

## 대학 평가지표들에 대한 상관분석과 변수선택에 의한 선형모형추정<sup>†</sup>

송필준<sup>1</sup> · 김중태<sup>2</sup>

<sup>1,2</sup>대구대학교 전산통계학과

접수 2012년 1월 16일, 수정 2012년 3월 7일, 게재확정 2012년 4월 19일

### 요약

본 연구의 목적은 한국대학교육협의회 대학정보공시센터의 '대학알리미'에서 주요 대학지표들을 분석하고, 지표들 간의 연관성과 통계적 모형을 추정하는데 있다. 먼저 상관관계수에 대한 통계적 검정을 이용하여 변수들 간의 통계적으로 유의한 상관성을 추정하고, 이들 주요 지표들의 모형을 추정하기 위해서 회귀분석 방법의 변수선택 방법을 이용하여 회귀 방정식을 추정하여 변수들 간의 연관성을 조사하였다. 변수선택의 판정기준에 따른 방법으로 전진선택법과 후진제거법, 단계별 회귀방법을 사용하였다.

주요용어: 대학평가지표, 변수선택, 상관분석, 회귀분석.

### 1. 서론

본 연구의 목적은 한국대학교육협의회 대학정보공시센터의 '대학알리미'에 있는 대학평가지표들을 분석하고, 지표들 간의 연관성과 통계적 모형을 추정하는데 있다.

대학평가지표들에 대한 연관성이 있는지를 분석하기 위하여 회귀분석의 변수선택에 의한 방법을 사용하여 반응변수에 영향을 미칠 수 있는 적절한 설명변수들을 선택하여 최적화된 회귀방정식 (best regression equation)을 추정하여 대학지표들 간에 모형을 추정한다. 변수선택의 판정기준에 따른 방법으로 전진선택법과 후진제거법, 단계별 회귀방법을 사용하였다.

대학 평가지표들의 모형을 추정하기 위한 방법 중 하나는 선형회귀방정식에 포함될 변수들을 일단 결정된 후 그 변수에 대한 적절한 다중회귀방정식을 추정하는 것이다. 변수선택에 의한 회귀방정식을 추정하기 전에 먼저 각 변수들 간의 피어슨 상관관계 계수 (correlation coefficient)에 대한 통계적 가설검정을 이용하여 변수들 간의 유의한 상관관계를 먼저 조사하는 상관분석을 시행한다. 두 번째는 다중선형회귀분석의 변수선택방법들을 이용하여 각 반응변수들에 대한 회귀방정식을 추정하여 예측변수들과 반응변수 간의 연관성을 조사하는 방법이다.

본 연구에서는 대학구조정의 영향을 가장 심각하게 받게 될 대구직할시와 경상북도의 4년제 24개 일반대학들과 23개 전문대학들에 관한 대학평가지표들을 가지고 분석한다. 대구·경북의 1개의 교육대학교와 1개의 산업대학에 대한 데이터는 대학의 성격이 다른 관계로 제외시켰다. 데이터의 수가

<sup>†</sup> 이 논문은 2010년도 대구대학교 학술연구비 지원에 의한 연구임.

<sup>1</sup> (712-714) 경상북도 경산시 진량면 내리동15, 대구대학교 전산통계학과, 교수.

<sup>2</sup> 교신저자: (712-714) 경상북도 경산시 진량면 내리동15, 대구대학교 전산통계학과 교수,  
E-mail: jtkim@daegu.ac.kr.

많지 않는 관계로 경북대와 안동대를 포함한 국립대학들을 사립대학과 함께 포함시켰고, 편제정원과 지역 등에 의한 영향에 대한 분석은 전국 단위의 대학 평가지표 분석에서 다루고자 한다. 또한 포항공대와 가야대학의 경우는 결측값을 가지는 정보가 있으므로 제외시켰다. 송필준과 김종태 (2010)와 윤용화와 김종태 (2010)는 초 중 고등 학생수 추계를 가지고 대학진학에 미치는 영향을 연구하였다.

본 연구에서의 상관분석과 변수선택에 의한 상관관계 분석의 결과는 조금은 다른 차이가 나타남을 보였다. 그리고 일반대학교와 전문대학들 사이에서도 주요지표 간에 상당히 다른 차이를 보였다.

Chatterjee 등 (2000)과 Christensen (1981), 박성현 (1998)에서 다중회귀모형에서 변수선택법 중에서 전진선택과 후진제거의 과정을 잘 설명하고 있다. 홍종선과 김명진 (2010)은 다중회귀모형에서 변수선택법 중에서 전진선택과 후진제거의 과정을 기하학적으로 표현하는 그래픽적 방법을 제안하였다. 최경호와 한동욱 (2010)은 대학입학 관련 주제별로 입학전형요소에 대하여 갖는 가중치 정도에 따른 계층적분석을 하였다.

2절에서는 대학 알리미에 제시된 많은 대학 평가 지표들 중 주요 지표들에 대한 상관관계수에 대한 통계적 가설검정을 하여 그 연관성을 조사하였고, 3절에서는 변수선택 방법을 이용하여 주요지표들에 대한 회귀식을 추정하였다. 변수선택의 방법에 따라 각 지표들에 대한 연관성을 알 수 있다. 4절에서는 결론으로 현재 대학에서 가장 크게 부각되고 있는 재학생 충원율과 건강보험을 가지는 취업률과 주요지표들과의 연관성을 분석한다.

## 2. 주요지표들의 상관분석

교육과학기술부 (2011) 구조개혁위원회는 ‘대학 구조개혁 추진 기본계획’을 의결하였고 오현숙 (2011)은 대학구조개혁위원회 제5차 오픈 포럼에서 고등교육 구조개혁의 방향과 과제를 발표했다. 대학구조개혁위원회의 계획에 따르면 전국 350개 대학 (전문대 146곳을 포함) 중 수도권과 그 밖의 지방에서 각각 대학평가지표가 부실한 하위 15%의 대학들에게 학자금 대출제한 및 이듬해부터 국가와 지방자치단체 재정지원 사업을 제한할 뿐 아니라 구조조정 대상 대학으로 선정이 될 수 있다는 정책을 구상하였다. 이 정책에 따르면, 수도권과 지방을 합친 350개 대학 가운데 하위 10%를 선정하고, 나머지 5%는 수도권 (119개) 대학과 지방 (231개) 대학을 구분하여 각각 하위 5%에 속한 대학들을 선정한다.

현재 대학 알리미에 제시된 많은 대학 평가 지표들 중에서 주요지표들을 정리 요약하면 (1) 신입생 최종등록률, (2) 신입생 경쟁률, (3) 재학생 충원률, (4) 신입생 일반계 고등학교 비율, (5) 신입생 전문계 고등학교 비율, (6) 상급학교 진학률, (7) 건강보험 DB연계 취업률, (8) 편제정원대비 전임교수 1인당 학생수, (9) 중도탈락 학생 비율, (10) 학생 1인당 교육비, (11) 재학생 1인당 장학금, (12) 편제정원 기준 교원 확보율과 같은 지표들 등을 구할 수 있다.

이 지표들에 대한 상관관계 검정을 이용하여 주요지표들 간의 상관관계를 조사하기 위하여, 모상관계수  $\rho$ 와 모상관계수의 점추정값 표본상관계수  $r$ 을 각각

$$\rho = \frac{Cov(x, y)}{\sqrt{var(x)}\sqrt{var(y)}}, \quad r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

으로 정의하자. 여기서, 모상관계수  $\rho = 0$ 일 때,  $t_0 = r\sqrt{(n-2)/(1-r^2)}$ 는 자유도가  $n-2$ 인  $t$ 분포를 따른다. 그러므로 상관관계수에 대한 통계적 가설검정은 다음과 같다.

$$H_0 : \rho = 0, \quad H_1 : \rho \neq 0.$$

대구 경북의 대학들의 지표들에 대한 상관관계 계수에 대한 검정 결과는 다음과 같다.

표 2.1 일반대학교 주요지표들의 상관분석표

일반대	최종 등록률	신입 경쟁률	재학생 충원율	일반계고 비율	전문계고 비율	진학률	취업률	1인당 학생수	중도 탈락생율	교육비	장학금	교원 확보율
최종등록률	1	.652**	.499**	.576**	-.498*	.136	.311	-.339	-.576**	.129	-.046	.058
신입경쟁률		1	.581**	.418	-.584**	.223	.498*	-.245	-.629**	.266	-.005	.362
재학생 충원율			1	.221	-.203	.263	.043	-.155	-.566**	.194	-.185	.591**
일반계고비율				1	-.612**	-.267	.356	-.170	-.486*	-.250	-.387	0.278
전문계고비율					1	-.431*	-.417	.425*	.808**	-.381	-.204	-.276
진학률						1	.179	-.373	-.508*	.898**	.747**	.682**
취업률							1	-.146	-.371	.385	.317	-.030
1인당학생수								1	.534**	-.270	-.290	-.289
중도탈락생율									1	-.375	-.102	-.463*
교육비										1	.861**	.623**
장학금											1	.275
교원확보율												1

표 2.1의 괄호 안의 숫자는 피어슨 상관계수를 나타내고, \*\*는 0.01 (양쪽) 수준에서 유의하고, \*는 0.05 (양쪽) 수준에서 유의함을 나타낸다.

4년제 일반대학교의 지표에서 신입생 최종등록률은 신입생 경쟁률 (.652\*\*), 재학생 충원율 (.499\*), 신입생 일반계 고등학교 비율 (.576\*)과는 통계적으로 유의한 양의 상관관계를 가지고, 신입생 전문계 고등학교 비율 (-.498\*)과 중도탈락 학생 비율 (-.576\*)에 대해서는 통계적으로 유의한 음의 상관관계를 가진다. 결론적으로 신입생 최종 등록률은 신입생 경쟁률, 재학생 충원율과 일반계 고등학교 학생들의 지원 비율에 대하여 양의 상관관계를 가짐으로서 신입생 최종 등록률과 이들 지표들에 대한 인과관계를 논리적으로 조사해 볼 필요가 있다. 또한 전문계 고등학교 학생의 지원 비율과 재학생의 중도탈락 비율은 신입생 최종 등록률과 음의 상관관계를 가짐으로서 이점에 대하여 경험적이고 논리적인 접근이 필요하리라 본다.

전문대학의 지표에서 신입생 최종등록률은 재학생 충원율 (.760\*\*)과 편제정원 기준 교원 확보율 (.578\*)과는 통계적으로 유의한 양의 상관관계를 가지고, 편제정원대비 전임교수 1인당 학생 수 (-.597\*)에 대해서는 통계적으로 유의한 음의 상관관계를 가진다.

4년제 일반대학교의 지표에서 재학생 충원율은 신입생 최종등록률 (.499\*), 신입생 경쟁률 (.581\*), 편제정원 기준 교원 확보율 (.591\*)과는 통계적으로 유의한 양의 상관관계를 가지고, 중도탈락 학생 비율 (-.566\*)에 대해서는 통계적으로 유의한 음의 상관관계를 가진다.

전문대학의 지표에서 재학생 충원율은 신입생 최종등록률 (.760\*\*), 신입생 경쟁률 (.440\*), 건강보험 DB연계 취업률 (.592\*)과 편제정원 기준 교원 확보율 (.446\*)과는 통계적으로 유의한 양의 상관관계를 가지고, 편제정원대비 전임교수 1인당 학생 수 (-.480\*)에 대해서는 통계적으로 유의한 음의 상관관계를 가진다.

4년제 일반대학교의 지표에서 건강보험 DB연계 취업률은 오직 신입생 경쟁률 (.498\*)과 통계적으로 유의한 양의 상관관계를 가진다.

전문대학의 지표에서 건강보험 DB연계 취업률은 재학생 충원율 (.592\*)과는 통계적으로 유의한 양의 상관관계를 가지고, 상급학교 진학률 (-.427\*), 편제정원대비 전임교수 1인당 학생 수 (-.436\*)와 중도탈락 학생 비율 (-.490\*)에 대해서는 통계적으로 유의한 음의 상관관계를 가진다.

4년제 일반대학교의 지표에서 중도탈락 학생 비율은 신입생 전문계 고등학교 비율 (.808\*\*)과 편제정원대비 전임교수 1인당 학생수 (.534\*)에 통계적으로 유의한 양의 상관관계를 가지고, 신입생 최종등록률 (-.576\*), 신입생 경쟁률 (-.629\*), 재학생 충원율 (-.566\*), 신입생 일반계 고등학교 비율 (-.486\*), 상급학교 진학률 (-.508\*), 편제정원 기준 교원 확보율 (-.463\*)과는 통계적으로 유의한 음의

표 2.2 전문대학 주요지표들의 상관분석표

전문대	최종 등록률	신입 경쟁률	재학생 충원율	일반계고 비율	전문계고 비율	진학률	취업률	1인당 학생수	중도 탈락생율	교육비	장학금	교원 확보율
최종등록률	1	.135	.760**	-.032	.286	.205	.191	-.597*	-.006	.156	-.157	.587**
신입경쟁률		1	.440*	.261	-.316	-.042	.249	.100	-.185	-.162	-.434*	-.161
재학생 충원율			1	.084	.157	-.066	.592**	-.480*	-.186	-.044	-.203	.446*
일반계고비율				1	-.688**	-.122	.364	-.215	-.489*	.099	-.235	.163
전문계고비율					1	.111	-.010	-.194	.337	.179	.082	.220
진학률						1	-.427*	.096	.371	.105	-.223	.030
취업률							1	-.436*	-.490*	.283	-.024	.398
1인당학생수								1	.198	-.321	.091	.939**
중도탈락생율									1	-.160	-.192	-.085
교육비										1	.164	.327
장학금											1	-.014
교원확보율												1

상관관계를 가지고 있다.

전문대학의 지표에서 중도탈락 학생 비율은 건강보험 DB연계 취업률 (-.490\*)과는 통계적으로 유의한 음의 상관관계를 가지고 있다.

4년제 일반대학교의 지표에서 편제정원 기준 교원 확보율은 재학생 충원율 (.591\*)과 상급학교 진학률 (.682\*\*), 학생 1인당 교육비 (.588\*)에 통계적으로 유의한 양의 상관관계를 가지고, 중도탈락 학생 비율 (-.463\*)과는 통계적으로 유의한 음의 상관관계를 가지고 있다.

전문대학의 지표에서 편제정원 기준 교원 확보율은 신입생 최종등록률 (.578\*), 재학생 충원율 (.446\*)과 통계적으로 유의한 양의 상관관계를 가지고, 편제정원대비 전임교수 1인당 학생 수 (-.939\*\*)와는 통계적으로 유의한 음의 상관관계를 가지고 있다.

### 3. 변수선택 방법에 의한 회귀모형 추정

변수선택 방법으로 단계별 회귀방법 (stepwise regression)과 후진제거법 (backward elimination method), 전진선택법 (forward selection method)을 사용하였다. 그러나 단계별 회귀방법의 결과는 앞으로부터 선택하는 방법의 결과와 같은 결과를 얻었다.

후진제거법은 모든 가능한 회귀방정식을 비교하지 않고, 다음과 같은 절차에 따라 변수를 선택하여 주는 방법이다. 박성현 (1998)의 후진제거법의 절차는 다음과 같다.

단계 1. 모든 변수를 포함한 회귀방정식을 적합 시킨다.

단계 2. 변수 하나하나씩에 대한 부분 F-검정치 (partial F-test value)  $F_0$ 를 구한다.

단계 3. 가장 작은 부분 F-검정치  $F_L$ 과 F-분포표로부터의 기각치  $F_c$ 를 비교한다.

단계 4. 만약  $F_L < F_c$ 이면  $F_L$ 의 값을 준 변수  $x_L$ 을 제거시키고 나머지 변수들만을 가지고 회귀방정식을 적합 시킨다. 그리고 위의 단계 2로 돌아간다. 만약  $F_L \geq F_c$ 이면 제거시킬 변수는 없으며, 현재 사용하고 있는 변수들을 선택한다.

앞으로부터의 단계적 회귀방법은 김충락과 강근석 (2010)에 따르면 다음과 같다.

단계 1. 현재모형이 한 개의 설명변수를 가지고 있다면 이는 단순선형회귀가 된다. 이때 기울기가 0인지에 대한 부분 F검정은

$$F_i^* = \frac{MSR_i}{s_i^2}, i = 1, \dots, k - 1$$

이 되고, 이 중에서 가장 큰  $F^*$  값이  $F_{\alpha}(1, n-2)$ 보다 크면 그 변수는 첨가되고 그렇지 않으면 영모형을 채택하게 된다.

**단계 2.** 만약 단계 1에서  $X_7$ 이 첨가되었다고 하자. 이제  $X_7$ 에 또 다른 설명변수가 추가된, 두 개의 설명변수에 대한 현재모형을 부분 F검정한다. 즉,

$$F_i^* = \frac{MSR(X_i|X_7)}{s^2(X_7, X_i)} = \left( \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right)^2, i \neq 7$$

이 되는데  $MSR(X_i|X_7)$ 은  $X_7$ 이 있을 때의  $X_i$ 에 의한 부분회귀제곱합이고,  $s^2(X_7, X_i)$ 는  $X_7$ 과  $X_i$ 로 구성된 현재모형에서의 오차평균제곱합이다.  $F_i^*$  중에서 가장 큰 값이  $F_{\alpha}(1, n - 3)$ 보다 크면 그 변수는 첨가되고, 그렇지 않으면 여기서 끝낸다.

**단계 3.** 만약 단계 2에서  $X_3$ 이 추가되었다고 하자. 즉, 현재모형에는  $X_7$ 과  $X_3$ 이 있다. 단계 3에서는 이들 두 변수에 대해 부분 F검정을 한다. 이미 단계 2에서  $X_3$ 에 대한 부분 F검정이 실시되었으므로 단계 3에서는  $X_7$ 에 대한 부분 F검정만 하면 된다. 즉,

$$F_7^* = \frac{MSR(X_7|X_3)}{s^2(X_3, X_7)}$$

인데 이 값이  $F_{\alpha}(1, n - 3)$ 보다 크면  $X_7$ 은 현재모형에 남아있게 되고 그렇지 않으면 제거된다.

**단계 4.** 만약 단계 3에서  $X_7$ 이 남아있게 된다고 하자. 이제  $X_3$ 과  $X_7$ 이 아닌 변수들에 대해 단계 2와 단계 3을 반복 적용한다.

2장에서 열거한 대학 알리미에 제시된 많은 지표들 중에서 (1) 신입생 최종등록률, (2) 신입생 경쟁률, (3) 재학생 충원률, (4) 신입생 일반계 고등학교 비율, (5) 신입생 전문계 고등학교 비율, (6) 상급학교 진학률, (7) 건강보험 DB연계 취업률, (8) 편제정원대비 전임교수 1인당 학생수, (9) 중도 탈락 학생 비율, (10) 학생 1인당 교육비, (11) 재학생 1인당 장학금, (12) 편제정원 기준 교원 확보율-중에서 교육과학기술부가 중요시하는 평가지표인 (1) 신입생 최종등록률, (3) 재학생 충원률, (7) 건강보험 DB연계 취업률과 (12) 편제정원 기준 교원 확보율에 대해서 지표들을 종속변수로 나머지 지표들을 독립변수로 두고 종속변수와 독립변수들의 인과관계를 고려하여 회귀모형을 추정하고자 한다.

최적화 회귀방정식을 추정하기 위한 방법으로 앞에서 제시한 회귀분석의 변수선택 방법인 단계별 회귀방법과 후진제거법을 사용하였다.

표 3.1의 신입생 최종등록률은 일반대의 경우, 단계별 회귀방법에서는 신입생 경쟁률의 계수가 한 단위씩 증가할 때마다 최종등록률은 4.468씩 증가하고, 후진제거법에서는 신입생 경쟁률의 계수는 선택된 다른 변수들이 고정되어 있을 때, 신입생 경쟁률이 한 단위씩 증가하면 신입생 최종등록률은 3.926씩 증가한다. 반면에 전임교원 1인당 학생수와 재학생 1인당 교육비와 편제정원 대비 교원확보율들의 계수에 대하여서는 자신 외의 선택되어진 다른 변수들을 고정시키고, 전임교원 1인당 학생수와 재학생 1인당 교육비와 편제정원 대비 교원확보율 한 단위씩 각각 증가시키면, 신입생 최종등록률은 각각 1.234, 0.002, 0.958씩 감소한다.

전문대의 경우에는 신입생 최종등록률은 재학생 충원율과 학생 1인당 교육비의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 재학생 충원율과 학생 1인당 교육비가 한 단위씩 각각

표 3.1 신입생 최종등록률에 대한 변수선택의 결과

신입생 최종등록률							
단계별 회귀방법				후진제거법			
일반대		전문대		일반대		전문대	
결정계수=.458		결정계수=.853		결정계수=.749		결정계수=.853	
수정된결정계수=.428		수정된결정계수=.821		수정된결정계수=.681		수정된결정계수=.821	
변수	계수	변수	계수	변수	계수	변수	계수
상수	72.223	상수	56.60	상수	192.489	상수	56.60
신입생경쟁률	4.468	재학생충원률	.804	신입생경쟁률	3.926	재학생충원률	.804
		취업률	-.797	전임교원학생수	-1.234	취업률	-.797
		전임교원학생수	-.647	학생1인교육비	-.002	전임교원학생수	-.647
		학생1인교육비	.004	교원확보율	-.958	학생1인교육비	.004

증가하면 신입생 최종등록률은 각각 0.804와 0.004 만큼 증가한다. 반면에 취업률과 전임교수 1인당 학생 수의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 취업률과 전임교수 1인당 학생 수가 한 단위씩 각각 증가하면 신입생 최종등록률은 각각 0.797와 0.647 만큼 감소한다.

표 3.2에서 일반대학의 경우에 재학생 충원률에 대한 단계별 회귀방법에서, 신입생 최종등록률과 전문계 고등학생 입학생 비율의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 신입생 최종등록률과 전문계 고등학생 입학생 비율이 한 단위씩 각각 증가하면 재학생 충원률은 각각 0.907와 2.014 만큼 증가한다. 반면에 중도 탈락학생 비율의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 중도 탈락학생 비율이 한 단위씩 각각 증가하면 재학생 충원률은 7.514 만큼 감소한다.

후진제거법의 경우에는 일반계 고등학생의 비율과 전문계 고등학생의 비율에 대한 계수는 선택되어진 다른 변수들이 고정되어 있을 때, 일반계 고등학생의 비율과 전문계 고등학생 입학생 비율이 한 단위씩 각각 증가하면 재학생 충원률은 각각 1.845와 3.395 만큼 증가한다. 반면에 중도 탈락학생 비율과 학생 1인당 교육비와 전임교원 1인당 학생수의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 각 계수가 한 단위씩 각각 증가하면 재학생 충원률은 탈락학생 비율과 학생 1인당 교육비와 전임교원 1인당 학생수에 대하여 각각 6.528, 0.04, 0.850씩 감소한다.

표 3.2 재학생 충원률에 대한 변수선택의 결과

재학생 충원률							
단계별 회귀방법				후진제거법			
일반대		전문대		일반대		전문대	
결정계수=.704		$R^2 = .900$		결정계수=.896		결정계수=.900	
수정된결정계수=.648		수정된결정계수=.878		수정된결정계수=.835		수정된결정계수=.878	
변수	계수	변수	계수	변수	계수	변수	계수
상수	27.968	상수	-4.019	상수	-42.944	상수	-4.019
신입생등록률	.907	신입생등록률	.833	일반계고비율	1.845	신입생등록률	.833
전문계입학비	2.014	취업률	.796	전문계고비율	3.395	취업률	.796
중도탈락율	-7.514	학생1인교육비	-.004	중도탈락율	-6.528	학생1인교육비	-.004
		신입생경쟁률	1.859	학생1인교육비	-.004	신입생경쟁률	1.859
				전임교원학생수	-.850		

재학생 충원률은 전문대학에 있어서 대학의 재정 및 대학의 인지도에 영향을 미치는 매우 중요한 지표이다. 전문대의 경우에 재학생 충원률은 신입생 등록률과 취업률, 신입생 경쟁률의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 재학생 충원률은 신입생 등록률과 취업률, 신입생 경쟁률이 한 단위씩 각각 증가하면 재학생 충원률은 각각 0.833와 0.796, 1.859 만큼 증가한다. 반면

에 학생 1인당 교육비의 계수는 자신 외에 선택되어진 다른 변수들이 고정되어 있을 때, 학생 1인당 교육비가 한 단위씩 각각 증가하면 재학생 충원률은 0.004씩 감소한다.

표 3.3 건강보험 DB연계 취업률에 대한 변수선택의 결과

건강보험 DB연계 취업률							
단계별 회귀방법				후진제거법			
일반대		전문대		일반대		전문대	
결정계수=.619		결정계수=.767		결정계수=.619		결정계수=.767	
수정된결정계수=.574		수정된결정계수=.716		수정된결정계수=.574		수정된결정계수=.716	
변수	계수	변수	계수	변수	계수	변수	계수
상수	35.90	상수	11.789	상수	21.552	상수	11.789
학생교육비	.006	재학생충원률	.672	학생교육비	.006	재학생충원률	.672
학생장학금	-.012	최종등록률	-.493	학생장학금	-.012	최종등록률	-.493
		학생교육비	.005			학생교육비	.005
		진학률	-1.456			진학률	-1.456

표 3.3에서 건강보험 DB연계 취업률은 일반대학의 경우에는 단계별 회귀방법과 후진제거법에서는 학생 1인당 교육비의 계수는 재학생 1인당 장학금이 고정되어 있을 때 학생 1인당 교육비가 한 단위씩 증가하면, 취업률은 0.006씩 증가한다. 반면에, 재학생 1인당 장학금의 계수는 학생 1인당 교육비가 고정되어 있을 때 재학생 1인당 장학금이 한 단위씩 증가하면, 취업률은 0.012씩 감소한다.

전문대의 경우, 건강보험 DB연계 취업률은 선택되어진 다른 변수들이 고정되어 있을 때, 재학생 충원률과 학생 1인당 교육비가 각각 한 단위씩 증가하면, 취업률은 각각 0.672와 0.005씩 증가하고, 반면에 타 대학으로의 진학률과 신입생 최종 등록률의 계수는 선택되어진 다른 변수들이 고정되어 있을 때, 진학률과 신입생 최종 등록률이 한 단위씩 증가하면 건강보험 DB연계 취업률이 각각 0.493과 1.456씩 감소하는 경향을 보인다.

표 3.4 편제정원 기준 교원 확보율에 대한 변수선택의 결과

편제정원 기준 교원 확보율							
단계별 회귀방법				후진제거법			
일반대		전문대		일반대		전문대	
결정계수=.928		결정계수=.881		결정계수=.928		결정계수=.881	
수정된결정계수=.920		수정된결정계수=.875		수정된결정계수=.920		수정된결정계수=.875	
변수	계수	변수	계수	변수	계수	변수	계수
상수	144.406	상수	93.5	상수	144.406	상수	93.5
전임교원학생수	-1.360	전임교원학생수	-.984	전임교원학생수	-1.360	전임교원학생수	-.984
최종등록률	-.352			최종등록률	-.352		

표 3.4에서 편제정원 기준 교원 확보율은 일반대학의 경우에는 단계별 회귀방법과 후진제거법에서 전임교원 1인당 학생수와 신입생 최종등록률 계수는 한쪽이 고정되어 있을 때 각 변수가 한 단위씩 증가하면 기준 교원확보율은 1.36과 0.352씩 감소한다

전문대의 경우에는 전임교원 1인당 학생수가 한 단위씩 증가할수록 기준 교원확보율은 0.984씩 감소한다

#### 4. 취업률과 재학생 충원률에 대한 결론

2절의 상관관계 검정에서, 4년제 일반대학교의 경우에 취업률은 오직 신입생 경쟁률과 통계적으로 유의한 양의 상관관계를 가지고, 3절에서의 변수선택에 있어서 취업률은 재학생 1인당 교육비와 재학

생 1인당 장학금에 영향을 받는다. 그러므로 취업률은 재학생 1인당 교육비와 재학생 1인당 장학금이 서로 한쪽이 고정되어 있을 때, 한 단위씩 증가하면 취업률은 각각 0.006과 -0.012씩 증가한다.

2절의 상관관계 검정에서, 전문대학 취업률은 재학생 충원률과는 통계적으로 유의한 양의 상관관계를 가지고, 상급학교 진학률과 편제정원대비 전임교수 1인당 학생 수와 중도탈락 학생 비율에 대해서는 통계적으로 유의한 음의 상관관계를 가진다. 3절에서의 변수선택에 있어서, 전문대학 취업률은 재학생 충원률과 학생 1인당 교육비가 각각 한 단위씩 증가하면, 취업률은 각각 0.672와 0.005씩 증가하고, 반면에 타 대학으로의 진학률과 신입생 최종 등록률의 계수는 선택되어진 다른 변수들이 고정되어 있을 때, 진학률과 신입생 최종 등록률이 한 단위씩 증가하면 건강보험 DB연계 취업률이 각각 0.493과 1.456씩 감소하는 경향을 보인다.

일반대학의 재학생 충원률은, 2절의 상관관계 검정에서는 신입생 최종등록률과 신입생 경쟁률 그리고 편제정원 기준 교원 확보율과는 통계적으로 유의한 양의 상관관계를 가지고, 중도탈락 학생 비율에 대해서는 통계적으로 유의한 음의 상관관계를 가진다. 3절에서의 재학생 충원률 변수선택에 있어서는 신입생 최종등록률과 전문계 고등학생 입학생 비율의 계수는 각각 0.907와 2.014로 나타나고, 반면에 중도 탈락학생 비율의 계수는 -7.514로 나타난다.

후진제거법의 경우에는 일반계 고등학생의 비율과 전문계 고등학생의 비율에 대한 계수는 1.845와 3.395 계수 값을 가지고, 반면에 중도 탈락학생 비율과 학생 1인당 교육비와 전임교원 1인당 학생수의 계수는 각각 -6.528, -0.04, -0.850씩 유의한 계수 값을 가진다.

전문대학의 재학생 충원률은 2절의 상관관계 검정에서, 신입생 최종등록률, 신입생 경쟁률 그리고 건강보험 DB연계 취업률과 편제정원 기준 교원 확보율과는 통계적으로 유의한 양의 상관관계를 가지고, 편제정원대비 전임교수 1인당 학생 수에 대해서는 통계적으로 유의한 음의 상관관계를 가진다. 3절에서의 변수선택에 있어서는 전문대의 재학생 충원률 경우에는 신입생 등록률과 취업률, 신입생 경쟁률의 계수는 각각 0.833와 0.796, 1.859 의 유의한 값을 가지고, 반면에 학생 1인당 교육비의 계수는 -0.004의 유의한 값을 가진다.

결론적으로 재학생 충원률은 일반대학의 경우, 재학생 중도 탈락률에 가장 큰 영향을 받고 있으므로 재학생들의 중도탈락을 방지하기 위한 모든 노력을 기울여야 할 것이다. 전문대의 경우에는 재학생 충원률은 취업률에 가장 밀접한 연관성을 가지고, 신입생 경쟁률과 신입생 충원률에 상당한 영향을 끼친다.

## 참고문헌

- 교육과학기술부 (2011). <“대학구조개혁위원회” 대학 구조개혁 기본 틀 마련>, 교육과학기술부, 서울
- 김충락, 강근석 (2010). <회귀분석>, 교우사, 서울.
- 박성현 (1998). <회귀분석>, 민영사, 서울.
- 송필준, 김종태 (2010). 로지스틱함수모형과 비례이동평균모형에 의한 학생 수 추계와 분석. <한국데이터정보과학회지>, **21**, 503-511.
- 오현숙 (2011). 고등교육 구조개혁의 방향과 과제. <대학구조개혁위원회 제5차 오픈 포럼 발제 자료>, 교육과학기술부, 서울.
- 윤용화, 김종태 (2010). 비선형 회귀모형을 이용한 학년별 학생수 추계. <한국데이터정보과학회지>, **23**, 71-77.
- 최경호, 한동욱 (2010). 입학사정 전형요소 상대적 중요도 결정에 대한 퍼지 계층분석적 접근방법 - X대학교 사례연구. <한국데이터정보과학회지>, **21**, 699-708.
- 홍중선, 김명진 (2010). 다중회귀모형에서 전진선택과 후진제거의 기하학적 표현 <한국데이터정보과학회지>, **21**, 901-908.
- Chatterjee, S., Hadi, A. S. and Price, B. (2000). *Regression analysis by example*, 3rd Ed., John Wiley & Sons, New York.
- Draper, N. and Smith, H. (1981). *Applied regression analysis*, 2nd Ed., Wiley, New York.



## The correlation and regression analyses based on variable selection for the university evaluation index<sup>†</sup>

Piljun Song<sup>1</sup> · Jongtae Kim<sup>2</sup>

<sup>1,2</sup>Department of Computing & Statistics, Daegu University

Received 16 January 2012, revised 7 March 2012, accepted 19 April 2012

### Abstract

The purpose of this study is to analyze the association between indicators and to find statistical models based on important indicators at 'College Notifier' in Korea Council for University Education. First, Pearson correlation coefficients are used to find statistically significant correlations. By variable selection method, the important indicators are selected and their coefficients are estimated. As variable selection method, backward and stepwise methods are employed.

*Keywords:* Correlation, regression analysis, university evaluation index, variable selection.

---

<sup>†</sup> This research was supported by the Daegu University Research Grant 2010.

<sup>1</sup> Professor, Department Computing & Statistics, Daegu University, Kyungbuk 712-714, Korea.

<sup>2</sup> Professor, Department Computing & Statistics, Daegu University, Kyungbuk 712-714, Korea.  
E-mail: jtkim@daegu.ac.kr.