

---

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 633 times
-