

# 『한국노동패널 가중치 연구』

2013. 12

- 책임 연구원 : 박민규(고려대학교 통계학과 교수)  
공동 연구원 : 김영원(숙명여자대학교 통계학과 교수)  
                  변종석(한신대학교 정보통계학과 교수)  
연구 보조원 : 송종호(고려대학교 통계학과 박사 수료)  
                  임찬수(성균관대학교 통계학과 박사수료)  
                  김사라(고려대학교 통계학과 석사 과정)



## <목 차>

|                                 |    |
|---------------------------------|----|
| I. 연구개요 .....                   | 1  |
| 1. 연구목적 .....                   | 1  |
| 2. 연구내용 및 방법 .....              | 2  |
| II. 원패널 가중치 산출 .....            | 4  |
| 1. 로지스틱 회귀분석을 이용한 응답모형 설정 ..... | 4  |
| 2. 원패널 가중치 산출과정 .....           | 15 |
| 3. 원패널 가중치 부여 결과 .....          | 21 |
| III. 통합패널 가중치 산출 .....          | 27 |
| 1. 통합패널 구축 현황 .....             | 27 |
| 2. 통합패널 1차 가중치 산출과정 .....       | 28 |
| 3. 로지스틱 회귀모형을 이용한 응답모형 설정 ..... | 33 |
| 4. 통합패널 가중치 산출과정 .....          | 39 |
| 5. 통합패널 차수별 가중치 산출 결과 .....     | 45 |
| IV. 노동패널 가중치 사용 방안 .....        | 47 |
| V. 참고문헌 .....                   | 49 |

## <표 목 차>

|   |    |
|---|----|
| <표 1> 원패널 2차년도 로지스틱 회귀분석 결과(가구소득포함) .....   | 5  |
| <표 2> 원패널 3차년도 로지스틱 회귀분석 결과(가구소득포함) .....   | 6  |
| <표 3> 원패널 2차년도 로지스틱 회귀분석 결과 .....           | 6  |
| <표 4> 원패널 3차년도 로지스틱 회귀분석 결과 .....           | 7  |
| <표 5> 원패널 4차년도 로지스틱 회귀분석 결과 .....           | 7  |
| <표 6> 원패널 5차년도 로지스틱 회귀분석 결과 .....           | 8  |
| <표 7> 원패널 6차년도 로지스틱 회귀분석 결과 .....           | 8  |
| <표 8> 원패널 7차년도 로지스틱 회귀분석 결과 .....           | 9  |
| <표 9> 원패널 8차년도 로지스틱 회귀분석 결과 .....           | 9  |
| <표 10> 원패널 9차년도 로지스틱 회귀분석 결과 .....          | 10 |
| <표 11> 원패널 10차년도 로지스틱 회귀분석 결과 .....         | 10 |
| <표 12> 원패널 11차년도 로지스틱 회귀분석 결과 .....         | 11 |
| <표 13> 원패널 12차년도 로지스틱 회귀분석 결과 .....         | 11 |
| <표 14> 원패널 13차년도 로지스틱 회귀분석 결과 .....         | 12 |
| <표 15> 원패널 14차년도 로지스틱 회귀분석 결과 .....         | 12 |
| <표 16> 원패널 15차년도 로지스틱 회귀분석 결과 .....         | 13 |
| <표 17> 원패널 15차 성별에 따른 응답 여부 .....           | 13 |
| <표 18> 원패널 15차 학력에 따른 응답 여부 .....           | 13 |
| <표 19> 원패널 15차 지역에 따른 응답 여부 .....           | 14 |
| <표 20> 원패널 15차 주된활동에 따른 응답 여부 .....         | 14 |
| <표 20> 원패널 15차 주된활동에 따른 응답 여부 .....         | 14 |
| <표 21> 신규진입가구원만으로 구성된 가구 정보(비원가구원 제외) ..... | 18 |
| <표 22> 추계 인구·가구 증가율 .....                   | 20 |
| <표 22> 1차 원패널 가중치의 기초통계량 .....              | 21 |
| <표 23> 2차 원패널 가중치의 기초통계량 .....              | 21 |
| <표 24> 3차 원패널 가중치의 기초통계량 .....              | 22 |
| <표 25> 4차 원패널 가중치의 기초통계량 .....              | 22 |
| <표 26> 5차 원패널 가중치의 기초통계량 .....              | 22 |
| <표 27> 6차 원패널 가중치의 기초통계량 .....              | 23 |
| <표 28> 7차 원패널 가중치의 기초통계량 .....              | 23 |
| <표 29> 8차 원패널 가중치의 기초통계량 .....              | 23 |
| <표 30> 9차 원패널 가중치의 기초통계량 .....              | 24 |
| <표 31> 10차 원패널 가중치의 기초통계량 .....             | 24 |
| <표 32> 11차 원패널 가중치의 기초통계량 .....             | 24 |
| <표 33> 12차 원패널 가중치의 기초통계량 .....             | 25 |
| <표 34> 13차 원패널 가중치의 기초통계량 .....             | 25 |
| <표 35> 14차 원패널 가중치의 기초통계량 .....             | 25 |
| <표 36> 15차 원패널 가중치의 기초통계량 .....             | 26 |

|  |    |
|--|----|
| <표 38> 시도별/동읍면부별 추가패널 구축 현황 .....                | 28 |
| <표 39> 통합패널 구축시 응답 가구여부 .....                    | 29 |
| <표 40> 통합 패널 가구내 원가구원수 분포 .....                  | 29 |
| <표 41> 통합 1차의 사후조정을 위한 추계인구 및 추계가구 .....         | 31 |
| <표 42> 통합패널 차수별 응답자의 성별 .....                    | 34 |
| <표 43> 통합패널 차수별 응답자의 거주지역 .....                  | 34 |
| <표 44> 통합패널 차수별 응답자의 학력 .....                    | 34 |
| <표 45> 통합패널 차수별 응답자의 주된활동 .....                  | 35 |
| <표 46> 통합패널 차수별 응답자의 만나이 .....                   | 35 |
| <표 47> 통합패널 2차년도 로지스틱 회귀분석 결과 .....              | 38 |
| <표 48> 통합패널 3차년도 로지스틱 회귀분석 결과 .....              | 38 |
| <표 49> 통합패널 4차년도 로지스틱 회귀분석 결과 .....              | 38 |
| <표 50> 통합 2~4차의 사후조정을 위한 인구주택총조사 인구 및 추계인구 ..... | 42 |
| <표 51> 통합 2~4차의 사후조정을 위한 인구주택총조사 가구 및 추계가구 ..... | 45 |
| <표 52> 통합패널 차수별 개인 가중치 기초통계량 .....               | 46 |
| <표 53> 통합패널 차수별 가구 가중치 기초통계량 .....               | 46 |

## <그림 목 차>

|                            |    |
|----------------------------|----|
| <그림 1> 원패널 가구원 구조 .....    | 15 |
| <그림 2> 통합패널 가중치 산출과정 ..... | 42 |
| <그림 3> 자료의 형태 .....        | 47 |

# I. 연구개요

## 1. 연구목적

2012년으로 15차를 맞는 한국노동패널조사는 사회·경제적으로 급격하게 변화하는 환경 속에서 노동, 가구 경제, 관광 정책 수요를 뒷받침 할 수 있는 중단자료 제공을 위한 조사로서 그 내용은 패널을 구성하고 있는 표본가구 내 가구 구성원의 경제활동, 노동이동, 고용, 실업, 가구 소득을 비롯한 각종 소득, 임금, 근로조건, 노사관계, 교육, 직업훈련 등 노동시장 전반에 걸친 주제들이다. 조사의 대상은 도시지역에 거주하는 5천 가구(1998년 기준)와 이들 가구로부터 분가한 분가가구 및 패널 노후화와 및 가구 대표성 보완을 위해 추가된 1,500가구(2009년) 및 가구원이다. 조사 방법은 개인면접조사(면접 타계식)로 면접원이 직접 응답자를 방문하여 해당 패널을 대상으로 설문에 대한 내용을 질문하고 그 답변을 기입해오는 방식을 이용하고 있다.

한국노동패널조사 자료의 통계적 타당성을 갖춘 중·횡단면 가구원 단위 분석과 가구 단위 분석을 위하여 1차부터 13차까지 매 년 가구원 및 가구 가중치가 부여되어 왔다. 가구원 가중치로는 종단면 가중치와 횡단면 가중치가 작성되어 왔고 가구 가중치로는 일종의 횡단면 가중치가 작성되어 왔다. 이러한 가중치는 기본적으로 최초 패널 구성 시점에 추출된 표본이 모집단을 대표할 수 있도록 부여된 가중치를 기반으로 매 해 조정되어 왔다. 매 해 패널 마모를 고려한 무응답 보정 가구원 종단면 가중치가 일차적으로 산출된 후 이를 바탕으로 가구 및 횡단면 가구원 가중치가 산출되어 왔다. 98년 구축된 원패널은 원년 이후 일관된 방법으로 그 가중치들이 생성되어 왔으나 2009년 표본 크기 및 모집단 대표성의 범위를 확보하기 위하여 추가된 패널과 원표본으로 구성된 통합 패널의 경우 그 가중치 산출에 있어 모집단과의 인구특성학적 변수 기준으로 차이가 발생하였다. 한국고용정보원(2012)은 이러한 통합 패널 가중치의 문제점을 파악하고 이의 수정을 위해 통합패널 가중치 안정화 방안에 대한 연구를 실시하였고 이를 바탕으로 모집단의 대표성을 확보하는 중·횡단면 가구원 가중치 및 가구 가중치 산출 방안을 제시하였다.

본 연구에서는 기존의 가중치 산출방안의 큰 틀은 유지하되 모집단의 시간에 따른 변화 및 대표성을 확보할 수 있는 가구 및 가구원의 가중치 산출 방안을 제시하고자 한다. 이는 기존의 가중치 산출 방안에 대한 자세한 기록이 부재하여 비록 산출된 가중치에 미치는 영향은 미미하나 매 년 가중치 산출 시 마다 서로 다른 조정 방안들이 사용되었을 것으로 예측되기 때문이다. 즉 본 연구의 목적은 매 해 원패널 및 통합패널 대표성을 확

보할 수 있는 중·횡단면 가구원 가중치 및 가구 가중치 산출 과정을 표준화하고 이에 맞추어 1차 년도부터 15차 년도까지의 원패널과 통합패널의 가중치를 산출하는 것에 있다.

## 2. 연구내용 및 방법

한국노동패널 자료 분석을 위한 가중치들은 총 4단계의 과정을 통해 산출된다. 제 1 단계에서는 표본마모로 인해 발생한 패널의 대표성 문제를 해결하기 위하여 무응답 보정을 실시하고 이를 바탕으로 종단면 가구원 가중치가 산출된다. 두 번째 단계에서는 1 단계에서 산출된 가구원 가중치의 가구 별 평균을 이용하여 가구 가중치가 부여되며 제 3단계에서는 산출된 가구 가중치가 가구원의 횡단면 가중치로 정의된다. 마지막 4단계에서는 가구원 및 가구 모집단의 증감을 반영하기 위한 중·횡단면 가구원 가중치와 가구 가중치의 스케일 조정이 이루어진다. 본 연구에서는 15차까지 이루어진 한국노동원 패널과 원패널과 더불어 2009년에 표본 추가를 통해 생성된 통합패널에 적용될 가중치들을 위의 4단계에 따라서 산출하는 것을 목적으로 다음과 같은 사항들을 연구한다.

### 1) 무응답 보정을 위한 로지스틱 회귀분석 모형 검토

무응답 보정을 위해서는 기본적으로 로지스틱 회귀분석을 통해 응답확률 혹은 응답 경향 점수(propensity score)를 추정하여 이를 종단면 가중치 산출에 적용하는 방안을 고려한다. 로지스틱 회귀 모형이 적합 되는 대상은 이 전년도에 응답한 원가구원이다. 무응답 보정의 통계적 타당성은 주어진 로지스틱 모형의 적합 여부에 의하여 결정된다. 즉 로지스틱 모형에서 사용된 설명변수들이 주어진 경우 응답여부는 MCAR(missing completely at random)조건을 만족해야 하며 모형 자체의 예측력은 높아야 할 것이다. 이러한 이유로 본 연구에서는 실제 이용 가능한, 즉 설명변수의 결측이 없는 그리고 지속적으로 모든 연차에 사용될 수 있는 로지스틱 회귀모형들이 검토되며 이 중 최종 모형을 제안할 것이다.

### 2) 가구원 중·횡단면 가중치 및 가구 가중치 산출 과정의 표준화

기존의 가구원 및 가구 가중치 산출과정은 언급된 4 단계에 의하여 이루어졌으나 산출과정에 대한 구체적인 설명의 부족으로 실제 가중치 산출에서 발생하는 여러 문제들은 일괄되게 처리되지 못한 부분이 있다. 예를 들어 스케일 조정의 경우 그 이론적 배경

이나 스케일 부여 과정에 대한 설명이 부족하여 매 해 연구자들의 판단에 의하여 이루어진 것으로 사료된다. 본 연구에서는 가중치 산출이 이루어지는 4단계 과정의 세부적인 설명을 통해 추 후 일관된 방법으로 가중치들이 산출될 수 있도록 가중치 산출 과정의 표준화 방안을 제시할 것이다.

### 3) 표준화된 절차에 따른 중·횡단면 가구원 가중치 및 가구 가중치 산출

표준화된 가중치 산출 절차에 따라서 1차에서 15차까지의 원패널 가중치 그리고 2009년부터 시작된 총 4개년도의 통합패널 가중치를 본 연구에서는 산출한다. 따라서 산출될 가중치 세트는 총  $15(\text{년차}) \times 3(\text{중·횡단면 가구원가중치, 가구 가중치}) + 4 \times 3 = 57$ 개이다.

### 4) 가중치 사용에 관한 제언

가중치 산출의 가장 큰 목적은 바로 그 사용에 있다. 중·횡단면 가구원 가중치 및 가구 가중치는 분석의 목적과 분석 방법에 맞추어서 적절하게 그러나 반드시 사용되는 것이 바람직하다. 그러나 분석의 목적 및 분석 방안의 모든 가능한 경우를 본 연구에서 고려할 수 없으므로 가중치 사용 방안에 대한 대략적인 가이드라인을 본 연구에서는 제시하고자 한다.



## II. 원패널 가중치 산출

가구원의 중·횡단면 가중치 및 가구 가중치 산출에 있어서 기본이 되는 원리는 원패널이 구성된 시점에서 추출된 표본이 대표하는 모집단의 변화를 그대로 반영할 수 있도록 하는 것이다. 즉 원년이후에 패널에서 누락되거나 신규로 진입하는 과정에서 각 가구 및 가구원에 부여되는 가중치가 원패널이 대표하던 모집단의 변화를 그대로 반영하도록 하는 것에 가중치 산출에 원리를 두고 있다. 예를 들어 원패널 가구에 속하지 않은 가구원이 결혼 등의 이유로 추 후 패널에 포함된 경우, 신규 유입된 비원가구원은 이미 원패널 가구원에 의하여 대표가 되도록 가중치가 산출되었기 때문에 이러한 비원가구원에 가중치를 부여할 때 추가로 신규 진입자에게 가중치가 부여되어 일종의 동일인이 중복되는 효과를 갖지 않도록 가중치가 산출되어야 한다. 이를 위하여 기본적으로 가정하는 것은 사망, 해외 이주 등과 같은 가구원 및 가구 소멸 및 출생(본 패널에서는 원가구 내에서 만 15세가 되는 가구원의 유입) 등의 모집단 내 변화는 모집단과 동일한 패턴으로 표본에서도 이루어진다는 것이다. 또한 이민 등을 통한 전혀 새로운 가구 및 가구원의 유입되어 발생하는 모집단의 변화는 패널 연구의 대상이 되지 않는다는 것이다. 즉 구성된 패널을 통하여 산출되는 결과는 원패널 구성 이후 모집단 외부로부터 신규 유입되는 가구 및 가구원은 대표할 수 없다는 것이다. 이러한 가정들 하에서 가구원 및 가구 가중치의 산출을 위한 내용들을 본 장에서는 살펴본다.

### 1. 로지스틱 회귀모형을 이용한 응답모형 설정

$t$ 차 년도에 가구원  $i$ 의 응답여부를 나타내는 변수  $R_{ti}$ 을 응답과 무응답의 가능한 두 범주를 갖는 이항 확률 변수로 정의하고  $p_{ti} = P[R_{ti} = 1(\text{응답})]$ 의 예측을 위하여 아래의 로지스틱 회귀모형을 고려한다.

$$R_{ti} \sim \text{Bernoulli}(p_{ti}), \quad (1)$$

$$\log\left(\frac{p_{ti}}{1-p_{ti}} \mid \mathbf{x}_{t-1}\right) = \alpha_t + \beta_t x_{t-1,1i} + \cdots + \beta_t x_{t-1,ki}$$

위의 로지스틱 회귀모형의 정의에서  $t$ 시점의 가구원의 응답확률은 전년도의 자료인  $\mathbf{x}_{t-1} = (x_{t-1,1}, \cdots, x_{t-1,k})'$ 을 이용하여 예측된다. 이는  $t$ 시점에 종단면 가중치가 원칙적으로 부여되는 대상은  $t$ 시점까지 모두 조사가 된 가구원이며 따라서 보정을 위해 추정되어야 하는 응답확률은 아래와 같기 때문이다.

$$\begin{aligned}
& P\{(R_{1i}=1)\cap(R_{2i}=1)\cap\cdots\cap(R_{ti}=1)\} \tag{2} \\
& = P(R_{1i}=1)\times P(R_{2i}=1 \mid R_{1i}=1)\times\cdots\times P(R_{ti}=1 \mid R_{1i}=1,\cdots,R_{t-1,i}=1) \\
& = P(R_{1i}=1)\times P(R_{2i}=1 \mid R_{1i}=1)\times\cdots\times P(R_{ti}=1 \mid R_{t-1,i}=1)
\end{aligned}$$

식 (2)의 각 조건부 응답확률은 식 (1)의 로지스틱 회귀모형을 적합하여 예측된다. 로지스틱 회귀모형의 통계적 타당성 확보를 위해서는 적절한 설명변수를 선정해야 한다. 기존의 모형에서는 성별, 지역, 학력, 주된 활동, 만 나이, 가구소득의 변수들이 로지스틱 회귀모형의 적합을 위하여 사용되었다. 그러나 가구소득의 경우 1차 년도에 382명, 2차 년도에 104명이 결측으로 나타나고 있으며, 이러한 현상은 거의 매 년 나타나고 있으며 이의 해결을 위해 다양한 대체 방안이 사용된 것으로 파악되고 있다. 응답률 예측을 위해 결측된 설명 변수의 대체가 이루어지는 경우 추가적인 변동이 발생하는 문제가 발생하며 더 심각하게는 설명 변수의 측정 오차로 인한 로지스틱 회귀분석 결과의 편향이 발생할 수 있다. 또한 기존의 설명변수들 중 다른 모든 변수가 가구원의 정보를 나타내는 반면 가구 소득만이 가구 특성을 나타내는 변수라는 것 역시 가구 소득을 설명 변수로 포함시키는 것이 타당하지 않은 것으로 파악된다.

<표 1>~<표2>는 가구소득을 설명변수로 포함 한 로지스틱 회귀모형의 적합결과의 일부를 나타내고 있다. 적합 결과를 통해 볼 때 가구소득 변수는 실제로도 각 가구원의 응답여부를 설명하는 데에 역할을 하지 않는 것으로 판단된다.

<표 1> 원패널 2차 로지스틱 회귀분석 결과(가구소득포함)

| 변수                          | 자유도 | 카이제곱       | 유의확률            |
|-----------------------------|-----|------------|-----------------|
| 성별                          | 1   | 12.144     | 0.0005          |
| 지역                          | 14  | 105.562    | <.0001          |
| 학력                          | 6   | 68.426     | <.0001          |
| 주된 활동                       | 5   | 82.969     | <.0001          |
| 만 나이                        | 1   | 52.312     | <.0001          |
| 가구소득                        | 1   | 0.439      | 0.5076          |
| likelihood $\chi^2=406.339$ |     | $n=13,321$ | $p-value<.0001$ |

<표 2> 원패널 3차 로지스틱 회귀분석 결과(가구소득포함)

| 변수                          | 자유도 | 카이제곱       | 유의확률             |
|-----------------------------|-----|------------|------------------|
| 성별                          | 1   | 6.551      | 0.0105           |
| 지역                          | 14  | 159.989    | <.0001           |
| 학력                          | 6   | 95.649     | <.0001           |
| 주된 활동                       | 5   | 56.302     | <.0001           |
| 만 나이                        | 1   | 17.762     | <.0001           |
| 가구소득                        | 1   | 0.172      | 0.6786           |
| likelihood $\chi^2=445.254$ |     | $n=12,037$ | $p-value <.0001$ |

<표 3>~<표 16>은 1차~14차 조사의 성별, 지역, 학력, 주된 활동, 만 나이를 설명변수로 2차~15차 조사의 응답여부를 종속변수로 사용한 로지스틱 회귀분석 결과를 나타내고 있다.

1차 조사에서 응답한 원가구원 13,321명을 대상으로 하여 종속변수를 2차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 3>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 424.1236이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 2차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 3> 원패널 2차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱       | 유의확률             |
|------------------------------|-----|------------|------------------|
| 성별                           | 1   | 11.7760    | 0.0006           |
| 지역                           | 14  | 105.2228   | <.0001           |
| 학력                           | 7   | 80.4856    | <.0001           |
| 주된 활동                        | 5   | 86.9154    | <.0001           |
| 만 나이                         | 1   | 51.7693    | <.0001           |
| likelihood $\chi^2=424.1236$ |     | $n=13,321$ | $p-value <.0001$ |

2차 조사에서 응답한 원가구원 12,037명을 대상으로 하여 종속변수를 3차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 4>와 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 454.7238이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 3차 조사 응답 여부를 잘 설명하고 있

는 것으로 판단된다.

<표 4> 원패널 3차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱       | 유의확률             |
|------------------------------|-----|------------|------------------|
| 성별                           | 1   | 6.4255     | 0.0112           |
| 지역                           | 15  | 158.6439   | <.0001           |
| 학력                           | 6   | 94.1776    | <.0001           |
| 주된 활동                        | 5   | 55.9027    | <.0001           |
| 만 나이                         | 1   | 20.5553    | <.0001           |
| likelihood $\chi^2=454.7238$ |     | $n=12,037$ | $p-value <.0001$ |

3차 조사에서 응답한 원가구원 10,346명을 대상으로 하여 종속변수를 4차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 5>와 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도 (likelihood) 카이제곱 검정통계량은 303.1503이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 4차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 5> 원패널 4차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱       | 유의확률             |
|------------------------------|-----|------------|------------------|
| 성별                           | 1   | 6.5876     | 0.0103           |
| 지역                           | 14  | 138.9051   | <.0001           |
| 학력                           | 6   | 45.0328    | <.0001           |
| 주된 활동                        | 5   | 60.4258    | <.0001           |
| 만 나이                         | 1   | 27.2381    | <.0001           |
| likelihood $\chi^2=303.1503$ |     | $n=10,346$ | $p-value <.0001$ |

4차 조사에서 응답한 원가구원 9,288명을 대상으로 하여 종속변수를 5차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 6>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도 (likelihood) 카이제곱 검정통계량은 341.2358이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 5차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 6> 원패널 5차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률            |
|------------------------------|-----|-----------|-----------------|
| 성별                           | 1   | 0.5960    | 0.4401          |
| 지역                           | 15  | 112.4537  | <.0001          |
| 학력                           | 6   | 59.1997   | <.0001          |
| 주된 활동                        | 5   | 65.1791   | <.0001          |
| 만 나이                         | 1   | 37.8036   | <.0001          |
| likelihood $\chi^2=341.2358$ |     | $n=9,288$ | $p-value<.0001$ |

5차 조사에서 응답한 원가구원 8,590명을 대상으로 하여 종속변수를 6차년도 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 7>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도 (likelihood) 카이제곱 검정통계량은 298.2779이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 6차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 7> 원패널 6차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률            |
|------------------------------|-----|-----------|-----------------|
| 성별                           | 1   | 25.9716   | <.0001          |
| 지역                           | 15  | 67.4533   | <.0001          |
| 학력                           | 7   | 21.9142   | 0.0026          |
| 주된 활동                        | 6   | 112.1359  | <.0001          |
| 만 나이                         | 1   | 30.1068   | <.0001          |
| likelihood $\chi^2=298.2779$ |     | $n=8,590$ | $p-value<.0001$ |

6차 조사에서 응답한 원가구원 8,135명을 대상으로 하여 종속변수를 7차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 8>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도 (likelihood) 카이제곱 검정통계량은 307.7101이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 7차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 8> 원패널 7차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 16.6025   | <.0001           |
| 지역                           | 15  | 82.8085   | <.0001           |
| 학력                           | 6   | 36.7295   | <.0001           |
| 주된 활동                        | 6   | 94.1178   | <.0001           |
| 만 나이                         | 1   | 39.8218   | <.0001           |
| likelihood $\chi^2=307.7101$ |     | $n=8,135$ | $p-value <.0001$ |

7차 조사에서 응답한 원가구원 7,775명을 대상으로 하여 종속변수를 8차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 9>와 같다. 지역을 제외한 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 지역 변수의 경우 비록 p-value가 0.05보다 크게 나타나나 그 값이 0.0555로 어느 정도 유의한 설명변수임을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 319.9556이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 8차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 9> 원패널 8차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 25.3427   | <.0001           |
| 지역                           | 15  | 24.6049   | 0.0555           |
| 학력                           | 6   | 46.0190   | <.0001           |
| 주된 활동                        | 6   | 103.9270  | <.0001           |
| 만 나이                         | 1   | 44.1817   | <.0001           |
| likelihood $\chi^2=319.9556$ |     | $n=7,775$ | $p-value <.0001$ |

8차 조사에서 응답한 원가구원 7,412명을 대상으로 하여 종속변수를 9차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 10>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 273.2110이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 9차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 10> 원패널 9차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 16.0625   | <.0001           |
| 지역                           | 15  | 36.4906   | 0.0015           |
| 학력                           | 6   | 71.3477   | <.0001           |
| 주된 활동                        | 6   | 65.3811   | <.0001           |
| 만 나이                         | 1   | 21.8924   | <.0001           |
| likelihood $\chi^2=273.2110$ |     | $n=7,412$ | $p-value <.0001$ |

9차 조사에서 응답한 원가구원 7,155명을 대상으로 하여 종속변수를 10차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 11>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 215.2555이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 10차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 11> 원패널 10차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 1.5999    | 0.2059           |
| 지역                           | 15  | 56.3676   | <.0001           |
| 학력                           | 6   | 50.0346   | <.0001           |
| 주된 활동                        | 6   | 46.5172   | <.0001           |
| 만 나이                         | 1   | 13.0659   | 0.0003           |
| likelihood $\chi^2=215.2555$ |     | $n=7,155$ | $p-value <.0001$ |

10차 조사에서 응답한 원가구원 6,937명을 대상으로 하여 종속변수를 11차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 12>와 같다. 지역을 제외한 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 312.9118이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 11차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 12> 원패널 11차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 11.4227   | 0.0007           |
| 지역                           | 15  | 17.3088   | 0.3007           |
| 학력                           | 6   | 65.5177   | <.0001           |
| 주된 활동                        | 7   | 89.3553   | <.0001           |
| 만 나이                         | 1   | 52.7763   | <.0001           |
| likelihood $\chi^2=312.9118$ |     | $n=6,937$ | $p-value <.0001$ |

11차 조사에서 응답한 원가구원 6,648명을 대상으로 하여 종속변수를 12차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 13>과 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도 (likelihood) 카이제곱 검정통계량은 274.8450이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 12차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 13> 원패널 12차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 12.1352   | 0.0005           |
| 지역                           | 15  | 39.0176   | 0.0006           |
| 학력                           | 6   | 47.9420   | <.0001           |
| 주된 활동                        | 7   | 62.2705   | <.0001           |
| 만 나이                         | 1   | 33.5546   | <.0001           |
| likelihood $\chi^2=274.8450$ |     | $n=6,648$ | $p-value <.0001$ |

12차 조사에서 응답한 원가구원 6,311명을 대상으로 하여 종속변수를 13차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 14>와 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도 (likelihood) 카이제곱 검정통계량은 354.0426이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 13차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.



<표 14> 원패널 13차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 19.8229   | <.0001           |
| 지역                           | 14  | 40.3033   | 0.0002           |
| 학력                           | 6   | 64.6776   | <.0001           |
| 주된 활동                        | 8   | 93.8192   | <.0001           |
| 만 나이                         | 1   | 25.8987   | <.0001           |
| likelihood $\chi^2=354.0426$ |     | $n=6,311$ | $p-value <.0001$ |

13차 조사에서 응답한 원가구원 5,956명을 대상으로 하여 종속변수를 14차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 15>와 같다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 294.2943이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 14차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 15> 원패널 14차 로지스틱 회귀분석 결과

| 변수                           | 자유도 | 카이제곱      | 유의확률             |
|------------------------------|-----|-----------|------------------|
| 성별                           | 1   | 27.1376   | <.0001           |
| 지역                           | 14  | 41.7994   | 0.0001           |
| 학력                           | 6   | 43.0999   | <.0001           |
| 주된 활동                        | 7   | 52.1722   | <.0001           |
| 만 나이                         | 1   | 40.9859   | <.0001           |
| likelihood $\chi^2=294.2943$ |     | $n=5,956$ | $p-value <.0001$ |

14차 조사에서 응답한 원가구원 5,670명을 대상으로 하여 종속변수를 15차 조사 응답여부로, 설명변수를 성, 지역, 학력, 주된 활동, 만 나이로 정의하여 원패널 로지스틱 회귀분석을 실시한 결과는 <표 16>과 같다. 기존의 15차 이전의 로지스틱 회귀분석에서는 모형에 상수항을 포함 하였으나, 15차 로지스틱 회귀분석에서는 디자인행렬이 full rank가 되도록 상수항을 고려하지 않았다. 고려된 모든 설명 변수가  $\alpha=0.05$  수준에서 유의함을 알 수 있다. 모형 적합도를 평가하기 위한 우도(likelihood) 카이제곱 검정통계량은 5410.2710이며 대응 p-value는 0.0001보다 작게 나타나 고려된 로지스틱 회귀모형이 15차 조사 응답 여부를 잘 설명하고 있는 것으로 판단된다.

<표 16> 원패널 15차 로지스틱 회귀분석 결과

| 변수                              | 자유도 | 카이제곱        | 유의확률               |
|---------------------------------|-----|-------------|--------------------|
| 성별                              | 1   | 27.9735     | <.0001             |
| 지역                              | 14  | 44.4425     | <.0001             |
| 학력                              | 6   | 91.3467     | <.0001             |
| 주된 활동                           | 7   | 74.6247     | <.0001             |
| 만 나이                            | 1   | 224.6288    | <.0001             |
| likelihood $\chi^2 = 5410.2710$ |     | $n = 5,670$ | $p - value <.0001$ |

15차 로지스틱 회귀모형을 설정하기 위하여 사용된 설명변수 중 범주형 설명변수 성별, 학력, 지역, 그리고 주된 활동의 응답유형에 따른 분포는 <표 17>~<표 20>과 같다. 이 중 <표 20>을 살펴보면, 일시휴직 범주에서 무응답인 가구원이 0명으로 모두 응답한 가구원만 존재하게 된다. 따라서 차수가 지남에 따라 각 범주형 범주에서 이러한 현상이 발생할 수 있으며, 심한 경우에는 로지스틱 모형의 적용이 불가할 수 있기 때문에 추 후 분석에서는 이러한 경우, 범주를 통합하는 방안을 고려하여야 할 것이다.

<표 17> 원패널 15차 성별에 따른 응답 여부

| 빈도 | 무응답 | (%)   | 응답    | (%)    | 전체    |
|----|-----|-------|-------|--------|-------|
| 남성 | 206 | (8.4) | 2,251 | (91.6) | 2,457 |
| 여성 | 180 | (5.6) | 3,033 | (94.4) | 3,213 |
| 합계 | 386 | (6.8) | 5,284 | (93.2) | 5,670 |

<표 18> 원패널 15차 학력에 따른 응답 여부

| 빈도   | 무응답 | (%)    | 응답    | (%)    | 전체    |
|------|-----|--------|-------|--------|-------|
| 무학   | 21  | (6.9)  | 284   | (93.1) | 305   |
| 초졸   | 38  | (4.1)  | 895   | (95.9) | 933   |
| 중졸   | 42  | (4.8)  | 826   | (95.2) | 868   |
| 고졸   | 113 | (5.5)  | 1943  | (94.5) | 2,056 |
| 전문대졸 | 50  | (9.4)  | 480   | (90.6) | 530   |
| 대졸   | 113 | (13.1) | 750   | (86.9) | 863   |
| 대학원졸 | 9   | (7.8)  | 106   | (92.2) | 115   |
| 합계   | 386 | (6.8)  | 5,284 | (93.2) | 5,670 |

<표 19> 원패널 15차 지역에 따른 응답 여부

| 빈도 | 무응답 | (%)    | 응답    | (%)    | 전체    |
|----|-----|--------|-------|--------|-------|
| 서울 | 107 | (9.5)  | 1,017 | (90.5) | 1,124 |
| 부산 | 34  | (5.6)  | 574   | (94.4) | 608   |
| 대구 | 12  | (4.1)  | 279   | (95.9) | 291   |
| 대전 | 5   | (2.8)  | 173   | (97.2) | 178   |
| 인천 | 20  | (5.7)  | 331   | (94.3) | 351   |
| 광주 | 9   | (6.7)  | 125   | (93.3) | 134   |
| 울산 | 10  | (4.8)  | 200   | (95.2) | 210   |
| 경기 | 89  | (8.1)  | 1,006 | (91.9) | 1,095 |
| 강원 | 17  | (13.8) | 106   | (86.2) | 123   |
| 충북 | 6   | (3.7)  | 157   | (96.3) | 163   |
| 충남 | 13  | (5.8)  | 210   | (94.2) | 223   |
| 전북 | 6   | (2.6)  | 226   | (97.4) | 232   |
| 전남 | 19  | (10.6) | 160   | (89.4) | 179   |
| 경북 | 21  | (7.4)  | 261   | (92.6) | 282   |
| 경남 | 18  | (3.8)  | 459   | (96.2) | 477   |
| 합계 | 386 | (6.8)  | 5,284 | (93.2) | 5,670 |

<표 20> 원패널 15차 주된활동에 따른 응답 여부

| 빈도   | 무응답 | (%)    | 응답    | (%)     | 전체    |
|------|-----|--------|-------|---------|-------|
| 일하였음 | 119 | (4.1)  | 2764  | (95.9)  | 2883  |
| 일시휴직 | 0   | (0.0)  | 13    | (100.0) | 13    |
| 구직활동 | 4   | (19.0) | 17    | (81.0)  | 21    |
| 가사   | 26  | (2.6)  | 976   | (97.4)  | 1002  |
| 학업   | 2   | (3.0)  | 65    | (97.0)  | 67    |
| 연로   | 141 | (18.2) | 635   | (81.8)  | 776   |
| 심신장애 | 2   | (11.1) | 16    | (88.9)  | 18    |
| 기타   | 92  | (10.3) | 798   | (89.7)  | 890   |
| 합계   | 386 | (6.8)  | 5,284 | (93.2)  | 5,670 |

## 2. 원패널 가중치 산출과정

원년도를 기준으로  $t$ 년도의 가구원 구조는 <그림 1>과 같다. <그림 1>을 바탕으로 중·횡단면 가구원 가중치 및 가구 가중치 부여 대상을 정리하면 다음과 같다.

- 개인가중치
  - 중단면 가중치 부여대상 - ①-1), ①-2), ②-1), ②-2)
  - 횡단면 가중치 부여대상 - 전체 가구원 (단, 해당 가구원이 속한 가구에 원가구원이 한명도 없는 경우는 “0”의 가중치가 부여됨. 즉 원가구 소멸로 인한 개인소멸로 간주됨.)
- 가구가중치 - 조사년도에 응답한 가구에 부여됨 (단, 응답된 가구에 원가구원이 한명도 없는 경우는 “0”의 가중치가 부여됨, 즉 원가구 소멸로 간주됨.)

<그림 1> 원패널 가구원 구조

| 1차년도   | ..... | $t-1$ 년도   | $t$ 년도        |
|--|-------|------------|---------------|
| 원가구원<br>: 1차시점 응답(13,319)                            |       | ① 원가구원     | ①-1) 원가구원     |
|  |       |            | ①-2) 원가구원     |
| 원가구원<br>:1차시점 무응답(534)<br>+만15세미만포함(3,650)<br>=4,184 |       | ② 신규진입원가구원 | ②-1) 원가구원     |
|  |       |            | ②-2) 신규진입원가구원 |
|  |       | ③ 비원가구원    | ③ 비원가구원       |
|  |       |            | ③ 비원가구원       |

$t$ 차년도의 원패널 가중치 산출은 총 3 단계(1단계 : 중단면 가구원 가중치 산출 → 2 단계 : 가구 가중치 산출 → 3단계 : 횡단면 가구원 가중치 산출)의 과정을 통해 이루어진다. 각 단계별 구체적인 설명은 다음과 같다.

## 1단계 : 종단면 가구원 가중치 산출

종단면 가구원 가중치 산출은 기본적으로 무응답 보정의 과정을 통해 이루어진다. 무응답 보정을 위해  $t$ 차년도의 각 가구원의 응답여부를 종속변수로,  $t-1$ 차 년도의 변수(성, 연령, 지역, 교육수준, 주된 활동, 취업 여부)을 설명변수로 하는 로지스틱 회귀분석을 실시하고 이를 통하여 무응답 보정 가중치를 산출하였다. 이 과정에서 생성된 가중치를 **기초가중치**로 정의한다. <그림 1>의 구조를 이용해 산출되는 기초가중치를 정의하면 다음과 같다.

- 1)  $t-1$ 년도 기준 원가구원에 해당하는 ①과 ②의 종단면 가중치에 로지스틱 모형을 이용한 무응답보정을 통하여 ①-1)과 ②-1)에 기초가중치를 부여한다. 즉 ①-1)과 ②-1)에 부여되는 기초가중치는 아래의 식 (3)과 같다. 여기서  $w_{pl,t-1,i}^*$ 는  $t-1$ 년도에  $i$ 번째 가구원에게 부여된 종단면 가구원 가중치를 나타내며  $\hat{p}_{ti}$ 는 모형 (1)을 이용하여 예측된  $t$ 년도에  $i$ 번째 응답확률을 의미한다.

$$w_{\text{base},t,i} = \hat{p}_{ti}^{-1} w_{pl,t-1,i}^* \quad \hat{p}_{ti} = [1 + \exp(\mathbf{x}'_{t-1,i} \hat{\boldsymbol{\beta}})] \quad (3)$$

- 2) ①-2)는 최근에 조사된 조사시점의 종단면 가구원 가중치로 기초가중치를 부여한다.
- 3) ③의 비원가구원에게는 “0”의 기초가중치를 부여한다.
- 4) ②-2)의 신규 진입된 원가구원의 경우 기초가중치 정의에서 제외된다. 즉 기초가중치를 부여하지 않는다.

위에서 정의된 기초가중치를 이용하여 생성되는 종단면 가구원 가중치는 다음과 같다. 먼저는 스케일 조정전의 종단면 가중치  $w_{pl,t,i}$ 을 생성한 후 최종 종단면 가중치  $w_{pl,t,i}^*$ 가 스케일 조정을 통해 산출된다. 그 과정은 다음과 같다.

- i) ①-1), ①-2) 그리고 ②-1)에 해당하는 가구원에게는 위에서 산출된 기초가중치를 종단면 가중치로 부여한다. 즉

$$w_{pl,t,i} = w_{\text{base},t,i}$$

- ii) ②-2)의 신규진입원가구원에는 스케일 보정전의 가구가중치,  $w_{h,t}$ 을 종단면 가중치로 부여한다. [식 (5) 참조] 즉

$$w_{pl,t,i} = w_{h,t}$$

iii) i)과 ii)를 통해 산출된 종단면 가중치에 추계인구증가율을 이용한 스케일 조정을 통해 최종 종단면 가중치를 산출한다. 즉 최종 종단면 가구원 가중치는 다음과 같다.

$$w_{pl,t,i}^* = w_{pl,t,i} \frac{\sum_i w_{pl,t-1,i}^*}{\sum_i w_{pl,t,i}} \times \text{추계인구증가율..} \quad (4)$$

## 2단계 : 가구 가중치 산출

가구 가중치 역시 1단계에서 계산된 기초가중치를 이용하여 산출한다. 가구 가중치는 기본적으로 1단계에서 산출된 기초 가중치들의 각 가구별 평균으로 정의된다. 즉 가구 가중치는 다음의 식 (5)로 정의되며  $n_h$ 는 가구  $h$ 의 가구원 수를 의미한다. 가구원 수는 가구 내 0 혹은 양의 기초가중치를 부여 받은 가구원의 수를 나타낸다.

$$w_{t,h} = \frac{\sum_{i \in h} w_{\text{base},t,i}}{n_h} \quad (5)$$

단, 가구에 속한 원가구원이 신규진입가구원만으로 구성된 경우 가구가중치는 0 또는 계산되어지지 않는다. 따라서 이때의 가구 가중치는 신규진입가구원이 속한 1차년도에 가구가중치로 정의한다. 다음의 <표 21>은 2차~15차 조사까지에서 발생된 가구에 속한 원가구원이 신규진입가구원만으로 구성된 가구의 정보이다.

<표 21> 신규진입가구원만으로 구성된 가구 정보(비원가구원 제외)

| 차수  | PID    | 현 차수 가구번호 | 1차 가구번호 | 1인 가구 여부 |
|-----|--------|-----------|---------|----------|
| 2차  | 90203  | 5011      | 901     | ×        |
| 3차  | 160205 | 5284      | 1602    | ×        |
|     | 445103 | 4451      | 4451    | ×        |
| 4차  | 303201 | 3032      | 3032    | ○        |
|     | 123503 | 5545      | 1235    | ○        |
|     | 78504  | 5145      | 785     | ×        |
|     | 450203 | 5438      | 4502    | ×        |
|     | 303201 | 3032      | 3032    | ×        |
| 5차  | 123503 | 5545      | 1235    | ×        |
|     | 435402 | 5796      | 4354    | ×        |
| 6차  | 433303 | 4333      | 4333    | ○        |
|     | 199402 | 5171      | 1994    | ○        |
|     | 259902 | 5914      | 2599    | ○        |
|     | 433303 | 4333      | 4333    | ×        |
|     | 75604  | 5144      | 756     | ×        |
|     | 268203 | 5402      | 2682    | ×        |
| 7차  | 94105  | 5874      | 941     | ×        |
|     | 462803 | 6237      | 4628    | ○        |
|     | 410404 | 6342      | 4104    | ○        |
|     | 128403 | 1284      | 1284    | ×        |
|     | 447702 | 4477      | 4477    | ×        |
| 8차  | 43306  | 6401      | 4330    | ×        |
|     | 305703 | 6692      | 3057    | ×        |
| 9차  | 191604 | 5890      | 1916    | ○        |
|     | 217103 | 6841      | 2171    | ○        |
|     | 456205 | 4562      | 4562    | ×        |
|     | 373803 | 5218      | 3738    | ×        |
| 11차 | 315501 | 3155      | 3155    | ×        |
|     | 311304 | 7363      | 3113    | ○        |
| 12차 | 102704 | 5153      | 1027    | ○        |
|     | 422903 | 5234      | 4229    | ○        |
|     | 405303 | 7882      | 4053    | ○        |
|     | 410406 | 7911      | 4104    | ○        |
| 13차 | 99507  | 6043      | 995     | ×        |
|     | 399103 | 7419      | 3991    | ○        |
|     | 66703  | 7554      | 667     | ○        |
|     | 269202 | 8095      | 2692    | ○        |
| 14차 | 306303 | 5409      | 3063    | ×        |

식 (5)의 가중치에 추계 가구 수 정보를 통해 산출된 스케일 조정 지수를 적용하여 최종 가구 가중치가 산출된다. 즉 최종 가구 가중치는 다음과 같이 산출된다. 여기서  $w_{h,t-1}^*$ 는  $t-1$ 시점의 최종 가구가중치를 나타낸다.

$$w_{t,h}^* = w_{t,h} \frac{\sum_j w_{t-1,j}^*}{\sum_j w_{t,j}} \times \text{추계가구증가율} \quad (6)$$

### 3단계 : 횡단면 가구원 가중치 산출

횡단면 가구원 가중치는  $t$ 시점에 원가구원을 한 명이상 포함하고 있는 가구에 속하는 모든 가구원에게 부여된다. 횡단면 가구원 가중치 산출을 위해서는 먼저 스케일 보정전의 가구가중치를 각 가구원에게 부여한다. 즉 스케일 조정 전 가구  $h$ 에 거주하는 가구원  $i$ 에게 부여되는 횡단면 가중치는 다음과 같으며 여기서  $w_{t,h}$ 는 2단계에서 사용된 가구 가중치이다.

$$w_{pc,t,i} = w_{t,h} \quad (7)$$

최종 횡단면 가구원 가중치는 위의 가중치 (7)에 스케일 조정을 적용하여 산출된다. 즉 산출된 최종 횡단면 가구원 가중치는 다음과 같다. 여기서 스케일 조정은 추계인구증가율이 고려된 종단면 가구원 가중치를 이용하여 이루어졌다. 가구 가중치 산출시 예외의 경우가 되는 신규진입가구원만이 원가구원으로 구성된 가구 가중치 때문에 종단면 가중치와 횡단면 가중치가 동일하지 않게 된다. 이러한 예외의 경우, 신규진입 원가구원의 종단면 가중치와 횡단면 가중치가 가구가중치로 같게 되나, 그 가구에 비원가구가 같이 존재하는 경우에는 비원가구에게도 횡단면 가중치를 가구가중치로 부여하기 때문에 횡단면 가중치의 합이 항상 종단면 가중치의 합보다 크게 된다. 여기서 고려하는 모집단은 종단면 가중치를 통하여 대표되는 것을 모집단으로 고려하기 때문에 종단면 가중치의 합과 동일하도록 스케일 조정을 실시한다. 즉, 식 (8)의 스케일 조정은 신규진입가구원만이 원가구원으로 구성된 가구에서 비원가구에게 횡단면 가중치를 부여함으로써 증가한 부분을 보정한 후 다시 추계인구증가율을 고려하여 조정하는 것이다.

$$w_{pc,t,i}^* = w_{pc,t,i} \frac{\sum_j w_{pl,t,j}^*}{\sum_j w_{pc,t,j}} \quad (8)$$

1차~15차 조사에서 중·횡단면 가구원 가중치 및 가구 가중치의 스케일 조정을 위한 추계 인구·가구 증가율은 <표 22>와 같다. 여기서 가구에 대한 2차, 3



차 추계증가율에 대한 정보가 존재하지 않아 기존의 가구증가율을 그대로 반영하였다.

<표 22> 추계 인구·가구 증가율

| 차수  | 년도   | 인구         |         | 가구         |         |
|-----|------|------------|---------|------------|---------|
|     |      | 인구수        | 인구증가율   | 가구수        | 가구증가율   |
| 1차  | 1998 | 46,286,503 |         |            |         |
| 2차  | 1999 | 46,616,677 | 1.00713 |            |         |
| 3차  | 2000 | 47,008,111 | 1.00840 | 14,507,010 |         |
| 4차  | 2001 | 47,357,362 | 1.00743 | 14,843,989 | 1.02323 |
| 5차  | 2002 | 47,622,179 | 1.00559 | 15,170,029 | 1.02196 |
| 6차  | 2003 | 47,859,311 | 1.00498 | 15,465,163 | 1.01946 |
| 7차  | 2004 | 48,039,415 | 1.00376 | 15,720,436 | 1.01651 |
| 8차  | 2005 | 48,138,077 | 1.00205 | 15,971,010 | 1.01594 |
| 9차  | 2006 | 48,371,946 | 1.00486 | 16,289,194 | 1.01992 |
| 10차 | 2007 | 48,597,652 | 1.00467 | 16,542,700 | 1.01556 |
| 11차 | 2008 | 48,948,698 | 1.00722 | 16,791,160 | 1.01502 |
| 12차 | 2009 | 49,182,038 | 1.00477 | 17,052,164 | 1.01554 |
| 13차 | 2010 | 49,410,366 | 1.00464 | 17,359,333 | 1.01801 |
| 14차 | 2011 | 49,779,440 | 1.00747 | 17,687,001 | 1.01888 |
| 15차 | 2012 | 50,004,441 | 1.00452 | 17,950,675 | 1.01491 |

### 3. 원패널 가중치 부여 결과

<표 23>~<표 37>는 1차 년도부터 15차 년도까지 산출된 가중치들의 기초통계량을 나타내고 있다.

1차년도 조사에서 정의된 원가구 5000가구와 그 원가구원 13,321에 대한 가구원 가중치의 분포는 <표 23>과 같다.

<표 23> 1차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차  | 최솟값     | 최댓값     |
|--------|-----|--------|--------------|---------|---------|-------|---------|---------|
| w01_p  | 가구  | 5,000  | 11,276,899.4 | 2,255.4 | 2,225.1 | 415.1 | 1,513.2 | 4,514.6 |
| w02pls | 가구원 | 13,321 | 30,039,370.6 | 2,255.0 | 2,225.1 | 416.8 | 1,513.2 | 4,514.6 |

2차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 11,708명과 비원가구원 323명 그리고 그 가구원들이 속한 4,503가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 24>와 같다. 실제 조사된 전체 4,507가구 중 4개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 6명으로 횡단면 가중치를 0으로 부여받았다.

<표 24> 2차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차  | 최솟값   | 최댓값     |
|--------|-----|--------|--------------|---------|---------|-------|-------|---------|
| w02hs  | 가구  | 4,503  | 11,453,518.5 | 2,543.5 | 2,513.9 | 561.2 | 472.1 | 5,987.0 |
| w02pls | 중단면 | 11,708 | 30,253,649.5 | 2,584.0 | 2,515.4 | 516.6 | 575.7 | 6,625.3 |
| w02pcs | 횡단면 | 12,031 | 30,253,649.5 | 2,514.6 | 2,492.9 | 562.6 | 468.0 | 5,934.9 |

3차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,797명과 비원가구원 397명 그리고 그 가구원들이 속한 4,260가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 25>와 같다. 실제 조사된 전체 4,266가구 중 6개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 11명으로 횡단면 가중치를 0으로 부여받았다.

<표 25> 3차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차  | 최솟값   | 최댓값     |
|--------|-----|--------|--------------|---------|---------|-------|-------|---------|
| w03hs  | 가구  | 4,260  | 11,632,900.5 | 2,730.7 | 2,676.7 | 702.5 | 458.4 | 7,400.8 |
| w03pls | 중단면 | 10,797 | 30,507,685.3 | 2,825.6 | 2,715.3 | 683.2 | 602.7 | 8,992.2 |
| w03pcs | 횡단면 | 11,194 | 30,507,685.3 | 2,725.4 | 2,675.0 | 699.7 | 456.5 | 7,370.0 |

4차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,500명과 비원가구원 521명 그리고 그 가구원들이 속한 4,230가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 26>과 같다. 실제 조사된 전체 4,247가구 중 17개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 30명으로 횡단면 가중치를 0으로 부여받았다.

<표 26> 4차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차  | 최솟값   | 최댓값     |
|--------|-----|--------|--------------|---------|---------|-------|-------|---------|
| w04hs  | 가구  | 4,230  | 11,903,117.6 | 2,814.0 | 2,757.7 | 839.6 | 283.8 | 7,983.2 |
| w04pls | 중단면 | 10,500 | 30,734,344.9 | 2,927.1 | 2,780.3 | 827.9 | 579.8 | 9,297.5 |
| w04pcs | 횡단면 | 11,021 | 30,734,344.9 | 2,788.7 | 2,735.2 | 824.2 | 279.9 | 7,874.4 |

5차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,296명과 비원가구원 628명 그리고 그 가구원들이 속한 4,275가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 27>과 같다. 실제 조사된 전체 4,298가구 중 23개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 42명으로 횡단면 가중치를 0으로 부여받았다.

<표 27> 5차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차  | 최솟값   | 최댓값      |
|--------|-----|--------|--------------|---------|---------|-------|-------|----------|
| w05hs  | 가구  | 4,275  | 12,164,563.0 | 2,845.5 | 2,757.3 | 960.2 | 108.6 | 8,241.9  |
| w05pls | 중단면 | 10,296 | 30,906,207.9 | 3,001.8 | 2,806.3 | 972.5 | 429.2 | 10,665.6 |
| w05pcs | 횡단면 | 10,924 | 30,906,207.9 | 2,829.2 | 2,747.8 | 935.9 | 107.3 | 8,141.1  |

6차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,658명과 비원가구원 843명 그리고 그 가구원들이 속한 4,571가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 28>과 같다. 실제 조사된 전체 4,592가구 중 21개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 40명으로 횡단면 가중치를 0으로 부여받았다.

<표 28> 6차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값   | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|-------|----------|
| w06hs  | 가구  | 4,571  | 12,401,225.5 | 2,713.0 | 2,607.0 | 1,028.1 | 97.0  | 9,239.1  |
| w06pls | 중단면 | 10,658 | 31,060,103.6 | 2,914.3 | 2,686.4 | 1,052.4 | 254.5 | 13,272.9 |
| w06pcs | 횡단면 | 11,501 | 31,060,103.6 | 2,700.6 | 2,602.1 | 998.1   | 95.7  | 9,113.5  |

7차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,695명과 비원가구원 959명 그리고 그 가구원들이 속한 4,756가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 29>와 같다. 실제 조사된 전체 4,761가구 중 5개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 7명으로 횡단면 가중치를 0으로 부여받았다.

<표 29> 7차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값  | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|------|----------|
| w07hs  | 가구  | 4,756  | 12,605,924.1 | 2,650.5 | 2,524.0 | 1,115.3 | 29.4 | 11,262.7 |
| w07pls | 중단면 | 10,695 | 31,176,988.9 | 2,915.1 | 2,649.4 | 1,187.2 | 99.9 | 22,628.3 |
| w07pcs | 횡단면 | 11,654 | 31,176,988.9 | 2,675.2 | 2,562.3 | 1,093.9 | 29.3 | 11,223.6 |

8차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,522명과 비원가구원 1,048명 그리고 그 가구원들이 속한 4,840가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 30>과 같다. 실제 조사된 전체 4,849가구 중 9개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 10명으로 횡단면 가중치를 0으로 부여받았다.

<표 30> 8차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값  | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|------|----------|
| w08hs  | 가구  | 4,840  | 12,806,854.7 | 2,646.0 | 2,506.5 | 1,228.8 | 27.2 | 14,569.8 |
| w08pls | 중단면 | 10,522 | 31,241,019.3 | 2,969.1 | 2,666.9 | 1,368.5 | 97.3 | 24,441.0 |
| w08pcs | 횡단면 | 11,570 | 31,241,019.3 | 2,700.2 | 2,575.2 | 1,216.8 | 27.3 | 14,607.9 |

9차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,591명과 비원가구원 1,150명 그리고 그 가구원들이 속한 4,990가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 31>과 같다. 실제 조사된 전체 5,001가구 중 11개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 15명으로 횡단면 가중치를 0으로 부여받았다.

<표 31> 9차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값  | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|------|----------|
| w09hs  | 가구  | 4,990  | 13,062,000.5 | 2,617.6 | 2,465.6 | 1,318.5 | 23.7 | 18,005.2 |
| w09pls | 중단면 | 10,591 | 31,392,797.4 | 2,964.1 | 2,625.2 | 1,501.5 | 47.6 | 32,839.4 |
| w09pcs | 횡단면 | 11,741 | 31,392,797.4 | 2,673.8 | 2,535.5 | 1,299.4 | 23.8 | 18,048.2 |

10차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,590명과 비원가구원 1,250명 그리고 그 가구원들이 속한 5,059가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 32>와 같다. 실제 조사된 전체 5,069가구 중 10개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 15명으로 횡단면 가중치를 0으로 부여받았다.

<표 32> 10차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값  | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|------|----------|
| w10hs  | 가구  | 5,059  | 13,265,282.3 | 2,622.1 | 2,453.7 | 1,428.4 | 10.7 | 24,077.6 |
| w10pls | 중단면 | 10,590 | 31,539,277.8 | 2,978.2 | 2,624.6 | 1,612.2 | 42.7 | 39,242.7 |
| w10pcs | 횡단면 | 11,840 | 31,539,277.8 | 2,663.8 | 2,506.2 | 1,384.1 | 10.7 | 24,041.2 |

11차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,411명과 비원가구원 1,309명 그리고 그 가구원들이 속한 5,105가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 33>과 같다. 실제 조사된 전체 5,116가구 중 11개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 14명으로 횡단면 가중치를 0으로 부여받았다.

<표 33> 11차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값  | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|------|----------|
| w11hs  | 가구  | 5,105  | 13,464,517.7 | 2,637.5 | 2,449.0 | 1,588.3 | 8.6  | 35,459.8 |
| w11pls | 중단면 | 10,411 | 31,767,102.4 | 3,051.3 | 2,632.8 | 1,909.4 | 25.8 | 67,843.0 |
| w11pcs | 횡단면 | 11,720 | 31,767,102.4 | 2,710.5 | 2,523.1 | 1,562.6 | 8.6  | 35,542.1 |

12차 조사에서 중·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,356명과 비원가구원 1,363명 그리고 그 가구원들이 속한 5,289가구의 가구가중치 및 가구원 중·횡단면 가중치의 분포는 <표 34>와 같다. 실제 조사된 전체 5,306가구 중 17개

의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구 소멸로 나타났으며 그로 인한 가구원소멸은 20명으로 횡단면 가중치를 0으로 부여받았다.

<표 34> 12차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값  | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|------|----------|
| w12hs  | 가구  | 5,289  | 13,673,811.9 | 2,585.3 | 2,364.3 | 1,721.1 | 2.5  | 42,376.2 |
| w12pls | 종단면 | 10,356 | 31,918,537.2 | 3,082.1 | 2,608.6 | 2,165.7 | 10.2 | 83,625.3 |
| w12pcs | 횡단면 | 11,719 | 31,918,537.2 | 2,723.7 | 2,511.4 | 1,731.3 | 2.5  | 43,422.9 |

13차 조사에서 종·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 10,133명과 비원가구원 1,427명 그리고 그 가구원들이 속한 5,361가구의 가구가중치 및 가구원 종·횡단면 가중치의 분포는 <표 35>와 같다. 실제 조사된 전체 5,379가구 중 18개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구 소멸로 나타났으며 그로 인한 가구원소멸은 22명으로 횡단면 가중치를 0으로 부여받았다.

<표 35> 13차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값 | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|-----|----------|
| w13hs  | 가구  | 5,361  | 13,920,125.0 | 2,596.6 | 2,307.1 | 1,868.9 | 2.6 | 44,732.2 |
| w13pls | 종단면 | 10,133 | 32,066,719.2 | 3,164.6 | 2,597.2 | 2,482.9 | 8.3 | 89,499.0 |
| w13pcs | 횡단면 | 11,560 | 32,066,719.2 | 2,773.9 | 2,484.7 | 1,923.9 | 2.7 | 46,338.8 |

14차 조사에서 종·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 9,877명과 비원가구원 1,470명 그리고 그 가구원들이 속한 5,380가구의 가구가중치 및 가구원 종·횡단면 가중치의 분포는 <표 36>과 같다. 실제 조사된 전체 5,403가구 중 23개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 30명으로 횡단면 가중치를 0으로 부여받았다.

<표 36> 14차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값 | 최댓값      |
|--------|-----|--------|--------------|---------|---------|---------|-----|----------|
| w14hs  | 가구  | 5,380  | 14,182,875.9 | 2,636.2 | 2,298.3 | 2,044.0 | 0.8 | 46,913.4 |
| w14pls | 종단면 | 9,877  | 32,306,243.7 | 3,270.9 | 2,635.1 | 2,836.4 | 2.5 | 94,662.8 |
| w14pcs | 횡단면 | 11,347 | 32,306,243.7 | 2,847.1 | 2,497.8 | 2,146.5 | 0.8 | 48,964.0 |

15차 조사에서 종·횡단면 가중치로 양의 값을 부여받은 가구원은 원가구원 9,861명

과 비원가구원 1,567명 그리고 그 가구원들이 속한 5,456가구의 가구가중치 및 가구원 중 · 횡단면 가중치의 분포는 <표 37>과 같다. 실제 조사된 전체 5,469가구 중 13개의 가구가 0의 가중치를 부여받아(가구에 소속된 원가구원이 한 명도 없는 경우) 가구소멸로 나타났으며 그로 인한 가구원소멸은 16명으로 횡단면 가중치를 0으로 부여받았다.

<표 37> 15차 원패널 가중치의 기초통계량

| 변수     | 라벨  | 관측치수   | 합계           | 평균값     | 중위수     | 표준편차    | 최솟값 | 최댓값       |
|--------|-----|--------|--------------|---------|---------|---------|-----|-----------|
| w15hs  | 가구  | 5,456  | 14,394,311.1 | 2,638.3 | 2,257.1 | 2,427.4 | 0.8 | 93,726.0  |
| w15pls | 종단면 | 9,861  | 32,452,266.6 | 3,291.0 | 2,578.2 | 3,552.4 | 2.6 | 191,803.6 |
| w15pcs | 횡단면 | 11,428 | 32,452,266.6 | 2,839.7 | 2,438.8 | 2,541.7 | 0.9 | 97,005.1  |

### III. 통합 패널 가중치 산출

#### 1. 통합 패널 구축 현황

한국노동패널조사의 통합패널은 1998년부터 수행되고 있는 원패널과 2009년부터 전국 모집단으로 확대하기 위해 추가된 추가패널을 통합한 패널을 의미한다. 통합패널은 전국 모집단으로 확대하기 위해 추가된 2009년을 통합1차 시점으로 정의한다.

통합패널의 필요성을 살펴보기 위해 1998년 원패널의 운영 현황과 한계점을 검토하고, 추가 패널의 필요성을 살펴볼 필요가 있다. 1998년 최초 설계되어 시작된 한국노동패널조사의 조사대상 지역과 표본규모는 도시지역에 거주하는 5000가구로 패널을 구성하였으며, 2008년(11차) 기준으로 11차 모두 응답한 가구수는 2,607가구로 최초 원표본 5,000가구의 52.1%에 해당된다. 이후 11차까지 조사가 진행됨에 따라 추적 실패 및 응답 거부 등의 사유로 인한 패널가구 마모와 더불어 패널가구가 시지역에서 군지역으로 이동하고, 설계 시점의 군지역에 거주하는 가구도 시지역으로 이주하게 되며, 군지역의 도시화 등의 변동으로 새롭게 시로 승격되는 지역이 발생하는 등 변화하는 모집단 특성을 반영하지 못함으로써 시지역 및 전국을 대표하는 데 한계가 있음이 지적되고 왔다.

이러한 한계를 보완하고, 전국 모집단으로의 확대 필요성에 따라 2009년 연구를 통해 추가 패널을 구축하게 된다. 통합패널의 가중치 산출을 위해 2009년 연구에서의 통합패널을 위한 설계와 통합 시점에서 부여된 가중치를 검토할 필요가 있다.

통합패널의 구축 현황과 가중치 산출 과정을 이해하기 위해 2009년 연구결과(김재광, 2009)를 요약해 정리하면 다음과 같다.

첫째, 추가패널 구축의 규모는 조사비용의 수용성을 감안하여 1998년의 패널조사 시작당시와 같은 규모인 5,000가구로 하였으며, 이를 기준으로 전국 모집단으로 확대하기 위해 11차 년도의 응답가구와 2005년 인구주택총조사 가구의 16개 시도별 동부/읍면부별로 분포를 비교한 후에 추가해야 할 가구수를 계산하였다. 추가 표본은 조사구당 5가구를 추출한다는 가정에서 추출해야 할 조사구수를 산출하였으며, 16개 시도별 동부/읍면부별로 추가표본규모와 조사구수는 <표 38>에 정리하였다.

둘째, 이를 바탕으로 통합패널을 분석하기 위한 가중치는 2009년 연구에서 부여되었으며, 2009년 통합시점에서 가중치 산출을 위한 조사모집단은 추가패널 구축의 기준이 되는 2005년 인구주택총조사 가구 데이터로 정의되었다. 가중치 산출 과정은 2절에서 설명하기로 한다.



<표 38> 시도별/동읍면부별 추가패널 구축 현황

| 시도 | 동부    |      |     | 읍면부  |      |     |
|----|-------|------|-----|------|------|-----|
|    | 현행패널  | 추가패널 | 조사구 | 현행패널 | 추가패널 | 조사구 |
| 서울 | 1,303 | 185  | 37  | 0    | 0    | 0   |
| 부산 | 461   | 65   | 13  | 9    | 0    | 0   |
| 대구 | 330   | 20   | 4   | 18   | 5    | 1   |
| 인천 | 319   | 40   | 8   | 7    | 0    | 0   |
| 광주 | 180   | 25   | 5   | 0    | 0    | 0   |
| 대전 | 168   | 30   | 6   | 0    | 0    | 0   |
| 울산 | 107   | 25   | 5   | 20   | 5    | 1   |
| 경기 | 1,004 | 240  | 48  | 194  | 50   | 10  |
| 강원 | 120   | 25   | 5   | 29   | 60   | 12  |
| 충북 | 108   | 30   | 6   | 17   | 70   | 14  |
| 충남 | 80    | 30   | 6   | 95   | 85   | 17  |
| 전북 | 161   | 25   | 5   | 60   | 30   | 6   |
| 전남 | 102   | 15   | 3   | 51   | 125  | 25  |
| 경북 | 158   | 50   | 10  | 139  | 65   | 13  |
| 경남 | 234   | 40   | 8   | 108  | 85   | 17  |
| 제주 | 2     | 55   | 11  | 1    | 20   | 4   |
| 합계 | 4,837 | 900  | 180 | 748  | 600  | 120 |

표본조사구의 추출은 동부는 가구원 수가 1~2인 가구수와 단독주택수를 기준으로 확률비례추출법을 적용하여 계통추출법을 통해 이루어졌고, 읍면부에서는 조사구내의 가구수를 기준으로 확률비례추출법을 통해 조사구가 추출되었다. 조사대상 가구의 추출은 선정된 표본조사구내에서 5가구를 계통추출법으로 이루어졌다.

<표 38>에서 읍면부가 없는 시도인 서울, 광주와 대전에서는 읍면부의 추가표본이 없으나 군지역이 있는 부산, 대구, 인천, 울산 중 부산과 인천은 1998년에 구축한 원패널 가구들이 잘 유지되어 추가패널을 고려하지 않고 있다.

## 2. 통합패널의 1차 가중치 산출 결과

한국노동패널조사의 통합패널 가중치 산출 과정은 II장에서 서술된 원패널 가중치 산출 방안과 그 순서와 구조는 기본적으로 동일하다.

통합패널의 경우 원가구의 정의는 표본이 추가된 2009년에 조사된 모든 가구원을 의미하고, 원가구는 조사된 가구원을 포함한 모든 가구로 정의되었다. 통

합패널에서는 2009년 이후 재진입하는 원패널 기준 원가구원은 만 15세가 되어 유입되는 원가구원을 위한 가중치 부여 방안을 적용해서 가구 및 가구원 가중치를 산출하였다.

통합패널은 원패널과 달리 전국 모집단을 대표하므로 원패널에서의 스케일 조정 대신에 인구 및 가구수에 대한 추계자료를 활용한 사후층화가 고려되었다. 추계인구의 지역별, 성별 및 연령대 인구수를, 추계가구의 지역별 가구수를 기준으로 사후조정하여 최종가중치를 산출하게 된다. 이에 대한 자세한 내용은 한국고용정보원(2012)의 연구를 참조할 수 있으며, 또한 본 보고서의 통합패널 세부 가중치의 산출과정에 대한 설명 부분에서 서술되었다.

### 1) 통합패널 가중치 산출을 위한 통합가구와 원가구원 정보

앞서 설명한 바와 같이 통합패널은 2009년 통합패널 구축 시 응답된 원패널의 가구와 가구내 가구원 정보를 원가구원으로 정의한다. 이에 우선 통합패널 1차 자료를 통해 통합패널의 가구와 가구원정보를 살펴보면 다음과 같다.

<표 39> 통합패널 구축시 응답 가구여부

|     | 빈도    | 백분율    |
|-----|-------|--------|
| 미응답 | 1,835 | 21.45  |
| 응답  | 6,721 | 78.55  |
| 전체  | 8,556 | 100.00 |

<표 39>는 통합패널이 구축된 시점의 가구 응답여부를 나타낸 결과이다. 변수명은 hwave12로 응답한 가구는 총 6,721가구로써 이 가구를 통합패널의 원가구로 간주한다. 그리고 6,721가구 내에서 가구원 정보가 응답된 가구원을 통합패널의 원가구원으로 정의한다. 즉, 가구내 가구원은 연령에 상관없이 모든 가구원을 포함하게 되므로 가구 내 가구원의 정보가 있으면 구축시점에서 무응답 가구원이라고 하더라도 이후 만 15세가 되어 새롭게 신규진입하는 조사대상으로 포함하는 것으로 정의한다. 따라서 가구 내 가구원 정보가 존재하는 가구원만을 원가구원으로 정의하였다. <표 40>은 통합패널의 가구의 가구별 원가구원의 기초통계량 결과이다.

<표 40> 통합 패널 가구내 원가구원수 분포

| 관측치수  | 합계     | 평균값  | 표준편차 | 중앙값 | 최소값 | 최대값 |
|-------|--------|------|------|-----|-----|-----|
| 6,721 | 19,436 | 2.89 | 1.34 | 3   | 1   | 10  |

6,721가구내 가구원은 총 19,436명이 가구원이 있으며, 통합패널 가구당 평균 2.89

명의 원가구원이 있는 것으로 나타났다. 따라서 통합패널 구축시점에서 조사 완료된 6,721가구의 19,436명을 통합패널의 원가구와 원가구원으로 정의한다.

## 2) 통합패널 1차 가중치의 사후조정된 가중치 산출 결과

통합패널은 전국 모집단을 대표하기 위해 기존 패널에 추가 패널을 포함하여 2009년에 통합하여 구축되었다. 1998년부터 구축되어 있던 원패널과 2009년에 새롭게 추가된 추가패널로 구성되어 있다. 이를 통합한 구축시점의 설계가중치는 기존 연구보고서인 한국고용정보원(2012)에 설명이 되어 있다.

통합패널은 통합시점인 2009년 모집단을 대표하도록 설계된 것이므로 1차년도로 설정한 통합시점에서의 가중치는 횡단 가중치와 종단 가중치는 동일하게 유지해야 한다. 그러므로 기존 연구에서 통합시점인 2009년에 부여된 개인가중치는 통합패널의 1차년도 가중치를 산출하고, 통합패널의 2차 이후 가중치 산출을 위해 다음과 같이 2009년 기준의 추계인구를 기준으로 통합 1차 개인 가중치를 조정한다.

$$w_{pl,1,i}^* = w_{pl,1,i} \frac{N_{region,gender,age}}{\sum_{region} \sum_{gender} \sum_{age} w_{pl,1,i}} \quad (9)$$

여기서  $w_{pl,1,i}$ 는 기존에 통합패널 구축 시 생성한  $i$ 번째 관측대상에 대한 1차 개인 종단가중치를 의미한다.

앞서 설명한 바와 같이 통합패널 구축 시점인 1차년도에는 개인의 종단가중치와 횡단가중치가 같다고 가정한다. 2009년의 사후조정된  $w_{pl,1,i}^*$ 를 통합 1차 패널의 개인 종단가중치로 시작한다. 또한 통합패널의 가구가중치도 2009년 추계가구로 사후조정한다.

$$w_{1,h}^* = w_{1,h} \frac{N_{region}}{\sum_{region} w_{1,h}} \quad (10)$$

여기서  $w_{1,h}$ 는 기존에 통합패널 구축 시 생성한  $h$ 번째 가구에 대한 1차 가구 가중치를 의미한다.

<표 41>은 통합패널의 1차연도인 2009년 모집단을 대표하는 가중치 산출을 위한 조

사모집단으로 정의된 추계인구 및 가구자료이다.

<표 41> 통합 1차의 사후조정을 위한 추계인구 및 추계가구

| 시도 | 연령대    | 2009년 추계인구 |           | 2009년 추계가구 |
|----|--------|------------|-----------|------------|
|    |        | 남자         | 여자        |            |
| 서울 | 15-29세 | 1,134,672  | 1,130,008 | 3,465,416  |
|    | 30-39세 | 870,128    | 853,326   |            |
|    | 40-49세 | 777,790    | 814,383   |            |
|    | 50-59세 | 634,457    | 694,072   |            |
|    | 60-69세 | 389,212    | 415,970   |            |
|    | 70세 이상 | 226,782    | 361,045   |            |
| 부산 | 15-29세 | 386,336    | 351,028   | 1,232,379  |
|    | 30-39세 | 251,579    | 255,001   |            |
|    | 40-49세 | 277,778    | 300,334   |            |
|    | 50-59세 | 261,681    | 281,879   |            |
|    | 60-69세 | 158,708    | 172,671   |            |
|    | 70세 이상 | 91,679     | 158,927   |            |
| 대구 | 15-29세 | 281,267    | 246,107   | 857,579    |
|    | 30-39세 | 187,989    | 196,312   |            |
|    | 40-49세 | 213,313    | 222,866   |            |
|    | 50-59세 | 163,347    | 171,361   |            |
|    | 60-69세 | 91,839     | 105,508   |            |
|    | 70세 이상 | 59,881     | 104,094   |            |
| 대전 | 15-29세 | 295,474    | 274,763   | 899,685    |
|    | 30-39세 | 224,410    | 219,628   |            |
|    | 40-49세 | 243,257    | 242,984   |            |
|    | 50-59세 | 170,582    | 165,547   |            |
|    | 60-69세 | 84,962     | 91,894    |            |
|    | 70세 이상 | 56,183     | 100,740   |            |
| 인천 | 15-29세 | 184,230    | 166,800   | 504,702    |
|    | 30-39세 | 122,178    | 123,946   |            |
|    | 40-49세 | 125,983    | 127,891   |            |
|    | 50-59세 | 92,168     | 92,121    |            |
|    | 60-69세 | 47,938     | 51,814    |            |
|    | 70세 이상 | 32,882     | 56,595    |            |
| 광주 | 15-29세 | 173,610    | 166,924   | 521,887    |
|    | 30-39세 | 119,430    | 122,282   |            |
|    | 40-49세 | 120,419    | 122,068   |            |
|    | 50-59세 | 83,601     | 86,815    |            |
|    | 60-69세 | 48,710     | 53,961    |            |
|    | 70세 이상 | 32,298     | 57,719    |            |
| 울산 | 15-29세 | 120,952    | 98,578    | 366,726    |
|    | 30-39세 | 92,564     | 91,373    |            |
|    | 40-49세 | 106,193    | 102,866   |            |
|    | 50-59세 | 74,074     | 69,276    |            |
|    | 60-69세 | 32,293     | 32,084    |            |
|    | 70세 이상 | 16,372     | 32,220    |            |

<표 41> 통합 1차의 사후조정을 위한 추계인구 및 추계가구 (계속)

| 시도 | 연령대    | 2009년 추계인구 |           | 2009년 추계가구 |
|----|--------|------------|-----------|------------|
|    |        | 남자         | 여자        |            |
| 경기 | 15-29세 | 1,183,072  | 1,114,801 | 3,730,743  |
|    | 30-39세 | 1,002,607  | 1,003,987 |            |
|    | 40-49세 | 1,025,028  | 986,245   |            |
|    | 50-59세 | 658,845    | 635,633   |            |
|    | 60-69세 | 341,295    | 384,343   |            |
|    | 70세 이상 | 251,753    | 421,740   |            |
| 강원 | 15-29세 | 157,446    | 126,326   | 550,326    |
|    | 30-39세 | 105,256    | 100,933   |            |
|    | 40-49세 | 122,950    | 117,410   |            |
|    | 50-59세 | 103,367    | 103,141   |            |
|    | 60-69세 | 66,201     | 76,563    |            |
|    | 70세 이상 | 60,154     | 99,345    |            |
| 충북 | 15-29세 | 165,129    | 142,755   | 548,078    |
|    | 30-39세 | 112,679    | 108,514   |            |
|    | 40-49세 | 125,568    | 119,369   |            |
|    | 50-59세 | 98,096     | 96,363    |            |
|    | 60-69세 | 59,435     | 67,914    |            |
|    | 70세 이상 | 55,421     | 94,902    |            |
| 충남 | 15-29세 | 217,748    | 187,659   | 731,202    |
|    | 30-39세 | 154,863    | 139,805   |            |
|    | 40-49세 | 159,683    | 143,629   |            |
|    | 50-59세 | 123,264    | 121,170   |            |
|    | 60-69세 | 85,759     | 98,133    |            |
|    | 70세 이상 | 86,983     | 144,218   |            |
| 전북 | 15-29세 | 184,044    | 158,905   | 651,948    |
|    | 30-39세 | 123,038    | 121,500   |            |
|    | 40-49세 | 137,611    | 134,584   |            |
|    | 50-59세 | 114,232    | 117,889   |            |
|    | 60-69세 | 83,076     | 97,936    |            |
|    | 70세 이상 | 76,902     | 137,256   |            |
| 전남 | 15-29세 | 154,505    | 127,976   | 678,409    |
|    | 30-39세 | 115,123    | 109,105   |            |
|    | 40-49세 | 137,693    | 127,216   |            |
|    | 50-59세 | 117,004    | 118,392   |            |
|    | 60-69세 | 93,345     | 112,015   |            |
|    | 70세 이상 | 93,040     | 173,925   |            |
| 경북 | 15-29세 | 272,449    | 222,907   | 992,047    |
|    | 30-39세 | 187,628    | 178,289   |            |
|    | 40-49세 | 208,670    | 199,005   |            |
|    | 50-59세 | 176,490    | 179,782   |            |
|    | 60-69세 | 120,051    | 141,324   |            |
|    | 70세 이상 | 114,844    | 203,653   |            |

<표 41> 통합 1차의 사후조정을 위한 추계인구 및 추계가구 (계속)

| 시도 | 연령대    | 2009년 추계인구 |         | 2009년 추계가구 |
|----|--------|------------|---------|------------|
|    |        | 남자         | 여자      |            |
| 경남 | 15-29세 | 319,853    | 266,152 | 1,132,139  |
|    | 30-39세 | 255,077    | 245,699 |            |
|    | 40-49세 | 274,364    | 262,774 |            |
|    | 50-59세 | 206,718    | 203,915 |            |
|    | 60-69세 | 118,931    | 140,749 |            |
|    | 70세 이상 | 95,119     | 183,846 |            |
| 제주 | 15-29세 | 52,818     | 47,829  | 185,699    |
|    | 30-39세 | 41,872     | 41,288  |            |
|    | 40-49세 | 45,413     | 43,275  |            |
|    | 50-59세 | 31,768     | 31,558  |            |
|    | 60-69세 | 21,135     | 23,540  |            |
|    | 70세 이상 | 15,483     | 34,058  |            |

### 3. 로지스틱 회귀모형을 이용한 응답모형 설정

한국노동패널조사에서는 응답자의 응답모형은 응답여부에 대한 로지스틱 회귀모형을 이용하고 있다. 통합패널에서의 응답모형도 원패널과 동일하므로 기본적인 응답모형 형태는 로지스틱 회귀모형을 이용한 응답모형으로 설정한다.

본 절에서는 통합패널에서 사용되는 로지스틱 회귀모형이 원패널의 응답모형과 다른 점만을 설명한다. 응답모형에서 응답여부에 대한 설명변수로 성별, 지역, 학력, 주된 활동, 만나이, 추가패널여부(통합패널 구축 시 1998년 원패널인지 2009년 추가패널인지 여부)를 분석에 이용한다.

세부적으로는 주된 활동과 학력, 만나이의 경우에는 응답범주를 고려하여 원자료를 일부 통합하여 이용한다. 학력은 석사, 박사인 경우 통합하여 ‘대학원졸’로, 주된 활동은 코드번호가 7, 8, 12, 13, 14, 15, 16, 18, 19, 20을 통합하여 ‘기타’로 변경하여 적용하였다. 만나이는 90세 이상인 경우는 ‘90세이상’으로 통합하여 적용한다. 학력과 주된 활동의 경우 각 차수마다 응답자가 적어서 통합하였고 만나이의 경우 90세 이상의 경우에는 이탈이 많아지는 문제가 존재하기에 응답범주를 통합하여 적용한다.

<표 42>~<표 46>은 통합패널의 차수별 응답자들의 설명변수에 대한 빈도 분포이다. 통합차수별 응답자의 정보 중 1차부터 3차까지만 로지스틱 회귀분석에 이용된다. 이는  $t$  시점의 응답률은  $t-1$  시점의 설명변수와 응답여부를 이용하여 계산하기 때문이다.

<표 42> 통합패널 차수별 응답자의 성별

|    | 통합 1차 |       | 통합 2차 |      | 통합 3차 |       | 통합 4차 |       |
|----|-------|-------|-------|------|-------|-------|-------|-------|
|    | 빈도    | 백분율   | 빈도    | 백분율  | 빈도    | 백분율   | 빈도    | 백분율   |
| 남자 | 6,952 | 47.98 | 6,762 | 47.9 | 6,641 | 47.78 | 6,726 | 48.04 |
| 여자 | 7,537 | 52.02 | 7,356 | 52.1 | 7,259 | 52.22 | 7,274 | 51.96 |

<표 43> 통합패널 차수별 응답자의 거주지역

|    | 통합 1차 |       | 통합 2차 |       | 통합 3차 |       | 통합 4차 |       |
|----|-------|-------|-------|-------|-------|-------|-------|-------|
|    | 빈도    | 백분율   | 빈도    | 백분율   | 빈도    | 백분율   | 빈도    | 백분율   |
| 서울 | 2,900 | 20.02 | 2,838 | 20.1  | 2,736 | 19.68 | 2,648 | 18.91 |
| 부산 | 1,130 | 7.80  | 1,109 | 7.86  | 1,082 | 7.78  | 1,082 | 7.73  |
| 대구 | 714   | 4.93  | 702   | 4.97  | 658   | 4.73  | 677   | 4.84  |
| 인천 | 413   | 2.85  | 412   | 2.92  | 388   | 2.79  | 382   | 2.73  |
| 광주 | 850   | 5.87  | 827   | 5.86  | 822   | 5.91  | 830   | 5.93  |
| 대전 | 342   | 2.36  | 349   | 2.47  | 319   | 2.29  | 340   | 2.43  |
| 울산 | 385   | 2.66  | 382   | 2.71  | 393   | 2.83  | 397   | 2.84  |
| 경기 | 3,257 | 22.48 | 3,043 | 21.55 | 3,191 | 22.96 | 3,249 | 23.21 |
| 강원 | 436   | 3.01  | 398   | 2.82  | 399   | 2.87  | 408   | 2.91  |
| 충북 | 470   | 3.24  | 487   | 3.45  | 476   | 3.42  | 470   | 3.36  |
| 충남 | 578   | 3.99  | 601   | 4.26  | 600   | 4.32  | 591   | 4.22  |
| 전북 | 540   | 3.73  | 529   | 3.75  | 516   | 3.71  | 529   | 3.78  |
| 전남 | 567   | 3.91  | 561   | 3.97  | 523   | 3.76  | 517   | 3.69  |
| 경북 | 782   | 5.40  | 771   | 5.46  | 672   | 4.83  | 707   | 5.05  |
| 경남 | 1,016 | 7.01  | 1,012 | 7.17  | 1,014 | 7.29  | 1,045 | 7.46  |
| 제주 | 109   | 0.75  | 97    | 0.69  | 111   | 0.80  | 128   | 0.91  |

<표 44> 통합패널 차수별 응답자의 학력

|     | 통합 1차 |       | 통합 2차 |       | 통합 3차 |       | 통합 4차 |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
|     | 빈도    | 백분율   | 빈도    | 백분율   | 빈도    | 백분율   | 빈도    | 백분율   |
| 무응답 | 2     | 0.01  | 4     | 0.03  | 3     | 0.02  | 4     | 0.03  |
| 2   | 811   | 5.60  | 744   | 5.27  | 698   | 5.02  | 663   | 4.74  |
| 3   | 1,944 | 13.42 | 1,886 | 13.36 | 1,831 | 13.17 | 1,788 | 12.77 |
| 4   | 1,719 | 11.86 | 1,675 | 11.86 | 1,633 | 11.75 | 1,638 | 11.70 |
| 5   | 5,007 | 34.56 | 4,838 | 34.27 | 4,706 | 33.86 | 4,700 | 33.57 |
| 6   | 1,744 | 12.04 | 1,728 | 12.24 | 1,722 | 12.39 | 1,777 | 12.69 |
| 7   | 2,833 | 19.55 | 2,842 | 20.13 | 2,889 | 20.78 | 2,987 | 21.34 |
| 8   | 361   | 2.49  | 341   | 2.42  | 350   | 2.52  | 365   | 2.61  |
| 9   | 68    | 0.47  | 60    | 0.42  | 68    | 0.49  | 78    | 0.56  |



<표 45> 통합패널 차수별 응답자의 주된 활동

|     | 통합 1차 |       | 통합 2차 |       | 통합 3차 |       | 통합 4차 |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
|     | 빈도    | 백분율   | 빈도    | 백분율   | 빈도    | 백분율   | 빈도    | 백분율   |
| 무응답 | -     | -     | 2     | 0.01  | 4     | 0.03  | -     | -     |
| 1   | 7583  | 52.34 | 7699  | 54.53 | 7670  | 55.18 | 7,820 | 55.86 |
| 2   | 76    | 0.52  | 45    | 0.32  | 37    | 0.27  | 37    | 0.26  |
| 3   | 122   | 0.84  | 78    | 0.55  | 94    | 0.68  | 99    | 0.71  |
| 4   | 2654  | 18.32 | 2382  | 16.87 | 2390  | 17.19 | 2,443 | 17.45 |
| 5   | 316   | 2.18  | 405   | 2.87  | 369   | 2.65  | 369   | 2.64  |
| 6   | 1435  | 9.90  | 1324  | 9.38  | 1235  | 8.88  | 1,222 | 8.73  |
| 7   | 30    | 0.21  | 31    | 0.22  | 30    | 0.22  | 28    | 0.20  |
| 8   | 47    | 0.32  | 53    | 0.38  | 43    | 0.31  | 48    | 0.34  |
| 9   | 683   | 4.71  | 616   | 4.36  | 591   | 4.25  | 533   | 3.81  |
| 10  | 290   | 2.00  | 281   | 1.99  | 279   | 2.01  | 275   | 1.96  |
| 11  | 126   | 0.87  | 97    | 0.69  | 94    | 0.68  | 109   | 0.78  |
| 12  | 41    | 0.28  | 32    | 0.23  | 36    | 0.26  | 33    | 0.24  |
| 13  | 25    | 0.17  | 20    | 0.14  | 25    | 0.18  | 23    | 0.16  |
| 14  | 10    | 0.07  | 4     | 0.03  | 3     | 0.02  | 5     | 0.04  |
| 15  | 9     | 0.06  | 9     | 0.06  | 10    | 0.07  | 9     | 0.06  |
| 17  | 992   | 6.85  | 1005  | 7.12  | 976   | 7.02  | 1     | 0.01  |
| 18  | 45    | 0.31  | 9     | 0.06  | 10    | 0.07  | 922   | 6.59  |
| 19  | -     | -     | 19    | 0.13  | -     | -     | 16    | 0.11  |
| 20  | 5     | 0.03  | 7     | 0.05  | 4     | 0.03  | -     | -     |

<표 46> 통합패널 차수별 응답자의 만나이

|     | 통합 1차 |      | 통합 2차 |      | 통합 3차 |      | 통합 4차 |      |
|-----|-------|------|-------|------|-------|------|-------|------|
|     | 빈도    | 백분율  | 빈도    | 백분율  | 빈도    | 백분율  | 빈도    | 백분율  |
| 무응답 | -     | -    | -     | -    | 1     | 0.01 | 1     | 0.01 |
| 15  | 210   | 1.45 | 200   | 1.42 | 204   | 1.47 | 195   | 1.39 |
| 16  | 238   | 1.64 | 207   | 1.47 | 197   | 1.42 | 196   | 1.40 |
| 17  | 231   | 1.59 | 228   | 1.61 | 199   | 1.43 | 191   | 1.36 |
| 18  | 227   | 1.57 | 210   | 1.49 | 216   | 1.55 | 185   | 1.32 |
| 19  | 191   | 1.32 | 181   | 1.28 | 178   | 1.28 | 181   | 1.29 |
| 20  | 149   | 1.03 | 137   | 0.97 | 133   | 0.96 | 128   | 0.91 |
| 21  | 106   | 0.73 | 120   | 0.85 | 123   | 0.88 | 128   | 0.91 |
| 22  | 123   | 0.85 | 123   | 0.87 | 134   | 0.96 | 146   | 1.04 |
| 23  | 147   | 1.01 | 144   | 1.02 | 127   | 0.91 | 163   | 1.16 |
| 24  | 176   | 1.21 | 156   | 1.10 | 140   | 1.01 | 123   | 0.88 |
| 25  | 196   | 1.35 | 162   | 1.15 | 149   | 1.07 | 135   | 0.96 |
| 26  | 201   | 1.39 | 179   | 1.27 | 153   | 1.10 | 142   | 1.01 |

<표 46> 통합패널 차수별 응답자의 만나이 (계속)

|    | 통합 1차 |      | 통합 2차 |      | 통합 3차 |      | 통합 4차 |      |
|----|-------|------|-------|------|-------|------|-------|------|
|    | 빈도    | 백분율  | 빈도    | 백분율  | 빈도    | 백분율  | 빈도    | 백분율  |
| 27 | 229   | 1.58 | 213   | 1.51 | 168   | 1.21 | 166   | 1.19 |
| 28 | 239   | 1.65 | 216   | 1.53 | 207   | 1.49 | 170   | 1.21 |
| 29 | 314   | 2.17 | 240   | 1.70 | 218   | 1.57 | 199   | 1.42 |
| 30 | 295   | 2.04 | 305   | 2.16 | 241   | 1.73 | 218   | 1.56 |
| 31 | 284   | 1.96 | 294   | 2.08 | 318   | 2.29 | 257   | 1.84 |
| 32 | 318   | 2.19 | 277   | 1.96 | 287   | 2.06 | 315   | 2.25 |
| 33 | 274   | 1.89 | 286   | 2.03 | 282   | 2.03 | 286   | 2.04 |
| 34 | 282   | 1.95 | 273   | 1.93 | 277   | 1.99 | 277   | 1.98 |
| 35 | 348   | 2.40 | 276   | 1.95 | 264   | 1.90 | 286   | 2.04 |
| 36 | 286   | 1.97 | 343   | 2.43 | 283   | 2.04 | 275   | 1.96 |
| 37 | 328   | 2.26 | 279   | 1.98 | 343   | 2.47 | 280   | 2.00 |
| 38 | 322   | 2.22 | 320   | 2.27 | 274   | 1.97 | 338   | 2.41 |
| 39 | 299   | 2.06 | 301   | 2.13 | 321   | 2.31 | 275   | 1.96 |
| 40 | 300   | 2.07 | 278   | 1.97 | 291   | 2.09 | 313   | 2.24 |
| 41 | 306   | 2.11 | 290   | 2.05 | 266   | 1.91 | 289   | 2.06 |
| 42 | 229   | 1.58 | 285   | 2.02 | 285   | 2.05 | 265   | 1.89 |
| 43 | 253   | 1.75 | 229   | 1.62 | 279   | 2.01 | 284   | 2.03 |
| 44 | 260   | 1.79 | 250   | 1.77 | 222   | 1.60 | 279   | 1.99 |
| 45 | 263   | 1.82 | 247   | 1.75 | 239   | 1.72 | 221   | 1.58 |
| 46 | 267   | 1.84 | 248   | 1.76 | 241   | 1.73 | 246   | 1.76 |
| 47 | 254   | 1.75 | 250   | 1.77 | 239   | 1.72 | 230   | 1.64 |
| 48 | 293   | 2.02 | 243   | 1.72 | 237   | 1.71 | 233   | 1.66 |
| 49 | 285   | 1.97 | 276   | 1.95 | 247   | 1.78 | 239   | 1.71 |
| 50 | 271   | 1.87 | 272   | 1.93 | 271   | 1.95 | 257   | 1.84 |
| 51 | 263   | 1.82 | 267   | 1.89 | 260   | 1.87 | 278   | 1.99 |
| 52 | 240   | 1.66 | 257   | 1.82 | 259   | 1.86 | 264   | 1.89 |
| 53 | 236   | 1.63 | 237   | 1.68 | 249   | 1.79 | 257   | 1.84 |
| 54 | 259   | 1.79 | 237   | 1.68 | 223   | 1.60 | 258   | 1.84 |
| 55 | 225   | 1.55 | 250   | 1.77 | 226   | 1.63 | 235   | 1.68 |
| 56 | 183   | 1.26 | 227   | 1.61 | 246   | 1.77 | 238   | 1.70 |
| 57 | 258   | 1.78 | 181   | 1.28 | 214   | 1.54 | 249   | 1.78 |
| 58 | 164   | 1.13 | 250   | 1.77 | 175   | 1.26 | 223   | 1.59 |
| 59 | 206   | 1.42 | 156   | 1.10 | 246   | 1.77 | 182   | 1.30 |
| 60 | 199   | 1.37 | 201   | 1.42 | 151   | 1.09 | 244   | 1.74 |
| 61 | 208   | 1.44 | 201   | 1.42 | 197   | 1.42 | 155   | 1.11 |
| 62 | 203   | 1.40 | 209   | 1.48 | 201   | 1.45 | 200   | 1.43 |
| 63 | 181   | 1.25 | 200   | 1.42 | 205   | 1.47 | 199   | 1.42 |
| 64 | 148   | 1.02 | 183   | 1.30 | 194   | 1.40 | 206   | 1.47 |

<표 46> 통합패널 차수별 응답자의 만나이 (계속)

|     | 통합 1차 |      | 통합 2차 |      | 통합 3차 |      | 통합 4차 |      |
|-----|-------|------|-------|------|-------|------|-------|------|
|     | 빈도    | 백분율  | 빈도    | 백분율  | 빈도    | 백분율  | 빈도    | 백분율  |
| 65  | 147   | 1.01 | 146   | 1.03 | 178   | 1.28 | 193   | 1.38 |
| 66  | 171   | 1.18 | 148   | 1.05 | 141   | 1.01 | 176   | 1.26 |
| 67  | 184   | 1.27 | 166   | 1.18 | 144   | 1.04 | 142   | 1.01 |
| 68  | 157   | 1.08 | 178   | 1.26 | 166   | 1.19 | 144   | 1.03 |
| 69  | 167   | 1.15 | 153   | 1.08 | 177   | 1.27 | 165   | 1.18 |
| 70  | 197   | 1.36 | 158   | 1.12 | 151   | 1.09 | 175   | 1.25 |
| 71  | 166   | 1.15 | 188   | 1.33 | 155   | 1.12 | 149   | 1.06 |
| 72  | 145   | 1.00 | 163   | 1.15 | 183   | 1.32 | 153   | 1.09 |
| 73  | 154   | 1.06 | 137   | 0.97 | 152   | 1.09 | 181   | 1.29 |
| 74  | 117   | 0.81 | 152   | 1.08 | 131   | 0.94 | 150   | 1.07 |
| 75  | 101   | 0.70 | 117   | 0.83 | 146   | 1.05 | 128   | 0.91 |
| 76  | 101   | 0.70 | 93    | 0.66 | 117   | 0.84 | 140   | 1.00 |
| 77  | 110   | 0.76 | 96    | 0.68 | 91    | 0.65 | 112   | 0.80 |
| 78  | 86    | 0.59 | 94    | 0.67 | 98    | 0.71 | 88    | 0.63 |
| 79  | 75    | 0.52 | 85    | 0.60 | 93    | 0.67 | 94    | 0.67 |
| 80  | 74    | 0.51 | 69    | 0.49 | 76    | 0.55 | 88    | 0.63 |
| 81  | 54    | 0.37 | 64    | 0.45 | 61    | 0.44 | 71    | 0.51 |
| 82  | 52    | 0.36 | 47    | 0.33 | 67    | 0.48 | 58    | 0.41 |
| 83  | 37    | 0.26 | 42    | 0.30 | 48    | 0.35 | 64    | 0.46 |
| 84  | 45    | 0.31 | 29    | 0.21 | 38    | 0.27 | 42    | 0.3  |
| 85  | 35    | 0.24 | 38    | 0.27 | 29    | 0.21 | 40    | 0.29 |
| 86  | 30    | 0.21 | 33    | 0.23 | 34    | 0.24 | 21    | 0.15 |
| 87  | 30    | 0.21 | 27    | 0.19 | 27    | 0.19 | 30    | 0.21 |
| 88  | 21    | 0.14 | 27    | 0.19 | 27    | 0.19 | 20    | 0.14 |
| 89  | 15    | 0.10 | 21    | 0.15 | 23    | 0.17 | 23    | 0.16 |
| 90  | 11    | 0.08 | 12    | 0.08 | 12    | 0.09 | 20    | 0.14 |
| 91  | 11    | 0.08 | 10    | 0.07 | 7     | 0.05 | 8     | 0.06 |
| 92  | 7     | 0.05 | 8     | 0.06 | 6     | 0.04 | 4     | 0.03 |
| 93  | 4     | 0.03 | 6     | 0.04 | 6     | 0.04 | 6     | 0.04 |
| 94  | 6     | 0.04 | 3     | 0.02 | 5     | 0.04 | 5     | 0.04 |
| 95  | 2     | 0.01 | 3     | 0.02 | 3     | 0.02 | 2     | 0.01 |
| 96  | 4     | 0.03 | -     | -    | 2     | 0.01 | 2     | 0.01 |
| 97  | 2     | 0.01 | 4     | 0.03 | 1     | 0.01 | 3     | 0.02 |
| 98  | 3     | 0.02 | 3     | 0.02 | 2     | 0.01 | -     | -    |
| 99  | 1     | 0.01 | 4     | 0.03 | 1     | 0.01 | 1     | 0.01 |
| 100 | -     | -    | -     | -    | 2     | 0.01 | 1     | 0.01 |
| 101 | -     | -    | -     | -    | -     | -    | 1     | 0.01 |

추가해 설명하면, 응답자의 학력, 주된 활동, 만나이에 대한 응답은 일부 통합패널 차수에서 무응답이 존재한다는 점이다. 응답모형 설정 과정에서 설명변수의 무응답이 존재하는 경우에는 무응답이 속한 다른 설명변수의 평균 응답률로 대체하였다. 예를 들면 학력이 무응답인 경우, 무응답이 속한 성별, 거주지역, 주된 활동, 만나이, 추가패널여부의 응답률의 평균을 응답률로 사용한다는 것이다.

<표 47> 통합패널 2차년도 로지스틱 회귀분석 결과

| 변수                             | 자유도 | 카이제곱         | 유의확률               |
|--------------------------------|-----|--------------|--------------------|
| 성별                             | 1   | 24.8259      | <.0001             |
| 지역                             | 15  | 120.3858     | <.0001             |
| 학력                             | 6   | 36.3657      | <.0001             |
| 주된 활동                          | 10  | 100.0739     | <.0001             |
| 만 나이                           | 75  | 250.2016     | <.0001             |
| 추가패널여부                         | 1   | 30.8440      | <.0001             |
| likelihood $\chi^2 = 900.5624$ |     | $n = 14,487$ | $p - value <.0001$ |

<표 48> 통합패널 3차년도 로지스틱 회귀분석 결과

| 변수                             | 자유도 | 카이제곱         | 유의확률               |
|--------------------------------|-----|--------------|--------------------|
| 성별                             | 1   | 23.3659      | <.0001             |
| 지역                             | 15  | 109.9560     | <.0001             |
| 학력                             | 6   | 6.0200       | <.0001             |
| 주된 활동                          | 10  | 77.5391      | <.0001             |
| 만 나이                           | 75  | 289.4463     | <.0001             |
| 추가패널여부                         | 1   | 0.0019       | 0.9656             |
| likelihood $\chi^2 = 762.6605$ |     | $n = 14,112$ | $p - value <.0001$ |

<표 49> 통합패널 4차년도 로지스틱 회귀분석 결과

| 변수                             | 자유도 | 카이제곱         | 유의확률               |
|--------------------------------|-----|--------------|--------------------|
| 성별                             | 1   | 12.3168      | 0.0004             |
| 지역                             | 15  | 58.2209      | <.0001             |
| 학력                             | 6   | 14.3828      | 0.0256             |
| 주된 활동                          | 10  | 73.0216      | <.0001             |
| 만 나이                           | 75  | 298.7603     | <.0001             |
| 추가패널여부                         | 1   | 0.3831       | 0.5359             |
| likelihood $\chi^2 = 853.7026$ |     | $n = 13,892$ | $p - value <.0001$ |

<표 47>에서 <표 49>까지는 통합패널의 차수별 로지스틱 회귀분석 결과이다. 실제로 통합 3차부터는 추가패널여부가 통계적으로 유의하지 않는 것으로 나타난다. 하지만 2

차의 경우에는 매우 유의함을 확인할 수 있었다. 그러나 초기 패널의 진행 상태를 유지하고자 추가패널여부를 설명변수로 사용하였음을 밝혀둔다.

#### 4. 통합패널 가중치 산출과정

통합패널의 가중치 산출은 기본적으로 3단계(1단계 : 종단면 가구원 가중치 산출 → 2단계 : 가구 가중치 산출 → 3단계 : 횡단면 가구원 가중치 산출)의 과정으로 구성된다. 가중치 산출 과정을 단계별로 살펴보면 다음과 같다.

##### 1단계 : 종단면 가구원 가중치 산출

한국노동패널조사의 종단면 가구원 가중치 산출은 기본적으로 무응답 보정의 과정을 통해 이루어진다. 무응답 보정을 위해  $t$ 차년도의 각 가구원의 응답여부를 종속변수로,  $t-1$ 차 년도의 변수(성별, 연령, 지역, 교육수준, 주된 활동, 추가패널여부)을 설명변수로 하는 로지스틱 회귀분석을 실시하고, 이를 통하여 무응답 보정 가중치를 산출하였다. 이 과정에서 생성된 가중치를 기초가중치로 정의한다.

- 1)  $t-1$ 년에 조사가 되고  $t$ 차에 조사된 가구원은 다음과 같이 방식으로  $t$ 차 기초 가중치를 계산한다.

$$w_{base,t,i} = \hat{p}_{ti}^{-1} w_{pl,t-1,i}^*, \hat{p}_{ti} = [1 + \exp(\mathbf{x}'_{t-1,i} \hat{\boldsymbol{\beta}})]$$

- 2) 원가구원이면서 기존에 재조사가 된 개인은 최근에 조사된 조사시점의 종단면 가구원 가중치를 기초가중치로 부여한다.
- 3) 비원가구원에게는 “0”의 기초가중치를 부여한다.
- 4) 신규 진입된 원가구원(만15세, 16세 이상)은 기초가중치 정의에서 제외한다. 즉, 기초가중치는 미부여한다.
- 5) 원가구원이나 통합 1차 이전(12차 이전)에 분가하였고, 통합 1차에 조사가 되지 않은 가구 및 가구원이 존재한다. 원가구원 여부에는 포함되나, 다른 가구를 구성하고 통합 1차에 조사되지 않는 가구 및 가구원은 결국 통합 1차 관점에서는 조사대상 가구가 아닌 것으로 정의하였다. 이에 이러한 경우에는 결측으로 처리한다.

위에서 정의된 기초 가중치를 이용하여 생성되는 중단면 가구원 가중값은 다음과 같다.

- i) 원가구원으로 연속적으로 조사된 경우와 통합 1차 이후 조사되었다가 재조사된 경우에는 위에서 산출한 기초가중치를 중단면 가중치로 부여한다. 즉,

$$w_{pl,t,i} = w_{base,t,i}$$

- ii) 원가구원이면서 신규진입 원가구원에게는 사후조정된 가구 가중치( $w_{h,t,i}$ )를 중단면 가중치로 부여한다. 이는 이후에 설명할 가구 가중치의 식 (12)를 참조하면 된다.

$$w_{pl,t,i} = w_{h,t}$$

- iii) i)과 ii)를 통해 산출된 중단면 가중치( $w_{pl,t,i}$ )를 최종적으로 통합 년도 기준 인구주택총조사 자료 혹은 추계인구를 이용하여 다음과 같이 사후조정하며 이를 최종 개인 중단면 가중값으로 정의한다.

$$w_{pl,t,i}^* = w_{pl,t,i} \frac{N_{region,gender,age}}{\sum_{region} \sum_{gender} \sum_{age} w_{pl,t,i}} \quad (11)$$

## 2단계 : 가구 가중치 산출

가구 가중치 역시 1단계에서 계산된 기초가중치를 이용하여 산출한다. 가구 가중치는 기본적으로 1단계에서 산출된 기초 가중치들의 각 가구별 평균으로 정의된다. 즉 가구 가중치는 다음의 식 (12)로 정의되며,  $n_h$ 는 가구  $h$ 의 가구원 수를 의미한다. 가구원 수는 가구 내 결측값, 0 혹은 양의 기초가중치를 부여 받은 가구원의 수를 나타낸다.

$$w_{t,h} = \frac{\sum_{i \in h} w_{base,t,i}}{n_h} \quad (12)$$

식 (12)의 가중치에 인구주택총조사 자료의 가구수 및 추계가구수 정보를 이용하여

사후조정된 가구 가중치를 최종적으로 다음과 같이 계산한다. 여기서,  $N_{region}$ 은 16개 시도별 가구수를 의미한다.

$$w_{t,h}^* = w_{t,h} \frac{N_{region}}{\sum_{region} w_{t,h}} \quad (13)$$

### 3단계 : 횡단면 가구원 가중치 산출

통합패널에서도 횡단면 가구원 가중치는  $t$ 시점에 원가구원을 한 명이상 포함하고 있는 가구에 속하는 모든 가구원에게 부여된다. 횡단면 가구원 가중치 산출을 위해서는 먼저 사후조정전의 가구가중치를 각 가구원에게 부여한다. 즉 사후조정 전 가구  $h$ 에 거주하는 가구원  $i$ 에게 부여되는 횡단면 가중치는 다음과 같으며 여기서  $w_{t,h}$ 는 2단계에서 사용된 가구 가중치이다.

$$w_{pc,t,i} = w_{t,h} \quad (14)$$

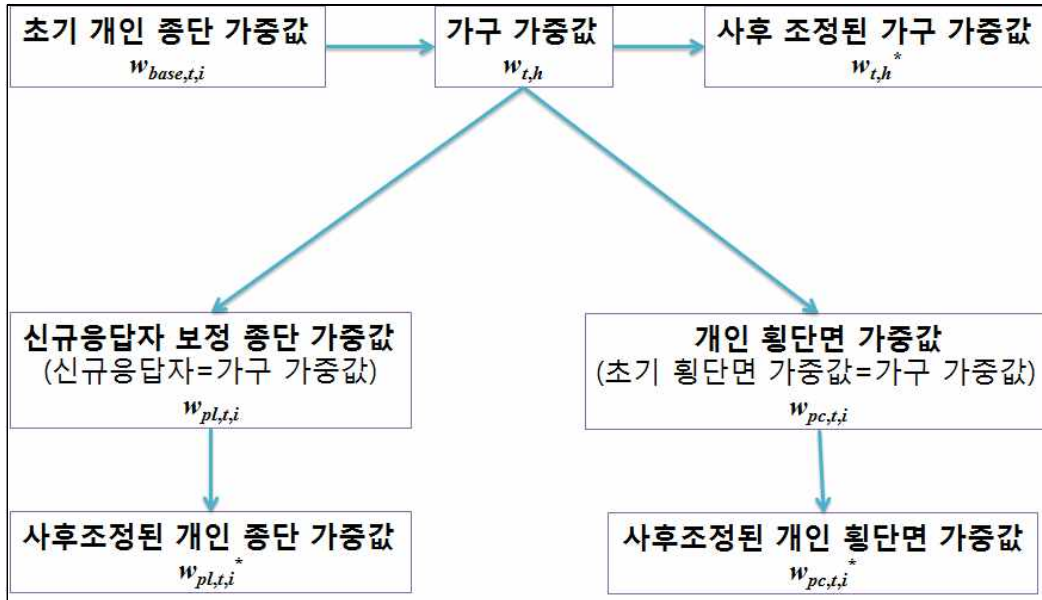
최종 횡단면 가구원 가중치는 위의 가중치 (14)에 사후조정을 적용하여 산출된다. 사후조정은 통합 2차(2010년), 통합 3차(2011년), 통합 4차(2012년)의 횡단면 가구원 가중치에 적용되는데, 사후조정의 경우 2010년에는 인구주택총조사의 인구수이며, 2011년과 2012년의 경우에는 각 년도의 추계인구를 기준으로 한다. 사후조정에 사용되는 항목은 시도, 성별, 연령대별 인구수이다. 이에 사후조정된 최종 횡단면 가구원 가중값은 다음과 같이 계산된다. 여기서  $N$ 은 인구주택총조사의 인구 혹은 통계청에서 공표한 추계인구를 의미한다.

$$w_{pc,t,i}^* = w_{pc,t,i} \frac{N_{region,gender,age}}{\sum_{region} \sum_{gender} \sum_{age} w_{pc,t,i}} \quad (15)$$

통합패널에서의 가중치 산출 과정을 요약해 표현하면 <그림2>와 같다. <그림2>를 보면, 초기 개인 가중치를 이용해 가구 가중치, 신규응답자 보정 및 개인 횡단면 가중치 등의 과정을 통해 최종 가중치 산출하는 과정을 그림으로 요약한 것이다.

그리고 <표 50>과 <표 51>은 사후조정에 이용된 인구주택총조사 혹은 추계인구수 및 가구수 현황 자료이다.

<그림 2> 통합패널 가중치 산출과정



<표 50> 통합 2~4차의 사후조정을 위한 인구주택총조사 인구 및 추계인구

| 시도 | 연령대    | 2010년 인구주택총조사 |         | 2011년 추계인구 |         | 2012년 추계인구 |         |
|----|--------|---------------|---------|------------|---------|------------|---------|
|    |        | 남자            | 여자      | 남자         | 여자      | 남자         | 여자      |
| 서울 | 15-19세 | 332,292       | 302,945 | 331,937    | 300,555 | 321,683    | 290,685 |
|    | 20-29세 | 778,790       | 804,136 | 806,776    | 830,159 | 787,703    | 808,662 |
|    | 30-39세 | 860,423       | 848,071 | 898,959    | 874,159 | 885,118    | 866,717 |
|    | 40-49세 | 774,354       | 807,473 | 829,150    | 841,471 | 823,430    | 834,599 |
|    | 50-59세 | 652,012       | 717,891 | 713,734    | 772,994 | 731,849    | 794,267 |
|    | 60-69세 | 400,422       | 430,680 | 418,522    | 450,746 | 424,817    | 461,087 |
|    | 70세 이상 | 227,582       | 339,108 | 248,167    | 356,710 | 267,302    | 375,823 |
| 부산 | 15-19세 | 130,771       | 111,643 | 129,050    | 110,475 | 123,735    | 106,225 |
|    | 20-29세 | 246,342       | 228,629 | 255,688    | 235,240 | 251,639    | 228,559 |
|    | 30-39세 | 247,178       | 249,618 | 256,175    | 252,138 | 251,575    | 247,484 |
|    | 40-49세 | 271,662       | 293,627 | 277,940    | 292,136 | 271,727    | 285,138 |
|    | 50-59세 | 269,146       | 291,988 | 282,032    | 301,912 | 287,403    | 310,114 |
|    | 60-69세 | 163,968       | 178,151 | 168,521    | 183,637 | 172,993    | 189,281 |
|    | 70세 이상 | 94,001        | 153,916 | 101,116    | 160,955 | 108,504    | 169,969 |
| 대구 | 15-19세 | 101,944       | 84,008  | 101,505    | 84,510  | 99,076     | 83,642  |
|    | 20-29세 | 175,492       | 156,414 | 181,852    | 160,096 | 180,792    | 155,466 |
|    | 30-39세 | 183,830       | 192,065 | 188,095    | 192,543 | 184,114    | 188,664 |
|    | 40-49세 | 212,981       | 223,628 | 218,115    | 223,843 | 215,182    | 222,105 |
|    | 50-59세 | 169,965       | 178,573 | 180,242    | 186,064 | 188,454    | 194,497 |
|    | 60-69세 | 94,539        | 108,123 | 97,278     | 110,850 | 100,245    | 113,895 |
|    | 70세 이상 | 60,958        | 101,050 | 65,112     | 105,542 | 69,649     | 111,748 |



<표 50> 통합 2~4차의 사후조정을 위한 인구주택총조사 인구 및 추계인구 (계속)

| 시도 | 연령대    | 2010년 인구주택총조사 |           | 2011년 추계인구 |           | 2012년 추계인구 |           |
|----|--------|---------------|-----------|------------|-----------|------------|-----------|
|    |        | 남자            | 여자        | 남자         | 여자        | 남자         | 여자        |
| 대전 | 15-19세 | 104,048       | 94,156    | 105,511    | 95,794    | 104,691    | 95,080    |
|    | 20-29세 | 192,755       | 180,949   | 204,791    | 187,177   | 204,206    | 186,035   |
|    | 30-39세 | 223,375       | 216,862   | 240,505    | 223,157   | 242,851    | 225,116   |
|    | 40-49세 | 243,532       | 245,466   | 257,277    | 250,948   | 257,297    | 251,437   |
|    | 50-59세 | 181,289       | 176,567   | 200,200    | 191,540   | 214,642    | 206,892   |
|    | 60-69세 | 88,771        | 94,825    | 93,367     | 98,457    | 97,944     | 102,165   |
|    | 70세 이상 | 56,012        | 96,244    | 60,892     | 101,044   | 65,382     | 107,183   |
| 인천 | 15-19세 | 62,759        | 57,851    | 63,453     | 57,772    | 62,724     | 56,886    |
|    | 20-29세 | 109,719       | 107,395   | 114,694    | 111,935   | 113,823    | 110,321   |
|    | 30-39세 | 118,965       | 121,893   | 123,319    | 123,557   | 122,924    | 122,886   |
|    | 40-49세 | 122,294       | 124,007   | 127,102    | 126,010   | 127,866    | 127,302   |
|    | 50-59세 | 87,523        | 91,069    | 93,667     | 95,458    | 98,786     | 100,730   |
|    | 60-69세 | 50,673        | 56,123    | 52,497     | 57,814    | 54,106     | 59,299    |
|    | 70세 이상 | 32,519        | 54,166    | 34,967     | 56,616    | 37,564     | 59,869    |
| 광주 | 15-19세 | 63,941        | 56,267    | 62,193     | 54,682    | 59,852     | 52,819    |
|    | 20-29세 | 120,681       | 109,387   | 127,809    | 115,543   | 129,013    | 115,495   |
|    | 30-39세 | 121,385       | 122,960   | 124,423    | 124,166   | 124,267    | 124,100   |
|    | 40-49세 | 126,730       | 128,882   | 130,485    | 129,586   | 130,576    | 130,018   |
|    | 50-59세 | 96,757        | 97,293    | 103,657    | 102,428   | 109,381    | 108,554   |
|    | 60-69세 | 49,838        | 53,471    | 51,509     | 54,675    | 53,656     | 56,674    |
|    | 70세 이상 | 32,643        | 53,268    | 34,932     | 55,357    | 37,311     | 58,398    |
| 울산 | 15-19세 | 46,769        | 36,632    | 47,775     | 38,447    | 48,019     | 39,885    |
|    | 20-29세 | 74,456        | 60,864    | 76,952     | 61,017    | 76,460     | 58,177    |
|    | 30-39세 | 91,353        | 89,115    | 96,563     | 89,533    | 97,344     | 88,980    |
|    | 40-49세 | 106,739       | 104,283   | 111,136    | 105,481   | 110,014    | 105,452   |
|    | 50-59세 | 78,406        | 74,005    | 85,599     | 78,683    | 91,268     | 84,077    |
|    | 60-69세 | 34,120        | 33,545    | 36,450     | 35,331    | 38,956     | 37,388    |
|    | 70세 이상 | 16,800        | 30,671    | 18,174     | 31,864    | 19,576     | 33,699    |
| 경기 | 15-19세 | 426,147       | 389,513   | 436,489    | 398,160   | 437,036    | 398,336   |
|    | 20-29세 | 773,613       | 733,069   | 832,340    | 763,876   | 829,101    | 758,288   |
|    | 30-39세 | 999,429       | 999,283   | 1,074,288  | 1,029,481 | 1,065,844  | 1,020,512 |
|    | 40-49세 | 1,041,049     | 1,011,356 | 1,122,023  | 1,058,086 | 1,134,173  | 1,074,551 |
|    | 50-59세 | 703,355       | 678,753   | 789,248    | 746,686   | 845,958    | 803,657   |
|    | 60-69세 | 354,097       | 395,987   | 374,124    | 412,747   | 390,350    | 425,360   |
|    | 70세 이상 | 253,139       | 405,247   | 274,778    | 427,351   | 293,698    | 453,303   |
| 강원 | 15-19세 | 55,000        | 47,384    | 54,717     | 47,782    | 53,135     | 47,184    |
|    | 20-29세 | 101,911       | 76,574    | 107,709    | 81,005    | 110,223    | 80,232    |
|    | 30-39세 | 102,717       | 98,844    | 104,264    | 98,507    | 101,726    | 96,299    |
|    | 40-49세 | 122,092       | 116,562   | 125,232    | 116,317   | 124,289    | 115,241   |
|    | 50-59세 | 108,693       | 108,344   | 116,876    | 113,642   | 122,206    | 118,862   |
|    | 60-69세 | 65,894        | 75,817    | 66,351     | 75,651    | 68,120     | 76,566    |
|    | 70세 이상 | 59,668        | 94,681    | 63,397     | 98,173    | 66,645     | 103,014   |
| 충북 | 15-19세 | 59,054        | 51,139    | 58,284     | 50,935    | 57,049     | 50,511    |
|    | 20-29세 | 105,541       | 90,570    | 113,096    | 95,476    | 113,225    | 93,826    |
|    | 30-39세 | 111,703       | 107,247   | 117,545    | 108,830   | 116,991    | 108,132   |
|    | 40-49세 | 126,183       | 120,361   | 130,687    | 121,294   | 130,469    | 121,325   |
|    | 50-59세 | 103,292       | 101,500   | 111,613    | 106,929   | 117,662    | 112,376   |
|    | 60-69세 | 59,696        | 67,389    | 61,016     | 68,013    | 63,312     | 69,246    |
|    | 70세 이상 | 54,928        | 90,479    | 57,725     | 93,543    | 60,377     | 97,875    |

<표 50> 통합 2~4차의 사후조정을 위한 인구주택총조사 인구 및 추계인구 (계속)

| 시도 | 연령대    | 2010년 인구주택총조사 |         | 2011년 추계인구 |         | 2012년 추계인구 |         |
|----|--------|---------------|---------|------------|---------|------------|---------|
|    |        | 남자            | 여자      | 남자         | 여자      | 남자         | 여자      |
| 충남 | 15-19세 | 75,768        | 69,256  | 76,211     | 68,850  | 73,535     | 65,968  |
|    | 20-29세 | 143,317       | 119,736 | 157,633    | 131,472 | 157,553    | 133,018 |
|    | 30-39세 | 156,220       | 140,324 | 170,551    | 145,112 | 172,855    | 146,115 |
|    | 40-49세 | 161,977       | 145,056 | 173,976    | 149,453 | 176,241    | 151,338 |
|    | 50-59세 | 128,416       | 126,071 | 141,552    | 133,805 | 150,145    | 140,444 |
|    | 60-69세 | 86,525        | 97,376  | 89,948     | 99,241  | 91,955     | 99,545  |
|    | 70세 이상 | 85,712        | 136,893 | 91,758     | 142,827 | 95,445     | 149,260 |
| 전북 | 15-19세 | 69,635        | 61,196  | 69,301     | 62,219  | 68,490     | 62,182  |
|    | 20-29세 | 112,114       | 95,473  | 118,423    | 100,951 | 118,080    | 99,069  |
|    | 30-39세 | 121,125       | 119,505 | 123,526    | 119,668 | 120,787    | 117,092 |
|    | 40-49세 | 138,007       | 134,191 | 141,905    | 135,128 | 141,909    | 134,866 |
|    | 50-59세 | 117,437       | 120,994 | 123,980    | 125,494 | 129,761    | 130,949 |
|    | 60-69세 | 84,300        | 97,937  | 85,420     | 98,335  | 85,855     | 97,387  |
|    | 70세 이상 | 75,682        | 129,706 | 79,437     | 133,697 | 82,992     | 139,813 |
| 전남 | 15-19세 | 61,450        | 53,564  | 63,567     | 56,710  | 64,738     | 58,570  |
|    | 20-29세 | 89,239        | 71,160  | 93,044     | 74,103  | 91,250     | 70,722  |
|    | 30-39세 | 111,564       | 105,743 | 115,298    | 106,048 | 111,953    | 102,863 |
|    | 40-49세 | 137,329       | 126,003 | 143,077    | 127,377 | 142,583    | 126,245 |
|    | 50-59세 | 119,401       | 120,346 | 126,780    | 124,280 | 131,885    | 128,415 |
|    | 60-69세 | 92,855        | 109,530 | 94,471     | 109,062 | 94,441     | 107,246 |
|    | 70세 이상 | 91,291        | 163,161 | 96,476     | 168,612 | 99,528     | 174,322 |
| 경북 | 15-19세 | 97,020        | 79,745  | 95,283     | 79,901  | 92,546     | 78,600  |
|    | 20-29세 | 172,242       | 138,792 | 183,638    | 146,666 | 182,543    | 143,694 |
|    | 30-39세 | 184,337       | 174,278 | 192,772    | 176,421 | 190,608    | 173,974 |
|    | 40-49세 | 208,808       | 198,541 | 214,790    | 199,074 | 213,345    | 197,468 |
|    | 50-59세 | 183,163       | 186,339 | 193,565    | 192,803 | 202,746    | 200,700 |
|    | 60-69세 | 121,077       | 140,026 | 123,446    | 141,223 | 126,319    | 142,081 |
|    | 70세 이상 | 112,959       | 192,851 | 118,278    | 199,259 | 122,784    | 207,307 |
| 경남 | 15-19세 | 118,993       | 98,360  | 122,421    | 102,634 | 123,239    | 105,127 |
|    | 20-29세 | 200,585       | 164,580 | 215,537    | 168,009 | 209,027    | 160,463 |
|    | 30-39세 | 252,425       | 241,920 | 271,983    | 244,438 | 269,980    | 240,615 |
|    | 40-49세 | 276,323       | 265,326 | 290,535    | 268,533 | 291,039    | 268,493 |
|    | 50-59세 | 216,790       | 213,557 | 234,227    | 224,669 | 247,472    | 237,076 |
|    | 60-69세 | 121,948       | 141,382 | 126,881    | 144,205 | 131,996    | 147,009 |
|    | 70세 이상 | 94,776        | 175,736 | 100,845    | 182,333 | 105,493    | 189,942 |
| 제주 | 15-19세 | 20,588        | 18,576  | 21,812     | 19,544  | 22,396     | 20,047  |
|    | 20-29세 | 31,379        | 28,465  | 33,369     | 30,061  | 33,122     | 29,213  |
|    | 30-39세 | 40,601        | 40,137  | 42,598     | 40,656  | 42,000     | 40,035  |
|    | 40-49세 | 46,012        | 43,947  | 49,135     | 45,355  | 50,074     | 46,314  |
|    | 50-59세 | 33,075        | 32,816  | 36,043     | 34,928  | 38,489     | 37,254  |
|    | 60-69세 | 21,554        | 23,765  | 22,645     | 24,483  | 23,011     | 24,667  |
|    | 70세 이상 | 15,694        | 30,958  | 17,278     | 32,516  | 18,403     | 33,939  |

<표 51> 통합 2~4차의 사후조정을 위한 인구주택총조사 가구 및 추계가구

| 시도 | 2010년 인구주택총조사<br>가구수 | 2011년 추계 가구수 | 2012년 추계 가구수 |
|----|----------------------|--------------|--------------|
| 서울 | 3,504,297            | 3,524,877    | 3,535,741    |
| 부산 | 1,243,880            | 1,263,159    | 1,272,041    |
| 대구 | 868,327              | 880,444      | 890,369      |
| 인천 | 918,850              | 942,184      | 966,165      |
| 광주 | 515,855              | 527,769      | 537,645      |
| 대전 | 532,643              | 546,195      | 558,233      |
| 울산 | 373,633              | 382,228      | 389,429      |
| 경기 | 3,831,134            | 3,948,096    | 4,036,972    |
| 강원 | 557,751              | 571,470      | 581,218      |
| 충북 | 558,796              | 570,865      | 582,687      |
| 충남 | 749,035              | 777,545      | 797,944      |
| 전북 | 659,946              | 670,676      | 679,028      |
| 전남 | 681,431              | 692,672      | 697,722      |
| 경북 | 1,005,349            | 1,023,891    | 1,039,017    |
| 경남 | 1,151,172            | 1,173,925    | 1,191,649    |
| 제주 | 187,323              | 191,005      | 194,815      |

## 5. 통합패널 차수별 가중치 산출 결과

한국노동패널조사의 가중치는 4절에서 설명한 가중치 산출 과정으로 부여된 개인 중 단면, 횡단면 가중치와 가구 가중치를 최종적으로 조사년도의 모집단인 추계인구 및 가구를 기준으로 사후조정하여 가구 및 가구원에 대한 최종 가중치를 부여하게 된다.

2009년에 전국 모집단을 대표하기 위해 패널을 추가한 통합패널에 대해 사후조정으로 조정된 통합패널의 차수별 최종 가중치를 요약하면 <표 52>, <표 53>과 같다.

한국노동패널조사의 개인 가중치는 조사년도의 모집단을 기준으로 사후조정하여 기본적으로 중단면 가중치와 횡단면 가중치를 동일하게 유지하도록 부여하고 있다.

첫째, 2009년의 통합 1차년도는 통합 시점에서의 가중치를 2009년 추계 인구를 기준으로 사후조정하여 최종 가중치를 산출함으로써 1차년도는 중단면 가중치와 횡단면 가중치가 동일하게 유지된다.

둘째, 통합 2차년도이후도 조사시점의 인구주택총조사 인구나 추계인구를 기준으로 사후조정하여 종단면 가중치와 횡단면 가중치를 동일하게 유지하게 되지만 앞서 설명한 바와 같이 조사년도별로 응답 모형에 이용되는 설명변수의 무응답 자료가 존재하는 경우는 약간 상이한 결과가 나타나게 된다. <표 52>를 보면, 통합 2차와 통합 3차의 종단면 가중치와 횡단면 가중치가 차이가 나타나는 데, 이는 통합 2차에서는 제주도에 15~19세 조사대상자가 없어서 사후조정에 반영을 못한 결과이며, 통합 3차에서는 응답자 중 연령이 결측으로 사후조정을 할 수 없기 때문에 발생한 결과이다. 하지만 설명 변수의 자료가 완전한 경우는 통합 4차와 같이 개인 종단면 가중치와 횡단면 가중치가 일치하게 됨을 확인할 수 있다.

<표 52> 통합패널 차수별 개인 가중치 기초통계량

| 차수   | 변수                  | N     | Sum      | Mean    | Std     | Min | Q1      | Median  | Q3      | Max      |
|------|---------------------|-------|----------|---------|---------|-----|---------|---------|---------|----------|
| 통합1차 | SW1_P <sup>1)</sup> | 14489 | 40077120 | 2766.04 | 1694.66 | 0   | 1643.39 | 2455.50 | 3461.46 | 25535.56 |
| 통합2차 | SW2_L_P             | 14064 | 40172409 | 2856.40 | 2013.43 | 0   | 1583.97 | 2491.22 | 3641.32 | 28465.00 |
|      | SW2_C_P             | 14064 | 40203788 | 2858.63 | 1998.12 | 0   | 1589.25 | 2477.91 | 3658.74 | 31379.00 |
| 통합3차 | SW3_L_P             | 13831 | 41975159 | 3034.86 | 2301.58 | 0   | 1614.98 | 2607.69 | 3890.74 | 29373.91 |
|      | SW3_C_P             | 13831 | 42008528 | 3037.27 | 2225.72 | 0   | 1610.80 | 2588.78 | 3922.17 | 34070.63 |
| 통합4차 | SW4_L_P             | 13929 | 42445378 | 3047.27 | 2486.67 | 0   | 1554.38 | 2557.76 | 3908.88 | 36734.16 |
|      | SW4_C_P             | 13929 | 42445378 | 3047.27 | 2340.61 | 0   | 1571.23 | 2541.22 | 3944.79 | 36385.05 |

<표 52>는 통합패널의 차수별 개인 가중치를 요약한 기초통계량 결과이다. 개인 가중치는 종단면과 횡단면 가중치로 나눌 수 있는데, 종단면의 경우에는 L, 횡단면은 C로 표기한 변수로 생성하였다.

<표 53>은 통합 차수별 모집단인 추계 가구로 사후조정된 통합패널 차수별 가구 최종 가중치의 기초통계량이다. <표 48>에서 N은 통합 차수별 응답된 가구를 의미하며, 가구 가중치가 '0' 인 가구는 원가구원이 한 명도 응답되지 않은 가구의 가중치를 의미한다.

<표 53> 통합패널 차수별 가구 가중치 기초통계량

| 차수   | 변수    | N    | Sum      | Mean    | Std     | Min | Q1      | Median  | Q3      | Max      |
|------|-------|------|----------|---------|---------|-----|---------|---------|---------|----------|
| 통합1차 | SW1_H | 6721 | 17048965 | 2536.67 | 1519.62 | 0   | 1569.89 | 2258.49 | 3147.35 | 18928.04 |
| 통합2차 | SW2_H | 6630 | 17339422 | 2615.30 | 1780.80 | 0   | 1500.92 | 2293.89 | 3335.12 | 21413.10 |
| 통합3차 | SW3_H | 6618 | 17687001 | 2672.56 | 1918.55 | 0   | 1465.23 | 2291.84 | 3440.77 | 29944.90 |
| 통합4차 | SW4_H | 6686 | 17950675 | 2684.82 | 2006.57 | 0   | 1425.79 | 2258.07 | 3473.63 | 28794.64 |

1) 2009년 시도 연령별 추계인구로 조정된 통합 1차 종단면-횡단면 가중치(종단면과 횡단면 같음으로 가정)

#### IV. 노동패널 가중치 사용 방안

본 연구에서는 각 년도 별 그리고 원패널과 통합패널 별로 종·횡단면 가구원 가중치와 가구 가중치의 총 3개의 가중치가 산출되었다. 생성된 3종류의 가중치는 분석의 목적에 따라 적절하게 사용되어야 한다. 기본적으로 종단면 가구원 가중치는 종단면 가구원 단위 분석을 위해서 그리고 횡단면 가구원 가중치는 횡단면 가구원 단위 분석을 위해 그리고 가구 가중치는 가구 단위 분석을 위해 사용되어야 한다.

종단면 가구원 단위 분석은 크게 두 가지로 구분할 수 있다. 그 첫 번째는 관심의 대상이 되는 모든 년도의 자료를 이용하여 분석되는 경우로 예를 들어 혼합선형 모형이 여기에 해당된다. 이 경우 분석을 위해 일반적으로 각 년도 별 자료는 세로 결합을 통해 병합되고 병합된 자료를 통해 고려된 모형이 적합 될 수 있다. 예를 들어 혼합 선형 모형이 고려되는 경우,  $t$ 년도의 가구원  $i$ 로부터 관측된 변수를 위한 모형은 아래와 같다.

$$y_{it} = \mathbf{x}_{it}\beta + t \alpha + \eta_i + \epsilon_{it}$$

여기서  $\mathbf{x}_{it}$ 는 분석에 사용되는 설명변수 벡터이며  $\alpha$ 는 시간효과를 나타내고  $\eta$ 와  $\epsilon$ 은 랜덤효과로 가구원 효과와 가구원 내 반복측정 된 값들의 랜덤효과를 나타낸다. 이 경우 자료의 형태는 다음의 <그림 3>과 같이 표현할 수 있다.

<그림 3> 자료의 형태

| $t$ | 설명변수           | 관심변수           | 가중치 벡터         |
|-----|----------------|----------------|----------------|
| 1   | $\mathbf{X}_1$ | $\mathbf{y}_1$ | $\mathbf{w}_1$ |
| 2   | $\mathbf{X}_2$ | $\mathbf{y}_2$ | $\mathbf{w}_2$ |
| ⋮   | ⋮              | ⋮              | ⋮              |
| $T$ | $\mathbf{X}_T$ | $\mathbf{y}_T$ | $\mathbf{w}_T$ |

여기서  $\mathbf{X}_t$ 는  $t$ 년차의 설명변수 행렬을,  $\mathbf{y}_t$ 는  $t$ 년차의 관심변수 벡터를 그리고  $\mathbf{w}_t$ 는 각 연차별 산출된 종단면 가중치 벡터를 나타낸다. 분석을 위한 자료의 행의 수는 따라서 각 년차에 관측된 가구원 수의 총  $T$ 년간의 총합이다. 즉 실제 분석을 위해

서 특별히  $\beta$  혹은  $\alpha$ 의 설명력을 평가하기 위하여 사용되는 관측치의 수 혹은 자유도는 랜덤효과인  $\eta$ 의 자유도가 사용되거나 반복을 고려한 전체 자료 수는 실제 자료를 제공하는 개인들의 수보다 크게 나타난다. 또한 이 경우 그림에 나와 있는 각년도 별 가중치를 분석을 위해서 직접 사용될 수 없다. 이 경우 가능한 가중치 사용방안은 각년도 별 가중치를 일종의 표준화한 가중치를 적용하는 것이다. 가능한 새로운 가중치는 산출된 가중치의 총합이 일정하며 각년도 별 상대적 크기를 고려하여 표준화하는 다음과 같은 것이 될 수 있다.

$$w_t^* = w_t \frac{\sum_i w_{ti}}{\sum_j \sum_i w_{ji}}$$

두 번째는 각 가구원 단위 년도 별 변화의 평균 추정과 같은 net change를 예측하기 위한 분석이다. 이 경우 분석 대상이 되는 가구원은 분석 대상 년도에 모두 응답한 가구원이며 이 경우 제공된 종단면 가중치에 추가적인 작업이 이루어져야 한다. 가능한 한 방법은 고려되는 년차들의 가중치 중 최근 생성된 종단면 가중치를 분석 대상이 되는 가구원 수에 맞추어서 스케일 조정을 하는 것이다. 이러한 방안은 매 번 분석의 대상이 되는 년차 별 자료에 따라 새로운 가중치를 생성해야 한다.

또 다른 방안으로는 위의 혼합선형모형과 같은 모형의 가정 하에서 관심변수, 예를 들어  $E(y_{it} - y_{it-1})$ 에 대한 일종의 최소제곱추정량을 구하는 것으로 이는 가중치의 재산출 없이 직접적으로 추정량의 산출이 가능하나 그 해석에 있어서는 주의가 필요하다.

횡단면 가구원 가중치 및 가구 가중치의 사용여부는 종단면 가중치의 경우와는 달리 직관적이며 이는 각년도 별 모집단 총합, 모비율, 모평균과 같은 관심 모수를 추정하기 위해 직접적으로 사용할 수 있다.

기본적으로는 종·횡단면 가구원 가중치 그리고 가구 가중치는 분석의 목적, 특별히 분석 결과를 적용하게 되는 추론의 대상 모집단의 정의에 따라 타당하게 사용되어야 하고 실제 패널 자료는 확률 표본 추출법에 의하여 구성되었기 때문에 이 가중치의 사용은 특별한 경우를 제외하고는 사용하는 것이 바람직하다.

## V. 참고문헌

- 강석훈 (2003), “KLIPS의 가중치 부여방안 연구.” 한국노동패널연구 2003-04.
- 김영원 · 김재광 · 이기재 · 조유미 (2005). “한국노동패널 표본의 대표성과 가중치 보정방법.” 제 6회 한국노동패널 학술대회 논문집, pp. 323-339.
- 김재광 (2003). “한국노동패널 오차분석 및 표본추가.” 한국노동패널연구 2003-02
- 김재광 (2009). “한국노동패널조사 표본추가 연구.” 미발간 보고서
- 박민규 · 김영원 · 최형아 (2011). “13차년도 한국노동패널 가중치 연구.” 한국고용정보원
- 이계오 · 임찬수 (2012). “한국노동패널 가중치 산출방안-원패널과 추가패널.” 한국노동패널 학술대회 논문집.
- Duncan, G. (1995), "A Simple Method for Weighting in Household Panel Survey", Working paper, Northwestern University.
- Kalton, G. and M. Brick (1994), "Weighting Schemes for Household Panel Survey", Proceedings of the Section on Survey research Method, American Statistical Association, 785-790.
- Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury Press.