

## 초·중·고생의 사교육비 지출에 대한 통계 분석

오만숙<sup>1</sup> · 김진희<sup>2</sup>

<sup>1</sup>이화여자대학교 통계학과, <sup>2</sup>이화여자대학교 통계학과

(2010년 8월 접수, 2010년 12월 채택)

### 요약

자녀의 사교육비 지출은 정치, 경제, 사회 등 국민 생활 전반에 걸쳐 막대한 영향을 미치고 그 부작용이 심각하여 한국사회의 중요한 이슈가 되고 있다. 본 논문에서는 통계청에서 수집한 2008년도 사교육비 실태조사 자료 중 일반사교육비 지출과 방과후 프로그램 참가비의 두변수에 대하여 지역, 초·중·고 등 학교급 구분, 가구소득, 학생 성적, 성별, 사교육 참가시간 등의 인구동태적 변수들의 영향을 알아보는 다중 선형 회귀분석을 수행 하였다.

분석결과 일반 사교육비 지출과 방과후 프로그램 참가비에 뚜렷한영향을 미치는 요소로는 지역과 학교급구분이고 나머지 변수들은 의미있는 영향력을 보여주지 못하였다. 일반 사교육비 지출에 대한 지역의 영향을 보면, 서울지역>광역시, 중소도시> 읍면지역 순으로 지출에 상당한 차이가 있음을 보여주었다. 방과후 참가비에 대한 지역의 영향을 보면 서울지역, 광역시, 중소도시> 읍면지역 의 순으로 지출이 많았는데 서울과 기타도시의 차이가 없다는 것이 일반 사교육비의 경우와 다른 특징이다.

학교급 구분에서는 일반사교육비의 경우 일반계 고교생에 대한 지출이 초·중생이나 전문계 고교생에 비하여 약 17% 지출이 더 많았다. 이는 대학입시 위주의 교육 풍토와 사교육 시장이 관련이 있음을 보여준다고 할 수 있다. 방과후 학교 참가비의 경우 일반계고, 초등학교> 전문계고> 중학교 의 순서이다. 초등학생은 사교육비의 부담을 덜기 위해서, 그리고 일반계 고등학생의 경우는 방과후 학교에서 내신에 도움을 주는 프로그램을 제공하기 때문에 추측할 수 있다.

주요용어: 사교육비, 방과후 학교 참가비, 다중 회귀분석.

### 1. 서론

사교육은 학교의 정규 교육과정을 중심으로 이루어지는 공교육에 대비되는 개념으로, 학교 안팎에서 행해지는 ‘학교에서의 정규 교육과정 이외의 다양한 교육’을 의미한다. 즉, 학부모나 학생은 학교의 정규 교육과정을 통하여 충족하지 못하는 다양한 교육적 필요나 욕구가 있으며, 그 욕구를 충족시키기 위한 목적으로 학교 안팎에서, 예컨대 방과 후 학교 또는 학원, 가정 등과 같은 곳에서 실시되는 다양한 교육 활동을 의미한다.

해마다 사교육으로 인한 지출비용은 계속해서 늘어나 우리나라 가계지출에 큰 비중을 차지하고 있으며 사회적 계층의 고착화, 부동산 양극화, 저출산율 등 많은 분야에서 심각한 영향을 미치고 있어 매우 중요한 사회적 이슈가 되고 있다. 때문에 정기적 또는 비정기적으로 통계청을 비롯해 한국교육개발원, 한국직업능력개발원, 한국노동연구원 등에서는 그 실태에 대해서 정기적으로 조사하고 있다.

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국학술진흥재단(또는 한국과학재단)의 지원을 받아 수행된 연구임(No. 2009-0071093)

<sup>1</sup>교신저자: (120-750) 서울시 서대문구 대현동 11-1, 이화여자대학교 통계학과, 교수. E-mail: ms0h@ewha.ac.kr

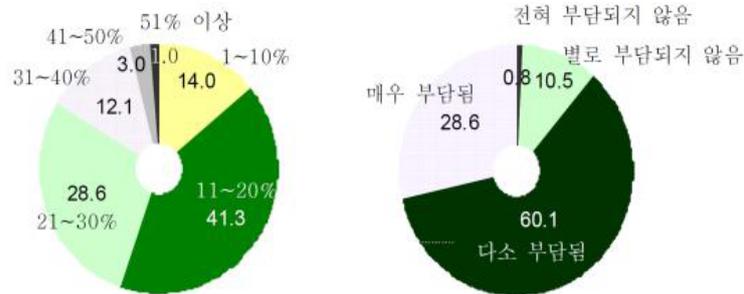


그림 1.1. 가구 지출 중 사교육비가 차지하는비중 및 사교육비 가계 부담 여부(N = 1,500)

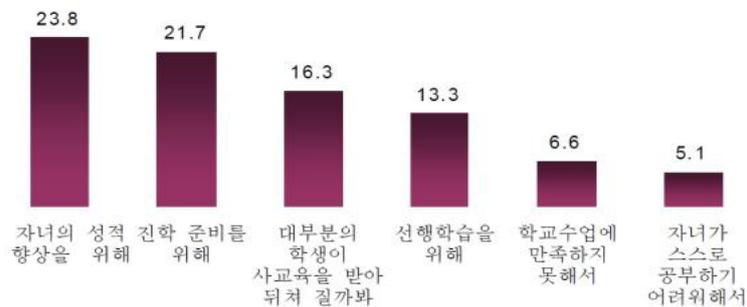


그림 1.2. 한국교육개발원에서 발표한 사교육 이유(단위: %)

2009년 교육과학기술부에서 조사한 ‘사교육비 및 사교육 의식 파악을 위한 학부모 여론 조사결과’ 그림 1.1에 따르면 자녀들의 교육을 위해 지출하는 사교육비의 가계 부담여부에 대해 10명 중 9명가량(88.7%)이 ‘부담된다’고 응답했으며, ‘매우 부담된다’는 응답도 28.6%로 나타났다(교육과학기술부, 2009). 가구 소득수준, 맞벌이 여부, 자녀의 학년을 불문하고 대부분의 계층에서 자녀의 사교육비가 부담된다는 응답이 압도적으로 높게 나타나 자녀의 사교육비 지출로 인한 가계의 부담이 상당한 것으로 보인다.

한국교육개발원에서 발표한 자료 그림 1.2에 의하면, 사교육을 하는 이유로는 ‘성적향상’이나 ‘진학준비’와 같은 이유 외에도 ‘대부분 학생이 사교육을 받아 뒤쳐질까봐’라는 응답이 비교적 높게 나타나고 있어 심리적 요인이 일부 작용하고 있는 것으로 조사되었다.

통계청 조사(통계청, 2009)에 의하면, 매년 계속해서 가구 당 사교육비 지출 비용이 증가하고 있다. 그림 1.3에 나타난 바와 같이 2009년 전국 초·중·고교 학생의 사교육비 전체 규모는 21조 6천억원으로 추정되며 2008년(20조 9천억원)보다 7000억원인 3.4% 증가하였고, 학생 1인당 월평균 사교육비는 24만 2천원으로 2008년(23만 3천원)보다 3.9% 증가하였다.

통계청에서는 사교육비 경감대책 및 공교육 내실화 등 교육정책수립에 활용할 수 있는 사교육비 실태를 체계적으로 조사하여 공신력있는 통계를 정기적으로 작성하고 제공하기 위해 2007년부터 매년 사교육비 실태조사를 수행하고 있다. 이 조사는 매년 2회씩 수행되고 있는데 매년 6월에 시행되는 제 1차 조사는 그 해 3월, 4월, 5월의 사교육비 지출을 조사하며, 매년 10월에 시행되는 제 2차 조사는 그 해 7월, 8월, 9월 사교육비 지출을 조사한다.

조사의 신뢰도 제고를 위해 전년도 조사에서 선정된 표본을 유지 하되, 지역규모별(서울, 광역시, 중소

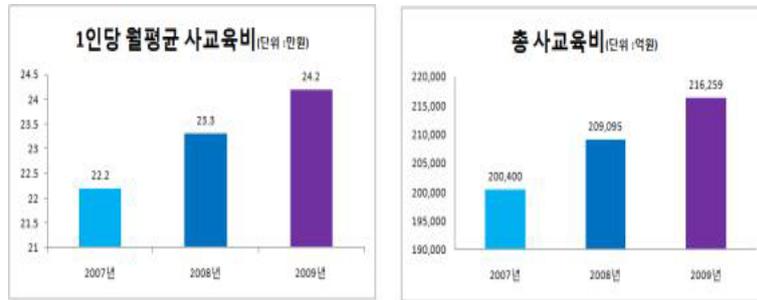


그림 1.3. 사교육비 규모 및 학생 1인당 월평균 사교육비

도시, 읍·면), 학교급별(초, 중, 일반고, 전문고)로 층화 후 조사대상 학교를 할당하고 조사대상 학교의 각 학년별로 1개 학급을 표본으로 추출하여 조사를 수행한다. 조사는 크게 사교육비와 방과 후 학교비용, EBS 교재비 및 어학연수비로 분리하여 수행한다. 사교육비는 주요 과목별로 학원비, 개인 및 그룹 과외비, 학습지 인터넷 및 통신 강의(EBS 제외) 과외비(교재비포함)와 일반교과 및 논술관련 사교육비, 예체능 및 취미·교양 관련 사교육비 그리고 취업 목적 관련 사교육비등으로 나누어 조사한다. 방과 후 학교 활동비, EBS 관련 교육비는 사교육비 경감효과의 정확한 분석을 위해 학교 밖에서 이루어지는 사교육비와 분리하여 조사한다.

그러나 통계청 조사 자료에 대한 분석 결과의 발표 내용을 보면, 사교육비 지출에 영향을 미치는 여러 변수들을 동시에 고려하기 보다는 관심있는 변수들의 단순 기술통계량의 비교 또는 두 변수들의 일대일 관계 등에 국한되고 있다. 예를 들면, 사교육비를 지출하지 않는 학생은 2007년 23%에서 2008년 24.9%로 증가하였으며, 과목별 사교육비는 2007년 대비 영어 11.8%, 수학 8.8% 증가한 반면, 국영수를 제외한 기타과목은 10% 감소하였다는 것 등이다. 많은 비용과 노력을 들여 수집한 자료에 대하여 단변량적 분석에 국한하는 것이 아니라 여러 변수들을 동시에 고려하는 다변량 분석을 이용하여 관측된 인구동태적 변수들이 사교육비 지출에 어떤 영향을 미치는지 고려하는 것이 본 연구의 목적이다.

본 연구에서는 넓은 의미의 사교육비 지출 중 ‘학교 밖의 일반 교과와 취미 교양을 위한 사교육비 지출’과 ‘학교 내의 방과 후 학교 프로그램을 위한 교육비’를 구분하여 분석하고자 한다. 이들 두 종류의 사교육비 지출액에 영향을 미치는 인구동태적 요인을 분석하여 향후 사교육비 경감대책 및 공교육 정책수립에 필요한 기초자료를 제공하고, 궁극적으로는 국민 가계 복지 증진 및 가계소비 정책에 기여하고자 한다.

## 2. 자료의 정리

### 2.1. 자료의 수집

본 연구의 분석 대상 자료는 통계청에서 사교육비 지출 규모와 사교육비 지출에 영향을 미치는 인구동태적 영향을 파악하여 사교육 경감대책 및 공교육 내실화 등 교육정책 수립에 활용하고, 사교육비 실태를 체계적으로 조사하여 공신력 있는 통계를 정기적으로 작성, 제공하기 위하여 2007년부터 통계청에서 매년 조사해온 사교육비실태조사 자료 중 2008년의 2차 자료이다. 통계청의 사교육비 실태 조사는 년 2회로 조사대상 기간으로는 1차와 2차로 나누어 실시가 되는데 제 1차 조사는 2008년 3월, 4월, 5월의 사교육비 지출을 조사했으며, 제 2차 조사로는 2008년 7월, 8월, 9월의 사교육비 지출을 조사하였다. 조사 실시 기간으로 제 1차 조사는 2008년 6월 2일~6월 13일, 제 2차 조사로는 2008년 10월 6일~10월

표 2.1. 추출틀 작성시 제외 조건

도서지역 및 특수 지역의 학교의 학생
학년 당 학급 수가 0인 학년이 있는 학교의 학생
전교에 6미만의 학급이 있는 초등학교의 학생
전교에 3미만의 학급이 있는 중·고등학교의 학생
학급당 평균 학생 수가 10명 미만인 학교의 학생

표 2.2. 층화과정

1단계	16개 시 도를 구분하여 층화
2단계	16개 시 도 내 학교를 도시지역과 읍·면지역으로 층화
3단계	학교급 별로 구분해 층화

17일에 실시되었다.

조사에 사용된 표본은 2007년 3월 기준 교육통계정보센터의 학교DB상의 학교를 사용하였다. 단, 조사의 현실성을 고려하여 표 2.1에 있는 학생들은 추출틀 작성 시 제외시켰다.

표본 추출을 위해서는 층화 확률 비례 추출 방법을 사용하였다. 즉, 사교육비에 영향을 주는 다양한 요인 중 지역과 학교급(초·중·고) 등이 주요변수에 해당되어 이들 변수를 층화변수로 사용하며, 층화과정은 표 2.2와 같이 크게 3단계로 정리된다.

조사 대상은 전국 초·중·고 273개 학교를 대상으로 하였다. 이 학교의 학부모 66,999명을 대상으로 실시되었으며, 조사지역으로는 서울, 광역시, 중소도시, 읍·면지역으로 이 지역에 거주 하고 있는 초등학교, 중학교, 일반계 고등학교, 전문계 고등학교 학생들에게 설문지(Questionnaire)를 나누어 주어 이를 학부모가 직접 작성하도록 하는 설문조사(Survey)를 이용 하였다 (통계청, 2008). 본 연구에 사용된 자료의 해당 조사기간은 2008년 7월부터 9월까지 이다.

## 2.2. 변수의 선정

원자료에 나타나 있는 변수 51개에 대해서 정리를 해보면 표 2.3과 같다. 본 연구는 사교육비 지출 규모와 사교육비에 영향을 미치는 인구동태적 영향을 파악하는 것이 목적이다. 여기서 사교육비란 일반교과의 연간 지출비용과 예체능·취미·교양의 연간 지출비용의 합으로 조사 대상이 초등학생, 중학생, 고등학생이므로 표 2.3에서 취업과 관련된 변수인 ‘취업관련 사교육 시간’과 ‘취업관련 금액’은 제외된 것으로 정의한다. 또한 ‘사교육 참여시간’ 변수는 일반교과 사교육시간과 예체능 사교육시간 그리고 취업관련 사교육시간 변수를 합한 것이기 때문에 여기에서 ‘취업관련 사교육시간’을 뺀 변수를 ‘사교육 참여시간’ 변수로 재 정의하여 사용하기로 한다.

각 변수 별로 빈도분석을 해본 결과 범주형으로 되어있는 변수인 일반교과 사교육 목적 구분(1순위), 일반교과 사교육 목적 구분(2순위), 예체능 사교육 목적 구분 1, 예체능 사교육 목적 구분 2 는 응답에 0이 너무 많아 분석대상에서 제거하기로 한다.

부의 연령과 부의 교육정도, 그리고 모의 연령과 모의 교육정도, 부모의 경제활동 변수에는 4% 이내의 결측값이 존재하는데 본 논문에서는 이들 결측 값을 각 변수의 범주 중 최빈값으로 대체하였다. 또한 단위가 금액으로 정의되어 있는 변수는 모두 만원 단위로 설정하였다.

이상의 자료 정리 과정 후에, 우리는 두 개의 관심있는 변수에 대하여 다른 변수들의 영향을 파악하는 분석을 수행하고자 한다. 첫째 관심변수는 일반교과와 예체능·취미·교양을 위한 사교육비 지출 금액을 합한 ‘사교육비’ 변수이다. 일반교과에 지출된 사교육비는 과목별로 국어, 영어, 수학, 사회·과학, 제

표 2.3. 원자료에 포함되어 있는 총변수

지역 구분	어학연수비 : 국내연수
학교급 구분	해외연수
일반교과 사교육 목적 구분(1순위)	일반교과 : 국어
일반교과 사교육 목적 구분(2순위)	영어
일반교과 사교육 시간	수학
예체능 사교육 목적 구분1	사회, 과학
예체능 사교육 목적 구분2	제 2외국어, 한문, 컴퓨터
예체능 사교육 시간	논술
취업관련 사교육 시간	일반교과 : 개인과외
사교육 참여시간	그룹과외
학생 성별	학원수강
학생 성적 구분	방문 학습지
부의 연령	유료 인터넷 및 통신강좌
부의 교육정도	예체능·취미·교양 : 음악
모의 연령	미술
모의 교육정도	체육
부모의 경제활동	취미, 교양
가구소득	예체능·취미·교양 : 개인과외
방과 후 학교 참여(유상+무상)	그룹과외
방과 후	학원수강
초등 방과 후 보육 프로그램	방문수업
특기적성 프로그램	교과 프로그램
기타	EBS 교재비
취업관련 금액	

2외국어·한문·컴퓨터, 논술 에 지출되는 비용의 합으로 주어지며 또한 사교육 형태별로 일반교과 개인과외, 일반교과 그룹과외, 일반교과 학원수강, 일반교과 방문학습지, 일반교과 유료인터넷 및 통신강좌 에 지출되는 비용의 합으로 주어진다. 예체능·취미·교양은 마찬가지로 표 2.3에서와 같이 과목별로 그리고 사교육 형태별로 조사되어 있다. 우리는 여기에서 일반교과에 지출되는 사교육비와 예체능·취미·교양에 지출되는 사교육비의 총합으로 정의되는 ‘사교육비’ 변수에 대하여 다중 회귀분석을 수행하고자 한다.

두 번째 관심변수는 방과후 학교 활동을 위해 지출하는 비용을 나타내는 ‘방과후’ 변수이다. 방과 후 학교는 기존의 공교육 시스템 내에서 이루어지는 보충 학습으로 상업적인 사교육에 대응하기 위해 정부 지원으로 운영하는 프로그램이므로, 이를 따로 분리하여 ‘방과후’에 대하여 다중회귀분석을 수행하여 방과 후 학교 프로그램에 지출하는 비용에 대한 인구동태적 변수들의 영향을 파악하고 이를 사교육비에 대한 다중회귀분석 결과와 비교해 본다면 최근 방과 후 학교 프로그램을 강화하려는 교육부의 정책 수립, 그리고 학부모들의 방과 후 학교 프로그램 선택에 도움이 될 것이다.

넓은 의미의 사교육비 지출에 포함되는 EBS 교재비와 국내외 어학연수비를 보면, EBS 교재비는 전체 66,999명의 조사 대상 중 81%가 교재비 지출이 없다고 응답하였고, 국내연수비와 해외연수비를 합한 변수인 어학연수비는 조사대상 중 99.3%가 어학연수비 지출이 없다고 응답하였으므로 이 두 변수는 분석 대상에서 제외하도록 한다.

설명변수들 중 부의 연령과 모의 연령은 상관계수가 0.58876로 높은 상관관계가 있음을 알 수 있다. 이 두 변수들을 모두 모형에 포함시키면 다중공선성 등의 중복으로 인한 문제를 야기할 수 있으므로, 두 변수 대신 부모의 평균 연령을 새로운 설명변수로 사용하기로 한다. 마찬가지로, 부의 교육정도와 모의 교

표 2.4. 설명변수

변수명	관측치
사교육 시간	일반교과와 예체능을 위한 사교육시간 (단위: 시간)
지역구분	서울(1), 광역시(2), 중소도시(3), 읍면지역(4)
학교급구분	초등학교(1), 중학교(2), 일반계 고등학교(3), 전문계 고등학교(4)
학생성별	남자(1), 여자(2)
학생성적구분	상위10%이내(1), 11~30%(2), 31~60%(3), 61~80%(4), 하위20%이내(5)
부모의 평균연령	20 30대(1), 40대(2), 50대 이상(3)
부모의 평균교육정도	중졸이하(1), 고졸(2), 대졸(3), 대학원졸(4)
부모의 경제활동	아버지(외별이)(1), 어머니(외별이)(2), 맞벌이(3), 경제활동 안함(4)
가구소득	100만원 미만(1), 100~199만원(2), 200~299만원(3), 300~399만원(4), 400~499만원(5), 500~599만원(6), 600~699만원(7), 700만원 이상(8)
방과후 학교참여	방과 후 학교 미참여(0), 방과 후 학교 참여(1)

표 2.5. 범주형 변수의 연속형 변수로의 변환

변수명	변환 전	변환 후
학생성적구분	상위 10% 이내(1)	95
	11~30%(2)	80
	31~60%(3)	55
	61~80%(4)	30
	하위 20% 이내(5)	15
부모의 평균연령	20 ~30대(1)	25
	40대(2)	45
	50대 이상(3)	55
부모의 평균교육정도	중졸 이하(1)	7.5
	고졸(2)	12
	대졸(3)	16
	대학원졸(4)	20
가구소득	100만원 미만(1)	50
	100~199만원(2)	150
	200~299만원(3)	250
	300~399만원(4)	350
	400~499만원(5)	450
	500~599만원(6)	550
	600~699만원(7)	650
	700만원 이상(8)	750

육정도는 상관계수가 0.63165이므로 이 두 변수들을 부모의 평균 교육정도로 대체하기로 한다.

이상의 정리 과정을 통하여 최종적으로 선택된 설명변수는 표 2.4에서 주어진 10개 변수이다. 이들 설명 변수 중 학생성적구분, 부모의 평균 연령, 부모의 평균 교육정도, 가구소득의 네 변수는 범주형으로 관측되었으나 양적인 의미를 지니므로 분석의 편의를 위해서 이들을 표 2.5와 같이 연속형 변수로 변환하여 사용하였다.

### 2.3. 변수의 변환

본격적인 분석에 들어가기에 앞서 각 변수가 대략적인 정규분포를 따르는지 정규성 검정을 수행할 필요

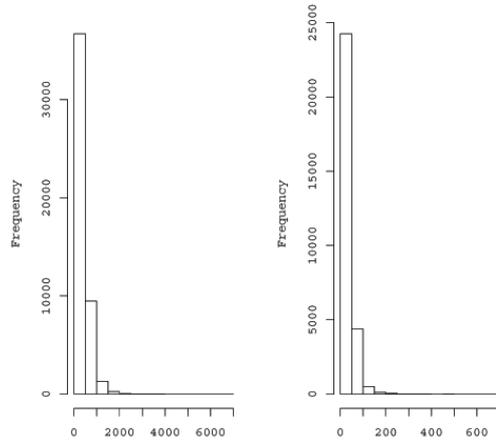


그림 2.1. 사교육비(왼쪽)와 방과후(오른쪽) 변수의 히스토그램

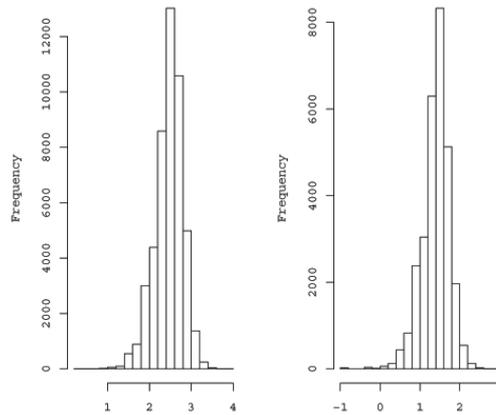


그림 2.2. log(사교육비)(왼쪽)와 log(방과 후)(오른쪽) 변수의 히스토그램

가 있다. 이를 위하여 분석에 포함되는 12개의 변수들 중 양적 변수들의 각각에 대하여 히스토그램을 그려보았다. 사교육시간, 학생성적구분, 부모의 평균 연령, 부모의 평균 교육정도, 가구소득의 양적 설명변수들은 모두 정규분포에서 크게 벗어나지 않는 것으로 관측되었다. 그러나 종속변수로 선택된 사교육비와 방과후 변수의 히스토그램을 살펴보면 그림 2.1과 같이 매우 비대칭적인 분포를 가짐을 알 수 있다.

이들 종속변수를 변환하여 정규성에 가깝도록 로그, 역수, 제곱, 루트 등의 값을 취한 변수를 다시 히스토그램을 그려본 결과, 로그 변환의 히스토그램이 그림 2.2와 같이 정규성에 가까움을 알 수 있다. 따라서 앞으로의 분석에는 사교육비와 방과후 변수에 로그변환을 한 변수를 종속 변수로 사용하도록 하겠다.

### 3. 다중선형 회귀분석

#### 3.1. 사교육비 지출액에 대한 다중 선형 회귀분석

분석에 앞서 사교육비 변수에 대한 자료를 살펴보면, 이 문항에 대하여 사교육비 지출이 없다고 답한 학

표 3.1. log(사교육비)에 대한 분산분석 결과

Source	Df	Sum of Squares	Mean Square	F-value	Pr > F
Model	16	2420.1459	151.2591	2622.40	<.0001
Error	47784	2756.1648	0.0577		
Corrected Total	47800	5176.3107			

표 3.2. log(사교육비)에 대한 회귀식의 결정계수 측정

Root MSE	0.24017
Dependent Mean	2.47019
Coeff. Variance	9.72257
$R^2$	0.4675
Adj $R^2$	0.4674

생이 19,198명으로 전체 66,999명 중 28.7%를 차지하고 있다. 따라서 전체 설문 대상 중 28.7%는 사교육비 지출이 전혀 없고 나머지 71.3%인 47,801명 만이 사교육비 지출이 있음을 의미한다. 본 분석에서는 사교육비 지출이 있을 경우 그 지출액에 영향을 미치는 요인들을 파악하는 것이 목적이므로 사교육비 지출이 있는 47,801개의 관측치만을 사용하여 분석한다. 참고로, 사교육 참여 여부에 영향을 미치는 요인들에 대한 분석은 추후 과제로 남겨 두기로 한다.

log(사교육비)에 대하여 위의 10개 설명변수의 교호작용 없이 일차항만을 가지는 다중회귀모형을 적합시켜 보고자 한다. 먼저 단계적 변수선택(Stepwise variable selection)방법을 사용하여 모형에 포함시킬 필요가 없는 변수들이 있는지 살펴 보았다. 그러나 모든 10개의 변수가 적어도 하나의 범주 이상이 유의한 변수로 선택되었기 때문에 모든 10개의 변수를 모형에 포함시켜 적합시키기로 하였다.

모형의 유의성 검정을 위해

$H_0$  : 다중 선형 회귀 모형의 계수가 모두 0이다.

$H_1$  : 다중 회귀 모형의 계수 중 적어도 하나는 0이 아니다.

가설을 검정해 보면 표 3.1과 같은 분산분석표를 얻을 수 있다. 모형의 평균 제곱합이 에러의 평균 제곱합 보다 월등히 커서  $F$  값이 매우 유의하므로 다중 선형 회귀 모형은 유의하다고 결론 지을 수 있다.

다중 회귀 모형을 적합시킨 결과 결정계수(coefficient of determination;  $R^2$ )는 0.4657 이었다(표 3.2). 결정계수는 전체 총 제곱합 중에서 회귀제곱합이 차지하는 비율을 나타내는 것으로, 총 변동 중 위의 다중회귀모형이 변동의 약 46%를 설명하는 것을 의미하므로 위 모형이 자료를 비교적 잘 적합한다고 볼 수 있다.

적합된 회귀계수를 살펴보면 표 3.3과 같다. 이들 계수 추정치를 살펴보면, 부모의 경제활동 변수에 대한 계수 외에는 모든 변수가 매우 유의함을 알 수 있다. 유의한 계수들의 추정값을 가지고 그 의미를 해석해 보면 다음과 같다.

먼저 지역을 보면, 서울이 다른 지역에 비하여 사교육비를 제일 많이 지출하고 그 다음이 광역시와 중소도시로 이 둘은 서로 비슷하며, 읍면 지역의 사교육비 지출이 가장 적다. 구체적으로, 서울 지역의 계수는 0.1334이므로, 다른 조건들이 동일하다고 할 때 서울지역에 사는 학생이 읍면지역에 사는 학생에 비해서 평균적으로 log(사교육비)가 0.13337만원 더 든다고 할 수 있다. 로그를 제거하면,  $e^{0.13337} = 1.14267$ 이므로 서울지역에 사는 학생의 사교육비 지출이 읍면지역에 사는 학생에 비하여 약 14% 정도 더 높다고 할 수 있다. 광역시와 중소도시의 계수는 서로 비슷한데 이 계수를 약 0.08로 보고 위와 같이

표 3.3. log(사교육비)에 대한 회귀식의 적합 결과

변수	자유도	계수추정치	표준오차	t-value	Pr >  t
Intercept	1	1.6954	0.0172	98.69	<.0001
지역구분(서울)	1	0.1334	0.0038	35.05	<.0001
지역구분(광역시)	1	0.0787	0.0035	22.80	<.0001
지역구분(중소도시)	1	0.0845	0.0035	24.33	<.0001
지역구분(읍면지역)		0.0000			
학교급구분(초등학교)	1	-0.0333	0.0066	-5.09	<.0001
학교급구분(중학교)	1	0.0317	0.0066	4.84	<.0001
학교급구분(일반계고교)	1	0.1604	0.0066	24.31	<.0001
학교급구분(전문계고교)		0.0000			
부모경제활동(부외별이)	1	0.0171	0.0125	1.37	0.1718
부모경제활동(모외별이)	1	0.0070	0.0133	0.52	0.5999
부모경제활동(맞별이)	1	-0.0146	0.0125	-1.17	0.2420
부모경제활동(없음)		0.0000			
사교육시간	1	0.0263	0.0002	134.11	<.0001
학생성별(여)	1	-0.0072	0.0022	-3.24	0.0012
학생성별(남)		0.0000			
학생성적구분	1	0.0009	0.0000	18.30	<.0001
부모의 평균연령	1	0.0006	0.0002	4.01	<.0001
부모의 평균교육정도	1	0.0136	0.0005	25.37	<.0001
가구소득	1	0.0004	0.0000	59.57	<.0001
방과후학교참여(유)	1	-0.0341	0.0024	-14.53	<.0001
방과후학교참여(무)		0.0000			

해석하면, 광역시 또는 중소도시에 사는 학생의 사교육비 지출이 읍면지역에 사는 학생에 비하여 약 8% 정도 더 높음을 나타낸다.

다음으로 학교급구분을 보면, 일반계 고교, 중학교, 전문계 고교, 초등학교 순으로 사교육비 지출이 낮아진다. 일반계 고교의 계수가 0.1604이므로 일반계 고등학교에 다니는 학생의 log(사교육비)가 전문계 고등학교에 다니는 학생에 비하여 평균적으로 0.1604만원 정도 더 든다. 로그를 제거하면, 일반계 고교생에 대한 사교육비 지출은 전문계 고교생에 비하여 1.1740배로 17.4% 더 많이 든다. 반면, 중학생은 전문계 고교생에 비하여 사교육비가 약 3% 더 들고 초등학생은 반대로 전문계 고교생에 비하여 약 3% 덜 든다. 이상을 종합하면, 일반계 고교생에 대한 사교육비 지출이 다른 세 그룹에 비하여 두드러지게 높음을 알 수 있다. 이는 우리나라 교육에서 대학입시를 최종 목표로 두고 있는 학부모들이 많음을 반영한다고 할 수 있다.

부모의 경제활동의 형태는 유의한 영향을 미치지 못하였다. 이는 부모가 맞별이나 외별이 또는 경제활동이 없음이 사교육비 지출에는 차별적인 영향을 주지 못함을 나타낸다.

사교육 시간에 대한 계수는 0.0263으로 사교육에 참가하는 시간은 당연히 사교육비 지출 증가에 유의한 영향을 줌을 나타낸다. 사교육 시간이 1시간 늘어남에 따라 지출은 약 2.6% 늘어나며 만약 5시간 늘어나면 약 14% 지출이 늘어난다.

학생 성별을 보면, 여학생에 대한 계수가 음수이므로 남학생에 대한 지출이 더 많음을 의미하나 계수값이 거의 0에 가까워 남녀 차이는 거의 없다고 볼 수 있다.

학생 성적을 보면, 앞에서 우리는 5개의 그룹으로 되어있는 원래의 성적 그룹을 0과 100 사이의 점수

로 변환하여 사용하였다. 각 그룹간의 점수차이는 15점 또는 25점 이었고 최상위 그룹과 최하위 그룹의 점수 차이는 80점 이었다. 최상위 그룹(상위 10% 이내)은 최하위 그룹(하위 20% 이내)에 비하여  $e^{0.0009*80} = 1.074$ 배의 사교육비 지출을 하며, 차상위 그룹(상위 11~30%)이 비해서는 1.026 배의 사교육비 지출을 한다. 이 결과를 보면, 최상위와 최하위 그룹의 차이도 7.4%에 불과할 만큼 차이가 그다지 크지 않음을 알 수 있다. 따라서 학생들의 성적이 좋고 나쁨에 따라 사교육비 지출이 크게 차이가 나지 않음을 의미한다.

부모의 평균 연령에 대한 계수는 거의 0에 가까워 부모의 연령에 의한 차이는 거의 없음을 나타낸다. 부모의 평균 교육 년도를 보면, 부모님의 교육정도가 높을수록 학생에게 지출하는 사교육비는 증가한다는 것을 알 수 있다. 부모의 평균 교육년도가 4년 증가하면 학생의 사교육비가  $e^{0.01356*4} = 1.0557$ 배 정도 더 많이 드는 것으로 나타났다. 즉, 5.57% 정도 사교육비가 더 많이 드는 것을 의미한다. 만약 부모의 평균 교육년도가 8년 증가하면 그 학생의 사교육비 지출비용은 약 11% 증가한다.

가구소득의 계수는 0에 가깝지만 단위가 만원으로 되어 있는 것을 감안하면 영향이 없다고 할 수 없다. 단위를 100만원으로 바꾸어 생각하면, 가구소득이 100만원 늘어나면 사교육비 지출은 약 4.5% 늘어나고 200만원 늘어나면 사교육비 지출은 약 9% 늘어남을 의미한다. 이는 소득 상위층이 사교육 투자가 더 많고 사교육의 효과를 더 누릴 수 있기 때문에 계층의 세습에 영향을 미치는 요인이 될 수 있음을 의미한다.

마지막으로 방과후 학습의 참여는 사교육비 지출 감소 효과를 나타낸다. 방과 후 학교에 참여하는 학생의 사교육비가 방과 후 학교에 참여하지 않는 학생의 사교육비 보다 3.4% 적다. 이는 방과후 프로그램이 사교육비에 대한 경제적 부담을 덜어주는 효과가 있음을 나타낸다. 그러나 두 그룹의 차이가 단지 3.4%로 효과는 매우 미미함을 알 수 있다.

이상의 분석을 종합하면 다음과 같이 요약할 수 있다. 사교육비 지출에 뚜렷한 영향을 미치는 요소로는 지역구분, 학교급구분, 가구소득, 사교육 시간, 방과후 학교 참여를 들 수 있다. 지역 구분에서는 서울 지역이 읍면지역에 비하여 14% 더 지출을 하여 가장 지출이 많고 다음이 광역시와 중소도시가 비슷하게 읍면 지역에 비하여 약 8% 지출이 많았으며 읍면 지역이 가장 낮았다. 이는 우리가 예상했던 바와 같이 도시에 거주하는 학생들이 사교육에 많이 노출되어 있으며 도시 가계의 사교육비 지출이 시골 지역 보다 더 높음을 나타낸다.

학교급 구분에서는 일반계 고교생에 대한 지출이 초중생이나 전문계 고교생에 비하여 약 17% 지출이 더 많아 사교육 지출이 상대적으로 일반계 고교생에 집중되어 있음을 알 수 있다. 이는 대학입시 위주의 교육 풍토와 사교육 시장이 관련이 있음을 보여준다고 할 수 있다.

가구소득이 높을수록 지출이 많았는데 100만원 증가할 때마다 약 4.5% 사교육비 지출이 늘어남을 보여 주고 있다. 그러나 가구소득의 영향은 지역구분이나 학교급구분의 경우 보다는 영향의 크기가 덜한 것으로 나타난다. 사교육 시간은 당연히 지출 증가에 영향을 미친다. 방과후 학교 참여는 사교육비 지출의 감소에 영향을 미치는 것으로 나타나지만 참여 그룹과 불참 그룹의 차이가 약 3.4%로 실질적으로 뚜렷한 차이라고 보기에는 어려움이 있다.

부모의 별이 형태, 성별, 학생의 성적, 부모의 평균 연령, 부모의 교육정도는 통계적으로는 유의하나 계수값이 거의 0에 가까워 실질적으로 의미있는 영향을 준다고 보기는 어렵다.

### 3.2. 방과후 학교 참가비에 대한 다중 선형 회귀분석

사교육비에 대한 분석과 같은 분석을 방과후 학교 참가비를 종속 변수로 하여 수행하고자 한다. 전체 66,999명의 대상 중 방과후 학교에 참가하는 학생은 29,324명으로 전체의 43%가 참여하고 나머지

표 3.4. log(방과후 학교 참가비)에 대한 분산분석 결과

Source	Df	Sum of Squares	Mean Square	F-value	Pr > F
Model	16	694.0834	43.3802	447.76	<.0001
Error	29307	2839.3474	0.0969		
Corrected Total	29323	3533.4309			

표 3.5. log(방과후 학교 참가비)에 대한 회귀식의 결정계수 측정

Root MSE	0.3113
Dependent Mean	1.3967
Coeff. Variance	22.2867
R <sup>2</sup>	0.1964
Adj R <sup>2</sup>	0.1960

57%는 방과후 학교에 참여하지 않았다. 본 분석은 방과후 학교 참여 학생을 대상으로 방과후 학교 참가비에 영향을 미치는 인구 동태적 요인들을 살펴보는 것이 목적이므로 참여하지 않는 학생들을 제외하고 방과후 학교에 참여하는 29,324명의 학생을 분석 대상으로 하였다. 참고로, 방과후 학교 참여 여부에 영향을 미치는 요인들에 대한 분석은 추후 과제로 남겨 두기로 한다.

앞서 2.3절에서 보여준 바와 같이 방과후 학교에 참여하는 29,324명의 방과후 학교 참가비의 분포를 그려본 결과 매우 치우친 분포를 보이므로 이를 로그 변환하여 log(사교육비)를 종속변수로 하는 다중 선형회귀 모형을 적합시킨다.

변수 선택 결과 모든 변수가 포함되므로 사교육비에서와 마찬가지로 모든 10개의 변수를 설명변수로 한 회귀모형을 적합시킨다. 모형의 유의성에 대한 분산분석 결과 표 3.4와 같이 다중 선형회귀 모형이 유의함을 알 수 있었다.

회귀식의 결정계수를 추정하여 보니  $R^2 = 0.1964$ 로(표 3.5) 사교육비에 비하여 다소 낮은 수치가 나왔지만 표본 사이즈가 큰 점을 감안할 때 모형이 적절하다고 판단된다.

범주형 설명변수에 대하여 적절한 더미 변수를 이용하여 회귀 모형을 적합시킨 결과 표 3.6과 같은 회귀계수의 추정치를 얻었다.

먼저 지역구분 변수를 보면, 서울, 광역시, 중소도시의 계수 추정값들이 별로 다르지 않음을 알 수 있다. 따라서 읍면 지역을 제외하고는 지역에 따라 방과후 학교 참가비는 별 차이가 없다고 결론 지을 수 있는데, 이는 일반 사교육비의 경우 서울이 광역시나 중소도시에 비하여 월등히 높았던 것과 대조적이다. 즉, 일반 사교육비 지출은 도시간 차이가 있지만 방과후 학교 참가비는 차이가 없다는 것인데 그 이유를 추측해 보면, 방과후 학교 참가비는 전국적으로 일률적인데 반해 일반 사교육비는 지역에 따라 비용이 차등적이어서 그럴 수도 있고, 또는 방과후 학교 참여 정도가 도시간에 비슷해서 그럴 수도 있다. 그러나 읍면지역은 다른 도시지역에 비하여 방과후 학교에 대한 지출이 유의하게 낮다.

초·중·고 등의 학교급 구분의 경우 일반 사교육비의 경우와는 매우 다른 양상을 보인다. 일반 사교육비의 경우는 일반계고> 중학교> 전문계고> 초등학교 순서였으며 일반계고가 다른 학교에 비하여 월등히 사교육비 지출이 높았던 것에 반해, 방과후 학교 참가비는 일반계고> 초등학교> 전문계고> 중학교의 순서이며 일반계고와 초등학생의 지출이 비슷하다. 초등학생은 사교육비의 부담을 덜기 위해서, 그리고 일반계 고등학생의 경우는 방과후 학교에서 내신에 도움을 주는 프로그램을 제공하기 때문이 아닌가 추측된다.

부모의 경제활동 형태(맞벌이, 외벌이 등)는 사교육비의 경우와 마찬가지로 계수들이 유의하지 않다.

표 3.6. log(방과후 학교 참가비)에 대한 회귀식의 적합결과

변수	자유도	계수추정치	표준오차	t-value	Pr >  t
Intercept	1	0.9925	0.3126	3.18	0.0015
지역구분(서울)	1	0.1777	0.0063	28.10	<.0001
지역구분(광역시)	1	0.1588	0.0052	30.60	<.0001
지역구분(중소도시)	1	0.1444	0.0055	27.44	<.0001
지역구분(읍면지역)		0.0000			
학교급구분(초등학교)	1	0.1341	0.0088	15.20	<.0001
학교급구분(중학교)	1	-0.1547	0.0088	-17.61	<.0001
학교급구분(일반계고교)	1	0.2017	0.0080	25.37	<.0001
학교급구분(전문계고교)		0.0000			
부모경제활동(부외별이)	1	0.0325	0.0179	1.81	0.0698
부모경제활동(모외별이)	1	0.0240	0.0190	1.27	0.2055
부모경제활동(맞별이)	1	0.0405	0.0179	2.27	0.0234
부모경제활동(없음)		0.0000			
사교육시간	1	-0.0033	0.0003	-9.91	<.0001
학생성별(여)	1	-0.0016	0.0037	-0.42	0.6715
학생성별(남)		0.0000			
학생성적구분	1	0.0002	0.0000	2.17	0.0301
부모의 평균연령	1	0.0010	0.0002	3.25	0.0012
부모의 평균교육정도	1	0.0052	0.0009	6.15	<.0001
가구소득	1	0.0000	0.0000	5.77	<.0001
방과후학교참여(유)	1	0.0156	0.3113	0.05	0.9600
방과후학교참여(무)		0.0000			

나머지 변수들인 사교육 시간, 학생성별, 학생성적구분, 부모의 평균연령, 부모의 교육정도, 가구소득 등은 계수가 통계적으로 유의하긴 하지만 그 값이 거의 0에 가까워 실질적으로 의미있는 결론을 유추하기에는 어려움이 있다.

#### 4. 요약 및 결론

본 논문에서는 통계청에서 조사한 2008년도 2차 사교육비 실태조사 자료 중 일반사교육비 지출과 방과후 프로그램 참가비를 종속 변수에 대하여 학교급구분, 가구소득, 학생 성적, 성별, 사교육 참가시간 등의 인구동태적 변수들의 영향을 알아보는 다중 선형 회귀분석을 수행 하였다.

종속 변수인 일반사교육비 지출과 방과후 참가비의 분포가 한쪽으로 지나치게 치우친 형태를 지니고 있기 때문에 이들을 로그 변환시켜 대략적으로 정규분포를 따르도록 한 후 로그 변환된 변수를 종속변수로 한 회귀분석을 실시하였다. 51개의 원래 변수 중 결측치, 중복성 등을 고려하여 10개의 설명변수를 선택하고 이들을 대상으로 단계적 변수선택을 수행한 결과 모든 10개의 변수가 적어도 하나의 범주 이상이 변수선택이 포함되었기 때문에 10개의 변수들을 설명변수에 포함시킨 분석을 수행하였다. 모형의 유의성 검정과 결정계수를 통하여 다중 선형모형이 유의함이 입증되었다.

회귀계수 추정치를 통한 분석을 종합하면 다음과 같이 요약할 수 있다. 일반 사교육비 지출과 방과후 프로그램 참가비에 뚜렷한 영향을 미치는 요소로는 지역구분, 학교급구분이고 나머지 변수들은 통계적으로 유의하지 않거나 유의하다 해도 계수값이 0에 가까워 실질적으로 의미있는 차이로 보기에는 어려움이 있다. 일반 사교육비 지출에 대한 지역 구분의 영향을 보면, 서울지역 > 광역시 > 중소도시 > 읍면지

역 순으로 지출이 많았는데 서울지역은 읍면에 비하여 14% 더 지출을 하며 광역시와 중소도시는 읍면 지역에 비하여 약 8% 지출이 많았다. 방과후 참가비에 대한 지역구분의 영향을 보면 서울지역, 광역시, 중소도시 > 읍면지역 의 순으로 지출이 많았는데 서울과 기타도시의 차이는 없었다는 것이 일반 사교육비의 경우와 다른 특징이다.

학교급 구분에서는 일반사교육비의 경우 일반계 고교생에 대한 지출이 초·중생이나 전문계 고교생에 비하여 약 17% 지출이 더 많아 사교육 지출이 상대적으로 일반계 고교생에 집중되어 있음을 알 수 있다. 이는 대학입시 위주의 교육 풍토와 사교육 시장이 관련이 있음을 보여준다고 할 수 있다. 방과후 학교 참가비의 경우 일반계고, 초등학교 > 전문계고 > 중학교 의 순서이다. 초등학생은 사교육비의 부담을 덜기 위해서, 그리고 일반계 고등학생의 경우는 방과후 학교에서 내신에 도움을 주는 프로그램을 제공하기 때문이 아닌가 추측된다. 그런데 여기에서 특징적인 점은 중학생의 방과후 참가비가 다른 그룹에 비하여 매우 낮다는 점이다. 중학교에서의 방과후 프로그램이 활성화되지 못했다는 점을 반영하는 것으로 볼 수 있는데 그 이유로는 프로그램이 만족할 만한 수준이 아니거나 개별적인 심화교육 위주의 사교육 시장으로 학생들이 이동하기 때문일 수도 있다.

본 논문에서는 사교육비와 방과후 프로그램에 참여하는 학생들을 대상으로 학생들이 지출하는 금액에 영향을 미치는 요인들을 살펴 보았다. 따라서 사교육이나 방과후 프로그램에 참석하지 않는 학생들은 분석대상에서 제외되었다. 사교육 지출이 전혀 없는 학생들은 전체의 약 30%, 방과후 지출이 전혀 없는 학생들은 전체의 약 60%로 상당한 부분을 차지하고 있는 만큼 사교육 참여 여부 그리고 방과후 참여 여부에 영향을 미치는 요인들을 분석하는 일을 추후 과제로 남겨 두기로 한다.

## 참고문헌

- 교육과학기술부 (2009). 사교육비 및 사교육 의식 파악을 위한 학부모 여론 조사, (주)코리아리서치센터.  
 통계청 (2008). 사교육비조사 보고서(11-1240000-00511-10).  
 통계청 (2009). 사교육비조사 보고서.

# Statistical Analysis of Private Education Expenses in Korea

Man-Suk Oh<sup>1</sup> · Jin Hee Kim<sup>2</sup>

<sup>1</sup>Department of Statistics, Ewha Woman's University

<sup>2</sup>Department of Statistics, Ewha Woman's University

---

## Abstract

Due to the great impact of private education expenses on many areas including economics and politics, reducing private education expenses is one of the key issues in Korea. In this paper, we analyze the data from a survey on private education expenses, conducted by Statistics, Korea, in 2008. We study the effect of some demographic variables on two dependent variables, the expenses for out-of-school private education (Private) and the expenses for after-school programs (Afterschool), by using a multiple linear regression model. The analysis results show that 'residential area' and 'school level' variables have a significant effect on the two dependent variables. 'Private' increases in the order of small town, middle town, or metropolitan city, and Seoul, by about 7%. On the other hand, 'Afterschool' are about the same for all areas except for the small town. In terms of the effect of 'school level', 'Private' for high school students is about 17% larger than all other students including professional high school students. This shows a strong correlation between university admission and private education, in Korea. 'Afterschool' is larger for high school and elementary school students and decreases in the order of professional school students and middle school students. It seems that after-school programs are alternatives to expensive private education programs for elementary school students, and that high school students are attracted to after-school programs to get a good GPA, which is important for university admissions.

**Keywords:** Private education, after-school program, multiple linear regression.

---

---

This work was supported the Korea Research Foundation(KRF) grant funded by the Korea government (MEST)2009-0071093.

<sup>1</sup>Corresponding author: Professor, Department of Statistics, Ewha Women's University, Daehyeon-dong, Seodaemun-gu, Seoul 120-750, Korea. E-mail: msoh@ewha.ac.kr