Subject: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Mon, 01 Sep 2014 10:56:09 GMT
View Forum Message <> Reply to Message

Good afternoon,

I have looked through the other discussions on merging but am still not confident enough to go ahead without checking that I have merged my files correctly.

I am using SPSS and working on the Tanzania 2011 AIS/MIS files. I am using point and click in SPSS as I normally use STATA so don't have experience of using syntax in SPSS. I would prefer to stick with the point and click for this if possible.

I have opened the HH member recode file and the Child recode files. In the HH member recode I have renamed the variables HV001, HV002 and HV003 by dropping the h on each. Now both files have v001, v002 and v003 variables.
I have sorted each dataset by these three variables (ascending sort) and saved two new files called 'hh member sort' and 'child sort'.
I have then merged these files using the child sort as the active file and choosing the hh member sort as the one to be merged to the active one.

My lack of confidence in what I have done is down to not being sure whether the v001, v002 and v003 will sufficiently identify the correct child. I have done the merge with these three as I am following the table in your guidance notes on merging datasets which suggest merging child and hh member files with v001 and v002 (I have added v003 to be sure that I am merging by the same respondent). However, in the child and hh member recodes v001, v002 and v003 relate to the respondent. If the respondent has more than one child then won't there be duplicates of these which actually relate to different children? Do I in fact need to merge by a variable which is a unique identifier for the child? I can't see an obvious variable to use for this. B16 in the child recode is described as the 'child's line number in the household' but b16 is not replicated in the hh member recode as far as I can see.

Do you have any advice?

With thanks!
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by Bridgette-DHS on Tue, 02 Sep 2014 15:35:14 GMT
View Forum Message <> Reply to Message

HV003 is the respondent to the household questionnaire, so merging the data with HV003 won't give you what you are expecting. You need to use HVIDX in the member's file to match women with their children. Please reference the table on user forum msg_2365.

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by Trevor-DHS on Sat, 06 Sep 2014 20:55:04 GMT

View Forum Message <> Reply to Message

Kate, it is a bad practice to use point and click commands in SPSS or Stata or any other software. One of the basic tenets of science is that results should be reproducible. When using point and click you are not able to reproduce your results and you have no documentation of your analysis. We recommend that you always use syntax files to document your analysis and allow it to be reproduced.

Also, if you are used to using Stata, why are you working in SPSS? All DHS datasets are available for both SPSS and Stata.

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Mon, 08 Sep 2014 09:58:42 GMT

View Forum Message <> Reply to Message

Thanks for the advice. I was using SPSS because its provided free my by institution and my stata licence has expired. However the point is well taken and I've decided just to buy a new licence and stick with stata.
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Mon, 08 Sep 2014 19:06:18 GMT

View Forum Message <> Reply to Message

Hello again Bridgette.

I hope you don't mind if I ask you another question as I move forward with trying to do this. I think the problem is I don't understand fully how the datasets are structured.

In the child recode as I understand it, each case is a child. So v001, v002 and v003 themselves don't identify each unique child - for example in cluster 1, hh number 32 (v001 = 1, v002 = 32) there are three cases all with v003 = '3', so the same woman with respondent line number '3' has her information on three different children broken down into three different cases, 1 per child. So far so good.

In the household member recode as I understand it, each case is a household member. So here we can see one case for v001 = 1, v002 = 32 and HDIX = 3. Presumably, then, all the information this woman has given about her three children is included in this one case. I don't see though how this is done in the variables. As far as I can see there are no variables with $ that indicate variables are repeated for each child, for example.

Most of the data I need are in the child recode and I will need to analyse by child. But I need to somehow merge into the child recode, the information on child's parasitaemia, whether they slept under a create net the previous night, and whether their household was sprayed with IRS in the past 12 months. Those pieces of information are in the hh member recode. At the moment I am struggling to see how a merge on the variables v001, v002 and v003/HDIX will give me a dataset in the format I need, i.e. with each case being an individual child. I am not sure how the merge I am trying to do will add the variables that I need against each child case in the child recode.

In any case when I try to merge as I think I am supposed to based on the table your referenced (matching v001 with hv001; v002 with hv002 and v003 with HDIX, with the appropriate renaming and sorting done first) I get the warning message that these my merge by variables don't uniquely identify cases. Which I can see when I look at the data itself but can't work out an alternative.

Sorry for additional questions and I will be much grateful for your guideance.

Best wishes
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by Trevor-DHS on Mon, 08 Sep 2014 19:26:13 GMT
View Forum Message <> Reply to Message

Hi Kate,

The matching variables you need to use are as follows:

v001 & hv001 - cluster number
v002 & hv002 - household number
b16 & hvidx - line number of child in household

Note that there are children listed in the KR file who do not have data in the PR file. This is because some children of the respondent do not live in the household, but live elsewhere.

[Note that there is no HDIX variable. If you meant HIDX then you are looking at the wrong variable. You need HVIDX from the PR file.]

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Mon, 08 Sep 2014 20:44:32 GMT
View Forum Message <> Reply to Message

Thank you very much for this.

I did the merge in SPSS (Still using this until end of the week) and indicated the following:

that I wanted to merge PR to KR rather than the other way round.
that I wanted to add variables rather than cases
that the key variables to match by were v001, v002 and idmerge (my new variable which represents b16 and hvidx in the relevant datasets)
I then also indicated that the the PR file was a keyed table, in order to -- I think --- add the relevant variables from the PR dataset to the existing cases in the KR dataset. The first time I tried the merge by selecting instead that 'both files provide cases; I had very many new cases (about 46,000) added.

Could you confirm that it is correct to specify the PR file as a 'keyed table' and that I haven't caused problems by doing this.

Thank you again for all your help
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by Trevor-DHS on Mon, 08 Sep 2014 20:49:00 GMT
View Forum Message <> Reply to Message

That's correct. In SPSS, you want to use the PR file as a keyed table. If you say that both files provide cases, you get a file with all cases from the PR file and all cases from the KR file. You only want to the cases from the KR file, with variables added from the PR file.

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Mon, 08 Sep 2014 21:02:07 GMT
View Forum Message <> Reply to Message

Thank you. Checking through it all I think I now have my correct merged dataset. Thanks a lot for the guidance.

Best wishes and have a good evening.
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Fri, 12 Sep 2014 10:11:49 GMT

Hello again,

Could you advise me on why I might have lost cases during my merge of datasets.

I merged the household member recode (PR) to the child recode (KR) using the chid recode as the active dataset and the HH member recode as the file to merge onto it.
I matched these as we discussed below.

You mentioned that the PR dataset may well include data for children who not appear in the KR dataset (as in the PR data on all children in the household were collected through the household questionnaire, whereas the KR dataset comprise data on children of the eligible women responding to the individual questionnaire). I assume that the way I have done the merge means that I will retain all cases from the KR dataset and merge additional variables available for these children from the PR dataset. I specified that cases should not be added and therefore any information on children which are included in the PR but not the KR will not be in my final dataset. I don't appear to have additional cases so I think I have achieved what I wanted.

However the original KR file has 8648 cases whereas my merged file has 8573 cases. What could be the reason for cases being lost from the file?

Thank you for your help
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Fri, 03 Oct 2014 08:04:11 GMT

Hello again,

Just checking if you could help me with the below.

Thanks a lot
Kate

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by Trevor-DHS on Fri, 03 Oct 2014 14:18:48 GMT

Hi Kate,

You should still have 8648 cases. I tried reproducing what you have done and I'm still left with 8648 cases. If you can send your syntax file I can try troubleshooting the problem.

Regards. Trevor

Subject: Re: Merging child's malaria results from HH member recode dataset with other child data in Child recode
Posted by kate2 on Thu, 09 Oct 2014 19:55:26 GMT
View Forum Message <> Reply to Message

Hi Trevor,

Thanks again for offering to help.

I have moved over to Stata.

The below is cut and paste from my log file. This has helped me understand the difference in the case numbers. From the below I can see that there are 8648 cases in the KR file and that 7914 of these are matched in the HH member file, the difference is those in the KR file that are not matched in the HH file. So I can see now that its my choice of whether to retain all 8648 or only those 7914 that are matched. I think for the merge I did in SPSS previously I must have used an command which automatically only retained those cases that were matched.

Anyhow - I now am clear on this...and just need to work out what I want to keep!

Thanks a lot
Kate

```
----------------------------------------------------------- -----------------------------------------------------------
-----------------------
     name:  <unnamed>
      log:   /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/Merging data.smcl
 log type:  smcl
 opened on:   3 Oct 2014, 10:19:05

. use " /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZPR6AF L.DTA "

. rename hv001 v001

. rename hv002 v002

. rename hvidx mergeid

. sort v001 v002 mergeid

. gen in_PR=1

. save " /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZPR6AF
L_renamed_sorted.dta "
```

file /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZPR6AF L_renamed_sorted.dta
saved

. clear

. use " /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZKR6AF L.DTA "

. rename b16 mergeid

. sort v001 v002 mergeid

. gen in_KR=1

. save " /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZKR6AF
L_renamed_sorted.dta "
file /Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZKR6AF L_renamed_sorted.dta
saved

. merge m:1 v001 v002 mergeid using "
/Users/admin/Dropbox/MPH/Dissertation/Datasets/Stata/TZPR6AF L_renamed_sorted.dta "

```
    Result                  # of obs.
   -----------------------------------------
   not matched                 46,840
       from master                734  (_merge==1)
       from using              46,106  (_merge==2)

   matched                      7,914  (_merge==3)
   -----------------------------------------
```

. tab1 _merge in_KR in_PR

-> tabulation of _merge

```
          _merge |     Freq.     Percent      Cum.
-----------------------+-----------------------------------
   master only (1) |      734        1.34       1.34
    using only (2) |   46,106       84.21      85.55
     matched (3) |    7,914       14.45     100.00
-----------------------+-----------------------------------
         Total |   54,754      100.00
```

-> tabulation of in_KR

```
    in_KR |     Freq.     Percent      Cum.
------------+-----------------------------------
        1 |    8,648      100.00     100.00
------------+-----------------------------------
```

```
       Total |     8,648     100.00
```

-> tabulation of in_PR

```
   in_PR |    Freq.    Percent      Cum.
------------+-----------------------------------
       1 |   54,020     100.00     100.00
------------+-----------------------------------
   Total |   54,020     100.00
```