
Subject: Primary sampling units (PSU)
Posted by [joy](#) on Wed, 12 Feb 2014 04:49:57 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear users,

I would like to check the number of PSUs in round 1, round 2 and round 3. I was looking at variable (V021).

In NHFS3, it corresponds to 3850 PSUs reported in the final report (I understand that it has been randomised, but it takes as many unique values).

But in earlier rounds (NFHS 2 and NFHS 1) I have much fewer unique values for this variable. According to the final reports, NFHS 2 should have 3165 PSUs and NFHS 1 should have 3522 area units. So I am thinking that V021 should have as many unique values.

Can anyone please confirm this? Or am I missing something?

Many thanks,
Jaai

Subject: Re: Primary sampling units (PSU)
Posted by [Liz-DHS](#) on Tue, 18 Feb 2014 04:24:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear User,

Here is a response from one of our experts, Fred Arnold:

The PSU numbers are unique within a single state, but they repeat across states, so you should look at the state code in conjunction with the PSU numbers when trying to determine how many PSUs there are.

Subject: Re: Primary sampling units (PSU)
Posted by [joy](#) on Mon, 24 Feb 2014 04:50:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

Many thanks, Liz and special thanks to Fred Arnold.
It is very good to get confirmation from the experts. Thanks.

Subject: Re: Primary sampling units (PSU)
Posted by [Liz-DHS](#) on Mon, 24 Feb 2014 12:32:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

you're welcome!

Subject: Re: Primary sampling units (PSU)
Posted by [vega25](#) on Mon, 06 Oct 2014 20:55:36 GMT
[View Forum Message](#) <> [Reply to Message](#)

May I ask what "looking at state code in conjunction with the PSU numbers" means for analysis. If I want to cluster at the level of the PSU, would the option cluster(psu) in Stata at the end of the regression equation not suffice, because the PSU numbers repeat in various states? If so, how do I generate unique numbers for each PSU across the states?

Thanks very much!

Subject: Re: Primary sampling units (PSU)
Posted by [Trevor-DHS](#) on Tue, 07 Oct 2014 14:51:56 GMT
[View Forum Message](#) <> [Reply to Message](#)

You probably need to combine state and psu. An easy way to do that is to use

```
egen newvar=group(state psu)
```

Subject: Re: Primary sampling units (PSU)
Posted by [Liz-DHS](#) on Tue, 07 Oct 2014 15:39:46 GMT
[View Forum Message](#) <> [Reply to Message](#)

Here is another response from one of our experts, Fred Arnold:
If you're using data for only a single state, you can use just the PSU number, but if you're using data across states or for the whole country you need to combine the state and PSU codes in order to obtain a unique code for each PSU. For example, every state will have a PSU=001 so that PSU code is not unique. If you add the PSU code at the end of the state code you will have a unique variable. For example, PSU 001 in state 22 will have the unique PSU number of 22001. PSU 001 in state 24 will have the unique PSU number of 24001. One way to generate the unique codes is to create a new variable, which =1,000*state code + PSU code.

Thank you for your post.

Subject: Re: Primary sampling units (PSU)
Posted by [vega25](#) on Thu, 13 Nov 2014 20:03:28 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you very much for your response, and please accept my apologies for the delay in communicating my thanks.

Interestingly I notice that a suffix <cluster(psu2)> in a regression statement in STATA for all-India

analysis, where psu2 has been calculated using the method that Trevor and Dr. Arnold described does not affect the results, even the standard errors. Is that normal? At my end, I'll look into the actual data file and the new psu2 variable that's been created a little more later this week, and report again.

Subject: Re: Primary sampling units (PSU)
Posted by [Trevor-DHS](#) on Sat, 15 Nov 2014 00:48:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

I'm not sure I'm following exactly what your question is, but in general I think you want to be using the svyset command to describe the sampling parameters and then svy: regress for your regression, and not trying to add other parameters to your regression. There are a number of other post on the forum about the use of svyset and svy.

For NFHS3 for example, you should be able to use:
svyset psu2 [pw=v005/1000000], strata(v022)

and then
svy: regress ...

Subject: Re: Primary sampling units (PSU)
Posted by [kbietsch](#) on Thu, 30 Mar 2017 14:41:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello Trevor,
I am looking for the strata variable (or which to combine to construct one). You suggest v022, but looking at the data file, this variable is missing. According to the sampling framework described in the final report (volume 2), the NFHSIII is not a 2 stage cluster sample, has both 2 and 3 stages (in rural and urban areas, respectively).

This is the code I generally use for analyzing DHS data (in R) with strata of urban/rural.

```
df$sampleweights <- df$v005/1000000  
design <- svydesign(ids=~v021+v002, strata=~v025, weights=~sampleweights, data=df)
```

How do you set the strata design for NFHSIII?
Thank you,
Kristin

Subject: Re: Primary sampling units (PSU)
Posted by [Trevor-DHS](#) on Thu, 30 Mar 2017 17:03:50 GMT

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

1) When I look at the NFHSIII dataset, I see values 1-73 for v022. The strata are urban slum and non-slum area and rural areas in each state. See table C.2 (Sample characteristics) in Volume II page 15 of the report.

2) For the svydesign the ids should be just the cluster level ids, not including the household id, so it should only be v021

Thus, I think the command should be:

```
df$sampleweights <- df$v005/1000000
```

```
design <- svydesign(ids=~v021, strata=~v025, weights=~sampleweights, data=df)
```

Subject: Re: Primary sampling units (PSU)

Posted by [kbietsch](#) on Fri, 31 Mar 2017 06:58:31 GMT

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

Thanks Trevor, I just downloaded the latest version of the data, and v022 is no longer missing. However, v021 is- though there is a note to use s021- which works perfectly.

The survey design is now set as:

```
df$sampleweights <- df$v005/1000000
```

```
design <- svydesign(ids=~s021, strata=~v022, weights=~sampleweights, data=df)
```

Cheers,

Kristin

Subject: Re: Primary sampling units (PSU)

Posted by [Trevor-DHS](#) on Fri, 31 Mar 2017 15:46:14 GMT

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

Looks good. I guess I'm behind the latest version.
