

성향 점수를 이용한 퍼지 매칭 방법: IBM SPSS 22 Ver.[†]

김소연¹ · 백종일²

¹²원광대학교 자연과학대학 수학교육통계학부

접수 2015년 12월 1일, 수정 2016년 1월 8일, 게재확정 2016년 1월 18일

요약

성향점수 매칭이란 선택편의가 존재 할 수 있는 두 집단의 데이터를 성향 점수로 매칭하여 비슷한 성향을 갖는 데이터를 추출하는 방법이다. 본 논문은 그 중 하나인 퍼지 매칭 방법을 제시하였다. 성향 점수를 만들기 위해 통제변수를 선정하는 방법과 로지스틱 회귀분석을 통하여 성향 점수를 구하는 방법을 제시하였으며, 이 점수로 퍼지 매칭을 통해 성향이 비슷한 실험 집단과 통제 집단을 추출할 수 있었다. 본 논문에서는 허용오차 범위를 달리하여 분포도와 표준화 차이를 통해 두 집단이 동일한 집단임을 증명했으며, 허용오차 범위 점수가 작아질수록 선택되어 지는 케이스 수도 작아지는 것을 확인할 수 있었다.

주요용어: 성향점수, 청소년 건강행태 온라인조사, 퍼지매칭.

1. 머리말

관찰연구 중에서 두 군을 비교하는 연구는 대부분 비무작위 표본 추출을 기준으로 진행되어 왔다. 하지만 비무작위 표본 추출은 선택편의를 통제할 수 없는 문제점을 가지고 있고, 결과의 인과관계에 대한 올바르게 못한 추론을 하게 하거나 결과를 과소 혹은 과대 추정하는 오류를 발생시킨다. Richard (1983)는 선택 편의가 입증되기 전까지 모든 사회학적 연구에 선택 편의가 존재하는 것으로 간주하고 이를 수정하기 위해 노력해야 한다고 설명하고 있다. 표본 추출 시 사전에 실험 집단과 통제 집단을 무작위로 배정해야 하는데 현실에서는 무작위로 배정하기 매우 어렵다. 따라서 설계 단계에서 혼란변수를 통제할 수 있는 변수를 선정하고, 이 후 통제된 혼란변수를 일정 공변량으로 산출한 후 이를 이용하여 실험 집단과 통제 집단을 유사하게 만드는 방법이 필요하다.

Paul과 Donald (1983)은 성향점수를 이용하여 평균 처리 효과를 제안하였고, 그들의 논문에서 관찰 자료의 인과관계를 보다 명확히 계산하기 위한 방법으로 성향 점수라는 용어를 처음 소개 하였다. 이 방법은 생물통계에서 처음 소개되었으며, 관측된 공변량 벡터가 주어졌을 때 특정한 처리에 배정될 조건부 확률로 정의하고 있다. 성향 점수는 실험 집단과 통제 집단간에 비교하고자 하는 처리변수 이외의 유사하지 않게 분포된 변수들을 통제하기 위해 사용되는 기법 중 하나로, 조건부 확률에 대한 예측은 모형화된 결과로부터 성향 점수를 추정함으로써 구할 수 있고, 준-랜덤화 실험과 비슷한 효과를 갖게 한다.

성향 점수 매칭방법을 적용한 사례를 보면, Park과 Kim (2012)은 실제 정책 효과 (국민기초생활보장제도의 효과분석)를 분석하는 데에 실험집단과 유사한 비교집단을 설정하고, STAT 프로그램을 통

[†] 이 논문은 2015년도 원광대학교의 박사학위 논문에서 발췌하였음.

¹ (54538) 전북 익산시 익산대로 460, 원광대학교 수학교육통계학부, 시간강사.

² 교신저자: (54538) 전북 익산시 익산대로 460, 원광대학교 수학교육통계학부, 교수.

E-mail: jibaek@wku.ac.kr

해 두 집단 간에 매칭 등의 방법을 통해 가능한 유사하게 구성한 뒤, 정책의 효과를 추정하였다. Kim (2011)은 국민건강영양조사 데이터를 가지고 SAS 프로그램을 통해 실험 집단과 통제 집단을 성향 점수를 통해 동일한 집단으로 만든 후 고혈압전기 위험 요인을 알아보았다. Jeong (2011)는 전향적 우울증 환자 코호트 자료를 이용하여 공변량 정보에 대한 SAS 프로그램을 통해 스칼라의 요약치를 제공하는 성향 점수를 이용하여 관찰 변수들이 동질적이라는 전제하에 우울증 환자가 인지한 사회적 지지는 급성기 우울증 치료 반응과 결과에 긍정적인 영향을 미친다는 연구에 사용하였다. Lee 등 (2007)은 당뇨환자를 대상으로 생활습관 개선 프로그램을 실시한 후 그 효과를 평가하기 위해 STAT 프로그램을 이용하여 선택편향을 줄이고 그 결과를 비교하는데 사용하였다. Gabriel 등 (2008)은 R 프로그램을 이용하여 만성 폐쇄성 폐질환 환자 중 양측을 폐 이식 받은 환자와 한쪽만 폐 이식을 받은 환자의 생존률에 관한 분석에서 가능한 선택편향을 줄이기 위해 성향 점수 대응을 이용하였다, 그 결과 양측을 이식한 환자가 한쪽을 이식한 환자보다 생존률이 더 높게 나타났다. William (2006)은 관찰 연구에서의 약물 유해성 평가를 하기위해 성향 점수를 사용하였다. 처리집단과 통제집단을 균형있게 만들기 위해 공변량을 예측 확률로 나타내서 성향 점수를 구한 뒤 비슷한 값을 갖는 환자들을 대응시켜 연구에 사용하였다. 또한, Chung 등 (2007)은 k-NN 과 회귀모형에서 통계적 매칭방법에 관한 연구를 하였으며, Park (2015)은 회귀모형에서 성향점수에 관한 연구를 하였다.

이와 같이 최근 많은 논문에서 성향 점수 매칭방법을 사용하고 있지만 대부분 R, SAS나 STAT 프로그램을 사용하고 있는 추세이다.

Felix (2012)은 과거에는 SPSS Ver.에서 사용하지 못한 성향 점수 매칭방법을 SPSS 20 Ver.에서 사용가능하게 연구 개발하였다. Felix가 개발한 성향 점수 매칭 방법은 R packages, Matchit, Rltools, Cem이라는 프로그램이 설치가 되어 있어야 하고, NN (Nearest Neighbor) 매칭, Caliper매칭, Kernel 매칭, LL (Local Linear) 매칭 방법 등을 구분해서 사용해야 한다. 하지만 Felix가 제시한 성향 점수 매칭 방법은 사용 시 통계적인 지식이 있어야 적절한 매칭 방법을 구분하여 사용할 수 있다. 반해 IBM에서 개발한 퍼지 매칭 방법은 Python 프로그램만 설치하면 되고, 캘리퍼 매칭을 토대로 1:1매칭을 기본으로 한 방법으로 통계전공자가 아닌 일반인들도 쉽게 이용할 수 있다. 따라서 본 연구는 SPSS 22 Ver.부터 새롭게 추가 되어진 퍼지 알고리즘을 이용한 성향 점수 매칭 방법을 시행하고자 한다.

2. 연구방법

2.1. 성향점수의 의미

성향점수 매칭 방법은 이분형 처치 할당에 대해 강한 무관성의 가정을 기초로 한다. 이는 공변량 X 가 주어졌을 때, 처치 할당의 여부는 반응변수 (Y_i)가 독립 (가정1)이고, 실험 집단과 p 의 각 구성원들 처치 할당 확률 분포는 공통의 영역에 존재한다 (가정 2)는 두 가지 가정을 만족시킬 때 충족된다.

가정 1) 조건부 독립성의 가정으로서 공변량 X 가 주어졌을 때, 처치 할당여부 반응변수 Z 는 Y_i 들과 독립적이라 가정한다. 이 반응변수와 관련한 차이는 관찰된 변수에 의해 통제가 가능하고 관찰되지 않은 어떤 특성도 반응변수에 영향을 주지 않는다는 것을 의미한다.

$$(Y_0, Y_1) \perp Z | X$$

가정 2) 공통 영역의 가정으로서 분석 대상의 처치 Z 의 할당 확률은 공통 영역 내에 있다고 가정한다.

$$0 < P(Z = 1 | X) < 1$$

각 그룹에 대한 특성을 측정할 수 있는 변수들이 충분히 존재할 경우 이들을 통제하는 것만으로 선택 편향이 없는 효과 추정치를 산출해 낼 수 있다. 따라서 성향 점수는 관찰된 공변량의 처리조건을 선택하는 방법을 표현한 것으로 실험 집단과 통제 집단이 일치하도록 사용한다. 성향점수를 이용한 매칭 방

법으로 가장 많이 이용하는 것은 특정 폭의 캘리퍼를 사용하는 그리디 매칭이다. 캘리퍼를 이용한 그리디 매칭 방법은 실험 집단을 중심으로 일정한 성향점수의 범위를 설정하여 이 범위에 들어오는 모든 통제 집단의 개체들 중에서 가장 가까운 개체를 선택하는 방법으로, 캘리퍼의 폭이 커질수록 동일하지 않은 대상자들을 매칭하게 되는 결과가 나올 수 있다. 이는 매칭된 집합에서 각 집단 간의 체계적인 차이를 크게 하기 때문에 효과를 추정함에 있어 편향이 또한 크게 작용하기 때문이다. 그러나 이것은 매칭된 짝의 수가 많아지게 하기 때문에 추정된 효과의 정밀도를 증가시킨다 (Peter, 2007). 다시 말해 캘리퍼의 성향 점수는 모든 설명변수를 조건으로 했을 때, 실험 집단(1)이 발생할 확률값을 이용하여 개발되었다. 즉, 실험 집단 값이 발생할 확률값이 통제 집단 (0)에서 가장 유사한 것이 바로 해당 실험 집단 개체와 대응되는 개체라고 할 수 있다. 보통 캘리퍼 허용오차는 정해진 기준은 없으나 통상적으로 0.01~0.00001이 사용되고 있고 (Kim, 2008), 캘리퍼 허용오차는 매칭 대상을 제한하는 오차범위로서 허용오차 크기가 0에 가까울수록 실험 집단의 개체수가 많이 상실되기 때문에 실험 집단의 표본 수와 효율성을 고려해서 사용해야 한다.

퍼지 매칭은 퍼지 알고리즘에 캘리퍼를 사용한 그리디 매칭을 토대로 만들어졌고, SPSS 20 Ver. 이상에서 사용 가능하다. 또한 파이썬 프로그램이 설치되어야 사용이 가능하다. 원래 퍼지 매칭은 질문의 문구나 단어들에 대한 번역이 필요할 때 의미상 거의 맞는 문장 (단어 포함)을 검색하는데 사용되지만, IBM 자체에서 퍼지 함수를 이용한 매칭 알고리즘을 개발하였다. 퍼지 매칭은 완전히 동일한 데이터가 아닌 비슷한 성향을 가진 데이터를 추출하게 되어 있고, 성향 점수 매칭과는 다르게 1:1매칭만을 허용하고 있다.

2.2. 성향 점수 매칭 분석 방법 및 질평가

(1) 데이터 탐색

첫 번째, 매칭 전 각 그룹 간 결과의 차이를 검증하는데 혼란을 줄 수 있는 변수를 선택하여 공변량으로 만들어야 한다. 매칭 전 차이가 있는 혼란 변수를 매칭을 통해 그 차이를 최소화 하는 것이 매칭 방법의 목적이다. 이 경우 공변량은 일반적인 변수 (예: 성별, 연령, 소득 등)로는 충분하지 않다. 따라서 연구자는 집단에 영향을 줄 수 있는 혼란변수를 찾기 위해 노력을 해야 한다. 예를 들면 Guanglei과 Stephen (2005)는 200개가 넘는 공변량이 고려된 연구를 내놓았지만 오히려 이 경우에는 관찰가능하지 않은 요인을 확인하기가 힘들다고 제시되고 있다. 아직까지 혼란 변수를 선택하는 방법은 확인되고 있지 않다 (Felix, 2012).

(2) 성향 점수 추정:

성향 점수는 1번에서 정해진 혼란 변수를 기반으로 추정해야 한다. 비교 집단 (예: 슬픔과 절망감 경험)을 종속변수로 하고, 이에 영향을 줄 수 있는 다양한 혼란 변수 (1번에서 선택되었던 변수)들을 독립변수로 하는 로지스틱 회귀분석을 이용하여 성향 점수를 추정하거나 혼란 변수 자체를 더미변수로 만들어 직접 사용한다.

(3) 매칭 방법 선택

매칭은 다양한 방법으로 수행 될 수 있지만, 여러 가지 매칭 방법들 가운데 어떤 방법을 선택해야 하는지는 어려운 문제이다. 표본 수가 크면 모든 방법들이 정확한 매칭에 가까워지고 거의 동일한 결과가 나타날 수 있지만, 표본 수가 작은 경우 매칭의 선택에 따라 표본의 수에 영향을 줌으로 중요하다. 따라서 공통영역 점검과 매칭의 질을 평가 후 매칭방법을 선택할 필요가 있다.

(4) 매칭의 질 평가

4번째 단계에서는 집단 간 성향 점수의 분포가 균형을 이루고 있는가를 평가하는 것이다. 추정된 성

향 점수를 기준으로 매칭을 한 후, 성향 점수 모형이 적절하게 지정되었을 때, 측정된 기저시점에서 각 집단의 공변량 분포가 유사하게 나타난다 (Daniel 등, 2007). 각 집단 간 균형을 평가하기 위한 방법들은 다양하다. 이 방법들은 각 집단 사이에 측정된 기저시점 공변량의 분포를 비교하기 위해 표준화 차이와 각 집단에 대한 상자 그림, 사분위 그림 등을 사용한다. 많은 연구자들은 성향 점수를 이용한 매칭에서 각 집단 사이의 균형을 평가하기 위해 표준화 차이를 이용한다 (Paul과 Donald, 1985).

연속형 공변량에 대해 표준화 차이는 다음과 같이 정의한다.

$$d = \frac{100|\bar{x}_{treatment} - \bar{x}_{control}|}{\sqrt{\frac{s_{treatment}^2 + s_{control}^2}{2}}}$$

여기서 $\bar{x}_{treatment}$ 와 $\bar{x}_{control}$ 은 각 집단의 공변량의 표본평균이다. $s_{treatment}^2$ 와 $s_{control}^2$ 은 각 집단의 공변량의 표본분산이다. 표준화차이계수 d 가 10% 이상의 값을 가지게 되면 대응된 표본의 두 군간의 변수가 균등하게 배분되었다고 볼 수 없다 (Bernhard과 Hans, 1986).

실제 성향 점수에 대한 전제로 각 집단은 측정된 기저시점의 특성치들이 같은 분포를 가져야 한다. 그러므로 성향 점수로 매칭 된 표본에서 각 집단 간의 연속형 변수들의 분포뿐만 아니라 평균도 유사하기 때문에 (Douglas 등, 2010), 각각의 두 그룹 간 공변량의 평균과 분산 모두 중요하다. 그러므로 표준화 차이뿐만 아니라 각각의 상자 그림이나 사분위 그림의 적용은 각각의 그룹 간의 연속형 기저시점의 공변량 비교에 유용하다.

(5) 결과 변수의 효과 추정

매칭된 변수들을 독립변수 (예: 슬픔 · 절망감여부)와 종속변수 (예: 자살시도 여부)의 추정하고 비교하여 검토하는 것이 필요하다.

본 연구에서는 기본 절차를 토대로 하여,

첫째, 허용오차 범위를 0.01, 0.001, 0.0001 세 가지로 다양하게 적용하고자 한다.

둘째, 상자도표와 교차분석 및 t -test를 통하여 성향 점수의 분포가 균등하게 분포되었는지 확인해보고자 한다.

셋째, 각 집단 사이의 균형을 평가하기 위해 표준화 차이를 이용하여 각 집단 사이의 차이를 확인해보고자 한다.

3. 연구결과

3.1. 조사 대상자 및 방법

본 연구는 ‘청소년의 슬픔 · 절망감 경험 (문항번호 33)’을 바탕으로 퍼지 매칭을 이용한 후향적 비교 연구이다. SPSS 22 Ver.을 이용하였고 허용오차 범위를 달리하여 퍼지 매칭 결과를 비교하였다. 본 연구논문에 사용된 ‘survey data: 청소년건강행태’ 데이터는 2013년에 시행한 원시자료이며, 이 자료는 인터넷이 가능한 학교 컴퓨터실에서 수업시간 1시간을 할애하여 일괄조사를 원칙으로 한 ‘익명성자기입식 온라인조사’이다. ‘슬픔 · 절망감 경험’을 독립변수로 두었고, 실험 집단 (예:4098명)과 통제 집단 (아니오:7914명)은 총 12,012명의 고3학생들을 대상으로 실시하였다. 혼란 변수는 성별, 거주형태, 음주 경험, 흡연 경험, 평상시 스트레스 정도, 학업 성적, 경제 상태를 포함하였다.

퍼지 매칭 방법을 위해 기존 연구에 참여한 모든 실험 집단과 통제 집단의 성향 점수는 SPSS 22 ver.에서 로지스틱 회귀분석을 이용하여 계산되었고, 표준화 차이는 Excel을 이용하여 계산하였다.

3.2. 연구 자료 분석 결과

(1) 퍼지 매칭 방법에 따른 분포도

매칭 전 대상자들의 성향 점수 분포를 Figure 3.1에 제시하였다. 매칭 전 통제 집단은 성향 점수 평균이 0.18에서 0.39사이에 몰려 있으며, 0.4이상인 대상자가 있는 것으로 많은 것으로 나타났고, 실험 집단은 성향 점수 평균이 0.27에서 0.57 사이로 넓게 퍼져 있었다. 이 두 그룹은 유사한 부분이 0.27에서 0.39사이로 적은 부분 겹쳐져 있어, 두 그룹 간 유사한 특징을 가진 사람의 수가 적음을 알 수 있다. 허용오차 점수를 0.01로 매칭 후 대상자들의 성향 점수 분포를 Figure 3.2에 제시하였다. 통제 집단과 실험 집단 각각 3602명씩 매칭 되었으며 두 그룹 간 성향 점수 평균이 약 0.39로 같고 분포는 두 그룹이 유사한 양상을 보였다. 허용오차 점수를 0.001로 매칭 후의 대상자들의 성향 점수 분포를 Figure 3.3에 제시하였다. 통제 집단과 실험 집단 각각 3480명씩 매칭 되었으며 그룹 간 성향 점수 평균이 약 0.39로 같고 분포는 두 그룹이 유사한 양상을 보였다. 허용오차 점수를 0.0001로 매칭 후의 대상자들의 성향 점수 분포를 Figure 3.4에 제시하였다. 통제 집단과 실험 집단 각각 3249명씩 매칭 되었으며 두 집단 성향 점수 평균이 약 0.39로 같고 분포는 두 그룹이 유사한 양상을 보였다.

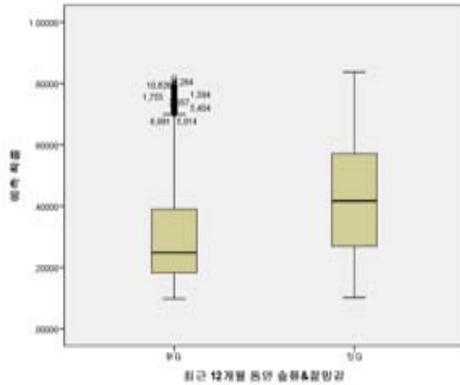


Figure 3.1 a box diagram before Fuzzy matching

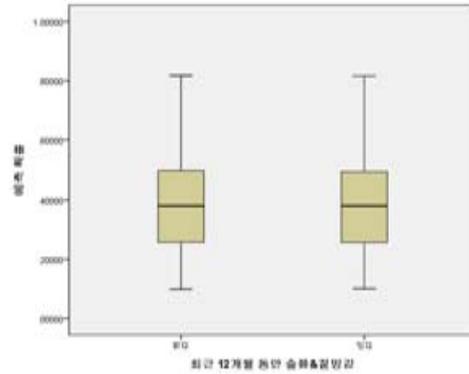


Figure 3.2 Distribution after Fuzzy matching (error tolerance: 0.01)

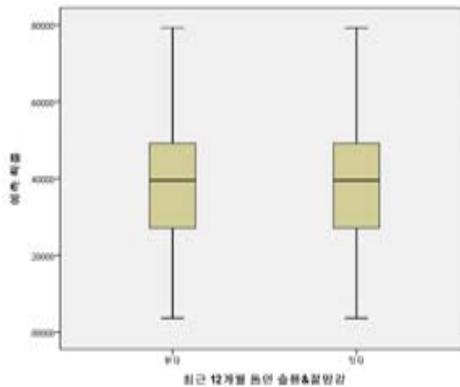


Figure 3.3 Distribution after Fuzzy matching (error tolerance: 0.001)

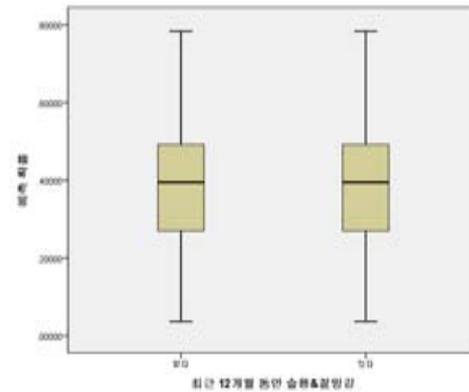


Figure 3.4 Distribution after Fuzzy matching (error tolerance: 0.0001)

(2) 퍼지 매칭 차이 분석

매칭 전 실험 집단은 4098명, 통제 집단은 7914명으로 나타났고, 모든 변수에서 통계적으로 유의하게 실험 집단과 통제 집단 사이에 차이를 보였다. 슬픔·절망감에 따른 스트레스, 학업성적, 경제상태의 분산은 가정되지 않음을 알 수 있다. 허용오차를 0.01로 설정한 후 매칭을 했을 때 실험 집단과 통제 집단은 각각 3602명씩 매칭 되었다. 이 때, 거주형태와 흡연여부에서만 통계적으로 유의한 차이가 있는 것으로 나타났고, 슬픔·절망감에 따른 경제상태의 분산은 두 집단이 동질하지 않음을 알 수 있었다. 허용오차를 0.001로 설정한 후 매칭을 했을 때 실험 집단과 통제 집단은 각각 3480명씩 매칭 되었다. 모든 변수에서 통계적으로 유의한 차이가 없는 것으로 나타났고, 슬픔과 절망감에 따른 경제상태의 분산은 가정되지 않음을 알 수 있다. 허용오차를 0.0001로 설정한 후 매칭을 했을 때 실험 집단과 통제 집단은 각각 3249명씩 매칭 되었다. 모든 변수에서 통계적으로 유의한 차이가 없는 것으로 나타났다. 또한 모든 변수에서 분산 또한 동질함을 알 수 있다.

Table 3.1 Analysis the difference between before Fuzzy matching and after Fuzzy matching (Error tolerance 0.01)

		State before Matching		Error tolerance 0.01			
		Sadness & Depression Experience		Sadness & Depression Experience			
		No	Yes		No	Yes	
		(N=7914)	(N=4098)	<i>p</i>	(N=3602)	(N=3602)	<i>p</i>
Sex	Female	3596 (60.2)	2375 (39.8)	<i>p</i> < .001	2037 (56.6)	1989 (55.2)	0.255
	Male	4318 (71.5)	1723 (28.5)		1565 (43.4)	1613 (44.8)	
Residence Type	Family	7331 (92.6)	3773 (92.1)	0.013	3384 (93.9)	3312 (91.9)	0.005
	Relatives	77 (1.0)	68 (1.7)		46 (1.3)	54 (1.5)	
	Boarding	461 (5.8)	232 (5.7)		150 (4.2)	214 (5.9)	
	House & Dormitory Orphanage	45 (0.6)	25 (0.6)		22 (0.6)	22 (0.6)	
Drinking Experience	No	2978 (37.6)	1173 (28.6)	<i>p</i> < .001	1166 (32.4)	1114 (30.9)	0.188
	Yes	4396 (62.4)	2925 (71.4)		2436 (67.6)	2488 (69.1)	
Smoking Experience	No	5742 (72.6)	2656 (64.8)	<i>p</i> < .001	2496 (69.3)	2380 (66.1)	0.003
	Yes	2172 (27.4)	1442 (35.2)		1106 (30.7)	1222 (33.9)	
Stress		2.76±0.88	2.13±0.85	<i>p</i> < .001*	3.75±0.79	3.74±0.80	0.615
Academic Grades		3.02±1.15	3.25±1.14	<i>p</i> < .001*	2.82±1.15	2.79±1.15	0.339
Economic Conditions		3.01±0.90	3.15±0.97	<i>p</i> < .001*	2.92±0.93	2.89±0.97	0.149*

* Equal variance is not assumed.

Table 3.2 Analysis the difference after Fuzzy matching (Error tolerance 0.001) and after Fuzzy matching (Error tolerance 0.0001)

		Error tolerance 0.001		Error tolerance 0.0001			
		Sadness & Depression Experience		Sadness & Depression Experience			
		No	Yes		No	Yes	
		(N=3480)	(N=3480)	<i>p</i>	(N=3249)	(N=3249)	<i>p</i>
Sex	Female	1939 (55.7)	1942 (55.8)	0.942	1818 (56.0)	1804 (55.5)	0.727
	Male	1541 (44.3)	1538 (44.2)		1431 (44.0)	1445 (44.5)	
Residence Type	Family	3262 (93.7)	3221 (92.6)	0.175	3103 (95.5)	3108 (95.7)	0.858
	Relatives	37 (1.1)	44 (1.3)		21 (0.6)	17 (0.5)	
	Boarding	158 (4.5)	195 (5.6)		118 (3.6)	119 (3.7)	
	House & Dormitory Orphanage	23 (0.7)	20 (0.6)		7 (0.2)	5 (0.2)	
Drinking Experience	No	1141 (32.8)	1077 (30.9)	0.1	1026 (31.6)	1002 (30.8)	0.521
	Yes	2339 (67.2)	2403 (69.1)		2223 (68.4)	2247 (69.2)	
Smoking Experience	No	2379 (68.4)	2359 (67.8)	0.607	2241 (69.0)	2221 (68.4)	0.593
	Yes	1101 (31.6)	1121 (32.2)		1008 (31.0)	1028 (31.6)	
Stress		3.74±0.79	3.72±0.81	0.47	3.72±0.78	3.72±0.78	0.987
Academic Grades		2.82±1.14	2.80±1.14	0.543	2.78±1.11	2.79±1.11	0.738
Economic Conditions		2.91±0.92	2.89±0.95	0.450*	2.88±0.90	2.89±0.90	0.535

* Equal variance is not assumed

(3) 연속형 공변량에 대한 표준화 차이 분석

각 집단 사이의 균형을 평가하기 위해 표준화 차이를 이용한 결과 다음과 같이 나타났다 (Figure 3.5). 매칭전 표준화 차이는 80.5%로 두 그룹 간에 균등하게 배분되었다고 볼 수 없고, 매칭 후 (허용오차값: 0.01, 0.001, 0.0001) 표준화 차이는 각각 0.2%, 0.0%, 0.0%로 두 그룹간에 균등하게 배분되었다고 볼 수 있다.

E2		fx		=(100*(C2-C3))/(SQRT((D2+D3)/2))		
A	B	C	D	E	F	G
		평균	분산	d		
매칭전	있다	0.427538	0.029	80.53845		
	없다	0.296431	0.024			
0.01	있다	0.393715	0.26	0.224847		
	없다	0.392569	0.26			
0.001	있다	0.392102	0.24	-0.0001		
	없다	0.392102	0.24			
0.0001	있다	0.389982	0.23	0.000104		
	없다	0.389981	0.23			

Figure 3.5 Difference of Standardize Testing using Excel

4. 결론 및 제언

성향 점수 매칭은 관측된 공변량 벡터가 주어졌을 때 특정한 처리에 할당될 조건부 확률로 정의되며 로지스틱 회귀분석으로 확률이 예측된다. 예측된 확률을 성향 점수로 이용하여 두 집단을 유사하게 만들어 분석에 이용한다. 여기서 두 집단을 유사하게 만든다는 것이 성향 점수 방법을 이용하는 목적이자 가장 큰 이점이다.

선행논문에서 제시 했듯이 최근 많은 연구논문에서 성향 점수 매칭을 사용하고 있지만, 보통 SAS (Kim, 2011; Jeong, 2011), R (Gabriel 등, 2008), STAT (Lee 등 2007; Park 등, 2012)를 사용하고 있음을 확인 할 수 있었다. 2012년 처음으로 Felix는 SPSS를 이용하여 성향 점수 매칭 방법을 제시한 바 있다. 본 연구에서는 SPSS 20 ver.부터 새롭게 추가된 퍼지 매칭 방법을 제시하였다.

첫째, 연구결과에서 허용오차 점수가 0에 가까워질수록 (0.01: 각 3602명, 0.001: 각 3480명, 0.0001: 각 3249명) 매칭이 되는 케이스 수가 작아짐을 알 수 있었다. 따라서, 대표본일 경우 허용오차 점수를 최대한 0에 가깝게 선택할 수 있지만 소표본일 경우에는 허용오차 점수가 작을수록 매칭되는 케이스 수가 작아지기 때문에 적절한 허용오차 값을 지정해줘야 될 필요성이 있다. 하지만 아직까지 케이스 수에 따른 적합한 허용오차 점수는 제시되어 있지 않은 상태이다

둘째, 매칭된 성향 점수를 이용한 상자도표를 확인해본 결과 두 그룹의 분포가 비슷한 것을 확인 할 수 있었고, χ^2 -test와 t-test 결과에서 빈도와 평균값의 분포를 확인해 본 결과 허용오차 값이 0에 가까울수록 평균과 분산에 차이가 없는 것으로 확인 할 수 있었다. 따라서 허용오차 값이 0에 가까울수록 정확하게 매칭이 되는 것을 확인 할 수 있었다.

셋째, 매칭 후 표준화 차이 만족도를 확인 해본 결과 구 그룹의 표준화 차이가 각각 10%이하로 나타났다. 두 그룹 간에 균등하게 배분되었다고 볼 수 있다.

일반적으로 실험 집단의 표본 수는 통제 집단의 표본수보다 항상 적은 수를 갖고 있다. 그리고 허용오차 값이 작을수록 매칭되는 케이스 수가 줄어드는 걸 확인할 수 있었다. 따라서 표본의 크기가 충분히

커야 성향 점수 매칭의 유용성이 커지기 때문에 케이스 수가 충분히 확보가 된 상태에서 매칭을 실시해야 한다.

매칭 방법은 공변량으로 인한 선택편의와 원인변수가 결과변수에 의해 영향을 받는 내생성이 존재하거나 측정오차가 존재하는 경우에는 한계가 존재한다. 따라서 퍼지 매칭에서도 측정하지 않은 공변수들에서 비롯된 추정값의 비뚤임을 피할 수 없으며 모형에 따라 다른 성향 점수가 나타나기 때문에 연구자들마다 결과가 다르게 나타날 수 있다. 따라서 퍼지 매칭을 사용하기 위해서는 실험 집단과 유사한 특징을 갖는 통제 집단을 충분히 확보할 수 있어야하고, 혼란변수가 충분히 확보되어 있는 ‘좋은 데이터’가 반드시 필요하다 (James 등, 1997).

퍼지 매칭의 장점은 파이톤 프로그램 외에 다른 데이터가 불필요하고, 쉽게 적용할 수 있으며, 빅 데이터를 이용할 경우 매칭 방법을 통해 일반적 특성이 동일한 변수들을 추출해 낼 수 있다. 반면, 퍼지 매칭 방법은 1:1매칭만 허용하고, 일치허용오차를 통해서만 매칭 대상을 제안하는 한계가 있다. 또한, 빅 데이터일 경우 컴퓨터 사양이 좋아야 SPSS에서 빠르게 구동이 된다. 만약 케이스 수가 클 경우 STAT, SAS, R로 실행하는게 좋을 듯하다.

SPSS는 현재 R, SAS, STAT와 같이 성향점수 매칭에 대한 여러 가지 형태로 변형이 가능한 매크로로 짜여있지 않다. SPSS는 R 플러그 인을 통해 사용할 수 있는 PS-matching방법 (Felix, 2012)과 퍼지 알고리즘을 통해 간단한 성향 점수 매칭을 사용할 수 있다.

따라서 SPSS ver.에서는 케이스 수와 연구 목적에 맞는 성향 점수 매칭 방법을 사용하기에 한계성이 있음을 알 수 있었다. 이 후 다른 프로그램과 마찬가지로 SPSS에서도 기본 매칭 알고리즘을 통해 여러 조건으로 변형 가능한 매칭 알고리즘을 개발하는 후속연구가 필요하다고 생각된다.

References

- Bernhard, K. F. and Hans, R. (1986). Standard distance in univariate and multivariate analysis. *The American Statistician*, **40**, 249-251.
- Chung, S. S., Kim, S. Y., Lee, S. S. and Lee, K. H. (2007). A Statistical Matching Method with k-NN and Regression. *Journal of the Korean Data & Information Science Society*, **18**, 879-890
- Daniel, E. H., Kosuke, I., Gary, K. and Elizabeth, A. S. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, **15**, 199-236.
- Douglas, E. F., Andrew, C. L., Josep, M. H. and Robert, L. O. (2010). *Analysis of observational health care data using SAS*, SAS Press, USA.
- Felix, J. T. (2012). *Propensity score matching in SPSS*, Cornell University Library, New York.
- Gabriel, T., Jason, D. C., Philippe, R., Yves, C., Olivier, B., Michel, F., Herve, M., Guy, L. and Raphael, P. (2008). Survival after bilateral versus single lung transplantation for patients with chronic obstructive pulmonary disease: A retrospective analysis of registry data. *The Lancet*, **371**, 744-751.
- Guanglei, H. and Stephen, W. R. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, **27**, 205-224.
- James, J. H., Hidehiko, I. and Petra, E. T. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic studies*, **64**, 605-654.
- Jeong S. H. (2011). *Effect of perceived social support on treatment responses and outcomes of depressive disorders using a propensity score matching method*, Catholic University, Seoul.
- Kim, E. S. (2008). *National basic livelihood security system and labour supply*, Korean labor & Income panel study, 457-471, Korea.
- Kim, M. H. (2011). *A study on the risk factors for prehypertension using the propensity score method*, Gyeongsang National University, Gyeongsangnamdo.
- Lee, S. J., Yoo, J. S., Sin, M. G., Park, C. G., Lee, H. C. and Choi, E. J. (2007). The use of propensity score matching for evaluation of the effects of nursing interventions. *Journal of Korean Academy of Nursing*, **37**, 414-421.
- Park, J. (2015). Performance study of propensity score methods against regression with covariate adjustment. *Journal of the Korean Data & Information Science Society*, **26**, 217-227.

- Park, S. H. and Kim T. I. (2012). Specific examples of how to apply and propensity score matching, *The Korean Association for Policy Studies Proceedings*, Korea.
- Paul, R. R. and Donald, B. R. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.
- Paul, R. R. and Donald, B. R. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, **39**, 33-38.
- Peter, C. A. (2007). Propensity score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. *Journal of thoracic and cardiovascular surgery*, **134**, 1128-1135.
- Richard, A. B. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, **48**, 386-398.
- William, R. H. (2006). Observational studies of drug safety-aprotinin and the absence of transparency. *New England Journal of Medicine*, **355**, 2171-2173.
- The ninth Korea youth risk behavior web-based survey. (2013). *Ministry of education, ministry of health and welfare*, Korea centers for disease control and prevention, Korea.

FUZZY matching using propensity score: IBM SPSS 22 Ver.[†]

So Youn Kim¹ · Jong Il Baek²

¹²Division of mathematics and informational statistics, Wonkwang University

Received 1 December 2015, revised 8 January 2016, accepted 18 January 2016

Abstract

Fuzzy matching is proposed to make propensities of two groups similar with their propensity scores and a way to select control variable to make propensity scores with a process that shows how to acquire propensity scores using logic regression analysis, is presented. With such scores, it was a method to obtain an experiment group and a control group that had similar propensity employing the Fuzzy Matching. In the study, it was proven that the two groups were the same but with a different distribution chart and standardization which made edge tolerance different and we realized that the number of chosen cases decreased when the edge tolerance score became smaller. So with the idea, we were able to determine that it is possible to merge groups using fuzzy matching without a precontrol and use them when data (big data) are used while to check the pros and cons of Fuzzy Matching were made possible.

Keywords: FUZZY mathcing, KYRBS, propensity score.

[†] This paper is extracted from doctorate thesis of Wonkwang University in 2015.

¹ Lecture Professor, Division of mathematics and informational statistics, Wonkwang University, Iksan 54538, Korea.

² Corresponding author: Professor, Division of mathematics and informational statistics, Wonkwang University, Iksan 54538, Korea. E-mail: jibaek@wku.ac.kr