Subject: Number of obs and population size not the same
Posted by Johanne Elgaard on Mon, 01 May 2023 12:00:07 GMT
View Forum Message <> Reply to Message

I'm working with the 2014 Ghana DHS. I am trying to calculate frequencies and percentages after weighting the data. My problem is that my number of observations are 945, but when using both svyset and [iweight] I get a population size of 925 and a number of obs of 945 in the output. When calculating the sum of the weighted frequencies together, it also adds up to 925. Can anyone explain the difference between number of obs and population size to me?

Thank you:)

Subject: Re: Number of obs and population size not the same
Posted by Bridgette-DHS on Mon, 01 May 2023 12:44:04 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

I believe you are just seeing the difference between the weighted and unweighted numbers of cases. In the IR file, v005 is scaled or normalized so that the mean value of v005/1000000 is 1. However, for subpopulations, or in the KR file, where v005 is attached to the woman's children, the mean weights don't necessarily retain this property. That's why the weighted and unweighted totals differ. If you think this is not the explanation, please let us know.

Subject: Re: Number of obs and population size not the same
Posted by Johanne Elgaard on Mon, 01 May 2023 13:10:36 GMT
View Forum Message <> Reply to Message

Thank you for your response :) That could be the explanation since I'm using the KR file. Should I then just weight the data when during my logistic regression but not interpret specifically on the frequencies since the weighted and unweighted numbers do not correspond here?

Subject: Re: Number of obs and population size not the same
Posted by Bridgette-DHS on Mon, 01 May 2023 16:45:22 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

You should not have to do any re-weighting. If you are using Stata, with svyset and [pweight=v005], Stata will always re-normalize so the weights have a mean of 1. As part of this,

the factor of 1,000,000 is removed. (That's why you get the same results with [pweight=v005] or with [pweight=v005/1000000]. I don't know about packages other than Stata.