
Subject: Merging data

Posted by [Lukresha](#) on Thu, 07 Jul 2016 10:18:54 GMT

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

I am working with data from KDHS and most of the variables I would like to use are in the PR file and MR file. I have seen that a previous question was already asked on the forum but I would like to know whether there is a way I can merge the data such that I do not lose part of my age(hv105)values.

Using advice from the previous posts, this is how I carried out my merging:

```
use "\\smbhome.uscs.susx.ac.uk\ln201\Desktop\KDHS\KEMR70FL.DTA", clear
rename mv001 hv001
rename mv002 hv002
rename mv003 hvidx
save "\\smbhome.uscs.susx.ac.uk\ln201\Desktop\KDHS\KEMR70FL.DTA", replace
file \\smbhome.uscs.susx.ac.uk\ln201\Desktop\KDHS\KEMR70FL.DTA saved
```

```
use "\\smbhome.uscs.susx.ac.uk\ln201\Desktop\KDHS\KEPR70FL.DTA", clear
sort hv001 hv002 hvidx
```

```
merge hv001 hv002 hvidx using MRtemp.dta
```

```
tab _merge
```

_merge	Freq.	Percent	Cum.
1	141,021	91.67	91.67
3	12,819	8.33	100.00
Total	153,840	100.00	

The output I get on age of household members before using keep if _merge==3 is:

```
sum hv105
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hv105	153,822	22.67255	19.12394	0	95

But after keep if _merge==3 is:

```
sum hv105
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hv105	12,819	30.07645	10.90915	15	54

I am kindly inquiring if there is a way I can merge such that hv105 will have values beginning from 0-15 as that's the desired age group I would like to focus on.

Thank you.

Subject: Re: Merging data

Posted by [Bridgette-DHS](#) on Thu, 07 Jul 2016 12:00:37 GMT

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

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

You are apparently trying to merge the MR and PR files. The MR file for this Kenya survey has 12, 819 cases and the age range is 15-54. No one is younger than 15 in the MR (or IR) files. It's not that you have lost hv105; rather, you have no cases with age<15 in the MR file.

Note that for the matched cases, there can be some discrepancies between age in the PR file (hv105) and age in the MR file (mv012). The age in the MR file comes directly from the man, and is assumed to be more accurate than the age in the PR file, which is provided by the household respondent, who MAY be the same person but may be someone else. (Similarly for the IR file and v012.)

I see the line "save "...KEMR70FL.DTA", replace". You should never replace one of the basic recode files with an altered version but the same name. This can cause you endless trouble.

Subject: Re: Merging data

Posted by [Lukresha](#) on Thu, 07 Jul 2016 12:27:25 GMT

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

Thank you for the response and the correction.

Kindly advise on how I can go about with my analysis as my study is focusing on children age 0-15 years but it seems merging with the MR is not feasible.

I need some variables I need like m714(currently working) m717(occupation-grouped) mv732(employment all year/seasonal) to add to the variables I have from the PR file. This is because I want to use them to construct a variable showing child labor as employment of parents in the household can determine whether a child needs to also work to supplement the income.

Subject: Re: Merging data

Posted by [Bridgette-DHS](#) on Thu, 07 Jul 2016 16:19:19 GMT

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

Another response from Tom Pullum:

The variables you name (m714, m717, mv732) are only available in the MR file, which includes for men 15-54. Similarly, v714, v717, v732 are only available in the IR file, which includes women 15-49. They are not in the PR file and not available for anyone below the age of 15. You can focus on ages 15-17 ("children" are defined by UNICEF to be anyone age 0-17) or on 15-19 (adolescents) or 15-24 (youth). Those are really your only options, so far as I can see.
