Subject: Adding wealth index to pre 2003 files
Posted by antrabhatt on Mon, 27 Mar 2023 21:14:30 GMT
View Forum Message <> Reply to Message

Dear Colleagues,

Does anyone have experience with merging the wealth index into the IR file for old DHS surveys. How do you identify and isolate the hv001 and hv002 from 'whhid'?

Thanks,
Antra

Subject: Re: Adding wealth index to pre 2003 files
Posted by Janet-DHS on Thu, 30 Mar 2023 19:50:18 GMT
View Forum Message <> Reply to Message

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

Which survey are you using? This has to be done separately for each survey because you can't be sure which columns of whhid are assigned to hv001 and hv002 without unpacking whhid. Examples have been given before but I can show the strategy if you specify one such survey.

Subject: Re: Adding wealth index to pre 2003 files
Posted by antrabhatt on Thu, 30 Mar 2023 20:05:10 GMT
View Forum Message <> Reply to Message

Hi, I am doing this exercise for 20 different countries so wanted to know what's the best way to recognize the strategy for isolating and understanding which columns of whhid are assigned to hv001 and hv002. The example we can take is Burkina Faso 1998 survey. The full list of countries is Benin, Burkina Faso, Colombia, Ethiopia, Ghana, Guatemala,
Haiti, Indonesia, Jordan, Malawi, Mali, Mauritania, Nigeria, Philippines, Rwanda, South Africa, Tanzania, Uganda and Zambia

Subject: Re: Adding wealth index to pre 2003 files
Posted by Janet-DHS on Fri, 31 Mar 2023 18:39:52 GMT
View Forum Message <> Reply to Message

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

The Burkina Faso 1998 survey is difficult to work with. For most surveys, you can unpack hhid and get hv001 hv002, but not this survey. However, you don't NEED to do that. The HR and PR files (e.g. BFPR31FL.dta) include hhid and the WI file (BRWI31FL.dta) includes whhid. In each file, enter "list *hhid if _n<=50, table clean" and you will see that they are the same. With "describe

*hhid" you see that both are 12-character strings. So you just need to rename one of them, sort both files, and merge. I'm sure there are other surveys on your list for which you can do the same and don't need to unpack hhid, especially to merge with the PR file.

The Ethiopia 2000 survey (ET41) is more typical. You may not actually need to unpack the columns of hhid in the PR file, but here's how you would do it:

* How to unpack an ID string, illustrated for hhid in the PR file for Ethiopia 2000, ET41

use "...ETPR41FL.DTA"

describe hhid
* We see that hhid is a 12-character string.

list hhid hv001 hv002 if _n<=50, table clean
* We see that hhid contains hv001 and hv002

* Check how many columns are needed for hv001 and hv002
summarize hv001 hv002
* We see that the max value of hv001 requires 3 digits, as does hv002

* Unpack the columns of hhid
forvalues lc=1/12 {
gen col`lc'=substr(hhid,`lc',1)
}

* List and inspect
list hhid col* hv001 hv002 if _n<20, table clean
* we see that hv001 is cols 7-9 of hhid and hv002 is cols 10-12

* Extract these substrings
gen hv001_test=substr(hhid,7,3)
gen hv002_test=substr(hhid,10,3)

* Convert from strings to numeric
destring *_test, replace

* Confirm
correlate hv001*
correlate hv002*

You then unpack whhid in the WI file to get hv001 and hv002 and then merge with the PR file to get hv270 and hv271. Or you can call the extracted variables v001 and v002 and merge with the IR file to get v190 and v191. Hope this is clear

There are other situations in which you have to unpack caseid, which includes line number.

If your goal is really to get the wealth index into the IR file, you can probably skip unpacking hhid by merging the PR and WI files using hhid and then merging with the IR file using hv001 hv002 hvidx in the PR file and v001 v002 v003 in the IR file.