

The Errors of Forecast Educational Statistics on Korean National Center for Education Statistics & Information¹⁾

Jongtae Kim²⁾

Abstract

Recently, Korean national center for education statistics & information exhibits the results of forecast educational statistics from 2007 to 2021 year. The purpose of this paper is to suggest the errors and problems on the results of forecast educational statistics and the distribution of the number of students.

Keywords : 교육예측통계, 교육통계서비스시스템, 로지스틱 성장 곡선 모형, 출산율추이

1. 연구의 배경과 서론

교육인적자원부 산하의 국가교육정보센터는 교육통계서비스시스템을 제공하고 있다. 교육통계서비스시스템(http://std.kedi.re.kr/jcgi-bin/pred_new/pred_frme.jsp)의 주요한 기능 중 하나는 교육예측통계인데 교육통계 자료를 기반으로 하여 향후 15년 동안의 주요 교육통계값을 예측하고 이를 장표와 그림 형태로 서비스하고 있다. 또한 예측값을 이용자가 손쉽게 재사용하거나 활용할 수 있도록 다운로드 기능도 제공하고 있다. 현재의 교육예측통계시스템은 2007년 이후부터 2021년까지의 초, 중, 고등학교의 학년별/시도별 학생수와 진급률, 진학률, 여학생 비율, 입학생과 졸업생수들을 추계하여 서비스를 제공해 놓았다. 또한 일반대학교의 재학생수, 계열비율, 여학생비율, 입학생수에 대한 교육통계 예측결과를 제공하고 있다. 이러한 예측결과는 우리나라 교육행정 및 지원, 교원의 임용, 교육환경과 설비 등 우리나라 교육정책의 수립과 계획에 매우 큰 영향을 미치는 중요한 자료로 활용되어진다. 이러한 이유로 교육통계의 예측은 보다 정확한 서비스를 제공해 줄 필요가 있다.

1) 이 논문은 2006년도 대구대학교 학술연구비 지원에 의한 연구임

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

본 연구는 국가교육정보센터의 교육통계 예측결과의 문제점과 오류들을 지적하고, 탐사적 방법으로 우리나라 초, 중, 고등학생수의 분포를 제시하고, 출생아 수의 분포(전광희 2006)와 초, 중, 고등학생수의 분포를 비교 분석하여 연도별 학생수의 변화를 규명하고, 보다 정확한 교육통계예측시스템에 도움을 주고자 하는데 그 목적이 있다.

교육통계 예측은 각 대학들의 입학생 유치 및 생존 전략에 많은 영향을 미친다. (참조: Kim(2006 2005a,b, 2004)). 다음절은 국가교육정보센터의 교육통계예측시스템에서 제공하는 예측과정과 예측이론을 소개한다. 3절에서는 교육통계의 예측결과의 오류와 문제점을 다루면서 우리나라 초, 중, 고등학생들의 연도별에 따른 분포를 제시한다. 4절에서는 결론 향후의 연구방향을 제시한다.

2. 교육통계예측시스템의 예측과정과 예측이론

교육통계예측시스템의 초, 중, 고 학생수의 예측은 다음과 같은 절차로 이루어진다. 초등학교 학생수의 예측을 위하여 통계청에서 발표한 장래인구추계자료를 사용하여 초등학교 취학연령인구의 수를 추정한다. 추정된 취학연령인구의 수와 추정된 취학률을 사용하여 초등학교 1학년 학생수를 예측하고, 예측된 초등학교 1학년 학생수를 사용하여 2학년 학생수를 예측한다. 이러한 과정을 순차적으로 거쳐서 초등학교 1학년 부터 6학년까지의 학생수를 예측한다. 이와 같은 예측작업은 시도별, 성별로 각각 진행된다.

중학교 학생수에 대한 예측값은 초등학교 6학년 학생수 예측값을 기준으로 하여 산출된다. 즉, 예측된 초등학교 6학년 학생수와 예측된 진학률을 사용하여 중학교 1학년 학생수가 예측되며, 이를 기준으로 하여 중학교 2학년, 3학년의 학생수가 예측된다. 이와 같은 예측작업은 시도별, 성별로 각각 진행된다.

고등학교 학생수는 일반계고등학교의 학생수와 실업계고등학교의 학생수로 구분하여 예측된다. 먼저, 예측된 중학교 3학년 학생수와 예측된 진학률을 사용하여 전체고등학교 1학년 학생수가 예측되며, 예측된 일반계고등학교 1학년 학생수와 실업계고등학교 1학년 학생수 비율에 의하여 전체고등학교 1학년 학생수 중에서 일반계고등학교와 실업계고등학교의 학생수가 예측된다. 각각의 1학년 학생수 예측값을 기준으로 하여 2학년, 3학년의 학생수가 예측된다. 이와 같은 예측작업은 시도별, 성별로 각각 진행된다.

교육인적자원부의 교육통계예측시스템의 초등학교, 중학교, 일반계고등학교, 실업계고등학교의 학년별 학생수에 대한 향후 15년간의 예측값을 제공하는 교육통계예측시스템에서는 다음과 같은 예측방법이 사용된다.

초·중·고등학교 학생수의 예측을 위한 일반모형을 설정한 후에 이를 기본으로 하여 학교급별·학년별 예측모형을 수립한다. t 년의 X_t 학생수는 다음 식(2.1)의 모형에 의해서 예측된다.

$$X_t = f(X_{t-1}, O_t, I_t). \quad (2.1)$$

여기서 X_{t-1} 은 X_t 의 시차변수로 $(t-1)$ 년도의 학생수, O_t 는 $(t-1)$ 년도에서 t 년도 사이에 사망, 질병, 품행, 유학/이민 등의 이유로 학교를 떠난 학생의 수, 그리고 I_t 는

($t-1$)년도에서 t 년 사이에 복학 및 재입학 등으로 학교에 들어온 학생의 수이다. 이때 X_{t-1} 은 외생적으로 결정된 비확률변수(non-stochastic variable)이고, O_t 와 I_t 는 ($t-1$)년도에서 t 년도 사이 1년 동안의 학생수 변동을 나타내는 확률변수이다.

식(2.1)의 예측모형은 다음 식(2.2)와 같은 선형함수의 형태를 갖는다고 가정되었다.

$$X_t = \beta X_{t-1} + \phi O_t + \theta I_t + \epsilon_t. \quad (2.2)$$

이때 식(2.2)를 사용하여 X_{t+i} , ($i > 0$)을 추정하기 위해서는 O_{t+i} 와 I_{t+i} 의 추정값이 필요하게 된다. ϵ_t 는 오차항이다. O_t 와 I_t 는 개별적으로 추정되지 않고 통합적으로 추정될 수 있도록 다음 식(2.3)의 식이 사용되었다.

$$X_t = \left(\beta + \frac{1}{X_{t-1}} (\phi O_t + \theta I_t) \right) X_{t-1} + \epsilon_t. \quad (2.3)$$

여기서 $\beta_t = \beta + \frac{1}{X_{t-1}} (\phi O_t + \theta I_t)$ 로 표현하면 위의 식(2.3)은 식(2.4)와 같이 간단한 형태가 된다.

$$X_t = \beta_t X_{t-1} + \epsilon_t. \quad (2.4)$$

이때 식(2.2)의 β 는 시간에 따라 일정한 값을 가지는 모수를 나타내고, β_t 는 시간에 따라 변화하는 확률변수가 된다. β_t 는 경험적인 판단으로 S자형 성장곡선모형을 가정하여 t 에 대하여 다음의 두 가지 방법으로 추정될 수 있다.

첫째, β_t 는 다음과 같은 로지스틱성장곡선 모형에 의하여 예측될 수 있다. 이 모형은 최소제곱법에 의하여 모수 a 와 b 를 추정할 수 있다.

$$\beta_t = u_1 + \frac{u_2}{1 + \exp^{(a+bt)}} \quad (2.5)$$

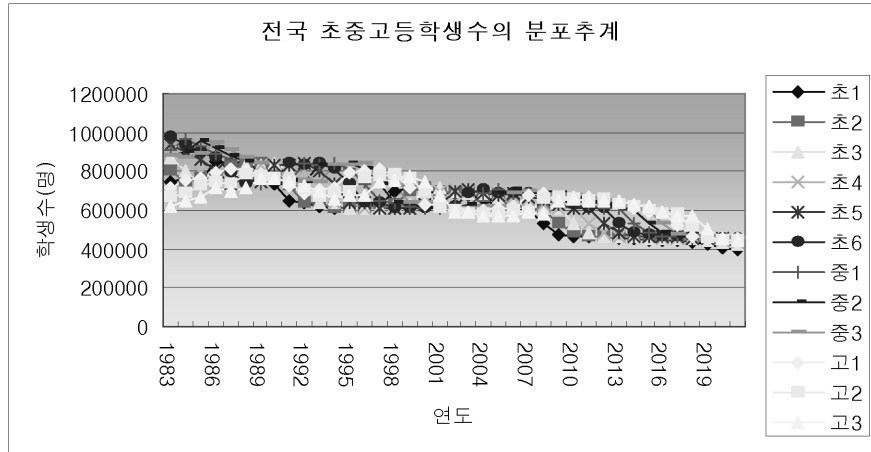
둘째, β_t 는 로지스틱 지수평활모형에 의하여 예측될 수 있다. β_t 를 로지스틱 변환하여 $u_t \left(\equiv \frac{u}{1 + \exp^{-\beta_t}} \right)$ 를 생성하고, 지수평활법을 이용하여 u_t 의 예측값을 구한 후에 이를 역변환하여 β_t 를 추정한다.

3. 교육통계 예측결과와 오류와 문제점

3.1 전국 연도별 학생수의 분포.

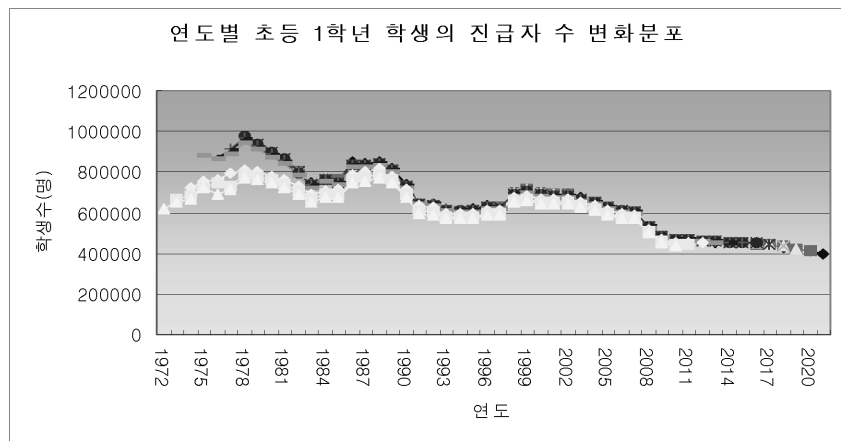
교육통계 예측결과(2007년-2021년)에 따르면 우리나라 고등학교 3학년 학생수는 2005년 569,844명을 저점으로 서서히 증가하다가 2010년 661,026명을 정점으로 감소하기 시작하여 2021년에는 440,992명으로 급격히 감소함을 보이고 있다. 아래의 그림

[3-1]은 국가교육정보센터의 예측결과를 가지고 우리나라 초등1학년에서 고등학교 3학년까지의 학생수의 변화 분포를 나타낸 것이다.



[그림3-1] 전국 초등1학년에서 고3학생수의 변화 분포

위의 [그림3-1]의 데이터에서, 2006년도 초등1학년은 2007년도에 초등2학년으로 진급하고, 2008년도엔 초등3학년으로 진급하고, 2017년에 고등학교 3학년(고3)으로 진급하게 된다. 이러한 진급자 수의 변화를 기준으로 데이터를 평행 이동하여 한해를 기준으로 재배열하면 아래의 [그림3-2]을 얻는다. [그림3-2]의 중요성은 [그림3-1]에서 복잡하게 표현되는 학생수분포를 당해 연도 초등1학년에 입학한 학생수가 고3까지 진학할 때까지 진학자 수를 당해 연도 기준으로 평행이동을 이용하여 일직선에 배치시킴으로써 당해 연도 초등1학년 학생들의 진급자수의 분포 변화를 명확하게 볼 수 있다.



[그림3-2] 연도별 초등1학년 학생의 진급자 수 변화 분포

3.2 교육통계 예측의 기초자료의 문제점

통계예측에 있어서 기초자료의 중요성은 말할 필요도 없다. 만약 기초자료가 오류가 많다면 기초자료를 기반으로 한 통계 예측은 그 신뢰성을 잃기 때문이다. 일반적으로 우리나라에서 태어난 출생아수는 초등학교 1학년 학생수보다 많아야 하고, 초등학교 1학년의 학생수는 이 학생들이 2학년, 3학년 등 상급학년으로 진학함에 따라 그 숫자가 주는 것이 일반적인 상식이다. 즉,

$$Y\text{년도 } A\text{학년 진급률} = \frac{Y\text{년도 } A\text{학년의 학생수}}{(Y-1)\text{학년도 } (A-1)\text{학년 학생수}}$$

의 진급률이 1이 넘어간다면, 기초자료 집계에서 이상점이 생긴 것으로 보는 것 옳다. 초등학교의 의무교육과정 하에서, 사회적인 심각한 변화 없이는 사망, 질병, 품행, 유학/이민 등의 이유로 학교를 떠난 학생수가 복학 및 재입학 등으로 학교에 들어온 학생수보다 일반적으로 많기 때문이다.

국가교육정보센터가 제공한 초등학교의 기초자료의 경우는 1980년, 1986년, 1992년, 1994-1997년, 1999년, 2000년, 2006년의 초등학교 학생수 집계표 조사에서 당해에 적어도 3개 학년이상 진급율이 1이 넘는 경우가 발생했다. 전국 학생수 데이터에서 진급한 학년의 학생수가 더 많아지는 이유와 원인에 대한 철저한 조사가 이루어져야 할 것이다. 이러한 오류는 초등학교 학생수의 추계 학생수 추정에 심각한 오류를 발생하게 만들었다.

3.3 교육통계 예측 결과의 두 가지 심각한 오류

아래의 <표3-1>은 국가교육정보센터 교육통계서비스시스템의 2007년부터 2010년까지 교육통계 예측결과(2007년-2021년)의 학생수 추정의 일부(초등학교 1학년 학생수 ~ 중학교 1학년 학생수)이다.

<표3-1> 교육통계서비스시스템의 초등1학년부터 중등1학년까지의
 2007년부터 2010년까지 교육통계 예측결과

| | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|----|---------|----------------|----------------|----------------|----------------|----------------|----------------|
| 초1 | 653,760 | 622,578 | 604,837 | 601,498 | 531,003 | 478,786 | 465,408 |
| 초2 | 663,512 | 652,036 | 626,676 | 607,896 | 604,614 | 533,682 | 481,227 |
| 초3 | 685,550 | 661,337 | 652,890 | 627,817 | 608,897 | 605,625 | 534,577 |
| 초4 | 684,657 | 683,179 | 665,454 | 653,626 | 628,429 | 609,402 | 606,062 |
| 초5 | 691,218 | 682,212 | 686,646 | 665,931 | 654,017 | 628,789 | 609,744 |
| 초6 | 703,702 | 688,764 | 685,413 | 687,059 | 666,167 | 654,141 | 628,878 |
| 중1 | 691,170 | 704,538 | 689,522 | 685,945 | 687,439 | 666,379 | 654,172 |

위의 <표3-1>의 진하게 표시된 숫자들을 살펴보면 교육통계 예측결과는 초등학교 학생수의 추정에 있어서 과다하게 학생수를 추정하는 오류를 범하고 있다. <표3-1>를 이용하여 당해 연도 초등1학년 학생들을 기준으로 상급학년 진급자 수로 재배열하

므로, 연도별 초등1학년 학생의 진급자 수 변화를 아래의 <표3-2>에 나타내었다.

상급학년으로 진급함에 따라 학생수는 감소하는 것이 일반적인 경우이지만, 초등1학년 학생수가 한 학년씩 진급함에 따라 그 숫자는 점점 더 늘어나서, 초등6학년 학생수가 초등1학년 학생수보다 많아지는 현상을 보이고 있다. 이러한 과다추정현상은 2004년부터 발생하여 2021년까지 이어진다. 적게는 약 4,000명 많게는 6,000명 이상이 초등1학년 학생수보다 초등6학년 학생수가 많이 추정되었다. 이는 초등학교 학생수 추계의 심각한 오류를 보여주고 중등과 고등학교 학생수의 추계를 심각하게 잘못 추정하게 된다.

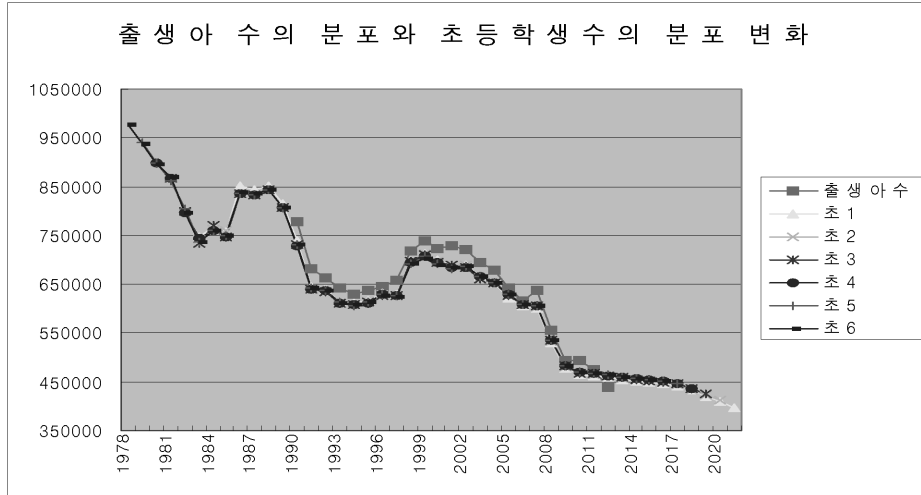
<표3-2>연도별 초등1학년 학생의 중등1학년까지의 진급자 수 변화

| | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 초1 | 653,760 | 622,578 | 604,837 | 601,498 | 531,003 | 478,786 | 465,408 |
| 초2 | 652,036 | 626,676 | 607,896 | 604,614 | 533,682 | 481,227 | 467,796 |
| 초3 | 652,890 | 627,817 | 608,897 | 605,625 | 534,577 | 482,069 | 468,653 |
| 초4 | 653,626 | 628,429 | 609,402 | 606,062 | 534,910 | 482,344 | 468,897 |
| 초5 | 654,017 | 628,789 | 609,744 | 606,418 | 535,243 | 482,683 | 469,281 |
| 초6 | 654,141 | 628,878 | 609,772 | 606,454 | 535,301 | 482,757 | 469,364 |
| 중1 | 654,172 | 628,769 | 609,498 | 606,077 | 534,889 | 482,305 | 468,862 |

두 번째의 심각한 추정오류는 결론적으로 출생아 수에 대한 잘못된 추정에서 발생한다. 당해 연도 출생아수의 분포는 6년 뒤의 초등1학년 학생수에 지대한 영향을 미친다. 그러므로 잘못된 출생아수의 추정은 잘못된 학생 추계를 예측하게 될 것이다.

아래의 [그림3-3]은 출생아수의 분포를 초등1학년 학생이 되는 6년 뒤로 평행이동시킨 후에 [그림3-2]의 분포와 비교하여 놓았다. 출생수의 분포는 전국 학생수의 분포와 매우 밀착된 분포 모양을 가진다. 2009년까지는 출생아 수의 분포와 초등학생수의 분포는 같은 형태를 이루고 있다.

그러나 2005년의 출생아 수 438,062명의 급격한 감소로 말미암아 2006년 출생아들이 초등학교 1학년이 되는 2012년에는 추정된 초등1학년 학생수(459,847명)가 자신들이 태어난 해의 출생아수보다 더 많아지는 기이한 현상을 보인다. 학생수의 추계는 인구의 추계와 마찬가지로 출생아 수에 대한 추정 계산부터 더 정확히 이루어질 필요성이 있다. (전광희 2006).



[그림3-3] 출생아수의 분포와 초등학생수의 분포변화

4. 결론

본 연구는 국가교육정보센터는 교육통계서비스시스템의 교육통계 전국 학생수 예측의 결과에 대한 문제점과 오류들을 관찰하였다 올바른 학생수의 예측을 위하여서는 출생아 수의 정확한 예측부터 다시 조명되어야 한다(전광휘, 2006). 출생아 수의 분포는 초등학생수의 분포에 절대적인 영향을 미치고 있기 때문이다. 또한 초등학교 학생수에 대한 정확한 교육인적자원부의 기초자료 통계집계가 이루어져야 한다. 상식적으로 초등학교에서 학년이 진급할수록 학생수가 질병/사망, 이민/유학 등의 이유로 학생수가 줄어드는 것이 원칙이다. 전국 초등학생수의 기초자료 통계적 집계에 보다 신중을 가질 필요가 있어 보인다. 정확한 학생수 예측 시스템의 개발은 차후의 연구로 미룬다.

참고문헌

1. 교육통계서비스시스템.(http://std.kedi.re.kr/jcgi-bin/pred_new/pred_frme.jsp)
2. 전광휘(2006). 한국의 혼인력과 출산력 추이와 전망. 대한통계협회, 제32권 1, 2호 통합본, p60-90.
3. Kim, J.T.(2004). Statistical Survey about the Rates of Application for the 2005 Susi Second Semester Admission to Universities in Daegu and Kyungbook *Journal of the Korean Data and Information Science Society*, Vol. 15, 845-853.
4. 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 and Information Science Society*, Vol. 16, 933-942.
5. 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 and Information Science Society*, Vol. 16, 969-977.
6. Kim, J.T.(2006). The Forecasting Analysis for the Numbers of High-school Graduate. *Proceedings of joint conference KDIS & KDAS*, pp.141-155

[2007년 1월 접수, 2007년 2월 채택]