Journal of the Korean Data & Information Science Society
2013, **24**(2), 311–320

http://dx.doi.org/10.7465/jkdi.2013.24.2.311
한국데이터정보과학회지

인구추계를 위한 가중비례추정모형†

윤용화[1] · 김종태[2]

[12]대구대학교 전산통계학과

접수 2013년 2월 1일, 수정 2013년 3월 5일, 게재확정 2013년 3월 13일

요 약

본 연구의 목적은 학생 (인구)수를 예측하기 위한 방법을 제시하는데 있다. 일반화된 가중비례모형들을 제시하고, 제시된 모형들을 이용하여 2029년까지의 학생 (인구)수를 추계하였다. 몬테칼로 모의실험의 결과 제시된 가중비례추정모형은 인구예측에 있어서 상당한 신뢰성과 예측력을 보인다. 결론적으로 추정된 고등학교 3학년 학생수는 대학들의 입학정원 수와 비교하면, 2019년부터 고3 학생수가 적게 나타나지만, 현재 대학들의 실제 입학자 수와 비교하면 2014년부터 입학생 수가 부족한 현상이 나타난다. 그러므로 입학정원과 입학정원 외의 신입생 수를 줄이지 않는 경우에는 2015년부터 입학생을 채우지 못하는 학교들이 생겨나게 될 것이다.

주요용어: 가중비례추정모형, 이동평균, 인구추정, 학생수추계.

1. 서론

인구를 추계하는 문제는 쉽지 않다. 데이터의 수집에 있어서 먼저 기존의 인구 데이터의 신뢰성 문제에 직면하게 된다. 신뢰성 문제를 해결하고 난 후에는 데이터 부재라는 또 다른 문제와 직면한다. 정부에서 공식적으로 발표한 통계 데이터도 최근 10년에서 15년 분량의 데이터만 있을 뿐, 1998년 이전 연령별 (각세별) 인구 데이터를 찾기가 쉽지 않다. 어쩌면 존재하지 않을 수도 있을 것이다. 이러한 문제를 극복해도 어떤 통계적인 방법들을 예측방법으로 사용하는가에 따라서 예측결과는 매우 큰 차이를 보인다. 그리고 사용된 통계적 방법의 예측 결과의 신뢰성을 검증할 문헌들을 찾기가 어렵다.

Kim (2005a, 2005b)은 대구와 경북지역을 중심으로 고3학생수를 예측하고, 대구와 경북지역의 대학 신입생 수와 비교 분석하였다. The Royal Society (1985)에서 Preece, Bunhill, Vetta, Bassett, Holt 등의 학생 수 추계에 대한 의견 및 토론들을 개제했고, Moore (1983)는 고등교육 (현재의 대학) 학생수 추계를 주제로 방법론을 제시했다. 인구수를 추정하기 위한 방법으로 Meade (1988)은 수정된 로지스틱 모형을 제시했고, Raeside (1988)는 인구수의 경향을 찾기 위한 모형으로 로지스틱 모형을 이용하였다. Kim (2009)은 학년진급률에 따른 학생 수 예측 방법으로 $m \times n$이동평균법에 의한 학생수 추정방법을 제시하고, 제시한 비례이동평균의 예측결과와 Korean Educational Development Institute (2005, 2006, 2007)에서 로지스틱함수를 이용하여 예측한 2005년, 2006년, 2007년의 고3학생 수 예측결과와 비교 분석하였다. Yoon과 Kim (2012)은 비선형 회귀모형을 이용한 학년별 학생수 추계를 하였다.

교육과학기술부 산하 Korean Educational Development Institute (2005, 2006, 2007)의 Educational Statistics & Information 시스템은 지방행정자치별 초 · 중 · 고등학교 학생 수에 대하여 2020년,

---

† 이 논문은 2010년도 대구대학교 학술연구비 지원에 의한 연구임.
[1] (712-714) 경상북도 경산시 진량면 내리동, 대구대학교 전산통계학과, 교수.
[2] 교신저자: (712-714) 경상북도 경산시 진량면 내리동, 대구대학교 전산통계학과, 교수.
E-mail: jtkim@daegu.ac.kr

2021년, 2022년까지 각각 학생 수의 예측결과를 제공하였다. 여기서 사용한 학생수 예측모형은 선형함수를 모델로 하여, 선형함수의 모수를 로지스틱성장곡선모형, 혹은 로지스틱지수평활 모형에 의하여 추정하였다. 그러나 시계열 예측에 있어서 로지스틱 함수를 적용할 경우에 초기값에 큰 영향을 받는 것으로 알려졌다 (Kim, 1994).

본 연구의 목적은 인구추정을 위하여 일반화된 추정모형인 가칭 p년차 가중비례모형을 제시하는데 있다. 위의 문헌들에서 제시한 대부분의 통계적 방법들은 1차 단순비례모형에서 선형모형 형태의 모수들을 추정하는 것이다. 본 연구는 2년차 ($p = 2$) 가중비례모형을 제시하고 장래 학생수를 추계하는데 목적이 있다. 제시된 2차 가중비례모형은 몬테칼로 모의실험의 결과 인구예측에 있어서 상당한 신뢰성과 예측력을 보인다.

본 연구에 제시된 가중비례추정모형은 예로써 향후 18년 동안 1세 인구부터 6세의 연령별 인구와 초등학교 1학년에서 고등학교 3학년 까지 학생수를 2030년까지 순차적으로 예측하였다.

2. 학생 (인구) 결합통계와 비례추정식

예측을 위하여 사용된 0세에서 6세까지의 연령별 인구통계는 연령별 (각세별) 주민등록인구 통계는 Statistics Korea (1998-2009)와 Ministry of Public Administration and Security (1998-2011)의 통계자료를 사용하였다. 또 다른 0세에서 6세까지의 연령별 인구통계로 Statistics Korea (2011)의 장래추계인구 통계를 사용할 수 있다. 그러나 통계청의 연령별 장래추계인구 통계에서는 2000년 이전의 연령별 인구 데이터를 구할 수 없다.

Korean Educational Development Institute (1982-2012)에서 1982년에서 2012년까지의 초 · 중 · 고등학교 학년별 학생수에 대한 통계 데이터를 구할 수 있다. 만 7세부터 초등학교 1학년에 입학하는 것으로 가정한다면 0세에서 6세까지의 연령별 데이터와 초등학교 1학년에서 고등학교 3학년까지의 학년별 데이터를 결합한 결합통계표를 사용하여도 무리가 없다고 가정한다.

N_y^d를 y (y=1982년, 1982년, $\cdots$, 2030년)년의 d (d=1, 2, $\cdots$, 19) 학생 (인구)수로 다음과 같이 정의하자.

$$N_y^d = y\text{년 } d\text{학년 학생수 (혹은 연령 인구수)}, \tag{2.1}$$

$$d = \begin{cases} 1\ (\mathrm{a0}), 2\ (\mathrm{a1}), 3\ (\mathrm{a2}), 4\ (\mathrm{a3}), 5\ (\mathrm{a4}), 6\ (\mathrm{a5}), 7\ (\mathrm{a6}), & \text{각각 0세, 1세, } \cdots\text{, 6세인구수,} \\ 8\ (\mathrm{e1}), 9\ (\mathrm{e2}), 10\ (\mathrm{e3}), 11\ (\mathrm{e4}), 12\ (\mathrm{e5}), 13\ (\mathrm{e6}), & \text{각각 초등 1, 2, } \cdots\text{, 6학년 학생수,} \\ 14\ (\mathrm{m1}), 15\ (\mathrm{m2}), 16\ (\mathrm{m3}), & \text{각각 중학 1, 2, 3학년 학생수,} \\ 17\ (\mathrm{h1}), 18\ (\mathrm{h2}), 19\ (\mathrm{h3}), & \text{각각 고등 1, 2, 3학년 학생수.} \end{cases}$$

다음의 y년도의 p년차 (진학 비율)들을 다음과 같이 정의하자. p=1, 2, $\cdots$, 대하여 $d-p>0$일 때,

$$y\text{년도의 } p\text{년차 단위 진학률} : P_{y,y-p}^{d,d-p} = N_y^d / N_{y-p}^{d-p}. \tag{2.2}$$

여기서 진학률 $P_{y,y-p}^{d,d-p}$의 의미는 $y-p$ 년 $d-p$학년 학생이 p 년이 지난 후에, y 년 d학년 학생이 되는 비율을 나타낸다. 예를 들어, 1, 2, 3년차의 단위 진학률은 다음과 같이 정의된다.

$$\begin{aligned} &\text{1년차 단위 진학률} : P_{y,y-1}^{d,d-1} = N_y^d / N_{y-1}^{d-1}, \\ &\text{2년차 단위 진학률} : P_{y,y-2}^{d,d-2} = N_y^d / N_{y-2}^{d-2}, \\ &\text{3년차 단위 진학률} : P_{y,y-3}^{d,d-3} = N_y^d / N_{y-3}^{d-3}. \end{aligned}$$

y가 미래의 시점으로 가정하고, $y-p$와 $d-p$에 대하여 N_{y-p}^{d-p}의 값들을 이미 알고 있다고 가정하면, 미지의 값 N_y^d을 추정하기 위해서 y년의 p년차 단위 진학률들은 y년의 1년차 진학률에 크게 벗어나지

않을 것이라는 가정을 하고, 다음과 같이 비례식을 두자.

$$p\text{년차 단위 진학비례식} : P^{d,d-p}_{y,y-p} \equiv P^{d,d-1}_{y,y-1} = N^{d-p}_{y-p} : N^{d-p}_{y-p-1} \equiv N^d_y : N^d_{y-1}. \tag{2.3}$$

예를 들어 $p = 1, 2, 3$ 에 대하여, 1, 2, 3년차 단위의 비례식은 다음과 같다.

$$1\text{년차 단위 진학비례식} : P^{d,d-1}_{y,y-1} \equiv P^{d-1,d-2}_{y-1,y-2} = N^d_y : N^{d-1}_{y-1} \equiv N^{d-1}_{y-1} : N^{d-2}_{y-2},$$

$$2\text{년차 단위 진학비례식} : P^{d,d-1}_{y,y-1} \equiv P^{d-2,d-3}_{y-2,y-3} = N^d_y : N^d_{y-1} \equiv N^{d-2}_{y-2} : N^{d-3}_{y-3},$$

$$3\text{년차 단위 진학비례식} : P^{d,d-1}_{y,y-1} \equiv P^{d-3,d-4}_{y-3,y-4} = N^d_y : N^d_{y-1} \equiv N^{d-3}_{y-3} : N^{d-4}_{y-4}.$$

식 (2.2)와 식 (2.3)에서 p년차 단위 진학비례식을 이용하여 p년차 단위 y년도 d학년 학생수 비례추정식 $\widehat{N}^d_y$은 다음과 같다.

$$\widehat{N}^d_y \equiv \frac{N^d_{y-1}}{N^{d-p}_{y-p-1}} N^{d-p}_{y-p} = P^{d,d-p}_{y-1,y-p-1} N^{d-p}_{y-p}. \tag{2.4}$$

예를 들어, 1, 2, 3년차 단위의 비례추정식은 다음과 같다.

1년차 단위 y년도 d학년 학생수 추정 $\widehat{N}^d_y$을 위한 비례추정 :

$$\widehat{N}^d_y \equiv \frac{N^d_{y-1}}{N^{d-1}_{y-2}} N^{d-1}_{y-1} = P^{d,d-1}_{y-1,y-2} N^{d-1}_{y-1}.$$

2년차 단위 y년도 d학년 학생수 추정 $\widehat{N}^d_y$을 위한 비례추정 :

$$\widehat{N}^d_y \equiv \frac{N^d_{y-1}}{N^{d-2}_{y-3}} N^{d-2}_{y-2} = P^{d,d-2}_{y-1,y-3} N^{d-2}_{y-2}.$$

3년차 단위 y년도 d학년 학생수 추정 $\widehat{N}^d_y$을 위한 비례추정 :

$$\widehat{N}^d_y \equiv \frac{N^d_{y-1}}{N^{d-3}_{y-4}} N^{d-3}_{y-3} = P^{d,d-3}_{y-1,y-3} N^{d-3}_{y-3}.$$

3. 일반화 가중비례추정모형과 학생수추계

2절의 식 (2.4)에서 구한 p년차 단위 y년도 d학년 학생수 비례추정식 $\widehat{N}^d_y$들을 $p = 1, 2, \cdots, m$일 때, 이들 m개의 식을 모두 합하여 구하면 다음과 같다.

$$\begin{aligned} mN^d_y &\equiv \frac{N^d_{y-1}}{N^{d-1}_{y-2}} N^{d-1}_{y-1} + \frac{N^d_{y-1}}{N^{d-2}_{y-3}} N^{d-2}_{y-2} + \cdots + \frac{N^d_{y-1}}{N^{d-m}_{y-m-1}} N^{d-m}_{y-m} + \epsilon \\ &\equiv P^{d,d-1}_{y-1,y-2} N^{d-1}_{y-1} + P^{d,d-2}_{y-1,y-3} N^{d-2}_{y-2} + \cdots + P^{d,d-m}_{y-1,y-m-1} N^{d-m}_{y-m} + \epsilon. \end{aligned} \tag{3.1}$$

식 (3.1)에서,

$$N^d_y \equiv \frac{1}{m} \sum_{k=1}^{m} P^{d,d-k}_{y-1,y-p-1} N^{d-k}_{y-k} + \epsilon. \tag{3.2}$$

위의 식 (3.2)에서 균등 가중치 m대신에 일반화 된 가중치 w_k를 사용함으로 일반화된 m차 가중비례추정모형을 설정할 수 있다. m차 일반화 가중비례추정 모형을 다음과 같이 정의하자.

$$N^d_y \equiv \sum_{k=1}^{m} (\frac{1}{w_k}) P^{d,d-k}_{y-1,y-p-1} N^{d-k}_{y-k} / (\sum_{k=1}^{m} \frac{1}{w_k}) + \epsilon. \tag{3.3}$$

여기서 $\sum_{k=1}^{m}(\frac{1}{w_k}/(\sum_{k=1}^{m}\frac{1}{w_k}))=1$ 이다.

$m=1$인 경우에, 식 (3.3)으로부터, 1차 단순비례추정모형은

$$N_y^d \equiv P_{y-1,y-2}^{d,d-1} N_{y-1}^{d-1} + \epsilon, \tag{3.4}$$

가 된다. 기존의 문헌들에서는 식 (3.4)의 $P_{y-1,y-2}^{d,d-1}$을 추정함으로써 N_y^d를 추정하는 다양한 방법들을 제시하였다. $P_{y-1,y-2}^{d,d-1}$의 추정방법으로 히든 마코브 체인, 로지스틱 회귀 모형, 시계열 모형, 이동평균 모형을 이용하는 방법 등이 사용되었다. 그러나 추정된 모수들의 작은 오차에도 추정된 값은 상당히 민감하게 작용하고, 오차들이 증폭됨으로서 인구수의 예측력이 낮아짐을 보였다.

기준연도를 $y=2013$년으로 두고 식 (2.1)의 d의 정의에 따르면, 고등학교 3학년부터 1세까지의 인구수를 추정하기 위한 1차 단순비례추정모형은 다음과 같다.

$$N_{2013}^{19} \equiv P_{2012,2011}^{19,18} N_{2012}^{18} + \epsilon,$$
$$N_{2013}^{18} \equiv P_{2012,2011}^{19,17} N_{2012}^{17} + \epsilon,$$
$$N_{2013}^{17} \equiv P_{2012,2011}^{19,16} N_{2012}^{16} + \epsilon.$$

의 순서로 $N_{2013}^{1} \equiv P_{2012,2011}^{19,1} N_{2012}^{1} + \epsilon$ 까지 계속해서 추정해 내간다. 이러한 추정은 2012년의 0세 인구 혹은 출생아수를 알고 있다는 가정 하에서 향후 18년 후의 고등학교 3학년 학생수를 추정할 수 있다. 더 나아가 2012년의 출생아수가 나이가 들어서 100세가 넘어서 인구가 다 소멸될 때까지 연령별 인구수를 계산해 낼 수 있을 것이다.

식 (3.3)에서 $m=2$인 경우에는 위의 1차 단순비례추정모형에서 2년차 단위 비례추정모형을 더하여 각각에 가중치를 두고, 평균제곱오차를 가장 적게하는 적합한 가중치를 추정하는 모형이다. 1년차 단위의 단순비례추정 모형보다도 추정값에 대한 판단기준방법들을 사용하여 비교해 볼 때, 가중비례추정모형이 더 우수한 추정력을 보인다.

$m=2$인 경우에, 식 (3.3)으로부터, 가중치 $\lambda_1 \equiv \frac{w_2}{w_1+w_2}$ 와 가중치 $\lambda_2 \equiv \frac{w_1}{w_1+w_2}$으로 두고, $m=2$차 비례추정모형은 다음과 같다.

$$\begin{aligned} N_y^d &\equiv (P_{y-1,y-p-1}^{d,d-1} N_{y-1}^{d-1}/w_1 + P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2}/w_2)/(\frac{1}{w_1}+\frac{1}{w_2}) + \epsilon \\ &\equiv \frac{w_2}{w_1+w_2} P_{y-1,y-p-1}^{d,d-1} N_{y-1}^{d-1} + \frac{w_1}{w_1+w_2} P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2} + \epsilon \\ &\equiv \lambda_1 P_{y-1,y-p-1}^{d,d-1} N_{y-1}^{d-1} + (1-\lambda_1) P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2} + \epsilon. \end{aligned} \tag{3.5}$$

여기서 적합한 λ_1의 값을 구하여 N_y^d의 추정값을 구할 수 있다. 본 연구에서는 가중치 λ_1 과 λ_2를 구하기 위하여 몬테칼로 모의실험을 이용하였다. 기존의 이미 알고 있는 인구수와 임의의 가중치 값 w_1과 w_2을 주고 구한 추정된 인구수들과의 차이를 오차할 때, 이들 오차항의 평균과 표준편차를 최소로 하는 w_1과 w_2의 값을 몬테칼로 모의실험을 이용하여 구하여, 가중치 λ_1 과 λ_2의 추정치를 구하였다. 이들의 모의실험 결과는 다음 절의 Table 4.1과 Table 4.2에서 보인다.

예로써 기준연도를 $y=2013$년 두면, 2차 가중비례추정모형은 다음과 같다.

$$N_{2013}^{19} \equiv \frac{w_2}{w_1+w_2} P_{2012,2011}^{19,18} N_{2012}^{18} + \frac{w_1}{w_1+w_2} P_{2011,2010}^{18,17} N_{2011}^{17} + \epsilon,$$
$$N_{2013}^{18} \equiv \frac{w_2}{w_1+w_2} P_{2012,2011}^{18,17} N_{2012}^{17} + \frac{w_1}{w_1+w_2} P_{2011,2010}^{17,16} N_{2011}^{16} + \epsilon.$$

계속해서 다음과 같이 $y = 2013$년의 인구수를 고등학교 3학년부터 1세 인구수까지 추계할 수 있다.

$$N^3_{2013} \equiv \frac{w_2}{w_1+w_2} P^{3,2}_{2012,2011} N^2_{2012} + \frac{w_1}{w_1+w_2} P^{2,1}_{2011,2010} N^1_{2011} + \epsilon.$$

그러나 여기서 $m = 2$인 2차 가중비례추정모형에서는 단순비례모형보다도 1세의 인구수 데이터를 더 필요한 이유로, 2013년의 1세의 인구수를 추정할 수 없다. 그러나 단순비례추정모형에서는 1세의 인구수의 추정이 가능하므로 단순비례추정 모형에서 추정된 1세의 인구수를 사용하여, 2013년의 1세의 인구수 추정값으로 사용한다. 2013년 인구수 추정값을 가지고, 2014년의 인구수를 위와 같은 방법으로 다시 추정할 수 있다.

$$N^{19}_{2014} \equiv \frac{w_2}{w_1+w_2} P^{19,18}_{2013,2012} N^{18}_{2013} + \frac{w_1}{w_1+w_2} P^{18,17}_{2012,2011} N^{17}_{2012} + \epsilon,$$

$$N^{18}_{2014} \equiv \frac{w_2}{w_1+w_2} P^{19,18}_{2013,2012} N^{18}_{2013} + \frac{w_1}{w_1+w_2} P^{18,17}_{2012,2011} N^{17}_{2012} + \epsilon,$$

같은 방법으로 2014년의 2세의 인구수까지 추정할 수 있다.

$$N^{3}_{2014} \equiv \frac{w_2}{w_1+w_2} P^{19,18}_{2013,2012} N^{18}_{2013} + \frac{w_1}{w_1+w_2} P^{18,17}_{2012,2011} N^{17}_{2012} + \epsilon,$$

현재 Population Statistics Based on Resident Registration는 2011년의 통계만 가지고 있기 때문에 주민등록인구 통계 자료를 사용하면, 18년 후인 2029년까지의 인구수를 추계할 수 있고, Statistics Korea의 장래추계인구를 사용한다면 2030년까지의 인구수를 추계할 수 있다. 그러나 장래추계인구의 추계 데이터에서 2013년 이후의 추계 인구수의 신뢰성에 대해서는 확실하지 않아서 2012년의 추계 데이터만을 사용하였다.

Table 3.1 The numbers of estimated students by 2nd weighted proportion estimation model using population data of statistical yearbooks

year	e1	e2	e3	e4	e5	e6	m1	m2	m3	h1	h2	h3
2013												634550
2014											624702	615556
2015										632882	614105	605115
2016									609995	603162	585267	576699
2017								593282	589972	583363	566056	557769
2018							600237	597874	594539	587879	570437	562087
2019						530554	528529	526449	523512	517648	502290	494937
2020					464882	464423	462651	460831	458260	453126	439683	433246
2021				473817	472808	472341	470539	468687	466072	460851	447178	440632
2022			455429	454650	453681	453233	451504	449727	447218	442208	429089	422807
2023		421011	420313	419595	418701	418287	416692	415052	412736	408113	396005	390207
2024	434747	434059	433340	432600	431678	431252	429607	427916	425528	420762	408278	402301
2025	479556	478797	478004	477187	476170	475700	473886	472020	469387	464129	450359	443766
2026	453688	452970	452220	451447	450485	450040	448323	446559	444068	439093	426066	419829
2027	434095	433408	432690	431950	431030	430605	428962	427274	424890	420131	407666	401698
2028	459869	459141	458381	457597	456622	456171	454431	452643	450117	445075	431871	425548
2029	463255	462522	461756	460966	459984	459530	457777	455975	453432	448352	435050	428681

Table 3.2 The numbers of estimated students by 2nd weighted proportion estimation model using future population estimation

year	e1	e2	e3	e4	e5	e6	m1	m2	m3	h1	h2	h3
2013												634550
2014											624702	615556
2015										632882	614105	605115
2016									609995	603162	585267	576699
2017								593282	589972	583363	566056	557769
2018							600237	597874	594539	587879	570437	562087
2019						530554	528529	526449	523512	517648	502290	494937
2020					464882	464423	462651	460831	458260	453126	439683	433246
2021				473817	472808	472341	470539	468687	466072	460851	447178	440632
2022			455429	454650	453681	453233	451504	449727	447218	442208	429089	422807
2023		420700	419979	419260	418366	417953	416359	414720	412406	407787	395688	389896
2024	413708	412750	412042	411336	410460	410055	408490	406883	404613	400080	388210	382527
2025	434538	433532	432789	432047	431127	430701	429058	427369	424985	420224	407757	401788
2026	461425	460356	459567	458780	457802	457350	455606	453812	451281	446226	432987	426648
2027	428682	427689	426956	426225	425316	424896	423276	421610	419257	414561	402262	396373
2028	422802	421823	421100	420379	419483	419069	417470	415827	413507	408875	396744	390936
2029	457607	456547	455765	454984	454014	453566	451836	450057	447547	442533	429404	423118
2030	428913	427920	427186	426454	425546	425126	423504	421837	419484	414785	402478	396587

4. 모의실험

2001년에서 2012년까지의 초등 4학년부터 고등학교 3학년까지의 학생수에 대한 기존의 데이터를 미지수로 두었다. 3절에서 제시한 가중비례추정모형으로 이들 미지의 데이터들을 추정하고, 추정된 데이터와 실제의 데이터와의 오차의 특성을 분석하여 가중비례추정모형이 추정력을 조사하고, 모의실험을 이용하여 적합한 가중치 λ을 추정한다.

모의실험을 위하여 먼저 2001년에서 2012년까지의 고등학교 3학년에서 초등학교 4학년까지의 학생수를 미지의 인구수로 둔다. 초등학교 4학년까지로 대상으로 한 이유는 $m=2$인 2차 가중비례추정 모형을 하려면 1998년부터 2000년까지의 3년간의 데이터가 필요한데 현재 정부에서 제공하는 연령별 데이터는 주민등록인구가 1998년부터이고, 장래추계인구통계가 2000년부터 기록되어 있다.

식 (3.3)에서 세 번째 식을 정리하면 다음과 같다.

$$\begin{aligned} N_y^d &\equiv \lambda_1 P_{y-1,y-p-1}^{d,d-1} N_{y-1}^{d-1} + (1-\lambda_1) P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2} + \epsilon, \\ &\equiv \lambda_1 (P_{y-1,y-p-1}^{d,d-1} N_{y-1}^{d-1} - P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2}) + P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2} + \epsilon, \end{aligned} \tag{4.1}$$

모형 (4.1)에서 가중치 λ를 추정하는 많은 방법들이 고안되어질 수 있다. 각각의 연령별 d에 따라서 다음과 같은 가중치 λ_1이 추정되어진다.

$$\lambda_1 = \frac{N_y^d - P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2}}{(P_{y-1,y-p-1}^{d,d-1} N_{y-1}^{d-1} - P_{y-1,y-p-1}^{d,d-2} N_{y-2}^{d-2})}$$

그러나 불행하게도 모의실험의 결과 실제 데이터의 변동에 의해서 λ_1의 값은 학년과 년도에 따라서 불규칙적인 큰 변동을 가지고 안정적이지 않다. 그 결과 λ_1의 값은 굉장히 큰 이상치의 값들을 갖는 경우가 발생한다. 이 문제를 해결하기 위해서는 중앙값을 이용한 로버스트 추정방법들을 이용하여 해결하는 방법이 있다. 그러나 이것에 대한 연구는 다음으로 미루고 본 연구에서는 몬테칼로 모의실험을 이용하여 평균제곱오차나 오차의 평균 혹은 오차의 표준편차를 가장 작게 하는 가중치 λ를 구한 다음

2013년 이후의 미래의 학생수를 추정하는데 사용하였다. 오차 $|N_y^d - \widehat{N}_y^d|$의 평균과 표준편차를 최소로 하는 모의실험을 이용한 가중치 $\widehat{w}_1$과 $\widehat{w}_2$의 결과는 다음과 같다.

Table 4.1은 1차 단순비례추정모형과 2차 비례추정모형 학생수추정에 대한 모의실험 결과의 오차의 평균과 표준편차를 나타낸 것이다. 모의실험에서 오차의 평균과 표준편차를 적게하는$\widehat{w}_1$과 $\widehat{w}_2$를 구해보면, $\widehat{w}_1 = 0.01$과 $\widehat{w}_2 = 0.138774$의 값으로 갈수록 오차의 평균이 거의 0에 접근하고, 오차의 표준편차는 2,707에 가까워짐을 보인다.

Table 4.1 Means and standard deviations of the errors for 1st · 2nd weighted proportion estimation models

	w_1	w_2	mean	SD
1st			-2,775	3,007
2nd	2	5	-10,619	6,270
	0.8	16	-4,316	3,497
	0.5	16	-3,769	3,305
	0.1	16	-2,982	3,063
	0.05	16	-2,897	3,035
	0.01	-0.5	-1,636	2,641
	0.01	-0.25	-1,329	2,736
	0.01	-0.2	-932	2,702
	0.01	-0.1	1,307	2,908
	0.01	-0.19	-825	2,697
	0.01	-0.18	-705	2,693
	0.01	-0.16	-414	2,691
	0.01	-0.14	-28	2,706
	0.01	-0.13	217	2,725
	0.01	-0.138	18	2,709
	0.01	-0.1385	6	2,708
	0.01	-0.1386	4	2,708
	0.01	-0.1387	2	2,707
	0.01	-0.1388	-0.67	2,707
	0.01	-0.13878	-0.149	2,707
	0.01	-0.13877	0.081	2,707
	0.01	-0.138773	0.01186	2,707
	0.01	-0.138774	-0.01106	2,707
	0.008	-0.1387735	-618	2,691

Table 4.2 Means and standard deviations of the errors for 3rd weighted proportion estimation models

	w_1	w_2	w_3	mean	SD
3th	0.01	-0.2	0.6	-1,541	2,775
	0.01	-0.2	0.5	-1,656	2,796
	0.01	-0.2	0.4	-1,826	2,832
	0.01	-0.2	-0.6	-268	2,706
	0.01	-0.2	-0.5	-128	2,718
	0.01	-0.2	-0.45	-33	2,729
	0.01	-0.2	-0.44	-12	2,732
	0.01	-0.2	-0.43	11	2,735
	0.01	-0.2	-0.4	87	2,745
	0.01	-0.2	-0.3	458	2,813
	0.01	-0.2	-0.2	1,255	3,046
	0.01	-0.2	2	-1,120	2,717
	0.01	-0.2	4	-1,027	2,709
	0.01	-0.2	10	-970	2,705
	0.01	-0.2	15	-957	2,704
	0.01	-0.2	50	-940	2,703
	0.01	-0.2	100	-936	2,703
	0.01	-0.25	-0.44	-452	2,704
	0.01	-0.22	-0.44	-214	2,715

Table 4.2는 식 (3.3)에서 $m = 3$인 3차 비례추정모형 학생수추정에 대한 모의실험 결과의 오차의 평균과 표준편차를 나타낸 것이다. 모의실험에서 오차의 평균과 표준편차를 적게하는 w_1과 w_2, w_3를 구해보면, 오차의 평균은 $\widehat{w}_1 = 0.01$과 $\widehat{w}_2 = 0.2$에서 $\widehat{w}_3 = -0.44$와 -0.43의 사이의 값으로 갈수록 오차의 평균이 가장 적게 접근한다. 오차의 표준편차는 $\widehat{w}_1 = 0.01$과 $\widehat{w}_2 = 0.2$, $\widehat{w}_3 = 50$일때, 2,703에 가까워짐을 보인다.

5. 고3학생수추정과 입학정원과 입학자수 비교

3절에서 제시된 가중비례추정모형에서 구한 전국 고등학교 3학년 학생수의 추계결과를 가지고, 전국 전문대학, 대학교 그리고 교육대학교의 입학정원수과 입학생수와 비교 분석한다.

Table 5.1 Differences of the numbers of estimated 3rd year high school students, the entrance quota and the new student for admission of universities and colleges

	1st(P)-quota	1st(P)-adm	2nd(P)-quota	2nd(P)-adm	2nd(F)-quota	2nd(F)-adm
2012	77,477	16,476	77,477	16,476	77,477	16,476
2013	78,219	17,218	79,735	18,734	79,735	18,734
2014	58,416	-2,585	60,741	-260	60,741	-260
2015	47,341	-13,660	50,300	-10,701	50,300	-10,701
2016	19,426	-41,575	21,884	-39,117	21,884	-39,117
2017	691	-60,310	2,954	-58,047	2,954	-58,047
2018	5,021	-55,980	7,272	-53,729	7,272	-53,729
2019	-62,017	-123,018	-59,878	-120,879	-59,878	-120,879
2020	-123,588	-184,589	-121,569	-182,570	-121,569	-182,570
2021	-116,365	-177,366	-114,183	-175,184	-114,183	-175,184
2022	-134,135	-195,136	-132,008	-193,009	-132,008	-193,009
2023	-166,971	-227,972	-164,608	-225,609	-164,919	-225,920
2024	-155,419	-216,420	-152,514	-213,515	-172,288	-233,289
2025	-114,076	-175,077	-111,049	-172,050	-153,027	-214,028
2026	-138,148	-199,149	-134,986	-195,987	-128,167	-189,168
2027	-156,294	-217,295	-153,117	-214,118	-158,442	-219,443
2028	-133,193	-194,194	-129,267	-190,268	-163,879	-224,880
2029	-135,217	-196,218	-126,134	-187,135	-131,697	-192,698

Table 5.1은 가중비례추정 모형에서 추정된 고3학생수 추정과 대학입학정원수의 합계와 입학자수를 비교한 것이다. Table 5.1에서 '1st(P)-quota'은 주민등록 0세에서 6세인구를 이용한 1차 단순비례추정모형에서 구한 고등학교 3학년 추정 학생수와 대학들 (전문대학, 일반대학, 교육대학들을 합한) 입학정원과의 차이를 의미한다. '1st(P)-adm'은 주민등록 0세에서 6세인구를 이용한 1차 단순비례추정모형에서 구한 고등학교 3학년 추정 학생수와 대학들의 입학자 수와의 차이를 의미한다. 입학자수와 입학정원과는 상당한 차이를 보인다. '2nd(P)-quota'은 주민등록 0세에서 6세인구를 이용한 2차 가중비례추정모형에서 구한 고등학교 3학년 추정 학생수와 대학들의 입학정원 수와의 차이를 의미한다. '2nd(P)-adm'은 주민등록 0세에서 6세인구를 이용한 2차 가중비례추정모형에서 구한 고등학교 3학년 추정 학생수와 대학들의 입학자 수와의 차이를 의미한다. '2nd(F)-quota'은 장래추계인구의 0세에서 6세인구를 이용한 2차 가중비례추정모형에서 구한 고등학교 3학년 추정 학생수와 대학들의 입학정원 수와의 차이를 의미한다. '2nd(F)-adm'은 장래추계인구의 0세에서 6세인구를 이용한 2차 가중비례추정모형에서 구한 고등학교 3학년 추정 학생수와 대학들의 입학자 수와의 차이를 의미한다.

결론적으로 Table 5.1의 결과에서 추정된 고등학교 3학년 학생수는 대학들의 입학정원과의 비교에서는 2019년부터 대학들의 입학정원 보다 고3 학생수가 적게 나타난다. 그러나 현재 대학들의 실제 입학자 수로 볼 때는 2014년부터 입학생 수가 부족한 현상이 나타난다. 그러므로 정원 외의 신입생 수를 줄이지 않는 경우에는 2015년부터 대학 입학생들을 채우지 못하는 미달 현상이 생겨나게 될 것이다.

참고문헌

Kim, J. T. (2005a). The forecasting about the numbers of the third graders in a high-school until 2022 year in Daegu. *Journal of the Korean Data & Information Science Society*, **16**, 933-942.

Kim, J. T. (2005b). The forecasting for the numbers of a high-school graduate and the number limit of matriculation in Kyungbook. *Journal of the Korean Data & Information Science Society*, **16**, 969-977.

Kim, J. T. (2009). The methods of forecasting for the number of student based on promotion proportion. *Journal of the Korean Data & Information Science Society*, **20**, 857-867.

Kim, Y. H. (1994). *Time series analysis*, Freeacademy, Seoul.

Korean Educational Development Institute (2005, 2006, 2007). *Educational statistics & forecasting*, Educational Statistics & Information, Seoul.

Korean Educational Development Institute (1982-2012). *Education statistical year book*, Educational Statistics & Information, Seoul.

Meade, N. (1988). A method logistic model applied to human population. *Journal of Royal Statistical Society A*, **151**, 491-498.

Ministry of Public Administration and Security (1998-2011). *Age-specific provincial population statistics based on resident registration*, Korean Statistical Information Service, Seoul.

Moore, P. G. (1983). Higher education: The next decade. *Journal of Royal Statistical Society A*, **146**, 213-245.

Raeside, R. (1988). The use of sigmoids in modelling and forecasting human population. *Journal of Royal Statistical Society A*, **151**, 499-513.

The Royal Society (1985). Projections of student numbers in higher education. *Journal of Royal Statistical Society A*, **148**, 175-213.

Statistics Korea (1998-2009). *Age-specific provincial population statistics based on resident registration*, Korean Statistical Information Service, Daejeon.

Statistics Korea (2011). *Population projections*, Korean Statistical Information Service, Daejeon.

Yoon, Y. H. and Kim, J. T. (2012). Estimations of the student numbers by nonlinear regression model. *Journal of the Korean Data & Information Science Society*, **23**, 71-77.

Journal of the Korean Data & Information Science Society
2013, **24**(2), 311–320

http://dx.doi.org/10.7465/jkdi.2013.24.2.311
한국데이터정보과학회지

The model of the weighted proportion estimation for forecasting the number of population[†]

Yong Hwa Yoon[1] · Jong Tae Kim[2]

[12]Department of Computing & Statistics, Daegu University

Received 1 February 2013, revised 5 March 2013, accepted 13 March 2013

Abstract

The purpose of this paper is to suggest the methods of forecasting the numbers of students. The generalized weighted proportion estimation models are suggested and used for forecasting the numbers of student until 2029. The results of the Monte Carlo simulation show that the suggested method is powerful for the forecasting. In conclusion, the numbers of the third grade high-school students will be less than the numbers of college admission quota from 2019.

Keywords: Forecasting of population, moving average, weighted proportion estimation model.

† This research was supported by the Daegu University Research Grant 2010.

[1] Professor, Department Computing & Statistics, Daegu University, Kyoungbook 712-714, Korea.

[2] Corresponding author: Professor, Department Computing & Statistics, Daegu University, Kyoungbook 712-714, Korea. E-mail: jtkim@daegu.ac.kr