

한국복지패널(KoWePS) 데이터 처리 및 머지 방법

신재동 (한국보건사회연구원)



데이터 머지 방법 소개에 앞서

• 한국복지패널(KoWePS) 데이터 공개

- 현재 www.koweps.re.kr에서 다운로드 가능 (3월 29일 업로드 완료)

- 13차 종단에러체크 완료



이용 전
Beta 버전을
확인해주세요!



개체 결합

개체 결합방식
(Case Merge)



동일한 패널 ID에 대해
관측치들을 옆으로 결합하는 방식

덧붙임 결합

덧붙임 결합방식
(Append Merge)



새로운 패널 ID에 대해
관측치들을 아래로 결합하는 방식

개체 결합

개체 결합방식
(Case Merge)



동일한 패널 ID에 대해
관측치들을 옆으로 결합하는 방식

1단계

키변수
지정

2단계

키변수를
동일명으로
변경

3단계

키변수로
정렬

4단계

키변수를
기준으로
머지

5단계

분석

덧붙임 결합방식의 단계

덧붙임
결합

덧붙임 결합방식
(Append Merge)



새로운 패널 ID에 대해
관측치들을 아래로 결합하는 방식

1단계

차수 변수
연도 변수
생성

2단계

변수명에서
차수 삭제

3단계

키변수를
기준으로
머지

4단계

키변수와
연도 변수로
정렬

5단계

분석

(가구용(h), 가구원용(p), 부가조사(c, wc, da))

+ (패널차수) + (조사영역별 일련번호) _

+ (각 조사영역에 속한 문항의 일련번호)

→ h0102_15, h0202_15, h0302_15, h0202_aq1, h0302_3aq1
h0402_4aq1, h0503_5aq1, h0708_7aq1, h0908_9aq1,
h1003_10aq1, h1103_11aq1, h1203_12aq1, h1308_13aq1

※ 1,2,3,4,5,6,7,8,9,10,11,12,13차의 각 조사영역에 속한 문항의 일련번호는 동일한 문항에 대해서 동일한 번호를 부여하였음.

※ 2차조사에서 추가하였거나 많이 변형된 문항은 aq를 붙였음.

※ 3차: 3aq, 4차: 4aq, 5차: 5aq, 6차: 6aq, 7차: 7aq, 8차: 8aq,
9차:9aq, 10차: 10aq, 11차: 11aq, 12차: 12aq, 13차: 13aq를 붙였음

1~13차 데이터 종류 및 구성

- 2019년 공개된 한국복지패널 데이터의 종류 및 구성



1, 4, 7, 10, 13차
데이터 종류



2, 5, 8, 11차
데이터 종류



3, 6, 9, 12차
데이터 종류

가구용

가구용

가구용

가구원용

가구원용

가구원용

아동 부가조사

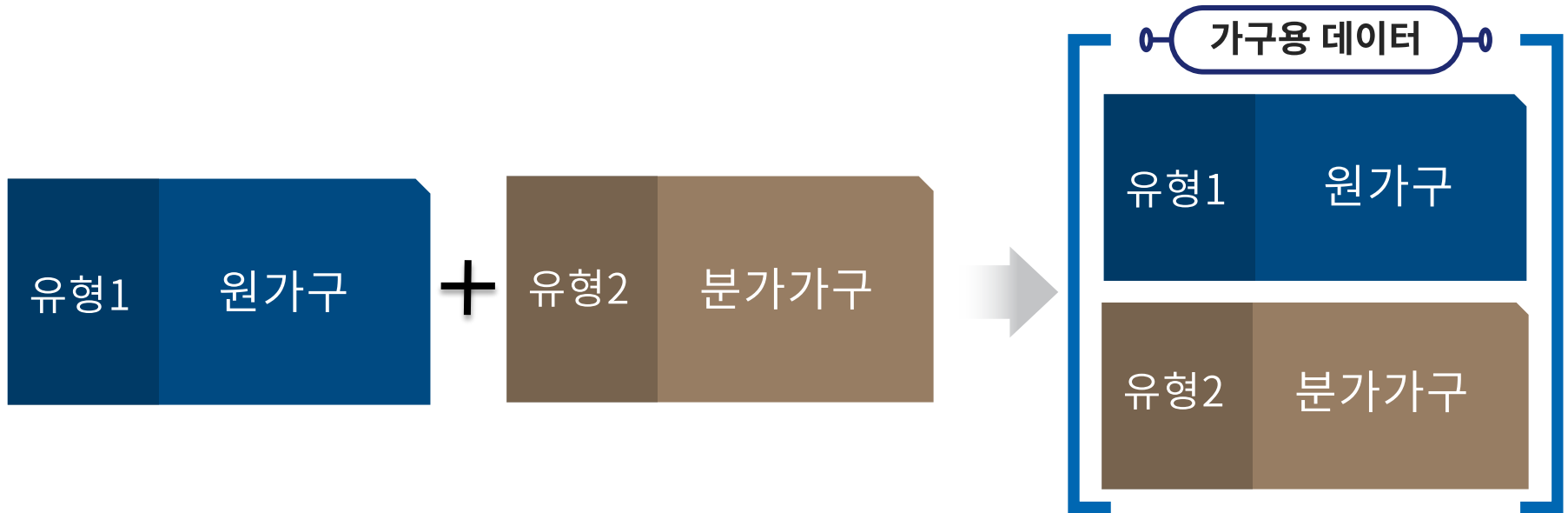
복지인식 부가조사

장애인 부가조사

2016년부터 1~13차 결합데이터(가구용, 머지용) 제공 중!

1~13차 데이터 종류 및 구성

- 13차 가구용 데이터 구성

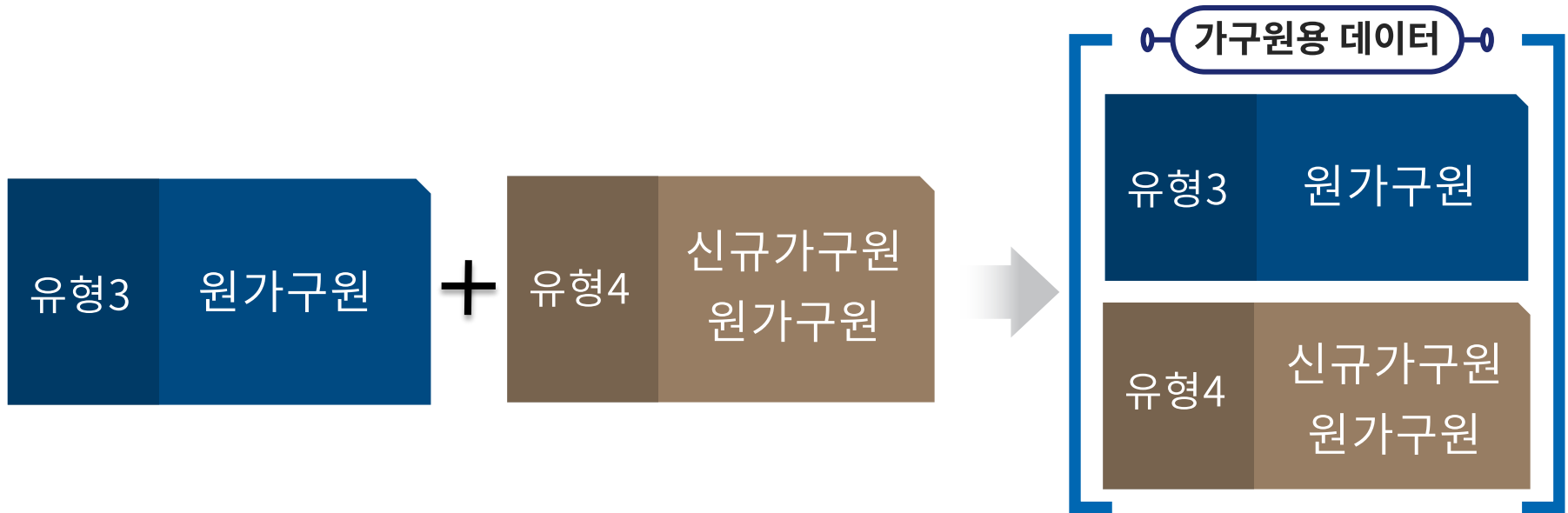


- 유형1(원가구)와 유형2(분가가구) 구별 방법: 가구생성차수(h13_ind)

- 유형1 : 원가구(1), 기존분가(2~12)
- 유형2 : 당해년도 분가가구(13)

1~13차 데이터 종류 및 구성

- 13차 가구원용 데이터 구성



- 유형3과 유형4 구별 방법: 설문유형구분변수(p13_tp)

- 유형3은 '3', 유형4는 '4'

- 원가구원과 신규가구원 구별방법: 가구원진입차수(p13_pind)

- 유형3 : 원가구원(1), 기존 신규가구원(2~12)

- 유형4 : 원가구원(1), 기존 신규가구원(2~12), 당해년도 신규가구원(13)

가구 및 개인 식별 변수

- 가구 식별변수

- 가구패널 ID, 가구생성차수, 가구분리일련번호

- ⇒ 가구머지키변수(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)

- 개인 식별변수

- 개인패널 ID

- ⇒ h01_pid, h02_pid, h03_pid, h04_pid, h05_pid, h06_pid, h07_pid, h08_pid, h09_pid, h10_pid, h11_pid, h12_pid, h13_pid

횡단면 분석시 다양한 형태의 머지 방법

- 가구용(개인단위변수) + 개인용(가구원용, 부가조사)

- Key Variable: 개인패널 ID(h01_pid, h02_pid, h03_pid, h04_pid, h05_pid, h06_pid, h07_pid, h08_pid, h09_pid, h10_pid, h11_pid, h12_pid, h13_pid)

※ 가구용 개인단위 변수들(Wide Type)의 Long Type 변환 절차

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

- 절차2: 가구원 번호에 해당하는 변수에 대해 식별(머지)변수명을 부여하고 덧붙임 결합방식으로 결합

- 가구원용 + 부가조사

- Key Variable: 개인패널 ID(h01_pid, h02_pid, h03_pid, h04_pid, h05_pid, h06_pid, h07_pid, h08_pid, h09_pid, h10_pid, h11_pid, h12_pid, h13_pid)

가구용 데이터 → 가구원용으로 Reshaping

표지	가구일반사항																		
	1번 가구원							2번 가구원							3번 가구원				
하위 관측치	가구패널 ID	가구원 진입차수	개인패널 ID	가구원 번호	가구주와의 관계	성 별	태어난 연도	가구원 진입차수	개인패널 ID	가구원 번호	가구주와의 관계	성 별	태어난 연도	가구원 진입차수	개인패널 ID	가구원 번호	가구주와의 관계	성 별	태어난 연도
변수명	h13_id	h13_pind1	h13_pid1	h1301_2	h1301_3	h1301_4	h1301_5	h13_pind2	h13_pid2	h1301_14	h1301_15	h1301_16	h1301_17	h13_pind3	h13_pid3	h1301_26	h1301_27	h1301_28	h1301_29
관측	1000	1	10001	1	010	1	1950	1	10002	2	020	2	1952	2	10003	3	011	1	1975
	2000	1	20001	1	010	1	1952	1	20002	2	020	2	1954						
	3000	1	30001	1	010	1	1948	1	30002	2	020	2	1950	2	30003	3	012	2	1973

↓

1번 가구원 데이터파일	하위 관측치 식별변수명	가구패널ID	가구원 진입차수	개인 패널 ID	가구원번호	가구주와의 관계	성별	태어난 연도	...
		h13_id	h13_pind	h13_pid	h13_g1	h13_g2	h13_g3	h13_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 식별변수명	가구패널ID	가구원 진입차수	개인 패널 ID	가구원번호	가구주와의 관계	성별	태어난 연도	...
		h13_id	h13_pind	h13_pid	h13_g1	h13_g2	h13_g3	h13_g4	...
	관측	100	1	10002	2	020	2	1952	...
		200	1	20002	2	020	2	1954	...
		300	1	30001	2	020	2	1950	...
+									
9번 가구원 데이터파일	하위 관측치 식별변수명	가구패널ID	가구원 진입차수	개인 패널 ID	가구원번호	가구주와의 관계	성별	태어난 연도	...
		h13_id	h13_pind	h13_pid	h13_g1	h13_g2	h13_g3	h13_g4	...
	관측	100	1	10009	9	015	1	1980	...
		300	1	20009	9	016	2	1985	...

가구용 데이터 → 가구원용으로 Reshaping

```

/* 1번 가구원 데이터셋 */
data a1;
  set h13_m_t;
  if h1301_2=1;
  rename h13_pind1=h13_pind
h13_pid1=h13_pid
h1301_2=h13_g1
h1301_3=h13_g2
h1301_4=h13_g3
h1301_5=h13_g4
h1301_6=h13_g6
h1301_7=h13_g7
h1301_8=h13_g8
h1301_9=h13_g9
h1301_11=h13_g10
h1301_12=h13_g11
h1301_13=h13_g12

h1302_1=h13_med1
h1302_2=h13_med2
h1302_3=h13_med3
h1302_4=h13_med4
h1302_5=h13_med5
h1302_6=h13_med6
h1302_7=h13_med7
h1302_8=h13_med8
h1302_aq1=h13_g9_1
h1302_9=h13_med9
h1302_3aq1=h13_med10

```

- 가구원 번호가 1인 것만 추출
- 설문지에서 매트릭스 형태를 머지용 데이터 변수명으로 바꿈

매트릭스 형태란?

- 모든 가구원을 대상으로 조사한 문항

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

- 위 작업을 가구원번호 1~9번까지 실행

가구용 데이터 → 가구원용으로 Reshaping

```
data h13_gagu;  
  set a1 a2 a3 a4 a5 a6 a7 a8 a9;  
  drop h13_pind1 h13_pid1 h1301_2  
        h13_pind2 h13_pid2 h1301_14  
        h13_pind3 h13_pid3 h1301_26  
        h13_pind4 h13_pid4 h1301_36  
        h13_pind5 h13_pid5 h1301_50  
        h13_pind6 h13_pid6 h1301_62  
        h13_pind7 h13_pid7 h1301_74  
        h13_pind8 h13_pid8 h1301_86  
        h13_pind9 h13_pid9 h1301_96  
  
        h1302_1 h1302_2 h1302_3  
        h1302_10 h1302_11 h1302_12  
        h1302_19 h1302_20 h1302_21  
        h1302_28 h1302_29 h1302_30  
        h1302_37 h1302_38 h1302_39  
        h1302_46 h1302_47 h1302_48  
        h1302_55 h1302_56 h1302_57  
        h1302_64 h1302_65 h1302_66  
        h1302_73 h1302_74 h1302_75  
  
        h1303_1 h1303_2 h1303_3  
        h1303_11 h1303_12 h1303_13  
        h1303_21 h1303_22 h1303_23  
        h1303_31 h1303_32 h1303_33  
        h1303_41 h1303_42 h1303_43  
        h1303_51 h1303_52 h1303_53  
        h1303_61 h1303_62 h1303_63  
        h1303_71 h1303_72 h1303_73  
        h1303_81 h1303_82 h1303_83
```

- 9개의 가구원별 데이터셋을 세로 결합
- 기존의 매트릭스 형태의 변수들을 삭제
- 이것으로 가구용 데이터가 개인단위로 바뀐 데이터셋을 구축 가능

가구용(개인단위조사) + 가구원용

```
proc sort data=h13_gagu;  
  by h13_pid;  
run;  
proc sort data=p13;  
  by h13_pid;  
run;  
data koweps_hp13;  
  merge h13_gagu p13;  
  by h13_pid;  
run;
```

```
/* (가구용 + 가구원용) + 부가조사 */  
proc sort data=koweps_hp13;  
  by h13_pid;  
run;  
proc sort data=c13;  
  by h13_pid;  
run;  
data koweps_hpc13;  
  merge koweps_hp13 c13;  
  by h13_pid;  
run;
```

- 가구용 (개인단위조사) + 가구원용을 개체결합 하기에 앞서 Key variable인 개인패널 ID로 정렬 후에 결합
- 가구용(개인단위조사)+가구원용으로 결합한 데이터를 부가조사와 개체결합
- 개체결합하기에 앞서 Key variable인 개인패널 ID로 정렬 후 결합
- 가구원용 + 부가조사도 동일한 방법
- 가구용 + 부가조사도 가구용 + 가구원용과 동일한 방법

종단면 분석시 다양한 형태의 머지 방법

1

1차 가구용 + 2차 가구용+3차 가구용+4차 가구용+5차 가구용+ 6차 가구용
+7차 가구용+8차 가구용+9차 가구용+10차 가구용+ 11차 가구용+12차 가구용
+13차 가구용

- Key Variable: 가구머지키변수

(h01_merkey,h02_merkey,h03_merkey, h04_merkey, h05_merkey,h06_merkey,
h07_merkey, h08_merkey, h09_merkey, h10_mekrey, h11_merkey, h12_merkey, h13_merkey)

2

1차 가구원용 + 2차 가구원용+3차 가구원용+4차 가구원용+5차 가구원용
+ 6차 가구원용+7차 가구원용+8차 가구원용+9차 가구원용+10차 가구원용
+ 11차 가구원용+12차 가구원용+13차 가구원용

- Key Variable: 개인패널 ID

(h01_pid,h02_pid,h03_pid, h04_pid, h05_pid,h06_pid, h07_pid, h08_pid,
h09_pid, h10_pid, h11_pid, h12_pid, h13_pid)

종단면 분석시 다양한 형태의 머지 방법

3

(1차 가구용(개인단위조사)+가구원용+아동) +
 (2차 가구용(개인단위조사)+가구원용+복지인식) +
 (3차 가구용(개인단위조사)+가구원용+장애인) +

...

(11차 가구용(개인단위조사)+가구원용+복지인식) +
 (12차 가구용(개인단위조사)+가구원용+장애인) +
 (13차 가구용(개인단위조사)+가구원용+아동)

- Key Variable: 개인패널 ID

(h01_pid, h02_pid, h03_pid, h04_pid, h05_pid, h06_pid, h07_pid, h08_pid,
 h09_pid, h10_pid, h11_pid, h12_pid, h13_pid)

1~13차 가구용을 종단면 데이터로 구축(균형패널)

- 1차가구용+2차가구용+3차가구용+4차가구용+5차가구용+...+10차가구용
+11차가구용+12차가구용+13차가구용

- 가구일반사항(경상소득, 가처분소득, 지역변수(5개권역, 7개권역), 가구구분, 가구원수, 가구형태, 경찰상태)을 종단면 데이터로 구축

```
data panel_h01;
  set w1.Koweps_h01_2006_beta15;
  keep h_merkey h01_merkey h01_reg5 h01_reg7 h01_dir
        h0101_110 h0103_4 h01_ws h01_wg;
  h_merkey=h01_merkey;
run;
```

```
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;
/* 13차 기준으로 1~13 모두 조사된 가구: 4170 */
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;
  by h_merkey;
  if h01_merkey=h02_merkey=h03_merkey=h04_merkey=h05_merk
     h06_merkey=h07_merkey=h08_merkey=h09_merkey=h10_merk
     h11_merkey=h12_merkey=h13_merkey;
run;
```

- 1차, 2차, 3차, 4차, 5차, 6차, 7차, 8차, 9차, 10차, 11차, 12차, 13차 가구용 데이터셋에서 분석할 변수만 남김
- h_merkey 생성(key variable)
- 1차, 2차, 3차, 4차, 5차, 6차, 7차, 8차, 9차, 10차, 11차, 12차, 13차를 개체 결합방식으로 가로 결합
- 1~13차에서 모두 조사된 가구만 남김

1~13차 가구용을 종단면 데이터로 구축(균형패널)

```
data k_panel_b_h;
  set k_panel_b_h0;
  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;
```

- time(시차)변수 생성(1차=1, 2차=2, 3차=3, 4차=4, 5차=5, 6차=6, 7차=7, 8차=8, 9차=9, 10차=10, 11차=11, 12차=12, 13차=13)
- hid(종단패널 ID) 생성(순서대로 재 생성)
- 새로운 공통변수 생성(각 차수에 있는 기존 변수를 이용)
ex) 경상소득(cin)은 1차~13차(h01_cin~h13_cin)
- 가로결합된 1~13차의 가구데이터를 다시 세로로 구성
- 1~13차 모두 조사된 가구: 4170
- 종단데이터는 54210(4170*13)

1~13차 가구용을 종단면 데이터로 구축(불균형패널)

```
/* 불균형 패널 */
/* 13차 기준으로 1~13 모두 조사된 가구: 9963(1차 7072 + 7차 추가 1800 + 1~13차 분가 1091) */
data k_panel_ub_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;
  by h_merkey;
run;
```

- 균형패널과 달리 조건없이 머지

```
data k_panel_ub_h1;
  set k_panel_ub_h0;
  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
- 단, 마지막에 각 차수별 조사가 안된 가구를 삭제해야 함 (86,526(129,519=9,963*13, 무응답 가구 삭제: 42,993))

```
/* 해당차수에 조사가 안된 가구 삭제 */
data k_panel_ub_h2;
  set k_panel_ub_h1;
  if h_hc=. then delete;
run;
```

1~13차 가구용을 종단면 데이터로 구축(불균형패널)

```
/* 가구에서 1~13차 동안 조사된 횟수 */  
data hid_count0;  
    set k_panel_ub_h2;  
    a=1;  
run;  


---

proc sort data=hid_count0; by h_merkey time; run;  


---

data hid_count;  
    set hid_count0;  
    by h_merkey;  
    hid_count+a;  
    if last.h_merkey then do;  
        output; hid_count=0;  
    end;  
    keep h_merkey hid_count;  
run;  


---

proc sort data=k_panel_ub_h2; by h_merkey; run;  
proc sort data=hid_count; by h_merkey; run;  


---

data k_panel_ub_h;  
    merge k_panel_ub_h2 hid_count;  
    by h_merkey;  
run;
```

- 1~13차 동안 조사된 횟수 변수 생성 (누적합)
- 가구데이터와 머지용데이터는 1~13차 결합용(개체 및 덧붙임) 데이터를 가구용과 머지용으로 제공함
- 불균형패널에서 조사 횟수 변수를 활용하여 균형패널을 만들 수 있음 (hid_count=13)

데이터활용: 15세 이상(미만) 가구원 수 구하기

- 가구용 데이터 사용

```

/* 나이변수 생성(age1~age9) */
data h13_age;
  set h13_r;
  age1=2017-h1301_5;
  age2=2017-h1301_17;
  age3=2017-h1301_29;
  age4=2017-h1301_41;
  age5=2017-h1301_53;
  age6=2017-h1301_65;
  age7=2017-h1301_77;
  age8=2017-h1301_89;
  age9=2017-h1301_101;
  keep h13_id--h1301_110 age1-age9;
run;

```

```

proc means data=h13_age min max; var age1-age9; run;

```

변수	최솟값	최댓값
age1	19.0000000	98.0000000
age2	2.0000000	100.0000000
age3	0	110.0000000
age4	0	96.0000000
age5	0	93.0000000
age6	1.0000000	90.0000000
age7	0	94.0000000
age8	5.0000000	22.0000000
age9	3.0000000	3.0000000

데이터활용: 15세 이상(미만) 가구원 수 구하기

```

/* 방법1 */
/* 15세 이상 가구원 수 */
data h13_age15_up;
  set h13_age;
  array x[9] age1-age9;
  array y[9] age15_u1-age15_u9;
  do i=1 to 9;
    if 15 <= x[i] <= 109 then y[i]=1;
    else y[i]=0;
  end;
  drop i ;
run;

data h13_age15_up_sum;
  set h13_age15_up;
  age15_usum=sum(of age15_u1-age15_u9);
run;

proc freq data=h13_age15_up_sum; tables age15_usum; run;

/* 15세 미만 가구원 수 */
data h13_age15_down;
  set h13_age;
  array x[9] age1-age9;
  array y[9] age15_d1-age15_d9;
  do i=1 to 9;
    if 0 <= x[i] < 15 then y[i]=1;
    else y[i]=0;
  end;
  drop i ;
run;

data h13_age15_down_sum;
  set h13_age15_down;
  age15_dsum=sum(of age15_d1-age15_d9);
run;

proc freq data=h13_age15_down_sum; tables age15_dsum; run;

```

방법1 결과

age15_usum	빈도	백분율	누적 빈도	누적 백분율
1	2121	32,76	2121	32,76
2	2780	42,94	4901	75,7
3	950	14,67	5851	90,38
4	505	7,8	6356	98,18
5	104	1,61	6460	99,78
6	12	0,19	6472	99,97
7	1	0,02	6473	99,98
8	1	0,02	6474	100

age15_dsum	빈도	백분율	누적 빈도	누적 백분율
0	5393	83,3	5393	83,3
1	500	7,72	5893	91,03
2	494	7,63	6387	98,66
3	75	1,16	6462	99,81
4	9	0,14	6471	99,95
5	3	0,05	6474	100

데이터활용: 15세 이상(미만) 가구원 수 구하기

```

/* 방법2 */
data age15_up_down;
  set h13_age;
  /* 15세 이상 */
  if 15 <= age1 <=109 then age15_up1=1; else age15_up1=0;
  if 15 <= age2 <=109 then age15_up2=1; else age15_up2=0;
  if 15 <= age3 <=109 then age15_up3=1; else age15_up3=0;
  if 15 <= age4 <=109 then age15_up4=1; else age15_up4=0;
  if 15 <= age5 <=109 then age15_up5=1; else age15_up5=0;
  if 15 <= age6 <=109 then age15_up6=1; else age15_up6=0;
  if 15 <= age7 <=109 then age15_up7=1; else age15_up7=0;
  if 15 <= age8 <=109 then age15_up8=1; else age15_up8=0;
  if 15 <= age9 <=109 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;

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

```

방법2 결과

age15_usum	빈도	백분율	누적 빈도	누적 백분율
1	2121	32,76	2121	32,76
2	2780	42,94	4901	75,7
3	950	14,67	5851	90,38
4	505	7,8	6356	98,18
5	104	1,61	6460	99,78
6	12	0,19	6472	99,97
7	1	0,02	6473	99,98
8	1	0,02	6474	100

age15_dsum	빈도	백분율	누적 빈도	누적 백분율
0	5393	83,3	5393	83,3
1	500	7,72	5893	91,03
2	494	7,63	6387	98,66
3	75	1,16	6462	99,81
4	9	0,14	6471	99,95
5	3	0,05	6474	100

데이터활용: 15세 이상(미만) 가구원 수 구하기

- 머지용 데이터 사용

```
/* 나이변수 생성(age) */
data hpc13_age;
  set hpc13_r;
  age = 2017-h13_g4;
  if 15 <= age <= 110 then age15_up=1;
  else age15_up=0;

  if 0 <= age < 15 then age15_down=1;
  else age15_down=0;
  keep h13_id--h1301_110 age age15_up age15_down;
run;
```

```
/* 15세 이상 */
proc sort data=hpc13_age; by h13_merkey h13_pid; run;
```

```
data hpc13_age15_up;
  set hpc13_age;
  by h13_merkey;
  if first.h13_merkey then age15_up_sum=0;
  age15_up_sum+age15_up;
  if last.h13_merkey;
run;
```

```
proc freq data=hpc13_age15_up; tables age15_up_sum; run;
```

```
/* 15세 미만 */
proc sort data=hpc13_age; by h13_merkey h13_pid; run;
```

```
data hpc13_age15_down;
  set hpc13_age;
  by h13_merkey;
  if first.h13_merkey then age15_down_sum=0;
  age15_down_sum+age15_down;
  if last.h13_merkey;
run;
```

```
proc freq data=hpc13_age15_down; tables age15_down_sum; run;
```

머지용 결과

age15_usum	빈도	백분율	누적 빈도	누적 백분율
1	2121	32,76	2121	32,76
2	2780	42,94	4901	75,7
3	950	14,67	5851	90,38
4	505	7,8	6356	98,18
5	104	1,61	6460	99,78
6	12	0,19	6472	99,97
7	1	0,02	6473	99,98
8	1	0,02	6474	100

age15_dsum	빈도	백분율	누적 빈도	누적 백분율
0	5393	83,3	5393	83,3
1	500	7,72	5893	91,03
2	494	7,63	6387	98,66
3	75	1,16	6462	99,81
4	9	0,14	6471	99,95
5	3	0,05	6474	100

데이터활용: 필요한 변수만 남기고 싶을 때

- keep, drop 활용

```
/* keep, drop 외에는 방법이 없음 */
/* 다만 sas에서는 proc sql을 사용하여 데이터의 변수 순서까지 변경이 가능 */
proc sql;
  create table hpc13_new as
  select h13_pid, h13_pind, h13_merkey, h_new, h13_cobf, p13_wgl, p13_wsl, p13_wgc, p13_wsc, h13_reg5, h13_reg7, h13_cin,
         h13_din, h13_hc, nh1301_1, nh1301_2, h1301_1, h1301_110, h13_g1, h13_g2, h13_g3, h13_g4, h13_g6, h13_g7, h13_g8,
         h13_g9, h13_g10, h13_g11, h13_g12
  from w13.Koweps_hpc13_2018_beta1;
quit;
```

```
/* SPSS 변환 */
PROC EXPORT DATA= w13.Koweps_h13_2018_beta1 /* 변환 하고자 하는 data명 */
  OUTFILE= "D:\#Kihasa\#Koweps\#panel data(1~13)\#2018년 13차 한국복지패널조사(koweps) 데이터 및 조사설계서(beta1)\#
           (2018년 13차 한국복지패널조사) 데이터(beta1)_spss\#Koweps_h13_2018_beta1"
  /* "저장하고자 하는 경로 및 데이터명 지정" */
  DBMS=SPSS REPLACE;
  /* 변환하고자 하는 패키지 */
RUN;
```

```
/* STATA 변환 */
PROC EXPORT DATA= w13.Koweps_h13_2018_beta1
  OUTFILE= "D:\#Kihasa\#Koweps\#panel data(1~13)\#2018년 13차 한국복지패널조사(koweps) 데이터 및 조사설계서(beta1)\#
           (2018년 13차 한국복지패널조사) 데이터(beta1)_stata\#Koweps_h13_2018_beta1"
  DBMS=STATA REPLACE;
RUN;
```