

한국노동패널연구 2003-04

## **KLIPS의 가중치 부여방안 연구**

2003. 7.

**강 석 훈**

(성신여대 경제학과 교수)

## 목 차

I. KLIPS표본추출과 실사과정 .....	1
1. 표본들과 모집단 .....	1
2. 1단계 조사구 추출과정 .....	2
3. 최종 조사가구 추출 .....	5
4. 실사단계에서의 표본조정 .....	6
II. 1차 웨이브 가중치 부여방법 .....	7
1. 1차 웨이브 가중치 부여방법 .....	7
2. 표본추출확률의 계산 .....	8
3. 응답률과 면접확률의 계산 .....	9
4. 1차 웨이브 가중치의 계산 .....	10
5. 계산결과 .....	11
III. 2차 웨이브 이후부터의 가중치 부여 방법론 .....	12
1. 종단면서베이에서의 가중치 부여에 대한 이슈 .....	12
2. 패널조사에서의 가중치 부여 방안에 대한 기존 연구 .....	13
3. KLIPS 2차 웨이브 이후부터의 가중치 부여 방법론 검토 .....	20
IV. 2차 웨이브 이후부터의 가중치 부여과정 .....	24
1. 2차년도 가중치 부여과정 .....	24
2. 3차년도 가중치 부여과정 .....	29
3. 4차년도 가중치 부여과정 .....	35
4. 각 웨이브에서의 가중치의 기초통계량 .....	40
참고문헌 .....	42

## I. KLIPS표본추출과 실사 과정

### 1. 표본들과 모집단

□ KLIPS 1차 웨이브에서는 1단계에서 조사구를 선정하고 2단계에서 가구를 선정하는 2단 층화집락계통추출법을 사용하였음.

□ (표본틀) 1997년 고용구조특별조사(이하 97고특)에서 사용한 1995년 인구주택총조사 10% 표본조사구인 전국의 21,938조사구 중에서 제주도 263조사구를 제외한 21,675조사구가 1차 대상임. 이 중에서 군부지역(2,650조사구)을 제외한 전국의 도시지역(시의 동부 17,353조사구와 시의 읍면부 1,672조사구)의 19,025 조사구가 표본틀임.

※ 97고특에서는 1995년 인구주택총조사의 10%표본조사구 22,049개 중에서 각 조사구별로 직업별, 취업자 구성을 분석하여 직업군인비율이 30%이상인 조사구(82개), 조사구당 가구수가 10가구 미만인 조사구(29개) 등을 삭제한 21,938개 조사구를 사용하였음.

○ 97고특에서 사용한 10%센서스 조사구에는 동부조사구가 17,353개(제주도 169개 동부 조사구 제외)있으므로, 나머지 1,672개가 시의 읍면부 조사구라고 할 수 있음. 이에 따라 10% 센서스표본조사구(제주도 제외)에 거주하는 가구는 동부조사구의 1,009,891가구와 읍면부 지역에 거주하는 100,141가구가 되고, 따라서 전체 도시조사구(시의 동부+시의 읍면부)에 19,025조사구에 거주하는 가구수는 1,110,032가구가 됨.

○ 10%인구주택총조사 표본임을 감안하면 결국 KLIPS표본의 모집단은 전국의 도시지역(동부와 시의 읍면부 포함) 거주하는 11,100,320가구임.

※ 이와 같은 모집단조사가구에 대한 계산은 1995년 인구주택총조사의 10%표본조사구는 모집단에서의 시의 동부와 시의 읍면부의 비율과 동일하게 추출되었으며, 모집단조사구와 표본조사구에서의 가구비율도 동일하다는 가정이 포함되어 있음.

## 2. 1단계 조사구 추출과정

□ (1차 층화) 지역을 우선으로 하고 동일지역 내에서는 1997년 고용구조특별조사의 층화방법을 사용하였음.

○ 지역은 서울특별시, 6개 광역시, 8개 도 등 15개로 구분하였음. 이렇게 구분된 지역 내에서 97고특의 층화방법을 이용하여 조사구를 층화하였음.

□ (조사구 계통추출) 1차적으로 1,000개의 조사구를 추출하고자 하였음.

○ 서울 및 6개 광역시의 경우 층화된 조사구들을 인구주택총조사 10%표본조사구에서의 전국도시조사구(19,025개) 대비 해당 지역의 도시조사구의 비율을 이용하여 추출할 도시조사구수를 결정하였음.

- 예를 들어 서울의 경우, 서울의 도시조사구수는 5,186개이고, 이는 인구주택총조사 10%표본에서의 전국도시조사구(19,025개)의 27.3%에 해당되므로 1,000개를 비례할당하는 경우 273개의 조사구가 추출됨.

○ 8개 도에서는 전국 도시조사구 대비 해당지역의 동부 및 시의 읍면부의 전국도시조사구 대비 비율을 이용하여 동부 및 시의 읍면부에서 추출할 조사구를 각각 결정하였음.

- 예를 들어 경기도 동부의 경우에는 인구주택총조사 10%표본에서의 전국도시조사구 19,025개 중에서 경기도의 동부조사구 2,837개가 차지하는 비중은 14.9%임. 1,000개를 비례할당하는 경우 149개의 동부조사구가 선정되어야 함. 경기도의 시의 읍면부조사구는 405개이고, 이는 인구주택총조사 10%표본의 전국도시조사구 19,025개의 2.1%이므로 1,000개를 비례할당하는 경우 21개의 시의 읍면부조사구가 표본으

로 추출됨.

□ (97고특조사구와의 matching) 계통추출방식으로 조사구를 선정할 때 계통추출된 조사구가 97고특조사구에 해당되면 선정된 조사구를 최종 표본조사구로 결정하였고, 97고특조사구에 해당되지 않으면 가장 근접한 97고특조사구를 선정하여 최종 표본조사구로 결정하였음.

○ 이러한 matching과정을 위해서는 <표 I-1>에서 결정한 조사구의 수가 97고특에서 사용한 조사구의 수보다 커야 하는데 일부 지역에서는 반대의 현상이 나타났음.

- 서울의 경우에는 273개 조사구를 추출하여야 하지만, 97고특조사구의 수는 227개에 불과하여 서울에서는 227개의 97고특조사구를 전부 추출하였음.

- 경기도 시의 읍면부의 경우에는 21개 조사구를 추출하여야 하지만, 97고특조사구에서의 시의 읍면부의 조사구수는 18개에 불과하여, 18개 전부를 추출하였음.

<표 1-1> 고특과의 matching을 감안한 최종 추출조사구수

지역	고특조사 구수	고특도시조 사구	최종표본조 사구	최종조사구 추출률(%)	최초 목표와 비교
서울	227	227	227	4.3772	고특조사구 최대 273개
부산	176	176	97	5.2689	동일
대구	155	155	64	5.2805	동일
인천	150	150	59	5.2867	동일
광주	106	106	32	5.2718	동일
대전	101	101	31	5.2365	동일
울산	70	70	24	5.2632	동일
경기	280	263	167	5.1511	고특읍면부에서 3개 감소
동부	245	245	153	5.3930	최초 149개에서 읍면부 4개이전
읍면부	35	14	14	3.4568	고특조사구 최대 14개
강원	165	121	26	5.1587	동일
동부	105	105	22	5.2009	동일
읍면부	60	16	4	4.9383	동일
충북	159	94	22	5.2133	동일
동부	84	84	20	5.2632	동일
읍면부	75	10	2	4.7619	동일
충남	122	73	25	5.1230	반올림의 문제로 1개감소
동부	40	42	13	5.5556	반올림의 문제로 1개감소
읍면부	82	31	12	4.7244	반올림의 문제로 1개증가
전북	170	119	40	5.2840	동일
동부	75	75	30	5.2174	동일
읍면부	95	44	10	5.4945	동일
전남	152	87	25	5.1335	반올림문제
동부	72	72	19	5.1351	동일
읍면부	80	15	6	5.1282	동일
경북	129	101	54	5.2889	동일
동부	75	75	37	5.3237	동일
읍면부	54	26	17	5.2147	동일
경남	220	182	58	5.3016	동일
동부	155	155	43	5.4282	반올림의 문제로 1개 감소
읍면부	65	27	15	4.9057	반올림의 문제로 1개 증가
전국	2,382	2,025	951	4.9987	

주 : 1) 고특조사구에서의 읍면부의 수는 전체 읍면부 조사구수를 의미하며, 고특도시조사에서의 읍면부의 수는 시에 속한 읍면부조사구의 수를 의미함.

2) 경남에서 기존에 동부에 포함되어 있던 894.경남 창원시 동읍, 912. 경남 마산시 내읍의 경우를 읍면부로 이전하였음. 이에 따라 동부는 45개에서 43개로 감소하고, 읍면부는 13개에서 15개로 증가하였음.

### 3. 최종 조사가구 추출

□ (조사가구 추출) 2단계에서 조사대상가구를 추출하였는데, 조사대상가구는 추출된 조사구 중에서 97고특에서 조사에 성공한 가구를 대상으로 서울 및 경기도의 읍면부 지역에서는 6가구, 기타지역은 5가구를 추출하였음.

○ 서울 및 경기도의 경우에는 추출하여야 할 조사구수보다 97고특에서 사용한 조사구수가 적었기 때문에 다른 조사구에 비해 1가구를 더 추출하였음. 한편, 경기도의 동부의 4개지역에서 5가구 대신 6가구를 추출하였음.

※ 이 들 지역은 640. 의정부시 장암동, 654. 평택시 서정동, 659. 동두천시 동암동, 682. 하남시 풍산동 등 4개 조사구임.

□ 최종 추출된 조사구와 조사가구의 수는 다음과 같이 요약가능함.

구 분	전체 모집단		10%표본조사구		표 본		추출률(%)	
	조사구	가구수	조사구	가구수	조사구	가구수	조사구	가구수
전국 시부	190,250	11,100,320	19,025	1,110,032	951	5,000	0.4999	0.0450
시의 동부	173,530	10,098,910	17,353	1,009,891	871	4,582	0.5019	0.0454
시의 읍면부	16,720	1,001,410	1,672	100,141	80	418	0.4785	0.0417

주 : 1) 10%표본조사구는 전국의 21,938조사구 중에서 제주도 지역 263개 조사구를 제외한 후 다시 군부의 읍면부 지역 2,650조사구를 제외하여 19,025개가 됨.

2) 10%표본조사구의 가구수는 97고특조사구에서 제시한 10%표본조사구에서의 조사구당 평균가구수를 이용하여 계산하였음. 구체적으로 97고특조사구에서 시의 동부는 17,522개의 조사구에 1,019,726가구가 있으므로 조사구당 평균가구수는 58.1969가구임. 한편, 97고특조사에서 10%표본조사구에 관한 자료는 시의 읍면부에 대한 자료가 제시되지 않고 전체 읍면부의 자료가 제시되고 있음. 이에 따라 전체 읍면부의 조사구당 가구수와 시의 읍면부의 조사구당 가구수가 일치한다고 가정하여 계산하였음. 구체적으로 10%표본조사구의 읍면부 조사구는 4,416개이고, 여기에 거주하는 가구수는 264,487가구이며, 따라서 평균가구수는 59.8929가구임.

3) 전체 모집단은 97고특에서 사용한 10%표본조사구에서 10을 곱하여 계산하였음.

4) 추출률은 전체 모집단 대비 추출률임.

#### 4. 실사단계에서의 표본조정

□ (최종응답가구 대체) 실제로 2단계에서 추출한 가구숫자는 목표표본가구였으며, 이러한 목표표본가구를 달성하기 위하여 조사구별로 접촉한 가구수는 5가구로부터 20가구까지 지역별로 매우 다양하게 나타나고 있음. 이와 같이 최종단계에서 대체되어 설문에 응답한 표본은 전체 조사가구의 24.5%인 1,227가구임.

□ (이사가구 추적) 최초에 추출된 조사대상가구 중에서 11.3%인 563가구가 이사를 하였음. 이러한 이사가구를 추적하여 이중에서 40.1%(전체 조사대상자 5,000가구의 4.5%, 무대체성공가구 3,773가구의 6.0%)인 226가구를 조사하였음.

<표 1-2> KLIPS 1차웨이브의 조사가구 개요

지역	총 접촉 가구수	최초 접촉 가구수	최초 성공 가구수	최초 성공률	대체 가구 접촉수	대체 성공 가구수	대체 접촉 성공률	최초성공 가구비율	대체성공 가구비율	최종 응답률
서울	2,109	1,362	954	70.0	747	408	54.6	70.0	30.0	64.6
부산	600	485	391	80.6	115	94	81.7	80.6	19.4	80.8
대구	452	320	239	74.7	132	81	61.4	74.7	25.3	70.8
인천	414	295	226	76.6	119	69	58.0	76.6	23.4	71.3
광주	197	160	134	83.8	37	26	70.3	83.8	16.3	81.2
대전	188	155	127	81.9	33	28	84.8	81.9	18.1	82.4
울산	150	120	100	83.3	30	20	66.7	83.3	16.7	80.0
경기	1,217	853	617	72.3	364	236	64.8	72.3	27.7	70.1
강원	200	130	94	72.3	70	36	51.4	72.3	27.7	65.0
충북	145	110	86	78.2	35	24	68.6	78.2	21.8	75.9
충남	152	125	107	85.6	27	18	66.7	85.6	14.4	82.2
전북	273	200	151	75.5	73	49	67.1	75.5	24.5	73.3
전남	164	125	101	80.8	39	24	61.5	80.8	19.2	76.2
경북	356	270	205	75.9	86	65	75.6	75.9	24.1	75.8
경남	359	290	241	83.1	69	49	71.0	83.1	16.9	80.8
전국	6,976	5,000	3,773	75.5	1,976	1,227	62.1	75.5	24.5	71.7



## II. 1차 웨이브의 가중치 부여방법

### 1. 1차 웨이브 가중치 부여 방법

□ 1차 웨이브의 경우에는 일반적인 횡단면조사에서의 가중치 부여 방안이 그대로 적용될 수 있음.

○ KLIPS는 패널조사이지만 1차 웨이브 자체만으로는 횡단면조사이므로 1차 웨이브의 가중치는 일반적인 횡단면조사에서의 가중치부여 방법을 사용할 수 있음. 그러나 2차 웨이브 이후에는 횡단면가중치(cross-sectional weights) 뿐만 아니라 종단면가중치(longitudinal weights)를 구하여야 하기 때문에 1차년도 가중치를 계산할 때에도 2차 웨이브 이후를 처음부터 고려해야 함.

○ 일반적인 횡단면조사에서 가중치 부여방법은 기본적으로 다음과 같은 3단계과정을 거쳐서 계산됨.

- (1단계 : 표본추출확률(selection probability)계산) : 대부분의 횡단면조사는 표본추출확률이 상이하므로 이를 조정하기 위해 1단계에서는 표본추출확률을 계산함. 이 때 표본의 모집단에 대한 부분포함(incomplete coverage)의 문제도 동시에 고려함.

- (2단계 : 무응답(nonresponse)조정) : 무응답의 경우를 조정하기 위해 사용될 수 있는 방법으로는 일반적으로 알려진 변수, 예를 들면 표본이 속한 지역에 대한 자료를 통해 표본을 분할하는 방법, 모든 표본의 알려진 성질을 이용하여 회귀분석( 또는 로짓분석)을 실시하는 방안 등이 사용됨.

- (3단계 : 사후층화(poststratification)) : 1단계와 2단계를 거친 후 나타난 표본의 성질을 외부적으로 알려진 전체모집단의 성질과 근사시키는 과정임.

- 최종적인 가중치는 위에서 구한 각 단계의 확률, 비율 등을 곱하여 산출함.

○ 그러나 모든 서베이에서 위와 같은 세 단계를 모두 거쳐서 가중치를 계산하는 것은 아니며 예를 들어 PSID는 1단계 및 2단계, BHPS, SLID는 세 단계를 모두 거치고 있음. 또한 각 단계에서 사용하는 방법도 서베이마다 다양한 방법이 사용되고 있음.

□ KLIPS의 경우 분석단위로서 개인과 가구가 모두 사용가능하며, 이에 따라 개인 가중치와 가구가중치의 관계를 어떻게 설정하느냐의 문제가 제기됨.

○ KLIPS의 경우 1차 웨이브에서는 가구와 그 가구에 속한 가구원의 추출확률이 동일하므로 가중치를 추출확률로만 계산하는 경우 개인가중치와 가구가중치는 동일함. KLIPS에서는 가구가 응답한 경우 대부분의 가구원이 응답하였으므로 응답확률을 고려하더라도 개인가중치와 가구가중치는 동일하게 됨.

## 2. 표본추출확률의 계산

□ Part 1.에서 설명한 표본추출방법 하에서 추출확률은 다음과 같이 계산할 수 있음.

□ (표본조사구 추출확률) 표본조사구로 선정되려면 먼저 1단계에서 센서스 10% 표본조사구로 선정되고 다시 2단계에서 KLIPS의 표본조사구로 선정됨. 1단계 센서스 10%표본조사구의 추출확률은 모든 지역에서 10%로 동일하다고 가정함. 2단계에서는 먼저 KLIPS표본조사구로 선정되고 다시 이 표본조사구가 고득조사구에 matching될 확률을 계산하여야 하는데 후자의 경우에는 확률계산이 불가능함. 이에 따라 조사구 추출확률은 다음과 같이 계산함.

○ 지역별 도시조사구의 추출확률은 지역별 전체조사구 중에서 도시조사구의 비율을 이용하여 계산함.

○ 도시조사구중에서 표본조사구로 추출될 확률은 도시조사구 중에서 표본조사구의 비율을 이용하여 계산함.

- 지역별 표본조사구 추출확률
  - 서울 및 6대 광역시 :  $0.1 * (\text{해당 시의 표본조사구수} / \text{해당 시의 조사구수})$
  - 도의 동부 :  $0.1 * (\text{해당 도의 동부 표본조사구수} / \text{해당 도의 동부 조사구수})$
  - 도의 시에 속한 읍면부 :  $0.1 * (\text{해당 도의 시에 속한 읍면부 표본조사구수} / \text{해당 도의 시에 속한 읍면부 조사구수})$

□ (표본조사구에서 최종표본가구로 추출될 확률) 엄밀한 의미에서 보면 특정 조사구에서 특정가구가 선정될 확률은 동일한 조사구내에서 특정가구를 접촉하기 전에 미리 접촉시도한 가구의 응답여부 또는 응답가구수에 따른 조건부확률로 계산하여야 함.

○ 예를 들어 서울지역의 어떤 조사구에서 7가구가 추출되었다면 6번째 가구까지는 동일한 조사구내에서 임의적으로 선정되었기 때문에 동일한 추출확률을 가지고 있지만, 일곱 번째 접촉시도한 가구가 추출될 확률은 이전까지 접촉 시도한 6가구 중에서 정확하게 1가구가 응답거절인 경우 7번째 가구가 선정될 조건부 확률임. 따라서 서울의 경우 6가구를 초과하는 가구의 추출확률, 기타지역의 경우 5가구를 초과할 경우의 가구추출확률은 상이하게 됨. 그러나 응답거절가구에 대한 정보가 전혀 없기 때문에 이러한 조건부확률을 구하기는 현실적으로 불가능함.

- 최종표본가구 추출확률
  - $(\text{총접촉 가구수} / 97\text{고특응답가구수}) * (97\text{고특응답가구수} / \text{ED내 전체 가구수})$

□ 가구추출확률 : 조사구추출확률과 최종표본가구 추출확률의 곱으로 계산됨.

### 3. 응답률과 면접확률의 계산

□ 일반적으로 응답률조정을 위해서는 무응답가구에 대한 특성과약이 전제되어야 함.

- KLIPS 1차 웨이브의 경우 무응답가구의 지역정보(광역시 또는 기타 道の 동부

및 읍면부) 또는 ED를 이용하여 응답률을 계산할 수 있음. 지역정보를 이용하는 경우 동일 지역 내에서 응답자와 무응답자의 특성이 같다는 것을 가정하는 것이며, ED 정보를 이용하는 경우 ED내의 응답자와 무응답자의 특성이 같다는 것을 가정하는 것과 같음.

※ 응답률을 고려하지 않는 경우는 전지역에서 응답자와 비응답자의 특성이 동일하다는 가정인바 이는 지역별 응답률 또는 ED내 응답률이 동일하다는 가정보다 강한 가정임.

- KLIPS 1차 웨이브의 응답률은 동일한 ED내에서 응답자와 무응답자의 특성이 같다는 가정 하에 ED내에서의 응답률을 면접확률로 이용함. 이에 따라 응답률은
  - 최종조사가구수/총접촉 가구수

#### 4. 1차 웨이브 가중치의 계산

추출확률과 응답확률을 모두 고려한 가구가중치는 다음과 같이 계산됨.

- 서울 및 6대 광역시 :  $0.1 \times (\text{표본조사가구수} / \text{도시조사가구수}) \times (97\text{고특조사가구수} / \text{ED내 전체 가구수}) \times (\text{총접촉 가구수} / 97\text{고특조사가구수}) \times (\text{최종조사가구수} / \text{총접촉 가구수})$
- 도의 동부 :  $0.1 \times (\text{해당 도의 동부 표본조사가구수} / \text{해당 도의 동부 조사가구수}) \times (97\text{고특조사가구수} / \text{ED내 전체 가구수}) \times (\text{총접촉 가구수} / 97\text{고특조사가구수}) \times (\text{최종조사가구수} / \text{총접촉 가구수})$
- 도의 읍면부 :  $0.1 \times (\text{해당 도의 시에 속한 표본읍면부 조사가구수} / \text{해당 도의 시에 속한 읍면부 조사가구수}) \times (97\text{고특조사가구수} / \text{ED내 전체 가구수}) \times (\text{총접촉 가구수} / 97\text{고특조사가구수}) \times (\text{최종조사가구수} / \text{총접촉 가구수})$

선출된 가구에서는 모든 가구원이 응답하였으므로 동일 가구 내에서는 가구가중치와 가구원가중치가 동일하게 됨.

※ 0.1을 곱하는 것은 전국 모집단의 동부 및 시의 읍면부의 비율과 10%표본에서의 동부 및 시의 읍면부 비율이 같다고 가정하고 있는 것임.

최종적으로 추출확률과 응답률을 감안하면 특정한 가구의 가중치는 가구추출확률과 가구가 속한 지역의 응답률의 곱의 역수로 나타남.

사후층화의 고려

○ 사후층화는 가구단위에서의 사후층화와 개인단위에서의 사후층화를 고려할 수 있음.

- 그러나 가구단위의 정보는 1995년 센서스에 의한 자료가 1차 웨이브에 가장 최근 자료인바 이를 이용하여 사후층화를 하는 것은 시의성이 떨어짐.

- 개인단위에서 사용할 수 있는 자료는 통계청의 인구추계자료가 있으나 이는 지역별, 연령별 세분화가 되어 있지 않음(서울 및 광역시 지역의 경우에는 연령별, 성별 인구분포가 있으나 기타지역의 경우에는 전체 지역에 대한 연령별, 성별추계는 있으나 시지역에 대한 추계는 없음).

- 기타 경제활동인구조사나 고용구조특별조사의 결과를 이용하여 사후층화를 고려할 수도 있으나 양 조사 모두 표본조사이므로 이들 자료를 이용하여 사후층화하는 것은 사후층화의 의의가 떨어짐.

이와 같은 점을 고려하여 사후층화는 하지 않음.

○ 사후층화를 고려하지 않음에 따라 결과적으로는 동일한 가구내의 가구원은 모두 동일한 가중치를 부여받게 됨.

## 5. 계산결과

가구가중치의 기초통계량(w1\_h)은 다음과 같음. Klips 1차 웨이브에서 1가구는 전국적으로 평균 2,255가구를 대표하는 것으로 나타났으며, 최소값은 1,513가구, 최대

값은 4,515가구로 나타났음.

□ 개인가중치의 기초통계량(w1\_p)은 다음과 같음. 평균은 2255.04, 표준편차는 416.76, 그리고 최대값, 최소값은 가구가중치와 같음.

※ w1\_h의 경우에는 가중치의 합이 11,276,900가구로 scale조정을 하였으나, w1\_p의 경우에는 가중치의 합에 대한 모수값(또는 추정모수값)을 알 수 없어서 scale조정을 하지 않았음.

### Ⅲ. 2차 웨이브 이후부터의 가중치 부여 방법론

#### 1. 종단면서베이에서의 가중치 부여에 대한 이슈

□ 종단면서베이의 1차 웨이브는 위에서 제시한 횡단면서베이에서의 가중치부여방법을 이용할 수 있으나, 2차 웨이브 이후부터는 횡단면분석 이외에 종단면분석이 이루어지기 위해서는 다음과 같은 네 가지의 가중치를 고려할 수 있음.

○ 횡단면가중치 : 횡단면가구가중치(cross-sectional household weights, CHW), 횡단면개인가중치(cross-sectional individual weights, CIW).

○ 종단면가중치 : 종단면가구가중치(longitudinal household weights, LHW), 종단면개인가중치(longitudinal individual weights, LIW)

□ 종단면서베이에서는 횡단면서베이에서는 발생하지 않는 사항들을 고려하여야 함.

○(표본탈락, panel attrition) 각 웨이브간에 표본탈락(panel attrition)이 발생함. 표본탈락은 가구차원에서 발생하거나 개인차원에서 발생할 수 있으며, 항목무응답

(item nonresponse)도 발생할 수 있음(항목무응답의 경우는 가중치부여 방법 보다는 보정(imputation)방법이 주로 사용됨).

○ (비표본가구원, nonsample household) 결혼, 이민 등으로 인해 표본가구에 새로운 개인이 포함되거나 탈락하기도 하며, 표본가구원 사이에서 새로운 가구원이 탄생하기도 함. 이에 따라 웨이브가 진행되는 경우 가구구성이 달라지게 되어 종단면가중치를 부여할 수 있는 가구의 개념이 모호해짐.

□ 이와 같은 개념상의 문제로 인하여 2차 웨이브 이후부터는 횡단면가구가중치와 횡단면개인가중치, 종단면개인가중치는 산출하나 종단면가구가중치에 대해서는 이를 구하는 경우와 구하지 않는 경우로 대별되고 있음.

○ BHPS, GSEP 등에서는 종단면가구가중치를 구하지 않고 있으나, PSID, SIPP 등에서는 종단면가구가중치를 구하고 있음.

## 2. 패널조사에서의 가중치 부여 방안에 대한 기존 연구<sup>1)</sup>

□ 2차 웨이브 이후의 가중치 부여방법에 대해서는 각 서베이마다 상이한 방법을 사용하고 있으며, 특히 비표본가구원(nonsample household members)에 대한 취급방법이 상이하게 나타나고 있음.

□ 1차 웨이브 이후의 횡단면가중치는 결국 비표본가구원의 추출확률과 밀접한 관련이 있으나 실제로는 이 추출확률을 구하지 않고도 불편추정량을 얻을 수 있는 가중치부여방법이 존재함.

---

1) 이 부분은 Kalton et al(1994)를 주로 참고하였음.

## 2.1 횡단면가구가중치

□ 모집단가구(모집단가구수  $H$ )의 변수  $Y$ 의 총합,  $Y = \sum_{i=1}^H Y_i$ , 을 구하는 추정량으로 다음을 고려할 수 있음.

$$\widehat{Y} = \sum_{i=1}^H w_i Y_i$$

○ 단  $w_i$ 는 가구  $i$ 가 표본에 속해 있을 때 1의 값을 가지고, 아니면 0의 값을 가지는 확률변수임. 이 때  $\widehat{Y}$ 가  $Y = \sum_{i=1}^H Y_i$ 의 불편추정량이 되기 위해서는  $E(w_i) = 1$ 가 만족되어야 함.

- 이러한 조건을 만족하는 방법으로 추출확률역수가중치 부여방법(inverse selection probability weighting scheme, ISPWS), 동등가구가중치부여방법(equal household weighting scheme, EHWS), 동등개인가중치부여방법(equal person weighting scheme, EPWS) 등이 있음.

□ ISPWS : 가구  $H_i$ 가 표본에 속할 확률은 다음과 같이 표현가능함.

$$\begin{aligned} P(H_i) &= P(h_j \cup h_k \cup h_l \dots) \\ &= \sum P(h_j) - \sum \sum P(h_j \cap h_k) + \sum \sum \sum P(h_j \cap h_k \cap h_l) \dots \end{aligned}$$

단,  $h_j, h_k, h_l, \dots$ 는 적어도 한명의 가구원이 현재의  $H_i$ 가구에 소속되어 있는 경우를 말함. 이 때의 가구가중치는 다음과 같음.

$$w_i = 1/P(H_i)$$

○ 이러한 가중치는  $E(w_i) = P(H_i)[1/P(H_i)] + [1 - P(H_i)]0 = 1$ 이 되어 불편성의 조건을 만족함. 그러나 ISPWS를 사용하려면 모든 현재 표본가구의 모든 구성원의 원시가구가 최초웨이브에서 추출될 확률과 이들의 결합추출확률을 알아야 하는데, 원시표본가구로 선정된 가구에 대해서는 이 확률이 알려져 있으며, 그렇지 않은 가구들의 경우와 결합확률은 알려져 있지 않기 때문에 현실적으로 사용하기가 어려움



(Rendtel(1992)은 결합확률을 무시하고, 로짓모형들을 이용하여 비표본가구의 추출 확률을 계산하는 방법을 제안한바 있음). 아래의 대안들을 비표본가구의 추출확률을 이용하지 않고 원시표본가구의 추출확률만을 이용하여 가중치를 구하는 방법임

□ EHWS : 다음과 같은 가중치를 고려함.

$$w_i = \sum_j \alpha_{ij} \cdot w'_{ij}$$

단  $w'_{ij}$ 은 적어도 하나의 가구구성원이 현재의 표본가구  $H_i$ 에 소속되어 있는 가구  $h_j$ 가 원시표본가구로 선정된 경우에는  $1/p_j$ 의 값을 가지고, 그렇지 않은 경우에는 0의 값을 가지는 변수임.  $p_j$ 는 가구  $h_j$ 의 추출확률, 즉  $p_j = P(h_j)$ 이며,  $\sum_j \alpha_{ij} = 1$ 임.

○ 이제  $\alpha_{ij}$ 의 값을 가구  $H_i$ 에 소속되어 있는 가구원들의 원시표본가구수를 이용하면 가중치는  $w_i = \sum w'_{ij}/C_i$ 가 됨. 단  $C_i$ 는 가구에 속해 있는 가구원들의 원시표본가구수를 의미함. 이 방법이 EHWS이며, Huang(1984)는 multiplicity approach라고 명명한 바 있음.

□ EPWS : 다음과 같은 가중치를 고려함.

$$w_i = \sum_j \sum_k \alpha_{ijk} \cdot w'_{ijk}$$

단  $w'_{ijk}$ 은 가구  $h_j$ 의 구성원  $k$ 가 원시표본가구원인 경우에는  $1/p_j$ 의 값을 가지고, 그렇지 않은 경우에는 0의 값을 가지는 변수임.  $p_j$ 는 가구  $h_j$ 의 추출확률, 즉  $p_j = P(h_j)$ 이며,  $\sum_j \sum_k \alpha_{ijk} = 1$ 임.

○ 이제  $\alpha_{ijk}$ 의 값을 원시표본으로 추출될 수 있었던 모든 가구원의 수를 이용하면, 가중치는  $w_i = \frac{1}{M_i} \sum M_{ij} w'_{ij}$ 가 됨. 단 가구  $H_i$ 의 가구원 중에서 1차 웨이브에서 추출된 가구와 동일한 구성원인 모든 경우에  $w'_{ij} = w'_{ijk}$ 은 상수이며,  $M_{ij}$ 는 가구  $H_i$ 에 소속되어 있는 가구원 중에서 가구  $h_j$ 로부터 온 가구원의 수를 의미하며,

$M_i = \sum_j M_{ij}$ 는 가구  $H_i$ 에 소속되어 있는 가구원 중에서 1차 웨이브에서 원시표본으로 추출될 수 있었던 개인의 수를 의미함.

○ 이 방법이 EPWS이며, Huang(1984), Ernst(1989)에서는 fair share weighting scheme으로 명명되고 있음. 이 방법은 SIPP와 SLID에서 사용하는 방법임.

※ 예시1 : 현재 표본가구  $i$ 에 소속되어 있는 가구의 구성원이, 원시표본가구 A로부터 온 가구원  $a_1, a_2$ 와 원시표본가구 B로부터 온 가구원  $b_1$ 과 비원시표본가구 C로부터 온 가구원  $c_1$ 으로 구성되었다고 가정하고, 원시웨이브에서 가구 A, B가 추출될 확률을 각각  $1/100, 1/200$ 이라고 가정함. 이 때 EHWS에서는  $w_{iA}=1/(1/100), w_{iB}=1/(1/200), w_{iC}=0$ 이며,  $C_i=2$ 가 됨. 따라서 가중치는  $w_i=(1/2)(100+200)$ 이 됨. 이 때 EPWS에서는  $w_{iAa1}=1/(1/100), w_{iAa2}=1/(1/100), w_{iBb1}=1/(1/200), w_{iCc1}=0$ 이며,  $M_{iA}=2, M_{iB}=1, M_{iC}=1$ 이며,  $M_i=2+1+1=4$ 가 됨. 따라서 가중치는  $w_i = (1/4)(2*(1/(1/100))+1*(1/(1/200))+1*0)$ 이 됨.

※ 예시2 : 두 개의 원시가구,  $h_1, h_2$ 의 추출확률을 각각  $p_1, p_2$ 라고 하고, 결합확률을  $p_{12}$ 라고 하면, ISPWS에서의 가구가중치는 다음과 같음.

$$w_i^* = \frac{1}{p_1 + p_2 - p_{12}}$$

EPWS에서는 만약  $h_1$ 만이 원시표본가구라면,  $w_i = P_1/p_1$ 이 되며, 만약  $h_2$ 만이 원시표본가구라면  $w_i = P_2/p_2$ 가 되며, 만약 가구  $h_1, h_2$ 각 모두 원시표본가구라면  $w_i = (P_1/p_1) + (P_2/p_2)$ 가 됨. 단  $P_1, P_2$ 는 각각 (모집단에서 신규진입가구원을 제외하고) 가구  $H_i$ 에 속해 있는 가구  $h_1, h_2$ 로부터 온 가구원의 비율을 의미함. 이 가중치의 기대값은 ISPWS의 가구가중치와 같다는 사실을 보일 수 있음. EHWS는  $P_1 = P_2 = 1/2$ 가 되며, 또한 기대값이 가구가중치와 같다는 사실을 보일 수 있음.

□ EPWS나 EHWS는 모두 현재 가구원 중에서 원시표본가구에 속한 가구의 추출확률만을 이용하여 가중치는 부여하는 방법으로 불편추정량이며, ISPWS의 경우보다 분산은 증가하나 현실적으로 적용가능함. 만약 원시표본가구가 유사한 수준의 동등확률로 추출되었다면 가구나 개인차원의 분석에서 EHWS가 이러한 분산증가를 제어하기 위한 준적절한(near optimal) 방법임. 한편, non-epsem표본인 경우에는 EHWS나 EPWS는 모두 하위적절한(suboptimal) 방법임. 그러나 이 경우에도 적절한 가중치부여방법은 알 수 없는 최초의 추출확률에 의존하기 때문에 EHWS나 EPWS가 모두 사용될 수 있음. 다만 이 경우에도 EHWS는 1차 웨이브에서의 가구 구성에 대한 정보(현재의 표본가구에 속해 있는 원시표본가구원들이 동일한 원시표본가구에서 추출되었는지의 여부, 즉 원시표본가구의 수)를 필요로 하나 EPWS에서는 (전체 원시표본가구원수만 필요하고, 원시표본가구수는) 필요하지 않기 때문에 EPWS가 보다 선호됨.

## 2.2 횡단면개인가중치

□ 개인가중치를 구할 때 모든 신규진입자를 제외하고 1차 웨이브부터 현재 웨이브까지 모두 응답한 개인들을 대상으로 분석할 것인가 또는 원시표본가구원을 포함하지 않는 가구에 거주하는 신규진입자들은 포함되지 않는다는 점을 받아들이면서 현재 웨이브에 존재하는 모든 개인을 분석대상으로 할 것인가를 결정하여야 함. 여기서는 후자의 접근방법을 취함.

○ 모집단개인의 총합,  $Y = \sum_{i=1}^H \sum_{k=1}^{N_i} Y_{ik}$ 에 대한 추정량으로 다음의  $\hat{Y}$ 을 고려할 수 있음.

$$\hat{Y} = \sum_{i=1}^H \sum_{k=1}^{N_i} w_{ik} Y_{ik}$$

단  $Y_{ik}$ 는 가구  $H_i$ 에 소속되어 있는 가구원  $k$ 의 변수값이며,  $H$ 는 전체가구의 수이고,  $N_i$ 는 가구  $H_i$ 의 가구원수임. 또한  $w_{ik}$ 는 가구  $H_i$ 에 소속되어 있는 가구원  $k$ 가 표본에 포함되어 있지 않으면 0의 값을 가지는 확률변수임. 이 때  $\hat{Y}$ 가  $Y$ 에 대

한 불편추정량이 되기 위한 조건은  $E(w_{ik})=1$ 임.

○ 이러한 불편성 조건을 만족하는 가장 단순한 방법으로 원시표본가구원은 추출 확률의 역수를 이용하고, 기타의 신규진입자의 가중치는 0으로 하는 경우임. 이 방법은 출생이나 응답가능연령에 도달한 경우에도 약간의 보정과정을 거쳐 적용될 수 있음. 예를 들어 원시표본가구원이 출생한 신생아의 경우에는 그 어머니의 가중치와 동일하게 부여할 수 있고, SIPP의 경우처럼 서베이의 모집단을 16세이상 개인으로 정의하고, 최초 표본가구원에 16세 이하의 가구원이 포함되어 있는 경우 이 가구원은 주어진 확률로 추출된 표본처럼 취급하고, 나중에 16세가 되어 응답가능한 연령이 되었을 때 분석대상에 포함시키는 방법을 사용할 수 있음. 분석대상을 원시표본가구원과 일정한 신규진입자(주로 원시표본가구에서 출생한 아동)으로 한정하는 방법을 PSID에서는 사용하고 있으며, 이 방법은 신규진입자의 정보를 이용하지 않는다는 점이 단점임.

○  $w_i$ 를 가구  $H_i$ 의 모든 구성원의 가중치라고 하면(동일한 가구내의 모든 가구원의 가중치는 동일),  $w_i$ 는 다음과 같이 쓸 수 있음.

$$w_i = \sum_j \sum_k \alpha_{ijk} w'_{ijk}$$

단  $w'_{ijk}$ 은 가구  $h_j$ 가 원시표본가구이면 원시표본가구의 추출확률의 역수, 즉  $1/p_j$ 를 갖고, 아니면 0의 값을 가지는 변수임. 만약  $\sum_j \sum_k \alpha_{ijk} = 1$ 이면 불편추정량이 됨.  $\alpha_{ijk}$ 를 모든 원시가구원에 대하여  $1/M_i$ 를 부여하는 경우가 EPHS방법이며, 이때 가구  $H_i$ 의 모든 가구원(신규진입자 포함)은  $w_i = \sum_j \sum_k w'_{ijk} / M_i$ 의 가중치를 부여 받게 됨. 단  $M_i = \sum_j M_{ij}$ 로서 가구  $H_i$ 에 존재하는 원시가구원의 수임. 만약 원시가구  $h_j$ 의 구성원에 대해  $\alpha_{ijk} = 1/C_i M_{ij}$ 를 부여하는 방법이 EHWS임. 단  $C_i$ 는 현재의 가구  $H_i$ 에 구성원이 있는 원시가구  $h_j$ 의 수임.

□ EPWS, EHWS의 두가지 방법 모두 불편추정량임. 이 접근방법은 기본적으로 표본가구원에 대하여 원시표본가구원인가 아니냐에 관계없이 동일한 가중치를 부여하는 방법임. 전자는 현재 가구에 속한 원시표본가구원의 가중치를 원시웨이브에서 응답가능했던 가구원으로 나눈 평균가중치를 현재의 모든 가구원에 적용하는 방법이며, 후자는 현재 가구에 속한 원시표본가구원의 가구가중치를 현재 가구원의 원시표본가구의 수로 나눈 평균가중치를 현재의 모든 가구원에 적용하는 방법임.

### 2.3 가구가중치와 개인가중치의 관계

□ 위에서 고려한 가중치부여방법은 기본적으로 가구가중치와 개인가중치가 동일하며, 무응답보정을 가구차원에서만 하는 경우에는 이러한 성질은 유효하고, 가구에 대한 별개의 가중치를 기록할 필요가 없음(가구원의 가중치와 동일하므로).

○ 그러나 개인차원의 무응답을 보정하거나 무응답보정을 가구차원에서만 하더라도 개인차원의 사후층화를 하게 되면 가구가중치와 개인가중치의 동일성은 유지되지 않으며, 동일가구라고 하더라도 가구원의 가중치가 모두 달라지게 됨.

○ 이 경우는 가구의 중요구성원(principal person)의 가중치를 가구가중치로 사용하는 방법을 사용할 수 있음(SIPP에서 사용하는 방법임). 이러한 방법을 통해 가구가중치를 별도로 계산해야 하는 부담을 회피할 수 있으나 서베이의 중요목적이 개인차원에서의 분석이라면 다음과 같은 문제점이 발생함. 첫째, 가중평균된 가구수 또는 중요변수의 분포가 알려진 모집단의 가구수나 분포와 일치하지 않게 될 수 있음. 둘째, 중요구성원의 가중치를 사용하는 경우에는 개인차원의 사후층화과정에서 특정한 계층의 가중치가 높게 결정되는 경우 가구의 분포가 왜곡될 수 있음. 따라서 가구차원의 분석이 서베이의 중요한 목적 중의 하나라면 별도의 가구차원의 가중치를 고려하는 것이 바람직함.

### 3. KLIPS 2차 웨이브 이후부터의 가중치 부여 방법론 검토

#### 3.1 Duncan 방법론의 장점

□ 본 연구에서는 Duncan(1995)의 방법론을 이용하고자 함. 이 방법을 사용하는 이유는 다음과 같이 요약 가능함.

○ KLIPS의 표본추적원칙이 PSID와 기본적으로 동일하며, 이에 따라 PSID의 가중치부여 방안을 이용하는 것이 논리적으로 일관성이 있음.

○ Duncan의 방법론이 사용하는 가정이 가장 약하다고 볼 수 있으며, 계산과정이 비교적 간단하여 쉽게 이해할 수 있음. 또한 비록 종단면서베이에서 가구의 개념이 문제시되기는 하나 가구단위의 분석이 현실적으로 이루어질 수밖에 없다는 점도 고려되어야 함. Duncan의 방법론은 2차 웨이브 이후에도 가구단위의 가중치를 제공하고 있음.

#### 3.2. Duncan방법론의 세부 내용

□ Duncan(1995)의 가중치 부여방법은 다음과 같은 구체적인 절차를 가짐.

① 최초 웨이브에서 가구차원의 가중치를 구함. 이때 표본추출과정에서 사용된 상이한 추출확률을 감안하여야 하며, 가능한 경우 상이한 응답률로 보정함. 또한 마지막 단계에서 외부적으로 이용가능한 모집단의 정보가 있다면 이러한 사항들을 비율조정(ratio estimating procedure)을 이용하여 적용함.-> 이 과정은 횡단면조사의 가중치 부여방안과 기본적으로 동일함. 1차 웨이브 가중치부여 방안에서 지적한 바와 같이 가구가 추출되면 그 가구에 속한 가구원은 대부분 응답하였기 때문에 가구차원의 무응답조정만을 실시함(ED를 하나의 weighting class로 간주하고 ED내에서의 무응답은 random이라고 가정함). 한편, 도시지역의 모집단에 대한 정보가 없으므로 사후층화는 하지 않음.

② 최초 웨이브에서 작성된 가구가중치를 연령이나 응답여부에 관계없이 모든 가구의 개인가중치로 사용함.-> 실제로 표본가구내에서 가구의 무응답이 거의 없는 경우에는 이러한 방법이 타당성을 가짐. 즉 KLIPS의 경우에는 비록 가구용설문서 응답자와 가구원 설문서 응답자가 상이하였으나 가구용설문서에 응답한 가구의 경우 대부분 가구원설문서에도 응답하였다는 점에서 가구와 가구원의 추출확률 및 응답확률이 동일하다고 인정할 수 있을 것임. 이 가중치는 응답한 15세 이상의 가구원뿐만 아니라 15세 이하의 모든 가구원에게도 동일하게 적용함.

③ 2차 웨이브 이후부터는 가구원들의 상이한 응답률을 이용하여 개인가중치를 조정함. 이 단계에서는 가구와 가구원의 정보를 모두 이용함. 예를 들어 응답여부를 나타내는 로짓모형을 설정하여 모형을 추정한 후 이 계수를 사용하여 모든 가구원들의 응답확률을 추정함. 이 응답확률의 역수를 최초 개인가중치에 곱하여 2차 웨이브에서 무응답조정가중치(nonresponse adjusted weights)를 산출함. 이 때 2차 웨이브에서는 존재하지만 1차 웨이브에서 응답하지 않았던 비표본가구원(nonsample adults)이나 1차 웨이브 이후 새롭게 태어난 가구원(new-borns)의 경우는 개인차원의 무응답조정과정에 포함하지 않음.

→ 로짓모형을 구성하는 경우 가구의 정보나 가구원의 개인정보를 모두 사용하여 응답확률을 계산할 수 있으며, 무응답과 관련하여 개인을 가능한한 최대도로 차별화하는 모형을 구성할 필요성이 있음

→ 무응답조정을 하는 대상이 전체 가구원이나 아니면 15세 이상 가구원이나에 대한 논란이 있음. Duncan의 경우 전체 가구원을 대상으로 무응답을 조정하여야 한다고 주장하지만, 실제 15세 이상의 가구원과 15세 미만 가구원의 응답패턴이 다름(즉 15세 미만의 경우에는 기록만으로 응답하는 것으로 간주됨). 이에 따라 분석대상 모집단을 15세 이상 인구로 한정하고 이들, 즉 응답대상자들만을 대상으로 무응답분석을 실시하는 것이 바람직한 방법으로 판단됨.

→ 만약 전체를 대상으로 한다면 BHPS의 기록가구원가중치가 될 것임. 이 때는 분석대상이 15세 이상 가구원이 아니라 전체 인구가 분석대상이 될 것인 바, KLIPS의 주된 분석목적이 전체 인구가 되기 보다는 15세 이상 가구원이므로 이러한 가중치의

효용성은 감소됨.

④ 2차 웨이브에서 산출된 개인가중치의 가구내 평균을 이용하여 2차 웨이브의 가구가중치를 산출함. 원시표본가구원과의 결혼, 동거 등의 사유로 새롭게 진입한 비표본가구원의 경우에는 0의 가중치를 사용함. 당해 연도에 새롭게 태어난 가구원은 이 평균계산에는 사용하지 않음.

※ 예를 들어 2차 웨이브에서 원시표본가구원(16세 이상)이 2명이고, 비원시표본가구원 1명(16세 이상)이 새롭게 동거하고, 1명이 새롭게 태어났다고 가정함. 이 때 2차 웨이브에서의 가구가중치는 원시표본가구원 2명의 개인가중치의 합을 새롭게 태어난 가구원을 제외한 현재 가구의 전체 가구원 3명으로 나누어 계산한 평균을 사용함. 이렇게 계산된 평균을 새롭게 태어난 아기의 개인가중치로 부여함. 따라서 개인가중치를 가지는 가구원은 원시표본가구원 2명과 새롭게 태어난 아이 등 3명이 되며, 비표본가구원 1명의 개인가중치는 0이 됨. 개인차원의 분석에서 개인가중치를 사용하게 되면 비표본가구원을 제외하고 분석하게 되며, 비표본가구원을 포함하여 분석하려면 (평균으로 계산된) 가구가중치를 사용함.

→ 이러한 가중치부여과정을 거치게 되면 가구 차원의 무응답조정가중치가 산정되며 최초 웨이브에 존재했던 가구원(즉 원시표본가구원)의 개인차원의 무응답조정가중치가 산출됨. 비표본가구원의 경우는 0의 개인가중치를 가지며 새롭게 태어난 가구원의 경우에는 이들이 태어난 웨이브에서의 가구가중치를 개인가중치로 부여받게 됨.

※ 아래의 경우는 15세 이상만을 대상으로 무응답분석을 실시하는 경우에 해당됨.

→ 15세의 경우가 문제가 되는데 15세가 되어 새롭게 표본에 진입한 첫 웨이브에는 가구가중치의 평균을 부여하고, 다음 웨이브부터 무응답분석에 포함하여 다른 원시표본가구와 같이 취급함.

→ 최초 웨이브에서 응답하지 않았지만 이후 웨이브부터 응답하였다면 응답을 개시한 웨이브에서는 그 가구의 가중치를 부여하고, 이후 웨이브부터 무응답분석에 포함함.

→ 웨이브에 포함되었다가 다시 탈락하고 다시 진입하는 원시표본가구원은 가장 최근의 가구가중치를 부여받음.



⑤ 3차 및 4차 웨이브에서도 동일한 방법을 이용함.

□ 개인차원의 횡단면분석이나 종단면분석의 경우 개인가중치를 사용하거나 가구 가중치를 사용할 수 있음.

○ 먼저 무응답조정된 개인가중치를 사용할 수 있음. 다만 이 경우에는 표본수가 적어진다(비표본가구원의 가중치가 0이므로 실제로 이들의 데이터를 사용하지 않음)는 단점이 있으나, 사용하는 가정이 상대적으로 약하다는 장점이 있음

※ 실제로 PSID에서 이러한 방법을 사용하는 경우 약 20%의 표본이 제외되게 됨 (Kalton(1989)).

○ 가구가중치를 사용하는 경우 모든 가구구성원(표본가구원+비표본가구원)에 대해 동일한 가구가중치를 사용하는 방법임. 이는 비표본가구원의 추출확률이 그가 속한 표본가구의 가구원과 같다는 가정을 사용하는 경우라고 볼 수 있으며, 표본수가 최대가 된다는 장점이 있음

<표 IV-1> Duncan의 가중치와 사용방법

구분		가중치	참고
개인차원	종단면분석	개인가중치	표본수 감소(비표본가구원 제외됨). 사용하는 가정이 약함.
		가구가중치	표본수 최대(비표본가구원 포함). 비표본가구원의 추출확률이 표본가구원과 동일하다고 가정
	횡단면분석	개인가중치	표본수 감소(비표본가구원 제외됨). 사용하는 가정이 약함.
		가구가중치	표본수 최대(비표본가구원 포함). 비표본가구원의 추출확률이 표본가구원과 동일하다고 가정
가구차원	종단면분석	가구가중치	종단면가구가 동일한 가구가 아닌 경우가 있기 때문에 동 개념이 정의되지 않는다는 주장이 있음.
	횡단면분석	가구가중치	

## IV. 2차 웨이브 이후부터의 가중치 부여 과정

### 1. 2차년도에 가중치 부여 과정

#### 1.1 기초 작업

(응답여부) tab at12

at12	Freq.	Percent	Cum.
0	2,084	15.64	15.64
1	11,237	84.36	100.00
Total	13,321	100.00	

(성별) tab at12 plsex

at12	남성	여성	Total
무응답 0	1,121	963	2,084
응답 1	5,348	5,889	11,237
Total	6,469	6,852	13,321

(연령별). tab at12, sum(plage)

at12	Mean	Std. Dev	Freq.
무응답 0	35.616603	15.364302	2,084
응답 1	39.878615	16.233589	11,237
Total	39.211846	16.174411	13,321

□ (지역별) tab plreg at12

plreg	무응답	응답	Total
서울 1	609	2,946	3,555
부산 2	176	1,146	1,322
대구 3	179	699	878
대전 4	78	346	424
인천 5	126	688	814
광주 6	117	321	438
울산 7	25	300	325
경기 8	262	1,920	2,182
강원 9	45	288	333
충북 10	41	268	309
충남 11	44	303	347
전북 12	108	459	567
전남 13	50	283	333
경북 14	95	587	682
경남 15	129	683	812
전체 Total	2,084	11,237	13,321

□ (교육수준별) tab pledu at12

pledu	0	1	Total
무학 2	80	777	857
초졸 3	173	1,546	1,719
중졸 4	213	1,561	1,774
고졸 5	856	4,628	5,484
전문대졸 6	204	789	993
대졸 7	514	1,732	2,246
대학원졸 8	44	204	248
전체 Total	2,084	11,237	13,321

□ (주된 활동별) tab plact at12

plact	0	1	Total
일하였음 1	901	5,378	6,279
일시휴직 2	25	181	206
구직활동 3	182	742	924
가사 4	315	1,944	2,259
학업 5	360	1,555	1,915
연로 6	117	929	1,046
심신장애 7	47	254	301
기타 8	137	254	391
전체 Total	2,084	11,237	13,321

(가구소득) tab at12, sum(plin)

at12		Mean	Std. Dev	Freq
무응답	0	1,881.7155	1,490.2069	2084
응답	1	1,801.9637	1,536.958	11,237
Total		1,814.4404	1,529.9576	13,321

(주거형태별) tab pltyp at12

pltyp		0	1	Total
자가	1	1,197	7,080	8,277
전세	2	642	2,985	3,627
월세	3	199	835	1,034
기타	4	46	337	383
Total		2,084	11,237	13,321

기초통계량

Variable	Obs	Mean	Std. Dev.	Min	Max
at12	13,321	.8435553	.3632901	0	1
pled1	13,321	.0643345	.2453571	0	1
pled2	13,321	.1290444	.3352616	0	1
pled3	13,321	.1331732	.3397746	0	1
pled4	13,321	.4116808	.4921564	0	1
pled5	13,321	.074544	.2626639	0	1
pled6	13,321	.168606	.3744176	0	1
pled7	13,321	.0186172	.1351739	0	1
plreg1	13,321	.2668719	.4423414	0	1
plreg2	13,321	.0992418	.2989976	0	1
plreg3	13,321	.065911	.2481357	0	1
plreg4	13,321	.0318294	.1755524	0	1
plreg5	13,321	.0611065	.2395346	0	1
plreg6	13,321	.0328804	.1783303	0	1
plreg7	13,321	.0243976	.1542858	0	1
plreg8	13,321	.1638015	.3701093	0	1
plreg9	13,321	.0249981	.1561251	0	1
plreg10	13,321	.0231965	.1505327	0	1
plreg11	13,321	.0260491	.1592873	0	1
plreg12	13,321	.0425644	.2018804	0	1
plreg13	13,321	.0249981	.1561251	0	1
plreg14	13,321	.0511974	.2204083	0	1
plreg15	13,321	.0609564	.2392593	0	1

Variable	Obs	Mean	Std. Dev.	Min	Max
p1num	13,321	3.949478	1.298619	1	9
p1hty1	13,321	.6213497	.485069	0	1
p1hty2	13,321	.2722769	.4451483	0	1
p1hty3	13,321	.0776218	.2675856	0	1
p1hty4	13,321	.0287516	.1671138	0	1
plin	13,321	1814.44	1529.958	0	32,000
p1sex1	13,321	.4856242	.4998121	0	1
p1sex2	13,321	.5143758	.4998121	0	1
plage	13,321	39.21185	16.17441	15	99
plact1	13,321	.471361	.4991979	0	1
plact2	13,321	.0154643	.1233949	0	1
plact3	13,321	.0693642	.2540819	0	1
plact4	13,321	.1695819	.3752791	0	1
plact5	13,321	.143758	.3508573	0	1
plact6	13,321	.0785226	.2690023	0	1
plact7	13,321	.0225959	.1486169	0	1
plact8	13,321	.0293522	.1687979	0	1

□ 추정결과

Logit estimates

Number of obs = 13,321  
 LR chi2(34) = 467.39  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0404

Log likelihood = -5543.9825

at12	Coef.	Std. Err.	z	P> z
plreg1	-.0751115	.1092268	-0.688	0.492
plreg2	.2352426	.1285919	1.829	0.067
plreg3	-.3525733	.1304384	-2.703	0.007
plreg4	-.2187869	.161054	-1.358	0.174
plreg5	-.0591624	.1391465	-0.425	0.671
plreg6	-.6843047	.1481843	-4.618	0.000
plreg7	.7750212	.2323173	3.336	0.001
plreg8	.289619	.119199	2.430	0.015
plreg9	.0726021	.1902512	0.382	0.703
plreg10	.1372291	.1969532	0.697	0.486
plreg11	.0411606	.1916865	0.215	0.830
plreg12	-.2917889	.1472157	-1.982	0.047
plreg13	-.0206166	.1847298	-0.112	0.911
plreg14	-.0461611	.149859	-0.308	0.758
plin	-.0000163	.0000164	-0.993	0.321
p1sex1	-.1829423	.0570014	-3.209	0.001
plage	.0138753	.0027527	5.041	0.000

at12	Coef.	Std. Err.	z	P> z
plhty2	-.2067025	.0574349	-3.599	0.000
plhty3	-.4159607	.0895556	-4.645	0.000
plhty4	.2240783	.165327	1.355	0.175
plnum	-.0112462	.0203729	-0.552	0.581
pledu2	.0270768	.1507507	0.180	0.857
pledu3	-.0425919	.1571945	-0.271	0.786
pledu4	-.2299249	.1540815	-1.492	0.136
pledu5	-.5012085	.1730641	-2.896	0.004
pledu6	-.672494	.1593684	-4.220	0.000
pledu7	-.4472544	.227395	-1.967	0.049
plact1	1.03383	.1171925	8.822	0.000
plact2	1.149133	.2433294	4.723	0.000
plact3	.6857321	.1385936	4.948	0.000
plact4	.8743574	.1346398	6.494	0.000
plact5	1.073585	.1287939	8.336	0.000
plact6	.5709078	.1746062	3.270	0.001
plact7	.5022762	.2002016	2.509	0.012
_cons	.8148557	.2538116	3.210	0.001

□ 응답확률의 추정

응답확률은 평균 0.8436으로 나타났으며, 최소값 0.3290, 최대값 0.9709로 나타났음.

Variable	Obs	Mean	Std. Dev.	Min	Max
hatp12_f	1,3321	.8435553	.0704685	.3290167	.970892

□ 이렇게 계산된 응답확률의 역수를 w1\_p에 곱하여 w2\_p를 구하였음.

- w2\_p의 평균은 2,674.707, 최소값 1,605.001, 최대값 8,095.803으로 나타났음.

## 1.2 가중치 부여 세부과정

□ 1, 2차 웨이브에 동시에 응답한 개인들 11,236명은 1차 웨이브 가중치에 응답확률의 역수를 곱하여 2차 웨이브 개인가중치로 부여하였음. 이렇게 계산된 가중치가 무응답조정 종단면 개인가중치이며, w2\_p로 명명됨.

□ 다음으로 2차 웨이브에 속한 모든 개인들의 가구가중치를 구함(즉 klips02p에 속한 11,236명과 klips02n에 속한 803명의 가구가중치를 구함). 다음은 케이스별로 가중치를 구하는 방법임.

- hhid99가 동일한 가구에서 원시표본가구의 1차 및 2차웨이브에 동시 응답한 개인들의 개인가중치(w2\_p)의 합을 구하고, 이를 가구원수로 나누어서 가구가중치 w2\_h를 구함. 이 때 분모의 가구원수에는 동일한 hhid99에 속하고 w2\_p가 존재하는 원시표본가구의 수와 동일한 hhid99에 속한 비표본가구의 수의 합을 의미함. 이 때 동일한 hhid99에 소속되어 있는 가구원일지라도 15세가 되어 최초로 응답가능연령이 되어 응답한 개인과 지난 웨이브에 응답하지 않고 이번 웨이브에 응답한 원시표본가구는 제외함.

- 이와 같은 방법으로 계산된 가구가중치를 동일한 가구에 속한 원시표본가구원, 비표본가구원, 15세 이상이 되어 새롭게 응답한 원시표본가구원, 지난 웨이브에 응답하지 않았지만 이번 웨이브에 응답한 원시표본가구원에게 동일하게 부여함(w2\_h).

- 다음으로 가구가중치(w2\_h)를 15세 이상이 되어 새롭게 응답한 원시표본가구원, 지난 웨이브에 응답하지 않았지만 이번 웨이브에 응답한 원시표본가구의 개인가중치(w2\_p)로 부여함.

## 2. 3차년도 가중치 부여 과정

### 2.1 기초작업

□ (응답여부)

at23	Freq.	Percent	Cum.
0	1,630	13.92	13.92
1	10,080	86.08	100.00
Total	11,710	100.00	

(성별) tab at23 p2sex

at23	1	2	Total
0	833	797	1630
1	4,797	5,283	10,080
Total	5,630	6,080	11,710

(연령별) tab at23, sum(p2age)

at23	Mean	Std. Dev.	Freq.
0	36.477301	16.057881	1,630
1	40.652183	16.553636	10,080
Total	40.07105	16.548076	11,710

(거주지역별) tab p2reg at23

p2reg	0	1	Total
1	488	2,583	3,071
2	92	1,087	1,179
3	140	595	735
4	38	319	357
5	73	651	724
6	64	269	333
7	14	289	303
8	343	1,705	2,048
9	38	254	292
10	42	227	269
11	35	285	320
12	96	374	470
13	37	261	298
14	78	529	607
15	52	652	704
Total	1,630	10,080	11,710



□ (교육수준별) tab p2edu at23

p2edu	0	1	Total
2	69	714	783
3	139	1,413	1,552
4	153	1,422	1,575
5	648	4,131	4,779
6	145	688	833
7	435	1,530	1,965
8	41	182	223
Total	1,630	10,080	11,710

□ (주된 활동별). tab p2act at23

p2act	0	1	Total
1	698	5,160	5,858
2	25	155	180
3	60	276	336
4	276	1,644	1,920
5	318	1,347	1,665
6	98	905	1,003
7	44	288	332
8	111	305	416
Total	1,630	10,080	11,710

□ (소득수준별) tab at23, sum(p2in)

at23	Mean	Std. Dev.	Freq.
0	1,961.6301	1,560.9432	1,630
1	1,843.1438	1,407.7001	10,080
Total	1,859.6368	1,430.5325	11,710

※ 2차 웨이브 응답률 계산모형에서 사용하였던 가구거주형태변수는 설명변수에서 제외함. 2차 웨이브 가구용설문서에서 원시표본가구원이 비표본가구원과 합쳐서 새롭게 가구를 구성한 경우에는, 거주형태를 질문하는 v02601에서 모두 missing 처리되었음. 여기에 해당하는 가구의 가구거주형태를 알 수 없기 때문에 설명변수에서 제외하였음.

□ (기초통계량)

Variable	Obs	Mean	Std. Dev.	Min	Max
at23	11,710	.8608027	.346167	0	1
p2reg1	11,710	.2622545	.4398791	0	1
p2reg2	11,710	.1006832	.3009216	0	1
p2reg3	11,710	.0627669	.2425535	0	1
p2reg4	11,710	.0304868	.1719298	0	1
p2reg5	11,710	.0618275	.2408523	0	1
p2reg6	11,710	.0284372	.1662255	0	1
p2reg7	11,710	.0258753	.1587701	0	1
p2reg8	11,710	.1748933	.379892	0	1
p2reg9	11,710	.024936	.1559366	0	1
p2reg10	11,710	.0229718	.14982	0	1
p2reg11	11,710	.0273271	.1630416	0	1
p2reg12	11,710	.0401366	.196288	0	1
p2reg13	11,710	.0254483	.1574892	0	1
p2reg14	11,710	.051836	.2217053	0	1
p2reg15	11,710	.0601196	.2377184	0	1
p2in	11,710	1859.637	1430.533	0	26,400
p2sex1	11,710	.4807857	.499652	0	1
p2sex2	11,710	.5192143	.499652	0	1
p2age	11,710	40.07105	16.54808	14	97
p2edu1	11,710	.0668659	.2498003	0	1
p2edu2	11,710	.1325363	.3390874	0	1
p2edu3	11,710	.1345004	.3412038	0	1
p2edu4	11,710	.4081127	.4915052	0	1
p2edu5	11,710	.0711358	.2570625	0	1
p2edu6	11,710	.1678053	.3737093	0	1
p2edu7	11,710	.0190436	.1366839	0	1
p2act1	11,710	.5002562	.5000213	0	1
p2act2	11,710	.0153715	.1230304	0	1
p2act3	11,710	.0286934	.1669506	0	1
p2act4	11,710	.1639624	.3702573	0	1
p2act5	11,710	.1421862	.3492559	0	1
p2act6	11,710	.0856533	.2798633	0	1
p2act7	11,710	.0283518	.165983	0	1
p2act8	11,710	.0355252	.185111	0	1

□ 추정결과

Logit estimates

Number of obs = 11,710  
 LR chi2(30) = 439.20  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0465

Log likelihood = -4505.4271

at23	Coef.	Std. Err.	z	P> z
p2reg2	.7555144	.120889	6.250	0.000
p2reg3	-.2713421	.108933	-2.491	0.013
p2reg4	.4802317	.1804977	2.661	0.008
p2reg5	.4678505	.1345795	3.476	0.001
p2reg6	-.1973557	.150735	-1.309	0.190
p2reg7	1.324967	.279842	4.735	0.000
p2reg8	-.1133371	.0784503	-1.445	0.149
p2reg9	.1233756	.1837183	0.672	0.502
p2reg10	-.0692694	.1776598	-0.390	0.697
p2reg11	.2182979	.1897109	1.151	0.250
p2reg12	-.3969066	.1276048	-3.110	0.002
p2reg13	.201167	.1852459	1.086	0.278
p2reg14	.0724897	.1341918	0.540	0.589
p2reg15	.824171	.1539833	5.352	0.000
p2edu2	.0745088	.1620043	0.460	0.646
p2edu3	.0404643	.1712076	0.236	0.813
p2edu4	-.205819	.1663668	-1.237	0.216
p2edu5	-.4924588	.1900531	-2.591	0.010
p2edu6	-.7420155	.1721524	-4.310	0.000
p2edu7	-.6147086	.2416141	-2.544	0.011
p2act2	-.199657	.2233094	-0.894	0.371
p2act3	-.4423512	.1518493	-2.913	0.004
p2act4	-.3843835	.0875579	-4.390	0.000
p2act5	-.3143393	.092475	-3.399	0.001
p2act6	-.2705142	.141603	-1.910	0.056
p2act7	-.4208517	.1727977	-2.436	0.015
p2act8	-.872604	.1222059	-7.140	0.000
p2sex1	-.1345918	.063683	-2.113	0.035
p2age	.0076098	.0029705	2.562	0.010
p2in	-3.20e-06	.0000197	-0.162	0.871
_cons	1.97419	.2312833	8.536	0.000

□ 응답확률 예측치의 기초 통계량 : 2차 웨이브와 3차 웨이브에 모두 응답한 원시 표본가구원 10,253명의 응답확률이 계산됨.

Variable	Obs	Mean	Std. Dev.	Min	Max
hatp23	11710	.8608027	.0683708	.5113549	.9794798
Summary of Pr(at23)					
at23	Mean		Std. Dev.	Freq.	
0	.82726831		.07735031	1630	
1	.86622546		.06520946	10080	
Total	.86080273		.06837082	11710	

※ 가구주와의 관계변수는 거의 유의하지 않았으며, 포함여부가 별다른 영향을 미치지 못했음.

※ 모형설정을 위하여 다양한 형태(interaction terms 포함)를 시도하여 보았으나, 모형별로 거의 차이가 없었음. 이에 따라 2차년도의 응답모형과 유사하게 3차년도 응답모형을 설정하였음.

## 2.2 가중치 부여 세부과정

### □ 세부과정

#### ○ 원시표본가구원

- 1,2,3차에 모두 응답한 경우 : 응답확률의 역수를 곱하여 개인가중치산출
- 1차 응답, 2차 무응답, 3차 응답한 경우 : 1차년도 개인가중치를 3차년도 개인가중치로 부여함.
- 3차에 처음으로 응답 : 평균으로 구한 가구가중치를 개인가중치로 부여함.

#### ○ 비표본가구원

- 평균으로 구한 가구가중치만을 부여함.

### 3. 4차년도 가중치 부여 과정

#### 3.1 기초작업

(응답여부). tab at34

at34	Freq.	Percent	Cum.
0	1,424	13.19	13.19
1	9,374	86.81	100.00
Total	10,798	100.00	

(성별)tab p3sex at34

p3sex	0	1	Total
1	737	4,452	5,189
2	687	4,922	5,609
Total	1,424	9,374	10,798

(연령별) tab at34, sum(p3age)

at34	Mean	Std. Dev.	Freq.
0	38.421348	17.409563	1,424
1	41.231598	16.62343	9,374
Total	40.860993	16.755393	10,798

□ (지역별) tab p3reg at34

p3reg	0	1	Total
1	453	2,292	2,745
2	103	1,054	1,157
3	122	510	632
4	37	320	357
5	118	585	703
6	24	247	271
7	25	290	315
8	272	1,582	1,854
9	49	223	272
10	15	228	243
11	29	279	308
12	36	362	398
13	14	258	272
14	69	478	547
15	58	666	724
Total	1,424	9,374	10,798

□ (교육수준별) tab p3edu at34

p3edu	0	1	Total
2	95	631	726
3	120	1,344	1,464
4	143	1,329	1,472
5	587	3,799	4,386
6	143	657	800
7	301	1,447	1,748
8	35	167	202
Total	1,424	9,374	10,798

□ (주된 활동별). tab p3act at34

p3act	0	1	Total
1	614	4,464	5,078
2	53	424	477
3	11	63	74
4	2	16	18
5	211	1,491	1,702
6	18	155	173
7	207	1,204	1,411
8	308	1,557	1,865
Total	1,424	9,374	10,798

□ (소득수준별) tab at34, sum(p3in)

at34	Mean	Std. Dev.	Freq.
0	2,077.2233	1,814.3871	1,424
1	2,024.803	1,547.9885	9,374
Total	2,031.716	1,585.6907	10,798

□ 변수들의 기초통계량

Variable	Obs	Mean	Std. Dev.	Min	Max
at34	10,798	.8681237	.3383719	0	1
p3age	10,798	40.86099	16.75539	14	98
p3in	10,798	2031.716	1585.691	0	27,612
p3reg1	10,798	.2542137	.4354385	0	1
p3reg2	10,798	.1071495	.3093175	0	1
p3reg3	10,798	.0585294	.2347526	0	1
p3reg4	10,798	.0330617	.1788059	0	1
p3reg5	10,798	.0651046	.2467218	0	1
p3reg6	10,798	.0250972	.1564277	0	1
p3reg7	10,798	.0291721	.1682964	0	1
p3reg8	10,798	.1716985	.3771356	0	1
p3reg9	10,798	.0251898	.1567086	0	1
p3reg10	10,798	.0225042	.1483232	0	1
p3reg11	10,798	.0285238	.1664715	0	1
p3reg12	10,798	.0368587	.1884235	0	1
p3reg13	10,798	.0251898	.1567086	0	1
p3reg14	10,798	.0506575	.2193075	0	1
p3reg15	10,798	.0670495	.2501192	0	1
p3sex1	10,798	.480552	.4996448	0	1
p3sex2	10,798	.519448	.4996448	0	1
p3edu1	10,798	.0672347	.2504396	0	1
p3edu2	10,798	.1355807	.3423586	0	1
p3edu3	10,798	.1363215	.3431456	0	1
p3edu4	10,798	.4061863	.4911429	0	1
p3edu5	10,798	.0740878	.2619258	0	1
p3edu6	10,798	.1618818	.3683594	0	1
p3edu7	10,798	.0187072	.1354951	0	1
p3act1	10,798	.4702723	.4991386	0	1
p3act2	10,798	.0441748	.2054929	0	1
p3act3	10,798	.0068531	.0825032	0	1
p3act4	10,798	.001667	.0407965	0	1
p3act5	10,798	.1576218	.3644029	0	1
p3act6	10,798	.0160215	.1255638	0	1
p3act7	10,798	.1306723	.3370573	0	1
p3act8	10,798	.1727172	.3780201	0	1



□ 응답확률 추정결과

Logit estimates

Number of obs = 10,798  
 LR chi2(29) = 271.05  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.0322  
 Log likelihood = -4075.0246

at34	Coef.	Std. Err.	z	P> z
p3reg3	-.388481	.1125457	-3.452	0.001
p3reg4	.3821588	.1806629	2.115	0.034
p3reg5	-.2105955	.1121059	-1.879	0.060
p3reg6	.539148	.2202104	2.448	0.014
p3reg7	.6591513	.2150004	3.066	0.002
p3reg8	-.0398637	.0808417	-0.493	0.622
p3reg9	-.3010143	.1666073	-1.807	0.071
p3reg10	.9389871	.2720979	3.451	0.001
p3reg11	.3737582	.2033181	1.838	0.066
p3reg12	.4508234	.1823812	2.472	0.013
p3reg13	1.077775	.2795816	3.855	0.000
p3reg14	.048508	.1394122	0.348	0.728
p3reg15	.6879054	.1456601	4.723	0.000
p3edu2	.6091004	.1536505	3.964	0.000
p3edu3	.5112526	.1609018	3.177	0.001
p3edu4	.2313073	.1557207	1.485	0.137
p3edu5	-.0728197	.1828659	-0.398	0.690
p3edu6	-.0612858	.1648426	-0.372	0.710
p3edu7	-.1213081	.2421716	-0.501	0.616
p3act2	-.1590703	.1604749	-0.991	0.322
p3act3	.1188931	.3364456	0.353	0.724
p3act4	.038213	.7619356	0.050	0.960
p3act5	-.2447953	.0984699	-2.486	0.013
p3act6	.1594928	.2594651	0.615	0.539
p3act7	.1134601	.1007045	1.127	0.260
p3act8	-.5683684	.0829481	-6.852	0.000
p3sex1	-.1622709	.0693815	-2.339	0.019
p3age	.0125258	.0027769	4.511	0.000
p3in	-5.16e-06	.0000191	-0.270	0.787
_cons	1.326785	.2273871	5.835	0.000

. sum hatp34

Variable	Obs	Mean	Std. Dev.	Min	Max
hatp34	10,798	.8681237	.0537263	.5896763	.9799005

  

at34	Mean	Std. Dev.	Freq.
0	.84684991	.05528721	1,424
1	.87135542	.05274266	9,374
Total	.86812373	.05372631	10,798

#### 4. 각 웨이브에서의 가중치의 기초통계량

1차 웨이브

Variable	Obs	Mean	Std. Dev.	Min	Max
w1_p	13,321	2,255.039	416.7548	1,513.22	4,514.563
w1_h	13,321	2,255.039	416.7548	1,513.22	4,514.563

2차 웨이브

Variable	Obs	Mean	Std. Dev.	Min	Max
pid	12,039	252,418.2	145127.8	101	513,101
w1_h	11,237	2,256.868	413.7408	1,513.22	4,514.563
w1_p	11,237	2,256.868	413.7408	1,513.22	4,514.563
w2_p	11,708	2,671.798	546.3235	611.9293	8,095.803
w2_h	12,032	2,600.247	588.4898	492.2492	6,160.998

※ w2\_h=0인 7명 제외

□ 3차 웨이브

Variable	Obs	Mean	Std. Dev.	Min	Max
pid	11,205	254,092.7	147,331.3	101	534,801
w1_h	10,141	2,264.563	417.4775	1,513.22	4,514.563
w1_p	10,141	2,264.563	417.4775	1,513.22	4,514.563
w2_p	10,080	2,667.529	544.3491	611.9293	8,095.803
w2_h	10,324	2,604.68	584.397	0	6,160.998
w3_p	10,798	3,072.339	758.9452	676.7079	10,699.03
w3_h	11,194	2,963.651	771.8895	482.1045	8,123.19

※ w3\_h가 0인 11명 제외

□ 4차 웨이브

Variable	Obs	Mean	Std. Dev.	Min	Max
pid	11,051	259,720.6	149,571.5	201	558,202
w1_h	8,957	2,264.824	418.4443	1,513.22	4,514.563
w2_p	9,317	2,667.034	544.2263	611.9293	8,095.803
w2_h	9,521	2,608.579	583.2122	0	6,160.998
w3_p	9,374	3,062.756	750.1324	676.7079	10,699.03
w3_h	9,674	2,968.351	764.0813	0	8,123.1
w4_p	10,499	3,450.922	980.4998	738.9341	12,094.37
w4_h	11,021	3,287.472	969.8315	522.0335	9,287.783

※ w4\_h가 0인 30명 제외

## 참 고 문 헌

- 강석훈 (1999), KLIPS 1차 웨이브의 가중치 부여방안에 대한 연구,
- Beckett, S. , W. Gould, L. Lillard, & F. Welch (1988), "The PSID after Fourteen Years : An Evaluation", *Journal of Labor Economics*, Vol 6:4, pp. 472-492
- Chapman, D., L. Baily, and D. Kasprzyk, (1986), "Nonresponse Adjustment Procedures at the U.S. Bureau of the Census", *Survey Methodology*, 12, pp. 161-180
- Duncan, G. (1995), "A Simple Method for Weighting in Household Panel Surveys", Working Paper, Northwestern University
- Duncan, G., J. Juster & T. Morgan (1984), "The Role of Panel Studies in a World of Scare Research Resources", In S. Sudman & M. Spath, ed. *The Collection and Analysis of Economic and Consumer Behavior data : In Memory of Robert Ferber*, Urbana : University of Illinois
- Earnst, L. (1989), "Weighting Issues for Longitudinal and Family Estimates", In Kasprzyk, D. , G. Duncan, M. Sing eds, *Panel Surveys*, Wiley, pp. 139-117
- Huang, H. (1984), "Obtaining Cross-sectional Estimates from a Longitudinal Survey : Experiences of the Income Survey Development program", *Proceedings of the section on Survey Research methods*, American Statistical Association, 670-675
- ISR(1972), *A panel Study of Income Dynamics*, University of Michigan
- Kalton, G. (1989), "Modeling Consideration from a Survey Sampling Prospective", In Kasprzyk, D., G. Duncan, M. Sing eds, *Panel Surveys*, Wiley, pp. 139-117
- Kalton, G. and M. Brick (1994), "Weighting Schemes for Household Panel Surveys", *Proceedings of the Section on Survey research Methods*, American Statistical Association, 785-790
- Kalton, G., J. Brick & Westat, Inc., (1994?), "Weighting Schemes for Household Panel Surveys", SIPP No. 199, Bureau of the Census

- Mack, S. and R. Petroni (1994), "Overview of SIPP nonresponse Research", Paper presented at the Fifth International Workshop on Household Survey Non-response
- Nelson, D. , C. Bowie, and A. Walker (1987), "SIPP Sample Loss and the Efforts to Reduce It", SIPP Working Paper, No. 8707, U.S. Bureau of Census
- Rendtal, U. (1992), "Weighting Procedures and Sampling Variance in Household Panels", Working Paper, German Institute for Economic Research
- Singh, R. and R. Petroni (1988), Nonresponse Adjustment Methods for Demographic Surveys at the U.S. Bureau of the Census, SIPP working Paper No. 8823, U.S. Bureau of the Census