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

Posted by [user_rm](#) on Tue, 01 Sep 2015 12:34:48 GMT

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

thanks for your reply. I have been looking at the variable v005 and I have a query relating to it. The weights are the same for all individuals in the district e.g for district Ahmadabad, the sample weight (v005) is 1148330 for all kids in that district. I wanted to ask if I should apply these weights for every individual (ie kid here) or does this number represent the total population for the entire district? I ask because I was using the collapse command to get the data at district level. I used `collapse(rawsum)v005/1000000 , by (district)` for the sample weight v005 as well to get 1 figure for each district. But in that case, if there are 25 kids in district, the collapse (rawsum) amount adds the weight 25 times. I was wondering if that is right? Or should I be using just 1148330 as the total weight for all kids in the same district?

Just to be clear I am running regressions at district level. So like you suggested in your previous email, I use collapse commands for controls and outcome variables. I am unsure about how I should use the survey weights? should I `collapse(rawsum) v005` or `collapse(mean) v005`? I apply these weights to my district level regression. Also when running regression should I use `aweights` or `pweights`? When I am collapsing data initially then should I use `aweights` or `pweights`?

Please clarify.

Thanks

R
