Subject: Re: Stata do file for U5 mortality analysis
Posted by Reduced-For(u)m on Mon, 22 Apr 2013 00:10:30 GMT
View Forum Message <> Reply to Message

Hey there. Sorry it has taken so long to get back to you. I don't have Stata code to do the synthetic cohort method, but I think maybe we can work through a little bit.

First, though...when you ask about CIs and tests...what are you trying to compare? If you want a confidence interval around the mortality rate, we might be able to come up with something, but I haven't ever actually seen that done for this synthetic method. But I think maybe we can get something if we make some assumptions. Let's try this:

First, we need some ingredients, and since I don't have the code, I'll just list them and you can grab the info from the recode files. When you're all done, maybe you can post the code here so we have it. Anyway...

Birthdate (year and month)
Survey date (year and month) - make these into numbers using the time commands in stata: gen bdate = ym(byear bmonth)
Age at measurement (compute from birthdate and measure date, to get "ages" for dead children (ugh, I hate writing "dead children")
Alive? (0/1 indicator)
Age at Death - make this something like 99 if the child is still living, computed maybe from deathdate - birthdate
X - any covariates you want to have around for cutting up the sample
sampling weight, strata, cluster (there is code on the DHS FAQs for this under "using data files" http://www.measuredhs.com/faq.cfm)
    *Also, do the "svyset" thing, as described.

*Note: I'm going to use "global" macros here to store some information, just know these will stay on your computer, and just in case, type "macro drop _all" at the beginning of your do file...so long as you don't have any macros you are keeping around for good measure.

*Now, a decision. I think you want to compute just 1 rate from the whole period, interpreted as a "current" mortality rate, yes? Well, then we have to decide whether, for instance, we want to use information for u1 mortality on children born a few years back. I'm going to assume we do. The "deep" assumption, here, is that child mortality hazards are not changing over the survey period. If we need to relax this, let me know.

Now, we are going to compute first the Prob(Death before age 1 month)

*an indicator for child having died
gen u1mort = alive==0 & deathage==1

*Now, we want to know, of the children who were recorded after they lived or would have lived to 1 month of age, what fraction died.
*And we want to weight everything all right and whatnot

```
svy: reg u1mort if measureage>=1
global P1b = _b[u1mort]
global P1se = _se[u1mort]
   *this may or may not end up working, but it might be nice to keep this handy
```

*Now were are going to do this for all the other ages in one loop

```
forvalues i=2/59 {
gen u`i'mort = deathage==`i'
reg u`i'mort if measureage>=`i'
global P`i'b = _b[u`i'mort]
global P`i'se = _se[u`i'mort]

}
```

*Now we'll compute all the rates you want

Under 1 year mortality
```
gen u1rate = $P1 * $P2 * ... * $P11
```

*That should basically allow you to compute any rate you want. Unless I made an obvious arithmetic error in here somewhere, which is possible.

Now, a few disclaimers. There are way more children in one survey who have measureage>1 than have measureage>=59. So your estimates for the lower ages are far more precise. You'll see that in your standard errors. But how to build a CI is a bit hard.

One thing you could do is ignore that and a few other things, and if you assume that all of these estimates are independent draws, you can use a standard formula: var(a*b) is described here: http://www.stata.com/statalist/archive/2005-12/msg00183.html

But I'm thinking that would get a little bit long. There is probably some better option that the DHS uses when it is computing it's synthetic cohorts that I just don't know, that accounts for the fact that the same kids are in all these regressions/means and the sample size issues (although, I think they do this some other way, because maybe they just use each kid once...so that for P1 you are using just kids born 2 months ago, and for P37 you are just using kids born 38 months ago...or something like that).

Real quick, though, you could get a U1 mortality rate and SE (and thus CI) using a slightly different method. For example:

```
keep if measureage>=12
gen u1mort = deathage<12
svy: reg dead
global u1rate = _b[dead]
global u1se = _se[dead]
*and for CI at 95%
global ubound = u1rate+1.96*$u1se
```

global lbound = u1rate - 1.96*$u1se

This method gives an average u1 mortality rate for all children in the survey who were (or would have been) at least 1 year old when the survey happened....so an average rate over the previous 5 years, but excluding the survey year. It's not the synthetic cohort approach, but it gives a nice standard error/CI.

OK. That was basically some ideas more than a real code. I hope it was helpful - I'm not totally sure that it was. I'm happy to iterate on this. It would be nice to have some clean code to post here at the end so that other people can have a template handy. Let me know if I can help anymore. I'll try to get back to you faster.

Anyone else wants to play with this code, re-organize it, or point out why I'm doing something really stupid, that would be appreciated.