Subject: Re: FGC--The Gambia: DHS 6--2013.

Posted by Jawla on Sat, 13 Aug 2016 02:42:32 GMT

View Forum Message <> Reply to Message

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	6,594.62	29 64.9	6 64.96
Stopped	3,392.3	17 33.4	98.38
8. Don't know	161.73842	23 1.59	99.97
9	3.193617	' 98 0.0	3 100.00
+			
Total	10,151.8	378 100.	.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 822 times