

---

Subject: Re: Domestic violence weight, denormalize, pooled cross-section, cross-tabulation

Posted by [Bridgette-DHS](#) on Mon, 20 Jun 2016 15:34:14 GMT

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

---

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

Pooling surveys into a single file is convenient for data processing and for calculating differences, but as you imply, the reference population is not well defined. I do not recommend calculating a mean (or something like a mean) for all surveys combined. But sometimes people do this and there's no law against it. If you decide to do this, I would recommend giving equal weight to each survey, which means re-scaling v005 or d005 in each survey so that the weighted total is the same in each survey. That is, if there are 10 surveys, and the UNWEIGHTED total number of women with d005<. in all 10 surveys is N, then re-scale d005 in each survey so that the WEIGHTED total in each survey is N/10. As I said, however, I would be reluctant to pool the surveys this way.

I would prefer to use the pooled data to do regressions that include "survey" as a categorical variable for fixed effects OR, if you have a lot of surveys, a random effect for the intercept. For such regressions, you do not need to re-scale the weights, but can leave them as they are in each survey. Then the total weighted number of cases will equal the total unweighted number of cases in each survey and for the combination of all surveys. From a statistical perspective, this is good because, as you say, the actual number of cases is what you need for a valid estimate of sampling error. And you are not producing an estimate of an overall mean (or proportion, etc.).

Yes, you can re-normalize d005 in the same way as v005. (I prefer "re-normalize" to "de-normalize". I don't think the latter term means the same thing for everyone.)

Stata recommends that you use the subpop option within svyset. I have done some checking and the difference between using subpop and NOT using subpop is always very small, much smaller than sampling error, but there are good theoretical reasons for using it. You refer to it but I don't see that option in your svyset statement.