Subject: SE calculation stratifications
Posted by Nahid on Mon, 23 Jan 2023 11:36:00 GMT
View Forum Message <> Reply to Message

Dear Expert,
In DHS datasets, hv022 and hv023 variables presented 'sample strata for sampling errors' and 'stratification used in sample design'. My understanding first one should use it while calculating SE, CV etc.

Later one (hv023), is clear how the samples were stratified.

Can you please explain a bit more about the technical/statistical reasons for SE stratification (hv022)? How is this stratification calculated?

Thanks, in advance!
Nahid

Subject: Re: SE calculation stratifications
Posted by Bridgette-DHS on Mon, 23 Jan 2023 13:19:31 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

The strata are constructed as part of the sample design, using characteristics of the sample clusters, which are usually census enumeration areas. Usually the strata are combinations of region and urban/rural residence (v024 x v025), but in some surveys a geographic structure with more detail than v024 is used. Before the sample is drawn, the allocation of clusters to strata is specified. For example, if it is estimated that 2% of the household population resides in a specific stratum, then it may be specified that 2% of the sample clusters will be in that stratum. (But often the allocation is different from the population distribution.)

Without stratification, the number of clusters selected in the different parts of the country would be random. Stratification (combined with weights) makes the sample more representative than it otherwise would be AND it reduces the potential variability across other (hypothetical) samples. Stratification tends to reduce the standard errors of the statistics calculated from the sample. This design effect is the opposite of what comes from using sample clusters. Because cases in the same cluster tend to be similar to one another, the cluster design tends to reduce the effective sample size, i.e. to increase the standard errors.

In the svy adjustment, the most important effect to include is the weights. When weights are included, the point estimates become unbiased. (But the weights tend to increase the standard errors, and for that reason some researchers do not use weights.) The adjustments for clusters and strata do not alter the point estimates. They only affect the estimates of the standard errors, but in opposite directions. I encourage users to occasionally run the same model repeatedly, with various combinations of the adjustments, to get a better sense of how much difference they make.

Sometimes they make a big difference. Sometimes they make surprisingly little difference, but we recommend always making the adjustments in your final models.

DHS samples are two-stage stratified cluster samples--you can look up more information about such samples.
