Subject: Merging old DHS variables from India NFHS 2 on IPUMS data for same period
Posted by preshit on Sun, 02 Jan 2022 04:51:55 GMT
View Forum Message <> Reply to Message

Hello everyone,
For my study, I m using all NFHS cycles so I have downloaded relevant harmonized variables from IPUMS and now trying to merge a few variables from respective DHS cycles. My unit of analysis is individual births so my objective is to merge old DHS variables such m15 (place of delivery), m14(ANC), m19(weight at birth) from the BR files to the IPUMS extracts. As of now, these variables are not available in IPUMS. Therefore, I have to go with the merging route. I am attempting the merge for each cycle separately starting with the NFHS-2. For this purpose, I have filtered out NFHS-2 data from IPUMS all NFHS extract. I am working with the IABR42.dta file which is the NFHS-2 birth recode file.

Attempt-1:
I have reviewed previous answers posted to similar questions like here:
https://www.idhsdata.org/idhs/linking.shtml. This webpage does not provide merge examples for birth recode files but I assume using only "idhspid" for birth file merging will not suffice as each respondent will have multiple births noted in the birth data. Unfortunately, NFHS 2 cycle does not contain the B16 variable in both datasets which is the line number for the child in the BR file. Instead, I used the BIDX variable available in both the data and attempted the 1:1 merge. However, I am getting the following results and observations are not getting merged:


sort idhspid bidx

merge 1:1 idhspid bidx using "IABR42_temp2.dta"

```
Result                  # of obs.
    -----------------------------------------
    not matched              537,758
        from master          268,879  (_merge==1)
        from using           268,879  (_merge==2)

    matched                        0  (_merge==3)
    -----------------------------------------
```


Attempt-2:
Suspecting that the variable structure might be different, I have recasted all the required variables, also making sure that their variable and value labels and storage type, etc. are exactly the same. But still, I could not get the observations matching. As a slight variation, I also tried the merge by including state-level identifier "geo_ia1992_2015" by creating a matching variable in old DHS but still could not succeed.

Attempt-3:
Out of curiosity, I downloaded matching variables in separate excel files and compared them with

each other. They look identical but I received "FALSE" results when I applied matching conditionality for comparison. As a last resort, in my Stata datasets, I created a numeric identifier as follows and used it for merging.

egen numindex = group(geo_ia1992_2015 idhspid bidx)

by sorting and using the above unique numeric identifier I was able to perform a 1:1 merge.


My questions are:
1. I am suspecting despite creating exact similar identifiers in both datasets something prevented them to match, maybe due to the differences in string structure?
2. Is my strategy acceptable to use BIDX in lieu of B16 for merging birth files?
3. Is my strategy acceptable to create a unique numeric id and perform the 1:1 merge to circumvent the issue?

Does anyone have faced similar challenges in merging similar datasets? Any insights will be highly appreciated. I wasn't able to log in to the IPUMS user forum using my DHS credentials, therefore I am posting my question here. Thank you.


Regards
Preshit