
Subject: FGC--The Gambia: DHS 6--2013.
Posted by [Jawla](#) on Wed, 03 Aug 2016 07:01:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

I am conducting analysis on attitudes towards FGC using the women data set with Stata. I have a question on weight.

Primary sampling unit v021

Sample strata v022

Women individual sample weight or the pw v005/1000000

fpc/10,233

Hence, this is a two stage stratified cluster sampling. Should I be worried about the secondary sampling unit and which variable is that? Do I need the household dataset for this? How relevant is the v020 in the women data set. Should I use the single stage or multi stage stata syntax.Thanks!

Subject: Re: FGC--The Gambia: DHS 6--2013.
Posted by [Liz-DHS](#) on Mon, 08 Aug 2016 13:43:13 GMT
[View Forum Message](#) <> [Reply to Message](#)

A response from Shireen Assaf:

Quote:

Hello Jawla,

In DHS we have a multi stage stratified sample design with one strata variable v022 or v023 as you indicated. There is no need to use fpc. You can set up the svyset for your analysis as follows:

```
gen wt =v005/1000000  
svyset v021 [pw=wt], strata(v022) singleunit(centered)
```

Thank you.

Best,

Shireen Assaf (DHS Senior Research Associate)

Subject: Re: FGC--The Gambia: DHS 6--2013.
Posted by [Jawla](#) on Tue, 09 Aug 2016 05:37:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much, Shireen, I very much appreciate your prompt response to my question.
Best, Jawla.

Shireen,

I played with the data after trying:
svyset v021 [pw=wt], strata(v022) singleunit(centered)

(Did not need to use -- gen wt =v005/1000000 -- as wt is already defined.)

after playing with some tabulation, the "population size" is still coming out approximately the same as the "number of observations." I think there is a problem -- I may be wrong. Evidence that I could be wrong is that a simple tab and a tab that includes the probability weights yield different answers. For example:

```
tab g119 [iweight=wt]
```

Female circumcision: continue or be stopped	Freq.	Percent	Cum.
-----+-----			
1. Continued	6,594.629	64.96	64.96
2. Stopped	3,392.317	33.42	98.38
8. Don't know	161.738423	1.59	99.97
9	3.19361798	0.03	100.00
-----+-----			
Total	10,151.878	100.00	

```
. tab g119
```

Female circumcision: continue or be stopped	Freq.	Percent	Cum.
-----+-----			
1. Continued	6,305	62.12	62.12
2. Stopped	3,674	36.20	98.32
8. Don't know	165	1.63	99.94
9	6	0.06	100.00
-----+-----			
Total	10,150	100.00	

However, as the Stata manual says better than I could:
pweights, or sampling weights, are weights that denote the inverse of the probability that the observation is included because of the sampling design.

So it seems to me that the population size reported in tabulations should be much higher than the

number of observations used to create that tabulation.

Why is that when you do a two-way tabulation (for example), the number of observations and the population size are reported to be very similar to one another?

Thank you and I look forward to hearing from you at your earliest convenience.
Best, Jawla.

File Attachments

1) [Gambia DHS 2013 women's dataset.docx](#), downloaded 820 times

Subject: Re: FGC--The Gambia: DHS 6--2013.
Posted by [Liz-DHS](#) on Mon, 29 Aug 2016 18:18:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

A response from Technical Specialist, Shireen Assaf:

Quote:

Dear Jawla,

In the DHS we perform a normalization so that the weighted and unweighted totals are the same. Perhaps the following DHS online video tutorials on sampling and weighting can help:

<https://www.youtube.com/watch?v=DD5npelwh80&list=PLagqLv-gqpTN8IZQBy7vAYw10NjynAn2Z>

Also in your code, you have set up the svyset but have not used it. If you run the svyset then you can use svy: ta g119 . There is no need to write ta g110 [iw=wt] if you use svy.

Best,

Shireen

Subject: Re: FGC--The Gambia: DHS 6--2013.
Posted by [Jawla](#) on Tue, 30 Aug 2016 02:11:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Shireen,

Thank you so much for your follow up on this matter. This is very much appreciated! I will keep you posted if I encounter any further issue with the DHS data.

Best, Jawla,
