Subject: Sample weight/Survey design
Posted by kusum on Tue, 24 Dec 2013 20:13:33 GMT
View Forum Message <> Reply to Message

Hi all,

I am using Nepal DHS 2011 dataset (child file) for a class project to examine the association between caste group and childhood stunting in Nepal.

To account for the survey design, I used the following codes after referring to the DHS notes.

gen finalwt= v005/1000000
svyset, clear
svyset v001 [pweight=finalwt], strata (v022)

v001 is cluster-- enumeration area (ward in rural, subward in urban)
v022 is the domain (13 ecoregions) by urban/rural (25 total)

When I do run the analysis, the population size is much small (see below). I just wanted to confirm you that I am using the sample weight correctly. Perhaps someone has encountered similar problem?

Thanks,
Kusum

svy, subpop (sample2):logit stunting i.femage
(running logit on estimation sample)

Survey: Logistic regression

Number of strata   =   25   Number of obs   =   5306
Number of PSUs   =   289   Population size   = 5391.3722
                              Subpop. no. of obs =   1134
                              Subpop. size   = 958.14927
                              Design df   =   264
                              F(3,   262)   =   1.66
                              Prob > F   =   0.1769

```
-------------------------------------------------------------- ------------------
         |          Linearized
 stunting |    Coef.  Std. Err.    t   P>|t|   [95% Conf. Interval]
----------+--------------------------------------------- ------------------
  femage |
      1 | -.0338201  .5256869  -0.06  0.949  -1.068893  1.001253
      2 |  .2319218  .570629    0.41  0.685   -.8916414  1.355485
      3 |  .4512506  .5719637   0.79  0.431   -.6749405  1.577442
        |
   _cons | -.3237562  .5264081  -0.62  0.539  -1.360249  .7127364
```

Subject: Re: Sample weight/Survey design
Posted by Reduced-For(u)m on Thu, 26 Dec 2013 19:48:17 GMT
View Forum Message <> Reply to Message

Hey kusum,

  Two quick questions:

1 - why is v001 before the [pweight=weight] bit? The DHS FAQ lists this code (following) and looking at the svy help file for STATA it doesn't seem like it should be there.

DHS FAQ code: svyset [pweight=weight], psu(v021) strata(strata)

2 - what is sample2, the subsample you list? My guess is that is where all your observations are going, but I don't know if that was intentional or not. If you type "sum femage if sample2==1" how many obs do you get?

One other thing - what is "femage"? Just because this is a thing of mine - older mothers will have older children (on average, just because of when you measure them), older children are more likely to be stunted due to the HAZ loss over the first few years (you know - poverty), and so you could get a spurious correlation here if femage is maternal age (or a mechanical effect if femage is child's age and sample2 is females, but maybe that's the point).

Anyway, if this isn't helpful, maybe I can provide better help if I understand 1 and 2.

Subject: Re: Sample weight/Survey design
Posted by user-rhs on Fri, 07 Feb 2014 06:04:38 GMT
View Forum Message <> Reply to Message

Better late than never

Reduced-For(u)m wrote on Thu, 26 December 2013 14:481 - why is v001 before the [pweight=weight] bit? The DHS FAQ lists this code (following) and looking at the svy help file for STATA it doesn't seem like it should be there.

DHS FAQ code: svyset [pweight=weight], psu(v021) strata(strata)

What Kusum had in his/her -svyset- specification is correct. The primary sampling unit (in this case, the EA/cluster-->v001) is specified before the pweight. See Stata documentation for -svyset-: http://www.stata.com/help.cgi?svyset

kusum wrote on Tue, 24 December 2013 15:13When I do run the analysis, the population size is much small (see below). I just wanted to confirm you that I am using the sample weight correctly. Perhaps someone has encountered similar problem?

Thanks,
Kusum

svy, subpop (sample2):logit stunting i.femage
(running logit on estimation sample)

Survey: Logistic regression

Number of strata = 25 Number of obs = 5306
Number of PSUs = 289 Population size = 5391.3722

Not sure what "small" is in relation to the total # of children in this dataset, but the svy: logit you ran was done only on the subset of your dataset where "sample2" == 1. I agree with Reducedform that you should do a svy: tab sample2,count to see what the # should be for sample2==1 and check against your weighted pop'n size from the regression output. From what I can see of your -svyset- command, you have set it correctly.

An important thing to note is that the weighting sometimes causes the pop'n size from your regression to be lower than the # of obs'ns. For example, if people living in Kathmandu were overrepresented in your sample relative to actual proportion of pop'n living in Kathmandu, their sampling weights would probably be <1 whereas ppl living in underrepresented regions would probably have sampling weights >1. Therefore, if you have many people from Kathmandu in the subpop you're running the regression on, your pop'n size may be < the # of obs'ns.

HTH,
rhs

Subject: Re: Sample weight/Survey design
Posted by mutia on Wed, 12 Feb 2014 07:54:33 GMT
View Forum Message <> Reply to Message

Hi
I'm Mutia FroM Indonesia.

How to order samples did not change the results of the analysis with SPSS crosstab after weigted?

Thanks

Subject: Re: Sample weight/Survey design
Posted by Bridgette-DHS on Thu, 20 Feb 2014 18:59:30 GMT
View Forum Message <> Reply to Message

Mutia, your posting is unclear. Please give more details, and say specifically what the problem is.

Thanks,

Bridgette-DHS