

제5장

경제활동인구조사의 자료수집방법 비교분석

박영실 · 정남수

제1절 서론

경제활동인구조사는 취업 및 실업, 노동력 등 경제활동에 관한 기초 자료를 제공하는 국가공식통계로, 빠르게 변화하는 사회상을 제대로 반영하기 위해 조사항목뿐 아니라 표본설계, 가중치 조정 등 다양한 영역에서 개선을 시도해 왔다. 또한 좀 더 정확하고, 신속하게 자료를 수집하기 위해 1990년대 후반부터는 컴퓨터를 이용한 자료수집방법을 도입하였다. 대면조사에 1999년에는 노트북을, 2004년에는 PDA(Personal Digital Assistants)를 적용하였으며, 2008년에는 일부 가구에 한해 인터넷조사, 2009년에는 CATI(Computer Assisted Telephone Interviewing)를 실시하고 있다. 그러나 컴퓨터를 기반으로 하는 자료수집방법이 종이조사표를 기반으로 하는 기존의 방법을 완전히 대체한 것은 아니며, 대면조사, 전화조사, 자기기입식조사에서 종이조사표와 전자조사표 모두를 사용하는 혼합방법조사(mixed mode survey)가 이루어지고 있다.

혼합방법조사는 응답거부가 심해지는 상황에서 점차 늘어나는 추세이다. 이는 다양한 응답방법을 제공함으로써 응답자들이 자신의 선호에 따라 응답을 할 경우에 응답률이 증가할 것이라는 가정에 따른 것이다(Dillman and Tarnai, 1988). 하지만 혼합방법조사를 실시할 경우 각각의 자료수집방법이 자료의 품질에 미치는 효과가 차별적이지는 않은지를 검토해야 한다. 응답률이 높아짐에 따라 무응답오차가 줄어들 것으

로 기대되고는 있으나(Hox and de Leeuw, 1994), 실제 무응답오차가 줄어들지는 확실한 바 없으며, 설령 혼합방법조사를 통해 높은 응답률을 얻었다고 하더라도 이것이 과연 덜 편향된 자료를 만드는가에 대해서는 의문이 제기되고 있다. 여러 가지 자료수집방법을 사용함으로써 오히려 다양한 유형의 측정오차가 발생할 개연성이 있다는 것이다 (Voogt and Saris, 2005).

이러한 문제제기에 따라 이 연구에서는 혼합방법조사에서 각각의 자료수집방법이 자료의 품질과 관련된 지표에 어떠한 효과를 미쳤는지를 경제활동인구조사자료를 이용하여 살펴보고자 한다. 앞서 보았듯이, 경제활동인구조사에는 여러 가지 자료수집방법이 혼합되어 있다. 지금까지는 새롭게 도입된 방법으로 자료를 수집하는 비중이 높지 않아 자료수집방법에 따른 효과를 분석하는 데 관심이 높은 편은 아니었을 것으로 생각되나, 향후 그 비중이 증가될 것으로 예상되는 가운데 자료수집방법효과(mode effect)에 대한 분석이 반드시 이루어져야 한다. 자료수집방법에 따라서 오차나 추정치에 차이가 생긴다면 그 차이를 보완할 수 있는 대안들이 마련되어야 하기 때문이다.

연구내용과 순서는 다음과 같다. 2절에서는 방법효과에 대한 기존 문헌 및 사례연구를 할 것이다. 사례연구는 우리나라의 경제활동인구조사와 유사한 노동력조사(Labour force survey, LFS)를 실시하고 있는 국가를 중심으로 진행할 것이다. 미국, 캐나다, 호주, 영국의 노동력조사 자료수집방법 현황과 미국 CPS(Current Population Survey)에서 1994년에 자료수집방법을 변경하면서 실시한 실험을 검토할 것이다. 3절에서는 2005~2008년까지의 경제활동인구조사 자료를 이용하여 방법효과를 분석하고, 4절에서는 사례연구 및 자료분석을 바탕으로 향후 혼합방법조사에 대한 시사점을 도출해 보고자 한다. 이는 비단 경제활동인구조사에만 국한되는 것은 아니며 혼합방법조사를 적용하고 있는 여타의 조사에도 분석방법이나 대안 등에 있어서 적용될 수 있을 것으로 기대된다.

제2절 문헌 및 사례연구

1. 문헌연구

자료수집방법이 자료의 품질에 영향을 미치는가에 관한 연구는 주로 다음의 세 분야에서 이루어졌다: 자료의 완전성, 사회적 바람직성 오차 (social desirability bias), 응답효과(response effect)(Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau, 2004).¹⁾ 항목무응답률이 나 내검불일치율(edit failure) 등으로 측정되는 자료의 완전성에 관한 초기 연구에 따르면, 전화조사나 우편조사를 통해 수집된 자료에 비해 대면조사를 통해 수집된 자료의 품질이 높은 것으로 보고되고 있다. 그러나 자료수집방법 간의 이러한 차이는 최근으로 올수록 줄어들고 있으며, 이는 전자적인 내검시스템을 활용하는 컴퓨터를 기반으로 한 방법 효과로 설명되고 있다(de Leeuw and van der Zouwen, 1988).

사회적 바람직성 오차란 응답자들이 민감한 질문, 바람직하지 않은 태도나 행동에 대해 과소보고하는 경향을 일컫는 것으로, 자료수집방법은 사회적 바람직성 오차에 영향을 미치는 것으로 나타났다. 민감한 질문에 대해서는 조사원 기입식보다는 응답자 기입식에서 사회적 바람직성 오차가 줄어든다는 것이다(Aquilino, 1992; Aquilino and Lo Sciuto, 1990 등).

자료수집방법은 응답분포에도 영향을 미치는 것으로 보고되고 있다. 일반적으로 응답자들은 처음에 나오는 응답범주를 선택하거나(primacy effect) 혹은 마지막 응답범주를 선택하는 경향(recency effect)이 있는데, 이것은 부분적으로 자료수집방법과 연결되어 있는 것으로 알려져 있다. 질문이 시각적으로 전달되는 조사에서는 앞쪽에 위치한 응답범주를 선택할 가능성이 높은 반면에 청각적으로 전달되는 조사에서는 뒤쪽에 위치한 응답범주를 선택할 가능성이 높다는 것이다(Groves et al., 2004).

방법효과에 대한 기존 문헌을 검토해 본 결과, 자료수집방법은 측정 오차에 영향을 미치며, 특히 사회적 바람직성 오차와 관련해서는 일관적인 연구결과들이 도출되고 있는 것으로 나타났다.

1) 좀 더 자세한 내용은 박영실·정남수(2009)를 참고할 것.

그런데 방법효과가 자기선택 오차(self-selection bias)에 의해 편향될 수 있음에 주의해야 한다. 자기선택 오차란 집단 간의 차이가 집단의 특성보다는 집단에 대한 자기선택의 결과라는 것을 생각하지 못하는 데서 비롯되는 것이다. 예컨대, 물리학을 전공하는 학생들이 수학점수가 높은 이유가 수학과목이 많이 배정되어 있는 프로그램 효과라고 결론지을 수도 있으나 또 다른 한편으로 볼 때 수학을 잘 하는 학생들이 물리학을 선택한 결과일 수도 있다는 것이다(Harvey, 2006). 따라서 자기선택 오차를 인지하지 못하고, 이를 보정하지 못한다면 엉뚱한 대안이 마련될 수도 있다. 이는 방법효과 분석에서도 마찬가지이다. 우편조사와 대면조사 중 우편조사로 수집된 자료에서 항목무응답률이 높게 나왔다고 하자. 그 원인이 대면조사는 조사원에 의해 항목무응답률이 통제되는 반면에 우편조사는 그렇지 않다는 방법적인 특성 때문일 수 있지만, 또 다른 한편으로는, 조사에 관심이 덜한 사람들이 우편조사를 선택했을 가능성에 의해 항목무응답률이 높게 나왔을 수도 있다. 그러므로 방법효과를 분석함에 있어서 후자와 같은 자기선택 오차를 통제해 주어야 한다.

물론, 자료수집방법이 표본 내에서 무작위로 할당될 경우 자기선택 오차에 대한 우려에서 어느 정도 자유로울 수 있다. 그러나 이것을 실제 구현하기란 쉽지 않다. 예컨대, 특정지역이 인터넷조사방법으로 할당되었다고 할지라도 그 지역 혹은 응답자가 인터넷조사를 할 수 있는 여건이 되지 않는다면, 해당 지역 및 응답자에게 적합한 자료수집방법이 제공될 수밖에 없기 때문이다. 그러므로 방법효과를 분석함에 있어서 자기선택 오차를 집단의 특성과 분리해내는 것은 무엇보다도 중요하다.

2. 사례연구

가. 해외 노동력조사의 자료수집방법 현황

패널조사에서는 비용 및 시간의 효율성을 높이기 위해 1회차(wave)에서는 대면조사를, 2회차 이후에서는 전화조사를 하는 경우가 많이 있다. 이러한 유형의 혼합방법조사가 LFS에서도 많이 채택되고 있다. 노

동력조사는 동일 가구(원)를 대상으로 일정 기간 동안 특정의 주제를 반복적으로 측정한다는 점에서 패널조사와 닮아있기 때문이다. 미국, 캐나다, 영국, 호주 등에서는 이와 같은 혼합방법조사를 실시하고 있다.²⁾

1) 미국

미국의 CPS는 약 72,000가구를 대상으로 매월 실시되는데, 표본으로 선정된 가구는 4개월간 응답을 하고(1~4회차), 8개월 동안 쉬었다가 이후 4개월 동안 다시 응답을 한다(5~8회차). 이러한 시스템 하에서 1회차와 5회차에는 대면조사가, 2~4회차와 6~8회차에는 응답자의 동의에 따라 전화조사가 이루어진다. 1994년 1월 자료수집방법에 컴퓨터가 도입되면서 대면조사와 전화조사는 각각 CAPI(Computer Assisted Personal Interviewing; CAPI)와 CATI로 변경되었다. 이 방법의 핵심은 컴퓨터 스크린 상에 나타나는 조사항목을 읽어가면서 면접을 진행한다는 것이다. 그런데 모든 전화조사가 CATI로 수행되는 것은 아니며 전화조사의 상당 부분은 현장조사원이 자신의 집에서 컴퓨터를 보면서 전화하는 방식으로 진행되었다.

각각의 자료수집방법을 통해 수집된 자료의 비율을 보면, 1회차와 5회차에 대면조사는 79.4%, 전화조사는 20.6%, 2~4회차와 6~8회차에 대면조사는 21.2%, 전화조사는 78.8%이다. 한편, 전화조사 중 CATI는 1987년에 300케이스로 시작한 것이 점차 증가하여 2004년 현재 약 7,000 케이스 정도이다(2004년 9월 기준).

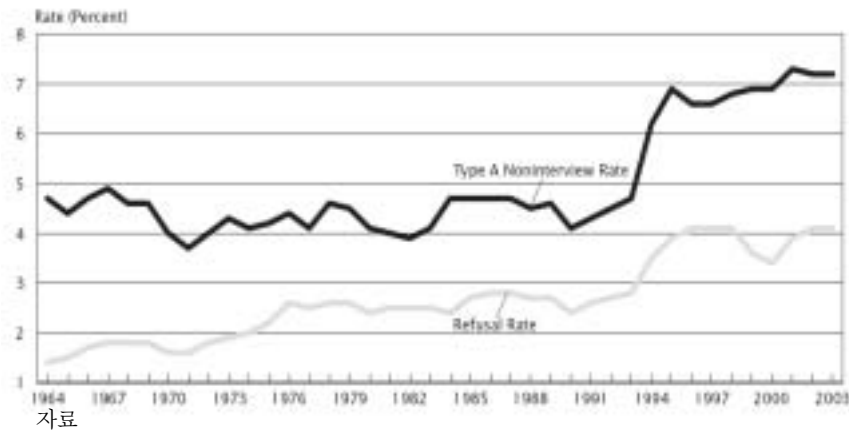
컴퓨터를 이용한 자료수집방법은 긍정적인 측면과 부정적인 측면을 모두 갖고 있는 것으로 평가되었다. 면접 과정 중에 이전 달의 정보를 활용하는 종속면접(dependent interviewing)을 함으로써 응답 부담과 응답 시간을 줄여 준 것으로 나타났으며, 응답자의 성, 연령, 이름에 따라 특성화된 면접을 할 수 있게 되었다.³⁾ 특히, CATI가 대도시 지역의

2) 노동력조사 자료수집방법의 소개는 각 국가의 조사 매뉴얼에 기반하였다.
—(U. S. Census Bureau, 2006; ABS, 2007; ONS, 2007; Statistics Canada, 2008, 2009)

3) 종속적인 면접은 회차에 따라 차별적으로 진행되었다. 8개월 동안 면접을 쉬었다가 다시 면접을 시작하는 5회차와, 그리고 4회차와 8회차에 부가적으로 조사되는 임금 등과 같은 항목에 대해서는 종속적인 면접을 하지 않았다.

표본을 중심으로 수행된다는 점은 조사원 고용과 관련하여 중요한 장점으로 부각되었다. 일반적으로 대도시 지역의 경우 대면조사를 수행하는 것이 어렵기 때문에 각 가구를 방문하여 조사하는 현장조사원의 고용에 어려움을 겪어왔다. 반면, CATI 조사원을 고용하는 것은 수월한 것으로 평가되었다.

그러나 컴퓨터 도입 이후 무응답률은 증가하였다. [그림 5-1]을 보면 거부 및 비접촉 등으로 구성되는 A 유형의 무응답률이 1994년 1월 이후 급격히 증가하였음을 볼 수 있다.⁴⁾



자료

: U.S. Census Bureau, 2006

[그림 5-1] CPS 연평균 전국 수준의 무응답률과 거부율: 1964~2003년

4) CPS에서는 A, B, C 세 가지 유형의 무응답을 제시하고 있다. A 유형은 조사는 가능하나 거부 및 비접촉 등으로 면접을 완료하지 못한 경우, B 유형과 C 유형은 해당 가구가 조사부적격 가구이기 때문에 발생하는 무응답을 일컫는다. 이 중 B 유형은 집을 내놓는 등의 이유로 잠시 비어 있거나 조사대상이 아닌 사람이 살고 있는 경우 등으로 향후에 조사가능 가구로 전환될 수 있는 경우, C 유형은 표본으로 뽑힌 가구가 상업용으로 바뀌었거나 가구 자체가 파손되어 향후 조사기간 내에 해당 표본이 조사가능가구로 들어올 가능성이 없는 경우이다. 2004년 9월 기준으로 무응답 유형의 분포를 보면 총 표본 중 A 유형이 차지하는 비중은 7.6%, B 유형은 16.2%, C 유형은 0.6%이다. 응답률은 총 표본 중 B와 C를 제외한 가구수에 대한 응답한 가구수의 비중으로 계산되었다.

1964~1993년까지는 무응답률이 평균 4~5% 가량으로 안정적이었으나, 1994년 이후 7~8% 가량으로 높아졌다. 이는 다음과 같은 이유 때문으로 추정되었다. 첫째는 자동화된 시스템과 관련된다. 종이조사표를 기반으로 했던 조사에서는 응답자가 조사를 거부할 경우에 핵심적인 질문만 하고 조사를 완료하는 것이 가능하였으나, 자동화된 시스템에서는 조사표 처음부터 끝까지 면접을 모두 진행해야 하므로 응답거부가 증가했을 것이다. 둘째는 조사기간의 문제이다. 처음에 CATI로 수행하기로 한 표본에서 이 방법으로 조사를 완료하지 못했을 경우 그 표본은 현장조사원에게 다시 넘어가게 되는데, 이는 현장조사기간이 그만큼 짧아졌음을 의미하는 것으로 현장조사원이 기간 내에 조사를 완료하는 데 어려움을 겪었을 것으로 추정된다.

2) 캐나다

캐나다의 LFS는 추출된 표본을 6개월간 매월 조사 후 완전 재교체하는 시스템을 따르고 있다. 표본은 회차에 따라서 이름을 달리하고 있는데, 1회차에 해당하는 1/6의 가구는 출생(birth) 가구로, 2~6회차에 해당하는 5/6의 가구는 지속(subsequent) 가구로 부르고 있다.

자료수집은 출생가구는 CAPI에 의해, 지속가구는 CATI에 의해 수행된다. 지속가구의 경우, 현장조사원에 의한 전화조사이었던 것이 2000년 중반 이후부터 중앙집중화된 면접시설에서 CATI로 진행되었다. CATI에서는 표본가구를 조사원에게 임의로 할당하는데, 따라서 출생가구의 대면조사를 담당했던 조사원이 계속해서 이 가구를 담당한다는 보장은 없다.

2004년에는 현장 방문비용을 줄이기 위해 자료수집방법에서 핵심적인 변화가 시도되었다. 출생가구 중 전화번호 수집이 가능한 도시지역의 가구에 대해 첫 번째 접촉부터 CATI를 이용하는 것이다(Telephone First Contact; TFC). 매달 진입하는 9,000가구 중 약 6,000가구가 도시지역에 거주하며, 이 중 60~70%에 해당하는 가구의 전화번호를 행정정보로부터 입수하는 것이 가능하였다. TFC로 분류된 가구는 자료수집기간의 처음 이틀(일요일, 월요일) 동안 전화를 걸어 접촉을 시도하며, 매

달 약 출생가구의 1/3 가량이 전화접촉 후 응답을 하는 것으로 보고되었다. 자료수집방법의 이와 같은 변화는 무응답률의 양과 패턴에 영향을 미쳤으며, 현재 무응답은 조사가능한 가구의 10% 정도인 것으로 나타났다.

캐나다 통계청은 높은 응답률을 유지하기 위해서 추출된 가구에 다양한 유형의 편지를 보내는 전략을 채택하고 있다. 처음으로 조사를 하는 가구에는 LFS에 대한 정보가 담긴 편지(introductionary letter)를, 응답을 꺼리는 가구에는 설득편지(refusal letters)를 보낸다. 한편, 70세 이상 노인들의 응답부담을 줄이기 위해 이전 면접의 응답내용을 다시 제공하고 있다. 이는 영국이나 미국에서 모든 연령 집단을 대상으로 종속 면접을 하고 있는 것과는 차별적이다.

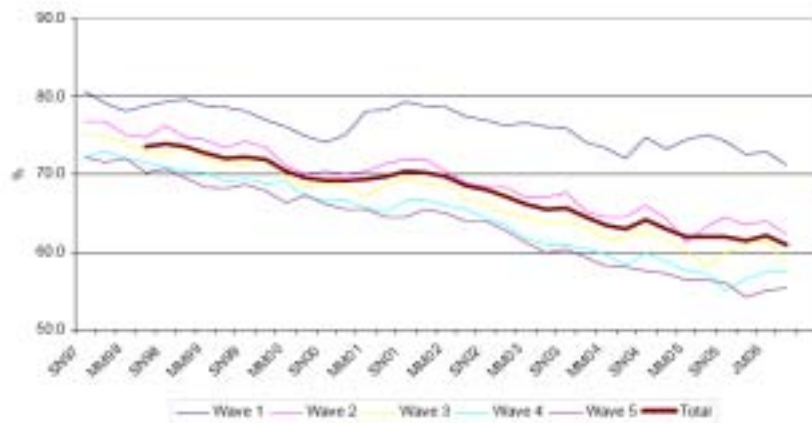
3) 영국

영국의 LFS는 분기별로 수행되며, 이 조사에 한번 표본으로 선정된 가구는 5분기 후에는 완전히 교체된다. 1회차 조사는 CAPI로, 2회차 이후의 조사는 응답자의 동의하에 CATI로 진행되며, 전 회차 평균 약 38%가 CAPI로, 62%가 CATI로 이루어지고 있다. 영국에서도 미국과 마찬가지로 종속면접을 실시하고 있다. 자동화된 조사 시스템은 이전 회차의 응답내용을 현재 면접에서 볼 수 있도록 하는 것을 가능하게 하였는데, 이는 매번 처음부터 조사를 하는 것보다 좀 더 정확한 결과를 제공하는 것으로 평가되고 있다.

1회차 면접에 앞서 모든 가구에 사전편지를 보내는데, 이 편지에는 해당 가구가 표본으로 선정된 방법 및 비밀보장 등에 대한 내용이 포함되어 있다. 조사 결과와 그의 활용에 관한 리플릿(Purpose Leaflet)을 보내기도 한다.

영국 LFS의 2006년 현재 응답률은 60%를 약간 웃도는 것으로 나타났다는데 이는 미국이나 캐나다의 응답률이 90% 수준인 것과 비교해 볼 때 낮은 수준이다([그림 5-2]). 이에 따라 영국 통계청은 무응답률을 최소화하기 위해 여러 가지 전략을 채택하고 있다. 접촉이 어려운 가구에 대해서는 최소 4번 이상 방문을 하되, 이 중 두 번은 반드시 저녁이나

주말을 포함시켜야 한다. 또한, 최근에는 모든 조사원들을 대상으로 응답 거부를 피하는 방법에 관한 교육(Avoidance Refusal Training)을 실시하고 있기도 하다.



자료

: ONS, 2007

[그림 5-2] 영국 LFS의 임퓨테이션 이전의 회차별 응답률: SN97-AJ06

4) 호주

호주에서도 앞의 세 국가와 마찬가지로 주로 대면조사와 전화조사를 통해 자료를 수집하고 있다. 한번 응답자를 선정하면 8개월을 연속적으로 조사하는데, 1회차에는 대면조사를, 2회차부터 8회차까지는 응답자의 동의하에 전화조사를 실시하고 있다.

전화조사는 1996년 8월부터 1997년 2월까지 단계적으로 도입되었다. 그 이전까지는 대면조사만으로 자료가 수집되었으나, 전화조사로 수집된 자료가 대면조사로 수집된 자료와 비교할 만한 수준일 뿐 아니라 전화조사가 이동시간을 줄여주기 때문에 결과적으로 비용감소의 효과가 있다는 평가 하에 전화조사가 도입되었다. 마찬가지로 2003년 10월부터 2004년 8월까지 자료수집방법에 컴퓨터 방식이 도입되었다.

호주의 LFS는 대면조사와 전화조사 이외에도 거부나 비접촉 등으로 면접이 어려운 가구에 한해 자기기입식 조사를 허용하고 있으며, 응답률은 약 96% 가량인 것으로 보고되고 있다.

나. 미국 CPS의 방법효과 실험 사례연구

LFS와 같이 동일한 내용을 반복적으로 측정하는 조사에서 자료수집 방법을 변경할 경우, 그것이 자료의 품질에 어떠한 영향을 미치는지를 검토해야 한다. 만일 방법변경 전과 후에 자료의 단절이 생긴다면 그것을 설명할 수 있는 근거를 제공해 주어야 한다. 여기에서는 미국 CPS에서 1994년 1월에 컴퓨터를 기반으로 한 방법을 도입하면서 도입 전과 후의 자료 연결을 위해서 실행한 일련의 과정들을 검토해 봄으로써 향후 우리나라의 통계조사에서 방법변경 시 고려할 점에 대한 시사점을 얻고자 한다.

1) 자료수집방법 변화의 배경

미국의 센서스국과 노동통계국은 1994년 1월, CPS의 조사표와 자료수집방법의 대대적인 변화를 시도하였다. 1967년 개편 이래 27년만의 일로 그동안 사회적으로나 조사방법론 상으로 많은 변화가 있어왔다. 서비스업 영역에서 고용이 증가했고, 노동시장에서 여성의 역할이 강조되었으며, 시간제와 자영업 등 다양한 유형의 고용형태가 등장하였다. 뿐만 아니라 조사표 설계 및 자료수집방법에서도 많은 발전이 있어왔다. CPS에서는 이러한 변화를 반영해 줄 수 있도록 조사표 내용을 개선하고, 자료수집방법을 컴퓨터를 이용한 방법으로 변경하는 개편안을 마련하였다.

이 개편은 자료의 품질 향상을 목적으로 하였다.⁵⁾ 조사표 개선을 통해 좀 더 정확한 응답자료를 얻어내고, 컴퓨터를 기반으로 한 방식으로 자료수집방법을 변경함으로써 조사원에 의해 발생하는 잠재적인 오차를 줄이고자 하였다(Bregger and Diplo, 1993).⁶⁾

가) 개편 진행 과정
자료수집방법 변경의 핵심은 CATI의 도입이다. CATI는 기존의 조

5) 1967년의 개편이 비용절감 및 자료처리 시간을 줄이기 위한 것이 목적이었다면, 1994년은 품질향상이라는 점에서 대비된다.

6) 본 연구의 목적상 자료수집방법 개편과 관련된 내용을 중심으로 소개하고자 한다.

사체계와는 차이가 있었다. 이전에는 현장조사원에 의해 개별적으로 전화조사가 이루어졌다면, CATI는 중앙집중화된 전화면접 시설에서 관리자의 감독 하에 조사가 이루어진다는 점에서 큰 차이가 있었다. 이에 따라서 센서스국과 노동통계국은 CATI 도입이 CPS 자료에 미치는 효과에 대한 검토를 하였다.

1985년에 CPS에 CATI가 처음 소개된 이래 몇 년 동안은 CATI 운영과 관련된 문제만을 집중적으로 파악하였으며, CATI 운영이 안정화 단계에 접어든 이후에야 업무량을 점차적으로 늘렸다. CATI가 조사비용 및 노동력 추정치에 미치는 효과에 대한 검토는 다음의 단계를 거쳐서 진행되었다. 먼저 1단계로 표본 내에서 소규모 실험(1991년 1월~1992년 12월)을 거친 후, 다음으로 전국적인 차원에서 CPS와 별도의 독립적인 조사를 실시하였다(1992년 7월~1993년 12월). 마지막으로 CATI 도입 전후 자료의 시계열 단절을 보완할 수 있는 보정요인을 개발하기 위해 추가적인 실험을 하였다(1994년 1월~1994년 5월)(U. S. Census Bureau, 2006).

(1) 1단계 CATI의 단계적 도입 프로젝트에서는 CATI의 중앙집중 면접효과와 컴퓨터를 이용한 전화면접 효과를 측정하기 위한 실험을 하였다.

실험은 CPS 표본 내에서 CATI 면접이 가능한 지역을 대상으로 이루어졌다. 이 지역을 몇 개의 하위집단으로 나눈 후 무작위로 통제집단과 실험집단으로 할당하였다. 두 집단 모두 1회차와 5회차에서는 종이와 연필에 기반한 대면조사를, 2~4회차와 6~8회차에는 전화조사를 실시하였다. 단, 통제집단에는 현장조사원에 의한 전화조사가, 실험집단에는 CATI가 수행되었다.

자료수집방법 간에 실업률과 고용률 추정결과에 차이가 있는가를 비교해 본 결과, CATI가 수행된 실험집단이 통제집단에 비해 실업률이 더 높은 것으로 나타났다(7.8% vs 7.0%). 즉, 컴퓨터를 이용하여 중앙집중화된 전화면접을 하는 경우가 종이조사표를 기반으로 현장조사원이 전화면접을 하는 것에 비해 실업률을 더 높게 예측하였다. 그러나 실제 CATI를 수행하기 위해 조사표를 만드는 과정에서 조사표의 내용이 전

자적인 형식에 맞게 약간 수정되었는데, 따라서 CATI 효과가 중앙집중 면접효과인지, 컴퓨터를 이용한 전화면접 효과인지, 수정된 조사표 효과 인지를 이 실험을 통해 분리해내기는 어려웠다(Shoemaker, 1993).

(2) 2단계

노동통계국과 센서스국은 실험결과를 일반화하기 위해 1992년 7월부터 1993년 12월까지 전국 규모의 프로젝트를 실시하였다. 종이조사표에 기반한 면접을 하는 CPS와는 독립적으로 새로운 조사표와 새로운 자료수집방법을 사용한 PS(Parallel Survey)를 실시하여 두 조사 사이에 연평균 노동력 추정치의 결과에 차이가 있는지를 검정하였다(Thompson, 1994). 두 조사의 표본설계를 비교해 보면 다음과 같다(<표 5-1>).

<표 5-1> CPS와 PS 표본설계의 차이점
구분

설계유형	주CPS	전국PS
월간표본규모		
일차표집단위 수	58,900	12,000
복합추정	예)729	아니오)283
사후층화 변수 수	주	주없음
월 실업률 표준오차	600+	559+
연 실업률 표준오차	0.11	0.25
비율	0.07	0.16
자료 CATI	9%	18%

: Miller, 1994

CPS는 전국 및 각 주를 대표하는 6만 가구를 다단계층화표집하였으며, 이 가구는 전국 729개의 지역(geographic area)에 위치하였다. 표본은 전국 단위에서 실업률 추정치에 매월 C.V를 1.7% 제공토록 고안되었다. PS는 전미범죄피해조사(National Crime Victimization Survey)에서 추출한 약 1만 2천 가구를 대상으로 실시했으며, 표본은 283개의 지역에 위치하도록 설계되었다. 통계치는 전국 수준에서 추정이 가능하도록

록 하였으며, 전국 실업률 추정치에 대한 월간 C.V는 3.5%이다. 이외에도 추정방법이나 사후층화에서 차이를 보였다.

자료수집방법과 조사표를 제외하고 두 조사에는 동일한 절차가 적용되었다. CPS는 종이조사표에 기반한 방법으로 수집된 반면에, PS는 수정된 조사표를 기반으로 컴퓨터를 이용한 자료수집방법으로 수집되었다. 즉, 현장조사원은 조사표가 내장된 랩탑컴퓨터를 사용하여 방문 및 전화조사를 하였다. 한편, CATI는 CPS와 PS 모두 CATI 가능지역에 한정해서 실시되었는데(CPS-CATI지역과 PS-CATI지역), CPS에서는 약 9%, PS에서는 약 18%가 CATI에 의해 수행되었다.

분석결과, 새로운 측정시스템이 적용된 PS에서 1993년 연평균 실업률은 CPS보다 0.5%p 더 높은 것으로 추정되었다. 특히, 여성의 경우에는 PS에서 실업률이 CPS에 비해 0.7%p 높은 것으로 나타났다. 고용률은 전체적으로는 두 조사에서의 차이가 거의 나타나지 않았으나 성별에 따라 남자의 경우 CPS보다 PS에서 0.6%p 낮았으며, 여성의 경우 0.7%p 높은 것으로 나타났다.

그렇다면 이러한 차이는 조사표에 의한 것인가? 자료수집방법에 의한 것인가(Polivka, 1994)? CATI 효과를 분리해서 살펴보기 위해 PS-CATI 지역의 표본을 무작위로 A와 B로 나눈 후 A지역은 CATI로, B지역은 현장조사원에 의한 전화조사(non-CATI)로 자료를 수집하였다. 두 집단 비교를 통해서 중앙집중화된 면접효과와 분산화된 면접효과(decentralized interviewing)의 차이를 비교해 볼 수 있다. 마찬가지로 CPS-CATI 지역의 표본을 무작위로 C와 D로 나눈 후 C지역에는 CATI로, D지역에는 현장조사원에 의한 전화조사로 자료를 수집하였다. 그런데 이 실험은 PS와는 다르게 조사표 효과가 혼재되어 있다. CPS-CATI 조사에 사용된 조사표는 컴퓨터 환경에 맞게 워딩이 변경되었으나, 현장조사원이 전화조사를 하는데 사용된 조사표는 변경 전의 것이기 때문이다.

먼저, PS 자료분석을 통해서 CATI에 의한 중앙면접효과가 발생했는지를 살펴보았으나, CATI와 non-CATI 사이에서 실업률과 고용률의 차이가 95% 신뢰수준에서 통계적으로 유의미하지 않은 것으로 나타났다. 이러한 패턴은 고용률에서도 마찬가지로 나타났다. 한편, CPS 자료 분

석에서는 고용률에서는 두 집단 간에 95% 신뢰수준에서 통계적으로 유의미한 차이가 나타나지 않았으나, 실업률에서는 CATI에서 non-CATI에 비해 실업률이 더 높게 나타났으며 그 차이는 동일한 신뢰수준에서 통계적으로 유의미하였다. 실업률에 대한 CATI 효과는 전체 집단뿐 아니라, 성 및 인종별로도 유사한 패턴이 나타났다. 이 결과는 CATI가 실업률에 영향을 미친다는 것으로 해석할 수 있으나, 그 효과가 중앙면접 효과인지 혹은 중앙면접을 사용한 조사표효과인지를 확실히 구분할 수 없다는 한계를 갖고 있다.

(3) 3단계

노동통계국과 센서스국은 조사체계의 변경이 실업률 추정 결과에 영향을 미칠 수 있다는 연구 결과에 따라 브릿지 팀(Bridge Team)을 구축하여 자료의 시계열 단절을 극복할 수 있는 대안을 마련하고자 부심하였다.

브릿지 팀은 먼저, 실업률 추정 결과의 차이가 CPS와 PS 표본설계상의 차이에 의해서 발생한 것인지에 대한 검토를 하였다(<표 5-1> 참고). 실업률 추정 결과의 차이가 CATI 비율 차이에서 기인했는지(Polivka, 1994), 표집오차와 사후층화 기준 모집단(population control)의 변경 때문인지(Miller, 1994), 이 요인들을 모두 포함하여 사후층화 요인 수, 복합추정, 계절조정에 의한 것인지(Kostanich and Cahoon, 1994)를 분해하였다. 측정 가능한 설계상의 차이를 모두 고려하여 가중치를 부여한 후에도 CPS와 PS에서 실업률의 추정 결과에 차이가 발생하였다. 이는 두 개의 조사로부터 관찰된 추정치가 표본설계에 의한 차이가 아니라는 결론을 이끌었다. 예컨대, 실업률에 대한 방법효과는 노인층에서 유의미한 것으로 나타났다. 새로운 방법론을 적용한 이후에 노인층에서 실업률이 높게 추정된 것은 자동화와 조사표의 리위딩(rewording) 효과로 풀이되었다. 구조조사표의 노동력관련 질문에는 은퇴라는 응답범주가 없어 은퇴한 사람들의 불평이 꽤 많았는데, 신조사표에서는 이 집단의 부담을 완화시키기 위해서 은퇴를 응답범주에 추가시켜 이후의 문항에 자동적으로 넘어가게 하였다(Polivka and Miller, 1994).

다음으로, 조사체계 변경에 따른 시계열 단절을 연결하기 위해 1994

년 1월부터 5월까지 추가적인 실험을 하였다. 2단계 실험과 반대로, 새로운 조사표와 자료수집방법이 적용된 CPS와는 독립적으로 기존의 조사표와 자료수집방법이 적용된 PS를 실시하였다. 이 두 자료와 1994년 이전의 두 자료(구 측정시스템을 적용한 CPS, 신 측정시스템을 적용한 PS)를 이용하여 자료수집방법을 주효과로 하는 선형모형을 구축한 후 모수를 추정, 보정요인을 개발하였다. 노동통계국과 센서스국은 이 보정요인을 CPS 시계열 자료를 이용하는 연구자 및 일반 사용자들에게 제공함으로써 시계열을 지속적으로 유지할 수 있게 하였다(Polivka and Miller, 1994).

나) 사례연구 결과 요약

미국과 캐나다, 영국, 호주의 자료수집방법 현황을 살펴본 결과, 우리나라와 여러 가지 측면에서 차이점이 있음을 알 수 있었다. 첫째, 네 국가 모두 1회차에는 대면조사를, 2회차 이후에는 응답자의 동의하에 전화조사를 하는 혼합방법조사를 하고 있었다. 우리나라의 경우, 조사회차와 무관하게 조사원이 각 가구를 방문하여 조사를 하는 대면조사를 원칙으로 하고 있다는 점에서 차이가 있는데, 과연 실제 조사현장에서 36회의 모든 회차에 대면조사를 해야 한다는 원칙이 어느 정도 보편적으로 적용되고 있는지에 대한 재검토가 필요하다.

둘째, CATI의 경우 중앙집중화된 시설에서 관리자의 감독 하에 별도로 고용된 조사원에 의해 면접이 이루어지고 있었다. 그런데 우리나라에서 도입한 CATI는 이와는 개념이 다르다. 해당 가구의 현장조사원이 컴퓨터 모니터 상에 나타나는 조사표를 보면서 개별적으로 면접을 하는 방식이다. 과연 이 방법이 컴퓨터 도입 이전의 전화조사와 비교해 어떠한 차별적인 효과가 있는가를 검토해 볼 필요가 있을 것이다.

셋째, 자료수집방법 변경은 응답률이나 추정치 등에 있어서 변화를 수반하였다. 중요한 것은 그러한 차이가 발생하게 된 원인을 찾고 그 차이를 보완할 수 있는 대안적인 방법을 모색하는 것이다. 이러한 면에서 볼 때, 미국 CPS에서 실시한 실험은 우리에게 자료수집방법 변경 시 고려할 점 등에 대한 시사점을 제시해 준다고 할 수 있다. 노동통계국과

센서스국은 방법 변경 과정에서 발생할 수 있는 시계열 단절을 극복하기 위해 여러 차례 시험조사를 실시하여 그 효과를 다각도에서 검토하였다. 표집오차뿐만 아니라 비표집오차를 포함하여 추정치에 영향을 미칠 수 있는 다양한 변인을 검토하였으며, 본문에서는 언급하지 않았으나 조사원 교육, 조사원 경험, 현장관리방법 등 세세한 모든 영역이 포함되어 있었다(Dippo, Kostanich, and Polivka, 1994). 이러한 일련의 과정들을 통해서 신뢰성 있는 통계자료를 생산할 수 있을 것이다.

제3절 자료분석

1. 분석자료

가. 자료수집 절차

경제활동인구조사는 약 32,000가구를 대상으로 하는 월간조사로 표본가구의 1/36씩인 900가구가 매월 새롭게 진입하며, 이 가구는 36회를 응답하면 탈락하게 되는 연동표본제(rotation sampling system)를 채택하고 있다. 따라서 2005년에 연동표본이 시작되고 3년이 지난 2007년 12월 조사 자료에는 이 시점에 처음 진입하여 1회차 응답을 한 가구부터 2005년 1월에 진입하여 36회차를 응답한 가구까지 모두 포함되어 있다.

자료수집 절차는 다음과 같다. 조사대상주간(매월 15일이 포함된 1주) 직전 1주간에 보조조사표를 배부하여 응답자로 하여금 근로시간이나 직업, 산업 등과 같은 기본적인 사항을 기입하도록 한다. 그리고 나서 조사대상주간 이후에 가구를 방문하여 보조조사표를 바탕으로 면접을 한다. 이 때 응답자를 만나는 것이 어렵거나 거부가 심할 경우, 혹은

7) 그런데 실제 조사과정에서는 새롭게 진입하는 가구 중에 연동이 아닌 다른 이유로 진입을 하는 가구가 있다. 표본가구가 36회 응답을 완료하기 이전에 전출이나 기타의 이유 등으로 탈락할 경우, 해당 가구에 진입해 오거나 대체되는 가구 등이 이 경우에 해당된다. 그런데 자료분석결과 2005~2008년까지 이러한 이유로 진입해 들어오는 개인의 비율이 점차 증가하고 있었다(1.6%(2005.1)→2.3%(2006.1)→2.8%(2007.1)→3.9%(2008.1)). 이 비율이 점차 증가한다는 것은 조사환경의 어려움을 간접적으로 보여주는 것이라고 할 수 있다.

응답자가 원하는 경우에 한해 전화조사나 자기기입식조사를 실시한다. 면접은 가구원마다 개별적으로 직접 면접을 하는 것이 원칙이나 부득이한 경우에는 가구내용을 가장 잘 알 수 있는 사람에 의한 대리응답을 허용하고 있다(통계청, 2008).

자료수집에 컴퓨터 방식이 도입된 것은 1999년 1월이다. 종이조사표를 기반으로 면접이 이루어진 대면조사가 컴퓨터 스크린 상에 나타나는 조사표를 읽어가면서 면접을 하는 CAPI 방식으로 바뀌었다. 2008년 1월에는 자기기입식조사에서도 컴퓨터 방식을 적용하는 인터넷조사가 도입되었다. 그런데 종이조사표를 이용한 자기기입식조사가 응답자가 원하거나 혹은 응답자를 만나기 어려운 경우에 실시되는 것과 달리, 인터넷조사는 경제활동상태가 안정적이며 가구사항이 충분히 파악된 가구를 대상으로 실시된다는 점에서 차이가 있다. 한편, 2007년부터 몇몇 조사구에서 일정 회차 이상 응답한 가구를 대상으로 CATI 시험조사를 실시했으며, 2009년에는 이 방법을 전국적으로 확대할 예정에 있다.⁸⁾

그러나 컴퓨터를 기반으로 한 자료수집방법이 종이조사표를 기반으로 한 전통적인 방법을 완전히 대체한 것은 아니다. 따라서 현재 경제활동인구조사에 사용되고 있는 대면조사, 전화조사, 자기기입식조사에서는 종이조사표와 전자조사표를 모두 이용하고 있었다. 종이조사표와 전자조사표의 내용 및 구성은 동일하다.

나. 자료수집방법 현황

1) 자료수집방법의 측정

자료수집방법은 [그림 5-3]과 같이 면접조사, 전화조사, 자기기입식조사 등으로 구분된다. 그러나 자료수집방법을 분류하는데 적용되는 기준에 모호한 점이 있다. 첫째, 응답범주 1번부터 5번까지는 자료수집방법을 묻고 있으나 응답범주 6번은 누가 응답했는지를 묻고 있다. 따라서 자료분석시 응답범주 6번은 무응답 처리 후 분석하였다.⁹⁾

8) 그러나 미국이나 캐나다, 영국 등의 CATI가 중앙집중화된 면접시설에서 별도의 조사원을 고용해서 실시한 것과는 달리, 우리나라의 경우 해당 가구의 담당 조사원에 의한 전화조사를 CATI로 부르고 있다.

조사원별 직립사항		
조사원	직립	직립
조사원직립	1. 면접조사	4. E-mail조사
직립방법	2. 전화조사	5. Fax조사
	3. 자기기입식조사	6. 단면조사
조사원 직립		
직립구분		
직립		

둘째, 면접조사, 전화조사, 자기기입식조사 등의 용어를 좀 더 분명히 할 필요가 있다. 여기에는 조사원의 참여도와 응답자와의 접촉여부라는 두 가지 기준이 포함되어 있는데, 조사원의 참여도에

[그림 5-3] 자료수집방법 구분 항목 따르면 조사원의 참여도가 높은 면접조사와 전화조사, 조사원의 참여도가 낮은 자기기입식조사로 분류되며 응답자와의 접촉여부에 따라서는 응답자와 직접 접촉하는 면접조사와 자기기입식조사, 응답자와 간접 접촉하는 전화조사로 구분된다(<표 5-2>). 자기기입식조사의 경우, 상황에 따라서 응답자와 간접 접촉하는 방법으로 포함시킬 수 있으나 응답자를 방문하였을 경우 응답자가 거부하거나 응답자가 원할 경우에 응답자 스스로 기입하도록 한다는 점에서 직접 접촉에 포함시켰다. 이 구분에 따르면 면접조사는 실제 조사원이 가구를 방문하여 응답자와 직접 만나서 진행하는 조사를 의미하는 것이므로 ‘face-to-face interviewing’을 번역한 대면조사로 바꾸어 표현하는 것이 더 적합할 것으로 판단된다. 또한 1999년 이후로 자료수집 방법에 컴퓨터 방식이 도입되었으며, 자기기입식조사의 경우에도 종이 조사표를 이용하는 경우와 전자조사표를 이용하는 경우가 공존하므로 향후에는 컴퓨터 사용여부를 분류기준으로 포함시켜야 할 것이다.

셋째 이메일조사나 팩스조사의 경우, 현재 거의 이용되지 않고 있으므로 향후 조사표 개선 시 이 두 범주는 삭제할 필요가 있다.

9) 2005~2008년까지 평균 약 5~6% 가량이 간접조사된 것으로 나타났는데, 향후에는 자료수집방법과 응답주체를 별도의 항목으로 구분해서 물어야 할 것이다. 직접보고와 대리보고 사이에는 내재적인 바이어스가 발견될 수 있기 때문이다. 노동력과 주요 인구사회학적 특성을 중심으로 직접보고와 대리보고를 비교해 보면 체계적으로 차이가 나타나는 경향이 있다(Census Bureau, 2006). 그러므로 조사표 개선을 통한 자료수집방법의 정확한 측정과 향후 이 분야에 대한 연구가 요구된다.

〈표 5-2〉 자료수집방법 구분

구분	조사원의 높은 참여		조사원의 낮은 참여	
	종이	컴퓨터	종이	컴퓨터
응답자와 직접 접촉	대면조사	CAPI	자기기입식	CASI
응답자와 간접 접촉	전화조사	CATI	자기기입식	인터넷조사

자료

주 : Biemer and Lyberg, 2003 . 는 조사원이 를 하기 위해 방문하였
 1) CASI(Computer-Assisted Self-Interviewing) 모니터를 응답자 방향으로 전환시켜
 응답자가 기입하도록 하게 하는 것을 말함 ,
 자기기입식조사는 응답자와 직접 접촉을 통해 이루어질 수도 있고 그렇지 않을
 수도 있음

2) 자료수집방법 분포 현황

자료수집방법 분포 현황은 대면조사 이외의 대안적인 자료수집방법이 실제 조사현장에서 어느 정도 사용되고 있는지, 그리고 응답회차에 따라서 수행되는 자료수집방법에 차이가 있는지 나누어 살펴보았다.

먼저, 2005~2008년 동안 경제활동인구조사 자료수집방법 현황을 조사시점별로 제시하였다(〈표 5-3〉). 자료수집에 이용된 방법을 보면 대면조사, 전화조사, 자기기입식조사 순이고, 대면조사가 90% 이상을 차지하고 있으며 대면조사 이외의 방법으로 수집된 비율은 매우 낮았다.

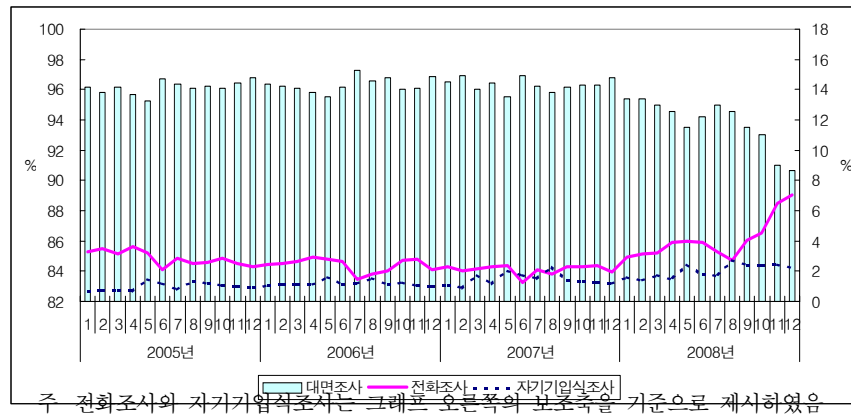
[그림 5-4]는 〈표 5-3〉을 월별로 분해한 것인데, 대면조사를 통해 자료가 수집된 비율은 2007년까지 96%를 중심으로 등락을 반복하다가 2008년 1월부터 감소 추세를 보이고 있다. 전화조사의 경우, 2007년까지 소폭으로 등락을 반복하나 전반적으로 감소하는 모습을 보이다가 2008년 이후 그 비율이 증가하였다.¹⁰⁾ 자기기입식조사를 통해 자료가 수집된 비율은 낮은 편이긴 하나, 2005년 이후 꾸준히 약간씩 증가하고 있음을 알 수 있다.

10) [그림 5-4]를 보면 전화조사로 자료가 수집된 비율이 2005년에는 6월, 2006년에는 7월, 2007년에는 6월에 가장 낮았다. 이 시점은 사회통계조사가 실시된 때라는 공통점이 있다. 사회통계조사를 하기 위해 방문하면서 전화조사를 하던 경제활동인구조사 가구에서 대면조사를 했을 가능성이 높다.

〈표 5-3〉 연도별 자료수집방법 현황

단위 명, %

	대면 조사	전화 조사	자기기입식 조사	기타	전체
2005년	768,089 (96.1)	22,970 (2.9)	7,728 (1.0)	145 (0.0)	798,932 (100.0)
2006년	770,205 (96.3)	19,306 (2.4)	9,533 (1.2)	607 (0.1)	799,651 (100.0)
2007년	749,499 (96.3)	16,425 (2.1)	11,414 (1.5)	733 (0.1)	778,071 (100.0)
2008년	710,199 (93.8)	31,083 (4.1)	15,349 (2.0)	511 (0.1)	757,142 (100.0)



주: 전화조사의 자기기입식조사는 그래프 오른쪽의 보조축을 기준으로 제시하였음

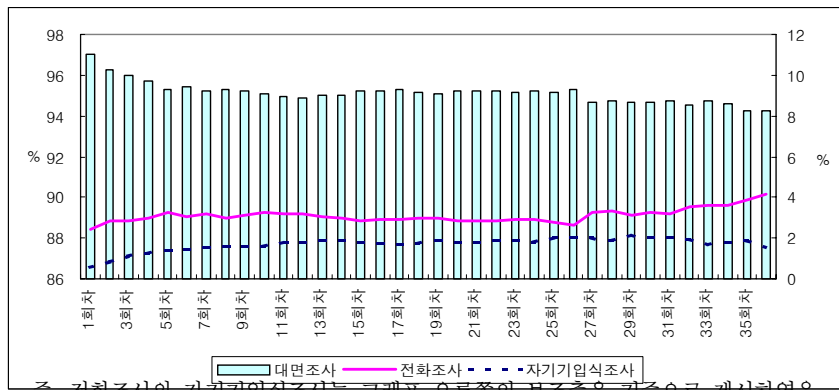
[그림 5-4] 월별 자료수집방법 현황

한편, 대면조사로 자료가 수집되는 비율이 감소하는 시점과 전화조사로 자료가 수집되는 비율이 증가하는 시점이 맞물리는 점에서 볼 때 전화조사가 대면조사의 대안적인 방법으로 활용되고 있음을 쉽게 예상해 볼 수 있다. 그런데 이것은 부분적으로는 부산과 울산에서 2008년에 실시한 CATI 시험조사의 영향이 작용한 것이다.

해당 표본이 몇 회차 응답을 할 차례인지에 따라서 사용되는 자료수집방법이 다를 수 있다. 우리나라의 경우 응답회차에 따라서 자료수집방법을 달리한다는 기준은 없으나 실제 조사현장에서 응답회차가 증가하면 조사원과 표본가구와의 안정적인 관계가 형성되기 때문에 대면조

사 이외의 방법으로 조사했을 가능성이 있다.

[그림 5-5]는 1회차, 2회차, ... , 36회차에 대면조사가 진행된 비율, 전화조사가 진행된 비율, 자기기입식조사가 진행된 비율을 2005~2008년까지 평균한 것이다. 1회차에서 대면조사가 진행된 비율은 평균 97.0%, 전화조사는 2.4%, 자기기입식조사는 0.5%로 나타났으며, 응답자가 표본으로 들어온 후 1년이 되는 시점인 12회차에서 대면조사는 94.9%, 전화조사는 3.2%, 자기기입식조사는 1.8%로 나타났다. 전반적으로 응답회차가 증가할수록 대면조사로 진행되는 비율은 낮아지고, 전화 및 자기기입식조사로 진행되는 비율은 높아졌다. 그러나 그 차이는 크지 않은 편이다.



주: 전화조사와 자기기입식조사는 그래프 오른쪽의 보조축을 기준으로 제시하였음

[그림 5-5] 응답회차별 자료수집방법 현황

그렇다면, 응답회차별 자료수집방법 비율 차이는 시계열적으로 어떻게 변화해왔는지 대면조사를 중심으로 살펴보도록 하겠다. 매월 새로 진입한 표본이 1회차, 2회차, ... , 36회차를 지나는 동안 해당 회차에서 대면조사로 수집된 비율을 종속변수로, 응답회차(1~36회차)를 독립변수로 하는 단순회귀분석을 하였다.

$$\text{대면조사 비율} = B(\text{응답회차}) + \text{상수}$$

이에 따르면, 응답회차가 1회차 증가할수록 대면조사로 수집된 비율

은 회귀계수인 B만큼 증가한다고 예측할 수 있다. <표 5-4>는 위 회귀식에서 응답회차의 비표준화계수와 이 계수에 대한 통계적인 유의성을 검증한 결과를 해당표본이 진입한 시점(t_n , n 은 2005.1, 2005.2, ... , 2007.12)별로 제시한 것이다.¹¹⁾

<표 5-4> 응답회차가 대면조사 비율에 미치는 효과

시점	B	P-	시점	B	P-
t1	0.042	0.001	t19	-0.157	0.000
t2	0.018	0.269	t20	-0.160	0.000
t3	0.047	0.000	t21	-0.206	0.000
t4	0.045	0.002	t22	-0.154	0.001
t5	0.034	0.003	t23	-0.121	0.000
t6	-0.042	0.048	t24	-0.181	0.000
t7	-0.019	0.122	t25	-0.339	0.000
t8	-0.018	0.149	t26	-0.302	0.000
t9	-0.026	0.153	t27	-0.166	0.000
t10	-0.044	0.023	t28	-0.173	0.000
t11	-0.062	0.002	t29	-0.480	0.000
t12	-0.044	0.025	t30	-0.443	0.000
t13	-0.150	0.000	t31	-0.339	0.000
t14	-0.136	0.000	t32	-0.346	0.000
t15	-0.115	0.000	t33	-0.422	0.000
t16	-0.126	0.000	t34	-0.598	0.000
t17	-0.200	0.000	t35	-0.509	0.000
t18	-0.194	0.000	t36	-0.561	0.000

<표 5-4>를 보면 t1(2005.1) ~ t5(2005.5)까지는 응답회차와 대면조사로 수집되는 비율간의 관계가 정(+)적이었으나, t6(2005.6)부터는 응답회차와 대면조사로 수집되는 비율 간의 관계가 부(-)적으로 바뀌었다. 즉, 응답회차가 증가할수록 대면조사로 수집되는 비율이 줄어들었다. 그런데 회귀계수를 보면 t12(2005.12)까지는 비표준화된 회귀계수가 0.1도 안 되는 수준으로 대면조사의 증가 혹은 감소비율이 매우 미미하였으나 최

11) 2007년 12월에 진입한 표본은 2008년 12월에 13회차까지 응답을 하게 된다. 일반적으로 회귀분석을 수행하기 위해 표본의 사례수가 독립변수의 20배 이상은 되어야 한다는 가정에 따라 2008년 1월 이후에 진입한 표본은 분석에서 제외하였다(성태제, 2007). 물론 2007년 6월 이후부터 표본의 크기가 20 이하로 떨어지기는 하나 독립변수와 종속변수간의 정규성 가정을 충족하므로 분석을 수행하였다.

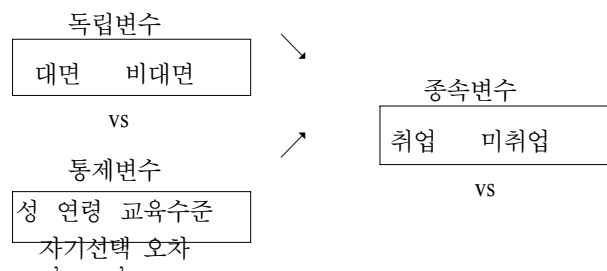
근으로 올수록 감소비율은 커졌다. t25인 2007년 1월의 경우 응답회차가 1단위 증가할수록 대면조사로 자료가 수집되는 비율이 줄어드는 정도는 0.339%였으며, t36인 2007년 12월의 경우 0.561%로 나타났다.

경제활동인구조사의 자료수집방법 현황을 조사시점 및 응답회차별로 살펴본 결과, 첫째 최근으로 올수록 대면조사 비율이 감소하고 전화조사 및 자기기입식조사의 비율이 증가하고 있음을 알 수 있었다. 둘째, 응답회차별로 보면 회차가 증가할수록 대면조사 비율은 감소하는 경향이 있으며 전화 및 자기기입식조사는 회차가 증가할수록 늘어나는 경향이 있었다. 셋째, 조사시점과 응답회차를 동시에 고려해 보면 응답회차가 증가할수록 대면조사 비율은 감소하는 경향이 있는데, 이런 경향은 최근으로 올수록 강화되었다. 이것은 부분적으로 일정 회차 이상의 응답자를 대상으로 실시하는 인터넷조사나 CATI의 영향이 작용한 것으로 판단된다.

다. 변수측정 및 분석방법

다음으로 이 연구의 주목적인 자료수집 방법효과 분석을 위한 변수 측정 및 분석방법에 대해서 설명하고자 한다.

〈표 5-5〉 분석모델



독립변수인 자료수집방법은 대면조사와 비대면조사로 구분된다. 자료수집방법은 대면조사, 전화조사, 자기기입식조사로 측정되었으나, 전화조사와 자기기입식조사의 비율이 낮아서 추정치를 편향시킬 수 있다

는 판단 하에 두 범주를 통합하였다. 또한 전화조사와 자기기입식조사는 대면조사의 대안적인 방법이며, 비대면조사라는 공통된 특성을 갖고 있다. 대면조사를 기준으로 비대면조사의 효과를 살펴보기 위해 비대면 조사를 1, 대면조사를 0으로 더미처리하였다.

종속변수는 경제활동인구의 주요 변수인 취업여부로 취업인 경우를 1, 취업이 아닌 경우를 0으로 더미처리하였다. 경제활동인구조사에서는 취업과 함께 실업이 중요한 지표이나 자료수집방법과 마찬가지로 전체 자료에서 실업이 차지하는 비율이 낮기 때문에 실업여부를 종속변수로 할 경우 추정치에 대한 왜곡된 해석의 우려가 있다. 따라서 이 연구에서는 취업여부만을 중심으로 방법효과를 살펴보고자 한다.

통제변수로는 취업에 영향을 미치는 이론적인 변수와 자기선택 오차 변수를 포함하였다. 취업에 영향을 미치는 변수로 성, 연령, 교육수준을 통제하였다. 일반적으로 여성이 남성에 비해 실업상태가 될 가능성이 높으며, 연령과 관련해서는 연령이 높은 사람이 연령이 낮은 사람에 비해 실업상태가 될 확률이 상대적으로 낮은 것으로 알려져 있다. 한편, 교육과 관련해서는 교육수준이 높은 사람일수록 직업 수행능력 및 생산성이 높다고 판단하여 교육수준이 낮은 사람에 비해서 취업가능성이 높은 것으로 알려져 있다(김순규, 2005). 성별은 남성은 1, 여성은 0으로 더미변수화했으며, 연령과 교육수준은 연속변수로 처리하였다.

이와 함께 자기선택 오차를 통제하였다. 자기선택 오차란 자료수집 방법이 무작위로 할당되지 않을 경우, 자료수집방법을 선택할 때 응답자의 개인적인 특성이 영향을 미침으로써 발생하는 오차를 말한다.

분석방법으로는 종속변수가 1과 0의 이분형 범주로 되어 있는 변수를 분석하기 적합한 로지스틱 회귀분석을 사용하였으며, SPSS 14.0 버전을 이용하였다. 2008년 자료를 이용하여 분석했으며, 제시되는 통계치는 가중치를 조정하기 전의 값이다.

2. 분석결과

가. 자료수집방법에 따른 고용률과 실업률의 차이

<표 5-6>은 대면조사와 비대면조사에서 고용률과 실업률을 제시한 것이다.¹²⁾ 전체적으로 볼 때 고용률은 대면조사로 수집된 자료에서 58.4%, 비대면조사로 수집된 자료에서 64.9%로 비대면조사로 수집된 자료에서 고용률이 더 높았다. 이는 성, 연령, 교육수준의 모든 하위집단에서 마찬가지였다. 표에는 신지 않았으나, 비대면조사를 전화조사와 자기기입식조사로 구분해서 살펴본 결과는 대면조사로 수집된 자료에서 고용률이 가장 낮게 나타났고, 다음으로 전화조사였으며, 자기기입식조사로 수집된 자료에서 고용률이 가장 높게 나타나 고용률에 있어서 전화조사는 대면조사와 자기기입식조사의 중간 정도의 성격을 갖고 있는 것으로 나타났다.

<표 5-6> 자료수집방법에 따른 고용률과 실업률: 2008년 단위

구분	고용률		실업률 (: %)	
	대면조사	비대면조사	대면조사	비대면조사
전체				
남성	58.4	64.9	2.8	3.1
여성	70.2	76.8	3.2	2.7
세	48.2	53.1	2.3	3.4
15-19세	5.7	6.5	10.6	15.8
20-24세	58.6	64.2	7.1	6.0
25-29세	71.1	74.6	3.1	2.8
30-34세	77.2	80.8	2.0	2.1
35-39세	70.0	72.0	1.8	2.7
40-44세	38.9	40.4	1.0	2.4
45-49세				
50세 이상	43.7	44.4	1.8	3.4
초대졸이하	62.3	66.7	3.5	3.6
초대졸이상	74.3	79.2	2.9	2.5

12) 고용률과 실업률은 각각 다음과 같이 계산되었다.
 — 고용률=(취업자/15세이상인구)*100, 실업률=(실업자/경제활동인구)*100

실업률은 대면조사와 비대면조사에서 고용률과는 다른 패턴을 보여 주었다. 전체적으로 볼 때 대면조사로 수집된 자료에서 실업률은 2.8%, 비대면조사로 수집된 자료에서 실업률은 3.1%로 비대면조사에서 약간 높았으나 이러한 패턴은 주요 하위집단에 따라서 차이가 있었다. 성별로 보면 남성에서는 대면조사로 수집된 자료에서 실업률이 더 높았으며, 여성에서는 비대면조사로 수집된 자료에서 실업률이 더 높았다. 연령별로 볼 경우에는 2-30대에서는 대면조사로 수집된 자료에서의 실업률이 비대면조사에 비해 높았으며 초대졸 이상에서도 대면조사로 수집된 자료에서 실업률이 비대면조사에 비해 높았다. 마찬가지로 표에는 신지 않았으나, 비대면조사를 전화조사와 자기기입식조사로 구분해서 살펴보았는데, 전화조사를 통해 수집된 자료에서 실업률이 가장 높았으며, 자기기입식조사를 통해 수집된 자료에서의 실업률이 가장 낮았다. 성, 연령, 교육수준별로 보더라도 자료수집방법에 따른 실업률의 패턴은 매우 유사하였다. 이는 전화조사와 자기기입식조사가 실업률에 있어서는 극단적인 성격을 띠고 있으며 이 두 집단을 비대면조사로 통합함으로써 비대면조사의 성격이 모호해졌을 가능성을 배제할 수 없음을 의미한다.

그러나 고용률에 있어서는 일관적으로 대면조사로 수집된 자료보다 비대면조사로 수집된 자료에서 고용률이 높게 나타나 대면조사와 비대면조사에 따른 차이의 가능성을 보여주었다. 이 결과를 통해 다음과 같은 질문이 제기될 수 있다. 혹시 비대면조사가 취업자와 관계가 있는 것은 아닐까? 즉, 취업자가 비대면조사를 선택할 가능성에서 비롯된 자기선택 오차 때문은 아닐까? 자료수집방법이 무작위로 할당되지 않았기 때문에 자기선택 오차의 개연성이 있다. 응답자의 개인적인 혹은 환경적인 특성에 따라서 비대면조사를 선택할 수 있기 때문이다. 이를 확인하기 위해 다음에서는 자기선택 오차를 통제한 이후에 자료수집방법이 취업자를 예측해내는 확률에 영향을 미치는지를 로지스틱 회귀분석을 통해 살펴보았다.

나. 취업예측확률에 대한 로지스틱 회귀분석

취업을 예측하는 데 자료수집방법이 미치는 효과를 분석하기 위해 1 단계로 자기선택 오차를 계산하여 변수화했으며, 2단계에서 자기선택 오차 변수를 통제한 후 자료수집방법효과를 살펴보았다.

자기선택 오차 변수를 만들기 위해 1차적으로 비대면조사 예측 모형을 구축한 후 이 모형 추정과정에서 생성되는 예측값(PRED_XB)을 이용하여 자기선택보정요인(self-selection correction factor)을 통계분석용 패키지에서 제공하는 변수계산프로그램을 이용하여 계산하였다(Harvey, 2006).

먼저, 자료수집방법을 선택하는 데 영향을 미치는 것으로 나타난 응답자 특성과 가구 특성인 성, 연령, 교육수준, 배우자유무, 농가여부 변수를 투입하여 비대면조사 예측모형을 만들었다(박영실·정남수, 2009).

$$\text{비대면조사} = B_1(\text{남성}) + B_2(\text{연령}) + B_3(\text{교육수준}) + B_4(\text{배우자있음}) \\ + B_5(\text{비농가}) + \text{상수}$$

위 모형에서 도출된 예측값을 이용하여, 자기선택 오차 변수인 람다(lambda)를 다음과 같은 공식으로 계산하였다.¹³⁾

$$\text{비대면조사}=1 \text{ 이면, } \lambda = \text{pdf}(\text{PRED_XB}) / \text{cdf}(\text{PRED_XB}) \\ \text{대면조사}=0 \text{ 이면, } \lambda = -\text{pdf}(\text{PRED_XB}) / (1 - \text{cdf}(\text{PRED_XB}))$$

여기서, pdf와 cdf는 예측값에 의해서 추정된, 표준정규분포의 확률밀도함수(probability density function)와 누적밀도함수(cumulative density function)이다.

다음으로 람다변수의 통제여부에 따라 2가지의 취업예측확률 모형을 구축하였다. 공통적으로 취업에 영향을 미치는 이론적인 변수로 성, 연령, 교육수준을 투입하였으며(김순규, 2005), 이 변수들을 통제한 후에 자료수집방법이 취업예측확률에 영향을 미치는지 살펴보았다(모형1). 모형2에는 모형1에 추가적으로 자기선택 오차인 람다변수를 통제하였다.

$$\text{모형1 : 취업} = B_1(\text{비대면조사}) + B_2(\text{성별}) + B_3(\text{연령}) \\ + B_4(\text{교육수준}) + \text{상수}$$

13) 이 부분에 대한 논의는 Harvey(2006)을 참고할 것.

$$\text{모형2 : 취업} = B_1(\text{비대면조사}) + B_2(\text{성별}) + B_3(\text{연령}) \\ + B_4(\text{교육수준}) + B_5(\text{람다}) + \text{상수}$$

람다변수는 자기선택 오차를 보정하기 위한 것으로 이 보정을 통해 자료수집방법이 취업에 미치는 효과가 편향되지 않을 수 있다. 람다변수가 통계적으로 유의미한지 여부에 따라 자기선택이 자료수집방법과 취업 사이의 관계에 영향을 미쳤는지를 결정할 수 있다. 만일, 람다변수가 정적인 방향으로 통계적으로 유의미하다면 취업자인 경우가 비대면조사를 선택할 가능성이 높다고 해석될 수 있으며, 실제 자기선택 오차를 보정하지 않은 분석결과에서 자료수집방법과 취업예측확률의 관계는 과대 추정되었음을 의미한다. 반대로 자기선택 오차가 통계적으로 유의미하되 그 방향이 부적이라면, 자기선택 오차가 보정되지 않은 분석에서 자료수집방법과 취업예측확률의 관계는 과소추정되었음을 뜻한다.

<표 5-7> 자료수집방법이 취업예측확률에 미치는 효과에 대한 로지스틱 회귀분석

구분	모형 1		모형 2	
	B (S.E.)	EXP(B)	B(S.E.)	EXP(B)
비대면조사				
람다	0.184(0.011) ***	1.202	-0.539(0.025) ***	0.583
주 는 임	-	-	0.008(0.000) ***	1.008

주 : 1) 4*모형에 성별, 연령, 교육수준을 통제하였음
2)

<표 5-7>은 취업예측확률에 자료수집방법이 미치는 효과를 분석한 결과로 모형 1은 자기선택 오차를 통제하기 전, 모형 2는 자기선택 오차를 통제한 후에 비대면조사 효과를 살펴본 것이다. 모형 1을 보면 대면조사에 비해 비대면조사에서 취업을 예측할 확률이 1.202배 증가하는 것으로 나타났다. 이것은 앞의 <표 5-6>에서 비대면조사를 수행한 표본에서의 고용률이 대면조사에 비해 높게 나온 것과 일치하는 결과이다. 그런데 자기선택 오차를 통제한 모형 2를 보면 비대면조사가 취업을 예측할 확률에 미치는 효과의 방향이 부(-)적으로 바뀌었음을 알 수 있다. 비대면조사에서 취업예측확률은 0.583배, 즉 비대면조사에서 취업예측확률이 대면조사에 비해 41.7% 감소하였다.

한편, 모형 2에서 람다변수의 계수는 95% 신뢰수준에서 통계적으로 유의미하였으며 그 방향이 정(+)적이었다. 즉, 자료수집방법과 취업자간에 자기선택 오차가 내재되어 있었으며, 취업자인 경우가 비대면조사를 선택할 확률이 높음을 알 수 있었다. 이는 조사원이 각 가구를 방문했을 경우, 취업자를 만날 가능성은 실업자나 비경제활동인구를 만날 가능성보다 낮기 때문인 것과 관계 된다. 따라서 취업자는 대면조사보다는 그 외의 대안적인 방법인 전화조사나 인터넷조사를 포함한 자기기입식조사를 통해 응답할 가능성이 높다.

다. 취업예측확률 차이 발생 원인

비대면조사와 취업자간의 자기선택 오차를 통제한 이후에 두 변수간의 관계를 살펴본 결과 비대면조사에서 취업을 예측할 확률은 대면조사에 비해 줄어드는 것으로 나타났다. 이것은 실제 조사현장과 일치하는 결과이다. 조사원들과의 인터뷰에 따르면, 가구를 방문하여 면접을 하는 경우, 응답자 자신은 취업자라고 생각하지 않으나 면접과정에서 취업자로 포착되는 경우가 많이 발생한다고 한다. 그렇다면, 자료수집방법에 따라서 이러한 차이가 발생하는 원인은 무엇인가? 딜만(Dillman, 2009)은 자료수집방법에 따른 조사원의 유무, 조사항목의 워딩이나 구성에 따라서 응답이 달라질 수 있다고 보고 있다. 또한 시각적인 설계와 레이아웃이 응답에 영향을 미칠 수도 있다. 경제활동인구조사의 경우, 모든 자료수집방법에서 동일한 조사항목과 워딩을 사용하였으므로 조사원에 의한 효과가 가장 큰 영향을 미친 것으로 보인다.

우리나라는 고맥락문화(high context)로 알려져 있다. 고맥락문화란 커뮤니케이션이 맥락적인 요소, 배경정보, 상호간의 큐(cue) 등에 의존하는 경향을 말하는 것이다. 반면에 저맥락(low context) 문화는 커뮤니케이션이 사실과 정보에 기반한다. 즉, 해당 정보가 언제, 어떻게 누구에 의해서 표현된다기보다는 메시지의 실제 내용이 더 중요한 것이다 (Hall and Hall, 1987, Hyunjoo and Sha, 2009에서 재인용). 이러한 문화적인 맥락은 우리나라의 조사체계에서 잘 드러나 있다. 조사원과 응답자간의 면접과정 중에서 부가적으로 파악되는 정보가 실제 조사결과에

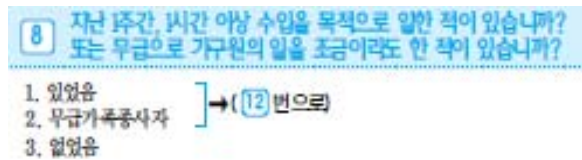
영향을 미치는 경우를 쉽게 접할 수 있다. 조사원의 캐어묻기(probing) 여부에 따라서 취업자와 실업자 및 비경제활동인구가 달라질 수 있는 것은 바로 이러한 문화적인 맥락이 반영된 것이다. 쉬운 예로, '지난 1주간 무엇을 하였습니다습니까?'라는 질문의 응답범주에 발령대기가 있는데, 이 경우 조사대상주간으로부터 1개월 이내에 발령이 난다고 하면 실업자로, 1개월 이상이면 비경제활동인구로 분류되는데, 조사표에는 이에 대한 내용이 전혀 언급되어 있지 않기 때문에 조사원이 없는 자기기입식조사에서는 이를 구분해 낼 수 없다. 동일한 조사표를 사용하여 조사가 진행된다고 할지라도 응답자와의 직접적인 접촉을 통해서 조사원이 면접을 통제할 수 있는 경우와 그렇지 않은 경우 사이에 이끌어낼 수 있는 정보의 양에는 차이가 발생할 수밖에 없다.

또한 자료수집방법에 따라 취업예측확률에서 차이가 나는 것은 조사표에서 정의하고 있는 취업자 개념과 응답자가 생각하는 취업자 개념의 불일치에서 기인하기도 한다. 경제활동인구조사에서 취업자에는 '① 조사대상주간에 수입을 목적으로 1시간 이상 일을 한 자, ② 동일 가구 내 가구원이 운영하는 농장이나 사업체의 수입을 위하여 주당 18시간 이상 일한 무급가족종사자, 그리고 ③ 직업 또는 사업체를 가지고 있으나 일시적인 병 또는 사고, 연가, 교육, 노사분규 등의 사유로 일을 하지 못한 일시휴직자가 포함된다.

각각은 다음의 조사항목으로 구체화되었다. [그림 5-6]은 지난 1주간 수입을 목적으로 1시간 이상 일한 자를 분류해 내기 위한 항목인데, 실제 이 항목은 응답자에게 매우 낯설게 느껴지는 것으로 나타났다. 우선 '지난 1주간'은 응답자에게 일반적이지 않은 개념이다. 우리나라 사람들은 1주일이라는 시간관념에 익숙하지 못하여 1주일에 적어도 한 시간 이상 수입을 목적으로 일한 경우를 취업으로 분류하는 조사의 취지를 이해하지 못한 경우가 많은 것으로 보고되고 있다(통계청, 2008). 조사대상자들과의 심층면접 결과, '지난주'라는 질문에 '이상하다', '애매하다'라는 반응을 보였으며, 이 표현이 '이제 지금은 다른 회사에 다닌다는 느낌을 준다', '지난주라고 하면 일을 하는 주가 있고 하지 않는 주가 있는 것 같다'라는 의견이 제시되었다.¹⁴⁾ 따라서 지난 1주간에 1시간 이

14) 경제활동인구조사 근로형태별 부가조사표 개선을 위한 인지면접 녹취 자료 참고.

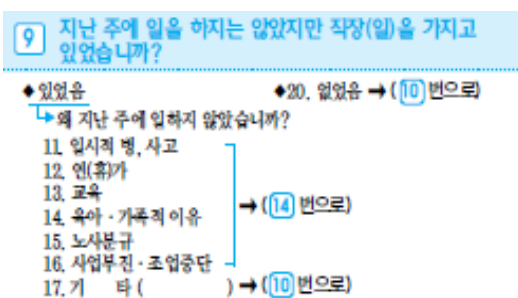
상 수입을 목적으로 일한 경우를 포착하는 것에 어려움이 생길 수 있다.



[그림 5-6] 지난 1주간 일한 자 및 무급가족종사자 구분 항목

[그림 5-6]에서 는 지난 1주간 일한 자와 함께 무급가족종사자를 묻고 있는데, '무급가족종사자'라는 말을 이해하지 못하는

계층의 경우, 조사원의 도움 없이 이 항목에 응답할 수가 없기 때문에 취업자가 아닌 실업자나 비경제활동인구로 분류될 가능성이 높다. 또한 두 가지 질문을 한 번에 함으로써 두 유형에 모두 해당하는 응답자의 경우 선택을 어렵게 만드는 경향이 있다. 18시간 미만의 무급가족종사자가 지난주에 수입 있는 일을 했을 경우를 생각해 보자. 이 사람은 지난주에 수입 있는 일을 했기 때문에 취업자이나 평소 무급으로 가구원의 일을 계속 해 왔기 때문에 무급가족종사자로 응답할 가능성이 높다. 그러나 무급가족종사로 일한 시간이 18시간 미만이기 때문에 다시 실업자나 비경제활동인구를 구분하는 항목으로 되돌아가게 된다. 따라서 실제 취업자로 포착되어야 함에도 불구하고 실업자나 비경제활동인구가 된다.



[그림 5-7] 일시휴직자 구분 항목
직자로 간주돼 취업자가 되나, 17번인 기타일 경우에는 실업자나 비경제활동인구로 포착된다.([그림 5-7]) 그런데 응답자와의 직접 접촉을 통해 면접을 하는 경우 캐어문기를 통해 기타 범주가 11번부터 16번의 응답범

일시휴직자를 구분해 내는 항목의 경우에는 다음과 같은 한계점을 갖고 있다. 지난주에 일을 하지는 않았지만 직장(일)을 가지고 있었느냐는 질문에 있었음이라고 하고 그 이유가 11번부터 16번까지일 경우에는 일시휴

주로 재분류될 가능성이 있다. 이는 취업자 예측에 영향을 미친다.

응답자와 조사자 간에 개념이 일치하지 않는 상황에서 응답자 판단 하에 자신을 실업자로 규정짓고 바로 바로 실업자 항목으로 이동할 수 있는데 이는 부분적으로 조사표 구성에 의해서도 영향을 받고 있는 것으로 판단된다.

[그림 5-8] 경제활동인구조사표

[그림 5-8]에서 보듯이 조사표에는 짧은 글씨로 취업자, 실업자, 비경제활동인구가 구분되어 있기 때문에 자신의 규정에 따라서 바로 해당 항목만 응답할 개연성이 높다. 대면조사에서는 조사원에 의해 응답자의 이런 행동이 통제될 수 있으나 비대면조사에서는 그럴 가능성이 매우 낮다. 물론 인터넷조사의 경우 조사표의 처음부터 응답을 하지 않으면 다음 항목으로 넘어가지 않기 때문에 이와 같은 가능성은 없다는 반론이 제기될 수 있으나, 인터넷조사의 비율은 매우 낮다. 인터넷조사가 도입되기 이전의 자기기입식조사 비율(1.5%)과 인터넷조사가 도입된 이후 인터넷조사를 포함한 자기기입식조사 비율(2.0%)의 차이는 0.5% 정도이다.

요약해 보면, 경제활동인구조사의 조사체계가 대면조사방식이었으나 최근에 인터넷조사나 CATI 등의 비대면조사방법이 도입되고 있는 가운데 자료수집방법에 따라 취업예측확률에 차이가 발생할 수 있는 것으로 분석되었다. 아직까지는 비대면조사의 비중이 높지 않은 편이나 향후 그 업무량이 증가될 것으로 예상되는 가운데 이에 대한 대안마련이 필요할 것으로 보인다.

제4절 결론

1. 결과 요약 및 시사점

응답률 하락 및 조사비용 증가 등에 대한 대안적인 방법으로 혼합방법조사가 증가하고 있으나, 동시에 이 조사에서 자료수집방법에 따른 비표본오차가 발생할 가능성이 있다는 문제가 지속적으로 제기되어 오고 있다. 이 때문에 자료수집방법을 변경하는 과정에서 방법효과를 검토하는 일은 중요하다. 방법효과가 나타날 경우, 이에 대한 원인을 찾고 그에 적절한 대안이 마련되어야 한다.

이 연구에서는 경제활동인구조사 자료를 이용해서 방법효과를 검토해 보았다. 경제활동인구조사는 시작된 이래로 대면조사를 원칙으로 하고 있는데 최근 대안적인 방법으로 전화조사나 자기기입식조사 등의 활

용이 꾸준히 증가하고 있다. 아직까지 그 비중이 높지는 않으나 앞으로 대안적인 방법의 활용이 증가할 것으로 예상되는 가운데, 이 연구는 그에 앞서 현재 수집된 자료를 이용하여 방법효과에 대한 분석을 실시해 봄으로써 차후 자료수집방법 개편 시 고려할 점에 대한 시사점을 얻고자 하였다. 연구내용은 해외의 노동력조사에 대한 사례연구와 경제활동 인구조사 자료의 방법효과 분석 두 가지이다.

가. 사례연구 결과 및 시사점

사례연구를 통해 파악된 주요 연구결과와 시사점은 다음과 같다. 첫째, 미국이나 캐나다, 영국, 호주 등의 노동력조사에서는 1회차에는 대면조사를, 2회차 이후에는 응답자의 동의하에 상당수의 전화조사가 이루어지고 있었다. 이는 응답회차와 무관하게 대면조사를 원칙으로 하고 있는 우리나라와 큰 차이를 보이는 부분이다. 검토한 해외사례에 따르면 표본으로 선정된 가구는 8회차를 초과하여 응답하지 않고 있는 상황에서 2회차 이후에는 전화조사를 하고 있으나, 36회차를 응답하는 우리나라에서 전 회차에 걸쳐 대면조사를 원칙으로 하고 있으며, 36회차까지 응답하는 동안 대면조사로 수행되는 비율이 90% 이상으로 나타났는데 그 결과의 신뢰성에 대해서 검토해 볼 필요가 있다. 자료수집방법에 대한 충분한 사례연구를 통해서 조사현실에 적합한 원칙이 마련되어야 할 것이다.

둘째, 전화조사는 현장조사를 담당했던 조사원에 의해 개별적으로 이루어지는 경우와 중앙집중화된 면접시설에서 관리자의 감독 하에 컴퓨터를 이용한 면접이 이루어지는 경우가 있었으며 미국이나 캐나다, 영국 등에서는 후자를 일컬어 CATI라고 하였다. 우리나라에서 실시하고 있는 CATI는 해당 표본의 담당 조사원이 개별적으로 각각의 사무소에서 컴퓨터 모니터 상에 나타나는 조사표를 보면서 전화면접하는 것으로 해외사례에서 살펴본 CATI와는 차이가 있다. 해외사례를 보면, 전화조사는 조사원의 이동거리 감소에 따른 비용효과 기대에서 출발하였으며 특히, CATI는 조사원의 오차를 줄여줌으로써 자료의 품질을 향상시키고자 하는 목적에서 도입되었다. 그렇다면, 우리나라의 경우 CATI가

기존의 전화조사에 비해 부가적으로 발생시키는 긍정적인 효과가 무엇인지, 조사원의 업무량이나 비용, 자료의 품질 측면에서 검토가 이루어져야 할 것이다.

셋째, 해외사례에 따르면 자료수집방법을 변경하는 과정에서 응답률이나 추정치에서의 변화가 수반되었으며 각 국가는 그 변화에 따라 발생하는 시계열 단절을 극복하기 위한 대안을 마련하는 데 많은 시간과 노력을 투자하였다. 미국은 CATI 도입 과정에서 CATI 운영상의 문제와 실업률이나 고용률 등 추정치 변화에 따른 시계열 단절 문제를 극복하기 위해 전국적인 차원에서 8년여 동안 몇 차례의 실험을 실시하였다. 이 경험적인 자료를 기반으로 차이가 발생하는 원인에 대해 다각적인 분석을 실시, 시계열 단절을 보정할 수 있는 보정요인을 개발하여 자료이용자에게 제공하였다. 본 연구에 따르면 대안적인 자료수집방법의 증가가 예상되며, 동시에 자료수집방법에 따른 차별적인 효과가 예상되는 바, 보정요인에 대한 개발은 향후 우리가 염두에 두어야 할 부분이다.

통계조사에서 자료수집방법변경 등을 결정함에 있어서 이와 같은 절차와 자료를 근거로 방향제시가 된다면 도입과정에서 발생할 수 있는 시행착오는 줄일 수 있을 것이다. 이러한 절차들이 신뢰성 있는 통계자료를 생산하는 데 기여할 것으로 생각된다. 특히, 고맥락 문화권에 속한 우리나라의 경우, 자료수집에서 조사원에 의한 효과가 크고, 지역에 따라서 차별적일 가능성이 있으므로 전국적인 차원에서의 검토가 중요할 것으로 생각된다.

나. 자료분석 결과 및 시사점

경제활동인구조사 자료의 분석을 통해서 다음과 같은 결과가 도출되었다. 자료수집방법이 주요 통계치에 영향을 미치는지를 자기선택 오차를 통제한 이후에 살펴본 결과 대면조사보다는 비대면조사에서 취업률 예측할 확률이 줄어들었다. 조사원이 응답자를 만나서 조사할 경우, 취업자로 포착해 내는 경우가 많다는 것이다. 조사자가 정의하는 취업자 개념과 응답자가 생각하는 취업자 개념이 일치하지 않는 상황에서, 조사체계는 조사원과 응답자 간의 상호작용에 의존해 있으므로 자료수

집방법에 따라 응답자를 분류해내는 데 차별적일 수밖에 없다.

따라서 조사자가 정의하고 있는 개념과 응답자가 생각하는 개념을 일치시킬 필요가 있다. 이를 위해 많이 활용되고 있는 방법으로 인지적인 접근법이 있다. 조사방법론 분야에서는 30여년 전부터 이 접근법이 활용됨으로써 응답자의 이해나 기억처리, 판단과정 등에 대한 이해도를 높이고 있다(Tourangeau, Rasinski, and Rips, 2000). 그동안 조사표가 연구자나 조사원 중심으로 설계되어 왔다면 여기에 응답자를 포함시킴으로써 응답자가 발생시킬 수 있는 응답오차를 최소화시킬 수 있다. 연구자나 조사자의 입장에서 설계된 조사표를 응답자가 제대로 이해하기란 쉬운 일이 아니며, 응답된 내용 또한 정확성을 담보할 수도 없을 것이다. 조사원의 참여도가 높은 대면조사와 달리 조사원과 응답자간의 상호작용이 배제되는 컴퓨터 방식의 조사나 자기기입식조사에서 응답자 친화적인 조사표의 중요성이 더욱 부각된다.

이와 연관되어, 조사원과 응답자간의 상호작용에 따라서 조사결과가 달라지는 것을 줄이기 위해 조사표에는 조사되어야 할 내용을 모두 포함시켜야 한다. 조사되어야 할 내용이 조사표가 아닌 조사원 교육과정에 전달된다면 이는 조사원에게 조사부담을 주는 일인 동시에, 조사원 뿐 아니라 자료수집방법(조사원기입식 vs 자기기입식)에 따라서 조사결과가 달라지는 원인이 될 것이다.

2. 자료의 한계 및 향후 과제

이 연구의 한계는 다음과 같으므로 연구결과 해석시 주의해야 할 것이다. 첫째, 자료수집방법을 구분하는 기준의 모호함으로 인해 일부 표본으로부터 자료수집방법의 정보를 얻을 수 없었다. 이후 조사표 개선 시에는 자료수집방법 구분 항목을 재설계해야 할 것이다. 자료수집방법과 응답주체를 구분해야 하고, 컴퓨터 사용여부 또한 기준으로 제시하여야 할 것이다.

둘째, 자료수집방법효과를 살펴보기 위해서는 자료수집방법 이외에 다른 요인은 모두 동일해야 하나 실험 상황에서 자료가 수집된 것이 아니므로 이에 대한 통제가 이루어지지 않았다. 예컨대, 컴퓨터를 이용한

방법의 경우 응답순서가 통제되는 반면에 대면조사의 경우 조사항목이나 응답범주의 순서가 조사원과 응답자의 상황에 따라서 바뀔 수 있다.

셋째, 자기선택 오차 변수를 계산하는 데 필요한 변수가 충분치 않았다. 자기선택 오차는 자료수집방법 결정 모델에서 도출되는 예측값을 이용하여 계산된다. 그런데 자료수집방법을 선택하는 데 있어서 개인적인 특성과 함께 가구의 특성 또한 중요한데 경제활동인구조사에서 가구 특성에 관한 변수가 충분치 않았다.

넷째, 대면조사 이외의 방법으로 조사된 비율이 매우 낮아 분석과정에서 전화조사와 자기기입식조사를 통합하였는데, 이는 각 자료수집방법이 갖는 고유한 특성에 대한 해석에 제약을 가하였다.

이러한 한계점들을 극복할 수 있는 방안은 방법효과를 검증하기 위해 확실적인 표본추출 절차에 근거한 시험조사를 하는 것이다. 해당 표본에 자료수집방법을 무작위로 할당한 후 동일한 조사절차를 통해서 자료를 수집하여 그 결과를 분석한다면 위에서 제기한 자료 및 분석 상의 한계점은 상당 부분 극복될 수 있을 것으로 보인다. 자료수집방법을 포함한 조사체계의 변경 과정에서 이와 같은 실험에 근거한 결정을 할 필요가 있을 것이다.

또한, 자료수집방법에 대한 연구와 함께 조사표 개선에 대한 연구가 동시에 이루어져야 할 것이다. 조사원과 응답자간의 상호작용에 의존하는 조사체계와 실제 현장과의 차이가 발생하고 있었다. 비록 그 비율이 작기는 하지만 조사현장에서는 자기기입식조사의 비율이 꾸준히 증가하고 있음을 자료를 통해서 확인할 수 있었다. 이 과정에서 조사표에 의해서 발생할 수 있는 측정오차의 개연성이 매우 크므로 향후 자료수집방법의 변경과 동시에 조사표에 대한 개선이 진행되어야 할 것이다.

참고문헌

- 김순규 (2005), 실업급여 수급권자의 재취업에 영향을 미치는 요인 사회 복지정책 22: pp.83-97.
- 박영실 · 정남수 (2009), 인구주택총조사 자료수집방법 비교분석 대면 · 인터넷 · 우편조사를 중심으로 통계개발원
- 성태제 (2007), SPSS/AMOS 를 이용한 알기쉬운 통계분석 기술통계에서 구조방정식모형까지 학지사
- 통계청 (2008), 경제활동인구조사 지침서
- ABS(2007), *Labour Statistics: Concepts, Sources and Methods*.
- 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.
- Biemer, P. and L. Lyberg(2003), *Introduction to Survey Quality*. Wiley-Interscience.
- Bregger, J. E. and C. S. Dipppo(1993), "Overhauling the Current Population Survey: Why is it necessary to change?," *Monthly Labour Review*: pp.1-9.
- de Leeuw, E. D. and J. van der Zouwen(1988), "Data Quality in Telephone and Face to face Surveys: a Comparative Meta Analysis," pp.283-299 in R. M. Groves(eds.), *Telephone Survey Methodology*, John Wiley and Sons, Inc.
- Dillman, D.(2009), "To Mix or Not to Mix Survey Modes," presented at 2009 AAPOR Conference.
- 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.
- Dippo, C., D. Kostanich, A., and Polivka(1994), "Effects of Methodological Change

- in the Current Population Survey,” <http://www.bls.gov/osmr/pdf/cp940030.pdf>.
- Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau(2004), *Survey Methodology*, John Wiley and Sons, Inc.
- Hall, E. T. and M. R. Hall(1987), *Hidden Differences: Doing Business with the Japanese*, New York: Anchor Books/Doubleday.
- Harvey, S.(2006), “Self-Selection Bias in Business Ethics Research,” *Business Ethics Quarterly*16(4): pp.559-577.
- 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.
- Hyunjoo P. and M. Sha(2009), “Cognitive Testing as a Method of Pre-testing Questionnaires in High and Low Context Cultures: a Comparison of Korean and English Cognitives Interview Data,” presented at 2009 AAPOR Conference.
- Kostanich, D. L. and L. S. Cahoon(1994), *Effect of Design Differences Between the Parrallel Survey and the New CPS*, CPS Bridge Team Technical Report 3.
- Miller, S. M.(1994), *What Would the Unemployment Rate Have Been Had the Redesigned Current Population Survey Been in Place from September 1992 to December 1993?: A Measurement Error Analysis*, CPS Bridge Team Technical Report 1.
- ONS(2007), *Labour Force Survey User Guide: Background and Methodology*.
- Polivka, A. E.(1994), *Comparisons of Labour Force Estimates from the Parallel Survey and the CPS during 1993: Major Labour Force Estimates*, CPS Overlap Analysis Team Technical Report 1.
- Polivka, A. E. and S. M. Miller(1994), “The CPS After the Redesign: Refocusing the Economic Lens”, presented at CRIW Labourstatistics Measurement Issues Conference.
- Shoemaker, H. H.(1993), “Results from the Current Population Survey CATI Phase-in Project”, *Proceedings of the Survey Research Methods*

Section, ASA.

Statistics Canada(2008), *Methodology of the Canadian Labour Force Survey.*

Statistics Canada(2009), *Guide to the Labour Force Survey.*

Thompson, J.(1994), *Mode Effects Analysis of Labour Force Estimates*, CPS
Overlap Analysis Team Technical Report 3.

Tourangeau, R., Rasinski, K., and Rips, L.(2000), *The Psychology of Survey
Response*, NewYork, NY: Cambridge University Press.

U. S. Census Bureau(2006), *Design and Methodology: Current Population
Survey*. Technical Paper 66.

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.