
Subject: Merging of NFHS-4 dataset

Posted by [DHS user](#) on Thu, 25 Jul 2019 20:29:13 GMT

[View Forum Message](#) <> [Reply to Message](#)

I am working with NFHS-4 data set. For my study i tried to merge PR, IR and MR files. I did the merging using hv024, hv001, hv002 and hvidx variables from PR file, v024, v001, v002 & v003 from IR file and mv024, mv001, mv002 & mv003 from MR file. Since, type of caste/tribe variable is not given in MR file therefore, i used the type of caste/tribe of the household's head. But after merging the values for this variable for different categories are not in line with the values given in report.

I am using table 3.1(background characteristics of respondents) to match the values. Initially, I made the list of all the required variables and for merging I kept the variables which were common in all these three files in PR file only and those variables also which were not given in MR file. Variables like caste, BMI, type of cooking fuel and other housing characteristics were not given in MR file. So, I merged the MR and IR files with PR file. I kept only those men and women who were in 15-54 and 15-49 age groups respectively and dropped the unmatched.

Following are the list of variables that I am using from PR file:

hvidx hv001 hv002 hv005 hv009 hv010 hv011 hv012 hv013 hv024 hv025 hv027 hv028 hv201 hv214 hv215 hv216 hv217 hv226 hv239 hv241 hv242 hv252 hv270 hv271 shdistri sh34 sh36 sh39 shnfhs2 shstruc hv104 hv105 hv108 hv115 shb16s shb16d shb23s shb23d shb27s shb27d shb70 ha40 hb40

Following are the list of variables that I am using from IR file:

V001 v002 v003 v024 ha40 v463a v463b v463c v463d v463f v463g v463x v463z v481 v501 v714 v716 v717 s707 s708 s710c s710e s716 s717 s718a s718b s718c s718d s718e s718x s723a s726a s726b s726c s726d s726e s726f s726g s726h s726i

Following are the list of variables that I am using from MR file:

mv001 mv002 mv003 mv024 mv138 mv463a mv463b mv463c mv463d mv463e mv463f mv463g mv463x mv463z mv464 mv481 mv714 mv716 mv717 sm606 sm609c sm609e sm615 sm616 sm617a sm617b sm617c sm617d sm617e sm617x sm622a sm626a sm626b sm626c sm626d sm626e sm626f sm626g sm626h sm626i smb70

But after merging the values for religion(sh34), caste(sh36), marital status(hv115), education completed in single year(hv108),type of residence(hv025) etc. are not matching with the report. Although the total values are in line with the report but values within the categories are not matching.

Subject: Re: Merging of NFHS-4 dataset

Posted by [Bridgette-DHS](#) on Thu, 25 Jul 2019 20:30:57 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I believe that table 3.1 was constructed from the IR and MR files without any merge with the PR file. When using the IR file, the weight is v005. When using the MR file, the weight is mv005. The age range for men is 15-54, so most of the results for men (other than the last two rows of table 3.1) must be restricted to mv012<=49. Please check whether this is correct, that is, whether you can match table 3.1 using just the IR and MR files, and let me know.

Subject: Re: Merging of NFHS-4 dataset
Posted by [DHS user](#) on Thu, 25 Jul 2019 20:32:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

Yes, table 3.1 is constructed from the IR and MR file without any merge with PR file. I am matching with the unweighted values and also tried to match values after applying weights but in both cases values are different from the report. I also restricted the age group for men to <=49 but here again values are different within categories.

Before merging, values (total and within categories) from IR and MR file are matching with the report. The problem arises only after the merging. Even after merging, the values for the variables which belong to IR and MR file are not getting disturbed. That is it is only for those variables which are coming from PR file that values are not matching.

Subject: Re: Merging of NFHS-4 dataset
Posted by [Bridgette-DHS](#) on Thu, 25 Jul 2019 20:33:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

I believe the problem is that sh34 and sh36 in the PR file are the religion and caste of the household head, not the respondent. For individual respondents they are given by v130 and v131 for women and mv130 and mv131 for men. Also, hv115 is an initial statement of marital status, during the household interview, but when the woman or man is interviewed individually, v501 and mv501 are considered to be more accurate than hv115. The following lines will give you a file that I think matches Table 3.1 Use v005 for weighted distributions for women and mv005 for men.

* Prepare IR and MR files

```
use v001 v002 v003 v005 v024 v013 v106 v130 v131 v501 using "C:\Users\26216\ICF\Analysis -  
Shared Resources\Data\DHSdata\IAIR74FL.DTA", clear  
rename v001 hv001  
rename v002 hv002  
rename v003 hvidx  
rename v024 hv024  
save e:\DHS\DHS_data\scratch\IAIR74temp.dta, replace
```

```
use mv001 mv002 mv003 mv005 mv024 mv013 mv106 mv130 mv131 mv501 using  
"C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAMR74FL.DTA", clear  
rename mv001 hv001  
rename mv002 hv002  
rename mv003 hvidx  
rename mv024 hv024  
append using e:\DHS\DHS_data\scratch\IAIR74temp.dta
```

```
sort hv024 hv001 hv002 hvidx  
save e:\DHS\DHS_data\scratch\IAIRMR74temp.dta, replace
```

* Prepare PR file and merge

```
use hv001 hv002 hvidx hv024 sh34 sh36 hv104 hv115 hv106 using "C:\Users\26216\ICF\Analysis  
- Shared Resources\Data\DHSdata\IAPR74FL.DTA", clear  
sort hv024 hv001 hv002 hvidx  
merge hv024 hv001 hv002 hvidx using e:\DHS\DHS_data\scratch\IAIRMR74temp.dta  
tab _merge
```

```
keep if _merge==3  
drop _merge
```

* Get distributions

```
tab1 v130 mv130 if mv013~=8  
tab1 v131 mv131 if mv013~=8  
tab1 v501 mv501 if mv013~=8
```