Subject: Reshaping Kenya's DHS dataset
Posted by Ritapriya Bandyopadhyay on Wed, 18 Oct 2023 06:55:18 GMT
View Forum Message <> Reply to Message

Hi, I am trying to reshape the Kenya DHS data.
Now, I believe the data is in wide format, so hv104_01-hv104_24 - represent the sex of each household member within a household? How to go forward with the reshaping?

Also, I want to create a unique household ID and a unique individual ID. The problem is, when I create a unique individual ID before reshaping to long - all individuals in the household gets the same unique ID, hence should I do this after reshaping to long?

Would greatly appreciate your help

Best
Ritapriya

File Attachments

1) Capture.PNG, downloaded 97 times
2) Capture.PNG, downloaded 95 times

Subject: Re: Reshaping Kenya's DHS dataset
Posted by Bridgette-DHS on Thu, 19 Oct 2023 12:33:48 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Staff Member, Tom Pullum:

You are apparently using the HR file, in which all the household information is on a single very wide record. You should use the PR file, in which cases are individual household members and the household-level information is on each record. The PR file is a reshaped version of the HR file. You do not need to do the reshaping--it has already been done.

Subject: Re: Reshaping Kenya's DHS dataset
Posted by Ritapriya Bandyopadhyay on Thu, 19 Oct 2023 13:04:24 GMT
View Forum Message <> Reply to Message

Thanks a lot! I had another question. I wanted to find out the number of students enrolled across different education levels per age. So to weight it, I am using the following command: svyset psu [pw=weight], strata(stratum) singleunit(scaled). To view my results I am using the following command: tab age school_level, count. I have used hv022 for stratum and hv021 for psu. I have divided the hv005 variable by 1000000 to arrive at the household weights.

However when the count is displayed I observed that "number of observations" is greater than "population size" - how is this possible? Because population size is supposed to be greater than

sample observations, right? I am working with adolescents between 10-24 year olds, but I am using household weights.

Subject: Re: Reshaping Kenya's DHS dataset
Posted by Bridgette-DHS on Thu, 19 Oct 2023 14:55:31 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Staff Member, Tom Pullum:

When you use svyset, the "number of observations" often shifts to some large number that's much different from the actual sample size. Frankly, I just ignore that number. Whatever it is, it's NOT the number of observations.

For what you are doing, you only need to adjust for the weights, not clustering or stratification. The adjustments for clustering and stratification only affect the standard errors of estimates, not the estimates themselves. If your command is just "tab age school_level [iweight=hv005/1000000]", I think you will get the same results. That table would give the weighted number of cases in the sample. The percentages describe both the sample and the population.

Subject: Re: Reshaping Kenya's DHS dataset
Posted by Ritapriya Bandyopadhyay on Thu, 26 Oct 2023 05:49:11 GMT
View Forum Message <> Reply to Message

Hi,
Thank you!

Just wanted to confirm once, I have added a snapshot - I am checking weighted frequency of school enrollment - the number of observations reduces as I add iweight (as shown in the snapshot) - you're saying this is possible?

Best,
Ritapriya

File Attachments

1) school enrollment.PNG, downloaded 83 times

Subject: Re: Reshaping Kenya's DHS dataset
Posted by Bridgette-DHS on Tue, 21 Nov 2023 14:16:10 GMT

Following is a response from Senior DHS Staff Member, Tom Pullum:

Yes, the weighted and unweighted totals are never exactly the same for subpopulations. The weighted total can be smaller or larger than the unweighted total. Usually within 10% but sometimes there is a larger difference.