
Subject: Merging Child and Household data for a new user of Stata

Posted by [rachelsb98](#) on Tue, 11 Apr 2023 18:59:54 GMT

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

Hello,

I am trying to merge the Child's dataset with the Household dataset. I have tried using this video as a guide: <https://www.youtube.com/watch?v=SJkJmtgaqBc>. I believe this merges on hvidx, hv001, hv002, and hc60. However, the variable "hvidx" is not in my dataset (currently using the Afghanistan 2015 dataset, but will eventually have to merge the data of over 50 countries). It does however include a variable called "hvidx_01" and subsequently hvidx_02, hvidx_03 . . . hvidx_21. How do I circumvent this issue? I am new to Stata and statistical modeling in general and would appreciate any help. Any advice on how to merge the child dataset and the household dataset is well appreciated.

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [Bridgette-DHS](#) on Wed, 12 Apr 2023 11:03:47 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

You should use the PR file rather than the HR file. The HR file has one very wide record for each household. The PR file has one record for each person in the household. In the PR file, the unique identifier for each person is hv001 hv002 hvidx. In the KR file, the identifier for each child is v001 v002 b16. Some children in the KR file have died or are not living with the mother, in which case b16 is NA or 0, and they are not in the PR file. There are also some children in the PR file who are not in the KR file, because their mothers have died or not living with the child. There have been many forum postings on this type of merge.

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [rachelsb98](#) on Wed, 12 Apr 2023 12:10:00 GMT

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

Thank you, I was wondering if you could help me with another issue I'm coming across. I followed the video, but am still receiving an error that variables "do not uniquely identify observations in the master data". I am using the following code

```
use "D:\Raw data sets\Albania\Albania 17-18 HH Raw\ALPR71FL.DTA"  
*Merging HH and CHild dataset using "https://www.youtube.com/watch?v=SJkJmtgaqBc"  
drop if hc60>12  
rename hvidx c_line  
rename hv001 cluster_line  
rename hv002 hh_line
```

```
rename hc60 mom_line
order c_line cluster_line hh_line mom_line
save "D:\Raw data sets\Albania\Albania 17-18 HH Raw\Albania 17-18 Merge.DTA", replace
```

```
use "D:\Raw data sets\Albania\Albania 17-18 Raw\Albania 17-18 Raw.DTA"
drop if b16==0
rename b16 c_line
rename v001 cluster_line
rename v002 hh_line
rename v003 mom_line
order c_line cluster_line hh_line mom_line
```

```
merge 1:1 c_line cluster_line hh_line mom_line using "D:\Raw data sets\Albania\Albania 17-18
HH Raw\Albania 17-18 Merge.DTA"
```

Do you know what could be the issue? Btw i tried looking at other posts, but couldn't find any with this issue. Thank you!

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [rachelsb98](#) on Wed, 12 Apr 2023 12:41:27 GMT

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

To be fair, I did find one post with this issue that recommended just using the "merge" command instead of "merge 1:1". But I still get an error listing "Master data not sorted"?

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [Bridgette-DHS](#) on Wed, 12 Apr 2023 12:43:47 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

I see two possible problems. First, replace "drop if b16==0" with "drop if b16==0 | b16==.". Second, where you have "order", you want "sort". Can you try with those changes?

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [rachelsb98](#) on Wed, 12 Apr 2023 15:28:27 GMT

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

I was able to do it with the ALbania dataset. However, with my Angola dataset, I was able to complete it. However, almost half of the data did not match. Is this okay? I just want to ensure I'm

not doing something wrong.

Here is my code for reference:

```
set more off  
clear
```

```
set maxvar 32000  
set more off  
clear
```

```
use "D:\Raw data sets\Angola\Angola 15-16 HH Raw\AOPR71FL.DTA"  
keep hc60 hvidx hv001 hv002 hv007 hv000 hv246 hv235 hv238a hv014  
drop if hc60>18  
rename hvidx c_line  
rename hv001 cluster_line  
rename hv002 hh_line  
rename hc60 mom_line  
order c_line cluster_line hh_line mom_line  
gsort c_line cluster_line hh_line mom_line  
save "D:\Raw data sets\Angola\Angola 15-16 HH Raw\Angola 15-16 Merge.DTA", replace
```

```
use "D:\Raw data sets\Angola\Angola 15-16 Raw\Angola 15-16 Raw.DTA"  
keep v000 v007 b16 v001 v002 v003 v465 v190 v012 v102 v106 m39a v113 v116 v160 v501  
v157 v158 v159 m70 h43  
drop if b16==0|b16==.  
rename b16 c_line  
rename v001 cluster_line  
rename v002 hh_line  
rename v003 mom_line  
order c_line cluster_line hh_line mom_line  
gsort c_line cluster_line hh_line mom_line
```

```
merge 1:1 c_line cluster_line hh_line mom_line using "D:\Raw data sets\Angola\Angola 15-16 HH  
Raw\Angola 15-16 Merge.DTA"
```

Thank you you've been very helpful

Subject: Re: Merging Child and Household data for a new user of Stata
Posted by [Bridgette-DHS](#) on Wed, 12 Apr 2023 19:06:16 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS staff member, Tom Pullum:

You should not need to include the match of v003 with h60 (mom_line). Within a household, the

match on b16 and hvidx should be sufficient. However, I think your merge is doing what you want.

If you are merging the KR file with the PR file, you can include (when preparing the PR file) a line such as "keep if hc1<." (there's a dot or period after the 1). Only those household members with hc1 less than a dot are age-eligible to be in the KR file.

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [rachelsb98](#) on Wed, 12 Apr 2023 20:08:14 GMT

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

I've been using the same system as above, minus the mom_line variable. But for the Columbia dataset, there are no matches? Any thoughts on why this is happening or any way to remedy this?

Subject: Re: Merging Child and Household data for a new user of Stata

Posted by [Bridgette-DHS](#) on Thu, 13 Apr 2023 11:15:30 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

I don't know for sure which Colombia survey you are using. I will insert below some Stata lines for this merge using the 2015-16 survey. I construct variables "in_PR" and "in_KR", which I sometimes find helpful for diagnosing problems with a merge. This program seems to work fine. It matches 11,231 children in both the PR and KR files.

*KR PR merge using Colombia 2015-16

* Specify workspace

```
cd e:\DHS\DHS_data\scratch
```

* prepare PR file

```
use "...COPR72FL.DTA" , clear
```

```
gen cluster=hv001
```

```
gen hh=hv002
```

```
gen line=hvidx
```

* This survey does not contain hw1. Reduce to children 0-4 years, plus an extra year

* for possible age discrepancies

```
drop if hv105>5
```

```
gen in_PR=1
```

```
sort cluster hh line
```

```
save COPRtemp.dta, replace
```

```
* prepare KR file and merge
use "...COKR72FL.DTA" , clear
drop if b16==0 | b16==.
gen cluster=v001
gen hh=v002
gen line=b16
gen in_KR=1
sort cluster hh line
merge cluster hh line using COPRtemp.dta
tab _merge
tab in*,m
```

Subject: Re: Merging Child and Household data for a new user of Stata
Posted by [rachelsb98](#) on Thu, 13 Apr 2023 11:33:42 GMT

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

Stata won't let me use the "merge" command without specifying. Should I do 1:1? or 1:m?

Subject: Re: Merging Child and Household data for a new user of Stata
Posted by [Bridgette-DHS](#) on Thu, 13 Apr 2023 12:07:12 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

This specification would be 1:1. The child appears once in the KR file and once in the PR file.