Subject: Clarification on Variables for svyset in STATA and generating stunting variable
Posted by 616blue on Wed, 29 Apr 2015 17:18:54 GMT
View Forum Message <> Reply to Message

Hello,

I've been through numerous threads on these two topics and would appreciate some additional clarification for my project.

#1. For using svyset in STATA, most threads note we need a psu, pweight and strata. There is consensus on psu=v021.
For strata, generally we would use group (v024, v025)-but this would differ by country what the strata is (and we would NOT use v022 in general). Where could I check what the strata details are specifically for my country of interest (ETHIOPIA)?
For sampwt, I've read both that we should use v005 directly/ use v005/1000000. Could I receive clarification on what I should use and why?

I have:
```
gen psu =    v021
egen strata = group(v024 v025), label
tab strata
gen sampwt = v005/1000000 //as per DHS  instruction//
svyset psu [pweight = sampwt], strata(strata)
```

#2. For generating the stunting variable, I previously used the coding:
```
codebook hw70
tab hw70 if hw70>9990,m
tab hw70 if hw70>9990,m nolabel
gen HAZ=hw70
replace HAZ=. if HAZ>=9996
histogram HAZ
gen stunted=.
replace stunted=0 if HAZ ~=.
replace stunted=1 if HAZ <-200
tab stunted
regress stunted
regress stunted [pweight=v005]
```

However, I found a response at: http://userforum.dhsprogram.com/index.php?t=tree&goto=3952&S=09ed0022cd4173b993b147e4fdc88183&srch=svy+stata #msg_3952
Re: svy, subpop [message #3952 is a reply to message #3942]
that suggests differently:

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

Here are the lines you need for a logit regression of stunting on the wealth index for children age 0-23 months.

```
* logit regression with stunting in months 0-23 as outcome
* use the PR file; KR file is limited to children living with mother
* usually limit to de facto residents, i.e. hv103=1
keep if hv103==1
* hc70 is the WHO haz score, already edited
gen stunting=0
replace stunting=1 if hc70<-2
replace stunting=. if hc70>600
* the cluster id is hv001=hv021
* stata will normalize the weights; no need to divide by 1000000
* hv022 may be the strata variable but always check
* always a good idea to include one of the singleunit options
svyset hv001 [pweight=hv005], strata(hv022) singleunit(centered)
* it is normal to use hc1 (=hv008-hc18) as age
svy: logit stunting i.hv270 if hc1<24
```

Questions are as follows:
Why would I use the PR file vs the BR file? Or asked differently, if I am interested in the mother's education on child's stunting outcomes, would it be correct to use the BR file-since it has the birth data for each child and mother info? Or would it even be the KR file?

Why would we limit to de facto residents? Is the svyset command not controlling for this?

Why would we use hc70<-2 vs hc<-200? I think on different threads I found that we need to multiply by 100.

I assume replace "stunting=. if hc70>600" is equivalent to "replace HAZ=. if HAZ>=9996" since the range for plausible values is up to 600 and anything else beyond is coded in the 9000's to indicate missing etc.?

The following four lines confuse me.

```
* the cluster id is hv001=hv021
* stata will normalize the weights; no need to divide by 1000000
* hv022 may be the strata variable but always check
* always a good idea to include one of the singleunit options
svyset hv001 [pweight=hv005], strata(hv022) singleunit(centered)
```

Why is the psu, weight and strata different from what is used in other examples, as illustrated in #1?
In what combination do we divide weights by 1000000 or not divide?

Thank you in advance!