Subject: Re: Creating panel
Posted by Reduced-For(u)m on Thu, 17 Aug 2017 19:05:57 GMT
View Forum Message <> Reply to Message

I don't think you gain much by collapsing the data in that way. Why not just use the individual-level observations?

Usually when you collapse data like that it is because you are doing something like bringing in external data that is merged to the cohort (say in your case, information on when schools were built; or maybe cohort variation in exposure to some mandatory schooling laws). Even then you don't necessarily need to collapse the data down, and can just merge the variables into the individual-level data.

As for the problem of getting the birth year wrong - does it really matter? Isn't a mother born in December of, say, 2005 very similar to one born January of 2006? It isn't clear you would be wrong to lump these two groups together in one time effect. But again, it depends a lot on the data setup, such as if Dec/Jan born women had different "exposures" to something important.

Also, once you collapse the data down into averages, it wouldn't be a 0/1 variable on the left - it would be a proportion. In which case the probit model wouldn't be right. You would want to run a probit on the individual-level data...so again I don't see the need to create this pseudo-panel. You could just run a least-squares regression of some sort on the aggregate data.

That said, if there is a reason to generate the pseudo-panel, it is straightforward to do using the "collapse" command in Stata and the "by()" option as "by(cohort)" or "by(cohort region)" or whatever is appropriate. You would also perhaps want to collapse using the DHS sample weights (to get representative estimates), which is explained in various places on the forum and on Stata help forums. But it doesn't seem clear that you really want or need to do that.