
Subject: Merging Children with Household
Posted by [befulina](#) on Thu, 17 Mar 2016 00:34:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,

I'm working with the 2012 Peru (continuous) data set and am having troubles merging the children's set with the household characteristics data. I want to apply the household characteristics to each child (and will use the child as the unit of analysis). This is my first time working with DHS data, and I am a novice Stata user (learning as I go).

Using advice from previous posts, I used V001 V002 V003 and b16 to attempt to uniquely identify the children. However, Stata informs me that "variables v001 v002 v003 hvidx do not uniquely identify observations in the master data". Below is my syntax for merging PR into the KR file.

```
use "C:\Users\Beth\Desktop\Peru Project\Data
Files\PE_2012_ContinuousDHS_10092015_2126_84061\pepr6idt\PEP R6IFL.DTA ", clear
rename hv001 v001
rename hv002 v002
rename hv003 v003
sort v001 v002 v003 hvidx
save "C:\Users\Beth\Desktop\Peru Project\Data
Files\PE_2012_ContinuousDHS_10092015_2126_84061\pepr6idt\PEP R6IFLTEMP1.DTA ",
replace
```

```
use "C:\Users\Beth\Desktop\Peru Project\Data
Files\PE_2012_ContinuousDHS_10092015_2126_84061\pekr6idt\PEK R6IFL.DTA ", clear
rename b16 hvidx
sort v001 v002 v003 hvidx
merge 1:1 v001 v002 v003 hvidx using "C:\Users\Beth\Desktop\Peru Project\Data
Files\PE_2012_ContinuousDHS_10092015_2126_84061\pepr6idt\PEP R6IFLTEMP1.DTA "
```

I appreciate any insight into this!

Subject: Re: Merging Children with Household
Posted by [Bridgette-DHS](#) on Tue, 22 Mar 2016 13:48:38 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

The following should work (I have shortened your filenames):

```
use PEPR6IFL.DTA, clear
rename hv001 v001
rename hv002 v002
rename hvidx b16
```

```
sort v001 v002 b16
save PEPR6IFLTEMP1.DTA, replace
```

```
use PEKR6IFL.DTA, clear
sort v001 v002 b16
merge 1:1 v001 v002 b16 using PEPR6IFLTEMP1.DTA
```

Subject: Re: Merging Children with Household
Posted by [befulina](#) on Tue, 22 Mar 2016 18:52:17 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for the response!
I copied and pasted in your suggested code (with my full file names), and am still receiving the same error:

```
"variables v001 v002 b16 do not uniquely identify observations in the master data"
```

What else can I try? Why would b16 identify multiple children in the household?

Thank you for your help!!

Subject: Re: Merging Children with Household
Posted by [user-rhs](#) on Sun, 27 Mar 2016 02:53:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

When you merge 2 datasets in Stata, Stata identifies one of the datasets as the "master" data and the other dataset as the "using" data. The master dataset is the one in memory, while the using dataset is the dataset that is identified in the -merge- command after 'using.'

If there are multiple observations in PEKR61FL.DTA with the same v001, v002, and b16 combination, you will get an error message after running

```
use PEKR6IFL.DTA, clear
sort v001 v002 b16
merge 1:1 v001 v002 b16 using PEPR6IFLTEMP1.DTA
```

because the merge command specified a 1:1 (one-to-one) merge, meaning: you are telling Stata that in both the master and using datasets, v001, v002, and b16 are unique identifiers. Obviously, the command will not work if v001, v002, and b16 do not uniquely identify observations in either or both datasets.

There was a slight typo/omission in the code Tom sent you below, which led to the error message. The merge using should be v001, v002, v003, and b16 instead of v001, v002, b16, since the child is identified via its cluster number (v001), household number (v002), mother's line number (v003),

and finally its line number in the HH roster (b16). Try it with the correction and I'm sure it will work this time.

HTH,
rhs

Subject: Re: Merging Children with Household
Posted by [befulina](#) on Sun, 27 Mar 2016 17:59:36 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks!

However, even including mother's line number (v003) did not uniquely identify the children in the KR set. Running a quick duplicates list shows that there are several children identified by these sets of supposed unique identifiers (26, to be exact). I'm quite at a loss of how to proceed!

duplicates list v001 v002 v003

group	obs	v001	v002	v003	b16

1	190	21	33	2	Not listed in household
1	191	21	33	2	Not listed in household
2	437	59	40	2	Not listed in household
2	438	59	40	2	Not listed in household
3	910	132	14	7	Not listed in household

3	911	132	14	7	Not listed in household
4	1027	150	23	2	.
4	1028	150	23	2	.
5	1768	256	31	2	3
5	1769	256	31	2	3

6	1876	269	1	2	.
6	1877	269	1	2	.
7	2296	338	44	2	4
7	2297	338	44	2	4
8	2343	345	57	2	4

8	2344	345	57	2	4
9	2793	447	26	4	Not listed in household
9	2794	447	26	4	Not listed in household
10	3179	478	2	2	Not listed in household
10	3180	478	2	2	Not listed in household

11	3664	552	149	2	4

11	3665	552	149	2	4
11	3666	552	149	2	4
12	3859	570	92	2	3
12	3860	570	92	2	3

13	4033	595	44	1	.
13	4034	595	44	1	.
14	4104	604	73	2	Not listed in household
14	4105	604	73	2	Not listed in household
15	4188	613	42	2	.

15	4189	613	42	2	.
16	4804	712	32	2	.
16	4805	712	32	2	.
17	4819	714	44	2	.
17	4820	714	44	2	.

18	5121	752	14	2	4
18	5122	752	14	2	4
18	5123	752	14	2	4
19	5638	841	94	4	.
19	5639	841	94	4	.

20	7541	1156	42	2	4
20	7542	1156	42	2	4
21	7588	1164	60	2	3
21	7589	1164	60	2	3
22	8448	1263	132	2	3

22	8449	1263	132	2	3
23	8506	1268	128	2	3
23	8507	1268	128	2	3
24	8561	1273	149	2	4
24	8562	1273	149	2	4

24	8563	1273	149	2	4
25	8855	1315	174	2	3
25	8856	1315	174	2	3
26	9247	1374	26	2	5
26	9248	1374	26	2	5

Subject: Re: Merging Children with Household
 Posted by [user-rhs](#) on Sun, 27 Mar 2016 18:46:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Befulina,

I stand corrected. You need to manually check the 26 duplicates and determine whether it makes sense to include them in the analysis. I noticed that some of the children don't have a line number in the HH roster, which probably means that they were dead at the time of interview (remember the birth hx questionnaire asks the woman of ALL births regardless of current vital status). I also saw that some children had "not listed in household" as the value for b16, which means that they belong to the woman being interviewed, but they do not live in the household with her. An example would be if the child lived with grandparents or if the woman was divorced/separated from the child's father and the child lives with him instead of her.

Here's a way to manually inspect the duplicates:

```
*Create a variable that flags duplicates using v001, v002, v003, b16  
bysort v001 v002 v003 b16: gen count=_N /*_N is one of two built-in count variables in Stata that  
indicates the highest count within a group, in this case (v001 v002 v003 b16) */
```

```
*See how many are duplicates  
tab count
```

```
*Visually inspect the duplicate records  
sort b5 v001 v002 v003 b16  
browse v001 v002 v003 b16 b5 b6 b8 if count>1
```

After you determine why the children don't have line numbers or are not listed in the HH roster, you can make a decision whether to include or exclude them. Remember that you can merge HH characteristics to the children without having their HH roster line number by merging using the HH unique identifier (v001 v002). If you're interested in all children born to a woman in the past 5 years, then it would be a good idea to include the dead children as well. If a child does not live with the mother, my inclination is that it would not make sense to link the mother's HH characteristics to the child, but again, that is your call to make.

HTH,
rhs

Subject: Re: Merging Children with Household
Posted by [Bridgette-DHS](#) on Mon, 28 Mar 2016 14:06:06 GMT
[View Forum Message](#) <> [Reply to Message](#)

Additional comment from Tom Pullum, in response to your post of 3/22/2016:

The lines I sent you would normally merge the PR and KR files successfully, but I see that they do not work for Peru 2012. There appears to be a sub-household identifier that must be specified in the merge. I am looking into this and will get back to you.

It could be data entry error too. Six duplicate sets were children who were not listed in the HH roster, although 3 pairs were reported as living with the respondent (Table 1). Seven duplicate pairs were of deceased children--those are straightforward (Table 2). The remaining 13 multiples sets (11 duplicate pairs and 2 triplicate sets) seem highly suspicious (Table 3). For example, check out the first triplicate set (552 149 2 4). You have a 3 living children: a female born 3/2008, a male born 1/2008, and another male born 7/2009. All are singlets. It's highly unlikely to have a child born 1/2008 and another child born 2 months later on 2/2008 (again, both singlets), unless the children were adopted, incorrectly classified as single pregnancies instead of twins, or the woman has a rare condition (e.g. double uterus). Notice also that the MIDX was the same for the female born 3/2008 and the male born 7/2009.

Table 1. Children not listed in household (b16=0)
 bysort b5: list v001 v002 v003 b16 b0 b1 b2 b4 b5 b9 midx if count>1 &b16 ==0, sepby(v001 v002 v003) noobs
 -> b5 = No
 -> b5 = Yes

v001	v002	v003	b16	b0	b1	b2	b4	b5	b9	midx
21	33	2	Not listed in household	1st of multiple	7	2007	Female	Yes	Lives elsewhere	3
21	33	2	Not listed in household	2nd of multiple	7	2007	Female	Yes	Lives elsewhere	2
59	40	2	Not listed in household	1st of multiple	3	2012	Male	Yes	Respondent	2
59	40	2	Not listed in household	2nd of multiple	3	2012	Male	Yes	Respondent	1
132	14	7	Not listed in household	Single birth	6	2008	Male	Yes	Respondent	3
132	14	7	Not listed in household	Single birth	1	2010	Female	Yes	Respondent	2
447	26	4	Not listed in household	1st of multiple	5	2010	Male	Yes	Respondent	2
447	26	4	Not listed in household	2nd of multiple	5	2010	Male	Yes	Respondent	1
478	2	2	Not listed in household	Single birth	7	2010	Male	Yes	Lives elsewhere	2
478	2	2	Not listed in household	Single birth	11	2011	Male	Yes	Lives elsewhere	

```

1 |
|-----|
| 604 73 2 Not listed in household Single birth 4 2008 Female Yes Lives elsewhere
2 |
| 604 73 2 Not listed in household Single birth 11 2009 Male Yes Lives elsewhere
1 |
+-----+

```

Table 2. Deceased children

. bysort b5: list v001 v002 v003 b16 b0 b1 b2 b4 b5 b9 midx if count>1 &b16 !=0, sepby(v001 v002 v003) noobs
-> b5 = No

```

+-----+
| v001 v002 v003 b16          b0 b1 b2 b4 b5 b9 midx |
|-----|
| 150 23 2 . 1st of multiple 10 2009 Female No . 2 |
| 150 23 2 . 2nd of multiple 10 2009 Female No . 1 |
|-----|
| 269 1 2 . Single birth 5 2007 Female No . 2 |
| 269 1 2 . Single birth 4 2009 Female No . 1 |
|-----|
| 595 44 1 . 1st of multiple 11 2007 Female No . 2 |
| 595 44 1 . 2nd of multiple 11 2007 Female No . 1 |
|-----|
| 613 42 2 . 1st of multiple 10 2011 Male No . 2 |
| 613 42 2 . 2nd of multiple 10 2011 Male No . 1 |
|-----|
| 712 32 2 . 1st of multiple 11 2007 Male No . 3 |
| 712 32 2 . 2nd of multiple 11 2007 Male No . 2 |
|-----|
| 714 44 2 . Single birth 5 2011 Male No . 1 |
| 714 44 2 . Single birth 12 2009 Female No . 2 |
|-----|
| 841 94 4 . Single birth 8 2011 Male No . 2 |
| 841 94 4 . Single birth 8 2012 Female No . 1 |
+-----+

```

Table 3. Children in household (b16 !=0), with different DOB but same line numbers

-> b5 = Yes

```

+-----+
| v001 v002 v003 b16          b0 b1 b2 b4 b5 b9 midx |
|-----|
| 256 31 2 3 Single birth 11 2009 Male Yes Respondent 1 |
| 256 31 2 3 Single birth 4 2008 Male Yes Respondent 1 |

```

338	44	2	4	Single birth	10	2009	Female	Yes	Respondent	1
338	44	2	4	Single birth	12	2008	Male	Yes	Respondent	1
345	57	2	4	Single birth	5	2010	Male	Yes	Respondent	1
345	57	2	4	Single birth	6	2009	Male	Yes	Respondent	1
552	149	2	4	Single birth	3	2008	Female	Yes	Respondent	1
552	149	2	4	Single birth	1	2008	Male	Yes	Respondent	2
552	149	2	4	Single birth	7	2009	Male	Yes	Respondent	1
570	92	2	3	Single birth	7	2008	Male	Yes	Respondent	2
570	92	2	3	Single birth	11	2009	Male	Yes	Respondent	1
752	14	2	4	Single birth	7	2011	Female	Yes	Respondent	1
752	14	2	4	Single birth	12	2008	Male	Yes	Respondent	1
752	14	2	4	Single birth	8	2009	Male	Yes	Respondent	1
1156	42	2	4	Single birth	11	2011	Male	Yes	Respondent	1
1156	42	2	4	Single birth	12	2010	Male	Yes	Respondent	1
1164	60	2	3	Single birth	3	2011	Male	Yes	Respondent	1
1164	60	2	3	Single birth	12	2011	Female	Yes	Respondent	1
1263	132	2	3	Single birth	1	2012	Female	Yes	Respondent	1
1263	132	2	3	Single birth	11	2008	Female	Yes	Respondent	1
1268	128	2	3	Single birth	6	2009	Female	Yes	Respondent	1
1268	128	2	3	Single birth	2	2010	Female	Yes	Respondent	1
1273	149	2	4	Single birth	2	2010	Male	Yes	Respondent	2
1273	149	2	4	Single birth	6	2008	Male	Yes	Respondent	1
1273	149	2	4	Single birth	9	2009	Female	Yes	Respondent	1
1315	174	2	3	Single birth	1	2010	Female	Yes	Respondent	1
1315	174	2	3	Single birth	7	2008	Female	Yes	Respondent	2
1374	26	2	5	Single birth	5	2010	Female	Yes	Respondent	1
1374	26	2	5	Single birth	10	2009	Female	Yes	Respondent	1

Subject: Re: Merging Children with Household

Posted by [Bridgette-DHS](#) on Wed, 30 Mar 2016 16:26:31 GMT

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

Additional Comments from Tom Pullum:

I have found that in this survey there are sometimes sub-households. The sub-household codes are not given by a variable but must be extracted as the 9th character in hhid and caseid. The following Stata lines will do the merge. At some points I do "tab subhh" and "tab b8" but those lines are only for checking. They are not needed for the merge. Let me know if there are still any problems.

```
use e:\DHS\DHS_data\PR_files\PEPR6IFL.DTA, clear
```

```
gen str2 subhh=substr(hhid,9,1)
destring subhh, replace
tab subhh,m
```

```
sort hv001 hv002 subhh hvidx
```

```
save e:\DHS\DHS_data\scratch\PEPR6IFLTEMP1.DTA, replace
```

```
use e:\DHS\DHS_data\KR_files\PEKR6IFL.DTA, clear
```

```
gen str2 subhh=substr(caseid,9,1)
destring subhh, replace
tab subhh,m
```

```
tab b8
```

```
drop if b16==0 | b16==.
```

```
rename v001 hv001
rename v002 hv002
rename b16 hvidx
```

```
sort hv001 hv002 subhh hvidx
```

```
merge 1:1 hv001 hv002 subhh hvidx using e:\DHS\DHS_data\scratch\PEPR6IFLTEMP1.DTA
```

```
tab b8 _merge,m
```

```
keep if _merge==3
drop _merge
```

```
tab b8
```

Subject: Re: Merging Children with Household
Posted by [user-rhs](#) on Thu, 31 Mar 2016 19:52:36 GMT
[View Forum Message](#) <> [Reply to Message](#)

Tom and Bridgette--Thanks for this follow-up post. I wasn't aware of sub-HH identifiers until now. As they say: you learn something new every day!

Subject: Re: Merging Children with Household
Posted by [Bridgette-DHS](#) on Fri, 29 Apr 2016 12:52:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

You are welcome.
