

주민등록자료와 인구주택총조사의 연령별 비교

서우석 (서울시립대 도시사회학과)

I. 서론

인구주택총조사의 조사환경이 악화되고 비용이 증대하면서 센서스 방식의 개선을 위한 논의들이 확산되고 있다(홍두승외, 2002; 김설희·강계화, 2005; 이명진외, 2006; 이건, 2007). 개선방안의 주요한 논의들 중 하나가 행정자료의 활용이다. 스웨덴, 핀란드, 덴마크 등 북구의 여러 나라들은 이미 1980년대부터 센서스에 행정자료를 활용하기 시작했으며 이제 별도의 실사를 수행하지 않는 단계에 이르렀다(홍두승외, 2002).

행정자료를 활용하여 센서스를 대체하는 경우 전 국민 대상의 조사에 수반되는 막대한 경제적 비용을 줄일 수 있다. 또한 행정정보의 통합관리에 따른 개인정보 보호를 위한 기술적 해결과 사회적 합의를 달성하면 조사 과정에서 나타나는 사생활 침해의 문제로 인한 사회적 갈등도 방지할 수 있다. 그리고 기존의 인구센서스에 비해 자료산출주기를 축소하고 자료수집에서부터 공표까지의 기간을 단축시킴으로써 자료의 시의성 측면에서도 이점을 기대할 수 있다(홍두승외, 2002; 이건, 2007). 또한 행정자료에 의한 센서스 대체가 아니라도 행정자료의 활용에 의하여 센서스의 효율성을 제고하는 것이 가능하다.

이 연구는 주민등록자료와 인구주택총조사의 연령별 비교를 통하여 주민등록자료의 활용가능성을 분석하고자 한다. 주민등록자료에는 인구주택총조사의 조사 항목과 중복되는 다수의 항목들이 있다(홍두승외, 2002; 김설희·강계화, 2005). 주민등록자료는 전체 국민대상을 포괄성하며 모든 국민에게 고유번호를 부여한다는 점에서 통계적 활용가능성의 이점을 갖고 있다.

하지만 주민등록의 활용가능성을 평가하기 위해서는 다음의 제한점들을 고려해야 한다. 우선 주민등록과 인구주택총조사는 대상 범위가 다르다. 인구주택총조사가 실제 거주인구를 조사하는데 비해 주민등록은 주소지 등록 인구를 대상으로 삼는다. 주민등록인구가 여행, 군 복무, 학업의 이유로 주소지를 벗어나는 경우 주민등록과 인구주택총조사 사이에는 차이가 발생한다.

다음으로 주민등록과 인구주택총조사 모두 오차를 안고 있어서 거주인구수에 대한 진실값을 알 수 없다는 점이다. 주민등록은 거주지 변동 후 미신고나 거짓신고, 주민등록 말소, 주민등록 미발급, 행정오류에 기인한 오차를 갖고 있다¹⁾. 인구주택총조사의 경우 조사대상의 누락이나 부정확한 정보 수집으로 인한 오차가 발생한다. 하지만 현실적으로는 주민등록과 인구주택총조사의 차이가 주민등록자료에 의한 인구주택총조사의 대체가능성을 평가하는 지표가

1) 이와 같은 주민등록의 오차를 줄이기 위해서 행정자치부는 주민등록 일제정리 사업을 수행하고 있다.

된다(이건, 2007). 주민등록과 인구주택총조사 사이의 비교에서 차이가 적을수록 주민등록의 활용가능성이 높은 것으로 평가된다. 또한 차이가 존재하더라도 원인 추정과 보정이 가능하다면 주민등록의 활용가능성을 높일 수 있다.

본 연구에서는 주민등록과 인구주택총조사의 차이를 연령별로 비교함으로써 주민등록의 활용 가능성을 평가하고 활용 가능성을 제고할 수 있는 방안을 모색하고자 한다. 이와 같은 목적을 위하여 주민등록과 인구주택총조사를 비교하는 과정에서 연령별 비교가 가지는 의의는 다음과 같다.

첫째, 주민등록과 인구주택총조사의 차이를 추정하고 주민등록의 통계적 활용가능성을 평가한다. 연령별 인구 집계는 인구통계의 핵심 항목이다. 주민등록을 통한 각 연령층 인구 집계와 인구주택총조사의 연령별 인구 집계 사이의 차이는 그 자체로서 주민등록을 통계적으로 활용할 수 있는 정도를 판단하는 중요한 지표가 된다.

또한 주민등록과 인구주택총조사 사이의 차이에 대한 연령별 및 지역수준별 비교 결과를 종합적으로 고려함으로써 주민등록의 활용 가능성을 더욱 정확히 추정할 수 있다. 연령대별 차이율을 고려하지 않고 전체 수준의 차이율만을 기준으로 삼을 경우 주민등록과 인구주택총조사 사이의 차이를 과소평가할 위험이 있다. 연령대나 지역별 차이가 서로 다른 방향으로 나타나서 차이들이 상쇄될 수 있기 때문이다.

둘째, 주민등록과 인구주택총조사 사이의 연령대별 차이를 지역수준별로 파악함으로써 오차 발생 현황과 원인에 대한 분석이 가능해진다. 연령대별 차이를 지역별로 비교하고 연령대별 차이율이 높은 지역들의 특성을 파악함으로써 오차의 발생원인을 어느 정도 파악하고자 한다.

II. 선행연구

선행연구는 크게 보아 두 가지로 구분된다. 하나는 행정자료의 인구주택총조사 활용의 필요성이나 방안을 논의하는 연구들이다. 다른 하나는 행정자료를 인구주택총조사와 실제 비교하여 분석한 연구들이다.

먼저 행정자료의 도입 방안에 대한 연구로서 홍두승외(2002)는 현재 센서스의 문제점을 제시하고 유럽 국가들의 행정자료 활용 실태를 소개하였다. 이 연구는 행정자료의 활용이 주민등록자료에서부터 단계적으로 실행하는 것이 적합하다는 의견을 제시하였다. 전산화가 완료되었고 개인의 비밀정보가 비교적 적기 때문에 개인정보 보호 측면에서도 사용이 덜 제한된다는 점을 지적하였다.

김설희·강계화(2005)는 행정정보와 GIS를 연계하여 통계생산에 활용하는 방안을 논의하면서 인구주택총조사의 항목들 중에서 행정자료 활용이 가능한

항목들을 제시하였다. 주민등록자료를 통해 정보 획득이 가능한 항목들은 성명, 가구주와의 관계, 성별, 연령, 혼인상태, 가구구분, 5년전거주지, 총출생아수였다.

다음으로 행정자료를 인구주택총조사와 실제 비교한 연구로서 김근영외(2003)는 1995년과 2000년의 인구주택총조사의 인구·가구 자료와 주민등록인구통계의 인구·세대자료를 비교하였다. 비교 결과 두 자료 사이에는 상당한 정도의 차이가 있는 것으로 보고되었다. 이러한 차이에 대한 설명으로서 위장전입에 의한 주민등록 과잉, 일시적 거주인구 유입에 따른 주민등록 과소, 역외거주인구의 센서스 대상 제외가 원인으로 제시되었다.

이명진외(2006)는 센서스 시험조사 자료를 바탕으로 가구명부-거처정보 자료와 건축물 대장의 일치 여부를 분석하였다. 주택유형과 건축물정보의 상용관계에 따라서 상이한 매칭 유형이 나타나며 이러한 매칭 유형에 따라서 건축물 대장 정보의 활용성이 상이하게 평가되었다. 매칭의 어려움은 특히 구주소 체계에서 가구와 건물에 대한 확인이 어려워 건축물대장과 조사자료를 일치시키기 힘들다는 사실에 있었다.

인구부문에 대한 분석을 수행한 연구로서 이건(2007)은 읍면동 수준에서 주민등록과 인구주택총조사의 집계자료를 비교하였다. 전국적으로 주민등록으로 집계된 인구수는 인구주택총조사에 의해 집계된 인구수 보다 약 174만명이 많았고 이 차이는 전체 인구수의 약 3.6%에 해당했다. 이는 1995년의 3.1%보다는 늘었고 2000년의 4.0%보다는 줄어든 것이다. 지역단위별로 비교하였을 때 대지역보다 소지역에서 차이율이 커졌으며, 이는 소지역에서의 오차가 서로 상쇄되는 효과가 발생한 것으로 파악되었다.

직접적으로 주민등록자료의 활용 가능성에 초점을 맞춘 것은 아니지만 이지연(2007)은 2005년도 인구주택총조사의 오차율을 연령별 지역별로 추정하기 위하여 주민등록 자료를 활용하였다. 주민등록인구를 기준으로 총조사에서 가장 많이 누락된 최대 누락집단은 35-39세 연령층이었다.

III. 분석

1. 자료 및 분석방법

이 연구에서는 2005년도 인구주택총조사의 집계자료와 이에 근접한 시점으로서 2005년 12월 31일 기준의 주민등록 집계자료를 비교하였다¹⁾. 연구에서 사용한 인구주택총조사와 주민등록의 연령별 집계자료는 모두 통계청이 홈페이지

1) 인구주택총조사와 주민등록자료 사이에는 2개월 사이의 시차가 존재한다. 이에 대한 논의로는 이건(2007) 참조.

이지(<http://www.nso.go.kr>)에 공표한 읍면동 수준의 인구집계이다.

이 연구에서는 연령별 비교를 전국, 광역시도, 시군구, 읍면동의 4가지 지역 수준에서 분석하였다. 통계청이 제시한 연령별 집계는 5세 연령층을 단위로 제시되었다. 이 연구에서도 5세 연령층을 기본 단위로 연령별 비교를 수행하였다. 모든 지역수준에서 연령층의 단위를 동일하게 사용하였으나 읍면동 수준에서는 70세 이상의 연령대를 통합하여 측정하였다. 각 지역 내 해당 연령대 인구 규모가 적기에 약간의 변화로도 차이율의 값이 크게 영향을 받기 때문이다.

주민등록과 인구주택총조사 집계자료를 비교하는 통계로서 ‘차이율’과 ‘연령대 절대차이율 평균’을 사용하였다. 차이율의 산식은 다음과 같다.

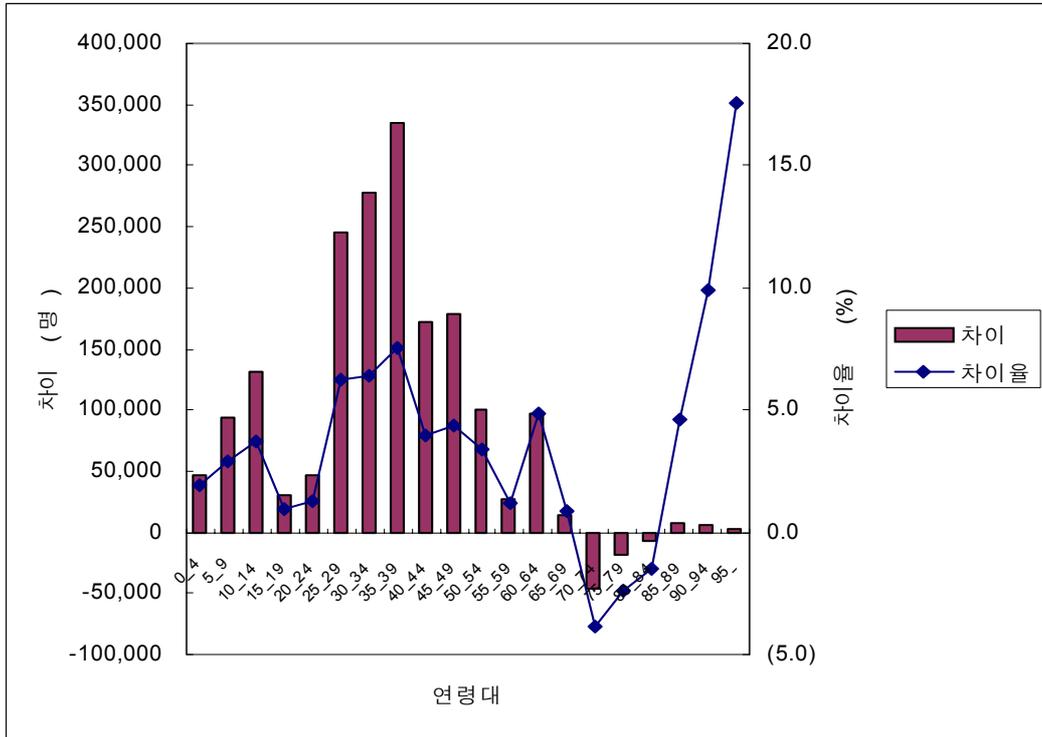
$$\text{차이율} = \{(\text{주민등록인구} - \text{총조사인구}) / \text{주민등록인구}\} * 100$$

그런데 연령대 차이율의 산술평균값은 각 연령대간의 상쇄효과가 있을 경우 차이의 규모를 정확히 반영할 수 없다. 이러한 문제를 해결하기 위해 각 지역의 연령대 차이율의 절대값을 평균한 연령대 절대차이율 평균을 제시하였다.

2. 전국 수준의 연령대별 차이율과 지역수준별 비교

전국 수준에서 주민등록인구와 총조사인구의 차이를 연령대별로 비교하면, <그림 1>에서와 같이 차이율이 연령대의 구간 구간별로 일정한 패턴을 가지는 것을 볼 수 있다. 차이율이 연령대에 따른 특정한 유형을 갖고 있다는 사실은 차이의 발생 원인에 대한 해석의 가능성을 제시한다.

<그림 4-1> 연령대별 차이와 차이율



연령대별 주민등록인구와 총조사의 차이를 보았을 때 다음과 같은 사실들을 발견할 수 있다. 우선 전체 인구수준에서 주민등록인구와 총조사인구를 비교할 때 차이율이 3.6%인데 연령대별로 차이율의 편차가 크게 나타나서 4% 대 이상의 높은 차이율을 보이는 연령대도 적지 않다. 또한 주민등록인구보다 총조사인구가 더 많은 연령층도 있어서 연령대들 사이에서 차이율의 상쇄효과가 발생한다.

전체적으로 생애주기에 따른 사회경제적 활동 및 인구이동의 변이와 인구주택총조사의 조사 가능성에 의해서 영향을 받은 것으로 해석된다. 예외적이라면 15-24세까지 연령기의 높지 않은 차이율이다. 학업을 수행하고 사회적 진출을 준비하는 기간의 연령층이지만 총조사를 통한 인구추정과 주민등록인구수가 비교적 근접하였다. 이후 20대 후반부터 30대까지는 경제적 사회적 활동이 가장 왕성하고 공간적 이동이 많은 시기로서 주민등록과 인구주택총조사의 차이가 크게 나타났다. 이 집단은 직장 근무지 때문에 발생하는 단기 이주, 유학이나 취업에 따른 해외 거주 등의 이유로 인해서 가구면접조사의 형태로 이루어지는 인구주택총조사에서 조사되지 않았을 가능성이 크다. 이와 같은 해석은 이동성이 강한 연령층에서 인구주택총조사의 오차가 상대적으로 크게 나타난다는 이지연(2007b)의 연구 결과에 의해서 뒷받침된다. 40세 이후의 연령대에서는 차이율이 낮아지는데 이는 청년층의 높은 차이율을 발생시킨 여러 가지 원인들이 약화된 결과로 해석할 수 있다.

노년층에서는 상반된 방향의 차이율이 발생하였다. 부의 차이율이 발생한 원인으로는 과거 행정등록의 미비로 인해 주민등록집계에 잡히지 않는 인구층의 존재를 추정할 수 있다. 반면 정의 차이율이 발생한 원인은 사망시 신원불명 등으로 인하여 주민등록이 정확하지 못하거나 복지시설 및 의료시설 수용으로 인해 조사 응답여건이 안 되어 총조사에서 제외된 경우를 추정할 수 있다.

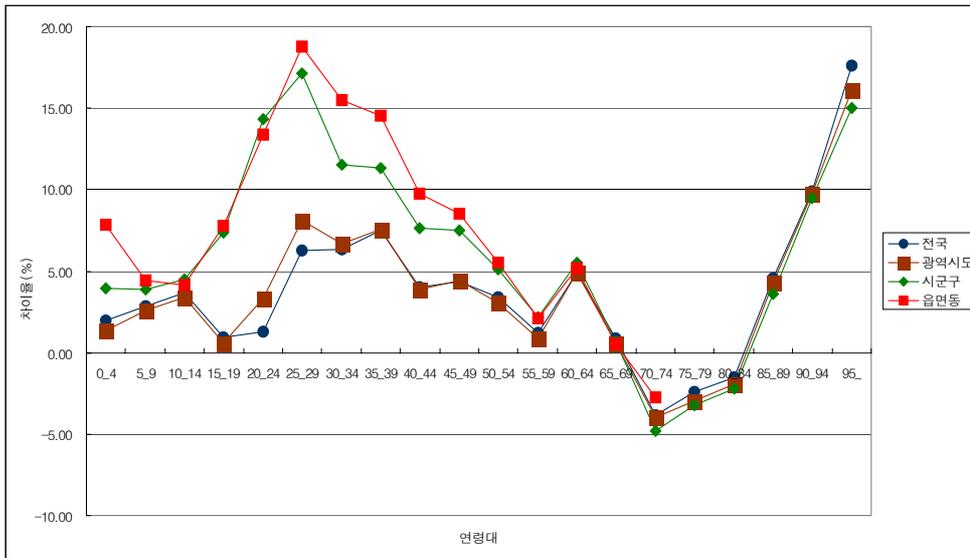
다음으로는 지역수준별로 차이율의 평균과 표준편차를 모아서 비교하였다. <그림 2>는 4가지 지역수준별로 연령대별 주민등록인구와 총조사인구의 차이율의 평균을 제시하였고 <그림 3>은 3가지 지역수준별로 연령대별 주민등록인구와 총조사인구의 차이율의 표준편차를 제시하였다.

연령대별 차이율 평균을 보면 연령대별 패턴 자체는 모든 지역단위 수준에서 유사하지만 규모는 지역수준별로 차이가 뚜렷했다. 이건(2007)의 결과와 유사하게 광역단위에서 소지역단위로 갈수록 차이율이 더 커졌다. 특히 광역시도와 시군구 수준 사이에서 차이율이 크게 차이났다.

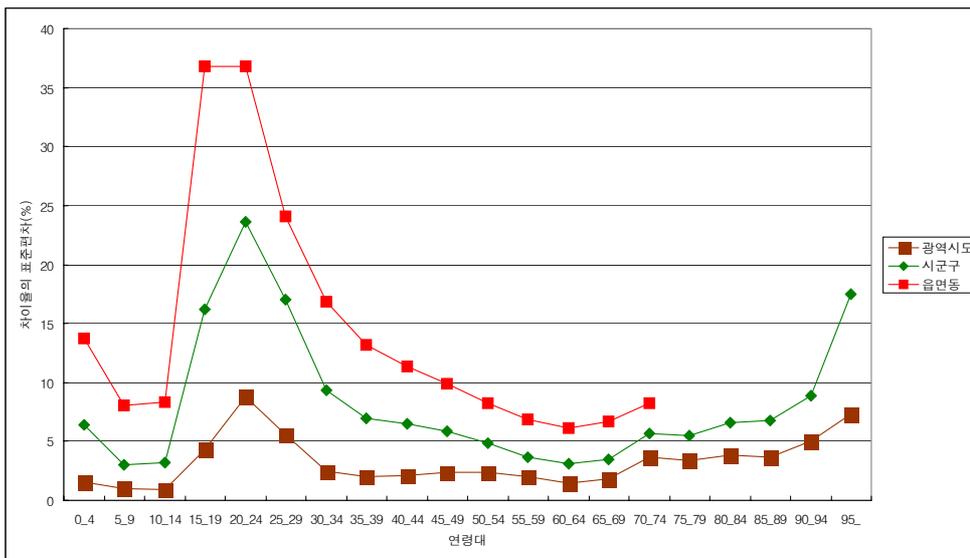
대체로 전국 수준에서 차이율이 높은 연령대들에서 지역수준별 차이도 컸다. 반면 15-19세와 20-24세의 연령대는 이러한 경향의 예외로서 주목할 만하다. 전국 수준이나 광역시도에서는 차이율이 높지 않았는데 시군구, 읍면동 수준에서는 차이율이 상당히 높았다. 전국수준의 차이율과 소지역 수준의 차이율 사이에 나타난 격차는 인구가동시 발생한 주민등록과 실주거지 사이의 괴리에서 발생한 것으로 보인다. 이 경우 특정 소지역에서 나타난 차이가 다른 소지역의 차이와 서로 상쇄되면 광역수준에서는 차이율이 축소된다.

연령대별 차이율의 표준편차를 보면 소지역수준으로 갈수록 표준편차가 컸다. 특히 15-19세, 20-24세의 연령대에서 표준편차가 가장 컸다. 반면 전국 수준에서 차이율이 가장 높았던 35-39세, 30-34세의 연령대에서는 차이율의 표준편차가 상대적으로 적었다. 이러한 결과는 차이의 지역별 분포가 연령대에 따라서 다른 형태로 나타난다는 사실을 의미한다. 30대의 경우 주민등록과 총조사인구의 차이가 전국적으로 비교적 균일하게 높은 반면, 반면 10대 후반에서 20대 초반의 차이는 지역별로 큰 편차를 가졌다.

<그림 2> 지역단위별 연령대별 차이율 평균 비교



<그림 3> 지역단위별 연령대별 차이율 표준편차 비교



3. 광역시도 비교 분석

<표 1>에서는 각 광역시도별로 차이율과 연령대 절대차이율 평균값을 제시하였다. 광역시도별로 비교해 보면 차이율이 높은 지역으로는 전라남도가 7.7%로 가장 높았고 다음으로 전라북도(5.7%), 제주도(4.8%), 충청남도(4.2%) 등이었다. 차이율이 낮은 지역으로는 광주광역시(-0.9%), 대전광역시(1.1%), 충청북도(1.8%) 등이 있었다. 서울특별시와 부산광역시는 각각 4.0%, 3.5%로서 평균적인 수준의 차이율을 보였다.

연령대 절대차이율 평균은 차이율과 상당히 다른 분포를 보였다. 전라남도처럼 차이율과 연령대 절대차이율 평균이 모두 높은 경우도 있었다. 하지만 광주광역시와 대전광역시의 경우 차이율은 전국에서 가장 낮은 수준이었으나 연령대 절대차이율 평균은 모두 4.5% 대 이상을 기록하여서 다른 지역과 거의 같은 수준이었다. 연령대별로 상이한 방향의 차이가 발생하여 서로 상쇄됨으로써 지역의 차이율이 낮게 나타난 것이었다. 이러한 사례들은 차이율이 낮다고 해서 주민등록의 활용가능성이 다른 시도보다 높다고 평가할 수 없음을 보여준다.

<표 1> 연령대별 차이율 평균 및 절대차이율 평균의 시도별 비교

행정구역명	주민등록 인구수	총조사 인구수	차이율 (%)	연령대 절대차이율 평균 (%)
서울특별시	10,167,344	9,762,546	3.98	6.38
부산광역시	3,638,293	3,512,547	3.46	4.25
대구광역시	2,511,306	2,456,016	2.20	4.14
인천광역시	2,600,495	2,517,680	3.18	4.80
광주광역시	1,401,745	1,413,644	(0.85)	4.53
대전광역시	1,454,638	1,438,551	1.11	4.68
울산광역시	1,087,648	1,044,934	3.93	4.48
경기도	10,697,215	10,341,006	3.33	4.34
강원도	1,513,110	1,460,770	3.46	6.08
충청북도	1,488,803	1,453,872	2.35	4.24
충청남도	1,962,646	1,879,417	4.24	5.04
전라북도	1,885,335	1,778,879	5.65	6.22
전라남도	1,967,205	1,815,174	7.73	8.11
경상북도	2,688,491	2,594,719	3.49	5.19
경상남도	3,160,431	3,040,993	3.78	4.49
제주도	557,569	530,686	4.82	4.43

<그림 4>와 <그림 5>를 보면 약간의 예외들을 제외하고 대체로는 광역시는 광역시끼리, 도는 도끼리 서로 유사한 연령대 차이율의 유형을 가진다. 광역시와 도 사이에 특히 두드러진 차이를 보이는 연령대는 15-29세와 70-84세의 연령대였다.

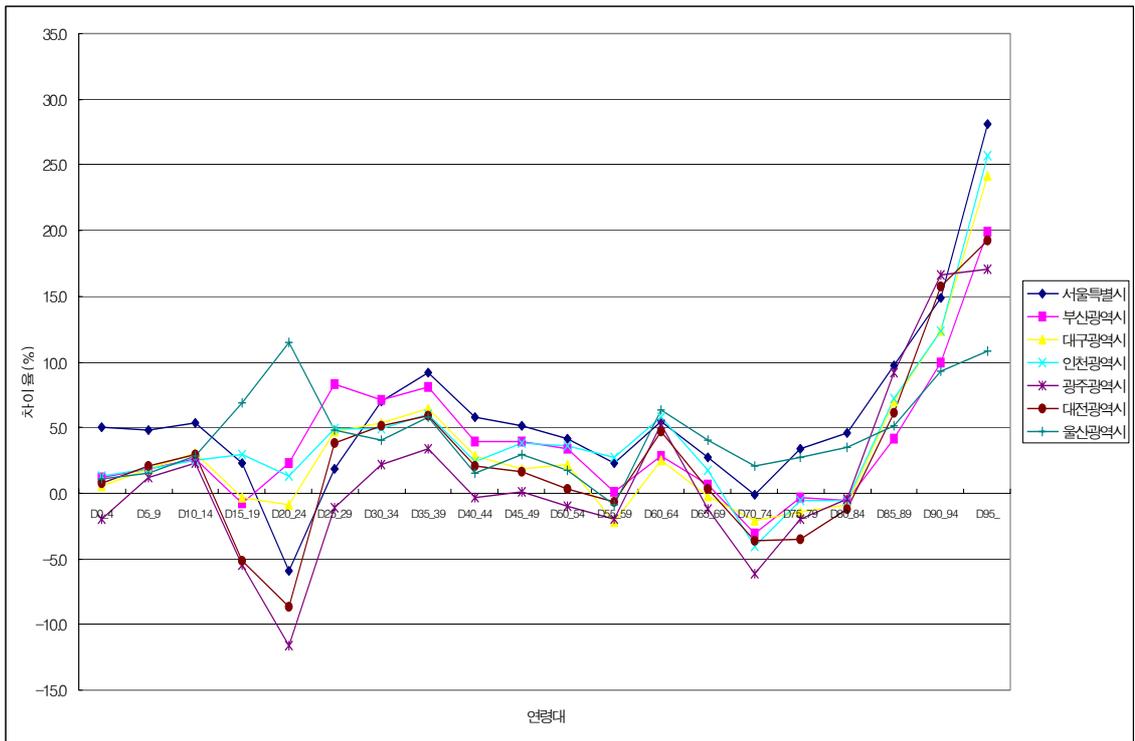
15-29세의 경우 광역시에서는 울산을 제외하고 부의 차이율을 보이거나 차이율의 절대값이 그리 크지 않았던 반면, 도에서는 경기도를 제외하고 매우 높은 정의 차이율을 보였다. 광역시에서는 상대적으로 젊은 연령층이 주민등록보다 많고 도에서는 젊은 연령층이 주민등록보다 적은 것이다. 이러한 현상

은 젊은 층이 농촌지역과 중소도시로부터 대도시로 이동한 결과로 보인다.

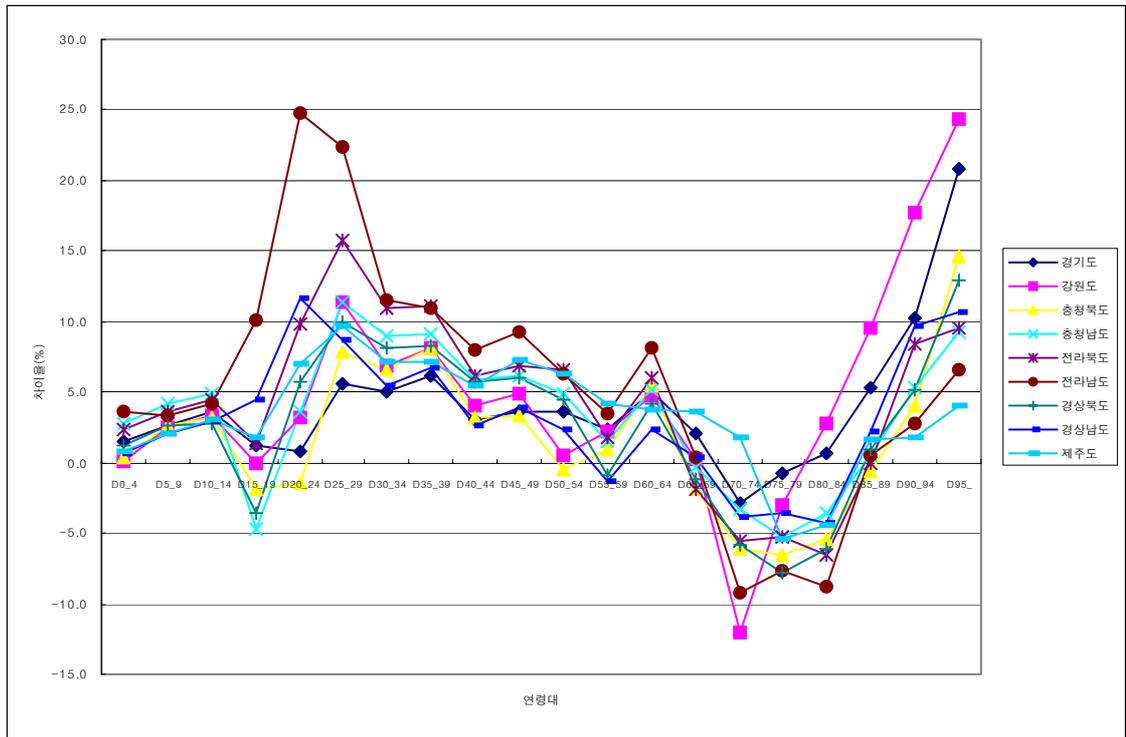
70-84세의 경우 광역시에서는 차이율의 절대값이 그리 크지 않았으나 도에서는 대체로 매우 낮은 부의 차이율을 보였다. 이들 노년층에서 주민등록보다 총조사인구가 더 많은 것은 행정등록의 미비나 주민등록 관리 소홀에 따른 것이며 이러한 문제가 도시지역보다 농촌지역에서 더 심한 것이다.

상대적으로 차이율이 높았던 30대나 80대 중반 이후의 경우에는 시도간 정도의 차이는 있지만 뚜렷한 유형의 차이는 나타나지 않았다.

<그림 4> 연령대별 차이율 : 광역시 비교



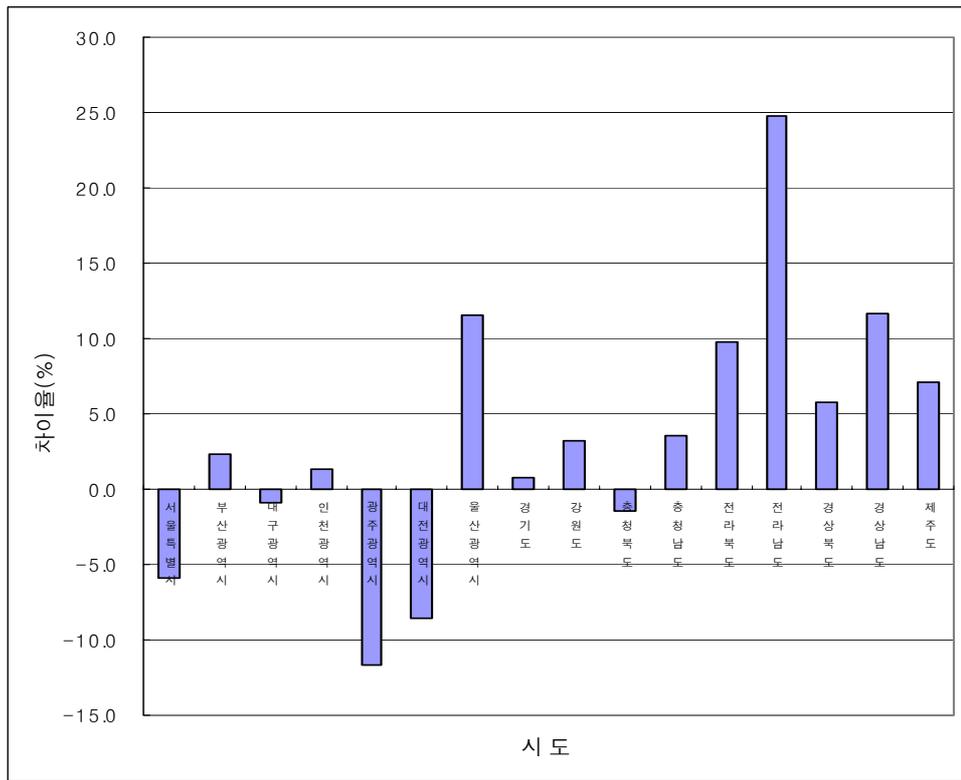
<그림 5> 연령대별 차이율 : 도 비교



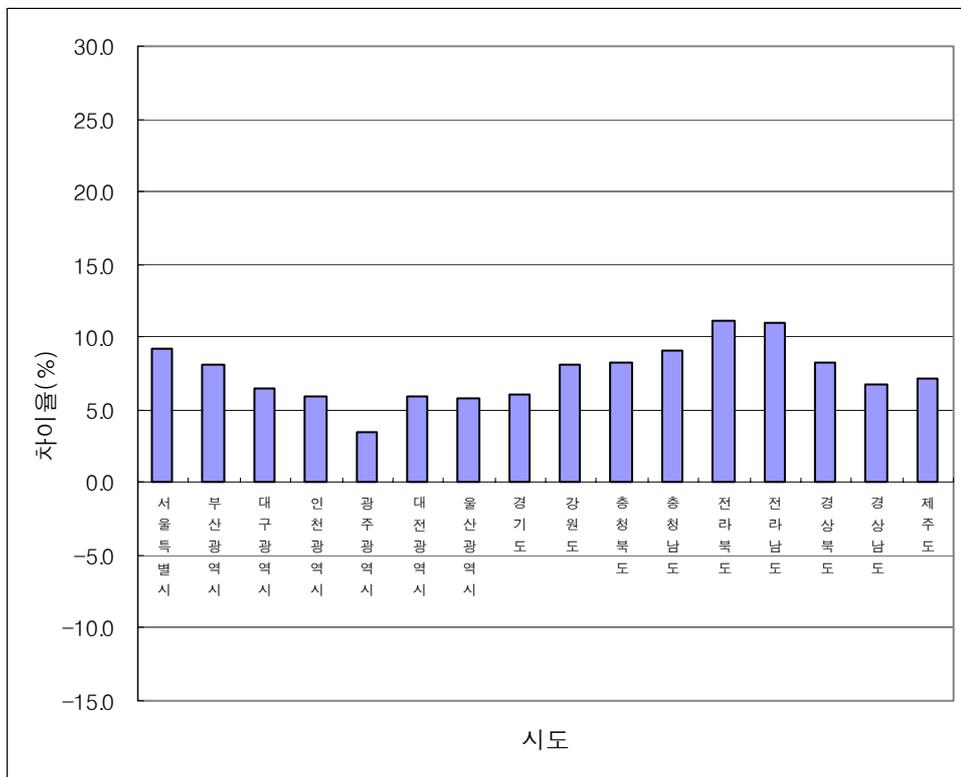
<그림 6>부터 <그림 8>은 20-24세, 35-39세, 70-74세의 연령층에서 광역시도별 차이율을 비교하였다. 이는 앞서 <그림 4>와 <그림 5>에서 제시한 시도간 비교 결과를 연령급간으로 세분하여 더욱 명확히 비교하기 위한 것이다. 분석의 대상이 된 연령대들 중에서 차이율의 시도간 차이가 가장 두드러진 것은 20-24세 연령대였다. 20-24세 연령대에서 주민등록인구와 총조사인구의 차이율을 보면 광역시도에 따라서 큰 차이가 났다. 차이율이 가장 높았던 지역은 전라남도였다. 반면 차이율이 최소값을 보인 지역은 광주광역시였다. 인접한 두 지역사이에 서로 상반된 방향의 차이율이 나타나는 현상은 전라남도 출신 청년층이 광주로 이주한 결과로 추정된다. 마찬가지로 현상이 규모는 작지만 충청남도와 대전광역시의 사이에서 나타나는 것으로 볼 수 있다.

70-74세 연령대에서는 대부분의 광역시도에서 총조사인구가 주민등록 보다 많았지만, 대체로 광역시 보다 도에서 차이율의 정도가 더 컸다. 농촌지역에서 주민등록의 정확성이 도시 지역 보다 더 미흡한 점을 고려할 때 이 연령대 차이율의 원인이 주민등록의 부정확성일 것이라는 추정을 뒷받침한다.

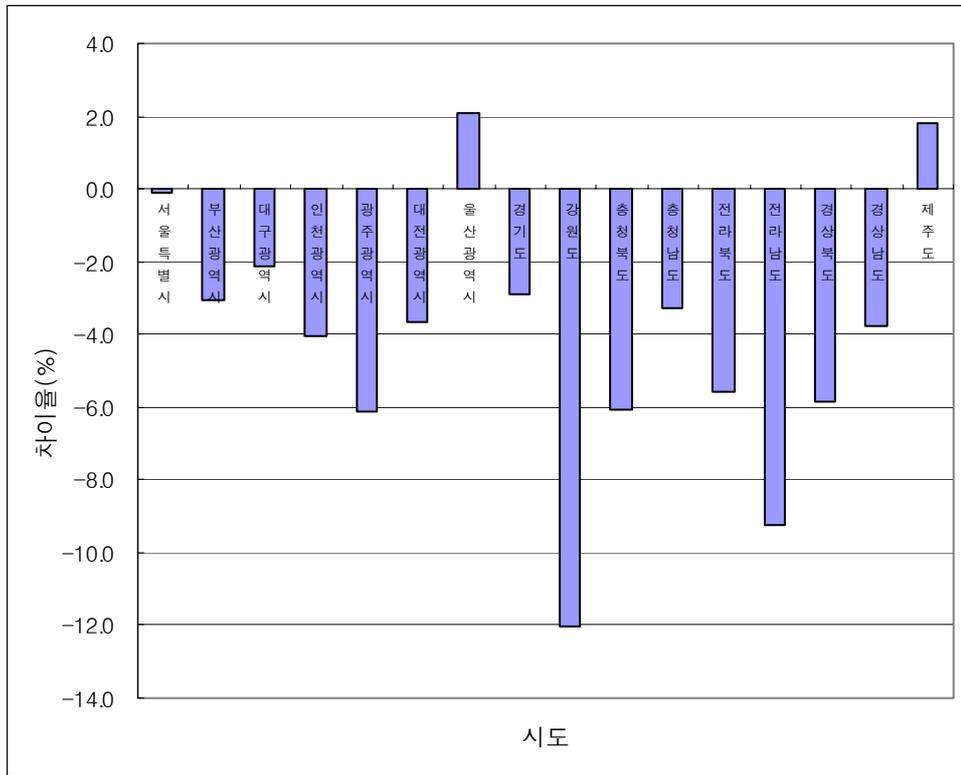
<그림 6> 20-24세 연령대 주민등록인구와 총조사인구 차이율



<그림 7> 35-39세 연령대 주민등록인구와 총조사인구 차이율



<그림 8> 70-74세 연령대 주민등록인구와 총조사인구 차이율



3. 시군구 및 읍면동 수준 분석

<표 2>와 <표 3>은 시군구와 읍면동 수준에서 연령대 절대차이율 평균 기준 상위 10개 지역의 조사 결과이다. 연령대 절대차이율 평균이 가장 큰 상위 10개 지역은 주민등록을 통계자료로서 활용하기 가장 곤란한 지역이다. 시군구의 경우 연령대 절대차이율 평균이 가장 높은 지역은 전라북도 임실군이었고, 다음으로는 인천광역시 옹진군, 경상남도 합천군, 강원도 평창군, 전라북도 순창군 등의 순서였다. 상위 10개 시군구는 모두 농어촌지역이었고 전라북도 지역이 4개로 가장 많았다. 모두 높은 수준의 정의 차이율을 보이는 지역들이었다.

반면 상위 10개 읍면동은 시군구 차원의 상위 10개 지역과 다른 양상을 보였다. 시군구 수준과는 달리 읍면동 수준의 상위 지역은 상위 10개 지역들 중에서 정의 차이율을 보인 곳이 2개 지역이었고 나머지 8개 지역이 부의 차이율을 보였다. 연령대 절대차이율 평균 기준의 상위 10개 읍면동 다수에서 총조사인구가 주민등록인구보다 많았다.

<표 2> 연령대 절대차이율 평균 기준 시군구별 차이율 : 상위 10개

시도	시군구	주민등록인구	총조사인구	차이율 (%)	연령대 절대차이율 평균(%)
전라북도	임실군	32,511	25,682	21.0	22.7
인천광역시	옹진군	15,609	12,261	21.4	22.2
경상남도	합천군	58,162	47,651	18.1	20.1
강원도	평창군	45,033	38,507	14.5	19.3
전라북도	순창군	32,012	26,250	18.0	19.2
충청북도	괴산군	38,595	32,208	16.5	18.0
전라북도	진안군	29,021	23,915	17.6	17.6
전라남도	신안군	46,451	38,429	17.3	17.5
전라북도	장수군	24,755	20,486	17.2	17.5
경상북도	의성군	64,457	56,522	12.3	17.0

<표 3> 연령대 절대차이율 평균 기준 읍면동 차이율 : 상위 10개

시도	시군구	읍면동	주민등록 인구	총조사 인구	차이율 (%)	연령대 절대차이율 평균(%)
강원도	원주시	홍업면	6402	10703	(67.2)	73.7
충청북도	충주시	달천동	5124	8699	(69.8)	69.2
인천광역시	서구	석남1동	13913	23031	(65.5)	64.4
서울특별시	성북구	월곡3동	702	233	66.8	63.8
경상북도	구미시	공단2동	4491	7255	(61.5)	55.4
경상남도	합천군	덕곡면	1713	909	46.9	54.3
경상북도	구미시	진미동	15504	22954	(48.1)	51.2
전라남도	목포시	충무동	3744	4138	(10.5)	50.6
부산광역시	영도구	동삼2동	4972	6812	(37.0)	49.6
경기도	안성시	대덕면	8922	14175	(58.9)	49.1

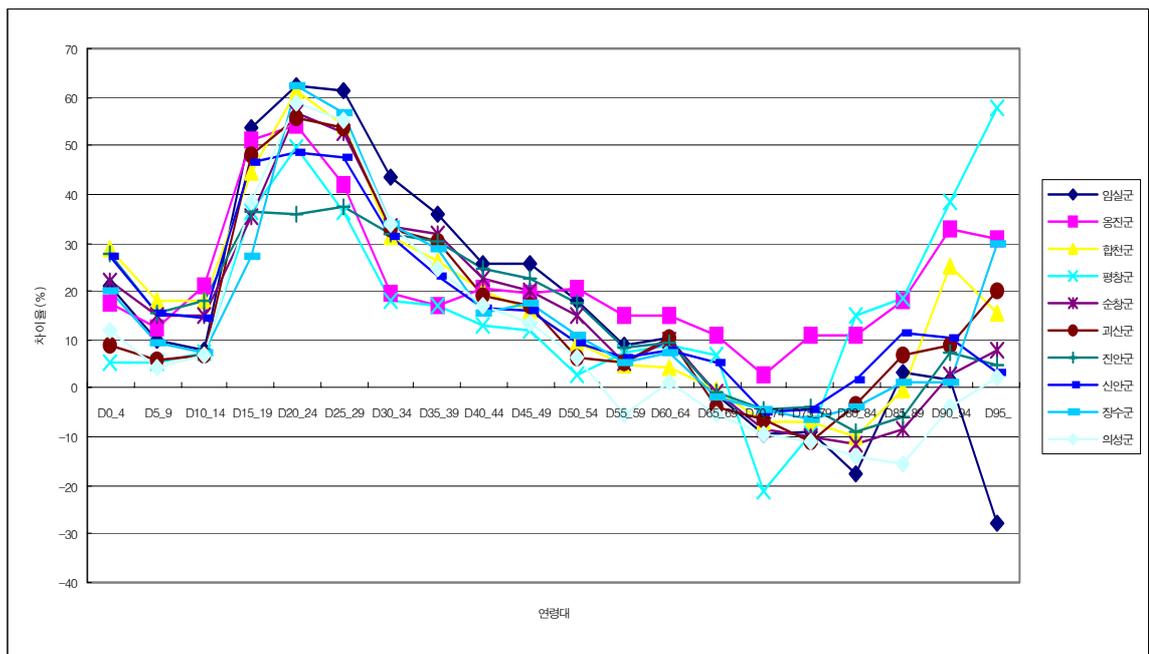
<그림 9>와 <그림 10>은 절대차이율 평균 기준 상위 10개 시군구와 읍면동의 연령대별 차이율을 제시하였다. 10개 시군구는 80대 이상의 고연령대를 제외하고는 대체로 유사한 형태였다. 특히 15세 이상 29세의 연령대에서는 모두 30%대 이상의 높은 차이율을 보였다. 이는 10대 후반과 20대가 주민등록 자료를 통계자료로서 활용하기 가장 어려운 문제 연령대라는 사실을 재확인해 준다.

상위 10개 읍면동 지역의 연령대별 차이율에서는 세 가지 유형을 구별할 수 있다. 10개의 읍면동 중 7개 지역은 15-19세, 20-24세의 연령대 차이율이 매우 낮은 경우였고, 2개 지역은 전체 연령대에서 차이율이 비교적 고르게 높은

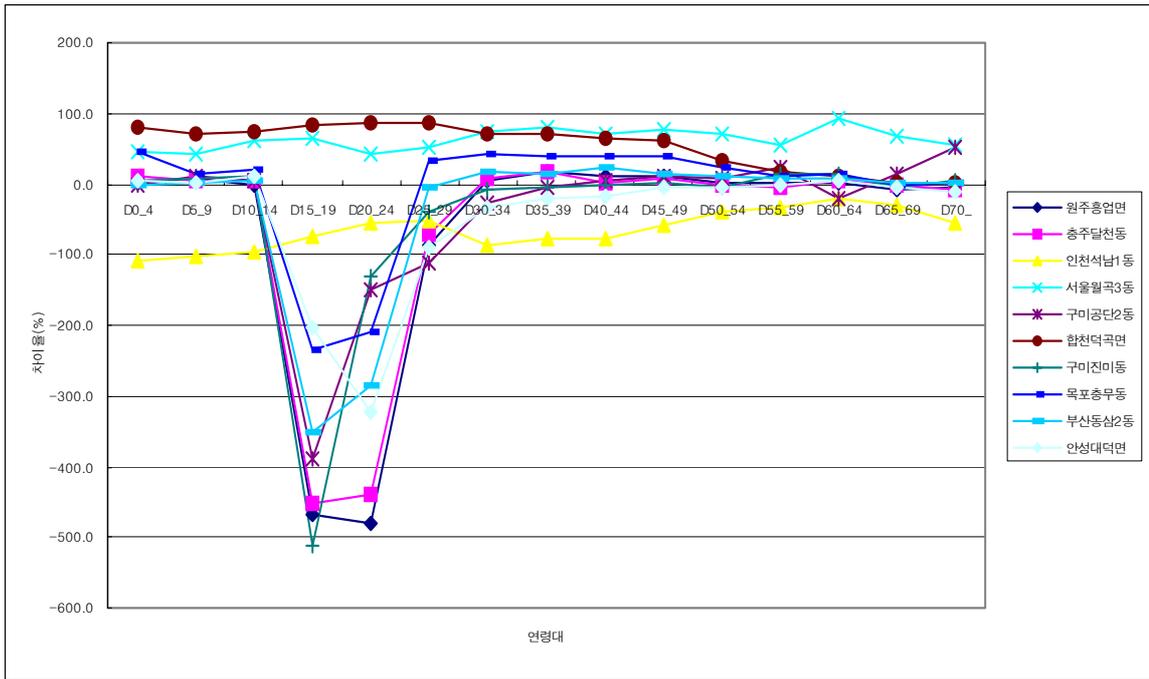
경우였으며, 나머지 1개 지역은 전체 연령대에서 차이율이 비교적 고르게 낮은 경우였다.

먼저 15-19세, 20-24세 연령대의 차이율이 매우 낮은 지역들을 보면 모두 대학 소재지와 공단 소재지의 특성을 갖고 있다. 원주시 흥업면, 충주시 달천동, 안성시 대덕면은 대학캠퍼스의 소재지로서 대학 타운이 형성되어 있다. 부산광역시 영도구 동삼2동과 목포시 충무동은 어촌 및 관광자원과 해양대라는 교육기관의 요소가 결합되어 있다. 구미시의 공단2동과 진미동은 구미시에 소재한 공단의 영향권에 속하였다. 다음으로 전체 연령대에서 차이율이 고르게 높았던 서울의 월곡동은 재개발 사업으로 인해서 주민수가 크게 줄어든 상황이며 그 결과 주민등록인구가 실거주자보다 전체 연령층에서 많은 것으로 보인다. 합천군 덕곡면의 경우는 재개발 지역은 아니나 전체적으로 개발이 부진하였다. 끝으로 전체 연령대에서 차이율이 고르게 낮았던 인천의 석남1동은 인근의 공업지역과 재개발이 복합적으로 관련된 것으로 보인다.

<그림 9> 절대차이율 평균 기준 상위 10개 시군구별 연령대별 차이율 꺾은선 그래프



<그림 10> 절대차이율 평균 기준 상위 10개 읍면동 연령대별 차이율 꺾은선 그래프



이와 같은 문제 연령대의 지역 분포에 대한 이해를 위하여 <표 4>에서는 20-24세 연령대를 기준으로 차이율이 가장 높은 시군구와 가장 낮은 시군구를 10개씩 제시하였다. 상위 시군구는 20-24세 연령대에서 총조사인구에 비해 주민등록이 훨씬 많은 경우이고 하위 시군구는 20-24세 연령대에서 주민등록에 비해 총조사인구가 훨씬 많은 경우이다.

상위 10개 시군구 중에서는 전라북도가 4개군이나 되어 가장 많았고 다음으로 전라남도와 경상북도가 각각 2개군이었다. 대부분 농어업 위주의 산업구조를 가지고 있는 농어촌지역들이었다. 또한 20-24세 연령대의 차이율이 높은 지역은 인접한 15-19세와 25-29세 연령대에서도 차이율이 높았다. 이는 젊은 연령대의 높은 차이율이 특정 연령급에 한정된 현상이 아니라라는 사실을 의미한다.

하위 10개 시군구들은 전국적으로 고르게 분포하였는데 대부분 대학이나 공단이 소재한 지역이었다. 이는 대학이나 공단에 유입된 젊은 층으로 인하여 주민등록과 인구주택총조사 사이에 큰 차이가 발생한 것임을 시사한다. 상위 10개와 하위 10개 시군구의 지역분포를 종합적으로 고려하면 젊은 층의 높은 차이율이 진학과 취업에 따른 인구이동에 의한 것으로 해석할 수 있다.

<표 4> 20-24세 연령대 차이율 기준 시군구별 차이율

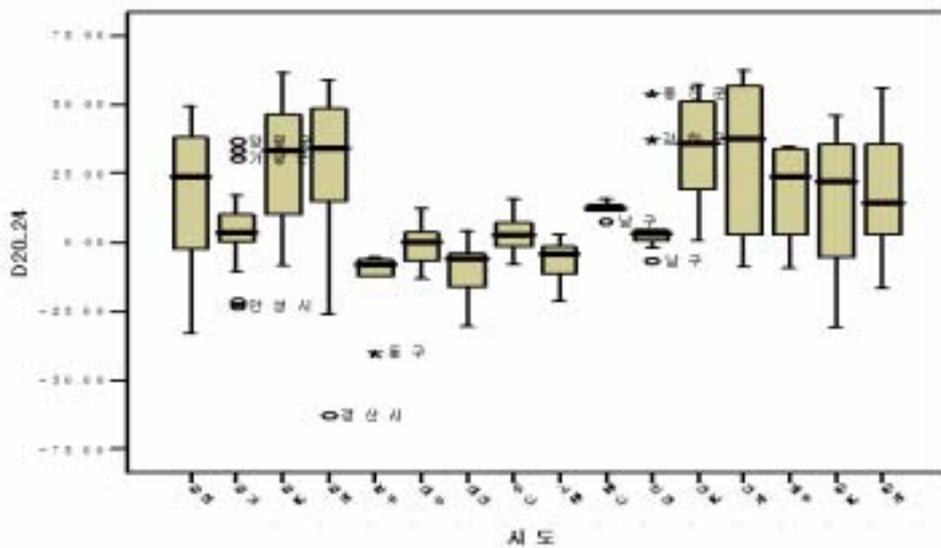
시도	시군구	10_14세	15_19세	20_24세	25_29세	30_34세	35_39세	70_74세
상위 10위								
전라북도	장수군	7.4	27.4	62.4	56.6	33.7	28.5	(4.3)
전라북도	임실군	8.1	53.6	62.2	61.4	43.6	35.8	(9.6)
경상남도	합천군	17.8	44.7	61.5	54.3	31.3	26.2	(7.0)
경상북도	의성군	6.7	38.5	58.9	55.3	33.9	24.7	(9.5)
경상북도	영양군	5.3	15.3	57.5	51.7	29.4	19.1	(7.3)
전라남도	진도군	5.9	39.1	57.2	43.5	18.6	13.4	(12.3)
전라남도	장흥군	4.2	34.7	56.9	44.1	20.2	18.0	(8.9)
전라북도	순창군	14.9	35.3	56.9	52.6	33.5	31.7	(8.2)
전라북도	고창군	6.8	21.8	56.8	47.0	25.1	21.1	(10.4)
충청북도	괴산군	6.6	48.1	55.6	53.8	33.0	30.5	(6.3)
하위 10위								
경상북도	경산시	(1.6)	(55.5)	(62.4)	(6.9)	4.3	5.9	(0.3)
광주광역시	동구	1.4	(30.4)	(40.0)	(7.3)	6.4	9.3	(9.3)
강원도	춘천시	3.1	(21.3)	(32.6)	(0.2)	4.8	5.5	(10.9)
충청남도	연기군	6.3	(22.1)	(30.7)	9.5	14.4	15.3	1.6
대전광역시	유성구	4.7	(19.3)	(30.1)	(4.5)	4.7	5.5	4.0
충청남도	아산시	5.7	(38.1)	(27.1)	(4.3)	5.5	8.8	1.3
경상북도	칠곡군	3.4	(18.4)	(26.0)	(12.4)	2.7	7.0	(1.7)
경기도	안성시	5.3	(17.1)	(22.9)	(0.4)	2.3	6.0	(6.8)
서울특별시	동작구	4.5	(6.5)	(21.4)	(8.9)	5.7	9.4	0.1
경기도	오산시	3.7	(17.3)	(21.3)	(3.7)	1.1	2.5	(5.0)

끝으로 <그림 11>부터 <그림 14>는 상자도표(box plot)를 사용하여 20-24세, 35-39세 연령대에서 광역시도별로 시군구와 읍면동의 차이율 분포를 제시하였다. <그림 11>과 <그림 12>의 분포를 보면 대체로 광역시의 경우 시군구별 차이가 적은 반면 도의 경우 시군구별 차이가 컸다. 하지만 경기도의 경우 도의 유형이 아니라 광역시와 유사한 형태의 분포를 보였다. 이와 같은 사실은 경기도가 다른 도에 비해서 도시화와 산업화의 정도가 높은 현실과 연관되는 것으로 보인다.

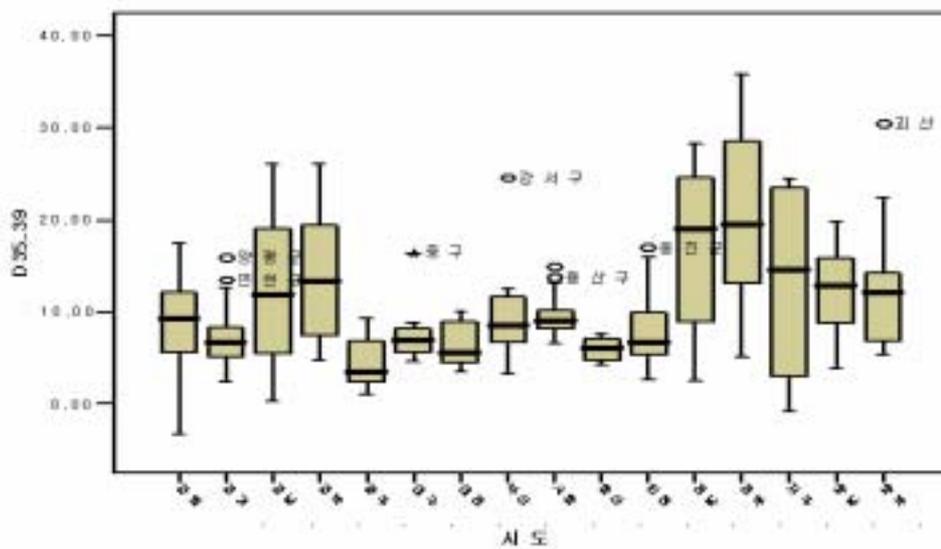
<그림 13>과 <그림 14>를 보면 읍면동별 차이율의 이상치(outlier)가 20-24세층에서 두드러짐을 알 수 있다. 이들 연령대에서는 중앙값에서 멀리 떨어져 있는 이상치들이 거의 모든 광역시도에서 나타났으며 이상치들이 중앙에서 떨어져 있는 정도도 매우 컸다. 특히 이상치들이 정의 차이율보다 부의 차이율

을 가지는 경우가 더 많았다는 점이 두드러졌다. 20-24세의 각 광역시도별 읍면동 차이율 분포를 보면 부의 방향으로 상당히 긴 꼬리를 가지고 중간값은 대체로 정의 방향에 위치해 있다. 긴 꼬리에 해당하는 지역들은 앞서 본 상위 10개 읍면동들처럼 젊은 연령층을 대거 흡수하는 기관의 소재지로 추정된다. 소수의 지역들이 젊은 층을 흡수하면서 높은 수준으로 부의 차이율을 보이는 반면 다수의 지역들은 젊은 층의 조사인구가 주민등록에 못 미치는 정의 차이율을 보였다. 이와 같이 상반된 방향의 차이율이 상쇄되는 까닭에 이 연령대의 읍면동 차이율이 상당한 규모임에도 불구하고 전국수준 차이율이 그리 높지 않았음을 알 수 있다.

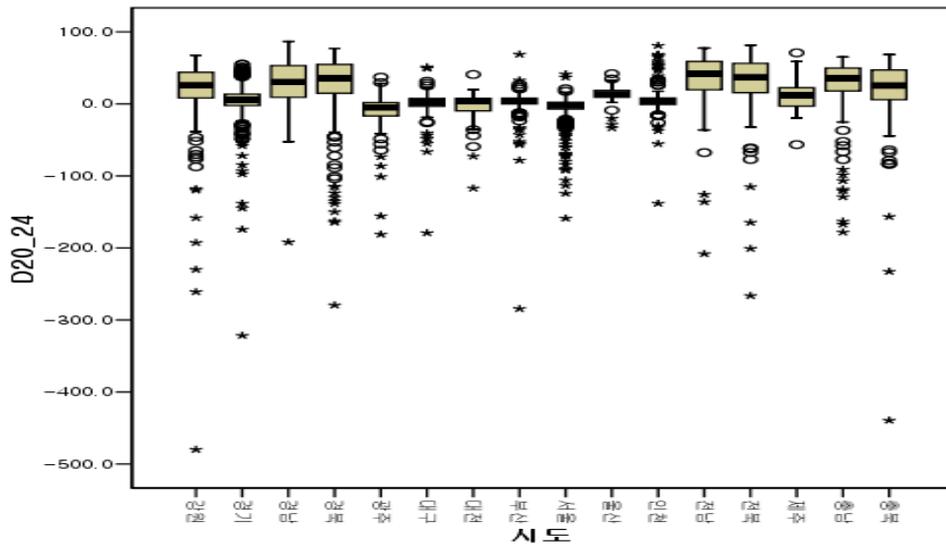
<그림 11> 시군구 비교 : 20-24세 연령대 주민등록인구와 총조사인구 차이율



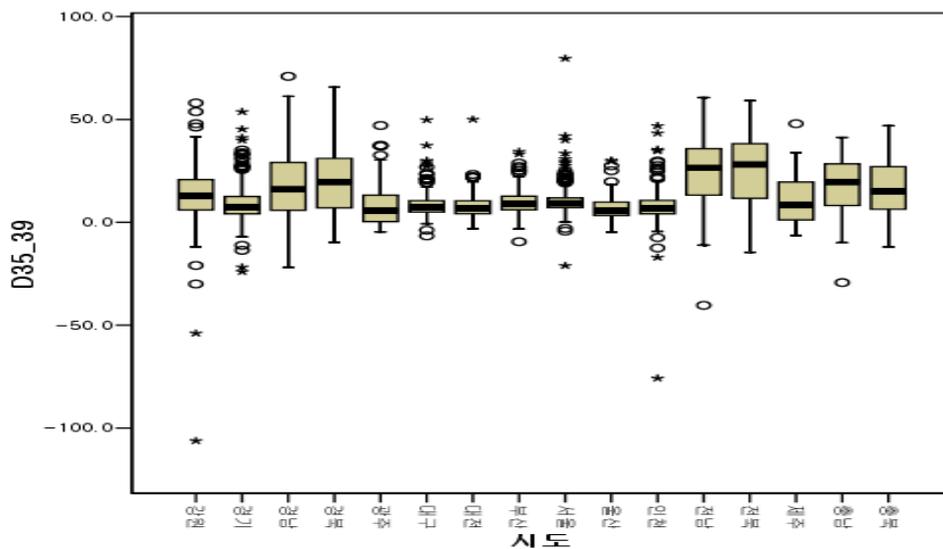
<그림 12> 시군구 비교 : 35-39세 연령대 주민등록인구와 총조사인구 차이율



<그림 13> 읍면동 비교 : 20-24세 주민등록인구와 총조사인구 차이율



<그림 14> 읍면동 비교 : 35-39세 주민등록인구와 총조사인구 차이율



III. 결론

이상의 분석은 연령대별로 차이율을 파악하는 것이 주민등록의 활용 가능성을 평가하는 데 매우 중요하다는 사실을 보여준다. 집계자료를 바탕으로 주민등록과 총조사인구의 차이를 파악하고자 할 경우 전국, 광역시도, 시군구, 읍면동의 수준에서 모두 연령대별 차이율을 세심하게 분석해야 한다. 연령대나 소지역을 고려하지 않고 특정지역에 대한 차이율을 계산할 경우 정의 차이율과 부의 차이율이 서로 상쇄됨으로써 차이율이 과소평가되는 위험이 발생한다.

또한 연령대 비교의 결과를 보면 주민등록인구와 총조사인구의 차이가 우연하게 발생하는 것이 아니었다. 차이율이 연령대에 따른 일정한 유형을 보인다는 것은 주민등록인구와 총조사인구의 차이가 생애주기와 밀접한 관련이 있음을 보여준다. 노년층에서 나타나는 주민등록 미비처럼 과거 특정한 행정현실이 과도기적으로 작용한 경우도 있다. 하지만 이러한 사례들의 비중이 현재도 양적으로는 그리 크지 않고 앞으로는 점차 감소할 것으로 예상된다. 따라서 전체적으로는 생애주기가 주민등록과 총조사인구의 차이에 대한 가장 중요한 설명 요인이 된다. 경제적, 사회적인 이유로 활동이 활발해질수록 해외거주, 조사누락 등으로 총조사에 포함되지 않게 됨으로써 주민등록과의 차이가 커지는 경우들이 증가한다. 그리고 노년층에서는 조사누락이나 사망신고 미비 등 주민등록신고의 부정확으로 인해서 차이율이 다시 높아진다.

가장 주목할 만한 연령대가 15-24세의 연령층이다. 이 연령층에서는 대학진학이나 경제활동을 시작하면서 대학과 공단 소재지로 집중 이동하는 상황이 발생한다. 이때의 이동이 주민등록이전을 수반하지 않는 경우가 많아서 주민등록과 총조사인구 사이에 큰 차이가 나타나는 지역들이 존재함을 여러 지역 수준에서 확인하였다.

이와 같이 주민등록인구와 총조사인구의 차이가 우연하게 나타나는 것이 아니라 구조적인 성격을 갖는다는 사실은 두 조사의 차이를 예측하고 설명할 수 있는 가능성을 의미한다. 또한 오차를 개선할 수 있는 가능성을 찾을 수 있다. 15-24세의 경우 대체로 젊은 층 유입 기관들에 의해서 오차가 발생한다는 점을 고려하면, 대학의 학적이나 공단의 인사기록 활용을 통하여 주민등록 집계 의 한계를 개선할 수 있을 것으로 예상된다.

끝으로 연령대별 차이를 보면 주민등록의 부정확성에 따른 오차 뿐만 아니라 인구주택총조사의 부정확성에 따른 오차의 가능성도 나타난다. 이와 같은 분석 결과는 주민등록을 통한 인구총조사의 대체가 아니라도 주민등록의 정보를 활용하여 인구주택총조사를 보완함으로써 오차를 줄이는 것이 현실적 대안이 될 수 있음을 시사한다.

참고문헌

- 김근영 · 김순관 · 이신해. 2003. "Comparing Census Data with Resistration Data with Respect to Population and Household in the Seoul Metropolitan Region." 『국토계획』 38(1): 237-249.
- 김설희 · 강계화. 2005. "GIS DB에 연계시킨 행정등록자료를 활용한 현장조사 품질제고 방안." 조사연구학회 춘계학술대회 발표문.
- 이건. 2007. "읍면동 수준에서 인구주택총조사와 주민등록의 집계인구 비교."

- 『통계』 33(1): 37-54.
- 이명진 · 서우석 · 이건. 2006. “주거부문 행정등록자료의 인구주택 총조사 활용 가능성에 대한 경험적 검토.” 『통계연구』 11(1): 57-87.
- 이지연. 2007. “2005년 인구주택총조사의 범위오차평가와 지역별 차이 연구.” 한국인구학회 춘계학술대회 발표문.
- 홍두승 · 이건 · 이명진 · 서우석 · 장원호. 2002. “인구주택 총조사 개선방안에 관한 연구-행정등록자료의 인구주택 총조사 활용방안.” 통계청 연구보고서.