

## 제5장

# 사회조사에서 자원자표본 인터넷조사 추정

김서영 · 안다영  
권순필 · 이승희

### 제1절 서론

#### 1. 연구배경

최근 들어 웹은 모집단 조사를 위한 현실성 있는 대안으로 주목받고 있다. 과거 수십 년에 걸쳐 조사 환경은 빠르게 변하고 있다. 종이조사표에 응답을 기입하는 전통적인 대면조사 방식 (paper and pencil interviewing: PAPI)은 컴퓨터의 발전과 함께 컴퓨터를 보조 도구로 활용하는 조사방법 (Computer assisted interviewing: CAI)으로 바뀌고 있다. 최근 들어 조사기관들은 컴퓨터를 이용한 대면조사 (CAPI), 전화조사 (CATI), 우편조사 (CASI)에서 웹 조사 방법으로 점점 전환하고 있다. 개념적 측면에서 이상적인 조사방법은 그 자료 수집 방법이 당해 시점에서 사회현상을 잘 반영할 수 있을 때 그 기능을 발휘할 수 있게 된다. 왜냐하면 조사방법은 조사하고자 하는 목적 변수들에 대해 종속적이기 때문이다.

웹 조사 방법의 대중화는 그리 놀라운 것은 아니다. 웹 조사는 다수의 응답 가능한 그룹에 대해 쉽게 접근할 수 있는 간단한 방법이다. 조사에 필요한 질문지는 매우 저렴한 비용으로 응답자에게 배부될 수 있다. 즉, 조사표 작성을 위한 조사원이 별도로 필요하지 않고, 우편 발송이나 질문지 인쇄에 들어가는 비용이 필요하지 않다. 이처럼 웹 조사는 조사에 소요되는 경제적·시간적 비용 측면에서 큰 이점이 있다. 뿐만 아니라 웹 조사는 다양한 멀티미디어 (그림, 애니메이션, 동영상 등)를 사용할 수 있다는 장점이 있다. 그러나 이러한 많은 장점에도 불구하고 웹 조사는 조사방법으로서 과학적 타당성에 대한 논쟁을 일으켰다 (Couper, 2000; Fricker 와 Schonlau, 2002; 등). 특히, 비확률표본에 기반 (non-probability based)한 웹 조사는 응답 대상자가 랜덤하게 선택되지 않고 목표 모집단 (target population) 자체가 확률표본이 아닌 편의 (convenience)표본으로 구성되기 때문에 문제시되고 있다 (Stephanie 와 Kea, 2010).



이처럼 자원자표본을 이용한 웹 조사 추정치는 과소포함, 비확률표본추출 및 무응답 편향으로부터 자유롭지 못하다. 이러한 편향을 줄이기 위해 성향점수 (propensity score) 조정을 이용한 가중치 방법이 이 분야 연구에 등장하였다 (Lee, 2004, 2006; Schonlau 등, 2004; Taylor 등, 2001). 이러한 방법들을 통해 편향들을 없애거나 줄일 수 있다는 관점에서 자원자표본을 이용한 웹 조사 (volunteer sample based web survey)는 실제 자료를 수집해야 하는 실무자들에게 큰 관심이 되고 있는 것 같다. 그러나 높은 관심만큼 대규모 조사 또는 정확성을 요구하는 조사에서는 비확률표본의 활용에 대한 우려도 높다. 네덜란드에서는 주요 여론조사에 자기선택 (self selected)적으로 조사에 참여하는 웹 패널을 사용하기도 하였다. 이 웹 패널 조사는 패널에 대해서 인구학적 변수에 대한 정보를 수집하기 때문에, 이들 변수들에 대한 웹 조사와 모집단과의 비교가 가능하다. 따라서 가중치 방법을 통해 특정 그룹에 대한 과소 (under-coverage) 또는 과대 (over-coverage) 포함에 따른 대표성 문제를 조정하고자 하였다. 그러나 지금까지 네덜란드에서 자기선택적 웹 조사가 정부에서 실시하는 공식통계 (official statistics)조사에 사용된 적은 없는 것 같다.

이러한 국제 사회 움직임과 함께, 한국 통계청에서는 2008년을 기점으로 자원자 패널을 이용한 웹 조사에 관심을 갖게 되었고, 이들 활용 방안을 찾기 위한 연구가 2009년에 처음 시도되었다 (통계청, 2009 용역보고서). 이 연구는 하나의 예로 국민의식조사에 관련한 사회조사에 그 적용 가능성을 모색하였다. 이 연구는 사회조사 자료에 대해 성향가중 (propensity weighting) 조정을 함으로써 기준이 되는 확률표본을 이용한 대면조사와 자원자표본을 이용한 웹 조사 간 추정치의 차이를 어느 정도 줄일 수 있다는 것을 보였다. 또한 림가중치 (rim weighting)와 성향점수 가중치 방법에 따른 조정 효과 차이는 없다는 것과 성향 조정변수에 따라 조정 효과는 다를 수 있다는 사실을 제시하였다. 뿐만 아니라, 국가통계에 비확률표본에 기반한 웹 조사 방법을 활용하기 위해서는 꾸준한 연구가 있어야 한다는 점을 강조하기도 하였다. 이로부터 2009년 연구에 이어 2010년 자료를 이용한 추가 및 확장 연구에 대한 필요성이 제기되었다. 본 연구는 사회조사에서 자원자표본 웹 조사의 활용 가능성을 찾기 위해 이론적, 경험적 연구를 시도하였다.

## 2. 연구 내용 및 방법

자원자 패널 웹 조사는 지금까지 언급한 바와 같이 많은 장단점을 가지고 있다. 뿐만 아니라, 지금까지 연구자들이 검토한 문헌 결과에 따르면, 국제적으로 볼 때 국가통계를 주도하는 통계 기관에서 자원자표본을 이용한 웹 조사 방법을 적용한 사례는 없는 것 같다. 그렇지만, 점점 어려워지는 조사환경, 신속한 정보 수요의 증가 및 조사비용에 대한 부담이 증가하면서 국가통계 기관에서도 자원자 패널 웹 조사에 대한 관심이 높아지

고 있는 것 같다. 특히, 학술계에서는 자원자 패널을 이용한 웹 조사에서 발생하는 다양한 원인의 편향을 극복하기 위한 방법론 연구가 꾸준히 진행되고 있다. 또한, 실용성 측면에 관한 다양한 경험적 연구결과들도 제시되고 있다. 특히 가중치 조정 (weighting adjustment)을 중심으로 자원자 패널 웹 조사의 편향을 감소시킬 수 있는 방법을 찾고 있지만, 모든 자료에 획일적으로 적용될 수 있는 방법이 없는 것 같다. 즉, 자료의 성격과 형태에 따라, 자료에 포함된 변수에 따라, 성향변수 선택에 따라 그 조정 효과는 매우 다르게 나타날 수 있다.

본 연구의 주목적은 2009년 연구 결과에 이어 (통계청, 2009), 2010년 사회조사 자료에 성향점수 가중치를 적용하고 그 효과를 검증하는데 있다. 그리고 성향점수 가중치 적용을 통해 웹 조사의 편향을 줄일 수 있는 지 여부를 살펴보고자 한다. 본 연구는 편향을 줄이기 위한 방법으로 기존 연구들과 마찬가지로 성향점수 조정 방법을 사용하였다. 사회조사는 매년 5개의 다른 주제에 대해 100여 개 이상의 항목에 대해 조사하고 있다. 또한 사회조사가 국민의식조사와 관련되어 있고 해당 가구원 모두 응답해야 된다는 점에서 응답률이 그다지 높은 편은 아니다. 이런 현실적 측면을 고려하여 웹 조사 방법의 실용성을 판단할 필요가 있을 것이다. 성향점수를 이용한 가중치 조정 방법의 관심사항들은 확률표본에 기반한 대면조사와 자원자표본 웹 조사 간의 차이를 얼마나 줄일 수 있을 것인가, 가중치 조정을 통해 비확률표본 웹 조사 추정치의 신뢰성을 얼마나 유지할 수 있을 것인가 하는 것이다. 이를 위해 본 연구는 다양한 방법론적 접근과 경험적 분석을 시도하였다.

연구목적 달성을 위한 구체적인 방법은 다음과 같다. 연구 과정은 1), 2), 3), 4)의 순서에 따라 진행되었다.

#### 1) 문헌 연구

- 자원자 패널 웹 조사의 특성을 중심으로 검토한다. 특히, 자원자 패널 웹 조사에서 기인한 자기선택 편향 (self selection bias), 비포함 편향 (non coverage bias), 무응답 편향 (nonresponse bias)이 이러한 웹 조사 추정에 어떤 영향을 주는지를 파악한다.
- 웹 참여 성향가중치 조정에 관한 방법론을 검토한다. 특히, 성향점수를 이용한 가중치 조정 방법의 특성을 파악한다.

#### 2) 성향변수 탐색

- 성향변수는 성향가중치 방법을 적용할 때 필요한 성향점수 계산에 이용된다.
- 성향점수를 이용한 가중치 조정에서 가장 중요한 핵심은 어떤 가중치 방법을 적용했느냐보다는 어떤 성향변수를 사용했는가 하는 것이다. 성향 조정변수 선정은 주로 전문가적 견해, 경험적 노하우, 통계적 분석에 근거한 과학적 방법 등이



이용되고 있다. 본 연구에서는 통계적 분석을 원칙으로 하되 전문가적 의견과 조사항목 및 공표 성격 그리고 경험적 노하우 등을 토대로 성향변수를 선택하였다. 통계적 방법은 1차적으로 조사방법 (대면조사 vs. 웹 조사)과 관심변수 (각 조사항목) 간의  $\chi^2$  분석을 통해 조사방법 간에 유의한 차이가 있는 변수를 선정하였다. 그리고 2차적으로 이들 변수를 로지스틱 회귀모형에 포함시켜 조사방법에 영향을 미치는 변수를 선정하였다 (유의수준 0.05). 이렇게 선정된 변수는 그 변수의 공표 가능성과 범위, 무응답 정도 등을 고려하여 몇 개 변수는 배제하기도 하였다.

### 3) 가중치 조정 방법 적용

- 성향점수를 이용한 성향점수 모형 (propensity score model)을 사용하였다.
- 성향점수 모형을 이용할 때, 본 연구에서 사용된 기본가중치 (base weight)는 “웹 패널 자원자 수/웹 표본 수”를 사용하였다. 이때 기본가중치 계산은 지역별 (16개 시도), 성별 (남/여), 연령별 (20대 이하, 30~40대, 50~60대, 60대 이상) 층을 사용하였다. Rim 가중치도 기본가중치로 사용해 보았다.

### 4) 사후 Calibration

- 웹 조사의 편향 조정을 목적으로 가중치를 적용할 경우, 가중치 적용 후에 대체로 인구·지리적 구성 분포가 흐트러지게 된다. 따라서 이를 다시 조정할 필요가 있다 (Lee, 2009). 본 연구에서는 이를 위해 성향조정 가중치가 붙은 웹 표본에 rim 가중치를 추가로 적용해 보고자 하였다.
- 이것은 일종의 칼리브레이션 절차로 (Deville 와 Sarndal, 1992), 2차 조정을 목적으로 사용된다. 칼리브레이션 절차는 조정된 웹 조사 표본이 목표 모집단과 유사하도록 보정해 준다. 칼리브레이션은 분산을 줄일 수 있고 성향조정에 의해 다루어지 않았던 무응답과 미포함 오차를 조정할 수 있다 (Kim, Li, 와 Valliant, 2007; Kott, 2006)고 알려졌다.
- 성향점수 방법으로 웹 표본의 공변량 (covariate) 분포가 준거조사에 근사하도록 조정하더라도 준거조사 자체는 여전히 칼리브레이션으로 조정할 수 있는 미포함 오차에 영향을 받을 수 있다. 이는 특히 웹 조사가 총수 (totals)를 추정할 때 중요하다. 칼리브레이션에 사용되는 공변량들은 성향점수 추정에 사용된 변수와 같지만, 반드시 그럴 필요는 없다 (Lee, 2009).
- 본 연구는 사후 칼리브레이션 적용을 고려하였으나, 결론적으로 결과치에 반영하지 않았다. 왜냐하면 사후 칼리브레이션을 하기 위해 모집단 분포가 알려진 몇 개 변수를 사용한 결과, 많은 변수들이 오히려 칼리브레이션 전에 비해 후에 준거조사 추정치로부터 차이가 더 커지는 것으로 나타났다. 이에 대한 원인은 본 연구와 별도로 향후 연구를 통해 살펴보고자 한다.

### 3. 한계점과 기대효과

본 연구에 사용된 사회조사는 2년에 걸쳐 매년 서로 다른 분야에 대해 조사하고 있다. 이런 점에 비추어 볼 때, 본 연구는 2010년 자료에 국한하여 성향점수 조정을 적용하였다는 점에서 사회조사 전체로 확대하여 해석하는 데는 한계가 있다. 특히 성향조정에서 중요한 성향변수의 경우, 2010년과 2009년 사회조사 자료에 포함된 변수가 다르기 때문에 본 연구에서 선택된 성향변수를 일관성 있게 적용하기는 어렵다. 향후 두 개 년도 자료에서 선택한 성향변수가 동시에 조사될 수 있다면 이러한 문제는 어느 정도 극복될 수 있을 것으로 판단된다.

본 연구는 대규모 통계조사를 통해서 자원자표본 웹 조사 방법을 검토하였다는 점에서 큰 의미가 있을 것이다. 특히, 사회조사와 같이 항목이 많고, 연차별로 다르게 조사되는 조사에서 이러한 조사방법의 실현 가능성을 판단하는데 많은 도움이 될 것으로 기대한다. 뿐만 아니라, 이러한 연구는 비확률표본 웹 조사 자료를 얻기가 힘들기 때문에 일반 연구자들이 관심과는 다르게 쉽게 접근하기 어려운 분야기도 하다. 이런 점에서 본 연구는 학술적 연구에도 크게 기여할 수 있을 것으로 기대한다.



## 제2절 자원자 패널 웹 조사

### 1. 기본 특성

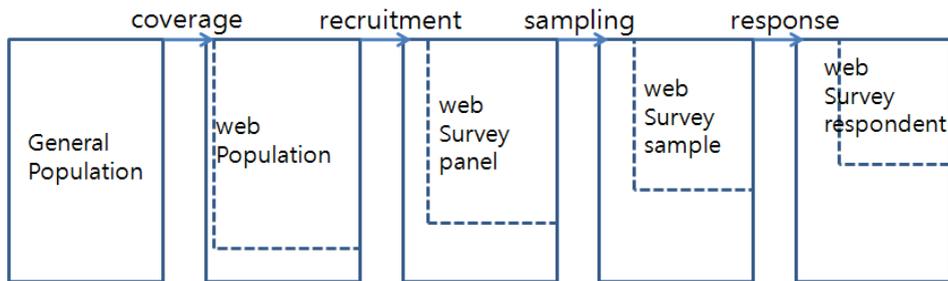
본 절에서는 자원자 패널을 이용한 웹 조사의 개념과 특성에 대해서 주로 Bethlehem (2010)과 Lee (2009, 2006)를 인용하고 이 외에도 관련된 선행 연구 문헌들을 정리하는 수준에서 설명하고자 하였다. Bethlehem (2010)은 두 가지 측면에서 웹 조사 결과에 대한 신뢰성을 얘기하고 있다. 모집단에 대한 과소포함과 조사 참여에 대한 자기선택 (self selection) 현상이 그것이다. 이 두 가지 현상은 편향된 추정치를 유도하기 쉽고, 따라서 수집된 자료로부터 잘못된 결론을 내릴 수 있다.

과소포함은 조사에 사용된 표본 선택 절차가 목표 모집단의 누군가는 선택되지 않는다는 것을 의미한다. 만약 자료가 인터넷에 의해 수집되었다면, 인터넷에 접근이 가능한 사람만이 질문지를 작성할 수 있다. 그러나 조사의 목표 모집단은 항상 인터넷 사용자보다 훨씬 더 넓은 범위를 포함한다. 이것은 인터넷 사용을 하지 않는 사람은 조사에서 배제된다는 것을 의미한다. Bethlehem (2010)은 평균적으로 볼 때 인터넷에 접근하는 사람은 접근하지 않는 사람과 그 특성이 다르다고 하였다. 결과적으로 웹 조사 결과는 전체

모집단이 아닌 단지 인터넷이 있는 부차 모집단에 대한 정보를 제공한다. 이렇게 얻어진 조사 결과는 전체 모집단 정보로 일반화하는 데는 사용될 수 없다. 웹 조사에 기초하여 얻어진 모집단 추정치는 편향되었다고 말할 수 있다.

자기선택은 조사에 참여할 것인지에 대한 여부를 완전히 개인 스스로가 결정한다는 것을 의미한다. 웹 조사의 경우, 조사 질문지는 단순히 웹에 올려지게 된다. 이 조사 질문지에 대한 응답 대상자들은 인터넷을 사용하는 사람들이고, 이들이 웹 사이트에 방문해서 조사에 참여할 것인가를 결정한다. 이때 조사 연구자는 인터넷에 접근하는 사람들이 조사에 대한 참여 여부를 판단하는 과정을 전혀 제어할 수 없다. 따라서 어떤 조사에서 조사 참여 여부 결정에 대해 자기선택적 방법을 적용한다는 것은 결국 확률표본추출의 원리를 따르지 않겠다는 것을 의미한다. 일반적으로 랜던표본추출일 경우에 확률이론이 적용될 수 있다. 이렇게 함으로써 조사추정치에 대해서 불편추정치를 얻을 수 있고, 추정치의 정확성을 측정할 수 있다. 확률표본추출 패러다임은 1940년대 이후 국가통계와 아카데미 통계에서 적용되어 왔고, 아주 드물게는 많은 상업적 시장조사에서도 사용되었다. 불행하게도, 많은 웹 조사는 확률표본추출 대신에 응답자의 자기선택에 의존한다. 이것은 조사결과의 품질에 심각한 영향을 미치게 된다. 이미 언급한 바와 같이 확률표본추출 이론이 적용될 수 없고 추정치는 편향되기 쉽다 (Bethlehem, 2010; Lee, 2006; 2009).

[그림 5-1]은 자원자 패널을 이용한 웹 조사에서 조사 참여자들에 대한 선택 절차를 나타낸다 (Lee, 2009). 자원자 패널 웹 조사는 가장 널리 사용되는 웹 조사 방법으로, [그림 5-1]을 이용하여 조사과정에서 포함과 자기선택 문제가 어떻게 발생하는지를 설명할 수 있다.



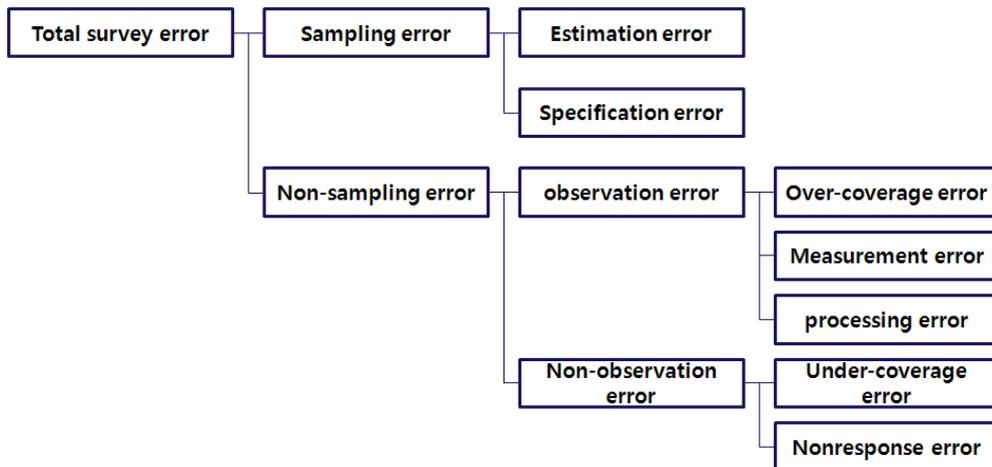
[그림 5-1] 자원자 패널 웹 조사 원칙 (Lee, 2006 Fig.1 발췌)

[그림 5-1]에서 보면, 일반 모집단의 모든 구성원이 웹 조사 모집단에 포함되지 않는다. 즉 웹 접근이 가능한 사람만 패널에 들어올 수 있다. 패널 모집은 다양한 유형의 광고 (배너, 팝업, 이메일 등)를 주로 이용한다. 조사 참여는 완전히 자발적으로 일어난다. 이때 자기선택에 의한 참여자는 자신에 대한 기본적인 정보를 제공하게 되지만 참여자의 정보나 신분을 확인하기는 어렵다. 패널 프레임이 구축되면 조사 연구자는 패널 프

레이아웃으로부터 패널의 기본정보를 이용하여 표본을 추출하고 이렇게 추출된 표본을 대상으로 조사를 실시한다 (Lee, 2006).

## 2. 총 조사 오차

전통적인 조사방법과 마찬가지로 자원자 패널 웹 조사는 전체적인 조사결과에 대해 많은 오차를 포함하고 있다. 이러한 오차에는 다양한 원인이 있을 수 있다. 오차들로부터 초래되는 궁극적인 결과는 조사추정치와 추정된 모집단 특성 간의 불일치이다. 이러한 불일치를 총 조사 오차 (total survey error)라 부른다. 총 조사 오차는 크게는 표본 오차와 비표본 오차로 나뉜다. 여기서 [그림 5-2]에 언급된 오차 중 주로 웹 조사와 관련된 오차에 대해서 설명한다. 나머지 개념들에 대해서는 (Lee, 2010; Bethlehem, 2010)을 참고할 수 있다.



[그림 5-2] 조사 오차 원리 (Bethlehem, 2010 Fig.1 발췌)

표본오차는 표본추출에 따른 오차로 표본의 크기가 모집단 크기와 같아진다면 표본 오차는 0이 된다. 표본오차는 추정오차 (estimation error)와 스펙오차 (specification error)로 구성된다. 추정오차는 랜덤추출 표본을 사용하기 때문에 발생한 결과이고, 추정오차는 확률이론을 통해 측정할 수 있고 표본설계를 통해 제어할 수 있다. 스펙오차는 표본 설계에서 정의된 선택확률과 실제 선택확률이 다를 때 발생한다. 스펙오차는 추출 프레임이 알려지지 않은 경우에 발생할 수 있다. 추출 프레임에서 선택된 웹 조사에서 스펙 오차는 발생할 수 있고, 자기선택인 경우 선택확률은 알 수조차 없다. 과대표함오차와

측정오차는 적당한 추출 프레임이 사용된 웹 조사에서도 발생할 수 있다. 자기선택적 웹 조사의 경우 과대표함 문제는 더 심각할 수 있다. 누가 질문지를 작성했는지조차 알 수 없기 때문이다. 또 면접원이 없고 조사에 대해 연구자 개입이 어렵기 때문에 조사설계 흐름상 심각한 측정오차가 발생할 수 있다 (Dillman 등, 2008; Couper, 2008). 과소포함은 자료수집 방법으로 인터넷이 사용되고 목표 모집단이 인터넷 모집단보다 더 범위가 넓다면 훨씬 심각하다. 또한 웹 조사는 조사관련자가 조사에 개입하기 어렵기 때문에 무응답에 대해 자유롭지 못하다.

요약하면 총 조사 오차 측면에서 자원자 패널 웹 조사는 모집단에 대한 과소·과대표함 오차와 스펙오차, 무응답 오차 등이 크게 부각되고 있다. 물론 확률표본 조사가 가지고 있는 모든 오차에 비확률표본에 따른 오차가 더해지기 때문에 확률표본에 의한 오차보다 훨씬 크게 나타날 수 있다. 또한 비확률표본에 따른 자기선택 편향은 확률 기반 조사와 비교 자체가 어렵다고 볼 수 있다.

### 3. 자원자 패널 웹 조사 처방

자원자 패널 웹 조사가 자료수집 방법으로서 기존의 전통적인 대면조사나 전화조사의 좋은 대안이긴 하지만, 조사의 총 오차와 관련해서 볼 때 과학적 자료의 질을 암시하지는 못한다 (Groves, 1989). 왜냐하면 [그림 5-1]에서처럼 한 단계에서 다른 단계로 이동할 때 선택 메카니즘이 잘 알려져 있지 않을 뿐더러 연구자가 각 단계를 제어할 수 없기 때문이다. 자원자 패널 웹 조사의 가장 큰 위험은 일반 모집단에 대해 불확실하고 불완전한 커버리지에 있다. 일반 모집단 중 조사 참여를 선택한 비율만이 웹에 접근한다는 사실은 일반 모집단에 대한 대표성을 의심스럽게 한다. 따라서 [그림 5-2]에서 살펴본 바와 같이 미포함, 비확률 추출, 무응답 등은 종합적으로 조사에 작용하여 웹 조사 추정치의 일반화를 어렵게 한다. 요약하면 이는 선택편향으로서 선택 메카니즘이 랜덤성을 보장하지 못하기 때문에 발생하는 문제이다.

웹 조사에서 발생하는 선택편향에 대한 문제를 해결하고 웹 조사 추정치의 질을 향상시키기 위해 다양한 가중치 기법들이 적용되어 왔다. 예를들어 사후층화 (post stratification) 나 성향점수 조정과 같은 방법이다. 사후층화는 전통적인 사후조사 조정에 해당하고, 이는 웹 조사의 편향을 보정하는데 한계가 있다. 대안으로 성향점수 조정이 등장하였고 웹 조사 자료의 사용을 쉽게 하기 위해서 필요한 방법이지만 여전히 많은 논의 중에 있다 (Lee 와 Vaillant, 2009; Loosveldt 와 Sonck, 2008; Schonlau 등, 2009; 등). 이는 사후적으로 편향을 조정한 후 웹 조사의 대표성이 향상되었는지에 대한 판단이 확실하지 않기 때문이다. 특히 자원자 패널 웹 조사에 대해 성향점수 조정을 통해 선택편향을 극복할 수 있

다는 많은 연구들 (Schonlau 등, 2004; Danielsson 2002; Taylor 2000)이 있는 반면, 좀 더 깊이 있는 연구가 필요하다는 연구들도 있다 (Lee, 2006; 2009; Steinmetz 와 Tijdens, 2010, Bethlehem, 2010 등). 이에 대해 특히 Lee (2006; 2009)는 웹 조사 편향 조정을 위한 성향점수 조정 방법이 등장하게 된 배경을 중심으로 1) 웹 조사와 성향점수 조정이 태어나게 된 상황 간의 유사성을 면밀하게 살필 필요가 있다고 하였고, 2) 웹 조사를 위한 성향점수 조정 뒤에 숨겨진 수학적 근거가 명백하게 제시될 필요가 있다는 의견을 제시하였다. 그리고 3) 성향점수 조정으로 인해 더 커지는 분산 성분에 대해 언급하고, 성향가중치 적용으로 인해 추정치의 분산이 더 커지게 됨으로써 추정치의 정도가 떨어질 수 있다고 하였다.

따라서 자원자 패널 웹 조사를 적용하는 데 있어서, 추정치에 대한 편향은 물론이고 분산도 고려되어야 할 것이다. 특히 국가통계와 같이 정확성이 요구되는 조사일수록 추정치의 신뢰성 유지에 대한 상세한 설명이 필요하다. 따라서 비확률표본 웹 조사 추정치에 대한 편향을 줄이는 대신 분산이 증가하게 된다면, 이는 실질적 적용에 있어서 신중하게 검토할 필요가 있을 것으로 보인다.

## 제3절 성향조정 방법

### 1. 기본 배경

확률 표본조사 이론이 과학적인 조사의 기본으로 자리매김하게 된 것은 확률표본을 사용할 경우 불편 추정치 (unbiased estimate)를 기대할 수 있다는 점이 크다고 하겠다. 이는 표본이 특정 변수의 영향에서 벗어나 랜덤하게 추출되기 때문이다. 확률표본 이론은 표본 추출틀 (sample frame)은 모집단에 대해 100% 커버리지 (coverage)를 충족하고, 100%의 응답을 얻는다는 전제에 기초하고 있다. 그러나 이러한 가정들은 표본 추출보다는 다른 요인들로부터 발생하는 편향을 포함하기 때문에 실제 표본조사에서는 그다지 현실성이 없다. 최근 이러한 가정들이 점점 효력을 잃어가면서 조사 연구자들은 발생 가능한 편향을 줄이는 방법을 찾기 위해 활발한 움직임을 보이고 있다. 가중치 조정 (weighting adjustment)은 그러한 편향에 대한 사후 처방책으로서 가장 널리 사용되는 방법 중 하나이다. 가중치 작업은 조사추정치의 정도를 높이고 편향을 줄일 수 있다 (Groves, 2006; Little, 1993; Little and Vartivarian, 2005). 이때 가중치 조정은 조정에 사용된 변수에 의존한다. 비록 가중치 조정이 조사 결과에 대해 가중치에 따른 변동을 더하게 됨으로써 조사 추정치의 분산을 증가시키지만 (Kalton and Flores-Cervantes, 2003), 편향 축소는 조사에 있어서 최우선 사항이기도 하다. 따라서 어떤 표본조사가 정확한 추론을 위한 가중치 변



수를 포함하기만 한다면 최신 대부분의 통계 소프트웨어 패키지들은 조사추정치를 구하는데 있어서 가중치를 쉽게 사용할 수 있는 툴들을 포함하고 있다.

최근, 성향점수 조정은 조사통계에서 가중치 조정을 위한 보조도구로 활용되고 있다. Lee 와 Valliant (2007)은 이 분야에서 최신 연구동향을 제시한 바 있다. 성향점수 조정은 관측자료들을 이용한 casual inference 에 관한 문헌에서 시작된다 (Rosenbaum 와 Rubin, 1983, 1984; D'Agostino, 1998). Causal inferences 는 실험 (experiment)이 필요하고, 그 실험은 확률 표본을 사용해야 한다는 것이 이상적인 전제 조건이다. 그러나, 실제 실험연구에 있어서 확률실험은 실현하기 어렵고, 비현실적이며 때로는 비윤리적일 수 있다. 그래서 연구자들은 종종 확률실험 자료 대신에 관찰 자료로 제한하여 사용하기도 한다. 대조군과 실험군에 포함된 표본 특성은 실험 요인뿐만 아니라 외적 다른 특성들에 의해서도 다를 수 있다. 두 그룹 간 결과를 비교할 때, 결과들은 편향을 수반하는 다른 특성들과 교락될 수 있다. 우리는 그룹 간 교락 요인들의 분포를 균형 있게 맞출 필요가 있다. 성향점수 조정은 공변량으로 적절한 교락 요인들에 대해 대조군에 대한 처리군의 성향을 모형화하는 것에서 출발한다. 처리군과 대조군의 결과가 조건부 성향점수에 비교될 때, 우리는 비교 결과가 선택편향에 덜 영향을 받을 것이라고 기대할 수 있다.

표본조사에서 무응답과 비포함 편향은, 그것들이 랜덤화 가정에 위반된 결과일 때, 관측연구에서 선택편향과 유사하게 보인다. 이러한 이유에서 잘 구축된 성향점수 조정이 조사 연구 문헌에 등장하게 되었고, 잠재적 무응답 (Iannacchione 등, 1991; Lepkowski 등, 1989; Smith 등, 2001), 늦은 응답 (Czajka 등, 1992), 및 비포함 편향 (Brick, Waksberg 와 Keeter, 1996; Duncan 과 Stasny, 2001; Hoaglin 과 Battaglia, 1996; Battaglia, 등, 1995; Garren 과 Chang, 2002)을 조정하기 위해 적용되었다. 예를 들어, 무응답 성향은 무응답자들의 연령, 성별, 도시성, 거주 여부와 같이 응답성향이 응답 결과와 관련이 있을 것으로 잘 알려진 특성들을 기초로 모형화한다. 그런 다음 이 적합 모형에 의해 무응답 성향점수가 각 표본개체에 대해 추정된다. 이때 적합 모형에는 응답자와 무응답자 모두가 포함된다. 추정된 점수가 이용가능하다면, 가중치 조정은 더 높은 무응답 성향점수를 갖는 개체가 더 높은 가중치를 받는 방식으로 조정된다. 이때 무응답자들은 모형에 포함된 특성변수들에 대해서 그 성향이 유사한 개체들이다. 논리자체는 사후층화와 같은 전통적인 조정 방법들과 다르지 않다. 우리가 조정에 여러 개의 변수를 포함시키고자 할 때, 성향점수 조정 방법은 사후층화 방법 대신에 사용해볼 만한 방법이다. 왜냐하면, 사후층화 방법은 너무 많은 층들이 붕괴 (collapsed)될 수 있기 때문이다. 성향점수 조정은 모든 교락 요인 (confounding factor)들이 하나의 크기(값) 형태로 요약되기 때문에 붕괴와 같은 문제는 신경쓰지 않아도 된다. 조정은 오직 한 개 점수만을 기초로 이루어진다.

성향점수 조정이 비확률 표본조사에 적용되었다. 비확률 표본조사는 주로 관측 연

구에서 선택편향과 같은 비확률 이슈와 관련된 자원자에 의한 웹 조사에 해당된다 (Terhanian 과 Bremer, 2000; Lee, 2006; Schonlau 등, 2009). 비확률표본과 확률표본에서의 무응답과 비포함이 갖는 한 가지 차이는 후자는 확률표본추출 내에서 랜덤화 이론이라는 명백한 가정에 위배되는 것이고, 반면에 전자는 랜덤화 원리를 전혀 포함하지 않는 것이다. 이런 견해에서 보면, 비확률표본에서 선택편향에 의한 영향은 심지어 확률표본에서 무응답과 비포함이 주는 영향보다 훨씬 크다. 게다가 무응답과 비포함의 크기는 자원자 웹 조사가 확률표본조사보다 훨씬 크다는 것이 알려져 있다 (Couper, 등, 2007).

이 분야 연구 문헌들은 꾸준히 증가하고 있으며 이들은 성향점수 조정은 편향을 줄이는데 효과적이라는 것을 보였다 (Lee, 2006; Harris Interactive, 2008; Schillewaert 와 Meulemeester, 2005; Schonlau 등, 2009; Taylor 등, 2001). 그러나 편향 축소가 모든 변수에 대해서 보편적인 현상은 아니다. 어떤 경우에는 편향이 조정 후에 더 커지는 경우도 있다. 이것은 성향점수 조정이 단지 임기응변적 처방이라는 점에서 보면 그리 놀라운 것은 아니다. 즉, 이 처방책은 그런 편향에 오염되지 않도록 이론적으로 보장된 확률표본과 달리, 선택편향을 줄이는데 도움이 되도록 의도된 것이다. 비록 자주 논의되는 것은 아니지만, 이 성향점수 조정의 큰 단점은 추정치의 변동이 증가하기 때문에 통계적 효율이 떨어진다는 것이다 (Lee 와 Valliant, 2009). 그런 의미에서, 성향점수 조정은 편향 축소 대신 분산 증가에 대한 희생이 따르게 된다. 편향과 분산을 모두 포함하는 조사의 총오차 관점에서 보면 성향점수 조정에 의한 이점은 편향 측면만 고려할 때에 비하면 그렇게 크지는 않다.

## 2. 성향점수 조정 방법

### 가. 준거조사

자원자 패널 웹 조사에서 성향점수 조정은 준거조사 (reference survey) 자료가 존재한다는 것을 가정한다. 준거조사는 웹 조사와 평행하게 실시되어야 한다. 자원자표본 웹 조사에 대한 성향점수 조정은 위에서 언급한 무응답 성향조정과 매우 유사한 메커니즘을 따른다. 유일한 차이는 이 경우는 준거조사라는 또 다른 조사가 필요하다는 것이다. 준거조사는 대조군 역할을 하고, 웹 조사는 처리군 역할을 한다. 이 아이디어는 벤치마크로서 준거조사를 사용하는 것이고, 준거조사는 가능하면 준거표본의 성향점수 분포와 비슷하도록 웹 조사 표본의 성향점수 분포를 만든다.

이런 이유에서, 준거조사는 표본 선택의 랜덤화, 무응답 및 비포함에 관해서 최상의 성격을 가져야 한다. 준거조사의 품질은 조정의 효과를 결정하는 요인 중 하나이다. 준



거조사는 웹 조사에 대한 자발적 참여 성향에 영향을 줄 것으로 기대되는 변수들을 반드시 포함해야 한다. 성향 모형이 많은 공변량을 포함할수록, 일관성 있는 공변량 집합이 나올 때까지 준거조사에서 많은 공변량에 관한 자료 수집을 고려하는 것이 바람직하다. 물론, 이 변수들이 성향을 잘 예측하는 만큼 조정 효과에 직접적으로 영향을 미친다.

준거조사는 표본 크기가 클 필요가 없다. 그렇지만 표본 크기가 작으면 이는 추정치의 변동을 크게 할 수 있고 성향 모형에 대한 잔차(residual)를 크게 할 수 있다. 준거조사는 목표 모집단이 같고 특별한 효과가 기대되지 않는 한 하나 이상의 웹 조사에 대해 사용될 수도 있다. 그러나 문맥효과(context effect)와 같은 또 다른 측정 이슈에 대해서 간과해서는 안 된다. 통계청 사회조사의 경우, 전통적인 방법으로 수집되는 대면조사가 준거조사에 해당된다. 준거조사의 표본 크기는 약 17,000가구로 확률층화계통추출 방법에 의해 추출되었다. 자세한 설명은 4절에서 하기로 한다.

#### 나. 자원자표본 웹 조사에서 성향점수 조정

성향점수는 누군가가 조사에 참여할 것인지 아닌지를 나타내는 변수를 모형화해서 얻게 된다. 보통 종속변수가 이항변수이고, 태도변수(attitudinal variable)가 종속변수인 로지스틱 회귀 모형(logistic regression model)이 사용된다. 이때 태도변수는 조사 참여자가 왜 조사에 참여하는지를 설명할 수 있을 것이라고 가정한다. 로지스틱 회귀 모형은 설명변수가 주어지면 참여확률 즉, 성향점수를 추정하게 된다.

##### 성향점수 추정

두 개의 표본이 있다고 가정하자. 하나는 자원자 패널 웹 조사 표본  $s^w$  이고, 이 조사의 표본 크기는  $n^w$ , 각 개체에 대한 기본가중치(base weight)를  $d_j^w$  라 하자.  $j=1, \dots, n^w$ . 다른 하나는 준거조사 표본  $s^r$  이고, 표본 크기는  $n^r$ , 기본가중치는  $d_k^r$ , ( $k=1, \dots, n^r$ )라 하자. 이때 자원자 패널 웹 조사 표본이 확률표본이 아니기 때문에 기본가중치  $d_j^w$  는 선택확률의 역수가 아닐 수 있다. 대체로 기본가중치  $d_j$  는 1인 경우가 많다. 그렇지만 사회조사에서와 같이 웹 조사 표본이 웹 조사에 자원한 사람들로 구성된 패널로부터 확률추출된 경우라면 패널에서 표본  $j$ 가 선택된 확률의 역수가 사용될 수 있다 즉, 웹 조사에 대한 기본가중치는  $\frac{N^w}{n^w}$  형태를 취한다. 여기서  $N^w$ 는 웹 패널 수를 나타낸다.

성향점수 추정을 위한 첫 단계는 준거 표본과 웹 조사 표본을 합해서 하나의 자료 셋을 만드는 것이다. 통합된 자료는  $s = s^w \cup s^r$  이고 통합된 자료의 크기는  $N = n^w + n^r$  과 같이 표현한다. 이렇게 구성된 자료  $s$ 로부터 성향점수를 추정할 수 있다.  $i$  ( $i=1, \dots, n$ )

번째 개체의 성향점수는 보조변수 (설명변수)가 주어졌을 때, 어떤 개체가 준거조사보다 자원자 패널 웹 조사에 참여할 가능성을 나타내고, 다음과 같이 표현될 수 있다.

$$ps(\mathbf{x}) = \Pr(i \in s^w | \mathbf{x}_i, i = 1, \dots, n)$$

성향점수 추정에는 웹 조사와 준거조사 모두에서 수집된 공변량을 포함한 회귀모형을 사용한다. 모형은 다음과 같다.

$$\log\left(\frac{ps(x_i)}{1 - ps(x_i)}\right) = \alpha + \beta' \mathbf{x}_i + \epsilon_i, \quad i = 1, \dots, n.$$

이렇게 성향점수가 추정되면, 다음 단계는 이들 성향점수를 이용하여 성향조정에서 사용될 가중치를 생성하는 것이다. 이는 그룹들 간의 균형을 맞추기 위해 사용된다. 방법으로는 추정된 성향 점수에 대해 매칭 (matching), 층화 (stratification), 공분산 조정 (covariance adjustment), 가중치 방법을 이용할 수 있다 (Steinmetz 와 Tijdens, 2010). 이 중에서는 우리는 층화방법과 가중치 방법을 이용하여 웹 조사의 성향을 조정하고자 한다. 따라서 다음에서는 여러 가지 방법 중 층화방법과 가중치 방법만 소개한다.

### 성향점수 가중치 작성

#### ① 층화방법

이 방법은 성향조정 방법으로 널리 사용되고 있는 방법 중 하나이다. 성향점수가 추정되면, 이 점수들은 모집단을 층화하는데 사용된다. 각 층은 거의 같은 성향점수를 갖는 개체들을 구성한다. 실제로 같은 층 내에 있는 모든 개체들의 응답 성향이 같다면, 웹 모집단에 있는 개체들만이 추정목적으로 사용될 경우 추정치에 대한 편향은 없을 것이다.

추정된 성향점수를 기초로 준거표본과 웹 표본이 모두 포함된 자료  $s$ 를  $\hat{ps}(\mathbf{x}_i)$ 를 크기 순으로 정렬한 후, 표본  $s$ 를  $c$ 개의 부그룹으로 나눈다. 이때 각 부그룹은 동일한 개체수를 포함하게 된다. Cochran (1968)은 편의상 사분위수 분기점 (quantile point)을 이용하여  $c = 5$ 를 제안하였다. 실제로 5개 보다 많은 그룹을 사용할 경우, 부그룹 구간이 더 좁게 형성되기 때문에 훨씬 동질적인 그룹이 될 수 있다. 본 연구는 5개의 부그룹을 사용하였다.

우선  $c$ 개의 부그룹으로 나누고,  $c$  그룹 내의 표본을  $s_c$ , 표본크기를  $n_c = n_c^w + n_c^r$ 라 하면, 표본  $s$ 에 포함된 개체 수는 각 그룹에 포함된 개체 수를 더한 것과 같다.

$$n = \sum_{c=1}^C (n_c^w + n_c^r) = \sum_{c=1}^C n_c .$$



다음으로 다음과 같은 조정계수 (adjustment factor)를 계산한다. 이때 조정계수는  $c$  번째 그룹의 웹 조사 자료 ( $s_c^w$ )에 포함된 모든 개체에 적용될 것이다.

$$f_c = \frac{\sum_{k \in s_c^r} d_k^r / \sum_{k \in s^r} d_k^r}{\sum_{j \in s_c^w} d_j^w / \sum_{j \in s^w} d_j^w} .$$

여기서  $s_c^r, s_c^w$ 는  $c$ 번째 그룹 내의 준거표본과 웹 표본을 나타낸다. 만약 위의 식에서 기본가중치  $d_k$ 가 선택확률의 역수라면 (즉, 확률표본), 조정계수는 다음과 같아진다.

$$f_c = \frac{\widehat{N}_c^r / \widehat{N}^r}{\widehat{N}_c^w / \widehat{N}^w} .$$

이때,  $\widehat{N}_c^l = \sum_{k \in (s_c^l)} d_k^l$  ( $l = w, r$ )는 준거 또는 웹 조사를 기초로 추정된 모집단 개수를 나타내고,  $\widehat{N}^l = \sum_c \widehat{N}_c^l$ 이다. 웹 표본의  $c$  그룹 내  $j$ 번째 개체에 대한 조정된 가중치는 웹 조사의 기본가중치에 조정계수를 곱하는 형태이다.

$$d_j^{w,psa} = f_c d_j^w = \frac{\widehat{N}_c^r / \widehat{N}^r}{\widehat{N}_c^w / \widehat{N}^w} \cdot d_j^w .$$

만약 기본가중치  $d_j$ 가 1이거나 이용할 수 없는 경우에는, 대안으로 다음과 같은 조정계수를 사용할 수 있다.

$$f_c = \frac{n_c^r / n^r}{n_c^w / n^w} .$$

그러나 이런 조정계수는 가중치가 적절하지 않기 때문에, 준거조사와 웹 조사 모두에 대한 모집단 크기가 알려지지 않았다면, 모집단 총수 추정에는 사용할 수 없다. 이렇게 계산된 가중치는 성향점수 개념에서 웹 조사 표본 분포와 준거조사 표본 분포를 같도록 한다. 예를 들면, 위에서 계산한 성향점수에 의해 조정된 가중치를 사용하여 웹 표본으로부터  $c$  그룹 내의 모집단 총수를 추정할 경우, 총수 추정은 다음과 같다.

$$\widehat{N}_c^{w,psa} = \sum_{j \in s_c^w} d_j^{w,psa} = \widehat{N}^w \frac{\widehat{N}_c^r}{\widehat{N}^r} .$$

관심변수  $y$ 에 대한 평균 추정은 다음 식을 이용할 수 있다.

$$\hat{y}_{est}^{w,psa} = \frac{\sum d_j^{w,psa} y_j}{\sum d_j^{w,psa}} .$$

## ② 가중치 방법

성향점수 조정을 다른 방법으로 이용할 수 있다. 이 방법은 가중치로서 성향점수의 역수를 사용하는 것이다 (Rosenbaum, 1984; Schonlau 등, 2007). 성향점수는 웹과 준거조사 응답자 모두에 관련하기 때문에 두 표본에 대한 성향점수 가중치는 직접적으로 다음과 같이 사용할 수 있다. 여기서  $ps_i$ 는 표본  $i$ 가 웹 조사에 참여할 확률로서 성향점수를 의미한다.

$$\begin{cases} (1-ps_i)/ps_i & , \text{ if 표본 } i \text{가 웹 표본에 포함} \\ (1-ps_i)/(1-ps_i) & , \text{ if 표본 } i \text{가 준거표본에 포함} \end{cases}$$

지금까지 언급한 성향점수 가중치 조정에 대해 몇 가지를 확인하고 추가하자면 다음과 같다. 웹 조사에 대한 준거조사가 필요하고, 이때 준거조사의 표본크기는 웹 표본 크기보다 더 작게 된다. 만약 웹 표본보다 준거 표본이 더 많이 필요하다면, 준거 표본은 모든 웹 조사에 대해 수행되어야 하고 웹 조사의 장점으로 부각되어 있는 비용에 대한 효과는 상실할 수 있다. 또한 일반적인 표본조사 자료의 분석에서 선택편향에 대한 성향점수 조정은 필요하지 않다. 왜냐하면 대부분 과학적 조사는 확률표본이기 때문에 불편추정량을 기대할 수 있다. 성향점수 조정은 조사통계 분야에서 그렇게 독특한 방법은 아니다. 특히 사후조사 조정에 사용된 성향점수 조정은 커버리지 문제, 늦은 응답, 무응답으로부터 발생하는 조사추정치의 편향을 줄이기 위한 조정 가중치를 생성하는데 사용된 것이다.

## 3. Rim 가중법

림 가중은 모집단의 인구사회학적 특성의 주변 분포를 알고 있는 경우 적용할 수 있는 가중치 보정 방법으로 각 특성 변수의 주변 분포를 이용하여 반복적으로 셀 내의 모집단 수를 구하는 방법이다. 이 방법은 모집단의 주변 분포만 알면 적용이 가능하므로 고려할 특성 변수가 많은 경우에도 활용이 가능하고, 빈 셀 문제에도 보다 적극적으로 대응할 수 있다.



림 가중치 방법은 Deming과 Stephan (1940)이 처음 제안한 이래 현실적인 유용성 때문에 실제 문제에 활용되기 시작하였다. 이 방법은 가중값을 반복적으로 수정해 가므로 수정된 가중치의 수렴 여부가 중요한데, 모든 칸의 표본수가 0이 아니라면 수렴한다는 결과가 있다 (Ireland와 Kullback, 1968). 그러나 만일 일부 칸에서 표본수가 관측되지 않으면 수렴하지 않을 수도 있다 (Thompson, 1981). 이 방법으로 구한 최종 가중치는 설계 가중치를 보조정보로 이용하여 수정한 형태이기 때문에 불편추정량도 일치 추정량 (consistent estimator)도 아니다. 그렇지만, 이 방법은 주요 보조변수에 관하여 표본 추정 결과를 모집단 비율과 일치시킴으로 추정의 관점에서 신뢰성을 확보하는 장점이 있다. 또한 단순임의추출에 의한 표본뿐 아니라 일반적인 추출법에 의한 표본에도 적용이 가능하다 (허명희 등, 2005).

림 가중치 작성을 위한 알고리즘은 다양하다. 여기서는 Little과 Wu (1991)의 방법에 대해 예를 들어 설명함으로써 독자들의 이해를 돕고자 한다.  $i \times j$  교차표에서 각 셀에 포함된 빈도를  $n_{ij}$ 라 하고,  $n$ 을 전체 표본 수라고 하자. 이에 대한 림 가중치 작성절차는 다음과 같다.

step1. 먼저,  $n_{ij}/n$  (=각 표본 셀의 빈도/표본 크기)과 같도록 가중치를 초기화 한다.

step2.  $\hat{w}_{ij}^{(1)} = w_i \times \frac{\hat{w}_{ij}^{(0)}}{\sum_i \hat{w}_i^{(0)}}$ . 기본가중치에 대해 행 가중치를 계산한다.

step3.  $\hat{w}_{ij}^{(2)} = w_j \times \frac{\hat{w}_{ij}^{(1)}}{\sum_j \hat{w}_j^{(1)}}$ . step1에서 계산된 결과에 대해 열 가중치를 계산한다.

step4. 모든  $i, j$ 에 대해서  $\sum_i \hat{w}_i = w_i$ ,  $\sum_j \hat{w}_j = w_j$ 이 될 때까지 step2와 step3을 반복한다.

예를들어 성별(2범주)×연령(2범주) 교차표가 있다고 하자 (<표 5-1>).

<표 5-1> 표본조사 결과

$i \backslash j$	10대	20대	30대	40대	50대 이상	가중합
여자	300	1,200	60	30	30	1,620
남자	150	1,080	90	30	30	1,380
가중합	450	2,280	150	60	60	3,000

주변합에 대한 모집단 총수는 여자는 1,510명, 남자는 1,490명, 10대 600명, 20대 2,120명, 30대 150명, 40대 100명, 50대 이상 30명 이라고 가정하자. 그러면,

Step1) 먼저 행을 조정한다. 각 행에  $\frac{\text{실제 모집단 행 총수}}{\text{추정된 행 모집단}}$  를 각각 곱한다. 즉, 여자 해당 행에는 1,510/1,620, 남자 해당 행에는 1,490/1,380을 곱한다.

$i \backslash j$	10대	20대	30대	40대	50대 이상	가중합
여자	279.63	1,118.52	55.93	27.96	27.96	1,510
남자	161.96	1,166.09	97.17	32.39	32.39	1,490
가중합	441.59	2,284.61	153.10	60.35	60.35	3,000

와 같이 행 주변합은 모집단 실제값과 같아진다. 그렇지만 열 주변합은 같아지지 않는다. 그래서 이제는 열에 대해서 동일한 과정을 거친다.

step2) 바로 위에서 조정된 교차표의 각 열에 대해서  $\frac{\text{실제 모집단 열 총수}}{\text{새롭게 추정된 열 모집단}}$  을 곱한다. 즉, 10대 열에는 600/441.59, 20대 열에는 2,120/2,284.61 등과 같이 모든 열에 대해서 동일한 방법으로 조정한다. 이렇게 하면 열 주변 합이 모집단 참값과 같아지고, 행 주변합은 다시 흐트러지게 된다.

$i \backslash j$	10대	20대	30대	40대	50대 이상	가중합
여자	379.94	1,037.93	54.79	46.33	13.90	1,532.90
남자	220.96	1,082.07	95.21	53.67	16.10	1,467.10
가중합	600	2,120	150	100	30	3,000

Step3) step2)의 결과를 다시 행에 대해 주변분포를 모집단과 같게 조정한다. 즉, 여자 행에는 1,510/1,532.90, 남자 행에는 1,490/1,467.10을 곱한다. 이렇게 함으로써 다음의 표와 같이 조정 전의 각 셀은 새로운 셀 값으로 채워지게 바뀌게 된다. 조정 전의 20대 남자는 1,080명으로, 조정 후 1,098.53명에 비해 약간 과소 추정된 것을 알 수 있다. 그래서 20대 남자 셀에 해당된 각 개인은  $1098.53/1080 \approx 1.017$ 의 가중치를 부여받는다.

$i \backslash j$	10대	20대	30대	40대	50대 이상	가중합
여자	375.59	1,021.47	53.72	45.56	13.67	1,510
남자	224.41	1,098.53	96.28	54.44	16.33	1,490
가중합	600	2,120	150	100	30	3,000

지금까지 설명한 림 가중치 방법은 수렴이 느릴 수 있고, 수렴 자체가 불가능한 경우가 있다는 단점이 있다.



## 제4절 실제 자료분석

본 절에서는 2010 사회조사 결과에 대해 웹 조사에 참여한 응답자들의 성향을 조정하고자 한다. 자료분석의 목적은 2010년 사회조사 전체 항목을 대상으로 성향점수 조정을 적용하고, 이를 통해 웹 조사 추정치가 준거조사 추정치에 어느 정도 근접해지는지 평가하는 데 있다.

### 1. 자료

사회조사는 국민의 삶의 질과 관련된 사회적 관심사와 주관적 의식에 관한 사항을 조사하여 국민의 삶의 수준과 사회적 변동을 파악하고, 이를 사회개발 정책 수립의 기초 자료로 제공하는 것을 목적으로 한다. 이 조사는 2개년에 걸쳐 10개 영역 조사를 목적으로 매년 5개 영역으로 나누어 조사하고 있다. 2009년에는 문화와 여가, 복지, 소득 및 소비, 노동, 사회참여 등 5개 영역을 조사하였고, 2010년에는 보건, 교육, 안전, 가족, 환경 등 5개 영역을 조사하였다. 사회조사는 전통적인 대면조사 방식으로, 조사원이 17,112 가구를 직접 방문하여 응답을 얻는다. 2010년 사회조사는 5월 17일부터 20일에 걸쳐 실시되었다.

그러나 조사환경이 날로 어려워지고, 사회 환경이 복잡해지면서 점점 늘어나는 1인 가구, 맞벌이 가구 등을 대면하기 어렵게 됨에 따라 응답률이 점점 줄어들고 있는 실정이다. 이러한 현상은 우리나라뿐만 아니라 국제적으로도 심각한 상황에 있다고 볼 수 있다. 이에 대한 대안으로 응답 가구를 직접 방문하지 않고 쉽게 응답을 얻을 수 있는 웹 조사 방식이 각광받고 있다. 게다가 응답대상자조차도 스스로 조사에 참여할 의사가 있는 가구를 선정하여 사용하는 자원자 패널 웹 조사에 대한 연구의 관심이 높다. 이런 연구를 위해 준거조사가 필요하고, 2010년 대면방식의 사회조사는 본 연구의 준거조사가 된다. 즉, 준거조사는 자원자 패널 웹 조사가 웹 참여 성향조정을 통해 얼마나 준거조사에 근접하는가를 평가하는데 사용된다.

한편, 사회조사는 2009년과 2010년 두 조사에서 자원자 패널 웹 조사를 병행하였다. 웹 조사는 연구를 목적으로 실시된 것으로 실제 사회조사 결과를 발표하는데 사용하지는 않았다. 웹 조사는 사회조사 실시기간에 동시에 진행되었다. 우선 웹 조사를 위해 이 조사에 참여할 의사가 있는 자원자를 다양한 홍보를 통해 응모하였다. 이렇게 구축된 자원자 패널 8,135가구는 비확률 기반 패널이라 할 수 있다. 이렇게 구축된 패널로부터 지역, 성, 연령별 확률적 층화에 의한 가중 리스트 계통추출법(systematic sampling from the weighted list)에 의해 1,500가구를 추출하였다. 표본크기 1,500에 대한 설정은

예산과 표본의 분산 (sample variance)을 고려한 것이다. 이때 추출 변수는 연령 (5개 범주), 성 (2개 범주), 동·읍면 (2개 범주), 시도 (16개 범주)를 사용하였다. 앞에서 나열한 추출변수 순서대로 정렬하여 계통추출 시 내재적 층화 효과를 얻고자 하였다.

## 2. 자료전처리 및 성향변수 선택

### 무응답 대체

자원자 패널 웹 조사결과의 조정을 위해 본 연구에서는 성향점수를 이용한 조정을 사용하였다. 성향점수 조정은 성향변수 선택이 매우 중요한 위치를 차지한다. 이때 성향변수는 무응답을 포함하지 않아야 한다. 무응답이 포함된 변수에 대해서는 성향점수를 계산할 수 없다. 따라서 무응답이 포함된 변수는 항목 대체를 실시한 후 사용하는 것이 바람직하다. 본 연구는 핫덱 (hot-deck) 방법을 사용하여 항목 무응답을 대체하였다. 무응답은 성별 (2개 범주) \* 연령별 (5개 범주) \* 시도별 (16개 범주) 변수를 고려하여 160개 층을 형성하고, 이 무응답층에 대해서 핫덱 무응답 대체를 실시하였다.

### 성향변수 선택

사회조사 주요 조사항목은 무응답 대체된 항목을 고려하여 총 102개에 해당된다. 여기서 102개 항목은 세부항목을 제외한 가장 상위의 항목만 해당하고, 15세 미만 가구원에 관한 세부 질문 항목은 제외되었다. 이 102개 변수는 성향변수 후보에 해당되고, 이 변수들을 대상으로 조사방법 간에 유의한 차이가 있는 변수들을 성향 조정변수로 고려하였다. 우선 조사방법(준거조사 vs. 웹 조사)과 변수 간 연관성 분석( $\chi^2$ ,  $\gamma$  통계량 사용)을 통해 유의한 차이가 있는 변수를 1차적으로 선정하였다. 여기에 조건부 변수라든가 변수의 성격상 매우 적은 수의 응답자의 의견을 반영하는 경우는 제외하였다. 최종 선정된 변수들에 대해서 로지스틱 회귀분석을 실시하고, 조사방법에 영향을 미치는 변수를 선정하였다. 그 결과는 <표 5-2>와 같다. 유의수준 0.05를 근거로 두 조사 방법 간에 차이가 있는 변수는 19개인 것으로 나타났다.

참고로 로지스틱 회귀분석을 수행할 때 분석에 사용된 모든 변수는 범주형 변수에 해당된다. 따라서 범주의 개수가 너무 많은 경우, 변수의 개수는 유의성 검정의 파워를 떨어뜨리는 경향이 있다. 즉, 변수의 개수 때문에 변수의 효과가 정확히 나타나지 않은 경우가 있을 수 있다. 따라서 범주형 변수들의 범주 개수를 줄여서 회귀분석을 수행하였다. 5점 리커트 척도 문항에 대해서는 3개 범주로 축소하였다. 예를들어 ‘매우 만족’부터 ‘매우 불만족’의 5점 척도의 경우는 ‘매우 만족’과 ‘만족’을 하나의 범주로 간주하여 ‘만족’, ‘매우 불만족’과 ‘불만족’을 합하여 ‘불만족’으로 만들었다. 즉, ‘만족’, ‘보통’, ‘불만



족' 등 3개 범주를 분석에 사용하였다. 리커트 척도가 아닌 변수에 대해서는 내용상 유사한 항목을 묶어서 하나의 범주로 사용하기도 하였다. 회귀분석에 사용된 변수의 범주 정보에 대한 변경전·후는 <부표 5-3>을 참고할 수 있다

이 질문지 항목 중에서 '학교급식 만족도'와 '학교주변식품 만족도'에 관한 두 질문 내용의 연관성이 매우 높다 ( $p=0.001$ ). 그래서 두 변수 중 학교급식 안전도 관한 "학교급식 만족도" 변수는 변수의 개수를 가급적 최소화한다는 차원에서 제외하기로 하였다. "인터넷조사 참여의향" 변수는 두 조사방법 간에 유의한 차이가 있었으나 성향조정 가중치의 극단값에 영향을 줄 것으로 판단되어 최종 성향변수에서 제외하였다. 뿐만 아니라 거처종류 변수는 통계적으로 유의한 차이가 있지만 이는 조사 참여자들의 성향보다는 일반적인 모집단 차이에 의한 원인이 크다고 판단되어 조정변수에서 제외하였고, 성별은 통계적으로 유의한 차이는 없지만, 이는 표본추출 시 모집단 분포가 사전에 고려된 주요 인구학적 특성 변수라는 점에서 성향 조정변수에 포함하기로 하였다. 따라서 최종적으로 선정된 변수는 17개에 해당된다. 추가로 시도 구분 변수를 성향변수로 포함해서 분석을 시도하였다. 시도에 따라 웹 조사 참여 정도가 다르기 때문이다. 그런데 많은 평가변수의 성향조정결과가 시도변수가 포함되기 전에 비해 웹 조사 추정치가 준거조사에서 더 차이가 생기는 것을 확인할 수 있었다. 따라서 이 변수는 성향 조정변수에 사용하지 않았다.

최종적으로 선택한 17개 성향변수는 다음과 같다.

<표 5-2> 인구학적 및 비인구학적 특성변수

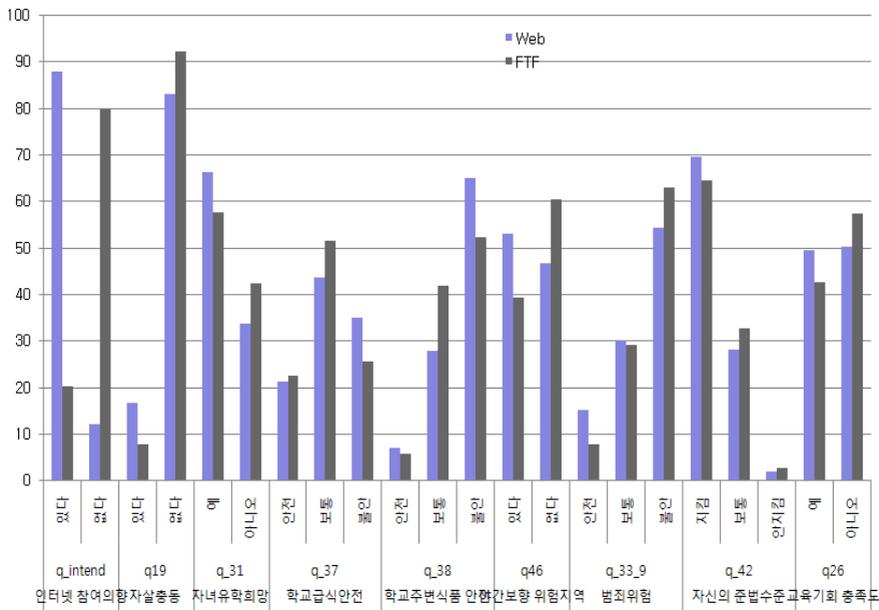
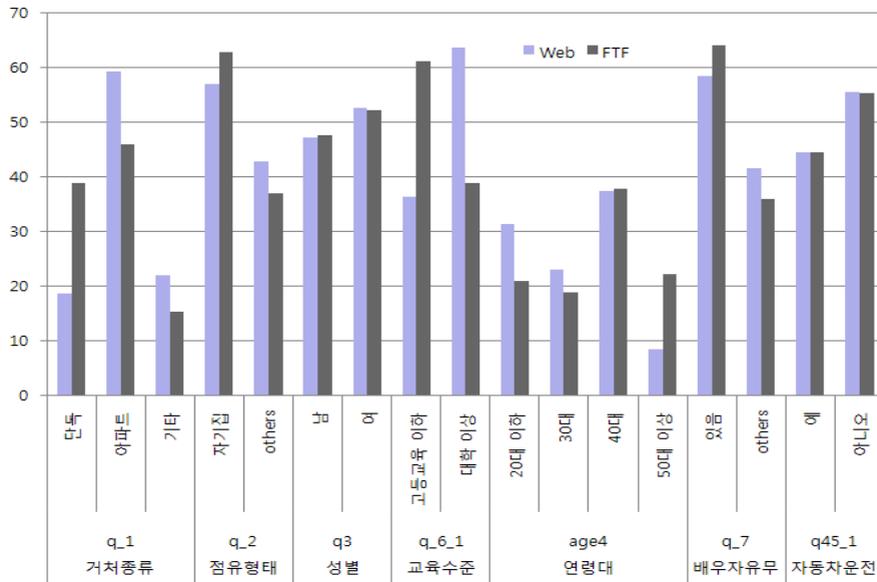
인구학적 및 비인구학적 특성변수			
질문내용	범주	질문내용	범주
1.연령대	5	10.자신의 준범수준인식	5
2.성별	2	11.범죄위험 인식정도	5
3.교육정도	8	12.부모부양 책임자에 대한 견해	5
4.배우자 유무	4	13.녹지환경변화 체감정도	5
5.자동차 운전여부	2	14.환경운동 참여노력정도	4
6.경제활동 행위여부	2	15.식품안전 인식도	5
7.가구소득수준	8	16.스트레스 체감정도	4
8.자살충동 경험여부	2	17.전반적 생활만족도	5
9.학교주변식품 안전인식	5		

〈표 5-3〉 로지스틱 모형에 의한 성향변수 선택

변수		odds	Pr ( $> z $ )	변수		odds	Pr ( $> z $ )
거처종류(ref. 단독)	아파트	1.151	0.200	자신의준법수준(ref. 잘지킴)	보통	0.959	0.649
	연립/다세대	1.731	0.000 ***		안지킴	0.572	0.068
점유형태(ref. 자택)	전세 등	0.957	0.623	범죄위험인식	보통	0.221	0.231
	성별(ref. 남자)	여자	1.096		0.497	안지킴	0.665
교육(ref. 고등 미만)	고등 이상	2.586	0.000 ***	교육기회의충족 (ref. 그렇다)	아니다	1.175	0.091
연령(ref. 20대 이하)	30대	1.342	0.102	부모부양책임 (ref. 본인, 가족)	정부, 사회	1.72	0.000 ***
	40대	1.781	0.003 ***	입양의향(ref. 의향 있음)	의향없음	1.052	0.539
	50대 이상	1.027	0.909	혼전동거가능(ref. 동의)	반대	1.003	0.969
혼인상태(ref. 유배우)	무배우	0.627	0.000 ***	가사분담책임 (ref. 절반이상 부인책임)	공평분담	1.116	0.238
	운전여부(ref. 예)	아니오	1.392	0.003 ***	절반 이상 남편책임	0.709	0.207
경제활동여부(ref. 예)	아니오	2.687	0.000 ***	녹지환경체감(ref. 좋아짐)	변함없다	0.745	0.002 **
	종사상지위 (ref. 임금근로자)	임금근로자 외	0.645	0.000 ***	나빠졌다	0.858	0.290
전문계고, 대학졸업여 부(ref. 예)	아니오	530.781	1.000	환경보호운동참여 (ref. 노력함)	노력안함	0.745	0.000 ***
직업전공일치도 (ref. 일치)	보통	0.922	0.553	전반적사회안전인식 (ref. 안전)	보통	0.824	0.156
	관계없음	1.158	0.217	불안	0.878	0.408	
가구소득 (ref. 200만원 미만)	200~400만 원 미만	1.311	0.012 **	식품안전인식 (ref. 안전)	보통	1.218	0.151
	400만원 이상	1.056	0.683	불안	1.41	0.019 **	
인터넷조사참여의향 (ref. 있다)	없다	0.033	0.000 ***	본인건강상태(ref. 좋은편)	보통	1.073	0.447
자살충동경험 (ref. 있다)	없다	0.599	0.000 ***	나쁜편	1.201	0.197	
교육비부담 (ref. 부담스럽다)	보통이다	0.727	0.061	전반적스트레스정도 (ref. 느낌)	느끼지 않는편	0.566	0.000 ***
	부담스럽지 않다	0.831	0.512	안느낌	0.311	0.002 **	
자녀유학소망(ref. 예)	아니오	0.983	0.892	학교생활만족도(ref. 만족)	보통	0.954	0.924
학교급식안전인식 (ref. 안전)	보통	1.367	0.005 **	불만족	1.724	0.477	
	불안전	1.435	0.004 **	생활만족도(ref. 만족)	보통	0.572	0.000 ***
학교주변식품안전 (ref. 안전)	보통	0.514	0.000 ***	불만족	0.645	0.000 ***	
	불안	0.626	0.012 *	흡연여부(ref. 피운다)	안피운다	1.155	0.110
야간보행안전 (ref. 두려움)	두렵지않음	1.142	0.120	지난1년간음주여부 (ref. 있다)	없다	0.854	0.173

\*\*\* p&lt;0.001, \*\* p&lt;0.01, \* p&lt;0.05

### 성향변수에 대한 준거조사와 웹 조사의 분포



[그림 5-3] 웹 조사와 준거조사 간의 성향 조정변수 결과 분포

17개 성향변수 중 몇 개 변수에 대해서 준거조사와 웹 조사 결과에 대한 분포를 살펴 보았다. “인터넷 참여의향 변수”도 참고로 삽입하였다. 이 변수는 준거조사 참여자와 웹 조사 참여자의 응답이 명확히 다른 특성을 나타낸다. 나머지 변수에 대한 정보는 성향조정 후 최종 결과표에서 확인할 수 있기 때문에 별도로 여기서는 제시하지 않기로 한다. [그림 5-3]에서 ‘FTF’는 본 연구에서 준거조사로 사용한 사회조사 대면조사를 의미한다.

[그림 5-3]에서 보면 성별과 “자동차운전”을 제외한 모든 변수에서 웹 조사와 준거조사 결과 사이에는 확실한 차이가 있는 것으로 나타났다. 연령변수는 준거조사와 웹 조사 모두에서 표본추출 변수로 사용되었지만, 웹 조사 자원자 패널의 연령분포가 젊은 층(20대, 30대)이 많고, 50대 이상이 적기 때문에 두 조사 간 분포는 차이가 있다고 볼 수 있다. 인터넷 참여의향 변수는 향후 자원자표본 웹 조사에 대한 참여의향이 있다고 응답한 사람들의 비율이 준거조사와 웹 조사에서 거의 반대 양상인 것으로 나타났다(준거조사 21% vs 웹 조사 88%). 자살충동에 대해서는 웹 조사에서 자살충동이 있다고 응답한 비율이 준거조사에 비해 약 10%p 높았다. 이처럼 그림에 제시된 변수 외에도 사회조사의 대부분 항목에서 두 조사 간의 응답 비율에 차이가 있었다. 이러한 응답성향의 차이를 제거할 수 있을 때 웹 조사 결과를 공표자료에 포함할 수 있을 것으로 보인다.

### 3. 성향점수 및 가중치 계산

여기에서는 성향점수 모형에 대해 최종 선택된 17개 성향변수에 대해 성향점수를 계산하고, 앞에서 설명한 두 가지 방법에 의해 성향점수 가중치를 작성하였다. 성향점수 계산은 로지스틱 회귀모형을 이용하였다. 성향점수 계산을 위해 우선 준거조사 자료와 웹 조사 자료를 하나의 자료셋으로 만들고, 이 통합된 자료에 대해서 로지스틱 모형을 적용하였다.

$$\log\left(\frac{ps}{1-ps}\right) = \alpha + \beta x.$$

여기서  $ps(x) = p(i=1|x)$ 로, 어떤 사람이 웹 조사에 참여할 확률을 의미한다. 성향점수 계산에 사용된 기본가중치는  $N_w/n_w$ 로, 시도(16개 특광역시), 성별(남여), 연령별(5개 범주)을 표본층으로 하여 추출률의 역수를 사용하였다. 한편 본 연구는 쉽게 사용할 수 있다는 점에서 기본가중치로 림 가중치도 사용해 보았다. 림 가중치 적용에 대해서는 그 결과만 뒤에서 간략하게 설명하기로 한다. 마지막으로 이렇게 추정된 성향점수를 이용하여 두 가지 방법으로 성향점수 가중치를 추정하였다.



① 층화방법

로지스틱 회귀모형으로 구한 성향점수를 크기순으로 정렬한 후, 최소값, 1사분위수, 중위수, 3사분위수, 최대값에 대해 구간을 나눈다. 각 구간 안에 포함된 개체들을 대상으로, 구간별 성향점수 조정 계수는

$$\text{구간의 성향점수 계수} = \frac{\text{구간에 포함된 준거조사 표본 수/전체 준거조사 표본 수}}{\text{구간에 포함된 웹표본 수/전체 웹표본 수}}$$

와 같이 구한다. 즉, 구간의 성향점수 가중치는 각 구간 안에서 총 표본수에 대한 웹 표본수의 비율로서, 동일한 성향점수를 갖는 표본 중에서 웹 표본이 차지하는 비율을 의미한다. 즉, 성향점수 가중치가 크다는 것은 해당 구간 안에서는 준거표본보다는 웹 표본일 가능성이 더 높다는 것을 의미한다.

② 성향점수의 역수방법

두 번째 방법은 성향점수의 역수를 이용하는 방법이다. 즉, 웹 표본에 적용될 성향점수 가중치는  $\frac{1-ps(x)}{ps(x)}$ , 준거표본에 적용될 성향점수는 1이다.

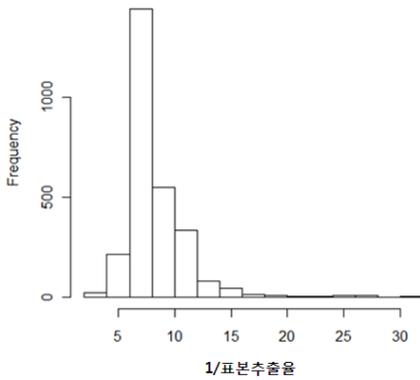
이 두 성향점수 가중치의 분포는 다음과 같다. 이때 가중치는 웹 조사 자료의 표본수의 크기로 조정된 값을 사용한 것이다. 즉, 가중치의 합이 실제 조사에 사용된 웹 조사의 표본크기가 된다.

<표 5-4> 두 성향가중치의 기초정보

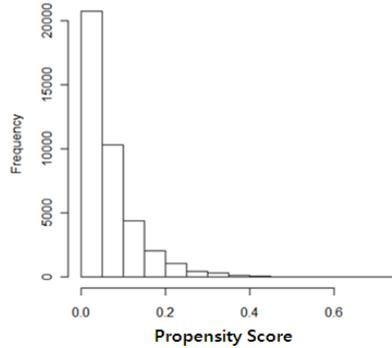
5개 층화 방법					성향점수 역수 방법				
최소	Q1	중위수	Q3	최대값	최소	Q1	중위수	Q3	최대값
0.11	0.35	0.58	1.21	15.77	0.40	4.32	7.95	15.04	230.50

먼저 층화방법의 경우, 성향점수 가중치의 분포는 최소값 0.11, 1사분위수(Q1) 0.35, 중위수 0.58, 3사분위수(Q3) 1.21, 최대값 15.77로 나타났다. 한편 성향점수 역수를 이용한 경우는 최소값 0.40, 1사분위수(Q1) 4.32, 중위수 7.95, 3사분위수(Q3) 15.04, 최대값 230.50으로 층화방법에 비해 범위가 넓고 오른쪽으로 심하게 치우친 경향이 있다.

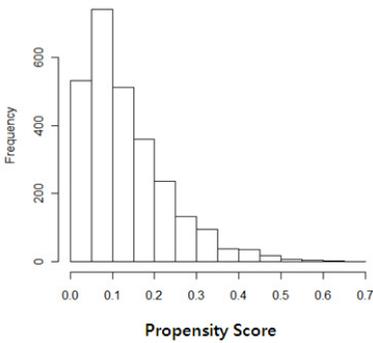
[그림 5-4]는 최종 성향가중치 분포, 추출률의 역수 및 성향점수 분포를 나타낸 것이다. [그림 5-4]의 (a)에서 추출률의 역수는 웹 조사의 기본가중치를 사용한 것으로, 약간 오른쪽으로 치우친 경향이 있다. 즉, 추출률이 매우 작은 값들이 존재한다는 것을 의미한다. 이는 이



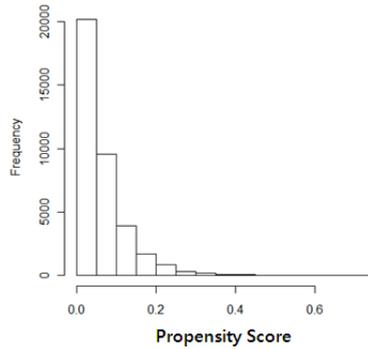
(a) 웹 조사 표본 추출률의 역수 분포



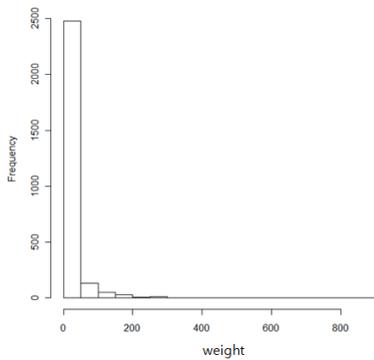
(b) 준거조사와 웹 조사에 대한 성향점수 분포



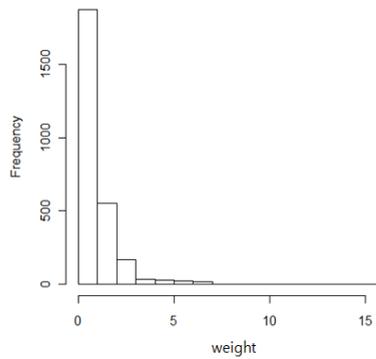
(c) 웹 조사에 대한 성향점수 분포



(d) 준거조사에 대한 성향점수 분포



(e) 성향점수 역수에 의한 성향가중치 분포



(f) 5개 층화방법에 의한 성향가중치 분포

[그림 5-4] 성향점수의 역수와 층화방법에 의한 성향가중치



들 표본층의 자원자가 매우 많지만 추출된 표본은 상대적으로 작은 경우에 해당될 수 있다. 사회조사의 경우, 대도시의 젊은 층에서 자원자가 많았다. 이러한 추출률 분포의 영향은 성향조정 가중치에 반영되어 나타났다. 그림 (b)~(d)는 성향점수 분포를 나타낸 것으로, (b)는 웹 조사와 준거조사 모두, (c)는 웹 조사, (e)는 준거조사 각각의 성향점수를 나타낸다. 그림 (e)와 (f)를 보면, 두 방법에 의한 성향가중치는 상대적으로 큰 값이 존재하는 것으로 나타났다. 이러한 기본가중치의 영향은 성향점수의 역수를 이용한 가중치 방법 (e)에서 더 크게 작용한 것으로 보인다. 이때 그림 (a)를 제외한 (b)~(f)에 사용된 수치는 가중치를 적용할 때, 가중치의 합이 각 조사자료의 표본크기가 되도록 크기를 조정하는 것이다.

이처럼 다른 값들에 비해 상대적으로 큰 가중치를 갖는 경우, 적절한 임계값 (cut value)을 사용하는 것이 일반적이다. 그러나 성향점수의 역수를 이용한 방법은 그 임계값으로 조정하기에 너무 큰 극값이 존재하기 때문에 근본적으로 발생하는 문제를 해결할 필요가 있다. 성향점수의 역수는  $\frac{1-ps}{ps}$  ( $p$ =어떤 사람이 웹 조사에 참여할 확률)와 같은 식을 사용하게 되고, 웹 조사에 참여할 확률  $ps$ 와 준거조사에 참여할 확률  $1-ps$ 의 차이가 매우 클 경우, 극값으로 나타날 가능성이 높다. 따라서 이러한 현상은 변수 선택이나 자료의 특성에 따라 민감하게 반응할 수 있을 것으로 예상된다. 반면에 5개 층화방법의 경우는 성향점수가 유사한 그룹에게 동일한 가중치를 부여하는 방법으로 기본가중치에 영향을 받는 것은 성향점수 역수 방법과 동일하지만, 상대적으로 그 효과는 작을 것으로 판단된다.

## 제5절 결 과

### 1. 평가방법

지금까지 웹 조사 결과를 준거조사에 대해 조정하기 위한 핵심 절차로서 3가지 단계를 거쳤다. ① 성향변수 선택, ② 성향점수 계산, ③ 성향가중치 계산 등이 이에 해당된다. 본 절에서는 성향조정 가중치에 의해 조정된 웹 조사 추정치가 조정 전에 비해 얼마나 준거조사 추정치에 근접해졌는가를 비교·평가한다. 즉, 준거조사 추정치는 웹 조사 추정치에 대해 벤치마킹 역할을 하게 된다. 평가척도로는 편향과 분산을 이용한다. 여기서 편향은 성향조정 전·후의 추정치 간의 차이를 보여주는 척도이고, 분산은 조정 전·후의 분산에 대한 증감 정도를 보여준다. 많은 선행 연구들에서는 분산보다는 편향을 조

정하는데 더 많은 관심을 가져왔던 것 같다. 그러나 가중치 적용 전에 비해 적용 후에 분산이 증가한다면, 이는 조사의 총 오차를 증가시키는 요인이 된다. 따라서 조사추정치에 대해 가중치를 부여하기 전과 후의 분산 변화는 신중하게 고려할 필요가 있을 것이다. 다음과 같은 절차에 따라 분석결과를 비교하였다.

첫째, 사회조사 설문지를 영역별로 나누고, 각 영역에 대해 가중치를 각각 적용한 후 그 결과를 비교 설명하였다. 지면의 한계상 103개 모든 항목에 대한 결과를 본문에서 설명하기는 어렵기 때문에 우선적으로 성향변수로 사용된 항목들에 대해 평가하고, 성향조정 효과를 검증하였다. 그리고 성향변수로 사용되지 않은 가구·개인 관련 사항과 보건, 교육, 가족, 안전, 환경 분야는 각 분야별로 각각 가중치 적용 후의 추정치를 계산하였지만, 본 보고서에는 ‘안전’ 부문에 대해서만 <부표 5-3>에 제시하기로 하였다.

둘째, 성향점수 가중치 작성 방법별로 분석결과를 설명한다. 본 연구에서는 성향가중치를 사용하는데 있어서 기본가중치로 웹 표본추출률의 역수와 림 가중치를 사용하고 분산과 편향을 통해 두 방법들의 특성을 비교해 보았다.

셋째, 성향점수 가중치를 계산할 때 층화방법과 성향점수의 역수 등의 두 가지 방법을 사용하였다. 즉, 두 가지 기본가중치 가운데 자료에 더 적합한 방법을 사용함으로써 좀더 성능이 좋은 성향점수 가중치를 작성하고자 하였다.

넷째, 웹 조사추정치에 가중치를 적용하기 전과 후의 결과를 평가하기 위한 척도로는 이미 설명한바와 같이 편향과 분산을 이용하였다. 해당 가중치를 조정된 후의 웹 조사 추정치가 준거조사 추정치에 더 가까워질수록, 즉 편향이 작을수록 그리고 분산이 작을수록 우수한 방법으로 평가하였다.

## 2. 평가결과

### 가. 기본가중치 선택

웹 조사에 참여한 사람들의 성향점수는 로지스틱 회귀모형을 사용하여 계산하였다. 이때 웹 조사에 참여한 표본은 표본이 자발적으로 조사에 참여하였기 때문에 할당추출(quota sampling)에 해당하는 비확률표본으로서 이런 경우는 표본이 추출된 확률은 계산할 수 없는 것이 일반적이다. 본 연구를 위해 사용된 사회조사의 경우 자발적 참여에 의한 자원자 패널을 우선적으로 구성하고, 그 자원자 패널로부터 확률층화비례계통 추출방법에 의해 표본을 추출하였다. 따라서 최종 웹 자원자 패널로부터 선택된 표본은 패널 안에서 확률표본으로 간주하여 추출률의 역수를 기본가중치로 사용하였다. 기본가중치를 모르는 경우는 일반적으로 기본가중치로 1을 사용한다. 따라서 본 연구는 기본가중치로 추출률의 역수를 사용하였다. 그리고 이 결과와 림 가중치를 적용한 결과를 비교하



였다. 림 가중치는 앞에서 설명한 바와 같이 모집단 구성비에 인구학적 특성비를 맞추어 사용하는 간단한 방법이다. 그러나 림 가중 조정으로 조정할 변수에 대해 모집단 분포를 사전에 알고 있어야 한다.

분석결과는 <표 5-5>와 같다. <표 5-5>는 성향조정 전 웹 조사의 추정치와 추출물의 역수를 사용한 경우와 림 가중치를 사용한 경우의 웹 조사 추정치를 제시하였다. 그리고 각 3가지 추정치에 대한 표준편차와 준거조사에 대한 각 추정치들의 차이를 제시하였다. 평가변수는 우선 17개 성향 조정변수에 대해서만 설명하였다. 그리고 <표 5-3>에 17개 중 7개 변수에 대한 결과만 제시하고, 나머지는 <부록>에 제시하였다.

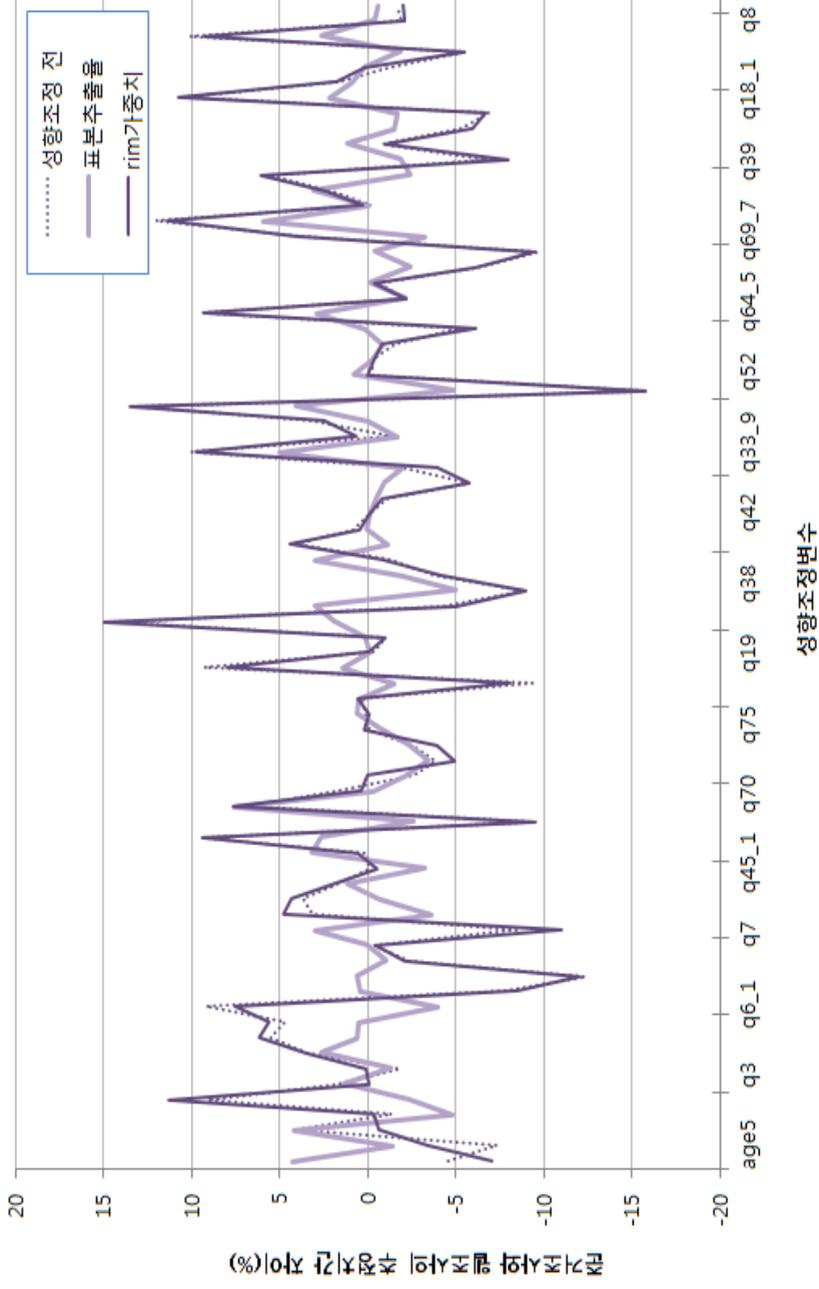
우선 [그림 5-5]를 통해 준거조사에 대해 조정 전과 각 방법에 의한 조정 후 추정치 간의 차이를 살펴보았다. 준거조사에 대해 성향조정 전과 후의 차이가 작을수록 좋다. 전체적으로 추출물의 역수를 사용한 성향조정 후 추정치와 준거조사 추정치 간의 차이가 가장 작을 것으로 나타났다. 그리고 림 가중치를 사용하여 조정한 경우는 원자료와 준거조사 간의 차이보다 약간 작긴 하지만 성향조정효과가 추출물의 역수를 사용한 경우에 비해 크지 않다는 것을 알 수 있다.

또한 분산을 통해 각 추정치들의 정도를 살펴보았다. [그림 5-6]은 성향조정 전과 추출물의 역수 및 림 가중치를 사용하여 조정한 추정치에 대한 표준편차를 각각 나타낸 것이다. 표준오차가 작을수록 성향조정 효과가 더 좋은 것으로 판단한다. 일반적으로 가중치를 적용하면 조정 전에 비해서 분산이 약간 증가하는 경향이 있다. [그림 5-6]에서도 조정 전 웹 조사 추정치의 표준오차가 가장 작다. 우리가 관심을 갖는 부분은 추출물의 역수를 사용한 경우와 림 가중치를 사용한 경우이다. 두 조정 방법에 의한 추정치의 표준오차가 대체로 비슷하거나 추출물의 역수를 사용하여 조정한 추정치의 경우가 표준오차가 약간 낮은 경향이 있다.

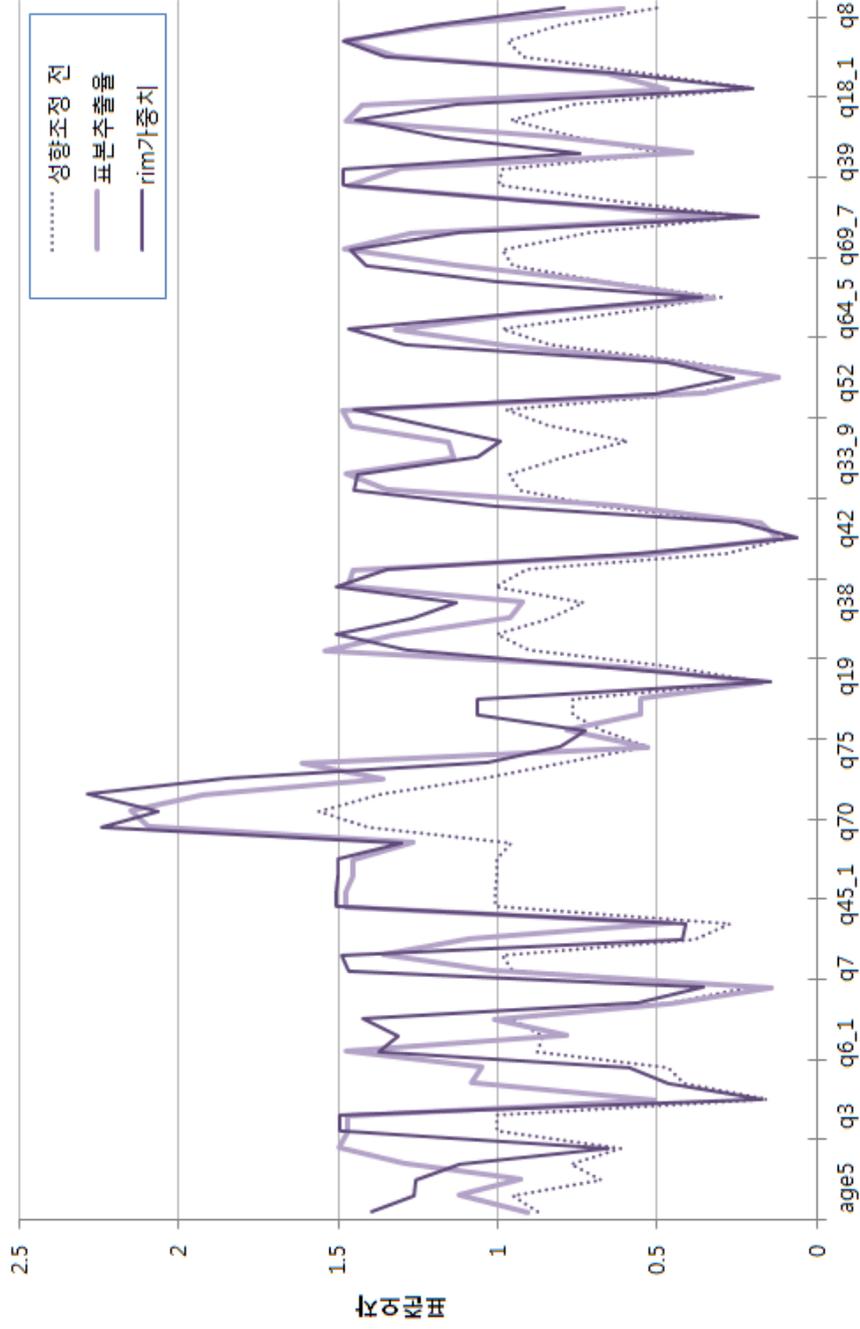
따라서 기본가중치 선택은 준거조사 추정치에 대해 추정치의 표준오차에 근거하여 추정치 간 차이와 표준오차가 작은 웹 조사 표본 추출물의 역수를 선택하였다. 어떤 기본가중치를 사용하느냐에 따라 조정효과에 있어서 큰 차이는 없을 것으로 생각된다. 그렇다하더라도 모집단의 분포 등 사전조건을 고려하면 추출물의 역수를 기본가중치로 사용하는 것이 성향조정 후 추정치의 편향과 분산 측면에서 더 우수한 결과를 줄 것으로 판단된다.

〈표 5-5〉 기본가중치 (표본추출률의 역수 vs. 림 가중치) 사용에 따라 성향조정 효과

설문문항	준거 조사	웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 차이		
		성향 조정 전	표본 추출률	림 가중치	성향 조정 전	표본 추출률	림 가중치	성향 조정 전	표본 추출률	림 가중치
연령	20대 이하	24.23	19.96	31.21	0.88	0.91	1.40	-4.53	4.27	-6.99
	30대	20.00	21.41	23.34	0.95	1.12	1.26	-7.33	-1.41	-3.34
	40대	20.69	16.63	21.28	0.68	0.93	1.26	4.06	4.24	-0.59
	50대	16.49	17.81	16.84	0.77	1.29	1.12	-1.32	-4.80	-0.35
	60대 이상	18.60	9.47	7.32	0.62	1.50	0.65	9.13	-2.30	11.28
성별	남자	49.19	47.38	49.29	1.00	1.47	1.50	1.80	1.33	-0.10
	여자	50.81	52.62	50.71	1.00	1.47	1.50	-1.80	-1.33	0.10
교육 정도	무학	4.04	0.56	1.43	0.16	0.51	0.17	3.48	2.61	3.55
	초등학교	9.72	4.20	3.56	0.42	1.08	0.46	5.52	0.66	6.16
	중학교	9.91	5.22	4.24	0.46	1.05	0.59	4.69	0.50	5.68
	고등학교	35.55	26.38	27.99	0.88	1.47	1.38	9.17	-3.95	7.56
	전문대	15.02	22.83	23.49	0.86	0.78	1.31	-7.81	0.44	-8.47
	4년대	22.18	34.53	34.18	0.96	1.01	1.42	-12.35	0.65	-12.00
	석사	2.79	5.14	4.86	0.45	0.46	0.56	-2.35	-1.02	-2.07
	박사	0.79	1.14	1.20	0.21	0.14	0.36	-0.35	0.11	-0.41
혼인 상태	미혼	27.02	35.44	37.98	0.95	1.02	1.47	-8.42	3.00	-10.96
	유배우	62.38	59.26	57.58	0.98	1.36	1.49	3.12	-3.57	4.80
	사별	7.07	3.37	2.70	0.38	1.09	0.42	3.69	-0.64	4.36
	이혼	3.53	1.92	1.74	0.27	0.46	0.41	1.61	1.22	1.79
윤전 여부	예	44.99	45.07	45.54	1.01	1.48	1.51	-0.09	-3.20	-0.56
	아니오	55.01	54.93	54.46	1.01	1.48	1.51	0.09	3.20	0.56
경제활동 여부	예	55.33	45.81	45.97	1.00	1.46	1.50	9.52	2.62	9.36
	아니오	44.67	54.19	54.03	1.00	1.46	1.50	-9.52	-2.62	-9.36
자살충동 경험	있다	7.69	16.98	15.70	0.76	0.55	1.06	-9.29	-1.50	-8.01
	없다	92.31	83.02	84.30	0.76	0.55	1.06	9.29	1.50	8.01



[그림 5-5] 림 가중치와 추출률의 역수를 이용한 성향조정 가중치 조정 후 추정치



성향조정변수

[그림 5-6] 림 가중치와 추출률의 역수를 이용한 성향조정 기중치 조정 후 추정치의 분산

## 나. 성향점수 가중치 적용 방법

본 절에서는 성향점수를 이용하여 가중치를 계산하였다. 계산방법은 성향점수를 5개 층으로 나누어 계산한 층화방법과 성향점수의 역수를 이용하였다. 각 방법에 대해 준거표본 결과와 비교하였다. 편향은 준거표본 추정치와 각 방법에 대한 성향점수 가중치 적용 후 추정치와의 차이를 통해 측정하였다. 차이값이 작을수록 더 좋은 방법으로 평가하였다. 그리고 각 성향점수 조정 가중치 적용 후의 추정치에 대한 분산이 작을수록 더 좋은 가중치 방법으로 평가하였다. <표 5-6>은 준거표본, 5개 층화방법, 성향점수의 역수를 이용한 추정치와 각각의 차이 그리고 각 추정치의 분산을 나타낸 것이다.

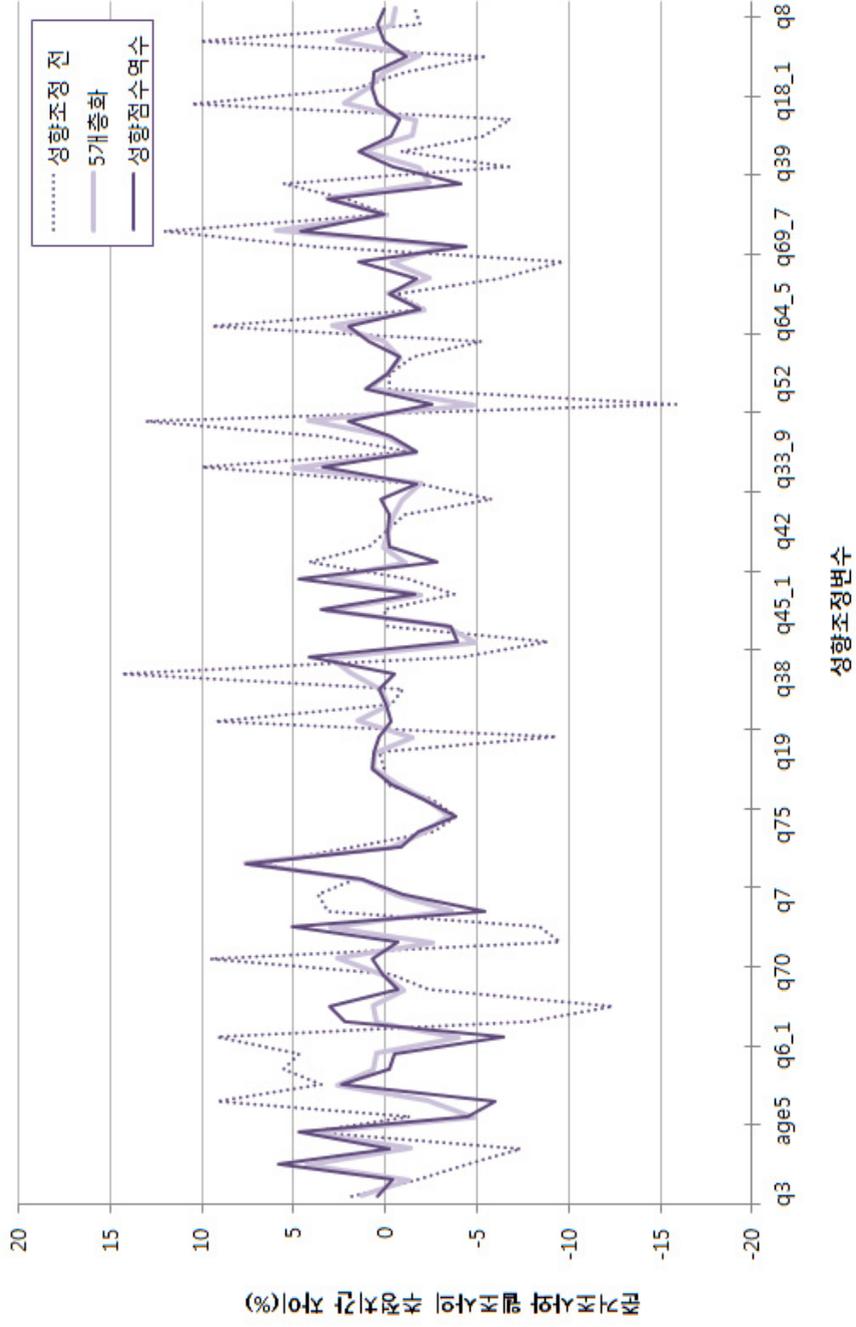
<표 5-6>에서 ①은 해당변수에 대해서 준거표본 추정치 ②~④는 가중치 조정 전 웹 조사추정치, 5개 층화방법에 의한 가중치 조정 후 추정치, 성향점수의 역수를 이용한 가중치 조정 후의 추정치를 나타낸다. ⑤~⑦은 각 방법에 대한 표준오차, ⑧~⑩은 각 추정치와 준거조사 추정치 간의 차이를 나타낸다. <표 5-6>과 [그림 5-7]을 보면, 대체로 5개 층화방법에 의한 추정치와 성향점수 역수에 의한 추정치와 준거조사 추정치 간의 차이는 거의 없어 보인다. 그리고 웹 조사의 가중치 조정 전의 추정에 비해 두 성향가중치 조정후의 추정치는 상대적으로 준거조사 추정치에 대해 훨씬 근사해지는 것을 알 수 있다. 즉, 성향점수 가중치 조정은 웹 조사 추정치에 대해 성향조정 효과가 있고, 편향 측면에서는 성향점수 가중치 방법 간에 큰 차이가 없다고 볼 수 있다.

이제 분산 측면에서 성향점수 가중치 조정 방법을 비교해 보자. [그림 5-8]은 <표 5-6>의 웹 조사 추정치, 5개 층화방법에 의한 가중치 조정 후 추정치, 성향점수 역수를 이용한 가중치 조정 후 추정치에 대한 표준오차를 그린 것이다. 그림에서 ‘성향조정 전’은 가중치 조정 전의 웹 조사 추정치, ‘5개 층화’는 5개 층화방법에 의한 가중치를 적용한 추정치, ‘성향점수 역수’는 성향점수 역수 이용방법을 이용한 가중치를 적용한 추정치를 의미한다. 전체적으로 성향점수 가중치 조정 후에 웹 조사 추정치의 분산은 조정 전의 웹 조사 추정치의 분산보다 커지는 경향이 있고, 5개 층화방법에 의한 가중치 조정 후 추정치의 분산(5개 층화)이 성향점수 역수를 이용한 가중치 조정 후 추정치의 분산(성향점수 역수)보다 약간 작다는 것을 알 수 있다.

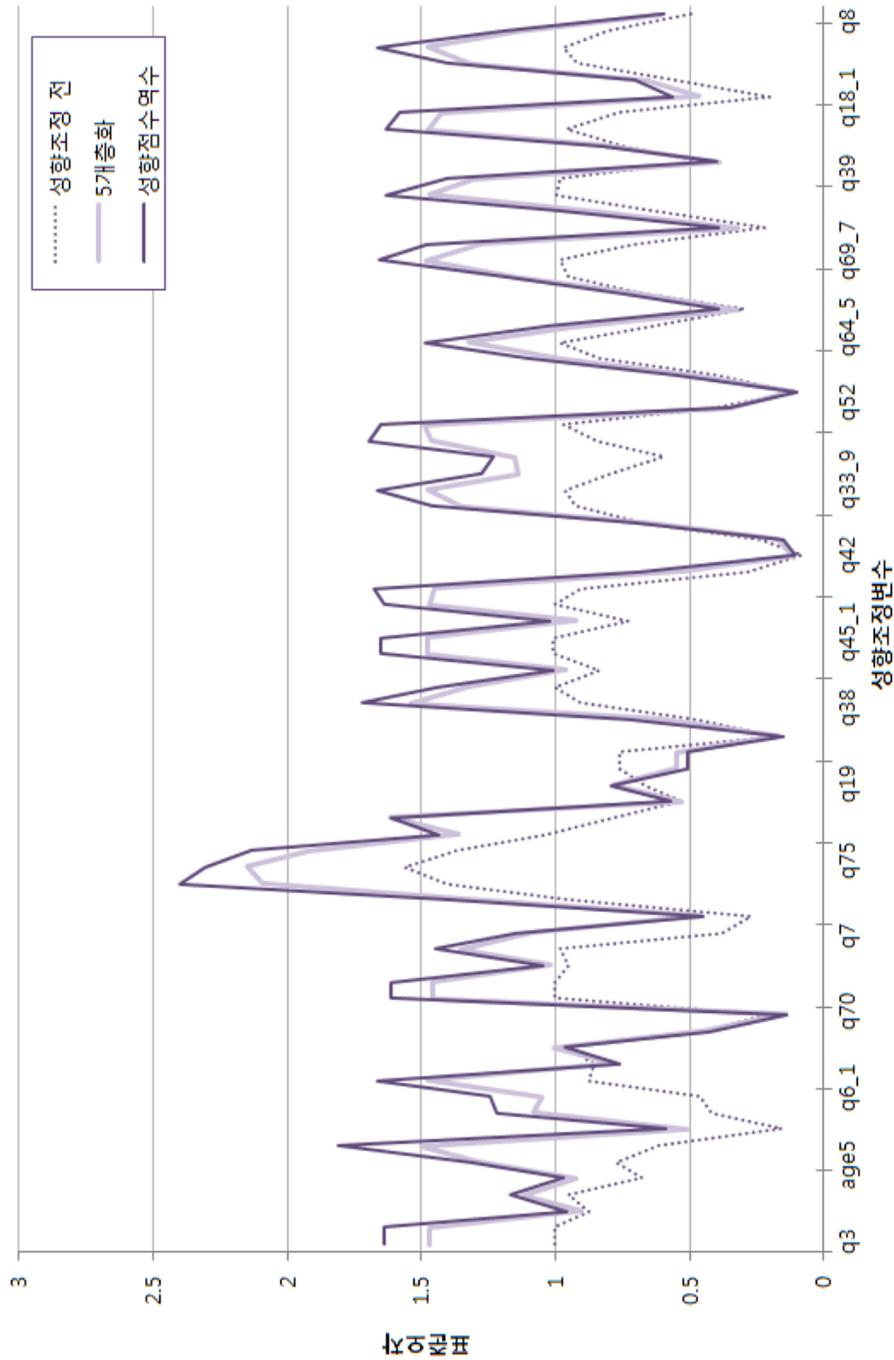
따라서 <표 5-6>, [그림 5-7], [그림 5-8]에 의하면 성향점수 가중치 조정방법에 있어서 5개 층화방법과 성향점수 역수 방법은 편향 측면에서는 큰 차이는 없지만, 분산이 상대적으로 작은 5개 층화방법을 본 연구에서는 선택하기로 하였다. 따라서 이후부터 본 연구의 설명은 5개 층화방법에 의한 가중치 방법을 성향조정 가중치 방법으로 사용하였다.

〈표 5-6〉 준거표본과 웹 성향조정 추정치의 분산과 편차

설문문항	준거조사 추정치			웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 차이		
	조정 전 원자료	5개 층화	성향점수 역수	조정 전 원자료	5개 층화	성향점수 역수	조정 전 원자료	5개 층화	성향점수 역수	조정 전 원자료	5개 층화	성향점수 역수
성별	남	49.19	47.38	48.80	1.00	1.47	1.64	1.80	1.33	1.80	1.33	0.39
	여	50.81	52.62	51.20	1.00	1.47	1.64	-1.80	-1.33	-1.80	-1.33	-0.39
연령	20대이하	24.23	28.76	18.36	0.88	0.91	0.96	-4.53	4.27	-4.53	4.27	5.87
	30대	20.00	27.33	20.18	0.95	1.12	1.17	-7.33	-1.41	-7.33	-1.41	-0.18
	40대	20.69	16.63	15.94	0.68	0.93	0.97	4.06	4.24	4.06	4.24	4.75
	50대	16.49	17.81	20.99	0.77	1.29	1.32	-1.32	-4.80	-1.32	-4.80	-4.50
	60대이상	18.60	9.47	24.53	0.62	1.50	1.81	9.13	-2.30	9.13	-2.30	-5.93
교육 정도	무학	4.04	0.56	1.43	0.16	0.51	0.59	3.48	2.61	3.48	2.61	2.36
	초등학교	9.72	4.20	9.94	0.42	1.08	1.22	5.52	0.66	5.52	0.66	-0.22
	중학교	9.91	5.22	10.38	0.46	1.05	1.25	4.69	0.50	4.69	0.50	-0.47
	고등학교	35.55	26.38	41.98	0.88	1.47	1.67	9.17	-3.95	9.17	-3.95	-6.44
	전문대	15.02	22.83	12.85	0.86	0.78	0.76	-7.81	0.44	-7.81	0.44	2.17
	4년대	22.18	34.53	19.09	0.96	1.01	0.97	-12.35	0.65	-12.35	0.65	3.08
	석사	2.79	5.14	3.45	0.45	0.46	0.43	-2.35	-1.02	-2.35	-1.02	-0.65
	박사	0.79	1.14	0.63	0.21	0.14	0.14	-0.35	0.11	-0.35	0.11	0.16
경제활동 여부	예	55.33	45.81	54.62	1.00	1.46	1.62	9.52	2.62	9.52	2.62	0.71
	아니오	44.67	54.19	45.38	1.00	1.46	1.62	-9.52	-2.62	-9.52	-2.62	-0.71
혼인 상태	미혼	27.02	35.44	21.90	0.95	1.02	1.05	-8.42	3.00	-8.42	3.00	5.12
	유배우	62.38	59.26	67.84	0.98	1.36	1.45	3.12	-3.57	3.12	-3.57	-5.46
	사별	7.07	3.37	8.04	0.38	1.09	1.15	3.69	-0.64	3.69	-0.64	-0.98
	이혼	3.53	1.92	2.22	0.27	0.46	0.45	1.61	1.22	1.61	1.22	1.31
자살충동 경험	있다	7.69	16.98	7.34	0.76	0.55	0.51	-9.29	-1.50	-9.29	-1.50	0.35
	없다	92.31	83.02	92.66	0.76	0.55	0.51	9.29	1.50	9.29	1.50	-0.35



[그림 5-7] 성향조정 방법에 따른 준거조사와 웹 조사 추정치 간 차이



[그림 5-8] 성향조정 방법에 따른 웹 조사 추정치의 표준오차

### 다. 5개 총화방법에 의한 성향점수 기중치 조정 효과

〈표 5-7〉 준거조사와 웹 조사 성향조정 전후 추정치와 표준오차 효과크기

		①준거 조사	②웹 조사				③표준오차		④효과크기#	
			성향조정 전		5개 총화		성향조정 전	5개 총화	성향조정 전	5개 총화
연령대	20대 이하	24.23	28.76	***	19.96	***	0.88	0.91	0.04	0.04
	30대	20.00	27.33	***	21.41		0.95	1.12	0.07	0.01
	40대	20.69	16.63	***	16.45	***	0.68	0.93	0.04	0.05
	50대	16.49	17.81	**	21.29	***	0.77	1.29	0.02	0.05
	60대 이상	18.60	9.47		20.90		0.62	1.50	0.11	0.02
성별	남	49.19	47.38		47.86		1.00	1.47	0.01	0.01
	여	50.81	52.62		52.14		1.00	1.47	0.01	0.01
교육정도	무학	4.04	0.56	***	1.43	**	0.16	0.51	0.09	0.06
	초등학교	9.72	4.20	***	9.05		0.42	1.08	0.09	0.01
	중학교	9.91	5.22		9.41		0.46	1.05	0.07	0.01
	고등학교	35.55	26.38	***	39.50	**	0.88	1.47	0.08	0.03
	전문대	15.02	22.83	***	14.58		0.86	0.78	0.09	0.01
	4년대	22.18	34.53	***	21.53		0.96	1.01	0.12	0.01
	석사	2.79	5.14	***	3.82	***	0.45	0.46	0.06	0.03
혼인상태	박사	0.79	1.14		0.68		0.21	0.14	0.02	0.01
	미혼	27.02	35.44	***	24.02		0.95	1.02	0.07	0.03
	유배우	62.38	59.26	***	65.95	*	0.98	1.36	0.02	0.03
	사별	7.07	3.37	***	7.71		0.38	1.09	0.07	0.01
운전여부	이혼	3.53	1.92		2.32		0.27	0.46	0.04	0.03
	예	44.99	45.07		48.19	*	1.01	1.48	0.00	0.02
경제활동 여부	아니오	55.01	54.93		51.81		1.01	1.48	0.00	0.02
	예	55.33	45.81	***	52.71		1.00	1.46	0.07	0.02
가구소득	아니오	44.67	54.19		47.29		1.00	1.46	0.08	0.02
	~100만원	15.62	8.79	***	7.98	***	0.96	1.27	0.09	0.10
	100~200만원	23.92	22.22		24.18		1.41	2.09	0.02	0.00
	200~300만원	24.08	26.71		26.24		1.56	2.15	0.03	0.02
	300~400만원	15.36	19.16	**	18.79	*	1.36	1.92	0.05	0.04
	400~500만원	8.43	10.93	**	10.73	*	1.04	1.36	0.04	0.03
	500~600만원	5.65	5.72		6.37		0.79	1.62	0.00	0.01
600~700만원	2.76	2.66		2.11		0.53	0.53	0.00	0.02	
자살충동 경험	700만원 이상	4.18	3.82		3.60		0.67	0.78	0.01	0.01
	있다	7.69	16.98	***	9.19	**	0.76	0.55	0.15	0.02
학교주변 식품 안전도	없다	92.31	83.02		90.81		0.76	0.55	0.08	0.01
	매우 안전	0.49	0.82		0.63		0.18	0.17	0.02	0.01
	안전	5.28	6.26		5.08		0.48	0.62	0.02	0.00
	보통	42.17	27.77	***	40.27		0.91	1.54	0.12	0.01
	불안	39.19	43.40		36.21	*	1.00	1.32	0.04	0.02
	매우 불안	12.87	21.76		17.81		0.84	0.96	0.11	0.06

		①준거 조사	②웹 조사		③표준오차		④효과크기#			
			성향조정 전	5개 층화	성향조정 전	5개 층화	성향조정 전	5개 층화		
자기준법 수준	아주 잘지킴	11.94	15.65		13.90		0.73	0.92	0.05	0.03
	비교적 잘지킴	52.89	54.17		49.85		1.01	1.47	0.01	0.02
	보통	32.29	28.10		33.36		0.91	1.45	0.04	0.01
	비교적 지키지않음	2.78	1.92	**	2.67		0.28	0.52	0.02	0.00
	전혀지키지않음	0.09	0.16		0.22		0.08	0.12	0.01	0.02
범죄위험 인식	매우 안전	0.51	1.57	***	0.94	**	0.24	0.18	0.06	0.03
	비교적 안전	7.66	13.42	***	8.55		0.68	0.63	0.09	0.01
	보통	28.6	30.52	***	30.58		0.93	1.35	0.02	0.02
	비교적 불안	46.06	36.02	***	41.07	***	0.97	1.47	0.08	0.04
부모부양 책임	매우 불안	17.17	18.48		18.87		0.79	1.14	0.01	0.02
	스스로 해결	12.68	9.44	***	12.66	*	0.59	1.15	0.04	0.00
	가족	36.05	23.02	***	31.89	**	0.84	1.46	0.11	0.03
	가족, 정부사회	47.39	63.20		52.22		0.97	1.49	0.12	0.04
	정부사회	3.85	4.11		3.02	***	0.40	0.35	0.01	0.02
녹지환경 체감	기타	0.03	0.23		0.22		0.11	0.12	0.04	0.04
	매우 좋아짐	2.85	4.34	*	3.64		0.40	0.47	0.04	0.02
	약간 좋아짐	16.95	22.18		16.82		0.84	0.97	0.06	0.00
	변함없음	70.01	60.57	***	67.10	*	0.99	1.32	0.07	0.02
	약간 나빠짐	8.58	10.73		10.71		0.63	0.93	0.03	0.03
환경보호 운동참여	매우 나빠짐	1.61	2.18		1.73		0.30	0.32	0.02	0.00
	매우 노력	6.61	12.81	***	9.01	***	0.67	0.71	0.11	0.04
	약간 노력	23.63	33.33	***	24.02		0.96	1.14	0.10	0.00
	별로 노력안함	42.59	38.74	***	45.79		0.98	1.48	0.03	0.02
식품 안전인식	전혀 노력안함	27.17	15.12		21.18		0.71	1.27	0.11	0.05
	매우 안전	1.20	1.30		1.25		0.22	0.32	0.00	0.00
	약간 안전	13.88	11.8	**	10.73	*	0.64	0.85	0.03	0.04
	보통	47.62	42.03	**	50.05		1.00	1.47	0.04	0.02
	약간 불안	32.15	38.89	**	34.02		0.99	1.30	0.06	0.02
전반적 스트레스 정도	매우 불안	5.15	5.98		3.95		0.47	0.39	0.02	0.02
	매우 많이 느낌	11.41	16.71	***	12.88	*	0.75	0.82	0.07	0.02
	느끼는 편	58.63	65.48	***	60.31		0.96	1.47	0.05	0.01
	느끼지 않는편	27.30	16.77	***	25.09		0.76	1.43	0.10	0.02
생활 만족도	전혀 느끼지 않음	2.65	1.04		1.72		0.20	0.47	0.05	0.03
	매우 만족	7.15	8.29		7.01		0.56	0.66	0.02	0.00
	약간 만족	23.62	29.12	**	25.50		0.92	1.32	0.05	0.02
	보통	45.70	35.57	***	43.04		0.97	1.48	0.08	0.02
	약간 불만족	18.53	20.37		18.91		0.81	1.15	0.02	0.00
	매우 불만족	5.00	6.65		5.54		0.50	0.61	0.03	0.01

주: 유의수준 \* 5%, \*\* 0.01%, \*\*\* 0.001%에서 준거조사 추정치와 웹 조사 추정치가 유의하게 다르다.

$$\# (\mu_{ref} - \mu_{web}) / std$$

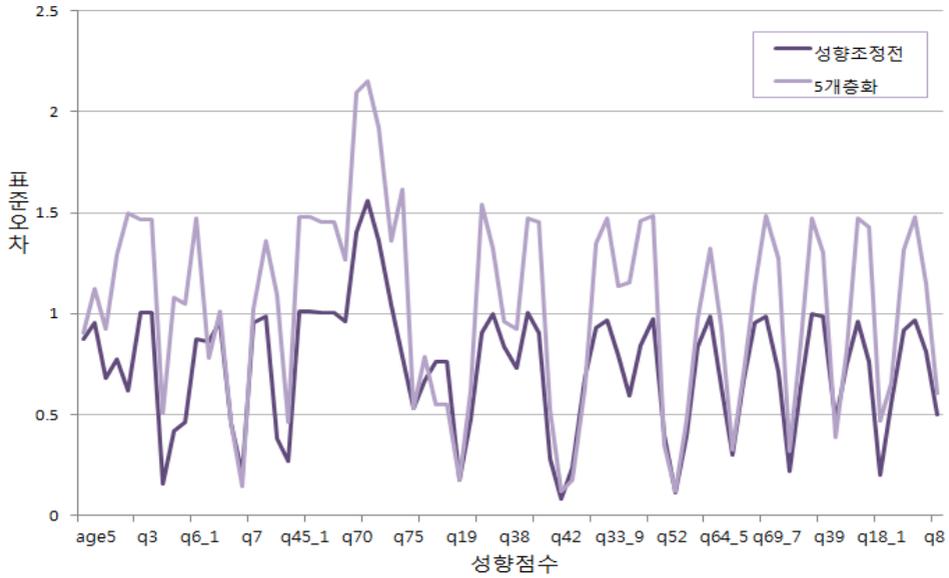


본 절에서는 5.나 절에서 선택한 성향점수 가중치 방법으로 5개 증화방법을 이용한 가중치 조정효과를 설명하고자 한다. 우선 성향 조정변수로 사용된 17개 변수에 대한 평가 결과를 설명하고, 2010년 사회조사 5개 분야 중 2개 분야를 임의로 선정하여 분야별 변수에 대한 평가결과를 설명하고자 한다. 이때 지면상 이 두 분야에 대한 설명은 본문에서 간단히 설명하고, 자세한 평가결과는 <부록>에 제시하였다. 표에서 ①은 준거조사 추정치 ②는 웹 조사 성향조정 전후 추정치 ③ 성향조정 전후 표준오차 ④는 준거조사 대비 성향조정 전후 추정치의 효과를 나타낸다. 이때 ④에 해당하는 효과크기는 준거조사추정치와 웹 조사 추정치 간 차이를 표준편차로 나눈 값을 나타낸다. 즉,  $(\mu_{ref} - \mu_{web}) / \sqrt{(\text{합동표본분산})}$ . 여기서 합동표본분산을 준거조사결과와 웹 조사결과 전체에 대한 분산을 의미한다.

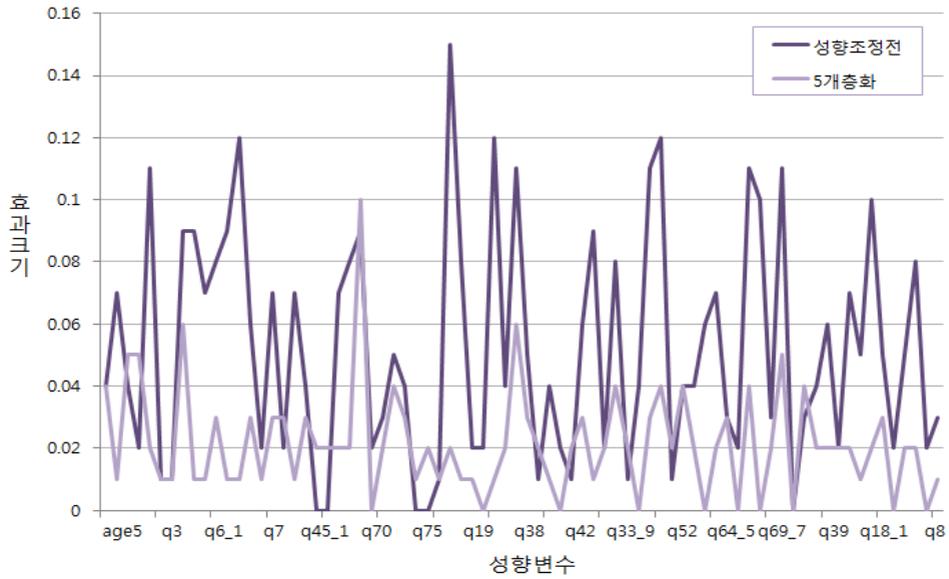
<표 5-7>에서 ②의 결과는 웹 자원자 성향조정 전과 후의 각 평가변수에 대한 추정치이고, 유의수준 0.05에서 성향조정 전과 후의 추정치 각각이 준거조사 추정치에 대해 얼마나 다른지를 검정한 결과이다. <표 5-7>의 ②에서 보면 성향조정 전에는 대부분 변수에서 준거조사와 웹 조사 추정치 간에 유의한 차이가 있는 것으로 나타났다. 이에 대해 성향조정 결과 상대적으로 17개 평가변수 중 10개 변수에서 대체로 준거조사와 웹 조사 간 추정치에 차이가 없는 것으로 나타났다. 하지만 여전히 7개 변수에서는 차이가 있는 범주들을 포함하고 있었다. 한편 7개 변수에서 이러한 차이는 조정 전에 비해 약간 줄어들기는 하지만 3개 변수는 유의성 정도(\*의 개수)가 변하지 않거나 조정 전에는 차이가 없던 범주들에서 차이가 나타나기도 하였다. 추가적으로 <부록>에 안전 분야에 검정결과를 제시하였다. 여기에 사용된 변수들은 성향 조정변수가 아닌 순수한 평가변수에 해당된다. 많은 변수들에서 조정 전에 비해 조정 후에는 준거조사와의 차이가 줄어든 것을 알 수 있다. 그렇지만 여전히 통계적 유의성 검정결과는 유의수준 0.05 수준에서 준거조사와 웹 조사의 추정치 간에는 차이가 있는 것으로 나타났다.

이처럼 성향가중치 조정으로 준거조사에 대한 웹 조사 추정치의 차이가 상당히 줄어드는 것은 사실이지만, 그 차이가 완전히 없어졌다고 보기 어렵고 특히 통계적 입장에서 여전히 두 조사추정치 간에는 차이가 있다고 볼 수 있다.

[그림 5-9]에서 가로축은 평가변수의 범주를 나타내고, 세로축은 표준오차값을 나타낸다. <표 5-7>과 [그림 5-9]를 보면, 웹 조사에 대해 성향점수 가중치를 적용한 후의 추정치의 표준오차는 가중치 적용 전 추정치에 비해 표준오차가 상당히 커지는 것을 알 수 있다. 또한 [그림 5-10]은 <표 5-7>의 효과크기를 그림으로 나타낸 것이다. [그림 5-10]에서 보면 준거조사 추정치에 대해 웹 조사의 성향조정 전에 비해 조정 후에 상당히 성향조정 효과가 있는 것을 알 수 있다. 즉, 성향조정 전의 효과크기에 비해 성향조정 후의 효과크기가 모든 변수의 대부분 범주에서 작아졌다는 것이다. 그리고 효과크기는 조정 전에는 최소 0에서 최대 0.15 이었던 것이 조정 후에는 최소 0에서 최대 0.1로 작아졌다.



[그림 5-9] 웹 조사의 성향조정 전과 후의 표준오차



[그림 5-10] 웹 조사 성향조정 전과 후의 준거조사 추정치에 대한 효과크기

종합적으로 볼 때, 성향점수 가중치 조정 후의 웹 조사 추정치는 준거조사 추정치에 보다 더 근사해지지만 추정치의 분산은 조정 전에 비해 상대적으로 커지는 경향이 있는 것으로 나타났다. 또한 성향조정 후의 웹 조사 추정치는 준거조사에 대한 편향이 줄어들긴 하지만, 여전히 조정 후에도 준거조사와 웹 조사 추정치 간 차이는 유의한 변수가 많은 것으로 나타났다. 따라서 지금까지의 성향조정 결과로 볼 때, 성향점수 조정에 의한 조정효과가 분명히 있지만 완벽한 조정이 이루어졌다고 보기는 어렵다. 게다가 성향조정 후 추정치의 분산이 증가하기 때문에 실질적 활용까지는 충분한 연구가 있어야 할 것이다.

## 제6절 결론 및 향후 연구

웹 조사는 인터넷의 발달과 함께 전통적인 조사방법의 대안으로 주목을 받아왔다. 더구나 자원자표본을 이용한 웹 조사는 조사의 비용과 신속성 측면에서 웹 조사의 큰 장점으로 평가되고 있다. 그러나 자원자표본을 이용한 웹 조사의 실질적 활용에 있어서는 다양한 의견들이 쏟아져 나오고 있다. 크게는 조사통계의 활용목적과 조사환경에 따라 웹 조사의 활용실태는 달라질 수 있을 것이다. 선행연구들은 정확성과 신뢰성이 요구되는 조사일수록 자원자표본을 이용한 웹 조사 방법을 신중하게 검토할 것을 권장하고 있다. 이런 측면에서 본 연구는 통계청 사회조사자료에 자원자표본을 이용한 웹 조사의 방법의 활용성을 평가해 보았다.

### 1. 요약

일반적으로 표본조사에서 확률표본이 갖는 장점은 매우 많다. 우선 통계적 추론에 대한 이론적 근거를 제시할 수 있다. 특히 확률표본의 경우, 추정치에 대한 불편성이 유지되고, 모집단 추정치에 대한 신뢰성을 평가할 수 있다. 그런 측면에서 볼 때 자원자표본은 비확률표본이라는 점에서 모집단에 대한 대표성 및 신뢰성 등을 측정하기 어렵고 추정치의 불편성이 보장되지 않는다. 이에 대해 많은 연구자들은 웹 조사의 단점을 극복하기 위한 여러 가지 대안들을 제시하였다. 그 중 자주 사용되는 방법 중 하나는 자원자들의 웹 조사 참여성향을 조정하는 성향점수를 이용한 가중치 조정방법이다.

본 연구에서는 성향가중치 조정방법을 사회조사 자료에 적용하고, 성향조정 효과를 분석하였다. 그리고 이로부터 자원자표본을 이용한 웹 조사 방법의 활용 가능성을 모색하고자 하였다. 이를 위해 다양한 성향조정 방법을 적용하고 그 결과를 비교함으로써 상대적으로 좋은 성향조정 방법을 선택하고자 하였다. 선택 기준은 편향과 분산을 사용하였다. 평가는 준거조사 자료를 기준으로 사용했을 때, 성향조정 가중치 조정 전과 조

정 후의 추정치의 편향이 얼마나 감소하였는지 그리고 추정치의 분산은 얼마나 증가 또는 감소하였는지를 살펴보았다. 지금까지의 연구내용과 결과를 요약하면 다음과 같다.

(1) 성향점수 모형과 자료 전처리

- 성향점수 가중치 모형은 로지스틱 회귀모형을 사용하였다.
- 성향조정을 위한 변수는 사회조사 102개 문항 중 준거조사와 웹 조사 간에 통계적으로 차이가 있는 변수 17개를 최종 선택하였다.
- 성향점수 계산을 위해 무응답이 포함되어 있는 변수는 핫텍 무응답 대체를 실시하였다.

(2) 성향점수 가중치 조정을 위한 기본가중치 설정

- 립 가중치와 웹 조사 표본 추출률의 역수 ( $N_w/n_w$ )를 사용한 결과 웹 조사 표본 추출률의 역수를 사용한 경우의 추정치 편향과 분산이 더 작았다.

(3) 5개 층화방법과 성향점수 역수 방법을 이용한 가중치 적용

- 두 방법을 적용한 결과, 5개 층화방법에 의한 성향점수 가중치 적용 결과가 추정치의 편향과 분산이 상대적으로 더 작았다.

(4) (1)~(3)에 의해 본 연구에서는 기본가중치로 웹 표본 추출률의 역수를 사용한 5개 층화방법을 최종 성향점수 가중치 방법으로 선택하였다.

(5) 종합하면, 성향점수를 이용한 가중치 조정은 웹 조사의 준거조사에 대한 편향을 줄이는데 큰 효과가 있는 것으로 나타났다. 전체적으로 성향조정에 사용된 변수를 평가한 결과는 조정 전에 비해 편향이 확실히 감소하는 것을 알 수 있다. 그렇지만 그 편향이 완전히 제거되는 것은 아니라는 점에 주목할 필요가 있다. 게다가 편향의 크기는 대체로 감소하였지만 조정 후에도 여전히 웹 조사와 준거조사 추정치 간에는 차이가 있는 변수들이 많은 것을 알 수 있다. 또한 성향가중치를 적용한 웹 조사 추정치의 분산이 조정 전에 비해 전체적으로 증가하였다. 따라서 웹 조사 참여 성향가중치 조정은 준거조사에 대해 편향을 줄이는 효과는 있지만, 분산이 커지는 대가를 치뤄야 한다.

## 2. 추가분석

본 연구는 지금까지 언급한 핵심적인 분석결과 외에도 여러 가지 추가분석을 수행하였다. 특히 성향변수를 사용하는 것과 관련하여 다양한 분석을 수행하였다. 연구자들은 선택한 성향변수를 인구학적 변수와 비인구학적 변수로 나누어 추정치의 편향이 어떻게 달라지는지를 살펴보았다. 전체적으로 인구학적 특성 변수만 사용할 경우는 웹 조사의 성향조정에 크게 기여하지 못한 것으로 나타났다. 한편 비인구학적 변수만 사용하여 조정된 경우



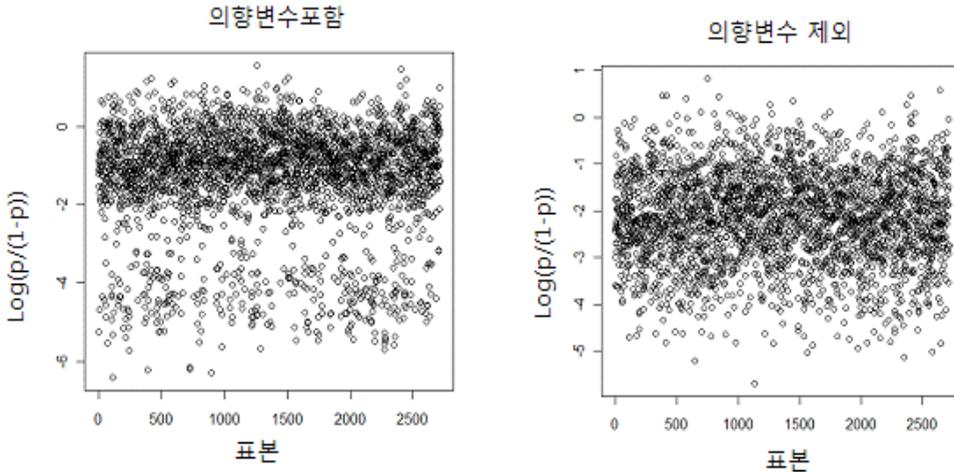
는 인구학적 변수만 사용한 경우에 비해 그 효과가 훨씬 큰 것으로 나타났다. 그렇지만 사회조사는 인구학적 변수와 비인구학적 변수를 모두 포함하고 있다. 특히 비인구학적 변수, 즉 사회의식과 관련한 변수가 많다는 것을 고려할 때 비인구학적 변수를 이용한 성향조정 의 효과가 큰 것은 당연한 결과라 할 수 있다. 따라서 대부분 사회의식관련 변수는 인구학적 특성과 독립적이지 않기 때문에 전체적으로 인구학적 변수와 비인구적 변수를 모두 사용한 경우가 웹 조사 추정치의 편향을 가장 줄이는 것을 알 수 있었다.

또한 본 연구는 최종적으로 사회조사 항목에 포함되어 있는 변수 중 17개 변수를 성향 조정변수로 사용하였다. 사회조사 항목 중에는 ‘인터넷 조사 참여의향’이라는 조사항목이 포함되어 있고, 이는 차기년도 조사에서 자원자표본으로서 웹 조사에 참여할 의향이 있는지를 준거조사와 웹 조사 모두에서 질문하는 항목이다. 연구자들은 이 항목이 준거조사인 대면조사와 웹 조사인 인터넷조사 간의 차이를 뚜렷하게 나타낼 것으로 예상하였고, 실제로 인터넷 참여의향 비율은 두 조사 간에 확실한 차이를 드러냈다. 준거조사 응답자 중에서 인터넷조사 참여의향이 있는 비율이 약 21%인 반면에 웹 조사에서 인터넷 참여의향이 있는 비율은 88%인 것으로 나타났다. 두 조사 간의 참여의향 비율이 거의 반대 양상을 보이는 것을 알 수 있다. 그리고 이 인터넷 참여의향 변수를 성향 조정변수들에 포함하여 성향조정된 결과 ‘참여의향이 있다’고 응답한 비율이 20.2%이고, 이 변수를 제외하고 조정 후 결과에서는 82.9%인 것으로 나타났다. 따라서 인터넷 참여의향 변수를 포함하여 그 변수를 평가할 경우 조정 결과는 거의 완벽하게 준거조사에 일치하는 것을 알 수 있다. 즉, 이 변수에 대해서만은 편향이 거의 완벽하게 줄어든다고 볼 수 있다. 그러나 이 인터넷 참여의향 변수를 포함할 경우, 다른 평가변수들에 대해서는 이 변수를 제외한 경우에 비해 오히려 준거조사와 차이가 발생하는 것으로 나타났다. 자세한 결과는 생략하기로 한다. 또한 조정 후 평가변수들에 대한 추정치의 분산이 상당히 커지는 것으로 나타났다.

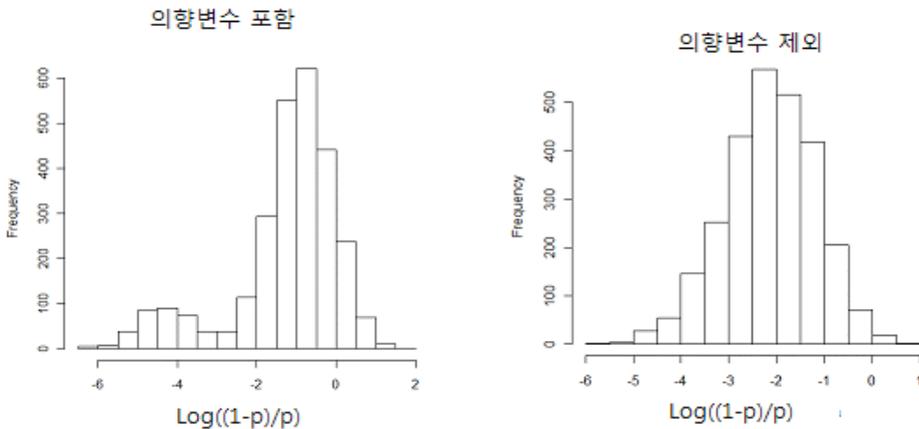
참여의향	준거조사	웹 조사		
		조정 전	의향변수 제외 후	의향변수 포함 후
있다	21.1 %	88.0 %	82.9 %	20.2
없다	78.9 %	12.0 %	17.1 %	79.8

[그림 5-11]과 [그림 5-12]는 웹 조사 성향조정을 위해 인터넷 참여의향 변수를 성향 변수로 포함시켰을 때와 그렇지 않았을 때에 대한 로지스틱 회귀분석을 실시한 결과를 각각 그림으로 나타낸 것이다. 그림에서 세로축은 표본이 웹 조사 대비 준거조사에 참여할 확률( $p/(1-p)$ )을 로그변환한 값이고 가로축은 표본을 나타낸다. 그림에서 보면 알 수 있듯이 인터넷 참여의향 변수는 준거조사와 웹 조사를 명확히 구분하고 있다. 전체자료

중에서 부그룹이 존재하는 것처럼 보인다. [그림 5-12]와 같이 이러한 웹 조사 참여의향 변수는 확률분포의 이중분포 형태를 하고 있으며 이는 실제로 성향가중치에서 극단값을 유도하는 결과를 초래하였다. 이렇게 가중치의 극단값이 존재할 경우, 변수에 따라서는 이 값에 의해 가중치 조정이 상당히 불안정할 수 있다. 이런 점을 고려하여 본 연구에서 웹 조사 참여의향 변수는 웹 조사와 준거조사를 뚜렷하게 구분하는 변수이긴 하지만 상대적으로 다른 변수들 조정에 불안정한 역할을 하기 때문에 성향 조정변수로 선택하지 않았다. 이 결과로부터 성향변수 선택이 얼마나 어렵고 신중해야 하는가를 다시 한 번 강조할 수 있을 것이다. 따라서 성향변수 선택과 관련한 연구가 별도로 진행될 필요가 있을 것으로 본다.



[그림 5-11] 웹 조사 참여의향 변수 포함 여부에 따른 웹 조사 참여확률 분포



[그림 5-12] 웹 조사 참여의향 변수 포함 여부에 따른 웹 조사 참여확률 히스토그램

### 3. 제언 및 향후 연구

자원자표본을 이용한 웹 조사는 관심 있는 사람은 누구나 조사를 완성할 수 있는 매우 간편한 조사방법임에는 틀림없다. 이러한 조사방법은 시장 여론조사 (marketing research)에서 빈번하게 사용되고 있다. 자원자표본 웹 조사가 모집단의 대표성, 선택편향 등의 문제에서 자유롭지 못한 것은 사실이다. 그렇지만 대표성에 관한 문제는 특히 우리나라와 같이 인터넷 보급률과 사용률이 높은 조사환경이라면 점차 해결될 수 있을 것으로 기대한다. 그러나 일반적으로 웹 조사에 참여하기 위해 자발적으로 자원한 사람들이 준 거조사에 참여한 사람들과 그 속성이 얼마나 유사할지에 대해서는 확신하기 어렵다. 이에 대한 확신이 없다면 비확률표본을 이용한 웹 조사 결과를 해석하는 데는 상당한 주의가 필요하게 될 것이다.

성향조정 가중치는 그 편향을 완전히 제거하지는 못하지만 어느 정도 줄이는 효과는 있다. 그렇지만 이에 대한 대가로 성향가중치를 적용한 추정치는 그 추정치의 분산이 커지는 효과가 발생한다. Lee (2009)에서도 성향가중치를 적용할 때 분산 증가에 대해 신중해야 한다는 것을 제안한 바 있다. 분산에 대한 문제가 없다 하더라도 현재 사회조사에 대한 성향가중치를 적용한 결과에 따르면, 성향모형에 의한 조정효과가 실질적으로 활용할 수 있을 만큼 충분하지 않다는 것이 문제이다. 특히 편향이 감소하기는 하지만, 얼마나 감소해야 좋을 것인지를 평가할 만한 객관적인 척도가 없다. 즉, 성향조정 효과를 측정할 만한 기준이 없다는 것이 문제이다. 본 연구에서는 성향조정 전과 후에 각각 준거조사에 대해 통계적으로 유의한 차이가 있는가를 검정함으로써 그 효과를 측정하였다. 그 결과 여전히 조정 후에도 준거조사와 웹 조사 추정치 간에는 차이가 있었다. 통계적 입장에 볼 때, 웹 조사 추정치가 준거조사에 충분히 근접해졌다고 보기 어렵다.

성향가중치 적용을 현실적으로 사용하는데 있어서 충분한 논의와 연구가 있어야 할 것으로 판단된다. 특히, 확률표본 기반의 정확성과 신뢰성이 요구되는 조사일수록 더욱 더 신중하게 판단해야 할 것이다. 왜냐하면 대면조사와 웹 조사를 혼합한 혼합조사방법을 사용할 경우, 웹 조사 추정치의 편향이 충분히 제거되지 않고 조정 후 추정치 분산이 증가하기 때문이다. 성향가중치 적용으로 웹 조사 결과가 준거조사에 동일하게 조정하기 어렵고 추정치의 신뢰성이 떨어질 수 있다는 우려가 있다.

본 연구에서 사용한 사회조사의 경우 매년 100여 개 이상의 많은 조사항목을 포함하고 있다. 게다가 100여 개의 항목은 모두 관심변수로서 핵심이 되는 변수가 없다는 것이 성향가중치 조정을 더욱 어렵게 만드는 요인이라 할 수 있다. 즉, 성향 조정변수에 따라서 나머지 평가항목의 조정효과가 얼마나 달라질 수 있느냐 하는 것이 문제이다. 성향가중치 모형이 갖는 장점은 사후층화방법이나 림 가중치 등의 방법과는 달리 여러 개의



성향변수를 동시에 사용할 수 있다는 것이다. 그러나 100여 개의 항목 중에서 전체 문항을 모두 효과적으로 조정할 수 있는 변수를 선택하는 것이 매우 어려운 작업이다. 조사 대상자가 15세 이상의 국민을 대상으로 하기 때문에 학생, 젊은 층, 성인층 등 다양한 응답층들을 동시에 포함한다는 것이 일반적인 성향조정을 어렵게 만드는 요인이다. 또 하나는 조건부 문항이 많다는 것이다. 즉, 1번 문항에서 1번에 응답한 사람만 2번을 응답하도록 하는 질문들이 많다. 이런 점을 극복하기 위해 다음과 같은 두 가지 방법을 생각해볼 수 있다.

첫째, 우선 다양한 층이 있는 경우에는 응답자를 몇 개의 층으로 나누어 성향 조정변수를 찾고, 각 층별로 성향조정 가중치를 적용하는 것이다. 사회조사는 의식조사인 만큼 가장 대별되는 층이 학생층과 성인층이라 할 수 있다. 이렇게 층을 구분하여 성향조정을 적용할 경우 전체를 하나의 층으로 간주하고 조정하는 경우에 비해 훨씬 조정효과가 있을 것으로 본다.

둘째, 사회조사는 매년 5개 분야에 대해 국민들의 의식관련 질문을 실시한다. 통계청(2009) 연구결과에서도 제안하였지만, 100여 개 문항을 동시에 조정하는 것보다 분야별로 성향 조정변수를 다르게 적용하는 방법을 적용할 수 있겠다. 실제로 본 연구에서 보여준 안전분야 평가 결과에 의하면 일부 항목(42-1번: 법을 지키지 않는 이유, 43번: 공공질서수준 6개 문항)에 대해서는 전혀 조정효과가 없었다. 이런 점을 극복하기 위해서 항목 특성을 명확히 파악하고 조정변수를 찾는 것이 중요할 것으로 보인다. 그렇지만 이렇게 할 경우 조사 분야 내 항목으로 성향 조정변수가 제한된다는 점과 한 조사 내에서 여러 개의 가중치를 적용해야 하는 부차적인 문제들이 발생할 수 있다.

지금까지 살펴본 연구결과와 선행연구의 결과에 근거해서 자원자표본을 이용한 웹 조사의 활용과 관련하여 실용적 측면과 학술적 측면으로 정리해 볼 수 있다. 자원자표본을 이용한 웹 조사 방법을 사회조사에 실질적 활용까지는 몇 가지 측면에서 제약이 있을 것으로 본다.

① 통계청 사회조사는 17,000여 가구에 대해 정확성과 신뢰성을 동시에 요구하는 조사이다. 뿐만 아니라 조사가 특정 항목에 초점이 맞춰진 것이 아닌 매우 폭넓은 분야에 걸쳐 조사되고 있다. ② 성향조정으로 인해 준거조사 대비 추정치의 편향은 어느 정도 줄일 수 있지만 그 효과를 평가할 만한 정확한 척도가 없다. 그런 면에서 볼 때 객관적 근거 하에서의 추가적 검증이 필요하지만 객관적 근거를 찾기가 쉽지 않다. ③ 성향조정 후 분산의 증가로 인해 추정치의 신뢰성이 떨어질 수 있다. 가중치 사용은 분산이 커지는 것이 일반적이지만 좋은 가중치를 사용할 수 있다면 오히려 분산을 줄일 수도 있다. 따라서 사회조사 통계의 신뢰성 증가를 위해서는 더 작은 분산을 위해 연구가 추가적으로 있어야 할 것이다. ④ 자원자표본을 이용한 웹 조사가 실용성 측면에서 매우 매력적

인 방법임에는 틀림없다. 소규모의 조사, 또는 특정 주요 항목이 몇 개로 제한되어 있는 그런 조사들에서는 그 사용을 고려해볼 만하다. 특히, 조사결과가 정확성 또는 신뢰성에 비해 저렴한 비용의 신속성이 더 중요한 요소라면 웹 조사는 성향가중치 조정을 통해 충분한 활용가치가 있을 것으로 기대한다. ⑤ 종합적으로 사회조사에서 자원자표본을 이용한 웹 조사 방법의 적용은 시기상조가 아닌가 싶다. 많은 조사항목과 넓은 조사범위 그리고 조사목적이 특정 항목에 국한되지 않는다는 점을 고려할 때 사회조사에 성향점수 가중치 적용은 매우 어려울 것으로 보인다.

본 연구는 학술적 측면에서는 여러 가지 의의가 있다.

학술 연구자들에게 있어서 자원자표본을 이용한 조사방법에 관한 연구는 많은 제약이 있기 마련이다. 이는 자원자표본을 이용한 웹 조사 자료를 얻기 어렵기 때문이다. 특히 준거조사가 없다는 것이 연구를 어렵게 한다. 그런 측면에서 보면 본 연구에 사용된 자료는 연구자들에게 있어서 매우 중요한 의미를 갖는다. 일반 여론조사 회사에서 자원자표본을 이용한 웹 조사를 자주 활용하고 있지만 이에 대한 준거조사는 거의 없다. 따라서 이러한 웹 조사 추정치의 정확성과 신뢰성을 상대적으로 비교할 만한 기준이 없다. 이러한 자료를 생산한다는 것만으로도 통계청은 학술 연구자들에게 큰 기여를 한다고 볼 수 있다. 학술적 측면에서 성향조정 모형을 이용한 연구는 매우 폭넓게 진행되어야 할 것이다. 성향가중치 조정은 조정 전에 비해 편향을 줄이는 효과는 분명하다. 이런 점에서 보다 정교한 성향가중치 생성을 위한 이론적 연구가 필요할 것이다. 그리고 사회조사의 경우와 같이 응답대상자가 성인 또는 학생 그룹으로 뚜렷하게 구분되는 형태의 부차 모집단(sub population)이 있는 자료에 대해서는 이를 고려하여 가중치를 적용할 수 있도록 추가연구가 있어야 할 것이다.

또한 성향조정에 있어서 가장 중요한 관건은 성향 조정변수의 사용이다. 실제로 웹 참여자의 성향을 조정할 수 있는 근본적인 항목을 조사에 포함시킴으로써 조정효과를 높일 수 있는 방법을 찾아야 한다. 그런데 본 연구에서 향후 웹 조사 참여의향과 같이 준거조사와 웹 조사를 확실하게 구분하는 변수는 의외로 역효과를 내는 경우도 있다. 이 변수는 웹 조사 참여자의 성향을 반영한다고 보기 어려운데, 현재 웹에 참여하고 있는 사람들은 당연히 다음 번에도 웹에 참여할 것이라고 기대할 수 있기 때문이다. 따라서 형식적으로 면접조사와 웹 조사 참여자의 성향을 드러내는 변수처럼 보이지만 이런 변수는 오히려 성향가중치의 극단값을 생성하게 되어 다른 평가변수들의 분산을 증가시키는 요인이 된다. 따라서 성향변수 선택과 관련해서 가중치 작성의 이론적 측면에 관한 연구가 필요해 보인다. 또한 사회조사는 2년에 걸쳐 10개 부문의 의식관련 조사가 실시되기 때문에 매년 조사로부터 찾은 유의한 성향변수를 시기를 교차하여 적용해 봄으로써 더 좋은 성향변수를 찾는 노력이 있어야 할 것이다.

이처럼 자원자표본을 정확성과 신뢰성을 요구하는 국가통계에 적용하기까지는 많은 시간과 연구 노력이 필요할 것으로 판단된다. 그렇지만 통계청 (2009)과 본 연구를 토대로 볼 때, 특정 응답층을 고려한 조사에 대해서는 성향조정 가중치 방법을 충분히 활용할 수 있을 것으로 보인다. 특히 자원자표본이 아닌 확률표본 기반의 웹 조사에 대해서는 방법선택에 따른 효과를 조정하는 차원에서 성향조정 가중치의 활용가치는 높을 것으로 기대한다. 성향가중치는 많은 변수를 자유롭게 사용하여 성향모형을 구축할 수 있다는 점에서 무응답 가중치 조정, 혼합조사방법에서 방법효과 조정 등에서 유용하게 활용될 수 있다.

자원자표본을 이용한 웹 조사는 비확률표본에 근거한다는 점에서 통계적 추론에 대해 약점이 있다. 이런 상황에서 응답률 또는 무응답 오차를 계산하는 것은 좋은 방법이 아니다. 확률적 기반 통계를 작성하는 이유가 바로 조사통계의 추정이 객관적 이론에 근거할 수 있다는 데 있다고 본다. 비확률표본에 기반한 웹 조사의 실질적 활용을 위한 핵심은 웹 조사에서 일반 모집단에 대한 대표성 있는 표본을 선택하는 문제, 인터넷에 접속하지 않는 사람들에 대한 대응 방법, 그리고 웹 조사의 무응답 편향을 최소화할 수 있는 방법을 찾는 문제 등이 될 것이다. 이러한 문제들은 웹 조사 추정치에 좋지 않은 영향을 미칠 수 있고, 넓고 방대한 모집단에 초점을 둔 웹 조사일수록 그런 문제들이 심각할 수 있다.

지금까지의 결과를 정리해 보자. 대규모 조사의 경우, 웹 조사 방법이 전통적 조사방법 (대면조사, 전화조사, 우편조사 등)에 비해 덜 비싸다는 이유들 때문에, 웹 조사는 하나의 매력적인 조사방법으로 등장하였다. 웹 조사비용은 대안적 방법이 무엇이나에 따라, 조사진행 속도에 따라, 조사장치의 정도에 따라 상대적일 수 있다. 그러나 웹 조사가 실현가능하고 신뢰할 만한 자료 수집 방법이라는 것을 제시할 만큼 조정 효과에 대한 검토는 완료되지 않았다. 특히 국가 공식통계에서처럼 대표성이 자료 수집에 필요한 핵심 요소라면, 자원자를 이용한 웹 조사는 조사방법의 역사적 견지에서나 이론적으로나 경험적으로 설명할 수 없는 품질의 자료를 제공할 수 있다. 인터넷을 이용한 조사 방법은 분명히 어려운 조사환경에 대한 좋은 대안일 수 있지만, 추정치의 정확성과 신뢰성 측면에서 보면 이것은 21세기 조사연구를 힘들게 하는 모든 요인들에 대한 만병통치약은 아님을 주지해야 할 것이다. 그만큼 국가통계에 있어서 비확률표본 기반 웹 조사 방법의 실제 적용은 충분한 시간과 여유를 가지고 접근해야 할 문제이다.



## 참고문헌

- 조성검, 허명희 (2009). <인터넷 자원자표본에 대한 추정방법 연구>, 통계청 용역보고서.
- Battaglia, M.P., Malec, D.J., Spencer, B., Hoaglin, D.C., and Sedransk, J. (1995). Adjusting for Noncoverage of Nontelephone Households in the National Immunization Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association.
- Brick, J.M., Waksberg, J., and Keeter, S. (1996). Using Data on Interruptions in Telephone Service as Coverage Adjustments. *Survey Methodology*, 22(2), 185-197.
- Cochran, W.G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies, *Biometrics*, 24, 205-213.
- Couper, M.P. (2000). *Web Surveys: A Review of Issues and Approaches*. Public Opinion
- Couper, M.P. (2008). *Designing effective web surveys*, Cambridge, UK: Cambridge University Press.
- Couper, M.P., Kapteyn, A., Schonlau, M., and Winter, J. (2007). Noncoverage and Nonresponse in an Internet Survey. *Social Science Research*, 36, 131-148.
- Czajka, J.L., Hirabayashi, S.M., Little, R.J.A., and Rubin, D.B. (1992). Projecting from Advance Data Using Propensity Modeling: An Application to Income and Tax Statistics. *Journal of Business and Economic Statistics*, 10(2), 117-132.
- D'Agostino, R.B. Jr. (1998). Propensity Score Methods for Bias Reduction for the Comparison of a Treatment to a Non-randomized Control Group. *Statistics in Medicine*, 17, 2265-2281.
- Danielsson, S. (2002). The propensity score and estimation in nonrandom surveys -An Overview, Available at <http://www.statistics.su.se/modernsurveys/publ/11.pdf>
- Deville, Sarndal (1992). Calibration estimators in survey sampling. *JASA*, 87, 376-82.
- Deming, W.E., Stephan, F.F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known, *Annals of Mathematical Statistics*, 11, 427-444.
- Dillman, D.A., Smyth, J.D. and Christian, L.M. (2008). *Internet, Mail and Mixed-Mode Surveys, The Tailored Design Method*, Hoboken, NJ: John Wiley & Sons.
- Duncan, K.B., and Stasny, E.A. (2001). Using Propensity Scores to Control Coverage Bias in Telephone Surveys. *Survey Methodology*, 27(2). 121-130.
- Fricker, R.D., and Schonlau, M. (2002). Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods*, 14(4): 347-365.
- Garren, S.T., and Chang, T.C. (2002). Improved Ratio Estimation in Telephone Surveys Adjusting for Noncoverage. *Survey Methodology*, 28(1), 63-76.

- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley and Sons.
- Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Harris Interactive (2008). Election Results Further Validate Efficacy of Harris Interactive's Online Methodology. Press Release from Harris Interactive, November 6, 2008.
- Hoaglin, D.C., and Battaglia, M.P. (1996). A Comparison of Two Methods of Adjusting for Noncoverage of Nontelephone Households in a Telephone Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Iannacchione, V.G. (2003). Sequential Weight Adjustments for the Location and Cooperation Propensity for the 1995 National Survey of Family Growth. *Journal of Official Statistics*, 19(1), 31-43.
- Iannacchione, V.G., Milne, J.G., Folsom, R.E. (1991). Response probability weight adjustments using logistic regression, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 637-642.
- Ireland, C.T., Kullback, S. (1968). Contingency tables with given marginals, *Biometrika*, 55, 179-188.
- Jelke Bethlehem (2010). Selection bias in web surveys, *International Statistical Review*, 78(2), 161-188.
- Kalton, G., and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- Kim, J. J., Li, J., and Valliant, R. (2007). Cell collapsing in poststratification, *Survey Methodology* 33, 139-150.
- Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors, *Survey Methodology*, 32, 133-142.
- Lee, S. (2004). *Statistical Estimation Methods in Volunteer Panel Web Surveys*, Unpublished Doctoral Dissertation. University of Maryland, Joint Program in Survey Methodology.
- Lee, S. (2006). Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys. *Journal of Official Statistics*, 22(2), 329-349.
- Lee, S., and Valliant, R. (2009). Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment. *Sociological Methods & Research*, 37, 319-343.
- Lee, S., and Valliant, R. (2007). Weighting Telephone Samples Using Propensity Scores. In: Lepkowski, J.M., Tucker C., Brick J.M., et al., eds. *Advances in Telephone Survey Methodology*. New York, NY: Wiley.
- Lepkowski, J., Kalton, G., and Kasprzyk, D. (1989). Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel. *Proceedings of the Section on Survey Research Methods*,



- American Statistical Association, 296-301.
- Little, R.J.A. (1993). Post-Stratification: A Modeler's Perspective. *Journal of the American Statistical Association*, 88, 1001-1012.
- Little, R.J.A., and Mei-Miau Wu. (1991). Models for Contingency Tables with Known Margins When Target and Sampled Populations Differ, *Journal of the American Statistical Association*, 86, 87-95.
- Little, R.J.A., and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, 31, 161-168.
- Loosveledt, G., Sonck, N. (2008). An evaluation of the weighting procedures for online access panel survey, *Survey Research Methods*, 2(2), 93-105.
- 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. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55.
- Schillewaert, N., and Meulemeester, P. (2005). Comparing Response Distributions of Offline and Online Data Collection Methods. *International Journal of Market Research*, 47, 163-178.
- Schonlau, M., van Soest, A., and Kapteyn, A. (2007). Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates From Web Surveys Using Propensity Scoring?, *Journal of the American Statistical Association*, 79, 516-524.
- Schonlau, M., Zapert, K., Simon L.P., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H., Turner, R., and Berry, S. (2004). A Comparison Between a Propensity Weighted Web Survey and an Identical RDD Survey. *Social Science Computer Review*, 22, 128-138.
- Schonlau, M., van Soest, A., Kapteyn, A., and Couper, M. (2009). Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods & Research*, 37(3), 291-318.
- Smith, P.J., Rao, J.N.K., Battaglia, M.P., Daniels, D., and Ezzati-Rice, T. (2001). Compensating for Provider Nonresponse Using Propensities to Form Adjustment Cells: The National Immunization Survey. *Vital and Health Statistics, Series 2, No. 133*, DHHS Publication No. (PHS) 2001-1333.
- Stephanie Steinmetz, Kea Tijdens (2010). Volunteer web surveys and propensity score adjustment the Wageindicator example, *The second International Workshop on Internet Survey Methods*, Statistical Center of Statistics Korea, Daejeon, South Korea, September 08-09, 2010.
- Taylor, H. (2000). Does Internet Research Work? Comparing online survey result with Telephone survey, *International Journal of Market Research*, 42, 58-63.

- Taylor, H., Bremer, J., Overmeyer, C., Siegel, J.W. and Terhanian, G. (2001). The Record of Internet-Based Opinion Polls in Predicting the Results of 72 Races in the November 2000 US Elections. *International Journal of Market Research*, 43(2), 127-135.
- Terhanian, G., and Bremer, J. (2000). Confronting the Selection-Bias and Learning Effects of Problems Associated with Internet Research. Research Paper: Harris Interactive.

## 〈부 록〉

〈부표 5-1〉 기본가중치 사용에 따른 성향조정 효과 비교

실문항	준거조사	웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 차이		
		성향조정 진	표본 추출률	rim 가중치	성향조정 진	표본 추출률	rim 가중치	성향조 정 진	표본 추출률	rim 가중치
연령대	20대 이하	24.23	19.96	31.21	0.88	0.91	1.40	-4.53	4.27	-6.99
	30대	20.00	21.41	23.34	0.95	1.12	1.26	-7.33	-1.41	-3.34
	40대	20.69	16.45	21.28	0.68	0.93	1.26	4.06	4.24	-0.59
	50대	16.49	21.29	16.84	0.77	1.29	1.12	-1.32	-4.80	-0.35
	60대 이상	18.60	9.47	7.32	0.62	1.50	0.65	9.13	-2.30	11.28
성별	남	49.19	47.38	47.86	1.00	1.47	1.50	1.80	1.33	-0.10
	여	50.81	52.62	50.71	1.00	1.47	1.50	-1.80	-1.33	0.10
교육정도	무학	4.04	0.56	1.43	0.16	0.51	0.17	3.48	2.61	3.55
	초등학교	9.72	4.20	3.56	0.42	1.08	0.46	5.52	0.66	6.16
	중학교	9.91	5.22	4.24	0.46	1.05	0.59	4.69	0.50	5.68
	고등학교	35.55	26.38	27.99	0.88	1.47	1.38	9.17	-3.95	7.56
	전문대	15.02	22.83	23.49	0.86	0.78	1.31	-7.81	0.44	-8.47
	4년대	22.18	34.53	34.18	0.96	1.01	1.42	-12.35	0.65	-12.00
	석사	2.79	5.14	4.86	0.45	0.46	0.56	-2.35	-1.02	-2.07
	박사	0.79	1.14	1.20	0.21	0.14	0.36	-0.35	0.11	-0.41
	미혼	27.02	35.44	37.98	0.95	1.02	1.47	-8.42	3.00	-10.96
	유배우	62.38	59.26	57.58	0.98	1.36	1.49	3.12	-3.57	4.80
혼인상태	사별	7.07	3.37	2.70	0.38	1.09	0.42	3.69	-0.64	4.36
	이혼	3.53	1.92	1.74	0.27	0.46	0.41	1.61	1.22	1.79
운전여부	예	44.99	45.07	45.54	1.01	1.48	1.51	-0.09	-3.20	-0.56
	아니오	55.01	54.93	54.46	1.01	1.48	1.51	0.09	3.20	0.56
경제활동 여부	예	55.33	45.81	45.97	1.00	1.46	1.50	9.52	2.62	9.36
	아니오	44.67	54.19	54.03	1.00	1.46	1.50	-9.52	-2.62	-9.36

〈부표 5-1〉 기본가중치 사용에 따른 영향조정 효과 비교(계속)

설문문항	준거 조사			웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 차이			
	성향조정 전	표본 추출률	rim 가중치	성향조정 전	표본 추출률	rim 가중치	성향조정 전	표본 추출률	rim 가중치	성향조정 전	표본 추출률	rim 가중치	
가구소득	15.62	7.98	8.01	0.96	1.27	1.3	6.83	7.64	7.61				
	23.92	24.18	23.51	1.41	2.09	2.24	1.7	-0.26	0.41				
	24.08	26.24	24	1.56	2.15	2.06	-2.63	-2.16	0.08				
	15.36	18.79	20.3	1.36	1.92	2.29	-3.8	-3.43	-4.94				
	8.43	10.73	12.32	1.04	1.36	1.84	-2.5	-2.29	-3.88				
	5.65	6.37	5.45	0.79	1.62	1.03	-0.07	-0.73	0.19				
	2.76	2.11	2.82	0.53	0.53	0.81	0.1	0.65	-0.06				
자살충동경험	4.18	3.6	3.58	0.67	0.78	0.73	0.36	0.58	0.6				
	7.69	9.19	15.7	0.76	0.55	1.06	-9.29	-1.5	-8.01				
	92.31	90.81	84.3	0.76	0.55	1.06	9.29	1.5	8.01				
	0.49	0.63	0.57	0.18	0.17	0.15	-0.33	-0.14	-0.08				
	5.28	5.08	6.26	0.48	0.62	0.72	-0.98	0.2	-0.98				
	42.17	40.27	27.22	0.91	1.54	1.29	14.41	1.9	14.96				
	39.19	36.21	44.15	1	1.32	1.51	-4.21	2.97	-4.96				
학교주변식품 안전도	12.87	17.81	21.8	0.84	0.96	1.27	-8.89	-4.94	-8.93				
	11.94	13.9	16.04	0.73	0.92	1.13	-3.71	-1.96	-4.1				
	52.89	49.85	53.64	1.01	1.47	1.51	-1.27	3.05	-0.75				
	32.29	33.36	27.88	0.91	1.45	1.35	4.19	-1.07	4.4				
	2.78	2.67	2.32	0.28	0.52	0.56	0.86	0.11	0.47				
	0.09	0.22	0.12	0.08	0.12	0.06	-0.07	-0.12	-0.03				
	0.51	0.94	1.32	0.24	0.18	0.25	-1.06	-0.42	-0.81				
자기준법수준	7.66	8.55	13.43	0.68	0.63	1.01	-5.76	-0.9	-5.77				
	28.6	30.58	32.47	0.93	1.35	1.45	-1.91	-1.97	-3.86				
	46.06	41.07	36.33	0.97	1.47	1.44	10.04	4.99	9.73				
	17.17	18.48	16.46	0.79	1.14	1.06	-1.31	-1.7	0.71				
	범죄위험인식	~100만원	7.98	8.01	0.96	1.27	1.3	6.83	7.64	7.61			
		100~200만원	24.18	23.51	1.41	2.09	2.24	1.7	-0.26	0.41			
		200~300만원	26.24	24	1.56	2.15	2.06	-2.63	-2.16	0.08			
300~400만원		18.79	20.3	1.36	1.92	2.29	-3.8	-3.43	-4.94				
400~500만원		10.73	12.32	1.04	1.36	1.84	-2.5	-2.29	-3.88				
500~600만원		6.37	5.45	0.79	1.62	1.03	-0.07	-0.73	0.19				
600~700만원		2.11	2.82	0.53	0.53	0.81	0.1	0.65	-0.06				

〈부표 5-1〉 기본가중치 사용에 따른 영향조정 효과 비교(계속)

실문항	준거조사 추정치			웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 차이			
	성향조정 전	표본 추출률	rim 가중치	성향조정 전	표본 추출률	rim 가중치	성향조정 전	표본 추출률	rim 가중치	성향조정 전	표본 추출률	rim 가중치	
부모부양 책임	스스로 해결	12.68	12.66	10.13	9.44	12.66	10.13	0.59	1.15	0.99	3.23	0.02	2.55
	가족	36.05	31.89	22.54	23.02	31.89	22.54	0.84	1.46	1.23	13.03	4.16	13.51
	가족, 정부사회	47.39	52.22	63.16	63.2	52.22	63.16	0.97	1.49	1.45	-15.82	-4.83	-15.78
복지환경 체감	정부사회	3.85	3.02	3.8	4.11	3.02	3.8	0.4	0.35	0.5	-0.25	0.83	0.06
	기타	0.03	0.22	0.37	0.23	0.22	0.37	0.11	0.12	0.26	-0.19	-0.18	-0.34
	매우 좋아짐	2.85	4.34	3.6	4.34	3.64	3.6	0.4	0.47	0.47	-1.48	-0.79	-0.75
환경보호 운동참여	약간 좋아짐	16.95	22.18	23.02	22.18	16.82	23.02	0.84	0.97	1.29	-5.23	0.13	-6.07
	변함없음	70.01	60.57	60.75	60.57	67.1	60.75	0.99	1.32	1.47	9.44	2.91	9.26
	약간 나빠짐	8.58	10.73	10.71	10.73	10.71	10.71	0.63	0.93	0.93	-2.15	-2.13	-2.13
환경보호 운동참여	매우 나빠짐	1.61	2.18	1.73	2.18	1.73	1.92	0.3	0.32	0.36	-0.58	-0.12	-0.31
	매우 노력	6.61	12.81	9.01	12.81	9.01	12.72	0.67	0.71	1.01	-6.2	-2.4	-6.11
	약간 노력	23.63	33.33	24.02	33.33	24.02	32.99	0.96	1.14	1.41	-9.71	-0.39	-9.36
식품안전 인식	별로 노력안함	42.59	38.74	45.79	38.74	45.79	38.49	0.98	1.48	1.46	3.85	-3.21	4.1
	전혀 노력안함	27.17	15.12	21.18	15.12	21.18	15.8	0.71	1.27	1.12	12.06	6	11.37
	매우 안전	1.2	1.3	0.88	1.3	1.25	0.88	0.22	0.32	0.19	-0.1	-0.05	0.32
진반적 스트레스 정도	약간 안전	13.88	11.8	10.73	11.8	10.73	11.45	0.64	0.85	0.91	2.08	3.15	2.43
	보통	47.62	42.03	50.05	42.03	50.05	41.5	1	1.47	1.48	5.59	-2.44	6.12
	약간 불안	32.15	38.89	34.02	38.89	34.02	40.06	0.99	1.3	1.49	-6.74	-1.88	-7.91
생활 만족도	매우 불안	5.15	5.98	3.95	5.98	3.95	6.11	0.47	0.39	0.74	-0.83	1.21	-0.96
	매우 많이 느낌	11.41	16.71	12.88	16.71	12.88	17.35	0.75	0.82	1.17	-5.3	-1.47	-5.94
	느끼는 편	58.63	65.48	60.31	65.48	60.31	65.28	0.96	1.47	1.44	-6.85	-1.67	-6.64
생활 만족도	느끼지 않는 편	27.3	16.77	25.09	16.77	25.09	16.5	0.76	1.43	1.12	10.53	2.21	10.8
	전혀 느끼지않음	2.65	1.04	0.88	1.04	0.88	0.88	0.2	0.47	0.2	1.61	0.94	1.78
	매우 만족	7.15	8.29	7.01	8.29	7.01	6.93	0.56	0.66	0.64	-1.14	0.14	0.22
생활 만족도	약간 만족	23.62	29.12	25.5	29.12	25.5	29.06	0.92	1.32	1.35	-5.5	-1.88	-5.44
	보통	45.7	35.57	43.04	35.57	43.04	36.43	0.97	1.48	1.48	10.13	2.66	9.27
	약간 불만족	18.53	20.37	18.91	20.37	18.91	20.59	0.81	1.15	1.21	-1.84	-0.38	-2.07
매우 불만족	5	6.65	5.54	6.65	5.54	6.98	0.5	0.61	0.8	-1.64	-0.54	-1.98	

〈부표 5-2〉 성향조정 방법 사용에 따른 성향조정 효과 비교

설문문항	준거조사 추정치			웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 간 차이		
	성향조정 전	5개 층화	성향점수역수	성향조정 전	5개 층화	성향점수역수	성향조정 전	5개 층화	성향점수역수	성향조정 전	5개 층화	성향점수역수
성별	남	47.38	47.86	48.8	1	1.47	1.64	1.8	1.33	0.39		
	여	52.62	52.14	51.2	1	1.47	1.64	-1.8	-1.33	-0.39		
연령대	20대 이하	28.76	19.96	18.36	0.88	0.91	0.96	-4.53	4.27	5.87		
	30대	27.33	21.41	20.18	0.95	1.12	1.17	-7.33	-1.41	-0.18		
	40대	16.63	16.45	15.94	0.68	0.93	0.97	4.06	4.24	4.75		
	50대	17.81	21.29	20.99	0.77	1.29	1.32	-1.32	-4.8	-4.5		
	60대 이상	9.47	20.9	24.53	0.62	1.5	1.81	9.13	-2.3	-5.93		
	무학	0.56	1.43	1.68	0.16	0.51	0.59	3.48	2.61	2.36		
교육 정도	초등학교	4.2	9.05	9.94	0.42	1.08	1.22	5.52	0.66	-0.22		
	중학교	5.22	9.41	10.38	0.46	1.05	1.25	4.69	0.5	-0.47		
	고등학교	26.38	39.5	41.98	0.88	1.47	1.67	9.17	-3.95	-6.44		
	전문대	22.83	14.58	12.85	0.86	0.78	0.76	-7.81	0.44	2.17		
	4년대	34.53	21.53	19.09	0.96	1.01	0.97	-12.35	0.65	3.08		
	석사	5.14	3.82	3.45	0.45	0.46	0.43	-2.35	-1.02	-0.65		
경제활동 여부	박사	1.14	0.68	0.63	0.21	0.14	0.14	-0.35	0.11	0.16		
	예	45.81	52.71	54.62	1	1.46	1.62	9.52	2.62	0.71		
	아니오	54.19	47.29	45.38	1	1.46	1.62	-9.52	-2.62	-0.71		
	미혼	35.44	24.02	21.9	0.95	1.02	1.05	-8.42	3	5.12		
	유배우	59.26	65.95	67.84	0.98	1.36	1.45	3.12	-3.57	-5.46		
	사별	3.37	7.71	8.04	0.38	1.09	1.15	3.69	-0.64	-0.98		
이혼	1.92	2.32	2.22	0.27	0.46	0.45	1.61	1.22	1.31			

〈부표 5-2〉 성향조정 방법 사용에 따른 성향조정 효과 비교(계속)

설문문항	준거조사 추정치			웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 간 차이		
	성향조정 전	5개 층화	성향점수	성향조정 전	5개 층화	성향점수	성향조정 전	5개 층화	성향점수	성향조정 전	5개 층화	성향점수
지살충동 경험	있다	16.98	9.19	7.34	0.76	0.55	0.76	0.55	0.51	-9.29	-1.50	0.35
	없다	83.02	90.81	92.66	0.76	0.55	0.76	0.55	0.51	9.29	1.50	-0.35
학교 주변 식품 안전도	매우 안전	0.82	0.63	0.50	0.18	0.17	0.18	0.17	0.16	-0.33	-0.14	-0.01
	안전	6.26	5.08	4.96	0.48	0.62	0.48	0.62	0.72	-0.98	0.20	0.32
	보통	27.77	40.27	42.71	0.91	1.54	0.91	1.54	1.72	14.41	1.90	-0.54
	불안	43.40	36.21	35.07	1.00	1.32	1.00	1.32	1.44	-4.21	2.97	4.12
운전 여부	매우 불안	21.76	17.81	16.77	0.84	0.96	0.84	0.96	1.01	-8.89	-4.94	-3.89
	예	45.07	48.19	48.51	1.01	1.48	1.01	1.48	1.65	-0.09	-3.20	-3.52
자기 준법 수준	아니오	54.93	51.81	51.49	1.01	1.48	1.01	1.48	1.65	0.09	3.20	3.52
	아주 잘지킴	15.65	13.90	13.55	0.73	0.92	0.73	0.92	1.02	-3.71	-1.96	-1.60
	비교적 잘지킴	54.17	49.85	48.19	1.01	1.47	1.01	1.47	1.64	-1.27	3.05	4.71
	보통	28.10	33.36	35.07	0.91	1.45	0.91	1.45	1.68	4.19	-1.07	-2.78
범죄 위험 인식	비교적 지킴 지키지않음	1.92	2.67	3.00	0.28	0.52	0.28	0.52	0.68	0.86	0.11	-0.22
	전혀 지키지않음	0.16	0.22	0.19	0.08	0.12	0.08	0.12	0.11	-0.07	-0.12	-0.10
	매우 안전	1.57	0.94	0.72	0.24	0.18	0.24	0.18	0.15	-1.06	-0.42	-0.21
	비교적 안전	13.42	8.55	7.37	0.68	0.63	0.68	0.63	0.7	-5.76	-0.90	0.29
	보통	30.52	30.58	30.34	0.93	1.35	0.93	1.35	1.46	-1.91	-1.97	-1.74
	비교적 불안	36.02	41.07	42.68	0.97	1.47	0.97	1.47	1.66	10.04	4.99	3.37
	매우 불안	18.48	18.87	18.88	0.79	1.14	0.79	1.14	1.28	-1.31	-1.7	-1.72

〈부표 5-2〉 성향조정 방법 사용에 따른 성향조정 효과 비교(계속)

설문문항	준거조사 추정치		웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 간 차이		
	성향조정 전	5개 층화	성향점수	성향점수	성향점수	성향조정 전	5개 층화	성향점수	성향조정 전	5개 층화	성향점수
환경 보호 운동 참여	매우 노력	12.81	9.01	8.30	0.67	0.71	0.75	-6.20	-2.40	-1.69	
	약간 노력	33.33	24.02	22.14	0.96	1.14	1.19	-9.71	-0.39	1.49	
	별로 노력안함	38.74	45.79	47.02	0.98	1.48	1.66	3.85	-3.21	-4.44	
	전혀 노력안함	15.12	21.18	22.54	0.71	1.27	1.48	12.06	6.00	4.64	
식품 안전 인식	매우 안전	1.30	1.25	1.17	0.22	0.32	0.39	-0.10	-0.05	0.03	
	약간 안전	11.80	10.73	10.77	0.64	0.85	0.99	2.08	3.15	3.11	
	보통	42.03	50.05	51.76	1.00	1.47	1.63	5.59	-2.44	-4.15	
	약간 불안	38.89	34.02	32.58	0.99	1.30	1.41	-6.74	-1.88	-0.43	
전반적 스트레스 정도	매우 불안	5.98	3.95	3.71	0.47	0.39	0.40	-0.83	1.21	1.44	
	매우 많이 느낌	16.71	12.88	11.71	0.75	0.82	0.83	-5.30	-1.47	-0.30	
	느끼는 편	65.48	60.31	59.45	0.96	1.47	1.64	-6.85	-1.67	-0.81	
	느끼지 않는 편	16.77	25.09	26.91	0.76	1.43	1.58	10.53	2.21	0.40	
생활 만족도	전혀 느끼지않음	1.04	1.72	1.94	0.20	0.47	0.56	1.61	0.94	0.71	
	매우 만족	8.29	7.01	6.55	0.56	0.66	0.71	-1.14	0.14	0.60	
	약간 만족	29.12	25.50	24.77	0.92	1.32	1.40	-5.50	-1.88	-1.15	
	보통	35.57	43.04	45.65	0.97	1.48	1.67	10.13	2.66	0.05	
매우 불만족	약간 불만족	20.37	18.91	18.09	0.81	1.15	1.19	-1.84	-0.38	0.44	
	매우 불만족	6.65	5.54	4.95	0.50	0.61	0.60	-1.64	-0.54	0.06	

〈부표 5-2〉 성향조정 방법 사용에 따른 성향조정 효과 비교(계속)

실문문항	준거조사 추정치		웹 조사 추정치			표준오차			준거조사와 웹 조사 추정치 간 차이		
	성향조정 전	5개 층화	성향점수	성향점수	성향조정 전	5개 층화	성향점수	성향조정 전	5개 층화	성향점수	
가구소득	~100만원	7.98	8.03	8.03	0.96	1.27	1.41	6.83	7.64	7.59	
	100~200만원	24.18	24.74	24.74	1.41	2.09	2.40	1.70	-0.26	-0.82	
	200~300만원	26.24	25.85	25.85	1.56	2.15	2.30	-2.63	-2.16	-1.77	
	300~400만원	18.79	19.23	19.23	1.36	1.92	2.14	-3.80	-3.43	-3.87	
	400~500만원	10.73	10.55	10.55	1.04	1.36	1.43	-2.50	-2.29	-2.12	
	500~600만원	6.37	6.00	6.00	0.79	1.62	1.62	-0.07	-0.73	-0.35	
	600~700만원	2.11	2.07	2.07	0.53	0.53	0.57	0.10	0.65	0.69	
스스로 해결 가족, 정부사회	700만원 이상	3.60	3.53	3.53	0.67	0.78	0.79	0.36	0.58	0.65	
	가족	12.66	12.97	12.97	0.59	1.15	1.23	3.23	0.02	-0.30	
	가족	31.89	34.07	34.07	0.84	1.46	1.70	13.03	4.16	1.98	
	정부사회	52.22	49.97	49.97	0.97	1.49	1.65	-15.82	-4.83	-2.58	
	정부사회	3.02	2.81	2.81	0.40	0.35	0.35	-0.25	0.83	1.05	
	기타	0.23	0.18	0.18	0.11	0.12	0.10	-0.19	-0.18	-0.15	
	매우 좋아짐	4.34	3.60	3.60	0.40	0.47	0.55	-1.48	-0.79	-0.75	
녹지환경 체감	약간 좋아짐	22.18	16.82	16.08	0.84	0.97	1.11	-5.23	0.13	0.87	
	변함없음	60.57	67.10	67.97	0.99	1.32	1.49	9.44	2.91	2.04	
	약간 나빠짐	10.73	10.71	10.51	0.63	0.93	1.03	-2.15	-2.13	-1.93	
	매우 나빠짐	2.18	1.73	1.84	0.30	0.32	0.39	-0.58	-0.12	-0.23	

〈부표 5-3〉 안전부문 항목 성향조정 결과

안전부문		ref	web				Std Err		effect_size	
			성향 조정전	5개층화		성향 조정전	5개 층화	성향 조정전	5개 층화	
사회 분야별 안전에 대한 인식도 문항(1~10)										
1. 국가 안보	매우 안전	1.00	2.51	***	1.73	**	0.30	0.29	0.07	0.03
	비교적 안전	13.93	21.74	***	17.41		0.83	0.99	0.09	0.04
	보통	32.59	30.59	***	35.11		0.93	1.47	0.02	0.02
	비교적 불안	43.28	36.42	***	36.60	***	0.97	1.41	0.05	0.05
	매우 불안	9.21	8.74	***	9.15	***	0.58	0.82	0.01	0.00
2. 자연 재해	매우 안전	1.02	2.28	***	2.08	***	0.30	0.40	0.05	0.04
	비교적 안전	17.74	25.82	***	21.53		0.87	1.09	0.09	0.04
	보통	45.43	39.23	***	45.57	*	0.99	1.49	0.05	0.00
	비교적 불안	31.62	27.96	***	26.97	***	0.90	1.27	0.03	0.04
	매우 불안	4.19	4.71	***	3.85	***	0.46	0.47	0.01	0.01
3. 건축물 및 시설물	매우 안전	1.40	2.83	***	2.40	*	0.33	0.39	0.05	0.04
	비교적 안전	22.04	26.51	**	24.49		0.88	1.20	0.04	0.02
	보통	54.63	44.48	***	49.51	***	1.00	1.47	0.07	0.04
	비교적 불안	19.75	21.73	***	20.23	**	0.83	1.13	0.02	0.01
	매우 불안	2.18	4.45	***	3.37	***	0.46	0.42	0.07	0.03
4. 교통 사고	매우 안전	0.31	1.28	***	0.88	**	0.22	0.25	0.07	0.04
	비교적 안전	7.66	11.06	**	9.13		0.63	0.75	0.05	0.02
	보통	40.71	33.19	***	36.63	***	0.95	1.45	0.06	0.03
	비교적 불안	41.37	40.52	***	41.14	***	0.99	1.45	0.01	0.00
	매우 불안	9.95	13.95	***	12.22	***	0.71	0.89	0.06	0.03
5. 화재 (산불 포함)	매우 안전	1.06	2.73	***	2.30	***	0.32	0.40	0.07	0.05
	비교적 안전	15.54	21.10	**	19.15		0.81	1.08	0.06	0.04
	보통	55.07	48.72	***	53.71	**	1.01	1.45	0.05	0.01
	비교적 불안	25.16	23.19	***	21.51	***	0.86	1.14	0.02	0.04
	매우 불안	3.17	4.26	***	3.34	***	0.44	0.47	0.03	0.00
6. 식량 안보	매우 안전	4.45	5.61	**	4.99		0.45	0.57	0.02	0.01
	비교적 안전	30.42	30.53	***	27.43	***	0.93	1.22	0.00	0.03
	보통	45.77	38.53	***	44.53	*	0.98	1.50	0.05	0.01
	비교적 불안	16.94	21.16	**	19.50		0.83	1.09	0.05	0.03
	매우 불안	2.43	4.17	***	3.56	***	0.42	0.51	0.05	0.03
7. 정보 보안	매우 안전	0.87	2.18	***	1.60	*	0.29	0.37	0.06	0.03
	비교적 안전	12.59	15.59	**	13.64		0.72	0.92	0.04	0.01
	보통	42.57	32.95	***	42.20	**	0.95	1.49	0.07	0.00
	비교적 불안	34.18	31.08	***	27.94	***	0.93	1.30	0.03	0.05
	매우 불안	9.80	18.20	***	14.61	***	0.80	0.88	0.12	0.07



<부표 5-3> 안전부문 항목 성향조정 결과(계속)

안전부문	ref	web				Std Err		effect size	
		성향 조정전		5개 층화		성향 조정전	5개 층화	성향 조정전	5개 층화
8. 신종 전염병	매우 안전	0.43	1.71 ***	1.17 ***		0.26	0.25	0.08	0.05
	비교적 안전	9.79	17.17 ***	14.71 ***		0.74	0.94	0.10	0.07
	보통	35.67	35.82 ***	38.41 **		0.97	1.45	0.00	0.02
	비교적 불안	43.37	33.52 ***	34.26 ***		0.95	1.41	0.08	0.07
	매우 불안	10.75	11.79 ***	11.46 ***		0.69	0.94	0.01	0.01
9. 범죄 위험	매우 안전	0.51	1.57 ***	0.94 **		0.24	0.18	0.07	0.03
	비교적 안전	7.66	13.42 ***	8.55		0.68	0.63	0.09	0.01
	보통	28.60	30.52 ***	30.58		0.93	1.35	0.02	0.02
	비교적 불안	46.06	36.02 ***	41.07 ***		0.97	1.47	0.08	0.04
	매우 불안	17.17	18.48 ***	18.87 ***		0.79	1.14	0.01	0.02
10. 전반적 사회 안전	매우 안전	0.56	1.53 ***	0.84		0.24	0.15	0.06	0.02
	비교적 안전	10.74	17.69 **	13.54		0.75	0.90	0.09	0.04
	보통	50.22	42.84 ***	47.30 ***		1.00	1.48	0.05	0.02
	비교적 불안	33.77	31.18 ***	32.45 **		0.93	1.39	0.02	0.01
	매우 불안	4.71	6.76 ***	5.87 ***		0.53	0.58	0.04	0.02
사회 불안 요인 -1순위	국가안보	28.78	30.09 **	29.06 **		0.91	1.28	0.01	0.00
	자연재해	5.10	4.16 **	4.08		0.41	0.58	0.02	0.02
	환경오염	6.05	7.61 ***	7.58 ***		0.53	0.68	0.03	0.03
	인재	5.58	4.93 **	4.69		0.45	0.59	0.01	0.02
	경제적 위험	15.36	17.69 ***	18.78 ***		0.77	1.24	0.03	0.04
	자원고갈	0.82	0.36 **	0.32 *		0.12	0.12	0.02	0.03
	도덕성 부족	6.96	6.94 **	6.00		0.50	0.63	0.00	0.02
	신종질병	5.74	2.59 ***	3.03 *		0.34	0.53	0.06	0.05
	범죄발생	21.12	20.74 **	22.86 ***		0.83	1.32	0.00	0.02
	계층갈등 기타	4.32 0.17	4.70 ** 0.18 ***	3.54 0.07 ***		0.42 0.10	0.41 0.04	0.01 0.00	0.02 0.01
사회 불안 요인 -2순위	국가안보	10.37	10.64 **	11.48		0.64	1.00	0.00	0.02
	자연재해	8.58	10.30 **	9.29		0.60	0.74	0.03	0.01
	환경오염	8.79	9.48 **	7.71		0.58	0.63	0.01	0.02
	인재	7.80	9.31 **	11.00 ***		0.59	0.98	0.02	0.05
	경제적 위험	14.14	16.14 **	16.95 *		0.74	1.10	0.02	0.04
	자원고갈	1.99	2.31 **	2.57		0.31	0.69	0.01	0.02
	도덕성 부족	10.10	10.94 **	10.52		0.65	0.87	0.01	0.01
	신종질병	11.10	7.95 **	8.42		0.57	0.99	0.04	0.04
	범죄발생	19.41	15.98 **	16.00		0.75	1.02	0.04	0.04
	계층갈등 기타	7.64 0.07	6.91 ** 0.04 ***	6.06 0.01 ***		0.51 0.04	0.64 0.01	0.01 0.01	0.03 0.01

〈부표 5-3〉 안전부문 항목 성향조정 결과(계속)

안전부문	ref	web		Std Err		effect size		
		성향 조정전	5개층화	성향 조정전	5개 층화	성향 조정전	5개 층화	
사회 불안 요인 -3순위	국가안보	11.04	11.41 **	12.00	0.67	1.06	0.01	0.02
	자연재해	6.40	7.83 **	8.93 **	0.55	0.93	0.03	0.06
	환경오염	10.71	12.49 **	10.59	0.69	0.79	0.03	0.00
	인재	8.04	8.70 **	8.38	0.59	0.76	0.01	0.01
	경제적 위험	11.06	13.08 **	13.74 *	0.70	1.14	0.03	0.05
	자원고갈	2.68	2.33 **	2.60	0.31	0.49	0.01	0.00
	도덕성 부족	10.56	11.29 **	11.16	0.64	0.97	0.01	0.01
	신종질병	10.65	8.46 **	8.97	0.56	0.82	0.04	0.03
	범죄발생	16.79	14.26 **	15.00	0.72	1.05	0.04	0.03
	계층갈등	11.98	10.09 **	8.57 **	0.63	0.77	0.03	0.06
기타	0.11	0.06 ***	0.07 ***	0.04	0.05	0.01	0.01	
현재 사회 안전	매우 안전	1.08	1.81 **	1.18	0.26	0.23	0.03	0.00
	약간 안전	15.18	14.84 **	12.61 **	0.72	0.91	0.00	0.03
	변화없다	37.17	31.53 ***	36.22	0.93	1.49	0.04	0.01
	약간 위험	39.23	40.77 **	40.76	0.99	1.42	0.01	0.01
	매우 위험	7.34	11.04 ***	9.23 ***	0.64	0.71	0.06	0.03
미래 사회 안전 전망	매우 안전	2.60	2.41 **	1.58 **	0.30	0.30	0.01	0.03
	약간 안전	22.71	22.45 **	20.07	0.84	1.10	0.00	0.03
	변화없다	44.27	34.81 ***	39.46	0.96	1.48	0.07	0.04
	약간 위험	24.71	32.03 ***	32.29 ***	0.94	1.37	0.07	0.08
	매우 위험	5.71	8.30 ***	6.61 ***	0.56	0.58	0.05	0.02
수입 식품 안전	매우 안전	0.63	1.65 ***	1.11 *	0.25	0.21	0.05	0.03
	약간 안전	6.58	8.34 **	7.99	0.54	0.79	0.03	0.02
	변화없다	34.12	28.41 ***	32.30 **	0.91	1.43	0.05	0.02
	약간 위험	44.63	44.63 ***	44.30 *	1.00	1.46	0.00	0.00
	매우 위험	14.03	16.97 ***	14.30 ***	0.75	0.87	0.04	0.00
수입 식품 불안 이유	정부 규제미흡	43.19	52.52 **	51.45	1.29	1.85	0.09	0.09
	식품업체 의식부족	26.51	22.33 **	22.43 *	1.08	1.53	0.05	0.05
	언론 불안감보도	10.50	5.80 ***	5.98 ***	0.59	0.83	0.09	0.09
	과거불안전	6.88	9.18 **	9.43 *	0.78	1.18	0.05	0.06
	식품사용경험	12.66	9.67 **	10.16 *	0.76	1.05	0.05	0.04
	수입국에 대한 불신	0.27	0.48 ***	0.55 ***	0.16	0.21	0.02	0.03
학교 급식 안전	매우 안전	3.02	2.61 **	2.20	0.31	0.36	0.01	0.02
	약간 안전	19.15	18.16 **	17.31	0.76	1.06	0.01	0.02
	보통	51.28	43.88 ***	50.62	1.00	1.47	0.05	0.01
	약간 불안	22.94	29.31 ***	25.84 **	0.92	1.21	0.06	0.03
	매우 불안	3.62	6.03 ***	4.03 ***	0.48	0.41	0.05	0.01



〈부표 5-3〉 안전부문 항목 성향조정 결과(계속)

안전부문	ref	web		Std Err		effect size		
		성향 조정전	5개 증화	성향 조정전	5개 증화	성향 조정전	5개 증화	
급식 불안 이유	정부 관리미흡	24.26	28.86 **	28.31	1.55	2.09	0.08	0.08
	관리자 의식부족	45.68	40.06 **	41.24	1.68	2.36	0.08	0.07
	언론 불안감보도	12.89	9.54 **	11.35	0.99	1.64	0.08	0.04
	실제경험	13.22	18.13 ***	15.14	1.29	1.46	0.12	0.05
	교육, 홍보부족	3.22	3.01 **	3.39	0.63	0.74	0.01	0.01
	기타	0.74	0.40 ***	0.57 ***	0.21	0.29	0.03	0.02
학교 주변 식품 안전	매우 안전	0.49	0.82 **	0.63	0.18	0.17	0.02	0.01
	약간 안전	5.28	6.26 **	5.08	0.48	0.62	0.02	0.00
	보통	42.17	27.77 ***	40.27	0.91	1.54	0.11	0.02
	약간 불안	39.19	43.40 **	36.21 *	1.00	1.32	0.03	0.02
	매우 불안	12.87	21.76 ***	17.81 ***	0.84	0.96	0.11	0.06
학교 주변 식품 불안 이유	정부 관리미흡	13.29	17.39 **	16.84 *	0.97	1.19	0.07	0.06
	판매자 의식부족	33.06	30.72 **	30.49	1.16	1.58	0.03	0.03
	언론 불안감보도	6.32	4.35 **	4.85	0.51	0.74	0.05	0.04
	실제경험	16.50	23.80 ***	22.32 **	1.07	1.33	0.11	0.09
	의식교육 부족	8.55	5.59 ***	5.69 **	0.55	0.69	0.06	0.06
	선입견	22.15	17.98 **	19.68	0.96	1.37	0.06	0.04
	기타	0.13	0.18 ***	0.13 ***	0.09	0.07	0.01	0.00
전반적 식품 안전	매우 안전	1.20	1.30 **	1.25	0.22	0.32	0.00	0.00
	약간 안전	13.88	11.80 **	10.73 *	0.64	0.85	0.03	0.04
	보통	47.62	42.03 **	50.05	1.00	1.47	0.04	0.02
	약간 불안	32.15	38.89 **	34.02	0.99	1.30	0.06	0.02
	매우 불안	5.15	5.98 ***	3.95 ***	0.47	0.39	0.02	0.02
전반적 식품 안전 불안 이유	정부 관리미흡	32.63	37.91 **	37.74	1.46	1.97	0.07	0.07
	판매자 의식부족	46.62	42.86 **	42.76	1.49	1.97	0.05	0.05
	언론 불안감보도	12.44	8.03 ***	7.73 ***	0.80	0.96	0.09	0.10
	실제경험	8.03	10.74 **	11.37 *	0.93	1.36	0.07	0.09
	기타	0.28	0.47 ***	0.40 ***	0.18	0.18	0.03	0.02
집중적 개선 요구 분야	식품안전기준	31.30	32.77 **	33.46	0.94	1.39	0.01	0.02
	법규위반자 처벌	39.30	41.26 **	40.62	1.00	1.45	0.02	0.01
	영업자재정, 기술지원	14.74	11.89 ***	12.39	0.65	1.02	0.03	0.03
	영업자 책임의식제고	10.80	10.51 **	10.08	0.63	0.83	0.00	0.01
	소비자	3.69	3.21 **	3.20	0.35	0.47	0.01	0.01
	안전식품선택교육, 홍보							
	기타	0.17	0.36 ***	0.24 ***	0.11	0.09	0.02	0.01

〈부표 5-3〉 안전부문 항목 성향조정 결과(계속)

안전부문	ref	web					Std Err		effect size	
		성향 조정전		5개층화		성향 조정전	5개층화	성향 조정전	5개층화	
타인의 준법 수준	아주 잘지킴	1.96	4.36 ***	4.63 ***	0.41	0.63	0.07	0.08		
	비교적 잘지킴	29.23	29.16 **	26.58 **	0.92	1.27	0.00	0.02		
	보통	47.30	45.23 ***	48.54	1.01	1.48	0.02	0.01		
	비교적 안지킴	20.61	20.21 **	19.27 **	0.80	1.05	0.00	0.01		
	전혀 안지킴	0.91	1.04 ***	0.98 ***	0.20	0.30	0.01	0.00		
자신의 준법 수준	아주 잘지킴	11.94	15.65 **	13.90	0.73	0.92	0.05	0.03		
	비교적 잘지킴	52.89	54.17 **	49.85	1.01	1.47	0.01	0.02		
	보통	32.29	28.10 **	33.36	0.91	1.45	0.04	0.01		
	비교적 안지킴	2.78	1.92 **	2.67	0.28	0.52	0.02	0.00		
	전혀 안지킴	0.09	0.16 ***	0.22 ***	0.08	0.12	0.01	0.02		
평소 법을 지키지 않는 이유	지키는게 손해	13.60	15.91 ***	22.24 ***	5.04	9.18	0.14	0.59		
	처벌규정 미약	8.03	8.98 ***	10.05 ***	3.94	5.91	0.09	0.20		
	다른사람도 지키지않아서	18.21	21.53 ***	16.23 ***	5.57	5.24	0.21	0.12		
	귀찮아서	44.11	43.30 ***	39.34 ***	7.27	9.24	0.04	0.23		
	단속하지 않아서	9.63	7.09 ***	4.76	3.51	2.90	0.25	0.47		
	준법교육이 미약	4.15	3.19 ***	7.38 ***	2.25	5.59	0.13	0.44		
	기타	2.27	100 ***	100 ***			13.78	17.21		

각 분야별 우리나라 국민 공공질서 수준(1~6)

1. 차레 지키기	아주 잘지킴	11.32	17.72 ***	16.09 **	0.76	1.00	0.08	0.06
	비교적 잘지킴	48.26	44.92 ***	42.58 ***	1.00	1.44	0.03	0.04
	보통	30.38	25.96 ***	29.85 *	0.90	1.44	0.04	0.00
	비교적 안지킴	9.13	9.83 **	10.20	0.60	0.91	0.01	0.02
	전혀 안지킴	0.91	1.56 ***	1.29 ***	0.25	0.26	0.03	0.02
2. 거리 환경 질서	아주 잘지킴	6.81	14.22 ***	12.09 ***	0.70	0.80	0.12	0.09
	비교적 잘지킴	21.89	23.28 **	21.34 *	0.85	1.16	0.01	0.01
	보통	32.31	25.49 ***	28.67 ***	0.89	1.45	0.06	0.03
	비교적 안지킴	31.04	27.65 ***	29.05 **	0.90	1.33	0.03	0.02
	전혀 안지킴	7.96	9.36 ***	8.85 ***	0.58	0.77	0.02	0.01
3. 보행 질서	아주 잘지킴	6.28	9.95 ***	8.56 **	0.59	0.70	0.06	0.04
	비교적 잘지킴	32.40	35.36 **	32.62	0.97	1.32	0.03	0.00
	보통	41.44	34.86 ***	39.38 *	0.96	1.50	0.05	0.02
	비교적 안지킴	18.29	17.79 **	17.67	0.77	1.11	0.01	0.01
	전혀 안지킴	1.58	2.05 ***	1.78 ***	0.29	0.31	0.02	0.01



〈부표 5-3〉 안전부문 항목 성향조정 결과(계속)

안전부문	ref	web					Std Err		effect size	
		성향 조정전			5개 층화		성향 조정전	5개 층화	성향 조정전	5개 층화
4. 운전자 교통 질서	아주 잘지킴	4.75	10.13	***	8.79	***	0.60	0.70	0.10	0.08
	비교적 잘지킴	22.81	25.08	**	22.69	*	0.88	1.17	0.02	0.00
	보통	45.31	36.13	***	38.85	***	0.97	1.46	0.07	0.05
	비교적 안지킴	24.57	25.31	***	26.75		0.87	1.37	0.01	0.02
	전혀 안지킴	2.56	3.35	***	2.92	***	0.36	0.43	0.02	0.01
5. 공공 장소 금연	아주 잘지킴	9.66	20.49	***	17.90	***	0.81	1.00	0.15	0.12
	비교적 잘지킴	25.40	21.74	***	20.68	***	0.82	1.15	0.03	0.05
	보통	34.96	26.54	***	31.83	***	0.91	1.50	0.07	0.03
	비교적 안지킴	23.69	21.88	***	20.98	***	0.83	1.13	0.02	0.03
	전혀 안지킴	6.29	9.35	***	8.61	***	0.59	0.77	0.05	0.04
6. 공공 장소 정숙	아주 잘지킴	5.79	12.21	***	10.01	***	0.65	0.74	0.11	0.08
	비교적 잘지킴	17.47	21.43	**	19.53		0.83	1.09	0.04	0.02
	보통	35.32	27.28	***	31.00	***	0.91	1.46	0.07	0.04
	비교적 안지킴	32.43	28.50	***	30.82	**	0.91	1.38	0.03	0.01
	전혀 안지킴	9.00	10.57	***	8.65	***	0.61	0.70	0.02	0.01
교통 사고 원인	운전자 부주의	26.46	25.24	**	25.62		0.87	1.35	0.01	0.01
	보행자 부주의	3.18	3.68	**	2.72		0.39	0.34	0.01	0.01
	교통혼잡	13.62	12.99	**	13.78		0.67	1.02	0.01	0.00
	도로구조 잘못	2.95	3.27	**	2.88		0.38	0.42	0.01	0.00
	신호체계 잘못	2.16	2.05	**	1.98		0.28	0.35	0.00	0.01
	질서의식 부족	51.46	52.58	**	52.91		1.01	1.47	0.01	0.01
	기타	0.17	0.18	***	0.11	***	0.08	0.05	0.00	0.01
교통 안전 시설 만족	매우 만족	4.94	4.61	**	4.26		0.41	0.54	0.01	0.01
	약간 만족	24.32	27.53	**	25.98		0.89	1.25	0.03	0.02
	보통	53.03	48.20	***	52.44		1.01	1.46	0.04	0.00
	약간 불만족	15.36	16.77	**	15.19		0.75	0.95	0.02	0.00
	매우 불만족	2.35	2.89	***	2.12	***	0.35	0.33	0.02	0.01
운전 여부	예	44.99	45.07	**	48.19	*	1.01	1.48	0.00	0.03
	아니오	55.01	54.93	***	51.81	***	1.01	1.48	0.00	0.02
야간 보행	두려움	40.59	52.85	***	46.72	***	1.01	1.46	0.10	0.05
	두렵지않음	59.41	47.15	***	53.28	***	1.01	1.46	0.09	0.05
야간 보행 위험 원인	가로등이 없어서	28.81	26.35	**	26.99		1.23	1.92	0.03	0.03
	우범지역이므로	7.57	7.43	**	7.08		0.73	0.97	0.00	0.01
	인적이 들몰어서	62.37	64.67	**	64.60		1.33	1.98	0.03	0.03
	기타	1.25	1.55	***	1.33	***	0.33	0.33	0.02	0.00

〈부표 5-3〉 안전부문 항목 성향조정 결과(계속)

안전부문	ref	web				Std Err		effect size	
		성향 조정전	**	5개 층화	*	성향 조정전	5개 층화	성향 조정전	5개 층화
무섭지만 그냥다님	56.22	47.11	**	48.03		1.38	1.99	0.10	0.09
조금 멀더라도	14.59	26.20	***	24.66	***	1.22	1.60	0.21	0.19
야간 피해다님									
보행시 집앞까지 택시이용	5.95	3.89	**	3.50	*	0.51	0.60	0.06	0.07
행동 아는사람과 동행	6.88	7.19	**	7.31		0.69	0.95	0.01	0.01
다음기회로 미룸	14.86	14.15	**	15.13		0.98	1.45	0.01	0.01
호신도구준비	0.37	0.34	**	0.15	*	0.14	0.06	0.00	0.02
기타	1.14	1.12	***	1.23	***	0.28	0.43	0.00	0.01

\*\*\* p&lt;0.001, \*\* p&lt;0.05, \* p&lt;0.1

