

---

Subject: Pooling cross country IPUMS-DHS data for all available surveys; using svset

Posted by [Rafi](#) on Wed, 03 Jul 2019 06:57:19 GMT

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

---

I want to do a cross-country study on child mortality using DHS data (birth recode). I am using IPUMS-DHS data (<https://www.idhsdata.org/idhs/>). My data extract from IPUMS-DHS consists of around 30 countries. All available waves for each country are included in the analysis.

I want to pool all the surveys across time and countries. I would be obliged if someone could tell me how to svyset the data in Stata. Some of the available methods which I could find so far suggest grouping the strata "egen stratid = group (year v024 v025), label" (<https://www.stata.com/statalist/archive/2014-01/msg00150.htm> I). However, the problem with this approach is that the IPUMS-DHS data does not have region variable (v024). One obvious reason why v024 is not there in IPUMS-DHS is that regions are different in different countries.

Using the usual svyset command "svyset idhspsu [pweight=v005], strata(idhsstrata)" gives the error "Note: Missing standard errors because of stratum with single sampling unit." I am using Stata/MP 15.1.

---

---

Subject: Re: Pooling cross country IPUMS-DHS data for all available surveys; using svset

Posted by [Bridgette-DHS](#) on Thu, 18 Jul 2019 12:34:34 GMT

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

---

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

There have been many postings on how to do this topic. Rather than repeat that information, I'd like to question what you mean by "pooling". Term can mean different things.

Something that many users do, and that we sometimes do here at DHS, is to combine several surveys into a single file, using the "append" command in Stata. Then you can loop through the analysis of each survey, one at a time, within a single combined file. You don't have to keep opening and closing separate files. This can make data processing more efficient but it doesn't affect the results. You can also use a combined file to look at differences between surveys.

However, sometimes the term means that you want to produce estimates that refer to some mega-population, such as "East Africa". I would only advise this for the analysis of some relatively rare event, such as fistula, for example, where you may not have enough cases (i.e. statistical power) in the separate surveys. But when you pool surveys this way, the reference time is blurred, you don't usually have a well-defined population, and the relative weight of each survey will be proportional to the sample size for that survey, which is arbitrary (unless you adjust the weights to make each survey count equally or to be proportional to the population of the country). With child mortality, you have enough cases for each survey and I wouldn't recommend this kind of pooling.

I wish I knew more about IPUMS, but I HOPE the files include a stratum variable even if they don't include region. Usually the stratum variable is just a crossing of all combinations of v024 and v025, and when it isn't, the stratum variable is MORE appropriate than the crossing of v024 and v025. If you replace "v024 v025" in your command "egen stratid = group (year v024 v025), label" with the name of the stratum variable, you will be ok.

To avoid the single sampling unit problem, just add "singleunit(centered)" at the end of the svyset command. Hope that works.

---