Subject: Reshaping Data
Posted by Anonymous 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 identiification 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 revisted) -- 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 99 times

---

Subject: Re: Reshaping Data
Posted by Bridgette-DHS on Tue, 19 Nov 2024 13:09:53 GMT
View Forum Message <> Reply to Message

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

The variable hv101 in the PR file gives the relationship of each person in the household head to the household head. It does not give the relationship of individuals to one another. It is confusing to me, at least, when you use the terms mother-in-law and daughter-in-law, because these are relationships to the household head. If H is the head of the household, usually a male, then his mother-in-law is his wife's mother and his daughter-in-law is his son's wife. I don't think these are the people you are talking about.

I think you are talking about three possible configurations. One type is male-headed households, and the two people are the wife of the head and the mother of the head. Then (a) the head of the household is male, (b) the spouse is present (and female), and (c) a female parent of the head is present. You want to put the education of person (c) onto the record of person (b).

In the second configuration, (a) the head is male, (b) the spouse is present (and female), and (c) the daughter-in-law of the head is present. You would have to assume that the daughter-in-law of the head is also the daughter-in-law of the head's spouse. Again, you would put the education of person (c) onto the record of person (b).

The third configuration is like the second, but (a) the head is female and (b) her daughter-in-law is present, This time you put the education of person (a) onto the record of person (c).

Some households will not have any such pairs. It is possible for a household to have multiple pairs, e.g. a woman-headed household that includes two adult married sons and their wives.

Does this sound correct? If so, let us know and I will show how to do that. If not, please clarify your question.

Subject: Re: Reshaping Data
Posted by Anonymous on Tue, 19 Nov 2024 19:07:24 GMT
View Forum Message <> Reply to Message

Hi Tom,

These are all great configurations but let me clarify what I am looking for. My research aligns closely with " The third configuration is like the second, but (a) the head is female and (b) her daughter-in-law is present, This time you put the education of person (a) onto the record of person (c)." However, I am more interested in having person (a) (female head of household's education) onto person (b) (daughter-in-law). If this doable please let me know.

Subject: Re: Reshaping Data
Posted by Bridgette-DHS on Wed, 20 Nov 2024 13:36:05 GMT
View Forum Message <> Reply to Message

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

Right--I had a typo and meant to say "... put the education of person (a) onto the record of person (b)." Glad you corrected that.

This (the "third configuration") is a situation in which the mother-in-law is the head of the household. There is fairly often an interest in attaching the education of the household head to everyone else in the household. I see that in the PR file for the Pakistan 2017-18 survey the head's sex and age are coded onto the records for all household members as hv219 and hv220 but the head's education is not.

There are 4 relevant education variables in the PR file: hv106, hv107, hv108, hv109. I will paste below a Stata program that does what I think you want to do. It includes a crosstab of hv106 for the two women.

The percentage of women who are in this kind of a pair is very small. It's an interesting topic but you may want to expand the relationships within the household. Let us know if you have other questions.

```
* Compare education of mother-in-law and daughter-in-law in Pakistan 2017-18 survey

* Mother-in-law is female head of household: hv101=1 and hv104=2
* Daughter-in-law: hv101=4 and hv104=2

use "...PKPR71FL.DTA", clear
describe hv101

* The label of hv101 is HV101
label list HV101

* How many pairs are there in the data?
tab hv219 if hv101==4 & hv104==2
* There are 623 pairs

lookfor education
* The education variables for all household members are hv106-hv109

* Construct a subfile of household heads with ID variables and education variables
keep if hv101==1
keep hv001 hv002 hv106-hv109
rename hv10* hv10*_head

* Merge the head's education variables onto every person in the household
merge 1:m hv001 hv002 using "...PKPR71FL.DTA"
tab _merge
drop _merge

* Identify women who are daughters-in-law of the household head
gen dtr_inlaw=0

* Daughter-in-law of male head
replace dtr_inlaw=1 if hv101==4 & hv104==2 & hv219==1


* Daughter-in-law of female head
replace dtr_inlaw=2 if hv101==4 & hv104==2 & hv219==2

label variable dtr_inlaw "Daughter in law of head"
label define dtr_inlaw 0 "No" 1 "Head is father in law" 2 "Head is mother in law"

tab dtr_inlaw

* The population of interest is cases with dtr_inlaw=2

* Simple comparison: crosstab of hv106 for the 623 pairs
```

tab hv106 hv106_head if dtr_inlaw==2