Subject: Merging DHS Pakistan 2017-2018
Posted by shabina1129 on Thu, 14 Nov 2024 22:08:07 GMT
View Forum Message <> Reply to Message

Hi,

My current research is focusing on mother-in-law's educational attainment and how that can alter their daughter-in-law's autonomy within the household. I have 2 questions.

1) I am trying to merge PR and IR files together. This is what I have done so far:

```
use "/Users/shabina/Desktop/Fall 2024 /PK_2017-18_DHS_08072024_1755_200211/Household Member/household 1.dta"

rename hv001 v001

rename hv002 v002

rename hvidx v003

sort v001 v002 v003

 save "/Users/shabina/Desktop/Fall 2024 /PK_2017-18_DHS_08072024_1755_200211/Household Member/household 1.dta", replace

 *open womens data
 save " /Users/shabina/Desktop/kristinbietsch-MIL-Analysis-c4d7973/w omen1.dta "

 use " /Users/shabina/Desktop/kristinbietsch-MIL-Analysis-c4d7973/w omen1.dta "

 sort v001 v002 v003

 keep if _merge==3


*Attempted to merge the household file onto the womens only file.
/*
    Result                      Number of obs
    -----------------------------------------
    Not matched                        85,801
        from master                         0  (_merge==1)
        from using                     85,801  (_merge==2)

    Matched                            15,068  (_merge==3)
    -----------------------------------------
```

This resulted in only 15,068 matches which the original womens only file contains, Leaving out around 30,000 of the sample. Is there another way to merge the datasets while keeping the sample I have which is 51,044?

I keep seeing that this is the only way to merge PR and IR but I wanted to know if there is another way to merge these datasets while still keeping the number of obs.

2) My second question is that if I were to try to run a regression on mother-in-law's education on standardized
autonomy there would either be no cases if I only run the regression on daughter-in-laws, or the regression would run solely on mother-in-laws since they are the only rows that have values for both the educational level obtained and standardized autonomy scale.
- Is there a way that we can construct the mother-in-law education variable to be available or on the same row as the daughter-in-law respondents so that regression analyses can be executed? If you know the codes for this. It would be so appreciated. Thanks.

Subject: Re: Merging DHS Pakistan 2017-2018
Posted by Bridgette-DHS on Fri, 15 Nov 2024 22:27:13 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

Your merge worked correctly. I did it a little differently and got the same result (see below).

The PR file includes all household members, including males of all ages and females who are not eligible for the women's interview. When I do this merge I first reduce the PR file to household members who are eligible for the women's interview. These are the cases with hv117=1.

By the way, the lines you gave were incomplete. You did not include the "merge" command. Also your first "save" line has the same filename as the file in the first "use" line. That command would over-write the first file, something to be avoided at all costs and something I'm pretty sure you did not do. Anyway, your merge did work correctly.

```
* Specify workspace
cd e:\DHS\DHS_data\scratch

use "...PKPR71FL.DTA", clear

* Reduce to women who are eligible for the IR file
keep if hv117==1
gen cluster=hv001
gen hh=hv002
gen line=hvidx
save temp.dta, replace

use "...PKIR71FL.DTA", clear
gen cluster=v001
gen hh=v002
gen line=v003
```

```
merge 1:1 cluster hh line using temp.dta
tab _merge

* women with _merge=2 were eligible but not interviewed; they count as nonresponse cases
keep if _merge==3
drop _merge

* Save the merged file with a different name
```

Subject: Re: Merging DHS Pakistan 2017-2018
Posted by shabina1129 on Fri, 15 Nov 2024 22:31:31 GMT
View Forum Message <> Reply to Message

Thank you for your response. I did do keep if _merge==3 I just forgot to include that. If you do know the answer to this question too that would be really helpful as well!

2) My second question is that if I were to try to run a regression on mother-in-law's education on standardized
autonomy there would either be no cases if I only run the regression on daughter-in-laws, or the regression would run solely on mother-in-laws since they are the only rows that have values for both the educational level obtained and standardized autonomy scale.
- Is there a way that we can construct the mother-in-law education variable to be available or on the same row as the daughter-in-law respondents so that regression analyses can be executed? If you know the codes for this. It would be so appreciated. Thanks.

Subject: Re: Merging DHS Pakistan 2017-2018
Posted by Bridgette-DHS on Fri, 13 Dec 2024 16:48:53 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

Sorry for there is a delay in this response (but I communicated directly, soon after you posted your question). Attached is a program that does what you want to do. Also, it uses a broader definition to find a mother-in-law, when she is either the household head herself or is the spouse of the head.

Let us know if you have other questions.

File Attachments

1) attach_moinlaw_characteristics_to_dtrinlaw.txt, downloaded

14 times