Subject: Duplicates in KR file and Empty values for HV005 after merging.
Posted by Mayank_Ag on Fri, 01 Jun 2018 10:09:29 GMT
View Forum Message <> Reply to Message

I am using DHS 15-16 for India and doing the analysis in SPSS.

1. While merging PR and KR files i found duplicate values in the KR file with same value for all indicators for the child. Can somebody please explain why is this happening and how to deal with them?

2. Further the merged file is having 3 missing values for HV005 (Household weight). This is after applying the following filters.

a) Child is alive (B5)
b) listed in the household(B16)
c) Removed the duplicates from KR file.

Is there any other filter i need to apply so that these empty values don't come in the final dataset? I have already gone through the previous posts on this forum but couldn't figure out a solution to this problem. Are these 3 cases coming because of the 6 duplicates?

Thanks in advance!!

Subject: Re: Duplicates in KR file and Empty values for HV005 after merging.
Posted by Bridgette-DHS on Tue, 05 Jun 2018 12:30:09 GMT
View Forum Message <> Reply to Message

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

I can only reply in terms of Stata. The following Stata program is something I use only for difficult KR/PR merges. It has a lot of redundancy. It includes the sex and age (in months) of the child and the line numbers of the child and the mother. It works on this survey. There will be 244,384 children in the merged file. There will be no cases with missing weights. I hope you can convert to SPSS.

```
set more off
*set maxvar 10000
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAKR73FL.DTA", clear

keep caseid v001 v002 v003 v005 v024 b4 b16 hw1
keep if b16>0 & b16<.
gen in_KR=1
gen sex_child=b4
gen age_child=hw1
gen line_number_of_child=b16
gen line_number_of_mother=v003
```

```
gen hv001=v001
gen hv002=v002
gen hv024=v024
sort hv024 hv001 hv002 sex_child age_child line_number_of_mother line_number_of_child
*list hv024 hv001 hv002 sex_child age_child line_number_of_mother line_number_of_child if
_n<=50, table clean
save e:\DHS\DHS_data\scratch\KRtemp.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR73FL.DTA" , clear
keep hhid hvidx hv001 hv002 hv003 hv005 hv024 hv112 hv104 hc1
keep if hc1<.
keep if hv112>0 & hv112<.
gen in_PR=1
gen sex_child=hv104
gen age_child=hc1
gen line_number_of_child=hvidx
gen line_number_of_mother=hv112
sort  hv024 hv001 hv002 sex_child age_child line_number_of_mother line_number_of_child
*list  hv024 hv001 hv002 sex_child age_child line_number_of_mother line_number_of_child if
_n<=50, table clean
merge 1:1 hv024 hv001 hv002 sex_child age_child line_number_of_mother line_number_of_child
using e:\DHS\DHS_data\scratch\KRtemp.dta

tab in*,m
keep if in_KR==1 & in_PR==1
drop in*
```

Subject: Which variable to use.
Posted by Mayank_Ag on Tue, 05 Jun 2018 18:10:21 GMT
View Forum Message <> Reply to Message

Thanks a lot for your reply. I merged KR with PR and did not remove the obs with missing HV005. I used it to calculate the estimates as presented in Table 10.1 of the final report. My estimates are matching except for the categorization according to Mother's Nutritional Status. I have tried more than 50 combinations but the estimates are coming out the same. I have attached a image for your reference.

I saw posts in the forums where it was mentioned that HA40 should be used. But with HA40 the number of valid obs is coming out to be 3 only (I don't know how). It was also mentioned in the posts that V445 and HA40 are the same variables but i found that they have different values for the obs. I tried using V445 for the estimates but they didn't match.

I applied all the filters as mentioned in the reports (Pregnancy(V213), Birth within 2 months(V222), Mother Interviewed(V015)). Can you please tell me what i am doing wrong here. Are the filters sufficient and right? Please tell me how to go forward with this.

Thanks in advance.

File Attachments

1) BMI.JPG, downloaded 526 times

Subject: Re: Which variable to use.
Posted by Bridgette-DHS on Thu, 07 Jun 2018 21:53:50 GMT
View Forum Message <> Reply to Message

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

I think you may be making this more complicated than necessary. This part of table 10.1 in the NFHS report is constructed directly from the KR file, with children as units. The following lines will give you, for example, the 45.8% in the image you attached. This is the percentage of children who are stunted, given that the mother is underweight. This could be modified to give more numbers, but for simplicity I just match that number. Can you try this approach?

```
gen stunted=0
replace stunted=100 if hw70<-200
replace stunted=. if hw70<-600 | hw70>600

gen mo_underwt=0
replace mo_underwt=1 if v445<1850
replace mo_underwt=0 if v445>=99988

summarize stunted if mo_underwt==1 [iweight=v005/1000000]
```

Subject: Re: Which variable to use.
Posted by Mayank_Ag on Fri, 08 Jun 2018 09:57:49 GMT
View Forum Message <> Reply to Message

I tried the same code. But i have some doubts.

1) The percentages are matching but the not the no. of obs. (Estimates Attached)
2) Further you have not used any of the conditions mentioned in the footnotes. (Images Attached)
   i) Pregnancy
   ii) Birth within preceding 2 months.
   iii) Slept at night

3) Why did you put 0 for the values greater than 9998 for the BMI?

Estimates

Underweight - 45.9; 51103
Normal- 38.2; 128260
Overweight 27.1; 32318

File Attachments

1) BMI 2.JPG, downloaded 517 times
2) BMI3.JPG, downloaded 549 times

Subject: Re: Which variable to use.
Posted by Mayank_Ag on Tue, 10 Jul 2018 06:34:31 GMT
View Forum Message <> Reply to Message

Can please someone look into this. I am still not able to figure out a way out of this.
