Subject: A n Urgent Request: how to merge data files
Posted by Jackie N on Wed, 27 Mar 2024 04:21:27 GMT
View Forum Message <> Reply to Message

Hello,

My name is Jackie and I am writing a thesis on the impact of the universal primary education policy (effective since 1997) on intergenerational education mobility. My paper is due on April 8.

The DHS surveys conveniently collect all the controls and the independent (parent's education and child's birth year) and dependent variables (child's education) I am interested in.

I ran into some trouble while merging the DHS 2016 data from Uganda. I have more than 65% NA values for my mother variables and more than 88% for my father variables. These are the steps I took:

1) I merged the women's survey with the household member survey. And did the same for men.

2) Then, using the merges above, I called individuals who answered HV101 with 1 or 2 parents and those with 3 or 11 children. HV101 = "relationship to head" where 1 = head, 2 = spouse, 3 = son/ daughter and 11 = adopted / fostered

3) Step #2 created 4 dataframes. I stacked the female children and male children and added their parents' information as columns. So I completely changed the structure and thus the unit of analysis of the original datasets. In my final dataset, the unit of analysis is children.

I am wondering if you have any ideas of why I have so much data missing. The merges were all successful and I am using R. Another worry I have is about my definition of parents: one could be the head of the household but not a parent. However, if I make it such that both conditions have to be true - so HV101 = 1 and HV101 = 2 - I might excluded single-parent households.

I have attached my code down below. I would greatly appreciate any suggestions and ideas you have. One I have right now is to use the couple's recode file and just attach the children's information.

Thank you very much for your time and consideration.

Sincerely,

Jackie Namala.

Subject: Re: A n Urgent Request: how to merge data files
Posted by Janet-DHS on Thu, 28 Mar 2024 19:15:51 GMT
View Forum Message <> Reply to Message

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

You have to use hv112 and hv114 to match children with their mothers and fathers. You can't do it with hv101.

Below I will paste a Stata program to do this. I don't use R but hope you can convert to R. It could be more efficient but it runs very quickly. This constructs a new file, with the variables renamed to have suffix "child", "mother" and "father". The cases are children age 0-17. The education variables are hv106-hv109 and hv121-hv124. A 2x2 table is produced describing the matches. Hope this helps.

```
* specify a workspace
cd e:\DHS\DHS_data\scratch

* The education variables are hv106-hv109, hv121-hv124

* Make a file of potential mothers
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\UGPR7BFL.DTA", clear
keep if hv104==2
drop if hv105<=17
keep hv*
rename hv001 cluster
rename hv002 hh
rename hvidx mo_line
rename hv114 fa_line
rename hv* hv*_mother
save mother.dta, replace
* Make a file of potential fathers
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\UGPR7BFL.DTA", clear
keep if hv104==1
drop if hv105<=17
keep hv*
rename hv001 cluster
rename hv002 hh
rename hvidx fa_line
rename hv* hv*_father
```

```
save father.dta, replace

* Make a file of children age 0-17
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\UGPR7BFL.DTA", clear
keep if hv105<=17
keep hv*
rename hv001 cluster
rename hv002 hh
rename hv112 mo_line
rename hv114 fa_line
rename hv* hv*_child

gen child=1
tab child
label define noyes 1 "No" 3 "Yes"
quietly merge m:1 cluster hh mo_line using mother.dta
rename _merge mother_matched
label values mother_matched noyes
keep if child==1
quietly merge m:1 cluster hh fa_line using father.dta
rename _merge father_matched
label values father_matched noyes
keep if child==1

tab *matched

. tab *matched

mother_mat |    father_matched
    ched |     No      Yes |   Total
-----------+----------------------+----------
      No |   10,095     3,095 |   13,190
     Yes |   10,828    26,454 |   37,282
-----------+----------------------+----------
   Total |   20,923    29,549 |   50,472

* save this file
```

Hi Tom and Janet,

Thank you both very much for your response. This is critically illuminating. I need the children to be at least 21 in 2016 because that way, they would have had an opportunity to pursue primary-tertiary education. It is just easier to draw conclusions when they've, theoretically,

"finished" school rather than when they are still going through the system.

Hopefully, this results in a sufficient sample size.

Much thanks once again,

Jackie.

Subject: Re: A n Urgent Request: how to merge data files
Posted by Janet-DHS on Tue, 02 Apr 2024 13:50:36 GMT
View Forum Message <> Reply to Message

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

Links to the mother and father with hv112 and hv114 are only available for children age 0-17 who are living in the same household as the parent(s). 0-17 is the definitional age range for children, established by UNICEF. Your age range of interest may be more like 15-24, or "youth".

A common pattern in parts of sub-Saharan Africa (I'm not sure about Uganda, specifically) is that children who do not have good access to secondary school locally may be fostered with kin who live in an area with better access. When this happens, and the child is not living with the parents, hv112 and hv114 will be NA (not applicable, a dot in Stata). If a child goes to a residential university, they will not be in the household population, and they will not be eligible to appear in a DHS survey.

Some studies have been done in which the education level of the household head, regardless of whether that person is a parent, is used as a predictor of the child's education.

You seem to be focused on whether the child proceeds to post-secondary education, but it will be easier if you shift to the transition to secondary education, which typically occurs before age 18.