
Subject: Merging DHS data in Stata

Posted by [DHS user](#) on Wed, 20 Feb 2013 16:45:11 GMT

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

Could you kindly advice me on how to merge the Individual, Male and HIV Recode files in Stata?

Subject: Re: Merging DHS data in Stata

Posted by [Bridgette-DHS](#) on Wed, 20 Feb 2013 16:46:28 GMT

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

Here is a response from one of our STATA experts Tom Pullum, that should answer your question.

Assuming you want a single file that includes men and women, as individuals, and has the HIV data merged onto the individual records. In Stata, I would go through the following steps:

Open the IR file and construct a new variables, sex=2 (female), for all cases. Save as file 1.

Open the MR file and construct a new variable, sex=1 (male), for all cases. For all variables you need that have an mv prefix, drop the m, so the prefix becomes v. Save as file 2.

Open the AR file and change the variable names exactly as you did. Sort on v001 v002 v003. Save as file 3.

Open file 1. Then APPEND file 2 to the end of file 1, getting a file with all the men and women as observations. Sort on v001 v002 v003. Merge with file 3, on v001 v002 v003. Drop any cases with hiv05 missing (these are cases with no HIV result). Save as file 4. This will be your working file.

In file 4, the preferred weight will be hiv05, rather than v005. The cluster variable will be v001 (it is duplicated in the v020's but we always use v001). There may be a variable that is identified as a stratum variable, e.g. v022, but we recommend that you use what is identified as the domain variable, e.g. v023. If domains are not given, you can construct a domain variable for virtually all the surveys as all combinations of region and urban/rural.

I hope this helps.

Bridgette-DHS

Subject: Re: Merging DHS data in Stata

Posted by [kf2349](#) on Tue, 26 Mar 2013 20:20:51 GMT

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

Do you have similar suggestion for merging DHS data in SAS? I would like to have a single data

file that includes men and women as individuals, as well as HIV status.

Karin

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Thu, 28 Mar 2013 13:25:54 GMT
[View Forum Message](#) <> [Reply to Message](#)

The suggestions are the same for merging in SAS. Please follow the same steps in SAS, and you can also reference the page on Merging Datasets, for additional information.

Bridgette-DHS

Subject: Re: Merging DHS data in Stata
Posted by [caragh](#) on Tue, 02 Apr 2013 22:46:19 GMT
[View Forum Message](#) <> [Reply to Message](#)

Can you please advise me if there are similar steps to be followed to merge the HIV test result with the Men's questionnaire only using SPSS? Thank you

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Tue, 09 Apr 2013 16:07:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Below is an example syntax for merging the HIV Test data with the Women's data in SPSS. You can modify this for other datasets.

```
GET FILE='C:\DATAUSER\ZMAR51FL.SAV'.  
SORT CASES BY  
HIVCLUST (A) HIVNUMB (A) HIVLINE (A).  
SAVE OUTFILE='C:\DATAUSER\HIV.sav'  
/COMPRESSED.
```

```
GET FILE='C:\DATAUSER\ZMIR51FL.SAV'.  
SORT CASES BY  
V001 (A) V002 (A) V003 (A) .  
SAVE OUTFILE='C:\DATAUSER\WOMEN.sav'  
/RENAME(V001 V002 V003=  
HIVCLUST HIVNUMB HIVLINE)  
/COMPRESSED.
```

```
GET FILE='C:\DATAUSER\WOMEN.sav'.  
MATCH FILES /FILE=*  
/TABLE='C:\DATAUSER\HIV.sav'
```

```
/BY HIVCLUST HIVNUMB HIVLINE.  
EXECUTE.  
SAVE OUTFILE='C:\DATAUSER\ZMAR_IR.SAV'  
/COMPRESSED.
```

Subject: Re: Merging DHS data in Stata
Posted by [owraza](#) on Sun, 29 Jun 2014 00:38:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

How do I know if I have merged files in STATA correctly or not? I am trying to merge IR file with PR file, I followed following codes:

```
use "C:\Users\Owais\Copy\DHS\DHS13 Datasets\Households\PKPR61FL.DTA", clear  
gen v001=hv001  
gen v002=hv002  
label variable v001 "Cluster number_copy"  
recast long v001  
format %12.0g v001  
label variable v002 "Household number_copy"  
recast int v002  
format %8.0g v002  
sort v001 v002
```

```
use "C:\Users\Owais\Copy\DHS\DHS13 Datasets\Women\PKIR61FL.DTA", clear  
sort v001 v002  
merge m:m v001 v002 using "C:\Users\Owais\Copy\DHS\DHS13  
Datasets\Households\PKPR61FL.DTA"
```

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Mon, 30 Jun 2014 17:20:12 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

You are right to be concerned about whether the merge was correct. It is very easy to make a mistake when merging. The way I would do it is given below. You can change the paths and run it.

During a merge, a variable called "_merge" is constructed. The codes for it are given in "help merge". In your case, you will want to keep the cases with _merge=3. In the lines below I construct "in_PR=1" and "in_IR=1" in the PR and IR files, respectively, and if you run the code you will see that the cases in the merged file with _merge=3 are exactly the same as those with in_PR=1 AND in_IR=1. I sometimes add this extra check to be absolutely sure that I have what I want.

You have to save the sorted PR file. You should save it with another name. Never over-write the basic recode files. I use old syntax for the merge command. I know some other people do too. The current syntax doesn't always mean what you think it means. I dropped the lines to recast v001 and v002. They are not needed. I used rename instead of gen to get v001 and v002 in the PR file. You can do things like that in temporary or scratch files.

Let us know if you have other questions.

```
use c:\DHS\DHS_data\PR_files\PKPR61FL.dta, clear
rename hv001 v001
rename hv002 v002
gen in_PR=1
sort v001 v002
```

```
save c:\DHS\DHS_data\scratch\temp.dta, replace
```

```
use c:\DHS\DHS_data\IR_files\PKIR61FL.dta, clear
gen in_IR=1
sort v001 v002
merge v001 v002 using c:\DHS\DHS_data\scratch\temp.dta
```

```
tab1 _merge in_PR in_IR
```

```
keep if in_PR==1 & in_IR==1
```

```
tab _merge
drop _merge
```

Subject: Re: Merging DHS data in Stata
Posted by [owraza](#) on Wed, 02 Jul 2014 11:24:17 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you very much Pullum & Bridgette. Your reply was helpful.

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Wed, 02 Jul 2014 14:21:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

You are welcome.

Subject: Re: Merging DHS data in Stata

Posted by [Malachi Arunda](#) on Sat, 16 May 2015 11:39:33 GMT

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

Tanzania 2011-12, Zambia 2011-12, Merging HIV data to women file (IR) on SPSS. I hope this "merging" syntax provided works in this case too? Since male and female are combined in IR, I will "merge" then create a new variable for only women. Will that be the right procedure? Thank you

.

Subject: Re: Merging DHS data in Stata

Posted by [Malachi Arunda](#) on Sat, 16 May 2015 11:46:47 GMT

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

Dear DHS expert,

Below is my syntax for merging HIV dataset AR to individual dataset IR using Tanzania 2011-12 dataset on SPSS. Please let know why I do not get the desired results. The combined dataset does not have HIV data for IR, rather the AR (HIV test results) data is scattered within the combined IR/AR dataset and seem to be no connection.

```
DATASET ACTIVATE DataSet4.  
SORT CASES BY HIVCLUST(A) HIVNUMB(A) HIV01(A).
```

```
DATASET ACTIVATE DataSet1.  
SORT CASES BY V001(A) V002(A) V003(A).  
SAVE
```

```
RENAME VARIABLES (V001 V002 V003 =  
HIVCLUST HIVNUMB HIVLINE).  
EXECUTE
```

```
DATASET ACTIVATE DataSet6.  
MATCH FILES /FILE=*  
  /FILE='DataSet4'  
  /BY HIVCLUST HIVNUMB HIVLINE.  
EXECUTE.  
Thank you
```

Subject: Re: Merging DHS data in Stata

Posted by [Bridgette-DHS](#) on Tue, 19 May 2015 15:14:08 GMT

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

In line 2 you are sorting by the wrong variable:

```
SORT CASES BY HIVCLUST(A) HIVNUMB(A) HIV01(A).
```

It should be

```
SORT CASES BY HIVCLUST(A) HIVNUMB(A) HIVLINE(A).
```

Subject: Re: Merging DHS data in SPSS

Posted by [Malachi Arunda](#) on Sun, 24 May 2015 12:29:07 GMT

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

Thank you,

I corrected that error but still the 2 data sets do not merge. Even after sorting cases and renaming variables HIVCLUST HIVNUMB HIVLINE properly. The case IDs still remain different and the data merged cannot link HIV status to any individual data, (SPSS).I will only consider mothers in this study. Could there be solution?Thank you.

Warm regards,
Malachi

Subject: Re: Merging DHS data in Stata

Posted by [Malachi Arunda](#) on Sat, 30 May 2015 17:46:10 GMT

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

Dear DHS experts,

Thank you once again for the guidance, however, I still seem to be having hitches with Tanzania SPSS 2011-12 datasets and the HIV datasets. I have tried to merge the two datasets using the syntax guideline you provided in the forum but they seem to merge with different case IDs, and the HIV dataset merges only as new cases with own variables with. Here is the syntax I used. Could I get further guidance. Thank you.

```
DATASET ACTIVATE DataSet4.  
SORT CASES BY hivCLUST(A) hivNUMB(A) hivLINE(A).
```

```
DATASET ACTIVATE DataSet1.  
SORT CASES BY V001(A) V002(A) V003(A).  
SAVE
```

```
RENAME VARIABLES (V001 V002 V003 =  
hivCLUST hivNUMB hivLINE).  
EXECUTE
```

```
DATASET ACTIVATE DataSet6.  
MATCH FILES /FILE=*  
/FILE='DataSet4'
```

/BY hivCLUST hivNUMB hivLINE.
EXECUTE.
Kind regards,
Malachi

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Mon, 01 Jun 2015 17:00:04 GMT
[View Forum Message](#) <> [Reply to Message](#)

Our SPSS Specialist, Ladys Ortiz, took a closer look at your syntax, and had the following comments:

You are missing a period after the "save" command, and the 1st "execute" command. Also, you are referencing "dataset6" where you are expected to be referencing "dataset1". Following is the syntax that you can use or compare to yours:

```
GET  
FILE='C:\Tanzania2011\TZAR6AFL.SAV'.  
SORT CASES BY hivCLUST(A) hivNUMB(A) hivLINE(A).  
SAVE OUTFILE='C:\Tanzania2011\TZAR6AFL.SAV'.
```

```
GET  
FILE='C:\Tanzania2011\TZIR6AFL.SAV'.  
SORT CASES BY V001(A) V002(A) V003(A).
```

```
RENAME VARIABLES (V001 V002 V003 =  
hivCLUST hivNUMB hivLINE).  
EXECUTE.
```

```
MATCH FILES /FILE=*  
/FILE='C:\Tanzania2011\TZAR6AFL.SAV'  
/BY hivCLUST hivNUMB hivLINE.  
EXECUTE.
```

Subject: Re: Merging DHS data in Stata
Posted by [Malachi Arunda](#) on Wed, 10 Jun 2015 03:21:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you very much,
This time the merging has worked.

Warm regards,

Malachi

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Wed, 10 Jun 2015 14:41:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

You are welcome.

Subject: Re: Merging DHS data in Stata
Posted by [Malachi Arunda](#) on Tue, 04 Aug 2015 21:55:04 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hallo Bridgette,experts,
Merging worked so well and the work is almost complete. However, incase I wanted to add v"age at death" variable from the children data variable to the already merged women and hiv dataset, how would I go about it? (The merged W + HIV datasets) assign NA to the 'age at death' variable) in spss.
Thank you,
Malachi

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Wed, 05 Aug 2015 14:04:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior Data Processing Specialist, Nouredine Abderrahim:

Tanzania 2011-12 is an AIDS Indicators Survey. In this survey we did not collect information about age at death of children, which explains why this variable is set to Not Applicable (NA).

Subject: Re: Merging DHS data in Stata
Posted by [Malachi Arunda](#) on Wed, 05 Aug 2015 17:22:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Alright, thank you.

Kind regards,
Malachi

Subject: Re: Merging DHS data in Stata
Posted by [Malachi Arunda](#) on Sun, 30 Aug 2015 10:21:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Bridgette, experts,
I looked at Tanzania 2003-2004 under-5 mortality frequencies below (34.7%) and I was awed.
Please encourage me that these figures are real or I made a mistake somewhere, perhaps during merging. Thank you. (Of course I restricted some variables)

Children born alive who died
Frequency Percent Valid Percent Cumulative Percent
Valid No 2704 65.3 65.3 65.3
Yes 1434 34.7 34.7 100.0
Total 4138 100.0 100.0

Kind regards,
Malachi

Subject: Re: Merging DHS data in Stata
Posted by [Reduced-For\(u\)m](#) on Sun, 30 Aug 2015 18:36:59 GMT
[View Forum Message](#) <> [Reply to Message](#)

I think your estimates are off by about triple. The DHS final report for Tanzania pegs the U5 mortality rate around 112/1,000, or about 10%.

Also, to just get the mortality rate, you don't need to merge anything, so... Is this just the HIV-positive sub-sample or something (glancing back over the thread)? In that case, given that your period would cover the late 1990's and early 2000's before lots of ART drugs were available, your 30% figure could be about right. But certainly it is too high for the whole sample.

Subject: Re: Merging DHS data in Stata
Posted by [Reduced-For\(u\)m](#) on Sun, 30 Aug 2015 18:37:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Sorry - link to final Tanzania 2004 report:

<http://dhsprogram.com/pubs/pdf/FR173/FR173-TZ04-05.pdf>

Subject: Re: Merging DHS data in Stata
Posted by [Malachi Arunda](#) on Mon, 31 Aug 2015 16:27:23 GMT

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

Thank you. Could I be mixing up reports, I can see the dhs/AIS 2003=4 report (<http://dhsprogram.com/pubs/pdf/AIS1/AIS1.pdf>) and then the 2004-5 report link you sent and I am using the 2003-4 dataset, could you please help clarify which one I could use. Like you say the mortality numbers are too high, I just obtained the raw frequencies, didn't calculate anything and that is what is what I have to use unless you advise me otherwise. And yes I merged the HIV data to the 2003-4 survey, however, I wanted to consider only last born children under-5, this variable is much easily selected in 2011-12 dataset but in 2003-4, I find it a bit difficult to select these cases. Any help?

Thank you very much,
Malachi

Subject: Re: Merging DHS data in Stata
Posted by [Reduced-For\(u\)m](#) on Mon, 31 Aug 2015 19:10:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

Are you using the DHS (2004-5) or the HIV/AIDS Indicator Survey (2003-4)? If the latter, someone else will need to chime in...

Subject: Re: Merging DHS data in Stata
Posted by [Malachi Arunda](#) on Mon, 31 Aug 2015 19:43:54 GMT
[View Forum Message](#) <> [Reply to Message](#)

Yes I am using standard HIV/AIDS Indicator Survey (2003-4). Thank you

Subject: Re: Merging DHS data in Stata
Posted by [PublicHealthMaster](#) on Fri, 11 Sep 2015 12:45:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

I am trying to evaluate the HIV status data in relationship to a few questions about family planning. Since there is no 1:1 identifier between the HIV status respondents and the questionnaire respondents, is the way to approximate that through the sample weights in both data sets when merging? Pardon my beginner question, but what's the term for this kind of analysis/method? I'm using SPSS and just need to be pointed in the right direction and I'll take it from there!

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Wed, 30 Sep 2015 11:31:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Data Processing Specialist, Nouredine Abderrahim:

For respondents in the HIV sample, there is a one to one relationship between the data in the main file and the data in the HIV Status file. Each respondent in the HIV subsample of the main survey has a record in the HIV file.

Subject: Re: Merging DHS data in Stata

Posted by kinsukmanisinha@gmail.com on Wed, 14 Oct 2015 13:38:03 GMT

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

Dear DHS experts,

I am trying to merge the IR and PR file. I have read the comments on this page, about how to use m:m merge (question by owraza and reply by DHS expert). I understand the entire procedure (use hv001 & hv002 from PR and v001 v002 from IR) but towards the beginning of the explanation the DHS expert mention that is is better to perform old merge as the new m:m merge may do something we dont want it to.

I have stata verion 12 and it does not allow me to use old merge, leaving m:m merge as the only option. However, the help file on stata suggest that I may use joinby, can I do this..???

So, if I use joinby

In the PR (household file):

```
rename hv001 v001  
rename hv002 v002  
sort v001 v002
```

In the IR (women file):

```
sort v001 v002
```

I am lost after this point, I cant use m:m merge as that is not advised and I dont know how to proceed further. I would appreciate any help.

Thanks a lot.

Subject: Re: Merging DHS data in Stata

Posted by [Bridgette-DHS](#) on Wed, 21 Oct 2015 17:55:59 GMT

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

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

I can do everything I want with the old version of merge, which did not include 1:1 or 1:m or m:1 or m:m. I prefer simplicity, so I stay with the old version, but it's not better or worse. You can use whatever version of the merge command that you want to.

You should only occasionally need to use joinby. It's a very powerful command but tends to produce files that are much much larger than the input files. It will give you all possible combinations of cases that match on whatever id code you are using. If you have 10 cases in file 1 with id=12345 and 8 cases in file 2 with id=12345, you will get 10*8 joined (or paired) cases with id=12345.

I think what you want to do would be as follows:

```
use PRXXXXFL.dta, clear
rename hv001 v001
rename hv002 v002
rename hvidx v003
sort v001 v002 v003
save temp.dta, replace
```

```
use IRXXXXFL.dta, clear
sort v001 v002 v003
```

```
merge v001 v002 v003 using temp.dta
tab _merge
keep if _merge==3
drop _merge
```

Subject: Re: Merging DHS data in Stata
Posted by [mianrashid](#) on Tue, 10 May 2016 14:01:44 GMT
[View Forum Message](#) <> [Reply to Message](#)

```
I tried to merge the file of PR and IR, by following,
use "C:\Users\Rashid Javed\Desktop\DHS Data
Set\Pakistan\Stata_PK_2012-13_DHS_01052016_926_87403\6_House hold Member
Recode_pkpr61
> dt\PKPR61FL.DTA", clear
```

```
rename hv001 v001

. rename hv002 v002

. gen in_PR=1

. sort v001 v002
```

```
save "C:\Users\Rashid Javed\Desktop\temp.dta
```

```
use "C:\Users\Rashid Javed\Desktop\DHS Data  
Set\Pakistan\Stata_PK_2012-13_DHS_01052016_926_87403\3_Individual Recode_pkir61dt\PKI  
> R61FL.DTA", clear
```

```
gen in_IR=1
```

```
. sort v001 v002
```

```
. merge v001 v002 using C:\Users\Rashid Javed\Desktop\temp.dta
```

After this finally received this error,

(note: you are using old merge syntax; see [D] merge for new syntax)
variables v001 v002 do not uniquely identify observations in the master data
file C:\Users\Rashid.dta not found
r(601);

Subject: Re: Merging DHS data in Stata
Posted by [owraza](#) on Tue, 10 May 2016 19:24:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

```
clear all
```

Renaming Household's unique variable

```
use "C:\Users\...\PKPR61FL.DTA"  
gen v001=hv001  
gen v002=hv002  
label variable v001 "Cluster number_copy"  
recast long v001  
format %12.0g v001  
label variable v002 "Household number_copy"  
recast int v002  
format %8.0g v002
```

```
sort v001 v002  
save "C:\Users\...\PKPR61FL_A.DTA"
```

Merging Women's file with Household's file

```
use "C:\Users\...\PKIR61FL.DTA", clear  
sort v001 v002  
merge m:m v001 v002 using "C:\Users\...\PKPR61FL_A.DTA"
```

* For checking merging has done correctly *

* v138=eligible women in women's questionnaire & hv010=eligible women in household's questionnaire*

tab v138 hv010

Subject: Re: Merging DHS data in Stata

Posted by [mianrashid](#) on Tue, 10 May 2016 21:04:14 GMT

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

Thank you.

How can I merge the IR, KR and MR files of Pakistan DHS data set?

Subject: Re: Merging DHS data in Stata

Posted by [owraza](#) on Wed, 11 May 2016 00:33:45 GMT

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

Since I haven't used MR file, so advising you anything might be risky. I am sure you can get help from here: <https://www.dhsprogram.com/data/Merging-Datasets.cfm>

Subject: Re: Merging DHS data in Stata

Posted by [Bridgette-DHS](#) on Thu, 12 May 2016 13:08:23 GMT

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

An additional response from Tom Pullum to the post: Quote:I tried to merge the file of PR and IR, by following.....

The problem is that you have not used the line numbers. The following should work. You do not really need in_IR and in_PR, but since you constructed them, I am keeping them.

```
use PKPR61FL.DTA, clear
rename hv001 v001
rename hv002 v002
rename hvidx v003
gen in_PR=1
sort v001 v002 v003
save temp.dta
use PKIR61FL.DTA, clear
gen in_IR=1
sort v001 v002 v003
merge v001 v002 v003 using temp.dta
```

```
keep if in_PR==1 & in_IR==1
drop _merge in_PR in_IR
```

Subject: Re: Merging DHS data in Stata
Posted by [mianrashid](#) on Tue, 17 May 2016 17:01:22 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,
Thanks you.

I merged the men, women, and children level data (IR, MR, KR) as a household unit. Now i want to declare survey for data set, Please let me know command is correct?

```
gen wgt=hv005/1000000
(880 missing values generated)
```

```
. svyset [pw=wgt], psu(hv021) strata(hv023)
??
```

Many Thanks

Rashid

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Thu, 19 May 2016 14:14:39 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

This should work. You will get the same results with `wgt=hv005/1000000` or with `wgt=hv005`. There is no need to divide by 1000000 because Stata automatically re-normalizes pweights to have a mean of 1. For the psu you can use hv021 or hv001. They are identical.

There is typically a small difference between the weight for the same woman in the PR file and in the IR files. That is, her value of hv005 and v005 can be a little different. Similarly for a child. Similarly, hv005 and mv005 for the same man may differ. This is because of adjustments for nonresponse in the IR and MR files. However, given how you have constructed your file, I think it will be ok to use hv005 as the weight for the household units.

Subject: Re: Merging DHS data in Stata
Posted by [phres110](#) on Wed, 02 Nov 2016 17:29:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Bridgette-DHS

kindly help me...i want to merge WI (pkwi21) file with IR (pkir21) file in PDHS 1990 with SPSS. I tried many time by using the same commands that are already discussed here but failed. Kindly help me where i am wrong.

GET

```
FILE='E:\ALL MY PAPER CS\1.DATA SETS\ALL IR SPSS DATA\pkwi21sv\PKWI21FL.SAV'.  
SORT CASES BY whhid withindf wlthind5.  
SAVE OUTFILE='E:\ALL MY PAPER CS\1.DATA SETS\ALL IR SPSS DATA\PKWI21FL.SAV'.
```

GET

```
FILE='E:\ALL MY PAPER CS\1.DATA SETS\ALL IR SPSS DATA\pkir21sv\PKIR21FL.SAV'.  
SORT CASES BY V001 V002 V003.
```

```
RENAME VARIABLES (V001 V002 V003 =  
whhid withindf wlthind5).  
EXECUTE.
```

MATCH FILES /FILE=*

```
/FILE='E:\ALL MY PAPER CS\1.DATA SETS\ALL IR SPSS DATA\pkwi21sv\PKWI21FL.SAV'.  
/BY whhid withindf wlthind5.  
EXECUTE.
```

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Fri, 11 Nov 2016 15:27:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Quote:You are using SPSS and I only work with Stata, so I cannot give you SPSS code. In Stata correct steps would be as follows. You want to merge on the household id code, which is whhid in the WI file and is the first 12 characters of caseid in the IR file. I hope than in SPSS you know a way to extract the first 12 characters of caseid.

```
use e:\DHS\DHS_data\WR_files\PKWR21FL.dta, clear  
sort whhid  
save e:\DHS\DHS_data\scratch\PKWRtemp.dta, replace
```

```
use e:\DHS\DHS_data\IR_files\PKIR21FL.dta, clear  
gen whhid=substr(caseid,1,12)  
sort whhid
```

```
merge whhid using e:\DHS\DHS_data\scratch\PKWRtemp.dta
tab _merge
keep if _merge==3
drop _merge
```

Subject: Re: Merging DHS data in Stata
Posted by [jack.murphy](#) on Tue, 15 Nov 2016 16:26:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for these instructions.

Regarding the domain variable, I have merged the Burkina Faso datasets (BFIR62, BFMR62, BFAR62) and have found that v023 is described as the "Stratification used in sample design." Is this still the domain variable I should use in svyset? If not, could you please give me steps on how to generate a domain variable?

At the moment, this is my svyset code:

```
svyst [pw=hiv05], psu(v001) strata(v023)
```

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Wed, 16 Nov 2016 14:32:43 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Quote:In the past, the terminology (domains vs. strata) has not always been consistent. Please interpret v023 as the stratification variable. That is what you should use in svyset, as in "svyset....., strata(v023)".

Sometimes v023 is empty. We are slowly producing a complete list of the stratification variable for all the surveys. In almost all cases, v023 (and/or the correct stratification variable) are the combinations of region (v024) and place of residence (v025). You can construct that with "egen stratumid=group(v024 v025)" and then "svyset....., strata(stratumid)". Actually I recommend "svyset....., strata(stratumid) singleunit(centered)", or another singleunit option, to avoid an error message.....

Subject: Re: Merging DHS data in Stata
Posted by [jack.murphy](#) on Thu, 17 Nov 2016 21:53:24 GMT
[View Forum Message](#) <> [Reply to Message](#)

Brilliant, this worked! Thank you for your help.

I have another question regarding using svyset for master dataset created by appending multiple countries together. I generated a variable for each country (called "country_ID" with strings for the names of each country included). Then I used the dropdown menus to make a 2-stage svyset codeline with country_ID as the higher level strata variable and stratumid (from your earlier post) as the lower level strata variable. Here is my code:

```
svyset v001 [pweight=hiv05], strata(country_ID) vce(linearized) singleunit(centered) || v001, strata(stratumid)
```

Is this accurate?

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Fri, 18 Nov 2016 15:31:12 GMT
[View Forum Message](#) <> [Reply to Message](#)

Another response from Tom Pullum:

Quote:

Hi Jack--The only change I would make to that line would be with v001. That's the cluster (PSU) id code and it is typically numbered 1, 2, 3, etc., within each survey. If you don't renumber the clusters in the combined file, Stata will think that cluster 1 in each survey is the same cluster, etc. You need to renumber the clusters in the same way that you renumbered the strata (in Stata the command would be "egen clusterid=group(survey v001)").

For a very few surveys (only the Egypt surveys, so far as I know) the PSU variable is v021 rather than v001. Usually v001 and v021 are exactly the same. If they are not, you would use v021 in svyset.

Subject: Re: Merging DHS data in Stata
Posted by [jack.murphy](#) on Fri, 18 Nov 2016 16:28:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks very much!

Subject: Re: Merging DHS data in Stata
Posted by [jack.murphy](#) on Fri, 18 Nov 2016 16:38:43 GMT
[View Forum Message](#) <> [Reply to Message](#)

Oh I forgot to ask: in the line of code you wrote,

egen clusterid=group(survey v001)

By "survey" did you mean the v000 variable that identifies the country and phase of the survey? Or would you recommend that I create a new variable called "survey" with a unique number for each country, the same way we created "sex" in the men's and women's recode files before appending them?

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Fri, 18 Nov 2016 19:36:59 GMT
[View Forum Message](#) <> [Reply to Message](#)

Another response from Tom Pullum:

Quote:Sorry about that. Yes, you need an identifier for each survey. You cannot use v000 for that purpose, because it can happen that DHS does two surveys in the same country with the same version of the core questionnaire. (The third character of v000 is often taken to be the phase of DHS--that's what I thought until relatively recently--but it actually identifies the version of the questionnaire.) You can find many instances of two successive surveys in the same country having the same value of v000. It is best to construct a unique identifier for each survey, as you describe. That's what I meant by "survey" in that line.

Subject: Re: Merging DHS data in Stata
Posted by [jack.murphy](#) on Fri, 18 Nov 2016 19:37:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

Excellent, thank you.

Subject: Re: Merging DHS data in Stata
Posted by [Hassen](#) on Sun, 06 May 2018 09:58:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear all Thank you,This posts are very helpful!!

Subject: Re: Merging DHS data in Stata
Posted by [pie](#) on Tue, 02 Jun 2020 10:21:22 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi all,

I have a similar question. I am using Cambodia DHS datasets 2014. I am seeking help.

I would like to test differences between sexually active men and women with regard to different

variables (age, education, residence, wealth, occupation, PTMCT knowledge, HIV stigma...etc.). I will use basic inferential statistical tests (chi-square, fisher, t-test or wilcoxon) to detect the differences between the two groups.

The IR dataset (women) contains: v005 (sample weight), v021 (sampling unit) and v022 (strata). The MR (men) contains: mv005, mv021 and mv022. These variables are needed to declare survey design in STATA when doing separate men and women analyses.

After appending/merging, what is the appropriate weight approach I should use to achieve my goal mentioned above? Please advise!

Thanks
Pie

Subject: Re: Merging DHS data in Stata
Posted by [Bridgette-DHS](#) on Fri, 05 Jun 2020 19:32:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

After any merging, if you have multiple versions of the *v021 or *v022 variables, it makes no difference which one you use. But if you have multiple versions of the *v005 variables, it does make a slight difference which one you use. The priority is like this: v005 has priority over hv005 for women and children; mv005 has priority over hv005 for men; mv005 has priority over v005 for couples; and d005 has priority over v005 when using the domestic violence module for women. I don't believe this survey included HIV testing, but for surveys that do, when using HIV test results, hiv05 has priority over any other version of the weight. These rules are based on the general pattern of nonresponse. Each weight includes an adjustment for nonresponse.
