



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

LASSO를 통한 고등학생의 주관적
학업성취도에 영향을 미치는 변수 탐색



한국교원대학교 대학원

교육과정 및 교육평가 전공

박진영

2018년 2월

LASSO를 통한 고등학생의 주관적
학업성취도에 영향을 미치는 변수 탐색

지도교수 유 진 은

이 논문을 교육학 석사학위 논문으로 제출함

한국교원대학교 대학원

교 육 과 정 및 교 육 평 가 전 공

박 진 영

2018년 2월

박진영의

교육학 석사학위 논문을 인준함

심사위원장 이 승 은 인

심사위원 최 성 욱 인

심사위원 유 진 은 인

한국교원대학교 대학원

2018년 2월

목 차

I. 서론	1
1. 연구의 필요성 및 목적	1
2. 연구문제	5
3. 용어의 정의	5
가. 학업성취도	5
나. 주관적 학업성취도	6
II. 이론적 배경	8
1. 학업성취도	8
가. KCYPS 자료를 활용한 선행연구 분석	9
나. 그 외 고등학교 학업성취 관련 선행연구 분석	20
2. 통계적 방법론	27
가. 로지스틱 회귀모형	28
나. 변수선택법	30
다. 모형평가	38
3. 데이터 마이닝	41
가. 빅데이터	42
나. 데이터 마이닝	44
다. 데이터 마이닝을 활용한 선행연구 분석	46
III. 연구 방법	50
1. 연구자료	50
가. 자료 정리	54
나. 반응 변수	60
다. 설명 변수	62

2. 분석방법	64
가. 분석과정	64
나. 분석도구 R	67
3. 모형평가 기준	68
 IV. 연구 결과	 70
1. LASSO 별점 회귀 모형 구축	70
가. 전체 성적 주관적 평가를 반응변수로 한 모형	70
나. 국어 성적 주관적 평가를 반응변수로 한 모형	75
다. 수학 성적 주관적 평가를 반응변수로 한 모형	78
라. 영어 성적 주관적 평가를 반응변수로 한 모형	81
2. LASSO 별점 회귀 모형 결과 분석	84
가. 전체 성적 주관적 평가에 영향을 미치는 변수 분석	84
나. 국어 성적 주관적 평가에 영향을 미치는 변수 분석	101
다. 수학 성적 주관적 평가에 영향을 미치는 변수 분석	104
라. 영어 성적 주관적 평가에 영향을 미치는 변수 분석	106
마. 각 모형별 중요도 순 10개 변수 비교 분석	108
 V. 논의	 116
1. 연구 방법 및 연구 과정에 대한 논의	116
2. 연구 결과에 대한 논의	122
 VI. 결론 및 제언	 126
1. 결론	126
2. 제언	128
 참고 문헌	 133
ABSTRACT	141
부록	143

표 목 차

〈표 I-1〉 성적 관련 변인	6
〈표 II-1〉 영역별 각 변수의 이용 빈도별 정리	11
〈표 II-2〉 정오분류표(confusion matrix)	39
〈표 II-3〉 데이터 마이닝 기법을 활용한 교육관련 선행연구(연도별) ...	46
〈표 III-1〉 KCYPS 조사문항 체계	52
〈표 III-2〉 삭제된 분 단위 응답 문항	56
〈표 III-3〉 삭제된 영근처 분산 가지는 문항	56
〈표 III-4〉 결측 비율이 10% 이상인 문항	58
〈표 III-5〉 전체 성적 주관적 평가 응답수	60
〈표 III-6〉 과목별(국어, 수학, 영어) 주관적 평가 응답수	61
〈표 III-7〉 각 반응변수별 ‘1(우수)’ 와 ‘0(미흡)’ 의 비율	62
〈표 IV-1〉 회귀계수가 0이 아닌 29개 설명변수와 척도(270개 변수사용) 89	
〈표 IV-2〉 중요도 순 변수 정리(270개 변수사용)	92
〈표 IV-3〉 회귀계수가 0이 아닌 41개 설명변수와 척도(266개 변수사용) 96	
〈표 IV-4〉 중요도 순 변수 정리(266개 변수사용)	99
〈표 IV-5〉 국어 성적 모형에 포함된 설명변수 및 척도	102
〈표 IV-6〉 수학 성적 모형에 포함된 설명변수 및 척도	104
〈표 IV-7〉 영어 성적 모형에 포함된 설명변수 및 척도	107
〈표 IV-8〉 각 모형별 중요도 순 10개 변수 정리	109
〈표 IV-9〉 중요도 순 10위 이내 학습습관 관련 변수	110
〈표 IV-10〉 중요도 순 10위 이내 학교적응에서 학습활동 변수	111

〈표Ⅳ-11〉 중요도 순 10위 이내 진로계획 관련 변수	112
〈표Ⅳ-12〉 중요도 순 10위 이내 자아인식 관련 변수	113
〈표Ⅳ-13〉 중요도 순 10위 이내 개인발달 영역 나머지 관련 변수	113
〈표Ⅳ-14〉 중요도 순 10위 이내 매체환경 관련 변수	114
〈표Ⅳ-15〉 중요도 순 10위 이내 발달환경 영역 나머지 변수	114
〈표Ⅳ-16〉 중요도 순 10위 이내 배경변인 영역 관련 변수	115
〈표Ⅵ-1〉 반응변수에 따른 모형의 성능 비교	126



그 립 목 차

[그림Ⅲ-1] 반응변수(INT1E) 코딩 R code	60
[그림Ⅲ-2] 영분산, 영근처 분산 가지는 변수 확인 R code	63
[그림Ⅲ-3] LASSO 로지스틱 회귀 모형 적합 R code	66
[그림Ⅲ-4] 2017년 가장 인기 있는 10대 분석도구	67
[그림Ⅲ-5] 모형 평가 및 변수의 중요도 출력 R code	69
[그림Ⅳ-1] 별점모수와 회귀계수	70
[그림Ⅳ-2] 이탈도와 회귀계수	71
[그림Ⅳ-3] 기준별 10-fold cv 결과 (전체 성적 모형)	72
[그림Ⅳ-4] 정오분류표 모형 평가 결과	73
[그림Ⅳ-5] ROC 곡선	74
[그림Ⅳ-6] 이탈도와 회귀계수, 별점모수와 회귀계수	75
[그림Ⅳ-7] 기준별 10-fold cv 결과 (국어 성적 모형)	76
[그림Ⅳ-8] 국어 성적 모형 평가 R code	77
[그림Ⅳ-9] 이탈도와 회귀계수, 별점모수와 회귀계수	78
[그림Ⅳ-10] 기준별 10-fold cv 결과 (수학 성적 모형)	79
[그림Ⅳ-11] 수학 성적 모형 평가 R code	80
[그림Ⅳ-12] 이탈도와 회귀계수, 별점모수와 회귀계수	81
[그림Ⅳ-13] 기준별 10-fold cv 결과 (영어 성적 모형)	82
[그림Ⅳ-14] 영어 성적 모형 평가 R code	83
[그림Ⅳ-15] 변수의 중요도 출력 R code	91
[그림Ⅳ-16] 변수의 중요도순 차트	93

논 문 요 약

LASSO를 통한 고등학생의 주관적 학업성취도에 영향을 미치는 변수 탐색

박 진 영

한국교원대학교 대학원 교육과정 및 교육평가 전공
(지도교수 유진은)

본 연구는 한국아동·청소년 패널조사 중1패널 6차년도 자료를 이용하여 고등학교 3학년의 주관적 학업성취도에 영향을 미치는 변수 탐색을 목적으로 하였다. 연구 방법으로 LASSO 로지스틱 회귀모형을 이용하였고, 분석도구로는 R 프로그램을 사용하였다.

전체 성적 주관적 평가, 국어 성적 주관적 평가, 수학 성적 주관적 평가, 영어 성적 주관적 평가를 각각 반응변수로 하여 모형을 적합한 결과, 전체 성적 주관적 평가 모형의 성능이 가장 뛰어났다. 전체 성적 모형은 데이터 전처리 후 남은 설명변수 270개 중 29개 변수가 선택되었으며, 이 때 모형의 예측 정확도는 85.98%였다.

모형에 포함된 변수 중 몸무게, 성취가치, 숙달목적 지향성, 행동통제, 학업시간 관리, 주의집중, 공격성, 우울, 자아존중감, 자아탄력성, 자아정체감, 장래희망 교육수준, 학업관련 시간, 양육방식, 학습활동, 교우관계, 교사관계, 매체환경, 모 최종학력, 보호자 삶의 만족도 관련 문항은 선행 연구에서 이용한 변수들이었다. 자아탄력성, 양육방식, 교우관계, 교사관

계를 제외하고는 대부분 선행연구 결과와 일치하였다.

학업성취도에 영향을 미치는 변수 중 학습습관에서 행동통제, 성취가치, 숙달목적 지향성, 학업시간 관리, 진로계획에서 장래희망 교육수준, 학교적응에서 학습활동 변수가 중요한 변수인 것으로 나타났다. 고등학교 3학년은 학교에서 하는 공부를 중요하게 생각하고 수업에 집중하는 학생들이일수록 학업성취도가 높은 것으로 나타났다. 선행연구에서 다루어지지 않았으나 별점 회귀모형에서 포함된 변수로 공동체 의식, 봉사활동 참여경험 유무, 형제 자매수, 학교규칙 준수와 관련되는 문항이었다.

시사점으로 첫째는 학업성취를 높이기 위한 방안으로 학생들의 행동통제 능력을 키워주는 방안을 마련하자는 것, 둘째는 ‘마을교육공동체’와 같은 학교-지역-가정이 연계하여 학생들이 자연스럽게 진로에 대해 생각하고 계획을 세울 수 있는 여건을 마련하자는 것, 셋째는 학교에서 교과외의 가치를 내면화할 수 있도록 가르쳐서 인성을 기를 수 있도록 하자는 것, 넷째는 교육 데이터에 통계기법을 활용하여 데이터 속에 숨겨진 패턴을 찾고 정책적으로 활용할 것을 제언하였다.

※ 이 논문은 2018년 2월 한국교원대학교 대학원위원회에 제출된 교육학 석사학위 논문임.

I. 서론

1. 연구의 필요성 및 목적

인간은 태어나서 아동·청소년기를 거쳐 성인기에 이르기까지 다양한 환경과의 상호작용을 통해 사고력, 정보처리능력, 기억 및 학습능력과 같은 지적 영역의 발달이 지속적으로 일어난다. 이때, 개인의 지적 발달은 물리적 환경과 심리적 환경 등에 따라 다양한 차이를 보인다. 지적 발달 중 학습정도를 나타내는 가장 대표적인 척도 중 하나는 학업성취도로 청소년의 진학 및 진로를 결정하는 여러 가지 변인 중 매우 중요한 요인 중 하나이다. 특히 고등학교에서의 학업성취 수준은 평생의 사회 경제적 지위, 사회적 계층 이동성을 예측할 수 있는 중요한 결정요인으로 볼 수 있다(김현주, 이병훈, 2007). 청소년기의 학교적응 실패, 낮은 학업성취는 그들이 성인이 되었을 때 직업적, 경제적 성과에까지 장기적으로 부정적인 영향을 미치기 때문에 성인기의 삶을 예측하는 주요 변인이기도 하다(김광혁, 차유림, 2007).

또한 학업성취도가 중요한 이유는 개인의 정신건강에 영향을 미치기 때문이기도 하다. 성적에 대한 실패 경험이 누적된 고등학생은 성공경험이 누적된 이들에 비해 정신적으로 건강하지 못하고, 다시 이런 학습된 무기력으로 인해 학업성취도에 좋지 않은 영향을 주게 된다(임규혁, 1991). 즉, 학업에서의 누적된 실패 경험은 학업적 효능감을 약화시키고 학습에 대한 동기를 떨어뜨림으로써 궁극적으로 무력감을 느끼도록 한다는 것이다(김지경, 백혜정, 임희진, 이계오, 2010). 이러한 입장에서 볼 때, 학업성취는 청소년기의 중요한 발달 과업 중의 하나로 고려되어야

할 것이다.

따라서 많은 연구자들은 학업성취에 영향을 주는 다양한 변인들을 살펴보고, 그 변인들이 학업성취에 어떻게 영향을 미치는가에 대한 탐색을 시도하였다. 기존 연구에서는 학업성취도를 성취해야 할 일련의 학습과제를 교수-학습의 과정 속에서 어느 정도 달성했는가를 말하는 것으로(신종호, 신태섭, 2006) 보고, 학업성취에 직접적으로 영향을 미칠 것으로 예상되는 변인을 선정하여 개인관련, 가족관련, 학교관련 요인으로 분류하여(김선숙, 고미선, 2007) 그 변인들 사이의 관계를 분석하는 연구가 주를 이룬다.

그러나 개인의 발달은 자신을 둘러싼 복잡한 환경체계 속에서 체계와의 상호작용을 통해 성장한다는 브론펜브레네(Bronfenbrenner, 1979)의 생태학적 관점으로 볼 때, 단순히 학업성취도와 직접적으로 연관되는 요소들에 의해 학업성취도가 결정된다는 생각에서 벗어나 개인발달적인 요소와 환경발달적인 요소들을 모두 살펴볼 필요가 있다(김지경 외, 2010).

최근 한국청소년정책연구소(NYPI)에서 실시한 한국아동·청소년 패널조사(Korean Children & Youth Panel Survey: 이하 KCYPS)에서는 생태학적 관점에 기초하여 개인발달, 발달환경, 배경변인에 대한 다양한 주제를 다루고 있다. 그러나 KCYPS 데이터를 이용한 선행연구에서도 대부분 이론적인 배경에 기초하여 일부 변수들만을 선정하여 학업성취도와 관련된 변인 간 관계를 연구한 것이 주를 이룬다. 기존의 양적연구법을 적용하는 연구, 즉 HLM(Hierarchical Linear Modeling), SEM(Structural Equation Modeling), LGM(Latent Growth Modelling) 등을 적용하는 연구에서는 선택된 변수를 수집하는 시간과 노력이 많이 들기 때문에 기존

이론을 토대로 유의할 것으로 예상되는 변수를 선택한다. 그러나 빅데이터 분석이 각광받는 요즘에는 기존 이론에 의거하지 않고, 이미 축적된 자료에서 데이터 마이닝 기법을 이용하여 새로운 변수까지도 찾아내어 연구를 확장할 수 있다(유진은, 2016).

따라서 본 연구에서는 데이터 마이닝 기법 중 하나인 LASSO(Least Absolute Shrinkage and Selection Operator) 별점회귀모형을 이용하여, KCYPS 자료에서 제공하는 사백 개가 넘는 변수 중 학업성취도에 영향을 미치는 변수를 찾아내고자 한다.

연구 대상에 있어서도 기존 연구에서는 중학생을 대상으로 한 연구가 대부분이다. 이에 고등학교 3학년의 학업성취도에 영향을 미치는 변수에 대한 연구는 의미가 있다. 또한 2015년에 실시한 KCYPS 6차년도 조사에 진로계획(진로계획, 진로관) 문항이 추가되어 진로계획이 학업성취도에 미치는 영향까지 살펴볼 수 있다.

본 연구에서 사용된 LASSO 별점회귀모형은 계수를 축소 추정하는 특징으로 인하여 빅데이터 분석 시 많이 이용되는 기법이다. 데이터마이닝은 현재도 활발히 연구되고 있는 분야이며, 대용량의 자료로부터 정보의 요약과 미래에 대한 예측을 그 목표로 하며 자료에 존재하는 관계, 패턴, 규칙 등을 탐색하고 이를 통계적으로 모형화함으로써 이전에는 알려지지 않은 유용한 지식을 추출하는 일련의 과정으로 정의할 수 있다. 앞으로 정보처리 및 저장 기술의 급격한 발전으로 교육학 분야에서도 교수학습 관련 데이터가 폭발적으로 증가할 것이다(이은경, 박도영, 최인봉, 2014).

본 연구에서 직접적인 학생들의 학업성적(점수)을 이용하지 않고 주관적 학업성취도를 이용하였다. 학업성취는 학생들이 자신의 능력 수준을

어떻게 생각하느냐에 따라 영향을 받게 된다. 자신이 공부를 잘 할 수 있다고 믿는 학생과 그렇지 않은 학생은 학업성취 수준에서 차이가 존재하게 된다. 학업성취도를 단순히 학생의 시험점수로 본 것이 아니라 학습한 것을 얼마나 알고 있는지에 더 초점을 맞춘 것이다. 시험 점수는 학생의 객관적인 실력이라고 흔히 생각하지만, 때때로 환경에 의한 실수 또는 운에 의해 영향을 받기 때문이다. 또한 본 연구에서 사용한 KCYPS 자료의 특성상 같은 시험을 일제히 치른 결과가 아니기 때문에 학교마다 수준의 차이가 있어 직접적인 점수를 이용하는 것보다 학생이 직접 자신의 학업수준에 대해 응답한 주관적 학업성취도를 사용하는 것은 의미가 있다(김선숙, 고미선, 2007). 이에 본 연구는 KCYPS 중1패널 6차년도 고등학교 3학년 학생들의 자료를 활용하여 고등학생의 ‘주관적 학업성취도’에 영향을 미치는 변수를 탐색해 보고자 한다.

우리나라 고등학생들에게 지금 하고 있는 공부에 대해서 질문한 결과, ‘사회인으로서 기본’ 이라고 생각은 하지만, ‘싫어도 해야 하는 것’, ‘입시에만 도움’ 된다는 부정적인 응답이 많았다(임희진, 2010). 따라서 LASSO를 통하여 KCYPS 자료가 제공하는 수백 개의 변수들을 최대한 활용함으로써 고등학생의 주관적 학업성취도에 영향을 미치는 변수를 탐색하고, 기존의 연구에서 사용된 변수와의 비교분석뿐만 아니라, 모형화를 통해 고등학생들이 학업에 관한 긍정적인 생각과 태도를 갖고 앞으로의 삶을 살아갈 수 있도록 시사점을 도출하는 것을 목적으로 한다. 급변하는 시대변화를 반영한 연구가 필요하다. 변화한 학생들에게 새롭게 학업성취도에 영향을 미치는 변수가 있을 것이다. 그 변수를 찾아내어 학생들이 학업에 더욱 정진할 수 있도록 환경을 마련해 주는 것이 필요할 것이다.

2. 연구문제

본 연구는 빅데이터에 적합한 방법인 LASSO를 활용하여 KCYPS에서 제공하는 400여 개의 변수들 중에서 학업성취도에 영향을 미치는 변수를 탐색하고자 한다. 고등학교 3학년 학업성취도에 영향을 미치는 변수들을 탐색하고, 그 결과에 따른 학생들의 학업 성장을 위한 시사점을 도출하고자 한다. 대학교 입시에 과도하게 집중되어 학업에 대한 부담을 가장 많이 받는 시기가 고등학교 3학년인 만큼 이 시기 이후에도 배움의 자세를 잃지 않고 자신이 하고자 하는 분야에서 계속 공부해 나갈 수 있도록 돕기 위해 중요한 요인이 무엇인지 살펴보고자 한다.

〈연구 문제〉 고등학교 3학년 학생의 주관적 학업성취도에 영향을 미치는 변수는 무엇인가?

3. 용어의 정의

가. 학업성취도

학업성취는 학교교육을 통하여 학습한 지식, 사회적 능력, 가치관, 예술능력, 운동기능 등 학습결과의 총칭이다. 그러므로 학업성취에는 지적 영역의 학습결과만이 아니라 비지적 영역의 학습결과도 포함된다(김신일, 2011). 이런 의미에서 학업성취도는 학생들의 성적은 물론 태도변화, 도덕성 함양 등 정의적 영역의 학습 결과까지 고려(반상진, 정성석, 양성관, 2005)하는 것이 바람직하다.

나. 주관적 학업성취도

본 연구에서 사용한 ‘주관적 학업성취도’는 학생이 본인의 성적이 어느 정도였는지 주관적으로 판단하여 직접 응답한 것이다. 구체적인 문항과 변인값은 <표 1-1>과 같다.

<표 1-1> 성적 관련 변인

변인	문항	변인값
전체 성적 주관적 평가	학생의 지난 학기 전 과목 성적은 다른 친구들과 비교할 때 다음 중 어디에 해당한다고 생각하십니까?	Likert: 1(매우 우수) ~ 7(매우 미흡)
과목별 주관적 평가	국어 성적 주관적 평가	Likert: 1(매우 잘했다) ~ 5(매우 못했다)
	수학 성적 주관적 평가	
	영어 성적 주관적 평가	

본 연구에서 사용하는 자료는 중1패널 6차년도 자료로 2015년에 고등학교 3학년이 된 학생들의 응답 자료이므로, 질문에서 지난 학기는 고등학교 3학년 1학기를 의미한다.

학업성적으로 자기보고식 평가결과(주관적 평가)를 사용하기 때문에 학생들의 과목에 대한 효능감이나 다른 학생들과의 비교 등에 의한 주관적 판단의 영향을 받을 여지가 있다. 그러나 청소년의 주관적인 자기평가를 분석 자료로 사용한 연구들은 실제 학기말 고사 평균점수와 학생들이 평정한 학업성취도간 상관분석을 실시한 결과 지각된 학업성취도는 비교적 신뢰할 만한 관계로 추정되었다고 보고하였다(추상엽, 임성문, 2007).

또한 학생들이 직접 자신의 성적에 대해 응답하는 것은 미래핵심역량

중심 학생평가 모델 중에 학습자로 하여금 자신의 학습 과정과 그에 대한 생각과 느낌을 되돌아보게 하는 과정 중심적 평가 방법인 ‘자기평가(self-evaluation)’ 라고 할 수 있다(조은순, 남민우, 조재윤, 김현진, 2016). 즉, 학습자 스스로 자신의 학습에 대해 돌아보고 결과를 해석해 보는 기회를 가지는 것이다. 이러한 관점에서 학업성취도로 지난 학기 성적을 돌아보고 스스로 자신의 성적에 대한 평가를 해 본다는 것은 의미가 있다. 점수 자체가 가장 공정한 학업 성취를 나타낸다고 생각할 수 있으나, 본 연구에서는 점수 자체가 나타내 주지 못하는 학생의 성취까지 포함해서 학생 스스로 평가한 학업성적을 ‘주관적 학업성취도’로 명명하여 사용하였다.

I 장에서는 연구의 필요성과 목적, 연구 문제, 용어를 정의하였고, II 장에서는 이론적 배경으로 학업성취도 관련 선행연구, 통계적 방법론 그리고 데이터 마이닝에 대해 살펴보았다. III 장에서는 연구 방법으로 연구 자료, 데이터 전처리 과정, 분석 방법 LASSO 및 분석도구 R, 분석 절차 그리고 모형평가 기준에 대해 설명하였다. IV 장은 연구 결과 부분으로 LASSO 별점 회귀 모형 구축 및 결과를 분석하였다. V 장은 연구에 대한 논의로 연구 방법 및 연구 과정, 연구 결과에 대한 논의를 하였다. 마지막으로 VI 장에서는 연구 결과를 요약하고 시사점을 도출하여 제언하였다.

II. 이론적 배경

1. 학업성취도

2015 청소년통계 자료에 의하면 2014년 13~19세 청소년의 49.5%가 ‘성적과 적성을 포함한 공부’에 대해 가장 고민하는 것으로 나타났으며 그 이전 2008년부터 2012년까지 2년 단위로 조사한 자료에서도 50% 이상의 청소년들이 가장 고민하는 문제가 ‘성적과 적성을 포함한 공부’라고 응답하였다(통계청, 2015). 전국 청소년상담복지센터 및 한국청소년상담복지개발원 등을 통한 청소년 고민, 고충 등 상담내용과 상담청소년의 대상에 대한 연도별 현황조사에서 고등학생을 대상으로 연도별 전국 청소년 상담내용¹⁾ 현황 중 ‘학업/진로’ 내용에 대한 상담 비율이 2009년부터 2014년까지 20%에서 28%까지 증가하다가 2015년과 2016년에는 20% 정도를 차지함을 보고하였다(여성가족부, 2017).

이것은 다른 중요한 것이 있음을 알지만 여전히 학업의 고민으로부터 자유롭지 못한 현실을 반영한 것이다. 학생이나 학부모의 관점에서 성적 즉 학업성취 수준은 가장 큰 관심거리일 수밖에 없다. 진학과 졸업, 취업 등 선발에 관련되는 것에는 가장 중시하는 판단 기준이 학생의 학업성취이기 때문이다. 또한 학업성취수준은 다음 단계의 과정을 학습할 준비가 되었는지를 나타내는 예측자료로도 이용되지만, 흔히 한 사람의 총체적 능력을 가늠하는 척도로도 사용되기 때문이다(김신일, 2011).

상식적으로 성적은 지적 능력(IQ 포함)과 노력여부, 학교의 교육여건에

1) 청소년들이 주로 상담을 받고자 호소하는 문제를 중심으로 총 11개 영역으로 구성되어 있는데 11개 영역으로는 대인 관계, 학업/진로, 정신 건강, 일탈/비행, 컴퓨터/인터넷, 가정, 성격, 단순 정보 제공, 성, 활동, 생활 습관 태도이다.

따라 결정된다고 생각한다. 그러나 성적은 생각보다 많은 요인의 영향을 받아 결정되기 때문에 많은 연구가 진행되어져 왔다. 특히 우리나라와 같이 입시가 차지하는 비중이 큰 나라에서는 학업성취 수준을 높이는 것에 대한 관심이 매우 높다. 성적이 상급학교진학에 큰 영향을 준다는 사실은 교육기회의 분배와 직결되는 요인이라는 의미이기도 하다. 교육 기회분배는 말할 필요도 없이 사회적 지위의 획득으로 연결되므로, 성적은 사회적 평등문제와도 관련된다(김신일, 2011). 따라서 학업성취도가 어떻게 결정되고 어떤 변인에 의해 영향을 받는지를 연구하는 것은 중요한 과제임에 틀림없다.

선행연구 분석은 KCYPS 자료를 활용한 연구와 그 외 고등학교의 학업성취도 관련 연구로 나누어 살펴보고자 한다.

가. KCYPS 자료를 활용한 선행연구 분석

KCYPS 자료는 개인발달 영역, 발달환경, 배경변인의 세 개의 대영역으로 구성되어 있다. 개인발달 영역은 아동·청소년이 자신을 둘러싼 주변 환경의 영향을 받으면서 성장 발달해가는 양상을 파악하기 위한 것으로서, 신체 발달, 지적 발달, 사회·정서적 발달, 진로계획, 비행, 생활시간의 6가지 영역이다. 발달환경 영역은 아동·청소년의 사회화과정에 중요한 영향을 미치는 다양한 환경요인으로서, 가정환경, 친구관계, 교육환경, 지역사회환경, 매체환경, 활동·문화 환경의 6가지 영역으로 구성되어 있다. 배경변인 영역의 조사내용은 성별, 지역, 교급, 학년, 보호자 학력/직업, 가정경제수준 등으로 데이터 분석에 필요한 청소년과 보호자

의 기본 속성 관련 항목들로 구성되어 있다.

RISS와 Google scholar에서 “KCYPs”를 키워드로 검색하였을 때 2017년 4월을 기준으로 국내학술지 논문을 450건 이상 찾을 수 있었다. 그중 학업성취도와 관련이 있는 논문을 찾기 위해 “KCYPs”와 “학업성취도”를 주제로 검색하여 그중 32편의 학술논문을 선정하였다. 각 논문에서 이용된 연구방법을 중복하여 세었을 때, 가장 많이 사용된 연구방법은 ANOVA를 포함한 일반 회귀분석으로, 총 15편의 연구가 위계적 회귀분석, 중다회귀분석, 매개회귀분석, 단순회귀분석, OLS회귀분석, ANOVA, 자기회귀교차지연모형(ARCL) 등을 이용하였다. 구조방정식모형(SEM)이 14편의 연구에서 이용되었으며, 잠재성장모형(LGM), 군집분석은 각각 2편의 연구에서 이용되었다. 위계적 선형모형분석(HLM), 위계적 일반화선형모형분석(HGLM), 경로분석, 다중집단분석, 교차지연 패널분석, 다차원 척도법, PAMS 모형, 프로빗 모형은 각각 1편의 연구에서 이용되었다.

연구에 주로 쓰인 변수는 학교(생활)적응, 부모의 양육방식, 학습습관, 자아인식(자아존중감, 진로정체감, 자아탄력성), 삶의 목표·만족도, 비만, 정서문제, 휴대전화, 월평균 교육비, 주관적 건강상태 등이었다. 배경변인으로는 성별, 부모의 사회경제적 지위에 관한 조사항목(연간 소득, 부모 학력/직업 등)도 연구에서 많이 사용되었다.

각 영역별로 변수의 이용 빈도를 정리하면 <표 II-1>과 같다. KCYPs를 활용한 선행연구 32편 중, 여러 변인의 사용 시 중복으로 헤아렸을 때, 개인발달 영역 변수는 38개, 발달환경 영역 변수는 33개, 배경변인 영역 변수는 15개 정도 이용되었다. 그중 가장 많이 사용된 변수는 발달환경 영역의 학교(생활)적응과 부모의 양육방식으로 각각 총 13편의 연구에서

이용되었다. 개인발달 영역에서는 학습습관이 12개, 자아인식이 11개가 이용되었다.

<표 II-1> 영역별 각 변수의 이용 빈도별 정리

조사영역			이용 빈도	조사항목
대영역	중영역	소영역		
개인 발달	지적발달	학습습관	12번	성취가치(학습가치), 행동통제(자기조절학습), 숙달목적 지향성, 학업시간 관리
	사회정서 발달	자아인식	11번	자아존중감(자존감), 자아탄력성, 자아정체감, 진로정체감
	사회정서 발달	삶의 목표·만족	5번	삶의 목표, 삶의 만족도
	신체발달	체격	4번	키, 몸무게 (체질량 지수 계산하여 이용)
	사회정서 발달	정서문제	4번	주의집중, 공격성, 신체증상, 사회적 위축, 우울
	신체발달	건강	1번	건강상태 평가, 체육시간 중 운동시간
	비행	비행, 비행친구	1번	연간 경험 유무, 연간 경험 횟수, 연간 비행경험 친구수
발달 환경	교육환경	학교(생활) 적응	13번	학습활동, 학교규칙, 교우관계, 교사관계
	가정환경	양육방식	13번	방임, 학대, 감독, 애정, 비밀관성, 과잉기대, 과잉간섭, 합리적 설명
	매체환경	휴대전화	3번	휴대전화 보유여부/종류, 이용빈도, 의존도
	교육환경	방과후 교육	2번	연간경험 유무, 방과후교육 받은 이유/받지 않은 이유, 월평균 교육비, 방과후 교육비 가계부담 정도
	매체환경	성인용 매체	1번	이용빈도, 몰입도
	매체환경	컴퓨터	1번	컴퓨터 사용여부, 사용시간, 사용장소, 이용빈도
배경변인			15번	성별, 교급, 보호자 학력/직업, 가정경제수준 등

학교(생활)적응은 하위범주로 학습활동, 학교규칙, 교우관계, 네 개로 이루어져 있으며, 8편의 연구에서 네 개의 하위 범주 20문항을 모두 사용하였다(금지현, 손찬희, 채수은, 강성국, 2013; 김영숙, 조한익, 2015;

김지혜 2015; 박경민, 정승원, 신안나, 2013; 배호중, 유비, 2015; 서지영, 2015; 이귀옥, 이미리, 2013; 진성희, 성은모, 2013). 김진아(2016)의 연구에서는 학교적응을 학교수업 또는 공부에 만족감을 느끼고 교사나 친구와의 관계에 있어서 조화로운 관계를 유지하는 것으로 정의하고 있어 학습활동, 교우관계, 교사관계 세 가지 측면만을 포함하여 사용하였고, 김충일, 이강이(2015)의 연구에서는 학습활동을 간접적 학업성취도로 간주하여 학습활동을 빼 나머지 세 가지 측면을 포함하였다. 김서현, 임혜림, 정익중(2015)의 연구에서는 교우관계와 교사관계 두 가지 측면만을 포함하여 학습의 자기주도성이 학업성취 이외에도 학교 내 대인관계에도 정적인 영향을 주는 것으로 나타났다. 권지은, 이은정, 노충래(2013)의 연구에서는 청소년이 지각한 방임과 학대가 학업성적에 미치는 영향을 알아보기 위해서 교사관계와 교우관계 두 가지 측면의 매개효과를 분석하였는데, 교우관계는 유의한 매개역할을 하지만 교사관계는 유의한 매개역할을 하지 못하는 것으로 나타났다. 임효진, 김재철(2014)의 연구에서는 학습활동을 학업적응이라는 용어로 학업성취와 관련이 가장 깊다고 생각되는 변수 하나를 선택해서 사용하였다. 그 결과 성취도와 학업적응은 정적인 영향을 주고 있는 것으로 나타났다. 대부분의 연구에서 학교적응과 학업성취도는 관련이 있는 것으로 나타났으며, 학교적응을 잘할수록 학업성취도가 높아지는 경향이 있으며, 또한 학업성취도가 높을수록 학교적응을 잘하는 것으로 나타났다.

부모의 양육방식은 방임, 학대, 감독, 애정, 비밀관성, 과잉기대, 과잉간섭, 합리적 설명으로 총 여덟 가지 하위 범주로 구성된다. 여덟 가지 하위 범주를 모두 포함한 연구는 박경민 외(2013)의 연구로서 부모 양육태도에 대한 대표 프로파일을 분석한 결과 합리-애정형, 방임형, 학대형

의 세 가지 차원으로 도출하여 부모 양육 태도로 사용하였다. 이 연구에서 합리-애정형 프로파일의 유사성 지수는 초·중학생의 학교 적응 및 학업성취도와 유의한 정적 상관, 방임형 프로파일은 유의한 부적 상관을 보였고, 학대형의 경우에는 중학생의 학업성취도와 유의한 부적 상관을 보이는 것으로 나타났다. 학업성취도가 높은 집단일수록 부모 양육태도의 평균척도 점수가 합리-애정형 프로파일에 가깝게 나타났다. 임선아(2014)의 연구에서는 방임, 학대를 제외한 나머지 여섯 하위 범주를 감독, 애정, 합리적 설명을 민주적 양육태도로, 비밀관성, 과잉기대, 과잉간섭을 비민주적 양육태도로 분류하여 사용하였다. 이 연구에서 부모의 민주적 양육태도는 자기조절학습능력과 학교적응력을 매개해 학업성취에 긍정적인 영향을 주는 것으로 나타났으나, 비민주적 양육태도는 매개하는 효과를 보이지 않는 것으로 나타났다. 박지영, 정현수(2016)의 연구에서도 여섯 가지 하위 범주를 탐색적 요인 분석을 통해 수용, 심리통제, 행동통제로 명명하여 사용하였고, 심리통제는 남녀 모두 학업성취도에 부적영향을 미치고, 수용과 행동통제는 남녀 모두 학업성취도에 정적영향을 미치는 것으로 나타났다. 남학생의 경우는 부모의 양육태도가 학업성취도를 매개로 하여 진로정체감에 유의한 영향을 미치나, 여학생의 경우 학업성취도의 매개효과가 없는 것으로 나타났다. 김충일, 이강이(2015)의 연구에서는 방임, 학대를 제외한 나머지 여섯 가지 하위 범주에 대하여 학교성적 중심의 성취지향적인 부모 양육태도를 자녀의 높은 학업성취도를 중심에 두고, 과잉기대와 통제를 하는 부모 양육태도라는 조작적 정의 아래 과잉간섭, 과잉기대, 일등지상주의라는 세 가지 잠재변수를 설정하여 사용하였다. 이 연구에서 성취지향적인 부모 양육태도는 아동의 학업성취도에 부정적인 영향을 주는 것으로 나타났다. 김진아

(2016)의 연구에서는 애정, 과잉기대, 과잉간섭 세 가지 하위 범주를 사용하였는데, 과잉간섭이 학교적응 발달에는 부적인 영향을 미치지만 학업성취도 발달에 정적으로 영향을 미치는 것으로 나타났고, 애정과 과잉기대는 학교 적응 초기치에 정적으로 유의하게 영향을 미치는 것으로 나타났다. 임양미(2014)는 감독과 애정 두 하위 범주를 사용하였고, 금지현 외(2013), 권지은 외(2013), 이귀옥, 이미리(2013), 손병덕, 허계형(2016)의 네 편의 연구에서는 방임과 학대 두 하위 범주를 사용하였다. 임양미(2014)의 연구에서는 부모의 감독 수준이 높을수록 영어·수학 학업성취도가 높게 나타났으며, 특히 남자 중학생의 경우 부모의 감독이 정적으로 영향을 미치는 것으로 나타났다.

금지현 외(2013)의 연구에서는 학업성취도 수준이 높다고 해서 학교생활에 적응하고 만족해하는 것이 아니라 개인의 심리적인 문제와 부모의 양육태도(학대, 방임)가 학교적응에 더욱 영향을 미치는 것으로 나타났다. 연구대상에 따라 중학생의 경우 방임과 학대 모두 학업성적에 부적인 영향을 미쳤으나(권지은 외, 2013), 초등학생의 경우는 방임요인만 학업성취수준에 통계적으로 유의하게 영향을 미치는 것으로 나타났다(손병덕, 허계형, 2016). 양혜린, 이재은, 한기주(2014), 박은정, 이유리, 이성훈(2015)은 방임, 그리고 이승진, 김경근(2015)은 애정 하나의 하위 범주에 대한 문항을 사용하였다. 이상에서 살펴본 바와 같이 연구마다 사용된 양육방식의 하위 범주는 상이하였으나, 연구결과는 부모의 양육방식이 긍정적이라고 인식할수록 학업성취도가 높고, 부정적이라고 인식할수록 학업성취도가 낮은 것으로 나타나 대체로 일치된 모습을 보였다. 그러나, 학령기의 특성으로 인해 부모의 양육 태도가 학업성취도에 미치는 영향이 다르게 나타나기도 한다(권지은 외, 2013; 손병덕, 허계형, 2016)

학습습관은 개인발달 하위 영역에서 가장 많이 사용된 변수로 성취가치, 행동통제, 숙달목적 지향성, 학업시간 관리의 네 개의 하위 범주로 구성되어 있다. 성취가치는 학교 공부가 나의 미래에 유용한 역할을 할 것이라고 생각하며 학습의 가치를 지각하는 것이고, 행동통제는 공부가 지루하고 재미없더라도 공부에 집중하는 것이다. 숙달목적 지향성은 무엇인가 배울 수 있는 것 자체를 좋아하는 것이고, 학업시간 관리는 학습 목표를 분명히 하고 시간계획을 세우는 것이다. 네 개의 범주를 모두 사용한 아홉 편의 연구에서 학습습관이 통계적으로 유의하게 학업성취도에 영향을 미치는 것으로 확인되었다(김서현 외, 2015; 박은정 외, 2015; 박은정, 이유리, 이성훈, 2016; 성은모, 진성희, 2012; 손병덕, 허계형, 2016; 이유리, 박은정, 이성훈, 2015; 임선아, 2014; 진성희, 성은모, 2013; 한희원, 문경숙, 2014). 임양미(2014), 임효진, 김재철(2014), 허균(2014)의 연구에서는 성취가치와 행동통제 두 개의 하위 범주를 사용하였고, 네 개의 범주를 사용한 아홉 편의 연구와 마찬가지로 학업성취도에 통계적으로 유의하게 영향을 미치는 것으로 나타났다. 중학교 3학년을 대상으로 한 임양미(2014)의 연구에서 남녀 중학생 모두 자기조절학습능력(학습습관)이 영어·수학 학업성취도를 가장 잘 예측하는 변인인 것으로 확인되었다. 또한 자기조절전략(행동통제와 학업시간관리)은 성취가치와 학업성취, 학업적응을 매개하는 역할을 하는 것으로 나타났다(임효진, 김재철, 2014; 임선아, 2014).

자아인식은 자아존중감, 자아탄력성, 자아정체감, 진로정체감 네 가지 하위 범주로 구성되어 있으나, 대부분의 연구에서 자아정체감을 제외한 나머지 세 개의 범주 중 한 개만을 사용하였다. 자아존중감(자존감)은 6 편의 연구에서 사용되었는데, 자아존중감이 학업성적에 미치는 영향은

통계적으로 유의한 것으로 나타났다(김서현 외, 2015; 김지혜, 2015; 송영주, 2016; 심도현, 강현구, 정수지, 2014; 이자형, 이기혜, 2015; 허균, 2014). 즉 자아존중감이 높을수록 학업성적이 높은 것으로 나타났다. 허균(2014)의 연구에서는 자아존중감이 학업성취에 직접적인 영향뿐만 아니라 자기 통제력과 함께 성취가치와 학업성취 사이의 관계에서 다중매개적 효과를 동시에 보이는 것으로 나타났다. 학업성적 만족도가 자아존중감에 미치는 영향은 후기 아동기보다 초기 청소년기에 더 낮게 나타났다(송영주, 2016), 자아존중감은 학업성취와 학교 내 대인관계 모두에 대해 정적영향을 미치는 것으로 나타났다(김서현 외, 2015). 다음으로 진로정체감은 3편의 연구에서 사용되었고, KCYPS 자료의 설문문항은 ‘장래에 내가 꼭 하고 싶은 직업 분야가 있다.’, ‘나는 장래에 어떤 인생을 살 것인가에 대해 대체로 방향을 정했다.’, ‘나는 어떤 직업 분야를 좋아하는데, 그 이유가 분명하다.’ 등의 8개 문항으로 구성되어 있다. 진로정체감은 자기조절학습을 매개로 하여 학업성취에 영향을 미치고(한희원, 문경숙, 2014), 부모로부터의 방임경험이 학업성취에 영향을 주는 과정에서 조절효과를 가지는 것으로 나타났다(양혜린 외, 2014). 남학생의 경우 부모의 양육태도가 학업성취도를 매개로 하여 진로정체감에 유의한 영향을 미치나, 여학생의 경우 학업성취도의 매개효과가 유의하지 않은 것으로 나타났다(박지영, 정현수, 2016).

자아인식의 하위 범주인 자아탄력성은 2편의 연구에서 사용된 변수로, KCYPS의 설문문항은 ‘나는 갑자기 놀라는 일을 당해도 금방 괜찮아지고 그것을 잘 이겨 낸다.’, ‘나는 평소에 잘 해보지 않았던 새로운 일을 해 보는 것을 좋아한다.’, ‘나는 매우 에너지(힘)가 넘치는 사람이다.’ 등의 14개 문항으로 구성되어 있다. 즉, 자아탄력성이 높은 사람들

은 상황이 변하거나 새로운 경험을 하여도 기본적으로 긍정적 정서를 유지하고, 낙관적 태도, 자신감, 자율적이고 생산적인 활동에 관심을 기울이는 경향이 있다(임효진, 김재철, 2014). 성취가치, 자기조절과 학업성취, 학업적응간의 구조적 관계의 탐색에서 자기조절, 성취도에 대한 성취가치의 효과는 자아탄력성이 더 클수록 높다는 결과가 나왔고(임효진, 김재철, 2014), 자아탄력성이 부모의 성취 지향적 양육태도가 학교생활적응, 학업성취에 미치는 영향력을 매개하는 것으로 나타났다(김충일, 이강이, 2015).

삶의 만족도는 삶의 목표·만족도의 하위 범주의 하나로 네 편의 연구에서 사용되었다(김영숙, 조한익, 2015; 이귀옥, 이미리, 2013; 이유리 외, 2015; 전영상, 최영신, 2017). 전영상, 최영신(2017)의 연구에서는 자녀의 성적수준보다 성적향상이 자녀뿐만 아니라 부모의 삶의 만족감과 변화에 유의한 영향을 미치는 것으로 나타났다. 중학생의 경우 학습습관 우수형은 학업성취와 삶의 만족도가 높은 반면, 학습습관 취약형은 가장 낮은 것으로 나타났다(이유리 외, 2015). 학교생활적응, 학업성취도, 주관적 안녕감(삶의 만족도)의 자기회귀지연계수들은 높게 나타났으며 학업성취도가 학교생활적응과 주관적 안녕감에 영향을 미치는 교차지연모형의 계수들은 통계적으로 유의한 것으로 나타났다(김영숙, 조한익, 2015). 삶의 목표(생애목표)는 박은정 외(2015)의 연구에서 군집분석을 통하여 생애목표 유형을 나누어 분석한 결과, 남학생의 경우 생애목표고조형과 성공목표추구형이 학업성취가 가장 높았으며, 생애목표저조형이 학업성취가 가장 낮게 나타났다. 여학생의 경우 가족성공추구형이 학업성취가 가장 높으며, 생애목표저조형의 학업성취가 가장 낮게 나타났다. 즉, 삶의 목표가 학업성취도에 긍정적인 영향을 미치고 있었다.

체격의 하위 범주인 키와 몸무게를 이용한 체질량지수를 비만의 기준으로 사용한 연구에서 비만이 학업성취도에 부정적인 영향을 주는 것으로 나타났다(배호중, 유비, 2015; 서지영, 2015; 이자형, 이기혜, 2015). 권용재, 이광현(2012)의 연구에서는 국어, 수학 성취도에는 비만이 부정적인 영향을 주는 것으로 보고하였으나, 영어 성취도의 경우 초등학교에서는 긍정적인 영향을, 중학교 단계에서는 부정적인 영향을 미치는 것으로 전환됨을 보고하였다.

정서문제는 주의집중, 공격성, 신체증상, 사회적 위축, 우울 5개의 하위 범주로 구성되어 있고, 3편의 연구에서 5개의 하위 범주를 모두 사용하였다(금지현 외, 2013; 서지영, 2015; 이귀옥, 이미리, 2013). 부모의 양육태도는 정서문제를 매개로 하여 학교적응에 영향을 미치는데, 학교적응이 학업성취도에 유의한 영향을 미치기 때문에 정서문제가 학업성취도에 직접적인 영향은 미치지 않으나 매개적인 역할을 할 것으로 보고하였다(금지현 외, 2013). 김지혜(2015)의 연구에서는 우울 범주를 사용하였고 우울이 학업성적에 미치는 직접적인 영향력은 통계적으로 유의하지 않다고 보고하였다.

방과후 교육의 하위 범주 중 하나인 월평균 교육비를 사용한 연구에서 남녀 모두에게 영어·수학 학업성취도에 영향을 미치나, 여자 중학생의 경우 월평균 사교육비가 영어·수학 학업성취도를 예측함으로써 성별차를 확인하는 것으로 보고하였다(임양미, 2014). 중학생의 경우 교육비와 가구소득이 통계적으로 유의하게 성취도에 영향을 주고 있었다(권용재, 이광현, 2012).

휴대전화 의존도, 휴대전화 사용목적(이용빈도), 오락적 휴대전화 이용을 사용한 연구에서 휴대전화 의존이 높을수록 오락적 휴대전화 이용의

빈도가 높을수록 학업성취에는 유의한 부적효과가 발생함을 보고하였다(김지혜, 2015; 성은모, 진성희, 2012; 이승진, 김경근, 2015).

건강(주관적 건강 상태 평가), 유해사이트 활용, 비행, 비행친구, 학습목적 컴퓨터 사용시간, 사교육 시간은 32편의 선행 연구 중 각각 한 편에서 사용된 변수들이다. 학업성취도는 건강 영향 요인에 채택되지 않는 것으로 나타났고(이자형, 이기혜, 2015), 남자 중학생이 여학생보다 유해사이트 활용빈도와 유해사이트 몰입도가 높은 것으로 나타났고, 유해사이트 몰입도가 높을수록 학업성취도가 낮은 것으로 확인되었다(진성희, 성은모, 2013). 학업성적은 비행과 비행친구 모두에 대하여 부적인 인과적 효과를 갖는 것으로 보고되었다(이종원, 2013). 즉 비행과 비행친구로 인하여 학업성적이 낮아진다고보다는 낮은 학업성적이 비행과 비행친구로 이어진다고 볼 수 있다. 학습목적 컴퓨터 활용은 고등학교 3학년에서 학습자가 주관적으로 느끼는 학업성취도 정도와 유의한 인과관계가 있다고 보고하였다(허균, 2017).

배경변인으로 성별, 교급, 보호자 학력/직업, 가정경제수준(가구소득) 등이 변수로 사용되었다. 부모의 교육수준, 가정경제수준(가구소득)은 학업성취에 강력한 정적효과를 가지는 것으로 나타났다(이승진, 김경근, 2015; 심도현 외, 2014).

선행 연구에서 살펴본 바와 같이 연구대상의 초·중고, 학년에 따라 각 변수들이 학업성취도에 미치는 영향이 상이한 경우가 있었다. 초·중학생과 중학생을 대상으로 한 연구가 대부분이고 연구에 사용한 영역 변수 중 발달환경 영역과 배경변인 영역에 관한 연구가 많은 비중을 차지한다. 청소년의 학업성취 변화에 영향을 주는 요인을 분석한 연구에서 중학교와 고등학교간의 학제간 차이가 있음이 나타났다(김선숙, 고미선,

2007). 즉, 청소년의 학업성취는 고등학교 진학이후로는 가족 및 학교관련 변인의 영향이 크지 않으며, 사교육비와 같은 학업과 관련된 직접적인 요인이나 청소년 개인이 스스로를 얼마나 존중하는가가 유의하게 영향을 미친다는 것이다.

이에 KCYPS 자료에서 제공하는 400개가 넘는 변수들을 이용하여 데이터 마이닝 기법을 통한 고등학교 3학년 학생들의 학업성취도에 영향을 미치는 변수의 탐색은 의미가 있다.

나. 그 외 고등학교 학업성취 관련 선행연구 분석

앞서 살펴 본 KCYPS 자료를 활용한 선행연구를 제외한 고등학생의 학업성취도와 관련된 연구를 분석하기 위하여, Google scholar에서 “고등학교 AND 학업성취도”를 키워드로 검색하였다. 2017년 10월 20일 기준으로 국내학술지 논문에서 약 12,000개라는 검색결과를 얻었고, “학업성취도”를 키워드로 검색하였을 때는 약 25,100개라는 검색결과를 얻었다. “high school AND academic achievement”로 검색하였을 때, 국외학술지 논문이 400만 건 이상 검색되었다. 학업성취는 다양한 맥락과 요인에 의해 영향을 받기 때문에 단순히 설명하기에는 한계가 있다. 따라서 학업성취에 관련된 연구들은 변인과 영역을 달리하여 끊임없이 지속되어 오고 있다(곽수란, 2006)는 것을 확인할 수 있었다. 방대한 양의 학술논문 중 KCYPS 자료의 조사영역에 포함되는 것을 중심으로 연구결과를 살펴보면 다음과 같다.

김선숙, 고미선(2007)은 잠재성장모형을 적용하여 청소년의 학업성취 변화에 영향을 주는 요인을 분석한 결과, 청소년의 학업성취는 고등학교

진학이후로는 가족 및 학교관련 변인의 영향이 크지 않으며, 사교육비와 같은 학업과 관련된 직접적인 요인이나 청소년 개인이 스스로를 얼마나 존중하는가에 유의하게 영향을 받는 것으로 나타났다.

임현정, 시기자, 김성은(2016)은 2014년 최초로 구축된 국가수준 학업성취도 평가 전수학생의 초-중-고 연계자료를 분석한 결과, 성별, 지역 규모, 가정환경에 따른 학력격차는 유지 또는 심화되는 경향이 있고, 학업적 효능감, 교과태도, 수업태도, 학교생활 행복도가 향상된 학생의 학업성취 역시 향상되었으나, 고등학교에 진학하면서 평균적으로 이런 정의적 특성들이 감소하는 경향을 나타내고 있음을 보고하였다. 또 부모와의 관계가 좋아질수록, 자기주도 학습시간, 독서시간이 증가하는 학생의 학업성취가 향상된 것으로 나타났다.

그러나 고등학교에서 독서시간의 증가가 학업성취도에 오히려 방해가 된다는 연구도 있다(김경근, 연보라, 장희원, 2014). 고등학생의 경우 교과공부와 직접적인 연관성이 적은 활동을 최소화하는 방향으로 시간 운용을 하는 것만이 학업성취 향상에 도움이 된다는 것이다. 김경근 외(2014)는 일반계 고등학생의 학업성취에 영향을 미치는 요인에 대한 연구에서 학업성취는 성별, 부모 학력, 가족 구조, 교우관계, 교사관계, 학습태도, 사교육 참여, 독서시간, 운동시간, 자기주도적 학습시간, 부모 기대교육수준, 학교평균 SES, 교사의 학생성취기대와 밀접하게 연관되어 있는 것으로 보고하였다.

학생은 학업성취의 주체이기 때문에 학습과 교육에 대한 열의 정도가 학업성취에 영향을 미친다는 것은 당연한 사실이다. 학업에 성공적인 학생들은 대체로 자신에 대해 좋은 평가를 내리고 있으며, 미래의 성취에 대해서도 낙관적이고, 일반적인 능력에 대해서도 확신을 갖는 경향을 보

인다(곽수란, 2006). 학생특성과 관련된 연구들에서 학생의 학업성취는 미래에 대한 계획과 그 계획을 달성하기 위해 스스로 목표를 가지고 지속적으로 학습하는 정도에 따라 그 성취정도가 결정된다는 공통된 결론을 보여준다. 또한 목표달성을 위해 투입하는 학습의 절대적인 양도 학업성취를 결정하는 중요한 요인으로 보고하였다.

자아탄력성이란 개인의 성격체제를 어려운 상황에서도 견딜 수 있도록 하거나 심리적인 적응 능력을 갖도록 자아구조를 연결시키는 정도이다(Block & Kermen, 1996). 즉, 자아탄력성이 높을수록 상황의 변화와 우발적인 사건에 유연하게 대처하며 상황적인 특성과 개인행동 사이의 적절한 분석력을 가지고 있어 새로운 상황에 대한 불안의 민감성을 없애고, 세상에 대한 긍정적인 참여와 경험에 대한 개방성을 갖도록 한다. 자아탄력성에 관한 연구에서 학업성취도에 긍정적인 영향을 준다는 결과(권지은, 2003; 조한익, 이나영, 2010)와 반대로 자아탄력성이 학업성취도에 부정적인 영향을 준다는 결과(Asendorpf & Aken, 1999)도 있다. 즉, 자아탄력성이 자아통제(ego-control)와 역전된 U-모양의 관계 패턴을 보이는데 자아통제를 잘하는 학생들의 학업성취도가 높기 때문에 자아탄력성이 높을수록 자아통제 능력이 떨어져 학업성취도에 부정적인 영향을 주는 것으로 보고한 것이다. 또한 IQ가 높은 사람과 자아탄력성이 높은 사람의 특성을 분석한 연구에서 자아탄력성이 높은 사람은 불분명한 장면이나 대인관계에서 보다 자신감 있고 유능함을 느끼는 경향을 나타낸 반면, IQ가 높은 사람은 분명한 장면에서 보다 효과적이고 생산적인 경향을 나타내었다(Block & Kermen, 1996).

학습자의 학습태도가 학업성취에 큰 영향을 미치는데 학습태도 관련 변인으로 진로계획, 즉 미래 직업에 대한 이해와 희망 전공과 관련된 구

체적인 활동들이 학습태도에 영향을 주는 것으로 나타났다(김경식, 이현철, 2009). 학생들이 자신의 미래 직업과 진로에 대하여 관심을 가지며 그에 따라 교육포부 수준을 설정하고 체계적으로 준비해나가는 과정이 자연스럽게 자신의 학습태도에 영향을 주고 있음을 보여주는 것이다. 김경식, 이현철(2009)은 학업성취가 학생들의 장애와 직결된다는 믿음을 여전히 간직한 채 생활하고 있는 우리의 교육현실에서 학업성적을 올리려면 과외보다는 학교수업에 충실하고 자기주도적 학습능력을 향상시키는 학습태도가 무엇보다도 중요하다는 것을 강조하였다. 박수원, 최병호(2017)의 연구에서 진로 선택가치를 사회공헌으로 선택한 경우, 다른 가치들을 선택한 경우보다 학생의 내재적 학습동기를 높이고, 학업성취와 긍정적인 관계를 가지는 것으로 나타났다.

김혜숙(2012)은 ICT 활용이 학업성취도에 미치는 영향에 대한 연구에서, 학교에서의 ICT 활용과 인터넷/오락 활용도는 성취도와 부적 관련성을, 고차원 ICT 과제에 대한 자신감과 컴퓨터에 대한 태도 변인은 정적 관련성을 가지는 것으로 보고하였다. 이수진, 신현주, 이지혜(2014)는 최근 사회적 문제가 되고 있는 인터넷 중독성향과 학업성취와의 관련성 연구 결과, 학업자신감이 높을수록 자기학습조절전략을 더 많이 사용하며 인터넷 중독성향이 낮아진다는 것을 보고하였다. 마시멜로 테스트(marshmallow test)로 유명한 월터 미셸(Walter Mischel)은 마시멜로 실험에 참여했던 아이들이 고등학생이 되었을 때, 그 중 633명에게 설문으로 그들의 학업성적에 대한 것을 포함한 특성에 대한 평가를 실시하였다. 결과는 빨리 종소리를 울렸던 즉, 자발적인 지연 행동을 못한 아이가 학교와 가정에서 문제행동을 일으킬 가능성이 더 큰 것으로 나타났고, 수학적능력시험(scholastic aptitude test; SAT)에서 낮은 점수를 받았다

(Matthew, 2013). 즉 자기통제 능력이 뛰어난 15분을 기다릴 수 있었던 아이는 SAT 점수에서 30초를 기다릴 수 있었던 아이들의 평균 점수보다 200점 이상 높았으며, 오래 기다렸던 아이들은 나중에 더 큰 성취를 이루었음을 확인하였다(Lehrer, 2009). 이것은 자기조절학습이 학업성취를 예측한다는 국내의 연구와 일치하는 결과이다(노현중, 손원숙, 2015; 양명희, 이정아, 2012). 노현중, 손원숙(2015)의 연구에서 자기조절학습이 수업 및 숙제태도를 매개하여 학업성취도에 통계적으로 유의한 효과를 주는데 특히 숙제에 대한 긍정적인 태도가 학업성취를 정적으로 예측하는 것으로 나타났다.

성별과 수학, 과학 교과 학업성취도와와의 관계에 대한 연구가 많이 수행되었는데 대체로 과학과 수학에서는 남학생이 강세를 보이지만, 국어와 영어에서는 여학생이 우세를 보이는 결과를 제시하고 있다(박경미, 최승현, 2002; 김혜숙, 2012). 하지만 최근에는 전반적으로 여학생이 남학생에 비해 학업성취에서 확고한 우위를 점하는 경향이 나타나고 있다(김경근, 2005). 사교육의 효과에 대한 연구에서는 학업성취에 유의한 영향을 미친다는 연구결과도 있지만, 사교육의 효과가 없다는 연구결과도 있었다(김경근, 2005; 김경근 외, 2014; 김현진, 2007; 임다희, 권기현, 2012). 따라서 학업성취에 대한 사교육의 효과에 대해서는 일관된 결과를 도출하기 어려운 실정이다.

학업성취에 영향을 미치는 요인으로 가정배경변인은 그 동안 국내외의 많은 연구에서 특별히 관심을 받아왔는데 부모가 형성하는 가정환경은 자녀의 인지적 성취 뿐 아니라 정의적 발달에 지속적인 영향을 미친다고 볼 수 있다. 특히 부모의 사회경제적 지위변인, 즉 부모의 학력, 수입, 직업 등은 자녀의 학업성취에 중요하게 영향을 미치는 것으로 나타

났다(곽수란, 2004, 2006; 김현진, 2007; 임다희, 권기현, 2012).

곽수란(2006)의 연구에서 인지적 성취(언어, 수리, 외국어 영역의 수능 표준점수 사용)와 비인지적 성취(태도, 자아개념, 가치관 정도 포함)에 영향을 미치는 요인들을 탐색하였는데 인지적 성취는 가정변인(부모교육, 가정수입, 가정관심, 교육기대)이 가장 높고, 비인지적 성취는 학생변인(학습정도, 교육열의, 교우영향 측정)이 가장 높은 영향력을 나타냈다. 학교요인은 비인지적 성취에서 전체효과만 유의할 뿐 인지적 성취에는 유의하지 않은 것으로 나타났다. 특히 교사특성이 학업성취에 미치는 영향이 인지적, 비인지적 성취에 유의하지 않은 것으로 나타났다. 교사 성취압력은 고등학생의 영어와 수학 과목 학업성취에 유의한 영향을 미치는 것으로 조사된 연구도 있다(김경근, 장희원, 2016). 학교교육에 있어서 교사의 중요성에 대한 인식은 팽배해 있는데 비하여 대규모 평가 자료를 활용한 교사와 교수·학습활동 효과 분석을 수행한 국내 연구는 거의 없으며, 수행된 선행 연구에서도 교사 관련 변인의 영향력은 미미한 것으로 나타나는 경우가 대부분이다(김양분, 임현정, 김난옥, 2012). 학생에게 교사 피드백의 영향은 학생의 성취도 수준 또는 성별 등 학생의 개인적 특성에 따라 차별적으로 나타날 수 있음과 고등학교에서 교사의 피드백 수행이 줄어들고 있음이 나타났다(노현중, 손원숙, 2015). 그러나 교사의 태도에 따라 학업성취도에 영향을 받는다는 연구결과도 있다(김경식, 이현철, 2010; 강명희, 유영란, 유지원, 2014).

정제영, 정예화(2015)에 따르면, 부모가 교육에 관여하는 수준이 학생들의 학업성취도에 긍정적인 영향을 미치는 것으로 나타났다. 특히 부모의 교육기대는 모든 학교급에서 높은 수준으로 학업성취도에 긍정적인 영향을 미치는 것으로 나타났다.

학업성취도를 반응변수로 한 선행연구에서 학업성취도를 국어, 수학, 영어 과목을 포함한 성적의 평균을 사용하거나, 국어, 수학, 영어 각 과목별 성적을 사용하였다. 과목별 성적을 학업성취도로 사용한 연구에서는 과목에 따른 학업성적에 영향을 미치는 요인들의 차이가 있었다. 예를 들면, 독서시간이나 사교육의 효과에 대해서 독서시간은 국어성적에 긍정적인 효과를 나타내는 반면, 수학성적에는 사교육이 긍정적인 효과를 나타내었다(임다희, 권기현, 2012; 김경근 외, 2014; 임현정 외, 2016).

연구 방법은 선행연구에서 의미가 있다고 밝혀진 변인들을 사용하여 그 인과관계를 밝히는 구조방정식 모형이 가장 많이 사용되었다. 이에 본 연구에서 전체 성적 주관적 평가와 국어, 수학, 영어 주관적 평가 변수 각각을 반응변수로 하고, 설명변수로 나머지 모든 변수를 사용하여 학업성취도에 영향을 미치는 변수를 탐색한다는 것은 기존 연구와 차별적인 의미가 있다. 기존 이론에서 밝혀지지 않은 새로운 변수를 탐색해서 그 관계를 살펴보는 것은 고등학생의 학업성취도 향상뿐만 아니라 고등학교 3학년의 특징과 거기에 따른 여러 가지 시사점을 얻을 수 있을 것이다.

2. 통계적 방법론

빅데이터의 분석에서 즉, 예측변수의 수가 많은 고차원의 데이터를 최소제곱법²⁾(ordinary least square method; OLS)을 적용하여 모형을 적합할 때 주의를 기울여야 한다. 변수들이 반응변수와 전혀 관련이 없어도, 모형에 포함되는 변수의 수가 증가함에 따라 결정계수(R^2)는 1로 증가하고, 이에 따라 훈련(training) 데이터의 평균제곱오차(mean squared error; MSE)는 0으로 줄어들기 때문이다. 즉 훈련 데이터에 과적합(overfitting)된 모형을 생성하게 된다. 따라서 적합모형이 모형적합에 사용되지 않은 새로운 자료에 대해서는 예측력이 떨어지는 경우가 발생한다. 즉, 모형의 일반화(generalization)가 잘 되지 않아, 새로운 자료인 검증 데이터(test set)에 대한 MSE는 모형에 포함되는 변수의 수가 증가함에 따라 극도로 커지게 된다. 그 이유는 추가적인 예측변수를 포함하게 됨으로써 계수 추정치들의 분산이 크게 증가하기 때문이다.

검증 데이터의 MSE를 살펴보면, 최고의 모형은 많아야 몇 개의 변수를 포함하는 것이 명백하다. 하지만, 단지 R^2 또는 훈련 데이터의 MSE만 조사한다면 가장 많은 수의 변수를 포함하는 모형이 최고의 모형이라고 잘못 판단할 수 있다. 이것은 많은 수의 변수를 가진 자료를 분석할 때 특별히 조심해야 하고 항상 독립적인 검증 데이터에 대해 모형의 성능을 평가한다는 것이 얼마나 중요한지를 알려준다. 따라서 고차원 설정(변수가 많은 경우)에 잘 맞는 다른 기법들이 필요하다(James, Witten, Hastie & Tibshirani, 2013).

2) 최소제곱법은 오차의 제곱합을 최소로 만드는 회귀 계수를 찾는 방법이다. 여기서 오차는 관측치에서 추정치를 뺀 값을 나타내는데 추정치가 실제 값에 가까울수록 오차가 작아져서 예측을 잘 하는 좋은 모형이라고 말할 수 있다.

본 연구에서는 손실함수에 벌점항(penalty term)을 추가하여 과적합의 문제를 피하여 일반화(regularization)가 잘 되도록 하는 LASSO 벌점 회귀모형을 사용하고자 한다. 로지스틱 회귀모형을 손실함수로 사용하고 벌점항은 l_1 -노름(norm)을 이용한 LASSO를 변수선택법으로 사용한다.

통계적 방법론의 이해를 돕기 위하여 로지스틱 회귀모형, 변수선택법, 모형평가 방법에 대한 설명은 다음과 같다.

가. 로지스틱 회귀모형

선형회귀(linear regression) 모형은 반응변수(종속변수)를 예측변수(독립변수)들의 선형(linear) 결합식(일차함수 모양)으로 다음 식과 같이 표현되는 모형이다.

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$
 (단, 오차항 ϵ 는 평균이 0, 분산이 σ^2 인 정규분포를 가정한다)

회귀모형은 주로 독립변수가 연속변수인 경우를 다루며, 장점으로는 연구모형을 검정하고 새로운 자료에 대해 예측까지 할 수 있다는 점이다(유진은, 2015a). 회귀모형이 종속변수의 분산을 설명하는 비율인 결정계수 또는 변수의 수에 벌점(penalty)을 주어 결정계수를 조정한 수정결정계수를 이용하여 모형의 성능을 평가할 수 있다. 다중회귀분석에서는 변수들 간의 다중공선성 문제가 발생하는 경우, 회귀분석의 결과를 신뢰하기 힘들게 된다(유진은, 2015a).

만일 반응변수 Y 가 두 집단을 나타내는 0과 1의 값을 갖는다고 하면 선형회귀 모형은 적합하지 않다. 집단 Y 의 예측값은 연속형 숫자가 되고, $[0, 1]$ 의 범위를 벗어나는 경우도 발생하게 된다. 이러한 선형회귀

모형의 문제점을 두 집단의 분류에 적합하도록 변형한 것이 로지스틱 회귀(logistic regression)이다(이정진, 2011).

로지스틱 회귀분석은 반응변수가 범주형 또는 순서형일 때 확률에 대한 로짓변환(logit transformation)을 하여 분석하는 방법이다. 로지스틱 회귀(logistic regression) 모형은 반응변수(종속변수)가 1과 0의 값 즉, 범주형인 경우, 각 $X=x$ 값에서 $Y=1$ 일 확률 $P(Y=1|x)$ 와 $Y=0$ 일 확률 $P(Y=0|x)$ 의 비율인 오즈비(odds ratio)의 로그값인 로그오즈(log-odds)³⁾를 반응변수로 하는 선형회귀식이다.

이진반응변수 Y 에 대하여, 다중(multiple) 로지스틱 회귀모형의 일반적 형태는 다음과 같다.

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k,$$

$$\pi(x) = P(Y=1|x), \quad x = (x_1, \cdots, x_k)$$

즉, $\pi(x)$ 는 관측치가 x 일 때, 성공확률이고 $1-\pi(x)$ 는 실패확률이다. 성공할 확률 $\pi(x)$ 에 대한 오즈(odds)는 성공확률 대 실패확률의 비율로 정의되며, 오즈에 로그를 취한 것이 로그오즈(log-odds)이다. 로지스틱 회귀모형은 오즈(odds)의 관점에서 해석될 수 있다는 장점을 가진다.⁴⁾

위 모형은 다음과 같이 $\pi(x)$ 에 관한 식으로 다시 표현할 수 있다.

3) “로지스틱 회귀모형은 종속변수가 예/아니오, 합격/불합격과 같은 이항분포를 따를 때 쓰인다. 오즈($\frac{\pi(x)}{1-\pi(x)}$)에 로그를 취한 것이 로그오즈 $\log\left(\frac{\pi(x)}{1-\pi(x)}\right)$ 다.” (유진은, 2015a, p. 381).

4) 예를 들어, $\exp(\beta_1)$ 의 의미는 나머지 변수(x_2, \cdots, x_k)가 주어질 때 x_1 이 한 단위 증가할 때마다 성공($Y=1$)의 오즈가 몇 배 증가하는지를 나타내는 값이다.

$$\begin{aligned}\pi(x) &= \frac{\exp(\alpha + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\alpha + \beta_1 x_1 + \cdots + \beta_k x_k)} \\ &= \frac{1}{1 + \exp\{-(\alpha + \beta_1 x_1 + \cdots + \beta_k x_k)\}}\end{aligned}$$

로지스틱 회귀가 분류의 목적으로 사용될 때는 기준값보다 크면 $Y=1$ 인 집단으로, 작으면 $Y=0$ 인 집단으로 분류하게 된다. 분류 기준값의 결정은 사전정보 또는 정오분류표를 이용한 측도(정분류율, 민감도, 특이도)를 동시에 고려하는 등의 다양한 방법이 사용된다(나종화, 2017a).

본 연구에서는 반응변수를 범주형(이진형) 자료로 코딩하여 분석하기 때문에 로지스틱 회귀모형을 이용하였다.

나. 변수선택법

다중회귀에서 변수선택은 출력(반응)변수(Y)에 대하여 여러 입력(예측)변수들(X)이 주어졌을 때 출력변수의 값에 대하여 설명력이 있는 입력변수들을 선택하는 것이다. 변수선택이 필요한 이유는 설명력이 없는 변수들이 모형에 포함되면 입력과 출력변수간의 올바른 관계를 알기가 어렵게 되며 올바른 모형에 비해 예측오차가 증가할 수 있기 때문이다(박창이, 김용대, 김진석, 송종우, 최호식, 2015). 즉, 출력변수의 변동(variance)을 많이 설명해 주는 입력변수를 선택할 때 추정된 모형에 대해 단순하지만 해석이 쉽고 예측도 뛰어난 모델을 적합하게 된다. 입력변수의 개수가 적을 때에는 가설검정을 통하여 유의한 변수를 찾아내지만 변수의 개수가 많아지면 검정 자체가 쉽지 않다. 거기에 변수들끼리

다중공선성(multicollinearity)⁵⁾이 존재하게 되면 더 복잡한 문제를 야기하게 되기 때문에 별도의 변수선택 기준이 필요하게 된다.

모형의 과적합을 피하는 방법으로 변수를 선택할 때 사용할 수 있는 방법으로 부분집합 선택법, 별점 축소 선택법, 차원 축소법 등이 있다. 차원의 수가 증가하면 분석에 필요한 자료의 수는 기하급수적으로 늘어나게 된다. 통계모형은 적은 수의 변수들을 가지고 더 해석이 용이한 모형이 선호되기 때문에 실제 분석 자료에서 차원축소나 변수선택은 필수적인 과정이다.

1) 부분집합 선택법(subset selection)

다중회귀에서 부분집합 선택법으로 최상의 부분집합 선택(best subset selection)과 순차적 모형 선택 방법(stepwise model selection)이 있다. 최상의 부분집합 선택(best subset selection) 과정은 처음에 절편항만 포함하는 가장 작은 모형에서 출발하여 1개, 2개, ..., p 개(예측변수의 수)의 예측변수들을 포함하는 각각의 모든 모형을 만들어, 같은 예측변수의 개수를 가지는 모형들끼리 비교 후 가장 작은 오차제곱합(RSS) 또는 가장 큰 결정계수(R^2)를 가지는 모형을 각각 하나씩 선택한다. 선택된 모형들 중 교차타당화(cross-validation)에 의한 예측 오차, C_p , BIC 또는 조정된 R^2 등을 사용하여 가장 좋은 모형을 선택하는 방법이다(Hastie, Tibshirani & Friedman, 2009). 단점으로 예측변수의 수가 증가함에 따라 고려해야 하는 모형의 수가 급격히 늘어나기 때문에 계산상의 문제로 적용이 불가능해진다. 또한 예측변수의 수가 많아지면 설명량은 계속해

5) 다중공선성(multicollinearity)은 상관성이 높은 예측변수가 존재하거나, 변수의 수가 자료(개체)의 수보다 많은 경우 등에 발생한다.

서 커지기 때문에 항상 모든 변수를 포함하는 모형이 선정될 가능성이 커지고, 거기에 따라 과적합문제가 발생하여 새로운 자료에 대한 예측력이 떨어지게 된다.

순차적 모형 선택 방법은 전진선택법(forward selection method), 후진제거법(backward elimination method), 단계별 선택법(stepwise selection method)이 있다. 전진선택법은 절편항만 포함하는 가장 작은 모형에서 반응변수에 가장 큰 영향을 주는 설명변수를 차례로 모형에 포함시켜 나가되 더 이상 의미 있는 변수가 없을 때 중단하는 방법이다. 후진제거법은 모든 설명변수를 포함하는 모형에서 기여도가 낮은 변수를 차례로 제거해 나가되 더 이상 제거할 변수가 없을 때 중단하는 방법이다. 전진선택법에서 한 번 선택된 변수는 다음 단계에서 제거될 기회를 갖지 못하고, 반대로 후진제거법의 경우는 앞에서 제거된 변수는 다음에 선택될 기회를 갖지 못한다는 단점이 있다. 즉, 최적모형을 선택하지 못할 수도 있으며 변수값이 작은 변동에도 그 결과가 크게 달라지는 불안정성이 있다. 이러한 단점을 보완한 방법인 단계별 선택법(stepwise selection method)은 먼저 선택된 변수도 다음 단계에서 제거될 수 있도록 변수선택의 단계마다 체크해 나가는 방법이다(나종화, 2017a, p. 33). 그러나 입력변수의 개수가 많아지면 계산량이 많아서 실제 자료 분석에서는 널리 사용되지 않는다(박창이 외, 2015).

정리하면, 부분 집합 선택법은 해석 가능한(interpretable) 모델을 제공하지만 변수의 선택이 모형에서 유지되거나 삭제되는 개별 과정이기 때문에 매우 가변적 일 수 있다. 데이터의 작은 변화에도 매우 다른 모형이 선택 될 수 있으며 이로 인해 예측 정확도(prediction accuracy)가 떨어질 수 있다.

2) 벌점화 축소추정법(penalization shrinkage estimation)

데이터 분석에서 예측 정확도와 해석 가능성은 중요하다. 예측 정확도는 종종 일부 계수를 축소(shrinkage)하거나 0으로 설정하여 향상시킬 수 있다. 즉, 일반회귀에서 계수값이 과다하게 추정되어 특정한 몇 개 변수의 기여도가 커지게 되면 새로운 자료에 대한 예측이 부정확해질 수 있는데, 벌점회귀는 모형에서 벌점항을 사용하여 특정 몇 개의 변수가 커지는 것을 막아 특정 변수의 과다한 기여를 제한함으로써, 상관된 다른 변수가 동시에 모형에 포함될 수 있도록 허용하는 회귀방법이다. 약간의 편의(bias)를 희생함으로써 예측값의 분산(variance)이 줄어들어 전반적인 예측 정확도를 향상시킬 수 있는 것이다. 또한 해석가능성은 가장 큰 효과를 나타내는 더 작은 부분집합(subset)을 변수로 선택하면 높아진다(Tibshirani, 1996).

벌점화 축소추정법에는 대표적으로 능형 회귀(ridge regression), LASSO, Elastic net이 있다. 계수 추정치를 0에 가깝게 축소함으로써 예측력을 높이는 방법인데 능형 회귀는 변수 선택은 하지 않고 계수 추정치를 0에 가깝게 축소하는 반면, LASSO와 Elastic net은 일부 계수를 0에 가깝게 축소하고 일부 계수는 0이 되도록 함으로써 변수선택과 모형적합을 동시에 하는 방법이다. LASSO는 많은 상관성 있는 변수들 중에서 하나의 변수만을 선택한다는 점이 특징이면서 단점이라고 할 수 있는데, 이것을 보완하여 상관관계에 있는 모든 변수들을 선택하도록 벌점함수를 구성한 것이 Elastic net이다.

먼저 능형회귀부터 살펴보면, 회귀계수를 과잉 추정하는 문제를 피하기 위해서 회귀계수의 평균 쪽으로 줄어들게 고안된 방법론이 능형회귀이다. 이러한 수축(shrinkage)은 모든 회귀계수의 효과들에 대하여 동시

에 적용된다.

능형회귀의 추정치는 제약조건(constraint) $\sum_{j=1}^p \beta_j^2 \leq t^2$ 하에서

$\min_{\beta} \left(\sum_{i=1}^n y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2$ 로 표현할 수 있다. 여기서 $t \geq 0$ 이고, $t = \infty$ 이면 모형은 최소제곱법과 동일하다. 능형회귀는 라그랑주 승수법(Lagrange multiplier)에 의해서 다음과 같이 표현할 수 있다.

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|y - X\beta\|^2 + \lambda \|\beta\|_2^2$$

여기서, $\lambda \geq 0$ 는 능형회귀의 조율모수(tuning parameter)이며, λ 를 포함하고 있는 식을 능형회귀의 벌점(penalty) 함수라고 한다. 조율모수 t 또는 λ 는 교차타당법(cross-validation)을 통해 흔히 선택되며, 이를 통해 얻어진 추정치는 최소제곱의 추정치보다 축소되어 편의(bias)는 있지만 분산(variance)을 줄이면서 추정된 모형으로부터 얻을 수 있는 예측오차를 줄일 수 있다는 장점을 지닌다(박민수, 김태현, 조은석, 김희발, 오희석, 2014). 능형 회귀는 계수를 축소하여 더 안정적인 연속 과정을 제공하지만, 계수를 0으로 설정하지 않고 모든 변수를 포함하는 모형을 적합하기 때문에 모형에 대한 해석이 용이하지 않다(Tibshirani, 1996).

다음으로 Tibshirani(1996)가 제안한 ‘최소 절대 수축 및 선택 연산자(least absolute shrinkage and selection operator)’인 LASSO는 몇몇 계수의 크기를 축소하고 다른 계수는 0으로 만들기 때문에 부분 집합 선택과 능형 회귀의 좋은 특성을 유지하면서 정확한 예측과 해석가능성을 높이기 위한 방법이다. 즉, 회귀계수를 추정할 때 계수축소와 변수선택을 동시에 실행함으로써 예측력을 향상시키고 적은 수의 변수만을 선택

하여 추정된 모형에 대한 해석 또한 용이하게 할 수 있다(Tibshirani, 1996).

LASSO의 추정치는 제약조건(constraint) $\sum_{j=1}^p |\beta_j| \leq t$ 하에서

$$\min_{\beta} \left(\sum_{i=1}^n y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \quad \text{로 표현되고 } t \geq 0 \text{이다.}$$

위의 식의 제약조건인 t 는 회귀계수 값에 대하여 축소 정도를 조절하는 조절모수(tuning parameter)이다. 이 조절모수 t 값이 줄어들면 중요하지 않은 변수의 회귀계수 값은 축소되면서 순서대로 0으로 만들어져 변수선택이 되는 효과가 생긴다. 조절모수 t 값이 충분히 커지면 회귀계수 값에 대한 제약이 없어지므로 최소제곱 부분만 남아 LASSO 회귀추정량이 최소제곱 추정량이 된다(박철용, 2013).

LASSO에서 사용되는 벌점함수는 $\lambda \|\beta\|_1$ 이고, LASSO 회귀의 추정량은 라그랑주 승수법⁶⁾(Lagrange multiplier)에 의해서 벌점함수로 다음과 같이 구할 수 있다.

$$\hat{\beta} = \operatorname{argmin}_{\beta} \|y - X\beta\|^2 + \lambda \|\beta\|_1$$

t 값이 충분히 작아짐에 따라 약간의 계수들은 정확하게 0으로 된다. 따라서 λ (penalty parameter)값이 변함에 따라 연속적인 변수선택이 이루어진다(Hastie et al., 2009). 이는 능형회귀는 모든 변수에 대해 0이 아닌 값으로 계수를 추정하는 것과 다르게, 계수를 축소시키되 일부의 변

6) 라그랑지 승수법(Lagrange multiplier method)은 제약(constraint)이 있는 최적화 문제를 푸는 방법으로 최적화하려는 값에 형식적인 라그랑지 승수(Lagrange multiplier)항을 더하여, 제약된 문제를 제약이 없는 문제로 바꾼다.

수만을 선택하여 종속변수를 추정하게 된다. LASSO는 상수보다 작은 계수의 절댓값을 합한 나머지 제곱합을 최소화하는 제약 조건(constraint)의 특성 때문에 정확히 0인 일부 계수를 생성하는 경향이 있어 해석하기 좋은 모형을 제공하게 되는 것이다(Tibshirani, 1996).

능형 회귀와 LASSO의 벌점함수를 절충안으로 사용하는 Elastic net은 자료의 크기에 비하여 높은 상관도를 지닌 많은 추정 변수가 있을 때 유용하게 사용되는 방법론이다.

요약하면, 벌점회귀(penalized regression)는 모형의 일반화의 성능이 우수한 회귀방법으로 모수에 대해 적절한 제약 조건하에서 최소제곱법을 사용한다. 다양한 제약조건하에서 오차제곱합을 최소화하는 회귀계수를 찾는 문제는 라그랑지 승수법을 통해 해결된다. 이때 라그랑지 승수로 불리는 조율모수인 λ 를 결정하는 것이 중요한데 교차분석법을 이용하여 선택한다. 벌점회귀는 안정적인 추정을 제공하여 추정된 모형이 새로운 자료에 대해 더욱 나은 일반화가 되는데 기여한다.

벌점화 축소추정법을 사용한 선행연구를 살펴보면 다음과 같다. 박민수 외(2014)는 요크셔 돼지 산자수와 SNP에 대한 적용 사례 연구에서 7가지 벌점화 축소추정 기법(능형회귀(Ridge regression), 능형회귀 BLUP, LASSO, adaptive LASSO, fused LASSO, elastic net, 그리고 SCAD)을 사용하여 모형을 구축하고 비교한 결과, LASSO와 elastic net이 일부 변수만을 사용하여 가장 좋은 예측력을 나타냈으며 가장 낮은 예측 오차를 갖는다는 사실을 보고하였다. 송상윤(2015)은 예대금리차 결정요인 모형의 예측력 비교 연구에서 Ridge, LASSO 및 Elastic Net 방법론을 사용하여 모형을 구축하고 예측력을 비교한 결과, 단계별 변수선택법을 사용한 회귀모형에 비하여 벌점화 축소추정법의 예측력이 전반적으로 우수하였

으며 축소추정법 중에서는 LASSO 회귀가 가장 우수한 예측력을 보임을 밝혔다. 유진은(2017)은 TIMSS(trends in mathematics and science study) 2011 데이터를 이용한 중학생의 수학 성취도 관련 변수 탐색 연구에서 LASSO를 이용하여 모형을 구축한 결과, 약 80% 정도의 모형 예측정확도를 나타내었다. 최소제곱법에 기반한 일반회귀모형으로는 몇 천 명의 응답자가 있다고 하더라도 몇 백 개의 변수를 모두 투입한 모형은 수렴의 문제로 인하여 모형화가 어려우며, 모형화된다고 하더라도 과적합문제 또한 심각할 수 있다는 단점을 가지게 되는데 이런 단점을 보완할 수 있는 방법임을 보고하였다(유진은, 2015c). 박철용(2013)의 연구에서 호흡곤란 환자 퇴원 결정을 위하여 일반 로지스틱 회귀모형과 별점 로지스틱 회귀모형을 이용하여 예측력을 비교한 결과, 별점 로지스틱 회귀모형의 예측력이 우수한 것으로 나타났다. 석경하, 이태우(2013)의 데일리 렌즈 고객들의 거래 데이터를 기반으로 한 연구에서 의사결정나무, 로지스틱 회귀모형과 같은 기존의 통계적 분류기법과 데이터 마이닝 기법 중 배깅, 부스팅, LASSO, 랜덤 포레스트 그리고 지지벡터기계의 분류 성능을 비교한 결과, 모형의 해석, 간명성 그리고 학습에 걸리는 시간을 고려하였을 때 LASSO 모형이 적합하다고 보고하였다. 김운대, 전치혁, 이혜선(2011)의 연구에서도 별점함수를 도입하여 분류분석을 실시한 결과가 기존의 분류분석방법을 사용한 것보다 더 나은 예측력을 나타내었다. 김용대, 조광현(2013)의 빅데이터와 통계학에 관한 연구에서 고차원 모형에서는 기존의 단계적 변수선택 방법보다 별점화 기법 중 변수선택과 예측을 동시에 할 수 있는 통계기법인 LASSO를 사용할 것을 언급하고 있다. 이에 본 연구에서는 별점화 축소추정법 중 예측력이 우수한 LASSO를 사용하여 모형을 적합하였다.

다. 모형평가

구축된 모형의 성능을 평가할 때 예측력(주어진 입력변수 값에 대하여 출력변수의 값을 얼마나 잘 예측하는가), 해석력(입력변수와 출력변수와의 관계를 얼마나 잘 설명하는가), 효율성(얼마나 작은 수의 입력변수로 모형을 구축했는가), 안정성(모집단 내의 다른 자료에 적용하는 경우 같은 결과를 주는가)의 4가지 요소를 고려할 수 있다. 이 중 가장 중요한 요소는 예측력인데 아무리 안정적이고 효율적이며 해석이 쉬워도 실제 문제에 적용했을 경우 잘못된 예측을 한다면 아무 의미가 없기 때문이다(박창이 외, 2015).

예측(분류) 모형에 대한 평가는 보통 훈련용 자료(training set)에 의해 구축된 모형을 검증용 자료(test set)에 대해 적용하여 평가한다. 반응변수가 범주형일 때, 정오분류표(confusion matrix)에 기반한 측도인 정분류율(correct classification rate or accuracy), 민감도(sensitivity), 특이도(specificity)들이 사용되며, 연속형 반응변수에 대해서는 평균절대오차, 평균제곱오차 등이 사용된다(나종화, 2017a).

1) 교차타당법(cross-validation method)

교차타당법은 데이터 셋을 모형구축에 사용될 훈련용 셋(training set)과 예측력 평가에 사용될 평가용 셋(validation or prediction set)으로 나누어 모형을 평가하는 방법이다. 본 연구에서는 두 데이터 셋의 비율을 70% : 30%로 랜덤하게 나누어 적용하였다.

k-중첩(fold) 교차타당법은 전체 데이터 셋을 동일한 크기의 k조각으로

나누고, 이 가운데 한 조각을 제외한 나머지 ($k-1$) 조각으로 모형을 구축한 뒤, 남겨둔 한 조각에 대해 예측을 수행한다. 남겨지는 조각을 바꾸어 가며 이 절차를 k 번 반복한다. 각 조각에 대한 제곱 예측오차를 더하여 교차타당법의 측도로 사용한다. 별점화 축소추정법에서 조율모수를 정할 때 보통 10-중첩(fold) 교차타당법을 이용하여 최적화한다.

LOO(leave-one-out) 교차타당법은 데이터 셋의 수가 적을 때 주로 사용하는 방법으로, k -중첩 교차타당법에서 $k=n$ 인 경우에 해당한다. 즉, 한 개의 데이터만 남기고 모형을 구축한 후 남겨진 한 개를 추정하는 과정을 반복한다.

2) 정오분류표(confusion matrix)

이진형(binary) 반응변수의 경우, 예측결과가 두 개의 집단(1: 발생, 0: 미발생)으로 이루어지는 경우 정오분류표(confusion matrix)는 <표 II -2>와 같다.

<표 II -2> 정오분류표(confusion matrix)

		예측집단	
		1(발생, TRUE)	0(미발생, FALSE)
실제집단	1(발생, TRUE)	TP(true positive)	FN(false negative)
	0(미발생, FALSE)	FP(false positive)	TN(true negative)

정분류율(correct classification rate) 또는 정확도(accuracy)는 전체에서 정확히 예측한 비율을 의미한다.

$$(\text{정분류율}) = \frac{TP + TN}{n}, \quad n: \text{객체수}$$

민감도(sensitivity)는 실제 발생(TRUE)한 것을 발생(TRUE)한 것으로 제대로 분류한 비율을 의미한다.

$$(\text{민감도}) = \frac{TP}{TP + FN}$$

특이도(specificity)는 실제 미발생(FALSE)을 미발생(FALSE)으로 제대로 분류한 비율을 의미한다.

$$(\text{특이도}) = \frac{TN}{FP + TN}$$

기준값을 변경할 때 민감도가 커지면 특이도는 작아지고, 반대로 민감도가 작아지면 특이도는 증가하기 때문에, 민감도와 특이도는 동시에 크게 할 수 없다. 실제 문제에서는 연구자가 적절한 민감도와 특이도를 갖는 기준값을 사용한다(박창이 외, 2015).

연속형 예측값에 대해 적용되는 평균절대오차, 평균제곱오차, 평균절대백분위오차는 다음과 같이 정의된다.

실제값을 y_i , 예측값을 \hat{y}_i 이라고 하면,

$$\text{평균절대오차(Mean Absolute Error)} : MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$\text{평균제곱오차(Mean Squared Error)} : MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

평균절대백분위오차(Mean Absolute Percentage Error) :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100$$

3) ROC(Receiver Operating Characteristic) 곡선

ROC 곡선은 로지스틱 회귀 또는 베이지 분류⁷⁾의 사후확률처럼 연속형의 값으로 주어질 때 유용하다. ROC 곡선의 원리는 다음과 같다. 검증용 자료에 대해 예측값을 내림차순으로 정렬한 뒤, 분류를 위해 기준값(cutoff)을 선택하면 정오분류표(confusion matrix)를 얻게 된다. ROC 곡선은 기준값을 0에서 1로 변화시키면서 해당 정오분류표로부터 (1-특이도)와 민감도의 값을 구하고, 이 값을 각각 x , y 값으로 하는 점을 좌표평면에 연결하여 그린 것이다. 연구자가 중요하게 생각하는 것에 따라 민감도와 특이도의 조절이 필요한 경우 기준값(cutoff)을 조절하여 예측에 활용할 수 있다. ROC 곡선에서 곡선 아래쪽의 면적(area under the ROC curve)을 c-통계량 또는 AUC라고 하는데, 이 값은 0.5이상 1이하의 값을 가진다. AUC값이 1에 가까울수록 모형의 성능이 우수하다고 할 수 있어서 여러 모형의 비교에 사용될 수 있다(나종화, 2017a).

3. 데이터 마이닝

대용량 데이터에서 유용한 정보와 관계를 탐색하고 모형화하여 지식을 발견하고자 하는 일련의 과정을 데이터 마이닝(data mining)이라고 한다. 컴퓨터와 네트워크 기술의 발전과 더불어 정보화 사회가 급진전되면서 개인, 정부기관 그리고 대기업 같은 조직들은 매일 대량의 데이터

7) 베이지 분류(Bayes classification)모형은 베이지 정리에 기반한 방법으로, 사후확률이 큰 집단으로 새로운 데이터를 분류하는 방법이다. 문서분류, 의료진단 등에 많이 사용된다(나종화, 2017a).

(massive data)를 만들고 이를 축적시키고 있다. 과거와 달리 자료의 양이 커지고 복잡해짐으로써 자동화된 새로운 분석이 필요하게 되었다.

가. 빅데이터

2010년 이후 우리 사회의 전반에서 큰 이슈로 등장한 용어 중 하나가 ‘빅데이터(big data)’이다. 빅데이터는 데이터의 규모가 매우 크거나 복잡하여 기존의 데이터 처리 방식으로는 다루기가 난해한 데이터 셋 자체를 의미하거나, 이들 데이터 셋을 처리하는 관련 기술과 인력을 통칭하는 의미로 사용되고 있다(나종화, 2017a). 대략 2010년을 기점으로 전 세계에서 한 해 동안 발생하는 디지털 데이터의 총량이 제타(zeta)바이트를 넘어섰으며, 이러한 데이터의 폭발적인 증가는 정보통신기술(ICT)의 주도권을 인프라, 기술, SW로부터 데이터(data)로의 전이를 가져오고 있다. 실제로 2017년 구글은 텐서플로우(tensorflow)⁸⁾를 개방함으로써 전 세계에서 가장 많은 데이터를 확보한 구글이 데이터에 얼마나 큰 가치를 두고 있는지 알 수 있다(강필성, 2017).

빅데이터는 기업에서 생성되는 고객 및 구매와 관련된 비즈니스 데이터, 각종 멀티미디어 데이터, 소셜미디어 데이터, 의생명 데이터, 사물인터넷(IoT) 데이터 등으로부터 기하급수적으로 축적되고 있다. 이와 같은 디지털 데이터의 폭발적인 증가로 ‘빅데이터’를 활용하여 새로운 가치를 창출할 수 있는 능력이 여러 분야의 중요한 요소로 대두되고 있다.

8) 텐서플로우는 기계학습과 딥러닝을 위해 구글이 개발한 C++ 기반의 오픈소스 라이브러리이다. 즉 구글(Google) 제품에 사용되는 머신러닝(기계학습)을 위한 소프트웨어인데 데이터만 있으면 인공지능을 구현할 수 있도록 만든 뼈대인데 구글이 이를 무료로 제공하고 있는 것이다.

이에 따라 세계적으로 빅데이터의 창의적인 활용이 미래 경쟁력의 중요한 역량임을 인식하고, 국가 차원에서 빅데이터 활용을 위한 노력을 기울이고 있다(권영옥, 2013). 빅데이터를 단순히 수집 축적하는 것이 중요한 것이 아니라 구조화되지 않은 대규모 데이터 속에서 숨겨진 패턴을 찾아내고 여러 변수들을 통합적으로 고려하면서 창의적으로 해석할 수 있는 분석능력이 더 중요해지고 있는 것이다(강만모, 김상락, 박상무, 2012).

교육 분야에도 대규모 패널 자료, 오프라인 및 온라인 교수 학습에 관한 수많은 정보, 학생들이 많이 사용하는 소셜미디어 데이터, 교육정책에 대한 언론보도 내용 등이 축적되어 있고 앞으로 교육관련 데이터가 폭발적으로 증가할 것이다(이은경 외, 2014; 유예림, 2017). 이미 전 세계적으로 빅데이터를 교육적으로 활용하기 위한 다양한 시도들이 이루어지고 있다. MIT Sloan Management Review의 아티클(Lavelle, Lesser, Shockley, Hopkins & Kruschwitz, 2011)에 따르면, 미국의 애틀란타에 위치한 한 공립 고등학교에서는 학생들의 학업 데이터를 분석한 결과 학생들의 졸업률이나 학업성과를 좌우하는 가장 중요한 과목이 수학(algebra 1)이고, 이 과목을 잘 하는 학생들이 창의적 글쓰기(creative writing) 과목도 잘 한다는 것을 알게 되어, 분석결과를 토대로 학생들이 글쓰기 과목에서 잘 할 수 있도록 도와줌으로써, 수학 과목을 성공적으로 이수하고 나아가 학교 전체의 졸업률을 높일 수 있었다고 한다(권영옥, 2013). 우리나라에서도 학생의 데이터를 수집하고 분석하여 데이터 속의 패턴과 경향을 발견하여, 학생들이 학습하는 방식에 대한 이론을 검증하고자 시도하고 있다(최제영, 박충식, 2013).

빅데이터의 특징을 4V로 표현할 수 있는데 속도(velocity), 다양성

(variety), 양(volume), 가치(value)로 표현할 수 있다. 오늘날 빅데이터에 관한 관심이 급속히 늘어나고 있는데 기존의 데이터 분석 관점을 버리지 않는 한 아무리 새로운 솔루션을 도입한다 해도 진정한 빅데이터의 가치를 얻어낼 수 없다(김용대, 조광현, 2013).

나. 데이터 마이닝

빅데이터는 관측 데이터의 수가 많다는 의미와 더불어, 데이터의 속성 변수도 많다는 의미도 포함하고 있다. 빅데이터는 전통적인 데이터 분석 방법으로 분석하는 것이 불가능하거나 혹은 의미가 없는 경우가 많다. 그래서 하드디스크와 같은 저장매체에 보관하여 놓고 이를 빠르게 처리하고 분석하는 데이터마이닝이라는 새로운 분야가 출현하게 되었다. 즉, 고차원이 될수록 차원의 저주(분석의 정확도를 일정 수준으로 유지하기 위해서는 변수의 수가 증가할 때 자료의 수가 지수적으로 증가한다는 것, 예를 들어 100개의 변수가 있으면, 필요한 자료의 수는 대략 2의 100제곱 정도 필요하다는 것)에 걸린다. 따라서 차원을 축소해 주는 방법이 빅데이터 시대에 크게 각광 받고 있는 것이다(김용대, 조광현, 2013). 기존 통계 방법에서는 표본 수가 변수 수보다 너무 많은 경우에 제1종 오류 확률이 높아질 수 있으므로 무선표집을 한 후 분석을 하거나, 또는 변수 수가 수백 개가 넘는 경우 처음부터 전체 변수 중 일부만을 이용하여 모형 구축을 하는 것이 일반적이다. 그러나 데이터 마이닝 기법을 이용하면 무수히 많은 전체 자료를 분석에 이용할 수 있다는 점이 큰 장점이라 할 수 있다(유진은, 2016).

마이닝(mining)이란 ‘광산에서 광물을 채광’ 한다는 뜻인데, 데이터

마이닝은 ‘빅데이터라는 거대 광산에서 유용한 보물(정보)을 찾는다’라는 비유로 사용된 단어이다. 이러한 비유적인 단어 대신에 ‘지식발견(knowledge discovery)’이라는 용어를 쓰기도 한다. 모형화를 통해 패턴을 찾아내고, 그 패턴에 대한 해석과 평가과정을 통해 새로운 지식(정보)를 만들어내는데 이 전체적인 과정을 데이터마이닝으로 볼 수도 있다(나종화, 2017a).

데이터마이닝은 빅데이터로부터 규칙이나 패턴을 찾아내는 과정으로, 통계학, 데이터베이스, 기계학습, 인공지능의 영역에서 발전된 다양한 기법들을 포함한다. 데이터 마이닝의 목적은 데이터 셋으로부터 정보를 추출하고, 이를 추후 사용을 위해 이해할 수 있는 구조로 변환하는 것이다. 현시대에 우리에게 주어진 ‘빅데이터 분석’이라는 공통적인 과제를 해결하기 위하여 여러 학문분야의 사람들이 협력하여 데이터마이닝이라는 새로운 분야를 만들었다(이정진, 2011).

데이터마이닝의 과정은 입력데이터가 주어질 때 먼저 데이터를 탐색하고 데이터마이닝을 위한 적당한 형태로 변환시켜 사용하게 된다. 이 과정을 데이터 탐색 및 전처리(preprocessing)라고 하는데 잡음(noise) 데이터나 중복 데이터 제거, 흩어져 있는 데이터를 통합하거나 필요한 경우에 정규화(normalization), 이산형화(discretization) 또는 수학적 변환(transformation)을 하고, 적합한 일부 데이터 추출(feature selection), 차원 축소(dimension reduction) 변환을 거치기도 한다. 전처리된 데이터를 이용하여 데이터마이닝 모형화(modeling)를 한다. 주어진 문제에 적합한 분석방법이 무엇인지 결정하고 모형을 수립한다. 이때 모형의 평가도 함께 이루어진다. 데이터마이닝 기법으로는 연관분석(association analysis), 분류분석으로 베이지 분류, 로지스틱 회귀, 의사결정나무, 인접이웃분류,

신경망, 서포트 벡터머신, 앙상블 모형이 있고, 군집분석 등이 있다. 다변량 분석기법 중 벌점화 축소법(능형 회귀, LASSO, Elastic net 등)을 이용한 회귀분석도 사용된다. 마지막으로 분석결과를 의사결정에 효율적으로 사용하기 위하여 시각화하거나 의미 있는 분석결과만 추출, 요약, 설명하는 작업을 통해 유용한 정보를 추출한다.

다. 데이터 마이닝을 활용한 선행연구 분석

교육 분야에서 빅데이터를 활용하거나, 데이터 마이닝 기법을 활용한 연구를 정리하면 <표Ⅱ-3>와 같다.

<표Ⅱ-3> 데이터 마이닝 기법을 활용한 교육관련 선행연구(연도별)

연구자 (연도)	분석 데이터	연구 내용	분석 방법
김동일, 홍성두, 이기정 (2008)	교사22명과 초등학생 668명의 ACCENT ⁹⁾ 인 지영역검사와 국어영역 검사 데이터	교사판단 기반 학습장 에 위험군의 핵심요인 탐색	의사결정나무 모형
김완섭 (2012)	송실대학교의 교양필수 과목인 '컴퓨터활용1'의 2011년 진행된 모든 강 좌의 수업환경 및 수강 생들의 정보 데이터	컴퓨터 교양교육 성과 의 요인에 대한 분석	로지스틱 회귀 분석, 의사결정 나무 분석
이혜주, 정의현 (2012a)	한국청소년패널조사 (KYPS)	초등학교 고학년의 교 사애착에 대한 예측변 인에서의 남녀 차이 연구	의사결정나무 분석
이혜주, 정의현 (2012b)	한국청소년패널조사 (KYPS)	초등학생의 학업스트 레스에 대한 예측모형 탐색	의사결정나무 분석

이혜주, 정의현 (2013)	한 국 청 소 년 패 널 조 사 (KYPs)	중학생의 학업성취 결 정요인 탐색	의사결정나무분 석
이혜주, 정의현 (2014)	한 국 청 소 년 패 널 조 사 (KYPs)	스트레스 결정요인의 연차별 추이 분석	의 사 결 정 나 무 분석
김 진, 용환승 (2014)	전국 특성화고등학교 정보공시 자료(2012)	특성화고등학교 교육 성과 분석	회귀분석, 의사 결정트리 모형
유진은 (2015c)	한 국 청 소 년 패 널 조 사 (KYPs)	두 가지 방법에 따른 진로선택 여부에 영향 을 미치는 변수 탐색 후 비교분석	의 사 결 정 나 무 모형, 랜덤 포레 스트 기법 ¹⁰⁾
유진은 (2016)	한 국 청 소 년 패 널 조 사 (KYPs)	청소년의 진로선택과 관련 있는 변수 탐색	LASSO 로지스 틱 회귀분석
안지혜, 윤유동, 임희석 (2016)	한국아동·청소년패널 조사(KCYPS)	중학생의 삶의 만족도 예측 요인 분석	의사결정나무분 석
유진은 (2017)	TIMSS2011 중2자료	중학생의 수학 성취도 관련 변수 탐색	LASSO 로지스 틱 회귀분석
유진은, 노민정 (2017)	한국아동·청소년패널 조사(KCYPS)	중학생의 삶의 만족도 에 영향을 미치는 변 수 탐색	group LASSO
김현주, 김원경 (2017)	한국교육종단연구 자 료(KEDI, 2005)	수학 교과의 자기효능 감과 내재적 동기에 영향을 미치는 주요 요인 분석	랜덤 포레스트 기법

이혜주, 정의현(2013)은 한국청소년 패널 조사 자료의 중2 학생 3449명
을 대상으로 학업 성취를 결정하는 요인을 의사결정나무 분석을 통해
탐색하였다. 그 결과, 사교육 시간, 자기 신뢰, 본인희망 교육, 개인공부

9) (AChievement-Cognitive ability ENdorsement Tests)

10) 랜덤 포레스트(random forests)는 붓스트랩 표본을 다수 생성하고 의사결정나
무 모형을 적용하여 그 결과를 종합하는 앙상블 방법(ensemble methods)

시간 등이 학업 성취에 영향을 미치는 결정 요인으로 나타났다.

또, 이해주, 정의현(2014)은 한국청소년패널조사 자료의 중2, 고1, 고3 학생들을 대상으로 스트레스 결정 요인을 의사결정나무 분석을 통해 탐색하였다. 그 결과, 중2 학생들은 생활만족도, 성적 등이, 고1 학생들은 사교육시간, 생활만족도 등이, 고3 학생들은 학업관련 스트레스, 개인 공부시간 등이 스트레스에 영향을 미치는 요인으로 나타났다.

유진은(2015c)은 한국청소년패널 자료를 이용하여 의사결정나무와 랜덤포레스트 기법의 예측오차를 비교 분석 하였다. 그 결과, 설명변수가 많고, 설명변수 간 상호 작용이 복잡한 자료를 분류 분석하는 경우 랜덤 포레스트 기법이 의사결정나무 기법보다 예측력과 안정성이 높은 것으로 나타났다. 이에 대규모의 교육데이터 분석에서 랜덤포레스트 기법의 활용을 제안하였다.

김현주, 김원경(2017)은 수학교과에 대한 자기효능감에 영향을 미치는 주요 요인을 밝히기 위하여 한국교육종단연구(2005) 자료를 랜덤 포레스트 기법을 이용하여 분석하였다. 연구 결과 주요 요인으로 내재적 동기, 통제 기대, 초인지이고, 내재적 동기에 영향을 미치는 주요 요인은 자기효능감, 노력과 끈기, 숙달 접근인 것으로 나타났다. 그리고 자기효능감과 내재적 동기는 수학 이해도에 영향을 미치는 것으로 보고하였다.

요약하면 교육과 관련된 데이터 마이닝 기법을 활용한 연구가 초기에는 대부분 의사결정나무 모형을 사용하다가 점차 데이터 마이닝 기법에 대해 관심을 가지고 별점화 축소기법인 LASSO, group LASSO와 랜덤 포레스트 방법 등을 연구에 사용하였다. 그동안 교육 분야에서 잘 활용되지 않았던 빅데이터 분석 기법을 활용해 기존 이론에서 간과되었던 새로운 변수나 변수 간 관계를 탐색함으로써 교육 연구의 지평을 넓혔다

는 점에서 의의가 있다(유진은, 2016).

본 연구에서 사용한 별점화 축소기법을 사용한 교육 분야를 제외한 선행 연구로는 마케팅과 시장예측 분야, 예금금리차 결정요인 모형 구축, 의학 분야, 컴퓨터 공학, 통계학 분야, 생물정보학, 생태학 등에서 많이 사용되고 있었다(김윤대 외, 2011; 박철용, 2013; 박철용, 계묘진, 2013; 석경하, 이태우, 2013; 박민수 외, 2014; 송상윤, 2015; 최소현, 이학래, 박천건, 이경은, 2017).



Ⅲ. 연구 방법

1. 연구 자료

본 연구에서는 고등학교 3학년의 주관적 학업성취도에 영향을 미치는 변수를 탐색하기 위해 KCYPS 중1패널 6차년도 자료를 사용하였다. KCYPS는 아동·청소년 성장과 발달의 다양한 양상을 종합적으로 파악하는 것을 목적으로 하고 있다. 2010년에 층화다단계집락표집¹¹⁾ 방식으로 표집된 초등학교 1학년(2,342명)과 4학년(2,378명), 중학교 1학년(2,351명)의 3개 패널 총 7,071명과 그 보호자를 대상으로 2016년까지 7개년에 걸쳐 추적조사를 실시하였으며, 매년 조사데이터의 일반 공개, 한국아동·청소년패널 학술대회 및 대학원생논문 경진대회 개최, 데이터 분석 보고서 발간 등을 통해 활발히 활용되고 있는 자료이다. 지금까지 데이터 활용 논문이 약 500여 편에 달하고, 다양한 청소년정책 수립에 기초 자료로 활용되고 있다.

KCYPS 패널 자료는 한국청소년정책연구원에서 국가승인을 받아 연구한 승인통계 자료이다. 국가승인통계는 통계법 제15조 제1항에 따라 지정된 통계작성지정기관이 작성하며, 통계작성지정기관 관련 지정요건(제18조), 지정신청 및 지정(제19조), 지정의 취소 관련 조항(제20조)이 통계법에 명시되어서 철저하게 관리된다는 특징이 있다(유진은, 2015b).

11) 층화추출법(stratified sampling)은 모집단을 적당한 개수의 동질적인 층(strata)으로 분할하여 각 층에서 정해진 크기의 표본을 추출하는 방법이다. 예를 들어, 남자와 여자로 구성된 모집단에서 평균임금에 관심이 있다면 모집단을 남자와 여자의 두 개의 층으로 나누어 각 층에서 단순확률추출로 표본을 뽑아 전체 평균임금을 추정하는 방법이다(이정진, 2011).

KCYPS는 아동·청소년들이 자신을 둘러싼 주변 환경에 영향을 받으면서 성장·발달해가는 양상을 종합적으로 파악할 수 있도록 조사내용을 구성하였다. 생활·행동·의식의 다양한 영역을 포괄하고 있어 전체적으로 문항 수가 많은 편이다. 2010년에 확정된 원표본을 대상으로 추적조사가 실시되었는데 2015년 12월 종료된 제6차 조사의 원표본 유지율은 87.5%로서 다른 패널 조사들에 비해 상대적으로 높은 수준이라 할 수 있다.

KCYPS의 조사 영역은 크게 개인발달 영역, 발달환경 영역과 배경변인으로 구분되며, 그 세부 구성은 <표Ⅲ-1>과 같다. 본 연구에서는 중1 패널 6차년도 자료로 2015년 고등학교 3학년이었던 학생 2,351명이 499개 문항에 응답한 결과를 활용하였다. 이 자료를 선택한 이유는 초·중학교와는 달리 고등학교는 자신의 인생에 대해 더욱 책임감을 가져야 하는 성년기로 들어가는 직전 시기이고, 대부분의 학생들이 대학 진학을 위해 학업에 가장 집중하는 시기라서 학업성취도에 어떤 변수들이 영향을 미치는지 알아보는 것이 의미가 있다. 또한 제6차 조사부터 진로계획 영역의 하위 범주로 ‘진로계획, 대학진학이유, 취업하려는 이유, 진로 관련 대화상대/빈도, 직업선택 요인 등’에 대한 문항 34개가 추가되어 진로계획 영역 변수까지 분석에 활용할 수 있다.

지식의 양이 엄청나게 폭발하는 정보화 사회에서 대학교 교육이 끝이 아니라 평생 교육에 대한 마음을 가지고 살아가야 할 것이다. 더군다나 고등학교 이후 대학과정에서 본인이 선택한 전공에 관한 학업도 계속 이어가야 할 것이다. 이런 급변하는 시기인 고등학교 3학년의 주관적 학업성취도에 영향을 미치는 요인을 파악하여 실질적으로 대학교와 이후 삶에 지속적인 발전을 이루며 살아갈 수 있도록 돕는 것이 필요하다. 대학교 2학년 학생들을 대상으로 대학생의 학업성취도와 그 영향요인들

간의 인과관계를 분석한 연구에서 학생들의 수학능력 시험 성적은 대학생의 학업성취도에 직접적인 영향은 주지 못하고, 대학만족도 및 교육과정 만족도 등이 수업태도에 영향을 주고, 궁극적으로 수업태도가 학업성취도에 결정적인 영향을 주는 것으로 나타났다(신정철, 정지선, 신택수, 2008). 자신이 있는 곳에 대한 만족하는 마음이나 수업태도라는 것은 금방 변화되거나 나타나는 것이 아니라 지속적인 학생의 학창시절 경험들과 연결이 되어 있다. 대학교 입시에 과도하게 집중되어 학업에 대한 부담을 가장 많이 받는 시기가 고등학교 3학년인 만큼 그들의 학업성취도에 영향을 미치는 변수 탐색을 통해 고등학교 이후 배움의 자세를 잃지 않고 자신이 하고자 하는 분야에서 계속 공부해 나갈 수 있도록 하는데 중요한 요인이 무엇인지에 대한 시사점을 얻을 수 있을 것이다.

〈표Ⅲ-1〉 KCYPS 조사문항 체계 (한국청소년정책연구원, 2017)

조사영역			조사항목
대영역	중영역	소영역	
개인 발달	신체발달	체격	키, 몸무게
		신체변화	[남자] 동정 시작 시기, [여자] 초경 시작 시기
		건강	건강상태 평가, 체육시간 중 운동시간
		건강문제	건강문제 유무, 지장받는 정도
	지적발달	성적	과목별 성적평가 여부, 과목별 점수, 과목별 성적 주관적 평가 등
		학습습관	성취가치, 숙달목적 지향성, 행동통제, 학업시간 관리
	사회정서 발달	정서문제	주의집중, 공격성, 신체증상, 사회적 위축, 우울
		자아인식	자아존중감, 자아탄력성, 자아정체감, 진로정체감
		삶의 목표·만족도	삶의 목표, 삶의 만족도
	진로계획	진로계획	장래 희망 교육수준, 장래 직업 결정 여부, 고교 졸업 후 진로계획
		진로관	진로관련 대화상대/빈도, 직업선택 요인, 대학 진학/취업 이유
	비행	비행	연간 경험 유무, 연간 경험 횟수
		비행 피해	연간 경험 유무, 연간 경험 횟수
		비행 친구	연간 비행경험 친구수

발달 환경	생활시간	수면 시간	취침 시간, 기상 시간
		학 업 관 련 시간	학원/과외 시간, 학교숙제 시간, 학원/과외숙제 시간 등
		여가시간	독서 시간, 컴퓨터/게임기 오락시간, TV/비디오 시청시간 등
	가정환경	가족구성	가족구성, 부모구성, 다문화가정 여부, 형제자매 유무/수
		양육방식	방입, 학대, 감독, 애정, 비밀관성, 과잉기대, 과잉간섭, 합리적 설명
		보호자 부재	방과후 보호자 부재 일수/부재 시간
		용돈	월평균 용돈, 용돈 주는 사람
	친구관계	또래애착	의사소통, 신뢰, 소외
		부모-친구관계	부모님의 친구 인지도/친구를 만난 경험/친구 선호도
		이성친구	이성친구 유무
	교육환경	전학·결석 경험	연간 전학경험 유무, 연간 결석경험 유무, 연간 결석일수
		학 교 생 활 적응	학습활동, 학교규칙, 교우관계, 교사관계
		방과후교육	연간 경험 유무, 방과후교육 받은 이유/받지 않은 이유, 월평균 교육비, 방과후 교육비 가계부담 정도
	지역사회 환경	지역사회	지역사회 인식, 공동체의식
		다문화	다문화 수용도
	매체환경	컴퓨터	컴퓨터 사용여부, 사용시간, 사용장소, 이용빈도
		휴대전화	휴대전화 보유여부/종류, 이용빈도, 의존도
		사이버비행	연간경험여부, 연간 경험횟수
		성인용 매체	이용빈도, 몰입도
	활동·문 화 환경	체험활동	연간 참여경험 유무, 연간 참여시간, 만족도, 희망 체험활동 등
		동아리활동	연간 참여경험 유무, 연간 참여시간, 만족도
		여행·문화 활동	가족/단체 여행 연간 경험횟수, 문화활동 연간 경험횟수
		해 외 방 문 경험	해외방문 연간 경험유무, 해외 교육·연수 연간 경험유무 등
		종교활동	종교 유무/종류, 종교시설 방문빈도
		팬덤활동	선호 연예인/운동선수 유무, 팬덤활동 빈도
배경변인			성별, 지역, 교급, 학년, 보호자 학력/직업, 가정경제수준 등
패널관리			ID, 패널 구분, 조사참여 여부, 가중치 등

가. 자료 정리

본 연구에서 사용한 KCYPS 중1 패널 제6차 자료는 원표본 2351명에서 조사에 성공한 표본은 2056명으로 원표본유지율은 87.5%이고, 499개의 변수(문항)를 가지는 데이터이다. 한국청소년정책연구원은 2009년부터 NYPI 한국 아동청소년 통합조사시스템을 구성하고 “NYPI 아동 청소년 데이터 아카이브(NYPI Youth and Children Data Archive)”를 통하여 한국 아동청소년의 의식 및 생활을 체계적으로 분석할 수 있는 통계자료를 제공한다. 그 중 하나인 KCYPS 자료는 EXCEL, SAS, SPSS, STATA 네 종류의 데이터 파일로 제공된다. 데이터 전처리(pre-processing)는 모형을 구축하기 전 단계에서 수행되며, 모형의 성능에 큰 영향을 끼친다. 여기에는 데이터 정제(cleaning), 정규화, 변환 및 변수 추출과 선택 등이 포함된다. 데이터 전처리 과정을 거치기 위해 SPSS데이터를 csv(comma separated value) 형태로 저장하여 활용하였다.

SPSS에서 자료를 불러와서 csv유형으로 저장할 때, 변수옵션(선택한 변수만이 지정된 파일로 저장됨)을 사용하여 학부모 또는 학생이 미응답한 관측치와 분석에 불필요하다고 판단된 변수(문항)를 선택해제하고 저장하였다. 선택해제(삭제)된 변수는 ID, 학교급, 가중치, 출생년, 출생월, 보호자 출생년도, 학교지역, 자택지역, 주택형태, 학교 일반 중 교급, 학년, 취업하려는 이유, 비행: 연간 행동경험 횟수, 비행: 연간 피해경험 횟수, 수면시간: 취침시간, 기상시간, 방과후교육, 사이버비행: 연간 경험횟수, 성인용 매체 몰입도, 체험활동: 연간 참여시간, 체험활동: 만족도, 희망 체험활동(범주가 9개인 변수), 봉사활동 참여이유, 동아리활동: 연간 참여시간, 동아리활동: 만족도, 해외 교육/연수 기간, 팬덤 활동: 빈도이

다. 삭제한 결과 499개의 변수(문항) 중 132개의 변수가 삭제되고 367개가 남았다.

그 다음부터 진행된 데이터 전처리 과정은 R 프로그램을 사용하였다. 먼저 R에서 파일(csv)을 불러와서 학생과 학부모가 6차 조사에 참여한 데이터만 추출하여 데이터프레임(data frame)¹²⁾을 만든 후, 조사 참여여부(SURVEY1, SURVEY2) 변수를 삭제하였다. 이 과정을 위한 R code는 [부록2]에 제시하였다.

본 연구의 목적 중 하나가 학생들이 고등학교 이후 학업을 계속할 때 지속적으로 영향을 미칠 것으로 예상되는 변수를 탐색하는 것이기 때문에 취업을 희망하거나 아직 결정하지 못한 학생들은 제외하였다. 고교 졸업 후 진로계획 영역에 ‘고등학교를 졸업한 뒤의 진로를 어떻게 계획하고 있습니까?’ 문항(FUR1C)에 ‘1(대학 진학 또는 진학 준비)’로 응답한 학생들만 대상으로 선택하였다. FUR1C 문항의 비율을 조사해 본 결과, ‘1’로 응답한 비율이 전체의 80.4% 였다.

최종학력(부친 HAK2A, 모친 HAK2B, 보호자 HAK2C), 직업: 근로여부(부친 JOB1A, 모친 JOB1B, 보호자 JOB1C) 문항에서 ‘해당사항 없음’으로 응답한 것과 진로계획 (장래 희망 교육수준 FUR1A01, [보호자]장래 희망 교육수준 FUR1A02) 문항에서 ‘아직 결정하지 않았다’로 응답한 자료를 결측치(NA)로 재코딩하였다.

개인발달 영역에서 생활시간 중 학업관련시간, 여가시간을 조사하는

12) 데이터프레임은 R의 대표적인 데이터 형태이다. 모양은 행렬구조로 되어 있지만 행렬과 다른 점이 있다. 행렬은 행렬의 원소 중에 하나라도 문자형 데이터가 있으면 행렬전체는 문자형으로 다 바뀌게 된다. 하지만 데이터 프레임은 다른 열에 있는 데이터의 유형이 다른 열에 있는 데이터의 유형에 아무런 영향도 줄 수 없다(이부일, 이서현, 2016, P. 102).

문항과 발달환경 영역에서 컴퓨터 사용시간을 조사하는 문항에서 시간 단위와 분 단위를 따로 응답하도록 되어 있다. 자료 탐색 결과 분단위의 응답율이 낮고 영향을 별로 미치지 않을 것으로 판단되어 분단위로 응답한 18문항을 삭제하였다. 삭제한 문항은 <표Ⅲ-2>와 같다.

<표Ⅲ-2> 삭제된 분 단위 응답 문항

대영역	조사영역	조사항목(등교일, 비등교일)	변인명
개인발달	학업관련시간	학원/과외시간(#분)	TIM2B, TIM2D
		학교숙제시간(#분)	TIM3B, TIM3D
		학원/과외숙제시간(#분)	TIM4B, TIM4D
		기타공부시간(#분)	TIM5B, TIM5D
	여가시간	독서시간(#분)	TIM6B, TIM6D
		컴퓨터/게임기 오락시간(#분)	TIM7B, TIM7D
		TV/비디오 시청시간(#분)	TIM8B, TIM8D
		친구들과 노는 시간(#분)	TIM9B, TIM9D
발달환경	컴퓨터	사용시간(주중, 주말)(#분)	MDA1B02, MDA1B04

영과 영근처 분산을 가지는 예측변수는 모형을 망가뜨리거나 불안정한 적합의 원인이 된다. 따라서 {caret} 패키지에서 제공하는 nearZeroVar() 함수를 사용하여 영근처 분산의 변수를 식별하여 해당되는 변수 34개를 삭제하였다. 삭제된 변수는 <표Ⅲ-3>과 같다. 삭제 후 남은 변수는 312개이고, 학생수(개체수)는 1559명이었다.

<표Ⅲ-3> 삭제된 영근처 분산 가지는 문항

대영역	조사영역	조사항목(주 응답)	변인명
개인 발달	학업 일반	학교재학여부(재학)	SCL1
	학교 일반	교급(고등학교)	SCHOOL

발달 환경	직업: 근로여부	부친(일을 하고 있다)	JOB1A
	비행: 연간행동경험 유무	가출, 심하게 놀리거나 조롱하기, 집단따돌림(왕따)시키기, 패싸움, 심하게 때리기, 협박하기, 뺑뺏기, 돈이나 물건 훔치기, 성관계, 성폭행이나 성희롱(없다)	DLQ1A04 ~14
	비행: 연간피해경험 유무	심한 놀림이나 조롱당하기, 집단따돌림(왕따)당하기, 심하게 맞기, 협박당하기, 뺑뺏기기, 성폭행이나 성희롱, 심한 욕설(없다)	DLQ2A01 ~07
	가족구성	다문화가정 여부(아니다)	FAM1C
	전학/결석 경험	연간 전학경험 유무(없다)	EDU1A
	휴대전화: 보유여부	휴대전화 보유여부(예)	MDA2A
	사이버비행: 연간 경험여부	채팅/게시판 등에 일부러 거짓정보 올리기, 다른 사람의 아이디나 주민번호를 허락없이 사용하기, 채팅하면서 성별이나 나이 속이기, 다른 사람 컴퓨터나 웹사이트 해킹하기, 채팅/게시판 등에 일부러 거짓정보 올리기(#회)(없다)	MDA3A01, 03~06
	체험활동: 연간 참여경험 유무	과학/정보 활동, 교류활동, 모험/개척 활동, 환경보존 활동(없다)	ACT1A02~04, 08
	동아리활동: 연간 참여경험 유무	교외 공식 동아리활동(없다)	ACT2A02

다변량 자료 분석에서 미응답한 관측치, 즉 결측값(missing values)은 모형 구축 전 처리를 해야 한다. 대부분의 방법은 결측값을 포함하는 케이스를 제외한 뒤 완전한 자료에 대해 분석을 진행한다. 특히 데이터의 크기가 크고, 결측 자료가 차지하는 비율이 작을 때는 분석의 편의(bias)를 없애기 위한 최선의 전략이 될 수 있다(나종화, 2017c). 본 연구에서도 완전 삭제 후 분석을 실시하였다. {VIM}¹³⁾ 패키지의 `aggr()` 함수를 사

용하여 먼저 결측값에 대한 집계를 하였다. 집계결과, 116개의 변수에 결측치가 포함되었다. 결측 비율이 높은 변수를 모두 모형에 포함시키는 경우 완전제거(listwise deletion) 후 유효한 사례 수가 전체 사례수의 5%도 되지 않았다. 따라서 결측이 응답의 10%이상인 변수 35개와 결측비율이 대략 9.4%인 부친의 직업(JOB4A)¹⁴⁾ 변수를 포함하여 36개의 변수가 삭제되었다. 삭제된 변수는 <표Ⅲ-4>와 같다.

<표Ⅲ-4> 결측 비율이 10% 이상인 문항

순	변수명	변인설명	비율	순	변수	변인설명	비율
1	HAK2C	최종학력: 보호자	0.996	19	INT1C05	과목별주관적평가: 사회	0.307
2	JOB1C	근로여부: 보호자	0.996	20	MDA1B01	컴퓨터사용시간: 주중(#시간)	0.278
3	ACT1C03	체험활동만족도: 교류활동	0.976	21	MDA1B03	컴퓨터사용시간: 주말(#시간)	0.278
4	ACT1C08	체험활동만족도: 환경보존 활동	0.976	22	MDA1C	컴퓨터사용장소	0.278
5	ACT1C04	체험활동만족도: 모험/개척 활동	0.974	23	MDA1D01	컴퓨터이용빈도: 공부 및 학습 관련 정보 검색	0.278
6	ACT1C02	체험활동만족도: 과학/정보 활동	0.963	24	MDA1D02	컴퓨터이용빈도: 학습 이외의 정보 검색 및 자료이용	0.278
7	ACT1C09	체험활동만족도: 자기(인성) 개발	0.945	25	MDA1D03	컴퓨터이용빈도: 게임 및 오락	0.278
8	ACT4B	해외방문: 해외 교육/연수 연간 경험유무	0.944	26	MDA1D04	컴퓨터이용빈도: 채팅이나 메신저	0.278
9	ACT1C05	체험활동만족도: 문화예술 활동	0.897	27	MDA1D05	컴퓨터이용빈도: 전자우편	0.278

13) 결측값의 패턴을 시각화하는데 유용하다. 본 연구에서 사용한 데이터의 양이 많아 시각화 결과는 알아보기 어려웠으나 개 변수값에 대한 결측값의 비율을 알 수 있었다.

14) 부친의 직업 변수의 변인값이 1(관리자), 2(전문가 및 관련종사자), 3(사무종사자), 4(서비스종사자), 5(판매종사자), 6(농림어업 숙련종사자), 7(기능원 및 관련기능 종사자), 8(장치, 기계조작 및 조립종사자), 9(단순노무종사자), 10(군인)이다. 선행연구에서 부친의 직업이 학업 성취도에 영향을 끼친다는 연구결과가 있으나 LASSO를 써서 분석하기에 적합하지 않아 삭제하였다.

10	ACT1C01	체험활동만족도: 건강/보건 활동	0.865	28	MDA1D06	컴퓨터이용빈도: 동호회, 카페, 커뮤니티 활동	0.278
11	ACT1C07	체험활동만족도: 직업체험 활동	0.848	29	MDA1D07	컴퓨터이용빈도: 개인 홈페이지 활동	0.278
12	ACT1C06	체험활동만족도: 봉사활동	0.774	30	MDA1D08	컴퓨터이용빈도: 온라인에서 사고팔기	0.278
13	MDA4B	성인용 매체: 이용횟수	0.703	31	MDA1D09	컴퓨터이용빈도: 댓글 달기	0.278
14	INT1C10	과목별주관적평가: 미술	0.647	32	MDA1D10	컴퓨터이용빈도: 19세 이상 사이트 이용	0.278
15	INT1C09	과목별주관적평가: 음악	0.643	33	INT1C11	과목별주관적평가: 체육	0.199
16	INT1C12	과목별주관적평가: 제2외국어	0.375	34	FUR1A02	[보호자]장래희망 교육수준	0.111
17	JOB4B	직업분류: 모친	0.347	35	FAM5A	용돈: 월평균 용돈(#만원)	0.100
18	INT1C04	과목별주관적평가: 과학	0.326	36	JOB4A	직업분류: 부친	0.094

변수 삭제 후 영과 영근처 분산을 가지는 변수를 다시 확인한 결과, 5개의 변수가 더 발견되어 삭제하였다. 응답분포가 극단적이어서 모형을 망가뜨릴 우려가 있는 영분산 또는 영근처 분산을 가져서 삭제된 변수는 [보호자]청소년과의 관계(PARENT), 비행: 연간 행동경험 유무 중 무단결석(DLQ1A03), 가족구성(FAM1A), 부모구성(FAM1B), 형제자매 유무(FAM1D) 문항이다. 마지막으로 na.omit() 함수를 이용하여 완전제거 후, 자료는 271개 문항에 대해 남학생 527명, 여학생 565명의 총 1092명(70%)¹⁵⁾으로 정리되었다. 본 연구에서는 여러 영역에서 역문항을 혼용해서 사용하고 있으나, 묶음문항을 사용하지 않았고 문항수가 많아서 역코딩을 하지 않고 설문지 척도를 그대로 사용하였다. 그래서 연구결과의 정확한 해석을 위해, 모형 적합 결과를 표로 제시할 때 설문지 척도와 회귀 계수를 함께 제시하였다.

15) 고교 졸업 후 진로계획(FUR1C) 문항에 ‘1대학 진학(또는 진학 준비)’에 응답한 학생 1559명을 전체로 하여 계산했을 때의 백분율이다.

나. 반응변수(종속변수)

본 연구에서 설정한 반응변수는 ‘주관적 학업 성취도’이다. 성적과 관련된 문항은 ‘과목별 주관적 평가’, ‘전체 성적 주관적 평가’, ‘전체 성적 만족도’이다. 이 가운데 데이터 전처리 후, 과목별 주관적 평가 변수로 남은 교과가 국어, 수학, 영어 주관적 평가 성적이었다. 전체 성적 만족도는 직접적인 학업성취라기보다는 학생 본인의 성적에 만족하느냐 안 하느냐를 묻는 문항이라고 판단하여, 반응변수로 ‘전체 성적 주관적 평가(INT1E)’를 사용하였다. 전체 학업성취와 국어, 수학, 영어 성적 주관적 학업성취에 영향을 미치는 변수의 비교를 위하여 세 교과 주관적 평가를 각각 반응변수로 하는 모형도 적합해 보았다.

‘전체 성적 주관적 평가(INT1E)’는 변인값이 ‘1(매우 우수)’, ‘2(우수)’, ‘3(약간 우수)’, ‘4(보통 수준)’, ‘5(약간 미흡)’, ‘6(미흡)’, ‘7(매우 미흡)’까지의 7단계로 응답하는 리커트 척도로 척도에 대한 응답수는 <표Ⅲ-5>와 같다. ‘4(보통수준)’의 응답수가 444명으로 전체 자료의 40% 이상을 차지하여 보통수준을 빼고 모형을 적합하는 것은 정보의 손실이 클 것으로 판단되었다. 따라서, 본 연구에서는 1(매우 우수) ~ 4(보통 수준)까지를 ‘1(우수)’로, 5(약간 미흡) ~ 7(매우 미흡)까지를 ‘0(미흡)’으로 코딩하였다. 학생이 자기보고식으로 응답한 자료에서 본인의 성적이 보통 수준이라고 응답한 것은 성적이 보통 이상은 된다고 판단하여 ‘1(우수)’로 코딩하였다. ‘1’로 코딩된 자료의 수는 ‘0’보다 3배 더 많았는데 진학을 희망하는 학생들로 자료를 추출하였기 때문인 것으로 판단된다.

〈표Ⅲ-5〉 전체 성적 주관적 평가 응답수

척도	1	2	3	4	5	6	7
	매우우수	우수	약간우수	보통수준	약간미흡	미흡	매우미흡
응답수	40	112	220	444	173	86	17

```

> TT6 <- K6
> TT6$ST <- ifelse(TT6$INT1E == 1 | TT6$INT1E == 2 |
+                 TT6$INT1E == 3 | TT6$INT1E == 4, 1, 0) #반응변수 코딩
> table(TT6$ST)

 0    1
276 816

# 0(미흡): 25%, 1(우수): 75%

```

[그림Ⅲ-1] 반응변수(INT1E) 코딩 R code

‘국어(INT1C01), 수학(INT1C02), 영어(INT1C03) 주관적 평가’ 변인값은 ‘1(매우 잘했다)’, ‘2(잘한 편이다)’, ‘3(보통이다)’, ‘4(못한 편이다)’, ‘5(매우 못했다)’의 5단계로 응답하는 리커트 척도이다. 응답수는 〈표Ⅲ-6〉에 제시하였다. 이분형 반응변수로 코딩하기 위하여 1(매우 잘했다) ~ 3(보통이다)까지를 ‘1(우수)’로, 4(못한 편이다) ~ 5(매우 못했다)를 ‘0(미흡)’으로 코딩하였다.

〈표Ⅲ-6〉 과목별(국어, 수학, 영어) 주관적 평가 응답수

척도	1	2	3	4	5
	매우 잘했다	잘한 편이다	보통이다	못한 편이다	매우 못했다
국어 응답수	64	275	558	175	20
수학 응답수	82	238	388	283	101
영어 응답수	75	214	467	292	44

국어 주관적 평가(INT1C01)는 ‘1(우수)’로 코딩된 자료는 897(82%), ‘0(미흡)’으로 코딩된 자료는 195(18%)이었다. 수학 주관적 평가(INT1C02)는 ‘1(우수)’로 코딩된 자료는 708(65%), ‘0(미흡)’으로 코딩된 자료는 384(35%)이었다. 영어 주관적 평가(INT1C03)는 ‘1(우수)’로 코딩된 자료는 756(69%), ‘0(미흡)’으로 코딩된 자료는 336(31%)이었다.

〈표Ⅲ-7〉 각 반응변수별 ‘1(우수)’와 ‘0(미흡)’의 비율

반응변수	‘1(우수)’	‘0(미흡)’
전체 성적 주관적 평가(INT1E)	816(75%)	276(25%)
국어 주관적 평가(INT1C01)	897(82%)	195(18%)
수학 주관적 평가(INT1C02)	708(65%)	384(35%)
영어 주관적 평가(INT1C03)	756(69%)	336(31%)

다. 설명변수(예측변수, 독립변수)

271개의 변수 중 반응변수로 사용된 변수를 제외한 270개의 변수를 설명변수로 사용하였다. 이분형 문항 15개는 다음과 같이 ‘0’과 ‘1’로 더미코딩하였다. 성별(GENDER)은 ‘남’은 ‘1’로, ‘여’는 ‘0’으로, 근로여부: 모친(JOB1B)은 ‘일을 하고 있다’는 ‘1’로 ‘일을 하고 있지 않다’는 ‘0’으로 코딩하였다. 비행: 연간 행동경험 유무 중 담배피우기(DLQ1A01), 술마시기(DLQ1A02)와 이성친구 유무(FRN3A), 사이버비행 연간 경험여부 중 불법 소프트웨어 다운받아 사용하기(MDA3A02), 체험활동 연간 참여경험 유무에서 건강/보건활동(ACT1A01), 문화/예술활동(ACT1A05), 봉사활동(ACT1A06), 직업체험 활동(ACT1A07),

자기(인성) 개발(ACT1A09), 동아리활동 연간 참여경험 유무 중 교내 공식 동아리활동(ACT2A01), 해외방문 연간 경험유무(ACT4A), 팬덤활동 선호 연예인/운동선수 유무(ACT6A) 문항은 ‘있다’는 ‘1’로, ‘없다’는 ‘0’으로 코딩하였다. 컴퓨터 사용여부(MDA1A)는 ‘사용한다’는 ‘1’로 ‘사용하지 않는다’는 ‘0’으로 코딩하였다.

범주형 변수 중 범주를 재구성하여도 맥락상 오류가 발생하지 않을 것으로 판단된 변수 3개의 응답을 0, 1로 다음과 같이 재구성하였다. 남녀공학 구분(COEDU)은 ‘남자학교’, ‘여자학교’를 ‘1’로, 나머지는 ‘0’으로 코딩하였다. 고등학교 유형(SCLTYP1)은 ‘일반 고등학교’, ‘자율형 고등학교’, ‘특수목적 고등학교(과학 외국어계열)’은 ‘1’로, 나머지는 ‘0’으로 코딩하였다. 휴대전화 종류(MDA2D)는 ‘스마트폰’은 ‘1’로 ‘피쳐폰’, ‘어떤 종류인지 모르겠다’는 ‘0’으로 코딩하였다. 자세한 R code는 [부록3]에 제시하였다.

코딩 후 {caret} 패키지에서 제공하는 nearZeroVar() 함수를 다시 사용하여 영분산이나 영근처 분산을 가지는 변수가 없음을 확인하였다.

```
> nzv<- nearZeroVar(TT6, saveMetrics = TRUE)
> summary(nzv)
```

freqRatio	percentUnique	zeroVar	nzv
Min. : 1.000	Min. :0.1832	Mode :logical	Mode :logical
1st Qu.: 1.272	1st Qu.:0.3663	FALSE:271	FALSE:271
Median : 1.770	Median :0.3663		
Mean : 2.441	Mean :0.4518		
3rd Qu.: 2.660	3rd Qu.:0.3663		
Max. :18.158	Max. :6.1355		

[그림 III-2] 영분산, 영근처 분산 가지는 변수 확인 R code

2. 분석 방법

LASSO는 제약 조건(constraint)의 성격 때문에 정확히 0인 계수를 생성하는 경향이 있어 변수의 선택을 통해 해석 가능한 모델을 생성하면서 능형 회귀(ridge regression)의 좋은 특성인 안정성을 유지한다(Tibshirani, 1996). 따라서, LASSO는 변수의 일부만을 포함하는 모델인 희소 모델(sparse model)을 생성한다. 별점함수의 도입은 다중공선성이 높은 데이터의 경우 편향을 줌으로써 분산의 감소를 통해 기존의 방법보다 더 나은 예측력을 갖게 할 수 있는 방법이다(김윤대 외, 2011). 별점회귀는 모형에서 특정 변수의 과도한 기여를 제한함으로써 즉, 큰 모수 값에 별점을 부여함으로써, 상관된 다른 변수가 동시에 모형에 포함될 수 있도록 허용하는 회귀방법이다(나종화, 2017b). 응답과 관련된 예측변수(predictor)의 수는 실제 데이터 셋의 경우에 가설이나 이론에 의해 선험적으로(priori) 알려져 있는 것이 아니다(James et al., 2013).

본 연구에서는 이러한 문제의식을 가지고 지금까지의 연구들에서 사용한 변수들뿐만 아니라 기존 연구에서 다루지 않았던 모든 변수들을 모형 적합에 사용할 수 있는 별점화 축소추정법 중 LASSO를 사용하여 모형 적합을 시도하였다.

가. 분석 과정

데이터 전처리 과정을 거쳐 분석에 적합하도록 데이터를 정제한 후, 반응변수를 이분형 변수로 코딩하였다. 설명변수는 이분형 변수는 더미 코딩하고 범주 재구성 및 응답 척도 수정을 하였다. R 프로그램의

{glmnet} 패키지를 사용하여 LASSO 로지스틱 회귀분석을 실시하였다.

반응변수가 주관적 학업성취도인데 ‘1(우수)’ 이냐 아니냐 하는 것이었으므로 로지스틱 회귀모형을 연구모형으로 이용하였다. 연구모형은 식 (1)과 같다. 이 때 β^T 는 설명변수에 대한 회귀계수의 벡터다.

$$\log \frac{P(G=1|X=x)}{P(G=0|X=x)} = \beta_0 + \beta^T X \quad (1)$$

LASSO는 변수의 척도에 민감하게 반응하므로, 코딩이 끝난 모든 변수를 표준화하여 추정하여야 하지만, R의 {glmnet} 패키지에서 모형 적합 전 표준화를 제공하고 있다(Hastie et al., 2009).

R의 {caTools} 패키지를 사용하여 모형 훈련에 사용할 훈련 데이터 셋(training data)과 모형 평가에 사용할 테스트 셋(test data)을 7:3의 비율로 분할하였다. 훈련 데이터 셋을 이용하여 {glmnet} 패키지에서 제공하는 glmnet() 함수로 모형을 적합하였다. 이때 주의할 점은 로지스틱 회귀를 분석모형으로 사용하기 때문에 family = “binomial” 옵션을 꼭 사용하여야 한다. 옵션으로 α 값을 지정해야 하는데, $\alpha = 1$ (디폴트)이 LASSO모형을 적합하는 것이다. 능형회귀를 적합할 때는 $\alpha = 0$ 으로 옵션을 바꾸어야 하고, elastic net에서는 $0 < \alpha < 1$ 인 α 를 사용한다.

LASSO 모형의 적합에서는 별점의 크기를 조절하는 조율모수(tuning parameter) λ 의 선택이 중요하다. 목적함수를 최소화하는 모수를 찾는 것으로 λ 가 0일 때는 최소제곱법의 결과와 동일하지만, λ 가 커질수록 목적함수를 최소화하는 회귀계수는 더 작은 값으로 축소(shrinkage)되어 추정된다. 자료 분석에서 적절한 λ 값의 선택을 위해 {glmnet} 패키지에

서 제공하는 `cv.glmnet()` 함수를 이용하여 교차타당화를 수행하였다. 옵션으로 제공되는 `type.measure`는 “dev”, “auc”, “class”, “mse”, “mae” 5종류¹⁶⁾이다. 본 연구에서는 훈련 데이터 셋으로 모델을 구축할 때, 이탈도(“dev”) 기준 10-fold 교차 타당화를 통해 예측오차를 가장 작게 하는 λ 의 값을 별점모수로 결정하였다. LASSO 로지스틱 회귀 모형 적합시 사용한 R 프로그램 code는 [그림 III-3]과 같다.

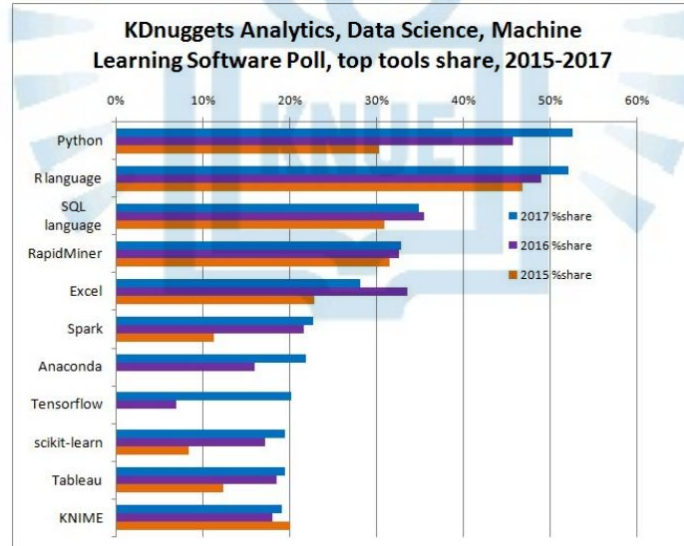
```
> library(caTools)
> set.seed(1) #데이터 셋 분할시 일정한 결과 얻기 위해 초깃값 지정
> split <- sample.split(TT6$ST, SplitRatio = 0.7)
> train <- subset(TT6, split == TRUE)
> test <- subset(TT6, split == FALSE)
> library(glmnet)
> x = model.matrix(ST ~.-1, data = train)
> y = train$ST
> fit = glmnet(x, y, family = "binomial")
> plot(fit, xvar = "lambda")
> plot(fit, xvar = "dev")
> set.seed(1) # 일정한 결과를 얻기 위해 초깃값 지정
> cvfit = cv.glmnet(x, y, family = "binomial", type.measure = "dev")
> plot(cvfit)
> s1 <- cvfit$lambda.min
> coef(fit, s = s1) # 변수 선택 및 회귀계수 출력
```

[그림 III-3] LASSO 로지스틱 회귀 모형 적합 R code

16) “dev”는 이탈도(deviance), “auc”는 AUC(area under the ROC curve), “class”는 오분류율(misclassification error), “mse”는 최소제곱오차(mean squared error), “mae”는 평균절대오차(mean absolute error)를 나타내는 척도이다.

나. 분석도구 R

본 연구에서 데이터 전처리 및 모형 적합, 평가 등은 통계 소프트웨어인 R을 통해서 구현하였으며, 사용한 패키지와 함수를 R코드와 함께 간략하게 소개하였다. R은 상업용 소프트웨어인 Splus를 기반으로 개발된 open source 프로그램으로 급증하는 데이터마이닝의 연구 추세에 대응하여 전 세계 연구자들에 의하여 개발된 강력한 분석도구로서 기능하고 있다(이재길, 2016). 미국의 KDnuggets 조사기관에서 데이터분석도구와 관련하여 “지난 12개월 동안 분석, 데이터마이닝, 데이터과학, 기계학습 프로젝트에 사용한 소프트웨어는 무엇입니까?” 라는 질문에 대한 투표 결과를 연중 발표하고 있다. [그림 III-4]와 같이 그 결과 R이 2016년에는 가장 인기 있었고, 2017년에는 두 번째로 인기 있는 분석도구였다.



[그림 III-4]¹⁷⁾ 2017년 가장 인기 있는 10대 분석도구

17) 출처: KDnuggets(<http://www.kdnuggets.com>)

3. 모형평가 기준

별점회귀모형에서는 통계적으로 유의한 회귀계수를 선택하는 것이 아니라, 모형 예측력을 최대화하는 방향으로 모형 선택이 이루어지기 때문에, 모형 예측에 기여하는 변수가 통계적으로 유의하지 않을 수 있다. 따라서 일반적인 회귀모형과 달리 상관관계나 결정계수 등으로 모형을 평가하지 않는다(유진은, 노민정, 2017).

LASSO 로지스틱 회귀모형으로 고등학교 3학년 주관적 학업성취도를 예측하고 모형을 평가하기 위하여 1,092명의 자료를 7:3의 비율로 나누어 각각을 훈련용 자료(training set)와 검증용 자료(testing data)로 이용하였다. 이 때 원자료에서의 주관적 학업 성취도 0과 1의 비율이 훈련용 자료와 검증용 자료에서도 유지되도록 두 집단으로 나눌 때 기준을 반응변수가 되도록 설정하였다. 모형평가를 위하여 검증용 자료를 이용하여 모형에 대입하여 추정치를 구하고 그 값과 실제값을 이용하여 정오분류표를 구하였다. R의 {caret} 패키지의 confusionMatrix() 함수를 이용하면 정분류율, 민감도, 특이도, Kappa 통계량 등을 구하는데 유용하다.

또 ROC 커브를 이용하여 AUC 값을 구하여 모형의 성능을 비교하였다. 연구자가 중요하게 생각하는 것을 기준으로 하여 민감도와 특이도를 조절할 수 있다. 이때 기준값(cutoff)에 따라 정분류율도 달라질 수 있는데 세 가지 조건이 잘 충족할 수 있도록 ROC 그래프를 활용할 수 있다.

마지막으로 R의 {caret} 패키지에서 제공하는 varImp() 함수를 이용하여 선택된 변수들의 중요도를 구할 수 있다. 앞서도 언급했듯이 별점회귀모형에서는 통계적으로 유의한 회귀계수를 선택하는 것이 아니라,

모형 예측력을 최대화하는 방향으로 모형 선택이 이루어지기 때문에 선택된 변수가 통계적으로 유의하지 않을 수 있다. 따라서 결과를 해석할 때, 얻어진 결과를 과장하지 않도록 주의해야 하고, 만들어질 수 있는 많은 예측 모델 가운데 하나로 여기고 추후 독립적인 변수 하나하나에 대한 검증이 필요하다.

마지막으로 검증용 자료로 정분류율(정확도), 민감도, 특이도를 구하였고, Kappa 통계량, AUC값을 구하여 모형을 평가하였다. 모형평가 및 중요도 출력을 위한 R code는 [그림 III-5]에 제시하였다.

```
> xx = model.matrix(ST ~.-1, data = test)
> yy = test$ST
> fit.pred <- predict(fit, s=s1, newx = xx, type = "class" )
  #s1은 교차타당화를 통한 최적의 별점모수값
> library(caret)
> confusionMatrix(data= fit.pred, yy, positive = '1')
# ROC 곡선과 AUC 값
> library(ROCR)
> fit.pred <- predict(fit, s=s1, newx = xx, type = "response" )
> ROC.pred = prediction(fit.pred, yy)
> ROC.pref = performance(ROC.pred, "tpr", "fpr")
> as.numeric(performance(ROC.pred, "auc")@y.values)
# variables importance
> importance <- varImp(fit, lambda = s1 )
> print(importance)
> order(importance, decreasing = TRUE) # 중요한 변수 순서로 배열
```

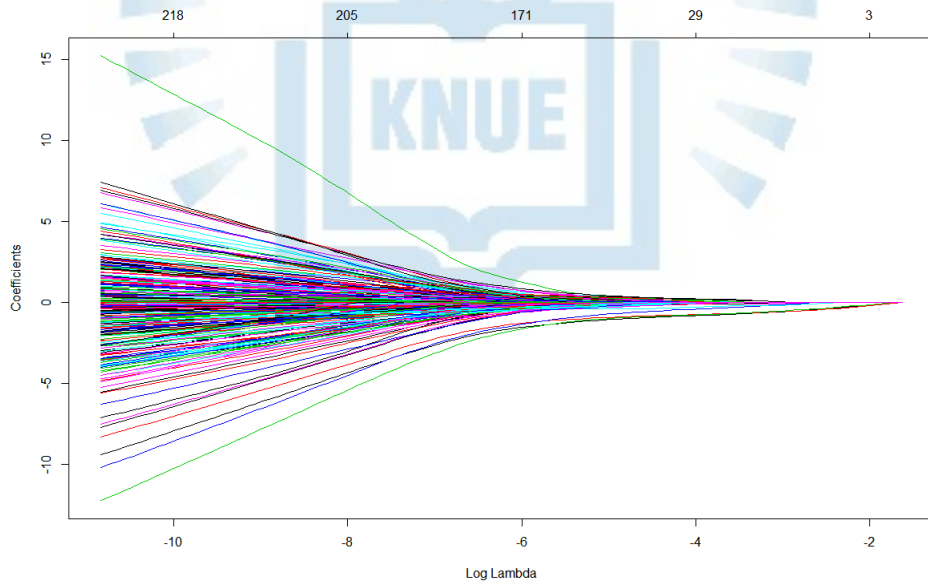
[그림 III-5] 모형 평가 및 변수의 중요도 출력 R code

IV. 연구 결과

1. LASSO 별점 회귀 모형 구축

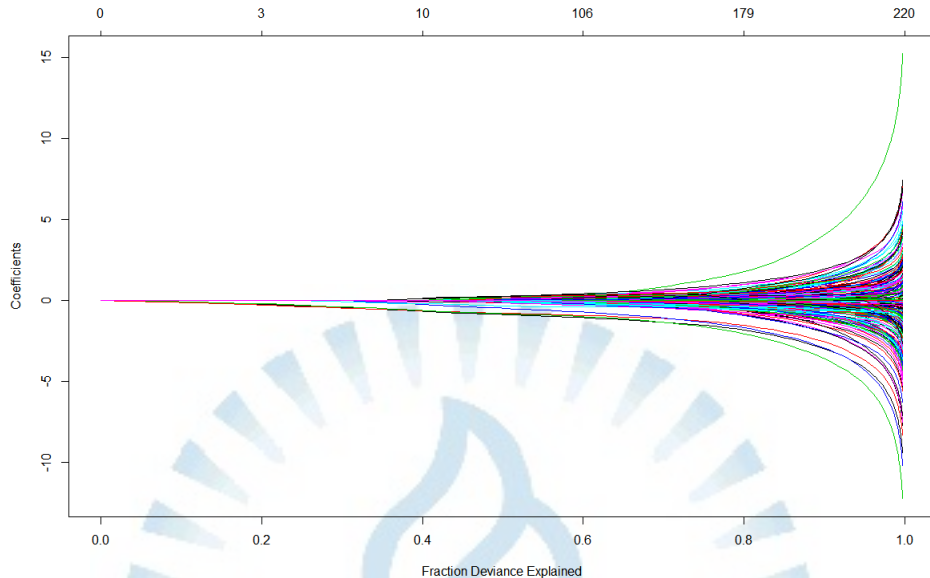
가. 전체 성적 주관적 평가를 반응변수로 한 모형

[그림 IV-1]은 KCYPS 자료에 대한 LASSO 계수 프로파일이다. 가로축은 별점모수에 자연로그를 취한 값이고 세로축은 그 때의 회귀계수 값이다. 각각의 선들은 270개 설명변수의 회귀계수가 별점모수(λ) 값이 증가함에 따라 0으로 축소되는 것을 보여준다. λ 값이 커지면 일부 계수가 급격히 0에 수렴하는 것을 확인할 수 있다. [그림 IV-1] 상자의 맨 위쪽에 표시된 숫자는 각 별점모수 값에 대한 설명변수의 개수를 나타낸다.



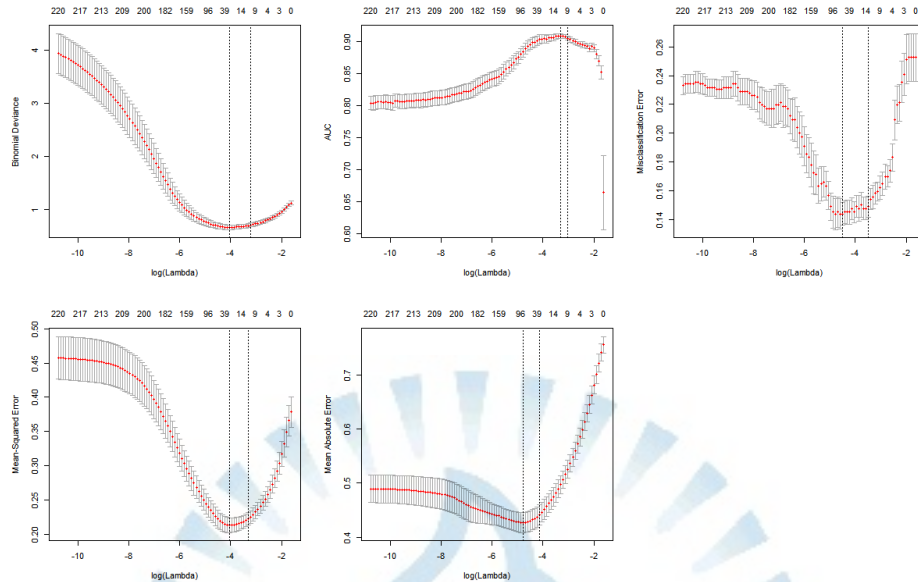
[그림 IV-1] 별점모수와 회귀계수

[그림 IV-2]는 이탈도와 회귀계수의 그래프를 나타낸 것이다.



[그림 IV-2] 이탈도와 회귀계수

다음은 별점모수 값을 찾기 위하여 교차 타당화(cross validation)를 이용하였다. [그림 IV-3]은 이탈도(deviance), AUC(area under the ROC curve), 오분류율(misclassification error), MSE(mean squared error), MAE(mean absolute error) 각각의 기준에 대한 10-fold 교차 타당화에 의해 LASSO 축소추정 양상을 보여주는 그래프이다. 가로축은 별점모수에 자연로그를 취한 값이고, 세로축은 각각의 기준에 대한 값을 나타낸다. [그림 IV-3] 상자의 맨 위쪽에 표시된 숫자는 선택되는 변수의 수를 나타낸 것이고, 수직으로 있는 두 개의 점선은 ‘1-standard error rule’ (1-표준오차 법칙)을 이용하여 각각의 기준에 의해 가장 작은 λ 값으로부터 1-표준오차에 해당되는 지점을 표시한다(유진은, 2017).



[그림 IV-3] 이탈도, UC, 오분류율, MSE, MAE 기준 10-fold CV 결과

LASSO에서는 특정한 데이터 셋에 가장 적합한 λ 값을 정하는 것이 매우 중요한데 `cv.glmnet()` 함수를 이용하면 10-fold 교차 타당화를 이용하여 교차타당오차를 최소로 하는 λ 를 알려준다. 이탈도(dev) 변화를 최소로 만들어주는 별점모수 λ 값 0.01754459를 이용하여 별점회귀모형을 적합하였다. 모형 적합 후, 그 모형에 검증용 자료를 이용하여 성능을 평가하였다. [그림 IV-4]는 ‘1’을 긍정(positive) 집단으로 할 때, {caret} 패키지의 `confusionMatrix()` 함수의 결과로 정분류율, 민감도, 특이도가 85.98%, 96.73%, 54.22%이라는 것을 보여준다. 원 데이터의 비율이 ‘1 (보통이상)’의 비율이 75%정도로 ‘0(못함)’의 3배정도를 차지하였기 때문에 민감도와 특이도의 차이가 크게 나타났다. 하지만 평균 정확도 (Balanced Accuracy)가 75.48%이므로 모형의 예측률은 높다고 할 수 있

다. 실제 예측을 위해 모형을 이용할 때는 연구자가 민감도와 특이도 중 중요하게 생각하는 값을 높이기 위해 기준값(cutoff)을 조정하여 활용할 수 있다. Kappa 통계량¹⁸⁾의 값은 0.5787로 0.4이상이면 일치도 수준이 적당하다고 볼 수 있다.

```
> confusionMatrix(data= fit.pred, yy, positive = '1')
Confusion Matrix and Statistics

      Reference
Prediction 0    1
0         45    8
1         38  237

      Accuracy : 0.8598
      95% CI   : (0.8174, 0.8955)
No Information Rate : 0.747
P-Value [Acc > NIR] : 4.388e-07

      Kappa : 0.5787
McNemar's Test P-Value : 1.904e-05

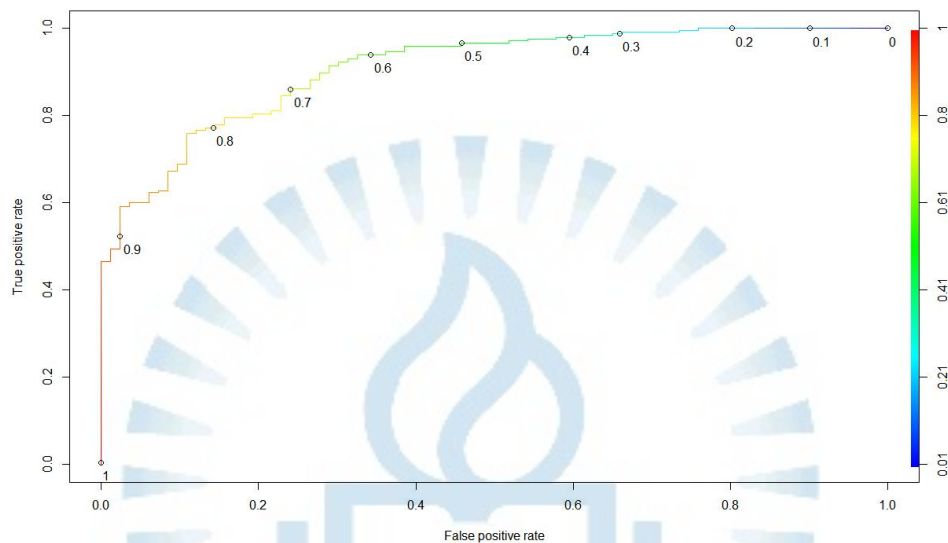
Sensitivity : 0.9673
Specificity : 0.5422
Pos Pred Value : 0.8618
Neg Pred Value : 0.8491
Prevalence : 0.7470
Detection Rate : 0.7226
Detection Prevalence : 0.8384
Balanced Accuracy : 0.7548

'Positive' Class : 1
```

[그림 IV-4] 정오분류표 모형 평가 결과

18) 카파 통계량(Kappa statistic)은 범주형으로 이루어지는 평가(또는 항목 분류)에서 여러 평가자들 간의 일치도의 신뢰성을 나타내는 통계적 척도이다. 0.4~0.6: 일치도 수준이 적당함, 0.6~0.8: 높은 일치, 0.8~1: 거의 완벽한 일치로 해석한다(나중화, 2017c).

[그림 IV -5]는 모형에 대한 ROC 곡선이다. 곡선 아래쪽의 넓이비인 AUC 값을 계산한 결과 0.907942이었다. AUC값은 0.5에서 1까지의 값을 가질 수 있는데, 1에 가까울수록 모형의 적합이 우수하다고 할 수 있다.

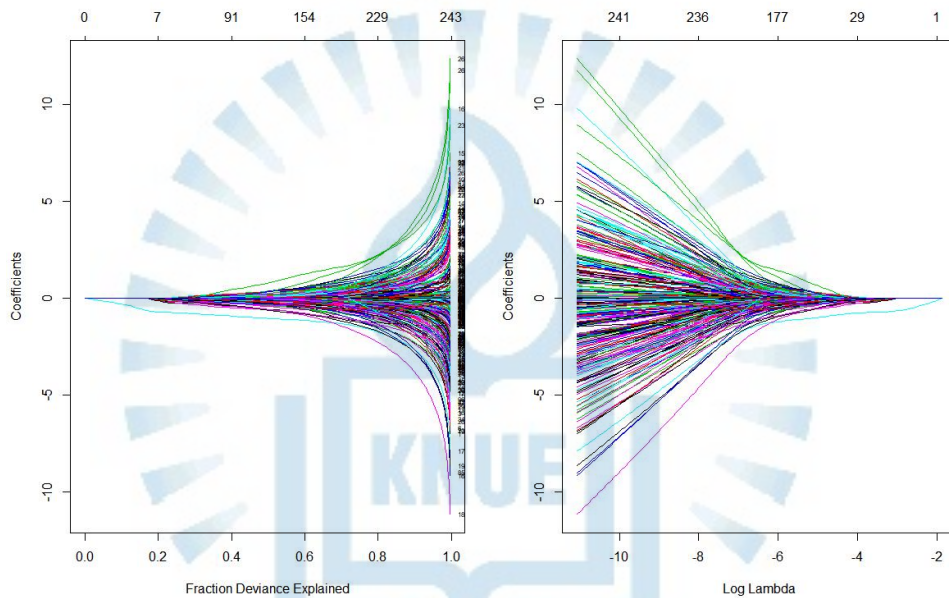


[그림 IV -5] ROC 곡선

모형 적합 과정에서 성적관련 변수인 과목별 국어, 수학, 영어 주관적 평가 변수와 성적 만족도 변수가 중요도가 높은 변수로 선택 되었다. 성적 관련 변수인 4개의 변수가 직접적으로 학생들의 성적을 나타내는 변수로 판단되어 제거 후, 설명변수 266개를 이용하여 모형을 적합해 보았다. 이탈도(dev)를 기준으로 얻은 별점모수 λ 값은 0.0202285였고 모형에 검증용 자료를 이용하여 모형을 평가한 결과, 정분류율, 민감도, 특이도는 80.49 %, 96.33%, 33.74%이었다. ROC 커브를 그려서 AUC 값을 계산한 결과 0.8091468이었다.

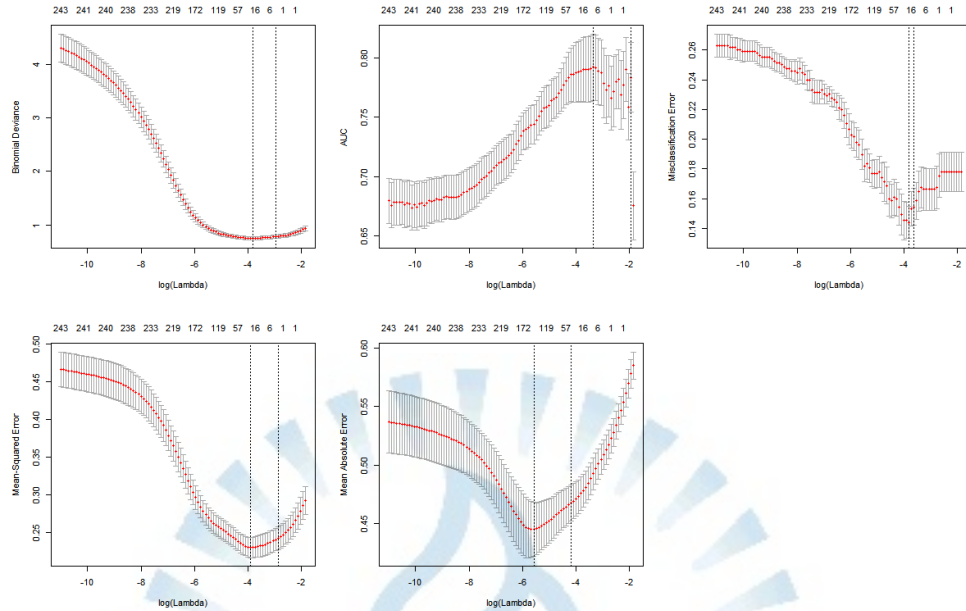
나. 국어 성적 주관적 평가를 반응변수로 한 모형

반응변수를 ‘국어 성적 주관적 평가(INT1C01)’로 하여 LASSO 별점 회귀 모형을 적합하였을 때 [그림 IV-6]은 LASSO 계수 프로파일을 나타내는 그래프로 순서대로 이탈도와 회귀계수, 별점모수와 회귀계수의 관계를 나타낸다.



[그림 IV-6] 이탈도와 회귀계수, 별점모수와 회귀계수

[그림 IV-7]은 이탈도(deviance), AUC(area under the ROC curve), 오분류율(misclassification error), MSE(mean squared error), MAE(mean absolute error) 각각의 기준에 의해 LASSO 축소추정 양상을 보여주는 그래프이다.



[그림Ⅳ-7] 이탈도, UC, 오분류율, MSE, MAE 기준 10-fold CV 결과

10-fold 교차 타당화를 이용하여 구한 최소의 별점모수 λ 값은 이탈도 (dev)를 기준으로 할 때 0.02207141였다. 적합된 모형에 테스트 셋을 이용하여 예측을 수행한 결과는 [그림Ⅳ-8]과 같다. 정분류율, 민감도, 특이도는 84.76%, 99.26%, 18.64%이었다. 민감도와 특이도의 비율이 차이가 많이 나는데, 원 데이터의 비율이 ‘1(보통이상)’의 비율이 82%, ‘0(못함)’의 비율이 18%로 두 집단의 비율이 4배 이상 차이가 나기 때문인 것으로 보인다. 기준값의 조정을 통해 민감도, 특이도의 조절이 가능하다. 평균정확도(Balanced Accuracy)는 58.95%이고, Kappa 통계량은 0.26으로 일치도 수준을 고려해 보아야 한다. ROC 커브를 그려서 AUC 값을 계산한 결과는 0.8238926이었다.

```

> confusionMatrix(data= fit.pred, yy, positive = '1')
      Confusion Matrix and Statistics

      Reference
Prediction  0   1
      0   11   2
      1   48 267

      Accuracy : 0.8476
      95% CI : (0.804, 0.8847)
      No Information Rate : 0.8201
      P-Value [Acc > NIR] : 0.1093

      Kappa : 0.2573
      Mcnemar's Test P-Value : 1.966e-10

      Sensitivity : 0.9926
      Specificity : 0.1864
      Pos Pred Value : 0.8476
      Neg Pred Value : 0.8462
      Prevalence : 0.8201
      Detection Rate : 0.8140
      Detection Prevalence : 0.9604
      Balanced Accuracy : 0.5895

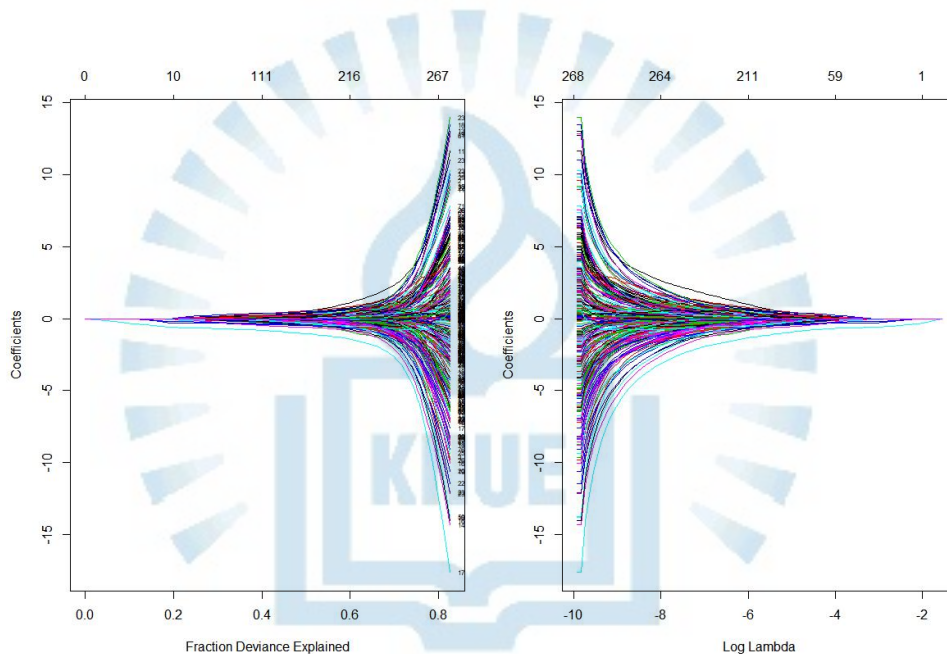
      'Positive' Class : 1
> library(ROCR)
> fit.pred <- predict(cvfit1, s=s1, newx = xx, type = "response" )
> ROC.pred = prediction(fit.pred, yy)
> ROC.pref = performance(ROC.pred, "tpr", "fpr")
> plot(ROC.pref)
> as.numeric(performance(ROC.pred, "auc")@y.values)
[1] 0.8238926

```

[그림 IV -8] 국어 성적 모형 평가 R code

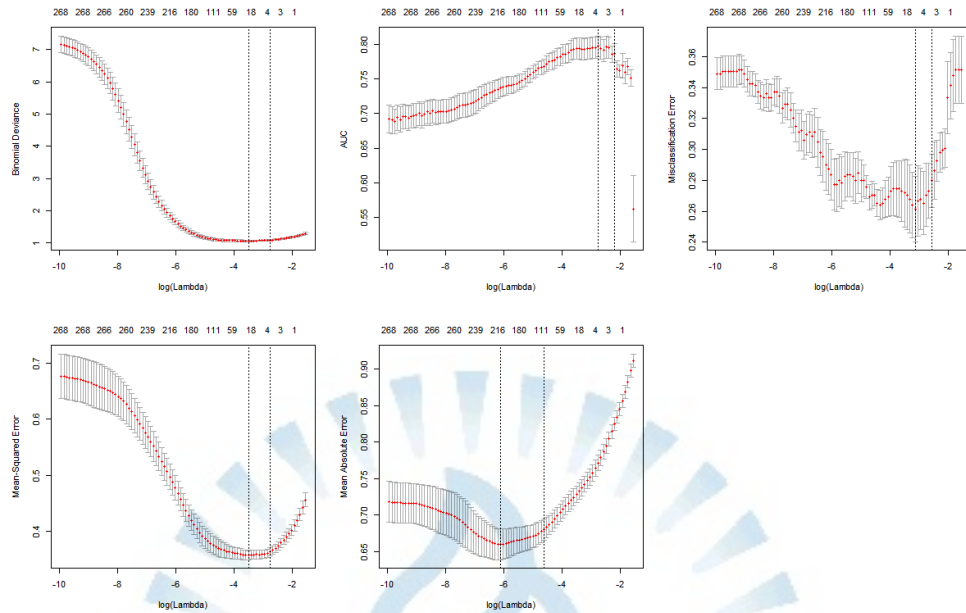
다. 수학 성적 주관적 평가를 반응변수로 한 모형

반응변수를 ‘수학 성적 주관적 평가(INT1C02)’로 하여 LASSO 별점 회귀 모형을 적합하였을 때 [그림 IV-9]은 LASSO 계수 프로파일을 나타내는 그래프로 순서대로 이탈도와 회귀계수, 별점모수와 회귀계수의 관계를 나타낸다.



[그림 IV-9] 이탈도와 회귀계수, 별점모수와 회귀계수

[그림 IV-10]은 이탈도(deviance), AUC(area under the ROC curve), 오분류율(misclassification error), MSE(mean squared error), MAE(mean absolute error) 각각의 기준에 의해 LASSO 축소추정 양상을 보여주는 그래프이다.



[그림 IV-10] 이탈도, UC, 오분류율, MSE, MAE 기준 10-fold CV 결과

10-fold 교차 타당화를 이용하여 구한 최소의 별점모수 λ 값은 이탈도 (dev)를 기준으로 할 때 0.03011595였다. 적합된 모형에 테스트 셋을 이용하여 예측을 수행한 결과는 [그림 IV-11]과 같다. 정분류율, 민감도, 특이도는 79.2%, 93.4%, 53.04%이었다. 평균정확도(Balanced Accuracy)는 73.22%이고, Kappa 통계량은 0.5045로 일치도 수준이 적당하다고 볼 수 있다. ROC 커브를 그려서 AUC 값을 계산한 결과는 0.8102133이었다.

```

> confusionMatrix(data= fit.pred, yy, positive = '1')
      Confusion Matrix and Statistics

      Reference
Prediction  0   1
      0  61  14
      1  54 198

      Accuracy : 0.792
      95% CI : (0.744, 0.8347)
      No Information Rate : 0.6483
      P-Value [Acc > NIR] : 1.017e-08

      Kappa : 0.5045
      McNemar's Test P-Value : 2.251e-06

      Sensitivity : 0.9340
      Specificity : 0.5304
      Pos Pred Value : 0.7857
      Neg Pred Value : 0.8133
      Prevalence : 0.6483
      Detection Rate : 0.6055
      Detection Prevalence : 0.7706
      Balanced Accuracy : 0.7322

      'Positive' Class : 1

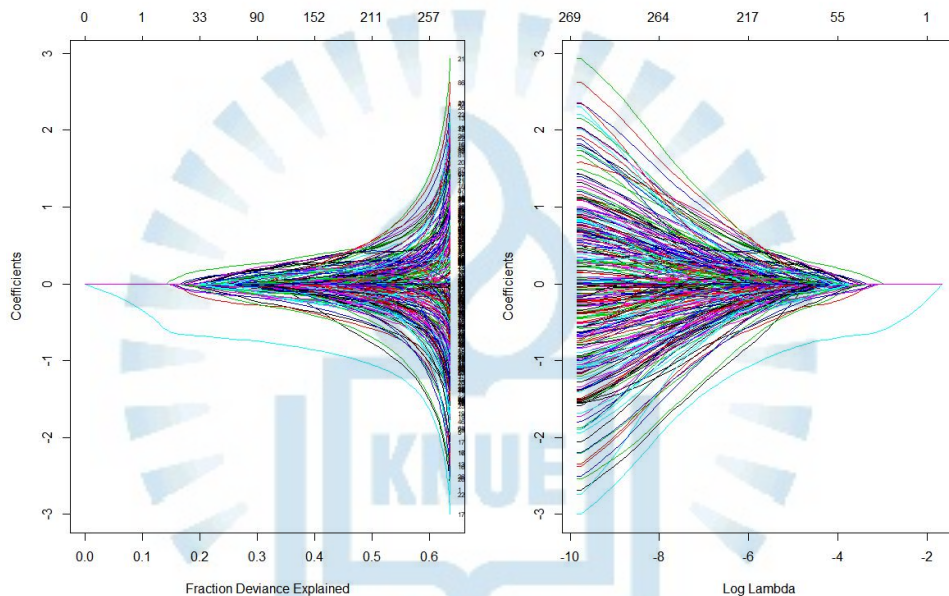
> as.numeric(performance(ROC.pred, "auc")@y.values)
[1] 0.8102133

```

[그림 IV-11] 수학 성적 모형 평가 R code

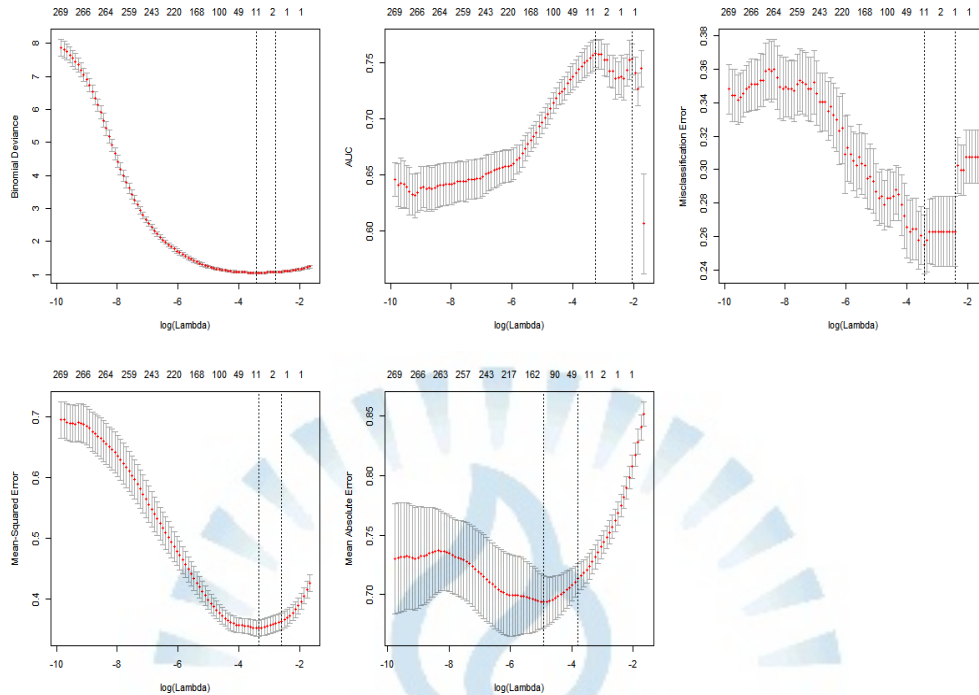
라. 영어 성적 주관적 평가를 반응변수로 한 모형

반응변수를 ‘영어 성적 주관적 평가(INT1C03)’로 하여 LASSO 별점 회귀 모형을 적합하였을 때 [그림 IV-12]은 LASSO 계수 프로파일을 나타내는 그래프로 순서대로 이탈도와 회귀계수, 별점모수와 회귀계수의 관계를 나타낸다.



[그림 IV-12] 이탈도와 회귀계수, 별점모수와 회귀계수

[그림 IV-13]은 모형의 성능평가를 위한 측도로 5가지 즉, 이탈도(deviance), AUC(area under the ROC curve), 오분류율(misclassification error), MSE(mean squared error), MAE(mean absolute error) 각각에 대해 LASSO 축소추정 양상을 보여주는 그래프이다.



[그림 IV-13] 이탈도, UC, 오분류율, MSE, MAE 기준 10-fold CV 결과

10-fold 교차 타당화를 이용하여 구한 최소의 별점모수 λ 값은 이탈도 (dev)를 기준으로 할 때 0.03250856였다. 적합된 모형에 테스트 셋을 이용하여 예측을 수행한 결과는 [그림 IV-14]와 같다. 정분류율, 민감도, 특이도는 77.44%, 95.59%, 36.63%이었다. 민감도와 특이도의 비율이 차이가 많이 나는데, 원 데이터의 비율이 ‘1(보통이상)’의 비율이 69%, ‘0(미흡)’의 비율이 31%로 두 집단의 비율이 2배 이상 차이가 나기 때문인 것으로 보인다. 평균정확도(Balanced Accuracy)는 66.11%이고, Kappa 통계량은 0.3784으로 일치도 수준을 고려해 보아야 한다. ROC 커브를 그려서 AUC 값을 계산한 결과는 0.7767697이었다.

```
> confusionMatrix(data= fit.pred, yy, positive = '1')
```

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	37	10
1	64	217

Accuracy : 0.7744

95% CI : (0.7253, 0.8185)

No Information Rate : 0.6921

P-Value [Acc > NIR] : 0.0005719

Kappa : 0.3784

Mcnemar's Test P-Value : 7.223e-10

Sensitivity : 0.9559

Specificity : 0.3663

Pos Pred Value : 0.7722

Neg Pred Value : 0.7872

Prevalence : 0.6921

Detection Rate : 0.6616

Detection Prevalence : 0.8567

Balanced Accuracy : 0.6611

'Positive' Class : 1

```
> as.numeric(performance(ROC.pred, "auc")@y.values)
```

[1] 0.7767697

[그림 IV-14] 영어 성적 모형 평가 R code

2. LASSO 별점 회귀 모형 결과 분석

가. 전체 성적 주관적 평가에 영향을 미치는 변수 분석

270개 설명변수를 모두 사용하여 이탈도(deviance)를 기준으로 LASSO 모형 적합 결과, 회귀계수가 0이 아닌 변수는 <표Ⅳ-1>과 같다. 270개 변수에서 모형에 포함된 변수는 개인발달 변수 18개, 발달환경 변수 11개로 모두 29개로 정리되었다. 배경변인 변수는 하나도 포함되지 않았다.

먼저 개인발달 영역으로 성적에서 국어, 수학, 영어 과목별 주관적 평가와 전체 성적 만족도, 학습습관에서 숙달목적 지향성과 행동통제, 정서문제에서 주의집중과 우울, 자아인식에서 자아탄력성과 자아정체감, 진로계획에서 장래 희망 교육수준, 진로관에서 대학진학이유와 직업선택이유, 학업관련 시간에서 등교일 학원/과외숙제 시간이 모형에 포함되었다. 국어, 수학, 영어 과목별 주관적 평가, 전체 성적 만족도가 높을수록, 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아할수록 학업성취도가 높은 것으로 나타났다. 특히 학습습관에 대한 행동통제 변수는 3문항이 선택되어 행동통제를 잘 할수록 학업성취도가 더 높았다. 이것은 공부가 지루하고 재미없더라도 수능을 준비하거나 내신을 위한 시험공부를 할 때 경험하는 엄청난 지루함을 이겨내는데, 자제력이 도움을 주기 때문일 것이다. 그리고 아마도 자제력이 강한 학생은 시험을 빨리 끝내고 싶은 충동에 사로잡혀 문제를 읽은 뒤 가장 먼저 머릿속에 떠오르는 답을 답안지에 적고 곧바로 다음 문제로 옮겨가는 대신에, 한 문제 한 문제에 대해 가능한 최선의 해답을 찾았다는 확신이 들 때까지 한눈을 팔지 않기 때문일 것이다(Matthew, 2013). 월터 미셸(Walter

Mischel)의 유명한 마시멜로와 종을 이용한 실험에서 네 살 때 더 오래 기다렸던 아이들은 수능시험(SAT)에서도 더 높은 점수를 받았고, 특히 실험자가 알아서 제 발로 돌아올 때까지 기다렸던 아이들은, 30초밖에 못 기다렸던 아이들보다 수능시험(SAT)에서 200점 이상 높은 점수를 받은 결과에서도 확인할 수 있다(Lehrer, 2009). 안젤라 덕워스(Angela Duckworth)가 조사한 바에 따르면, 내신 성적은 지능지수보다 욕구 충족을 미룰 수 있는 능력에 의해 더 잘 예측되는 것으로 나타났다(Duckworth & Seligman, 2005). 국내의 선행연구에서도 학습태도가 학업성취에 영향을 미치는 중요한 변수임을 보고한 연구결과와 일치한다(김경식, 이현철, 2009; 김서현 외, 2015; 노현중, 손원숙, 2015; 박은정 외, 2015; 성은모, 진성희, 2012; 손병덕 외, 2016; 이수진 외, 2014; 이유리 외, 2015; 임선아, 2014; 한희원, 문경숙, 2014).

정서문제로 주의집중에서 연필이나 지우개 등, 학용품을 잘 안 잃어버릴수록, 우울에서 울기를 잘하지 않을수록 학업성취도가 높은 것으로 나타났다. 정서적 경험과 지적 수행의 관계에 대한 연구에서 긍정적 정서는 사고 및 의사결정 능력의 향상과 관련이 되며 작업기억 능력도 향상시키는 것으로 나타났다(Matthew, 2013). 그러나 김지혜(2015)는 우울이 학업성취에 미치는 직접적인 영향력은 통계적으로 유의하지 않다고 보고하였고, 금지현 외(2013)는 정서문제가 학업성취도에 직접적인 영향은 미치지 않으나 매개적인 역할을 한다고 보고하였다. 즉, 학업성취도가 높은 학생들은 주의집중을 잘하고, 우울한 성향이 낮은 정서적인 특징을 갖는 것으로 해석할 수 있다.

자아인식으로 자아탄력성에서 같은 장소를 갈 때도 늘 가던 길보다는 다른 길로 가 보는 것을 좋아하지 않을수록, 자아정체감에서 타인의 말

과 행동에 영향을 쉽게 받지 않을수록, 낯선 사람을 만나는 것을 꺼려할수록 학업성취도가 높은 것으로 나타났다. 자아탄력성이 높을수록 학업성취도가 높다는 기존연구(권지은, 2003; 임효진, 김재철, 2014; 조한익, 이나영, 2010)와 상반된 결과이나, 자아탄력성이 자아통제(ego-control)와 역전된 U-모양의 관계 패턴을 보인다는 연구결과와는 다소 일치하는 결과이다(Asendorpf & Aken, 1999; Block & Kermen, 1996).

6차 조사에 새롭게 조사된 문항인 진로계획 영역에서 장래 희망 교육수준, 대학진학이유, 직업선택 요인이 포함되었다. 장래 희망 교육수준이 높을수록, 대학 진학 이유로 고졸 학력으로는 취업이 어려워서와 부모님/가족이 원하기 때문이라고 생각할수록, 직업선택 요인으로 국가/사회에 기여할 수 있다는 것을 중요하게 생각할수록 학업성취도가 높은 것으로 나타났다. 자신의 미래에 대한 포부나 진로계획이 학생의 내적인 동기를 유발하여 학습태도에 긍정적인 효과를 준다는 연구결과와 일치한다(김경근 외, 2014; 김경식, 이현철, 2009). 여학생의 경우 가족성공 추구형이 학업성취가 가장 높게 나타났는데 대학 진학 이유 중 부모님/가족이 원하기 때문이라고 생각한 학생들의 학업성취도가 높게 나온 것과 연결된다(박은정 외, 2015). 진로선택 가치로 사회공헌을 이유로 희망직업을 선택했다는 응답자들이 다른 가치들을 선택한 학생들에 비해 내재적 조절동기를 높여 학업성취와 긍정적인 관계를 가진다는 연구결과와 일치한다(박수원, 최병호, 2017).

학업관련 시간에 대한 학원/과외숙제시간은 등교일 기준으로 많을수록 학업성취도가 높은 것으로 나타났다. 고등학생의 학업성취도를 높이기 위해서는 학생 스스로 자기 속도에 맞춘 학습에 시간을 할애하여야 한다는 연구(허은정, 이재덕, 2014)와 스스로 공부하는 시간을 많이 가질수

록 학업성취도가 높아진다는 것, 목표달성을 위해 투입하는 학습의 절대적인 양이 학업성취를 결정하는 중요한 요인임을 보고한 연구(김경근 외, 2014)와 부합된다. 고등학생의 경우 교과공부와 직접적인 연관성이 적은 활동을 최소화하는 방향으로 시간 운용을 하는 것이 학업성취 제고에 도움이 된다고 볼 수 있다(김경근 외, 2014).

다음은 발달환경 영역으로 양육방식에서 합리적 설명과 방임, 학교적응에서 학습활동과 교사관계, 지역사회관에서 공동체의식, 매체환경에서 컴퓨터 사용여부와 사용장소, 휴대전화 의존도가 포함되었다. 체험활동, 동아리 활동의 연간 참여경험 유무와 문화 활동 연간 경험횟수도 포함되었다.

양육방식에서 합리적 설명 문항의 결과는 기존 연구(박경민 외, 2013)와 다른 양상을 보이는데 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 설명해 주지 않을수록 학업성취도가 높은 것으로 나타났다. 고등학교 3학년이라는 학령기의 특성으로 인해 부모의 양육태도가 학업성취에 미치는 영향이 다르게 나타나기도 한다는 연구결과와 일치한다(권지은 외, 2013; 손병덕 외, 2016). 방임과 관련해서는 다른 일(직장이나 바깥일)보다 나를 더 중요하게 생각하실수록 학업성취도가 높게 나타났다. 부모가 자녀를 더 중요하게 생각할수록 교육지원행동이 증가하고, 더 많은 부모의 교육지원행동과 교육 기대, 부모의 지원과 격려는 자녀의 학업적 자기효능감을 높여 학업성취가 증가한다는 연구결과와 부합된다(금지현 외, 2013; 임다희, 권기현, 2012; 정제영, 정예화, 2015; 추상엽, 임성문, 2008).

학교적응으로 학습활동에서 수업시간에 배운 내용을 잘 알고 있을수록, 공부 시간에 딴 짓을 하지 않을수록 학업성취도가 높게 나타났다.

학업 성적을 올리려면 학교수업에 충실하고 자기주도적 학습능력을 향상시키는 학습태도가 중요하다는 기존 연구결과와 부합된다(김경식, 이현철, 2009). 한편 교사관계에서 학교 밖에서 선생님을 만나면 반갑지 않을수록 학업성취도가 높은 것으로 나타났다. 교사요인이 학생의 학업성취도에 미치는 영향은 거의 없고 관련 선행연구 검토에서도 교사 효과가 미미하다고 보고한 연구(김양분, 임현정, 김난옥, 2012)와 지각된 교사태도는 학교생활만족도와 학습태도에 직접적인 영향을 미치며, 학습태도를 통해 학교생활만족도와 학업성취도의 향상에 간접적인 영향을 미치는 것으로 보고된 연구(강명희, 유영란, 유지원, 2014)도 있다. 기존의 자료를 활용하였으므로 제한된 문항으로 변인을 측정할 수밖에 없다는 현실적인 한계를 지니고 있다. 따라서 추후연구에서는 교사관계를 설명할 수 있는 다양한 변인들과 관련 변인을 보다 포괄적으로 측정할 수 있는 방법을 사용하여 연구할 필요가 있다.

봉사활동과 관련하여 공휴일에 복지기관에서 자원봉사 활동을 하지 않을수록, 봉사활동 연간 참여경험이 없을수록, 컴퓨터 사용을 하지 않을수록 학업성취도가 높은 것으로 나타났다. 이것은 고등학교 3학년의 경우 입시준비로 인하여 교과공부와 직접적인 연관성이 적은 활동을 최소화하는 방향으로 시간 운용을 하는 것만이 학업성취를 높이는 방법(김경근 외, 2014)임을 나타내는 것으로 판단된다. 하지만 고등학교 3학년의 삶이 지독한 경쟁 속에서 공동체를 잊고 자기 자신의 목표만을 향해 나아가고 있는 것은 아닌지 돌아볼 필요가 있다.

휴대전화 의존도에서는 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없을수록 학업성취도가 높게 나타났다. 휴대전화 의존이 높을수록 오락적 휴대전화 이용의 빈도가 높을수록 학업성취에는 유의한 부정적

효과가 발생함을 보고한 연구결과와 상반되는 결과이다(김지혜, 2015; 성은모, 진성희, 2012; 이승진, 김경근, 2015). 하지만 본 연구에서 반응 변수의 코딩 단계에서 ‘4(보통수준)’을 ‘1’에 포함시켜 변수선택을 했기 때문에 나온 결과로 분석되었다¹⁹⁾. 즉, 휴대전화 의존도가 높은 것은 보통수준인 학생들의 특징으로 해석할 수 있다.

교내 공식 동아리활동을 참여할수록 학업성취도가 높게 나타난 것은 고등학교에서 다양한 학습, 진로 관련 동아리를 운영하고 있어 학습에 도움을 받는 것이라고 판단된다. 문화활동 연간 경험횟수가 많을수록 학업성취도가 높게 나타났다.

<표Ⅳ-1> 회귀계수가 0이 아닌 29개 설명변수와 척도(270개 변수사용)

변수	변수설명	코딩	계수
INT1C01	과목별 주관적 평가: 국어	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.834
INT1C02	과목별 주관적 평가: 수학	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.770
INT1C03	과목별 주관적 평가: 영어	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.792
INT1D	전체 성적 만족도	Likert: 1(매우 만족한다) ~ 4(전혀 만족하지 않는다)	-0.432
INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.118

19) 보통수준으로 응답한 학생수가 1092명 중 444명으로 40%를 넘는 데이터의 손실을 막기 위하여 본 연구에서는 ‘4(보통수준)’를 ‘1’로 코딩하여 분석하였다. 그러나 분석과정에서 보통수준 학생을 빼고 분석해 보는 것도 의미가 있을 것으로 판단되어 보통수준으로 응답한 학생을 빼고 1(매우 우수) ~ 3(약간 우수)로 응답한 학생을 ‘1(우수)’, 5(약간 미흡) ~ 7(매우 미흡)으로 응답한 학생을 ‘0(미흡)’으로 코딩하여 모형을 적합하였다. 결과는 [부록4]에 제시하였다.

INT2C01	학습습관_행동통제: 나는 공부가 지루하고 재미없더라도 끝까지 다한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.118
INT2C02	학습습관_행동통제: 나는 하던 공부를 끝낼 때까지 공부에 집중한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.117
INT2C05	학습습관_행동통제: 나는 공부를 하려면 쓸데없는 생각 때문에 집중을 못한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.162
PSY1A04	정서문제_주의집중: 연필이나 지우개 등, 학용품을 잘 잃어버린다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.016
PSY1E05	정서문제_우울: 울기를 잘한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.031
PSY2B07	자아인식_자아탄력성: 나는 같은 장소에 갈 때도 늘 가던 길 보다는 다른 길로 가 보는 것을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.083
PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.092
PSY2C07	자아인식_자아정체감: 나는 낯선 사람을 만나는 것을 꺼려한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.056
FUR1A01	진로계획_장래 희망 교육수준	Likert: 1(고등학교 졸업) ~ 5(대학원 박사학위 취득)	0.219
FUR2A01	대학진학이유: 고졸 학력으로는 취업이 어려워서	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.125
FUR2A03	대학진학이유: 부모님/가족이 원해서	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.004
FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.117
TIM4A	학업관련시간_학원/과외숙제시간: 등교일	#시간으로 응답	0.001
FAM2F03	양육방식Ⅰ_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안되는지 그 이유를 말씀해 주신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.098
FAM3A01	양육방식Ⅱ_방임: 다른 일(직장이나 바깥일)보다 나를 더 중요하게 생각하신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.038

EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.304
EDU2A05	학교적응_학습활동: 공부 시간에 딴 짓을 한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.189
EDU2D03	학교적응_교사관계: 학교 밖에서 선생님을 만나면 반갑다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.019
COM2A02	공동체의식: 공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.100
MDA1A	컴퓨터_사용여부	1(사용한다), 0(사용하지 않는다)	-0.002
MDA2C05	휴대전화_의존도: 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.043
ACT1A06	체험활동_연간 참여경험 유무: 봉사활동	1(있다), 0(없다)	-0.017
ACT2A01	동아리활동_연간 참여경험 유무: 교내 공식 동아리활동	1(있다), 0(없다)	0.047
ACT3B	여행/문화활동: 문화활동 연간 경험횟수	#회로 응답	0.008

* 변수는 KCYPS 자료 입력 순으로 정리되었다.

변수의 중요도를 알아보기 위하여 R의 {caret} 패키지에서 varImp() 함수를 이용하였다. R code는 [그림 IV-15]에서 제시하였고, 변수의 중요도 결과는 <표 IV-2>에서, 시각화는 [그림 IV-16] 차트로 제시하였다.

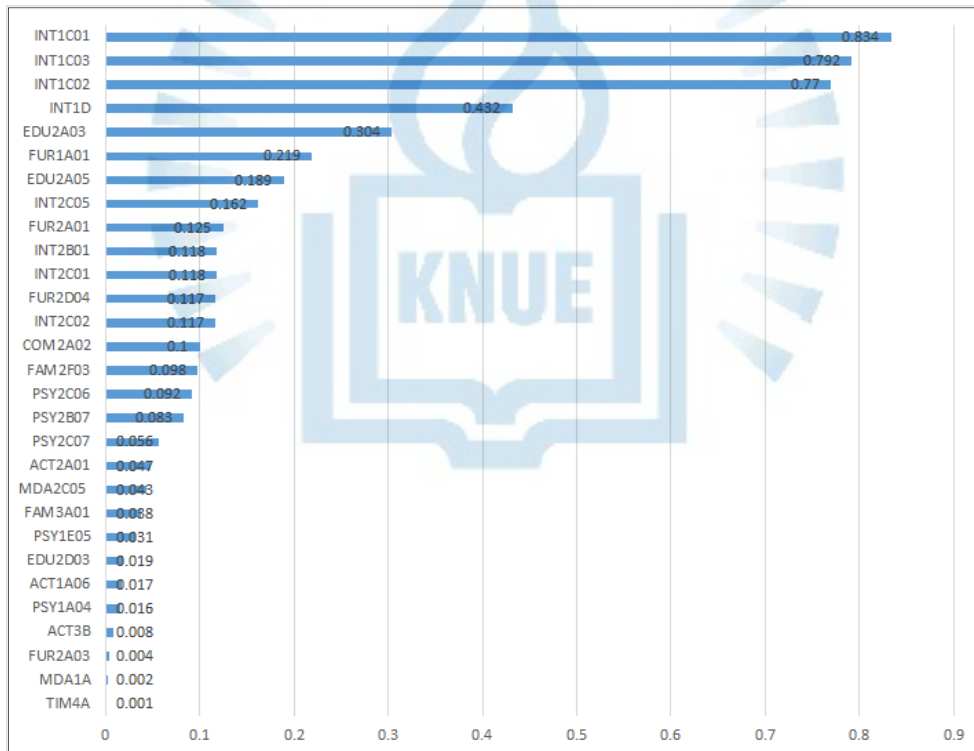
```
> importance <- varImp(fit, lambda = s1 )
> order(importance, decreasing = TRUE) # 중요한 변수 순서로 배열
[1] 15 17 16 18 208 116 210 32 118 26 28 135 29 233
186 102 89 103 266 256 187 67 223 263 40 268 120 241 150
중략
```

[그림 IV-15] 변수의 중요도 출력 R code

〈표Ⅳ-2〉 중요도 순 변수 정리(270개 변수 사용)

순	변수	변수설명 (결과)	순	변수	변수설명 (결과)
1	INT1C01	과목별 주관적 평가: 국어 (잘 했다고 할수록)	16	PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다 (그렇지 않을수록)
2	INT1C03	과목별 주관적 평가: 영어 (잘 했다고 할수록)	17	PSY2B07	자아인식_자아탄력성: 나는 같은 장소에 갈 때도 늘 가던 길 보다는 다른 길로 가보는 것을 좋아한다 (그렇지 않을수록)
3	INT1C02	과목별 주관적 평가: 수학 (잘 했다고 할수록)	18	PSY2C07	자아인식_자아정체감: 나는 낯선 사람을 만나는 것을 꺼려한다 (그럴수록)
4	INT1D	전체 성적 만족도 (만족할수록)	19	ACT2A01	동아리활동_연간 참여경험 유무: 교내 공식 동아리활동 (있을수록)
5	EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다 (그럴수록)	20	MDA2C05	휴대전화_의존도: 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없다 (그럴수록)
6	FUR1A01	진로계획_장래 희망 교육수준 (희망 교육수준이 높을수록)	21	FAM3A01	양육방식Ⅱ_방임: 다른 일(직장이나 바깥일)보다 나를 더 중요하게 생각하신다 (그럴수록)
7	EDU2A05	학교적응_학습활동: 공부 시간에 딴 짓을 한다 (그렇지 않을수록)	22	PSY1E05	정서문제_우울: 울기를 잘한다 (그렇지 않을수록)
8	INT2C05	학습습관_행동통제: 나는 공부를 하려면 쓸데없는 생각 때문에 집중을 못한다 (그렇지 않을수록)	23	EDU2D03	학교적응_교사관계: 학교 밖에서 선생님을 만나면 반갑다 (그렇지 않을수록)
9	FUR2A01	대학진학이유: 고졸 학력으로 는 취업이 어려워서 (그럴수록)	24	ACT1A06	체험활동_연간 참여경험 유무: 봉사활동 (없을수록)
10	INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다 (그럴수록)	25	PSY1A04	정서문제_주의집중: 연필이나 지우개 등, 학용품을 잘 잃어버린다 (그렇지 않을수록)
11	INT2C01	학습습관_행동통제: 나는 공부 가 지루하고 재미없더라도 끝까지 다한다 (그럴수록)	26	ACT3B	여행/문화활동: 문화활동 연간 경험횟수 (많을수록)

12	FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다 (중요할수록)	27	FUR2A03	대학진학이유: 부모님/가족이 원해서 (그럴수록)
13	INT2C02	학습습관_행동통제: 나는 하던 공부를 끝낼 때까지 공부에 집중한다 (그럴수록)	28	MDA1A	컴퓨터_사용여부 (사용하지 않을수록)
14	COM2A02	공동체의식: 공휴일에 쉬지 못 하더라도 복지기관에서 자원 봉사 활동을 할 수 있다 (그렇지 않을수록)	29	TIM4A	학업관련시간_학원/과외숙제 시간: 등교일 (많을수록)
15	FAM2F03	양육방식 I_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주신다 (그렇지 않을수록)	공 란		



[그림 VI-16] 변수의 중요도순 차트

성적관련 변수 4개가 중요도 4위 안에 모두 포함된 것을 확인하고 성적관련 변수 4개를 삭제 후, 266개 변수로 LASSO 모형을 적합하였다. 결과는 <표Ⅳ-3>와 같이 41개 변수가 선택되었고, 배경변인 변수 2개, 개인발달 변수 22개, 발달환경 변수 17개로 정리되었다. LASSO는 많은 상관성 있는 변수들 중에서 하나의 변수만을 선택한다는 점이 특징이면서 단점이라고 할 수 있는데 이런 특징으로 인하여 4개의 변수를 삭제한 후 모형을 적합한 결과, 같은 영역인 경우 다른 변수가 선택되기도 하였다.

더 선택된 변수 중심으로 살펴보면 배경변인에서 모친의 최종학력이 높을수록, 보호자 특성에서 삶의 만족도가 높을수록 학업성취도가 높게 나타났다. 기존 연구에서 가정배경이 학업성취도에 미치는 영향이 크다는 연구결과들과 일치한다(곽수란, 2004; 김재철, 2004; 김현진, 2007; 심도현 외, 2014; 이중섭, 이용교, 2011; 이승진, 김경근, 2015; 임다희, 권기현, 2012).

개인발달 영역에서 체격에는 몸무게가 적게 나갈수록, 학습습관 영역에는 성취가치 변수 중 학교생활이 미래에 상당한 역할을 할 것으로 생각할수록, 학업시간 관리 변수 중 시험 계획을 세우고 그에 따라 공부를 할수록 학업성취도가 높은 것으로 나타났다. 비만이 학업성취도에 부정적인 영향을 준다는 연구결과와 일치한다(배호중, 유비, 2015; 서지영, 2015; 이자형, 이기혜, 2015).

정서문제 영역에서는 공격성 변수 중 별 것 아닌 일로 싸우지 않을수록, 자아인식으로 자아존중감에서 남들만큼의 일은 할 수 있다고 생각할수록 학업성취도가 높은 것으로 나타났다. 학업에 성공적인 학생들이 대체로 자신에 대해 좋은 평가를 내린다는 연구결과와 일치한다(곽수란,

2006).

발달환경 영역으로 양육방식 중 과잉간섭에서 나에게 무엇을 하든지 항상 이겨야 한다는 걸 강조하지 않고, 방임에서 내가 많이 아프면 적절한 치료를 받게 하지 않을수록 학업성취도가 높은 것으로 나타났다. 270개 변수를 모두 설명변수로 사용하였을 때 방임과 관련하여 다른 일보다 나를 더 중요하게 생각하실수록 학업성취도가 높게 나타난 것과 다른 결과이다. 보통수준으로 응답한 자료를 제외하고 재코딩한 모형에서도 같은 결과가 나왔는데 이것은 나를 중요하게 생각하는 것은 학업성취에 긍정적인 영향을 미치나, 아플 때 적절한 치료를 받게 하는 것은 부정적인 영향을 미치는 것으로 해석할 수 있다. 아마도 고등학교 3학년이라는 학령기의 특징이 반영된 결과라고 판단된다.

학교적응에는 학교규칙 중 휴지나 쓰레기를 버릴 때 꼭 휴지통에 버리지 않을수록, 교우관계 중 친구와 다투었을 때 먼저 사과하지 않을수록, 놀이나 모둠활동을 할 때 친구들이 내 말을 잘 따라줄수록 학업성취도가 높게 나타났다. 하지만 보통수준으로 응답한 자료를 제외하고 재코딩한 모형에서는 학교규칙과 교우관계 변수가 포함되지 않았다. 보통수준 학생의 특징으로 판단된다. 그러나 학교규칙 중 휴지나 쓰레기 버리는 문항과 교우관계에서 친구와 다투었을 때 먼저 사과하는 문항은 추후 연구가 더 필요한 부분이다.

매체환경에서는 휴대전화를 음악 듣는 용도로 자주 사용하지 않을수록, 성인용 매체를 이용하지 않을수록 학업성취도가 높게 나타났다. 팬덤활동에는 선호 연예인/운동선수가 없을수록 학업성취도가 높은 것으로 나타났다.

〈표 IV-3〉 회귀계수가 0이 아닌 29개 설명변수와 척도(270개 변수사용)

변수	변수설명	코딩	계수
HAK2B	최종학력: 모친	Likert: 1(중졸 이하) ~ 5(대학원 졸)	0.106
PAR2	보호자특성: [보호자]삶의 만족도	Likert: 1(매우 만족한다) ~ 4(전혀 만족하지 못한다)	-0.064
PHY1B	체격: 몸무게	#kg으로 응답	-0.001
INT2A04	학습습관_성취가치: 학교생활이 나의 미래에 상당한 역할을 할 것이다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.071
INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.112
INT2C01	학습습관_행동통제: 나는 공부가지루하고 재미없더라도 끝까지 다 한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.461
INT2C02	학습습관_행동통제: 나는 하던 공부를 끝낼 때까지 공부에 집중한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.181
INT2C05	학습습관_행동통제: 나는 공부를 하려면 쓸데없는 생각 때문에 집중을 못한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.171
INT2D04	학습습관_학업시간 관리: 나는 시험 전에 계획을 세우고 그에 따라 공부한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.089
PSY1A05	정서문제_주의집중: 주의를 기울이지 않아서 실수를 하거나 사고를 낸다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.009
PSY1B04	정서문제_공격성: 별 것 아닌 일로 싸우곤 한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.006
PSY1E05	정서문제_우울: 울기를 잘한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.088
PSY1E06	정서문제_우울: 어떤 일이 잘못 되었을 때 나 때문이라는 생각을 자주 한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.014
PSY2A04	자아인식_자아존중감: 나는 남들만큼의 일은 할 수 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.123

PSY2B07	자아인식_자아탄력성: 나는 같은 장소에 갈 때도 늘 가던 길 보다는 다른 길로 가 보는 것을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.068
PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.191
FUR1A01	진로계획_장래 희망 교육수준	Likert: 1(고등학교 졸업) ~ 5(대학원 박사학위 취득)	0.328
FUR2A03	대학진학이유: 부모님/가족이 원해서	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.086
FUR2D03	직업선택 요인: 사회적 명예(위신)가 높다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.067
FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.073
FUR2D10	직업선택 요인: 직장 동료와의 경쟁이 심하지 않다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.016
FUR2D12	직업선택 요인: 다양한 경험을 할 수 있다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	0.091
TIM4A	학업관련시간_학원/과외숙제시간: 등교일	#시간으로 응답	0.001
FAM2E02	양육방식 I_과잉간섭: 나에게 무엇을 하든지 항상 이겨야 한다는 걸 강조하신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.002
FAM2F03	양육방식 I_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.073
FAM3A04	양육방식 II_방임: 내가 많이 아프면 적절한 치료를 받게 하신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.102
EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.627
EDU2A04	학교적응_학습활동: 모르는 것이 있을 때 다른 사람(부모님이나 선생님 또는 친구들)에게 물어 본다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.140
EDU2A05	학교적응_학습활동: 공부 시간에 딴 짓을 한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.191

EDU2B05	학교적응_학교규칙: 휴지나 쓰레기를 버릴 때 꼭 휴지통에 버린다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.041
EDU2C02	학교적응_교우관계: 친구와 다투었을 때 먼저 사과한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.018
EDU2C05	학교적응_교우관계: 놀이나 모듬활동을 할 때 친구들이 내 말을 잘 따라준다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.016
EDU2D03	학교적응_교사관계: 학교 밖에서 선생님을 만나면 반갑다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.006
COM2A02	공동체의식: 공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.129
MDA2B08	휴대전화_이용빈도: 음악듣기	Likert: 1(자주 사용한다) ~ 4(전혀 사용하지 않는다)	0.006
MDA2C05	휴대전화_의존도: 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.132
MDA4A	성인용 매체_이용빈도: 이용주기	Likert: 1(전혀 없다), 2(일년 평균) ~ 4(일주일 평균)	-0.115
ACT1A06	체험활동_연간 참여경험 유무: 봉사활동	1(있다), 0(없다)	-0.073
ACT2A01	동아리활동_연간 참여경험 유무: 교내 공식 동아리활동	1(있다), 0(없다)	0.160
ACT3B	여행/문화활동: 문화활동 연간 경험횟수	#회로 응답	0.004
ACT6A	팬덤활동: 선호 연예인/운동선수 유무	1(있다), 0(없다)	-0.036

* 변수는 KCYPS 자료 입력 순으로 정리되었다.

변수의 중요도를 분석하기 위하여 R의 {caret} 패키지에서 varImp() 함수를 이용하였다. 결과는 <표 IV-4>와 같다.

<표 IV-4> 중요도 순 변수 정리(266개 변수사용)

순	변수	변수설명 (결과)	순	변수	변수설명 (결과)
1	EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다(그럴수록)	22	FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다 (중요할수록)
2	INT2C01	학습습관_행동통제: 나는 공부가지루하고 재미없더라도 끝까지 다한다 (그럴수록)	23	ACT1A06	체험활동_연간 참여경험 유무: 봉사활동 (경험이 없을수록)
3	FUR1A01	진로계획_장래 희망 교육수준 (높을수록)	24	INT2A04	학습습관_성취가치: 학교생활이 나의 미래에 상당한 역할을 할 것이다 (그럴수록)
4	EDU2A05	학교적응_학습활동: 공부 시간에 딴 짓을 한다 (그렇지 않을수록)	25	PSY2B07	자아인식_자아탄력성: 나는 같은 장소에 갈 때도 늘 가던 길 보다는 다른 길로 가보는 것을 좋아한다 (그렇지 않을수록)
5	PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다 (그렇지 않을수록)	26	FUR2D03	직업선택 요인: 사회적 명예(위신)가 높다 (중요할수록)
6	INT2C02	학습습관_행동통제: 나는 하던 공부를 끝낼 때까지 공부에 집중한다 (그럴수록)	27	PAR2	보호자특성: [보호자]삶의 만족도 (만족할수록)
7	INT2C05	학습습관_행동통제: 나는 공부를 하려면 쓸데없는 생각 때문에 집중을 못한다 (그렇지 않을수록)	28	EDU2B05	학교적응_학교규칙: 휴지나 쓰레기를 버릴 때 꼭 휴지통에 버린다 (그렇지 않을수록)
8	ACT2A01	동아리활동_연간 참여경험 유무: 교내 공식 동아리활동 (있을수록)	29	ACT6A	팬덤활동: 선호 연예인/운동선수 유무 (없을수록)
9	EDU2A04	학교적응_학습활동: 모르는 것이 있을 때 다른 사람(부모님이나 선생님 또는 친구들)에게 물어본다 (그럴수록)	30	EDU2C02	학교적응_교우관계: 친구와 다투었을 때 먼저 사과한다 (그렇지 않을수록)
10	MDA2C05	휴대전화_의존도: 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없다 (그럴수록)	31	EDU2C05	학교적응_교우관계: 놀이나 모둠활동을 할 때 친구들이 내 말을 잘 따라준다 (그럴수록)
11	COM2A02	공동체의식: 공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있다 (그렇지 않을수록)	32	FUR2D10	직업선택 요인: 직장 동료와의 경쟁이 심하지 않다 (중요하지 않을수록)

12	PSY2A04	자아인식_자아존중감: 나는 남들만큼의 일은 할 수 있다 (그렇지 않을수록)	33	PSY1E06	정서문제_우울: 어떤 일이 잘못 되었을 때 나 때문이라는 생각을 자주 한다(그렇지 않을수록)
13	MDA4A	성인용 매체_이용빈도: 이용주기 (이용하지 않을수록)	34	PSY1A05	정서문제_주의집중: 주의를 기울이지 않아서 실수를 하거나 사고를 낸다 (그렇지 않을수록)
14	INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다 (그렇지 않을수록)	35	MDA2B08	휴대전화_이용빈도: 음악듣기 (사용하지 않을수록)
15	HAK2B	최종학력: 모친 (높을수록)	36	EDU2D03	학교적응_교사관계: 학교 밖에서 선생님을 만나면 반갑다 (그렇지 않을수록)
16	FAM3A04	양육방식Ⅱ_방임: 내가 많이 아프면 적절한 치료를 받게 하신다 (그렇지 않을수록)	37	PSY1B04	정서문제_공격성: 별 것 아닌 일로 싸우곤 한다 (그렇지 않을수록)
17	FUR2D12	직업선택 요인: 다양한 경험을 할 수 있다 (중요하지 않을수록)	38	ACT3B	여행/문화활동: 문화활동 연간 경험횟수 (많을수록)
18	INT2D04	학습습관_학습시간 관리: 나는 시험 전에 계획을 세우고 그에 따라 공부한다 (그렇지 않을수록)	39	FAM2E02	양육방식Ⅰ_과잉간섭: 나에게 무엇을 하든지 항상 이겨야 한다는 걸 강조하신다 (그렇지 않을수록)
19	PSY1E05	정서문제_우울: 울기를 잘한다 (그렇지 않을수록)	40	PHY1B	체격: 몸무게 (가벼울수록)
20	FUR2A03	대학진학이유: 부모님/가족이 원해서 (그렇지 않을수록)	41	TIM4A	학업관련시간_학원/과외숙제 시간: 등교일 (많을수록)
21	FAM2F03	양육방식Ⅰ_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주신다 (그렇지 않을수록)	공 란		

나. 국어 성적 주관적 평가에 영향을 미치는 변수 분석

모형 적합 결과, <표Ⅳ-5>와 같이 270개 변수 중 배경변인 2개, 개인 발달 변인 11개, 발달환경 변인 4개로 모두 17개 변수가 선택되었다.

전체 성적 주관적 평가에 영향을 미치는 변수 중 자아인식으로 자아 정체감에서 낮은 사람을 만나는 것을 꺼려할수록, 양육방식으로 방임에서 내가 많이 아프면 적절한 치료를 받게 하지 않을수록 학업성취도가 높다는 결과와 반대로 국어교과에는 낮은 사람을 꺼리지 않을수록, 아프면 적절한 치료를 받게 할수록 학업성취도가 높은 것으로 나타났다. 두 변수의 중요도가 낮은 편에 속하는데 이것은 국어 교과를 잘하는 학생의 특성과 관련되어 나타난 결과로 해석할 수 있다. 새롭게 포함된 변수는 배경변인으로 성별과 모친의 직업 근로 여부, 개인발달 변인으로 체육시간 중 운동시간, 진로관련 대화상대/빈도, 비행연간 행동경험 유무, 발달환경 변인으로 가족구성 형제자매수, 사이버 비행 연간경험횟수, 성인용 매체 이용 빈도가 포함되었다.

배경변인으로 성별은 여학생일수록, 모친이 일을 하고 있지 않을수록 국어 학업성취도가 높은 것으로 나타났다. 국어와 영어에서 여학생이 우세를 보이는 결과를 제시한 기존 연구결과와 일치한다(김혜숙, 2012; 박경미, 최승현, 2002).

개인발달 변인으로 체육시간 중 운동시간이 적을수록, 친척·친지와 학교선생님과 진로관련 대화를 자주 할수록, 술을 안 마실수록 국어 학업성취도가 높은 것으로 나타났다. 고등학생의 경우 교과공부와 직접적인 연관성이 적은 활동을 최소화하는 방향으로 시간 운용을 하는 것만이 학업성취 향상에 도움이 된다는 연구결과(김경근 외, 2014), 비행 행동을

하지 않을수록 학업성취도가 높다는 연구결과(이종원, 2013)와 일치한다.

발달환경 변인으로 누나/언니의 수가 적을수록, 불법 소프트웨어 다운 받아 사용한 경험이 없을수록, 성인용 매체 이용을 하지 않을수록 국어 학업성취도가 높은 것으로 나타났다. 가족구성에 대한 선행연구에서 주로 부모 구성에 대한 연구가 주를 이루는데 ‘형제 자매수’는 새롭게 나타난 변수이다. 성인용 매체 이용을 하지 않을수록 학업성취도가 높다는 결과는 유해사이트 몰입도가 높을수록 학업성취도가 낮은 것으로 확인된 선행연구와 일치한다(진성희, 성은모, 2013).

<표Ⅳ-5> 국어 성적 모형에 포함된 설명변수 및 척도

변수	변수설명	코딩	계수
GENDER	성별	1(남), 0(여)	-0.2551
JOB1B	직업_근로여부: 모친	1(일을 하고 있다) 0(일을 하고 있지 않다)	-0.0332
PHY3C	건강_체육시간 중 운동시간	Likert: 1(없다) ~ 5(4시간)	-0.0302
INT1E	전체 성적 주관적 평가	Likert: 1(매우 우수) ~ 7(매우 미흡)	-0.7367
INT1D	전체 성적 만족도	Likert: 1(매우 만족한다) ~ 4(전혀 만족하지 않는다)	-0.2115
INT2A01	학습습관_성취가치: 학교 공부는 나에게 중요한 의미를 지닌다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.1619
INT2A06	학습습관_성취가치: 학교에서 배우는 내용은 내가 살아가는 데 유용할 것이다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.0862
INT2D02	학습습관_학업시간 관리: 나는 효율적으로 공부하기 위해 시간 계획을 세운다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.0971

PSY2C07	자아인식_자아정체감: 나는 낯선 사람을 만나는 것을 꺼 려한다	Likert: 1(매우 그렇 다) ~ 4(전혀 그렇지 않다)	0.0344
FUR2A07	대학진학이유: 사회에 진출하 기에는 이르기 때문에	Likert: 1(매우 그렇 다) ~ 4(전혀 그렇지 않다)	0.0463
FUR2C04	진로관련대화상대/빈도: 친적 ·친지	Likert: 1(자주 대화한 다) ~ 4(전혀 대화하 지 않는다)	-0.0570
FUR2C05	진로관련대화상대/빈도: 학교 선생님	Likert: 1(자주 대화한 다) ~ 4(전혀 대화하 지 않는다)	-0.0339
DLQ1A02	비행_연간 행동경험 유무: 술 마시기	1(있다), 0(없다)	-0.1522
FAM1E02	가족구성_형제자매수: 누나/ 언니	#명으로 응답	-0.0661
FAM3A04	양육방식II_방임: 내가 많이 아프면 적절한 치료를 받게 하신다	Likert: 1(매우 그렇 다) ~ 4(전혀 그렇지 않다)	-0.0184
MDA3A02	사이버비행_연간경험횟수: 불 법 소프트웨어 다운받아 사 용하기	1(있다), 0(없다)	-0.0856
MDA4A	성인용 매체_이용빈도: 이용 주기	Likert: 1(전혀 없다), 2(일 년 평균) ~ 4(일 주일 평균)	-0.0737

* 변수는 KCYPS 자료 입력 순으로 정리되었다.

다. 수학 성적 주관적 평가에 영향을 미치는 변수 분석

모형 적합 결과는 <표Ⅳ-6>과 같이 270개 변수 중 배경변인 3개, 개인 발달 변인 15개, 발달환경 변인 2개로 모두 20개 변수가 선택되었다.

새롭게 포함된 변수는 배경변인으로 소득에서 학생이 체감하는 가정의 경제수준 평가, 학업관련시간에서 학교숙제시간(등교일)과 기타공부시간(비등교일)이다. 가정의 경제수준이 잘 사는 편이라고 응답할수록, 학업관련시간이 많을수록 수학 학업성취도가 높은 것으로 나타났다. 모학력이 높을수록, 가구소득이 높을수록 수학 성적이 높은 경향을 나타낸다는 연구결과들과 일치한다(김영희, 김재웅, 류한구, 2007; 이현숙, 신진아, 김경희, 2013; 한수경, 박재범, 손형국, 양정호, 2015).

정서문제에서 사회적 위축에 관한 두 변수의 결과는 [부록 4]의 모형의 결과와 반대인 것으로 나타났다. 즉, 부끄럼을 많이 타지 않을수록, 사람들 앞에 나서기를 싫어하지 않을수록 수학 학업성취도가 높은 것으로 나타났다. 이것은 고등학교 수학 교과목의 특성이 반영된 결과로 모르는 것을 물어보고 자신의 생각을 부끄러워하지 않고 말할 수 있을 때 수학적 학업능력이 올라가는 경향이 있는 것으로 분석된다. 학업을 위해 투입하는 학습의 절대적인 양과 스스로 학습하는 정도에 따라 그 성취정도가 결정된다는 연구결과와 일치한다(김경근 외, 2014).

<표Ⅳ-6> 수학 성적 모형에 포함된 설명변수 및 척도

변수	변수설명	코딩	계수
HAK2B	최종학력: 모친	Likert: 1(중졸 이하) ~ 5(대학원 졸)	0.0204
INCOME1	소득: 가정의 경제수준 평가	Likert: 1(매우 잘 사는 편) ~ 7(매우 못 사는 편)	-0.0741

PAR2	보호자특성: [보호자]삶의 만족도	Likert: 1(매우 만족한다) ~ 4(전혀 만족하지 못한다)	-0.0307
INT1E	전체 성적 주관적 평가	Likert: 1(매우 우수) ~ 7(매우 미흡)	-0.6357
INT2A05	학습습관_성취가치: 학교공부는 내가 미래의 직업을 선택하는 데 커다란 역할을 할 것이다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.1203
INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.0662
INT2C01	학습습관_행동통제: 나는 공부가 지루하고 재미없더라도 끝까지 다한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.3092
PSY1B06	정서문제_공격성: 아무 이유 없이 울 때가 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.0780
PSY1D02	정서문제_사회적 위축: 부끄럼을 많이 탄다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.0124
PSY1D05	정서문제_사회적 위축: 사람들 앞에 나서기를 싫어한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.0250
PSY1E09	정서문제_우울: 장래가 희망적이지 않은 것 같다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.0237
PSY2D06	자아인식_진로정체감: 현재 나는 어떤 직업 분야를 좋아하는데, 그 이유가 분명하다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.0070
FUR2C03	진로관련 대화상대/빈도: 형제·자매	Likert: 1(자주 대화한다) ~ 4(전혀 대화하지 않는다)	-0.0164
FUR2D05	직업선택 요인: 자신의 소질과 적성을 살릴 수 있다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	0.1100
DLQ1A01	비행_연간 행동경험 유무: 담배피우기	1(있다), 0(없다)	-0.0320
TIM3A	학업관련시간_학교숙제시간: 등교일	#시간으로 응답	0.0706
TIM4A	학업관련시간_학원/과외숙제시간: 등교일	#시간으로 응답	0.0342
TIM5C	학업관련시간_기타공부시간: 비등교일	#시간으로 응답	0.0052
EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.2113
EDU2A04	학교적응_학습활동: 모르는 것이 있을 때 다른 사람(부모님이나 선생님 또는 친구들)에게 물어본다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.0529

* 변수는 KCYPS 자료 입력 순으로 정리되었다.

라. 영어 성적 주관적 평가 결과에 영향을 미치는 변수 분석

모형 적합 결과는 다음 <표Ⅳ-7>과 같이 270개 변수 중 배경변인 1개, 개인발달 변인 7개, 발달환경 변인 3개로 모두 11개 변수가 선택되었다.

배경변인으로 가구 연간 소득이 많을수록, 자아인식에서 진로정체감으로 대학에 가서 전공하고 싶은 구체적인 분야가 있을수록, 진로계획에서 장래 직업이 정해놓은 것이 있을수록 영어 학업성취가 높은 것으로 나타났다. 수학에 영향을 미치는 변수에서 자아정체감으로 현재 자신이 좋아하는 직업분야에 대한 이유가 분명하지 않을수록 학업성취가 높게 나온 것과 대조가 된다. 가족구성에서 여동생의 수가 작을수록 성취도가 높은 것으로 나타났는데 국어 모형과는 유사한 결과이나 [부록 4] 모형과는 대조적인 결과이다. 추후 가족구성 중 ‘형제자매수’와 학업성취도와의 관계에 대한 연구를 통해 흥미로운 결과를 얻을 수 있을 것으로 기대된다.

<표 IV-7> 영어 성적 모형에 포함된 설명변수 및 척도

변수	변수설명	코딩	계수
INCOME	가구 연간 소득	#만원으로 응답	0.00005
INT1C02	과목별 주관적 평가: 수학	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.01050
INT1E	전체 성적 주관적 평가	Likert: 1(매우 우수) ~ 7(매우 미흡)	-0.65983
INT1D	전체 성적 만족도	Likert: 1(매우 만족한다) ~ 4(전혀 만족하지 않는다)	-0.07121
INT2D04	학습습관_학습시간 관리: 나는 시험 전에 계획을 세우고 그에 따라 공부한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.01894
PSY2B05	자아인식_자아탄력성: 나는 새로운 음식을 먹어 보는 것을 즐긴다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.11417
PSY2D04	자아인식_진로정체감: 대학에 가서 전공하고 싶은 구체적인 분야가 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.01437
FUR1B	진로계획: 장래 직업 결정 여부	Likert: 1(구체적으로 정해놓은 직업이 있다) ~ 3(아직 하고 싶은 직업을 정하지 않았다)	-0.01093
FAM1E03	가족구성_형제자매수: 여동생	#명으로 응답	-0.07427
FAM2D03	양육방식 I_과잉기대: 나에게 관한 한 다른 어떤 일보다 공부에 더 열성적이었다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.05584
MDA2B08	휴대전화_이용빈도: 음악듣기	Likert: 1(자주 사용한다) ~ 4(전혀 사용하지 않는다)	0.01655

* 변수는 KCYPS 자료 입력 순으로 정리되었다.

마. 각 모형별 중요도 순 10개 변수 비교 분석

<표Ⅳ-8>는 중요도 순으로 각 모형에 포함된 변수를 10위까지 정리한 것이다. 상위 10위 요인을 대영역으로 구분하여 살펴보면, 개인발달 영역의 변수는 34번, 발달환경 영역은 15번, 배경변인은 1번 포함되었다. 개인발달 영역 중 학습습관과 관련된 요인이 가장 많았는데, 행동통제는 5번, 성취가치는 3번, 숙달목적 지향성 2번, 학업시간 관리가 2번으로 총 12번 포함되었다. 다음으로 성적과 관련된 요인은 총 10번 포함되었는데, 전체성적 만족도와 전체성적 주관적 평가가 각각 3번씩 포함되었고, 수학 성적 주관적 평가는 2번, 국어와 영어 성적 주관적 평가는 각각 1번씩 포함되었다. 발달환경 영역에서는 학교적응과 관련된 요인이 가장 많았는데, 교우관계나 교사관계는 선택되지 않았고, 학습활동 관련 요인만 7번 선택되었다. 개인발달 영역에서 진로계획과 관련해서는 장래희망 교육 수준이 2번, 장래직업 결정 여부가 1번 나타났고, 진로관에 해당하는 직업선택요인, 진로관련 상대/빈도, 대학진학 이유는 각각 1번씩, 총 6번 선택되었다. 자아인식과 관련된 영역으로는 자아탄력성, 자아정체감과 진로정체감이 1번씩, 총 3번 포함되었다. 정서문제 중 공격성, 학업관련시간 중 학교숙제시간, 비행 영역 중 술 마신 경험 유무가 각각 1번씩 포함되었다. 발달환경 영역 중 가족구성에서 여동생수, 누나/언니수, 양육방식 중 과잉기대, 동아리 활동 중 연간참여경험 유무, 매체 환경 영역에는 사이버비행 중 불법소프트웨어 다운받아 사용한 연간경험횟수, 성인용 매체 이용 빈도, 휴대전화 영역 중 의존도, 음악듣기 이용 빈도가 각각 1번씩 나타났다. 배경변인에서는 소득 영역 중 가정경제수준이 포함되었다.

〈표Ⅳ-8〉 각 모형별 중요도 순 10개 변수 정리

반응 변수 순서	전체 성적 주관적 평가		국어 성적 주관적 평가	수학 성적 주관적 평가	영어 성적 주관적 평가
	270개 사용	266개 사용			
1	INT1C01	EDU2A03	INT1E	INT1E	INT1E
	국어성적 평가	학교적응_ 학습활동	전체성적 평가	전체성적 평가	전체성적 평가
2	INT1C03	INT2C01	INT1D	INT2C01	PSY2B05
	영어성적 평가	학습습관_ 행동통제	전체성적 만족도	학습습관_ 행동통제	자아인식_ 자아탄력성
3	INT1C02	FUR1A01	INT2A01	EDU2A03	FAM1E03
	수학성적 평가	장래희망 교육수준	학습습관_ 성취가치	학 교 적 응_ 학습활동	가족구성_ 여동생수
4	INT1D	EDU2A05	DLQ1A02	INT2A05	INT1D
	전체성적 만족도	학교적응_ 학습활동	비행_ 술마시기	학습습관_ 성취가치	전체성적 만족도
5	EDU2A03	PSY2C06	INT2D02	FUR2D05	FAM2D03
	학 교 적 응_ 학습활동	자아인식_ 자아정체감	학습습관_ 학 업시간 관리	직업선택요 인_소질적성 살리기	양육방식_과 잉기대
6	FUR1A01	INT2C02	INT2A06	PSY1B06	INT2D04
	장래희망 교육수준	학습습관_ 행동통제	학습습관_ 성취가치	정서문제_ 공격성	학습습관_ 학 업시간 관리
7	EDU2A05	INT2C05	MDA3A02	INCOME1	MDA2B08
	학교적응_ 학습활동	학습습관_행 동통제	사이버비행_ 불법소프트 웨어 다운	소득_가정의 경제 수준평가	휴대전화_ 음악듣기 이용빈도
8	INT2C05	ACT2A01	MDA4A	TIM3A	PSY2D04
	학습습관_ 행동통제	동아리활동_ 연간참여경 험유무	성인용 매체 이용 빈도	학업관련시 간_학교숙제 시간	자아인식_ 진로정체감
9	FUR2A01	EDU2A04	FAM1E02	INT2B01	FUR1B
	대학진학 이유	학 교 적 응_ 학습활동	가족구성_ 누나/언니수	학습습관_ 숙달목적 지향성	진로계획_ 장래직업 결정 여부
10	INT2B01	MDA2C05	FUR2C04	EDU2A04	INT1C02
	학습습관_ 숙달목적 지향성	휴대전화 의존도	진로관련대 화상대/빈도_ 친척·친지	학교적응_ 학습활동	수학성적 평가

개인발달 영역 중 학습습관 관련 변수는 다음 <표Ⅳ-9>와 같다. 별점 회귀 분석에서는 계수를 축소추정하기 때문에 부호가 중요하다고 판단하여 계수를 부호로만 나타내었다.

<표Ⅳ-9> 중요도 순 10위 이내 학습습관 관련 변수

영역	변수	변수설명	코딩	계수
행동 통제	INT2C05	나는 공부를 하려면 쓸데없는 생각 때문에 집중을 못한다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	+
	INT2C01	나는 공부가 지루하고 재미없더라도 끝까지 다한다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	INT2C02	나는 하던 공부를 끝낼 때까지 공부에 집중한다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
성취 가치	INT2A01	학교 공부는 나에게 중요한 의미를 지닌다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	INT2A06	학교에서 배우는 내용은 내가 살아가는 데 유용할 것이다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	INT2A05	학교공부는 내가 미래의 직업을 선택하는 데 커다란 역할을 할 것이다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
학업 시간 관리	INT2D02	나는 효율적으로 공부하기 위해 시간 계획을 세운다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	INT2D04	나는 시험 전에 계획을 세우고 그에 따라 공부한다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
숙달 목적 지향	INT2B01	나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-

중요도 순 10위 이내에 포함된 학교적응에서 학습활동 변수로 선택된 것은 <표Ⅳ-10>과 같다. 중복된 횟수까지 헤아려서 총 7번 나타났다. 고등학교 3학년 학생들은 공부 시간에 딴 짓을 하지 않을수록, 수업시간에 배운 내용을 잘 알고 있을수록 학업성적이 높다는 것을 알 수 있다. 학교에서 내신을 위해 계획된 시간만을 채우고 학업성취 향상을 위한 학습은 학교 밖에서 하고 있다는 우려를 나타낸 연구결과(곽수란, 2006)와

상반된 결과로 볼 수 있다. 모르는 것이 있을 때 다른 사람(부모님이나 선생님 또는 친구)에게 잘 물어 볼수록 학업성적이 높은 것으로 나타났다.

〈표Ⅳ-10〉 중요도 순 10위 이내 학교적응에서 학습활동 변수

영역	변수	변수설명	코딩	계수
학습 활동	EDU2A03	수업 시간에 배운 내용을 잘 알고 있다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	EDU2A04	모르는 것이 있을 때 다른 사람(부모님이나 선생님 또는 친구들)에게 물어 본다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	EDU2A05	공부 시간에 딴 짓을 한다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	+

수업태도나 학생의 교육기대수준, 주당 혼자 공부하는 시간 등 학생개인의 노력 또는 학생이 자신의 수업태도에 대한 평가 또는 인지가 중요함을 확인시켜 준 것으로 사교육이나 가정의 지원만으로는 학업성취도에 큰 영향을 주지 못하며, 학업지원에서 자기주도적 학습 형태의 지원이 필요하다는 것을 의미한다(한수경 외, 2015). 자기주도 학습시간처럼 학생노력변인들이 다른 변인들 보다 더 높은 영향력을 가지고 있는 것으로 나타나, 수학의 학업성취에 사교육이나 가정배경이 영향을 미치고 있기도 하지만, 학생 스스로가 노력하는 것이 학업성취에 더 큰 영향력이 있다는 연구결과를 지지한다(한수경 외, 2015).

중요도 순 10위 이내에 포함된 진로계획 관련 변수는 〈표Ⅳ-11〉과 같다. 장래 희망 교육수준을 높게 생각할수록, 장래 직업을 정할수록, 대학 진학이유를 고졸 학력으로는 취업이 어려워서라고 생각하지 않을수록, 친척·친지와 진로관련대화를 자주 할수록 학업성취가 높게 나타났다. 수

학 성적에 영향을 미치는 상위 9위 변수로 포함된 변수인 직업선택 요인으로 자신의 소질과 적성을 살릴 수 있는 것이 중요하지 않을수록 학업성취가 높게 나타났다.

<표Ⅳ-11> 중요도 순 10위 이내 진로계획 관련 변수

영역	변수	변수설명	코딩	계수
진로 계획	FUR1A01	장래 희망 교육수준	1(고등학교 졸업) ~5(대학원 박사학위취득)	+
	FUR1B	장래 직업 결정 여부	1(구체적으로 정해놓은 직업이 있다)~3(아직 하고 싶은 직업을 정하지 않았다)	-
진로 관	FUR2A01	대학진학이유: 고졸 학력으로는 취업이 어려워서	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	FUR2C04	진로관련대화상대/빈도: 친적·친지	1(자주 대화한다) ~4(전혀 대화하지않는다)	-
	FUR2D05	직업선택 요인: 자신의 소질과 적성을 살릴 수 있다	1(매우 중요하다) ~ 4(전혀 중요하지 않다)	+

중요도 순 10위 이내에 포함된 자아인식 관련 변수는 <표Ⅳ-12>와 같이 자아탄력성, 자아정체감, 진로정체감이었다. 새로운 음식을 먹어 보는 것을 즐기지 않을수록, 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받지 않을수록, 대학에 가서 전공하고 싶은 구체적인 분야가 있을수록 학업성취도가 높게 나타났다. 즉, 학업성취도가 높을수록 자아탄력성이 낮다는 해석이 가능한데, 앞으로 급변하는 시대를 살아갈 고등학생들의 사회 적응력에 우려를 하게 만드는 결과라고 볼 수 있다.

〈표Ⅳ-12〉 중요도 순 10위 이내 자아인식 관련 변수

영역	변수	변수설명	코딩	계수
자아탄력성	PSY2B05	나는 새로운 음식을 먹어 보는 것을 즐긴다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	+
자아정체감	PSY2C06	남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	+
진로정체감	PSY2D04	대학에 가서 전공하고 싶은 구체적인 분야가 있다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-

중요도 순 10위 이내에 포함된 개인발달 영역에 속하는 나머지 변수는 〈표Ⅳ-13〉과 같다. 아무 이유 없이 울지 않을수록, 등교일에 학교 숙제하는 시간이 많을수록, 술을 안 마실수록 학업성취도가 높게 나타났다. 정서적으로 우울하지 않을수록 학업성취도가 높다고 분석할 수 있다. 비행과 관련하여 술 마시기, 담배피우기를 하지 않을수록 학업성취도가 높은 것으로 해석할 수 있다.

〈표Ⅳ-13〉 중요도 순 10위 이내 개인발달 나머지 관련 변수

영역	변수	변수설명	코딩	계수
정서문제	PSY1B06	공격성: 아무 이유 없이 울 때가 있다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	+
학업 관련시간	TIM3A	학교숙제시간: 등교일	#시간으로 응답	+
비행	DLQ1A02	연간행동경험유무: 술마시기	1(있다), 0(없다)	-

중요도 순 10이내에 포함된 매체환경 관련 변수는 〈표Ⅳ-14〉와 같다. 불법 소프트웨어를 다운받아 사용하지 않을수록, 성인용 매체 이용빈도가 적을수록 학업성취도가 높은 것으로 나타났다. 휴대전화 의존도에서 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수 없을수록, 음악듣기

빈도는 낮을수록 학업성취도가 높게 나타났다. 휴대전화 의존도 변수는 보통수준 응답을 빼고 재코딩한 [부록4]의 모형에는 포함되지 않은 변수이다. 즉, 보통수준 학생들의 특징으로 해석할 수 있다.

〈표Ⅳ-14〉 중요도 순 10위 이내 매체환경 관련 변수

영역	변수	변수설명	코딩	계수
사이버 비행	MDA3A02	연간경험횟수: 불법 소프트웨어 다운받아 사용하기	1(있다), 0(없다)	-
성인용 매체	MDA4A	이용빈도: 이용주기	1(전혀 없다), 2(일년 평균)~ 4(일주일 평균)	-
휴대 전화	MDA2C05	의존도: 혼자 있을 때 휴대 전화가 없으면 심심해서 건딜 수가 없다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
	MDA2B08	이용빈도: 음악듣기	1(자주 사용한다) ~ 4(전혀 사용하지 않는다)	+

중요도 순 10위 이내에 포함된 발달환경 영역에 속하는 나머지 관련 변수는 〈표Ⅳ-15〉와 같다. 부모의 양육방식이 나에 관한 한 다른 어떤 일보다 공부에 더 열성적이실수록, 교내 공식 동아리활동에 참여할수록 학업성취도가 높은 것으로 나타났다. 가족구성 영역에서는 형제자매수가 적을수록 학업성취도가 높은 것으로 나타났다.

〈표Ⅳ-15〉 중요도 순 10위 이내 발달환경 나머지 관련 변수

영역	변수	변수설명	코딩	계수
양육 방식	FAM2D03	과잉기대: 나에 관한 한 다른 어떤 일보다 공부에 더 열성적이시다	1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-
동아리 활동	ACT2A01	연간 참여경험 유무: 교내 공식 동아리활동	1(있다), 0(없다)	+
가족 구성	FAM1E02	형제자매수: 누나/언니	#명으로 응답	-
	FAM1E03	형제자매수: 여동생	#명으로 응답	-

연구에서 사용한 데이터는 전처리 과정을 거치면서 형제자매가 0명인 데이터는 모두 삭제가 되어 실제 분석에 사용한 데이터는 모두 형제자매가 있는 자료였다. 기존의 가족구성에 관한 연구에서는 대부분 양부모, 편부모인 경우에 관심을 기울였는데 본 데이터의 마지막 전처리 과정에서 부모구성(FAM1B) 변수가 영근처 분산을 가지는 것으로 밝혀져 변수에서 삭제되었다. 따라서 기존 연구에서 부모구성의 영향이 고등학교 3학년 학생에게도 있는지는 알 수 없었다.

중요도 순 상위 10위 이내에 포함된 배경변인 관련 변수는 <표Ⅳ-16>과 같다. 수학 학업성취도와 관련하여 선택된 변수인데, 배경변인 중 가정의 경제수준을 평가할 때 잘 산다고 응답할수록 학업성취도가 높은 것으로 나타났다.

<표Ⅳ-16> 중요도 순 10위 이내에 포함된 배경변인 관련 변수

영역	변수	변수설명	코딩	계수
소득	INCOME1	가정의 경제수준 평가	1(매우 잘 사는 편) ~ 7(매우 못 사는 편)	—

V. 논의

1. 연구 방법 및 연구 과정에 대한 논의

본 연구에서는 LASSO 별점 회귀 모형을 이용하여 KCYPS에서 제공하는 수백 개의 변수들을 최대한 활용하여, 고등학교 3학년의 주관적 학업 성취도에 영향을 미치는 변수를 탐색하고, 새로운 자료에 대한 모형 예측력을 높이는데 초점을 맞추어 모형을 구축하였다. LASSO는 별점화 축소추정법 중 하나로, 계수축소와 변수선택을 동시에 실행함으로써 예측력을 향상시키고 적은 수의 변수만을 선택하여 추정된 모형에 대한 해석을 용이하게 하는 방법이다. 비별점 회귀모형을 사용하면 투입 변수가 많아서 회귀계수의 표준오차가 팽창하여 모형이 불안정해질 수 있는데, 별점 회귀 모형은 비별점 회귀모형의 안정성이 떨어지는 문제에 대해 해결책을 제시한다.

또한 예측변수들 간에 다중공선성 문제가 있으면 모형에 포함될 올바른 예측변수의 선택을 어렵게 할 수 있으며, 각 예측변수의 정확한 효과를 파악하는데 장애가 될 수 있다(나종화, 2017b). 그러나, LASSO 별점 회귀를 이용함으로써 400개가 넘는 변수를 모두 활용하여도 모형을 적합할 때, 모형에서 특정 변수의 과다한 기여를 제한함으로써(큰 모수 값에 별점을 부여함으로써), 상관된 다른 변수가 동시에 모형에 포함될 수 있도록 허용하여 주관적 학업성취도에 영향을 미치는 변수선택을 할 수 있었고, 모형의 예측력도 우수하였다. 앞으로 기존의 대규모의 패널 데이터뿐만 아니라 기하급수적으로 늘어나고 있는 교육관련 데이터에 본 연구 방법을 적용하여 변인들 사이의 관계를 분석하는데 사용할 수 있

을 것이다.

손실함수로 반응변수의 범주를 이진형(binary)으로 코딩하여 로지스틱 회귀모형을 이용하였다. 이 과정에서 전체 성적 주관적 평가 응답 척도가 7개의 단계였는데, 범주를 이진형으로 코딩하기 위하여 1(매우 우수) ~ 4(보통 수준)까지를 '1'로, 5(약간 미흡) ~ 7(매우 미흡)까지를 '0'으로 코딩하였다. 이 과정에서 원래 데이터의 정보가 손실되는 현상이 있었다. 이것을 보완하기 위하여 [부록 4]에서 제시한 것과 같이 4(보통 수준) 데이터를 제거한 후, 1(매우 우수) ~ 3(약간 우수)를 '1'로, 5(약간 미흡) ~ 7(매우 미흡)까지를 '0'으로 재코딩하여 모형을 적합하였다. [부록 4]의 방법에서는 4(보통 수준)로 응답한 444명의 정보 손실이 있어 개체 정보의 손실이 더 큰 것으로 판단하여 본 연구에서와 같이 코딩하여 사용하였다. 7개의 척도를 효율적으로 사용하기 위해서 이진형 반응변수를 3개의 범주로 나눈 후, 다범주(multi-category) 로지스틱 회귀모형을 사용할 수 있을 것이다. 본 연구에서는 학업성취도에 영향을 미치는 변수 선택에 목적을 두고 즉, 예측력(predictability)에 초점을 맞추어 주관적 학업성취도가 우수하나, 아니냐 하는 것을 판단하기 위해 이진형으로 코딩하였다.

LASSO는 상관이 높은 예측변수 가운데 일부만을 선택하는 경향을 가진다. 본 연구에서도 반응변수가 달라짐에 따라 영향을 미치는 같은 영역의 변수 선택에서 다른 예측변수를 선택하였다. 별점 회귀 중 Elastic Net 회귀는 능형회귀와 LASSO 회귀의 절충적인 방법으로 상관이 높은 예측변수들이 모두 선택되도록 하는 특징을 가지고 있기 때문에 변수들 간의 관계를 살펴볼 때 사용하기에 적절하다. 하지만 더 많은 수의 변수가 선택되어 모형의 설명력이 다소 떨어지는 경향을 가질 수 있다. 본

연구에서도 Elastic net 회귀 방법을 이용해 본 결과 LASSO를 이용한 결과와 거의 유사하였으나, 상관이 높은 변수들을 모두 선택하여 더 많은 수의 변수가 선택되었다. 선행연구에서도 별점 회귀 모형 중 LASSO 회귀의 예측력이 뛰어난 결과를 나타내었다는 연구 결과(송상운, 2015)와 김용대 외(2013)의 빅데이터와 통계학에 관한 연구에서 고차원 모형에서는 기존의 단계적 변수선택 방법보다 별점화 기법 중 변수선택과 예측을 동시에 할 수 있는 통계기법인 LASSO를 사용할 것을 언급하고 있다.

이에 본 연구에서는 별점화 축소추정법 중 예측력이 우수한 LASSO를 통하여 모형을 구축하고 학업성취도에 영향을 미치는 변수를 탐색하였다. 그동안 사회 과학 분야에서 잘 활용하지 않았던 별점화 축소기법을 활용해, 기존 이론에 근거하지 않고 새로운 변수나 변수 간 관계를 탐색해 보았다는데 연구의 의의가 있다. 하지만 양적 연구방법의 특성상 연구자의 주관에 의한 판단이 연구 과정 중 개입되기 때문에 그 해석에 있어서 유의할 필요가 있다. 또한 그 결과를 다른 다양한 방법들을 통해 검증할 필요가 있다.

분석에는 R에서 제공하는 패키지²⁰⁾ {car}, {caret}, {caTools}, {glmnet}, {prettyR}, {ROCR}, {VIM}을 사용하였다. R은 통계분석, 데이터 시각화, 빅데이터 분석 등에 강점을 가지고 있고, 다양한 전공 분야에서 활발히 이용되고 있는 open source 프로그램이다.

모형 적합 전 데이터 전처리(preprocessing) 과정을 거치면서 여러 개의 변수들이 삭제되었고 재코딩되었다. 삭제된 변수들은 분석에 불필요하다고 판단된 변수, 즉 ID, 학교급(모두 고등학교), 가중치, 출생년, 출

20) R에는 패키지(package)라는 추가 기능이 있는데, 이 패키지 안에는 함수, 예제 데이터, 예제 프로그램, 도움말이 있으며, 수시로 새로운 패키지들이 개발되고 배포되어 유용하게 사용할 수 있다(이부일, 이서현, 2016).

생월, 보호자 출생년도, 학년(모두 3학년)이고, 학생과 학부모가 미응답한 결측치였다. 또 전체 응답율이 지나치게 낮아서 결측률이 매우 높은 문항들은 처음부터 삭제하였다. 영과 영근처 분산을 가지는 예측변수는 모형을 망가뜨리거나 불안정한 적합의 원인이 되기 때문에 데이터 전처리 과정에서 삭제하였다. LASSO는 학교지역, 자택지역, 주택형태, 부친의 직업분류와 같이 더미코딩이 필요한 범주형 변수의 경우 사용하기 적절하지 않아 삭제하였다. 예를 들어 아버지의 직업분류를 관리자, 전문가 및 관련종사자, 사무종사자, 서비스종사자, 판매종사자, 농림어업숙련종사자, 기능원 및 관련 기능 종사자, 장치, 기계조작 및 조립종사자, 단순노무종사자, 군인의 열 개 범주로 자료를 수집하였다. 이 경우 직업분류는 총 아홉 개의 더미변수로 코딩될 수 있으며, 이 아홉 개의 더미변수가 하나의 범주형 변수로부터 도출되었다는 변수 집단의 구조적 특성이 축소 추정시 고려되어야 한다. 그런데, LASSO의 경우 아홉 개의 더미 변수가 하나의 범주형 변수에서 생성되었다는 것을 고려하지 않는다는 문제가 있다. 이러한 문제를 해결하기 위해 group LASSO를 사용할 수 있다(유진은, 노민정, 2017). 그러나, 몇 가지 문제점에도 불구하고 LASSO는 대체로 고차원인 유한표본 자료에서의 예측력이 다른 방법에 비해 좋기 때문에 여전히 많은 분야에서 사용된다(박창이 외, 2015). 추후 보완된 다른 LASSO 방법을 사용한 결과와 비교하여 살펴보는 것도 의미가 있을 것이다.

데이터 전처리 과정에서 학업성취도에 영향을 미치는 변수를 살펴보기 위해 진로를 대학 진학 또는 진학 준비로 응답한 80.4%의 학생만 선택하였다. 큰 의미로 보면 학업성취도라는 것이 진학을 희망하든 취업을 희망하든 평생 교육의 관점으로 비추어 볼 때 배움을 지속할 수 있는

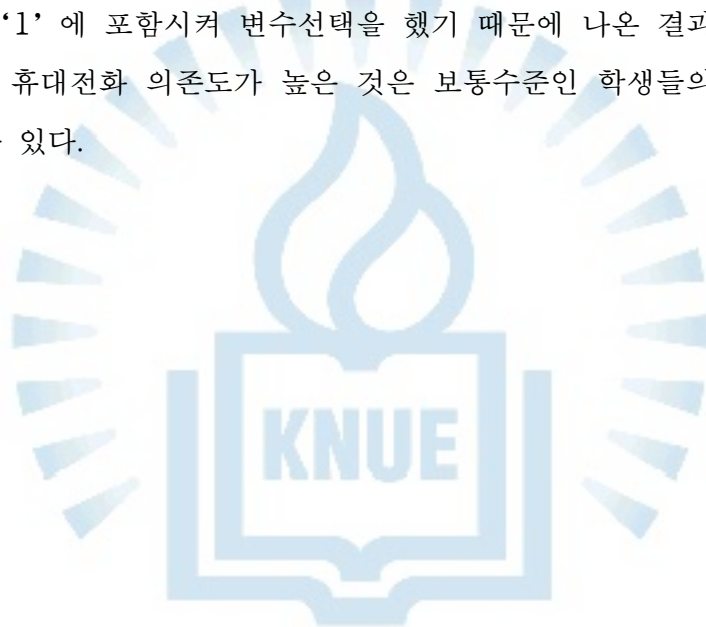
능력이 어느 정도인지를 의미한다고 봐야 할 것이다. 그러나 본 연구에서는 현재 우리나라 고등학교 3학년의 학업성취도라는 것이 대학 진학을 목표로 하는, 즉 입시위주의 교육에서 얼마나 성취를 하고 있는지를 의미하는 것으로 제한하여 학업을 계속 이어나갈 때 지속적으로 영향을 미칠 것으로 예상되는 변수를 탐색하고자 하였다. 연구자 본인의 학교 현장 경험을 통해 진학을 희망하지 않는 학생들은 학업에 거의 손을 놓은 상태라고도 볼 수 있기 때문에 진학을 희망하는 학생들을 대상으로 하는 것이 더 타당하다고 판단한 것이다. 추후 취업을 희망하거나 아직 결정하지 못한 학생들을 대상으로 모형을 적합해 보고 비교하는 것도 의미가 있다고 생각한다.

진로계획 문항에서 장래 희망 교육수준을 ‘아직 결정하지 않았다’로 응답한 자료를 결측치로 재코딩하였는데 이것은 장래 희망 교육수준을 결정하지 못해서 응답하지 않은 것과 같은 것으로 판단되었기 때문이다. 결측비율이 10%를 넘는 문항은 삭제하였고, 마지막으로 결측치를 하나라도 포함하고 있는 개체를 완전 삭제하여 271개 문항에 대해 남학생 527명, 여학생 565명의 총 1092명으로 정리되었다.

전처리 과정을 거친 데이터를 사용하여 크게 4개의 모형을 적합하였다. 전체 성적 주관적 평가, 국어 성적 주관적 평가, 수학 성적 주관적 평가, 영어 성적 주관적 평가를 각각 반응변수로 하여 모형을 구축하였다. 이렇게 반응변수를 전체 성적, 교과로 나누어 모형을 구축한 이유는 주관적 학업성취도에 영향을 미치는 변수에 차이가 있는지를 살펴보고자 한 것이다.

모형 적합 과정에서 기존 연구 결과와 상이하게 선택된 변수가 있어서 전체 성적 주관적 평가에 응답한 값 중 4(보통 수준)을 제거하고 반

응변수를 재코딩한 후 모형을 적합해 보았다. 그 결과 앞선 연구 결과와 다른 결과를 보이기도 하였다. 예를 들면 휴대전화 의존도에서 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없을수록 학업성취도가 높게 나타났다. 휴대전화 의존이 높을수록 오락적 휴대전화 이용의 빈도가 높을수록 학업성취에는 유의한 부적효과가 발생함을 보고한 연구결과와 상반되는 결과이다(김지혜, 2015; 성은모, 진성희, 2012; 이승진, 김경근, 2015). 하지만 본 연구에서 반응변수의 코딩 단계에서 ‘4(보통수준)’을 ‘1’에 포함시켜 변수선택을 했기 때문에 나온 결과로 분석되었다. 즉, 휴대전화 의존도가 높은 것은 보통수준인 학생들의 특징으로 해석할 수 있다.



2. 연구 결과에 대한 논의

고등학교 3학년의 학업성취는 직접적으로 대학입시에 영향을 미친다. 우리 사회에서 어떤 대학, 전공을 선택하느냐가 앞으로 살아갈 삶의 질에 큰 영향을 준다. 따라서 우리나라 고등학교 3학년들은 무한 경쟁 속에서 개인적으로 그 고통을 감내하며 학업에만 열중하기를 강요당하는 것이 우리의 현실이다. 이에 본 연구에서는 KCYPS 중1패널 6차년도 자료를 이용하여 고등학교 3학년의 학업성취도에 영향을 미치는 변수를 탐색하였다. 변수 탐색을 통해 고등학교 3학년의 특징과 우리나라 교육 현실을 들여다 볼 수 있었다.

전체 성적 주관적 평가를 반응변수로 사용한 모형에서는 선행연구에서 학업성취도에 영향을 미치는 것으로 밝혀진 대부분의 변수들이 선택되었다. 개인발달 영역에는 체격(몸무게), 학습습관(성취가치, 숙달목적 지향성, 행동통제, 학업시간 관리), 정서문제(주의집중, 공격성, 우울), 자아인식(자아존중감, 자아탄력성, 자아정체감), 진로계획(장래 희망 교육수준, 대학진학 이유, 직업선택 요인), 학업관련 시간(학원/과외숙제 시간)이 포함되었다. 발달환경 영역에는 양육방식(과잉간섭, 합리적 설명, 방임), 학교적응(학습활동, 학교규칙, 교우관계, 교사관계), 매체환경(컴퓨터 사용여부, 휴대전화 용도별 이용 빈도와 의존도, 성인용 매체 이용 빈도), 체험활동(연간 참여경험 유무), 동아리 활동(연간 참여경험 유무), 여행/문화활동(문화활동 연간 경험횟수), 팬덤활동(선호 연예인/운동선수 유무)이 포함되었다. 배경변인에는 보호자(모) 최종학력, 보호자 특성(삶의 만족도)이 포함되었다.

기존 연구와 다른 결과를 가지는 변수는 자아탄력성이 낮을수록, 양육

방식에서 합리적 설명을 하지 않을수록, 과잉간섭에서 무엇을 하든지 항상 이겨야 한다는 걸 강조하지 않을수록, 교사관계에서 선생님께 반갑게 인사하지 않을수록, 교우관계에서 친구와 다투었을 때 먼저 사과하지 않을수록 학업성취도가 높게 나타난 것이다.

반면 기존 연구에서 사용하지 않았던 새롭게 선택된 변수로는 공동체 의식, 봉사활동 참여경험유무, 가족구성 중 형제 자매수가 있었다. 공동체 의식에서 ‘공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있다’와 체험활동에서 ‘봉사활동 연간 참여경험 유무’는 자원봉사 활동을 하지 않을수록, 봉사활동에 참여경험이 없을수록 학업성취도가 높은 것으로 나타났다. 이것은 인과관계로 해석하기 보다는 고등학생의 경우 교과공부와 직접적인 연관성이 적은 활동을 최소화하는 방향으로 시간 운용을 하는 것만이 학업성취 향상에 도움이 되기 때문에 나타난 특징으로 해석하는 것이 더 타당하다(김경근 외, 2014). 또한 고등학교의 교과 및 비교과 활동은 대학입시 및 대학진학에 초점이 맞춰져 있는 현실을 반영한 결과이기도 하다.

모형에 따라 결과에 차이가 있는 변수는 양육방식에서 방임인데 전체 성적 주관적 평가 모형에서는 내가 많이 아프면 적절한 치료를 받게 하지 않을수록 학업성취가 높다는 결과였으나, 국어 성적 주관적 평가 모형에서는 적절한 치료를 받게 하실수록 학업성취도가 높은 것으로 나타났다. 또 전체 성적 모형에서는 자아인식에서 자아정체감과 관련해서 낯선 사람을 만나는 것을 꺼려할수록 학업성취도가 높다는 결과였으나, 국어 성적 모형에서는 반대로 낯선 사람을 만나는 것을 꺼려하지 않을수록 국어 성적이 높은 것으로 나타났다.

전체 성적 모형과 비교하여 국어 성적 모형에 새롭게 포함된 변수는

성별, 모친의 근로여부, 체육시간 중 운동시간, 진로관련대화상대/빈도, 비행연간 행동경험 유무, 가족구성 형제자매수, 사이버 비행 연간경험횟수, 성인용 매체 이용 빈도 변수였다. 수학 성적 모형에 새롭게 포함된 변수는 소득에서 학생이 체감하는 가정의 경제수준 평가, 자아인식에서 진로정체감, 학업관련시간에서 학교숙제시간(등교일)과 기타공부시간(비등교일)이었다. 영어 성적 모형에는 가구 연간 소득, 장래 직업 결정 여부, 양육방식에서 과잉기대가 새롭게 포함되었다. 영어 성적 모형에서 진로정체감으로 대학에 가서 전공하고 싶은 구체적인 분야가 있을수록, 진로계획에서 장래 직업이 정해놓은 것이 있을수록 영어 학업성취가 높은 것으로 나타난 것과 다르게 수학 모형에서는 자아정체감으로 현재 자신이 좋아하는 직업분야에 대한 이유가 분명하지 않을수록 학업성취가 높은 것으로 나타났다. 고등학교 3학년에서 공부하는 각 교과목의 학문적인 특성이 반영된 것으로 해석할 수 있다. 각 교과목별 모형이 전체 성적 모형에 포함된 변수를 포함하지 않는 것도 있었다.

교사관계와 교우관계 변수에서 나타난 결과는 고등학교에서 학업성취도가 높은 학생들의 특징으로 볼 수 있다. 학교 공부는 나에게 중요한 의미를 지니고 미래의 직업을 선택하는 데 커다란 역할을 할 것으로 생각할수록, 학교에서 배우는 내용은 내가 살아가는데 유용할 것이라고 생각할수록, 수업 시간에 배우는 내용을 잘 알고 있기 위해 수업 시간에 딴 짓을 안 할수록 학업성취도가 높은 것으로 나타났다. 이것은 학교에서 이루어지는 학습활동에는 큰 관심을 가지고 중요하게 생각하고 있다는 것을 나타낸다. 그러나 인지적 영역인 학업 이외의 사회적 관계 즉, 교사, 친구들과의 관계는 중요하게 생각하고 있지 않다는 것을 드러내는 것이라고 판단된다.

학교가 학생들에게 제대로 된 교육의 경험을 제공할 때 학생들은 교육을 통한 좋은 내면적인 것과 지식적인 것의 성장을 경험하게 될 것이다. 이런 경험을 통해 살아가면서 마주하게 되는 새로운 경험을 두려워하지 않고, 그 변화에 유연하게 대처하며 사회 구성원으로 당당하게 살아갈 수 있을 것이다.

또한 학생과 교사가 학교 울타리 안에서 함께 배우고 성장할 수 있는 분위기 속에서 교육적 성장을 이뤄갈 수 있는 풍토 조성이 시급하다고 판단된다. 학교에서 정답만 맞추는 훈련이 아니라, 진정한 의미의 교육이 일어나고 바로 그곳의 경험을 통해서 학생들의 교육적 능력을 키우는 통로를 열어야 할 것이다. 안 해본 것을 해 보게 될 때 느끼는 즐거움, 어려움이 있으나 알게 되었을 때 느낄 수 있는 보람과 같이 교육이 즐겁고 신나는 놀이이며, 기쁨이 된다는 것을 경험하는 것이 꼭 필요하다(장상호, 2009).



VI. 결론 및 제언

1. 결론

‘전체성적 주관적 평가’, ‘국어, 수학, 영어 성적 주관적 평가’를 각각 반응변수로 하여 LASSO 별점회귀 모형을 적합한 결과, <표 VI-1>와 같이 설명변수 270개를 모두 사용하여 ‘전체성적 주관적 평가’를 반응변수로 사용한 모형이 가장 예측력이 뛰어난 모형이었다.

<표 VI-1> 반응변수에 따른 모형의 성능 비교

반응변수 통계량	전체성적 주관적 평가		국어 성적 주관적 평가	수학 성적 주관적 평가	영어 성적 주관적 평가
	설명변수 모두 사용	국,영,수, 성적만족도 제외			
정분류율(%)	85.98	80.49	84.76	79.2	77.44
Kappa 값	0.5787	0.3681	0.2600	0.5045	0.3784
AUC 값	0.9079	0.8091	0.8238	0.8102	0.7768

고등학교 3학년 학생의 주관적 학업성취도에 중요하게 영향을 미치는 영역은 개인발달 영역이었고, 그 중 학습습관에서 행동통제를 잘 할수록, 성취가치를 인식하고 있을수록, 숙달목적 지향성이 높을수록, 학업시간 관리를 잘 할수록 학업성취도가 높은 것으로 나타났다. 진로계획에서 장래희망 교육수준이 높을수록, 중요하게 생각하는 대학진학이유를 가지고 있고, 진로와 관련한 대화를 자주 나눌수록 학업 성취도가 높았다. 정서적으로 안정되어 있어 우울하지 않으며, 공격성이 없을수록, 비행

행동을 하지 않을수록, 학업관련 숙제시간, 공부시간이 많을수록 학업성취도가 높은 것으로 나타났다. 다음으로 중요하게 영향을 미치는 것은 발달환경 영역인데, 그 중 학교적응 변수가 중요하였다. 즉, 학습활동에서 학교 수업시간에 배운 내용을 잘 알고 있고, 모르는 것이 있을 때 다른 사람에게 질문을 잘 할수록, 공부 시간에 딴 짓을 안 할수록 학업성취도가 높은 것으로 나타났다. 마지막으로 배경변인은 다른 학령기에 비해 영향력이 적은 것으로 나타났는데, 중요변수로 10위 안에 든 변수는 수학 성적 모형에서 가정의 경제수준 평가를 잘 산다고 인식 할수록 수학 학업성취도가 높은 것으로 나타났다.

별점 회귀 모형 적합 결과 해석 시 주의할 점은 모형 예측력을 최대화하는 방향으로 모형 선택이 이루어지기 때문에 선택된 변수가 통계적으로 유의하지 않을 수 있다. 따라서 결과를 해석할 때, 얻어진 결과를 과장하지 않도록 주의해야 하며 추후 독립적인 변수 하나하나에 대한 검증이 필요하다는 것이다.

2. 제언

본 연구는 KCYPS에서 제공하는 많은 변수들을 활용하여 데이터 마이닝 기법인 LASSO 로지스틱 회귀분석을 이용하여 고등학교 3학년의 학업성취도에 영향을 미치는 변수들을 탐색하였다. 모형에 포함된 변수들을 통해 우리나라 고등학교 3학년의 특징들도 살펴 볼 수 있었다.

본 연구의 결과를 바탕으로 다음과 같이 제언하고자 한다.

첫째, 학업성취를 높이기 위한 방안으로 학생들의 행동통제 능력을 키워주는 것이 필요하다. 어릴 때부터 가정에서 기다리는 훈련, 자신의 욕망을 이길 수 있는 노력을 할 수 있도록 가르쳐서 그것이 습관이 된 고등학생들은 학업성취도가 높았다(Lehrer, 2009). 어릴 때 행동통제 능력을 키우지 못하였더라도 운동을 통해 근육을 강화할 수 있는 것과 같이 행동통제 능력도 연습을 통해 강화될 수 있다(Matthew, 2013). 가정과 학교에서 행동통제 능력을 키워주는 방안을 마련할 필요가 있다.

둘째, 자신의 진로를 계획하고 그것을 이루기 위해 얼마만큼의 교육이 필요할 것인지에 대한 생각을 할 수 있는 여건을 마련해 줄 필요가 있다. 최근 서울시교육청과 경기도교육청이 ‘마을교육공동체’를 통해 학교-지역-가정을 연계한 개념과 실천의 움직임이 일어나고 있다(김유리, 김현철, 박진옥, 권선향, 김누리, 2016). 시대는 점점 더 빠르게 변화해 가고 변화는 학교와 학생에게 지식을 공급하는 전문화된 길을 걸으라고 주문하면서 마을과 경계가 생겼고, 학교의 울타리는 높아졌다(이승훈, 2017). 학부모와 학생들도 무한 경쟁 사회에서 자신의 불안과 두려움을 해소하기 위해 교육 소비자가 되어 점점 더 개인적인 서비스와 개별적

인 공간을 찾게 된 것이다. 마을교육공동체를 통해 다양한 삶의 모습들을 지켜보고 저마다의 길이 있다는 것을 삶으로 배우고 일터를 경험하며 진로에 대한 구체적인 계획을 갖게 될 것이다. 학교에서 진로교육이 중요하다며 전문가 특강이나 프로그램 등을 운영하지만 그것이 오히려 학생들에게 성공하지 않으면 살기 힘들고 돈을 많이 벌지 않으면 이 사회에서 인정받지 못한다는 생각을 부추기는 결과는 아닌지 돌아볼 문제이다. 배움과 역량중심교육이 강조되고 있는 오늘날 학생들의 역량이란 결국 미래를 설계하고 개척해 나가는데 필요한 능력이 아닐까 생각한다. 이러한 의미에서 마을교육공동체가 학습플랫폼으로 정보와 지식이 모이고, 사람들이 소통하고, 이를 통해 문제를 해결하는 기회이자 공간이 될 수 있다. 이제 더 이상 학교가 지식을 전달하는 장소가 아니라 학생들이 미래를 설계하는데 필요한 정보와 지식을 습득·공유하고 사람들과 소통하며 문제를 해결하기 위한 기회를 제공하는 곳이 되어야 할 것이다(김용련, 2017). 다행인 것은 아직까지 고등학교 3학년 학생들은 학교생활을 중요하게 생각하고 있고 학교에서 하는 학습활동이 자신의 미래에 중요하다고 생각하고 있기에 학교가 새로운 패러다임을 가지고 변화할 필요가 있다고 본다.

셋째, 학교 수업을 교과와 본래적 의미를 올바르게 이해하고 교과와 가치를 내면화할 수 있도록 올바른 방향으로 교과를 가르치자는 것이다. 그렇게 될 때 교과 공부가 실제 삶과 유리되지 않고 교과에 포함된 가치들을 통해 자연스럽게 학생들의 생각이 깊어지고 바른 인성이 키워질 것이다(유재봉, 2016). 그러기 위해서는 자연스럽게 평가의 변화도 함께 이루어져야 한다. 수능은 원래 자격고사 개념으로 설계되었는데 현재 수능은 학생들을 변별해서 대학에서 학생을 선발하는 도구로 변질되었다

(박도순, 2017). 변별력을 강조하다 보니 수학 같은 경우 아무리 수학적 재능이 있는 학생이라도 100분에 30문제를 푼다는 것은 유형별 반복 학습이 되어 있지 않으면 주어진 시간에 다 풀 수가 없다. 수학의 본질은 생각의 힘을 키워주는 것인데 반복학습을 통해 생각할 겨를도 없이 기계적으로 문제를 풀어야 하는 시험이다 보니 흥미는 급격히 떨어지고 호기심은 어딘가로 사라져 더 이상 생각하지 않게 되는 것이다(박형주, 2017). 본 연구결과에서도 학업성취도가 높은 학생들이 자아탄력성이 떨어지고 새로운 낯선 환경을 선호하지 않는 것으로 나타났다. 공식의 기계적 적용이 아니라 문제의 본질을 보고 해결방안을 찾을 수 있도록 평가의 변화가 있어야 할 것이다.

본 연구에서 학업성취도가 높은 학생들의 도덕성, 사회성에 우려가 되는 결과가 나타났다. 보통 수준 이상의 학업성취도를 가진 학생들의 특징이 교사를 만나서 반갑게 인사하지 않고, 친구와 싸웠을 때 먼저 사과하지 않고, 휴지나 쓰레기를 버릴 때 휴지통에 버리지 않는 것, 학교 밖에서 교사를 만나면 반갑지 않다는 등의 변수들이 여러 개 선택되었다. 학교급별 학생들의 인성 수준을 진단한 연구결과에 따르면 초등학교의 인성수준이 가장 높고, 중학교의 인성수준이 가장 낮으며, 고등학교의 인성수준은 중학교에 비해 높지만 초등학교에 비해 인지적 역량이나 사회적 역량이 증가함에도 불구하고 인성수준이 초등학교보다 낮았다고 보고했다(윤영돈, 2015). 아마도 고등학교의 교과 및 비교과 활동이 대학 입시 및 대학진학에 초점이 맞춰져 있기 때문인 것으로 생각된다. 인성 중 봉사(배려) 및 협력 부분에서 봉사의 중요성은 알고 있지만 그것이 학창시절에 실천되지 않고 경험되지 않는다면 과연 대학입시의 압박에서 벗어나 사회인이 되었을 때 봉사를 실천하며 협력하는 인간으로 살

수 있을까 하는 생각이 든다. 지금 하고 있는 공부에만 몰입하여 나와 다른 생각을 가진 친구의 마음을 이해하려고 노력하지 않고 새로운 환경을 두려워하며 안정된 여건 가운데서 다른 모든 활동을 자제하고 공부에만 매달려야 좋은 성적, 좋은 대학을 갈 수 있는 고등학교 3학년의 현실을 적나라하게 보여주는 것이라고 판단된다. 모든 교육이 학교에서만 이뤄져야 된다는 생각에는 동의하지 않지만 오랜 시간을 학교에서 보내고 있는 고등학생들에게는 학교에서 제대로 된 인성을 가르쳐야 하는데 그것이 다른 특별한 프로그램을 운영해서 일회성으로 끝나게 할 것이 아니라 교과와 내재적 가치를 가르침으로써 자연스럽게 도덕성, 인성이 길러지도록 할 수 있을 것이다. 또한 앞에서도 언급했듯이 가정과 학교와 사회의 공동 노력을 통하여 자연스럽게 인성을 습득할 수 있는 건강한 사회문화적 풍토 조성 및 입시 위주의 교육 시스템 개선이 따라와야 할 것이다. 덧붙여 건전한 역할 모델로서의 교사의 도덕적이고 민주적인 학습 분위기 조성도 필요하다(진의남, 2012).

넷째, 교육 데이터에 관심을 가지고 통계 기법을 익혀 교육현상을 객관적으로 조망하고 정책에 활용할 수 있을 것이다. 오늘날 데이터가 시시각각 변하고 시대에 따라 다양하게 전달되기 때문에 통계적인 생각을 가지고 분석할 필요가 있다. 교육 분야에도 대규모 패널 자료, 오프라인 및 온라인 교수 학습에 관한 수많은 정보, 학생들이 사용하는 소셜미디어 데이터 등이 이미 축적되어 있고, 앞으로는 더욱 많은 교육관련 데이터가 폭발적으로 증가할 것이다(이은경 외, 2014). 따라서 데이터 속에서 숨겨진 패턴을 찾아내고 여러 변수들을 통합적으로 고려하면서 창의적으로 해석할 수 있는 분석능력이 앞으로 더욱 중요해질 것이다(강만모 외, 2012). 본 연구에서도 KCYPS에서 제공하는 모든 변수들을 활용하여

학업성취도에 영향을 미치는 변수를 탐색함으로써 새로운 변수의 선택이 가능했고, 그것을 통해 고등학교 3학년 학생들의 특징도 살펴볼 수 있었다.

본 연구를 통해 고등학교 3학년의 현실을 조금이나마 살펴볼 수 있었고, 올바른 가르침과 배움의 공간으로 학교가 거듭날 수 있도록 학교, 지역, 가정이 힘을 합쳐야 할 것이다. 급변하는 시대에 고등학생의 주관적 학업성취도를 키울 수 있도록 여건을 마련하되 인지적 영역에서만이 아니라 정의적 영역에서도 바람직한 교육하는 인간으로 키워내야 할 것이다.



참 고 문 헌

- 강만모, 김상락, 박상무(2012). 빅데이터의 분석과 활용. **정보과학회지**, 30(6), 25-32.
- 강명희, 유영란, 유지원(2014). 국가수준 학업성취도 평가에 나타난 고등학생의 지각된 교사태도, 학습태도, 학업성취도, 학교생활만족도 간의 구조적 관계 규명. **교육과학연구**, 45(1), 181-203.
- 강필성(2017). 텐서플로에 올라 탄 한국 AI, 구글은 왜 기술을 개방했나?. (사) 한국인공지능협회, AI웹진. <<https://www.koraia.org/single-post/2017/03/08>>에서 2017년 11월 1일 검색.
- 곽수란(2004). 가족배경이 인문계 고등학교 학생의 학업성취에 미치는 영향. **교육연구**, 17, 51-60.
- 곽수란(2006). 인문계 학생의 학업성취 결정요인. **교육사회학연구**, 16(2), 1-29.
- 권영옥(2013). 빅데이터를 활용한 맞춤형 교육 서비스 활성화 방안연구. **지능정보연구**, 19(2), 87-99.
- 권용재, 이광현(2012). 비만이 학업성취도에 미치는 영향 연구: 2차년도 한국아동·청소년 패널조사를 이용하여. **한국청소년정책연구원 학술대회자료집**, 60-76.
- 권지은(2003). 부모 및 또래 애착, 문제해결방식과 자아탄력성의 관계. 이화여대 대학원 석사학위논문.
- 권지은, 이은정, 노충래(2013). 청소년이 지각한 방임과 학대가 학업성적에 미치는 영향: 교사관계 및 교우관계 매개효과 분석. **한국아동복지학**, 42, 29-54.
- 금지현, 손찬희, 채수은, 강성국(2013). 중학생의 학교생활적응과 가정 및 개인 관련 변인 간의 관계. **청소년학연구**, 20(2), 119-144.
- 김경근(2005). 한국사회 교육격차의 실태 및 결정요인. **교육사회학연구**, 15(3), 1-27.
- 김경근, 연보라, 장희원(2014). 서울시 중고등학생의 학업성취 영향요인 및 그 함의. **교육사회학연구**, 24(4), 1-29.
- 김경근, 장희원(2016). 교사의 성취압력이 학업성취에 미치는 영향: 가정배경에 따른 차별적 효과중심으로. **한국교육학연구**, 22(2), 55-76.
- 김경식, 이현철(2009). 학습태도 결정에 영향을 주는 학생 내·외적 요인 분석: 인문계 고등학생을 중심으로. **중등교육연구**, 57(2), 181-207.

- 김경식, 이현철(2010). 사회적 자본으로서의 부모, 교사, 교우 관계와 학업성취 간의 종단적 관계 분석. **중등교육연구**, 58(1), 203-225.
- 김광혁, 차유림(2007). 가구빈곤이 청소년의 학교적응에 영향을 미치는 과정분석: 부모의 교육적 관여(투자)와 부모-자녀 관계의 매개효과를 중심으로. **한국청소년연구**, 18(1), 29-51.
- 김서현, 임혜림, 정익중(2015). 학습의 자기주도성은 학업성취 이외에 대인관계에도 영향을 미치는가?. **청소년복지연구**, 17(1), 305-332.
- 김선숙, 고미선(2007). 청소년의 학업성취 변화에 영향을 주는 요인: 잠재성장모형을 적용하여. **한국청소년연구**, 18(3), 5-29.
- 김신일(2011). **교육사회학(제4판)**. 파주: 교육과학사.
- 김양분, 임현정, 김난옥(2012). 학업성취에 대한 학급 및 교사 요인의 영향. **한국교육**, 39(2), 157-179.
- 김영숙, 조한익(2015). 학교생활적응, 학업성취도 및 주관적 안녕감의 종단적 인과관계 분석. **교육심리연구**, 29(4), 845-871.
- 김용대, 조광현(2013). 빅데이터와 통계학. **한국데이터정보과학지**, 24(5), 959-974.
- 김용련(2017). 마을교육공동체의 미래: 지역과 교사가 중심이 되는 학습플랫폼. **좋은교사**, 11, 20-28.
- 김유리, 김현철, 박진옥, 권선향, 김누리(2016). 인성교육의 새로운 패러다임. **한국교육문제연구**, 34(4), 1-20.
- 김윤대, 전치혁, 이해선(2011). 별점 부분최소자승법을 이용한 분류방법. **한국데이터정보과학회지**, 22(5), 931-940.
- 김완섭(2012). 로지스틱 회귀분석과 데이터마이닝 분석을 이용한 컴퓨터 교양교육 성과의 요인에 대한 연구. **교양교육연구**, 6(3), 743-767.
- 김지경, 백혜정, 임희진, 이계오(2010). 한국아동·청소년패널조사2010 I. **한국청소년정책연구원 연구보고서**, 1-236.
- 김지혜(2015). 청소년의 자아존중감과 우울이 휴대전화 의존을 매개로 학업성적과 학교부적응에 미치는 영향. **청소년시설환경**, 13(1), 65-72.
- 김진아(2016). 부모양육태도에 따른 청소년의 학교적응과 학업성취도 간의 종단적 관계. **상담학연구**, 17(2), 303-326.
- 김충일, 이강이(2015). 아동이 지각한 부모의 성취지향적 양육태도가 자아탄력성을 매개로 학교생활적응과 학업성취도에 미치는 영향. **대한가정학회**, 53(5),

503-517.

- 김현주, 이병훈(2007). 자녀의 학업성취에 미치는 가족배경, 사회자본 및 문화자본의 영향. **한국인구학**, 30(1), 125-148.
- 김현주, 김원경(2017). 데이터 마이닝을 이용한 수학 교과의 자기효능감과 내재적 동기에 영향을 미치는 주요 요인 분석. **교원교육**, 33(3), 77-104.
- 김현진(2007). 가정배경과 학교교육 그리고 사교육이 학업성취에 미치는 영향 분석. **교육행정학연구**, 25(4), 485-508.
- 김혜숙(2012). ICT 활용이 학업성취도에 미치는 영향:PISA 2009 한국 자료를 중심으로. **아시아교육연구**, 13(1), 1-22.
- 나중화(2017a). R 데이터마이닝. 파주: 자유아카데미.
- 나중화(2017b). R 응용회귀분석. 파주: 자유아카데미.
- 나중화(2017c). R 응용다변량분석. 파주: 자유아카데미.
- 노현종, 손원숙(2015). 교사의 숙제 피드백이 학생의 자기조절학습, 과제가치, 학습태도 및 학업성취도에 미치는 영향. **교육평가연구**, 28(3), 879-902.
- 문은식, 배정희(2010). 학습사 및 사회적·동기적 변인과 학습된 무기력의 관계: 남고생과 여고생의 비교. **교육심리연구**, 24(1), 183-205.
- 박경미, 최승현(2002). 학업성취도 국제 비교 연구(PISA)에 나타난 수학적 소양의 성별 차이에 대한 고찰. **수학교육**, 41(3), 319-328.
- 박경민, 정승원, 신안나(2013). 초·중학생 부모의 양육태도에 대한 프로파일 분석: 학교적응 및 학업성취도. **한국청소년연구**, 24(2), 91-118.
- 박도순(2017). 수능은 변별을 위해 만든 시험이 아닙니다. **좋은교사**, 11월호, 58-68.
- 박민수, 김태현, 조은석, 김희발, 오희석(2014). R을 이용한 별점화 축소추정 기법 비교연구: 요크셔 돼지 산자수와 SNP에 대한 적용 사례. **농업생명과학연구**, 48(3), 147-155.
- 박수원, 최병호(2017). 진로선택 가치와 학습동기 및 학업성취 간 관계: 사회공헌 가치를 중심으로. **미래청소년학회지**, 14(2), 1-20.
- 박은정, 이유리, 이성훈(2015). 부모의 방임과 아동의 학업성취의 관계: 자기조절학습능력과 공동체의식의 매개효과. **한국생활과학회지**, 24(6), 755-768.
- 박은정, 이유리, 이성훈(2016). 남녀 고등학생의 생애목표 유형에 따른 자기조절학습과 학업성취의 차이. **교육종합연구**, 14(3), 1-27.
- 박지영, 정현수(2016). 남녀중학생이 지각한 부모양육태도가 진로정체감에 미치

- 는 영향: 학업성취도의 매개효과. **한국생애학회지**, 6(1), 21-37.
- 박창이, 김용대, 김진석, 송종우, 최호식(2015). **R을 이용한 데이터마이닝(개정판)**. 서울: 교우사.
- 박철용, 계묘진(2013). 호흡곤란 환자 퇴원 결정을 위한 별점 로지스틱 회귀모형. **한국데이터정보과학회지**, 24(1), 125-133.
- 박철용(2013). 라소를 이용한 간편한 주성분분석. **한국데이터정보과학회지**, 24(3), 533-541.
- 박형주(2017). 미래세대에게 언제까지 죽은 수학을 가르칠 건가. NEWSIS <<http://v.media.daum.net/v/20171202090405551?rcmd=rn>>에서 2017년 12월 2일 검색.
- 반상진, 정성석, 양성관(2005). 과외가 학습 성취에 미치는 영향 분석. **제1회 한국교육고용패널 학습대회 논문집**. 483-517.
- 배호중, 유비(2015). 중학생의 비만과 학교생활: 학교적응도 및 학업성취도를 중심으로. **한국청소년연구**, 26(1), 79-109.
- 서지영(2015). 중학생의 BMI(체질량 지수) 정도에 따른 학교적응 영향요인. **한국데이터정보과학회지**, 26(2), 429-440.
- 석경하, 이태우(2013). 데일리 렌즈 데이터를 사용한 데이터마이닝 기법 비교. **한국데이터정보과학회지**, 24(6), 1341-1348.
- 성은모, 진성희(2012). 청소년의 성별에 따른 휴대폰 활용이 자기조절학습능력 및 학업성취도에 미치는 영향. **교육정보미디어연구**, 18(4), 441-467.
- 손병덕, 허계형(2016). 방임·학대 경험아동의 성취가치와 자기관리 형성이 학업성취 수준에 미치는 영향 연구. **청소년학연구**, 23(9), 305-323.
- 송상윤(2015). 예대금리차 결정요인 모형의 예측력 비교 연구: Ridge, LASSO 및 Elastic Net 방법론을 중심으로. **금융지식연구**, 13(3), 41-65.
- 송영주(2016). 학업성적 만족도와 또래애착이 자아존중감에 미치는 영향: 후기 아동기와 초기 청소년기 비교. **한국생애학회지**, 6(3), 1-15.
- 신정철, 정지선, 신태수(2008). 대학생의 학업성취도와 그 영향요인들 간의 인과관계 분석. **교육행정학연구**, 26(1), 287-313.
- 신종호, 신태섭(2006). 고등학생의 학업성취와 학업적 자기효능감, 지각된 교사기대, 가정환경요인 간의 관계 연구. **아동교육**, 15(1), 5-23.
- 심도현, 강현구, 정수지(2014). 아동·청소년의 사회경제적 지위가 학업성취도, 건강상태, 자아존중감에 미치는 영향: 연간가구소득 차이를 중심으로. **한**

- 국청소년정책연구원 학술대회자료집, 591-635.
- 양명희, 이경아(2012). 자기조절학습과 학업성취도의 관련성. **교육과학연구**, 43(2), 175-195.
- 양혜린, 이재은, 한기주(2014). 방임경험이 청소년의 학업성적에 미치는 영향: 진로정체감의 조절효과를 중심으로. **사회과학연구**, 38(2), 41-62.
- 여성가족부(2017). 청소년백서 및 한국청소년상담복지개발원 상담통계. <<http://kostat.go.kr>>에서 2017년 9월 3일 검색.
- 유예림(2017). 빅데이터 분석 기법을 활용한 2015 개정 교육 과정 정책에 대한 언론 보도 분석. **서울대학교 대학원 박사학위논문**.
- 유재봉(2016). 학교 인성교육의 문제점과 방향. **교육철학연구**, 38(3), 99-119.
- 유진은(2015a). **한 학기에 끝내는 양적연구방법과 통계분석**. 서울: 학지사.
- 유진은(2015b). 우리나라 교육 승인통계에 대한 평가와 향후 과제. **아시아교육연구**, 16(2), 175-199.
- 유진은(2015c). 랜덤 포레스트: 의사결정나무의 대안으로서의 데이터 마이닝 기법. **교육평가연구**, 28(2), 427-448.
- 유진은(2016). 데이터 마이닝 기법을 통한 교육 패널데이터 분석: 별점 회귀모형과 KYPs 자료. **아시아교육연구**, 17(3), 1-19.
- 유진은(2017). 기계학습을 통한 TIMSS 2011 중학생의 수학 성취도 관련 변수 탐색. **교원교육**, 33(1), 43-56.
- 유진은, 노민정(2017). Group lasso를 통한 중학생의 삶의 만족도에 영향을 미치는 변수 탐색. **한국청소년연구**, 28, 127-149.
- 윤영돈(2015). 학교급별 인성 수준과 고교 인성교육의 방향설정. **한국교원교육연구**, 32(4), 153-172.
- 이귀옥, 이미리(2013). 초등학교 저학년, 고학년, 그리고 중학생의 학교적응에 미치는 개인, 가족, 학교 변인 탐색. **생애학회**, 3(1), 13-28.
- 이부일, 이서현(2016). **R과 함께한 데이터 여행**. 서울: 경문사.
- 이수진, 신현주, 이지혜(2014). 중·고등학생의 학업자신감, 자기조절학습전략, 학업성취도와 인터넷중독성향의 구조적 관계. **청소년학연구**, 21(2), 349-374.
- 이승진, 김경근(2015). 가정배경, 양육태도, 교우관계, 오락적 휴대전화 이용행태와 학업성취 간 구조적 관계. **한국교육학연구**, 21(2), 41-66.
- (2017). 학교, 마을의 느티나무를 꿈꿀 수 있을까?. **좋은교사**, 11월호, 37-43.

- 이유리, 박은정, 이성훈(2015). 중학생의 학습습관 유형에 따른 학업성취와 삶의 만족도의 차이. **학습자중심교과교육연구**, 15(11), 621-641
- 이은경, 박도영, 최인봉(2014). 교육 빅데이터 관련 연구 동향. **한국컴퓨터정보학회 학술발표 논문집**, 22(2), 175-176.
- 이자형, 이기혜(2015). 청소년의 건강불평등 영향요인 분석: 주관적 건강상태와 비만여부를 중심으로. **한국교육학연구**, 21(2), 217-243.
- 이정진(2011). **R, SAS, MS-SQL을 활용한 데이터마이닝**. 파주: 자유아카데미.
- 이종원(2013). 청소년의 학업성적과 비행, 비행친구의 관계에 대한 중단연구. **한국청소년정책연구원 연구보고서**, 1-82.
- 이중섭, 이용교(2011). 부모의 학력이 청소년 자녀의 학업성취에 미치는 영향. **한국가족복지학회**, 16(1), 65-88.
- 이재길(2016). **R 프로그램에 기반한 다변량분석 및 데이터마이닝**. 서울: 황소걸음아카데미.
- 이현숙, 신진아, 김경희(2013). 다층 구조방정식모형을 활용한 교육 맥락변인과 학업성취도의 관계 분석. **교육평가연구**, 26(2), 477-506.
- 이혜주, 정의현(2013). 데이터마이닝을 이용한 학업성취 결정요인 탐색. **아동교육**, 22(2), 5-18.
- 임규혁(1991). 학업성취의 누적적 경험과 정신건강. **교육문제연구**, 3, 23-90.
- 임다희, 권기현(2012). 서울시 고등학생의 학업성취도 결정요인. **한국정책분석평가학회**, 22(2), 43-63.
- 임선아(2014). (구조관계모형 검증을 통한) 부모의 민주적, 비민주적 양육태도가 초등학교 아동의 자기조절능력과 학교적응력 및 학업성취에 미치는 영향 분석. **초등교육연구**, 27(2), 73-93.
- 임양미(2014). 부모의 사교육비 및 감독·애정, 자녀의 학습가치와 자기조절학습능력이 학업성취도에 미치는 영향: 중학생의 성별 비교를 중심으로. **한국가정과교육학회지**, 26(3), 113-131.
- 임현정, 시기자, 김성은(2016). 학생 학업성취 변화의 영향요인 탐색: 국가수준 학업성취도 평가 세 시점 연계자료 분석. **교육평가연구**, 29(1), 123-145.
- 임효진, 김재철(2014). 중학교 학생들의 성취가치, 자기조절, 학업적응과 학업성취의 구조적 관계: 자아탄력성의 조절효과. **청소년학연구**, 21(6), 161-190.
- 임희진(2010). **한국 고등학생의 학업에 관한 조사**. 서울: 한국청소년정책연구원.

- 장상호(2009). **학문과 교육(중 II): 교육본위의 삶**. 서울: 서울대학교출판문화원.
- 전영상, 최영신(2017). 학업성적이 부모와 자신의 행복에 미치는 영향. **청소년학연구**, 24(2), 473-490.
- 정제영, 정예화(2015). 부모의 교육적 관여 수준이 학생의 학업성취도에 미치는 영향. **청소년연구**, 22(7), 73-93.
- 조은순, 남민우, 조재운, 김현진(2016). **미래핵심역량 중심 세종형 학생평가 모델 개발**(정책 373.7-16-3). 세종: 세종특별자치시교육청.
- 조한익, 이나영(2010). 청소년의 학업성취와 자아탄력성과의 관계. **청소년학연구**, 17(6), 1-20.
- 진성희, 성은모(2013). 중학생의 성별에 따른 유해사이트 활용이 학습 및 학교생활에 미치는 영향. **아시아교육연구**, 14(3), 293-318.
- 진의남(2012). 학교 교육에서 인성교육의 인식과 개선 요구. **실과교육연구**, 18(3), 145-169.
- 추상엽, 임성문(2007). 가족구조와 고등학생의 학업성취 간의 관계: 부모소득, 학업지원행동의 매개효과. **청소년학연구**, 14(3), 135-157.
- 추상엽, 임성문(2008). 부모의 성취압력과 학업성취 간의 관계. **청소년학연구**, 15(7), 347-368.
- 최소현, 이학래, 박천건, 이경은(2017). 사례연구: 대구 파티마 병원 폐렴 입원 환자 수에 영향을 미치는 날씨 변수 선택. **한국데이터정보과학회지**, 28(1), 131-142.
- 최제영, 박충식(2013). 빅데이터와 교육. **한국컴퓨터정보학회지**, 21(1), 9-23.
- 통계청(2015). 청소년 통계 보도자료. <<http://kostat.go.kr>>에서 2017년 9월 3일 검색.
- 한국청소년정책연구원(2017). **한국아동·청소년 패널조사 1-7차 조사 데이터 유저 가이드**. 서울: 한국청소년정책연구원.
- 한수경, 박재범, 손형국, 양정호(2015). 사교육과 자기주도학습이 서울시 일반계 고등학생의 수학학업성취에 미치는 영향. **한국교육학연구**, 21(1), 197-233.
- 한희원, 문경숙(2014). 중학생의 자기조절학습을 매개로한 진로정체감이 학업성취에 미치는 영향. **교육종합연구**, 12(4), 25-42.
- 허균(2014). 중학생의 성취가치와 학업성취 관계에서 자기통제력과 자아존중감의 다중매개 효과분석. **사고개발**, 10(2), 113-126.
- 허균(2017). 고등학생들의 학습 목적 컴퓨터 사용과 고3의 주관적 학업성취도 관계 연구. **수산해양교육연구**, 29(1), 180-187.

- 허은정, 이재덕(2014). 고등학생의 학습유형별 참여 시간이 학업성취도 및 학업 자아개념에 미치는 영향: 자기주도학습 시간의 매개효과를 중심으로. **한국 교육문제연구**, 21(4), 121-139.
- Asendorpf, J. B., & Aken, M. A. G. (1999). Resilient, overcontrolled, and undercontrolled personality prototypes in childhood: replicability, predictive power, and the trait-type issue. *Journal of Personality and Social Psychology*, 77(4), 815-832.
- Block, J., & Kermen, A. M. (1996). IQ and Ego-resiliency: Conceptual and empirical connections and separateness. *Journal of Personality and social psychology*, 70(2), 349-361.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939-944.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.)*. New York: Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. New York: Springer.
- Lehrer, J. (2009). DON'T! The secret of self-control. *New York, May 18*, 26-32.
- Lavelle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 5(2), 21-32.
- Matthew, D. L. (2013). *Social: why our brains are wired to connect*. New York: Brockman. 최호영(역)(2015). **사회적 뇌**. 서울: 시공사.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of Royal Statistics Society B*, 58, 267-288.

A B S T R A C T

Predictor exploration via LASSO: Focusing on
high school student's subjective academic achievement

Park, Jin Young

Major in Curriculum and Educational Evaluation
Graduate School of Korea National University of Education
Chung-Buk, Korea

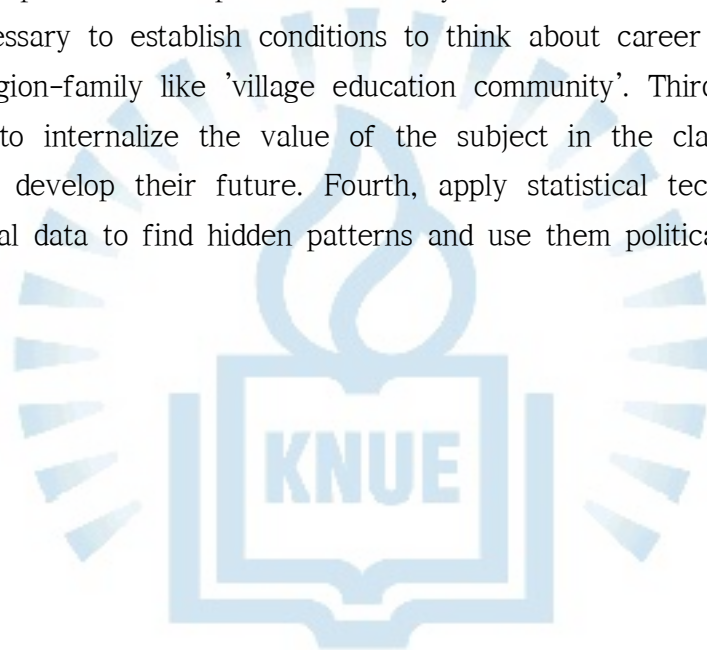
Supervised by Professor Yoo, Jin Eun, Ph.D.

The purpose of this study is to explore predictors affecting the subjective academic achievement of high school third graders, via LASSO, using Middle School 1 Panel data of Korean Children & Youth Panel Survey(KCYPS). The R program was used as an analysis tool. Among 270 KCYPS predictors after data cleaning had been conducted, a total of 29 were selected in the LASSO model. The prediction accuracy was determined to be 85.98%.

Weight, achievement value, proficiency goal orientation, behavior control, academic time management, attention, depression, self-esteem, ego-resilience, self-identity, future education level, academic time, parental rearing styles, learning activities, the life satisfaction of the guardian were found to have an effect on academic achievement, consistent with the findings of previous research. Community

consciousness, volunteer activities experience, the number of brother and sister, and school rules compliance were newly explored predictors, not yet investigated in previous research. Among the selected predictors, important predictors were behavioral control, achievement value, proficiency orientation, academic time management, future hope level of education, and school adjustment.

The implications of this study are as follows. First, it is necessary to prepare a plan to develop students' ability to control behavior. Second, it is necessary to establish conditions to think about career by linking school-region-family like 'village education community'. Third, teaching students to internalize the value of the subject in the class so that they can develop their future. Fourth, apply statistical techniques to educational data to find hidden patterns and use them politically.



※ A thesis submitted to the Committee of the Graduate School of Korea National University of Education in partial fulfillment of the requirements for the degree of Master of Education in February, 2018.

부 록

[부록 1] KCYPS 자료를 활용한 학업성취도 관련 선행연구 목록

[부록 2] 데이터 전처리를 위한 R 프로그램 code

[부록 3] 설명변수(예측변수) 데이터 전처리 R 프로그램 code

[부록 4] ‘전체 성적 주관적 평가’ 재코딩 후 모형 적합 결과



[부록 1] KCYPs 자료를 활용한 학업성취도 관련 선행연구 목록

순	연구 제목	연구자 (연도)	분석 자료	분석 방법
1	비만이 학업성취도에 미치는 영향 연구: 2차년도 한국아동·청소년 패널조사를 이용하여	권용재, 이광현 (2012)	초4패널, 중1패널 각 2차	중 다 회 귀 분석 프로빗 모 형
2	청소년이 지각한 방임과 학대가 학업 성적에 미치는 영 향: 교사관계 및 교우관계 매개효과 분석	권지은, 이은정, 노충래 (2013)	중1패널 2차	SEM
3	중학생의 학교생활적응과 가정 및 개인 관련 변인 간의 관계	금지현, 손찬희, 채수은 (2013)	중1패널 2차	SEM
4	학습의 자기주도성은 학업성취 이외에 대인관계에도 영 향을 미치는가?	김서현, 임혜림, 정익중 (2015)	중1패널 3차	SEM
5	학교생활적응, 학업성취도 및 주관적 안녕감의 종단적 인 과관계 분석	김영숙, 조한익 (2015)	초4패널 2~5차	자 기 회 귀 교차지연모 형(ARCL)
6	청소년의 자아존중감과 우울이 휴대전화 의존을 매개로 학업성적과 학교부적응에 미치는 영향	김지혜 (2015)	중1패널 3차	SEM
7	부모양육태도에 따른 청소년의 학교적응과 학업성취도 간의 종단적 관계	김진아 (2016)	중1패널 1~4차	LGM
8	아동이 지각한 부모의 성취지향적 양육태도가 자아탄력 성을 매개로 학교생활적응과 학업성취도에 미치는 영향	김충일, 이강이 (2015)	초4패널 1차	SEM
9	초·중학생 부모의 양육태도에 대한 프로파일 분석: 학 교적응 및 학업성취도	박경민, 정승원, 신안나 (2013)	초4패널, 중1패널 각1,2차	다 차 원 척 도법 PAMS 모 형
10	부모의 방임과 아동의 학업성취의 관계: 자기조절학습능 력과 공동체적응의 매개효과	박은정, 이유리, 이성훈 (2015)	초4패널 3차	매 개 회 귀 분석
11	남녀 고등학생의 생애목표 유형에 따른 자기조절학습과 학업성취의 차이	박은정, 이유리, 이성훈 (2016)	중1패널 5차	군집분석 ANOVA
12	남녀중학생이 지각한 부모양육태도가 진로정체감에 미치 는 영향: 학업성취도의 매개효과	박지영, 정현수 (2016)	초4패널 4차	SEM
13	중학생의 비만과 학교생활: 학교적응도 및 학업성취도를 중심으로	배호중, 유비 (2015)	중1패널 1, 2차	다 중 회 귀 분석 OLS회귀분 석
14	중학생의 BMI정도에 따른 학교적응 영향요인	서지영 (2015)	중1패널 2차	다 중 회 귀 분석
15	청소년의 성별에 따른 휴대폰 활용이 자기조절학습능력 및 학업성취도에 미치는 영향	성은모, 진성희 (2012)	중1패널 2차	다 중 회 귀 분석

16	방임·학대 경험아동의 성취가치와 자기관리 형성이 학업성취 수준에 미치는 영향 연구	손병덕, 허계형 (2016)	초1패널 5차	다 중 회 귀 분석
17	학업성적 만족도와 또래애착이 자아존중감에 미치는 영향: 후기 아동기와 초기 청소년기 비교	송영주 (2016)	초4패널 2, 4차	경로분석 다 중 집 단 분석
18	아동·청소년의 사회경제적 지위가 학업성취도, 건강상태, 자아존중감에 미치는 영향: 연간가구소득 차이를 중심으로	심도현, 강현구, 정수지 (2014)	초1패널 3차, 초4 패널 2,4차, 중1패 널 3차	SEM
19	방임경험이 청소년의 학업성적에 미치는 영향: 진로정체감의 조절효과를 중심으로	양혜린, 이재은, 한기주 (2014)	중1패널 3차	중 다 회 귀 분석
20	초등학교 저학년, 고학년, 그리고 중학생의 학교적응에 미치는 개인, 가족, 학교 변인 탐색	이귀옥, 이미리 (2013)	초1패널, 초4패널, 중1패널 각 2차	단 계 적 중 다 회 귀 분 석
21	가정배경, 양육태도, 교우관계, 오락적 휴대전화 이용행태와 학업성취 간 구조적 관계	이승진, 김경근 (2015)	중1패널 2차	SEM
22	중학생의 학습습관 유형에 따른 학업성취와 삶의 만족도의 차이	이유리, 박은정, 이성훈 (2015)	중1패널 3차	군집분석 ANOVA
23	청소년의 건강불평등 영향요인 분석: 주관적 건강상태와 비만여부를 중심으로	이자형, 이기혜 (2015)	중1패널 3차	HLM HGLM
24	청소년의 학업성적과 비행, 비행친구의 관계에 대한 중단 연구	이종원 (2013)	중1패널 2,3차	교 차 지 연 패 널 분 석, SEM
25	부모의 민주적, 비민주적 양육태도가 초등학교 아동의 자기조절능력과 학교적응력 및 학업성취에 미치는 영향 분석	임선아 (2014)	초4패널 1~3차	SEM
26	부모의 사교육비 및 감독·애정, 자녀의 학습가치와 자기조절학습능력이 학업성취도에 미치는 영향: 중학생의 성별 비교를 중심으로	임양미 (2014)	중1패널 3차	위계적회귀 분석
27	중학교 학생들의 성취가치, 자기조절, 학업적응과 학업성취의 구조적 관계: 자아탄력성의 조절효과	임효진, 김재철 (2014)	중1패널 3차	SEM
28	학업성적이 부모와 자신의 행복에 미치는 영향	전영상, 최영신 (2017)	초4패널 2~6차	LGM
29	중학생의 성별에 따른 유해사이트 활용이 학습 및 학교생활에 미치는 영향	진성희, 성은모 (2013)	중1패널 2차	ANOVA 단 순 회 귀 분석
30	중학생의 자기조절학습을 매개로한 진로정체감이 학업성취에 미치는 영향	한희원, 문경숙 (2014)	중1패널 3차	SEM
31	중학생의 성취가치와 학업성취 관계에서 자기통제력과 자아존중감의 다중매개 효과분석	허 군 (2014)	중1패널 1,3차	SEM
32	고등학생들의 학습목적 컴퓨터 사용과 고3의 주관적 학업성취도 관계 연구	허 군 (2017)	중1패널 6차	SEM

[부록 2] 데이터 전처리를 위한 R 프로그램 code

가. 학부모 또는 학생이 미응답한 관측치와 분석에 불필요하다고 판단된 변수(문항) 삭제 후 R 프로그램을 이용하여 학생, 학부모가 6차 조사에 참여한 데이터 추출 후 조사 참여여부 변수 삭제함
: 2351명의 499개 변수 → 1938명의 367개 변수

```
> K <- read.csv("K16thNA.csv") #csv 파일을 로딩하여 K에 저장
> str(K)
'data.frame':      2351 obs. of  367 variables:
 $ SURVEY1w6: int  1 2 1 2 1 1 1 1 1 1 ...
 $ SURVEY2w6: int  1 2 1 2 1 1 1 1 1 1 ...
      중략
> K1<- subset(K, SURVEY1w6==1 & SURVEY2w6 == 1) #응답한 자료만 가져옴
> colnames(K1) <-gsub("w6", "", colnames(K1)) #데이터명 변경
> str(K1)
'data.frame':      1938 obs. of  367 variables:
 $ SURVEY1: int  1 1 1 1 1 1 1 1 1 1 ...
 $ SURVEY2: int  1 1 1 1 1 1 1 1 1 1 ...
      중략
> K1[, 1:2]<- NULL #SURVEY1, SURVEY2 변수 제거
```

나. 고교 졸업 후 진로계획 영역에 대학 진학 또는 진학 준비에 응답한 학생들 데이터 추출: 전체의 80.4% 데이터 추출

```
> library(prettyR)
> freq(K1$FUR1C) #결측치는 없고, 1(진학 희망)의 응답율은 80.4%임
Frequencies for K1$FUR1C
      1      2      3      NA
1559  274  105      0
%    80.4 14.1  5.4    0
%!NA 80.4 14.1  5.4
> K2 <- subset(K1, FUR1C==1) #진학 희망 데이터 추출
```

```
> which(colnames(K2)=="FUR1C") # "FUR1C" 데이터의 위치
```

```
[1] 134
```

```
> K2[, 134] <- NULL # "FUR1C" 데이터 삭제
```

```
> str(K2)
```

```
'data.frame':      1559 obs. of  364 variables:
```

```
 $ GENDER : int  1 1 1 1 1 1 2 1 2 1 ...
```

```
 $ PARENT : int  1 1 2 1 1 1 1 1 1 1 ...
```

중략

다. 최종학력(부친 HAK2A, 모친 HAK2B, 보호자 HAK2C), 직업: 근로 여부 (부친 JOB1A, 모친 JOB1B, 보호자 JOB1C) 문항에서 ‘해당사항 없음’ 으로 응답한 것, 진로계획 (장래 희망 교육수준 FUR1A01, [보호자]장래 희망 교육수준 FUR1A02) 문항에서 ‘아직 결정하지 않았다’ 로 응답한 자료를 결측치(NA)로 재코딩

```
> library(car) #'NA'로 재코딩
```

```
> K2$HAK2A <- recode(K2$HAK2A, "6=NA")
```

```
> K2$HAK2B <- recode(K2$HAK2B, "6=NA")
```

```
> K2$HAK2C <- recode(K2$HAK2C, "6=NA")
```

```
> K2$JOB1A <- recode(K2$JOB1A, "3=NA")
```

```
> K2$JOB1B <- recode(K2$JOB1B, "3=NA")
```

```
> K2$JOB1C <- recode(K2$JOB1C, "3=NA")
```

```
> K2$FUR1A01 <- recode(K2$FUR1A01, "6=NA")
```

```
> K2$FUR1A02 <- recode(K2$FUR1A02, "6=NA")
```

라. 학업관련시간, 여가시간 조사 문항과 컴퓨터 사용시간 조사 문항 중 분단위로 응답하도록 한 문항 삭제

```
> K3 <- subset(K2, select = -c(TIM2B, TIM2D, TIM3B, TIM3D, TIM4B, TIM4D,  
+                               TIM5B, TIM5D, TIM6B, TIM6D, TIM7B, TIM7D,  
+                               TIM8B, TIM8D, TIM9B, TIM9D, MDA1B02,  
+                               MDA1B04))
```

마. 영과 영근처 분산을 가지는 변수 삭제

```
> library(caret)
> dim(K3)
[1] 1559 346
> nzv <- nearZeroVar(K3)
> nzv #문제가 되는 변수의 위치 반환
[1] 3 4 10 163 164 165 166 167 168 169 170 171 172 173
[15] 174 175 176 177 178 179 180 199 246 296 314 316 317 318
[29] 319 323 324 325 329 341
> filteredK3 <- K3[, -nzv] #영-근처 분산 가지는 변수 제거
> K33 <- filteredK3
> dim(K33)
[1] 1559 312
```

바. 결측값이 10% 넘는 변수 삭제 후 결측값을 포함하는 개체 삭제
: 271개 문항에 대해 1092명으로 정리

```
> library(VIM)
> aggr(K33, numbers=TRUE, sortVars=TRUE)
Variables sorted by number of missings:
Variable      Count      Variable      Count
HAK2C 0.9955099423  INT1C05 0.3072482360
JOB1C 0.9955099423  MDA1B01 0.2777421424
ACT1C03 0.9762668377  MDA1B03 0.2777421424
ACT1C08 0.9762668377  MDA1C 0.2777421424
ACT1C04 0.9743425273  MDA1D01 0.2777421424
ACT1C02 0.9627966645  MDA1D02 0.2777421424
ACT1C09 0.9454778704  MDA1D03 0.2777421424
ACT4B 0.9441949968  MDA1D04 0.2777421424
ACT1C05 0.8967286722  MDA1D05 0.2777421424
ACT1C01 0.8646568313  MDA1D06 0.2777421424
ACT1C07 0.8479794740  MDA1D07 0.2777421424
ACT1C06 0.7742142399  MDA1D08 0.2777421424
MDA4B 0.7030147530  MDA1D09 0.2777421424
INT1C10 0.6472097498  MDA1D10 0.2777421424
```

INT1C09 0.6427196921	INT1C11 0.1988454137
INT1C12 0.3745991020	FUR1A02 0.1109685696
JOB4B 0.3470173188	FAM5A 0.1000641437
INT1C04 0.3258499038	JOB4A 0.0936497755

중략

```
> K4 <- subset(K33, select = -c(HAK2C, JOB1C, JOB4B, INT1C04,INT1C05,
  INT1C09,INT1C10, INT1C12, MDA4B, ACT1C01, ACT1C02, ACT1C03,
  ACT1C04, ACT1C05, ACT1C06, ACT1C07, ACT1C08, ACT1C09,
  MDA1D01, ACT4B, MDA1B01, MDA1B03, MDA1C, MDA1D02,
  MDA1D03, MDA1D04, MDA1D05, MDA1D06, MDA1D07,MDA1D08,
  MDA1D09 , MDA1D10,INT1C11,FUR1A02, FAM5A, JOB4A))
> K5 <- na.omit(K4)
> dim(K5)
[1] 1092 271
> K6 <- K5
```



[부록 3] 설명변수(예측변수) 데이터 전처리 R 프로그램 code

가. 이분형 문항 ‘0’ 과 ‘1’ 로 더미코딩

```
> K6$GENDER <- ifelse(K6$GENDER == 1, 0, 1)
> K6$JOB1B <- ifelse(K6$JOB1B == 1, 1, 0)
> K6$DLQ1A01 <- ifelse(K6$DLQ1A01 == 1, 1, 0)
> K6$DLQ1A02 <- ifelse(K6$DLQ1A02 == 1, 1, 0)
> K6$FRN3A <- ifelse(K6$FRN3A == 1, 1, 0)
> K6$MDA3A02 <- ifelse(K6$MDA3A02 == 1, 1, 0)
> K6$ACT1A01 <- ifelse(K6$ACT1A01 == 1, 1, 0)
> K6$ACT1A05 <- ifelse(K6$ACT1A05 == 1, 1, 0)
> K6$ACT1A06 <- ifelse(K6$ACT1A06 == 1, 1, 0)
> K6$ACT1A07 <- ifelse(K6$ACT1A07 == 1, 1, 0)
> K6$ACT1A09 <- ifelse(K6$ACT1A09 == 1, 1, 0)
> K6$ACT2A01 <- ifelse(K6$ACT2A01 == 1, 1, 0)
> K6$ACT4A <- ifelse(K6$ACT4A == 1, 1, 0)
> K6$ACT6A <- ifelse(K6$ACT6A == 1, 1, 0)
> K6$MDA1A <- ifelse(K6$MDA1A == 1, 1, 0)
```

나. 범주 재구성/성인용 매체 이용 빈도 응답 척도 수정

```
> K6$COEDU <- ifelse(K6$COEDU == 1 | K6$COEDU == 2, 1, 0 )
> K6$SCLTYP1 <- ifelse(K6$SCLTYP1 == 1 | K6$SCLTYP1 == 2
  | K6$SCLTYP1 == 4, 1, 0 )

# MDA4A (성인용 매체 이용빈도) 2 ->4, 4 -> 2
> K6$MDA4A <- ifelse(K6$MDA4A == 1, 1,
  ifelse(K6$MDA4A ==2, 4,
    ifelse(K6$MDA4A == 3, 3, 2)))
```


[부록 4] ‘전체 성적 주관적 평가’ 재코딩 후 모형 적합 결과
 ▶ ‘1(우수)’: 1(매우 우수) ~ 3(약간 우수), ‘0(미흡)’: 5(약간 미흡) ~ 7(매우 미흡)

가. 270개 변수를 모두 사용한 모형 적합 결과

가정배경 변인 1개, 개인발달 변인 13개, 발달환경 변인 12개로 모두 29개의 변수로 정리되었고 포함된 변수는 <표4-1>과 같다. 추가된 변수로는 가정배경 변인으로 모친의 최종학력과 개인발달 변인으로 정서문제에서 사회적 위축, 발달환경 변인으로 양육방식에서 과잉기대와 과잉간섭, 지역사회 인식, 휴대전화 종류, 팬덤활동 이었다. 추가된 변수와 삭제된 변수를 <표4-2>에 정리하였다.

<표4-1> 모형에 포함된 변수와 계수(270개 설명변수 사용)

변수	변수설명	코딩	계수
HAK2B	최종학력: 모친	Likert: 1(중졸 이하) ~ 5(대학원 졸)	0.048
INT1C01	과목별 주관적 평가: 국어	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.913
INT1C02	과목별 주관적 평가: 수학	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.807
INT1C03	과목별 주관적 평가: 영어	Likert: 1(매우 잘 했다) ~ 5(매우 못했다)	-0.691
INT1D	전체 성적 만족도	Likert: 1(매우 만족한다) ~ 4(전혀 만족하지 않는다)	-0.593
INT2A01	학습습관_성취가치: 학교 공부는 나에게 중요한 의미를 지닌다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.028
INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.094
INT2C01	학습습관_행동통제: 나는 공부가 지루하고 재미없더라도 끝까지 다한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.243
PSY1A03	정서문제_주의집중: 오랫동안 집중해야 하는 과제는 하고 싶지 않다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.007
PSY1C07	정서문제_신체증상: 숨쉬기 어려울 때가 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.008
PSY1D02	정서문제_사회적 위축: 부끄럼을 많이 탄다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.043

PSY1D04	정서문제_사회적 위축: 수줍어한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.003
PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.094
PSY2C07	자아인식_자아정체감: 나는 낯선 사람을 만나는 것을 꺼려한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.100
FUR1A01	진로계획_장래 희망 교육수준	Likert: 1(고등학교 졸업) ~ 5(대학원 박사학위 취득)	0.367
FUR2A03	대학 진학 이유: 부모님/가족이 원해서	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.163
FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.115
FAM1E03	가족구성_형제자매수: 여동생	#명으로 응답	0.048
FAM2D01	양육방식 I_과잉기대: 부모님(보호자)의 기대가 항상 내 능력 이상이어서 부담스럽다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.098
FAM2E02	양육방식 I_과잉간섭: 나에게 무엇을 하든지 항상 이겨야 한다는 걸 강조하신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.018
FAM2F01	양육방식 I_합리적 설명: 부모님(보호자)의 결정을 무조건 따르게 하기 보다는 왜 그래야 하는지 설명해 주신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.008
FAM2F03	양육방식 I_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.113
EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.724
EDU2D06	학교적응_교사관계: 졸업한 다음에도 담임선생님을 찾아 뵙고 싶다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.027
COM1A01	지역사회 인식: 나는 우리 동네 사람들 대부분을 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.023
COM2A02	공동체意識: 공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.028
MDA2D	휴대전화 종류	1(스마트폰), 0(그 외)	-0.028
ACT3B	여행/문화활동: 문화활동 연간 경험 횟수	#회로 응답	0.022
ACT6A	팬덤활동: 선호 연예인/운동선수 유무	1(있다), 0(없다)	-0.114

〈표4-2〉 재코딩 후 모형에서 추가된 변수와 빠진 변수

구분	변수 (변수 설명)	해석
추가 된 변수	HAK2B (최종학력: 모친)	학력이 높을수록
	INT2A01(학습습관_성취가치: 학교 공부는 나에게 중요한 의미를 지닌다)	그럴수록
	PSY1C07 (정서문제_신체증상: 숨쉬기 어려울 때가 있다)	그렇지 않을수록
	PSY1D02(정서문제_사회적 위축: 부끄럼을 많이 탄다) PSY1D04(정서문제_사회적 위축: 수줍어한다)	그럴수록
	FAM1E03(가족구성_형제자매수: 여동생)	많을수록
	FAM2D01(양육방식_과잉기대: 부모님(보호자)의 기대가 항상 내 능력 이상이어서 부담스럽다)	그럴수록
	FAM2E02(양육방식_과잉간섭: 나에게 무엇을 하든지 항상 이겨야 한다는 걸 강조하신다)	그럴수록
	COM1A01(지역사회 인식: 나는 우리 동네 사람들 대부분을 알고 있다)	그렇지 않을수록
	MDA2D (휴대전화 종류)	스마트폰 아닐수록
	ACT6A(팬덤활동: 선호 연예인/운동선수 유무)	없을수록
빠진 변수	PSY1E05(정서문제_우울: 울기를 잘한다)	그렇지 않을수록
	PSY2B07(자아인식_자아탄력성: 나는 같은 장소에 갈 때도 늘 가던 길 보다는 다른 길로 가 보는 것을 좋아한다)	그렇지 않을수록
	TIM4A(학업관련시간_학원/과외숙제시간: 등교일)	많을수록
	FAM3A01(양육방식Ⅱ_방임: 다른 일(직장이나 바깥일)보다 나를 더 중요하게 생각하신다)	그럴수록
	MDA1A(컴퓨터_사용여부)	사용하지 않을수록
	MDA2C05(휴대전화_의존도: 혼자 있을 때 휴대전화가 없으면 심심해서 견딜 수가 없다)	그럴수록
	ACT1A06(체험활동_연간 참여경험 유무: 봉사활동)	없을수록
	ACT2A01(동아리활동_연간 참여경험 유무: 교내 공식 동아리활동)	있을수록

〈표4-3〉 중요도 순으로 변수 정리

순	변수	변수설명 (결과)	순	변수	변수설명 (결과)
1	INT1C01	과목별 주관적 평가: 국어 (잘 할수록)	16	HAK2B	최종학력: 모친 (학력이 높을수록)
2	INT1C02	과목별 주관적 평가: 수학 (잘 할수록)	17	FAM1E03	가족구성_형제자매수: 여동생 (많을수록)
3	EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다 (그럴수록)	18	PSY1D02	정서문제_사회적 위축: 부끄럼 을 많이 탄다 (그럴수록)
4	INT1C03	과목별 주관적 평가: 영어 (잘 할수록)	19	COM2A02	공동체의식: 공휴일에 쉬지 못하더라도 복지기관에서 자 원봉사 활동을 할 수 있다 (그렇지 않을수록)
5	NT1D	전체 성적 만족도 (만족할수록)	20	MDA2D	휴대전화 종류 (스마트폰이 아닐수록)
6	FUR1A01	진로계획_장래 희망 교육수준 (교육수준이 높을수록)	21	INT2A01	학습습관_성취가치: 학교 공부 는 나에게 중요한 의미를 지 닌다 (그럴수록)
7	INT2C01	학습습관_행동통제: 나는 공부가 지루하고 재미없더라도 끝까지 다한다 (그럴수록)	22	EDU2D06	학교적응_교사관계: 졸업한 다 음에도 담임선생님을 찾아 뵈 고 싶다 (그렇지 않을수록)
8	FUR2A03	대학 진학 이유: 부모님/가족이 원해서 (그럴수록)	23	COM1A01	지역사회 인식: 나는 우리 동 네 사람들 대부분을 알고 있 다 (그렇지 않을수록)
9	FUR2D04	직업선택 요인: 국가/사회에 기 여할 수 있다 (그럴수록)	24	ACT3B	여행/문화활동: 문화활동 연간 경험횟수 (많을수록)
10	ACT6A	팬덤활동: 선호 연예인/운동선수 유무 (없을수록)	25	FAM2E02	양육방식 I_과잉간섭: 나에게 무엇을 하든지 항상 이겨야 한 다는 걸 강조하신다 (그럴수록)
11	FAM2F03	양육방식 I_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주 신다 (그렇지 않을수록)	26	PSY1C07	정서문제_신체증상: 숨쉬기 어 려울 때가 있다 (그렇지 않을 수록)
12	PSY2C07	자아인식_자아정체감: 나는 낯선 사람을 만나는 것을 꺼려한다 (그럴수록)	27	FAM2F01	양육방식 I_합리적 설명: 부모 님(보호자)의 결정을 무조건 따르게 하기 보다는 왜 그래 야 하는지 설명해 주신다 (그 렇지 않을수록)
13	FAM2D01	양육방식 I_과잉기대: 부모님(보 호자)의 기대가 항상 내 능력 이 상이어서 부담스럽다 (그럴수록)	28	PSY1A03	정서문제_주의집중: 오랫동안 집중해야 하는 과제는 하고 싶지 않다 (그렇지 않을수록)

14	INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다 (그럴수록)	29	PSY1D04	정서문제_사회적 위축: 수줍어한다 (그럴수록)
15	PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다 (그렇지 않을수록)	공 란		

모형 평가 적합 결과는 [그림4-1]과 같다. 정분류율, 민감도, 특이도, Kappa 통계량, AUC 값은 90.77%, 96.43%, 83.13%, 0.8082, 0.9782702로 성능이 아주 좋은 모형으로 판단된다.

```
> confusionMatrix(data= fit.pred, yy, positive = '1')
```

Confusion Matrix and Statistics

```

      Reference
Prediction 0  1
0      69  4
1      14 108

```

Accuracy : 0.9077

95% CI : (0.858, 0.9444)

No Information Rate : 0.5744

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.8082

Mcnemar's Test P-Value : 0.03389

Sensitivity : 0.9643

Specificity : 0.8313

Pos Pred Value : 0.8852

Neg Pred Value : 0.9452

Prevalence : 0.5744

Detection Rate : 0.5538

Detection Prevalence : 0.6256

Balanced Accuracy : 0.8978

'Positive' Class : 1

```
> as.numeric(performance(ROC.pred, "auc")@y.values)
```

[1] 0.9782702

[그림4-1] 보통수준을 빼고 재코딩한 모형 평가 R code

나. 266개 변수를 사용한 모형 적합 결과

성적 관련 변수인 국어, 수학, 영어 성적 주관적 평가와 전체 성적 만족도 변수를 삭제 후 재코딩한 반응변수로 모형을 적합한 결과는 <표 4-4>와 같다. 266개 변수 중 배경변인 1개, 개인발달 13개, 발달환경 10개로 24개의 변수가 선택되었다.

<표4-4> 모형적합 포함된 변수와 계수(266개 설명변수 사용)

변수	변수설명	코딩	계수
HAK2B	최종학력: 모친	Likert: 1(중졸 이하) ~ 5(대학원 졸)	0.076
INT2A01	학습습관_성취가치: 학교 공부는 나에게 중요한 의미를 지닌다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.137
INT2A04	학습습관_성취가치: 학교생활이 나의 미래에 상당한 역할을 할 것이다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.067
INT2B01	학습습관_숙달목적 지향성: 나는 실수를 하더라도 무엇인가를 배울 수 있는 어려운 내용을 좋아한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.165
INT2C01	학습습관_행동통제: 나는 공부가 지루하고 재미없더라도 끝까지 다한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.636
INT2C02	학습습관_행동통제: 나는 하던 공부를 끝낼 때까지 공부에 집중한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.072
INT2D04	학습습관_학업시간 관리: 나는 시험 전에 계획을 세우고 그에 따라 공부한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.093
PSY2A04	자아인식_자아존중감: 나는 남들만큼의 일은 할 수 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.098
PSY2B02	자아인식_자아탄력성: 나는 갑자기 놀라는 일을 당해도 금방 괜찮아지고 그것을 잘 이겨 낸다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.012
PSY2C06	자아인식_자아정체감: 남의 말을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.133
FUR1A01	진로계획_장래 희망 교육수준	Likert: 1(고등학교 졸업) ~ 5(대학원 박사학위 취득)	0.532
FUR2A03	대학진학이유: 부모님/가족이 원해	Likert: 1(매우 그렇다) ~	-0.057

	서	4(전혀 그렇지 않다)	
FUR2D03	직업선택 요인: 사회적 명예(위신)가 높다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.009
FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다	Likert: 1(매우 중요하다) ~ 4(전혀 중요하지 않다)	-0.076
FAM2F01	양육방식 I_합리적 설명: 부모님(보호자)의 결정을 무조건 따르게 하기 보다는 왜 그래야 하는지 설명해 주신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.030
FAM2F03	양육방식 I_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.109
FAM3A04	양육방식 II_방임: 내가 많이 아프면 적절한 치료를 받게 하신다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.009
EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-1.072
EDU2A04	학교적응_학습활동: 모르는 것이 있을 때 다른 사람(부모님이나 선생님 또는 친구들)에게 물어본다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	-0.043
EDU2D01	학교적응_교사관계: 선생님을 만나면 반갑게 인사한다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.054
COM1A01	지역사회 인식: 나는 우리 동네 사람들 대부분을 알고 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.076
COM2A02	공동체 의식: 공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있다	Likert: 1(매우 그렇다) ~ 4(전혀 그렇지 않다)	0.026
MDA2B08	휴대전화_이용빈도: 음악듣기	Likert: 1(자주 사용한다) ~ 4(전혀 사용하지 않는다)	0.018
ACT2A01	동아리활동_연간 참여경험 유무: 교내 공식 동아리활동	1(있다), 0(없다)	0.108

변수의 중요도 순으로 정리한 결과는 <표4-5>와 같다.

<표4-5> 중요도 순으로 변수 정리(266개 변수 사용)

순	변수	변수설명 (결과)	순	변수	변수설명 (결과)
1	EDU2A03	학교적응_학습활동: 수업 시간에 배운 내용을 잘 알고 있다 (그럴수록)	13	HAK2B	최종학력: 모친 (학력이 높을수록)
2	INT2C01	학습습관_행동통제: 나는 공부 가 지루하고 재미없더라도 끝 까지 다한다 (그럴수록)	14	INT2C02	학습습관_행동통제: 나는 하 던 공부를 끝낼 때까지 공 부에 집중한다 (그럴수록)
3	FUR1A01	진로계획_장래 희망 교육수준 (희망 교육수준이 높을수록)	15	INT2A04	학습습관_성취가치: 학교생 활이 나의 미래에 상당한 역 할을 할 것이다 (그럴수록)
4	INT2B01	학습습관_숙달목적 지향성: 나 는 실수를 하더라도 무엇인가 를 배울 수 있는 어려운 내용 을 좋아한다 (좋아할수록)	16	FUR2A03	대학진학이유: 부모님/가족 이 위해서 (그럴수록)
5	INT2A01	학습습관_성취가치: 학교 공부 는 나에게 중요한 의미를 지닌 다 (그럴수록)	17	EDU2D01	학교적응_교사관계: 선생님 을 만나면 반갑게 인사한다 (그렇지 않을수록)
6	PSY2C06	자아인식_자아정체감: 남의 말 을 잘 받아들이고 타인의 말과 행동에 영향 받기 쉽다 (그렇지 않을수록)	18	EDU2A04	학교적응_학습활동: 모르는 것이 있을 때 다른 사람(부 모님이나 선생님 또는 친구 들)에게 물어본다 (그럴수록)
7	FAM2F03	양육방식Ⅰ_합리적 설명: 내가 무리한 요구를 하면 그것이 왜 안 되는지 그 이유를 말씀해 주신다 (그렇지 않을수록)	19	FAM2F01	양육방식Ⅰ_합리적 설명: 보 모님(보호자)의 결정을 무조 건 따르게 하기 보다는 왜 그래야 하는지 설명해 주신 다 (그렇지 않을수록)
8	ACT2A01	동아리활동_연간 참여경험 유 무: 교내 공식 동아리활동 (있 을수록)	20	COM2A02	공동체의식: 공휴일에 쉬지 못하더라도 복지기관에서 자원봉사 활동을 할 수 있 다 (그렇지 않을수록)
9	PSY2A04	자아인식_자아존중감: 나는 남 들만큼의 일은 할 수 있다 (그 려ლ수록)	21	MDA2B08	휴대전화_이용빈도: 음악듣 기 (사용하지 않을수록)
10	INT2D04	학습습관_학업시간 관리: 나는 시험 전에 계획을 세우고 그에 따라 공부한다 (그럴수록)	22	PSY2B02	자아인식_자아탄력성: 나는 갑자기 놀라는 일을 당해도 금방 괜찮아지고 그것을 잘 이겨 낸다 (그렇지 않을수록)
11	COM1A01	지역사회 인식: 나는 우리 동네 사람들 대부분을 알고 있다 (그 려지 않을수록)	23	FAM3A04	양육방식Ⅱ_방임: 내가 많이 아프면 적절한 치료를 받게 하신다 (그렇지 않을수록)

12	FUR2D04	직업선택 요인: 국가/사회에 기여할 수 있다 (중요할수록)	24	FUR2D03	직업선택 요인: 사회적 명예(위신)가 높다 (중요할수록)
----	---------	----------------------------------	----	---------	---------------------------------

모형 평가 적합 결과는 [그림4-2]와 같다. 정분류율, 민감도, 특이도, Kappa 통계량, AUC 값은 81.54%, 87.50%, 73.49%, 0.6176, 0.8800559였다.

```
> confusionMatrix(data= fit.pred, yy, positive = '1')
```

Confusion Matrix and Statistics

Reference

Prediction 0 1

0 61 14

1 22 98

Accuracy : 0.8154

95% CI : (0.7537, 0.8672)

No Information Rate : 0.5744

P-Value [Acc > NIR] : 7.655e-13

Kappa : 0.6176

Mcnemar's Test P-Value : 0.2433

Sensitivity : 0.8750

Specificity : 0.7349

Pos Pred Value : 0.8167

Neg Pred Value : 0.8133

Prevalence : 0.5744

Detection Rate : 0.5026

Detection Prevalence : 0.6154

Balanced Accuracy : 0.8050

'Positive' Class : 1

```
> as.numeric(performance(ROC.pred, "auc")@y.values)
```

```
[1] 0.8800559
```

[그림4-2] 재코딩한 모형평가 R code (반응변수 266개 사용)