진관동	-	-	36,484	51,593
혜화동	11,207	11,829	12,347	22,942
회기동	12,543	12,416	10,748	14,317
회현동	5,619	5,063	4,127	5,501

센서스 조사항목의 개념/정의라는 관점에서 반드시 언급해야 할 또 다른 항목은 경제활동 항목이다. 먼저 적용연령은 이 항목을 처음 도입한 1960년부터 1980년까지는 14세 이상을 대상으로 조사를 하였다. 하지만 1985년 이후는 15세 이상으로 대상을 조정하였는데, 이는 당시 경제발전이 따라 중학교 진학률이 99%를 상회했기 때문이다.

접근법에서도 변화가 있었다. 경제활동인구는 조사대상 기간의 차이에 따라 두 개의 경제활동인구로 구분이 된다. 첫 번째로 정상상태활동인구(currently active population)로 1주일, 1일과 같이 짧은 기간에 적용하는 것으로 노동력(labour force)이라고 하고, 또 다른 접근법은 평상상태활동인구(usually active population)로 1년과 같이 장기간을 측정하는 것이다. 전자는 정상상태접근법(또는 노동력접근법), 후자는 평상상태접근법이라고 한다. 그밖에도 조사대상기간이 없이 직업이나 사업이 있는 사람만을 파악하는 유업자(gainfully occupied population) 접근법도 있다(김민경, 2000; 2020).

그런데 경제활동의 규모, 구조는 접근법에 따라 크게 달라진다. 우리나라는 <표 3>에서 보듯이 센서스 마다 경제활동인구를 접근하는 방법이 상이하기 때문에 시계열 분석이 쉽지 않다. 특히 2015년의 경우 구직활동과 일을 할 수 있는지 여부를 질문하지 않아 실업 상태 여부는 알 수가 없는 문제점이 있다.

<표 3> 센서스의 경제활동인구 조사 방식 변화

연도	접근법	대상 기간	구직 문항	취업 가능 여부 문항
1960	노동력 접근법	지난 1주일간	○	
1966	노동력 접근법	지난 1주일간	○	
1970	노동력 접근법	지난 1주일간	○	
1975	노동력 접근법	지난 1주일간	○	
1980	노동력 접근법	지난 1주일간	○	
1985	유업자 접근법	대상기간 없음	○	
1990	평상상태 접근법	지난 1년 중 30일	○	
1995	노동력 접근법의 변형	지난 1개월간	○	
2000	노동력 접근법	지난 1주일간	○	○
2005	노동력 접근법	지난 1주일간	○	○
2010	노동력 접근법	지난 1주일간	○	○
2015	노동력 접근법	지난 1주일간		

4) 분류

통계분류(statistical classification)의 변화 또한 인구센서스 자료를 시계열로 만들어 분석하기 어렵게 만드는 장애 요인 중의 하나이다. 여기서는 행정구역의 변경이나 산업 및 직업분류와 같이 한국표준분류의 변화에 대해서는 언급하지 않기로 한다. 왜냐하면 이러한 표준분류의 변화는 반드시 인구센서스에 국한 되는 문제가 아니라 모든 국가통계에 공통되는 문제이기 때문이다.

(1) 가구주와의 관계

가구주와의 관계(relationship to the head of the household)는 가구형태를 구성하는 항목으로 모든 항목 중에 가장 핵심이 되는 항목 중 하나이다. 유엔은 가구주(head of household)라는 개념이 가족 구성원 간의 평등 원칙에 맞지 않기 때문에 가구를 대표하는 참고인(reference)을 선정하고 이 참고인과의 관계를 조사하도록 권고하고 있다(United Nations, 2017). 하지만 여전히 많은 비서구권 국가에서는 가구주라는 개념을 사용하고 있고, 가구 구조나 형태를 이해하는데 중요한 역할을 하고 있다.

1960년 인구센서스에서는 0.집단가구원/1.가구주(집단가구대표자)/2.배우자/3.아들/4.딸/5.손/6.부모/7.조부모/8.친족/9.동거인/미상으로 구성되어 있다. 집단가구원을 사전에 배제하지 않고 조사에 포함했다.

1966년에는 11.가구주/12.배우자/13.자녀들/14.부모/15.손자녀/16.조부모와 그의 형제, 자매/17.가구주의 형제, 자매/18.삼촌, 고모, 기타 친척/19.하숙인, 가정부, 고용인(6명 미만)/21.1인 가구(하숙인, 호텔 거주자)/22.집단가구(양로원, 고아원, 수도원)/23.집단가구(양로원, 고아원, 수도원)의 구성원/24.집단가구(6명이상 고용원을 포함)/25.집단가구(6명이상 고용원을 포함)의 구성원/미상으로 구분하였다. 1인가구, 시설가구, 집단가구를 구분하였다.

1970년에는 1.가구주/2.배우자/3.미혼자녀/4.미혼 형제, 자매/5.부모/6.조부모/7.배우자의 형제, 자매/8.배우자의 부모/9.기혼 아들/10.기타 친척/11.하숙생, 가정부(식모), 고용인/12.1인가구의 가구주/13.기숙, 사회 시설의 관리자/14.기숙, 사회 시설의 수용인/15.가구주의 처 이외에 결혼해서 생존한 아이를 데리고 있는 여인/21.20의 자녀/30.제

3의 경우/31.30의 자녀/미상으로 분류하며, 1975년도에는 1970년 응답 범주에서 동거인, 가사 및 영업사원과 가정부를 분리하였다. 이전 센서스와 비교할 때, 기혼 아들(며느리 및 손자녀)을 새로이 분류하였는데, 장남과 동거하는 시대상을 반영한 것으로 보인다. 특히 가구주의 처 이외에 결혼해서 생존한 아이를 데리고 있는 여성과 그 자녀를 별도 분리해 새로이 포함한 것이 이채롭다고 할 수 있다.

1980년 및 1985년에는 오늘날처럼 주요한 관계 위주로 간단하게 조정이 되었다. 1.가구주/2.배우자/3.미혼자녀/4.미혼 형제, 자매/5.부모/6.조부모/7.배우자의 형제, 매/8.배우자의 부모/9.기혼자녀, 그 배우자 및 자녀/10.기타친척/11.기타 동거인/91.단독가구/미상으로 구분하였다. 가구주와 배우자를 구분하여 그들의 부모와 형제, 자매를 파악한다거나, 기혼자녀를 아들, 딸을 구분하지 않는 등 관계 분류에 성 평등주의적 사고를 반영하려고 노력하였다. 한편, 시설 및 집단가구 여부는 1980년 가구주와의 관계 분류부터 없어진 것을 알 수 있다.

1990년은 1985년 분류에서 1인가구를 제외하였다. 이후 센서스에서 1인가구는 가구주와의 관계에서 사라졌다.

1995년 및 2000년은 1.가구주/2.배우자/3.자녀/4.자녀의 배우자/5.가구주의 부모/6.배우자의 부모/7.손자녀, 그 배우자/8.증손자녀, 그 배우자/9.조부모/10.형제자매, 그 배우자/11.형제자매의 자녀, 그 배우자/12.기타 친인척/13.기타 동거인으로 구분한다. 자녀의 배우자를 별개 항목으로 구분했고, 손자녀와 배우자, 증손자녀와 배우자, 형제자매의 자녀와 배우자 등 기타 친척에 속하던 관계들을 다수 응답 범주로 분류했다.

2005년은 1.가구주/2.배우자/3.자녀/4.자녀의 배우자/5.가구주의 부모/6.손자녀 및 그 배우자, 증

손자녀 및 그 배우자/7.형제자매, 그 배우자/8.기타 친인척 (형제자매의 자녀 및 그 배우자 + 부모의 형제자매 및 그 배우자 + 기타친인척)/9.기타 동거인으로 1960년과 함께 가장 간단하게 가구주와의 관계가 범주화되었다. 사실 2005년 가구주와의 관계 분류는 “최악”이라고 평가할 수 밖에 없는 정도로, 사회변동을 연구하는 장기적인 프로젝트에서는 많은 시계열 단절이 2005년에 발생한다.

2010년은 2000년 분류 중 기타친인척에서 부모의 형제자매·그 배우자를 분리한 것 이외에는 동일하다.

2015년에는 2010년 분류 중 손자녀 및 그 배우자를 친손자녀인지 외손자녀인지를 추가로 구분하는 분류를 적용했다. 이는 아동보육을 친가에서 하는지 외가에서 하는지를 보기 위한 의도라고 해석할 수 있다.

가구주와의 관계 변천을 보면, 시설가구나, 집단가구, 1인가구 등 가구주와의 관계와 관련 없는 분류들이 1960년부터 1985년 사이에 있다가 없어진 것을 알 수 있었고, 1970년부터는 부모나 형제자매의 경우 가구주 관련인지 배우자 관련인지를 구분하였다. 특히 2010년 이후에는 손자녀, 증손자녀, 형제자매, 부모의 형제자매에 배우자를 모두 포함하는 등 기타 친인척에 대한 구분을 강화하였다.

이렇듯 연도마다 다른 가구주와의 관계를 시계열로 생성하기 위해서는 공통된 분류로 묶을 수 밖에 없는데, 이 연구에서는 다음 항목으로 검토할 가구형태(living arrangement)와 세대 수 분류가 가능하도록 가구주와의 분류를 유지하는 것이 중요하다고 본다.

위의 가구주와의 관계 항목의 활용도를 통계 이용자의 관점에 재검토할 때, 시계열 자료는 구비해야 할 최소 분류는 다음과 같이 정리할 수 있다.

먼저 가구주와 배우자, 미혼자녀, 부모의 구분은 필수적이다. 그런데 1960년, 1966년 및 1995

년부터 2015년까지의 센서스는 자녀여부만 질문하였기 때문에 그 자녀들의 혼인상태와 다시 결합하여 미혼여부를 확인하여 다시 분류할 필요가 있다. 부모의 경우에는 대부분 가구주의 부모인지 배우자의 부모인지 구분이 되나, 1960년 및 1966년 그리고 2005년은 가구주 및 배우자 부모의 구분이 없다. 이러한 구분을 유지하는 것은 가족 동거형태에서 부계 및 모계 연구에는 유용하다.

조부모는 2005년을 제외하고 모든 연도에 구분되어 있었으며, 1966년에는 조부모와 그의 형제자매까지 포함하고 있다. 따라서 세대구분에 필요할 뿐만 아니라 시계열 유지가 용이한 분류라고 할 수 있다.

손자녀는 1960년, 1966년에 조사하다가 이후 한동안 포함되지 않았다가 1995년 이후 2015년까지 다시 분류되기 시작한다. 1995년 이후는 2005년을 제외하고는 모두 손자녀와 증손자녀를 구분하는데, 특히 2015년 이후에는 친손자이지 외손자인지도 구분하여 아동보육에서 부계와 모계의 역할을 살펴보도록 하였다. 손자녀와 증손자녀의 구분은 세대 구성을 연구하는데 도움이 되지만 이는 1995년 이후로 제한해야 한다.

형제자매는 1966년에 처음 등장하지만 가구주의 형제자매만을 특정했다. 1970년부터 1990년까지는 가구주의 형제자매는 미혼을 명시한 반면, 배우자의 형제자매는 혼인상태를 고려하지 않고 분류했다. 1995년부터 2015년은 혼인상태나 가구주 및 배우자의 형제자매인지 여부를 고려하지 않고 분류하였다.

자녀의 배우자는 1995년 이후 2015년까지 분류되고 있으나, 가구형태나 세대 구성 측면에서는 중요한 정보는 아니다.

따라서 이 연구에서는 1960년부터 2015년까지의 가구주와의 시계열 분석을 위해 다음과 같은 재분류가 필요하다고 본다.

- 1) 가구주
- 2) 배우자
- 3) 미혼 자녀
- 4) 부/모
- 5) 손자녀
- 6) 조부/모
- 7) 형제자매
- 8) 기타

<표 4> 가구주와의 관계 시계열: 1960~2015년

	가구주	배우자	미혼 자녀	부모	손자녀	조부모	형제 자매
1960	O	O	O ¹	O	X	O	X
1966	O	O	O ¹	O	X	O ⁵	X
1970	O	O	O	O	X	O	O ⁶
1975	O	O	O	O	X	O	O ⁶
1980	O	O	O	O	X	O	O ⁶
1985	O	O	O	O	X	O	O ⁶
1990	O	O	O	O	X	O	O ⁶
1995	O	O	O ¹	O ²	O ³	O	O ⁷
2000	O	O	O ¹	O ²	O ³	O	O ⁷
2005	O	O	O ¹	O	O ³	X	O ⁷
2010	O	O	O ¹	O ²	O ³	O	O ⁷
2015	O	O	O ¹	O ²	O ⁴	O	O ⁷

- 주) 1. 자녀만 조사하여 혼인상태 결합필요
 2. 가구주의 부모와 배우자의 부모 구분
 3. 증손자 및 배우자 포함
 4. 친, 외손자 구분 가능하며 배우자 포함
 5. 형제자매 포함
 6. 미혼
 7. 기혼 및 배우자 포함

(2) 연령

1960년 조사에서 연령(age)은 만나이가 아닌 “세는 나이”로 조사하였다. 1세와 2세 유아의 경우에는 생년월일을 조사하였다. 따라서 0세 인구가 없으며, 전체 연령이 1세 정도 상향되어 집계되는 특이한 센서스 통계가 작성되었다.

연령조사는 1960년에는 세는 나이 한 가지만 질문하였으며, 1966년에는 세는 나이와 출생년월일을 동시에 조사하였다. 하지만 이러한 방식의 질문은 양력이나 음력 한 가지로만 나이나 생년월일을 기억하는 사람들의 만나이를 정확하게 조사할 수가 없었다. 이에 따라 1970년부터는 양음력과 띠를 같이 조사하고 있다.

2010년부터는 연령의 정확성이 개선이 되어 띠로 확인할 필요성이 없어지면서, 띠에 대해서는 더 이상 질문을 하지 않고 있다.

한편, 우리나라 인구센서스의 연령은 특정 숫자 선호(digit preference) 없이 비교적 정확한 것으로 평가되고 있어, 이 항목의 세부항목 변경에 따른 영향은 거의 없다고 볼 수 있을 것이다.

<표 5> 연령 세부항목: 1960~2015년

	세는 나이	출생년월일	양음력	띠	만 나이
1960	○				
1966	○	○			
1970	○	○	○	○	
1975	○	○	○	○	○
1985	○	○	○	○	○
1990	○	○	○	○	○
1995	○	○	○	○	
2000	○	○	○	○	
2005	○	○	○	○	
2010	○	○	○		
2015	○	○	○		

(3) 혼인상태

혼인상태(marital status)는 응답배열이 센서스마다 두 가지로 나타나 코드 값으로 시계열 작성 시 주의가 필요하다. 1960년, 1966년, 1970년, 1980년은 유배우, 사별, 이혼, 미혼의 순서이고, 다른 연도는 미혼, 유배우, 사별, 이혼의 순서이다.

<표 6> 센서스 혼인상태 코드값: 1960~2015년

코드값	1960년, 1966년, 1970년, 1980년	1975년, 1985년~2015년
1	유배우	미혼
2	사별	유배우
3	이혼	사별
4	미혼	이혼
8 또는 9	미상	미상

혼인상태 응답대상은 1960년, 1966년, 1970년은 모든 인구를 대상으로 질문하였으며, 1975년부터 1990년까지는 만 15세 이상 그리고 1995년부터 2015년까지는 세는 나이 16세 이상으로 포괄범위가 다소 다르다. 세는 나이 16세 이상으로 제한하기 때문에 조사대상의 응답 부담은 줄일 수 있지만, 실제 세는 나이 15세 이하에서 혼인한 사람은 포착할 수 없는 문제도 간과할 수는 없다. 또한 세는 나이 16세와 만 나이 15세도 엄밀하게 말하면 포괄범위가 다르지만 결과 집계 과정에서 만 나이 15세 이상 기준으로 다시 환산하기 때문에 시계열 상에 문제는 그리 생겨나지 않을 것이라고 볼 수 있다.

(4) 경제활동 취업자와 비(非)경제활동인구 구분

경제활동 항목은 센서스 간 개념과 정의의 변화는 물론 응답범주에도 많은 변화가 있어 시계열 자료로 사용할 때 많은 주의를 필요로 한다.

경제활동인구(economically active population)는 크게 취업자, 실업자, 비경제활동인구로 구분된다. 취업자는 현재 일을 하거나, 가사, 학교(학원) 등 다른 활동을 하면서 틈틈이 일하였거나, 일을 하여 왔으나, 휴가 등으로 잠시 쉬고 있는 경우로 구분된다. 1960년과 2005년은 취업을 하나로만 질문한 반면 나머지 대부분 연도는 취학이나 가사를 돌보면서 틈틈이 일하거나, 일이 있지만 사정

상 잠시 쉬고 있는 경우를 취업자에서 빠뜨리지 않기 위해 세분해서 질문하였다.

비경제활동은 가사, 통학 위주로 질문하였다. 1970년과 1975년에는 연로와 불구도 분리해서 자세하게 질문했다. 반면 1985년 이후는 비경제활동의 세부 범주에 대해 전혀 질문을 하지 않았다.

경제활동 항목은 특히 2000년 이후 항목을 축소하려는 경향이 보이는데, 이는 조사결과가 월간 경상조사인 경제활동인구조사와의 불가피하게 차이가 나기 때문에 소극적으로 이 항목을 다루었기 때문이 아닌가 추측을 할 수 있을 뿐이다.

경제활동은 인구의 사회경제적인 지위나 신분은 물론 소득과도 밀접하게 관련이 되어 있기 때문에 그들의 사회적인 행동이나 판단, 선택을 이해하는데 매우 중요한 변수이다. 인구센서스의 많은 변수들을 경제활동과 연계하면 각 항목들을 보다 심층적으로 분석할 수 있다.

그런데 이 항목이 중요성에 비해 관리되는 덜 되고 있다는 느낌이다. 특히 실업자의 경우 경제활동인구조사에서는 지난 4주간 구직활동을 했는지, 또 일이 있었다면 할 수 있었는지를 물어 봐서 실업자를 확정하지만, 인구센서스에서는 2000년~2010년을 제외한 대부분 연도에서 구직활동 여부만 물어 봤다. 2015년의 경우에는 이마저도 질문을 하지 않았다. 무급가족종사자의 경우에도 36시간 이상 여부를 질문하여 취업자와 비경제활동으로 재분류해야 하는데 이러한 문항은 아예 없었다.

비경제활동의 종류도 중요한 정보 중 하나이다. 가사, 통학은 물론 연로, 불구 등 다양하게 질문한 연도는 1970년과 1975년에 불과하고 2005년 이후는 아예 파악조차 하지 않고 있다.

(5) 교육정도

교육정도는 지난 60년간 변화가 많은 항목 중 하나이다. 1960년은 학교 급별로 파악하지 않고 교육년수로 조사하기 때문에 다른 연도와 시계열로 연결하려면, 초등 6년, 중학 3년, 고등 3년, 대학 4년과 같이 교육년수를 재분류해서 사용할 수 밖에 없다.

1966년부터는 학교 급별로 조사했다. 안다녔음 코드는 1970년과 1985년은 없고, 1985년은 6번으로 가장 마지막에 배치한 것을 제외하면 모두 1번으로 가장 먼저 배치했다.

학교급은 1966년은 중학교가 신제중학교와 구제중학교)로 구분된다. 1970년부터 1985년까지는 초등학교, 중학교, 고등학교, 초급대·전문대, 대학 이상으로 구분했다. 다만 1980년에는 고등학교를 인문고등학교와 실업고등학교로 세분해서 질문하였다.

<표 7> 센서스 취업상태 코드값: 1960~2015년

	취업			비경				
	취업	주1	주2	가사	통학	연로	불구	기타
1960	○			○	○			○
1966	○	○	○	○	○			○
1970	○		○	○	○	○	○	○
1975	○		○	○	○	○	○	○
1985	○	○	○	○	○			○
1990	○	○	○	○	○			
1995	○	○	○	○	○			
2000	○		○	○				
2005	○							
2010	○	○	○					
2015	○	○	○					

- 주) 1. 가사, 학업(학교, 학원) 등 다른 활동을 하면서 틈틈이 일하였음.
- 2. 일자리를 갖고 있으나, 휴가 등으로 잠시 쉬고 있음.

3) 일본 강점기에 설치된 5년제 중등교육기관으로 제2차대전 후 6년제로 유지되다가 1951년 폐지되었다.

<표 8> 교육 정도 항목: 1966~2015년

	초등	중학	고등	대학		대학원	
				(2년)	(4년)	석사	박사
1966	○	○ ¹	○	○	○		
1970	○	○	○	○	○		
1975	○	○	○	○	○		
1980	○	○	○ ²	○	○		
1985	○	○	○	○	○		
1990	○	○	○	○	○		○
1995	○	○	○	○	○		○
2000	○	○	○	○	○	○	○
2005	○	○	○	○	○		○
2010	○	○	○	○	○	○	○
2015	○	○	○	○	○	○	○

주) 1. 구제중학교와 신제중학교로 구분
 2. 인문고등학교와 실업고등학교로 구분

1990년과 1995년은 대학원이 처음으로 포함되었고, 2000년부터 2015년까지(2005년 제외)는 대학원을 석사과정과 박사과정으로 분리해서 파악하였다. 다만 2005년은 대학도 2년제와 4년제를 구분하지 않았고, 대학원도 석사 및 박사과정을 구분하지 않았다. 조사표에는 대학의 경우 4년제 미만과 이상이 구분되며, 대학원도 석박사가 분리되어 조사된 것을 감안하면, 이는 통계청의 범국민 이용용 2% 센서스 개별자료의 제작 착오로 볼 수 있다.

교육정도 항목은 이수여부에 대해 1966년부터 1980년까지 졸업, 재학, 중퇴로 구분하여 조사하였다. 이후 1995년에는 휴학이 추가되었고, 2000년 이후에는 중퇴 문항도 추가 되었다.

이수여부 항목은 교육정도를 교육년수로 환산할 때 유용하게 사용할 수 있으며, 보통 교육정도는 졸업을 해야지만 해당 학교급을 마친 것으로 보고 다른 상태는 직전 학교 급별로 조정하여 사용한다.

<표 9> 교육 이수여부 항목: 1966~2015년

	졸업	재학	수료	휴학	중퇴
1966	○	○	○		
1970	○	○	○		
1975	○	○	○		
1980	○	○	○		
1985	○	○	○		
1990	○	○	○		
1995	○	○	○	○	
2000	○	○	○	○	○
2005	○	○	○	○	○
2010	○	○	○	○	○
2015	○	○	○	○	○

(6) 총 출생아 수

총 출생아 수는 전수조사만 실시한 1985년을 제외하고는 모두 조사하였다. 성별로 구분하여 조사한 것은 1975년 이후로, 인구동태통계 특히 출생통계의 완전성에 흠결이 많은 시기에 해당한다.

1960년은 성별 구분 없이, 2000년은 성별 구분하여 총출생아수만 질문하였다. 1966년은 성별 구분 없이 총출생아 수, 사망아 수, 생존아 수를 물었다. 1985년과 1990년은 성별 구분하여 총출생아 수, 사망아 수, 생존아 수를 파악하였고, 2005년은 생존아 수 대신 타지거주 자녀와 동거자녀로 변경하여 질문하였다. 2010년과 2015년은 성별 구분하여 총출생아 수 및 사망자녀 수만 질문하였다. 조사대상은 1985년까지는 15세, 1990년은 만 15세, 2000년부터는 16세 이상의 기혼 여성을 대상으로 하였으며, 미혼모는 공식적으로 조사하지 않았다.

UN(2017)에서는 총출생아수와 생존자녀수, 사망자녀수를 구분하여 질문할 것을 권고하고 있다. 총출생아수 및 사망자녀수를 물어 두 수치의 차이로 생존자녀수를 계산하는 방식은 사망자가 덜 잡히기 때문이다.

<표 10> 총 출생아 수: 1960~2015년

	총출생아수			사망 자녀			생존 자녀			타지 거주			동거 자녀		
	남	여	계	남	여	계	남	여	계	남	여	계	남	여	계
1960			0												
1966			0			0			0						
1970			0						0						
1975			0	0	0		0	0		0	0				
1985	0	0	0	0	0	0	0	0	0						
1990	0	0	0	0	0	0	0	0	0						
1995															
2000	0	0	0												
2005	0	0		0	0					0	0		0	0	
2010	0	0		0	0										
2015	0	0		0	0										

우리나라의 총출생아수 세부항목은 연도마다 달라 시계열로 구성 시 미세하지만 정확도가 다르다는 논란이 있을 수도 있다. 이 항목은 우리나라처럼 인구동태통계가 발달한 국가에서는 절대적인 출산력 측정보다는 다른 변수하고의 관계나 적생아법(Own Children Method) 및 출산경험재구성법(Birth History Reconstruction Method)을 이용한 합계출산력 추정이나 차별 출산력 연구 등 2차 가공분석에 더 가치가 있는 항목이다.

5) 파생변수의 일관성

파생변수(derived variables)는 센서스 조사표에 있는 문항을 토대로 사후적으로 만든 항목들이다. 통계청에서는 2% 센서스 마이크데이터를 통계이용자들에게 제공할 때 조사된 내용을 토대로 의미 있는 파생변수들을 만들고, 이들을 기준 변수로 하여 자료를 제공하고 있다.

주요 파생항목으로는 가구주와의 관계, 혼인상태 등을 조합하여 만든 세대유형, 응답한 가구원수로 집계한 가구원 수, 동거형태로 재집계한 가구 핵 수 및 세대 수 등이 있다.

(1) 가구원 수

가구원 수는 고출산/고사망에서 저출산/저사망으로의 인구변천의 과정에서 농경사회의 대가족에서 산업화 및 도시화로 핵가족, 소가족으로 가구가 분화되고 탈가족화(defamiliarization) 되어 가는 모습을 파악하는 중요한 변수이다.

가구원 수는 센서스 조사표에서 명시적인 조사항목으로 포함되기 보다는 조사표 난외사항으로 수집되는 정보이다. 1966년, 1970년은 인구항목으로 제공하고 있으며, 1990년부터 2010년까지는 가구/주택 항목으로 제공되고 있다. 1985년도에는 인구와 가구/주택 양쪽 데이터 세트에서 모두 가구원 수 항목이 제공되어, 자료제공 방식에 문 제점이 있다고 본다.

반면 1960년, 1975년, 1980년, 2015년은 통계청에서 가구원 수 정보를 제공하지 않아 이용자가 가구원 수를 직접 계산해야 한다. 생성하는 방법은 가구번호가 같은 가구원 수를 카운트하는 방식이다. 이 방식으로 1975년, 1980년, 2015년은 가구원수의 계산이 가능하다. 하지만 1960년 센서스 자료의 경우 2% 표본의 표집이 가구 단위가 아니

라 개인 단위로 이루어져 가구단위로 가구원 수를 카운트 할 수 없어 가구원 수 생성이 원천적으로 불가능한 문제점이 있다.

2015년 센서스의 경우에는 난외사항으로 가구원 수를 조사까지 했으면서도 가구원 수 같은 중요한 정보를 제공하지 않고 있어, 가구와 가족을 시계열적으로 비교하는 연구자들에게 어려움을 주고 있어서 이에 개선이 필요하다고 본다.

<표 11> 가구원 수: 1960~2015년

	인구 항목	가구 항목	생성 필요	생성 불가	조사표 난외사항
1960				○	
1966	○ ¹				○
1970	○ ¹				○
1975			○		
1980			○		
1985	○	○			○
1990		○			○
1995		○			○
2000		○			○
2005		○			○
2010		○			○
2015			○		○

주) 1. 남녀 구분해서 제공

(2) 핵 수

가에서 핵은 가구를 형성하는 최소의 단위이다. 유엔에 따르면 가구 핵은 부부 또는 부부+미혼자녀, 편부+미혼자녀, 편모+미혼 자녀로 이루어지는 가구를 구성하는 최소 단위이다(United Nations, 2017). 확대가구(extended household)는 가구 핵과 친척 또는 친척인 복수의 가구 핵 또는 친척으로만 이루어진 가구이며, 복합가구(composite household)는 비친척이 가구 핵과 다양한 형태로 혼합되거나 남남으로만 이루어진 가구이다.

가구 핵(household nucleus)이 하나로만 이루어진 가구를 핵가족 가구라고 한다. 가구의 분화나 소가족화의 정도를 측정하는 방법으로 우리는 핵가족 가구 비율(nuclear household prevalence)을 이용하기도 한다.

가구에 핵이 몇 개인지를 집계하는 것은 보통 조사가 끝나고 센서스 통계의 제표작업을 위한 전산처리 과정에서 사후적으로 집계하여 파생변수로 제공을 한다.

센서스 마이크로 데이터의 검토결과 가구 핵 수는 1975년 이후 파생변수로 매년 제공되고 있으나, 2005년 및 2015년에는 제공되지 않았다. 가구 핵 수는 가구의 구성이 혈연과 비(非)혈연 간의 조합에서 나타나는 변화를 파악하는데 반드시 필요한 항목이므로, 모든 센서스 실시연도에서 해당 자료를 제공하는 것이 중요하다.

(3) 세대유형

세대유형은 가구원의 동거형태(living arrangement)를 의미한다. 가구주와의 관계와 혼인상태를 바탕으로 개인의 가구에서의 지위로 유형을 파악한다.

앞에서 살펴 본 바와 같이 가구형태는 가구 핵(household nucleus)과 이들과의 관계에서 이루어진다. 가구 핵이 부부 또는 부부+미혼자녀, 편부+미혼자녀, 편모+미혼 자녀로 이루어지기 때문에, 이 형태의 가구를 파악하는 것이 가장 기본이다. 그 다음은 가구핵과 친척/인척으로 이루어지는 확대가구, 가구핵과 남남으로 이루어지는 복합가구를 파악해야 한다.

우선 1960년 센서스에는 세대유형에 대한 정보가 없어 가구주와의 관계를 이용해 세대유형을 별도로 집계해야 한다. 1966년과 1970년은 인구 항목으로 집계하여 제공하고 있으며, 이후는 가구

항목으로 제공하고 있다.

우리나라의 세대 유형은 핵가구, 확대가구, 혼합가구라는 UN(2017)의 방식을 따르지 않고 세대(generation)의 수와 세대(household) 유형이 혼합되어 다소 복잡하다.

기본적으로 모든 연도에서 1, 2, 3, 4세대 이상과 같이 세대의 수로 일관되게 세대유형을 구분하고 있다. 하지만 오늘날 중요하게 여기는 1인가구나 비친족가구는 연도마다 상이하게 구분하여 중요하게 취급하지 않았다. 예를 들어 1966년, 1995년, 2000년은 단독가구와 비친족가구(또는 집단가구)를 구분하지 않고 하나로 묶어서 분류했으며, 1970년과 1980년은 단독가구와 집단가구를 구분하지 않고 함께 분류했다. 여기서 집단가구가 비친족가구를 포함하고 있는지도 분명하지가 않다. 아울러 2005년도에는 아예 단독가구와 비친족가구를 분류조차 하지 않았다. 두 유형의 가구가 분리해서 집계된 것은 1985년, 1990년, 2010년과 2015년에 불과하다. 그것도 1985년과 2010년은 집단가구와 단독가구를 분리했고, 1990년과 2015년은 비친족가구(비혈연가구)와 분리하였다. 1975년도에는 비혈연이나 집단가구 모두 분류에 포함되지 않았다.

통계청(2020)에 따르면, “남남이 사는 5인 이하의 가구”는 일반가구로 분류하며 세대유형은 비친족가구이다. 반면 “남남이 사는 6인 이상의 가구”는 집단가구, “기숙사나 보육원, 요양원 등 사회시설에 집단으로 살고 있는 가구”는 집단시설가구로 분류한다. 이러한 기준에서 볼 때, 지금까지의 우리나라 센서스에서 연도마다 1인가구, 비친족가구, 집단가구 등의 이름으로 가구 수를 카운트하는 것의 범위가 혼합되어 분명하지 않은 것으로 나타났다.

현재 가구의 분화와 함께 탈가족화가 급속히 진행되고 있어 가구유형이 날이 복잡 다양해 지

고 있다. 세대(generation)로 구분할 수 없는 1인가구, 비친족가구, 집단가구, 집단시설가구 등에 대해서도 중요한 가구유형으로 간주하여 일관된 기준으로 집계할 필요가 있다고 본다.

특히 고시원은 “취사와 취침을 공동으로 하는 가”라는 가구 정의를 적용하면, 이론적으로 이용자가 각각 1인 가구로 조사가 될 수도 있고, 집단가구(6인 이상 20인 미만인 경우)로 분류되거나, 집단시설가구(20인 이상 또는 기숙시설이 있는 경우)로 분류될 수도 있다. 만약 센서스마다 조사지침이 다르다면 1인 가구 수가 영향을 받을 수도 있기 때문에 1인 가구, 집단가구, 집단시설가구의 기준이나 정의에 대해서는 특별한 관리와 통계작성이 필요하다고 본다(전광희, 2016).

<표 12> 기타 세대유형 구분: 1960~2015년

	1~4세대	기타유형 1	기타유형 2
1960			
1966	○	1인가구, 비친족가구	준가구
1970	○	1인가구, 집단가구	
1975	○	단독가구	
1980	○	단독가구, 집단가구	
1985	○	단독가구	집단가구
1990	○	단독가구	비혈연가구
1995	○	1인가구, 비혈연가구	
2000	○	1인가구, 비혈연가구	
2005	○		
2010	○	1인가구	집단가구
2015	○	1인가구	비친족가구

다음은 1세대부터 3세대까지(4세대 이상은 세부 분류 없음) 세부적인 분류의 변천을 알아보자.

① 1세대

1995년 이전까지는 부부인지 형제자매나 그 밖의 유형이지만 분류했는데, 2000년 이후에는 형제자매도 미혼으로 한정하였고, 부부가 아닌 가구

주와 형제자매나 기타친인척의 동거 유형도 구분하고 있다.

<표 13> 세대유형(1세대): 1966~2015년

1966	1970~1995	2000~2015
부부	부부	부부
부부+부부의 형제자매	기타	부부+미혼 형제자매
그 외 1세대가구		부부+기타친인척
		가구주+미혼형제자매
		가구주+기타친인척
		1세대 기타

② 2세대

2세대는 가구의 주된 유형이기 때문에 세대유형이 가장 다양하다. 기본적으로 가구 핵에 해당하는 부부+자녀, 편부모+자녀가 기본적으로 모든 연도에 들어 있다. 하지만 편부와 편모로 구분한 것은 1966년과 1995년부터 2015년까지의 센서스에 해당하고, 1970년부터 1990년 센서스는 편부모를 하나의 분류로 묶었다. 필요하다면 편부모의 성별 자료와 결합하여 재분류가 가능하다. 부모에 대해서는 1966년부터 2015년까지 양친이나 편친인가만을 구분하고 편친이 부인지 모인지는 구분하지 않아서, 자료의 활용 가치를 떨어뜨리고 있다. 부부의 형제자매는 모든 연도에 구분이 되어 중요한 구성임을 보여 주고 있다.

조손가구에 대해서는 1985년에 분류를 하였는데, 편부모+손자, 손녀만 구분하고 특이하게 부모+손자, 손녀는 분류가 없었다. 1995년 이후는 조손가구에 대한 분류가 본격적으로 이루어졌는데, 특히 2010년 이후에는 조부모가 모두 있는지 아니면 조부모 중 한쪽만 있는지도 구분하였다.

2010년부터는 부부+자녀에 대해 미혼자녀를 명시하여 가구 핵을 명확하게 하였다. 과거 분류들은 2005년 이전은 미혼여부를 명시하지 않아 기

혼자녀가 섞여져 분류되었는지 명확하지 않다.

<표 14> 세대유형(2세대): 1966~2015년

1966	1970~1980
부부+자녀	부부, 자녀
부부+자녀+부부의 형제자매	편부모, 자녀
부+자녀	부부, 양친
모+자녀	부부, 편부모
부모+기혼부부(기혼부부의 형제자매 포함)	부부, 자녀, 부부의 형제자매
편부모+기혼부부(기혼부부의 형제자매 포함)	기타
그 외 2세대가구	
1985	1990
부부+자녀	부부, 자녀
편부모+자녀	편부모, 자녀
부부+양친	부부, 양친
부부+편친	부부, 편부모
부부+자녀+부부의 형제 또는 자매	부부, 자녀, 부부의 형제자매
부부+손자, 손녀	기타
편부모+손자, 손녀	
2세대 기타	
1995~2005	2010~2015
부부+자녀	부부+미혼자녀
편부모+자녀	부+미혼자녀
편모+자녀	모+미혼자녀
부부+양친	부부+양친
부부+편부모	부부+한부모
부부+자녀+부부의 형제자매	부부+미혼자녀+부부의 미혼형제자매
조부모+손자녀	조부모+미혼손자녀
2세대 기타	조부 또는 조모+미혼 손자녀
	2세대 기타

이에 대해서는 검증을 위해서는 모든 자료에 대해 혼인상태를 대조하는 체계적인 작업이 필요하다고 본다.

③ 3세대

3세대는 부부와 자녀, 부모가 함께 사는 유형이다. 2005년 이전에는 부모가 양친인지 편부모인지만 구분했는데, 2010년 이후에는 편부모의 경우 부친인지 모친인지 구분을 한다. 자녀의 경우는 2010년 이후에는 미혼 자녀로 명시했다. 과거 자료의 경우 미혼만 포함했는지는 알 수가 없으며, 정확한 진단을 위해서는 혼인상태 변수와 대조가 필요할 것이다.

그러다가 최근에는 인구파일과 가구/주택 파일을 분리하여 자료를 제공하고 있어, 자료를 이용자가 인구 및 가구주택 세트로 분리할 필요는 없어졌다. 하지만 여전히 개인의 가구 특성정보(예, 개인별이 자신이 속한 가구의 가구원 수, 가구형태, 점유형태, 주택의 종류 등)를 사용하기 위해서는 두 자료세트를 공통키를 이용해 통합(merge)해야 한다.

<표 15> 세대유형(3세대): 1966~2015년

1966	1970~2005	2010~2015
부모+기혼부부+ 자녀(기혼부부의 형제자매 포함)	부부, 자녀, 양친	부부+미혼 자녀+양친
편부모+기혼부부+ 자녀(기혼부부의 형제자매 포함)	부부, 자녀, 편부모	부부+미혼 자녀+부친
그 외 3세대가구	3세대 기타	부부+미혼 자녀+모친 3세대 기타

<표 16> 가구일련번호 부여 방식: 1975~2015년

	전체 가구 일련번호 제공	시도별 가구 일련번호 제공
1975	○	
1980	○	
1985	○	
1990	○	
1995		○
2000		○
2005	○	
2010	○	
2015	○	

6) 인구항목 및 가구주택 항목 연결 방법의 일관성

과거 자료 크기가 제한적인 시기에는 가구 자료와 가구원 자료 곧 개인자료를 혼합해서 제공했다. 이때 자료를 구분하는 방법은 가구 내 가구원 일련번호가 “000”이면 가구자료이고 “001”부터는 가구원 자료라는 점을 감안해 가구원 번호로 먼저 인구와 가구주택 자료를 분리하여 두 개의 데이터 세트로 만드는 것이다. 다음은 이를 다시 두 데이터 세트의 공통키(common key)로 다시 통합하여 개인별로 가구 특성을 부여하여 자료를 분석해야만 했다(Harron, Goldstein, & Dibben 2015),

인구 및 가구주택 데이터세트가 분리해서 제공되는 연도는 1975년부터 2015년까지이다. 가구일련번호는 인구와 가구주택 데이터 세트에 각각 부여하는데, 이 일련번호는 1990년과 2000년을 제외하고 1975년부터 2015년까지는 전체 2%세트에 하나씩 부여함으로써 고유한 키가 되지만, 1995년 및 2000년은 시도가 바뀔 때마다 새로 가구일련번호를 부여해서 이 일련번호가 고유한 키가 되지 못한다. 즉, 시도마다 동일한 가구번호가 있기 때문에 인구와 가구주택 데이터세트를 바로 통합할 수가 없다. 이 경우에는 “시도코드+(시군구코드+읍면동코드)+가구일련번호”와 같이 새로운 변수를 생성해서 사용해야 한다.

이에 따라 이용자의 편의를 위해 최소한 가구번

호라도 고유한 키로 사용할 수 있게 전체 자료 기준으로 일련번호를 제공해야 한다.

4. 결과 및 토의

유엔의 공식통계(official statistics) 기본원칙 중 하나인 “관련성, 공정성 및 동등한 접근 기회 (Relevance, Impartiality and Equal Access)”에 따르면 국가 공식통계는 민주사회의 중요한 공공재로서 사용자의 요구에 부합해야 한다고 한다(United Nations, 2015). 또한 실용성을 확보하기 위해서는 이러한 통계의 관련성과 품질이 적절하고 쉽고 올바른 사용이 용이해야 하며, 이를 위해 다양한 도구를 이용해 사용자의 요구를 지속적으로 파악하고 통계작성 과정에서 지속적으로 사용자와 협의할 것을 권고하고 있다.

많은 국가통계 중에서도 인구주택총조사는 가장 자원이 많이 들어가는 국가 제1의 기본통계조사이다. 1925년부터 90년간 매5년마다 주기적으로 조사를 해 왔기 때문에 숫자로 보는 우리사회의 역사이기도 하다.

이처럼 중요한 국가기본통계가 활용성이 극대화되어야 한다는 것은 너무나 자명한 일이다. 하지만 국민에게 서비스하는 인구센서스 2% 개별자료가 표본조사를 병행했음에도 불구하고, 일부 연도(1975년 및 1980년)는 전수자료에서 추출해서 시리즈를 만들어 활용성이 크게 떨어지고 있다. 1960년부터 주택조사가 병합되어 실시됨에도 불구하고 1960년, 1966년, 1970년은 인구자료만 제공하고 있다.

항목에서는 먼저 조사주기의 일관성이 다소 부족하다. 인구센서스는 90년 이상 조사해 왔기 때문에 시대에 맞게 당시 필요로 하는 항목들을 조사하는 것이 맞기 때문에 항목에 따라서는 어느

정도 시계열 단절은 불가피하다고 본다. 이에 따라 항목의 특성별로 매년 조사해야 하는 항목 10년 주기 항목, 15년 주기 항목, 1회성 항목으로 분류하여 전략적으로 관리해야 한다.

개념 및 조사방법, 접근법에서는 조사통계와 행정자료 기반의 등록센서스 간의 시계열 단절에 대해 지속적인 검토와 연구를 제안한다. 장점이 많아 채택한 조사방법이지만 90년 역사를 가진 인구센서스 자료의 시계열 단절을 최소화해야 한다. 경제활동 항목에 대해서도 본질적인 검토가 필요하다. UN(2017)도 권고하고 대부분 나라에서도 조사하는 경제활동 항목이 월간 경상조사인 경제활동인구조사의 실질적인 모집단으로 활용되기 위해서는, 세부항목 수를 늘려 두 조사의 개념을 일치시킬 필요가 있다.

분류 측면에서는 행정구역이나 산업 및 직업처럼 조사 외적인 요인으로 한국표준분류 자체가 바뀌는 경우에는 불가피하다고 할 수 밖에 없겠지만, 응답 카테고리나 순서 같은 부분은 최대한 일관성/정합성을 유지하여 이용자의 편의성과 활용성을 극대화하는 방안을 찾아보아야 할 것이다.

센서스 조사표에서 명시적으로 조사항목으로 선정하여 실제로 조사한 자료를 이용해 새롭게 만든 파생변수에 대해서도 좀 더 신중하게 관리해야 할 필요가 있을 것이다. 예를 들어 가구원 수 같은 항목은 가구자료에 포함되어 있지만 인구의 특성을 이해하는데도 매우 중요하다. 이런 변수는 이용자가 스스로 만들어 사용하여 이용자 간에 편차가 발생하지 않도록 센서스 통계작성 시점에서 사전에 계획하여 제공할 필요가 있다고 본다. 2015년에 가구원 수 변수가 아예 없는 것도 대단히 아쉽다고 할 수 밖에 없다. 1975년부터 제공하던 핵 수도 핵가족화 등 가족구조를 이해하는데 매우 중요한 항목이다. 이용자가 스스로 계산하기에는 복잡한 내용이기 때문에 파생변수로

사전에 만들어 제공하는 것이 좋은데, 2005년 및 2015년 2% 마이크로데이터에는 없어 이용하기에 불편하다.

마지막으로 2% 센서스 마이크로 자료는 인구와 가구주택 데이터 세트로 구분이 되어 있다. 그런데 인구분석에서 인구의 가구특성이 중요하기 때문에 이용자들이 두 데이터 세트를 통합(merge)해야만 한다. 이때 공통 키로 사용하는 것이 가구 일련번호인데, 이 일련번호가 1995년 및 2000년에는 시도마다 새롭게 시작해서 결과적으로 시도마다 가구번호의 중복이 생겨 키로 사용할 수가 없다. 이 때문에 이용자들이 새롭게 가구 고유 키를 생성(시도번호+가구일련번호)해서 사용하고 있다. 이용자들이 보다 편하게 센서스 자료를 활용할 수 있도록 통계청에서 가구 일련번호 생성 시 배려할 필요가 있다.

통계청에서 인구주택총조사를 실시하면서 이용자의 의견과 청내외 전문가회의를 수차례 거치고

국가통계위원회의 심의를 받는 등 폭 넓게 사용자와 협의를 하고 있는 것으로 안다. 하지만 정작 자료를 직접 핸들링하고 분석하는 이용자의 불편에 대해서는 의견 수렴이 부족한 것으로 보인다. 통계청조차 자료를 이용하기 보다는 작성하고 제공하는 입장이다 보니 자료 이용자의 세부적인 불편과 한계를 충분히 인식하지 못하고 있는 것 같다. 전문가는 물론 일반인을 포함하는 통계이용자들을 대상으로 하여 보다 폭 넓게 이용자의 의견을 수렴할 것을 제안한다.

마지막으로 이 연구에서 파악한 우리나라 인구센서스의 조사방법, 조사항목, 개념 및 정의의 변천과 특징 및 한계점 등 연구결과를 토대로 인구 및 사회학적 맥락에서 인구센서스가 우리나라 사회변화를 어떻게 얼마나 잘 반영해 왔는지 심층 분석할 필요가 있어 이를 후속연구로 제안한다.

참 고 문 헌

- 김민경 (2000). <인구센서스의 이해>. 도서출판 글로벌.
- 김민경 (2020). <한국의 인구주택총조사: 1925-2015>. 통계청.
- 박명규 · 서호철 (2003). <식민권력과 통계-조선총독부의 통계체계와 센서스>. 서울대학교출판부.
- 박계수 (1989). <표본조사법>. 법문사.
- 전광희 (2014). 등록센서스를 위한 보완행정인구의 시도별 규모에 대한 평가와 과제. <충남대학교 사회과학연구소 사회과학연구>, 25(2), 355-378.
- 전광희 (2017). 등록가구와 조사가구의 정합성 개선 방법론 연구. <충남대학교 사회과학연구소 사회과학연구>, 27(1), 339-363.
- 통계청 (2020). <2020년 인구주택총조사 시범예행조사 지침서>.
- 통계청 (각년도). <인구주택총조사 조사표>.
- 한국인구학회 (2016). <인구대사전>. 전면개정판. 통계청.
- Harron, Kattie, Harvey Goldstein, Chris Dibben (2015). *Methodological Developments in Data Linkage*, New York: Wiley.

UN (2015). *United Nations Fundamental Principle of Official Statistics*, Implementation Guidelines: New York.

UN (2017). *Principles and Recommendations for Population and Housing Censuses*, Revision 3, ST/ESA/STAT/SER.M/67/Rev.3: New York.

Wallgren, Anders, Britt Wallgren (2007). *Register-based Statistics: Administrative Data for Statistical Purposes*, New York: Wiley.

Review of the Consistency of Korea's Population Census for Longitudinal Analysis

Ka-Eun Kim, Kwang-Hee Jun

Chungnam National University

Korea has conducted a population census since 1925 every five years to determine the size, structure, distribution, and characteristics of the population. Statistics Korea provides 2% individual census data from 1960 to 2015 and the data are widely used by many people. This paper examined data consistency among censuses in order to satisfy UN's national statistical principles. The evaluation focused on the consistency of microdata, periodicity of items, changes in concepts and methods, approaches, classification, derivative variables, and linkage of population and household items, etc. This study reveals that the 2% individual data of 1975 and 1980, unlike other years, were not based on sample enumeration which contains more items than completed enumeration. Data consistency was partly problematic due to the shift from traditional to register-based census, instability of economic activity items, and frequent changes in item structure. Further, inconvenience emerged in some years when integrating population data with household data. These findings suggest that persistent efforts be made to improve the consistency and coherence of Korea's 2% public use sample data to enhance the utilization of the population census.

Keywords: 2% Public Use Sample Data of Korea's Population Census, Consistency of Population Census, Coherence of Population Census