
Subject: Issue merging household and individual (women's) file

Posted by [ec2017](#) on Tue, 28 Nov 2017 00:28:29 GMT

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

I am having some trouble interpreting the results of my merge of the household and individual files for Tanzania 2015-2016. According to the DHS Final Report for Tanzania, there were 12,563 households that were successfully interviewed. At these households, 13,266 women were successfully interviewed. When I merged the household and women's data files, all 13,266 women matched to a household (as expected). However, the merge reported that there were 3,033 households for which no woman was found.

Problems:

- This figure (3,033) does not match the number of households with 0 women present; according to the household file, there should be 2,907 households for which no woman was found.
- When I load my merged file, it shows that there are 16,299 observations instead of the 13,266 women I expected to see. It appears that somehow, 13,266 and 3,033 are getting added together to result in 16,299.
- The mystery 3,033 observations have missing values for almost all variables.

I have two questions. Why are there 3,033 households for which no woman was found instead of 2,907? How should I handle the 3,033 observations?

Here is my Stata code for the merge in case it is helpful:

```
*import household file to rename variables
clear
use "C:\Users\student\Desktop\TZHR7HFL.DTA"

*generate variables for the match in the household file
gen v001=hv001
gen v002=hv002

*sort household file
sort v001
sort v002

*save revised household dataset
save "C:\Users\student\Desktop\TZHR7HFL_sorted.DTA", replace

*open base data file (women's)
clear
use "C:\Users\student\Desktop\TZIR7HFL.DTA"

*sort women's file
sort v001
sort v001
```

```
*save revised women's file
save "C:\Users\student\Desktop\TZIR7HFL_sorted.DTA", replace
```

```
*merge the two files
clear
use "C:\Users\student\Desktop\TZIR7HFL_sorted.DTA"
```

```
*merge many to one using v001 and v002
merge m:1 v001 v002 using "C:\Users\student\Desktop\TZHR7HFL_sorted.DTA"
```

```
*check the merge
tab _merge
```

Thank you very much for any assistance you can provide.

Subject: Re: Issue merging household and individual (women's) file
Posted by [Bridgette-DHS](#) on Mon, 08 Jan 2018 20:50:14 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

In this survey, there were 2907 households with no women in them. However, some eligible women were not actually interviewed. This can happen because of a refusal, or unavailability, or age is later determined to be out of range, etc. I suggest that you do your merge a little differently, merging the IR file with the PR file rather than the HR file. If you run the lines given below, you will find that there were 126 households in which there was one or more eligible women who did not make it into the IR file. $2907+126=3033$; that's where the 3033 comes from. There were 116 households in which there was one eligible women and she dropped out; 7 households in which there were 2 eligible women and both of them dropped out; and 3 households in which there were 3 eligible women and all three dropped out.

```
set more off
set maxvar 10000
use e:\DHS\DHS_data\IR_files\TZIR7HFL.dta, clear
sort v001 v002 v003
save e:\DHS\DHS_data\scratch\TZPRtemp.dta, replace

use e:\DHS\DHS_data\PR_files\TZPR7HFL.dta, clear
gen v001=hv001
gen v002=hv002
gen v003=hvidx
sort v001 v002 v003
merge v001 v002 v003 using e:\DHS\DHS_data\scratch\TZPRtemp.dta

tab hv117 _merge
```

```
gen elig_women_not_interviewed=1 if hv117==1 & _merge==1  
gen elig_women_interviewed=1 if hv117==1 & _merge==3
```

```
collapse (sum) elig*, by(v001 v002)  
tab elig_women_not_interviewed if elig_women_interviewed==0
```

Subject: Re: Issue merging household and individual (women's) file
Posted by [ec2017](#) on Fri, 12 Jan 2018 18:55:26 GMT

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

Thank you for the response. I'll review this and see if I can implement!
