Subject: sample weights when using a subsample
Posted by amil on Tue, 27 Sep 2016 21:30:10 GMT
View Forum Message <> Reply to Message

I would like to kindly ask one quick info related to the use of sample weights. I am using a pooled dataset of India 1992, 1998, and 2005.

In my analysis, i only consider a sub-sample of women: those who at least one child, less than 9 children, aged 15 to 39, and exclude women for whom some key variables are missing.

Do I need to use sample weights when I do descriptive stats and regressions with this subsample? I am unsure what it means to use sample weights when you are using a subsample of women in the IR file (and the subsample of women with children is certainly not balanced across the strata, for example urban/rural)...

Also, some authors use weights, others do not. Are there clear guidelines on this?

Many thanks for any assistance on this, highly appreciated!

Subject: Re: sample weights when using a subsample
Posted by Bridgette-DHS on Wed, 28 Sep 2016 12:54:28 GMT
View Forum Message <> Reply to Message

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

I recommend that you always use the sampling weights, even for a subsample such as you described. The reason for using the weights is that they correct for relative over-sampling and under-sampling of geographically defined strata and they correct for different levels of nonresponse. The weights are intended to produce unbiased estimates of proportions, means, rates, etc. Without using the weights, the estimates will be biased toward areas that were over-sampled or had the highest response rates. Comparisons between the India surveys, for example, to estimate changes, will be meaningless if you do not use weights.

The only exceptions to using weights, so far as I am concerned, would be for checking data quality, checking recodes, and a few other situations where you are just testing a Stata command or program, listing cases, etc.

I recommend using weights for all analyses, including statistical models such as logit regression. I know some people do not use weights, or make other adjustments for clustering and stratification in the sample design, which can affect the standard errors of the estimates (but not the estimates themselves). I would be interested in whether any users of the DHS forum would take that position, and what reasons they would give.