

User's Guide : smart_nastab_v2

A.1 패키지 설치절차

smart_nastab_v2 패키지 사용하기 위해서는 표 1-1의 파일들이 필요하다. *.ado 파일은 명령어가 코딩된 파일이다. *.dlg 파일은 명령어를 대화창으로 실행하는 데 필요한 dialog box 디자인 파일이다. 표 1의 모든 파일이 영문자 알파벳 “s”로 시작하기 때문에 c:\ado\s 폴더에 복사하면 설치작업이 완료된다.* c:\ado\s 폴더에 이미 같은 이름의 파일이 있다면 덮어쓰기(overwrite)를 하는 것을 추천한다.

<표 1> Stata 파일목록

ado file	dlg file & sthlp file
smart_nastab_h_v2.ado smart_nastab_p_v2.ado smart_nastab_m_v2.ado smart_nastab_v2.ado smart_nastab_add_p_v2.ado smart_nastab_add_h_v2.ado smart_nastab_search_v2.ado	smart_nastab_v2.dlg

smart_nastab_v2 관련 명령어를 실행하기 위해서는 NaSTaB 원 데이터가 필요하다. 원 데이터는 반드시 *.dta 파일 형식을 취해야 하고 표 2와 같은 파일을 이름을 가져야 한다. 2021년 2월 현재 조세재정연구원 재정패널 홈페이지 (<http://panel.kipf.re.kr/>)에서 내려받을 수 있는 wave는 1차 ~ 12차 가구와 가구원 데이터 파일이다.

* c:\ado\s 폴더가 사용자 컴퓨터에 없다면 임의로 만들면 된다.

<표 2> 가구와 가구원 데이터 파일목록

wave	가구 (H)	가구원 (P)
1차 (2008년)	NaSTaB01H.dta	NaSTaB01P.dta
2차 (2009년)	NaSTaB02H.dta	NaSTaB02P.dta
3차 (2010년)	NaSTaB03H.dta	NaSTaB03P.dta
4차 (2011년)	NaSTaB04H.dta	NaSTaB04P.dta
5차 (2012년)	NaSTaB05H.dta	NaSTaB05P.dta
6차 (2013년)	NaSTaB06H.dta	NaSTaB06P.dta
7차 (2014년)	NaSTaB07H.dta	NaSTaB07P.dta
8차 (2015년)	NaSTaB08H.dta	NaSTaB08P.dta
9차 (2016년)	NaSTaB09H.dta	NaSTaB09P.dta
10차 (2017년)	NaSTaB10H.dta	NaSTaB10P.dta
11차 (2018년)	NaSTaB11H.dta	NaSTaB11P.dta
12차 (2019년)	NaSTaB12H.dta	NaSTaB11P.dta
총 22개 데이터 파일		

smart_nastab 관련 모듈을 실행하기 위해서는 반드시 Stata 14.0 이상 버전이 필요하다. Stata 13 버전에서는 실행되지 않음에 주의해야 한다. 13버전 데이터 파일은 unicode encoding 작업을 통해 14버전으로 전환해야 한다. 자세한 내용은 A.4을 참고하라.

<표 3> 필요한 Stata 버전

Stata 14/15 버전 (IC, SE 또는 MP) 이상

A.2 패키지 사용방법 및 주의사항

smart_nastab_v2 패키지를 사용하기 위해서는 A.1에서 설명한 대로 필요한 파일을 설치해야 한다. 그런 다음 대화창(dialog box)을 불러오는 명령어를 입력한다.

```
. db smart_nastab_v2
```

그림 1. smart_nastab_v2 대화창

먼저 select command에서 연구자가 사용하고자 하는 명령어를 선택한다. “smart_nastab_v2”를 선택하면 저자가 미리 코딩해 놓은 변수를 선택하여 패널데이터 파일로 만들 수 있다. 가구 레벨 변수와 가구원 레벨 변수에서 연구자가 임의로 변수를 선택한다. 선택 가능한 가구와 가구원 변수에 대한 자세한 설명은 smart_nastab_v2_codebook.xlsx 파일을 참고하라.

add variables 옵션에서는 연구자가 (저자가 미리 세팅해 놓은 변수 외) 추가하고자 변수가 있는 경우 사용한다. 가구 레벨(NaSTaB__H)에서 가져오는 변수인 경우 add variables(가구레벨)에서 입력한다. 가구원 레벨(NaSTaB__P)

에서 가져오는 경우 add variables(가구원레벨)에서 입력한다.

add variables 옵션에서 변수 이름을 쓰는 규칙을 이해할 필요가 있다. NaSTaB 변수이름은 wave가 포함되어 있다. wave에 해당하는 부분을 제외한 나머지 이름을 그대로 써주면 된다. 가령 1차~12차 가구레벨 데이터에 있는 h01fb030~h12fb030과 h01fb031~h12fb031 변수를 가져오고자 하면 hfb030와 hfb0301이라고 입력하면 된다. NaSTaB_P(가구원 레벨)에 있는 변수에 대해서도 같은 방식을 적용한다.

h01fb030	h02fb030	h03fb030	h04fb030	h05fb030	hfb030
h06fb030	h07fb030	h08fb030	h09fb030	h10fb030	
h11fb030	h12fb030				
h01fb031	h02fb031	h03fb031	h04fb031	h05fb031	hfb031
h06fb031	h07fb031	h08fb031	h09fb031	h10fb031	
h11fb031	h12fb031				

add variables를 입력할 때 연구자가 입력한 변수와 wave가 꼭 일치할 필요는 없다. 예를 들어 hfb021(주택 이외의 보유부동산: 토지) 변수는 1차 ~ 4차에는 조사항목이 없고 5차 ~ 12차에만 존재한다. hfb021이라고 쓰고 wave(01 02 03 04 05 06 07 08 09 10 12) 옵션을 사용하면 hfb021 변수는 1차 ~ 4차 wave에서는 결측치로 처리되고 5차 ~ 12차 wave에서 조사된 값을 가져온다. 연구자가 추가하고자 하는 변수가 어떤 wave에 존재하는지 미리 확인할 필요가 없다는 장점이 있다.

wave 옵션에서는 연구자가 패널데이터로 만들고자 하는 wave를 지정한다. wd 옵션에서는 <표 2>에 제시된 NaSTaB *.dta 파일이 저장되어 있는 폴더를 지정한다.* 실행하면 해당 명령문이 실행되고 마지막에 생성된 패널데이터는 **nastab_final_add.dta** 이름을 갖게 된다. save file 옵션은 저장된 패널데이터 이름을 연구자가 원하는 대로 바꾸고자 하는 경우 사용한다. save file 옵션에서 nastab_20190525라고 입력하면 패널데이터는 nastab_20190525.dta 파일로 저장된다.

* wd 옵션을 사용하는 대신 Stata 명령문으로 working directory를 변경하는 명령문을 먼저 실행한 후 smart_nastab_v2를 사용해도 된다. 가령 c:\temp에 표 2 데이터가 저장되어 있다면 다음과 같이 입력한다.

· cd c:\temp

smart_nastab_search_v2 명령어를 선택하면 NaSTaB 1차 ~ 12차 데이터에서 연구자에게 필요한 변수 이름을 확인한다. 필요한 하는 변수가 가구 레벨에 있으면 “household”를 선택하고 가구원 데이터에 있다면 “person”을 선택한다. word 옵션에서는 찾고자 하는 변수 레이블(variable label) 일부분을 입력하면 된다. 가령 “소득”이라고 입력하면 변수레이블이 “소득”을 포함하는 모든 변수 이름을 각 wave별로 보여준다. 변수 레이블뿐 아니라 변수이름 일부분을 입력해도 관련된 변수를 찾을 수 있다. 가령 “hch001”이라고 입력하면 각 wave에서 h_ch001 변수를 리스트해서 보여준다.

A.3 예제 코드

그림 2 예제에서는 저자가 미리 세팅한 가구 레벨 변수만을 선택하여 패널데이터를 만드는 명령문이다. working directory 는 미리 지정하였다고 c:\temp (예시)로 지정한다. describe 명령문 결과에서 알 수 있듯이 가구 변수 이름은 모두 “h_”로 시작하도록 코딩되어 있다.

그림 2. 가구 패널데이터

```
. cd c:\temp
. smart_nastab_v2 hage hgen hmar asset_fb asset_fc
deb_fc , wave( 07 08 09 10 11 )

. de
Contains data from nastab_final_add.dta
  obs:      24,212
  vars:       9
  size:    532,664
26 May 2019 16:05
```

variable name	storage type	display format	value label
hhid	long	%15.0g	가 구 id
wave	byte	%9.0g	서 베 이 차 수
year	int	%9.0g	조 사 년 도
h_asset_fa	float	%9.0g	[자산] 금융 자산 (단위 : 만 원)
h_asset_fb	float	%9.0g	[자산] 부동산 자산 및 기타 자산 (단위 : 만 원)
h_debt_fc	float	%9.0g	[부채] 부채 총액 (단위 : 만 원)
h_hage	byte	%9.0g	[가구주] (만)연령
h_hgen	byte	%9.0g	[가구주] 성별 (남=1,여=2)
h_hmar	byte	%9.0g	[가구주] 작년 말 기준 혼인상태 (배우자있음=1,배우자없음/미혼=2)

그림 3 예제에서는 저자가 미리 세팅한 가구원 레벨 변수만을 선택하여 패널데이터를 만드는 명령문이다. describe 명령문 결과에서 알 수 있듯이 가구원 변수 이름은 모두 “p_”로 시작하도록 코딩되어 있다.

그림 3. 가구원 패널데이터

```
. smart_nastab_v2 pgen page prel pedu aa200 aa005 ,
wave( 07 08 09 10 11 )
```

```
. de
```

Contains data from nastab_final_add.dta

obs: 67,522
vars: 10
size: 1,215,396

26 May 2019 16:11

variable name	storage type	display format	value label	variable label
hhid	long	%15.0g		가구 id
pid	long	%15.0g		가구원 id
wave	byte	%9.0g		서베이차수
year	int	%9.0g		조사년도
p_aa005	byte	%9.0g		[경찰상태] (현재, 조사년도 5월 1일 기준) 총사상지위 (1=상용직 2=원시/일용직 3=자영업 4=무급가족종사자)
p_aa200	byte	%9.0g		[경찰상태] (현재, 조사년도 5월 1일 기준) 취업상태 (1=취업 2=전업주부 3=무직 4=학생)
p_page	int	%9.0g		[가구원] 연명
p_pedu	byte	%9.0g		[가구원] 최종 교육수준 (1=중졸이하, 2=고졸, 3=전문대졸이상)
p_pgen	byte	%9.0g		[가구원] 성별
p_prel	byte	%9.0g		[가구원] 가구주와의 관계코드

그림 4 예제에서는 가구 레벨과 가구원 레벨 변수를 동시에 지정하는 경우이다. 순서와 상관없이 저자가 미리 저장해 놓은 가구와 가구원 변수 이름을 입력하면 된다. 가구 데이터에 생성한 변수는 “h_”로 시작하고 가구원 데이터에서 생성한 변수는 “p_” 시작하는 것을 알 수 있다. 같은 가구에 속한 가구원에게 가구 변수는 당연히 같은 변수값을 갖게 된다.

그림 4. 가구+가구원 패널데이터

```
. smart_nastab_v2 hage hgen hmar asset_fa asset_fb  
debt_fc pgen page prel pedu aa200 aa005 , wave( 07 08  
09 10 11 )
```

```
. de
```

Contains data from nastab_final_add.dta

obs: 67,716
vars: 16
size: 2,234,628

26 May 2019 16:24

variable name	storage type	display format	value label	variable label
hhid	long	%15.0g		가구 id
pid	long	%15.0g		가구원 id
wave	byte	%9.0g		서베이차수
year	int	%9.0g		조사년도
h_asset_fa	float	%9.0g		[자산] 금융자산 (단위: 만원)
h_asset_fb	float	%9.0g		[자산] 부동산 자산 및 기타자산 (단위: 만원)
h_debt_fc	float	%9.0g		[부채] 부채총액 (단위: 만원)
h_hage	byte	%9.0g		[가구주] (만)연령
h_hgen	byte	%9.0g		[가구주] 성별 (남=1,여=2)
h_hmar	byte	%9.0g		[가구주] 직년월 기준 혼인상태 (배우자있음=1,배우자없음/미혼=2)
p_aa005	byte	%9.0g		[경찰상태] (현재, 조사년도 5월 1일 기준) 총사상지위 (1=상용직 2=원시/일용직 3=자영업 4=무급가족종사자)
p_aa200	byte	%9.0g		[경찰상태] (현재, 조사년도 5월 1일 기준) 취업상태 (1=취업 2=전업주부 3=무직 4=학생)
p_page	int	%9.0g		[가구원] 연명
p_pedu	byte	%9.0g		[가구원] 최종 교육수준 (1=중졸이하, 2=고졸, 3=전문대졸이상)
p_pgen	byte	%9.0g		[가구원] 성별
p_prel	byte	%9.0g		[가구원] 가구주와의 관계코드

그림 5에서는 add variables 옵션을 사용한 예제이다. 가구 변수인 hgen, hage, hedu 변수를 지정하고 add_h() 옵션에서 haa013(주말부부 배우자 연간소득), wmar01(작년말 기준 혼인상태_가구원 1)을 추가한다. 가구원 변수인 pbb001(근로소득 유무) 변수를 역시 추가한다.

그림 5. add variables 옵션

```
. smart_nastab_v2 hgen hage hedu , wave(07 08 09 10 11) add_h(haa013 wmar01 ) add_p(pbb001 )
```



```
. de
```

Contains data from **nastab_final_add.dta**

obs: 38,784
vars: 10 26 May 2019 16:31
size: 659,328

variable name	storage type	display format	value label	variable label
hhid	long	%15.0g		가 구 id
pid	long	%15.0g		가 구 원 id
wave	byte	%9.0g		서 배 이 차 수
year	int	%9.0g		조 사 년 도
h_hage	byte	%9.0g		[가 구 주] {만 }연령
h_hedu	byte	%9.0g		[가 구 주] 교 육 수 준 (1=중 졸 이 하 , 2=고 졸 , 3=전 문 대 졸 이 상)
h_hgen	byte	%9.0g		[가 구 주] 성 별 (남 =1,여 =2)
haa013	byte	%8.0g	h07aa013	[aa.가 구 기 본 현 황] 주 말 부 부 배 우 자 연 간 소 득
wmar01	byte	%8.0g	w07mar01	[ab.가 구 월 인 직 현 황] 작 년 말 기 준 혼 인 상 태 - 가 구 원 01
pbb001	byte	%8.0g	p07bb001	[bb.연 간 소 득] 근 로 소 득 유 무

그림 6 예제에서는 smart_nastab_search_v2 명령문을 보여 준다. wave를 지정하고 hp(h)는 가구 데이터에서 변수이름을 찾고자 할 때 사용한다. word 옵션에서 변수 레이블 또는 변수 이름 일부분을 지정한다. 실행결과를 확인하면 NaSTaB10H.dta 파일 내에 있는 변수 중 “근로소득” 레이블을 포함한 변수이름을 확인할 수 있다.

그림 6. smart_nastab_search_v2 예제

```
. smart_nastab_search_v2 , wave(10 11) hp(h) word(근로소득)
```


===== Na5TaB10H.dta 파일에서 근로소득률 (율) 포함한 변수 리스트 =====				
variable name	storage type	display format	value label	variable label
h10ka005	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 01
h10ka010	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 02
h10ka015	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 03
h10ka020	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 04
h10ka025	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 05
h10ka030	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 06
h10ka035	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 07
h10ka041	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 08
h10ka045	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 09
h10ka082	byte	%8.0g	labfi	[ka.가 구 설 문 종 료] 근로 소득 연 말 정 산 및 종 합 소 득 신고 여부 - 가 구 원 10

● dialog box tip

select command에서 특정 명령어를 실행한 후 그 다음 명령어를 실행하기 위해서는 이미 입력했던 내용을 모두 지워야 하는 불편함이 있다. dialbox 창의 왼쪽 맨 아래에 보면



위와 같은 3가지 선택메뉴가 있다. “R”을 클릭하면 대화창에서 입력했던 내용이 모두 지워진 후 초기화면으로 넘어간다. 반복적으로 dialog box를 사용하는 경우 “R”(reset) 기능을 적절히 활용할 수 있다.

A.4 unicode encoding 작업

재정패널 홈페이지에서 제공하는 Stata *.dta 파일은 13버전 형식으로 저장되어 있다. smart_nastab_v2 패키지를 사용하기 위해서는 14버전 Stata *.dta 파일이 필요하다. 표 2의 데이터 파일을 unicode encoding 하여 14버전을 저장하는 작업이 선행되어야 한다. 이를 위해 다음과 같은 명령문을 실행한다. NaSTaB 원 데이터가 저장된 폴더를 working directory로 지정한 후 unicode encoding 명령어를 실행한다. 아래 명령문을 실행하면 c:\temp 폴더에 14버전 데이터 파일이 자동으로 생성된다.

```
clear  
cd c:\temp // working directory 예시  
unicode encoding set korean  
unicode translate *.dta , invalid
```