

GLM을 이용한 대학학업성취도 분석 *

최국렬¹⁾ 이동석²⁾

요 약

대학의 학업성취도를 나타내는 평점평균은 많은 요인의 영향을 받는 것으로 알려져 있다. 본 논문은 현행입시제도하에서 획득 가능한 요인과 평점평균과의 관련성을 일반화선형모형(GLM)을 이용하여 통계적으로 분석·평가하고자 한다. 여기서 얻어진 결과는 요즈음 변화가 모색되고 있는 입시제도나 대학교육제도의 개선에 적으나마 도움이 되는 기초자료를 제공할 수 있으리라 믿어진다. 분석자료는 1996, 97학년도 인제대학교에 입학한 학생들의 입시자료와 96, 97학년도의 평점평균, 기숙사입사 여부 등을 대상으로 삼았다.

주요용어: 학업성취도, GLM, 수학능력시험.

1. 서론

대학에서 학생 개인의 학업성취도를 나타내는 평점평균(Grade Point Average, GPA)은 많은 요인의 영향을 받는 것으로 알려져 있다. 대표적인 요인으로는 대학의 학생선발 지표인 입학성적과, 사회인구학적 특성을 나타내는 성별, 출신지역 등을 들 수 있다. 그 외에도 학생들의 생활양식이나 입시전형 방법, 기숙사 입사여부 등도 영향을 미칠 것으로 예상된다. 이러한 요인들은 대학의 입시관계자들이나 구성원 모두에게 많은 관심의 대상이 되어 왔다. 특히 무시험 입학전형을 비롯하여 혁명적이라 할 만한 대학 입시제도의 개혁과 학부제 등 교육제도의 변화가 시도되고 있는 요즈음 이러한 요인들에 대한 분석은 필요하리라 여겨진다. 더구나 해방이후 50여년간 우리 나라의 대학입시제도는 너무 잦은 변경으로 교육관계자나 수험생들에게 적지 않은 혼선을 빚어왔으며, 이러한 변경이 충분한 과학적 검증 없이 이루어짐으로써 적지 않은 시행착오와 바람직한 교육활동이 저해되는 경우가 많았다. 그 결과 교육관계자는 물론 국민도 교육정책을 불신하는 상황까지 이르렀다. 대학의 입시제도와 교육제도는 고등학교 교과과정은 물론 국민적 교육방향에 심대한 영향을 미친다는 점을 고려할 때 대학교육 개혁은 충분한 과학적 데이터를 축적하면서 이를 바탕으로 예측되는 부작용을 최소한으로 줄이는 방향에서 서서히 진행되는 것이 바람직하다.

본 연구는 이런 관점에서 대학의 학업성취도를 나타내는 평점평균(GPA)과 이에 영향을 미칠 수 있는 요인변수를 통계적으로 분석하여 평점평균(GPA)과 어느 정도 연관관계를 갖고 있는지 평가하고자 한다. 여기에 대한 선행 연구로는 “허명희 외1(1996)”의 고려

* 본 논문은 1998학년도 인제대학교 학술연구조성비 보조에 의한 것임.

1) (621-749), 경상남도 김해시 어방동 607, 인제대학교 데이터정보학과 부교수

E-mail: choi@stat.inje.ac.kr

2) (621-749), 경상남도 김해시 어방동 607, 인제대학교 교무처장

대학교에 대한 결과가 있으나 본 논문에서는 수학능력시험 성적을 영역별로 세분화하였으며, 학생들의 생활 스타일에 영향을 미칠 수 있는 기숙사 입사 여부 등을 고려하였다. 또한 인재대학교는 중위권 대학이라는 점에서 이와 유사한 수준의 대학들에 대한 특성을 파악하는데 도움이 될 수 있으리라 믿어진다.

본 연구의 목적을 위하여 96, 97학년도 인재대학교에 입학한 학생들을 대상으로 삼았으며 96학년도 입학생은 특차모집과 정시모집 전부를 분석에 포함시켰으나 97학년도 입학생의 경우는 특차모집에서 고교내신성적을 반영하지 않았기 때문에 96학년도 입학생들과 동일한 기준으로 비교하기 위하여 정시모집만 분석 대상으로 삼았다. 또한 사회체육학과와 체육 특기생은 신입생 선발방법이나 GPA 평가방법이 일반학생들과 다르기 때문에 일반적 특성을 왜곡시킬 가능성이 있으므로 분석대상에서 배제하였다.

2. 분석모형 및 방법

본 연구의 목적을 위하여 사용한 변수는 96, 97학년도 입학생에 대한 각각 첫4개 학기, 첫2개 학기의 대학 평점평균(GPA)과 여기에 영향을 미칠 수 있는 변수로 대학의 학생선발 지표인 수학능력시험의 각 분야별 성적(외국어, 언어, 수리탐구1, 수리탐구2)과 고교내신 성적, 사회인구학적 특성으로 성별, 고교출신지역, 전형별(특차모집, 정시모집), 기숙사입사 여부 그리고 전공학과를 고려하였다. 여기서 고교내신등급은 실질반영 점수로 환산하여 사용하였으며 고교출신지역은 인재대학교의 지역적 특성을 고려하여 부산광역시, 경남, 기타지역으로 나누어 분석하였다. 부산광역시는 아직 고교평준화 지역으로 남아 있지만 경남은 대부분 고교입사가 부활되어 비평준화 지역으로 분류되는 특성을 갖고 있다. 또한 인재대학교는 전교생의 약1/3이 기숙사 생활을 하고 있으며 기숙사입사 여부가 통학거리, 기숙사 고유의 면학분위기 등 생활 스타일에 많은 영향을 미칠 것으로 예상되어 이를 분석에 포함시켰다. 분석방법으로는 각 학년도별로 입학성적이나 학업성적의 기준이 다르기 때문에 입학 년도별로 나누어 통계분석을 시도하였으며 각 학년별 매 학기 평점평균(GPA)을 반응 변수로 하고 나머지 변수를 설명변수로 하였다. 입학성적의 실질 반영점수는 96학년도 입학생의 경우 외국어 영역:40점, 수리탐구1 영역:40점, 수리탐구2 영역:60점, 언어 영역:60점, 고교내신성적:200점 만점 기준이었으며, 97학년도 입학생의 경우 외국어 영역:80점, 수리탐구1 영역:80점, 수리탐구2 영역:120점, 언어 영역:120점, 고교내신성적:290점 만점 기준이었다.

통계적 분석모형으로는 설명변수가 정성적 변수(성별, 고교출신지역, 입시전형별, 기숙사 입사여부, 전공학과)와 정량적 변수(각 영역별 수학능력시험성적, 고교내신성적)를 동시에 포함하고 있기 때문에 아래와 같은 일반선형모형을 이용한 공분산분석을 시도하였다.

$$\begin{aligned} \text{GPA} = & \text{정성적 변수(성별, 입학전형, 전공학과, 기숙사입사여부, 고교출신지역)} \\ & + \text{정량적 변수(언어, 외국어, 수리탐구1,2, 고교내신성적)} \\ & + \text{오차항} \end{aligned}$$

위 자료에 대한 선형회귀모형식은 다음과 같이 표현할 수 있다.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i, i = 1, \cdots, n \quad (2.1)$$

표 2.1: 입학성적에 대한 상관분석결과

	96학년도				97학년도			
	외국어	언어	수리탐구1	수리탐구2	외국어	언어	수리탐구1	수리탐구2
언어	0.2499 (0.000)				0.2104 (0.000)			
수리탐구1	0.1175 (0.000)	0.0916 (0.000)			0.2191 (0.000)	0.0337 (0.277)		
수리탐구2	0.2309 (0.000)	0.2749 (0.000)	0.2763 (0.000)		0.1923 (0.000)	0.2619 (0.000)	0.2453 (0.000)	
고교내신	0.1016 (0.000)	0.0673 (0.006)	0.1039 (0.000)	0.0890 (0.000)	0.3535 (0.000)	0.1111 (0.000)	0.2952 (0.000)	0.3426 (0.000)

()안은 p-값

위 모형에서 Y 는 반응변수의 관측값을 나타내며 X_k ($k = 1, \dots, p$)는 설명변수를 나타낸다. β_k ($k = 0, \dots, p$)는 회귀계수로 불리며 ϵ 은 오차항으로 $N(0, \sigma^2)$ 인 정규분포를 따르는 것으로 가정하였다. 위 2.1식 모형에 대한 적합은 최소자승법을 사용하였다. 이때에 익명의 한 심사위원이 지적한대로 정량적 설명변수간에 다중공선성(Multicollinearity)이 존재하면 심각한 문제를 야기할 수 있다. 이를 확인하기 위하여 정량적 설명변수에 대한 상관분석을 실시하였다. 표2.1의 결과를 보면 알 수 있듯이 분석결과 관측치의 수가 큰 관계로 p-값은 대부분 유의하게 나왔지만 다중공선성을 의심할 정도로 상관계수는 높지 않았다. 여기에 실지는 않았지만 다중공선성 검토에 대한 분산팽창요인(variance inflation factor)의 값도 모두 1.5 이내로 위 모형을 사용하는데 다중공선성은 문제되지 않았다.

분석 프로그램으로는 SAS의 GLM procedure를 이용하여 선형회귀계수를 추정하여 설명변수의 실질 변화에 따른 반응변수의 변화추이를 검토하였다. 이때에 정량적 변수에 대한 회귀계수 추정치는 설명변수가 한 단위 증가함에 따라 반응변수가 어느 정도 영향을 받는지를 평가하는데 중요한 역할을 하지만 정성적 변수에 대한 회귀계수 추정치는 그룹내 비교를 위해서는 의미가 있지만 반응변수에 대한 기여도를 평가하는 데는 부적절하다. 또한 정량적 변수에 대한 회귀계수 추정치도 각 변수의 단위가 다르기 때문에 상대적 기여도를 평가하는 데는 부족하다. 따라서 각 설명변수의 상대적 중요도를 알아보기 위하여 표준화 회귀계수를 이용하여야 한다. 표준화 회귀계수는 선형회귀모형 2.1식에서 변수변환

$$Y_i' = \frac{Y_i - \bar{Y}}{\sqrt{\sum(Y_i - \bar{Y})^2}}, \quad X_i' = \frac{X_i - \bar{X}}{\sqrt{\sum(X_i - \bar{X})^2}} \quad (2.2)$$

을 이용하여 아래와 같은 표준화 선형회귀식을 구할 수 있다.

$$Y_i' = \beta_1' X_{i1}' + \dots + \beta_p' X_{ip}' + \epsilon_i' \quad (2.3)$$

위 모형에서 절편항(intercept term)이 없는 이유는 절편항에 대한 회귀계수 추정치가 항상 0이 되기 때문이다. 2.2식의 변환을 상관변환(correlation transformation)이라고도 하는데 "Neter 외3(1996)"이 지적한 것처럼 회귀계수 추정치의 행렬이 설명변수들의 상관행렬로

구성되어 중간계산 과정에서 반올림 오차에 의한 영향을 적게 받는 장점을 갖고 있다. 또한 2.3식에서 얻어진 표준화 회귀계수는 2.1식의 회귀계수 추정치와 다음 관계를 갖고 있다.

$$\beta_k = \frac{s_y}{s_k} \beta_k' \quad (k = 1, 2, \dots, p) \quad (2.4)$$

$$\text{단 } s_y = \sqrt{\frac{\Sigma(Y_i - \bar{Y})^2}{n-1}}, s_k = \sqrt{\frac{\Sigma(X_{ik} - \bar{X}_k)^2}{n-1}}$$

이러한 표준화 회귀계수는 대부분의 통계패키지에서 옵션을 지정하여 쉽게 구해지며 몇몇 통계패키지(Statistical Package)에서는 베타계수로도 불리어 진다. 표준화 회귀계수(베타계수)는 “SAS(Strategic Application System; SAS/STAT User’s Guide 1991)”의 GLM procedure를 사용하는 경우 Type III SS를 이용하면 쉽게 구할 수 있으며, 이는 정성적 변수에 대한 상대적 기여도를 평가하는데도 이용될 수 있다.

3. 분석결과

1996학년도 입학생들에 대한 통계분석 결과는 표3.1에 나타나 있으며, 1997학년도 입학생들에 대한 통계분석 결과는 표3.2에 정리되어 있다.

먼저 96학년도 입학생들에 대하여 살펴보면 GPA에 상대적으로 가장 큰 영향을 미치는 변수는 성별임을 알 수 있다. 성별에 대한 회귀계수 추정치가 제1학기 0.4313, 제2학기 0.5244, 제3학기 0.4605, 제4학기 0.4139로 여학생들이 남학생들보다 줄 곧 우수한 성적을 얻은 것으로 나타났으며 상대적 기여도를 의미하는 베타계수도 제4학기모두 매우 높은 것으로 나타났다. 이는 “혁명회 외1(1996)”의 고려대학교 경우와 비슷한 경향을 보여주고 있지만 인제대학교의 경우가 다소 높게 나타났다. 이러한 분석결과는 여학생들의 학습능력이 남학생들 보다 뛰어났다기 보다는 생활스타일이나 학점관리에 임하는 자세의 차이라고 보는 것이 타당할 것 같다. 실제 여기서는 보이지 않았지만 남녀간의 입학성적을 비교한 결과 수학능력시험, 고교내신성적 모두 평균점수의 차가 0.2점 이하로 나타나 남녀간의 유의한 차이가 인정되지 않았던 점을 고려하면 매우 의미 있는 결과로 해석된다. 다음으로 학생선발 지표인 입학성적을 살펴보면 고교내신성적이 수학능력시험성적 보다 중요한 영향을 미치고 있는 것으로 볼 수 있다. 고교내신성적은 96학년도 입학생의 경우 제1학년에 해당하는 첫1,2학기는 비교적 큰 영향을 주다가 전공기초과정이 시작되는 제3,4학기에는 차츰 그 영향이 줄어들고 있음을 알 수 있다. 또한 수학능력시험성적도 학기에 따른 변화추이는 고교내신성적과 유사한 경향을 보이고 있지만 외국어 영역이 제1,2학기에 미치는 영향을 제외하고는 GPA에 큰 영향을 미치지 못하고 있다. 특히 외국어나 언어영역에 비하여 수리탐구 영역은 전 학기에 걸쳐서 베타계수가 0에 가깝다. 이는 선 시험 후 지원이라는 현행입시체제하에서 동일학과에 입학하는 학생들의 수학능력시험성적이 좁은 구간 안에 밀집되어 있는 영향도 있겠지만 최소점과 최대점의 차이가 대부분의 학과에서 20점 이상이기 때문에 고교내신성적에 비하여 실질 반영점수의 차가 크다는 점을 고려하면 인제대학교에 입학하는 학생들의 수준에서는 수학능력시험성적의 변별력이 그리 높지 않음을 시사하

표 3.1: 96학년도 입학생들의 GLM 분석결과

반응 변수		제1학기 GPA				제2학기 GPA			
		회귀계수 (b)	SE	베타 (β)	p-값	회귀계수 (b)	SE	베타 (β)	p-값
설명변수									
수 능 성 적	외국어	0.0321	0.0040	0.0385	0.0001	0.0251	0.0044	0.0206	0.0001
	언어	-0.0139	0.0038	-0.0081	0.0003	-0.0131	0.0042	-0.0061	0.0021
	수리탐구1	0.0000	0.0043	0.0000	0.9999	0.0045	0.0047	0.0006	0.3432
	수리탐구2	0.0073	0.0040	0.0020	0.0713	0.0095	0.0044	0.0030	0.0323
고교내신		0.0264	0.0030	0.0460	0.0001	0.0284	0.0033	0.0452	0.0001
성별(여)		0.4313	0.0389	0.0712	0.0001	0.5244	0.0426	0.0888	0.0001
입학전형(특)		0.1002	0.0338	0.0054	0.0030	0.0956	0.0372	0.0043	0.0102
고교출신지역				0.0072	0.0002			0.0072	0.0102
전공학과				0.0457	0.0001			0.0509	0.0001
결정계수		0.2895(n=1637)				0.2689(n=1569)			

반응 변수		제3학기 GPA				제4학기 GPA			
		회귀계수 (b)	SE	베타 (β)	p-값	회귀계수 (b)	SE	베타 (β)	p-값
설명변수									
수 능 성 적	외국어	0.0081	0.0046	0.0026	0.0810	0.0066	0.0046	0.0020	0.1518
	언어	-0.0054	0.0044	-0.0013	0.2241	-0.0118	0.0045	-0.0071	0.0072
	수리탐구1	-0.0064	0.0048	-0.0016	0.1779	-0.0043	0.0048	-0.0008	0.3658
	수리탐구2	0.0022	0.0046	0.0002	0.6213	0.0001	0.0046	0.0000	0.9847
고교내신		0.0188	0.0034	0.0255	0.0001	0.0134	0.0034	0.0154	0.0001
성별(여)		0.4605	0.0460	0.0804	0.0001	0.4139	0.0489	0.0649	0.0001
입학전형(특)		0.0210	0.0387	0.0003	0.5881	0.0173	0.0386	0.0004	0.5126
기숙사입사		0.4479	0.0491	0.0667	0.0001	0.3715	0.0440	0.0660	0.0001
고교출신지역				0.0060	0.0299			0.0074	0.0235
전공학과				0.0667	0.0001			0.0355	0.0041
결정계수		0.3216(n=1195)				0.2372(n=1046)			

표 3.2: 97학년도 입학생들의 GLM 분석결과

반응 변수		제1학기 GPA				제2학기 GPA			
		회귀계수 (b)	SE	베타 (β)	p-값	회귀계수 (b)	SE	베타 (β)	p-값
설명변수	수								
	능력								
	성								
	적								
	외국어	0.0078	0.0028	0.0078	0.0048	0.0120	0.0028	0.0182	0.0001
	언어	-0.0082	0.0026	-0.0099	0.0015	-0.0130	0.0026	-0.0254	0.0001
	수리탐구1	0.0011	0.0029	0.0001	0.6991	0.0054	0.0030	0.0032	0.0764
	수리탐구2	0.0025	0.0028	0.0008	0.3784	0.0021	0.0029	0.0005	0.4705
	고교내신	0.1133	0.0116	0.0858	0.0001	0.1097	0.0121	0.0779	0.0001
	성별(여)	0.4135	0.0516	0.0593	0.0001	0.3361	0.0536	0.0388	0.0001
	기숙사입사	0.2361	0.0472	0.0281	0.0001	0.4083	0.0493	0.0660	0.0001
	고교출신지역			0.0084	0.0033			0.0005	0.5000
	전공학과			0.0720	0.0001			0.0623	0.0001
	결정계수	0.2884(n=1047)				0.3420(n=1000)			

는 듯하다. 다음으로 전공학과가 상당한 영향을 갖고 있는 것으로 보이는데 이는 전공학과에 따라 학점을 평가하는 기준이나 면학분위기에 어느 정도 차이가 있는 것으로 해석할 수 있다. 마지막으로 기숙사 입사여부가 GPA에 상당한 영향을 미치는 것으로 나타났다. 96학년도 입학생의 경우 제1,2학기는 기숙사 입사자료를 입수할 수 없었으나 제3,4학기는 남녀간의 차이 다음으로 베타계수가 높았다. 기숙사생들은 어느 정도 통제된 생활 속에서 다양한 교육프로그램의 제공으로 면학분위기 조성은 물론 통학시간의 단축 등으로 많은 이점을 얻고 있기 때문으로 풀이된다. 물론 위 결과를 해석하는 과정에서 휴학생으로 인한 자료의 수가 제1,2,3,4 학기 각각 $n=1637, 1569, 1195, 1049$ 로 줄어들고 있음을 유념하여야 할 것이다. 특히 남학생들은 1학년을 마치고 군 입대를 위한 휴학을 많이 하고 있는 경향이 있다.

다음으로 97학년도 입학생들에 대한 표3.2를 살펴보면 성별에 대한 회귀계수가 제1,2학기 각각 0.4135, 0.3361로 높게 나타나 여학생들이 남학생들 보다 높은 학업성취도를 보여주고 있다. 이는 96학년도 입학생들과도 유사한 경향을 보이는 것으로 성별에 대한 흥미 있는 특성으로 해석하여도 큰 무리는 없을 것으로 여겨진다. 한편 상대적 기여도를 나타내는 베타계수는 고교내신성적이 제1,2학기 모두 가장 높게 나타나 GPA에 가장 큰 영향을 미치고 있음을 알 수 있다. 이는 96학년도 제1,2학기의 분석결과와도 어느 정도 일치하여 입학년도에 상관없이 고교내신성적의 기여도가 수학능력시험성적 보다 높음을 알 수 있다. 그 이유는 수학능력시험이 1회성 평가인데 반하여 고교내신성적은 고등학교 전과정을 반영하고 있어 예측타당도가 높을 뿐만 아니라 학생 개인의 성실도가 어느 정도 반영되어 있는 결과로 해석할 수 있다. 특히 수학능력시험의 수리탐구 영역은 외국어나 언어영역에 비하여 96학년도와 마찬가지로 영향력이 극히 미미한 것으로 나타났다. 다음으로 전공학과와 성

별이 큰 영향을 미치고 있는 것으로 나타났으며 나머지 변수도 96학년도와 유사한 경향을 보여주고 있다.

4. 결론

대학의 학생선발 지표인 입학성적이 GPA에 큰 영향을 미칠 것으로 예상되었으나 실제 통계분석결과 수학능력시험성적은 고교내신성적에 비하여 예측타당도(성태제, 1994 참조)가 매우 낮은 것으로 나타났다. 이는 “허명희 외1(1996)”의 고려대학교에 대한 연구결과와 어느 정도 일치하는 경향을 보여주고 있다. 특히 인제대학교에 입학하는 학생들의 경우 수리탐구 영역이 GPA에 미치는 영향력은 극히 미미한 것으로 보인다. 물론 분석대상을 수리탐구 영역에 관련된 교과목을 수강하는 학과와 그렇지 않은 학과로 구분하여 분석한다면 좀 더 자세한 결과를 얻을 수 있을지 모르겠지만 자연계와 인문계로 나누었을 때는 거의 차이가 없었다. 전반적으로 대학의 학업성취도를 나타내는 GPA는 학기가 지남에 따라 입학성적 보다는 외적인 요인 성별, 기숙사입사 여부 등의 영향을 많이 받는 것으로 풀이된다. 좀더 정확한 분석결과나 전공학과에 따른 변화 추이를 얻기 위해서는 학기가 바뀔 때마다 휴학생들로 인하여 발생하는 결측치 문제가 우선적으로 해결되어야 할 것 같다. 그 외에도 추가적인 적절한 설명변수를 찾아 설명력(결정계수)을 높이는 문제도 중요하리라 여겨진다. 이러한 문제는 앞으로의 연구과제가 될 수 있을 것이다.

참고문헌

- [1] 성태제(1994). 대학별고사를 위한 문항분석, 표준점수, 검사동등화. 〈한국통계학회논문집〉.제1권1호 206-214.
- [2] 허명희, 전성연(1996). 입학선발지표와 대학학업성취도의 관련성. 〈교육학 연구〉. 34권 5호, 299-309.
- [3] Neter J., Kuter M. H., Nachtsheim J. and Wasserman (1996). *Applied Linear Regression Models*. Irwin.
- [4] SAS/STAT User's Guide (1991). *SAS Institute Inc.*

[1998년 12월 접수, 2000년 1월 채택]

The Impact of Admission Indices and Social-Demographic Features on Grade Point Average *

Kook-Lyeol Choi ¹⁾ Dong Seok Lee ²⁾

ABSTRACT

The impact of two admission indices, HSG(High School Grade) and SAT(Scholastic Aptitude Test), and social-demographic features on the GPA(Grade Point Average) , was evaluated by using an GLM(General Linear Model) model with the GPA from the year 1996 and 1997 at the Inje university. The results was obtained from the students obtained admission in 1996 and 1997.

Keywords: GLM; GPA; SAT.

* This work was supported by grant from Inje University, 1998.

1) Dept. of Data Science, Inje University. E-rmail: choi@stat.inje.ac.kr

2) Dean of Academic Affairs, Inje University.