

응답 거부와 부재율이 무응답 오차에 미치는 영향: 경제활동인구조사를 중심으로

김서영¹ · 권순필²

¹통계청 통계개발원, ²통계청 통계개발원

(2009년 1월 접수, 2009년 4월 채택)

요약

본 논문은 조사원이 기록한 가구의 현장방문내용을 분석함으로써 응답자의 부재(noncontact)와 거부(refusal)가 무응답 오차에 미치는 영향을 살피고자 한다. 통계 자료의 품질은 무응답률과 무응답 오차에 의해 결정되고, 응답률은 통계자료의 신뢰성 평가를 위한 척도로 사용되기도 한다. 가구를 대상으로 하는 조사에서 무응답 유형은 크게 부재와 거부의 형태로 나뉘고, 이들 무응답 유형이 무응답 오차에 미치는 영향은 다르다. 이에 본 논문은 무응답 유형에 따른 무응답 편향을 측정하고, 무응답 오차에 영향을 미치는 가구의 무응답 유형을 찾는데 그 목적이 있다.

주요용어: 응답 거부, 부재, 무응답률, 무응답 오차.

1. 서론

본 연구는 좋은 통계 자료를 생산하는 것과 관련하여 무응답률과 무응답 오차에 대한 탐색적 분석을 수행하는 것을 목표로 한다. 무응답률은 어떤 이유에서든(거부, 부재, 기타 등) 실제로 조사에 포함되지 않은 표본들의 비율을 나타낸다. 자료품질은 성공적으로 조사된 표본에 의한 모수 추정치와 모수값 간의 차이에 의해서 측정된다. 이러한 자료품질에 대한 정의는 상대적으로 엄격하고 통상의 조사 연구에서 받아들이기는 쉽지 않다. 왜냐하면 참값은 특별한 경우를 제외하고는 일반적으로 알려져 있지 않고 만약 알려져 있다면 조사를 할 필요가 없기 때문이다.

조사통계의 이론적 측면은 무응답이 없다는 것을 전제로 하고 있다. 그러나 선진국들의 경우 수십 년 전부터 가구 대상 조사에서의 무응답이 증가해 왔다(De Leeuw와 De Heer, 2001). 무응답의 증가로 많은 연구자들은 무응답을 줄이기 위한 방법을 찾는데 공통적인 관심을 갖게 되었다. 왜냐하면 무응답은 조사 추정치에 심각한 편향을 일으킬 수 있기 때문이다.

그러나 적정 응답률에 대한 유일한 정답은 없다. 다만, 연구자 또는 연구 분야 및 조사 특성에 따라 다양한 의견들이 존재할 뿐이다. Alreck과 Settle (1995)은 가능한 무응답을 줄이고 적정한 응답률을 유지하는 것이 중요하다고 하였다. Babbie (2007)는 사회과학 연구의 경우, 자료분석을 위한 응답률에 있어서, 적어도 50%의 응답률은 유지되어야 하고, 60%의 응답률은 좋은 편, 70%의 응답률은 매우 좋다는 결과를 제시하였다. Singleton과 Straits (2005)는 응답률에 관심을 갖는 것이 중요하며, 면접조사의 경우 적어도 85%의 응답률이 적정하고, 70% 이하는 심각한 편향을 가져올 수 있다고 하였다. 이처럼 각 연구자들이 제시하는 적정 응답률의 수준은 다르지만 무응답률을 줄여야 한다는 공통적인 메시지를 전

¹교신저자: (302-701) 대전광역시 서구 둔산동 신사로 139, 통계청 통계개발원, E-mail: sykim2217@nso.go.kr

하고 있다. 응답률 향상에 대한 이런 관심과 더불어 한편에서는 무응답률의 변화가 반드시 조사 추정치의 변화를 초래하는 것은 아니라는 연구들이 증가하고 있다 (Lynn 등, 2002; Stoop, 2005; Keeter 등, 2000; Curtin 등, 2000). 이들은 무응답 오차는 무응답률에만 의존하는 것은 아니며, 응답률이 높다고 해서 반드시 자료품질이 좋다고 확인할 수 없다는 의견들을 제시하고 있다.

그렇다면 무응답률과 무응답 오차는 어떤 관계가 있을까. Heerwegh 등 (2007)은 무응답 오차는 무응답률과 무관하지 않고, 게다가 무응답 유형이 무응답 오차에 기여하는 정도는 다르다고 하였다. 이들은 비율 측면에서는 응답 거부율이 부재율보다 높지만, 오차 측면에서는 부재로 인한 무응답 오차가 거부로 인한 오차보다 2.6배 이상 높다고 보고하고 있다. 이는 무응답 오차는 무응답률과 응답자와 무응답자간의 차와의 함수라는 점을 잘 설명하는 예이다. 즉, 무응답률만으로는 자료의 품질을 평가하기 어렵다는 것이다 (Curtin 등, 2000; Groves 등, 2004). 무응답률과 무응답 오차와의 관계는 조사의 성격, 응답자의 성향, 국민성 및 사회와 문화적 환경에 따라 다를 수 있다. 따라서 무응답 오차를 구성하는 요소와 그 크기를 측정하는 것은 통계 자료의 품질향상에 있어서 중요한 의미를 갖는다.

본 연구는 조사현장에서 발생하는 무응답 현황을 파악하고, 무응답 유형별 오차를 평가함으로써 무응답 오차에 기여도가 높은 무응답 유형을 찾고자 한다. 이를 위해 분석자료는 통계청의 경제활동인구조사를 사용하였다. 본 연구는 가구의 현장방문기록을 근거로 무응답 현황을 분석했다는 점에서 그 의미가 있다 하겠다. 본 연구의 구성은 다음과 같다. 2장에서는 분석에 사용된 자료의 특성과 자료 탐색 결과를 설명한다. 3장에서는 부재와 거부 등에 따른 무응답 유형별 오차분석 결과를 제시하고, 마지막으로 결론을 제시한다.

2. 탐색적 자료분석

2.1. 자료설명

본 연구는 통계청의 경제활동인구조사의 2007년 10월 신표본을 대상으로 하였다. 이 조사는 매월 약 32,000여 가구를 대상으로 우리나라 국민의 경제활동 현황을 파악하는 것을 목적으로 하고 있다. 이 조사는 매월 전체 표본의 약 1/36(900여 가구)을 새로운 표본으로 교체하는 연동표본을 사용하고 있다. 따라서 한번 표본으로 선정된 가구는 3년 동안 조사에 임하게 된다. 이 조사가 연속조사인만큼 조사원들에게 있어서 새로운 가구가 처음으로 표본이 되는 시점에서 조사에 비협조적인 표본 가구를 조사에 참여시키는 것이 중요하다. 조사원들의 방문에 대한 가구의 협조 상황을 살펴보기 위해, 본 연구는 조사원들로 하여금 해당 월에 처음으로 표본이 되는 신표본 가구의 방문현황을 기록하게 하였다. 그리고 조사에 대해 늦게 응답한 가구의 조사결과가 전체 추정치에 미치는 효과는 가구의 방문기록자료와 경제활동인구조사 자료를 연계하여 분석하였다. 무응답 유형별 오차계산은 경제활동인구조사의 표본명부자료를 사용하였다 (통계청, 2007).

조사원들이 작성한 가구의 현장방문기록의 구체적 내용은 다음과 같다. 조사항목은 최초 대면시 응답 거부 여부, 방문일시(요일), 총 방문횟수(그 중 부재수), 최종응답여부 등을 사용하였다. 방문일시(시간, 요일)는 최초와 마지막 방문에 대해서만 작성토록 하여 조사원들의 응답부담을 적게 하였다. 방문대상가구는 전국 47개 지방청(출장소)의 2007년 10월 신표본 가구인 970개로 하였다. 조사기간은 신표본 가구를 처음 방문한 첫날부터 경찰조사 본조사가 시작되기 바로 직전 주말까지로 정했다.

표본명부 자료는 무응답 오차 계산에 필요한 핵심 변수 선정에 사용되었다. 표본명부는 경제활동인구조사 표본추출을 위해 사용된 특성지표를 포함하고 있으며, 특히 목표 추정치(실업자수, 취업자수) 등에 영향을 주는 변수들을 포함하고 있다. 무응답 오차 측정은 주택유형(단독주택, 아파트, 연립주택, 다세대 주택, 기타 등), 주택소유형태, 가구주산업, 가구주직업 등의 변수들을 사용할 수 있다. 본 연구

표 2.1. 조사원의 방문기록에 의한 가구의 응답 유형별 분포

무응답	조사대상 전체표본		이용가능 표본	
	n	%	n	%
응답	872	89.25	872	91.69
부재	25	2.56	25	2.63
거부	54	5.53	54	5.68
기타	4	0.41	-	-
이용불가	22	2.25	-	-
총합	977	100	951	100

표 2.2. 2007년 10월 실태조사 가구의 응답현황

		n	%
응답자	초기 응답자	588	61.8
	응답전환자	284	29.9
	부분 합	872	91.7
무응답자	부재로 인한 무응답	25	2.6
	거부로 인한 무응답	54	5.7
	부분 합	79	8.3
	총합	951	100

는 주택유형과 주택소유형태를 핵심변수로 사용하였다. 그 이유로서 하나는 주택유형과 주택소유형태는 경제활동인구조사의 표본 설계시에 사용되는 주요 변수라는 것과 다른 하나는 이 두 변수는 가구 방문과 같은 면접조사의 경우 응답유형에 영향을 미치는 변수로 작용한다는 점을 들 수 있다.

표 2.1은 조사원이 방문한 총 977가구들의 응답현황을 나타낸 것이다. 977가구 중 기타(빈집, 4가구)는 응답이 되지 않았고, 22가구는 표본명부에 주어진 코드와 연계되지 않아 분석에 이용되지 않았다.

2.2. 방문가구의 응답현황 분석

2.2.1. 가구의 유형별 응답분포 표 2.2는 조사 대상 가구들의 응답과 무응답 현황을 나타낸다. 표 2.2에서 ‘초기 응답자’는 조사원이 가구를 방문하여 처음으로 가구의 누군가를 대면하였을 때 거부하지 않고 조사에 응답한 자, ‘응답 전환자’는 대면에서 최소한 한 번 이상 거부를 한 후에 재방문을 통해 조사에 응답한 자로 정의한다. 그리고 이 두 그룹을 총칭하여 응답자라 한다. 무응답자는 부재, 거부, 기타(질병, 장기출타 등)의 이유로 최종적으로 조사에 응답하지 않은 자를 말한다. 여기서는 무응답 유형으로 부재와 거부만을 고려하였다. 기타는 해당 가구가 극소수(4가구)이기 때문에 무응답 유형에서 제외되었다.

응답률 계산 방법은 다양하며, 조사목적과 자료의 특성에 따라 다르게 적용될 수 있다. 본 연구는 일반적으로 널리 사용되는 방법인 조사에 참여한 유효 표본 가운데 조사에 응답한 가구의 비율로 정의하였다. 응답률은 전체 951개의 유효 표본 가구의 91.7%로 실태조사의 경우 경제활동인구조사에 참여하겠다는 비율은 대체로 높게 나타났다(표 2.2). 이들 응답자 중 한 번도 거부하지 않고 조사에 협조한 ‘초기 응답자’ 비율은 61.8%로 최종 응답률 91.7%에 비하면 그다지 높은 편은 아니다. 응답자의 29.9%는 초기 대면에서 응답에 거부한 후 재방문에서 응답으로 전환되었고, 이는 조사원들의 노력의 대가라 할 수 있다. 응답전환 비율은 초기 대면에서 응답하지 않은 363(= 응답전환자 + 부재 + 거부)가구 중 약 78.2%(284가구)로 전환율이 상당히 높은 편이다. 한편, 무응답률은 전체 951가구 중 79가구

표 2.3. 평균 방문횟수, 부재횟수, 설득기간: 평균(표준편차)

	전체	응답자		무응답자	
		초기응답자	응답전환자	부재자	거부자
방문횟수	3.09(1.68)	2.44(1.47)	3.96(1.37)	4.68(1.41)	4.92(1.54)
부재횟수	1.63(1.52)	1.36(1.42)	1.73(1.24)	4.56(1.47)	2.83(1.78)
설득기간(일)	8.23(12.07)	6.36(11.18)	10.11(12.02)	14.60(12.26)	15.68(15.80)
총합(명)	951	588	284	25	54

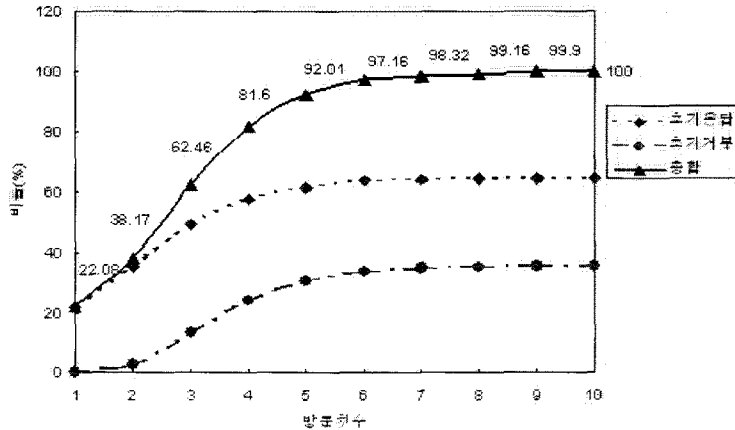


그림 2.1. 방문횟수에 따른 누적 응답률

로 약 8.3%를 차지하였다. 이중 거부율은 5.7%로 전체 무응답률의 68.7%, 부재율은 2.6%로 전체 무응답률의 31.3%를 각각 차지하였다. 예상대로 전체 무응답 중 거부로 인한 무응답이 부재보다 높게 나타났다.

2.2.2. 응답률 향상을 위한 노력 표 2.3은 조사원들의 가구 설득을 위한 노력을 양적으로 파악한 것이다. 이는 조사원들의 가구에 대한 평균 방문횟수, 부재횟수, 최종 응답여부가 결정되기까지 걸리는 기간을 나타낸다. 이 표에 의하면 조사원들의 전체 가구당 평균 방문횟수는 3회(3.09)로 그 중 부재횟수는 평균 2회(1.63)인 것으로 나타났다. 초기 응답자의 경우 방문회수는 약 2회이고 이중 부재회수는 약 1회로 방문 가구가 조사에 협조적임을 알 수 있다. 무응답자의 경우는 부재와 거부자 모두에서 방문회수가 약 5회로 전체 평균 3회보다 2회 정도 많다. 이 중에서 부재의 경우는 5회(4.68회)를 방문하고도 결국 조사를 완성하지 못하는 실정이다. 거부자의 경우는 약 5회 방문(4.92)중 3회(2.83)는 가구의 부재, 2회는 거부인 것으로 나타났다. 또한 가구의 조사 협조를 위한 설득 기간은 평균 8일로, 응답자(초기응답 = 약 6일, 응답전환 = 약 10일)보다는 무응답자(부재자 & 거부자 = 약 15일)를 설득하는데 더 많은 기간이 소요되고 있음을 알 수 있다. 설득기간에 대한 차이는 지방청(출장소)마다 또는 조사원마다 조사 착수 일정이 다르기 때문에 발생하는 부분도 무시할 수는 없다.

그림 2.1은 방문횟수에 따른 응답자와 거부자의 누적비율을 나타낸다. 그림 2.1에서 보는 바와 같이 4번째 방문까지는 방문가구의 응답 비율이 급격하게 증가하는 반면에 5번째부터는 그 증가폭이 완만한 것을 알 수 있다. 즉, 4번째 방문까지로 조사를 완성하는 가구의 비율은 약 82%정도이다. 또한 전체 약 62.5%가 평균 3번 이하의 방문으로 응답여부가 결정되는 것을 알 수 있다. 이 결과에 비추어 표 2.2에

표 2.4. 최초 응답자와 응답 전환자 실업률 비교

	최초 응답자	응답 전환자	전체 응답자
실업률 추정치(%)	3.21	4.75	3.60
취업자 수(명)	755	341	1,096
실업자 수(명)	25	17	42
가구 수(가구)	576	266	842

서 설명한 총 응답률 91.7%는 최대 5회 방문으로 달성할 수 있음을 알 수 있다. 즉, 그 이상의 재방문 시도는 응답률 향상에 크게 도움이 되지 않을 수도 있음을 짐작케 한다. 자료수집상의 여러 가지 제약 사항을 고려하면 본 연구 결과를 모든 통계조사에 대해 일반화시키는 것은 어렵다. 다만, 응답률 향상을 위한 조사원들의 최적 방문 횟수는 어떤 기준에서든 결정될 필요가 있다.

2.2.3. 응답전환자의 응답특성 표 2.4는 초기응답자와 응답전환자의 조사 추정치를 나타낸 것으로, 이를 통해 응답전환자들의 응답특성을 살펴보고자 한다. 참고로 표 2.4는 2007년 10월 경제활동인구조사의 신표본만의 조사 결과를 분석한 것으로, 설계 가중치 등을 적용하지 않은 실업률임을 밝혀둔다. 표에서 응답 전환된 가구들의 실업률 추정치는 4.75%로 전체 실업률 3.60%보다 상당히 높다. 또한 응답 전환된 가구들은 초기 응답자들의 실업률 3.21%와도 큰 차이를 보이고 있다. 이는 실업자가 많은 가구일수록 조사에 대해 부담을 갖거나 조사를 거부할 가능성이 높다는 것을 짐작케 한다. 뿐만 아니라 이는 응답을 거부하는 가구일수록 전체 응답자의 성향과 다를 수 있음을 시사한다.

3. 무응답 유형별 오차

지금까지는 가구의 무응답 유형별 응답률과 유형별 가구의 응답 성향을 살펴보았다. 본 절에서는 이를 기본으로 무응답 오차가 무응답 유형별로 어떻게 구성되어 있고, 유형별 무응답자의 성향은 응답자와 어떻게 다른지에 대해서 구체적으로 살펴보고자 한다.

3.1. 무응답 오차의 개념

무응답이 실제로 조사결과에 어떤 영향을 미칠 것인가에 대해서 명확하게 증명하기는 어렵다. 그렇지만 무응답은 조사단계에서의 노력으로 충분히 줄일 수 있고 무응답 오차가 발생하는 원인이 반드시 존재한다는 것만은 확실하다. 모집단의 부차그룹(subgroup)별로 응답률을 조사하는 방법은 무응답 편향의 정도를 결정하는데 직접적인 도움이 되는 것은 아니다. 그러나 이는 무응답 편향에 대한 존재 여부는 결정할 수 있다. 만약 무응답률이 부차그룹에 대해 매우 유사하다면 무응답 편향이 존재하더라도 이 무응답은 조사 자료에는 심각한 영향을 주지 않을 가능성이 높다.

최근 연구들은 무응답 편향이 무응답 유형에 따라 그 정도가 다르다고 보고하고 있다 (Groves, 2006; Heerwegh 등, 2007). 이들에 따르면 무응답 유형은 가구의 어떤 특성에 의해 기인된 것으로, 이들 원인에 따른 무응답 편향을 측정하는 것은 자료의 품질 향상을 위해 중요하다. 따라서 무응답 원인별로 무응답 오차를 계산하고, 무응답률과 오차와의 관계를 파악할 필요가 있다.

오차계산을 위한 핵심변수 선정은 어려운 문제이며, 일반적으로는 핵심 추정치와 관련된 변수를 선정한다 (조사통계연구회, 2000; Heerwegh 등, 2007). 본 연구는 핵심변수로 주택유형과 주택소유형태를 사용하였다. 주택유형에서는 주택, 아파트, 기타(연립, 다세대 등)별로 차지하는 가구의 비율, 주택소유형태에서는 자가와 전월세가 차지하는 가구의 비율을 고려하였다. 가구를 대상으로 하는 조사에서 주택유

형은 부재율과 거부율에 영향을 줄 수 있기 때문에 보다 넓은 의미에서 본다면 우리의 관심 모수가 될 수 있다 (Groves와 Couper, 1998). 따라서 이 핵심변수의 추정치에 대한 오차는 우리의 관심사항이며, 이 오차는 전체적인 자료 품질에 대해 설명할 수 있다. 뿐만 아니라 이를 통해 자료를 수집하는데 들어가는 노력이 자료 품질과 어느 정도 연관성이 있는지에 대한 비전을 제시할 수 있다.

이제 무응답 오차를 정의해 보자. 표본설계에 의해 추출된 각 표본에 대해서 전체 표본의 평균, \bar{y}_n 과 그 중에서 응답자의 평균, \bar{y}_r 과 무응답자의 평균, \bar{y}_m 의 관계는 다음과 같다 (Groves와 Couper, 1998).

$$\bar{y}_n = \left(\frac{r}{n}\right)\bar{y}_r + \left(\frac{m}{n}\right)\bar{y}_m.$$

이것을 응답자 평균에 대해서 다시 정리하면 다음과 같다.

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

여기서 m 은 표본에서 무응답자 수, r 은 표본에서의 응답자 수, $n(r+m)$ 은 전체 표본의 수를 나타낸다. 결국, 전체 표본평균에 대한 응답자 평균의 편차는 무응답률 m/n 과 응답자와 무응답자간의 평균차($\bar{y}_r - \bar{y}_m$)와의 함수로 나타난다. 만약 전체 모집단으로부터 단순 임의추출에 의해 표본을 추출한다면, 이 편차는 표본 내의 응답자에 대해서만 관측할 수 있게 된다. 이때 구한 응답자 평균 \bar{y}_r 과 이 응답자 평균을 사용하여 전체 모집단 평균을 추정한다면, 이는 응답자 평균에 대한 편향으로 다음의 근사식으로 표현할 수 있다 (김재광, 2008).

$$\text{Bias}(\bar{y}_r) = \left(\frac{M}{N}\right)(\bar{Y}_r - \bar{Y}_m), \quad \text{Var}(\bar{y}_r) = \frac{1}{r}S_r^2,$$

여기서 Var 는 조사결과에 대한 응답자 평균의 분산, S_r^2 은 모집단에서의 응답자들의 분산, M 은 모집단에서의 무응답자 수, N 은 모집단 전체 수를 나타낸다. 이러한 근사식을 사용할 경우 편향된 추정을 하게 되고, 표본 수($r < n$) 감소로 인해 추정의 효율이 떨어진다는 문제가 발생하게 된다. 이렇게 발생하는 편향을 보정하고 효율을 높이기 위해서는 적절한 무응답 처리 방안을 찾는 것이 필요하다. 뿐만 아니라 무응답 오차에 영향을 미치는 요인은 무응답 유형에 따라 다를 수 있고, 이에 따라 대처방안도 달라져야 한다는 점에 주의해야 한다. 이처럼 세부적인 무응답 오차 평가는 추정치의 편향을 줄이고 효율을 높이는 데 중요하다.

3.2. 무응답 오차의 구성

표 3.1은 주택유형별 표본가구들의 분포를 나타낸다. 조사 대상가구 중 951개의 유효 표본 중에서 39.85%는 주택(단독주택), 44.79%는 아파트, 나머지 15.35%는 기타유형(상가주택, 다세대, 연립주택, 주거지 외 거처)으로 나타났다. 응답한 표본 가구에 대해서만 주택유형을 살펴보면, 응답 가구 872가구 중 주택에 거주하는 비율은 40.71%(355가구), 아파트에 거주하는 비율은 44.27%(386가구)이고, 나머지 기타유형에 15.23%(131가구)가 거주하는 것으로 나타났다. 표 3.1에 근거한 응답자의 응답유형별(응답/부재/거부) 주택유형(주택/아파트/기타)의 분포는 유의수준 0.05에서 통계적으로 큰 차이를 보이지 않았다($\chi^2 = 5.1, df = 4, p = 0.2733$). 이로부터 표본설계에 기인한 주택유형에 대한 편향은 거의 없다고 볼 수 있다. 한편, 전체 표본 구성에 있어서 부재와 거부가 차지하는 비중은 전체의 약 8.3%(79/951)로 응답이 차지하는 비중에 비해 적은 편이다. 뿐만 아니라, 응답 유형별 오차 분석과 관련하여 응답유형을 응답과 무응답(부재/거절)의 두 개 그룹으로 구분하여 분석할 필요가 있다. 이에 응답유형을 응답과 무응답으로 나누어 χ^2 검정한 결과, 응답유형(응답/무응답)에 따른 주택유형의 분포는 유의수준 0.1에서 통계적으로 다소 유의한 것으로 나타났다($\chi^2 = 4.6, df = 2, p = 0.0980$).

표 3.1. 무응답 원인과 주택유형별 가구 분포

응답 유형	주택		아파트		기타		전체	
	n	%	n	%	n	%	n	%
응답	355	40.71	386	44.27	131	15.02	872	100
부재	10	40.00	10	40.00	5	20.00	25	100
거부	14	25.92	30	55.56	10	18.52	54	100
	379	39.85	426	44.79	146	15.35	951	100

이제 특정 추정치에 대해서 무응답 유형을 고려한 무응답 오차를 구체적으로 계산해 보자. 만약 주택에 살고 있는 가구의 비율이 목표 추정치일 경우, 이 추정치에 대한 무응답 오차는 응답자 평균과 전체 표본 평균과의 차에 의해 구할 수 있다. 즉, 무응답 오차는 $\bar{y}_r - \bar{y}_n = 0.4071 - 0.3985 = 0.0086$ 과 같다. 여기서 \bar{y}_r 는 응답자 중 주택 가구에 거주하는 비율, \bar{y}_n 는 전체 대상 표본 가구 중 주택에 거주하는 가구의 비율(모수에 해당)을 나타낸다. 계산을 위해 식 (3.1)을 다시 쓰면 다음과 같다.

$$\bar{y}_r = \bar{y}_n + (\text{무응답 오차}).$$

즉, 응답자 평균은 전체 표본평균과 무응답 오차의 합으로 구성된다. 식 (3.1)에 무응답 오차 0.0086, $m = 79$, $n = 951$ 을 대입하여 계산하면 무응답자의 평균은 $\bar{y}_m = 0.3036$ 이 된다. 이는 무응답 가구의 30.36%가 주택에 거주하고 있다는 것을 의미한다. 다시 말하면, 응답하지 않은 79(부재 + 거부 = 25 + 54)가구 중 약 30.36%에 해당하는 24가구가 주택에 살고 있다는 것이다. 이 결과는 응답가구의 40.71%가 주택에 거주한다는 것과는 약간 차이를 보인다. 물론 χ^2 검정 결과 전체 표본에 대해서 응답유형에 따른 주택유형의 분포는 통계적으로 명백한 차이는 없지만, 특정 그룹에 해당하는 주택 거주 비율을 추정하는데 있어서 응답가구들과 무응답 가구들 간의 주택 거주 비율에는 약간 차이가 있다고 볼 수 있다.

또한 이를 확장하여 무응답 그룹이 서로 다른 세부그룹으로 분할될 때 응답과 무응답 그룹 간에 발생하는 추가적인 차이를 볼 수 있다 (Mayer와 Pratt, 1966). 무응답 유형별로 오차 특성을 살펴보기 위해 식 (3.1)을 다음과 같이 확장할 수 있다 (Groves와 Couper, 1998).

$$\bar{y}_r = \bar{y}_n + \left(\frac{m_{ref}}{n}\right) (\bar{y}_r - \bar{y}_{m_{ref}}) + \left(\frac{m_{nc}}{n}\right) (\bar{y}_r - \bar{y}_{m_{nc}}), \tag{3.2}$$

여기서, m_{ref} , m_{nc} 는 각각 거부와 부재에 의해 발생한 무응답자 수를 나타낸다. 식 (3.2)에 의한 구체적인 계산은 다음과 같다.

$$0.4071 = 0.3985 + \left(\frac{54}{951}\right) (0.4071 - 0.2593) + \left(\frac{25}{951}\right) (0.4071 - 0.4000).$$

거부에 의한 오차는 0.0084(97.67%)로 무응답 오차의 대부분을 차지하고, 부재에 의한 오차는 0.0002(2.33%)로 전체 오차의 아주 적은 부분만을 차지한다는 것을 알 수 있다. 추정 대상이 주택이 아니라 아파트나 기타 유형인 경우에 대해서도 유사한 방법으로 계산할 수 있다. 단독주택에서와 마찬가지로 아파트 비율과 기타 비율을 추정할 경우 거부에 의한 오차가 각각 85.3%와 60.6%로 전체 무응답 오차의 대부분을 차지한다.

본 연구에서 무응답 오차는 크게 거부와 부재에 의한 오차로 나뉘고, 이 중 거부로 인한 오차가 대부분을 차지하였다. 결과적으로 무응답 오차를 구성하는 무응답율은 거부율이 5.7%로 부재율 2.6%보다 높고(표 2.2), 무응답 오차 또한 거부가 부재로 인한 무응답 오차보다 더 크게 나타났다. 따라서 무응답율

표 3.2. 무응답 원인과 주택소유 여부별 가구 분포

응답 유형	소유		무소유		전체	
	n	%	n	%	n	%
응답	595	68.23	277	31.77	872	100
부재	13	52.00	12	48.00	25	100
거부	27	50.00	27	50.00	54	100
	635	66.77	316	33.23	951	100

이나 오차 크기 측면에서 거부로 인한 무응답이 추정치의 편향을 가져올 수 있는 가능성이 상당히 잠재되어 있다고 볼 수 있다. 이처럼 거부로 인한 무응답 층이 전체 무응답 구성의 위협요인으로 다가오는 것은 주택유형 추정에 있어서는 부재자가 거부자보다 응답자의 성향을 더 닮았다고 볼 수 있기 때문이다.

추가적으로 다른 변수에서도 동일한 현상이 발생하는지를 살펴보았다. 주택소유형태(자가/전월세)변수에 대해서도 주택유형과 동일한 방법으로 무응답 유형에 따른 오차를 계산하였다. 표 3.2는 주택소유형태에 대한 가구 분포를 나타낸다. 표 3.2와 같이 전체적으로 주택 소유자의 응답비율이 93.7%로 무소유자의 응답비율 87.66% 보다 높다. 무소유자의 경우는 부재보다 거부에 의한 무응답이 상대적으로 높게 나타났다. 전체 표본 중 주택소유자는 66.77%이고 응답자 중 주택소유자는 68.23%로 조사되었다. 주택 소유자 비율이 목표 추정값일 경우, 표본에 대한 응답자 평균과 전체 평균은 각각 $\bar{y}_r = 0.6823$, $\bar{y}_n = 0.6677$ 과 같다. 식 (3.2)에 의해 무응답 오차를 계산할 때, 거부에 의한 무응답 오차는 0.0104(70.75%), 부재에 의한 오차는 0.0043(29.25%)으로, 주택유형에서와 마찬가지로 거부에 의한 오차가 전체 오차의 대부분을 차지하였다. 주택소유의 경우, 무응답 오차는 표본을 설계할 때 주택소유자에 더 많은 비중의 표본(전체의 66.77%, 635가구)이 배정된 영향도 무시할 수 없다.

4. 결론 및 논의

본 연구는 경제활동인구조사의 표본 가구를 대상으로 통계조사에서 발생하는 무응답 현황과 특성을 분석하고, 이를 바탕으로 무응답 유형에 따른 무응답 오차를 평가하였다. 표본 가구의 경우, 최초 대면에서 거부하지 않고 응답에 협조한 가구가 60% 이상이고, 조사원의 재방문을 통해 응답을 전환한 경우가 약 30% 이상인 것으로 나타났다.

무응답 유형별 무응답률은 거부율이 부재율보다 높았으며, 무응답 오차는 부재보다 거부에 의한 오차가 더 큰 비중을 차지하였다. 이를 통해 부재자는 응답자와 응답 성향이 더 닮았으며, 부재자에서 응답을 얻지 못하더라도 이는 조사 추정치에 큰 영향을 주지 않을 것으로 짐작할 수 있다. 그러나 무응답 오차의 대부분을 차지하는 거부자는 응답 성향이 응답자와 다르기 때문에 거부율이 줄어든다 하더라도 추정치에는 여전히 심각한 편향이 발생할 수 있다. 따라서 경제활동인구조사 추정치의 신뢰성 확보를 위해서는 거부 가구들에 대한 심층 분석을 통해 신중하게 대처할 필요가 있다.

한편 조사원들의 조사부담은 조사결과에 좋지 않은 영향을 미칠 수 있다. 자료분석 결과에 의하면 조사원들은 부재 또는 거부 가구에 대한 재방문을 통해 응답률 향상을 꾀하고 있다. 이때 재방문 일정은 조사원이 스스로 결정하는 경향이 있으며, 당초 계획과는 달리 야간이나 주말 시간을 이용하여 가구를 방문하는 경우도 많다. 그런데 실제로 조사에 비협조적인 거부 가구들은 여러 번의 재방문에도 결국은 조사에 협조하지 않거나 잦은 방문이 귀찮아서 성의 없는 응답을 하는 경우도 종종 있다. 이처럼 조사원이 가구의 특성을 제대로 파악하지 않고 재방문 일정을 변경하여 조사하는 것은 노력만큼의 소득은 없을 수 있다. 오히려 재방문을 통한 응답률 향상은 조사 관리자의 철저한 표본 가구 모니터링을 통해 체계적으

로 이루어져야 할 문제라 할 수 있다.

본 연구는 무응답 오차는 각 유형별 무응답률 뿐만 아니라 그에 따른 오차의 크기를 동시에 고려해야 한다는 것을 강조하고 있다. 또한 무응답 오차에 대한 평가 없이 거부율과 부재율에 의해 거부 또는 부재 중 어느 한쪽의 오차가 더 클 것이라는 결론을 내리는 잘못을 범할 수 있음을 지적하고 있다. 무응답 유형별 가구의 특성은 조사의 성격, 또는 사회/문화적 배경에 따라 다를 수 있지만, 이는 반드시 자료품질 향상과 관련하여 분석되어야 할 부분이다. 본 연구는 방문기록내용, 조사기간 및 자료의 양적 측면에서 다소 부족한 점이 있기 때문에, 이 결과를 일반화시키기에는 한계가 있을 수 있다. 그렇지만 조사원들의 가구방문기록을 근거로 무응답 실태를 분석하였다는 점에서 큰 의미가 있다 하겠다. 본 연구를 통해서 무응답 유형과 조사 자료와의 관계를 충분히 이해하고, 향후 이 연구가 무응답 유형별 오차 연구의 근간이 될 수 있을 것으로 기대한다.

참고문헌

- 김재광 (2008). <표본조사론>, 자유아카데미.
- 조사통계연구회 (2000). <무응답 오차>, 자유아카데미.
- 통계청 (2007). <가구부문 표본개원 결과 2007>, p.6.
- Alreck, L. P. and Settle, R. (1995). *The Survey Research Handbook*, Wiley John & Sons, New York.
- Babbie, E. (2007). *The Practice of Social Research*, 11th ed., Wadsworth Publishing, Belmont.
- Curtin, R., Presser, S. and Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment, *Public Opinion Quarterly*, **64**, 413-428.
- De Leeuw, E. and De Heer, W. (2001). *Trends in Households Survey Non-Response: A Longitudinal and International Comparison*, In *Survey Non-Response*, Wiley John & Sons, New York, 41-54.
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys, *Public Opinion Quarterly*, **70**, 646-675.
- Groves, R. M. and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*, Wiley John & Sons, New York.
- Groves, R. M., Presser, S. and Dipko, S. (2004). The role of topic interest in survey participation decisions, *Public Opinion Quarterly*, **68**, 2-31.
- Heerwegh, D., Abts, K. and Loosveldt, G. (2007). Minimizing survey refusal and noncontact rates: Do our efforts pay off?, *Survey Research Methods*, **1**, 3-10.
- Keeter, S., Miller, C., Kohut, A., Groves, R. M. and Presser, S. (2000). Consequences of reducing nonresponse in a telephone survey, *Public Opinion Quarterly*, **64**, 125-148.
- Lynn, P., Clarke, P., Martin, P. and Sturgis, P. (2002). *The effects of extended interviewer efforts on non-response bias*, In R. Groves, D. Dillman, J. Eltinge and R. Little (Eds.), *survey nonresponse*. 135-147, Wiley, New York.
- Mayer, C. S. and Pratt, Jr. R. W. (1996). A note on nonresponse in a mail survey, *Public Opinion Quarterly*, **30**, 637-646.
- Singleton, R. and Straits, B. (2005). *Approaches to Social Research*, 4th ed., Oxford University Press, New York.
- Stoop, I. (2005). *Nonresponse in Sample Survey. The Hunt for the Last Respondent*, The Hague: Social and Cultural Planning Office.

The Effect of Survey Refusal and Noncontact on Nonresponse Error: For Economically Active Population Survey

Seo Young Kim¹ · Soon Pil Kwon²

¹Statistics Research Institute; ²Statistics Research Institute

(Received January 2009; accepted April 2009)

Abstract

This study investigates the effect of survey refusal and noncontact on the nonresponse error in the household survey. For this purpose we analyzed the data of the interviewer's field work report. The survey data quality is affected by nonresponse rate and nonresponse error, and also nonresponse rate measures the reliability of the survey data. The household survey mainly contains two types of nonresponses of refusals and noncontacts. These refusals and noncontacts have different effect on the nonresponse error. This could be a venue for future research interested in decreasing the error due to noncontacts and refusals.

Keywords: Refusal, noncontact, nonresponse rate, nonresponse error.

¹Corresponding author: Statistics Research Institute, KNSO, 139 Seonsaro, Seo-gu, Daejeon 302-701, Korea.
E-mail: sykim2217@nso.go.kr