Subject: Do I need to weight data for cohort analysis?
Posted by AmsP on Wed, 06 Apr 2016 10:54:12 GMT
View Forum Message <> Reply to Message

Good morning,

I plan to use women datasets to calculate average schooling years for each 10-year birth cohort (e.g. people born in 1970-1979 and in 1980-1989) for a particular country or a sub-country region (across several survey rounds). I am not quite sure if it is necessary to use sampling weight, strata/cluster adjustment and de-normalization for calculating cohort-based value.

Sampling weight is to adjust different sampling sizes across regions with different population sizes, so it is a cross-sectional/regional adjustment. But a cohort analysis is a time-series issue, so I am not sure if a cohort analysis of the DHS data should also be adjusted by sampling weight, strata/cluster and de-normalization.

Thank you very much in advance!

Subject: Re: Do I need to weight data for cohort analysis?
Posted by Reduced-For(u)m on Thu, 07 Apr 2016 00:24:15 GMT
View Forum Message <> Reply to Message

The weights are not just for differing sample sizes and population (sorta) but for probability of being sampled - some groups of people are over/under sampled by design. So if you want to actually estimate levels (such as, the average schooling of a group of people) you do in fact need the weights. The clustering is also necessary to get correct standard error estimates, so you have to do that too. Neither of these should be difficult.

The time-series/cohort dimension might require you to do more with your standard errors (in the sense of clustering at a higher level) but it would depend on what exactly you are estimating and how. But the way you describe it - which is just to calculate average schooling by cohort - there is no real time dimension to your analysis, you are just picking your sub-groups based on cohort.

Now - if you wanted to estimate the determinants of schooling using variation in cohort exposure to some variable X, that would be a different story about the adjustments you want to make. But otherwise, you can just grab each cohort you want from the data, calculate the means using weights and the CI/p-val using clustered standard errors, and you are fine.

For Stata - to do that in a regression context, just "svy: reg Schooling if cohort==c" for each cohort c and the constant will give you the mean and you can use the confidence intervals Stata spits out. Before that you have to do the "svyset" command that is described in many other posts and on the DHS website, but the exact right specification of the "svyset" command depends on the particular survey you are using.

Subject: Re: Do I need to weight data for cohort analysis?
Posted by AmsP on Mon, 11 Apr 2016 10:55:09 GMT
View Forum Message <> Reply to Message

Thank you very much for help! Actually, what I need is just a simple mean value of the female population in a region, for regression analysis. So maybe I do not need to take into consideration standard error (and thus cluster and strata adjustment).