

제2장

2010 인구주택총조사 자료수집방법 비교분석 -대면·인터넷·우편조사를 중심으로-

박영실 · 정남수

제1절 서론

1. 연구배경

급변하는 조사환경 속에서 응답률 하락에 대한 우려의 목소리가 증가하고 있다. 이는 비단 어느 한 기관 및 국가에서만 나타나고 있는 문제는 아니며, 전 세계적으로 일어나는 보편적인 현상이라는 것이 여러 연구에서 지적되고 있다(Atrostic et al, 2001 등).

이러한 환경은 자료수집방법(data collection mode)에서의 변화를 요구하고 있는데, 그 대안 중 하나로 조사 대상의 하위 그룹에 서로 다른 자료수집방법을 적용하는 혼합방법조사(mixed-mode surveys)가 증가하고 있다(de Leeuw, 2005). 대면조사나 전화조사, 우편조사 등 어느 한 가지 방법을 이용하여 자료를 수집하기보다는 여러가지 방법을 혼합하여 조사함으로써 각 방법의 장점을 취하여 응답률을 향상시키는 대안으로 활용하고 있다. 예컨대, 우편조사를 한 후 무응답자에 대해서 대면조사를 한다든가, 패널조사(pannel survey)의 경우 첫 번째 웨이브에서는 대면조사를 한 후 두 번째 웨이브부터는 전화조사를 하는 경우 등이 그러한 예이다. 현재 이 방법은 미국, 오스트레일리아, 캐나다 등 세계 여러 국가에서 그 활용빈도가 늘어나고 있는 추세이다.

우리나라도 예외는 아니다. 통계청에서 주관하고 있는 조사를 예로 들어보면, 50개 조사 중 26개인 52.0%에서 두 가지 이상의 자료수집방

법을 혼합하여 사용하고 있는 것으로 나타났다. 조사종류별로는 사업체 조사의 47.1%, 가구조사의 87.5%, 그리고 농수산조사의 32.0%가 여기에 해당된다(통계청, 2008a). 통계청의 대표적인 조사라고 할 수 있는 인구주택총조사의 경우, 전통적으로 조사원이 가구에 방문하여 응답자와 면접을 하는 대면조사를 원칙으로 하였으나, 2005년부터는 부분적으로 인터넷조사를 도입하였으며, 향후 인터넷조사의 확대와 함께 우편조사 등을 추가로 도입할 예정에 있다.

응답률 향상에 대한 대안으로 혼합방법조사가 사용되는 것은 다음과 같은 가정에 따른 것이다(Dillman and Tarnai, 1988). ‘사람들은 개인마다 선호하는 응답방법이 다르다. 따라서 하나의 응답방법을 거절한 사람은 다른 방법으로 응답을 하게 될 것이므로 대안적인 응답방법을 제공할 경우 무응답이 줄어들게 된다.’ 혼합방법조사는 응답률 향상에 부가적으로 자료의 품질향상 및 비용감소 효과를 제공해 줄 것으로 기대되고 있다. 일반적으로 응답률이 높을 때 심각한 무응답 오차가 발생하지 않음이 암묵적으로 가정되는 상황에서(Curtin, Presser, and Singer, 2000), 응답률 향상은 무응답 오차에 대한 감소를 의미하므로 품질향상을 뜻하는 것으로 해석된다. 그리고, 높은 응답률에서 무응답 오차가 적을 것이라는 가정에 따라 무응답을 줄이기 위해 많은 비용을 투입하게 되는데 혼합방법조사는 응답률을 증가시켜 줌으로써 무응답추적을 위한 조사 비용 또한 줄여줄 수 있을 것이다(Hox and de Leeuw, 1994).

그러나, 혼합방법조사가 당초의 목적처럼 응답률 증가와 이에 따른 품질향상 및 비용감소 효과를 가져오는지에 대해서는 충분한 검토가 필요하다. 혼합방법조사를 사용하여 높은 응답률을 얻었다고 하더라도 이것이 과연 덜 편향된 자료를 만들었는지에 대해서는 의문이 제기될 수 있다(Voogt and Saris, 2005). 앞서, 응답률 향상이 무응답 오차의 감소로 이어질 수 있다고 하였으나 그러한 관계는 검증된 것이 아니다. 무응답 오차는 무응답률뿐 아니라 관심대상 변수에 대한 응답자와 무응답자 특성의 차이에 의해서도 영향을 받으므로 높은 응답률 하에서도 무응답 오차가 발생할 가능성은 잠재되어 있다. 또한 혼합방법 그 자체에 측정 오차가 발생할 개연성이 있다. 서로 다른 자료수집방법에 따라서 응답자 혹은 조사원의 반응, 조사표 설계 및 자료처리방법에 차이가 있을 수 있

고 이는 자료수집방법에 따른 차별적인 오차를 발생시킬 수 있다. 혼합방법조사에 대한 통합관리시스템이 부재한 상황에서 이러한 오차를 통제하기란 쉽지 않다. 따라서, 혼합방법조사를 실시할 경우 조사설계 단계부터 자료수집방법효과(mode effect)를 고려해야 한다. 그러나 우리나라에서는 혼합방법을 이용한 자료수집이 빈번하게 이루어지고 있음에도 불구하고 그 현황에 대한 정확한 보고나 방법효과에 관한 연구들은 빈곤한 실정이다. 이는 우리나라의 조사현실과도 어느 정도 관련이 있는 것으로 보여진다. 예컨대, 대면조사가 원칙인 조사에서 대면조사 이외의 다른 자료수집방법을 사용할 경우 이것이 조사원에 대한 평가나 업무량 배분 및 수당과 연결될 수 있기 때문에 제대로 된 보고가 이루어지지 않고 있다.

2. 연구목적 및 내용

본 연구에서는 이와 같은 연구배경 하에 ‘2010 인구주택총조사 제2차 시험조사’ 자료를 이용하여 방법효과를 검토해 보고자 한다. 과연, 자료수집방법에 따라서 무응답 오차나 측정 오차 등의 차이가 있는가? 이러한 효과가 발생할 경우, 자료수집방법에 따른 차별적인 전략을 세움으로써 전체 자료의 품질향상을 위한 대안을 마련할 수 있을 것으로 기대된다.

연구에 포함된 내용은 다음과 같다. 첫째는 혼합방법조사에서 나타나는 방법효과와 관련된 문헌연구이다. 혼합방법조사의 역사 및 주요 유형 등에 대해서 개괄적으로 소개한 다음, 혼합방법조사가 자료의 품질에 미치는 효과에 대한 문헌연구를 하였다. 둘째는 미국 센서스¹⁾의 혼합방법조사 사례 검토이다. 미국은 1990년 센서스 이후에 응답률 향상 및 비용 절감을 목적으로 혼합방법조사 실험을 지속적으로 실시해오고 있다. 물론 미국과 우리나라의 조사환경은 크게 다르므로 그 내용을 직접적으로 도입하는 데에는 무리가 있으나 실험방법 및 주요 결과는 혼합방법조사 확대를 앞두고 있는 우리에게 시사하는 점이 있을 것으로

1) 우리나라는 인구주택총조사, 미국은 센서스라는 용어를 그대로 사용하고자 한다.

기대된다. 셋째는 ‘2010 인구주택총조사 2차 시험조사’ 자료를 이용한 자료분석이다. 마지막은 문헌 및 사례연구, 분석결과를 통해 얻은 시사점을 종합하여 향후 인구주택총조사의 자료수집방법 및 이와 관련된 제반사항에 대한 개선안 제안이다.

제2절 문헌 및 사례연구

1. 혼합방법조사의 개요

가. 혼합방법조사의 역사

전통적으로 중요한 세 가지 자료수집방법은 우편조사, 대면조사, 전화조사이다. 최초의 우편조사는 Sinclair 경이 스코틀랜드 교회의 모든 교구에 있는 목사에게 우편으로 조사표를 보낸 1788년으로 기록되어 있다. 대면조사는 Bowley 경이 1912년 영국의 5개 도시에 있는 노동자계급의 상태를 연구하기 위해서 시민들을 표본으로 하여 구조화된 면접을 실시한 것이 첫 번째이다. 전화조사는 이보다 늦은 1960년대 초에 소개되었으며, 1970년대경 마케팅 분야에서 CATI(Computer Assisted Telephone Interviewing)가 실시되면서 미국 내의 자료수집방법에서 우위를 점하기 시작하였다. 컴퓨터 기술의 발전은 자료수집에서의 급격한 변화를 야기시켰다. CATI 이외에도 컴퓨터를 보조적으로 활용한 방법으로 CAPI(Computer Assisted Personal Interviewing), TDE(Touch-tone data entry) 등이 새로운 방법으로 자리를 잡았으며, 가장 최근에는 웹 혹은 인터넷 조사가 인기를 얻고 있다. 인터넷조사는 커머리거나 무응답과 관련해서는 한계를 갖고 있으나 비용과 시간 측면에서 매우 효율적인 것으로 평가받으면서 단시간 내에 인기있는 방법으로 부상하였다(de Leeuw and Collins, 1997; de Leeuw, 2005).

새로운 자료수집방법의 부흥으로 인해 광범위한 범위에서 혼합방법 조사가 이루어졌다. 물론 혼합방법조사가 최근에 시작된 것은 아니다. 우편조사와 대면조사의 역사에서도 보듯이 두 방법의 혼합형태는 자료수집이 이루어지기 시작한 초기까지 거슬러 올라가며, 전화조사가 급증

하면서 대면조사와의 혼합방법이 인기를 끌었다. 미국 센서스국에서는 1960년에 우편에 의한 송부와 인편에 의한 회수(personal pick-up) 혹은 그 반대의 경우인 인편송부와 우편회수와 같은 방법을 혼합하여 센서스에 사용하기도 했다. 같은 시기에 오스트레일리아 통계국의 농가경제조사(Agricultural Finance Survey)에서도 우편조사와 대면조사 방법을 혼합한 바 있다(Groves et al., 2004). Hochstim(1967)의 연구는 혼합방법조사에 관한 기원 중의 하나로 꼽히는데, 그는 건강에 관한 자료를 수집하는 데 있어서 우편조사와 전화조사, 대면조사의 세 가지 방법을 혼합하여 사용한 후 그 결과를 비교하였다. 추출된 표본에 무작위로 세 가지 자료수집 방법을 할당하여 응답을 얻도록 하였으며 만일 한 방법에서 응답을 얻지 못했을 경우에는 다른 방법을 할당하였다(Groves et al., 2004에서 재인용).

혼합방법조사는 1980년 이후에 확장되기 시작했으며, 많은 국가의 정부조사에서 채택되었다. 특히, 미국의 CPS(Current Population Survey), 영국의 LFS(Labour Force Survey), 핀란드의 LFS 등과 같은 노동력 조사에서는 첫 번째 웨이브에서는 대면조사를, 두 번째 웨이브 이후에는 전화조사를 하는 방식의 혼합방법조사를 사용하는 것으로 잘 알려져 있다.

나. 혼합방법조사의 유형

혼합방법조사의 유형은 다양하다. 초기에는 주로 자료수집측면에서 그 유형을 세분화하였으나(Dillman and Tarnai, 1988), 최근에는 응답자와 커뮤니케이션을 하는 모든 과정을 대상으로 어떤 방법이 사용되느냐에 따라서 혼합방법조사 유형을 구분하고 있다(de Leeuw, 2005). 이것은 조사원들이 응답자와 서로 다른 시점에서 다른 방법을 사용하여 커뮤니케이션을 한다는 것에 기반한 것으로 혼합방법을 단순히 자료수집에 한정시켜서 바라보는 것에서 벗어나 하나의 체계 즉 혼합방법체계(mixed- or multi-mode system)로 보는 것을 의미한다. 이에 따라서 사전접촉단계, 응답단계, 추적단계에서의 혼합조사방법을 그 목적과 자료품질에 미치는 효과에 따라서 유형화하였다.2) 이 중 본 연구에서는 실제 자료수집단계에서의 혼합방법을 중심으로 살펴보고자 한다(<표 2-1>).

〈표 2-1〉 혼합방법조사의 유형

유형	목적
· 한 조사시점에 단일표본과 단일조사표를 사용하나 응답자에 따라서 다른 자료수집방법을 적용	· 비용 감소 · 커버리지 향상 · 응답률 향상
· 한 조사시점에 단일표본을 대상으로 하며, 조사의 항목에 따라서 다른 자료수집방법을 적용	· 프라이버시 향상 · 사회적 바람직성 오차 감소
· 한 표본을 대상으로 조사시점에 따라서 다른 자료수집방법을 적용	· 비용 감소
· 표본에 따라서 다른 자료수집방법을 적용하며, 심지어 조사시점이나 조사표가 다른 경우도 있음	· 비교 연구 · 다른 연구 전통 · 다른 커버리지 · 다른 비용 구조

자료: de Leeuw, 2005: 238에서 재구성.

혼합방법조사의 첫 번째 유형은 같은 자료를 수집하기 위해서 두 가지 이상의 자료수집방법을 사용하는 것이다. 이는 표본에 따라서 자료수집방법을 다르게 적용하는 것으로 예컨대, 표본의 일부에는 대면조사를, 표본의 또 다른 부분에는 우편조사를 실시하여 자료를 수집하는 것이다. 이러한 혼합방법조사는 주로 커버리지를 향상시키고, 합리적인 비용 수준에서 조사를 완료하고자 하는 목적에서 수행된다. 무응답을 줄이는데 있어서 좀 더 일반적이고 효율적인 방법은 시차를 둔(sequential) 혼합방법조사이다. 일반적으로 비용이 적게 드는 방법이 주 방법으로 사용되고, 비용이 드는 방법이 응답률을 향상시키기 위한 무응답 추적조사에

2) 본조사에서 전화조사를 실시하더라도 사전에 응답자를 접촉하는 과정에서는 편지 등을 보내는 방법이 사전접촉단계에서 사용하는 혼합방법조사의 예이다. 추적단계에서 사용되는 혼합방법으로써 독촉편지는 응답률을 향상시키는 데 효율적인 도구로 알려져 있다. 대면조사를 꺼렸던 응답자에게 설득편지를 보냄으로써 응답을 독려할 수 있다. 이처럼 사전접촉단계나 추적조사단계에서 사용하는 혼합방법조사는 대부분 윈-윈(win-win) 전략의 일환으로 사용된다. 반면에 자료수집방법에서 사용하는 혼합방법조사는 매우 까다롭다.

사용되는 경우이다. 이 방법은 무응답을 감소시키는 데에는 효율적인 것으로 평가되나, 동전에도 양면이 있듯이 측정 오차가 개입되기 마련이다.

두 번째 유형은 특정한 단일 시점에 하나의 표본을 추출하여 자료를 수집하나 조사표의 항목별로 다른 자료수집방법을 적용하는 경우이다. 예컨대, 조사원 기입식의 대면조사가 원칙일지라도 민감한 항목에서 대해서는 응답자가 기입하도록 하는 방법을 취할 수 있다. 이는 응답자의 프라이버시를 증진시켜 줌으로써 사회적인 바람직성 오차(social desirability bias)를 줄이기 위한 것이다.

세 번째 유형은 시점에 따라서 다른 방법을 사용하는 것으로 동일한 응답자를 다른 시점에 여러 번 조사하는 패널조사(pannel study)가 여기에 해당한다. 패널조사에서 혼합방법조사가 사용되는 주요한 이유는 현실적인 고려와 비용의 문제 때문이다. 예컨대, 접촉단계에서 주소는 알고 있으나 전화번호나 이메일 주소 등은 모를 경우, 첫 번째 웨이브에서는 대면조사를 하고 이후에는 전화나 이메일조사를 한다. 대면조사는 협조를 얻어내거나 응답자에 대한 적절한 스크리닝 등을 하기 위해서 처음 단계에서 일반적으로 선호되는 방법이기도 하다. 비용을 줄이기 위해서 첫 번째 웨이브 이후에는 비용이 덜 드는 방법을 이용하기도 한다.

네 번째 유형은 다른 모집단 혹은 하위집단에 다른 자료수집방법을 사용하는 경우이다. 이 접근법을 사용하는 전형적인 영역은 국가 간 혹은 지역 간 비교이다. 국가 간에 자료수집과 관련한 전통이 다르거나 현실적인 제약에서 차이가 있기 때문이다. 인구가 밀집되어 있는 국가에서는 대면조사가 가능하지만, 덜 밀집되어 있는 경우에는 전화조사나 우편조사가 더 수월할 수 있다. 한 국가의 문맹률 또한 자료수집방법을 제약하는 요인이 되기도 한다.

2. 방법효과에 대한 문헌연구

흔히 혼합방법조사를 사용함으로써 응답률이 향상되거나 비용이 줄어 들 것이라고 기대하는데, 이에 대한 대가로써 측정 오차 등이 발생할 수 있다는 점에 주의를 해야 한다. 이와 함께 혼합방법조사에서 상대적

으로 강점으로 생각되던 커버리지 오차나 무응답 오차 또한 하나의 오차의 원천으로 제기되고 있다.

가. 측정 오차

방법 효과에서 측정 오차는 자료의 완전성, 사회적 바람직성 오차, 그리고 응답효과(response effect)의 세 가지 측면에서 많이 다루어지고 있다(Groves et al., 2004). 자료의 완전성은 조사항목에 응답하지 않은 정도와 관련된 오차를 의미하며, 사회적 바람직성 오차는 응답자들이 바람직한 행위에 대해서는 과대보고를 하고 바람직하지 않은 행위에 대해서는 과소보고를 하는 경향에 대한 것이다. 응답효과는 정확한 워딩, 응답 범주 및 질문 순서 등으로 조사 항목의 측정과 관련된 것이다. 이 지표들 외에도 내검실패율(edit failure rates), 코딩오차율(coding error rates) 등이 사용되고 있다(Biemer, 2001; de Leeuw and der Zouwen, 2001).

1) 자료의 완전성

대면조사와 전화조사에 관한 28개 연구(1952년~1986년)의 메타분석(de Leeuw and van der Zouwen, 1988)결과, 조사를 통해서 얻은 정보의 양은 두 가지 자료수집방법에서 차이가 있었다. 즉, 대면조사가 전화조사에 비해서 항목무응답률이 낮고 정확성 등이 높은 것으로 나타났다. 그러나 두 방법의 자료품질에서의 차이가 최근으로 올수록 줄어들었는데, 이는 자료수집방법의 발전이 방법효과를 줄여줄 수도 있음을 보여주는 결과라고 할 수 있겠다.

Catlin과 Ingram(1998)은 캐나다 통계청에서 실시한 노동력조사에서 종이조사표를 이용한 대면조사와 CATI가 자료의 품질에 미치는 효과를 평가하였다. 자료의 품질지표인 완전성은 항목무응답률과 오차율로 측정되었다. 노동력조사와 동일한 조사대상에서 RDD 방법으로 표본을 추출한 후 표본 내에서 무작위로 대면조사와 CATI 방법을 배분하였다. 조사내용은 크게 가구원의 인구학적 특성에 관한 질문과 노동력에 관한 질문의 두 가지였다. 가구항목에 대해 자료수집방법별 자료의 완전성을 보면, 항목무응답률이나 오차율은 대면조사와 CATI에서 유사한 것으로

나타났다. 그러나 노동력과 관련된 질문에서는 CATI에 비해 대면조사에서 오차율이 2배 이상으로 높게 나타났다. 이는 가구항목의 경우 질문구조가 비교적 평이한데 반해, 노동력항목은 복잡한 질문구조를 띠고 있는 상황에서 CATI의 경우, 자동적으로 질문의 순서가 통제되기 때문이라고 추정되었다.

우편조사와 이메일조사에 대한 Couper와 그의 동료들(1999)의 연구에서도 항목무응답으로 측정된 자료의 품질이 그 방법에 따라서 차이가 있는 것으로 나타났다. 전체 항목에 대한 분석에서는 방법효과와 차이가 뚜렷하지 않았으나, 인구사회학적 배경항목과 태도항목으로 나누어 분석해 본 결과, 전자에서는 우편조사의 항목무응답률이 이메일조사에 비해서 더 높았다. 그러나, 우편조사와 이메일조사에서의 방법효과에 대한 결과는 여러 연구에서 일관적으로 지지되지는 않고 있다. 이메일조사에서 항목무응답이 더 많다는 연구가 있는 반면에 우편조사에서 더 많다는 연구도 있다(Schaeffer and Dillman, 1998).

2) 사회적 바람직성 오차

일반적으로 응답자들은 민감한 질문, 바람직하지 않은 태도나 행동에 대해서는 과소보고하는 경향이 있는 것으로 나타났다. 왜냐하면 자신이 부정적으로 평가되는 것에 대한 우려로 인해서 사회적인 규범에 어긋나는 태도나 행동을 정직하게 보고하는 것을 꺼리는 경향이 있기 때문이다(Knapp and Kirk, 2003). 그런데 이러한 사회적 바람직성 오차에 자료수집방법이 영향을 미치는 것으로 나타났다. 민감한 질문에 대해서는 조사원 기입식보다는 응답자 기입식에서 그 오차가 줄어든다는 것이다(Aquilino, 1992; Aquilino and Lo Sciuto, 1990 등). 대면조사의 경우, 자신의 응답을 다른 사람이 보는 앞에서 보고하는 반면에 우편조사는 조사원과의 개인적인 접촉없이 프라이버시를 존중받으면서 응답할 수 있다는 데 그 차이가 있다. 따라서, 민감한 자료를 수집하는데 자기기입을 사용하는 장점은 여러 연구에서 지지되고 있다. 특히 약물이나 알코올과 같은 문제행동에 대한 분야에서 연구가 많이 이루어졌다.

3) 응답효과

응답효과는 질문순서(question order effects)나 응답범주순서(response order effects)에 따라서 다르게 발생한다. 일반적으로, 응답자들은 처음에 나오는 응답범주를 선택하거나(primacy effect) 혹은 마지막 응답범주를 선택하는 경향(recency effect)이 있는데, 이것은 부분적으로 자료수집 방법과 연결되어 있는 것으로 알려져 있다. 질문이 시각적으로 전달될 때에는 첫 번째 응답범주를 선택할 가능성이 높은 반면에 청각적으로 전달될 때에는 마지막 응답범주를 선택할 가능성이 높은 것으로 나타났다(Groves et al., 2004).

방법효과에 대한 문헌연구들을 통해서 다음과 같은 결론을 얻었다. 첫째, 대면조사와 전화조사, 우편조사 사이의 응답 차이는 자주 관찰되고 있으며 실질적일 때도 있다. 둘째, 사회적 바람직성 오차가 조사원 기입식보다는 자기 기입식에서 더 적게 나타난다는 것은 문헌연구에서 가장 많은 지지를 얻고 있다(Dillman, 2000).

그러나 방법효과의 결과를 해석함에 있어서 그 차이가 단순히 자료수집방법의 차이에서 기인하는 것은 아닐 수도 있다는 점에 주의를 해야 한다. 실험상황이 아닌 이상 순수한 방법효과를 찾아내는 것은 사회과학의 현실에서는 쉬운 일이 아니다. 따라서, 그 결과를 해석함에 있어서 응답자의 성별이나 연령, 가구규모 등과 같이 방법효과에 영향을 미칠 수 있는 변수를 적절히 통제하는 것이 중요하다(Biemer, 1988; Wright et al., 2001).

나. 커버리지 오차

커버리지 오차란 모집단의 구성원을 누락 혹은 중복, 잘못 조사함으로써 발생하는 오차를 말한다. 그런데, 이러한 커버리지 오차의 문제는 혼합방법조사를 사용하면 감소될 수 있다. 커버리지 범위가 서로 다른 자료수집방법을 혼합함으로써 한 가지 방법으로 자료를 수집할 때보다는 누락 측면에서 커버리지를 향상시킬 수 있기 때문이다. 그러나 혼합방법조사에서 새롭게 대두되고 있는 커버리지 오차의 문제는 누락보다

는 중복 조사의 문제이다. 조사설계를 정교하게 하지 않을 경우 한 응답자가 여러가지 방법에 동시에 응답하는 사례가 발생할 수도 있다.

미국 센서스의 경우에도 2000년 센서스부터 우편조사 이외에 다양한 자료수집방법을 허용함으로써 중복조사에 대한 대안을 마련해야 했다. 센서스국에서는 중복조사의 사례를 파악하기 위해서 자동매칭기술을 사용하였다. 센서스의 모든 거처는 고유의 식별자(unique identifier)인 센서스 ID를 가지고 있다. 만일, ID가 없는 조사표가 들어오게 되면 이 조사표를 마스터주소파일(Master Address File)과 연결시킨 후에 센서스 ID를 할당하였다. 이 ID에 의해서 중복응답을 판별하였다. 매칭규칙은 동일인 혹은 동일가구를 식별해 낼 수 있도록 매우 엄격하게 적용하였다(Census Bureau, 2000).

다. 무응답 오차

무응답 오차는 무응답률과 응답자와 무응답자의 특성 차이의 함수로 표현된다. 따라서 혼합방법조사를 통해서 응답률이 증가하였다고 해도 이것이 바로 무응답 오차의 감소를 의미하는 것은 아니다. 그러나 지금까지의 연구들을 보면 무응답 오차가 응답자와 무응답자의 특성 차이에 의해서도 영향을 받음에도 불구하고 무응답 오차에 관한 연구는 대부분 응답률에 초점이 맞추어져 왔는데 그 이유는 무응답자의 특성이 알려져 있지 않기 때문이다.

3. 사례연구: 미국 센서스의 혼합방법조사 실험

미국은 1960년 이래로 우편조사를 이용해 센서스를 실시하고 있으며, 우편조사에 응답하지 않은 가구에 대해서는 전화 혹은 대면조사를 통해서 무응답 추적조사를 하고 있다(Guarino, 2001). 그러나, 최근에는 무응답 추적조사 이전에 우편조사 이외에도 다양한 자료수집방법을 활용하여 자료를 수집하고 있다. 이는 다음과 같은 조사환경의 변화에 따른 것이다.

1980년에 75%이었던 우편응답률(mail response rate)³⁾이 1990년에

64%로 하락함에 따라 2000년 센서스에서 우편응답률을 높이기 위한 방안을 마련해야 했다. 왜냐하면 우편응답률이 하락한 만큼, 우편조사에 응답하지 않은 가구에 대한 무응답 추적조사의 비율이 증가하게 되고 이는 많은 비용을 필요로 하기 때문이다. 1990년 센서스의 무응답 추적 비용을 보면, 우편응답률이 1% 하락할 경우 이를 보충하기 위해 조사원이 방문하여 조사표를 완성하는데 1,700만 달러가 소요되는 것으로 추정되었다(Dillman, West and Clark, 1994).

이러한 상황에서 센서스국은 응답자의 협조를 증가시키는 동시에 효율적으로 자료를 수집하는 기술을 발전시켜야 한다는 요구에 직면하였다. 센서스국은 증가하는 센서스 비용을 줄이기 위해서 공공의 협조를 증가시키는 프로그램을 연구하였으며,⁴⁾ 그 중 하나가 바로 다양한 응답 방법(response mode)을 제공하는 것이었다. 이에 따라 센서스국에서는 혼합방법조사에 대한 여러가지 실험을 하였다. 여기에서는 이 실험내용을 자세히 검토해 봄으로써 우리나라에 대한 시사점을 얻고자 한다.

가. 미국 센서스 자료수집 방법 및 절차

미국 센서스와 우리나라 인구주택총조사 자료수집과정에서 두드러진 차이는 미국은 우편조사를, 우리나라는 대면조사를 기본으로 하고 있다는 점이다(<표 2-2>). 이 기본적인 방법에 의한 두 국가의 응답률을 가장 최근의 조사를 기준으로 보면, 미국은 74.0%이며, 우리나라는 99.1% 이다.

3) 센서스국은 우편조사에 의한 응답률을 우편응답률(mail response rate)과 우편회수율(mail return rate)의 두 가지로 측정하고 있다. 우편응답률과 우편회수율을 계산하기 위한 기준은 여러가지가 있는데 주요한 차이는 우편응답률의 경우, 분모에 미국의 우편 시스템과 센서스국의 스템에 의해서 빈집으로 판명된 거처, 그리고 최종적으로 존재하지 않는 것으로 결정된 센서스 2000에서의 주소가 포함되어 있다는 것이다. 우편응답률과 회수율 모두 우편회수지역을 대상으로 하되 우편시스템 및 조사원에 의한 우편전달, 인터넷조사, Be Counted program, 전화조사, 커버리지 내검 추적조사 절차에서 우편으로 회수된 경우를 대상으로 한다(Treat, 2003).

4) 주요 테스트를 통해서 높은 응답률과 관련된 것으로 파악된 요인은 다음과 같다. 응답자 친화적인 조사표의 구성(3.4%), 약간 짧아진 조사표 형태(4.6%), 조사항목의 재배치(10.4%), 사전편지(6.4%), 사전편지 이후의 감사 및 독촉엽서(6.3%)이다. 이 모든 요인을 사용하였을 경우에 추정되는 응답률은 71.4%이었으며, 이런 처치 없이 기존의 조사표 패키지만 보냈을 경우의 응답률은 30.3%였다(Dillman, West, and Clark, 1994).

〈표 2-2〉 미국 센서스 자료수집방법 및 실시 기간

		미국		우리나라		
		주요 내용	시점	주요 내용	시점	
준비단계		사전편지발송		준비조사	10.29~10.31	
		감사/독촉엽서발송				3.20
자료수집 (I)	기본 방법	우편 조사	우편발송/ 우편회수	3.13 ~4.18	대면조사	11.1~11.15
			조사원전달/ 우편회수	3.3 ~4.18		
		대면조사		1.31~5.30		
	부가 방법	인터넷조사		3.3~4.18	인터넷조사	10.29~11.12
		전화조사		3.3~6.30		
Be Counted Program		3.31~4.17				
자료수집(II) :무응답 추적		대면조사		4.27~6.26	대면조사 ⁵⁾	11.28~11.30

자료: 통계청, 2006b, 2006c; U. S. Census Bureau, 2000.

주: 미국은 2000년 4월 1일, 우리나라는 2005년 11월 1일을 기준으로 한 것임.

자료수집방법별 응답률에 대한 기존 문헌검토 결과를 상기해 보면, 우편조사는 대면조사와 전화조사에 비해 응답률이 낮은 방법이다. 그럼에도 불구하고, 우편조사의 응답률이 다른 방법에 비해서 선호될 수 있는 것은 미국에서 센서스가 차지하는 위상과 관련이 있는 것으로 생각된다. 첫째, 센서스는 미국 내에서 의석수 및 예산을 배분하기 위해 해당 조사연도의 12월말까지 대통령에게 보고해야 하므로 법적인 강제력이 매우 강하다. 둘째, 우편 시스템의 체계적인 정비로 인해서 각 가구에 우편이 정확히 도달될 수 있다. 셋째, 이와 관련하여 센서스국에서는 우편조사의 가능성 여부를 조사지역(Enumeration Area) 설정에 있어서도 중요한 기준으로 제시하고 있다. 센서스국은 미국을 대략적으로 7.2백만 개의 블록(block)으로 나누고 있으며, 각 블록을 다시 주소 유형(address type), 센서스 조사표 전달방법 혹은 특별조사절차의 필요성 등에 따라 6개 조사지역으로 구분하였다(<표 2-3>). 6개 조사지역은 Mailout/Mailback,

5) 우리나라의 경우, 조사기간 중 부재나 불응 등으로 인해서 조사대상의 20% 이상이 무응답인 조사지역에 대해서 추가조사를 실시하였다. 센서스 본조사가 완료된 후 약 2주 후에 해당 지역을 방문하여 조사표를 다시 작성하거나 부재나 불응 등으로 조사가 불가능한 경우에는 최소한의 정보라도 파악하도록 하였다(통계청, 2006c).

Update/Leave, UrbanUpdate/Leave, Update/Enumerate, List/Enumerate, RemoteAlaska로 분류되며 앞의 세 지역을 합하여 우편회수지역(Mailback Area)이라고 부른다.⁶⁾

〈표 2-3〉 조사지역 유형별 자료수집방법

(단위: %)

조사지역유형	자료수집방법	블록 비중	가구 비중
Mailout/Mailback	우편발송+우편회수	52.7	79.8
Update/Leave	조사원전달+우편회수	41.7	18.8
UrbanUpdate/Leave	조사원전달+우편회수	0.2	0.2
Update/Enumerate	대면조사	2.3	0.8
List/Enumerate	대면조사	3.0	0.3
RemoteAlaska	대면조사	0.1	0.0

자료: Treat, 2003.

가구번호와 거리이름이 있는 도시유형의 주소(예, 801 Main Street)로 정비되어 있는 Mailout/Mailback 지역에는 우편시스템을 통해서 조사표를 발송하고 회수하며, 우편조사 마감기일인 4월 18일까지 조사표가 회수되지 않은 경우에 조사원이 방문하여 대면조사를 실시한다. 우편시스템에 의해서 조사표의 정확한 전달이 어려운 Update/Leave와 UrbanUpdate/Leave 지역에 대해서는 센서스국 스태프가 직접 해당 주소에 방문하여 조사표를 전달하며, 응답자들이 조사표를 기입한 후 우편으로 보낸다. 이 지역에서 역시 4월 18일까지 조사표가 회수되지 않을 경우 조사원 방문조사를 하였다. 이 중 Update/Leave 지역은 도시유형의 주소와 비도시유형의 주소(예, Rural Route 7, Box 4.)가 혼재되어 있는 지역이며, UrbanUpdate/Leave는 본래 Mailout/Mailback 지역이었으나 도시 내부에 다가구(multi-unit) 구조를 포함하고 있어 조사표의 정확한 전달이 어려운 지역이다. 한편, 가구들 간 거리가 매우 멀고, 인적이 드문 지역의 경우에는 조사원이 직접 방문하여 면접을 하는 방법으로 조사가 진행되었는데, 여기에는 Update/Enumerate, List/Enumerate, RemoteAlaska가 있다(U. S. Census Bureau, 2000).

6) 6개 지역에 대한 자세한 설명은 treat(2003)을 참고할 것.

센서스국은 2000년도에 우편조사의 응답률을 증가시키기 위해서 다 단계우편접촉 전략을 채택하였다. 제일 먼저 응답가구에 센서스 조사표에 대한 설명을 제공하는 사전편지(advance letter)를, 다음으로 센서스 조사표를 발송하였다. 세 번째로는 감사/독촉엽서(reminder postcard)를, 최종적으로 조사표가 회수되지 않은 가구에 대해서 교체조사표(replacement questionnaire)를 보냈다(NRC, 2004).

우편조사와 대면조사 이외에도 2000년 센서스에서는 인터넷조사, 전화조사(CATI), Be Counted Program⁷⁾을 통해서 응답자들이 좀 더 쉽게 응답을 할 수 있는 방법을 채택하였다.

센서스 표본조사를 대체하고 있는 미국지역사회조사의 경우에도 전체적인 흐름은 전수조사와 유사하다.⁸⁾ 표본가구에 우편으로 조사표를 보내고 이것을 다시 우편으로 회수하는 것을 원칙으로 하며, 우편조사 이후 무응답가구에 대해서 컴퓨터를 보조적으로 활용한 전화조사와 방문조사를 실시한다. 우편이 발송된 지 6주가 지나면 상업벤처(vendor)로부터 전화번호를 넘겨받아서 CATI를 실시하며 다시 3주 이후에는 무응답자들의 하위표본을 대상으로 CAPI를 수행한다. 2000년 조사의 경우 우편응답률은 약 53%, 무응답 추적조사 이후 최종 응답률은 96.5% 가량인 것으로 나타났다(Griffin et al., 2001).⁹⁾

나. 1993년 센서스 전국 시험조사: MTMT

1) 목적

우편응답률을 향상시키기 위한 혼합방법조사 실험은 1993년 전국시험조사(National Census Test)에서 다음과 같은 세 가지 목적을 갖고 처음

7) Be Counted Program 이라 함은 주소가 지정되어 있지 않은 종이조사표를 해당 지역의 목표지점에 배포해 놓으면 센서스에서 조사되지 않았다고 생각하는 응답자가 그 조사표를 직접 가져가서 완성한 후에 우편으로 보내도록 하는 것이다.

8) 미국지역사회조사는 미국 센서스국이 인구, 가구, 사회경제학적 자료를 수집하기 위해 개혁적으로 발전시킨 새로운 개념의 조사로 10년 주기로 작성되던 센서스 표본조사 결과를 모든 주, 카운티, 시티, 메트로폴리탄, 인구 65,000명 이상의 지역에 대해서 매년 추정치를 생산하는 목적으로 계획되었다.

9) 2005년의 경우에는 면접이 가능한 ACS 표본 주소 중에서 51%는 우편에 의해서, 9%는 CATI에 의해서, 38%는 CAPI에 의해서 수집되었다(U.S. Census Bureau, 2006).

으로 실시되었다. 첫째는 우편조사와 전화조사 각각에 대한 응답자의 선호도를 알아보는 것이고, 둘째는 대안적인 자료수집방법을 제공하는 것이 응답률 향상을 가져오는지를 살펴보는 것이며, 셋째는 전화조사에 의해 수집된 자료의 품질을 검토하는 것이다(Bates, 1993).

2) 방법론

MTMT(Mail and Telephone Mode Test)는 우편조사의 각 단계마다 전화조사를 대안으로 제시하여 우편조사와 전화조사 응답률을 비교할 수 있도록 설계되었다. 우편조사 는 사전편지→조사표→감사/독촉엽서→(추적편지)→교체조사표 단계로 이루어졌으며, 이 중 어느 단계에 전화조사 대안을 제공하느냐에 따라서 총 5개의 실험집단을 구성하였다.

첫 번째 실험집단은 통제집단으로서 오직 우편조사에 의한 응답만 가능하였다. 두 번째 집단은 감사/독촉엽서 단계에서, 세 번째 집단은 감사/독촉엽서와 추적편지 단계에서, 네 번째 집단은 감사/독촉엽서, 추적편지, 교체조사표의 세 단계에서 전화조사 응답이 가능하다는 설명을 들을 수 있었다. 다섯 번째, 집단은 우편조사의 모든 단계마다 전화조사 대안을 소개받았다(<표 2-4>).

<표 2-4> MTMT 실험집단 처치 내용

구분	집단1	집단2	집단3	집단4	집단5
사전편지	M	M	M	M	M+T
조사표	M	M	M	M	M+T
독촉/감사엽서	M	M+T	M+T	M+T	M+T
추적편지	-	-	M+T	M+T	-
교체조사표	M	-	-	M+T	M+T

자료: Dillman et al., 1994.

주: M은 우편조사를, T는 전화조사를 의미함.

MTMT 모집단은 센서스 조사지역 중 우편회수지역에 위치한 가구 중 post-1990 센서스 테스트에 포함된 가구를 제외한 88,817,070가구이다. 이 가구들은 소수인종의 비율과 1990년 센서스 응답률을 기준으로 응답이 낮은 지역(LRA, Low Response Areas)과 응답이 높은 지역(HRA, High Response Areas)으로 층화되었다. 최종적으로 22,500가구가 추출되

었으며 LRA와 HRA 각각에 11,250가구가 배분되었다. 각 층에서 다시 2,250가구를 계통추출하여 5개 실험집단에 무작위로 할당하였다.

3) 분석결과

여기에서는 자료수집방법별 응답률 비교(Dillman et al., 1994)와 항목 무응답률 비교(Bates, 1993) 결과를 중심으로 살펴보도록 하겠다.

가) 자료수집방법과 응답률

자료수집방법에 따른 응답률의 효과를 살펴보기 위해서 <표 2-5>와 같이 실험기간을 나눈 후에 해당 기간의 자료수집방법(우편 및 전화)별 완성률¹⁰⁾을 제시하였다.

첫째, 사람들은 우편과 전화 중 어느 방법으로 응답하는 것을 선호하였는가? 모든 단계에서 우편응답과 전화응답 대안을 동일하게 제공한 집단5의 완성률을 분해해 보면 69.3%인 전체 완성률 중 5.6%가 전화, 63.7%가 우편에 의한 것으로 응답자들은 전화보다는 우편으로 응답하는 것을 선호함을 알 수 있다.

둘째, 그렇다면 과연 전화조사를 응답방법의 대안으로 제공한 것이 응답자들의 협조를 증가시켰는가? 통제집단인 집단1과 우편접촉의 모든 단계에서 전화응답 대안을 제공한 집단5의 완성률을 비교해 보면, 통제집단의 완성률이 집단5에 비해서 높기는 하나 그 차이는 매우 미미하였다(70.6% 대 69.3%). 다음으로 집단1과 집단2의 비교를 통해서 전화응답 대안이 응답을 증가시켰는지는 살펴보았다. 교체조사표 발송 전까지 집단2에만 감사/독촉엽서에 전화응답을 대안으로 제공하였다는 점을 제외하고 두 집단에 처치된 모든 절차는 동일하였다. 비교결과, 두 집단의 완성률 차이(58.2% 대 59.3%)는 표집오차 내에서의 차이로 전화응답 대안이 응답자의 협조를 이끌어내는데 유의미한 효과를 미친다는 결론을 내리기는 어려웠다. 집단2(59.3%)와 집단5(57.4%)를 비교해 보면 사전편지와 조사표를 보냈을 때에만 전화응답 대안을 제공한 집단5의 효

10) 응답률은 그 계산방식에 따라서 완성률, 협조율 등 여러가지 용어로 사용되나 본 연구에서는 얼마나 많은 가구가 응답을 했는지가 관심대상이므로 각 용어가 포함하고 있는 구체적인 특성에 대한 것은 논외로 하고자 한다.

과를 분리해 낼 수 있다. 마찬가지로, 이 기간 동안에 집단5가 집단2에 비해서 완성률이 높지 않은 것으로 나타나 전화조사에 대한 응답기회의 제공이 응답을 고양시키지 않았음을 알 수 있다(Dillman et al., 1994).

셋째, 전화응답이 과연 우편응답을 대체하였는가? 결론부터 말하자면 그렇다. 전화조사에 의해 완성률이 증가한 만큼 우편조사에 의한 완성률은 감소하였다. 집단1과 집단5를 비교해 보면, 우편조사에 의해서 얻은 완성률은 각각 70.6%와 63.7%이었으나 집단5에서 전화조사에 의한 완성률이 6% 가량으로 나타나 전체적인 완성률에서는 유의미한 차이가 발생하지 않았다(70.6% vs. 69.3%). 시작일부터 추적편지 발송 전까지의 기간 동안 집단5와 집단2, 3, 4를 비교해보더라도 역시 전체적인 완성률에 있어서는 유의미한 차이가 없었다. 교체조사표 발송부터 마감 직전까지의 기간 동안 집단1과 집단4, 5의 비교에서도 마찬가지로의 결과를 보여줌으로써 전화응답이 우편응답을 대체하고 있다고 볼 수 있다(Dillman et al., 1994).

〈표 2-5〉 자료수집방법별 완성률 비교

(단위: %)

구분	집단1	집단2	집단3	집단4	집단5
시작 ~ 추적편지 발송 전					
우편조사	20.3	21.6	19.8	18.8	17.0
전화조사	0.0	0.5	0.4	0.3	2.7
합계	20.3	22.1	20.2	19.2	19.7
추적편지 발송 후 ~ 교체조사표 발송 전					
우편조사	37.9	36.1	38.9	38.8	35.7
전화조사	0.0	1.1	3.6	4.0	2.0
합계	37.9	37.2	42.5	42.8	37.7
교체조사표 발송 후 ~ 교체조사표 마감 전					
우편조사	12.4	3.3	3.2	9.3	11.0
전화조사	0.0	0.1	0.1	0.9	0.8
합계	12.4	3.4	3.3	10.1	11.8
시작 ~ 마감					
우편조사	70.6	61.1	61.9	66.7	63.7
전화조사	0.0	1.7	4.1	5.2	5.6
합계	70.6	62.7	66.0	72.2	69.3

자료: Dillman et al., 1994.

주: 라운딩 에러로 합계가 맞지 않을 수도 있음.

주목할 점은 대안적인 응답방법을 제안한 횟수에 따라서 완성률이 달라진다는 것이었다. 전화응답 대안이 한 번 제공되었을 경우(집단2)에는 1.7%, 그 횟수가 두 번(집단3), 세 번(집단4), 네 번(집단5)으로 증가할 경우에 4.1%, 5.2%, 5.6%로 완성률이 증가하였다. 또한 조사 시작부터 교체조사표 발송 전까지의 기간 동안에 추적편지가 완성률을 증가시켰는지를 집단2와 집단3, 4를 비교해 보면 추적편지에 의한 완성률이 2.8% 증가했으며, 이 중 0.2%는 우편에 의한 것이고 나머지 2.6%는 전화에 의한 것이었다.

이 실험을 통해서 다음과 같은 점이 발견되었다. 자료수집방법의 선호도는 전화보다 우편에서 높게 나타났으며 그 차이는 큰 편이었다. 그러나 추적편지를 보냈을 경우에 우편보다는 전화에 의한 추가응답이 더 높았던 점으로 미루어 볼 때 전화조사가 잠재적으로는 응답률을 향상해 줄 수 있을 것으로 생각해 볼 수 있다.

나) 자료수집방법과 항목무응답률

항목무응답률은 해당 항목에 응답하지 않은 사례수의 비율로 측정되었으며, 6개의 인구항목과 8개의 주택항목으로 나누어서 분석이 진행되었다. 그 결과를 보면, 인구항목의 대부분은 전화조사에 비해서 우편조사에서 수집된 자료에서 항목무응답률이 일관적으로 낮게 나타났다. 전화의 경우 1% 혹은 그 이하의 항목무응답률이 나타난 반면 우편의 경우 특정 항목에서는 9%에 가까운 항목무응답률이 나타났다.

주택항목에서도 인구항목과 마찬가지로 우편조사보다는 전화조사에서 항목무응답률이 낮았다. 그러나, 집세와 자산가치 항목의 무응답률에서는 자료수집방법 간에 유의미한 차이가 나타나지 않았다. 이는 두 항목 모두 재산과 관련된 민감한 질문으로 응답을 꺼렸거나 혹은 응답자가 가구주가 아닐 경우에 대답하기 어려운 항목이었기 때문으로 추정되었다.

분석결과를 종합해 보면, 전화는 우편보다 더 완전한 응답을 얻어내는 것으로 나타났다. 그러나, 이 실험의 경우 우편으로 응답할지, 전화로 응답할지가 응답자들의 자발적인 선택에 달려 있었으므로 자료수집방법

에 따라서 응답자 집단의 특성에서 차이가 있을 수 있다는 한계를 갖는다. 예를 들면, 기존의 자료수집방법에는 응답을 하지 않았으나 대안으로 제안한 전화조사에 응답을 했다는 것은 전화응답자가 우편응답자에 비해서 좀 더 동기화되었을 경향이 있으며 이에 따라서 평균응답자보다 좀 더 성의껏 응답했을 가능성이 있다는 것이다.¹¹⁾

다. 2000년 센서스: RMIE

1) 목적

자료수집방법에도 컴퓨터 기술이 보급되면서 인터넷조사, CATI, IVR 등 새로운 방법이 출현하게 되었다. 이에 따라 센서스국은 2000년 센서스에서 다음과 같은 목적 하에 컴퓨터 관련 자료수집방법에 대한 실험(RMIE; Response Mode and Incentive Experiment)을 진행하였다. 첫째, 인센티브가 이 자료수집방법의 활용을 증진시키는가? 둘째, 컴퓨터 관련 자료수집방법을 제공할 경우 응답자들이 이 방법을 얼마나 선호하는가? 셋째, 이 방법을 사용할 경우 자료의 품질은 향상되는가? (Schneider et al., 2002)

2) 방법론

RMIE를 하기 위해서 센서스 2000의 전수조사 가구 중 35,376 가구를 확률추출하였다. 이 표본 가구는 세 가지 대안적인 자료수집방법과 인센티브 유무에 따라서 여섯 가지 실험집단과 하나의 통제집단에 무작위로 할당되었다. 우편조사 이외에 컴퓨터를 보조적으로 활용하는 자료수집방법의 대안은 인터넷조사, CATI, IVR이다. 실험집단의 반에는 인센티브 조건을 제시하였으며, 이 조건 하에 있는 가구가 컴퓨터 관련 방법으로 응답을 할 경우 30분용 국내 전화카드를 받았다. 표본은 MTMT에서와 마찬가지로 응답률의 높고 낮음에 따라서 두 개로 층화되었다.

11) 전화응답자와 우편응답자의 특성을 비교해 본 결과, 전체적인 응답자의 연령분포는 방법들 간에 유사한 것으로 나타났다. 그러나, LRA와 HRA 층별 분석을 해 본 결과, LRA에서 전화조사 응답자가 우편조사에 비해서 연령의 양극단(19-29, 70세 이상)에 위치한 것으로 나타났다(Bates, 1993).

센서스와 동일한 절차로 실험을 진행하기 위해서 표본가구에 센서스 2000 전수조사표와 커버레터를 2000년 3월 13일에 발송하였다. 여기에는 전통적인 방법으로써 종이조사표를 완성한 후 우편으로 회신할 수 있다는 것과 그 대안으로 컴퓨터를 보조적으로 활용한 방법으로도 응답할 수 있다는 내용이 함께 포함되어 있다. 이 중 집단1과 집단4에는 CATI로 연결되는 무료전화번호를, 집단2와 집단5에는 IVR로 연결되는 무료전화번호를 기재하였으며, 집단3과 집단6에는 인터넷조사에 대한 응답방법을 소개하였다. 그리고 집단4, 5, 6에는 컴퓨터와 관련된 방법으로 응답을 할 경우에 전화카드를 인센티브로 제공한다는 것을 명시하였다. 실험집단과 함께 센서스통제집단을 구성하였는데 이 집단에는 컴퓨터 관련 자료수집방법에 대한 것뿐 아니라 인센티브에 대한 어떠한 처치도 실행하지 않았다([그림 2-1]).



자료: U.S. Census, 2004.

[그림 2-1] 2000년 센서스: RMIE 설계

그런데 자료수집결과, 자료의 대표성과 관련하여 한계가 드러났다. 집단2에서 응답률이 낮게 나타났을 뿐 아니라 이 집단에서 우편배달이 불가능한 주소(UAA, Undeliverable As Addressed)가 다른 집단에 비해

서 현저하게 높았다. 따라서 집단2가 분석과정에서 인위적인 효과(artificial effect)를 발생시킬 가능성이 있다는 판단 하에 다음과 같은 두 가지 분석전략을 채택하였다. 첫째는 집단2와 집단5를 제외한 네 집단으로만 분석을 실시하는 것이다. 집단5는 요인설계(factorial design)의 균형을 유지하기 위해서 함께 제외되었다. 또 다른 전략은 여섯 집단을 모두 분석에 포함시키되, UAA 비율이 높은 문제 지역을 제외한 후 분석하는 것이다. 그러나, 이 분석결과는 전국수준으로 일반화할 수 없다는 한계를 갖는다. 따라서, 본 연구에서는 집단2와 집단5를 제외한 분석결과를 중심으로 소개하되, 자료수집방법별 항목무응답률 결과는 두 집단을 포함시킨 후 분석한 결과를 살펴보도록 하겠다.

3) 분석결과

가) 자료수집방법과 응답률

응답률은 전체 응답률(overall response rate)과 할당방법 응답률(assigned mode response rates)의 두 가지로 측정되었다. 전자는 자료수집 방법과 무관하게 각 집단 내에서 완성된 전체 조사표 수로 계산된 응답률을 말하며, 후자는 우편조사를 제외한 대안적인 방법에 의해서만 완성된 조사표 수로 계산된 응답률을 말한다.

(1) 인센티브가 응답률에 미치는 효과

인센티브는 전체 응답률과 할당방법 응답률의 향상에 영향을 미쳤는가? 인센티브를 제공받은 집단과 그렇지 않은 집단의 응답률을 비교해보면 후자에서 응답률이 더 높은 것으로 나타났다(71.7% 대 73.7%). 로지스틱회귀분석을 통해서 응답방법 및 층화지역을 통제한 후 인센티브가 전체 응답에 영향을 미치는지를 본 결과, 인센티브의 효과가 유의미한 것으로 나타났다. 그 방향은 앞서와 마찬가지로 인센티브를 받지 않은 집단이 인센티브를 받은 집단에 비해서 응답할 가능성이 더 높았다.

할당방법 응답률의 경우에는 인센티브를 제공받은 집단이 그렇지 않은 집단에 비해서 응답률이 높았다(16.5% 대 2.7%). 이 효과를 자료수집방법에 따라서 구분해 본 결과, 인센티브 유무에 따른 CATI 응답

률은 각각 17.7%와 1.5%이며, 인터넷 응답률은 각각 15.4%와 4.0%로 인센티브가 대안적인 방법으로 응답하도록 유인하는 경향이 있었으며 방법별로는 인터넷보다는 CATI에서 더 많이 증가시켰다. 앞서와 마찬가지로 응답방법과 층화지역을 통제한 후 인센티브가 할당방법 응답 여부에 미치는 효과에 대한 로지스틱 회귀분석을 실시해 본 결과, 인센티브는 응답률이 낮은 층과 높은 층 모두에서 할당방법 응답가능성을 증가시켰다. 즉, 인센티브가 CATI 혹은 인터넷을 사용하여 응답하도록 유인하는 경향이 있었으며, 그것은 인터넷보다는 CATI에서 더욱 효과적이었다. 그러나, 대안적인 방법에 의한 응답률의 증가는 우편에 의한 응답률 감소에 의해서 상쇄되었다. 전체 응답률에서는 오히려 인센티브를 제공하지 않은 집단에서 더 높게 나타났음이 이를 증명해준다.

여기에는 여러가지 설명이 가능한데 그 중에서도 가장 설득력이 있는 것은 대안적인 방법으로 응답을 하고자 했던 응답자들 중 일부는 첫 번째 시도에서 응답을 하지 못했으며 이후에 조사를 하기 위해 다시 CATI나 인터넷으로 되돌아오지 않았다는 것이다. 또 다른 설명은 인센티브로 제공한 전화카드를 활성화시키는 방법을 몰랐던 응답자들에게 오히려 인센티브가 조사동기를 감소시키는 데 기여했을 수도 있다는 것이다. 한편, 인센티브효과가 CATI와 인터넷에서 차별적으로 나타난 것은 CATI가 인터넷에 비해서 좀 더 쉽게 접속이 가능했기 때문으로 추정되었다.

(2) 자료수집방법과 응답률

CATI와 인터넷을 대안적인 방법으로 제공한 것이 응답률에 어떠한 영향을 미쳤는가? 만일 영향을 미쳤다면, 그 영향은 인센티브조건과 층화조건에서 동일하였는가? 먼저, 전체 응답률에 대한 응답방법 효과를 보도록 하자. 통제집단과 비교해 보면, 인센티브를 제공하지 않은 CATI 집단(73.5%)과 인터넷집단(73.9%)의 전체 응답률은 통제집단(71.4%)보다 높은 것으로 나타났다.

그렇다면 대안적인 응답집단에서 전체 응답률이 통제집단에 비해 높게 나온 것이 우편조사에 의한 것인가, 대안적인 방법에 의한 것인가? 이를 위해서 인센티브를 제공하지 않은 CATI 집단과 인터넷 집단에서

우편조사에 의한 응답률과 통제집단(우편조사) 응답률을 비교해 보았다. 그 결과, CATI 집단에서의 우편조사 응답률(72.0%)은 통제집단의 응답률(71.4%)과 통계적으로 유의미한 차이를 보이지 않은 반면에 인터넷 집단에서의 우편조사응답률(69.6%)은 통제집단의 응답률(71.4%)에 비해서 유의미하게 낮다는 것을 알 수 있었다. 비록 그 차이가 작기는 하지만 인터넷과 CATI 모두 우편조사에 비해 높은 응답률을 얻도록 하는 효과를 갖는 것으로 나타났음을 보여주었다.

할당방법 응답률에 영향을 미치는 응답방법의 효과를 보면, CATI집단(집단1과 집단4: 9.6%)과 인터넷집단(집단3과 집단6: 9.8%) 사이에 통계적으로 유의미한 차이를 보이지는 않았다. 그러나 이것은 인센티브 제공 여부에 따라서 차이가 발생하였다. 인센티브를 제공하였을 경우에는 CATI집단의 응답률은 인터넷집단보다 높았던 반면에 인센티브를 제공하지 않았을 경우에는 인터넷집단의 응답률이 CATI집단보다 높았다. 이를 통해서 인센티브가 인터넷 보다는 CATI를 사용하도록 유인하는 경향이 있음을 알 수 있었다.

나) 자료수집방법과 항목무응답률

기존의 연구에 따르면, 컴퓨터를 이용한 자료수집방법이 종이조사 표에 의한 자료수집방법보다 더 나은 자료품질을 제공하였다. 이것은 전자의 방법에 일반적으로 자동스킵 및 내검프로그램이 장착되어 있기 때문이다. 이 실험에서 사용된 컴퓨터 관련 방법의 경우에도 인터넷을 제외하고 CATI와 ASQ방법에 이 기능을 추가하였다.

가장 높은 항목무응답률을 보인 자료수집방법은 IVR로, 그 비율을 보면 11.8%에 달하였다. 다음으로 우편조사, 인터넷조사, CATI 순으로 항목무응답률이 낮았다. IVR에서 항목무응답률이 높게 나타난 이유는 응답자들이 조사가 완료되기 전에 끊어버리거나, 컴퓨터 프로그램에 할당된 시간 내에서 응답을 하지 못했기 때문으로 분석되었다. 조사항목별로 보면 가구소유, 히스패닉/인종과 같은 특정 항목에서 무응답률이 우편조사에서 높게 나타나고 대안적인 방법에서 항목무응답률이 낮게 나타났는데 이것은 CATI의 자동화된 시스템 때문인 것으로 해석되었다.¹²⁾

실험결과를 종합해 보면 대안적인 방법으로서의 인터넷조사나 CATI가 응답률 및 항목무응답률에 어느 정도 긍정적인 효과를 보이고 있는 것으로 나타났다. 그러나 이 방법을 도입함에 있어서는 오차뿐 아니라 비용을 고려해야 한다. CATI나 인터넷조사 모두 자료수집의 특정 부분 예컨대, 조사표의 인쇄나 우편발송 및 회수, 자료의 캡처 등과 같은 부분에서 비용을 줄여줄 수 있다. 다만, 변동성이 강한 이러한 비용과 함께 소프트웨어와 하드웨어에 대한 고정비용을 동시에 고려해야 비용효과성에 대한 적절한 평가가 가능할 것이다.

라. 2003년 센서스 전국 시험조사

1) 목적

2003년 전국시험조사에서 역시 자료수집방법의 다양화를 검토하였다. 우편조사에 추가적으로, 조사원의 도움 없이 표본대상이 스스로 응답할 수 있는 방법(self-response mode)을 제공함으로써 이것이 과연 응답자의 협조율과 자료품질을 증가시키는지 검토하였다. 이 실험은 응답자 기입식의 방법을 통해서 응답률을 증가시킬 수 있는 최상의 전략을 찾음으로써 궁극적으로 무응답추적조사의 업무량을 감소시키고자 하는 목적을 가지고 있다(Brady et al., 2003; Treat, 2005).

2) 방법론

우편조사만으로 수행되는 통제집단과 함께 대안적인 응답방법과 그 응답방법을 선택함에 있어서 강제조건이 부여되었는지의 여부에 따라서 5개 실험집단이 구성되었다. 우편에 의한 응답방법과 함께 대안적인 응답방법으로 인터넷 혹은 IVR을 제공하였다. 그리고 이 방법으로 응답하

12) 그렇다면 왜 인터넷조사에서 자동스킵 시스템이 없음에도 불구하고 항목무응답률이 낮은가? 이에 대한 분석은 실시되지 않았지만 MTMT 결과에서 유추해 보면, 인터넷조사 응답자가 우편조사 응답자에 비해서 좀 더 동기화되어 있었을 가능성이 있다. 즉, 인터넷 응답자의 경우 자발적인 응답자일 가능성이 높아 좀 더 성실히 응답했을 수 있다.

는 것을 자유롭게 선택할 수 있는지(choice) 혹은 반드시 그 방법으로 응답을 해야 하는지(push)를 다시 구분하였다(<표 2-6>). 강제집단의 경우 처음에는 종이조사표를 받지 않는 대신에 인터넷 혹은 IVR 응답방법에 대한 지침만을 제공받은 반면에 선택집단은 종이조사표와 함께 인터넷과 IVR 지침을 동시에 제공받았다.

<표 2-6> 2003년 센서스 전국 시험조사 실험 집단

구분	집단1	집단2	집단3	집단4	집단5
응답방법	IVR	IVR	인터넷	IVR+ 인터넷	IVR+ 인터넷
강제조건	강제	선택	선택	강제	선택

실험대상은 2000년 센서스 Mailout/Mailback에 위치한 가구에서 무작위로 추출되었다. 통제집단에는 20,000가구가, 실험집단 각각에는 10,000가구가 배분되었다. 이전의 실험들과 마찬가지로 표집에 앞서 센서스 트랙은 2000년 센서스의 우편응답률에 따라서 응답률이 높은 지역과 낮은 지역으로 층화되었다.

접촉전략은 실험집단에 따라서 차별적으로 이루어졌다. 통제집단과 실험집단 모두 첫 단계에서 조사를 예고하는 사전편지를 받았다. 조사표 패키지는 다음 단계에서 보냈는데, 이 때 통제집단, 선택집단, 강제집단에 따라서 패키지 안에는 다른 내용이 포함되었다. 통제집단 혹은 선택집단 중 하나는 종이조사표를 받았으며, 실험집단의 처치 내용에 따라서 선택집단에 보내진 조사표 패키지에는 웹사이트 주소나 IVR 전화번호 혹은 두 가지 방법 모두에 대한 지침이 포함되어 있다. 그리고 두 강제집단에는 IVR 혹은 인터넷과 IVR 중 하나로 센서스 조사표를 완성하라는 지침이 제공되었다. 세 번째는 독촉엽서 단계이다. 만일, 이미 응답을 했다면 감사의 엽서를, 아직 응답하지 않았다면 응답을 독려하는 내용을 담고 있는 엽서를 받게 되어 있다. 독촉내용을 담고 있는 엽서에는 선택과 강제 집단 모두에 대해서 두 번째 우편에서 보냈던 내용과 같은 내용의 응답방법으로 응답을 하도록 독려하였다. 마지막으로 응답을 하지 않은 모든 가구에 교체조사표를 보냈다. 이 경우, 통제집단과 선택집단에는 처음에 보낸 조사표 패키지와 같은 것을 보낸 반면에

강제집단에는 종이조사표를 함께 보내기는 하나 가능한 최초에 할당받은 방법으로 응답하기를 독려하였다.

3) 분석결과

가) 자료수집방법과 응답률

대안적인 응답방법에 따라서 협조율에 차이가 있는가? 분석결과(<표 2-7>), 강제조건을 부여한 집단1과 집단4가 통제집단에 비해서 협조율이 낮게 나타났는데 이것은 전국수준뿐 아니라 층화지역에서도 마찬가지였다. 즉, IVR에 응답하도록 하거나 또는 IVR 혹은 인터넷에 응답하도록 하는 것이 강제사항일 경우에 통제집단에 비해 각각 4.9%와 5.7% 감소하였다. 반면에 선택조건을 부여받은 집단2, 3, 5의 경우에는 통제집단과의 협조율 차이가 전국수준이나 층화지역 모두에서 유의미한 차이를 보여주지 않았다. 응답방법을 선택할 수 있는 기회를 제공하는 것이 응답을 하도록 유인하기보다는 응답방법들 사이에서 이동하도록 만드는 경향이 있는 것으로 나타났다. 즉, 대안적인 응답방법에 대한 선택기회가 협조율을 증가시키지도, 그리고 감소시키지도 않았다. 표에는 제시하지 않았으나, 이동의 폭을 보면 집단2의 경우 전체 협조율 65.9% 중 61.4%가 우편으로 응답을 한 경우이며, IVR로 응답을 한 경우는 3.7%, 기타의 경우는 0.8%이었다. 집단3의 경우에는 67.3% 중 60.6%가 우편, 6.7%가 인터넷이었다. 집단5는 66.4%의 협조율 중 58.7%가 우편, 4.7%가 인터넷, 그리고 나머지가 IVR 및 기타였다.

<표 2-7> 통제집단과 실험집단의 협조율 차이 비교

구분	전국		HRC		LRC	
	차이	90% C.I	차이	90% C.I	차이	90% C.I
집단1-통제	-4.9*	-6.61 ~3.19	-4.1*	-6.02 ~ - 2.18	-7.1*	-8.99 ~-5.21
집단2-통제	-1.2	-2.80 ~0.40	-1.6	-3.45 ~ 0.25	-0.2	-2.27 ~1.87
집단3-통제	0.2	-1.40 ~1.80	0.8	-0.89 ~ 2.49	-1.8	-3.97 ~0.37
집단4-통제	-5.7*	-7.50 ~3.90	-4.9*	-6.95 ~ - 2.85	-8.1*	-10.06 ~-6.14
집단5-통제	-0.7	-2.41 ~1.01	-0.4	-2.34 ~ - 1.54	-1.7	-3.77 ~0.37

자료: Treat, 2005.

그렇다면 강제집단과 선택집단의 협조율은 차이가 있는가(<표 2-8>). 강제집단의 IVR(집단1)은 선택집단의 IVR(집단2)에 비해서 전국수준뿐 아니라 두 층화지역에서도 모두 협조율이 낮게 나타났으며, 강제집단의 IVR 및 인터넷(집단4)은 동일한 응답방법에 대한 선택집단(집단5)보다 협조율이 낮았다. 종합적으로, 우편으로 응답을 할 것인지와 IVR 및 인터넷으로 응답할 것인지 선택할 수 있는 집단은 IVR 및 인터넷으로만 응답하도록 되어있는 강제집단보다 협조율이 유의미하게 더 높다고 볼 수 있다.

<표 2-8> 강제집단과 선택집단의 협조율 차이 비교

구분	전국		HRC		LRC	
	차이	90% C.I	차이	90% C.I	차이	90% C.I
M1-M2	-3.7*	-5.07~-2.33	-2.5*	-4.21~-0.79	-6.9*	-8.80~-5.00
M4-M5	-5.1*	-6.75~-3.45	-4.5*	-6.64~-2.36	-6.4*	-8.34~-4.46

자료: Treat, 2005.

추가적으로, 모든 응답자들이 실험의 전 기간 동안에 우편, IVR, 인터넷의 세 가지 방법에 대해서 동일하게 응답할 확률을 갖고 있는 집단5에 대해서 응답자들의 인구학적 특성을 분석해 보았다. 그 결과, 남성이 IVR(41%)보다는 우편(61.2%)이나 인터넷(73.4%)에 더 많이 응답하였으며, 세입자가 우편(25.4%)보다 IVR(35.3%)에 더 많이 응답하였다. 또한, 젊은 응답자들이 인터넷을 더 사용하는 경향이 있었다. 25~44세 연령그룹 중에 인터넷(55.0%)이 우편(31.3%)이나 IVR(34.1%)에 비해 높았으며, 65세 이상에서는 인터넷(7.0%)보다는 우편(26.8%)이나 IVR(28.7%)이 높았다.

나) 자료수집방법과 항목무응답률

항목무응답률은 5개의 인구항목(가구주와의 관계, 성별, 연령, 히스패닉/인종)과 2개의 가구항목(보유기간, 인구집계)으로 나누어 전국수준에서 분석되었다. 층화지역에 따른 분석은 예상대로 응답률이 낮은 층에서의 항목무응답이 그렇지 않은 층에 비해서 높게 나타났다. 자료수

집방법에 따른 항목무응답률을 보면, 인구수준의 모든 항목 중에서 우편조사보다 IVR에서 항목무응답률이 유의미하게 높은 것으로 나타났다. 이것은 항목 그 자체보다는 IVR 시스템을 사용하는 것이 어려웠기 때문으로 추정되었다.

〈표 2-9〉 자료수집방법별 항목무응답률 비교

항목	우편 (통계집단)	우편 (집단1-집단5)	IVR	인터넷	기타
관계	2.1(0.17)	2.4(0.16)	12.8(0.85)*	0.5(0.13)*	1.9(0.40)
성별	1.8(0.12)	1.9(0.10)	8.0(0.55)*	1.2(0.15)*	1.5(0.28)
연령	1.7(0.12)	1.9(0.11)	8.4(0.55)*	1.5(0.24)	1.5(0.31)
히스패닉	3.8(0.17)	4.0(0.15)	9.0(0.56)*	1.5(0.16)*	1.4(0.26)*
인종	3.8(0.20)	4.1(0.16)	8.7(0.58)*	1.5(0.17)*	1.5(0.28)*
보유기간	2.0(0.14)	1.6(0.09)*	0.8(0.13)*	0.2(0.09)*	0.4(0.21)*
인구집계	2.7(0.17)	1.1(0.08)*	0.0(0.0)*	0.0(0.0)*	0.0(0.0)*

자료: Treat, 2005.

주1: 관계항목에서 첫 번째 가구원은 제외하였음.

주2: *, p<0.10임.

한편, 거의 모든 항목에서 인터넷조사에 의한 항목무응답률이 통계 집단의 우편조사에 의한 것보다 유의미하게 낮게 나타났다. 특히, 히스패닉과 인종에 대해서 통계집단의 항목무응답률이 인터넷조사에서의 항목무응답률보다 2배 이상 높았는데, 이것은 부분적으로 인터넷조사의 어플리케이션에 있는 내검절차에서 기인한 것이다(〈표 2-10〉).

마. 2000년~2001년 미국지역사회조사의 인터넷조사방법 실험

1) 목적

미국지역사회조사는 우편조사, 전화조사, 대면조사 방법을 사용하여 자료를 수집하고 있는데, 여기에 인터넷조사를 추가적인 대안으로 제공하고자 하였다. 이것은 우편에 응답하지 않는 사람들 중 인터넷 응답을 선호하는 사람이 응답을 할 경우 전반적으로 응답률이 향상될 것

이라는 기대에 따른 것이다. 설령, 우편 응답자를 인터넷 응답자로 이 동시시킬 가능성도 배제할 수 없으나 이 경우 또한 비용절감으로 이어질 수 있다는 장점을 갖는다. 인터넷조사는 또한 자료의 품질을 높여줄 수도 있다. 종이조사표와 달리 인터넷조사표에서는 스킵 패턴이나 응답 범주 등을 통제함으로써 일관된 응답을 얻어낼 수 있는 가능성이 높기 때문이다. 또한 도움말을 통해서 왜 질문을 하는지에 대한 자세한 지침을 줄 수 있기도 하다(Griffin et al., 2001).

2) 방법론

센서스국은 미국지역사회조사의 정기표본(regular production sample)과는 별도의 표본을 구성하여 이 실험을 수행하였다. 기존의 조사표본에 이 대안을 제공할 경우 오차가 발생할 가능성이 있기 때문이다. 대신에 정기표본과 동일한 절차에 따라서 자료를 수집하였으며, 다만 실험집단의 경우 최초의 조사표 패키지에 인터넷 응답 방법에 관한 정보와 지침서를 함께 동봉하였다는 차이만을 두었다.

3) 분석결과

가) 자료수집방법과 응답률

인터넷조사를 대안으로 제공하는 것이 과연 응답률¹³⁾을 향상시켰는가? 통제집단과 실험집단의 응답률을 비교해 보면 우편조사에 의해서만 수행된 통제집단의 응답률은 43.6%, 우편조사와 인터넷조사 두 가지 대안을 제시한 실험집단의 응답률은 37.8%이었다(<표 2-10>). 실험집단의 응답률을 분해해 보면, 전체 37.8% 중 약 2%만이 인터넷으로 응답한 것으로 나타났다. 통제집단과 실험집단의 우편 응답률의 차이

13) 응답률은 정기표본의 무응답자 추적조사가 시작되기 이전에 도착한 조사표에 대해서만 계산되었다. 정기표본의 경우 우편조사 이후 무응답가구에 대한 CATI 추적조사의 목적으로 전화접촉을 시도하는데 이것이 우편응답률에 정적인 효과를 주는 것으로 발견되었다. 전화접촉 그 자체가 응답자에게 독촉의 기능을 함으로써 응답자들이 조사표를 작성하여 우편으로 보내게 만드는 것이다. 이 실험에서는 CATI 추적은 계획되지 않았기 때문에 응답률을 정확하게 비교하기 위해서 정기표본의 무응답자 추적조사 이전까지의 응답률만 계산하였다.

가 8%인 것을 감안한다면, 우편 응답률의 차이가 단지 인터넷으로의 이동에 의해서만 설명되는 것은 아님을 알 수 있다.

〈표 2-10〉 통제집단과 실험집단의 응답률 비교

통제집단	실험집단	
43.6%	37.8%	
	우편	인터넷
	35.6%	2.2%

인터넷 응답률은 기존 실험과 달리 낮게 나타났다. 인터넷 응답률이 2000년의 경우 인센티브 조건에 따라서 많게는 15.4%, 적게는 4.0%이었으며, 2003년의 경우 우편과 인터넷 선택조건에서는 6.7%, 우편과 인터넷, 그리고 IVR 선택조건에서는 4.7%이었던 점을 볼 때 이번 실험의 인터넷 응답률이 2%인 것은 낮은 수준임을 알 수 있다.

그렇다면, 인터넷 응답률이 낮은 이유는 무엇인가?¹⁴⁾ 인터넷 응답자들의 응답행동을 분석한 결과, 집단별로 응답자의 행동 패턴에서 차이가 발견되는데, 첫 주 이후 통제집단의 응답률이 실험집단의 응답률에 비해서 급격히 높아졌다. 다음으로, 실험집단에서 인터넷 응답자, 실험집단에서 우편응답자, 통제집단에서 우편응답자를 비교해 본 결과, 인터넷 응답자의 응답은 주로 초기에 이루어진 반면에 통제 및 실험 집단에서의 우편 응답은 2주 이후에 급격하게 증가하는 패턴을 보였다. 이것은 인터넷으로 초대받은 응답자가 바로 응답을 하지 않을 경우 이후에 인터넷으로 응답할 가능성이 적으며 우편으로 응답할 가능성이 높다는 것을 시사하는 것으로 해석되었다.

14) 이에 대한 첫 번째 가설은 인터넷조사를 중도에 포기한 응답자가 우편조사로 되돌아 오지 않았다는 것이다. 두 번째 가설은 인터넷 응답을 대안으로 제공한 것이 어떤 응답자들에게는 당장에 응답하고자 하는 것을 제쳐두게 하는 원인이 됨으로써 응답과정을 깨고 무응답을 촉발하였다는 것이다. 세 번째 가설은 인터넷 응답이 가능하다는 지침 그 자체가 어떤 응답자들에게는 프라이버시 보호에 대한 의문을 제기하도록 만든다는 것이다. 그러나 인터넷 접속률 자체가 낮은 상태에서 이 가설들을 모두 검증하는 것은 어려우며 두 번째 가설에 초점을 맞추어서 자료를 분석하였다.

나) 자료수집방법과 내검실패율

응답률 이외에 내검실패율을 통해서 자료수집방법별로 품질을 평가해 보았다(<표 2-11>). 미국지역사회조사에서는 회수된 조사표에 대해서 무응답 혹은 일관적이지 않은 응답이 있는지를 식별해내기 위해 자동내검을 실시한다. 내검과정에서 문제가 있는 것으로 나타난 비율을 통제집단 대 실험집단, 우편조사 대 인터넷조사별로 비교해 보았다.

〈표 2-11〉 실험집단별 내검실패율 비교

(단위: %)

구분	내검실패율	표준오차
통제집단	33.2	0.16
실험집단	32.1	0.77
실험(인터넷)집단	11.3	2.46
실험(우편)집단	33.2	0.80
통제집단-실험집단	1.2*	0.79
통제집단-실험(인터넷)집단	21.9*	0.55

주: *, $p < 0.1$ 임.

<표 2-11>을 보면 통제집단과 실험집단에서의 내검실패율의 차이는 크지는 않은 것으로 나타났다(33.2% 대 32.1%). 그러나, 실험집단이 아닌 자료수집방법별로 비교해 보면 그 차이는 확연하게 드러났다. 실험집단 내에서 우편조사와 인터넷조사에 의한 내검실패율은 각각 33.2%와 11.3%로 통계적으로 유의미한 차이가 나타났다. 이것을 볼 때 앞서 통제집단과 실험집단에서 내검 실패율의 차이가 크지 않았던 것은 실험집단에서 인터넷으로 회수된 비율이 낮았기 때문으로 볼 수 있다. 종합적으로 보면, 인터넷조사에서의 내검실패율이 우편조사에서의 내검실패율보다 낮았다. 이는 인터넷조사표의 로직과 기본정보, 조사항목별 도움 스크린 등이 응답 오차를 줄여주었기 때문으로 해석되었다. 그러나, 자발적으로 구성된 인터넷 응답자들이 우편 응답자들에 비해서 더 주의깊고, 협조적일 것이라는 가능성을 전혀 배제할 수는 없다.

지금까지 미국 센서스의 혼합자료수집방법 실험의 방법론과 그 결과에 대해서 살펴보았으며, 다음과 같은 결론이 도출되었다. 첫째, 응답자에게 다양한 응답방법을 제공하는 것이 전체적인 응답률을 증가시켜 주지는 않았다. 즉, 혼합방법조사가 응답률을 높여주기 보다는 종이조사표에 응답을 하는 기존 응답자를 대안적인 응답방법에 응답하도록 이동시키는 역할을 하고 있었다. 둘째, 종이조사표보다는 컴퓨터를 이용한 자료수집방법에서 항목무응답률이 더 낮은 것으로 나타났다. 이는 자동스킵이나 내검 등의 어플리케이션 효과에서 기인한 것으로 보인다. 셋째, 비용을 정확하게 계산하기는 어렵지만 유사한 응답률을 두고 자료의 품질이 향상된다면 대안적인 방법이 선호될 수 있는데, 이때 초기비용을 반드시 고려해야 한다.

그렇다면, 우리나라 인구주택총조사의 혼합방법조사 시험조사 결과는 어떠한가? 미국의 사례와 비교할 때 어떠한 공통점과 차이점이 있으며 이러한 특성이 발생하는 원인은 무엇인지, 그리고 이것이 갖는 함의는 어떤 것이 있는지를 살펴보겠다.

제3절 자료분석

1. 연구자료

가. 조사지역선정 및 조사항목

2010 인구주택총조사에서는 응답자들의 조사편의를 도모하고자 기존의 자료수집방법인 대면조사 외에도 인터넷 및 우편조사 등을 대안적인 방법으로 도입할 계획에 있다. 이는 다양한 응답기회를 제공함으로써 적은 비용을 들이고서도 응답을 효율적으로 얻어내기 위한 것이다(통계청, 2007). 이러한 목적 하에 제1차 시험조사(2007년)에서는 인터넷조사와 우편조사를, 제2차 시험조사(2008년)에서는 여기에 ARS 및 CATI를 추가하여 도입가능성을 검토하였다.

제2차 시험조사 지역은 해당지역의 인구 및 가구규모, 도시 특성, 조

사 난이도 등을 고려하여 부산광역시, 경기도, 강원도를 선정하였으며, 각 지역에서 2개 읍면동을 임의로 추출하였다. 부산광역시에서는 동래구 명장1동과 부산진구 범천1동이, 경기도에서는 광주시 광남동과 용인시 기흥동이, 강원도에서는 동해시 묵호동과 인제군 인제읍이 선정되었다. 선정된 조사지역은 확률추출을 통해 조사지역의 90%에는 전수조사를, 10%에는 표본조사를 실시하였다.

각 지역에 대한 구체적인 특성을 2005년 인구주택총조사 결과를 통해서 살펴보면(<표 2-12>), 평균연령이 가장 낮은 지역은 용인시 기흥동(32.4세)이며, 가장 높은 지역은 동해시 묵호동(47.6세)이었다. 성비를 보면, 용인시 기흥동의 남성비율이 가장 높으며, 동해시 묵호동의 여성비율이 가장 높았다.

<표 2-12> 제2차 시험조사지역의 주요 특성

지역	인구수 (명)	가구수 (가구)	주택수 (호)	평균연령 (세)	성비 (%)	
전국	47,041,434	15,988,274	13,222,641	35.6	99.5	
부 산	부산진구 범천1동	10,118	4,009	3,254	37.9	96.9
	동래구 명장1동	18,963	6,064	4,171	36.8	96.5
경 기	광주시 광남동	28,383	8,431	8,053	33.4	103.2
	용인시 기흥동	9,632	2,688	2,974	32.4	132.2
강 원	동해시 묵호동	3,956	1,690	1,884	47.6	94.0
	인제군 인제읍	9,213	3,121	3,016	37.0	108.1

자료: 통계청 홈페이지(www.kosis.kr).

주: 인구수는 외국인을 제외한 수치임.

조사기준일은 4월 14일 0시이며, 조사지역 내에 거주하는 모든 사람과 이들이 살고 있는 주택에 대해 조사하였다. 조사항목은 인구, 가구, 주택으로 구분되었다. 전수조사항목을 중심으로 살펴보면, 인구항목은 성명, 성별, 나이, 가구주와의 관계, 교육정도, 혼인상태 등이며, 가구항목은 가구구분, 사용방수, 주거시설형태, 건물 및 거주층, 점유형태, 타지주택소유 및 주인가구 여부 등, 주택항목은 거처의 종류, 주거용연면적, 총방수, 건축연도, 편익시설수 등이다(통계청, 2008b).

조사결과 최종 분석대상은 총 28,082가구이다. 이는 조사과정에서 빈 집 등의 이유로 조사가 적격하지 않은 것으로 판명된 243가구와 내검과정에서 통합되어야 하는 것으로 나타난 5가구를 제외한 것이다. 28,082가구 중 27,626가구에서 조사를 완료하였으며, 이 중 전수조사는 25,283가구, 표본조사는 2,799가구이다.

나. 자료수집방법

2005 인구주택총조사의 기본적인 자료수집방법이 조사원이 가구를 방문하여 면접을 하는 대면조사였다면, 시험조사에서는 이 방법과 함께 인터넷조사, 우편조사, ARS, CATI 등이 대안으로 제시되었다. 이 방법을 비방문조사로 범주화하였으며, 이와 대비하여 대면조사를 방문조사로 정의하였다. 제2차 시험조사 자료수집은 비방문조사를 먼저 한 후에 무응답자를 대상으로 방문조사를 하는 순으로 진행되었다(<표 2-13>).

비방문조사(1단계)는 준비조사기간을 포함하여 12일간(4월 11일~22일) 실시되었다. 인터넷조사는 인구주택총조사 홈페이지에서 실명인증을 받고 기본정보사항을 입력한 후 조사표를 입력하면 된다. 우편조사는 조사원이 준비조사 기간 중에 배부한 조사표를 작성하여 우편으로 응답하는 것인데, 아파트와 같은 공동주택의 경우 관리사무소 또는 경비실에 비치된 회수함을 이용하였다. 전화조사 시스템으로 직접 전화를 걸어 응답하는 ARS와 CATI 방식 또한 동시에 채택되었다. 비방문조사가 완료되면 4월 23일과 24일 동안 비방문조사현황을 정리하여 응답/미응답가구명부를 정리하고, 조사원별로 방문조사 업무량을 배정하였다.

비방문조사 시 응답하지 않은 가구를 대상으로는 2주간(4월 25일~5월 8일) 조사원이 직접 가구를 방문하여 대면조사를 실시하였다(2단계). 이 경우 응답자가 스스로 기입을 원할 경우에는 조사표를 배포하고 나서 추후에 조사표를 회수하는 배포조사 방법을 동시에 채택하였다.

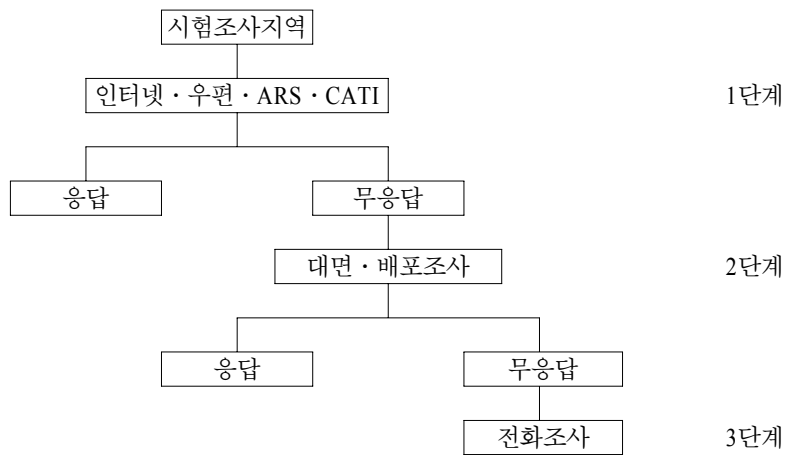
대면조사를 하기 위해서 2번 이상 방문하였음에도 불구하고 면접을 완료하지 못한 가구에는 조사원이 직접 전화를 걸어서 방문시기를 확인하거나 조사할 수 있게 하였다(3단계).¹⁵⁾ 추후 분석과정에서는 조사원

15) 방문조사기간에도 응답자의 선호에 따라서 인터넷이나 우편조사에 의한 응답이 이루어

전화조사 단계를 최종무응답 추적조사단계로 분류하였다.

조사를 마친 조사표는 조사원과 조사관리자가 조사표의 내용을 검토하여, 조사항목의 누락을 확인·보완하였다(통계청, 2008b).

〈표 2-13〉 제2차 시험조사 자료수집방법 단계



한편, 제2차 시험조사에서는 대안적으로 도입된 자료수집방법의 참여율을 높이고자 인센티브를 제공하였다. 인센티브는 추후에 추첨을 통해서 경품을 제공하는 것이며, 이러한 사실을 사전에 홍보했는지의 여부에 따라서 홍보를 한 A지역과 홍보를 하지 않은 대조군 B지역으로 나누었다. 인센티브 사전홍보지역은 명장1동, 광남동, 인제읍이다. 경품추첨은 시기를 나누어 지방자치단체와 통계청에서 차별적으로 실시하였다. 비방문조사의 중간시점(4월 17일)과 최종시점(4월 24일)을 기준으로 추첨을 실시하되, 중간시점에는 A지역에서만, 최종시점에는 A와 B지역 모두에서 경품을 추첨하였다. 한편, 각 지방자치단체에서는 두 시기 모두, 통계청에서는 최종 시기에 한 번만 경품을 추첨하였다(통계청, 2008c).

여기기는 하였으나 그 사례는 매우 적은 편이다.

2. 변수측정 및 분석방법

가. 변수정의 및 측정

자료수집방법이 자료의 품질에 미치는 효과를 분석하기 위해 자료의 품질과 관련된 지표로서 응답률, 내검불일치율, 항목무응답률을 측정하였다.

1) 응답률

미국여론조사협회(AAPOR, American Association for Public Opinion Research)(2006)에 의하면 응답률은 계산방법에 따라서 여섯 가지로 구분되는데, 이 연구에서는 조사원이 가구에 방문한 결과 빈집으로 판명된 것을 제외하고 전체 표본 가구 중에서 면접을 완료한 조사 가구가 차지하는 비율을 응답률로 정의하였다.

$$R = \frac{I}{(I + P) + (R + NC + O)}$$

여기에서 R은 응답률(Response Rate), P는 부분면접(Partial Interview), I는 완성된 면접(Complete Interview), R은 거부(Refusal), NC는 비접촉(Non-contact), O는 기타(Other)를 뜻한다.

응답률은 단계별로 구분하여 측정하였다. 1단계는 비방문조사를 통해서 얻은 응답률, 2단계는 방문조사를 통해서 얻은 응답률, 3단계는 최종무응답 추적조사를 통해서 얻은 응답률이다. 한편, 자료수집단계를 고려하지 않고 무응답 추적조사 이전까지의 응답률을 전체 응답률로 정의하였다.

2) 내검불일치율

자료입력이 완료되면 자체 내용검토 프로그램을 이용하여 자료의 비일관성을 체크한 후 그에 따라서 자료를 수정하여 최종 데이터베이스를 구축하게 되는데, 이 연구에서는 내검프로그램을 수행하기 전과 후의

자료를 비교하여 전체 레코드 수 중에서 일치하지 않는 레코드 수가 차지하는 비율을 내검불일치율로 측정하였다.

$$\text{내검불일치율} = \frac{\text{특정 항목의 불일치 레코드 수}}{\text{전체 레코드 수}} \times 100$$

인구, 가구, 주택항목으로 구분하여 내검불일치율을 살펴보았으며, 전체 레코드수는 각각의 항목에 대해서 인구수, 가구수, 주택수를 의미한다. 내검불일치율은 전수조사항목에 대해서만 계산되었다.

3) 항목무응답률

항목무응답률은 특정항목에서 응답하지 않은 비율로 정의되며, 전체 레코드 중에서 해당 항목의 무응답 레코드수가 차지하는 비율로 측정되었다.

$$\text{항목무응답률} = \frac{\text{특정 항목의 무응답 레코드 수}}{\text{전체 레코드 수}} \times 100$$

내검불일치율과 마찬가지로 전수항목에 대해서 인구, 가구, 주택항목을 구분하여 항목무응답률을 계산하였다. 항목무응답률은 전수조사항목에 대해서만 계산되었다.

4) 자료수집방법

자료수집방법은 응답자와의 접촉정도·조사원의 참여정도·컴퓨터 보조정도 등의 세 가지 측면에 따라서 분류될 수 있다(Biemer and Lyberg, 2003). 이 기준에 따라 제2차 시험조사에서 사용된 자료수집방법을 분류해 보면 <표 2-14>와 같다.

응답자와의 접촉이 가장 많고 자료수집자의 참여가 가장 높은 방법으로 대면조사, 응답자와는 간접적으로 접촉하나 비교적 조사원의 참여가 높은 전화조사, 그리고 응답자와는 간접적으로 접촉하며 조사원의

참여 또한 낮은 우편조사 및 인터넷조사, 배포조사¹⁶⁾로 구분될 수 있다. 전화조사는 컴퓨터를 활용하느냐 하지 않느냐에 따라서 CATI와 ARS, 그리고 조사원에 의한 전화조사로 구분된다. ARS의 경우에는 CATI나 조사원 전화조사와 달리, 조사원의 참여가 낮은 방법으로 분류될 수 있다. 한편, 우편조사의 경우 회수방법에 따라서 우편배달시스템에 의한 회수와 우편회수함에 의한 조사원 회수로 구분된다.

〈표 2-14〉 제2차 시험조사 자료수집방법 종류

구분	조사원의 높은 참여		조사원의 낮은 참여	
	종이	컴퓨터	종이	컴퓨터
응답자와 직접 접촉	대면조사	-	-	-
응답자와 간접 접촉	전화조사	CATI	배포조사	인터넷조사 ARS
			우편조사	

자료: Biemer and Lyberg, 2003 수정.

5) 통제변수

자료수집방법에 따른 오차를 검토하기 위해서는 종속변수에 영향을 미칠 수 있는 다양한 요인을 통제할 필요가 있다(Biemer, 1988). 단일 표본에서 응답자들의 선호에 따라서 자료수집방법을 선택하게 한 이번 시험조사에서 요인의 통제는 더욱 중요하다. 왜냐하면, 자료수집방법에 따라서 응답자의 특성이 선별적이기 때문이다(Bates, 1993). 예컨대 전화조사를 선호하는 응답자의 특성과 인터넷조사를 선호하는 응답자의 특성은 차이가 있을 것이다. 그러므로 순수한 방법효과를 살펴보기 위해서는 응답자의 특성을 통제한 후 살펴보아야 한다.

응답률, 항목무응답률, 내검불일치율 중에서 지금까지 상대적으로 많은 연구가 진행되어온 분야는 응답률이다. 여기에서는 이에 대한 연구결과를 중심으로 검토해 보았다. 문헌연구 결과 응답률에 영향을 미치는

16) 조사원이 가구에 방문했으나 응답자가 스스로 기입을 원할 경우, 가구에 조사표를 배포한 후 추후에 회수하는 방법을 허용하였다.

변수들로는 사회적인 분위기, 조사 설계상의 속성, 표본의 속성, 조사원의 특성 등 여러가지가 있다. 자료의 한계상 응답자와 관련한 변수들을 중심으로 살펴보면, 응답자 수준에서는 일반적으로 가구주의 인구사회학적 특성으로 성별, 연령, 건강상태, 소득변수가 고려되었다. 인과관계를 확신할 수는 없으나 그 관계가 응답자의 심리상태에 영향을 미칠 것으로 간주되는 변수로는 가구구조 및 가구특성, 가구원 수 등 가구특성과 관련된 변수가 있었다(Groves, Cialdini, and Couper, 1992). 이와 함께 최근에는 지역 수준 변수 또한 응답률에 영향을 미친다는 결과가 있다(Johnson et al., 2006).

이 연구에서는 응답자 수준과 지역 수준을 모두 고려하되, 응답자 수준은 응답자의 개인적인 특성과 가구 특성을 구분하였다. 기존의 연구에서 응답자의 개인적인 특성 대신에 가구주 특성이 측정된 것과 비교할 때, 좀 더 정확한 분석을 할 수 있을 것으로 기대된다. 응답자의 개인적인 특성으로는 성별, 연령, 교육수준, 혼인상태 변수를 고려하였다. 성별은 여성인 경우는 1, 남성인 경우는 0으로 더미변수처리하였으며, 연령은 나이를 그대로 변수화하였다. 교육수준은 졸업여부와 무관하게 미취학은 1, 초등학교는 2, 중학교는 3, 고등학교는 4, 4년제 미만 대학은 5, 4년제 대학이상은 6, 대학원 석사과정은 7, 대학원 박사과정은 8로 코딩하였다. 혼인상태는 배우자가 있는 경우는 1, 없는 경우는 0으로 처리하였다. 가구의 특성으로는 가구내 여성비율, 가구원의 평균연령, 가구원의 평균교육정도를 고려하였다. 이와 함께 일인가구여부, 25~64세 인구비율, 아파트거주여부, 자가소유여부, 비농림어가여부를 측정하였으며 이 변수들은 모두 일인가구인 경우를 1, 아파트에 거주하는 경우를 1, 자가소유인 경우를 1, 비농림어가인 경우를 1로 하고 그렇지 않은 경우를 0으로 처리하였다.

지역 수준 변수로는 가구 수준 변수를 지역 수준에서 평균한 것을 이용하였다. 이 때, 지역의 범위는 조사구 수준으로 하였다. 조사구란 전국의 모든 지역에 대해서 식별이 명확한 지형지물을 기준으로 지도상에서 일정한 가구수가 포함되도록 분할한 조사담당구역으로 평균 60가구로 구성되어 있다(통계청, 2008b).¹⁷⁾ 이와 함께, Sampson과 동료들(1997)이 개념화한 집합적 효능감(collective efficacy)을 측정하였다. 이것은 공동선

(common good)을 위한 활동에 참여하고자 하는 시민의 의지로 정의되는데, 사회조사에 대한 참여 또한 지역에 대한 관여의 형태로 해석될 수 있기 때문이다(Couper, Singer, and Kulka, 1998, Johnson et al., 2006에서 재인용). 여러가지 측정변수 중에서 Johnson과 동료들(2006)의 연구에서 유의미한 것으로 나온 것은 다음과 같다.

동일한 주소에서 지난 5년간 거주한 인구의 비율로 측정된 거주지 안정성 지표(indicator of residential stability), 전문직 혹은 관리직 직업을 가진 성인의 비율로 측정된 부의 집중도 지표(indicator of concentrated affluence), 빈곤선 이하의 인구비율로 측정된 취약함의 집중도 지표(indicator of concentrated disadvantage), 인구밀도로 측정된 도시화의 지표(indicator of urbanicity)이다. 본 연구에서는 비이동자 비율과 함께 거주기간 변수를 함께 사용하였으며, 빈곤선 이하 대신에 중하위소득가구의 비율을, 부의 집중도에 대한 대리지표로 임금근로자 비율을 측정하였다. 또한 인터넷조사 비율의 중요한 예측지표로 기대되는 컴퓨터 보유비율을 추가하여 측정하였다. 그러나, 인구밀도를 제외한 집합적 효능감 지표와 컴퓨터 보유비율은 표본조사 항목에만 있는 변수이기 때문에 실제 분석상에서는 제약이 발생하였다.

나. 자료분석방법

연구문제를 설명하기 위해서 다음과 같은 자료분석방법을 이용하였다. 첫째, 자료수집방법에 따라서 응답률, 내검불일치율, 항목무응답률에 차이가 있는지를 교차분석 및 분산분석을 통해서 살펴보았으며, 차이가 있는 경우에 한해, 집단 간 차이를 보기 위해 사페 검정(scheffe test)을 추가하였다. 둘째, 응답자 및 지역 수준에서의 특성을 통제한 이후에도 이러한 효과가 지속되고 있는지는 로지스틱 회귀분석을 통해서 살펴보았다. 통계분석프로그램은 SPSS 15.0을 이용하였다.

17) 인구주택총조사에서 조사구는 아파트조사구, 보통조사구, 섬조사구, 기숙시설조사구, 특수사회시설조사구, 관광호텔 및 외국인 거주지역 조사구로 구분된다. 이 중 제2차 시험조사에서는 아파트조사구와 보통조사구만을 대상으로 하였다.

3. 분석결과

분석결과는 첫째, 자료수집방법과 응답률, 둘째, 자료수집방법과 내검불일치율 및 항목무응답률, 셋째, 자료수집방법별 응답자 특성 순으로 제시하였다. 응답률에서는 인센티브와 응답률, 자료수집방법과 응답률을 분석하였는데, 인센티브의 경우 응답률에 영향을 미치는 중요한 실험조건이므로 자료수집방법과 응답률을 분석함에 있어서 고려되어야 한다. 분석결과는 전수조사와 표본조사를 구분하되, 두 조사에서 큰 차이가 없을 경우에는 전수조사 결과 위주로 설명하였다.

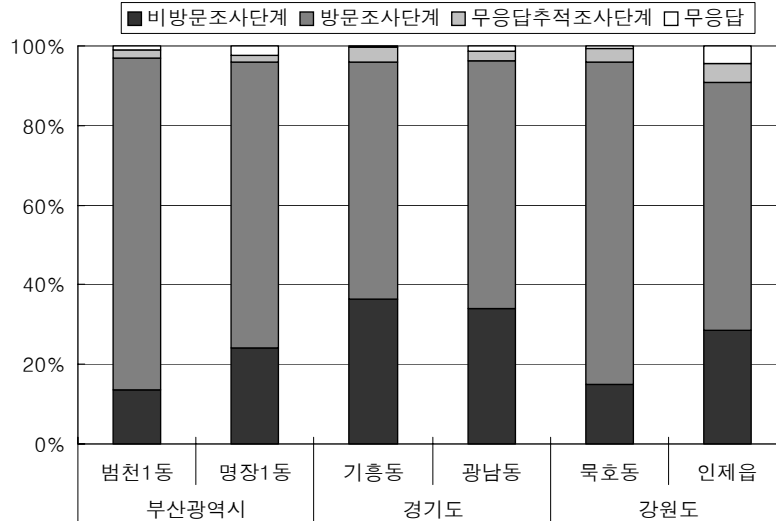
가. 응답률

제2차 시험조사 지역의 비방문조사에 의한 응답률은 27.0%이며 방문조사에 의한 응답률은 68.8%로 전체 응답률이 95.8%로 나타났다(<표 2-15>). 방문조사 기간에 부재 등으로 인해서 조사를 완료하지 못한 무응답 가구에 대해 실시한 전화조사를 통해 2.6% 추가 응답을 얻었다. 응답률은 전수조사인지, 표본조사인지에 따라서 차이가 있는데 전수조사가 표본조사에 비해서 비방문조사 응답률은 높았으나, 방문조사 응답률은 낮았다($\chi^2=21.16$, $p<0.00$).

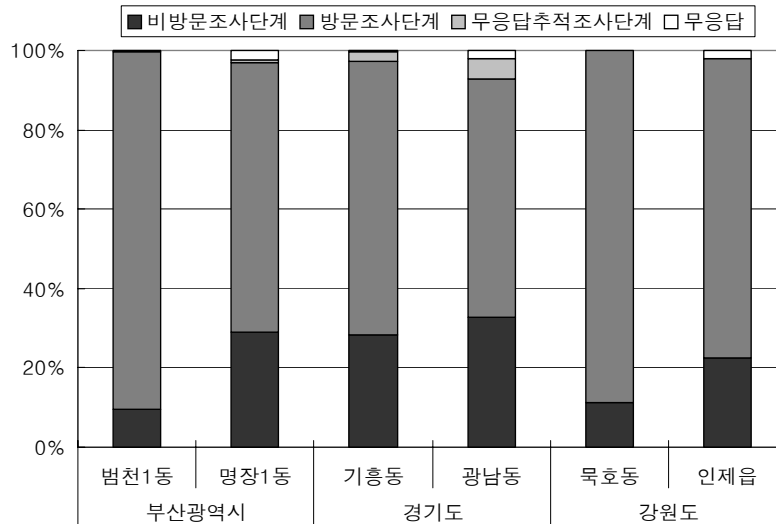
<표 2-15> 제2차 시험조사 단계별 응답 및 무응답 분포

(단위: 가구, %)

구분	응답			무응답	전체
	비방문조사	방문조사	무응답추적조사		
전체	7,575 (27.0)	19,323 (68.8)	728 (2.6)	456 (1.6)	28,082 (100.0)
전수조사	6,894 (27.3)	17,296 (68.4)	675 (2.7)	418 (1.7)	25,283 (100.0)
표본조사	681 (24.3)	2,027 (72.4)	53 (1.9)	38 (1.4)	2,799 (100.0)



[그림 2-2] 전수조사의 지역별 응답 및 무응답 분포



[그림 2-3] 표본조사의 지역별 응답 및 무응답 분포

응답률은 시험조사 지역별로도 차이가 있었다([그림 2-2], [그림 2-3]). 먼저, 전수조사 응답률을 보면 다른 지역에 비해서 비방문조사 응답률이 높은 곳은 기흥동(36.6%)과 광남동(34.2%)이었으며, 비방문조사 응답률이 낮은 곳은 범천1동(13.59%)과 목호동(14.8%)이었다. 이러한 지역별 차이는 통계적으로 유의미한 것으로 나타났으며($\chi^2=1195.74$, $p<0.00$), 이는 표본조사에서도 마찬가지였다($\chi^2=227.03$, $p<0.00$). 전반적인 경향을 보면, 비방문조사 응답률이 높은 곳은 방문조사 응답률이 낮고, 비방문조사 응답률이 낮은 곳은 방문조사 응답률이 높았다.

1) 인센티브와 응답률

제2차 시험조사에서는 비방문조사에서 응답한 가구를 대상으로 추첨을 통해서 경품을 제공하였다. 이 때 경품제공을 사전에 홍보한 경우와 그렇지 않은 경우의 두 집단으로 구분한 후 홍보지역에서는 비방문조사 중간단계와 최종단계에서 모두 추첨을 하였으며, 미홍보지역에서는 최종단계에서만 추첨하였다. 과연 인센티브를 홍보한 경우는 그렇지 않은 경우에 비해서 응답률이 향상되었는가? 구체적으로 그 효과는 경품추첨여부에 따라서 달라졌는가? 만일 차이가 있다면 어느 자료수집방법에서 효과적이었는가?

인센티브는 응답률을 향상시켰는가?

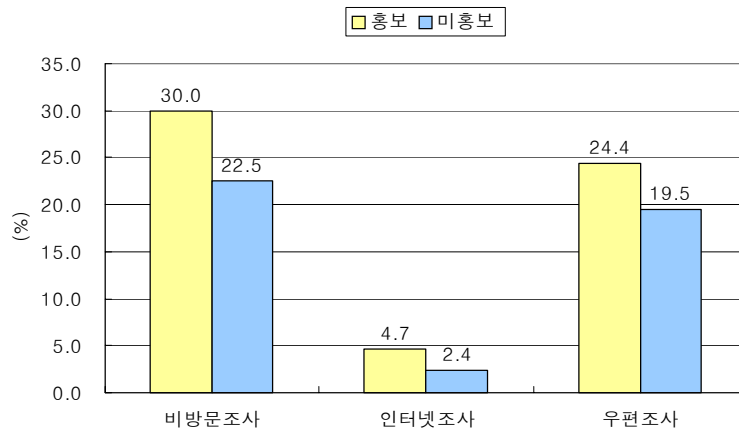
먼저, 무응답추적조사 이전까지의 전체 응답률을 인센티브 홍보여부에 따라서 살펴보았다(<표 2-16>).

<표 2-16> 인센티브 홍보여부별 응답 및 무응답 분포

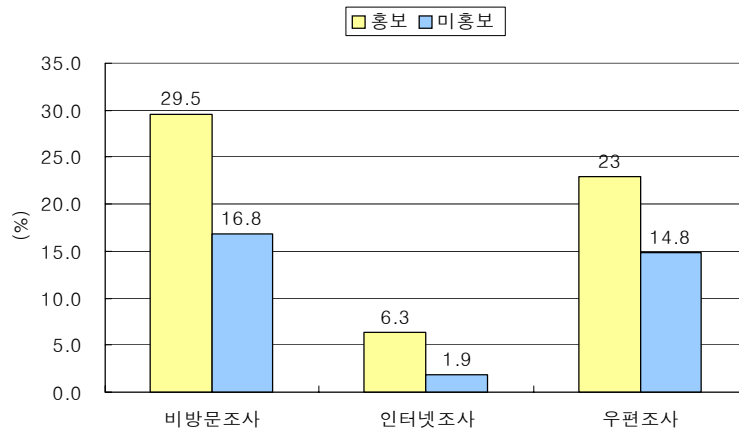
(단위: 가구, %)

	전수		표본	
	홍보	미홍보	홍보	미홍보
응답	15,242 (95.3)	8,948 (96.4)	1,576 (95.2)	1,132 (99.0)
무응답	758 (4.7)	335 (3.6)	79 (4.8)	12 (1.0)
전체	16,000 (100.0)	9,283 (100.0)	1,655 (100.0)	1,144 (100.0)

전수조사의 경우 홍보지역의 응답률은 95.3%인 반면, 미홍보지역의 응답률은 96.4%로 홍보지역 응답률이 더 낮게 나타났다($\chi^2=18.10$, $p<0.00$). 표본조사에서 역시 홍보지역의 응답률은 95.2%인 반면에 미홍보지역의 응답률은 99.0%로 미홍보지역에서의 응답률이 더 높았다($\chi^2=29.83$, $p<0.00$). 그러나, 전체 응답률에 대한 인센티브의 부적인 효과는 자료수집방법을 고려할 경우, 비방문조사에 대해서는 정적인 효과를 미치는 것으로 분석되었다.



[그림 2-4] 인센티브 홍보여부에 따른 자료수집방법별 응답률: 전수



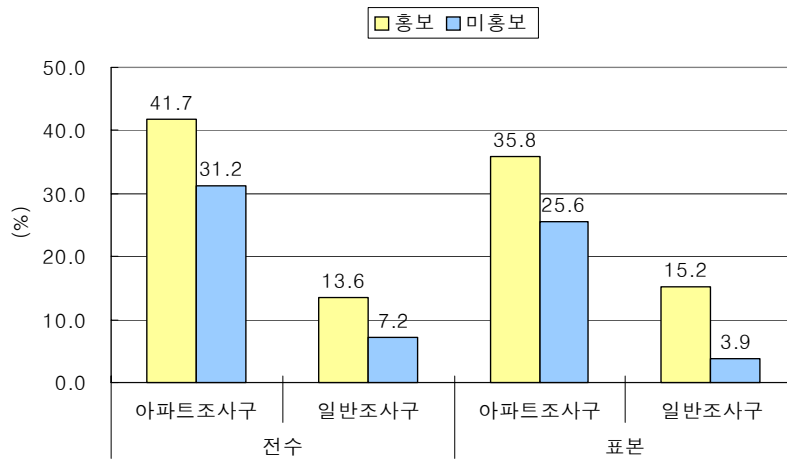
[그림 2-5] 인센티브 홍보여부에 따른 자료수집방법별 응답률: 표본

[그림 2-4]와 [그림 2-5]는 전수조사와 표본조사 각각에 대해서 자료 수집방법을 고려한 후의 홍보여부에 따른 응답률을 제시한 것이다. 비방문조사의 경우 홍보지역에서의 응답률이 미홍보지역에 비해서 7.5%p ($\chi^2=164.83$, $p<0.00$) 높았다. 구체적으로, 인터넷조사의 경우에는 2.3%p($\chi^2=77.32$, $p<0.00$), 우편조사의 경우에는 4.9%p 더 높은 응답률을 보여주었다($\chi^2=81.04$, $p<0.00$). 표본조사의 경우에도 마찬가지로 홍보지역에서의 비방문조사, 인터넷조사, 우편조사 응답률이 미홍보지역에 비해서 높게 나타났다. 비방문조사의 경우에는 12.7%p($\chi^2=59.86$, $p<0.00$), 인터넷조사의 경우에는 4.4%p($\chi^2=30.49$, $p<0.00$), 우편조사의 경우에는 8.2%p($\chi^2=29.07$, $p<0.00$) 더 높게 나타났다. 전수조사와 표본조사를 비교해 볼 경우에는 인터넷조사나 우편조사 모두 홍보지역과 미홍보지역에서의 응답률의 차이가 표본조사가 전수조사에 비해 약 2배가량 더 높은 것으로 나타나 표본조사 지역에서의 인센티브 홍보효과가 눈에 띄었다.

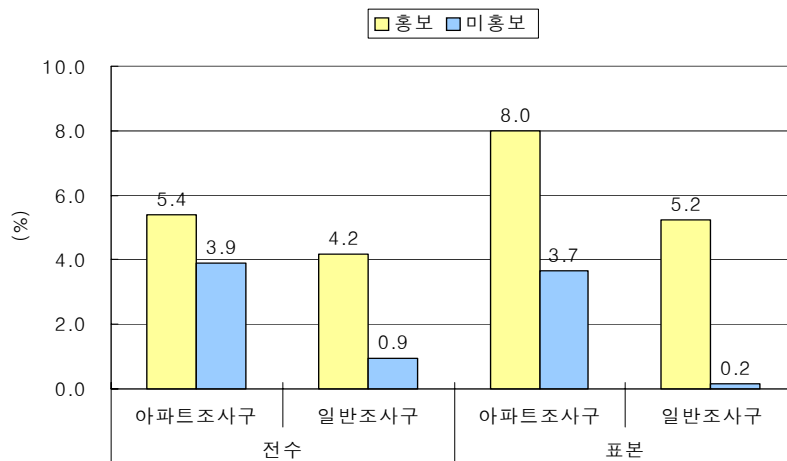
한편, 우편조사의 경우에는 일반조사구와 우편조사구로 구분해서 홍보효과를 살펴볼 필요가 있다. 왜냐하면, 우편조사의 회수방식은 전통적인 방법으로써 우편배달에 의한 것과 아파트 조사구를 중심으로 설치한 우편회수함에 의한 것으로 구분되는데 이는 응답자의 접근가능성에 차별적으로 작용했을 것으로 보이기 때문이다.¹⁸⁾ 먼저 전수조사를 살펴보면, 아파트 조사구 중 인센티브 홍보지역의 우편조사 응답률이 미홍보지역에 비해서 10.5%p 더 높았으며($\chi^2=125.79$, $p<0.00$), 일반조사구의 홍보지역은 같은 조사구의 미홍보지역에 비해서 우편조사 응답률이 6.4%p 가량 더 높았다($\chi^2=122.77$, $p<0.00$). 표본조사의 경우에는 아파트조사구 중 인센티브 홍보지역의 우편조사 응답률이 미홍보지역에 비해서 10.2%p 더 높았으며($\chi^2=14.80$, $p<0.00$), 일반조사구에서는 홍보지역의 응답률이 미홍보지역보다 11.3%p 더 높았다($\chi^2=47.72$, $p<0.00$). 전수조사나 표본조사 모두 홍보효과가 아파트 조사구에서 약 10%p 가량 높게 나타났는데, 이

18) 우편회수함을 설치한 경우라도 응답자의 선호에 따라서 우편배달방식을 택할 수도 있으므로 엄밀히 말해서 아파트조사구라고 해서 모두 우편회수함을 통해서 조사표를 회수하였다고 단정지를 수는 없다. 마찬가지로 일반조사구의 아파트에서도 회수함이 설치되었을 수 있다. 다만, 아파트조사구인 경우가 일반조사구인 경우에 비해서 회수함에 의한 조사표 수거의 양이 많을 것이라는 가정 하에 이후의 분석이 진행되었다.

는 전수조사의 일반조사구(6.4%p)보다는 높으나, 표본조사의 일반조사구(12.3%p)보다는 약간 낮은 수치이다. 한편, 인터넷조사의 경우에는 유사한 경향이 발견되었으나 일반조사구와 아파트조사구에서의 홍보효과 차이가 우편조사보다는 크지 않아 아파트조사구의 우편조사에서 홍보효과가 더 크다는 것을 간접적으로 지지해주었다(그림 2-6).



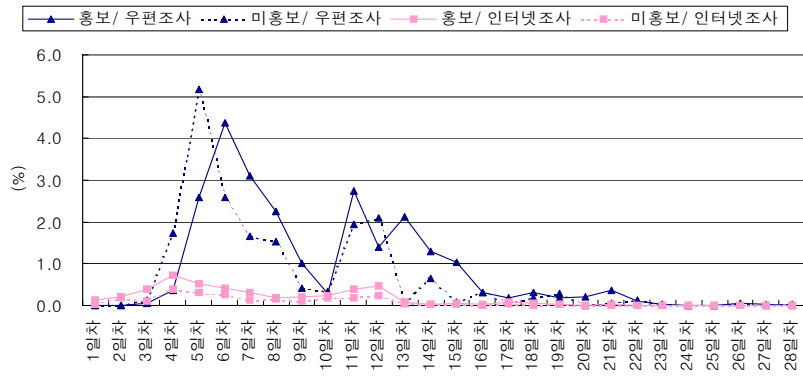
[그림 2-6] 인센티브 홍보여부에 따른 조사구유형별 우편조사 응답률



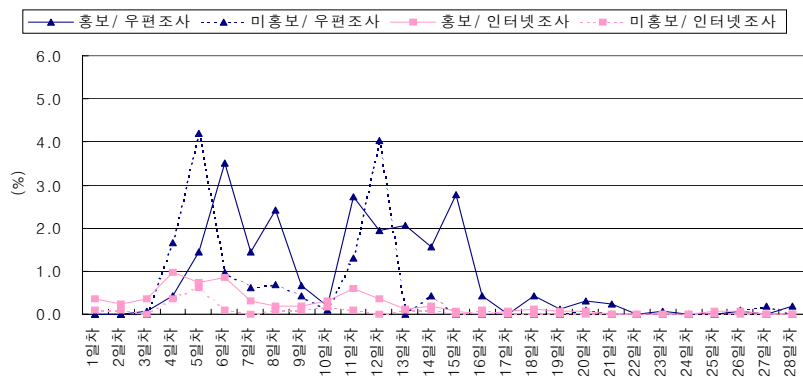
[그림 2-7] 인센티브 홍보여부에 따른 조사구유형별 인터넷조사 응답률

경품추첨은 응답률을 향상시켰는가?

경품추첨에 의한 효과를 살펴보기 위해서 [그림 2-8]과 [그림 2-9]에 전수조사와 표본조사 각각에 대한 인터넷조사와 우편조사 일일응답률을 홍보여부에 따라서 제시하였다. 그래프에서 4일차는 조사 기준일인 4월 14일이며, 7일차는 1차 경품추첨기준일인 4월 17일, 12일차는 비방문조사 완료시점인 4월 22일, 14일차는 2차 경품추첨기준일인 4월 24일을 뜻한다.



[그림 2-8] 인센티브 홍보 여부에 따른 인터넷·우편조사 일일응답률: 전수



[그림 2-9] 인센티브 홍보 여부에 따른 인터넷·우편조사 일일응답률: 표본

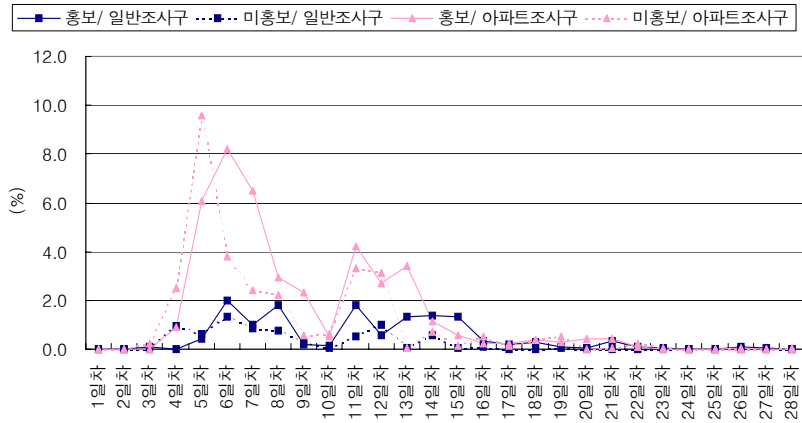
전반적으로 인터넷조사의 응답률이 우편조사에 비해서 낮은 가운데, 인터넷조사의 일일응답률이 가장 높은 지점은 조사기준일인 4일차이며 점점 하락하다가 비방문조사 완료시점인 12일차에 다시 한 번 상승 이후 하락하였다. 이러한 패턴은 홍보지역과 미홍보 지역에서 유사하였다. 우편조사의 경우에는 경품추첨과 관련해서 주목해 볼 만한 점이 있다. 먼저, 홍보지역에서 일일응답률 그래프는 M자형과 유사한데, 6일차에서 가장 높은 일일응답률을 보였으며 점차 하락하다가 11일차와 13일차에서 다시 높아졌으며, 이 지점은 경품추첨 바로 직전이라는 것을 알 수 있다.

한편, 미홍보 지역의 경우에도 유사한 시점에서 일일응답률의 상승이 있었다. 표본조사의 경우 그 양상이 조금 더 복잡하게 나타나는 것처럼 보이기는 하나, 일일응답률이 가장 높은 지점들이 경품추첨과 인접한 시기라는 공통점이 있다. 이를 통해서 경품추첨에 의한 효과는 인터넷조사보다는 우편조사에서 더 두드러지게 나타남을 알 수 있다.

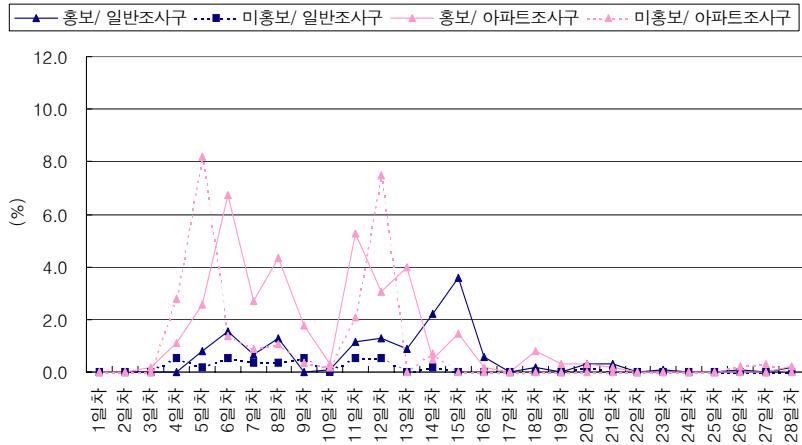
구체적으로 우편조사를 조사구유형별로 구분하여 일일응답률을 살펴보았다([그림 2-10], [그림 2-11]). 일반조사구와 아파트조사구의 홍보지역에서의 일일응답률을 비교해 보면, 일반조사구의 경우 일일응답률이 높은 지점은 6일차, 8일차, 11일차이며, 아파트조사구의 경우에는 6일차, 11일차, 13일차였다. 아파트조사구와는 달리 일반조사구의 경우, 특별히 경품추첨시점인 7일차와 14일차에 인접하여 응답률이 증가하였다는 점을 발견하기는 어려웠다. 표본조사의 경우에는 오히려 15일차에 가장 높은 일일응답률을 보이기도 하였다.

그런데, 전반적으로 볼 때 일일응답률의 증가 및 감소 패턴이 홍보지역과 미홍보 지역에서 유사하게 나타나는 것이 눈에 띈다. 특히, 1차 경품추첨시기에 추첨을 하지 않은 미홍보 지역에서도 홍보지역과 마찬가지로 경품추첨 시기 직전에 일일응답률이 높았다. 이는 우편조사 날짜가 회수날짜로 기입되었고, 우편회수가 응답자의 응답시점과 무관하게 조사원에 의한 회수에 의해서 일괄적으로 이루어졌기 때문으로 추정된다. 그러므로 경품추첨 직전에 일일응답률이 높다고 해서 반드시 경품추첨의 효과가 있다고 보기에는 무리가 있을 가능성이 있다. 따라서 경품추첨 구간별로 누적응답률을 함께 살펴보았다. 구간은 다음과 같이

나누었다. 첫 번째는 준비조사부터 1차 경품 추첨(4월 11일~4월 17일), 두 번째는 1차 경품 추첨 후부터 2차 경품 추첨(4월 18일~4월 24일), 세 번째는 2차 경품 추첨 이후부터 조사완료(4월 25일~5월 8일) 기간이다.



[그림 2-10] 조사구 유형별 우편조사 일일 응답률: 전수

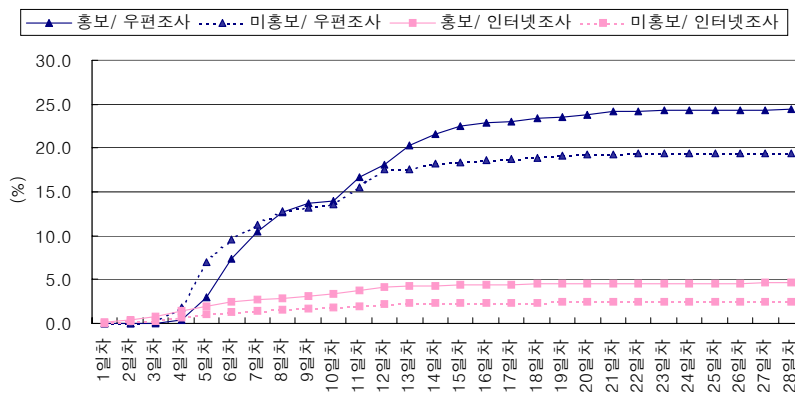


[그림 2-11] 조사구 유형별 우편조사 일일 응답률: 표본

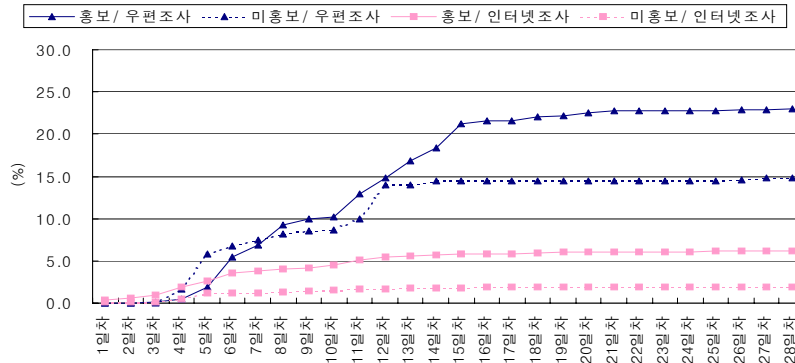
<표 2-17>을 보면, 첫 번째 시기의 홍보지역과 미홍보지역의 응답률 차이에 비해서 두 번째 시기의 홍보지역과 미홍보지역의 응답률 차이가 두드러졌다. 전수조사의 경우에는 첫 번째 시기의 응답률 차이가 0.5%p에 불과했으나, 두 번째 시기에는 4.8%p까지 높아졌으며, 같은 시기 표본조사의 경우에는 2.0%p이었던 응답률 차이가 5.9%p로 증가하였음을 알 수 있다. 이는 홍보 그 자체보다는 경품추첨에 의한 효과가 더 큼을 보여준다. 실제로 일자별 누적응답률을 보면 우편조사의 경우에는 1차 경품추첨 전까지 미홍보지역의 응답률이 홍보지역의 응답률보다 더 높게 나타나고 있다([그림 2-12]~[그림 2-15]).

<표 2-17> 경품추첨 시기별 인터넷·우편조사 응답률 증가량 (단위: %, %p)

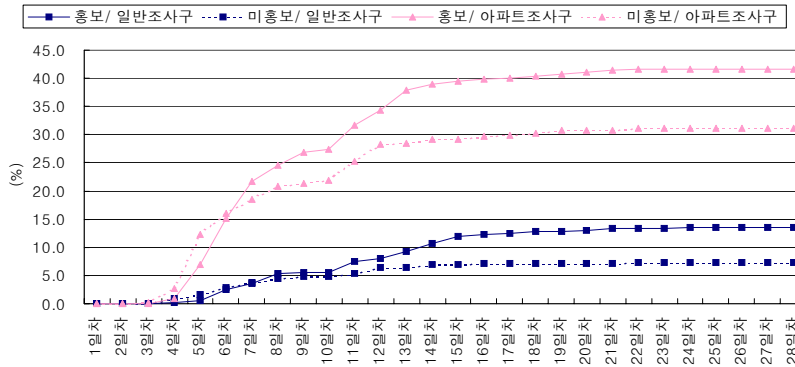
구분	전수			표본		
	홍보 (N=16,000)	미홍보 (N=9,283)	차이	홍보 (N=1,655)	미홍보 (N=1,144)	차이
4.11~4.17	13.2	12.7	0.5	10.7	8.7	2.0
1차 경품	추첨	미추첨	-	추첨	미추첨	-
4.18~4.24	12.7	7.9	4.8	13.5	7.6	5.9
2차 경품	추첨	추첨	-	추첨	추첨	-
4.25~5.8	3.2	1.3	1.9	5.1	0.4	4.7



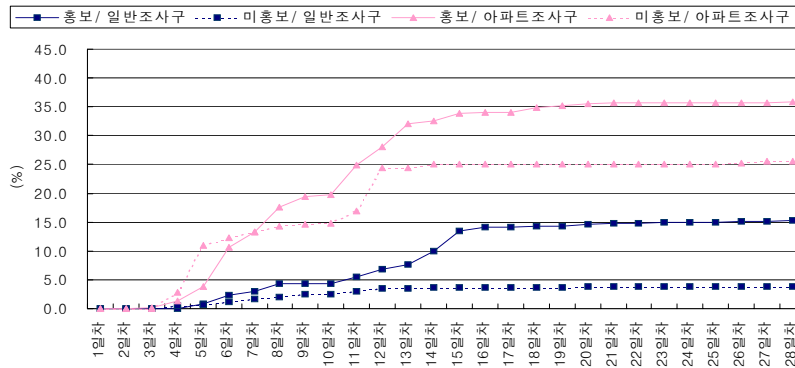
[그림 2-12] 인센티브 홍보여부에 따른 인터넷·우편조사 누적응답률: 전수



[그림 2-13] 인센티브 홍보여부에 따른 인터넷·우편조사 누적응답률: 표본



[그림 2-14] 조사구유형별 우편조사 누적응답률: 전수



[그림 2-15] 조사구유형별 우편조사 누적응답률: 표본

인센티브가 응답률에 미치는 효과에 대한 분석결과는 다음과 같이 요약될 수 있다. 첫째, 인센티브 홍보는 전체 응답률에서는 부적인 효과를 갖는 것으로 나타났으나 대안적인 응답방법으로써의 비방문조사에 대해서는 정적인 효과를 보여주었다. 둘째, 인센티브 홍보가 대안적인 방법을 선택하도록 동기화시키는 경향이 있는데, 인터넷조사보다는 우편조사를, 일반조사구에서의 우편조사보다는 아파트조사구에서의 우편조사를 더욱 독려하는 경향이 있었다. 셋째, 인센티브 홍보와 함께 실제 경품 추첨은 홍보 지역에서의 응답률 증가를 더욱 강화시켰다.

분석결과를 통해서 인센티브 제공에 대한 홍보가 인터넷이나 우편을 통해서 응답하도록 만드는 경향이 있다는 것이 어느 정도는 지지되었다. 그러나, 전체 응답률에서는 오히려 미홍보 지역의 응답률이 더 높은 것으로 나온 것에서 볼 수 있듯이, 증가된 비방문조사 응답률은 방문조사 응답률의 감소에 의해서 상쇄되어 두 가지의 순효과는 오히려 인센티브 지역에서 전체 응답률의 감소로 이어졌다.¹⁹⁾ 이에 대해서는 여러가지 설명이 가능할 수 있겠으나, 첫째는 비방문조사 응답자를 대상으로 한 인센티브가 잠재적인 비방문조사 응답자들에게 비방문조사 기간 이후에는 조사 자체를 하지 않게 만드는 경향이 있을 수 있을 것으로 추정된다. 반대로 미홍보 지역에서는 잠재적인 응답자가 비록 비방문조사는 하지 못했더라도 응답에 대한 의지가 방문조사에서 표출되었을 가능성이 있다. 또는 인센티브로 제공된 경품에서 추첨될 가능성에 대한 기대가 높지 않았기 때문에 아예 응답 자체를 포기했을 가능성이 있다. 또 한편으로는 인센티브 제공과 무관하게 조사지역 자체의 특성에서 기인했을 가능성도 배제할 수 없다. 한편, 비방문조사 중에서도 인터넷조사에 비해서 우편조사에 더 큰 효과를 보여준 것은 응답자들이 우편조사에 좀 더

19) 이러한 결과는 2000년 미국 센서스의 RMIE에서와 유사하다. 이에 대해서 센서스국의 가정은 인터넷이나 CATI와 같은 대안적인 방법에 응답하고자 하는 응답자가 첫 번째 시도에서 시스템의 문제로 조사를 완료하지 못했을 경우에 다시 되돌아오지 않기 때문일 것으로 추정하였다. 또 다른 가능성은 당시에 제공한 인센티브의 문제에 대한 것이었다. 제공한 인센티브는 30분 무료 전화카드였는데 이것은 비방문조사를 완료했을 경우에 활성화가 가능했다. 그럼에도 불구하고 인센티브에 대한 이해가 부족했던 응답자가 조사를 완료하기 전에 인센티브를 먼저 사용하려고 시도했다가 그것에 실패하자 아예 조사 자체에 대한 신뢰를 하지 못하고 포기해버리는 경우가 많다는 것이었다.

쉽게 접근할 수 있었기 때문으로 보여진다. 아파트 조사구나 일반조사구 모두에서 우편조사가 인터넷조사보다 응답률이 높게 나타난 것은 응답자가 인터넷조사에 응답하기 위해서 컴퓨터를 켜는 것보다는 조사표를 작성한 후 회수함에 넣거나 우체함에 넣는 것을 좀 더 쉽게 생각할 수 있기 때문이다. 우편조사의 경우, 일반조사구보다는 아파트 조사구에서 응답률 향상이 더 높았다는 것과 인터넷조사의 경우에는 이러한 경향이 나타나지 않았다는 점이 이를 지지해 준다.

2) 자료수집방법과 응답률

여기에서는 자료수집방법이 응답률에 미치는 효과를 중심으로 살펴 보았다. 구체적인 연구질문은 첫째, 응답자들이 가장 많이 선택한 자료수집방법은 무엇인가? 둘째, 비방문조사가 전체 응답률을 향상시키는가? 구체적으로 비방문조사 중 인터넷조사와 우편조사, 전화조사가 전체 응답률 향상에 미치는 효과는 각각 어떠한가? 셋째, 자료수집방법별 응답률을 결정하는 요인은 무엇인가?

응답자들이 가장 많이 선택한 자료수집방법은 무엇인가?

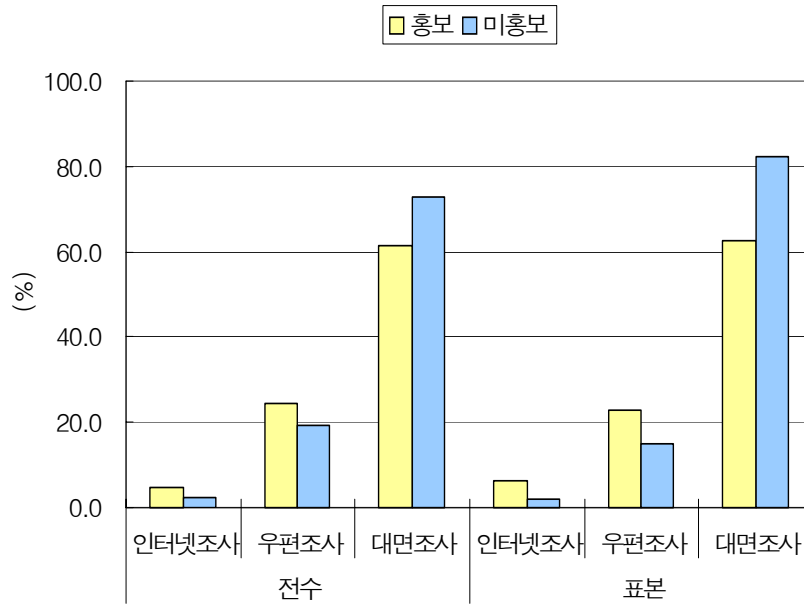
<표 2-18>을 보면 전수조사와 표본조사 모두에서 응답률이 가장 높은 자료수집방법은 대면조사이며, 다음으로 우편조사, 인터넷조사 순이었다. 비방문조사를 먼저 실시한 후 방문조사를 실시하였음에도 불구하고, 대면조사의 응답률이 가장 높게 나타났다는 점에서 주목해 볼 만하다. 대안으로 제시한 인터넷조사, 우편조사, ARS, CATI 중 응답자가 가장 많이 선택한 응답방법은 우편조사 → 인터넷조사 → CATI, ARS 순이었다. 특히, ARS와 CATI의 응답률은 모두 0.5% 미만으로 매우 낮았다.

이러한 응답순서는 인센티브의 홍보여부(그림 2-16) 뿐 아니라 시험조사지역(<표 2-19>)별로도 차이가 없었다. 특히 우편조사를 조사구 유형별로 보았을 경우에 일반조사구에서 실시한 우편조사 응답률의 경우에도 기흥구를 제외하고는 인터넷조사 응답률에 비해서도 높은 것으로 나타났다.

〈표 2-18〉 자료수집방법별 응답률

(단위: 가구, %)

구분	비방문				방문		무응답 추적	무응답	전체
	인터넷	우편	ARS	CATI	대면	배포			
전수	971 (3.8)	5,710 (22.6)	99 (0.4)	114 (0.5)	16,572 (65.6)	724 (2.9)	675 (2.7)	418 (1.7)	25,283 (100.0)
표본	126 (4.5)	550 (19.6)	- -	5 (0.2)	1,978 (70.7)	49 (1.8)	53 (1.9)	38 (1.4)	2,799 (100.0)



[그림 2-16] 인센티브 홍보여부에 따른 자료수집방법별 응답률

<표 2-19> 시험조사지역에 따른 자료수집방법별 응답률

(단위: %)

구분	인터넷 조사	우편 조사	ARS	CATI	대면 조사	배포 조사
전수						
범천1동(N=4,181)	1.1	12.1 (3.9)	0.0	0.2	81.9	1.3
명장1동(N=5,354)	3.7	19.4 (9.7)	0.5	0.7	67.8	3.8
기흥동(N=3,535)	4.5	31.1 (2.4)	0.4	0.5	58.4	1.1
광남동(N=8,121)	5.7	27.6 (6.0)	0.5	0.4	56.9	5.1
목호동(N=1,567)	1.5	12.8 (5.5)	0.1	0.4	80.7	0.5
인제읍(N=2,525)	3.5	24.5(14.6)	0.4	0.4	61.9	0.2
표본						
범천1동(n=516)	1.0	8.5 (2.9)	-	0.2	90.1	0.0
명장1동(n=563)	7.6	21.0(15.3)	-	0.4	68.0	0.0
기흥동(n=416)	3.4	25.0 (1.0)	-	0.0	69.0	0.0
광남동(n=799)	6.1	26.5 (5.0)	-	0.0	54.4	5.9
목호동(n=212)	1.4	9.9 (1.4)	-	0.0	88.7	0.0
인제읍(n=293)	4.1	17.4(11.8)	-	0.7	75.1	0.7

주: 괄호안은 일반조사구의 우편조사 응답률임.

<표 2-20> 비방문조사 및 전체 응답률의 지역별 비교

(단위: %)

구분	전수		표본	
	비방문조사 응답률(S.E)	전체 응답률(S.E)	비방문조사 응답률(S.E)	전체 응답률(S.E)
인센티브 사전 홍보지역				
명장1동	24.3 (0.6)	95.9 (0.3)	29.0 (1.9)	97.0 (17.1)
광남동	34.2 (0.5)	96.2 (0.2)	32.7 (1.7)	93.0 (25.5)
인제읍	28.8 (0.9)	90.9 (0.6)	22.5 (2.4)	98.0 (14.2)
F-값	77.46*	64.22*	5.76*	8.77*
인센티브 사전 미홍보지역				
범천1동	13.6 (0.5)	96.8 (0.3)	9.7 (1.3)	99.8 (4.1)
기흥동	36.5 (0.8)	96.0 (0.3)	28.4 (2.2)	97.4 (16.6)
목호동	14.9 (0.9)	96.0 (0.5)	11.3 (2.2)	100.0 (0.0)
F-값	344.76*	2.23	33.29*	8.14*

주: * ; p<0.05

비방문조사가 전체 응답률을 향상시키는가?

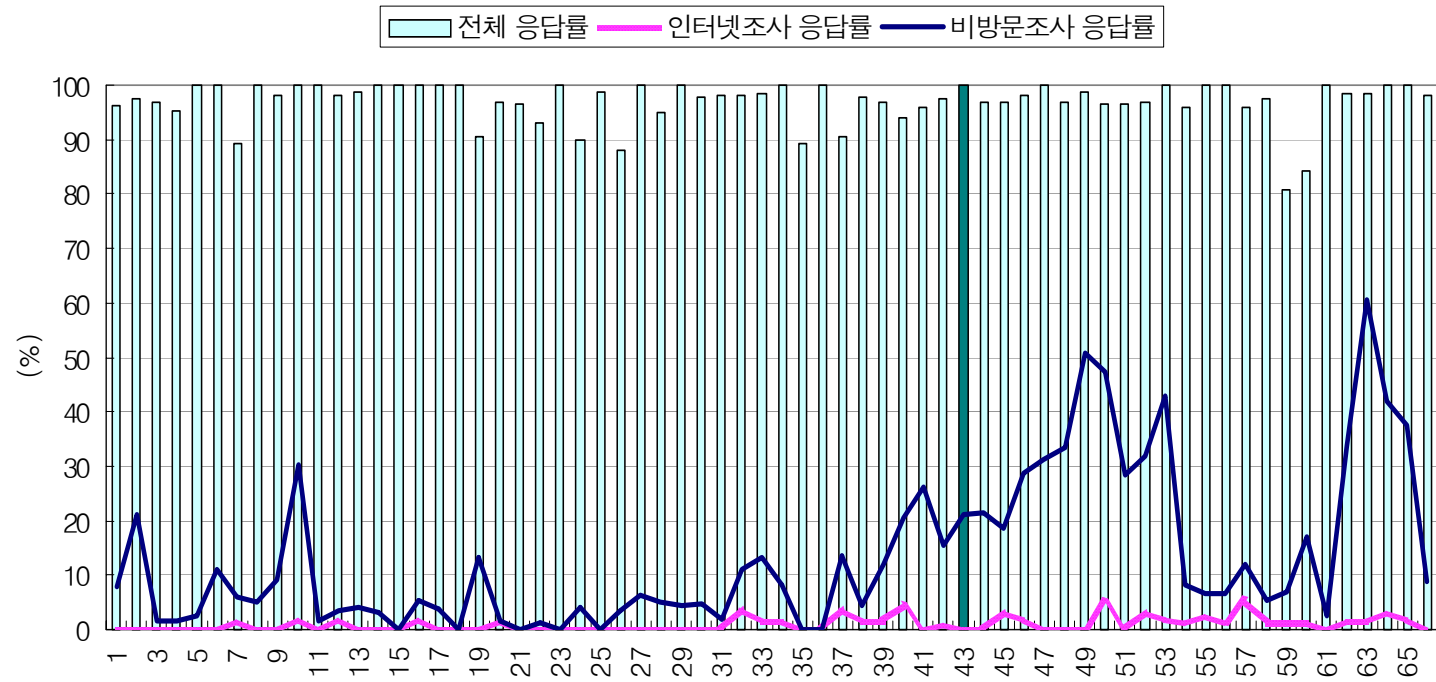
그렇다면, 비방문조사가 전체 응답률 향상에 기여했는가? 조사의 전 기간 동안 방문조사만으로 이루어진 통제집단이 없기 때문에 비방문조사가 전체 응답률을 향상시켰는지에 대해서 검증하기는 어렵다. 다만, 여기에서는 비방문조사 응답률과 전체 응답률 간의 관계를 살펴봄으로써 대안적인 답을 찾아보았다. 먼저 읍면동 수준에서 비방문조사 응답률과 전체 응답률의 관계를 보도록 하자.

<표 2-20>은 비방문조사 및 전체 응답률을 지역별로 비교하기 위한 일원분산분석 결과이다.²⁰⁾ 전수조사를 중심으로 살펴보면, 인센티브 홍보지역의 경우 비방문조사 응답률과 전체 응답률 모두 지역 간 차이가 통계적으로 유의미하였다. 미홍보지역의 경우에는 비방문조사 응답률의 지역 간 차이가 유의미하였으나 전체 응답률의 지역 간 차이는 유의미하지 않았다. 눈에 띄는 점은 비방문조사 응답률이 높은 곳이 반드시 전체 응답률이 높다는 근거는 없으며 특히 이것은 미홍보지역에서 뚜렷하게 나타났다. 표본조사의 경우에는 오히려 반대로 비방문조사 응답률이 높은 곳이 전체 응답률이 낮은 경향을 보이기도 하였다. 조사구 수준에서의 응답률 경향을 살펴보면 좀 더 뚜렷한 양상이 나타났다([그림 2-17] ~[그림 2-23]).

[그림 2-17]~[그림 2-22]는 전수조사구를 6개 시험조사 지역별로 나타낸 것이며, [그림 2-23]은 지역과 무관하게 표본조사구를 하나의 그림에

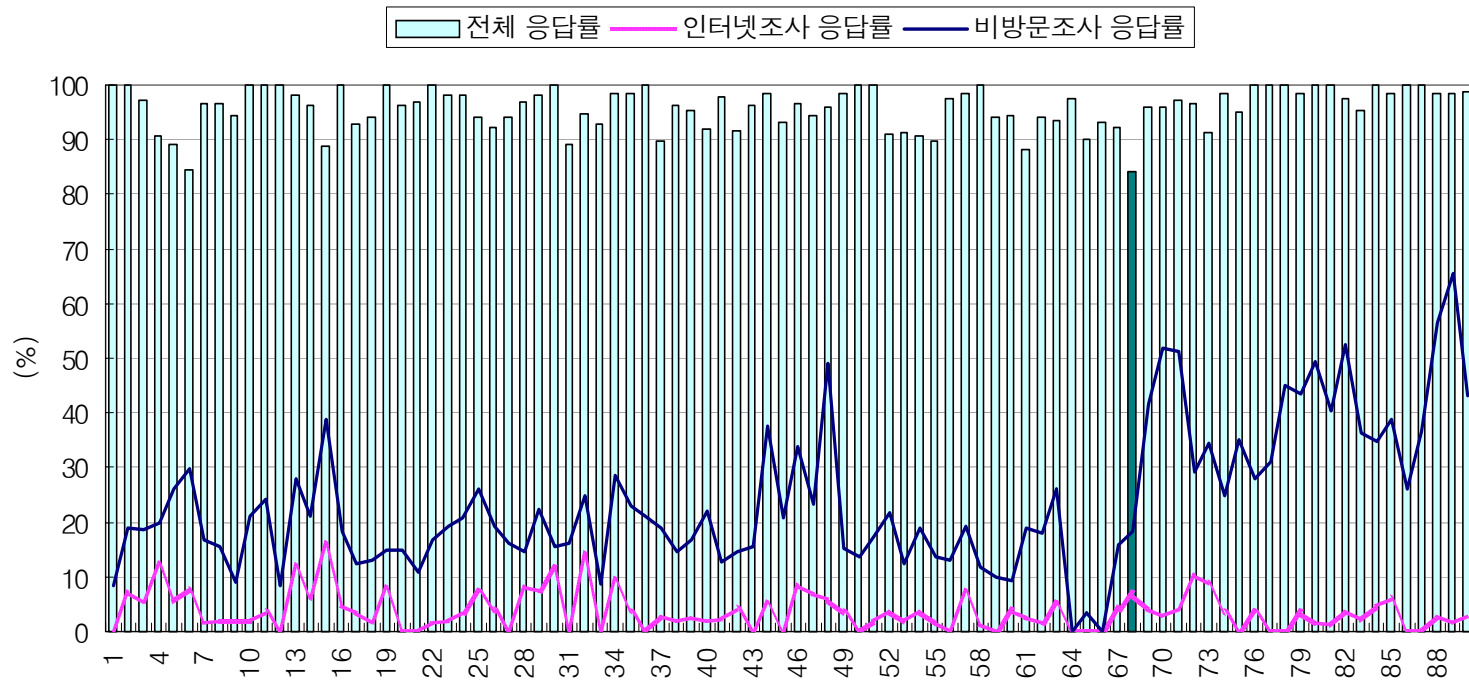
20) 표에는 제시하지 않았으나, 인센티브 사전홍보 여부를 구분하지 않고 비방문조사 응답률과 전체 응답률의 비교를 6개 지역 모두에 대해서 실시하였을 때 전수조사의 경우 비방문조사 응답률이 13.6%에서부터 36.5%까지 지역별로 그 차이가 큰 것으로 나타났다. 이러한 지역별 차이를 F-값을 통해서 검정해 본 결과 95% 유의수준에서 통계적으로 유의미한 것으로 나타났다. 반면에 전체 응답률에서는 시험조사 지역 간 응답률의 차이가 감소하였음을 알 수 있다. 물론, F-값은 95% 유의수준에서 통계적으로 유의미한 것으로 나타나기는 하였으나, 어느 지역 간의 응답률의 차이가 유의한지에 대한 사후검정결과를 보면 전체 응답률에서는 지역 간 차이가 줄어들었음을 알 수 있다. 6개 지역을 두 지역씩 비교를 하게 될 경우 총 15개 비교쌍이 생성되는데, 비교쌍의 응답률 차이가 통계적으로 유의한지를 살펴본 결과, 전수조사의 비방문조사 응답률 경우에는 2개쌍을 제외한 모든 비교쌍의 응답률 차이가 통계적으로 유의미하였던 반면에 전체 응답률에서는 통계적으로 유의미한 비교쌍이 5개로 줄어들었다. 이러한 경향은 표본조사에서도 크게 다르지 않게 나타났다(부록 1, 2 참고).

제시한 것이다. 여기에서 그래프의 조사구 순서는 실제 조사구 번호와는 차이가 있다. 범천1동, 기흥동, 묵호동 등 인센티브 사전미홍보지역을 중심으로 그 경향을 보면, 비방문조사의 응답률이 일반조사구보다 아파트 조사구에서 더 높게 나타나는 경향이 있다는 점과 비방문조사 응답률이 높은 곳에서 전체 응답률이 높지 않으며 비방문조사 응답률이 낮은 곳이 전체 응답률이 낮지 않다는 점이 눈에 띄었다.



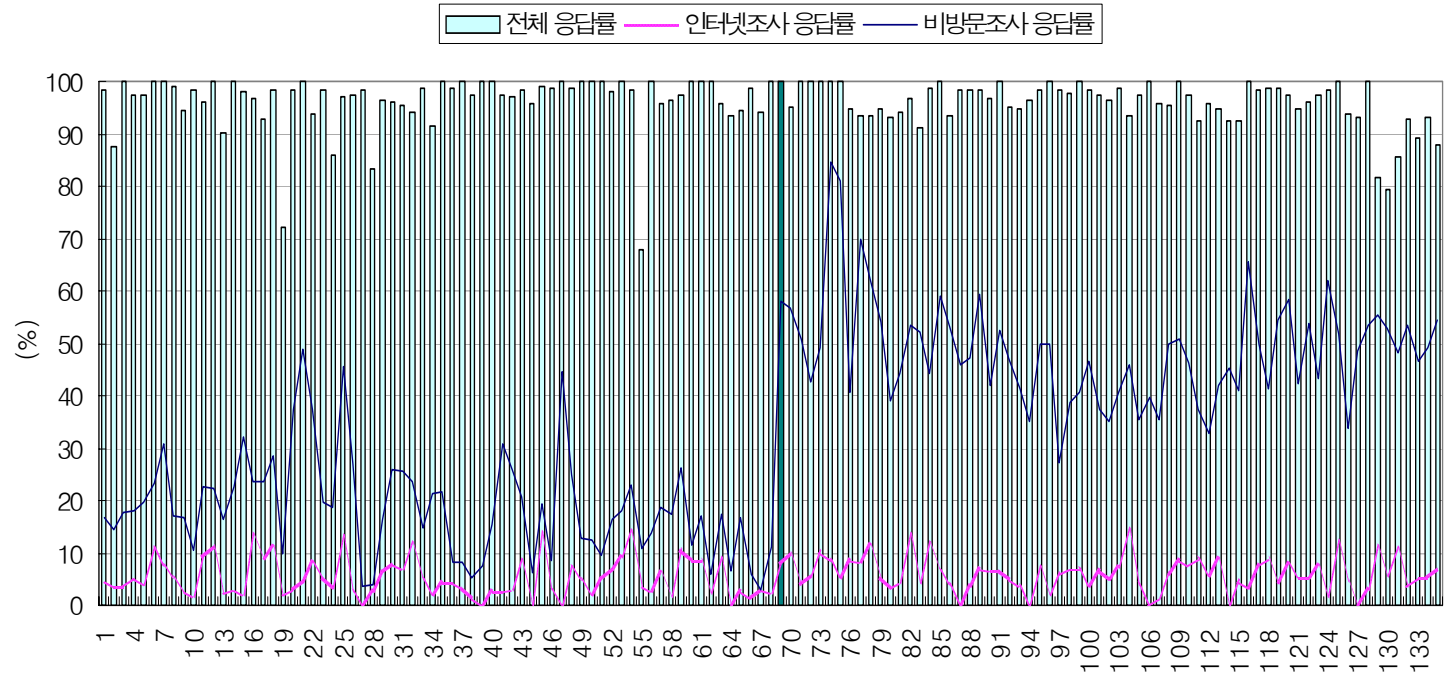
주: 1~42번까지는 일반조사구, 43번부터는 아파트조사구임.

[그림 2-17] 전수조사구별 응답률: 범천1동



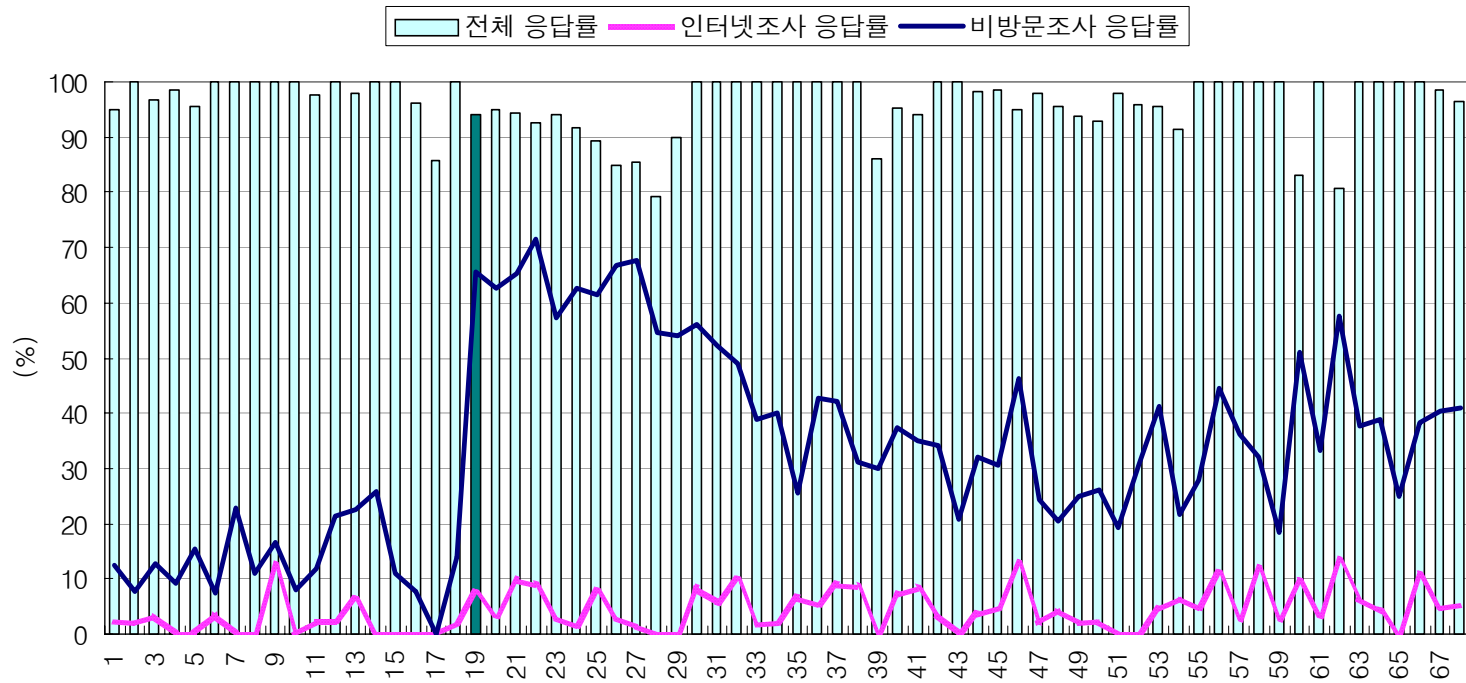
주: 1~67번까지는 일반조사구, 68번부터는 아파트조사구임.

[그림 2-18] 전수조사구별 응답률: 명장1동



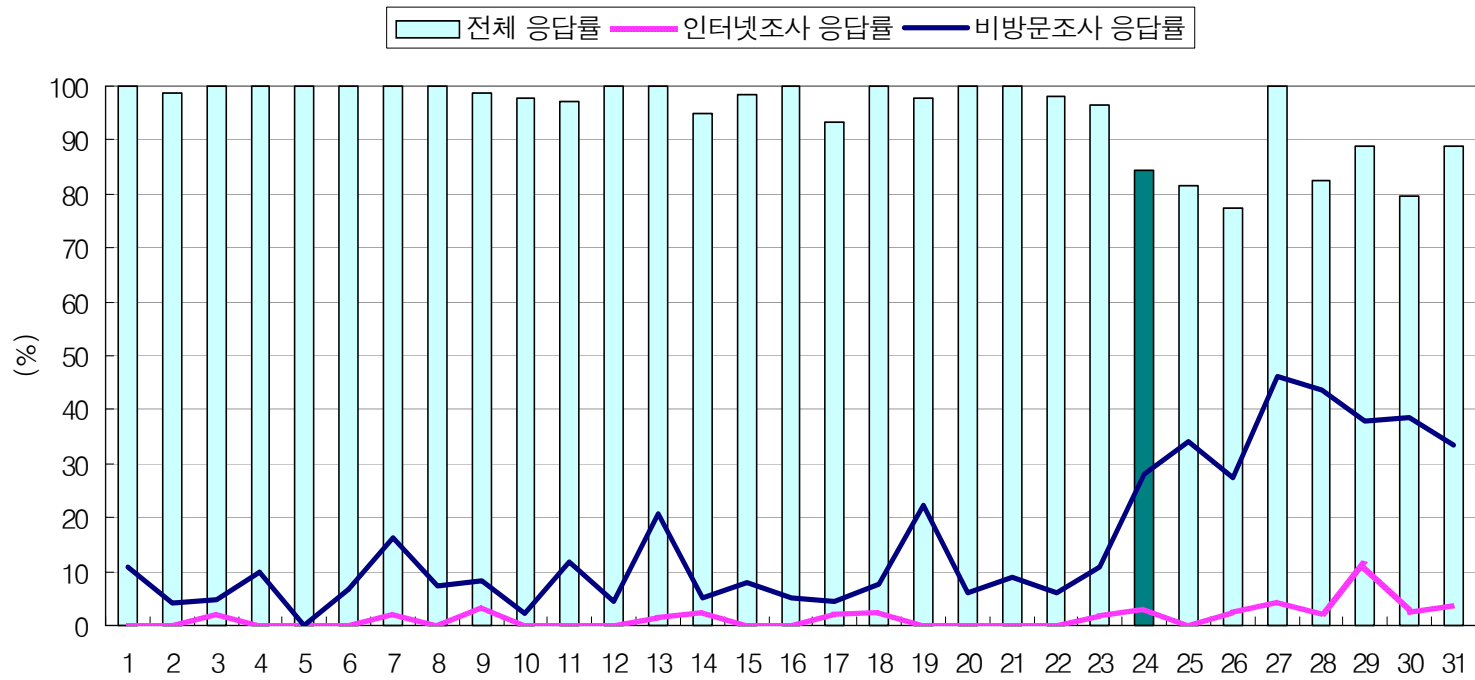
주: 1~68번까지는 일반조사구, 69번부터는 아파트조사구임.

[그림 2-19] 전수조사구별 응답률: 광남동



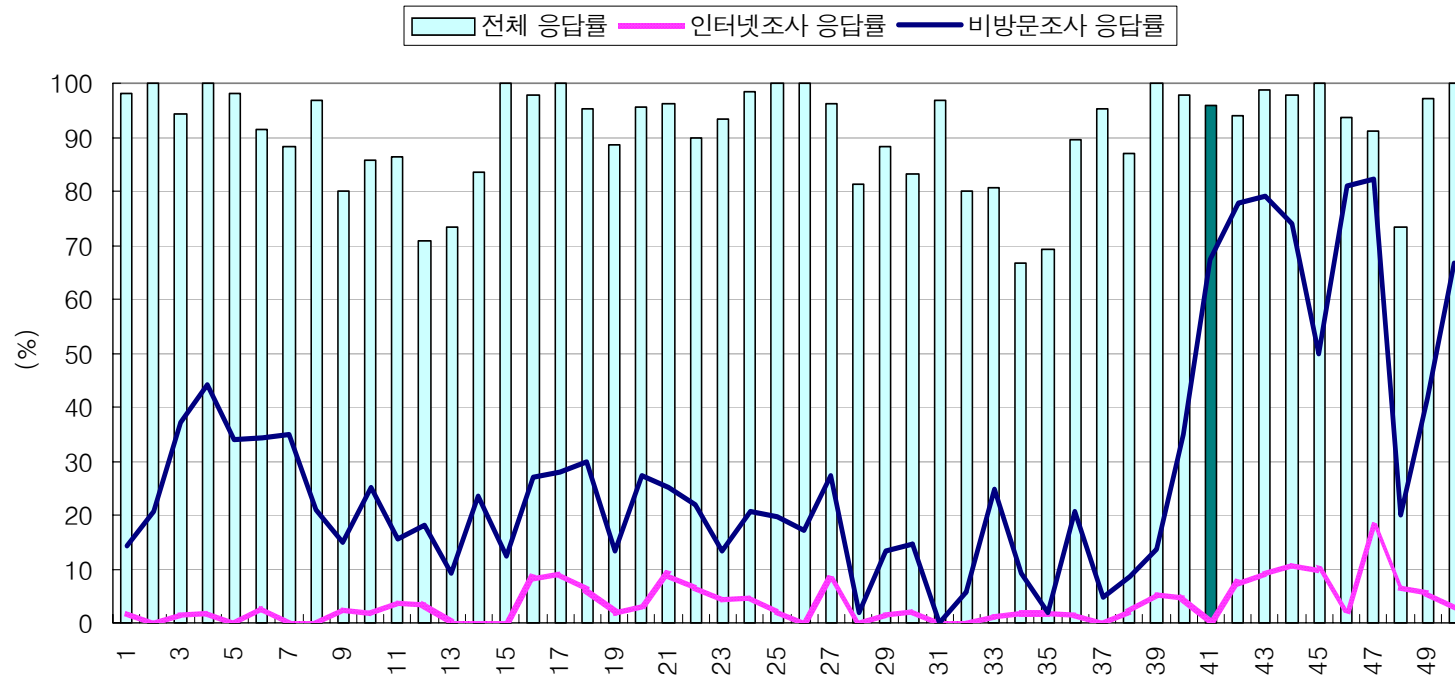
주: 1~18번까지는 일반조사구, 19번부터는 아파트조사구임.

[그림 2-20] 전수조사구별 응답률: 기흥동



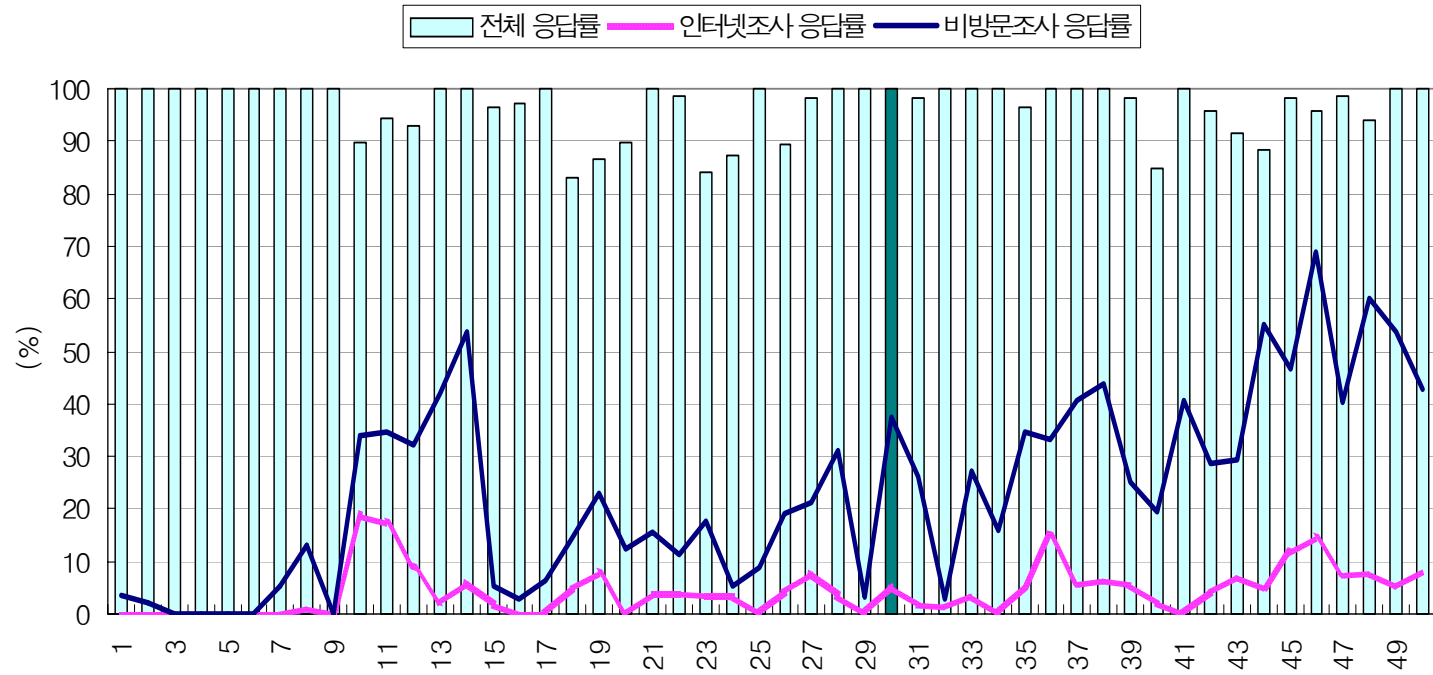
주: 1~23번까지는 일반조사구, 24번부터는 아파트조사구임.

[그림 2-21] 전수조사구별 응답률: 묵호동



주: 1~40번까지는 일반조사구, 41번부터는 아파트조사구임.

[그림 2-22] 전수조사구별 응답률: 인제읍



주: 1~29번까지는 일반조사구, 30번부터는 아파트조사구임.

[그림 2-23] 표본조사구별 응답률: 전체지역

<표 2-21>은 조사구 수준에서 응답률 간 상관분석을 실시한 결과이다. 전체 응답률과 비방문조사 응답률 외에도 인터넷조사 응답률과 우편조사 응답률 간의 관계를 함께 분석해 보았다. 그 결과, 전체 응답률과 인터넷조사 응답률 사이에 음의 상관관계가 나타났다. 물론 이 관계는 95% 신뢰수준에서 통계적으로 유의미하지는 않으나 비방문조사 응답률이 전체 응답률을 향상시켜 줄 것이라는 기대에는 부응하지 못한다는 점에서 눈여겨 볼 만하다. 한편, 인센티브 사전 홍보여부에 따른 하위지역별 상관분석을 실시해 본 결과, 인센티브 홍보지역에서는 비방문조사 응답률과 전체 응답률 간에 정적인 상관관계가 나타난 반면에 미홍보지역에서는 부적인 관계가 나타났다.

<표 2-21> 전체 응답률과 비방문조사 응답률 간 상관관계

	전체 응답률 ¹⁾	비방문조사 응답률 ²⁾	인터넷조사 응답률 ³⁾	우편조사 응답률 ⁴⁾
전체(N=490)				
1)	1.00	-	-	-
2)	-0.01	1.00	-	-
3)	-0.02	0.47*	1.00	-
4)	0.01	0.98*	0.28*	1.00
인센티브 사전 홍보지역(N=305)				
1)	1.00	-	-	-
2)	0.18*	1.00	-	-
3)	0.07	0.37*	1.00	-
4)	0.18*	0.97*	0.15*	1.00
인센티브 사전 미홍보지역(N=185)				
1)	1.00	-	-	-
2)	-0.28*	1.00	-	-
3)	-0.11	0.58*	1.00	-
4)	-0.28*	0.99*	0.45*	1.00

주: *, $p < 0.05$ 임.

이 분석결과 비방문조사 응답률이 전체 응답률 향상에 영향을 미칠 것이라는 당초의 질문에 긍정적인 답을 할 수 있는 것은 인센티브 사전 홍보지역에서만이다. 그러나, 조사구 수준에서 비방문조사 응답률의 변이가 매우 큼에도 불구하고 전체 응답률에서는 차이가 나타나지 않았다는 것은 미국 센서스의 혼합방법조사 실험결과에서 보았던 것처럼 응답자들의 이동이 일어난 것에 기인했음을 추정해 볼 수 있다. 즉, 기존의 방문조사 응답자들이 비방문조사 응답으로 이동을 하면서 결과적으로 전체 응답률에 영향을 미치지 못하게 된 것이다.

자료수집방법별 응답률 결정요인은 무엇인가?

그렇다면, 조사구 수준에서 어떤 요인들이 응답률에 영향을 미쳤을까? 여기에서는 조사구 특성 변수들을 독립변수로, 조사구 수준에서 각각의 응답률, 즉 전체 응답률, 비방문조사 응답률, 그리고 여기에 부가적으로 인터넷 및 우편조사 응답률을 종속변수로 설정한 후 중다회귀분석을 실시하였다.²¹⁾

먼저, <표 2-22>는 전체 응답률에 영향을 미치는 조사구 특성 변수들에 대한 중다회귀분석결과이다. 그 결과, 조사구 내 여성비율, 단독가구 비율, 비농림어가비율, 인센티브여부가 95% 신뢰수준에서 통계적으로 유의미한 효과를 보여주었으며, 그 방향은 여성비율과 비농림어가비율은 정적으로, 1인가구비율과 인센티브여부는 부적으로 나타났다. 독립변수들의 종속변수에 대한 기여도를 표준화된 회귀계수를 통해서 살펴보면, 비농림어가비율, 인센티브여부, 단독가구비율, 여성비율 순이었다. 한편, 단순상관관계에서는 유의미한 효과를 보여주는 것으로 나타났던 25~64세의 노동가능인구비율과 자가소유비율, 그리고 비농림어가비율은 다른 변수를 통제하였을 경우에 그 효과가 사라졌다. 이 모형은 F-값이 8.71로 95% 신뢰수준에서 통계적으로 유의미하였으며, 전체 응답률의

21) 표본조사구와 전수조사구를 구분하여 각각의 집단 내에서 상관분석과 회귀분석을 실시해 보았으나 이것은 전체 조사구(490개)를 대상으로 전수조사여부에 따른 더미 처리를 한 것과 큰 차이가 없으므로 여기에서는 전체집단을 대상으로 한 분석결과만 제시하였다.

총 분산 중 약 17%를 설명하였다. 한편, 앞서 인센티브 효과에 대한 분석에서 나타났듯이 여타의 변수를 통제한 이후에도 인센티브 홍보여부가 전체 응답률과 부적인 관계를 보고하였다는 점은 주목해 볼 만하다.

<표 2-23>은 동일한 모형을 비방문조사 응답률에 적용한 중다회귀분석결과이다. 이 모형의 F-값은 89.16으로 95% 신뢰수준에서 통계적으로 유의미하였으며, 비방문조사 응답률 분산의 67% 가량을 설명하였다. 개별 독립변수의 종속변수에 대한 기여도와 통계적인 유의성을 검정해 본 결과, 95% 신뢰수준에서 비방문조사 응답률에 영향을 미치는 독립변수는 조사구 내 노동가능인구비율, 1인가구비율, 비농림어가비율, 인구밀도, 아파트조사구여부, 인센티브사전홍보여부이며, 아파트조사구여부, 1인가구비율, 인센티브홍보여부 순으로 영향을 크게 미치는 것으로 나타났다.

비방문조사 내에서 인터넷조사와 우편조사 응답률과 조사구 특성과는 어떠한 관계가 있는가? <표 2-24>는 인터넷조사 응답률에 영향을 미치는 요인에 대한 중다회귀분석결과이다. 95% 유의수준에서 통계적으로 유의미한 효과를 보이는 변수는 조사구 내 평균연령, 1인가구비율, 비농림어가비율, 인구밀도, 인센티브사전홍보여부, 전수조사구여부이다. 평균연령이 낮을수록, 1인가구비율이 낮을수록, 인구밀도가 낮을수록 인터넷조사의 응답률이 높아졌으며, 비농림어가비율이 높을수록, 인센티브를 사전에 홍보한 경우가, 그리고 표본조사구인 경우에 인터넷조사의 응답률이 높아졌다. 우편조사의 경우에는(<표 2-25>), 조사구 내 노동가능인구비율, 1인가구비율, 인구밀도, 아파트조사구여부, 인센티브홍보여부가 영향을 미쳤으며 이 중에서 아파트조사구여부 변수의 영향력이 가장 큰 것으로 나타났다.

〈표 2-22〉 전체 응답률 결정 요인에 대한 중다회귀분석결과

구분	단순 상관계수	비표준화된 계수		표준화된 계수	t-값
		B	S.E.		
여성비율	0.19 *	0.14	0.05	0.13	2.72 *
평균연령	0.04	0.00	0.00	-0.02	-0.26
평균교육수준	-0.06	-0.01	0.01	-0.06	-0.79
25~64세인구비율	-0.14 *	-0.07	0.05	-0.09	-1.46
1인가구비율	-0.13 *	-0.05	0.02	-0.17	-2.65 *
자가소유비율	0.08 *	0.02	0.01	0.11	1.85
비농림어가비율	0.22 *	0.12	0.02	0.29	5.12 *
인구밀도	0.11 *	0.00	0.00	0.07	1.31
아파트조사구여부	0.01	-0.01	0.01	-0.11	-1.84
인센티브사전홍보여부	-0.13 *	-0.02	0.01	-0.19	-3.48 *
전수조사구여부	-0.06	-0.01	0.01	-0.04	-0.98
상수		0.87	0.06		14.72 *
F-값		8.71*			
R ²		0.17			

주: *, P<0.05.

〈표 2-23〉 비방문조사 응답률 결정 요인에 대한 중다회귀분석결과

구분	단순 상관계수	비표준화된 계수		표준화된 계수	t-값
		B	S.E.		
여성비율	-0.13 *	-0.12	0.11	-0.03	-1.08
평균연령	-0.52 *	0.00	0.00	-0.05	-1.10
평균교육수준	0.29 *	0.02	0.02	0.07	1.47
25~64세인구비율	0.06	-0.21	0.09	-0.09	-2.22 *
1인가구비율	-0.52 *	-0.25	0.04	-0.23	-5.99 *
자가소유비율	0.16 *	0.01	0.03	0.02	0.43
비농림어가비율	0.26 *	0.09	0.05	0.07	1.84 *
인구밀도	-0.26 *	0.00	0.00	-0.08	-2.35 *
아파트조사구여부	0.71 *	0.22	0.01	0.60	16.32 *
인센티브사전홍보여부	0.20 *	0.07	0.01	0.18	5.40 *
전수조사구여부	0.05	0.02	0.02	0.04	1.40
상수		0.25	0.12		2.10 *
F-값		89.16*			
R ²		0.67			

주: *, P<0.05.

〈표 2-24〉 인터넷조사 응답률 결정 요인에 대한 중다회귀분석결과

독립변수	단순 상관계수	비표준화된 계수		표준화된 계수	t-값
		B	S.E.		
여성비율	-0.16 *	0.00	0.03	0.00	-0.08
평균연령	-0.37 *	0.00	0.00	-0.14	-1.87 *
평균교육수준	0.18 *	0.00	0.00	0.02	0.25
25~64세인구비율	0.13 *	0.02	0.03	0.04	0.69
1인가구비율	-0.37 *	-0.05	0.01	-0.22	-3.82 *
자가소유비율	0.06	-0.01	0.01	-0.05	-0.91
비농림어가비율	0.21 *	0.03	0.01	0.13	2.36 *
인구밀도	-0.19 *	0.00	0.00	-0.14	-2.89 *
아파트조사구여부	0.23 *	0.00	0.00	0.05	0.93
인센티브사전홍보여부	0.30 *	0.02	0.00	0.22	4.53 *
전수조사구여부	-0.05	-0.01	0.00	-0.07	-1.72 *
상수		0.04	0.04		0.97
F-값		16.53*			
R ²		0.28			

주: *, P<0.05.

〈표 2-25〉 우편조사 응답률 결정 요인에 대한 중다회귀분석결과

독립변수	단순 상관계수	비표준화된 계수		표준화된 계수	t-값
		B	S.E.		
여성비율	-0.10 *	-0.11	0.10	-0.03	-1.08
평균연령	-0.47 *	0.00	0.00	-0.02	-0.32
평균교육수준	0.27 *	0.02	0.01	0.08	1.57
25~64세인구비율	0.03	-0.24	0.09	-0.11	-2.71 *
1인가구비율	-0.48 *	-0.19	0.04	-0.20	-4.78 *
자가소유비율	0.16 *	0.02	0.03	0.03	0.83
비농림어가비율	0.23 *	0.06	0.05	0.05	1.32
인구밀도	-0.24 *	0.00	0.00	-0.06	-1.76 *
아파트조사구여부	0.72 *	0.22	0.01	0.64	16.75 *
인센티브사전홍보여부	0.15 *	0.05	0.01	0.14	4.10 *
전수조사구여부	0.05	0.02	0.02	0.04	1.61
상수		0.20	0.12		1.72 *
F-값		77.41*			
R ²		0.64			

주: *, P<0.05.

각각의 응답률에 영향을 미치는 조사구 특성변수에 대한 회귀분석결과, 전체 응답률과 비방문조사 응답률에 공통적으로 영향을 미치는 1인가구비율과 비농림어가비율을 제외하면 자료수집방법에 따라서 영향을 미치는 조사구 특성 변수가 달라지는 것으로 나타났다(<표 2-26>).

<표 2-26> 중다회귀분석결과 요약

구분	전체 응답률	비방문조사 응답률	인터넷조사 응답률	우편조사 응답률
여성비율	+			
평균연령			-	
평균교육수준				
25~64세인구비율		-		-
1인가구비율	-	-	-	-
자가소유비율				
비농림어가비율	+	+	+	
인구밀도		-	-	-
아파트조사구여부		+		+
인센티브홍보여부	-	+	+	+
전수조사구여부			-	

주: +는 정적인 효과를, -는 부적인 효과를, 공란은 유의미한 효과가 없음을 의미함.

첫째, 전체 응답률에서는 유의미하였으나 비방문조사 응답률에서는 유의미하지 않게 나타난 변수는 여성비율이다. 일반적으로 성별, 즉 여성은 조사에 응답할 가능성에 정적인 영향을 미치는 것으로 보고되고 있는데, 이 분석결과에 따르면 성별효과가 자료수집방법에 따라서 달라질 수도 있음을 알 수 있다. 둘째, 전체 응답률에서는 유의미하지 않았으나 비방문조사 응답률에서는 유의미한 것으로 나온 변수는 노동가능인구비율, 인구밀도, 아파트조사구여부이다. 이 중 아파트 조사구는 다른 변수들에 비해서 큰 효과를 갖는 것으로 나타났다. 셋째 유의미한 변수 중에서 그 효과가 반대로 나온 것은 인센티브 홍보여부이다. 전체 응답률에서는 부적인 효과를 갖고 있었던 인센티브 홍보효과가 비방문조사

응답률에서는 정적인 효과를 보여주었다. 넷째, 인터넷조사 응답률과 우편조사 응답률에 영향을 미치는 변수도 차별적으로 나타났다. 인터넷조사 응답률에는 영향을 미치나 우편조사 응답률에는 영향을 미치지 않는 변수로써 평균연령, 비농림어가비율, 전수조사구여부가, 반대의 경우에는 노동가능인구비율, 아파트조사구여부가 있다. 인터넷조사 응답률은 조사구 내 평균연령이 낮을수록 높으며, 우편조사 응답률은 아파트조사구인 경우가 일반조사구에 비해서 높다라는 점이 주목해 볼 특성이다. 그리고 두 조사의 응답률에 공통적으로 영향을 미치는 변수로는 1인가구비율, 인구밀도, 인센티브홍보여부가 있었다.

한편, <표 2-27>은 Sampson과 그의 동료들이 집합적 효능감으로 정의한 변수가 응답률에 미치는 효과를 검증해 보기 위해서 실시한 단순상관분석결과이다. 회귀분석을 수행하기 위해서는 표본의 사례수가 최소한 독립변수의 20배 이상은 되어야 한다는 가정(성태제, 2007)에 비추어 볼 때 표본조사구에서 도출이 가능한 집합적 효능감 변수에 대한 중다회귀분석은 무리가 있으므로 단순상관관계만을 살펴보았다. 이는 향후 응답률 결정 모형의 정교화 단계에서 추가적인 지표를 설정함에 있어서 함의를 제공해 줄 수 있을 것이다.

<표 2-27> 집합적 효능감 변수와 응답률 간의 상관관계(n=50)

변수	전체 응답률	비방문조사 응답률	인터넷조사 응답률	우편조사 응답률
임금근로자비율	-0.35 *	0.19	0.07	0.21
비이동자비율	0.14	-0.31 *	-0.02	-0.35 *
가구평균거주기간	0.23	-0.51 *	-0.26 *	-0.51 *
중하위층소득가구비율	0.09	-0.56 *	-0.47 *	-0.51 *
컴퓨터보유비율	-0.15	0.65 *	0.44 *	0.62 *

주 : *, P<0.05.

상관분석결과, 전체 응답률과 통계적으로 유의미한 상관이 있는 변수는 임금근로자비율로 나타났으며, 비이동자비율, 가구평균거주기간, 중하위층소득가구비율, 컴퓨터비율은 비방문조사 응답률과 관계가 있는 것으로 나타났다. 비이동자의 비율이 높을수록 조사를 거부할 가능성이

높다는 기존의 연구결과(Johnson et al., 2006)에 비추어 본다면 우리나라에서 비이동자의 비율이 응답률과 부적인 관계를 갖는 것은 상식적으로 이해할 만하다고 하겠다. 또한, Johnson과 동료들의 연구에서 빈곤선이 하 인구 비율이 응답여부와 부적인 관계를 갖고 있었던 것도 우리나라에서 지지되었다. 단, 차이점은 이 변수들이 전체 응답률이 아닌 비방문조사 응답률에만 영향을 미치고 있다는 점이다.

자료수집방법이 응답률에 미치는 효과에 대한 지금까지의 분석결과를 요약해 보면 다음과 같다. 첫째, 응답자들이 가장 많이 선택한 방법은 대면조사이다. 대안적인 응답방법 중에서는 우편조사 → 인터넷조사 순으로 응답을 많이 하였다.

둘째, 비방문조사가 전체 응답률 향상에 기여했는가에 대한 질문에는 조사의 전 기간 동안 방문조사만으로 이루어진 통제집단이 없기 때문에 답을 하기에는 어려움이 있다. 다만 비방문조사 응답률과 전체 응답률 간의 관계를 살펴봄으로써 대안적인 답을 찾아보면 비방문조사 응답률이 높은 곳이 전체 응답률이 높다는 일관적인 결과를 발견하기는 어려웠다. 오히려 특정 지역에서는 비방문조사에 의해 증가된 응답률이 방문조사 응답률의 감소에 의해서 상쇄되어 전반적으로 전체 응답률이 감소하는 경향이 나타나기도 하였다.

셋째, 이에 따라서 조사구 수준에서 응답률을 결정하는 요인에 대한 회귀분석을 실시했는데, 그 결과 전체 응답률, 비방문조사 응답률, 인터넷 및 우편조사 응답률에 영향을 미치는 요인들이 차별적인 것으로 나타났다. 여성비율은 전체 응답률에만 통계적으로 유의미한 영향을 미쳤으며 노동가능인구비율, 인구밀도, 아파트조사구여부는 비방문조사 응답률에만 영향을 주었다. 한편, 인터넷조사 응답률에는 평균연령이, 우편조사 응답률에는 아파트조사구의 영향력이 상대적으로 큰 것으로 나타났다. 이러한 차별적인 효과는 비방문조사 응답률이 높다고 해서 반드시 전체 응답률이 높지 않다는 것을 지지해주는 것으로 해석될 수 있다.

넷째, 그러나 조사표에 있는 항목으로만 응답률을 추정하는 데에는 한계가 있다. 여전히 응답률 변량의 상당부분이 설명되지 않은 채 남아

있으므로 추가적인 변수에 대한 고민이 필요할 것으로 보인다. 특히 표본조사 항목 중에서 인구이동 및 소득, 컴퓨터 소유 여부 등과 같은 변수가 고려될 수 있을 것이다.

나. 내검불일치율과 항목무응답률

컴퓨터를 보조적으로 활용하는 자료수집방법은 일반적으로 자체 내검프로그램을 장착하고 있기 때문에 종이와 연필에 의한 자료수집방법보다 더 나은 품질을 산출하는 것으로 알려져 있다(Schneider, 2002). 그렇다면, 우리나라의 경우에는 어떠한가? 이것을 살펴보기 위해서 자료수집방법 간에 내검불일치율과 항목무응답률에 차이가 있는가를 검토하였다. 여기에서는 인구·가구·주택항목을 대상으로 인터넷 및 우편조사를 포함하여 실제 시험조사에서 사용된 모든 자료수집방법 간 내검불일치율과 항목무응답률을 비교하였다.

1) 자료수집방법과 내검불일치율

가) 인구항목

<표 2-28>은 인구항목에 대한 내검불일치율을 보여주고 있다. 인구항목 전체 중에서 내검불일치율이 가장 높은 항목은 출생월로 전수조사와 표본조사 모두에서 50% 이상의 내검불일치율을 보였다. 내검불일치율이 높게 나타나는 데 기여한 주요한 원인은 5월로 기입된 출생월이 내검과정에서 05월로 처리되었기 때문이다. 내검불일치율이 가장 낮은 항목은 전수조사의 경우는 성별, 표본조사의 경우에는 교육정도로 두 변수 모두 0.3%의 내검불일치율을 보여 자료의 정확도가 다른 항목에 비해서 높은 것으로 나타났다. 이와 함께 나이와 교육정도에 있어서도 내검불일치율이 전수조사와 표본조사 모두에서 1% 미만으로 나타났다.

자료수집방법별로 보면, 전수항목 전체에서 내검불일치율은 인터넷조사에서 가장 낮고, CATI에서 가장 높았다. 인터넷조사의 경우에는 모든 항목에서 가장 낮은 내검불일치율을 보이고 있는 반면에 CATI는 나이, 생월, 생일, 가구주와의 관계에서 내검불일치율이 높게 나타났다. 여타항목에서는 성명의 경우에는 ARS, 성별, 생년, 교육수준, 혼인상태의

경우에는 조사원에 의한 전화조사에서 내검불일치율이 높게 나타났다. 내검불일치율이 높은 자료수집방법의 공통점은 이들이 모두 전화를 이용한다는 점이다. 개인적인 정보를 전화로 말하는 것에 대한 부담감이 작용한 것으로 추정되었다. 자료수집방법에 따른 내검불일치율의 차이는 95% 신뢰수준에서 모두 통계적으로 유의미한 것으로 나타났다. 표본 조사의 경우에도 특정 항목을 제외하고 전반적인 패턴은 유사하였다.

〈표 2-28〉 자료수집방법별 인구항목 내검불일치율 비교

(단위: %)

전수 (사례수)	전체 (70,317)	인터넷 (3,146)	우편 (18,113)	대면 (44,454)	ARS (292)	CATI (368)	전화 (1,600)	배포 (2,344)	F-값
성명	1.8	0.1	1.9	1.9	4.1	1.1	1.6	1.8	10.71 *
성별	0.3	0.1	0.4	0.3	0.7	0.8	1.1	0.5	7.65 *
나이	0.6	0.3	0.5	0.7	0.7	1.6	0.8	0.3	3.59 *
생년	1.4	0.3	0.8	1.5	0.7	0.8	6.0	0.6	57.57 *
생월	52.8	37.7	52.8	53.4	55.8	75.5	64.7	48.2	81.23 *
생일	19.2	15.0	18.4	19.4	25.7	33.2	24.9	18.1	22.55 *
관계	0.9	0.1	0.8	0.9	1.0	4.1	1.9	1.4	15.70 *
교육수준	0.8	0.1	0.6	0.8	1.0	0.8	1.6	1.0	6.75 *
졸업여부	2.2	0.2	3.4	1.8	1.4	1.1	2.1	4.7	45.52 *
혼인상태	0.9	0.2	0.9	0.9	1.4	0.5	2.5	1.7	12.75 *
전체	80.8	54.0	80.5	81.6	92.5	119.6	107.1	78.3	104.89 *
표본 (사례수)	전체 (7,582)	인터넷 (450)	우편 (1,707)	대면 (5,117)	ARS (-)	CATI (19)	전화 (122)	배포 (167)	F-값
성명	3.9	0.2	3.3	4.6	-	0.0	1.6	1.2	5.87 *
성별	0.57	0.0	0.6	0.6	-	0.0	0.0	1.2	0.94
나이	0.78	0.0	1.1	0.8	-	0.0	0.0	0.0	1.69
생년	1.45	0.2	1.7	1.5	-	0.0	1.6	0.6	1.35
생월	54.39	38.9	49.2	57.4	-	84.2	65.6	45.5	20.04 *
생일	21.29	12.9	18.8	22.7	-	26.3	30.3	18.6	7.80 *
관계	1.48	0.0	2.1	1.2	-	0.0	3.3	6.0	7.85 *
교육수준	0.33	0.0	0.6	0.2	-	0.0	0.8	0.6	1.87 *
졸업여부	1.13	0.0	1.6	1.1	-	0.0	0.0	1.2	2.15 *
혼인상태	1.7	0.0	1.9	1.7	-	0.0	3.3	2.4	2.16 *
전체	87.0	52.2	80.9	91.8	-	110.5	106.6	77.3	22.55 *

주: *, P<0.05임.

나) 가구항목

가구항목의 경우, 전수조사와 표본조사 모두에서 내검불일치율이 가장 높은 것은 층구분 항목으로 각각 6.6%와 5.4%의 내검불일치율을 보였다(<표 2-29>). 이는 조사표 설계와 관련된 것으로 보인다. [그림 2-24]에서 보듯이 층구분을 묻는 항목은 거주하는 건물의 층수와 실제 거주 층을 묻는 질문 사이에 놓여져 있어서 응답자들이 응답을 하지 않고 건너뛰기 쉬운 구조임을 알 수 있다. 특히 무응답을 허용하지 않는 인터넷 조사에서 이 항목에 대한 내검불일치율이 가장 낮고 응답자 기입식의 한 방법인 배포조사에서 내검불일치율이 높게 나왔음은 이를 반증해주는 것이다. 조사표 설계에 따라서 발생할 수 있는 내검불일치율이 높게 나타나는 또 다른 유형은 주항목과 부가항목의 설계이다. 전반적으로 부엌, 화장실, 목욕시설 등의 형태보다는 이 시설들의 단독사용여부를 묻는 부가항목에서 내검불일치율이 더 높게 나타났다([그림 2-25]).

가구항목의 경우에 자료수집방법별로 내검불일치율에 차이가 있는가? 전체적으로 보면 전수조사의 경우 내검불일치율이 ARS에서 가장 높고, 조사원에 의한 전화조사에서 가장 낮았다. 전화조사 다음으로 낮은 방법은 대면조사이다. 표본조사의 경우에는 배포조사와 인터넷조사에서 내검불일치율이 상대적으로 높게 나타났으며 전화조사에서 낮게 나타났다. 이를 통해서 알 수 있는 것은 인구항목과 달리, 가구항목에서는 내검불일치율이 응답자 기입방식의 자료수집방법에서 높으며, 조사원 기입방식의 자료수집방법에서 낮다는 것이다. 특히, 전수조사나 표본조사에서 모두 전화조사에서 내검불일치율이 낮게 나타난 이유는 인구항목과 달리, 조사원이 인접가구로부터 가구에 대한 사항을 대략적으로 파악한 후 응답자들과 전화통화를 시도했기 때문인 것으로 추정된다.

〈표 2-29〉 자료수집방법별 가구항목 내검불일치율 비교

(단위: %)

전수 (사례수)	전체 (24,865)	인터넷 (971)	우편 (5,710)	대면 (16,572)	ARS (99)	CATI (114)	전화 (675)	배포 (724)	F-값
가구구분	2.7	2.2	2.3	2.8	2.0	0.9	2.7	4.0	2.06
침실수	0.8	1.4	1.0	0.8	2.0	0.0	0.6	1.0	1.64
침실외방수	0.8	1.6	1.1	0.6	8.1	0.9	0.3	0.6	16.02 *
거실수	0.5	0.8	0.5	0.5	1.0	0.9	0.0	0.6	1.09
식당수	2.5	28.5	2.6	0.8	18.2	1.8	0.4	2.9	578.91 *
부업형태	0.3	0.2	0.3	0.3	0.0	0.0	0.0	0.3	0.53
부업단독사용여부	1.3	1.2	1.8	1.1	4.0	0.0	1.2	1.7	4.59 *
수도형태	0.3	0.1	0.3	0.3	0.0	0.0	0.0	0.3	0.80
화장실형태	0.5	0.5	0.4	0.5	1.0	1.8	0.1	0.4	1.33
화장실단독사용여부	1.2	1.1	1.4	1.2	1.0	2.6	0.4	1.4	1.29
목욕시설형태	0.4	0.1	0.4	0.4	0.0	0.0	0.0	0.8	1.58
목욕시설단독사용여부	1.4	1.2	1.5	1.3	0.0	0.0	1.0	2.2	1.39
건물층수	0.7	1.0	0.6	0.7	1.0	0.9	0.3	0.6	0.77
층구분	6.6	0.2	6.7	6.8	3.0	3.5	3.9	14.0	23.56 *
거주층수	0.5	0.5	0.3	0.6	0.0	0.0	0.3	0.6	1.19
주거전용여부	1.9	0.5	1.9	1.9	1.0	0.0	2.2	3.7	4.33 *
점유형태	0.8	0.1	0.9	0.8	3.0	5.3	1.2	2.5	10.26 *
타지주택소유여부	1.6	0.1	1.5	1.5	25.3	9.6	1.5	2.9	71.36 *
주인가구여부	5.5	12.2	4.9	4.6	47.5	48.2	6.7	6.1	147.96 *
전체	30.2	53.7	30.4	27.5	118.2	76.3	22.8	46.3	36.45 *
표본 (사례수)	전체 (2,761)	인터넷 (126)	우편 (550)	대면 (1,978)	ARS (-)	CATI (5)	전화 (53)	배포 (49)	F-값
가구구분	1.6	1.6	2.5	1.3	-	0.0	0.0	4.1	1.87
침실수	0.3	0.8	0.2	0.4	-	0.0	0.0	0.0	0.39
침실외방수	0.7	0.0	2.2	0.4	-	0.0	0.0	0.0	5.21 *
거실수	0.4	0.0	0.9	0.3	-	0.0	0.0	0.0	1.51
식당수	4.9	33.1	7.5	2.4	-	0.0	0.0	10.2	70.69 *
부업형태	0.1	0.0	0.4	0.1	-	0.0	0.0	0.0	1.04
부업단독사용여부	1.1	0.0	3.1	0.7	-	0.0	0.0	0.0	6.25 *
수도형태	3.6	1.6	1.6	4.1	-	0.0	1.9	14.3	6.40 *
화장실형태	0.2	1.6	0.4	0.1	-	0.0	0.0	0.0	3.20 *
화장실단독사용여부	0.9	0.0	1.8	0.7	-	0.0	0.0	2.0	2.08
목욕시설형태	0.6	1.6	0.7	0.5	-	0.0	0.0	0.0	0.79
목욕시설단독사용여부	1.3	0.8	2.4	1.1	-	0.0	0.0	0.0	1.74
건물층수	0.5	0.8	0.7	0.4	-	0.0	0.0	2.0	1.11
층구분	5.7	0.0	8.4	4.9	-	20.0	3.8	6.1	4.62 *
거주층수	1.1	0.0	0.4	1.4	-	0.0	0.0	4.1	2.45 *
주거전용여부	0.8	0.0	1.5	0.7	-	0.0	0.0	2.0	1.47
점유형태	1.1	0.8	0.9	1.1	-	0.0	0.0	6.1	3.00 *
타지주택소유여부	0.5	0.0	0.4	0.4	-	40.0	0.0	2.0	1.02
주인가구여부	2.8	10.2	2.9	2.2	-	40.0	1.9	2.0	7.38 *
전체	27.9	53.2	38.7	22.9	-	100.0	7.5	55.1	13.13 *

주1) *, p<0.05 인.

주2) 표본조사의 CATI 경우, 사례수가 작아서 F-검정에서는 제외하였음.

■ 이 가구가 살고 있는 건물은 몇 층이며, 어디에 살고있습니까?

층 건물중

- 지하(반지하)
- 지상 층
- 옥상(옥탑)

[그림 2-24] 건물 및 거주층 항목의 조사표 예

■ 이 가구의 부엌, 수도, 화장실, 목욕 시설은 어떤 형태이며, 어떻게 사용하고 있습니까?

화장실	<ul style="list-style-type: none"> <input type="radio"/> 수세식 <input type="radio"/> 채래식 <input type="radio"/> 없음 	<input type="text"/> 이며	<ul style="list-style-type: none"> <input type="radio"/> 단독사용 <input type="radio"/> 공동사용
목욕시설	<ul style="list-style-type: none"> <input type="radio"/> 온수시설 <input type="radio"/> 비온수시설 <input type="radio"/> 없음 	<input type="text"/> 이며	<ul style="list-style-type: none"> <input type="radio"/> 단독사용 <input type="radio"/> 공동사용

[그림 2-25] 주거시설형태 항목의 조사표 예

다) 주택항목

주택항목의 경우(<표 2-30>), 전반적으로 보면 인구나 가구항목에 비해서 내검불일치율이 4, 5%에 근접한 항목이 눈에 많이 띄었다. 그 중에서도 총방수가 전수조사와 표본조사에서 가장 높은 내검불일치율을 보여주었다. 내검불일치율이 가장 낮은 항목은 단독주택유형과 기타거처 항목으로 1% 미만이었었는데 이것은 이 두 항목에 반드시 응답해야 하는 가구수가 매우 적었기 때문이었다.

자료수집방법에 따라서 주택항목의 내검불일치율은 차이가 있는가? 전수조사의 경우 모든 항목에서 내검불일치율이 가장 높게 나타난 자료수집방법은 ARS인 반면에 우편조사와 배포조사에서 내검불일치율이 낮게 나타났다. 표본조사의 경우에는 인터넷조사에서 내검불일치율이 높았으며, 우편조사에서의 내검불일치율이 낮게 나타났다. 전반적으로 전수조사보다 표본조사에서 내검불일치율이 높았는데 이것은 표본조사의 주택항목의 위치가 전수조사의 주택항목의 위치에 비해서 훨씬 더 뒤쪽에 위치하기 때문에 응답자들이 피로감을 느꼈기 때문으로 보인다.

〈표 2-30〉 자료수집방법별 주택항목 내검불일치율 비교

(단위: %)

전수 (사례수)	전체 (21,727)	인터넷 (905)	우편 (5,466)	대면 (13,940)	ARS (91)	CATI (104)	전화 (575)	배포 (646)	F-값
거처종류	1.3	2.3	1.0	1.4	5.5	1.9	1.4	1.2	3.75 *
단독거처	0.9	0.7	0.8	0.8	3.3	1.0	1.2	1.2	1.52
기타거처	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	2.00
주거용면적	1.8	5.4	1.3	1.7	8.8	3.9	2.3	1.2	17.06 *
총방수	7.1	11.4	5.4	7.4	16.5	12.5	8.0	4.0	13.26 *
총거실수	5.1	9.9	3.4	5.5	11.0	7.7	6.3	2.9	15.59 *
총식당수	4.6	10.6	3.3	4.7	12.1	4.8	4.5	3.3	18.54 *
건축연도	2.5	3.1	3.6	1.9	13.2	5.8	4.9	1.6	18.43 *
부엌수	6.7	9.3	4.4	7.4	14.3	5.8	7.8	3.3	15.23 *
화장실수	5.8	9.0	3.7	6.5	13.2	7.7	7.5	2.9	15.99 *
출입구수	5.9	9.2	3.8	6.6	13.2	6.7	6.4	3.4	14.84 *
전체	41.7	70.8	30.9	43.9	111.0	57.0	50.3	25.1	18.83 *
표본 (사례수)	전체 (2,338)	인터넷 (118)	우편 (505)	대면 (1,628)	ARS (-)	CATI (5)	전화 (38)	배포 (44)	F-값
거처종류	0.7	0.9	0.4	0.8	-	0.0	0.0	0.0	0.38
단독거처	0.6	0.9	0.4	0.7	-	0.0	0.0	0.0	0.27
기타거처	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	-
주거용면적	2.2	5.1	1.4	2.3	-	0.0	0.0	2.3	1.77
총방수	5.5	12.7	4.0	5.3	-	0.0	10.5	6.8	4.04 *
총거실수	4.0	8.5	2.6	4.1	-	0.0	7.9	2.3	2.69 *
총식당수	6.4	15.3	11.1	3.5	-	0.0	13.2	31.8	27.90 *
건축연도	1.9	0.0	0.6	2.5	-	0.0	0.0	0.0	2.99 *
부엌수	4.1	8.5	2.2	4.4	-	0.0	5.3	2.3	2.84 *
화장실수	3.5	7.6	2.2	3.7	-	0.0	5.3	0.0	2.66 *
출입구수	3.4	7.6	2.2	3.6	-	0.0	5.3	0.0	2.68 *
전체	32.3	67.5	26.9	30.8	-	0.0	47.4	45.5	3.26 *

주1) *, p<0.05 임.

주2) 표본조사의 CATI 경우, 사례수가 작아서 F-검정에서는 제외하였음.

2) 항목무응답률

자료수집방법에 따라서 항목무응답률은 어떠한 차이가 있는가? 자료수집방법을 구분하지 않은 전체의 항목무응답률은 전반적으로 1% 미만으로 낮게 나타났는데, 이는 항목무응답률을 인정하지 않는 조사관행과 이 자료가 1차적으로 내검을 거친 자료라는 점에서 기인하는 것으로 생각된다.²²⁾ 이러한 한계가 있는 상황에서 항목무응답률의 높고 낮음을 비교하는 데에는 무리가 있으며 다만, 자료수집방법별로 그 차이가 있는지만을 전수조사를 중심으로 개략적으로 살펴보겠다.²³⁾

<표 2-31>은 인구 항목무응답률을 자료수집방법별로 제시한 것이다. 전수조사의 경우, 항목무응답률은 졸업여부에서 가장 높았으며, 나이에서 가장 낮았다. 자료수집방법별로 인구항목의 무응답률을 비교해 본 결과 인터넷조사에서 항목무응답률은 0%로 나타났다. 이는 인터넷조사의 경우, 특정 항목에 응답하지 않을 경우에 다음 항목으로 넘어갈 수 없도록 한 조사표 설계 때문이다. 반면, 항목무응답률이 가장 높게 나온 자료수집방법은 조사원에 의한 전화조사로 성명과 나이 항목을 제외한 모든 인구항목에서 가장 높은 항목무응답률을 보여주었다. 조사원 전화조사를 제외하고는 ARS에서 항목무응답률이 높게 나타났다. 동일한 항목에 대해서 미국의 2003년 센서스 시험조사 항목무응답률을 비교해 보면, 성별의 경우 우편조사에서는 1.8%, 인터넷조사에서는 1.2%, 나이의 경우 우편조사에서는 1.7%, 인터넷의 경우에는 1.5%, 가구주와의 관계는 우편조사에서는 2.1%, 인터넷에서는 0.5%로 나타났다(Treat, 2005).

22) 따라서, 내검불일치율에는 항목무응답률 또한 반영이 되어 있다고 볼 수 있다.

23) 표본조사의 경우 인구, 가구, 주택의 항목무응답률을 자료수집방법별로 구분했을 경우에 항목무응답률이 거의 0%에 가깝게 나와서 실질적으로 비교의 의미가 없다.

〈표 2-31〉 자료수집방법별 인구항목 무응답률 비교

(단위: %)

구분 (사례수)	전체 (70,317)	인터넷 (3,146)	우편 (18,113)	대면 (44,454)	ARS (292)	CATI (368)	전화 (1,600)	배포 (2,344)	F-값
성명	0.1	0.0	0.0	0.1	3.8	0.0	0.3	0.0	79.83 *
성별	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	3.27 *
나이	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	51.26 *
생년	0.3	0.0	0.1	0.3	1.0	0.0	2.3	0.3	38.42 *
생월	0.5	0.0	0.2	0.6	1.0	0.0	3.1	0.4	45.78 *
생일	0.5	0.0	0.2	0.6	1.0	0.0	3.1	0.4	44.30 *
관계	0.0	0.0	0.0	0.0	1.0	0.0	0.2	0.0	17.94 *
교육수준	0.5	0.0	0.2	0.5	1.0	0.0	2.9	0.6	42.54 *
졸업여부	0.8	0.0	0.7	0.8	0.3	0.0	3.3	1.7	28.60 *
혼인상태	0.1	0.0	0.1	0.1	0.7	0.3	0.3	0.1	3.52 *
전체	2.9	0.0	1.6	3.1	11.0	0.3	15.6	3.6	58.41 *

주: *, p<0.05 임

<표 2-32>는 가구항목의 무응답률을 나타낸 것으로 자료수집방법별로 항목무응답률을 살펴보면, ARS에서 항목무응답률이 높게, 인터넷조사에서 항목무응답률이 낮게 나타났다. 주택항목의 경우(<표 2-33>)에는 ARS에서 항목무응답률이 높게, 대면조사에서 항목무응답률이 낮게 나타났다. 주택항목의 경우는 인구나 가구항목과 달리 항목무응답률이 다른 조사에 비해서 높은 편이다. 이것은 주택항목 그 자체가 어려워서라기 보다는 조사가 완료된 이후 내검과정에서 발생한 것으로 파악되었다. 조사 당시에는 세들어 살고 있는 가구이기 때문에 주택항목에 모두 응답을 하지 않았으나 내검과정에서 대표가구로 코드가 변경되면서 이후의 항목이 모두 무응답 처리가 되는 경우가 발생한 것이다. 인터넷조사의 경우에도 이와 같은 논리에 의해서 무응답률이 발생하였으며, 실제로 이 부분을 걸러줄 경우 항목무응답률이 사라지는 것을 확인할 수 있었다. 따라서 이것은 실제적으로 자료수집방법에 의한 것이라고 보기에 어려울 수도 있으나 동시에 응답자 기입방법의 한계로 볼 수도 있을 것이다.

〈표 2-32〉 자료수집방법별 가구항목 무응답률 비교

(단위: %)

구분 (사례수)	전체 (24,865)	인터넷 (971)	우편 (57,10)	대면 (16,572)	ARS (99)	CATI (114)	전화 (675)	배포 (724)	F-값
가구구분	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	1.03
침실수	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	42.11 *
침실외방수	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	41.55 *
거실수	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	13.83 *
식당수	0.1	0.0	0.2	0.1	2.0	0.0	0.0	0.4	7.02 *
부엌형태	0.5	0.0	0.5	0.5	1.0	0.0	1.0	1.4	3.67 *
_단독사용여부	0.5	0.0	0.4	0.4	1.0	0.0	1.0	1.4	3.97 *
수도형태	0.7	0.0	0.7	0.7	3.0	0.0	1.0	1.8	4.83 *
화장실형태	0.6	0.0	0.7	0.6	1.0	0.0	1.0	1.7	3.57 *
_단독사용여부	0.5	0.0	0.5	0.5	1.0	0.0	1.0	1.4	3.69 *
목욕시설형태	0.8	0.0	0.8	0.8	1.0	0.0	1.5	1.9	3.86 *
_단독사용여부	0.8	0.0	0.7	0.8	1.0	0.0	1.3	1.8	3.63 *
건물층수	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	42.1 *
층구분	0.1	0.0	0.1	0.1	1.0	0.0	0.0	0.1	1.54
거주층수	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	42.1 *
주거전용여부	0.2	0.0	0.2	0.2	1.0	0.0	0.4	0.3	1.60
점유형태	0.1	0.0	0.1	0.1	1.0	0.0	0.3	0.1	1.49
타지주택소유여부	0.4	0.0	0.2	0.4	7.1	2.6	0.9	0.1	23.52 *
주인가구여부	0.2	0.0	0.2	0.2	1.0	0.0	0.4	0.1	1.65
전체	5.6	0.0	5.5	5.3	29.3	2.6	10.2	12.6	7.72 *

주: *, p<0.05 임.

〈표 2-33〉 자료수집방법별 주택항목 무응답률 비교

(단위: %)

전체 (사례수)	전체 (21,727)	인터넷 (905)	우편 (5,466)	대면 (13,940)	ARS (91)	CATI (104)	전화 (575)	배포 (646)	F-값
거처종류	1.0	3.9	0.8	0.8	3.3	1.9	0.5	0.8	16.09 *
단독거처	1.0	4.0	0.8	0.8	3.3	1.9	0.5	0.8	16.72 *
기타거처	0.1	0.0	0.1	0.2	0.0	0.0	0.4	0.0	1.23
주거용연면적	0.8	4.0	0.7	0.6	5.5	1.9	0.5	0.8	24.77 *
총방수	0.9	4.0	0.8	0.7	5.5	2.9	0.7	0.8	23.48 *
총거실수	0.9	4.0	0.8	0.7	6.6	2.9	0.7	0.8	25.41 *
총식당수	1.5	5.3	1.4	1.2	8.8	5.8	1.9	1.4	24.82 *
건축연도	1.2	4.0	1.0	1.1	6.6	5.8	1.6	0.8	17.51 *
부엌수	0.9	4.0	0.8	0.7	5.5	3.9	0.7	0.8	23.41 *
화장실수	0.9	4.0	0.8	0.7	5.5	3.9	0.7	0.8	23.41 *
출입구수	0.9	4.0	0.8	0.7	5.5	3.9	0.7	0.8	23.41 *
전체	9.9	40.1	8.7	8.0	56.0	34.6	8.9	8.4	24.00 *

주: *, p<0.05 임.

자료수집방법에 따른 내검불일치율과 항목무응답률을 분석해 본 결과를 요약해 보면 다음과 같다. 전반적으로, 내검불일치율 및 항목무응답률이 가장 높은 자료수집방법은 ARS로 나타났다. ARS에서 내검불일치율이나 항목무응답률이 높게 나타난 이유는 생년월일 등 개인정보를 전화라는 매체를 이용해서 응답하는 것에 대해서 부담감을 느낀 것으로 추정되는 동시에, ARS 음성인식 방식의 경우 음성인식이 불완전하고 정확하지 않을 경우 조사결과에 대한 인식율이 떨어졌기 때문이다. 이러한 결과는 미국 센서스국의 결과와도 유사한데, 2003년 센서스 시험 조사 결과에 의하면 IVR에서 그 비율이 높게 나타났다. 이것은 특정 조사방법에서 항목이 어렵다기 보다는 시스템 상의 문제와 연결되는 것으로 해석되어 있다(Treat, 2005).

그런데, 내검불일치율이나 항목무응답률이 상대적으로 낮은 자료수집방법에서는 그 경향이 인구, 가구 주택항목 유형별로 차별적인 것으

로 나타났다. 인구항목의 경우에는 인터넷조사에서 내검불일치율과 항목무응답률이 가장 낮았으나 가구항목과 주택항목에서는 전화조사나 대면조사에서 낮게 나타났다. 이러한 차이가 발생하는 이유는 조사표 상의 배열문제가 관련된 것으로 보인다. 인터넷조사 자체에 내검프로그램을 장착했기 때문에 전반적으로 인터넷조사에서 내검불일치율이나 항목무응답률이 낮게 나타나기는 하지만, 조사표 뒤로 갈수록 응답에 대한 성실도는 조사원 기입식의 자료수집방법에서 좀 더 높은 것으로 나타났다. 이러한 이유와 함께 가구나 주택항목의 경우, 그 개념이 어려워서 응답자 혼자서 조사표를 완성하는 것이 쉽지 않았던 것으로 생각된다.

다. 자료수집방법별 응답자 특성 분석

지금까지 자료수집방법에 따라서 응답률, 내검불일치율, 그리고 항목무응답률에 차이가 있음을 살펴보았다. 여기에서는 마지막으로 자료수집방법별 응답자 특성을 분석해 보았다. 이는 자료수집방법에 따라서 커버하는 집단에 차이가 있을 수 있다는 기존의 연구결과(Aquilino and Lo Sciuto, 1990 등)와 응답자들마다 선호하는 응답방식에서 차이가 있을 것이라는 가정(Dillman and Tarnai, 1988)에 따른 것이다. 특히 후자는 여러가지 자료수집방법이 동시에 제공되었을 때, 특정의 방법을 선호하는 집단은 그렇지 않은 집단과는 선별적인 특성을 갖게 됨을 의미하는 것으로 방법효과 분석에서 이 점을 고려해야 한다는 것은 앞선 연구들에서 이미 지적되어온 바 있다(Bates, 1995; Biemer, 2001; Wright et al, 2001 등). 이에 자료수집방법에 따라서 응답자들의 특성이 어떻게 다른지 살펴보고, 이 특성들이 응답률, 내검불일치율 및 항목무응답률을 결정하는데 어떠한 영향을 미치는지 살펴보고자 한다. 자료수집방법은 인구주택총조사에서 기본적인 방법이 될 인터넷조사와 우편조사, 그리고 대면조사를 중심으로 살펴보았다.

분석은 각각의 자료수집방법 응답자의 특성을 개인 특성과 가구 특성으로 구분한 후 이들의 평균을 비교해보고, 차이가 있는 변수들을 중심으로 로지스틱회귀분석을 실시하는 방향으로 진행하였다. 조사표에 응답자가 누구인지를 기록하지 않은 사례는 분석에서 제외하였다. 전수

조사와 표본조사 모두 동일한 경향이 발견되어 분석은 전체를 대상으로 실시되었고, 로지스틱 회귀분석에서는 전수조사 여부를 더미처리하였다.

자료수집방법에 따라서 응답자의 특성에 차이가 있는가?

<표 2-34>에 따르면, 성별, 연령, 교육수준, 혼인상태 등 응답자의 개인적인 특성과 이 응답자들이 속한 가구의 특성은 자료수집방법 간에 뚜렷한 차이를 보였다. 인터넷조사, 우편조사, 대면조사 가운데 응답자가 여성인 경우가 가장 많은 조사는 대면조사이며, 다음으로 우편조사였다. 한편 남성인 경우가 많은 조사는 인터넷조사로 나타났다. 이러한 결과는 우편조사와 인터넷조사를 비교했을 경우 인터넷조사에서 남성비율이 우편조사보다 더 높았던 2000년과 2003년 미국 센서스 시험조사 결과와 유사하다(Schneider et al., 2005; Treat, 2005). 연령별로 보더라도 그 차이는 뚜렷한데, 대면조사, 우편조사, 인터넷조사 순으로 응답자들의 나이가 많은 것으로 나타났다. Treat(2005)의 연구에서 보면, 나이가 젊을수록 인터넷조사에 응답할 가능성이 높은 것으로 나타났다. Schneider와 동료들의 연구(2005)에서도 유사한 경향이 나타났는데 인터넷조사 응답자의 평균연령은 43.9세이며, 우편조사는 51.4세로 우리나라와 비교해 보면 우리나라의 해당 자료수집방법 응답자의 평균연령이 약간 더 낮은 것으로 나타났다. 반면에 교육수준은 성별이나 연령효과와는 역순이었다. 인터넷조사 응답자의 교육수준이 가장 높았으며 다음으로 우편조사와 대면조사 순이었다. 인터넷조사 응답자들의 평균 교육수준은 4년제 미만의 대학 정도였으며, 우편조사는 고등학교, 대면조사는 중학교 정도 수준이었다. 이는 대면조사 응답자의 평균연령이 높은 것과도 연관된 것으로 보인다. 혼인상태에서 또한 차이가 발견되었는데 그 비율로 환산해 보면 배우자가 있는 경우가 우편조사에서는 80% 정도로 가장 높게 나타났으며 다음으로 인터넷조사와 대면조사 순으로 배우자가 있는 비율이 높았다.

〈표 2-34〉 인터넷조사, 우편조사, 대면조사 응답자 및 가구 특성 비교

구분	잠재적 무응답자	응답자	응답자			F-값
			인터넷조사	우편조사	대면조사	
여성비율	0.47 (0.02)	0.57 (0.00)	0.43 (0.02)	0.55 (0.01)	0.59 (0.00)	65.46 *
연령	46.20 (0.56)	47.31 (0.10)	39.72 (0.33)	46.77 (0.18)	48.16 (0.13)	165.86 *
교육수준	4.22 (0.06)	4.08 (0.01)	5.04 (0.04)	4.40 (0.02)	3.88 (0.01)	529.60 *
배우자유무	0.56 (0.02)	0.66 (0.00)	0.70 (0.01)	0.80 (0.01)	0.61 (0.00)	348.87 *
가구내 여성비	0.48 (0.01)	0.51 (0.00)	0.50 (0.01)	0.51 (0.00)	0.52 (0.00)	2.36
가구내 평균연령	40.91 (0.53)	41.76 (0.10)	35.23 (0.37)	39.78 (0.19)	43.07 (0.12)	199.18 *
가구내 평균교육수준	4.08 (0.05)	3.75 (0.01)	4.07 (0.03)	3.83 (0.01)	3.69 (0.01)	73.11 *
가구내 25~64세 인구비율	0.77 (0.01)	0.64 (0.00)	0.67 (0.01)	0.63 (0.00)	0.65 (0.00)	8.20 *
일인가구비	0.35 (0.02)	0.20 (0.00)	0.11 (0.01)	0.10 (0.00)	0.25 (0.00)	368.02 *
자가소유비	0.44 (0.02)	0.55 (0.00)	0.60 (0.01)	0.63 (0.01)	0.51 (0.00)	132.97 *
아파트거주비	0.57 (0.02)	0.48 (0.00)	0.57 (0.02)	0.71 (0.01)	0.38 (0.00)	1059.67 *
비농림어가비	0.91 (0.01)	0.95 (0.00)	0.97 (0.00)	0.97 (0.00)	0.94 (0.00)	52.99 *

주1) *, p<0.05 임.

주2) 인터넷조사 1,097가구; 우편조사 6,260가구; 대면조사 18,550가구 중 응답자의 수는 항목에 따라서 인터넷조사는 최소1,086~최대1,097가구이며, 우편조사는 최소5,569~최대6,260가구이며, 대면조사는 최소16,985~최대18,550가구임.

응답자가 속한 가구의 특성도 가구내 여성비율을 제외하고는 자료수집방법에 따라서 통계적으로 유의미한 차이가 나타났다. 가구내 평균연령은 대면조사가 43.1세로 가장 높았으며 인터넷조사가 35.2세로 가장 낮았다. 가구내 평균교육수준은 응답자 수준보다는 약간 낮아졌으나 여전히 인터넷조사에서 교육수준이 가장 높았으며 대면조사가 가장 낮았다. 단독가구여부도 중요한 변수였는데 대면조사에서 일인가구인 경우가 25%가량으로 가장 높았으며 인터넷이나 우편조사에서는 약 10%대로 낮았다. 자가소유여부의 경우에도 우편조사에서 자가소유비율이 가장 높았으며 대면조사에서 가장 낮았다. 아파트 거주여부 또한 중요한 구별 변수 중의 하나인데 우편조사에서 아파트 거주비율이 71% 가량으로 높았으나 대면조사에서는 38%로 낮았다.

한편, 잠재적인 무응답자라고 할 수 있는 전화조사 응답자의 특성을 전체 응답자 집단과 함께 표에 제시해 보았다. 이것은 혼합방법조사가 과연 무응답오차를 줄여줄 수 있는가를 탐색해 보기 위한 것이다. 두드러진 차이가 나타나는 변수를 중심으로 살펴보면, 잠재적인 무응답자 집단은 응답자집단에 비해서 여성인 경우가 적고, 배우자가 있는 경우 또한 적었다. 그리고 가구내 평균교육수준과 일인가구의 비율이 응답자집단에 비해서 높았으며, 자가소유비율은 약 11%가량 더 적었다. 구체적으로 인터넷조사, 우편조사, 대면조사 응답자집단과 잠재적인 무응답자 집단을 비교해 보면, 성별, 연령, 교육수준, 가구내 평균연령, 아파트거주여부 등은 세 응답자 집단의 범위에 포함되어 있다. 이 중에서도 성별과 연령, 아파트 거주여부는 인터넷조사와 우편조사의 중간에, 교육수준과 가구내 평균연령은 우편조사와 대면조사의 중간에 위치하되 우편조사에 근접한 특성을 보이고 있다. 반면에, 배우자유무, 가구내 여성비율, 가구내 평균교육수준, 가구내 25~64세 인구비율, 일인가구여부, 자가소유여부, 비농림어가여부는 세 응답자 집단의 범주에 포함되지 않는 구별되는 특성을 가지고 있음을 알 수 있었다.

응답자의 특성을 통제한 이후에도 자료수집방법이 내검불일치율에 미치는 효과는 유의미한가?

그렇다면, 응답자의 이러한 특성이 자료의 품질지표인 내검불일치율이나 항목무응답률에 영향을 미치는가? 앞서 자료수집방법에 따라 내검불일치율과 항목무응답률에 차이가 있는 것으로 나타났는데 각 집단이 갖고 있는 선별적인 특성을 통제하면 자료수집방법의 효과는 사라지는가? 여기에서는 위의 응답자 특성과 함께 자료수집방법을 독립변수로, 내검불일치여부를 종속변수로 투입한 후 로지스틱회귀분석을 실시하였다. 내검불일치여부는 가구항목에 한정하였다. 그 결과, 자료수집방법의 효과가 95% 신뢰수준에서 통계적으로 유의미한 것으로 나타난 항목은 <표 2-35>와 같다. 이 항목들은 <표 2-29>에서 응답자 변수 및 가구 특성을 통제하지 않았을 때와 유사함을 알 수 있다. 이것은 응답자 및 가구 특성과 무관하게 내검불일치율이 자료수집방법에 따라서 차이가 있음을 의미한다.

<표 2-35> 내검불일치여부에 대한 로지스틱 회귀분석결과 요약

구분	인터넷조사	우편조사	대면조사	배포조사
가구구분	×	○	○	○
침실외방수	×	×	○	×
식당수	○	×	○	×
목욕시설단독사용여부	×	×	×	○
층구분	○	○	○	○
주거전용여부	○	×	×	○
점유형태	○	○	○	○
타지주택소유여부	○	○	○	×
주인가구여부	○	○	○	○

주: 자료수집방법에 대한 기준변수는 CATI, ARS, 전화조사를 통합한 변수임.

제4절 요약 및 결론

1. 분석결과 요약

지금까지 혼합방법조사에서 발생할 수 있는 방법효과 검토의 일환으로 2010 인구주택총조사 제2차 시험조사 자료를 이용하여 자료수집방법 별로 응답률, 내검불일치율, 항목무응답률에 어떠한 차이가 있는지를 살펴보았다. 분석을 통해서 나타난 주요 결과는 다음과 같다.

첫째, 응답자들의 조사편의 및 비용감소를 목적으로 제공한 비방문조사(인터넷조사, 우편조사, CATI, ARS) 중 응답자들은 우편조사 → 인터넷조사 순으로 응답을 많이 하는 것으로 나타났다. 조사의 전 기간 동안 방문조사만으로 이루어진 통제집단이 없기 때문에 비방문조사가 전체 응답률을 향상시켰는지에 대해서 검증하기는 어렵다. 다만, 여기에서는 비방문조사 응답률과 전체 응답률 간의 관계를 살펴봄으로써 대안적인 답을 찾아보면 비방문조사 응답률이 높은 곳이 전체 응답률이 높다는 일관적인 결과를 발견하기는 어려웠다. 오히려 특정 지역에서는 비방문조사 응답률이 높은 곳에서 방문조사 응답률이 낮게 나타나기도 하였다. 이는 비방문조사 응답률이 전체 응답률을 향상시켰는가에 대해서는 오히려 비방문조사에 의해 증가된 응답률이 방문조사 응답률의 감소에 의해서 상쇄되는 경향을 보여주는 것이다. 이에 따라서 조사구 수준에서 응답률을 결정하는 요인에 대한 중다회귀분석을 실시해 보았는데, 예상대로 전체 응답률과 비방문조사 응답률을 결정하는 요인이 차별적인 것으로 나타났다. 여성비율은 전체 응답률에서만, 노동가능인구비율, 인구밀도, 아파트조사구여부는 비방문조사 응답률에서만 유의미한 영향을 미쳤다. 이 중 아파트 조사구의 효과는 다른 변수들에 비해서 큰 효과를 갖는 것으로 나타났다. 이러한 차별적인 효과는 비방문조사 응답률이 높다고 해서 반드시 전체 응답률이 높지 않다는 것을 지지해주는 것으로 해석될 수 있다. 그러나, 우리나라의 경우 전반적으로 응답률이 매우 높은 상태에서 비방문조사 응답률이 높을 경우 방문조사 응답률이 낮아지는 경향은 어느 정도 받아들여질 수도 있는 부분이므로 해석상의 주의를 요한다. 한편, 비방문조사 응답률은 인센티브를 제공한 지역이 그렇

지 않은 지역보다 응답률이 높았으며, 인터넷조사와 우편조사 중에서는 우편조사 응답률이 미치는 효과가 더 큰 것으로 나타났다. 특히 아파트 조사구의 우편조사 응답률의 효과가 컸다.

둘째, 내검불일치율 및 항목무응답률은 자료수집방법과 항목의 유형에 따라서 차이가 있었다. 전반적으로 내검불일치율 및 항목무응답률이 높게 나타난 자료수집방법은 ARS 및 CATI 등의 전화조사이다. 이것은 실제 응답자들이 개인정보를 전화로 응답한다는 것에 상당한 부담감을 느꼈으며, 특히 음성인식 방식의 ARS는 음성인식이 완전하지 않을 경우에 조사결과에 대한 인식율이 떨어진 것으로 보여진다. 내검불일치율 및 항목무응답률이 낮게 나타난 자료수집방법은 항목의 유형에 따라서 차별적이었다. 인구항목의 경우, 인터넷조사에서 내검일치율 및 항목무응답률이 낮게 나타났다. 이는 인터넷조사에 자동내검프로그램을 장착하였기 때문이다. 반면에 인구항목과 달리 가구항목 및 주택항목에 대해서는 대면조사나 전화조사에서 내검불일치율 및 항목무응답률이 높게 나타났는데 이는 조사표의 항목배열상 뒤쪽에 있는 이 항목에 대한 응답의 완전성이 인터넷 등과 같은 비방문조사에서 떨어질 우려가 있을 뿐 아니라 조사항목에 대한 이해 부족으로 인해서 응답자 혼자 조사표를 작성하는 데 있어서의 어려움 때문으로 추정된다.

셋째, 응답자들마다 선호하는 응답방식이 차이가 있으므로 자료수집 방법별 응답자 특성이 차별적일 것이며 방법효과를 분석함에 있어서 이들의 특성을 통제해야 한다는 기존의 연구 결과에 따라서 인터넷조사, 우편조사, 대면조사 응답자의 개인 및 가구 수준 특성을 살펴보았다. 그 결과, 응답자의 성별, 연령, 교육수준, 혼인상태에서 자료수집방법 간에 뚜렷한 차이를 보였다. 인터넷조사, 우편조사, 대면조사 가운데에서 응답자가 여성인 경우가 가장 많은 조사는 대면조사이며, 남성인 경우가 많은 조사는 인터넷조사로 나타났다. 연령별로 보더라도 그 차이는 뚜렷한데, 대면조사, 우편조사, 인터넷조사 순으로 응답자들의 나이가 많았다. 반면에 교육수준은 인터넷조사가 대면조사에 비해서 높았다. 응답자가 속한 가구의 특성도 자료수집방법에 따라서 통계적으로 유의미한 차이가 나타났다. 1인가구비율은 대면조사에서 25%가량으로 가장 높았으며 인터넷이나 우편조사에서는 약 10%대로 낮게 나타났다. 아파트 거주

여부 또한 중요한 구별 변수 중의 하나인데 우편조사에서 아파트 거주 비율이 71% 가량으로 높게 나타났으며 대면조사에서 38%로 낮게 나타났다.

그렇다면, 응답자의 이러한 특성이 내검불일치율에 영향을 미치는가? 앞서 자료수집방법에 따라 내검불일치율에 차이가 있는 것으로 나타났다는데 각 집단이 갖고 있는 선별적인 특성을 통제하면 자료수집방법의 효과는 사라지는가? 응답자 특성과 함께 자료수집방법을 독립변수로, 내검불일치여부를 종속변수로 투입한 후 로지스틱회귀분석을 실시해 본 결과 여전히 자료수집방법이 내검불일치여부에 영향을 미치는 것으로 나타났다. 한편, 잠재적인 무응답자라고 할 수 있는 전화조사 응답자의 특성을 살펴본 결과, 응답자 집단에 비해서 여성인 경우가 적고, 배우자가 있는 경우가 적었다. 그리고 가구내 평균교육수준과 일인가구의 비율이 응답자집단에 비해서 높았으며, 자가소유비율은 약 11%가량 더 적었다.

2. 정책적 함의

이러한 결과는 다음과 같은 정책적 함의를 내포하고 있다. 먼저, 비방문조사 방법 도입의 타당성 측면을 생각해 보자. 첫째, 이번 분석결과는 ARS나 CATI와 같은 조사시스템을 도입하기에는 아직 여러가지 측면에서 어려움이 따른다는 것을 시사하고 있다. 두 조사 모두 약 0.5% 정도의 응답률을 얻지 못했을 뿐 아니라 다른 자료수집방법과 비교해 볼 경우 내검불일치율이나 항목무응답률도 높은 것으로 나타났다. 둘째, 인터넷 및 우편조사의 경우 내검불일치율 및 항목무응답률이 높은 편은 아니다. 또한 두 자료수집방법에 의한 응답률이 6개 시험조사지역 평균 26%가량으로 나타났다. 비록, 비방문조사 응답률이 전체 응답률 향상에 기여한다는 일관적인 근거를 찾기는 어려우나 전체적인 분석결과를 볼 때 인터넷 및 우편조사 도입에 따른 응답률 향상으로 인한 무응답 오차 감소보다는 자료의 품질개선 및 비용감소의 효과를 기대할 수 있을 것으로 생각된다. 단, 여기에는 비방문조사를 위한 비용을 고려해야 할 것이다. 비방문조사 시스템 구축에 초기 비용이 많이 들어간다면 이 초기

비용을 상쇄할 만큼의 응답률 향상이 있어야 한다는 점을 염두해 두어야 할 것이다.

둘째, 자료의 품질향상을 위한 조사표 개선 측면에서 살펴보자. 인터넷조사에서 인구항목과 달리 가구나 주택항목에서는 상대적으로 내검불일치율이나 항목무응답이 많이 나온 것과 대면조사에서 가구나 주택항목에서 그 비율이 낮게 나온 것은 조사표 개선의 필요성을 보여주는 것이다. 가구나 주택항목의 경우 인구항목에 비해서 상대적으로 뒤쪽에 위치해 있기 때문에 응답 피로가 누적되어 있을 수도 있겠으나, 조사원 기입 방식인 대면조사나 전화조사에 비해서도 내검불일치율 및 항목무응답률이 높다는 것은 해당 내용을 인터넷을 통해서 스스로 기입하는데에는 어려움이 있다는 것을 시사하는 것이므로 인터넷조사표 환경 개선의 필요성을 제기한다. 이와 함께 내검불일치율 및 항목무응답률이 높게 나온 항목을 중심으로 조사표 개선이 이루어져야 할 것이다. 청구분 항목의 경우 배포 및 대면조사 등에서 내검불일치율이 높게 나타나고 있으므로 전반적인 자료의 품질향상을 위해서 이 부분에 대한 개선이 필요할 것으로 보인다.

셋째, 이 연구가 효율적인 운영을 위한 비방문조사 응답률 추정모형에 대해서 갖는 함의는 어떠한가? 비방문조사 응답률이 비용감소로 연결되기 위해서는 비방문조사 시스템을 구축하기 위한 초기비용과 방문조사 업무량에 대한 정확한 측정을 통해 조사원 채용 및 기타 비용을 사전에 고려해야 한다. 이와 관련하여 비방문조사 응답률에 대한 예측이 중요한 과제로 부가된다. 비방문조사 응답률은 조사구에 따라서 그 변이가 매우 상이하며, 조사원의 업무량이 조사구 단위로 배분되므로 조사구 수준에서 응답률을 추정해야 한다. 조사표 항목 중심으로 회귀식을 도출해 본 결과, 인터넷조사 응답률의 28%, 우편조사 응답률의 64%가량이 설명되는 것으로 나타나 조사표에 있는 항목만으로 응답률을 추정하는 데에는 한계가 있음을 알 수 있다. 그러므로 향후 시험조사를 통해서 좀 더 정교한 응답률 추정공식의 개발이 필요하며 무엇보다도 인터넷 및 우편조사 응답률의 분산 중 설명되지 않은 부분에 대해서 설명력이 강한 핵심지표를 찾아내는 일이 시급하다. 본 연구에서 표본조사 항목 중에서 개략적으로 검토해 본 바 있는 컴퓨터 보유여부나

비이동률 등이 그 예가 될 수 있다.

3. 연구의 한계

분석결과와 관련한 자료의 한계점은 다음과 같다. 첫째는 조사지역의 대표성과 관련된 문제이다. 시험조사지역으로 선정된 3개 시도의 6개 시군구 분석결과를 전국으로 일반화하는 데에는 한계가 있다. 또한 시험조사과정은 여러가지 방법론들이 테스트 되므로 인구주택총조사 본조사와 동일한 세팅과정을 갖고 있지 않다. 따라서 자료를 해석함에 있어서 이 부분을 충분히 감안해야 할 것이다.

둘째, 조사설계과정에서 비방문조사가 과연 응답률을 향상시켜주었는지 등에 대한 질문에 대한 평가가 이루어지기 위해서는 통제집단과의 비교를 해야 한다. 그러나 조사설계과정에서 비방문조사에 대한 통제집단이 설정되지 않아서 이러한 효과를 분석하는 데에는 한계가 있었다.

셋째, 본 연구에서 자료의 품질지표로 측정하고자 하는 항목무응답률이나 내검불일치율의 경우, 자료수집방법별로 자체 내검 로직을 갖고 있는 경우도 있고 그렇지 않은 경우도 있다. 항목무응답률의 경우 1차 내검과정을 거친 자료에서 측정하였으므로 과소측정되었을 가능성이 있다. 자료입력방법에 따른 입력과정의 오류가 함께 반영되어 있으므로 자료수집방법이 자료의 품질에 미치는 효과를 분석하여 그 결과를 해석함에 있어서 주의를 기울여야 할 것이다.

참고문헌

- 강현철 · 한상태 · 김지연 · 정용찬 · 허명희(2008), RDD 전화조사와 주요 결과. 조사연구9(1): pp.1-22.
- 성태제(2007), SPSS/AMOS를 이용한 알기쉬운 통계분석-기술통계에서 구조방정식모형까지. 학지사.
- 통계청(2006a), 인터넷조사방법 평가조사 결과 보고서. 내부자료.
- _____ (2006b), 2005 인구주택총조사 종합평가보고서. 내부자료.
- _____ (2006c), 2005 인구주택총조사 종합평가보고서. 내부자료.
- _____ (2007), 2010 인구주택총조사 제1차 시험조사 추진 기본계획(안). 내부자료.
- _____ (2008a), 효율적인 현장조사를 위한 조사방법 개선 추진계획. 내부문서.
- _____ (2008b), 2010 인구주택총조사 조사지침서: 제2차 시험조사.
- _____ (2008c), 제2차 시험조사 인센티브 적용결과. 내부문서.
- AAPOR(2006), *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys, 4th edition*, Lenexa, Kansas: AAPOR.
- Aquilino, W. S.(1992), "Telephone versus face-to-face interviewing for household drug use surveys." *International Journal of the Addictions* 27: pp.71-91.
- Aquilino, W.S. and L.A. Lo Sciuto(1990), "Effects of interview mode on self-reported drug use." *Public Opinion Quarterly* 54: pp.362-395.
- Atrostic, B. K., N. Bates, G. Burt, and A. Silberstein(2001), "Nonresponse in U.S. government household surveys: consistent measures, recent trends, and new insights." *Journal of Official Statistics* 17(2): pp.209-226.
- Bates, N.(1993), "Data quality issues in a multi-mode Census: results from the Mail and Telephone Mode Test(MTMT)." *Joint Statistical Meetings-Section on Survey Research Method*, pp.736-741

- Biemer, P. P.(1988), “Measuring data quality.” *Telephone Survey Methodology*, pp.273-282.
- Biemer, P. and L. Lyberg(2003), *Introduction to Survey Quality*. Wiley-Interscience.
- Brady, S. E., C. N. Stapleton, J. A. Bouffard and J. D. Imel(2003), “Effects of alternative data collection modes on cooperation rates and data quality.” *Joint Statistical Meetings-Section on Survey Research Methods*: pp.693-700.
- Catlin, G. and S. Ingram(1988), “The effects of CATI on costs and data quality: a comparison of CATI and paper methods in centralized interviewing.” *Telephone Survey Methodology*, pp.437-450.
- Cook, C., F. Heath, and R. Thompson(2000), “A meta-analysis of response rates in web-or internet- based surveys.” *Educational and Psychological Measurement* 60(6): pp.821-836.
- Couper, M., J. Blair, and T. Triplett(1999), “A comparison of mail and e-mail for a survey of employees in U.S. Statistical Agencies.” *Journal of Official Statistics* 15(1): pp.39-56.
- Couper, M. P., E. Singer, and R. A. Kulka(1998), “Participation in the 1990 Decennial Census: Politics, Privacy, Pressures.” *American Politics Quarterly* 26: pp.59-80.
- Curtin, R., S. Presser, and E. Singer(2002), “The Effects of Response Rate Changes on the Index of Consumer Sentiment.” *Public Opinion Quarterly* 64: pp.413-428.
- de Leeuw, E. D. and M. Collins(1997), “Data collection methods and survey quality: an overview.”, pp.199-220 in Lyberg et al.(eds.). *Survey Measurement and Process Quality*. A Wiley-Interscience Publications.
- de Leeuw, E. D. and J. van der Zouwen(1988), “Data quality in telephone and face to face surveys: a comparative meta analysis.” *Telephone Survey Methodology*, pp.283-299.
- de Leeuw, E. D.(2005), “To mix or not to mix data collection modes in surveys.” *Journal of Official Statistics* 21(2): pp.233-255.

- Dillman, D. A. and J. Tarnai(1998), "Administrative issues in mixed mode surveys.", pp.509-528 in R. M. Groves(eds.) *Telephone Survey Methodology*. John Wiley and Sons, Inc.
- Dillman, D. A. , K. K. West, and J. R. Clark(1994), "Influence of an invitation to answer by telephone on response to census questionnaire." *Public Opinion Quarterly* 58: pp.557-568.
- Dillman, D. A.(2001), *Mail and Internet Surveys: The tailored design method* second edition. John Wiley and Sons, Inc.
- Griffin, D. H., D. P. Fischer, M. T. Morgan(2001), "Testing an internet response option for the American Community Survey." Presented at the annual conference for the American Association for Public Opinion Research, Montreal, Quebec, Canada May 17-20, 2001.
- Groves, R. M., F. J. Flower, Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau(2004), *Survey Methodology*. John Wiley and Sons, Inc.
- Groves, R. M., R. B. Cialdini, and M. P. Couper(1992), "Understanding the decision to participated in a survey." *Public Opinion Quarterly* 56: pp.475-495.
- Guarino, J.(2001), *Assessing the Impact of Differential Incentives and Alternative Data Collection Modes on Census Response*. Census 2000 Testing, Experimentation, and Evaluation Program Final Report. U.S. Census Bureau.
- Hochstim, J.(1967), "A Critical comparison of the strategies of collecting data from households." *Journal of the American Statistical Association* 62: pp.976-989.
- Hox, J. J. and E. D. de Leeuw(1994), "A comparison of nonresponse in mail, telephone, and face-to-face surveys." *Quality and Quantity* 28: pp.329-344.
- Johnson, T. P. Cho, Young Ik, R. T. Campbell, and A. L. Holbrook(2006), "Using community-level correlates to evaluate nonresponse effects in a telephone survey." *Public Opinion Quarterly* 70(5): pp.704-719.

- Krysan, M., H. Schuman, L. Scott, and P. Beatty.(1994) “Response rates and response content in mail versus face-to-face surveys.” *Public Opinion Quarterly* 58: pp.381-399.
- NRC(2004), *The 2000 Census Counting Under Adversity*. National Academies Press.
- _____(2004), *Reengineering the 2010 Census: Risks and Challenges*. National Academies Press.
- Roberts, C.(2007), “Mixing modes of data collection in surveys: a methodological review.” *ESRC National Center for Research Methods Review Papers*. NCRM/008.
- Sampson, R. J., J. D. Morenoff, and F. Earls.(1999), “Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children.” *American Sociological Review* 64: pp.633-660.
- Schneider, S., D. Canter, P. Segel, C. Arieira, L. Nguyen(2002), *Response Mode and Incentive Experiment for Census 2000 : Census 2000 Testing, Experimentation, and Evaluation Program*. U.S. Census Bureau.
- Treat, J. B.(2003), *Response Rates and Behavior Analysis*. Census 2000 Testing, Experimentation, and Evaluation Program. U.S. Census Bureau.
- Treat, J. B.(2005), “Research into the use of electronic data collection modes for the 2010 United States Decennial Census.” <http://panda.hyperlink.cz/cestapdf/pdf05c6/treat.pdf>.
- U. S. Census Bureau(2000), *Census 2000 Operational Plan*. U.S. Census Bureau.
- _____(2003), *Meeting 21st Century Demographic Data Needs-Implementing the American Community Survey*. U.S. Census Bureau.
- _____(2004), *Results from the Response Mode and Incentive Experiment in 2000*. Census 2000 Synthesis Report NO. 18. U.S. Census Bureau.

_____ (2006), *Design and Methodology: American Community Survey*. Technical Paper 67-unedited Version. U.S. Census Bureau.

Voogt, R. J. and W. E. Saris(2005), "Mixed mode designs: finding the balance between nonresponse bias and mode effects." *Journal of Official Statistics* 21(3): pp.367-387.

Wright, D. L., W. S. Aquilino, and A. J. Supple(2001), "A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use." *Public Opinion Quarterly* 62: pp.331-353.

통계청 홈페이지(www.nso.go.kr).

<http://qb.soc.surrey.ac.uk/docs/methods.htm>

<부 록>

<부록 1> 비방문조사 응답률의 지역별 비교

구분	전수		표본	
	차이	95% C.I.	차이	95% C.I.
기흥동 - 광남동	2.35	-0.58~5.29	-4.30	-12.77~4.17
기흥동 - 인제읍	7.82*	4.03~11.62	5.58	-5.09~16.25
기흥동 - 명장1동	12.29*	9.13~15.44	-0.59	-9.64~8.47
기흥동 - 목호동	21.68*	17.26~26.10	17.05*	5.23~28.86
기흥동 - 범천1동	22.96*	19.64~26.29	18.68*	9.45~27.90
광남동 - 인제읍	5.47*	2.15~8.79	9.88*	0.32~19.43
광남동 - 명장1동	9.93*	7.37~12.50	3.71	-3.99~11.42
광남동 - 목호동	19.33*	15.31~23.34	21.35*	10.53~32.16
광남동 - 범천1동	20.61*	17.84~23.38	22.98*	15.07~30.89
인제읍 - 명장1동	4.46*	0.95~7.98	-6.16	-16.24~3.91
인제읍 - 목호동	13.86*	9.17~18.54	11.47	-1.15~24.09
인제읍 - 범천1동	15.14*	11.47~18.81	13.10*	2.87~23.33
명장1동 - 목호동	9.39*	5.21~13.57	17.63*	6.35~28.92
명장1동 - 범천1동	10.68*	7.67~13.68	19.26*	10.73~27.80
목호동 - 범천1동	1.28	-3.03~5.60	1.63	-9.79~13.06

〈부록 2〉 전체 응답률의 지역별 비교

	전수		표본	
	차이	95% C.I.	차이	95% C.I.
범천1동 - 목호동	0.80	-1.20~2.80	-0.19	-4.96~4.58
범천1동 - 기흥동	0.83	-0.71~2.37	2.45	-1.40~6.30
범천1동 - 명장1동	0.91	-0.48~2.31	2.83	-0.74~6.39
범천1동 - 광남동	0.65	-0.64~1.93	6.82*	3.51~10.12
범천1동 - 인제읍	5.96*	4.26~7.66	1.85	-2.43~6.12
목호동 - 기흥동	0.03	-2.01~2.08	2.64	-2.29~7.58
목호동 - 명장1동	0.12	-1.82~2.05	3.02	-1.69~7.73
목호동 - 광남동	-0.15	-2.01~1.71	7.01*	2.49~11.53
목호동 - 인제읍	5.16*	2.99~7.32	2.04	-3.23~7.31
기흥동 - 명장1동	0.08	-1.38~1.54	0.38	-3.40~4.16
기흥동 - 광남동	-0.18	-1.54~1.18	4.36*	0.83~7.90
기흥동 - 인제읍	5.12*	3.37~6.88	-0.60	-5.06~3.85
명장1동 - 광남동	-0.27	-1.45~0.92	3.99*	0.77~7.21
명장1동 - 인제읍	5.04*	3.41~6.67	-0.98	-5.19~3.23
광남동 - 인제읍	5.31*	3.77~6.84	-4.97*	-8.96~-0.98

〈부록 3〉 응답률과 조사구 특성 변수와의 상관계수

	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)	12)	13)	14)	15)	16)	17)	18)	19)	20)
1) 전체 응답률	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2) 비방문조사 응답률	-0.01	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3) 인터넷조사응답률	-0.02	0.47*	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4) 우편조사응답률	0.01	0.98*	0.28*	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5) 여성비율	0.19*	-0.13*	-0.16*	-0.10*	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6) 평균연령	0.04	-0.52*	-0.37	-0.47*	0.17*	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7) 평균교육수준	-0.06	0.29*	0.18*	0.27*	-0.03	-0.66*	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-
8) 노동가능인구비율	-0.14*	0.06	0.13*	0.03	-0.19*	-0.47*	0.70*	1.00	-	-	-	-	-	-	-	-	-	-	-	-
9) 1인가구비율	-0.13*	-0.52*	-0.37*	-0.48*	0.20*	0.20*	0.13*	0.21*	1.00	-	-	-	-	-	-	-	-	-	-	-
10) 자가소유비율	0.08	0.16*	0.06	0.16*	0.03	0.32*	-0.33*	-0.33*	-0.50*	1.00	-	-	-	-	-	-	-	-	-	-
11) 비농림어가비율	0.22*	0.26*	0.21*	0.23*	0.03	-0.51*	0.50*	0.33*	-0.08	-0.33*	1.00	-	-	-	-	-	-	-	-	-
12) 인구밀도	0.11*	-0.26*	-0.19*	-0.24*	0.29*	0.14*	0.11*	0.12*	0.31*	-0.22*	0.28*	1.00	-	-	-	-	-	-	-	-
13) 아파트조사구여부	0.01	0.71*	0.23*	0.72*	0.02	-0.57*	0.45*	0.17*	-0.26*	-0.03	0.31*	-0.16*	1.00	-	-	-	-	-	-	-
14) 인센티브사전홍보여부	-0.12*	0.20*	0.30*	0.15*	-0.22*	-0.14*	-0.06	0.04	-0.38*	0.32*	-0.08	-0.03	-0.13*	1.00	-	-	-	-	-	-
15) 전수조사여부	-0.05	0.05	-0.05	0.05	-0.07	-0.03	0.03	0.03	-0.03	0.01	0.00	0.00	0.00	0.02	1.00	-	-	-	-	-
16) 임금근로자비율	-0.35*	0.19	0.07	0.21	-	-	-	-	-	-	-	-	-	-	-	1.00	-	-	-	-
17) 비이동차비율	0.14	-0.31*	-0.02	-0.35*	-	-	-	-	-	-	-	-	-	-	-	-0.72*	1.00	-	-	-
18) 평균거주기간	0.23	-0.51*	-0.26	-0.51*	-	-	-	-	-	-	-	-	-	-	-	-0.73*	0.89*	1.00	-	-
19) 중하위층가구비율	0.09	-0.56*	-0.47*	-0.51*	-	-	-	-	-	-	-	-	-	-	-	-0.51*	0.56*	0.71	1.00	-
20) 컴퓨터보유비율	-0.15	0.65*	0.44*	0.62*	-	-	-	-	-	-	-	-	-	-	-	0.53*	-0.62*	-0.80*	-0.89*	1.00

주: *, p<0.05 임.