
Subject: Pakistan 2014 - Using correct weights for domestic violence

Posted by [gspek](#) on Mon, 02 Nov 2015 04:50:57 GMT

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

I am using the Pakistan 2014 women's dataset for analysis of associations between women's empowerment (and individual and community level) and domestic violence experiences (at individual level). I am currently struggling with a couple of issues:

1. How do I correctly weight community-level independent variables? I tried to create community-level weighted means for average percentage of women working in a community, average decision making index value, and percentage in community accepting of any domestic violence justification scenarios (with participant's values excluded). However, I'm not sure how to approach the weights in regressions where both the community and individual level independent variables are included.

```
foreach var of varlist works dm_sum dvaccept_any{
  gen `var'_w = `var'*wgt if !missing(`var') & nocomm == 0
  bys v001: gen `var'_agg_w = sum(`var'_w) if nocomm == 0
  gen `var'_comm_w = (`var'_agg_w - `var'_w)/(clustersz - 1) if nocomm == 0
}
```

2. For stata logic analysis (and summary statistics), do I use iweights or pweights?

Thank you!

Subject: Re: Pakistan 2014 - Using correct weights for domestic violence

Posted by [Bridgette-DHS](#) on Mon, 09 Nov 2015 12:51:12 GMT

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

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

If your units are the individual women, then you should use the individual-level weights for the women. In general this would be v005. However, for models or combinations of variables that include the DV variables, use dv005.

If you have attached cluster-level variables to the records of the individual women, then what I just said continues to hold.

If you were using clusters as units--which I doubt--then you could construct cluster-level weights by adding up the weights for the women in the cluster (e.g. with the collapse command).

You would never multiply the variables by the weights. This is not how weighting works. You use the weight option.

Use pweights when you can. If you can't, use iweights. For example, "summarize x [iweight=v005/1000000]". Otherwise, do not use iweights.

Subject: Re: Pakistan 2014 - Using correct weights for domestic violence
Posted by [gspek](#) on Tue, 10 Nov 2015 16:53:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Tom,

Thank you for the detailed response. I have a couple of follow up questions:

1. When using the svy: commands, is clustering by psu and weighting enough, or does the data also require stratification? I saw a paper that used the dataset mention stratification due to the data collection design, but it did not discuss what the strata should be. Could you please clarify?
2. If multiplying by the weight is incorrect in that situation, how would you then go about creating weighted community-level averages with the respondent's own value excluded?
3. If we were to merge in male data (probably on community level), how would that alter the weights?

Thank you!

Subject: Re: Pakistan 2014 - Using correct weights for domestic violence
Posted by [Bridgette-DHS](#) on Thu, 12 Nov 2015 14:46:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Yes, you should use the strata. Because stratification makes the sample more efficient, the stratum adjustment will usually slightly reduce the standard errors and confidence intervals. The stratum variable is v022 or v023, usually the combinations of urban/rural and region. Sometimes there is ambiguity about the stratum variable. DHS is preparing a list for all the surveys.

If you merge the male data with the DV data then you are probably talking about a couples file. Usually the recommendation is that you use the male weights (mv005) for couples. In your situation, it could be better to use dv005. The general rule is to use the weights that have been calculated for the subsample with the highest level of nonresponse. Men have more nonresponse than women. Among women, there is a higher level of nonresponse for the DV questions than for non-DV questions. It could be a judgment call, whether you would use the men's weights or the DV weights for these couples. If there is ambiguity about what weights to use, I suggest you do some analysis with one set and then repeat with the other set. I would expect very little difference.

To construct your cluster-level variables, you can do a collapse, within which you can use the weight option, and then attach the cluster-level variables to the individual-level records. You should use the weight option rather than multiplying the variables by the weights, which you did in some earlier code.

Subject: Re: Pakistan 2014 - Using correct weights for domestic violence
Posted by [gspek](#) on Fri, 13 Nov 2015 05:58:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much! That was incredibly helpful.
