Subject: Unique ID Error
Posted by l.kincaide on Tue, 30 Jan 2018 09:28:27 GMT
View Forum Message <> Reply to Message

Hello,
I am using STATA to work with DHS data from all countries in sub-Saharan Africa. The variables that the DHS says uniquely identify observations (v001 v002 v003 for the individual recode and mv001 mv002 mv003 for the men's recode) apparently do not uniquely identify observations for Cameroon Round 3, Chad Round 3, Mali Round 3, Mali Round 4, Niger Round 3, Senegal Round 4, and Togo Round 3. The command I am using to check this is "isid". Has anyone else had this problem and know a way around it?
Thanks a lot!

Subject: Re: Unique ID Error
Posted by Bridgette-DHS on Tue, 30 Jan 2018 17:32:18 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

The following Stata lines illustrate how to do this merge for the Mali 2001 survey. These surveys contain a variable that gives a sub-household code but the general approach would be to unpack hhid and caseid to get that code, which I call "subhh". Since hhid and caseid are not always constructed in exactly the same way, you have to be careful that you pull out (using "substr") the correct columns.

The surveys you are looking used a definition of households that is different from the standard that is now used. I don't believe that any surveys after DHS-4 have this issue.

* Strategy to find how hhid and caseid match up in a survey that includes
*   sub-household numbers, call them subhh

use e:\DHS\DHS_data\PR_files\MLPR41FL.dta, clear
keep if hvidx==1

* We need to find the specific columns of hhid that match with hv001, hv002, and subhh
local li=1
while `li'<=12 {
gen hhid_`li'=substr(hhid,`li',1)
destring hhid_`li', replace
local li=`li'+1
}

list hhid hv001 hv002 hhid_* if _n<=10, table clean
* Apparently hv001 is columns 5-7 of hhid,
*   hv002 is columns 11-12, and

```
*   subhh is columns 8-10

* Confirm the correspondence for hv001 and hv002
gen hv001_test=substr(hhid,5,3)
gen hv002_test=substr(hhid,11,2)
destring *test, replace
correlate hv001*
correlate hv002*

* I no longer need the extracted values of hv001 and hv002 but I do need subhh

* Prepare the PR file for merge
use e:\DHS\DHS_data\PR_files\MLPR41FL.dta, clear
gen subhh=substr(hhid,8,3)
destring subhh, replace
sort hv001 hv002 subhh hvidx
save e:\DHS\DHS_data\scratch\MLPR_temp.dta, replace

* Prepare the IR file for merge
use e:\DHS\DHS_data\IR_files\MLIR41FL.dta, clear
summarize v001 v002
list caseid v001 v002 v003 if _n<=10, table clean
codebook caseid

* Figure out how to destring caseid
local li=1
while `li'<=15 {
gen caseid_`li'=substr(caseid,`li',1)
destring caseid_`li', replace
local li=`li'+1
}

list caseid v001 v002 v003 caseid_* if _n<=10, table clean
* Apparently v001 is columns 5-7 of caseid,
*  v002 is columns 11-12, and
*  v002 is columns 14-15, and
*  subhh is columns 8-10

* Confirm the correspondence for v001, v002, and v003
gen v001_test=substr(caseid,5,3)
gen v002_test=substr(caseid,11,2)
gen v003_test=substr(caseid,14,2)
destring *test, replace
correlate v001*
correlate v002*
correlate v003*
drop *test
```

```
gen subhh=substr(caseid,8,3)
destring subhh,replace
rename v001 hv001
rename v002 hv002
rename v003 hvidx
sort hv001 hv002 subhh hvidx

merge hv001 hv002 subhh hvidx using e:\DHS\DHS_data\scratch\MLPR_temp.dta
tab _merge
```

Subject: Re: Unique ID Error
Posted by kehinde.atoloye@gmail.com on Wed, 19 Oct 2022 09:13:36 GMT
View Forum Message <> Reply to Message

DHS-7 Nigeria has this problem. Thankfully, your code did a great job as I have been struggling to do this for child and household member recode. I successfully merged both files. Thanks a million.