한국복지패널에서 응답형태에 따른 패널가구의 특성 비교연구

손 창 균[1)]

요약

본 논문은 한국복지패널 표본가구의 응답패턴을 가구의 인구사회학적 특성에 따라 분석하였다. 조사원의 표본가구 방문회수별 조사 성공율과 패널가구로 구축된 표본가구의 원표본 유지율 등을 가구의 특징에 따라 분석함으로서 과학적인 조사체계를 구축할 수 있는 계기가 되고자 한다.

주요용어 : 응답패턴, 원표본 유지율, 패널조사.

1. 서론

다양한 조사방법과 노력에도 불구하고 전체 표본대상가구에 대해 완전한 응답을 얻어내는 것은 여전히 불가능한 문제이다. 응답을 꺼리는 응답자들에 대해 설득과 참여를 유도하기위해 조사비용과 더불어 오차의 발생이 증가하고 있으며, 연구자들은 높은 품질의 조사를 담보하기위해 상당한 노력을 경주하거나 무응답오차를 줄이기 위해 사후조사와 같은 기법을 수없이 적용해야 한다.

조사에서 필연적으로 발생하는 단위무응답에 대해 사후 조정을 통한 무응답 오차를 조정하려는 노력은 Kalton(1983), Lessler and Kalsbeek(1992), Little and Rubin(1987), Sarndal and Lundstorm(2005) 등 많은 학자들에 의해 수행되었으며, 국내에서도 여러 학자들에 의해 방법론적인 연구가 수행되었다.

이러한 대부분의 연구에서는 주로 무응답조정을 위해 제시된 추정량의 통계적 성질을 관찰하는데 초점이 맞춰져 있으며, 응답자의 성향에 따른 모형화에 대한 연구는 사실상 등한시된 것이 사실이다. 즉, 응답자의 응답패턴에 따라 응답성향 모형의 설정이나, 응답자가 조사에 참여하는 근본적인 원인이 무엇인지, 왜 사람들이 조사에 응답하는 지 등에 관한 응답자의 행태에 관한 연구는 여전히 진행되어오고 있는 연구 분야이다.

대체로 조사원이 표본가구에 접근하였을 때, 1차 접촉에서의 성공이 응답을 이끌어내기가 가장 수월한 것으로 알려져 있으며(최봉호, 2005), 이때 응답한 자료의 신뢰성이 재 접촉에 의한 자료의 신뢰성 보다 높은 것으로 알려져 있다. 또한 조사 참여율 역시 1차 접촉에서 성공한 가구가 재 접촉에 의한 가구보다 더 높게 나타남으로서 응답가구의 응답패턴이 향후 조사에서 표본가구의 조사에 대한 충성도를 측정하는 좋은 지표가 될 것으로 판단된다.

이러한 측면에서 본 연구에서는 패널조사에 참여하는 응답자들의 다양한 사회 경제학적 변인들에 의해 응답형태가 어떻게 다른지를 살펴봄으로서 무응답 가능성이 있는 패널을 보다 중점적으로 관리하여 궁극적으로 패널마모를 최소화 할 수 있는 전략을 수립하고, 조사과정에서 조사원들이 어떠한 노력에 의해 응답을 얻는지를 분석하고자 한다. 이를 위해 2006년 조사를 수행한 한국복지패널조사의 1차 조사 자료를 분석하고자 한다.

1) 한국 보건사회연구원 보건사회통계센터 부연구위원

2. 한국복지패널 자료의 분석 개요

2.1 조사개요

한국복지패널은 외환위기 이후 빈곤층(또는 working poor)및 차상위계층의 가구형태, 소득수준, 취업상태가 급격히 변화하고 있어, 이들의 규모와 상태변화를 동적으로 파악하여 정책지원을 위한 기초 자료를 생산하고, 소득계층별 경제활동 상태별, 연령별 등 각 인구집단의 생활실태와 복지욕구 등을 역동적으로 파악하고 정책의 효과를 평가함으로써 정책형성과 피드백에 기여하기위한 조사이다. 따라서 도시의 일반 가구를 대상으로 하는 KLIPS와는 지향하는 정책관점이 서로 다르다고 할 수 있으며, 조사대상가구는 일반가구와 저소득층 가구를 각각 50%씩 추출하여 이들에 관한 다양한 복지실태와 욕구 등에 관한 조사를 하고자 한다. 한국복지패널조사는 가구용 설문지와 가구원용 설문지(15세 이상), 아동용 설문지로 구성되어있으며, 가구원용 설문지는 15세 이상 중 고등학생을 제외한 경제활동인구 모두에 대해 응답을 받도록 하고 있다. 또한 아동용 설문지는 1차 패널조사에서만 조사되는 부가용 설문지로서 향후 3년마다 수행할 예정이다.

총 7,000가구(일반가구: 3,500가구, 저소득가구:3,500가구)를 대상으로 조사를 실시한 결과는 다음과 같이 나타났다. 먼저 총 조사된 가구수는 총 7,072가구로 나타나 전체 표본의 101%가 완료된 결과를 보였다. 가구의 총 조사 대상 가구원수는 14,469명으로 나타났다. 아동용 부가조사의 응답대상이었던 초등학교 4~6학년 아동의 수는 총 759명이었다. 각 지역별로 일반가구, 저소득가구의 조사현황을 살펴보면 〈표 2-1〉과 같다.

〈표 2-1〉 지역별 조사현황

(단위: 개, 가구, 명)

지역	표본 가구수		조사 완료 가구수		조사 완료 가구원수		조사완료 아동수	
	일반 가구	저소득 가구	일반 가구	저소득 가구	일반 가구	저소득 가구	일반 가구	저소득 가구
서 울	811	506	886	449	2024	823	92	35
부 산	254	272	277	251	651	439	33	20
대 구	187	227	204	218	478	395	25	12
인 천	228	193	247	190	559	356	39	19
광 주	114	130	123	121	305	214	19	13
대 전	118	89	132	76	298	137	16	5
울 산	120	82	124	79	290	134	22	4
경 기	644	471	706	426	1556	768	105	46
강 원	102	131	118	118	264	203	16	7
충 북	108	113	113	107	259	198	22	15
충 남	153	168	161	160	375	287	23	10
전 북	138	209	140	207	311	346	14	11
전 남	104	273	109	268	255	454	21	28
경 북	152	339	164	329	363	579	25	20
경 남	229	254	240	248	558	418	28	9
제 주	38	43	45	36	100	72	5	0
계	3,500	3,500	3,789	3,283	8,646	5,823	505	254

2.2 가구자료의 분석 개요

본 연구의 목적에 따라 응답 가구의 다양한 사회경제적 특성이 응답패턴에 어떠한 영향을 미치는 지를 분석하기위해 먼저 가구자료를 중심으로 가구주의 인구학적 특성과

가구규모별 분포를 살펴보면 다음 <표 2-2>와 같다. 전체적으로 가구원수가 4명 이하 가구가 90% 이상을 차지하였으며, 이중 가장 많은 비중을 차지하는 가구규모는 4인 가구로 전체 가구의 28.6%를 차지하였다. 그 다음은 2명(23.3%), 3명(21.3%), 1명(17.1%) 순이었다.

저소득가구와 일반가구의 가구원수별 분포는 큰 차이를 보이는 것으로 나타났다. 일반가구의 경우 4인 가구의 비중이 34.4%로 매우 높게 나타나고, 다음으로 3인 가구 23.5%, 2인 가구 21.3%의 순인데 비해, 1인 단독가구(11.9%)의 비중은 상대적으로 낮게 나타난다. 이에 비해, 저소득가구의 경우 2인 가구의 비중이 35.3%, 1인 가구 32.2%로 1,2인 가구의 비율이 가장 높게 나타나고, 다음으로 3인 가구 15.0%, 4인 가구 11.7%의 순이었다. 또한, 일반가구는 저소득가구에 비해 5명 이상 가구규모를 가진 가구비중도 상대적으로 높은 것으로 나타나 소득수준과 가구규모가 어느 정도 연관성이 있음을 알 수 있다.

〈표 2-2〉 가구규모

(단위: %)

구분 / 특성	전체	저소득	일반
1명	17.1	32.1	11.9
2명	23.3	35.3	19.1
3명	21.3	15.0	23.5
4명	28.6	11.7	34.4
5명	7.6	4.3	8.7
6명	1.8	1.4	2.0
7명 이상	0.4	0.2	0.4
계	100.0	100.0	100.0

다음으로 가구주의 인구사회학적 특성을 소득집단별로 구분하여 살펴본 결과는 다음 <표 2-3>과 같다. 우선 전체적으로 남성 가구주(81.8%)가 높은 비중을 차지하는 것으로 나타났다. 소득집단 별로는 저소득가구의 경우 남성가구주가 61.9%, 여성가구주가 38.1%, 일반가구의 경우 남성가구주가 88.6%, 여성가구주가 11.4%로 저소득가구의 여성가구주 비중이 일반가구에 비해 상대적으로 높게 나타나고 있다.

가구주의 연령별 분포를 살펴보면, 전체적으로 40대 가구주가 25.3%, 30대 가구주가 24.5%로 비슷한 비중을 차지하였으며, 65세 이상 노인 가구주도 19.3%로 높은 비중을 차지하는 것으로 나타났다. 소득집단별로는 일반가구의 경우 30대 가구주가 30.0%, 40대 가구주가 28.6%로 근로활동이 활발한 30, 40대 가장의 비율이 높은 비중을 차지하는 반면, 저소득가구의 경우 65세 이상 노인 가장이 46.7%로 매우 높은 비중을 차지하고 있다.

한편 가구주의 학력분포를 보면, 전체적으로는 고졸 이하(35.1%)가 가장 높은 비중을 차지하였으며, 다음으로 대졸(22.2%), 초졸(20.8), 중졸(11.6%)의 순으로 나타났다. 소득집단별로는 일반가구의 경우 고졸(39.0%), 대졸(27.4%), 중졸(10.6%), 초졸(10.4%)의 순으로 상대적으로 고학력자의 비중이 높게 나타났다. 이에 비해 저소득가구의 경우 초졸 이하가 51.4%로 매우 높은 비중을 차지하였으며, 다음으로 고졸(23.6%), 중졸(14.6%), 대졸(7.0%)의 순이었

다. 이와 같이, 저소득가구 가구주가 일반가구 가구주에 비해 상대적으로 연령이 높고 저학력임을 알 수 있다.

〈표 2-3〉 가구주의 성별, 연령, 학력별 분포

(단위: %)

특성 \ 구분		전체	저소득	일반
성별	남성	81.8	61.9	88.6
	여성	18.2	38.1	11.4
연령	20세 미만	0.1	0.2	0.1
	20~30세 미만	6.1	1.9	7.5
	30~40세 미만	24.5	9.8	30.0
	40~50세 미만	25.3	15.7	28.6
	50~60세 미만	16.5	13.6	17.4
	60~65세 미만	8.2	12.1	6.8
	65세 이상	19.3	46.7	10.0
학력	초등졸 이하	20.8	51.4	10.4
	중학교졸 이하	11.6	14.6	10.6
	고등학교졸 이하	35.1	23.6	39.0
	전문대졸 이하	6.3	2.4	7.6
	대학교졸 이하	22.2	7.0	27.4
	대학원졸 이상	4.0	1.0	5.1

가구주의 경제활동 참여상태는 취업자가 76.5%, 미취업자가 23.5%로 나타났다. 소득집단별로 보면, 가구주의 취업비율이 저소득층 가구에서는 48.7%, 일반가구에서는 86.4%로 저소득층 가구의 취업 가구주 비율이 낮은 것으로 나타났다. 취업 가구주 중 임금근로자 비율은 저소득 가구보다는 일반가구에서 높게 나타났다.

〈표 2-4〉 가구주의 경제활동 참여상태 및 종사상 지위

(단위: %)

특성 \ 구분		전체	저소득	일반
취업자	임금근로자	53.9	27.2	63.4
	비임금근로자	22.6	21.5	23.0
미취업자	실업자	3.3	5.7	2.5
	비경제활동인구	20.2	45.6	11.1
모름/무응답		0.0	0.0	0.0
계		100.0	100.0	100.0

임금 근로자 가구주의 고용형태를 살펴보면 다음 <표 2-5>와 같다. 전체적으로 정규직 근로자 비율이 59.9%로 가장 높게 나타났으며, 그 다음으로 한시적 근로자(26.3%), 비전형 근로자(13.2%)의 순서로 나타났다. 소득계층별로 보면 일반 가구의 임금근로자 가구주 중 65.6%가 정규직인 반면, 저소득 가구의 임금근로자 가구주 중 22.7%만이 정규직인 것으로 나타났다.

〈표 2-5〉 임금 근로자 가구주의 고용형태

(단위: %)

구분 / 특성	전체	저소득	일반
한시적	26.3	44.9	23.5
비전형	13.2	31.0	10.4
정규직	59.9	22.7	65.6
모름/무응답	0.6	1.4	0.5
계	100.0	100.0	100.0

2.3 표본가구의 응답형태에 따른 기술통계

패널가구의 응답패턴을 분석해보면 먼저 1차 접촉에서 3,238가구(45.8%)를 조사완료 하였으며, 나머지 3,834가구(55.2%)는 조사를 완료하지 못하였다. 특히 이들 가구 중에서 조사거부에 의해 조사를 완료하지 못한 가구는 101(2.63%)가구로 나타났다.

〈표 2-6〉 미완사유별 방문회수

(단위: 가구)

방문회수 / 미완사유	1차	2차	3차	4차	5차	계
완료가구수	3,238	1,756	1,059	547	472	7,072
늦은귀가	366	210	113	61	-	750
장기출타	26	25	15	7	-	73
부재중	3,097	1,648	752	333	-	5,830
일부미완	159	95	74	34	-	362
조사거부	101	60	42	23	-	226
기타사유	85	40	23	14	-	162

다음으로 가구주 성별에 따라 미완사유를 분석해본 결과 1차 방문에서 완료한 3,238가구 중 남성 가구주는 2,398가구이고, 여성가구주는 840 가구인 것으로 나타났다. 특히 1차 조사에서 조사를 거부한 가구101가구 중에서 남성 가구주 가구는 80가구 여성가구주 가구는 21가구로 나타났다. 1차 접촉에서 여성가구 1,711 가구 중에서 늦은 귀가, 장기출타, 부재인 가구는 47.7%로 나타났으며, 남성가구주의 경우는 약 50%인 것으로 분석되어 가구주의 성별의 차이가 미완사유에 영향을 미치지 않는 것으로 나타났다. 2차 접촉에서는 여성가구주의 50%가 늦은 귀가, 장기출타, 부재로 조사를 완료하지 못하였고 남성가구주는 48.8%가 이와 같은 이유로 조사를 완료하지 못한 것으로 나타났다. 3차 조사에서는 여성가구주 가구의 46.3%가 남성가구주 가구의 41%가 같은 이유로 조사를 완료하지 못한 것으로 나타났다.

〈표 2-7〉 가구주성별에 따른 가구의 미완사유

(단위: 가구)

방문회수	가구주성별	완료가구	늦은귀가	장기출타	부재중	일부미완	조사거부	기타	계
1차	남	2,398	276	16	2,381	145	80	65	5,361
	여	840	90	10	716	14	21	20	1,711
2차	남	1,345	165	18	1,263	85	54	33	2,963
	여	411	45	7	385	10	6	7	871
3차	남	828	93	9	565	69	35	19	1,618
	여	231	20	6	187	5	7	4	460
4차	남	424	50	4	250	31	20	11	790
	여	123	11	3	83	3	3	3	229
5차	남	366	-	-	-	-	-	-	366
	여	106	-	-	-	-	-	-	106
계		7,072	750	73	5,830	362	226	162	14,475

다음으로 <표2-8>에 제시된 바와 같이 소득계층별로 분석해보면, 1차 접촉에서 조사를 완료한 가구의 22%가 일반가구이고, 23.8%가 저소득가구로 분석되었다. 미완사유를 분석해보면, 먼저 부재중으로 인해 가구를 접촉하지 못한 가구는 22.5% 가 일반가구이고, 21.3%가 저소득 가구로 나타났다. 도한 2차 접촉에서는 총 3834가구 중에서 가구원 부재로 인해 조사를 완료하지 못한 가구는 각각 23%가 일반가구이고, 19.9%가 저소득 가구로 나타났다. 이러한 분석을 통해 살펴보면, 저소득가구가 일반가구에 비해 가구원 부재로 인한 조사미완의 가능성이 낮은 것으로 판단되며, 상대적으로 저소득 가구의 가구원 접촉이 용이한 것을 알 수 있다. 또한 일반가구와 저소득가구의 조사 거부율을 살펴보면 1차 접촉시 일반가구는 약 0.8% 이고, 저소득 가구는 0.65%로 나타났으며, 2차 접촉시에는 동일한 비율로 나타났으며, 3차 접촉시에는 1.34%가 일반가구이고, 0.67%가 저소득가구로 나타났다.

〈표 2-8〉 소득계층별 가구의 미완사유

(단위: 가구)

방문회수	소득계층	완료가구	늦은귀가	장기출타	부재중	일부미완	조사거부	기타	계
1차	일반	1,558	199	9	1,591	100	55	43	3,555
	저소득	1,680	167	17	1,506	59	46	42	3,517
2차	일반	884	119	11	882	48	30	23	1,997
	저소득	872	91	14	766	47	30	17	1,837
3차	일반	543	68	6	404	51	28	13	1,113
	저소득	516	45	9	348	23	14	10	965
4차	일반	302	38	2	189	19	11	9	570
	저소득	245	23	5	144	15	12	5	449
5차	일반	268	-	-	-	-	-	-	268
	저소득	204	-	-	-	-	-	-	204
계		7,072	750	73	5,830	362	226	162	14,475

〈표 2-9〉 방문회수에 따른 가구형태별 가구의 미완사유

(단위: 가구)

방문회수	가구형태	완료가구	늦은귀가	장기출타	부재중	일부미완	조사거부	기타	계
1차	1인가구	775	87	8	620	6	10	14	1,520
	2인가구	1,042	69	8	883	38	34	21	2,095
	3인가구	608	67	3	600	40	25	15	1,358
	4인가구	600	98	6	750	53	24	24	1,555
	5인가구	166	36	1	198	13	6	10	430
	6인가구	40	8	0	37	8	2	1	96
	7인이상	7	1	0	9	1	0	0	18
	소계	3,238	366	26	3,097	159	101	85	7,072
2차	1인가구	339	39	9	344	2	5	7	745
	2인가구	542	48	8	410	15	19	11	1,053
	3인가구	336	39	4	324	27	14	6	750
	4인가구	408	63	3	432	29	12	8	955
	5인가구	107	14	1	109	17	9	7	264
	6인가구	22	6	0	24	3	1	0	56
	7인이상	2	1	0	5	2	0	1	11
	소계	1,756	210	25	1,648	95	60	40	3,834
3차	1인가구	204	25	3	165	0	6	3	406
	2인가구	281	24	4	175	13	9	5	511
	3인가구	209	17	2	153	14	15	4	414
	4인가구	259	32	5	203	31	8	9	547
	5인가구	86	8	1	43	13	4	2	157
	6인가구	14	7	0	11	2	0	0	34
	7인이상	6	0	0	2	1	0	0	9
	소계	1,059	113	15	752	74	42	23	2,078
4차	1인가구	112	11	2	72	0	2	3	202
	2인가구	132	9	3	73	8	3	2	230
	3인가구	114	10	1	63	8	5	4	205
	4인가구	133	22	0	103	15	11	4	288
	5인가구	42	6	1	17	2	2	1	71
	6인가구	13	3	0	4	0	0	0	20
	7인이상	1	0	0	1	1	0	0	3
	소계	547	61	7	333	34	23	14	1,019
5차	1인가구	90	-	-	-	-	-	-	90
	2인가구	98	-	-	-	-	-	-	98
	3인가구	91	-	-	-	-	-	-	91
	4인가구	155	-	-	-	-	-	-	155
	5인가구	29	-	-	-	-	-	-	29
	6인가구	7	-	-	-	-	-	-	7
	7인이상	2	-	-	-	-	-	-	2
	소계	472	0	0	0	0	0	0	472
계		7,072	750	73	5,830	362	226	162	14,475

<표2-9>는 가구형태에 따라 응답패턴을 나타내주는 것으로서 1차 접촉시 전체 7,072 가구 중에서 조사 완료된 3,238 가구 중 2인가구가 1042 가구인 32.2%로 가장 높은 비율을 차지하고 있으며, 다음으로 1인가구가 775가구인 24% 로 나머지는 3인 가구, 4인 가구, 5인 가구 순으로 나타났다. 이러한 응답패턴은 2차 이상의 가구형태별 완료율을 비교해보면 동일한 패턴으로 나타나며, 2인가구의 구성 형태상 젊은층의 부부 또는 상대적으로 노년부부 층으로 구성된 경우가 많기 때문에 조사가 용이했던 것으로 판단된다.

〈표 2-10〉 방문회수에 따른 가구주 연령별 가구의 미완사유

(단위: 가구)

방문회수	가구주연령	완료가구	늦은귀가	장기출타	부재중	일부미완	조사거부	기타	계
1차	20세미만	1	0	0	2	0	0	0	3
	20~30세미만	111	20	0	128	4	4	1	268
	30~40세미만	517	70	3	619	32	16	18	1,275
	40~50세미만	585	98	3	697	35	30	19	1,467
	50~60세미만	473	65	5	514	40	13	19	1,129
	60~65세미만	315	34	3	253	14	12	4	635
	65세이상	1,236	79	12	884	34	26	24	2,295
2차	20세미만	2	0	0	0	0	0	0	2
	20~30세미만	61	13	0	76	4	1	2	157
	30~40세미만	325	46	3	347	17	11	9	758
	40~50세미만	369	59	4	404	20	17	9	882
	50~60세미만	296	39	5	270	31	10	5	656
	60~65세미만	148	19	3	136	6	6	2	320
	65세이상	555	34	10	415	17	15	13	1,059
3차	20세미만	0	0	0	0	0	0	0	0
	20~30세미만	44	7	0	42	1	2	0	96
	30~40세미만	224	29	1	154	12	8	5	433
	40~50세미만	228	30	4	212	21	10	8	513
	50~60세미만	176	20	4	127	19	9	5	360
	60~65세미만	96	7	2	53	5	7	2	172
	65세이상	291	20	4	164	16	6	3	504
4차	20세미만	0	0	0	0	0	0	0	0
	20~30세미만	32	2	0	16	2	0	0	52
	30~40세미만	91	25	1	83	4	3	2	209
	40~50세미만	140	17	0	106	12	7	3	285
	50~60세미만	100	10	2	54	5	7	6	184
	60~65세미만	44	1	1	22	2	5	1	76
	65세이상	140	6	3	52	9	1	2	213
5차	20세미만	0	-	-	-	-	-	-	0
	20~30세미만	20	-	-	-	-	-	-	20
	30~40세미만	118	-	-	-	-	-	-	118
	40~50세미만	145	-	-	-	-	-	-	145
	50~60세미만	84	-	-	-	-	-	-	84
	60~65세미만	32	-	-	-	-	-	-	32
	65세이상	73	-	-	-	-	-	-	73
계		7,072	750	73	5830	362	226	162	14,475

<표2-10>은 가구주의 연령에 따른 응답패턴을 나타내주는 것으로서 1차 접촉시 전체 7,072 가구 중에서 조사 완료된 3,238 가구 중 가구주 연령이 65세 이상인 가구가 1,236 가구인 38.2%로 가장 높은 비율을 차지하고 있으며, 다음으로 40~50세 미만 인 가구가 585가구인 18.1% 로 나머지는 30~40세미만 가구, 50~60세미만 인 가구, 20~30세 미만 인 가구 순으로 나타났다. 이러한 응답패턴은 2차 접촉에서도 완료율을 비교해보면 동일한 패턴으로 나타나며, 이러한 접촉회수로 볼때 65세 이상의 가구원인 가구의 접촉성공률이 다른 연령대의 가구보다 높음을 알 수 있다. 한편 조사거부율을 살펴보면, 40~50세 미만의 가구주 가구에서 상대적으로 높게 나타나고 있다. 또한 늦은 귀가로 인한 조사 미완가구를 연령대별로 살펴보면 경제활동이 활발한 연령대인 30대~60세 미만의 가구주 가구에서 높게 나타나고

있다.

다음으로 가구주의 학력에 따른 가구의 미완사유를 분석해보면 상대적으로 가구주의 학력이 낮은 계층의 조사 완료율이 학력이 높은 계층에 비해 높게 나타나고 있다. 한편 조사 거부율을 살펴보면 고졸 학력의 가구주 가구원의 거부율이 가장 높게 나타나며, 다음으로 대졸자, 중졸, 전문대졸 순으로 나타나고 있다.

〈표 2-11〉 방문회수에 따른 가구주 학력별 가구의 미완사유

(단위: 가구)

방문회수	가구주학력	완료가구	늦은귀가	장기출타	부재중	일부미완	조사거부	기타	계
1차	무학	455	38	7	337	8	7	11	863
	초등졸	742	53	7	604	30	15	12	1,463
	중졸	386	39	4	402	20	13	13	877
	고졸	910	142	5	997	57	37	32	2,180
	전문대졸	143	15	1	183	9	4	3	358
	대졸	47	5	0	55	0	2	0	109
	대학원졸	94	16	0	69	6	1	2	188
2차	무학	226	13	5	154	3	2	5	408
	초등졸	348	35	8	300	18	6	6	721
	중졸	215	21	2	217	21	10	5	491
	고졸	556	85	5	554	32	24	14	1,270
	전문대졸	81	13	1	110	5	3	2	215
	대졸	22	1	0	33	4	1	1	62
	대학원졸	49	10	0	34	0	0	1	94
3차	무학	114	7	3	52	2	3	1	182
	초등졸	210	15	2	128	11	4	3	373
	중졸	131	17	3	107	9	6	3	276
	고졸	353	41	6	263	26	18	7	714
	전문대졸	68	6	0	52	6	1	1	134
	대졸	16	1	0	19	2	1	1	40
	대학원졸	20	6	0	14	0	4	1	45
4차	무학	48	2	1	15	1	1	0	68
	초등졸	94	6	3	48	5	4	3	163
	중졸	69	11	1	51	7	4	2	145
	고졸	184	24	2	122	11	12	6	361
	전문대졸	36	5	0	24	1	0	0	66
	대졸	14	0	0	10	0	0	0	24
	대학원졸	13	0	0	11	1	0	0	25
5차	무학	20	-	-	-	-	-	-	20
	초등졸	69	-	-	-	-	-	-	69
	중졸	76	-	-	-	-	-	-	76
	고졸	177	-	-	-	-	-	-	177
	전문대졸	30	-	-	-	-	-	-	30
	대졸	10	-	-	-	-	-	-	10
	대학원졸	12	-	-	-	-	-	-	12
계		6,038	627	66	4965	295	183	135	12,309

3. 가구특성에 따른 응답 패턴에 대한 분석

2절에서는 표본가구별 인구 사회학적 특성에 따른 응답패턴을 단순한 기술통계학적인 분석을 수행하였다. 이러한 기본적인 분석을 통해 통계적으로 어떤 의미를 가지는 지를 보다 심도 있는 분석을 하고자 한다. 즉, 가구의 인구사회학적인 특성에 따라 응답패턴이 어떻게 변화하는지에 대한 가설을 설정하여 통계적인 유의미성을 검증하고, 이를 바탕으로 표본가구에 대한 보다 과학적인 관리가 이루어 질 수 있는 관리 체계를 마련하고 한다.

3.1 분석모형

본 절에서는 접촉 차수별(1차~4차)로 가구의 특성과 가구주의 특성에 따라 조사완료확률이 어떻게 변화하는 지를 살펴보고자 한다. 이를 위한 연구가설 들은 먼저, 개별 방문 차수별로 표본가구의 인구사회학적인 특성에 따라 조사 완료률에 어떠한 영향을 끼치는지를 분석하고, 둘째, 가구의 전체 방문회수와 가구의 인구사회학적인 특성에 따라 원표본 유지율에 어떠한 영향을 미치는지를 검증하고, 셋째, 원표본과 대체표본 간에 가구의 특성에 따라 성공률이 어떠한 차이가 있는지를 분석하고자 한다.

이를 위해 복지패널 1차 자료(연구용)에서 분석변수를 다음과 같이 정의하였다.

먼저 조사가구에 방문하는 회수를 나타내는 변수로서,

a41 : 1차 방문 결과 (1= 완료, 2=미완)

a43 : 2차 방문 결과 (1= 완료, 2=미완)

a45 : 3차 방문 결과 (1= 완료, 2=미완)

a47 : 4차 방문 결과 (1= 완료, 2=미완)

aa1 : 대체여부 (0=원표본, 1=대체표본)

로 정의하였다.

다음으로 설명변수들로서는 다음과 같다.

a12 : 가구의 경제적 특성(1=일반가구, 2=저소득가구)

a39 : 방문회수 (1~10, 10회 이상은 10으로 처리)

a60 : 가구원 수(1=1인 가구, 2=2인 가구 … 7=7인 이상 가구)

b08 : 가구주 성별 (1=남, 2=여)

age(group) : 가구주 나이

grade : 가구주 학력(1=무학, 2=초졸, 3=중졸, 4=고졸, 5=전문대졸, 6=대졸, 7=대학원졸)

b24 : 혼인상태(1= 유배우, 2=사별, 3=이혼, 4=별거, 5=미혼)

b20 : 장애종류

b22 : 장애등급

b23 : 만성질환

위와 같이 설정된 연구가설을 검증하기 위해 다음과 같이 로짓모형을 이용하기로 한다. p 차원의 설명변수와 $i=1,2,\cdots,n$ 개의 관측치로부터 조사 완료율에 대한 로짓모형은 다음과 같이 설정할 수 있다.

$$\log\left(\frac{p_{ki}}{1-p_{ki}}\right)= \alpha+\beta_{k1}X_{i1}+\beta_{k2}X_{i2}+\cdots+\beta_{kp}X_{ip}$$

즉, $k=1$ 에 대해 1차 방문시 조사가 완료된 가구는 $y_i=1$을 갖고, 그렇지 않은 가구에 대해서는 $y_i=0$의 값을 가지며, 이에 대한 확률을 p_i로 표현할 수 있다. 같은 방법으로 $k=2$ 에 대해 2차 방문시 조사가 완료된 가구는 $y_i=1$을 갖고, 그렇지 않은 가구에 대해서는 $y_i=0$의 값을 갖는다($k=1,2,3,4$). 한편 원표본 유지율에 대한 확률을 추정할 경우 첨자 k를 생략한 모형으로 고려한다.

3.2 분석결과

먼저 1차 방문에서 조사 완료율이 가구의 특성에 따라 어떠한 영향을 보이는 지를 분석해 보면 <표3-1>에서와 같이 1차 방문에서 조사 완료율은 가구원수와 가구주연령, 가구주혼인상태가 조사 완료율에 영향을 미치는 것으로 분석되었다.

〈표 3-1〉 1차 방문에서 가구의 인구학적 특성에 따른 로짓분석

변 수	자유도	추정계수	표준오차	Wald Chi-Square	odds 비
상 수	1	-0.4068	0.2735	2.2132	
소득계층	1	-0.00712	0.0573	0.0154	0.993
가구원수	1	-0.133***	0.0279	22.9631	0.875
가주주성별	1	0.0196	0.0719	0.0744	1.020
가구주연령	1	0.0109***	0.00250	19.2398	1.011
가구주학력	1	0.0226	0.0245	0.8519	1.023
가구주혼인상태	1	-0.0670*	0.0309	4.7069	0.935
-2 Log L	-	8222.809	***: P<0.001, **: P<0.01, *: P<0.05		
Wald 통계량	6	106.8557***			

즉, 가구원수가 많은 가구일수록 1차 접촉에서 조사를 완료할 확률이 낮으며, 가구주의 혼인사태에서는 유배우 이외의 상태일수록 완료율이 낮아지는 결과를 나타내고 있다. 반면에 가구주연령이 높을수록 조사를 완료할 가능성은 높아지는 것으로 분석되었다.

〈표 3-2〉 2차 방문에서 가구의 인구학적 특성에 따른 로짓분석

변 수	자유도	추정계수	표준오차	Wald Chi-Square	odds 비
상 수	1	0.1619	0.3671	0.1946	
소득계층	1	0.0351	0.0767	0.2092	1.036
가구원수	1	-0.1108***	0.0367	9.1350	0.895
가주주성별	1	-0.0248	0.0983	0.0636	0.976
가구주연령	1	0.00408	0.00334	1.4929	1.004
가구주학력	1	-0.0297	0.0340	0.7624	0.971
가구주혼인상태	1	-0.1067***	0.0409	6.8099	0.899
-2 Log L	-	4462.940	***: P<0.001, **: P<0.01, *: P<0.05		
Wald 통계량	6	35.4151***			

<표3-2>에서는 2차 방문시 가구의 인구학적 특성에 따른 로짓 분석 결과로서 1차 방문과 마찬가지로 가구원수가 많은 가구일수록 완료율이 낮아지며, 가구주의 혼인상태가 유배우일수록 완료율이 높아지는 것으로 분석되었다.

앞에서와 같은 방법으로 3차 방문과 4차 방문에서 조사 완료율에 영향을 미치는 가구의 인구학적 특성변수를 분석해보면, 3차의 경우 가구주의 학력(grade)을 제외한 나머지 변수들은 별다른 영향을 미치지 않았으며, 4차 방문의 경우 2차 방문의 분석결과와 유사하게 가구원수(-0.1835**)와 가구주의 혼인상태(-0.02278**)에 영향을 받는 것으로 분석되었다.

분석 결과를 종합해 볼 때, 표본가구를 방문하여 가구원수가 많은 가구일수록, 가구주의 혼인상태가 불안정할수록 다른 가구형태에 비해 방문회수가 증가되며, 따라서 접촉할 경우 보다 세심한 주의를 기울일 필요가 있으며, 반대로 가구주의 연령층이 낮을수록 집에서 활동하는 시간이 적은 관계로 역시 방문회수가 증가할 가능성이 크기 때문에 연령층에 따라 적절한 접근 방식을 고안할 필요가 있다.

다음으로 표본설계 당시에 원 표본으로 선정된 단위와 조사과정에서 응답거절 등의 사유로 대체된 단위간의 원표본 유지율이 가구의 특성에 어떠한 영향을 미치는 지를 분석하였다. <표3-3>으로부터 저소득계층일수록 원표본 가구의 유지율이 높아지며, 가구원수가 많을수록 원표본 유지율이 높아지는 것으로 분석되었다. 한편, 가구주의 성별이 남자일수록, 가구주학력이 높을수록 원표본의 유지율을 떨어지는 것으로 분석되었고, 가구주의 장애 유무와 등급, 만성질환은 원표본 유지율에는 별 영향을 끼치지 않는 것으로 나타났다.

〈표 3-3〉 가구의 인구학적 특성에 따른 완료 및 대체가구의 영향에 대한 로짓분석

변 수	자유도	추정계수	표준오차	Wald Chi-Square	odds 비
상 수	1	0.3061	0.2875	1.1338	
소득계층	1	0.6626***	0.0643	106.3041	1.940
가구원수	1	0.1415***	0.0315	20.1522	1.152
가구주성별	1	-0.1547*	0.0814	3.6122	0.857
가구주 연령 층	1	-0.0214	0.0276	0.6028	0.979
가구주학력	1	-0.1473***	0.0265	30.8193	0.863
가구주혼인상태	1	0.00459	0.0333	0.0191	1.005
가구주장애종류	1	0.0239	0.0202	1.3987	1.024
가구주장애등급	1	-0.0179	0.0274	0.4249	0.982
가구주만성질환	1	0.0255	0.0245	1.0888	1.026
-2 Log L	-	6972.173	***: P<0.001, **: P<0.01, *: P<0.05		
Wald 통계량	9	214.5797***			

다음은 가구에 방문한 회수를 모형에 투입하여 가구특성과 함께 분석한 결과 당연한 결과이지만, 가구에 대해 방문회수가 증가할수록 완료율이 높아지는 것으로 나타나며, 소득계층이 저소득일수록, 가구원수가 증가할수록 원표본 유지율은 높아지는 것으로 분석되며, 가구주의 성별이 남성일수록, 가구주의 학력이 높을 수록 원표본 유지율은 낮아지는 것으로 분석되었다.

〈표 3-4〉 가구방문회수와 가구의 인구학적 특성에 따른 완료 및 대체가구의 영향에 대한 로짓분석

변 수	자유도	추정계수	표준오차	Wald Chi-Square	odds 비
상수	1	-1.1175	0.2905	14.7993	
방문회수	1	0.5683***	0.0311	334.4531	1.765
소득계층	1	0.7127***	0.0663	115.6615	2.040
가구원수	1	0.1185***	0.0290	16.7239	1.126
가구주성별	1	-0.1740*	0.0799	4.7427	0.840
가구주연령	1	0.00517*	0.00261	3.9373	1.005
가구주학력	1	-0.1454***	0.0278	27.4403	0.865
-2 Log L		6537.788	***: P<0.001, **: P<0.01, *: P<0.05		
Wald	6	521.5186***			

방문차수에 따른 분석과는 달리 가구원수가 많을 경우 원표본 유지율이 높아지는 현상은 표본가구에 제공되는 인센티브의 영향이 클 것으로 사료되며, 대체로 저소득 계층의 원표본 유지율이 높은 이유 또한 이러한 영향인 것으로 판단된다.

4. 결론

통상적으로 가구면접조사의 경우 조사원이 표본가구를 방문하여 일대일 면접을 통한 조사를 수행하게 된다. 한편 패널조사와 같이 지속적으로 조사대상이 되는 표본가구에 대해 조사 설계시 표본으로 추출된 가구를 한번에 패널표본으로 구축하기란 현실적으로 불가능하다. 이러한 관점에서 본 연구에서는 가구를 방문하여 조사를 수행할 때 가구의 특성에 따라 조사완료율과 원표본 유지율이 어떻게 영향을 받는지를 통계적으로 분석하여, 이에 대한 적절한 대응방안을 모색하고자 하였다.

분설결과 우선 가구의 특성에 따른 방문회수에 영향을 주는 요소들로는 가구주의 연령, 가구원수, 가구주의 학력, 가구주의 혼인상태 등으로 분석되었고, 원표본 유지율 측면에서는 가구의 소득, 가구원수, 가주의 학력 등이 영향을 끼치는 것으로 분석되었다. 한편 가구주의 경제활동 상태나 업종, 직종에 따른 방문회수별 영향은 없는 것으로 나타났으며, 원표본 유지율에 대한 분석에서도 이러한 변수에는 영향을 받지 않는 것으로 나타났다.

향후 이러한 연구를 바탕으로 조사원에 대한 교육자료 활용이 가능할 것으로 사료되며, 무응답가구에 대한 가구의 특성에 대해서는 향후 연구과제로 남기고자 한다.

참고문헌

한국복지패널 1차년도 기초보고서(2006), 한국보건사회연구원.

최봉호(2006), 통계조사에서의 오차 및 대응방안, 한국조사연구학회 추계학술발표논문집, pp.219-235.

Allison, P, (1999), Logistic Regression,: Using the SAS system, John & Wiley International.

Groves, R.(1989), Survey Errors and Survey Cost, John & Wiley

Groves, R. and Couper, M.(1998), Nonresponse in Household Interview Surveys, John & Wiley Sons.

Kish, L.(1965), Survey Sampling, John & Wiley Sons.

Kauff, J., Olsen, R., and Fraker, T.(2002), Nonrespondents and Nonresponse bias : Evidence from a Survey of Former Welfare Recipients in Iowa, Methematica Policy Research, Inc.

Lessler, J. and Kalsbeek, W.(1992), Nonsampling Error in Surveys, John & Wiley Sons.

Littel, R. J.A. (1986), Rubin, D.(1987), Satistical Analysis with Missing Data, John & Wiley Sons.

Sarndal, C.E. and Lundstorm, S.(2005), Estimation in Surveys with Nonresponse, John & Wiley Sons.

Schrapler, J-U., Schupp, J.,and Wagner, G.(2005), An Analysis Based on a New Subsample of the German Soci-Economic Panel(SOEP) including Microgeographic Characteristics and Survey-Based interviewer Characteristics, SOEP working paper.