
Subject: recoding v034X (line number fo husband)
Posted by [mm4599](#) on Tue, 31 May 2016 15:12:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

I am using the 2003/04 Tanzania HIV/AIDS Indicator Survey, the 2007/08 Tanzania AIDS Indicator Survey (AIS) and the 2011/12 AIS

The sample size in the 2003/04 individual recode file (IR): tzir4afl.dta is 12,522. The line number of husband variable (v034X) variable is wide and there are no observations in v034_5 - v034_8. Tabulating v034_1 - v034_4 reveals that:

```
v034_1 v034_2 v034_3 v034_4
0 677 203 22 4 906
1 3,541 2 3543
2 2,635 36 6 2677
3 199 38 3 240
4 119 4 4 127
5 50 3 1 54
6 40 4 44
7 35 2 37
8 33 2 35
9 22 3 1 26
10 21 1 22
11 14 14
12 10 10
13 5 1 6
14 6 6
15 3 3
16 2 2
17 2 2
20 1 1
22 1 1
31 1 1
35 1 1
36 2 2
39 1 1
42 1 1
43 1 1
0
Total 7,422 300 36 5 7,763
```

When you reshape the file (`gen id = _n; reshape long v034_ , i(id) j(linenum)`) and tabulate you get `v034_ = 7,763` but the dataset increases to 50,088 records with 42,325 missing observations for the `v034_` variable.

What do I need to do so that the sample size for the reshaped file is NOT 50,088?

When I create a variable (`gen v034= .`) and use the replace command to group the multiple

husband line number categories, I get 12,522 records (7422+ 5100), which is not correct, tab v034_

v034_ Freq. Percent Cum

0	677	9.12	9.12
1	3,541	47.71	56.83
2	2,635	35.5	92.33
3	199	2.68	95.01
4	119	1.6	96.62
5	50	0.67	97.29
6	40	0.54	97.83
7	35	0.47	98.3
8	33	0.44	98.75
9	22	0.3	99.04
10	21	0.28	99.33
11	14	0.19	99.51
12	10	0.13	99.65
13	5	0.07	99.72
14	6	0.08	99.8
15	3	0.04	99.84
16	2	0.03	99.87
17	2	0.03	99.89
20	1	0.01	99.91
31	1	0.01	99.92
35	1	0.01	99.93
36	2	0.03	99.96
39	1	0.01	99.97
42	1	0.01	99.99
43	1	0.01	100
Total	7,422	100	

Another thing I tried was to spilt the dataset into 4 files: one for v034_1, one for v034_2, one for v034_3, and one for v034_4 and then merge the files. Doing so I obtain 12,863 records (7763 + 5100 missing or 12522+341)

Which is the correct way to reformat this dataset so that I only have one variable v034?

I am attaching the do file.

Thanks

File Attachments

1) [v034X Recode_Reshape.do](#), downloaded 594 times

Subject: Re: recoding v034X (line number fo husband)

Posted by [Trevor-DHS](#) on Mon, 06 Jun 2016 15:06:09 GMT

[View Forum Message](#) <> [Reply to Message](#)

What are you trying to do exactly? Why do you need to reshape v034x from wide to long? The only reason I can see for doing this would be to create a couples file where you match husbands and wives together. If that is the case, then, after reshaping v034x, you can simply drop all cases where v034x==. This will leave you with a file of women and men that you can then merge with the original file (v001 v002 v003 in the original file will match with v001 v002 v034x in the reshaped file). You will also have to do quite a lot of either renaming or cloning of variables for the characteristics of the couples so you have separate variables for each half of the couple.

Subject: Re: recoding v034X (line number fo husband)

Posted by [mm4599](#) on Fri, 10 Jun 2016 13:53:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi Trevor

Thanks for the response. Yes, I am trying to create a couples file and then, after linking HIV status info, I would like to determine if couples are discordant. I am using the 2003/04 Tanzania HIV/AIDS Indicator Survey, the 2007/08 Tanzania AIDS Indicator Survey (AIS) and the 2011/12 AIS.

The v034 variable only needed to be reshaped in the 2003/04 AIS. As you suggest, after reshaping (using the following two commands: 1) gen id = _n; 2) reshape long v034_, i(id) j(linenum), I dropped dropped the cases where v034x==. and then merged this file with the original file.

My next step is to do an individual file and HIV file merge as follows

* Step 1: open AR file

use "xxAR61FL.DTA", clear

* Step 2: rename identifying variables

renvars hivclust hivnumb hivline / v001 v002 v003

* Step 3: sort by a unique identifier which I constructed from identifying variables (v001 v002 v003) as follows uid= v001*100000 + v002*100 + v003.

sort uid

* Step 4: save results

save "xxAR61FL_mergeprep.DTA", replace

* Step 5: open IR file

use "xxCR61FL.DTA", clear

* Step 6: sort by identifying variables

sort uid

* Step 7: merge!

merge uid using "xxAR61FL_mergeprep.DTA"

* Step 8: Complete the merge

drop if _merge==2

*Step 9: Split the merged dataset into two datasets, one for women and one for men

*Step 10: Rename the added hiv variable in the female dataset to so that it is unique for women (rename hiv03 hiv03f) and unique to men in the male dataset (rename hiv03 hiv03m)

To match couples, is the next step to merge both files into one doing the merge on v001 v002 and v034? Or do you have another suggestion?

Thanks for your assistance.

Best regards

Subject: Re: recoding v034X (line number fo husband)
Posted by [Trevor-DHS](#) on Mon, 27 Jun 2016 20:34:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

Try the following code:

```
* Step 1: open AR file
use "TZAR4AFL.DTA", clear
* Step 2: rename identifying variables
rename hivclust v001
rename hivnumb v002
rename hivline v003
* Step 3: sort according to ID vars
sort v001 v002 v003
* Step 4: save results
save "TZAR4AFL_mergeprep.DTA", replace
```

```
* Step 5: open IR file
use "TZIR4AFL.DTA", clear
* Step 6: sort by identifying variables
sort v001 v002 v003
* Step 7: merge!
merge 1:1 v001 v002 v003 using "TZAR4AFL_mergeprep.DTA"
* Step 8: Complete the merge
drop if _merge!=3
* drop the merge variable
drop _merge
* Step 9: save women and men data with HIV results added
save "TZIR4AFL_merged.DTA", replace
```

```
*Step 10: Split the merged dataset into two datasets, one for men and one for women
* first men
use "TZIR4AFL_merged.DTA", clear
keep if aidsex==1
* rename variables to names for men, and drop a few unnecessary ones
rename v* mv*
rename s* sm*
rename h* mh*
```

```
drop awfact*
* rename back the ID variables used for matching
rename mv001 v001
rename mv002 v002
* create man's line number var for matching
clonevar v034=mv003
* sort on the ID variables
sort v001 v002 v034
save "TZIR4AFL_merged_men.DTA", replace

* second women
use "TZIR4AFL_merged.DTA", clear
keep if aidsex==2
* create husband's line number var for matching
clonevar v034=v034_1
* drop women who are unmarried or whose partner does not live in the household
drop if v034==. | v034==0
* sort and save
sort v001 v002 v034
save "TZIR4AFL_merged_women.DTA", replace

*Step 11: Merge women and men as couples
merge m:1 v001 v002 v034 using "TZIR4AFL_merged_men.DTA"
* keep only the couples who matched
drop if _merge!=3
save "TZIR4AFL_merged_couples.DTA", replace
```
