Subject: Re: Weight in Indonesia DHS 2002 and 2007
Posted by Ihuri on Sat, 11 Jan 2020 11:31:30 GMT
View Forum Message <> Reply to Message

Thanks a lot Bridgette and Tom,

It works now.

On a related note, I just want to make sure about the stratification used to weight the data in 2002, 2007, 2012, and 2017 Indonesia DHS.

According to the DHS manual (https://dhsprogram.com/Data/Guide-to-DHS-Statistics/Analyzin g_DHS_Data.htm), it's said that "Typically, DHS samples are stratified by geographic region and by urban/rural areas within each region." I could follow this concept for 2012 and 2017 because, in these datasets, v022 consists of the number of all urban-rural area for each province in the survey year. For example, v022 in 2017 dataset contains 67 strata which represent urban and rural area in each of 34 provinces minus rural Jakarta, because Jakarta is the capital city and entirely considered as urban area). This also matches the explanation in the 2017 report "The sample of the 2017 IDHS was stratified by province and by urban and rural areas, and implicitly stratified by welfare concentration" (appendix B, p351).

However, in the 2002 and 2007 dataset, v022 shows that the number of strata is 680 and 751, respectively. I'm aware that there was a change of province number in Indonesia over time (30 provinces in 2002 survey and 33 in 2007 survey) and the appendix B in 2002 report (p267) does mention additional info about the five districts in East Java covered in the Safe Motherhood Project (SMP). However, I still don't quite understand where the 680 strata came from.

So there are several things I need to ask/clarify:

1. Does v022 in each survey year represent the number of strata by provinces and urban-rural residence?

2. If the sample of each survey year is stratified by provinces and urban-rural residence, why do 2002 and 2007 datasets have hundreds of sample? What do these numbers represent?

3. Why do v022 and v023 have the same unique values in the 2017 and 2012 datasets, but not in 2007 and 2002 datasets? For example, in 2017, v022 and v023 both have 67 unique values. Meanwhile, in the 2002, v022 (which is labeled as number of stratum) has 680 values and v023 (labeled as sample domain) has 26 values.

4. I've been following the guidelines including the Youtube tutorial and using "gen wgt = v005/1000000" followed by "svyset[pw=wgt], psu(v021) strata (v022)" to apply weights in the 2012 and 2017 dataset. To run analyses, I simply use svy commands like "svy: tab var", "svy: poisson var1 i.var2 i.var3, irr", etc. I haven't encountered any problem so far, but I'm just checking if this is the correct approach.

5. I tried adding 'singleset' option in dataset 2017 but it didn't change the results (no difference with the results generated by commands without 'singleset' option. Is it because there's no

singleton PSU is 2017 strata?

Apologies if I have missed any guidelines, and for the long list of questions.
Thank you and have a nice weekend.