Journal of the Korean Data & Information Science Society
2009, **20**(6), 1145–1153

한국데이터정보과학회지

주민등록 0세-6세 인구의 역 추정과 기존 인구통계와의 출생아수 비교†

김종태[1]

대구대학교 전산통계학과

접수 2009년 9월 12일, 수정 2009년 11월 20일, 게재확정 2009년 11월 24일

요 약

1998년부터 집계된 연령별 주민등록인구통계의 오류는 0세 혹은 출생아가 연령상승에 따라 인구수가 점점 늘어나서 6세 혹은 7세가 되는 해에 인구가 최대가 되고, 6세 혹은 7세 이후로부터 인구가 점차 감소하는 현상이 나타난다는 것이다. 본 연구는 장래의 인구추론에 가장 많은 영향을 미치는 0세-6세의 각 연령별 주민등록인구에서 미래에 1세-7세가 되는 인구를 예측하고, 최대치를 기준으로 0세-6세 인구를 '완전생명표'를 이용하여 0세-6세의 역순으로 역 추정 한다. 결론적으로 0세-6세까지의 제시된 역 추정 인구수는 기존의 인구통계데이터들 보다도 신뢰할 수 있다.

주요용어: 완전생명표, 이동평균, 장래인구추계, 주민등록인구.

1. 서론

최근 저출산율과 인구감소, 초고령화 사회로의 진입으로 연금문제, 노인복지문제, 등에 대한 것이 국가적으로 매우 심각한 문제로 떠오르고 있다. 이러한 문제들은 기존의 인구통계를 기초로 하여 문제점들을 제기하고 발표한다. 그러므로 문제 제기를 위한 정확한 추정과 예측을 위해서는 기초자료인 인구통계가 신뢰할 수 있는 통계가 되어야 한다.

가장 많이 사용되는 연도별 우리나라 인구통계 데이터는 다음 세 가지가 있다. 첫 번째는 5년마다 추계되는 통계청 (2006)의 '장래인구추계'이고, 두 번째는 통계청이 매년 12월 31일을 기점으로 발표하는 '연령별 주민등록인구읍면동'이고, 세 번째가 통계청 (2008a)의 '성 모의연령 출산순위별 출생'이다.

장래인구를 정확히 예측하기 위해서는 현재의 연령별 인구수에 대한 정확한 추계와 장래에 태어날 출생아수에 대한 정확한 추정과 예측이 가장 필수적인 요소이다. 출생아수가 정해지면, '완전생명표' 혹은 연령별 사망률의 추정에 의해서 이들의 향후 90년간의 각 연령별 인구수가 예측되기 때문이다. 출생아수의 정확한 추정은 인구통계에 있어 가장 중요한 요소이다. 실제로 0세-6세 인구는 '장래인구추계'에 결정적인 영향력을 가짐으로 0세-6세 연령별 인구수에 대한 신뢰할만한 추정이 필요하다.

기존의 인구통계데이터에서 가장 신뢰할 수 있는 출생아수는 무엇인가라는 문제점이 제시된다. '장래인구추계'의 0세 인구통계데이터인가? '주민등록인구'의 0세 인구통계인가? 아니면 '성 모의연령 출산순위별 출생' 인구통계인가? 이들 인구통계를 살펴보면 위의 모든 것이 답이 아님을 알 수 있다.

이들 인구통계들을 살펴보면 통계청 (2008a)의 '성 모의연령 출산순위별 출생'에 나오는 출생아수가 가장 근접한 답이 된다. 그러나 이 답 역시 다른 인구통계데이터와 비교하면 이상한 점들이 발생한다.

† 이 논문은 2008년도 대구대학교 학술연구비 지원에 의한 연구임.
[1] (712-714) 경상북도 경산시 진량면 내리동15, 대구대학교 전산통계학과, 교수. E-mail: jtkim@daegu.ac.kr

즉, '주민등록인구'의 인구가 최고가 되는 7세의 인구가 7세들이 태어난 해에 보고된 성 모의연령 출산순위별 출생에 나오는 출생아수 보다 많다는 것이다. 다음 절들에서 이 문제를 다룰 것이다.

'주민등록인구'는 우리나라 국민이면 모두가 출생과 동시에 생성하고, 사망과 동시에 필히 소멸시켜야 하는 법적인 의무를 지니고 있다. 그럼에도 불구하고 흥미로운 사실은 '주민등록인구'에서 0세-6세의 인구는 연령상승에 따라서 6세 혹은 7세가 되면 인구수가 최대 인구수가 된다는 것이다. 없었던 인구가 어디에서 생긴 것인가? 왜 6세-7세 인구수가 최대인가라는 의문을 가진다. 이것에 대한 답은 또 하나의 법적의무인 초등학교 의무교육에서 찾을 수 있다. 6세-7세에 자녀들을 초등학교에 입학시키지 않으면 법적인 문제가 발생하고, 초등학교 입학을 위해서는 주민등록이 필요하기 때문일 것이라는 분석이다.

6세-7세 인구수가 최대이라면 역으로 과연 0세-6세 사이의 연령별 인구수는 얼마인가? 하는 의문을 가진다. 연도별 연령별 인구통계는 '주민등록인구'와 '장래인구추계'가 전부이기 때문에 그 답을 통계청이 제시한 '장래인구추계'에서만 찾을 수밖에 없다.

'장래인구추계'는 5년마다 실시되는 인구센서스가 끝나는 다음 해에 '인구센서스'통계와 '주민등록인구'와 '출생률', '인구동태통계' 등을 이용하여 추계함으로 매우 정확한 인구통계처럼 보인다. 그러나 통계청 (2006)의 '장래인구추계'는 다음 인구센서스가 있는 5년 후까지의 인구는 예측통계이므로. 2006년에서 현재까지의 추계된 출생아수들과 비교할 때, 정확성이 매우 떨어질 뿐 아니라 심한 오차를 가지고 있다. 이들 오차가 조정된 '장래인구추계'를 얻기 위해서는 2010년의 인구센서스가 끝나고 '장래인구추계'를 하는 2011년까지 기다려야한다. 심지어 2006년 이전의 조정된 '장래인구추계'에서도 많은 통계적인 오류들이 보인다.

본 연구에서는 기존의 '주민등록인구'를 이용하여 미래의 1세-6세의 인구를 예측하고, 최대인구수를 가지는 6세-7세를 기점으로 다시 0세-6세까지 우리나라 인구를 다시 역 추정하여 새로운 0세-6세의 데이터를 제시하고, 추정된 데이터와 '장래인구추계'와 '성 모의연령 출산순위별 출생'에서의 출생아수에 대한 비교를 통해서 적절한 출생아수를 추정하는데 목적이 있다.

2절에서는 '주민등록인구'의 문제점과 연령상승에 따른 인구이동의 표기방법을 정의하고, 3절에서는 이동평균을 이용하여 미래의 인구를 예측하는 방법과 0세-6세의 인구를 역 추정하는 알고리즘을 제시하였다. 4절에는 제시된 '역 추정된 주민등록인구'와 기존의 인구통계들인 '성 모의연령 출산순위별 출생'에서의 출생아수의 분포와 '장래인구추계'의 0세인구의 분포들과 비교분석 하였다.

2. 주민등록인구와 0세-6세의 연령상승에 따른 인구이동

'주민등록인구'는 크게 다음과 같은 문제점을 가지고 있다.

첫 번째는 연령별 '주민등록인구'는 1998년-2008년까지만 제공되고 있고 1998년 이전의 '연령별 주민등록인구'는 없다. 단지 5세별 인구데이터 밖에는 알 길이 없다. 그러므로 10년 남짓한 '주민등록인구'를 가지고 연령별 분석을 할 수밖에 없다. 인구통계는 중요한 통계임에도 불구하고 데이터의 부재는 이해할 수 없는 부분이다.

두 번째는 2003년의 5세를 제외하고 나머지 연령은 1998년에서 2008년까지의 0세 인구 혹은 출생아는 6세 혹은 7세가 될 때까지 계속해서 인구가 증가하여 최대치를 이루고 난 뒤 7세 이후부터는 점진적으로 인구수가 줄어들기 시작한다는 것이다. 이론적으로는 출생아수가 최대로 가장 많아야하고, 연령상승에 따라 사망 혹은 이민 등으로 인해서 인구수가 줄어야 한다.

세 번째는 국외 장기체류자 혹은 외국 영주권자들도 우리나라 '주민등록인구'에 포함이 되어 있다는 것이다. 그러므로 '주민등록인구'는 실제 국내에 거주하는 인구 보다 훨씬 더 많은 인구수를 나타내고 있다. 정확한 인구통계를 위해서는 한국출입국사무소나 법무부의 국외 장기체류자나 외국 영주권자들

에 대한 연령별 인구에 대한 통계가 필요하다.

네 번째는 2008년부터 '주민등록인구'가 통계청에서 행정자치부로 옮겨가면서 2008년 '주민등록인구'에서 24세에서 43세까지, 즉 20개의 연령층의 인구들이 일제히 2007년의 연령상승되기 전의 인구보다 증가된 것이다. 데이터 집계오류가 아니고, 탈북자 혹은 외국인이 우리나라 사람과의 결혼으로 인하여 증가된 것이라면, 여기에 대한 통계자료의 집계 역시 중요할 것이다.

y 를 출생연도, d ($d = 0, 1, \cdots$,)는 각 연령이라고 할 때, $X_{d,y+d}$를 y를 출생연도로 하는 연령 d의 인구수로 정의하자. d연령이 m년만큼 연령상승이 생긴다면, m년 후의 인구수를 $X^M_{d+m,y+d+m}$으로 정의하자.

예를 들면, $y = 1998$년도의 $d = 0$ 세 인구, $X_{d,y+d} = X_{0,1998}$은 $m = 6$년 후인 2004년에는 6세의 인구, $X^M_{d+m,y+d+m} = X^M_{6,2004}$가 된다. 연령상승에 따른 연도의 이동을 표로 만들어 보면 다음과 같다.

표 2.1 출생연도를 기준한, 학년 연령 증가에 따른 상승 이동표

연령 d	출생연도 y			
	1998년	1999년	$\cdots$	2008년
0세	$X_{0,1998}$	$X_{0,1999}$	$\cdots$	$X_{0,2008}$
1세	$X^M_{1,1999}$	$X^M_{1,2000}$	$\cdots$	$X^M_{1,2009}$
2세	$X^M_{2,2000}$	$X^M_{2,2001}$	$\cdots$	$X^M_{2,2010}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
m세	$X^M_{d+m,1998+d+m}$	$X^M_{d+m,1999+d+m}$	$\cdots$	$X^M_{d+m,2008+d+m}$

출생아수는 모든 연령상승에 따른 인구수 $X^M_{d+m,y+d+m}$보다 큰 값을 가져야 한다. 연령상승이 되면 연령상승 이전의 인구수보다 사망이나 이민 등의 이유로 인구수는 줄어야 하고, 연령상승되기 전과 된 후의 인구수의 차이는 천재지변이나 전쟁 등으로 많은 인구가 줄어들지 않는 한 그 차이는 이상치를 넘어서는 안 된다. 이러한 이유를 기반으로 다음의 가정을 정의하자.

가정 1. $X_{0,y} > X^M_{d+m,y+d+m}$, 모든 $d = 0, 1, 2, \cdots$, $m = 1, 2, \cdots$에 대하여.
가정 2. $X^M_{d+m,y+d+m} < X^M_{d+m-1,y+d+m-1}$, 모든 $m = 1, 2, \cdots$에 대하여.
가정 3. $\left|X^M_{d+m,y+d+m} - X^M_{d+m-1,y+d+m-1}\right| < C$, 여기서 C는 이상치의 한계치이다.

위의 가정들을 가지고 표 2.2의 '주민등록인구' 데이터를 살펴보자.

표 2.2 0세에서 7세까지의 주민등록인구수 (원본)

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
0세	607,766	588,010	608,060	532,700	470,251	469,270	451,219	413,839	424,737	470,258	444,421
1세	671,158	**638,899**	**618,444**	**637,932**	**557,596**	**492,091**	**490,783**	**471,089**	**433,122**	**446,103**	**491,498**
2세	686,229	**672,557**	**640,444**	**619,315**	**638,559**	**558,093**	**493,125**	**491,759**	**471,970**	**433,946**	**446,884**
3세	709,382	**687,069**	**673,589**	**640,716**	**619,744**	**638,482**	**558,553**	**493,809**	**492,427**	**472,617**	**434,320**
4세	716,143	**710,221**	**687,471**	**673,714**	**640,894**	**620,038**	**639,050**	**559,075**	**494,287**	**492,842**	**472,935**
5세	713,230	**716,847**	**711,135**	**688,057**	**673,758**	640,643	**620,183**	**639,477**	**559,423**	**494,527**	**492,914**
6세	729,241	**714,354**	**718,080**	**711,496**	**688,504**	**673,822**	**641,385**	**620,534**	**639,775**	**559,670**	**494,660**
7세	707,547	**729,999**	**715,126**	718,029	**711,597**	688,306	**673,988**	**641,472**	**620,666**	**639,800**	559,549
8세	644,055	707,154	729,723	714,271	717,392	710,791	687,792	673,531	641,047	620,368	639,584

※ 짙은 색으로 표시된 부분은 가정 1과 2에 위반되는 주민등록인구이다.

표 2.2에서 색칠된 부분들은 가정 1과 2를 모두 위반하고 있다. 2003년의 5세를 제외하고 나머지 연령은 1998년-2008년까지의 0세 인구는 6세 혹은 7세가 될 때까지 계속해서 인구가 증가하여 최대치를

이루고 난 뒤 7세 혹은 8세 이후부터는 점진적으로 인구수가 줄어들기 시작한다는 것이다. y연도 출생한 출생아들이 법적 의무교육인 초등1학년을 입학 할 때까지 여러 가지 사정으로 출생신고를 늦춘 것이 이유로 보인다.

표 2.2가 의미하는 바는 매우 크다. 적어도 우리나라의 출생아수는 6세 혹은 7세에서 최대치를 이루는 인구수보다는 많아야 한다는 것이다. 그렇다면 우리나라 출생아수는 실제로 얼마인가? 그리고 0세-6세의 각 연령별 실제 인구수는 얼마인가라는 문제에 부닥치게 된다.

다음 절에서는 미래의 1세-7세 인구를 예측하고, 최대치를 가지는 6세 혹은 7세의 인구를 기점으로 0세-6세까지의 주민등록인구를 '완전생명표'를 가지고 0세-6세의 역순으로 역 추정하는 알고리즘을 제시한다. 이들 추정을 통해서 출생아수를 포함한 보다 정확한 연령별 추정치를 얻을 수 있다.

3. 제시된 인구 예측방법과 역추정 알고리즘

'주민등록인구'의 0세-6세의 인구수 추정을 위하여 다음과 같은 절차를 수행한다.

단계 1에서 단계 5까지는 주민등록인구에서 미래의 연령별 인구에 대한 추정이 이루어진다. 즉, 2008세의 0세 인구가 7년 후인 2015년에 7세가 되기까지의 인구수, 2008년 1세의 인구가 6년 후인 2014년에 7세가 되기까지의 인구수, 2008년 2세의 인구가 5년 후인 2013년에 7세가 되기까지의 인구수, 등 계속해서, 2008년 6세의 인구가 1년 후인 2009년에 7세가 되기까지의 인구수에 대한 추정이 이루어진다. 이는 기존의 주민등록인구를 기반으로 2009년에서 2015년까지의 미래의 0세-6세까지의 인구수를 예측하는 알고리즘이다.

단계 1. 1998년부터 2008년까지의 '주민등록인구'를 연령상승에 따른 데이터로 표 2.1과 같이 전환시킨다.

단계 2. 연령상승에 따른 데이터에서 인구수의 최대치를 찾는다. 이 때 최대치는 6세 혹은 7세가 되는 해에 최대로 나타난다.

단계 3. 연령상승표에서 1992년에 출생하여 1998년 6세와 1999년에 7세의 연령상승 인구증감률 기점으로 2007년에 0세와 2008년의 1세 인구가 되기까지의 연령상승 인구증감률들을 포함하는 모든 인구증감률을 다음과 같이 계산한다.

$$P_{(d+1,\ y+d+1)} = \frac{X^M_{d+1,y+d+1}}{X^M_{d\,y+d}}. \quad y = 1992, 1993, \cdots, 2008, \quad d = 0, 1, \cdots, 6. \tag{3.1}$$

단계 4. 미래의 연령상승 인구증감률 $P_{(d+1,[y+1]+d+1)}$의 추정치 $\widehat{P}_{(d+1,[y+1]+d+1)}$을 구하기 위한 방법으로 다음의 모형들을 고려한다. 이미 알려진 직전 연도의 증감률을 포함하는 최근 m개의 기지의 증감률들에 대한 이동평균을 다음 연도의 증감률의 추정치로 사용하는 것이다. 즉, m은 $d+1$연령의 최근 증감률을 포함해서 사용되는 증감률의 m개의 수를 의미한다. 여기서 $m = 4$을 사용하였다. m이동평균 방법은 쉽고, 단순하면서도 우수한 예측력을 지니고 있다 (김종태, 2009).

$$\widehat{P}_{(d+1,\ [y+1]+d+1)} = \frac{1}{m}\sum_{k=0}^{m-1} P_{(d+1,[y-k]+d+1)}, \quad d = 0, 1, \cdots, 6 \tag{3.2}$$

단계 5. $X_{d+1,\ [y+1]+d+1}$을 다음 해인 $[y+1]$의 미래의 $d+1$연령 인구수라고 두자. $d+1$연령의 미래 인구증감률 $P_{(d+1,[y+1]+d+1)}$에 대한 추정치 $\widehat{P}_{(m,d+1,\ [y+1]+d+1)}$를 추정함으로서

$X_{d+1,\ [y+1]+d+1}$는 다음과 같이 추정할 수 있다.

$$\widehat{X}^{M}_{d+1,\ [y+1]+d+1} = \widehat{P}_{d+1,\ [y+1]+d+1)} \cdot X^{M}_{d,\ y+d+1}, \quad d = 0, 1, \cdots, 6. \tag{3.3}$$

다음의 단계 6부터는 주민등록인구에 나타나는 6세 혹은 7세의 최대인구수를 가지고 통계청 (2008b)과 구자흥 (1996, 2002)에서 제공한 '완전생명표'를 이용하여 0세-6세 연령별 인구수를 역 추정하여, 원래의 0세-6세의 인구수를 재구성 한다.

단계 6. 통계청 (2008b)에서 제공한 표 3.1의 '완전생명표'를 이용한다. '완전생명표'의 각 셀의 사망확률을 $P^{D}_{(d,y+d)}$로 두자. 1995년 사망확률은 구자흥 (2002) 74쪽의 자료이고, 1996년, 1998년, 2000년의 사망확률은 '완전생명표'에서 보간법을 사용하여 구하였다 (참고; 김기환 등, 2006).

표 3.1 완전생명표

각세별	1995 사망확률 (전체)	1997 사망확률 (전체)	1999 사망확률 (전체)	2001 사망확률 (전체)	2002 사망확률 (전체)	2003 사망확률 (전체)	2004 사망확률 (전체)	2005 사망확률 (전체)	2006 사망확률 (전체)	2007 사망확률 (전체)
0세	0.00852	0.00716	0.00602	0.00572	0.00551	0.00531	0.00512	0.00493	0.00454	0.00413
1세	0.00086	0.00076	0.00061	0.00058	0.00052	0.00050	0.00047	0.00045	0.00040	0.00039
2세	0.00071	0.00060	0.00050	0.00047	0.00042	0.00040	0.00037	0.00034	0.00031	0.00029
3세	0.00059	0.00050	0.00042	0.00039	0.00035	0.00033	0.00030	0.00028	0.00025	0.00023
4세	0.00050	0.00043	0.00037	0.00033	0.00030	0.00029	0.00026	0.00023	0.00020	0.00019
5세	0.00047	0.00039	0.00034	0.00029	0.00027	0.00026	0.00023	0.00021	0.00018	0.00018
6세	0.00044	0.00036	0.0003	0.00027	0.00024	0.00023	0.00021	0.00020	0.00017	0.00017
7세	0.00042	0.00033	0.00027	0.00024	0.00022	0.00021	0.00019	0.00018	0.00015	0.00015

※ 1995년 사망확률은 구자흥 (2002) 74쪽의 자료임.

단계 7. 단계 5에서 구한 $\widehat{X}^{M}_{m,\ d+1,y+d+2}$의 값들 중 6세 혹은 7세에서의 최대치를 기준으로 통계청 (2008b)에서 제공한 표 3.1의 '완전생명표' 사망확률 $P^{D}_{(d,y+d)}$을 이용하여 0세에서 6세의 인구수 $\widehat{X}^{INV}_{m,\ d,y+d}$을 역 추정한다.

$$\widehat{X}^{INV}_{d,y+d} = \widehat{X}^{M}_{d+1,y+d+1}/(1 - P^{D}_{(d,y+d)}), \quad d = 0, 1, \cdots, 6. \tag{3.4}$$

단계 1에서 단계 7까지의 알고리즘을 거쳐서 표 3.2의 주민등록 0세-6세 인구의 역 추정된 인구수를 구할 수 있다.

이러한 역 추정 방법은 '주민등록인구'가 가지는 오류인 0세의 인구수가 6세나 7세가 될 때까지 계속해서 인구수가 늘어나는 오류들을 수정한 것으로 '주민등록인구'에 기록된 최대인구수를 가지고, 0세-6세까지의 인구를 추정하는 것이다.

표 3.2에서는 0세-7세까지의 전체인구를 역 추정한 것이고, 표 3.3은 '주민등록인구' 0세-7세까지의 여자 인구를 같은 방법으로 역 추정한 것이다.

2절의 표 2.2의 원래의 '주민등록인구'와 표 3,2의 '역 추정된 주민등록인구'와는 상당한 차이가 있다. 예를 들어, 2008년 0세의 경우에 '주민등록인구'인 경우에 444,421명이지만 역 추정된 2008년 주민등록 0세 인구는 469,873명이 된다.

다음 4절에서는 표 3.2에서 '역 추정된 주민등록인구'가 타당성이 있는지를 통계청 (2008a)의 '성 모의연령 출산순위별 출생'에 나오는 출생아수의 분포와 비교 분석한다.

표 3.2 '역 추정된 주민등록인구' 전체 인구

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
0세	647,503	625,990	645,124	564,032	498,442	496,781	476,674	437,954	450,868	496,542	469,873
1세	676,039	643,264	622,244	641,359	560,824	495,710	494,157	474,246	435,805	448,831	494,500
2세	686,229	675,526	642,823	621,865	640,978	560,499	495,453	493,910	474,023	435,609	448,651
3세	709,382	687,069	675,121	642,470	621,554	640,667	560,235	495,245	493,712	473,848	435,461
4세	716,143	710,221	687,471	674,784	642,175	621,293	640,408	560,017	495,072	493,550	473,705
5세	713,230	716,847	711,135	688,057	674,494	641,918	621,063	640,184	559,832	494,923	493,407
6세	729,241	714,354	718,080	711,496	688,504	674,231	641,684	620,852	639,982	559,670	494,789
7세	707,547	729,999	715,126	718,029	711,597	688,306	673,988	641,472	620,666	639,800	559,549

표 3.3 '역 추정된 주민등록인구' 0세-6세 여자 인구

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
0세	308,293	298,904	307,398	269,982	237,779	238,609	229,368	211,110	217,637	241,183	228,062
1세	324,861	306,333	297,148	305,704	268,594	236,593	237,455	228,293	210,150	216,737	240,263
2세	324400	324,627	306,135	296,978	305,533	268,446	236,473	237,334	228,185	210,057	216,653
3세	332917	324890	324,446	305,981	296,845	305,400	268,333	236,383	237,246	228,105	209,988
4세	333258	333267	325148	324,303	305,857	296,735	305,292	268,242	236,312	237,177	228,044
5세	331681	333656	333801	325421	324,183	305,751	296,640	305,200	268,167	236,251	237,118
6세	342231	332240	334322	334101	325687	324,076	305,657	296,554	305,118	268,100	236,196
7세	334249	342688	332691	334267	334102	325619	323,979	305,574	296,483	305,049	268,050

4. 역 추정된 데이터와 기존의 인구통계와의 비교 분석

1절의 서론에서 제시한 기존의 인구통계데이터에서 우리나라의 믿을 수 있는 출생아수는 어느 것인가라는 문제로 돌아가 보자. 결론적으로 '역 추정된 주민등록인구' 0세의 인구수는 현존하는 인구통계 데이터들 중에서 우리나라 출생아수에 가장 가까운 데이터이라고 분석된다.

그 이유를 알아보기 위하여, 출생아수를 통계청 (2008a)에 발표한 '성 모의연령 출산순위별 출생'과 '장래인구추계'와 '주민등록인구', 그리고 '역 추정된 주민등록인구' 0세와 비교해 보자. 표 4.1은 0세 인구통계 데이터들을 비교해 놓은 것이다.

표 4.1의 출생아수는 통계청 (2008a)의 '성 모의연령 출산순위별 출생'에서의 출생아수로서 1997년의 출생아수는 668,020명이다. 반면에 '주민등록인구'에서 연령상승에 따라 최대인구수는 7세 인구 673,988명이다. 이는 1997년에 673,988명 보다 더 많은 수가 출생했을 것이라는 증거이다. 그러므로 '주민등록인구'의 최대인구수에서 '완전생명표'의 0세-6세까지 연령별 사망률을 이용하여 0세-6세까지 인구를 역순으로 역 추정하면 '역 추정된 주민등록인구'의 0세 인구수는 680,880명으로 추정된다. 7세가 최대가 될 것이라는 가정 하에 '주민등록인구'를 역 추정해 보면, 표 4.1에서 '역 추정된 주민등록' 0세 인구와 통계청의 출생아수가 거의 일치해 감을 알 수 있다.

2008년의 출생아수는 통계청 (2009)의 2008년도 1월부터 12월까지의 월별 출생아수들을 합친 값은 466,000명으로 추산되고, 이는 '역 추정된 주민등록인구' 0세 인구의 469,873명과 거의 비슷하다. 유아사망률이나 해외입양아수를 고려하면 역 추정된 출생아수가 더 타당성이 있다.

통계청 (2006)의 '장래추계인구'는 2006년 이전의 통계는 '인구센서스' 및 '주민등록인구' 등 많은 인구통계자료들을 이용하였고, 코호트 계산방법을 사용하여 보정작업을 거쳐서 만든 통계이다. 그러나 1998년부터 2005년까지의 '장래인구추계' 0세 인구는 과대추정이 된 반면에 2006년에서 2008년의 0세 인구는 매우 과소추정 되었다. 실제로 2006년 이후의 '장래인구추계' 데이터는 예측된 데이터로서 출생아수에 대한 예측력이 과소추정 되었음을 알 수 있다.

그러므로 본 연구에서 제시한 '주민등록인구'의 역 추정방법은 '성 모의연령출산순위별 출생'이나 '장

래인구추계'에서 제공되는 출생아수의 결과보다도 믿을만한 추정치를 가진다고 볼 수 있다.

그림 4.1을 보면 '성 모의연령 출산순위별 출생'의 출생아수의 분포와 '역 추정된 주민등록인구'의 분포가 거의 같음을 알 수 있다. 이와는 달리 '장래인구추계'의 0세 인구는 매우 다른 분포모양을 보인다. 이 사실에서 '성 모의연령 출산순위별 출생'과 '주민등록인구'와 '역 추정된 주민등록인구' 이 세 가지가 모두 잘못되었던지 아니면 '장래인구추계'가 잘못되었다고 볼 수 있다.

실제로 '장래인구추계'를 가지고 인구 예측할 때에는, '주민등록인구'를 기초로 예측한 것보다 예측력이 훨씬 떨어짐이 모의실험에서 발견된다.

표 4.2와 그림 4.2는 우리나라 여성에 대하여 분석한 것이다. 분석 방법은 우리나라 남·여를 합친 전체 인구데이터의 분석방법과 동일하며 결론 또한 같다. 그림 4.2를 보면 우리나의 '장래인구추계'가 상당히 잘못되어 있을 알 수 있다. 이와 같은 결론은 비록 주민등록인구가 가지는 약점인 재외국민을 포함하여, 국외 장기체류자의 변수를 다 포함한다고 가정을 하더라도 통계청 (2008a)에 발표한 '성 모의연령 (각세) 출산순위별 출생'에서의 출생아수의 분포와 '장래인구추계'의 0세인구의 분포가 다르다는 것은 납득하기 어렵다.

그러므로 5년마다 실시하는 '장래인구추계'에는 많은 예산과 거의 모든 인구관련 데이터를 가지고 만들어짐에도 불구하고 상당히 많은 문제점을 가지고 있음을 알 수 있다. 비록 본 연구의 결과는 '주민등록인구'만을 가지고 출생아수를 추정하였지만. 통계청 (2008a)에 발표한 '성 모의연령 출산순위별 출생'에서의 출생아수와 비교해 볼 때 매우 타당하다고 판단한다.

표 4.1 0세 혹은 출생아수 전체 인구통계데이터 비교

	장래인구 추계 0세	주민등록 0세	주민등록 최대값	주민등록 최대값연령	출생아수	역추정주민 등록 0세
1997	675,996		673,988	7	668,020	680,880
1998	650,449	607,766	641,472	7	634,790	647,503
1999	623,394	588,010	620,666	7	614,233	625,990
2000	622,098	608,060	639,800	7	634,501	645,124
2001	596,105	532,700	559,670	6	554,895	564,032
2002	533,117	470,251	494,660	6	492,111	498,442
2003	494,291	469,270	492,914	5	490,543	496,781
2004	480,092	451,219	472,935	4	472,761	476,674
2005	453,778	413,839	434,320	3	435,031	437,954
2006	442,831	424,737	446,884	2	448,153	450,868
2007	449,027	470,258	491,498	1	493,189	496,542
2008	446,738	444,421	444,421	0	466,000	469,873

표 4.2 0세 여자 인구통계 데이터 비교

	장래인구 추계 0세	주민등록 0세	주민등록 최대값	주민등록 최대값연령	출생아수	추정주민 등록 0세
1998	311,011	289,106	305,574	7	301,987	308,293
1999	297,086	280,415	296,483	7	293,039	298,904
2000	296,242	289,685	305,049	7	301,854	307,398
2001	285,285	254,560	268,100	6	265,434	269,982
2002	254,712	224,081	236,129	6	234,290	237,779
2003	236,532	225,049	236,839	5	235,028	238,609
2004	230,532	216,761	227,632	4	227,056	229,368
2005	218,571	199,420	209,384	3	209,346	211,110
2006	213,282	204,794	215,710	2	215,980	217,637
2007	215,946	228,180	238,671	1	239,190	241,183
2008	215,019	215,444	215,444	0	225,773	228,062

그림 4.1 0세 혹은 출생아수 전체 인구통계 분포 비교

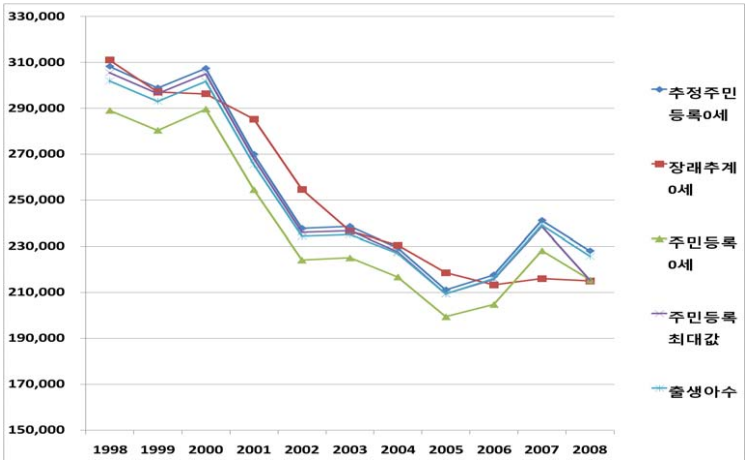

그림 4.2 0세 여자 인구통계 분포

참고문헌

구자흥 (1996). 한국인의 생명 표에 관한 해석적 연구 - 특히 간이생명표 작성을 중심으로 -. <응용통계연구>, **9**, 31-43.
구자흥 (2002). <인구통계학의 이론과 실제>, 교우사, 서울.
김기환, 이동희, 정승환 (2006). 간이생명표 확장 기법을 통한 사망확률 계산. <한국데이터정보과학회 춘계학술대회논문집>, 77-85.
김종태 (2009). 학년진급률에 따른 학생수 예측방법. <한국데이터정보과학회지>, **20**, 857-867.
통계청 (2006). <장래인구특별추계결과>, 통계정보시스템, 대전.
통계청 (2008a). <성 모의연령(각세) 출산순위별 출생>, 통계정보시스템, 대전.
통계청 (2008b). <완전생명표>, 통계정보시스템, 대전.
통계청 (2009). <월간 출생아수>, 통계정보시스템, 대전.

Journal of the Korean Data & Information Science Society
2009, **20**(6), 1145–1153

한국데이터정보과학회지

The re-projections under 7 age on Korean registered population and a comparison with the reported population statistical data†

Jongtae Kim[1]

Department of Computing & Statistics, Daegu University

Received 12 September 2009, revised 20 November 2009, accepted 24 November 2009

Abstract

The problem of Korean registered population is that population classified by age increases as one grows older until 6 age or 7 age. This paper is to suggest an algorithm of the re-projection under 7 age on Korean registered population and to analysis of comparison with the reported population statistical data. As the result, the re-projections population is trusted in the number of 0 age on the comparison of other reported population statistical data.

Keywords: Korean registered population, life table, moving average.

† This research was supported by the Daegu University Research Grant 2008.
[1] Professor, Department Computing & Statistics, Daegu University, Kyoungsan 712-714, Korea.
E-mail: jtkim@daegu.ac.kr