

Effects of Financial College Tuition Support by Korean Parents using a Hierarchical Bayes Model

Man-Suk Oh^{a,1} · Hyun Sook Oh^b · Min Jung Oh^a

^aDepartment of Statistics, Ewha Womans University

^bDepartment of Applied Statistics, Gachon University

(Received January 8, 2013; Revised February 21, 2013; Accepted March 5, 2013)

Abstract

College tuition is a significant economic, social, and political issue in Korea. We conduct a Bayesian analysis of a hierarchical model to address the factors related to college tuition based on a survey data collected by Statistics Korea. A binary response variable is selected depending on if more than 70% of tuition costs are supported by parents, and a hierarchical Probit model is constructed with areas as groups. A set of explanatory variables is selected from a factor analysis of available variables in the survey. A Markov chain Monte Carlo algorithm is used to estimate parameters. From the analysis results, income and stress are significantly related to college tuition support from parents. Parents with high income tend to support children's college tuition and students with parents' financial support tend to be mentally less stressed; subsequently, this shows that the economic status of parents significantly affects the mental health of college students. Gender, a healthy life style, and college satisfaction are not significant factors. Comparing areas in terms of the degrees of correlation between stress/income and tuition support from parents, students in Kangwon-do are the most mentally stressed when parents' support is limited; in addition, the positive correlation between parents support and income is stronger in big cities compared to provincial areas.

Keywords: Hierarchical model, Probit model, factor analysis, Markov chain Monte Carlo.

1. 서론

대학등록금 문제는 IMF 이후 지속적으로 한국사회에서 경제적, 사회적, 정치적 주요 현안 문제로 대두되어 왔다. 대학 등록금이 꾸준히 오르면서 대학등록금에 대한 경제적 부담이 더욱 가중되고 있으며 대학생과 학부모는 대학등록금 수준이 지나치게 높다고 인식하고 있다. 고교 졸업생의 대학 진학률이 OECD 최고수준인 74%에 달하면서 대학교육이 한국 사회에서 보편적 교육으로 자리 잡아 대학 등록금이 거의 전 국민에 해당되는 경제적 부담이 되고 있다. 또한 대학교육에 교육 서비스 정신이 도입되면서 각 대학들이 질 좋은 교육 서비스를 위한 교육 투자를 늘리면서 등록금이 꾸준히 인상되어 왔고 이로 인한 일반 서민들의 경제적 부담이 증가하였다.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology(No. 2010-0010285).

¹Corresponding author: Professor, Department of Statistics, Ewha Womans University, 21 Daehyun Dong, Seodaemun Gu, Seoul 120-750, Korea. E-mail: msoh@ewha.ac.kr

Table 1.1. College tuitions (2004~2005)

Country	GDP (\$)	College tuition(\$)		Ratio of college tuition to GDP(%)	
		public	private	public	private
Korea	22,666	3,883	7,406	17.1	32.7
USA	41,571	5,027	18,604	12.1	44.8
Japan	31,406	3,920	6,117	12.5	19.5
Canada	34,550	3,464	-	10.0	-
Australia	31,318	3,855	7,452	12.3	23.8
England	30,277	-	1,737	-	5.7
Italy	29,218	1,017	3,520	3.5	12.0

한국의 대학등록금 수준을 다른 나라와 비교하여 살펴보기 위하여 미국, 일본, 캐나다 등 주요 선진국의 대학등록금 수준과 비교해 보았다. Table 1.1은 각국의 2004~2005년도 대학등록금을 구매력(Purchasing Power Parities; PPP)지수로 환산한 가격과 1인당 2005년도 GDP 대비 등록금 비율을 제시하고 있다 (OECD, 2008; IMF, 2008). PPP 지수를 보면, 국공립대학은 평균 \$3,883, 사립대학은 평균 \$7,406으로 사립대학의 등록금이 국공립대에 비하여 약 2배 더 높다. 비교 국가 중 국공립대 및 사립대 모두 미국 대학등록금이 가장 비싸며, 다음으로 국공립대는 일본, 한국, 호주 순이었고, 사립대는 호주, 한국 일본 순이었다. 그리고 국공립대와 사립대의 등록금이 3배 이상 차이가 나는 국가는 미국과 이탈리아이다. 이렇게 본다면 우리나라 대학등록금 수준이 다른 선진국에 비해서도 결코 낮지 않은 수준임을 알 수 있다.

또한 국민 개개인의 경제능력에 비해 대학등록금 수준이 어느 정도인지를 알려주는 1인당 GDP 대비 대학등록금 비율을 보면, 한국의 국민 1인당 GDP 대비 대학등록금이 차지하는 비율은 국공립대는 17.1%, 사립대는 32.7%이었다. 특히 한국 대학교육의 대부분을 차지하고 있는 사립대의 등록금을 보면 미국 사립대 등록금이 1인당 GDP 대비 44.8%인 것을 제외하고 가장 높은 수치로 미국을 제외한 다른 선진국과 비교해도 한국의 대학등록금이 매우 높은 수준임을 알 수 있다.

한편, 최근 들어 대학등록금 부담이 가계경제에 더욱 부담을 주고 있음을 고려하여 우리나라 가구당 가계소득에 대한 대학등록금 비율을 Figure 1.1에 나타내 보았다 (Cheon, 1997; Statistics Korea, 2009; Choi, 2006; Bahn, 2008). 이 비율은 1980년 이후로 감소하다가 2000년대 이후 다시 증가하는 경향을 보여주고 있다. 특히 2003년 이후의 추세를 보면 국공립대학인 경우 2003년에 8.5%에서 2007년도에는 11.2%까지 증가했고, 사립대학인 경우 2003년에 17.5%에서 2007년도에는 22%까지 증가하여 대학등록금에 대한 가계부담이 증가하는 것으로 나타났다.

본 논문에서는 대학 등록금 부담에 대한 최근 한국 사회의 관심을 반영하여, 2010년 통계청에서 주관한 ‘2010년 사회조사’ 설문지 자료를 기반으로 대학 등록금 조달 재원에 대한 분석을 수행하고자 한다. “귀하는 2010년 1학기 대학교 총 등록금을 어떠한 방법으로 마련하셨습니다? 해당되는 곳 모두에 각각 기입하여 주십시오.”라는 설문지 질문에 응답한 대학생들은 5개의 항목 (부모님, 대출, 스스로 벌어서 마련, 장학금, 기타) 중에서 총합이 100%가 되도록 해당되는 곳 모두에 %로 기입하였다. 한국의 현재 상황에서 부모님이 경제적 능력이 있어 등록금을 조달할 경우 학생이 경제적 부담없이 안정적으로 학업을 수행할 수 있다고 판단되어, 부모님이 등록금의 70% 이상을 조달하는 경우와 그렇지 않은 경우를 분류하여 이항반응 자료로 놓고 분석을 수행하였다. 기준을 70%로 택한 것은 대학 등록금을 학기당 대략 400만원으로 간주하여 30%인 120만원까지는 대출이나 아르바이트, 장학금 등으로 충당하기에 무리가 없는 액수로 보았기 때문이다. 즉, 학기당 120만원 이상을 부모님의 경제적 지원 아닌 다른 재원에서 충

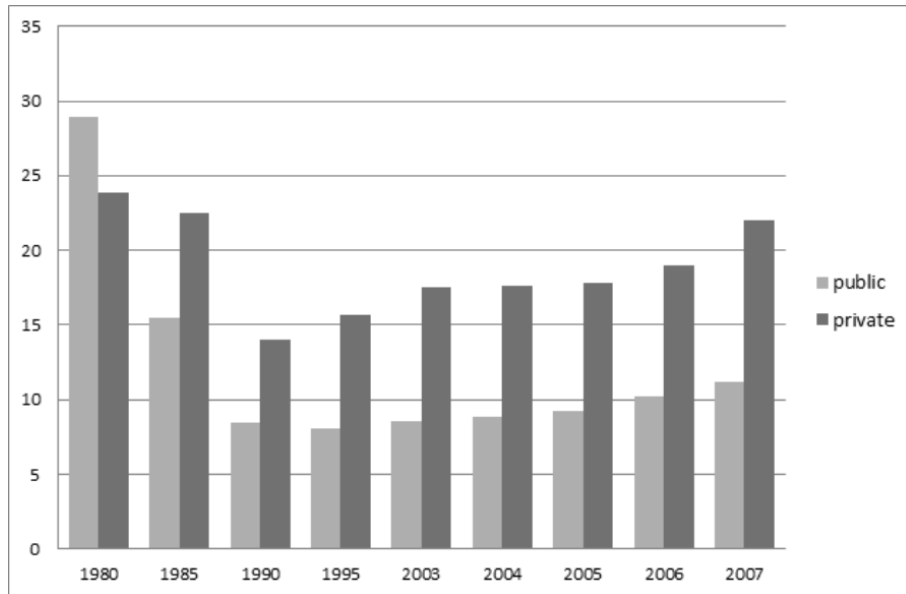


Figure 1.1. Ratio(%) of college tuition to household income.

당해야 하는 경우를 등록금으로 인한 경제적 부담이 과중한 경우로 간주하여 분석을 수행하였다.

분석의 중요한 목적 중의 하나는 위의 반응변수와 관련있는 변수들이 어떤 것들이 있는지 살펴보고자 하는 것이다. 2010년 사회조사 설문 자료에는 매우 많은 변수들이 존재하는데, 다수의 변수들이 서로 높은 상관성을 지녀 다중 공선성을 유발할 수 있기 때문에 요인분석을 통하여 변수들 내에 존재하는 구조를 관찰한 결과 중요한 5개의 요인(정신건강, 성별, 생활건강, 소득, 학교생활만족도)을 발견할 수 있었다. 이들 5개의 요인들은 서로 다른 특징을 뚜렷이 보여 주어 요인분석이 매우 유용함을 볼 수 있었다.

지역별로 소득이 상당한 차이가 있고 또한 서울을 제외한 지역에서 다수의 학생들이 출신 고장에 있는 대학에 진학한다는 점을 고려하여 지역변수를 설명변수에 포함시키지 않고 지역을 하나의 계층으로 하는 계층 모형을 설정하고 모수에 대한 베이지안 분석을 수행하였다. 이러한 계층모형에 대한 베이지안 분석은 지역간 공통적인 특성과 지역별로 고유한 특성을 동시에 고려하는 합리적인 분석결과를 제공하는 장점이 있다.

분석결과 지역별 회귀계수 추정치를 살펴보면, 대다수의 지역에서 정신건강지수와 소득은 유의한 요인으로 나타났다. 소득요인에는 주택점유 형태와 가구소득 변수가 포함되는데, 이는 부모님의 경제적 능력을 나타내는 지표로서 부모님의 경제적 능력이 높을수록 부모님이 등록금의 대부분을 책임지는 것으로 나타났다. 또한 부모님이 경제지원을 할수록 학생이 느끼는 정신적 스트레스가 한결 덜함을 알 수 있었다. 반면, 성별의 경우 인천지역을 제외한 모든 지역에서 유의한 관련이 없는 것으로 분석결과가 나타나 부모님의 등록금 조달에 있어서 자녀의 성별에는 차이는 없는 것으로 보인다. 또한 아침식사여부, 적정수면, 규칙적 운동, 정기건강검진 등의 생활건강 요인은 모든 지역에서 유의하지 않게 나타나 등록금에 대한 부담이 정신적 건강에는 영향을 미치지만 다행히 생활 건강에는 의미있는 영향을 미치지 않음을 알 수 있었다. 마지막으로, 학교생활 만족도는 전북지역을 제외한 나머지 지역에서 유의한 관련성이 없는 것으로 나타나 전반적으로 등록금에 대한 부담이 학교에 대한 인식에 영향을 주지는 않는 것으로 나타났다.

지역별 차이를 살펴보면, 지역별로 부모님의 등록금 지원과 관련된 요인들에 대한 계수 추정치에 대한 유의성과 계수추정치 절대적 크기에 차이가 있음을 알 수 있었다. 특히 정신적 스트레스와 부모님의 지원간의 관계를 보면 강원도에서 다른 지역에 비하여 월등히 높은 관련성을 보였고 반면 서울, 부산, 대구, 대전, 울산 등의 대도시에서는 약한 관련성을 보였다. 소득과 부모님의 지원간의 관계에서는 서울, 부산, 광주, 울산 등의 대도시에서 높은 관련성을 보이고 대부분의 행정도에서는 유의하지 않은 관련성을 보여 대조를 이루었다.

본 논문의 구성은 다음과 같다. 2절에서는 분석에 사용된 계층적 프로빗 모형을 기술하고 3절에서는 계층적 프로빗 모형에 대한 베이지안 분석 수행을 위한 깃스 표본기법을 설명한다. 4절에서는 2010 사회조사 자료에서 요인분석을 통하여 중요한 요인들을 선별하고 이 요인들을 설명변수로 하는 계층적 프로빗 모형의 베이지안 분석결과를 제시하고 해석한다. 5절에서는 요약과 결론으로 끝을 맺는다.

2. 계층적 프로빗 모형

K 개의 소집단이 존재하고, 소집단 i , $1 \leq i \leq K$ 에서 n_i 개의 관측치 $(\mathbf{x}_{ij}, y_{ij})$, $j = 1, \dots, n_i$ 가 주어진다고 가정하자. 여기에서 \mathbf{x}_{ij} 는 p 차원 벡터로, i 번째 소집단에서 j 번째 개체의 설명변수들의 관측값이며 y_{ij} 는 \mathbf{x}_{ij} 에 대응하는 반응변수 Y_{ij} 의 관측값으로 0 또는 1을 갖는 이항자료이다.

반응변수 Y_{ij} 각각에 대하여 독립적으로 베르누이 분포 $\text{Ber}(p_{ij})$ 를 가정한다. 소집단 i 에서 확률 p_{ij} 와 설명변수 \mathbf{x}_{ij} 를 잇는 연결함수로 $g = \Phi$ 를 사용하고, 소집단 내에서는 동일한 회귀계수를 가정하면 $p_{ij} = \Phi(\mathbf{x}'_{ij}\boldsymbol{\beta}_i)$ 이다. 정리하면,

$$\begin{aligned} Y_{ij} &\sim \text{Ber}(p_{ij}), \\ p_{ij} &= \Phi(\mathbf{x}'_{ij}\boldsymbol{\beta}_i) \end{aligned} \quad (2.1)$$

의 프로빗 모형을 가정한다.

소집단의 회귀계수들 $\boldsymbol{\beta}_i$, $i = 1, \dots, K$ 에 대하여 다음과 같은 공통 사전분포를 가정하고

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \Sigma). \quad (2.2)$$

사전분포의 초모수 $(\boldsymbol{\beta}, \Sigma)$ 에 대하여 다음과 같은 반공액 사전분포를 가정한다.

$$\begin{aligned} \boldsymbol{\beta} | \Sigma &\sim N\left(\boldsymbol{\beta}_0, \frac{1}{k_0}\Sigma\right), \\ \Sigma &\sim \text{IW}(\alpha_0, \Sigma_0). \end{aligned} \quad (2.3)$$

여기에서 IW는 역위샷트 분포를 나타낸다. 위의 반공액 사전분포 대신 $\boldsymbol{\beta}$ 와 Σ 에 대하여 다음과 같은 무정보사전분포를 가정할 수도 있다.

$$\pi(\boldsymbol{\beta}, \Sigma) = |\Sigma|^{-\frac{(p+1)}{2}}.$$

3. 마코브 체인 몬테칼로

$\boldsymbol{\beta}_i$ 에 대하여 식 (2.2)와 같은 정규 사전분포를 가정한 경우 식 (2.1)의 우도함수와 사전밀도함수는 짝 관계가 아니어서 사후분포가 편리한 형태로 주어지지 않고 따라서 베이지안 추론이 쉽지 않다. 그러나 다음과 같은 잠재변수를 고려하면 깃스표본기법을 사용하여 쉽게 베이지안 추론을 이끌어낼 수 있다. 잠

재변수 Z_{ij} 를 다음과 같이 정의한다 (Albert와 Chib, 1993).

$$Z_{ij} \sim N(\mathbf{x}'_{ij}\boldsymbol{\beta}_i, 1),$$

$$Y_{ij} = \begin{cases} 1, & \text{if } Z_{ij} > 0, \\ 0, & \text{if } Z_{ij} \leq 0. \end{cases}$$

위의 정의에 의하면 이항변수 Y_{ij} 가 1인 사건은 연속변수 Z_{ij} 가 0 보다 큰 사건과 동일하다. Z_{ij} 는 관측된 변수가 아니므로 모수로 취급하여 새로운 우도함수를 구해보면

$$L(\boldsymbol{\beta}_i, \mathbf{z}_i | \mathbf{y}_i) = \prod_{j=1}^{n_i} \phi(z_{ij} | \mathbf{x}'_{ij}\boldsymbol{\beta}_i, 1) [I(z_{ij} > 0, y_{ij} = 1) + I(z_{ij} \leq 0, y_{ij} = 0)]$$

이며, 여기에서 $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$, $\mathbf{z}_i = (z_{i1}, \dots, z_{in_i})$, $\phi(z_{ij} | \mathbf{x}'_{ij}\boldsymbol{\beta}_i, 1)$ 는 $N(\mathbf{x}'_i\boldsymbol{\beta}_i, 1)$ 분포의 z_{ij} 에서의 확률밀도함수 값이다.

이 우도함수를 $\boldsymbol{\beta}_i$ 의 정규사전분포와 결합한 후 조건부 사후분포를 구해보면, $\mathbf{X}_i = (\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \dots, \mathbf{x}'_{in_i})'$ 일 때,

$$\begin{aligned} \pi(\boldsymbol{\beta}_i | \mathbf{z}_i, \boldsymbol{\beta}, \Sigma, \mathbf{y}_i) &\propto \prod_{j=1}^{n_i} f(z_{ij} | \boldsymbol{\beta}_i, \boldsymbol{\beta}, \Sigma) \pi(\boldsymbol{\beta}_i | \boldsymbol{\beta}, \Sigma) \\ &\propto \exp \left[-\frac{1}{2} \{ (\mathbf{z}_i - \mathbf{X}_i\boldsymbol{\beta}_i)' (\mathbf{z}_i - \mathbf{X}_i\boldsymbol{\beta}_i) + (\boldsymbol{\beta}_i - \boldsymbol{\beta})' \Sigma^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\beta}) \} \right] \\ &\propto \exp \left[-\frac{1}{2} \{ \boldsymbol{\beta}_i' (\mathbf{X}'_i\mathbf{X}_i + \Sigma^{-1}) \boldsymbol{\beta}_i - 2\boldsymbol{\beta}_i' (\mathbf{X}'_i\mathbf{z}_i + \Sigma^{-1}\boldsymbol{\beta}) \} \right] \end{aligned}$$

이므로

$$\begin{aligned} \boldsymbol{\beta}_i | \mathbf{z}_i, \boldsymbol{\beta}, \Sigma, \mathbf{y}_i &\sim N(\boldsymbol{\mu}_\pi, \Sigma_\pi), \\ \Sigma_\pi &= (\mathbf{X}'_i\mathbf{X}_i + \Sigma^{-1})^{-1} \\ \boldsymbol{\mu}_\pi &= \Sigma_\pi (\mathbf{X}'_i\mathbf{z}_i + \Sigma^{-1}\boldsymbol{\beta}) \end{aligned}$$

이다.

잠재변수 Z_{ij} 의 조건부 사후분포는

$$Z_{ij} | \boldsymbol{\beta}_i, \boldsymbol{\beta}, \Sigma, \mathbf{y}_i \sim \begin{cases} N(\mathbf{x}'_{ij}\boldsymbol{\beta}_i, 1)I(Z_{ij} > 0), & \text{if } y_{ij} = 1, \\ N(\mathbf{x}'_{ij}\boldsymbol{\beta}_i, 1)I(Z_{ij} \leq 0), & \text{if } y_{ij} = 0 \end{cases}$$

으로, 절단된 정규분포를 따른다.

$\boldsymbol{\beta}$ 와 Σ 의 사전분포를 식 (2.3)과 같이 가정했을 때, $\boldsymbol{\beta}$ 의 조건부 사후분포와 Σ 의 조건부 사후분포는 다음과 같이 구할 수 있다.

$$\begin{aligned} \boldsymbol{\beta} | \boldsymbol{\beta}_i, \mathbf{z}_i, \Sigma, \mathbf{y}_i &\sim \prod_{i=1}^K \pi(\boldsymbol{\beta}_i | \boldsymbol{\beta}, \Sigma) \pi(\boldsymbol{\beta} | \Sigma) \\ &\sim N \left(\frac{K\bar{\boldsymbol{\beta}} + k_0\boldsymbol{\beta}_0}{K + k_0}, \frac{1}{K + k_0} \Sigma \right), \end{aligned}$$

$$\begin{aligned}\Sigma | \boldsymbol{\beta}_i, \mathbf{z}_i, \boldsymbol{\beta}, \mathbf{y}_i &\sim \text{IW}(K + \alpha_0, \Lambda_n), \\ \Lambda_n &= \Sigma_0 + S + \frac{Kk_0}{K + k_0} (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)', \\ S &= \sum_{i=1}^K (\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}) (\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})'.\end{aligned}$$

만약 $\boldsymbol{\beta}$ 와 Σ 의 사전분포를 무정보사전분포 $\pi(\boldsymbol{\beta}, \Sigma) = |\Sigma|^{-(p+1)/2}$ 로 가정하면, $\boldsymbol{\beta}$ 의 조건부 사후분포와 Σ 의 조건부 사후분포는 아래와 같이 달라진다.

$$\begin{aligned}\boldsymbol{\beta} | \boldsymbol{\beta}_i, \mathbf{z}_i, \Sigma, \mathbf{y}_i &\sim N\left(\bar{\boldsymbol{\beta}}, \frac{1}{K}\right), \\ \Sigma | \boldsymbol{\beta}_i, \mathbf{z}_i, \boldsymbol{\beta}, \mathbf{y}_i &\sim \text{IW}(K - 1, S).\end{aligned}$$

이상과 같이 $\boldsymbol{\beta}_i$ 와 Z_{ij} , $\boldsymbol{\beta}$, Σ 의 조건부 사후분포가 표본생성에 편리한 형태로 주어지므로 김스표본기법을 이용하여 $(\boldsymbol{\beta}_i, Z_{ij}, \boldsymbol{\beta}, \Sigma)$ 의 사후표본을 얻을 수 있고 사후표본을 이용하여 사후평균, 사후분산 등 추론에 필요한 모수 추정치를 얻을 수 있다.

4. 대학 등록금 조달 재원자료의 분석

4.1. 자료 요약

앞에서 소개한 모형과 베이지안 추론을 바탕으로 실제로 통계청에서 주관한 ‘2010년도 사회조사’의 설문지 자료를 분석한다. 설문 조사기간은 2010년 5월 17일부터 5월 29일까지이며 전국 약 17000 표본 가구의 만 15세 이상 가구원 37000명을 대상으로 하였다. 우리는 이 중에서 대학생 2136명의 응답결과를 분석에 사용하였다.

등록금 조달원에 대한 설문은 다음과 같다. “귀하는 2010년 1학기 대학교 총 등록금을 어떠한 방법으로 마련하셨습니다? 해당되는 곳 모두에 각각 기입하여 주십시오.”라는 설문지 질문에 응답한 대학생들은 5개의 항목(부모님의 도움, 대출, 스스로 벌어서 마련, 장학금, 기타)에 대하여 총합이 100%가 되도록 비율(%)을 기입하도록 하였다.

2010학년도 1학기 대학 등록금의 평균을 대략 400만원으로 간주하여 30%인 120만원까지는 대출이나 아르바이트, 장학금 등으로 충당하기에 무리가 없는 액수로 보고 반응변수 Y 를 등록금의 70%이상을 부모님으로부터 도움을 받는 학생인 경우 1, 그렇지 않은 경우 0 값을 가지는 이항변수로 정의하였다. 즉, 학기당 30% 이상을 부모님이 아닌 다른 재원에서 충당해야 하는 경우를 등록금으로 인한 경제적 부담이 과중한 경우로 간주하여($Y = 0$) 분석을 수행하였다.

전체 설문 응답자 중 Y 가 1인 경우가 66.62%이고, 0인 경우가 33.38%로 나타났다. 즉, 설문에 응답한 대학생 중 1/3이 등록금의 30% 이상을 스스로 해결해야 하는 경제적 부담을 지닌 것으로 나타났다.

Table 4.1는 등록금 조달 재원 외에 대학생들이 설문에서 응답한 변수들을 요약한 것이다. 지역을 보면 7개의 대도시와 9개의 행정도로 구분되어 있다. 응답 대학생의 15.2%가 서울, 10.7%가 경기도, 울산을 제외한 나머지 대도시에 각각 6.9%–8.5%가 거주하고 있으며 울산과 9개 행정도에는 각각 2.3%–4.8%가 거주하고 있다. 지역별 분포가 차이가 많이 나고 대도시를 제외한 행정도에는 타 지역에서 유입되는 학생들이 거의 없을 것으로 판단되어 지역을 소집단으로 하는 계층모형을 설정하였다.

Table 4.1의 나머지 변수들에 대하여 등록금 부담과 유의한 관련성을 알아보기 위하여 이 변수들을 설명변수로 하는 계층적 프로빗 모형을 적합 시키고자 한다. 단, ‘학교생활 만족도’ 변수는 결측값이 17.2%가 있어서 평균값인 3.5로 대체하였다.

Table 4.1. Explanatory variables.

Variable	Category
area	Seoul, Pusan, Daegu, Incheon, Kwangju, Daejeon, Ulsan, Kyungki, Kangwon, Chungbuk, Chungnam, Chonbuk, Chonnam, Kyungbuk, Kyungnam, Jeju
residence	free, rent, rent-charter, charter, own
sex	male, female
personal satisfaction	1(strongly unsatisfied)~5(strongly satisfied)
health	1(very bad)~5(very good)
health_eat breakfast	no, yes
health_sound sleep	no, yes
health_regular exercise	no, yes
health_regular screening	no, yes
smoking	no, yes
stress_family	1(strong)~4(none)
stress_general	1(strong)~4(none)
suicidal intention	no, yes
school satisfaction	1(strongly unsatisfied)~5(strongly satisfied), nonresponse
frequency of drink	less than once per month, 2~3 times per month, 1~2 times per week, 3~4 times per week, almost everyday, do not drink
household income (₩)/month	< 1000000, 1000000~2000000, . . . , 6000000~7000000, > 7000000

Table 4.2. KMO measure and Bartlett test.

	Kaiser-Meyer-Olkin measure	.674
	chi-square	3273.009
Bartlett test	d.f.	105
	<i>p</i> -value	.000

4.2. 요인분석

Table 4.1에서 지역을 제외한 16개의 변수를 모두 설명변수로 하기에는 변수의 수가 많고 또한 변수간 상관관계로 인하여 다중공선성 문제가 발생할 수 있으므로 요인분석을 통해 변수들 내에 존재하는 상관 구조를 발견하여 설명변수의 수를 줄이고자 한다. 요인분석은 여러 변수들 사이의 상관관계를 기초로 하여 정보의 손실을 최소화하면서 변수의 개수보다 적은 수의 요인(factor)으로 자료변동을 설명하는 다변량 기법이다.

자료가 요인분석에 적합하지 여부를 판단하기 위하여 KMO(Kaiser-Meyer-Olkin)와 Bartlett의 검정을 수행한 결과 적절성을 나타내는 KMO값이 0.674로 0.5보다 크므로 자료는 요인분석에 적합하다고 할 수 있다 (Table 4.2). 마찬가지로 Bartlett의 구형성 검정은 변수간의 상관행렬이 단위행렬인지 여부를 판단하는 검정방법이다. 여기서 유의확률이 0.000이므로 변수간의 행렬이 단위행렬이라는 귀무가설은 기각된다. 즉 변수들의 상관관계가 통계적으로 유의하다고 볼 수 있어 요인분석을 적용할 수 있다.

다음 유의한 요인의 수를 결정하기 위하여 요인이 설명할 수 있는 변수들의 분산 크기를 나타내는 고유값의 변화를 Figure 4.1의 스크리 차트(scree chart)로 알아보았다. 고유값이 1보다 크다는 것은 하나의 요인이 변수 1개 이상의 분산을 설명해 준다는 것을 의미한다. 따라서 고유값이 1 이상인 경우를 기본으로 해서 요인수를 결정하게 되면 이 그래프에서는 5개가 적당한 것으로 보인다.

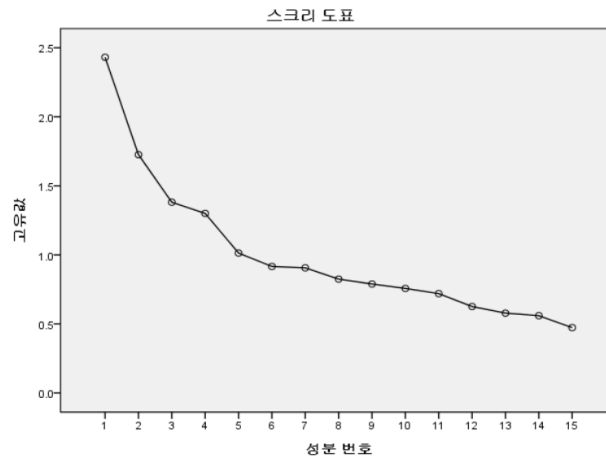


Figure 4.1. Scree chart.

Table 4.3. Rotated factors.

	factor				
	1	2	3	4	5
residence	.009	-.025	.052	.815	-.042
sex	-.202	-.752	-.107	-.029	-.070
personal satisfaction	.568	-.065	.127	.187	.377
health	.584	.089	.069	.039	.283
health_eat breakfast	.055	-.149	.618	.194	-.047
health_sound sleep	.165	-.097	.551	-.045	-.137
health_regular exercise	.050	.289	.633	-.059	.181
health_regular screening	-.145	-.006	.480	-.007	.457
smoking	-.031	.748	-.076	-.084	.061
stress_family	.740	.028	.037	-.057	-.047
stress_general	.737	.000	.140	-.077	-.049
suicidal intention	-.515	-.047	.049	-.043	.042
school satisfaction	.117	-.051	-.087	-.011	.775
household income	.016	-.025	.012	.815	.056
frequency of drink	-.025	.615	-.070	.000	-.198

선택된 5개의 요인을 해석하기 위하여 회전된 성분행렬을 살펴보았다. 이 성분행렬은 회전된 요인 부하량(factor loading)을 보여주는데 요인 부하량은 각 변수와 요인 사이의 상관관계 정도를 나타내므로, 각 변수는 요인 부하량이 가장 높은 요인에 속하게 된다. 사실 요인 부하량의 제곱값은 결정계수를 의미하므로, 요인 부하량은 요인이 해당변수를 설명해 주는 정도를 의미한다. 일반적으로 요인 부하량의 절대값이 0.4 이상이면 유의한 변수로 간주하고 0.5를 넘으면 아주 중요한 변수라고 할 수 있다. 그리고 요인의 회전을 하는 이유는 변수의 설명축인 요인들은 회전시킴으로써 요인의 해석을 돕는 것이다. 여러 가지 요인 회전방법이 있으나 여기서는 가장 많이 사용하는 베리맥스 직각회전방법을 사용하였다.

Table 4.3의 성분행렬을 보면 요인 1은 5개의 변수(주관적만족감, 건강평가, 스트레스정도, 가정, 전반적인 스트레스정도, 자살충동)에 높은 값을 가지고 있고, 요인 2는 3개의 변수(성별, 흡연여부, 음주횟수), 요인 3은 4개의 변수(건강-아침식사, 건강-적정수면, 건강-규칙적운동, 건강-정기건강검진), 요인

4는 2개의 변수(점유형태, 가구소득), 요인 5는 1개의 변수(학교생활만족도)에 대하여 높은 값을 가진다. 각 요인에 해당하는 변수들은 공통점을 지니고 있는 것으로 보인다. 이 공통점을 고려하여 각 요인을 명명하면, 요인 1은 정신건강(mental health) 요인, 요인 2는 성별(sex) 요인, 요인 3은 생활건강(healthy life) 요인, 요인 4는 소득(income) 요인, 요인 5는 학교생활만족도(school satisfaction) 요인으로 부를 수 있겠다.

이상의 5개 요인을 설명변수로 하여 계층적 프로빗 모형을 적합한다. 단, 정신건강 요인은 숫자가 높을수록 낮은 스트레스를 나타내므로 편의상 부호를 바꾸어 숫자가 낮을수록 스트레스가 높은 것을 나타내도록 하였다. 이 계층적 프로빗 모형에 대하여 피어슨 적합도 검정을 해보면 유의확률이 0.3924로 모형이 잘 적합 한다고 할 수 있다.

4.3. 베이지안 분석 결과

모수 β_i 에 대하여 $\beta_i \sim N(\beta, \Sigma)$ 의 사전분포를 가정하였다. β_i 에 대한 정규성을 알아보기 위하여 β_i 의 각 원소에 대하여 지역별로 구한 16개의 최우추정치(MLE), $\hat{\beta}_i^{MLE}$ 의 분포를 히스토그램을 통하여 살펴본 결과 정규성 가정이 크게 어긋나지 않음을 알 수 있었다.

다음, β_i 의 원소들 간에 상관관계를 알아보기 위하여 $\hat{\beta}_i^{MLE}$ 의 원소들 간의 공분산행렬을 구해본 결과 대각행렬이 아니므로 Σ 의 분포를 $IW(\alpha_0, \Sigma_0)$ 로 가정하는 것이 합리적임을 알 수 있었다. 참고로, $\hat{\beta}_i^{MLE}$ 의 원소들 간의 공분산행렬이 대각행렬이라면 원소들 간에 상관성이 없으므로 $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ 로 놓고 각 σ_m^2 , $m = 1, \dots, p$ 에 대하여 역감마 분포를 가정하는 것이 더 효율적일 것이다.

모형에 대한 베이지안 추론을 위하여 3절에서 제시한 깃스표본기법을 이용하여 모수 $(\beta_i, z_{ij}, \beta, \Sigma)$ 의 사후표본을 추출하고 이를 이용하여 모수의 사후 추정치를 구하였다. 깃스표본 알고리즘을 구현하기 위한 초기치 설정은 다음과 같다.

β 와 Σ 에 대한 반공액 사전분포 식 (2.3)에서 $k_0 = 1$, $\alpha_0 = 9$ 로 선택하였다. β 의 사전평균 β_0 의 초기치로는 지역별로 프로빗 모형을 적합하여 얻은 계수의 최우 추정치 $\hat{\beta}_i^{MLE}$ 의 평균 $\bar{\beta}$ 를 β_0 의 초기치로 설정하였다. 다음, $\Sigma \sim IW(\alpha_0, \Sigma_0)$ 의 사전평균 $\Sigma_0/(\alpha_0 - p - 1)$ 가

$$\frac{1}{K-1} \sum_i (\hat{\beta}_i^{MLE} - \bar{\beta}) (\hat{\beta}_i^{MLE} - \bar{\beta})'$$

와 일치하도록 Σ_0 의 초기치를 정하였다.

Table 4.4는 5000번의 초기반복 이후 10000개의 표본들을 사용한 모형의 추정값과 표준오차(괄호안)를 보여 주고 있다. 95% 신뢰수준으로 0과 유의하게 다른 계수값은 진하게 표시하였다. 지역별로 유의한 변수를 살펴보면 서울, 광주, 경기, 강원, 경북지역은 정신건강과 소득이 유의하다. 소득에 대한 사후평균값은 (0.3920, 0.4588, 0.2854, 0.3858, 0.3871)로 양의 효과를 보이며 이는 소득이 높을수록 부모님도움으로 등록금을 내는 학생일 가능성이 크다는 것을 의미한다. 정신건강에 대한 사후평균값은 (-0.1739, -0.3793, -0.3942, -0.5409, -0.2402)로 음의 효과를 보이며 이는 부모님도움으로 등록금을 내는 학생일수록 정신적 스트레스가 적다는 것을 의미한다.

인천지역은 요인 1인 정신건강과 요인 2인 성별이 유의하다. 정신건강에 대한 사후평균값은 0.2612로 양의 효과를 보인다. 성별에 대한 사후평균값은 0.2395로 양의 효과를 보이며 이는 남자가 여자보다 부모님도움으로 등록금을 내는 가능성이 크다는 것을 의미한다.

부산, 울산 지역은 요인 4인 소득만 유의하다. 소득에 대한 사후평균값은 (0.2872, 0.3359)로 양의 효과

Table 4.4. Posterior mean and SE of $\hat{\beta}_{ip}$.

	Seoul	Pusan	Daegu	Inchon	Kwangju	Daejeon	Ulsan
(Intercept)	0.5163 (0.0758)	0.2926 (0.0974)	0.3999 (0.0996)	0.4430 (0.1072)	0.4792 (0.1096)	0.5020 (0.1022)	0.5565 (0.1475)
mental health	0.1739 (0.0753)	0.0409 (0.0977)	0.2201 (0.1021)	0.2612 (0.1126)	0.3793 (0.1037)	0.2321 (0.0979)	0.1878 (0.1334)
sex	0.0168 (0.0755)	0.0234 (0.0959)	0.1385 (0.1020)	0.2395 (0.1134)	0.0270 (0.1033)	-0.1487 (0.1176)	-0.0162 (0.1566)
healthy life	-0.1270 (0.0770)	0.0347 (0.0952)	-0.0070 (0.0994)	-0.1852 (0.1028)	-0.0517 (0.1114)	0.0321 (0.1058)	-0.0773 (0.1563)
income	0.3920 (0.0694)	0.2872 (0.1168)	0.1441 (0.1088)	0.2230 (0.1104)	0.4588 (0.1191)	0.1182 (0.0980)	0.3359 (0.1475)
school satisfaction	-0.1441 (0.0775)	-0.0976 (0.1096)	-0.0456 (0.0914)	0.0138 (0.1070)	0.0129 (0.1154)	-0.0732 (0.1001)	-0.2998 (0.1619)
	Kyungki	Kangwon	Chungbuk	Chungnam	Chonbuk	Chonnam	Kyungbuk
(Intercept)	0.3961 (0.0886)	0.1261 (0.1366)	0.5911 (0.1329)	0.5758 (0.1371)	0.5778 (0.1284)	0.6219 (0.1505)	0.7319 (0.1469)
mental health	0.3942 (0.1015)	0.5409 (0.1433)	0.2636 (0.1432)	0.2630 (0.1345)	0.1879 (0.1427)	0.2304 (0.1740)	0.2402 (0.1222)
sex	-0.0164 (0.0940)	0.0444 (0.1454)	0.1022 (0.1298)	0.0671 (0.1278)	-0.0463 (0.1160)	0.1559 (0.1511)	0.0167 (0.1241)
healthy life	0.0644 (0.0827)	-0.0662 (0.1423)	-0.0976 (0.1376)	0.0311 (0.1233)	-0.0037 (0.1404)	-0.2123 (0.1622)	-0.0423 (0.1434)
income	0.2854 (0.0950)	0.3858 (0.1544)	0.0458 (0.1185)	0.1237 (0.1467)	0.0152 (0.1294)	0.1449 (0.1660)	0.3871 (0.1205)
school satisfaction	-0.0228 (0.0892)	0.0079 (0.1303)	-0.1020 (0.1160)	-0.1214 (0.1340)	0.2874 (0.1269)	0.0162 (0.1414)	-0.0599 (0.1500)
	Kyungnam	Jeju	mean($\hat{\beta}_p$)	$\hat{\beta}_p^{MLE}$			
(Intercept)	0.4599 (0.1292)	0.1285 (0.1836)	0.4624 (0.1667)	0.4562 (0.2585)			
mental health	0.3453 (0.1424)	0.1597 (0.1668)	0.2575 (0.1149)	0.2513 (0.1570)			
sex	0.0372 (0.1342)	0.0956 (0.1489)	0.0460 (0.0904)	0.0485 (0.1124)			
healthy life	0.1622 (0.1423)	0.0036 (0.1801)	-0.0339 (0.0942)	-0.0396 (0.1185)			
income	0.1329 (0.1689)	0.2509 (0.1675)	0.2332 (0.1350)	0.2278 (0.1808)			
school satisfaction	-0.1504 (0.1425)	0.0539 (0.1764)	-0.0453 (0.1237)	-0.0495 (0.1656)			

를 보인다.

대구, 대전, 경남지역은 요인 1인 정신건강 요인만 유의하다. 정신건강에 대한 사후평균값은 (0.2201, 0.2321, 0.3453)로 양의 효과를 보인다.

충북, 전남, 제주지역은 유의한 요인이 하나도 없었다.

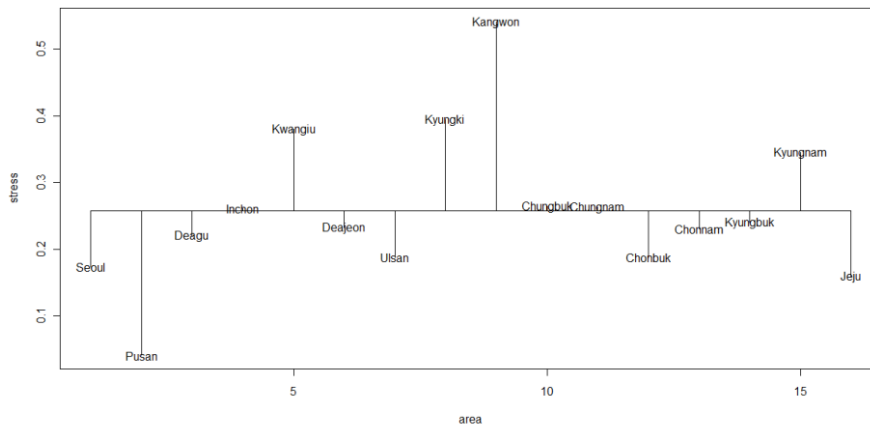


Figure 4.2. Estimated coefficients of regional mental health.

전북지역은 학교생활만족도만 유의하다. 학교생활만족에 대한 사후평균값은 0.2874로 양의 효과를 보이며 이는 부모님도움으로 등록금을 내는 학생일수록 학교생활만족도점수가 높다는 것을 의미한다.

이상의 결과를 종합해 보면 다음과 같다. 대다수의 지역에서 정신건강과 소득이 유의한 요인으로 나타났다. 소득요인에는 주택점유 형태와 가구소득 변수가 포함되는데, 이는 부모님의 경제적 능력을 나타내는 지표로서 부모님의 경제적 능력이 높을수록 부모님이 등록금의 대부분을 책임지는 것으로 나타났다. 또한 부모님의 경제적 지원이 있는 경우 학생이 느끼는 정신적 스트레스가 한결 덜함을 알 수 있었다. 이는 우리가 일반적으로 예상할 수 있는 것과 자료의 분석결과가 다르지 않음을 보여주며 경제적 문제가 대학생들에게 주관적 만족감, 가정생활에서의 스트레스, 전반적인 스트레스, 자살충동과 관련있는 정신적 스트레스에 많은 영향을 준다는 것을 보여주는 결과이다.

반면, 성별의 경우 인천지역을 제외한 모든 지역에서 유의한 관련이 없는 것으로 분석결과가 나타나 부모님의 등록금 조달에 있어서 자녀의 성별에는 차이는 없는 것으로 보인다.

또한 아침식사여부, 적정수면, 규칙적 운동, 정기건강검진 등의 생활건강 지수는 모든 지역에서 유의하지 않게 나타나 등록금에 대한 부담이 정신적 건강에는 영향을 미치지만 다행히 신체적 건강에는 의미 있는 영향을 미치지 않음을 알 수 있었다.

마지막으로, 학교생활 만족도는 전북지역을 제외한 나머지 지역에서 유의한 관련성이 없는 것으로 나타나 전반적으로 등록금에 대한 부담이 학교에 대한 인식에 영향을 주지는 않는 것으로 나타났다.

다음으로 지역별 차이를 알아보기 위하여 정신건강과 소득에 대한 지역별 회귀계수 추정치를 Figure 4.2와 Figure 4.3에 그려보았다. 참고로, 정신건강에 대한 회귀계수는 음의 값을 가지는데 시각적 편의를 위하여 절댓값을 그림에 표시하였다. 중앙의 가로선은 지역별 회귀계수 추정치의 평균을 나타낸다. 먼저 지역별 정신건강 회귀계수를 살펴보면, 강원 지역은 부모님의 등록금 지원과 정신적 스트레스의 관련성이 다른 지역에 비하여 월등히 높음을 알 수 있다. 다음으로 광주, 경기, 경남 지역이 그 뒤를 따르고 있다. 반면, 부산 지역은 부모님의 등록금 지원과 정신적 스트레스의 관련성이 다른 지역에 비하여 월등히 낮다. 대학생들이 아르바이트 등 자신의 힘으로 등록금을 충당할 기회가 상대적으로 작은 강원도 학생들이 부모님으로부터 지원이 없을 경우 정신적 스트레스를 상대적으로 심하게 받는 것이 아닌가 추측해볼 수 있다.

지역별 소득에 대한 회귀계수를 살펴보면, 광주 지역의 경우 부모님의 등록금 지원과 소득의 관련성이

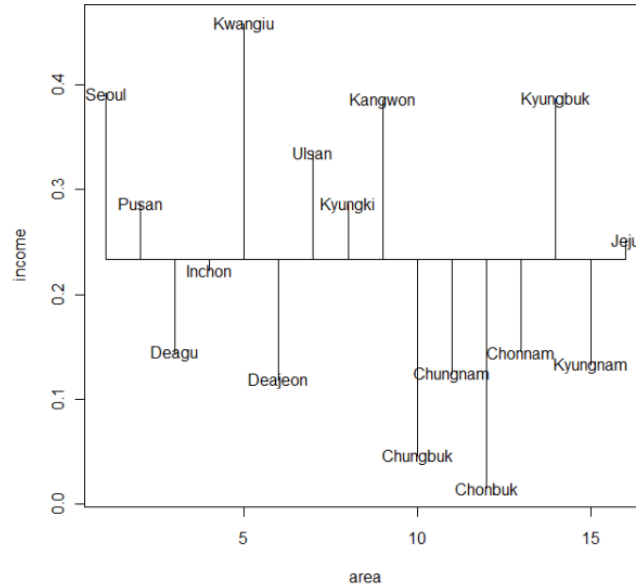


Figure 4.3. Estimated coefficients of regional income.

다른 지역에 비하여 월등히 높고 그 뒤를 서울, 강원, 경북 지역이 따르고 있다. 반면, 충북, 전북 지역은 부모님의 등록금 지원과 소득의 관련성이 다른 지역에 비하여 상대적으로 낮다. 소득의 경우 Figure 4.3에서 평균치에 근접한 제주도와 평균치 보다 작은 회귀계수 값을 갖는 모든 지역에서 소득에 대한 회귀계수가 유의하지 않았다. 즉, Figure 4.3에서 중앙 가로선 가까이 있거나 아래에 있는 지역에서는 소득이 많은 부모님이 소득이 적은 부모님 보다 더 등록금을 지원하는 경향을 보이지는 않는다는 것이다. 이 지역들은 대구와 대전을 제외하고 대도시가 아닌 행정도에 해당된다. 소득 요인에는 두 가지 요소, 가구 소득과 주택의 점유형태가 주로 포함되는데, 부동산인 주택 점유형태가 부모님의 등록금 지원과 관련성이 적고 대도시가 아닌 행정도에서는 대부분 자가 주택을 소유하고 있어 실질 가구소득이 적음에도 불구하고 주택점유형태에서 높은 점수를 받음으로써 소득이 높게 나타나기 때문이 아닌가 의심된다. 이에 대한 자세한 분석은 추후 수행하여 보기로 한다.

5. 결론

본 논문에서는 통계청에서 수집한 ‘2010년 사회조사’ 자료를 기반으로 대학 등록금에 대한 경제적 부담과 관련 있는 요인들에 대한 분석을 수행하였다. 등록금액의 70% 이상을 부모님으로부터 도움을 받는지 여부에 대한 이항변수에 대하여 지역을 계층으로 하는 계층적 프로빗 모형을 적용하여 베이지안 추론을 수행하였다. 지역과 등록금 조달 변수를 제외한 16개의 변수에 대한 요인분석 결과 얻어진 5개의 요인을 설명변수로 사용하여 다중 공선성 문제를 해결하고 보다 경제적인 간단한 모형을 제시하였다. 또한 베이지안 사후 추정치를 용이하게 구하기 위하여 적절한 잠재변수를 도입하고 마코브 체인 몬테 칼로 기법을 사용하였다.

자료의 분석 결과, 많은 지역에서 소득과 정신적 스트레스 요인이 부모님의 등록금에 대한 경제적 지원과 유의 관련이 있음을 보여주었다. 소득이 높은 부모일수록 자녀의 대학 등록금을 지원하며 부모로부

더 경제적 지원을 받는 학생일수록 정신적 스트레스를 덜 받는 것으로 나타났다. 따라서 부모의 소득이 대학생 자녀의 정신건강에 유의한 영향을 미침을 보여 주었다. 반면에, 성별, 생활건강, 학교 만족도는 대부분의 지역에서 부모님의 등록금 지원과 유의한 관련이 없었다.

지역별 차이를 보면, 강원도 지역에서 정신적 스트레스와 부모님의 지원 간에 높은 관련성을 보였다. 대학생의 아르바이트 등 학생의 경제활동 기회가 적은 강원도의 특성이 반영된 것이 아닌가 추측된다. 소득과 부모님의 지원 간의 관련성에 있어서는 대도시가 행정도에 비하여 관련성이 높게 나타났고 대부분의 행정도에서는 관련성이 약하거나 유의하지 않았다.

논문에 제시된 결과는 공액 사전분포를 사용한 결과이다. 사전분포의 영향을 비교하기 위하여 무정보 사전분포를 이용한 분석을 수행한 결과 거의 비슷한 결과를 얻어 본 논문의 결과는 사전분포에 민감하지 않음을 알 수 있었다.

References

- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, **88**, 669-679.
- Bahn, S. J. (2008). *High Education Economy*, Jipmoondang, Seoul.
- Cheon, S. Y. (1997). Index to Korean education, *Education Development Summer*, **108**, 55-56.
- Choi, S. Y. (2006). Analysis report of financial status for college tuition in Korea, '06 National Audit Policies Kit.
- IMF (2008). World Economic Outlook Database: www.imf.org/external/ns/cs.aspx?id=29.
- OECD (2008). Education at a Glance 2008: OECD Indicator, www.oecd.org/publishing/corrigenda.
- Statistics Korea (2009). KOSIS National Statistics Portal, www.kosis.kr.

계층적 베이지스 모형을 이용한 대학등록금에 대한 부모님의 경제적 지원 영향 분석

오만숙^{a,1} · 오현숙^b · 오민정^a

^a이화여자대학교 통계학과, ^b가천대학교 응용통계학과

(2013년 1월 8일 접수, 2013년 2월 21일 수정, 2013년 3월 5일 채택)

요약

최근 한국 사회에서 경제적, 정치적, 사회적 이슈가 되고 있는 대학 등록금의 경제적 부담에 영향을 미치는 요인들에 대한 분석을 위하여 통계청에서 실시한 '2010년도 사회조사'에서 수집된 자료를 기반으로 지역을 계층으로 하는 베이지안 계층모형을 이용한 분석을 수행하였다. 등록금의 70% 이상을 부모님이 지원하는가에 대한 이항 반응변수에 대하여 계층적 프로빗 모형을 설정한 후 설명변수들에 대한 요인분석을 실시하여 설명변수를 압축하고 마코브체인 몬테칼로 기법을 적용하여 모수를 추정하였다. 자료의 분석 결과, 많은 지역에서 소득과 정신적 스트레스 요인이 부모님의 등록금에 대한 경제적 지원과 유의한 관련이 있음을 보여주었다. 소득이 높은 부모일수록 자녀의 대학 등록금을 지원하며 부모로부터 경제적 지원을 받는 학생일수록 정신적 스트레스를 덜 받는 것으로 나타나 부모의 소득이 자녀의 정신건강에 유의한 영향을 미침을 보여 주었다. 반면에, 성별, 생활건강, 학교 만족도는 대부분의 지역에서 부모님의 등록금 지원과 유의한 관련이 없었다. 스트레스 또는 소득과 부모님의 지원에 대한 지역별 차이를 보면, 강원도 지역 학생들이 부모님의 지원이 낮을 경우 가장 정신적 스트레스를 많이 받는 것으로 나타났으며 소득이 많을수록 부모님의 지원 가능성이 높아지는 경향은 지방 행정도에 비하여 대도시에서 더 뚜렷하게 나타남을 알 수 있었다.

주요용어: 계층모형, 프로빗모형, 요인분석, 마코브체인 몬테칼로.

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행한 기초연구사업임 (No. 2010-0010285).

¹교신저자: (120-750) 서울시 서대문구 대현동 21, 이화여자대학교 통계학과, 교수. E-mail: msoh@ewha.ac.kr