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

Posted by [RenaM](#) on Mon, 20 Jun 2016 09:37:09 GMT

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

Dear DHS users,

I explore associations between experience of domestic violence, including marital controlling behavior, and health (measured using anthropometry data and information on anemia status). I will use all individual-level DHS survey data sets that feature a complete domestic violence module and measurements for BMI and anemia for the same women, yielding 22 datasets (2000 to 2014, I ignored older data sets).

The main reason I pooled the data is to have enough observations to be able to look at different combinations of violence and marital controlling behavior (e.g. any violence AND controlling behavior, physical violence AND controlling behavior, physical violence AND NO controlling behavior). I always use a completely "clean comparison group" consisting of women who did not experience any violence at all. Before running logistic regressions (binary dep.var: "BMI below 18.5", or "Has any anemia"), I would like to create meaningful cross-tabulations (mostly oneway and twoway for anthropometric indicators, dv indicators and main control variables).

Now I'm confused about denormalized weights for pooled data and results of descriptive statistics.

- Is it valid to denormalize domestic violence weights even if they only apply to a subgroup (those selected for dv module)?
(DHS forum posts apply to v005 and the like)

- In my cross-tabulations, frequencies are large numbers and seem to be estimates of the total population of my survey countries. This seems dodgy because the numbers do not represent the actual number of observations in my pooled sample anymore. I've never seen any tabulation like that in any paper. Should I have applied any other adjustments to the domestic violence weight?

- For a pooled sample including countries from several regions (South and Central Asia, Africa, Latin America), is there any other meaningful way to weight the data? Or may I ignore weights and svy adjustments in cross-tabulations for pooled data altogether since my dataset anyway does not represent a certain world region?

So far, I followed other DHS forum posts on the topic, and did the following in Stata for each country BEFORE pooling the data:

1) divided d005 by 1000000 (=d005_pw)

2) De-normalized weights:

```
gen d005_pwpool=d005_pw*(total population of women, age 15-49, at the time of the survey/number of women in the resulting domestic violence subsample)
sum d005_pwpool, detail
```

[d005_pwpool sums up to the total pop of women, 15-49, at the time of the survey]

3) I executed the svy command in the pooled dataset, including unique codes for each country's psu and strata:

```
svyset [pweight=d005_pwpool], psu(psu_pool) strata(strata_pool)
```

Then, I ran cross-tabulations using tabout and svy, e.g:

```
tabout agecat residence indexwealth edulevel occup_pool reli_pool ///  
if subgroup==1 using pool0.xls, ///  
svy oneway cells (freq) clab(_ _ _) ///  
format(0c) layout(rb) h3(nil) npos(row) replace
```

"Subgroup" is an indicator for the subsample of observations without missing values and excludes pregnant women/women who gave birth in the preceding 2 months.

Looking forward to any helpful hints, thank you very much.

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.

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

Posted by [RenaM](#) on Wed, 29 Jun 2016 21:55:26 GMT

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

Dear Bridgette, dear Tom,

Thanks for your clarification. I tried the options you suggested and I think I found a good solution to handle the data (in short: I go with DHS survey datasets from African countries only; comparison of equally weighted and unweighted pooled regressions does not make a huge difference in my case, but is informative).

Thanks a ton!
