

---

Subject: Re: Weighting variables in DHS India data (1992 and 1998)

Posted by [Reduced-For\(u\)m](#) on Sun, 16 Aug 2015 21:11:25 GMT

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

---

The problem with doing the weighting manually (instead of via the "svy" prefix) is not in the weighting, but in the stratification and clustering. That is, your command will weight the observations just fine, but won't account for other aspects of the survey design.

The clustering at PSU could be done manually ", cluster(psu)", but I am not sure there is a good, easy way to manually account for the stratification (in the sampling procedure).

Your version (just the manually used weights) will produce standard errors that are too small (your p-values will be too small, and you'll reject a true null hypothesis too often). Adding in clustering at the PSU level will produce almost right standard errors (or p-values, or CIs) but they may be a little too big since they don't account for the stratification. Using the "svy" prefix guarantees not just that you get good population level estimates of means/differences (weighting) but also more appropriate statistical inference properties (rejection rates, CI coverage rates, however you want to think of that).

---