Subject: DHS Burundi 1987
Posted by tnkunzimana on Thu, 06 Nov 2014 14:15:47 GMT
View Forum Message <> Reply to Message

Dear all,

I want to use DHS of 1987 for anthropometric measurements.
Who can help for the stata code for results by province or by region?
As I have a dataset on Height and Weight Scores - WHO Child Growth Standards
(BUHW01FL.dta)with 'hwcaseid' var., how can merge this dataset with the one of the household
raw?
Many thanks for your help.

Tharcisse

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Sat, 08 Nov 2014 02:47:08 GMT
View Forum Message <> Reply to Message

You need to use the height and weight dataset with the children's recode (KR) file.

The following code will link the two together:

```
* rename variables and sort data for merging
use "BUHW01FL.DTA", clear
ren hwcaseid caseid
ren hwline midx
sort caseid midx
save "BUHW01_temp.dta", replace

* Sort the children's recode data and then merge
use "BUKR01FL.DTA", clear
sort caseid midx
merge 1:1 caseid midx using "BUHW01_temp.dta"

* Example tabulation of stunted by region
gen stunted = (hc70<-200) if hc70 <9990
tab v101 stunted [iw=v005/1000000]
```

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by tnkunzimana on Mon, 10 Nov 2014 12:53:53 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

Many thanks for your reply.
When using the stata code sent, I do not have the same total number of children as the technical report.
I find in total 1,917.81 instead 1929.5 children in the report.
a) What should be the causes of this discrepancy?
b) Is it possible to merge my new dataset "BUHW01_temp.dta" with the the one with the household(Hhs) characteristics "BUHH01FL.dta"?
This will help me to figure out the main characteristics of the Hhs where children are stunted.
Many thanks and kind regards,
Tharcisse

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by Trevor-DHS on Mon, 10 Nov 2014 16:15:33 GMT
View Forum Message <> Reply to Message

The 1929 cases are based on using the old CDC/NCHS/WHO reference, not the WHO standard from 2006. If you use hw4 (from the old reference) instead of hc70 (based on WHO 2006 standard) in the tabulation below, you will get 1929 cases.

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by tnkunzimana on Mon, 10 Nov 2014 16:24:23 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

many thanks for your feed-back.
Tharcisse

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by Trevor-DHS on Mon, 10 Nov 2014 19:01:09 GMT
View Forum Message <> Reply to Message

I forgot to answer your second question. You can add the variables from the BUHH01 dataset using code like this added to the previous code:

```
* Drop merge variable from previous run
drop _merge
* Copy cluster and household variables
clonevar hhseg=v001
clonevar hhrugo=v002
* Merge data
merge m:1 hhseg hhrugo using "BUHH01FL.dta", keep(match)
```

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by tnkunzimana on Tue, 11 Nov 2014 04:10:57 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

Thank you very much for this response of my second question. I will try to use this and will come to you if necessary.

As I want to analyse the changes from 1987 to 2010, do I have to do the same exercise to retrieve anthropometric measures (Stunted, Wasted and underweigthed children)for DHS2010? I mean do I have to merge children data set with the one of Households (Hhs) in order to assess the characteristics of Hhs where there are severe stunting, wasting and underweighted children? Thanks for your help.
Regards,
Tharcisse

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by tnkunzimana on Tue, 11 Nov 2014 09:19:32 GMT
View Forum Message <> Reply to Message

Hi Trevor-DHS,

You forget to attach the tabulation you talked about in your previous response:
"The 1929 cases are based on using the old CDC/NCHS/WHO reference, not the WHO standard from 2006. If you use hw4 (from the old reference) instead of hc70 (based on WHO 2006 standard) in the tabulation below, you will get 1929 cases." message of Mon, 10 November 2014 17:15.
I didn't find any tabulation to be used with hw4 instead hc70.
Can I use the same formulation for wasted and under-weighted children?
Many thanks,
Tharcisse

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by Trevor-DHS on Tue, 11 Nov 2014 16:46:10 GMT
View Forum Message <> Reply to Message

If comparing to the 2010 data, if you want to use comparable data, look at HW70-HW72 in the KR file. Some household characteristics are already merged with this file, but some may not be, in which case you would need to merge the data in the same way.

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by Trevor-DHS on Tue, 11 Nov 2014 16:47:23 GMT
View Forum Message <> Reply to Message

I didn't include the tabulation. Just replace hc70 by hw4 in the example tabulation already provided.

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by Trevor-DHS on Tue, 11 Nov 2014 16:50:17 GMT
View Forum Message <> Reply to Message

Correction: not hw4, but hw5. And yes, you can use the same formulation with hc71 and hc72, or with hw8 and hw11.

Subject: Re: DHS Burundi 1987/ another REQUEST
Posted by tnkunzimana on Thu, 13 Nov 2014 09:07:05 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

Many thanks for this useful information.
Will come to you if necessary,
Kind regards,
Tharcisse

Subject: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Wed, 19 Nov 2014 19:29:16 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

I an struggling trying to find out a variable in DHS1987 (Burundi) which is similar to 'v023'= 'stratification used in sample design' recoded in Burundi DHS2010. Do you have any suggestion so that I can make my trend analysis in terms of province (administrative level 2)?
Regards,
Tharcisse

Subject: Re: DHS Burundi 1987 and 2010
Posted by Trevor-DHS on Thu, 20 Nov 2014 23:26:08 GMT
View Forum Message <> Reply to Message

It is not possible to reconstruct a variable for the stratification exactly as used in the sample

design from the Burundi 1987 survey as all of the information needed is not included in the dataset. Chapter 1 (page 1) of the survey report describes the 11 natural regions that formed the rural strata, and the grouping of those 11 regions into 5 groups. Appendix A provides more information on the stratification as well as defining the two urban strata of Bujumbura and Gitega. We don't have the information for the 11 rural strata in the dataset, but we do have the 5 rural groups, and we can use the following as an approximation to the stratification:

egen v023 = group(v102 v101)
This will create strata formed by cross classifying the 5 regions by the urban and rural areas and will result in two urban strata plus the five rural groups as strata.

Subject: Re: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Fri, 28 Nov 2014 05:59:54 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

Many thanks for your reply.
So, that means it is impossible to retrieve a variable relate on provinces instead regions(11) and urban/rural.
Best regards,
Tharcisse

Subject: Re: DHS Burundi 1987 and 2010
Posted by Trevor-DHS on Fri, 28 Nov 2014 17:37:17 GMT
View Forum Message <> Reply to Message

Yes, as best I can tell there is only information for the 5 natural regions, and no information on the provinces in which the household resided, and I don't believe it is possible to retrieve more information for this survey.

Subject: Re: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Fri, 06 Mar 2015 18:32:58 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

I am struggling trying to merge data sets (HR, household) and (KR, children) in DHS2010 for Burundi. I tried many times but still have this in my results:
"variables v001 v002 do not uniquely identify observations in the master data" Can you tell me how you come up with your problem ?
Here is my codes:
****

use "BUHR61FL.DTA", clear

```
ren hv001 v001
ren hv002 v002
sort v001 v002
save "BUHW6FL_temp.dta", replace

use "BUKR61FL.DTA", clear
sort v001 v002
merge 1:1 v001 v002 using "BUHW6FL_temp.dta"
drop _merge
save "BUHW6FL_BUKR6FL.dta", replace
 ************************************************************* *****
```

Many thanks in advance,
Tharcisse

Subject: Re: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Fri, 06 Mar 2015 18:36:03 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,

I will need to do the same exercise for Rwanda (2010), Tanzania (2010), Kenya (2008-2009) and
Uganda (2011)
many thanks for you help!
Regards,
Tharcisse

Subject: Re: DHS Burundi 1987 and 2010
Posted by Reduced-For(u)m on Sat, 07 Mar 2015 22:23:12 GMT
View Forum Message <> Reply to Message

Two things:

1 - this is not a 1:1 merge - there are often more than one child in each household. Try m:1 or
1:m, depending on which one you start with. There is help here:

http://www.stata.com/help.cgi?merge

2 - what variables do you need from the HH file that are not in the child recode?

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Sat, 07 Mar 2015 22:41:08 GMT
View Forum Message <> Reply to Message

Reduced-for(u)m is right. You need to use
merge m:1 v001 v002 using "BUHW6FL_temp.dta"
in the example you give below as there are potentially multiple children from each household.

Subject: Re: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Mon, 09 Mar 2015 09:30:41 GMT
View Forum Message <> Reply to Message

Many thanks for your reply.
Yes, as it is possible to have many children for one household, I used this new code with Stata 13 and it worked!
Kind regards,
Tharcisse

Subject: Re: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Mon, 09 Mar 2015 16:09:32 GMT
View Forum Message <> Reply to Message

Dear Mr/Ms.,
I am using the following stata codes to retrieve the Stunting figures by province in Burundi (DHS2010).
****
gen weight = v005/1000000
tab hw70 if hw70>9990,m nolabel
gen HAZscore=hw70
replace HAZscore=. if HAZ>=9996
gen stunted=.
replace stunted=0 if HAZscore ~=.
replace stunted=1 if HAZscore <-200
gen eligch = 0
replace eligch = 1 if v135==1 (control for usual resid.)
tab province stunted [iw=weight] if eligch==1, row
****

Unfortunately, I am getting the total number of under five (3528) different form the one in the formal DHS report (3590).
Do I have to control another element? I get also 57.92 % of stunted children under-5 instead 57.7% from the report.
Many thanks for your help.
Regards,
Tharcisse

Subject: Re: DHS Burundi 1987 and 2010

Posted by Trevor-DHS on Tue, 10 Mar 2015 16:39:51 GMT
View Forum Message <> Reply to Message

For the Burundi 2010 survey, the anthropometric indicators are calculated for all children in the household, not just the children of women interviewed. To match the numbers of cases you need to use the Person's Recode (PR) file and use the HC variables rather than the HW variables.

Subject: Re: DHS Burundi 1987 and 2010
Posted by tnkunzimana on Tue, 10 Mar 2015 18:50:41 GMT
View Forum Message <> Reply to Message

Dear Trevor-DHS,
Many thanks for your reply. As, from the PR data set I don't have some characteristics I need for child under-5 (like vaccination, vitamin A,...), I will merge then PR and KR data sets.
Regards,
T

Subject: Re: DHS Burundi 1987
Posted by id709nvz on Mon, 28 Sep 2020 21:14:42 GMT
View Forum Message <> Reply to Message

Dear,

I have used this code to merge Ethiopian HW with children's KR file (DHS 2000 AND 2005). The merge doesn't seem ok. could you please highlight on how can I get the desired results? Here is an example of the code I used.

Height/weight measurement for Ethiopian DHS 2000
*Height/Weight measurement
use ETHW41FL.DTA, clear
rename hwhhid caseid
rename hwline midx
sort caseid midx
save ETHW41FL.DTA, replace

*Children-KR file
use ETKR41FL.DTA, clear
sort caseid midx
merge 1:1 caseid midx using "ETHW41FL.DTA"

Result   # of obs.
    ----
    not matched   21,220
      from master   10,873 (_merge==1)
      from using   10,347 (_merge==2)

matched   0 (_merge==3)

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Mon, 05 Oct 2020 18:53:51 GMT
View Forum Message <> Reply to Message

You need to first merge the data with the PR file (which is the natural file to use for anthropometry for this survey), and then based on that to then merge the PR file with the KR file. Try the following code:

```
* First merge with PR file as we need the mother's line number plus the cluster and household number
use ETHW41FL.DTA, clear
rename hwhhid hhid
rename hwline hvidx
sort hhid hvidx
save ETHW41FL_sort.DTA, replace

*Members-PR file
use ETPR41FL.DTA, clear
sort hhid hvidx
merge 1:1 hhid hvidx using ETHW41FL_sort.DTA
keep if _merge==3

* Second merge resulting file with KR file
rename hv001 v001
rename hv002 v002
rename hc60  v003
rename hvidx b16
keep hhid v001 v002 v003 b16 hc7*
save ETHW41FL_sort2.DTA, replace

*Children-KR file
use ETKR41FL.DTA, clear
sort v001 v002 b16
* drop cases with no matching entry in the PR file
drop if b16==. | b16==0
merge 1:1 v001 v002 b16 using ETHW41FL_sort2.DTA
```

Note that there will be some cases in the using file that do not match. These are children listed (and weighed and measured) in the PR file that were not children of interviewed women.

Subject: Re: DHS Burundi 1987

Posted by id709nvz on Mon, 05 Oct 2020 20:52:39 GMT
View Forum Message <> Reply to Message

Dear,
Thank you very much.
Since I want to combine KR file of 2000, 2005, 2011 and 2016, one issue just wind up.
height and weight measurement for the 2000 and 2005 is HC7* while for 2011 and 2016 is HW7*.
I am thinking this is because for the earlier all children in the household is measured and for the
later the measurement is only for children of interviewed women. If this is correct can I

keep if _merge==3 and then rename HC7* HW7*

I am mainly interested combined data of children H/W of interviewed women (mother).

Kind regards

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Mon, 05 Oct 2020 22:17:29 GMT
View Forum Message <> Reply to Message

I think you are misunderstanding the data. In all of the surveys that you are using the
anthropometry data were collected for children listed in the household schedule (the PR file). For
the earlier surveys (2000 and 2005), for which you need to use the HW files, the WHO growth
standards did not exist at the time of those surveys, and so these HW files were constructed later.
These HW data files are designed to be used with the PR file, and include the hc7* series of
variables. These variables already exist if the PR file for the later surveys (2011 and 2016). If you
wish to analyze anthropometry we recommend using the PR file.

The KR files include the hw7* series of variables which are a copy of the hc7* series of variables
from the PR file, but restricted just to children of interviewed women. Thus analyzing
anthropometry using the KR file will give you somewhat biased estimates as it is only based on
children of women interviewed, and thus likely excludes orphaned and fostered children. Thus we
recommend using the PR file.

Why do you need to restrict your analysis just to children of interviewed women?

If you do need to restrict your analysis and need to use the KR file, then, yes, you can rename
hc7* to hw7* when you merge the data.

Subject: Re: DHS Burundi 1987
Posted by id709nvz on Tue, 06 Oct 2020 15:55:43 GMT
View Forum Message <> Reply to Message

Dear,
Thanks for your valuable comments.

I am analyzing the effect of variable X on variable Y (that is related to mother's), and now I am being suggested to look at the impact of the same X on newborn birth weight in order to understand intergenerational effects of the variable X. This is why I am restricting the sample to interviewed women (guessing they are the mothers).
Aren't this a valid approach to follow?
If I may add to this question, is it correct to do the weighting after pooling all survey together? or should I first do the weighting for each survey first and then do pooling? i.e gen wgt= v001/1000,000 (for 2000, 2005, 2011, and 2016 individually first than).

Kind regards,

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Tue, 06 Oct 2020 17:10:50 GMT
View Forum Message <> Reply to Message

Thanks - that makes sense. I just wanted to make sure that you were using the KR file for a good reason.

Weighting is applied when you analyze the data after you have pooled the data. Note that you cannot just combine the datasets and use them as they are with the weights given for each survey as the probability of selection of a woman in each survey varies. You need to adjust the weighting to take into account that the probability of selection differed. The sample sizes for the surveys were:

Year : Sample : Total women 15-49* : Probability of selection
2000 : 15367 : 14,619,000   : 1 in 951.324
2005 : 14070 : 16,861,000   : 1 in 1198.365
2011 : 16515 : 20,659,000   : 1 in 1250.923
2016 : 15683 : 24,949,000   : 1 in 1590.831
* Total women 15-49 is taken from the UN's World Population Prospects, but you could also use other estimates, such as census estimates.

See other posts on the forum about pooling.

Using gen wgt = v005/1000000 is not applying the weights, just calculating a weight variable to be used later (note that it uses v005, not v001). The weights are applied later when analyzing the data, and should be adjusted to take into account the probabilities of selection as above.

Subject: Re: DHS Burundi 1987
Posted by id709nvz on Wed, 07 Oct 2020 01:39:06 GMT
View Forum Message <> Reply to Message

Dear,
thanks for your insight and helpful comments. After going through the forum Q&A I follow the following weighting. Could you please evaluate it?

*After opening KR file of 2000 survey
gen wgt=v005/1000000
gen weight=(wgt*FEMALE POPULATION)/FEMALE SAMPLE -For KR 2000; FEMALE
POPULATION=14619 and FEMALE POPULATION=15367
gen survey=1

*Same procedure for 2005, 2011 and 2016. Except that I used different total number of female
population, female sample and with gen survey 2, 3, and 4 respectively. Then I appended the data
starting from the latest survey (2016). Afterward, I did the following.

egen cluster=group(survey v021)   *V021 is PSU and survey identifies year of survey (coded 1,
2, 3 and 4)
egen stratum=group(survey v022)   *v022 is sample strata for sampling error
svyset cluster [pw=wgt], strata(stratum) singleunit(centered)

Is it correct to use: tab x1 x2 [iweight=wgt] or tab x1 x2 [iweight=weight]
how about svy: reg Y x1 x2 (Will this understand the complex nature of the survey).
Added to this: svy is not supported with ivreg2 stata command is there an alternative to this? I
simply used ivreg2 with cluster option at v001.

Thank you in advance for any comments.
Kind regards!

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Wed, 07 Oct 2020 04:31:40 GMT
View Forum Message <> Reply to Message

Yes, this mostly looks good. A couple of notes:
1) For 2000, FEMALE POPULATION=14619 and FEMALE SAMPLE=15367. I assume this is just
a typo in the message.
2) For the svyset it should use [pw=weight]. wgt is for use with a single survey, weight would be
used for the pooled dataset.
3) For tabulating the combined pooled file, you should use tab x1 x2 [iweight=weight], not wgt
4) The svy: reg will use whatever you have defined in the svyset, so that should be weight for the
pooled dataset, or wgt if you are using only a single survey.
5) I don't know ivreg2, so I can't advise on this, but typically not taking the complex sampling into
account (with svy) means that significance tests and confidence intervals are not accurate.

Subject: Re: DHS Burundi 1987
Posted by id709nvz on Wed, 07 Oct 2020 08:47:11 GMT
View Forum Message <> Reply to Message

Dear,

I thank you very much for your comments.

As for total population it is not by mistake even if puzzling I took it from world economic perspective (from the link you provided).
Even you pointed out such stat in your earlier post (when showing probability of being selected in each survey). I assume I have to look for another sources. Or maybe I calculated it wrongly (I simply sum over the entire row of 15-49 population).

Thanks for your comments.
Kind regards
Zecharias

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Wed, 07 Oct 2020 14:01:31 GMT
View Forum Message <> Reply to Message

Your calculations should be fine, but your example gave
2000: FEMALE POPULATION=14619 and FEMALE POPULATION=15367
I corrected this to
2000: FEMALE POPULATION=14619 and FEMALE SAMPLE=15367
(also strictly speaking the population is actually 14619000 but if you drop the 000 from each population then the relative weights are fine.)

Subject: Re: DHS Burundi 1987
Posted by id709nvz on Wed, 07 Oct 2020 14:18:31 GMT
View Forum Message <> Reply to Message

Ahh, thanks so much. Now, I corrected it.

I can't thank you enough for your valuable time and comments!

kind regards

Subject: Re: DHS Burundi 1987
Posted by Adetunji on Thu, 08 Aug 2024 17:19:16 GMT
View Forum Message <> Reply to Message

Thanks for the code. I am doing the same analysis on Nigeria DHS and I need to merge the HW data with the KR file for DHS IV. While the code you shared worked well with no error code, it failed to produce any match. I think this because the caseid in the KR file has three characters in the string, whereas that of HW is only two. Could that be the reason and what do you advise I do?

Here is the code and the screenshot of the output:

```
use "${hw2003}", clear
ren hwhhid caseid
ren hwline midx
sort caseid midx
save "${kr2003}_temp.dta", replace

* Sort the children's recode data and then merge
use "${kr2003}", clear
sort caseid midx
merge 1:1 caseid midx using "${kr2003}_temp.dta"
```

File Attachments
1) Screenshot (424).png, downloaded 212 times

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Thu, 08 Aug 2024 18:01:01 GMT
View Forum Message <> Reply to Message

Hi Jacob

A couple of problems with your code: 1) caseid in the women's data is 15 characters, whereas hwhhid is just 12 characters for the household ID. You need to drop the woman's line number part of caseid which is the last 3 characters. 2) midx is the variable that links from the KR file to the birth history, not to the household schedule. B16 is the variable that links to the household schedule.

I think the following code should do what you need:
```
* Sort the children's recode data and then merge
use "${kr2003}", clear
gen hwhhid = substr(caseid,1,12)
clonevar hwline = b16
sort hwhhid hwline
merge m:1 hwhhid hwline using "${hw2003}"
```
and then just use those cases with _merge==3 (it looks like there are 6 cases that do not match properly, with _merge==2)

Cheers. Trevor

Subject: Re: DHS Burundi 1987
Posted by Adetunji on Mon, 12 Aug 2024 17:41:54 GMT
View Forum Message <> Reply to Message

Thanks, Trevor.

The code worked for me. However, there is no b16, child line number in Nigeria's 1990 dataset, and I am unable to merge the HW and the KR files.

Any suggestion as to what I should do, please?

Subject: Re: DHS Burundi 1987
Posted by Trevor-DHS on Mon, 12 Aug 2024 21:33:27 GMT
View Forum Message <> Reply to Message

Hi Jacob

There are two forms of the HW file: 1) Where the anthropometric data collection was based on children in the household roster, 2) Where the anthropometric data collection was based on children in the mother's birth history. The example just above for Nigeria 2003 is for the first kind, and for Burundi 1987 it is the second kind (actually the oldest kind from the earliest surveys). You can tell the difference by the variable names. If you find hwhhid then this is for the first kind based on the household roster, and if you find hwcaseid this is the second kind based on the mother's birth history.

The code for Burundi 1987 is given at the top of the thread here:
[url=https://userforum.dhsprogram.com/index.php?t=msg&th =1642&goto=3219&#msg_3219][/url]

Regards. Trevor