
Subject: NFHS3 - Weights and Survey command
Posted by [rsuri](#) on Fri, 14 Aug 2015 11:04:37 GMT

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

Hey,

I am analysing the health indicators of children under the age of 5 from the NFHS-3 dataset and would be running ordered logit, binary logit and count models. I am using individual and the children's recode files and had a few doubts-

1. For the summary statistics, which type of weights do i use- aweights or iweights with the v005 variable? The latter gives me slightly different standard deviations from the former but the difference is of a few decimal values only. Also, Stata 12.0 gives me an error while using pweights, so I am assuming that using pweights can be ruled out. Is that okay?

2. While running the regressions, is it advisable to use the survey command? I came across an article online which said that its use for maximum likelihood estimation techniques should be avoided. If i can use the command, can you tell me the specific information i would need to set for the same to be stored in stata? Also, if i use the survey command, does it work for summary statistics as well in the sense that i don't have to specify the weights?

Thanks.

-RS

Subject: Re: NFHS3 - Weights and Survey command
Posted by [Bridgette-DHS](#) on Sat, 15 Aug 2015 14:44:14 GMT

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

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

For summary statistics, you can usually manage with iweights. You should virtually never use aweights. The definition "analytical" for the "a" is very misleading. Read "help aweight"--they can be a big help if you are combining means that are based on different numbers of cases but not otherwise. Use, for example, "summarize x [iweight=v005/1000000]"

For virtually all estimation models, you can use svy, with pweights. I am not aware of the warning you mention. Some people, usually economists, are opposed to using any weights in estimation models. Perhaps that's where the warning came from, but it would not just be for ML estimation.

You can get the syntax for svyset with "help svyset". It has changed slightly with version 14. In DHS surveys, the cluster variable is v001 or v021, they should be identical. You use "pweight=v005"; no need to divide by 1000000. Stata automatically normalizes pweights to have a mean of 1. See other posts for the stratum variable; it is not always the same. After specifying svyset, to apply it, you put "svy:" at the beginning of an estimation command, e.g. "svy: regress y

x".

Summary statistics with weights can be difficult. For example, "tab A B, summarize(x)" doesn't like iweights, let alone pweights. The "collapse" command can be problematic. However, if all else fails, for such commands, you can use "[fweight=v005]" and just ignore the huge frequencies; means will be correctly weighted.

Subject: Re: NFHS3 - Weights and Survey command
Posted by [lucianabrondi](#) on Wed, 07 Oct 2015 12:11:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear DHS FOrum colleagues,

I am also using the same dataset, DHS2005_06 India, using Stata 13 and looking at care seeking outcomes for diseases in children. Each Child is my unit of analysis and I would like to be sure that I am declaring the data as survey data properly before running descriptive analysis and regression. I have two questions.

1. have put the V005 as the weight like this (see under), does that make sense? Is that enough?

```
svyset _n [pweight=v005], vce(linearized) singleunit(missing)
```

2. If I want to account for intracluster correlation or household correlation, can I use another command? Should I do it?

Thanks, Luciana

Subject: Re: NFHS3 - Weights and Survey command
Posted by [Bridgette-DHS](#) on Thu, 08 Oct 2015 12:02:32 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

As you may know, "_n", or "underscore n", is assigned by Stata to be the ordinal number of the case in the file you are currently using. That is, if you have N cases, then the first case has _n=1, the second case has _n=2, the last case has _n=N, etc. For most purposes, if you want to use or save that number, you have to give it another name, for example with "gen id=_n".

What you want in that position of the svyset command is the cluster variable. You can use v001 or v021 for that purpose. They are always (so far as I know) exactly the same. That's how to include the cluster adjustment.

Thus, instead of "svyset _n [pweight=v005], vce(linearized) singleunit(missing)", you probably

want "svyset v001 [pweight=v005], vce(linearized) singleunit(missing)".

Subject: Re: NFHS3 - Weights and Survey command
Posted by [lucianabrondi](#) on Thu, 08 Oct 2015 12:58:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks a lot,
Luciana

Subject: Re: NFHS3 - Weights and Survey command
Posted by [ambudon](#) on Wed, 22 Mar 2017 11:06:12 GMT
[View Forum Message](#) <> [Reply to Message](#)

This discussion is useful. I think, I have done the right thing. But let me confirm with the experts here.

So I am analysing woman's questionnaire - specifically, the section where questions are asked to married women about partner's education and occupation. Hence, one has to use women weight variable.

In this case, I have used the commands of the following nature:

```
tab2 VARx VARy if (some condition added) [fweight=v005], column OR  
tab2 VARx VARy if (some condition added) [fweight=v005s], column OR
```

Both weights give me the same results. And as pointed out, the abs # of frequencies are quite high. But it's the % which are imp. in this case.

Am I right?

Subject: Re: NFHS3 - Weights and Survey command
Posted by [Bridgette-DHS](#) on Wed, 22 Mar 2017 11:50:17 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

- (1) You refer to v005s, as an alternative to v005. I don't know what is the difference. Sorry if you defined v005s and I missed it.
 - (2) Instead of tab2 you should use tab. Enter "help tab" to find out the difference.
 - (3) I don't know of any option "OR" that can be used with the tab commands. There is an option "or", for odds ratios, that can be used with logit.
 - (4) Yes, with fweight your percentages will be correct. Ignore the frequencies, which are too high by a factor of one million (10 to the 6th).
-
-

Subject: Re: NFHS3 - Weights and Survey command
Posted by [ambudon](#) on Thu, 23 Mar 2017 05:22:34 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Tom,

Thanks for your reply.
