Subject: Appending multiple waves of the NFHS annd PSU codes
Posted by Niranjana on Thu, 03 Oct 2019 11:41:53 GMT
View Forum Message <> Reply to Message

Hello,

I am working with the women's module of the Indian NFHS. I would like to create a repeated cross sectional dataset using all four rounds of the survey using STATA ie append all 4 datasets. I have pulled the individual datasets for all 4 rounds and have kept only the variables I am looking at and recoded them. I would like to confirm that if I append these datasets, it would be a sound step to get the cross sectional dataset I am looking for.

Also, what is the difference between PSU numbers in NFHS 2 and 3 vs Cluster nos in NFHS 1 and 4? I want to use -svyset- on the data while running summary stats and analytics and since PSU is a key variable in that, I am not sure how to proceed with varying PSU numbers across waves. For example:

PSU ranges between

2015-16 : 10001 to 360482

2005-06 : 1001 to 33214

1998-99 : 1001 to 33214

1992-93 : 4 to 341

I would really appreciate if the experts at DHS could provide some guidance on that front. Thank you!

Niranjana

Thanks a lot!

Niranjana

Subject: Re: Appending multiple waves of the NFHS annd PSU codes
Posted by Bridgette-DHS on Fri, 18 Oct 2019 17:42:45 GMT
View Forum Message <> Reply to Message

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

These four surveys have much different dates and sample sizes. You can append them as repeated cross-sections in order to make comparisons but I think it would be meaningless to regard them as a single cross-section, if that is what you are suggesting. For example, a CPR using all cases in the four surveys would not have a clear reference population. In general, in

DHS datasets, the PSU is given both as v001 and v021 (or hv001 and hv021). The two almost always agree exactly. However, if they do not agree exactly, use v021 (or hv021). For a small number of surveys, only one or the other is included, in which case you use the one that is included.

Subject: Re: Appending multiple waves of the NFHS annd PSU codes
Posted by Niranjana on Thu, 24 Oct 2019 11:10:13 GMT
View Forum Message <> Reply to Message

Thank you Bridgette and Tom.

I intend to use the data as repeated cross-sections in order to make comparisons. But given the differences in district level data being non-comparable between 1992-93/2005-06 vs 1998-99/2015-16, I hope to use the state level variations instead.

I'm assuming all trend analysis should also be restricted to the NFHS 2,3 and 4 rounds given intra-state administrative changes.

Do let me know if this seems like a reasonable step. Thanks again.

Subject: svy and svyset after appending multiple waves of the NFHS
Posted by Niranjana on Mon, 04 Nov 2019 10:19:08 GMT
View Forum Message <> Reply to Message

Dear Bridgette and Tom,

While working with the appended (all four waves have been appended to form a repeated cross-sectional dataset) version of the DHS India women's folders, I've found that the variable for strata (V023) is coded differently across all years from 1992-93 to 2015-16. I would like to work with -svy- and -svyset- codes but am concerned about the variable definition varying significantly between survey waves. Currently, this is how I was hoping to run the codes to obtain SEs, CIs.

```
gen wt=v005/1000000

egen stratum  = group(v024 v025)
svyset v021 [pw=wt], strata(stratum)
br stratum


svy : mean var_interest, over(v024)
```

However, I am unable to produce any standard errors or CI after this and I get the following error message:

 Missing standard errors because of stratum with single sampling unit

I am also not sure if I need to generate a different weight variable for each year of the India women's survey rounds.

Do let me know how to proceed on this front.

Thank you,

Niranjana

Subject: Re: svy and svyset after appending multiple waves of the NFHS
Posted by Bridgette-DHS on Mon, 11 Nov 2019 13:52:45 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Specialist, Kerry MacQuarrie:

When appending surveys, you will have to do the following two steps:
1. Harmonize the strata variable;
2. Ensure that variable has unique values across all surveys.
    (Same for the PSU variable).

Step 1:
The appropriate strata variable is usually v023 but not always. In our workshops, we train participants to match it by examining Appendix A. You may want to create a new variable, called "strata", before appending that is equal to v024 x v025 for NFHS-1 and NFHS-2, and equal to v022 for NFHS-3 and NFHS-4.

Step 2:
There may be equivalent values on this variable in 2 different surveys that refer to different strata (or PSUs). I typically handle this by adding a prefix, e.g. add 100 or 1,000 to the strata variable in the first survey, 200 or 2,000 to the strata variable in the second survey, etc. This should be done BEFORE appending. (Note: Because of the number of strata and PSUs in all of India's surveys, but especially NFHS-4, you may need more digits!)

Subject: Re: svy and svyset after appending multiple waves of the NFHS
Posted by Bridgette-DHS on Mon, 11 Nov 2019 14:30:17 GMT

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

We apologize for the long delay with this reply. There is no need to append the files. You can append them if you want, of course, but if you do append them, the file will be very very large because of the size of the NFHS's.

You need to include v024 (or hv024) as part of the id if you do any merging of the India files (within a single round).

In DHS surveys in general, the PSU and cluster are usually the same, but not always. Usually the cluster is v001 and the psu is v021 and you can confirm that v001=v021. In the India surveys, as I recall, you have v021 and not v001. If you have both v001 and v021 and they are different, then svyset should use v021.

Subject: Re: svy and svyset after appending multiple waves of the NFHS
Posted by Niranjana on Tue, 12 Nov 2019 10:46:15 GMT

Thank you for the response.

I am appending the files to get a repeated cross-sectional dataset so it is necessary in this context that I append it.

I followed the advice of Kerry MacQuarrie for both NFHS 1 & 2 to create a strata variable, however they are not unique within waves. V024 is the states variable and V025 is they type of residence

gen strata = v024*v025
tab strata v024

I got the following output after following the 2 steps mentioned by Kerry.

I am not certain if this was the expected output but as you can see, two states have the same two strata variable. So adding the prefix depending on NFHS 1 or 2 would not work. Do let me know if there is a way to rectify this before appending.

File Attachments
1) strata.png, downloaded 2092 times

Subject: Re: svy and svyset after appending multiple waves of the NFHS
Posted by Bridgette-DHS on Mon, 18 Nov 2019 12:38:54 GMT
View Forum Message <> Reply to Message

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

In the appended file, I suggest that you includ a variable "survey" that is numbered 1, 2, 3, 4 for the successive NFHS's. You then construct the combined PSU number with one of the "egen" commands, specifically "group". That is, if the PSU is given by v021, the command would be "egen PSU_all=group(survey v021)". You need to do something similar for the strata, for example as "egen strata_all=group(survey v023)". There are other ways to develop unique id codes for the clusters and strata but this is the easiest way.