
Subject: Problem with gender variable merging household recode with women's recode

Posted by tiampalermo@gmail.com on Fri, 19 Jun 2015 07:23:15 GMT

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

I am using the Tanzania AIS 2011-2012 data and trying to merge the household recode (PR) with the women's recode (IR). I created a gender variable based on hv104 in the PR dataset, created cluster, household and indiv (hvidx) variables with consistent names to merge across datasets, and then merged 1:1 on cluster hh id with the women's recode (IR). However, when I tab the _merge variable based on gender, there are n=8352 matches (_merge==3) among those who are supposedly male according to the PR dataset. Why is the gender information in the PR dataset inconsistent with who is interviewed in the women's recode (IR)? I would appreciate any insight on where I'm going wrong.

Thank you. Code is below:

```
. use "$z\TZPR6AFL.DTA", clear
```

```
. *gender
```

```
. tab hv104, missing
```

Sex of household member	Freq.	Percent	Cum.
Male	26,132	48.37	48.37
Female	27,888	51.63	100.00
-----+			
Total	54,020	100.00	

```
. tab hv104, nol
```

Sex of household member	Freq.	Percent	Cum.
1	26,132	48.37	48.37
2	27,888	51.63	100.00
-----+			
Total	54,020	100.00	

```
. gen female=hv104==2
```

```
. gen male=hv104==1
```

```
.
```

```
. tab female
```

female	Freq.	Percent	Cum.
-----+-----			
0	26,132	48.37	48.37
1	27,888	51.63	100.00
-----+-----			
Total	54,020	100.00	

```
.
.
.
. ***generate variables for merging with same names across datasets
. gen cluster=hv001

. gen hh=hv002

. gen id=hvidx

.
. /*****MERGES*****/
.
. *women's questionnaire
. preserve

. use "$z\TZIR6AFL.DTA", clear

. count
19319

. gen cluster=v001

. gen hh=v002

. gen id=v003

. gen female2=1

. sort cluster hh id

. save "$z\women_temp.DTA", replace

. restore

.
. tab female
```

female	Freq.	Percent	Cum.
--------	-------	---------	------

-----+-----			
0	26,132	48.37	48.37
1	27,888	51.63	100.00
-----+-----			
Total	54,020	100.00	

. sort cluster hh id

. merge 1:1 cluster hh id using "\$z\women_temp.DTA"

Result	# of obs.

not matched	34,701
from master	34,701 (_merge==1)
from using	0 (_merge==2)
matched	19,319 (_merge==3)

. tab _m

_merge	Freq.	Percent	Cum.
-----+-----			
master only (1)	34,701	64.24	64.24
matched (3)	19,319	35.76	100.00
-----+-----			
Total	54,020	100.00	

. bys female: tab _m

-> female = 0

_merge	Freq.	Percent	Cum.
-----+-----			
master only (1)	17,780	68.04	68.04
matched (3)	8,352	31.96	100.00
-----+-----			
Total	26,132	100.00	

-> female = 1

_merge	Freq.	Percent	Cum.
-----+-----			
master only (1)	16,921	60.67	60.67

matched (3)	10,967	39.33	100.00
-----+-----			
Total	27,888	100.00	

. bys female2: tab _m

-> female2 = 1

_merge	Freq.	Percent	Cum.
-----+-----			
matched (3)	19,319	100.00	100.00
-----+-----			
Total	19,319	100.00	

-> female2 = .

_merge	Freq.	Percent	Cum.
-----+-----			
master only (1)	34,701	100.00	100.00
-----+-----			
Total	34,701	100.00	

. tab female female2, missing

	female2		
female	1	.	Total
-----+-----			
0	8,352	17,780	26,132
1	10,967	16,921	27,888
-----+-----			
Total	19,319	34,701	54,020

. drop _m

. * n=8352 matched among supposed males---investigate