

연구노트

생활시간 조사를 이용한 가중치 부여방법:

인터넷 조사에 대한 적용 가능성 검토*

Using Time Use Data for Weighting Internet Survey Results

조성겸** · 박아현*** · 허명희****

Sung Kyum Cho · Ah Hyun Park · Myung-Hoe Huh

인터넷 조사에 가중치를 적용하여 그 일반성을 높이고자 하는 시도가 다양하게 이루어졌지만, 안정적으로 적용될 수 있는 방법이 아직까지 개발되지는 않았고 이것이 인터넷 조사의 활용을 제약하고 있다. 인터넷 조사를 위한 가중치 개발에는 준거조사 결과가 필요하지만, 센서스나 정부기관에서 제공되는 통계 중 일부 인구학적 속성을 제외하면 이러한 준거조사 자료를 구하기 어렵다는 것이 중요 요인이다.

본 연구는 생활시간 조사를 이용해서 가중치를 개발하여 적용할 때, 인터넷 조사의 일반성을 높일 수 있는지를 검토해 보았다. 생활시간 조사는 정부기관에 의해 정기적으로 조사될 뿐만 아니라 센서스와는 달리 라이프스타일에 관련된 다양한 내용을 조사하기 때문이다. 2009년에 실시된 통계청 사회조사의 온라인 버전을 2004년 생활시간 조사결과를 이용해 가중치를 적용한 결과, 2009년에 동일한 설문지를 이용해 실시된 면접원방문 사회조사결과에 보다 일치하는 방향으로 작용하는 것으로 나타났다. 즉 생활시간 조사에서 나타난 취침 시간, 귀가시간, 인터넷 이용시간 등을 이용해서 가중치를 적용했을 때 단지 인구학적 속성만을 이용해서 가중치를 적용하는 경우보다, 온라인 조사결과와 오프라인 조사결과가 보다 근접해지는 경향이 있다는 것이다. 이러한 본 연구의 결과는 생활시간 조사자료가 인터넷 조사의 가중치 개발에 활용될 수 있다는 점을 보여준다.

* 이 논문은 한국연구재단 2010 한국사회 기반연구사업(SSK)의 연구비 지원(NRF-2010-330-B00280 "미디어 융합시대 새로운 사회조사 방법의 개발")과 통계청의 '인터넷 자원자 표본에 대한 추정방법 연구' 지원을 받아 이루어졌으며, 통계청 용역보고서로 발간된 내용을 보완하여 작성되었다.

** 교신저자(corresponding author): 충남대학교 언론정보학과 교수 조성겸.
E-mail: skcho99@gmail.com

*** 서울대학교 언론정보학과 박사과정 수료

**** 고려대학교 정경대학 통계학과 교수

주제어: 인터넷 조사, 가중치, 생활시간 조사, 자원자 표본

There are limitations as to how internet surveys can be used. Applying various weighting procedures has not always resulted in error reduction. A good reference survey would increase the effects of the weighting method, but it is very difficult to get a reference survey for non-demographic weighting variables, which restricts the use of internet surveys.

We hypothesized that time use variables could be employed as weighting variables. The time use survey is conducted regularly by KOSTAT and includes various time-related variables. We tested our hypothesis using the 2009 survey results, which had been gathered by KOSTAT in 2009. When we applied weighting variables which were based on the 2004 time use survey results to the online version of the 2009 KOSTAT social survey, the gap between the online and off-line versions was slightly reduced. This result shows that we could use time use survey results to develop a better weighting scheme.

key words: time use survey, internet survey, voluntary sample, weighting

I. 연구의 배경과 목적

온라인 사용이 보편화되면서, 온라인을 통한 설문조사가 각광을 받고 있다. 기존의 면대면 조사나 전화조사 혹은 설문지를 통한 조사들에 비해 인터넷을 통한 온라인 조사는 비용이 저렴하며, 도달률을 높일 수 있어 매우 매력적인 대안으로 떠오르고 있다. 구체적으로 온라인 조사는 면대면 조사나 전화조사에 비해 비용이 저렴하다는 최대 장점이 있으며, 기존의 면대면 조사나 전화조사가 지니고 있던 조사대상자들에 대한 접근의 어려움을 극복하는 데도 탁월한 강점을 지닌다. 실제로 가구를 대상으로 하는 면접조사의 경우는 일과 시간 중에 가정에 머무르는 가구원이 없는 경우가 대부분이어서 한밤중이나 주말 등을 이용한 조사가 이루어져야 한다거나 유선전화의 응답률이 떨어지는 등 조사대상자들에 대한 접근 자체가 어려운데 반해 인터넷을 통한 온라인 설문조사의 경우 이를 극복할 수 있는 대안으로 여겨지고 있다. 하지만 인터넷 조사의 경우 인터넷 사용자

들로 조사 대상자들이 국한되기 때문에 인터넷 조사가 기존의 면대면 조사나 전화조사의 단점을 대체할 만한 대안이 되기 위해서는 인터넷 조사의 대상자 표본이 기존의 일반 조사 대상자를 대표할 수 있기 위한 보정 작업이 필요하게 된다.

인터넷 조사의 대표성을 높이기 위한 방안에는 여러 가지가 있지만 그 중에서 가중치를 부여함으로써 그 대표성을 높이고자 하는 방안이 국내·외에서 다양하게 연구되어 왔다 (김영원·배예영 2010; 허명희·조성겸 2010; Schonlau et al. 2004; Duffy et al. 2005; Schonlau et al. 2009). 이들 연구들은 그 결과가 제 각기 다르게 나타나기 때문에 가중법 사용의 효과에 대해 일반적으로 단언하기는 힘들다. 예컨대 Duffy et al. (2005)의 연구에서는 가중치가 일부 항목에서는 면대면 조사와 인터넷 조사와의 차이를 줄였으나 이러한 가중법의 효과가 모든 항목에서 나타나지는 않았다. Schonlau et al. (2009)의 연구에서도 가중치가 인터넷 조사의 표본편향을 제거하는 데에는 충분하지 않았다. 그렇지만 이러한 연구들은 가중치 산출에 사용된 변인들이 제한되었다는 한계가 있다. 허명희·조성겸(2010) 그리고 김영원·배예영(2010)의 연구에서는 가중치 산정에 인구통계학적 변인 외에 다른 추가적인 변인들을 포함하는 경우, 가중치 적용의 효과가 보다 긍정적인 것으로 나타났다. 즉 중요한 것은 가중법 그 자체라기보다는 어떤 변인을 가중치 산정에 포함하느냐의 문제라고 볼 수 있다.

따라서 이 연구는 기존에 논의되었던 인구통계학적인 요인 외에 인터넷 조사의 편향을 줄일 수 있는 다른 요인을 찾아보고자 하는 데서 시작되었다. 실제로 허명희·조성겸(2010)은 2009년 통계청의 온라인 사회조사자료(응답자수 2,903명)을 대상으로 광역시 및 도, 지역 크기(동·읍·면), 성별, 연령대, 교육정도, 소비생활 만족도, 소득분배 견해, 일반 신문 보기, 인터넷 신문 보기와 같은 인구통계학적 변인과 라이프스타일 변인을 섞어 9개의 변인을 조정 기준 변수로 사용하여 같은 해에 면대면으로 실시된 통계청 사회조사 자료(응답자수 37,049명)에 성향 조정 가중치를 부여하여 두 데이터를 비교·분석하는 연구를 진행하였다. 이 결과 106개의 평가 문항 가운데 80% 정도의 문항에서 인구통계학적 변수에 의한 기본 가중치 보정에 비하여 향상된 추정치를 제공하는 것으로 나타났다.

이 연구는 기존의 허명희·조성겸(2010)의 연구에서 나아가, 인구통계학적 변인과 시간 사용 변인을 조정 기준 변수로 사용하여 가중치를 부여하는 것이 면대면 조사와 인터넷 조사 사이의 차이를 보다 잘 보정할 수 있을 것이라는 가정에서 출발하였다.

II. 인구통계학적 변인 가중치 부여의 한계

가중법은 특정한 특성이 일치되도록 새로운 설문조사의 결과를 조정하는 것이다. 따라서 모집단에 대한 정확한 정보를 연구자가 가지고 있어야 하며, 이러한 정보를 얻을 수 있는 방법은 다른 설문조사나 인구총조사 등이 된다. 실제로 이러한 준거조사(reference survey) 자료가 제대로 존재하지 않을 경우 가중치를 부여하는 방법은 의미가 없게 된다. 따라서 이러한 준거조사는 기존에 수행된 조사들과 별도로 이루어진다. 인터넷 설문조사에서 편향을 줄이기 위해 가중치를 사용하는 경우, 대개 인터넷이 아닌 다른 방법을 사용하곤 한다. 예를 들어, 준거조사로 전화조사 방법이 쓰였을 경우, 이 전화조사 방법은 무선 표집 방법으로 완벽하지는 않다 해도 상대적으로 보다 정확한 결과를 얻을 수 있는 것으로 인정되어 왔다. 현재까지 전화조사가 인터넷 조사보다 더 신뢰할 수 있다고 보았으며, 전화조사 결과나 다른 종류의 설문조사 결과가 인터넷 설문조사의 준거조사로 사용되는 경우가 대부분이었다.

하지만, 이렇게 준거조사를 수행하는 것은 비용적인 면이나 방법적인 면에서 결코 쉬운 일은 아니다. 인구총조사나 다른 정부기관에서 운영하는 설문조사를 사용할 경우 비용 면에서는 그렇게 비싼 것은 아니나, 이런 공식적인 조사에서 다뤄지는 항목이 한정적이기 때문에 이들 조사에서 다뤄지지 않은 요인을 변인으로 삼아 특정 설문조사에서 가중치를 부여하기는 어렵게 된다.

따라서 가중법을 사용하기 위해서는 신뢰할 수 있는 준거조사를 확보하는 것이 매우 중요하다. 이를 위해 온라인 패널을 통한 설문조사를 실시한 미국의 해리스 인터랙티브(Harris Interactive)는 기존 인터넷 설문조사에서 물었던 항목을 포함한 전화조사를 통해 신뢰할 만한 준거 자료를 확보하였고, 이를 바탕으로 인터넷 조사에서 성향 점수 가중법을 사용한 선구자로 인식되고 있다. 실제로 해리스 인터랙티브가 준거조사로 실시한 전화조사는 인터넷 설문조사에서 물었던 항목을 포함하고 있어서 모집단을 비교할 수 있으며, 인터넷 조사 모집단을 전화조사 모집단과 같은 특성을 지니도록 가중치를 부여할 수 있었다. 따라서 전화조사가 해리스 인터랙티브의 많은 인터넷 조사의 준거조사로 사용될 수 있었던 것이다. 그러나 이처럼 준거조사 자료로 활용하기 위한 추가 조사를 실시하는 것 자체가 추가 비용 등의 이유로 쉽지 않다.

준거조사로 사용하기 위해 선택된 설문조사는 가중법을 적용하는 데 있어 매우 결정적인 역할을 한다. 만약 준거조사의 데이터가 정확하지 않다면, 이를 준거조사로 사용한 설문조사 결과 역시 부정확한 결과를 낳을 수밖에 없게 된다. 심지어 인구총조사 데이터가 준거조사로 사용된 경우라도 이 역시 부정확성에 대한 우려를 떨쳐버릴 수는 없다. 이는 그만큼 모집단에 대해 완벽하게 정확한 데이터를 얻기가 쉽지 않다는 것을 의미한다. 예를 들어, 어떤 사람이 실제로는 다른 곳으로 이사를 갔으나 주민등록 상의 주소는 여전히 자신의 부모와 동거하는 것으로 나타나는 경우, 이 사람의 주민등록 상의 주소지가 과연 정확한 것인지에 대한 문제가 남게 된다. 또한 설문조사에서 소득 수준에 대한 질문을 받은 응답자가 자신 혹은 가구의 소득에 대한 정확한 정보를 알지 못하거나, 자신들의 소득 정도를 과장 혹은 축소하는 등 다양한 이유로 소득 수준에 대한 부정확한 정보를 제공할 가능성이 있다. 주소나 소득 수준 외에도 교육 수준 역시 수집된 데이터가 적절한 가중치 변인으로 사용될 수 있느냐는 데 의문을 갖지 않을 수 없다.

이전의 세대에 비해 현격하게 대학 진학률이 높아져 학력 인플레이션이 나타난 젊은 세대의 상황에서 학력을 설문조사의 가중치 변인으로 사용했을 경우, 이러한 가중치가 얼마나 유용한 효과를 가져올 수 있을 것인지에 대해서는 다시 한 번 고민해 볼 필요가 있다. 또 다른 변인으로 들 수 있는 직업의 경우 응답자들이 정확한 정보를 주느냐의 문제를 넘어서, 어떠한 방법으로 혹은 어느 정도 세분화 하여 유형을 나눌 것이며, 이를 어떻게 수량화 할 것인지가 매우 어려운 작업으로 남게 된다.

인구총조사를 준거조사로 사용하는 데 있어 나타난 이러한 어려움에도 불구하고 인구통계학적 요인들만으로 편향을 제거하기에 충분하다는 확신이 든다면, 이는 충분히 받아들일 만한 선택일 것이다. 실제로 한국에서 진행된 수많은 설문조사들은 나이와 성별, 거주 지역과 같은 인구통계학적 변인들을 대상으로 가중치를 부여하고 있으며, 따라서 인구총조사를 준거조사로 사용하는 것은 일견 타당해 보였다. 하지만 응답자들의 대담에 영향을 주는 보다 복합적이고 다양한 요인들이 있음을 고려한다면, 이러한 인구통계학적인 요인에만 기대어 가중치를 부여하는 것은 충분하지 않아 보인다.

따라서 앞서 언급되었던 학력이나 소득수준, 직업과 같이 기존에 논의되었던 인구통계학적인 요인 외에 인터넷을 통한 온라인 조사에서 편향을 줄일 수 있는 다른 요소를 찾아보는 데 이 연구의 초점이 맞춰져 있다. 물론 라이프스타일의 어떤 측면이 온라인 조사에서 응답자들의 응답에 영향을 미치는지에 대한 확신은 없으나, 적어도 인구통계학적인 요인만으로는 인터넷 조사에서 오는 편향을 제거하는 데 충분치 않다고 판단되기

때문이다. 그러므로 이 연구는 설문조사 결과에 가중치를 부여하는 추가적인 변인이 필요하다고 판단하고, 인구총조사에 포함되지 않은 항목들 중에서 어떤 항목이 가중치를 부여하는 데 사용될 수 있을 것인가에 주목하고자 한다.

예를 들어 인구총조사는 사람들이 온라인 구매를 얼마나 자주 하는지를 묻지 않지만, 이러한 데이터가 실제 조사 결과의 가중법에서는 훨씬 더 유용할 수도 있다. 해리스 인터랙티브와 같은 회사는 실제로 많은 인터넷 조사들에 대한 준거조사로 사용하기 위한 전화 조사에서 비인구통계학적인 질문들을 포함시켜 데이터를 구축하였다. 하지만 준거조사 생성을 위한 이러한 방법은 수행하기까지 시간과 비용 측면에서 과도한 부담이 들기 때문에 개인 연구자가 추가적인 준거조사를 실시하는 것은 사실상 거의 불가능해 보인다. 또한 연구자들은 준거조사 데이터와 가중치를 부여하는 데 사용할 데이터 변수들이 결과에 영향을 미칠 것이라는 확신이 필요하다.

Ⅲ. 준거조사로서의 시간 사용 조사

앞서 살펴본 것처럼 일반 연구자가 준거 자료의 필요성 때문에 새로운 조사를 실시하는 것은 시간과 비용의 문제가 따르기 때문에, 이 연구는 연구자가 실제로 자신의 준거 데이터를 생성하는 대신 준거조사로 시간 사용 조사를 사용함으로써 성향 점수 가중법이나 다른 유형의 가중법을 인터넷 기반 조사에 적용하는 것을 제안하고자 한다. 시간 사용 조사는 대부분의 선진국에서 정부 기관에 의해 실시되고 있다. 국내의 경우 생활시간 조사라고 하여 매 5년마다 한국인들의 생활시간 사용실태 조사가 약 8천 가구를 대상으로 이루어지는데, 10살 이상의 가구원을 대상으로 지난 이틀 동안 자신이 무엇을 했는지를 기록하도록 하는 방식으로 진행된다. 사람들의 생활시간 이용은 계절에 따라 달라지기 때문에 이러한 생활시간 조사는 봄과 가을에 두 번 시행하게 되고, 이 두 개의 데이터를 합쳐서 발표하게 된다. 물론 이러한 생활시간 조사가 개인의 기록에 의존하기 때문에 기억의 왜곡에서 오는 문제점을 제기할 수 있으나, 생활시간 조사의 설계 과정에서 이러한 문제점을 어느 정도 극복할 수 있을 것이라 판단했다.

본 연구는 이러한 생활시간 조사가 준거조사로 사용될 만큼 충분히 정확하고 신뢰할 수 있다고 본다. 따라서 시간 사용에 대한 설문문항을 포함하여 조사를 실시할 경우, 그에 대한 준거조사로 생활시간 조사를 사용할 수 있을 것이다. 일반적으로 시간 사용 조

사는 사람들이 하는 다양한 활동들을 추적하고자 응답자들에게 몇 가지 구체적인 질문들을 통해 시간을 어떻게 사용하는지 간단히 답할 수 있게끔 하고 있다. 따라서 이러한 문항을 사용할 경우, 공식적인 생활시간 조사와 설문 응답을 큰 어려움 없이 비교하여 이들 정보를 코딩할 수 있다. 시간은 숫자를 사용하여 하루 중 시간을 언급하거나 실제 시간의 길이를 나타내는 방식을 사용하여 측정되기 때문에 시간 사용에 대한 설문은 일반적으로 어떤 일을 하는 데 얼마나 시간이 걸리는지, 혹은 보통 때 어떤 일을 하는 데 얼마나 시간이 걸리는지 등으로 간단히 물어보면 된다. 이렇게 함으로써 생활시간 조사에 의해 수집된 정보와 비교하는 것이 가능하다.

이러한 시간 사용 조사는 정보를 수집하는 것이 용이할 뿐만 아니라 사람들의 세계관이나 의견에 영향을 미치는 습관과 같은 통찰력을 제공할 수 있다는 점에서 매우 가치 있다고 판단된다. 예를 들어 누군가 밤늦게 집에 귀가한 경우, 이 사람은 문화활동을 즐겼을 수도 있고 외식을 했을 수도 있다. 다른 한편으로 이 사람은 밤늦게까지 야근을 해야만 했을 수도 있는 것이다. 누군가의 취침시간 역시 그 사람의 라이프스타일을 파악할 수 있는 귀한 정보를 제공한다. 만약 늦게까지 깨어 있는 것을 좋아하는 사람이나 일찍 잠자리에 드는 걸 좋아하는 사람을 상상하는 경우, 우리는 다른 종류의 사람을 생각해 볼 수 있다. 이런 두 유형의 사람 사이의 정확한 차이를 이해하지 않고도 우리는 여전히 그들의 취침 시간을 알고 있다면, 이 정보를 통해 우리의 결과에 가중치를 부여할 수 있는 것이다.

또한, 인터넷 사용에 대한 질문은 그것을 통하여 응답하는 사람들의 성향을 어느 정도 알 수 있기 때문에 인터넷 사용유형과 시간을 파악하는 것은 매우 중요하다. 예를 들어 우리는 응답자의 성향을 평가하기 위해 인터넷 사용을 정보 검색, 이메일이나 채팅을 포함해 사회 활동 참여, 컴퓨터 게임, 홈페이지나 블로그와 같은 웹사이트 관리(여기에 온라인 쇼핑도 포함) 등의 네 가지 범주로 유형화해서 질문할 수 있고, 나아가 응답자들에게 각각의 범주에 대해 온라인에서 보내는 시간의 양을 응답하도록 할 수 있다.

이러한 인터넷 사용에 대한 응답은 비단 인터넷 사용에 대한 것뿐 아니라 추가적인 정보를 제공해 주기도 한다. 실제로 컴퓨터 게임을 많이 한다고 응답한 사람의 경우 극장이나 공원에 가는 것과 같은 야외활동보다 컴퓨터 앞에서 보내는 여가시간이 훨씬 더 많다는 것을 보여준다. 물론 이러한 라이프스타일의 차이가 사람들의 견해에 영향을 미친다는 명확한 증거는 없다 하더라도 그것이 사람들의 의견이나 생각에 영향을 미칠 수 있는 가능성은 충분히 존재하는 것이다. 인터넷을 많이 사용하는 사람은 인터넷에서 시

사 문제에 대한 정보를 얻는 반면, 그렇지 않는 사람들은 신문이나 텔레비전 뉴스 프로그램을 통해 정보를 얻게 된다. 이처럼 정보를 얻는 정보원의 차이가 사람들의 사고에 영향을 미칠 수 있는 가능성은 얼마든지 있는 것이다.

실제로 개인의 활동을 분석하고 온라인으로 응답하는 사람들의 관점과 성향에 영향을 미치는 요인을 연구하는 것보다, 시간 사용 조사의 항목들을 사용하여 상대적으로 간단한 질문을 함으로써 라이프스타일의 차이를 이해하는 것과 같은 혜택을 얻을 수 있다. 이에 따라 우리는 취침시간과 귀가시간과 같은 정보가 표본의 편향을 줄이는 데 훨씬 더 효과적일 것이라는 가정을 해 볼 수 있는 것이다.

이 연구는 통계청에서 실시한 기존 면대면 조사의 온라인 버전을 사용하여 이러한 시간 변인들이 편향을 줄이는 데 도움을 줄 수 있는지 확인해 보고자 했다. 온라인 버전에는 귀가 시간, 취침 시간, 앞에서 언급했던 인터넷 사용 유형에 따른 시간과 관련된 항목들이 추가되었다. 온라인 설문조사 결과를 조정하기 위한 준거조사로는 역시 통계청에서 실시한 한국인 생활시간 조사를 이용하였으며 가중치는 림 가중법(rim weighting)을 사용하였다. 즉 온라인 설문결과에 생활시간조사 결과에 입각한 가중치를 부여했을 때 조사결과가 공식버전과 좀더 일치하는 방향으로 변화되는지를 검토하였다.

IV. 연구방법

이 연구에서 사용된 인터넷 설문조사 자료는 통계청에서 매년 실시하는 사회조사와 병행하여 실시해 수집된 것이다. 통계청에서 실시하는 사회조사는 면대면 조사로, 사람들의 일상의 다양한 측면들을 담고 있으며, 사람들의 삶의 질과 사회 변화를 이해하고자 하는 목적으로 수행되고 있다. 2009년도에 실시된 조사는 사회 복지와 노동, 소득과 지출, 여가 활동 등에 대한 질문으로 구성되었다. 이 조사의 온라인 버전은 면대면 조사 버전과 동일한 질문 외에 인터넷 이용시간 등과 같은 생활시간 사용관련 질문을 추가한 것이다. 온라인 조사는 인터넷 자원자 표본을 이용한 추정방법을 연구하기 위한 것이었다¹⁾.

실제로 이 조사를 위해 통계청을 포함하여 인터넷 버전의 설문조사 응답자를 찾는 광고가 여러 웹 사이트에 게재되었으며, 응답자들에게는 인센티브를 주는 방식으로 설문조

1) 조사방법 및 연구결과는 조성경·허명희(2009) 참조.

사를 실시하였다. 참가자들은 설문조사를 완료하기 위해 다른 가족 구성원의 참가를 독려하기도 하였으나, 본 조사는 전 세대원이 참가했는지 여부와 무관하게 모든 응답을 인정하였다.

면대면 버전 조사는 일반적인 방법으로 진행되어 셀 가중법을 사용해 인구조사 데이터를 기반으로 결과에 가중치를 부여하였다. 일반적으로 통계청에서 사회조사에 사용하는 가중치 변인으로는 인구통계학적 변인 세 가지, 즉 나이, 성별, 거주 지역을 사용한다. 따라서 이 연구에서도 이들 세 가지 인구통계학적 변인에 가중치를 부여하고, 인터넷 조사와 면대면 조사결과를 비교하였다. 여기에 몇 가지 추가 변인들을 사용하여 인터넷 조사 표본에 림 가중법을 적용해 보았다. 인터넷 조사에는 사람들의 귀가시간과 취침 시간, 인터넷 사용 유형에 따른 이용 시간이라는 시간 변인들이 포함되어 있어 이들을 가중 변인으로 사용하기로 한 것이다.

이 연구는 생활시간 조사를 이용해서 가중치를 개발한 경우 인터넷 조사의 일반성을 높일 수 있는지를 분석하기 위해 오프라인 데이터와 가중치를 부여한 온라인 데이터를 비교·분석하였다. 그리고 2004년 생활시간 조사를 데이터에 가중치를 부여하는 준거조사로 사용되었다. 하지만 실제 생활시간 조사와 온라인 버전 간에는 생활시간을 묻는 문항에 다소 차이가 있었다. 예를 들어 우리가 사용한 조사는 온라인 게임을 얼마나 하는지 서술하게끔 하였는데, 생활시간 조사는 컴퓨터 게임을 얼마나 하느냐고 묻는 식이다. 즉, 컴퓨터 게임에는 인터넷 접속 여부와 상관없이 컴퓨터에 내장된 아케이드 게임을 하는 경우가 포함되어 있어 온라인 게임을 하는 데 보내는 시간과는 다른 결과를 가져올 수 있다.

하지만 이보다 더 중요한 것은 사람들이 직장에서 온라인을 통해 보내는 시간에 대한 고려가 없었다는 점이다. 직장에서 온라인을 통해 작업을 하는 경우는 생활시간 조사에서 의미하는 여가시간으로의 온라인 이용 시간과는 차이가 있게 됨에도 불구하고, 우리가 실시한 조사의 질문은 직장에서 온라인을 이용하는 시간을 제외하지 않았기 때문에 온라인 이용 시간을 그대로 이용하기 어려운 점이 있었다. 또한 2004년 생활시간 조사의 경우 블로그나 웹 사이트 운영과 같은 항목이 별도로 마련되어 있지 않았기 때문에 데이터가 약간 부정확할 수 있는 여지가 존재한다. 따라서 결론적으로 인터넷 이용 항목의 경우 정보 검색과 사회 참여(이메일, 채팅 포함), 컴퓨터 게임 이용이라는 세 가지 유형만 사용하기로 했다.

분석의 결과 온라인 조사 응답자의 인터넷 사용과 생활시간 조사 응답자의 인터넷 이용에는 상당한 차이가 발견되었다. 예를 들어 생활시간 조사 응답자의 94.9%가 사회적 목적으로 인터넷을 이용한 적이 없다고 응답한 반면, 온라인 조사 응답자의 경우 47.4%만이 사회적 목적으로 인터넷을 이용한 적이 없다고 답한 것이다. 따라서 인터넷 사용을 코딩함에 있어서 특정 목적으로 인터넷을 사용했는지 여부만을 고려하기로 했다. 마찬가지로 정보 검색을 위해서는 생활시간 조사 응답자의 83.4%, 온라인 조사 응답자의 25.3%가 인터넷을 사용한 적이 없다고 답했다. 컴퓨터 게임의 경우 차이는 좀 줄어들었지만 생활시간 조사 응답자의 86.7%, 온라인 조사 응답자의 73.3%가 인터넷을 사용한 적이 없다는 비슷한 결과를 보였다. 따라서 실제로 본 연구에서는 인터넷 이용 시간이 아니라 이들 행동을 했는지 안 했는지 여부만 고려하기로 했다.

또한 온라인 조사는 귀가시간과 취침시간 항목에 가중치를 부여하였는데, 귀가시간의 경우 저녁 6시 전, 6시에서 8시 사이, 8시에서 10시 사이, 밤 10시 이후의 네 그룹으로 나누고, 취침시간은 밤 10시 전, 10시에서 12시 사이, 12시 이후의 세 그룹으로 나누었다. 앞에서 말한 나이, 성별, 거주 지역과 같은 인구통계학적 요인과 마찬가지로 이들 요인들 역시 림 가중법을 사용하여 가중치를 부여하였다. 그리고 공식적인 면대면 조사와 온라인 조사를 인구통계학적 요인만 가중치를 부여했을 경우와 시간 요인도 가중치를 부여했을 경우로 나누어 비교·분석하였다.

V. 연구결과

연구결과는 그다지 고무적이지는 않았다. 우리는 사회조사 중 106개 항목에 가중치를 부여했다. 이러한 가중치 부여가 인구통계학적 요인만을 이용해 가중치를 부여했을 경우보다 다소 나은 결과를 보여주시는 했지만 그 크기가 시간 변인 가중법의 효과를 주장하기에는 다소 부족하게 나타났다. <표 1>에서 보듯이 시간 변인을 이용해 가중치를 부여했을 경우 평균적으로 오차가 감소하는 경향을 보이고는 있지만 효과의 크기가 충분하다고는 볼 수 없었다.

그러나 이러한 조정의 대부분은 충분히 큰 것은 아니었지만, 적어도 올바른 방향으로 이동했다는 점에서 의미 있다고 판단된다. 이러한 시간 사용 가중치는 <표 2>에서 보듯이 여가 등과 관련된 항목 및 매스 미디어 사용 항목에서 보다 더 효과적인 것으로 나타났다. 삶의 질과 관련된 항목에서도 좋은 결과를 보여주었다. 예컨대 TV시청시간, 여

행빈도, 여가활동, 그리고 다양한 측면에서의 생활만족도 항목에서 긍정적인 결과가 나타났다. 따라서 시간변인을 가중치에 포함하는 것은 적용분야에 따라 좋은 결과를 나타낼 수도 있다는 점을 보여준다. 특히 준거조사로 활용된 생활시간 조사결과가 2004년도에 수집된 것이라는 점을 감안하면 이러한 분석결과는 충분히 생활시간 변인의 활용가능성을 보여주고 있다고 해석된다.

〈표 1〉 기본 가중 모형과 시간 변인 가중 모형의 오차 크기 비교

	문항수	최소오차	최대오차	오차크기 평균	표준편차
기본가중모형	106	.00	.70	.0592	.07908
시간변인 림가중	106	.00	.61	.0553	.07569

* 출처: 조성겸·허명희(2009), 인터넷 자원자 표본에 대한 추정방법 연구, 통계청 정책연구용역 보고서, p.47 재구성.

〈표 2〉 문항 유형별 오차

		문항수	오차 평균	표준 편차
매체 접촉	기본 가중 모형	11	.109	.20
	시간 가중 모형	11	.091	.18
관광, 여가	기본 가중 모형	31	.050	.04
	시간 가중 모형	31	.037	.05
인간 관계	기본 가중 모형	8	.067	.04
	시간 가중 모형	8	.050	.02
소득 소비	기본 가중 모형	12	.068	.08
	시간 가중 모형	12	.093	.07
사회관계	기본 가중 모형	7	.090	.06
	시간 가중 모형	7	.075	.06
봉사, 자선활동	기본 가중 모형	18	.021	.01
	시간 가중 모형	18	.042	.06
생활 만족	기본 가중 모형	19	.061	.06
	시간 가중 모형	19	.048	.05

* 출처: 조성겸·허명희(2009), 인터넷 자원자 표본에 대한 추정방법 연구, 통계청 정책연구용역 보고서, pp.48~49 재구성.

따라서 우리는 이 결과가 생활시간 조사를 이용한 시간 사용 가중법이 인터넷 조사의 가중법으로 가능성을 보여주었다는 점에 주목할 필요가 있다고 본다. 물론 이러한 조정 가중법의 효과가 상대적으로 경미하다고 해도 말이다. 따라서 앞으로는 이러한 시간 사용 가중법이 보다 더 일치한 결과를 가져올 수 있도록 하기 위한 수정이 필요하며, 좀더 정밀한 가중치에 대한 시도 또한 필요할 것으로 판단된다. 앞서 언급했던 것처럼 인터넷 이용에 대한 가중치 부여과정에서 3개 유형밖에 사용하지 못한 것이 한계로 남는다. 따라서 보다 다양한 범주의 인터넷 이용 유형에 따른 이용 시간에 대한 조사가 병행된다면 보다 정확한 결과를 가져올 수도 있을 것이다.

VI. 후속 연구를 위한 제언

이 연구는 온라인 조사와 면대면 조사의 표본 집단 차이가 가져오는 편향을 줄이고자 기존의 인구통계학적 변인을 바탕으로 하는 가중치 외에 시간 사용 변인을 추가한 가중법의 효용성을 검토해 보고자 하였다. 하지만 시간 사용 변인을 위해 사용한 데이터가 2004년 생활시간 조사로, 면대면 조사와 온라인 조사가 행해졌던 2009년의 데이터와는 차이가 있어 정확한 비교가 어려웠다는 한계를 지닌다. 앞으로는 2009년 시간조사 자료를 활용할 필요가 있다. 특히 향후에 진행되는 설문문의 경우에는 시간 활용의 다양한 측면에 관한 문항을 포함함으로써 보다 정교한 가중치를 개발할 필요가 있다.

이 연구가 사용한 시간 사용 조사 가중법은 인터넷 설문문의 가중치를 통한 준거조사에서 가장 유망하고 잠재적인 가능성이 높은 조사라 할 수 있다. 이는 접근이 쉽고, 정기적으로 수행되고 있으며, 신뢰도가 높은 방법일 뿐 아니라 사람들의 라이프스타일에 대한 통찰을 제공해 주기 때문이다. 따라서 이와 같은 가중법을 이용한 더 많은 후속 연구들을 진행하여 시간 이용 조사 가중법의 실효성을 검토해 보는 것이 필요할 것이다.

참고문헌

- 김영원·배예영. 2010. “인터넷 선거여론조사 가중치보정을 위한 성향점수의 활용.” 《한국통계학회논문집》 17(1): 55-66.

- 조성겸·허명희. 2009. 《인터넷 자원자 표본에 대한 추정방법 연구》. 통계청 정책연구용역 보고서.
- 허명희·조성겸. 2010. “자원자 패널에 의한 인터넷 조사의 성향조정 가중화.” 《조사연구》 11(2): 1-28.
- Duffy, Bobby, Kate Smith, George Terhanian and John Bremer. 2005. “Comparing Data from Online and Face-to-Face Surveys.” *International Journal of Market Research* 47(6): 615-639.
- Schonlau, Matthias, Arthur van Soest, Arie Kapteyn and Mick Couper. 2009. “Selection Bias in Web Surveys and the Use of Propensity Scores.” *Sociological Methods Research* 37: 291-318.
- Shonlau, Matthias, Kinga Zapert, Lisa Payne Simon, Katherine Haynes Sanstad, Sue M. Marcus, John Adams, Mark Spranca, Hongjun Kan, Rachel Turner and Sandra H. Berry. 2004. “A Comparison Between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey.” *Social Science Computer Review* 22(1): 128-138.
- Taylor, Humphrey, John Bremer, Cary Overmeyer, Jonathan W. Siegel and George Terhanian. 2001. “The Record of Internet-Based Opinion Polls in Predicting the Results of 72 Races in the November 2000 U.S. Elections” *International Journal of Market Research* Vol. 43 (2):127-138.

<접수 2010/2/23, 1차수정 2011/3/15, 2차수정 2011/7/6, 게재확정 2011/7/7>