

**한국노동패널 1-22차년도 조사자료**

Korean Labor and Income Panel Study

# User's Guide

KLIPS



# 목 차

I . 한국노동패널(KLIPS) 소개 .....	1
1. KLIPS 조사개요 .....	1
2. 현장실사 및 면접방식 .....	4
3. 조사결과 .....	8
II . 표본설계 및 가중치 .....	16
1. 표본설계방법 및 표본추출과정 .....	16
2. 가중치 부여방식 및 사용방법 .....	27
III . 자료의 주요 특성 .....	40
1. 설문지 구성 .....	40
2. 자료 구성 .....	44
3. 조사내용 .....	46
IV . 변수 가이드 .....	49
1. key변수의 이해와 이용 .....	50
2. 오픈 코드(open code) .....	61
3. 무응답 대체 데이터(Imputation) .....	65
4. 직업코드 및 산업코드 .....	65
V . 주제별 가이드 .....	69
1. 가구자료 .....	69
2. 개인자료 .....	82
3. 직업력 자료 .....	99
4. 부가조사 자료 .....	106

FAQ .....	130
〈부록〉 SMART_KLIPS User's Guide .....	263
찾아보기 .....	270

## 표 목 차

<표 I - 1> 각 연도별 조사방식(면접, 유치, 전화조사) 분포 .....	6
<표 I - 2> 연도별 조사 결과 비교 .....	10
<표 I - 3> 표본가구 구성 .....	14
<표 II-1> 최종 추출된 조사구와 조사가구 .....	18
<표 II-2> KLIPS 1차년도 조사가구 개요 .....	19
<표 II-3> 표본추가 후 KLIPS의 동부·읍면부별 분포 .....	20
<표 II-4> 최종 배정 표본 조사구 .....	21
<표 II-5> 1~18차 조사결과 .....	23
<표 II-6> 패널 추가 및 17예비표본의 가구 규모 .....	24
<표 II-7> 98표본 가구 가중치 특성 .....	29
<표 II-8> 98표본 개인 가중치 특성 .....	29
<표 II-9> 09통합표본 가구 가중치 특성 .....	34
<표 II-10> 09통합표본 개인 가중치 특성 .....	35
<표 II-11> 18통합표본 가구 가중치 특성 .....	37
<표 II-12> 18통합표본 개인 가중치 특성 .....	37
<표 II-13> 가중치 변수명 .....	38
<표 III- 1> 22차년도 KLIPS 설문지의 구성 .....	41
<표 III- 2> 1-22차년도 KLIPS 응답자용 설문지의 구성 .....	42
<표 III- 3> 일자리 지속을 통한 유형설문의 판별 .....	43
<표 III- 4> 1-22차년도 KLIPS 자료의 구성 .....	45
<표 III- 5> 1-22차년도 조사 내용 (공통) - 응답자용 설문 기준 .....	46
<표 III- 6> 개인별 한번만 조사하는 내용 - 응답자용 설문 기준 .....	47
<표 III- 7> 각 차수별 차별된 조사 내용 - 변수명 일치 자료 기준 .....	48

<표 IV-1> 가구의 고유번호 .....	52
<표 IV-2> 1-22차년도 연령 변수 .....	55
<표 IV-3> 각 연도별 학력 변수 .....	56
<표 IV-4> 경제활동상태의 구분 .....	59
<표 IV-5> 각 연도별 실업자 및 비경제활동인구 구분변수 .....	60
<표 IV-6> 1-22차년도 종사상 지위 변수 .....	60
<표 IV-7> 현재(최종) 일자리의 직업 · 산업코드 변수 .....	68
<표 V-1> 1-22차년도 가구원 정보 변수 .....	70
<표 V-2> 1-22차년도 가구소득 관련 변수 .....	72
<표 V-3> 1-22차년도 가구소득 세부항목 .....	74
<표 V-4> 1-22차년도 가구소비 관련 변수 .....	75
<표 V-5> 2-22차년도 부동산 관련 변수 .....	76
<표 V-6> 2-22차년도 금융자산 관련 변수 .....	77
<표 V-7> 2-22차년도 가구부채 관련 변수 .....	78
<표 V-8> 1-22차년도 주거지 관련 변수 .....	78
<표 V-9> 사교육 및 보육기관 이용실태 조사내용 .....	81
<표 V-10> 평소의 경제활동관련 변수 .....	84
<표 V-11> 1-22차년도 구직활동 관련 변수 .....	85
<표 V-12> 1-22차년도 근로계약 관련 변수 .....	86
<표 V-13> 비정규직 측정과 관련된 다양한 변수들 .....	87
<표 V-14> 고용형태 자료의 변수 .....	88
<표 V-15> 1-22차 근로시간 관련 변수 .....	90
<표 V-16> 1-22차년도 임금근로자의 임금관련 변수 .....	91
<표 V-17> 1-22차년도 비임금근로자의 소득관련 변수 .....	92
<표 V-18> 1-22차년도 사회보험 관련 변수 .....	94
<표 V-19> 1-22차년도 요인별 직무만족도(facet measure) 변수 .....	95
<표 V-20> 단일항목으로 구성된 전반적 직무만족도 변수 .....	96
<표 V-21> 5개 항목으로 구성된 전반적 직무만족도 변수 .....	96

<표 V-22> 조직몰입도의 측정 변수 .....	97
<표 V-23> 1-22차년도 기타 변수 .....	98
<표 V-24> 1-22차년도 신규조사자 관련 변수 .....	99
<표 V-25> 직업력 자료의 변수 .....	104
<표 V-26> 3차년도 청년층 부가조사 자료의 변수 .....	107
<표 V-27> 4차년도 건강과 은퇴 부가조사 자료의 변수 .....	109
<표 V-28> 6차년도 중고령자 부가조사 자료의 변수 .....	110
<표 V-29> 7차년도 근로시간과 여가 부가조사 자료의 변수 .....	112
<표 V-30> 8차년도 노동조합과 노사관계 부가조사 자료의 변수 .....	113
<표 V-31> 9차년도 청년층 부가조사 자료의 변수 .....	114
<표 V-32> 10차년도 고용형태 자료의 변수 .....	115
<표 V-33> 11차년도 교육 부가조사 자료의 변수 .....	116
<표 V-34> 17차년도 시간사용과 삶의 질 부가조사 자료의 변수 .....	118
<표 V-35> 18차년도 삶의 인식 부가조사 자료의 변수 .....	120
<표 V-36> 18차년도 행복 관련 부가조사 자료의 변수 .....	120
<표 V-37> 18차년도 장년층 부가조사 자료의 변수 .....	122
<표 V-38> 19차년도 청년층 부가조사 자료의 변수 .....	124
<표 V-39> 20차년도 의사결정 부가조사 자료의 변수 .....	126
<표 V-40> 21차년도 고용형태 부가조사 자료의 변수 .....	128
<표 V-41> 22차년도 비임금근로자 부가조사 자료의 변수 .....	129

## 그림목차

[그림 I -1] 통합표본의 개요 .....	12
[그림 II -1] 21차(2018년) 조사대상 규모 .....	26
[그림 II -2] 09통합표본에서의 가중치 산출 과정 .....	33
[그림 V -1] 직업력 자료의 구조 .....	100



# I. 한국노동패널(KLIPS) 소개

## 1. KLIPS 조사개요

현재 우리나라는 급격한 사회·경제적 변동과 함께 노동시장도 질적·양적 변화를 겪고 있으며, 이에 따라 보다 구체적이고 미시적인 노동정책에 대한 수요가 증대되고 있다. 올바른 노동정책의 수립과 실행을 위해서는 신뢰할 수 있는 통계 데이터의 뒷받침이 필수적이지만 우리나라에서 실시하고 있는 주요 노동관련 통계 조사만으로는 역부족인 실정이다. 특히 노동시장 정책의 효과는 일시적이고 단기적인 것도 있지만 그 효과가 장기적·지속적인 것도 있다. 따라서 정책의 효과 및 효율성을 이해하기 위해서는 상당기간 동안 지속적으로 개인 및 가구의 경제활동을 관찰할 필요성이 생긴다.

한국노동패널조사(Korean Labor and Income Panel Study: 이하 KLIPS)는 국내유일의 노동관련 가구패널조사로 횡단면 자료와 시계열 자료의 장점을 모두 갖고 있는 자료이다. KLIPS는 도시지역에 거주하는 한국의 5,000가구와 가구원을 대표하는 패널표본 구성원(5,000가구에 거주하는 모든 가구원)을 대상으로 1년에 1회씩 조사를 실시하고 있으며, 1998년 1차 조사를 시작으로 22차 조사(2019년)까지 완료되었다. 패널 조사는 매년 동일한 가구 및 가구원에 대해 동일한 조사를 반복 실시하기 때문에 원표본 가구 유지율이 매우 중요하다. KLIPS의 각 조사년도별 원표본가구 유지율을 살펴보면, 2차년도(1999년) 87.6%, 3차년도(2000년) 80.9%, 4차년도(2001년) 77.3% 이후 표본 유지율은 안정되어 매해 약 1%포인트 감소하여, 22차년도(2019년)에 65.3%의 표본유지율을 나타냈다.

KLIPS 자료는 크게 가구를 조사 대상으로 한 가구용 자료와 가구에 속한 만15세 이상의 가구원을 조사 대상으로 한 개인용 자료로 구분된다. 가구용 자료의 내용은 가구원의 인적 사항, 변동 가구원 관련 사항, 가족관계와 세대간 경제적 자원 교류, 주거상태, 자녀교육과 보육, 가구의 소득과 소비, 가구의 자산과 부채, 가구의 경제 상태 및 가계에 부담을 느끼는 소비 항목 등의 내용을 담고 있으며, 개인용 자료는 개인의 경제활동상태, 소득활동 및 소비, 교육 및 직업 훈련, 고용상의 특성, 근로시간, 직무만족 및 생

활만족, 구직활동, 노동시장 이동 등의 다양한 내용을 담고 있다. 한편 노동패널 표본에 처음으로 편입된 개인을 조사 대상으로 한 신규 조사자용 자료는 개인용 자료의 모든 정보와 함께 1차 조사 당시 파악했던 기본적인 인적 사항을 추가적으로 수록하고 있다. 회고적(retrospective) 일자리를 포함한 개인의 모든 일자리 정보를 토대로 구성된 직업력 자료는 기업 형태 및 규모, 고용형태, 취업시기 및 퇴직시기, 업종과 직종, 종사상 지위, 근로시간 형태, 임금 및 소득 등 개인의 일자리 관련 정보를 제공한다.

부가조사는 3차, 4차, 6차, 7차, 8차, 9차, 10차, 11차, 17차, 18차, 19차, 20차, 21차, 22차년도에 각 실시되었다. 3차년도에는 만15세 이상 30세 미만의 청년층을 대상으로 청년층 부가조사를 실시하였고, 4차년도에는 만45세 이상의 개인을 대상으로 건강과 은퇴 부가조사를 실시하였다. 5차년도에는 별도의 부가조사를 실시하지 않았으며, 6차년도에는 만50세 이상의 중고령자를 대상으로 은퇴 및 노후생활과 관련된 부가조사를 실시하였다. 7차년도에는 2004년 7월 1일부터 시행된 법정근로시간단축에 따라 응답자들의 근로시간 변화와 여가사용에 대해 조사하였으며, 8차년도 조사에서는 노동조합과 노사관계에 관한 부가조사를 실시하였다. 9차년도 조사에서는 청년층을 대상으로 성장과정, 학업 및 학교생활을 조사하는 청년층 부가조사를 실시하였으며, 10차년도 조사에서는 임금근로자와 일부 자영업자를 대상으로 계약 형태 등을 심층 조사하는 고용형태 부가조사를 실시하였다. 11차년도 조사에서는 중고등학교 및 대학교 재학경험과 사교육 등을 심층 조사하는 교육 부가조사를 실시하였다. 17차년도 조사에는 만15세 이상 만74세 이하 모든 응답자를 대상으로 시간사용과 삶의 질 부가조사를 실시하였다. 18차년도 조사에서는 두가지 부가조사가 실시되었는데, 삶의 인식부가조사에서는 만 15세 이상 모든 응답자를 대상으로 개인의 성격과 태도를 파악하는 비인지적 능력을 측정하였고, 장년층 부가조사에서는 만 50세 이상 만 74세 이하 모든 응답자를 대상으로 장년층의 고용과 은퇴 후 경제적 상태, 사회관계 및 건강 등을 파악하였다. 19차년도 조사에는 만15세 이상 만 34세 이하 청년층을 대상으로 취업과 NEET 및 주관적 인식 등의 부가조사를 실시하였다. 20차년도 조사에서는 개인의 전략적 사고, 합리성 및 사회성을 측정하였다. 21차년도 조사에서는 다양한 형태의 고용 증가를 파악하기 위해 고용형태를 측정하였다. 22차년도에서는 고용주, 자영업자, 무급가족 종사자를 대상으로 비임금 설문을 진행하였다.

한편, 한국노동패널의 표본추가는 12차년도(2009년), 21차년도(2018년) 두차례에 걸쳐서 이루어졌다. 12차년도(2009년)에는 표본의 전국대표성 확보를 위해 1,415가구 표본을

추가하였다. 1998년 추출된 표본(이하 KLIPS 98표본)은 제주도를 제외한 도시가구를 대상으로 표집하였기 때문에, 전국을 아우르는 분석을 하는데 한계가 있었다. 이에 12차년도(2009년)에 1,415개의 가구 표본을 추가하여, 이러한 문제를 보완코자 하였다. 21차년도(2018년) 표본추가는 20년동안 진행된 조사기간 중 탈락하는 가구가 많아지면서, 전체 모집단의 성격을 반영하기 못하는 것을 보완하기 위해, 5,044가구를 추가하였다.

KLIPS 자료를 통하여 기대되는 바는 보다 합리적이고 정확한 고용정책의 수립과 실행이 가능하다는 점이다. KLIPS의 등장 이전에 노동시장 분석 및 정책 결정에 주로 사용되어 왔던 자료는 횡단면자료(cross-section data) 또는 시계열자료(time-series data)였다. 횡단면 자료란 일정한 시점에서 각각의 표본들이 어떤 값을 갖는가를 보여주는 자료로 미시적인 경제 분석에 주로 사용되는 반면 시계열 자료는 주어진 표본들이 시간의 흐름에 따라 어떤 값을 갖는지에 대한 동태적인 변화를 보여준다. 반면 패널 자료는 노동시장의 분석에 일반적으로 사용되는 횡단면 자료의 모든 장점을 갖추고 있으며, 더불어 횡단면 자료로는 파악할 수 없는 추가적인 장점을 지니고 있다. 관측할 수 없는 개인효과(unobservable individual effects)를 통제함으로써 시간의 경과 혹은 정부정책 등 환경의 변화에 따라 개인 또는 가구의 의사결정이나 행동양식이 어떻게 변화해나가는지에 관한 동태적 변화를 정확하게 포착할 수 있으며, 횡단면 자료로는 파악이 불가능한 상태 간 변이과정(transition between states)을 분석할 수 있다. 특정기간효과(period specific effect), 특정연령효과(age specific effect), 그리고 동시생태효과(cohort effect) 등에 대한 분석이 가능하다는 점 또한 패널 자료의 장점으로 꼽힌다.

KLIPS 자료는 가구특성, 경제활동 및 노동시장이동, 소득활동 및 소비, 교육 및 직업훈련 등 수많은 정보를 담고 있기 때문에 학문적 연구와 정책적 발전에 기여할 수 있는 매우 유용한 자료이다. KLIPS를 이용하여 수행할 수 있는 과제의 예로는 학교를 졸업한 개인이 최초의 직장으로부터 전직·이직 과정을 거쳐 은퇴하기까지 생애에 걸친 직업경로에 대한 이해, 여성의 노동시장 진입과 탈퇴과정과 이에 영향을 미치는 요인들의 효과, 개인이 교육, 훈련, 연수 등을 통하여 인적자본(human capital)을 쌓아 가는 과정과 이것이 일자리 진입에 미치는 영향, 실업자 재취업훈련 등 정부의 고용정책이 인적자본의 형성에 미치는 영향, 가구 경제의 동태적 변화와 세대 변화 연구 등을 들 수 있으며, 이러한 주제들은 KLIPS 자료와 같은 패널자료를 활용함으로써 보다 정밀한 분석이 가능하다.

KLIPS는 다른 일회성 조사들과는 달리 전국의 12,000여 가구와 그 가구에 속한 23,000

여명의 개인을 매년 추적 조사하는 대규모 패널조사이기 때문에 조사과정과 자료생산과정에 많은 어려움을 겪는다. 우선, 표본 추출, 설문지의 구성 등과 같은 조사설계(survey design)가 결정되면 조사가 진행되는 한 이를 변경하기 어렵다. 따라서 표본 추출, 설문 설계 등과 같은 사전 준비에 많은 노력과 비용이 투입되어야 하며, 본 조사 이전에 철저한 예비조사(pilot survey)를 하는 것이 일반적이다. 둘째, 조사 대상인 표본이 결정되면 패널 조사가 진행되는 한 동일한 표본이 계속 유지되어야 한다. 즉 매년 동일한 가구 및 개인을 대상으로 조사하기 때문에 이들이 응답거부하지 않도록 설득하고, 이사 또는 분가 등의 이유로 표본에서 탈락하지 않도록 하는 표본유지 및 관리에 많은 노력과 비용이 소요된다. 셋째, 설문 내용이 방대하며 내용 또한 개인의 소득, 고용상태 등 민감한 사안들로 구성되어 있기 때문에 실사과정에서 면접원들이 조사하는데 어려움을 겪는다. 넷째, 매년 방대한 양의 데이터가 지속적으로 축적되기 때문에 자료를 데이터베이스화하고 관리하는 데에도 상당한 시간과 비용이 소요된다. 특히 조사된 자료는 클리닝(cleaning)과정과 가공과정을 거쳐서 제공되기 때문에 KLIPS 1개년도 자료를 생산하기 위해서는 이후 1년 이상의 기간이 소요된다. 한편, 이러한 과정을 거쳐 생산된 KLIPS 자료의 학문적 연구 및 정책 수립을 위한 기초 자료로서의 유용성을 보다 높이기 위해, 매년 정기적인 학술대회를 개최하고 있다.

## 2. 현장실사 및 면접방식

### 2.1 조사대상과 조사방법

#### 2.1.1 가구용 조사대상과 조사방법

- ☐ KLIPS는 가구패널 조사로 기본적으로 가구를 기준으로 추적 조사를 실시한다. KLIPS 원가구(혹은 가구 표본)는 세 가지로 구분된다. 1998년 최초 조사 당시 표본가구로 구성된 98원표본과, 2009년(12차년도) 당시 응답가구를 원표본으로 보는 09통합표본, 2018년(21차년도) 당시 응답가구를 원표본으로 보는 18통합표본이다.
- ☐ 따라서, KLIPS의 가구 조사 대상은 첫째, 98원표본 가구인 5,000가구와 원가구에서

분가한 분가가구이며, 둘째, 12차년도(2009년) 당시 응답가구(6,721가구)를 원표본으로 하는 09통합표본 원가구와 분가가구, 셋째, 21차년도(2018년) 당시 응답가구(12,134가구)를 원표본으로 하는 18통합표본 원가구와 분가가구이다.

- 설문은 반드시 직접 면접을 통하여 이루어진다. 이때, 면접원은 주로 가구주 본인이나 가구주의 배우자를 대상으로 면접 하는 것을 원칙으로 한다.
- 이중 1차년도 조사당시 원가구원들이 분가하여 새로운 가구를 구성하였을 경우에는 다음과 같은 원칙에 따라 추적조사한다.

- 분가가구의 조사추적 원칙

- 반드시 추적해야 하는 가구원인 원가구원(98표본 원가구원과 그 자녀, 09통합표본 원가구원과 그 자녀, 18통합표본 원가구와 그 자녀)이 있는 가구는 모두 조사대상이다.

예) 1998년 서울에 사는 한 가구가 원가구로 선정되어 조사됨

→ 2000년 12월 형과 아우가 같이 대전으로 분가하여 나왔음

→ 2001년 3월(4차년도 조사일 이전) 형이 또 부산으로 분가하여 나감

→ 이 경우 대전에 살고 있는 동생은 물론 부산으로 분가한 형이 살고 있는 가구도 추적조사해야 함

- 추적하지 않아도 되는 가구원 즉, 원가구원이 한명도 없는 가구는 조사대상이 아니다.

예) 1998년 서울에서 사는 한 가구가 원가구로 선정되어 조사됨

→ 2001년 12월 가구주의 아들과 며느리가 잠시 함께 살게 됨(1998년 당시 따로 살고 있던 아들과 며느리는 가구원이 아님)

→ 2002년 3월 가구주의 아들과 며느리가 분가하여 나감

→ 이 경우 아들과 며느리는 원가구원이 아니므로 이 가구는 추적조사하지 않음

### 2.1.2 개인용 설문의 조사대상과 조사방법

- 개인용 설문 대상은 각 연도별 조사대상 가구에 소속된 만15세 이상의 개인이다.

□ 한국노동패널조사가 실시되는 기간 동안 유학, 출장, 여행, 입원, 실종, 행상 등의 이유로 함께 거주하지 않아 면접조사에 응할 수 없는 가구구성원에 대해서는 다음과 같은 원칙에 따라 조사를 실시한다.

- 실사기간 중 가구로 일시 또는 완전 복귀하는 사람 → 직접조사
- 가구로 복귀하지 않은 사람 → 전화조사
- 면접이나 전화조사가 어려우면서 현재 경제활동을 하고 있는 가구원 → 대리응답
- 정신박약, 허약, 치매, 노환 등 정신적, 신체적 장애로 응답할 수 없는 성인 → 대리 응답
- 면접이나 전화조사가 어려운 경우, 조사일 현재 비경제활동인구에 속하는 가구원
  - 예 : 영내거주 군인 및 전투경찰, 취학으로 외지에 상주하는 가구원, 교도소, 소년원, 요양원 등 사회보호시설 수용자 등

<표 I -1> 각 연도별 조사방식(면접,유치,전화조사)분포

(단위: 명,%)

	개인 응답자수	전체 비중	면접	유치	전화	면접+ 전화	유치+ 전화	면접+ 유치	면접+유 치+전화	웹
1차년도	13,321	96.5	64.4	21.8	1.5	2.6	3.7	2.6	0.1	-
2차년도	12,037	100.0	77.8	7.6	7.4	4.7	2.5	0.1	0.0	-
3차년도	11,205	99.8	83.0	5.1	5.3	3.9	1.5	0.9	0.2	-
4차년도	11,051	100.0	83.7	3.4	4.6	4.9	1.2	1.8	0.5	-
5차년도	10,965	100.0	83.8	2.6	3.9	6.4	1.0	1.0	1.3	-
6차년도	11,540	100.0	86.3	1.7	3.8	5.7	0.9	1.3	0.4	-
7차년도	11,660	100.0	88.0	1.8	4.1	3.7	0.4	1.6	0.5	-
8차년도	11,580	100.0	85.4	1.6	2.8	6.1	1.4	1.7	1.0	-
9차년도	11,756	100.0	89.7	1.5	2.9	2.8	0.7	2.3	0.3	-
10차년도	11,855	100.0	91.3	1.1	3.2	3.1	0.6	0.6	0.1	-
11차년도	11,734	100.0	95.7	0.6	1.8	1.3	0.3	0.2	0.1	-
12차년도	14,489	100.0	93.2	1.5	2.3	2.1	0.5	0.3	0.1	-
13차년도	14,118	100.0	94.5	0.4	2.5	2.2	0.3	0.1	0.1	-
14차년도	13,899	100.0	93.6	0.8	2.9	2.0	0.4	0.3	0.1	-
15차년도	13,998	100.0	94.0	0.4	4.3	1.2	0.1	0.0	0.1	-
16차년도	13,887	100.0	95.6	0.4	2.8	0.9	0.1	0.1	0.0	-
17차년도	13,168	100.0	96.8	0.3	2.1	0.6	0.1	0.1	0.0	-
18차년도	14,011	100.0	95.4	0.5	3.1	0.7	0.2	0.1	0.1	-
19차년도	14,202	100.0	96.5	0.4	1.9	0.6	0.5	0.1	0.0	-
20차년도	14,475	100.0	96.9	0.3	2.1	0.3	0.4	0.0	0.0	-
21차년도	23,972	100.0	95.6	0.3	2.9	0.5	0.3	0.0	0.4	-
22차년도	23,225	100.0	97.1	0.3	1.8	0.6	0.1	0.0	0.0	0.1

→ 개인용 설문지의 응답대상에서 제외 (단, 이들의 인적사항은 가구용 설문지에서 질문함)

□ 질문형태는 면접원이 설문문항을 응답자에게 구두로 질문하고, 이에 대한 응답을 면접원이 기록하는 “**면접타계식**”을 원칙으로 한다. 그러나 심야귀가 등의 원인으로 면접원이 조사대상자를 만나기 어려운 경우에 한하여 유치조사방법을 사용한다. 본 조사 기간 중 응답하지 않은 개인에 대하여 추가적으로 웹조사를 실시한다.

□ <표 I -1>은 이러한 조사원칙이 각 연도별로 실제로 얼마나 지켜지고 있는지를 보여주고 있다.

- 1차년도 당시 면접조사의 비중은 64.4%에 그치고 있으나, 매년 지속적으로 상승하여 22차년도에는 97.1%에 이르렀다.
- 반면 유치조사는 1차년도에 21.8%나 되었지만, 22차년도에는 0.3%로 그 비중이 대폭 감소하였다.
- 22차에 시범 도입된 웹조사는 설문조사 기간 중 응답을 거절한 개인을 대상으로 이루어졌으며, 0.1%(26건)가 자기기입 방식인 웹으로 응답하였다<sup>1)</sup>.

### 2.1.3 신규조사자용 설문대상과 조사방법

□ 신규조사자란, KLIPS에 처음으로 조사된 개인을 의미한다. 신규조사자의 조사 대상은 다음과 같이 분류되며 조사방법은 개인용과 동일하다.

- 대상 1 : 기존가구원 중 작년까지는 만14세였다가 올해 만15세가 된 개인
- 대상 2 : 조사대상이었는데 군복무 등의 이유로 한번도 조사를 하지 못한 개인
- 대상 3 : 분가로 발생하게 된 신규가구원 중 만15세 이상인 개인
- 대상 4 : 기존 가구에 새롭게 추가된 만15세 이상의 개인

---

1) 웹조사와 면접조사 결과 검증에 대한 자세한 내용은 김유빈 외(2020), 『패널자료 품질개선연구(X)』 제5장 「웹을 활용한 조사 자료 검증: 노동패널자료를 중심으로」(이지은·채민희 저술)을 참조하기 바란다. 또한 웹조사 시범 실시 전에 이루어졌던, 웹을 활용한 실험조사는 김유빈 외(2017), 『패널자료 품질개선연구(VII)』 제7장 「웹을 활용한 조사의 효과에 대한 실증연구: 노동패널 CAWI 실험 데이터의 분석」(이지은·신선옥·이혜정·김기홍 저술)과 김유빈 외(2018), 『패널자료 품질개선연구(VIII)』 제3장 「웹을 활용한 조사의 효과에 대한 실증연구 II: 노동패널 CAWI 1차, 2차 실험 연구 결과 분석」(이지은·이혜정·신선옥·정현상 저술)을 참고하기 바란다.

## 2.2 실사 개요

- 한국노동패널의 현장 실사는 전문 조사업체인 한국리서치가 1차년도부터 지속적으로 조사를 담당하고 있다. 조사 기간은 매년 4월부터 9월까지 약 6개월 정도가 소요되며 연인원 100여명 정도의 전문 면접원이 투입되어 조사가 진행된다. 면접원 교육은 매년 조사가 시작되기 전에 해당 지역별로 설문지의 내용에 관한 체계적인 교육을 실시한다. 교육에 투입되는 면접원의 수는 실질적으로 조사수행에 필요한 면접원수보다 20-30%가량 많은데, 그 이유는 교육을 받은 면접원 중에서 20-30%가량의 면접원이 패널조사의 어려움으로 인해 중도에 탈락하기 때문이다.
- 보다 정확한 자료를 수집하기 위하여 1차년도부터 22차년도까지 매년 설문지 재검토과정을 통해 응답이 불성실하거나 지나친 무응답, 응답 간 논리적인 오류 등이 있는지를 체크하여 담당 면접자에게 다시 재설문하도록 하였다.

## 3. 조사결과

### 3.1 조사결과 개요

- 22차년도 조사결과
  - 2019년 22차 조사에 성공한 총 가구수는 11,270가구이며, 이중 98원표본가구는 3,266가구로 65.3%의 표본유지율을 보이고 있으며, 09통합 원표본가구는 5,517가구로 82.1%의 표본유지율을 보인다. 18통합 원표본가구는 12,083가구로 94.5%의 표본유지율을 보인다.
  - 한편, 22차 조사에 총 응답 개인은 23,225명이며, 이중 98표본은 11,880명, 09통합 표본은 13,704명, 18통합표본은 23,011명다.



□ 공표방법

- KLIPS 조사 결과는 매년 공표되며, 공표 시기는 조사기준년도 다음해 12월경이다. 공표 방법은 한국노동패널조사 홈페이지(<https://www.kli.re.kr/klips/index.do>)를 통하여 무료로 제공한다.

<표 I-2> 연도별 조사결과 비교

조사 차수	98표본					09통합표본					18통합표본					전체 성공 가구원 응답자 수	전체 성공 가구원 응답자 수	실사 기간
	조사 대상 원가구- 소멸 가구	조사 성공 가구 수 <sup>1)</sup>	원표본 가구수 2)	원표본 유지율 3)	가구원 응답자 수 <sup>4)</sup>	조사 대상 원가구- 소멸 가구	조사 성공 가구수 5)	원표본 가구수 6)	원표본 유지율 7)	가구원 응답자 수 <sup>8)</sup>	조사 대상 원가구- 소멸 가구	조사 성공 가구수 9)	원표본 가구수 10)	원표본 유지율 11)	가구원 응답자 수 <sup>12)</sup>			
1차('98)	5,000	5,000	5,000	100.0	13,321											5,000	13,321	6~10월
2차('99)	5,000	4,507	4,378	87.6	12,037											4,507	12,037	7~12월
3차('00)	5,000	4,266	4,044	80.9	11,205											4,266	11,205	4~10월
4차('01)	5,000	4,248	3,866	77.3	11,051											4,248	11,051	4~10월
5차('02)	5,000	4,298	3,798	76.0	10,965											4,298	10,965	4~10월
6차('03)	4,991	4,592	3,862	77.2	11,540											4,592	11,540	5~10월
7차('04)	4,947	4,761	3,862	77.2	11,660											4,761	11,660	5~10월
8차('05)	4,935	4,849	3,822	76.4	11,580											4,849	11,580	4~10월
9차('06)	4,914	5,001	3,820	76.4	11,756											5,001	11,756	4~11월
10차('07)	4,899	5,069	3,775	75.5	11,855											5,069	11,855	4~9월
11차('08)	4,882	5,116	3,710	74.2	11,734											5,116	11,734	4~10월
12차('09)	4,833	5,306	3,658	73.2	11,739	6,721	6,721	6,721	100.0	14,489						6,721	14,489	3~익년1월
13차('10)	4,803	5,379	3,607	72.1	11,582	6,691	6,398	6,232	92.7	13,641						6,683	14,118	7~12월
14차('11)	4,787	5,404	3,528	70.6	11,376	6,673	6,388	6,082	90.5	13,409						6,686	13,899	7~익년2월
15차('12)	4,771	5,469	3,517	70.3	11,442	6,640	6,433	6,016	89.5	13,424						6,753	13,998	6~12월
16차('13)	4,742	5,501	3,472	69.4	11,330	6,596	6,455	5,904	87.8	13,300						6,785	13,887	6~12월
17차('14)	4,742	5,552	3,451	69.0	10,756	6,588	6,490	5,840	86.9	12,591						6,838	13,168	3~10월
18차('15)	4,702	5,632	3,421	68.4	11,445	6,530	6,575	5,793	86.2	13,369						6,934	14,011	3~11월
19차('16)	4,687	5,714	3,393	67.9	11,652	6,505	6,632	5,728	85.2	13,517						7,012	14,202	4~11월
20차('17)	4,670	5,761	3,355	67.1	11,879	6,475	6,683	5,674	84.4	13,768						7,066	14,475	3~11월
21차('18)	4,634	5,796	3,309	66.2	11,886	6,430	6,709	5,598	83.3	13,738	12,134	12,134	12,134	100.0	23,972	12,134	23,972	3~12월
22차('19)	4,604	5,823	3,266	65.3	11,880	6,386	6,729	5,517	82.1	13,704	12,083	11,583	11,471	94.5	23,011	11,720	23,225	3~11월

- 주: 1) ‘조사성공가구’는 98표본 원표본가구 중에서 실사가 성공적으로 이루어진 유효표본가구와 조사 성공한 분가가구를 포함한 총 조사성공가구임.
- 주: 2) ‘원표본가구’는 98표본 1차년도 조사 당시 원표본이었던 5,000가구 중 해당 연도에 실사가 성공적으로 이루어진 가구를 의미함.
- 주: 3) ‘원표본유지율’은 98표본 1차년도에 조사에 성공한 원표본 5,000가구 중에서 해당 연도에 조사 성공한 원표본가구의 비율임.
- 주: 4) 98표본 중 15세 이상 가구원으로 면접에 성공한 가구원 수.
- 주: 5) ‘조사성공가구’는 09통합표본 원표본가구(2009년) 중에서 실사가 성공적으로 이루어진 유효표본 가구와 조사 성공한 분가가구를 포함한 총 조사성공가구임.
- 주: 6) ‘원표본가구’는 09통합표본 1차년도(2009년) 조사 당시 원표본이었던 6,721가구 중 해당 연도에 실사가 성공적으로 이루어진 가구를 의미함.
- 주: 7) ‘원표본유지율’은 09통합표본 1차년도(2009년)에 조사 성공한 원표본 6,721가구 중에서 해당 연도에 조사 성공한 원표본가구의 비율임.
- 주: 8) 09통합표본 중 15세 이상 가구원으로 면접에 성공한 가구원 수.
- 주: 9) ‘조사성공가구’는 18통합표본 원표본가구(2018년) 중에서 실사가 성공적으로 이루어진 유효표본 가구와 조사 성공한 분가가구를 포함한 총 조사성공가구임.
- 주: 10) ‘원표본가구’는 18통합표본 1차년도(2018년) 조사 당시 원표본이었던 12,134가구 중 해당 연도에 실사가 성공적으로 이루어진 가구를 의미함.
- 주: 11) ‘원표본유지율’은 18통합표본 1차년도(2018년)에 조사 성공한 원표본 12,134가구 중에서 해당 연도에 조사 성공한 원표본가구의 비율임.
- 주: 12) 18통합표본 중 15세 이상 가구원으로 면접에 성공한 가구원 수.

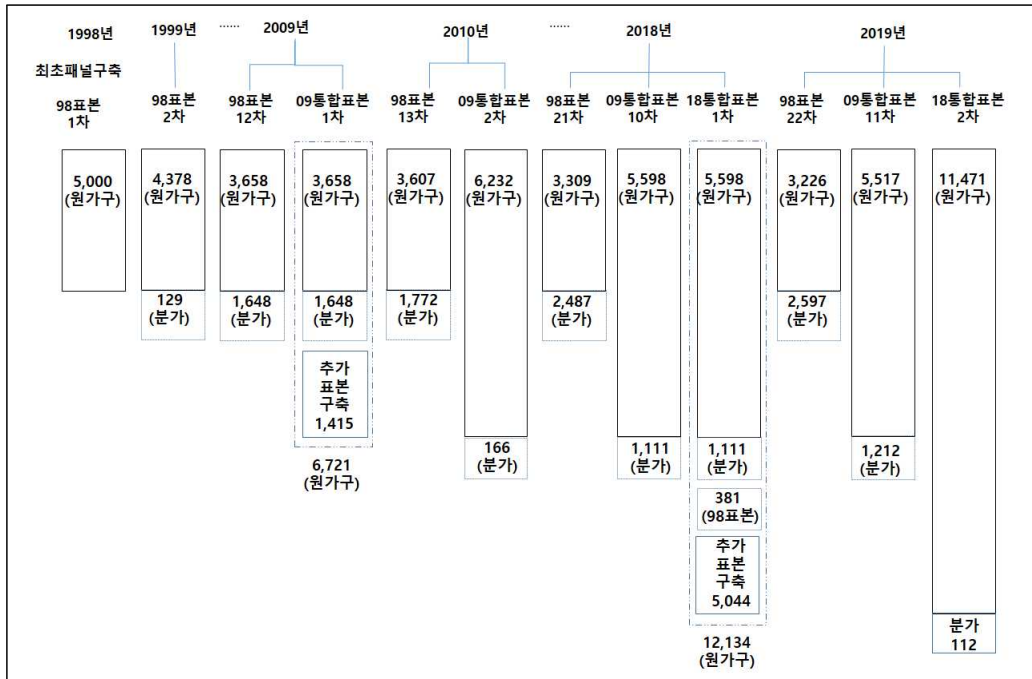
### 3.2 09통합표본, 18통합표본 개요

□ [그림 I-1]은 KLIPS 원표본의 개념을 쉽게 이해할 수 있도록 도식화한 것이다.

- 2019년 현재 KLIPS는 98표본, 09통합표본, 18통합표본으로 구성되어 있다.
- 최초 표본인 98표본은 1998년 5,000가구를 원가구로 구축되었고, 해당 표본은 지속적 추적조사를 통해 2019년까지 총 22차에 걸쳐 조사가 실시되었다.
- 09통합표본은 2009년 조사 당시 추가표집된 1,415가구를 포함하여, 당해연도 응답 가구(6,721가구) 전체를 원가구로 하는 표본을 의미한다.
- 09통합표본은 2009년(12차년도)에 기존 조사대상가구 중 당해연도 응답가구인 5,306가구(98표본 원가구(3,658가구) + 98표본 분가가구(1,648가구))와 추가 표집된 가구인 1,415가구를 더한 6,721가구를 원가구로 하는 표본이다.
- 09통합표본가구에는 98원표본이 다수 포함되어 있지만, 2009년 당시 응답하지

[그림 I-1] 통합표본의 개요

(단위: 가구)



주: 98표본 원가구 중, 2009년도에 조사되지 않은 원가가 다음해인 2010년에 조사된다면, 98표본에는 포함되지만, 09통합표본의 원가구에 속하지 않았으므로 09통합표본에는 포함되지 않음. 따라서 통합표본은 2009년도에 조사된 가구만을 대상으로 추적 조사하는 패널임에 유의해야 함. 18통합패널도 마찬가지로 2018년도 응답한 가구만을 원가구로 정의하였기 때문에, 2018년도에 응답하지 않고, 2019년도 이후에 응답한 가구는 98표본이나, 09통합패널일수는 있지만, 18통합패널에는 속하지 않음에 유의해야 함.

않고, 2010년 이후 다시 응답한 98표본은 포함되지 않으므로, 사용에 주의가 요구된다).

- 18통합표본은 2018년 조사 당시 추가 표집된 5,044가구를 포함하여, 당해연도 응답 가구 전체(12,134가구)를 원가구로 하는 표본을 의미한다.
- 18통합표본은 2018년(21차년도)에 기존 조사대상가구 중 당해연도 응답가구인

2) 09통합패널 원년인 2009년도 당시 응답하지 않고, 이후에 응답하여 09통합표본에는 포함되지 않지만, 98표본에는 포함되는 가구를 찾는 방법은 다음과 같다.

(sample98=1 or sample98=2) and sample09=3

이들 가구는 기본적으로 98표본이지만, 09통합표본은 아니므로, 98표본 가중치는 있지만, 09통합표본 가중치는 주어지지 않는다.

- 7,090가구와 추가표집된 5,044가구를 더한 12,134가구를 원가구로 하는 표본이다.
- 2018년도 기존 응답가구 7,090가구는 09통합표본 원가구(5,598가구) + 09통합표본 분가가구(1,111가구) + 09통합표본에 포함되지 않는 98표본(381가구)이다.
  - 18통합표본가구에는 98원표본과 09통합패널이 다수 포함되어 있지만, 2018년 당시 응답하지 않고, 2019년 이후 다시 응답한 98표본, 09통합표본은 포함되지 않으므로, 사용에 주의가 요구된다<sup>3)</sup>.

□ 22차년도 기준 KLIPS 가구 데이터의 표본가구 구성은 <표 1-3>과 같다.

- 98원가구이면서 09통합표본 원가구인 경우는 3,658가구였으며, 98원표본이지만, 12차년도 당시 비응답가구로 통합표본의 조사대상가구가 아닌 경우가 1,342가구였다.
- 98표본 분가가구 중 12차년도에 응답하여 09통합표본 원표본이 된 가구는 1,648가구였다.
- 2009년~2019년까지 총 11차례에 걸친 조사에서 09통합표본에서 분가한 가구는 1,449가구인데, 이중 1,212가구는 98표본 분가가구이기도 하다.
- 한편, 2009년 당시 추가표집된 1,415가구와 이들 가구에서 분가한 237가구는 98표본에는 속하지 않는다.
- 98표본과 18통합표본의 관계, 09통합표본과 18통합표본의 관계도 동일하게 해석할 수 있다.
- 참고로, 표에는 표시하지 않았지만, 2009년도에 응답하지 않아서 09통합표본에는 포함되지 않지만, 2019년도에 응답한 98표본은 387가구이다. 마찬가지로 개념으로

---

3) 18통합패널 원년인 2018년도 당시 응답하지 않고, 이후에 응답하여 18통합표본에는 포함되지 않지만, 98표본에는 포함되는 가구를 찾는 방법은 다음과 같다.

(sample98=1 or sample98=2) and sample18=3

이들 가구는 기본적으로 98표본이지만, 18통합표본은 아니므로, 98표본 가중치는 있지만, 18통합표본 가중치는 주어지지 않는다.

18통합패널 원년인 2018년도 당시 응답하지 않고, 이후에 응답하여 18통합표본에는 포함되지 않지만, 09통합표본에는 포함되는 가구를 찾는 방법은 다음과 같다.

(sample09=1 or sample09=2) and sample18=3

이들 가구는 기본적으로 09통합표본이지만, 18통합표본은 아니므로, 09통합표본 가중치는 있지만, 18통합표본 가중치는 주어지지 않는다.

2018년도에 응답하지 않아 18통합표본에는 포함되지 않지만, 2019년도 응답한 98표본은 121가구, 09통합표본은 120가구이다

<표 I-3> 표본가구 구성

(단위: 가구)

		09통합표본			전체
		원가구	분가가구	조사대상아님	
98표본	원가구	3,658	0	1,342	5,000
	분가가구	1,648	1,212	801	3,661
	조사대상아님	1,415	237	0	1,652
전체		6,721	1,449	2,143	10,313
		18통합표본			전체
		원가구	분가가구	조사대상아님	
98표본	원가구	3,309	0	1,691	5,000
	분가가구	2,487	65	1,109	3,661
	조사대상아님	6,338	47	0	6,385
전체		12,134	112	2,800	15,046
		18통합표본			전체
		원가구	분가가구	조사대상아님	
09통합표본	원가구	5,598	0	1,123	6,721
	분가가구	1,111	80	258	1,449
	조사대상아님	5,425	32	0	5,457
전체		12,134	112	1,381	13,627

### 3.3 분가가구 및 신규조사자

□ 분가가구란 원표본가구에 속한 가구원이 혼인 등의 사유로 분가하여 새로운 가구를 형성한 신생 가구를 의미한다.

- 분가가구는 원표본가구원이 1인이더라도 포함된 경우에만 조사하는 것을 원칙으로 한다.
- KLIPS의 경우 분가가구는 세 가지 형태로, 98원표본가구에서 분가한 경우와 09통합표본에서 분가한 경우, 18통합표본에서 분가한 경우이다. 물론 98원표본가구이면서 통합표본가구인 가구가 존재하므로, 이들 가구의 경우에는 양쪽 모두의 분가가

구가 되므로, 주의가 요구된다.

□ 분가가구 실사 결과

- 먼저, 98표본을 기준으로 할 때, KLIPS 22차년도 조사성공가구(5,823가구) 중 2차년도 이후 분가하여 새롭게 조사에 진입한 비원표본가구는 2,557가구였다.
- 이들 분가가구는 매해 100가구 내외 증가하고 있으며, 22차년도 98표본 조사성공가구의 43.9%에 달한다.
- 한편, 09통합표본의 경우에는 22차년도 조사성공가구 6,729가구 중 분가가구가 1,212가구인 것으로 나타났다.
- 18표본의 경우에는 22년도 조사성공가구 12,134가구 중 분가가구가 112가구이다.

□ 신규조사자란 KLIPS에 처음으로 조사되는 개인으로서, 만15세가 되어 조사에 처음으로 포함되었거나 전년도까지 조사되지 못하여 처음으로 조사된 개인 혹은 분가 등의 사유로 처음으로 조사된 개인을 의미한다.

- 주된 사유는 만 15세가 되어 처음 개인조사에 응답하는 경우와 혼인이나 합가로 인해 조사년도에 새롭게 조사대상자로 추가되는 경우이다.

## Ⅱ. 표본설계 및 가중치

### 1. 표본설계방법 및 표본추출과정

#### 1.1 표본틀과 모집단

##### ☐ 표본틀

- KLIPS는 1995년 인구주택총조사의 10%표본조사구인 전국의 21,938조사구 중에서 제주도의 263조사구와 군부지역 2,650조사구를 제외한 전국 도시지역(시의 동부 17,353조사구와 시의 읍면부 1,672조사구) 19,025조사구를 표본틀로 삼았다.

##### ☐ 모집단

- KLIPS표본의 모집단은 우리나라 도시 거주 가구와 15세 이상 개인이다.
- 제주도와 병역의무이행자, 시설거주가구(원)는 조사대상에서 제외된다.

#### 1.2 표본추출과정

##### 1.2.1 98표본추출과정<sup>4)</sup>

- ☐ KLIPS의 표본추출방법은 1단계에서 조사구를 선정하고 2단계에서 가구를 선정하는 2단계 층화집락계통추출법을 사용하였다.

- 이를 위해 먼저 제주도를 제외한 전국의 시부만을 대상으로 1,000개의 조사구를 선정하고 각 조사구 내에서는 5가구를 무작위 추출(random sampling)하는 원칙을 정하였다.
- 예산제약의 문제로 말미암아 1995년 인구주택총조사를 이용해 최종 조사가구를

---

4) 「한국 가구와 개인의 경제활동-한국노동패널 1차년도 자료분석-」 보고서(방하남, 1999년)를 요약 정리함



선정하는 과정에서 1997년 통계청이 수행한 고용구조특별조사의 조사가구가 선정되도록 매칭하였다.

□ 표본추출 과정은 다음과 같다.

- 첫째, 지역을 기준으로 1997년 고용구조특별조사(이하 97고특)와 똑같은 방법으로 층(strata)을 구분하였다. 97고특의 층화방법을 따르지만, 시부로 구성된 층만 포함된다.
- 둘째, 각 층에서 조사할 조사구수는 97고특을 따르지 않고, 인구주택총조사 10% 표본조사구에서의 전국도시조사구(19,025개) 대비 해당 지역의 도시조사구의 비율을 이용하여 추출할 도시조사구수를 결정하였다. 97고특은 인구크기비례, 실업자수의 상대오차, 예산 등을 고려한 보다 복잡한 절차에 따라 층별 조사구수를 할당하였으므로, 인구크기비례만 고려한 썸인 KLIPS와는 각 층별로 필요한 조사구수가 다르다. 이를테면, 서울지역의 경우 고용구조특별조사의 조사구수보다 더 많은 조사구수가 필요하였다.
- 각 층 안에서 조사구를 추출할 때에는 집락추출의 원리가 이용되었다. 이를테면, 한 층에 100개의 조사구가 있다면 모든 조사구에서 한두가구씩 최종 조사가구를 선정한 것이 아니라 일부 조사구만 선택해 그 안에 있는 가구를 조사대상으로 선정했다는 의미이다. 이는 조사비용의 절감을 위해 통계청 등 대부분의 가구조사에서 활용하는 방식이다.
- 셋째, 각 층에서 조사구를 선택하는 방법으로는 계통추출법(systemic sampling)을 이용하였다. 계통추출법은 조사구를 줄세운 다음 등간격으로 일정한 수의 조사구를 표집하는 방식이다. 조사구를 줄세우려면 번호를 할당해야 하는데, 이 번호할당방식도 97고특을 그대로 따랐다. 추출된 조사구가 97고특의 조사구였으면 조사 대상 조사구로 선정했고, 97고특의 조사구가 아닌 경우에는 가장 가까운 97고특의 조사구로 조사 대상 조사구를 대체 선정하였다.
- 넷째, 추출된 조사구 중에서 97고특에서 조사에 성공한 가구를 대상으로 서울 및 경기도의 읍면부 지역에서는 6가구, 기타지역은 5가구를 임의추출(random sampling)하였다.

- 서울 및 경기도의 경우에는 추출하여야 할 조사구수보다 97고특에서 사용한 조사구수가 적었기 때문에 다른 조사구에 비해 1가구를 더 추출하였으며, 추가하여 경기도의 동부의 4개 지역에서 5가구 대신 6가구를 추출하였다.
- 참고로, 97고특은 서울, 부산, 대구, 인천의 경우 조사구의 총 가구수가 40가구 이하인 경우 40가구 전체를 모두 조사하였고, 총 가구수가 40가구 이상인 경우 약 10개 정도의 가구를 하나의 집락으로 묶어 평균 6개 정도의 집락으로 한 조사구를 분할한 후 임의추출로 1개 집락을 선택해 그 집락 북쪽에 있는 세 집락 전체를 조사했다. 나머지 지역은 한 조사구당 50가구 기준으로 위와 같이 했다.

□ 최종적으로 추출된 KLIPS의 5,000가구는 시의 동부 4,582가구 + 시의 읍면부 418가구로 구성된다.

<표 II-1> 최종 추출된 조사구와 조사가구

	전체 모집단		10%표본조사구		표본		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

□ 2단계에서 추출한 목표표본가구 5,000가구 중 11.3%인 563가구는 이사를 하였으며, 이중 40.1%는 추적 조사에 성공하였다.

※ 최초 표본 추출 단계에서 군부는 제외되었으나, 이후 이사가구 추적 조사 과정에서 군부로의 이주 가구가 발생하여 최종 표본에는 군부지역 가구가 포함되었다.

□ 5,000가구 중 75.5%인 3,773가구를 조사 성공하였다. 강력거절, 이사 후 추적불가 등의 사유에 따라 대체된 가구는 1,227가구인데, 대체를 위해 1,976가구를 접촉하였다. 대체가구 성공률은 62.1%였다. <표 II-2>는 대체과정을 거친 조사구별 최종 응답가구의 가구수와 가구비율이다.

<표 II-2> KLIPS 1차년도 조사가구 개요

지역	총접촉 가구수	최초접촉 가구수	최초성공 가구수	대체가구 접촉수	대체성공 가구수	최초성공 가구비율	대체성공 가구비율
서울	2,109	1,362	954	747	408	70.0	30.0
부산	600	485	391	115	94	80.6	19.4
대구	452	320	239	132	81	74.7	25.3
인천	414	295	226	119	69	76.6	23.4
광주	197	160	134	37	26	83.8	16.3
대전	188	155	127	33	28	81.9	18.1
울산	150	120	100	30	20	83.3	16.7
경기	1,217	853	617	364	236	72.3	27.7
강원	200	130	94	70	36	72.3	27.7
충북	145	110	86	35	24	78.2	21.8
충남	152	125	107	27	18	85.6	14.4
전북	273	200	151	73	49	75.5	24.5
전남	164	125	101	39	24	80.8	19.2
경북	356	270	205	86	65	75.9	24.1
경남	359	290	241	69	49	83.1	16.9
전국	6,976	5,000	3,773	1,976	1,227	75.5	24.5

### 1.2.2 09통합표본추출과정<sup>5)6)</sup>

□ 2008년 11차년도 조사에서는 원가구 5,000가구 및 분가가구 2,531가구의 총 7,531패널가구 중 소멸가구와 5차년도 이후 무응답을 제외한 나머지 6,027가구가 패널 표본으로 유지되어 80.03%의 표본유지율을 기록하고 있다.

- 표본이탈 문제는 결국 현 표본의 대표성에 한계를 나타내었고 또한 지역 통계 생산 및 보다 심층적인 분석을 위해서는 현행 표본의 규모가 부족한 점이 있다. 이를 해결하기 위해 현재 모집단에 대한 대표성을 보완하도록 표본을 추가할 필요성이 있다.
- 또한 지금까지 KLIPS의 목표 모집단을 도시지역 가구로 한정하여, 현행 표본이 비도시에서 도시로 이동하는 가구에 대한 대표성을 확보하지 못한다는 단점이 있다.

5) 「표본추가 연구」 보고서(김재광, 2009년 3월)를 요약 정리함

6) 참고로 「한국노동패널 오차분석 및 표본추가」 보고서(김재광, 2003년 6월)에는 표본 추가 작업과 관련하여 필요한 여러 통계학적인 사항을 점검함

따라서 목표 모집단을 도시지역 가구가 아닌 전국 가구로 확장하여 그 모집단의 정의에 대해 대표성을 가질 수 있도록 표본을 추가할 필요가 있다.

□ 2009년(12차년도) 추가표본의 조사 모집단 및 조사 대상 가구 설정은 다음과 같다.

- 조사 모집단은 2009년 3월 현재 대한민국 영토(제주도 포함) 내의 일반 가구에 거주하고 있는 가구원이다.
- 이러한 조사 모집단에 대한 추출 프레임으로는 2005년 센서스 조사구 자료를 사용하였다.

□ 표본수의 결정 및 표본 배정은 다음과 같다.

- 먼저 센서스 모집단 자료와 KLIPS 자료에서 동부와 읍면부의 분포를 동일하게 맞춰 주기 위해서, 동부에서 900가구 및 읍면부에서 600가구의 표본 추가가 필요하였다. <표 II-3>을 보면 목표 표본수는 7,085가구(전국)이다.

<표 II-3> 표본추가 후 KLIPS의 동부·읍면부별 분포

	가구수	%
동부	5,737	80.97
읍면부	1,348	19.03
전국	7,085	100.0

- 이후 지역별(동/읍면부, 시도)분포를 고려하여 비례 배정으로 표본을 배정하였다. 센서스 조사구를 1차 표본 추출 단위로 하고 가구를 2차 표본 추출 단위로 하는 2단계 집락 추출을 사용하였으므로 실제로 배정된 조사구 단위 표본수는 <표 II-4>와 같다. 여기서는 각 조사구당 5개의 표본 가구를 추출하는 것을 원칙으로 하였다.

<표 II-4> 최종 배정 표본 조사구

(단위: 가구)

광역시/도	동부	읍면부
서울특별시	37	0
부산광역시	13	0
대구광역시	4	1
대전광역시	6	0
인천광역시	8	0
광주광역시	5	0
울산광역시	5	1
경기도	48	10
강원도	5	12
충청북도	6	14
충청남도	6	17
전라북도	5	6
전라남도	3	25
경상북도	10	13
경상남도	8	17
제주도	11	4
총조사구수	180	120

□ 표본 추출 과정은 다음과 같다.

- 표본 추출은 동부와 읍면부에서 다르게 적용하였다. 동부에서는 기존 자료의 가구 원수 분포와 주택유형별 분포가 센서스의 분포와 지나치게 차이가 나므로 이를 보정해주는 표본 추출법을 사용하였다.

- 1단계 : 2005년 센서스 자료에서 조사구 단위로 지역(시도), 총가구수, 가구규모별 가구수, 주택 유형별 가구수를 얻어낸다. 조사구별 특성치는  $((1\sim 2인\ 가구수) + (단독주택\ 가구수))$ 로 구한다.
- 2단계 : 지역 내에서 모집단 조사구를 지역 내에서 특성치에 비례하는 PPS 추출(Probability Proportionate Sampling, 확률비례추출)을 실시한다. 즉, 각 지역 내에서 조사구가 추출될 확률이 특성치에 비례하도록 추출한다.
- 3단계 : 이렇게 해서 얻어진 표본 조사구에서 5개의 가구를 계통 추출로 추출하여 본표본으로 지정하고 나머지에서 10개 가구를 예비 표본으로 지정한다.

- 읍면부 추출 역시 층화 2단계 집락 추출을 사용하였으나 동부에서 사용한 특성치를 사용하지 않고 조사구별 가구수에 비례하는 PPS 추출을 사용하였다. 각 표본 조사구에서 5개의 가구를 계통 추출로 추출하여 본표본으로 지정하고 나머지에서 10개 가구를 예비 표본으로 지정한다.

□ 추가 표본으로 최종 1,415가구가 구축되었다.

### 1.2.3 18통합표본추출과정<sup>7)</sup>

□ 18통합표본을 위한 추가 표본추출은 2017년 예비표본추출과 2018년 추가 표본추출로 나누어져 진행되었고 그 중 2017년 예비표본은 다음의 과정을 통해 추출되었다.

□ 2015년 18차년도 조사 기준 가구 표본의 현황은 [표 II-5]와 같다.

- 20여년 이상의 조사기간 동안의 표본이탈의 문제 등으로 표본이 모집단을 반영하지 못하는 일이 발생하면서, 이를 해결하기 위하여 또 한 번의 표본 추가를 통한 패널 확대가 요구되었고, 2017년과 2018년에 걸쳐 표본 추가 연구 및 표본추출이 이루어져 조사가 진행되었다.

- 2015년 조사의 표본 현황을 표본별로 보면, ① 98표본의 원표본유지율은 2015년(18차) 조사에서 68.4%를 유지하고 있으며, ② 09통합표본의 원표본유지율은 86.2%이다.

□ 2018년(21차년도) 추가표본의 조사 모집단 및 조사 대상 가구 설정은 다음과 같다.

- 조사 모집단은 2016년 11월 현재 대한민국 영토(제주도 포함) 내의 일반 가구에 거주하고 있는 가구원이다.
- 이러한 조사 모집단에 대한 추출 프레임으로는 2016년 등록 센서스 조사구 자료를 사용하였다.

7) 「패널자료 품질개선 연구(IX) 2장」 보고서(변종석·이혜정, 2019년 12월)

<표 II-5> 1~18차 조사결과

조사 차수	98표본					09통합표본				
	조사대상 원가구- 소멸가구	조사 성공 가구수1)	원표본 가구수2)	원표본 유지율3)	분가 가구	조사대상 원가구- 소멸가구	조사 성공 가구수5)	원표본 가구수6)	원표본 유지율7)	분가 가구
1차('98)	5,000	5,000	5,000	100.0	-					
2차('99)	5,000	4,507	4,378	87.6	129					
3차('00)	5,000	4,266	4,044	80.9	222					
4차('01)	5,000	4,247	3,866	77.3	382					
5차('02)	5,000	4,298	3,798	76.0	500					
6차('03)	4,991	4,592	3,862	77.2	730					
7차('04)	4,947	4,761	3,862	77.2	899					
8차('05)	4,935	4,849	3,822	76.4	1,027					
9차('06)	4,914	5,001	3,820	76.4	1,181					
10차('07)	4,899	5,069	3,775	75.5	1,294					
11차('08)	4,862	5,116	3,710	74.2	1,406					
12차('09)	4,833	5,306	3,658	73.2	1,648	6,721	6,721	6,721	100.0	-
13차('10)	4,803	5,379	3,607	72.1	1,772	6,691	6,398	6,232	92.7	166
14차('11)	4,787	5,404	3,528	70.6	1,876	6,673	6,388	6,082	90.5	306
15차('12)	4,771	5,469	3,517	70.3	1,952	6,640	6,433	6,016	89.5	417
16차('13)	4,742	5,501	3,472	69.4	2,029	6,596	6,455	5,904	87.8	551
17차('14)	4,742	5,552	3,451	69.0	2,101	6,588	6,490	5,840	86.9	650
18차('15)	4,702	5,632	3,421	68.4	2,211	6,530	6,575	5,793	86.2	782

□ 표본수의 결정 및 표본 배정은 다음과 같다.

- 변화된 모집단 특성을 반영하고, 변형된 패널을 보완하기 위해 다음과 같은 과정으로 표본 크기를 결정하고, 배분한다.

과정 1: 전체 패널의 규모는 13,000가구로 설정한다. 이는 현재 약 7,000가구가 운영되는데, 고용 및 노동 관련하여 연구진 및 외부, 정책적 합의 도출을 위해 요구되는 수준을 고려하여 약 6,000가구를 추가하기로 결정된 바를 반영한 규모이다.

과정 2: 전체 패널 13,000가구에 대한 층별 필요한 패널 규모를 배분한다. 층은 앞서 설명한 바와 같이 광역시도x동읍면부x주택유형별로 가구수 규모의 비례배분으로 표본을 배분한다.

과정 3: 현재 유지되는 층별 패널 규모를 파악한다.

과정 4: 과정 2를 통해 배분된 층별 표본 규모에서 과정 3에서 파악한 현재 유지되고 있는 패널 규모와의 과부족을 계산하여 새롭게 추가해야 할 표본 규모를 결정한다. 이때, 초과하는 부분은 계속 유지하도록 하고, 부족한 층에 대해서만 추가 표본 규모로 확정한다.

- 최종 패널 가구의 분포는 다음의 표와 같다.

<표 II-6> 패널 추가 및 17예비표본의 가구 규모

광역시도 및 읍면부		모집단			현행 패널 규모			예비표본 추출 규모			17예비표본 규모		
		가구수	일반	아파트	계	일반	아파트	계	일반	아파트	계	일반	아파트
전국	소계	19,100,262	9,866,097	9,234,165	7,012	3,615	3,397	6,000	3,110	2,890	13,012	6,725	6,287
	동부	15,597,553	7,407,000	8,190,553	5,673	2,684	2,989	4,930	2,330	2,600	10,603	5,014	5,589
	읍면부	3,502,709	2,459,097	1,043,612	1,339	931	408	1,070	780	290	2,409	1,711	698

□ 표본 추출 과정은 다음과 같다.

- 새로운 표본설계에서의 추출단위는 다음과 같다. 새로운 표본설계에서는 층별로 조사구를 추출하고, 조사구에서 10가구를 추출하기로 한다. 필요한 조사구의 추출 규모는 약 600개 조사구이다.

1차 추출단위: 조사구

2차 추출단위: 가구

#### (1) 1단계 추출 과정

- 1차 추출단위인 조사구의 추출 과정은 다음과 같은 확률비례계통추출법으로 추출한다.
- 1단계: 광역시도, 동읍면부, 및 주택유형의 층별로 조사구는 지역순(동읍면부-구시군-동읍면)-가구수의 순으로 정렬한다.
- 2단계: 조사구의 가구수 규모에 확률비례하는 확률비례계통추출법으로 조사구



를 추출한다.

(2) 2단계 추출 과정

- 2차 추출단위인 가구는 표본 조사구에서 가구 명부를 작성하여 계통추출법으로 10가구를 추출한다. 이때, 표본가구 접촉 실패 혹은 조사 참여를 거부하는 등이 패널 확보에 실패하는 경우는 바로 인접한 가구로 대체하도록 한다.
- 조사 환경이 악화되어 조사구별 10가구 확보가 어려우며, 또한 조사구 교체의 최소화를 위해 조사구내에서 확보해야 할 가구 수는 조사구의 규모에 따라 6~14가구를 확보하도록 한다. 조사구에서 5가구 이하이면 조사구를 교체하도록 하는 기준을 설정하였다.

□ 2017년 예비표본으로 최종 6,034가구<sup>8)</sup>가 구축되었다.

□ 2018년도 추가 추출 사유와 구축과정은 다음과 같다.

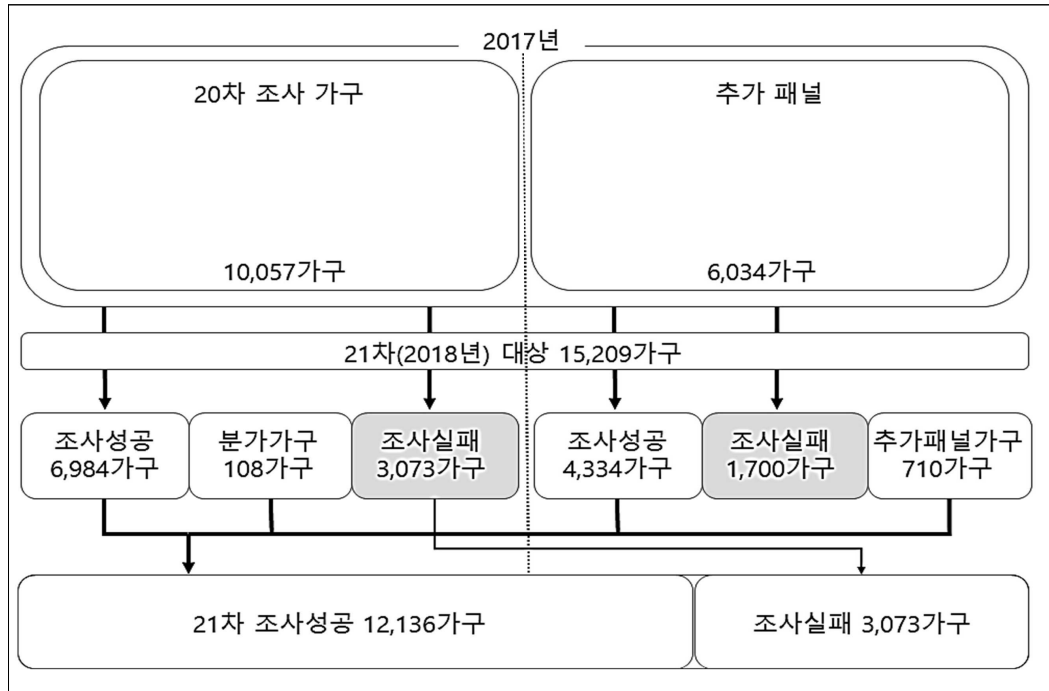
- 고용노동부와 한국노동연구원은 통합패널 구축과정에서 6,000가구를 추가하기로 논의하였고 2017년 예비조사에 6,034가구가 추가 표본으로 구축되었다. 그러나 2018년 21차 조사에서 2017년 예비표본(6,034가구)을 대상으로 재접촉하였으나 조사 환경의 악화, 이사 및 재조사 거절 등의 이유로 응답률이 72%에 그쳐, 4,334가구만이 확보되었다. 이를 감안하여 21차 조사의 통합패널 구축을 위한 조정된 최소 신규패널 규모를 5,000가구로 재설정하였으나 최소 목표로 재설정한 5,000가구를 확보하지 못하는 상황이 발생되었다.
- 고용노동부와 한국노동연구원은 재설정된 목표 규모인 5,000가구를 확보하기 위해 2018년에 710가구를 추가하였고 최종적으로 18통합표본의 1차년도(2018년 21차조사)에 확보한 신규표본은 5,044가구로 확정되었다(그림 II-1 참고).
- 2018년도 추가 표본(710가구) 추출방법은 207년 예비표본 추출 방법과 동일하다.

8)

광역시도 및 읍면부		계	일반	아파트
전국	소계	6,034	3,063	2,971
	동부	4,958	2,279	2,679
	읍면부	1,076	784	292

□ 18통합표본을 위한 최종 추가 표본은 5,044가구이다.

[그림 II-1] 21차(2018년) 조사대상 규모



## 2. 가중치 부여방식 및 사용방법<sup>9)</sup>

### 2.1 98표본 가중치 산출

#### 2.1.1 로지스틱 회귀분석을 이용한 응답모형 설정

- $t$ 차년도에 가구원의 응답여부를 나타내는 변수를 응답과 무응답의 가능한 두 범주를 갖는 이항 확률 변수로 정의하고, 응답의 예측을 위하여 로지스틱 회귀모형을 고려한다.  $t$ 시점의 가구원의 응답확률은 전년도의 자료를 이용하여 예측된다. 이는  $t$ 시점에 종단면 가중치가 원칙적으로 부여되는 대상은  $t$ 시점까지 모두 조사가 된 가구원이다. 각 조건부 응답확률은 로지스틱 회귀모형을 적합하여 예측된다.
- 성별, 지역, 학력, 주된 활동, 만 나이의 변수들이 로지스틱 회귀모형의 적합을 위하여 사용되었다.

#### 2.1.2 98표본 가중치 산출과정

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

##### - 1단계 : 종단면 가구원 가중치 산출

- 무응답 보정을 위해 로지스틱 회귀분석을 실시하고 이를 통하여 무응답 보정 가중치를 산출한다. 이 과정에서 생성된 가중치를 기초가중치로 정의하며 이를 이용하여 스케일 조정전의 종단면 가중치를 생성한다. 여기서 산출된 종단면 가중치에 추계인구증가율을 이용한 스케일 조정을 통해 최종 종단면 가중치가 산출된다.

##### - 2단계 : 가구 가중치 산출

- 기본적으로 1단계에서 산출된 기초 가중치들의 각 가구별 평균으로 정의한다.

---

9) 「한국노동패널 가중치 연구」 보고서(박민규 외, 2013년 12월)를 요약 정리함

여기서 산출된 가구 가중치에 추계 가구 수 정보를 이용한 스케일 조정 지수를 적용하여 최종 가구 가중치가 산출된다.

- 3단계 : 횡단면 가구원 가중치 산출

- 횡단면 가구원 가중치는  $t$ 시점에 원가구원을 한 명이상 포함하고 있는 가구에 속하는 모든 가구원에게 부여된다.
- 먼저 스케일 보정 전의 가구 가중치를 각 가구원에게 부여한 다음, 이 값에 스케일 조정을 적용하여 최종 횡단면 가구원 가중치를 산출한다. 여기서 스케일 조정은 추계인구증가율이 고려된 종단면 가구원 가중치를 이용하였다.

□ 이러한 계산과정을 거친 최종 가중치 값은 <표 II-7> 및 <표 II-8>와 같다.

- <표 II-7>를 통해 가구 가중치의 특성을 살펴보면, 1차년도 당시 가구당 평균적으로 2,255가구 정도를, 22차년도는 2,834가구를 대표하는 것으로 나타났다.

- 조사차수가 늘어날수록 표준편차가 커지는 경향이 있다.

- <표 II-8>에서는 개인 가중치의 특성을 확인할 수 있다.

- 각각의 가중치가 복원하는 모집단의 수(아래 표에서의 합계치)는 동일하다.
- 가구 가중치와 마찬가지로 개인 가중치의 표준편차 역시 조사차수가 늘어날수록 커지는 경향이 있다.

<표 II-7> 98표본 가구 가중치 특성

조사차수	사례수	합계	평균	표준편차
98표본 1차	5,000	11,276,899	2,255.38	415.05
98표본 2차	4,507	11,527,038	2,559.86	565.09
98표본 3차	4,266	11,820,903	2,774.86	714.99
98표본 4차	4,247	12,039,541	2,846.23	852.02
98표본 5차	4,298	12,311,848	2,879.96	975.85
98표본 6차	4,592	12,575,967	2,751.25	1,047.69
98표본 7차	4,761	12,819,131	2,695.93	1,140.03
98표본 8차	4,849	13,069,116	2,700.23	1,261.61
98표본 9차	5,001	13,299,913	2,665.31	1,349.71
98표본 10차	5,069	13,527,761	2,674.00	1,463.65
98표본 11차	5,116	13,757,297	2,694.87	1,632.85
98표본 12차	5,306	13,989,092	2,644.94	1,770.06
98표본 13차	5,379	14,255,545	2,659.12	1,922.62
98표본 14차	5,404	14,568,797	2,706.95	2,105.84
98표본 15차	5,469	14,763,703	2,705.96	2,464.89
98표본 16차	5,501	14,983,690	2,732.26	2,374.35
98표본 17차	5,552	15,241,577	2,767.17	2,607.61
98표본 18차	5,632	15,492,482	2,770.47	2,889.90
98표본 19차	5,714	15,713,937	2,773.37	3,207.79
98표본 20차	5,761	15,908,615	2,786.10	3,503.80
98표본 21차	5,796	16,094,578	2,808.83	3,884.59
98표본 22차	5,823	16,273,478	2,834.61	4,037.15

주: 가중치가 0인 경우는 제외한 후, 평균 및 표준편차 산출

<표 II-8> 98표본 개인 가중치 특성

조사차수	라벨	사례수	합계	평균	표준편차
98표본 1차	가구원	13,321	30,039,802	2,255.07	416.77
98표본 2차	종단면	12,037	30,254,084	2,584.05	517.32
	횡단면	12,037	30,254,084	2,514.68	562.78
98표본 3차	종단면	11,205	30,508,123	2,825.61	685.82
	횡단면	11,205	30,508,123	2,725.40	700.41
98표본 4차	종단면	11,051	30,743,095	2,927.91	833.23
	횡단면	11,051	30,743,095	2,789.50	826.60
98표본 5차	종단면	10,965	30,921,291	3,003.23	979.95
	횡단면	10,965	30,921,291	2,830.84	939.88
98표본 6차	종단면	11,540	31,081,979	2,915.21	1,060.09
	횡단면	11,540	31,081,979	2,702.78	1,002.30
98표본 7차	종단면	11,660	31,205,411	2,916.39	1,196.22
	횡단면	11,660	31,205,411	2,678.35	1,099.05

조사차수	라벨	사례수	합계	평균	표준편차
98표본 8차	종단면	11,580	31,271,636	2,972.02	1,378.23
	횡단면	11,580	31,271,636	2,702.82	1,224.10
98표본 9차	종단면	11,756	31,436,307	2,967.93	1,510.65
	횡단면	11,756	31,436,307	2,677.48	1,307.10
98표본 10차	종단면	11,855	31,595,535	2,982.96	1,620.79
	횡단면	11,855	31,595,535	2,668.54	1,390.80
98표본 11차	종단면	11,734	31,836,359	3,057.37	1,925.83
	횡단면	11,734	31,836,359	2,716.41	1,574.50
98표본 12차	종단면	11,739	32,000,637	3,089.76	2,177.76
	횡단면	11,739	32,000,637	2,730.66	1,743.13
98표본 13차	종단면	11,582	32,160,470	3,173.84	2,493.84
	횡단면	11,582	32,160,470	2,782.05	1,935.59
98표본 14차	종단면	11,376	32,408,729	3,280.57	2,843.88
	횡단면	11,376	32,408,729	2,856.15	2,156.93
98표본 15차	종단면	11,442	32,579,554	3,303.21	3,524.14
	횡단면	11,442	32,579,554	2,851.35	2,528.97
98표본 16차	종단면	11,330	32,728,201	3,369.87	3,393.88
	횡단면	11,330	32,728,201	2,894.00	2,479.83
98표본 17차	종단면	10,756	32,934,430	3,588.80	3,854.39
	횡단면	10,756	32,934,430	3,076.26	2,844.28
98표본 18차	종단면	11,445	33,108,548	3,429.87	4,269.83
	횡단면	11,445	33,108,548	2,906.81	3,040.40
98표본 19차	종단면	11,652	33,240,201	3,410.65	4,482.92
	횡단면	11,652	33,240,201	2,869.24	3,201.06
98표본 20차	종단면	11,879	33,333,727	3,378.30	4,914.37
	횡단면	11,879	33,333,727	2,823.70	3,442.52
98표본 21차	종단면	11,886	33,492,550	3,407.87	5,767.52
	횡단면	11,886	33,492,550	2,839.80	3,858.45
98표본 22차	종단면	11,880	33,559,050	3,441.60	6,204.09
	횡단면	11,880	33,559,050	2,849.54	4,088.52

주: 가중치가 0인 경우는 제외한 후, 평균 및 표준편차 산출

## 2.2 09통합표본 가중치 산출

### 2.2.1 09통합표본 구축 현황

□ 종단적 조사가 진행되면서 표본탈락(panel attrition)이 발생한다. 표본탈락은 가구 차원에서 발생하거나 개인차원에서 발생할 수 있으며, 이와 함께 항목무응답(item nonresponse)도 발생할 수 있다. 항목무응답의 경우는 가중치부여 방법보다는 보

정(imputation)방법이 주로 사용된다.

- 결혼, 이민 등으로 인해 표본가구에 새로운 개인이 포함되거나 탈락하기도 하며, 표본가구원 사이에서 새로운 가구원인 비표본가구원(nonsample household)이 탄생하기도 한다. 이에 따라 종단적으로 조사가 진행되면서 가구원 구성이 달라지어 종단면가중치를 부여할 수 있는 가구의 개념이 모호해진다.

### 2.2.2 09통합표본의 1차년도 가중치 산출 결과

- 09통합표본 가중치 산출 과정은 2.1에서 설명된 98표본 가중치 산출 방안과 그 순서와 구조는 기본적으로 동일하다.
  - 09통합표본 가중치는 표본을 추가한 시점인 2009년부터 산출하였다.
  - 09통합표본은 98표본과 달리 전국 모집단을 대표하므로 98표본에서의 스케일 조정 대신에 인구수 및 가구수에 대한 추계자료를 활용한 사후층화가 고려되었다.
- 09통합표본 구축 시점인 1차년도에는 개인의 종단가중치와 횡단가중치가 같다고 가정한다. 2009년 추계인구를 기준으로 사후조정된 값을 09통합표본 1차년도 개인 종단가중치로 정의한다. 또한 09통합표본의 가구가중치도 2009년 추계가구로 사후조정한다.

### 2.2.3 로지스틱 회귀분석을 이용한 응답 모형 설정

- 09통합표본에서의 응답모형도 98표본과 동일하므로 기본적인 응답 모형 형태는 로지스틱 회귀모형을 사용하였다.
  - 응답 모형에 대한 설명변수로 성별, 지역, 학력, 주된 활동, 만나이, 추가패널여부(09통합표본 구축 시 1998년 또는 2009년 패널인지 여부)를 분석에 이용한다.
  - 일부 차수에서 응답자의 학력, 주된 활동, 만나이에 대해 무응답이 존재한다. 응답 모형 설정 과정에서 설명변수의 무응답이 존재하는 경우에는 무응답이 속한 다른 설명변수의 평균 응답률로 대체하였다.

#### 2.2.4 09통합표본 가중치 산출과정

□ 09통합표본의 가중치 산출은 기본적으로 3단계(1단계 : 종단면 가구원 가중치 산출 → 2단계 : 가구 가중치 산출 → 3단계 : 횡단면 가구원 가중치 산출)의 과정으로 구성된다.

- 1단계 : 종단면 가구원 가중치 산출

- 로지스틱 회귀분석을 통하여 무응답 보정 가중치를 산출하고, 이 과정에서 생성된 가중치를 기초가중치로 정의하며 이를 이용하여 종단면 가구원 가중치를 생성한다. 여기서 산출된 종단면 가중치를 최종적으로 통합연도 기준 인구주택총조사(2010년) 자료 혹은 추계인구를 이용하여 사후조정하며 이를 최종 개인 종단면 가중값으로 정의한다.

- 2단계 : 가구 가중치 산출

- 가구 가중치는 기본적으로 1단계에서 산출된 기초 가중치들의 각 가구별 평균으로 정의된다. 이 가중치에 인구주택총조사 자료의 가구수 및 추계가구수 정보를 이용하여 사후조정한 가구 가중치를 최종적으로 계산한다.

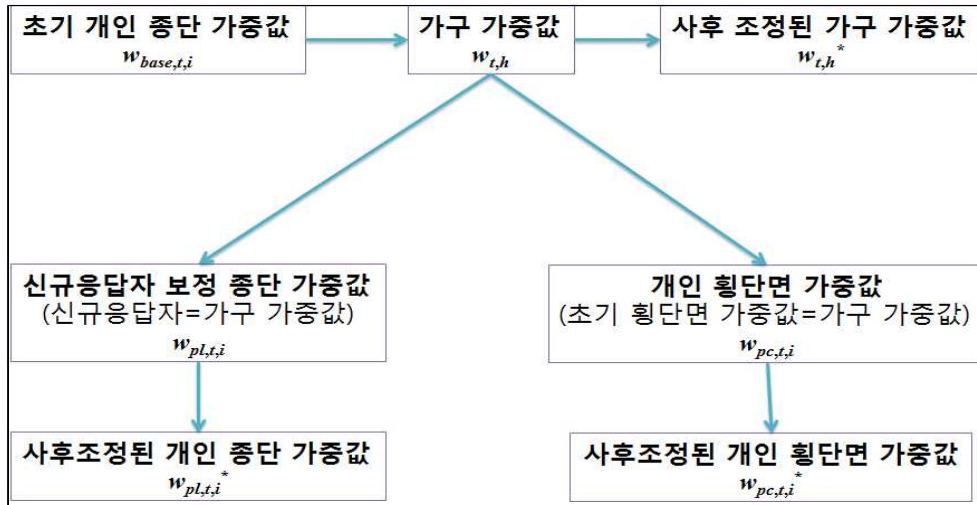
- 3단계 : 횡단면 가구원 가중치 산출

- 횡단면 가구원 가중치는  $t$ 시점에 원가구원을 한 명이상 포함하고 있는 가구에 속하는 모든 가구원에게 부여된다.
- 먼저 사후조정전의 가구가중치를 각 가구원에게 부여한 다음, 이 값에 사후조정을 적용하여 산출된다. 사후조정은 09통합표본 2~9차년도의 횡단면 가구원 가중치에 적용되는데, 사후조정의 경우 2차년도 및 7차년도는 인구주택총조사의 인구수이고 그 외 조사차수는 각 연도의 추계인구를 기준으로 한다.

□ 다음은 09통합표본에서의 가중치 산출 과정을 보여 주는 그림이다.



[그림 II-2] 09통합표본에서의 가중치 산출 과정



□ 이러한 계산과정을 거친 최종 가중치 값은 <표 II-9> 및 <표 II-10>과 같다.

- 최종 가중치 값은 사후조정을 적용하였다.
  - 사후조정은 09통합표본 1차년도부터 9차년도에 모두 실시하였으며, 사후조정 시 고려한 항목은 시도별·성별·연령대별 인구수이다.
  - 인구주택총조사(통계청 자료 사용)에 대한 정보를 사용할 수 있는 조사 차수인 09통합표본 2차년도(2010년) 및 7차년도(2015년)의 경우에는 2010 인구주택총조사 및 2015 인구주택총조사를, 그 외 조사차수에서는 각 연도의 장래추계인구를 기준으로 산출하였다(통계청 장래가구추계 및 장래인구추계 자료 사용).
- <표 II-9>를 통해 가구 가중치의 특성을 살펴보면, 09통합표본 1차년도 당시 가구당 평균적으로 2,559가구 정도를 대표하는 것으로 나타났으나, 09통합표본 11차년도에는 3,002가구를 대표하는 것으로 나타난다.
  - 조사차수가 늘어날수록 표준편차가 커지는 경향이 있다.

<표 II-9> 09통합표본 가구 가중치 특성

조사차수	사례수	합계	평균	표준편차
09통합표본 1차	6,721	17,167,884	2,559.31	1,408.69
09통합표본 2차	6,398	17,339,422	2,715.22	1,629.12
09통합표본 3차	6,388	17,879,317	2,804.16	1,714.27
09통합표본 4차	6,433	18,118,513	2,822.20	1,787.77
09통합표본 5차	6,455	18,388,488	2,854.03	1,869.46
09통합표본 6차	6,490	18,704,975	2,890.14	2,025.11
09통합표본 7차	6,575	19,560,603	2,980.89	2,305.71
09통합표본 8차	6,632	19,281,443	2,914.81	2,093.26
09통합표본 9차	6,683	19,570,670	2,936.33	2,139.87
09통합표본 10차	6,709	19,871,152	2,971.17	2,197.62
09통합표본 11차	6,729	20,115,725	3,002.35	2,218.88

주: 가중치가 0인 경우는 제외한 후, 평균 및 표준편차 산출

- <표 II-10>에서는 개인 가중치의 특성을 확인할 수 있다.
- 차수별로 종단면 가중치와 횡단면 가중치가 일치한다(아래 표에서 합계치). 그러나 09통합표본 2, 3, 7차는 제주도 20~29세 남자 그룹이 1명(비원가구원)으로, 종단면 가중치는 0으로 부여되어 종단면과 횡단면 가중치의 합계가 일치하지 않는다.
- 가구 가중치와 마찬가지로 개인 가중치의 표준편차 역시 조사차수가 늘어날수록 커지는 경향이 있다.

<표 II-10> 09통합표본 개인 가중치 특성

조사차수	라벨	사례수	합계	평균	표준편차
09통합표본 1차	횡단면 <sup>10)</sup>	14,489	41,078,571	2,838.88	1,638.66
09통합표본 2차	종단면	13,641	40,172,409	2,981.92	1,803.61
	횡단면	13,641	40,203,788	2,952.25	1,806.21
09통합표본 3차	종단면	13,409	42,165,178	3,210.63	2,056.64
	횡단면	13,409	42,165,178	3,150.18	1,979.88
09통합표본 4차	종단면	13,424	42,622,622	3,272.12	2,188.18
	횡단면	13,424	42,622,622	3,181.03	2,063.84
09통합표본 5차	종단면	13,300	43,036,656	3,356.73	2,305.78
	횡단면	13,300	43,036,656	3,241.69	2,169.00
09통합표본 6차	종단면	12,591	43,532,966	3,601.04	2,853.69
	횡단면	12,591	43,532,966	3,464.35	2,678.22
09통합표본 7차	종단면	13,369	44,077,062	3,460.55	2,699.30
	횡단면	13,369	44,115,062	3,305.98	2,512.34
09통합표본 8차	종단면	13,517	44,353,240	3,471.88	2,740.34
	횡단면	13,517	44,353,240	3,289.57	2,437.90
09통합표본 9차	종단면	13,768	44,637,628	3,456.26	2,691.63
	횡단면	13,768	44,637,628	3,249.45	2,400.67
09통합표본 10차	종단면	13,738	45,017,245	3,514.23	2,743.86
	횡단면	13,738	45,017,245	3,285.45	2,439.39
09통합표본 11차	종단면	13,704	45,274,471	3,571.11	2,890.68
	횡단면	13,704	45,274,471	3,314.87	2,521.81

주: 가중치가 0인 경우는 제외한 후, 평균 및 표준편차 산출

## 2.3 18통합표본 가중치 산출<sup>11)</sup>

### 2.3.1 18통합표본의 1차년도 가중치 산출 결과

□ 18통합표본 가중치 산출과정은 2.1에서 설명된 98표본 가중치 산출방안과 그 순서와 구조는 기본적으로 유사하다.

- 18통합표본 가중치는 표본을 추가한 시점인 2018년부터 산출하였다.
- 18통합표본은 2018년에 응답한 모든 가구와 가구원, 2018년에 추가된 표본 가구와 가구원으로 구성되며 가중치는 해당 가구와 가구원에게 부여된다.

10) 2009년 시도별·성별·연령별 추계인구로 조정된 통합 1차년도 종단면-횡단면 가중치(종단면과 횡단면 같음으로 가정)

11) 「한국노동패널조사(KLIPS) 추가 표본에 대한 가중치 연구」 보고서(박민규 외, 2020년 9월)를 요약 정리함

□ 18통합표본 구축 시점인 1차년도 가중치는 두 단계를 통해 이루어진다. 먼저 98표본 및 09통합패널 가중치, 그리고 2018년에 추가된 표본 가구와 가구원 가중치가 산출되고, 이 두 가중치의 조정을 통해 최종 가중치가 작성된다. 98표본 및 09통합표본 횡단면 가중치는 2.2절에 설명된 절차에 따라 작성되고, 추가되는 표본 가구와 가구원 가중치는 1) 설계 가중치 산출, 그리고 2) 사후 층화 조정의 순서에 따라 작성된다. 98표본 및 09통합표본 가중치와 2018년 통합표본 가중치 모두 모집단을 대표하도록 가중치가 작성되기 때문에 최종 가중치 작성을 위해서는 전 단계에서 작성된 가중치를 조정하게 되며 조정 계수는 각 개체(가구, 가구원)의 층 내 상대 빈도이다.

□ 18통합표본 구축 시점인 1차년도에 대응하는 개인의 종단가중치와 횡단가중치가 이론적으로 같다. 2018년 추계인구를 기준으로 사후조정된 값을 18통합표본 1차년도 개인 종단가중치로 정의한다. 또한, 통합표본의 가구 가중치도 2018년 추계 가구 정보로 사후 조정한다.

### 2.3.2 로지스틱 회귀분석을 이용한 응답 모형 설정

□ 18통합표본에서의 응답모형도 09통합표본과 같게 가정하고, 로지스틱 회귀모형을 통해 적합하였다.

- 응답모형에 대한 설명변수로 성별, 지역, 학력, 주된 활동, 만나이, 추가패널여부(18통합표본 구축 시 09통합표본 여부)를 분석에 이용한다.

### 2.3.3 18통합표본 2차년도 가중치 산출과정

□ 18통합표본의 가중치 산출은 기본적으로 3단계(1단계 : 종단면 가구원 가중치 산출 → 2단계 : 가구 가중치 산출 → 3단계 : 횡단면 가구원 가중치 산출)의 과정으로 구성된다.

- 1단계 : 종단면 가구원 가중치 산출

- 로지스틱 회귀분석을 통하여 무응답 보정 가중치를 산출하고, 이 과정에서 생성된 가중치를 기초가중치로 정의하며 이를 이용하여 종단면 가구원 가중치를 생성한다. 여기서 산출된 종단면 가중치를 최종적으로 통합연도 기준 추계인구를 이용하여 사후조정하며 이를 최종 개인 종단면 가중값으로 정의한다.

- 2단계 : 가구 가중치 산출

- 가구 가중치는 기본적으로 1단계에서 산출된 기초 가중치들의 각 가구별 평균으로 정의된다. 이 가중치에 인구주택총조사 자료의 가구수 및 추계가구수 정보를 이용하여 사후조정된 가구 가중치를 최종적으로 계산한다.

- 3단계 : 횡단면 가구원 가중치 산출

- 횡단면 가구원 가중치는  $t$ 시점에 원가구원을 한 명이상 포함하고 있는 가구에 속하는 모든 가구원에게 부여된다.
- 먼저 사후조정전의 가구가중치를 각 가구원에게 부여한 다음, 이 값에 사후조정을 적용하여 산출된다. 사후조정은 18통합표본 2차년도의 횡단면 가구원 가중치에 적용되는데, 사후조정의 경우 2차년도 추계인구를 기준으로 한다.

<표 II-11> 18통합표본 가구 가중치 특성

조사차수	사례수	합계	평균	표준편차
18통합표본 1차	12,134	19,871,152	1,637.64	1,021.69
18통합표본 2차	11,584	20,115,725	1,736.51	1,120.92

주: 가중치가 0인 경우는 제외한 후, 평균 및 표준편차 산출

<표 II-12> 18통합표본 개인 가중치 특성

조사차수	라벨	사례수	합계	평균	표준편차
18통합표본 1차	횡단면 <sup>12)</sup>	23,972	45,017,245	1,877.91	1,173.92
18통합표본 2차	종단면	23,013	45,274,471	1,976.02	1,304.86
	횡단면	23,013	45,274,471	1,967.34	1,320.44

주: 가중치가 0인 경우는 제외한 후, 평균 및 표준편차 산출

12) 2018년 시도별·성별·연령별 추계인구로 조정된 통합 1차년도 종단면-횡단면 가중치(종단면과 횡단면 같음으로 가정)

## 2.4 가중치 변수명 및 제공형태

□ 98표본, 09통합표본, 18통합표본 가중치의 변수명은 다음과 같다.

<표 II-13> 가중치 변수명

자료명	가중치		
	98표본	09통합표본	18통합표본
가구용	w_h	sw_h	nw_h
개인용	w_p_l(종단면가중치) w_p_c(횡단면가중치) 단, 1차년도 98표본에는 w01p라는 하나의 가중치만 있음에 주의	sw_p_l(종단면가중치) sw_p_c(횡단면가중치) 단, 12차년도 09통합표본에는 sw12p라는 하나의 가중치만 있음에 주의	nw_p_l(종단면가중치) nw_p_c(횡단면가중치) 단, 21차년도 18통합표본에는 nw21p라는 하나의 가중치만 있음에 주의

## 2.5 가중치 사용 방안

□ 기본적으로 종단면 가구원 가중치는 종단면 가구원 단위분석을 위해서, 횡단면 가구원 가중치는 횡단면 가구원 단위분석을 위해 사용되어야 한다. 그리고 가구 가중치는 가구 단위 분석을 위해 사용한다.

- 종단면 가구원 단위 분석은 크게 두 가지로 구분할 수 있다.

- 첫 번째는 관심의 대상이 되는 모든 연도의 자료를 이용하여 분석되는 경우이다. 혼합선형모형이 해당되며, 이 분석을 위해 일반적으로 각 연도별 자료는 서로 결합을 통해 병합되고 병합된 자료를 통해 고려된 모형이 적합될 수 있다.
- 두 번째는 각 가구원 단위 연도별 변화의 평균 추정과 같은 net change를 예측하기 위한 분석이다. 이 경우 분석 대상이 되는 가구원은 분석 대상 연도에 모두 응답한 가구원이며 이 경우 제공된 종단면 가중치에 추가적인 작업이 이루어져야 한다. 가능한 한 방법은 고려되는 연차들의 가중치 중 최근 생성된 종단면 가중치를 분석 대상이 되는 가구원 수에 맞추어서 스케일 조정을 하는

것이다. 이러한 방안은 매번 분석의 대상이 되는 연차별 자료에 따라 새로운 가중치를 생성해야 한다.

- 횡단면 가구원 가중치 및 가구 가중치의 사용여부는 종단면 가중치의 경우와는 달리 직관적이며 이는 각 연도별 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.

### Ⅲ. 자료의 주요 특성

#### 1. 설문지 구성

- KLIPS의 설문지는 크게 가구용과 개인용으로 나뉜다. 개인용 설문은 취업자와 미취업자용 설문지로 나뉘며, 취업자용 설문은 다시 임금근로자와 비임금근로자용 설문으로 나뉜다.
- 10차년도 공개용 자료부터는 이전차수 변수명을 일치시킨 자료가 제공됨에 따라, 설문지 또한 통합설문지 형태로 제공된다. 통합설문지는 1차년도부터 22차년도까지 설문내용을 총망라하고 있는 하나의 설문지로, 각 설문 문항별로 해당 변수명과 응답차수가 표시되어 있다. 통합설문은 크게 가구용 설문과 개인용 설문, 부가조사 설문으로 구분되며, 개인용 설문은 유형1-유형8, 취업자용 설문과 미취업자용 설문으로 구성된다. 통합설문지는 데이터 사용자들의 편의를 위해 하나의 설문으로 재구성한 것으로, 실사에서 사용하는 설문과는 다소 차이가 있다.
- 실제 조사에서 사용된 설문은 1998년 1차년도에는 가구용, 임금근로자, 비임금근로자, 미취업자 설문이었다. 1999년 2차년도 조사는 가구용 설문과 개인용 설문, 신규용 설문으로 구성되었으며, 개인용 설문은 유형설문과 개인공통 설문으로 구성됐다. 2000년 3차년도에는 가구용, 개인용, 신규용, 부가조사(청년층)가 실시되었으며, 개인용 설문은 유형설문과 임금근로자, 비임금근로자, 미취업자 설문으로 구성됐다. 2001년 4차년도 이후로는 가구용, 개인용(유형설문, 취업자, 미취업자용), 신규용, 부가조사 설문으로 진행되고 있다. 각 차수별 부가조사는 4차년도에는 건강과 은퇴, 6차년도 중고령자, 7차년도 근로시간과 여가, 8차년도 노동조합과 노사관계, 9차년도 청년층, 10차년도 고용형태, 11차년도 교육, 17차년도 시간사용과 삶의 질, 18차년도 장년층 및 삶의 인식, 19차년도 청년층, 20차년도 의사결정, 21차년도 고용형태, 22차년도 비임금근로자 및 무급가족종사자에 관한 내용이었다.



- 1999년 2차년도 조사부터 개인의 노동시장의 진입과 탈퇴 및 이동을 면밀하게 관찰할 수 있는 일자리별 유형 설문을 개발하였다. 유형설문은 지난 조사시의 일자리와 현재 일자리 지속여부를 토대로 여덟 가지 유형으로 구성된다. 가구설문이 가구에 관한 정보를 담고 개인설문이 개인 정보를 담고 있다면 유형설문은 개인이 가졌거나 가지고 있는 일자리를 응답 대상으로 하며, 개인의 일자리에 관한 정보만을 담고 있다.

<표 III-1> 22차년도 KLIPS 설문지의 구성

가구용 설문		
개인용 설문 (기존응답자 및 신규조사자)		유형 ①~⑧ 취업자 개인 미취업자 개인
부가조사 설문	3차년도	청년층
	4차년도	건강과은퇴
	6차년도	중고령자
	7차년도	근로시간과 여가
	8차년도	노동조합과 노사관계
	9차년도	청년층
	10차년도	고용형태
	11차년도	교육
	17차년도	시간사용과 삶의 질
	18차년도	삶의 인식
		장년층
	19차년도	청년층
	20차년도	의사결정
	21차년도	고용형태
	22차년도	비임금근로자

<표 III-2> 1-22차년도 KLIPS 응답자용 설문지의 구성

구 성		
1차년도(1998)	가구용 설문	
	개인용 설문	임금근로자 개인 비임금근로자 개인 미취업자 개인
2차년도(1999)	가구용 설문	
	개인용 설문	유형 ①~⑧ 개인공통
	신규용 설문	
3차년도(2000)	가구용 설문	
	개인용 설문	유형 ①~⑧ 임금근로자 개인 비임금근로자 개인 미취업자 개인
	신규용 설문	
	부가조사 설문(청년층)	
4차년도(2001) ~ 22차년도(2019)	가구용 설문	
	개인용 설문	유형 ①~⑧ 취업자 개인 미취업자 개인
	신규용 설문	
	부가조사 설문	4차년도 건강과은퇴
		6차년도 중고령자
		7차년도 근로시간과 여가
		8차년도 노동조합과 노사관계
		9차년도 청년층
		10차년도 고용형태
		11차년도 교육
		17차년도 시간사용과 삶의 질
		18차년도 장년층 및 삶의 인식
		19차년도 청년층
		20차년도 의사결정
		21차년도 고용형태
		22차년도 비임금근로자

□ <표 III-3>은 일자리 지속을 통한 유형설문의 판별 방법이다.

- 지난 조사시 보유하고 현재에도 지속되는 일자리에 대해서는 임금근로자는 유형 ①, 비임금근로자는 유형③의 설문을 받는다.
- 지난 조사시 보유하고 현재는 지속되지 않는 일자리에 대해서는 임금근로자는 유형②, 비임금근로자는 유형④의 설문을 받는다.
- 지난 조사 이후 시작되었고 현재도 지속되는 일자리에 대해서는 임금근로자는 유형⑤, 비임금근로자는 유형⑦의 설문을 받으며,
- 지난 조사 이후 시작되었고 현재 조사 이전에 끝난 일자리에 대해서는 임금근로자는 유형⑥, 비임금근로자는 유형⑧의 설문을 받는다.

<표 III-3> 일자리 지속을 통한 유형설문의 판별

지난 조사 당시 일자리 여부	종사상 지위	현재 계속 여부	유형
있었다	임금	현재도 하고 있다.	유형 ①
		지금은 그만두었다.	유형 ②
	비임금	현재도 하고 있다.	유형 ③
		지금은 그만두었다.	유형 ④
없었다	임금	현재도 하고 있다.	유형 ⑤
		지금은 그만두었다.	유형 ⑥
	비임금	현재도 하고 있다.	유형 ⑦
		지금은 그만두었다.	유형 ⑧

◎User notes : 유형 설문지는 개인단위가 아닌 일자리 단위로 성립되는 설문이다. 따라서 개인의 일자리 경력에 따라 복수의 유형 설문이 가능하다. 유형설문의 내용 중 현재 지속되는 주된 일자리의 정보는 개인 데이터의 “주된 일자리”에 수록되며, 유형설문의 모든 내용은 직업력 자료에 수록된다. 이때 “주된 일자리”는 가장 많은 시간을 일한 일자리를 의미한다.

## 2. 자료 구성

□ KLIPS 조사는 설문내용이 워낙 방대하고 그 구성 또한 복잡하기 때문에 사용자들이 자료를 보다 편리하게 이용할 수 있도록 자료를 재구성(가공)하였다. 이중 가구용 설문과 부가조사 설문은 각각 가구용 자료, 부가조사 자료와 그 구조가 거의 일치한다. 신규용설문과 개인용 설문-유형설문과 개인공통(취업자용, 미취업자용) 설문-의 내용은 개인용 자료와 직업력 자료로 구성된다.

□ KLIPS의 각 연도별 자료의 구성은 다음과 같다.

- 가구자료 : 모든 조사년도별로 만들어지며 가구조사용 설문지를 토대로 구성된다.
- 개인자료 : 모든 조사년도별로 만들어지며, 유형설문지, 취업자와 미취업자 개인 설문지, 신규조사자용 설문지 등을 토대로 구성된다.
- 부가조사자료 : 3차~4차년도 자료와 6차~11차년도 자료, 17차~22차년도 자료에 포함되며, 부가조사 설문지를 토대로 구성된다.
- 직업력자료 : 직업력 자료는 유형 설문지를 통해 조사된 개인의 모든 직업력 (work history)를 누적한 형태의 자료이므로 다른 자료들과는 달리 연도별 자료가 별도로 존재하는 것이 아니라 1개의 자료만이 존재한다.

※ 1~22차년도 공개용 자료는 개인자료에 신규설문의 응답내용이 모두 포함되어 있다. 개인응답자의 회고적 자료, 즉 부모의 학력, 14세 당시의 성장지 등과 같은 변수들은 1차년도에는 모든 응답자의 값이 존재하지만, 2차년도 이후로는 신규응답자들에 한해서 응답내용이 존재한다.

<표 III-4> 1-22차년도 KLIPS 자료의 구성

		1차 (1998)	2차 (1999)	3차 (2000)	4차 (2001)	5차 (2002)	6차~11차 (2003~2008)	12차~16차 (2009~2013)	17차~22차 (2014~2019)
가구용 자료		KLIPS01H	KLIPS02H	KLIPS03H	KLIPS04H	KLIPS05H	KLIPS06H~ KLIPS11H	KLIPS12H~ KLIPS16H	KLIPS17H~ KLIPS22H
<div>개인 공통</div>	개인용 자료	KLIPS01P	KLIPS02P	KLIPS03P	KLIPS04P	KLIPS05P	KLIPS06P~ KLIPS11P	KLIPS12P~ KLIPS16P	KLIPS17P~ KLIPS22P
	유형 ①~⑧	×	○	○	○	○	○	○	○
	<div>취 업 자</div>	임금 근로	○	○	○	○	○	○	○
		비임금 근로		○					
		미취업자		○	○	○	○	○	○
	신규조사자		×	○	○	○	○	○	○
부가조사 자료		×	×	KLIPS03A	KLIPS04A	×	KLIPS06A~ KLIPS11A	×	KLIPS17A~ KLIPS22A
직업력 자료		KLIPS22W							

### 3. 조사내용

#### 3.1 조사내용

□ KLIPS는 가구와 개인의 경제활동과 노동시장의 동태적 변화를 관찰하기 위해 설계된 조사이므로 기본적으로 매년 동일한 내용의 설문 질문이 반복되는 형식을 갖추고 있다. 다만 부분적으로 조사년도별로 조사 내용에 약간의 차이를 보인다.

□ <표 III-5>는 1차년도부터 22차년도까지 지속적으로 조사된 설문 내용이다.

<표 III-5> 1-22차년도 조사 내용 (공통) - 응답자용 설문 기준

설문지 구성	내 용	
가구용	<ul style="list-style-type: none"> <li>▪ 가구원의 인적 사항</li> <li>▪ 변동 가구원(추가, 분가, 사망) 관련 사항</li> <li>▪ 가족관계와 부모와의 경제적 교류 관련 사항</li> <li>▪ 주거 관련 사항</li> <li>▪ 자녀교육 및 자녀 보육 관련 사항</li> <li>▪ 가구의 소득, 소비, 자산과 부채</li> <li>▪ 가구의 경제 상태 및 가계에 부담을 느끼는 소비 항목</li> </ul>	
예비조사	<ul style="list-style-type: none"> <li>▪ 지난 조사시 일자리 응답내용 확인과 현재 지속여부</li> <li>▪ 지난 조사 이후 새로운 일자리의 종사상 지위와 현재 지속여부</li> </ul>	
유형①, 유형② 유형⑤, 유형⑥	<ul style="list-style-type: none"> <li>▪ 지난조사 이후 일자리관련 변경사항</li> <li>▪ 일자리 형태 및 규모</li> <li>▪ 근로계약관련 사항</li> <li>▪ 근로시간관련 사항</li> <li>▪ 임금관련 사항</li> </ul>	<ul style="list-style-type: none"> <li>▪ 일자리를 통한 사회보험 가입여부</li> <li>▪ 일자리의 노동조합</li> <li>▪ 일자리 그만둘 당시의 상황</li> <li>▪ 일자리의 직업 및 산업</li> <li>▪ 구직활동 관련 사항</li> </ul>
유형③, 유형④ 유형⑦, 유형⑧	<ul style="list-style-type: none"> <li>▪ 지난 조사 이후 일자리 관련 변경사항</li> <li>▪ 일자리의 종업원 수와 가족종사자의 수</li> <li>▪ 일자리의 근로시간</li> <li>▪ 일자리의 매출액과 소득</li> <li>▪ 일자리를 그만둘 당시 상황</li> <li>▪ 구직활동관련 사항</li> <li>▪ 일자리의 직업 및 산업</li> <li>▪ 창업과정의 어려움</li> </ul>	
취업자 개인	<ul style="list-style-type: none"> <li>▪ 경제활동상태</li> <li>▪ 부가급여 및 복리후생</li> <li>▪ 직무만족도</li> <li>▪ 구직활동 및 희망하는 일자리의 특성, 구직시 어려움</li> <li>▪ 교육 및 직업훈련</li> <li>▪ 정규교육 관련 사항</li> <li>▪ 생활만족도 및 경제적 여건</li> <li>▪ 종교 및 종교활동(12차년도 이후)</li> </ul>	<ul style="list-style-type: none"> <li>▪ 현재 하는 일의 고용상의 특성</li> <li>▪ 근로시간관련 사항</li> <li>▪ 현재 하는 일의 교육, 기술 수준 적합도</li> <li>▪ 사회보험 수급관련 사항</li> <li>▪ 혼인상태</li> <li>▪ 고용형태</li> </ul>

<표 III-5>의 계속

설문지 구성	내 용
미취업자 개인	<ul style="list-style-type: none"> <li>▪ 경제활동상태</li> <li>▪ 구직활동, 구직방법, 희망하는 일자리의 특성, 구직시 어려움</li> <li>▪ 교육 및 직업훈련</li> <li>▪ 사회보험 수급관련 사항</li> <li>▪ 정규교육 관련 문항</li> <li>▪ 혼인상태</li> <li>▪ 생활만족도 및 경제적 여건</li> <li>▪ 종교 및 종교활동(12차년도 이후)</li> </ul>
신규	▪ 개인 설문지의 모든 내용 + 회고적 일자리

□ <표 III-6>은 모든 응답차수에 걸쳐 각 개인별로 한번만 응답하도록 한 설문을 정리한 것이다. 즉 특정년도에 모든 개인응답자에게 질문한 후, 이후 신규응답자만 조사한 문항들이다.<sup>13)</sup>

<표 III-6> 개인별 한번만 조사하는 내용 - 응답자용 설문 기준

설문 내용	응답 차수 및 응답대상
▪ 교육력	1차년도, 4차년도는 모든 개인응답자 다른 차수는 신규 응답자
▪ 혼인력	1차년도, 2차년도는 모든 응답자의 생애 전체 혼인력 3차 이후 모든 응답자 변화를 중심으로 조사 3차 이후 신규 응답자 최초결혼이후 혼인력을 모두 조사
▪ 출산력	1차년도 모든 응답자 2차 이후 신규 응답자
▪ 출생지      ▪ 성장지 ▪ 성장기 아버지의 교육수준 및 경제활동 ▪ 군복무	1차년도 모든 응답자 2차 이후 신규 응답자
▪ 종교	1~11차년도 신규응답자 12차 이후는 모든 응답자
▪ 형제자매	6차년도 모든 응답자 7차 이후 신규 응답자
▪ 성장기 어머니의 교육수준 및 경제활동	4차년도 모든 응답자 5차 이후 신규 응답자
▪ 성장기 경제적형편	8차년도 모든 응답자 9차 이후 신규 응답자
▪ 만 15세 이후 일자리 경력	1차년도 모든 응답자 2차 이후 신규 응답자

13) 다만 예외적으로 상당한 시간이 흐른 후 변화를 살펴보기 위한 목적으로 모든 응답자에게 재 질문을 실시한 경우가 있을 수 있음.

□ <표 III-7>은 각 차수별로 부가조사와는 별도로 가구용 혹은 개인용 설문에서 한 시적으로 질문한 문항들이다.

<표 III-7> 각 차수별 차별된 조사 내용 - 변수명 일치 자료 기준

조사차수	해당자료	조사내용
1차년도	가구용	▪ 최근의 경제위기(IMF)에 관한 견해    ▪ (가구원) 전공분야 ▪ 탁아모관련문항    ▪ 사회보험 수급관련문항
	개인용	▪ 부업관련 문항    ▪ 취업 중 구직활동 ▪ 경제위기 후 변화사항    ▪ 취업 중 구직어려움
2차년도	가구용	▪ (가구원) 전공분야    ▪ (가구원) 건강상태
	개인용	▪ 취업 중 구직활동    ▪ 취업 중 구직어려움
3차년도	가구용	▪ 탁아모관련문항
	개인용	▪ 고용계약 관련문항    ▪ 이직 관련문항 ▪ 사회적 지위 및 인맥    ▪ 취업중 구직활동 ▪ 취업 중 구직어려움
4차년도	가구용	▪ 탁아모관련문항
	개인용	▪ 취업 중 구직어려움    ▪ 창업훈련 관련 문항 ▪ 비임금근로자의 사업체 운영관련 문항
5차년도	개인용	▪ 취업 중 구직어려움    ▪ 비임금근로자의 사업장 가치 관련 문항 ▪ 입시성적 관련 문항
6차년도	개인용	▪ 학업성취 관련 문항    ▪ 질병관련 문항
7차년도	개인용	▪ 차별관련 문항    ▪ 리스크 관련문항    ▪ 질병관련 문항
8차년도	개인용	▪ 근로시간 단축 (8차이후 지속)    ▪ 키, 몸무게, 혈액형
9차년도	개인용	▪ 장애관련문항
10차년도	가구용	▪ 자녀의 영어교육관련문항    ▪ 유동성제약관련문항
	개인용	▪ 어학연수 및 영어능력관련문항    ▪ 리스크 관련문항    ▪ 키, 몸무게
11차년도	개인용	▪ 키, 몸무게



## IV. 변수 가이드

- 2010년 10차년도 공개용 자료부터 변수명이 차수에 걸쳐 일치된 형태로 제공된다.
- 예를들어, 2009년까지 기존 자료의 경우에는 ‘개인응답자의 종사상지위’ 변수가 다 음과 같이 차수별로 조금씩 달랐다.

1차	2차	3차	4차	5차	6차	7차	8차	9차
p01026	p02351	p03363	p04746	p05748	p06741	p07736	p08736	p09736

- ☐ 그러나 10차년도 공개용 자료 이후로는 이들 변수명이 각 차수별로 모두 p\_\_0314로 일치되어 제공된다. 이때 중간의 \_\_에는 해당 차수가 들어가므로, 22차년도의 경우 개인응답자의 종사상 지위는 p220314이며, 1차년도의 경우에는 p010314이다.
- ☐ 변수명이 부여되는 규칙은 다음과 같다.

예:                    h                    22                    01                    41  
                         설문종류   조사차수   SET번호   변수번호

- 설문종류 : 가구용의 경우 h, 개인용의 경우 p, 부가조사의 경우 a이며, 직업력 데이터는 j로 시작한다. 단, 10차년도 공개용자료 이후로는 별도의 신규용 데이터가 제공되지 않고, 개인용에 포함되어 제공된다.
- 조사차수 : 설문종류 다음 두 자리는 해당하는 조사 차수를 의미한다.
- set 번호 : 여러 개의 관련 문항을 하나의 set으로 구성하여, 관련 문항들은 동일한 set 번호를 갖는다. 예를 들어 가구소득과 관련된 문항은 set 번호가 21~22번이므로, 해당 set을 찾으신 후 해당변수를 찾으시면 된다. (set번호 리스트는 노동패널 홈페이지를 참조.)

- 변수번호 : 해당하는 변수의 일련번호를 의미한다.

□ KLIPS 자료의 변수 종류는 다음과 같다.

- 키(key)변수 : KLIPS는 매년 동일 가구와 동일 개인을 반복 추적조사하기 때문에 패널구성원의 이름이라고 할 수 있는 key변수가 부여된다. KLIPS 자료는 여러 가지 자료의 묶음으로 구성되어 있기 때문에 자료간 연결을 할 때 key변수는 유용하게 사용된다. key변수의 예로 hhid, pid 등을 들 수 있다.
- 그 외의 주제별 변수 : 키변수와 구성변수를 제외한 모든 변수들로 설문지의 내용을 반영한 변수들이다.

◎User notes : KLIPS 자료의 변수명은 자료별로 다음과 같은 고유의 영문 이니셜을 갖고 있으며, 영문 이니셜 뒤에는 일련의 숫자로 구성된다.

h_____ : 가구자료	p_____ : 개인자료
j_____ : 직업력 자료	a_____ : 부가조사자료

## 1. key변수의 이해와 이용

- KLIPS자료를 사용하기 위해서는 key변수를 포함한 주요 변수들에 대한 이해가 필요하다. 여기서는 각 자료별 key변수를 고유번호와 그 외의 key변수들로 나누어 설명하고 있으며, 그 외의 key변수들로는 자료를 사용할 때 가장 빈번히 사용되는 구성변수와 인구학적인 변수를 중심으로 구성하였다.
- KLIPS의 key변수들은 해당 년도마다 동일한 형태를 띠는 것도 있지만, 조사 년도에 따라 약간씩 다른 형태를 띠는 변수들도 있다. 따라서 주요한 인구학적 key변수들의 1-22차년도 변수명을 정리하고 이 변수들의 이용방법을 설명하고자 한다.

## 1.1 가구용 key변수

### 1.1.1 가구의 고유번호

□ KLIPS 가구용 자료에 담겨 있는 가구의 고유번호는 22차년도를 기준으로 했을 때 각 차수별로 hhid01-hhid22이며, orghid98, orghid09, orghid18는 패널고유가구번호이다.

- orghid98(98표본)는 1998년 처음 조사가 시작된 연도의 가구번호를 의미하며, orghid09(09통합표본)은 2009년에 표본추가로 구축된 1,415가구를 포함하여 2009년 당시 응답 가구 전체의 원가구번호를 의미하며, orghid18(18통합표본)은 2018년 표본추가로 구축된 5,044가구를 포함하여 2018년 당시 응답가구 전체의 원가구번호를 의미한다. 나머지 가구의 고유번호인 hhid01-hhid22는 해당연도의 가구번호를 의미한다.

- KLIPS 조사의 응답 가구는 매년 고정적이지 않으며, 분가 혹은 합가 등의 이유로 새로운 가구가 생성되거나 혹은 기존의 가구가 소멸되는 등의 변화를 보인다.

- 따라서, 당해연도 가구번호가 missing인 경우는 당해연도에 조사되지 않은 가구를 의미한다. 단, orghid98, orghid09, orghid18의 경우에는 분가가구가 분가이전에 어떤 원가구에 소속되었는지를 알 수 있도록 응답여부에 관계없이 값이 입력되어 있다.

[illegible]

### 1.1.2. 그 외의 key 변수

□ KLIPS의 가구용 자료에는 가구의 고유번호 외에 key가 되는 구성변수들이 포함되어 있다. 고유번호 외에 key가 되는 변수는 hwave01-hwave22, hwaveent, sample98, sample09, sample18 등이다.

- 여기서 hwave는 각 해당연도의 응답여부를 의미한다. 즉 hwave01은 KLIPS 1차년도(1998년) 가구의 응답여부이며 hwave02는 KLIPS 2차년도(1999년) 가구의 응답 여부이다. sample98(98표본) 및 sample09(09통합표본), sample18(18통합표본)는 원가구 여부를 의미한다.

□ KLIPS 가구용 자료는 당해연도 응답여부와 관계없이 한번이라도 응답한 적이 있는 가구라면 가구용 자료상에 하나의 case로 입력된다.

- 예컨대, 22차년도 가구용 자료는 총 12,687가구로 구성되어 있으며, 이 중에서 22차년도에 응답한 가구는 11,720가구(hwave22=1)이다.

## 1.2 개인용 key변수

### 1.2.1 개인의 고유번호

□ KLIPS 개인용 자료에 담겨 있는 개인의 고유번호관련 변수는 다음의 <표IV-1>과 같다. 개인용 자료에서는 key변수로 개인의 고유번호인 pid와 각 해당연도의 가구번호와 가구원 번호가 있다. pid는 KLIPS 응답 대상인 개인의 이름이라고 할 수 있는 것으로 조사년도와 상관없이 불변하는 번호이다.

- pid는 개인이 처음으로 응답한 해의 가구번호와 그 가구에서의 일련번호로 구성된다. 즉,

pid = 개인이 처음으로 응답한 해의 가구번호\*100 + 몇 번째 가구원  
이다.

◎User notes : 가구 및 개인의 고유번호는 자료간 연결을 할 때 유용하게 사용된다. KLIPS자료의 자료 간 연결은 다음의 네 가지 형태가 있을 수 있으며, 각각의 자료간의 연결 방법은 다음과 같다.

가. 가구자료와 개인자료의 연결 : 해당연도의 가구번호를 통해서 두 자료를 연결한다.  
     ex) 22차년도 가구와 개인 자료의 연결시 hhid22을 이용

나. 가구자료의 시계열 연결 : hhid\*\*를 이용하여 두 자료를 연결한다.  
     ex) 22차년도 가구와 21차년도 가구 자료의 연결시 hhid22와 hhid21를 이용.

다. 개인자료의 시계열 연결 : pid를 이용하여 두 자료를 연결한다.

라. 개인자료와 개인자료(부가자료)의 연결 : pid를 이용하여 두 자료를 연결한다.

자세한 사항은 FAQ Q6, Q7, Q8참조.

### 1.2.2 그 외의 key 변수

□ 개인용 자료의 고유번호 외의 key가 되는 변수로는 성별, 가구주와의 관계, 연령, 학력, 1차 원가구원 여부, 취업상태, 고용형태, 근로시간형태, 종사상 지위 등을 들 수 있다. 이러한 변수들은 자료의 사용 시 가장 기본이 되는 사항으로 가장 빈번히 사용되는 변수들이다. 이 변수들을 중심으로 각각의 변수에 관한 내용을 살펴 보겠다.

#### ● 가구주와의 관계

- 가구주와의 관계는 기본적으로 가구자료에서 파악되지만 자료상으로는 개인자료에도 key변수로 들어가 있다.
- 이때 가구주란 ‘가구를 실질적으로 대표하는 사람’을 의미하는데 이는 응답자의 주관적 판단에 의한 것이며 가구주가 사망하였을 경우 다른 가구원이 가구주로 대체된다. 따라서 가구주와의 관계의 변수 값은 조사 연도에 따라 동일개인에 대하여 고정적이지 않다.
- KLIPS자료에서는 가구주와의 관계변수를 통하여 한 가구의 가족 구성형태 및 이에 관한 정보를 얻을 수 있다. 가구주와의 관계 변수는 p\*\*0102이며, 응답항목 값은 코드북의 open code를 참조하면 된다.

● 연령

- KLIPS 자료에서 연령을 구성할 수 있는 변수로는 생년월일을 이용하는 방법과 만나이를 이용하는 두 가지의 방법이 있다. 생년월일은 지난 조사 시 응답 오류가 발생할 수 있으므로 매년 확인을 거쳐 자료가 구성된다. 연령을 구성하는데 있어서 생년월일과 만나이 어떤 것을 사용해도 무방하다. 각 차수 변수명은 <표 IV-2>와 같다.

<표 IV-2> 1-22차년도 연령 변수

		변수명
생 년 월 일	양(음)력여부	p**0103
	생년	p**0104
	생월	p**0105
	생일	p**0106
만나이		p**0107

● 혼인상태

- 혼인상태는 개인 및 신규설문 모두에서 조사되었다. 다만, 조사설계시 1차년도부터 4차년도까지는 혼인과 관련한 문항에 변동이 많았다. 1차년도와 2차년도에는 혼인력 형태의 설문이 진행되었으며, 3차년도의 경우에는 신규응답자에게만 물어보았고, 개인용 설문의 경우에는 변경여부만을 확인하였다. 또한 4차년도에는 혼인상태를 가구용 설문에 질문하였다. 그러나, 데이터 상에는 이용자들의 편의를 돕기 위해 차수별로 동일한 항목으로 구성된 혼인상태 변수가 들어있다. 또한 혼인상태의 변화내용과 변화시기가 보정된 변수가 제공된다.
- 혼인상태 변수는 p\*\*5501이며, 응답항목은 (1) 미혼, (2) 기혼유배우, (3) 별거, (4)이혼, (5) 사별을 의미한다.
- 3차년도와 4차년도의 혼인상태는 이전차수 응답내용이나 다음차수 응답내용을 끌어와 붙여서 생성시킨 값으로, missing 값이 다소 존재한다.

● 학력

- KLIPS의 학력 변수는 설문지상 가구용 조사에서 질문되고 있으며 자료상으로는 가구자료와 개인자료에서 모두 제공된다. 학력변수는 학교, 이수여부, 학년 변수로 구성되므로 범주형과 연속형 변수로의 구성이 모두 가능하다.
- 학력 관련 변수는 연도별로 약간씩의 차이를 보인다. 1, 2차년도에는 학교, 이수여부, 학년, 전공계열의 네 가지 변수로 구성되었으나 3차년도 이후로는, 전공계열 변수는 삭제되었다. 대신 4차년도 개인용 자료부터는 개인의 정규교육과 관련된 자세한 정보를 담고 있다.
- 1, 2차의 전공 계열의 자료 구성은 약간 다르다. 1차년도에는 고등학교에 재학중인 개인에 한하여 고등학교 전공계열을 조사했으며, 2차년도에는 고등학교와 전문대 이상에 재학중인 개인을 대상으로 전공계열을 조사하였다. 따라서 1, 2차년도의 전공계열의 변수값은 다르다. 또한 1차년도 자료의 학력은 최고 대학원 석사까지 조사하였으나 2차년도부터는 대학원 박사항목을 추가하여 조사하고 있다.
- ‘학년’은 졸업을 제외한 수료, 중퇴, 재학중, 휴학중인 개인을 대상으로만 조사하였다.
- 한편 졸업·수료·중퇴 시기를 묻는 질문이 5차년도부터 가구용 설문에 추가되었다.

<표 IV-3> 각 연도별 학력 변수

	변수명	응답차수				
		1차	2차	3차	4차	5차-22차
학교	p**0110	○				
이수 여부	p**0111	○				
학년	p**0112	○				
전공계열	p**0113	○	○	-	-	-
졸업·수료·중퇴 시기(년)	p**0114	-	-	-	-	○
졸업·수료·중퇴 시기(월)	p**0115	-	-	-		○



◎User notes : 학력변수의 구성 예는 다음과 같다.(11차년도 기준)

예1) 연속형(교육년수)

p\*\*0110(학력)=5, p\*\*0111(이수여부)=4, p\*\*0112(학년)=2의 값을 갖는 개인의 학력은 고등학교 2학년에 재학중이다. 따라서 이 개인의 교육년수는 9+2=11이 된다.

예2) 범주형

p\*\*0110(학력)=4, p\*\*0111(이수여부)=1의 값을 갖는 개인의 학력은 중졸로 연구자가 임의의 값을 부여하면 된다.

◎ 1차 원가구원 여부(98표본)

- 98표본 1차 원가구원이라 함은 1차년도(1998년) 표본으로 지정된 가구에 속한 개인을 뜻한다. KLIPS는 원가구로 선정된 5,000가구와 그 가구에 속한 개인을 기본적인 조사대상으로 한다. 그러나 조사가 진행되면서 분가 혹은 혼인 등의 이유로 개인이 탈락 혹은 진입하기 때문에 ‘비원가구원’이 개인 조사 대상으로 편입된다.
- ‘원가구원’은 신규용 자료에 응답한 개인 중에도 존재한다. 즉, 1차 원가구원이면서 1차 조사 당시 개인용 조사가 누락된 개인 혹은 만15세가 되어 새로 조사된 개인은 원가구원에 속한다. 따라서 원가구원을 대상으로 분석하고자 할 때는 신규조사자 중 원가구원도 고려해야 한다.
- 개인(신규) 자료에서 원가구원만을 분석에 사용하고자 한다면, ‘각 연도별 종단면 가중치 값이 0인 개인’을 제외하면 된다. 종단면 가중치를 산출할 때 비원가구원은 0의 값을 갖도록 설계되었기 때문이다.
- 변수명은 p\*\*orig98이다.

◎ 1차 원가구원 여부(09통합표본)

- 변수명은 p\*\*orig09이다.
- 09통합표본의 1차 원가구원은 09통합표본 1차년도(2009년)에 조사된 6,451가구에 속한 개인을 뜻한다.

◎ 1차 원가구원 여부(18통합표본)

- 변수명은 p\*\*orig18이다.
- 18통합표본의 1차 원가구원은 18통합표본 1차년도(2018년)에 조사된 12,134가구에 속한 개인을 뜻한다.

◎User notes : KLIPS는 표본의 대표성을 갖는 동일 가구, 동일 개인을 추적 조사하는 패널조사이기 때문에 이런 원칙상의 이유로 원가구원 여부가 갖는 의미는 매우 크다고 할 수 있다. 그러나 개인 분석대상을 원가구원만으로 한정할 것인지 아니면 모든 개인을 분석대상으로 할 것인지는 연구자의 자의적인 판단에 따른다. 다만 가중치를 사용할 경우 이런 문제는 어느정도 해결된다. 종단면 가중치의 경우 비원가구원에 대해 '0'의 가중치가 부여되는 한편, 횡단면 가중치의 경우에도 비원가구원은 동일한 가구내의 원가구원에 비해 상대적으로 작은 가중치를 부여받기 때문이다.

#### ◎ 취업/미취업 상태

- KLIPS 자료에서는 다음의 정의에 따라 개인의 취업상태를 파악하고 있다.
  - 취업자: (1) 조사주간 중 수입을 목적으로 1시간 이상 일한 자
  - (2) 자기에게 직접적으로는 이득이나 수입이 오지 않더라도 가구단위에서 경영하는 사업체의 수입을 높이는데 도움을 준 가족종사자로 주당 18시간 이상 일한 취업자
  - (3) 직업을 가지고 있으나 일시적인 병·사고, 휴가 또는 연가, 노사분규, 조업 중단 등으로 일시적으로 쉬고 있는 휴직자
  - 미취업자 : 지난 1주간 수입을 목적으로 일을 한 적이 없으며, 일거리나 직장을 가지고 있지도 않은 경우, 또는 가족이나 친지가 사업주인 사업체에서 평소 일주일에 18시간 미만으로 일을 도운 경우를 말한다.
- 각 차수별로 설문문의 형태는 조금씩 달랐지만, 변수통합 작업을 통하여 모두 동일한 형태를 갖도록 변수명이 조정되었다. 해당 변수는 p\*\*0201이다.

#### ◎ 경제활동상태

- ☐ KLIPS의 개인용 설문문항 중 개인의 경제활동상태와 관련된 문항들은 기본적인

로 통계청의 「경제활동인구조사」와 비교가 가능하도록 구성되었다. 먼저 경제활동인구분류와 경제활동상태를 규정하는 기본적인 개념들을 살펴보면 다음과 같다.

<표 IV-4> 경제활동상태의 구분

15세 이상 인구		
경제활동인구		비경제활동인구
취업자	실업자	-가사 혹은 육아 -학업 -연로 -심신장애 -기타
- 수입을 목적으로 1시간 이상 일한 자 - 주당 18시간 이상 일한 무급가족종사자 - 일시휴직자		

- 경제활동인구 : 만15세 이상 인구 중 조사대상 주간동안 재화나 용역을 생산하기 위하여 노동을 제공한 사람과 제공할 의사와 능력이 있는 사람을 말하며 취업자와 실업자를 합하여 경제활동인구라고 한다.
- 비경제활동인구 : 만15세 이상 인구 중 취업자도 실업자도 아닌 사람 즉 일할 능력은 있으나 일할 의지가 없거나(적극적인 구직활동을 하지 않은 경우) 혹은 일할 능력이 없는 경우이다.
- 실업자 : 현재 미취업자이면서, 지난 1주간 구직활동을 하였고 지난주에 일이 있었다면 일을 할 수 있었던 경우, 또는 지난 1주 동안은 구직활동을 하지 않았지만 지난 1개월 동안 구직활동을 한 적이 있고, 지난 1주 동안 적당한 일자리가 있었다면 일을 할 의사가 있는 경우를 말한다.

□ KLIPS 자료를 통하여 개인의 경제활동상태를 구분할 수 있다. 개인의 경제활동상태는 취업자와 실업자를 포함한 경제활동인구와 비경제활동인구로 구분될 수 있다.

<표 IV-4>는 각 조사년도별 경제활동상태 구분과 관련된 변수를 정리한 것이다.

□ 개인의 경제활동상태를 구분하기 위해서는 각 연도별 취업상태 변수를 통해 취업자와 미취업자를 구분한 다음, 미취업자의 해당 변수를 통해 실업자를 구분하면 된다.

<표 IV-5> 각 연도별 실업자 및 비경제활동인구 구분변수

미취업자		p**0201
실업자	지난 1주간 구직여부	p**2801
	지난 1개월간 구직여부	p**2802
	취업 가능성 여부	p**2806
비경제활동인구		미취업자에서 실업자를 제외한 나머지 경우.

● 종사상 지위

- 임금근로자의 종사상의 지위는 상용직, 임시직, 일용직으로 구분된다. 1차년도 자료에는 종사상 지위변수가 임금근로자만을 대상으로 하고 있으나, 2차년도 이후 자료부터는 모든 취업자를 대상으로 주된 일자리의 종사상 지위변수가 구성되었기 때문에 변수 값이 5가지로 구성된다.
- 1차 자료를 이용하여 굳이 비임금근로자의 종사상 지위를 구성하려면 p010402(종업원 수)와 p013121(사업체 대표) 변수를 이용하면 된다.

<표 IV-6> 1-22차년도 종사상 지위 변수

	1차년도	2차년도 이후
종사상 지위	p**0314	
응답항목	1=상용직 2=임시직 3=일용직	1=상용직 2=임시직 3=일용직 4=고용주/자영업자 5=무급가족종사자

● 주관적 판단에 따른 고용형태(정규/비정규 여부)

- 고용형태(정규/비정규 여부) 변수는 응답자들의 주관적 인식에 따라 정규직·비정규직 여부를 구분하는 변수이다. 이 항목은 설문지상에는 1차년도에는 존재했고, 2차년도에는 전년도와 비교한 변경여부만 물었으며, 3차년도에는 삭제되었다가 4차년도에 다시 조사되었다.

- 자료상에는 3차년도를 제외한 모든 차수에 존재한다. 따라서 3차년도 임금근로자의 고용형태를 구성할 때는 전년도 자료에서 해당 변수를 연결(merge)하여 사용하여야 한다.

※ 단, 6차년도 정규/비정규 여부 변수는 missing 값이 다른 연도보다 높게 나타나고 있다는 점에 유의하기 바란다. 6차년도 조사 당시 새로 시작한 임금일자리, 즉 '유형5'에 해당하는 응답 대상자들에게 정규/비정규의 변화여부만 질문한 후 고용형태가 변하지 않은 응답자들에 대해서는 정규직인지, 비정규직인지 여부를 확인하지 않았다. 이들의 고용형태 7차년도 조사를 통해 다시 확인하여 사후적으로 코딩한 것이다. 따라서, 7차년도 조사에 응하지 않은 응답자의 고용형태는 복원하지 못했으며, 이들의 정규/비정규 여부는 missing으로 처리되었다.

- 주관적 판단에 따른 고용형태(정규/비정규) 여부 변수는 p\*\*0317이다.

#### ● 주관적 판단에 따른 근로시간 형태

- KLIPS에서는 임금근로자를 대상으로 근로시간 형태를 묻고 있다. 근로시간 형태는 시간제, 전일제로 구분하였다.
- 근로시간 형태 변수는 p\*\*0315이며, (1)시간제 (2) 전일제를 의미한다.

## 2. 오픈 코드(open code)

- 각 변수의 변수값이 너무 많은 경우 모두 코드화할 수 없기 때문에 오픈코드로 처리하였다. 자료별 오픈코드의 종류와 내용을 살펴보면 다음과 같다.

### 2.1 가구용 자료의 오픈 코드

- 가구용 자료에서는 가구주와의 관계를 오픈 코드로 처리했다. 가구주와의 관계는 기본적으로 가구조사에서 파악되며 개인용 자료에도 포함된다. 가구주를 기준으로 (코드값 '10') 다른 가구원과의 관계를 통한 고유 코드값이 지정되고 이 값들을 각 가구원에게 부여된다.

## 2.2 개인용 자료의 오픈 코드

□ 각 연도별 개인용 자료의 오픈 코드는 다음과 같다.

- 1차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명, 직업훈련 종류, 직업훈련기관, 설문지 거절이유, 학과, 직위
- 2차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명, 자격증 종류, 자격증 급수
- 3차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명
- 4차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명, 직업훈련 분야, 1년이내받고싶은 직업훈련내용, 자격증 종류, 자격증 급수, 지병종류(병명), 고등학교명
- 5차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명, 직업훈련 분야, 1년이내받고싶은 직업훈련내용, 자격증 종류, 자격증 급수, 고등학교명
- 6차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명, 직업훈련 분야, 직업훈련 기관, 직업훈련 외 교육, 1년이내받고싶은 직업훈련내용, 자격증 종류, 자격증 급수, 기타 질병, 고등학교명
- 7~22차년도 : 가구주와의 관계, 산업, 직업, 대리이유, 대학교명, 시군구명, 직업 훈련 분야(7~9차), 직업훈련 외 교육(7차), 기타질병(7차)

□ 22차년도 자료를 기준으로 문자변수로 제공되는 변수는 다음과 같다.

- 직위, 직업훈련이름, 자격증종류, 고등학교명, 고등학교 및 대학교 재학중 취업내용, (신규조사자)부모-14세 무렵 경찰상태(직위)
- 직위는 1차년도는 코드변수, 2차년도부터 21차년도까지 문자변수로 제공하고 있다.
- 자격증 종류는 1차, 3차년도에는 조사되지 않았으며, 2차, 4차~6차년도에는 코드변수로, 7차년도 이후에는 문자변수로 제공하고 있다.
- 고등학교명은 4차~6차년도에는 코드변수이며, 7차년도 이후에는 문자변수로 제공하고 있다.
- 직업훈련이름은 5차년도부터 문자변수로 제공하고 있다.

- ☐ 각 연도별 개인용 자료에 공통으로 담긴 오픈 코드는 가구주와의 관계, 산업, 직업, 직위, 대학교명, 시군구코드이다.

### 2.2.1 직업 훈련 및 자격증 코드

#### ● 직업훈련 코드

- ☐ 직업훈련 종류(분야)는 1-2차년도와 4-22차년도에 조사되었다. 그러나 직업훈련 종류(분야) 오픈코드의 구성은 1차년도와 4, 5-22차년도가 다른 코드로 구성되었으나, 응답항목을 통합하였다.
  - 1차년도의 직업훈련 종류는 707종류로 구성되었으나, 기타항목으로 통합 처리하였다.
  - 4-9차년도의 직업훈련 분야는 최대 208번까지의 코드를 부여하였으나, 기타항목으로 통합 처리하였다.
- ☐ 직업훈련 기관은 3차년도를 제외하고 1차, 2차, 4-22차년도에 조사되었다. 1차년도 조사에서 46종류로 구성되었으나, 기타항목으로 통합 처리하였다. 6차-9차년도 직업훈련 기관은 93번까지 코드를 부여하였으나 기타항목으로 통합처리하였다.
- ☐ 자격증 종류는 2차, 4-22차에 조사되었다. 2차, 4차-6차년도에는 코드변수로, 7차년도 이후에는 문자변수로 제공한다.

### 2.2.2 학과 코드

- ☐ 정규교육 관련 문항은 1-22차년도 자료에 모두 산재해 있으나 학과는 1차년도 개인조사, 2차년도 신규조사자, 3차년도 신규조사자들을 대상으로만 조사하였다.
  - 해당 변수명은 p\*\*5245, p\*\*5265, p\*\*5285, p\*\*5305이다.

◎User notes : 학과 코드의 이용

노동패널의 학과코드는 한국교육개발원에서 개발한 학과분류코드와 연계시킬 수 있다. 교육개발원의 학과 코드는 대분류, 중분류, 소분류까지 다섯자리로 구성되며, 노동패널자료에서 전환가능한 분류는 소분류(2-3자리)까지이다. 학과코드를 전환하는 프로그램은 직업 및 산업전환프로그램과 마찬가지로 한국노동패널 홈페이지에서 '패치파일'메뉴로 들어가면 SAS program file 및 SPSS syntax file을 다운로드 받을 수 있다.

※ 한국노동패널 홈페이지 패치파일 주소 : <http://www.kli.re.kr/klips/>

### 2.2.3 대학교명 코드

□ 대학교명은 1~6차년도에는 구코드를, 7-22년도에서는 신코드로 제공하며, 통합코드북에서 신코드와 구코드 매칭된 코드를 제공하고 있다.

- 해당 변수명은 p\*\*5241, p\*\*5261, p\*\*5281, p\*\*5301, p\*\*5321이다.

### 2.2.4 시군구명 코드

□ 10차년도 공개용 자료부터 시군구코드를 포함해서 제공하고 있다. 각 연도별로 행정구역분류 변경이 된 곳도 있으므로 통합코드북을 참조하기 바란다. 예를 들어, 경상남도 마산시는 2010년 창원시와 통합하여 창원시 마산합포구, 창원시 마산회원구로 분구되었다.

- 해당 변수명은 h\*\*0142이다.

## 2.3 부가자료의 오픈 코드

□ 3, 4차년도에 실시된 부가조사 자료의 오픈 코드는 다음과 같다.

- 청년층 자료(3차) : 가구주와의 관계, 자격증 종류, 산업, 직업, 직업
- 건강과 은퇴 자료(4차) : 가구주와의 관계, 병명

□ 청년층 자료(3차)의 오픈 코드 중 가구주와의 관계, 자격증 종류, 산업, 직업, 직업



는 개인용 자료의 오픈 코드와 동일하다.

- 건강과 은퇴(4차)의 오픈코드는 중 가구주와의 관계는 개인용과 동일하며 병명 오픈코드는 지병 종류(a04061)와 병명(a04095)변수에 대한 오픈코드이다.

### 3. 무응답 대체 데이터(Imputation)

- 가구, 개인, 및 직업력 데이터의 항목 무응답(item nonresponse)에 대한 무응답 대체가 적용된 데이터가 추가로 제공된다(노동패널 홈페이지에서 제공). 무응답 대체 데이터는 원본 데이터와 동일한 변수들이 기록되어 있지만 원본에서 무응답이었던 케이스들의 대체치가 들어가 있으며, 무응답 대체 대상이 되었던 변수들에 대해서는 각 변수에 해당하는 대체여부 표시변수가 추가된다. 대체여부 표시변수를 활용하면 데이터를 따로 열 필요 없이 대체 적용/미적용 비교를 간단히 수행할 수 있다.

- 예를 들어 17차년도 가구데이터 원본파일인 klips17h 에서 ‘연건평’ 변수인 h171409는 ‘모름/무응답’에 해당하는 코드인 ‘-1’이 2 케이스 있다. 동일한 변수를 17차년도 가구데이터 무응답 대체 파일인 klips17h\_i에서 살펴보면 ‘-1’이 하나도 없으며, 그 대신 대체여부 표시변수인 i\_h171409를 보면 원본 자료에서 대체된 케이스만 ‘1’로 기록되어 있음을 확인할 수 있다.

### 4. 직업코드 및 산업코드

- 한국노동패널 원자료 입력시 1-22차년도 조사자료의 경우 산업과 직업의 분류는 한국표준산업분류(8차; 2000년 개정)와 한국표준직업분류(5차; 2000년 개정)의 소분류를 기준으로 하였으며, 3자리로 코드화하는 것을 원칙으로 하였으며, 모름/무응답은 ‘-1’로 처리하였다.
- 한국노동패널 12~22차는 2000년에 개정된 한국표준산업분류(8차, 2000년 개정), 한국표준직업분류(5차, 2000년 개정)와 함께 2007년에 개정된 한국표준산업분류(9차, 2007년 개정), 한국직업분류(6차, 2007년 개정)도 3자리수로 코드화하여 제공하

고 있다. 추가로 2017년에 개정된 한국표준산업분류(10차, 2017년 개정)와 한국표준직업분류(7차, 2017년 개정)도 3자리수로 코드화하여 자료를 제공한다.<sup>14)</sup>

● 직업·산업 분류 과정과 방법은 다음과 같다.

□ 한국노동패널조사의 산업과 직업 코딩 작업은 실사 과정에서 산업과 직업을 전문가가 판별할 수 있게끔 정확한 정보를 수집하고 수집된 정보를 바탕으로 전문가가 코드를 부여하는 방식을 취하고 있다.

□ 산업과 직업 코딩 작업은 다음과 같은 절차를 통해 이뤄진다.

- 첫째, 조사에 들어가기 전에 면접원을 대상으로 통계청의 표준분류 전문가가 직·산업 분류에 필요한 정보를 어떻게 수집해야 하는가에 관한 교육을 진행한다.
- 둘째, 실사가 진행되면 면접원은 사업체의 이름, 사업체의 사업내용, 구체적으로 하는 일과 직위·직책을 Open Code로 상세히 조사한다.
- 셋째, 조사가 완료된 뒤 통계청에서 오랫동안 해당 업무 경험이 있으며 노동패널 직·산업코딩 작업을 계속 수행한 전문가가 Open Code의 내용을 보고 코딩을 부여한다.
- 마지막으로 전문가에 의해 코딩된 자료를 검토하여 누락되거나 잘못 입력된 부분을 확인한 후 입력한다.

□ 산업과 직업 코딩 작업은 원칙적으로 한국표준산업분류(제8차, 2000년 개정)와 한국표준직업분류(제5차, 2000년 개정)에 따라 이루어 졌다. 노동패널 12차 데이터(2009년)부터는 최근 개정한 한국표준산업분류(한국표준산업직업분류 10차, 2017년 개정)와 한국표준직업분류(한국표준직업분류 7차, 2017년 개정)도 함께 제공하여 연구자들의 요구에 부응하고 있다.

14) 21차 이전 노동패널자료를 사용했던 연구자는 직산업 변수명이 변경되었음에 주의를 요함

구분		20차 이전 공개용 자료		21차 이후 공개용 자료	
		개인자료	직업력자료	개인자료	직업력자료
업 종	한국표준산업분류 8차(KSIC8)	p**0330	j630	p**0340	j640
	한국표준산업분류 9차(KSIC9)(12차부터 제공)	p**0331	j631	p**0341	j641
	한국표준산업분류 10차(KSIC10)(12차부터 제공)	-	-	p**0342	j642
직 종	한국표준직업분류 5차(KSCO5)	p**0332	j632	p**0350	j650
	한국표준직업분류 6차(KSCO6)(12차부터 제공)	p**0333	j633	p**0351	j651
	한국표준직업분류 7차(KSCO7)(12차부터 제공)	-	-	p**0352	j652

산업 코드: 6 5 1



한국표준산업분류에 따른 소분류

직업 코드: 5 2 1



한국표준직업분류에 따른 소분류

- 산업코드는 990번이하, 직업코드는 999번이하의 3자리 숫자로 입력되어 있는 경우에는 표준산업분류와 표준직업분류의 소분류(3-digit)와 동일한 값이 입력되어 있다.
- 다만 산업의 경우에 농업, 임업, 어업, 광업은 다른 대분류와는 달리 두 자리가 아닌 한 자리 숫자로 되어 있어(앞자리는 0) 자료상으로는 유일하게 2자리 소분류로 구성된다.
- 한편 한국표준직업분류에서 군인은 다른 대분류와는 달리 숫자가 아닌 알파벳 A로 되어 있어 별도의 3자리 숫자 코드를 부여했다.

□ 직·산업 코딩 과정에서 제공된 Open Code의 정보만으로 소분류(3-digit)까지 분류가 불가능할 경우에는 대분류(2-digit)에 해당되는 값에 0을 붙여 세자리 숫자로 맞추었다.

산업 코드: 6 5 0



한국표준산업분류에 따른 대분류

직업 코드: 5 2 0



한국표준직업분류에 따른 대분류

□ 노동패널 각 연도별 개인자료에서 직업 및 산업코드를 부여받는 변수들은 매우 다양하다. 예컨대, 최종 일자리의 직업산업 코드, 최초 혹은 변경전 코드, 희망하는 일자리의 코드 등등을 들 수 있다. 또한, 자료상에서도 개인자료, 부가조사 자료, 직업력 자료 등 다양한 자료들에 분포되어 있다.

- 여기서는 자료 이용자들이 가장 많이 사용하는 현재(최종) 일자리의 직업 및 산업코드 변수를 소개한다.

<표 IV-7> 현재(최종) 일자리의 직업·산업코드 변수

산업코드						직업코드					
개인자료			직업력자료			개인자료			직업력자료		
구분류	신분류	신분류	구분류	신분류	신분류	구분류	신분류	신분류	구분류	신분류	신분류
8차개정 (2000년 code)	9차개정 (2007년 code)	10차개정 (2017년 code)	8차개정 (2000년 code)	9차개정 (2007년 code)	10차개정 (2017년 code)	5차개정 (2000년 code)	6차개정 (2007년 code)	7차개정 (2017년 code)	5차개정 (2000년 code)	6차개정 (2007년 code)	7차개정 (2017년 code)
p**0340	p**0341	p**0342	j640	j641	j642	p**0350	p**0351	p**0352	j650	j651	j652

## V. 주제별 가이드

### 1. 가구자료

#### 1.1 가구자료의 구조

- ☐ KLIPS는 가구자료와 개인자료로 구성되며, 가구자료의 구조를 이해하는 것이 패널 자료를 분석하기 위한 기본적인 단계가 된다. 가구자료를 이해하기 위해서 우선 가구자료의 전체적인 구조를 살펴보고 가구자료의 주된 내용을 주제별로 제시하면 다음과 같다.
- ☐ IV. 변수 가이드에서도 살펴보았듯이 KLIPS 자료는 key변수, 구성변수, 그 외의 주제별 변수로 구성되며, 22차년도를 기준으로 가구 자료의 key변수와 주제별 변수는 다음과 같다.
  - key 변수 : orghid98, orghid09, orghid18, hhid01-hhid22, hwave01-hwave22, hwaveent, sample98, sample09, sample18
  - 주제별 변수 : key변수를 제외한 모든 변수
- ☐ 22차년도를 기준으로 했을 때 가구자료의 전체적인 구조는 key변수, 구성변수, 그 외의 주제별 변수의 순서대로 구성된다. key변수와 구성변수 다음에 위치하는 주제별 변수는 우선 가구원 모두에 대한 인구학적 정보가 제시되고 설문지의 흐름에 따라 주제별 변수들이 제시된다.
- ☐ 가구용 자료를 분석할 때 유의해야 할 점은 가구자료는 가구가 기본 분석 단위이기 때문에 가구원에 관한 정보는 한 가지 주제에 관하여 변수1에서 변수15까지로 구성된다는 점이다. 예를 들어 22차년도 자료를 기준으로 성별변수는 1번부터 15

번가구원까지의 정보를 담은 h\*\*0241-h\*\*0255의 15개의 변수로 구성되며, 해당 가구의 가구원이 3명이라면 가구원의 성별변수는 h\*\*0241-h\*\*0255 중 해당되는 세 개의 변수에 유효한 값이 들어있다. 그러나 가구원 정보가 아닌 가구단위의 변수(예를 들어 가구소득, 생활비 등)는 한 가구당 하나의 변수로 구성된다.

◎User notes : KLIPS 가구용 자료는 조사된 개인을 모두 자료에 포함하는 개인자료와는 달리 해당 연도에 조사되지 않은 가구에 대한 기본적인 정보도 포함한다. 따라서 각 연도의 가구정보를 분석할 때에는 hwave\*\*변수를 이용하여 해당 연도에 응답하지 않은 가구는 분석에서 제외시켜야 한다.

## 1.2 가구자료의 내용

### 1.2.1 가구원 기본정보

□ 가구자료는 가구원들의 개인고유번호(pid), 성별, 가구주와의 관계, 연령, 학력, 분가시기 등의 기본적인 인적정보들을 담고 있으며, 각 연도별 한 가구당 최대 가구원수는 15명이 할당되어 있다.

□ <표 V-1>은 각 조사연도별 가구원 기본 정보 변수이다.

<표 V-1> 1-22차년도 가구원 정보 변수

	변수명	1차	2차	3차	4차	5-22차
성별	h**0241-h**0255	○				
가구주와의 관계	h**0261-h**0275					
양/음력	h**0281-h**0295					
생년	h**0301-h**0315					
생월	h**0321-h**0335					
생일	h**0341-h**0355					
만나이	h**0361-h**0375					
동거여부	h**0381-h**0395					
비동거이유	h**0401-h**0415					
학력(학교)	h**0661-h**0675					
학력(이수여부)	h**0681-h**0695					
학력(학년)	h**0701-h**0715					
학력(전공계열)	h**0721-h**0731	○	○	-	-	-
졸업/수료/중퇴시기(년)	h**0741-h**0755	-	-	-	-	○
졸업/수료/중퇴시기(월)	h**0761-h**0775	-	-	-	-	○
혼인상태	h**0781-h**0795	-	-	-	○	-
건강상태	h**0801-h**0815	-	○	○	-	-

## 1.2.2 가구소득

□ KLIPS에서 가구소득은 다음의 6가지 항목으로 구성되며 기본적으로 '작년 한해'의 소득을 기준으로 한다. 또한, 4차년도 조사부터는 '지난 한 달간 가구소득'이 추가되었다. 각각의 정의는 다음과 같다.

- 근로소득 : 근로(일)의 대가로 벌어들인 수입을 말하며, 직장 또는 일자리에서 받은 임금이나 봉급, 자영업 소득 등을 포함한다.
- 금융소득 : 금융자산을 통하여 벌어들인 수입으로 예금의 이자소득, 주식의 배당금, 사채 또는 채권이자, 주식의 매매차익 등을 포함한다. 그러나 자산의 가치가 변화하였더라도 실제로 매매를 하지 않은 경우(주식의 시세는 높아졌지만 실제로 주식을 사고 팔지는 않은 경우)는 금융소득에 포함되지 않는다.
- 부동산소득 : 부동산에서 발생하는 수입으로 집세, 토지임대료, 부동산 매매차익 등을 포함한다. 그러나 부동산을 임대해주고 받는 전세금은 다시 갚아야 하는 돈이므로 부동산 소득에 포함되지 않는다.
- 사회보험 : 국민연금, 특수직역연금, 산재보험, 보훈연금, 실업급여, 육아휴직급여, 출산전후휴가급여와 같은 사회보험 수급액을 포함한다.
- 이전소득 : 생활비나 교육비 명목으로 친척/친지로부터 받는 지원금과 정부 및 사회단체로부터 대가없이 받는 보조금(현물 포함)을 포함한다. 재해나 재난 등으로 인해 1회적으로 정부기관에서 받은 소득은 기타소득에 포함된다.
- 기타소득 : 보험금 지급(사고나 질병 등으로 인한 보험금 지급이나 보험을 들었다가 중도해약하고 받은 돈)이나 퇴직금(위의 사회보험에서 받은 연금 제외), 복권 탄 돈, 증여 또는 상속, 축의금, 부의금, 분실물 찾아주고 받은 보상금, 이주민 주거대책비, 폐품매각대금 등과 같이 다른 소득에 포함되지 않은 소득이다.

□ 1-22차년도 가구소득 항목 변경사항은 다음과 같다.

- 지난 한해 동안의 총 근로소득은 1차년도에는 '지난 1년간 연간총소득'으로, 2차년도

이후부터는 ‘**작년 한해의 연간총소득**’으로 단위가 변경되었다.

- 사회보험은 1차년도에는 가구원수의 제한을 두지 않고 가구원 각각의 1회 수급액을 조사했으나, 2, 3차년도에는 가구원 최대 3명까지 조사되었으며, 4차와 5차년도에는 4명까지, 6차년도부터 자료상에는 최대 5명의 가구원까지 기록할 수 있도록 하였다.
- 이전소득은 1차년도에는 지난 1년간의 월평균으로 조사되었으며, 2차년도 이후 조사부터는 작년 한해동안의 연평균으로 조사되었다. 9차년도 이후 친척/친지보조금 항목을 ‘따로 사는 부모님 도움, 따로 사는 자녀의 도움, 그 외 친척/친지 보조금’으로 세분화시켰으며, 각 항목에 대해 유무 여부로 설문 구조를 변경하였다.
- 기타소득 역시 1차년도에는 지난 1년간의 월평균으로 조사되었으며, 2차년도 이후 조사부터는 연평균으로 조사되었다. 9차년도에는 보험금 항목을 ‘손해보험, 저축성보험, 종신보험’으로 세분화하였으며, ‘축의금, 조의금, 복권/경마 등, 교통사고 보상금 등’의 기타 수입을 기타소득으로 조사하도록 항목을 추가하였으며, 각 항목에 대해 유무 여부를 묻는 것으로 설문구조를 변경하였다.

□ <표 V-2>는 1-22차년도 가구소득 관련변수이다.

<표 V-2> 1-22차년도 가구소득 관련 변수

		변수명	1차-22차
지 난 해 소 득	총근로소득	h**2101-h**2102	○
	금융소득	h**2111-h**2117	
	부동산소득	h**2121-h**2127	
	사회보험	h**2131-h**2142	
	이전소득	h**2151-h**2161	
	기타	h**2181-h**2192	
지 난 달 소 득	근로소득	h**2201-h**2202	4차-22차
	금융소득	h**2203-h**2204	
	부동산소득	h**2205-h**2206	
	사회보험	h**2207-h**2208	
	이전소득	h**2209-h**2210	
	기타	h**2211-h**2212	



□ 가구소득 관련 변수 사용 시에는 다음의 사항들에 유의해야 한다.

● KLIPS의 모든 소득항목은 ‘세후소득’을 기준으로 한다.

- 가구소득을 추정하는 대표적인 국내의 통계조사로는 통계청의 『가구소비실태조사』와 『도시가계조사』가 있다. 이 조사의 경우 소득을 ‘세전소득’으로 조사하고 있는데 반해 KLIPS는 ‘조세’를 제외한 ‘세후소득’을 측정한다.

● KLIPS 가구소득의 추정기간은 이전 연도의 소득을 기준으로 한다.

- KLIPS의 가구소득은 조사 연도의 이전 한해를 기준으로 조사되었다. 단 1차년도 조사의 경우 응답일로부터 지난 1년간의 소득을 응답하도록 하였다.

● 각 연도별 측정의 기간이 동일하지 않다.

- 이전소득과 기타소득 또한 1차년도에는 월평균으로 조사하였으나 2차년도 이후부터는 ‘작년 한해’의 소득으로 조사되었다.

● 각 연도별 가구소득의 세부 항목은 차이가 있다.

- 1차년도 가구소득은 세부 항목이 없이 한달 평균 액수만을 측정하였다. 그러나 2차년도 조사부터 가구소득은 다음과 같은 세부항목으로 조사되었다. 각각의 세부 항목 중 부동산 소득의 ‘토지를 도지 준 것’은 3차년도에 추가되었고, 이전 소득의 ‘친척/친지 보조금’은 9차년도에 ‘따로사는 부모님(배우자 부모님 포함) 도움, 따로 사는 자녀의 도움, 그 외 친척/친지 보조금’으로 항목을 세분화하였다. 기타 소득의 ‘증여 및 상속’은 4차년도에 추가되었으며, 부동산 소득의 ‘권리금’은 5차년도 추가되었으며, ‘보험금’은 ‘손해보험, 저축성보험, 종신보험’으로 9차년도에 세분화시켰으며, ‘축의금/조의금, 당첨상금/상으로 받은 상금, 재해보상금’은 9차년도에 추가되었다.

● 가구소득의 세부 항목은 통계청 “1차 가계수지항목분류<sup>15)</sup>”와 차이가 있다.

- 통계청 “1차 가계수지항목분류”에 따르면 가계수지는 경상소득, 비경상소득, 기

---

15) 가계수지항목분류 제1차 개정결과 보고, 통계청(2020. 12)

타 수입, 소비지출, 비소비지출, 기타 지출, 자산, 부채 등 8개 대분류 체계로 구성되어있다. 반면 노동패널조사에서는 해당 분류체계로 조사되지 않기 때문에 통계청 가계소득 관련 조사와 비교·연구를 위해서는 몇몇 항목에 주의해야하는데 가장 차이나는 것은 다음 두 가지이다.

- 노동패널조사의 금융소득 항목 중 “이자소득”과 “배당금” 등은 통계청 기준에서 경상소득의 재산소득에 포함되지만 “주식 및 매매차익”은 기타소득에 포함된다.
- 또한 “부동산 매매차익”은 노동패널조사에서 부동산소득 항목에 포함되지만 통계청 분류기준에 따르면 기타 수입으로 자산변동으로 인한 수입에 해당된다.

<표 V-3> 1-22차년도 가구소득 세부항목

	세부항목
금융소득	은행 및 금융기관 이자소득, 사채이자수입, 주식 및 채권의 매매차익, 배당금, 기타
부동산소득	월세 등 임대료, 부동산 매매차익, 토지를 도지 준 것(3차 추가), 권리금, 기타
사회보험	1회 수급액
이전소득	국민기초생활보호대상, 기타정부보조금, 사회단체 보조금, 친척/친지 보조금(9차 항목세분 : 따로 사는 부모님(배우자 부모님 포함) 도움, 따로 사는 자녀의 도움, 그 외 친척/친지 보조금), 기타 보조금
기타소득	보험금(9차 항목세분 : 손해보험, 저축성보험, 종신보험), 퇴직금, 증여 및 상속(4차 추가), 축의금/조의금, 당첨상금/상으로 받은 상금, 재해보상금(9차 추가), 기타
	근로장려금(13차 추가), 자녀장려금(19차 추가)

◎User notes : KLIPS 자료를 이용하여 가구총소득을 산출할 때는 가구소득의 각각의 항목을 합하면 된다. 단 가구소득 각각의 항목을 합할 때는 우선 해당 항목의 소득 유무를 파악하고 소득이 있는 가구만 선택하여 소득액을 합한다. 또한 가구총소득 산출 시 유의해야 할 점은 각각의 변수마다 'missing'값이 포함(-1로 표기)되어 있으므로 missing값을 통제해주어야 한다. FAQ Q14 참조

### 1.2.3 가구소비

□ KLIPS의 가구소비는 생활비 항목과 저축으로 구성된다. <표 V-4>는 1-22차년도 가구소비 관련 변수이다. 연도별 가구소비 관련 변수를 살펴보면,

- 1, 3차년도에는 생활비를 세부항목으로 조사하지 않고 월평균 총액만 조사하였고 2, 4, 5차년도 이후부터는 총액과 함께 세부항목별 생활비를 조사하였다.
- 또한 세부항목 중 ‘통신비’ 항목은 4차년도에, ‘용돈’ 항목은 5차년도에, ‘피복비’

<표 V-4> 1-22차년도 가구소비 관련 변수

			변수명	1차	2차	3차	4차-22차
월 평 균 생 활 비	식비 (주식비와 부식비)	h**2311	-	○	-		4차-22차
	외식비	h**2312		○			4차-22차
	공교육비(학교등록금,납입금 등)	h**2313		○			4차-22차
	사교육비(학원, 과외비 등)	h**2314		○			4차-22차
	차량유지비(자동차세, 차량보험료 포함)	h**2315		○			4차-22차
	주거비(주택관리비, 월세, 냉난방비 등)	h**2316		○			4차-22차
	경조사비	h**2317		○			4차-22차
	보건의료비	h**2318		○			4차-22차
	문화비(TV시청료, 취미·여가 등에 들어가는 돈)	h**2319		○			4차-22차
	내구재(전자제품이나 가구구입비)	h**2320		○			4차-22차
	통신비(전화, 인터넷, 휴대폰 사용료 등)	h**2321		-			4차-22차
	용돈	h**2322		-			5차-8차
	같이 사는 부모님 용돈	h**2323		-			9차-22차
	자녀용돈	h**2324		-			9차-22차
	그 외 가구원 용돈	h**2325		-			9차-22차
	피복비(옷, 신발 등의 구입비)	h**2326		-			7차-22차
	현금 및 각종기부금	h**2327		-			8차-22차
	국민연금 (및 의료보험료)	h**2328		-			8차-22차
	건강보험료(의료보험료)	h**2332		-			11차-22차
	대중교통비	h**2329		-			8차-22차
	생필품 구입비	h**2330		-			10차-22차
	기타	h**2331		-			4차-22차
	월평균 총액	h**2301					
저 축	예·적금(주택청약부금 포함)	h**2412	-	-	-		4차-22차
	개인연금	h**2414					4차-22차
	보장성보험(생명보험, 상해보험 등)	h**2416					4차-22차
	종신보험	h**2418					9차-22차
	적립식펀드	h**2426					10차-22차
	저축성보험(태테크보험, 교육보험 등)	h**2420					4차-22차
	계(→갯돈 불입금)	h**2422					4차-22차
	기타	h**2424					4차-22차
	월평균 총액	h**2402					

주: 경조사비, 현금 및 각종 기부금, 국민연금과 건강보험료는 비소비지출에 해당함.

항목은 6차년도에 각각 추가되었다. 그리고 7차년도에는 ‘헌금 및 각종기부금’, ‘국민연금 및 의료보험료’, ‘대중교통비’ 등이 추가되었다. 9차년도에는 ‘문화비’는 ‘교양오락비’로 항목 변경되었고, ‘용돈’은 ‘같이 사는 부모님 용돈, 자녀 용돈, 그 외 가구원의 용돈’으로 항목을 세분화시켰고, ‘생필품 구입비’는 추가되었다.

- 저축은 1, 2, 3차년도에는 월평균 총액만 조사하였고 4·8차년도에는 월평균 총액과 함께 세부항목별로 조사하였다. 9차년도에는 보장성 보험을 ‘보장성 보험과 종신보험’으로 세분하였다. 10차년도에는 ‘적립식편드’가 추가되었다.
- 지역 의료보험 가입자의 정보 수집을 위하여 11차년도에는 ‘국민연금 및 의료보험료’ 항목을 세분화하여 ‘건강보험료(의료보험료)’ 항목이 추가되었다.

☐ 가구소비는 가구소득과 마찬가지로 ‘작년 한해’를 기준으로 조사하였다. 단 1차년도의 경우 응답일로부터 지난 1년간의 소득을 조사했다.

#### 1.2.4 자산

☐ 자산과 관련된 항목은 거주주택을 제외한 부동산과 금융자산으로 구성된다. 자산 항목은 1차년도에는 조사되지 않았으며, 2차년도부터 조사되었다. 자산의 측정은 가구소비와 가구소득과는 달리 지난 한해가 아닌 ‘조사 당시’를 기준으로 한다.

☐ 10차년도부터 기타자산 항목인 ‘자동차 소유 여부’ 및 ‘구매 여부와 금액’에 대하여 추가 조사되었다. 자동차 자산의 측정은 ‘작년 한해’를 기준으로 한다. 11차년도에는 자산에 대한 많은 정보를 수집하기 위하여 ‘기타자산 보유 여부’ 및 ‘기타자산의 현재시가’에 대하여 추가되었다.

☐ <표 V-5>는 자산 항목 중 2·22차년도 부동산 관련 변수들이다.

<표 V-5> 2·22차년도 부동산 관련 변수

	변수명	2차-22차
거주주택 외 부동산	h**2501-h**2513	○
부동산 타인 임대	h**2521-h**2524	
부동산 임차	h**2531-h**2543	

□ 부동산 관련 변수를 사용할 때 유의할 점은 다른 자산 항목들과는 달리 시가 총액 또는 보증금 총액을 측정하는데 있어서 연속형으로 우선 측정한 다음 ‘잘모르겠다’라고 응답한 가구에 대하여 범주형으로 다시 응답하도록 하였다. 또한 범주형의 응답 범주는 차수마다 응답항목이 달랐으나 변수통합 자료에는 응답항목도 10차년도 기준으로 맞추었다.

□ <표 V-6>은 자산항목 중 2-22차년도 금융자산 관련 변수이다. 금융자산 관련 변수 중 기타 금융자산종류는 1, 2차년도에는 별도로 코드화하지 않았지만 3차년도 이후부터는 별도로 코드화하여 하나의 변수로 포함되었다.

- 보장성 보험은 17차년도(2014년)부터 추가되었으며, ‘저축성 보험 및 보장성 보험’에 포함하여 조사하고 있다.

<표 V-6> 2-22차년도 금융자산 관련 변수

	변수명	2-22차
은행예금	h**2561-h**2562	○
주식, 채권, 신탁	h**2563-h**2564	
저축성 보험 및 보장성 보험	h**2565-h**2566	
아직 타지 않은 게	h**2567-h**2568	
개인적으로 다른 사람에게 빌려준 돈	h**2569-h**2570	
기타	h**2571-h**2572	

### 1.2.5 부채

□ 2-22차년도 가구부채 관련 변수는 <표 V-7>과 같다. 가구부채에 대한 조사는 조사연도별로 변동이 있다.

- 1차년도에는 부채의 세부 항목의 구분 없이 부채여부, 총 부채액, 월평균 부채 상환금만을 조사했으나, 2차년도부터는 다음의 총 6개 항목으로 구분하여 조사하였으며, 총 부채액은 별도로 조사하지 않았다.

<표 V-7> 2-22차년도 가구부채 관련 변수

	변수명	2-22차
금융기관 부채	h**2601-h**2603	○
비금융기관 부채	h**2604-h**2606	
개인적으로 빌린 돈	h**2607-h**2609	
전세금 임대보증금	h**2610-h**2612	
부어야 할 게	h**2613-h**2615	
기타	h**2616-h**2618	

### 1.2.6 주거지

□ 주거지는 이사여부, 주거지 변동사항, 입주형태, 주택의 종류, 주택의 평수, 주거지의 시가, 거주시작 시기 등에 관하여 조사되었다. 22차년도를 기준으로 주거지의 세부적인 조사 항목은 다음과 같다.

- 입주형태 : 자가, 전세, 월세, 기타
- 주택종류 : 단독주택, 아파트, 연립주택, 다세대(다가구)주택, 상가주택, 기타
- 주택의 평수 : 대지면적, 연건평, 실제사용평수
- 거주시작시기 : 거주시작 년, 월
- 주거지의 시가, 임대보증금 혹은 전(월)세금

<표 V-8> 1-22차년도 주거지 관련 변수

	변수명	1-22차
현재 거주지 이사 동기	h**1420	11-22차
이사여부	h**1401	2-22차
변동사항	h**1402~h**1405	
입주형태	h**1406	1-22차
주택종류	h**1407	
대지면적, 연건평, 평수	h**1408-h**1411	
시가	h**1412	
보증금, 월세	h**1413-h**1414	
거주 시작시기	h**1415-h**1416	

□ 1-22차년도 주거지 항목 변경사항은 다음과 같다.

- 이사여부와 주거지 변동 사항은 2차년도부터 추가되었다.
- 주택의 평수는 1-3차년도에는 대지면적과 연건평을 구분하여 조사하였고, 4차년도 이후에는 대지면적, 연건평, 실제평수의 세 가지 항목으로 구분하여 조사하였다.
- 주택 종류는 2차년도에는 5개의 항목, 3차년도에는 9개의 항목, 4·7차년도에는 5개의 항목, 8차년도부터는 6개의 항목으로 조사되었고 각각의 항목의 내용에도 변동이 있다.
- 11차년도에는 지역 노동시장의 이동을 살피고자 현재 거주지 이사 동기 문항을 추가하였다.

◎User notes : KLIPS 가구용 자료의 주거지 관련 변수를 사용할 때 주의할 점은 2차와 3차 자료에서 입주 형태, 주택의 종류, 주택의 평수, 주거지의 시가, 거주시기는 전년도 조사 이후 이사하였거나 이사하지 않았더라도 변동사항이 있을 경우에만 응답하도록 하였다. 따라서 기존의 2, 3차 자료를 이용하여 주거지에 관한 분석을 할 때에는 전년도 자료에서 해당 변수를 연결하여 사용해야했다. 그러나 8차 자료부터는 해당변수를 가공하여 붙여두었으므로, 별도의 작업을 하지 않고 해당변수를 사용하면 된다. FAQ Q19 참조

### 1.2.7 자녀교육

□ 자녀교육은 ‘사교육 및 보육기관 이용실태와 비용’과 ‘탁아모 이용실태와 비용’으로 구분되어 조사되었다.

- 사교육 및 보육기관 이용실태와 비용은 2차년도를 제외한 모든 연도에서, 탁아모 이용실태와 비용은 1, 3, 4차년도에만 조사되었다.

□ <표 V-9>는 사교육 및 보육기관 이용실태 조사 내용이다. 조사 내용은 차수별로 차이가 있다.

- 1차년도에는 조사대상을 ‘초등학교 3학년 이하 자녀’로 한정하였으나 3차년도 이후로는 ‘0세-고등학생 이하(재수생 포함) 자녀’를 조사 대상으로 한다. 그러나 16차년도부터는 ‘0세-미취학 자녀’와 ‘취학-재수생 자녀’로 세분화하여 미취학자녀의 자녀 보육 및 양육에 관한 질문을 강화하였다.
- 7차년도 조사부터는 대학생 이상 자녀에 대한 교육비도 조사하였다.
- 이용기관 종류의 항목은 연도별로 차이가 있다.

☐ 사교육 및 보육기관 이용에 대한 경제적 부담정도는 3차년도 이후에 조사되었다.

◎User notes : 자녀의 사교육 및 보육과 관련하여 6차년도 조사 이후 가장 큰 변동사항은 사교육 이용기관 종류별로 조사되었던 항목들이 자녀 당 항목으로 대폭 축소되었다는 점이다. 예컨대, 5차년도 조사에서는 자녀가 2명이고 두 명이 각각 3개씩의 학원에 다닐 경우 그 가구에 대해서는 총 6개의 사교육종류별로 이용횟수와 이용시간, 월평균 수업료 등을 질문하게 된다. 그러나 6차년도 조사이후에는 동일한 상황에 대해 총 2개의 월평균수업료만을 파악하게 된다. 한편 7차년도 조사의 큰 변화로는 대학생이상 자녀의 사교육비를 조사하였다는 점을 들 수 있다. 대학생 이상 자녀의 수, 정규학교 등록금, 그 외 지출비용 등이 조사되었다.

☐ 탁아모 이용실태의 조사 내용은 사교육 및 보육기관 이용과 유사하며, 그 내용은 탁아모 종류, 탁아모 이용 횟수와 시간, 비용, 경제적 부담정도 등이다.

☐ 탁아모 이용실태는 1차년도에는 ‘초등학교 3학년 이하’의 자녀만을 조사내용으로 하고 있으나 3, 4차년도에는 ‘초등학교 6학년 이하’의 자녀로 조사대상으로 확대하였다. 탁아모 이용으로 인한 경제적 부담정도는 3, 4차년도에만 조사되었다. 그러나, 5차년도 조사부터는 탁아모 이용실태와 관련한 조사가 실시되지 않았다.

☐ 자녀교육 분석에서 유의할 점은 변수구성에 있어서 한 자녀가 이용하는 하나의 기관(혹은 탁아모)단위로 변수가 구성된다는 점이다. 또한 자녀교육 문항에는 교육기관을 이용하는 자녀의 번호와 엄마의 번호를 기재하도록 하였는데 여기서 자녀번호와 엄마번호는 자녀와 엄마의 가구내 개인번호를 의미한다.

- 자녀번호를 가구원 정보와 연결하여 사용할 경우, 가구번호와 자녀(엄마)번호를 조합하여 merge한다. FAQ Q12 참조

☐ 16차년도 조사부터는 ‘0세 이상~고등학생이하(재수생 포함) 자녀’에 대한 문항이 ‘0세~미취학 자녀’와 ‘취학~재수생 자녀’로 구분하였으며, 미취학자녀의 자녀 보육 및 양육에 대한 문항이 새롭게 추가되었다.



<표 V-9> 사교육 및 보육기관 이용실태 조사내용

조사차수	조사 대상	조사 내용
1차	초등학교 3학년 이하 자녀	이용기관 종류 이용횟수와 이용시간 월평균 수업료
3차 4차 5차	0세-고등학생 이하(재수생 포함)	이용기관 종류 종류별 이용횟수와 이용시간 종류별 월평균 수업료 가구 전체에 대한 경제적 부담정도
6차		자녀당 이용여부 자녀당 월평균 수업료 가구 전체에 대한 경제적 부담정도
7/15차	0세-고등학생 이하(재수생 포함) 대학생 이상 자녀(휴학 중인 자녀 포함)	고등학교이하 자녀당 이용개수 고등학교이하 자녀당 월평균 수업료 대학생이상 자녀당 정규학교 등록금 대학생이상 자녀당 그 외 지출 비용 가구 전체에 대한 경제적 부담정도
16차-22차	0세-미취학 취학-재수생 대학생 이상 자녀(휴학 중인 자녀 포함)	종일제 아이돌보미 서비스 이용여부 아이사랑카드 사용여부 자녀양육 수당 지급 받고 있는지 여부 월평균 지급 받는 양육수당 금액 이용하는 방과 후 교육 및 보육서비스 비용 정부지원 여부 고등학교이하 자녀당 이용여부 고등학교이하 자녀당 사교육종류, 교육 및 보육비 대학생이상 자녀당 정규학교 등록금 대학생이상 자녀당 그 외 지출 비용 가구 전체에 대한 경제적 부담정도

□ 2017년 0세-미취학 자녀의 보육 및 양육관련 회고 조사를 실시하였다.

- 16차년도(2013년) 조사에서 미취학자녀의 보육 및 양육에 대한 문항이 추가되면서, 설문 구조와 문항을 수정하는 과정에서 보육연령에 대한 오류를 19차년도(2017년) 조사 때 발견하였다.
- 16-18차년도 조사(2013-2015년)에서 당시 보육연령 5세(취학 전 1년)인 자녀를 둔 미응답자에 대해서 자녀 보육 및 양육 관련 회고 조사를 실시하였다.
- 수집된 회고 조사는 가구자료에 동일한 형태로 들어 있다. 단, 자녀 보육 및 양육에 대한 '회고 조사 여부'를 나타내는 flag 변수를 추가하였다.

◎User notes : ‘보육 및 양육 회고 조사 flag 변수’는 해당하는 자녀를 알려주는 변수이다. 회고 조사 응답값을 가지는 자녀의 경우에는 (1) 해당함으로 표기하였다. 변수명은 ‘fh’로 시작하여, 가구자료의 변수명과 다르다는 점을 주의해야 한다. \*\*는 해당하는 조사 차수(16-18차)를 의미한다.

fh\*\*1511 : [보육 및 양육 회고 조사 flag 변수] 자녀번호 - 1

fh\*\*1611 : [보육 및 양육 회고 조사 flag 변수] 자녀번호 - 2

fh\*\*1711 : [보육 및 양육 회고 조사 flag 변수] 자녀번호 - 3

fh\*\*1811 : [보육 및 양육 회고 조사 flag 변수] 자녀번호 - 4

fh\*\*1911 : [보육 및 양육 회고 조사 flag 변수] 자녀번호 - 5

- 회고조사 자료의 특징과 응답값의 신뢰도를 검토한 보고서(『패널자료품질개선연구(VII)』, “KLIPS 보육관련 설문조사의 문제와 데이터 품질 제고 방안:”, 최효미 (2017))가 작성되었다.
- 전반적으로 16-18차년도에 걸친 회고 조사 자료는 조사 과정과 결과에 있어서 큰 문제가 없다는 것을 연구를 통해 밝혔다.
- 회고 조사와 별개로 추가로 언급한 내용이 있다. 16차년도부터 미취학 자녀의 보육 및 양육에 대한 문항을 추가하면서 발생한 기존 설문구조와의 관계이다. 15차년도 이전 자료까지는 사교육비가 포함되는 대신 개별돌봄서비스 이용비용은 과소하게 책정되었을 가능성이 높은 반면에, 16차년도 이후 자료는 사교육비가 아예 반영되지 않은 대신 개별돌봄서비스 비용은 크게 잡혀 있는 상황이 발생할 수 있다는 점이다.

## 2. 개인자료

### 2.1 개인자료의 구조

- KLIPS 가구용 자료가 가구를 분석단위로 하고 있다면 개인용 자료는 가구에 속한 15세 이상의 가구원을 분석단위로 한다. 개인자료는 가구자료와 동일하게 key변수와 주제별 변수로 구성된다.
- 21차년도를 기준으로 개인용 자료의 key변수와 구성변수, 그 외의 주제별 변수는 다음과 같다.

- key변수 : pid, hhid01 ~ hhid22, hmem22
- 주제별 변수 : key변수를 제외한 모든 변수

□ 주제별 변수는 그 변수가 어떤 설문지에서 만들어졌는가에 따라 변수레이블 앞에 ‘(가구정보)’, ‘(주된 일자리)’, ‘(모든 응답자)’, ‘(취업)’, ‘(미취업)’, ‘(신규조사자)’라는 라벨이 붙는다.

#### ● 주제별 변수의 구성

- (가구정보) : 가구자료에서 조사된 가구원 정보를 개인자료에 연결한 변수이다. 개인의 주된 인구학적 특성을 담고 있다.
- (주된 일자리) : 취업자가 응답한 유형 설문의 내용 중 ‘주된 일자리’라고 응답한 일자리 관련 변수들이다.
- (모든응답자) : 취업자 개인과 미취업자 개인 설문 중 취업자 미취업자 모두에게 공통적으로 질문한 설문내용에 관한 변수로 각 변수에는 취업자와 미취업자 모두의 응답내용이 들어 있다.
- (취업) : 취업자 설문 중 취업자에게만 해당하는 내용에 관한 변수이다.
- (미취업) : 미취업자 설문 중 미취업자에게만 해당하는 내용에 관한 변수이다.
- (신규조사자) : 1차년도 응답자 및 만 15세가 되어 처음 조사되거나 합가 등의 이유로 추가된 가구원에게만 해당하는 내용에 관한 변수들이다.

□ 개인 자료에서 변수레이블 앞에 ‘(주된 일자리)’라고 입력된 변수들은 취업자설문과 유형설문을 토대로 별도로 가공된 자료이다. 이 변수들은 다음과 같이 구성된다.

- KLIPS는 개인이 여러 가지 일자리를 가지고 있을 경우 유형 설문을 통해 모든 일자리에 관한 응답을 받고 있다. 단 취업자 개인 설문은 주된 일자리를 기준으로 응답받는다.
- 주된 일자리는 현재의 일자리(유형1, 3, 5, 7 설문) 중에서 응답자가 판단하기에 가장 중요한 일자리로 정의된다.

□ 개인용 자료의 (주된 일자리)변수들이 담고 있는 내용은 다음과 같다.

- 취업(사업)시작 시기
- 산업, 직업, 직위
- 기업의 형태 및 규모

- 고용형태와 근로계약유무 및 기간, 근로시간 형태, 종사상 지위
- 근로시간, 임금
- 사회보험 가입여부 및 수급
- 노동조합
- 농업일자리(12차년도부터 추가)

## 2.2 개인자료의 내용

### 2.2.1 평소의 경제활동

- ☐ 평소의 경제활동은 개인용 자료의 (모든응답자)변수로 지난주 주요 활동과 지난주 소득 활동, 지난주 직장유무, 지난주 일 하지 않은 이유를 묻고 있다.
- ☐ 평소 경제활동 관련 문항은 1차년도에는 가구용 설문지에 있었으나 자료상에는 개인용에서 제공된다. 3-22차년도에는 개인용 설문에서 묻고 있으며 개인용 자료에서 제공된다.

<표 V-10> 평소의 경제활동관련 변수

	변수명	1차-22차
평소 활동	p**0202	3차-22차
지난 1주 주요활동	p**0203	1차-22차
지난 1주간 소득활동	p**0204	
지난주 일자리 소유여부	p**0205	
지난주 일하지 않은 이유	p**0206	

### 2.2.2 구직활동

- ☐ 구직활동은 취업자의 구직활동과 미취업자의 구직활동으로 구분된다.
  - 1차년도 개인용 자료에서 취업자의 구직활동 관련해서는 희망 종사상 지위, 희망 근로시간, 희망하는 일의 종류, 구직방법, 구직시 어려움, 구직기간 등의 자세한 내용을 담고 있다. 미취업자의 구직활동은 실업자의 구직활동과 비경제활동

인구의 구직활동에 관한 내용으로 구분된다.

- 2, 3차년도의 구직 관련 변수는 취업자에 대해서는 현재 직장의 구직 과정 관련 사항과 취업 중 구직활동(on the job search)에 관한 사항을 묻고 있다. 미취업자의 경우에는 실업자를 대상으로 구직활동에 관한 내용을 묻고 있다.
- 4차~9차년도의 구직활동 관련 변수구성은 3차와 동일하다. 단 4차년도부터 새로운 일자리에 대한 구직 방법 외에 성공한 구직방법과 구체적인 일자리 진입방법 변수가 추가되었고, 구직방법에 대한 세부 항목이 구체화되었다. 9차년도에는 미취업자의 구직활동 변수 중 지난 조사이후 구직활동여부, 마지막 구직활동시기(년, 월), 지난주 일자리 제의 여부, 제의받은 일자리에 취업하지 않은 이유에 관한 내용이 추가되었다.

<표 V-11> 1-22차년도 구직활동 관련 변수

	변수명	1차-22차
취업자의 구직활동	p**2901~p**3020	1차-4차
미취업자의 구직활동	p**2801~p**2885	1차-22차

### 2.2.3 고용형태(정규/비정규여부) 및 근로계약

- ☐ KLIPS에서 임금근로자의 고용형태는 정규직·비정규직 여부를 의미한다. 정규직 여부는 개인의 주관적인 판단 하에 자기 응답식으로 측정되었다. 그밖에 임금근로자의 고용형태 관련 항목은 IV. 변수 가이드의 1.key 변수의 이해와 활용에 제시되어 있다.
- ☐ 임금근로자의 근로계약관련변수는 근로계약기간유무, 근로계약기간, 계속근로가능여부, 근로계약문서 수령여부, 근로계약갱신여부, 향후근무예상기간으로 구성된다. 근로계약관련변수의 연도별 변화사항은 다음과 같다.
  - 1-22차년도 조사에서 근로계약기간 유무와 근로계약기간은 동일하게 조사되었다.
  - 계속 근로 가능 여부는 2차를 제외한 1, 3차년도에 조사되었다.
  - 근로계약문서 수령여부는 3-22차년도에 조사되었다.
  - 근로계약 갱신여부는 3차, 5-22차년도에 조사되었다.

<표 V-12> 1-22차년도 근로계약 관련 변수

	변수명	1차-22차
근로 계약 기간 유무	p**0501	1차-22차
근로 계약 기간	p**0502 ~ p**0504	
근로 계약 갱신 여부	p**0505	5차-22차
불규칙적인 일자리 여부	p**0508	5차-22차
근로 계약문서 수령	p**0509	3차-22차

□ 노동패널조사에서 ‘향후 이 일자리에서 계속 근로가 가능한지 여부’는 3차년도 조사에서는 임금근로자의 유형설문에서 조사되었다. 동일한 문항에 대해 5차년도에는 통계청의 경제활동인구조사 중 2002년 비정규직 관련 부가조사를 통해 유사한 문항이 설문되었다. 노동패널 조사에서도 이를 반영하기 위해서 5차년도 조사에서 부터는 취업자 공통 설문에서 이를 조사하였다.

- 이중 향후 계속 근로가능기간에 대해서도 4차년도 조사에서 ‘1년 이상’과 ‘1년 미만’으로 나누어 조사하였다. 즉, 기간제(단기 계약직) 근로자에 대해서는 5차년도 이전조사에서도 파악이 가능하였다고 볼 수 있다.
- 5차~11차년도에는 향후 근로가능 기간에 대해 경제활동인구조사와 응답항목을 통일시켜 ‘(1) 1년 이하(개월), (2) 1년-3년 이하 (3) 3년 초과’와 같이 세 가지 형태로, 12차년도 이후에는 (1) 1년 이하, (2) 1년 초과~2년 이하, (3) 1년 초과~3년 이하, (4) 3년 초과로 네 가지로 나누어 질문하였다.
- 그 외에 호출 근로나 시간제 근로와 같은 형태는 노동패널내의 기존 유형설문에서 판별이 가능하였기 때문에, <표 V-13>과 같이 파견/용역근로여부, 독립도급 여부, 가내 근로 여부 등에 대해서만 추가로 조사하였다.

<표 V-13> 비정규직 측정과 관련된 다양한 변수들

	변수명	3차	4차	5차-22차
계속 근로가능 여부	p**0601	○	-	○
향후 계속근로가능기간	p**0603~p**0604	○		
계속근로불가능이유	p**0605	-		
파견/용역여부	p**0611	○		
독립도급여부	p**0612	-		
가내(제택)근로 여부	p**0613	-		

□ 10차년도 부가조사에서 조사되었던 ‘고용형태’ 문항들은 11차년도부터 개인용 취업자 설문에 포함시켜 조사하였다. 고용형태 자료 내용은 <표V-14>와 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 임시적 고용 : 근로계약 기간이 임시적이거나 업무의 성격이 임시적이기 때문에 불안정성을 가지고 있는 고용형태이다. 여기에는 한시근로, 프로젝트베이스, 대체근로, 성과계약제, 임시직, 계절근로 등이 포함된다.
- 간접고용 : 일하는 곳과 임금을 지급하는 곳이 다른 고용형태에는 파견/용역근로나 사내하청 등이 포함된다.
- 일일/호출근로 : 근로계약의 유무에 제약받지 않고 ‘필요할 때 연락을 받고 며칠 또는 몇 주 동안 고용’되거나 ‘인력시장, 직업소개소 등 일정한 장소에서 대기하고 있다가 고용주나 고용할 업체 등에서 일할 사람을 선택하면 일일 단위로 일하는’ 경우 등을 의미한다.
- 특수고용 : 임금과 비임금의 경계선상에 있는 특수형태 고용은 임금근로자 뿐만 아니라 ‘일정한 사업장이 없는 비임금근로자’ 중에서 보험설계사, 학습지교사, 쿼터서비스 배달기사 등과 같이 임금근로자와 자영자의 경계가 모호한 고용형태를 의미한다.
- 시간제 근로 : 파트타임·아르바이트로 일하거나, 같은 업무에 종사하는 사람들보다 적은 시간 동안 일하거나, 임금이 시간 단위로 지급되는 경우로 정의된다.

- 17차년도 조사부터는 기존의 방대한 비정규직 판별문항을 대폭 삭제하여 간소화하였다.

<표 V-14> 고용형태 자료의 변수

내 용	10차(부가조사)	11차-16차(취업자)	17차-22차(취업자)
고용형태 판별	a105101-a105121	pa**5101-pa**5121	pa**5101-pa**5103 pa**5107, pa**5108 pa**5900-pa**5911
한시근로 및 임시적 계약형태	a105200-a105223	pa**5200-pa**5223	pa**5200-pa**5202, pa**5211, pa**5215, pa**5216
간접고용	a105300-a105369	pa**5300-pa**5369	삭제
일일호출근로	a105400-a105405	pa**5400-pa**5405	삭제
특수형태고용	a105500-a105535	pa**5500-pa**5535	pa**5500-pa**5504, pa**5509
시간제 근로	a105600-a105618	pa**5600-pa**5618	삭제
직장상황 및 적응	a105700-a105705	pa**5700-pa**5705	삭제

◎User notes : 비정규직 근로자의 판별(FAQ Q23참조)

5차년도 조사부터는 통계청의 경제활동인구조사의 비정규직 관련 부가조사(2002년)의 내용들을 반영함으로써 보다 다양한 형태의 비정규직 구분이 가능하게 되었다. 한국노동패널에서 측정가능한 비정규직의 개념들을 살펴보면 다음과 같다.

1. 종사상 지위 : 상용직, 임시직, 일용직 기준 p\*\*0314를 참조할 것.
2. 주관적 판단에 의한 비정규직 여부 <표 V-13>을 참조할 것.
3. 다양한 형태의 대안적 근로(단기계약, 호출근로, 파견근로, 용역근로, 독립도급근로, 가내근로, 시간제 근로)

※ '다양한 형태의 대안적 근로(alternative work arrangements)'와 관련된 분석에 사용된 변수들을 제시하면 다음과 같다.

항 목	변수명
한시적 근로	
기간제근로	p**0501(=1)
비기간제근로	p**0501(=2) & p**0601(=2) & p**0605(=1,2,3,4,5,6), p**0501(=2) & p**0602(=2)
시간제근로	p**0315(=1)
비전형근로	
파견근로	p**0611(=2)
용역근로	p**0611(=3)
독립도급근로	p**0612(=1)
가정내근로	p**0613(=1)
일일(단기)근로	p**0508(=1)

통계청에서 실시한 「경제활동인구조사」의 비정규직 부가조사의 내용을 비교하기 위해서는 **홍민기 외 2015, 노동패널자료 연구(I), 제3장 한국노동패널조사에서 나타난 비정규직 현황과 추세(성재민)**,을 참조하기 바란다.



## 2.2.4 근로시간

- 정규근로시간은 임금근로자에게만 해당하는 변수이다. 정규근로시간과 관련해서는 정규근로시간유무여부, 주당정규근로시간 및 일수, 초과근로여부, 주당초과근로시간 및 일수 등을 묻고 있다.
- 근로시간과 관련하여 1-22차년도별 차이점은 다음과 같다.
  - 1차년도에는 주당정규근로시간과 주당초과근로시간, 주당정규근로일수에 관한 변수만 있고 주당초과근로일수에 관한 정보는 없다.
- 근로시간관련 변수를 사용할 때의 유의사항과 근로시간관련 추가 사항은 다음과 같다.
  - 정규근로시간의 유무에 따라 근로시간 응답 항목이 달라지기 때문에 총근로시간을 구성할 때는 이점에 유의해야 한다.
  - 1차년도에는 지난 1년 동안 ‘가장 많이 일한 달’과 ‘가장 적게 일한 달’의 기간에 관한 내용이 추가되었다.
  - 2차년도에는 지난 1주간의 근로시간이 평상시와 비교하여 어땠는지에 관한 내용이 추가 되었다.
  - 1-6차년도에는 정규근로시간 이외에 각 요일별 근무시간에 관한 정보를 담고 있다.
  - 3-4차년도에는 근로시간단축 견해에 관한 정보를 담고 있다.
  - 8차년도부터 근로시간을 정하는 방법, 주5일제 확산에 따른 근로시간 변화여부, 근로시간 단축으로 생긴 여유시간을 어디에 활용하고 있는지에 대한 정보를 담고 있다(p\*\*1151 ~ p\*\*1224).

<표 V-15> 1-22차 근로시간 관련 변수

	변수명	1차-22차
정규근로시간 유무	p**1003	○
주당정규근로시간 및 일수	p**1006~p**1007	
주당평균근로시간 및 일수(임금)	p**1004~p**1005	
주당평균근로시간 및 일수(비임금)	p**1031~p**1032	
초과근로여부 및 구분	p**1011, p**1019	
주당초과근로시간 및 일수	p**10121~p**1013	
초과근로수당 지급여부	p**1015	
초과근로수당 지급방법	p**1016~p**1017	
월평균 초과근로수당	p**1018	

◎User notes : 임금근로자의 근로시간은 정규근로시간의 유무에 따라 해당 문항이 달라진다. 따라서 임금근로자의 주당 평균 근로시간을 산출할 때는 정규 근로 시간이 있다고 응답한 경우에는 주당 정규근로시간을, 정규 근로 시간이 없다고 응답한 경우에는 주당 평균 근무시간을 참조하여 주당 평균 근로시간을 산출해야 한다. FAQ Q24 참조

## 2.2.5 임금 및 소득

□ 임금 및 소득과 관련된 정보는 ‘(주된 일자리)’를 기준으로 구성되었다.

- 비임금근로자의 소득은 연간매출액과 소득액을 측정하였으며, 임금근로자의 임금은 임금지급주기, 임금결정방식, 임금 수준, 성과급적용여부, 임금내용 등의 자세한 항목으로 구성된다.

□ 임금근로자의 임금관련 변수는 <표 V-16>와 같으며, 임금관련 변수의 년도별 변동사항과 유의 사항은 다음과 같다.

- 임금수준은 해당하는 임금결정방식에 따라 응답하도록 하였으며 설문지 구성과 자료의 구성에는 약간의 차이가 있다.
- 임금수준의 자료의 구성은 1차년도에는 연봉제, 일당제, 시간급제의 세 가지 변수로 구성되었고 2차년도에는 도급제를 제외한 6개의 변수로 구성되었다. 3차년

도에는 도급제와 기타를 제외한 5개의 변수로 구성되었으며 4차년도에는 기타를 제외한 6개의 변수로 구성되었다.

- 임금 내용은 기본급, 상여금, 기타수당 구분 여부와 각각의 액수에 관한 내용으로 구성되었으며 1차년도에만 조사되었다.
- 월평균 임금은 임금유무의 임금 액수로 구성되었다. 7차년도부터 공제되는 세금액이 추가되었다.
- 9차년도부터 임금수준 변수 중 연봉액과 월급액 변수를 기존의 월평균 임금 변수로도 가능하기 때문에 중복되므로 삭제하였다.

<표 V-16> 1-22차년도 임금근로자의 임금관련 변수

	변수명	1차-22차
임금지급방법	p**1601	○
임금결정방식	p**1602	
임금수준	p**1611 ~ p**1617	
성과급 관련	p**1621 ~ p**1624	
임금내용	p**1632 ~ p**1634	
월평균임금	p**1641 ~ p**1642	
임금에서 공제되는 세금	p**1643	

□ 1-22차년도 비임금근로자의 소득 관련 변수는 <표 V-17>과 같다. 비임금근로자의 소득은 연간매출액과 월평균 소득으로 조사되었다. 비임금근로자의 소득과 관련하여 1-9차년도 변화사항은 다음과 같다.

- 1차-3차년도까지 연간 매출액은 개방형으로 측정하였고 모름일 경우 범주형으로 측정하였다. 그러나 4차년도부터는 개방형의 문항으로 측정하였다.
- 연간매출액의 범주형의 세부항목 중 2, 3차년도에는 '(14)모르겠다'가 추가되었다.
- 월평균 소득관련 변수는 1-9차년도가 동일하다.
- 8차년도부터 적자액 정보가 추가되었다.
- 21차년도부터 연간 임대료와 연간인건비, 항목별 투자액, 부동산비용, 권리금을 추가되었다.

<표 V-17> 1-22차년도 비임금근로자의 소득관련 변수

	변수명	1차-22차
연간매출액	p**1662	1차-22차
월평균소득	p**1671~p**1672	
월평균 적자액	p**1673	8차-22차
연간 임대료	p**1691	22차
연간 인건비	p**1692	22차
총투자액	p**3211~p**3217	22차
연간 부동산 비용	p**1693~p**1694	22차
권리금	p**1697	22차

◎User notes : 비임금 근로자의 월평균 소득은 임금근로자의 월평균 소득과 비교할 때 missing 값이 많다는 질문이 많다. 그러나, 비임금 근로자 중에서 종사상 지위가 '무급가족종사자'인 경우에는 응답 대상이 아니다. 또한, '적자를 보고 있다'고 응답한 경우도 상당수에 이른다. 이런 사례들을 통제하고 나면 비임금 근로자의 소득에도 missing값이 거의 나타나지 않음을 확인할 수 있다.  
FAQ Q25참조

## 2.2.6 직업훈련 및 자격증

□ 직업훈련은 3차년도를 제외한 1, 2, 4차-22차년도에 조사되었으며 조사년도별로 내용의 구성상 많은 차이가 있다. 공통된 내용은 직업훈련여부, 직업훈련 시기 및 기간, 직업훈련 종류, 비용부담, 희망하는 직업 훈련 등이다.

□ 조사년도별 직업훈련 관련 사항의 차이점은 다음과 같다.

- 1, 2차년도에는 가장 최근에 받은 직업훈련 한 가지에 관하여 조사하였다.
- 1차년도에는 직업훈련의 개수는 조사되지 않았다.
- 4-9차년도에는 가장 최근에 받은 직업훈련부터 총 세 가지 직업훈련에 관하여 조사하였다.
- 응답 기간을 제한하는데 있어서 1차년도에는 기간의 제한이 없었으나 2차년도에는 지난 조사 이후 받은 직업훈련에 한하여 응답받았으며, 4차년도에는 3차년도에 조사가 누락되었던 관계로 1999년 1월 이후에 받은 직업훈련에 관하여 응답하도록 하였다. 그러나, 5차년도부터는 다시 '지난 조사 이후'에 받은 직업훈련에 관하여 조사하였다.

- 직업훈련의 비용부담관련해서 1, 2차년도에는 비용부담의 주체가 누구인지에 대해서만 조사하였으나, 4, 5차년도에는 비용부담주체 뿐만 아니라, 본인이 교육훈련을 받는데 직접 지불한 비용까지 조사하였다.

- 직업훈련 관련 변수명은 p\*\*4501-p\*\*4644이다.

□ 17차년도 조사부터 직업교육 및 직업훈련의 문항 수정되었다. 16차년도까지는 받은 교육훈련이 2개 이상일 경우 가장 최근의 것부터 3개까지 응답받았으나, 17차년도 조사부터는 가장 중요한 한 가지만 응답받았다.

□ 자격증과 관련된 조사 내용은 다음과 같다.

- 자격증 관련 내용은 자격증보유 여부, 자격증 보유개수, 자격증 유형, 자격증 종류, 취득시기에 대하여 조사되었다.

- 2차와 4, 5차년도에 조사되었으며, 2차년도에는 최대 9개까지, 4, 5차년도에는 최대 6개까지 자격증의 종류와 급수, 취득시기에 관하여 조사하였다.

- 2차년도에는 기간의 제한 없이 개인이 소지한 모든 자격증에 관하여 응답하게 하였고, 4차년도에는 1999년 1월 이후 취득한 자격증에 한하여 응답하게 하였으며, 5차년도에는 ‘지난조사 이후’ 취득한 자격증에 대해 응답하게 하였다.

- 5차년도 조사부터는 소지한 자격증의 유형이 국가 자격증인지, 민간 또는 외국 자격증인지 여부를 구분할 수 있도록 항목을 추가하였다.

- 7차년도부터 자격증 유형에 대한 정보가 추가되었다. 이는 기존에 있던 자격증의 급수 정보를 자격증 제도가 운영되는 현실에 맞춰 대체한 것이다.

- 자격증 관련 변수명은 p\*\*4701-p\*\*4796이다.

- 17차년도에 추가된 자격증 취득목적 변수명은 p\*\*4797이며, 19차년도부터 추가된 변수명은 p\*\*4717(자격증 취득목적 1), p\*\*4727(자격증 취득목적 2), p\*\*4737(자격증 취득목적 3), p\*\*4747(자격증 취득목적 4), p\*\*4757(자격증 취득목적5)이다.

- 17차년도 이후 직업훈련 외에 받고 싶은 교육과 직업훈련이외 받고 싶은 교육 문항은 삭제하였다.

## 2.2.7 사회보험

□ 사회보험 관련 사항은 설문지상 가구용과 개인공통, 유형 설문지에 존재하며 자료 상에는 가구자료와 개인자료에 있다.

- 가구용 설문지에서는 한 가구 당 최대 3명의 가구원에 한하여 사회보험수급자 유무, 사회보험 수급자 이름과 가구원 번호, 연간 총수급액을 묻고 있다.
- 개인공통 설문지에서는 사회보험 수급자를 대상으로 사회보험 종류, 수급기간, 수급방식, 1회 수급액을 묻고 있다.
- 유형설문지에서는 직장을 통한 사회보험 가입여부를 조사하고 있다. 따라서, 관련 문항은 개인자료의 '주된 일자리'에 아래 표와 같은 변수형태로 입력되어 있다. 또한, 개인이 1개 이상의 임금 일자리를 갖고 있을 경우, 모든 임금 일자리에 대해 사회보험 가입여부를 조사하므로 동일한 내용이 직업력 자료에도 입력되어 있다.

□ <표 V-18>은 1-22차 사회보험 관련 변수이다. 사회보험 관련 사항은 1차년도에는 가구용 자료에만 있으며 2-22차에는 가구용과 개인용 자료에 모두 존재한다.

<표 V-18> 1-22차년도 사회보험 관련 변수

	변수명	차수
가구자료	h**2131-h**2142	2차-22차
주된 일자리	p**2101-p**2105	2차-22차
취업자 개인	p**2141-p**2190	3차-22차
직업력 자료	j401-j405	2차-22차

## 2.2.8 직무만족도(Job Satisfaction)의 측정

□ 한국노동패널의 직무만족도 측정과 관련된 자세한 논의사항은 “이상호, 노동패널의 직무만족도 측정, 노동리뷰 2006년 5월호, 한국노동연구원”(www.kli.re.kr/klips 테크니컬 리포트)을 참고하기 바란다.

# ① 요인별 직무만족도(facet measure)

□ 요인별 직무만족도의 측정은 1차년도 조사부터 현재까지 진행되고 있는 직무만족도 측정문항이다.

- 설문지의 개발은 미네소타대학교의 직업심리연구소(Vocational Psychology Research) 미네소타만족지수(Minnesota Satisfaction Questionnaire, MSQ)와 Smith, Kendall & Hulin(1969)등이 개발한 직무기술지표(Job Description Index: JDI) 등에 근거하여 이중 가장 사용되고 있는 항목들을 한국의 실정에 맞게 일부 수정하여 축약형(reduced form)으로 작성한 것이다.

<표 V-19> 1-22차년도 요인별 직무만족도(facet measure) 변수

응답항목	변수명	1차-22차
(1) 임금 또는 소득	p**4311	○
(2) 취업의 안정성	p**4312	
(3) 하고 있는 일의 내용	p**4313	
(4) 근로환경	p**4314	
(5) 근로시간	p**4315	
(6) 개인의 발전가능성	p**4316	
(7) 의사소통 및 인간관계	p**4317	
(8) 인사고과의 공정성(비임금제외)	p**4318	
(9) 복지후생(비임금제외)	p**4319	

- 이 중에서 ①-⑦까지의 항목은 취업자 전체를 대상으로 조사되었으며, ⑧, ⑨번 항목은 임금근로자에 대해서만 조사되었다.
- 단, 1차년도의 경우 임금근로자에 대해서 ‘근로시간’을 제외한 8가지 항목에 대해 조사하였으며, 비임금근로자에 대해서는 ‘의사소통 및 인간관계’를 제외한 6개 항목에 대해 조사하였다.

## ② 전반적 직무만족도(global measure)

- ☐ 전반적 직무만족도에 대한 측정은 두 가지 형태로 이루어지고 있다. 첫 번째는 단일항목으로 이루어진 측정이며, 두 번째는 5개의 항목들로 구성된 전반적 직무만족도의 측정이다.
- ☐ 우선 단일항목으로 측정되는 ‘전반적 직무만족도’ 설문은 3차년도부터 조사되어 왔다. 측정항목은 ‘(1) 매우 만족스럽다’에서 ‘(5) 매우 불만족스럽다’까지 5점 척도로 구성되어 있다.

<표 V-20> 단일항목으로 구성된 전반적 직무만족도 변수

설문 항목	변수명	3차-22차
_____님께서는 현재 주로 하는 일자리에 대해 전반적으로 얼마나 만족하고 계십니까.	p**4321	○

- ☐ 5개의 항목으로 구성된 ‘전반적 직무만족도’ 설문은 5차년도부터 조사되고 있다. 문항의 개발은 Brayfield & Rothe (1951)의 척도에서 5개 문항을 선별한 것으로, 질문에 대해 ‘(1) 전혀 그렇지 않다’부터 ‘(5) 아주 그렇다’까지 역시 5점 척도로 구성되어 있다.

<표 V-21> 5개 항목으로 구성된 전반적 직무만족도 변수

응답 항목	변수명	5차-22차
① 나는 현재 직장에서 하고 있는 일에 만족하고 있다.	p**4301	○
② 나는 현재 직장에서 하고 있는 일을 열정적으로 하고 있다.	p**4302	
③ 나는 현재 직장에서 하고 있는 일을 즐겁게 하고 있다.	p**4303	
④ 나는 현재 직장에서 하고 있는 일을 보람을 느끼면서 한다.	p**4304	
⑤ 별다른 일이 없는 한 현재 하고 있는 일을 계속하고 싶다.	p**4305	



## 2.2.9 조직몰입도(organizational commitment)의 측정

- 조직몰입도에 대한 측정 역시 5차년도 조사부터 이루어지기 시작했다. 설문은 개발은 Porter et al. (1974)의 OCQ(Organizational Commitment Questionnaires)에 기초하여 우수한 문항들을 선별한 것이다. 응답항목은 ‘(1) 매우만족’부터 (5) 매우 불만족’까지 5점 척도로 구성되어 있다.

<표 V-22> 조직몰입도의 측정 변수

응답 항목	변수명	5차-22차
① 지금 근무하고 있는 직장은 다닐만한 좋은 직장이다.	p**4201	○
② 나는 이 직장에 들어 온 것을 기쁘게 생각한다.	p**4202	
③ 직장을 찾고 있는 친구가 있으면 나는 이 직장을 추천하고 싶다.	p**4203	
④ 나는 내가 다니고 있는 직장을 다른 사람들에게 자랑할 수 있다.	p**4204	
⑤ 별다른 일이 없는 한 이 직장을 계속 다니고 싶다.	p**4205	

## 2.2.10 기타 변수

- 앞에서 언급된 변수 외에 매년 특별한 주제(건강상태, 생활만족도, 장애 및 장애, 영어학습 등)에 관한 여러 문항이 조사되었다.
- <표 V-23>은 1-22차년도 기타 변수들이다. 이 변수들의 년도별 차이점과 유의사항은 다음과 같다.
- 생활만족도는 1차-22차년도에 모두 조사되었다.
  - 8차년도부터 건강유지를 위해 평소 하는 일, 운동, 흡연 및 음주습관에 대한 정보가 추가되었다.
  - 장애관련 문항은 4차년도(2001년) 건강관련 부가조사에서 조사하였다. 이후 관련 문항들을 조사하지 않았으나 9차년도에 장애 관련 문항을 추가하였다.
  - 8차년도 부가조사인 ‘노동조합과 노사관계’ 설문 중 노동조합에 대한 견해에 관한 3개 문항을 9차년도 조사에 추가하였다.
  - 10차년도에는 영어학습 및 영어능력 문항을 추가하였다.

- 11차년도에는 취업자(비임금근로자)를 대상으로 국민연금 보험료 납부 여부 및 국민연금 가입형태 문항을 추가하였으며, 고용주/자영업자를 대상으로 현재 사업체의 산재보험 가입여부 문항도 추가하였다. 또한 임금근로자에게는 특수고용 직업 선택 문항도 추가로 조사되었다. 모든 응답자들에게는 금연일자 및 금주일자에 대한 문항도 추가되었다.
- 16차년도에는 취업자 설문에서 비임금근로자에게만 응답받을 수 있도록 ‘고용보험 가입 여부’ 문항이 추가되었다.
- 20차년도에는 삶의 만족도와 관련된 3개 문항을 취업자 설문, 미취업자 설문, 신규조사자 설문에 추가되었다.

<표 V-23> 1-22차년도 기타 변수

	변수명	차수
장애 또는 장애여부	p**6201-p**6262	9차
생활만족도	p**6501-p**6506	1차-22차
차별	p**6301-p**6376	7차
위험인식	p**6401-p**6405	7차, 9차
건강유지와 운동, 흡연, 음주습관	p**6151-p**6163	8차-22차
노동조합에 대한 견해	p**6602-p**6604	9차-22차
영어학습 및 영어능력	p**4901-p**4977	10차
작업장 환경	p**4461-p**4465	10차
고용보험 가입여부	p**2109	16차-22차
국민연금 보험료 가입여부 및 가입형태	p**2107-p**2108	11차-22차
사업체의 산재보험 가입여부	p**2106	11차-22차
금연일자, 금주일자	p**6166-p**6167 p**6164-p**6165	11차-22차
특수고용 직업 선택	p**0616	11차-22차
노사협의회	pa**3601, pa**3621, pa**3624	16차-22차
삶의 만족도	pa**8140, pa**8141, pa**8143	20차-22차

## 2.2.11 1차응답자 및 신규조사자 자료

□ 1차년도 응답자 및 신규조사자 자료는 조사시점 당시 처음으로 개인용 조사에 진 입한 응답자에게 해당되는 설문임. 따라서 1차년도에는 ‘모든 개인을 대상으로 아 래 문항이 조사되었으며, 2차년도 이후부터는 신규조사자만 응답함. 여기서 신규 조사자란 아래의 경우가 해당된다.

- ① 만15세가 된 원표본 및 비원표본 가구원
- ② 이전 조사에 누락되었던 원표본가구원(장기 해외 출장, 유학, 군입대 등)
- ③ 혼인 등으로 인해서 새롭게 조사에 포함된 비원표본가구원

<표 V-24> 1-22차년도 신규조사자 관련 변수

	변수명	1차-22차
출생지 및 14세무렵 성장지	p**9001-p**9006	○
형제자매 수 및 학력	p**9021-p**9027	○
종교, 종교참여도	p**9031-p**9032	12차 이후는 모든 응답자 응답
군복무 관련	p**9041-p**9045	○
부모님의 학력과 직업	p**9051-p**9054	○
부모-14세무렵 생활상태	p**9055-p**9059	○
출산력	p**9071-p**9075	○

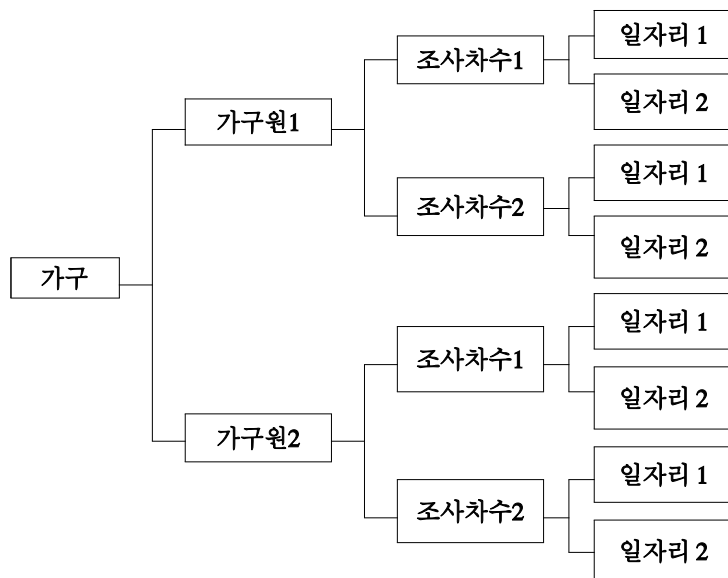
## 3. 직업력 자료

### 3.1 직업력 자료의 구조

□ KLIPS의 직업력(work history) 자료는 1차 조사 시점에서 과거에 가졌던 직업을 기 술한 회고적(retrospective) 정보와 1차 조사 이후로 조사된 개인의 모든 일자리에 관련된 정보가 담겨 있는 자료이다. 따라서 직업력 자료를 통해 개인이 최초로 노 동시장에 진입한 이래 가졌던 모든 일자리의 역사를 종합적으로 볼 수 있다.

- 가구용 자료가 한 가구를 기본 단위로 하고 있고 개인용 자료가 한 개인을 기본 단위로 하고 있다면 직업력 자료는 개인의 직업을 자료 구성의 기본 단위로 한다. 직업력 자료는 개인 조사의 유형설문지를 토대로 구성되며 개인 자료의 주된 일자리 정보 또한 직업력 자료에 포함된다.
- 직업력 자료에 포함되는 모든 자료는 조사된 개인의 모든 직업을 포함한다. 즉 원 가구원은 물론 비원가구원의 모든 일자리 정보가 수록되어 있다.
- 직업력 자료의 변수는 key 변수, 구성변수, 그 외의 주제별 변수로 구성되며 그 내용은 다음과 같다.
  - key 변수 : pid, hhid01-hhid22, hmem01-hmem22
  - 구성변수 : jobwave, jobseq, jobnum, jubnumc, jobcens, jobtype, jobclass, mainjob
  - 그 외의 주제별 변수 : key 변수와 구성변수를 제외한 모든 변수
- 직업력 자료의 구조를 [그림 V-1]을 통해 살펴보자. 한 가구에 속한 가구원1과 가구원2는 각각 조사차수 1, 2에 두 개씩의 일자리를 가지므로 직업력 자료에 이들의 일자리 정보는 8개의 행을 차지하게 된다.

[그림 V-1] 직업력 자료의 구조



### 3.2 직업력 자료의 내용

□ 직업력 자료에서 key가 되는 주요 변수는 다음과 같다.

- jobwave

- ‘일자리 발견 차수’를 의미하며 해당 직업이 조사된 차수를 나타낸다.
- 1부터 22까지는 각 일자리가 조사된 차수를 의미한다.

- jobseq

- ‘일자리가 조사된 순서’를 의미하며, 보통 일자리를 가졌던 순서와 동일하다.
- 동일한 일자리는 항상 동일한 jobseq를 갖는다.

- jobnum

- ‘일자리 번호’를 의미한다.
- 회고적 일자리는 처음으로 패널 조사에 응하기 전에 가졌던 일자리로, 순서대로 한자리의 값이 부여된다.
- 처음으로 패널 조사에 응하기 바로 직전에 그만둔 일자리에 대해서는 ‘100’ 값이 부여된다. 다만, 12차년도 조사에 추가된 표본의 경우에는 ‘1200’ 값이 부여되었다.
- 회고적 일자리를 제외한 일자리는 세자리의 값을 가진다. 백단위는 조사 차수를 의미하며 십·일단위는 조사 시점에서 일자리가 발견된 순서대로의 일련번호이다.

- jubnumc

- ‘지난 조사 당시의 일자리 번호(jobnum)’를 의미한다.

- jobcens

- 이전 조사와 현 조사 시점의 일자리의 보유/미보유 여부를 의미한다.
- ‘0’은 회고적 일자리에 관한 자료이다.
- ‘1’은 이전 조사 시 보유하고, 현재도 보유된 일자리를 의미한다.
- ‘2’는 이전 조사 시 보유하고, 현재 미보유된 일자리를 의미한다.
- ‘3’은 이전 조사 시 미보유했고, 현재 보유된 일자리를 의미한다.
- ‘4’는 이전 조사 시 미보유했고, 현재 미보유된 일자리를 의미한다.

- jobtype

- ‘일자리의 형태(임금/비임금)’를 의미한다.
- ‘1’은 임금근로자, ‘2’는 비임금근로자, ‘3’은 모름/무응답 혹은 구분이 불가능한 경우

이다.

- jobclass

- ‘일자리 유형’을 의미한다.
- 해당 일자리가 유형 1 - 유형 8 중 어디에 해당하는가를 의미한다. 유형 설문지는 2차년도부터 조사되었으므로 1차년도에 조사된 일자리와 신규조사자에게서 추출된 일자리는 이 변수가 적용되지 않는다.

- mainjob

- 해당 jobwave에서의 ‘주된 일자리 여부’를 의미한다.
- ‘1’은 주된 일자리를 의미하며, ‘0’은 기타 일자리를 의미한다.

□ 앞에서 살펴본 주요 변수 외에 직업력 자료를 구성하는 변수들과 그 내용은 <표 V-25>와 같으며 변수 사용 시 유의사항은 다음과 같다.

- 변수에 따라 임금근로자에게만 해당되거나 혹은 비임금근로자에게만 해당되는 변수가 있는 반면 모든 취업자에게 해당되는 변수도 있다.
- ‘업종(한국표준산업분류 8차(KSIC8))’(j640)와 ‘직종(한국표준직업분류 5차(KSCO5))’(j650)은 1차부터 제공한다. 21차 학술대회용 버전부터 변수명을 변경하였다.<sup>16)</sup>
- ‘업종(한국표준산업분류 9차(KSIC9))’(j641)와 ‘직종(한국표준직업분류 6차(KSCO6))’(j651)은 12차부터 조사되었다. 21차 학술대회용버전부터 변수명을 변경하였다.<sup>17)</sup>
- ‘업종(한국표준산업분류 10차(KSIC10))’(j642)와 ‘직종(한국표준직업분류 7차(KSCO7))’(j652)은 12차 자료부터 소급적용하여 제공한다.

16) 21차 이전 노동패널자료를 사용했던 연구자는 직산업 변수명이 변경되었음에 주의를 요함

구분		20차 이전 공개용 자료		21차 이후 공개용 자료	
		개인자료	직업력자료	개인자료	직업력자료
업종	한국표준산업분류 8차(KSIC8)	p**0330	j630	p**0340	j640
	한국표준산업분류 9차(KSIC9)(12차부터 제공)	p**0331	j631	p**0341	j641
	한국표준산업분류 10차(KSIC10)(12차부터 제공)	-	-	p**0342	j642
직종	한국표준직업분류 5차(KSCO5)	p**0332	j632	p**0350	j650
	한국표준직업분류 6차(KSCO6)(12차부터 제공)	p**0333	j633	p**0351	j651
	한국표준직업분류 7차(KSCO7)(12차부터 제공)	-	-	p**0352	j652

17) 21차 이전 노동패널자료를 사용했던 연구자는 직산업 변수명이 변경되었음에 주의를 요함

- ‘초과근로시간 단위구분’(j804)은 2차부터 조사되었다.
- ‘월평균임금의 세금액’(j730)과 ‘취업당시 월평균임금의 세금액’(j731)은 7차년도부터 추가되었다.
- ‘비임금 근로자의 월평균 적자액’(j802) 및 ‘시작 당시 월평균 적자액’(j803)은 7차년도부터 추가되었다.
- 사회보험 세부 항목 중 ‘특수직역연금’(j402)과 산재보험(j405)은 3차년도부터 조사되고, ‘(비임금) 산재보험 가입여부’(j1103)은 11차부터 조사되었다.
- ‘일자리의 주된 장소’(j1101)와 ‘조직형태’(j1102)은 11차년도부터 조사되었다.
- ‘노조가입경로’(j510)와 ‘노조비가입 이유’(j511)는 3차년도부터 조사되었고, ‘노조탈퇴의향’(j901)’은 9차년도부터 조사되었다.
- 실업급여 신청여부, 액수, 수령기간 등(j611-j616)은 5차년도부터 추가되었다.
- 교대제 유무, 형태, 근로시간 변경 주기(j621-j623)는 5차년도부터 추가되었다.
- 월평균 임금 및 처음월평균임금의 세금액(j730, j731)은 7차년도부터 추가되었다.
- 실업급여 관련 설문(j611-j616)이 비임금근로자에게도 17차년도부터 추가되었다.
- 정부 일자리 대책 관련 설문(j1301-j1302)이 17차년도부터 추가되었다.

<표 V-25> 직업력 자료의 변수

내 용	변수
취업시기 및 퇴직시기(임금)	j001-j006
구직활동 관련(임금/비임금)	j007-j015
업종, 직종(임금/비임금) 및 직위(임금)	j640-j642, j650-j652, j136-j139
고용형태(임금)	j142-j145
종사상 지위(임금/비임금)	j147-j150
근로시간 형태(임금)	j152-j155
근로계약기간(임금)	j156-j159
근로시간 및 초과수당 관련(임금)	j201-j212, j804
임금(임금) 및 소득(비임금) 관련 사항	j301-j318, j322-j327
(취업당시/시작당시)월평균임금의 세금액	j730-j731
비임금 근로자의 (처음)월평균 적자액	j802-j803
사회보험관련	j401-j405, j1103
정부 일자리 대책 관련	j1301-j1302
기업 형태 및 규모(임금)	j501-j504
일자리 주된 장소 및 조직 형태(비임금)	j1101-j1102
가족종사자의 수(비임금)	j505
노동조합 관련(임금)	j507-j511, j901
퇴직 및 퇴직금 관련(임금)	j601-j608
실업급여 관련(임금)	j611-j616
교대제 관련(임금)	j621-j623

◎User notes : 취업시기 및 퇴직시기가 서로 논리적으로 맞지 않는 경우(예를 들어 취업시기는 1999년 5월 3일인데 퇴직시기는 1999년 3월 1일로 퇴직시기가 취업시기보다 앞서는 경우)가 존재한다. 이는 실사과정에서 발생한 응답자 혹은 면접원 오류로 확인된다.

□ 이상에 언급된 내용은 토대로 다음의 <예시>를 통해 직업력 자료의 구조를 보다 자세히 살펴보겠다.

- <예시>의 11개의 행이 동일한 pid를 가지는 것으로 보아 동일인의 자료임을 알 수 있다.

- jobseq(일자리 번호)를 보면 3까지의 값을 가지므로 이 사람은 조사가 진행되는 동안 3개의 일직리를 가졌음을 알 수 있다. 3번을 가지는 일직리들은 모두 동일한 일직



리들이 다음 차수에도 이어지는 것을 의미한다.

- jobwave(일자리 발견차수)를 보면 첫 번째, 두 번째, 세 번째 일자리는 '1'의 값을 갖는다. 따라서 이 일자리는 1차년도에 조사된 일자리임을 알 수 있다. 이후에도 2, 3, 4, 5, 6, 7, 8, 9까지는 각각 2차년도부터 9차년도까지 조사된 일자리임을 알 수 있다.
- jobnum(일자리 번호) 중 '1'과 '100'의 값을 갖는 두 일자리는 1차년도에 조사된 회고적 일자리이며, 특히 '100'은 1차년도 당시 가장 최근에 그만두었다고 응답한 일자리임을 알 수 있다. '101'은 1차년도에 발견된 첫 번째 일자리, '201'은 2차년도에 발견된 첫 번째 일자리이다. 이런식으로 '301', '401', '501'까지 이 개인은 매년 한 개씩의 일자리를 가지고 있었던 것으로 나타난다.
- jobnumc(지난 조사당시의 일자리번호)를 통해서 이 개인이 1차년도부터 9차년도까지 매년 동일한 일자리를 계속 지속하고 있음을 알 수 있다. jobnum '201'옆의 칼럼이 '101'이므로 이것은 현재 일자리 번호가 '201'번인 일자리가 지난조사 당시에는 '101'번이었던 일자리였음을 알 수 있다. 이는 jobseq가 1차년도부터 9차년도까지 동일한 '3'값을 갖는 것으로도 알 수 있다.
- jobcens(일자리 지속 여부)는 위의 두 개의 일자리는 '0'의 값을 가지며, 세 번째 행에서는 '3'값을 가지며 이후부터는 '1'값을 가지고 있다. 따라서 위의 두 개의 일자리는 회고적 자료임을 알 수 있고, 세 번째 일자리는 이전조사 시 미보유했고 1차년도 조사 당시 가진 일자리임을 의미한다. 2차년도에는 이전조사 시 보유하고 2차년도 조사당시에도 보유하고 있는 일자리이므로 '1'값을 가지며, 이는 3-8차년도까지 반복된다.
- jobtype(일자리 형태)는 모두 '2'의 값을 가진 것을 보아 세 일자리 모두 비임금형태의 일자리인 것을 알 수 있다.
- jobclass(일자리 유형)은 2차년도부터 '3'의 값을 갖는다. 따라서 이 일자리는 유형3, 즉 비임금근로 중 이전조사 시 보유하고 있었고 현재에도 보유하고 있는 일자리이다. 전년도까지의 값이 'missing'인 이유는 당연하게도 '유형설문지' 자체가 2차년도이전에는 존재하지 않았기 때문이다.
- mainjob(주된 일자리 여부)는 1차년도부터 시작된 jobseq '3'의 일자리가 모두 주된 일자리임을 알 수 있다.

〈예 시〉

pid	jobwave	jobseq	jobnum	jobnumc	jobcens	jobtype	jobclass	mainjob
13801	1	1	1	.	0	2	.	0
13801	1	2	100	.	0	2	.	0
13801	1	3	101	.	3	2	.	1
13801	2	3	201	101	1	2	3	1
13801	3	3	301	201	1	2	3	1
13801	4	3	401	301	1	2	3	1
13801	5	3	501	401	1	2	3	1
13801	6	3	601	501	1	2	3	1
13801	7	3	701	601	1	2	3	1
13801	8	3	801	701	1	2	3	1
13801	9	3	901	801	2	2	4	0

## 4. 부가조사 자료

### 4.1 3차년도 부가조사 - 청년층 자료

#### 4.1.1 청년층 자료의 구조

- 조사시기 : 3차년도(2000년) 조사
- 조사대상 : 만15세 이상 - 30세 미만의 청년층으로 1970년 5월 1일 이후부터 1985년 4월 30일까지 출생한 개인
- 조사목적 : 청년 실업 문제와 청년 유희인력 문제에 대한 심층적인 규명과 학교에서 직업 세계로의 이행에 관한 조사의 필요성이 제기되어 실시됨.
- 조사내용 및 자료의 구조 : 청년층 설문지의 조사 내용을 담고 있으며, 기본적으로 개인자료와 동일한 구조를 갖고 있다.

#### 4.1.2 청년층 자료의 내용

□ 청년층 자료의 내용은 <표 V-26>와 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 학력관련 변수는 고등학교, 전문대학, 대학교(4년제), 대학원(석사과정), 대학원(박사과정)으로 나누어서 각각의 유형과 계열/전공, 학교 이름, 소재지, 졸업/그만둔 시기와 재학여부에 관한 내용을 묻고 있다.
- 졸업 당시의 건강상태는 5점 척도로 측정하였다.
- 자격증은 재학 중 혹은 졸업하면서 취득한 자격증 여부와 자격증 종류를 묻고 있다.
- 재학 중 직업훈련은 재학 중 취업을 위해 받은 훈련, 교육, 실습에 관한 내용을 담고 있다.
- 재학 중 취업(아르바이트)은 일한 내용과 일한 기간에 관한 내용을 담고 있다.
- 구직활동과 취업 경험은 졸업/중퇴/휴학 이후 구직활동방법과 기간, 취업제의 경험 유무와 취업 제의 거절 이유, 2개월 이상 취업 경험에 관한 내용을 담고 있다.
- 첫 일자리는 최종적인 학업 종료 이후 첫 일자리의 업종, 직종, 직위, 종사상 지위, 일자리 규모 등의 내용을 담고 있다.

<표 V-26> 3차년도 청년층 부가조사 자료의 변수

내 용	3차
학력	a030101-a030173
자격증 및 훈련	a030201-a030274
구직활동	a030301-a030342
첫 일자리	a030401-a033344

◎User notes : 청년층 자료에서 학력변수(a030101-a030105)는 최종학력을 응답하는 것을 기준으로 하였고, 현재 재학 중인 경우에는 재학 중인 학력을, 졸업이나 중퇴 한 경우 졸업이나 중퇴한 최종 학력을 기재하도록 했다. 그러나 응답자에 따라서 최종 학력 이전의 모든 학업 스토리 혹은 부분적 학업 스토리를 기재한 경우가 있다. 따라서 응답자가 기재할 수 있는 최대치인 5개의 학력 변수가 생성되었다.

청년층 자료를 다른 개인 데이터와 연계하여 사용할 때는 pid를 key변수로 사용하면 된다.

## 4.2 4차년도 부가조사 - 건강과 은퇴 자료

### 4.2.1 건강과 은퇴 자료의 구조

- 조사시기 : 4차년도(2001) 조사
- 조사대상 : 건강 - 15세 이상 모든 개인 설문 조사 대상자  
은퇴·노후생활 - 만45세 이상 응답자
- 조사목적 : 국민의 생활수준이 향상됨에 따라 건강에 대한 관심이 고조되고, 인구 고령화에 따른 고령자의 은퇴 및 노후생활에 대한 실태 파악을 위해 조사하였다.
- 조사내용 및 자료의 구조 : 건강과 은퇴 설문지의 조사 내용을 담고 있으며, 기본적으로 개인자료와 동일한 구조를 갖고 있다.

### 4.2.2 건강과 은퇴 자료의 내용

□ 건강과 은퇴 자료의 내용은 <표 V-27>과 같다.

- 건강상태는 현재와 1년 전의 건강상태를 5점 척도로 측정하였고, 건강상의 이유로 근로에 제한을 받았던 경험을 묻고 있다.
- 장애/장해는 장애/장해 여부와 종류, 장애/장해 판정 여부와 급수, 장애와 관련된 근로의 제약 등에 관한 내용을 담고 있다.
- 지병은 지병의 유무와 종류, 지병 시작 시기 등을 묻고 있다.
- 입원 경험과 건강진단은 입원 경험 및 입원 이유와 정기적인 건강진단에 관한 사항을 묻고 있다.
- 의료기관 이용은 의료기관 이용경험과 이용 의료기관 형태 및 이용 의료 서비스 종류를 묻고 있다.
- 건강보험은 국민 건강 보험 가입 여부와 보험료 등에 관한 사항을 묻고 있다.
- 건강관련 사항은 흡연과 음주, 운동 관련 사항을 묻고 있다.
- 피부양과 부양은 피부양 여부와 비용 등에 관한 내용을 묻고 있다.
- 건강상의 문제와 근로는 건강상의 문제로 일하는데 제약이 있었는지의 여부와 출산휴가 사용 여부 등을 묻고 있다.

- 은퇴와 노후 생활은 은퇴 상태와 은퇴 계획, 은퇴 시기, 은퇴자의 생활비 조달 방법 등을 묻고 있다.

<표 V-27> 4차년도 건강과 은퇴 부가조사 자료의 변수

내 용	4차
건강과 생활습관	a041101-a041104
장애	a041201-a041264
질병	a041301-a041356
국민건강보험	a041401-a041428
건강 및 생활습관	a041111-a041142
건강상 문제로 인한 부양 및 피부양	a041501-a041524
근로	a041601-a041644
은퇴	a041701-a041772

### 4.3 6차년도 중고령자 부가조사

#### 4.3.1 중고령자 부가조사 자료의 구조

- 조사시기 : 6차년도(2003) 조사
- 조사대상 : 만 50세 이상(1953년 4월 30일 이전에 출생) 모든 개인
- 조사목적 : 고령화 사회로의 급속한 진전에 따라 고령자들의 노후 대책 및 근로실태 파악을 위해 조사설계되었다.
- 조사내용 및 자료의 구조 : 가구정보와 중고령자 부가조사 설문지의 조사 내용을 담고 있으며, 기본적으로 개인자료와 동일한 구조를 갖고 있다.

#### 4.3.2 중고령자 부가조사 자료의 내용

- ☐ 중고령자 부가조사 자료의 내용은 <표 V-28>과 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 은퇴 여부와 관련한 내용은 ‘은퇴자’를 정의하고 은퇴 연령 및 은퇴이유, 그리고 은퇴 계획을 조사하기 위해 설계되었다. 또한, 은퇴할 당시 배우자의 경제활동 상태에 대해서도 질문하였다.
- 생애에서 가장 중요한 일자리의 특성은 현재 하고 있는 일자리가 은퇴한 후에 다시 재취업한 일자리인지 아니면 여전히 주된 일자리인지를 판별한다. 그 다음으로는 생애에서 가장 주된 일자리의 산업, 직업, 재직기간, 종사상지위, 근로시간 형태, 사업체의 규모, 근로시간, 임금 등을 조사하였다.
- 개인의 생애주기 전체에서 볼 때 일반적으로 가장 왕성한 경제활동을 하는 시기를 45세 전후로 보고 이 당시의 취업활동을 조사하였다. 여기서는 종사상 지위와 산업, 직업에 대해 간략하게 조사하였다.
- 고령자들의 소득원에 대해서도 본인 및 배우자의 근로소득, 연금소득, 금융자산소득, 부동산소득, 사회보험수급, 공적이전, 사적이전 등에 대해 파악하였다. 이미 노동패널 조사에서는 가구자료에서 가구단위의 소득을 파악하고 있을 뿐만 아니라, 개인자료에서도 개인단위의 근로소득을 조사하고 있으므로 중고령자 자료의 소득자료까지 분석에 적용할 경우 고령자의 소득원천에 대해 보다 상세한 파악이 가능할 것으로 예상된다.
- 고령자의 경제생활은 가족내의 경제적 교류에 직접적으로 의존하므로 이를 파악하였다. 자녀의 수 및 가족내 어떤 구성원으로부터 도움을 받는지, 혹은 반대로 오히려 도움을 주고 있는지를 조사하였다.
- 부모님에 대한 부양을 누가 담당하는지, 혹은 했는지도 조사하였다. 친구 및 친지와의 교류 정도, 친목모임, 종교모임 등 노후 활동에 대한 조사도 실시하였다.
- 마지막으로 지난 2주 동안의 건강상태와 질병여부에 대해서도 조사하였다.

<표 V-28> 6차년도 중고령자 부가조사 자료의 변수

내 용	6차
은퇴	a061702-a061704
중고령자 주된 일자리	a061801-a061836
소득	a061901-a061952
경제적 교류 및 사회관계 등	a062001-a062084

## 4.4 7차년도 근로시간과 여가 부가조사

### 4.4.1 근로시간과 여가 부가조사 자료의 구조

- 조사시기 : 7차년도(2004) 조사
- 조사대상 : 모든 개인
- 조사목적 : 2004년 7월부터 실시되는 ‘법정근로시간단축’이 경제생활 및 여가 생활의 패턴에 어떤 영향을 미치는지를 연구하기위해 설계되었다.
- 조사내용 및 자료의 구조 : 근로시간과 여가 설문지의 조사 내용을 담고 있으며, 기본적으로 개인자료와 동일한 구조를 갖고 있다.

### 4.4.2 근로시간과 여가 부가조사 자료의 내용

□ 근로시간과 여가 자료의 내용은 <표 V-29>과 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 임금근로자의 근로자의 근로시간 및 휴가 : 직장에서의 근무시간, 정기휴일, 휴무(법정공휴일, 명절, 기타 휴가)에 대해 직장에서의 제공여부와 실제로 응답자가 원날, 연월차 휴가 및 수당 여부, 하계휴가 및 생리휴가 사용 여부 및 회수
- 비임금근로자의 근로시간 및 휴가 : 근무시간의 규칙성 여부, 평일 근무 중 시간 배분, 주5일제의 효과, 휴일(법정공휴일 및 명절, 기타 휴무, 하계휴가)관련 사항
- 근로시간 단축에 관한 의견(임금/비임금) : 근로소득, 휴가일수, 실근로시간, 취업 기회 등의 증감 여부, 근무시간과 노동강도 증감 여부, 2002년 1월 이후 정규근로시간 단축 여부
- 여성의 출산휴가 : 출산 경험여부, 출산 당시 취업형태, 출산휴가 여부
- 일상의 시간배분 : 평소의 일상적인 활동에 대한 시간배분을 평일, 토요일, 일요일로 나누어 조사함.
- 여가관련 사항 : 지난 주 여가활동여부/소요시간/경비 등, 여가 활동의 순위, 지난 1년간의 독서량, 신문 구독량, TV시청시간, 종교활동 참여수준

<표 V-29> 7차년도 근로시간과 여가 부가조사 자료의 변수

내 용	7차
임금 근로자의 근로시간 및 휴가	a072501-a072658
비임금 근로자의 근로시간 및 휴가	a072701-a072851
근로시간 단축 및 출산휴가	a072901-a072936
시간배분	a073001-a073051
여가활동	a073101-a073182

#### 4.5 8차년도 노동조합과 노사관계 부가조사

##### 4.5.1 노동조합과 노사관계 부가조사 자료의 구조

- 조사시기 : 8차년도(2005) 조사
- 조사대상 : 모든 임금근로자 및 비임금근로자 일부
- 조사목적 : 노동조합 및 근로자대표제, 다양한 형태의 노사의사소통체계 등의 실태와 의식을 사업체 단위가 아닌 근로자를 통해 조사함으로써 노동시장 및 노사관계 연구의 발전에 기여한다.
- 조사내용 및 자료의 구조 : 노동조합과 노사관계 설문지의 조사 내용을 담고 있으며, 기본적으로 개인자료와 동일한 구조를 갖고 있다.

##### 4.5.2 노동조합과 노사관계 부가조사 자료의 내용

□ 노동조합과 노사관계 자료의 내용은 <표 V-30>와 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 노동조합 가입과 참여 : 직장의 노조 존재여부, 노동조합 가입여부, 노조 가입과 탈퇴 의사 여부, 노동조합 활동에 대한 참여수준, 노동조합과 사용자간의 협력수준, 단체협상의 적용 수준, 노조 비가입 이유, 이전 일자리에서의 노조가입 경험, 자발적 이직 여부
- 노사협의회 운영 : 노사협의회 운영여부, 노사협의회 운영에 대한 관심수준 및



의식, 근로자 대표 선출 방식, 노동조합과 노사협의회에 대한 의식

- 인적자원관리와 근로자 대표제도 : 노사협의회와 노동조합을 제외한 공식적·비 공식적 의사소통제도의 존재여부, 운영실태에 대한 의견, 직무상의 의사결정 참여와 영향력 수준, 임금, 근로시간, 노동강도 등에 대한 불만 및 의사소통경로, 고용주와의 신뢰관계 수준
- 우리나라 노동조합에 대한 견해 : 노동조합의 근로조건향상에 기여하는지에 대한 견해, 노동조합의 사회적 영향력에 대한 견해, 노동조합의 경제적 역할에 대한 견해, 노동조합에 대한 부정적 견해들에 대한 응답자의 판단
- 응답자의 정치·경제·사회적 가치관 : 우리나라 고용주에 대한 응답자의 태도, 우리나라의 사회적 형평성과 공정성에 대한 견해, 응답자 개인의 사회적 가치관

<표 V-30> 8차년도 노동조합과 노사관계 부가조사 자료의 변수

내 용	8차
노동조합	a083501-a083554
노사협의회	a083601-a083624
인적자원관리 및 근로자대표제도	a083701-a083824
노동조합에 대한 태도 및 정치·경제·사회 가치관	a083901-a083964

## 4.6 9차년도 청년층 부가조사

### 4.6.1 청년층 부가조사 자료의 구조

- 조사시기 : 9차년도(2006) 조사
- 조사대상 : 만 15세 이상 35세 이하 청년층(1971-1991년 출생자)
- 조사목적 : 청년층의 성장당시 가족 및 학교생활과 현재의 구직활동, 취업, 임금 등과 같은 노동시장 행태와의 연관성을 파악하고자 하였다.
- 조사내용 : 취학전, 취학이후 초등학교부터 대학교까지의 사교육, 성적, 일자리 경험, 첫 일자리의 특성, 가족들과의 일상적 관계, 청년층 무업자(NEET: Not in Education, Employment or Training)의 행태 등.

#### 4.6.2 청년층 부가조사 자료의 내용

□ 청년층 자료 내용은 <표 V-31>과 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 유소년기부터 청소년기까지의 학교생활 및 교육 : 유치원, 초등학교, 중학교, 고등학교 당시의 사교육 경험, 사교육 종류, 어학연수 경험, 성적, 학업력(Educational history)
- 대학교 및 대학원 진학 : 대학교 입학여부, 전공, 편입여부, 학점, 졸업여부, 대학원 입학여부, 전공, 진학 이유 등
- 졸업 후 첫 일자리 경험 : 졸업 후 구직활동, 졸업후 첫 일자리 종사상지위, 근무형태, 고용형태, 기업형태, 소득, 근로시간, 전공 및 교육수준 일치도, 그만둔 시기 및 이유
- 성장기 가족관계 및 생활환경 : 성장기 당시 부모와의 동거여부, 모친의 직업유무, 14세 당시의 집안형편, 문화활동 참여, 가족관계
- 무직자의 행태 및 청소년기의 일탈행위 : 무직자의 행태, 중학교 및 고등학교 시절의 휴학, 지각, 결석, 폭행, 술, 담배 등의 일탈행위

<표 V-31> 9차년도 청년층 부가조사 자료의 변수

내 용	9차
유소년기부터 청소년기까지의 학교생활 및 교육	a094101-a094354
대학교 진학	a094401-a094438
대학원 진학	a094439-a094473
졸업 후 첫 일자리 경험	a090311-a090506
성장기 가족관계 및 생활환경	a094501-a094566
청소년기의 일탈행위	a094601-a094680

## 4.7 10차년도 고용형태 부가조사

### 4.7.1 고용형태 부가조사 자료의 구조

- 조사시기 : 10차년도(2007) 조사
- 조사대상 : 임금근로자 4,327명과 고용주/자영자 432명
- 조사목적 : 고용형태의 특성을 파악하는 것이 목적이다.
- 조사내용 : 임시적 고용, 간접고용, 일일/호출근로, 특수고용, 시간제 근로 등

<표 V-32> 10차년도 고용형태 자료의 변수

내 용	10차
고용형태 판별	a105101-a105121
한시근로 및 임시적 계약형태	a105200-a105223
간접고용	a105300-a105369
일일호출근로	a105400-a105405
특수형태고용	a105500-a105535
시간제 근로	a105600-a105618
직장상황 및 적응	a105700-a105705

## 4.8 11차년도 교육 부가조사

### 4.8.1 교육 부가조사 자료의 구조

- 조사시기 : 11차년도(2008) 조사
- 조사대상 : 만 15세 이상 65세 응답자(만15세로 중학교 3학년 재학 중인 응답자는 조사대상에서 제외함)
- 조사목적 : 개인의 사교육 및 교육과 관련된 중학교부터 대학교까지의 회고적 정보, 본인 및 자녀의 현재 사교육에 대한 정보수집을 위함이다.
- 조사내용 : 중·고등·대학교 다녔는지 여부, 다니지 않았다면 그 이유, 검정고시 여부, 학교 생활 만족도, 학교 활동, 성적 성취도, 사교육, 공부시간 등과 자녀교육 관련 문항, 군복무 관련 등.

#### 4.8.2 교육 부가조사 자료의 내용

□ 교육 부가조사 자료 내용은 <표 V-33>과 같으며 각각의 내용에 대한 설명은 다음과 같다.

- 중학교 및 중학교 생활 : 중학교에 다녔는지 여부, 검정고시 여부, 입학년도, 졸업년도, 중학교 이름, 위치, 학교형태, 남녀공학 여부, 중학교 생활 만족도, 사교육 경험 여부, 공부 시간, 성적 성취도, 중퇴한 이유
- 고등학교 및 고등학교 생활 : 고등학교에 다녔는지 여부, 검정고시 여부, 입학년도, 졸업년도, 중학교 이름, 위치, 학교형태, 남녀공학 여부, 인문계/실업계 여부, 고등학교 생활 만족도, 자율학습 여부, 사교육 경험 여부, 공부 시간, 성적 성취도, 중퇴한 이유
- 고등학교 재학 생활 : 학습시간 및 수면시간, 컴퓨터 사용시간, 사교육 경험 여부 및 사교육 종류, 학교 성적
- 대학교 및 대학 생활 : 대학교에 다녔는지 여부, 대학과 전공 선택 여부, 성적, 대학교 생활 만족도, 등록금 충당 방법 및 장학금 수혜 여부, 학습시간 및 수면시간, 국가시험 준비 여부, 해외체류 경험 여부
- 대학생 생활: 학교생활 만족도, 등록금 충당 방법, 장학금 수혜 여부 등
- 자녀 교육 : 자녀들과의 대화 정도, 이민 및 해외유학 고려 여부, 자녀들의 학교 만족도, 자녀의 현재 성적, 자녀의 기대 성적
- 군복무 : 군복무 여부, 군복무 하지 않은 이유, 군복무 시기

<표 V-33> 11차년도 교육 부가조사 자료의 변수

내 용	11차
중학교 및 중학교 생활	a116101-a116178
고등학교 및 고등학교 생활	a116201-a116298
고등학교 재학생 관련	a116301-a116344
대학교 및 대학교 생활 관련	a116401-a116456
대학교 재학생(휴학생 포함) 관련	a116501-a116552
자녀교육 관련	a116601-a116661
군복무 관련	a116711-a116719

## 4.9 17차년도 시간사용과 삶의 질 부가조사

### 4.9.1 시간사용과 삶의 질 부가조사 자료의 구조

- 조사시기 : 17차년도(2014) 조사
- 조사대상 : 만 15세 이상 74세 이하 응답자
- 조사목적 : 본 부가조사는 ① 법정 근로시간 단축을 넘어선 실질 근로시간 단축, ② 근로시간 단축과 유연성에 대한 근로자의 결정권, ③ 여성을 대상으로 한 법정 모성보호 제도 활용 등을 파악하고 이러한 시간 사용 상태가 일·생활 균형 및 삶의 질과 어떤 연관성을 갖는지를 파악하고자 하는 목적을 가진다.
- 조사내용 : 근로시간 현황 및 단축, 일상적인 시간 배분 및 선호, 시간제 근로, 가사노동시간, 삶의 질, 여가시간, 휴일·휴가 사용 등

### 4.9.2 시간사용과 삶의 질 부가조사 자료의 내용

□ 시간사용과 삶의 질 부가조사 자료 내용은 <표 V-34>와 같으며, 각각의 내용에 대한 설명은 다음과 같다.

- 근로시간 현황 및 단축 : 실제 근무시간, 업무 내용에 따른 시간 사용, 근로시간 연장 및 단축 가능 여부, 업무량과 정규근로시간, 업무량과 전체 근로시간, 근로시간 단축 혹은 연장 희망, 희망 주당 평균근무일, 현행 출퇴근 시간, 희망 출퇴근 시간, 출퇴근 시간 변경 희망 이유, 출퇴근 시간 변경 허용 여부, 근무시간 중 개인 용무 가능여부
- 휴일·휴가 사용 : 연월차 휴가, 연월차 휴가 수당, 미사용 연월차 휴가 처리 방법, 휴일에 쉬지 않고 근무한 날 및 그 이유
- 시간제 근로 : 시간제 유형, 전일제로의 전환 의향, 전일제로 변경 여부, 시간제 근로 필요 여부, 시간제 근무 의향
- 미취업자의 경제활동 참가 유인 : 취업 의향, 희망 근무일수, 희망 출퇴근시간, 희망 월평균 소득
- 일상적인 시간 배분 : 24시간 활동 내용 및 시간
- 여가 시간 : 휴무일, TV 시청 시간, 인터넷 이용 시간, 여가 시간 활동 내용·횟

수·이유·동행자

- 성역할 및 삶의 질 : 가족과 함께 저녁식사 횟수, 집안일 하는 정도, 성역할 4점 척도<sup>18)</sup>19), 사용시간 수준, 일중독 5점 척도<sup>20)</sup>

<표 V-34> 17차년도 시간사용과 삶의 질 부가조사 자료의 변수

내 용	17차
근로시간 현황 및 단축	a177101-a177139
휴일·휴가 사용	a177201-a177213
시간제 근로	a177301-a177308 a177140-a177141 a177408-a177409
미취업자의 경제활동 참가 유인	a177401-a177407
일상적인 시간 배분	a177501-a177553
여가 시간	a177554-a177672
성역할 및 삶의 질	a177673-a177729

#### 4.10 18차년도 삶의 인식 부가조사

##### 4.10.1 삶의 인식 부가조사 자료의 구조

- 조사시기 : 18차년도(2015) 조사
- 조사대상 : 만 15세 이상 모든 응답자
- 조사목적 : 본 부가조사는 ① 최근 개인의 사회 경제적 성과(교육수준, 취업, 임금 등)의 중요한 요인으로 지목되는 비인지적 능력(개인의 성격 및 태도)을 파악, ② 인생 행복의 기준이 되는 ‘삶의 만족도’ 문항을 본 조사 문항에서 보다 세부적으로 파악 하고자 하는 목적을 가진다.

18) Anderson, Robert et al.(2009), Second European Quality of Life Survey : Overview. Office for Official Publications of the European Communities.

19) Brown M.J. and Glasdstone(2012), “Development of a Short Version of the Gender Role Beliefs Scale,” International Journal of Psychology and Behavioral Sciences 2(5), pp.54~158

20) Aziz et al.(2013), “The Workaholism Analysis Questionnaire: Emphasizing Work-Life Imbalance and Addiction in the Measurement of Workholism.”

- 조사내용 : 개인의 성격 태도(Big Five), 통제소재(Locus of control), 상호성(호혜성, Reciprocity), 신뢰, 위험 회피, 참을성, 충동, 삶의 만족도, 일에 대한 가치, 삶의 상태(사다리), 삶의 행복 정도, 5년 전 삶의 만족도, 5년 후 삶의 만족도

#### 4.10.2 삶의 인식 부가조사 자료의 내용

□ 삶의 인식 부가조사 자료 내용은 <표 V-35>, <표 V-36>과 같으며, 각각의 내용에 대한 설명은 다음과 같다.

##### ○ 비인지적 능력<sup>21)</sup>

- 개인의 성격 태도(Big Five) : 개인의 성격은 개방성, 성실성, 외향성, 친화성, 신경증으로 구분하며, 각각 요인별로 3개 문항씩 총 15개 문항, 7점 척도 사용
- 통제 소재 : 개인의 행동이나 사건이 자기 내부에서 기인한 것, 아니면 반대로 통제할 수 없는 바깥 영역에서 기인한 것이라고 인식하는지를 알아보는 것으로, 내적 통제(4개 문항)와 외적 통제(6개 문항)로 구분하며, 7점 척도 사용
- 상호성(호혜성) : 다른 사람의 행동에 대해 반응하는 경향을 측정하는 것으로, 긍정과 부정 각각 3문항으로 구성되며, 7점 척도 사용
- 신뢰, 위험 회피, 참을성, 충동 : 신뢰 3개 문항, 그 외 각각 1개 문항, 7점 척도 사용

##### ○ 행복<sup>22)</sup>

- 삶의 만족도, 일에 대한 가치, 삶의 상태(사다리), 행복정도, 5년전·후의 삶의 만족도 : 0~10점까지 11점 척도 사용하여 측정

21) David Richter, Maria Metzing, Michael Weinhardt, Jürgen Schupp. 2013. SOEP Scales Manual. SOEP Survey Papers 138: Series C. Berlin: DIW/SOEP  
[http://panel.gsoep.de/soep-docs/surveypapers/diw\\_ssp0138.pdf](http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0138.pdf)

22) OECD (2013), OECD Guidelines on Measuring Subjective Well-being, OECD Publishing.  
<http://dx.doi.org/10.1787/9789264191655-en>

<표 V-35> 18차년도 삶의 인식 부가조사 자료의 변수

내용		문항	18차 변수
성격 (Big 5)	개방성	[문1] 4, 9, 14	a188104, a188109, a188114
	성실성	[문1] 1, 7(역항목), 11	a188101, a188107, a188111
	외향성	[문1] 2, 8, 12(역항목)	a188102, a188108, a188112
	친화성	[문1] 3(역항목), 6, 13	a188103, a188106, a188113
	신경증	[문1] 5, 10, 15(역항목)	a188105, a188110, a188115
통제 소재	내적	[문2] 1, 4, 6, 7(역항목)	a188116, a188119, a188121-2
	외적	[문2] 2, 3, 5, 8, 9, 10	a188117-8, a188120, a188123-5
상호성(호혜성)	긍정	[문3] 1, 2, 3	a188126-a188128
	부정	[문3] 4, 5, 6	a188129-a188131
신뢰		[문4] 1(역항목), 2, 3	a188132-a188134
위험 회피		[문5]	a188135
참을성		[문6]	a188136
충동성		[문7]	a188137

<표 V-36> 18차년도 행복 관련 부가조사 자료의 변수

내 용	18차 변수
전반적인 삶의 만족도	a188138
일에 대한 가치	a188139
삶의 상태(사다리)	a188140
모든 것을 고려할 때, 행복 정도	a188141
5년전, 삶의 만족도	a188142
5년후, 삶의 만족도	a188143

## 4.11 18차년도 장년층 부가조사

### 4.11.1 장년층 부가조사 자료의 구조

- 조사시기 : 18차년도(2015) 조사
- 조사대상 : 만 50세 이상 만 74세 이하 모든 응답자
- 조사목적 : 본 부가조사는 장년층의 고용과 은퇴 후 경제적 상태, 사회관계 및 건강 등을 파악 하고자 하는 목적을 가진다.



- 조사내용 : 은퇴상태, 생애 주된 일자리 정보, 퇴직금 및 퇴직연금, 은퇴준비 및 은퇴생활, 창업 또는 재취업 계획, 은퇴 후 경제적 준비, 가족관계, 사회관계, 건강 등

#### 4.11.2 장년층 부가조사 자료의 내용

□ 장년층 부가조사 자료 내용은 <표 V-37>과 같으며, 각각의 내용에 대한 설명은 다음과 같다.

- 은퇴상태 : 은퇴여부, 은퇴시기, 은퇴 이유, 은퇴 시 배우자 여부, 은퇴 시 배우자 경제활동상태, 은퇴 계획 여부, 예상 은퇴 연령
- 생애 주된 일자리 : 일자리 여부, 일자리 정년 여부, 산업, 직업, 취업시기, 퇴직시기, 종사상 지위, 근로시간형태, 사업장 종업원수, 주당 근로시간, 주당 근로일수, 월평균 임금 또는 소득(세후), 퇴직이유, 퇴직 자발성 여부
- 퇴직금 및 퇴직연금 : 퇴직급여제도, 퇴직연금 형태, 퇴직급여(액수, 일시), 퇴직금 및 퇴직연금 사용 계획, 최소한의 노후 소득이나 수입, 적정한 노후 소득이나 수입
- 은퇴준비 및 은퇴 후 생활 : 퇴직 후 재취업 또는 창업 계획, 준비 분야, 기술·기능·지식 등의 수준, 재취업 또는 창업 프로그램, 재취업 또는 창업 준비 계획여부, 재취업 장애 요인, 향후 소일거리 여부, 소일거리 이유, 은퇴를 위한 경제적 준비 여부 및 준비 방법, 경제적 준비 부족 이유, 은퇴후 부족한 생활비 마련 방법, 은퇴후 최고 지출 항목, 은퇴 후 생활안정을 위한 정부의 지원정책, 은퇴 후 이사 계획 여부 및 이유
- 가족 관계 : 자녀 유무, 경제적으로 도움 받는 여부 및 액수, 경제적으로 도움 주는 여부 및 액수
- 사회관계 및 건강: 친구·이웃·친척과 교류 횟수, 교양과 취미활동 시간, 종교모임이나 친목모임, 동호회 참여 횟수, 특정한 단체 가입여부 및 유형, 지난 2주 동안 하루라도 아파서 쉬어야 했던 경험 및 기간, 40대 이후 3개월 이상 활동에 지장이 있었던 여부 및 시기, 원인

<표 V-37> 18차년도 장년층 부가조사 자료의 변수

내 용		18차
은퇴 상태		a188201 - a188210
생애 주된 일자리		a188301 - a188322
퇴직금 및 퇴직 연금		a188401 - a188417
은퇴준비 및 은퇴 후 생활		a188501 - a188532
가족관계		a188601 - a188641
사회관계 및 건강		a188701 - a188717
개인자료 결합	생애 주된 일자리가 현재 일자리인 경우	p180340-2, p180350-2, p180301-2 p180314-5, p180402-3, p180405-7, p181004-5, p181031-2, p181642, p181672
	혼인 상태	p185501

## 4.12 19차년도 청년층 부가조사

### 4.12.1 청년층 부가조사 자료의 구조

- 조사시기 : 19차년도(2016) 조사
- 조사대상 : 만 15세 이상 만 34세 이하 모든 응답자
- 조사목적 : 본 부가조사는 청년층의 교육과 취업, 청년 고용 정책 및 주관적 인식 등을 파악 하고자 하는 목적을 가진다.
- 조사내용 : 교육, 취업준비, 취업의사, 첫 번째 일자리, 현재 상태(NEET), 청년 고용정책, 주관적 인식 등

### 4.12.2 청년층 부가조사 자료의 내용

□ 청년층 부가조사 자료 내용은 <표 V-38>과 같으며, 각각의 내용에 대한 설명은 다음과 같다.

- 교육: 학업력, 전공, 대학원 진학동기
- 취업 준비: 시작 시기, 입사원서 회수, 면접 회수, 창업준비여부, 취업준비 내용

- 및 기간, 비용, 공무원 준비 및 합격여부, 국가전문자격증 준비 및 합격 여부
- 취업의사: 취업여부<sup>23)</sup>, 이직 의사 여부 및 사유, 중소기업 취업 의사, 해외취업 의사, 직업훈련 의사, 창업 의사, 시간제 일자리 의사, 희망 일자리, 희망 수입, 취업시 고려 사항, 취업준비 지속 기간
- 첫 번째 일자리: 첫 일자리 시작시기, 산업, 직업, 종사상지위, 근무형태, 비정규직 여부, 기업유형, 규모, 소득, 근로시간, 전공 및 교육수준 일치도, 희망직업과 일치 여부, 그만둔 시기 및 이유
- 현재상태: 현재상태(NEET) 지속 여부, 구직 여부, 비구직 사유, 사회적 관계, 수면시간
- 청년 고용정책: 관심도, 인지도, 취업에 도움정도, 청년층 취업난의 원인, 개선 방향
- 주관적 인식 및 기타 문항: 부모 동거 여부, 생활비 마련, 청년활력지수<sup>24)</sup>, 한국 사회에 대한 청년의 인식<sup>25)</sup>, 자존감<sup>26)</sup>, 사회적지지<sup>27)</sup>

---

23) 취업준비를 한 사람들에게 한해 취업여부를 질문하였기 때문에, 개인자료에 있는 취업여부와 상이하므로 사용시 유의할 것

24) 한겨레경제사회연구원(2015), 개원 심포지엄: 「이 땅에서 청년으로 산다는 것」

25) 한겨레경제사회연구원(2015), 개원 심포지엄: 「이 땅에서 청년으로 산다는 것」에서 발췌

26) 자존감 문항은 Rosenberg(1965)가 개발하고, 김문주(1988)가 번안하여 윤유경 (1996)이 사용한 척도를 5점척도로 수정하여 사용함(심리척도핸드북 I (1999) 176-177쪽 참조).

원출처: Rosenberg, M.(1965). Society and the adolescent self image. Princeton, N. J.: Princeton University press.

27) 사회적지지 문항은 Abbey, Abraims와 Caplan(1985)이 개발한 척도를 전지아(1990)가 번안하여, 유성은(1997)이 사용한 것을 사용함(심리척도핸드북Ⅱ(2000) 146-147쪽 참조).

원출처: Abbey, A., Abraims, D.J., & Caplan, R. D.(1985). Effects of different sources of social support and social conflict on emotional well-being. Basic & Applied Social Psychology, 6(2) 111-129

<표 V-38> 19차년도 청년층 부가조사 자료의 변수

내 용		19차	
교육		a199101 - a199154	
취업준비		a199201 - a199294	
취업의사		a199301 - a199325	
첫 번째 일자리		a199401 - a199426	
현재상태(NEET)		a199501 - a199518	
청년 고용 정책		a199601 - a199637	
기타	부모동거여부, 경제적 독립	a199701 - a199703	
주관적 인식	청년활력지수		a199704 - a199708
	한국사회에 대한 청년의 인식	공정성	a199709 - a199710
		연대·협동지수	a199711 - a199712
		세대연대지수	a199713 - a199714
		청년세대 자존감지수	a199715 - a199716
	자존감 [문 55] (역항목: 2,3,4,6,7,8,9,10,11)		a199717 - a199727
	사회적지지 [문 56] (역항목: 7,8,9,10,11)		a199728 - a199738

#### 4.13 20차년도 의사결정 부가조사<sup>28)</sup>

##### 4.13.1 의사결정 부가조사 자료의 구조

- 조사시기 : 20차년도(2017) 조사
- 조사대상 : 만 15세 이상 만 74세 이하 응답자<sup>29)</sup>
- 조사목적 : 본 부가조사는 개인의 전략적 사고, 합리성 및 사회성을 측정하여, 경제 사회적 성과와 결합하여 분석하고자 하는 목적을 가진다.
- 조사내용 : 후진 귀납법 능력, 전략적 사고, 마음을 이해하는 능력

28) 본 부가조사는 노동패널팀과 서울대학교 경제학부 실험행동사회과학센터 최승주 교수팀이 함께 개발함.

29) 예산상 제약으로 노동패널 응답자 15세 이상 74세 이하 가구원이 두명 이상 존재하는 가구 중 800가구를 지역별로 할당하여 조사함. 800가구 중 400가구는 면접원이 개인 조사를 마치고, 테블릿 PC를 제공하여 방문 당시 진행하였고, 400가구에 대하여는 이메일만 수집하여 패널 본조사를 마치고 추후 이메일로 조사 사이트를 안내하고 자율적으로 응답하도록 독려함.

#### 4.13.2 의사결정 부가조사 자료의 내용

□ 의사결정 부가조사 자료 내용은 <표 V-39>와 같으며, 각각의 내용에 대한 설명은 다음과 같다.

- 후진 귀납법 능력(Backward induction)<sup>30)</sup>

\* 방법: 주어진 목표숫자에서, 부분 게임 완전 내쉬 균형 (subgame perfect Nash equilibrium) 전략이 무엇인지를 파악하는 능력을 측정. 응답자가 한번의 실수도 없이 게임을 진행하면 승리하고, 그러한 경우에 1로 코딩 되어있고, 패배하는 경우에 0으로 코딩 되어 있음. 목표숫자는 5, 11, 14, 17로 총 4번의 게임이 진행됨. 즉 4번의 게임 결과의 합이 높을수록 후진 귀납법의 능력이 높음.

\* 결과: 부분 게임 완전 내쉬 전략

= 목표숫자가 5인 경우에는 (1,5)을 순차적으로 선택하는 것

= 목표숫자가 11인 경우에는 (3,7,11)을 순차적으로 선택하는 것

= 목표숫자가 14인 경우에는 (2,6,10,14)을 순차적으로 선택하는 것

= 목표숫자가 17인 경우에는 (1,5,9,13,17)을 순차적으로 선택하는 것

- 전략적 사고 (Introspective strategic thinking) 능력<sup>31)</sup>

\* 방법: 응답자의 위치(position)에 따라 답례금을 선택(choice)하는 게임임. 응답자는 자신이 응답한 답례금을 기본적으로 받고 응답자의 상대방보다 (A는 상대방이 없고, B의 상대방은 A, C의 상대방은 B, D의 상대방은 C, E의 상대방은 D) 10만큼 더 적게 선택하면 추가로 보너스를 10만원을 더 받음. 총 5번의 라운드를 위치를 옮겨가면서(position A, B, C, D, E) 진행됨. 즉 5번의 라운드에서 결정된 답례금의 합이 높을수록 전략적 사고 능력이 높음.

\* 결과: 내쉬 균형 이론

= 포지션 A에서는 50, B에서는 40, C에서는 30, D에서는 20, E에서는 10을 선택하는 것

30) 참고문헌: Gneezy, U., A. Rustichini, and A. Vostroknutov (2010) "Experience and insight in the Race game," *Journal of Economic Behavior & Organization*, 75(2), 144-155.

31) 참고문헌: Arad, A. and A. Rubinstein (2012) "The 11-20 Money Request Game: A level-k Reasoning Study," *American Economic Review*, 102(7), 3561-3573.

- 마음을 이해하는 능력<sup>32)</sup>

\* 방법: 상대방의 눈을 보고 해당 상태가 어떤 감정인지 읽어내는 능력을 측정하는 것으로 총 28문항으로 구성. 맞으면 1, 틀리면 0으로 코딩되어, 28문항 결과의 합이 높을수록 다른 사람의 마음을 이해하는 능력이 높음.

\* 결과: 각 문항별로 많은 사람이 응답한 것이 정답임.

<표 V-39> 20차년도 의사결정 부가조사 자료의 변수

내 용	20차
후진 귀납법의 능력	a200101 - a200104
전략적 사고 능력	a200201 - a200210
마음을 이해하는 능력	a200301
자율응답 여부	a200401

#### 4.14 21차년도 고용형태 부가조사<sup>33)</sup>

##### 4.14.1 고용형태 부가조사 자료의 구조

- 조사시기 : 21차년도(2018) 조사
- 조사대상 : 만 15세 이상 모든 응답자
- 조사목적 : 본 부가조사는 근래 다양한 형태의 고용이 증가하면서 기존의 비정규직 유형과 규모를 파악하기 위한 설문이 현실을 포착하기 어려워졌다는 인식 하에, 특고 등 임금노동자와 자영업자의 특징을 모두 가진 일자리 규모 파악을 목적으로 한다.
- 조사내용 : 자영업여부, 임금결정방식, 근로시간과 장소, 컴퓨터 사용여부, 소득

32) 참고문헌: Barron-Cohen, S., S. Wheelwright, and T. Jolliffe (1997) "Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome," *Visual Cognition*, 4(3), 311-331.

33) 본 부가조사는 다양한 고용형태를 파악하기 위해 작성된 것으로, 본조사에서 응답되지 않은 일자리도 파악하려고 노력하였음. 따라서 본조사에서 응답되지 않고, 부가조사에서만 응답된 일자리는 따로 일자리고유번호(jobnum)는 결측처리 하였음에 주의를 요함.

#### 4.14.2 고용형태 부가조사 자료의 내용

□ 고용형태 부가조사 자료 내용은 <표 V-40>과 같으며, 각각의 내용에 대한 설명은 다음과 같다.

- 자영업자와 임금근로자 구별 문항: 개인사업자여부, 사업자등록증소유여부, 근로자 고용 방법, 상품가격 결정 여부, 근로형태 규제 여부, 임금결정방식

\* 명백한 자영업자: a211201=1 & a211202=1 & a211203=1 & a211204=1  
a211205=2

\* 명백한 임금근로자: a211201=2 & a211204=4 & a211205=1 & a211206=1

\* 이후 설문은 명백한 자영업자와 명백한 임금근로자를 제외하고 질문되었음.

- 일의 종류 및 장소: 일의 종류, 일하는 장소
- 컴퓨터 사용 여부: 고객을 얻기 위해 컴퓨터, 핸드폰, PDA 등 기기 사용 여부, 일에서 컴퓨터, 핸드폰, PDA 등 기기 이용 여부
- 근로 기간: 근무 일수, 하루 평균 근로시간, 대기시간
- 소득 및 세금: 근로 소득<sup>34)</sup>, 소득 출처, 납부 세금 종류

---

34) 부가조사설문과 개인조사설문의 소득 설문문항이 상이하여 응답값에 차이가 있을 수 있음에 유의해야 한다.

부가조사 설문(a211215): 지난 3개월간 이 일자리를 통해서 월평균 얼마정도의 소득을 올리셨습니까?

개인조사 설문(p\*\*1642): 현재 이 일자리에서 \_\_\_\_\_님의 임금은 얼마나 됩니까?

<표 V-40> 21차년도 고용형태 부가조사 자료의 변수

내 용		21차
일자리 여부		a211101 - a211102
일자리 번호		jobnum1 - jobnum3
첫 번째 일자리	임금근로자 자영업자 구별 문항	a211201 - a211206
	일의 종류 및 장소	a211207 - a211208
	컴퓨터 사용 여부	a211209 - a211210
	근로 기간	a211211 - a211214
	소득 및 세금	a211215 - a211217
두번째 일자리	임금근로자 자영업자 구별 문항	a211301 - a211306
	일의 종류 및 장소	a211307 - a211308
	컴퓨터 사용 여부	a211309 - a211310
	근로 기간	a211311 - a211314
	소득 및 세금	a211315 - a211317
세번째 일자리	임금근로자 자영업자 구별 문항	a211401 - a211406
	일의 종류 및 장소	a211407 - a211408
	컴퓨터 사용 여부	a211409 - a211410
	근로 기간	a211411 - a211414
	소득 및 세금	a211415 - a211417

#### 4.15 22차년도 비임금근로자 부가조사

##### 4.15.1 비임금근로자 부가조사 자료의 구조

- 조사시기 : 22차년도(2019) 조사
- 조사대상 : 비임금근로자(유형3, 유형7, 신규용 문3번의 (2) 응답자)
- 조사목적 : 본 부가조사는 비임금근로자인 고용주, 자영자, 무급가족종사자의 고용형태, 근로형태, 소득, 결근 등에 관해 파악하고자 하는 목적을 가진다.
- 조사내용 : 일자리 기본 정보, 일자리 선택 및 전망, 매출액 및 영업비용, 결근



#### 4.12.2 비임금근로자 부가조사 자료의 내용

□ 비임금근로자 부가조사 자료 내용은 <표 V-41>과 같으며, 각각의 내용에 대한 설명은 다음과 같다.

- 일자리 기본 정보: 시작시기, 평균 근로일 및 근로시간, 조직형태, 프랜차이즈 가맹점 여부, 영업시간, 휴무일수, 종사자수, 최저임금 근로자수, 사회보험 가입 근로자수
- 일자리 선택 및 전망: 비임금근로자 선택 이유, 임금근로자 취업 여부, 유보임금, 일자리 준비기간, 사업체 지속 여부, 그만 둘 시기, 그만 두려는 이유
- 일자리 매출액 및 영업비용: 주된 장소, 건물계약형태, 임대차 비용, 매출액영업 비용, 월평균 영업이익, 전자상거래 여부, 전자상거래 비중
- 결론: 결론 여부, 결론일수, 업무관련 여부, 몸이 아픈데도 불구하고 근로 여부, 근로일수

<표 V-41> 22차년도 비임금근로자 부가조사 자료의 변수

내 용		22차
일자리 번호		jobnum
일자리 기본정보	고용형태	a221101
	일자리 시작 시기	a221102- a221104
	평균 근로시간 및 근로일	a221105 - a221106
	조직형태	a221201
	프랜차이즈 가맹점 여부	a221202
	영업시간 및 영업기간	a221203 - a221211
	휴무일 및 휴무일수	a221212 - a221213
	종사자 수	a221214 - a221216
일자리 선택 및 전망	비임금근로 선택이유	a221301
	임금근로 취업 의사 여부 및 유보임금	a221302 - a221303
	일자리 준비기간	a221304
	사업체 지속 여부 및 기타	a221305 - a221307
매출액 및 영업비용	주된 장소 및 건물형태	a221401 - a221413
	월평균 매출액 및 영업비용	a221414 - a221423
	전자상거래	a221424 - a221426
결론	결론 여부 및 결론 일수	a221501 - a221504
	몸이 아픈데도 근로 여부 및 근로일수	a221505 - a221506

노동패널자료의 조사차수가 늘어나면서 자료의 한계로 인해 할 수 없었던 많은 실증연구들이 가능해졌다. 그러나 패널자료 분석에 뛰어들었던 많은 연구자들이 이내 분석을 포기하고 마는 경우가 종종 있는데, 이는 변수가 워낙 많고 자료도 복잡하기 때문이다. 19차 조사 자료에만 총 2천여개의 변수들이 개인용·가구용·부가조사용·직업력 자료를 구성하고 있다. 이 변수들 중에서 분석모형에 사용할 변수들을 일일이 찾아내야 한다. 고생 끝에 필요한 변수들을 찾아낸다 하더라도 어떤 것을 선택해야 적절한지 난감할 수 있다.

당연한 말이겠지만 패널자료를 잘 쓰기 위해서는 인내심을 갖고 느긋하게 자료를 이리저리 만져보는 것이 필요하다. 코드북과 데이터만 갖고서는 제대로 된 분석을 하기 어렵다. **동일한 변수라도 각 연도별로 항목값이 다른 경우가 많으므로 반드시 각각의 변수에 대해 설문지를 확인해야 한다.** 만일 노동패널을 처음 사용하는 연구자라면 자료를 만지기 전에 아래의 절차를 밟기를 추천한다.

1. 먼저 유저가이드 1장 한국노동패널소개와 3장 자료의 주요 특성 부분을 읽어본다.
2. 자료를 열어보기 전에 먼저 설문지를 전체적으로 훑어본다. 특히 관련 문항들의 전후구조에 유의한다.
3. 다음으로 설문지에 분석주제와 관련된 문항들이 있는지를 하나하나 체크한다.
4. 설문지에서 체크한 문항들이 실제로 코드북의 변수값과 일치하는지를 체크한다.
  - 이때 해당 주제와 관련한 내용이 유저가이드에 언급되어 있는지를 확인한다. 유저가이드의 변수가이드와 주제별 가이드에는 연구자들이 놓치기 쉬운 데이터의 핵심 내용을 최소한으로 요약한 것이므로 참조가 될 것이다.
5. 데이터 기초분석을 통해 관련 변수들의 빈도값이 코드북에 나와있는 빈도값과 일치하는지를 확인한다. 이때 missing값의 처리에 주의한다.

◎ FAQ에 사용된 예제 프로그램에 대해

1. 아래의 예제 프로그램들은 노동패널 22차년도 학술대회용 자료를 기초로 작성되었다. 주의할 점은 자료를 불러오기 위해서는 경로를 지정해 주어야 된다는 점이다. 아래의 예제에서는 다음과 같이 지정하였다.

※ SAS library명은 다음과 같다: a-data 폴더, new-새폴더.

```
libname a 'C:\klips';  
libname new 'C:\klips\new';
```

※ STATA 는 cd로 파일경로를 지정하고, klips 데이터는 용량이 크므로 충분한 메모리를 할당 시켰다.

```
cd "c:\klips"  
clear  
set memory 700m, permanently
```

2. R에서 사용된 패키지는 최초 설치 후에는 재설치 할 필요가 없다. R studio 가동 시 1회만 실행해주면 된다. 아래의 예제 프로그램들에서는 총 5개의 패키지가 사용되었다. “Reshape2”, “dplyr”, “foreign”, “plyr”, “gmodels”.
3. 아래의 예제 프로그램들은 단지 ‘예제’일 뿐임을 주지하기 바란다. 즉, 해당주제와 분석목적에 따라 노동패널팀 필자들과는 얼마든지 다른 정의와 구분을 토대로 프로그램을 작성할 수 있다. 프로그램의 효율성 면에서도 얼마든지 훨씬 효율적인 프로그램을 만들 수 있을 것이다.

## Q 1

### 노동패널자료를 좀 더 빨리 사용할 수 있을까?

연구자라면 누구나 자료에 대한 욕심을 갖기 마련이다. 다른 사람들보다 먼저 자료를 받고 싶고 다른 사람들은 이용할 수 없는 정보들에 더욱 접근하고 싶어 한다. 노동패널도 마찬가지다. 벌써부터 왜 23차년도 자료가 아직까지 안 나오는지에 대한 문의가 많다. 노동패널 팀에서는 이제 겨우 22차년도 자료를 release하려고 하는 상황이다.

우선 하루라도 빨리 자료를 받아보고 싶은 연구자들을 위해 노동패널 자료가 만들어지는 과정을 눈여겨보시라고 말씀드리고 싶다. 아래는 22차년도 자료가 release되기까지의 과정이다.

- ① 조사설계 : 2019년 1-3월
- ② 현장실사 : 2019년 4-11월
- ③ 원자료 생성 및 1차 클리닝(실사업체) : 2019년 12월-2020년 5월
- ④ 원자료 가공 및 2차 클리닝(연구팀) : 2020년 1-6월
- ⑤ 학술대회용 자료 공개 : 2020년 6월<sup>35)</sup>
- ⑥ 자료 업로드 준비, 유저가이드, 코드북 제작 : 2020년 7-12월
- ⑦ release버전 자료 공개 : 2020년 12월

이상과 같이 약 2년의 과정을 거쳐야 올바른 자료를 생산할 수 있다. 물론, 이 과정에서 다른 년도의 조사 및 데이터 발간 작업이 복잡하게 이루어지기 때문에 이러한 데이터 준비과정을 앞당기기란 쉽지 않다는 점에 대한 양해가 필요하다.

어떠한 경우에도 자료를 가장 빨리 받을 수 있는 방법은 학술대회에 연구계획서를 신청하는 것뿐이다. 학술대회 연구계획서 접수는 매년 4월 초에 공고가 나가 6월 중순까지 접수를 받으며, 일정한 통과절차를 거치면 학술대회용 데이터를 제공받게 된다.

---

35) 2020년에는 내부사정으로 인해 학술대회용 자료는 2020년 9월에 공개되었다.

## Q 2

### SAS, SPSS, Stata, R, 엑셀 패키지 간 파일변환

노동패널 자료는 SAS 9.1, SPSS 13.0, Stata 9.0 이상의 버전에서 만들어졌다. 만일 SPSS나 Stata 사용자라면 낮은 버전의 패키지를 사용하더라도 자료를 불러오거나 다른 패키지로 전환할 때 별도의 주의가 필요 없다. 그러나, SAS의 경우 각 버전마다 데이터파일의 확장자가 다르기 때문에 이를 확인한 후에 변환해야 한다. 만일 이런 문제가 없는데도 자료를 불러올 수 없다면 자료에 에러가 발생한 것이므로 다시 자료를 수령해야 한다.

SAS, SPSS나 Stata이외에 다른 통계패키지를 사용하는 연구자들은 우선 데이터를 ASCII 파일이나 data파일로 바꾼 후, 다시 해당 패키지 파일로 바꾸어야 할 것이다. 요즘에는 많은 통계패키지들이 SAS, SPSS, Stata, 엑셀 간에 파일을 변환해주는 기능을 갖고 있기 때문에 이런 작업들이 필요 없는 경우가 대부분이다. R을 사용하는 경우에는 다음과 같은 두 가지 방법으로 노동 패널 데이터를 불러올 수 있다.

1. R studio를 이용하는 경우, 상단 메뉴바의 File → Import Data Set → 'From Spss', 'From SAS', 'From STATA' 메뉴를 통해 쉽게 데이터 파일을 불러올 수 있다.
2. R studio를 이용하지 않거나 명령어로 데이터를 불러오고 싶을 때는 'foreign'패키지를 통해 불러올 수 있다.

```
install.packages("foreign") # foreign 패키지를 설치해준다.
```

```
library(foreign)           # foreign 패키지를 실행한다.
```

그 다음 read.\*\*\* 명령어를 통해 원하는 파일을 불러온다.

## Q 3

### 98표본, 09통합표본, 18통합표본의 차이

노동패널은 현재 한 개의 데이터 안에 세 가지의 패널이 함께 존재한다. 98표본은 1998년 1차년도 당시에 조사된 5,000가구를 원가구로 해서 22차년도(2019년)까지 이어진 조사이다. 한국노동패널의 표본추가는 12차년도(2009년), 21차년도(2018년) 두차례에 걸쳐서 이루어졌다. 09통합표본은 2009년 패널 가구의 마모 및 도시지역 표집에 대한 한계를 극복하고, 패널데이터의 대표성을 확보하기 위해 추가로 포집된 가구를 포함하여, 2009년 1차 자료를 시작으로 2019년 11차년도(98표본기준 22차년도) 조사가 완료되었다<sup>36)</sup>. 2009년(12차 조사당시) 당시 응답한 모든 가구를 원표본으로 하는 자료이다. 따라서 09통합표본은 2009년 추가 포집된 1,415가구와 기존 조사 대상가구(98원표본 + 분가가구) 중 12차년도에 응답한 5,306개 가구를 더한 6,721가구로 구성되어 있다. 21차년도(2018년) 표본추가는 20년동안 진행된 조사기간 중 탈락하는 가구가 많아지면서 전체 모집단의 성격을 반영하기 못하는 것을 보완하기 위해 5,044가구를 추가하였으며, 2018년 1차 자료를 시작으로 2019년 2차년도(98표본 기준 22차년도) 조사가 완료되었다. 2018년(21차 조사당시) 당시 응답한 모두 가구를 원표본으로 하는 자료이다. 따라서 18통합표본은 2018년에 추가 포집된 5,044가구와 기존 조사 대상가구(09통합표본 + 09통합표본 분가가구 + 09통합표본에 포함되지 않은 98표본) 7,090개를 더한 12,134 가구로 구성되어 있다.

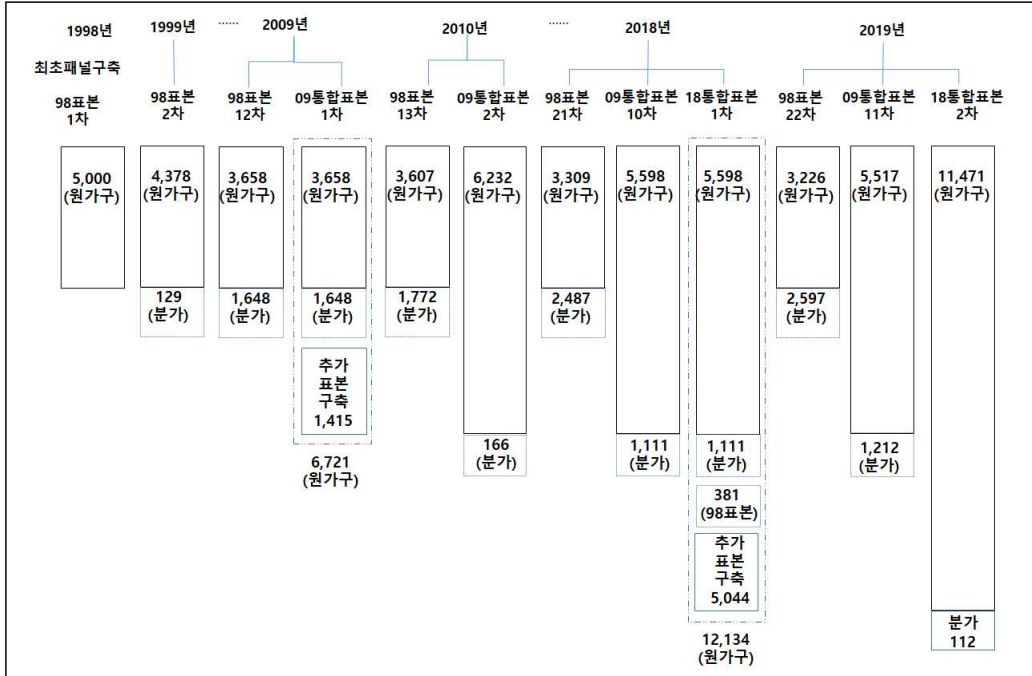
09통합표본과 18통합표본의 개념을 이해하기 위해 아래 [그림]을 살펴보면, 98표본은 1998년부터 2019년까지 22년 동안 원가구 5,000가구를 대상으로 추적 조사가 실시된 패널로, 22차년도에 원가구 3,266, 분가가구 2,557, 총 5,823가구가 조사되었다. 09통합표본은 2009년 당시 98표본 원가구 5,000가구 중 조사를 성공한 3,658가구와 그에 따른 분가가구 1,648가구, 추가 표본 1,415가구를 모두 합한 6,721가구가 통합표본 1차 원가구로 형성되어 2019년까지 11년 동안 조사가 실시되었다. 09통합표본 11차년도에는 원가구 5,517, 분가가구 1,212, 총 6,729가구가 조사되었다. 18통합표본은 2018년 당시 조사를 성공한 09통합표본 원가구 6,232가구 중 조사를 성공한 5,598가구와 그에 따른 분가가구 1,111가구, 09통합표본에는 포함되어 있지 않은 98표본 381가구, 추가 표본 5,044가구를 모두 합한 12,134가구가 통합표본 1차 원가구로 형

36) 『한국노동패널 22차년도 기초분석보고서』, 2장 제6절 09통합표본 개요 참조.

성되어 2019년까지 2년 동안 조사가 실시되었다. 18통합표본 2차년도에는 원가구 11,471, 분가가구 112, 총 11,583가구가 조사되었다.

[KLIPS 표본 구조 개요]

(단위:가구)



주: 98표본 원가구 중, 2009년도에 조사되지 않은 원가가구 다음해인 2010년에 조사된다면, 98표본에는 포함되지 않지만, 09통합표본의 원가구에 속하지 않았으므로 09통합표본에는 포함되지 않음. 따라서 통합표본은 2009년도에 조사된 가구만을 대상으로 추적 조사하는 패널임에 유의해야 함. 18통합패널도 마찬가지로 개념으로 2018년도 응답한 가구만을 원가구로 정의하였기 때문에, 2018년도에 응답하지 않고, 2019년도 이후에 응답한 가구는 98표본이나, 09통합패널일수는 있지만, 18통합패널에는 속하지 않음에 유의해야 함.

연구자는 연구목적에 따라 조사 기간이 긴 98표본을 사용할 것인지, 차수는 짧지만 상대적으로 표본의 대표성을 확대하여 표본의 숫자가 많은 통합표본을 사용할 것인지 결정해야 한다.

98표본을 사용하고자 할 경우에는 sample98변수의 값이 '1'(98표본 원가구)과 '2'(98표본 분가가구)인 케이스를 선택하고, 09통합표본을 사용하고자 할 경우에는 sample09 변수의

값이 '1'(09통합표본 원가구)과 '2'(09통합표본 분가가구)인 케이스를 선택하여야 하며, 18통합표본을 사용하고자 할 경우에는 sample18 변수의 값이 '1'(18통합표본 원가구)과 '2'(18통합표본 분가가구)인 케이스를 선택하여야 한다.

다만, 횡단면 자료를 사용하는 경우에는 표본의 구분 없이 사용하여도 무방하다.

<98표본, 09통합표본, 18통합표본 사용시 유의점>

자료명	가중치		
	98표본	09통합표본	18통합표본
가구용	w_h	sw_h	nw_h
개인용	w_p_l(종단면가중치) w_p_c(횡단면가중치) 단, 1차년도 98표본에는 w01p라는 하나의 가중치만 있음에 주의	sw_p_l(종단면가중치) sw_p_c(횡단면가중치) 단, 12차년도 09통합표본에는 sw12p라는 하나의 가중치만 있음에 주의	nw_p_l(종단면가중치) nw_p_c(횡단면가중치) 단, 21차년도 18통합표본에는 nw21p라는 하나의 가중치만 있음에 주의



## Q 4

### 원가구와 분가가구의 차이

노동패널조사는 가구패널조사로 1차년도 당시에 조사된 가구를 원가구로 해서 매해 추적 조사한다. 이후 원가구에서 분가한 가구가 발생한 경우, 어떠한 가구에서 분가한 것인지 알 수 있도록 노동패널의 모든 가구에 「원가구번호」를 부여하고 있다. 98표본의 「원가구 번호」는 orghid98 변수이고, 98년도에 최초로 조사된 가구인 1번부터 5000번 가구까지의 번호를 갖고 있다. 09통합표본의 「원가구 번호」는 orghid09 변수이다. orghid09변수는 2009년도 당시 통합표본의 원가구 번호를 의미하며, 1-101,415번까지의 값을 갖는다. 18통합표본의 「원가구 번호」는 orghid18 변수이다. orghid18변수는 2018년도 당시 통합표본의 원가구 번호를 의미하며, 2-205,044번까지의 값을 갖는다.

98표본을 중심으로 분가와 소멸을 설명하면 다음과 같다. 원가구는 시간이 지남에 따라 분가·합가·소멸 등의 이유로 사라지기도 하고, 새롭게 생기기도 한다. 따라서 결혼이나 분가 등의 이유로 원가구에서 분가한 가구에 대해서는 5001번부터 새로운 가구번호가 부여된다. 이 때 주의할 점은 분가가구라 하더라도, 1차년도 당시에 조사된 가구(즉 원가구)에 있던 가구원(즉 원가구원)이 분가하였을 경우에만 분가가구번호가 부여된다는 점이다. 따라서, 분가가구 내에 다른 가구원이 존재한다 하더라도 원가구원이 한명도 없는 경우에는, 가구가 소멸된 것으로 간주하고 있다. 즉, 원가구일 경우에는 원가구번호인 orghid98과 당해 연도 가구번호인 hhid22(22차)는 같은 번호를 갖게 되지만, 분가가구의 경우에는 orghid98와 hhid22는 다른 가구번호를 갖게 된다.

참고로 98표본의 각 년도별 분가가구번호를 살펴보면, 5000번 이내는 1차(1998) 원가구의 번호이며, 5,001-5,258번은 2차(1999), 5,259-5,444번은 3차(2000), 5,445-5,742번은 4차(2001), 5,743-5,982번은 5차(2002), 5,983-6,261번은 6차(2003), 6,262-6,525번은 7차(2004), 6,526-6,818번은 8차(2005), 6,819-7,132번은 9차(2006), 7,133-7,421번은 10차(2007), 7,422-7,780번은 11차(2008), 7,781-8,020번은 12차(2009), 8,021~8,294번은 13차(2010), 8,295-8,499번은 14차(2011), 8,500~8,718번은 15차(2012), 8,719-8,907번은 16차(2013), 8,908-9,066번은 17차(2014), 9,067-9,264번은 18차(2015), 9,265-9,456번은 19차(2016), 9,457-9,597번은 20차(2017), 9,598-9,687번은 21차(2018), 9,689-9,855번은 22차(2019) 분가가구 번호이다.

한편, 2009년 추가 표집 된 1,415가구는 원가구 번호가 100,001부터 시작하여 101,415번까

지를 부여하였으며, 이들 가구 중 다음 차수(2010년)에 분가가구가 발생한 경우 110,001번부터 순차적으로 가구번호가 부여되었다. 또한, 2018년 추가 표집 된 5,044가구는 원가구 번호가 200,001부터 시작하여 205,044번까지를 부여하였으며, 이들 가구 중 다음 차수(2019년)에 분가가구가 발생한 경우 210,001번부터 순차적으로 가구번호가 부여되었다.

**Q 5****개인자료와 부가자료 합치기  
기준 변수 PID를 이용한 Merge**

부가조사 자료를 분석하기 위해서는 개인에 있는 핵심 변수들을 붙여야 하는 경우가 대부분이다. 예컨대, 어떤 연구자가 11차년도 교육 부가조사 중 “고2 때 수면시간” 변수를 개인자료에 붙여서 사용할 경우에는 다음과 같이 하면 된다.

① 11차년도 교육 부가조사에서 가져올 변수들을 keep 한다. (본 예시에서는 “고2때 수면시간” 변수인 a116257 와 merge 에 필요한 pid를 남긴다)

② 22차년도 개인자료와 11차년도 부가조사 자료를 각각 pid 를 기준으로 정렬(“sort”)한다.

③ 정렬시킨 두 자료를 merge 한다.

**<SAS>**

\* Merge문의 사용 ;

```
ibname a 'C:\klips'
```

```
data a11; set a.klips11a;
```

```
keep pid a116257;
```

```
proc sort; by pid;run;
```

```
data p22; set a.klips22p;
```

```
proc sort; by pid; run;
```

```
data new.pa2211; merge p22 a11; by pid;
```

```
proc freq; table a116257;
```

```
run;
```

## < SPSS >

GET FILE='C:\klips\klips11a.sav'/keep pid a116257.

DATASET NAME a10.

SORT CASES BY PID.

GET FILE='C:\klips\klips22p.sav'.

SORT CASES BY PID.

MATCH FILES /FILE=\*

/FILE=a10

/BY PID.

EXE.

fre a116257.

SAVE OUTFILE='C:\klips\k22p11a.sav'.

## < STATA >

/\* Merge문의 사용 \*/

/\* data가 C:\klips\에 있다고 가정

메뉴바 이용법도 가능하다

1. 아래의 "use-" 줄까지 수행해서 자료를 불러놓은 다음

2. 메뉴바의 "data → combine datasets

→ merge two datasets → one-to-one on key variables

→ 파일 선택에서 C:\klips\klips11a.dta로 경로지정하고

→ key variables에는 pid 지정

→ sort key variables when necessary에 체크표시 한 후 ok

위와 같이 하면 아래의 "merge-" 부분이 실행됨

아래와 같이 명령문을 실행하면 \_merge라는 변수가 생성됨

1=원데이터에는 값이 없고 머지하고자 하는 데이터에는 값이 있는 경우

2=원데이터에는 값이 있으나 머지하고자하는 데이터에는 값이 없는 경우

3=양쪽에 값이 있어 제대로 붙은 경우 \*/

```
use klips11a
```

```
sort pid
```

```
merge pid using klips22p, unique sort
```

```
tab _merge //머지 결과의 확인
```

```
tab a116257
```

위의 프로그램은 SAS, Stata의 merge문과 SPSS의 match문을 이용하여 variable merge를 한 결과이다.

<R>

```
# Merge문의 사용 #
```

```
#데이터를 불러온다.
```

```
klips11a <- read.spss(file = "C:\\klips\\klips11a.sav", use.value.labels = FALSE,  
to.data.frame = TRUE )
```

```
klips22p <- read.spss(file = "C:\\klips\\klips22p.sav", use.value.labels = FALSE,  
to.data.frame = TRUE )
```

#11차년도 교육 부가조사에서 가져올 변수들을 keep한다. 데이터 가공을 위한 패키지 dplyr를 로드한다.

```
install.packages("dplyr") #패키지는 1회 설치 후에는 다시 설치할 필요가 없다.
```

```
library(dplyr)
```

```
a11 <- select(klips11a, pid, a116257)
```

#22차년도 개인자료와 11차년도 부가조사 자료를 각각 pid 기준으로 정렬한 후 병합한다.

```
pa2211 <- merge(klips22p, a11, by = "pid", all = TRUE)
```

```
head(pa2211)
table(pa2211$a116257)
View(pa2211)
```

## Q 6

### 가구자료와 개인자료 합치기 기준 변수 HHID\*\*를 이용한 Merge

노동패널자료에서는 가구자료와 개인자료를 병합해서 사용해야 할 경우가 많다. 예컨대, 22차년도 자료를 이용하여 가구총소비를 종속변수로 하는 회귀분석 모형을 구성한다고 하자. 그리고 설명변수를 가구총소득, 자산, 부채, 총가구원수, 가구주의 성별, 연령, 가구주의 경제활동상태 등의 변수들로 구성한다.

이 때 다른 모든 변수들은 가구자료 내에서 자체적으로 구성할 수 있지만, 가구주의 경제활동상태는 개인자료에서 별도로 구성한 다음 가구자료로 붙여야 한다.

이런 경우 다음과 같은 절차를 따르면 된다.

- ① FAQ 22에서와 같이 개인의 경제활동상태 변수를 구성한다.
- ② 가구주의 경제활동상태만을 "keep"한다. 개인자료의 경우 'p220102 = 10'인 경우가 가구에 해당한다.
- ③ 22차년도 개인자료와 가구자료를 각각 hhid22를 기준으로 정렬("sort")한다.
- ④ 가구자료와 개인자료의 병합은 앞서 FAQ Q5에서 다룬 '기준변수'를 이용한 variable merge를 이용하면 된다.

만일 21차년도 가구자료와 개인자료를 merge한다면 → hhid21

만일 20차년도 가구자료와 개인자료를 merge한다면 → hhid20

이와 같이 기준변수를 정하여 merge하면 된다.

## Q 7

### 직업력자료에 개인의 성별, 연령 붙이기

개인단위의 일자리 정보를 모두 수록한 자료가 바로 직업력 자료이다. 그런데, 이 자료를 사용하는 많은 사람들이 ‘왜 직업력 자료에는 개인의 성별, 연령, 학력 등과 같은 인구학적 특성 정보가 자료에 붙어있지 않은지에 대한 불편함을 호소하는 경우가 있다.

직업력 자료는 기본적으로 일자리 단위로 케이스가 만들어진 데이터 셋이다. 즉 한 개인이 1차 조사부터 지금까지 총 20개의 일자리를 가졌었거나 가졌다면 20개의 ‘행’이 생성된다. 더구나, 그 개인이 15세이후부터 1차조사 시점 직전까지 5개의 일자리를 가졌었다면, 20개의 일자리에 5개를 더하여 총 25개의 일자리 ‘행’이 생성되는 것이다.

이상과 같이 직업력 자료의 크기가 이미 방대하기 때문에 직업력 자료에는 가급적 변수를 최소화하였다. 즉, 개인의 인적정보는 연구자가 개인(신규)자료에서 붙여 쓰는 것이 훨씬 더 효율적이라고 판단한 것이다. 직업력 자료에 개인의 인적정보를 붙이는 작업은 앞서 Q5에서 다룬 ‘기준변수를 이용한 데이터 merge’를 이용하여 간단하게 실행할 수 있다.

#### < SAS >

```
/* Merge문의 사용 */  
libname a 'C:\klips'  
  
data p22; set a.klips22p(keep= pid p220101 p220107);  
rename p220101=sex p220107=age;  
proc sort; by pid;  
  
data w22; set a.klips22w;  
if jobwave=22; /*22차에 응답한 가구 추출*/  
proc sort; by pid;
```



```
data new.work:merge p22 w22; by pid; if jobwave=22;
run;
```

### < SPSS >

```
GET FILE='C:\klips\klips22p.sav'/KEEP pid p220101
p220107/RENAME(p220101=sex)(p220107=age).
DATASET NAME p22.
SORT CASES BY PID.
```

```
GET FILE='C:\klips\klips22w.sav'.
select if(jobwave=22). /*22차에 조사된 사람만 남기기*/
SORT CASES BY PID.
```

```
MATCH FILES /FILE=*
/table=p22
/BY PID.
```

```
select if (jobwave=22).
fre sex.
```

```
SAVE OUTFILE='C:\klips\k22wp.sav'
```

### < STATA >

```
/* merge문의 사용 */
/* 직업력과 개인자료를 머지할 때는 직업력에는 같은 사람이 여러 번 나타날 수 있고
개인자료에는 같은 사람이 단 한번만 나타난다는 점을 감안해야 한다.
이를 위한 머지는 메뉴바에서 다음과 같이 하면 된다
FAQ 4번에서 one-to-one on key variables 대신
```

many-to-one on key variables를 선택하고 나머지는 모두 똑같이 하면 된다. \*/

```
use klips22p
keep pid p220101 p220107
rename p220101 sex
rename p220107 age
sort pid
save k22p, replace

use klips22w, clear
keep if jobwave==22 /* 22차에 조사된 일자리에만 성별 연령 머지하기 */
sort pid
merge pid using k22p, uniqueness
keep if jobwave==22
table _merge //merge 결과의 확인
```

<R>

```
# Merge문의 사용 #
install.packages("foreign") #패키지 설치는 1회만 해주면 된다.
library(foreign)
klips22p <- read.spss(file = "C:\\klips\\klips22p.sav", use.value.labels = FALSE,
to.data.frame = TRUE )
klips22w <- read.spss(file = "C:\\klips\\klips22w.sav", use.value.labels = FALSE,
to.data.frame = TRUE )

#개인 자료 중 기준변수(pid)와 성별, 연령을 추출한다.
p22 <- klips22p[,c("pid", "p220101", "p220107")]

#성별과 연령 변수명을 바꿔준다.
```

```
names(p22) <- c("pid", "sex", "age")
```

```
#직업력 데이터 중 22차에 응답한 가구를 추출한다.
```

```
w22 <- klips22w[klips22w$jobwave == 22 ,]
```

```
#pid 를 기준으로 성별,연령 변수를 직업력 데이터에 병합한다.
```

```
new.work <- merge(p22, w22, by = "pid")
```

```
table(new.work$sex)
```

## Q 8

orghid98-hhid22 중에서 무엇을 가구 id로 사용해야 할까?  
SAS, SPSS, Stata문 활용 예 : 가구데이터의 연도간 merge

위의 Q6에서 본 바와 같이, 노동패널의 가구번호는 크게 당해 연도 가구번호와 원가구 번호로 구성된다. 노동패널 가구용 데이터에는 가구번호가 여러 개 존재하는데, 이는 노동패널 조사에 "한번이라도 응답한 적이 있는 모든 가구"를 기준으로 데이터가 생성되어 있기 때문이다.

22차년도 가구용 데이터에는 orghid98, orghid09, orghid21, hhid02-hhid22 과 같이 22개의 가구번호가 있다. 이들은 각각 1998년 원가구번호(orghid98), 2009년 추가표본 패널고유가구번호(orghid09), 2018년 추가표본 패널고유가구번호(orghid18), 2차가구번호(hhid02) - 22차가구번호(hhid22)를 의미한다. 원가구번호는 Q6에서 설명한 바와 같이 1차년도 조사당시의 가구번호를 의미하며, 2-22차 가구번호는 각 해마다 조사된 가구번호를 의미한다. 따라서 해당연도에 조사되지 않은 가구일 경우 해당연도의 가구번호가 missing값을 가진다.

예를 들어, 1차년도에 3962번이었던 가구가 3차년도에 5273번 가구로 분가한 후, 5차년도에는 조사에 실패했다면, 이 가구의 가구번호는 orghid98은 3962, hhid02는 3962, hhid03은 5273, hhid04는 5273, hhid05는 missing을 갖게 되는 것이다.

만일 연구자가 동일한 가구끼리 연도 간 merge를 할 경우에는, 연구의 목적에 따라 가구번호를 사용하면 된다. 그러나 가구의 변동과 같은 특정상황을 분석하고자 하는 것이 아니라면, 일반적으로 원가구번호가 아닌 당해 연도 가구번호를 기준으로 merge하는 것이 대부분이다. 이와 관련한 예제를 살펴보면 아래와 같다.

## < SAS >

\* 가구용 데이터의 년도간 merge ;

```
libname a 'C:\klips'
```

```
data h21; set a.klips21h;
```

```
hid=hhid21;
```

```
proc sort; by hid;
```

```
data h22; set a.klips22h;
```

```
hid=hhid22;
```

```
proc sort; by hid;
```

```
data new.h21h22; merge h21h22;by hid;
```

```
if hwave22=1; /*21차에 응답한 가구 추출*/
```

```
run;
```

## < SPSS >

```
GET FILE='C:\klips\klips21h.sav'.
```

```
dataset name h21.
```

```
compute hid=hhid21.
```

```
sort cases by hid.
```

```
GET FILE='C:\klips\klips22nd.sav'.
```

```
compute hid=hhid22.
```

```
sort cases by hid.
```

```

MATCH FILES /FILE=*
  /FILE=h21
  /BY hid.
select if (hwave22=1). /*22차에 조사된 가구만 남기기*/
fre hwave22.

SAVE OUTFILE='C:\klips\k21h22h.sav'.

```

### < STATA >

```

/* 가구용 데이터의 연도간 merge */

use klips22h
drop if hhid22==.
gen hid=hhid22
sort hid
save k22h, replace
use klips22h, clear
keep if hwave22==1
gen hid=hhid22 //21차에 조사된 가구만 남기기
sort hid
merge hid using k22h, unique sort
table _merge
table hwave22

```

### < R >

```

# 가구용 데이터의 연도간 merge

#데이터를 불러온다.
library(foreign)

```

```
klips21h <- read.spss(file = "C:\\klips\\klips21h.sav", use.value.labels = FALSE,
to.data.frame = TRUE )
klips22h <- read.spss(file = "C:\\klips\\klips22h.sav", use.value.labels = FALSE,
to.data.frame = TRUE )
```

#21차, 22차 년도의 가구데이터 가구번호(해당연도 가구번호) 통일

```
library(plyr)
library(dplyr)
```

```
klips21h <- klips21h %>% filter(is.na(klips21h$hhid22) == FALSE)
h21 <- plyr::rename(klips21h, c("hhid21" = "hid"))
```

```
klips22h <- klips22h %>% filter(hwave22 == 1)
h22 <- plyr::rename(klips22h, c("hhid22" = "hid"))
```

#hid 를 기준변수로 하여 두 가구 데이터 병합.

```
h21h22 <- merge(h21, h22, by = "hid")
```

```
table(h21h22$hwave22)
```

**Q 9****15세 이상 가구원 수, 15세 미만 가구원수 구하기**

가구자료를 이용한 분석모형을 구성할 때 가구원수는 중요한 변수로 사용된다. 예컨대, 소득불평등 지표를 구성할 때 가구원수는 ‘조정계수’로 사용된다. 총가구원수는 가구자료에 변수로 들어가 있다. 그러나 연구자에 따라서는 경제활동가능인구인 15세 이상 가구원수 같은 변수들이 중요할 수 있다. 여성노동시장 참가여부 모형을 구성할 때 5세미만 자녀수, 혹은 3세미만 자녀수 등이 중요한 도구변수로 사용되기도 한다. 이런 경우 특정 조건을 만족하는 가구원 수는 별도의 프로그램을 만들어 구성해야 한다. 22차년도 가구자료를 이용하여 15세 이상/미만 가구원수를 구해보자.

**<SAS>**

\* 가구원수 만들기 (Array 문의 활용) :

```
libname a 'C:\klips'
data h22; set a.klips22h;
if hwave22=1;          /*22차에 응답한 가구 추출*/
array p[15] lonum1-lonum15;    /*15×1 dummy array 생성(15세 미만)*/
array q[15] hinum1-hinum15;    /*15×1 dummy array 생성(15세 이상)*/
array y[15] h220361-h220375;   /*첫가구원의 만나이-15번째 가구원의 만나이*/
array c[15] h220421-h220435;   /*가구원 개인의 추가탈락여부*/

do i=1 to 15;
if y[i]<15 and 1<=c[i]<=2 then p[i]=1; else p[i]=0;
if y[i]>=15 and 1<=c[i]<=2 then q[i]=1;else q[i]=0;
end;

lonum=sum(of lonum1-lonum15);
hinum=sum(of hinum1-hinum15);
```



```
label lonum='15세미만 가구원수' hinum='15세 이상 가구원수';
proc means; var h220150 lonum hinum ;
proc freq; table lonum hinum ;
run;
```

< SPSS >

\* Count 문의 활용 .

\*추가탈락여부=change\*

```
GET FILE='C:\klips\klips22h.sav'.
select if (hwave22=1).
```

```
compute lonum=0.
compute hinum=0.
```

```
do repeat
change=h220421 to h220435/
age=h220361 to h220375.
if((change=1 or change=2) and (age<15)) lonum=lonum+1.
if((change=1 or change=2) and (age>=15)) hinum=hinum+1.
end repeat.
```

```
var lab lonum '15세 미만 가구원수' hinum '15세 이상 가구원수'.
des h220150 lonum hinum/stat=mean std min max.
```

## < STATA >

```
/* Stata - 가구원수 만들기 */
```

```
/* 15세 미만 가구원수 = lonum */
```

```
/* 15세 이상 가구원수 = hinum */
```

```
use klips22h
```

```
keep if hwave22==1 /* 21차에 조사된 가구만 남기기 */
```

```
forvalue i=1/15 {
```

```
    local a=`i'+60
```

```
    local b=`i'+20
```

```
    gen ppp`i'=(h2203`a'<15 & (h2204`b'==1| h2204`b'==2))
```

```
    gen qq`i'=(h2203`a'>=15 & (h2204`b'==1| h2204`b'==2))
```

```
}
```

```
egen lonum=rowtotal(ppp*)
```

```
egen hinum=rowtotal(qq*)
```

```
sum h220150 lonum hinum
```

```
tab1 lonum hinum
```

<R>

# R - 가구원수 만들기 #

```
library(foreign)
klips22h <- read.spss(file = "C:\\klips\\klips22h.sav",
use.value.labels = FALSE, to.data.frame = TRUE )
h22 <- klips22h[klips22h$hwave22==1,] #21차 응답 가구만 추출.
dim(h21)
```

```
library(dplyr)
xx <- select(h22, h220361:h220375)
yy <- select(h22, h220421:h220435)
```

```
lo <- data.frame(ifelse((xx < 15) & (yy == 1|2), 1, 0))
hi <- data.frame(ifelse((xx >= 15) & (yy == 1|2), 1, 0))
total <- lo + hi
```

```
lonum <- rowSums(lo, na.rm = TRUE)
hinum <- rowSums(hi, na.rm = TRUE)
num <- rowSums(total, na.rm = TRUE)
```

```
summary(lonum)
summary(hinum)
summary(num)
```

```
install.packages("psych") #패키지 설치는 1회만 해주면 된다.
library(psych)
describe(lonum)
describe(hinum)
describe(num)
```

위의 프로그램을 실행시키면 22차년도 응답가구인 총 11,720케이스의 평균값이 아래와 같이 나올 것이다. 22차년도 노동패널조사에서 평균 가구원 수는 2.6명이며, 15세 이상 가구원수는 2.2명이다. 또한, 15세미만 가구원수는 0.4명으로 나타난다.

	평균값	표준편차	최소값	최대값
총가구원수	2.6	1.3	1	8
15세이상가구원수	2.2	1.0	1	7
15세미만가구원수	0.4	0.7	0	5

## Q 10

### 가구의 성별, 연령, 학력 등 인구학적 변수 추출하기 SAS Array 문, SPSS Repeat 문, Stata Foreach문 이용

가구자료를 이용하는 분석모형에서 가구의 성별, 연령, 학력 등의 인구학적 특성은 핵심적인 변수로 사용되곤 한다. 그러나 가구자료는 이들 정보에 대해 1가구 내 최대 15명의 정보를 담고 있기 때문에 몇 번째 가구원의 정보를 어떻게 취해야 하는지에 대한 문제에 부딪히게 된다. 이런 경우, SAS 사용자는 array문을, SPSS 사용자는 repeat문을 사용하여 쉽게 가구의 인구학적 특성 변수들을 구성할 수 있다.

#### <SAS>

\* Array문의 사용 ;

```
libname a 'C:\klips'
data h22; set a.klips22h;
if hwave22=1;
array h[15] h220261-h220275; /*가구주와의 관계-값이 10이면 본인(가구주)*/
array s[15] h220241-h220255; /*가구원의 성별*/
array a[15] h220361-h220375; /*가구원의 나이*/

do i=1 to 15;
if h[i]=10 then hsex=s[i];      /*가구주와의 성별*/
if h[i]=10 then hage=a[i];     /*가구주의 나이*/
end;

proc freq; table hsex;
proc means; var hage;
run;
```

## < SPSS >

\* Repeat문의 사용.

```
GET FILE='C:\klips\klips22h.sav'.
```

```
do repeat  
head=h220261 to h220275/  
sex=h220241 to h220255/  
age=h220361 to h220375.  
if(head=10) hsex=sex.  
if(head=10) hage=age.  
end repeat.
```

```
fre hsex.  
des hage/stat=mean std min max.
```

## < STATA >

/\* foreach문의 사용 \*/

```
use klips22h  
keep if hwave22==1
```

```
local i=1  
foreach head of varlist h220261-h220275 {  
  rename `head' head`i'  
  local i=`i'+1  
}  
local j=1
```

```

foreach sex of varlist h220241-h220255 {
  rename `sex' sex`j'
  local j=`j'+1
}
local k=1
foreach age of varlist h220361-h220375 {
  rename `age' age`k'
  local k=`k'+1
}
gen hsex=.
gen hage=.
foreach n of num 1/15 {
  replace hsex=sex`n' if head`n'==10
  replace hage=age`n' if head`n'==10
}
tab hsex
sum hage

```

<R>

# melt문의 사용 #

```

library(foreign)
klips22h <- read.spss(file = "C:\\klips\\klips22h.sav",
  use.value.labels = FALSE, to.data.frame = TRUE )
h22 <- klips22h[klips22h$hwave22==1,] #22차 응답 가구만 추출.
library(dplyr)
h <- select(h22,hhid22, h220261:h220275)
s <- select(h22,hhid22, h220241:h220255)
a <- select(h22,hhid22, h220361:h220375)

```

```

install.packages("reshape2") # 최초 1회 reshape2 패키지를 설치해준다.
library(reshape2)
hh ← melt(h, id.vars = c("hhid22"), variable.name = "number.h", value.name =
"relationship")
ss ← melt(s, id.vars = c("hhid22"), variable.name = "number.s", value.name = "sex")
aa ← melt(a, id.vars = c("hhid22"), variable.name = "number.a", value.name = "age")

hsa ← cbind(hh ,ss ,aa)
head(hsa) #전체 가구원의 성별, 나이 데이터를 담은 데이터 프레임 생성.

hsa.owner ← na.omit(hsa[hsa$relationship == 10, c("relationship", "sex", "age")])

table(hsa.owner$sex)
summary(hsa.owner$age)

```



## Q 11


### 가구자료에서 모든 가구원의 성별, 연령, 가구주와의 관계 추출 SAS Transpose 문, SPSS Restructure , Stata Reshape 기능 이용

가구자료에 수록된 가구원의 성별, 연령, 가구주와의 관계, 학력 등의 정보들은 노동패널 팀에서 자체 가공하여 개인자료에 “(가구정보)”라는 라벨이 붙은 변수로 제공하고 있다. 따라서 개인자료를 분석하는 일반적인 상황이라면 가구자료에서 가구원의 인적특성 정보들을 별도로 추출할 필요가 없다. 그러나 연구상황에 따라 개인자료에 포함되지 않은 개인의 인적 특성 등이 필요한 경우가 발생 할 수 있다. 대부분의 범용 통계패키지들은 행단위의 연단보다는 열단위(변수단위)의 연산에 적합하도록 만들어졌다. 따라서 이런 문제를 풀기 위해서는 자료의 적절한 조작이 필요하다.

가구자료의 성별, 연령, 가구주와의 관계는 아래 그림과 같이 1가구에 소속된 가구원일 경우 1개 case내에 붙어 있다. 예컨대, 1가구에 소속된 가구원이 10명이라면 1부터 15번째 가구원의 성별변수들 사이에 흩어져서 각각의 값이 들어가 있다. 우리의 목적은 이렇게 가로형태로 붙어있는 성별변수를 개인 케이스별로 잘라내어 다시 붙이는 것이다. 이렇게 하면 케이스는 조사된 개인 수에 따라 늘어나겠지만 성별변수는 원래 15개에서 1개로 줄일 수 있다.

<가구자료의 가구원 정보>

가구원의 인적정보(최대 15명)					
h**0241	h**0242	...	h**0361	h**0362	...
1번째 가구원의 성별	2번째 가구원의 성별	...	1번째 가구원의 만나이	2번째 가구원의 만나이	...



성별	연령	가구주와의 관계	...
1번째 가구원			
2번째 가구원			
...			
1번째 가구원			
2번째 가구원			
...			

아래 프로그램들은 가로로 15개씩 나열되어있는 데이터를 하나의 변수로 세로 형태로 전치(transpose) 시켜 묶는 것을 기본으로 하고 있다.

- ① 22차년도 가구자료로부터 1부터 15번째 가구원들의 성별, 가구주와의 관계, 연령변수를 추출하여 4개의 취합된 data set로 묶는다.
- ② 이렇게 생성된 4개의 데이터를 case merge하여, hhid22 기준으로 붙인다.

SAS는 이를 위해 transpose 문을, SPSS는 restructure 기능을, Stata는 reshape문을 사용한다. 세 코드를 비교해보면, 결국 동일한 변환과정임을 쉽게 알 수 있다.

### < SAS >

\* Transpose 문의 사용;

```
libname a 'c:\klips'
data h22; set a. klips22h;
keep hhid22 hwave22
h220221-h220235 /*pid*/
h220241-h220255 /*sex*/
h220261-h220275 /*relation*/
h220361-h220375 /*age*/
h220421-h220435 /*추가탈락여부*/;

if hwave22=1;

proc sort; by hhid22;
run;

data h22; set h22;
proc transpose data=h22 out=trans1(rename=(col1=pid)); by hhid22;
var h220221-h220235;
```

```

proc transpose data=h22 out=trans2(rename=(col1=sex)); by hhid22;
    var h220241-h220255;
proc transpose data=h22 out=trans3(rename=(col1=rel)); by hhid22;
    var h220261-h220275;
proc transpose data=h22 out=trans4(rename=(col1=age)); by hhid22;
    var h220361-h220375;
proc transpose data=h22 out=trans5(rename=(col1=change)); by hhid22;
    var h220421-h220435;

data fnl; merge trans1 trans2 trans3 trans4 trans5; by hhid21;
if 1<=change<=2; /*이번 차수에서 조사된 가구와 가구원만 남기기(기존+신규)*/
drop _name_ _label_; /*불필요한 변수 탈락*/

proc means; var sex rel age;
proc freq; table sex rel age change;
run;

```

## < SPSS >

\* Restructure 기능의 사용.

```

GET FILE='C:\klips\klips22h.sav'
/keep hhid22 hwave22 h220221 to h220235 h220241 to h220255 h220261 to h220275
h220361 to h220375 h220421 to h220435.

```

```

*변수설명: h220221-h220235 /*pid*/ h220241-h220255 /*sex*/ h220261-h220275 /*가구
주와의 관계*/
h220361-h220375;/*만나이*/ h220421-h220435/*추가탈락여부*/
/* data restructure (variables into cases) 메뉴에서 실행하는 법*/
/* data → restructure → variables into cases → more than one (5) : pid, sex, rel, age,
change */

```

```
/* 각각의 1~15가구원 변수들을 pid, sex, rel, age, change 로 묶어 지정. */
/* hhid22을 fixed variable 로 지정한다 */
```

```
VARSTOCASES /ID = id
/MAKE pid FROM h220221 to h220235
/MAKE sex FROM h220241 to h220255
/MAKE rel FROM h220261 to h220275
/MAKE age FROM h220361 to h220375
/MAKE change FROM h220421 to h220435
/INDEX = Index1(15) /KEEP = hhid22 hwave22 /NULL = KEEP.
```

```
/* 이번 차수 조사된 가구와 가구원만 남기기 */
```

```
select if (change=1 or change=2).
```

```
fre sex rel.
```

```
des age/stat=mean std min max.
```

## < STATA >

```
/* reshape 기능의 사용 */
```

```
use klips22h
keep if hwave22==1
keep hhid22 h220221-h220235 h220241-h220255 h220261-h220275 h220361-h220375
h220421-h220435
local i=1
foreach rel of varlist h220261-h220275 {
rename `rel' rel`i'
local ++i
}
```

```

local j=1
foreach sex of varlist h220241-h220255 {
  rename `sex' sex`j'
  local ++j
}
local k=1
foreach age of varlist h220361-h220375 {
  rename `age' age`k'
  local ++k
}
local l=1
foreach pid of varlist h220221-h220235 {
  rename `pid' pid`l'
  local ++l
}
local m=1
foreach change of varlist h220421-h220435 {
  rename `change' change`m'
  local ++m
}
reshape long pid sex rel age change, i(hhid22) j(orderInHH)
keep if change==1 | change==2 /* 22차에 조사된 개인만 남기기*/

tab1 sex rel age

```

**<R>**

# melt, arrange 문의 사용 #

#Q.10 에서 이미 적절한 형태의 자료를 생성하였다(melt문 이용).  
head(hsa)

```
#가구별 hhid22로 정렬시켜준다.  
hsa.22 <- na.omit(hsa[, -c(2,4,5,7,8)]) # 불필요 컬럼 제거  
head(hsa.22)  
hsa.22 <- arrange(hsa.22, hhid22) # hhid 기준으로 정렬  
head(hsa.22)  
  
table(hsa.22$sex) # 성별 빈도  
summary(hsa.22$age) # 연령 통계량
```

## Q 12

### 사교육 자녀번호와 엄마번호 사용하기 - 가구원정보와 자녀(엄마)번호 연결하기

노동패널 가구용 자료는 사교육에 관한 변수들을 여러 가지 포함하고 있다. 이러한 사교육변수를 가구원 정보와 연결하여 사용하면, 보다 풍부한 분석이 가능해진다. 그렇다면, 가구원 정보와 사교육에 응답한 자녀번호(혹은 엄마번호)는 어떻게 연결될 수 있을까? 이는 자녀번호와 엄마번호가 부여되는 방식에 대한 이해가 갖춰지면 간단하게 알 수 있다. 자녀번호와 엄마번호는 각각 한 가구당 5쌍씩 있다. 22차의 경우, 자녀번호-1 (h221511)은 첫 번째 자녀가 그 가구의 몇 번째 가구원인지를 기록하며, 엄마번호-1 (h221512)은 자녀-1의 엄마가 그 가구의 몇 번째 가구원인지를 기록하고 있다. 보다 구체적인 예를 들어보자.

노동패널 22차년도 가구자료(klips22h)를 열어서 22차년도 가구번호(hhid22)가 5016인 가구를 살펴보자. FAQ Q8에서도 보았듯이, 이 가구는 원가구번호가 3618이며, 2차년도에 분가한 가구이다. 그런데, 2차년도 분가 당시 원가구 3618번 가구에서 총 4명의 가구원이 함께 분가하였다. 이때 분가한 가구원들의 pid는 순서대로 501602, 501603, 501604였다 (한 개인의 pid는 절대 변하지 않음을 기억하자). 이 가구의 첫 자녀번호(h221511)를 살펴보면 4이다. 즉, 5016 가구의 4번째 가구원임을 알 수 있다. 한 가구의 pid는 h220221부터 h220235까지 15개의 값이 옆으로 기록되어 있다. 따라서 이 아이의 pid는 4번째 가구원의 pid(h220223)에서 찾으려 하는 것이다. 이제 이 아이의 엄마에 대해 알아보자. 첫 아이의 엄마의 번호 (h221512)는 1이므로, 엄마의 pid는 두 번째 가구원의 pid (h220221)에서 찾으려 한다. 5016 가구에는 아이가 둘이고, 각각의 pid가 501603와 501604임을 직접 확인 해 보자.

참고로 98표본의 각 년도별 분가가구번호를 살펴보면, 5000번 이내는 1차(1998) 원가구의 번호이며, 5001-5258번은 2차(1999)에 분가한 가구, 5259-5444번은 3차(2000)에 분가한 가구, 5445-5742번은 4차(2001), 5743-5982번은 5차(2002), 5983-6261번은 6차(2003), 6262-6525번은 7차(2004), 6526-6818번은 8차(2005), 6819-7132번은 9차(2006), 7133-7421번은 10차(2007), 7422-7780번은 11차(2008), 7781~8020번은 12차(2009), 8021~8294번은 13차(2010), 8295~8499번은 14차(2011), 8500~8718번은 15차(2012), 8719-8907번은 16차(2013), 8908-9066번은

17차(2014), 9067-9264번은 18차(2015), 9265-9456번은 19차(2016), 9457-9597번은 20차(2017) 분가가구 번호이고, 9,598-9,687번은 21차(2018) 분가가구 번호이고, 9689-9855번은 22차(2019) 분가가구 번호이다.

이와 같이, 자녀번호와 엄마번호를 이용하면 5쌍의 자녀/엄마 각각의 정보를 추출하는 것이 가능해진다. 여러 변수에 대해 반복적인 작업을 할 때에는 array (SAS) 혹은 vector (SPSS)를 이용해서 작업할 변수들을 묶은 뒤 do 문을 이용하면 편리하다. 아래 예제들에서는 이러한 방법으로 필요한 가구원 정보를 추출한 뒤, Q10에서와 마찬가지로 transpose (SAS) 혹은 restructure (SPSS)를 사용하여 취합하는 과정을 거치고 있다. 두 예제는 구현방식만 다를 뿐 논리적 구성은 똑같음을 확인 할 수 있다. 아래 두 예제는 각 case(아이) 별로 아이의 성별, 나이, 엄마 정보, 그리고 사교육정보를 가로로 추출한다. 엄마의 pid 가 기록되어있으므로 엄마의 개인설문 정보와의 merge 도 손쉽게 할 수 있다.

#### < SAS >

\* 자녀별 성별, 나이, 엄마정보, 사교육정보 추출하기: Array 문, Transpose 문 사용;

```
libname a 'C:\klips'
data h22; set a.klips22h;
if h221501=1;
run;
```

```
data h22_1; set h22;
array ppid[15] h220221-h220235;
array ssex[15] h220241-h220255;
array aage[15] h220361-h220375;
```

/\*예) i=4일때: h221611=4 즉 두번째 자녀 번호가 4(두번째 아이가 4번째 가구원)인 경우, 그 아이의 pid(chpid2)는 4번째 가구원의 pid,  
그 아이의 성별(sex)은 4번째 가구원의 성별, 그 아이의 나이(chage2)는 4번째 가구원의 만나이로 기



록\*/

```
do i=1 to 15;
if h221511=i then do; chpid1=ppid[i]; sex1=ssex[i]; chage1=aage[i]; end;
if h221611=i then do; chpid2=ppid[i]; sex2=ssex[i]; chage2=aage[i]; end;
if h221711=i then do; chpid3=ppid[i]; sex3=ssex[i]; chage3=aage[i]; end;
if h221811=i then do; chpid4=ppid[i]; sex4=ssex[i]; chage4=aage[i]; end;
if h221911=i then do; chpid5=ppid[i]; sex5=ssex[i]; chage5=aage[i]; end;
```

```
if h221512=i then do; mompid1=ppid[i]; momage1=aage[i]; end;
if h221612=i then do; mompid2=ppid[i]; momage2=aage[i]; end;
if h221712=i then do; mompid3=ppid[i]; momage3=aage[i]; end;
if h221812=i then do; mompid4=ppid[i]; momage4=aage[i]; end;
if h221912=i then do; mompid5=ppid[i]; momage5=aage[i]; end;
end;
run;
```

```
data h22_1; set h22_1;
proc sort; by hhid22;
/*transpose 문의 활용. 5개씩 데이터를 하나로 묶되, hhid21기준으로 정렬한다*/
proc transpose data=h22_1 out=trans1(rename=(col1=chpid));by hhid22;
    var chpid1-chpid5;    /*자녀 pid*/
proc transpose data=h22_1 out=trans2(rename=(col1=sex));by hhid22;
    var sex1-sex5;    /*자녀 성별*/
proc transpose data=h22_1 out=trans3(rename=(col1=age));by hhid22;
    var chage1-chage5;    /*자녀 만나이*/
proc transpose data=h22_1 out=trans4(rename=(col1=mompid));by hhid22;
    var mompid1-mompid5;    /*엄마 pid*/
proc transpose data=h22_1 out=trans5(rename=(col1=momage));by hhid22;
    var momage1-momage5;    /*엄마 만나이 */
proc transpose data=h22_1 out=trans6(rename=(col1=type1)); by hhid22;
```

```

var h221516 h221616 h221716 h221816 h221916; /* 사교육 종류 (1번째) */
proc transpose data=h22_1 out=trans7(rename=(col1=cost1)); by hhid22;
var h221519 h221619 h221719 h221819 h221919; /* 사교육 비용 (1번째) */
proc transpose data=h22_1 out=trans8(rename=(col1=type2)); by hhid22;
var h221520 h221620 h221720 h221820 h221920; /* 사교육 종류 (2번째) */
proc transpose data=h22_1 out=trans9(rename=(col1=cost2)); by hhid22;
var h221523 h221623 h221723 h221823 h221923; /* 사교육 비용 (2번째) */
proc transpose data=h22_1 out=trans10(rename=(col1=type3)); by hhid22;
var h221524 h221624 h221724 h221824 h221924; /* 사교육 종류 (3번째) */
proc transpose data=h22_1 out=trans11(rename=(col1=cost3)); by hhid22;
var h221527 h221627 h221727 h221827 h221927; /* 사교육 비용 (3번째) */
proc transpose data=h22_1 out=trans12(rename=(col1=type4)); by hhid22;
var h221528 h221628 h221728 h221828 h221928; /* 사교육 종류 (4번째) */
proc transpose data=h22_1 out=trans13(rename=(col1=cost4)); by hhid22;
var h221531 h221631 h221731 h221831 h221931; /* 사교육 비용 (4번째) */
proc transpose data=h22_1 out=trans14(rename=(col1=type5)); by hhid22;
var h221532 h221632 h221732 h221832 h221932; /* 사교육 종류 (5번째) */
proc transpose data=h22_1 out=trans15(rename=(col1=cost5)); by hhid22;
var h221535 h221635 h221735 h221835 h221935; /* 사교육 비용 (5번째) */
proc transpose data=h22_1 out=trans16(rename=(col1=type6)); by hhid22;
var h221536 h221636 h201736; /* 사교육 종류 (6번째) */
proc transpose data=h22_1 out=trans17(rename=(col1=cost6)); by hhid22;
var h221537 h221637 h221737; /* 사교육 비용 (6번째) */
proc transpose data=h22_1 out=trans18(rename=(col1=type7)); by hhid22;
var h221538 h221638 h221738; /* 사교육 종류 (7번째) */
proc transpose data=h22_1 out=trans19(rename=(col1=cost7)); by hhid22;
var h221539 h221639 h221739; /* 사교육 비용(7번째)*/

```

```

data edu;merge trans1 trans2 trans3 trans4 trans5 trans6 trans7 trans8 trans9 trans10
trans11 trans12 trans13 trans14 trans15 trans16 trans17 trans18 trans19; by hhid21;
drop _name_ _label_;

```

```

if chpid^=.: /*chpid결측치 처리*/
array c[7] cost1-cost7; /*사교육 비용 결측 처리*/
do i=1 to 7;
if c[i]=-1 then c[i]=.;
end;

```

```

totcost =sum(of cost1-cost7); /*한아이당 총 사교육 비용*/

```

```

array t[7] t1-t7; /*사교육이용 여부*/
array ty[7] type1-type7;
do i=1 to 7;
if ty[i]=. then t[i]=0;
if ty[i]>0 then t[i]=1;
end;

```

```

numb=sum(of t1-t7); /*한아이당 총 사교육기관 이용 갯수*/

```

```

proc univariate; var totcost;
proc freq; table numb;
run;

```

< SPSS >

\* Vector 문, Restructure 기능의 사용.

GET FILE = 'C:\klips\klips22h.sav'.

```

VECTOR pid (5) sex (5) age (5)
/V1=h220221 to h220235
/V2=h220241 to h220255
/V3=h220361 to h220375 .

```

```

LOOP #i=1 to 15.
if (h221511=#i) chpid1 = V1(#i). /*1511 자녀번호1*/
if (h221511=#i) sex1 = V2(#i).
if (h221511=#i) chage1 = V3(#i).
if (h221611=#i) chpid2 = V1(#i)./*1611 자녀번호3*/
if (h221611=#i) sex2 = V2(#i).
if (h221611=#i) chage2 = V3(#i).
if (h221711=#i) chpid3 = V1(#i)./*1811 자녀번호3*/
if (h221711=#i) sex3 = V2(#i).
if (h221711=#i) chage3 = V3(#i).
if (h221811=#i) chpid4 = V1(#i)./*1811 자녀번호4*/
if (h221811=#i) sex4 = V2(#i).
if (h221811=#i) chage4 = V3(#i).
if (h221911=#i) chpid5 = V1(#i)./*1911 자녀번호5*/
if (h221911=#i) sex5 = V2(#i).
if (h221911=#i) chage5 = V3(#i).
if (h221512=#i) mompid1 = V1(#i)./*1512 엄마번호1*/
if (h221512=#i) momage1 = V3(#i).
if (h221612=#i) mompid2 = V1(#i)./*1612 엄마번호2*/
if (h221612=#i) momage2 = V3(#i).
if (h221712=#i) mompid3 = V1(#i)./*1812 엄마번호3*/
if (h221712=#i) momage3 = V3(#i).
if (h221812=#i) mompid4 = V1(#i)./*1812 엄마번호4*/
if (h221812=#i) momage4 = V3(#i).
if (h221912=#i) mompid5 = V1(#i)./*1912 엄마번호5*/
if (h221912=#i) momage5 = V3(#i).
END LOOP.
EXECUTE.

```

```

VARSTOCASES /ID = id

```

```

/MAKE chpid FROM chpid1 chpid2 chpid3 chpid4 chpid5 /* 자녀 pid */

```

```

/MAKE mompid FROM mompid1 mompid2 mompid3 mompid4 mompid5 /* 엄마 pid */
/MAKE sex FROM sex1 sex2 sex3 sex4 sex5 /* 자녀 성별 */
/MAKE chage FROM chage1 chage2 chage3 chage4 chage5 /* 자녀 나이 */
/MAKE momage FROM momage1 momage2 momage3 momage4 momage5 /* 엄마 나이 */
/MAKE type1 FROM h221516 h221616 h221716 h221816 h221916 /* 사교육 종류 (1번째) */
/MAKE type2 FROM h221520 h221620 h221720 h221820 h221920 /* 사교육 종류 (2번째) */
/MAKE type3 FROM h221524 h221624 h221724 h221824 h221924 /* 사교육 종류 (3번째) */
/MAKE type4 FROM h221528 h221628 h221728 h221828 h221928 /* 사교육 종류 (4번째) */
/MAKE type5 FROM h221532 h221632 h221732 h221832 h221932 /* 사교육 종류 (5번째) */
/MAKE type6 FROM h221536 h221636 h221736 h221836 h221936 /* 사교육 종류 (6번째) */
/MAKE type7 FROM h221538 h221638 h221738 h221838 h221938 /* 사교육 종류 (7번째) */
/MAKE cost1 FROM h221519 h221619 h221719 h221819 h221919 /* 사교육 비용 (1번째) */
/MAKE cost2 FROM h221523 h221623 h221723 h221823 h221923 /* 사교육 비용 (2번째) */
/MAKE cost3 FROM h221527 h221627 h221727 h221827 h221927 /* 사교육 비용 (3번째) */
/MAKE cost4 FROM h221531 h221631 h221731 h221831 h221931 /* 사교육 비용 (4번째) */
/MAKE cost5 FROM h221535 h221635 h221735 h221835 h221935 /* 사교육 비용 (5번째) */
/MAKE cost6 FROM h221537 h221637 h221737 h221837 h221937 /* 사교육 비용 (6번째) */
/MAKE cost7 FROM h221539 h221639 h221739 h221839 h221939 /* 사교육 비용 (7번째) */
/INDEX = Index1(5)
/KEEP = hhid20
/NULL = KEEP.
execute.

```

```

recode cost1 to cost7(-1=sysmis). /*비용 결측처리*/
count numb=type1 type2 type3 type4 type5 type6 type7(1 thru high). /*한아이당 이용
사교육기관 갯수*/
compute totcost=sum(cost1 to cost7). /* 한 아이당 총 사교육비용 */

```

```

select if (chpid>0). /* 결측치 처리 */

```

```

des totcost/stat=mean std min max.

```

fre numb.

SAVE OUTFILE= 'C:\klips\klips22h.sav'.

<STATA>

/\* 사교육 자녀번호와 엄마번호 \*/

use klips22h

keep if hwave22==1

local j=1

foreach sex of varlist h220241-h220255 {

rename `sex' sex`j'

local ++j

}

local k=1

foreach age of varlist h220361-h220375 {

rename `age' age`k'

local ++k

}

local l=1

foreach pid of varlist h220221-h220235 {

rename `pid' pid`l'

local ++l

}

foreach n of numlist 1/5 {

gen chpid`n' =.

gen chsex`n' =.

gen chage`n' =.

gen mompid`n' =.

```

}
foreach n of numlist 1/15 {
replace chpid1=pid`n' if h221511==`n'
replace chsex1=sex`n' if h221511==`n'
replace chage1=age`n' if h221511==`n'
replace mompid1=pid`n' if h221512==`n'
}
foreach n of numlist 1/15 {
replace chpid2=pid`n' if h221611==`n'
replace chsex2=sex`n' if h221611==`n'
replace chage2=age`n' if h221611==`n'
replace mompid2=pid`n' if h221612==`n'
}
foreach n of numlist 1/15 {
replace chpid3=pid`n' if h221711==`n'
replace chsex3=sex`n' if h221711==`n'
replace chage3=age`n' if h221711==`n'
replace mompid3=pid`n' if h221712==`n'
}
foreach n of numlist 1/15 {
replace chpid4=pid`n' if h221811==`n'
replace chsex4=sex`n' if h221811==`n'
replace chage4=age`n' if h221811==`n'
replace mompid4=pid`n' if h221812==`n'
}
foreach n of numlist 1/15 {
replace chpid5=pid`n' if h221911==`n'
replace chsex5=sex`n' if h221911==`n'
replace chage5=age`n' if h221911==`n'
replace mompid5=pid`n' if h221912==`n'
}

```

```

local i=1
foreach type1 of varlist h221516 h221616 h221716 h221816 h221916 {
  rename `type1' type1`i'
  local ++i
}
local i=1
foreach type2 of varlist h221520 h221620 h221720 h221820 h221920 {
  rename `type2' type2`i'
  local ++i
}
local i=1
foreach type3 of varlist h221524 h221624 h221724 h221824 h221924 {
  rename `type3' type3`i'
  local ++i
}
local i=1
foreach type4 of varlist h221528 h221628 h221728 h221828 h221928 {
  rename `type4' type4`i'
  local ++i
}
local i=1
foreach type5 of varlist h221532 h221632 h221732 h221832 h221932 {
  rename `type5' type5`i'
  local ++i
}
local i=1
foreach type6 of varlist h221536 h221636 h221736 h221836 h221936 {
  rename `type6' type6`i'
  local ++i
}
local i=1

```



```

foreach type7 of varlist h221538 h221638 h221738 h221838 h221938 {
  rename `type7' type7`i'
  local ++i
}
local i=1
foreach cost1 of varlist h221519 h221619 h221719 h221819 h221919 {
  rename `cost1' cost1`i'
  local ++i
}
local i=1
foreach cost2 of varlist h221523 h221623 h221723 h221823 h221923 {
  rename `cost2' cost2`i'
  local ++i
}
local i=1
foreach cost3 of varlist h221527 h221627 h221727 h221827 h221927 {
  rename `cost3' cost3`i'
  local ++i
}
local i=1
foreach cost4 of varlist h221531 h221631 h221731 h221831 h221931 {
  rename `cost4' cost4`i'
  local ++i
}
local i=1
foreach cost5 of varlist h221535 h221635 h221735 h221835 h221935 {
  rename `cost5' cost5`i'
  local ++i
}
local i=1
foreach cost6 of varlist h221537 h221637 h221737 h221837 h221937 {

```

```

rename `cost6' cost6`i'
local ++i
}
local i=1
foreach cost7 of varlist h221539 h221639 h221739 h221839 h221939 {
  rename `cost7' cost7`i'
  local ++i
}
#delimit;
keep hhid22 chpid1-mompid5
type11 type12 type13 type14 type15
type21 type22 type23 type24 type25
type31 type32 type33 type34 type35
type41 type42 type43 type44 type45
type51 type52 type53 type54 type55
type61 type62 type63
type71 type72 type73
cost11 cost12 cost13 cost14 cost15
cost21 cost22 cost23 cost24 cost25
cost31 cost32 cost33 cost34 cost35
cost41 cost42 cost43 cost44 cost45
cost51 cost52 cost53 cost54 cost55
cost61 cost62 cost63
cost71 cost72 cost73;
#delimit cr

gen type64=.
gen type65=.
gen type74=.
gen type75=.
gen cost64=.

```

```
gen cost65=.
```

```
gen cost74=.
```

```
gen cost75=.
```

```
#delimit;
```

```
reshape long chpid chsex chage mompid
```

```
type1 cost1 type2 cost2 type3 cost3 type4 cost4 type5 cost5 type6 cost6 type7 cost7,
```

```
i(hhid22) ;
```

```
#delimit cr
```

```
recode cost*(-1=.)
```

```
egen totcost=rowtotal(cost*), missing /*사교육 비용*/
```

```
egen numb=anycount(type*), values(0/21) /*사교육 종류*/
```

```
drop if chpid==.
```

```
sum totcost
```

```
tab numb
```

```
save k22ch, replace
```

<R>

```
* ifelse 문의 사용 #
```

```
klips22h <- read.spss(file = "C:\\klips\\klips22h.sav",
```

```
use.value.labels = FALSE, to.data.frame = TRUE )
```

```
h22 <- filter(klips22h, h221501 == 1) #자녀 있는 경우만 추출
```

```
first.chnum <- select(h22, h221511) # 첫번째 자녀 가구원 넘버
```

```
second.chnum <- select(h22, h221611)
```

```
third.chnum <- select(h22, h221711)
```

```

fourth.chnum <- select(h22, h221811)
fifth.chnum <- select(h22, h221911)

first.momnum <- select(h22, h221512) #첫번째 자녀 어머니의 가구원 넘버
second.momnum <- select(h22, h221612)
third.momnum <- select(h22, h221712)
fourth.momnum <- select(h22, h221812)
fifth.momnum <- select(h22, h221912)

first.num <- cbind(h22$hhid22, first.chnum, first.momnum, order = c(1)) # 첫번째 자녀 및
어머니의 가구원 넘버. order변수 추가
second.num <- cbind(h22$hhid22,second.chnum, second.momnum, order = c(2))
third.num <- cbind(h22$hhid22,third.chnum, third.momnum, order = c(3))
fourth.num <- cbind(h22$hhid22,fourth.chnum, fourth.momnum, order = c(4))
fifth.num <- cbind(h22$hhid22,fifth.chnum, fifth.momnum, order =c(5))

names(first.num) <- c("hhid22", "chnum", "momnum", "order") #합치기 위해 변수명 변경
names(second.num) <- c("hhid22", "chnum", "momnum", "order")
names(third.num) <- c("hhid22", "chnum", "momnum", "order")
names(fourth.num) <- c("hhid22", "chnum", "momnum", "order")
names(fifth.num) <- c("hhid22", "chnum", "momnum", "order")

num <- rbind(first.num,second.num,third.num, fourth.num, fifth.num)#데이터 병합

#자녀의 pid변수 생성
num$child.pid <- ifelse(num$chnum==1, h22$h220221,
                        ifelse(num$chnum==2,h22$h220222,
                                ifelse(num$chnum==3,h22$h220223,
                                        ifelse(num$chnum==4,h22$h220224,
                                                ifelse(num$chnum==5,h22$h220225,

```

```

        ifelse(num$chnum==6,h22$h220226,
              ifelse(num$chnum==7,h22$h220227,
                    ifelse(num$chnum==8,h22$h220228,
                          ifelse(num$chnum==9,h22$h220229,
                                ifelse(num$chnum==10,h22$h220230,

ifelse(num$chnum==11,h22$h220231,

ifelse(num$chnum==12,h22$h220232,

ifelse(num$chnum==13,h22$h220233,

ifelse(num$chnum==14,h22$h220234,

ifelse(num$chnum==15,h22$h220235,NA)))))))))))))

```

#자녀의 성별 변수 생성

```

num$child.sex <- ifelse(num$chnum==1, h22$h220241,
                      ifelse(num$chnum==2,h22$h220242,
                            ifelse(num$chnum==3,h22$h220243,
                                  ifelse(num$chnum==4,h22$h220244,
                                        ifelse(num$chnum==5,h22$h220245,
                                              ifelse(num$chnum==6,h22$h220246,
                                                    ifelse(num$chnum==7,h22$h220247,
                                                          ifelse(num$chnum==8,h22$h220248,
                                                                ifelse(num$chnum==9,h22$h220249,
                                                                      ifelse(num$chnum==10,h22$h220250,

ifelse(num$chnum==11,h22$h220251,

ifelse(num$chnum==12,h22$h220252,

```

```
ifelse(num$chnum==13,h22$h220253,
```

```
ifelse(num$chnum==14,h22$h220254,
```

```
ifelse(num$chnum==15,h22$h220255,NA)))))))))))))
```

#자녀의 나이 변수 생성

```
num$child.age <- ifelse(num$chnum==1, h22$h220361,
```

```
ifelse(num$chnum==2,h22$h220362,
```

```
ifelse(num$chnum==3,h22$h220363,
```

```
ifelse(num$chnum==4,h22$h220364,
```

```
ifelse(num$chnum==5,h22$h220365,
```

```
ifelse(num$chnum==6,h21$h220366,
```

```
ifelse(num$chnum==7,h21$h220367,
```

```
ifelse(num$chnum==8,h21$h220368,
```

```
ifelse(num$chnum==9,h21$h220369,
```

```
ifelse(num$chnum==10,h21$h220370,
```

```
ifelse(num$chnum==11,h21$h220371,
```

```
ifelse(num$chnum==12,h21$h220372,
```

```
ifelse(num$chnum==13,h21$h220373,
```

```
ifelse(num$chnum==14,h21$h220374,
```

```
ifelse(num$chnum==15,h21$h220375,NA)))))))))))))
```

#어머니의 pid 변수 생성

```

num$mom.pid <- ifelse(num$momnum==1, h22$h220221,
  ifelse(num$momnum==2,h22$h220222,
    ifelse(num$momnum==3,h22$h220223,
      ifelse(num$momnum==4,h22$h220224,
        ifelse(num$momnum==5,h22$h220225,
          ifelse(num$momnum==6,h22$h220226,
            ifelse(num$momnum==7,h22$h220227,
              ifelse(num$momnum==8,h22$h220228,
                ifelse(num$momnum==9,h22$h220229,
                  ifelse(num$momnum==10,h22$h220230,

```

```

ifelse(num$momnum==11,h22$h220231,

```

```

ifelse(num$momnum==12,h22$h220232,

```

```

ifelse(num$momnum==13,h22$h220233,

```

```

ifelse(num$momnum==14,h22$h220234,

```

```

ifelse(num$momnum==15,h22$h220235,NA)))))))))))))

```

#어머니의 나이 변수 생성

```

num$mom.age <- ifelse(num$momnum==1, h22$h220361,
  ifelse(num$momnum==2,h22$h220362,
    ifelse(num$momnum==3,h22$h220363,
      ifelse(num$momnum==4,h22$h220364,
        ifelse(num$momnum==5,h22$h220365,
          ifelse(num$momnum==6,h22$h220366,
            ifelse(num$momnum==7,h22$h220367,
              ifelse(num$momnum==8,h22$h220368,

```

```

                                ifelse(num$momnum==9,h22$h220369,
                                ifelse(num$momnum==10,h22$h220370,

ifelse(num$momnum==11,h22$h220371,

ifelse(num$momnum==12,h22$h220372,

ifelse(num$momnum==13,h22$h220373,

ifelse(num$momnum==14,h22$h220374,

ifelse(num$momnum==15,h22$h220375,NA)))))))))))))

```

###사교육 데이터 정리

# 1번째 사교육 종류 및 비용

```

num$edu1 <- ifelse(num$order == 1, h22$h221516,
                  ifelse(num$order ==2, h22$h221616,
                  ifelse(num$order ==3, h22$h221716,
                  ifelse(num$order ==4, h22$h221816,
                  ifelse(num$order == 5, h22$h221916,NA )))))
num$educost1 <- ifelse(num$order ==1, h22$h221519,
                  ifelse(num$order == 2, h22$h221619,
                  ifelse(num$order ==3, h22$h221719,
                  ifelse(num$order ==4, h22$h221819,
                  ifelse(num$order ==5, h22$h221919, NA))))))

```

# 2번째 사교육 종류 및 비용

```

num$edu2 <- ifelse(num$order == 1, h22$h221520,
                  ifelse(num$order ==2, h22$h221620,

```



```

        ifelse(num$order ==3, h22$h221720,
              ifelse(num$order ==4, h22$h221820,
                    ifelse(num$order == 5, h22$h221920,NA )))))
num$educost2 <- ifelse(num$order ==1, h22$h221523,
                    ifelse(num$order == 2, h22$h221623,
                          ifelse(num$order ==3, h22$h221723,
                                ifelse(num$order ==4, h22$h221823,
                                      ifelse(num$order ==5, h22$h221923, NA))))))

```

# 3번째 사교육 종류 및 비용

```

num$edu3 <- ifelse(num$order == 1, h22$h221524,
                  ifelse(num$order ==2, h22$h221624,
                        ifelse(num$order ==3, h22$h221724,
                              ifelse(num$order ==4, h22$h221824,
                                    ifelse(num$order == 5, h22$h221924,NA )))))
num$educost3 <- ifelse(num$order ==1, h22$h221527,
                    ifelse(num$order == 2, h22$h221627,
                          ifelse(num$order ==3, h22$h221727,
                                ifelse(num$order ==4, h22$h221827,
                                      ifelse(num$order ==5, h22$h221927, NA))))))

```

# 4번째 사교육 종류 및 비용

```

num$edu4 <- ifelse(num$order == 1, h22$h221528,
                  ifelse(num$order ==2, h22$h221628,
                        ifelse(num$order ==3, h22$h221728,
                              ifelse(num$order ==4, h22$h221828,
                                    ifelse(num$order == 5, h22$h221928,NA )))))

```

```

num$educost4 <- ifelse(num$order ==1, h22$h221531,
                      ifelse(num$order == 2, h22$h221631,
                            ifelse(num$order ==3, h22$h221731,
                                  ifelse(num$order ==4, h22$h221831,
                                        ifelse(num$order ==5, h22$h221931, NA))))))

```

# 5번째 사교육 종류 및 비용

```

num$edu5 <- ifelse(num$order == 1, h22$h221532,
                  ifelse(num$order ==2, h22$h221632,
                        ifelse(num$order ==3, h22$h221732,
                              ifelse(num$order ==4, h22$h221832,
                                    ifelse(num$order == 5, h22$h221932,NA )))))
num$educost5 <- ifelse(num$order ==1, h22$h221535,
                      ifelse(num$order == 2, h22$h221635,
                            ifelse(num$order ==3, h22$h221735,
                                  ifelse(num$order ==4, h22$h221835,
                                        ifelse(num$order ==5, h22$h221935, NA))))))

```

# 6번째 사교육 종류 및 비용

```

num$edu6 <- ifelse(num$order == 1, h22$h221536,
                  ifelse(num$order ==2, h22$h221636,
                        ifelse(num$order ==3, h22$h221736,
                              ifelse(num$order ==4, h22$h221836,
                                    ifelse(num$order == 5, h22$h221936,NA )))))
num$educost6 <- ifelse(num$order ==1, h22$h221537,
                      ifelse(num$order == 2, h22$h221637,
                            ifelse(num$order ==3, h22$h221737,

```

```

        ifelse(num$order ==4, h22$h221837,
              ifelse(num$order ==5, h22$h221937, NA))))

```

# 7번째 사교육 종류 및 비용

```

num$edu7 <- ifelse(num$order == 1, h22$h221538,
                  ifelse(num$order ==2, h22$h221638,
                        ifelse(num$order ==3, h22$h221738,
                              ifelse(num$order ==4, h22$h221838,
                                    ifelse(num$order == 5, h22$h221938,NA )))))
num$educost7 <- ifelse(num$order ==1, h22$h221539,
                      ifelse(num$order == 2, h22$h221639,
                            ifelse(num$order ==3, h22$h221739,
                                  ifelse(num$order ==4, h22$h221839,
                                        ifelse(num$order ==5, h22$h221939, NA))))))

```

# 자녀별 사교육 총비용

```

num$totalcost <-as.numeric((ifelse(is.na(num$child.pid), NA,
                                   (ifelse(is.na(num$edu1), NA,
                                             (rowSums(num[
c("educost1","educost2","educost3","educost4","educost5","educost6","educost7"),
na.rm=TRUE))
))))))

```

# 자녀별 사교육 이용 여부

```

num$edu <- ifelse((is.na(num$chnum)) & (is.na(num$edu1)) & (is.na(num$edu2)) &
(is.na(num$edu3)) & (is.na(num$edu4)) & (is.na(num$edu5)) & (is.na(num$edu6)) &
(is.na(num$edu7)), "NA",
                  ifelse((is.na(num$edu1)) & (is.na(num$edu2)) & (is.na(num$edu3)) &

```

```
(is.na(num$edu4))      &      (is.na(num$edu5))      &      (is.na(num$edu6))      &
(is.na(num$edu7)), "No", "Yes"))
```

# 자녀별 사교육 이용 갯수

```
num$edu1[num$edu1==0] <- NA #표기 오류 결측값 수정
```

```
num$edu2[num$edu2==0] <- NA
```

```
num$edu3[num$edu3==0] <- NA
```

```
num$edu4[num$edu4==0] <- NA
```

```
num$edu5[num$edu5==0] <- NA
```

```
num$edu6[num$edu6==0] <- NA
```

```
num$edu7[num$edu7==0] <- NA
```

```
k      <-      (as.numeric(is.na(num$edu1))      +      as.numeric(is.na(num$edu2))      +
as.numeric(is.na(num$edu3))
      +      as.numeric(is.na(num$edu4))      +      as.numeric(is.na(num$edu5))      +
as.numeric(is.na(num$edu6))
      + as.numeric(is.na(num$edu7))) # 임시 변수 k 생성
```

```
num$numedu <- ifelse(is.na(num$chnum), NA, (c(7)-k)) # 리코딩하여 변수 생성
```

#가구별로 정렬된 자녀와 어머니의 가구 번호. order는 몇번째 자녀인지를 나타냄.

```
num.hhid22 <- arrange(num, hhid22)
```

```
head(num.hhid22)
```

# 순서대로 가구번호, 자녀의 가구원번호, 어머니의 가구원 번호, 자녀 차례, 자녀의pid, 자녀의 성  
별,

# 자녀의 나이, 어머니의 pid, 어머니의 나이, 7개의 사교육 종류와 비용, 자녀별 총 사교육 비용, 사  
교육 여부, 사교육 이용 갯수를

# 나타낸다.

```
summary(num.hhid22$totalcost)
```

```
table(num.hhid22$numedu)
```

# Q 13

## 사교육 이용기관의 변동사항 : 연도별 차이

노동패널 가구용 자료의 사교육에 관한 변수들을 해마다 조금씩 차이가 난다. 특히 사교육 이용기관은 해마다 사회적 변화를 적당히 반영하기 위하여, 그 항목이 변경·추가되어 왔으므로, 이용에 주의를 요한다. 사교육 이용기관과 관련된 변수의 변화는 아래의 표와 같다. 사교육기관의 항목은 그 숫자도 많은데다가 연도별로 계속 항목변동이 이루어져 왔다. 또 어린이방(어린이집)과 국공립 어린이집·민간어린이집, 놀이방과 사설놀이방, 5차부터 사라진 탁아소와 유아원 등의 항목 간 구분이 명확하게 이루어지지 않는 상태이다. 따라서 연구자가 아래의 항목들을 연구의 목적에 맞추어 사용해야 한다(관련항목 : V. 주제별 가이드, <표 5-9> 참조).

		1차	2차	3차	4차	5차	6차	7차	8차 이후
주요 특징	사교육관련 문항의 유무	○	×	○	○	○	○	○	○
	'탁아모' 관련 문항의 유무	○	×	○	○	×	×	×	×
이용 기관	탁아소	○	×	○	○	×	×	×	×
	어린이방 (어린이집)	○	×	○	○	○	○	○	○
						○	○	○	○
	국공립(국립·사립·구립등) 어린이집 민간 어린이집								
	놀이방 (혹은 사설놀이방)	○	×	○	○	○	○	○	○
	유아원	○	×	○	○	×	×	×	×
	직장보육시설	○	×	○	○	○	○	○	○
	시간제유치원(:정규시간외에보육을 맡아주는 경우1)	○	×	○	○	○	○	○	○
	종일제유치원(:정규시간외에보육을 맡아주는 경우1)							○	○
	정규유치원(:정규시간만을 담당하는 경우)	○	×	○	○	○	○	○	○
	(보육/교육기관) 기타	×	×	○	○	○	○	○	○
	학원	○	×	○	○	○	○	○	○
		○							
	보습학원								
	예체능학원								
	개인·그룹과의	×	×	○	○	○	○	○	○
	학습지	×	×	○	○	○	○	○	○
	방과후교내보충학습	×	×	○	○	○	○	○	○
	방과후교실	×	×	○	○	○	○	○	○
	동거하고있는 가족/친지	○	×	○	○	×	×	○	○
	비동거하고있는 가족/친지	○	×	○	○	×	×	○	○
	동거하는 타인	○	×	○	○	×	×	○	○
	비동거하는 타인	○	×	○	○	×	×	○	○
	(탁아모 관련) 기타	○	×	○	○	×	×	×	○
	어학연수	×	×	×	×	×	×	○	○
	인터넷(온라인 교육) 유료강좌	×	×	×	×	×	×	○	○
	문화센터	×	×	×	×	×	×	○	○
항목수		14	0	18	18	12	12	19	20

주: 1차에는 놀이방으로 조사되었으나, 5차부터는 놀이방이 삭제되고 사설놀이방으로 변경

## Q 14

### 가구소득을 구성하는 방법

#### SAS Array문 활용 예 : 가구총소득 구성하기

노동패널 가구용 자료에서 조사되는 가구소득 항목은 ① 근로소득 ② 금융소득 ③ 부동산 소득 ④ 사회보험 수혜금 ⑤ 이전소득 ⑥ 기타소득 이다.

여기서 주의할 점은 가구용 자료에 있는 소득들은 이미 가구 단위로 합산이 되어 있는 값이라는 점이다. 예를 들어 가구용 자료의 근로소득이란 근로소득이 있는 모든 가구원의 근로소득을 합한 값을 의미한다. 만일 연구자가 가구원 각각의 소득을 알고 싶다면, 개인용 자료에서 가구원 각각의 근로소득을 추출하여야 한다.

만일 가구의 총소득을 구하고자 할 때에는 6가지 항목별 소득을 모두 합산하면 된다. 다만, 근로소득을 제외한 다른 소득들은 다시 세분하여 질문하고 있으므로, 이들을 각각 합산하여야 한다. 예컨대 금융소득의 경우 ㉠ 은행 등 금융기관 이자/투자소득 ㉡ 사채 등 비금융기관 이자수입 ㉢ 주식/채권 매매차익 ㉣ 배당금 ㉤ 기타 로 구분되어 있으므로, 가구의 금융소득은 이들을 모두 합한 값이 된다.

가구소득의 세부 항목은 통계청 “1차 가계수지항목분류<sup>37)</sup>”와 차이가 있다. 통계청 “1차 가계수지항목분류”에 따르면 가계수지는 경상소득, 비경상소득, 기타 수입, 소비지출, 비소비지출, 기타 지출, 자산, 부채 등 8개 대분류 체계로 구성되어있다. 반면 노동패널조사에서는 해당 분류체제로 조사되지 않았기 때문에 통계청 가계소득 관련 조사와 교·연구를 위해서는 몇몇 항목에 주의해야하는데 가장 차이나는 것은 다음 두 가지이다. 첫째, 노동패널조사의 금융소득 항목 중 “이자소득”과 “배당금” 등은 통계청 기준에서 경상소득의 재산소득에 포함되지만 “주식 및 매매차익”은 기타소득에 포함된다. 둘째, “부동산 매매차익”은 노동패널조사에서 부동산소득 항목에 포함되지만 통계청 분류기준에 따르면 기타 수입으로 자산변동으로 인한 수입에 해당된다.

37) 가계수지항목분류 제1차 개정결과 보고, 통계청(2020. 12)

## < SAS >

\* 가구 총소득 구성하기: Array 문 활용하기;

/\*13차부터 제도변화에 따라 근로장려세제를 추가로 질문하였고, 19차부터는 자녀장려금 지원금을 추가로 질문하였음. 총가구소득시에는 이전소득으로 포함하여 계산함\*/

```
libname a 'C:\klips'
```

```
data h22; set a.klips22h;
```

```
/*결측치 처리*/
```

```
array h[*] h222102 /*작년 한해 총 근로소득*/
```

```
h222112-h222116 /*작년 한해 총 금융소득*/
```

```
h222122-h222126 /*작년 한해 총 부동산소득*/
```

```
h222134 h222136 h222138 h222140 h222142 /*작년 한해 사회보험소득*/
```

```
h222152 h222153 h222155 h222157-h222160 /*작년 한해 총 이전소득*/
```

```
h222183-h222191 /*작년 한해 총 기타소득*/
```

```
h224002 h224004; /*작년 한해 총 근로장려세제*/
```

```
do i=1 to dim(h);
```

```
if h[i]=-1 then h[i]=.;
```

```
end;
```

```
/*항목별 소득총계*/
```

```
inc_e=h222102; /*근로소득*/
```

```
inc_m=sum(of h222112-h222116); /*금융소득*/
```

```
inc_p=sum(of h222122-h222126); /*부동산소득*/
```

```
inc_i=sum(of h222134, h222136, h222138, h222140, h222142); /*사회보험소득*/
```

```
inc_t=sum(of h222152 h222153 h222155 h222157-h222160 h224002 h224004); /*이전소*/
```

```
inc_o=sum(of h222183-h222191); /*기타소득*/
```

```
/*소득유무*/
```

```

inc_ey=h222101;
inc_my=h222111;
inc_py=h222121;
inc_iy=h222131;
inc_ty=0;

if h222151=1 or h224001=1 then inc_ty=1;
if h222151=2 and h224001=2 then inc_ty=2;
if inc_ty=0 then inc_ty=.;
inc_oy=h222181;

/*가구 총소득*/
inc=sum(of inc_e, inc_m, inc_p, inc_i, inc_t, inc_o);

proc freq; table inc_ey inc_my inc_py inc_iy inc_ty inc_oy;
proc univariate; var inc;
run;

```

< SPSS >

\* 가구 총소득 구성하기.

GET FILE='C:\klips\klips22h.sav'.

/\*13차부터 제도변화에 따라 근로장려세제를 추가로 질문하였고, 19차부터는 자녀장려금 지원금을 추가로 질문하였음. 총가구소득시에는 이전소득으로 포함하여 계산함\*/

/\* 결측치 처리 (transform → recode into same variables 메뉴 사용) \*/

RECODE

h222102 h222112 h222113 h222115 h222116 h222122 to h222126 h222134 h222136



```

h222138 h222140 h222142
h222152 h222153 h222155 h222157 to h222160 h222183 to h222191 h224002
h224004(-1=SYSMIS) .
EXECUTE.

```

```

/*항목별 소득총계*/
compute inc_e=h222102. /*근로소득*/
compute inc_m=sum(h222112 to h222116). /*금융소득*/
compute inc_p=sum(h222122 to h222126). /*부동산소득*/
compute inc_i=sum(h222134, h222136, h222138, h222140, h222142). /*사회보험소득*/
compute inc_t=sum(h222152, h222153, h222155, h222157 to h222160, h224002, h224004).
/*이전소득*/
compute inc_o=sum(h222183 to h222191). /*기타소득*/

```

```

/*소득유무*/
compute inc_ey=h222101.
compute inc_my=h222111.
compute inc_py=h222121.
compute inc_iy=h222131.
compute inc_ty=0.
if(h222151=1 or h224001=1) inc_ty=1.
if(h222151=2 and h224001=2) inc_ty=2.
recode inc_ty (0=sysmis).
compute inc_oy=h222181.

```

```

/*가구 총소득*/
compute inc=sum(inc_e, inc_m, inc_p, inc_i, inc_t, inc_o).
execute.

```

```

fre inc_ey inc_my inc_py inc_iy inc_ty inc_oy.
des inc/stat=mean std min max.

```

## < STATA >

/\* 가구 총소득 구성하기 \*/

/\* 13차부터 제도변화에 따라 근로장려세제를 추가로 질문하였고, 19차부터는 자녀장려금 지원금을 추가로 질문하였음. 총가구소득시에는 이전소득으로 포함하여 계산함\*/

/\*결측치 처리\*/

use klips22h

keep if hwave22==1

#delimit;

recode h222102 h222112-h222116 h222122-h222126

h222134 h222136 h222138 h222140 h222142

h222152 h222153 h222155 h222157-h222160 h222183-h222191 h224002 (-1=.);

#delimit cr

/\*항목별 소득총계\*/

gen inc\_e=h222102 /\*근로소득\*/

egen inc\_m=rowtotal(h222112-h222116)/\*금융소득\*/

egen inc\_p=rowtotal(h222122-h222126)/\*부동산소득\*/

egen inc\_i=rowtotal(h222134 h222136 h222138 h222140 h222142) /\*사회보험소득\*/

egen inc\_t=rowtotal(h222152 h222153 h222155 h222157-h222160 h224002 h224004) /\*이전소득\*/

egen inc\_o=rowtotal(h222183-h222191) /\*기타소득\*/

/\*소득유무\*/

gen inc\_ey=h222101

gen inc\_my=h222111

gen inc\_py=h222121

gen inc\_iy=h222131

```

gen inc_ty=1 if h222151==1| h224001==1
replace inc_ty=2 if h222151==2 & h224001==2
gen inc_oy=h222181

/*가구 총소득*/
egen inc=rowtotal(inc_e inc_m inc_p inc_i inc_t inc_o),mis

tab1 inc_ey inc_my inc_py inc_iy inc_ty inc_oy
sum inc

```

<R>

```

# 가구 총소득 구성하기 #

#결측치 처리
library(foreign)
h22 <- read.spss(file = "C:\\klips\\klips22h.sav",
use.value.labels = FALSE, to.data.frame = TRUE )
h22[h22== (-1)] <- NA
install.packages("dplyr")
library(dplyr)
# 소득 계산
inc_e <- h22[, "h222102"] # 근로소득
m <- select(h22,h222112:h222116)
inc_m <- rowSums(m , na.rm = TRUE) # 금융소득
p <- select(h22,h222122:h222126)
inc_p <- rowSums(p , na.rm = TRUE) # 부동산소득
i <- select(h22,h222134,h222136,h222138,h222140,h222142)
inc_i <- rowSums(i , na.rm = TRUE) # 사회보험소득
t <- select(h22,h222152,h222153,h222155, h222157:h222160,h224002,h224004)
inc_t <- rowSums(t, na.rm = TRUE) # 이전소득

```

```

o <- select(h22,h222183:h222191)
inc_o <- rowSums(o, na.rm = TRUE) # 기타소득
inc_table <- data.frame(inc_e,inc_m,inc_p,inc_i,inc_t,inc_o)
inc_table # 하나의 데이터로 병합
inc_table$inc <- rowSums(inc_table[,1:6], na.rm=TRUE) # 연간 총소득
inc_table$inc <- ifelse(inc_table$inc ==0 , NA, inc_table$inc)
inc_table

# 소득 유무
inc_ey <- select(h22,h222101)
inc_my <- select(h22,h222111)
inc_py <- select(h22,h222121)
inc_iy <- select(h22,h222131)
inc_ty <- ifelse((h22$h222151==1 | h22$h224001 ==1), 1,
                ifelse((h22$h222151==2 & h22$h224001 ==2), 2, 0))
inc_oy <- select(h22,h222181)

inc_tabley <- data.frame(inc_ey,inc_my,inc_py,inc_iy,inc_ty,inc_oy)
names(inc_tabley) <- c("inc_ey","inc_my","inc_py","inc_iy","inc_ty","inc_oy")
inc_tabley #소득 유무에 대한 데이터프레임으로 병합
inc_freqy <- data.frame(table(inc_ey) ,table(inc_my) ,table(inc_py), table(inc_iy)
,table(inc_ty) , table(inc_oy))
inc_freqy #각 소득 유무 변수별 빈도수

summary(inc_table$inc) # 가구 총소득 관련 통계치
sd(inc_table$inc, na.rm=TRUE)

```

## Q 15

### 연평균가구소득과 지난 한 달간 소득의 차이점

노동패널 가구용 조사에서는 지난해 연평균 가구소득과 지난 한달간 소득을 구분하여 질문하고 있다. ‘연평균가구소득’과 ‘지난한달간 소득’은 말 그대로 가구 전체의 소득을 의미하며 단지 시점이 다를 뿐이다.

그렇다면 왜 동일한 소득항목을 두 가지 방식으로 질문하는 것일까. 예컨대 어떤 가구가 2007년에 분가하였다고 하자. 혹은 지난해 가구내 모든 구성원들이 실직상태였다가 올해부터 일을 시작한 가구를 생각할 수도 있다. 이런 가구들은 지난해에는 소득이 없었으나, 지난 한달 간 소득은 존재하는 가구들이다. 반대로 지난해에는 정상적인 가구소득이 관찰되었지만 두 달 전에 실직한 가구원이 있다면, 지난해 연평균가구소득을 월평균으로 바꾼 값이 지난 한 달간 소득에 비해 클 수도 있는 것이다.

한편 분석상의 시의성이라는 측면도 고려할 수 있다. 노동패널조사는 연단위로 행해지는 조사이며, 조사가 종료된 후에도 연구자들이 자료를 이용할 수 있는 시점까지는 상당기간이 소요된다. 따라서 ‘지난해 연평균가구소득만’을 질문할 경우, 가장 최근 자료를 사용한다하더라도 2년 전의 연평균가구소득을 사용하게 되는 것이다. (예: 2020도 release data 이용시 → 22차년도 가구용 자료 이용 가능(2019년에 조사한 데이터) → 2018년도 한 해 동안의 연평균 가구의 근로소득이 파악됨). 그러나 ‘지난 한 달간 소득’은 소득 추정기간이 짧아, 지난한달 동안만 가구소득이 유난히 높았을(혹은 낮았을) 경우를 통제할 수 없다는 문제점이 있다. 또한, 연평균가구소득이 각각의 소득에 대해 보다 세부적 항목 단위로 소득파악이 가능한 반면, 지난 한 달간 소득의 경우에는 각각의 소득 총액만이 파악 가능하다.

그러므로 연구자가 연구의 목적에 맞추어 [지난한해 연평균 가구소득]이나 [지난 한 달간 소득] 중 하나를 선택하여 사용하면 된다. 1-22차년도 가구소득 관련변수에 대한 자세한 사항은 본문 V. 주제별 가이드, 1.2.2 가구소득 부분을 참고하기 바란다.

## Q 16

### 생활비 항목에 대한 설명

생활비는 지난 한 해 동안의 가구전체가 지출한 생활비 총액을 월단위로 환산한 것이다. 가구자료에서 생활비란 식비, 주거비, 의류비와 같이 생활하는데 드는 돈만을 의미하므로, 가구 전체의 소비수준과는 다소 다른 개념이 될 수 있음을 주의하기 바란다. 즉, 혼수 장만이 나 교통사고 보상금 지불과 같이 특별한 일로 인해 지불한 돈의 액수는 파악이 불가능하다. 다만 저축과 부채에 대한 원금상환액 등은 별도의 문항으로 조사하고 있으므로 관련 설문을 참조하기 바란다.

생활비 항목은 1차년도와 3차년도에는 조사되지 않았고, 매해 항목이 수정되어 왔으며 21차년도에는 총 20개 항목에 대해 조사되었다(본문 V. 주제별 가이드; 1.2.3 가구소비; <표 V-4> 1-22차년도 가구소비 관련 변수 참조).

또한 생활비의 항목별 액수 뿐 아니라 전체 월평균 생활비 액수 또한 묻고 있으므로, 생활비 총액만 필요할 경우 ‘생활비 총액’ 변수(22차년도 : h222301)만을 사용해도 된다. 다만 ‘생활비 항목별 액수의 총합’과 ‘생활비 총액’과는 다소 다를 수 있다. 왜냐하면 실사과정에서 면접원들에게 두 항목이 대략적인 일치는 권고하지만 이를 강제적 지침으로 규정할 경우 오히려 편향된 응답을 얻을 수 있다는 판단을 내렸기 때문이다.

## Q 17

### 금융자산과 부채 변수 구성하는 방법 - 자산과 부채액 구성하기

#### 1. 금융 자산

금융자산과 금융소득을 혼동하는 연구자들이 종종 있다. 우선 용어의 정의를 살펴보면, ‘금융자산’이란 금융기관 혹은 개인적으로 빌려준 돈과 같이 가지고 있는 돈의 총액을 의미하는 저량(stock) 변수인 반면, ‘금융소득’은 이러한 자산을 통해 발생한 소득(예: 이자소득, 배당금)을 의미하는 유량(flow) 변수이다.

특히 금융자산과 관련하여 실제 가구용 설문에서는 금융자산의 소유여부를 각 항목별로 질문하여 [예]라고 응답한 경우 총액을 질문한다. 그러므로, 연구자가 가구의 ‘금융자산’ 전체 액수를 필요로 한다면 각 항목에 대한 총액을 모두 합산해서 사용해야 한다(※ 주의: 조사시점을 기준으로 현재 시세).

#### 2. 가구 부채

‘가구부채’는 ‘금융자산’과 유사한 구조로 조사되고 있다. 즉 항목별로 부채유무를 질문한 후, [예]라고 응답한 경우 현재 상환해야 할 잔액을 묻는다. 따라서, 현재 가구에 남아있는 부채의 총액은 각 항목의 잔액을 모두 합산하면 된다. 다만, 부채항목에는 이외에도 ‘현재까지 상환한 원금과 이자’의 액수를 각각의 항목에 대해 묻고 있으므로, ‘잔액’ 및 ‘원금과 이자상환금’을 합산할 경우 처음 돈을 빌릴 당시의 부채액도 구할 수 있다.

## <SAS>

\* 가구 자산과 부채 구성하기 ;

```
libname a 'C:\klips'
```

```
data h22; set a.klips22h;
```

```
/*금융자산*/
```

```
/*결측치 처리*/
```

```
array h[6] h222562 h222564 h222566 h222568 h222570 h222572;
```

```
do i=1 to 6;
```

```
if h[i]=-1 then h[i]= .;
```

```
end;
```

```
if h222561=1 or h222563=1 or h222565=1 or h222567=1 or h222569=1 or h222571=1 then
```

```
fin_y=1; /*금융자산 여부*/
```

```
fin1 = h222562; /* 은행예금 액수*/
```

```
fin2 = h222564; /* 주식, 채권, 신탁 액수*/
```

```
fin3 = h222566; /* 저축성보험 액수*/
```

```
fin4 = h222568; /* 아직 타지 않은 계 액수*/
```

```
fin5 = h222570; /* 개인적으로 다른 사람에게 빌려준 돈 액수*/
```

```
fin6 = h222572; /* 기타 금융자산 액수*/
```

```
fin = sum(of fin1-fin6); /* 금융자산 총액*/
```

```
/*부채*/
```

```
/*결측치 처리*/
```

```
array d[6] h222602 h222605 h222608 h222611 h222614 h222617;
```

```
do i=1 to 6;
```

```
if d[i]=-1 then d[i]= .;
```

```
end;
```



```

if h222601=1 or h222604=1 or h222607=1 or h222610=1 or h222613=1 or h222616=1 then
deb_y=1; /*부채여부*/
deb1 = h222602; /*금융기관부채*/
deb2 = h222605; /*비금융기관 부채*/
deb3 = h222608; /*개인적으로 빌린 돈*/
deb4 = h222611; /*전세금, 임대보증금 받은 것*/
deb5 = h222614; /*머리 타고 앞으로 부어야 할 계*/
deb6 = h222617; /*기타 부채*/
deb= sum(of deb1-deb6); /*부채 총액*/

proc freq; table fin_y deb_y;
proc univariate; var fin deb;
run;

```

< SPSS >

\* 가구 자산과 부채 구성하기.

GET FILE='C:\klips\klips22h.sav'.

\*금융자산\*

```

RECODE h222562 h222564 h222566 h222568 h222570 h222572 (-1=SYSMIS) ./*결측치 처리
*/
if (h222561=1 or h222563=1 or h222565=1 or h222567=1 or h222569=1 or h222571=1)
fin_y=1./*금융자산 여부*/
compute fin1 = h222562. /* 은행예금 액수*/
compute fin2 = h222564. /* 주식, 채권, 신탁 액수*/
compute fin3 = h222566. /* 저축정보험 액수*/
compute fin4 = h222568. /* 아직 타지 않은 계 액수*/
compute fin5 = h222570. /* 개인적으로 다른 사람에게 빌려준 돈 액수*/

```

```

compute fin6 = h222572. /* 기타 금융자산 액수*/
compute fin = sum(fin1 to fin6). /* 금융자산 총액*/

/*부채*/
RECODE h222602 h222605 h222608 h222611 h222614 h222617 (-1=SYSMIS) ./*결측치 처리
*/
if (h222601=1 or h222604=1 or h222607=1 or h222610=1 or h222613=1 or h222616=1)
deb_y=1./*부채 여부*/
compute deb1 = h222602. /*금융기관부채*/
compute deb2 = h222605. /*비금융기관 부채*/
compute deb3 = h222608. /*개인적으로 빌린 돈*/
compute deb4 = h222611. /*전세금, 임대보증금 받은 것*/
compute deb5 = h222614. /*미리 타고 앞으로 부여야 할 계*/
compute deb6 = h222617. /*기타 부채*/
compute deb=sum(deb1 to deb6)./*부채액*/
execute.

fre fin_y deb_y.
des fin deb/stat=mean std min max.

```

## < STATA >

```
/* 가구 자산과 부채 구성하기 */
```

```
use klips22h
```

```
/*금융자산*/
```

```

recode h222562 h222564 h222566 h222568 h222570 h222572 (-1=.)/*결측치 처리*/
gen fin_y=1 if h222561==1 | h222563==1 | h222565==1 | h222567==1 | h222569==1 |
h222571==1 /*금융자산 여부*/

```

```

gen fin1 = h222562 /* 은행예금 액수*/
gen fin2 = h222564 /* 주식, 채권, 신탁 액수*/
gen fin3 = h222566 /* 저축성보험 액수*/
gen fin4 = h222568 /* 아직 타지 않은 계 액수*/
gen fin5 = h222570 /* 개인적으로 다른 사람에게 빌려준 돈 액수*/
gen fin6 = h222572 /* 기타 금융자산 액수*/
egen fin = rowtotal(fin1-fin6), m /* 금융자산 총액*/

/*부채*/
recode h222602 h222605 h222608 h222611 h222614 h222617(-1=.) /*결측치 처리*/
gen deb_y=1 if h222601==1 | h222604==1 | h222607==1 | h222610==1 | h222613==1 |
h222616==1 /*부채 여부*/
gen deb1 = h222602 /*금융기관부채*/
gen deb2 = h222605 /*비금융기관 부채*/
gen deb3 = h222608 /*개인적으로 빌린 돈*/
gen deb4 = h222611 /*전세금, 임대보증금 받은 것*/
gen deb5 = h222614 /*미리 타고 앞으로 부어야 할 계*/
gen deb6 = h222617 /*기타 부채*/
egen deb=rowtotal(deb1-deb6), m/*부채액*/

tab1 fin_y deb_y
sum fin deb

```

<R>

```

# 가구 자산과 부채 구성하기 #

library(foreign)
h22 <- read.spss(file = "C:\\klips\\klips22h.sav",
use.value.labels = FALSE, to.data.frame = TRUE )

```

```
##### 금융 자산 #####
```

```
#결측값 처리
```

```
h22[, c("h222562","h222564","h222566", "h222568", "h222570", "h222572")][h22[,  
c("h222562","h222564","h222566", "h222568", "h222570", "h222572")] == (-1)] <- NA
```

```
#금융자산 보유 여부 (1: 보유)
```

```
fin_y <- ifelse((h22$h222561 ==1 | h22$h222563 ==1 | h22$h222565 ==1 | h22$h222567 ==1  
| h22$h222569 ==1 | h22$h222571 ==1), 1,NA)
```

```
#금융 자산 액수
```

```
fin1 <- h22$h222562 # 은행 예금 액수
```

```
fin2 <- h22$h222564 # 주식, 채권, 신탁 액수
```

```
fin3 <- h22$h222566 # 저축성보험 액수
```

```
fin4 <- h22$h222568 # 아직 타지 않은 게 액수
```

```
fin5 <- h22$h222570 # 개인적으로 다른 사람에게 빌려준 돈 액수
```

```
fin6 <- h22$h222572 # 기타 금융자산 액수
```

```
fintable <- data.frame(fin1,fin2,fin3,fin4,fin5,fin6)
```

```
fintable$fin <- rowSums(fintable,na.rm=TRUE) #금융 자산 총액을 fintable 에 추가  
fintable
```

```
##### 부채 #####
```

```
#결측값 처리
```

```
h22[, c("h222602","h222605","h222608", "h222611", "h222614", "h222617")][h22[,  
c("h222602","h222605","h222608", "h222611", "h222614", "h222617")] == (-1)] <- NA
```

```
table(h20$h202562)
```

```
#부채 여부 (보유:1)
```

```
deb_y <- ifelse((h22$h222601 ==1 | h22$h222604 ==1 | h22$h222607 ==1 | h22$h222610  
==1 | h22$h222613 ==1 | h22$h222616 ==1), 1,NA)
```

```

#부채 액수
deb1 <- h22$h222602 #금융기관부채
deb2 <- h22$h222605 #비금융기관부채
deb3 <- h22$h222608 #개인적으로 빌린돈
deb4 <- h22$h222611 #전세금, 임대보증금 받은 것
deb5 <- h22$h222614 #미리 타고 앞으로 부어야 할 계
deb6 <- h22$h222617 #기타 부채
debtable <- data.frame(deb1,deb2,deb3,deb4,deb5,deb6)
debtable$deb <- rowSums(debtable,na.rm=TRUE) #부채 총액

```

#금융 자산, 부채 소유 유무 빈도

```

table(fin_y)
table(deb_y)
summary(fin_y)
summary(deb_y)

```

```

table(fintable$fin)
fintable$fin[fintable$fin==0] <- NA
summary(fintable$fin)

```

```

table(debtable$deb)
debtable$deb[debtable$deb==0] <- NA
summary(debtable$deb)

```

## Q 18

### 부동산 자산의 시가 관련 문항의 구조

‘부동산자산’에 대해서는 ① ‘현재 집을 제외한 부동산’ ② ‘현재 살고 있는 집을 포함한 부동산 임대’ ③ ‘현재 살고 있는 집 이외의 부동산 임대’에 대해서 묻고 있다(※ 현재 ‘본인의 가족만 살고 있는 집’이나 ‘전세나 월세로 사는 경우의 보증금’은 주거관련 문항에서 파악이 가능하다).

구체적인 설문구조를 살펴보면, 일단 각 문항에 대해 해당하는지 여부를 질문한 후 자산이 있을 경우 그 종류와 시가 총액을 질문한다. 그런데, 응답자가 부동산의 시가총액을 정확히 알지 못하는 경우가 빈번하므로, [잘모르겠다]고 응답한 경우 범주로 응답하도록 하고 있다. 실제로 범주화 된 문항에 응답하는 응답자가 많으므로, 시가총액을 [잘모르겠다]고 응답한 경우를 분석에서 모두 제외시키면, 실제와는 상당히 다른 결과가 나타날 수 있다. 따라서, 연구자가 연속형 부동산 시가총액을 범주화하거나, 범주형 부동산 시가를 연속형으로 바꾸어 사용해야 한다.

## Q 19

### 2, 3차년도 가구자료의 주택종류와 평수

2차년도와 3차년도 자료에서 입주형태, 주택의 종류, 주택의 평수, 주거지의 시가, 거주시키는 전년도 조사이후에 이사하였거나 이사하지는 않았더라도 변동사항이 있을 경우에만 응답하도록 하였다. 그러므로 1차년도 조사 이후에 주거사항이 변화가 없는 가구라면, 2차-3차 가구자료에서 주택종류와 평수가 missing 값을 갖게 된다. 따라서 기존의 2차-3차 가구 자료를 이용하여 주거지에 관한 분석을 할 때에는 전년도 자료에서 해당 변수를 연결하여 사용해야 한다.

그러나, 8차년도 release 자료부터는 해당 변수를 가공하여 해당차수의 데이터에 붙여두었으므로, 별도의 작업을 거치지 않고도 해당변수를 사용하면 된다.

**Q 20****개인의 인적특성(성별, 연령, 학력) 변수들이 필요할 때**

개인의 학력, 성별, 연령, 가구주와의 관계 등 인적특성 정보는 가구설문지를 통해 조사하고 있다. 그렇지만, 가구설문지를 통해 조사된 이와 같은 인적특성 항목들은 개인, 신규, 부가 자료에도 붙여서 함께 제공하고 있으므로, 이들 자료를 이용하여 분석할 때에는 개인의 인적특성 변수를 가구자료로부터 추출하여 붙이기 위해 고생할 필요가 없다.

다만, 개인 및 신규 설문지는 15세 이상인 가구원들에게만 배포되므로, 15세 미만 학령기 아동의 인적 특성과 같은 주제에 관심이 있을 경우에는 가구자료를 이용해 개인정보를 추출하여야 한다(FAQ Q 11을 이용). 가구자료의 개인정보 관련변수는 h\*\*0201부터 h\*\*0775까지이다.



## Q 21

### 종사상 지위, 임금/비임금 근로 구분의 기준

#### 1. 종사상 지위의 구분

노동패널자료에서 종사상 지위 즉, 상용·임시·일용·자영업자/고용주·가족종사자를 구분하는 조사항목은 유형설문에 있다. 유형마다 문항의 형태는 조금씩 다르다.

유형 1, 2는 지난 조사와 비교해 상용직·임시직·일용직 여부가 바뀌었는지를 묻는다.

유형 3, 4도 지난 조사와 비교해 종사상 지위가 바뀌었는지를 묻는다.

유형 5, 6은 현재 종사상 지위가 무엇인지를 묻는다.

유형 7, 8도 현재 종사상 지위가 무엇인지를 묻는다.

유형 1-4는 지난 조사로부터 이어져온 일자리이므로 "지난 조사와 비교해 변했는지"만을 질문하면 충분하다. 유형 5-8은 새로 시작한 일자리이므로 현재 종사상 지위를 질문해야 한다. 분석을 위해 종사상 지위 변수가 필요할 때에는 "p\*\*0314 : (주된일자리)종사상지위-현재(또는 최종)" 변수를 이용하면 된다. 직업력을 이용할 경우 "j150 : 종사상의 지위-현재(또는 최종)"을 이용하면 된다. 유형 1-4에 응답한 경우 이전년도로부터 관련 자료를 가져와 채워 놓았다.

#### 2. 임금/비임금의 구분

임금/비임금 근로를 구분할 경우 유형설문을 통해 조사된 종사상 지위 변수 외에도 취업자 개인 설문지를 통해 조사되는 "p\*\*0211 : (공통)취업형태" 변수를 이용할 수 있다. 이 변수는 "문4 그렇다면 \_\_\_\_님께서 하시는 이 일은 다음 중 어디에 해당되니까?"에 대한 응답이다. 그런데, 취업자 개인 설문을 통해 조사되는 취업형태 변수와 유형설문을 통해 조사되는 종사상 지위 변수가 일치하는 응답을 주어야 하지만, 불행히도 그렇지 못한 문제가 있다. 이는 한국노동패널조사와 같은 미시자료를 다룰 때 가장 빈번하게 부딪히는 "응답의 비밀관성

” 문제이다. 자영업과 임금근로의 경계가 불분명한 일자리들이 상당히 존재하는 것이 현실이어서, 이런 현실이 그대로 설문조사에 투영되는 경우라고 볼 수 있다. 이를 클리닝 과정에서 최소화하기 위해 노력하고 있지만, 아직까지 완전히 제거되지 못하고 있다. 따라서 노동패널팀에서는 임금/비임금 근로를 구별할 때에는 취업형태(p\*\*0211) 변수보다는 종사상 지위(p\*\*0314) 변수를 이용하는 것을 권하고 있다. 이렇게 하는 것이 일자리 관련 변수를 조금이라도 더 확보하여 결측치(missing value)를 줄이는 방법이다.

## Q 22

### 경제활동상태 변수 만들기

경제활동상태의 정의에 대해서는 이미 “IV. 변수 가이드”에서 이미 소개한 바 있다. 여기서는 예제를 통해 경제활동상태를 구성하는 방법을 살펴보자. 아래 예제는 21차년도 개인자료를 이용한 경제활동상태의 구성이다.

여기서 유의할 점은 개인의 종사상지위와 경제활동상태가 반드시 일치하지 않을 수 있다는 점이다. 개인의 종사상 지위는 유형설문에서 이루어지는 반면, 취업/미취업 여부, 그리고 경제활동상태는 취업자공통과 미취업자공통 설문을 통해 구성된다(FAQ Q 21 참조).

이 경우 개인의 경제활동상태는 기본적으로 취업자 및 미취업자 공통 설문변수를 기준으로 만들 것을 권고한다.

#### <SAS>

```
libname a 'c:\klips'
```

```
data p22; set a.klips22p;
```

```
/*현재 일자리의 종사상 지위 */
```

```
/* 1=상용, 2=임시, 3=일용, 4=고용주/자영업자, 5=무급가족종사자*/
```

```
empst=p220314;
```

```
/*현재 경제활동 상태 */
```

```
econst=3; /* 비경제활동인구*/
```

```
if p220201=1 then econst =1; /* 취업자 */
```

```
if p222801=1 and p222806=1 then econst=2; /* ILO 기준 실업자*/
```

```
/* 고용원 있는 자영업자, 고용원 없는 자영업자 구분하기*/
```

```
if empst=4 then self=1; /*고용원 있는 자영업자*/
```

```
if empst=4 and p220409=2 then self=2; /*고용원 없는 자영업자*/
```

```
proc freq; table empst econst self;
```

run:

## < SPSS >

GET FILE='C:\klips\klips22p.sav'.

/\*현재 일자리의 종사상 지위 \*/

/\* 임금근로자 1=상용, 2=임시, 3=일용, 4=고용주/자영업자, 5=무급가족종사자\*/

compute empst=p220314.

/\*현재 경제활동 상태 \*/

compute econst=3. /\* 비경제활동인구\*/

if (p220201=1) econst =1. /\* 취업자 \*/

if (p222801=1 and p222806=1) econst=2. /\* ILO 기준 실업자\*/

fre empst econst.

/\* 고용원 있는 자영업자, 고용원 없는 자영업자 구분하기\*/

if (empst=4) self=1. /\*고용원 있는 자영업자\*/

if (empst=4 and p220409=2) self=2. /\*고용원 없는 자영업자\*/

## < STATA >

use klips22p

\*변수설명

\*econ: 1)임금근로자, 2)비임금근로자, 3)미취업자, 4)비경제활동인구

\*econst: 1)경제활동인구, 2)비경제활동인구 (dummy 변수 방법1)

\*econst2: 1)경제활동인구, 2)비경제활동인구 (dummy 변수 방법2)

\*emp: 1)취업자, 2)미취업자

\*empst: (종사상지위) 1)상용, 2)임시, 3)일용, 4)고용주/자영업자, 5)무급가족종사자

\*self : 1)고용원 있는 자영업자, 2)고용원 없는 자영업자

gen econ = 4 /\*비경활\*/

replace econ =1 if jobclass == 1 | jobclass ==5 /\*임금\*/

replace econ =2 if jobclass == 3 | jobclass ==7 /\*비임금\*/

replace econ =3 if(p222801==1 | p222802==1) & p222806==1 /\*OECD 기준, 경활 기준\*/

/\*실업자: 지난 1주일간 구직여부 -예, 지난 한달간 구직여부 - 예, 일할 수 있었다\*/

\*replace econ =3 if p222801==1 & p222806==1 /\*ILO기준\*/

/\*실업자: 지난 1주일간 구직여부 -예; \* 일할 수 있었다;\*/

recode econ(1/3=1)(4=0), gen(econst) /\*경활/비경활 더미 변수 만들기 방법 1\*/

gen econst2 =(econ >= 1 & econ <=3) /\*더미 변수 만들기 방법 2\*/

recode econ(1/2=1)(3=2)(4=.), gen(emp) /\*경활 중 취업여부 더미 변수 만들기 \*/

gen empst=p220314

gen self=1 if empst==4 /\*고용원 있는 자영업자\*/

replace self=2 if empst==4 & p220409==2 /\*고용원 없는 자영업자\*/

<R>

klips22p ← read.spss(file = "C:\\klips\\klips22p.sav",

use.value.labels = FALSE, to.data.frame = TRUE )

p22 <-klips22p

```

#현재 일자리의 종사상 지위.
#임금근로자 1=상용, 2=임시, 3=일용, 4=고용주/자영업자, 5=무급가족종사자.
empst <- p22$p220314
#결측값 -1이 있으므로 NA 로 바꿔준다.
empst[empst==(-1)] <- NA

#현재 경제활동 상태
econst <- ifelse(p22$p220201 == 1 , 1,
                ifelse((p22$p222801 == 1 & p22$p222806 == 1), 2, 3))

table(empst)
table(econst)

## 원하는 경우 gmodels 패키지의 Crosstable 을 통해 빈도수와 퍼센티지를 알 수 있다.
install.packages("gmodels")
library(gmodels)
CrossTable(empst,econst)

# 고용원 있는 자영업자, 고용원 없는 자영업자 구분하기
self <- ifelse(p22$empst == 4, 1, NA)  #고용원 있는 자영업자
self <- ifelse(self == 1 & p22$p200409 == 2, 2, self) #고용원 없는 자영업자
table(self)

```

## Q 23

노동패널자료를 이용하여 비정규직 규모를 추산하는 방법은?

### 1. 노사정위원회 합의안에 근거한 비정규직

우리나라의 비정규직 분류는 2002년 7월 노사정위원회에서 합의된 바에 근거한다. 즉, 고용형태가 한시적근로자, 시간제근로자, 비전형근로자에 해당하는 노동자가 비정규직으로 분류된다. 여기서 비전형근로자는 파견근로자, 용역근로자, 특수형태근로종사자, 가정내 근로자, 일일(단기)근로자를 포함한다.

한국노동패널조사는 5차(2002)년도부터 미국에서 주로 사용하는 한시적근로자뿐만 아니라 파견, 용역, 독립도급, 가내근로 등 대안적 고용형태를 측정할 수 있는 문항들을 추가하였으며, 10차(2007)년도부터는 관련 문항을 보강하여 다양한 형태의 비정규직 규모를 추산할 수 있게 되었다(고용형태 부가설문의 구조와 내용에 대한 자세한 사항은 이상호, 「노동패널 10차년도 고용형태 부가조사의 개요 및 주요 결과」, 노동리뷰 2008년 7월호를 참고.)

통계청의 「경제활동인구 부가조사」에 따르면, 우리나라의 임금근로자 대비 비정규직 비율은 2005년 8월 36.6%에서 점차 하락하여 2019년 8월 36.4%이다. 한국노동패널조사는 이와 다소 차이를 보이고 있다. 통계청과 동일한 기준으로 추산된 한국노동패널조사의 임금근로자 대비 비정규직 비율은 2009년 31.4%에서 점증하여 2019년 36.8%이다. 수치상으로는 큰 차이를 보이고 있지 않으나 추세는 다르게 나타나고 있다. 이러한 차이는 조사방법의 차이에 기인한 것으로 보인다(자세한 내용은 노동패널자료 연구(Ⅰ) 제3장을 참조). 그러나 한국노동패널조사는 취업률, 실업률, 산업별 및 직종별 취업자 비율, 노동시간, 평균임금 등 대부분의 핵심변수들이 기준통계자료들과 큰 차이를 보이지 않고 있어 여전히 그 신뢰성이 높다고 볼 수 있다.

### 2. 종사상 지위에 따른 비정규직

한국노동패널조사에서는 또한 종사상 지위에 따른 비정규직 규모를 추산할 수 있다. 그러

나 한국노동패널조사의 종사상 지위는 고용계약기간만을 기준으로 결정되는 반면, 「경제활동인구 부가조사」에서는 보다 엄격한 기준을 적용하므로 양 조사 간 차이가 존재한다. 보다 자세히 설명하자면, 「경제활동인구 부가조사」는 먼저 고용계약기간에 따라 종사상 지위를 상용직, 임시직, 일용직으로 구분한다. 이후 상용직 중에서 동일한 사내 규범의 적용 여부, 퇴직금 및 상여금 수령 여부, 부가급여 지급 여부 등을 고려할 때 차별적 대우를 받고 있는 노동자를 다시 임시직 혹은 일용직으로 재분류하고 있다.

### 3. 자기 선언적 비정규직

한국노동패널조사에서는 자기 선언적 비정규직 여부를 추가로 조사하고 있다. 자기 선언적 비정규직 여부는 응답자에게 비정규직의 정의에 대해 최소한의 정의만을 제시하고 고용형태에 대한 판단을 맡기는 것이다. 따라서 앞서 종사상 지위가 임시직 또는 일용직인 응답자들은 대부분 이에 대해 비정규직으로 응답한다. 실제로 매년 80% 이상의 자기 선언적 비정규직이 임시직 혹은 일용직인 것으로 나타나고 있다. 단, 자기 선언적 비정규직 여부는 3차(2000)년도에 조사되지 않았다.

#### < SAS >

```
libname a 'C:\klips';
```

```
data p22; set a.klips22p;
```

```
* 한시적근로자 ;
```

```
IF p220501 = 1 THEN foe111 = 1 ; * 기간제근로자 > 한시적근로자 ;
```

```
ELSE foe111 = 0 ;
```

```
IF p220501 = 2 AND p220601 = 2 AND p220605 in (1,2,3,4,5,6) THEN foe112 = 1 ; /* 비기간  
제근로자 > 한시적근로자 */
```

```
ELSE IF p220501 = 2 AND p220602 = 2 THEN foe112 = 1 ; /* 비기간제근로자 > 한시적근로자  
*/
```

```
ELSE foe112 = 0 ;
```



\* 한시적근로자 ;

IF foe111+foe112 > 0 THEN foe11 = 1 ;

ELSE foe11 = 0 ;

\* 시간제근로자 ;

IF p220315 = 1 THEN foe12 = 1 ;

ELSE foe12 = 0 ;

\* 파견근로자 > 비전형근로자 ;

IF p220611 = 2 THEN foe131 = 1 ;

ELSE foe131 = 0 ;

\* 용역근로자 > 비전형근로자 ;

IF p220611 = 3 THEN foe132 = 1 ;

ELSE foe132 = 0 ;

\* 특수형태노동종사자 > 비전형근로자 ;

IF p220612 = 1 THEN foe133 = 1 ;

ELSE foe133 = 0 ;

\* 가정내근로자 > 비전형근로자 ;

IF p220613 = 1 THEN foe134 = 1 ;

ELSE foe134 = 0 ;

\* 일일(단기)근로자 > 비전형근로자 ;

IF p220508 = 1 THEN foe135 = 1 ;

ELSE foe135 = 0 ;

\* 비전형근로자 ;

IF foe131+foe132+foe133+foe134+foe135 > 0 THEN foe13 = 1 ;

ELSE foe13 = 0 ;

\* 노사정위원회 기준 비정규직 ;

IF foe11 + foe12 + foe13 > 0 THEN foe1 = 1 ;

ELSE foe1 = 0 ;

\* 종사상 지위 기준 ;

IF p220314 = 1 THEN foe2 = 0 ;

IF p220314 in (2,3) THEN foe2 = 1 ;

\* 자기 선언적 비정규직 여부 기준 ;

foe3 = p220317-1 ;

proc freq; table foe3; run;

< SPSS >

GET FILE='C:\klips\klips22p.sav'.

compute foe111=0.

if (p220501=1) foe111=1.

var label foe111 '기간제 근로자'.

compute foe112=0.

if (p220501=2 and p220601=1 and p220602=2) foe112=1.

if (p220501=2 and p220601=2 and (p220605=1 or p220605=2 or p220605=3 or p220605=4 or  
p220605=5 or p220605=6)) foe112=1.

var label foe112 '비기간제 근로자'.

compute foe11=0.

if (foe111=1 or foe112=1) foe11=1.

var label foe11 '한시적 근로자'.

compute foe12=0.

if (p220315=1) foe12=1.

var label foe12 '시간제근로자'.

compute foe131=0.

if (p220611=2) foe131=1.

var label foe131 '파견근로자'.

compute foe132=0.

if (p220611=3) foe132=1.

var label foe132 '용역근로자'.

compute foe133=0.

if (p220612=1) foe133=1.

var label foe133 '특수형태근로종사자'.

compute foe134=0.

if (p220613=1) foe134=1.

var label foe134 '가정내근로자'.

compute foe135=0.

if (p220508=1) foe135=1.

var label foe135 '일일(단기)근로자'.

```
compute foe13=0.
```

```
if (foe131=1 or foe132=1 or foe133=1 or foe134=1 or foe135=1) foe13=1.
```

```
var label foe13 '비전형근로자'.
```

```
compute foe1=0.
```

```
if (foe11=1 or foe12=1 or foe13=1) foe1=1.
```

```
var label foe1 '노사정 위원회 기준 비정규직'.
```

```
compute foe2=0.
```

```
if (p220314=2 or p210314=3) foe2=1.
```

```
var label foe2 '종사상지위 비정규직'.
```

```
compute foe3=0.
```

```
if (p220317=2) foe3=1.
```

```
var label foe3 '자기선언적 비정규직'.
```

```
execute.
```

**< STATA >**

```
use klips22p
```

```
qui gen foe111=0
```

```
qui replace foe111=1 if p220501==1
```

```
label var foe111 "기간제근로자"
```

```
qui gen foe112=0
```

```
qui replace foe112=1 if p220501==2 & p220601==1 & p220602==2
qui replace foe112=1 if p220501==2 & p220601==2 & (p220605==1 | p220605==2 |
p220605==3 | p220605==4 | p220605==5 | p220605==6)
label var foe112 "비기간제근로자"
```

```
qui gen foe11 = 0
qui replace foe11 = 1 if foe111+foe112 > 0
label var foe11 "한시적근로자"
```

```
qui gen foe12=0
qui replace foe12=1 if p220315==1
label var foe12 "시간제근로자"
```

```
qui gen foe131=0
qui replace foe131=1 if p220611==2
label var foe131 "파견근로자"
```

```
qui gen foe132=0
qui replace foe132=1 if p220611==3
label var foe132 "용역근로자"
```

```
qui gen foe133=0
qui replace foe133=1 if p220612==1
label var foe133 "특수형태근로종사자"
```

```
qui gen foe134=0
qui replace foe134=1 if p220613==1
label var foe134 "가정내근로자"
```

```
qui gen foe135=0
qui replace foe135=1 if p220508==1
```

```
label var foe135 "일일단기근로자"
```

```
qui gen foe13=0
```

```
qui replace foe13 = 1 if foe131+foe132+foe133+foe134+foe135 > 0
```

```
label var foe13 "비전형근로자"
```

```
qui gen foe1=0
```

```
qui replace foe1=1 if foe11+foe12+foe13>0
```

```
label var foe1 "노사정 위원회 기준 비정규직"
```

```
qui gen foe2=0
```

```
qui replace foe2=1 if p220314==2 | p220314==3
```

```
label var foe2 "종사상지위 비정규직"
```

```
qui gen foe3=0
```

```
qui replace foe3=1 if p220317==2
```

```
label var foe3 "자기선언적 비정규직"
```

<R>

```
library(foreign)
```

```
klips22p <- read.spss(file = "C:\\klips\\klips22p.sav",
```

```
use.value.labels = FALSE, to.data.frame = TRUE )
```

```
p22 <- klips22p
```

```
#기간제 근로자
```

```
table(p22$p210501)
```

```
foe111 <- ifelse(p22$p220501 == 1, 1, 0)
```

```
"기간제 근로자" <- foe111
```

#비기간제 근로자

```
foe112 <- ifelse((p22$p220501==2 & p22$p220601 ==1 & p22$p220602 ==2), 1,  
                ifelse((p22$p220501 ==2 & p22$p220601 ==2 &  
                        (p22$p220605==1 | p22$p220605==2 | p22$p220605==3| p22$p220605==4|  
p22$p220605==5| p22$p220605==6))  
                ,1,0 ))
```

"비기간제 근로자" <-foe112

#한시적 근로자

```
foe11 <- ifelse((foe111 ==1 | foe112 == 1) , 1 , 0)
```

"한시적 근로자" <- foe11

#시간제 근로자

```
foe12 <- ifelse(p22$p220315 ==1, 1, 0)
```

"시간제 근로자" <- foe12

#파견 근로자

```
foe131 <- ifelse(p22$p220611 == 2, 1, 0)
```

"파견 근로자" <- foe131

#용역 근로자

```
foe132 <- ifelse(p22$p220611 == 3, 1, 0)
```

"용역 근로자" <- foe132

#특수형태근로종사자

```
foe133 <- ifelse(p22$p220612 == 1, 1, 0)
```

```
"특수형태근로종사자" <- foe133
```

```
#가정내근로자
```

```
foe134 <- ifelse(p22$p220613 == 1, 1, 0)
```

```
"가정내근로자" <- foe134
```

```
#일일(단기)근로자
```

```
foe135 <- ifelse(p22$p220508 == 1, 1, 0)
```

```
"일일(단기)근로자" <- foe135
```

```
#비전형근로자
```

```
foe13 <- ifelse((foe131 ==1 | foe132 ==1 | foe133 ==1 | foe134 ==1 | foe135 ==1), 1, 0)
```

```
"비전형 근로자" <-foe13
```

```
#노사정 위원회 기준 비정규직
```

```
foe1 <- ifelse((foe11 ==1 | foe12 ==2 | foe13 ==1), 1, 0)
```

```
"노사정 위원회 기준 비정규직" <-foe1
```

```
#종사상지위 비정규직
```

```
foe2 <- ifelse((p22$p220314 ==2 | p22$p220314 ==3), 1, 0)
```

```
"종사상지위 비정규직" <- foe2
```

```
#자기선언적 비정규직
```

```
foe3 <- ifelse(p22$p220317 ==2 , 1, 0)
```

```
"자기선언적 비정규직" <- foe3
```



## Q 24

임금근로자의 근로시간변수에 결측치가 너무 많이 나온다면?

이는 노동패널의 근로시간 변수를 제대로 구성하지 못했기 때문에 발생한 실수이다. 주로 "p\*\*1004 : (주된일자리)주당 평균 근무시간(임금)" 변수만을 이용하였을 때 이런 현상이 나타난다.

노동패널에서는 유형설문을 통해 주당 근로시간을 조사한다. 유형설문에서의 근로시간 관련 항목은 임금근로자일 경우와 비임금근로자일 경우 각각 질문 형태가 다르다. 임금근로자의 경우

- ① 정규근로시간이 정해져 있는지,
- ② 정해져 있지 않을 경우 식사시간을 제외하고 일주일에 평균 몇 시간 일했는지
- ③ 정해져 있을 경우 정규근로시간이 식사시간을 제외하고 일주일에 몇 시간인지, 정규근로시간 이외에 초과로 근로하는 시간이 있었는지, 있었다면 일주일 평균 몇 시간이나 되는지를 질문하는 방식으로 구성되어 있다.

비임금근로자일 경우 한 문항(p\*\*1031)으로 간단하게 일주일에 몇 시간이나 일하는지 질문한다. 아래는 임금근로자의 근로시간 변수 생성방법 예제이다.

<SAS>

```
libname a 'C:\klips';  
data p22; set a.klips22p;
```

/\*변수설명

p221003 : 주된일자리 정규근로시간 여부 y/n = 1/2

p221004 : 주된일자리 주당 평균 근로시간

p221006 : 주된일자리 주당 정규 근로시간

p221011 : 주된일자리 초과근로여부 y/n = 2/1

p221012 : 주된일자리 주당 평균 초과근로시간\*/

array w[3] p221004 p221006 p221012; /\*결측치 처리\*/

do i=1 to 3;

if w[i] =-1 then w[i]=.;

end;

if p221003=1 and p221011=1 then worktime=p221006; /\*주당 정규근로시간\*/

if p221003=1 and (p221011=2 & p221019=1) then worktime=sum(of p221006, p221012); /\*  
주당 정규근로시간+주당 초과근로시간\*/

if p221003=1 and (p221011=2 & p221019=2) then worktime=sum(of p221006, p221012/4.3);  
/\*주당 정규근로시간+(월평균 초과근로시간/4.3=주당 초과근로시간)\*/

if p221003=2 then worktime=p221004; /\*정규근로시간이 정해져있지 않은 경우 주당 평균 근로시  
간 사용\*/

/\*불가능한 초과근로시간 확인\*/

proc freq; table p221004 p221006 p221012;

proc print; where p221012>168; var pid p221012;

run;

data hours; set p21(keep=worktime); /\*근로시간 변수 그래프\*/

proc chart data=hours; vbar worktime; title 'Hours worked per week';

run;

< SPSS >

GET FILE='D:\klips\klips22p.sav'.

recode p221004 p221006 p221012(-1=sysmis).

if (p221003=1 and p221011=1) worktime=p221006./\*주당 정규근로시간\*/

```

if (p221003=1 and (p221011=2 and p221019=1)) worktime=sum(p221006, p221012)./*주당
정규근로시간+주당 초과근로시간*/
if (p221003=1 and (p221011=2 and p221019=2)) worktime=sum(p221006, p221012/4.3). /*
주당 정규근로시간+(월평균 초과근로시간/4.3=주당 초과근로시간)*/
if (p221003=2) worktime=p221004. /*정규근로시간이 정해져있지 않은 경우 주당 평균 근로시간
사용*/

```

```

GRAPH /HISTOGRAM=worktime.

```

```

/*불가능한 초과근로시간 확인*/
select if (p221012>168).
list pid p221012.

```

## < STATA >

```

use klips22p

```

```

recode p221004 p221006 p221012(-1=.)
gen worktime=p221006 if p221003==1 & p221011==1 /*주당 정규근로시간*/
replace worktime=p221006 + p221012 if p221003==1 & (p221011==2 & p221019==1) /*주당
정규근로시간+주당 초과근로시간*/
replace worktime=p221006 + p221012/4.3 if p221003==1 & (p221011==2 & p221019==2) /*
주당 정규근로시간+(월평균 초과근로시간/4.3=주당 초과근로시간)*/
replace worktime=p221004 if p221003==2 /*정규근로시간이 정해져있지 않은 경우 주당 평균 근
로시간 사용*/

```

```

histogram worktime

```

```

/*불가능한 초과근로시간 확인*/

```

```

keep if p221012>168

```

```
tab pid p221012
```

< R >

```
p22 <- read.spss(file = "C:\\klips\\klips22p.sav",  
use.value.labels = FALSE, to.data.frame = TRUE )
```

#변수 설명

#p221003 : 주된일자리 정규근로시간 여부 y/n = 1/2

#p221004 : 주된일자리 주당 평균 근로 시간

#p221006 : 주된일자리 주당 정규 근로 시간

#p221011 : 주된일자리 초과근로여부 y/n =2/1

#p221012 : 주된일자리 주당 평균 초과근로시간

#결측치 처리

```
p22$p221004[p22$p221004 == (-1)] <- NA
```

```
p22$p221006[p22$p221006 == (-1)] <- NA
```

```
p22$p221012[p22$p221012 == (-1)] <- NA
```

#평균 근로 시간

```
sum <- data.frame(p22$p221006, p22$p221012)
```

```
rowtotal <- rowSums(sum, na.rm=TRUE) #변수 활용을 위해 합을 구함
```

```
worktime <- ifelse(p22$p221003 ==2 , p22$p221004,  
  ifelse((p22$p221003 ==1 & p22$p221011 == 1), p22$p221006,  
    ifelse((p22$p221003 ==1 & p22$p221011 ==2), rowtotal, NA)))
```

```
summary(worktime)
```

```
hist(worktime) #worktime 히스토그램
```

#불가능한 초과근로시간 확인

```

over <- filter(p22, p221012 > 168)
over <- p22[p22$p221012 > 168, c("pid", "p221012")] #pid 까지 확인하고 싶은 경우.
table(over)

worktime <- ifelse(p22$p221003==1 & p22$p221011==1, p22$p221006,
  ifelse(p22$p221003==1 & p22$p221011==2 & p22$p221019==1, p22$p221006 +
p22$p221012,
    ifelse(p22$p221003==1 & p22$p221011==2 & p22$p221019==2, p22$p221006 +
(p22$p221012/4.3),
      ifelse(p22$p221003==2, p22$p221004, NA))))
summary(worktime)

```

**Q 25**

임금 및 소득 변수의 평균값 추이가 들락날락하고, 너무 큰 값이 나올 때

임금, 소득 변수의 평균값을 낼 때 가장 주의해야 할 점이 무응답의 처리이다. 노동패널의 임금 및 소득 변수는 모두 6자리로 top-coding을 하며, 무응답일 경우 ‘-1’로 처리하고 있다. 따라서, 무응답을 결측치로 바꿔 처리하지 않을 경우 값이 커지거나 추이가 안 나타나는 등의 문제가 나타난다.

노동패널팀에서 일괄적으로 무응답을 결측치로 처리하여 제공하는 것이 데이터 분석자의 입장에서 편할 수도 있겠지만, 보다 정확한 연구를 위해서는 무응답과 결측값이 완전히 다른 의미를 갖는 문제이므로 -1을 결측치로 재처리해 제공하지는 않고 있다. 또한, 아직까지 100억에 달하는 월평균임금을 보고한 사람들은 없었으므로, top-coding이 실제로 이루어진 사례는 없었다.

임금, 소득 관련 변수를 이용할 때에는 반드시 기초통계량을 구하여 코드북에 기재된 빈도값과 일치하는지를 확인한 후 무응답 값을 처리하기를 권장한다.

## Q 26

개인자료에 부업 관련 정보를 붙이고자 할 때

개인자료에는 주업과 관련된 정보만 들어 있다. 따라서 부업 관련 정보를 필요로 할 경우 직업력을 이용해 부업 관련 정보를 찾아낸 후 이를 개인자료에 다시 붙여 이용해야 한다.

예를 들어 22 년도에 조사된 개인의 부업을 찾는 상황을 생각해 보자. 직업력에서 아래의 세 가지 조건을 동시에 만족하는 일자리가 이들 개인의 부업이 된다.

- ① 조사년도를 의미하는 jobwave 변수가 22
- ② 부업을 의미하는 mainjob이 0,
- ③ 일자리의 현재 지속여부를 알 수 있는 jobclass가 1, 3, 5, 7.(1, 3, 5, 7만 남기는 이유는 이들 일자리가 조사시점 현재 지속 중인 일자리이기 때문이다.)

이상의 절차를 통해 걸러진 데이터를 개인자료와 pid를 기준으로 붙이면 된다.

## Q 27

### 지니계수 구하기

소득 불평등을 나타내는데 가장 대표적인 지표는 지니계수이다. 노동패널 자료를 이용하여 지니계수를 구할 때는 가구의 시장 소득을 구성해야 한다.

지니계수와 관련한 자세한 내용은 노동패널자료연구(Ⅰ) 제1장(홍민기 저)을 참조하기 바란다. 아래는 STATA를 이용한 지니계수 산출프로그램 예제이다. SAS와 SPSS, R 프로그램은 생략한다.

#### < STATA >

```
use klips22h
```

```
/* stata - 가구 시장 소득 구성하기 */
```

```
keep if hwave22==1
```

```
gen hwgt=w22h //가구가중치
```

```
#delimit:
```

```
recode h222102 h222112-h222116 h222122-h222126
```

```
h222134 h222136 h222138 h222140 h222142 h222152-h222160
```

```
h222183-h222191 (999999=. );
```

```
recode h222102 h222112-h222116 h222122-h222126
```

```
h222134 h222136 h222138 h222140 h222142 h222152-h222160
```

```
h222183-h222191 (-1=. );
```

```
#delimit cr
```

```
gen inc_labor=h222102 //근로소득
```

```
egen inc_fin=rowtotal(h222112-h222116) //금융소득
```

```
egen inc_est=rowtotal(h222122-h222126) //부동산소득
```

```
egen inc_tran=rowtotal(h222155 h222157-h222160) //사적이전소득
```



```
egen inc_market=rowtotal(inc_labor inc_fin inc_est inc_tran) //시장소득
```

```
gen hsize=h220150 //가구원수
```

```
gen inc_equiv=inc_market/sqrt(hsize) //균등화가구소득
```

```
gen pwgt=hsize*hwgt //인구가중치=가구원수*가구가중치
```

```
ineqdec0 inc_equiv [w=pwgt] //지니계수
```

**Q 28****학력 및 교육년수 변수를 구성 할 때**

임금에 관한 분석모형을 구성할 때 학력은 중요한 변수로 사용된다. 보통 학력 변수의 경우 중학교 졸업 이하, 고등학교 졸업, 전문대 졸업, 대학교 졸업, 대학원 졸업의 최종 학력의 더미변수를 구성하여 활용하기도 하지만, 총 교육년수를 활용하여 사용되기도 한다. 다음은 개인용 데이터를 활용하여 학력 및 총 교육년수 변수를 구성하는 방법의 각 통계 프로그램 별 예제이다.

**<SAS>**

```
libname a 'C:\klips';
data p22; set a.klips22p;

/* 학력 더미변수 만들기 (h**0661) */
/* 중졸이하 = 1 고졸 =2 전문대졸=3 대졸=4 석사이상=5*/
if p220110 in (1,2,3,4) then edu=1;
if p220110 =5 then edu =2;
if p220110 =6 then edu =3;
if p220110 =7 then edu =4;
if p220110 in (8,9) then edu =5;
/* 졸업=1, 수료&중퇴&재학중&휴학중=2 */
if p220111 =1 then degree =1;
if p220111 in (2,3,4,5) then degree =2;

if edu =2 and degree~=1 then edu =1;
if edu =3 and degree~=1 then edu =2;
if edu =4 and degree~=1 then edu =3 ;
if edu =5 and degree~=1 then edu =4 ;
if edu =-1 then edu =.;
```

```
/* 한 변수를 가지고 다수의 더미변수를 만드는 명령문*/
```

```
array dum{5} edu_dum_1-edu_dum_5;
```

```
do j=1 to 5;
```

```
if edu = j then do i = 1 to 5;
```

```
if i = j then dum{i} = 1;
```

```
else dum{i} = 0;
```

```
end;
```

```
end;
```

```
/* level = 학년 */
```

```
level = p210112;
```

```
/* 졸업&수료를 제외한 중퇴&재학중&휴학중의 경우 학년 추가*/
```

```
if p220110 in (1,2) then edu_scale =0;
```

```
if p220110 =3 then edu_scale =6;
```

```
if p220110 =4 then edu_scale =9;
```

```
if p220110 =5 then edu_scale =12;
```

```
if p220110 =6 then edu_scale =14;
```

```
if p220110 =7 then edu_scale =16;
```

```
if p220110 =8 then edu_scale =18;
```

```
if p220110 =9 then edu_scale =21;
```

```
if p220110 =3 and level =1 then edu_scale =1;
```

```
if p220110 =3 and level =2 then edu_scale =2;
```

```
if p220110 =3 and level =3 then edu_scale =3;
```

```
if p220110 =3 and level =4 then edu_scale =4;
```

```
if p220110 =3 and level =5 then edu_scale =5;
```

```
if p220110 =3 and level =6 then edu_scale =6;
```

```
if p220110 =4 and level =1 then edu_scale =7;
```

```
if p220110 =4 and level =2 then edu_scale =8;
```

```
if p220110 =4 and level =3 then edu_scale =9;
```

```
if p220110 =5 and level =1 then edu_scale =10;
if p220110 =5 and level =2 then edu_scale =11;
if p220110 =5 and level =3 then edu_scale =12;
```

```
if p220110 =6 and level =1 then edu_scale =13;
if p220110 =6 and level =2 then edu_scale =14;
```

```
if p220110 =7 and level =1 then edu_scale =13;
if p220110 =7 and level =2 then edu_scale =14;
if p220110 =7 and level =3 then edu_scale =15;
if p220110 =7 and level =4 then edu_scale =16;
```

```
if p220110 =8 and level =1 then edu_scale =17;
if p220110 =8 and level =2 then edu_scale =18;
```

```
if p220110 =9 and level =1 then edu_scale =19;
if p220110 =9 and level =2 then edu_scale =20;
if p220110 =9 and level =3 then edu_scale =21;
proc freq; table edu degree edu_scale;
run;
```

< SPSS >

```
GET FILE='C:\klips\klips22p.sav'.
```

```
if (p220110 =1 or p220110 =2 or p220110 =3 or p220110 =4) edu=1.
if (p220110 =5) edu=2.
if (p220110 =6) edu=3.
if (p220110 =7) edu=4.
if (p220110 =8 or p220110 =9) edu=5.
```

```
if (p220111 =1) degree=1.  
if (p220111 =2 or p220111 =3 or p220111 =4 or p220111 =5) degree =2.
```

```
if (edu =2 and degree >1) edu =1.  
if (edu =3 and degree >1) edu =2.  
if (edu =4 and degree >1) edu =3.  
if (edu =5 and degree >1) edu =4.
```

```
recode edu(-1=sysmiss).  
recode edu(1=1) (else=0) into edu_dum_1.  
recode edu(2=1) (else=0) into edu_dum_2.  
recode edu(3=1) (else=0) into edu_dum_3.  
recode edu(4=1) (else=0) into edu_dum_4.  
recode edu(5=1) (else=0) into edu_dum_5.  
fre edu.  
execute.
```

```
if (p220110 =1 or p220110 =2) edu_scale =0.  
if (p220110 =3) edu_scale =6.  
if (p220110 =4) edu_scale =9.  
if (p220110 =5) edu_scale =12.  
if (p220110 =6) edu_scale =14.  
if (p220110 =7) edu_scale =16.  
if (p220110 =8) edu_scale =18.  
if (p220110 =9) edu_scale =21.
```

```
compute level =p220112.  
if (p220110 =3 and level =1) edu_scale =1.  
if (p220110 =3 and level =2) edu_scale =2.  
if (p220110 =3 and level =3) edu_scale =3.  
if (p220110 =3 and level =4) edu_scale =4.
```

if (p220110 =3 and level =5) edu\_scale =5.

if (p220110 =3 and level =6) edu\_scale =6.

if (p220110 =4 and level =1) edu\_scale =7.

if (p220110 =4 and level =2) edu\_scale =8.

if (p220110 =4 and level =3) edu\_scale =9.

if (p220110 =5 and level =1) edu\_scale =10.

if (p220110 =5 and level =2) edu\_scale =11.

if (p220110 =5 and level =3) edu\_scale =12.

if (p220110 =6 and level =1) edu\_scale =13.

if (p220110 =6 and level =2) edu\_scale =14.

if (p220110 =7 and level =1) edu\_scale =13.

if (p220110 =7 and level =2) edu\_scale =14.

if (p220110 =7 and level =3) edu\_scale =15.

if (p220110 =7 and level =4) edu\_scale =16.

if (p220110 =8 and level =1) edu\_scale =17.

if (p220110 =8 and level =2) edu\_scale =18.

if (p220110 =9 and level =1) edu\_scale =19.

if (p220110 =9 and level =2) edu\_scale =20.

if (p220110 =9 and level =3) edu\_scale =21.

recode edu\_scale(-1=sysmiss).

fre edu\_scale.

execute.

## < STATA

```
use klips22p
```

```
/* 학력 더미변수 만들기 (h**0661) */
```

```
/* 중졸이하= 1 고졸=2 전문대졸=3 대졸=4 석사이상=5*/
```

```
recode p220110(1/4=1)(5=2)(6=3)(7=4)(8/9=5), gen(edu)
```

```
recode p220111(1=1)(2/5=2), gen(degree)/* 졸업=1, 수료&중퇴&재학중&휴학중=2 */
```

```
replace edu =1 if edu==2 & degree!=1 /* 졸업을 제외한 나머지는 하위학력코딩*/
```

```
replace edu =2 if edu==3 & degree!=1 /* "~"를 사용해도 됨*/
```

```
replace edu =3 if edu==4 & degree!=1
```

```
replace edu =4 if edu==5 & degree!=1
```

```
mvdecode edu, mv(-1) /*-1값을 결측치로 전환*/
```

```
tab edu, gen(edu_dum_) /* 한 변수를 가지고 다수의 더미변수를 만드는 명령문*/
```

```
/* 학력 숫자변수 만들기 */
```

```
recode p220110(1/2=0)(3=6)(4=9)(5=12)(6=14)(7=16)(8=18)(9=21), gen(edu_scale)
```

```
gen level = p220112 /* level = 학년 */
```

```
replace edu_scale=1 if p220110==3 & level==1
```

```
/* 졸업&수료를 제외한 중퇴&재학중&휴학중의 경우 학년 추가*/
```

```
replace edu_scale=2 if p220110==3 & level==2
```

```
replace edu_scale=3 if p220110==3 & level==3
```

```
replace edu_scale=4 if p220110==3 & level==4
```

```
replace edu_scale=5 if p220110==3 & level==5
```

```
replace edu_scale=6 if p220110==3 & level==6
```

```
replace edu_scale=7 if p220110==4 & level==1
```

```
replace edu_scale=8 if p220110==4 & level==2
```

```
replace edu_scale=9 if p220110==4 & level==3
```

```
replace edu_scale=10 if p220110==5 & level==1
replace edu_scale=11 if p220110==5 & level==2
replace edu_scale=12 if p220110==5 & level==3
```

```
replace edu_scale=13 if p220110==6 & level==1
replace edu_scale=14 if p220110==6 & level==2
```

```
replace edu_scale=13 if p220110==7 & level==1
replace edu_scale=14 if p220110==7 & level==2
replace edu_scale=15 if p220110==7 & level==3
replace edu_scale=16 if p220110==7 & level==4
```

```
replace edu_scale=17 if p220110==8 & level==1
replace edu_scale=18 if p220110==8 & level==2
```

```
replace edu_scale=19 if p220110==9 & level==1
replace edu_scale=20 if p220110==9 & level==2
replace edu_scale=21 if p220110==9 & level==3
mvdecode edu_scale, mv(-1)
tab edu_scale
```

**< R >**

```
p22$edu <- ifelse(p22$p220110 ==1 | p22$p220110 ==2 | p22$p220110 ==3 | p22$p220110
==4, 1,
  ifelse(p22$p220110 ==5, 2,
    ifelse(p22$p220110 ==6, 3,
      ifelse(p22$p220110 ==7, 4,
        ifelse(p22$p220110 ==8| p22$p220110 ==9, 5, NA))))))
```



```
p22$degree <- ifelse(p22$p220111 ==1, 1,
  ifelse(p22$p220111 ==2 | p22$p220111 ==3 | p22$p220111 ==4 | p22$p220111
==5, 2, NA))
```

```
p22$edu <- ifelse(p22$edu ==2 & p22$degree !=1, 1,
  ifelse(p22$edu ==3 & p22$degree !=1, 2,
    ifelse(p22$edu ==4 & p22$degree !=1, 3,
      ifelse(p22$edu ==5 & p22$degree !=1, 4, NA))))
```

```
p22$edu[p22$edu == (-1)] <- NA
```

```
p22$dum1 <- as.numeric(p22$edu ==1)
p22$dum2 <- as.numeric(p22$edu ==2)
p22$dum3 <- as.numeric(p22$edu ==3)
p22$dum4 <- as.numeric(p22$edu ==4)
p22$dum5 <- as.numeric(p22$edu ==5)
```

```
p22$edu_scale <- ifelse(p22$p220110 ==1 | p22$p220110 ==2, 0,
  ifelse(p22$p220110 ==3, 6,
    ifelse(p22$p220110 ==4, 9,
      ifelse(p22$p220110 ==5, 12,
        ifelse(p22$p220110 ==7, 16,
          ifelse(p22$p220110 ==8, 18,
            ifelse(p22$p220110 ==9, 21, NA)))))))))
```

```
level <- p22[, "p210112"]
p22$level <-level
```

```

p22$edu_scale <- ifelse(p22$p220110 ==3 & level ==1, 1,
                        ifelse(p22$p220110 ==3 & level ==2, 2,
                              ifelse(p22$p220110 ==3 & level ==3, 3,
                                    ifelse(p22$p220110 ==3 & level ==4, 4,
                                          ifelse(p22$p220110 ==3 & level ==5, 6,
                                                ifelse(p22$p220110 ==3 & level ==6, 6,
                                                      ifelse(p22$p220110 ==4 & level ==1, 7,
                                                            ifelse(p22$p220110 ==4 & level ==2, 8,
                                                                  ifelse(p22$p220110 ==4 & level ==3, 9,
                                                                        ifelse(p22$p220110 ==5 & level ==1,
10,
                                                                ifelse(p22$p220110 ==5 & level
==2, 11,
                                                                ifelse(p22$p220110 ==5 &
level ==3, 12,
                                                                ifelse(p22$p220110 ==6
& level ==1, 13,
                                                                ifelse(p22$p220110
==6 & level ==2, 14,
                                                                ifelse(p22$p220110 ==7 & level ==1, 13,
                                                                ifelse(p22$p220110 ==7 & level ==2, 14,
                                                                ifelse(p22$p220110 ==7 & level ==3, 15,
                                                                ifelse(p22$p220110 ==7 & level ==5, 16,
                                                                ifelse(p22$p220110 ==8 & level ==1, 17,
                                                                ifelse(p22$p220110 ==8 & level ==2, 18,

```

```
ifelse(p22$p220110 ==9 & level ==1, 19,
```

```
ifelse(p22$p220110 ==9 & level ==2, 20,
```

```
ifelse(p22$p220110 ==9 & level ==3, 21,NA
```

```
))))))))))))))))))
```

```
p22$edu_scale[p22$edu_scale == (-1)] <- NA
```

## Q 29

### 기업규모 변수 구성 할 때

개인용 데이터를 활용한 임금의 결정에 관한 분석모형을 구성할 때 기업 규모 및 사업체 규모는 자주 사용되는 변수 중 하나이다. 노동패널조사에서는 종사하는 일자리의 전체 종업원 수 (p\*\*0402)와 현재 실제로 근무하고 있는 사업장의 종업원 수 (p\*\*0405)를 묻는 항목이 있다. 또한, 정확한 종업원 수를 모르는 경우에 위의 두 항목에 대해서 종업원 수에 해당하는 범주를 선택하는 항목이 추가로 있다. (p\*\*0403, p\*\*0406). 따라서 범주형 기업 규모 및 사업체 규모 변수를 구성하기 위해서는 위의 p\*\*0403 혹은 p\*\*0406과 함께 정확한 종업원 수를 응답한 p\*\*0402 및 p\*\*0405의 항목을 범주화하여 변수를 구성해야 한다. 다음은 개인용 데이터를 활용한 범주형 기업 규모 변수 구성의 예제이다.

#### <SAS>

```
libname a 'C:\klips';
data p22; set a.klips22p;

/* 기업규모 */
/**tip *****.
* 기업형태가 (5) 정부기관인경우는 기업규모가 missing이므로 주의.
*****/

if p220211 ~=1 then delete;
if p220403 ~=. and p220211 =1 then fsize = p220403;
if p220211 =1 then temp =p220402;

/* 연속형변수->범주화 */
if temp >= 1 and temp < 5 then fsize =1;
if temp >= 5 and temp < 10 then fsize =2;
if temp >= 10 and temp < 30 then fsize =3;
```

```

if temp >= 30 and temp < 50 then fsize =4;
if temp >= 50 and temp <70 then fsize =5;
if temp >= 70 and temp <100 then fsize =6;
if temp >= 100 and temp <300 then fsize =7;
if temp >= 300 and temp < 500 then fsize =8;
if temp >= 500 and temp < 1000 then fsize =9;
if temp >= 1000 and temp < 100000 then fsize =10;
drop temp;

```

```

/* 기업체 종류 - (주된일자리)기업형태 활용 */
if p220211 =1 then fkind = p220401;

```

```

/* 임금 */
array d{4} p221642 p221643 p221702 p221703 ;
do i = 1 to 4;
if d{i} =-1 then d{i} =.;
end;
/*현재의 월평균 소득(세금공제후)*/
mwage = p221642;
/*현재 월평균 공제세금*/
mtax = p221643;
/*지난해의 세전소득(개인설문)*/
pretwage = p221702/12;
/*지난해의 세후소득 */
postwage = p221703/12;

```

```

proc freq; table fsize fkind;
proc univariate; var mwage mtax pretwage postwage;
run;

```

< SPSS >

GET FILE='C:\klips\klips22p.sav'.

select if (p220211 =1).

if (p220211 =1) temp =p220402.

recode p220403 (sysmis=-1).

if (p220403 =-1 and p220211 =1) fsize =p220403.

if (temp >=1 and temp <5) fsize = 1.

if (temp >=5 and temp <10) fsize = 2.

if (temp >=10 and temp <30) fsize = 3.

if (temp >=30 and temp <50) fsize = 4.

if (temp >=50 and temp <70) fsize = 5.

if (temp >=70 and temp <100) fsize =6 .

if (temp >=100 and temp <300) fsize = 7.

if (temp >=300 and temp <500) fsize = 8.

if (temp >= 500 and temp <1000) fsize = 9.

if (temp >=1000 and temp <100000) fsize = 10.

execute.

recode p221642 p221643 p221702 p221703 (-1=sysmiss).

compute mwage =p221642.

compute mtax =p221643.

compute pretwage =p221702.

compute postwage =p221703.

execute.

## < STATA >

```
use klips22p
```

```
keep if p220211 == 1 /*임금근로자만 추출*/
```

```
/* 기업규모 */
```

```
/**tip *****.
```

```
* 기업형태가 (5) 정부기관인경우는 기업규모가 missing이므로 주의.
```

```
*****/
```

```
/* 임금근로자만 추출 할때 사용하는 변수 p210211==1 */
```

```
/* 기존 범주형변수 */
```

```
gen fsize = p220403 if p220403 ~= . & p220211==1
```

```
/* 범주형으로 만들기.
```

```
연속형변수->범주화 */
```

```
gen temp = p220402 if p220211==1
```

```
replace fsize = 1 if temp >= 1 & temp < 5
```

```
replace fsize = 2 if temp >= 5 & temp < 10
```

```
replace fsize = 3 if temp >= 10 & temp < 30
```

```
replace fsize = 4 if temp >= 30 & temp < 50
```

```
replace fsize = 5 if temp >= 50 & temp < 70
```

```
replace fsize = 6 if temp >= 70 & temp < 100
```

```
replace fsize = 7 if temp >= 100 & temp < 300
```

```
replace fsize = 8 if temp >= 300 & temp < 500
```

```
replace fsize = 9 if temp >= 500 & temp < 1000
```

```
replace fsize =10 if temp >=1000 & temp <100000
```

drop temp

/\* 기업체 종류 - (주된일자리)기업형태 활용 \*/

/\*

gen fkind = p220401 if p220211==1

제거 \*/

/\* 임금 \*/

mvdecode p221642 p221643 p221702 p221703, mv(-1)

gen mwage = p221642 /\*현재의 월평균 소득(세금공제후)\*/

gen mtax = p221643 /\*현재 월평균 공제세금\*/

gen pretwage = p221702/12 /\*지난해의 세전소득(개인설문)\*/

gen postwage = p221703/12 /\*지난해의 세후소득 \*/

/\* 사용시, 소비자 물가지수를 곱해 실질임금으로 변환 후 사용 \*/

보류 \*/

tab1 fsize

sum mwage mtax pretwage postwage

<R>

library(foreign)

library(dplyr)

klips22 <- read.spss(file = "C:\\klips\\klips22h.sav",

use.value.labels = FALSE, to.data.frame = TRUE )

q29\$fsize <- ifelse(q29\$p220403 !=NA | q29\$p220211 ==1, q29\$p220403, NA)

q29\$temp <- ifelse(q29\$p220211 ==1, q29\$p220402, NA)



```

q29$fsize <- ifelse(q29$temp>=1 & q29$temp <5, fsize ==1,
  ifelse(q29$temp>=5 & q29$temp <10, fsize ==2,
    ifelse(q29$temp>=10 & q29$temp <30, fsize ==3,
      ifelse(q29$temp>=30 & q29$temp <50, fsize ==4,
        ifelse(q29$temp>=50 & q29$temp <70, fsize ==5,
          ifelse(q29$temp>=70 & q29$temp <100, fsize ==6,
            ifelse(q29$temp>=100 & q29$temp <300, fsize ==7,
              ifelse(q29$temp>=300 & q29$temp <500, fsize ==8,
                ifelse(q29$temp>=500 & q29$temp <1000, fsize ==9,
                  ifelse(q29$temp>=1000 & q29$temp <100000, fsize
==10, NA

```

```

)))))))))

```

```

q29$temp <- NULL

```

```

q29[, c("p221642", "p221643", "p221702", "p221703")][q29[, c("p221642", "p221643", "p221702",
"p221703")] == (-1)] <- NA

```

```

level <- klips21p[ , "p220112"]

```

```

q29$m wage <- q29[ , "p221642"]

```

```

q29$mtax <- q29[ , "p221643"]

```

```

q29$pret wage <- q29[ , "p221702"]

```

```

q29$post wage <- q29[ , "p221703"]

```

**Q 30****동시 일자리 변수 구성 할 때**

취업자 개인 질문 항목에는 주된 일자리 외에 또 다른 일(직장)이나, 부업/아르바이트 여부에 관한 질문 항목이 있다. 이를 바탕으로 직업력데이터에는 동시 일자리에 관한 데이터를 제공하고 있다. 직업력 데이터에서 동시 일자리가 있는 경우 각 pid별로 종으로 데이터가 정렬되어 있으며, 각 일자리 번호의 jobnum 변수를 통해 주된 일자리 여부를 알 수 있다. 연구를 수행하면서 동시 일자리의 개수 혹은 동시 일자리를 횡으로 정렬하여 연구해야 하는 경우가 발생할 수 있다. 다음은 각 프로그램 별로 한해년도의 동시 일자리를 횡으로 데이터를 정렬하는 예제와 전 생애에 걸친 (회고적 일자리 제외) 일자리에 관한 정보를 pid와 일자리가 조사된 순서인 jobseq별로 횡으로 데이터를 정렬하는 예제이다.

**<SAS>**

```
/* 22차년도에 동시에 몇개의 일자리를 가지고 있는가? (pid를 기준으로 reshape) */
```

```
libname a 'C:\klips'
```

```
data w22; set a.klips22w;
```

```
/* 22차년도 응답한 일자리만 남김*/
```

```
if jobwave ~= 22 then delete;
```

```
if jobcens = 0 then delete;
```

```
if jobclass in (2, 4, 6, 8) then delete;
```

```
/*종사상지위 */
```

```
jobst = j150;
```

```
/*무급가족종사자*/
```

```
if j150 =5 then jobst =6;
```

```
/*자영업: 종업원수가 0명인 경우*/
```

```
if j150 =4 and (j502 =2 or j502 =0) then jobst =5;
```

```
keep pid jobwave jobseq jobnum jobtype jobclass mainjob jobcens jobst;
```

```
run;
```

```
proc sort; by pid; run;
```

```
proc freq; table pid/out=count1(keep=pid count); run;
```

```
data ex1;
```

```
merge w22 count1;
```

```
by pid;
```

```
retain count 0;
```

```
if pid ~=lag(pid) then n = 1;
```

```
else n+1;
```

```
proc print;
```

```
run;
```

```
proc transpose data=ex1 out=wide1 prefix = jobwave;
```

```
by pid
```

```
id n;
```

```
var jobwave;
```

```
run;
```

```
proc transpose data=ex1 out=wide2 prefix = jobseq;
```

```
by pid;
```

```
id n;
```

```
var jobseq;
```

```
run;
```

```
proc transpose data=ex1 out=wide3 prefix = jobcens;
```

```
by pid;
```

```
id n;
```

```
var jobcens;
```

```
run;
```

```
proc transpose data=ex1 out=wide4 prefix = jobtype;  
    by pid;  
    id n;  
    var jobtype;  
run;
```

```
proc transpose data=ex1 out=wide5 prefix = jobclass;  
    by pid;  
    id n;  
    var jobclass;  
run;
```

```
proc transpose data=ex1 out=wide6 prefix = jobnum;  
    by pid;  
    id n;  
    var jobnum;  
run;
```

```
proc transpose data=ex1 out=wide7 prefix = mainjob;  
    by pid;  
    id n;  
    var mainjob;  
run;
```

```
proc transpose data=ex1 out=wide8 prefix = jobst;  
    by pid;  
    id n;  
    var jobst;  
run;
```

```

data ex_j1;
    merge    wide1(drop=_name_)    wide2(drop=_name_)    wide3(drop=_name_)
wide4(drop=_name_) wide5(drop=_name_)    wide6(drop=_name_)    wide7(drop=_name_)
wide8(drop=_name_);
    by pid;
    drop _LABEL_;
    /*일자리 갯수*/
    mjob=2- nmiss(jobnum1, jobnum2, jobnum3);
run;

```

```

/* 일생동안 몇개의 일자리를 가지고 있는가? (pid와 jobseq를 기준으로 reshape) */

```

```

data w22; set a.klips22w;

```

```

if jobcens =0 then delete;

```

```

if jobwave in (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11) then delete;

```

```

/*종사상지위 */

```

```

jobst = j150;

```

```

/*무급가족종사자*/

```

```

if j150 =5 then jobst =6;

```

```

/*자영업: 종업원수가 0명인 경우*/

```

```

if j150 =4 and (j502 =2 or j502 =0) then jobst =5;

```

```

digits1 = int(log10(jobseq))+1;

```

```

if digits1=1 then jobseqa=compress('00000')||compress(jobseq);

```

```

else if digits1=2 then jobseqa=compress('0000')||compress(jobseq);

```

```

else if digits1=3 then jobseqa=compress('000')||compress(jobseq);

```

```

digits2 = int(log10(pid))+1;

```

```

if digits2=3 then pida=compress('000')||compress(pid);

```

```

else if digits2=4 then pida=compress('00')||compress(pid);

```

```
else if digits2=5 then pida=compress('0')||compress(pid);
```

```
retain pida jobseqa;
```

```
pidseq = compress(pida)||compress(jobseqa);
```

```
keep pid jobwave jobseq jobnum jobtype jobclass mainjob jobcens jobst digits1 digits2
```

```
jobseqa pida pidseq;
```

```
proc sort; by pidseq; run;
```

```
proc freq; table pidseq/out=count2(keep=pidseq count);
```

```
run;
```

```
data ex2;
```

```
merge w22 count2;
```

```
by pidseq;
```

```
retain count 0;
```

```
if pidseq ~=lag(pidseq) then n = 1;
```

```
else n+1;
```

```
run;
```

```
proc transpose data=ex2 out=wide1 prefix = jobwave;
```

```
by pidseq;
```

```
id n;
```

```
var jobwave;
```

```
run;
```

```
proc transpose data=ex2 out=wide2 prefix = jobcens;
```

```
by pidseq;
```

```
id n;
```

```

        var jobcens;
run;

```

```

proc transpose data=ex2 out=wide3 prefix = jobtype;
    by pidseq;
    id n;
    var jobtype;
run;

```

```

proc transpose data=ex2 out=wide4 prefix = mainjob;
    by pidseq;
    id n;
    var mainjob;
run;

```

```

proc transpose data=ex2 out=wide5 prefix = jobst;
    by pidseq;
    id n;
    var jobst;
run;

```

```

data ex_j2;
    merge    wide1(drop=_name_)    wide2(drop=_name_)    wide3(drop=_name_)
wide4(drop=_name_) wide5(drop=_name_);
    by pidseq;
    drop _LABEL_;
    /*일자리 갯수*/
    mjob=10- nmiss(jobwave1-jobwave10);
run;

```

< SPSS >

/\* 22차년도에 동시에 몇개의 일자리를 가지고 있는가? (pid를 기준으로 reshape) \*/

GET FILE='C:\klips\klips22w.sav'.

select if (jobwave =22).

select if not (jobcens =0).

select if (jobclass =1 or jobclass =3 or jobclass =5 or jobclass =7).

compute jobst =4.

if (j150 =1 or j150 =3) jobst =j150.

if (j150 = 5) jobst =6.

if (j150 =4 and (j502 =2 or j503 =0)) jobst =5.

format jobst(F9.0).

MATCH FILES file =\*

/KEEP = pid jobwave jobseq jobnum.jobtype jobclass mainjob jobcens jobst.

if \$casenum = 1 or lag(pid) ne pid count1 = 1.

if pid = lag(pid) count1 = sum(lag(count1),1).

format count1(F9.0).

fre count1.

casestovars

/id = pid

/index = count1.

execute.

compute mjob =nvalid(jobnum.1, jobnum.2, jobnum.3).



format mjob(F9.0).

execute.

SAVE OUTFILE= 'D:\22nd\users guide\22nd\ex\_j1.sav'.

GET FILE='D:\22nd\users guide\22nd\klips22w.sav'.

select if (jobwave =12 or jobwave =13 or jobwave =14 or jobwave =15 or jobwave =16 or jobwave =17  
or jobwave =18 or jobwave =19 or jobwave =20 or jobwave =21 or jobwave =22).

select if not (jobcens =0).

select if (jobclass =1 or jobclass =3 or jobclass =5 or jobclass =7).

compute jobst =4.

if (j150 =1 or j150 =3) jobst =j150.

if (j150 = 5) jobst =6.

if (j150 =4 and (j502 =2 or j503 =0)) jobst =5.

format jobst(F9.0).

MATCH FILES file =\*

/KEEP = pid jobwave jobseq jobnum.jobtype jobclass mainjob jobcens jobst.

SORT CASES BY PID, jobseq.

if \$casenum = 1 or lag(pid) <> pid or lag(jobseq) <> jobseq count2 = 1.

if pid = lag(pid) and jobseq = lag(jobseq) count2 = sum(lag(count2),1).

execute.

format count2(F9.0).

fre count2.

casestovars

/id = pid jobseq

/index = count2.

execute.

```
compute mjob =nvalid(jobwave.1, jobwave.2, jobwave.3, jobwave.4, jobwave.5, jobwave.6, jobwave.7,  
jobwave.8, jobwave.9, jobwave.10).  
format mjob(F9.0).  
execute.
```

SAVE OUTFILE= 'D:\22nd\users guide\22nd\ex\_j2.sav'.

## < STATA >

\*\* 22차년도에 동시에 몇개의 일자리를 가지고 있는가? (pid를 기준으로 reshape) \*\*

```
use klips22w
```

```
sort pid jobwave
```

```
tab jobwave
```

```
keep if jobwave==22 /* 22차년도 응답한 일자리만 남김*/
```

```
drop if jobcens==0 /* 회고적 자료는 제외(신규)*/
```

```
keep if inlist(jobclass,1,3,5,7) /*22차 당시 진행중인 일자리만 추출*/
```

\* 종사상지위만들기.

```
gen jobst=4 /*고용주*/
```

```
replace jobst = j150 if inrange(j150,1,3) /*임금근로자*/
```

```
replace jobst=6 if j150 ==5 /*무급가족종사자*/
```

```
replace jobst=5 if j150 ==4 & (j502 ==2 | j503 ==0) /*자영업: 종업원수가 0명인 경우*/
```

```
tab jobst
```

```
sort pid jobseq
```

```
keep pid jobwave jobseq jobnum jobtype jobclass mainjob jobcens jobst
tab jobwave
```

\* 동일한 개인에 대해서 옆으로 데이터 변환.

```
bysort pid: gen sub = _n /* _n: 동일한 pid에 대해 반복되는 pid숫자를 모두 산출함 */
tab sub
```

```
reshape wide jobwave jobseq jobcens jobtype jobclass jobnum mainjob jobst , i( pid ) j(sub)
```

\* 일자리 개수.

```
egen mjob=rownonmiss(jobnum1 jobnum2 jobnum3) /* missing이 아닌 행의 숫자를 산출 */
tab mjob
save ex_j1, replace
```

\*\* 일생동안 몇개의 일자리를 가지고 있는가? (pid와 jobseq를 기준으로 reshape) \*\*

```
use klips22w, clear
set more off
```

```
drop if jobcens==0 /* 회고적 자료는 제외(신규) */
keep if inlist(jobwave, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22) /* 12-22차 당시 진행중인 일자리만 추출 */
```

```
sort pid jobseq
tab1 jobwave jobseq
```

\* 종사상지위만들기.

```
gen jobst=4 /* 고용주 */
replace jobst = j150 if inrange(j150,1,3) /* 임금근로자 */
replace jobst=6 if j150==5 /* 무급가족종사자 */
replace jobst=5 if j150==4 & (j502==2 | j503==0) /* 자영업: 종업원수가 0명인 경우 */
```

```
sort pid jobseq
keep jobwave jobtype mainjob jobcens jobst pid jobseq
```

\* 동일한 일자리를 옆으로 데이터 변환.

```
bysort pid jobseq : gen sub = _n /* _n: 동일한 pidseq에 대해 반복되는 pidseq숫자를 모두 산출함 */
tab sub
```

```
reshape wide jobwave jobcens jobtype mainjob jobst , i( pid jobseq ) j(sub)
```

\*일자리 갯수.

```
egen mjob=rownonmiss(jobwave1 jobwave2 jobwave3 jobwave4 jobwave5 jobwave6 jobwave7
jobwave8 jobwave9 jobwave10) /*(위의 반복되는 pid숫자를 기반으로 구성) missing이 아닌 행의 숫자를 산출
*/
tab mjob
```

```
save ex_j2, replace
```

<R>

```
library(foreign)
library(dplyr)
klips22w <- read.spss(file = "C:\\klips\\klips22p.sav",
use.value.labels = FALSE, to.data.frame = TRUE )

# 22차년도에 동시에 몇개의 일자리를 가지고 있는가? (pid를 기준으로 reshape)

q301 <- subset(klips22w, jobwave ==22 & jobcens !=0 & (jobclass == 1 | jobclass ==3 | jobclass ==5 |
jobclass ==7))

q301$jobst <- ifelse(q301$j150 ==1 | q301$j150 ==2, q301$j150,
```

```

        ifelse(q301$j150 ==5, 6,
              ifelse(q301$j150 ==4 & (q301$j502 ==2 | q301$j503 ==0), 5, 4)))

q301 <- select(q301, pid, jobwave, jobseq, jobnum, jobtype, jobclass, mainjob, jobcens, jobst)
q301 <- data.frame(q301, sub = as.factor(unlist(lapply(table(q301$pid), seq_len))))

q301 <- reshape(q301,idvar=c("pid"), timevar="sub", direction="wide")

q3011 <- select(q301, jobnum.1, jobnum.2, jobnum.3)
cc <- apply(q3011, 1, function(x) length(which(!is.na(x))))
q301$mjob <- cc

# 일생동안 몇개의 일자리를 가지고 있는가? (pid와 jobseq를 기준으로 reshape)

q302 <- subset(klips22w, jobcens !=0 &
              (jobwave ==12 | jobwave ==13 | jobwave ==14 | jobwave ==15 | jobwave ==16 | jobwave ==17 |
               jobwave ==18 | jobwave ==19 | jobwave ==20 | jobwave == 21 | jobwave == 22))

q302$jobst <- ifelse(q302$j150 ==1 | q302$j150 ==2, q302$j150,
                  ifelse(q302$j150 ==5, 6,
                        ifelse(q302$j150 ==4 & (q302$j502 ==2 | q302$j503 ==0), 5, 4)))
q302 <- select(q302,pid, jobwave, jobseq, jobnum, jobtype, jobclass, mainjob, jobcens, jobst)
q302 <- q302[order(q302$pid, q302$jobseq),]
q302$pida <-sprintf("%010d", q302$pid)
q302$jobseqa <-sprintf("%03d", q302$jobseq)
q302$pidjobseq <-paste0(q302$pida, q302$jobseqa)
q302 <- q302[order(q302$pidjobseq),]
q302 <- data.frame(q302, sub2 = as.factor(unlist(lapply(table(q302$pidjobseq), seq_len))))
q302 <- select(q302, pid, sub2, jobwave, jobseq, jobnum, jobtype, jobclass, mainjob, jobcens, jobst)
q302 <- reshape(q302,idvar=c("pid", "jobseq"), timevar="sub2", direction="wide")

```

```
q3021 <- select(q302, jobwave.1, jobwave.2, jobwave.3, jobwave.4, jobwave.5, jobwave.6, jobwave.7,  
jobwave.8, jobwave.9, jobwave.10)  
cc <- apply(q3021, 1, function(x) length(which(!is.na(x))))  
q302$mjob <- cc
```

## <부록> 한국노동패널 1~22차년도 조사자료 SMART\_KLIPS User's Guide<sup>38)</sup>

### A.1 모듈 설치절차

SMART\_KLIPS 모듈과 SMART\_KLIPS\_ADD 모듈을 사용하기 위해서는 표 1-1의 파일들이 필요하다. \*.ado 파일은 명령어가 코딩된 파일이고 \*.dlg 파일은 명령어를 대화창으로 실행하기 위해 필요한 dialog box 디자인 파일이다. <부표 1-1>의 모든 파일이 영문자 알파벳 "s"로 시작하기 때문에 c:\ado\s 폴더에 복사하면 설치작업이 완료된다.<sup>39)</sup> c:\ado\s 폴더에 이미 같은 이름의 파일이 있다면 덮어쓰기(overwrite)를 하는 것을 추천한다.

<부표1-1> Stata 파일목록

ado file	그 외 file
smart_h.ado smart_p.ado smart_klips.ado smart_klips_add.ado smart_search.ado	smart_klips.dlg smart_klips.sthlp smartklips_users_guide.pdf

SMART\_KLIPS 모듈을 실행하기 위해서는 원 데이터가 필요하다. 원 데이터는 반드시 \*.dta 파일형식이여야 하고 <부표 1-2>와 같은 파일을 이름을 가져야 한다.<sup>40)</sup> 2020년 9월 현재 한국노동연구원 노동패널 홈페이지(<http://www.kli.re.kr/klips>)에서 학술대회 신청자에 한하여 1차~22차 가구와 가구원 데이터를 다운받을 수 있다.

38) 민인식교수(경희대학교 경제학과) 연구용역으로 개발된 것이다.

39) c:\ado\s 폴더가 사용자 컴퓨터에 없다면 임의로 만들면 된다.

40) 14버전 이상 Stata \*.dta 파일로 저장하는 것을 추천한다.

<부표 1-2> 가구와 가구원 데이터 파일목록

wave	가구	가구원
1차 (1998년)	klips01h.dta	klips01p.dta
2차 (1999년)	klips02h.dta	klips02p.dta
3차 (2000년)	klips03h.dta	klips03p.dta
4차 (2001년)	klips04h.dta	klips04p.dta
5차 (2002년)	klips05h.dta	klips05p.dta
6차 (2003년)	klips06h.dta	klips06p.dta
7차 (2004년)	klips07h.dta	klips07p.dta
8차 (2005년)	klips08h.dta	klips08p.dta
9차 (2006년)	klips09h.dta	klips09p.dta
10차 (2007년)	klips10h.dta	klips10p.dta
11차 (2008년)	klips11h.dta	klips11p.dta
12차 (2009년)	klips12h.dta	klips12p.dta
13차 (2010년)	klips13h.dta	klips13p.dta
14차 (2011년)	klips14h.dta	klips14p.dta
15차 (2012년)	klips15h.dta	klips15p.dta
16차 (2013년)	klips16h.dta	klips16p.dta
17차 (2014년)	klips17h.dta	klips17p.dta
18차 (2015년)	klips18h.dta	klips18p.dta
19차 (2016년)	klips19h.dta	klips19p.dta
20차 (2017년)	klips20h.dta	klips20p.dta
21차 (2018년)	klips21h.dta	klips21p.dta
22차 (2019년)	klips22h.dta	klips22p.dta
총 44개 데이터 파일		

SMART\_KLIPS 모듈을 실행하기 위해서는 반드시 Stata 14.0 이상 버전이 필요하다.<sup>41)</sup> Stata 11, 12, 13 버전에서는 실행되지 않음에 주의해야 한다.

**Stata14.0버전(IC,SE또는MP)이상**

41) 13버전 사용자는 노동패널팀([klips@kli.re.kr](mailto:klips@kli.re.kr))로 문의.



## A.2 모듈 사용방법 및 주의사항

SMART\_KLIPS 모듈을 사용하기 위해서는 A.1에서 설명한대로 필요한 파일을 설치해야 한다. 그런 다음 [그림 2-1]와 같은 대화창(dialog box)을 불러오는 명령어를 command 창에 아래와 같이 입력한다<sup>42)</sup>.

```
. db smart_klips
```

[그림 2-1] SMART KLIPS 대화창

KLIPS 패널데이터 만들기 v.2020-12-30

select command : ☒ smart\_klips ☐ smart\_search ☐ smart\_klips\_add

**smart\_klips**

가구레벨변수

가구원레벨변수

save file  ☐ wide panel

**smart\_search**

h or p (REQUIRED) ☒ household ☐ person word:

**smart\_klips\_add**

mastertype (REQUIRED) ☒ household ☐ person

addtype (REQUIRED) ☒ household ☐ person

추가하고자 하는 변수이름 (REQUIRED)

master file  save file

Waves: 01~22 (REQUIRED)

wd:

website (http:// 로 시작) :

?

42) stata 프로그램은 소문자와 대문자를 구분하므로, 꼭 소문자로 입력해야 한다.

## 1) smart\_klips

대화창 상단에 위치한 select command에서 연구자가 사용하고자 하는 명령어를 선택한다. "smart\_klips"를 선택하면 저자가 미리 코딩해 놓은 변수를 선택하여 데이터 파일로 만들 수 있다. 가구레벨 변수와 가구원 레벨 변수에서 연구자가 임의로 변수를 선택하면 된다. 그런 후 wave 옵션에서 패널데이터로 만들고자 하는 wave를 선택해 주면 된다. wd 옵션에서는 <표 1-2>의 klips\*.dta 파일이 저장되어 있는 폴더를 지정해 주면 된다<sup>43)</sup> 완성된 파일은 klips\_final.dta 파일로 저장되어 있다.

## 2) smart\_search

대화창 상단에 위치한 select command에서 smart\_search 명령어를 선택하면, KLIPS 1차 ~ 22차 데이터에서 연구자가 찾고자 하는 변수이름을 확인할 수 있다. 찾고자 하는 변수가 가구데이터에 있으면 "household"를 선택하고 가구원 데이터에 있다면 "person"을 선택한다. word 옵션에서는 찾고자 하는 변수레이블(variable label)의 일부분을 입력하면 된다. 가령 "소득"이라고 입력하면 변수레이블이 "소득"을 포함하는 모든 변수이름을 보여준다. 또한 wave( ) 옵션에서 wave를 지정해주면 해당 wave에서 "소득"을 포함한 변수리스트를 화면에 보여준다. 주의할 점은 word 옵션에서 단어를 입력할 때 띄어쓰기를 하면 에러가 발생한다. 가령 "부동산 소득"이라고 쓰면 명령어가 실행되지 않는다. "부동산 소득" 대신 "부동산소득"이라고 입력해야 한다.<sup>44)45)</sup>

---

43) wd 옵션을 사용하는 대신 Stata 명령문으로 working directory를 변경하는 명령어를 먼저 실행한 후 smart\_klips를 사용해도 된다. 가령 c:\temp에 KLIPS \*.dta 데이터가 저장되어 있다면 다음과 같이 입력한다.

· cd c:\temp

44) 설사 원 데이터의 변수레이블에 "부동산 소득"이라도 되어 있을지라도 word 옵션에서 "부동산 소득"이라고 입력하면 해당 변수를 찾아주도록 코딩되어 있다.

45) 결과창에 한글이 깨져서 나오는 경우에는 tool bar에서 화면 설정을 변경해 주어야 한다. Edit ☐ Preferences ☐ General preferences ☐ Color scheme을 Classic으로 변경한다.

### 3) smart\_klips\_add

대화창 상단의 select command에서 "smart\_klips\_add"를 선택하면 연구자가 임의로 선택한 변수를 추가하여 패널데이터를 완성할 수 있다. mastertype() 옵션에서는 smart\_klips 명령어로 만든 klips\_final.dta 파일이 가구레벨인지 가구원레벨인지를 지정한다. 가구레벨이면 household를 선택하고 가구원레벨 데이터이면 person을 선택한다.<sup>46)</sup> addtype() 옵션에서는 연구자가 임의로 추가하고자 변수가 가구레벨 데이터(klips\_h)에서 가져오는지 가구원레벨 데이터(klips\_p)에서 가져오는지를 지정한다. klips\_h 데이터에서 가져오고자 하면 household를 선택하고 klips\_p 데이터에서 가져오고자 하면 person을 선택한다. 앞선 명령어들과 마찬가지로 wave 옵션을 지정해주면 해당 wave에서 추가하고자 변수를 가져와서 패널데이터로 만들어 준다. 마지막에 저장되는 파일명은 klips\_final\_add.dta 파일이 된다.

"smart\_klips\_add"에서 가장 중요한 부분은 추가하고자하는 변수명을 입력하는 방법이다. KLIPS 원 데이터에서 변수이름은 wave가 포함되어 있다. wave에 해당하는 부분을 제외한 나머지 이름을 그대로 써주면 된다. 가령 1차 ~ 5차 가구레벨 데이터에 있는 h010142~h050142와 h011415~h051415 변수를 가져오고자 하면 h0142와 h1415라고 입력하면 된다. klips\_p(가구원 데이터)에 있는 변수에 대해서 같은 방식을 적용한다.

h010142 h020142 h030142 h040142 h050142	h0142
h011415 h021415 h031415 h041415 h051415	h1415

주의할 점은 연구자가 입력한 변수와 wave가 일치해야 한다는 것이다. 즉 h0142라고 쓰고 wave(01 02 03 04 05 06) 옵션을 사용하면 h010142~h060142 변수가 반드시 원 데이터에 존재해야 한다. 만약 h0142 변수가 1차 ~ 5차 조사에만 존재하고 6차에는 존재하지 않는다면 wave(01 02 03 04 05)라고 입력해야 에러가 발생하지 않는다. 연구자가 추가하고자 하는 변수가 어떤 wave에 있는지 정확히 파악한 후 **smart\_klips\_add**를 사용해야 한다. 먼저 smart\_search를 이용하여 찾고자 변수이름을 먼저 확인한 후 **smart\_klips\_add**를 사용

46) 가구레벨과 가구원레벨을 동시에 포함된 데이터는 당연히 가구원레벨 데이터로 간주한다.

하는 것을 추천한다.




마지막으로 **smart\_klips\_add**를 반복적으로 사용하는 경우를 설명하고자 한다. **smart\_klips\_add**에서는 `addtype( )`에서 `household` 또는 `person` 둘 중 하나만 지정할 수 있다. 만약 `klips_final.dta` 파일에 가구와 가구원 변수를 동시에 임의로 추가하고자 하면 명령문을 어떻게 작성해야 하는가? 다음과 같이 **smart\_klips\_add**를 반복적으로 사용할 수 있다.

먼저 `smart_klips` 명령어를 이용하여 `klips_final.dta` 파일을 생성한다. 두 번째 명령문에서는 가구레벨 데이터에서 `h0142`와 `h1414` 변수를 7차 ~ 12차까지 가져오고자 한다. `mastertype(h)`와 `addtype(h)`를 사용한다. 중요한 옵션은 `save(klips_add_1)`이다. `h0142`와 `h1415`가 추가된 패널데이터를 `klisp_add_1.dta` 이름으로 저장한다. 세 번째 명령문에서는 `p0107`과 `p0301` 변수를 7차 ~ 12차에서 가져와 추가하고자 한다. `addtype(p)`라고 입력해야 함에 주의해야 한다. 여기서 중요한 옵션은 `master(klips_add_1)`이다. 두 번째 저장한 파일 이름을 `master( )` 옵션에서 사용해야 한다. `save(klips_add_1)`은 세 번째 명령문까지 실행된 마지막 패널데이터를 `klips_add_1.dta`로 저장한다는 의미이다. 두 번째 명령문에서와 같은 파일을 계속 지정해도 무관하다. 기존 파일을 덮어쓰기(`overwrite`)하기 때문이다.




```
. smart_klips h_hsex h_hage , wave( 07 08 09 10 11 12)

. smart_klips_add h0142 h1415 , wave(07 08 09 10 11 12) mastertype(h)
  addtype(h) save(klips_add_1)

. smart_klips_add p0107 p0301 , wave(07 08 09 10 11 12) mastertype(h)
  addtype(p) master(klips_add_1) save(klips_add_1)
```

○ SMART\_KLIPS 대화창 TIP(    )

대화창의 왼쪽 맨 아래에 보면 위와 같은 3가지 선택메뉴가 있다.

	user's guide 연결되어 SMART_KLIPS 사용방법을 알 수 있다.
	select command에서 특정 명령어를 실행한 후 그 다음 명령어를 실행하기 위해서는 이미 입력했던 내용을 모두 지워야 하는 불편함이 있다."R"을 클릭하면 대화창에서 입력했던 내용이 모두 지워진 후 초기화면으로 넘어간다. 반복적으로 dialog box를 사용하고자 하면 "R"(reset) 기능을 적절히 활용할 수 있다.
	copy 기능으로 실행한 명령문을 복사할 수 있다. 유사한 작업을 계속하는 경우 이 기능을 활용하여, command 창에 직접 입력할 수도 있다.

## 찾아보기

(ㄱ)		대체가구접촉수	19
가구소득	71	대체성공가구비율	19
가구소비	74	대체성공가구수	19
가구용 설문	40	대학교명 코드	64
가구원 기본정보	70	동부	18
가구자료	44, 69	동시생태효과(cohort effect)	3
가구주와의 관계	54		
가중치	27	(ㄴ)	
개인용 설문	40	만나이	55
개인자료	44, 82	면접원 오류	104
개인효과(individual effects)	3	면접타계식	7
건강과 은퇴 자료	108	모집단	16
경제활동상태	57	목표표본가구	18
경제활동인구	59	무응답 대체 데이터(imputation)	65
고용형태 부가조사	115, 126	무작위 추출(random sampling)	16
교육 부가조사	115	미취업자	58
교육년수	57		
군부	16	(ㄷ)	
근로시간과 여가 부가조사	111	보육기관	79
금융자산	77	부가조사자료	44
		부동산	76
(ㄴ)		부채	77
노동조합과 노사관계 부가조사	112	분가가구	14
		비경제활동인구	59
(ㄷ)		비임금근로자 부가조사	128
대리응답	6		

( 人 )		전반적 직무만족도	96
사교육	79	전화조사	6
사회보험	71, 94	정규/비정규 여부	60
삶의 인식 부가조사	118	조사 기간	7
삶의 질 부가조사	117	조사구	16
상태 간 변이과정(transition between states)	3	조사내용	46
생활만족도	97	조사대상	4
세후소득	73	조사방법	4
시계열자료(time-series data)	1, 3	조사설계(survey design)	4
신규조사자	14	조사추적 원칙	5
실업자	59	조직몰입도	97
		종단면 가중치	27
( 〇 )		종사상 지위	60
연령	55	주거지	78
예비조사(pilot survey)	4	주택종류	78
오픈 코드(open code)	61	중고령자 부가조사	109
요인별 직무만족도	95	직무만족도(Job Satisfaction)	94
원표본가구	10	직업력(work history)	99
유치조사	7	직업력자료	44
유형설문	40	직업훈련	92
의사결정 부가조사	124	직업훈련 코드	66
이수여부	56	직접조사	6
입주형태	78		
		( ㄱ )	
( ㄱ )		청년층 부가조사	113, 122
자격증	63, 92	청년층 자료	106
자녀교육	79	총접촉가구수	19
자산	76	최초성공가구비율	19
장년층 부가조사	120	최초성공가구수	19
		최초접촉가구수	19

취업/미취업 상태	58	(1)	
취업자	58	18통합표본	4, 11
		1차 원가구원 여부	57
(ㄷ)		(2)	
특정기간효과(period specific effect)	3	2단계 층화집락계통추출법	16
특정연령효과(age effect)	3		
(ㄹ)		(9)	
표본설계방법	16	98표본	4
표본추출과정	16		
표본탈락(panel attrition)	30	(J)	
표본틀	16	jobcens	101
		jobclass	102
(ㅎ)		jobnum	101
학과 코드	56, 64	jobnumc	101
학교	56	jobseq	101
학년	56	jobtype	101
학력	56	jobwave	101
한국리서치	8	jubnumc	101
한국표준산업분류	65		
한국표준직업분류	65	(K)	
항목무응답(item nonresponse)	30	key 변수	50
혼인상태	55		
회고적(retrospective) 정보	99	(M)	
횡단면자료(cross-section data)	3	mainjob	102
(0)		(P)	
09통합표본	4, 11	pid	53



◆ 執筆陣

- 김유빈(한국노동연구원 연구위원)
- 이지은(한국노동연구원 전문위원)
- 신선옥(한국노동연구원 전문위원)
- 정현상(한국노동연구원 책임연구원)
- 권익성(한국노동연구원 연구원)