

별점회귀모형에서의 통계적 추론: 청소년의 스마트폰 의존 관련 변수를 중심으로*

노민정** · 유진은***

초 록

본 연구의 목적은 청소년의 스마트폰 의존 관련 변수를 탐색하여 스마트폰 의존을 예방하는 교육적 조치 및 대책 마련에 필요한 근거 자료를 제공하는 것이다. 스마트폰 의존 관련 변수를 분석하는 선행연구가 활발히 이루어져 왔으나, 기존 전통적 통계기법의 한계로 인하여 동시에 많은 수의 변수를 모형에 투입하여 다양한 변수들 간의 관계를 고려한 연구는 아직 실행된 바 없다. 따라서 본 연구는 스마트폰 의존과 관련된 수많은 변수를 하나의 모형에서 분석하고자 한국아동·청소년패널조사(2018) 2차년도 자료에 기계학습 기법인 별점회귀모형을 적용하였다. 뿐만 아니라, 별점회귀모형에서 선택된 변수에 대하여 통계적 추론을 실시하여 변수의 통계적 유의성까지 검정하였다. 자료를 무선편할한 후 모형 적합과 통계적 추론을 100회 반복한 결과, 통계적 유의성이 50회 이상 확인된 17개 변수를 최종 변수로 선택하였다. 최종 선택 변수는 부모님과 대화시간, 여가시간, 학업성취 만족도, 학업무기력, 주의집중, 창의적 성격, 그릇, 운동시간, 스마트폰 사용 목적별 빈도 등이었다. 이 17개 변수는 단순히 탐색적 기법만을 고려한 결과가 아니라, 전통적 기법에서 중시하는 통계적 추론까지 통과한 변수로서 의의가 있다. 선행연구에 비추어 최종 선택 변수와 청소년의 스마트폰 의존 간의 관계를 논의하였으며, 다양한 사회과학 분야 대용량 자료에의 적용과 모의실험 등 후속 연구 주제를 제안하였다. 본 연구는 사회과학 대용량자료 분석 시 별점회귀모형에서의 통계적 추론을 수행한 첫 번째 연구라는 방법론적 의의를 지닌다.

주제어: 스마트폰 의존, 기계학습, 별점회귀모형, 변수 선택 후 통계적 추론, 규제화

* 제9회 한국아동·청소년패널학술대회(2020.11.20.)에서 발표한 동제목 논문을 수정·보완한 것임.

이 논문은 한국청소년정책연구원에서 수집한 한국아동·청소년패널조사 데이터를 활용하였음.

** 한국교원대학교 박사

*** 한국교원대학교 교수, 교신저자, jeyoo@knue.ac.kr

I. 서론

스마트폰의 접근성과 휴대성은 우리 삶의 편의성을 획기적으로 향상시켰다. 우리는 스마트폰을 이용하여 언제 어디서든 원하는 정보를 얻고, 사람들과 교류하고, 즐거움을 얻는다(Bian & Leung, 2015; Lin et al., 2014). 스마트폰 사용자는 전세계적으로 급증하고 있으며, 스마트폰은 연령을 불문하고 많은 사람에게 없어서는 안 될 삶의 일부가 되었다(Samaha & Hawi, 2016). 특히 우리나라의 스마트폰 보급률은 이미 90%를 돌파하였으며(한국갤럽, 2019.8.29.), 세계 27개국을 대상으로 한 스마트폰 보급률 조사에서 1위를 차지하기도 하였다(KBS 뉴스, 2019.2.11.). 그러나 급격한 스마트폰 확산은 과도한 스마트폰 사용으로 인한 심각한 행위 중독 문제를 야기하였다(Choi et al., 2015; Kwon, Kim, Cho, & Yang, 2013).

스마트폰 의존은 스마트폰을 지나치게 많이 사용하고, 사용을 스스로 통제하지 못하며 일상생활에서 신체적·정신적 불편을 겪는 것을 의미한다(서보경, 차성수, 2020). 최근 수행된 여성가족부의 ‘청소년 인터넷·스마트폰 이용습관 진단조사’ 결과에 따르면 스마트폰 의존 현상은 청소년층에서 두드러지고 있다. 스마트폰 의존 고위험군 및 잠재적 위험군 비율은 성인층보다 청소년층에서 높았으며, 청소년층에서도 중학생이 유치원생, 초등학생, 고등학생, 대학생에 비하여 스마트폰 의존에 취약한 것으로 나타났다(과학기술정보통신부, 한국인터넷진흥원, 2019). 이러한 실태는 신체적·정서적·인지적 발달의 균형을 이루어야 하는 청소년의 발달적 특성을 고려할 때 심각한 사회적 문제라고 할 수 있다.

청소년의 스마트폰 의존에 대한 문제의식을 바탕으로 예방 및 개선책 마련을 위한 관련 변수 파악 연구가 진행되고 있다. 선행연구들은 주로 성별, 우울이나 자기통제력과 같은 개인 변수, 부모의 양육태도, 부모-자녀 간 의사소통, 학교생활적응, 또래애착과 같은 환경 변수들에 주목하여 이들 변수와 스마트폰 의존 간의 관계를 밝히는데 주력하였다. 대부분의 연구에서 남학생보다는 여학생의 스마트폰 의존도가 높게 나타났다(이규현, 권윤희, 2014; 이하나, 양승목, 2018; 조운주, 2019), 스마트폰 의존 청소년들은 일반군보다 우울하거나 부모의 양육태도가 부정적이고(노충래, 김소연, 2016; 문호경, 2020; 박소연, 유미숙, 2019), 또래애착 정도가 낮았다(김보경, 백용매, 허창구, 2016; 신선임, 2017). 그러나 일부 변수에서 선행연구 결과가 일관되지 않았는데,

이는 선행연구가 사용한 자료와 변수의 종류 및 조합이 상이하기 때문일 수 있다. 실제로 이수진, 문혁준(2013)은 스마트폰 중독에 대한 연구에서 일부 변수를 선택하는 단편적 접근보다 다양한 변수들 간의 관계를 밝힐 수 있도록 다양한 변수를 이용한 연구가 필요하다는 주장을 제기하였다.

이에 따라 본 연구는 별점회귀모형을 분석 기법으로 선택하여 청소년의 스마트폰 의존과 관련된 변수를 전체적인 관점에서 탐색하였다. 기계학습 기법인 별점회귀모형은 별점항을 이용하여 반응 변수와 관련이 적은 변수의 회귀계수를 축소하는 방식으로 변수를 선택하기 때문에 전통적 통계 기법과 달리 수백, 수천여 개의 변수를 동시에 분석할 수 있다(유진은, 2016). 이렇듯 별점회귀모형은 모형 투입 변수의 개수에 제한이 없으나 전통적 통계 기법에서와 같은 통계적 추론이 어렵다는 특징이 있다. 이는 별점회귀모형이 예측 성능을 향상시키기 위하여 분산을 감소시키는 대신 편향된 추정치를 산출하기 때문이다. 교육학 분야의 대용량자료에 별점회귀모형을 적용한 연구(노민정, 유진은, 2019a, 2020; 서은경, 김현철, 황수진, 이철원, 2018; 유진은, 2016; 유진은, 노민정, 2017)가 다수 존재함에도 통계적 추론을 수행한 연구가 아직 보고되지 않았다는 점에서 그러한 특징을 확인할 수 있다. 본 연구는 별점회귀모형의 통계적 추론을 위하여 최근 제안된 기법인 Post-Selection Inference(Lee, Sun, Sun, & Taylor, 2016; 이하 PSI)를 연구에 적용하였다. 정리하자면, 본 연구의 목적은 청소년의 스마트폰 의존과 관련된 변수들을 탐색하고 변수 간 관계를 확인하는 것이며, 이를 위하여 별점회귀모형으로 한국아동·청소년패널조사(Korean Children and Youth Panel Survey; 이하 KCYPS) 자료를 분석하고 별점회귀모형 결과에 대한 통계적 추론을 실시하였다.

II. 이론적 배경

1. 청소년과 스마트폰 의존

스마트폰 의존은 과거 ‘스마트폰 중독’이라는 용어로 사용되었으나, 스마트폰에 대한 의존 문제를 기존 중독 문제들과 구별하기 위하여 ‘스마트폰 의존’이라는 용어로 재명명되었다(한국정보화진흥원, 2017). ‘중독’은 정신병리학적 용어로 상담과 치료가

수반되어야 하는 질병인데 현대 사회의 일상에 보편화 되어 있는 기기인 스마트폰의 과도한 사용을 ‘중독’으로 취급하기에는 무리가 있다고 판단하였기 때문이다. 즉, 스마트폰 의존에는 스마트폰의 과도한 사용을 ‘중독’이 아닌 의존으로 재정의하여 치료가 아닌 선용을 계도하겠다는 의미가 부여되어 있다(서보경, 차성수, 2020).

스마트폰 의존은 과도한 스마트폰 이용으로 스마트폰에 대한 현저성(salience), 조절 실패(self-control failure)와 문제적 결과(serious consequences)를 경험하는 상태를 이룬다(과학기술정보통신부, 한국정보화진흥원, 2019). 현저성은 개인의 삶에서 스마트폰을 이용하는 생활 패턴이 다른 형태보다 두드러지고 가장 중요한 활동이 되는 것이다. 조절실패는 이용자가 스마트폰 이용을 본인의 자율적 의지로 통제하는 능력이 낮은 것을 의미하며, 문제적 결과는 스마트폰 사용으로 인해 신체적·심리적·사회적으로 부정적인 결과를 경험함에도 스마트폰을 지속적으로 사용하는 것을 의미한다. 이 중 청소년이 가장 많이 경험하는 스마트폰 의존 유형은 조절실패였다(과학기술정보통신부, 한국인터넷진흥원, 2019).

2. 스마트폰 의존 관련 변수

스마트폰 의존에 대한 연구는 초등학교부터 대학생, 성인에 이르기까지 다양한 연령대에서 수행되었다. 청소년층에는 초등학교, 중학교, 고등학교가 모두 포함되나, 청소년층은 학령에 따른 특성이 이질적인 편이므로 학교급을 구분하여 살펴볼 필요가 있다. 따라서 이 절에서 제시한 선행연구는 본 연구의 분석 대상인 중학생의 스마트폰 의존을 분석한 연구로 한정하였음을 밝힌다. 스마트폰 의존과 관련되어 연구된 변수는 KCYPs 2018의 분류를 따라 개인발달 변수와 발달환경 변수로 구분하여 제시하였다.

1) 개인발달 변수

선행연구들은 성별, 자기통제, 자기효능감, 자아존중감과 같은 심리 특성 변수나 우울, 공격성, 충동성과 같은 부정적 정서 변수, 스트레스 등의 개인발달 변수를 사용하였다. 가장 높은 빈도로 연구된 변수는 우울이었으며, 대부분의 연구에서 우울과 스마트폰 의존 간의 정적 상관관계가 확인되었다(경승구, 김진욱, 2019; 노충래, 김소연,

2016; 문호경, 2020; 이웅택, 오승근, 백진현, 이승봉, 2020; 정민, 2015; 조윤주, 2019). 그러나 이하나, 양승목(2018)의 연구에서는 중학생을 제외한 고등학생과 대학생에서만 우울과 스마트폰 의존 간의 통계적 유의성이 확인되었다. 이하나, 양승목(2018)의 연구에서 학령에 상관없이 통계적으로 유의한 변수는 공격성이었다. 노충래, 김소연(2016)도 공격성이 부모의 양육태도와 스마트폰 의존 간의 관계를 매개하고 있음을 보여주었다. 그러나 경승구, 김진옥(2019)의 연구에서는 공격성과 스마트폰 의존 간의 관계가 통계적으로 유의하지 않았다.

경승구와 김진옥(2019)의 연구에서 가장 큰 영향력을 나타낸 변수는 스트레스였다. 스트레스는 행위중독을 예측하는 위험요인 중 하나이며(Van Deursen, Bolle, Hegner, & Komdmers, 2015), 학교 및 가정 스트레스는 스마트폰 중독에 영향을 미친다. 스트레스는 자극 자체보다 개인의 대처 능력이 중요하며(이혜진, 2017), 이러한 관점에서 자기통제력, 자기효능감, 자아존중감과 같은 심리 변수가 스마트폰 의존과 함께 연구되었다. 대부분의 선행연구에서 자기통제력(김상립, 2018; 이수진, 문혁준, 2013; 정민, 2015), 자기효능감(이규현, 권윤희, 2014; 이윤미, 권윤희, 채규만, 2019), 자아존중감(이하나, 양승목, 2018; 문호경, 2020)과 스마트폰 의존 간의 부적 상관관계가 확인되었다.

그 외 개인발달 변수에는 성별, 성적, 여가 만족, 체육 활동이 있었다. 대체로 여학생이 남학생보다 스마트폰 의존도가 높게 나타났으나(경승구, 김진옥, 2019; 조윤주, 2019), 성별에 따른 스마트폰 의존도 차이가 통계적으로 유의하지 않은 연구도 있었다(이웅택 외, 2020). 성적과 스마트폰 의존에 대한 선행연구 중에는 부적 상관을 보고한 연구(이규현, 권윤희, 2014)와 통계적으로 유의하지 않음을 보고한 연구(이웅택 외, 2020)가 있었다. 여가만족, 운동 및 체육 활동에서는 여가 활동에 만족할수록(정인경, 김정현, 2015), 체육 활동에 적극적일수록 스마트폰 의존도가 낮게 나타났다(조건상, 이영익, 2018).

2) 발달환경 변수

많은 연구자들이 주목한 발달환경 변수는 부모 양육태도였으며, 대부분의 연구에서 부모의 양육태도가 부정적일수록 스마트폰 의존도가 높게 나타났다(노충래, 김소연, 2016; 문호경, 2020; 박소연, 유미숙, 2019; 이규현, 권윤희, 2014; 조윤주, 2019). 그

러나 부모의 양육태도와 스마트폰 의존 간의 관계가 통계적으로 유의하지 않은 연구(이하나, 양승목, 2018)도 존재하였다.

부모 양육태도 외에 부모 또는 가족과 관련된 선행연구 변수에는 부모 애착과 의사소통이 있었다. 대부분의 연구에서 부모와의 애착 관계는 스마트폰 의존에 영향을 미치는 것으로 확인되었다(오현희, 김현진, 2014; 이윤미 외, 2019). 그러나 부모 애착과 스마트폰 의존 간에는 직접적 관계가 없다고 설정하거나(정규석, 2017), 통계적 유의성이 존재하지 않는다고 보고한 연구도 있었다(이용택 외, 2020). 부모와의 애착 문제는 대인관계나 자녀 간 의사소통에도 전이되어 스마트폰 의존에 영향을 미칠 수 있다(김보경 외, 2016). 부모 애착과 더불어 부모 또는 가족 의사소통과 스마트폰 의존 간의 관계를 분석한 연구들은 개방적 의사소통이 스마트폰 의존도를 감소시킨다고 보고하였다(고은혜, 김근영, 2017; 김상림, 2018; 오현희, 김현진, 2014; 이수진, 문혁준, 2013; 정민, 2015).

학교생활만족도와 또래애착은 부모 양육 태도와 더불어 높은 빈도로 연구된 발달환경 변수였다. 대부분의 연구에서 학교생활만족도가 높을수록 스마트폰 의존도가 낮게 나타났다(이규현, 권윤희, 2014; 이수진, 문혁준, 2013; 정민, 2015; 조윤주, 2019; 주석진, 조성심, 2015). 김상림(2018)에서는 학교생활만족도가 통계적으로 유의하지 않았다. 마찬가지로, 또래애착의 경우에도 또래애착이 높아 사회적 위축이나 소외감을 경험하지 않을수록 스마트폰 의존도가 낮은 것으로 보고한 연구(경승구, 김진욱, 2019; 고은혜, 김근영, 2017; 오현희, 김현진, 2014; 정민, 2015; 주석진, 조성심, 2015)가 있는 반면, 통계적으로 유의하지 않은 연구(이용택 외, 2020)도 있었다.

스마트폰 의존과 관련된 변수에는 스마트폰 사용 유형(배성만, 2018; 이용택 외, 2020; 이하나, 양승목, 2018), 스마트폰 사용 시간(주석진, 조성심, 2015), 게임중독(배성만, 2018), SNS중독(배성만, 2018)이 있었다. 성별과 상관없이 게임중독과 SNS중독 수준이 높을수록 스마트폰 의존 수준 또한 높게 나타났다. 스마트폰 사용 용도는 성별에 따라 달랐으며, 여학생은 사진 촬영 및 동영상, 남학생은 게임을 많이 하는 것으로 나타났다(이하나, 양승목, 2018).

이상과 같이 선행연구는 여러 가지 변수를 사용하였으며, 다양한 변수 조합으로 변수와 스마트폰 의존 간의 관계를 분석하였다. 이로 인하여 변수의 직접적, 매개 및 조절 관계 설정이 연구마다 상이하였으며, 분석 결과 통계적 유의성이 일치하지 않는

경우가 드물지 않았다. 본 연구는 이와 같은 선행연구 분석을 바탕으로 동시에 많은 변수를 분석할 수 있는 별점회귀모형을 이용하여 스마트폰 의존 관련 변수를 탐색하였다.

3. 별점회귀모형: LASSO

기계학습 기법인 별점회귀모형은 변수 선택과 동시에 계수를 추정할 수 있는 방법으로 특히 빅데이터 또는 고차원 자료 분석에서 각광받고 있다. 주로 보건 및 의학(Futoma, Morris, & Lucas, 2015; Kim, Banerjee, & Park, 2016), 유전학(Ghosh & Chinnaiyan, 2005; Lu et al., 2017; Ma, Song, & Huang, 2007), 생명정보학(Ma & Huang, 2008; Wu et al., 2009), 공학(Fan, Feng, & Wu, 2009; Rida, Jiang, & Marcialis, 2015; Zou, Jiang, & Tsung, 2011), 경제학(Panagiotidis, Stengos, & Vravosinos, 2018; Wang, Xu, & Zhou, 2015) 등의 분야에서 규제화(regularization) 기법으로 사용되고 있다. 사례 수에 비하여 상대적으로 변수 수가 많은 자료를 기존 회귀모형의 최소제곱추정법을 사용하여 분석할 때 비수렴 또는 과적합으로 인한 예측력 감소 등의 문제가 발생할 수 있다(박창이, 김용대, 김진석, 송종우, 최호식, 2011; 유진은, 2016; Tibshirani, 1996). 반면, 규제화 기법인 별점회귀모형은 변수 수가 사례 수보다 많은 자료에도 변수 선택 기법으로 활용될 수 있으며, 이를테면 교차타당화를 통하여 예측 오차를 줄일 수 있다.

대표적인 별점회귀모형인 LASSO(Tibshirani, 1996)는 별점잔차제곱합을 최소로 하는 회귀계수를 추정한다. 식(1)의 첫 번째 항은 최소제곱추정량이며, 두 번째 항이 별점항이다. 별점항의 조율모수 λ 가 회귀계수의 축소 정도를 조절한다. 즉, 조율모수 λ 가 커지면 회귀계수가 많이 축소되어 0에 가까워지며, 회귀계수가 0인 변수는 모형에서 배제된다.

$$\hat{\beta}^L = \arg \min_{\beta} \| Y - \sum_{j=1}^p X_j \beta_j \|_2^2 + \lambda \sum_{j=1}^p |\beta_j|. \quad \text{식(1)}$$

4. 벌점회귀모형에서의 통계적 추론: PSI

벌점회귀모형은 예측오차를 최소화하는 모형을 구축하기 위하여 편향된 추정치를 도출하므로(유진은, 2016) 벌점회귀모형에서의 통계적 추론은 개념적으로나 분석적으로나 어려운 과제였다(Hastie, Tibshirani, & Wainwright, 2015). 벌점회귀모형에서의 통계적 추론을 수행하기 위한 여러 연구가 진행되었으며, 최근 PSI(post-selection inference; Lee et al., 2016)와 debiasing(Javanmard & Montanari, 2014)과 같은 기법들이 제안되었다.

PSI와 debiasing의 특징을 비교하면 다음과 같다. 먼저 debiasing은 희소성(sparsity) 가정 하에 규제화된 추정치를 얻기 위하여 전체 모형을 활용한다. 따라서 하위 모형(submodel)을 선택할 필요가 없으며 추정 결과로 도출된 신뢰구간이 지나치게 넓지 않다는 장점이 있다(Hastie et al., 2015). 그러나 전체 모형을 활용하여 추론하므로 다중검정(multiple testing) 시 상대적으로 보수적인 검정 결과를 도출한다(Hastie et al., 2015; Lee et al., 2016). 반면, PSI는 LASSO가 0이 아닌 모든 참변수를 찾았다고 가정하고 LASSO 결과로 선택된 변수에 대하여 조건부 추론을 실시한다. 따라서 debiasing의 문제인 다중검정오류 확률이 줄어든다는 장점이 있다. 단, PSI에서는 ‘LASSO가 0이 아닌 모든 참변수를 찾았다’는 가정을 충족해야 하는데, LASSO와 같은 불록벌점합수를 교차타당화와 함께 쓸 때 참변수를 모두 선택한다는 모의실험연구가 다수 있다(Bühlmann & Mandozzi, 2014; Fan & Lv, 2008; Yoo & Rho, in press). 특히 소규모자료 분석보다 대용량자료 분석 시 그 가정을 충족할 확률이 더 높아진다고 할 수 있다. 따라서 사회과학 대용량자료를 분석하는 본 연구는 PSI를 선택하였다.

PSI 단계를 설명하면 다음과 같다. 먼저 LASSO를 수행하여 모형을 구축한다(식 1). LASSO에서 선택된 변수가 활성 세트($\hat{M}$)에 들어가고, 선택되지 않은 변수는 비활성 세트($-\hat{M}$)에 들어간다(식 2). LASSO로 구축된 모형은 Karush-Kuhn-Tucker(KKT) 조건을 충족하며, 식 3과 같이 정리된다. 즉, 활성 세트에 포함되는 변수 w 의 부호가 s 가 되고, 활성 세트에 포함되지 않는 변수 벡터 u 의 노음은 1보다 작게 된다. 식 3을 만족하는 w 와 u 는 식 4 우변의 부등식과 같다. 식 4의 M 과 s 는 활성화 제약조건인 A_1 과 b_1 , 비활성 제약조건인 A_0 와 b_0 로 나뉘며, 이는 LASSO 모형에서 선택된 변수 집합 $\{\hat{M} = M, \hat{s} = s\}$ 이 다면체(polyhedron) 형태를 띠는 식 5로 유도된다.

그 결과, PSI의 조건부 분포(식 6)는 절단된 정규분포(truncated normal distribution)를 따르게 되며 이 분포를 바탕으로 통계적 추론이 이루어진다. PSI에 대한 더 자세한 설명은 Lee et al.(2016)을 참고하기 바란다.

$$\hat{M} = \{j : \hat{\beta}_j \neq 0\} \quad \text{식(2)}$$

$$\begin{aligned} X_M^T(X_M \mathbf{w} - y) + \lambda \mathbf{s} &= 0, \text{ sign}(\mathbf{w}) = \mathbf{s}, \\ X_{-M}^T(X_M \mathbf{w} - y) + \lambda \mathbf{u} &= 0, \|\mathbf{u}\|_\infty < 1. \end{aligned} \quad \text{식(3)}$$

$$\{\text{sign}(\mathbf{w}) = \mathbf{s}, \|\mathbf{u}\|_\infty < 1\} = \left\{ \begin{pmatrix} A_0(M, \mathbf{s}) \\ A_1(M, \mathbf{s}) \end{pmatrix} y < \begin{pmatrix} b_0(M, \mathbf{s}) \\ b_1(M, \mathbf{s}) \end{pmatrix} \right\} \quad \text{식(4)}$$

$$\{\hat{M} = M, \hat{\mathbf{s}} = \mathbf{s}\} = \{A(M, \mathbf{s})y \leq b(M, \mathbf{s})\} \quad \text{식(5)}$$

$$\eta_M^T y \mid \{\hat{M} = M\} \quad \text{식(6)}$$

III. 연구 방법

1. 분석 자료

1) 자료 정리

한국아동·청소년패널조사 2018은 2018년 기준 전국의 초등학교 4학년, 중학교 1학년 재학생을 다단계층화집락표집 방식으로 추출하여 추적 조사하는 종단 연구 자료이다. 본 연구는 중학생의 스마트폰 의존 관련 변수를 탐색하기 위하여 KCYPS 2018 중1패널 2차년도 자료를 사용하였다. 해당 자료는 2019년 조사 당시 중학교 2학년이었던 학생 2,438명(남학생 1,318명, 여학생 1,120명)의 개인 발달 및 발달환경 378개 문항에 대한 응답 값이다.

자료는 다음과 같이 정리하였다. 성별(YGENDER), 선호 연예인/운동선수 유무(YACT3A00)

와 같은 이분형 변수의 응답 값은 0 또는 1로 재코딩하였다. 하위범주별 응답 비율이 매우 불균형한 스마트폰 사용(YMDA1A00) 변수는 소유자와 상관없이 스마트폰을 사용하고 있으면 '1', 스마트폰을 사용하지 않으면 '0'으로 변환하였다. 범주와 서열이 혼합되어 있는 변수는 특정 값을 결측값으로 변환하였다. 예를 들면, 전교과 성취도 주관적 평가(YINT1A00), 학업성취 만족도(YINT1B00), 학교생활만족도(YEDU1A00)의 '6: 잘 모르겠음'과 장래희망 교육수준(YFUR1A00)의 '5: 잘 모르겠음', 진로관련 대화 빈도(YFUR2A) 변수의 '8: 해당 없음'을 결측값으로 변환하였다. 또한 형제자매 중 본인 순서(YFAM3A00) 변수의 결측값은 0으로 변환하였다.

응답값 및 결측값 변환 후 총 99개의 변수를 삭제하였다. 삭제한 변수는 학교 코드, 가중치와 같은 관리 변수, 분석에 불필요하다고 판단된 '시간(분)' 변수, 30개의 창의적 성격 문항에서 생성되어 다중공선성의 우려가 있는 창의적 성격 점수(YPSY6Asc), 분산이 0에 가까운 변수였다. 또한 PSI에서는 범주형 변수를 처리할 수 없어 학교 및 주택 소재지(ARA1, ARA2) 변수를 삭제하였다. 변수 삭제 후 남아 있는 결측값은 k-NN 대체법을 이용하여 대체하였다. K-NN 대체법은 비모수적 방법으로 알고리즘에 대한 이해 및 적용이 쉬우면서 대체 결과 또한 우수한 기법으로 알려져 있다(노민정, 유진은, 2019b; 박소현, 방성완, 전명식, 2011; Jönsson & Wohlin, 2004). K는 Jönsson & Wohlin(2004)을 따라 완전관측 사례수의 제곱근 값인 39를 사용하였다.

표 1

KCYPs 2018 아동·청소년 조사 영역(중등)

조사영역			조사항목
대영역	중영역	소영역	
개인 발달	생활시간	하루일과	수면시간(취침/기상)과 질, 부모님 대화시간, 학습시간, 여가시간
	지적발달	학업성취	전 교과 성취도의 주관적 평가, 학업성취 만족도
		학업태도	학업 열의, 학업 무기력
	진로	진로계획	장래희망 교육수준, 장래직업 결정여부
		진로관	진로 관련 대화상대/빈도
	사회/정서/역량발달	만족도	삶의 만족도
		행복	행복감
		자아인식	자아존중감

조사영역			조사항목
대영역	중영역	소영역	
		정서문제	주의집중, 공격성, 신체증상, 사회적 위축, 우울
		협동의식	협동심
		창의성	창의적 성격
		끈기	그릿(Grit)
비행		비행1	현실비행(비행경험 유무 및 빈도)
		비행2	사이버비행(비행경험 유무 및 빈도)
신체발달		건강	건강상태 평가, 운동시간, 아침식사 횟수
		체격	키, 몸무게
매체		스마트폰	사용여부, 사용 목적별 빈도, 의존도
발달 환경	활동/문화 환경	청소년활동	연간 참여 횟수 및 만족도
		동아리활동	연간 참여 횟수 및 만족도
	학교	팬덤활동	선호 연예인/운동선수 유무, 팬덤 활동 빈도
		학교생활	학교생활 만족도
가정		친구	친구관계
		교사	교사관계
		부모관계	부모님과 함께 보내는 시간
		양육태도	따스함, 거부, 자율성지지, 강요, 구조제공, 비밀관성

2) 설명변수 및 반응변수

자료 정리 과정을 거쳐 분석에 사용한 최종 자료의 학생 수는 2,438명, 설명변수는 264개였다. 설명변수는 개인발달 변수와 발달환경 변수로 구분되며 자세한 내용은 표 1과 같다(하형석 외, 2018). 구체적으로 살펴보면, 개인발달 변수에는 생활시간, 지적 발달, 진로, 사회·정서·역량 발달, 비행, 신체발달이 있으며, 발달환경 변수에는 매체, 활동·문화환경, 학교, 가정 관련 변수가 있다.

반응변수인 스마트폰 의존도는 KCYPS 2018의 스마트폰 의존도 15개 문항의 평균 값이며, 각 문항은 ‘1; 전혀 그렇지 않다’에서 ‘4; 매우 그렇다’까지 4점 리커트형 척도이다. 15개 문항 중 5, 10, 15번째 문항 3개를 역변환하였으며, 점수가 높을수록 스마트폰 의존도가 높음을 의미한다. 반응변수의 최솟값과 최댓값은 1과 3.87, 평균은 2.14, 표준편차는 0.47, 크론바흐 알파값은 0.87이었다.

2. 분석 방법

자료 분석 과정은 다음과 같다. 먼저 10-fold 교차검증(cross-validation; CV)을 실시한 후, 1-표준편차를 기준으로 조율모수 λ 를 선택하여 LASSO 모형을 구축하고(Hastie et al., 2015), 구축된 LASSO 모형에 대하여 PSI를 실행하여 변수들의 통계적 유의성을 확인하였다. CV는 자료를 동일한 크기로 분할한 후, 분할한 자료를 다시 훈련 및 타당화 자료로 나누어 모형을 구축하고 타당화 오차를 산출하는 것을 반복하며 모형을 평가하는 방법이다(유진은, 2016). 또한, 본 연구에서는 모형 구축과 PSI를 100번 반복한 후, Meinshausen & Bühlmann(2010)과 Shevade & Keerthi(2003)를 참고하여 100번의 반복에서 각 설명변수의 선택 및 통계적 유의성 확인 횟수를 계산하였다. 이 값을 선택계수(selection counts; 유진은, 김형관, 노민정, 2020; Yoo & Rho, 2020)라 하고, 1, 25, 50, 75, 100번을 기준으로 정리하여 제시하였다. 선택계수를 사용한 이유는 CV 활용 시 자료 분할에 따라 조율모수 λ 가 달라지는 문제 및 일반적으로 모형 선택 일치성이 충족되지 않는 것으로 알려져 있는 LASSO의 특성(Huang, Ma, & Zhang, 2008; Zhao & Yu, 2006)을 고려하기 위함이다.

자료 정리 및 통계 분석 시 사용한 프로그램은 R이다. 결측값 대체에는 VIM(Kowarik, & Templ, 2016), 벌점회귀모형 구축과 PSI에는 각각 glmnet(Friedman, Hastie, & Tibshirani, 2010)과 selectiveInference(Tibshirani et al., 2019) 패키지를 사용하였다.

IV. 연구 결과

1. LASSO와 PSI 결과

최종 분석 자료에서 100개의 LASSO 모형을 생성하였으며, 각 LASSO 모형에서 선택된 변수를 대상으로 PSI를 실행하였다. 표 2와 표 3은 구축된 100개의 LASSO 모형의 크기와 선택계수를 정리한 것이다. 표 2와 같이 LASSO 모형 크기의 최솟값은 58, 최댓값은 78, 평균(표준편차)은 67.48(5.33)이었다. 표 3과 같이 100개의 LASSO 모형에서 1번이라도 선택된 변수는 78개였으며, 25번 이상 선택된 변수는 70개, 50번 또

는 75번 이상 선택된 변수는 64개, 100번 모두 선택된 변수는 58개였다.

표 2
반복 구축된 LASSO 모형 크기

	최솟값	1Q	2Q	평균 (표준편차)	3Q	최댓값
변수 개수	58	64	64	67.48 (5.33)	70	78

표 3
LASSO 모형의 선택계수

반복 선택 횟수	≥1	≥25	≥50	≥75	=100
변수 개수	78	70	64	64	58

표 4는 생성된 LASSO 모형에서 선택된 변수들을 대상으로 PSI를 실행한 후 확인한 통계적 유의성의 기술통계를 제시한 것이다. LASSO 모형에서 선택된 변수 중 통계적 유의성이 확인된 변수 개수의 최솟값은 12개, 최댓값은 20개, 평균(표준편차)은 15.80(1.63)개였다. 표 5는 100번 반복 수행된 PSI 결과에서 선택계수를 산출한 것이다. 통계적 유의성이 확인된 횟수가 1번 이상인 변수는 26개, 25번 이상인 변수는 21개, 50번 이상인 변수는 17개, 75번 이상인 변수는 10개, 100번인 변수는 5개였다.

표 4
반복 수행된 PSI 결과

	최솟값	1Q	2Q	평균 (표준편차)	3Q	최댓값
변수 개수	12	14	16	15.80 (1.63)	17	20

표 5
PSI 선택계수

통계적 유의성 확인 횟수	≥1	≥25	≥50	≥75	=100
변수 개수	26	21	17	10	5

2. 최종 선택 변수

본 연구는 100개의 LASSO 모형에서 100번 모두 선택된 변수 중 통계적 유의성이 50회 이상 확인된 변수 17개를 최종 변수로 선택하였다. 이 중 14개는 생활시간, 지적발달, 사회/정서/역량발달, 신체발달에 대한 개인발달 영역의 변수였으며, 나머지 3개 변수는 발달환경 영역의 매체 변수였다. 표 6은 17개 최종 변수의 변수명, 변수 내용, 응답 범주, 회귀계수의 평균(표준편차), PSI 결과를 제시한 것이다.

표 6

최종 선택 변수

영역	순	변수명	변수 내용	범주	회귀계수 평균(표준편차)	PSI
개인 발달	1	YTIM1D01	[부모님 대화시간-평일_부모님 대화시간	A	-0.02(0.0008)	100
	2	YTIM1N01	[여가시간-평일_이외에 친구들과 노는 시간		-0.01(0.0024)	68
	3	YTIM1K01	[여가시간-평일_스마트폰을 가지고 노는 시간		0.06(0.0016)	80
	4	YTIM1K02	[여가시간-주말_스마트폰을 가지고 노는 시간		0.03(0.0005)	65
	5	YINT1B00	[학업성취 만족도-지난 학기 학교성적 만족 수준	B	-0.02(0.0012)	93
	6	YINT2B01	[학업 무기력-실력 차이를 내 힘으로 극복할 수는 없다고 생각한다	C	0.02(0.0001)	100
	7	YINT2B15	[학업 무기력-수업에 적극적으로 참여하지 않는다		0.02(0.0005)	84
	8	YPSY4A06	[주의집중-공부할 때 차분하게 앉아 있기 힘들다		0.03(0.0007)	71
	9	YPSY4A07	[주의집중-글자를 잘 빠뜨리고 쓰는 편이다	D	0.02(0.0002)	53
	10	YPSY6A30	[창의적 성격-의심이 많은		-0.02(0.0021)	100
	11	YPSY7A01	[그릿(Grit)-나는 무엇을 하다가 다른 생각이 나면 집중하기가 어렵다	C	0.02(0.0002)	73
	12	YPSY7A03	[그릿(Grit)-나는 어떤 문제에 잠깐 집중하다가 곧 흥미를 잃은 적이 있다		0.03(0.0003)	100
	13	YPSY7A07	[그릿(Grit)-나는 시작하면 무조건 끝낸다		-0.03(0.0002)	82
	14	YPHY1B00	[운동시간-지난 일주일 간 땀이 날 정도로 운동한 시간	E	-0.02(0.0002)	80

영역	순	변수명	변수 내용	범주	회귀계수 평균(표준편차)	PSI
발달 환경	15	YMDA1B05	[사용 목적별 빈도]-SNS 이용	F	0.02(0.0008)	100
	16	YMDA1B06	[사용 목적별 빈도]-게임		0.03(0.0012)	65
	17	YMDA1B07	[사용 목적별 빈도]-사진/동영상 촬영		0.03(0.0009)	70

A: 1, 전혀 안함; 2, 30분 미만; 3, 30분~1시간 미만; 4, 1시간~2시간 미만; 5, 2시간~3시간 미만; 6, 3시간~4시간 미만; 7, 4시간 이상

B: 1, 매우 불만족; 2, 불만족; 3, 보통; 4, 만족; 5, 매우 만족

C: 1, 전혀 그렇지 않다; 2, 그렇지 않은 편이다; 3, 그런 편이다; 4, 매우 그렇다

D: 0, 선택 안 함; 1, 선택함

E: 1, 없다; 2, 1시간; 3, 2시간; 4, 3시간; 5, 4시간 이상

F: 1, 전혀 사용하지 않는다; 2, 거의 사용하지 않는다; 3, 가끔 사용한다; 4, 자주 사용한다

생활시간 영역에서는 하루일과 중 부모님과 대화시간, 여가시간 변수가 3개 선택되었다. 평일에 부모님과 대화하는 시간이 많을수록(YTIM1D01), 평일에 친구들과 노는 시간이 많을수록(YTIM1N01) 스마트폰 의존도가 낮게 나타났다. 친구들과 노는 시간은 부모님과 대화, 공부 및 독서, 운동, 스마트폰이나 컴퓨터를 하는 시간, TV 시청 시간을 제외한 시간이다. 그리고 평일, 주말에 스마트폰을 가지고 노는 시간(YTIM1K01, YTIM1K02)이 많을수록 스마트폰 의존도가 높게 나타났다.

지적발달 영역의 학업성취 영역에서 학업성취 만족도, 학업 태도 영역에서 학업 무기력 변수가 선택되었다. 지난 학기 학교 성적에 대한 만족도가 낮을수록(YINT1B00), 실력 차이를 내 힘으로 극복할 수 없다고 생각할수록(YINT2B01), 수업에 적극적으로 참여하지 않을수록(YINT2B15) 스마트폰 의존도가 높게 나타났다.

사회/정서/역량발달 영역에서는 정서문제-주의집중, 창의성, 끈기 변수가 선택되었다. 공부할 때 차분하게 앉아 있기 힘들다고 응답하였을수록(YPSY4A06), 글자를 잘 빠뜨리고 쓰는 편이라고 응답하였을수록(YPSY4A07) 스마트폰 의존도가 높게 나타났다. 창의성에서는 창의적 성격 변수 중 의심이 많다고 응답하였을수록(YPSY6A30) 스마트폰 의존도가 낮게 나타났다. 끈기에서는 그릇 변수가 3개가 선택되었으며, 무엇을 하다가 다른 생각이 나면 집중하기 어렵다고 응답하였을수록(YPSY7A01), 어떤 문제에 잠깐 집중하다가 곧 흥미를 잃는다고 응답하였을수록(YPSY7A03) 스마트폰 의존도가 높게, 어떤 일을 시작하면 무조건 끝낸다고 응답하였을수록(YPSY7A07) 스마트폰 의존도가 낮게 나타났다.

신체발달 영역에서는 운동시간 변수가 선택되었다. 지난 일주일 간 땀이 날 정도로 운동한 시간이 많을수록(YPHY1B00) 스마트폰 의존도가 낮게 나타났다. 마지막으로 발달환경에서는 매체 관련 변수가 3개 선택되었다. 스마트폰 사용 목적별 빈도에서 SNS, 게임, 사진/동영상 촬영을 많이 할수록(YMDA1B05, YMDA1B06, YMDA1B07) 스마트폰 의존도가 높게 나타났다.

V. 논의 및 제언

1. 최종 선택 변수에 대한 논의

본 연구는 KCYPS 2018 2차년도 자료에 별점회귀모형을 적용하여 중학생의 스마트폰 의존도와 관련된 변수를 탐색하고 통계적 추론을 진행하였다. 100번의 반복에서 모두 선택된 변수 중 50번 이상 통계적 유의성이 확인된 총 17개 변수를 선행연구와 비교하여 논의하면 다음과 같다.

먼저 선행연구에 사용된 변수이면서 본 연구에서도 선택된 변수는 스마트폰 사용목적별 빈도 및 사용 시간, 학업 무기력, 주의집중이다. 스마트폰 사용 시간이 많을수록 스마트폰 의존 경향이 높게 나타나며(류세인, 조인숙, 2015), 이는 본 연구 결과에서도 확인되었다. 스마트폰 사용 목적의 경우에는 SNS나 게임, 사진 및 동영상 촬영에 많이 사용할수록 스마트폰 의존도가 높은 것으로 나타났다. 본 연구에서 성별과 스마트폰 의존 간의 통계적 유의성은 확인되지 않았으나, 성별에 따라 스마트폰 사용 목적이 다른 경향을 보인 연구(이하나, 양승목, 2018)를 염두에 두고, 스마트폰 사용 목적에 따른 차별적 접근 대책을 마련할 필요가 있음을 제기한다.

학업 무기력과 주의집중도 스마트폰 의존과 관계가 있는 것으로 알려져 있다(이은경, 안지영, 김지신, 2020; 허균, 2013). 학업 무기력은 학업 상황에서 겪는 학습된 무기력을 특정한 것이며(박병기, 노시언, 김진아, 황진숙, 2015), 학습된 무기력은 학업 성취에 영향을 미치는 것으로 알려져 있다(김아영, 주지은, 1999). 성적은 이규현, 권윤희(2014)의 연구에서 스마트폰 의존과 부적 관계가 있는 것으로 나타났으며, 본 연

구에서는 학업성취도에 대한 만족도가 선택되어 성적과 스마트폰 의존 간의 부적 관계를 보고한 선행연구와 관련 지을 수 있었다. 선행연구와 구별되는 점은 KCYPS 2018에 성적에 대한 주관적 평가 문항이 존재함에도 해당 문항이 아닌 성적에 대한 만족도가 선택되었다는 것이다. 이는 성적 그 자체보다 성적에 대한 학생의 인식이 중요함을 의미한다고 볼 수 있으며, 성적에 대한 만족은 학업에 대한 스트레스가 낮은 상황이라고 할 수 있기 때문에 이는 학업 스트레스가 스마트폰 의존과 정적 상관을 맺고 있다고 보고한 선행연구(정인경, 김정현, 2017)와 일맥상통한다.

다음은 선행연구에서 사용되었으나 본 연구에서는 선택되지 않은 변수이다. 부모 양육태도와 우울, 또래애착 변수는 선행연구에서 매우 높은 빈도로 사용되었으나 본 연구에서 선택되지 않았다. KCYPS 2018은 학생이 인식하는 부모의 양육태도를 애착, 거부, 자율성지지, 거부 등으로 조사하고 있으나 본 연구에서 해당 문항들은 선택되지 않았다. 본 연구에서 선택된 부모 관련 변수는 부모와의 대화시간이었으며, 부모와 자녀 간의 대화가 개방적일 경우 폐쇄적일 때보다 대화 시간이 많을 것이므로 부모와 자녀 간의 의사소통 유형을 사용한 선행연구(노충래, 김소연, 2016)와 같은 맥락에 있다고 볼 수 있다.

이미 여러 선행연구에서 우울과 스마트폰 의존의 상관관계가 연구된 바 있음에도 PSI 결과에서 우울은 통계적으로 유의한 변수로 선택되지 않았다. 이에 대한 몇 가지 이유를 고찰하면 다음과 같다. 먼저 선행연구와 본 연구의 ‘우울’ 구인이 기존 연구에서 사용된 ‘우울’ 구인과 이름은 동일하나 그 구조는 상이할 수 있다. 본 연구가 개별 문항을 각각의 변수로 취급하였는데, 기존 연구는 문항 묶음의 평균값을 우울 구인으로 사용했기 때문에 같은 이름의 구인이라도 그 구인을 측정하는 문항 묶음을 구성하는 개별 문항들이 다를 수 있는 것이다(노민정, 유진은, 2019a, 2020). 또 다른 이유로 분석 모형에 투입된 변수 세트의 차이를 들 수 있다. 선행연구에서 우울과 함께 사용된 설명변수는 주로 부모양육태도와 학교생활만족도였는데, 본 연구에서는 수백 개에 달하는 개인발달 및 환경발달 변수를 한 모형에 투입하였고, 이로 인하여 변수 간 관계가 달라졌을 수 있다. ‘우울’ 변수에 대한 본 연구 결과는 가능한 많은 변수를 하나의 모형에 투입하여 전체적인 관점에서 변수 간 관계를 파악하려는 본 연구 목적과 닿아 있다. 또한, 이미 많은 연구가 수행되어 일견 새로운 연구가 불필요해 보이는 주제에도 별점회귀모형을 적용하는 것의 의의를 보여준다고 할 수 있다.

부모와 관련된 변수 중 선택된 것은 부모와의 평일 대화시간이었으며, 부정적 정서에서는 주의집중이, 사회적 관계에서는 또래애착이 아닌 평일 이외에 친구들과 노는 시간이 선택되었다. 친구들과 어울리는 시간이 많을수록 스마트폰 및 SNS 사용이 높아지므로(염동문, 2017) 친구들과 어울리는 시간은 감소시켜야 한다고 주장하는 연구도 있었으나, 본 연구에서 선택된 ‘친구들과 노는 시간’은 학습, 운동 및 신체활동, 스마트폰 및 컴퓨터 사용 시간을 제외하고 친구들과 어울리는 시간을 의미한다. 따라서 중학생들이 친구들과 어울리는 시간을 온/오프라인을 구분하여 파악할 필요가 있겠다.

체육 활동 정도와 스마트폰 의존 간의 부적 관계는 양영미, 이선영(2019)의 연구에서 확인되었다. 본 연구 결과가 선행연구와 구별되는 점은 단순히 체육 활동 유무가 아니라 어떠한 운동을 했느냐에 대한 정보가 제시되었다는 점이다. KCYPS 2018에는 학교 수업 시간 외 운동이나 체육 활동 시간을 묻는 문항이 있었는데 본 연구에서는 일주일 간 땀이 날 정도로 운동한 시간을 묻는 문항이 선택되었다. 운동 강도에 따른 신체 활동 정도가 스마트폰 의존과 어떠한 관계가 있는지에 대한 후속 연구가 필요하겠다.

마지막으로 창의적 성격과 그릿(grit; Duckworth et al., 2007)을 측정하는 문항 일부가 새롭게 탐색되었는데, 이 변수들이 통계적으로도 유의한 변수였다는 점을 주목할 만하다. 구체적으로 창의적 성격을 측정하는 문항 중 의심이 많다고 답할수록 스마트폰 의존도가 낮았고, 그릿 문항 중 쉽게 흥미를 잃고 노력 지속 시간이 짧다고 답할수록 스마트폰 의존도가 높았다. 그릿의 하위 범주인 흥미유지와 노력지속에서 각각 한 문항씩 선택된 점 또한 흥미롭다. 본 연구만으로 창의적 성격 또는 그릿이 스마트폰 의존을 예측한다고 단언할 수 없으나, 본 연구에서 새롭게 탐색된 변수들은 100개 모형에서 100번 모두 중요한 예측변수로 선택된 변수임과 동시에 통계적 유의성 또한 검증된 변수로 의의가 있다. 창의성과 그릿 모두 KCYPS 2018에 새로 추가된 구인이므로 향후 관련 연구가 활발할 것으로 기대한다. 특히 창의성 또는 그릿과 스마트폰 의존 간 관계에 대한 후속 연구를 제안한다.

2. 연구 의의 및 제언

1996년 LASSO가 발표된 이래 별점회귀모형에서의 통계적 추론은 통계학자들의 당면 연구 과제 중 하나였다(Hastie et al., 2015; Yoo & Rho, in press). 본 연구는 사

회과학 대용량자료 분석 시 기계학습 모형에 통계적 추론까지 수행한 국내 최초의 연구로 방법론적 의의를 지닌다. 별점회귀모형에서의 통계적 추론 기법인 PSI를 사회과학 대용량자료인 KCYPS 자료 분석에 적용함으로써 선택된 17개 변수는 단순히 탐색적 기법만을 고려한 결과가 아니라, 전통적 기법에서 중시하는 통계적 추론까지 통한 변수다. 다시 말해, 이 변수들은 탐색적 기법과 통계적 추론을 모두 고려하여 얻은 결과로서 의의가 있다.

PSI는 LASSO로 선택된 변수에 대하여 다시 통계적 유의성을 검정하기 때문에 탐색적 기법만 적용한 연구보다 보수적인 특징을 지닌다. 본 연구의 PSI 결과로 선택된 다수의 변수가 선행연구에서도 통계적으로 유의한 변수로 연구된 바 있다. 즉, PSI를 적용함으로써 결과적으로 불필요한 변수가 선택될 확률이 줄어들었으나, 같은 맥락에서 새로운 변수가 탐색될 확률 역시 감소한 것으로 보인다. 다양한 사회과학 대용량자료 분석에 PSI를 적용하는 후속 연구가 진행될 필요가 있다. KCYPS 뿐만 아니라 KEEP(Korean Education and Employment Panel; 한국교육고용패널조사), KELS(Korean Education Longitudinal Study; 한국교육종단연구) 등의 대용량자료가 공개되고 있다. 통계적 추론까지 고려하는 별점회귀모형 기법을 다양한 유형의 사회과학 대용량자료 분석에 적용하고, 그 결과를 바탕으로 모의실험 연구를 설계하고 분석할 것을 제안한다.

실험 연구를 통하여 스마트폰 의존 관련 변수들의 인과 관계를 밝히는 것은 불가능에 가깝다. 연구자들은 스마트폰 의존과 관련 있는 변수들의 관계를 간접적으로 확인함으로써 스마트폰 의존을 예방하기 위한 대책 마련에 힘쓰고 있다. 본 연구는 수백 개의 변수 중에서 스마트폰 의존 관련 변수를 별점회귀모형을 이용하여 탐색하고, 선택된 소수의 변수에 대하여 통계적 유의성까지 검정하였다. 수많은 소규모 연구들을 통해 밝혀진 결과 중 상당 부분을 본 단일 연구를 통해서 확인할 수 있었다는 점에서 연구방법의 유용성을 확인할 수 있었으며, 본 연구 결과 선별된 변수는 중학생의 스마트폰 의존을 예방하는 교육적 조치 및 대책 마련에 유용한 자료로 기능할 수 있을 것이다. 사회과학 대용량자료 분석에 PSI를 적용한 첫 번째 연구인 본 연구를 시작으로 향후 PSI 기법을 적용하는 연구가 확산되기를 기대한다.

참고 문헌

- 정승구, 김진욱 (2019). 청소년의 스마트폰 중독에 관한 연구: 스마트폰 과의존, 게임 중독, SNS 중독 영향요인의 비교를 중심으로. **생명연구**, 52, 179-200. doi:10.17924/solc.2019.52.179
- 고은혜, 김근영 (2017). 중학생의 수줍음과 대인불안이 스마트폰 중독에 미치는 영향: 부모-자녀 간 폐쇄형 의사소통과 또래관계의 매개효과를 중심으로. **청소년학연구**, 24(5), 55-78. doi:10.21509/KJYS.2017.05.24.5.55
- 과학기술정보통신부, 한국정보화진흥원 (2019). **2019 스마트폰 과의존 실태조사 (NIA VIII-RSE-C-19067)**. 세종: 과학기술정보통신부.
- 김동일, 정여주, 이주영, 김명찬, 이윤희, 강은비, 금창민, 남지은 (2012). 성인용 간략형 스마트폰 중독 자가진단 척도 개발. **상담학연구**, 13(2), 629-644. doi:10.15703/kjc.13.2.201204.629
- 김보경, 백용매, 허창구 (2016). 애착과 스마트폰 및 SNS 중독의 관계에서 대인관계 문제의 매개효과. **청소년학연구** 23(3), 483-502. doi:10.21509/KJYS.2016.03.23.3.483
- 김상립 (2018). 청소년의 자기통제력과 스마트폰 과의존 간 관계에서 부모-자녀의사소통 및 학교생활적응의 매개효과. **한국생활과학회지**, 27(2), 91-102. doi:10.5934/kjhe.2018.27.2.91
- 김아영, 주지은 (1999). 학습된 무기력, 실패내성과 학업성취간의 관계. **교육과학연구**, 29, 157-176.
- 노민정, 유진은 (2019a). Adaptive LASSO를 통한 진로결정 관련 변수 탐색. **열린교육연구**, 27(4), 133-155. doi:10.18230/tjye.2019.27.4.133
- 노민정, 유진은 (2019b). 사회과학 대용량 자료 분석을 위한 벌점회귀모형과 결측처리 기법의 성능 비교: 몬테카를로 모의실험. **교육평가연구**, 32(4), 755-776. doi:10.31158/JEEV.2019.32.4.755
- 노민정, 유진은 (2020). Group Mnet 기계학습 기법을 이용한 고등학생의 진로개발역량 관련 변수 탐색. **직업능력개발연구**, 23(2), 83-109.
- 노충래, 김소연 (2016). 부모양육태도가 청소년의 스마트폰 의존에 미치는 영향: 청소년의 우울과 공격성의 매개효과. **한국청소년연구**, 27(4), 87-114. doi:10.14816/sky.2016.27.4.87

- 류세인, 조인숙 (2015). 초등학생의 스마트폰 중독에 미치는 영향요인. **한국산학기술학회논문지**, 16(9), 6180-6189. doi:10.5762/KAIS.2015.16.9.6180
- 문호경 (2020). 부모양육태도가 청소년의 스마트폰의존에 미치는 영향: 자아존중감과 우울의 직렬 이중매개효과. **학교사회복지**, 49, 25-43. doi:10.20993/jSSW.49.2
- 박병기, 노시연, 김진아, 황진숙 (2015). 학업무기력 척도의 개발 및 타당화. **아동교육**, 24(4), 5-29. doi:10.17643/KJCE.2015.24.4.01
- 박소연, 유미숙 (2019). 방임경험이 청소년의 스마트폰 의존에 영향을 미치는 경로: 자아탄력성, 소외감의 매개효과. **재활치료과학**, 8(1), 27-39. doi:10.22683/tsnr.2019.8.1.027
- 박소현, 방성완, 전명식 (2011). 순차 적응 최근접 이웃을 활용한 결측값 대체법. **응용통계연구**, 24(6), 1249-1257. doi:10.5351/KJAS.2011.24.6.1249
- 박창이, 김용대, 김진석, 송종우, 최호식 (2011). **R을 이용한 데이터마이닝**. 서울: 교우사.
- 배성만 (2018). 스마트폰 사용유형과 스마트폰에 대한 과도한 기대가 청소년의 스마트폰 과의존에 미치는 영향. **정보화정책**, 25(4), 65-83. doi:10.22693/NIAIP.2018.25.4.065
- 서보경, 차성수 (2020). 대학생의 우울이 스마트폰 과의존에 미치는 영향: 자기효능감의 매개효과를 중심으로. **한국콘텐츠학회논문지**, 20(2), 47-486. doi:10.5392/JKCA.2020.20.02.478
- 서은경, 김현철, 황수진, 이철원 (2018). 특목고 진학계획의 사교육비 지출에 대한 인과 효과 분석. **교육학연구**, 56(2), 161-190. doi:10.30916/KERA.56.2.161
- 신선임 (2017). 부모애착이 스마트폰 중독에 미치는 영향. **청소년학연구** 24(5), 131-154. doi:10.21509/KJYS.2017.05.24.5.131
- 양영미, 이선영 (2019). 아동의 우울과 스마트폰 과의존 간의 관계-체육활동 참여의 조절효과-. **한국아동복지학**, 67, 35-59. doi:10.24300/jkscw.2019.09.67.35
- 염동문 (2017). 초등학생의 스마트폰 의존도에 관한 잠재프로파일분석. **재활복지공학회논문지**, 11(2), 107-114. doi:10.21288/resko.2017.11.2.107
- 오현희, 김현진 (2014). 중·고등학생이 지각하는 부·모애착과 인터넷 및 스마트폰 중독의 관계에서 가족의사소통의 매개효과 분석. **한국청소년연구**, 25(4), 35-57. doi:10.14816/sky.2014.11.75.35

- 유진은 (2016). 데이터 마이닝 기법을 통한 교육 패널데이터 분석: 벌점회귀모형과 KYPs 자료. **아시아교육연구**, 17(3), 1-19. doi:10.14816/sky.2014.11.75.35
- 유진은, 김형관, 노민정 (2020). Group Mnet 기계학습 기법을 통한 중학생의 끈기(grit) 관련 변수 탐색. **한국청소년연구**, 31(1), 157-182. doi:10.14816/sky.2020.31.1.157
- 유진은, 노민정 (2017). Group lasso를 통한 중학생의 삶의 만족도에 영향을 미치는 변수 탐색. **한국청소년연구**, 28(1), 127-149. doi:10.14816/sky.2017.28.1.127
- 이규현, 권윤희 (2014). 중학생의 스마트폰 중독 영향요인. **한국생활환경학회지**, 21(5), 783-790.
- 이수진, 문혁준 (2013). 중학생의 자기통제, 부모-자녀 간의 의사소통 및 학교생활만족도가 스마트폰 중독에 미치는 영향. **한국생활과학회지**, 22(6), 587-598. doi:10.5934/kjhe.2013.22.6.587
- 이윤미, 권윤나, 채규만 (2019). 중학생의 부모 애착이 스마트폰 중독에 미치는 영향: 자기통제력의 조절 효과를 중심으로. **청소년문화포럼**, 59, 55-79. doi:10.17854/ffyc.2019.07.59.55
- 이은경, 안지영, 김지신 (2020). 중학생의 사회적 위축과 스마트폰 의존도 사이에서 부정적인 친구관계와 학업 무기력의 직렬매개효과. **한국가족복지학**, 25(3), 271-289. doi:10.13049/kfwa.2020.25.3.2
- 이용택, 오승근, 백진현, 이승봉 (2020). 청소년기 스마트폰 사용에서의 기본심리욕구와 스마트폰 중독의 관계에서 자기효능감의 조절효과. **청소년문화포럼**, 61, 67-94. doi:10.17854/ffyc.2020.01.61.67
- 이하나, 양승목 (2018). 청소년의 스마트폰 의존도와 영향 변인의 관계에 대한 학령 및 성별 비교 분석: 부모의 양육태도, 청소년의 심리적 특성, 스마트폰 이용 용도를 중심으로. **한국언론학보**, 62(5), 175-214. doi:10.20879/kjics.2018.62.5.006
- 이혜진 (2017). 청소년의 스트레스와 인터넷·스마트폰 중독의 관계에서 다차원적 경험회피의 매개효과. **스트레스연구**, 25(4), 279-285. doi:10.17547/kjsr.2017.25.4.279
- 정규석 (2017). 부모애착과 전념이 중학생의 스마트폰 중독에 미치는 영향: 자기통제의 매개효과. **학교사회복지**, 38, 1-19.
- 정인경, 김정현 (2015). 중학생의 여가만족 및 여가제약이 스마트폰 중독에 미치는 영향: 자기효능감의 매개효과를 중심으로. **한국여가레크리에이션학회지**, 39(1), 67-84.

- 정인경, 김정현 (2017). 중학생들의 학업스트레스와 학업소진이 스마트폰 중독에 미치는 영향. **한국지역사회생활과학회지**, 28(2), 289-300. doi:10.7856/kjcls.2017.28.2.289
- 정민 (2015). 중학생의 스마트폰 중독에 영향을 미치는 변인 간의 구조적 관계: 생태 체계적 관점을 중심으로. **한국청소년연구**, 26(3), 103-131. doi:10.14816/sky.2015.26.3.103
- 조건상, 이영익 (2018). 청소년의 체육수업활동과 스마트폰 의존도 및 학교생활적응의 관계. **한국융합학회논문지**, 9(10), 355-362. doi:10.15207/JKCS.2018.9.10.355
- 조운주 (2019). 학령후기 아동의 스마트폰 의존에 대한 종단적 변화와 예측요인. **아동학회지**, 40(6), 49-62. doi:10.5723/kjcs.2019.40.6.49
- 주석진, 조성심 (2015). 청소년의 스마트폰 중독 판별요인에 관한 연구: 일평균 스마트폰 사용시간, 충동성, 교우관계, 모의 권위주의형 부모양육태도, 학교부적응을 중심으로. **청소년복지연구**, 17(1), 97-118.
- 하형석, 이종원, 정은진, 김성은, 한지형, 구분호 (2018). **한국아동·청소년 패널조사 IX: 사업보고서**. 세종: 한국청소년정책연구원.
- 한국갤럽 (2019.9.29.) 2012-2019 스마트폰 사용률, 브랜드, 스마트워치, 손목시계에 대한 조사. <https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=1041>에서 2020년 9월 15일 인출
- 한국정보화진흥원 (2017). 2016년 스마트폰 과의존 실태조사(NIA V-RER-C-16026). 대구: 한국정보화진흥원.
- 허균 (2013). 고차잠재성장모형을 적용한 휴대폰 의존도 변화와 자기조절능력 및 주의집중의 구조관계 연구. **교육정보미디어연구**, 19(4), 861-879.
- KBS 뉴스 (2019.2.11.). 국민 95%가 스마트폰 사용...보급률 1위 국가는?. <https://news.kbs.co.kr/news/view.do?ncd=4135732>에서 2020년 09월 15일 인출.
- Bian, M., & Leung, L. (2015). Linking loneliness, shyness, smartphone addiction symptoms, and patterns of smartphone use to social capital. *Social Science Computer Review*, 33(1), 61-79. doi:10.1177/0894439314528779
- Bühlmann, P., & Mandozzi, J. (2014). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Computational Statistics*, 29(3-4), 407-430. doi:10.1007/s00180-013-0436-3

- Choi, S. W., Kim, D. J., Choi, J. S., Ahn, H., Choi, E. J., Song, W. Y., ... & Youn, H. (2015). Comparison of risk and protective factors associated with smartphone addiction and Internet addiction. *Journal of Behavioral Addictions*, 4(4), 308-314. doi:10.1556/2006.4.2015.043
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long term goals. *Journal of Personality and Social Psychology*, 92(6), 1087-1101. doi:10.1037/0022-3514.92.6.1087
- Fan, J., Feng, Y., & Wu, Y. (2009). Network exploration via the adaptive LASSO and SCAD penalties. *The Annals of Applied Statistics*, 3(2), 521. doi:10.1214/08-AOAS215SUPP
- Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849-911. doi:10.1111/j.1467-9868.2008.00674.x
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1-22. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2929880/>
- Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, 56, 229-238. doi:10.1016/j.jbi.2015.05.016
- Ghosh, D., & Chinnaiyan, A. M. (2005). Classification and selection of biomarkers in genomic data using LASSO. *Journal of Biomedicine and Biotechnology*, 2005(2), 147-154.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. Boca Raton, FL: CRC press.
- Huang, J., Ma, S., & Zhang, C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, 18, 1603-1618. <https://www.jstor.org/stable/24308572>
- Javanmard, A., & Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1), 2869-2909.

- Jönsson, P., & Wohlin, C. (2004). An evaluation of k-nearest neighbour imputation using likert data. *In Proceedings of 10th International Symposium on Software Metrics, IEEE*, 108-118. doi:10.1109/METRIC.2004.1357895.
- Kim, M. H., Banerjee, S., Park, S. M., & Pathak, J. (2017). Improving risk prediction for depression via Elastic Net regression - Results from Korea National Health Insurance Services Data. *AMIA... Annual Symposium proceedings. AMIA Symposium, 2016*, 1860-1869.
- Kowarik, A., & Templ, M. (2016). Imputation with the R Package VIM. *Journal of Statistical Software*, 74(7), 1-16. doi:10.21256/zhaw-4233
- Kwon, M., Kim, D. J., Cho, H., & Yang, S. (2013). The smartphone addiction scale: Development and validation of a short version for adolescents. *PloS one*, 8(12), e83558. doi:10.1371/journal.pone.0083558
- Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3), 907-927. doi:10.1214/15-AOS1371
- Lin, Y. H., Chang, L. R., Lee, Y. H., Tseng, H. W., Kuo, T. B., & Chen, S. H. (2014). Development and validation of the smartphone addiction inventory (SPAII). *PloS one*, 9(6), e98312. doi:10.1371/journal.pone.0098312
- Lu, M., Zhou, J., Naylor, C., Kirkpatrick, B. D., Haque, R., Petri, W. A., & Ma, J. Z. (2017). Application of penalized linear regression methods to the selection of environmental enteropathy biomarkers. *Biomarker Research*, 5(1), 9. doi:10.1186/s40364-017-0089-4
- Ma, S., & Huang, J. (2008). Penalized feature selection and classification in bioinformatics. *Briefings in Bioinformatics*, 9(5), 392-403. doi:10.1093/bib/bbn027
- Ma, S., Song, X., & Huang, J. (2007). Supervised group Lasso with applications to microarray data analysis. *BMC Bioinformatics*, 8, 60. doi:10.1186/1471-2105-8-60
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), 417-473. doi:10.1111/j.1467-9868.2010.00740.x

- Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. *Finance Research Letters*, 27, 235-240. doi:10.1016/j.frl.2018.03.016
- Rida, I., Jiang, X., & Marcialis, G. L. (2015). Human body part selection by group lasso of motion for model-free gait recognition. *IEEE Signal Processing Letters*, 23(1), 154-158. doi:10.1109/LSP.2015.2507200
- Samaha, M., & Hawi, N. S. (2016). Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. *Computers in Human Behavior*, 57, 321-325. doi:10.1016/j.chb.2015.12.045
- Shevade, S. K., & Keerthi, S. S. (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19(17), 2246-2253. doi:10.1093/bioinformatics/btg308
- Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25), 7629-7634. doi:10.1073/pnas.1507583112
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Tibshirani, R., Tibshirani, R., Taylor, J., Loftus, J., Reid, S., & Markovic, J. (2019). SelectiveInference: tools for post-selection inference. R package version 1.2.5. <https://CRAN.R-project.org/package=selectiveInference>
- Van Deursen, A. J., Bolle, C. L., Hegner, S. M., & Kommers, P. A. (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in Human Behavior*, 45, 411-420. doi:10.1016/j.chb.2014.12.039
- Wang, H., Xu, Q., & Zhou, L. (2015). Large unbalanced credit scoring using lasso-logistic regression ensemble. *PloS one*, 10(2), e0117844. doi:10.1371/journal.pone.0117844

- Wu, T. T., Chen, Y. F., Hastie, T., Sobel, E., & Lange, K. (2009). Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6), 714-721. doi:10.1093/bioinformatics/btp041
- Yoo, J. E., & Rho, M. (2020). Exploration of predictors for Korean teacher job satisfaction via a machine learning technique, Group Mnet. *Frontiers in Psychology*, 11, 441. doi:10.3389/fpsyg.2020.00441
- Yoo, J. E., & Rho, M. (in press). Large-scale survey data analysis with penalized regression: A Monte Carlo simulation on missing categorical predictors. *Multivariate Behavioral Research*, 56.
- Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7(Nov), 2541-2563.
- Zou, C., Jiang, W., & Tsung, F. (2011). A lasso-based diagnostic framework for multivariate statistical process control. *Technometrics*, 53(3), 297-309. doi:10.1198/TECH.2011.10034

ABSTRACT

Statistical inference after variable selection in penalized regression: Focusing on variables relating to adolescents' smartphone reliance

Rho Minjeong* · Yoo Jin Eun**

The purpose of this research was to explore variables predicting adolescents' smartphone reliance and to provide research-based evidences for possible educational interventions. In particular, this study employed post-selection inference after penalized regression in analyzing KCYPS 2018 2nd wave data. In order to consider the inconsistency issues of LASSO with respect to variable selection, this study repeated 100 times of model selection and post-selection inference. A total of 17 variables were identified as statistically significant 50% of the iterations; the variables included predictors relating to conversation with parents, leisure time, satisfaction on academic achievement, academic helplessness, attention, amount of workout, creative personality, and grit. This study bears methodological implications as one of the first studies to apply post-selection inference to social science large-scale data. Future research directions are suggested.

Key Words: smartphone reliance, machine learning, penalized regression, post-selection inference, regularization

투고일: 2020. 12. 7, 심사일: 2020. 12. 17, 심사완료일: 2021. 1. 13

* Korea National University of Education(Doctor, Department of Education, E-mail: minjeong019@gmail.com)

** Korea National University of Education(Professor, Department of Education, E-mail: jeyoo@knue.ac.kr)