

제3장

회귀분석을 이용한 Imputation 기법활용 연구

- 사회통계조사 항목무응답을 중심으로 -

이 내 성

제1절 서 론

1. 연구배경 및 목적

오늘날 현대사회는 도덕이나 법률, 사회제도와 같은 사회구조와 인간
간의 습관이나 태도, 환경에 대응하는 방식 등 광범위한 영역에서 빠른
속도로 변화하고 있다. 같은 사회 안에서도 가치관과 생활양식이 다양
한 행태로 나타나고 있다. 이와 같이 급변하는 사회변화에 대응하기 위
해 통계청에서는 가족, 보건, 복지, 주거와 교통 등 각 부문에 대한 국민
들의 주관적 의식, 사회적 관심사, 삶의 질에 관련된 항목들을 조사하여
삶의 수준과 사회상태의 변동을 파악하고 이를 사회개발정책에 기초자
료로 제공하기 위하여 사회통계조사를 실시하고 있다.

그런데 우리가 통계조사에서 직면하는 문제들 중에 하나가 표본으로
선정된 모든 사람들로부터 전체 설문문항에 대해 유용한 자료를 얻을
수 없다는 것이다. 외국의 응답률을 살펴보면 일본 51%(2005년), 미국
70%(2004년), 영국 74%(2006년), 캐나다 80%(2006)로 나타난다. 유럽의
사회조사(the European Social Survey)에서는 조사에 참여하는 국가들에

대한 기대응답률을 70% 정도로 보고 있다(Loosveldt, Carton, and Billiet, 2004). 우리나라의 사회통계조사 응답률은 95%(2007년)로 매우 높게 나타난다.

이는 현재 사회통계조사가 무응답을 허용하고 있지 않는 시스템 구조로 되어 있기 때문에 나타나는 현상으로 볼 수 있다. 사회통계조사와 동일한 조사대상가구에 대해 실시된 2000년 가구소비실태조사(82%)와 외국의 사례를 보면 향후 우리나라의 사회통계조사도 무응답에 필연적으로 직면하게 될 것이라는 점을 시사하고 있다.

무응답 및 결측자료는 정보의 손실, 표본탈락, 또는 설문문항에 대한 무응답 등으로 인해 발생한다. 무응답 및 결측값을 가진 자료를 분석할 때 우리는 보편적으로 적절한 가정하에(예, Missing Completely at Random: Rubin, 1976) 불완전한 자료는 무시하고 완전하게 관측된 자료만을 분석한다. 이는 분석이 간편하다는 장점이 있는 반면에 가정이 맞지 않을 경우에는 통계적 추론이 틀릴 수 있다는 문제가 있다.

즉, 무응답이 생기면 원래 의도했던 표본의 크기를 축소시키고, 그에 따라 통계자료 이용의 효율성을 떨어뜨릴 뿐만 아니라 완전한 자료를 바탕으로 하는 표준적인 통계적 방법을 사용하는 데 있어서도 평균추정량, 분산추정량 등 모집단에 대한 여러 가지 추정결과에 편향(bias)된 통계적 추론을 구하게 되어 문제가 발생한다.

[그림 3-1] 커피향과 각설탕



* 커피에 각설탕이 더해지면 향긋한 커피향에 더욱 감미로운 맛을 즐길 수 있듯, 무응답 항목에 대해 imputation 기법 등을 활용한 기술적 보완은 통계자료 이용의 효율성을 높인다.

본 연구의 목적은 첫째, 수집된 자료에 남아 있는 무응답 및 결측값을 Imputation기법을 활용하여 효율적으로 처리함으로써 통계자료의 유용성을 높이는 데 있다. 둘째, 사회통계조사를 중심으로 Imputation을 처리단계별로 기술함으로써 실무에 직접적으로 적용이 가능하도록 한다. 셋째, 적합한 Imputation기법을 선택할 수 있도록 평균대체법, 최근방대체법과 비교하여 제시한다. 넷째, Imputation 적용에 따른 효과평가를 통해 그 중요성을 재인식시킨다.

본 연구의 의의는 항목무응답 대체에 대해 표준적인 지침을 마련함으로써 사회통계조사의 정확성과 효율성을 높이고 신뢰도가 높은 통계자료생산을 위한 기반을 제공할 수 있다. 따라서 1인 가구, 맞벌이가구 및 노인인구 증가와 개인의 비밀보호 의식의 강화 등 사회변화로 인한 낮은 응답률로부터 표본조사의 정도 저하를 줄일 수 있어 통계자료의 유용성을 높일 수 있다. 또한 인구주택총조사뿐만 아니라 가계조사, 가계자산조사 등 가구조사에 대한 Imputation기법 선행연구로서 활용할 수 있을 것이다.

2. 연구범위

가. 2007 사회통계조사

본 연구에서는 조사 후에 마지막으로 남아 있는 무응답 및 결측값을 채워 넣는(fill-in, impute, replace) 방법인 Imputation에 대해 2007 사회통계조사를 중심으로 적용·사례연구를 한다.

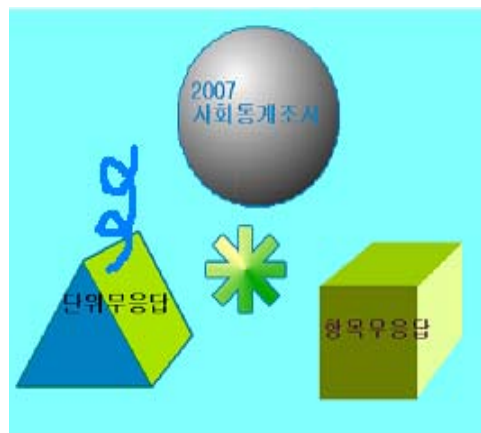
본 연구에서 이용한 통계자료는 내검 전 자료이기 때문에 내검단계를 거쳐 완결된 최종 자료인 2007 사회통계조사 공표자료와 다소 차이가 있다.

나. 항목무응답

무응답은 단위무응답(unit nonresponse)과 항목무응답(item nonresponse)으로 대별할 수 있다. 단위무응답은 여러 개의 조사항목으로 구성되어 있는 조사단위 전체에 대해 응답하지 않은 경우를 일컫는다. 이와 달리

항목무응답은 조사단위에는 응답을 하였으나 그 조사단위를 구성하고 있는 응답항목들 중에서 몇 개의 항목에 대해 응답을 하지 않은 경우를 일컫는다.

[그림 3-2] 연구범위



* 2007 사회통계조사 조사결과를 활용하여 항목무응답에 대한 Imputation 적용사례 연구를 한다.

본 연구는 항목무응답에 대한 Imputation에 한정한다. 결측값은 사실 전혀 모르는 값이기 때문에 어떤 값을 대체값으로 결정하느냐 하는 문제보다는 대체값을 어떤 합리적인 방법으로 구하느냐에 초점을 맞추는 것이 최선의 방법일 것이다. 따라서 사회통계조사 항목무응답을 중심으로 대체처리단계별로 기술한다.

3. 연구방법

Greenless(1982)와 David(1982)는 미국 경상인구조사(Current Population Survey)에서 발생하는 결측값에 회귀대체를 이용하고 그 결과를 분석하면서, 대체된 값과 실제값의 평균절대편차(mean absolute deviation)를 비교할 때 다른 대체방법에 비해 회귀대체가 매우 적합함을 보인다고 논하고 있다.¹⁾

1) 조사통계연구회, 「무응답오차」, 자유아카데미, 2000, p.88.

본 연구에서는 이에 따라 회귀대체법을 중심으로 기술한다. 대체결과와 신뢰성 확보와 타당성 검토를 위해 응답자별로 무응답처리 결과와 실제값의 일치율을 비교한다. 또한 회귀대체기법의 효율성 검증은 위해 회귀대체법과 같은 단일대체법의 하나인 평균대체법 및 최근방대체법과 비교한다.

4. 자료분석 도구 및 절차

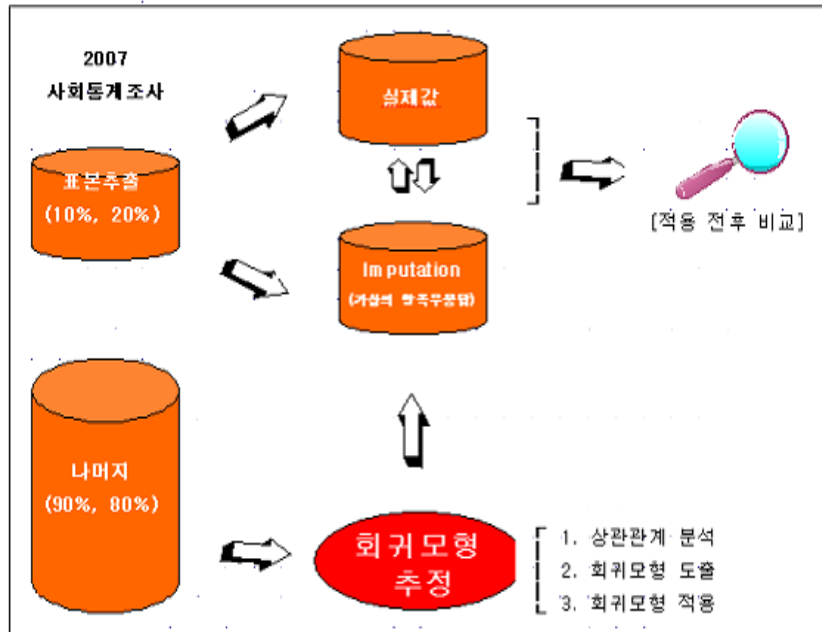
가. 자료분석 도구 : sas 9.1을 사용

나. 자료분석 절차

자료분석은 다음과 같은 단계로 진행한다.

- (1) 2007 사회통계조사에서 조사하기 가장 어렵고 힘들었던 조사항목에 대해 3개 지방통계청(서울, 부산, 광주전남) 조사팀장을 통해 조사담당자의 의견 조사
- (2) (1)을 바탕으로 가상의 무응답항목 결정
- (3) 항목무응답/결측값이 무응답자와 응답자에 따라 특성이 변수간에 다른지, 또는 이용할 수 있는 정보가 있는지 알기 위하여 무응답되기 쉬운 변수와 관련이 있는 변수들을 통계적으로 조사
- (4) 표본추출(10%, 20%)한 후 가상의 결측값을 만들
- (5) 표본추출 자료를 제외한 자료(90%, 80%)로부터 기본항목과 조사항목을 설명변수로 예측력을 높일 수 있는 모형을 탐색
- (6) 이 모형하에서 응답자료(90%, 80%)를 바탕으로 회귀계수를 추정하여 자료의 특성에 맞는 대체회귀모형을 도출
- (7) 가상으로 만들어진 결측값을 가진 표본추출된 자료(10%, 20%)에 도출된 항목무응답 대체회귀모형을 적용하여 Imputation
- (8) Imputation값과 실제값을 비교 평가
- (9) 회귀대체와 같은 단일대체법의 하나인 평균대체 및 최근방대체 결과와 비교 평가

[그림 3-3] 자료처리 흐름도



제2절 우리나라의 사회통계조사와 조사환경

1. 우리나라의 사회통계조사

사회통계조사는 통계법 제4조 1항 및 제10조에 의해 승인된 지정통계(승인번호 10118호)로 1977년 3월 「한국의 사회지표」 체계구성을 위한 기초자료 수집목적으로 소득·소비, 교육, 보건, 주택·환경, 사회에 관한 사항을 조사하였다. 1978년에는 사회지표체계를 조사통계국과 한국개발연구원 공동연구로 수립하였으며, 1979년부터 1984년까지 총 8개 부문 중 매년 4~5개 부문을 선정하고 부문당 평균 5개 항목을 조사하였다. 1987년에는 1차 「한국의 사회지표」 체계 개편이 있었다.

1985부터 1996년까지는 심층조사를 위해 조사부문을 2~3개 부문으로 축소하는 대신 부문당 항목수를 확대하여 조사를 실시하였고, 1995

년 2차 「한국의 사회지표」 체계 개편이 있었다. 1997년부터 연 2회(4월, 9월) 조사로 변경(각 2개 부문 조사)하였으며, 1998년에는 연 1회 조사로 환원(3개 부문 조사)하였다. 2006년부터는 매년 3~4개 부문씩 부문별로 3년 주기로 조사주기를 전환하였다.

연도별 조사실시시기 및 조사부문을 살펴보면 다음과 같다.,

1996. 9.15 ~ 9.24	: 문화와 여가, 교육
1997. 4.20 ~ 4.29	: 정보와 통신, 안전
1997. 9.21 ~ 9.30	: 주거와 교통, 환경
1998.10.18 ~ 10.27	: 가족, 복지, 노동
1999.10.17 ~ 10.26	: 보건, 사회참여, 소득과 소비
2000. 7.16 ~ 7.25	: 문화와 여가, 교육, 정보와 통신
2001. 9.16 ~ 9.25	: 주거와 교통, 환경, 안전
2002. 9.23 ~ 10. 2	: 가족, 복지, 노동
2003. 9.21 ~ 9.30	: 보건, 사회참여, 소득과 소비
2004. 6.20 ~ 6.29	: 주거와 교통, 문화와 여가, 교육
2005. 6.19 ~ 6.28	: 복지, 안전, 환경
2006. 7.16 ~ 7.25	: 가족, 보건, 사회참여, 노동
2007. 6.17 ~ 6.27	: 복지, 문화와 여가, 소득과 소비

2. 조사환경²⁾

현재 사회통계조사는 전국 약 33,000 표본가구 내의 만 15세 이상 가구원을 조사대상으로 하며, 확률표본을 통한 면접조사 방식을 택하고 있다.

다변하고 복잡한 현대사회는 기존의 사회통계조사 전반에 대해 재고의 필요성을 요구하고 있다. 사회통계조사에 변화를 요구하고 있는 조사환경은 크게 첫 번째는 1인 가구, 맞벌이가구 및 노인인구 증가 등 조사환경의 악화, 두 번째는 표본변경에 따른 변화, 세 번째는 사회통계조사의 확대요구로 살펴볼 수 있다. 이러한 조사환경은 조사상황에 영향을 미침으로써 무응답의 발생요인으로 작용한다.

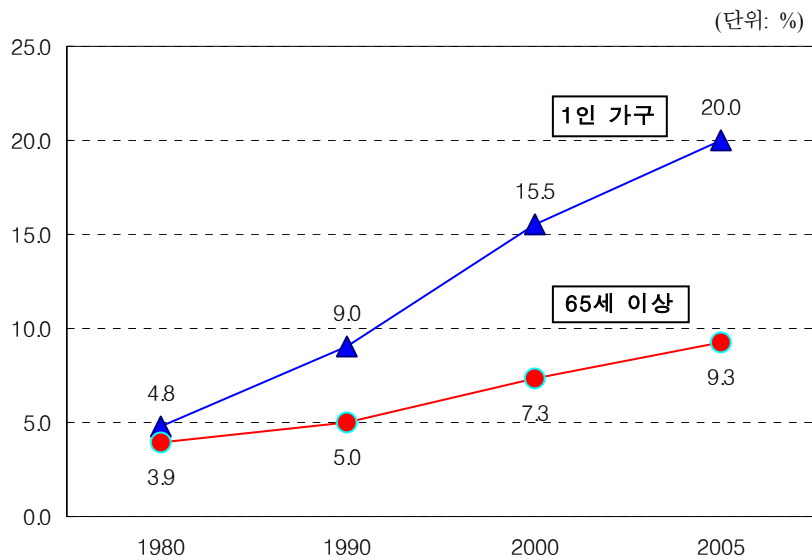
2) 사회통계조사 무응답층에 대한 대체기법 개발 관련 실무자 회의(2007.4.11).

가. 1인 가구 및 노인인구 증가 등 조사환경의 악화

사회통계조사는 가능한 모든 조사대상자가 응답하기 쉽도록 문항을 설계하고 있으나 1인 가구, 맞벌이가구, 노인인구의 증가 등으로 조사의 어려움이 가중되고 있다. 1인 가구의 경우 2000년에 222만 가구로 전체의 15.5%를 차지하였으나 2005년에는 95만 가구가 늘어난 317만 가구로 전체의 20.0%를 차지하고 있다. 또한 65세 이상 인구의 경우 2000년에 337만 명으로 전체의 7.3%를 차지하였으나 2005년에는 99만 명 늘어난 436만 명으로 전체의 9.3%를 차지하고 있다.

또한 통계법 제32조3)가 통계응답자의 성실의무를 규정하고 있으나, 개인비밀보호의식의 강화로 자살충동 경험, 이혼 등 민감한 조사항목에 대한 조사는 더욱 어려워지고 있다.

[그림 3-4] 1인 가구 및 노인인구 추이



자료: 통계청, 「인구주택총조사」, 각년도.

- 3) 통계법 제32조(통계응답자의 성실의무) 통계응답자는 통계의 작성에 관한 사무에 중사하는 자료부터 통계의 작성을 목적으로 질문 또는 자료제출 등의 요구를 받은 때에는 신뢰성 있는 통계가 작성될 수 있도록 조사사항에 대하여 성실하게 응답하여야 한다.

조사자 및 응답자에 대한 응답강제는 법률적, 도덕적으로 문제발생 소지가 내재되어 있다. 이는 조사자뿐만 아니라 응답자로 하여금 자의적인 조사나 응답을 하도록 유도하여 결과에 왜곡을 초래할 수 있다. 조사응답⁴⁾은 응답자의 선택사항이나 일단의 응답은 성실이 담보되어야 한다.

〈표 3-1〉 1인 가구 및 연령별 인구

(단위: 가구, 명, %)

구분		1980년		1990년	
			구성비		구성비
일반 가구	전체	7,969,201	100.0	11,354,540	100.0
	1인 가구	382,743	4.8	1,021,481	9.0
	2인 이상	7,586,458	95.2	10,333,059	91.0
연령별 인구	전체	37,406,815	100.0	43,390,374	100.0
	0~14세	12,655,775	33.8	11,134,215	25.7
	15세 이상	24,751,040	66.2	32,256,159	74.3
	(65세 이상)	(1,446,114)	(3.9)	(2,162,239)	(5.0)

구분		2000년		2005년	
			구성비		구성비
일반 가구	전체	14,311,807	100.0	15,887,128	100.0
	1인 가구	2,224,433	15.5	3,170,675	20.0
	2인 이상	12,087,374	84.5	12,716,453	80.0
연령별 인구	전체	45,985,289	100.0	47,041,434	100.0
	0~14세	9,638,756	21.0	8,986,128	19.1
	15세 이상	36,346,533	79.0	38,055,306	80.9
	(65세 이상)	(3,371,806)	(7.3)	(4,365,218)	(9.3)

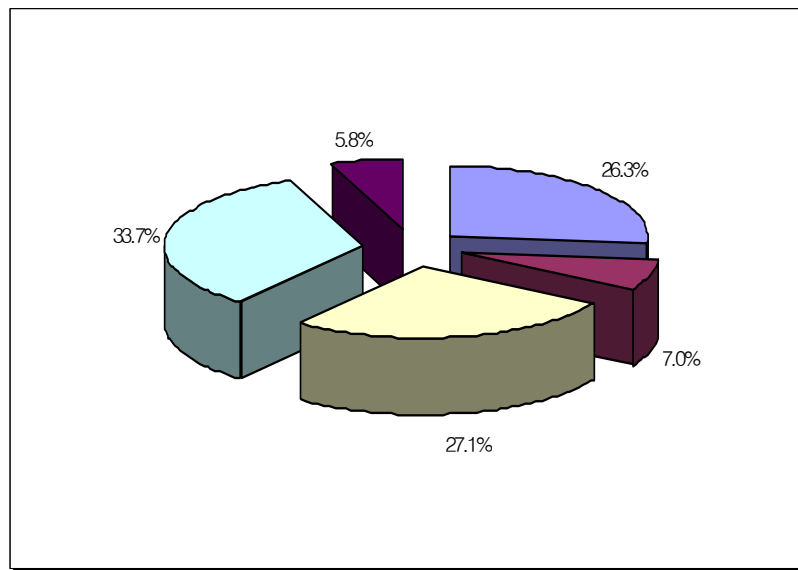
자료: 통계청, 「인구주택총조사」, 각년도.

4) 인지응답부담과 실질응답부담을 말하며, 인지응답부담은 개인들 간에 다르게 나타나는 주관적, 질적인 부담, 실질응답부담은 시간 또는 비용으로 측정 가능한 객관적이고, 정량적인 부담을 일컫음. Willeboordse(1998), Rainer(2006), Hedlin et al.(2005).

나. 표본변경에 따른 변화

현재 사회통계조사는 경제활동인구조사, 가계조사, 소비자전망조사, 집세조사, 사회통계조사, 생활시간조사, 가계자산조사 등과 같이 동일한 표본에서 조사하는 다목적 표본이나, 2008년 사회통계조사부터는 단독(독립)표본으로 변경하여 실시할 계획이다. 따라서 기존에 동일한 가구에서 여러 통계를 조사하던 방식에서 벗어나 표본이 변경된 다수의 신규조사가구를 대상으로 조사하여야만 한다. 따라서 그동안 고운 정 미운 정으로 쌓아 왔던 친숙성으로 조사가 보다 용이하였으나, 신규조사대상 가구의 경우 낯선 조사에 대한 거부감, 조사문항에 대한 이해의 어려움 등으로 조사불응이나 항목무응답이 나타날 개연성이 매우 높을 것이다.

[그림 3-5] 조사대상 가구원 중 불응 및 불응 사유



자료: 서울지방통계청, 「2007년 강남구의 사회통계조사 용역사업 결과보고서」, 2007. 8.

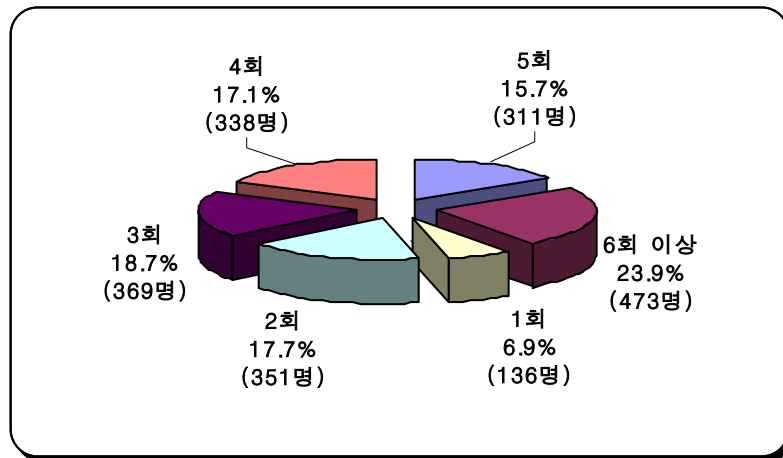
서울시 강남구가 사회개발 정책의 기초자료로 활용하고자 실시한 강남구 사회통계조사에 대한 「2007년 강남구의 사회통계조사 용역사업

결과보고서」⁵⁾(2007. 8)에 의하면 조사대상가구원 4,583명 중 불응·불능가구원이 498명으로 불응·불능률이 10.7%로 나타났다. 불응·불능가구원 498명의 불응·불능 사유 구성비를 보면 바빠서(33.7%), 사생활 노출기피(27.1%), 출장이나 여행으로 인한 부재중(26.3%)의 순으로 나타났다. ‘바빠서’, ‘사생활 노출기피’ 등 의도성 조사회피가 전체의 60.8%를 차지하고 있다.

다. 사회통계조사의 확대요구

최근의 인구, 복지, 환경 등 다양한 정책수립을 위하여 심층적인 정보수요가 증가하고 있음에 따라 사회통계조사의 확대 필요성은 배증되고 있다. 또한 경제구조변화에 따른 상대적 빈곤 심화와 급속한 고령화

[그림 3-6] 방문횟수 결과



자료: [그림 3-5]와 동일.

5) 서울지방통계청이 서울시 강남구로부터 용역 수주하여 실시한 서울시 강남구 사회통계조사 개요
 · 법적근거: 일반통계 승인번호 제52301호
 · 조사대상: 서울시 강남구 관내 100개조사구의 2,000 표본가구 내의 만 15세 이상 가구원(외국인 제외)
 · 조사기준시점: 2007. 3. 1.(0시 기준)
 · 조사기간: 2007. 3. 5. ~ 3. 19.(기간 중 13일)

뿐만 아니라 저출산 등 전통적 가족제도의 변화에 따른 새로운 복지정책의 필요성 등으로 각 기관에서 새로운 조사항목 추가를 요구하고 있다.

그러나 현재의 부족한 조사인력과 예산 등을 감안할 때 가족, 보건, 사회참여 등 11개 부문에 대하여 3~4개 부문을 연 1회 조사-부문별 평균 10개, 1회당 40개 항목을 조사-하는 현재의 사회통계방식으로는 대응하기에 역부족인 상태이다.

「2007년 강남구의 사회통계조사 용역사업 결과보고서」에 의하면 1개 통계조사에 소요된 조사원의 방문횟수는 6회 이상 방문 비율이 전체의 23.9%를 차지하고 있고, 1회 방문으로 조사가 완료된 경우는 전체의 6.9%에 불과한 실정이다.

제3절 해외의 대체법 활용 현황 및 이론적 고찰

1. 해외의 대체법 활용 현황

최근, OECD 국가들의 사회정책 수립에 주요한 기초자료로 사용되는 사회통계조사인 가구조사에 대한 대체법 활용에 대해 조사한 결과는 다음과 같다.

소득에 대한 정보 기재를 의무화하는 등록제도(registry)를 실시하는 몇몇 나라(네덜란드, 노르웨이, 포르투갈)를 제외하면, 최근 실시된 가구 조사는 30%(스위스)와 85%(멕시코, 캐나다) 사이의 응답률을 보였다. 유의한 분석이 가능한 14개국과 유로연합의 17개 가구조사 중 그리스, 멕시코, 체코, 핀란드, 헝가리의 통계조사는 대체법을 적용하지 않고 완전하게 관측된 자료만을 분석에 이용하고 있는 것으로 나타났다.

많은 국가에서 단순 대체법이 결측/무응답의 처리에 적용되고 있었으며, 세부적으로 Rubin(1987)의 hot-deck 방법(오스트리아, 유로연합, 한국, 호주)과 regression 방법(유로연합, 이탈리아, 일본)이 가장 널리 사용되고 있었다. 다음으로는 nearest-neighbor 방법(뉴질랜드, 캐나다)과 row-and-column 방법(독일)이 사용되고 있었으며, 미국은 regression을 이용한 다중 대체법을 적용하고 있었다.

〈표 3-2〉 해외 가구조사 현황

(단위: 가구, %)

국가명	조사명	표본크기	응답률	사용 대치법
한국	Survey of Household Income and Expenditure (2000)	27,001	82	hot deck
그리스	Household Budget Survey (1999)	6,258	-	사용 안함
네덜란드	Income Panel Survey (2000)	82,000	100	불필요
노르웨이	The Income Distribution Survey (2000)	12,919	100	불필요
뉴질랜드	Household Economic Survey (2004)	2,854	73	nearest neighbor
덴마크	Household Budget Survey (2005)	2,449	55	-
독일	Socio-Economic Panel (2007)	12,000	50	row-and-column
멕시코	Survey of Household Income and Expenditure (2000)	20,252	85	사용 안함
미국	Consumer Expenditure Survey (2004) Survey of Consumer Finances (1989)	7,500 4,522	81 70	regression
영국	Family Resources Survey (2006)	28,029		hot deck
오스트리아	Household Budget Survey (2005)	8,400	42	-
이탈리아	Bank of Italy Survey of Household Income and Wealth (2005)	8,012	38	regression
일본	Comprehensive Survey of Living Condition of the People on Health and Welfare (2004)	16,267	76	regression
캐나다	Survey of Consumer Finances (2004) Survey of Labour and Income Dynamics (2005)	33,843 33,461	85 85	nearest neighbor
핀란드	Household Budget Survey (2002)	8,960	63	사용 안함
호주	Household Expenditure Survey (2004)	9,753	71	hot deck

2. 이론적 고찰

무응답/결측값을 가지는 자료를 분석할 수 있는 모형은 최근 10여 년 사이에 많은 연구가 이루어졌다(Rubin, 1987; Rubin, 1996; Schafer, 1999). 무응답/결측값을 무시하지 않고 분석할 수 있는 통계 방법론의 하나인 대체법(imputation)이 이에 속하며, 특히 추론의 효율성, 복잡성 및 일치성 문제를 보완해 주는 다중대체법은 현재 많은 통계학자들에 의해 활발히 연구가 진행 중에 있다(Collins, Schafer, & Kam, 2001; Sinharay, Ster, Russell, 2001; Schafer & Graam, 2002; Fichman & Cummings, 2003).

Ford(1983)는 표본 관찰값이 독립인 경우 평균대체와 핫덱대체, 최근 방대체의 분산을 비교하였다. Kalton & Kish(1984)는 무응답률이 증가할 때 복원 핫덱대체와 비복원 핫덱대체의 분산을 비교하였으며, 무응답률이 증가함에 따라 복원의 분산이 비복원의 경우보다 더 증가함을 제시하고 있다.

가. 주관적 통계조사인 사회통계조사와 Imputation

사회통계조사의 큰 특징은 국민의 삶의 질과 관련된 사회적 관심사와 주관적 의식에 관한 사항을 조사하여 삶의 수준과 사회적 변동을 파악하는 주관적 항목에 관한 조사이다. 일반적으로 주관적 항목은 사회적인 바람직성(socially desirability)으로 대답할 가능성이 높으나, 어느 특정 가구원이 대신하여 응답할 수 있는 조사가 아니며 조사대상자 본인을 직접 조사해야 한다. 민감한 조사항목이 포함된 조사표의 경우에는 가구원(가족)간에도 사생활 보호를 우선시하여 작성되어야 한다. 이러한 특징을 지닌 사회통계조사의 주관적 항목에 대한 대체기법 연구에서 Penn(2007)⁶⁾은 ‘부모의 생활수준이 주관적인 행복에 미치는 영향’에 관한 무응답 대체에 관한 연구에서 다음과 같은 주요한 결과를 제시하였다.

첫째, 조사항목간에 통계적으로 분석한 결과 부모의 상대적 생활수준은 주관적 행복과 의미있는 관계를 가진다는 것이다. 이는 과학적이고 실증적인 분석 결과에서도 보여지고 있다.

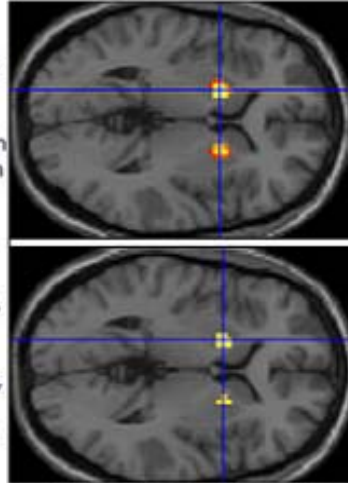
6) D. A. Penn, "Estimating missing values from the general social survey: An application of multiple imputation", *Social Science Quarterly*, 88(2), 2007, pp.573-584.

[그림 3-7] 행복에 대한 과학적 분석

'Brain scan'

Using magnetic resonance tomographs, the researchers examined the volunteers' blood circulation throughout the activities. High blood flow indicated that the nerve cells in the respective part of the brain were particularly active.

Neuroscientist Dr Bernd Weber explains: "One area in particular, the ventral striatum, is the region where part of what we call the 'reward system' is located. In this area, we observed an activation when the player completed his task correctly."



Brain scans showing high activation (top) and low activation (bottom).

A wrong answer, and no

- 주: 1) 독일 본 대학교의 연구팀은 38쌍의 남성을 대상으로 한 게임을 통해 상대보다 자신이 돈을 더 벌었다는 사실을 알게 되면, 배쪽 줄무늬체(ventral striatum)가 활성화한다는 것을 관찰했다. 즉, 사람은 상대에 비해 더 많은 돈을 벌었다는 사실을 알았을 때 더 큰 행복을 맛본다고 하였다(사이언스, 2007.11.23).
- 2) 영국 텔레그래프는 “친구보다 우위에 있을 때 진정으로 보상을 받았다고 느끼는 것”으로 나타났으며, 이는 “행복이 상대적인 것”이라는 사실을 보여준다고 평가하였다.

둘째, 결측값이 제외된 완전한 자료만을 사용하는 모형에서 제시된 것보다 대체법을 사용하여 결측값을 보완한 자료가 표준오차를 줄일 수 있어 보다 중요한 역할을 한다고 하였다.

여기서 Penn은 대체법이 연구자들로 하여금 주어진 자료 내에서 보다 많은 자료를 분석할 수 있게 함으로써 무응답을 가진 자료를 모두 무시할 때 발생할 수 있는 편향(bias)의 문제를 줄일 수 있다고 강조하여 설명하고 있다.

나. Imputation의 종류

대체법의 종류⁷⁾를 살펴보면 대체법은 한 개의 값으로 대체하는 단일 대체법(single imputation)과 여러 개의 값으로 대체하는 다중대체법(multiple imputation)으로 구분된다. 단일대체법은 다시 결정적 대체법(deterministic imputation)과 확률적 대체법(stochastic imputation)으로 구분된다.

결정적 대체법은 결측값에 대한 대체값이 유일하게 결정되는 대체법으로 평균대체(mean imputation), 비대체(ratio imputation), 회귀대체(regression imputation), 최근방대체(nearest neighbor imputation), 연역적 대체(deductive imputation), 과거자료대체(historical imputation) 등이 있다.

확률적 대체법은 대체값이 확률적으로 정해지는 방법으로 핫덱대체(hot deck imputation), 가중핫덱대체(weighted hot deck imputation), 랜덤비대체(random ratio imputation), 랜덤회귀대체(random regression imputation) 등을 들 수 있는데 조사의 규모와 비용, 성격 등을 고려하여 적절한 대체방법을 선택하여야 하며 그 내용은 다음과 같다.

1) 평균대체법(mean imputation)

평균대체법은 전체 표본을 몇 개의 대체군으로 분류한 후 각 층에서의 응답 평균값으로 대체하는 방법이다. 이 대체법은 사용의 간편성과 비용측면을 감안하여 실제적으로 사회통계조사에서 많이 사용되고 있으며, 항목변수가 양적변수이고 구하고자 하는 통계량이 평균일 때 유용하다. 그러나 대체 후 값들은 평균값의 빈도수가 지나치게 많아져 응답값들의 분포가 상당부분 왜곡될 가능성이 있다.

2) 최근방대체법(nearest neighbor imputation)

최근방대체법은 Brooks & Bailer(1978)에 의해 미국의 경상인구조사(Current Population Survey)에서 발생한 결측자료의 처리에 사용된 방법으로, 전체 표본을 몇 개의 대체군으로 분류하여 각 층에서의 응답자료를 순서대로 정리한 후 결측값 바로 이전의 응답을 결측값으로 대체한다. 편의성이나 비용측면에서 상당히 유용해 보이지만 같은 응답값이

7) 조사통계연구회, 「무응답오차」, 자유아카데미, 2000, pp.87-89.

여러 번 사용될 수 있고 무응답률이 높은 경우 평균, 분산 등의 모수추정에 심각한 편향을 가져올 수 있다는 단점이 있다.

3) 콜드덱대체법(cold-deck imputation)

콜드덱대체법은 결측값을 기존에 실시된 표본조사에서의 유사한 항목의 응답값으로 대체하는 방법이다. 동일한 질문이라 하더라도 현재와 과거라는 시간적 차이에 의한 의견 조정을 무시하고 과거자료를 사용함으로써 응답의 일치성에 문제점을 지니고 있다.

4) 조사단위대체법(substitution)

조사단위대체법은 무응답된 대상을 표본으로 추출되지 않은 다른 대상으로 대체시키는 방법이다.

5) 회귀대체법(regression imputation)

회귀대체법은 무응답이 있는 항목(반응변수) y 에 응답이 있는 y 의 보조변수(설명변수) X_1, X_2, \dots, X_n 을 회귀모형에 적합시키는 방법이다. 이 방법은 관심변수와 보조변수가 양적변수이며 두 변수간에 상관관계가 높아서 회귀모형(regression model)이 잘 적합하여야 한다.

6) 이월대체법(carry-over imputation)

이월대체법은 Kalton & Kasprzyk(1982)가 제안한 대체방법으로 조사시점 순서로 표본정렬 후 무응답 t 시점의 항목 y_t 에 가장 가까운 과거 u 시점 응답값 y_u 를 회귀모형에 적합시킨다. 이 방법은 조사시점과 조사항목 간에 높은 상관관계가 있을 때 적용하면 효과적이다.

7) 핫덱대체법(hot-deck imputation)

핫덱대체법은 대체층 내에서 대체값을 확률추출에 의해 랜덤하게 선택하여 결측값에 대체하는 방법이다. 평균대체방법이 y 의 분포를 왜곡시킬 수 있다는 문제점을 완화시킬 수 있다.

제4절 항목무응답 모의실험

1. 항목무응답의 발생원인과 결측항목의 결정

가. 항목무응답의 발생원인

항목무응답 발생원인은 ① 일부 문항에 대한 응답누락의 경우, ② 불일치 혹은 유효하지 않은 응답의 경우로 크게 구분할 수 있다.

첫 번째, 일부 문항에 대한 응답누락의 경우는 문항 작성시 불명확한 개념이나 부정확 혹은 이해하기 어려운 용어의 사용(소비: 소비의 포함 범위, 소득: 보너스 및 기타소득의 포함여부 등), 민감한 질문에 대한 응답(소득, 이혼사유 등), 길거나 지루한 설문 혹은 문항의 수가 많은 경우(이중형 설문 포함), 과거 기억에 의존하여 응답하게 하는 질문, 고의적인 응답누락 혹은 응답기피 등이 있다.

두 번째는 불일치 혹은 유효하지 않은 응답의 경우로 기억이나 경험에 근거하여 응답이 어려운 질문문항 내용에 대한 이해 부족 혹은 오류·잘못 작성된 질문, 질문을 다 읽지 않은 경우, 응답 선택 지문의 누락, 응답자의 응답 오류, 획득 자료의 분실 등이 있다.

나. 결측항목의 결정

실제 사회통계조사 결과와 가장 유사한 가상의 항목무응답 및 결측값을 구하기 위해, 2007 사회통계조사에 실제로 투입되어 조사활동을 하였던 조사담당자를 대상으로 3개 지방청의 사회조사팀장을 통하여 전자우편조사를 실시하였다. 실시된 설문조사 결과를 토대로 [7번 문항], [7-3번 문항] 및 [40번 문항]을 결측항목으로 하였다.

□ 조사개요

- 조사내용 : “2007 사회통계조사를 수행하면서 가장 어렵고 힘들었던 조사문항은 어느 항목이었습니까?”

- 조사대상 : 서울/부산/광주전남지방통계청 조사담당자
- 조사대상기간 : 2007. 9. 18~19(2일간)
- 조사방법 : 전자우편조사
- 조사결과(가장 어렵고 힘들었던 조사문항)
 - ① 사회보험료 부담여부[7번 문항]
 - ② 사회보험료부담액에 대한 인식-고용·산재보험[7-3번 문항]
 - ③ 장래소득에 대한 기대[40번 문항]
 - ④ 기타의견
 - 6.생활여건의 변화, 8.향후 늘려야 할 공공시설, 10.장애인 복지카드 소유 여부, 15.장애인 복지사업, 18.본인의 노후준비, 27.독서, 33.국내 관광여행 횟수, 28.레저시설 이용 횟수, 29.문화예술 및 스포츠 관람, 38.소득만족도, 41.소득분배에 대한 견해, 43.가구소득

2. 표본추출

가. 표본추출 방법

가상의 무응답 결측값을 만들기 위한 표본추출은 모집단 전체에 걸쳐 공평하게 표본을 추출할 수 있어 대표성이 높고 표본추출이 용이한 계통적 표본추출 방법(systematic sampling method)으로 하였다.

나. 표본추출 절차

10% 표본추출의 경우는 리스트에서 5번째의 요소를 1번 요소로 추출한 후 매 $[5+10*i]$ 번째의 요소를 추출하였다. 또한 20% 표본추출의 경우는 리스트에서 2번째의 요소를 1번 요소로 추출한 후 매 $[2+5*i]$ 번째의 요소를 추출하였다.

다. 표본추출 결과

2007 사회통계조사 조사대상가구원수 73,946명 중 10%인 7,395명 (20%의 경우는 14,789명)이 모의실험을 위한 가상의 항목무응답자로 표

본추출되었다. 성별, 연령별, 가구주관계 등 인구특성별 표본추출 결과는 다음과 같다.

〈표 3-3〉 표본추출 결과

(단위: 명, %)

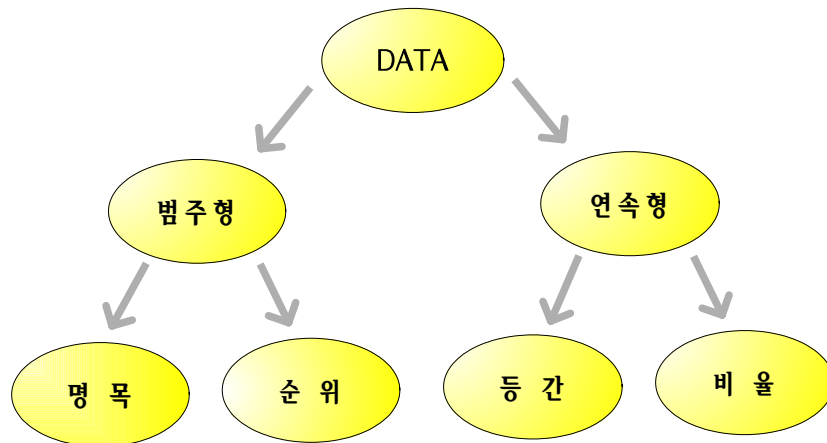
구분		전체	10% 표본		20% 표본		
			구성비	구성비	구성비	구성비	
계		73,946	100.0	7,395	100.0	14,789	100.0
성별	남자	35,033	47.4	3,480	47.1	6,935	46.9
	여자	38,913	52.6	3,915	52.9	7,854	53.1
연령별	15세 미만	7,335	9.9	730	9.9	1,488	10.1
	15세~19세	5,284	7.1	521	7.0	1,061	7.2
	20세~24세	3,888	5.3	403	5.4	737	5.0
	25세~29세	5,185	7.0	548	7.4	1,024	6.9
	30세~34세	5,844	7.9	582	7.9	1,163	7.9
	35세~39세	7,236	9.8	738	10.0	1,392	9.4
	40세~44세	7,275	9.8	741	10.0	1,508	10.2
	45세~49세	7,669	10.4	780	10.5	1,550	10.5
	50세~54세	6,003	8.1	585	7.9	1,164	7.9
	55세~59세	4,682	6.3	475	6.4	952	6.4
	60세~64세	3,869	5.2	372	5.0	761	5.1
65세 이상	9,676	13.1	920	12.4	1,989	13.4	
가구주 관계	가구주	30,449	41.2	2,978	40.3	6,042	40.9
	배우자	19,551	26.4	2,017	27.3	3,994	27.0
	미혼자녀	18,610	25.2	1,886	25.5	3,679	24.9
	기혼자녀/배우자	778	1.1	85	1.1	155	1.0
	손자녀/배우자	597	0.8	52	0.7	120	0.8
	부모	2,550	3.4	236	3.2	523	3.5
	조부모	45	0.1	6	0.1	11	0.1
	미혼형제자매	676	0.9	69	0.9	112	0.8
	기타친인척	600	0.8	57	0.8	134	0.9
기타동거인	90	0.1	9	0.1	19	0.1	
혼인상태	미혼	16,014	21.7	1,623	21.9	3,120	21.1
	배우자있음	43,777	59.2	4,280	57.9	8,572	58.0
	사별	5,870	7.9	568	7.7	1,219	8.2
	이혼	2,122	2.9	224	3.0	419	2.8
	기타	7,163	9.7	710	9.6	1,459	9.9
학력별	초졸 이하	20,820	28.2	2,023	27.4	4,216	28.5
	중졸	10,759	14.5	1,074	14.5	2,202	14.9
	고졸	25,714	34.8	2,637	35.7	5,031	34.0
	대졸 이상	16,653	22.5	1,661	22.5	3,340	22.6

3. 연관성 분석 및 회귀분석

2007 사회통계조사 조사담당자에 대한 설문조사 결과를 이용하여 선정된 가상의 항목무응답/결측값이 무응답자와 응답자에 따라 특성이 변수간에 다른지, 또는 이용할 수 있는 정보가 있는지 알기 위하여 기본항목 12개(성별, 시도, 점유형태, 연령, 교육정도, 혼인상태 등)와 조사항목 184개(생활여건의 변화, 사회보험료 부담에 대한 인식 등)에 대해 통계적으로 유의한지 연관성을 조사하였다.

연관성 분석은 변수간의 관계에 따라 분석방법을 달리하여야 하나 Imputation의 실용성 확보를 위해 반응변수를 중심으로 [7번 문항]은 Chi-square Test, [7-3번 문항]과 [40번 문항]은 spearman 상관분석을 실시하였다.⁸⁾

[그림 3-8] 자료의 종류

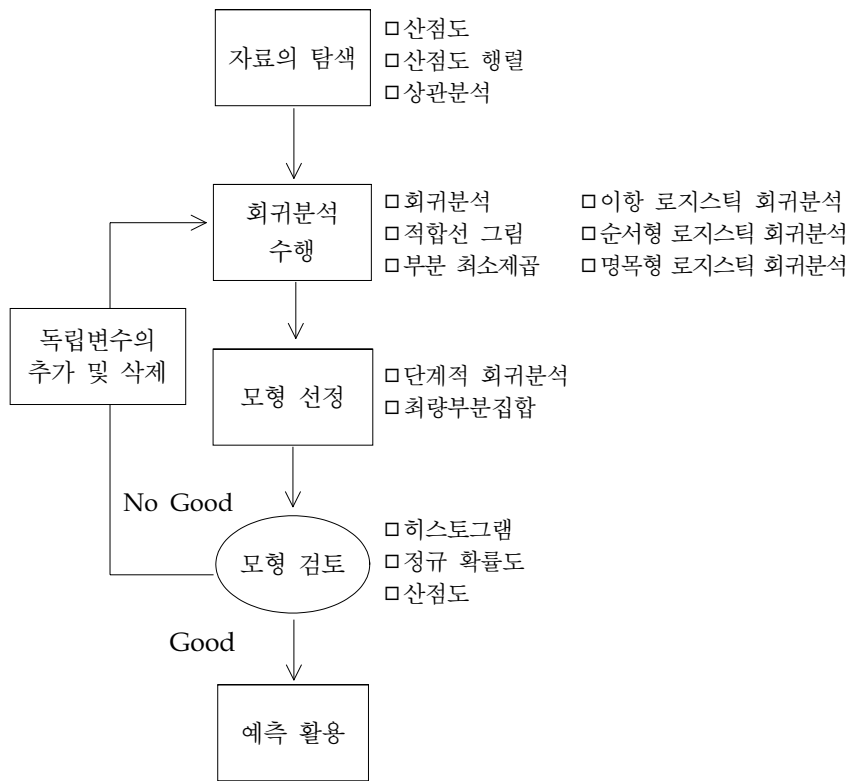


8) Contingency Table Chi-square Test는 $r \times c$ 교차분석표에 있어서 명목척도로 측정된 변수들의 독립성을 검증하기 위하여 적용. 2×2 교차분석표에 있어서는 $\phi(\phi)$ 를 사용하고 그 밖의 교차분석에서는 Cramer's v 를 사용할 수 있음. Spearman's Rank Correlation Coefficient(ρ)는 순위척도로 측정된 변수 사이에 존재하는 연관성의 정도를 나타내는 지수로 rho라고도 일컬음(유동근, “제9장 마케팅스쿨”, 2007: <http://www.marketingschool.com>, 11월 28일자).

<표 3-4> 자료의 종류

구분	설명	예시
명목자료	범주로만 의미	성별, 혈액형
순위자료	명목+대소관계	찬성/중립/반대
등간자료	절대영점(×)	온도(섭씨/화씨)
비율자료	절대영점(○)	매출액

[그림 3-9] 회귀분석¹⁾



주: 1) 미니탭, “미니탭 활용법과 기초통계”, 2007: <http://www.minitab.co.kr>, 11월 28일자.

4. 항목무응답대체

가. [7번 문항]에 대한 대체

사회보험료 납부 여부 (1. 있다 2. 없다)

[그림 3-10] 사회보험료 부담에 대한 인식

사회보험료 부담에 대한 인식

7 귀하가 현재 납부하고 있는 사회보험료(건강보험, 국민연금, 고용보험 등)가 있습니까? 있다면 납부하고 있는 사회보험료 각각에 대하여 귀하의 소득에 비해 어느 정도 부담이 된다고 생각하시는지 체크하여 주십시오.

※여기에서 사회보험료를 납부한다는 것은 본인의 소득에서 공제되는 것을 의미합니다.

있다 **없다**

	● 매우 부담됨	● 약간 부담됨	● 보통임	● 별로 부담안됨	● 전혀 부담안됨	● 해당없음
1. 건강보험						
2. 국민연금						
3. 고용·신체보험						
4. 기타 ()						

없다

회귀분석 모형은 단지 두 개의 값만을 가지는 종속변수(있다/없다)와 독립변수들 간의 인과관계를 이용하여 추정하는 통계기법인 로지스틱 회귀분석(logistic regression)⁹⁾을 이용하였다. 또한 항목무응답/결측값을 설명할 수 있는 이용가능한 변수를 찾기 위해 결측항목(7번 문항)과 기본항목, 타 조사항목과 카이제곱 검정을 실시하였다.

카이제곱 검정 결과 기본항목 중 ‘종사상의 지위’, ‘직업’ 등과 ‘38.소득만족도 중 소득의 유무’, ‘18.노후준비방법 중 준비여부’ 등이 유의한 것으로 나타났다. 연관성이 높게 나타난 기본항목과 조사항목은 다음과 같다.

9) 종속변수가 연속형이 아닌 경우의 회귀분석으로 반응변수가 이진형/순서형/분류형인 경우 사용할 수 있으며, 명목형/범주형 자료분석에 이용됨(권세형, “제9장 로지스틱회귀분석”, 「Regression」, 2007: [@2005 Spring, 11월 28일자](http://wolfpack.hannam.ac.kr)).

- 기본항목

설명변수	표본크기	X^2 값	유의확률	크래머의 V
종사상의 지위	59,946	31,843	<.0001	0.7288
직업	59,946	26,359	<.0001	0.6631
산업	59,946	25,164	<.0001	0.6479
경제활동상태	59,946	22,605	<.0001	0.6141
연령	59,946	10,729	<.0001	0.4231
성별	59,946	8,809	<.0001	0.3673

- 조사항목

설명변수	표본크기	X^2 값	유의확률	크래머의 V
소득 만족도 중 소득의 유무[38번]	59,946	22,879	<.0001	0.6178
노후 준비방법 중 준비 유무[18번]	56,803	11,040	<.0001	0.4409
현재 소득수준에 대한 인식[39번]	38,388	5,448	<.0001	0.3767

[그림 3-11] 본인의 노후 준비 방법

본인의 노후 준비 방법

18 만 18세 이상(1989. 6. 15.이전 출생자만 기입하여 주십시오. 귀하는 노후를 위하여 어떤 준비를 하고 계십니까? 준비하고 계신다면 그 중 가장 주된 것과 부수적인 것을 기입하여 주십시오.

☑ 준비하고 있다

가장 주된 것 부수적인 것

- ① 국민연금
- ② 기타 공적연금(공무원, 군인, 교원)
- ③ 사적 연금(은행, 보험회사 등을 통한 개인연금)
- ④ 퇴직금
- ⑤ 예금, 적금, 저축성보험
- ⑥ 부동산 운용
- ⑦ 주식, 채권 등
- ⑧ 기타

☐ 준비하고 있지 않다

- ① 아직 생각하고 있지 않다
- ② 앞으로 준비 할 계획이다
- ③ 준비할 능력이 없다
- ④ 자녀에게 의지한다

[그림 3-12] 소득 만족도



[그림 3-13] 현재 소득수준에 대한 인식



표본추출(10%, 20%)을 제외한 90% 및 80% 자료에 대해 stepwise 변수선택 방법을 적용하여 변수(4개)를 선택하였다. 선택된 변수의 도수분포표(90%)¹⁰⁾는 다음과 같다.

- 기본항목: 종사상의 지위[bun_j], 연령[bun_g], 성별[bun_s]
- 조사항목: 현재 소득의 유무[h38], 노후 준비 유무[h18]

<표 3-5> [7번 문항] 관련 통계 도수분포표

(단위: 명)

구 분	대상자수	7번 조사항목		구성비 ¹⁾	
		있음(1)	없음(0)		
합 계	59,946	28,587	31,359	0.4769	
종사상의 지위	상용근로자	12,409	12,392	17	0.9986
	임시근로자	7,325	4,597	2,728	0.6276
	일용근로자	3,470	1,646	1,824	0.4744
	고용주	2,369	2,125	244	0.8970
	자영자	7,840	5,128	2,712	0.6541
	무급가족종사자	2,767	324	2,443	0.1171
	기타	23,766	2,375	2,139	0.9990

10) 전체 자료 중 가상의 결측값을 만들기 위해 표본추출한 10%를 제외한 자료를 말함.

<표 3-5>의 계속

(단위: 명)

구 분		대상자수	7번 조사항목		구성비 ¹⁾
			있음(1)	없음(0)	
연령별	15세~19세	4,763	68	4,695	0.0011
	20세~24세	3,485	1,037	2,448	0.2976
	25세~29세	4,637	2,664	1,973	0.5745
	30세~34세	5,262	3,177	2,805	0.6038
	35세~39세	6,498	4,113	2,385	0.6330
	40세~44세	6,534	4,304	2,230	0.6587
	45세~49세	6,889	4,461	2,428	0.6476
	50세~54세	5,418	3,636	2,055	0.6207
	55세~59세	4,207	2,293	1,914	0.5450
	60세~64세	3,497	1,383	2,114	0.3955
	65세 이상	8,756	1,724	7,032	0.1969
성 별	1. 남자	28,070	18,874	9,196	0.6724
	2. 여자	61,876	9,713	22,163	0.3047
현재소득 유무		59,946	28,587	31,359	0.5231
노후준비 유무		56,803	28,578	28,225	0.5031

주: 1) 구성비 = 있음/대상자수.

먼저 통계적으로 유의한 것으로 나타난 변수들을 이용하여 로지스틱 회귀모형을 도출하였다. 도출된 회귀모형에 가상의 결측값(표본추출한 10% 및 20%)을 종속변수로 하여 대체값을 산출하였다. 구한 대체값과 실제값을 비교한 결과 다음과 같이 나타났다.

- 10%인 경우: 응답자별 실제값과 대체값 일치율 85.7%
(표 3-6 참조)
- 20%인 경우: 응답자별 실제값과 대체값 일치율 85.3%
(표 3-7 참조)

<표 3-6> [7번 문항] 통계량 및 Imputation 전·후 비교(1)

Parameter	DF	Estimate	Standard		Wald
			Error	Chi-Square	Pr > ChiSq
Intercept	1	5.5092	0.3853	204.4783	<.0001
bun_j_1	1	-7.1736	0.2479	837.5628	<.0001
bun_j_2	1	-1.4373	0.0490	861.0414	<.0001
bun_j_3	1	-0.6735	0.0557	146.3250	<.0001
bun_j_4	1	-2.1921	0.0818	717.4607	<.0001
bun_j_5	1	-1.1044	0.0462	571.5108	<.0001
bun_j_6	1	-0.00192	0.0684	0.0008	0.9776
bun_g_1	1	1.5594	0.1895	67.7042	<.0001
bun_g_2	1	0.1558	0.0760	4.2003	0.0404
bun_g_3	1	-0.5057	0.0663	58.1577	<.0001
bun_g_4	1	-0.8803	0.0612	206.7794	<.0001
bun_g_5	1	-1.1822	0.0554	455.0051	<.0001
bun_g_6	1	-1.3453	0.0537	628.1095	<.0001
bun_g_7	1	-1.4159	0.0522	736.0340	<.0001
bun_g_8	1	-1.4567	0.0546	710.5020	<.0001
bun_g_9	1	-1.2371	0.0566	477.4368	<.0001
bun_g_10	1	-0.6189	0.0578	114.5654	<.0001
bun_s_1	1	-1.5763	0.0292	2,911.6742	<.0001
h38_1	1	-1.3303	0.0483	757.4733	<.0001
h18_1	1	-2.2262	0.3838	33.6499	<.0001

실제값				
Imputation	빈도 백분율 행 백분율 칼럼 백분율	0	1	총합
		0	2824 42.37 89.91 81.62	317 4.76 10.09 9.89
1	636 9.54 18.05 18.38	2,888 43.33 81.95 90.11	3,524 52.87	
총합	3460 51.91	3205 48.09	6665 100.00	

<표 3-7> [7번 문항] 통계량 및 Imputation 전·후 비교(2)

Parameter	DF	Estimate	Standard		Wald
			Error	Chi-Square	Pr > ChiSq
Intercept	1	5.5123	0.3843	205.7012	<.0001
bun_j_1	1	-7.1283	0.2636	731.0601	<.0001
bun_j_2	1	-1.4054	0.0516	740.8644	<.0001
bun_j_3	1	-0.6347	0.0588	116.6002	<.0001
bun_j_4	1	-2.1710	0.0879	610.5106	<.0001
bun_j_5	1	-1.0562	0.0488	467.7668	<.0001
bun_j_6	1	-0.00010	0.0724	0.0000	0.9989
bun_g_1	1	1.4075	0.1875	56.3455	<.0001
bun_g_2	1	0.1221	0.0795	2.3570	0.1247
bun_g_3	1	-0.4728	0.0705	44.9224	<.0001
bun_g_4	1	-0.9184	0.0647	201.4603	<.0001
bun_g_5	1	-1.1865	0.0587	408.7415	<.0001
bun_g_6	1	-1.2867	0.0571	507.2471	<.0001
bun_g_7	1	-1.3912	0.0556	626.2719	<.0001
bun_g_8	1	-1.4311	0.0579	610.3142	<.0001
bun_g_9	1	-1.2313	0.0603	417.5470	<.0001
bun_g_10	1	-0.6202	0.0616	101.4735	<.0001
bun_s_1	1	-1.5779	0.0309	2,600.0428	<.0001
h38_1	1	-1.4119	0.0512	759.7749	<.0001
h18_1	1	-2.2060	0.3827	33.2238	<.0001

		실제값		
Imputation	빈도 백분율 행 백분율 칼럼 백분율	0	1	총합
	0	0	5720	634
		43.00	4.77	47.77
		90.02	9.98	
		81.28	10.12	
1	1	1,317	5,630	6947
		9.90	42.33	52.23
		18.96	81.04	
		18.72	89.88	
총합		3460	3205	13301
		51.91	48.09	100.00

[7번 문항]에 대한 분석 결과 응답자별 실제값과 대체값 일치율은 10% 가상결측의 경우 85.7%, 20% 가상결측의 경우 85.3%로 나타났다.

도출된 회귀분석의 의미는 “「사회보험료 납부 여부」는 종사상의 지위와 연령별 및 성별뿐만 아니라 현재 소득의 유무, 노후준비 유무와 밀접한 관련이 있다”는 것을 나타내고 있다.

나. [7-3번 문항]에 대한 대체

□ 「고용·산재보험」 부담에 대한 인식(6점 척도)

[그림 3-14] 「고용·산재보험」 부담에 대한 인식

	① 매우 부담됨	② 약간 부담됨	③ 보통임	④ 별로 부담안됨	⑤ 전혀 부담안됨	⑥ 해당없음
1. 건강보험						
2. 국민연금						
3. 고용·산재보험						
4. 기타()						

회귀모형은 순위척도(1.매우 부담됨~5.전혀 부담안됨)와 명목척도(6.해당없음)로 이루어진 문항으로 구성되어 있다. 먼저 「6.해당없음」을 구분해 내기 위해 두 개의 값만을 가지는 종속변수(있다/없다)와 독립변수들 간의 인과관계를 이용하여 추정하는 통계기법인 로지스틱 회귀분석(logistic regression)을 이용하였다. 다음에 순위척도 분석을 위해 순위형 로지스틱 분석을 실시하였다.

항목무응답/결측값을 설명할 수 있는 이용 가능한 변수를 찾기 위해 결측항목(7-3번 문항)을 기본항목, 타 조사항목과 상관분석을 실시하였다. 그 결과 기본항목에서는 ‘종사상의 지위’, ‘연령’ 및 ‘혼인상태’, 조사항목에서는 ‘39.현재 소득 수준에 대한 인식 중 주된 소득’이 유의한 것으로 나타났다.

- 기본항목

설명변수	표본크기	상관계수	유의확률
종사상의 지위	31,792	0.5512	<.0001
연령	31,792	0.3456	<.0001
혼인상태	31,792	0.2353	<.0001

- 조사항목

설명변수	표본크기	상관계수	유의확률
현재 소득수준에 대한 인식 중 주된 소득 [39]	30,253	0.4573	<.0001

<1차 작업> 결과 선택된 변수에 대한 도수분포표(90%)를 살펴보면 다음과 같다.

<표 3-8> [7-3번 문항] 관련 통계 도수분포표

(단위: 명)

구분		대상자수	7-3번 조사항목		구성비 ¹⁾
			6번	1~5번	
합 계		28,587	14,783	13,804	0.5171
종사상의 지위	상용근로자	12,392	2,509	9,883	0.2025
	임시근로자	4,597	2,160	2,437	0.4699
	일용근로자	1,646	1,307	339	0.7940
	고용주	2,125	1,567	558	0.7374
	자영자	5,128	4,762	366	0.9286
	무급가족종사자	324	284	40	0.8765
	기타	2,375	2,194	181	0.9238
연령	15세~19세	68	18	50	0.2647
	20세~24세	1,037	158	879	0.1524
	25세~29세	2,664	543	2,121	0.2038
	30세~34세	3,177	1,802	1,375	0.3406
	35세~39세	4,113	1,797	2,316	0.4369
	40세~44세	4,304	2,218	2,086	0.5153

〈표 3-8〉의 계속

(단위: 명)

구분		대상자수	7~3번 조사항목		구성비 ¹⁾
			6번	1~5번	
연령	45세~49세	4,461	2,649	1,812	0.5938
	50세~54세	3,363	2,139	726	0.6360
	55세~59세	2,293	1,567	726	0.6834
	60세~64세	1,383	1,028	355	0.7483
	65세 이상	1,724	1,584	140	0.9188
혼인상태	미혼	5,024	1,375	3,649	0.2737
	배우자 있음(법률혼)	20,137	10,942	9,195	0.5434
	배우자 있음(동거)	599	362	237	0.6043
	사별	1,453	1,144	309	0.7873
	이혼	1,374	960	414	0.6987
주된소득	근로소득	18,962	6,267	12,695	0.3305
	사업소득	6,888	6,010	878	0.8725
	재산소득	466	419	47	0.8991
	기타소득(이전소득, 연금 등)	867	814	53	0.9389

주: 1) 구성비 = 있음/대상자수.

먼저 통계적으로 유의한 것으로 나타난 변수들을 이용하여 로지스틱 회귀모형을 도출하였다. 도출된 회귀모형에 가상의 결측값(표본추출한 10% 및 20%)을 종속변수로 하여 대체값을 산출하였다. 구한 대체값과 실제값을 비교한 결과 다음과 같이 나타났다.

□ 10%인 경우

〈표 3-9〉 [7-3번 문항] 통계량 및 Imputation 전·후 비교(1)

Parameter	DF	Estimate	Standard		Wald
			Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.4266	0.0252	286.3729	<.0001
bun_j_6	0	1.0111	0.0242	1,743.9275	<.0001
bun_g_6	0	0.2889	0.0163	313.8894	<.0001
bun_m_6	0	0.2267	0.0217	109.4716	<.0001
h39_6	0	0.4574	0.0273	280.9451	<.0001

		실제값		
Imputation	빈도 백분율 행 백분율 칼럼 백분율	0	6	총합
	0	1,414 46.06 73.42 89.72	512 16.68 26.58 34.27	1,926 62.74
	6	162 5.28 14.16 10.28	982 31.99 85.84 65.73	1,144 37.26
	총합	1,576 51.34	1,494 48.66	3,070 100.00

□ 20%인 경우

<표 3-10> [7-3번 문항] 통계량 및 Imputation 전·후 비교(2)

Parameter	DF	Estimate	Standard		Wald
			Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.4216	0.0268	248.3630	<.0001
bun_j_6	0	1.0122	0.0255	1,574.4775	<.0001
bun_g_6	0	0.2856	0.0172	274.4615	<.0001
bun_m_6	0	0.2407	0.0230	109.9937	<.0001
h39_6	0	0.4626	0.0288	257.9035	<.0001

		실제값		
Imputation	빈도 백분율 행 백분율 칼럼 백분율	0	6	총합
	0	2,664 44.89 73.07 89.25	982 16.55 26.93 33.29	3,646 61.43
	6	321 5.41 14.02 10.75	1,968 33.16 85.98 66.71	2,289 38.57
	총합	2,985 50.29	2,950 49.71	5,935 100.00

<2차 작업>은 1차 작업에서 선별된 선택항목 ‘6. 해당없음’을 제외한 자료를 이용하였다. [1번~5번]항목은 순위형 로지스틱분석을 통하여 회귀모형을 탐색하였다.

탐색된 회귀모형에 가상의 결측값(표본추출한 10% 및 20%)을 종속 변수로 하여 대체값을 산출하였다. 구한 대체값과 실제값을 비교한 결과 다음과 같이 나타났다.

□ 10%인 경우 응답자별 실제값과 대체값 일치율 55.8%

<표 3-11> [7-3번 문항] 통계량 및 Imputation 전·후 비교(3)

Parameter	DF	Estimate	Standard Wald 95% Confidence			Pr>ChiSq		
			Error	Limits	Square			
Intercept	1	2.9398	0.1660	2.6146	3.2651	313.78	<.0001	
bun_j	710	1	0.1747	0.1083	-0.0377	0.3870	2.60	0.1069
bun_j	720	1	0.1223	0.1093	-0.0920	0.3366	1.25	0.2633
bun_j	730	1	0.1194	0.1177	-0.1113	0.3500	1.03	0.3104
bun_j	740	1	-0.1686	0.1211	-0.4059	0.0686	1.94	0.1636
bun_j	750	1	0.1209	0.1217	-0.1177	0.3594	0.99	0.3208
bun_j	760	1	-0.1768	0.2141	-0.5964	0.2427	0.68	0.4088
bun_g	105	1	-0.3445	0.1529	-0.6442	-0.0447	5.07	0.0243
bun_g	115	1	-0.2797	0.0837	-0.4438	-0.1157	11.18	0.0008
bun_g	120	1	-0.2090	0.0796	-0.3651	-0.0530	6.89	0.0086
bun_g	130	1	-0.1768	0.0780	-0.3297	-0.0239	5.14	0.0234
bun_g	135	1	-0.1574	0.0773	-0.3089	-0.0059	4.14	0.0418
bun_g	145	1	-0.1641	0.0774	-0.3158	-0.0123	4.49	0.0341
bun_g	150	1	-0.1718	0.0776	-0.3238	-0.0197	4.90	0.0268
bun_g	160	1	-0.1383	0.0787	-0.2925	0.0158	3.09	0.0786
bun_g	165	1	-0.1331	0.0810	-0.2919	0.0258	2.70	0.1006
bun_g	170	1	0.1132	0.0875	-0.0582	0.2846	1.68	0.1956
bun_m	510	1	0.0215	0.0485	-0.0735	0.1165	0.20	0.6571
bun_m	520	1	0.0069	0.0435	-0.0783	0.0922	0.03	0.8731
bun_m	530	1	0.0762	0.0707	-0.0623	0.2148	1.16	0.2808
bun_m	540	1	0.0122	0.0660	-0.1171	0.1416	0.03	0.8528
h39	1	1	-0.2138	0.1218	-0.4524	0.0249	3.08	0.0792
h39	2	1	-0.3897	0.1310	-0.6465	-0.1330	8.85	0.0029
h39	3	1	-0.3657	0.1734	-0.7056	-0.0257	4.44	0.0350

		실제값						
Imputation	빈도							총합
	백분율	1	2	3	4	5	6	
2	칼럼 백분율	0	0	1	0	0	0	1
	행 백분율	0.00	0.00	0.03	0.00	0.00	0.00	0.03
	총합	0.00	0.00	100.00	0.00	0.00	0.00	0.00
	비율	0.00	0.00	0.12	0.00	0.00	0.00	0.00
3	칼럼 백분율	124	349	732	166	42	512	1,925
	행 백분율	4.04	11.37	23.84	5.41	1.37	16.68	62.70
	총합	6.44	18.13	38.03	8.62	2.18	26.60	
	비율	81.58	88.58	91.27	91.21	91.30	34.27	
6	칼럼 백분율	28	45	69	16	4	982	1,144
	행 백분율	0.91	1.47	2.25	0.52	0.13	31.99	37.26
	총합	2.45	3.93	6.03	1.40	0.35	85.84	
	비율	18.42	11.42	8.60	8.79	8.70	65.73	
총합	칼럼 백분율	152	394	802	182	46	1,494	3,070
	행 백분율	4.95	12.83	26.12	5.93	1.50	48.66	100.00

□ 20%인 경우: 응답자별 실제값과 대체값 일치율 55.8%

<표 3-12> [7-3번 문항] 통계량 및 Imputation 전·후 비교(4)

Parameter	DF	Estimate	Standard Wald 95% Confidence			Pr > ChiSq		
			Error	Limits	Square			
Intercept	1	2.9067	0.1753	2.5631	3.2503	274.88	<.0001	
bun_j	710	1	0.2524	0.1170	0.0230	0.4818	4.65	0.0311
bun_j	720	1	0.1859	0.1181	-0.0455	0.4173	2.48	0.1153
bun_j	730	1	0.1603	0.1268	-0.0881	0.4088	1.60	0.2060
bun_j	740	1	-0.0952	0.1308	-0.3517	0.1612	0.53	0.4668
bun_j	750	1	0.2103	0.1311	-0.0466	0.4673	2.57	0.1086
bun_j	760	1	0.0100	0.2089	-0.3994	0.4194	0.00	0.9619
bun_g	105	1	-0.2882	0.1587	-0.5992	0.0229	3.30	0.0694
bun_g	115	1	-0.2752	0.0871	-0.4459	-0.1045	9.99	0.0016
bun_g	120	1	-0.1868	0.0829	-0.3493	-0.0243	5.07	0.0243
bun_g	130	1	-0.1592	0.0813	-0.3185	0.0001	3.84	0.0501
bun_g	135	1	-0.1376	0.0805	-0.2953	0.0202	2.92	0.0874
bun_g	145	1	-0.1468	0.0807	-0.3050	0.0113	3.31	0.0688
bun_g	150	1	-0.1491	0.0809	-0.3076	0.0094	3.40	0.0652

<표 3-12>의 계속

Parameter	DF	Estimate	Standard Wald 95% Confidence				Pr > ChiSq	
			Error	Limits		Square		
bun_g	160	1	-0.1395	0.0820	-0.3001	0.0211	2.90	0.0887
bun_g	165	1	-0.1074	0.0844	-0.2729	0.0580	1.62	0.2032
bun_g	170	1	0.1410	0.0909	-0.0371	0.3191	2.41	0.1207
bun_m	510	1	0.0227	0.0509	-0.0772	0.1225	0.20	0.6564
bun_m	520	1	-0.0129	0.0458	-0.1027	0.0770	0.08	0.7786
bun_m	530	1	0.0726	0.0736	-0.0717	0.2169	0.97	0.3241
bun_m	540	1	-0.0430	0.0683	-0.1768	0.0908	0.40	0.5289
h39	1	1	-0.2600	0.1249	-0.5049	-0.0151	4.33	0.0374
h39	2	1	-0.4320	0.1346	-0.6957	-0.1682	10.30	0.0013
h39	3	1	-0.5011	0.1834	-0.8605	-0.1417	7.47	0.0063

		실제값						
Imputation	빈도							총합
	백분율	1	2	3	4	5	6	
Imputation	칼럼 백분율	1	2	3	4	5	6	총합
	2	0	1	3	0	0	2	6
		0.00	0.02	0.05	0.00	0.00	0.03	0.10
		0.00	16.67	50.00	0.00	0.00	33.33	
		0.00	0.12	0.20	0.00	0.00	0.07	
	3	220	700	1,344	325	71	980	3,640
		3.71	11.79	22.65	5.48	1.20	16.51	61.33
		6.04	19.23	36.92	8.93	1.95	26.92	
		82.40	86.85	90.81	92.86	86.59	33.22	
	6	47	105	133	25	11	1,968	1,144
		0.79	1.77	2.24	0.42	0.19	33.16	37.26
		2.05	4.59	5.81	1.09	0.48	85.98	
	17.60	13.03	8.99	7.14	13.41	66.71		
총합	267	806	1,480	350	82	2,950	5,935	
	4.50	13.58	24.94	5.90	1.38	49.71	100.00	

[7-3번 문항]에 대한 분석 결과 응답자별 실제값과 대체값의 일치율은 10% 가상결측의 경우 55.7%, 20% 가상결측의 경우 55.8%로 나타났다.

회귀모형의 의미는 “「고용·산재보험 부담이 본인의 소득에 대한 부담여부」는 종사상의 지위, 연령, 혼인상태, 주된 소득과 관계가 있다”는 것을 나타내고 있다.

다. [40번 문항]에 대한 대체

□ 「장래소득에 대한 기대」(5점 척도)

순위척도(1.크게 증가~5.크게 감소)로 이루어진 문항으로 구성되어 있어 순위형 로지스틱분석을 실시하였다.

항목무응답/결측값을 설명할 수 있는 이용 가능한 변수를 찾기 위해 결측항목(40번 문항)을 기본항목, 타 조사항목과 연관분석을 실시한 결과 기본항목에서는 ‘연령’, ‘종사상의 지위’, ‘교육정도’, 조사항목에서는 ‘6-1. 전반적인 생활여건의 변화’, ‘독서 경험유무’, ‘문화예술 및 스포츠 경험유무’, ‘신문구독 유무’ 등이 유의한 것으로 나타났다.

[그림 3-15] 장래소득에 대한 기대



- 기본항목

설명변수	표본크기	상관계수	유의확률
연령	42,663	0.4719	<.0001
종사상의 지위	42,663	0.3086	<.0001
교육정도	42,663	-0.3881	<.0001

- 조사항목

설명변수	표본크기	상관계수	유의확률
독서(교과서 및 참고서 제외) 경험 유무[27]	42,663	0.2986	<.0001
문화예술 및 스포츠 관람 경험 유무[29]	42,663	0.2934	<.0001
신문(인터넷 신문 포함) 구독 유무[24]	42,663	0.2718	<.0001
전반적인 생활여건의 변화[6-1]	42,663	0.2499	<.0001

[그림 3-16] 전반적인 생활여건의 변화

생활여건의 변화

6 3년전(2004년)과 대비하여 다음 분야의 생활여건은 어느 정도 변화되었다고 보십니까?
각 분야마다 해당하는 것 하나에 체크하여 주십시오.

	① 많이 좋아짐	② 약간 좋아짐	③ 변화 없음	④ 약간 나빠짐	⑤ 많이 나빠짐
1. 전반적인 생활여건					
2. 보건 의료 서비스					
3. 사회보장제도(건강보험·국민연금 등)					
4. 문화·여가생활 할수여건					

[그림 3-17] 신문

신문(인터넷 신문 포함)

24 귀하는 지난 1개월(2007. 5. 15.~6. 14.)동안 2주일에 1회 이상 신문(인터넷 신문 포함)을 보신 적이 있습니까? 있다면 어떤 형태의 신문을 얼마나 자주 보셨습니까? (신문지 및 인터넷으로 볼 때 보는 경우 각각에 체크하여 주십시오.)

☐ 보지 않음

신문형태	보는 횟수			
	① 거의 매일	② 1주일에 3~4회	③ 1주일에 1~2회	④ 2주일에 1회
1.일반신문				
2.인터넷신문				

☐ 보지 않음

[그림 3-18] 독서

독서

27 귀하는 지난 1년 동안(2006. 6. 15. ~ 2007. 6. 14.) 참고서 및 참고서 제외를 읽어 보신 적이 있습니까? 있다면 각각에 대해 몇 권인지 모두 기입하여 주십시오.

☐ 보지 않음

- ① 공자유(주간, 월간, 기타) _____ 권
- ② 교양서적 (공과, 소설, 시집, 역사, 예술서적 등) _____ 권
- ③ 직업(직무)과 관련한 서적류 _____ 권
- ④ 생활·취미·정보서적 (육아, 꽃꽂이, 정원, 낚시서적 등) _____ 권
- ⑤ 기타(당화, 무협지 등) _____ 권

☐ 보지 않음

[그림 3-19] 문화예술 및 스포츠 관람

문화예술 및 스포츠 관람

2번 귀하는 지난 1년 동안 (2006. 6. 15. ~ 2007. 6. 14.) 다음의 공연 및 스포츠를 관람하신 적이 있습니까? 있다면 각각에 대해 관람횟수를 모두 기입하여 주십시오.

☐ 있다

① 음악·연주회	_____	□□	번
② 연극·마당극·뮤지컬	_____	□□	번
③ 무용	_____	□□	번
④ 영화	_____	□□	번
⑤ 박물관(유물전시회 등 포함)	_____	□□	번
⑥ 미술관(공예, 서화, 전시회 포함)	_____	□□	번
⑦ 스포츠	_____	□□	번

☐ 없다

상기 변수를 이용한 회귀분석에 앞서 표본추출을 제외한 90% 및 80% 자료에 대해 stepwise 변수선택 방법을 적용하여 변수를 선택하였으며, 선택된 변수(6개)는 다음과 같다.

- 기본항목:

· 연령[bun_g], 종사상의 지위[bun_j], 교육정도[bun_h]

- 조사항목

· 전반적인 생활여건의 변화[h6-1], 독서[h27], 문화예술 및 스포츠 [h29], 신문구독[h24]

먼저 통계적으로 유의한 것으로 나타난 변수들을 이용하여 로지스틱 회귀모형을 도출하였다. 도출된 회귀모형에 가상의 결측값(표본추출한 10% 및 20%)을 종속변수로 하여 대체값을 산출하였다. 구한 대체값과 실제값을 비교한 결과 다음과 같이 나타났다.

□ 10%: 응답자별 실제값과 대체값 일치율 46.9%

<표 3-13> [40번 문항] 통계량 및 Imputation 전·후 비교(1)

Parameter	DF	Estimate	Standard Wald 95% Confidence			Pr > ChiSq	
			Error	Limits	Square		
Intercept	1	3.5355	0.0261	3.4844	3.5866	18398.6	<.0001
bun_g	105	-1.1005	0.0471	-1.1929	-1.0081	544.99	<.0001
bun_g	115	-0.9898	0.0266	-1.0419	-0.9378	1388.51	<.0001
bun_g	120	-0.8969	0.0232	-0.9423	-0.8514	1494.52	<.0001
bun_g	130	-0.8218	0.0222	-0.8652	-0.7784	1375.21	<.0001
bun_g	135	-0.7048	0.0210	-0.7459	-0.6637	1130.28	<.0001
bun_g	145	-0.5565	0.0203	-0.5962	-0.5168	753.21	<.0001
bun_g	150	-0.3976	0.0194	-0.4357	-0.3596	418.65	<.0001
bun_g	160	-0.2136	0.0198	-0.2525	-0.1748	116.18	<.0001
bun_g	165	-0.0287	0.0206	-0.0691	0.0117	1.93	0.1643
bun_g	170	-0.0162	0.0214	-0.0582	0.0257	0.58	0.4478
bun_j	710	0.0608	0.0171	0.0274	0.0942	12.70	0.0004
bun_j	720	0.1364	0.0174	0.1023	0.1704	61.56	<.0001
bun_j	730	0.2476	0.0199	0.2086	0.2866	154.51	<.0001
bun_j	740	0.0973	0.0227	0.0528	0.1419	18.38	<.0001
bun_j	750	0.2925	0.0163	0.2605	0.3245	321.21	<.0001
bun_j	760	0.2071	0.0404	0.1278	0.2863	26.24	<.0001
bun_h	250	0.1425	0.0185	0.1063	0.1788	59.48	<.0001
bun_h	260	0.2022	0.0174	0.1680	0.2363	134.61	<.0001
bun_h	270	0.1273	0.0112	0.1054	0.1492	129.63	<.0001
h27	1	-0.0850	0.0108	-0.1062	-0.0637	61.51	<.0001
h29	1	-0.0846	0.0102	-0.1046	-0.0645	68.29	<.0001
h24	1	-0.0689	0.0114	-0.0913	-0.0465	36.41	<.0001
h6E1	1	-0.6303	0.0280	-0.6852	-0.5755	506.78	<.0001
h6E1	2	-0.4797	0.0178	-0.5146	-0.4448	726.64	<.0001
h6E1	3	-0.2950	0.0168	-0.3278	-0.2621	309.54	<.0001
h6E1	4	-0.1379	0.0184	-0.1739	-0.1019	56.45	<.0001

		실제값					
Imputation	빈도 백분율 행 백분율 칼럼 백분율	1	2	3	4	5	총합
	2	2	124	743	305	44	14
		2.90	17.38	7.13	1.03	0.33	28.77
		10.08	60.41	24.80	3.58	1.14	
		64.25	49.87	19.24	6.33	4.49	
3	3	68	715	1,057	445	196	2,481
		1.59	16.73	24.73	10.41	4.58	58.04
		2.74	28.82	42.60	17.94	7.90	
		35.23	47.99	66.69	64.03	62.82	
4	4	1	32	223	206	102	564
		0.02	0.75	5.22	4.82	2.39	13.19
		0.18	5.67	39.54	36.52	18.09	
		0.52	2.15	14.07	29.64	32.69	
총합		193	1,490	1,585	695	312	4,275
		4.51	34.85	37.08	16.26	7.30	100.00

□ 20% : 응답자별 실제값과 대체값 일치율 48.2%

<표 3-14> [40번 문항] 통계량 및 Imputation 전·후 비교(2)

Parameter	DF	Estimate	Standard Wald 95% Confidence			Pr > ChiSq	
			Error	Limits	Square		
Intercept	1	3.5472	0.0276	3.4931	3.6012	16,544.9	<.0001
bun_g	105	-1.0551	0.0499	-1.1528	-0.9573	447.62	<.0001
bun_g	115	-0.9789	0.0281	-1.0339	-0.9239	1,216.13	<.0001
bun_g	120	-0.8980	0.0246	-0.9462	-0.8497	1,329.82	<.0001
bun_g	130	-0.8203	0.0235	-0.8664	-0.7742	1,215.55	<.0001
bun_g	135	-0.7059	0.0223	-0.7495	-0.6623	1,006.40	<.0001
bun_g	145	-0.5670	0.0217	-0.6094	-0.5246	685.74	<.0001
bun_g	150	-0.3841	0.0207	-0.4247	-0.3435	343.45	<.0001
bun_g	160	-0.2081	0.0210	-0.2494	-0.1669	98.01	<.0001
bun_g	165	-0.0248	0.0219	-0.0678	0.0182	1.28	0.2584
bun_g	170	0.0061	0.0227	-0.0385	0.0507	0.07	0.7878
bun_j	710	0.0473	0.0181	0.0118	0.0827	6.82	0.0090

〈표 3-14〉의 계속

Parameter	DF	Estimate	Standard Wald 95% Confidence				Pr > ChiSq	
			Error	Limits		Square		
bun_j	720	1	0.1254	0.0184	0.0893	0.1615	46.30	<.0001
bun_j	730	1	0.2520	0.0212	0.2105	0.2935	141.81	<.0001
bun_j	740	1	0.0927	0.0243	0.0451	0.1403	14.58	0.0001
bun_j	750	1	0.2899	0.0174	0.2559	0.3239	278.97	<.0001
bun_j	760	1	0.1993	0.0425	0.1160	0.2825	22.00	<.0001
bun_h	250	1	0.1422	0.0196	0.1037	0.1806	52.40	<.0001
bun_h	260	1	0.2002	0.0185	0.1639	0.2364	117.13	<.0001
bun_h	270	1	0.1308	0.0118	0.1076	0.1540	121.91	<.0001
h27	1	1	-0.0738	0.0115	-0.0964	-0.0513	41.26	<.0001
h29	1	1	-0.0840	0.0109	-0.1053	-0.0627	59.87	<.0001
h24	1	1	-0.0698	0.0122	-0.0936	-0.0460	32.96	<.0001
h6E1	1	1	-0.6327	0.0298	-0.6911	-0.5743	451.04	<.0001
h6E1	2	1	-0.4994	0.0188	-0.5362	-0.4626	707.36	<.0001
h6E1	3	1	-0.3071	0.0177	-0.3418	-0.2725	301.36	<.0001
h6E1	4	1	-0.1590	0.0194	-0.1969	-0.1210	67.33	<.0001

실제값							
	행 백분율						총합
	칼럼 백분율	1	2	3	4	5	
Imputation	2	225	1470	579	90	24	2,388
		2.67	17.44	6.87	1.07	0.28	28.33
		9.42	61.56	24.25	3.77	1.01	
		60.48	51.38	18.32	6.11	4.26	
	3	137	1,322	2,132	919	339	4,849
		1.63	15.68	25.29	10.90	4.02	57.53
		2.83	27.26	43.97	18.95	6.99	
		36.83	46.21	67.45	62.43	60.21	
	4	10	69	450	463	200	1,192
		0.12	0.82	5.34	5.49	2.37	14.14
		0.84	5.79	37.75	38.84	16.78	
		2.69	2.41	14.24	31.45	35.52	
	총합	372	2,861	3,161	1,472	563	8,429
		4.41	33.94	37.50	17.46	6.68	100.00

[40번 문항]에 대한 분석 결과 응답자별 실제값과 대체값 일치율은 10% 가상결측의 경우 46.9%, 20% 가상결측의 경우 48.2%이다. 회귀모형의 의미는 “「장래 소득수준의 변화에 대한 기대」는 연령, 교육정도, 종사상의 지위, 전반적인 생활여건의 변화와 신문구독, 독서 경험유무 등 정보매체 접근성, 문화예술 및 스포츠 관람 등 여가와 관련이 있다”는 것을 나타내고 있다.

5. Imputation 전·후 비교

7. 귀하가 매월 납부하고 있는 사회보험료(건강보험, 국민연금, 고용보험 등)가 있습니까?

<표 3-15> [7번 문항] 사회보험료 납부여부

(단위: %)

구분	2007 조사결과	10% 대체	20% 대체
1. 있다	47.7	48.2	48.8
2. 없다	52.3	51.8	51.2

7-3. 고용·산재보험에 대하여 귀하의 소득에 비해 어느 정도 부담이 된다고 생각하시는지 체크하여 주십시오.

<표 3-16> [7-3번 문항] 고용·산재보험 부담에 대한 인식

(단위: %)

구분	2007 조사결과	10% 대체	20% 대체
1. 매우 부담됨	4.5	4.0	3.7
2. 약간 부담됨	12.8	11.6	10.3
3. 보통임	24.5	28.0	31.3
4. 별로 부담안됨	5.4	4.9	4.3
5. 전혀 부담안됨	1.1	1.0	0.9
6. 해당 없음	51.6	50.5	49.5

40. 장래 귀하의 소득수준은 어떻게 변화할 것이라고 생각하십니까?

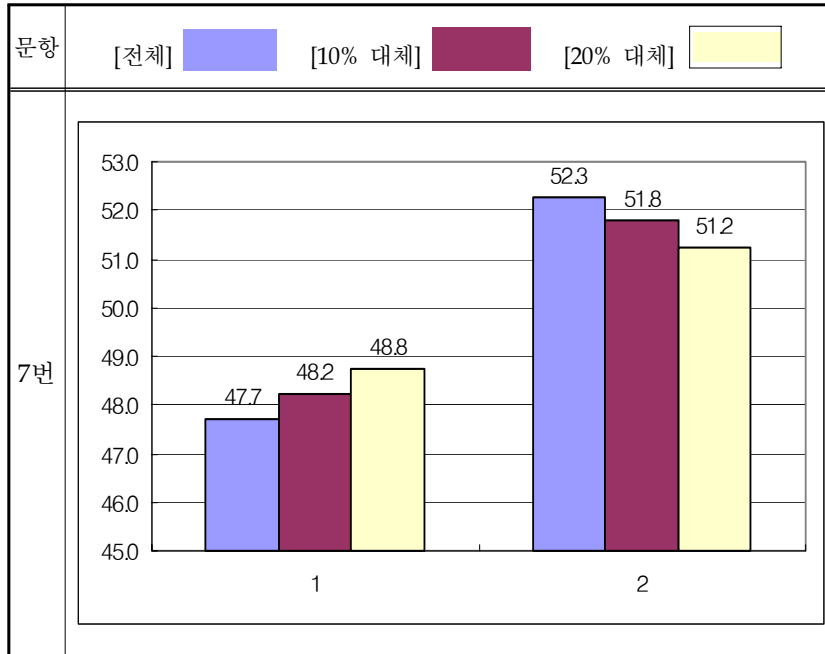
<표 3-17> [40번 문항] 장래소득에 대한 기대

(단위: %)

구 분	2007 조사결과	10% 대체	20% 대체
1. 크게 증가	4.7	4.2	3.8
2. 약간 증가	33.7	33.1	32.6
3. 동일	37.4	39.5	41.3
4. 약간 감소	17.4	17.1	16.7
5. 크게 감소	6.9	6.2	5.6

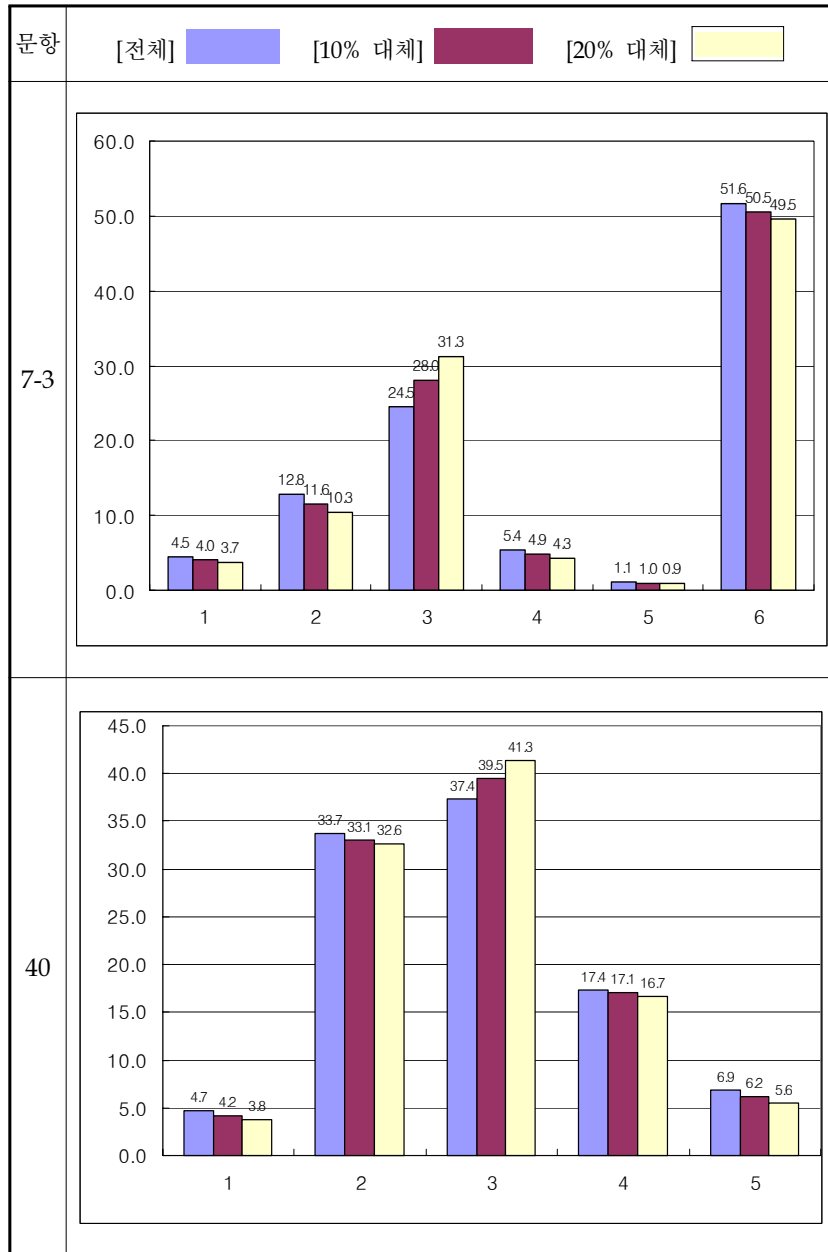
[그림 3-20] Imputation 전·후 비교

(단위: %)



[그림 3-20]의 계속

(단위: %)



제5절 최근방대체·평균대체 결과 비교

1. 비교 대체방법

대체법 중 회귀대체와 같은 단일대체방법의 하나인 평균대체, 최근방대체를 이용하여 비교하였다. 비교의 객관성 확보를 위해 서로 동일한 기본항목과 조사항목을 적용하였다.

〈표 3-18〉 기본항목과 조사항목

문항	기본항목		조사항목
	공통	추가	
7. 사회보험료 납부여부	중사상 지위	성별	현재소득 유무 노후준비 유무
7-3. 「고용·산재보험」 부담에 대한 인식		혼인상태	현재의 주된 소득
40. 「장래소득」에 대한 기대		연령	교육정도

2. 대체결과 비교

가. 카이제곱 검정

실제값과 대체값 간에는 대체방법에 관계없이 모두 유의한 관계가 있고, 관련성이 높은 것으로 나타났다. 카이제곱값을 비교해 보면 [7번 문항]은 평균대체법, 회귀대체법, 최근방대체법의 순으로, [7-3번 문항]은 평균대체법, 회귀대체법, 최근방대체법의 순으로, [40번 문항]은 회귀대체법, 평균대체법, 최근방대체법의 순으로 크게 나타났다.

〈표 3-19〉 대체방법별 카이제곱값 비교

문항	구분	표본크기	회귀대체	최근방대체	평균대체
7번	10%	6,665	3,435	2,796	3,442
	20%	13,301	6,727	5,658	6,734
7-3번	10%	3,070	1,017	794	1,191
	20%	5,935	1,974	1,418	2,265
40번	10%	4,275	1,014	377	841
	20%	8,429	2,115	719	1,869

주: 유의확률 <.0001.

나. 응답자별 실제값=대체값 일치율

응답자별 실제값과 대체값 간의 일치율을 비교하면 [7번 문항]은 평균대체법, 회귀대체법, 최근방대체법의 순으로, [7-3번 문항]은 회귀대체법, 최근방대체법, 평균대체법의 순으로, [40번 문항]은 회귀대체법, 평균대체법, 최근방대체법의 순으로 높게 나타났다.

〈표 3-20〉 대체방법별 일치율 비교

(단위: %)

문항	가상 결측값 10%			가상 결측값 20%		
	회귀대체	최근방대체	평균대체	회귀대체	최근방대체	평균대체
7	85.7	82.4	86.8	85.3	82.7	86.5
7-3	55.8	51.6	38.7	55.8	51.6	40.1
40	46.9	38.2	45.4	48.2	37.8	47.4

다. 평균 및 표준편차

최근방대체는 회귀대체 및 평균대체보다 표준편차가 다소 크게 나타났다.

〈표 3-21〉 대체방법별 평균 및 표준편차 비교

문항	구분	실제값	회귀대체	최근방대체	평균대체
7번	평균	1.48 / 1.47	1.48 / 1.47	1.53 / 1.52	1.52 / 1.54
	표준편차	0.50 / 0.50	0.50 / 0.50	0.50 / 0.50	0.50 / 0.50
7-3번	평균	4.32 / 4.35	4.12 / 4.16	4.38 / 4.42	4.28 / 4.29
	표준편차	1.75 / 1.75	1.45 / 1.46	1.78 / 1.74	1.17 / 1.20
40번	평균	2.87 / 2.88	2.76 / 2.77	2.88 / 2.89	2.80 / 2.83
	표준편차	0.98 / 0.97	0.64 / 0.64	0.99 / 0.98	0.65 / 0.67

주: 가상 결측값 10% / 가상 결측값 20%.

제6절 결론 및 향후 과제

1. 결론

사회통계조사 항목무응답/결측값에 대한 Imputation기법활용 연구를 통해 다음과 같은 결론을 제시하고자 한다.

첫째, imputation기법 활용은 일관된 수정논리에 따라 무응답/결측값에 대해 대체가 가능하여짐으로써 효율적인 자료처리로 인해 통계자료의 유용성을 높일 수 있다.

둘째, 여러 번 방문에서도 응답을 거부한 응답자에 대해 조사방문 횟수를 줄일 수 있어 재조사에 따른 시간과 비용을 절감할 수 있어 표본변경과 조사의 확대 등에 맞춰 효율적인 조사로의 유도가 가능하며, 자료집계가 빨라져 통계자료이용의 시의성을 높일 수 있다.

John Kovar & Eric Rancourt에 의하면 Editing 비용은 기본 인건비, 재조사 비용 등으로 미국의 경우 총조사 비용의 20~40%에 달한다고 한다(우리나라의 경우 10% 내외 수준 부담).¹¹⁾ 2005년 인구주택총조사의 경우를 살펴보면 총 1,290억원 중 약 70%(894억원)가 인건비이며, 인건

11) 변중석, "Introduction to Data Edition", 「2007년 통계의 날 기념 Workshop 자료집」, 통계청, 2007.

비 중 내검요원에 소요된 비용은 24억원으로 2.7%를 차지하였다.¹²⁾

셋째, 응답기피항목에 대한 항목무응답 허용은 응답자에게 응답강제로부터 벗어나게 함으로써 사회적인 바람직성이 아닌 본인의 의사를 정확하게 응답할 수 있다. 또한 타 조사항목에 대해 이로 인한 영향력을 배제할 수 있어 통계자료에 대한 신뢰성과 정확성을 높이는 등 조사자료의 품질향상을 도모할 수 있다.

넷째, 회귀분석을 통한 대체법은 주관적 의식조사인 사회통계조사의 경우 인과성 도출에 유용하여 시험조사에서 나타나지 않은 특정항목에 대한 무응답 원인을 심도있게 분석할 수 있고, 도출된 정보를 이용하여 조사표 설계에 유용하게 활용할 수 있다.

다섯째, 사회통계조사 항목무응답을 중심으로 처리단계별로 기술하여 사회통계조사 등 실무에 직접 적용할 수 있다.

여섯째, 인구주택총조사뿐만 아니라 가계조사, 가계자산조사 등 가구조사에 대한 Imputation기법 선행연구로서 활용할 수 있다.

2. 향후 과제

앞으로 해결 및 연구하여야 할 과제는 다음과 같다.

첫째, 본 연구가 기존 조사결과를 활용하여 임의적인 무응답항목을 설정하여 분석하는 제한적인 이론적 연구였다면, 앞으로의 연구는 실제 조사에서 발생한 무응답층에 대한 1차 조사 결과와 내검단계를 거쳐 수정·보완된 최종자료를 가지고 분석하는 실증적 연구가 되어야 한다.

둘째, 항목무응답을 허용하는 경우 조사표 회수율이 급격히 떨어질 수 있는 문제점이 있으므로 항목무응답 최소화 방안에 대한 연구가 필요하다

셋째, 무응답항목에 대한 원인 분석을 통하여 해결방안 모색이 필요하다. 즉, 어떤 문제에 대해 특별한 관심을 갖는 집단에게 응답이 잠재적으로 이익이 될 수 있다는 동기유발 및 홍보방안 등에 대한 연구가 필요하다.

12) 통계청, 「2005 인구주택총조사 종합평가보고서(1)」, 2006.

참고문헌

- 김규성(2000), “무응답 대체방법과 대체 효과”.
- 김규성 · 이기재 · 김진(2005), “농어가경제조사에서 가중하택 무응답 대체법의 활용”.
- 김동욱 · 노영화(2003), “대체방법 G--추정량 비교”.
- 김미정 · 변종석 · 남궁평, “패널조사에서 Wav-무응답의 조정장법 비교”.
- 김세미 · 이석훈 · 윤연옥(2007), “다변량회귀나무를 이용한 무응답 가중치 조정층 형성”, www.nso.go.kr/attach/journal/10-2-8.pdf, 2007.11.28.
- 김영원 · 조선경(1996), “표본조사에서 항목 무응답 대체 방법”.
- 김유진 · 이승욱(1994), “표본조사에서 무응답처리를 위한 통계기법 고찰”.
- 김재광 · 한근식 · 윤연옥(2004), “가계조사 무응답 처리기법 연구”.
- 김진(2004), “농가경제조사에 대한 대체법 비교”.
- 나혜정(1993), “뉴욕시립대학 개방입학정책 연구에 관한 자료 중 무응답 자료 분석에 -M 연산방식을 이용한 사례”.
- 박태성 · 이승연(1998), “무응답을 포함하는 범주형 자료의 분석”.
- 배성우 · 오형재(1999), “표본조사에서 무응답자료처리를 위한 임putation 방법 등의 비교 연구”.
- 변종석(2007), “Introduction to Data Edition”, 「2007년 통계의 날 기념 Workshop 자료집」, 통계청.
- 서울지방통계청(2007), 「2007년 강남구의 사회통계조사 용역사업 결과 보고서」.
- 손창균 · 정훈조(2000), “무응답 상황하에서 최적추정량에 관한 연구”.
- 손창균 · 홍기학 · 이기성(2003), “단위무응답 보정에서 보조변수의 선택에 관한 연구”.
- 심미선 · 최현철(1997), “선거여론조사에서 나타난 무응답 연구”.
- 염준근 · 손창균(1998), “층화표본에서 단위무응답에 대한 가중치 조정 방법”.

- 윤성철(2004), “결측값의 대처법”.
- 이성용(1997), “사회조사의 총오차와 오차 유형에 따른 사회통계분석 모형들”.
- 이승욱 · 김유진(1997), “보건학 관련 범주형 자료에서 무응답처리를 위한 통계기법”.
- 이원철(2000), “로지스틱 회귀분석”, 「대한예방의학회 하계 워크샵 및 전공의 연수교육집」.
- 이정택 · 이병주 · 성수련 · 남궁문(2002), “노선배정을 이용한 무응답 조사 데이터 보정의 유효성”.
- 이진희 · 김진 · 이기재(2006), “표본조사에서 공간변수를 이용한 결측 대체의 효율성 비교”.
- 정영해(1995), “무응답에 대한 고찰: 조사연구기관의 표본조사 사례보고”.
- 정재구(1999), “무응답보정을 위한 「래킹」 비추정방법에 관한 연구”.
- 조사통계연구회(2000), 「무응답오차」, 자유아카데미.
- 통계청(2007), 「사회통계조사 실시계획안」.
- 한상열 · 성화경 · 최관(1998), “이선택형 가상가치평가법에서의 무응답 편익에 관한 실증적 규명”.
- 허석열 · 석정봉 · 한기준(1990), “무응답 Spanning Tr-- 구성 알고리즘”.
- Collins, L. M., J. L. Schafer, & C. M. Kam(2001), “A comparison of inclusive and restrictive strategies in modern missing data procedures”, *Psychological Methods*, 6, pp.330-351.
- Fichmam, M. & J. N. Cummings(2003), “Multiple imputation for missing data: Making the most of what you know”, *Organizational Research Methods*, 6, pp.282-308.
- Fisher, J.(2006), *Income Imputation and the Analysis of Expenditure Data in the Consumer Expenditure Survey*, Bureau of Labor Statistics.
- Harel, O. & J. L. Schafer, Multiple imputation in two stages.(in press)
- Kennickell, A. B.(1998), *Multiple Imputation in the Survey of Consumer Finances*, Board of Governors of the Federal Reserve System.

- Paulin, G., J. Fisher, & S. Reyes-Morales(2006), "User's guide to income imputation in the CE", Bureau of Labor Statistics.
- Penn, D. A.(2007), "Estimating missing values from the general social survey: An application of multiple imputation", *Social Science Quarterly*, 88(2), pp.573-584.
- Schafer, J. L.(1999), "Multiple imputation: a primer", *Statistical Methods in Medical Research*, 8, pp.3-15.
- Schafer, J. L. & M. K. Olsen(1998), "Multiple imputation for multivariate missing-data problems: A data analyst's perspective", *Multivariate Behavioral Research*, 33(4), pp.545-571.
- Sinharay. S., H. S. Ster, & D. Russell(2001), "The use of multiple imputation for the analysis of missing data", *Psychological Methods*, 6(4), pp.317-329.

< 부 록 >

1. 2007 사회통계조사 조사문항 및 설명

조사문항	설 명
	조사구번호
	구역번호
	거처번호
	가구번호
	가구원 일련번호
	성명
	가구주와의 관계
	성별
	생년월일
	나이
	교육정도
	혼인상태:경제활동상태
	혼인상태
	혼인신고여부
	경제활동상태
	산업
	직업
	종사상의 지위
	불응여부
6-1	6.생활여건의 변화 - 1.전반적인 생활여건
6-2	6.생활여건의 변화 - 2.보건의료서비스
6-3	6.생활여건의 변화 - 3.사회보장제도
6-4	6.생활여건의 변화 - 4.문화 여가생활 향유여건
7	7.사회보험료 부담에 대한 인식
7-1	7.사회보험료 부담에 대한 인식 - 1.건강보험
7-2	7.사회보험료 부담에 대한 인식 - 2.국민연금
7-3	7.사회보험료 부담에 대한 인식 - 3.고용·산재보험
7-4	7.사회보험료 부담에 대한 인식 - 4.기타
7D1	7.사회보험료 부담에 대한 인식 - 기타서술
8-1	8.향후 늘려야 할 공공시설 - 1순위
8-2	8.향후 늘려야 할 공공시설 - 2순위
8-3	8.향후 늘려야 할 공공시설 - 3순위
8D1	8.향후 늘려야 할 공공시설 - 기타서술

1-1. 2007 사회통계조사 조사문항 및 설명 (계속)

조사문항	설 명
9-1	9.향후 늘려야 할 복지서비스 - 1순위
9-2	9.향후 늘려야 할 복지서비스 - 2순위
9-3	9.향후 늘려야 할 복지서비스 - 3순위
9D1	9.향후 늘려야 할 복지서비스 - 기타서술
10	10.장애인 복지카드 소유여부
10-1	10.장애인 복지카드 소유여부 - 장애등급
11	11.장애인 차별정도 - 우리 사회
12	12.장애인 차별정도 - 자신
13	13.장애인에 대한 견해
14	14.장애인 관련 시설
15	15.장애인 복지사업 - 확대에 대한 견해
15-1-1	15-1.장애인 복지사업 - 우선 확대 사업 - 1순위
15-1-2	15-1.장애인 복지사업 - 우선 확대 사업 - 2순위
15-1-3	15-1.장애인 복지사업 - 우선 확대 사업 - 3순위
15-1D1	15-1.장애인 복지사업 - 우선 확대 사업 - 기타서술
16	16.장애인을 위한 자원봉사활동
16-1	16.장애인을 위한 자원봉사활동 - 1.복지시설 등
16-2	16.장애인을 위한 자원봉사활동 - 2.일반 가정 등
16-3	16.장애인을 위한 자원봉사활동 - 3.병원 등
16-4	16.장애인을 위한 자원봉사활동 - 4.기타
16D1	16.장애인을 위한 자원봉사활동 - 기타서술
17	17.향후 봉사활동 의사
18	18.본인의 노후 준비 방법
18-11	18.본인의 노후 준비 방법 - 1.준비하고(되어)있다 - 가장 주된 것
18-12	18.본인의 노후 준비 방법 - 1.준비하고(되어)있다 - 부수적인 것
18D1	18.본인의 노후 준비 방법 - 1.준비하고(되어)있다 - 기타서술
18-2	18.본인의 노후 준비 방법 - 2.준비하고(되어)있지 않다
19-1	19.노인이 경험하는 어려움 - 1순위
19-2	19.노인이 경험하는 어려움 - 2순위
19-3	19.노인이 경험하는 어려움 - 3순위
19D1	19.노인이 경험하는 어려움 - 기타서술
20	20.생활비 마련방법
20-1	20.생활비 마련방법 - 1.본인 및 배우자 부담
20-2	20.생활비 마련방법 - 2.자녀 또는 친척 지원
20D1	20.생활비 마련방법 - 기타서술

1-2. 2007 사회통계조사 조사문항 및 설명 (계속)

조사문항	설 명
21	21.현재 자녀와 동거 여부 및 이유
21-1	21.현재 자녀와 동거 여부 및 이유 - 1순위
21-2	21.현재 자녀와 동거 여부 및 이유 - 2순위
21-3	21.현재 자녀와 동거 여부 및 이유 - 3순위
21D1	21.현재 자녀와 동거 여부 및 이유 - 기타서술
22	22.향후 자녀와 동거 의향
22-1	22.향후 자녀와 동거 의향 - 2.같이 살고 싶지 않다
22D1	22.향후 자녀와 동거 의향 - 2.같이 살고 싶지 않다 - 기타서술
23	23.받고 싶은 복지서비스
23-1	23.받고 싶은 복지서비스 - 1순위
23-2	23.받고 싶은 복지서비스 - 2순위
23-3	23.받고 싶은 복지서비스 - 3순위
23D1	23.받고 싶은 복지서비스 - 기타서술
24	24.신문(인터넷 신문 포함)
24-1	24.신문(인터넷 신문 포함) - 1.일반신문
24-2	24.신문(인터넷 신문 포함) - 2.인터넷신문
25	25.TV시청
25-1	25.TV시청 - 1.평일
25-2	25.TV시청 - 2.토요일
25-3	25.TV시청 - 3.일요일 공휴일
26	26.비디오(DVD포함)
26-1	26.비디오(DVD포함) - 1.오락용
26-2	26.비디오(DVD포함) - 2.교육용
26-3	26.비디오(DVD포함) - 3.교양용
26-4	26.비디오(DVD포함) - 4.기타
26D1	26.비디오(DVD포함) - 기타서술
27	27.독서
27-1	27.독서 - 1.잡지류
27-2	27.독서 - 2.교양서적
27-3	27.독서 - 3.직업(직무)과 관련한 서적류
27-4	27.독서 - 4.생활 취미 정보서적
27-5	27.독서 - 5.기타
28	28.레저시설 이용횟수
28-1	28.레저시설 이용횟수 - 1.관광명소
28-2	28.레저시설 이용횟수 - 2.온천장 및 스파

1-3. 2007 사회통계조사 조사문항 및 설명 (계속)

조사문항	설 명
28-3	28.레저시설 이용횟수 - 3.골프장
28-4	28.레저시설 이용횟수 - 4.스키장
28-5	28.레저시설 이용횟수 - 5.해수욕장
28-6	28.레저시설 이용횟수 - 6.산림욕장
28-7	28.레저시설 이용횟수 - 7.놀이공원
28-8	28.레저시설 이용횟수 - 8.기타
28D1	28.레저시설 이용횟수 - 기타서술
29	29.문화예술 및 스포츠 관람
29-1	29.문화예술 및 스포츠 관람 - 1.음악 연주회
29-2	29.문화예술 및 스포츠 관람 - 2.연극 마당극 뮤지컬
29-3	29.문화예술 및 스포츠 관람 - 3.무용
29-4	29.문화예술 및 스포츠 관람 - 4.영화
29-5	29.문화예술 및 스포츠 관람 - 5.박물관
29-6	29.문화예술 및 스포츠 관람 - 6.미술관
29-7	29.문화예술 및 스포츠 관람 - 7.스포츠
30-1	30.주말이나 휴일의 여가활용 - 1순위
30-2	30.주말이나 휴일의 여가활용 - 2순위
30-3	30.주말이나 휴일의 여가활용 - 3순위
30D1	30.주말이나 휴일의 여가활용 - 기타서술
30-1	30-1.주말이나 휴일의 여가활용 - 누구와 함께
30-ID1	30-1.주말이나 휴일의 여가활용 - 누구와 함께 - 기타서술
31-1	31.앞으로 하고 싶은 여가활동 - 1순위
31-2	31.앞으로 하고 싶은 여가활동 - 2순위
31-3	31.앞으로 하고 싶은 여가활동 - 3순위
31D1	31.앞으로 하고 싶은 여가활동 - 기타서술
32	32.여가활용 만족 여부
32-1	32.여가활용 만족 여부 - 2.불만족 이유
32D1	32.여가활용 만족 여부 - 2.불만족 이유 - 기타서술
33	33.국내 관광여행 횟수
33-1	33.국내 관광여행 횟수 - 1.숙박여행
33-11	33.국내 관광여행 횟수 - 1.숙박여행 - 1회당 평균
33-2	33.국내 관광여행 횟수 - 2.당일여행
34	34.여행지에서 불편한 점
34-1	34.여행지에서 불편한 점 - 1순위
34-2	34.여행지에서 불편한 점 - 2순위

1-4. 2007 사회통계조사 조사문항 및 설명 (계속)

조사문항	설 명
34-3	34.여행지에서 불편한 점 - 3순위
34D1	34.여행지에서 불편한 점 - 기타서술
35	35.해외여행 경험 및 횟수
35-11	35.해외여행 경험 및 횟수 - 1.관광
35-12	35.해외여행 경험 및 횟수 - 2.가사
35-13	35.해외여행 경험 및 횟수 - 3.업무
35-14	35.해외여행 경험 및 횟수 - 4.어학연수(6개월 미만)
35-15	35.해외여행 경험 및 횟수 - 5.어학연수(6개월 이상)
35-2	35.해외여행 경험 및 횟수 - 과거 해외여행 경험
36	36.지역문화 예술행사
36-1	36.지역문화 예술행사 - 만족도
36-1-1	36-1.지역문화 예술행사 - 참여하지 않는 이유 - 1순위
36-1-2	36-1.지역문화 예술행사 - 참여하지 않는 이유 - 2순위
36-1-3	36-1.지역문화 예술행사 - 참여하지 않는 이유 - 3순위
36-ID1	36-1.지역문화 예술행사 - 참여하지 않는 이유 - 기타서술
36-2-1	36-2.지역문화 예술행사 - 불만족 이유 - 1순위
36-2-2	36-2.지역문화 예술행사 - 불만족 이유 - 2순위
36-2-3	36-2.지역문화 예술행사 - 불만족 이유 - 3순위
36-2D1	36-2.지역문화 예술행사 - 불만족 이유 - 기타서술
37	37.전통문화에 대한 관심
37-1	37.전통문화에 대한 관심 - 가장 관심 있는 분야
37D1	37.전통문화에 대한 관심 - 가장 관심 있는 분야 - 기타서술
38	38.소득만족도 - 소득유무
38-1	38.소득만족도 - 만족도
39	39.현재 소득 수준에 대한 인식 - 주된 소득
39-1	39-1.현재 소득 수준에 대한 인식 - 소득 수준 인식
40	40.장래 소득에 대한 기대
41	41.소득분배에 대한 견해
42	42.소비생활만족도
43	43.관리사항(가구소득)
44	44.아동수
44-1	44.아동수 - 총명수
GRAD-	44.아동수 - 학년
45	45.방과후 함께 지내는 사람
45-1	45.방과후 함께 지내는 사람 - 주로 돌보는 사람

1-5. 2007 사회통계조사 조사문항 및 설명 (계속)

조사문항	설 명
45D1	45.방과후 함께 지내는 사람 - 기타서술
46-1	46.방과후 시간의 활용 - 1.학습관련 활동
46-2	46.방과 후 시간의 활용 - 2.학습 외 활동
47	47.학습 방법
47D1	47.학습 방법 - 기타서술
48	48.해외여행 경험
48-1	48.해외여행 경험 - 1.관광 및 여행
48-2	48.해외여행 경험 - 2.어학연수
48-3	48.해외여행 경험 - 3.부모의 근무지 이동
48-4	48.해외여행 경험 - 4.기타
48D1	48.해외여행 경험 - 기타서술
49-1	49.놀이 방법 및 장소 - 놀이 - 1순위
49-2	49.놀이 방법 및 장소 - 놀이 - 2순위
49-3	49.놀이 방법 및 장소 - 놀이 - 3순위
49D1	49.놀이 방법 및 장소 - 놀이 - 기타서술
49-1	49-1.놀이 방법 및 장소 - 장소
49-1D1	49-1.놀이 방법 및 장소 - 장소 - 기타서술
50	50.결식 여부 및 식사제공 기관 - 결식 경험
50-1-1	50-1.결식 여부 및 식사제공 기관 - 1.학기 중
50-1D1	50-1.결식 여부 및 식사제공 기관 - 1.학기 중 - 기타서술
50-1-2	50-1.결식 여부 및 식사제공 기관 - 2.방학 휴일
50-1D2	50-1.결식 여부 및 식사제공 기관 - 2.방학 휴일 - 기타서술
51	51.아동 어머니의 동거 및 취업 여부 - 동거
51-1	51-1.아동 어머니의 동거 및 취업 여부 - 취업
51-1D1	51-1.아동 어머니의 동거 및 취업 여부 - 취업 - 기타서술
52	52.안전한 놀이 장소
53-1	53.아동 양육의 어려운 점 - 1순위
53-2	53.아동 양육의 어려운 점 - 2순위
53-3	53.아동 양육의 어려운 점 - 3순위
53D1	53.아동 양육의 어려운 점 - 기타서술

2. 카이제곱 검정 - 반응변수: [7번 문항]

2-1. 기본항목

설명변수	카이제곱값	유의확률	크래머의 V	표본크기
종사상의 지위	31,843	<.0001	0.7288	59,946
직업	26,359	<.0001	0.6631	59,946
산업	25,164	<.0001	0.6479	59,946
경제활동상태	22,605	<.0001	0.6141	59,946
연령	10,729	<.0001	0.4231	59,946
성별	8,089	<.0001	0.3673	59,946
교육정도	5,973	<.0001	0.3157	59,946
혼인상태	2,816	<.0001	0.2167	59,946
가구소득	1,894	<.0001	0.1777	59,946
세대구분	331	<.0001	0.0743	59,946
주택형태	310	<.0001	0.0719	59,946
가구원수	194	<.0001	0.0568	59,946
점유형태	186	<.0001	0.0557	59,946
시도	131	<.0001	0.0468	59,946

2-2. 조사항목

설명변수	카이제곱값	유의확률	크래머의 V	표본크기
38	22,879	<.0001	0.6178	59,946
18	11,040	<.0001	0.4409	56,803
39	5,448	<.0001	0.3767	38,388
27-3	2,804	<.0001	0.2894	33,480
24	2,514	<.0001	0.2048	59,946
18-11	2,366	<.0001	0.2596	35,098
30-1	1,981	<.0001	0.1818	59,946
43	1,894	<.0001	0.1777	59,946
31-1	1,806	<.0001	0.1736	59,946
36_1-1	1,734	<.0001	0.1964	44,979
25-1	1,425	<.0001	0.1573	57,607
32-1	1,370	<.0001	0.1714	46,645
40	1,235	<.0001	0.1794	38,388
8-1	1,110	<.0001	0.1361	59,946
17	1,109	<.0001	0.136	59,946

2-2-1. 조사항목 (계속)

설명변수	카이제곱값	유의확률	크래머의 V	표본크기
20	1,109	<.0001	0.3008	12,258
31-2	834	<.0001	0.1296	49,663
35-13	816	<.0001	0.3278	7,592
44	734	<.0001	0.1633	27,520
18-2	692	<.0001	0.1785	21,705
30-2	661	<.0001	0.1147	50,189
37-1	580	<.0001	0.1328	32,875
15	545	<.0001	0.0954	59,946
24-1	527	<.0001	0.1364	28,337
34	483	<.0001	0.1209	33,039
28-3	481	<.0001	0.1121	38,241
27-5	472	<.0001	0.1187	33,480
35	463	<.0001	0.0879	59,946
18-12	441	<.0001	0.139	22,816
27-2	426	<.0001	0.1128	33,480
9-1	406	<.0001	0.0823	59,946
36_1-2	386	<.0001	0.1213	26,230
28	385	<.0001	0.0802	59,946
28-6	380	<.0001	0.0996	38,241
27	377	<.0001	0.0793	59,946
25-2	373	<.0001	0.0804	57,607
33	372	<.0001	0.0787	59,946
6-1	356	<.0001	0.0771	59,946
33-2	346	<.0001	0.1024	33,039
24-2	345	<.0001	0.125	22,102
30_1	339	<.0001	0.0752	59,946
6-3	332	<.0001	0.0745	59,946
29-7	316	<.0001	0.1008	31,041
28-1	301	<.0001	0.0887	38,241
15_1-1	271	<.0001	0.0736	50,059
20-1	264	<.0001	0.1868	7,571
19-1	261	<.0001	0.146	12,258
29	258	<.0001	0.0657	59,946

2-2-2. 조사항목 (계속)

설명변수	카이제곱값	유의확률	크래머의 V	표본크기
37	252	<.0001	0.0649	59,946
8-2	251	<.0001	0.0698	51,431
33-1	247	<.0001	0.0864	33,039
35-11	245	<.0001	0.1797	7,592
23-1	245	<.0001	0.1624	9,285
28-4	227	<.0001	0.0771	38,241
6-4	226	<.0001	0.0614	59,946
31-3	203	<.0001	0.0896	25,254
26-2	192	<.0001	0.118	13,827
34-1	175	<.0001	0.0825	25,736
11	166	<.0001	0.0526	59,946

3. 스피어만 상관계수 - 반응변수 : [7-3번 문항]/[40번 문항]

3-1. 기본항목

7-3번 문항			40번 문항		
설명변수	표본크기	ρ	설명변수	표본크기	ρ
종사상지위	31,792	0.5512	연령	42,663	0.4719
연령	31,792	0.3456	종사상지위	42,663	0.3086
혼인상태	31,792	0.2353	직업	42,663	0.2742
경제활동상태	31,792	0.1778	혼인상태	42,663	0.2631
직업	31,792	0.1445	경제활동상태	42,663	0.0988
시도	31,792	0.0802	시도	42,663	0.0923
접유형태	31,792	-0.0067	성별	42,663	0.0322
주택유형	31,792	-0.0320	접유형태	42,663	-0.1176
가구주관계	31,792	-0.0882	세대구분	42,663	-0.1299
교육정도	31,792	-0.1878	가구주관계	42,663	-0.1401
가구소득	31,792	-0.2139	가구소득	42,663	-0.2381
성별	31,792	-0.2159	교육정도	42,663	-0.3881

3-2. 조사항목

7-3번 문항			40번 문항		
설명변수	표본크기	ρ	설명변수	표본크기	ρ
39	30,253	0.4573	18-2	12,078	0.3113
7-2	31,792	0.4052	27	42,663	0.2986
18-2	5,408	0.2260	29	42,663	0.2934
18	31,782	0.2118	24	42,663	0.2718
40	30,253	0.1931	6-1	42,663	0.2499
20-1	2,968	0.1889	26	42,663	0.2294
29	31,792	0.1688	39	42,663	0.2228
38	31,792	0.1657	42	42,663	0.2098
27	31,792	0.1467	28	42,663	0.1951
20	3,430	0.1455	6-4	42,663	0.1943
26	7,742	0.1450	7-3	30,253	0.1931
24	13,207	0.1363	17	42,663	0.1881
6-1	31,792	0.1244	38-1	42,663	0.1684
28	31,792	0.1166	44	28,264	0.1569
21	3,430	0.1118	34	24,353	0.1563
25-1	30,620	0.0975	39-1	42,663	0.1563
34	18,859	0.0973	33	42,663	0.1525
18-11	26,374	0.0932	7	42,663	0.1418
33	31,792	0.0894	18	42,591	0.1289
17	31,792	0.0864	15	42,663	0.1246
26-3	7,742	0.0798	7-4	30,253	0.1174
35-11	5,008	0.0796	25-1	41,080	0.1158
44	22,826	0.0795	37	42,663	0.1053
42	31,792	0.0790	31-2	35,521	0.1010
6-4	31,792	0.0711	52	5,349	0.1010
29-1	17,551	0.0655	26-3	9,413	0.0986
52	5,007	0.0650	36	42,663	0.0986
18-12	18,187	0.0609	7-1	30,253	-0.0773
36-1-1	23,532	0.0603	32-1	33,263	-0.0773

3-2-1. 조사항목 (계속)

7-3번 문항			40번 문항		
설명변수	표본크기	ρ	설명변수	표본크기	ρ
35-12	5,008	0.0568	15-1-1	36,172	-0.0787
35	31,792	0.0551	28-6	27,805	-0.0837
15	31,792	0.0527	26-1	9,413	-0.0913
12	31,792	-0.0523	27-5	23,517	-0.0971
28-3	21,589	-0.0582	9-1	42,663	-0.1149
27-5	19,063	-0.0587	19-1	8,208	-0.1194
9-1	31,792	-0.0605	11	42,663	-0.1303
21-2	1,458	-0.0622	20-2	1,724	-0.1340
27-3	19,063	-0.0660	37-1	23,785	-0.1418
37-1	18,507	-0.0769	20	8,208	-0.1552
28-5	21,589	-0.0825	28-4	27,805	-0.1605
11	31,792	-0.0939	20-1	5,996	-0.1672
33-1	18,859	-0.0960	28-5	27,805	-0.1816
28-4	21,589	-0.1002	35-13	6,079	-0.1869
28-7	21,589	-0.1277	33-1	24,353	-0.2047
35-13	5,008	-0.1516	27-3	23,517	-0.2109
7-4	31,792	-0.1643	28-7	27,805	-0.2213
29-4	17,551	-0.1738	29-4	21,885	-0.2400