Subject: Districts as cluster-level for multi-level model
Posted by dgodha on Sun, 23 Feb 2020 10:28:13 GMT
View Forum Message <> Reply to Message

Hello,

I will appreciate your expert guidance on my query. We usually use 'psu' as the cluster level in DHS data. In my case, the group size is too small if I use 'psu'.

```
Group Variable |     #Groups    Minimum    Average    Maximum
           psu |     25,063          1        3.1         16
```

Since NFHS-4 is representative at the district level and we have to anyway create a variable for the cluster-weight, I am wondering if it is possible to use district as the cluster-level. I tried changing my weighting command for psu to district but as you can see in the output, I don't get the p-values and CIs.

```
*Rescaling of weights
 gen wt=v005/1000000

*Level 1 weights using scaling method 1: New weights sum to district sample size
 gen sqw = wt*wt
 egen sumsqw = sum(sqw), by(sdistri)
 egen sumw = sum(wt), by(sdistri)
 gen pwt11 = wt*sumw/sumsqw

* Survey setting
 gen wt2=1
 svyset sdistri, weight(wt2) strata(v023) , singleunit(centered) || _n, weight(pwt11)

*Output
*******
Number of strata   =     2,509               Number of obs     =  1,538,126
Number of PSUs     =     2,509                 Population size   =  1,438,715
Subpop. no. obs    =     78,446
Subpop. size       =  73,653.12
Design df          =         0
F(   0,      0)    =         .
Prob > F           =         .


Linearized
 y          Coef.    Std. Err.      t      P>t     [95% Conf. Interval]

_cons     -1.585093   .0192937   -82.16       .           .          .

sdistri
var(_cons) .1527032   .0153514                      .          .
```

Note: 5 strata omitted because they contain no subpopulation members.
Note: Strata with single sampling unit centered at overall mean.

I am not sure what is going wrong and will appreciate any understanding.
Thank you
Deepali

Subject: Re: Districts as cluster-level for multi-level model
Posted by Bridgette-DHS on Tue, 24 Mar 2020 18:36:14 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The purpose of the svy adjustment is to compensate for the similarities of respondents within clusters, the under- and over-weighting of clusters, and stratification.

I would not recommend that you change to districts as the sampling units. The adjustments for clustering and sampling weights will be seriously thrown off.

The clusters, by definition, are the primary sampling units. If you shift to districts you will capture some if the intra-class correlation that goes into the svy calculation, but not nearly all of it, and the weighting adjustment, no matter how you do it, will be incorrect. The new weights would affect all the estimates and tabulations.

Subject: Re: Districts as cluster-level for multi-level model
Posted by dgodha on Wed, 25 Mar 2020 08:57:14 GMT
View Forum Message <> Reply to Message

Many thanks for your response.
I do have a follow-up question. If I don't use survey weights, then I can go ahead with using districts as clusters. Is that correct? I need to use districts because 85% of my PSUs have 5 or less observations.

Subject: Re: Districts as cluster-level for multi-level model
Posted by Bridgette-DHS on Mon, 30 Mar 2020 20:05:06 GMT
View Forum Message <> Reply to Message

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

If you ignore the weights entirely, your estimates won't mean anything. They will not be corrected for the under- and over-sampling in the survey design. They will not be unbiased estimates of population values.

I often ignore the survey design for a data quality assessment or for initial data exploration or for testing a program. However, if you want to do more than that, you need to use the weights to get unbiased estimates and use the clustering and stratification adjustments to get robust standard errors.

In other words, I recommend that you do not treat districts as clusters.
