

연구논문

매개모형에서 경향점수를 활용한 사전 특성 통제방법*

이진실**

본 논문은 조사연구에서 많이 활용되고 있는 매개모형에서 사전 특성을 통제하기 위한 방법으로 경향점수 역확률가중치를 활용한 방법을 비교하고 검증하였다. 이를 위해 모의실험에서는 4가지 모형(경향점수를 적용하지 않은 통제 전 모형, 처치변인에 대한 경향점수만 추정하여 활용한 모형, 매개변인에 대한 경향점수만 활용한 모형, 처치변인과 매개변인에 대한 경향점수를 각각 추정하여 곱 가중치를 통해 활용한 모형)에서의 효과 추정치의 정확성을 비교하였다. 연구의 결과, 처치변인과 매개변인에 대한 경향점수를 모두 추정하여 곱하는 방식을 통해 가중치를 적용한 모형에서 효과 추정치의 편이가 가장 작게 나타났다. 실제자료 분석에서도 처치변인과 매개변인에 대한 경향점수를 추정한 후 가중치를 곱하여 적용했을 때 사전 특성의 체계적인 차이가 크게 줄어드는 것으로 확인되었다. 연구결과를 바탕으로 매개모형에서 사전 특성 통제의 중요성 및 경향점수의 활용방법을 제안하였다.

주제어: 매개모형, 경향점수, 역확률가중치, 사전 특성 통제

* 본 연구는 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2019S1A5B5A07109438).

** 서울대학교 강사(leejinsil326@gmail.com).

I. 서론

조사연구를 통해 특정한 처치나 변인의 효과를 경험적으로 확인하고자 할 때가 많이 있다. 예를 들어, 진로지도 프로그램이 학생의 진로성숙도를 향상시키는가?, 봉사활동에 참여하면 공동체意識이 높아지는가? 등은 처치효과에 관심을 둔 연구 질문들이다. 처치효과를 분석할 때 인과적인 효과를 확인하기 위해서는 처치에 참여하기 전에는 두 집단이 동질적이어야 하는데, 실험연구에서는 피험자들을 처치집단과 통제집단에 무작위로 할당함으로써 비교하고자 하는 두 집단의 동질성을 확보하고자 한다(Murnane & Willett 2011). 그러나 사회과학 연구에서는 윤리적 문제 등으로 무선배치(random assignment) 실험연구가 불가능한 경우가 많고, 조사연구를 통해 효과를 추론해야 하는 경우가 많다.

조사연구 자료를 활용해 인과효과를 분석하고자 할 때 사전 특성을 통제하는 것은 매우 중요한 문제이다. 사전 특성이 통제되지 못하면 분석된 효과에는 처치가 이루어지기 전 이미 존재하는 집단 간 차이도 함께 반영되기 때문에, 추정된 효과가 진짜 처치로 인한 것인지 확신하기 어렵다. 예를 들어, 봉사활동의 효과를 분석하고자 할 때, 사전 특성을 고려하지 않고서 봉사활동 참여 집단과 비참여 집단을 단순 비교할 경우, 두 집단 간 차이에는 봉사활동의 효과뿐만 아니라 봉사활동에 참여하기 전 이미 가지고 있었던 특성까지 반영될 가능성이 높다. 실제로 선행연구들에 의하면 가정의 소득이 더 높은 경우, 부모가 봉사활동에 참여한 경험이 있는 경우, 종교를 가진 경우 봉사활동에 더 많이 참여하는 것으로 보고된 바 있다(김옥·송미영 2006; 오혜정·손병덕·임희수 2018; 최영창·장연심 2008; 홍은진 2006). 따라서 사전 특성을 통제하지 않고 추정된 봉사활동의 효과에는 가정환경이나 개인 특성에 의한 차이까지 혼재되어 있을 가능성이 높다.

한편 조사연구에서 변인들 간의 인과관계를 분석하기 위해 많이 활용되는 방법으로 구조방정식모형을 들 수 있다. 구조방정식모형은 인과모형(causal modeling)으로 불리기도 하는데(Kline 2005), 이는 자료를 분석할 때 변인들 간 인과관계를 모형으로 설정하여 분석하기 때문에 붙여진 이름으로 볼 수 있다. 하지만 구조방정식모형을 통해 자료를 분석하더라도 사전 특성이 통제되지 못한 경우에는, 분석된 그

관계가 허구상관으로 인한 관계일 수 있다. 구조방정식모형에서, 특히 매개변인 (mediator)을 포함하여 분석하는 경우에는 사전 특성들을 통제하기가 더 복잡하고 어려워진다(이진실 2013). 매개변인이 포함된 구조방정식모형에서 사전 특성들을 통제하기 위해 통제변인을 모형에 함께 투입하게 되면, 처치변인과 통제변인의 관계, 매개변인과 통제변인의 관계, 결과변인과 통제변인의 관계, 심지어 통제변인들 간의 관계까지도 모두 적절하게 설정해주어야 하며, 통제해야 할 변인의 수가 늘어나면 모형이 급격하게 복잡해지는 문제가 발생한다. 표본크기가 한정된 상황에서 추정해야 할 모수의 수가 늘어나게 되면 모수를 안정적으로 추정하기 어려워지는 문제도 생긴다. 이러한 한계로 인해 구조방정식모형을 활용하여 매개효과를 분석한 기존의 연구들에서는 사전 특성을 통제하지 않고 분석하거나, 사전 특성을 통제하더라도 5개 이하의 일부 변인만을 포함하여 분석한 경우가 많았다(김지혜 2012; 김선숙·안재진 2012; 노충래·김소연 2016; 박재숙 2011; 조복순·안상근 2017; 홍나미 2015 등).

앞서 언급하였듯이 사전 특성이 잘 통제되지 못하면 분석된 효과가 인과적인 효과로 해석될 수 없으며, 따라서 매개변인이 포함된 모형에서 사전 특성을 효과적으로 통제할 방법을 모색하는 것이 필요하다. 이때 대안으로 고려될 방법으로 경향점수(propensity score)를 생각할 수 있다. 경향점수는 조사연구를 통해 수집된 관찰 자료를 활용해 인과효과를 추론하기 위해 Rosenbaum & Rubin(1983)이 제안한 방법으로, 사전 특성 정보를 활용하여 각 사례가 ‘처지를 받을 확률, 즉 경향점수’를 추정하고 처지-통제 집단에서 경향점수가 비슷해지도록 조정하여 집단 간 동질성을 확보하고자 한다. 경향점수의 활용은 통제하고 싶은 여러 변인들을 1차원으로 함축하여 통제할 수 있다는 점에서 장점을 지닌다. 또한 각각의 사전 특성을 모형에 직접 투입하여 통제하는 방법에서는 사전 특성들의 분포가 겹치지 않을 때 외삽에 의존해야 하지만, 경향점수를 활용한 방법에서는 사전 특성의 분포가 겹치는 사례만 분석에 활용함으로써 분석의 정확성을 높일 수 있다는 장점도 있다.

경향점수를 활용하여 처치효과를 분석하고자 한 연구들은 그동안 많이 제시되어 왔지만(김준엽·정혜경·Seltzer 2008; 박현정·이진실·이용석 2015; 변상민·신중휘·양준혁 2018 등), 매개변인이 포함된 모형에 경향점수를 적용하고자 한 연구는 아직까지 많지 않다. 매개변인이 포함된 구조방정식모형에 경향점수를 적용하고자 한 Coffman(2011)의 연구에서는 ‘매개변인’에 대한 경향점수를 적용하였을 때 ‘매개변인이 결과변인에 미치는 효과’가 정확하게 추정되는지를 모의실험을 통해 확인하였

다. 그러나 이 연구에서는 ‘처치변인이 매개변인에 미치는 효과’와 ‘처치변인이 결과변인에 미치는 직접효과’는 고정하고서 연구를 진행했기 때문에, 고정한 다른 효과들에 대해서는 확인하지 못했다. 이준호·박현정(2012)과 하여진(2016)의 연구에서는 ‘처치변인’에 대한 경향점수를 추정하여 매개모형에 적용하고자 하였다. 그러나 이진실(2013)의 연구에 따르면, ‘처치변인’에 대해서만 경향점수를 추정하여 활용하는 방법에서는 ‘처치변인이 매개변인에 미치는 효과’와 ‘처치변인이 결과변인에 미치는 직접효과’는 정확하게 추정되지만, ‘매개변인이 결과변인에 미치는 효과’에서는 편의가 발생하는 것으로 나타났다. 이처럼 선행연구들에서는 주로 처치변인 또는 매개변인 중 하나만을 기준으로 경향점수를 활용하였고, 매개모형에서 추정되는 3가지 유형의 개별 효과 추정치(처치변인이 매개변인에 미치는 효과, 매개변인이 결과변인에 미치는 효과, 처치변인이 결과변인에 미치는 직접효과) 중에서 일부는 사전 특성이 잘 통제되지 못하는 한계가 있었다.

이 연구에서는 이러한 한계점을 보완할 수 있는 방법으로 처치변인 및 매개변인 각각에 대해 경향점수를 모두 추정한 후 이를 곱해서 활용하는 방법을 검증하고 비교하고자 한다. 경향점수를 곱해서 활용하는 방법은 종단연구 상황에 적용되어 온 방법으로, 처치가 한 시점에서만 이루어지는 것이 아니라 여러 시점에 걸쳐 반복적으로 이루어지는 경우 각 처치에 대한 경향점수를 곱하여 활용해 왔다(Robins, Greenland, & Hu 1999; Robins, Hernan, & Brumback 2000; 박현정·이준호 2012 등). 이 연구에서는 경향점수를 곱해서 사용하는 방법이 처치변인 또는 매개변인 하나만 기준으로 경향점수를 추정하여 적용한 모형에 비해 장점을 가지는지 확인하고자 한다. 또한 실제자료의 분석 예시도 함께 제시하여, 매개모형을 분석하는 연구자들에게 경향점수를 활용하여 사전 특성을 통제하는 방법을 안내하고자 한다.

Ⅲ. 이론적 배경

1. 경향점수를 활용한 사전 특성 통제방법

경향점수는 ‘처치에 참여할 확률(처치집단에 할당될 확률)’을 의미하며, 경향점수를 활용한 인과추론 방법에서는 처치집단과 통제집단의 경향점수의 분포가 비슷해

지도록 조정해줌으로써 사전 특성의 영향을 통제한다(Rosenbaum & Rubin 1983). 경향점수의 추정에는 관찰된 사전 특성들을 독립변인으로, 처치변인을 종속변인으로 하는 로지스틱 회귀분석 또는 프로빗 회귀분석을 통해 추정될 수 있다(김준엽 외 2008). 로지스틱 회귀분석을 통해 사전 특성들을 고려했을 때 각 사례가 처치에 참여할 확률-경향점수를 저장하여 이를 경향점수로 활용할 수 있다.

경향점수를 활용하는 방법은 매칭, 층화, 역확률가중치 등의 방법이 있으며, 특히 역확률가중치(Inverse Probability Weighting; IPW) 방법은 자료의 구조나 변수의 특성에 제한을 적게 받는다는 장점(Curtis et al. 2007) 때문에 광범위하게 활용되고 있다. 역확률가중치 방법은 각 피험자가 해당 집단에 속할 확률의 역수를 가중치로 활용하여, 가중치를 통해 처치집단과 통제집단의 사전 특성의 분포를 비슷하게 조정해주는 방법이다. 예를 들어, 봉사활동의 효과를 검증하고자 할 때 가구소득이 봉사활동 참여에 영향을 미친다고 생각해보자. 이 때 봉사활동 참여 집단에는 가구소득을 고려했을 때 봉사활동에 참여할 기대 확률(\hat{p})이 높은 학생들(즉 가구소득이 높은 학생들)이 과대표집되어 있을 것이고, 봉사활동 비참여 집단에는 가구소득을 고려했을 때 봉사활동에 참여하지 않을 확률($1-\hat{p}$)이 높은 학생들(즉 가구소득이 낮은 학생들)이 과대표집되어 있을 것이다. 이러한 경우 참여 집단에서는 $\frac{1}{\hat{p}}$ 로 가중치를 부여하고, 비참여 집단에서는 $\frac{1}{1-\hat{p}}$ 로 가중치를 부여하게 되면, 특정 조건의 학생들이 과대표집되어 있는 비율을 조정해 줄 수 있기 때문에, 사전 특성에 있어서 집단 간 균형을 맞출 수 있다.

처치변인이 참여/비참여로만 구분된 이분형 변인이 아닌, 3개 이상의 범주를 가진 다분형 변인이거나 연속형 변인일 경우에는 더 확장된 형태인 일반화 경향점수(Generalized Propensity Score; GPS)를 활용할 수 있다(Hirano & Imbens 2004). 연속형 변인에 대한 경향점수는 사전 특성을 고려했을 때 각 사례가 해당하는 수준(값)의 처치를 받을 확률을 의미하며, 사전 특성들을 독립변인으로 처치변인을 종속변인으로 하는 회귀분석을 실시한 후 정규분포의 확률밀도함수를 활용하여 추정할 수 있다.

경향점수 역확률가중치 방법은 사전 특성을 모형에 직접 투입하여 통제하는 방법에 비해 다음과 같은 장점을 가진다. 사전 특성을 모형에 직접 투입하는 방법에서는 통제해야 하는 사전 특성의 개수가 많아질수록 연구모형이 복잡해지는 데 반해, 경향점수 역확률가중치 방법에서는 모형의 간명성을 유지하면서 사전 특성의

영향력을 제거할 수 있다. 또한 경향점수를 활용한 방법에서는 처치집단과 통제집단의 경향점수의 분포를 확인한 후 경향점수의 분포가 겹치는 사례, 즉 공통영역(common support)에 해당하는 사례만 추출하여 분석할 수 있기 때문에 비현실적인 외삽에 의존하지 않아도 된다는 장점이 있다. 또한 처치효과를 분석하기 전 집단 간 사전 특성의 분포가 잘 조정이 되었는지를 밸런스 테스트(balance test)를 통해 미리 검증할 수 있으므로 분석의 정확성을 높일 수 있다.

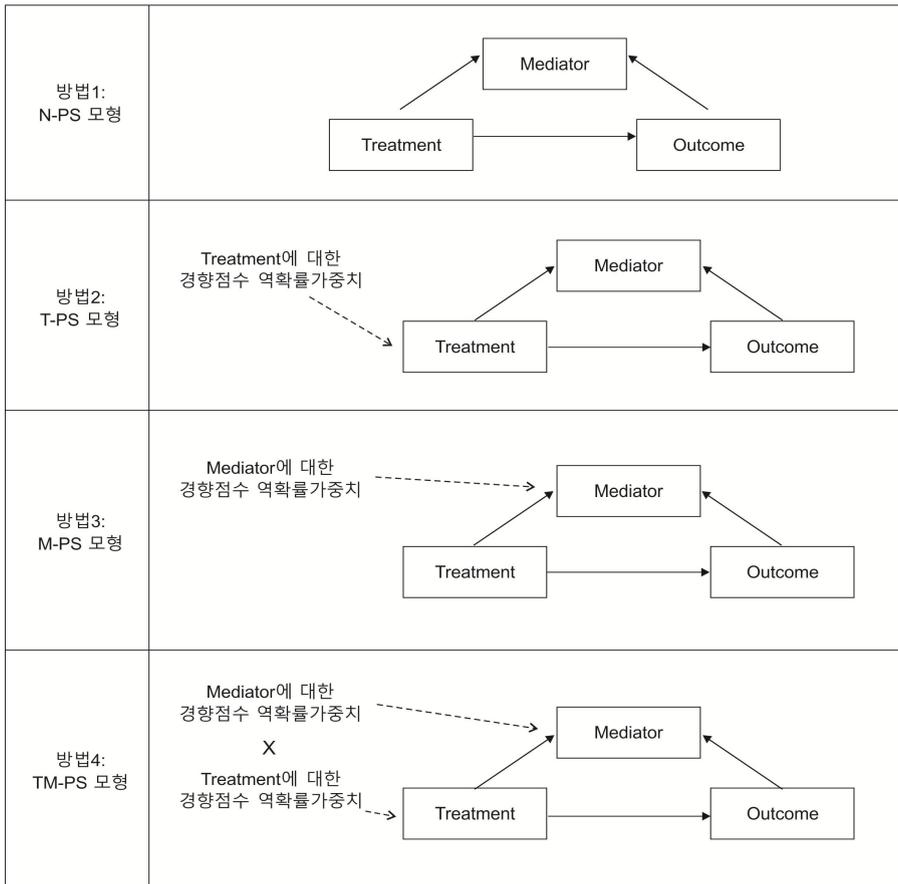
경향점수 역확률가중치 방법의 단점은 극단적인 가중치에 영향을 받을 수 있는데, 이를 보완하기 위해 안정화 역확률가중치(stabilized IPW)를 활용하는 것이 권장된다. 안정화 역확률가중치는 사전 특성을 고려하지 않고 구해진 경향점수를 분자에 활용하는 방법으로, 안정화 역확률가중치를 활용하면 분자와 분모의 극단적인 차이를 줄여줄 수 있다(Cole & Hernan 2008; Xu et al. 2010).

2. 매개모형에서 경향점수 역확률가중치 활용방법

매개모형에서는 처치변인뿐만 아니라 매개변인에 대해서도 경향점수를 추정할 수 있고, 처치변인에 대한 경향점수와 매개변인에 대한 경향점수를 어떻게 활용할 지에 따라 <그림 1>과 같이 활용방법이 구분될 수 있다. 매개모형에서 경향점수 역확률가중치를 활용하는 방법은 크게 처치변인에 대한 경향점수만 활용하는 방법, 매개변인에 대한 경향점수만 활용하는 방법, 처치변인과 매개변인에 대한 경향점수를 동시에 활용하는 방법으로 구분해볼 수 있다.

<그림 1>의 방법1(N-PS 모형)은 경향점수 역확률가중치를 적용하지 않은 모형으로 비교를 위해 제시하였다. 경향점수를 적용하지 않고서 매개모형을 분석하는 방법1에서는 사전 특성의 영향력이 통제되지 못하여, 효과 추정치가 편향되어 추정될 가능성이 있다. 방법2(T-PS 모형)는 처치변인에 대해서만 경향점수를 추정한 후 경향점수 역확률가중치를 적용하는 방법이다. 이 방법은 이준호·박현정(2012), 이진실(2013), 하여진(2016)의 연구에서 적용된 바 있다. 이러한 방법에서는 매개변인에 대한 사전 특성의 영향력은 교정되지 못하기 때문에, 선행연구(이진실 2013)에 따르면 ‘매개변인이 결과변인에 미치는 효과’는 부정확하게 추정되는 것으로 알려져 있다. 방법3(M-PS 모형)은 매개변인에 대해서만 경향점수를 추정한 후 경향점수 역확률가중치를 적용하는 방법으로, 이 방법은 처치변인에 대한 사전 특성의 영향력은 교정되지 못할 가능성이 있다.

이러한 두 방법의 한계를 보완하기 위해 처치변인과 매개변인을 동시에 활용하는 방법을 고려할 수 있으며, 방법4(TM-PS 모형)는 처치변인과 매개변인 각각에 대해 경향점수를 추정한 후 역확률가중치를 서로 곱하여 최종 가중치를 산출하여 활용하는 방법이다. 이 방법은 신중휘(2020)의 연구에서 적용된 바 있다. 하지만 해당 연구는 적용 연구였기 때문에, 경향점수 역확률가중치 방법에 대한 방법론적인 검토는 깊이 다루어지지 않았다. 이에 본 연구에서는 처치변인과 매개변인에 대한 경향점수를 모두 추정하여 곱 가중치를 활용한 방법이 처치변인이나 매개변인 하나에 대해서만 경향점수를 추정하여 활용하는 방법에 비해 더 나은 수행을 보이는지를 검증하고, 매개모형에서 경향점수 역확률가중치의 활용방법에 따라 효과 추정치의 정확성에 어떤 차이가 있는지를 확인하고자 한다.



<그림 1> 매개모형에서 경향점수 역확률가중치 활용방법

Ⅲ. 모의실험

매개모형에서 경향점수의 활용방법에 따른 효과 추정치의 정확성을 비교하기 위해 모의실험을 실시하였다. 모의실험에서는 효과의 참값(true effect)을 연구자가 지정할 수 있기 때문에, 해당 모형에서 효과가 정확하게 추정되는지 여부를 검증할 수 있다. 모의실험을 통해 매개모형에서 경향점수의 활용방법에 따라 효과 추정치가 정확성이 어떻게 달라지는지 비교하였다.

1. 모의실험의 자료생성 조건

모의실험의 자료생성 조건으로, 먼저 표본크기는 Kline(2005)을 참고하여 구조방정식모형에서 작은 표본에 해당하는 100명과, Breckler(1990)의 기준을 참고하여 중간 정도인 200명, 표본크기가 충분한 경우인 500명을 조건으로 포함하였다. 처치 변인은 2개의 집단으로 분할된 이분형 변인으로, 매개변인과 결과변인은 정규분포를 따르는 연속형 변인으로 설정하여 생성하였다. 효과의 참 값(true effect)은 0으로 설정하여, 실제 효과가 없을 때 사전 특성에 의해 효과가 얼마나 편향되어 나타나는지를 확인하였다. 통제해야 하는 사전 특성 변인은 관련 선행연구(이진실 2013; 하여진 2015; Coffman 2011)에서의 조건을 참고하여 다변량 정규분포를 바탕으로 4개의 변인(X1, X2, X3, X4)을 생성하였고, 사전 특성이 연구변인에 미치는 영향력은 실제자료의 상관 및 Cohen(1977)의 효과크기 r에 대한 기준을 참고하여 2개의 변인(X1, X2)은 작은 효과크기에 해당하는 .1의 영향력을 가지고, 다른 2개의 변인(X3, X4)은 중간 효과크기에 해당하는 .3의 영향력을 갖는 것으로 설정하였다.

<표 1> 모의실험 자료생성 조건

구분	조건	
표본크기	$n=100, 200, 500$	
효과의 참 값(true effect)	0	
사전 특성	사전 특성 변인의 개수	4개 (X1, X2, X3, X4)
	사전 특성 간 상관	0.2
	연구변인에 미치는 영향력	0.1, 0.3

모의실험 자료를 생성한 절차는 다음과 같다. 사전 특성의 생성은 아래 식과 같은 상관행렬을 바탕으로 다변량 정규분포에서 난수를 발생하여 생성하였다. 사회과학 연구 상황에서 사전 특성들 간에도 상관관계가 있을 수 있으므로 사전 특성들 간 상관을 0.2로 설정하여 자료를 생성하였다.

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{pmatrix} \right)$$

처치변인(T), 매개변인(M), 결과변인(Y)은 사전 특성 변인과의 관계 및 효과의 크기를 반영하여 아래와 같은 방법으로 생성하였다. 처치변인의 경우 이분형 변인이기 때문에 잠재적인 연속변인을 먼저 생성한 후 생성된 자료에서 잠재적인 연속변인의 값이 평균(=0)보다 크면 1의 값을, 잠재적인 연속변인의 값이 평균보다 작으면 0의 값을 가지도록 변환하였다.

$$T^* = 0.1X_1 + 0.1X_2 + 0.3X_3 + 0.3X_4 + e_T, \quad e_T \sim N(0,1)$$

$$T = 1 \quad \text{if } T_1^* \geq 0$$

$$T = 0 \quad \text{if } T_1^* < 0$$

$$M = 0 \cdot T + 0.1X_1 + 0.1X_2 + 0.3X_3 + 0.3X_4 + e_M, \quad e_M \sim N(0,1)$$

$$Y = 0 \cdot T + 0 \cdot M + 0.1X_1 + 0.1X_2 + 0.3X_3 + 0.3X_4 + e_Y, \quad e_Y \sim N(0,1)$$

각 자료에서 표본크기는 조건에 따라 $n=100, 200, 500$ 으로 변화시켜가며 모의실험 자료를 생성하였다. 모의실험은 조건별로 10,000번 반복 시행(replication)하였고, 자료 생성 및 분석에는 R3.4.3.프로그램을 활용하였다.

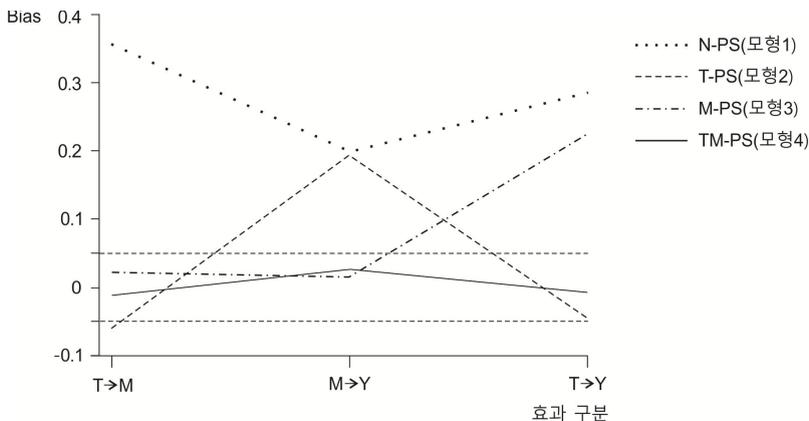
2. 모의실험의 분석모형

모의실험에서는 경향점수 활용방법에 따라 4개의 모형을 분석한 후 비교하였다 (<그림 1> 참고). 모형1은 경향점수를 적용하지 않은 통제 전 모형(N-PS 모형), 모형2는 처치변인에 대한 경향점수만 활용한 모형(T-PS 모형), 모형3은 매개변인에 대한 경향점수만 활용한 모형(M-PS 모형), 모형4는 처치변인과 매개변인 각각에 대해 경향점수를 추정한 후 곱 가중치를 활용한 모형(TM-PS 모형)에 해당한다.

경향점수 추정 시, 처치변인(T)은 이분형 변인으로 생성했으므로 로지스틱 회귀분석을 통해 처치변인에 대한 경향점수를 추정하였고, 매개변인(M)은 연속형 변인이므로 일반화 경향점수를 적용했다. 모형2에서는 처치변인에 대한 경향점수를 기준으로, 모형3에서는 매개변인에 대한 경향점수를 기준으로 안정화 역확률가중치가 적용되었고, 모형4에서는 두 역확률가중치를 서로 곱해 산출된 가중치가 적용되었다.

3. 모의실험 결과

모의실험을 통해 각 연구모형별(모형1~모형4)로 매개모형에서 추정되는 3가지 유형의 개별 효과 추정치(처치변인이 매개변인에 미치는 효과(T→M), 매개변인이 결과변인에 미치는 효과(M→Y), 처치변인이 결과변인에 미치는 직접효과(T→Y))의 정확성을 평가하였다. 효과 추정치의 정확성은 편의(bias)라는 지수를 기준으로 평가하였다. 편의(bias)는 각 시행에서 추정된 효과의 평균이 효과의 참 값과 얼마나 차이가 있는지를 요약하는 지수로, 편의의 값이 0에 가까울수록 효과 추정치의 정확성이 높다고 평가할 수 있다(Hoogland & Boomsma 1998; Mooney 1997). 모의 실험에서 각 시행별 차이도 참고할 수 있도록, 추정치들의 표준편차를 함께 반영하는 지수인 표준화 편의(standardized bias)와 평균제곱오차(MSE)도 <표 2>에 보고하였다.¹⁾



<그림 2> 경향점수 활용방법에 따른 효과 추정치의 편의 비교($n=200$ 일 때)

1) 표준화 편의(standardized bias)는 반복시행에서 얻은 추정치들의 평균과 참 값의 차이(=편의)를 추정치들의 표준편차로 나누어주어 계산된다. 평균제곱오차(MSE)는 편이의 제곱과 추정치들의 분산을 더하여 계산될 수 있다.

<표 2> 경향점수 활용방법에 따른 모의실험 결과

표본 크기	구분	N-PS 모형 (모형1)	T-PS 모형 (모형2)	M-PS 모형 (모형3)	TM-PS 모형 (모형4)	
n=100	T→M	Bias	.355	-.085	.001	-.011
		Std. Bias	1.853	-.382	.005	-.042
		MSE	.134	.009	.010	.009
	M→Y	Bias	.197	.177	.003	.010
		Std. Bias	2.016	1.428	.018	.069
		MSE	.041	.044	.004	.005
	T→Y	Bias	.285	-.067	.196	-.021
		Std. Bias	1.490	-.303	.871	-.082
		MSE	.088	.009	.074	.010
* mean of valid n^2 : 모형1=100, 모형2=89, 모형3=71, 모형4=68						
n=200	T→M	Bias	.355	-.060	.022	-.014
		Std. Bias	2.604	-.402	.150	-.088
		MSE	.145	.026	.023	.026
	M→Y	Bias	.198	.191	.016	.026
		Std. Bias	2.884	2.219	.176	.264
		MSE	.044	.044	.009	.010
	T→Y	Bias	.286	-.046	.226	-.009
		Std. Bias	2.083	-.302	1.520	-.053
		MSE	.101	.025	.073	.029
* mean of valid n : 모형1=200, 모형2=187, 모형3=165, 모형4=161						
n=500	T→M	Bias	.356	-.032	.044	-.012
		Std. Bias	4.159	-.355	.501	-.126
		MSE	.134	.009	.010	.009
	M→Y	Bias	.198	.202	.029	.040
		Std. Bias	4.566	3.819	.526	.683
		MSE	.041	.044	.004	.005
	T→Y	Bias	.284	-.027	.258	-.001
		Std. Bias	3.301	-.298	2.902	-.006
		MSE	.088	.009	.074	.010
* mean of valid n : 모형1=500, 모형2=483, 모형3=452, 모형4=449						

모의실험 결과, 사전 특성을 통제하지 않은 N-PS 모형(모형1)에서는 처치변인이 매개변인에 미치는 효과 추정치($T \rightarrow M$), 매개변인이 결과변인에 미치는 효과 추정치($M \rightarrow Y$), 처치변인이 결과변인에 미치는 직접효과 추정치($T \rightarrow Y$)가 모두 크게 편향되어 추정되었다. N-PS 모형(모형1)의 결과를 통해 매개모형에서 사전 특성을 통제하지 않으면 매개모형에서 각 개별 효과가 부정확하게 추정됨을 알 수 있다.

T-PS 모형(모형2)은 처치변인에 대해서만 경향점수를 추정하여 활용한 모형으로, 처치변인이 매개변인에 미치는 효과 추정치($T \rightarrow M$), 처치변인이 결과변인에 미치는 직접효과 추정치($T \rightarrow Y$)의 편향은 작은 편이었으나, 매개변인이 결과변인에 미치는 효과 추정치($M \rightarrow Y$)의 편향은 크게 나타났다. 이는 이진실(2013)의 결과와 일치하는 결과이다. T-PS 모형(방법2)에서 직접효과는 비교적 정확하게 추정될 수 있으나, 매개변인이 결과변인에 미치는 효과($M \rightarrow Y$)가 편향되므로 $(T \rightarrow Y) \times (M \rightarrow Y)$ 로 계산되는 매개효과의 크기는 부정확해질 수 있다.

M-PS 모형(모형3)은 매개변인에 대해서만 경향점수를 추정하여 역확률가중치를 적용한 모형으로, M-PS 모형(모형3)에서는 처치변인이 매개변인에 미치는 효과 추정치($T \rightarrow M$)와 매개변인이 결과변인에 미치는 효과 추정치($M \rightarrow Y$)의 편향은 0에 가까웠으나, 처치변인이 결과변인에 미치는 직접효과 추정치($T \rightarrow Y$)의 편향은 크게 나타났다. 즉, M-PS 모형(모형3)에서는 매개효과의 크기는 비교적 정확하게 추정될 수 있으나, 직접효과 추정치는 정확하게 추정되지 못하는 것으로 나타났다.

이에 비해 TM-PS 모형(모형4)은 처치변인과 매개변인에 대한 경향점수를 각각 추정한 후 역확률가중치의 곱을 활용한 모형으로, 처치변인이 매개변인에 미치는 효과 추정치($T \rightarrow M$), 매개변인이 결과변인에 미치는 효과 추정치($M \rightarrow Y$), 처치변인이 결과변인에 미치는 직접효과 추정치($T \rightarrow Y$)의 편향이 모두 0에 가깝게 확인되었다. 이를 통해 처치변인과 매개변인에 대한 경향점수를 동시에 활용하는 것이 처치변인 또는 매개변인 하나에 대한 경향점수만 활용하는 것보다 더 나은 수행을 보임을 확인하였다.

2) 모의실험 자료 분석 시 경향점수의 분포가 겹치는 사례, 즉 공통영역에 해당하는 사례만 추출하여 활용하였는데, 제시된 유효 사례 수(mean of valid n)는 각 시행에서 최종 분석에 포함된 사례 수의 평균값에 해당한다.

IV. 실제자료 분석

1. 분석에 활용한 데이터 및 주요 변인 소개

실제자료 적용 예시에서는 ‘한국 아동·청소년 패널조사(Korean Children & Youth Panel Survey: KCYPS)’ 초1패널 자료를 활용하였고, 봉사활동 참여를 처치변인, 공동체의식을 매개변인, 삶의 만족도를 결과변인으로 설정하여 매개모형을 분석하였다. 연구변인(봉사활동 참여, 공동체의식, 삶의 만족도)은 초1패널 7차년도 자료를 활용하였고 이는 조사대상이 중학교 1학년 시기에 응답한 자료에 해당한다. 사전 특성으로는 성별, 부모의 교육수준, 가정의 경제적 수준, 부모의 양육태도, 학업 성적, 종교, 사전 공동체의식, 사전 삶의 만족도를 통제하고자 하였고, 사전 특성에 대한 정보는 처치에 참여하기 전에 조사된 4~6차년도 자료를 활용하였다. 분석에 활용한 변인의 기술통계치는 다음과 같다.

<표 3> 분석에 활용한 변인의 기술통계치(n=1,841)

	구분	평균	표준편차	최소값	최대값
독립변인	봉사활동 참여(비참여=0, 참여=1)	.56	.50	.00	1.00
매개변인	공동체의식	3.16	.55	1.25	4.00
결과변인	삶의 만족도	3.23	.61	1.00	4.00
	성별(남=0, 여=1)	.49	.50	.00	1.00
	부모교육연수	13.91	1.78	9.00	18.00
	로그가구소득	8.44	.47	5.70	1.31
	부모의 양육태도_애정	3.35	.60	1.00	4.00
	부모의 양육태도_합리적 설명	3.16	.67	1.00	4.00
사전 특성	부모의 양육태도_돌봄	3.41	.52	1.00	4.00
	부모의 양육태도_학대	1.55	.56	1.00	4.00
	학업성적	3.82	.69	1.00	5.00
	종교 유무(없음=0, 있음=1)	.50	.50	.00	1.00
	사전 공동체의식	3.09	.60	1.00	4.00
	사전 삶의 만족도	3.36	.61	1.00	4.00

2. 경향점수 역확률가중치 추정

1) 경향점수 추정 및 공통영역 확인

치치변인인 봉사활동 참여 여부는 이분형 범주변인이므로, 아래와 같은 로지스틱 회귀분석을 실시한 후 각 사례별로 사전 특성을 고려했을 때 봉사활동에 참여할 예측 확률($\hat{P}(\text{봉사활동 참여} = 1 | X)$)을 저장하여 경향점수를 추정하였다.

$$\ln \frac{\hat{P}(\text{봉사활동 참여} = 1)}{\hat{P}(\text{봉사활동 참여} = 0)} = \beta_0 + \beta_1(\text{성별}) + \dots + \beta_{11}(\text{사전 삶의 만족도})$$

매개변인에 해당하는 공동체의식은 연속형 변인이므로 일반화 경향점수를 추정하였다. 사전 특성을 독립변인으로, 공동체의식을 종속변인으로 하는 회귀분석을 실시한 후, 정규분포의 확률밀도함수를 활용하여 사전 특성들을 고려했을 때 각 사례의 공동체의식 점수(회귀분석에서의 공동체의식의 예측 점수)가 관찰될 확률을 계산하여 경향점수를 추정하였다.

$$\widehat{\text{공동체의식}} = \beta_0 + \beta_1(\text{성별}) + \dots + \beta_{11}(\text{사전 삶의 만족도})$$

$$\hat{p}(\text{공동체의식} = m | X) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} e^{-\frac{(m - X\hat{\beta})^2}{2\hat{\sigma}^2}}$$

<경향점수 추정을 위한 R코드>

```
# 치치변인(이분형 변인)에 대한 경향점수 추정
t_ps_model=glm(치치변인~사전 특성1+사전 특성2+..., data=데이터명, family="binomial")
데이터명$t_ps<-predict(t_ps_model, data=데이터명 type="response")

# 매개변인(연속형 변인)에 대한 경향점수 추정
m_ps_model=lm(매개변인~사전 특성1+사전 특성2+..., data=데이터명)
데이터명$m_pred<-fitted(m_ps_model)
데이터명$m_ps<-dnorm(데이터명$매개변인, 데이터명$m_pred, sd(데이터명$매개변인))
```

경향점수를 추정된 후에는, 경향점수가 서로 겹치는 공통영역을 확인하고, 공통 영역에 포함되는 사례만을 추출하였다. 이분 처치변인의 경우에는 처치집단과 통제 집단의 경향점수 범위를 살펴봄으로써 공통영역을 확인할 수 있다. <표 4>를 통해 봉사활동 참여(처치변인)에 대한 경향점수의 분포를 살펴보면, 비참여 집단에서 경향점수는 .299~.730의 범위를 보였고, 참여 집단에서 경향점수는 .322~.741의 범위를 보였다. 따라서 경향점수의 범위가 겹치는 .322~.730에 해당하는 사례만 추출하였다. 9명의 사례가 공통영역을 벗어나는 것으로 확인되었고, 공통영역에 포함되는 사례 수는 1,832명이었다.

<표 4> 처치변인에 대한 경향점수의 공통영역 확인

집단구분	최소값	최대값	평균	표준편차
비참여 집단(n=810)	.299	.730	.548	.072
참여 집단(n=1,031)	.322	.741	.569	.071
전체(n=1,841)	.299	.741	.560	.072

다음으로 매개변인에 대한 공통영역도 확인하였다. 연속형 변인의 경우에는 Kluge et al.(2012)이 제안한 방법을 적용할 수 있다. 이를 적용하여 공동체의식의 30백분위, 70백분위를 기준으로 3개 집단(집단1=30백분위 미만, 집단2=30백분위 이상 70백분위 미만, 집단3=70백분위 이상)을 구분하였고, 각 집단의 중앙값에서의 일반화 경향점수의 범위를 비교하였다. 집단1의 중앙값(2.5점으로 확인)에 대한 일반화 경향점수(GPS1), 집단2의 중앙값(3.0점)에 대한 일반화 경향점수(GPS2), 집단3의 중앙값(4.0점)에 대한 일반화 경향점수(GPS3)를 각각 구하여, 집단별로 GPS1, GPS2, GPS3의 범위를 비교한 결과는 <표 5>와 같다. GPS1의 공통영역에 해당하는 범위는 .104~.723, GPS2의 공통영역에 해당하는 범위는 .414~.726, GPS3의 공통영역에 해당하는 범위는 .017~.544로 확인되었다. 이를 모두 충족하는 사례만 추출하였을 때, 1,832명 중에서 20명의 사례가 공동체의식의 경향점수에 대해 공통영역을 벗어나는 것으로 확인되었고, 최종적으로 공통영역에 포함되는 사례 수는 1,812명이었다.

<표 5> 매개변인에 대한 경향점수의 공통영역 확인

구분		최소값	최대값	평균	표준편차
GPS1	집단1(n=519)	.104	.726	.441	.147
	집단2(n=742)	.097	.726	.375	.141
	집단3(n=571)	.097	.723	.284	.139
GPS2	집단1(n=519)	.414	.726	.676	.147
	집단2(n=742)	.398	.726	.663	.141
	집단3(n=571)	.399	.726	.607	.139
GPS3	집단1(n=519)	.012	.544	.182	.113
	집단2(n=742)	.017	.559	.231	.121
	집단3(n=571)	.014	.558	.318	.135

2) 역확률가중치 계산 및 사전 특성의 균형 검토

공통영역 가정을 모두 만족하는 1,812명의 사례에 대해, 추정된 경향점수를 활용하여 아래의 공식을 통해 각 사례에 대한 가중치를 계산하였다.

$$w = \frac{\hat{p}(T=t)}{\hat{p}(T=t|X)} \times \frac{\hat{p}(M=m)}{\hat{p}(M=m|X)}$$

위의 식에서 T는 봉사활동 참여, M은 공동체의식, X는 사전 특성을 나타내는 기호로 사용되었다. 위의 식을 적용하여 처치변인에 대한 안정화 역확률가중치와 매개변인에 대한 안정화 역확률가중치를 곱한 최종 가중치를 산출하였다. 이러한 방법으로 최종 가중치를 구했을 때 가중치의 값이 10을 넘어서는 극단적인 가중치를 보이는 사례는 없었다.

<역확률가중치 계산을 위한 R코드>

```
# 처치변인(이분형 변인(참여=1, 비참여=0))에 대한 안정화 역확률가중치 계산
##는 처치변인을 의미, data는 데이터명을 의미, ps_t는 처치변인에 대한 경향점수를 의미.
data$w_t=data$t*(mean(data$t)/data$ps_t)+(1-data$t)*((1-mean(data$t))/(1-data$ps_t))
```

```
# 매개변인(연속형 변인)에 대한 안정화 역확률가중치 계산
#m은 매개변인을 의미, ps_m은 매개변인에 대한 경향점수를 의미.
data$m_ps = dnorm(data$m, mean(data$m), sd(data$m))/data$m

# 두 가중치의 곱을 통해 최종 가중치 계산
data$tm = data$t * data$m
```

그런 다음 경향점수 역확률가중치를 적용했을 때 사전 특성의 균형이 잘 확보되는지 밸런스 테스트를 통해 검토하였다. 이를 위해 처치변인 또는 매개변인에 따라 각 사전 특성에 차이가 있는지를 일변량 가중회귀분석을 통해 가중치 적용 전·후를 비교하였다. <표 6>의 왼쪽에는 처치변인(봉사활동 참여)에 대한 밸런스 테스트 결과가, <표 6>의 오른쪽에는 매개변인(공동체의식)에 대한 밸런스 테스트 결과가 제시되어 있다.

<표 6> 경향점수 적용 전·후 사전 특성의 균형 비교

사전 특성	봉사활동 참여에 대한 사전 특성의 균형 비교				공동체의식에 대한 사전 특성의 균형 비교			
	가중치 적용 전		가중치 적용 후		가중치 적용 전		가중치 적용 후	
	β	p	β	p	β	p	β	p
성별	.053	.024	.011	.631	.025	.281	.012	.641
부모교육년수	.071	.002	-.007	.758	.043	.064	.016	.508
로그가구소득	.047	.043	-.006	.814	.060	.009	.018	.468
부모의 양육태도_애정	.049	.034	-.015	.529	.200	.000	.046	.071
부모의 양육태도_합리적 설명	.048	.039	-.024	.304	.163	.000	.152	.545
부모의 양육태도_돌봄	.063	.006	.002	.924	.162	.000	.057	.022
부모의 양육태도_학대	-.058	.012	-.010	.666	-.079	.001	-.022	.386
학업성적	.093	.000	-.014	.547	.180	.000	.022	.367
종교 유무	.016	.498	-.002	.949	.046	.050	.011	.668
사전 공동체의식	.096	.000	-.048	.041	.385	.000	.026	.306
사전 삶의 만족도	.055	.019	-.015	.544	.199	.000	.038	.136

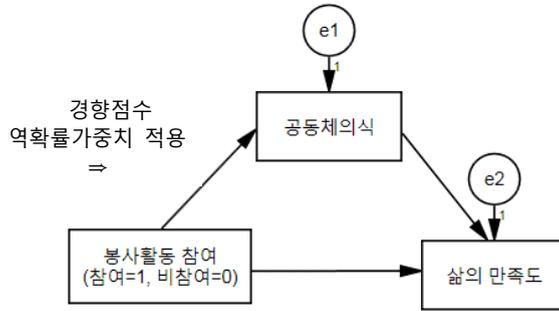
처치변인에 대한 밸런스 테스트를 실시하였을 때, 경향점수 역확률가중치를 적용하기 전에는 봉사활동 참여 집단과 비참여 집단 간에 성별 분포, 부모교육년수, 로그가구소득, 부모의 애정, 부모의 합리적 설명, 부모의 돌봄, 부모의 학대, 학업성적, 사전 공동체의식, 사전 삶의 만족도에 있어 통계적으로 유의한 차이가 있었으나, 경향점수 역확률가중치를 적용한 후에는 두 집단 간 사전 특성 차이가 대부분 통계적으로 유의하지 않은 것을 확인할 수 있었다.³⁾ 경향점수를 적용하기 전에는 봉사활동 참여 집단은 참여하지 않은 집단에 비해 여학생, 부모교육년수가 높은 학생, 가구소득이 높은 학생, 부모의 양육방식에서 애정, 합리적 설명, 돌봄이 높고 학대가 낮은 가정의 학생, 학업성적이 높은 학생, 초등학교 시기 공동체의식이 높은 학생, 초등학교 시기 삶의 만족도가 높은 학생이 더 많이 분포되어 있었으나, 경향점수를 적용한 후에는 이러한 차이가 조정되어 집단 간 차이가 크게 줄어들거나 통계적으로 유의하지 않았다.

매개변인인 공동체의식에 대해서도 사전 특성 균형이 잘 이루어졌는지 확인했을 때, 경향점수 역확률가중치를 적용하기 전에는 공동체의식 수준에 따라 로그가구소득, 부모의 애정, 부모의 합리적 설명, 부모의 돌봄, 부모의 학대, 학업성적, 사전 공동체의식, 사전 삶의 만족도에 통계적으로 유의한 차이가 있었으나, 경향점수 역확률가중치를 적용한 후에는 통계적으로 유의하지 않거나 표준화 계수의 크기가 크게 줄어드는 것을 확인할 수 있었다.

3. 경향점수 역확률가중치를 적용하여 매개모형 분석

최종적으로, 경향점수 역확률가중치를 적용하여 <그림 3>의 매개모형⁴⁾을 분석하였다.

-
- 3) 가중치를 적용한 후에도 사전 공동체의식의 경우에는 집단 간 차이가 통계적으로는 유의하였으나, Leite(2016)는 사전 특성의 균형을 검토할 때 가설검증보다 표준화 계수에 주목할 것을 제안한 바 있다. 사전 공동체의식에 대한 표준화 계수의 크기가 가중치 적용 전보다 가중치 적용 후 크게 줄어들었고, 표준화 계수의 크기가 .1보다 작으며 .05보다도 작기 때문에 공변인 균형에 크게 문제가 없는 것으로 판단할 수 있다.
- 4) 구조방정식모형에서 매개모형은 <그림 3>과 같이 측정변인들로만 구성된 경로분석(path analysis)의 형태로도 표현될 수 있고, 잠재변인이 포함된 형태로도 설정될 수 있다. 이 연구에서는 보다 간단한 경로분석을 적용하여 매개모형을 분석하였으며, 잠재변인이 포함된 더 복잡한 모형으로의 확장은 추후 연구과제로 다루고자 한다.



<그림 3> 실제자료 분석의 매개모형

```

<가중치를 적용하여 매개모형 분석을 위한 R코드>

library("lavaan.survey")
model <- 'y~t+m
          m~t'
result<-lavaan.survey(sem(model, data=data), svydesign(id=~ID, weights=~w_tm, data=data))
# w_tm은 가중치에 해당하는 변수명을 의미.
summary(result, fit.measures=T, standardized=T)
    
```

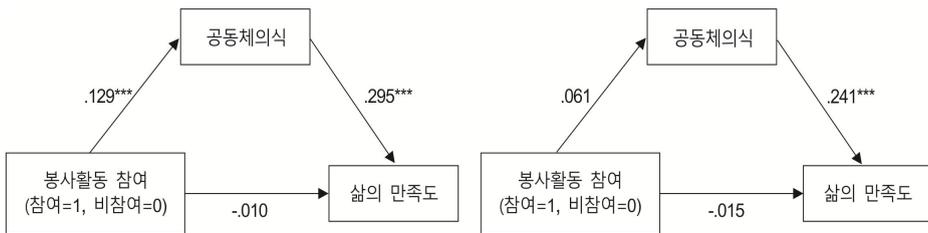
<그림 3>의 매개모형을 분석한 결과는 <표 7>에 제시하였고, 비교를 위해 경향점수 역확률가중치를 적용하기 전의 결과도 함께 제시하였다. 경향점수 역확률가중치를 적용하기 전과 후를 비교해보면, 경향점수 역확률가중치를 적용한 후에 효과 추정치의 크기가 줄어들거나 통계적 유의성이 사라지는 것을 확인할 수 있다. 특히, 봉사활동 참여가 공동체의식에 미치는 효과의 경우, 경향점수 역확률가중치를 적용하기 전에는 해당 효과가 통계적으로 유의하였으나, 경향점수 역확률가중치를 적용한 후에는 통계적으로 유의하지 않은 것으로 확인되었다. 즉, 경향점수 역확률가중치를 적용하기 전 결과에서는 봉사활동에 참여한 학생들이 공동체의식이 높아지는 것으로 나타났지만, 이는 사전 특성의 영향력이 혼재되었기 때문에 나타난 결과임을 알 수 있다. 경향점수 역확률가중치를 통해 사전 특성의 영향력을 통제해준 후에는 봉사활동에 참여하더라도 공동체의식이 통계적으로 유의하게 높아지지 않는

것으로 나타났고, 사전 특성을 통제하지 않으면 봉사활동 참여가 공동체의식에 미치는 효과가 과대추정될 수 있음을 <표 7>의 결과를 통해 알 수 있다. 공동체의식이 삶의 만족도에 미치는 효과의 경우에는 경향점수 역확률가중치 적용 전과 후 모두 해당 효과가 통계적으로 유의하였으나, 경향점수 역확률가중치를 적용했을 때 효과의 크기가 줄어드는 것으로 확인되었다. 봉사활동 참여가 삶의 만족도에 미치는 직접효과의 경우에는 경향점수 역확률가중치 적용 전과 후 모두 통계적으로 유의하지 않게 나타났다.

<표 7> 경향점수 적용 전·후 매개모형 분석결과

효과 구분	경향점수 역확률가중치 적용 전			경향점수 역확률가중치 적용 후(연구모형)		
	B	SE	β	B	SE	β
봉사활동 참여 → 공동체의식 (T→M)	.143***	.026	.129	.065	.035	.061
공동체의식 → 삶의 만족도 (M→Y)	.327***	.025	.295	.273***	.031	.241
봉사활동 참여 → 삶의 만족도 (T→Y)	-.012	.028	-.010	-.018	.031	-.015

* $p < .05$, ** $p < .01$, *** $p < .001$.



<그림 4> 경향점수 적용 전·후 매개모형 분석결과(표준화 계수 기준)

V. 결론

이 연구에서는 조사연구에서 많이 활용되는 매개모형에서 경향점수를 활용하여 사전 특성을 통제할 수 있는 방법을 비교하고 검증하였다. 이를 위해 모의실험 및 실제자료 분석을 실시하였다. 모의실험에서는 경향점수 활용방법에 따라 4가지 모

형을 설정하여 효과 추정치의 정확성을 확인하였다. 경향점수를 적용하지 않은 모형(N-PS 모형), 처치변인에 대한 경향점수만 활용하는 모형(T-PS 모형), 매개변인에 대한 경향점수만 활용하는 모형(M-PS 모형), 처치변인 및 매개변인에 대한 경향점수를 각각 추정 후 곱하는 방식을 통해 최종 가중치를 산출하여 활용하는 모형(TM-PS 모형)을 비교하였고, 어떤 방법에서 효과 추정치의 편의가 가장 작은지를 평가하였다. 실제자료 분석에서는 모의실험 결과 가장 나은 수행을 보인 방법을 적용하여 실제 데이터를 활용하여 매개모형을 분석하였다. 실제자료 분석을 통해 경향점수 역확률가중치를 적용했을 때 사전 특성의 균형이 실제로 잘 확보되는지를 검증하였고, 경향점수 역확률가중치를 적용하기 전과 후 매개모형 분석결과가 얼마나 차이가 있는지를 비교하였다.

연구의 결과, 모의실험을 통해 경향점수 활용방법에 따른 효과 추정치의 정확성을 검증하였을 때 경향점수를 적용하지 않은 모형에서는 매개모형에서 추정되는 3가지 유형의 개별 효과 추정치(처치변인이 매개변인에 미치는 효과, 매개변인이 결과변인에 미치는 효과, 처치변인이 결과변인에 미치는 직접효과)에서 모두 편의가 크게 발생하였다. 이를 통해 사전 특성을 통제하지 않으면 매개모형에서 효과가 정확하게 추정되지 못한다는 점을 알 수 있다.

처치변인에 대한 경향점수만 활용하는 방법에서는 ‘매개변인이 결과변인에 미치는 효과’ 추정치의 편의가 크게 나타났다. 처치변인에 대한 경향점수만 활용하는 방법은 처치변인에 대해서만 사전 특성의 차이를 조정해주기 때문에 매개변인에 대해서는 사전 특성의 차이를 조정해주지 못한다. 때문에 처치변인에서 출발하는 화살표에 해당하는 경로($T \rightarrow M$, $T \rightarrow Y$)에서는 편의가 발생하지 않지만, 다른 경로($M \rightarrow Y$)는 부정확하게 추정되는 것으로 볼 수 있다.

반면 매개변인에 대한 경향점수만 활용하는 방법에서는 ‘처치변인이 결과변인에 미치는 직접효과’ 추정치에서 편의가 크게 나타났다. 매개변인에 대한 경향점수만 활용하는 방법에서는 사전 특성의 차이를 매개변인에 대해서만 조정해주었기 때문에 매개변인에서 출발하거나 매개변인으로 도착하는 경로($M \rightarrow Y$, $T \rightarrow M$)의 효과 추정치는 비교적 정확하게 추정되지만, 다른 경로($T \rightarrow Y$)에서는 편의가 발생하는 것으로 볼 수 있다.

처치변인과 매개변인에 대한 경향점수를 모두 추정하여 곱해서 활용하는 방법에서는 모든 효과 추정치(처치변인이 매개변인에 미치는 효과, 매개변인이 결과변인에 미치는 효과, 처치변인이 결과변인에 미치는 직접효과)에서 편의가 0에 가까운

것으로 확인되었다. 이를 통해 처치변인과 매개변인에 대한 경향점수를 곱해서 활용하는 방법에서는 처치변인 및 매개변인 모두에 대해서 사전 특성의 차이가 비교적 잘 조정됨을 알 수 있다. 매개모형에서 경향점수를 활용할 때 처치변인 또는 매개변인 하나만 기준으로 경향점수를 추정하여 활용하는 방법보다, 처치변인과 매개변인에 대한 경향점수를 모두 추정한 후 활용하는 것이 더 나은 수행을 보임을 확인할 수 있었다.

실제자료를 활용하여 봉사활동 참여, 공동체의식, 삶의 만족도 간의 관계를 분석하였을 때에도, 처치변인(봉사활동 참여)에 대한 경향점수와 매개변인(공동체의식)에 대한 경향점수를 추정하여 가중치를 곱해서 활용하는 방법을 적용하였을 때 사전 특성의 균형이 잘 확보되는 것으로 나타났다. 경향점수 역확률가중치를 적용하기 전에는 봉사활동 참여여부 또는 공동체의식에 따라서 사전 특성에 유의미한 차이가 있었는데, 경향점수 역확률가중치를 적용한 이후에는 사전 특성의 차이가 크게 줄어들거나 사라지는 것으로 나타났다. 또한 매개모형에서의 개별 효과 추정치(경로계수)의 크기도 경향점수를 적용하기 전과 후의 결과가 차이를 보였고, 특히 봉사활동 참여가 공동체의식에 미치는 효과의 경우 사전 특성을 통제하지 않으면 크게 과대추정되는 것으로 나타나, 사전 특성 통제가 매우 중요함을 확인할 수 있었다.

이 연구는 매개모형에서 사전 특성을 통제하기 위한 방법으로 경향점수 활용방법을 비교하고 이를 소개했다는 점에서 의의가 있다. 매개모형을 분석한 기존의 선행 연구들에서는 사전 특성 통제가 중요함에도 불구하고 모형의 간명성을 유지하기 위해 사전 특성을 충분히 통제하지 못한 채 자료를 분석한 경우가 많았다. 본 연구에서는 매개모형에서 경향점수의 활용 가능성을 확인하고, 구체적으로 경향점수를 어떻게 활용할 수 있을지를 제시했다는 점에서 의의가 있다. 본 연구의 결과는 조사연구 자료를 분석하는 연구자들에게 적절한 연구방법을 안내하는 자료가 될 수 있을 것이라 기대된다.

연구의 제한점 및 후속연구를 위한 제언으로는, 이 연구에서는 매개변인의 수가 1개인 비교적 단순한 매개모형에 경향점수를 적용하였지만, 매개변인의 수가 2개 이상인 다중매개모형에서도 해당 방법이 좋은 수행을 보이는지 확인할 필요가 있다. 여러 변인에 대한 경향점수 역확률가중치를 곱해서 활용하는 방법은 매개변인의 수가 늘어나면 가중치가 여러 번 곱해지기 때문에 사전 특성이 동시에 잘 조정되지 못할 가능성도 존재한다. 추후 연구에서는 매개모형이 보다 복잡한 형태로 확

장되었을 때에도 경향점수를 활용한 방법이 효과 추정치의 편의를 줄여줄 수 있는지 추가로 더 검증할 필요가 있다. 또한 이 연구에서는 일부 조건에서만 모의실험을 실시하였는데 다른 조건의 연구 상황에서는 다른 결과를 보일 가능성도 있다. 후속 연구에서는 모의실험 조건을 더 다양하게 설정하여, 보다 다양한 조건(효과의 크기, 공변인 조건, 처치-통제집단의 비율, 매개변인 특성, 결과변인의 분포, 결측치 조건 등) 아래에서 결과를 재확인하는 것이 필요할 것이라 생각된다.

마지막으로 이 연구에서는 경향점수 활용방법을 비교할 때 ‘효과 추정치’의 정확성을 기준으로 비교하였다. 하지만 <표 7>의 결과를 자세히 살펴보면 가중치 적용 후의 표준오차 값이 약간 더 커지는 경향을 보였는데, 후속 연구에서는 이에 대한 추가적인 검토도 필요할 것으로 생각된다. 경향점수 방법에서는 공통영역을 벗어나는 사례를 제외하여 분석하기 때문에 사례 수가 줄어들면서 나타나는 현상일 수 있으나, 경향점수 가중치를 적용한 방법에서 표준오차가 과대 추정될 가능성도 있을 수 있다. 이에 추후 연구에서는 추정치의 정확성뿐만 아니라 ‘표준오차’의 정확성이 적용방법에 따라 달라지는지도 다양한 조건에서 추가적으로 확인해 볼 필요가 있을 것이다.

참고문헌

- 김선숙·안재진. 2012. “청소년 자원봉사활동이 공동체 의식에 미치는 영향.” 《청소년복지연구》 19(1): 339-363.
- 김육·송미영. 2006. “대학생의 자원봉사 참여도에 영향을 미치는 요인에 관한 연구.” 《청소년학연구》 13(6): 330-360.
- 김준엽·정혜경. M.H. Seltzer. 2008. “Drawing Causal Inferences Using Propensity Score Methods in Educational Research.” 《교육평가연구》 21(3): 219-242.
- 김지혜. 2012. “청소년의 봉사활동이 자아존중감과 자아탄력성을 매개로 공동체 의식과 삶의 만족도에 미치는 영향: 봉사활동 시간과 주관적 만족을 중심으로.” 《청소년복지연구》 14(1): 41-62.
- 노충래·김소연. 2016. “부모양육태도가 청소년의 스마트폰 의존에 미치는 영향.” 《한국청소년연구》 27(4): 87-114.
- 박재숙. 2011. “초기 청소년의 온, 오프라인 참여활동이 삶의 만족도에 미치는 영향: 공동

- 체의식의 매개효과를 중심으로.” 《한국청소년복지학회지》 13(4): 47-69.
- 박현정·이준호. 2012. “방과후학교의 연속적 처치(sequential treatment) 효과 분석: 고1~고2 시기 학업성취도를 중심으로” 제4회 국가수준 학업성취도 평가와 대학수학능력시험 자료 활용 분석 심포지엄 자료집, 248-257.
- 박현정·이진실·이용석. 2015. “자발적 봉사활동 참여가 공동체 의식에 미치는 효과 분석.” 《교육행정학연구》 33(2): 235-257.
- 변상민·신중휘·양준혁. 2018. “사교육 참여 여부 및 시간이 중학생의 학업스트레스에 미치는 영향.” 《아시아교육연구》 19(4): 913-944.
- 신중휘. 2020. “중학생의 방과 후 수학 학습시간과 수학 성취도의 관계 분석: 일반화경향 점수 및 종단적 구조모형의 적용.” 서울대학교 석사학위논문.
- 오혜정·손병덕·임희수. 2018. “청소년의 봉사활동 자발성, 만족도, 참여 어려움에 영향을 미치는 요인.” 《청소년학연구》 25(12): 365-387.
- 이준호·박현정. 2012. “맞벌이의 교육적 기회비용.” 《아시아교육연구》 13(3): 27-59.
- 이진실. 2013. “사교육 효과 추정을 위한 구조방정식모형에서 공변인 통제방법 비교” 서울대학교 석사학위논문.
- 조복순·안상근. 2017. “청소년이 지각한 부모 양육태도와 스트레스 대처방식과의 관계에서 자아탄력성의 매개효과.” 《한국청소년연구》 28(4): 225-253.
- 최영창·장연심. 2008. “청소년 자원봉사활동의 지속적 참여 및 추천의도에 영향을 미치는 요인.” 《한국청소년연구》 19(2): 87-109.
- 하여진. 2015. 교육연구에서 다중처치의 선택편의를 조정하기 위한 경향점수의 활용. 서울대학교 박사학위논문.
- 하여진. 2016. “직무일치가 직무만족을 매개로 이직의도에 미치는 효과.” 《한국교육재정경제학회》 25(2): 29-52.
- 홍나미. 2015. “청소년 봉사활동이 학교적응에 미치는 영향에서 자아존중감과 공동체의식의 매개효과.” 《청소년 문화포럼》 44: 157-180.
- 홍은진. 2006. “자원봉사의 참여에 영향을 미치는 요인에 관한 연구.” 《한국비영리연구》 5(1): 33-71.
- Breckler, S.J. 1990. “Application of Covariance Structure Modeling in Psychology: Cause for Concern?” *Psychological Bulletin* 107: 260-273.
- Coffman, D.L. 2011. “Estimating Causal Effects in Mediation Analysis Using Propensity Scores.” *Structural Equation Modeling* 18: 357-369.
- Cohen, J. 1977. *Statistical Power Analysis for the Behavioral Sciences*. (2nd ed.). NY: Academic Press.
- Cole, S.R. and M.A. Heman. 2008. Constructing Inverse Probability Weight for Marginal Structural Models. *American Journal of Epidemiology* 168(6): 656-664.

- Curtis, L.H., B.G. Hammill, E.L. Eisenstein, J.M. Kramer, and K.J. Anstrom, 2007. "Using Inverse Probability-weighted Estimators in Comparative Effectiveness Analyses with Observational Databases." *Medical Care* 45(10): 103-107.
- Hirano, K. and G.W. Imbens. 2004. "The Propensity Score with Continuous Treatments." *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspective* 226164, 73-84.
- Hoogland, J.J. and A. Boomsma. 1998. "Robustness Studies in Covariance Structure Modeling." *Sociological Methods & Research* 26(3): 329-367.
- Kline, R.B. 2005. *Principles and Practice of Structural Equation Modeling*(2nd ed.). NY: Guilford Press.
- Kluve, J., H. Schneider, A. Uhlendorff, and Z. Zhao. 2012. "Evaluating Continuous Training Programmes by Using the Generalized Propensity Score." *Journal of the Royal Statistical Society: Series A(Statistics in Society)* 175(2): 587-617.
- Leite, W. 2016. *Practical Propensity Score Methods Using R*. L.A.: Sage Publications.
- Mooney, C.Z. 1997. *Monte Carlo Simulation*. LA: Sage Publications.
- Murnane, R.J. and J.B. Willett. 2011. *Methods Matter*. NY: Oxford University Press.
- Robins, J.M., S. Greenland, and F.C. Hu. 1999. "Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of Repeated Binary Outcome." *Journal of the American Statistical Association* 94: 687-700.
- Robins, J.M., M.A. Hernan, and B. Brumback. 2000. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 11: 550-560.
- Rosenbaum, P.R. and D.B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70: 41-55.
- Xu, S., C. Ross, M.A. Raebel, S. Shetterly, C. Blanchette, and D. Smith. 2010. "Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals." *Value in Health* 13(2): 273-277.

<접수 2020.04.17; 수정 2020.06.23; 게재확정 2020.06.29>

Methods for Controlling Confounders Using Propensity Score in Mediation Analysis*

Jinsil Lee

The purpose of this study was to compare propensity score methods for controlling confounders in mediation analysis. To do this, simulation and real data analysis was conducted. In simulation study, four models (N-PS model, T-PS model, M-PS model, TM-PS model) were compared. The results showed that when N-PS model was applied, estimates of effects were biased. In T-PS model, estimates of effects of mediator on outcome were biased. In M-PS model, estimates of effects of treatment on outcome were biased. When TM-PS model was applied, estimates of effect were unbiased. In real data analysis, propensity score weighting estimated through TM-PS model showed good performance in terms of balancing confounders. Based on this results, importance of controlling confounders and how to use propensity score in mediation analysis were suggested.

Key words: mediation analysis, propensity score, inverse probability weighting, controlling confounders

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2019S1A5B5A07109438).