

# 성향점수매칭 방법을 사용한 로지스틱 회귀분석에 관한 연구\*

김소연 · 백종일<sup>†</sup>

원광대학교 자연과학대학 수학과통계학부

## On Logistic Regression Analysis Using Propensity Score Matching\*

So Youn Kim · Jong Il Baek<sup>†</sup>

Division of Mathematics and Informational Statistics, Wonkwang University

**Purpose:** Recently, propensity score matching method is used in a large number of research paper, nonetheless, there is no research using fitness test of before and after propensity score matching. Therefore, comparing fitness of before and after propensity score matching by logistic regression analysis using data from ‘online survey of adolescent health’ is the main significance of this research.

**Method:** Data that has similar propensity in two groups is extracted by using propensity score matching then implement logistic regression analysis on before and after matching separately.

**Results:** To test fitness of logistic regression analysis model, we use Model summary, -2Log Likelihood and Hosmer-Lomeshow methods. As a result, it is confirmed that the data after matching is more suitable for logistic regression analysis than data before matching.

**Conclusion:** Therefore, better result which has appropriate fitness will be shown by using propensity score matching shows better result which has better fitness.

**Keywords:** Propensity Score, Logistic Regression Analysis, Fitness For Logistic Model Assessment, KYRBS

### 1. 서론

관찰연구 중에서 두 군을 비교하는 연구는 대부분 비무작위 표본 추출을 기준으로 진행되어 왔지만 비무작위 표본 추출은 선택편의(selection bias)를 통제할 수 없는 문제점을 가지고 있다. 선택편의는 집단

간 이질성으로 인해 발생하는 것으로, 결과의 인과관계에 대한 올바르지 못한 추론을 하게 하거나 결과를 과소 혹은 과대 추정하는 오류를 발생시킨다. Berk[1]는 선택편의가 입증되기 전까지 모든 사회학적 연구에서는 선택 편의가 존재하는 것으로 간주하고 이를 수정하기 위해 노력해야 한다고 설명하고 있다. 따라

\* 이 논문은 2016년도 원광대학교 교내 연구비에 의해 연구되었음.

<sup>†</sup> 교신저자 jibaek@wku.ac.kr

2016년 9월 30일 접수; 2016년 12월 7일 수정본 접수; 2016년 12월 16일 게재 확정.

서 집단을 선별할 때에는 공변량(covariate)들의 불균형(unbalanced)을 통제할 수 있는 대상을 선정하여 선택편의를 최소화할 방법을 찾아야 한다[2].

이에 Rosenbaum and Rubin[3]은 성향점수를 이용한 평균 처리 효과(average treatment effect, ATE)를 제안하였고, 그들의 논문에서 관찰 자료의 인과관계를 보다 명확히 계산하기 위한 방법으로 성향 점수라는 용어를 처음 소개 하였다. 이 방법은 관측된 공변량 벡터가 주어졌을 때 특정한 처리에 배정될 조건부 확률로 정의하고 있다. 성향 점수는 실험집단과 통제집단 간에 비교하고자 하는 처리변수 이외의 유사하지 않게 분포된 변수들을 통제하는 기법 중 하나이다 조건부 확률에 대한 예측은 모형화된 결과로부터 성향 점수를 추정함으로써 구할 수 있고, 준-랜덤화 실험과 비슷한 효과를 갖게 한다. 두 실험 대상이 같은 성향 점수를 갖는다면 한 명은 실험 집단이 되고 다른 한명은 통제 집단이 되는데 이 과정에서 두 실험대상이 각각의 집단에 무작위로 배정된다[4].

성향 점수 매칭 방법을 적용한 사례를 보면, Park and Kim[5]은 실제 정책 효과(국민기초생활보장제도의 효과분석)를 분석하는 데에 실험집단과 유사한 비교집단을 설정하고, 두 집단 간을 성향점수 매칭을 통해 가능한 유사한 집단으로 배정한 뒤 정책의 효과를 추정하였다. Jeong[6]은 전향적 우울증 환자 코호트 자료를 이용하여 공변량 정보에 대하여 스칼라의 요약치를 제공하는 성향 점수를 이용하여 분석하였다. Lee *et al.*[7]은 당뇨환자를 대상으로 생활습관 개선 프로그램을 실시한 후 그 효과를 평가하기 위해 실험 집단과 비교집단의 선택편중을 줄이고 그 결과를 비교하는데 사용하였다. Thabut *et al.*[8]은 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. Hiatt[9]은 관찰 연구에서의 약물 유해성 평가를 하기위해 성향 점수를 사용하였다. 처리집단과 통제 집단을 균형있게 만들기 위해 공변량을 예측확률로 나타내서 성향 점수를 구한 뒤 비슷한 값을 갖는 환자들을 대응시켜 연구에 사용하였다.

이와 같이 최근 많은 연구논문에서 성향 점수 매칭 방법을 사용하고 있다. 하지만 성향 점수 매칭 방법을 많이 사용하고 있음에도 불구하고 성향 점수 매칭전과 후의 적합도 검사를 사용한 논문은 없었다. 따라서

본 연구논문은 ‘청소년 건강행태 온라인 조사 데이터’를 이용하여 로지스틱 회귀분석을 통한 성향점수 매칭 전과 후의 여부에 따라 어느 것이 적합도가 좋은지를 비교하고자 한다. 로지스틱 회귀분석의 모형 적합도 검사는 Model summary, -2Log Likelihood, Hosmer-Lomeshow 기법 방법을 이용하여, 다음과 같이 성향점수 전과 후의 로지스틱 회귀분석 비교방법에서 적합도 검사를 바탕으로 본 연구의 연구가설을 설정하였다.

첫째, 성향 점수 매칭 후 Model summary를 통해 우도 비율 검사 결과 통계량 값이 낮아질 것이다.

둘째, 성향 점수 매칭 후 -2Log Likelihood 값이 작아질 것이다.

셋째, 성향 점수 매칭 후 독립변수의 종속변수에 대한 설명력을 파악하기 위해서 성향 점수 매칭 후 Hosmer-Lomeshow 검정의 p-value 값이 높아질 것이다.

## 2. 연구방법

### 2.1 성향점수 의미

성향점수 매칭 방법은 이분형 처리 할당에 대해 강한 무관성의 가정을 기초로 한다. 이는 공변량  $X$ 가 주어졌을 때, 처리 할당의 여부는 반응변수 ( $Y_i$ )가 독립(가정 1)이고, 실험 집단의 각 구성원들 처리 할당 확률 분포  $p$ 는 공통의 영역에 존재한다(가정 2)는 두 가지 가정을 만족시킬 때 충족된다.

가정 1) 조건부 독립성의 가정으로서 공변량  $X$ 가 주어졌을 때, 처리 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 처리  $Z$ 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 방법으로 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[3]. 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[10], 캘리퍼 허용오차는 매칭 대상을 제한하는 오차범위로서 허용오차 크기가 0에 가까울수록 실험 집단의 개체수가 많이 상실되기 때문에 실험 집단의 표본의 수와 효율성을 고려해서 사용해야 한다.

## 2.2 로지스틱 회귀분석의 정의

회귀분석은 근본적으로 연속형 자료를 종속변수로 갖는다. 하지만 질병의 유무라는 변수는 범주형 자료이기 때문에 회귀분석에 그대로 적용할 수가 없다. 그래서 회귀분석을 확장하여 종속변수  $y$ 를  $f(x)$ 라는 함수로 치환한 일반화 선형모형을 이용한다면 보다 폭 넓은 현상들을 회귀모형으로 설명하는 것이 가능해진다. 로지스틱회귀분석은 이 일반화 선형모형의 한 형태이다. 예를 들어 질병이 있을 확률을  $p$ 라고 할 때 로지스틱 회귀분석에서는  $f(x) = \ln \frac{p}{1-p}$ 라는 함수를 일반화 선형 모형에 대입하여 적용한다.

즉, 위험요인(예: 체중)의 각 수준에 따른 질병(예: 고혈압)이 있을 확률  $p$ 를 로짓변환해주면 logit  $p$ 값은  $-\infty \sim \infty$ 의 연속형 변수의 형태를 띠게 된다. 그리고 이 logit  $p$ 를 대상으로 회귀분석에 적용한 것이 로지스틱 회귀분석이다.

## 2.3 로지스틱 회귀분석 모형의 적합도 평가

### 2.3.1 Model Summary(Step $\chi^2$ -test)

선형회귀분석에서 F-검정을 이용하여 추정된 회귀모형의 유의성을 검정했듯이 로지스틱회귀분석에서는 모형계수 전체 테스트(Model  $\chi^2$ -test)를 통해 추정된 회귀모형의 유의성을 검정한다. 이 Model  $\chi^2$ -test는 우도 비율 검사의 결과로 검정통계량은 자유도가 1인  $\chi^2$ -test 분포를 따른다. Step, Block, Model 값이 나타나 있는데, Step은 이전 단계와 현 단계의 -2Log Likelihood(이하 -2LL)값의 차이이며, Block은 분석 시 블록을 여러 개 설정했을 때 이전 블록과 현재 블록의 -2LL값의 차이이다. Model은 절편만 있는 기저모델의 -2LL값과 독립변수들을 포함한 분석모델의 -2LL값의 차이이다. 이때 자유도는 비교되는 두 모델에서 추정해야 할 미지수이다. 따라서 Step  $\chi^2$ -test 통계량 값이 작을수록 좋은 모형이다[10].

### 2.3.2 -2Log Likelihood

모수의 추정값이 주어졌을 때 표본에서 결과가 실제로 얼마나 발생하는지를 검사하는 방법이다. 이때 모수의 추정값이 주어졌을 때 관측된 결과의 확률을 우도 함수로 표현한다. 우도 함수의 값은 1보다 작다. 그러므로 여기에 Log를 취하고 -2를 곱한 -2LL로 추정된 모형이 자료를 얼마나 잘 적합시키는지 보는 척도로 사용한다. (-)를 곱해주는 이유는 지수가 0과 1 사이인 Log값은 음(-)의 값을 가지기 때문에 부호를 반대로 해주기 위해서이며, 2배를 해준 것은 로그우도값이  $\chi^2$ 분포에 가깝게 하려고 하기 때문이다. 따라서 -2LL의 값은 작을수록 좋은 모형이다[18].

### 2.3.3 Hosmer-Lemeshow 검정

선형 회귀분석에서는 독립변수의 종속변수에 대한 설명력을 파악하기 위해 결정계수 값  $R^2$ 을 이용한다. 로지스틱 회귀분석에서도 로그 우도함수 값을 이용해 계산한 결정계수  $R^2$ 값이 있다. 로지스틱 회

귀분석에서는 선형 회귀분석과 달리 두 가지 면에서  $R^2$ 을 해석하는데 첫째, 종속변수가 범주형 변수인 로지스틱회귀분석에서는 오차의 등분산성 가정이 충족되지 않는다. 로지스틱회귀분석에서 오차분석을 구하는 공식은  $var(e_i) = p_i(1-p_i)$ 이다. 따라서 오차분산은 동일하지 않고 예측 확률( $p_i$ )에 따라 달라지기 때문에 종속변수에 따라 오차분산이 달라지고  $R^2$  값도 변하므로 로지스틱 회귀분석에서는 총분산 중 설명된 분산이라고 해석하기 어렵다[3]. 둘째, Cox and Snell[12]이 개발한  $R^2$ 과 Nagelkerke[13]가 개발한  $R^2$ 의 공식을 제시하고자 한다. 우선, Cox와 Snell의  $R^2$ 을 계산하는 공식은 다음과 같다.

$$R_{cs}^2 = 1 - \left( \frac{L_0}{L_A} \right)^{-2/N}$$

$L_0$ : 기저모델의 우도함수

$L_A$ : 해당 분석모델의 우도함수

$N$ : 표본수

Cox와 Snell이 개발한  $R^2$ 의 문제점은 최대치가 1이 되지 않는다는 점이다. 이 문제점을 교정하기 위해 개발한 Nagelkerke의  $R^2$ 값이 있으며, 이를 구하는 공식은 다음과 같다.

$$R_N^2 = \frac{R_{cs}^2}{1 - (L_0)^{2/N}}$$

$R_{cs}^2$ : Cox와 Snell의  $R^2$

$L_0$ : 기저모델의 우도함수

$N$ : 표본수

그런데, Hosmer and Lemeshow[14]는 Hosmer-Lemeshow 검정 방법을 제시 했는데 이는 로지스틱 회귀분석에서 구한  $R^2$ 값은 대개 낮게 나오는 경향이 있어 독립변수의 종속변수에 대한 설명력을 파악하기 위해  $R^2$ 값을 너무 신뢰할 필요가 없다고 제시 하였다. 따라서 본 연구는 Hosmer와 Lemeshow가 제시한 모형의 적합도를 높이기 위해 Hosmer-Lemeshow 검정 방법을 사용하였다. 본 연구는, Hosmer-Lemeshow 검정법의 적합도를 계산하기 위해 예측 확률값을 서열화하여 각 집단별로 다시 분류하는 방법을 제안하였다. 따라

서 개별화된 사례의 형태를 가진 표본에 대해 보다 정확한 모형의 적합성 검증을 수행할 수 있다는 장점을 가지고 있음을 알 수 있었다. 이 검정법은 표본의 수가 충분히 클 때만 사용해야 하며, 귀무가설은 “모형이 적합하다”이기 때문에 p-value는 클수록 좋다.

### 3. 연구 결과

#### 3.1 조사 대상자 및 방법

본 연구는 성향점수 매칭을 이용한 로지스틱 회귀 분석 결과를 비교하는 후향적 연구이다.

본 연구논문에서 사용된 ‘survey data: 청소년건강상태’ 데이터는 2014년에 시행한 원시자료이며, 이 자료는 인터넷이 가능한 학교 컴퓨터실에서 수업시간 1시간을 할애하여 일괄조사를 원칙으로 한 ‘익명성자기기입식 온라인조사’이다.

성향점수 매칭 방법에 사용되는 성향 점수는 로지스틱 회귀분석을 이용하여 ‘슬픔·절망감 경험’을 독립변수로 두었고, 혼란 변수는 성별, 거주형태, 음주 경험, 흡연 경험, 평상시 스트레스 정도, 학업 성적, 경제 상태를 포함하여 예측확률변수를 산출하였다.

혼란변수 선정은 Kwon *et al.*[15] ‘우울경험에 따른 청소년들의 자살시도 위험 요인분석에서 쓰인 요인들을 참고로 선정하였다.

‘자살시도’를 종속변수로 두고, ‘슬픔·절망감 경험, 성별, 거주형태, 음주 경험, 흡연 경험, 평상시 스트레스 정도, 학업 성적, 경제 상태를 독립변수로 로지스틱회귀분석을 실시하여 성향 점수 매칭 전과 후의 결과를 비교하였다. 총 72,060명의 학생들을 대상으로 실시하였고, 분석 프로그램은 SPSS 23 Ver.을 이용하였다.

#### 3.2 퍼지매칭 차이 분석

본 연구는 우울과 비우울 집단이 자살시도에 영향을 주는지 확인해보기 위하여 혼란변수를 통제하였다.

매칭 전 ‘우울 집단’은 19,174명, ‘비우울 집단’은 52,886명으로 나타났고, 모든 변수에서 통계적으로 유의하게 우울 집단과 비우울 집단 사이에 인구학적 특성에 차이가 있음을 알 수 있었다.

**Table 1** Analysis the difference between before PSM and after PSM(Error tolerance 0.0001)

		State before Matching			Error tolerance 0.01		
		Sadness & Depression Experience			Sadness & Depression Experience		
		No (N = 52886)	Yes (N = 19174)	P	No (N = 16770)	Yes (N = 16770)	P
Sex	Female	28474(53.8)	7996(41.7)	.000	7394(44.1)	7346(43.8)	.597
	Male	24412(46.2)	11178(58.3)		9376(55.9)	9424(56.2)	
Residence Type	Family	50799(96.1)	18045(94.1)	.000	16056(95.7)	16041(95.7)	.936
	Relatives	387(0.7)	256(1.3)		134(0.8)	140(0.8)	
	Boarding House & Dormitory	1483(2.8)	736(3.8)		509(3.0)	522(3.1)	
	Orphanage	217(0.4)	137(0.7)		71(0.4)	67(0.4)	
Drinking Experience	No	32481(61.4)	9013(47.0)	.000	8390(50.0)	8411(50.2)	.819
	Yes	20405(38.6)	10161(53.0)		8390(50.0)	8359(49.8)	
Smoking Experience	No	43940(83.1)	14153(73.8)	.000	12841(76.6)	12895(76.9)	.485
	Yes	8946(16.9)	5021(26.2)		3929(23.4)	3875(23.1)	
Stress	None	2355(4.5)	195(1.0)	.000	186(1.1)	177(1.1)	.947
	Rarely	10934(20.7)	891(4.6)		853(5.1)	880(5.2)	
	Sometime	25272(47.8)	5709(29.8)		5671(33.8)	5655(33.7)	
	Almost Always	11634(22.0)	7988(41.7)		7600(45.3)	7583(45.2)	
	Always	2691(5.1)	4391(22.9)		2460(14.7)	2475(14.8)	
Academic Grades	Excellent	6925(13.1)	1924(10.0)	.000	1767(10.5)	1769(10.5)	1.000
	Very Good	13615(25.7)	4259(22.2)		3899(23.2)	3898(23.2)	
	Good	15129(28.6)	5033(26.2)		4555(27.2)	4562(27.2)	
	Average	12168(23.0)	5136(26.8)		4403(26.3)	4403(26.3)	
	Poor	5049(9.5)	2822(14.7)		2146(12.8)	2138(12.7)	
Economic Conditions	Excellent	4240(8.0)	1372(7.2)	.000	1229(7.3)	1231(7.3)	.703
	Very Good	13855(26.2)	4478(23.4)		4034(24.1)	4022(24.0)	
	Good	26383(49.9)	8657(45.1)		7863(46.9)	7837(46.7)	
	Average	6850(13.0)	3503(18.3)		2876(17.1)	2855(17.0)	
	Poor	1558(2.9)	1164(6.1)		768(4.6)	825(4.9)	

허용오차를 0.0001로 설정하여 매칭한 결과 우울 집단과 비우울 집단은 각각 16,770명씩 매칭 되었다. 모든 변수에서 통계적으로 유의한 차이가 없으므로 나타났고, 우울여부에 따라서 인구학적 특성이 동일한 집단으로 매칭되었음을 알 수 있다.

### 3.3 모형적합도 검증

#### 3.3.1 Model summary

매칭 전 Model summary의 카이제곱 값은 3800.581, 매칭 후 Model summary의 카이제곱 값은 1465.805로 우도 비율 검사의 결과에 따른 매칭 후 로지스틱회귀

분석의 검정 통계량 값이 좋음을 알 수 있다.

#### 3.3.2 -2Log Likelihood

매칭 전 -2LL값은 15215.016, 매칭 후 -2LL값은 11074.165로 나타났고, -2LL의 값은 작을수록 좋은 모형이므로 매칭 후 로지스틱회귀분석 모형이 더 좋은 모형임을 알 수 있다.

#### 3.3.3 Hosmer-Lemeshow 검증

매칭 전 Cox와 Snell  $R^2$ 값은 .051, Nagelkerke  $R^2$ 값은 .222로 나타났고, 매칭 후 Cox와 Snell  $R^2$ 값은 .043, Nagelkerke  $R^2$  .137로 나타났다.  $R^2$ 값은 매칭 전  $R^2$

**Table 2** Logistic Regression the difference between before PSM and after PSM

		before PSM OR(95% CI)	after PSM OR(95% CI)
Melancholy	No	-	-
	Yes	8.140(7.220-9.176)***	6.299(5.462-7.265)***
Sex	Male	-	-
	Female	1.416(1.284-1.562)***	1.353(1.208-1.515)***
Residence Type	a	-	-
	b	2.233(1.654-3.015)***	1.692(1.066-2.686)*
	c	1.001(.785-1.277)	1.022(.747-1.397)
	d	3.844(2.758-5.359)***	3.038(1.873-4.926)***
Drinking Experience	No	-	-
	Yes	1.136(1.025-1.259)*	1.158(1.030-1.302)*
Smoking Experience	No	-	-
	Yes	1.838(1.647-2.052)***	1.735(1.523-1.976)***
Stress	None	-	-
	Rarely	.334(.241-.463)***	.297(.200-.443)***
	Sometime	.383(.291-.505)***	.210(.149-.295)***
	Almost Always	.647(.492-.850)**	.335(.240-.468)***
	Always	1.280(1.025-1.259)	.716(.510-1.003)
Academic Grades	Excellent	-	-
	Very Good	.801(.675-.952)*	.848(.698-1.032)
	Good	.825(.696-.977)*	.836(.688-1.016)
	Average	.958(.811-1.131)	.983(.812-1.191)
	Poor	1.079(.904-1.288)	1.071(.869-1.319)
Economic Conditions	Excellent	-	-
	Very Good	.594(.499-.708)***	.554(.453-.677)***
	Good	.520(.441-.613)***	.512(.424-.618)***
	Average	.572(.476-.688)***	.587(.475-.725)***
	Poor	.908(.735-1.122)	.929(.723-1.193)
Model summary		3800.581***	1465.805***
-2LL		15215.016	1465.805
Cox and Snell $R^2$ /Nagelkerke $R^2$		.051/.222	.043/.137
Hosmer and Lemeshow		253.704**	8.112

Residence Type a: With family, b: With relatives, c: Boarding house, dormitory, etc, d: Nursery facilities.

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

값이 매칭 후  $R^2$ 보다 높게 나오는 것으로 확인할 수 있으나, Hosmer와 Lemeshow에 의해 독립변수의 종속 변수에 대한 설명력을 자세히 파악하기 위해  $R^2$ 값 대신에, 매칭 전 Hosmer-Lemeshow p-value값이 .003으로 모형이 적합하지 않다는 대립가설을 채택하였고, 매칭 후 Hosmer-Lemeshow p-value값이 .423으로 모형이 적합하다는 귀무가설을 채택하였다. 따라서 매칭 후 로지스틱 모형이 적합해졌음을 알 수 있다.

#### 4. 결론 및 제언

본 연구는 ‘survey data: 2014년도 청소년건강행태’ 데이터를 이용하여 성향점수 매칭 전과 후의 로지스틱 회귀분석 결과를 비교하였다.

매칭 전 ‘우울 집단’은 19,174명, ‘비우울 집단’은 52,886명으로 나타났고, 매칭 후 ‘우울 집단’과 ‘비우울 집단’ 각각 16,770명씩 매칭 되었다. 매칭 후 우울

여부에 따라서 인구학적 특성이 동일한 집단으로 매칭되었음을 알 수 있다.

로지스틱 회귀분석 결과 첫째, Model summary의 카이제곱 값인 우도 비율 검사 결과 매칭 전보다 매칭 후 검정 통계량 값이 좋아짐을 알 수 있다.

둘째, -2LL은 값은 작을수록 좋은 모형이므로 매칭 후 로지스틱회귀분석 모형이 더 적합한 모형임을 알 수 있다.

셋째, 매칭 전 Hosmer-Lemeshow 값에 의해 ‘모형이 적합하지 않다’로 나왔고, 매칭 후 Hosmer-Lemeshow 값에 의해 ‘모형이 적합하다’는 귀무가설을 채택하였다. 따라서 매칭 후 로지스틱 모형이 적합해졌음을 알 수 있다.

따라서 로지스틱 적합도 검사인 3가지 방법 모두 매칭 후에 적합도가 좋아졌음을 알 수 있었다.

몇 가지 제언에 관하여, 로지스틱회귀분석은 5 미만의 빈도를 가지는 범주의 수가 많으면 적합도 지수에 심각한 왜곡이 생길 수 있기 때문에 표본의 크기가 커야 한다[16]고 하는데. 이것은 단순히 연구에서 수집된 전체 사례수가 많아야 한다는 것을 의미하는 것이 아니라 독립변수의 각 값에 대하여 빈도분포가 균일하게 분포되도록 계획적으로 표집해야 한다는 것이다.

또한 성향점수 매칭 분석도 허용오차 값이 작을수록 매칭되는 케이스 수가 줄어들기 때문에 표본의 수가 충분히 확보된 상태에서 매칭이 실시되어야 한다 또한 성향점수 매칭은 혼란변수에 의해 매칭되는 표본이 틀려지므로 사전에 충분히 문헌고찰을 하여 혼란변수를 충분히 확보해야 한다. 문헌고찰을 통해 검증되지 않은 혼란변수는 이후 로지스틱 회귀분석 시 많은 문제점을 야기 시킬 수 있다. 따라서 본 연구를 바탕으로 분석을 실시할 경우 표본의 수가 충분히 확보하여야 하고, 또한 문헌을 충분히 살펴본 뒤 성향점수 매칭방법을 사용하여야 하고, 그 이후 로지스틱 회귀분석을 실시하는게 바람직하다.

끝으로, 로지스틱 분석의  $R^2$ 값에 대해서, Menard (2000)은 여러 개의 측정치를 통해  $R^2$ 값을 검토하였으며, 최근 Paul D. Allison 교수[18]는 ‘statistical horizons’ 사이트에 ‘What’s the best R-squared for logistic regression?’ 이란 글을 올렸다. 이 글을 보면 로지스틱에 사용되는  $R^2$ 값에 대한 기술이 상세히 기술되어 있는데, Cox와 Snell의  $R^2$ 값과 Nagelkerke[13]이 제안한  $R^2$ 값 중에서 Cox와 Snell의  $R^2$ 값을 선호한다고 되어

있고, Nagelkerke의  $R^2$ 값은 Cox와 Snell의  $R^2$ 값을 수정하여 사용하는 값이며, Cox와 Snell의  $R^2$ 값이 가지고 있는 이론적 장점을 심각하게 훼손한다고 보고하였다. Paul 교수는 최근 Tjur[19]가 개발한 새로운  $R^2$ 을 제안하였는데 이것은 계산이 쉽다는 직관적인 장점이 있고, 상한값이 1.0이며, 선형 모형에서의  $R^2$ 값과 밀접한 관련이 되어 있다고 하였다. Tjur의  $R^2$ 값은 종속변수의 두 가지 범주 각각에 대해 사건의 예측확률의 평균값을 계산한 후 두 평균값의 차이를 구하는 것이다. 따라서 어떤 모형이 좋은 예측을 한다면, 사건이 발생한 사례들은 높은 예측값을 가져야 하고 사건이 발생하지 않은 사례들은 낮은 예측값을 가져야 한다. 또한 Tjur의  $R^2$ 의 공식은 산술평균과 동일하며, 잔차제곱에 기반한 기하평균과 동일하나 Tjur의  $R^2$ 값도 한계점이 있음을 알 수 있다. 그것은 순위로지스틱이나 다항로지스틱으로 일반화 할 수 없다는 것이다. 따라서 로지스틱 회귀분석에 대한  $R^2$ 값을 계산하는 방법은 많은데 어떠한 것이 최선인지에 대한 합의는 아직까지 나타나 있지 않다.

## References

- [1] Berk, R. A. (1983). “An Introduction to Sample Selection Bias in Sociological Data”. American sociological review, Vol. 48, No. 3, pp. 386-398.
- [2] No, S. Y. (2008). “Reassessment of risk factors for the development of liver cirrhosis based on propensity score matching methods”. Yonsei University, Seoul.
- [3] Rosenbaum, P. R and Rubin, D. B. (1983). “The central role of the propensity score in observational studies for causal effects”. Biometrika, Vol. 70, No. 1, pp. 41-55.
- [4] D’agostino, R. B. (1998). “Tutorial in Biostatistics Propensity Score methods for Bias Reduction in The Comparison of a Treatment to a Non-randomized Control Group”. Statistics in medicine, Vol. 17. pp. 2265-2281.
- [5] 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.
- [6] 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.
- [7] Lee, S. J. *et al.* (2007). "The use of propensity score matching for evaluation of the effects of nursing interventions". *Journal of Korean Academy of nursing*, Vol. 37, pp. 414-421.
- [8] Thabut, G. *et al.* (2008). "Survival after bilateral versus single lung transplantation for patients with chronic obstructive pulmonary disease: a retrospective analysis of registry data". *The Lancet*, Vol. 371, No. 9614, pp. 744-751.
- [9] Hiatt, W. R. (2006). "Observational Studies of Drug Safety-Aprotinin and the Absence of transparency". *New England Journal of Medicine*, Vol. 355, pp. 2171-2173 .
- [10] Kim, E. S. (2008). "National basic livelihood security system and labour supply". *Korean labor & Income panel study*, pp. 457-471.
- [11] Chai, G. M. (2013). "Advanced Statistic using SPSS and AMOS". Yangseowon.
- [12] Cox, D. R. and Snell, E. J. (1989). "Analysis of Binary Data". CRC Press.
- [13] Nagelkerke, N. J. D. (1991). "A note on a General Definition of the coefficient of Determination". *Biometrika*, Vol. 78, No. 3, pp. 691-692.
- [14] Hosmer, D. and Lemeshow (2000). "Applied Logistic Regression". New York: John Wiley & Sons.
- [15] Kwon, Y. M. and Kim S. Y. and Baek, J. I. (2015). "The Risk Factors Analysis of Adolescent Suicide due to Depression Experience". *Journal of Applied Reliability*, Vol. 15, No. 2, pp. 76-83.
- [16] Gottman, J. M. and Roy, A. K. (1990). "Sequential analysis: A guide for behavioral researchers". Cambridge University Press.
- [17] Menard, S. (2000). "Coefficients of determination for multiple logistic regression analysis". *The American Statistician*. Vol. 54, pp. 17-24.
- [18] Paul Allison (2013). <http://www.statisticalhorizons.com/r2logistic>.
- [19] Tjur, T. (2009). "Coefficients of determination in logistic regression models-A new proposal: The coefficient of discrimination". *The American Statistician* Vol. 63, pp. 366-372.