Subject: Need Help to Clarify
Posted by Aamna on Thu, 13 Oct 2022 22:09:39 GMT
View Forum Message <> Reply to Message

Hello,

I am working on the India women dataset (2019-2021). I am analyzing data for women who gave birth in the last 5 years. I noticed that some variables have 80% or more missing data, e.g., v169a, v170, v743a. I am unsure if I have a corrupt file due to which missing data has occurred or if missing data was originally present in the file. Can you kindly help and guide me about it? I appreciate the help.

Aamna

Subject: Re: Need Help to Clarify
Posted by Janet-DHS on Tue, 18 Oct 2022 20:40:39 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member Tom Pullum:

This survey had a 1/6 subsample of men for the men's interview. (In 1/3 of the clusters, half of the households were selected.) In the PR file there is a variable hv027 that is 1 if the household was selected for the men's interview. This subsampling had implications for the women's interview. For example, the questions for the variables you mention (v169a, v170, v743a) were only asked in the households selected for the men's interview. The design of the survey anticipated that researchers would want to relate them to characteristics of the husband.

The PR/IR merge below confirms that v170, for example, is Not Applicable (coded with a dot) for women in the households that were not selected for the men's interview.

* specify workspace
cd e:\DHS\DHS_data\scratch

* Prepare the IR file for the merge
use "...IAIR7DFL.DTA", clear

keep v001 v002 v003 v025 v170
rename v025 state
rename v001 cluster
rename v002 hh
rename v003 line
sort state cluster hh line
save IAtemp.dta, replace

```
* Prepare the PR file and do the merge
use "...IAPR7DFL.DTA", clear

keep if hv117==1
keep hv001 hv002 hvidx hv025 hv027
rename hv025 state
rename hv001 cluster
rename hv002 hh
rename hvidx line
sort state cluster hh line
merge state cluster hh line using IAtemp.dta
tab _merge
keep if _merge==3

* Check the pattern of NA on v170 with the subsampling
tab hv027 v170,m
```

Subject: Re: Need Help to Clarify
Posted by Aamna on Wed, 19 Oct 2022 04:42:15 GMT
View Forum Message <> Reply to Message

Hello,

Thank you for the reply and explanation. I understood the reason for the missing data. I ran the codes in your reply and got the following results. I want to be sure that my results are accurate or the same as yours. I also want to ask that I only need to work with women's data corresponding to men's survey data observations (26,985) because the data for some variables is only available for these observations. If I am wrong in assuming it, then can you kindly guide me? I appreciate your help again.

Aamna

tab hv027 v170, mi

household selected for male interview	has an account in a bank or other financial institution			Total
	no	yes	.	
not selected	0	0	149,892	149,892
men's survey	5,436	21,549	0	26,985
Total	5,436	21,549	149,892	176,877

```
tab _merge

        _merge |     Freq.     Percent      Cum.
---------------+-----------------------------------
master only (1) |   570,299       76.33     76.33
   matched (3) |   176,877       23.67    100.00
---------------+-----------------------------------
         Total |   747,176      100.00
```

Following is a response from DHS staff member Tom Pullum:

Your results look fine. Yes, if you are making a work file, you can limit it to the cases you describe. Other cases would be dropped from the analysis anyway.