

The Errors of Population Projections for Korea on Korean Information Statistical System¹⁾

Yong-Hwa Yoon²⁾, Jongtae Kim³⁾

Abstract

Recently, Korean National Statistical Office submits the results of population projections for Korea from 1960 to 2050 year. The purpose of this paper is to suggest the reasonable assumptions for the survey of population, and then to detect the errors of the surveyed population (1960-2005) on Korean Information Statistical System.

Keywords : 연령별 전국추계인구, 연령이동인구모형, 이상치 추정, 통계정보시스템

1. 서론

통계청의 통계정보시스템(KOSIS; <http://kosis.nso.go.kr/>)은 1960년부터 2005년까지의 확정인구데이터를 기초로 하여, 2006년부터 2050년까지 '연령별 전국추계인구'를 2006년 11월에 산출하여 홈페이지에 제공하고 있다.

인구추계의 분석모형으로 시계열분석방법에 의한 추세연장법(trend extrapolation method), 평활법, ARIMA모형, 벡터자기회귀모형(VAR: Vector Auto-regression) 등이 있는데, 주로 사용하는 추세연장법은 사용대상에 따라서 인구추세, 비율추세, 차이추세, 밀도추세로 세분화 된다.(오재석 (2003)).

추계인구추정을 위한 현재까지의 기존 연구들은 주로 성장곡선을 사용하여 연도의 변화에 따라서 연령 이동변화에 의존하는 '횡적 추세연장법적인 관점에서 추정방법을 연구한 것이다. 이러한 추세연장법의 모형으로 자주 사용하는 것은 시간의 흐름에 따라 누적자료를 설명하기 위한 성장곡선모형이며, 이는 크게 선형모형과 비선형모형으로 나누어진다. 선형모형에는 Mansfield-Blackman 모형, Linear Gompertz 모형, Weibull 모형, Nonsymmetc Reponding Logistic (NSRL)모형, Harvey 모형 등이 있으며 비선형모형에는 지수기하모형, 수정지수곡선모형, Logistic 모형, Probit 모형,

1) 이 논문은 2006년도 대구대학교 학술연구비 지원에 의한 연구임

2) 경북 경산시 진량면 내리동 대구대학교 전산통계학과 교수

3) 교신저자: 경북 경산시 진량면 내리동 대구대학교 전산통계학과 교수
E-mail: jtkim@daegu.ac.kr

Gompertz 모형 등이 있다. (최종후의 3인 (2000), Young (1993)).

인구추계의 중요성은 사회, 정치, 행정, 경제, 산업, 보건, 복지, 교육 문화 등, 모든 측면에서 말 할 필요 없이 중요하다. 실제로 각 분야의 전문가들은 인구와 관련된 연구를 함에 있어서 통계청의 '연령별 전국추계인구' 데이터를 기초자료로 사용하여 왔다. 예로서 저출산 고령화의 예측, 국민연금에 비롯한 각종 연금의 고갈, 지역행정구역의 인구의 변동대책, 교육행정의 변화, 산업의 입지 조건, 문화 예술, 체육 시설 건립 등을 들 수 있다. (참조; 전광희 (2006)).

현 단계에서는 통계청이 집계하고 발표한 '인구추계'데이터 외에는 신뢰할만한 인구 통계 데이터를 찾기가 쉽지가 않다. 따라서 만약에 '연령별 전국추계인구' 데이터가 데이터 그 자체로서의 신뢰성을 잃어버리거나 오류들이 있다면 각 분야의 예측 결과들은 잘못된 결과를 추정하게 되고 그 결과의 신뢰성도 상당히 낮아지게 될 것이다. 그러므로 통계청이 발표한 장래의 추계인구 추정의 신뢰성을 가지기 위하여 통계청이 발표한 확정인구 통계집계 데이터의 분석과 검증이 필요하다. (Kim (2006)).

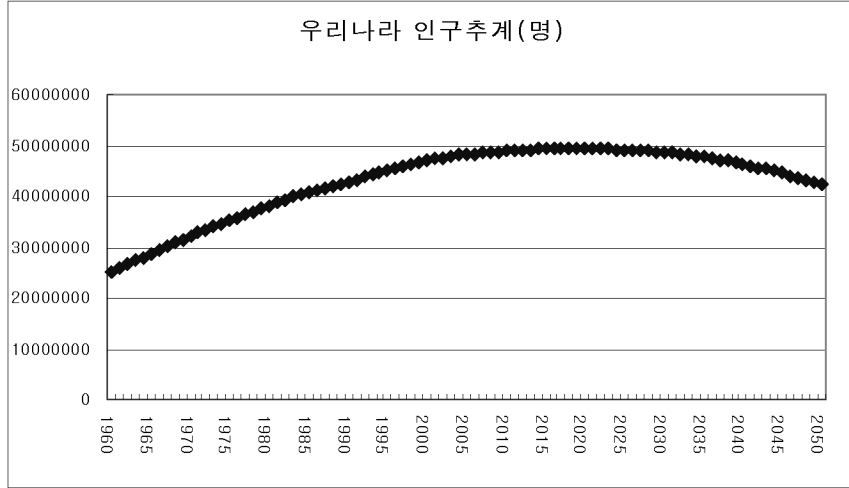
본 연구는 통계청이 발표한 확정인구 통계집계 데이터의 분석과 검증을 기초로 인구통계집계의 합리적인 인구집계모형의 가정들을 제시하고, 기존의 확정인구(1960년-2005년) 데이터를 분석하여, 통계집계에 대한 오류들과 이상점, 특성들을 조사하는데 목적을 둔다. 실제로 잘못된 기존 인구집계 데이터의 오류들은 2006년부터 2050년까지의 인구추계에서도 커다란 영향을 미치고 있음이 발견되었다.

2. 전국 연령별인구추계 데이터분석

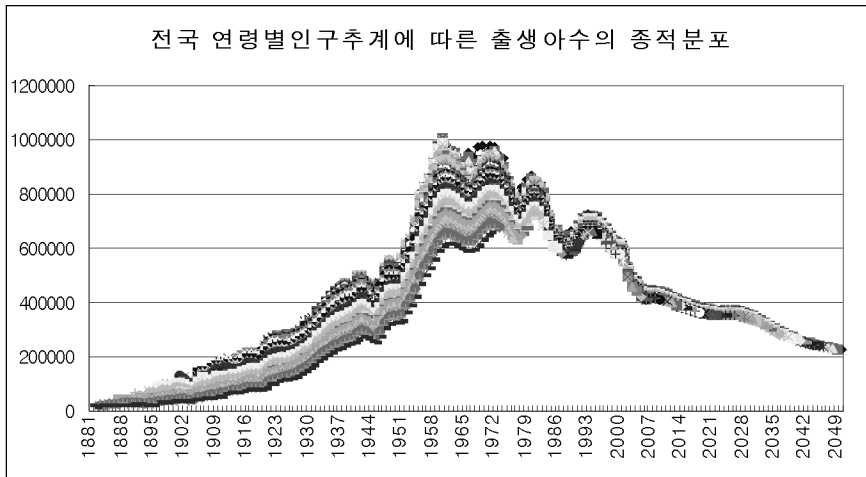
통계청에서 2006년 11월에 발표한 '전국 연령별 전국추계인구' 데이터는 다음의 두 가지로 나눌 수 있는데, 1960년부터 2005년까지의 확정인구와 2006년부터 2050년까지의 추계인구로 나눈다.

다음의 [그림2-1]은 우리나라 총인구수에 따른 인구추계분포를 나타낸 것이다. 통계청에 따르면 2005년의 확정인구는 48,138,077명으로 집계되고, 2018년 추계인구 49,340,350명으로 5천만 명을 넘기지 못하고 2018년을 기점으로 인구가 감소하는 것으로 전망하였다. 2050년에는 추계인구 42,342,769명으로 1991년의 확정인구 수준정도가 될 것으로 전망하고 있다.

다음의 [그림2-2]는 전국 연령별 인구추계에 따른 출생아수의 분포를 추정해 나타내었다. 1960년의 1세의 인구수는 1959년의 0세의 인구수를 대변할 것이고, 이 인구수의 변동률을 가지고, 다시 역으로 1959년 0세의 수를 추정할 수 있다. 이러한 작업을 반복해 가면, 1881년의 0세의 인구수까지 역 추정할 수 있다. 이러한 작업 후에 다시 나타나는 전국 연령별인구추계에 따르는 출생아수의 분포 모양은 [그림2-2]에서 1960년 이전의 원래의 0세(출생아수)를 다시 추정해서 그려 넣는다면 [그림2-2]의 왼쪽 상위부분으로 더 올라간 두터운 형태를 지니게 되고 1966을 기점으로 좌우대칭의 형태로 보여 진다. 우리는 1966년을 기점으로 출생아수의 증가 속도와 감소 속도를 비교해 볼 수 있는 결과를 가지게 될 것이다.



[그림2-1] 우리나라 총인구수에 따른 인구추계분포

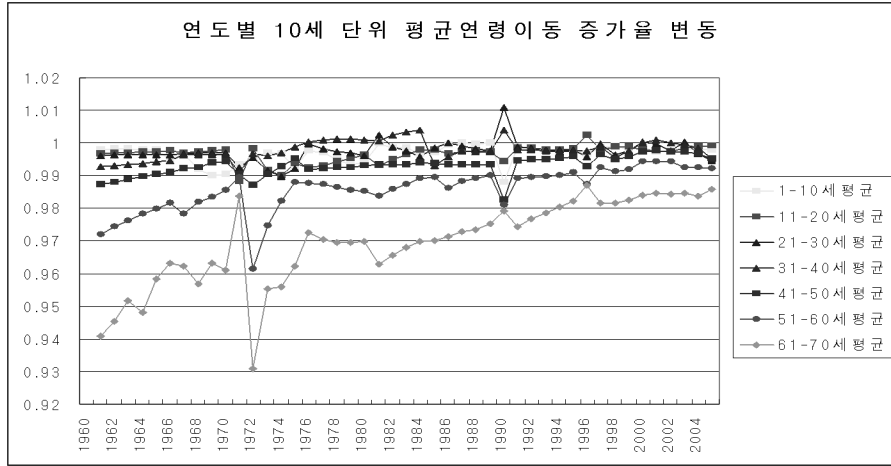


[그림2-2] 전국 연령별 인구추계에 따른 출생아수의 증적분포

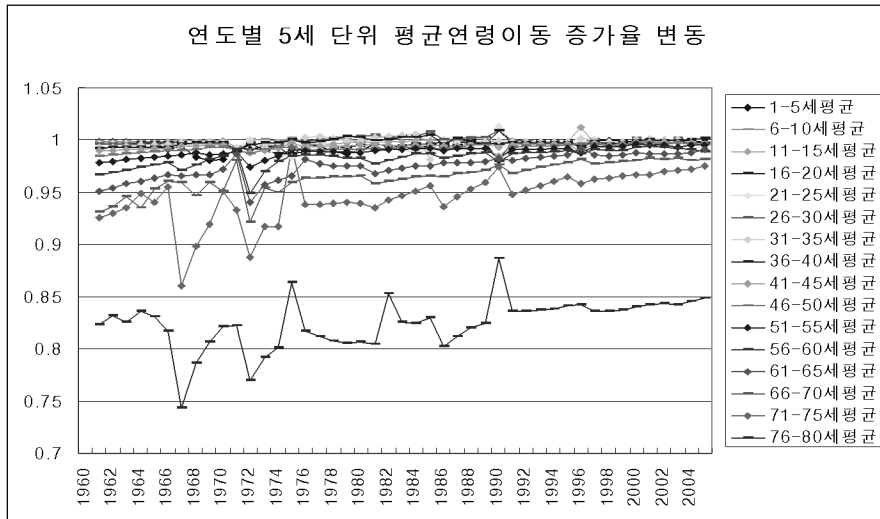
그러나 2006년부터의 출생아수의 분포가 통계청에서 제시한 추계에 대한 신뢰성을 가질지는 의문이다. 인구추계의 증적 영향력을 가지는 출생률은 인구성장곡선에 기초한 획적인 인구추정과는 전혀 다른 사회적, 심리적, 경제적, 환경의 요인에 영향을 받는 독립성을 가지기 때문이다. 이러한 예는 2006년의 합계 출산율이 1.6명 이하로 떨어지지 않을 것이라는 전문가들의 예상과 달리 2006년의 합계출산율은 1.08명으로 떨어졌고, 획적 인구추계의 결과는 심각한 오류를 낳게 되었다. 그리고 특단의 대책이 없는 한 출생아수의 감소는 통계청이 제시한 [그림2-2]보다 더 빠른 속도로 감소될 것으로 보인다.

[그림2-3]와 [그림2-4]는 연도별 평균연령이동 증가율 변동분포를 1960년-2005년까지

지 확정인구에 대하여 10세 단위와 5세 단위로 각각 나타낸 것이다. 전국의 추계인구를 추정하는데 있어서 [그림2-3]와 [그림2-4]를 살펴보면, 직관적으로 5년마다 전국총인구조사가 실시되는 0년, 5년이 되는 해와 그 해를 중심으로 전후 해에 평균연령이동 인구의 변화가 두드러짐이 보인다. 이러한 특성은 많은 인구추정 전문가들이 시계열 통계분석으로 장래 인구추계를 하고자 하는 기반을 마련해 준다.



[그림2-3] 연도별 10세 단위 평균연령이동 증가율 변동분포



[그림2-4] 연도별 5세 단위 평균연령이동 증가율 변동분포

그러나 전국인구총조사가 실시되는 해와 그렇지 않은 해와의 인구변동이 크게 나타나는 특성은 오히려 확정인구데이터의 신뢰성에 대한 의구심을 가진다. 예를 들어, 정치, 행정, 사회, 복지, 교육적인 행정구역에 따른 지역적인 인구통계의 과대 혹은 과소

집계의 필요성, 혹은 집계의 오류에 의해 통계집계의 문제점들이 발생할 수 있다.

3. 전국 연령별인구추계 대한 오류

1960년부터 2005년까지의 확정인구 통계집계표에 대한 검정을 하기 위하여 아래와 같은 일반적인 정의들을 수립한다. (단, 국내거주 외국인수는 인구통계집계에 포함하지 않는다.)

[가정1]. Y년도의 G세 인구는 (Y+1)년도의 (G+1)세 인구보다 많아야 한다. 즉,

$$\frac{(Y+1)\text{년도의 } (G+1)\text{세 인구수}}{(Y)\text{년도의 } (G)\text{세 인구수}} < 1.$$

사망과 해외이민 등으로 인하여 Y년도의 G세 인구는 (Y+1)년도의 (G+1)세가 되었을 때, 일반적으로 그 수가 감소한다. 한국국적을 가지는 귀화자수가 많지 않을 경우에 위의 정의는 타당하다.

[가정2]. (Y+k)년도의 (G+k)세 인구는 (Y+k+1)년도의 (G+k+1)세 인구와 비교할 때, 인구 증감의 변동이 특별한 이유가 없이 매우 클 수 없다.

위의 [가정2]를 토대로 이상점 판정을 위한 상수 C와 $k=0,1,\dots$ 에 대하여,

$$\left| \frac{(Y+k+1)\text{년도의 } (G+k+1)\text{세 인구}}{(Y+k)\text{년도의 } (G+k)\text{세 인구}} \right| > C.$$

이면, 인구통계집계에 이상치가 발생한 것으로 본다. 특별히 국가적 재앙이나 전쟁, 전염병, 대규모 해외이주와 같은 대규모의 인구 변동이 있지 않는 한은 (Y+k)년도의 (G+k)세 인구는 (Y+k+1)년도의 (G+k+1)세 인구와 비교할 때, 그 인구 증감의 변동이 커서는 안 된다.

위의 두 가지 가정은 추계인구를 추정하기 위한 연도의 변화에 따른 연령 이동을 기준으로 사망률과 같은 횡적변화에 대한 것이다. 그러나 인구추계에 있어서 또 하나의 중요한 요인은 출생아 수나 대규모 이민자 수의 변동에 따른 '종적'변화의 영향이다. 추계인구의 추정은 사망률과 같은 횡적변화의 추정정보는 종적변화인 출생율의 변화에 보다 많은 영향을 받는다. 이러한 종적인 변화의 추정은 사회, 심리, 경제 등의 복잡한 변인들로 인하여 추정이 보다 어렵다.

아래의 <표3-1>은 지면상의 제약으로 인해 연령이동에 따른 인구표의 일부분만을 예로서 표시하여 놓았다. [가정1]에서 Y년도의 G세 인구는 (Y+1)년도의 (G+1)세 인구보다 많아야 한다. 즉, 예를 들어 1969년에 0살이었던 인구수가 1975년에 910,414명이었는데 해가 거듭하여 나이를 더 할수록 인구수가 늘어나 10세가 되는 해는 8,538명이 더 늘어나 918,952명이 된다. 그러면 이 8,538명은 어디에서 생겨난 것인가? 이것은 모순이다. 나이가 들수록, 사고나 질병, 이민 등으로 인하여 그 수가 줄어야 함에도 불구하고 연령이동 인구가 늘어난다는 것은 통계집계방법이나 시기 등의 문제점이

있다는 것을 지적한다. <표3-1>에서의 110개의 셀 중에 64개 셀(58%)이 [가정1]을 위반하는 것으로 드러난다. 또한 이러한 연령이동인구의 증가는 1964년 24세에서 1965년 25세 차이에서 25,410명이 갑자기 증가한 것으로 보인다. 1952년생이 33세 되는 해인 1985년에는 1984년 32세의 인구보다 20,175명이 늘어나 약3.5%의 연령이동 인구증가율을 보인다. 더욱 심각한 이러한 모순은 장래 추계인구의 추정에서 성장곡선을 이용하여 추정하는 경우 인구추정의 과대추정을 하는 오류를 범하게 된다.

<표3-1>연령이동에 따른 인구표 (일부분)

0세연도	6세(연도)	7세(연도)	8세(연도)	9세(연도)	10세(연도)
1961	962046(1967)	961363(1968)	960681(1969)	960006(1970)	953550(1971)
1962	950462(1968)	950288(1969)	950105(1970)	943880(1971)	943710(1972)
1963	941167(1969)	940494(1970)	934711(1971)	935408(1972)	936067(1973)
1964	903058(1970)	897444(1971)	898049(1972)	898626(1973)	899277(1974)
1965	870437(1971)	870694(1972)	870925(1973)	871222(1974)	873703(1975)
1966	878657(1972)	878889(1973)	879191(1974)	881695(1975)	876680(1976)
1967	874353(1973)	874136(1974)	876113(1975)	872684(1976)	869268(1977)
1968	887679(1974)	886851(1975)	886151(1976)	885452(1977)	884753(1978)
1969	910414(1975)	912541(1976)	914673(1977)	916810(1978)	918952(1979)
1970	926074(1976)	929074(1977)	932088(1978)	935114(1979)	938152(1980)
1971	943130(1977)	943146(1978)	943177(1979)	943214(1980)	941180(1981)
1972	939932(1978)	940529(1979)	941147(1980)	937971(1981)	936671(1982)
1973	890813(1979)	886803(1980)	898960(1981)	901638(1982)	904392(1983)
1974	868865(1980)	870325(1981)	871672(1982)	873087(1983)	874470(1984)
1975	822200(1981)	825908(1982)	829695(1983)	833466(1984)	836284(1985)
1976	790874(1982)	794643(1983)	798398(1984)	801241(1985)	797464(1986)
1977	779201(1983)	780114(1984)	780122(1985)	777549(1986)	775235(1987)
1978	779239(1984)	777395(1985)	773455(1986)	769783(1987)	765822(1988)
1979	793029(1985)	796127(1986)	799496(1987)	802556(1988)	805427(1989)
1980	822207(1986)	828851(1987)	835213(1988)	841415(1989)	822685(1990)
1981	850276(1987)	852304(1988)	854125(1989)	842028(1990)	840977(1991)
1982	836839(1988)	838654(1989)	832284(1990)	831482(1991)	830560(1992)

위의 <표3-1>에서 나타난 [가정1]에 대한 기존 통계집계의 오류는 장래인구추계에 서도 똑 같은 오류를 범하고 있다. 실제로 다음의 <표3-2>에서와 같이 통계청이 제시한 2006년에서 2050년의 장래인구 추정에서 18세, 23세, 26세, 27세의 경우에 2006년부터 2050년까지의 모든 연도에 있어서 [가정1]을 위배하는 과대추정을 하고 있다.

<표3-2> 통계청 연령별인구추계의 연령이동에 대한 과대추정의 예

	18세	23세	26세	27세
2046추계인구	349575	353368	358469	360506
2046연령이동증감분	68	391	200	205
2046연령이동증가율	1.00019456	1.00110772	1.00055824	1.00056897
2047추계인구	344248	354356	353506	358675
2047연령이동증감분	69	394	199	206
2047연령이동증가율	1.00020048	1.00111311	1.00056325	1.00057467
2048추계인구	337496	354651	350601	353711
2048연령이동증감분	68	396	199	205
2048연령이동증가율	1.00020152	1.00111784	1.00056792	1.00057991
2049추계인구	329589	353842	351852	350806
2049연령이동증감분	68	397	202	205
2049연령이동증가율	1.00020636	1.00112323	1.00057443	1.00058471
2050추계인구	320782	351653	352841	352059
2050연령이동증감분	66	397	203	207
2050연령이동증가율	1.00020579	1.00113023	1.00057566	1.00058832

위의 [가정2]에 따라서 (Y+k)년도의 (G+k)세 인구는 (Y+k+1)년도의 (G+k+1)세 인구와 비교할 때, 연령이동 인구증가율 변동이 특별한 이유가 없이 매우 클 수 없다. 그러므로 탐사적자료분석에 의해 각 연령의 이동증가율에 대한 이상치(outlier)를 다음과 같이 구할 수 있다.

$$\text{연도별 G세 증가율} < (\text{G세 증가율의 제1사분위수} - 1.5 \times \text{사분위수범위})$$

이거나 혹은

$$\text{연도별 G세 증가율} > (\text{G세 증가율의 제3사분위수} + 1.5 \times \text{사분위수범위})$$

인 경우에는 해당연도의 G세 증가율은 이상치로 간주한다.

<표3-3> 통계청 연령별인구추계의 연령이동 인구증가율 변동의 이상치(outlier)

	1세	2세	3세	4세	5세	6세	7세	8세	9세	10세
1981	0.984	0.993	0.999	1.001	1.005	1.005	1.002	1.014	0.997	0.998
1982	0.991	0.988	0.993	0.999	1.001	1.005	1.005	1.002	1.003	0.999
1983	0.992	0.997	0.988	0.993	0.999	1.001	1.005	1.005	1.002	1.003
1984	0.992	0.995	0.997	0.978	0.993	0.999	1.001	1.005	1.005	1.002
1985	0.986	0.970	1.003	1.016	0.998	0.992	0.998	1.000	1.004	1.003
1986	0.996	0.995	1.000	1.002	1.002	1.008	1.004	0.995	0.997	0.995
1987	0.997	0.997	0.996	1.000	1.003	1.003	1.008	1.004	0.995	0.997
1988	0.997	0.999	0.996	0.995	1.000	1.002	1.002	1.008	1.004	0.995
1989	0.997	0.999	0.999	0.996	0.995	1.000	1.002	1.002	1.007	1.004
1990	0.964	0.971	0.983	0.971	1.013	1.017	1.003	0.992	0.986	0.978

위의 <표3-3>의 짙은 부분은 [가정2]을 위배한 이상치가 발생한 곳을 표시한 부분들이다. 또한 희미하게 짙은 부분은 [가정1]을 위배한 부분들을 표시하였다. 통계청의 확정인구 데이터를 보면 많은 통계집계 셀들의 오류를 볼 수 있고, 이는 통계청 데이터의 통계집계에 대한 신뢰성을 떨어뜨리게 된다.

4. 결론

본 연구는 인구통계집계의 합리적인 인구집계모형의 가정들을 제시하고, 기존의 확정인구(1960년-2005년) 데이터를 분석하여, 통계집계에 대한 이상점들과 특성들을 조사하는데 있다. 인구에 관한 많은 문헌들이 있었지만 이미 조사된 확정인구 통계집계에 관한 신뢰성을 검증하는 논문이나 보고서들은 찾기가 어려웠다. 이는 현실적으로 통계청이 발표한 확정인구의 통계데이터의 문제점을 지적하기란 쉽지 않기 때문일 것이다. 인구집계 통계데이터의 신뢰성 있는 공식통계데이터의 집계를 위해서는 여러 분야의 전문가들의 협동연구가 필요하다. 상식적이고, 합리적인 측면에서의 접근해 보면 기존 인구집계통계 데이터의 신뢰성을 높이는데 큰 어려움은 없다. 잘못된 기존 인구집계 데이터의 오류들은 2006년부터 2050년까지의 인구추계에서도 커다란 영향을 미치고 있음이 발견된다. 본 연구의 결과가 앞으로의 인구통계집계와 장래추계인구의 신뢰성이 있는 추정에 도움이 되었으면 한다.

참고 문헌

1. 오재석 (2003). 통계적 인구추정모형의 정립과 검증에 관한 연구, 조선대학교, 석사과정학위논문.
2. 전광희 (2006). 한국의 혼인력과 출산력 추이와 전망. 대한통계협회, 제32권 1, 2호 통합본, p60-90.
3. 최종후, 최봉호, 양우성, 김유진 (2000). 성장곡선모형에 의한 인구예측시스템, 한국인구학회, 제23권, 제1호, 197-215.
4. Kim, J.T. (2006). The Forecasting Analysis for the Numbers of

- High-school Graduate. *Proceedings of joint conference KDIS & KDAS*, 141-155.
5. Young P. (1993). Technological Growth Curve: A Competition of Forecasting Models, *Technological Forecasting and Social Change*, Vol.44, 375-389.

[2007년 3월 접수, 2007년 4월 채택]