

---

Subject: Re: Merging data files in Stata

Posted by [Bridgette-DHS](#) on Fri, 22 Nov 2013 18:38:01 GMT

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

---

Here is a response from DHS Stata expert Tom Pullum:

It can be useful to pool surveys and/or ignore weights for exploratory work or model building. However, you have to be careful in a final analysis. The general goal that I would have is to produce rates, means, proportions, coefficients, etc. that are unbiased estimates of what you would get if you had complete data from a well-defined population at a specific point in time.

If you pool multiple surveys from a single country, the time points or time intervals should be taken into account, at least as covariates.

If you pool surveys from multiple countries, then the pooled sample probably cannot be described as a sample of a well-defined population. The dates of data collection will differ, but beyond that there will be many historical, cultural, ethnic, etc. differences. You probably will not have data from ALL the countries in a recognized region, e.g. "East Africa".

If you pool, consider using a multi-level model, and allow for variation in the effects, not just in the intercepts.

If you CAN justify pooling, then you have to re-weight somehow, at least to compensate for the arbitrary variation in the sizes of the samples. The two main options would be to re-weight so that the total weight is the same for each country or survey (as with the U.S. Senate) or the total weight for each survey is proportional to the population of the country, e.g. the estimated number of women 15-49 at the time of the survey (as with the U.S. House of Representatives).