

제2장

가계금융조사를 이용한 무응답 조정방법 연구

심규호 · 백지선

제1절 서론

1. 연구배경 및 목적

통계 조사에서의 오차는 크게 표본오차와 비표본오차로 구분할 수 있다. 표본오차는 모집단과 표본추출 방법에 따라 달라지며 비표본오차는 표본오차를 제외한 모든 오차를 말한다. 무응답은 비표본오차를 발생시키는 많은 원인중 하나이다. 통계조사를 수행할 때 표본을 모두 조사하려고 노력하는 것이 우선이지만 조사를 수행하다보면 많은 경우 완전한 자료를 얻지 못하는 경우가 발생하게 된다. 무응답은 표본의 크기를 축소시켜 추정값의 효율을 떨어뜨린다. 또한 무응답에 포함되는 표본들이 시스템적인 차이를 발생시킨다면 무응답 편향이 발생하여 추정값의 신뢰도에 문제가 생길 수 있다(송주현 외, 2009). 통계청 가계자산조사에 대해 김재광(2004)과 박성인(2007)은 무응답 처리방법에 대한 연구를 수행하여 무응답 편향을 조정한바 있다.

통계청에서는 가구의 실물 및 금융자산, 부채 등의 구성과 분포를 조사하여 각종 경제, 사회, 금융정책이나 학문연구의 기초자료로 제공하기 위해 2010년 1차 가계금융조사를 시작하였다. 이 조사는 패널조사로 가계부문의 미시적 재무건전성을 주기적으로 파악하여 정부와 금융당국의 부동산 정책, 통화 및 금리정책, 가계신용관리, 금융시스템 발전 등을 위한 기본정보로 활용하게 된다. 조사주기는 1년이며 모집단은 2005년 인구주택총조사 결과 일반조사구와 총조사 이후 신축된 아파트가 대상이 된다. 표본규모는 전국의 약 10,000가구이며 주거유형, 주택평균연건평으로 비례하여 확률비례계통추출을



하였다. 이후 2011년 2차 가계금융조사를 실시하였으며, 2012년 새로운 2만 가구를 표본으로 확대, 개편되어 가계금융·복지조사로 수행하게 되었다.

본 연구에서는 무응답 가구 특성과 편향 정도, 무응답 조정 효과를 파악하기 위하여 1, 2차 가계금융조사 자료를 이용하였다. 2차 가계금융조사의 응답정보를 이용하여 1차 가계금융조사에서 비협조 성향이 강한 응답가구를 무응답가구로 재분류하는 연구설계와 무응답 조정을 실시하고, 1차 가계금융조사 1만 가구 대상 결과와 비교분석을 실시하였다. 무응답으로 가정된 가구는 2010년 응답을 하고 1년 뒤 조사에서 무응답으로 전환된 가구로 처음부터 응답하지 않은 가구보다는 무응답 성향이 약할 수 있다는 한계점을 가지지만, 무응답에 대한 계량적 분석을 통해 무응답 조정에 유의미한 정보를 제공할 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 1절에서 자료와 연구설계 등 연구방법을 설명하였다. 2절에서는 무응답 가구의 특성에 대해 알아보았으며 연구설계에 의해 얻어진 각각의 무응답 조정방법에 대한 추정값을 비교 분석하였다. 3절에서는 연구내용을 요약하고 결론을 제시하였다.

2. 연구 방법

무응답 조정에 유의미한 변수와 효과를 분석해 보기 위하여 본 연구에서는 다음과 같은 실험설계를 하였다. 가계금융조사가 패널조사임을 이용하여 2차 가계금융조사에서 무응답한 가구에 해당하는 1차 가계금융조사의 응답 가구를 비교적 비협조 성향이 강한 가구로 보고 무응답 가구로 설정하였다. 즉, 2010년에 조사된 가계금융조사 결과에 동일 가구에 대한 2011년 가계금융조사 응답정보를 이용하여 무응답 가구를 생성하였다. 이러한 실험설계를 통하여 2010년 1만 가구에 대한 가계금융조사를 준거(reference)로 두고, 1만 가구 중 응답가구는 8,918가구, 무응답 가구는 1,026가구로 분류하였다.

연구내용은 크게 3가지 단계로 구분된다. 1단계에서 무응답 가구의 특성과 가구의 응답여부와 관련된 가구 특성 및 금융변수를 파악하였다. 먼저 무응답 가구의 특성을 파악하기 위해 응답가구와 무응답가구에 대해 가구주 성별, 연령, 동거여부, 학력 등 가구주 특성변수와 금융변수인 자산, 부채, 소득 분포를 비교하였다. 무응답 조정에 필요한 응답여부와 관련된 가구 특성변수를 파악하기 위하여 가구특성과 응답유형의 분포에 대한 카이제곱 독립성 검정을 실시하고, 가구특성별 금융변수 평균을 비교하여 가구 특성변수와 금융변수의 관계를 살펴보았다. 2단계는 무응답 발생과 관련이 있을 것으로 판단되는 변수들(지역, 주택유형, 가구주 정보, 금융 정보)을 이용하여 무응답 조정에 대한

분석을 실시하였다. 무응답 조정층 작성을 위하여 표본 추출정보인 지역정보와 주택유형을 사용하였다. 다음으로 가구주 특성변수의 무응답 조정 효과를 분석하기 위하여 가구주 정보를 이용하여 무응답 조정층을 작성하였으며, 마지막으로 가구주 정보와 금융정보를 동시에 활용하여 무응답층을 작성하였다. 최종적으로 6가지의 무응답 층 작성 방법, “지역변수(수도권, 비수도권)와 주택유형, CHAID”, “지역변수(16개 시도)와 주택유형, CHAID”, “가구주정보, CHAID”, “가구주정보 와 금융정보, CHAID”, “금융정보, CHAID”, “지역변수(수도권, 비수도권)와 주택유형, Logistic”로 무응답 조정을 실시하였다. 3단계로 무응답 조정 방법에 따른 가중값 분포와 금융변수 추정결과를 비교하였다. 또한 무응답 조정 전과 6가지 무응답 조정 후 추정결과를 비교하여 무응답 조정으로 인한 편향 조정과 추정결과에 미치는 효과를 검토하였다.

제2절 가계금융조사 무응답 가중값 적용

가계금융조사의 무응답 가구 특성, 무응답 조정효과 등을 분석하기 위하여 본 연구에서는 2010년 가계금융조사(1만 가구)를 준거(reference)로 정의하고 2011년 가계금융조사 응답정보를 이용하여 2010년 가계금융조사의 응답가구와 무응답 가구를 생성하였다. 이러한 실험 하에서 우리는 응답가구와 무응답 가구의 특성을 분석하였다. 일반적으로 무응답에 영향을 미치는 가구 또는 가구주의 정보가 있을 것으로 기대한다. 조사자료를 바탕으로 한 가구의 특성에는 가구주 성별, 나이, 직업, 교육정도, 혼인상태 등과 가구의 외형적인 특성인 주택유형과 거주지역이 있다. 이 중에 가구주 정보는 조사를 직접 수행하여야만 얻을 수 있는 정보이고 주택유형과 거주지역과 같은 정보는 조사를 직접 하지 않아도 얻을 수 있는 정보이다. 무응답은 조사에 응답하지 않은 가구에서 발생하므로 가구주의 특성에 해당하는 다양한 정보는 사실상 얻을 수 없다.

그러나 본 연구에서는 2010년과 2011년의 자료를 모두 가지고 2011년의 응답 정보를 이용하여 2010년 자료에 무응답 가구를 생성하였기 때문에 가구와 가구주 정보를 온전히 사용하여 무응답 가구 특성과 응답여부에 영향을 주는 변수, 무응답 조정 효과를 파악할 수 있다. 앞서서도 언급하였듯이 여기서 무응답 가구는 2010년 응답을 하고 1년 뒤 조사에서 무응답으로 전환된 가구로 실제 무응답 가구와 유사한 성향을 가질 것으로 판단하고 분석을 실시하였다.

가계금융조사에서 주요 관심은 가구의 금융정보인 즉 자산, 소득, 부채 수준이다. 무응답 가구를 분석하고자 하는 목적은 응답가구의 금융수준과 무응답가구의 금융수



준이 차이가 있는지와 그 차이가 금융정보 추정결과에 얼마나 영향을 미치는 가를 알고자 하는데 있다. 또한 추정단계에서 두 가구 집단 간의 차이로 인한 영향을 최소화하기 위하여 실시하는 무응답 조정에 적절한 변수와 방법을 파악하는데 있다. 여기서 중요한 것은 금융정보에 대한 응답여부가 가구의 특성에 따라 발생할 수도 있지만 가구의 금융수준에 따른 차이에서도 발생할 수 있다는 것이다. 예를 들면 부채가 많고 소득이 적은 가구의 경우 통계조사에 응하고 싶지 않을 수도 있고 반대로 소득과 자산이 상대적으로 큰 가구의 경우도 세금과 같은 문제 때문에 금융상태의 노출을 꺼릴 수도 있다. 또한 금융상태가 좋은 가구는 면접조사와 같은 방식은 접촉 자체가 어려울 수도 있다. 따라서 무응답 조정층을 작성하는데 금융정보를 이용하는 것도 필요할 것으로 판단된다.

이 절에서는 가구주의 특성에 따른 무응답 가구의 특성을 분석하고 이 가구 및 가구주 특성에 따른 무응답층을 결정하여 무응답 가중값을 작성해 보고자 한다. 또한 금융변수를 활용한 무응답 가중값도 작성하여 어떠한 무응답 가중값이 무응답 편향을 효과적으로 줄이는지에 대해 분석해 보고자 한다.

실험설계와 분석대상은 아래 <표 2-1>과 같다. 1차 웨이브에서 응답한 1만 가구의 중 2차 웨이브 조사에서 제외된 56가구는 분석에서 제외하고 조사대상인 9,944가구를 분석 대상으로 하였다. 2차 웨이브에서 응답한 8,918가구는 협조성향을 가지는 가구로, 행방 불명되거나 불응하거나 부재인 가구 1,026가구는 비협조성향 가구로 판단하여, 협조성향의 8,918가구를 2010년에서 응답한 가구로, 비협조성향의 1,026가구를 무응답 가구로 정의하였다.

<표 2-1> 실험설계와 분석대상

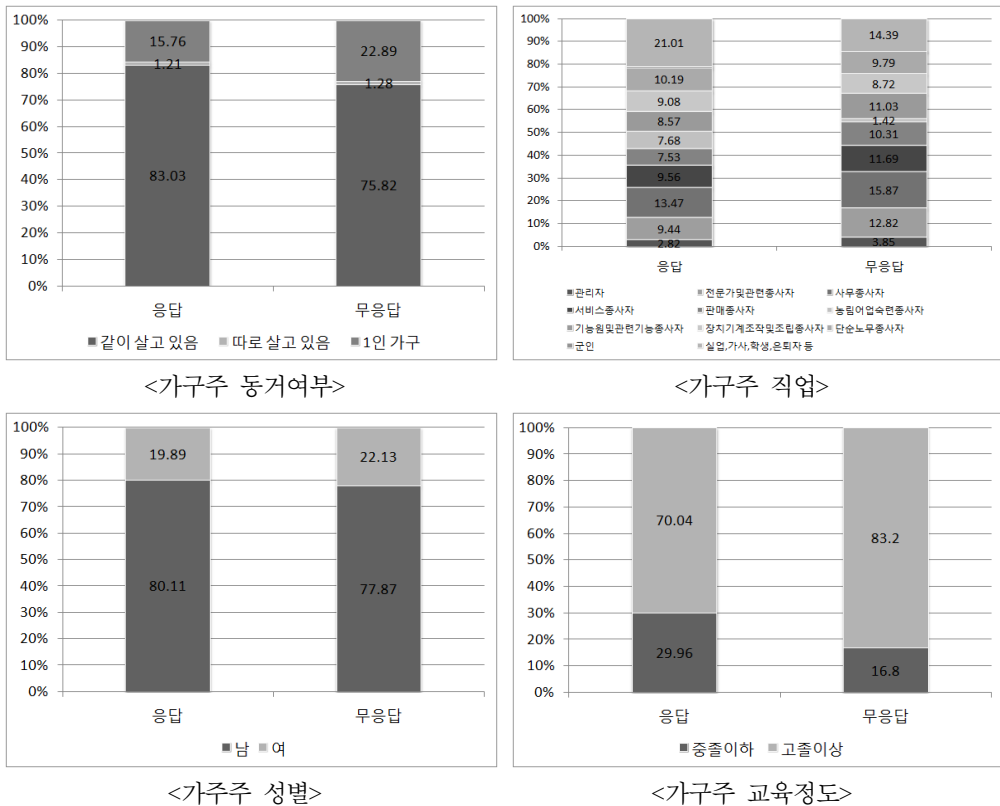
2차 웨이브 가구상태*		1차 웨이브 가구 분류	
조사대상		분석대상	9,944
유지(응답)	➡	응답가구(협조성향)	8,918
(행방불명, 불응, 부재)(무응답)	➡	무응답가구 (비협조성향)	1,026
(합가, 부적격, 기타)(조사제외)		분석제외	56
		합가 부적격, 기타	
		합계	10,000

* 1차 웨이브 조사대상 1만 가구 2차 웨이브 가구 상태임

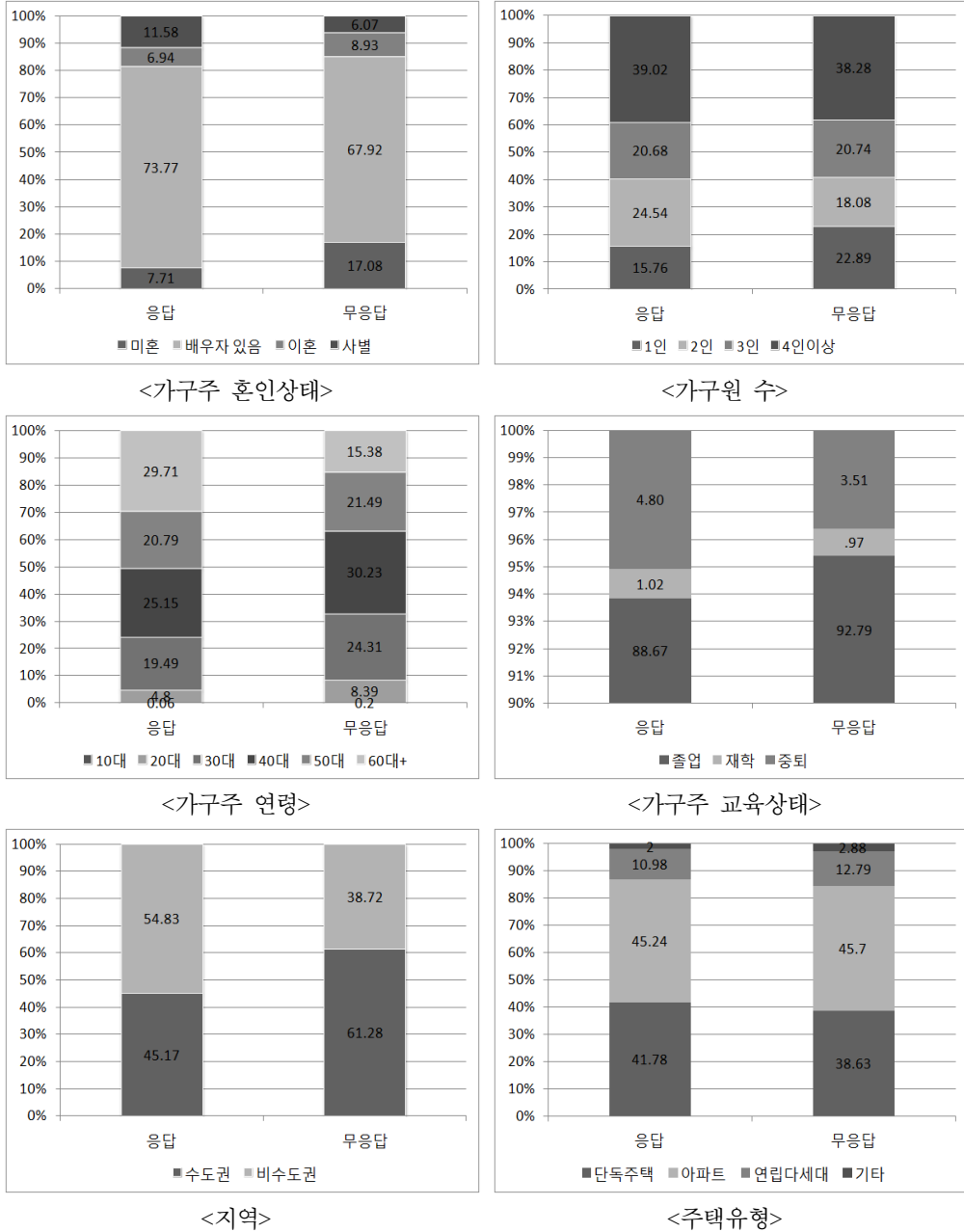
1. 무응답 가구 특성 분석

가. 응답가구와 무응답가구의 가구 특성

무응답 가구의 특성을 파악하기 위하여 무응답 가구와 응답가구의 가구주 정보 분포를 비교하였다(그림 2-1) 참조). 가계금융조사의 가구주 정보로 사용할 수 있는 특성 변수는 성별, 연령, 동거여부, 혼인상태, 가구원수, 교육정도, 가구주 직업, 가구주 종사상 지위, 가구주 종사 산업 등의 변수가 있다. 가구주 특성 외에 거주지역과 주택유형을 사용하였는데 지역 같은 경우 지역 간의 효과를 효율적으로 분석하기 위해 수도권(서울, 경기, 인천), 비수도권(기타 지역)으로 구분하였다. 응답가구와 무응답 가구를 비교해보았을 때 대부분의 가구주의 특성변수에서 차이가 발생하였다. 특히 지역(수도권, 비수도권), 연령, 교육정도, 혼인상태에서 비율의 차이가 큰 것으로 판단되며, 응답가구보다 무응답가구에서 여성, 수도권, 20대와 30대, 1인 가구, 미혼, 고졸이상의 비율이 높다는 것을 알 수 있다.



[그림 2-1] 응답여부에 따른 가구주 특성



[그림 2-1] 응답여부에 따른 가구주 특성 - 계속

(가구주 동거여부, 가구주 성별, 가구주 교육정도, 가구주 혼인상태, 가구주 산업, 가구원수, 가구주 연령, 가구주 직업, 시도, 가구주 교육상태, 수도권 비수도권, 주택유형)

나. 응답가구와 무응답가구의 금융 특성

가구의 금융수준과 응답여부와와의 관계를 알아보기 위하여 응답가구와 무응답가구의 금융정보의 평균과 분포를 분석하였다(<표 2-2>, <부록 3> 참조). 모든 금융항목에서 무응답가구의 평균값이 높으며, 금융부채를 제외한 자산 및 소득에서 유의미한 차이가 있는 것으로 나타났다. 따라서 자산이나 부채가 많은 가구의 무응답은 과소 추정을 초래하는 요인이 될 것이며, 응답여부에 영향을 미치면서 가구 금융변수와 관련이 있는 변수를 이용한 무응답 보정이 필요할 것으로 판단된다.

<표 2-2> 응답정보와 금융정보의 평균비교

(단위 : 만 원)

항목	금융항목	응답가구	무응답가구	t 값	유의확률
자산	자산총액	29981.26	41004.69	-5.41	0.000
	실물자산	23774.20	32503.28	-4.68	0.000
	금융자산	6207.06	8501.41	-5.67	0.000
	기타실물자산	10470.07	12477.46	-1.33	0.181
부채	부채총액	4962.82	6553.13	-2.59	0.009
	금융부채	3396.95	4142.20	-1.35	0.176
	임대보증금	1565.88	2410.94	-4.09	0.000
소득	경상소득	3876.56	4785.52	-7.06	0.000

다. 가구특성과 응답여부

앞에서 응답가구와 무응답가구의 가구 특성 분포 비교를 통해 두 가구 집단 간 가구 특성 구성비에 차이가 있음을 살펴보았다. 이제 무응답 조정에 필요한 응답여부에 영향을 주는 정보를 파악하기 위하여 가구 특성(가구주 특성, 지역, 주택유형)과 응답여부의 연관성 분석을 실시하였다.

가구 특성에 따른 응답과 무응답 비율에서 대부분의 가구 특성변수에서 차이가 있었으며, 카이제곱 검정을 통한 유의성 검정 결과 가구주 성별을 제외한 대부분의 가구주 특성변수에서 응답과 무응답의 비율이 통계적으로 유의하게 차이가 있는 것으로 분석되었다(<표 2-3>, <부록 1>, <부록 2> 참조). 유의수준 5%에서 성별을 제외한 가구의 특성에서 응답여부에 유의미한 차이가 있으며, 응답과 무응답에 영향을 미치는 것으로 판단되었다. 만약, 응답여부에 유의미한 가구특성의 금융 수준에서 차이가 있다면, 편향된 추정 결과를 초래할 수 있으므로 가구 특성변수를 이용한 조정이 필요할 것으로 판단된다.



〈표 2-3〉 가구특성과 응답여부에 대한 카이제곱 검정

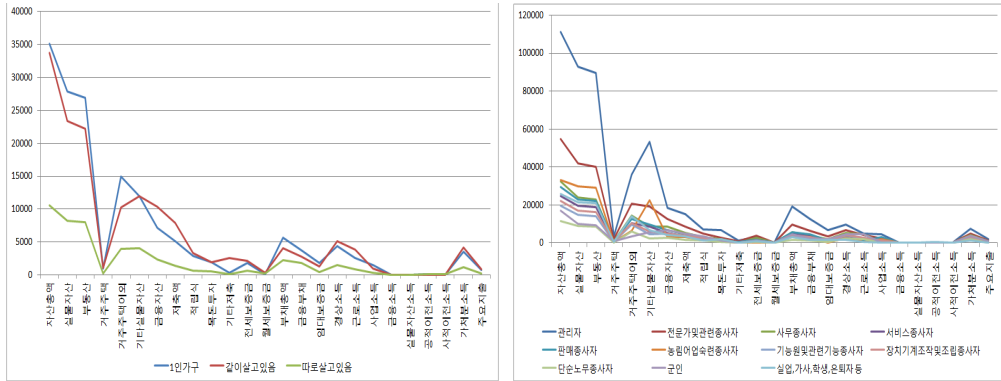
가구 특성	카이제곱 검정값	유의확률
가구주 동거여부**	32.45	0.000
가구주 직업**	115.86	0.000
가구주 성별	2.54	0.061
가구주 교육정도**	90.93	0.000
가구주 혼인상태**	17.06	0.000
가구주 산업**	151.15	0.000
가구원 수*	3.87	0.049
가구주 연령**	100.01	0.000
가구주 교육상태**	18.64	0.000
수도권,비수도권**	106.00	0.000
주택유형*	8.63	0.035

* : 유의수준 5%에서 통계적으로 유의함, ** : 유의수준 1%에서 통계적으로 유의함

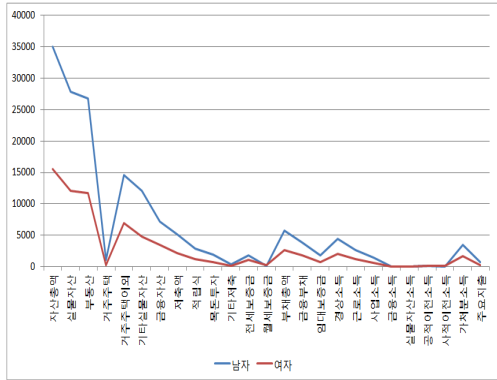
라. 가구 특성과 금융정보

가구특성과 가구 금융정보에 대해 응답가구와 무응답가구를 비교한 결과, 대부분에서 차이가 있음을 확인하였으며, 이 절에서는 가구 특성에 따른 가구 금융정보 수준에 차이가 있는지를 살펴보았다. 가구주 특성변수에 대해 금융정보별 평균값은 지역, 가구원수, 결혼여부, 학력 등에서 차이가 큰 것으로 나타났으며, [그림 2-2]에 수록하였다. 지역의 경우 수도권 지역이 대부분의 금융정보에서 큰 평균값을 가지며, 가구원수는 당연한 결과일수도 있겠지만, 1인 가구에서 대부분의 금융정보의 평균이 낮다. 학력의 경우도 고졸이상이 중졸이하보다는 금융정보의 평균이 높다. 혼인상태는 배우자가 있는 가구가 금융정보의 평균이 높고 사별한 경우의 평균이 두 번째로 높은 것을 볼 수 있다.

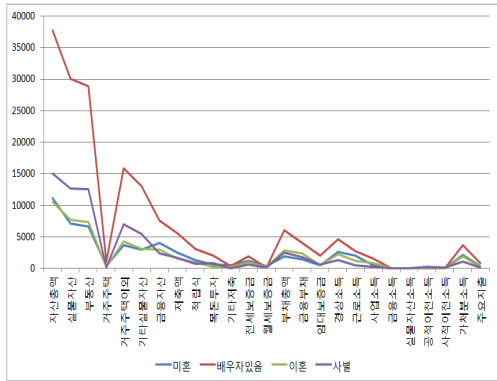
이러한 차이는 무응답 가구에 높은 금융수준을 가지는 가구들의 비율이 높을 경우, 예를 들어 무응답가구에서 수도권의 비율 또는 고졸이상 가구주 비율 높으면, 조사 결과는 무응답 편향을 가지게 되며, 과소 추정될 가능성이 높아진다. 따라서 무응답 조정에 주요 추정변수인 금융정보와 상관이 높은 가구 정보나 그 외 활용 가능한 보조정보를 최대한 이용한다면, 무응답으로 인한 편향 조정효과를 얻을 수 있을 것으로 판단된다.



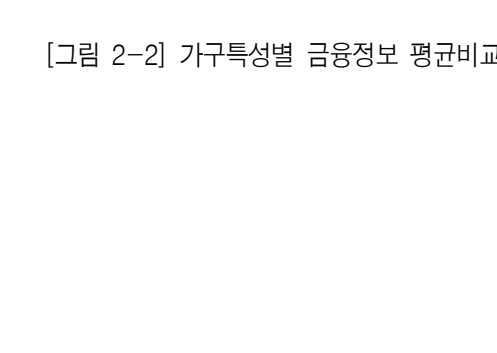
<가구주 동거여부>



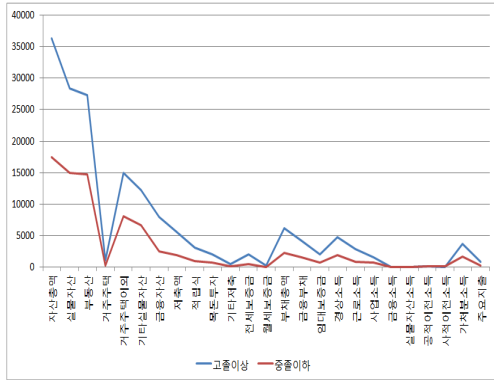
<가구주 성별>



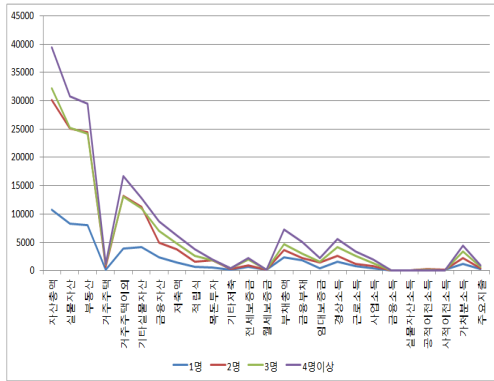
<가구주 혼인상태>



<가구주 직업>

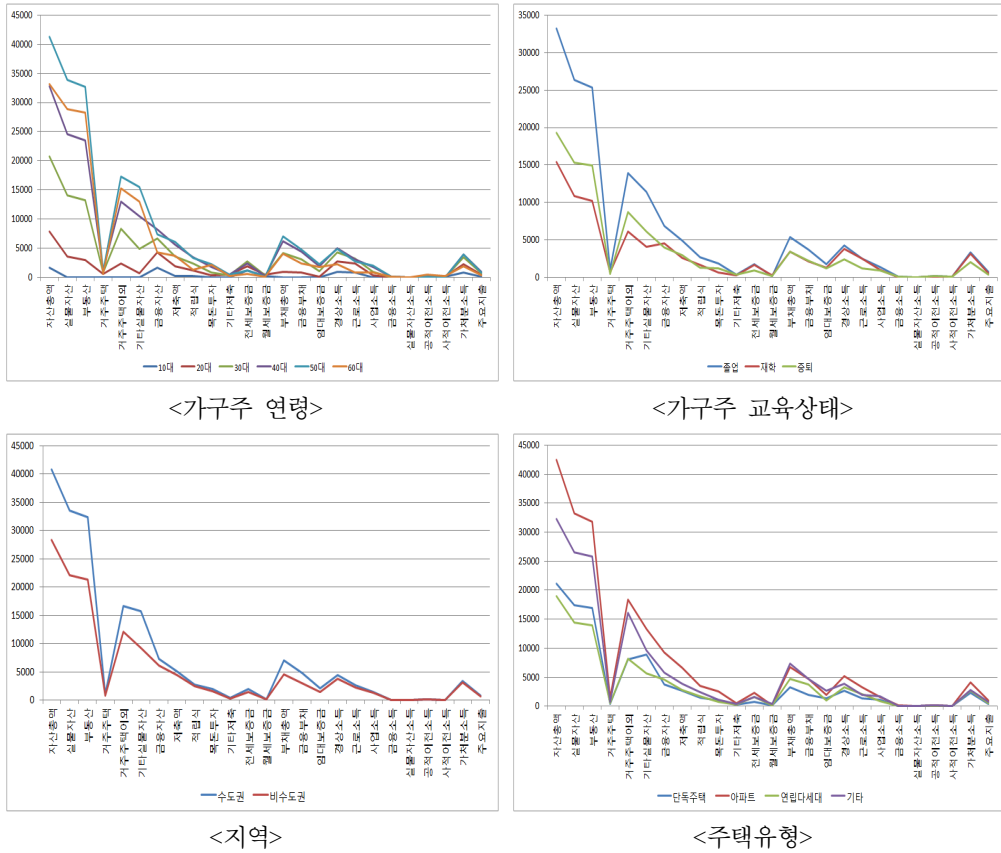


<가구주 교육정도>



<가구원 수>

[그림 2-2] 가구특성별 금융정보 평균비교



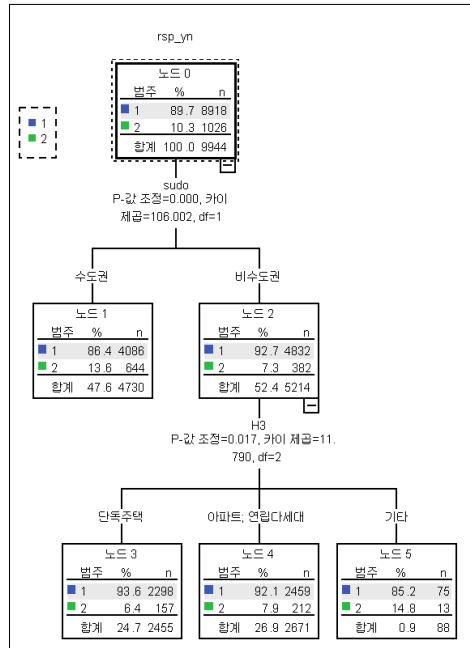
[그림 2-2] 가구특성별 금융정보 평균비교 - 계속

2. 무응답층의 결정과 조정률 결정

본 절에서는 앞에서의 응답관련 변수에 대한 기초 분석 결과를 바탕으로 무응답 조정층과 조정률을 산출하였다. 무응답 조정 변수에 따른 무응답 조정 효과를 파악하고자, 무응답 조정 변수는 이용 가능성을 기준으로 (표본 추출 정보), (가구주 정보), (가구 금융 정보)로 구분하였다. 표본추출정보는 시도와 주택유형이며, 편의상 가구주 정보와 추출정보를 가구 정보로 하였다.

가. CHAID(지역-수도권과 비수도권, 주택유형) 분석 (A)

가계금융조사의 경우 무응답 가중값을 작성할 수 있는 정보는 가구로부터 얻을 수 있는 정보로 한정되어 있기 때문에 기본적으로 시도와 주택유형만 사용할 수 있다. 시도 전체를 변수로 사용할 경우 유의하지 않은 케이스가 발생할 수 있으므로 우선적으로 수도권(서울, 경기, 인천)과 비수도권(기타 지역)으로 구분하여 새로운 변수로 사용하였다. 시도와 주택유형을 사용한 CHAID 분석 결과 수도권이 하나의 무응답층으로 구분되고 비수도권에서 단독주택이 하나의 층, 아파트·연립다세대가 하나의 층, 기타 주택유형의 층 4개의 층으로 구분되었다([그림 2-3], <표 2-4>, <표 2-5> 참조)



[그림 2-3] 수도권, 비수도권, 주택유형을 이용한 CHAID 분석 결과

<표 2-4> 수도권, 비수도권, 주택유형에 의한 무응답 층

층	지역	주택유형
1층	비수도권	기타
2층	비수도권	아파트;연립다세대
3층	비수도권	단독주택
4층	수도권	

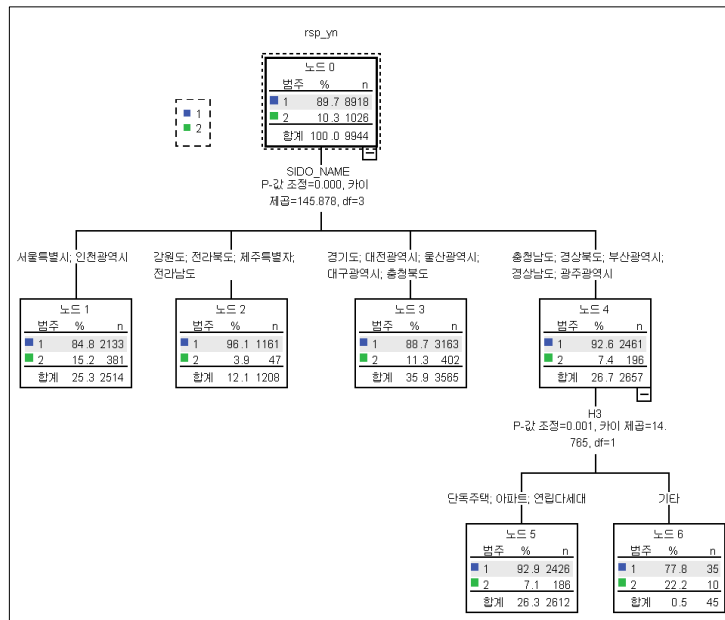


<표 2-5> 수도권, 비수도권, 주택유형에 의한 무응답 조정 비율

층	표본수				무응답조정비율		
	계	응답가구	무응답	비율	응답 승수합	무응답 승수합	조정률
1층	88	75	13	14.77	130825	22270	1.17
2층	2671	2459	212	7.94	4118537	345936	1.08
3층	2455	2298	157	6.40	4010081	276037	1.07
4층	4730	4086	644	13.62	6871944	1065909	1.16

나. CHAID(지역-16개시도, 주택유형) 분석 (B)

지역을 수도권, 비수도권으로 구분한 것과 별개로 16개 지역 모두를 무응답층으로 사용한 효과를 분석하기 위해 16개 시도를 모두 사용하여 주택유형과 CHAID 분석을 하였다. 수도권 중에 서울특별시와 인천광역시가 하나의 층으로 구성되었으며 강원도, 대전광역시, 울산광역시, 대구광역시, 충청북도가 하나의 층으로 구성되었다. 충청남도, 경상북도, 부산광역시, 경상남도, 광주광역시는 안에서 다시 단독주택,아파트;연립다세대와 기타 주택유형으로 구분되었다([그림 2-4], <표 2-6>, <표 2-7> 참조)



[그림 2-4] 전지역, 주택유형을 이용한 CHAID 분석 결과

〈표 2-6〉 전지역, 주택유형에 의한 무응답 층

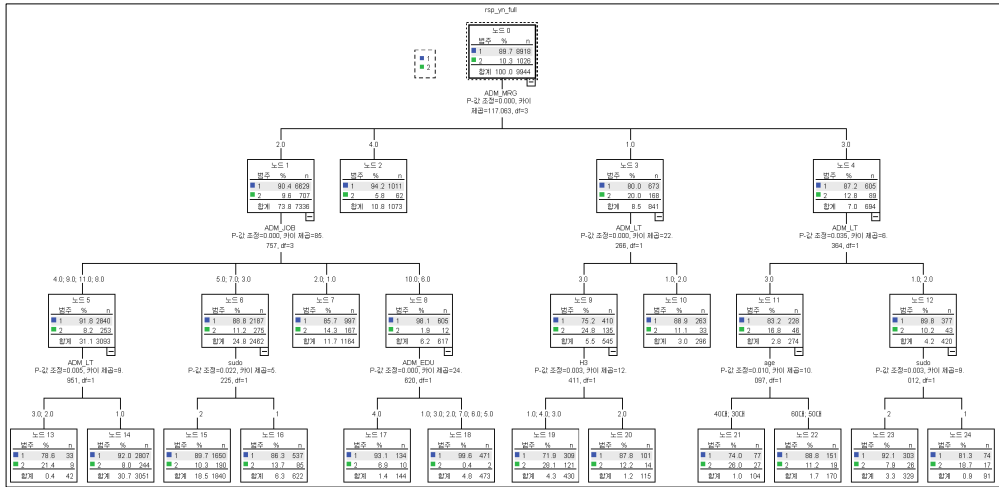
층	지역	주택유형
1층	충청남도;경상북도;부산광역시;경상남도;광주광역시	기타
2층	충청남도;경상북도;부산광역시;경상남도;광주광역시	단독주택;아파트;연립다세대
3층	경기도;대전광역시;울산광역시;대구광역시;충청북도	
4층	강원도;전라북도;제주특별자치;전라남도	
5층	서울특별시;인천광역시	

〈표 2-7〉 전지역, 주택유형에 의한 무응답 층

층	표본수				무응답조정비율		
	계	응답가구	무응답	비율	응답 승수합	무응답 승수합	조정률
1층	45	35	10	22.22	61201	17269	1.28
2층	2612	2426	186	7.12	4174774	310693	1.07
3층	3565	3163	402	11.28	5313203	660627	1.12
4층	1208	1161	47	3.89	1958577	81532	1.04
5층	2514	2133	381	15.16	3623630	640030	1.17

다. CHAID(가구정보) 분석 (C)

가계금융조사에서의 무응답 가구에서는 가구주 정보가 쉽게 사용가능한 정보일 것으로 생각된다. 하지만 가구주 정보 역시 무응답가구에 대해서는 사실상 조사가 어려우며, 인구주택총조사 자료나 행정자료 정보를 활용한다 하더라도 정확하다고 확신할 수 있는 정보는 아니다. 본 연구에서는 정보사용의 문제점을 떠나 가구주 정보를 무응답 층 설정에 활용하면 추정값에 어떠한 영향을 미치는지 알아보기 위해 가구 정보(주택유형과 지역 그리고 가구주 정보)를 활용한 CHAID분석을 하였다. 모든 가구 정보를 사용하였을 때 혼인상태, 동거여부, 직업, 주택유형, 나이, 지역(수도권, 비수도권)만이 실제 무응답 층 생성에 사용되었다([그림 2-5], <표 2-8> 참조).



* (입력변수) 가구주 나이, 교육정도, 직업, 동거여부, 결혼여부, 성별과 주택유형, 지역(수도권, 비수도권)

[그림 2-5] 가구 정보에 의한 CHAID 분석

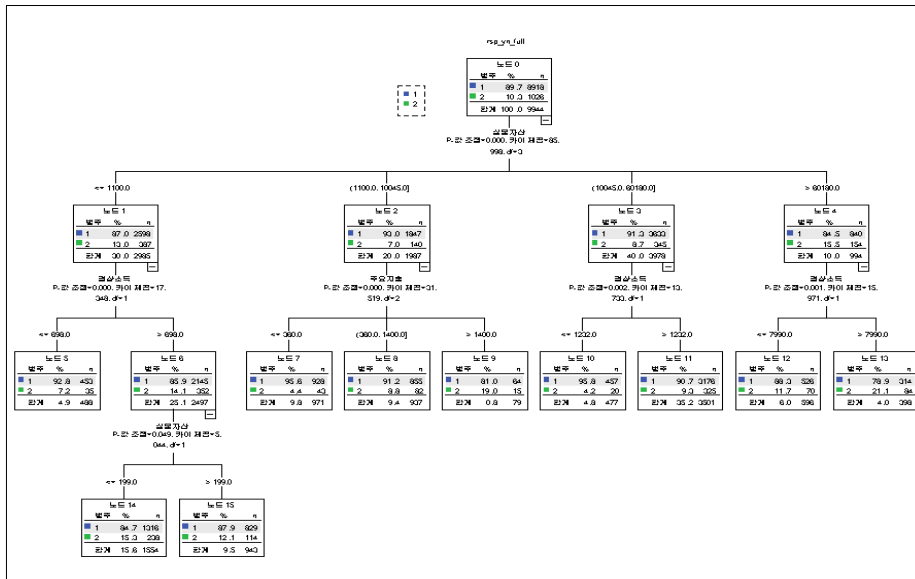
<표 2-8> 가구 정보(혼인상태, 동거여부, 직업, 주택유형, 나이, 수도권비수도권)에 의한 무응답 층

층	무응답조정비율			층	무응답조정비율		
	응답승수합	무응답승수합	조정률		응답승수합	무응답승수합	조정률
1층	1743064.40	104478.07	1.06	9층	820755.34	3517.94	1.00
2층	1566993.85	241938.13	1.15	10층	534630.81	209587.78	1.39
3층	452511.69	57265.56	1.13	11층	174516.21	24384.90	1.14
4층	57509.17	15554.80	1.27	12층	133600.11	46921.39	1.35
5층	4785331.71	418142.30	1.09	13층	260398.20	32787.52	1.13
6층	2822389.26	322872.10	1.11	14층	526992.99	45345.05	1.09
7층	890818.03	142287.86	1.16	15층	126415.79	27927.27	1.22
8층	235461.33	17142.78	1.07				

라. CHAID(금융정보) 분석 (D)

응답가구와 무응답 가구간의 금융정보의 평균에는 차이가 발생하였으며, 전반적으로 무응답 가구의 자산과 부채의 평균이 높은 것으로 분석되었다. 따라서 무응답층을 구성할 때 금융정보를 이용하여 무응답을 보정하는 것이 효율적일 수 있다. 2010년, 2011년 가계금융조사를 이용한 본 연구에서는 실험설계상 금융정보의 이용이 가능하지만 향후

금융정보를 이용하려면 조사 이외의 보조적인 정보의 추가 구성이 필요할 것으로 예상된다. CHAID분석을 위해 실물자산 등의 주요 금융변수를 입력하였지만 실제로 무응답층 구성에 사용된 변수는 실물자산, 경상소득, 주요지출 세 변수만 사용되었다. 이 세 변수가 실제 무응답에 영향을 많이 주는 변수일 것으로 예상된다([그림 2-6], <표 2-9> 참조)



* (입력변수) 실물자산, 금융자산, 부채총액, 경상소득, 가치분소득, 주요지출

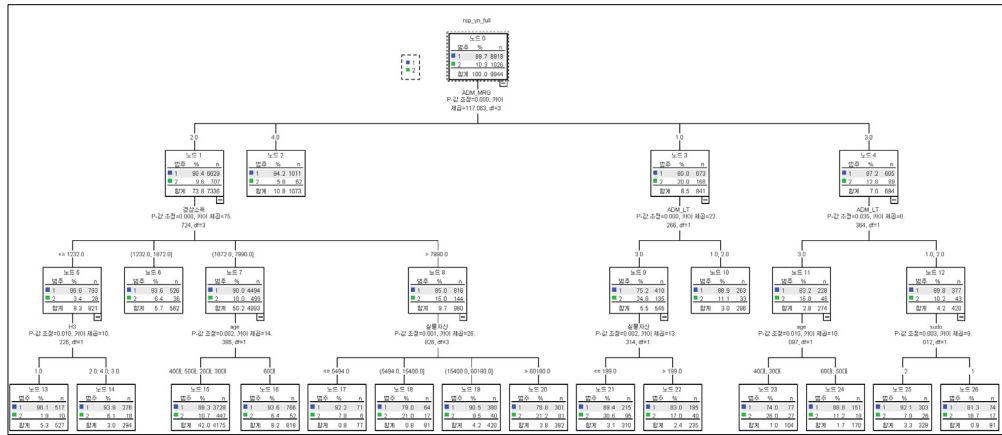
[그림 2-6] 주요 금융정보를 이용한 CHAID 분석 결과

<표 2-9> 실물자산, 경상소득, 주요지출에 의한 무응답 층

층	무응답조정비율		
	응답승수합	무응답 승수합	조정률
1층	784607.14	60832.52	1.0775
2층	1612346.74	74859.53	1.0464
3층	1489873.85	140835.56	1.0945
4층	107665.13	23905.33	1.2220
5층	791961.50	34695.06	1.0438
6층	5355586.06	545169.88	1.1018
7층	835915.64	111151.06	1.1330
8층	441143.35	108863.34	1.2468
9층	2278916.68	412154.85	1.1809

마. CHAID(금융변수, 가구정보) 분석 (E)

추가적인 정보가 요구되는 주요 금융변수의 사용은 쉽지 않으며 가구 정보에 보조적으로 사용하는 것이 효율적일 수 있을 것이다. 따라서 주요 금융변수와 가구 정보를 함께 이용하면 무응답층에 더 많은 정보를 제공할 수 있을 것으로 기대한다. 모든 주요 금융변수와 가구 정보를 이용하여 CHAID 분석을 수행한 결과 가구주 결혼여부, 가구주동거여부, 지역(수도권, 비수도권), 나이, 실물자산, 경상소득, 주택유형이 무응답층을 결정하는 변수로 사용되었다(그림 2-7, <표 2-10> 참조).



* (입력변수) 주요지출, 경상소득, 가처분소득, 부채총액, 금융자산, 실물자산, 가구주 나이, 교육정도, 산업, 직업, 동거여부, 결혼여부, 성별, 주택유형, 지역(수도권,비수도권)

[그림 2-7] 주요 금융변수와 가구 정보에 의한 CHAID 분석

<표 2-10> 주요 금융변수(경상소득, 실물자산)와 가구 정보(혼인상태, 동거여부, 주택유형, 나이, 수도권비수도권)에 의한 무응답 층

층	무응답조정비율			층	무응답조정비율		
	응답승수합	무응답승수합	조정률		응답승수합	무응답 승수합	조정률
1층	1847542.47	1743064.40	1.0599	10층	669336.48	603924.37	1.1083
2층	967308.52	906760.76	1.0668	11층	525481.94	420845.51	1.2486
3층	509777.25	452511.69	1.1266	12층	536749.49	372058.60	1.4426
4층	916652.18	899158.78	1.0195	13층	406370.23	337088.43	1.2055
5층	504203.21	473213.35	1.0655	14층	180521.50	133600.11	1.3512
6층	7134707.07	6376022.02	1.1190	15층	293185.72	260398.20	1.1259
7층	1359766.36	1273227.51	1.0680	16층	572338.04	526992.99	1.0860
8층	127755.20	117343.01	1.0887	17층	154343.06	126415.79	1.2209
9층	135503.64	108763.39	1.2459				

바. 로지스틱(지역-수도권과 비수도권, 주택유형) 분석 (F)

CHAID와 함께 가장 쉽게 사용되는 무응답층 생성 기법이 로지스틱 분석이다. 로지스틱회귀분석은 종속변수를 응답과 무응답으로 구분되는 지시변수를 사용하여 각각의 응답과 무응답에 대한 확률을 계산하여 준다. CHAID방법과 무응답 조정 효과가 어떻게 다른지 비교하는 대상이 된다. 주택유형과 지역(수도권, 비수도권)을 이용하여 로지스틱 분석을 수행하였다. 로지스틱 분석 결과 주택유형과 수도권, 비수도권 정보가 응답률과 통계적으로 유의한 관계가 있으므로 분석되었다 (<표 2-11> 참조).

<표 2-11> 로지스틱 모형 적합 결과

독립변수		추정값	표준오차	유의확률
주택유형*	H3			.067
단독주택**	H3(1)	0.186	0.070	.008
아파트	H3(2)	0.084	0.067	.211
연립 및 다세대	H3(3)	-0.012	0.089	.890
지역(수도권, 비수도권)**	SUDO			.015
수도권**	SUDO(1)	-0.932	0.038	.015
상수항***		2.009	0.061	<.0001

(***) 유의확률 <0.0001, (**) 유의확률 <0.05, (*) 유의확률 <0.1

3. 무응답 조정 적용 분석

본 절에서는 무응답 조정 효과를 측정하고자 다음과 같이 9가지 경우에 대해 추정결과에 대한 비교분석을 실시하였다.

- 2010년 1만 가구의 표본자료
- 2010년 1만 가구 표본자료와 설계 가중값
- 2010년 8,918 응답가구와 설계 가중값
- 2010년 8,918 응답가구와 6가지 무응답 조정 가중값



무응답 조정 가중값은 2.2절에서 결정된 6가지 무응답 조정방법 - A.CHAID(지역- 수도권과 비수도권, 주택유형), B.CHAID(지역-16개 시도, 주택유형), C. CHAID(가구주정보), D. CHAID(금융정보), E.CHAID(금융정보, 가구정보), F.로지스틱(지역 - 수도권과 비수도권, 주택유형)으로 산출하였다. 실험설계에 따라 2010년 1만 가구 표본자료에 대해 설계 가중값을 이용하여 추정된 결과를 준거(reference)로 하였다.

가. 무응답 조정 방법별 가중값 분포 비교

2.2절에서 적용한 6가지 무응답 조정방법의 가중값 분포에 대한 기초분석을 실시하였다(<표 2-12> 참조). 무응답 조정된 가중값이 설계 가중값에 비해 평균값과 표준편차가 커지는 것을 알 수 있다. 설계가중값의 평균이 1693.70으로 가장 작으며 로지스틱 방법의 평균이 1877.02로 가장 크다. 표준편차는 CHAID(금융정보와 가구정보)가 266.10으로 가장 크며 CHAD(수도권 비수도권, 주택유형)가 240.71로 가장 낮다. 변동계수(CV)의 값은 설계가중값이 12.56으로 가장 작으며 CHAID(금융정보+가구정보)가 14.27로 가장 커지는 것으로 나타났다.

<표 2-12> 가중값 분포 비교

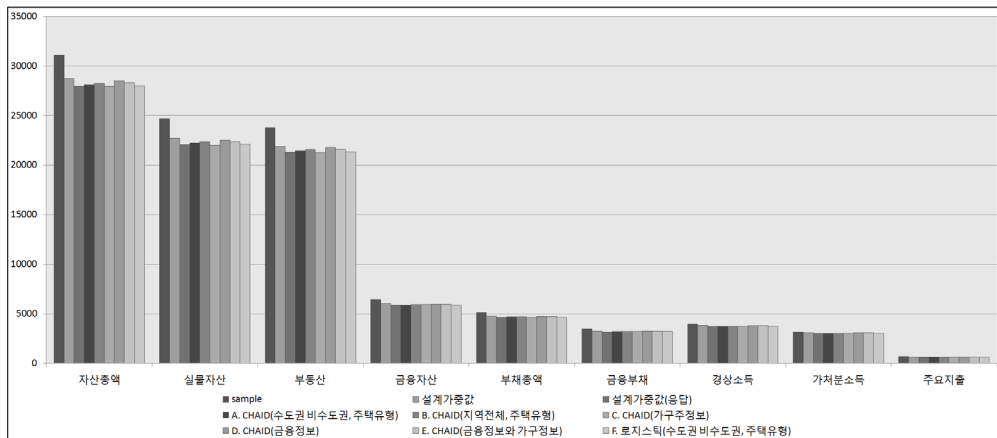
가중값	평균	표준편차	CV	최대값	최소값	중앙값
설계가중값*	1693.70	212.66	12.56	2167.45	258.10	1722.47
A. CHAID (지역-수도권과 비수도권, 주택유형)	1837.39	240.71	13.10	2514.24	258.10	1873.93
B. CHAID(지역-16개시도, 주택유형)	1857.83	250.60	13.49	2535.92	258.10	1890.84
C. CHAID(가구주정보)	1864.71	265.15	14.22	2818.77	258.10	1897.11
D. CHAID(금융정보)	1864.71	249.31	13.37	2648.62	258.10	1897.24
E. CHAID(금융정보와 가구정보)	1864.72	266.10	14.27	2818.77	258.10	1877.57
F. 로지스틱 (수도권 비수도권, 주택유형)	1877.02	245.04	13.05	2465.83	258.10	1925.29

* 2010년 8,918 응답가구의 설계가중값

나. 무응답 가중값을 적용한 금융변수 추정값 비교

무응답 가중값을 포함하여 9가지 방법에 대한 금융변수 추정결과를 [그림 2-8]과

<표 2-14>에 수록하였다. 설계가중값을 적용한 준거 추정결과와 비교했을 때, 무응답 조정을 실시한 모든 항목에서 과소 추정되는 것으로 나타났으며, 무응답 조정변수로 금융 정보 변수를 이용하였을 때, 과소추정 정도가 가장 작다는 것을 알 수 있다. 무응답 조정을 실시하지 않고 응답가구에 설계가중값만을 적용하여 추정한 결과는 과소추정 정도가 가장 높은 것으로 나타나, 무응답 조정 효과가 어느 정도 있다는 것을 알 수 있다.



[그림 2-8] 가중값에 의한 추정값 비교

다. 무응답 조정 효과 분석

무응답 가중값 적용으로 인한 조정 효과를 알아보기 위하여 측도로 무응답 편향(bias)을 사용하였다. 표본에서 응답자 평균에 대한 무응답 편향과 상대편향은 무응답률 및 응답가구와 무응답 가구의 평균값 차이에 영향을 받으며, 무응답 편향은 평균값 차이가 크더라도 다음과 같이 정의된다.

$$\text{무응답편향} = \text{무응답률} \times (\bar{y}_r - \bar{y}_m), \tag{1}$$

$$\text{무응답상대편향} = \text{무응답률} \times \left(\frac{\bar{y}_r - \bar{y}_m}{\bar{y}_m} \right), \tag{2}$$

여기서 \bar{y}_r 는 표본에서 응답자 평균, \bar{y}_m 는 표본에서 무응답자 평균이다.



본 연구에서는 무응답 조정으로 인한 편향축소 정도를 파악하고자 다음과 같이 무응답 조정된 응답 평균에 대한 편향을 정의하였다. 즉, 무응답 조정으로 인한 응답가구와 무응답가구의 주요변수 추정값의 편향이 표본 또는 무응답 조정을 적용하지 않은 경우에 비해 얼마나 축소되었지를 파악하고자 한다.

$$\text{편향1} = \left\{ \frac{(\bar{y}_{r(NR)} - \bar{y}_{m(D)})}{\bar{y}_{m(D)}} \right\} \times 100(\%)$$

$$\text{편향2} = \text{무응답률} \times (\bar{y}_{r(NR)} - \bar{y}_{m(D)})$$

$$\text{편향3} = \text{무응답률} \times \left\{ \frac{(\bar{y}_{r(NR)} - \bar{y}_{m(D)})}{\bar{y}_{m(D)}} \right\}$$

여기서 $\bar{y}_{r(NR)}$ 은 무응답조정을 실시한 응답가구의 평균, $\bar{y}_{m(D)}$ 는 설계가중값을 적용한 무응답 가구의 평균이다.

모든 경우에서 응답표본에 대한 무응답 편향보다는 편향정도가 줄어들음을 알 수 있다. 자산총액의 경우 Bias1의 CHAID(금융정보)를 이용한 추정값이 -20.5로 설계적용된 무응답 추정값에 가장 근사하며 CHAID(가구정보)의 Bias1값이 -22.1로 가장 차이가 큰 것으로 분석되었다. Bias2 역시 자산총액의 경우 CHAID(금융정보)의 값이 -759.8로 가장 작고, CHAID(가구정보)의 값이 -817.4로 가장 크게 나타났다. Bias3 역시 자산총액의 경우 CHAID(금융정보)의 값이 -2.1로 가장 작고, CHAID(가구정보)의 값이 -2.3으로 가장 크게 분석되었다.

무응답 조정된 응답 평균에 대한 편향 분석에서 무응답 조정으로 응답가구와 무응답가구의 차이가 축소됨을 알 수 있다. 적용한 방법 중 금융정보를 이용한 경우에 편향이 가장 크게 감소하였으며, 가구정보를 이용한 경우는 지역이나 주택유형을 이용한 경우보다 편향이 오히려 커지는 것으로 나타났다.

다음은 무응답 조정이 추정값에 미치는 영향력을 비교하기 위하여 무응답 조정 방법별 추정값과 준거(reference) 추정값을 비교하여 <표 2-14>에 수록하였다. 비교 측도는 무응답 조정된 응답가구의 추정값과 준거(reference, 2010년 가계금융조사(1만 가구)에 설계가중값 적용한 경우) 추정값의 차이와 비율을 사용하였으며, 아래와 같다.

$$\text{차이} = \bar{y}_{r(NR)} - \bar{y}_{Ref}$$

$$\text{비율} = \frac{\bar{y}_{r(NR)} - \bar{y}_{Ref}}{\bar{y}_{Ref}} \times 100$$

여기서 \bar{y}_{Ref} 2010년 가계금융조사(1만 가구)에 설계가중값 적용한 평균이다.

준거(reference) 추정결과와 비교하였을 때, 무응답 조정으로 어느 정도 추정값의 편향을 축소할 수는 있으나, 모든 방법에서 과소 추정되는 것으로 나타나 무응답 조정의 한계가 있음을 알 수 있다. 그러나 무응답 조정을 실시한 추정값이 준거 추정값에 더 근사하였으며, 자산총액에서 설계가중값을 적용한 경우 준거 추정값에 비해 -2.79(-803만원)로 과소 추정되며, 금융정보와 가구정보를 이용한 경우 -1.57(-452만원), 금융정보만을 이용한 경우 -0.9(-260만원)로 과소 추정되고 있다. 가구 정보만을 이용한 경우는 -2.85(-819만원)로 지역이나 주택유형만을 이용하는 것보다는 좋지 않았다. 자산과 부채는 금융정보만을 이용한 경우가 준거추정결과에 가장 근사한 추정값을 제공하고 있으며 경상소득에서는 금융정보와 가구정보를 같이 이용하는 경우에 가장 근사하게 추정되는 것으로 나타났다. 따라서 무응답 조정 시 금융정보와 관련이 있는 보조정보의 이용을 고려할 필요가 있을 것으로 판단된다.

그러나 기대했던 것과는 달리 가구정보(추출정보와 가구주정보)만을 이용한 경우에, 무응답 조정 효과가 좋지 않았다. 특히 자산에서 가구 정보만을 이용한 경우는 무응답 조정을 실시하지 않은 경우보다 오히려 더 과소 추정되는 결과를 가져왔다. 또한 적용 방법에 따라 차이는 있지만 모든 방법에서 과소 추정되는 결과를 보이고 있다. 이러한 결과에 대해서는 주요 금융변수인 자산, 부채에 대한 무응답 발생 메커니즘이 가구정보 변수에 대해 임의 결측(Missing At Random, MAR)을 따르지 않을 수 있다는 것을 생각할 수 있다. 즉, 앞에서 응답가구와 무응답가구의 금융변수 평균값을 보면, 이미 무응답한 가구의 모든 금융변수 평균값이 높고, 그 차이가 또한 큰 것으로 볼 때, 가구별 금융변수의 무응답 메커니즘이 비임의 결측(Not Missing At random, NMAR)을 따를 가능성이 크다는 점이다.

이 부분에 대해서는 추후 더 심도 있는 연구가 필요할 것이며, 본 연구결과로부터 부채, 자산, 소득과 같은 금융조사에서 무응답 조정변수로는 가구주 정보보다는 지역과 주택유형 변수가 더 유용하며, 가능하다면 금융정보와 관련이 깊은 보조변수를 사용하는 것이 바람직한 것으로 판단된다.

〈표 2-13〉 방법별 금융변수 추정값과 편향

항목	추정값										편향							
	sample	Reference	실제가중값 (응답)	A	B	C	D	E	F	Sample	실제가중값 (응답)	A	B	C	D	E	F	
자산	자산총액	31119.0	28769.0	27966.0	28135.7	28282.4	27950.3	28509.4	28316.7	28012.5	-26.9	-22.0	-21.6	-21.2	-22.1	-20.5	-21.1	-21.9
	실물자산	24675.0	22711.0	22087.0	22240.7	22346.7	22028.3	22535.4	22352.0	22124.7	-26.9	-21.8	-21.2	-20.8	-22.0	-20.2	-20.8	-21.6
	부동산	23772.0	21894.0	21320.0	21469.3	21576.4	21254.9	21755.1	21568.6	21355.2	-25.9	-21.0	-20.4	-20.0	-21.2	-19.3	-20.0	-20.8
	거주주택	903.0	818.0	768.0	771.4	770.3	773.4	780.3	783.4	769.5	-46.4	-39.0	-38.7	-38.8	-38.5	-38.0	-37.7	-38.8
	거주주택 이외	13095.0	12117.0	11630.0	11697.1	11820.6	11614.5	11729.8	11684.1	11652.4	-32.5	-29.2	-28.8	-28.0	-29.3	-28.6	-28.9	-29.1
	기타실물자산	10677.0	9776.0	9689.0	9772.2	9755.8	9640.4	10025.3	9884.5	9702.8	-16.1	-8.1	-7.4	-7.5	-8.6	-5.0	-6.3	-8.0
	금융자산	6444.0	6057.0	5878.0	5895.0	5935.7	5922.0	5974.0	5964.7	5887.8	-27.0	-23.1	-22.9	-22.3	-22.5	-21.8	-21.9	-22.9
	저축액	4617.0	4286.0	4205.0	4212.7	4221.4	4223.1	4268.5	4269.1	4210.1	-20.9	-15.9	-15.7	-15.5	-15.5	-14.6	-14.6	-15.8
	적립식	2527.0	2398.0	2350.0	2356.1	2361.2	2365.5	2384.3	2392.5	2353.5	-18.6	-16.6	-16.4	-16.2	-16.1	-15.4	-15.1	-16.5
	목돈투자	1718.0	1536.0	1520.0	1520.9	1524.6	1517.0	1544.4	1534.2	1521.0	-20.8	-9.7	-9.6	-9.4	-9.9	-8.2	-8.8	-9.6
기타저축	372.0	352.0	335.0	335.7	335.5	340.6	339.7	342.4	335.5	-33.6	-32.5	-32.3	-32.3	-31.3	-31.5	-31.0	-32.4	
부채	전세보증금	1638.0	1580.0	1492.0	1500.5	1530.8	1513.5	1519.2	1510.1	1496.1	-39.9	-36.9	-36.6	-35.3	-36.0	-35.8	-36.2	-36.8
	월세보증금	188.0	191.0	182.0	181.7	183.5	185.4	186.4	185.5	181.6	-33.7	-34.3	-34.4	-33.7	-33.1	-32.7	-33.0	-34.4
	부채총액	5126.9	4782.4	4652.0	4687.2	4722.1	4665.8	4761.9	4729.8	4665.2	-24.3	-21.6	-21.0	-20.4	-21.4	-19.8	-20.3	-21.4
	금융부채	3473.8	3249.1	3186.0	3209.9	3222.7	3190.1	3262.4	3240.8	3196.6	-18.0	-16.3	-15.7	-15.3	-16.2	-14.3	-14.9	-16.0
소득	임대보증금	1653.1	1533.3	1466.0	1477.2	1499.4	1475.7	1499.5	1489.0	1468.5	-35.1	-31.1	-30.6	-29.6	-30.7	-29.6	-30.1	-31.0
	경상소득	3970.3	3828.0	3754.4	3764.5	3771.1	3776.2	3803.3	3811.6	3759.4	-19.0	-16.2	-15.9	-15.8	-15.7	-15.1	-14.9	-16.1
	근로소득	2344.3	2289.2	2238.8	2246.4	2251.8	2266.1	2263.5	2275.4	2243.3	-18.9	-18.1	-17.8	-17.6	-17.1	-17.2	-16.8	-18.0
	사업소득	1319.4	1246.4	1216.7	1220.2	1221.6	1216.2	1241.6	1242.9	1217.6	-24.8	-19.4	-19.1	-19.1	-19.4	-17.7	-17.6	-19.3
	금융소득	66.5	58.4	56.5	56.5	56.6	56.6	57.4	57.2	56.6	-35.4	-24.7	-24.6	-24.6	-24.6	-23.5	-23.8	-24.6
	실물자산소득	9.8	8.6	9.2	9.2	9.0	9.2	9.4	9.3	9.2	182.5	136.0	135.9	132.6	136.0	141.0	139.0	135.5
지출	공적이전소득	153.3	150.0	156.1	155.3	154.9	152.3	155.1	151.7	155.8	53.5	63.3	62.5	62.0	59.3	62.3	58.7	63.0
	사적이전소득	77.0	75.4	77.1	76.9	77.2	75.8	76.3	75.2	77.0	27.6	27.0	26.7	27.3	24.9	25.7	23.9	26.9
	가처분소득	3179.3	3082.6	3037.9	3044.5	3048.3	3053.8	3073.7	3081.9	3041.3	-15.5	-12.7	-12.5	-12.4	-12.2	-11.6	-11.4	-12.6
	주요지출	686.3	652.0	628.8	630.8	632.6	634.2	637.9	639.2	629.9	-29.8	-26.6	-26.4	-26.2	-26.0	-25.6	-25.4	-26.5



항목	편향2						편향3										
	표본	체계기울임 (응답)	A	B	C	D	E	F	sample	체계기울임	A	B	C	D	E	F	
자산	자산총액	-1137.4	-815.8	-798.3	-783.2	-817.4	-759.8	-779.6	-811.0	-2.8	-2.3	-2.2	-2.2	-2.3	-2.1	-2.2	-2.3
	실물자산	-900.6	-634.0	-618.2	-607.2	-640.1	-587.8	-606.7	-630.1	-2.8	-2.2	-2.2	-2.2	-2.3	-2.1	-2.1	-2.2
	부동산	-826.5	-583.4	-568.0	-556.9	-590.1	-538.5	-557.7	-579.7	-2.7	-2.2	-2.1	-2.1	-2.2	-2.0	-2.1	-2.1
	거주주택	-74.2	-50.6	-50.2	-50.3	-50.0	-49.3	-49.0	-50.4	-4.8	-4.0	-4.0	-4.0	-4.0	-3.9	-3.9	-4.0
	거주주택이외	-619.4	-494.8	-487.9	-475.2	-496.4	-484.5	-489.3	-492.5	-3.4	-3.0	-3.0	-2.9	-3.0	-2.9	-3.0	-3.0
	기타실물자산	-207.1	-88.6	-80.0	-81.7	-93.7	-53.9	-68.5	-87.2	-1.7	-0.8	-0.8	-0.8	-0.9	-0.5	-0.6	-0.8
	금융자산	-236.7	-181.9	-180.2	-176.0	-177.3	-172.0	-173.0	-180.9	-2.8	-2.4	-2.4	-2.3	-2.3	-2.3	-2.3	-2.4
	저축액	-122.3	-81.8	-81.0	-80.1	-79.9	-75.3	-75.2	-81.3	-2.2	-1.6	-1.6	-1.6	-1.6	-1.5	-1.5	-1.6
	적립식	-58.4	-48.4	-47.8	-47.2	-46.8	-44.8	-44.0	-48.0	-1.9	-1.7	-1.7	-1.7	-1.7	-1.6	-1.6	-1.7
	복합투자	-45.4	-16.8	-16.7	-16.3	-17.1	-14.3	-15.4	-16.7	-2.1	-1.0	-1.0	-1.0	-1.0	-0.8	-0.9	-1.0
기타저축	-18.5	-16.6	-16.5	-16.6	-16.0	-16.1	-15.9	-16.6	-3.5	-3.3	-3.3	-3.3	-3.2	-3.3	-3.2	-3.3	
전세보증금	-105.1	-90.2	-89.3	-86.2	-87.9	-87.4	-88.3	-89.8	-4.1	-3.8	-3.8	-3.6	-3.7	-3.7	-3.7	-3.8	
월세보증금	-9.4	-9.8	-9.8	-9.6	-9.5	-9.4	-9.4	-9.8	-3.5	-3.5	-3.6	-3.5	-3.4	-3.4	-3.4	-3.6	
부채	부채총액	-164.1	-132.4	-128.8	-125.2	-131.0	-121.1	-124.4	-131.1	-2.5	-2.2	-2.2	-2.1	-2.2	-2.0	-2.1	-2.2
	금융부채	-76.9	-64.1	-61.6	-60.3	-63.6	-56.2	-58.4	-63.0	-1.9	-1.7	-1.6	-1.6	-1.7	-1.5	-1.5	-1.7
	임대보증금	-87.2	-68.4	-67.2	-64.9	-67.4	-64.9	-66.0	-68.1	-3.6	-3.2	-3.2	-3.1	-3.2	-3.0	-3.1	-3.2
	경상소득	-93.8	-74.7	-73.7	-73.0	-72.5	-69.7	-68.8	-74.2	-2.0	-1.7	-1.6	-1.6	-1.6	-1.6	-1.5	-1.7
	근로소득	-55.0	-51.1	-50.3	-49.8	-48.3	-48.6	-47.4	-50.7	-1.9	-1.9	-1.8	-1.8	-1.8	-1.8	-1.7	-1.9
소득	사업소득	-43.4	-30.2	-29.8	-29.7	-30.2	-27.6	-27.5	-30.1	-2.6	-2.0	-2.0	-2.0	-2.0	-1.8	-1.8	-2.0
	금융소득	-3.6	-1.9	-1.9	-1.9	-1.9	-1.8	-1.8	-1.9	-3.7	-2.5	-2.5	-2.5	-2.5	-2.4	-2.5	-2.5
	실물자산소득	0.7	0.5	0.5	0.5	0.5	0.6	0.6	0.5	18.8	14.0	14.0	13.7	14.0	14.5	14.3	14.0
	공적이전소득	5.7	6.2	6.2	6.1	5.9	6.1	5.8	6.2	5.5	6.5	6.4	6.4	6.1	6.4	6.1	6.5
	사적이전소득	1.8	1.7	1.7	1.7	1.6	1.6	1.5	1.7	2.8	2.8	2.8	2.8	2.6	2.7	2.5	2.8
지출	가치분소득	-58.9	-45.5	-44.8	-44.4	-43.8	-41.8	-40.9	-45.1	-1.6	-1.3	-1.3	-1.3	-1.3	-1.2	-1.2	-1.3
	주요지출	-28.9	-23.5	-23.3	-23.1	-23.0	-22.6	-22.5	-23.4	-3.1	-2.7	-2.7	-2.7	-2.7	-2.6	-2.6	-2.7

〈표 2-14〉 조정된 무응답 가중값의 영향력 비교

(단위 : 만원, %)

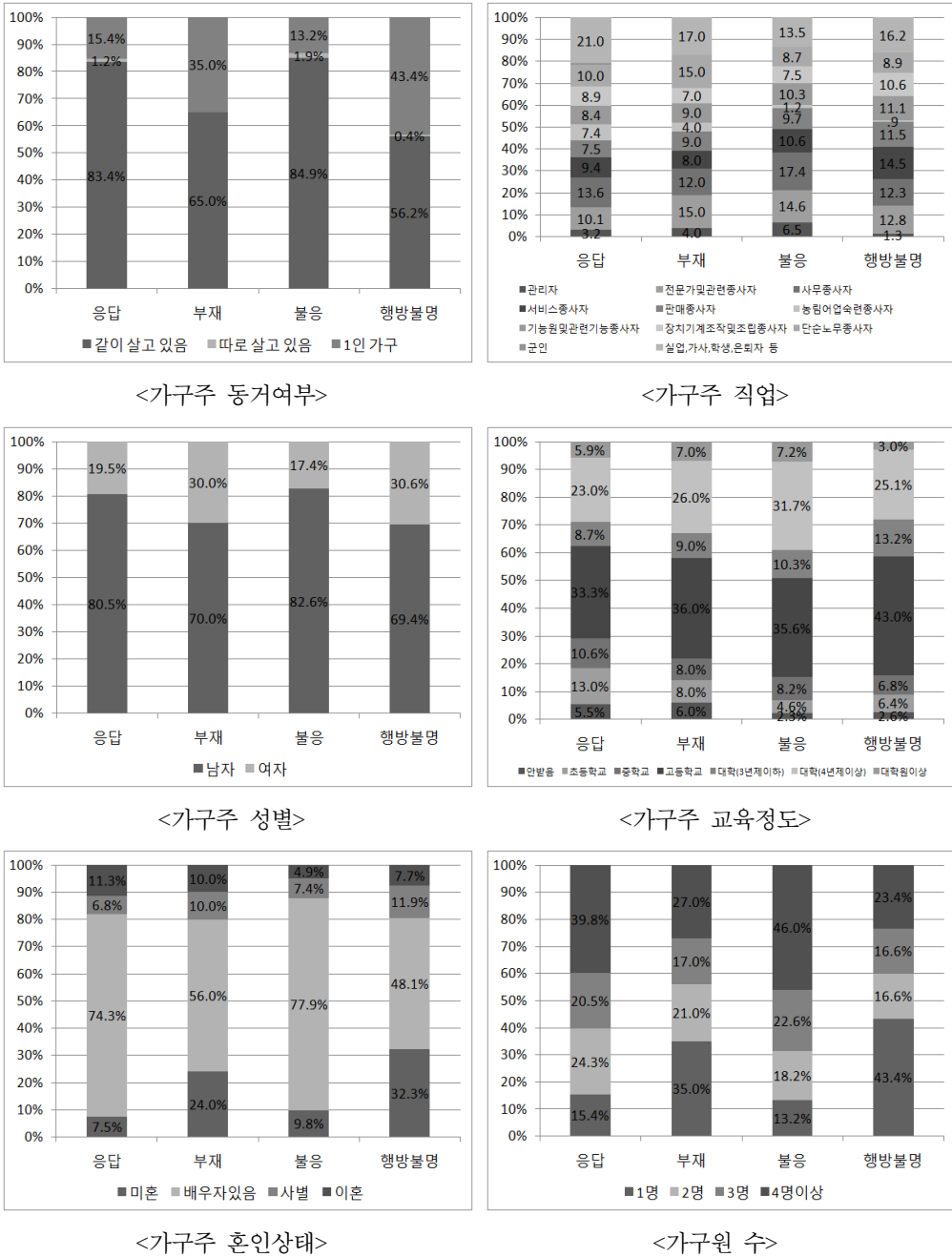
항 목	Reference	설계가중값 (응답)	A	B	C	D	E	F
자산총액(만원)	28769.00	27966.00	28135.69	28282.42	27950.33	28509.39	28316.71	28012.47
비율(%)		-2.79	-2.20	-1.69	-2.85	-0.90	-1.57	-2.63
차이(만원)		-803.00	-633.31	-486.58	-818.67	-259.61	-452.29	-756.53
부채총액(만원)	4782.36	4652.01	4687.16	4722.11	4665.81	4761.92	4729.76	4665.18
비율(%)		-2.73	-1.99	-1.26	-2.44	-0.43	-1.10	-2.45
차이(만원)		-130.35	-95.20	-60.25	-116.55	-20.44	-52.60	-117.18
경상소득(만원)	3827.96	3754.41	3764.55	3771.13	3776.23	3803.31	3811.62	3759.42
비율(%)		-1.92	-1.66	-1.48	-1.35	-0.64	-0.43	-1.79
차이(만원)		-73.55	-63.41	-56.83	-51.73	-24.65	-16.34	-68.54

4. 무응답 유형별 특성 분석

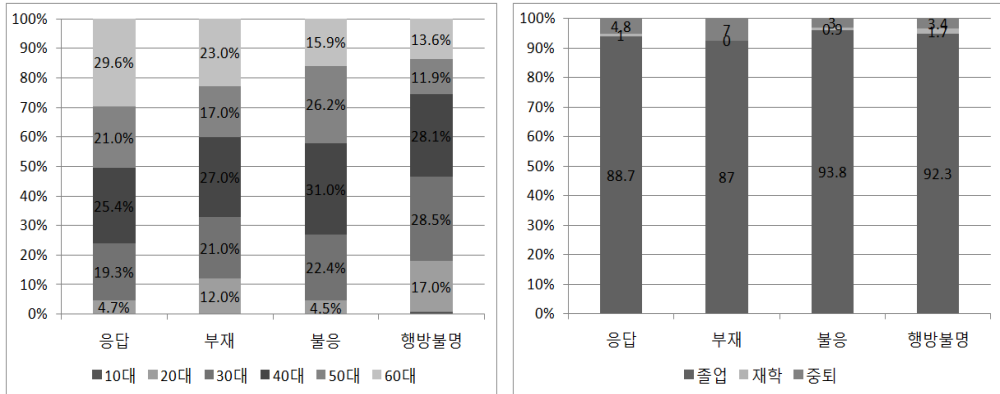
2.3절의 무응답 조정결과에서 가구 정보를 이용한 경우가 추출정보인 시도와 주택유형을 이용한 경우보다 효과적이지 않은 것으로 나타났다. 본 절에서는 이에 대해 생각해 볼 수 있는 또 다른 원인으로 무응답가구의 유형(부채, 불응, 행방불명)에 따른 특성을 살펴보았다. 무응답 유형별 가구특성과 금융정보 수준 간에 차이가 있다면, 이로 인해 무응답 조정에서 적용하는 완전 임의결측이나, 임의결측에 대한 가정은 더욱 적절하지 않을 수 있을 것이다.

무응답 가구의 유형은 부채, 불응, 행방불명으로 구분되며 응답 유형별 가구특성 분포와 금융변수 평균값을 각각 [그림 2-9]와 <표 2-15>에 나타내었다. 가구특성 분포를 살펴보면, 가구주 동거여부, 성별, 교육정도, 혼인상태, 연령과 주택유형에서 무응답 유형별 분포에 차이가 있음을 확인할 수 있으며, 행방불명 가구는 부채, 불응가구에 비해 1인 가구, 미혼, 여성, 20-30대, 단독주택 거주 비율이 높은 것으로 나타났다. 금융변수에서도 유형별 평균값이 차이가 있으며, 행방불명 가구에 비해 불응과 부채가구의 금융변수 평균값이 높은 것을 알 수 있다. 이것은 무응답 가구의 분포가 무응답 유형에 따라 다를 수 있다는 것을 의미하며, 무응답 조정에 이러한 정보를 포함시킬 필요가 있음을 시사한다고 볼 수 있다. 이에 대해서는 추후 연구과제로 남겨 두고자 한다.



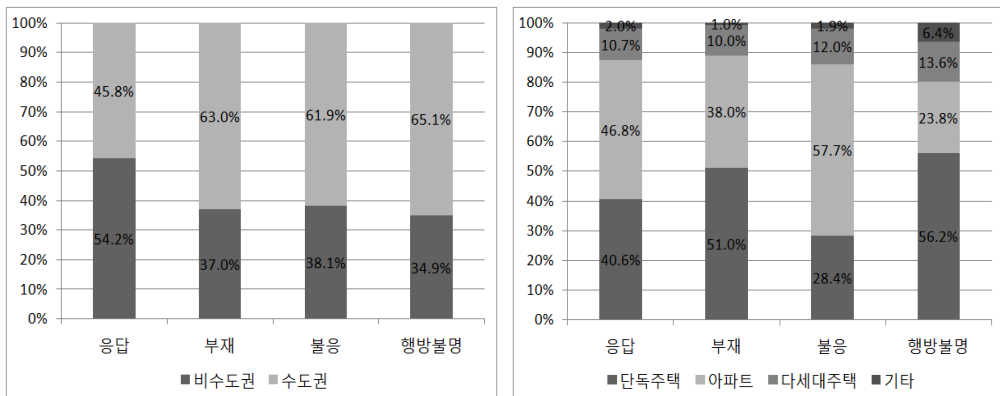


[그림 2-9] 응답유형(응답, 부재, 불응, 행방불명)에 따른 가구주 특성



<가구주 연령>

<가구주 교육상태>



<지역>

<주택유형>

[그림 2-9] 응답유형(응답, 부재, 불응, 행방불명)에 따른 가구주 특성 - 계속
(가구주 동거여부, 가구주 성별, 가구주 교육정도, 가구주 혼인상태, 가구주
산업, 가구원수, 가구주 연령, 가구주 직업, 시도, 가구주 교육상태, 수도권
비수도권, 주택유형)

〈표 2-15〉 응답유형별 금융정보 평균 비교

응답 유형	통계	자산 총액	가치분 소득	거주 주택	거주주택 이외	경상 소득	공적이전 소득	근로 소득	금융 부채	금융 소득	금융 자산	기타실물 자산	기타 저축	목돈 투자
응답	평균	29981.26	3120.47	828.77	12475.37	3876.56	159.03	2289.34	3396.95	62.90	6207.06	10470.07	353.39	1672.84
	표준편차	59942.19	3136.67	2196.35	21656.54	3734.92	486.26	2776.47	17101.06	340.61	11414.37	46124.44	3130.59	6459.89
행방 불명	평균	17177.58	2649.28	701.37	6731.49	3299.49	73.77	2070.28	2371.64	80.68	5714.08	4030.64	92.11	851.79
	표준편차	32379.72	2418.83	1734.82	18012.53	3026.04	333.37	2095.08	8180.90	573.30	10809.21	14449.88	525.07	3420.67
불응	평균	48729.32	4058.38	1852.81	21895.14	5348.99	108.30	3148.24	5183.43	96.88	9453.91	15527.45	736.42	2422.45
	표준편차	83037.30	4450.39	7250.45	42219.16	5550.94	434.25	3642.20	14764.97	621.58	20208.82	45305.07	9849.27	10081.90
부계	평균	43621.17	3600.88	1429.32	22470.00	4384.04	141.52	2338.72	1108.07	140.72	8469.85	11252.00	153.70	2938.50
	표준편차	83902.30	4491.07	4496.86	54511.10	5293.63	1003.26	3424.56	2803.12	585.07	15784.62	36606.86	766.83	10354.88
응답 유형	통계	부동산	부채총액	사업소득	사적이전 소득	실물자산	실물자산 소득	월세 보증금	임대 보증금	저축액	적립식	전세 보증금	주요지출	응답수
응답	평균	22945.43	4962.82	1276.03	78.73	23774.20	10.52	179.08	1565.88	4495.11	2468.88	1532.86	657.48	8918
	표준편차	54548.45	18595.98	3098.90	299.46	55235.78	174.82	761.17	5794.90	9919.36	4651.33	4612.12	838.92	6459.89
행방 불명	평균	10762.13	3429.17	989.83	79.11	11463.50	5.83	373.40	1057.53	2759.40	1815.50	2581.28	565.87	235
	표준편차	24915.91	11196.91	2556.29	310.89	25976.38	59.17	769.59	4085.19	5380.18	3321.65	7882.74	651.00	3420.67
불응	평균	37422.60	7944.21	1940.88	51.15	39275.41	3.55	214.98	2760.78	6705.65	3546.78	2533.28	1081.56	691
	표준편차	69558.07	20051.16	4852.78	274.49	73052.82	51.98	988.77	8721.53	18824.43	6618.51	6779.53	1643.87	10081.90
부계	평균	33722.00	4282.07	1669.20	93.88	35151.32	0.00	406.60	3174.00	5452.25	2360.05	2611.00	811.79	100
	표준편차	71276.76	18700.71	4071.25	411.01	73605.83	0.00	1035.55	18483.23	11837.38	3665.60	9140.89	1145.64	10354.88

제3절 결 론

본 연구에서는 2010년과 2011년 가계금융조사의 동일 가구의 연속된 조사자료를 이용하여 무응답 조정 연구를 진행하였다. 2010년 조사에 응답한 1만 가구 중 2011년 조사 제외 대상인 56가구는 분석대상에서 제외하고 나머지 조사대상인 9,944가구를 분석대상으로 하여, 2010년과 2011년 모두 조사에 응답한 협조성향을 보이는 8,918 가구를 응답가구로 정의하고 2010년 조사에는 응답하고 2011년에는 행방불명, 불응, 부재인 비협조성향의 1,026가구를 무응답가구로 정의하였다. 이러한 실험을 통해 무응답 가구의 특성과 무응답 조정 효과 등 계량적 분석을 실시하였으며, 무응답 조정에 대한 유의미한 정보와 시사점을 찾고자 하였다.

연구결과를 요약하면 다음과 같다. 첫째, 무응답가구와 응답가구의 특성 비교에서 무응답가구는 시도, 주택유형, 가구주 성별, 교육수준 등 가구특성과 자산, 소득, 부채관련 금융변수 모두에서 응답가구와 차이가 있다. 무응답가구는 응답가구에 비해 자산, 부채, 소득이 높으며, 수도권, 여성, 20대와 30대, 1인 가구, 미혼, 고졸이상의 비율이 높은 것으로 나타났다. 둘째, 성별을 제외한 대부분의 가구특성에서 응답분포에 유의미한 차이가 있으며, 가구 특성에 따른 가구 금융변수의 평균에 차이가 있는 것으로 나타났다. 셋째, 가구 특성변수와 금융변수를 이용하여 무응답 조정 변수로 (표본 추출 정보), (가구 정보), (금융 정보)를 구분하고, 로지스틱모형과 CHAID 의사결정나무 모형을 이용하였다. 최종적으로 6가지의 무응답 층 작성 방법, “지역변수(수도권, 비수도권)와 주택유형, CHAID”, “지역변수(16개 시도)와 주택유형, CHAID”, “가구정보, CHAID”, “가구정보와 금융정보, CHAID”, “금융정보, CHAID”, “지역변수(수도권, 비수도권)와 주택유형, Logistic”로 무응답 조정을 실시하였다. 동일한 조정변수- 지역(수도권, 비수도권)과 주택유형을 이용한 로지스틱 모형과 CHAID 모형 결과를 비교하였을 때, 로지스틱 모형에서 가중값의 평균과 표준편차가 약간 커지며, 추정결과 과소추정 정도가 큰 것으로 나타나 CHAID 모형이 더 적절한 것으로 판단하였다. 넷째, 무응답 조정에 따른 추정결과를 준거 추정값(2010년 모두 응답한 1만 가구 대상으로 설계가중값 적용)에 대해 조정 전 설계가중값만을 이용한 경우와 비교하여 편향 조정과 추정결과에 미치는 효과를 살펴보았다.

준거 추정결과와 비교하였을 때, 응답한 가구만을 이용하여 추정한 결과는 모든 방법에서 과소 추정되며, 응답한 가구에 대해 설계가중값을 이용하여 추정한 결과보다는 무응답 조정을 적용하였을 때, 과소 추정 정도가 축소되는 것으로 나타나 무응답 조정 효과를 확인할 수 있었다. 또한 적용한 방법에서 조정효과는 가구정보, 추출정보인 시도



와 주택유형, 가구정보와 금융정보, 금융정보 순서로 크게 나타나는 것으로 분석되었다. 특히 자산과 부채에서 가구정보를 이용한 무응답 조정 효과는 좋지 않은 것으로 나타났다. 다섯째, 무응답 조정에 영향을 미칠 수 있는 요인으로 무응답 가구 유형(부채, 불응, 행방불명)에 따른 가구변수와 금융변수 특성을 살펴보았다.

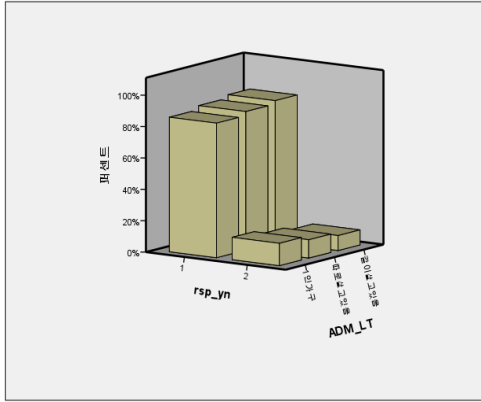
가구주 동거여부, 성별, 교육정도, 혼인상태, 연령, 주택유형에서 무응답 유형별 분포에 차이가 있음을 확인할 수 있으며, 행방불명 가구는 부채, 불응가구에 비해 1인 가구, 미혼, 여성, 20-30대, 단독주택 거주 비율이 높은 것으로 나타났다. 금융변수에서도 유형별 평균값이 차이가 있으며, 행방불명가구에 비해 불응과 부채가구의 금융변수 평균값이 높은 것으로 나타나 무응답 가구 내에 무응답 유형별 분포에 차이가 있음을 확인할 수 있었다.

연구결과로부터 응답가구와 무응답가구는 가구 특성과 금융수준에 있어서 차이가 있으며, 응답가구만으로 추정하여 발생하는 과소 추정에 대해 무응답 조정이 필요하다는 것과 부채, 자산, 소득과 같은 금융조사에서 무응답 조정변수로는 가구주 정보보다는 지역과 주택유형 변수가 더 유용하며, 가능하다면 금융정보와 관련이 깊은 보조변수를 사용하는 것이 바람직하다는 결론을 내릴 수 있었다. 또한 무응답 조정 결과로 비추어 볼 때, 금융변수에 대한 무응답 발생 메커니즘이 임의결측을 따르지 않으며, 비임의결측을 따를 가능성이 크다는 점과 무응답 유형별(불응, 부채, 행방불명) 가구변수와 금융변수의 분포가 다르다는 것을 확인할 수 있었다. 이 부분과 관련한 무응답 조정방법에 대해서는 앞으로 심도 깊은 연구가 진행되어야 할 것으로 사료되며, 추후 연구과제로 남겨 두고자 한다.

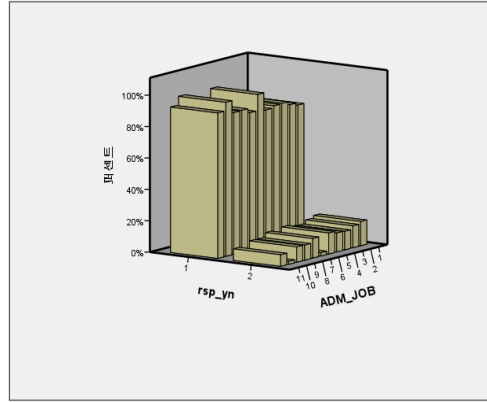
참고문헌

- 송주현, 안형진(2009), 무응답 자료처리 및 분석 정리, 통계교육원 교재, <통계교육원>
김재광(2004), 가계조사 무응답처리 기법 연구, 「통계연구」. 제9권 제1호, pp. 79-102, 통계청
박성인(2007), 가계자산조사의 무응답처리기법연구, 석사학위논문, 한남대학교

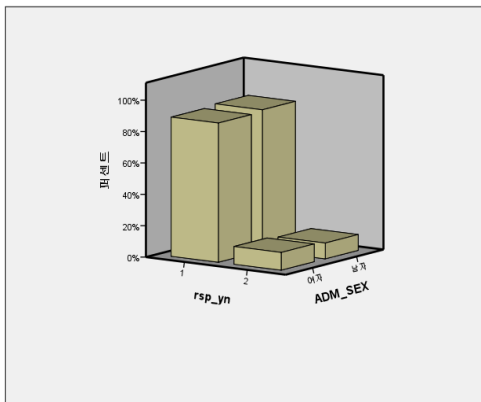




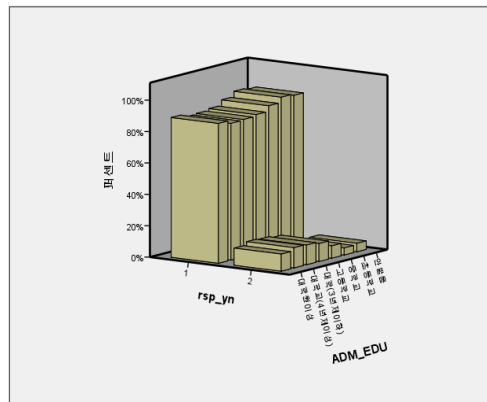
<**가구주 동거여부>



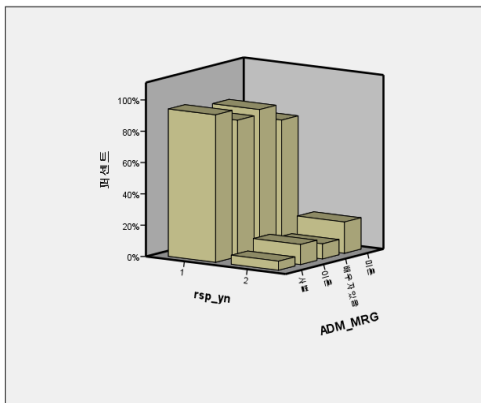
<**가구주 직업>



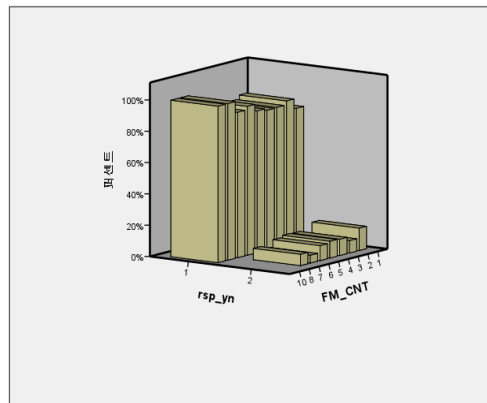
<가구주 성별>



<**가구주 교육정도>

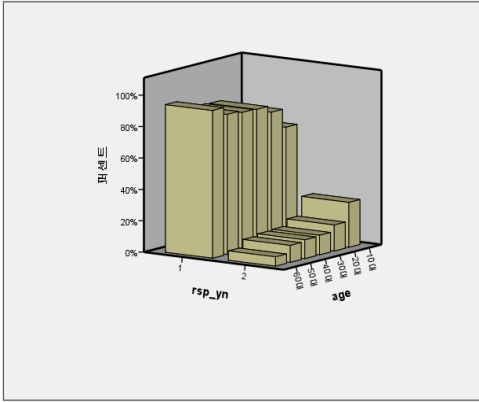


<**가구주 혼인상태>

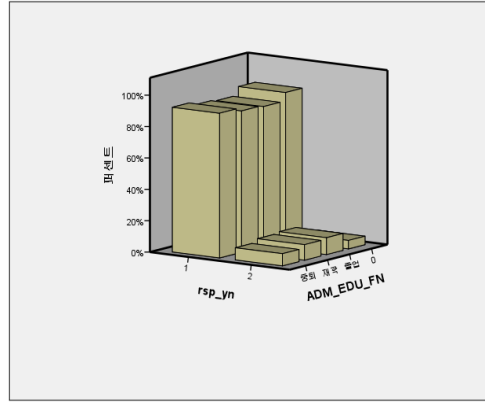


<**가구원 수>

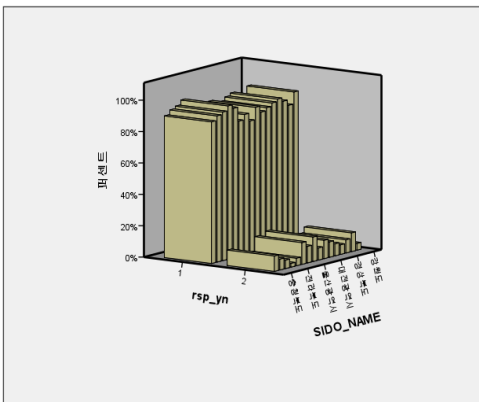
[부록 1] 가구주 특성에 따른 응답 분포



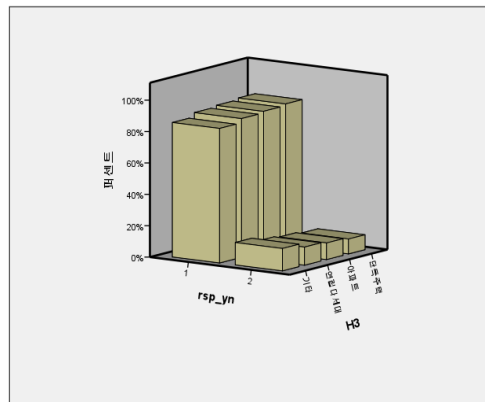
< **가구주 연령 >



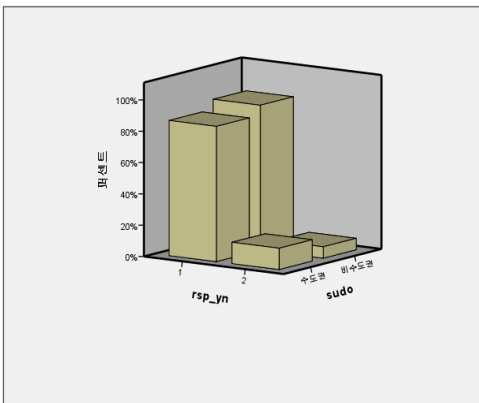
< **가구주 교육상태 >



< **시도 >



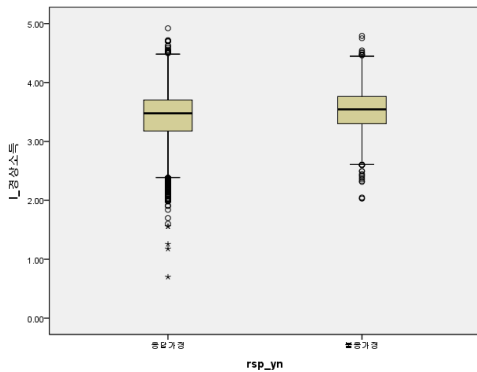
< **주택유형 >



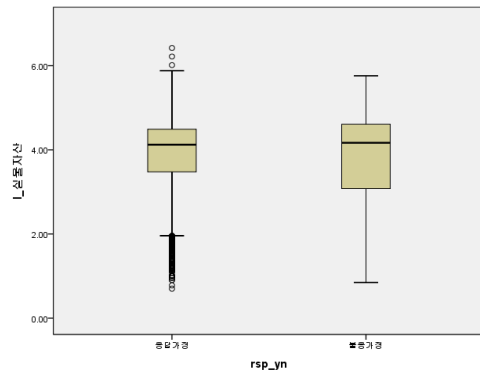
< **수도권 비수도권 >

[부록 1] 가구주 특성에 따른 응답 분포 - 계속

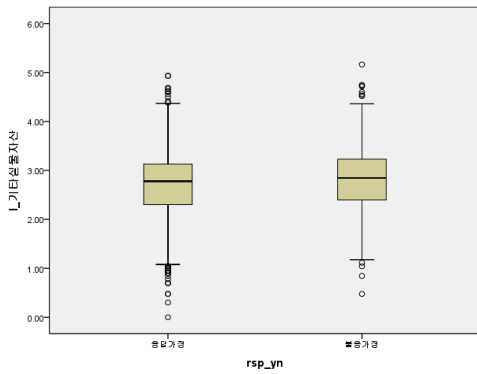
		가구주 연령						전체
		10대	20대	30대	40대	50대	60대	
응답 여부	응답	5	417	1717	2266	1876	2637	8918
	%	71.4%	83.4%	87.6%	88.1%	89.2%	94.1%	89.7%
	무응답	2	83	243	307	226	165	1026
	%	28.6%	16.6%	12.4%	11.9%	10.8%	5.9%	10.3%
전체	빈도	7	500	1960	2573	2102	2802	9944
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%



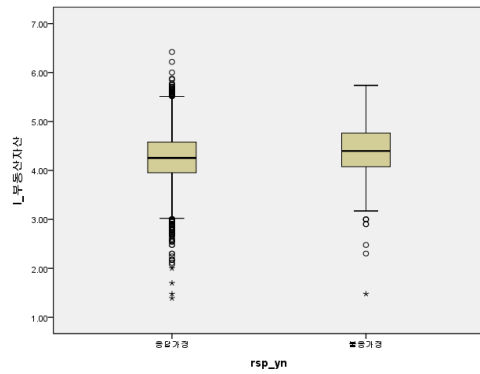
<경상소득>



<실물자산>

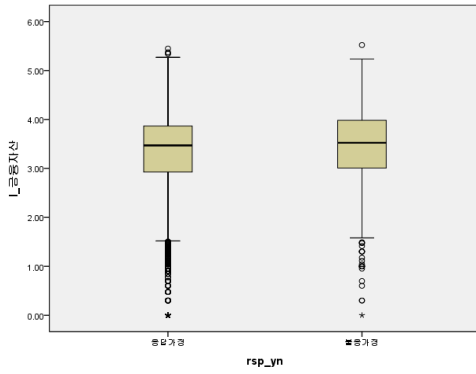


<기타실물자산>

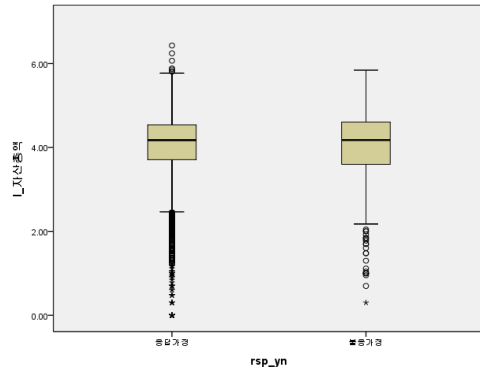


<부동산자산>

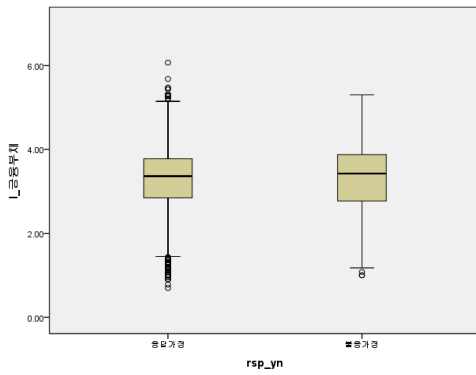
[부록 3] 가구의 응답여부에 따른 금융정보의 분포



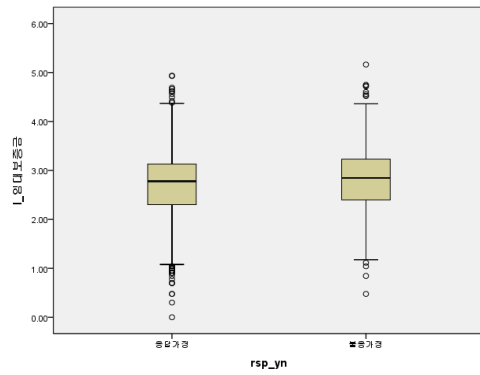
<금융자산>



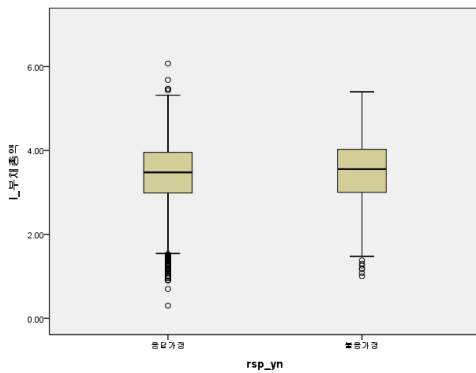
<자산총액>



<금융부채>



<임대보증금>



<부채총액>

[부록 3] 가구의 응답여부에 따른 금융정보의 분포 - 계속