
Subject: The Gambia DHS, 2013

Posted by [Jawla](#) on Sat, 06 Aug 2016 02:19:54 GMT

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

Hi DHS experts:

According to the DHS Household and Sampling Manual, the DHS surveys are two stage surveys. The sampling units for the first stage of selection is called the Primary Sampling Unit (PSU); the sampling unit for the second stage of selection is called the Secondary Sampling Unit (SSU).

This is the survey syntax for stratified two stage design from Stata 14 documentation:

```
svyset psu [pweight=weight], strata (strata) fpc (fpc1) || ssu, fpc (fpc2)
```

```
svyset v021 [pweight=v005], strata (v022) fpc (fpc1) || ssu, fpc (fpc2)
```

What is the secondary sampling unit(ssu) variable in DHS data set?

Any suggestion will be appreciated. Thanks

Subject: Re: The Gambia DHS, 2013

Posted by [Jawla](#) on Sun, 07 Aug 2016 19:59:24 GMT

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

I am assuming that I might have to merge the women's data set and that of HH data set to be able to use the ssu(which I think is the household). Right?

Subject: Re: The Gambia DHS, 2013

Posted by [Liz-DHS](#) on Mon, 15 Aug 2016 14:17:50 GMT

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

Dear User,

A response from Dr. Ruilin Ren, sampling expert:

Quote:

In the STATA survey set, you need to declare the PSU (DHS variable V001 or V021), but not necessary to declare the SSU (DHS variable V002), if your aim is to produce sampling errors. You do not need to declare the fpc because it is not provided in the DHS data.

Subject: Re: The Gambia DHS, 2013

Posted by [Jawla](#) on Mon, 22 Aug 2016 03:48:16 GMT

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

Thanks Dr. Ruilin for your kind answer to my question. Is very much appreciated!

Best, Jawla.

Subject: Re: The Gambia DHS, 2013
Posted by [ab803](#) on Tue, 26 Sep 2017 19:03:23 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Dr. Ren,

If the FPC is not provided, how do we set up the second stage, which requires an FPC for stata to recognize it. I'm trying to account for the fact that multiple children are selected from a household (and are non-independent), but am unable to set up hv002 as ssu.

Many thanks!

Subject: Re: The Gambia DHS, 2013
Posted by [Liz-DHS](#) on Tue, 10 Oct 2017 12:45:01 GMT
[View Forum Message](#) <> [Reply to Message](#)

A response from technical expert, Dr. Tom Pullum:

Quote:

As has been noted in other posts, we are unable to separate out the probabilities of selection for the cluster (v001 or v021)) and the household (v002). What we have is the product (the second probability is conditional on the selection of a cluster). There is no way to separate the two. That's unfortunate, and it means that svyset cannot be properly constructed for multi-level models. The basic reason for this is that if the separate probabilities were available, then the first one, with a little additional information, would allow one to identify the PSU. DHS does everything possible to protect the privacy and confidentiality of the respondents. Apart from that, the required information is deleted after the weights have been constructed. We couldn't retrieve it even if we wanted to.

Subject: Re: The Gambia DHS, 2013
Posted by [ab803](#) on Fri, 10 Nov 2017 05:55:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much, this is really helpful!

Any advice on how to account for the clustering of multiple children within a household?

With gratitude!

Subject: Re: The Gambia DHS, 2013
Posted by [Liz-DHS](#) on Wed, 29 Nov 2017 15:59:41 GMT

Dr. Tom Pullum responded:

Quote:

There is definitely some clustering of child-level outcomes by the mother and/or household. At DHS we do not adjust for that, but you certainly may want to do so. A complication is that the number of children per woman is low, especially if you are focusing on outcomes within a recent time period and on children in a narrow age range such as 0-4. Other users may be able to suggest how to specify svyset for such outcomes.
