Subject: Correct Way to Do Sample Weighting for Cross-Sectional Pooled Data from Kenya
Posted by subanara on Sun, 18 Mar 2018 17:30:14 GMT
View Forum Message <> Reply to Message

Hi,

I am using a pooled sample of the Kenya 2003, 2008-09, and 2014 datasets for only women who answered the domestic violence module. So far I have done my regressions unweighted, but I know that this likely causes error towards the samples in 2014 (since the total sample size is much larger in this dataset) and have used a control variable indicating the interview year to control for interview variation.

I was wondering what the correct procedure (in STATA) to generate weights for this type of pooled sample, I'm assuming I use the dv005 weight.

Thanks so much!

Subject: Re: Correct Way to Do Sample Weighting for Cross-Sectional Pooled Data from Kenya
Posted by Bridgette-DHS on Tue, 20 Mar 2018 12:55:58 GMT
View Forum Message <> Reply to Message

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

Yes, you should use d005. Then I recommend that you give equal weight to each survey. There are various ways to do this, e.g. the following. Refer to the three surveys in the pooled file with survey=1, survey=2, survey=3. Calculate the sum of the DV weights in each survey as D1, D2, and D3, and the sum D=D1+D2+D3. Then in the first survey calculate the revised d005 as d005r=(d005/D1)*(D/3). In the second survey, d005r=(d005/D2)*(D/3). In the third survey, d005r=(d005/D3)*(D/3).

The following lines will accomplish this (change the paths, of course):

```
set more off
use e:\DHS\DHS_data\IR_files\KEIR42FL.dta, clear
keep caseid v0* d*
gen survey=1

save e:\DHS\DHS_data\scratch\KEdvtemp.dta, replace

append using e:\DHS\DHS_data\IR_files\KEIR52FL.dta
keep caseid v0* d* survey
replace survey=2 if survey==.

append using e:\DHS\DHS_data\IR_files\KEIR70FL.dta
```

```
keep caseid v0* d* survey
replace survey=3 if survey==.

drop if d005==.

summarize d005 if survey==1
scalar D1=r(N)*r(mean)

summarize d005 if survey==2
scalar D2=r(N)*r(mean)

summarize d005 if survey==3
scalar D3=r(N)*r(mean)

scalar D=D1+D2+D3
scalar list D1 D2 D3 D

gen d005r=d005*(D/3)

replace d005r=d005r/D1 if survey==1
replace d005r=d005r/D2 if survey==2
replace d005r=d005r/D3 if survey==3

summarize d005*

save, replace
```