Subject: Confirming correct survey design
Posted by vbrar4 on Sun, 21 Aug 2022 23:54:57 GMT
View Forum Message <> Reply to Message

Once a survey design has been set, what would be the best course of action for determining if it has been set correctly for a dataset that has its final report in its volume 1 (volume 2 not released yet). I am working with the NFHS-5 dataset, more specifically the mens subset and recently set the svy design with the code "generate weight = mv005/1000000" & "svyset [pw=weight], psu(mv001) strata(mv022)" as recommened by experts on this forum. However, the final report seems to be lacking an appendix and i cannot seem to find any mention of the percise number of strata and PSU's (just a quick mention of 30198 PSU's at the beginning of the report). I wonder this because anytime i tabulate anything using the command "svy: tab predictorname" i am presented number of PSU's and strata in the top left corner of STATA but, the number of PSU's is stated to be 9102.

Am i making an error or approaching this the wrong way because i see no other way to confirm the survey design and am not getting anything consistent to the final report.

Subject: Re: Confirming correct survey design
Posted by vbrar4 on Tue, 23 Aug 2022 00:26:33 GMT
View Forum Message <> Reply to Message

Hi!

I'm not exactly sure how this helps me confirm it. Doing this still makes STATA output provide the exact same results (e.g., for a tabulation, logistic regression etc.,). It also did not change the number of strata and PSU's STATA mentions when any command is ran.

I have been running "generate weight = mv005/1000000" & "cluster_ID=group(mv024 mv001)" & "svyset [pw=weight], psu(cluster_ID) strata(mv022) singleunit(centered)".

I understand how this may be the "correct" method because clusters are numbered within states, but this method does not change any output and once again provides nothing close to the PSU's mentioned in the final report.

Subject: Re: Confirming correct survey design
Posted by Bridgette-DHS on Tue, 23 Aug 2022 13:37:17 GMT
View Forum Message <> Reply to Message

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

I made a mistake yesterday when I said that you needed to construct a new cluster id, and we are taking down that response to avoid confusion. v001 has already been constructed to be unique

across states. Using egen group, as I suggested, will not change the results (as you found out).

If you enter "codebook mv001 mv022" you will see that there are indeed 9,102 clusters (unique values of v001) and 2,681 strata (unique values of v022). Those numbers, which Stata repeats for any command with svy and your svyset, are correct.

I see that on page 2 of the final report the number of PSUs is given as 30,198, as you say, a number that does not match the data files. We will post an explanation of the difference between 9,102 and 30,198. However, what matters now is that there is nothing wrong with your svyset command.

Subject: Re: Confirming correct survey design
Posted by vbrar4 on Tue, 23 Aug 2022 17:13:11 GMT
View Forum Message <> Reply to Message

Thank you for clearing up this confusing! Where could I expect to see this explanation whenever it is posted?

Subject: Re: Confirming correct survey design
Posted by Bridgette-DHS on Thu, 25 Aug 2022 11:23:02 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum (Fred Arnold contributed to this response):

The NFHS-5 had a 15% subsample for men and for some topics in the survey of women. To achieve a representative 15% subsample, men were interviewed in alternate households in 30 percent of the randomly selected clusters. That's the reason why there were about 30,000 clusters in the entire survey, and that many distinct values of hv001 in the HR and PR files and v001 in the IR, KR, and BR files, but only about 9,000 distinct values of mv001 in the MR file. There should be no problem with svyset and svy, or with merges. This design is unusual but it makes sense as a strategy to get a 15% subsample in such a large survey.