Subject: Sample weights disappear after merge household with childs recode
Posted by RKelders on Wed, 29 Jun 2016 15:13:50 GMT
View Forum Message <> Reply to Message

Good afternoon,

After merging the files from household recode and children recode, some of the sampling weights for the household data disappear. I am doing analyses for which I need both sampling weights as sometimes the analyses are on the level of the individual child and other times these are on the level of the household. However, it seems that for a lot of cases the household sampling weight has disappeared after merging these files. It seems to be the case for households with more children where one child gets the household sampling weight in the merged file and the other children from the same household get a missing value.

Could you please help me understand what is happening here and how I can fix this problem?

Thank you very much in advance.

Kind regards,
Rachel Kelders

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Wed, 29 Jun 2016 17:25:34 GMT
View Forum Message <> Reply to Message

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

Are you doing a merge of the KR and PR files? The KR file includes children who have died (b5=0) and other children who have survived but are not living in the same household as the mother (b16 is "0" or "."). These cases would not match with the PR file and would not have a value of hv005. To be in both files, a child would have _merge=3.

Taking those possibilities into account, do you still have cases that are missing hv005?

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Fri, 01 Jul 2016 12:56:40 GMT
View Forum Message <> Reply to Message

Dear Tom,

Thank you very much for your swift reply. Yes, I am doing a merge of the PR and KR files. I just looked at B5 and B16. Looking at B5, there are 218 children not alive and looking at B16, there

are 25 missing values + the 218 children that have died.
However, when I merge these files I am looking at 1329 missing values for HV005. So B5 and B16 do not explain the total number of missing HV005 cases. Is there anything else I could check for now? It looks like it only merges the HV005 data for one child of the same household. For example, I have checked some cases and they all are alive and they live in the household of the respondent, but, these are all cases of which another child comes from the same household (equal cluster and household number). So, the first child has both the sample weight for children and household, but the second child from the same household only has a sampling weight for children, and nothing for household.

Any thoughts?

Thanks again and kind regards,
Rachel Kelders

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Fri, 01 Jul 2016 16:25:51 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

I'm not sure what survey you are working with. I arbitrarily picked the 2013 survey of Nigeria. I will paste below the Stata lines to do what I think you are trying to do. There are no missing values of weights for the children who are in both surveys. Please tell me if I have misunderstood what you want to do.

```
set more off
use e:\DHS\DHS_data\KR_files\NGKR6AFL.dta, clear
keep v001 v002 v005 b5 b16
rename v001 hv001
rename v002 hv002
rename b16 hvidx
sort hv001 hv002 hvidx

save e:\DHS\scratch\NGKR6Atemp.dta, replace

use e:\DHS\DHS_data\PR_files\NGPR6AFL.dta, clear
keep hv001 hv002 hvidx hv005
sort hv001 hv002 hvidx
merge hv001 hv002 hvidx using e:\DHS\scratch\NGKR6Atemp.dta

codebook *v005 if _merge==3

* no missing values of v005 or hv005 for the children who are both files
```

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Sat, 02 Jul 2016 14:36:42 GMT
View Forum Message <> Reply to Message

Dear Tom,

I am using the Bangladesh 2007 DHS KR and PR recode files in SPSS. As I am looking at data on the household level as well as on the children's level, I need both sample weights in the merged file. I will use the sample weights from the KR file when I am analyzing data on the children's level and I will use the household sample weights from the PR file when I am working with household level data. I also need to know from which households the children's are to check for certain correlations, hence the need for a merged file where children are linked to their household's data.

I merged the files on V001 and V002 like you showed in your example but I used the KR file as the basic file, so I merged the PR data into the KR data file (because there can be more children in one household). I do get the error 5132 below after the merge. I tried deleting the cases from the PR file without children before the merge to see if that helped, but it did not. Perhaps the missing values for HV005 after the merge have something to do with this error?

```
GET
  FILE='C:\Users\Rachel\Dropbox\WASH\Bangladesh\SPSS niet-werkbestanden\2007\Childs
recode 2007 - Bangladesh.sav'.
SORT CASES BY V001(A) V002(A).
DATASET ACTIVATE DataSet1.
MATCH FILES /FILE=*
  /FILE='C:\Users\Rachel\Documents\HH for merge 2007.sav'
  /BY V001 V002.
EXECUTE.
File #1
     KEY:       1       11

>Warning # 5132
>Duplicate key in a file.  The BY variables do not uniquely identify each case
>on the indicated file.  Please check the results carefully.

FREQUENCIES VARIABLES=HV005
  /ORDER=ANALYSIS.

 Statistics
Sample weight
N Valid 5183
 Missing 1921
```

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Tue, 05 Jul 2016 17:12:26 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

I don't use SPSS--just Stata. If hv005 is missing for some children in the merged file, they will be children who are in the KR file but not in the PR file. They will be children who died OR children who do not live in the household, even though their mother is in the household. The KR file includes children born in the past five years to the women who are in the IR file. Some of those children live elsewhere.

I would not expect you to lose hv005 just because of the sequence of the two files in the merge. Maybe that can happen in SPSS. I don't think it would happen in Stata. However, my normal practice is to start with the larger file and then merge the smaller file with it. Thus, in this case I would start with the PR file and then merge with the KR file, not the reverse.

You should get unique identifiers if you match hv001 hv002 hvidx in the PR file with v001 v002 b16 in the KR file.

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Wed, 06 Jul 2016 11:03:18 GMT
View Forum Message <> Reply to Message

Thank you for your response!

I now started with the PR file and then merged it with the KR file. This seems to have worked! I now have 949 missing values for the sampling weights but these are all children that also have missing values for B5 or are not alive.

However, I do have one more question. How do I use the B16/HVIDX variable as a unique key identifier? I can change the name of the B16 variable in the KR file to HVIDX, but then it does not match with the HVIDX of the PR file as these are coded as HVIDX$02, HVIDX$03, HVIDX$04 etc. I tried naming them all HVIDX but this does not work.

Thanks again,
Rachel

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Thu, 07 Jul 2016 14:09:44 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

I'm glad the latest attempt to do that merge worked!

You should only get subscripts for hvidx if you are using the HR file (they appear in Stata as _01, etc., and in SPSS as $01, etc.). The HR file has one record per household and the subscripts identify the line number of the person in the household. (That means that hvidx_01=1, hvidx_02=2, etc.) Most users hardly ever need the HR file. I don't understand how you can be getting subscripts for hvidx in the PR file. Are you getting subscripts for other variables too? This is mysterious.

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Sun, 10 Jul 2016 13:04:39 GMT
View Forum Message <> Reply to Message

Hello again,

Apologies,I made a mistake when I confirmed that I am using the PR file because I am not. I am using the HR file, hence I am getting the subscripts. However, I think I do need to use the household file because I will be needing data on the level of the household.

I resolved the error 5132 in SPSS by using another syntax, and this worked. Now I can merge the HR file with the KR file and this seems to have worked!

No further questions at this point..

Thanks for your help!

Regards,
Rachel

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Sun, 17 Jul 2016 14:09:38 GMT
View Forum Message <> Reply to Message

Dear Tom,

Now that I am working with the children's file (KR file), I am using the sample weights for

calculating percentages on stunting, wasting and underweight of the children in Bangladesh. I computed a new variable to be able to work with the sample weights: V005/1000000.

For the years 2004, 2007 and 2014 Bangladesh I had no problems, but for Bangladesh 2011 I am getting an error saying that not all cases have sample weights or these are zero or negative. So I ran a frequency table to check this but there seems to be a positive sample weight for each and every case. I tried this again and again but I am still getting the same error (see below) and I do not understand why that is. Could you please advise what to do?

>Warning # 3211
>On at least one case, the value of the weight variable was zero, negative, or
>missing. Such cases are invisible to statistical procedures and graphs which
>need positively weighted cases, but remain on the file and are processed by
>non-statistical facilities such as LIST and SAVE.

I hope you can help! Thanks again,
Rachel

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Mon, 18 Jul 2016 15:36:57 GMT
View Forum Message <> Reply to Message

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

You did not tell me the name of your new weight variable, but let's say it is "wt". All I can suggest is that after you get that error message, you have a line such as this: "list v001 v002 v003 v005 wt if wt<=0 | wt==., table clean". If you are lucky, this listing will lead you to the problem. Please let me know if this does or does not get you to an answer.

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Tue, 19 Jul 2016 13:34:48 GMT
View Forum Message <> Reply to Message

The name of the new variable is rweightCH. I am not getting the line you prescribed. Just this error message:

>Warning # 3211
>On at least one case, the value of the weight variable was zero, negative, or
>missing. Such cases are invisible to statistical procedures and graphs which
>need positively weighted cases, but remain on the file and are processed by
>non-statistical facilities such as LIST and SAVE.

When I run a frequency table I get the numbers below. As you can see, there are no missings.

Statistics
rweightCH
N Valid 8395
 Missing 0

Anything else you can think of that might cause this error?

Thanks!

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Mon, 10 Oct 2016 12:18:26 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

Quote:I think the warning is related to the fact that the weight variable is hv005 in the household file and v005 in the KR file. The KR file includes some children who are not in the PR file (because they have died or are not living with their mother (and for them hv005 will be missing in the merge). The PR file includes some children who are not in the KR file (because their mother is not in the household and they are missing in the merge). I am pretty sure that the error message simply reflects that. You could check which values are missing with something like "summarize hv005 if hv005<=0 | hv005==.".

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Wed, 12 Oct 2016 16:08:20 GMT
View Forum Message <> Reply to Message

Dear Tom,

I am not working with the merged file at the moment, just in the KR file. I ran a frequency table on the V005 and there were no missing values. All values were above 0. Would you mind trying this in the STATA programme to see if you get the same error? I am using the Bangladesh, 2011, KR file. Below is my syntax:

```
COMPUTE rweightCH=V005/1000000.
EXECUTE.

FREQUENCIES VARIABLES=rweightCH
  /ORDER=ANALYSIS.

WEIGHT by rweightCH
```

And when I want to analyse my data using the weights, I get the following error:

>Warning # 3211
>On at least one case, the value of the weight variable was zero, negative, or
>missing. Such cases are invisible to statistical procedures and graphs which
>need positively weighted cases, but remain on the file and are processed by
>non-statistical facilities such as LIST and SAVE.

I hope you can help!

Thank you in advance,
Rachel

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by Bridgette-DHS on Thu, 13 Oct 2016 23:08:18 GMT
View Forum Message <> Reply to Message

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

Quote:I don't use SPSS, but I'm pretty sure I know what is happening. Thanks for the extra detail.

For many purposes, software requires weights to be integers. Why, I don't know, but that's why the DHS weights are multiplied by 1,000,000. Instead of averaging to 1, they will average to 1,000,000. Then there are so many significant digits to the left of the decimal point that the weight can be treated as an integer. When you divide v005 by 1,000,000 and treat that as a weight, I think SPSS is either rounding to the nearest integer or dropping everything to the right of the decimal place. In this KR file there are 8753 children (I just checked). If SPSS is rounding, then 1207 of the values will become 0. If it is truncating, then 4794 of the values will become 0. It's a good thing that you get a warning, because if you didn't, then you would be dropping all of those cases from your analysis without knowing it. This will not show up in the distribution of the variable you call rweightCH. It will only happen when that variable is used by the WEIGHT procedure.

So--If you use SPSS, you have to figure some way to get around this default. I hope other forum users can suggest how to do this.

Subject: Re: Sample weights disappear after merge household with childs recode
Posted by RKelders on Wed, 26 Oct 2016 15:19:14 GMT
View Forum Message <> Reply to Message

Thank you! You were right. Luckily I did not have this problem with all the other datasets and since this matter only occurred for 11 cases, I simply excluded these from analysis. These cases had a sample weight of 0,18 which SPSS analyzed as 0.

Kind regards,
Rachel