
Subject: Pooling 3 rounds of DHS Nepal -- weights?
Posted by [LeahBevis](#) on Tue, 08 Oct 2019 17:38:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

I am working on a project where we pool 3 rounds (2006, 2011, 2016) of Nepal's DHS data in the terai region only, doing analysis in Stata with all observations at once. While I know how to use DHS weights for a single cross-section, I am not fully sure how to adjust the weights, PSU and strata variables in order to have correct SEs in all analysis. Three questions.

1. It looks like I should generate weights in this way, where 2497.704, 2660.37095, and 2773.8584 are the total population of Nepal in each round. (I obtained those populations by first creating wt as below (v005/1000000), then using tab year [iweight=wt].)

```
gen wt = v005/1000000
replace wt=wt/2770.8256 if year==2006
replace wt=wt/2660.37095 if year==2011
replace wt=wt/2773.8584 if year==2016
```

After doing this, I find that the weighted-average of my variable wt is 1 in each round. I think this is correct?

2. I am unsure how to change the stratum within Nepal, in part because I'm not clear on the construction of the stratum in each year, and also because I don't know if I would ideally want these stratum to be unique by year or to change by year. Right now, the stratum in the terai range from 9-13 in 2006 and 2011, but from 1-14 in 2016. If stratum should be the SAME in all years, then I need to recode the 2016 stratum to match the previous years. (And I would need to know, from DHS, how to do this so the locations were consistent.) If stratum are supposed to be UNIQUE by year, I simply need to differentiate the numbers 9-13 from each round.

3. Similarly, the primary sampling unit IDs (held in variable v021) changed in 2016. The IDs range up to 7,000 in 2006 and 2011, but stop at 400 in 2016. Similar to the question above, I'm not really sure what the goal is here... do I want PSUs to be unique by round, or the same across rounds? If the same across rounds, I need to know how to recode the 2016 PSUs into the 2006 and 2011 PSUs. (Also, is it a problem that while many PSU IDs are given in both 2006 and 2011, there are also many PSU IDs that are only in 2006, or only in 2011?)

4. And just to be sure, after having created an adjusted PSU var and an adjusted strata var, I believe I use the wt variable defined above, and then run:

```
svyset v021 [pw=wt], psu(psu_adjusted) strata(stratum_adjusted)
```

Thank you!
Leah

Subject: Re: Pooling 3 rounds of DHS Nepal -- weights?

Posted by [Bridgette-DHS](#) on Mon, 21 Oct 2019 10:59:05 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

Say that you have combined the three surveys (by appending them) and you use v021 as the psu id and v023 as the stratum id. Then run these two lines:

```
egen clusterid=group(survey v021)
egen stratumid=group(survey v023)
```

This is the easiest way to get unique ids in the combined file for clusters and strata. There is no need to do any other kind of spatial reconciliation.

What you have done with the weights seems ok to me. Yes, the means will be 1 within each survey and overall. However, there is a problem of scale. You divided v005 by 1000000, but when you divide again by 2770.8256, etc. you will have a number which is much less than 1. You should avoid having decimal points or anything to the right of a decimal point in a weight. There are some kinds of weights (such as fweight in Stata) that require a weight to be an integer. I have encountered weighting procedures that will truncate a weight and ignore anything to the right of the decimal point, which means that a weight between 0 and 1 will be treated like 0--that is, the case will be ignored entirely! You might want to rescale your weights somehow with an arbitrary multiplier such as 1000000.

Subject: Re: Pooling 3 rounds of DHS Nepal -- weights?

Posted by [LeahBevis](#) on Wed, 22 Jul 2020 19:46:45 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hey, I'm so sorry for the delay, but could you clarify what the variable "survey" is, there? Is this a categorical variable indicating DHS round/survey? I.e., survey =1 for 2006, survey=2 for 2011, survey=3 for 2016?

Subject: Re: Pooling 3 rounds of DHS Nepal -- weights?

Posted by [LeahBevis](#) on Wed, 22 Jul 2020 20:12:30 GMT

[View Forum Message](#) <> [Reply to Message](#)

Additionally, actually, when I try this there seems to be a problem with the stratum variable created.

Ignoring the issue of relative weighting of the 3 survey for a moment, I ran:

```
gen svyweight = v005/1000000
gen survey=1 if year==2006
replace survey=2 if year==2011
```

```
replace survey=3 if year==2016
egen clust=group(survey v021)
egen stratum=group(survey v023)
```

Then I can successfully run this:

```
svyset clust [pw=svyweight],
svy: reg HAZ X $CONTROLS
```

But THIS does not work, when I add the stratum:

```
svyset clust [pw=svyweight], strata(stratum)
svy: reg HAZ X $CONTROLS
```

So there seems to be a problem with creating the stratum as a grouped variable in that way. Can you please advise? Thanks!

Subject: Re: Pooling 3 rounds of DHS Nepal -- weights?
Posted by [Bridgette-DHS](#) on Sat, 25 Jul 2020 16:22:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

I assume you have converted the dates? All the dates in the Nepal surveys are in the Nepalese calendar.

I suggest that you add "singleunit(centered)" at the end of the svyset line. Please let us know if that doesn't work.

Subject: Re: Pooling 3 rounds of DHS Nepal -- weights?
Posted by [LeahBevis](#) on Mon, 27 Jul 2020 21:45:38 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

Thanks for the response. Yes, we converted those dates. I realized that the problem is actually the 2016 strata.

We are working only with the districts in the Terai region of Nepal -- the southern-most districts along the Indian border.

In 2006 and 2011, the 13 strata contain 5 strata that uniquely defined the Terai, allowing us to have a representative Terai sub-sample. Great.

But in 2016 the stratification changed: now 10 strata define the urban and rural parts of the 5 provinces shown in the attached map. The problem is, that leaves us with very few PSU in the Terai part of the rural/urban strata of Provinces 3 and 4 -- circled in yellow on the map. In Province 3, only 1 rural PSU exists, and this is why Stata can't use svyset with those strata. However, only 4 urban PSUs exist in Province 3, and only 3 rural and 3 urban PSUs exist in Province 4. And I don't understand how Stata's svyset interprets weights with respect to their stratum, so I don't know if these very-few-PSU-strata are a problem.

So, ideally, I want our sample to be representative of the Terai. I realize that's not possible with the 2016 sampling strategy, but how should I best deal with the PSUs in Provinces 3 and 4, in order to be as close as possible? Should I...

--- Drop the single rural PSU in Province 3 and proceed as normal?

--- Re-assign Province 3 IDs to Province 2 IDs (since Chitwan is next to Province 2) and re-assign Province 4 IDs to Province 5 (since Nawalpur is next to Province 5)? This seems geographically logical, but is a bad idea if weights are properly interpreted only w/ respect to their correct stratum.

--- Something else?

Thanks in advance for your advice.

Leah

File Attachments

1) [Nepal_prov2016.png](#), downloaded 329 times
