
Subject: Re: PSU and strata identifiers - Egypt
Posted by [Bridgette-DHS](#) on Fri, 08 Jul 2016 11:57:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Specialists, Tom Pullum and Trevor Croft:

To be clear, the files are EGIR33 (the 1995 survey), EGIR51 (the 2005 survey), and EGIR61 (the 2014 survey).

In most DHS surveys, PSUs and clusters are the same, so either v021 or v001 can be used as the cluster id. However, for most surveys of Egypt, including the three you are using, there are two clusters per PSU. You should use v021 as the PSU id, as you are doing, but with the following modification for the first survey. For that survey, v021 is incorrect and should be reconstructed with "v021 = int(v001/10000)". After that change, you can use v021 as the id for PSU in all three surveys.

In all three surveys the sample strata are urban and rural areas within each governorate. That must be constructed in the first two surveys. For example, in those two surveys you could use these two lines: "rename v023 v023_original" and "egen v023 = group(v023_original v025)". For the third survey, just use v023 as it is.

We would never recommend that you replace the original file with revised variables. If you make changes using original variable names, as just suggested, you should save with another file name or you should just put those changes in a program and not save the data file.

Yes, you can pool the surveys into a single file. You can include a variable "survey" or you can distinguish them with v000. In the three surveys, v000 is coded EG3, EG4, and EG6, respectively. You could reduce the files to just the variables you need before appending them.

You could get unique identifiers in the pooled file, after you have reconstructed v021 and v023 as described above, with "egen cluster=group(v000 v021)" and "egen stratum=group(v000 v023)" and then "svyset cluster [pweight=wt_dv], strata(stratum) singleunit(centered)". There have been many postings on possible rescaling of the weights when surveys are pooled. In this case I would leave the weights alone, because I assume (and hope) you are really just looking at changes between surveys rather than, say calculating means or percentages that combine all three surveys.