

제5장

가구 무응답 대응을 위한 무응답 가구 특성 분석

- 총조사자료와 지역별고용조사 가구관리종합표자료 연계 -

김서영 · 박라나 · 이선영

제5장



제1절 서론

1. 연구배경

우리나라의 경제·사회 환경의 변화는 통계조사 환경의 변화에도 적지 않은 영향을 주었다. 가구 단위 조사의 경우 가구구조의 변화, 즉 1인 가구의 증가, 맞벌이 가구의 증가, 노인 가구주 가구의 증가는 조사에 대한 참여도를 떨어뜨리는 주요 원인이 되고 있다. 따라서 조사 대상 가구로부터 조사에 대한 참여와 협조를 구하기 위해서는 무응답 가구의 성향을 분석하여 그 특성을 이해하는 것이 중요하다. 이는 통계조사를 기획하는 단계에서 무응답 가구에 대한 대처 방법을 찾는 데 중요한 정보가 될 수 있기 때문이다. 한편, 무응답 가구는 조사에 협조하지 않는 가구로서 이 가구의 속성을 파악하기가 쉽지 않다. 무응답 가구들의 특성을 파악하고 무응답 편향을 측정해낼 수 있다면 이는 무응답 가구에 대한 추가 조사 없이 통계 오차를 줄일 수 있는 방법을 찾는 데 크게 기여할 수 있게 된다.

우리나라 전국 가구에 대한 일반적인 속성을 가장 잘 파악할 수 있는 자료로 5년마다 실시되는 인구주택총조사(총조사)를 꼽을 수 있다. 뿐만 아니라 행정자료도 좋은 예가 될 수 있다. 표본조사에서 잘 협조하지 않은 가구도 우리나라 국민 전체를 대상으로 하는 총조사에서는 조사에 협조하는 경향이 있다. 우리나라 총조사 자료는 가구를 대상으로 하는 대부분 표본조사의 모집단으로서 가구 특성을 이해할 수 있는 가장 전형적인 자료라 할 수 있다. 가장 최근의 총조사는 2010년 10월말에서 11월초에 걸쳐 실시되었다. 이와 비슷한 시점인 2010년 9월에 가구를 대상으로 하는 대규모 표본조사인 “지역별

고용조사”가 실시되었다. 지역별고용조사는 우리나라 시군의 고용동향을 파악하기 위한 조사로서 전국의 대도시, 소도시, 농·어촌 지역의 가구를 골고루 포함하고 있다. 따라서 대규모 표본조사인 이 지역별고용조사의 무응답 가구들의 정보를 분석할 수 있다면 가구 단위 조사에서의 무응답 가구의 성향을 이해하는데 큰 도움이 될 것이다.

우리는 다행히 지역별고용조사의 가구관리종합표 자료를 통해 무응답 가구 목록을 얻을 수 있다. 지역별고용조사가 총조사와 유사한 시기에 실시되었다는 점에서 이 두 자료의 연계는 연구자료로서 중요한 의미를 갖을 수 있다. 이 두 조사의 대상가구에 대해 두 조사에서 동시에 관측되는 정보에는 “주소”가 있다. 이 주소 변수를 통해 동일가구에 대해 두 조사 자료를 연계할 수 있다. 가구를 중심으로 두 조사 자료를 연계하면 우리는 표본조사 대상 가구에 대해 총조사의 정보를 이용할 수 있게 된다. 이렇게 연계된 무응답 가구의 정보는 무응답 가구에 대한 성향을 이해하고 무응답 편향 측정 및 무응답에 영향을 미치는 변수를 탐색하는데 중요한 역할을 한다. 특히 표본조사에서 무응답 편향을 측정하는 것은 응답률이 낮거나 정확성이 요구되는 자료일수록 필요한 분석과정이라 할 수 있다. 그렇지만 실제로 무응답 가구에 대한 특성을 파악하고, 이들로부터 무응답 편향을 계량화하는 것은 쉽지 않은 일이다. 그럼에도 불구하고 많은 연구자들이 여전히 무응답 편향에 관심이 갖는 이유는 그 측정이 어려운데 반해 무응답 편향이 통계품질에 많은 영향을 미친다는 것에는 이견이 없기 때문일 것이다.

2. 연구내용 및 방법

본 연구의 가장 큰 목표는 무응답 가구의 성향을 이해하고 이로부터 무응답 가구에 대한 대응 방안을 마련하기 위해 필요한 정보를 제공하는 데 있다. 즉 어떤 가구가 조사에 협조적이고, 어떤 가구가 조사에 비협조적인지를 파악함으로써 실제조사에서 이 정보를 활용할 수 있도록 하고자 한다. 본 연구는 다음의 세 가지 내용을 중심으로 분석하였다.

- ① 무응답 가구의 성향을 파악하기 위해 주요 총조사 변수들에 대해 빈도분석을 실시한다. 조사가 완료된 시점을 기준으로 조사 대상 가구를 응답 가구와 무응답 가구 두 개 그룹으로 분류하고 각 유형별 가구의 특성을 분석한다. 무응답 가구에 대해 다시 부재(not-at-home), 거부(refusal) 및 기타(others)가구로 분류하였다. 그렇지만 실제 분석에서는 부재(비접촉)와 거부 두 개 유형으로만 무응답 유형을 분류하였다. 그 이유는 기타 유형에 속하는 무응답 가구의 비중이 전체 표본에 대해 차지하는 비중이 매우 작고, 특히 본 연구에서는 기타와 불능 가구를 명확히 구분하기 쉽지 않았기 때문이다.

- ② 무응답 편향을 계산한다. 무응답 가구 부재와 거부 가구에 대해 편향을 측정한다. 또한 대도시(경기도를 포함한 특광역시)와 소도시(경기도를 제외한 8개 도)지역으로 구분하여 무응답 편향을 계산한다. 그 이유는 지역별고용조사의 표본에 소도시 및 농·어촌 지역의 표본이 많이 포함되어 있다는 점과 대도시와 소도시 지역에서의 조사 대상 가구들의 특성이 다를 것으로 판단되었기 때문이다(실무자와 연구자들의 경험적 판단에 근거).
- ③ 가구들의 응답 여부에 영향을 미칠 수 있는 변수들을 찾는다. 이는 무응답 가중치 조정 변수를 찾는 것과 같다. 이러한 변수들은 무응답 편향을 줄이는 데 기여할 수 있으며, 의사결정나무분석 또는 로지스틱회귀분석과 같은 다양한 분석방법들이 사용될 수 있다.

정리하면 본 연구의 핵심은 지역별고용조사 대상가구를 2010년 센서스 자료에 연계하여 표본조사 대상가구들에서 발생하는 무응답 가구의 특성을 파악하고, 이로부터 무응답 가구에 대한 대응 논리를 마련하는 것이다. 또한 이 연구 결과를 토대로 향후 가구 단위 조사에서 무응답 가구 대응 논리를 마련하는데 필요한 기초 정보를 제공할 수 있다. 추가적으로 본 연구는 총조사 자료의 변수가 무응답 가구의 특성을 분석하고, 무응답 편향을 측정하고, 무응답 보정을 위한 가중치 변수를 탐색하는데 얼마나 중요한 정보를 제공하는지 확인할 수 있는 계기가 될 것이다.

3. 기대효과 및 한계점

우리나라에서 실시되는 가구 대상 표본조사의 경우, 무응답 가구에 대한 다양한 의견들이 있지만 실제로 무응답 가구가 어떤 성향의 가구들인지에 대해서 분석한 연구는 많지 않은 것 같다. 이는 무응답 가구에 대한 정보를 수집하기가 쉽지 않았기 때문이었을 것이다. 그런 점에서 볼 때 무응답 가구에 대한 성향 분석을 실시한 본 연구는 상당한 의미가 있다고 볼 수 있다.

- 조사자료와 총조사자료를 연계하여 무응답 가구에 대한 정보를 파악할 수 있는 계기를 마련할 수 있다는 점에서 큰 의미가 있다.
- 무응답 가구에 대한 성향을 이해하여 향후 현장조사에서 필요한 무응답 가구 대응 전략을 세우는데 기여할 수 있다.
- 무응답 가구 가중치 적용에 필요한 변수를 파악하고 이로부터 무응답 편향을 줄이는데 기여할 수 있다. 무응답 조정에 효과가 있을 것으로 판단되는 변수는 표본조사 질문지 등에 포함될 수 있도록 요구할 수 있다.

그러나 많은 연구에서 그렇듯이 본 연구에서도 몇 가지 한계는 있다.



- 본 연구의 가장 어려운 점은 조사자료와 총조사자료를 연계할 만한 연계 키 변수(matching key variable)가 충분하지 않다는 것이었다. 실제 사용될 수 있는 키 변수는 가구 주소 밖에 없다. 이 주소도 가구에서 기입할 때 표본조사와 총조사자료에 대해 동일하게 입력되지 않은 경우가 있고, 다가구 또는 다세대 거주지가 많은 우리나라의 주거 특성상 주택가구들의 연계가 상당히 어려웠다.
- 표본조사와 센서스가 비슷한 시점에서 이루어졌지만 약 1달 반 정도의 시차가 있고, 그 기간 동안에 발생한 인구이동 정도에 대해서는 파악하기가 어렵다. 즉, 가구 속성에 대해 매칭오류가 발생할 수 있다. 주소가 정확히 연계된 경우라도 주소에 거주한 가구의 속성은 두 자료에서 다를 수도 있다.

제2절 분석자료 구성 및 자료연계

본 연구는 2010년 센서스(11월 1일 실시)와 비슷한 시점에 실시된 지역별고용조사의 가구관리종합표자료를 사용하였다. 가구관리종합표자료는 조사원이 가구를 방문할 때 얻을 수 있는 정보를 기록하는 것으로써 가구의 거주유형, 가구 방문회차별 방문일시 및 응답상태(응답, 부재, 거부 등)에 관한 정보를 포함하고 있다. 본 연구에 사용된 총조사 자료는 총조사 전수항목과 표본항목을 모두 포함하고 있다. 총조사 자료와 가구관리종합표 두 자료를 연계하면 다음과 같은 정보를 포함하는 하나의 자료셋을 얻을 수 있다.

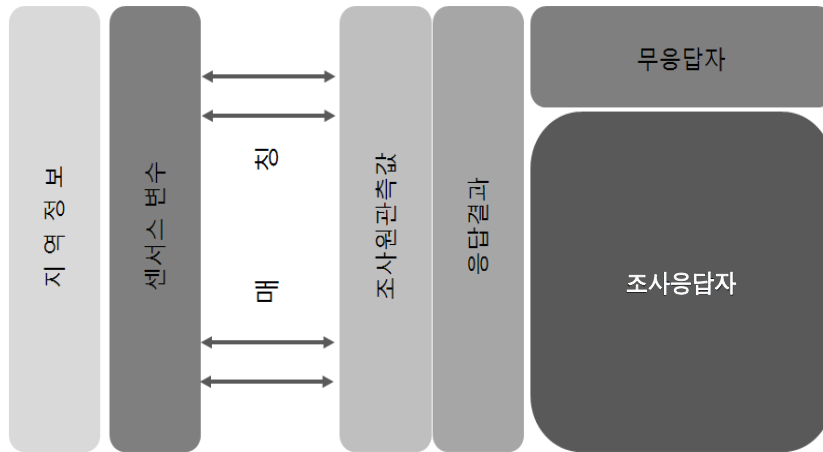
- 조사를 위한 추출된 지역 정보(행정구역단위)
- 가구와 주택 정보 및 가구주 인구학적 특성 정보
- 조사원의 방문회차에 관한 기록(회차별 가구 응답 : 응답/거부/부재/기타)

1. 자료연계과정

가. 연구자료생성

본 연구에서 자료 연계는 행정구역을 중심으로 총조사 변수와 조사원이 관측한 변수를 연계하였다. [그림 5-1]은 총조사 자료와 가구관리종합표 자료의 연계를 통해 얻어지는 자료의 상태를 나타낸다. 두 개 이상의 자료를 연계할 때 가장 중요한 것이 연계 키가 있느냐 하는 것이다. 본 연구의 경우 연계 키로 사용될 수 있는 변수는 가구의 주소로 조사시에 가구 또는 조사원이 직접 기입하는 정보이다. 일반적으로 주소를 연계 키로 사용할 때 어려운 점은 동일한 가구에 대해 두 자료에 기입된 주소가 일치하지 않을 수

있다는 점이다. 우리나라의 경우, 가구의 거처에 대한 주소 특히 번지와 건물명을 입력하는 방법이 다양하게 나타나고 있다. 우리나라의 거주 특성상 주택은 한 번지 내에 여러 가구가 거주하는 다가구 또는 다세대 가구 형태가 존재하고 있다. 이는 주소 키를 이용한 자료 연계에 있어서 많은 어려움을 초래하는 요인이라 할 수 있다.



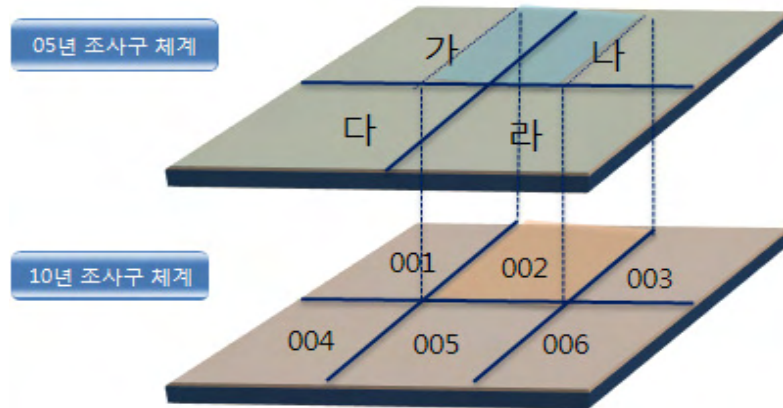
[그림 5-1] 연구자료 셋 구성

나. 연계방법

연계 대상이 되는 자료들의 조사시점과 표본 추출틀이 동일할수록 연계된 자료의 정확성은 높다고 볼 수 있다. 지역별고용조사는 2005년 총조사 자료를 표본추출틀로 사용하였다. 2010년 총조사에서 조사구 체계가 변경되었다. 따라서 조사구 체계가 다른 경우 두 조사자료의 조사구내에 속해있는 가구의 연계를 쉽게 하기 위해 2005년과 2010년도의 총조사 조사구 연계를 우선적으로 시도해 볼 수 있을 것이다.

1) 조사구 연계

조사구 연계를 위해 연구자들은 GIS 시스템의 UNION 기능을 활용하였다. 이는 지도 위에 2005년 조사구와 2010년 조사구를 중첩시키는 방법으로서, 2005년 조사구에 대해 2010년 조사구의 면적이 겹치는 비율을 사용하여 두 조사구의 일치성 여부를 판단한다. 이 방법은 겹치는 면적비율이 90% 이상이면 두 조사구는 일치 조사구로 판단한다. 본 연구에서는 전체 약 80~90만 센서스 조사구 중 약 10% 정도가 일치 조사구로 연계되었다. 조사구 연계 비율이 낮지만 연계된 조사구에 대해서는 보다 쉽고 정확하게 가구 연계를 시도할 수 있었다.



[그림 5-2] 총조사 일정 구획에서의 조사구 분할

위의 [그림 5-2]는 적당한 크기의 구획에 대한 조사구 분할을 예로 든 것이다. 그림의 위쪽은 2005년 조사구 체계, 아래쪽은 2010년 조사구 체계라 하자. 그림에서 보듯이 동일한 구획에 대해 조사구가 서로 다르게 분할되어 있다. 이 조사구들의 일치성을 파악하기 위해 GIS의 UNION 기능을 사용하면 2010년 ‘001’ 조사구는 2005년 ‘가’ 조사구와 완전히 일치하지만, 2010년 ‘002’ 조사구는 2005년 ‘가’와 ‘나’ 조사구에 걸쳐 있고, 각각의 겹치는 비율은 50% 정도이다. 따라서 90% 이상 겹치는 비율 규칙을 적용할 때, 2010년 ‘001’ 조사구와 2005년 ‘가’ 조사구는 동일한 조사구로 간주하지만, ‘002’ 조사구는 조사구 ‘가’와 ‘나’ 어디에도 속한다고 보기 어렵다.

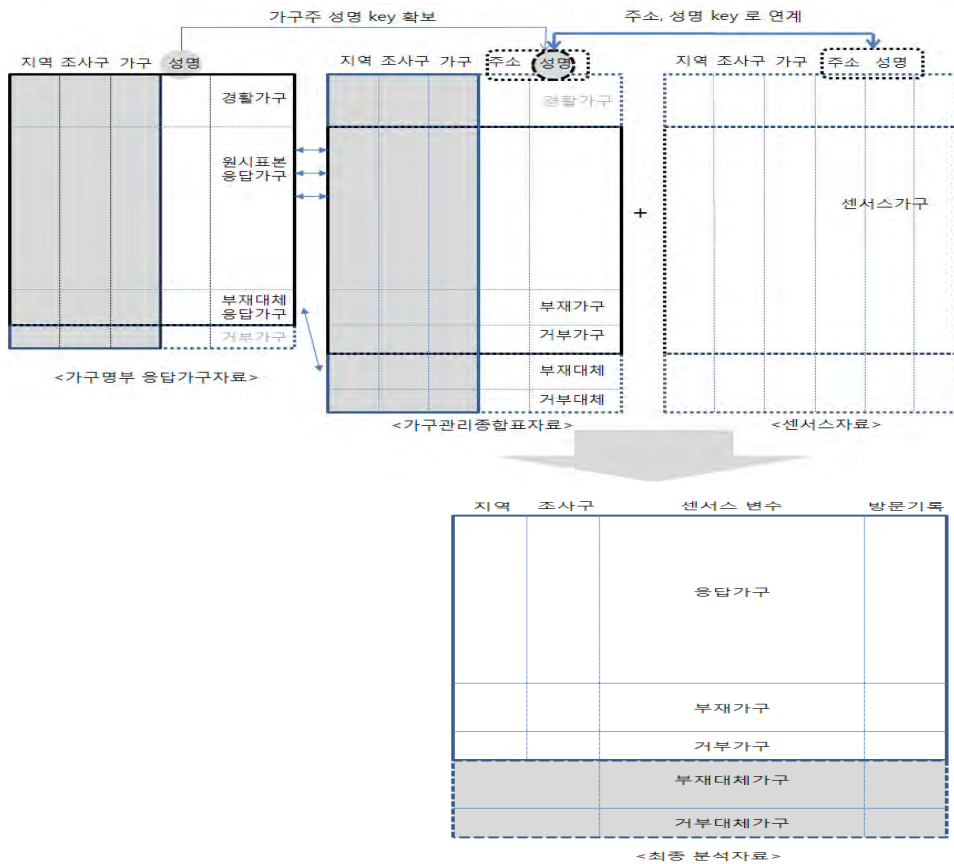
2) 최종자료 셋 구성 방법

[그림 5-3]은 지역별고용조사의 가구관리종합표 자료와 총조사 자료가 연계되는 과정을 나타낸 것이다. 자료 연계는 가구관리종합표 자료를 기본으로 하고 여기에 총조사 자료를 연계하는 방식을 사용하였다. 연계 키는 가구의 주소를 사용하였고, 응답을 얻은 가구에 대해서는 가구명부에 나타난 가구주 성명도 이용하였다. 이렇게 복잡한 작업을 통해 두 자료를 가능한 정확하게 연계하고자 하였다.

최종 자료셋을 만드는 과정에 대해 좀 더 살펴보기로 하자. 연계작업의 핵심은 가능한 정확하면서 많은 자료를 연계하는 것이라 할 수 있다. [그림 5-3]과 [그림 5-4]는 가구에 대해 주소를 연계하는 구체적인 과정을 나타낸 것이다. 분석결과에 대한 신뢰성을 확보하기 위해서 자료 연계에 대한 이해가 충분해야 할 것이다. 연계과정을 설명하면 다음과 같다.

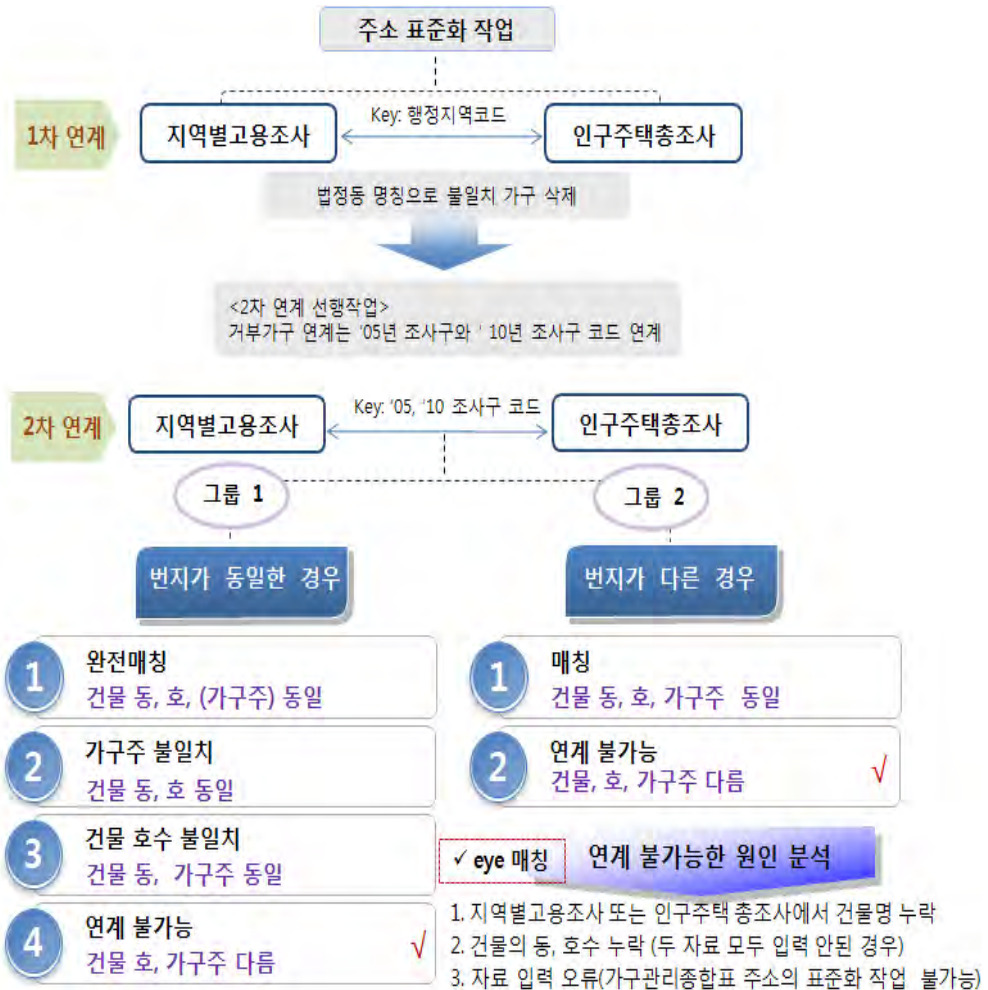
< 연계 절차 >

- 단계 1. 가구관리종합표 주소를 각 특성별로 표준화 작업 실시(등록센서스과)
 - 총조사자의 가구명부는 변수별(건물명, 동, 호수, 번지, 통, 반 등)로 특성화 된 주소를 저장하고 있으나, 가구관리종합표는 주소 전체가 한 변수명으로 설정
- 단계 2. 표준화된 주소로 총조사 가구명부와 가구관리종합표 자료를 연계
 - 법정동 명칭으로 불일치 가구는 삭제한다.
- 단계 3. 행정구역(읍·면·동)내, 거처 번지의 일치성에 따라 2개 그룹으로 분리
 - 각 그룹별로 건물명, 동, 호수 또는 가구주 성명 등 일치여부 고려
 - 동일한 건물명이 서로 다르게 입력된 경우는 텍스트 매칭을 고려하여 건물명 중에서 일치하는 글자수를 고려하여 판단



[그림 5-3] 자료연계 방법





[그림 5-4] 구체적인 연계 절차

제3절 자료 및 분석결과

1. 무응답 가구 유형

본 연구에서 분석대상이 되고 있는 지역별고용조사는 총 4번의 방문을 통해 조사를 실시하고 있다. 조사가 완료된 시점에서 조사대상 가구의 응답 상태는 크게 응답과 무응

답 가구로 분류할 수 있다. 특히 무응답 가구에 대해서는 연구자들과 현장 실무경험자들의 판단에 의해 부재(비접촉), 거부, 기타 유형의 가구로 다음과 같이 분류하였다. 이 분류는 본 연구에서 사용한 방법으로 연구자에 따라 다르게 정의할 수도 있다.

〈 조사대상 가구에 대한 최종 응답 상태 정의 〉

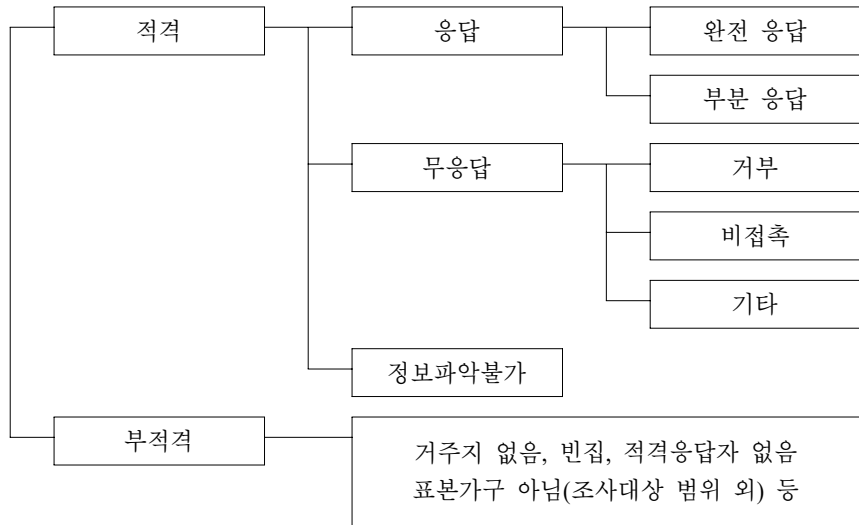
상태	정 의
응답	총 4번 방문 중, 가구의 조사표가 정상적으로 완료된 경우
부재	총 4번 방문 중, 가구내 응답이 가능한 가구원을 만나지 못한 경우
거부	총 4번 방문 중, 접촉이 된 가구 중에서 1번 이상 조사를 거부한 가구
기타	언어장애, 신체활동상의 이유로 조사에 임할 수 없는 경우

거부 가구에 대한 정의는 상황에 따라 조금씩 다를 수 있다. 1번의 방문으로 그치지 않고 여러 번 가구를 방문하는 조사의 경우는 매 방문시마다 가구의 접촉여부 및 거부여부가 다를 수 있다. 부재와 거부가 번갈아 일어나는 가구에 대해서는 사실 명확하게 그 가구가 어떤 유형의 무응답 가구인지를 분류하기가 어렵다. 나라마다 관심을 갖는 무응답 유형이 다를 수 있다. 많은 나라에서 거부 가구에 대해 관심이 많은 것은 실제로 가구를 접촉하기가 어렵기 때문에 접촉이 된 가구에 대해서는 최대한 응답을 얻고자 하기 때문이다. 응답하기를 꺼려하는 가구들에 대해 그런 유형의 가구 특성을 분석하여 그 특성에 맞게 거부 전환 전략을 세우는 것은 응답률을 올리기 위한 하나의 방법으로 널리 사용되고 있다.

국제사회에서는 조사에 대해 보다 객관성과 타당성을 갖기 위해 다양한 노력을 하고 있다. 응답 유형에 대해 AAPOR(2010)에서는 <표 5-1>과 같이 응답 유형을 분류하고 있다 (김서영과 안다영, 2010). <표 5-1>에서 적격표본 중 ‘정보과악불가(unknown eligibility)’는 표본가구로의 적격성 여부가 분명하지 않거나 적격가구이나 적격대상자 여부가 명확하지 않은 경우에 해당한다. 이런 경우 조사에 대한 응답여부를 파악할 수 없지만 표본 자체는 적격표본으로서 모집단에 명백하게 포함된 것이기 때문에 응답유형 그 자체를 추정하여 사용할 수 있다. 부적격표본은 표본 자체를 이용할 수 없는 경우로, 대상표본이 아니거나 (out of sample), 주거지역(hot-a-housing-unit)이 아닌 회사, 정부기관, 요양소, 군대 등 시설이거나 장기적 혹은 정기적으로 비어있는 집(트레일러, 별장 등) 또는 적격대상자가 없는 가구(housing unit with no eligible respondents) 등이 해당된다. 응답 유형의 정확한 분류는 응답률 계산과 밀접한 관련이 있다. 본 연구는 조사자료의 특성을 고려하여 적격표본에 대해서 정보과악이 불가한 경우는 없고, 부분응답 가구도 없는 것으로 간주하였다.



<표 5-1> 유형별 응답형태(AAPOR, 2010)



2. 기초자료 분석

가. 응답유형별 분포

<표 5-2>는 가구관리종합표 자료에 의한 응답 유형별 분포를 나타낸다. 표에서 부재 가구 비율이 거부 가구 비율보다 약 2배 정도 높은 것을 알 수 있다. 빈집 비율이 높은 것은 이 조사가 시군 지역을 대상으로 한다는 점에서 시도 단위 설계의 조사에 비해 농·어촌 가구가 많이 포함되어 있다는 점을 생각해 볼 수 있다.

<표 5-2> 전체 조사대상 가구 응답 유형별 분포

	전체		원시표본	
	가구수	%	가구수	%
응답	118,971	76.81	92,216	78.45
거부	5,990	3.87	4,404	3.75
부재	12,650	8.17	8,648	7.36
빈집	16,170	10.44	11,458	9.75
기타	1,105	0.71	825	0.70
전체 대상수	154,886	100.0	117,551	100.0

나. 응답률 계산

응답률은 연계되기 전의 가구관리종합표 원시표본에 의해서 계산되었다. 어떤 조사에서든 조사가 완료된 시점에서 응답률을 계산하게 된다. 이는 응답률은 조사오차에 무관하지 않고 응답률이 낮을수록 무응답 편향이 커질 수 있기 때문이다. 무응답 편향은 그 조사의 무응답 정도(무응답률)와 응답자로부터 관측된 값과 무응답자로부터 관측되지 않은 값과의 차로 구성되어 있다. 무응답 정도를 파악할 때, ‘완료율’ 과 ‘응답률’이 서로 다른 개념으로 적용되기도 한다(Brick, Bruke and Le, 2000). 참고로 완료율은 각 표본추출 단계에서 조사에 직접 참여하는 개체들의 비율을 말한다. 응답률은 조사에 참여한 전체 표본비율을 말하고 모든 표본추출 단계를 설명할 수 있다. 응답률 평가는 무응답과 관련된 제반 연구들의 출발점이 될 수 있다.

<표 5-3>은 우리나라를 서울·경기, 광역시 및 도의 3개 권역으로 분류하여 각 권역별로 응답률, 접촉률, 협조율, 거부율을 계산하였다. 계산은 APPOR(2010)을 따른다. <표 5-3>은 적격가구만을 대상으로 한 것으로 <표 5-2>에서의 빈집 등 부적격 가구들은 계산에 포함되지 않는다.

제5장



< 가구 수준에서의 응답률, 접촉률, 협조율, 거부율(AAPOR, 2010) >

- 응답률(RR5) =
$$\frac{I}{(I+P) + (R+NC+O)}$$
- 접촉률(CON3) =
$$\frac{(I+P)+R+O}{(I+P)+R+O+NC}$$
- 협조율(COOP1) =
$$\frac{I}{(I+P)+R+O}$$
- 거부율(REF3) =
$$\frac{R}{(I+P) + (R+NC+O)}$$

* R: 거부자, I: 응답자, P: 부분응답자, NC: 비접촉자, O: 기타

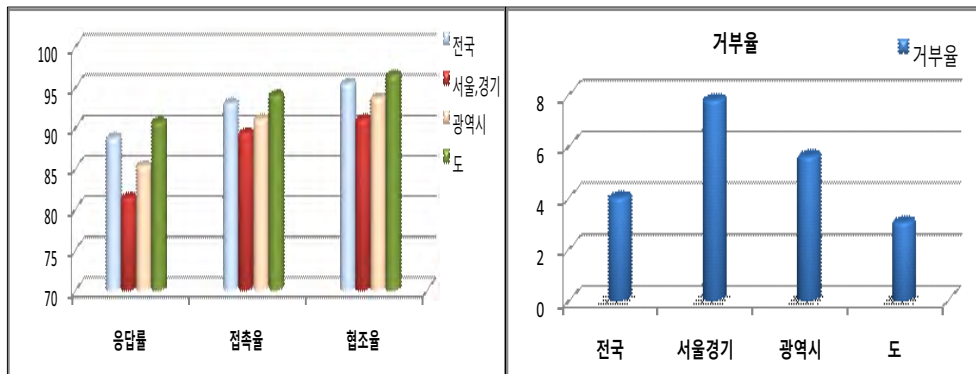
서울·경기 등 수도권 지역은 다른 권역에 비해 응답률, 접촉률이 낮고, 거부율은 상당히 높다([그림 5-5]). 거부율의 경우, 서울·경기는 전국수준보다 2배 정도 높고, 도 지역은 전국 수준보다 약간 낮다. 광역시는 서울·경기 지역보다 협조율은 높고 거부율은

낮아보이지만, 대도시에서 조사하기가 어렵다는 것은 쉽게 짐작할 수 있다.

〈표 5-3〉 권역별 응답률, 접촉률, 협조율, 거부율

(단위 : %)

권역	응답률	접촉률	협조율	거부율
전국	88.82	93.03	95.47	4.01
서울, 경기	81.42	89.39	91.08	7.80
광역시	85.33	91.06	93.70	5.61
도	90.76	96.52	96.52	3.05



[그림 5-5] 권역별 응답률, 접촉률, 협조율, 거부율

3. 연계자료 분석결과

본 연구에서 분석 변수로는 가구 특성을 나타내는 가구구분(1인 가구 여부), 거주유형(주택, 아파트, 연립주택 등), 점유형태(자가, 전세 등), 가구원수, 동거아동수, 주거연면적, 가구주 특성으로 성별, 연령, 학력, 혼인상태(유배우 여부) 변수를 사용하였다. 분석 결과는 거주 유형별 자료 연계 정도와 국가통계가 지역 단위로 공표된다는 점을 고려하여 설명하고자 하였다.

앞에서 설명한 바와 같이 본 연구에서 핵심은 무응답 가구 성향을 분석하고 이를 기초로 무응답 편향을 측정하고 무응답에 영향을 미치는 변수를 찾아냄으로써 향후 무응답 가구에 대한 대응 전략을 세우는데 기초정보를 제공하기 위함이다. 요약하면 다음과 같다.

(1) 무응답 가구 성향 분석

- 응답 가구, 부재 가구, 거부 가구에 대해 센서스 변수 개념에서 특성 비교

(2) 무응답 편향 측정

- 센서스 특성 개념에서 다른 부차 그룹들 간의 과대 또는 과소 대표성
- 계산식

$$\begin{aligned} bias(\bar{y}_r) &= \bar{y}_r - \bar{y}_n = \frac{m}{n}(\bar{y}_r - \bar{y}_m) \\ &= \frac{m_{ref}}{n}(\bar{y}_r - \bar{y}_{m_{ref}}) + \frac{m_{nc}}{n}(\bar{y}_r - \bar{y}_{m_{nc}}) + \frac{m_{oth}}{n}(\bar{y}_r - \bar{y}_{m_{oth}}) \end{aligned} \quad (2.1)$$

여기서 r 은 응답 가구수, n 은 전체 표본 가구수, m 은 무응답 가구수, m_{ref} 는 거부에 의한 무응답 가구수, m_{nc} 는 비접촉에 의한 무응답 가구수, m_{oth} 는 기타 무응답 가구수 각각 나타냄

(3) 무응답에 영향을 미치는 변수 탐색

- 응답결과들에 대해 센서스 변수들간의 상대적 영향력을 결정하기 위함



가. 최종 연계 결과

<표 5-4>는 두 자료의 연계율을 계산한 것이다. 연계 대상 가구는 전체 117,551 가구로 이 중에서 아파트는 34,881 가구, 주택은 82,670 가구에 해당한다. 이 가구들을 대상으로 연계한 결과 전체 가구의 연계율은 74.6%, 아파트 가구의 연계율은 94.1%, 주택 가구의 연계율은 66.4%이다. 주택 가구의 연계율이 아파트 가구에 비해 낮다. 이미 예상한 바와 같이, 이와 같은 현상은 연계 키가 거처의 주소라는 점에서 주택 가구의 주소가 두 자료에서 약간 다르게 기입되었다거나 한 주소 내에 여러 가구가 거주하여 주소 연계가 불가능한 경우가 있기 때문이라 볼 수 있다.

<표 5-4> 조사대상 가구 연계 정도

	연계 대상 가구수			연계된 가구수			연계비율(%)		
	전체	아파트	주택	전체	아파트	주택	전체	아파트	주택
전국	117,551	34,881	82,670	87,730	32,823	54,907	74.6	94.1	66.4
특광역시,경기	28,947	12,016	16,931	23,929	11,647	12,282	82.7	96.9	72.5
8개도	88,604	22,865	65,739	63,801	21,176	42,625	72.0	92.6	64.8

나. 응답 가구와 무응답 가구의 특성 비교

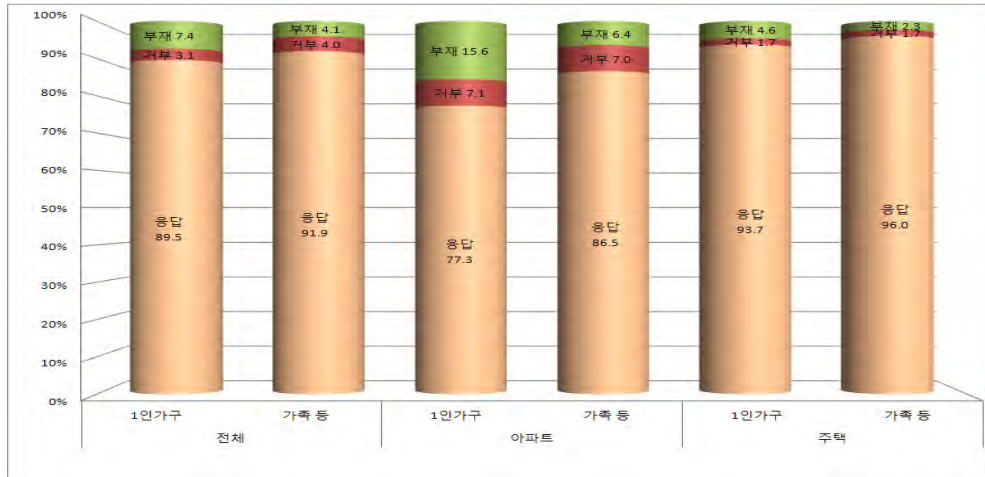
1) 응답 가구와 무응답 가구 비교

아파트 가구의 연계율은 94.1%로 매우 높고, 주택 가구의 연계율은 66.4%로 아파트 가구에 비해 상대적으로 낮다(<표 5-4>). 자료의 연계율을 고려하여 거주 유형별로 나누어 분석을 실시하였다. 이는 최대한 자료분석결과를 정확하게 이해하기 위한 것이다. 지역별로는 서울과 6대 광역시, 경기도를 하나의 권역(이하 대도시로 통칭)으로 보고, 나머지 8개 도지역(이하 소도시로 통칭)을 합쳐서 총 2개 권역으로만 나누어 분석하였다. 본 절에서 분석의 목적은 우선 응답 가구와 무응답 가구가 가구의 특성을 나타내는 변수에 대해 차이가 있는지를 검증하는데 있다. 이를 위해 각 변수에 대해 빈도분석과 카이제곱 검증 실시하였다. 이때 유의수준 0.05를 기준으로 통계적 유의성을 판단하였다. 대부분 빈도분석 결과로 응답 가구와 무응답 가구의 성향을 설명할 수 있고, 경우에 따라서 카이제곱 검증 결과를 이용하였다.

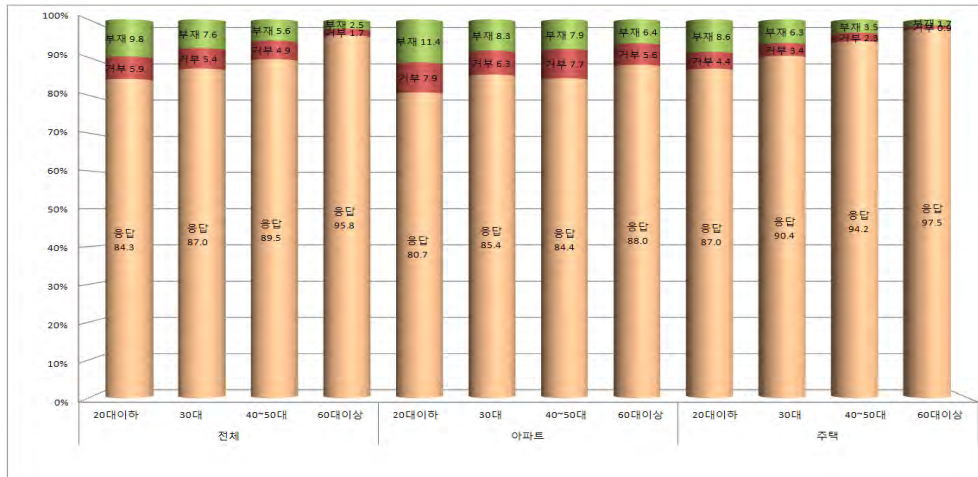
전국의 총 가구를 대상으로 분석한 결과 응답 가구와 무응답 가구 간에는 대부분의 분석 변수에서 통계적으로 유의한 차이를 보였다. 전국 가구에 대한 구체적 분석 결과는 지면의 한계상 모든 변수들에 대한 구체적인 설명은 생략하였다. 응답과 무응답 가구 간에 두드러진 차이를 나타내는 변수, 가구 구분과 가구주 연령 변수에 대해서만 설명하고자 한다. [그림 5-6]은 전체 가구, 아파트 가구, 주택 가구에 대해서 가구 구분별 응답률, 부재율, 거부율을 나타낸 것이다. 가구구분 항목에 있어서 먼저 거처 유형의 경우, 아파트 가구의 응답률이 낮고 주택가구의 응답률이 높다. 가구 구분의 경우, 1인 가구의 응답률이 낮고 가족 등 가구의 응답률이 높다. 거부율은 전체적으로 가구 구분의 경우, 1인 가구의 거부율은 3.1%, 가족 등 가구의 거부율은 4.0%로 가족 등으로 구성된 가구에서 거부율이 약간 높다. 거처 유형별로 살펴보자. 아파트 가구의 경우, 거부율은 1인 가구나 가족 등 가구에서 큰 차이가 없는 반면, 부재율은 큰 차이가 있다. 전체 가구 중의 1인 가구의 부재율은 7.4%이고 가족 등 가구에서는 4.1%로 나타났다. 전체적으로 볼 때, 거처 유형에 관계없이 가족 등 가구의 부재율(아파트 6.4%, 주택 2.3%)보다 1인 가구의 부재율(아파트 15.6%, 주택 4.6%)이 두 배 이상 높은 것을 알 수 있다.

가구주 연령별 응답률 분포는 [그림 5-7]과 같다. 전체적으로 가구주의 연령이 많을수록 응답률이 높고, 이 현상은 주택 가구에서 더 뚜렷하게 나타나고 있다. 한편, 아파트 가구는 주택 가구와 다르게 가구주의 연령이 40~50대에서 응답률이 다른 연령 그룹에 비해 낮게 나타났다. 거부율은 대체로 나이가 많을수록 낮다. 아파트 가구의 경우 40~50대의 연령은 거부율과 부재율이 거의 비슷한 수준이고 다른 연령그룹에서는 거부율이 부재율보다 약간 낮은 경향이 있다. 부재율은 연령이 많을수록 낮고 연령이 적을수록 높다.

이러한 현상은 아파트 가구에서 더 크게 나타나는 것을 알 수 있다.



[그림 5-6] 전국의 가구구분 항목별 응답현황 분포



[그림 5-7] 전국의 가구주 연령 항목별 응답현황 분포

이제 전국을 대도시와 소도시로 나누어 유사한 분석결과를 살펴보자. 앞에서 설명한 바와 같이 연구에 사용된 조사자료의 대상가구에 놓여진 가구가 상당수 포함되어 있기 때문에 도시규모를 고려하여 대상가구들의 특성을 살펴볼 필요가 있다. 게다가 우리나라의 거처 유형은 시 지역에서 아파트 가구가 많고, 군 지역에서 주택 가구가 많기 때문에 도시규모와 거처 유형별 특성이 어느 정도 겹치는 경향이 있다고 볼 수 있다. 따라서 도시규모별로 나누어 거처 유형별 무응답 가구의 특성을 분석함으로써 보다 무응답 가

구에 대한 정확한 이해를 할 수 있을 것이다. <표 5-5>와 <표 5-6>은 대도시(특광역시·경기도)와 소도시(8개 도지역)의 응답 유형(응답 가구, 거부 가구, 부재 가구)에 대해 거처 유형별로 나누어 응답률, 거부율, 부재율을 나타낸다. 각 표에서 ‘유의성’을 나타내는 두 개의 열 중 ‘응답/거부’는 응답 가구와 거부 가구 간의 차이, ‘응답/부재’는 응답 가구와 부재 가구의 차이를 분석한 것으로, 카이제곱 분석을 이용하였다. 즉 각 변수에 따라 응답 유형이 차이가 있는지를 살펴보고자 하였다.

[대도시의 특성]

<표 5-5>의 분석결과에 의하면 대도시의 아파트에 거주하는 가구의 경우, 응답 가구와 거부 가구 간의 차이가 있는 변수는 거처 점유형태와 가구주 연령, 가구주 학력에서 차이가 있는 것으로 나타났다. 우선 아파트 가구에 대해 구체적으로 살펴보자. 아파트 가구의 경우 전세 등(전세, 월세, 사글세 포함)의 가구에서 부재 비율보다 거부 비율이 더 높게 나타났다. 가구주 연령이 20대 이하이거나 40~50대에서 다른 연령 그룹에 비해 응답률도 낮고 거부 비율도 높은 경향이 있다. 가구주 학력이 대학교 이상인 그룹에서 다른 학력 그룹에 비해 거부 비율이 높은 경향을 보이고 있다.

대도시 아파트 가구에 대해 응답 가구와 부재 가구 간 차이를 분석해 보면, 가구 구분, 점유 형태, 가구주 성별, 가구주 연령 변수에서 응답 가구와 부재 가구 간에 통계적으로 유의한 차이가 있다. 1인 가구, 전세 등 가구, 여성 가구주 가구, 20~30대 연령 가구주 가구, 미혼 가구주 가구에서 부재 비율이 높게 나타났다.

대도시 주택 가구의 경우, 아파트 가구의 분석과 유사한 결과를 보였다. 응답률 측면에서 보면 주택 응답률이 아파트 응답률보다 높아서 수치적으로 그 비율에는 차이가 있지만 각 변수에 대한 대도시 주택 가구의 성향은 대도시 아파트 가구의 성향과 매우 비슷한 것을 알 수 있다. 우선 각 변수에 대해 응답 가구와 거부 가구 간 성향의 차이를 보면, 가구주 연령과 학력 변수에서 통계적으로 유의한 차이가 있는 것으로 나타났다. 특히 주택에 거주하는 20대 가구주 가구의 부재 비율은 11.1%로 아파트 가구의 20대 연령 그룹의 부재 비율 9.5%보다 높고, 다른 연령 그룹에 비해 월등히 높게 나타났다. 주택 가구에서 가구주 학력이 대학교 이상인 가구는 부재 비율이 7.9%로 거부 비율 4.3에 비해 약 2배 정도 높다. 이는 가구주의 학력이 대학교 이상인 가구에서 거부 비율이 부재 비율보다 더 높게 나타난 아파트 가구와는 반대 현상인 것을 알 수 있다.

전체적으로 대도시 가구의 성향을 요약하면 다음과 같다. 조사 대상 가구 중 대도시에 거주하는 응답 가구와 거부 가구는 가구주 연령과 학력에서 유의한 차이가 있고, 다른 변수들에 대해서는 응답 가구와 거부 가구는 유사한 성향이 있다고 볼 수 있다. 그리고 부재 가구와 응답 가구는 거의 모든 변수에서 두 그룹 간에 차이가 있는 것을 알

수 있었다. 특히 1인 가구, 전·월세 등에 거주하는 가구, 주거 연면적이 적고, 가구주의 연령이 젊은 여성 가구, 학력이 높은 가구의 부재 비율이 높다고 볼 수 있다. 특이한 사실은 일반적으로 부유층일수록 조사에 응답하지 않은 경향이 있다는 의견에 대해, 분석 결과는 주거 연면적이 클수록 거부율이 약간 높긴 하지만, 통계적으로는 유의한 차이가 없는 것을 알 수 있다. 분석 결과를 통해 대도시 아파트에 거주하는 거부 가구는 응답 가구에 대해, 대부분의 변수들에서 통계적 차이가 없는 것으로 볼 때, 유사한 성향이 있다고 볼 수 있다. 이에 비해 부재 가구는 응답 가구와 그 성향이 대체로 다르다고 볼 수 있으며, 이는 부재 가구의 조사 누락이 발생할 경우 무응답 편향이 크게 발생할 수 있을 것으로 짐작할 수 있다.

〈표 5-5〉 대도시 거주유형별 연관성 분석

(단위 : %)

변수	구분	아파트(n=11,925)					주택(n=11,618)				
		응답유형			유의성		응답유형			유의성	
		응답	거부	부재	응답/거부	응답/부재	응답	거부	부재	응답/거부	응답/부재
가구구분	1인 가구	76.7	8.4	14.9		***	86.3	3.8	10.0		***
	가족 등	84.2	8.7	7.1			92.2	3.8	4.1		
접유형태	자가	84.3	8.2	7.5	**	**	91.9	4.0	4.1		***
	전세 등	81.4	9.5	9.1			89.5	3.6	7.0		
주거 연면적	40이하(18평이하)	82.4	8.1	9.5			89.6	3.2	7.2		
	40~60(18~24평)	83.8	8.5	7.7			90.8	3.6	5.6		***
	60~85(24~34평)	83.7	8.5	7.9			91.1	3.9	5.0		
	85초과(34평초과)	81.6	10.0	8.4			91.0	4.8	4.2		
가구주 성별	남성	84.0	8.5	7.5		***	91.3	3.7	5.0		***
	여성	80.4	9.3	10.3			88.7	4.0	7.3		
가구주 연령	20대이하	80.8	9.7	9.5			84.6	4.3	11.1		
	30대	83.3	7.4	9.2	**	**	87.7	4.5	7.8	**	***
	40~50대	82.4	9.6	8.0			90.6	4.0	5.4		
	60대이상	86.1	7.4	6.5			93.4	3.0	3.6		
가구주 학력	중등이하	84.8	7.8	7.4			93.3	2.9	3.8		
	고등	84.2	7.9	7.9	*		90.5	4.0	5.4	**	***
	대학교	82.3	9.4	8.3			87.8	4.3	7.9		

주 : χ^2 검정 결과 - *** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

[소도시 특성]

여기서는 소도시에 거주하는 가구들의 응답 성향을 거부와 부재로 나누고 이것을 아파트 가구와 주택 가구별로 살펴보기로 하자(<표 5-6>). 소도시 아파트에 거주하는 가구 중 거부 가구는 1인 가구에서의 거부 비율이 6.5%, 가족 등 가구에서의 거부 비율이



6.0%로, 1인 가구의 거부 비율이 더 높은 것을 알 수 있다. 또한 소도시 아파트 가구에 거주하는 20대와 4~50대의 가구주로 구성된 가구의 거부 비율이 다른 연령 그룹에 비해 약간 높게 나타났다. 소도시 아파트에 거주하는 가구 중 부재 가구는 가구주 학력을 제외한 모든 변수에서 응답 가구와 부재 가구 간에 통계적인 차이가 있는 것으로 나타났다. 1인 가구의 부재 비율이 가족 등 가구에 비해 높고, 전·월세에 거주, 주거 연면적이 작은 가구, 가구주 연령이 적은 가구에서 각각 다른 그룹에 비해 부재 비율이 높은 것으로 나타났다. 소도시 주택가구의 경우, 거부 가구는 응답 가구에 대해 가구 구분과 성별을 제외한 모두 변수에서 통계적으로 유의한 차이가 있는 것으로 나타났다. 부재 가구는 응답 가구에 대해 모든 변수에서 차이가 있는 것으로 나타났다.

〈표 5-6〉 소도시 거주유형별 연관성 분석

(단위 : %)

항목	gubun	아파트(n=21,372)					주택(n=40,411)				
		응답	거부	부재	유의성		응답	거부	부재	유의성	
					응답 /거부	응답 /부재				응답 /거부	응답 /부재
가구구분	1인 가구	77.6	6.5	15.9	**	***	95.6	1.2	3.2		***
	가족 등	87.9	6.0	6.1			97.2	1.1	1.8		
점유형태	자가	87.3	6.0	6.7		***	97.5	0.8	1.6	***	***
	전세 등	84.0	6.4	9.6			92.8	2.5	4.7		
주거 연면적	40이하(18평이하)	82.0	5.8	12.2			95.2	1.6	3.2		
	40~60(18~24평)	85.8	6.2	8.0		***	96.7	0.9	2.4	**	***
	60~85(24~34평)	87.9	5.8	6.3			96.9	1.1	2.0		
	85초과(34평초과)	86.2	7.5	6.3			97.1	1.0	1.9		
가구주 성별	남성	87.0	6.0	7.0	*	***	96.8	1.1	2.1		*
	여성	82.7	6.7	10.6			96.4	1.1	2.4		
가구주 연령	20대이하	80.6	7.3	12.1			89.0	4.5	6.4		
	30대	86.4	5.8	7.8	***	***	92.7	2.4	4.9	***	***
	40~50대	85.6	6.7	7.7			95.5	1.6	2.9		
	60대이상	89.3	4.4	6.3			98.1	0.5	1.4		
가구주 학력	중등이하	86.7	5.4	7.9			97.8	0.7	1.5		
	고등	86.1	6.3	7.6			95.3	1.5	3.1	***	***
	대학교	86.0	6.2	7.8			93.7	2.6	3.7		

주 : χ^2 검정 결과 - *** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

지금까지의 응답 가구와 무응답 가구들의 성향을 비교해 본 결과를 종합해 보면 다음과 같다. 응답 가구와 비교한 거부 가구와 부재 가구의 특성은 도시별, 주택유형에 따라 조금씩 그 성향이 다르고, 거부 가구보다는 부재 가구에서 응답 가구와 더 뚜렷한 차이를 나타내고 있다. 거부 가구는 가구들의 고용상태를 나타내는 변수들과 관련이 있을

것으로 판단되는 가구주 연령, 가구주 학력 등의 변수에서 응답 가구와 차이가 있었다. 반면, 부재 가구는 1인 가구, 전·월세 거주자, 거주 연면적이 작은 가구, 가구주 연령이 낮은 가구에서 응답 가구와 차이가 있었다. 부재 가구는 조사에서 접촉하기 어려운 가구로서 가구의 특성상 대부분 바쁘게 생활하는 가구일 가능성이 높다. 이런 가구들은 조사에서 누락될 가능성이 높고, 특히 그 성향이 응답 가구와 다르다는 점에서 통계자료에서 편향을 발생시킬 우려가 있다고 볼 수 있다.

2) 늦게 응답한 가구와 무응답 가구(거부, 부재) 비교

<표 5-7>과 <표 5-8>은 늦게 응답하는 가구는 무응답 가구에 대해 그 성향이 다른지를 살펴보기 위한 것이다. 조사기간 중 거부 경험이 있는 가구가 마지막(4회차) 방문에서 응답한 가구를 거부 성향이 있는 가구라 하고, 4번 방문 중 3번 가구의 가구원을 만나지 못하다가 마지막 방문을 통해 응답을 얻은 가구를 부재 성향이 있는 가구라 하였다. 이 거부 성향 가구와 부재 성향 가구를 각각 최종 거부 가구와 부재 가구에 비교하고자 하였다.

<표 5-7>에서 보면, 대도시 아파트 가구에서 거부 성향의 가구(4회 응답)는 최종 거부 가구에 대해 모든 변수에서 통계적으로 차이가 없고, 대도시 주택 가구에서는 가구 구분과 가구주 성별 변수에서 통계적으로 차이가 있는 것으로 나타났다. 소도시의 경우는 대도시와 거의 유사하지만 아파트 가구에서 주거 연면적에 따라 거부 성향 가구와 거부 가구 간의 차이가 있고, 주택에서는 가구 구분과 주거 연면적에 따라 각각 거부 성향 가구와 거부 가구 간에 차이가 있는 것으로 나타났다. 이처럼 몇 가지 경우를 제외하면 거부 성향 가구와 거부 가구 간에는 가구 특성에 따라 큰 차이가 없다고 볼 수 있다.

한편 <표 5-7>에서 거부 성향 가구와 거부 가구의 분포를 통해 어떤 가구에서 거부 전환이 더 잘 일어나는지를 이해할 수 있다. 지역규모에 관계없이 아파트와 주택 가구별로 살펴보면, 아파트 가구의 경우는 대체로 대도시 1인 가구의 거부 성향 비율은 16.8%, 소도시 1인 가구의 거부 성향 비율은 20.5%로 각 그룹의 거부 비율인 11.1%, 18.3% 보다 높다. 동일한 방법으로 비교해보면, 18평 이하 거주 가구, 30대 연령의 가구주 가구에서도 거부 성향 가구 비율이 거부 가구 비율보다 높게 나타났다. 주택의 경우는 아파트와 달리 대도시에서 가족 등 가구의 거부 성향 가구 비율이 84.6%, 거부 가구 비율이 73.1%로 거부 성향 가구 비율이 거부 가구 비율보다 10% 포인트 이상 높게 나타났다. 소도시에서도 이와 비슷하게 가족 등 가구에서 거부 성향 가구 비율이 79.8%, 거부 가구 비율이 68.5%로 거부 성향 비율이 거부 가구 비율보다 높은 것을 알 수 있다. 주택의 경우는 점유형태가 자가인 경우에 대해서도 비슷한 결과를 보였다. 실제 현장에서 거부 가구를 설득해서 응답을 얻어내는 것이 쉽지는 않지만, 접촉에 성공했을 경우, 아파트에서



는 혼자 거주하는 가구, 주거 연면적이 적은 가구가, 주택에서는 가족 등으로 구성된 가구이나 연면적이 너무 적지 않은 가구에서 설득이 잘 이루어진다고 볼 수 있다. 이처럼 가구 또는 가구주의 속성을 현장 조사에 착수하기 전에 미리 알 수 있다면 조사에 대한 성공률이 훨씬 높아질 것으로 짐작할 수 있다.

〈표 5-7〉 거부 성향 가구와 거부 가구 비교

(단위 : %)

항목	구분	대도시						소도시					
		아파트			주택			아파트			주택		
		4회 응답	거부	유약성	4회 응답	거부	유약성	4회 응답	거부	유약성	4회 응답	거부	유약성
가구구분	1인 가구	16.8	11.1		15.4	26.9	*	20.5	18.3		20.2	31.5	*
	가족 등	83.2	88.9		84.6	73.1		79.5	81.7		79.8	68.5	
점유형태	자가	68.1	62.4		51.3	48.6		55.5	61.8		65.7	60.8	
	전세 등	31.9	37.6		48.7	51.4		44.5	38.2		34.3	39.2	
주거 연면적	40이하(18평이하)	12.6	7.9		15.4	22.0		15.0	9.3		11.1	16.4	
	40~60(18~24평)	28.6	31.6		37.6	32.1		59.0	52.1	**	27.3	19.0	*
	60~85(24~34평)	43.7	43.8		29.9	27.2		20.5	29.6		25.3	37.8	
	85초과(34평초과)	15.1	16.7		17.1	18.8		5.5	9.0		36.4	26.8	
가구주 성별	남성	76.5	79.3		80.3	70.4	*	76.5	77.5		76.8	72.1	
	여성	23.5	20.7		19.7	29.6		23.5	22.5		23.2	27.9	
가구주 연령	20대이하	3.4	3.9		10.3	8.9		6.5	6.4		8.1	11.3	
	30대	22.7	18.3		16.2	18.8		28.5	24.7		6.1	11.0	
	40~50대	50.4	60.3		48.7	46.7		56.0	57.5		52.5	49.8	
	60대이상	23.5	17.5		24.8	25.7		9.0	11.4		33.3	27.9	
가구주 학력	중등이하	16.0	15.0		23.9	26.3		16.6	17.4		40.8	39.5	
	고등	29.4	29.1		38.5	41.0		41.2	39.4		30.6	32.2	
	대학교	54.6	55.9		37.6	32.8		42.2	43.2		28.6	28.4	

주 : χ^2 검정 결과 - *** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

이제 부재 성향 가구와 부재 가구 간의 차이를 살펴보자(〈표 5-8〉). 부재 성향 가구와 부재 가구는 지역 규모에 상관없이 대부분의 변수에 따라 그 성향에 차이가 있는 것을 알 수 있다. 이는 거부 성향 가구와 거부 가구가 대부분의 변수에 따라 차이가 없다는 결과(〈표 5-7〉)와는 상당히 다른 결과임을 알 수 있다. 특히 소도시 아파트 가구는 모든 변수에 따라 부재 성향 가구와 부재 가구 간에 차이가 있고, 대도시 주택 가구는 모든 가구주 관련 특성 항목에서 두 그룹 간에 차이가 있는 것을 알 수 있다. 지역규모와 거주 유형에 관계없이 가구 구분, 가구주 성별에 따라 부재 성향 가구와 부재 가구 간에는 차이가 있는 것으로 나타났다.

부재 성향 가구와 부재 가구 간의 특성에 대해서 구체적으로 비교해 보자. 대도시 아

파트 가구에서 가구 구분 항목에 해당하는 1인 가구 비율을 보면, 부재 성향 가구는 14.2% 이지만 부재 가구에서는 22.9%로 부재 가구에서의 1인 가구 비율이 부재 성향 가구에서보다 더 높다. 이는 재접촉 시도를 통해서 1인 가구는 접촉하기가 어렵다는 것을 의미한다고 볼 수 있다. 1인 가구의 비율은 도시별, 거주 유형에 상관없이 부재 성향 가구에 비해 부재 가구의 비율이 높게 나타나고 있다. 이 외에는 재촉이 어려운 가구로 자가 소유 가구보다는 전세 가구, 남성 가구주보다는 여성 가구주, 대도시 아파트에서는 주거 연면적이 큰 가구, 소도시 아파트에서는 주거 연면적이 적은 가구를 들 수 있다.

전체적으로 가구 특성 변수들에 대해서 부재 성향 가구는 부재 가구 간 특성에 있어서 차이가 있다. 이러한 가구들의 성향을 나타내는 항목들이 가구의 고용특성과 상관성이 있을 경우, 이 가구들에 대한 조사 누락은 통계 편향을 발생시킬 우려가 있다.

〈표 5-8〉 부재 성향 가구와 부재 가구 비교

(단위 : %)

항목	구분	대도시						소도시					
		아파트			주택			아파트			주택		
		4회 응답	부재	유여성	4회 응답	부재	유여성	4회 응답	부재	유여성	4회 응답	부재	유여성
가구구분	1인 가구	14.2	22.9	***	33.0	46.7	***	19.7	34.5	***	32.2	43.3	***
	가족 등	85.8	77.1		67.0	53.3		80.3	65.5		67.8	56.7	
접유형태	자가	61.8	61.8		37.0	33.1		60.3	54.7	**	67.2	61.9	*
	전세 등	38.2	38.2		63.0	66.9		39.7	45.3		32.8	38.1	
주거 연면적	40이하(18평이하)	9.1	9.0		30.3	34.1		11.2	15.0		18.8	16.6	
	40~60(18~24평)	37.1	31.4	*	35.5	31.6		48.2	53.7	***	24.8	25.8	
	60~85(24~34평)	39.8	43.2		22.0	22.5		31.6	25.3		30.5	34.0	
	85초과(34평초과)	13.9	16.5		12.2	11.8		9.0	6.0		25.9	23.6	
가구주 성별	남성	80.7	74.6	**	68.7	63.1	*	78.3	72.5	**	72.6	68.2	*
	여성	19.3	25.4		31.3	36.9		21.7	27.5		27.4	31.8	
가구주 연령	20대이하	5.2	4.0		11.2	16.3		5.8	8.2		9.1	8.1	
	30대	23.6	24.4		18.9	20.9	**	27.3	26.5	*	8.8	11.2	*
	40~50대	55.3	55.0		46.3	42.1		55.1	52.5		40.7	44.7	
	60대이상	15.9	16.6		23.6	20.8		11.8	12.7		41.4	36.0	
가구주 학력	중학교 이하	14.3	14.5		29.4	23.0		16.6	19.7		49.1	45.2	
	고등학교	33.6	31.6		40.0	36.3	***	37.3	37.5	*	29.9	33.7	
	대학교	52.1	53.8		30.6	40.7		46.1	42.8		21.1	21.1	

주 : χ^2 검정 결과 - *** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

3) 초기 응답 가구와 늦은 응답 가구의 특성 비교

조사기간 동안 초기에 응답한 가구와 늦게 응답한 가구들을 비교 분석해 보자. 이는



늦게 응답하는 가구들이 조사 초기에 응답하는 가구들에 대해 어떻게 다른가를 파악하기 위한 것이다. 이러한 분석은 현장에서 조사 대상 가구들에 대한 접촉 전략 또는 설득 전략을 세우는데 중요한 정보가 될 수 있다는 점에서 의미가 있다. 비교를 위해 방문회차 간 응답 가구들의 분포를 비교하고, 방문회차내에서 응답 가구와 무응답 가구의 분포를 비교하였다. 방문회차는 총 4회를 사용하였다. 초기 응답 가구와 늦은 응답 가구 간의 분석결과는 지역규모 또는 거처 유형에 따라 비슷한 결과를 보였다. 여기서는 그 중 하나의 결과로서 대도시 아파트 가구에 대한 분석결과만 제시하기로 한다. 대도시 아파트 가구는 현장에서 조사하기가 가장 어려운 대상으로 알려져 있기 때문에 이 가구들에 대한 분석결과를 중심으로 설명하고자 하였다.

<표 5-9>에서 보면 1회 방문에서 응답한 가구와 2회 방문에서 응답한 가구들(1회 응답 vs 2회 응답)은 주거 연면적과 가구주 학력을 제외한 모든 변수들에 따라 유의한 차이가 없었다. 그렇지만 방문회차가 증가할수록 가구주 성별을 제외한 모든 변수에서 1회 방문에서 응답한 가구와 3회 방문에서 응답한 가구 또는 1회 방문에서 응답한 가구와 4회 방문에서 응답한 가구들의 분포는 차이가 있는 것으로 나타났다. 한편 가구주 성별은 모든 방문회차 간 비교에서 유의하지 않았고, 이로부터 가구주 성별은 방문회차에 따라 응답 성향에 영향을 주지 않는다고 볼 수 있다. 이처럼 방문회차가 증가함에 따라 초기 응답 가구와 이후 방문회차간 응답 가구들의 분포에 차이가 있다는 것은 응답을 꺼려하는 가구들은 조사가 진행됨에 따라 여전히 조사에 협조하지 않으려는 성향이 있다는 것을 의미한다.

한편 방문회차내에서 응답 가구와 무응답 가구 간의 분포차이를 비교해 보면, 1회 방문에서 응답 가구와 무응답 가구 간에는 모든 변수에서 통계적으로 유의한 차이가 있는 것으로 나타났다. 이는 1회 방문에서는 쉽게 응답해 주는 가구와 무응답하는 가구 간에는 그 특성이 분명히 다르다는 것을 보여주는 결과라 할 수 있다. 방문회차가 증가함에 따라 이러한 차이는 약해지는 경향이 있지만, 가구 구분, 점유형태, 주거 연면적과 같은 변수들에서는 여전히 유의한 차이가 있어 보인다. 전체적으로, 1, 2회차 방문회차 내에서 응답 가구와 무응답 가구 간에는 모든 변수에 따라 그 성향이 다르다는 것을 알 수 있다. 그렇지만 방문회차가 3회, 4회로 갈수록 응답자와 무응답자의 성향이 유사해지는 경향이 있어 보인다. 이는 거부 가구에 대한 설득 또는 부재 가구에 대한 재접촉 시도 등에 따른 조사의 노력으로 각 회차내의 무응답 가구와 응답 가구 간의 분포가 유사해지는 것이라 할 수 있다. 방문회차에 따라 응답 가구와 무응답 가구 간의 가구 특성에 따른 분포가 유사해진다는 것은 응답 전환이 특정 가구에서만 이루어지는 것이 아니라 전체 가구에서 랜덤하게 이루어지는 현상이라 할 수 있다. 이러한 현상은 응답률 향상이 전체 표본에 대해 랜덤하게 이루어진다는 점에서 통계자료 입장에서는 긍정적인 효과를 줄 수 있을 것이다.

〈표 5-9〉 대도시 아파트 가구의 방문회차별 응답과 무응답 비교

항목	구분	방문회차 간 응답 가구 비교			방문회차 내 응답 가구와 무응답 가구 비교			
		1회응답 VS 2회응답	1회응답 VS 3회응답	1회응답 VS 4회응답	1회응답 VS 1회무응답	2회응답 VS 2회무응답	3회응답 VS 3회무응답	4회응답 VS 4회무응답
		가구구분	1인 가구		**	***	***	***
	가족 등							
점유형태	자가		**	***	***	***	*	
	전세 등							
주거 연면적	40이하(18평이하)							
	40~60(18~24평)	**	**	***	***	**		**
	60~85(24~34평)							
	85초과(34평초과)							
가구주 성별	남성							
	여성				*		*	**
가구주 연령	20대이하							
	30대							
	40~50대		***	***	***	***		
	60대이상							
가구주 학력	중등이하							
	고등	*	**	**	***			
	대학교							

주 : χ^2 검정 결과 - *** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

다. 무응답 편향 측정

1) 무응답 편향 정의

지금까지 응답 가구와 무응답 가구 간에는 어떤 특성이 있고 그 차이는 무엇인지에 대해 살펴보았다. 표본조사에서 가장 관심이 되는 것 중에 하나가 응답률이다. 그렇다면 왜 조사 연구자들은 응답률에 관심을 갖는 것일까? 과거에는 응답률이 곧 통계의 신뢰성을 대표하기도 하였다. 그와 함께 더 관심을 가져야 할 사항은 모든 통계에서 무응답이 발생한다는 것이고, 이러한 무응답이 실제 현장에서는 랜덤하게 발생하지 않는다는 것이다. 무응답이 랜덤하게 발생하지 않는다는 것은 어떤 특별한 성향을 갖는 가구들이 응답을 해주지 않는다는 것이다. 이와 같은 현상은 응답률에 크게 관계없이 통계 편향을 유발할 수 있게 되고, 응답률이 낮게 되면 더 큰 편향을 초래하게 된다는 것이 많은 경험으로부터 얻은 결과일 것이다. 따라서 조사통계를 작성함에 있어서 일정한 비율의 무응답이 발생할 경우 무응답 편향은 반드시 측정되어야 할 사항이 되어 버렸다.



편향은 조사값(조사추정치)과 모수값(참값)간에 발생하는 차이를 말한다. 표본조사에서 편향은 소위 조사의 대상이 되는 모든 표본을 바탕으로 계산되며 조사값과 모수값간에 발생하는 차의 기댓값으로 생각할 수 있다. 추정치와 관련이 있는 무응답 편향은 두 개의 성분으로 구성되어 있다. 무응답의 양적(무응답률)크기와 응답자 추정치와 무응답자 추정치 차이가 그것이다. 어떤 추정치가 갖는 편향은 수학적으로는 두 요소들 간의 관계를 보여주기 위한 것으로 표현될 수 있으며, 그 식은 다음과 같다.

$$Bias(\hat{y}_r) = p_m \{E(\hat{y}_r - \hat{y}_m)\}$$

여기서 \hat{y}_r 은 응답자들만에 의해 추정된 특성값이고, p_m 은 무응답률, \hat{y}_m 은 무응답자들만으로 추정된 특성값, E 는 모든 가능한 표본에 대한 평균을 의미하는 것으로 기댓값 연산자를 나타낸다(Nolin et al., 2000). 위의 식에 따르면 무응답 편향은 응답자 평균이 모집단 평균에 대해 얼마나 다른지에 대한 정도를 측정할 수 있다. 이와 같은 식은 조사에서 측정오차와 같은 편향을 일으킬 수 있는 다른 근원은 없다고 가정한다(Groves, 2006).

무응답 편향은 위의 식에서 보는 것과 같이 낮은 응답률과 응답자 추정치와 무응답자 추정치 간에 발생하는 차이 둘 모두에 관계되어 있다. 이러한 편향을 피할 수 있는 좋은 방법은 강력한 거부전환 전략, 좋은 인센티브, 혼합조사모드를 이용한 자료수집, 탄력적 조사 스케줄 및 조사원 훈련 등과 같은 방법을 사용하여 응답률을 올리는 것이다. 그러나 이러한 다양한 노력들에도 불구하고 무응답은 발생하기 마련이다. 그런 경우에 우리는 무응답 자료를 보충하기 위해 확률적 방법에 의한 가중치를 통해 조사자료를 조정하기도 한다. 그렇지만 무응답 가중치 조정을 통해 어느 정도의 무응답 편향을 줄일 수 있지만 완전히 제거할 수는 없다.

2) 실업률과 연관성 있는 변수 탐색

본 연구에서 사용된 조사대상 가구는 지역별로 가구의 고용현황 파악을 목적으로 사용되었다. 이 조사의 무응답 편향을 보다 정확하게 측정하기 위해서는 조사의 목적과 부합하는 고용상태에 관련된 변수를 이용하는 것이 가장 좋다. 그렇지만, 연구에 사용된 자료가 총조사 변수를 포함하고 있지만, 이는 총조사 전수 항목으로 고용 관련 변수는 전수항목에 포함되지 않는다. 즉, 고용상태를 직접적으로 측정할 수 있는 변수는 없다는 것이다. 이런 경우 무응답 편향은 조사목적에 직접 달성할 수 있는 변수와 연관성이 있는 변수를 이용할 수 있다(Heerweh et al., 2007). 본 연구에서는 고용상태(실업률, 고용률)와 관련이 있는 변수를 이용하여 무응답 편향을 측정하고자 하였다.

고용상태와 연관성이 있는 변수는 상관분석 및 회귀분석을 통해 탐색하였다. 이때

각 변수에 대해 유의수준 0.05를 기준으로 통계적 유의성을 검증하였다. 고용상태와 연관성이 있는 변수 탐색은 고용항목을 포함하고 있는 2010 총조사 10% 자료를 이용하였다. 무응답 편향은 가구 특성 변수의 각 범주에 해당하는 가구의 비율을 대상으로 할 것이기 때문에, 각 특성변수의 범주에 포함되는 가구의 비율 변수를 시군구 수준에서 새롭게 생성하였다. 구체적인 변수들로는 1인 가구, 아파트 거주 비율, 자기집 소유 비율, 동거아동수 있는 가구 비율, 18평~34평대 거주하는 가구 비율, 가구주 연령이 4-50대 가구 비율, 가구주 학력이 대학교 이상 가구 비율, 가구주가 유배우자인 가구 비율을 사용하였다. 이 변수들과 실업률 또는 고용률과의 상관분석을 실시한 결과, 모든 변수들은 실업률에 대해 유의수준 0.05에서 유의한 상관관계를 갖는 것으로 나타났다.

〈표 5-10〉 실업률과 센서스 항목의 상관분석

	1인 가구	아파트	자기집	18-34평	4-50대	대학 이상	유배우	동거아동 있음	남자
상관계수	-0.617	0.622	-0.774	0.644	0.473	0.659	0.157	-0.578	0.354

가구 특성 변수는 실업률에 대해 일차원적이 아닌 다차원적으로 영향을 미칠 가능성이 있기 때문에 여러개의 변수를 동시에 고려한 다중회귀분석을 실시하였다. 단계적 회귀분석(stepwise regression)을 실시한 결과는 <표 5-11>과 같다. 최종 선택된 변수들에 대한 회귀분석에서 1인 가구, 아파트 가구, 자기집 소유, 거주 연면적이 18-34평대, 가구주 연령이 4-50대, 가구주 학력이 대학교 이상, 가구주 혼인상태가 유배우자인 가구 등과 같은 변수는 실업률에 영향력 있는 변수인 것으로 나타났다.

〈표 5-11〉 실업률과 총조사 변수와의 단계적 회귀분석 결과

모형	회귀계수	표준오차	t값	유의확률
(상수)	4.253	.430	9.900	.000
1인 가구 비율	-.021	.006	-3.569	.000
아파트 가구 비율	-.009	.002	-4.690	.000
자기집 비율	-.014	.002	-7.292	.000
18-34평 비율	-.005	.002	-3.119	.002
40-50대 비율	.018	.004	4.170	.000
대학교 이상 비율	.004	.002	2.044	.042
유배우 비율	-.023	.004	-5.066	.000



3) 무응답 편향 계산

위에서 실업률과 연관성이 있는 변수로 1인 가구, 아파트 가구, 자기집 소유, 거주 연면적이 18-34평대, 가구주 연령이 4-50대, 가구주 학력이 대학교 이상, 가구주 혼인상태가 유배우자인 가구 비율 변수가 선택되었다. 이 변수들에 대해 무응답 편향을 측정하였다. 실업률과 연관성이 있는 변수들에서 발생하는 편향의 정도는 실제로 실업률 측정에서 나타나는 무응답 편향과 무관하지 않다고 볼 수 있다. 각 변수에 대한 무응답 편향은 무응답 유형을 거부와 부재로 나누어 측정하였다. 무응답 편향 중 거부 무응답에 의한 편향과 부재 무응답에 의한 편향이 차지하는 비중이 다를 수 있고, 이로부터 거부 편향과 부재 편향 중 어떤 것이 무응답 편향에 더 크게 발생하는지 알 수 있다.

앞의 3.나 절에 의하면 응답유형별 가구 특성변수들의 분포는 유의수준 0.05에서 통계적으로 유의한 차이가 있었다. <부표 I>은 가구 특성 변수에 대한 응답률, 부재율, 거부율을 나타낸 것이다. 만약 우리가 가구 특성 변수 중 가구구분 변수에서 1인 가구 비율을 추정하고자 할 때, 전체 유효표본가구(불능가구 제외)의 23.77%가 1인 가구이고 나머지(76.2%)는 가족 등으로 구성된 가구이다. 이 중에서 응답 가구 중 23.3%는 1인 가구이고 나머지는 76.7%이다. 전체 가구의 1인 가구 비율에 대해 응답 가구의 1인 가구 비율이 동일하지는 않다. 이는 1인 가구 비율이 조사를 통해 작게 추정될 수 있음을 의미한다.

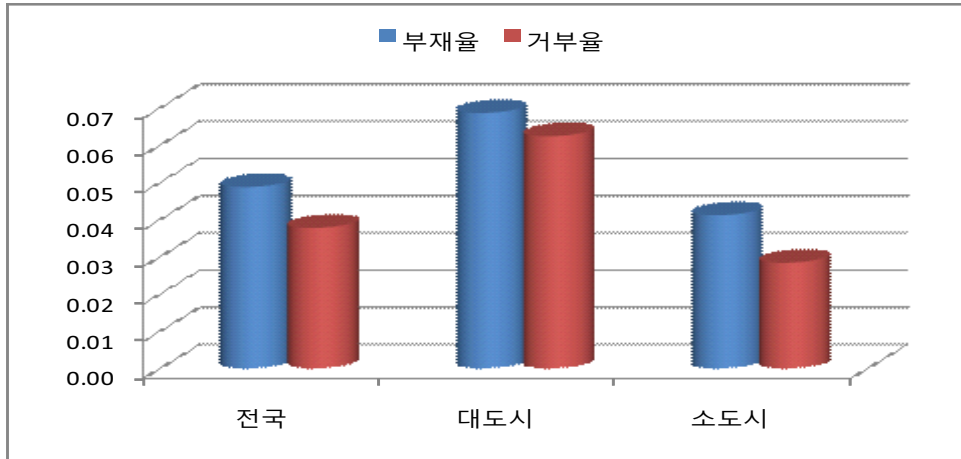
무응답 편향을 구체적으로 계산해 보자. 만약 표본조사를 통한 우리의 목표 추정치가 1인 가구 비율일 경우, 이에 무응답 편향은 -0.0046 또는 -0.46%(23.30%-23.77%)이다. 이 계산은 식(2.1)을 이용한 것으로 다시 써보면 식(3.1)과 같다.

$$\bar{y}_r = \bar{y}_n + \left(\frac{m}{n}\right)[\bar{y}_r - \bar{y}_m] \quad (3.1)$$

여기서 r 은 응답자수, n 은 총 표본수, m 은 무응답수를 말하고, 여기서 \bar{y}_r (0.2330)은 \bar{y}_n (0.2377)에 무응답 편향 -0.0046을 더한 것과 같다. 따라서 \bar{y}_m 은 0.2862와 같다. 이것은 유효표본의 응답하지 않은 가구의 28.62%는 1인 가구라는 것을 의미한다. 이는 응답 가구의 23.30%가 1인 가구라는 것과 차이가 있다.

무응답 유형별 편향의 정도를 살펴보기 위해 무응답 가구를 부재와 거부 가구로 나누어 계산해 보았다. 식(2.1)의 구체적인 식을 다시 이용하면 부재에 의한 편향은 -0.0062, 거부에 의한 편향은 0.0015로 이 두 편향을 더한 총 편향은 -0.0047이다. 1인 가구 비율을 추정한다고 했을 때, 응답 가구만으로 모수를 추정할 경우 0.47% 포인트 정도 1인 가구 비율이 과소하게 추정될 가능성이 있다. 총 무응답 편향 중에서는 부재에 의한 편향이 거부에 의한 편향보다 더 많은 비중을 차지하는 것을 알 수 있다.

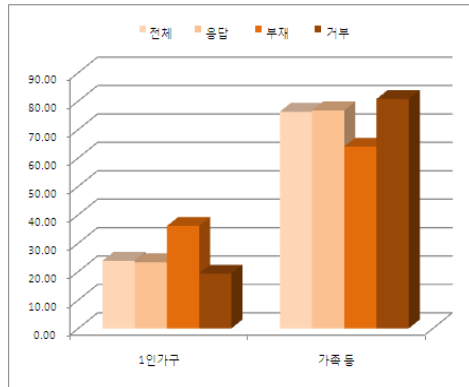
총 무응답률에 대해서 부재율과 거부율은 2절에서 이미 살펴보았다. [그림 5-8]은 2절의 결과를 다시 표현한 것으로 전체적으로 이 조사는 부재율이 거부율보다 높고, 이런 차이는 소도시 이하(경기도를 제외한 8개 도)에서 더 크게 나타났다.



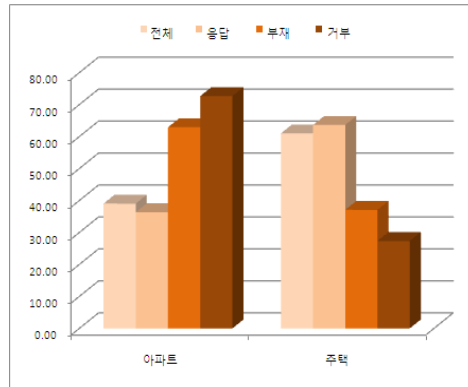
[그림 5-8] 전국, 대도시, 소도시 이하 지역의 부재율과 거부율

<부표 I>을 통해서 응답 가구와 부재 가구 또는 응답 가구와 거부 가구 간에는 그 성향에 있어서 어느 정도 차이가 있는지 살펴보았다. <부표 I>은 전국, 대도시, 소도시 별로 응답 유형에 따라 차지하는 가구들의 특성에 대한 분포를 나타낸 것이다. [그림 5-9_1]은 가지 주요 변수에 대해서만 그 분포를 그림으로 나타낸 것이다. 전체적으로 부재 가구의 1인 가구 비율이 전체 가구 또는 응답 가구의 1인 가구 비율보다 높다. 부재 가구의 아파트 가구 비율, 주거 연면적이 24평 이하 가구 비율, 가구주 연령이 40-50대 가구 비율, 가구주 학력이 대학교 이상의 가구 비율이 응답 가구의 동일 특성에 비해 그 비율이 높은 것으로 나타났다. 마찬가지로 거부 가구의 가족 등 가구 비율, 아파트 가구 비율, 주거 연면적이 24-34평 가구 비율, 가구주 연령이 40-50대 가구 비율, 가구주 학력이 대학교 이상 가구 비율이 응답 가구의 동일 특성 비율보다 높게 나타났다.

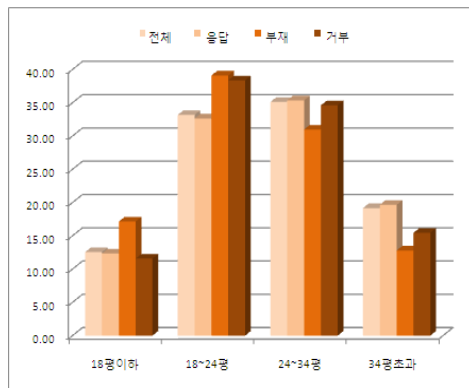
이러한 현상은 거주 유형별로 약간 다른 경향이 있다. <부표 I>과 [그림 5-9_2]에 의하면, 아파트 가구의 부재 가구에서 1인 가구 비율이 월등히 높고, 거부 가구의 가족 등 가구 비율이 높다. 한편 전체 가구의 성향과 달리 아파트 거부 가구 중 주거 연면적이 34평을 초과하여 거주하는 가구의 비율이 응답 가구 중 34평을 초과하여 거주하는 가구의 비율에 비해 높고, 아파트 가구내에서 거부 가구와 부재 가구는 가구주 학력에 따라서는 큰 차이가 없는 것으로 나타났다. 이 외의 다른 항목에 대한 구체적인 수치는 <부표 I>을 참고할 수 있다.



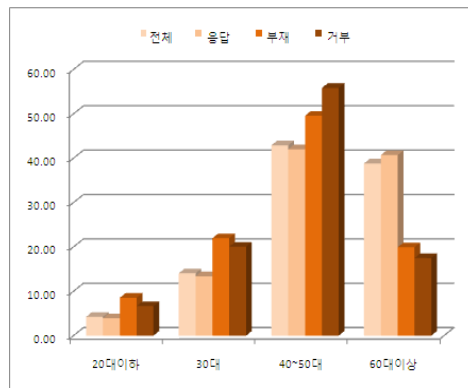
〈전체〉 가구구분



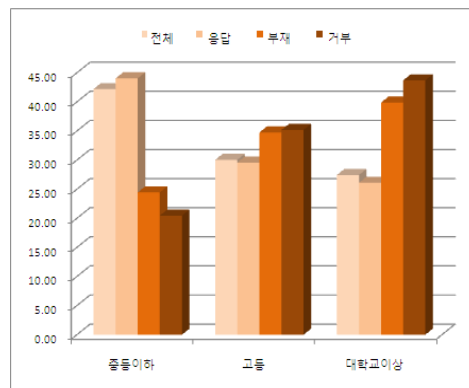
〈전체〉 주택유형



〈전체〉 주거연면적

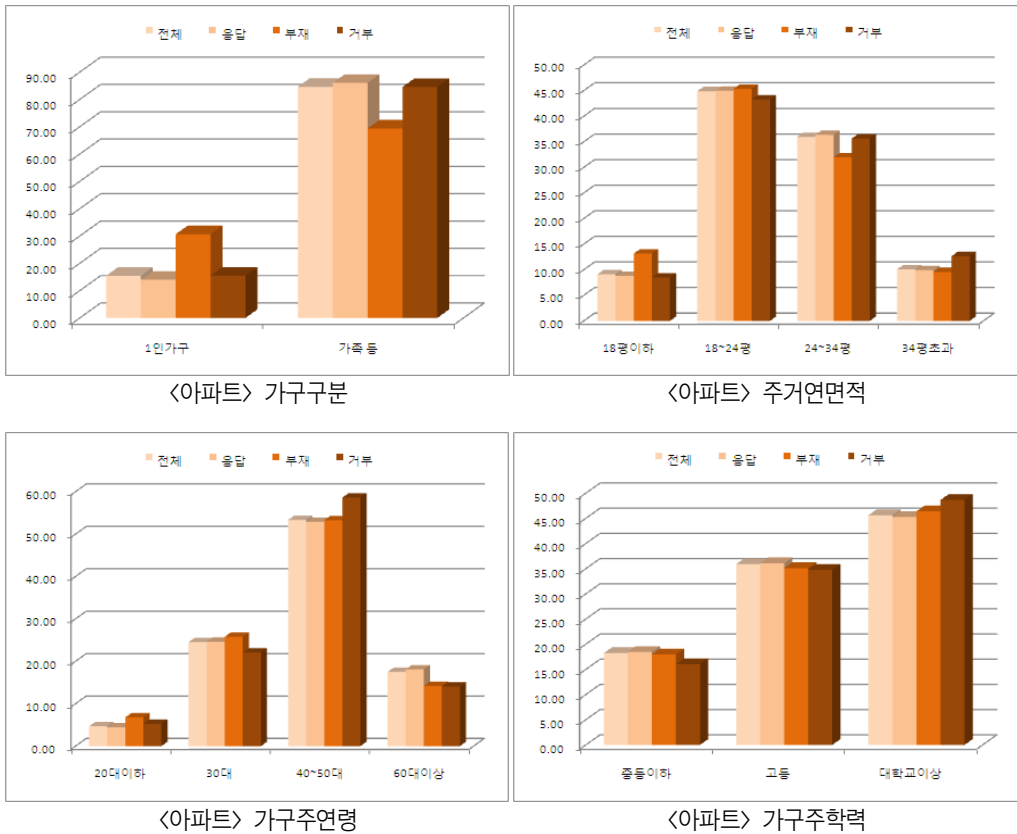


〈전체〉 가구주연령



〈전체〉 가구주학력

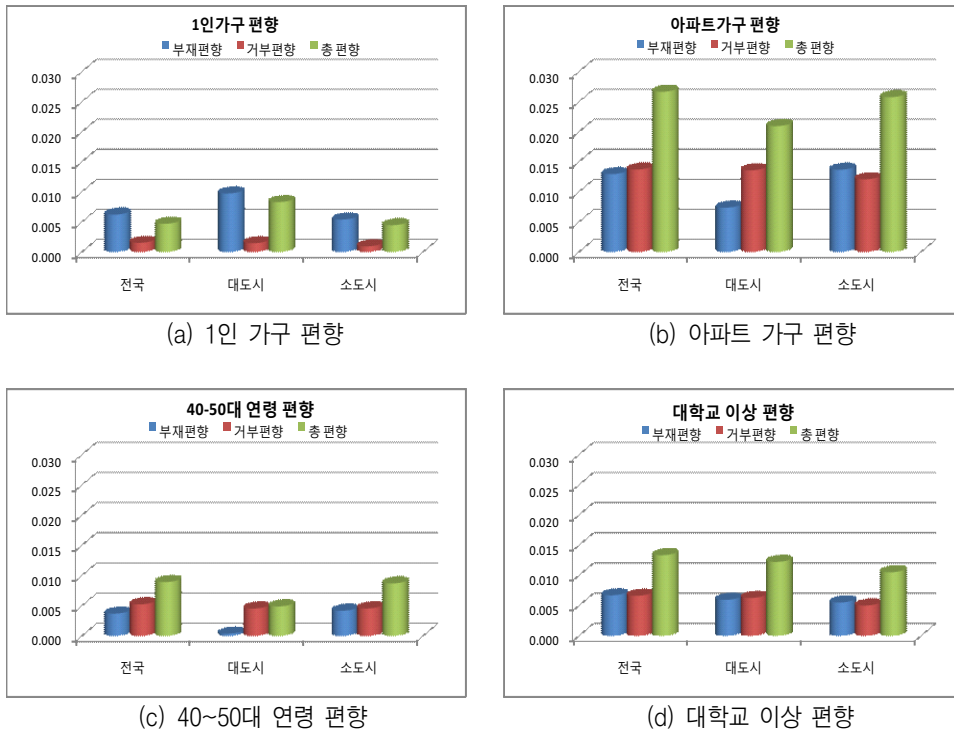
[그림 5-9_1] 전체, 응답, 부재, 거부 가구의 가구 특성별 분포



[그림 5-9_2] 아파트 가구에서 전체, 응답, 부재, 거부 가구의 가구 특성별 분포

[그림 5-10]은 지역규모별 가구 특성에 해당하는 몇 개 변수들의 무응답 편향을 무응답 유형별로 나타낸 것이다. 예를 들어 이 조사에서 1인 가구 비율을 추정하는 것이 목표라고 하면, 이 조사에서는 무응답 가구의 발생으로 인해 무응답 편향이 초래된다고 볼 수 있다. [그림 5-10]의 (a)를 보면 1인 가구의 비율을 추정하는데 있어서 부재에 의한 편향이 거부에 의한 편향보다 더 크게 발생하고, 이러한 현상은 소도시에서보다 대도시에서 더 크게 나타난다. 이 그림 (b)에서는 아파트 비율이 목표 추정치인 경우에 해당한다. 전체적으로 부재 편향과 거부 편향은 비슷한 수준으로 발생한다. 그렇지만 대도시에서는 거부 편향이, 소도시에서는 부재 편향이 더 크게 발생하여 지역에 따라 무응답 성향이 조금 다르게 나타난다고 볼 수 있다. 가구주 연령이 40-50대인 비율을 추정한다고 할 때 전체적으로 거부 편향이 크고, 대도시의 경우 무응답 편향은 대부분 거부에 의한 편향이라는 것을 알 수 있다([그림 5-10]의 (c)). 가구주 학력이 대학교 이상인 가구의 비율을 추정한다고 할 때, 부재와 거부 편향은 비슷한 수준으로 총 편향에 영향을 미치는 것을 알 수

있다([그림 5-10]의 (d)). 이 외의 변수들에 대해서도 유사한 방법으로 계산하여 설명해 볼 수 있다. 지면상 본문에서는 그 설명을 생략하고, 부록의 <부표 II>을 참고할 수 있다.

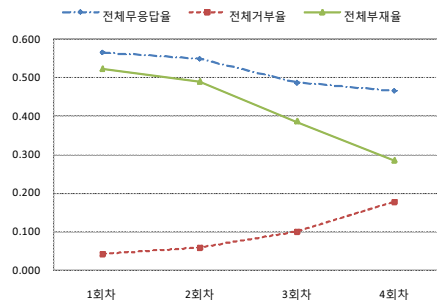


[그림 5-10] 도시규모별 몇 개의 가구특성에 대한 무응답 편향

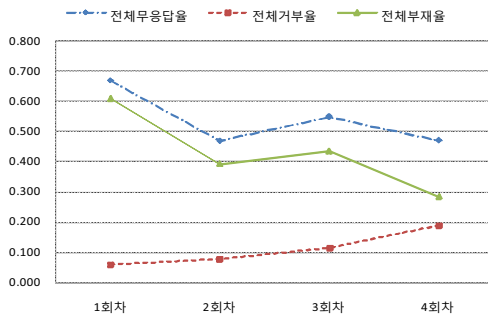
[그림5-10]과 <부표 II>에 비추어 종합해 볼 때, 이 조사의 경우 아파트 가구 비율과 가구주 연령이 40-50대인 가구의 비율을 제외하면 총 무응답 편향 중에서 부재에 의한 편향이 거부에 의한 편향보다 더 크고, 부재율이 거부율보다 더 높다는 것을 알 수 있다. 즉, 부재 가구의 특성값과 응답 가구의 특성값과의 차이와 높은 부재 비율 모두가 부재 편향을 크게 하는 원인이 되었다고 볼 수 있다. 거부 편향의 경우를 보면, 아파트 가구 비율을 추정하고자 할 때, 거부율이 부재율보다 낮지만, 무응답 편향 측면에서는 거부 편향이 부재 편향보다 더 큰 것으로 보아 거부 가구의 아파트 가구 비율과 응답 가구의 아파트 가구 비율의 차이가 부재 가구에서보다 거부 가구에서 더 크다고 볼 수 있다. 이 결과를 토대로 우리나라 가구단위 조사는 부재율이 거부율보다 높고, 몇 가지 특성 가구 변수를 제외하면 부재에 의한 편향이 총 편향에 기여하는 정도가 더 크다고 볼 수 있다.

우리나라 대부분의 표본조사의 조사대상 가구가 총조사를 추출틀로 사용한다고 할 때, 거의 모든 가구 단위 조사에서 발생할 수 있는 현상이라고 조심스럽게 예상할 수 있다.

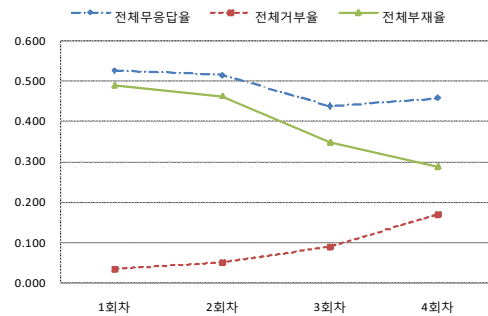
이제 가구에 대한 방문회수가 무응답 편향에는 어떤 영향을 미치는가를 살펴보도록 하자. 이에 대해 여러 가지 의견들이 있지만 일관성이 있는 것은 아닌 것 같다. 즉, 나라마다 조사마다 그리고 변수마다 그 무응답 현상은 다를 수 있다는 것이다. [그림 5-11] ~ [그림 5-14]는 가구 특성을 설명할 수 있는 4개 변수에 대해 가구 방문회수에 따른 거부, 부재, 총 편향과 각각의 무응답 비율을 거쳐 유형별로 나타낸 것이다. 다른 변수들에 대한 무응답 편향은 <부표 III>에 제시하였다.



(a) 전국 회차별 무응답률, 거부율, 부재율



(b) 대도시 회차별 무응답률, 거부율, 부재율

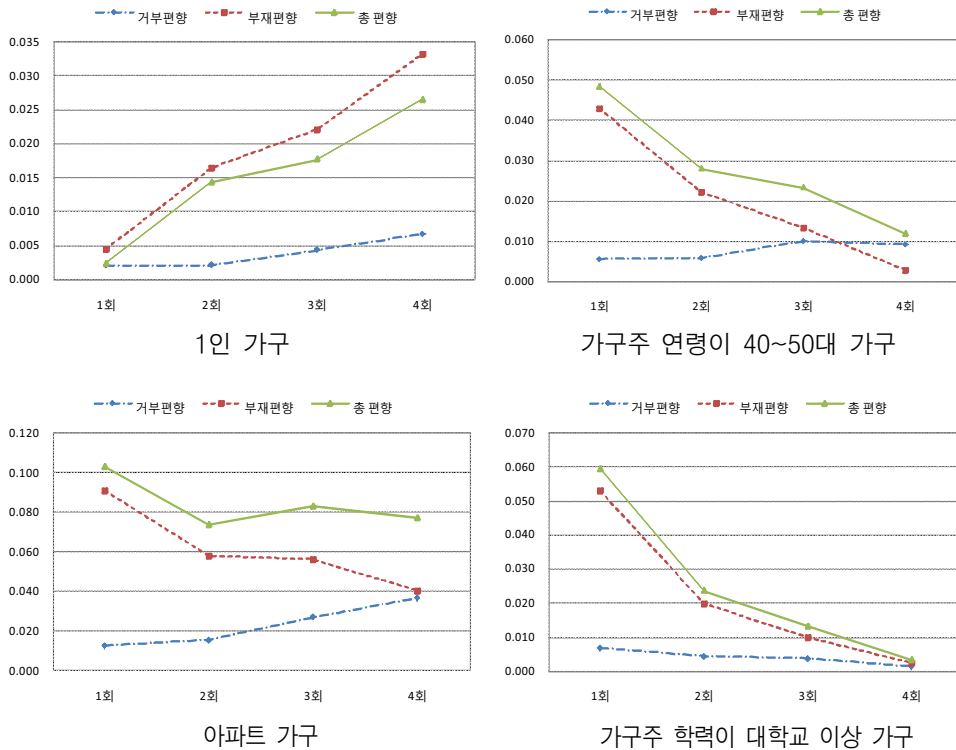


(c) 소도시 회차별 무응답률, 거부율, 부재율

[그림 5-11] 회차별 무응답률, 거부율, 부재율

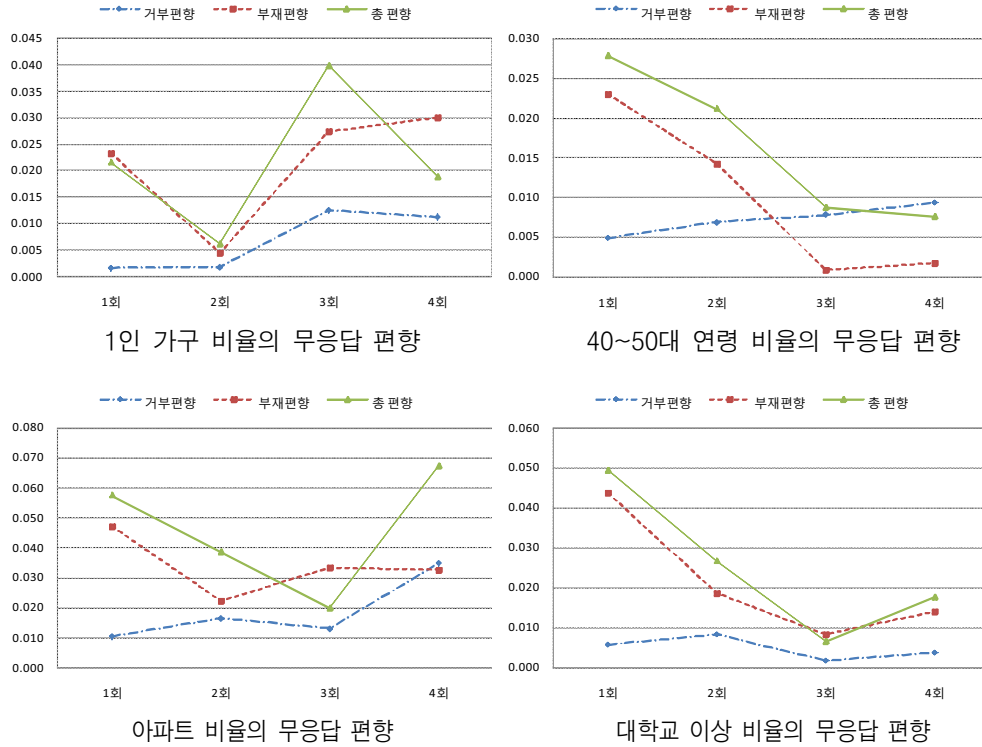
[그림 5-11]은 방문회수가 증가함에 따라 전체 무응답률이 감소하는 경향이 있다. 또한 방문회수가 많아질수록 부재율은 감소하지만 거부율은 증가하는 경향이 있다. 이 그림으로부터 방문회수가 많아질수록 조사에 대한 응답률은 향상되지만 그에 대해 거부율도 증가한다는 것을 알 수 있다. [그림 5-12] ~ [그림 5-14]는 방문회수에 따라 무응답 편향을 구성하는 정도가 다르다는 것을 보여준다. [그림 5-12]에서 1인 가구 비율을 제외하

면 대체로 방문회수가 증가할수록 무응답 총 편향과 부재 편향은 감소하고 거부 편향은 약간 증가하는 경향이 있다. 1인 가구 비율 추정에서는 방문회수가 증가함에 따라 부재 편향은 거부 편향에 비해 급격하게 증가하는 것을 알 수 있다. 이와 같은 현상은 1인 가구를 그 만큼 접촉하기 어렵다는 것을 의미하고 방문회수가 증가할수록 1인 가구의 부재율이 증가하게 되어 부재 편향이 커지는데 기여하게 되는 셈이다.



[그림 5-12] 전국 1인 가구, 40-50대, 아파트, 대학교 이상 비율에 대한 무응답 편향

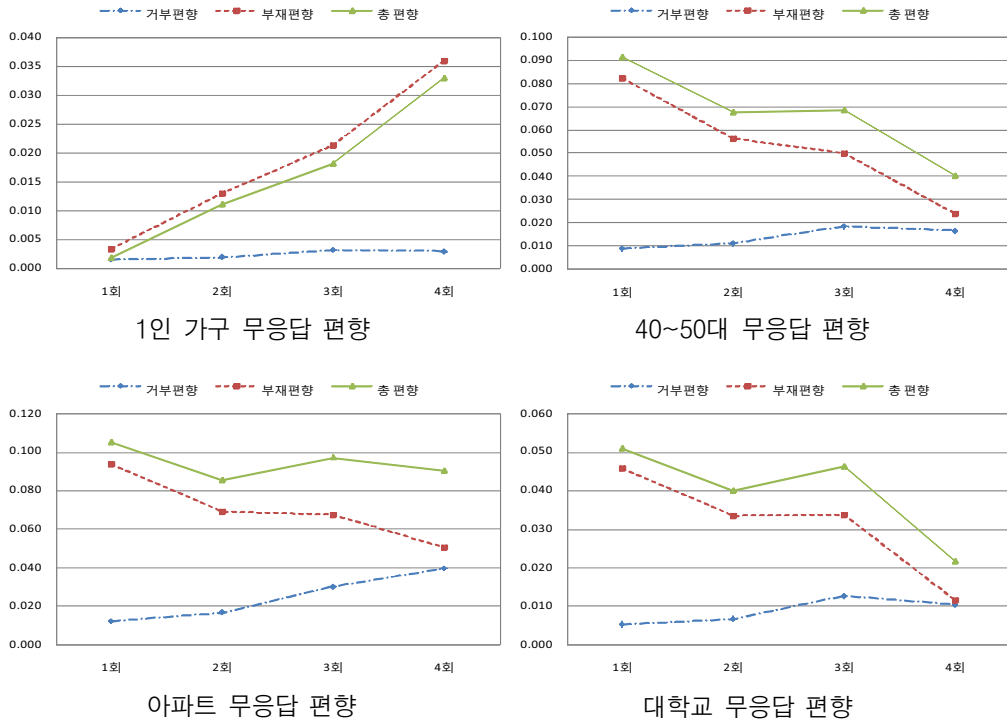
그리고 대도시의 아파트거주자와 대학교 이상의 가구주가 있는 가구에서는 마지막 회차에서 거부, 부재, 총 편향 모두 증가하는 경향이 있어 보인다([그림 5-13]). 대도시의 1인 가구 비율에 대한 무응답 편향은 전체 경향과 달리 방문회수가 증가함에 따라 그 정도가 감소하는 경향이 있다. 대도시는 대체로 방문회수가 4회째 접어들면서 무응답 편향이 증가하는 경향이 있다. 이 조사에서 방문회수가 4회로 제한되어 있기 때문에 이 결과만으로 일반적인 결론을 내리기는 어렵지만, 대도시(경기도를 포함한 특광역시)에서는 방문회수가 많아질수록 보다 신중하게 무응답 편향을 고려할 필요가 있어 보인다.



[그림 5-13] 특광역시 + 경기도 등 대도시 1인 가구, 40~50대, 아파트, 대학교 이상 비율

소도시의 경우([그림 5-14])는 전체 경향과 매우 유사한 경향을 보인다. 전체 경향에서 1인 가구 비율에 대한 무응답 편향이 방문회수가 증가함에 따라 커지는 현상은 대도시보다는 소도시의 무응답 경향 때문이라고 볼 수 있다. 즉, 1인 가구의 비율이 대도시보다 농·어촌을 다수 포함하고 있는 소도시에서 더 높고, 이들 1인 가구에서 부재 가능성은 거부보다 더 높다는 것을 이미 앞에서 살펴보았다. 따라서 방문회수가 증가할 수록 1인 가구는 접촉하기가 더 어렵고 이러한 현상은 대도시에서보다 소도시에서 더 심각하게 발생하기 때문에 1인 가구 편향이 더 크게 발생한다고 볼 수 있다.

종합적으로 [그림 5-11] ~ [그림 5-14]에 의하면 방문회수가 증가함에 따라 전체적으로 무응답 편향은 감소하는 경향이 있고, 거부 편향은 약간 증가하기도 하지만 부재 편향은 확실하게 감소하는 경향이 있다. 다만 이 조사가 농·어촌이 다수 포함되어 있는 조사라는 점에서 볼 때, 소도시에서 연령이 높은 1인 가구 비율이 많고, 1인 가구의 부재율이 높다는 점에서 무응답 편향이 크게 나타나는 현상으로 이해할 수 있다.



[그림 5-14] 8개 도(경기도 제외) 소도시 1인 가구, 40~50대, 아파트, 대학교 이상 비율

라. 응답여부에 영향을 미치는 변수 탐색

본 절에서는 조사대상 가구의 응답 여부에 영향을 미치는 변수를 탐색하고자 한다. 이를 위해 널리 사용되고 있는 의사결정나무 분석을 이용하였다. 의사결정나무 분석에 사용할 수 있는 알고리즘은 주로 카이제곱 통계량, 지니지수, 엔트로피 지수 등이 있고, 일반 상용 프로그램을 통해 쉽게 사용할 수 있다. 우리가 사용하는 자료의 변수가 모두 범주형이라는 점에서 카이제곱 통계량에 의한 CHAID(Chi-squared Automatic Interaction Detection) 알고리즘을 사용하였다. 최종 응답 여부 변수는 ‘응답’과 ‘무응답’ 두 개의 범주를 갖는 이항 범주형 변수이다. 이 변수와 상관도 높은 변수들을 찾기 위해 가구구분, 가구주 학력, 가구주 혼인상태, 거처종류, 점유형태, 동거아동수, 주거연면적, 거주주 성별, 가구주 연령을 분류변수로 사용하였다. 이 외에도 여러 가지 다른 변수들도 고려하였지만, 분류변수들 간에 상관관계가 높거나 의미적으로 유사한 변수들은 사전 분석과정에서 제외하였다. 예를들면 가구원 수(1명/2명/3명 이상, 3개 범주) 변수는 가구 구분(1인 가구/가족가구, 2개 범주) 변수와 거의 동일한 의미를 갖는 변수에 해당한다.

우선 전국 수준에서의 분류변수들과 결과변수(응답 여부)간의 연관성 정도를 살펴보았다. 다음으로 전국을 대도시(특광역시, 경기도)와 소도시(8개 도)로 크게 나누어 분류 분석하였다. 지역규모별 구분은 향후 무응답 가중치 적용을 고려한 것으로 통계 공표 범위 또는 수준에 따라 다르게 분류할 수도 있다. 즉 수도권(서울, 경기, 인천), 인천을 제외한 광역시, 경기도를 제외한 도 지역 등과 같이 구분할 수도 있다. 지역규모에 대한 세부적인 구분은 향후 연구를 통해 수행하고자 한다.

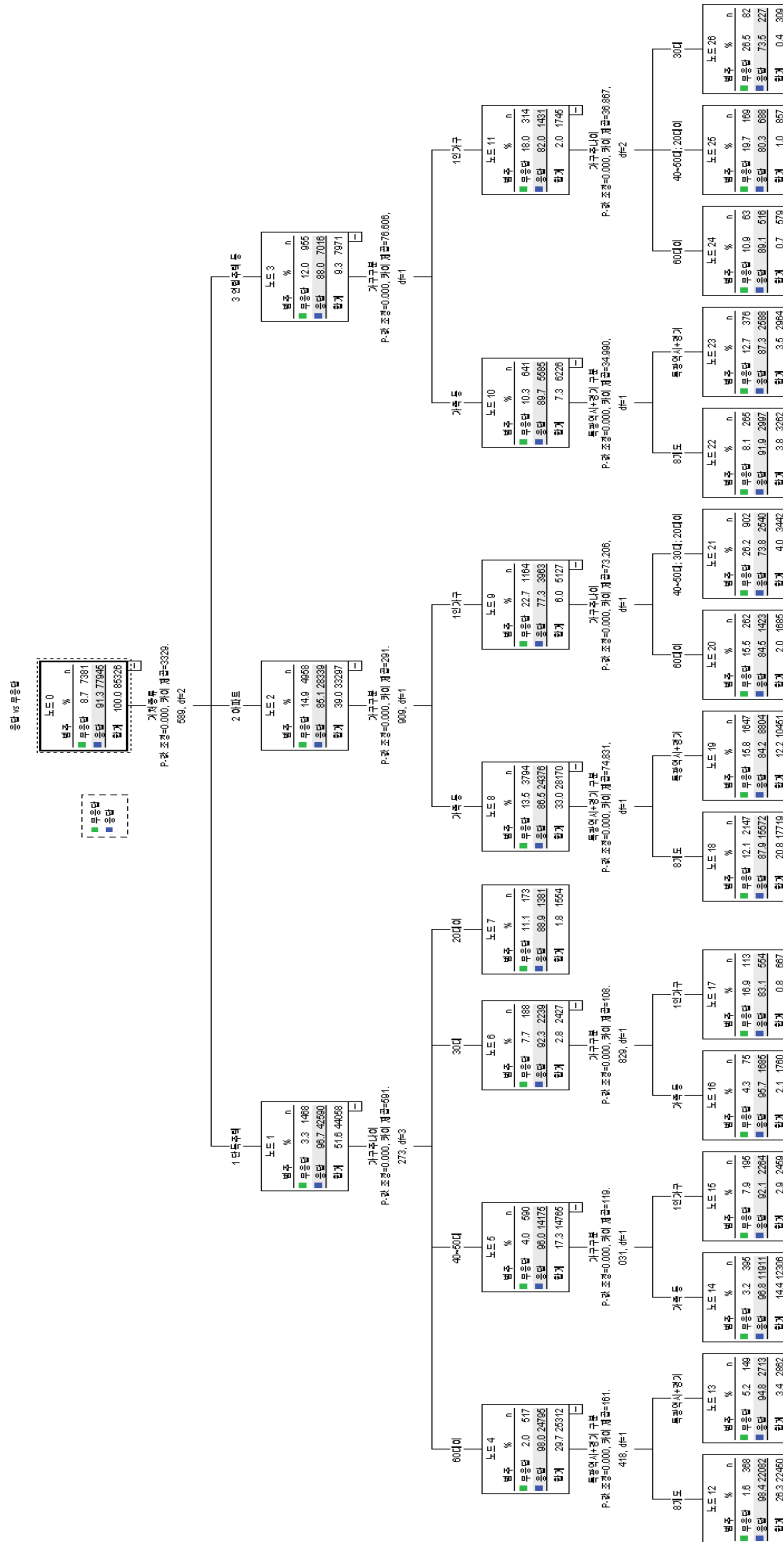
[그림 5-15]는 전국 수준에서 분석한 결과이다. 응답여부에 가장 상관도가 높은 변수는 거주 유형(아파트, 단독주택, 연립주택 등)으로 나타났다. 이렇게 분리된 후에 아파트 및 연립주택 등에서 결과변수와 상관도가 높은 변수는 1인 가구 여부이고, 단독주택내에서 결과변수와 상관도가 높은 변수는 4개 범주를 갖는 연령 변수인 것으로 나타났다. 그 이후 노드에서는 상위 노드에 대해서 조금씩 다른 경향이 있긴 하지만, 연령, 지역구분, 1인 가구 여부 등의 3개 변수가 각각 상위 노드에 영향이 있는 것을 알 수 있다.

[그림 5-16]은 경기도와 특광역시를 포함한 대도시 지역에 대해 분류한 결과이다. 전국 수준에서와 마찬가지로 결과변수와 상관도가 가장 높은 변수는 거주 유형인 것으로 나타났다. 그 다음은 1인 가구 여부가 결과변수에 대해 영향력이 있는 것으로 나타났다. 마지막 노드에서는 연립주택을 제외하면 1인 가구의 경우 가구주 연령, 가족 등 가구에서는 가구주 성별이 영향력 있는 변수로 나타났다. 연립주택의 마지막 노드에서는 1인 가구의 경우 가구주 학력, 가족 등 가구에서는 동거아동수 유무가 결과변수에 영향을 주는 것을 알 수 있다. 소도시의 경우([그림 5-17])도 대도시에 비해 크게 다르지 않다.

[그림 5-17]은 소도시에서의 의사결정나무 분석결과를 나타낸 것이다. 소도시에서 응답 여부에 가장 영향력 있는 변수는 1차 노드에 해당하는 거주 유형 변수로 나타났다. 2차 노드의 거주 유형 중, 단독주택의 경우 가구주 연령이 결과변수와 상관도가 높은 반면, 그 외 유형에서는 1인 가구 여부가 결과변수와 상관도가 높은 것으로 나타났다. 마지막 노드에서는 상위 노드에 대해 가구주 연령, 1인 가구 여부, 점유형태와 같은 변수들이 상관도가 있는 것으로 나타났다.

우리나라의 경우 아파트와 주택에 거주하는 각각의 가구들은 응답, 부재 또는 거부 성향이 조금씩 다르다고 볼 수 있다. 이와 같은 현상은 지역, 가구 구분, 가구주 연령, 학력 등과 같은 가구내의 속성을 나타내는 변수들이 서로 다를 수 있다는 데에서 비롯되는 것 같다. 이런 점을 고려하여 거처 유형별로 나누어 동일한 분석을 실시하였다. 거처 유형을 아파트와 단독주택으로 나누어 살펴보았다. 아파트와 단독주택 거주 가구에 따라 결과변수와 상관도가 높은 변수가 각각 다른 것을 알 수 있다. 첫 번째 노드에서 아파트는 1인 가구 여부가 결과변수와 상관도가 가장 높고, 단독주택에서는 가구주 연령이 상관도가 높은 변수인 것으로 나타났다.





[그림 5-15] 전국 수준에서의 의사결정나무

지금까지의 의사결정나무 분석 결과를 종합해 보면, 전체적으로 응답 여부와 상관도가 가장 높은 변수는 거처 유형과 1인 가구 여부(또는 가구원수)변수라 할 수 있다. 그 다음으로 지역, 가구주 연령과 같은 변수들이 응답 여부에 대해 영향력 있는 변수라 할 수 있다. 지역 규모를 고려해 볼 때 대도시에서는 거처 유형 및 1인 가구, 소도시에서는 거처 유형 및 1인 가구 또는 가구주 연령이 가구의 응답 여부에 영향을 준다고 볼 수 있다. 거처 유형별로 보면, 아파트 가구의 응답여부에 가장 영향을 주는 변수는 1인 가구 여부이고, 단독주택 가구의 응답 여부에 가장 영향을 주는 변수는 가구주 연령인 것으로 나타났다. 이러한 변수들은 무응답 가구의 성향과 밀접한 관련이 있는 변수들로서 이런 변수들이 무응답 가구들에 대해서 조사될 수 있다면, 무응답 편향을 보정하는데 중요한 역할을 하게 될 것이다.

제4절 결론 및 논의

지금까지 가구단위 표본조사에서 발생하는 무응답 가구의 특성에 대해 살펴보았다. 본 연구의 주요 목적은 무응답 가구의 특성을 파악하고 이 무응답 가구가 자료에 어떤 영향을 미치는가를 파악하는 것이었다. 특히 무응답 가구의 특성을 설명할 수 있는 변수들에 대해 첫째, 무응답 가구와 응답 가구의 분포가 어떻게 다른가? 둘째, 무응답 가구에 의해 발생할 수 있는 무응답 편향은 어느 정도이고 거부 편향과 부재 편향 중 어느 것이 더 크게 발생하고 있는가? 셋째, 가구들의 응답과 무응답 여부에 어떤 변수들이 영향을 주고 있는가? 를 중심으로 살펴보았다. 분석대상 가구는 2010년 지역별고용조사 가구를 대상으로 하였고, 무응답 가구의 특성을 설명할 수 있는 변수는 2010년 센서스 자료를 지역별고용조사의 가구관리 종합표 자료에 연계하여 사용하였다.

1. 연구결과 요약

무응답 가구의 특성을 분석하는 것은 상당히 어려운 작업이다. 왜냐하면 표본조사 대상가구에 대해 조사를 실시할 때, 우선 가구를 접촉하기가 어렵고 접촉하더라도 조사에 협조하지 않는 가구가 많을 뿐만 아니라, 이들 가구들에 대해서 어떠한 정보도 파악하기가 어렵기 때문이다. 본 연구에서는 파악하기 어려운 무응답 가구들에 대한 정보를 비슷한 시점에 실시된 총조사 자료를 통해 접근하고자 하였다. 이 자료를 바탕으로 무응답 가구에 대해서 다양하게 분석하였고, 이로부터 몇 가지 중요한 사실을 확인할 수 있었다.

① 무응답 가구는 응답 가구와 비교하였을 때 그 성향이 다르다. 무응답 가구 중에서 거



부 가구는 응답 가구와 그 특성이 유사하지만, 부재 가구는 그 성향이 응답 가구와 상당히 다르다고 볼 수 있다.

- 가구 또는 가구를 대표하는 가구주 특성 변수를 중심으로 분석한 결과, 가구가 1인 가구, 아파트 또는 연립주택 가구, 전세 가구, 주거 연면적이 작은 가구(24평 이하), 가구주 연령이 낮은 가구(30대 이하), 가구주 학력이 높은 가구(대학교 이상)에서 주로 조사에 대해 무응답하는 경향이 높다.
- 1인 가구, 주택 거주 가구, 가구주 연령이 30대인 가구, 주거 연면적이 작은 가구, 가구주가 미혼인 가구는 부재 성향이 강하고, 전세 가구, 주거 연면적이 넓은 가구, 가구주 학력이 높은 가구에서 거부하는 경향이 높다.
- 지역적으로는 1인 가구에서의 무응답은 대도시보다 소도시에서 더 많고, 아파트 거주 가구보다 주택 거주 가구에서 더 많이 발생한다. 이는 이 조사의 표본이 농·어촌지역 가구들을 많이 포함하고 있고, 우리나라의 특성상 농·어촌 지역의 거주자는 고령이면서 주택에 거주하는 경우가 많을 뿐만 아니라, 고령의 1인 가구가 상당히 많다는 것과 무관하지 않다.

② 늦게 응답하는 가구는 무응답 가구의 성향과 유사하다.

- 대도시의 아파트 가구는 총 방문 중 마지막 방문에서 응답한 가구와 거부 가구는 가구 특성 변수에 따라 통계적으로 유의한 차이가 없고, 주택 거주자는 가구 구분과 가구주 성별 변수에서 유의한 차이가 있었다. 이와 같은 현상은 소도시에서도 비슷하여 전체적으로 늦게 응답하는 가구는 최종적으로 거부 가구와 그 특성이 비슷한 것을 알 수 있다.
- 조사초기에 응답한 가구와 조사기간 마지막에 응답한 가구는 그 성향에 있어서 통계적으로 유의한 차이가 있다. 대도시의 아파트 가구의 경우 가구주 성별을 제외한 다른 변수들에 따라 3회째 방문 또는 4회째 방문에서의 응답 가구는 1회째 방문에서의 응답 가구와는 그 성향이 다른 것으로 나타났다.
- 방문회수가 증가할수록 전체적인 무응답 편향이 줄어든다는 점에서 늦게 응답하는 가구의 자료가 반드시 품질이 낮다고 보기는 어렵다. 거부 가구에 의한 편향이 방문회수 따라 증가하는 경향이 있지만, 방문회수가 제한적이기 때문에 일반화하는데는 한계가 있다.
- 다만 늦게 응답하는 가구에 대해서 그 가구들의 속성을 분석하고, 이들 가구들로부터 좋은 자료를 얻기 위해서 어떤 방법으로 접촉하고, 설득할 것인지에 대한 고민이 필요하다 하겠다.

③ 부재 가구에 의한 편향이 거부 가구에 의한 편향보다 총 무응답 편향에 더 많이 기여

한다. 무응답률은 부재율이 거부율보다 높다. 거부 편향이 더 작다는 측면에서 거부 가구는 응답 가구에 대해 그 성향이 닮았다고 볼 수 있다.

④ 더 많은 접촉은 그 정도는 다르지만 무응답 편향을 줄이는데 효과가 있다. 접촉회수의 증가는 부재 편향을 줄이는 효과가 있지만 거부 편향을 커지게 하기도 한다.

- 무응답에 대한 총 편향이 접촉회수 증가에 대해 크게 감소하지 않는 이유를 부재에 대한 관측 패턴에서 찾아볼 수 있다. [그림 5-14]와 <부표 III>에서 보면 1인 가구의 경우 접촉회수가 많을수록 부재율은 감소하지만 여전히 비접촉으로 남아있는 그룹은 접촉이 이루어진 그룹들과 그 특성값(아파트비율, 1인 가구 비율)이 더 다르다는 것을 알 수 있다. 이 부재율과 응답자와 부재자의 특성값 차가 곱해지면서 총 무응답 편향에 기여하기 때문에 부재율이 감소하더라도 두 그룹간의 차이가 크면 무응답 편향은 줄어들지 않게 된다.
- 접촉회수가 증가함에 따라 무응답 편향이 크게 감소하는 이유는 부재 비율도 감소하고 여전히 부재로 남아있는 그룹과 접촉이 된 가구들 간의 성향이 비슷하기 때문으로 볼 수 있다.
- 이 조사는 방문회수가 4번으로 제한되어 있지만, 더 많은 접촉을 허용한다면 무응답 편향을 더 줄일 수 있을 것이다. 방문회수의 증가는 예산 증가에도 영향을 미친다. 따라서 예산과 무응답 편향을 고려한 적정 방문회수를 정하는 것이 필요할 것이다. 총 무응답 편향의 감소 추세와 거부 편향의 증가 추세를 적절히 고려하면 무응답 편향을 고려한 적정 방문회수를 정할 수 있을 것이다.

⑤ 가구의 특성에 따라 조사에 대해 협조하는 경향이 다르다. 이는 조사대상 가구의 응답 여부에 영향을 미치는 변수를 통해 파악할 수 있다.

- 거주유형(주택, 아파트, 연립주택 등)과 가구구분(1인 가구 여부)변수가 가구의 응답 여부에 가장 영향력 있는 변수로 파악되었다. 그 다음으로는 가구주 연령, 성별, 학력과 같은 변수들이 영향력 있는 것으로 나타났다.
- 아파트 가구 또는 주택 가구 여부에 대해서는 응답 여부에 영향을 미치는 변수가 약간 다르다. 아파트 가구의 경우 1인 가구 여부를 나타내는 가구 구분, 주택의 경우는 가구주 연령이 응답 여부에 대해 가장 영향력 있는 변수로 파악되었다.
- 이들 변수들이 자료처리 과정에서 무응답 조정 변수로 사용되기 위해서는 이들 정보들을 우선적으로 파악할 수 있어야 한다. 1인 가구 여부가 무응답 편향을 조정하는데 좋은 정보라 하더라도 조사 현장에서 어떤 가구가 1인 가구인지를 파악하는 것은 쉽지 않고, 이를 위해서는 별도의 노력이 필요하게 될 것이다.



2. 무응답 가구 대응을 위한 방법

통계의 질, 무응답 증가, 조사비용 증가, 짧은 조사기간, 그리고 조사현장에서 직면하는 현실적인 문제들, 이러한 상황들은 서로 어떤 관계를 형성하고 있을까? 이는 조사기관의 가장 큰 고민거리가 아닌가 싶다. 이들 중 어느 하나 통계 품질에 긍정적으로 작용하는 것도 없는 것 같다. 본 연구 결과를 바탕으로, 중요한 것은 가구를 대상으로 하는 조사에서 무응답 가구로 인해 무응답 편향이 발생하고 있고 이 편향이 경우에 따라서는 통계 자료에 심각한 영향을 미칠 수 있다는 것이다. 뿐만 아니라 무응답 편향에 영향을 미치는 요인이 주로 높은 부채율과 부채 가구와 응답 가구 간의 성향이 서로 다르다는 데 있다는 것도 주목할 만하다. 거부 가구는 부채 가구에 비해 상대적으로 무응답 편향에 기여하는 정도가 작지만, 방문회수가 증가할수록 거부 편향이 증가하는 경향이 있다.

무응답 편향에 관한 연구는 수십 년전에도 있었고 지금도 조사통계에서 큰 관심을 받고 있는 분야 중 하나이다. 조사에 대한 협조도가 낮은 나라일수록 무응답 가구에 대한 관심이 높다. 이들 나라들은 단순 설득이나 응답률 향상뿐만 아니라, 통계 품질과 조사 비용을 동시에 고려하는 방법으로 무응답 가구에 대한 해결책을 찾는 것 같다. 여기서 조사 비용이 관심 사항이 되는 이유는, 통계작성에 들어가는 비용은 매년 엄청나게 증가하고 있는 반면, 통계 예산은 점점 삭감되는 추세에 있다는 데 있다. 이러한 상황에서라도 우리는 좋은 통계를 만들어야 할 책임이 있고, 우리나라의 조사환경 안에서 무응답 가구에 대한 대처 방법을 강구할 필요가 있다.

가. 실무적 대응

- ① 거부 가구와 부채 가구에 따른 맞춤형 전략이 필요하다. 부채 가구는 지속적인 접촉을 시도하고, 거부 가구는 조사원들의 전문적인 설득이 중요하다.
 - 연구결과에 따르면 무응답률 중 부채율이 높고, 부채 가구는 응답 가구에 대해 그 성향이 다른 경향이 있기 때문에 지속적인 접촉을 통해 부채 가구로부터 응답을 얻어 낼 필요가 있다. 지속적인 방문은 부채율을 감소시키고, 부채 편향을 줄일 수 있다는 점에서 총 무응답 편향을 줄이는데 상당히 기여할 수 있을 것이다.
 - 전체적으로 거부율은 낮고 거부 편향도 작다. 그렇지만 방문회수가 증가할수록 거부율과 거부 편향은 증가하는 경향이 있다. 따라서 편향을 고려할 때, 거부 가구에 대한 설득은 전략적으로 이루어져야 하며, 전문 조사원을 적정한 시점에 투입함으로써 응답의 질과 거부율을 줄일 수 있을 것이다.
 - 무응답 가구를 대응함에 있어서 우선 가구를 접촉할 수 있는 방법을 찾아야 하고,가

구를 접촉한 후에는 그 가구가 조사에 협조할 수 있도록 설득하는 기술이 필요하다.

- ② 조사기간 중에 부재 가구와 접촉하기 위해서는, ‘사전 알림장’과 같은 형태를 효과적으로 활용해야 한다. 지역의 특성(대도시, 중소도시, 농·어촌 등)에 따라 조사대상 가구의 성향을 고려하는 것도 중요하다.

- 우리나라의 가구들이 조사에 협조하는 정도는 지역 규모 또는 성격에 따라 다를 수 있다. 지방청(소)에서는 해당 지역내 가구들의 성향을 충분히 파악하여 응답 가구가 관심을 보일 수 있는 차별화된 방법을 찾아야 한다. 특히 주택 또는 아파트 가구에 따라 ‘사전알림’의 방법은 달라야 하며, 지방청간에 상호의 노하우를 공유하는 것도 중요하다 하겠다.

- ③ 1인 가구와 그렇지 않은 가구들에 대해 차별화된 전략으로 접근해야 한다.

- 본 연구의 경우, 1인 가구 중에는 소도시내 농·어촌 지역, 주택에 거주하는 60대 이상의 고령 가구가 많다([그림 5-18])은 점을 고려하여 이들 가구를 접근하는 것이 효과적이다. 대도시의 1인 가구는 젊은 층, 고학력, 전세 거주, 연립주택 및 아파트 거주가 상대적으로 많다는 점에서 소도시의 1인 가구의 성향과는 약간 다르다.
- 가구를 방문할 때, 가구주 학력, 가구주 연령, 점유형태 등을 고려하여 접근방법을 달리하기란 쉽지 않다. 최소한 조사원이 외관적으로 파악할 수 있는 정보는 지역내의 거처 유형 정도일 것이다. 정보의 개수도 중요하지만 정보의 질 또한 무시할 수 없다. 지역내 거처 유형은 무응답 편향을 줄이는데 필요한 핵심 정보이고, 이 정보를 모든 가구단위 조사에서 파악할 필요가 있다.

- ④ 접촉에 성공한 가구에 대해 응답을 얻어내는 것은 조사원 경험과 면접 스킬에 따라 다를 수 있다.

- 조사기간 중에 1번이라도 조사에 대해 거부한 경험이 있는 가구는 그 다음 접촉 시에 새로운 방법으로 면접을 시도해야 한다. 이를 위해서는 그 가구가 조사에 대해 거부하는 이유를 반드시 파악할 수 있어야 한다. 이는 다음 접촉에서 조사에 대한 성공여부를 결정할 수 있는 핵심요소가 될 수 있다.
- 가구에 대해 재접촉을 시도할 할 경우 슈퍼바이저의 판단이 매우 중요하다. 전문 조사원의 투입시기와 재접촉 시점은 응답 전환 성공과 밀접한 관련이 있다.
- 조사원들이 조사 가이드라인에 따를 수 있는 체제를 갖출 필요가 있다. 조사현장에서 발생하는 모든 정보들은 조사 오차를 최소화하는데 좋은 정보가 될 수 있다.





[그림 5-18] 도시규모에 따른 1인 가구내 조사항목별 분포

나. 통계적 대응 방법

- ① 무응답 가중치 조정 방법은 무응답 편향을 줄일 수 있는 좋은 대안이다. 조사기간 중 최대한 조사에 대해 협조를 구하고, 그 이후의 무응답 가구에 대해서는 조사에 협조한 가구들의 자료를 이용하여 통계적을 보정하는 과정을 거침으로써 무응답 편향을 줄일 수 있다.
 - 무응답 가중치 조정은 추가조사에 대한 노력을 하지 않기 때문에 조사원 부담이나 조사 비용측면에서 상당히 효율적인 방법이다. 이를 위해서 무응답 조정에 필요한 정보를 반드시 수집할 수 있어야 한다.
- ② 응답자 맞춤형 자료수집방법을 통해 응답률과 자료품질을 동시에 향상시킬 수 있다.
 - 1인 가구, 노인 가구, 맞벌이 가구가 증가한다는 것은 그만큼 조사대상이 다양해지고 있다는 것이다. 따라서 이들 각각이 선호하는 방법으로 자료를 수집할 수 있어야 한다.

- 자료수집방법에 따라 통계작성에 필요한 방법론 연구를 병행해야 한다. 혼합조사 방법에 의해 자료를 수집할 경우, 자료수집방법에 따른 효과를 고려하지 못할 경우 그로 인해 통계 편향이 발생할 수 있다. 하나의 조사에서 다양한 자료수집방법을 사용하고자 한다면 각 수집방법별 효과를 충분히 분석하고, 이를 자료처리과정에 반영함으로써 통계 오차를 줄일 수 있어야 한다.
- ③ 파라데이터(para-data)는 조사과정에서 발생하는 다양한 상황들을 파악함으로써 무응답 편향 축소는 물론이고 조사의 질적 향상을 가져올 수 있는 가장 좋은 수단이다.
- 파라데이터는 현장조사과정을 보여줄 수 있는 자료로서 조사과정에서 발생하는 통계오차를 보정하는데 필요한 정보를 제공할 수 있다. 이 자료는 조사표 항목의 난이도 및 이해도 증진을 위한 수정·보완, 항목 무응답 또는 단위 무응답 보정을 통한 무응답 편향 축소, 조사원 및 응답자와 관련된 현장의 실태 파악에 절실한 자료이다.

3. 향후연구

본 연구에서 사용된 자료는 가구에 대한 방문회수를 4회로 제한하고 있다. 따라서 무응답 편향과 방문회수간의 관계를 정확하게 분석하기는 어렵다. 가구에 대한 방문회수에 따라 조사기간과 조사비용이 영향을 받을 수 있다는 점에서 적정 방문회수를 정하는 것도 필요해 보인다. 거부 가구에 대한 대응 방법을 찾기 위해서는 그 가구가 조사에 거부하는 이유를 살펴볼 필요가 있다. 첫 대면에서 가구가 조사에 대해 거부하는 이유를 통해 거부 가구에 대한 일반화된 전환 전략을 세울 수 있을 것이다. 선행연구들은 접촉이 이루어진 가구에서 첫 대면은 매우 중요하며 그 가구의 거부시 반응을 파악하는 것이 필요하다고 주장하고 있다. 거부시 가구의 반응은 조사원이 재방문시 어떻게 그 가구를 대처할 것인가를 판단하는데 도움이 된다고 한다.

조사통계를 작성하는 과정에서 무응답은 점점 더 관심사항이 될 수밖에 없는 것 같다. 양적·질적으로 우수한 통계 자료에 대한 수요는 빠른 속도로 증가하는 반면, 통계조사에 대한 대상자들의 협조도는 빠른 속도로 감소하고 있다. 이런 점을 고려할 때 단위 무응답이 존재한다는 전제하에 좋은 통계를 작성할 수 방법을 모색해야 할 것이다. 따라서 단위 무응답에 관한 연구는 매우 폭넓게 꾸준히 진행되어야 할 분야라 할 수 있다.



참고문헌

- 김서영, 안다영 (2010). 가구면접조사에서 무응답률과 무응답 편향, 통계개발원 2010년 상반기 연구보고서 II, 111-164.
- 통계청 (2010). 2010년 지역별고용조사 조사지침서, 통계청.
- AAPOR. (2004). Standard definitions: Final dispositions of cas codes and outcome rates for surveys. AAPOR.
- Brick, J.M., J.Bruke and T.Le (2000). Analysis of Nonresponse Bias in the Base Year, Early Childhood Logitudinal Study: Kindergarten Class of 1998-99, unpublished report, Washington, DC: National Center For Education Statistics.
- D.Heerwegh, K.Abts and G.Loosveldt (2007). Minimizing survey refusal and noncontact rates: do our efforts pay off ?, Survey Research Methods, 1, 3-10.
- Groves, R.M., Presser, S. and Dipko, S. (2004). The role of topic interest in survey participation decisions. Public Opinion Quarterly, 68, 2-31.

〈부표 1〉 지역 내 응답유형별 가구 특성 분포

(단위 : %)

가구구분	전국												대도시				소도시																																																																																																																																																																																																																																																																																																		
	1인 가구			가족 등			아파트			주택			자가			전세 등			1명			2명			3명이상			없음			있음			18평이하			18~24평			24~34평			34평초과			남성			여성			20대이하			30대			40~50대			60대이상			중등이하			고등			대학교			유배우자			미혼 등																																																																																																																																																																																																																																							
	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부	전체	응답	부재	거부																																																																																																																																																																																																																																																																							
가구구분	23.77	23.30	36.04	19.27	19.62	18.79	32.98	16.47	25.35	24.90	37.99	21.62	76.23	76.70	63.96	80.73	83.53	74.65	75.10	62.01	78.38	39.02	36.36	62.92	72.66	70.32	50.65	48.56	59.28	70.32	34.59	32.02	65.24	74.61	60.98	63.64	37.08	27.34	29.68	65.41	67.98	34.76	25.39	70.32	71.60	54.36	59.93	55.82	56.12	49.75	58.34	75.84	77.11	57.29	61.27	29.68	28.40	45.64	40.07	44.18	43.88	50.25	41.66	24.16	22.89	42.71	38.73	23.77	23.30	36.04	19.27	19.62	18.79	32.98	16.47	25.35	24.90	37.99	21.62	30.80	31.38	24.93	24.18	23.43	23.62	21.97	22.46	33.60	34.15	26.82	25.61	45.44	45.32	39.02	56.55	56.95	57.59	45.05	61.06	41.05	40.95	35.19	52.77	78.04	78.13	80.93	72.25	74.63	74.25	80.14	73.93	79.34	79.51	81.44	70.85	21.96	21.87	19.07	27.75	25.37	25.75	19.86	26.07	20.66	20.49	18.56	29.15	12.54	12.34	17.14	11.58	17.32	17.52	19.49	12.12	10.72	10.49	15.65	11.12	33.17	32.64	39.10	38.36	32.47	32.60	31.62	31.59	33.43	32.65	43.85	44.04	35.11	35.36	30.97	34.61	35.20	34.98	35.02	38.39	35.08	35.49	28.39	31.43	19.17	19.67	12.79	15.46	15.01	14.89	13.86	17.90	20.76	21.36	12.11	13.41	74.88	75.06	70.50	76.23	75.97	76.43	69.99	76.17	74.47	74.57	70.82	76.27	25.12	24.94	29.50	23.77	24.03	23.57	30.01	23.83	25.53	25.43	29.18	23.73	4.24	3.91	8.54	6.64	5.82	5.59	9.03	5.51	3.63	3.31	8.22	7.59	14.04	13.37	21.98	20.05	18.48	18.10	23.27	18.52	12.36	11.69	21.16	21.33	42.91	42.02	49.58	55.83	49.40	48.91	49.44	56.16	40.43	39.56	49.67	55.56	38.81	40.70	19.91	17.47	26.30	27.40	18.25	19.81	43.58	45.43	20.96	15.52	42.20	44.05	24.45	20.45	24.61	25.62	17.95	17.84	48.91	50.61	28.59	22.65	30.04	29.58	34.74	35.20	34.60	34.87	33.17	32.40	28.30	27.69	35.75	37.54	27.44	26.10	39.89	43.70	40.24	39.01	47.71	49.08	22.56	21.50	34.92	39.19	69.23	69.65	59.46	71.66	70.08	70.86	58.79	71.75	68.90	69.21	59.89	71.59	30.77	30.35	40.54	28.34	29.92	29.14	41.21	28.25	31.10	30.79	40.11	28.41



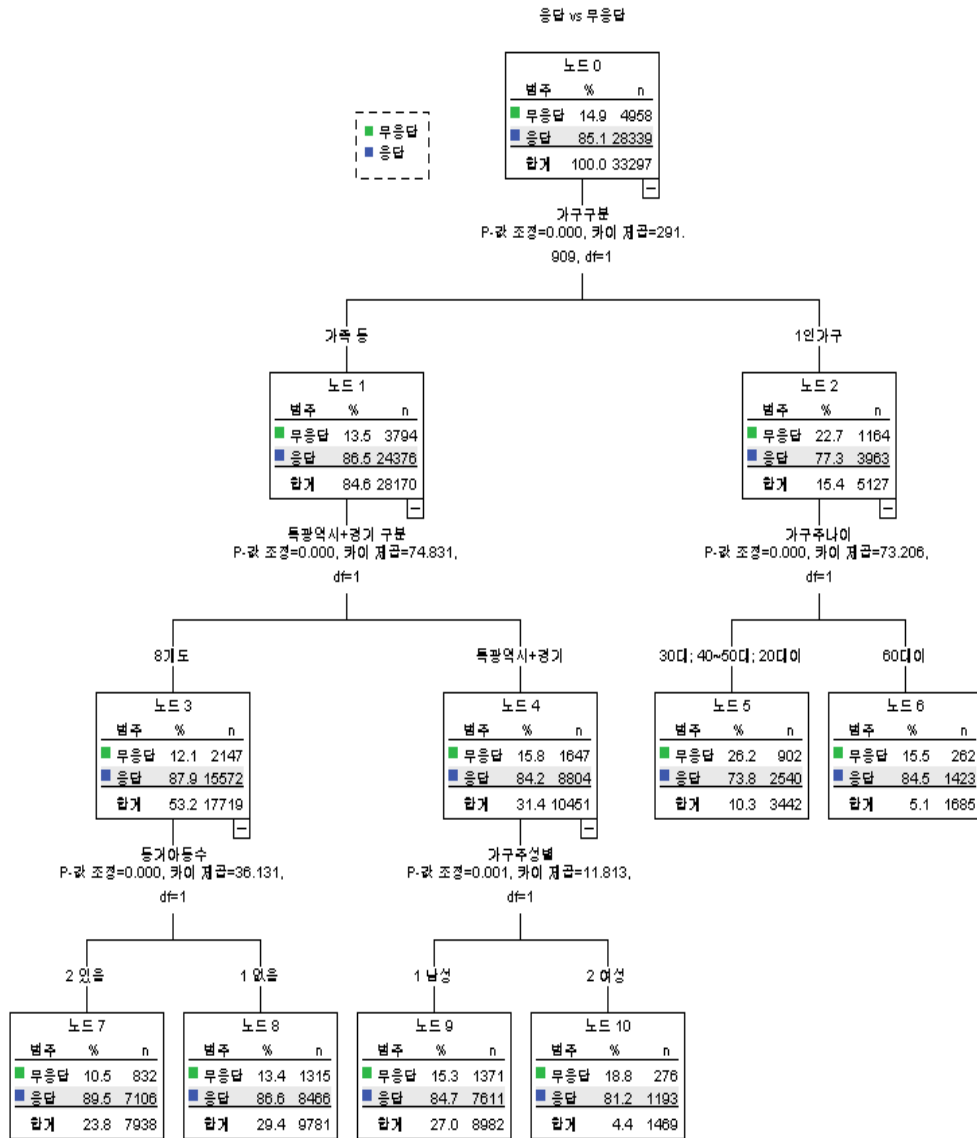
<부표 II> 가구 특성 비율에 대한 무응답 편향 절대값

전국	1인 가구	아파트	자가	동거아동수유	40~60(18~24평)	남성	40~50대	대학교	유배우자
부재편향	0.0062123	0.0129490	0.0084002	0.0013677	0.0031476	0.0022241	0.0036867	0.0067217	0.0049639
거부편향	0.0015193	0.0137069	0.0044053	0.0022179	0.0021610	0.0004400	0.0052183	0.0066461	0.0007621
편향합계	0.0046930	0.0266559	0.0128055	0.0008502	0.0053086	0.0017840	0.0089051	0.0133678	0.0042017
대도시	1인 가구	아파트	자가	동거아동수유	40~60(18~24평)	남성	40~50대	4 대학교	유배우자
부재편향	0.0097421	0.0073611	0.0043706	0.0040405	0.0006741	0.0044219	0.0003659	0.0059709	0.0082851
거부편향	0.0014451	0.0135787	0.0013847	0.0002008	0.0006347	0.0001594	0.0045242	0.0062829	0.0005567
편향합계	0.0082970	0.0209398	0.0029859	0.0038397	0.0013089	0.0045813	0.0048901	0.0122538	0.0077284
소도시	1인 가구	아파트	자가	동거아동수유	40~60(18~24평)	남성	40~50대	4 대학교	유배우자
부재편향	0.0053860	0.0136742	0.0081545	0.0007955	0.0046079	0.0015441	0.0041587	0.0055218	0.0038378
거부편향	0.0009310	0.0120868	0.0044943	0.0024562	0.0032312	0.0004812	0.0045397	0.0050181	0.0006746
편향합계	0.0044550	0.0257611	0.0126488	0.0016608	0.0078390	0.0010630	0.0086985	0.0105399	0.0031632

<부표 III> 가구 방문 회차별 가구 특성에 대한 무응답 편향 절대값

visit1	1인 가구	2 아파트	1 자가	동거아동수 유	2 40~60(18~24평)	1 남성	40~50대	4 대학교	1 유배우자
거부편향	0.002065677	0.012772634	0.004806197	0.003020333	0.002913346	0.001234469	0.005406972	0.006576171	0.001205664
부재편향	0.004348269	0.090580211	0.054170005	0.017445466	0.023260621	0.002514662	0.042906416	0.053039043	0.005448434
총 편향	0.002282592	0.103352845	0.058976202	0.020465799	0.026173967	0.00374913	0.048313388	0.059615214	0.004242769
visit2	1인 가구	2 아파트	1 자가	동거아동수 유	2 40~60(18~24평)	1 남성	40~50대	4 대학교	1 유배우자
거부편향	0.002104974	0.015565117	0.005584029	0.002537664	0.002537664	0.000994016	0.00370682	0.004162021	0.00100571
부재편향	0.016415877	0.057921095	0.044053354	0.002156115	0.011858634	0.003283175	0.022296191	0.019591129	0.013916744
총 편향	0.014310903	0.073486211	0.049637383	0.000381548	0.014396297	0.00228916	0.028003011	0.02375315	0.012911034
visit3	1인 가구	2 아파트	1 자가	동거아동수 유	2 40~60(18~24평)	1 남성	40~50대	4 대학교	1 유배우자
거부편향	0.004372093	0.026967865	0.005991755	0.004103805	0.004630444	0.001748837	0.009956448	0.003590764	0.002772304
부재편향	0.022015606	0.055811493	0.044262956	0.003630644	0.013016244	0.011200923	0.013363859	0.009636916	0.020393404
총 편향	0.017643513	0.082779358	0.050254711	0.000473162	0.017646688	0.009452085	0.023320308	0.01322768	0.017621099
visit4	1인 가구	2 아파트	1 자가	동거아동수 유	2 40~60(18~24평)	1 남성	40~50대	4 대학교	1 유배우자
거부편향	0.00669132	0.036704121	0.002551623	0.003675765	0.000820802	0.001088455	0.009135882	0.001334941	0.002373188
부재편향	0.033194567	0.040215946	0.015584036	0.01498465	0.00599386	0.013728793	0.002825677	0.002089694	0.024118151
총 편향	0.026503247	0.076920067	0.013032412	0.011308885	0.006814662	0.014817248	0.011961559	0.003424634	0.021744962

[부그림 1] 아파트 가구의 응답/무응답 의사결정나무



[부그림 II] 주택가구의 응답·무응답 의사결정나무

