연구노트

성향점수를 이용한 무응답 보정 연구*

A Study on Nonresponse Adjustment by Using Propensity Scores

이계오**
Kay-O Lee

본 연구는 사회조사에서 무응답으로 인한 편향을 축소하는 방안으로 성향점수를 이용하는 방법과 사례를 설명하기 위해서 성향점수 방법의 이론적인 개념과 배경을 정리하였다. 또한 성향점수 방법을 처음으로 적용한 역학적인 관찰연구에서 성향점수 모형의 정의와 이론적 배경을 살펴보았고, 추정에서 편향의 축소방법으로 이용되는 3가지 성향점수 방법을 정리하였다. 성향점수로 짝짓기는 통제그룹의 데이터가 상대적으로 많을 경우에 이용되고 부차분류법은 통제그룹의 모든 데이터를 이용할 수 있으며, 회귀모형을 이용한 보정은 다중공변량에서도 사용할 수 있을 뿐만 아니라, 각 관찰단위에 성향점수 값을 산출하여 사용할 수 있는 특징이 있다. 그리고 사회여론조사에서 항목무응답으로 인한 편향을 축소하는 데 성향점수 가중법을 적용하는 절차를 제안하고 기존의 데이터를 이용하여 실제 적용에 대한 가능성을 검토하였다.

주제어: 성향점수, 성향점수가중, 관찰연구, 편향보정, 부차분류법

The propensity score method is used to minimize the bias level in social survey, which comes from nonresponse. The theoretical concept and the background of the propensity score method is discussed first. The propensity score method was first applied in the epidemiology observational study. I have summarized the process of the three propensity score methods that were used to reduce estimation bias in this study. Matching by propensity score is applied to the relatively large control group. Subclassification has the advantage of using whole control group data and regression adjustment is applied to multiple covariates as well as propensity score of each unit is computable and usable. Lastly, the application procedures of propensity score method to reduce the nonresponse bias is suggested and its applicability to real situation is reviewed with the existing data.

* 2008년도 한남대학교 교비학술연구조성비지원에 의하여 연구되었음(과제번호:2008A106)

** 한남대학교 정보통계학과 교수 이계오.
E-mail: kayolee@hnu.kr

Key words : propensity score, propensity score weight, observational study, bias adjustment, subclassification

Ⅰ. 서론

사회실태조사 또는 선거여론조사 등에서, 주요 문항에 대한 응답거부 또는 허위응답 등의 이유로 인해 모집단의 특성을 추정할 때 편향(bias)이 발생하는 사례는 다양한 분야의 통계조사에서 볼 수 있다. 특히 선거여론조사의 경우에는 지지후보를 묻는 문항이나 특정 후보의 지지 여부를 묻는 문항에서 응답을 거부하는 경우가 대부분의 조사에서 발생한다. 무응답한 사람을 제외하고 응답한 사람들의 데이터만으로 후보의 지지율이나 선호율을 추정한 경우, 만일에 응답을 거부한 사람과 응답한 사람들의 후보지지 성향이 다르다면 모수 추정은 편향으로 인하여 정확도가 떨어질 것이다.

특히, 선거여론조사는 주로 전화조사를 통해서 데이터를 수집하고 있는데 전화조사에서 발생하는 무응답은 두 종류이다. 조사 자체를 완전히 거부하거나 조사대상자를 접촉할 수 없는 경우와 같은 단위무응답(unit nonresponse)과 조사내용 중에서 일부 문항에 대한 응답을 거부하는 항목무응답(item nonresponse)으로 구분된다. 단위무응답의 특성이나 예방조치 또는 대처방법에 대해서는 많은 연구가 있었다. 항목무응답에 대해서도 통계조사에서 항목무응답을 최소화하는 방안과 항목무응답의 대체(imputation)방법 등에 관한 연구들이 다양한 분야에서 이루어졌다(조사통계연구회 2000). 그러나 기존의 무응답에 대한 대체 방법들은 응답자와 무응답자 간의 특성이 유사하다는 가정을 하고 있어서 특성이 다른 경우에 대한 연구는 미흡한 실정이다.

본 연구에서는, 기존의 연구와는 달리 무응답으로 인한 편향을 줄이는 성향점수(propensity score) 가중법에 관한 내용을 다루려 한다. 성향점수를 이용하여 단위 무응답에서 가중치를 보정하고, 항목 무응답에 대해서는 대체를 통해 완전한 데이터 세트를 구축하여 모수를 추정하는 방법을 연구한다. 실제로 사회여론조사에서 적용할 수 있는 성향점수 적용절차를 제시하고, 기존의 데이터를 통해 적용가능성을 수치적으로 살펴본다.

먼저 성향점수 방법에 대한 정의와 배경은 2절에서 설명하며 성향점수방법을 처음 사용한 역학분야에서 적용한 3가지 성향점수 방법의 절차와 개념을 3절에서 설명한다. 사회여론조사에서 무응답으로 인한 편향을 보정하는 데 성향점수 가중법을 적용하는 절차를 4절에서 제시하고 결론에서는 연구결과를 요약한 후에 사회여론조사 등에서 무응답으로 인한

편향을 축소할 때 사용하는 성향점수 가중법의 적용과정에서 해결해야 할 과제 등에 대해서 언급하였다.

Ⅱ. 성향점수의 정의와 배경

1. 성향점수의 정의

성향점수 또는 성향가중(propensity weighting)이라는 단어가 선거여론조사 분야에서 사용된 것은 2000년 미국대통령 선거여론조사의 예측방송이었다. 2000년 미국 대통령선거는 공화당 후보인 부시와 민주당 후보인 고어간의 경합이 치열하여 예측조사에서 혼란이 있었을 뿐 아니라, 투표에서도 전체적인 득표율은 고어가 앞섰지만 선거인단의 확보에서는 부시가 승리하였다. 미국 대통령선거제도에 대한 관심도 높았지만 사전 선거여론조사에서 결과를 얼마나 정확하게 예측했는지에 대한 논란이 있었는데 여기서 Harris 조사회사의 인터넷을 이용한 예측결과가 다른 어떤 조사방법에 의한 예측결과보다 정확했던 것으로 알려졌다. Harris 조사회사는 인터넷 사용자들로부터 자발적인 선거여론조사의 참여자를 모집한 후에 조사한 데이터의 특성과 전화조사의 특성을 비교·분석하여 성향가중모형을 추정하고 이를 사용하여 모집단의 특성을 추론하는 방법으로 선거결과를 예측했다(Traugott 2001).

사회여론조사의 목적 중의 하나는 그룹 간의 특성을 비교·분석하여 차이의 유의성과 원인을 규명하여 적합한 대책을 수립하는 것이다. 선거여론조사에서 1위와 2위 후보 간의 지지율의 차이가 통계적으로 유의한지 또는, 차이가 생긴 원인이 무엇인가를 분석하여 결과를 예측하는 것 등이다.

또다른 분야인 역학적 관찰 데이터의 비교·분석에서도 유의성과 원인을 규명하는 데 성향점수가 이용되고 있다. 예를 들면 폐암의 발병률과 흡연실태간의 인과적인 효과를 파악하기 위해서 흡연자 그룹과 비흡연자 그룹 간의 폐암발병률의 차이를 분석하게 될 것이다. 여기서 심층적인 분석을 하면 비흡연그룹에서의 폐암발병률은 흡연그룹의 폐암발병률보다 월등하게 낮기 때문에 폐암발병률을 줄이기 위해서 흡연율을 낮추어야 할 것이라는 데에는 이론(異論)의 여지가 없을 것이다. 그러나 다른 요인인 생활습관, 연령, 성별, 종족 또는 건강상태 등에 대하여 통제하지 않은 상황에서 두 그룹의 폐암발병률에 대한 비교·분석을 통한 인과해석은 통계학적으로 문제가 있을 수 있다. 이와 같은 관찰연구에서 범하기 쉬운

오류는 확률화를 통한 두 그룹의 동일한 실험조건을 만들지 않았기 때문에 특수한 여건의 현상을 전체 모집단으로 확대 해석할 때 생기는 것이다. 위의 경우에는 사람들에게 흡연하는 실험을 랜덤하게 지정하여 할 수는 없을 것이다. 만일에 인위적으로 랜덤하게 처리를 할당할 수 없다면 다른 요인들의 효과와 혼재되어 정확하게 인과분석을 할 수 없을 것이다. 그러나 혼재된 요인들을 다른 통계적인 방법으로 식별할 수 있다면 확률화 방법으로 처리하지 못한 실험데이터에서 혼재된 요인들의 효과를 통제하여 분석할 수 있는 방법 중의 하나가 성향점수법이다.

성향점수법은 관찰연구와 같이 사전처리변수들에 대한 그룹 간의 편향을 보정하는 방법으로 어렵지 않게 관찰할 수 있는 공변량변수(covariate)을 이용하는 것이다. 성향점수는 조건부확률 모형을 이용하여 아래와 같이 정의된다(Rosenbaum & Rubin 1983).

$$e(x_i) = \Pr(Z_i = 1 | X_i = x_i) \ , \ i = 1, \cdots, N \qquad (1)$$

여기서 Z는 처리변수(treatment variable)로서 흡연과 폐암발병률의 연구에서는 흡연여부를 나타내는 것이며 X는 공변량변수들의 벡터로서 관찰이 용이한 변수들로 구성된다.

성향점수가 같은 사람은 사전처리변수인 흡연여부와 상관없이 폐암발병률이 동일한 집단으로 구분할 수 있기 때문에 성향점수가 유사한 실험집단이나 통제집단은 동일한 환경을 갖는 것으로 가정할 수 있게 되어 흡연그룹과 비흡연그룹 간의 폐암발병률을 편향 없이 비교·분석할 수 있다.

식(1)에 주어진 성향점수식에서 공변량변수를 이용하여 성향점수를 계산하는 방법에 대한 연구는 별도의 과제이지만 본 연구에서는 로지스틱 회귀모형을 이용한 성향점수 계산방법에 대해서 간단하게 살펴보겠다. 먼저 성향점수 $e(x)$가 주어진 경우에 처리변수와 공변량 간에 조건부 독립이 성립한다고 가정한다면, 조사된 데이터에 대한 성향점수를 아래와 같이 표현할 수 있다.

$$\Pr(Z_1 = z_1, \cdots, Z_n = z_n \mid X_1 = x_1, \cdots, X_n = x_n) = \Pi_{i=1} e(x_i)^{z_i} \left(1 - e(x_i)^{1-z_i}\right) \qquad (2)$$

식(1)과 식(2)로 주어진 확률모형에서 성향점수 방법을 이용하기 위해서는 관찰된 데이터를 사용하여 성향점수 모형을 추정해야 한다. 원칙적으로 성향점수 모형은 전체 모집단으로부터 도출되어야하지만 현실적으로 실현되기 어렵기 때문에 표본조사 데이터를 사용하여 추정한 성향점수 모형을 적용할 수 있다(Rubin & Thomas 1992).

성향점수 모형을 추정하는 데 사용되는 추정법으로는 로지스틱 회귀모형, 프로빗 모형,

일반화 선형모형, 일반화 가법모형과 분류수형모형 등이 있다. 이들 중에서 가장 일반적으로 적용되는 모형은 로지스틱 회귀모형이다.

성향점수의 적용방법을 충분히 이해하기 위해서는 성향점수 모형의 생성배경을 살펴보는 것이 유용할 것이므로 성향점수 모형의 배경을 살펴보겠다.

2. 성향점수 모형의 배경

역학적인 관찰연구에서는 실험대상을 처리그룹과 통제그룹으로 랜덤하게 할당하는 것이 현실적으로 불가능하다. 관찰된 데이터와 공변량을 사용한 그룹 간의 특성 비교에서 초모집단(super－population)을 기반으로 한 편향 축소방법이 연구되었다. 처리효과의 특성은 처리그룹과 통제그룹 내의 개체들의 초모집단에서 이론적인 평균으로 정의할 수 있다. 즉 τ_1은 실험조건에서 각 개체들의 평균이고 τ_o는 통제조건에서 각 개체들의 평균이면 처리효과는 $\tau = \tau_1 - \tau_o$로 계산될 수 있다.

이론적으로는 i번째 개체의 처리효과는 $\tau_i = \tau_{1i} - \tau_{oi}$로 계산가능하고, 모든 개체들에 대한 전체적인 처리효과는 각 개체들의 처리효과의 평균으로 나타낼 수 있다. 그러나 실제 상황에서는, 유한모집단에서 처리효과를 각 개체별 처리효과 t_i들의 평균으로 계산할 수 있다.

이론적으로는 유한모집단에서 모든 개체들이 실험조건과 통제조건에 동시에 노출되어야 각 개체별로 처리효과를 계산할 수 있지만, 실제상황에서는 모집단에서 선정한 표본개체에 대해서만 관찰될 수 있고 이들 개체들은 처리조건 또는, 통제조건 중의 하나에만 노출될 것이다. 표본으로 선정되지 않은 개체에 대해서는 아무런 관찰값을 얻을 수 없다. 이론적인 상황에서 가정한 처리효과는 직접 계산할 수 없으므로 관찰된 데이터에서 추론되어야 한다. 두 개의 독립적인 단순임의표본들에 대해서 실험이 주어지고, 크기가 n_1인 표본에 대해서는 실험조건에서 관찰되고 크기가 n_o인 표본에 대해서는 통제조건에서 관찰된다면 처리효과의 추정치는 두 개 독립표본들의 평균들의 차이로 계산할 수 있다.

처리효과가 모집단의 특성을 반영하기 위해서는 관찰되지 않은 성분과 표본으로 선정되지 않은 단위들에 대한 가정이 필요하다.

먼저 모집단의 모든 단위들에 대해서 실험과 통제를 반복적으로 무한번 적용하는 것을 체계 M이라고 표기하자.

M 체계 하에서는 $E_M(t_1) = \tau_1$, $E_M(t_o) = \tau_{o,}$ $E_M(t_1 - t_o) = \tau$ 이 성립하고 기댓값 E_M은 체계 M에 대해서 계산되며 체계 M에서는 $N \to \infty$이 성립된다고 가정한다.

표본조사의 추정치 $\hat{t_1}$과 $\hat{t_o}$를 초모집단 특성치 τ_1과 τ_o의 근사값인 유한모집단의 특성치인 t_1과 t_o사이를 연계시키는 방안은 처리 할당을 랜덤하게 하는 것이다. 처리 할당분포를 π로 나타내면, $E_\pi(\hat{t_1})=t_1$, $E_\pi(\hat{t_o})=t_o$, $E_\pi(\hat{t_1}-\hat{t_o})=t$ 가 성립된다. 동일표본 내에 있는 i단위와 j단위 간에 종속되지 않는다는 조건 하에서 처리효과의 평균은 다음과 같이 나타낼 수 있다.

$$E_M E_\pi(\hat{t_1}-\hat{t_o})=\tau \tag{3}$$

여기서 $E_\pi(\cdot)$는 랜덤 할당체계에 대한 기댓값이고 $E_M(\cdot)$는 유한모집단과 초모집단 간의 관계로부터 주어진 기댓값이다.

효과 τ는 실제효과를 나타낸 것이 아니라 잠재된(intented) 효과이므로, 불완전한 확률화와 불완벽성 때문에 실제효과에는 의도하지 않은 효과가 포함될 수 있음에 유의해야 한다.

3. 공변량을 이용한 편향보정

연구대상에 할당된 처리특성보다 공변량 X에 대해서 체계적으로 다를 경우에는 처리그룹에 속한 i단위와 통제그룹 내의 j단위에 대한 관찰값을 아래와 같이 표현할 수 있다.

$$t_{1i}=\tau_1+u(X_{1i})+e_{1i},\ t_{0j}=\tau_0+u(X_{0j})+e_{0j} \tag{4}$$

단, $u(X)$는 공변량들의 행렬 X의 함수이고 e_{1i}와 e_{0j}는 평균 0인 랜덤오차이다. t_{1i}는 i번째 단위의 결과치로서, 동일그룹 내의 처리의 참값인 τ_1과는 $u(X_{1i})$와 e_{1i}만큼의 차이가 있으며, t_{0j} 관찰값에 대해서도 특성이 유사하다고 가정한다.

처리결과의 비교·분석에서 그룹별 특성을 고려해야 할 것이며, 그렇지 않을 경우에는 X_1과 X_0 분포의 불균형성이 비교결과에 혼재되어 있을 것이다. 따라서 공변량의 혼재효과를 조정하지 않으면 처리효과의 기댓값이 아래와 같이 편향될 것이다.

$$E_M(t_1)-E_M(t_0)=\tau_1-\tau_0+(\overline{u_1}-\overline{u_0})=\tau+(\overline{u_1}-\overline{u_0}) \tag{5}$$

단, $\bar{u}_1$와 $\bar{u}_0$는 각각 처리그룹과 통제그룹에서 공변량 함수 $u(X)$의 기댓값이다.

식(5)에서 기댓값은 단위들의 반복적인 처리결과로부터 계산되며, 기대효과 계산에서 처리효과와 공변량 간에는 교호작용이 없는 것으로 가정한다.

식(3)과 식(5)로부터 처리효과 추정에서 $|\overline{u_1}-\overline{u_0}|$ 만큼 편향이 있음을 볼 수 있다. 따라서 두 그룹 간의 공변량의 불균형을 보정한다면 편향이 제거되거나 축소됨을 알 수 있다. 공변량 X의 차원이 클 경우에는 공변량 분포의 불균형을 보정하는데 이론적으로는 동등한 분포를 갖도록 계산할 수 있으나 현실적으로는 용이하지 않다. 한 가지 대안은 성향점수와 같은 요약측도로 보정하거나 불균형을 조정하는 방법으로 모든 공변량들을 하나의 척도로 축약하는 것이다. 성향점수 보정은 이런 목적으로 사용하는 데 직관적이고 효과적인 방법이며, 최소의 가정조건 하에서 필요한 공변량의 정보를 이용한다면 각 단위별로 스칼라 값으로 산출할 수 있다.

Ⅲ 성향점수 보정 적용방법

성향점수를 이용하여 추정에서 편향을 축소시키는 방법 중에서, 짝짓기 방법·부차분류법·회귀모형법 등을 다양한 분야에 적용시킬 수 있도록 돕기 위해 그 내용들을 정리하면 아래와 같다.

1. 성향점수로 짝짓기

처리그룹에서 하나의 개체와 통제그룹에서 하나의 개체를 선택하여 성향점수를 기준으로 쌍으로 연결하는 것을 성향점수로 짝짓기(matching by propensity scores)라고 한다. 실험비용이 비싸고 통제조건에서 데이터를 대량으로 이용할 수 있는 역학관찰연구에서 짝짓기 방법이 많이 이용되고 있다. 그 이유는 역학관찰연구의 인과분석연구의 경우엔 실험조건에 노출된 사람들은 극소수이고 통제그룹에 속하는 사람들은 처리그룹에 비해서 월등하게 많기 때문이다. 짝짓기의 기본개념은 실험 처리된 모든 단위들과 공변량의 분포가 유사한 통제된 단위들을 짝짓는 것이다.

예를 들면, 단일 공변량 X을 이용하는 데 처리그룹 P_1의 모집단에서 랜덤 표본 S_1과 통제그룹 P_0의 모집단에서 크기 $m=kn(k \geq 1)$인 대규모의 랜덤표본 S_0가 있다고 가정한다(Rubin 1973). S_1과 S_0에 있는 모든 단위들에 대해서 공변량 X가 기록되었다고 가정하고, S_1에 있는 모든 단위들은 S_0에서 선정한 개체들과 X를 근거로 짝짓기를 한다. 먼저 S_0에서 크기 n인 부차 표본을 추출하고 이를 S_0^*로 표시하며 S_0^*내의 각 단위들은 X에서 동일한 값을 갖는 S_1에 있는 단위들과 짝을 짓는다. 처리효과는 S_1과 S_0^*에서

추정된다. 만일에 $k=1$이면, S_0^*는 P_0와 동일한 랜덤 본이므로 목적에 맞는 짝짓기는 불가능하다. 이런 경우에는 X의 불균형성 때문에 생기는 편향이 그대로 남는다. 만일에 $k \to \infty$이면, S_1과 S_0^* 간의 짝짓기가 가능하므로 편향은 축소될 수 있으나 제거될 수는 없다.

짝짓기를 위해서 S_0^*를 구성하는 절차는 $\{s_{1j}\} \in S_1$ 내의 모든 단위에 대해 $\{s_{0j}\} \in S_0$ 내의 연결되지 않은 단위들 중 X을 기준으로 가장 근접한 단위들을 연결한다. S_0^*의 선정체계는 $\{s_{0j}\}$의 순서를 정하는 방법에 의해서 결정되며 순서를 결정하는 방법은 기존연구에 소개되었다(Rubin 1973).

2. 성향점수를 이용한 부차분류법

처리그룹과 통제그룹의 모든 단위들을 한 집합으로 모은 후에 공변량의 분포를 기준으로 몇 개의 부차클래스로 나누고 동일한 부차클래스에 속하는 단위들은 이용하여 편향을 보정하는 방법이 부차분류법이다. 임상실험 자료분석에서 성향점수를 이용한 부차분류법(subclassification by propensity scores)의 적용 사례들이 많이 언급되었다(Rosenbaum & Rubin 1984; Lavori & Keller 1988). 부차분류는 짝짓기 보다는 쉽게 적용할 수 있고 통제그룹의 단위가 처리그룹의 단위보다 많을 필요가 없으며, 짝짓기에서는 통제그룹 내의 단위들 중에서 짝짓기에 사용하지 않은 것은 버리지만 부차분류법은 모든 단위들을 사용하기 때문에 이의 활용성은 매우 높다.

부차분류법을 적용하기 위해서는 모든 단위들을 X의 값에 의해서 정렬할 수 있고 c번째 부차클래스에 대한 X의 경계점을 X_{c-1}과 X_c라고 가정하자. c번째 클래스에 속하는 두 그룹 내의 관찰값들의 표본평균을 t_{1c}와 t_{0c}라고 한다면 실험그룹과 통제그룹으로부터 기대되는 결과는 다음과 같은 식으로 표현할 수 있다.

$$E_M E_\pi(t_{1c}) = \tau_1 + \bar{u}_{1c},\ E_M E_\pi(t_{0c}) = \tau_0 + \overline{u_{0c}} \tag{6}$$

여기서 $\bar{u}_{1c}$와 $\bar{u}_{0c}$는 각각 공변량의 함수 $u(X)$에 대한 기댓값을 의미하고 c는 부차클래스를 나타낸다.

처리효과 평균의 주된 편향은 $\bar{u}_1 - \bar{u}_0 = \sum_{c=1}^{C} (\bar{u}_{1c} - \bar{u}_{0c})$으로 표현할 수 있다.

한 부차클래스에 있는 모든 단위들은 성향점수 모형에 포함된 공변량을 기준으로 비교

할 수 있다. 각 부차클래스에 적절하게 가중치를 할당함으로써 전체적 처리효과를 보정할 수 있으며, 처리효과는 주어진 부차클래스에서 처리그룹과 통제그룹 간의 차이들의 가중평균이다. 예를 들면 처리그룹 또는 통제그룹에서 각 부차클래스의 구성비를 기준으로 각 부차클래스에 대한 가중치를 산출할 수 있다.

X의 분포적 차이에 대한 보정을 한 후에 남은 편향은 $\sum_{c=1}^{C} w_c(\bar{u}_{1c} - \bar{u}_{0c})$이다. 여기서 w_c는 c번째 부차클래스에 부여된 가중치이며 보정에 의해서 축소된 편향의 비는 아래 식으로 표현할 수 있다.

$$Q = 100 \times \left[1 - \frac{\sum_{c=1}^{C} w_c(\bar{u}_{1c} - \bar{u}_{0c})}{\bar{u}_1 - \bar{u}_0} \right] \qquad (7)$$

부차분류 보정에 의한 편향의 축소는 ① 공변량 $u(X)$의 함수, ② 분포함수 $\phi_0(X)$와 $\phi_1(X)$의 형태, ③ 부차클래스의 수, ④ X의 경계점과 ⑤ 가중치 w_c의 선택에 의해서 결정된다.

실제로 하나 이상의 공변량에 대해서도 편향축소는 유사하게 적용되지만, 다중공선성을 갖는 변수들에 대한 부차분류는 공변량의 개수 증가와 부차클래스의 증가에 따라서 부차클래스의 수가 기하급수적으로 증가하므로 쉽게 적용할 수가 없다. 다중공변량 대신에 성향점수의 이용은 모형에 포함된 모든 공변량을 대표할 수 있도록 부차클래스를 구성할 수 있게 된다.

추정된 성향점수를 이용한 부차분류에서도 단일변량 X에서 부차분류에 적용한 절차들을 적용할 수 있다. 각 부차클래스에서 모든 단위들이 동일한 성향점수를 갖도록 세분화한다면 무수히 많은 부차클래스를 생성할 수 있다. Cochran(1968)은 5개의 부차클래스만으로도 편향의 90% 이상을 제거할 수 있고, 5개 이상의 부차클래스에서는 편향축소가 크지 않다는 것을 언급하였다.

부차클래스의 크기가 거의 같도록 생성하는 것이 보다 합리적이며, 각 부차클래스 내에는 할당 무시가능성의 조건들을 충족하는 단위들이 최소한 하나 이상 있어야 한다. 또한 안정된 가중치를 가질 수 있도록 충분히 많은 관찰값들이 각 부차클래스에 포함되어야 한다.

부차분류법에 의한 성향점수 보정을 사후층화와 비교·분석한다면 표본조사 추정에서 주
된 오차를 $\left(\sum_{h=1}^{H} w_h \overline{y_h} - \sum_{h=1}^{H} W_h \overline{Y_h} \right)$로 표현할 수 있다. 여기서 $\overline{y_h}$와 $\overline{Y_h}$는 각각 h층에서 표본추정치와 모집단 특성을 나타내고 w_h와 W_h는 h층에서 표본과 모집단의 구성비이다. 사후

층화의 목적은 w_h와 W_h를 유사하게 만드는 것이므로 오차는 다음과 같다(Kish 1965: 92).

$$\sum_{h=1}^{H} W_h(\overline{y_h}-\overline{Y_h}) \tag{8}$$

식 (8)에 의하면 오차의 크기는 가중치 크기의 선택과 표본추정치와 모집단 특성 간의 차이에 관련되어 있다. 표본추출이 랜덤하다면 $(\overline{y_h}-\overline{Y_h})$는 편향이 무시될 수 있으며, 랜덤하게 표본추출이 이루어지지 않는다면 많은 편향이 남아 있을 것이다. 비록 $w_h = W_h$가 성립하더라도 $\overline{y_h}$가 $\overline{Y_h}$와 근사적으로 같다는 보장이 없기 때문에 오차는 클 수 있다. 성향점수 부차분류법에서는 $\sum_{c=1}^{C} w_c(\overline{u}_{1c}-\overline{u}_{0c})$를 최소화하는 가중치 w_c를 산출하도록 해야 한다. 만일에 처리그룹과 통제그룹에서 함수 $u(X)$에 포함된 공변량들이 다를 경우에는 전체적인 편향축소가 클 것으로 기대할 수 없다. 사후층화와의 차이점은 정형화된 모형을 적용하여 가중치를 계산하는 것이므로 성향점수를 기반으로 한 부차분류는 모형기반 사후층화라고 부를 수 있고 사후층화보다는 효율이 높은 것이 특징이다.

3. 성향점수를 이용한 회귀보정

성향점수를 이용한 회귀보정으로 처리효과의 편향을 축소할 수 있으며 이 방법은 반응의 기댓값을 아래와 같이 모형화하는 것이다.

$$E_M E_\pi(t_1)=\tau_1+\beta_1 x_1,\ E_M E_\pi(t_0)=\tau_0+\beta_0 x_0 \tag{9}$$

기대처리효과는 다음과 같이 회귀모형식으로 나타낼 수 있다.

$$E(r_1-r_0)=\tau_1-\tau_0+\beta_1 x_1-\beta_0 x_0=\tau+\beta(x_1-x_0) \tag{10}$$

회귀모형에서 두 회귀직선이 평행 $(\beta_1=\beta_0)$이면 처리효과의 편향은 $\beta(x_1-x_0)$로 나타낼 수 있고 $x_1=x_0$이면 편향은 제거될 수 있다.

다중공변량을 이용하게 되면 성향점수는 적절한 대안을 찾을 수 있으며 여기서는 단지 처리그룹과 통제그룹에서 성향점수에 관한 반응회귀식의 추정만이 필요하고 이 회귀모형을 이용하여 처리효과를 추정할 수 있게 된다. 식(9)에서 X 대신에 성향점수가 사용된다면 기대처리효과에서 성향점수가 주어진 경우에는 편향은 없어질 것이다.

Ⅳ. 성향점수 가중법을 이용한 무응답 보정 절차

앞에서 살펴본 성향점수 가중법을 사회여론조사의 항목무응답 보정에 적용하는 절차를 살펴봄으로써 실제 사회여론조사의 핵심문항에서 무응답으로 인한 영향이 심각할 경우에 정확한 모수추정에 도움이 되리라 생각된다.

본 절에서 성향점수 가중법을 항목무응답의 보정에 적용할 수 있는 세부적인 절차를 설명한 후에 Lorenc(2002)의 연구결과를 적용가능 사례로 보임으로써 실제 사회여론조사에서의 활용가능성을 보이고자 한다.

1. 조사설계와 조사항목 분석

사회여론조사에 적용한 조사과정과 조사항목을 파악하여 핵심변수 및 핵심변수와 상관성이 깊은 조사항목(보조변수 또는 공변량)들을 식별하고 특성을 파악한다. 조사에서 특정 후보에 대한 지지여부 또는 선호여부가 핵심변수인 경우에는 성별(X_1), 연령(X_2), 교육정도(X_3) 또는, 소득수준(X_4) 등이 보조변수로 고려될 수 있다. 여기서는 지지성향에 따라서 응답률이 다르고 보조변수들이 응답률에 영향을 미치고 있다고 가정한다.

조사설계에서는 모집단의 구조적 특성을 보조변수를 기준으로 교차빈도분석을 하고 핵심변수와 보조변수들 간의 관계를 파악하며 성향점수 가중법을 사용한 무응답 보정에 도움이 될 수 있도록 분석한다.

2. 핵심변수에서 무응답 특성 파악

지지여부를 묻는 문항에서 응답자와 무응답자의 특성이 보조변수와 어떤 관계가 있는지를 빈도분석이나 상관분석을 통해서 파악한다. 이를 기준으로 성향점수 모형에 사용할 보조변수를 선정한다. 예를 들어 성별과 연령대별로 무응답률에서 차이가 있는가를 파악한다. 만일에 특정 연령대에서 무응답률이 높게 나타난다면 연령대라는 보조변수가 응답성향에 영향을 주므로 결국은 지지성향에도 영향을 주게 된다. 응답한 데이터만을 이용하여 분석하면 모수추정에서 편향이 발생하게 될 것이다. 보조변수가 범주형 데이터로 주어진 경우에는 응답자와 무응답자에 대해 보조변수별로 빈도분석을 한다. 또는 보조변수들 간의 교차빈도분석을 통해서 이들 간에 발생하는 교호작용 효과의 유의성을 검증한다. 검증결과에 따라서 성향점수 모형을 분석할 때 포함되어야 할 보조변수를 선택하게 된다.

3. 성향점수 모형 적합

항목무응답 보정에서 무응답 대체층을 이용하는 대체법들이 자주 적용되고 있다. 무응답 대체층의 구성에서 다수의 보조변수를 사용한다면 층의 구분 작업이 쉽지 않을 것이다. 만일에 다수의 보조변수를 사용하는 경우에도 단일차원의 성향점수값을 사용할 수 있다면 편리할 것이다. 단일차원의 성향점수 값을 기준으로 구분한 층을 무응답 보정에 사용하기 위해서는 다음 가정을 충족해야 한다. 동일 층에 속한 조사대상자는 보조변수들이 어떤 값을 갖더라도, 지지여부의 문항에 대한 응답여부에 관계없이 동일한 지지성향을 갖는다고 가정한다.

우선 성향점수의 모형으로는 Rosenbaum & Rubin(1984)이 조건부 확률로 정의한 모형을 고려하자. 모형에 사용할 공변량 변수는 성별(X_1), 연령(X_2), 교육정도(X_3)와 소득수준(X_4) 등으로 하고 반응변수는 응답여부를 나타내는 변수(Z)로 한다. 성향점수는 공변량 값이 $X=x$로 주어졌을 때 지지여부를 묻는 문항에서 무응답일 조건부 확률로 아래 식으로 나타낸다.

$$e(x)=\Pr(Z=1|X=x)$$

모집단의 성향점수를 기준으로 X의 분포가 균형화 되도록 조사대상자들을 층으로 분류한다. 즉 $e(x)$가 동질적인 층(subclass) 내에서는 응답자이거나 또는 무응답자이건 관계없이 공변량 X의 분포가 동일하다는 의미이다.

조건부 확률모형에서 공변량 변수 X와 처리상태 변수 Z는 성향점수 $e(x)$가 주어졌다는 조건에서 독립이므로 아래와 같이 표현할 수 있다.

$$\Pr(X,Z|e(x))=\Pr(X|e(x))\cdot\Pr(Z|e(x))$$

공변량 변수들이 주어지고 조사대상자가 지지여부 문항에서 무응답일 확률이 성향점수이다. 이는 Z에 대한 공변량 X의 로지스틱 회귀모형으로부터 추정할 수 있다.

공변량 변수들이 다수인 경우에는 적절한 변수선정법으로 로지스틱 회귀모형에 포함될 공변량 변수들을 선택하고 1차 교호작용항까지 고려한다.

최종적으로 적합한 로지스틱 회귀모형은 연령과 교육정도 및 1차 교호작용항도 포함된 것으로 가정하면 다음 식으로 표현할 수 있다.

$$\log\left(\frac{e(x)}{1-e(x)}\right)=\hat{\alpha}+\widehat{\beta_2}X_2+\widehat{\beta_3}X_3+\widehat{\beta_{23}}X_2\cdot X_3 \tag{11}$$

4. 성향점수에 의한 층 구분

최종적으로 적합한 성향점수 모형은 식 (11)에 주어진 로지스틱 회귀식을 사용하여 아래와 같이 나타낼 수 있다.

$$e(x) = \frac{\exp[\hat{\alpha} + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \widehat{\beta_{23}} X_2 \cdot X_3]}{1 + \exp[\hat{\alpha} + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \widehat{\beta_{23}} X_2 \cdot X_3]} \tag{12}$$

모든 조사대상자에 대한 성향점수는 보조변수들의 값을 식(12)에 대입하여 산출하고 이를 기준으로 층을 분류한다. 층이 세분화되면 층 내의 성향점수값들이 유사하게 되지만 층에 포함될 대상자 수가 적을 수 있기 때문에, 기존 연구결과에 의하면 4개 또는 5개의 층으로 구분하는 것이 적절하다. 통상적으로는 성향점수 분포의 4분위수를 이용하여 조사대상자들을 4개의 계층으로 구분한다. 이는 모집단 분포의 4분위수와 경험적 분포의 4분위수가 일치하지 않을 경우에 가중치를 사용하여 이들의 분포를 조정할 수 있기 때문이다. 여기서 사용된 가중치를 성향점수 가중치라 하며, 가중치는 무응답자에 대해서는 모든 층에서 동일한 값이지만 응답자 층에서는 응답자 수에 따라서 다른 값을 갖게 된다. 응답자에 대한 가중치는 해당층에 속한 응답자들이 동일한 가중치를 갖는다는 가정에서 무응답층의 가중치를 층에 포함된 응답자 수로 나누고 무응답자 수를 곱하여 산출한다.

층 구분에 대한 정확성은, 응답자와 무응답자로 나누어 실제 조사된 데이터를 경험적으로 분류하였을 때의 층의 일치 정도에 따라 식(12)의 성향점수 모형의 적합성을 판단한다.

성향점수 모형을 이용한 층 구분결과를 아래 표와 같이 정리하여 가중치 계산과 모수추정에 이용한다.

〈표 1〉 성향점수 기준층 분류결과

성향점수 구간	경계 값	응답자	무응답자	합 계
$0.0 \sim ps_1$		n_{11}	n_{21}	n_{+1}
$ps_1 \sim ps_2$	q_1	n_{12}	n_{22}	n_{+2}
$ps_2 \sim ps_3$	q_2	n_{13}	n_{23}	n_{+3}
$ps_3 \sim 1.00$	q_3	n_{14}	n_{24}	n_{+4}
합계		n_{1+}	n_{2+}	n

〈표 1〉에 주어진 경계값은 무응답자의 빈도가 모든 층에서 가능한 한 균등하도록 층을 구분하는 성향점수값을 의미하고 응답자와 무응답자는 조사 데이터에서 계수한 것이다.

5. 성향점수 가중치를 이용한 모수추정

동일한 층에 속한 조사대상자들은 핵심변수에서 동일한 특성을 갖고 있다는 가정에서 표본구성비와 모집단구성비를 비교하여 가중치를 계산한다. 이는 일종의 사후층화법을 사용한 모수추정법으로 볼 수 있으며 가중치는 무응답자와 응답자를 구별하여 계산한다. 무응답자를 기준으로 한 층별 성향점수의 가중치는, 〈표 1〉에 주어진 4개 층 모두가 동일한 가중치를 갖도록 층을 구분했다면, 모집단규모를 무응답자 수로 나눈 값이다. 응답자의 가중치는 포함된 층의 무응답자의 가중치를 응답자 수로 나누고 여기에 무응답자 수를 곱하여 산출한다.

가중치 계산을 수식적으로 설명하기 위해서 모집단크기를 N이라 하면 〈표 1〉에서 h층의 무응답자와 응답자에 대한 가중치는 각각 아래와 같다.

$$W_{2h} = \frac{N}{n_{2+}}, \qquad W_{1h} = \frac{W_{2h} \cdot n_{2h}}{n_{1h}} \tag{13}$$

응답자의 층과 무응답자의 층의 가중치를 모두 합하면 2N이 된다. 핵심변수인 특정 후보의 지지여부를 묻는 문항에 대한 응답자층의 평균과 각 층의 성향점수 가중치를 사용한 모수추정식은 아래와 같다.

$$\widehat{P} = \sum_{h=1}^{H} \frac{n_{+h} W_{+h}}{2N} \cdot \overline{y_h} \tag{14}$$

여기서 W_{+h}는 h층에서 무응답자와 응답자를 통합한 가중치이며, h층에서 무응답자 수와 응답자 수 및 각각의 성향가중치의 관계식 $n_{+h} W_{+h} = n_{1h} W_{1h} + n_{2h} W_{2h}$ 에서 계산한다. 그리고 $\overline{y_h}$는 h층에서 응답자의 표본평균이다.

식(14)에 주어진 추정량의 분산추정은 사후층화추정법의 분산추정식을 사용하여 계산할 수 있다.

6. 성향점수 가중치의 적용사례

앞에 설명한 성향점수 가중법을 항목무응답 보정에 적용할 수 있는지에 대한 사례로 Lorenc(2002)의 조사데이터에서 계산된 결과를 살펴보겠다.

모집단의 규모는 6,865,519명이고 표본규모는 1,784명이며, 이 중에서 응답자는 1,581명, 무응답자는 203명이다. 핵심변수는 지지여부를 묻는 문항이며, 보조변수로는 성별(X_1), 연령(X_2), 교육정도(X_3)와 소득수준(X_4) 등이 고려되었다. 응답자와 무응답자간의 성별, 연령별 분포에서 유의한 차이가 있고, 무응답자와 응답자의 지지성향이 상이한 것으로 가정하자. 성향점수 모형을 지지여부에 대한 응답여부와 보조변수를 사용하여 로지스틱 회귀모형에 적합한 결과가 아래 식과 같다.

$$\log\left(\frac{e(x)}{1-e(x)}\right) = -2.192 + 0.228X_2 - 2.359X_3 + 0.275X_2 \cdot X_3 \tag{15}$$

식(15)에 주어진 적합 로지스틱 회귀모형을 이용하여 1,784명의 지지여부를 예측한 결과는 실제 응답한 1,581명 중에서 1,560명은 정확하게 예측하였으나 21명은 틀리게 예측하였다. 무응답자인 203명 중에서는 154명은 정확하게 무응답자로 예측하였으나, 24.1%인 49명은 틀리게 예측하였다. 정확도가 96.1%이므로 식(15)에 주어진 모형은 타당성이 있는 것으로 생각할 수 있으므로 이를 이용한 성향점수 모형은 아래 식으로 나타낼 수 있을 것이다.

$$e(x) = \frac{\exp[-2.192 + 0.228X_2 - 2.359X_3 + 0.275X_2 \cdot X_3]}{1+\exp[-2.192 + 0.228X_2 - 2.359X_3 + 0.275X_2 \cdot X_3]} \tag{16}$$

식(16)으로 주어진 성향점수 모형을 이용하여 1,784명에 대한 성향점수를 계산하고 이를 기준으로 무응답자의 분포가 같도록 층을 구분한 내용이 〈표 2〉에 주어졌다.

〈표2〉 성향점수에 의한 4개 층 구분

성향점수구간	경계값	응답자	무응답자	합 계
0~0.49362		1,560	49	1,609
0.49363~0.55055	0.50	5	48	53
0.55056~0.70398	0.57	5	57	62
0.70399~1.00	0.75	11	49	60
합 계		1,581	203	1,784

〈표 3〉 층별 계산된 성향점수 가중치

	층1	층2	층3	층4	합계
가중치	33820.291	33820.291	33820.291	33820.291	
빈도	49	48	57	49	203
합계	1657194.241	1623373.951	1927756.566	1657194.241	6865519
가중치	1062.304	234676.790	385551.313	150564.022	
빈도	1560	5	5	11	1581
합계	1657194.241	1623373.951	1927756.566	1657194.241	6865519

〈표 2〉에 주어진 무응답자 층을 기준으로 계산한 가중치는 33,820.291이고, 응답자에 대한 가중치는 식(13)을 사용하여 계산하였으며 결과는 〈표 3〉에 정리하였다.

〈표 3〉과 식(14)을 이용한 지지여부에 대한 모수추정치와 분산의 추정값은 각각 0.5892와 0.5386×10^{-3}으로 주어졌다.

이상에서 사회여론조사에서 항목무응답에 대한 편향을 보정하기 위해서 식(12)으로 주어진 성향점수 모형을 적용할 수 있음을 보였다.

Ⅴ. 결론

사회의식조사와 선거여론조사 등에서 개인주의의 팽배 등으로 인해 아예 응답자체를 거부하고 거짓으로 응답하거나 또는, 핵심적인 문항에서 응답하지 않는 등의 요인 때문에 정확한 응답을 얻지 못하거나 결측값으로 인한 항목무응답에 대한 보정방법 중의 하나로 성향점수 가중치를 적용하는 방법에 대해서 살펴보았다. 성향점수 가중치를 이용한 추정의 편향을 보정하는 방법에 대한 종합적인 설명을 위해서 성향점수의 개념을 공변량의 조건부 확률을 이용하여 정의하였다. 성향점수가 처음에는 처리집단과 통제집단 간의 비교·분석에서 잠재적인 편향을 제거하여 정확하게 인과관계를 설명하는 데 사용할 목적으로 연구되었다는 역사적인 배경을 세부적으로 설명하여 앞으로 성향점수 가중치의 적용을 다양한 분야와 주제로 확대할 수 있도록 하였다.

역학분야에서 처리그룹에 대한 데이터는 수집도 제한적이고 비용이 과다하게 소요되는 반면에 통제그룹의 데이터는 상대적으로 용이하게 확보할 수 있다. 또는 다량의 DB형태로 축적될 수 있기 때문에 처리의 효과를 정확하게 측정하기 위해서는 처리그룹과 통제그룹의 비교·분석을 통해서 인과관계를 해석할 수 있어야 한다. 하지만 처리그룹과 통제그룹의 단

순한 비교·분석으로는 다른 변량이나 환경 등의 영향으로 편향이 처리효과와 혼재될 수 있음을 초모집단과 유한모집단의 모형을 이용하여 이론적으로 설명하였다. 혼재된 편향을 축소 또는, 제거하기 위해서 성향점수 가중치를 적용하는 방법으로 성향점수로 짝짓기, 성향점수를 이용한 부차분류법과 성향점수를 이용한 공분산 및 회귀보정 등의 세 가지를 설명하였다.

사회여론조사에서 항목무응답은 발생할 수밖에 없다. 응답자와 무응답자의 연구변수에 대한 특성에서 차이가 있을 경우에 무응답을 무시하고 응답한 데이터만으로 추론할 경우에는 모수추정에서 편향이 발생할 수 있다. 이를 축소하기 위한 방안으로 공변량 변수를 이용한 성향점수 보정법의 절차를 설명하고 수치적인 사례를 통해서 선거여론조사 또는 사회실태조사에서 항목무응답으로 인한 편향을 보정할 수 있다는 것을 실증적으로 보임으로써 성향점수 가중치 보정법의 적용에 대한 관심을 높이는 계기가 되었다.

그러나 성향점수 가중치를 편향 보정에 사용하기 위해서는 성향점수 모형의 정의에서 전제조건은 주어진 공변량에 대해서 처리그룹의 성향점수는 조건부 독립성을 만족시켜야 한다. 만일에 조건부 독립성이 성립하지 않는다면 성향점수 모형의 로지스틱 회귀모형의 형태가 선형으로 주어질 수 없다. 따라서 조사된 데이터를 이용하여 적합한 로지스틱 회귀모형은 최적의 모형이 되지 않을 수 있다. 이에 대한 명확한 검증이나 확인 없이 단순하게 조건부 독립성이 충족된 것으로 가정한 연구나 적용사례가 대부분이다. 그러나 정확한 적용을 위해서는 반드시 주어진 공변량의 값에서 성향점수는 조건부 독립성이 성립한다는 것을 검증하고 사용해야 한다. 앞으로 연구되어야 할 핵심과제는 주어진 공변량에서 성향점수가 조건부 독립성 모형으로 표현될 수 있는 공변량 변수를 찾아내는 것이다. 조건부 독립성을 만족할 수 있는 공변량에 관한 문항을 개발하는 연구는 질문하는 문항의 내용에 따라 다를 수 있으므로 시행착오적인 방법이나 이론적인 모형을 통한 방법 등으로 추진되어야 할 것이다.

성향점수 가중치법이 사회여론조사뿐만 아니라 역학적인 관찰연구에서도 활성화되어 열악한 조사환경에서 생길 수 있는 편향을 축소함으로써, 정확하고 시의성이 높은 통계를 생산하여 활용함으로써 지식기반 정보화 사회로 발전하는 데 기여할 수 있기를 바란다.

참고문헌

조사통계연구회. 2000.《무응답 오차》. 서울. 자유아카데미
Cochran, W.G. 1968. " The Effectiveness of Adjustment by Subclassification in Removing Bias

in Observational Studies." *Biometrics* 24 : 295−313.

Kish, L. 1965. *Survey Sampling*. New York : John Wiley and Sons.

Lavori, P.W. and Keller, M .N. 1988. "Improving the Aggregate Performance of Psychiatric Diagnostic Methods When Not All Subjects Receive the Standard Test." *Statistics in Medicine* 7: 727−737.

Lorenc, B. 2002. "A Demonstration of Some Weighting Techniques." Unpublished Research Paper.

Rosenbaum, P.R., and Rubin, D.B. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1) : 41−55.

Rosenbaum, P.R., and Rubin, D.B. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79(387) : 516−524.

Rosenbaum, P.R., and Rubin, D.B. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39(1) : 33−38.

Rubin, D.B. 1973. "Matching to Remove Bias in Observational Studies." Biometrics 29(1) : 159−183.

Rubin, D.B., and Thomas, N. 1992. "Characterizing the Effects of Matching Using Linear Propensity Score Methods with Normal Distributions." *Biometrika* 79(4) : 797−804.

Traugott, M.W. 2001. "Assessing Poll Performance the 2000 Campaign." *Public Opinion Quarterly* 65(30) : 389−419.

[접수 2009/1/28, 1차수정 2009/2/12, 2차수정 2009/2/25, 게재확정 2009/3/2]