

---

Subject: Re: Weighting Data for Pooled Indonesian DHS Dataset

Posted by [Bridgette-DHS](#) on Mon, 20 Jun 2016 12:13:25 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

When you pool the surveys, I recommend that you construct a variable called "survey" with survey=1 for the 2007 survey and survey=2 for the 2012 survey. There are other ways to distinguish between the surveys but this an easy way.

You then construct a unique cluster code with "egen cluster=group(survey v001)". Note that v001 and v021 are the same and you can use v021 here if you prefer.

You construct a unique stratum code with "egen stratum=group(survey v024 v0025)". In the 2012 survey, v022 and v023 are identical and they are a crossing of v024 and v025. In the 2007 survey, both v022 and v023 are defective and the ONLY way to get the strata is by crossing v024 and v025. (The 2007 survey also has a defective label for v007, year of the survey.) When you pool surveys, the clusters and strata id codes must be different for the surveys. The strategy of adding 1000 or 2000, etc., to the codes is obsolete; "egen group" is much better, and as you note it can include labels.

The label for v024 (region) is called HV024 in the 2007 survey and hv024 in the 2012 survey. There are some differences in the labels that I think are misspellings or changes in official names. You must be very careful when pooling surveys because codes can change, especially for country-specific variables. The labels for the last survey in the append sequence will over-write the previous labels and you will not be alerted to any differences or changes.

Apparently you want to re-weight the surveys in such a way that the weighted sample sizes are proportional to the number of women in the population. You can do that either before or after the appending or pooling. Or, if you just want the weighted number of cases to be the same in each survey, you can do that. Your proposed recodes should work, but you need to confirm at the end that the total weights have the property you were trying to achieve. You also want to make sure that the average weight has lots of digits before the decimal point. The mean weight of the original hv005 or v005 will be close to 1000000. You definitely do not want the weights to end up with only one or two digits before the decimal place. Preferably 6 to 8.....

---