Subject: PSU and strata identifiers - Egypt
Posted by Nora on Thu, 07 Jul 2016 17:28:18 GMT
View Forum Message <> Reply to Message

I am trying to use the svyset option to weight my data. I am using pooled data for Egypt for years 1995,2005 and 2014 and restricting my sample to women who participated in the domestic violence module.

1.I am not sure what variable to specify in the PSU and strata. As v021 is labelled PSU and v022 is labelled as sample strata for the three surveys, would this syntax be correct?

svyset [pw= domestic violence wgt], psu(v021) strata(v022)

What does this result mean:
pweight: wt_dv
VCE: linearized
Single unit: missing
Strata 1: v022
SU 1: v021
FPC 1: <zero>

2.Also, is it okay to run this code with pooled data from three surveys?

Thank you.

Subject: Re: PSU and strata identifiers - Egypt
Posted by Bridgette-DHS on Fri, 08 Jul 2016 11:57:10 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Specialists, Tom Pullum and Trevor Croft:

To be clear, the files are EGIR33 (the 1995 survey), EGIR51 (the 2005 survey), and EGIR61 (the 2014 survey).

In most DHS surveys, PSUs and clusters are the same, so either v021 or v001 can be used as the cluster id. However, for most surveys of Egypt, including the three you are using, there are two clusters per PSU. You should use v021 as the PSU id, as you are doing, but with the following modification for the first survey. For that survey, v021 is incorrect and should be reconstructed with "v021 = int(v001/10000)". After that change, you can use v021 as the id for PSU in all three surveys.

In all three surveys the sample strata are urban and rural areas within each governorate. That must be constructed in the first two surveys. For example, in those two surveys you could use these two lines: "rename v023 v023_original" and "egen v023 = group(v023_original v025)". For the third survey, just use v023 as it is.

We would never recommend that you replace the original file with revised variables. If you make changes using original variable names, as just suggested, you should save with another file name or you should just put those changes in a program and not save the data file.

Yes, you can pool the surveys into a single file. You can include a variable "survey" or you can distinguish them with v000. In the three surveys, v000 is coded EG3, EG4, and EG6, respectively. You could reduce the files to just the variables you need before appending them.

You could get unique identifiers in the pooled file, after you have reconstructed v021 and v023 as described above, with "egen cluster=group(v000 v021)" and "egen stratum=group(v000 v023)" and then "svyset cluster [pweight=wt_dv], strata(stratum) singleunit(centered)" . There have been many postings on possible rescaling of the weights when surveys are pooled. In this case I would leave the weights alone, because I assume (and hope) you are really just looking at changes between surveys rather than, say calculating means or percentages that combine all three surveys.

Subject: Re: PSU and strata identifiers - Egypt
Posted by Nora on Wed, 13 Jul 2016 21:38:08 GMT
View Forum Message <> Reply to Message

Thank you for clarifying the steps.

Just for confirmation, doing then these steps then running the regression as follows:
svy: reg y x1 x2 x3 for each year should be enough without the need to add an option of clusters, am I correct?

Also, I would like to ask if there is a method to draw a ROC curve as well as a test for weak instruments that can be combined with the svy command for binary response models as it seems that svy cannot be combined with many tests.

Thank you.

Subject: Re: PSU and strata identifiers - Egypt
Posted by Nora on Thu, 14 Jul 2016 05:23:15 GMT
View Forum Message <> Reply to Message

Similarly, is there an alternative way for using svy: ivreg2 as ivreg2 doesn't seem to be compatible with the svy command.

Thank you.

Subject: Re: PSU and strata identifiers - Egypt
Posted by Bridgette-DHS on Thu, 14 Jul 2016 16:40:12 GMT
View Forum Message <> Reply to Message

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

Yes, that command would work if you have already specified the clusters in the svyset command. Your other two questions will have to be answered by other forum users.

Subject: Re: PSU and strata identifiers - Egypt
Posted by laurajbrown@exwh.org on Fri, 05 Oct 2018 14:35:07 GMT
View Forum Message <> Reply to Message

Hi DHS users,

I have a similar issue to Nora but for the 1988 survey. I am unsure how I should add weights in this instance. For most other DHS surveys I am analysing, I use the code below in Stata:

generate wgt = v005/1000000
svyset[pw=wgt], psu(v021) strata(v022) singleunit(centered)

However neither the 1988 Egypt IR file nor the HH file contain suitable variables for psu or strata. Using the replies to Nora's post I figured I could construct the psu as follows:
gen v021=int(v001/10000)

but I am stumped on the strata variable. Is it based on urban and rural in this survey as previous replies to this thread would suggest? And if so, do I just use the v102 "type of place of residence" variable to construct it like so:
gen v022=v102

making my final weighting code as follows:

generate wgt = v005/1000000
gen v021=int(v001/10000)
gen v022=v102
svyset[pw=wgt], psu(v021) strata(v022) singleunit(centered)

Or am I missing something?

Thanks for your help!

Laura

Subject: Re: PSU and strata identifiers - Egypt
Posted by Bridgette-DHS on Wed, 17 Oct 2018 14:25:30 GMT
View Forum Message <> Reply to Message

Following is a response from Senior Sampling Specialist, Mahmoud Elkasabi:

In general, you need to check the sampling design appendix to figure out the stratification variables and then construct it in a proper way.

Regarding the 1988 EDHS survey, see page 174 in the final report: "All list of PSUs allocated according to governorate and residential sector (urban/rural)" This means that similar to the other surveys, governorates by urban/rural were used as design strata for the 1988 EDHS.