Subject: Phase I surveys
Posted by rejone4@emory.edu on Tue, 11 Nov 2014 16:34:47 GMT
View Forum Message <> Reply to Message

I'm working with a few surveys from phase I of the DHS and was wondering how to coded in STATA for weights. I'm specifically looking at Egypt 1988 right now. There is no v021 or v023/v024 variable for primary sampling unit or sample domain.

I've been coding other phases as the following:
gen psu = v021
gen strata = v023
gen weight1 = v005 / 1000000
svyset psu [pweight = weight1], strata(strata)

Any recommendations?

Subject: Re: Phase I surveys
Posted by Bridgette-DHS on Wed, 12 Nov 2014 15:16:35 GMT
View Forum Message <> Reply to Message

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

V001 and v021 are identical, and both are the cluster id. If v021 is missing, just use v001. (Personally, I always use v001 rather than v021.)

The strata are almost always the combinations of region and urban/rural. There are typically around 20 to 30 strata. Unfortunately, in many of the early surveys, the strata are not actually coded, and the variable that says "sample stratum number" is not what it says.

I just looked at the 1988 and 1992 surveys of Egypt. Your interest is mainly in 1988 but you have to do some recoding for 1992 as well.

In the 1992 survey, v023 (domains) and v024 (regions) are exactly the same. They take only 5 values. The alleged stratum variable (v022) has 178 different values--too many to be the correct stratum variable. It consists mostly of groups of 2 or 3 clusters.

The key to the strata is found by looking at the pattern of variation in the weights. I quickly found that v005 takes only 21 different values, corresponding to groupings of values of v022. These are the strata. The clusters are numbered consecutively within strata. To construct the range of codes for v022 that go into each of the 21 strata, you could use the following lines in Stata:

use ....EGIR21FL.DTA , clear
collapse (mean) v005, by(v022)
list, table clean
egen stratum=group(v005)
gen v022_min=v022

```
gen v022_max=v022
sort stratum v022
collapse (first) v022_min (last) v022_max, by(stratum)
list stratum v022_min v022_max, table clean
```

To repeat, this will give you the information to construct a stratum id variable from v022.

In the 1988 survey, you should use v001 in place of v021.

I see that in the 1988 survey v001 has 454 values, ranging from 1005 to 228008. This must be a code for the enumeration areas that was not converted to integers 1 through 454. I see that v005 has only 7 values. These must correspond to the 7 strata. Again, you need to work out a recode of v001 to construct the stratum id variable. It is easily found that the clusters that go into the strata are not just ranges. You will have to work out the recode; the following lines will help.

```
use ....EGIR01FL.DTA, clear
tab v001 v005
```