

2025년 한국복지패널 데이터설명회



한국복지패널 데이터 소개

이원진 한국보건사회연구원



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

LEADING THE WAY IN IMPROVING
PEOPLE'S QUALITY OF LIFE AND SOCIAL SECURITY

한국복지패널(KoWePS) 조사 목적

- 2006년부터 사회·경제적 환경변화에 따른 가구와 개인의 경제와 생활실태 변화
 - ✓ 국민의 생애주기별 삶의 역동성과 욕구의 동태적 변화
 - ✓ 저소득층의 소득, 경제활동 상태와 생활실태 동태적 변화
 - ✓ 정부의 다양한 복지 정책과 정책 변화를 조사
 - 공공부조, 사회보험, 사회서비스의 정책 변화 및 긴급재난지원금 조사



1. 한국복지패널 조사 목적

한국복지패널(KoWePS) 조사 목적

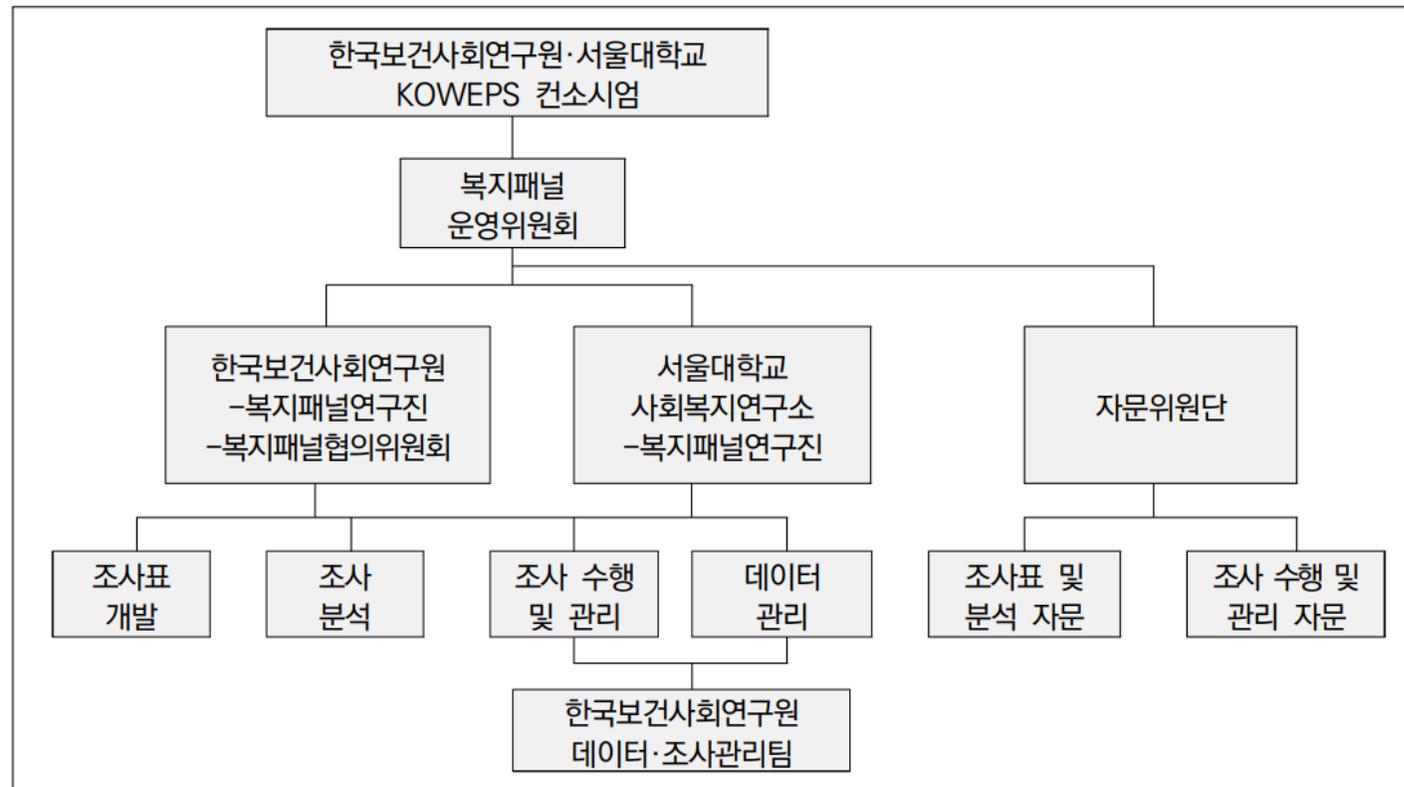
- 다양한 인구집단별 생활실태와 복지욕구 파악
 - ✓ 연령, 소득계층, 경제활동상태별 생활 및 복지욕구 파악
 - ✓ 변화하는 정부정책의 효과성 평가
- 새로운 제도형성과 정책 개선에 기여



2. 한국복지패널의 관리 및 운영체계

한국복지패널(KoWePS) 관리 및 운영체계

- 한국복지패널은 한국보건사회연구원과 서울대학교 사회복지연구소의 컨소시엄을 통해 시작하여 조사 연구의 시너지 효과를 기대



3. 한국복지패널 데이터 특징

한국복지패널(KoWePS) 데이터 특징(1)

- 전국 지역 대표성 높음
 - ✓ 지역적으로 제주도를 포함
 - ✓ 가구유형은 농어가 가구를 포함
- 종단면 데이터
 - ✓ 사회환경변화에 따른 개인이나 가구의 동태적 특성 분석이 가능

3. 한국복지패널 데이터 특징

한국복지패널(KoWePS) 데이터 특징(2)

- 저소득층 연구에 적합
 - ✓ 저소득층을 전체 표본의 약 50%로 할당하여 표본 추출
- 충분한 표본수
 - ✓ 다양한 집단으로 세분화하여 분석이 가능
 - ✓ 통계학적으로 표본오차가 작아지는 장점

3. 한국복지패널 데이터 특징

한국복지패널(KoWePS) 데이터 특징(3)

- 다양한 사회보장제도 관련 정보 조사
 - ✓ 사회보험, 공공부조, 사회서비스 등 사회보장 정보 조사
- 학제간 연구가 가능
 - ✓ 사회과학분야의 다양한 영역의 학자가 연구 및 분석 가능
- 공신력 있는 기관에서 데이터 생산 및 제공
 - ✓ 신뢰할 수 있고 일관성 있는 통계자료를 생산

감사합니다.



KIHASA
한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

LEADING THE WAY IN IMPROVING
PEOPLE'S QUALITY OF LIFE AND SOCIAL SECURITY

2025년 한국복지패널 데이터설명회



19차년도 한국복지패널 조사자료 설문지 구성

한국보건사회연구원 강예은 연구원



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

LEADING THE WAY IN IMPROVING
PEOPLE'S QUALITY OF LIFE AND SOCIAL SECURITY

CONTENTS.

CHAPTER.1

조사표의 구성

CHAPTER.2

가구 및 가구원의 개념

CHAPTER.3

조사내용

CHAPTER.4

자료 접근과 활용 시 유의 사항

1. 조사표의 구성

가구용

- I. 가구 일반 사항
- II. 건강 및 의료 A
- III. 경제활동 상태
- IV. 사회보험, 퇴직(연금), 개인연금 가입
- V. 의료 B
- VI. 주거
- VII. 생활비
- VIII. 소득
- IX. 부채, 이자
- X. 재산
- XI. 생활 여건
- XII. 국민기초생활보장
- XII-1. 근로(자녀)장려세제
- XIII. 가구의 복지 서비스 이용
- XIV. 노인 가구의 복지 서비스 이용
- XV. 아동 가구의 복지 서비스 이용
- XVI. 장애인 가구의 복지 서비스 이용
- XVII. 가족

가구원용

- A. 사회보험, 퇴직금, 개인연금 수급
- B. 근로
- C. 생활실태·만족 및 의식
- D. 사회적 환경에 대한 의식
- E. 생활 습관, 가족 관계 및 정신건강
- F. 교육
- G. 개인사



부가조사용 (아동)

- A. 현재 상태
- B. 나의 학교생활
- C. 나의 생각과 행동
- D. 우리 부모님은?
- E. 나의 친구는?
- F. 나의 건강 및 생활
- G. 나의 아르바이트 경험 및 활동
- H. 나의 진로는?



1. 조사표의 구성

| | 조사 대상 | 조사기준 연도 및 시점 |
|--|--|--|
|  가구용 | 가구주 또는 가구주의 배우자 | 2023. 1. 1. ~ 2023. 12. 31. 2023. 12. 31. |
|  가구원용 | 만 15세 이상 가구원 전체 <ul style="list-style-type: none">• 중학생, 고등학생 제외• 단, 2023.2월 고등학교 졸업한 가구원 포함 | 2023. 1. 1. ~ 2023. 12. 31. 2023. 12. 31. |
|  부가조사용 (아동) | 19차 아동 부가조사 응답자 대상 <p>조사대상: 18차연도 조사 완료된 가구원 중 4~6학년인 아동</p> <p>*아동부가조사: 1차/4차/7차/10차/13차/16차/19차</p> | 조사일 현재 |



2. 가구 및 가구원 개념

원가구

전년도 조사완료된 가구



신규가구

조사기준연도 9월 30일 이전에 분가·결혼·직장·학업·위탁·이혼·별거 등의 사유로 분리된 가구



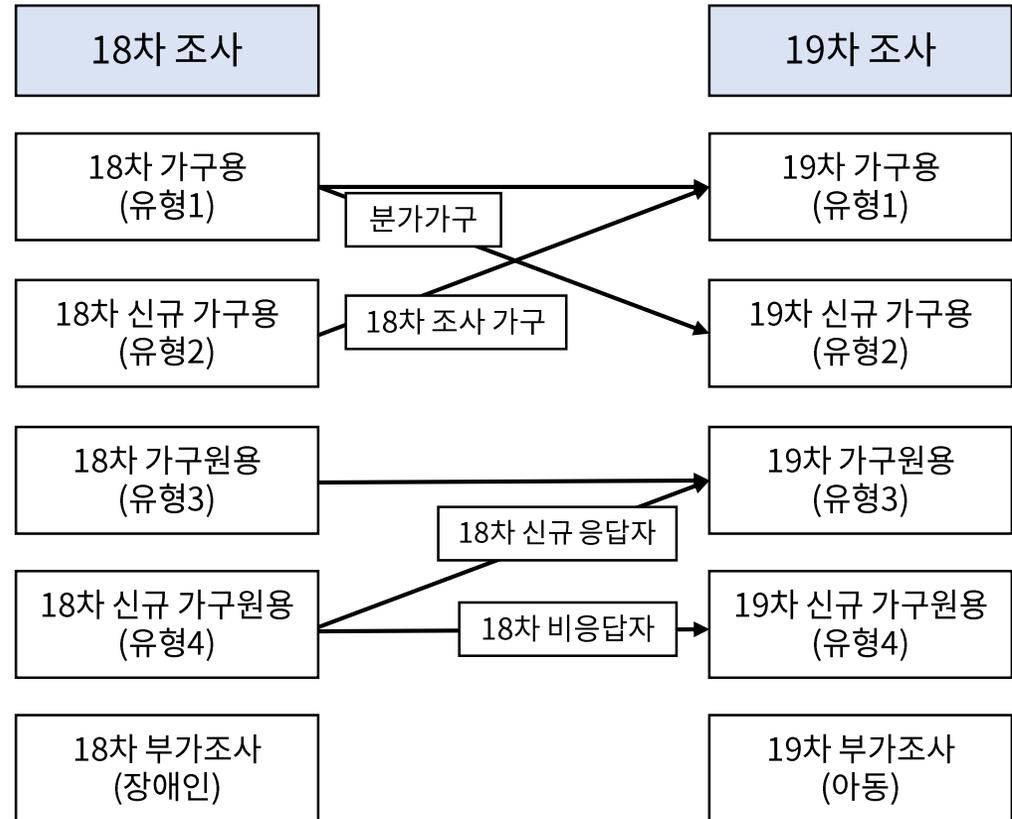
원가구원

전년도 가구원 조사완료된 가구원

신규가구원

전년도 가구원 조사에 응답하지 않은 가구원이면서, 다음에 해당하는 자

- ① 조사 기준 연도에 만 15세가 된 가구원이면서 중고등학생이 아닌 자
- ② 조사 기준 연도 2월에 고등학교를 졸업(중퇴 포함)한 가구원
- ③ 조사 기준 연도에 원가구 혹은 신규가구에 새로 진입한 가구원



3. 조사내용

3-1. 조사 영역

| 조사표 구분 | | 조사영역 | | |
|----------------|------------------|--|--|--|
| 가구용 (유형1,2) | 공통조사영역 (1~3차) | I. 가구일반사항 II. 건강 및 의료 A III. 경제활동상태 IV. 사회보험, 퇴직(연금), 개인연금 가입 V. 의료 B VI. 주거 | VII. 생활비 VIII. 소득 IX. 부채, 이자, 주관적 최저생계비 X. 재산 XI. 생활여건 XII. 국민기초생활보장 | XIII. 가구의 복지서비스 이용 XIV. 노인가구의 복지서비스 이용 XV. 아동가구의 복지서비스 이용 XVI. 장애인가구의 복지서비스 이용 XVII. 가족 |
| | 4차년도 변경영역 | XII-1. 근로장려세제(추가) | | |
| | 6차년도 변경영역 | XII-2. 희망키움통장(근로소득장려금)(추가) | | |
| | 7차년도 변경영역 | XII-2. 희망키움통장(근로소득장려금)(삭제) XII-2. 가구의 경제상황 및 복지수준에 대한 인식 및 전망(추가) | | |
| | 8차년도 변경영역 | XII-2. 가구의 경제상황 및 복지수준에 대한 인식 및 전망(삭제) | | |
| | 11차년도 변경영역 | XII-1. 국민기초생활보장(맞춤형급여, 2015년 7월 이후 기준)(추가) | | |
| | 12차년도 변경영역 | XII. 국민기초생활보장(2015년 7월 이전 기준)(삭제) | | |

| | | | | |
|-----------------|-----------|--|--|-----------------------------------|
| 가구원용 (유형3,4) | 1차년도 조사영역 | A. 사회보험, 퇴직금 개인연금 수급 B. 근로 | C. 생활실태 · 만족 및 의식 D. 사회적 환경에 대한 의식 | E. 생활습관, 가족관계 및 정신건강 F. 개인사 |
| | 2차년도 변경영역 | F. 개인사('유형3:가구원용'에서 삭제, '유형4:신규가구원용'에만 포함) | | |
| | 3차년도 변경영역 | F. 교육('유형4:신규가구원용'에 추가) → G. 개인사 | | |

| | | | |
|------------------|---|--|--------------------------------------|
| 아동 부가조사 (유형5) | A. 현재 상태 B. 나의 학교생활 C. 나의 생각과 행동 | D. 우리 부모님은? E. 나의 친구는? F. 나의 건강 및 생활 | G. 나의 아르바이트 경험 및 활동 H. 나의 진로는? |
|------------------|---|--|--------------------------------------|

3-2 조사 주제



인구사회학적배경

- ✓ 가구원 일반사항
- ✓ 부모세대
- ✓ 개인사



경제활동 상태

- ✓ 직업이력
- ✓ 경제활동상태
- ✓ 고용지원프로그램수급



가구 및 개인의 경제 상황

- ✓ 소득
- ✓ 재산 및 부채, 이자
- ✓ 지출 및 저축



사회보장제도 가입 및 수급 상태

- ✓ 사회보험
 - 건강보험, 고용보험
 - 산재보험, 국민연금
 - 개인연금, 퇴직연금 등
- ✓ 국민기초생활보장제도
- ✓ 근로(자녀)장려세제
- ✓ 기타 사회복지서비스
 - 노인, 장애인, 아동가구 등



가구여건 (복지욕구)

- ✓ 보육 및 교육
- ✓ 건강 및 의료
- ✓ 주거
- ✓ 가족관계, 사회적관계망
- ✓ 생활습관 및 정신건강
- ✓ 복지, 사회, 정치문제인식 등

4. 자료 접근과 활용 시 유의 사항

유의 사항

[15차 일시삭제 문항 유의]

- 코로나 19로 문항 축소 조사, 16차에 복원함

[17차 신규표본 추가]

- 추가패널 대상자는 모두 신규가구(원)용 조사 응답
- 신규가구원용 교육(F) 및 개인사(G) 영역 조사 안함

[3년 주기 문항 유의]

- (가구용) 건강 및 의료A 주요병명
- (가구용) 의료B 건강보험 만족도
- (가구원용) 사회적 자본
- (가구원용) 성역할



4. 자료 접근과 활용 시 유의 사항

자료 접근과 데이터 파일에 대한 이해 (1)

- 한국복지패널 홈페이지에서 차수별 데이터와 결합데이터를 다운로드 받을 수 있음.
- 파일 형식은 SAS, SPSS, STATA로 제공

[데이터 파일 설명 (19차)]

- 1. 가구용: [koweps_h19_2024_Beta1](#)
- 2. 가구원용: [koweps_p19_2024_Beta1](#)
- 3. 아동설문: [koweps_c19_2024_Beta1](#)
- 4. 가구용, 가구원용, 장애인설문용 머지데이터 [koweps_hpc19_2024_Beta1](#)

13차 웨이브 >

- (2018년 13차 한국복지패널조사) 데이터 (beta7)_stata, zip
- (2018년 13차 한국복지패널조사) 데이터 (beta7)_spss, zip
- (2018년 13차 한국복지패널조사) 데이터 (beta7)_sas, zip

14차 웨이브 >

- (2019년 14차 한국복지패널조사) 데이터 (beta6)_stata, zip
- (2019년 14차 한국복지패널조사) 데이터 (beta6)_spss, zip
- (2019년 14차 한국복지패널조사) 데이터 (beta6)_sas, zip

15차 웨이브 >

- (2020년 15차 한국복지패널조사) 데이터 (beta5)_stata, zip
- (2020년 15차 한국복지패널조사) 데이터 (beta5)_spss, zip
- (2020년 15차 한국복지패널조사) 데이터 (beta5)_sas, zip

16차 웨이브 >

- (2021년 16차 한국복지패널조사) 데이터 (beta4)_stata, zip
- (2021년 16차 한국복지패널조사) 데이터 (beta4)_spss, zip
- (2021년 16차 한국복지패널조사) 데이터 (beta4)_sas, zip

17차 웨이브 >

- (2022년 17차 한국복지패널조사) 데이터 (beta3)_stata, zip
- (2022년 17차 한국복지패널조사) 데이터 (beta3)_spss, zip
- (2022년 17차 한국복지패널조사) 데이터 (beta3)_sas, zip

18차 웨이브 >

- (2023년 18차 한국복지패널조사) 데이터 (beta2)_stata, zip
- (2023년 18차 한국복지패널조사) 데이터 (beta2)_spss, zip
- (2023년 18차 한국복지패널조사) 데이터 (beta2)_sas, zip

19차 웨이브 >

- (2024년 19차 한국복지패널조사) 데이터 (beta1)_stata, zip
- (2024년 19차 한국복지패널조사) 데이터 (beta1)_spss, zip
- (2024년 19차 한국복지패널조사) 데이터 (beta1)_sas, zip

차수별 데이터(1-19차) >

- 한국복지패널_1_19차_차수별_데이터 (release250327)stata, zip, zip
- 한국복지패널_1_19차_차수별_데이터 (release250327)spss, zip, zip
- 한국복지패널_1_19차_차수별_데이터 (release250327)sas, zip, zip

결합 데이터(1-19차) >

- 한국복지패널_1_19차_결합데이터 _STATA, zip
- 한국복지패널_1_19차_결합데이터 _SPSS, zip
- 한국복지패널_1_19차_결합데이터 _SAS, zip

4. 자료 접근과 활용 시 유의 사항

자료 접근과 데이터 파일에 대한 이해 (2)

[결합데이터 파일 설명]

- 1. 1~19차 가구데이터를 세로 결합한(밑으로 붙인) 파일:
[koweps_h01_19_long_240327](#)
- 2. 1~19차 가구데이터를 가로 결합한(옆으로 붙인) 파일:
[koweps_h01_19_wide_240327](#)
- 3. 1~19차 머지데이터(가구+가구원+부가조사 데이터의 결합본)를 세로 결합한(밑으로 붙인) 파일:
[koweps_hp01_19_long_240327](#)
- 4. 1~19차 머지데이터(가구+가구원+부가조사 데이터의 결합본)를 가로 결합한(옆으로 붙인) 파일:
[koweps_hp01_19_wide_240327](#)



4. 자료 접근과 활용 시 유의 사항

유저가이드 활용

[유저가이드 변화]

- 1~19차 가구(원)용 조사표 변경내역과 1~19차 변수이력 표는 **엑셀파일**로 별도 제공
- 변수 활용 시 유의사항은 기존 유저가이드 참조

[유저가이드의 구성]

- 제1장 조사개요
- 제2장 표본설계 및 가중치
- 제3장 설문 구성 및 내용
- 제4장 데이터 설명



4. 자료 접근과 활용 시 유의 사항

유저가이드 활용

[유저가이드의 데이터 설명]

- 기본 팁
- N차 이후 달라진 점
- 변수 활용시 유의점
- 모름/무응답 값

2. 건강 및 의료 A

[기본 팁]

Koweps 18차년도 자료에서 건강 및 의료 A 항목은 '2022년 1년 중('22.01.01~12.31)'시점을 기준으로 조사하였음. 단, 건강상태(h1802_2)와 민간의료보험 가입전수(h1802_3aq1)의 경우는 2022년 12월 31일을 기준시점으로 조사하였음.

[1,2차년도 이후 달라진 점]

Koweps 3차년도 자료에서부터 건강 및 의료 A의 문7) 주요 병명의 코드분류 항목을 추가하였음. 이 문항은 14차년도부터 원가구 3년 주기로 변경하여 원가구의 주요 병명은 조사하지 않았으며(단, 신규가구는 조사함), 15차년도부터는 신규가구도 원가구와 동일하게 3년 주기 문항으로 변경하여 원가구, 신규가구 모두 조사하지 않음. 지난 16차에 원가구와 신규가구 모두 조사하였으며, 17~18차에는 조사하지 않음. 한편, 3차 때부터 문8) 민간의료보험 가입전수 문항이 추가되었음.

[변수 활용시 유의점]

- SYSMIS=, , 모름/무응답=9 또는 99 또는 999
- 외래진료횟수(h1802_3)의 경우는 하루에 동일병원 각각 다른 과를 2회 이상 진료하면 1회로 기록하며, 2회 이상이 병원은 1회로 기록하며, 2회로 기록한(병원)은 기준으로 한, 또한

4. 자료 접근과 활용 시 유의 사항

유저가이드 활용

[유저가이드(엑셀 파일)의 구성]

- 가구용 조사표 변경사항
- 가구원용 조사표 변경사항
- 부가조사 조사표 변경사항
- 가구용데이터 변수이력표
- 가구원용데이터 변수이력표
- 아동 부가조사 변수이력표



4. 자료 접근과 활용 시 유의 사항

원척도가 있는 변수의 활용

[생활습관, 가족관계 및 정신건강]

✓ 음주, 우울, 자아존중감

[예시: 음주]

✓ 세계보건기구의 AUDIT(Alcohol Use Disorder Identification Test)를 사용

✓ 코딩 변경한 10개 변수의 값을 합하여, 총점이 높을수록 위험하고 해로운 음주행위를 할 가능성이 높다고 평가함.

| 원척도(AUDIT) 구분 | | 변수명 | 0점 | 1점 | 2점 | 3점 | 4점 |
|---------------|---------------|------------|----|----|----|----|----|
| 해로운 음주행 동 | 음주의 빈도 | p1705_2 | ⑤ | ① | ② | ③ | ④ |
| | 음주량 | p1705_3 | ① | ② | ③ | ④ | ⑤ |
| | 고위험음주의빈도 | p1705_4aq1 | ① | ② | ③ | ④ | ⑤ |
| 알코올 의존 | 음주에 대한 통제력 상실 | p1705_4aq2 | ① | ② | ③ | ④ | ⑤ |
| | 음주에 대한 증대된 동기 | p1705_4aq3 | ① | ② | ③ | ④ | ⑤ |
| | 해장술 | p1705_4aq4 | ① | ② | ③ | ④ | ⑤ |
| 위험한 음주 | 음주 후 후회감 | p1705_4aq5 | ① | ② | ③ | ④ | ⑤ |
| | 취중의 일을 기억 못 함 | p1705_4aq6 | ① | ② | ③ | ④ | ⑤ |
| | 음주관련상해 | p1705_4aq7 | ① | - | ② | - | ③ |
| | 음주관련 문제경험 | p1705_4aq8 | ① | - | ② | - | ③ |

감사합니다.



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

LEADING THE WAY IN IMPROVING
PEOPLE'S QUALITY OF LIFE AND SOCIAL SECURITY

한국복지패널 가중치 조정 과정

신재동 (한국보건사회연구원)

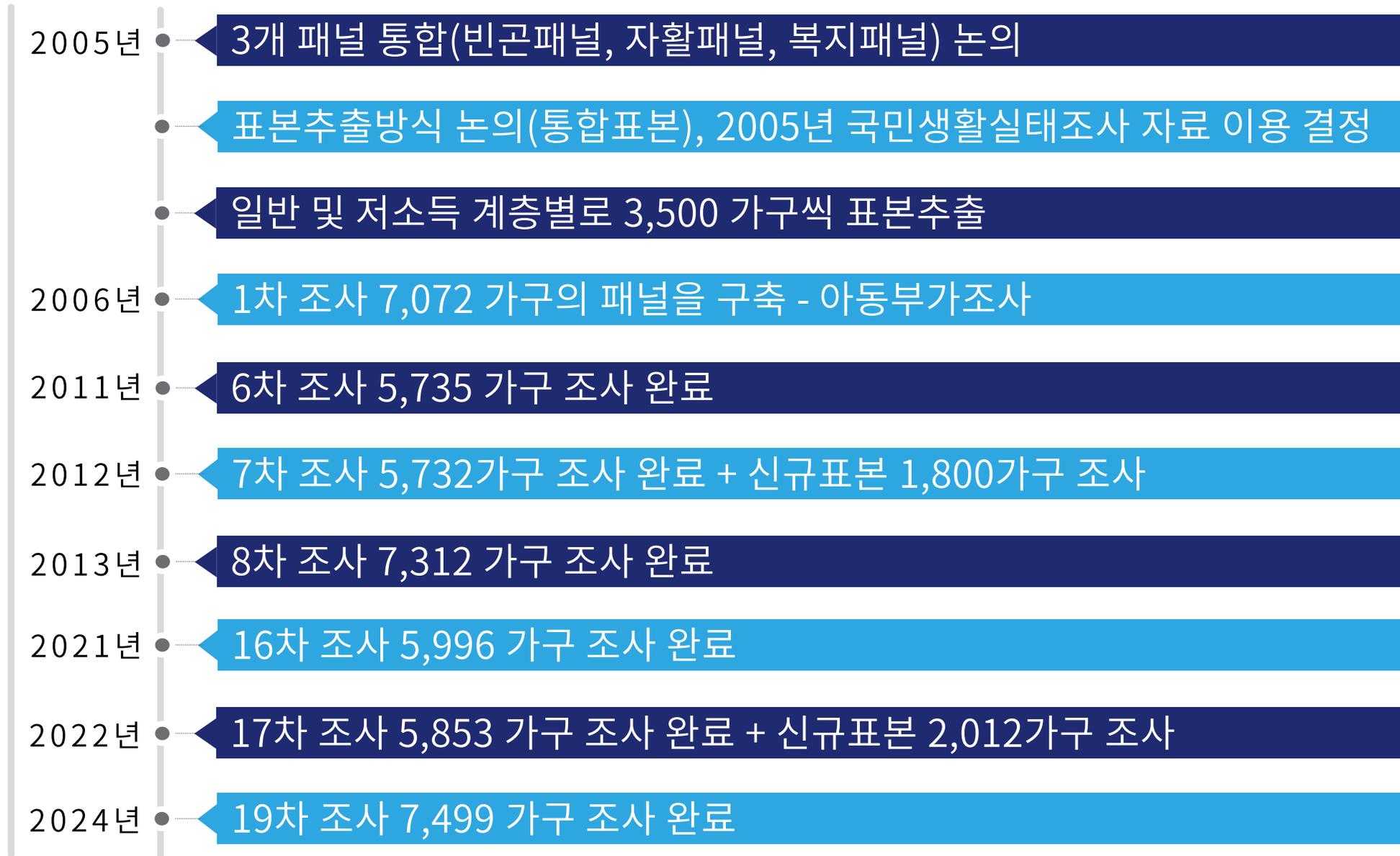


한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

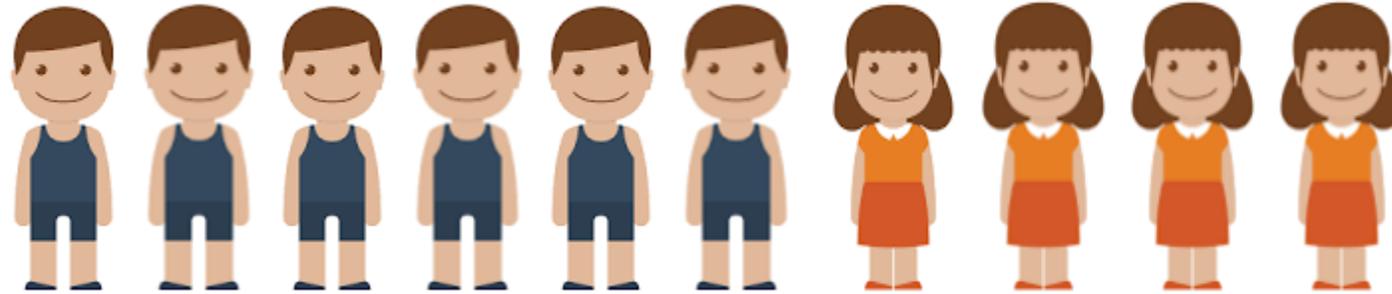


1. 한국복지패널 가중치 부여체계





모집단
N : 10



표본
N : 6



모수 가중치

[남 : 1.5]

[여 : 2]

[총합 : 10]

표본 가중치

[남 : 0.9]

[여 : 1.2]

[총합 : 6]



가중치 부여 절차

- 가구가중치 (household weight)는 횡단면과 종단면 가중치를 구분하지 않고 단일가중치를 부여하며, 개인가중치 (person weight) 만 횡단면 가중치와 종단면 가중치로 구분하여 부여
- 1차 웨이브에 부여된 개인가중치를 바탕으로 개인조사표에 응답한 가구원에 대해 가중치 조정을 수행
- 종단면 가중치는 로지스틱 회귀분석을 통해 각 개인의 응답확률을 추정하여 적용

개인가중치 산출 과정

- 2차년도 이후에는 t-1차년도 종단면 가중치를 기본 가중치로 고려하여 t차년도 원표본 해당여부 판정
- 무응답 및 개인별 변동 상황을 반영하여 각 년도 종단면 기본 가중치 보정
- 사후조정

개인 종단가중치 산출 과정

- 패널 탈락으로 인한 무응답 보정을 위해, t차 웨이브 종단면 응답여부 변수와 t-1차 웨이브 변수들의 관계를 로지스틱 회귀모형을 이용하여 응답확률을 추정
- 로지스틱 회귀모형의 설명변수로 응답자의 성별, 연령, 지역, 교육수준, 경제활동상태 사용
- t차 웨이브 종단면 응답여부 변수는 t차와 t-1차 웨이브 모두 응답한 경우 1, 그렇지 않은 경우는 0을 부여
- t차 웨이브의 개인 종단면 기본가중치에 로지스틱 회귀모형으로 추정된 응답확률의 역수를 곱하여, t차 웨이브의 개인 종단면 가중치를 조정

개인 종단가중치 산출 과정

- 예를 들어, 19차 웨이브의 기본가중치를 w_{19} , 로지스틱 회귀분석을 통해 예측된 추정 응답확률을 \hat{p} 라고 하면 무응답 보정을 통해 얻어지는 가중치는

$$w_{19} \times \frac{1}{\hat{p}}$$

- 개인별 변동 상황에 따라 t차년도 종단면 가중치를 조정
(신규 가구원의 경우 개인별 종단면 가중치 0값 존재)
- 지역 및 응답자의 인구학적 특성에 따라 t차년도 조사 기준연도 인구 추계 값을 이용하여 사후조정을 실시

개인 횡단가중치 산출 과정

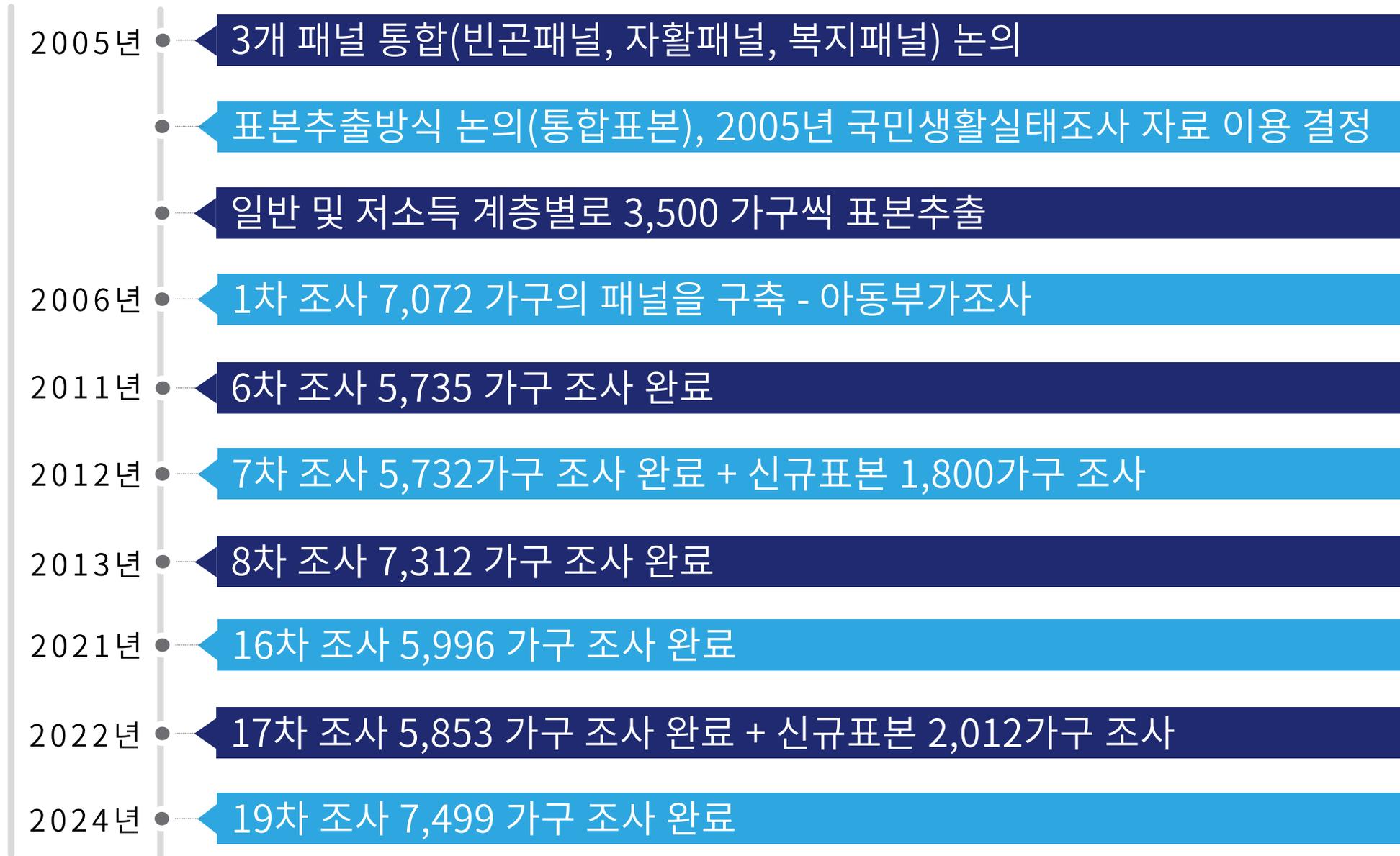
- t차 웨이브의 개인 종단면 기본가중치에 로지스틱 회귀모형으로 추정된 응답확률의 역수를 곱하여 t차 웨이브의 개인 종단면 가중치를 조정
- 변동 상황에 따라 따라 t차년도 종단면 가중치를 조정
- t차년도 종단면 가중치 값이 0인 가중치에 대해 가구별 평균가중치를 적용
- 지역 및 응답자의 인구학적 특성(성별, 연령)에 따라 조사 기준연도 인구 추계 값을 기준으로 사후조정

가구가중치 산출 과정

- t차 웨이브 가구가중치는 t차 웨이브 개인 종단면 가중치의 가구 내 평균값을 계산하여 산출
- 웨이브별로 새로 생성된 신규 가구의 경우, 가구 내 원표본 가구원의 가중치 평균을 적용하고, 원표본 가구원만으로 구성된 가구는 이들의 평균으로 가구가중치를 산출
- 조사 기준연도 가구추계 가구수를 기준으로 지역 및 가구규모별로 사후조정

2. 신규표본 추가에 따른 가중치 조정





가중치 조정과정

- 1차 조사 이후 19차 조사까지 구축된 기존의 패널 표본은 18차 조사의 응답률을 로지스틱 회귀모형으로 응답 확률을 추정하고, 18차 조사의 종단면 가중치에 기초하여 19차 조사에서 조사된 개인 종단면 가중치를 조정
- 2012년 7차 조사에 새롭게 추가된 신규 패널은 ‘2011년 복지욕구 실태조사’의 최종가중치를 기본 가중치로 고려하였으며, 1,800가구에 대한 추출 확률의 역수를 곱하여 설계 가중치를 조정 후 18차 조사까지 응답한 개인에 대해 로지스틱 회귀모형을 이용하여 종단 가중치를 조정
- 2022년 신규로 구축한 패널에 대해 2023년 종단가중치는 2022년 횡단 가중치를 기본가중치로 고려하여 2023년 응답에 대해 로지스틱 회귀모형으로 응답확률을 추정한 후 가중치를 조정

가구 통합 가중치 산출 과정

- 18차 조사의 가구 통합 가중치는 2006년, 2012년 추가 패널, 2022년 추가 패널의 가구 가중치를 결합한 새로운 통합 가중치를 산출해야 하는데 아래와 같은 가중치 계산식을 통해 구함

$$w_{hik}^{CBH} = \begin{cases} \lambda_{1h} w_{ahik}, & hik \in s_a \\ (1 - \lambda_{1h}) w_{bhik}, & hik \in s_b \\ (1 - \lambda_{1h} - \lambda_{2h}) w_{chik}, & hik \in s_c \end{cases}$$

- λ_{1h} 는 2006 패널의 가구 가중치, λ_{2h} 는 2012년 추가 패널의 가구 가중치, λ_{3h} 는 2022년 추가패널의 가중치를 나타냄
- $\lambda_{1h} < \lambda_{2h} < \lambda_{3h}$ 는 가구통합 가중치를 정의하게 되며, 이는 기존 패널과 2012년 및 2022년 추가 표본의 추정량을 선형형태로 결합하게 됨

가구 통합 가중치 산출 과정

- 최적의 혼합계수 λ 는 혼합 추정량의 분산을 최소화 하는 것이라 할 수 있고, 모비율 추정을 기준으로 σ_{ah} σ_{bh} σ_{ch} 이 각각 2006년, 2012년, 2022년 표본에서의 비율추정량의 분산이라면, 혼합계수 λ 는 아래와 같이 정의할 수 있음

$$\lambda_{1h} = \frac{\sigma_{ah}^2}{\sigma_{ah}^2 + \sigma_{bh}^2 + \sigma_{ch}^2} \approx \frac{n_{ah}}{n_{ah} + n_{bh} + n_{ch}}$$

- 최종적으로 가구 통합 가중치는 3종의 가중치를 선형 결합한 형태의 가중치 에
2022년 추계 가구 수에 따라 지역별, 가구원 수별 레이킹 비 조정을 통해 구함

개인 통합 종단 가중치 산출 과정

- 18차 조사의 개인 통합 종단 가중치는 2006년, 2012년 추가 패널, 2022년 추가 패널의 개인 종단 가중치를 결합한 새로운 개인 통합 종단 가중치도 가구 통합 패널 가중치 산정과 동일한 방법으로 다음과 같이 산출됨

$$w_{hik}^{CBPL} = \begin{cases} \lambda_{1h} w_{ahik}, & hik \in s_a \\ (1 - \lambda_{1h}) w_{bhik}, & hik \in s_b \\ (1 - \lambda_{1h} - \lambda_{2h}) w_{chik}, & hik \in s_c \end{cases}$$

- 최종적으로 개인 통합 종단 가중치는 3종의 가중치를 선형 결합한 형태의 가중치에 2022년 추계 인구 자료를 이용하여 지역별, 성별, 연령별의 레이킹 비 조정을 통해 구함

개인 통합 횡단 가중치 산출 과정

- 18차 조사의 개인 통합 횡단 가중치는 2006년, 2012년 추가 패널, 2022년 추가 패널의 개인 횡단 가중치를 결합한 새로운 개인 통합 횡단 가중치도 가구 통합 패널 가중치 산정과 동일한 방법으로 다음과 같이 산출됨

$$w_{hik}^{CBP} = \begin{cases} \lambda_{1h} w_{ahik}, & hik \in s_a \\ (1 - \lambda_{1h}) w_{bhik}, & hik \in s_b \\ (1 - \lambda_{1h} - \lambda_{2h}) w_{chik}, & hik \in s_c \end{cases}$$

- 최종적으로 개인 통합 횡단 가중치는 3종의 가중치를 선형 결합한 형태의 가중치에 2022년 추계 인구 자료를 이용하여 지역별, 성별, 연령별의 레이킹 비 조정을 통해 구함

극단값 조정

- 극단 가중치의 경우, 추정량의 분산을 크게 만들어 추정의 정확도에 영향을 미치므로 조정이 필요함
- 조사 기준년도 인구 추계값을 기준으로 레이킹을 통한 사후 조정 후, 지나치게 큰 가중치(상하위 5%)를 찾아 상하위 5% 경계값으로 절단(trimming)
- 이를 가중치 평균으로 대체한 후 재조정하여 최종 가중치로 산정

3. Q & A



통계 분석을 할 때 가중치를 꼭 써야하나요?

표준가중치와 일반가중치는 각각 어떠한 경우에 쓰나요?

머지데이터의 가중치는 우리나라 전체인구를 대표하는 것인가요?

2025년 한국복지패널(KoWePS) 데이터 설명회

데이터 처리 및 머지 방법

신재동 (한국보건사회연구원)



K I H A S A
한국보건사회연구원



1

복지패널 데이터 소개

2

1~19차 데이터 종류 및 구성

3

자료의 결합 방법

4

횡단면 분석을 위한 결합 방법

5

종단면 분석을 위한 결합 방법

6

데이터 활용 예시(SAS)



1. 복지패널 데이터 소개

■ 한국복지패널(KoWePS) 데이터 공개

- 현재 www.koweps.re.kr:442에서 다운로드 가능

- 19차 종단에러체크 완료

이용 전
Beta 버전을
확인해주세요!



2. 1~19차 데이터 종류 및 구성

- 2025년 공개된 한국복지패널 데이터의 종류 및 구성



2016년부터 결합데이터(1~19차) 제공 중!



19차 가구용 데이터 구성

| 가구용 데이터 | |
|---------|---|
| 구분 | 내용 |
| 원가구 | 전년도 조사당시 조사 완료된 패널가구 (예: 18차년도 조사완료가구 → 19차년도 원가구) |
| 신규가구 | <p>조사기준연도(예: 19차년도 조사기준연도 2023년) 9월 30일 이전에 분가, 결혼, 직장, 학업, 위탁, 이혼, 별거 등의 이유로 원가구(가구주 기준)로부터 분리된 가구</p> <p>※ 조사기준연도 10월 1일 이후에 분리된 가구원은 원가구에 속하는 가구원(9개월 이상 생계를 함께 한 가구원)이므로, 신규가구가 아니라 원가구의 가구원으로 조사</p> |

- ✓ 유형1(원가구, 기존분가)와 유형2(분가 가구) 구별 방법 : 가구생성차수(h19_ind)
 - 유형1 : 원가구(1), 기존분가(2~18)
 - 유형2 : 조사기준년도 분가 가구(19)



19차 가구원용 데이터 구성

가구원용 데이터

| 구분 | 내용 |
|-------|--|
| 원가구원 | 전년도 가구원 조사 응답 가구원 (예: 18차년도 가구원조사 응답가구원 → 19차년도 원가구원) |
| 신규가구원 | 전년도 가구원 조사에 응답하지 않은 가구원이면서, 다음에 해당하는 자 1) 조사기준연도에 만 15세가 된 가구원이면서 중고등학생이 아닌 자 2) 조사기준연도 2월에 고등학교를 졸업(중퇴 포함)한 가구원 3) 조사기준연도에 원가구 혹은 신규가구에 새로 진입한 가구원 |

- ✓ 유형3과 유형4 구별 방법 : 설문유형 구분변수(p19_tq)
 - 유형3은 '3', 유형4는 '4'
- ✓ 원가구원과 신규가구원 구별 방법 : 가구원 진입차수(p19_pind)
 - 유형3 : 원가구원(1), 기존 신규가구원(2~18)
 - 유형4 : 원가구원(1), 기존 신규가구원(2~18), 당해년도 신규가구원(19)



2. 1~19차 데이터 종류 및 구성(계속)

✓ 가구 식별 변수(h##_merkey)

| 변수명 | 내용 |
|------------------------|---|
| h##_id | <ul style="list-style-type: none"> - 가구패널ID로 1차년도 조사대상 가구에 부여된 패널 ID(5자리) - 원가구(1~7072), 7차 추가패널가구(8001~9800), 17차 추가패널가구(10000~12011) - 신규가구의 경우도 원가구와 동일한 가구패널ID를 가짐 |
| h##_ind | <ul style="list-style-type: none"> - 가구생성차수(2자리) - 원가구로부터 분리된 패널 차수를 의미 - 원가구는 1~19차 까지 '01', 신규가구의 경우 '2이상'의 숫자를 가짐 |
| h##_sn | <ul style="list-style-type: none"> - 가구분리 일련번호(2자리) - 원가구로부터 분리된 신규가구의 일련번호를 의미 - 원가구는 1~19차까지 '01'로 유지 - Ex) 2차 조사에서 한 가구로 부터 2개의 신규가구가 생성된 경우 첫 번째 신규가구는 '01', 두 번째 신규가구는 '02'로 부여됨 |
| h##_merkey(KEY) | <ul style="list-style-type: none"> - 패널간 가구 머지 키변수(9자리) - 가구패널ID(5자리) + 가구생성차수(2자리) + 가구분리 일련번호(2자리) - 가구머지 키변수(h01_merkey, h02_merkey, ..., h19_merkey) |

✓ 가구원 식별 변수(h##_pid)

- 개인패널ID(h01_pid, h02_pid, ..., h19_pid)

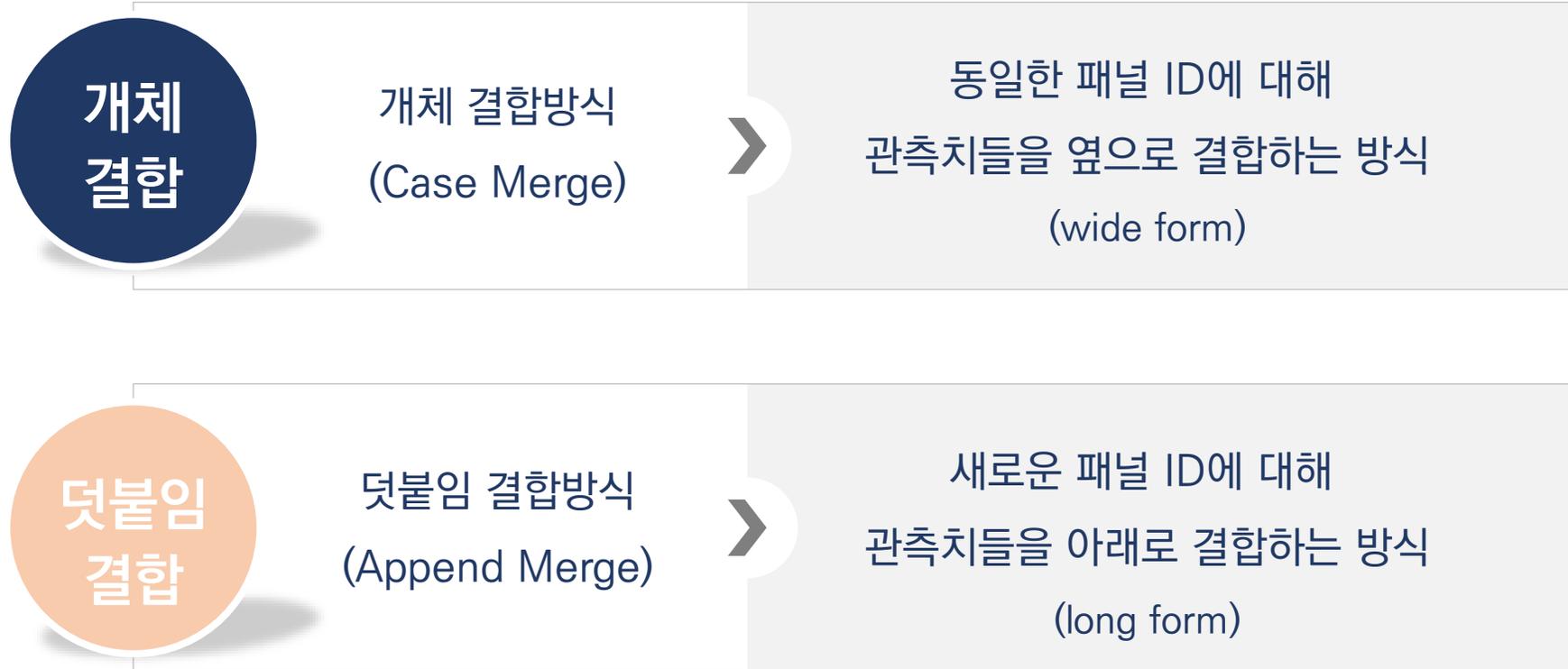


■ 변수명 구조

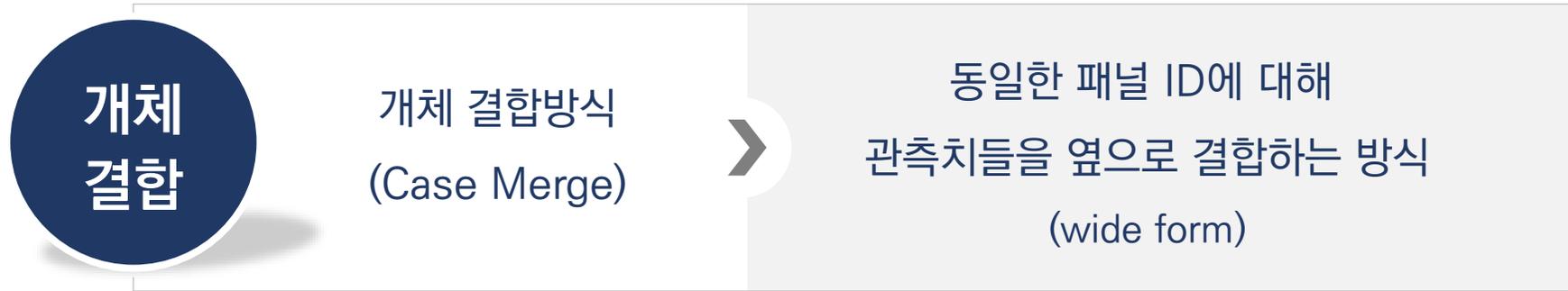


ex) h0102_11, h0202_11, ..., h1902_11
 p0101_1, p0201_1, ..., p1901_1
 c0101_1, c0401_1, ..., c1901_1
 wc02_1, wc05_1, ..., wc17_1
 da03_1, da06_1, ..., da18_1
 h0202_aq1, h0302_3aq1, ..., h1908_19aq1

- 1,2, ..., 19차의 각 조사영역에 속한 문항의 일련번호는 동일한 문항에 대해서 동일한 번호를 부여함.
- 2차 조사에서 추가하였거나 많이 변형된 문항은 변수명에 'aq'를 붙여 구분하였음.
- 3차: 3aq, 4차: 4aq, ..., 19차: 19aq를 붙였음.



- 개체 결합 방식의 단계



- 덧붙임 결합 방식의 단계



- **가구용(개인단위로 변환) + 개인용(가구원용, 부가조사)**

- 키변수(key variable) : 개인패널ID(h01_pid, h02_pid, ..., h19_pid)

※ Wide type 가구용 데이터의 개인단위 Long type 데이터로의 변환 절차

절차1 : 가구용 데이터를 가구원 번호(1~9)로 나누어 9개의 분리데이터를 생성

절차2 : 가구원 번호에 해당하는 변수에 대해 식별(머지) 변수를 부여하고, '덧붙임' 결합방식으로 결합

- **가구원용 + 부가조사**

- 키변수(key variable) : 개인패널ID(h01_pid, h02_pid, ..., h19_pid)



4. 횡단면 분석을 위한 결합 방법(계속)

✓ 가구용 데이터 → 가구원용으로 Reshaping

| 표지 | 가구원별사항 | | | | | | | | | | | | | | | | | | | |
|--------|---------|-----------|----------|---------|----------|---------|---------|-----------|----------|----------|----------|----------|----------|-----|-----------|----------|----------|----------|-----------|-----------|
| | 1번 가구원 | | | | | | | 2번 가구원 | | | | | | | ... | 9번 가구원 | | | | |
| 하위 관측치 | 가구패널 ID | 가구원 진입차수 | 개인패널 ID | 가구원 번호 | 가구주와의 관계 | 성별 | 태어난 연도 | 가구원 진입차수 | 개인패널 ID | 가구원 번호 | 가구주와의 관계 | 성별 | 태어난 연도 | ... | 가구원 진입차수 | 개인패널 ID | 가구원 번호 | 가구주와의 관계 | 성별 | 태어난 연도 |
| 변수명 | h19_id | h19_pind1 | h19_pid1 | h1901_2 | h1901_3 | h1901_4 | h1901_5 | h19_pind2 | h19_pid2 | h1901_14 | h1901_15 | h1901_16 | h1901_17 | ... | h19_pind9 | h19_pid9 | h1901_98 | h1901_99 | h1901_100 | h1901_101 |
| 관측 | 100 | 1 | 10001 | 1 | 010 | 1 | 1950 | 1 | 10002 | 2 | 020 | 2 | 1952 | ... | 2 | 10009 | 9 | 015 | 1 | 1980 |
| | 200 | 1 | 20001 | 1 | 010 | 1 | 1952 | 1 | 20002 | 2 | 020 | 2 | 1954 | ... | | | | | | |
| | 300 | 1 | 30001 | 1 | 010 | 1 | 1948 | 1 | 30002 | 2 | 020 | 2 | 1950 | ... | 2 | 30009 | 9 | 016 | 2 | 1985 |



| | | | | | | | | | |
|-----------------|--------|--------|----------|----------|--------|----------|--------|--------|-----|
| 1번 가구원 데이터파일 | 하위 관측치 | 가구패널ID | 가구원 진입차수 | 개인 패널 ID | 가구원번호 | 가구주와의 관계 | 성별 | 태어난 연도 | ... |
| | 식별변수명 | h19_id | h19_pind | h19_pid | h19_g1 | h19_g2 | h19_g3 | h19_g4 | ... |
| | 관측 | 100 | 1 | 10001 | 1 | 010 | 1 | 1950 | ... |
| | | 200 | 1 | 20001 | 1 | 010 | 1 | 1952 | ... |
| | | 300 | 1 | 30001 | 1 | 010 | 1 | 1948 | ... |
| 2번 가구원 데이터파일 | 하위 관측치 | 가구패널ID | 가구원 진입차수 | 개인 패널 ID | 가구원번호 | 가구주와의 관계 | 성별 | 태어난 연도 | ... |
| | 식별변수명 | h19_id | h19_pind | h19_pid | h19_g1 | h19_g2 | h19_g3 | h19_g4 | ... |
| | 관측 | 100 | 1 | 10002 | 2 | 020 | 2 | 1952 | ... |
| | | 200 | 1 | 20002 | 2 | 020 | 2 | 1954 | ... |
| | | 300 | 1 | 30002 | 2 | 020 | 2 | 1950 | ... |
| 9번 가구원 데이터파일 | 하위 관측치 | 가구패널ID | 가구원 진입차수 | 개인 패널 ID | 가구원번호 | 가구주와의 관계 | 성별 | 태어난 연도 | ... |
| | 식별변수명 | h19_id | h19_pind | h19_pid | h19_g1 | h19_g2 | h19_g3 | h19_g4 | ... |
| | 관측 | 100 | 2 | 10009 | 9 | 015 | 1 | 1980 | ... |
| | | 300 | 2 | 30009 | 9 | 016 | 2 | 1985 | ... |



4. 횡단면 분석을 위한 결합 방법(계속)

✓ 가구용 데이터 → 가구원용으로 Reshaping(계속)

/* 1번 가구원 데이터셋 */

data a1;

set h19;

if h1901_2=1;

rename h19_pind1=h19_pind

h19_pid1=h19_pid

h1901_2=h19_g1

h1901_3=h19_g2

h1901_4=h19_g3

h1901_5=h19_g4

h1901_6=h19_g6

h1901_7=h19_g7

h1901_8=h19_g8

h1901_9=h19_g9

h1901_11=h19_g10

h1901_12=h19_g11

h1901_13=h19_g12

1. 가구일반사항

h1902_1=h19_med1

h1902_2=h19_med2

h1902_3=h19_med3

h1902_4=h19_med4

h1902_5=h19_med5

h1902_6=h19_med6

h1902_7=h19_med7

h1902_8=h19_med8

h1902_aq1=h19_g9_1

h1902_9=h19_med9

h1902_3aq1=h19_med10

2. 건강 및 의료

➤ 가구원 번호가 1인 변수 추출

➤ 설문지에서 매트릭스 형태 변수를 머지용 데이터 변수명으로 변환

매트릭스 형태란?

- 모든 가구원을 대상으로 조사한 문항

1. 가구일반사항
2. 건강 및 의료
3. 경제활동
4. 사회보험
5. 소득

➤ 위 작업을 가구원 번호 1~9번까지 실행



4. 횡단면 분석을 위한 결합 방법(계속)

✓ 가구용 데이터 → 가구원용으로 Reshaping(계속)

```
data h19_gagu;  
set a1 a2 a3 a4 a5 a6 a7 a8 a9;
```

} ➤ 9개의 가구원별 데이터셋을 덧붙임 결합함

```
drop h19_pind1 h19_pid1 h1901_2 h1901_3 h1901_4  
h19_pind2 h19_pid2 h1901_14 h1901_15 h1901_16  
h19_pind3 h19_pid3 h1901_26 h1901_27 h1901_28  
h19_pind4 h19_pid4 h1901_38 h1901_39 h1901_40  
h19_pind5 h19_pid5 h1901_50 h1901_51 h1901_52  
h19_pind6 h19_pid6 h1901_62 h1901_63 h1901_64  
h19_pind7 h19_pid7 h1901_74 h1901_75 h1901_76  
h19_pind8 h19_pid8 h1901_86 h1901_87 h1901_88  
h19_pind9 h19_pid9 h1901_98 h1901_99 h1901_100
```

```
h1902_1 h1902_2 h1902_3 h1902_4 h1902_5  
h1902_10 h1902_11 h1902_12 h1902_13 h1902_14  
h1902_19 h1902_20 h1902_21 h1902_22 h1902_23  
h1902_28 h1902_29 h1902_30 h1902_31 h1902_32  
h1902_37 h1902_38 h1902_39 h1902_40 h1902_41  
h1902_46 h1902_47 h1902_48 h1902_49 h1902_50  
h1902_55 h1902_56 h1902_57 h1902_58 h1902_59  
h1902_64 h1902_65 h1902_66 h1902_67 h1902_68  
h1902_73 h1902_74 h1902_75 h1902_76 h1902_77
```

➤ 기존의 매트릭스 형태의 변수들을 삭제



가구용 데이터를 개인단위로 바꾼 데이터셋으로 변환



■ 개인단위로 변환된 가구 데이터와 가구원용 데이터 결합

```
/* Reshaping 된 가구용 + 가구원용 개체결합 */  
proc sort data=h19_gagu;  
  by h19_pid;  
run;  
  
proc sort data=p19;  
  by h19_pid;  
run;  
  
data koweps_hp19;  
  merge h19_gagu p19;  
  by h19_pid;  
run;
```

- 개인단위로 변환된 가구데이터+가구원용데이터를 개체결합하기에 앞서 Key variable인 개인패널ID로 정렬 후 결합
- 개인단위로 변환된 가구데이터+가구원용데이터로 개체 결합

```
/* (가구용 + 가구원용) + 부가조사 */  
proc sort data=koweps_hp19;  
  by h19_pid;  
run;  
proc sort data=c19;  
  by h19_pid;  
run;  
data koweps_hpc19;  
  merge koweps_hp19 c19;  
  by h19_pid;  
run;
```

- 개체결합하기에 앞서 Key variable인 개인패널ID로 정렬 후 결합
- (개인단위로 변환된 가구데이터+가구원용데이터)+부가조사로 개체 결합



- 가구원용+부가조사도 동일한 방법
- 가구용(개인단위)+부가조사도 동일한 방법

■ 종단면 분석을 위한 다양한 결합 형태 예시

✓ 1차 가구용+ 2차 가구용 + ... + 19차 가구용

- 키변수(key variable) : 가구머지키 변수(h01_merkey, h02_merkey, ..., h19_merkey)

※ 균형패널

1차부터 19차까지 한번도 빠짐없이 모두 조사한
가구만 있는 데이터

VS

※ 불균형패널

1차부터 19차까지 가구의 탈락과 재진입이 섞여
있는 데이터

✓ 1차 가구원용+ 2차 가구원용 + ... + 19차 가구원용

- 키변수(key variable) : 개인패널ID(h01_pid, h02_pid, ..., h19_pid)

- ✓ 1차 {가구용(개인단위로 변환)+ 가구원용 + 아동} +
2차 {가구용(개인단위로 변환)+ 가구원용 + 복지인식} +
3차 {가구용(개인단위로 변환)+ 가구원용 + 장애인} +
...

19차 {가구용(개인단위로 변환)+ 가구원용 + 아동}

- 키변수(key variable) : 개인패널ID(h01_pid, h02_pid, ..., h19_pid)



✓ 1~19차 가구용 데이터의 종단면 데이터로의 변환(균형패널)

➤ 1차 가구용 + 2차 가구용 + ... + 19차 가구용

- 가구일반사항(경상소득, 가처분소득, 지역변수(5개 권역, 7개 권역), 가구구분, 가구원수, 가구형태, 가구의 경제 활동 참여상태, 표준 가중치, 일반가중치)을 종단면 데이터로 구축

/* 1차 가구용 구성 */

```
data panel_h01;
  set w1.koweps_h01_2006_beta21;
  keep h_merkey h01_merkey h01_reg5 h01_reg7
       h01_din h01_cin h01_hc h0101_1
       h0101_110 h0103_4 h01_ws h01_wg;
  h_merkey=h01_merkey;
run;
```

➤ 1차, 2차, ..., 19차 가구용 데이터에서 분석할 변수만 남기고, Key variable인 h_merkey를 생성

```
proc sort data=panel_h01; by h_merkey; run;
proc sort data=panel_h02; by h_merkey; run;
proc sort data=panel_h03; by h_merkey; run;
proc sort data=panel_h04; by h_merkey; run;
proc sort data=panel_h05; by h_merkey; run;
proc sort data=panel_h06; by h_merkey; run;
proc sort data=panel_h07; by h_merkey; run;
proc sort data=panel_h08; by h_merkey; run;
proc sort data=panel_h09; by h_merkey; run;
proc sort data=panel_h10; by h_merkey; run;
proc sort data=panel_h11; by h_merkey; run;
proc sort data=panel_h12; by h_merkey; run;
proc sort data=panel_h13; by h_merkey; run;
proc sort data=panel_h14; by h_merkey; run;
proc sort data=panel_h15; by h_merkey; run;
proc sort data=panel_h16; by h_merkey; run;
proc sort data=panel_h17; by h_merkey; run;
proc sort data=panel_h18; by h_merkey; run;
proc sort data=panel_h19; by h_merkey; run;
```

➤ Key variable인 h_merkey로 정렬 후 결합

/* 19차 기준으로 1~19 모두 조사된 가구 */

```
data k_panel_b_h0;
  merge panel_h01 panel_h02 panel_h03 panel_h04 panel_h05 panel_h06 panel_h07
        panel_h08 panel_h09 panel_h10 panel_h11 panel_h12 panel_h13 panel_h14
        panel_h15 panel_h16 panel_h17 panel_h18 panel_h19;
  by h_merkey;
  if h01_merkey=h02_merkey=h03_merkey=h04_merkey=h05_merkey=h06_merkey=h07_merkey=
  h08_merkey=h09_merkey=h10_merkey=h11_merkey=h12_merkey=h13_merkey=h14_merkey=
  h15_merkey=h16_merkey=h17_merkey=h18_merkey=h19_merkey;
run;
```

➤ 1차, 2차, ..., 19차를 개체 결합함, 1~19차 모두 조사된 가구만 남김



✓ 1~19차 가구용 데이터의 종단면 데이터로의 변환(균형패널)(계속)

```
data k_panel_b_h;
  set k_panel_b_h0;
  hid=_n_;
```

➤ Hid(종단패널 ID)를 생성(순서대로 재 생성)

```
time=1;
reg5=h01_reg5;
reg7=h01_reg7;
cin=h01_cin;
din=h01_din;
h_hc=h01_hc;
f_num=h0101_1;
f_form=h0101_110;
work=h0103_4;
h_ws=h01_ws;
h_wg=h01_wg;
output;
```

➤ Time(시차)변수 생성(1차=1, 2차=2, ..., 19차=19)
 ➤ 새로운 공통변수 생성(각 차수에 있는 기존 변수를 이용)
 Ex) 경상소득(cin)은 1차~19차(h01_cin~h19_cin)

```
time=2;
reg5=h02_reg5;
reg7=h02_reg7;
cin=h02_cin;
din=h02_din;
h_hc=h02_hc;
f_num=h0201_1;
f_form=h0201_110;
work=h0203_4;
h_ws=h02_ws;
h_wg=h02_wg;
output;
```

➤ 개체 결합된 1차~19차의 가구데이터를 다시 Long form으로 재구성
 1~19차 모두 조사된 가구 : 3,295가구
 종단데이터는 62,605(=3,295*19)

✓ 1~19차 가구용 데이터의 종단면 데이터로의 변환(불균형패널)

```

/* 불균형 패널 */
/* 1~19차 동안 한번 이상 조사된 가구 */
data k_panelLub_h0;
  merge panel_h01 panel_h02 panel_h03 panel_h04 panel_h05 panel_h06 panel_h07
        panel_h08 panel_h09 panel_h10 panel_h11 panel_h12 panel_h13 panel_h14
        panel_h15 panel_h16 panel_h17 panel_h18 panel_h19;
  by h_merkey;
run;
    
```

➤ 균형패널과 다르게 조건 없이 결합

```

data k_panelLub_h1
  set k_panelLub_h
  hid=_n_;

  time=1;
  reg5=h01_reg5;
  reg7=h01_reg7;
  cin=h01_cin;
  din=h01_din;
  h_hc=h01_hc;
  f_num=h0101_1;
  f_form=h0101_110;
  work=h0103_4;
  h_ws=h01_ws;
  h_wg=h01_wg;
  output;

  time=2;
  reg5=h02_reg5;
  reg7=h02_reg7;
  cin=h02_cin;
  din=h02_din;
  h_hc=h02_hc;
  f_num=h0201_1;
  f_form=h0201_110;
  work=h0203_4;
  h_ws=h02_ws;
  h_wg=h02_wg;
  output;
    
```

➤ 균형패널과 동일하게 종단데이터 구축

```

/* 해당차수에 조사가 안된 가구 삭제 */
data k_panelLub_h2;
  set k_panelLub_h1;
  if f_num=. then delete;
run;
    
```

➤ 마지막으로 종단데이터 구축 후 각 차수별 무응답 가구를 삭제해야 함.

(238,621 (12,559*19=238,621), 무응답 가구 삭제 : 127,900)



✓ 1~19차 가구용 데이터의 종단면 데이터로의 변환(불균형패널)(계속)

```
/* 가구에서 1~19차 동안 조사된 횟수 */  
proc sql;  
  create table hid_count1_1 as  
  select h_merkey, count(h_merkey) as count_h  
  from k_panel_lub_h2  
  group by h_merkey;  
quit;
```

```
proc sql;  
  create table hid_count1_2 as  
  select *, b.count_h from k_panel_lub_h2 as a  
  left join hid_count1_1 as b  
  on a.h_merkey=b.h_merkey;  
quit;
```

```
/* 1차~19차 균형패널 추출 */  
proc sql;  
  create table hid_count as  
  select * from hid_count1_2 where count_h=19;  
quit;
```

➤ 1~19차 동안 조사된 횟수 변수 생성(count_h)

➤ h_merkey를 기준으로 좌측결합(left join)함

➤ 불균형 패널에서 조사 횟수 변수를 활용하여 균형패널을 만들 수
있음 (count_h=19)



- ✓ 15세 이상(미만) 가구원 수 구하기(가구데이터기준)

```

/* 나이변수 생성(age1~age9) */
data h19_age;
  set h19_r;
  age1=2023-h1901_5;
  age2=2023-h1901_17;
  age3=2023-h1901_29;
  age4=2023-h1901_41;
  age5=2023-h1901_53;
  age6=2023-h1901_65;
  age7=2023-h1901_77;
  age8=2023-h1901_89;
  age9=2023-h1901_101;
  keep h19_id--h1901_110 age1-age9;
run;

```



MEANS 프로시저

| 변수 | 최솟값 | 최댓값 | 중위수 |
|------|-------|--------|-------|
| age1 | 19.00 | 97.00 | 68.00 |
| age2 | 1.00 | 117.00 | 58.00 |
| age3 | 0.00 | 101.00 | 21.00 |
| age4 | 0.00 | 92.00 | 14.00 |
| age5 | 1.00 | 95.00 | 12.00 |
| age6 | 1.00 | 77.00 | 9.00 |
| age7 | 1.00 | 67.00 | 7.00 |
| age8 | 5.00 | 5.00 | 5.00 |
| age9 | . | . | . |

```

proc means data=h19_age min max median maxdec=2 var age1-age9; run;

```



✓ 15세 이상(미만) 가구원 수 구하기(가구데이터기준)(계속)

```

data age15_up_down;
  set h19_age;
  /* 15세 이상 */
  if 15 <= age1 <= 116 then age15_up1=1; else age15_up1=0;
  if 15 <= age2 <= 116 then age15_up2=1; else age15_up2=0;
  if 15 <= age3 <= 116 then age15_up3=1; else age15_up3=0;
  if 15 <= age4 <= 116 then age15_up4=1; else age15_up4=0;
  if 15 <= age5 <= 116 then age15_up5=1; else age15_up5=0;
  if 15 <= age6 <= 116 then age15_up6=1; else age15_up6=0;
  if 15 <= age7 <= 116 then age15_up7=1; else age15_up7=0;
  if 15 <= age8 <= 116 then age15_up8=1; else age15_up8=0;
  if 15 <= age9 <= 116 then age15_up9=1; else age15_up9=0;
  age15_up_sum=sum(of age15_up1-age15_up9);
  /* 15세 미만 */
  if 0 <= age1 < 15 then age15_down1=1; else age15_down1=0;
  if 0 <= age2 < 15 then age15_down2=1; else age15_down2=0;
  if 0 <= age3 < 15 then age15_down3=1; else age15_down3=0;
  if 0 <= age4 < 15 then age15_down4=1; else age15_down4=0;
  if 0 <= age5 < 15 then age15_down5=1; else age15_down5=0;
  if 0 <= age6 < 15 then age15_down6=1; else age15_down6=0;
  if 0 <= age7 < 15 then age15_down7=1; else age15_down7=0;
  if 0 <= age8 < 15 then age15_down8=1; else age15_down8=0;
  if 0 <= age9 < 15 then age15_down9=1; else age15_down9=0;
  age15_down_sum=sum(of age15_down1-age15_down9);
run;
  
```



FREQ 프로시저

| age15_up_sum | 빈도 | 백분율 | 누적 빈도 | 누적 백분율 |
|--------------|------|-------|-------|--------|
| 1 | 2986 | 39.82 | 2986 | 39.82 |
| 2 | 3073 | 40.98 | 6059 | 80.80 |
| 3 | 935 | 12.47 | 6994 | 93.27 |
| 4 | 434 | 5.79 | 7428 | 99.05 |
| 5 | 67 | 0.89 | 7495 | 99.95 |
| 6 | 4 | 0.05 | 7499 | 100.00 |

| age15_down_sum | 빈도 | 백분율 | 누적 빈도 | 누적 백분율 |
|----------------|------|-------|-------|--------|
| 0 | 6623 | 88.32 | 6623 | 88.32 |
| 1 | 437 | 5.83 | 7060 | 94.15 |
| 2 | 370 | 4.93 | 7430 | 99.08 |
| 3 | 64 | 0.85 | 7494 | 99.93 |
| 4 | 5 | 0.07 | 7499 | 100.00 |

```

proc freq data=age15_up_down; tables age15_up_sum age15_down_sum; run;
  
```

6. 데이터 활용 예시(SAS)(계속)

✓ 15세 이상(미만) 가구원 수 구하기(개인단위 결합데이터기준)

```

/* 나이변수 생성(age) */
data hpc19_age;
  set hpc19_r;
  age = 2023-h19_g4;
  if 15 <= age <= 116 then age15_up=1;
  else age15_up=0;

  if 0 <= age < 15 then age15_down=1;
  else age15_down=0;
  keep h19_id--h1901_110 age age15_up age15_down;
run;

/* 15세 이상 */
proc sort data=hpc19_age; by h19_merkey h19_pid; run;

data hpc19_age15_up;
  set hpc19_age;
  by h19_merkey;
  if first.h19_merkey then age15_up_sum=0;
  age15_up_sum+age15_up;
  if last.h19_merkey;
run;

proc freq data=hpc19_age15_up; tables age15_up_sum; run;

/* 15세 미만 */
proc sort data=hpc19_age; by h19_merkey h19_pid; run;

data hpc19_age15_down;
  set hpc19_age;
  by h19_merkey;
  if first.h19_merkey then age15_down_sum=0;
  age15_down_sum+age15_down;
  if last.h19_merkey;
run;

proc freq data=hpc19_age15_down; tables age15_down_sum; run;

```



FREQ 프로시저

| age15_up_sum | 빈도 | 백분율 | 누적 빈도 | 누적 백분율 |
|--------------|------|-------|-------|--------|
| 1 | 2986 | 39.82 | 2986 | 39.82 |
| 2 | 3073 | 40.98 | 6059 | 80.80 |
| 3 | 935 | 12.47 | 6994 | 93.27 |
| 4 | 434 | 5.79 | 7428 | 99.05 |
| 5 | 67 | 0.89 | 7495 | 99.95 |
| 6 | 4 | 0.05 | 7499 | 100.00 |

| age15_down_sum | 빈도 | 백분율 | 누적 빈도 | 누적 백분율 |
|----------------|------|-------|-------|--------|
| 0 | 6623 | 88.32 | 6623 | 88.32 |
| 1 | 437 | 5.83 | 7060 | 94.15 |
| 2 | 370 | 4.93 | 7430 | 99.08 |
| 3 | 64 | 0.85 | 7494 | 99.93 |
| 4 | 5 | 0.07 | 7499 | 100.00 |

■ 필요한 변수만을 남기고 싶을 때

- Keep, Drop 옵션을 활용하여 필요 변수 선택 가능
- SQL 프로시저를 활용하여 필요 변수 선택 가능(기재된 변수 순서에 따라 데이터가 생성됨)

proc sql;

```
create table hpc19_new as
```

```
select h19_pid, h19_pind, h19_merkey, h_new, h19_cobf, p19_wgl, p19_wsl, p19_wgc, p19_wsc, h19_reg5, h19_reg7, h19_cin,  
       h19_din, h19_hc, nh1901_1, nh1901_2, h1901_1, h1901_110, h19_g1, h19_g2, h19_g3, h19_g4, h19_g6, h19_g7, h19_g8,  
       h19_g9, h19_g10, h19_g11, h19_g12
```

```
from w19.koweps_hpc19_2024_beta1;
```

quit;

■ Proc export

```
/* SPSS 변환 */
```

```
PROC EXPORT DATA= w19.koweps_h19_2024_beta1 /* 변환 하고자 하는 data명 */
```

```
OUTFILE= "D:\₩KOWEPS₩2025₩2025복지패널데이터설명회₩sample_spss"
```

```
/* "저장하고자 하는 경로 및 데이터명 지정" */
```

```
DBMS=SPSS REPLACE;
```

```
/* 변환하고자 하는 패키지 */
```

RUN;

```
/* STATA 변환(v13) */
```

```
PROC EXPORT DATA= w19.koweps_h19_2024_beta1
```

```
OUTFILE= "D:\₩KOWEPS₩2025₩2025복지패널데이터설명회₩sample_stata"
```

```
DBMS=STATA REPLACE;
```

RUN;



감사합니다

