
Subject: Reshaping Data

Posted by [shabina1129](#) on Mon, 18 Nov 2024 22:52:49 GMT

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

Hello,

I am currently working with the DHS Pakistan data 2017-2018 using PR and IR file. I would like to know how I can reshape my dataset to construct the mother-in-law education variable on the same row of the daughter-in-law respondents. Here is my do-file with more indepth explanations.

```
* import household dataset before starting*
```

```
* Trying same code with female restricted from the beginning *
```

```
sort hhid hvidx  
list hhid hvidx
```

```
* making inlaw have 0-1 and restricting data to women only *
```

```
keep if hv104 == 2
```

```
gen inlaw =.
```

```
replace inlaw = 1 if hv101 == 2
```

```
replace inlaw = 0 if inlaw==.
```

```
tab inlaw
```

```
quietly by hhid: generate suminlaw=sum(inlaw)
```

```
list hhid inlaw suminlaw
```

```
quietly by hhid: replace suminlaw=suminlaw[_N]
```

```
/* There are still multiple mother in laws in a single household.
```

Variable	Obs	Mean	Std. dev.	Min	Max
-----+-----					
suminlaw	51,044	.8580244	.4011059	0	4

We browsed hhid hvidx inlaw and suminlaw and noticed that there are households where suminlaw exceeded 1 (instances of 2, 3, or 4 MILs)

```
browse hhid hvidx inlaw suminlaw
```

There might be speculation that this may be a result of male head of households having multiple wives, which in turn each wife is stating that they are the mother-in-laws.

```
*/
```

```
/* Household characteristics available in the household data include age and highest level of education obtained. Employment status is not asked.
```

*/

* age *
tab hv105

* highest level of education attained *
tab hv106

* Merging attempt *

/* Following an online recommended table and description of how to merge the Pakistan DHS household dataset to the individual dataset of ever-married women aged 15-49).

The post recommended that the variables in the household dataset should be renamed to match those in the individual dataset. And then the data should be sorted to match the same sorting configuration of the individual dataset.

```
rename hv001 v001 - Cluster number  
rename hv002 v002 - Household number  
rename hvidx v003 - Respondent's line number answering HH questionnaire
```

```
sort v001 v002 v003
```

*/

```
rename hv001 v001  
rename hv002 v002  
rename hvidx v003  
sort v001 v002 v003
```

/* We then use the following steps to merge the datasets.

1. Save the newly configured household dataset.
2. Open the Pakistan DHS individual ever-married women dataset.
3. Merge the datasets on the three configured identification variables.

*/

```
merge 1:1 v001 v002 v003 using "The Configured HH Dataset"
```

/* If we use the professors code, then the merge will have the following results. (We kept only females/Mother in Law's so N = 51,044)

Result	Number of obs
--------	---------------

Not matched	35,976
from master	0 (_merge==1)
from using	35,976 (_merge==2)
Matched	15,068 (_merge==3)

Instead -- if we use the professor's code, without restricting the household dataset to only women, we get the following results.

Result	Number of obs

Not matched	85,801
from master	85,801 (_merge==1)
from using	0 (_merge==2)
Matched	15,068 (_merge==3)

In either case, there seems to only be 15,068 women that have matchable data in both the household and individual Pakistan DHS datasets.

So as of right now, the working sample would be 15,068 ever-married women aged 15-49

Of the 15,068 sample of ever-married women aged 15-49

Mother-in-law (hv101 == 2 for women only) N = 7,841

Daughter-in-law (hv101 == 4 for women only) N = 4,300

Dataset and Analytical Concerns

1. Assuming the merge was done correctly (this should be revisited) -- then the `suminlaw` variable the Professor created is incorrectly counting the cumulative number of MILs in a household (there seems to be some households that have 2, 3, or 4 mother-in-laws).

There could be an instance culturally that husbands have multiple wives in the household who all state that they are the wives (and therefore are the mother-in-laws to the children in the home).

2. Analytically, due to the way that the data is collected in stored, regression analyses are not possible. Specifically there is an issue with variables not being able to be on the same row as a respondent, (for instance, mother-in-law education is missing for respondents who are the daughter in law).

So a RQ such as how does mother-in-law's level of education affect the daughter-in-law's autonomy is impossible to do, solely because there are 0 cases of the mother-in-law education variable available to be regressed on the same line as the daughter-in-law respondent.

Thus, the next step would be to figure out how to construct the mother-in-law education variable on the same row of the daughter-in-law respondents.

See attached word document with a snapshot of the browse function of the identification variables, mother-in-law constructed education variable, and the standardized autonomy variable. (I uploaded this to the forum)

- Note how there is missing for mother-in-law education on the lines in which the respondent is a daughter in law, making analyses impossible on the daughter-in-laws specifically.

Please let me know !

File Attachments

1) [Screenshot 2024-11-18 at 5.50.04?PM.png](#), downloaded 19 times
