Subject: variables for svyset in Stata (Bangladesh 2011)
Posted by kmorris on Fri, 09 Jan 2015 18:31:31 GMT
View Forum Message <> Reply to Message

Hello,
I am currently trying to analyze c-section by wealth quintile (lower 2 and upper 3 as their own groupings).

I am using Stata 12, and when trying to set svyset for the data, I have been encountering an issue with the strata that is affecting my results in analysis.

```
gen birth5=0
replace birth5=1 if v208>0
label var birth5 "have a live birth in past 5 years"
label define yesno 0 "no" 1 "yes"
label values birth5 yesno
keep if birth5 == 1

gen wt=v005/1000000

gen csect = . if m17 == .
replace csect = 1 if m17 == 1
replace csect = 0 if m17 == 0

label variable csect "C-Section"
label define csect 0 "no" 1 "yes"
label value csect csect

gen psu=v021

gen strata=v023
**note, I generated the weight earlier, above)**

svyset psu [pweight=wt], strata(strata)

**now analyzing using poisson for the lower 2 wealth quintiles**
svy: poisson csect v190 if v190<3
```

And this is where I run into errors, see results as written below
(running poisson on estimation sample)

```
Survey: Poisson regression

Number of strata   =        19              Number of obs      =      3645
Number of PSUs     =       499              Population size    = 3867.0893
                                            Design df          =       480
                                            F(   0,    480)    =         .
```

```
                                        Prob > F          =          .

------------------------------------------------------------------------------
             |             Linearized
       csect |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        v190 |   1.011505          .        .       .            .           .
       _cons |  -4.721347          .        .       .            .           .
------------------------------------------------------------------------------
Note: missing standard errors because of stratum with single sampling unit.
```

Did I do something wrong with the sampling unit? Is this error unique to this dataset or will I have to change my syntax?

Thank you!
Kate Morris

Subject: Re: variables for svyset in Stata (Bangladesh 2011)
Posted by Reduced-For(u)m on Fri, 09 Jan 2015 21:17:56 GMT
View Forum Message <> Reply to Message

I think "v023" is the wrong stratification variable. Strata are defined in v022 according to the recode manual - http://dhsprogram.com/pubs/pdf/DHSG4/Recode5DHS_23August2012 .pdf

You can also see in the final report that stratification was done by urban/rural and then by 7 administrative divisions, so you could generate your own stratification variable and then check it against v023 to confirm (see page 5 of the introduction).

http://dhsprogram.com/pubs/pdf/FR265/FR265.pdf

Subject: Re: variables for svyset in Stata (Bangladesh 2011)
Posted by mmr-UMICH on Tue, 03 Feb 2015 18:49:57 GMT
View Forum Message <> Reply to Message

To avoid such an error/warning :"Note: missing standard errors because of stratum with single sampling unit.", we can use singleunit() in svyset command line that specifies how to handle strata with one sampling unit. By default, svyset uses singleunit(missing) that results in missing values for the standard errors. I usually use to handle such situations by using method singleunit(centered) that specifies that strata with one sampling unit are centered at the grand

mean instead of the stratum mean.
e.g., for your code: svyset psu [pweight=wt], strata(strata) singleunit(centered)
Other methods are:
singleunit(certainty) causes strata with single sampling units to be treated as certainty units. Certainty units contribute nothing to the standard error.
singleunit(scaled) results in a scaled version of singleunit(certainty). The scaling factor comes from using the average of the variances from the strata with multiple sampling units for each stratum with one sampling unit.

Thanks,

Subject: Re: variables for svyset in Stata (Bangladesh 2011)
Posted by mmr-UMICH on Tue, 03 Feb 2015 18:50:32 GMT
View Forum Message <> Reply to Message

To avoid such an error/warning :"Note: missing standard errors because of stratum with single sampling unit.", we can use singleunit() in svyset command line that specifies how to handle strata with one sampling unit. By default, svyset uses singleunit(missing) that results in missing values for the standard errors. I usually use to handle such situations by using method singleunit(centered) that specifies that strata with one sampling unit are centered at the grand mean instead of the stratum mean.
e.g., for your code: svyset psu [pweight=wt], strata(strata) singleunit(centered)
Other methods are:
singleunit(certainty) causes strata with single sampling units to be treated as certainty units. Certainty units contribute nothing to the standard error.
singleunit(scaled) results in a scaled version of singleunit(certainty). The scaling factor comes from using the average of the variances from the strata with multiple sampling units for each stratum with one sampling unit.

Thanks,

Subject: Re: variables for svyset in Stata (Bangladesh 2011)
Posted by Trevor-DHS on Wed, 04 Feb 2015 00:57:05 GMT
View Forum Message <> Reply to Message

I want to correct a few miscomprehensions here:
1) For this survey v022 and v023 are identical, so it doesn't mater which you use.
2) The strata for this survey are not just urban/rural within the 7 administrative divisions, but are actually 3 separate groups within each division:
a) Urban city corporations
b) Other urban areas
c) Rural areas
There are a total of 20 strata as there is no city corporation strata for Rangpur.

3) While using the singleunit(centered) parameter is one way around the problem of a stratum with a single unit, in general we recommend to regroup the strata with a similar strata and not use the singleunit parameter. In this case I would regroup strata 5 (Rajshahi city corp.) which only has the one cluster (after the selection for this estimation) with strata 11 (Rajshahi other urban). The difference in the results between the two approaches however will be tiny.

Subject: Re: variables for svyset in Stata (Bangladesh 2011)
Posted by mmr-UMICH on Wed, 04 Feb 2015 03:58:29 GMT
View Forum Message <> Reply to Message

Thank you, Trevor.

I will cut and paste some lines whenever requires from original message and try to clarify concretely the reasons behind that error/warnings and its possible solution:

now analyzing using poisson for the lower 2 wealth quintiles
svy: poisson csect v190 if v190<3

The above svy: command is not recommended [and not correctly handle the domain concept] as "if v190<3" in svy: subsets the data (i.e., also deleting design information [aside: full sample design information is important for correctly calculating the sampling errors]) prior to run the poisson regression.
This svy: command uses such a subset data and as a result analysis sample (i.e. "estimation sample" in Stata wording) lacks one strata and 101 PSUs (see below output (cut and paste):
----------start----
(running poisson on estimation sample)

Survey: Poisson regression

Number of strata = 19 Number of obs = 3645
Number of PSUs = 499 Population size = 3867.0893
Design df = 480
-------- end ------

We have to create a variable, say: mydomain = 1 if v190 < 3, otherwise, mydomain = 0, then use svy command:
svy, subpop(mydomain): poisson csect v190

I hope this run will not encounter such issue and also does not require singleunit(centered)* svyset option. And the output will show the same # of obs and population size, but changed others stats such as # of strata, PSUs and degrees of freedom (df).

I verified that strata 5 and 11 have 5 and 23 PSUs respectively; so this svy, subpop(): that form "analytic" domain/subpopulation/subgroup will not be an issue of singleton-strata from full sample data.

*note that singleunit(method) is kind of practically recommended for "analytic" subgroup and/or subclass analysis which sometimes encounter singleton-strata. This specification also appropriately calculates the degrees of freedom, which is prim important for statistical inferences, e.g, confidence intervals and p-values estimation.

Thank you all again.

Moshiur Rahman

Subject: Re: variables for svyset in Stata (Bangladesh 2011)
Posted by Trevor-DHS on Wed, 04 Feb 2015 13:09:24 GMT
View Forum Message <> Reply to Message

Good point, Moshiur, I should have mentioned the subpop.
You can also write it without using another variable, as follows:
svy, subpop(if v190 < 3): poisson csect v190
