재 정 패 널 조 사

NaSTaB (National Survey of Tax and Benefit)

데이터 가공법 소개

2014. 8.

한국조세재정연구원

목　　차

※본 코드 사용 시 D드라이브에 데이터 저장 후 사용 권고

1. 데이터 합치기

가. 횡으로 결합 : 가구원용 + 가구용

○ STATA

```
use "d:\NaSTaB06P.dta", clear              /* 6차년도 가구원 데이터 열기*/
  rename hid06 hid                         /* key 변수 변수명 변경 : hid06에서 hid*/
  sort hid                                 /* key 변수 정렬*/
save "d:\a1.dta", replace                  /* a1 데이터 저장*/

use "d:\NaSTaB06H.dta", clear              /* 6차년도 가구 데이터 열기*/
  rename hid06 hid                         /* key 변수 변수명 변경 : hid06에서 hid*/
  sort hid                                 /* key 변수 정렬*/
save "d:\a2.dta", replace                  /* a2 데이터 저장*/

use "d:\a1.dta", clear                     /* a1 데이터 열기*/
  merge hid using "d:\a2.dta"              /* a2 데이터 가로 병합 */
save "d:\a3.dta", replace                  /* a1과 a2 가로 병합 데이터 a3.dta 데이터로 저장*/

  tab _merge                               /*병합 결과 보기*/
```

○ SPSS

```
get file="d:\NaSTaB06P.sav".               /* 6차년도 가구원 데이터 열기*/
dataset name a1 window=front.
   rename variables (hid06=hid).           /* key 변수 변수명 변경 : hid06에서 hid*/
   sort cases by hid.                      /* key 변수 정렬*/
   save out ='d:\a1.sav'                   /* a1로 데이터 저장*/
   compressed.

get file="d:\NaSTaB06H.sav".                /* 6차년도 가구 데이터 열기*/
dataset name a2.
   rename variables (hid06=hid).            /* key 변수 변수명 변경 : hid06에서 hid*/
   sort cases by hid.                       /* key 변수 정렬*/
   save out ='d:\a2.sav'.                   /* a2.로 데이터 저장*/

dataset active a1.                          /* a1 데이터 열기*/
match files /file=*                         /* a2와 종으로 결합*/
  /table="a2"
  /by hid.
   save out ='d:\a3.sav'.                   /* a1과 a2 종으로 결합한 데이터 a3로 저장*/
```

○ SAS

```
libname nastab "d:\";                          /* nastab 라이브러리 생성*/

data nastab.a1;                                /* 데이터 a1 생성*/
   set nastab.nastab06p;                       /* 6차년도 가구원 데이터 열기*/
   rename hid06=hid; run;                      /* key 변수 변수명 변경 : hid06에서 hid*/

data nastab.a2;                                /* 데이터 a2 생성*/
   set nastab.nastab06h;                       /* 6차년도 가구 데이터 열기*/
   rename hid06=hid; run;                      /* key 변수 변수명 변경 : hid06에서 hid*/

proc sort data=nastab.a1;                      /* a1 데이터 key 변수 정렬*/
   by hid; run;
proc sort data=nastab.a2;                      /* a2 데이터 key 변수 정렬*/
   by hid; run;

data nastab.a3;                                /* 데이터 a3 생성*/
   merge nastab.a1 nastab.a2; by hid; run;     /* a1과 a2 횡으로 결합 */
```

나. 횡으로 결합 : 가구용 + 가구용

○ STATA

```
use "d:\NaSTaB05H.dta", clear                 /* 5차년도 가구 데이터 열기*/
  rename hid05 hid                            /* key 변수 변수명 변경 : hid05에서 hid*/
  sort hid                                    /* key 변수 정렬*/
save "d:\b1.dta", replace                     /* b1 데이터 저장*/

use "d:\NaSTaB06H.dta", clear                 /* 6차년도 가구 데이터 열기*/
  rename hid06 hid                            /* key 변수 변수명 변경 : hid06에서 hid*/
  sort hid                                    /* key 변수 정렬*/
save "d:\b2.dta", replace                     /* b2 데이터 저장*/

use "d:\b1.dta", clear                        /* b1 데이터 열기*/
  merge hid using "d:\b2.dta"                 /* b2 데이터 가로 병합 */
save "d:\b3.dta", replace                     /* b1과 b2 가로 병합 데이터 a3.dta 데이터로 저장*/

  tab _merge                                  /* 병합 결과 보기*/
```

○ SPSS

```
get file="d:\NaSTaB05H.sav".              /* 5차년도 가구 데이터 열기*/
dataset name b1
   rename variables (hid05=hid).          /* key 변수 변수명 변경 : hid05에서 hid*/
   sort cases by hid.                     /* key 변수 정렬*/
   save out ='d:\b1.sav'                  /* b1로 데이터 저장*/
   compressed.

get file="d:\NaSTaB06H.sav".              /* 6차년도 가구 데이터 열기*/
dataset name b2.
   rename variables (hid06=hid).          /* key 변수 변수명 변경 : hid06에서 hid*/
   sort cases by hid.                     /* key 변수 정렬*/
   save out ='d:\b2.sav'.                 /* b2로 데이터 저장*/

dataset active b1.                        /* b1 데이터 열기*/
match files /file=*                       /* b2와 종으로 결합*/
  /table="b2"
  /by hid.
   save out ='d:\b3.sav'.                 /* b1과 b2 종으로 결합한 데이터 b3로 저장*/
```

○ SAS

```
libname nastab "d:\";                     /* nastab 라이브러리 생성*/

data nastab.b1;                           /* 데이터 b1 생성*/
   set nastab.nastab05h;                  /* 5차년도 가구 데이터 열기*/
   rename hid05=hid; run;                 /* key 변수 변수명 변경 : hid05에서 hid*/

data nastab.b2;                           /* 데이터 b2 생성*/
   set nastab.nastab06h;                  /* 6차년도 가구 데이터 열기*/
   rename hid06=hid; run;                 /* key 변수 변수명 변경 : hid06에서 hid*/

proc sort data=nastab.b1;                 /* b1 데이터 key 변수 정렬*/
   by hid; run;
proc sort data=nastab.b2;                 /* b2 데이터 key 변수 정렬*/
   by hid; run;

data nastab.b3;                           /* 데이터 b3 생성*/
    merge nastab.b1 nastab.b2; by hid; run;  /* b1과 b2 횡으로 결합 */
```

다. 종으로 결합 : 가구용 + 가구용

○ STATA

```
use "d:\NasTaB05H.dta", clear                  /* 5차년도 가구 데이터 열기*/
     keep hid05 h05ba001 h05ba002 h05ba003 /* 병합 변수 외 다른 변수 삭제*/
     gen time=2012                          /* 조사 년도 2012년 변수 생성*/
     rename hid05 hid                       /* 변수 명 변경*/
     rename h05ba001 ba001
     rename h05ba002 ba002
     rename h05ba003 ba003
save "d:\c1.dta", replace                       /* c1로 데이터 저장*/

use "d:\NasTaB06H.dta", clear                  /* 6차년도 가구 데이터 열기*/
     keep hid06 h06ba001 h06ba002 h06ba003 /* 병합 변수 외 다른 변수 삭제*/
     gen time=2013                           /* 조사 년도 2013년 변수 생성*/
     rename hid06 hid                        /* 변수 명 변경*/
     rename h06ba001 ba001
     rename h06ba002 ba002
     rename h06ba003 ba003
save "d:\c2.dta", replace                      /* c2로 데이터 저장*/

use "d:\c1.dta", clear                         /* c1 데이터 열기*/
     append using "d:\c2.dta"                 /* c2와 종으로 결합*/
save "d:\c3.dta", replace                      /* c1과 c2 횡으로 결합한 데이터 c3로 저장*/
```

○ SPSS

```
get file="d:\NaSTaB05H.sav"                    /* 5차년도 가구 데이터 열기*/
   /keep=hid05 h05ba001 h05ba002 h05ba003    /* 병합 변수 외 다른 변수 삭제*/
   /rename=(hid05 h05ba001 h05ba002 h05ba003=hid ba001 ba002 ba003). /* 변수 명 변경*/
dataset name c1.
   compute time=2012.                          /* 조사 년도 2012년 변수 생성*/
   save out ='d:\c1.sav'.                      /* c1로 데이터 저장*/

get file="d:\NaSTaB06H.sav"                    /* 6차년도 가구 데이터 열기*/
   /keep=hid06 h06ba001 h06ba002 h06ba003    /* 병합 변수 외 다른 변수 삭제*/
   /rename=(hid06 h06ba001 h06ba002 h06ba003=hid ba001 ba002 ba003). /* 변수 명 변경*/
dataset name c2.
   compute time=2013.                          /* 조사 년도 2013년 변수 생성*/
   save out ='d:\c2.sav'.                      /* c2로 데이터 저장*/
```

(계속)

```
dataset active c1.                      /* c1 데이터 열기*/
add files /file=*                       /* c2와 종으로 결합*/
          /file='c2'.
  save out ='d:\c3.sav'.                /* c1과 c2 횡으로 결합한 데이터 c3로 저장*/
```

○ SAS

```
libname nastab "d:\";                   /* nastab 라이브러리 생성*/

data nastab.c1;                         /* 데이터 c1 생성*/
  set nastab.nastab05h;                 /* 5차년도 가구 데이터 열기*/
  time=2012                             /* 조사 년도 2012년 변수 생성*/
  keep hid05 h05ba001 h05ba002 h05ba003; /* 병합 변수 외 다른 변수 삭제*/
  rename hid05=hid;                     /* 변수 명 변경*/
  rename h05ba001=ba001;
  rename h05ba002=ba002;
  rename h05ba003=ba003; run;
data nastab.c2;                         /* 데이터 c2 생성*/
  set nastab.nastab06h;                 /* 6차년도 가구 데이터 열기*/
  keep hid06 h06ba001 h06ba002 h06ba003; /* 병합 변수 외 다른 변수 삭제*/
  time=2013                             /* 조사 년도 2013년 변수 생성*/
  rename hid06=hid;                     /* 변수 명 변경*/
  rename h06ba001=ba001;
  rename h06ba002=ba002;
  rename h06ba003=ba003; run;
data nastab.c3;                         /* 데이터 c3 생성*/
  set nastab.c1 nastab.c2; run;         /* c1과 c2 횡으로 결합*/
```

2. 구조변환

가. 가구 데이터의 가구원 정보 구조변환

○ STATA

```
use "d:\NaSTaB06H.dta", clear                /* 6차년도 가구 데이터 열기*/

    /* 구조변환 : pid, 생년월일, 성별, 가구주와의 관계 변수 가구원 단위로 구조 변환 */
    reshape long hpid0 w06byr0 w06gen0 w06rel0 , i(hid06) j(num)

    rename hpid0 pid06                       /* 변수 명 변경*/
    sort pid06                               /* key 변수 정렬*/
    keep pid06 hid06 w06byr0 w06gen0 w06rel0 num /*구조변환 변수 외 변수 제외*/
    drop if pid06==.                         /* key 변수 결측치 제외*/

save "d:\long.dta", replace                  /* 구조변환 결과 long으로 저장*/
```

○ SPSS

```
get file="d:\NaSTaB06H.sav".                 /* 6차년도 가구 데이터 열기*/
dataset name long.                           /* 데이터 명 long 설정*/

 /* 구조변환 : pid, 생년월일, 성별, 가구주와의 관계 변수 가구원 단위로 구조 변환 */
varstocases
   /make pid06 from hpid01 hpid02 hpid03 hpid04 hpid05 hpid06 hpid07 hpid08
hpid09
   /make w06gen from w06gen01 w06gen02 w06gen03 w06gen04 w06gen05 w06gen06
w06gen07 w06gen08 w06gen09
   /make w06rel from w06rel01 w06rel02 w06rel03 w06rel04 w06rel05 w06rel06 w06rel07
w06rel08 w06rel09

   /index=index1(9)
   /keep=hid06
   /null =keep .
   select if pid06~=sysmis(pid06).           /* key 변수 결측치 제외*/
   save out ='d:\long.sav'. /*297cases.      /* 구조변환 결과 long 데이터 저장*/
```

○ SAS

```
libname nastab "d:\";                  /* nastab 라이브러리 생성*/

data nastab.long;                      /* 데이터 long 생성*/
     set nastab.nastab06h;             /* 6차년도 가구 데이터 열기*/

     array a[9] w06byr01-w06byr09;     /*출생년도 변수 배열 선언*/
     array b[9] w06gen01-w06gen09;     /*성별 변수 배열 선언*/
     array c[9] w06rel01-w06rel09;     /*가구주와의 관계 변수 배열 선언*/
     array d[9] hpid01-hpid09;         /*생년월일 변수 배열 선언*/

       do num=1 to 9;                  /*1~9번 가구원 정보 구조 변환*/
       w06byr=a[num]; w06gen=a[num]; w06rel=a[num]; hpid=a[num];
       if hpid ne . then output; end;  /* key 변수 결측치 제외*/
     keep hid06 num hpid w06byr; run;  /*구조변환 변수 외 변수 제외*/
```

나. 가구원 데이터의 가구 정보 구조변환

○ STATA

```
use "d:\NaSTaB06P.dta", clear          /* 6차년도 가구원 데이터 열기*/
     keep pid06 hid06 ps06a04 p06bb002 /*구조변환 변수 외 변수 제외*/
     /* 구조변환 : pid, p06bb002 변수 가구단위로 구조 변환 */
     reshape wide  pid06  p06bb002 , i( hid06) j(ps06a04)
     sort hid06                        /* key 변수 정렬*/
save "d:\wide.dta", replace            /* 구조변환 결과 wide 데이터 저장*/
```

○ SPSS

```
get file="d:\NaSTaB06P.sav"            /* 6차년도 가구원 데이터 열기*/
   /keep=pid06 hid06 ps06a04 p06bb002. /*구조변환 변수 외 변수 제외*/
dataset name wide.                     /*데이터 wide 생성*./
   sort cases by  hid06 ps06a04.       /* key 변수 정렬*/
/* 구조변환 : 가구단위로 구조 변환 */
casestovars
  /id=hid06
  /index=ps06a04
  /count=num "생성 가구원".
   save out ='d:\wide.sav'.            /* 구조변환 결과 wide 데이터 저장*/
```

○ SAS

```
data nastab.wide/* 데이터 wide 생성*/
(keep= hid06  pid061-pid069 ps06a04 p06aa0011-p06aa0019 p06aa0021-p06aa0029);
/* 구조변환 : pid, p06bb002 변수 가구단위로 구조 변환
   1번 ~9번 가구원에 대해 각각 데이터 분리 후 key 변수인 hid06 기준으로 횡으로 결합*/
merge
nastab.nastab06p
(rename=(pid06 =pid061 p06aa001=p06aa0011 p06aa002=p06aa0021) where=(ps06a04=1))
nastab.nastab06p
(rename=(pid06 =pid062 p06aa001=p06aa0012 p06aa002=p06aa0022) where=(ps06a04=2))
nastab.nastab06p
(rename=(pid06 =pid063 p06aa001=p06aa0013 p06aa002=p06aa0023) where=(ps06a04=3))
nastab.nastab06p
(rename=(pid06 =pid064 p06aa001=p06aa0014 p06aa002=p06aa0024) where=(ps06a04=4))
nastab.nastab06p
(rename=(pid06 =pid065 p06aa001=p06aa0015 p06aa002=p06aa0025) where=(ps06a04=5))
nastab.nastab06p
(rename=(pid06 =pid066 p06aa001=p06aa0016 p06aa002=p06aa0026) where=(ps06a04=6))
nastab.nastab06p
(rename=(pid06 =pid067 p06aa001=p06aa0017 p06aa002=p06aa0027) where=(ps06a04=7))
nastab.nastab06p
(rename=(pid06 =pid068 p06aa001=p06aa0018 p06aa002=p06aa0028) where=(ps06a04=8))
nastab.nastab06p
(rename=(pid06 =pid069 p06aa001=p06aa0019 p06aa002=p06aa0029) where=(ps06a04=9))
; by hid06; run;
```

3. 기초 통계 분석

○ STATA

```
/*기초 통계량산출*/
use "d:\a1.dta", clear
     sum h06ca004 [iweight=h06cwt]          /*h06ca004 평균, 분산, 최대값 최소값*/
     sum h06ca004 [aweight=h06cwt] , d      /*h06ca004 분포: 가중치 aweight사용*/
/*백분위 산출*/
sort h06ca004
pctile  h06ca004pct =  h06ca004 [aweight=h06cwt], nq(5)
list  h06ca004pct in 1/5
/*빈도 분석*/
use "d:\NaSTaB06P.dta", clear
     tab p06aa005
     tab p06aa006 [iweight=p06cwt]
     tab p06aa005 p06aa006 [iweight=p06cwt]
```

○ SPSS

```
/*기초 통계량산출 : 평균, 최대값, 최소값*/
dataset active a1.
weight by h06cwt.
des h06ca004
   /stat=mean min max .
/*기초 통계량산출 : 중앙값*/
fre h06ca004
   /stat=median
   /format=notable.
weight off.
/*백분위 산출*/
rank ps06a13 (a)
   /ntiles(5)
   /ties=mean.
```

○ SAS

```
/* 평균, 중앙값, 분포*/
proc univariate data=nastab.nastab06h; var h06ca004; weight h06cwt; run;
/* 빈도 분석*/
proc freq data=nastab.nastab06h; table W06gen01;  run;
proc freq data=nastab.nastab06h; weight h06cwt;  table p06aa005;  run;
proc freq data=nastab.nastab06h; weight h06cwt;  table p06aa005 * p06aa006;  run;
```

4. 패널 데이터 분석 (STATA)

○ 패널 데이터로 쌓기

```
/*Append*/
use "d:\NasTaB05HPanel.dta", clear              /*5차년도 가구 데이터 열기*/
     append using "d:\NasTaB06HPanel.dta"       /*6차년도 가구 데이터 종으로 결합*/
save "d:\NasTaBHPanel.dta", replace             /*결과 NasTaBHPanel로 저장*/

/*Panel Data setting*/
use "d:\NasTaBHPanel.dta", clear                /*5차년도 가구 데이터 열기*/

by hid, sort: gen time=_n                       /*key변수 정렬, 시간 변수 생성*/
xtset hid time                                  /*key 변수 hid, time 변수로 패널 데이터 설정*/
xtdes                                           /*패널 데이터 description*/
```

※ NasTaB05HPanel/NasTaB06HPanel 데이터는 동일한 변수명을 갖도록 변수명을 변경하여 생성

○ 패널 분석

```
/*기초통계량 분석*/
xtsum x                                         /* 변수 x 에 대한 기초 통계량 분석*/
xttab x                                         /* 변수 x 에 빈도 분석*/

/*Analysis : y = x'β, i = 1,2...N, t = 1,2...T */
xtreg y x, fe                                   /*Fixed Effect Model*/
estimate store fe                               /*Fixed Effect Model 결과 저장*/
xtreg y x, re                                   /*Random Effect Model*/
estimate store re                               /*Random Effect Model 결과 저장 */
hausman fe re, sigmamore                        /*Hausman Test*/
```