

Understanding Society User Support - Support #145

working with two waves and weights

05/08/2013 07:07 PM - Giulia Montresor

Status:	Closed	Start date:	05/08/2013
Priority:	High	% Done:	100%
Assignee:	Giulia Montresor		
Category:	Data analysis		
Description			
<p>Hello,</p> <p>I want to use both waves 1-2 of US. In such case I think I need to keep only those individuals that responded in both waves. I have to estimate the mean of life satisfaction for different groups of immigrants over the two years. Therefore I need to use the longitudinal self-completion weight, b_indscus_lw.</p> <p>I don't understand one thing:</p> <p>The weight regards only observations of wave 2, therefore I need to keep only these observations? doing so I end up with a miserable number of individuals, and I cannot estimate the life satisfaction mean for the groups because the observations have zero weight. I attached my do file. I look forward to your kind reply, thanks</p>			

History

#1 - 05/08/2013 07:17 PM - Redmine Admin

Giulia,

We don't allow attachments, but you can copy and paste a table and maybe a snippet of the syntax to illustrate the problem [hit "update" at the bottom of the post].

Jakob

#2 - 05/09/2013 09:11 AM - Giulia Montresor

Hello,

thank you for the prompt reply. After joining or replacing the variables with the relative feed forward, I append the two waves of "indresp" and recode the missings, and this is a sketch of my results:

. summ macob pacob ukborn sclfsato

Variable	Obs	Mean	Std. Dev.	Min	Max
macob	55835	11.35709	25.02096	1	97
pacob	55587	11.62329	25.18556	1	97
ukborn	58755	2.051247	1.609738	1	5
sclfsato	82982	5.231713	1.475936	1	7

. * 1st gen immigrants

. gen gen1=0

. replace gen1=1 if macob>4 & pacob>4 & ukborn==5 & ukborn!=. & macob!=. & pacob!=.

(8905 real changes made)

. *Keep only individuals from wave 2, who responded in both waves (?)

. bysort pidp: gen q = _N

. keep if q==2

(28815 observations deleted)

. keep if wave==2

(38388 observations deleted)

. tab gen1

gen1	Freq.	Percent	Cum.
0	38,230	99.59	99.59
1	158	0.41	100.00
Total	38,388	100.00	

*Life satisfaction mean estimation

. svyset, clear

. svyset psu [pweight=indscus_lw], strata(strata)

```
pweight: indscus_lw
      VCE: linearized
Single unit: missing
Strata 1: strata
SU 1: psu
FPC 1: <zero>;
```

```
. svy, subpop (if gen1==1):mean sclfsato
(running mean on estimation sample)
all observations in subpop() subpopulation have zero weights
r(461);
```

#3 - 05/09/2013 04:36 PM - Redmine Admin

- % Done changed from 0 to 10

Born outside UK with both parents born outside UK - from Wave 1, 2 and both.
Looks like your variable gen1 could be the problem?

```
use pidp a_macob a_pacob a_ukborn using a_indresp, clear
merge 1:1 pidp using b_indresp, keepus(b_macob b_pacob b_ukborn)
g x=(a_ukborn==5 & a_macob>4 & a_pacob>4) | (b_ukborn==5 & b_macob>4 & b_pacob>4)
```

```
. ta _m x
```

		x		
	_merge	0	1	Total
master only (1)		9,846	2,760	12,606
using only (2)		15,639	570	16,209
matched (3)		32,833	5,555	38,388
Total		58,318	8,885	67,203

#4 - 05/09/2013 05:54 PM - Giulia Montresor

Hi,

what do you mean for the variable gen1 being the problem? I created this variable to indicate first generation immigrants.

Do you then confirm that, in order to estimate with the weights, I have to keep only observations from wave 2?

I also noted that the longitudinal self-completion weight b_indscus_lw is not positive for all the positive observations of life satisfaction. Shouldn't it be positive for all the non missing observations related to the self-completion questionnaire?

Thanks

#5 - 05/09/2013 07:29 PM - Redmine Admin

Giulia,

So far, I have just tried to reproduce the first few lines and there seems to many more that would fit the definition when looking at Wave 1, 2 or both.

Jakob

#6 - 05/16/2013 09:34 AM - Giulia Montresor

Hello,

I have 8905 observations for gen1 after appending the two indresp files, as I joined b_pacob with its feed forward variable b_ff_pacob.

However, since I want to work with both waves, I need to keep only those individuals that responded in both waves. By doing this 28815 observations are deleted.

Moreover, from what I understood, if I want to estimate using the longitudinal weight, I have to keep only the observations from wave 2. In this case, other 38388 observations are deleted.

This is why I end up with so few observations.

Here below you can see my stata output in major detail.

```
. use pidp b_psu b_strata b_indscus_lw b_macob b_pacob b_ff_pacob b_ff_macob b_ukborn b_sclfsato ///
      using "$dirdata\b_indresp", clear

. gen wave=2

. rename b_pacob b_pacob_new

. gen b_pacob=.
(54597 missing values generated)

. replace b_pacob=b_pacob_new if b_ff_pacob<0
```

```
(49675 real changes made)

. replace b_pacob=b_ff_pacob if b_pacob_new<0
(47410 real changes made)

. rename b_macob b_macob_new

. gen b_macob=.
(54597 missing values generated)

. replace b_macob=b_pacob_new if b_ff_macob<0
(49660 real changes made)

. replace b_macob=b_ff_macob if b_macob_new<0
(47397 real changes made)

.

. drop b_ff_pacob b_pacob_new b_ff_macob b_macob_new

.

. renpfix b_

. compress
wave was float now byte
pacob was float now byte
macob was float now byte

. save "$dirresults\bind_junk", replace
file C:\Users\gmontr\Desktop\Dropbox\Research\bind_junk.dta saved

.

. use pidp a_hidp a_psu a_strata a_macob a_pacob a_ukborn a_sclfsato ///

    using "$dirdata\a_indresp", clear

.

. gen wave=1

.

. renpfix a_

. append using "$dirresults\bind_junk"

. compress
wave was float now byte

. save "$dirresults\abind_long", replace
file C:\Users\gmontr\Desktop\Dropbox\Research\abind_long.dta saved

. tsset pidp wave
panel variable: pidp (unbalanced)
time variable: wave, 1 to 2
delta: 1 unit

.

. mvdecode _all, mv(-9/-1)
ukborn: 46836 missing values generated
pacob: 50004 missing values generated
macob: 49756 missing values generated
sclfsato: 22609 missing values generated

.

. summ macob pacob ukborn sclfsato
```

Variable	Obs	Mean	Std. Dev.	Min	Max
macob	55835	11.35709	25.02096	1	97
pacob	55587	11.62329	25.18556	1	97
ukborn	58755	2.051247	1.609738	1	5
sclfsato	82982	5.231713	1.475936	1	7

```
. *Immigrant groups
```

```
. * 1st gen
. gen gen1=0

. replace gen1=1 if macob>4 & pacob>4 & ukborn==5 & ukborn!=. & macob!=. & pacob!=.
(8905 real changes made)
```

```
.
. tab gen1
```

gen1	Freq.	Percent	Cum.
0	96,686	91.57	91.57
1	8,905	8.43	100.00
Total	105,591	100.00	

```
.
. *Keep only individuals from wave 2, who responded in both waves
. bysort pidp: gen q = _N
```

```
. *browse pidp wave q
. keep if q==2
(28815 observations deleted)
```

```
. keep if wave==2
(38388 observations deleted)
```

```
.
. tab gen1
```

gen1	Freq.	Percent	Cum.
0	38,230	99.59	99.59
1	158	0.41	100.00
Total	38,388	100.00	

```
. *Statistics representative of UK population
. svyset, clear
```

```
. svyset psu [pweight=indscus_lw], strata(strata)
```

```
pweight: indscus_lw
      VCE: linearized
Single unit: missing
  Strata 1: strata
      SU 1: psu
      FPC 1: <zero>
```

```
. svy, subpop (if gen1==1):mean scfsato
(running mean on estimation sample)
all observations in subpop() subpopulation have zero weights
r(461);
```

#7 - 05/17/2013 01:29 PM - Olena Kaminska

Dear Giulia,

Your problem seem to be less related to weights and more related to the variables you use. `b_pacob` is asked only to those people who have not responded in previous wave (see wave 2 questionnaire - and pay attention to the universe for the question). This variable does not change over time, so you should combine information from all waves.

If you are using only `pacob` variable you can conduct cross-sectional analysis and use cross-sectional weight `b_indinus_xw`; if you are interested in change - use longitudinal weight `b_indinus_lw`.

Some loss of respondents when you merge waves 1 and 2 is due to BHPS being present only in wave 2, and many respondents who are TSMs are not followed. You don't need to worry about it as long as you use correct weight - it will take care of picking the right respondents for you.

Hope this helps,
Olena

#8 - 05/30/2013 11:49 AM - Redmine Admin

- Status changed from New to Closed

- % Done changed from 10 to 100

#9 - 11/10/2015 02:36 PM - Gundi Knies

- *Target version set to X M*