Subject: Dropping mothers' records without dropping the child's
Posted by Alinoor on Sat, 06 Feb 2016 08:32:29 GMT
View Forum Message <> Reply to Message

hey there,
i'am using 2011 Ethiopia data set particularly child record. as we know it's one record per child and much of the mother's information is attached to the child's record. if i want to drop an observation based on one of the maternal characteristics without dropping child's, how can i do it? for instance i tried to keep the height measurements of those mothers with only one child but ended up dropping child records too:

replace v438=1 if (hwidx~=1)
tab v438
drop if v438==1
tab v438

thank u

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Reduced-For(u)m on Wed, 10 Feb 2016 20:22:58 GMT
View Forum Message <> Reply to Message

"Drop" in your code will drop observations, not variables. Instead, you could just replace the value of the maternal variable you don't want with "." instead of dropping the observation.

Note: be careful with your summary stats/analysis after this, because in some cases it will average over all the observations, and sometimes it will ignore the observations with missing variables, and you could accidentally end up comparing different samples in different places. See the Stata documentation for how to deal with that.

Note note: looking at your explanation, I'm not exactly sure what you want. You want mom's with only 1 kid to have height measures, and no height measures for the other moms, but you still want their kids in the sample? Why?

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Alinoor on Thu, 11 Feb 2016 07:49:02 GMT
View Forum Message <> Reply to Message

Thank you for your reply. What i understand is, for instance if a mother has three kids she is repeated thrice . Therefore , my question is instead of having one child per record, can I have one mother with three kids or two kids kind of record.

Thank u

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Bridgette-DHS on Thu, 11 Feb 2016 14:12:40 GMT
View Forum Message <> Reply to Message

Here is another response from Senior DHS Stata Specialist, Tom Pullum:

Yes, if you use "drop" then you will lose the record entirely. In your example, it sounds like women should be the unit of analysis and you should be using the IR file. The number of children born in the past five years is v208. Perhaps you want a command such as this: summarize v438 if v208>1. Note that v438 includes special codes in the range 9994 to 9999, so it would be better to say summarize v438 if v438<9994 & v208>1. Another possibility would be tab v208 if v438<9994, summarize(v438), which would include women with no births or one birth, but you could just focus on women with more than one birth.

If you use the KR or BR files to study women, then each woman (only mothers) will appear once for every child born in the past five years. I doubt that you want that kind of multiple counting of women. Or perhaps you want to relate a characteristic of the child to a characteristic of the mother. For example, you could be studying the relationship between the mother's BMI and whether the child is stunted, underweight, overweight, or wasted. If so, you could remove the women who had one child by including "...if hwidx>1" in the estimation command, before the comma.

In these examples I am omitting the weights; in real life you would want to include them. Note that in the KR and BR files, bidx, hidx, midx, and hwidx are the same.

If there is some other reason why you are using the KR or BR file to analyze women, I'd be interested to hear what it is. Don't hesitate to ask again if I am not answering your question.

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Alinoor on Thu, 11 Feb 2016 18:29:35 GMT
View Forum Message <> Reply to Message

Thank you very much for your swift and constructive response. I found it useful. Actually my unit of analysis is child. My concern was not to inflate maternal mean BMI , height...etc as a result of repeated mothers. I can temporarily drop mothers with more than one child and calculate their mean BMI..etc and come back to the data for the rest of analysis.

Thank you once again

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Reduced-For(u)m on Thu, 11 Feb 2016 21:15:44 GMT

You can have children who have mothers with the same height (the same mom) and it will not bias your point estimates. It is likely, though, to bias your standard errors (because you aren't getting a full amount of new information by adding new children to the same mother). The good news is that if you use the DHS clustering method (the one using svyset) it will account for this. The basic intuition is that, absent DHS survey design, you'd want to "cluster" on the household (all those with the same mom). But, since DHS requires you to cluster at the PSU level, and all children in a household are in the same PSU, you are already taking the within-household serial correlation in maternal BMI into account (you can always cluster at a higher level and it will subsume the lower level).

So basically, you should be fine. If you were fitting a hierarchical model, you could add separate variance parameters for households and PSUs, but if you follow the DHS recommended strategies you should be just fine even though multiple children have the same mom.

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Alinoor on Fri, 12 Feb 2016 03:12:16 GMT

Thank you.

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Bridgette-DHS on Thu, 10 Mar 2016 14:17:36 GMT

Additional comments from Tom Pullum:

If you are using the KR file and looking at the relationship between characteristics of the mother and characteristics of the child, such as anemia, underweight, etc., you have two options for defining mother-child pairs. One is to use all the children and repeat the mothers. From the child's point of view--if, say, you believe the mother's status is a determinant of the child's status--I think that option is preferable. The direction of causation is ambiguous, but that direction may dominate. You could then be multi-level, with the child as the level 1 unit, the mother as level 2, and the cluster as level 3. The repeats of the mothers does not have to be a reason for throwing out a high percentage of the observations.

The second option is the one you are taking, matching the mother with only one of her children. If you do that, you need to give a good reason for doing so. And if you do that, it is not a good idea to just match the woman with the youngest child. That will introduce a bias (see https://www.dhsprogram.com/pubs/pdf/MR14/MR14.pdf) because the youngest child born in the

past five years tends to be the ONLY child born in the past five years, and therefore tends to be healthier than other children born in the past five years. It would be better to select the matched child at random--although that means no one will be able to match your results exactly.

I suppose a third option would be to use all mother child pairs but to re-weight, for example dividing v005 by 2 if the woman has two children in the KR file, dividing by 3 if she has three, etc.

These are just suggestions, but I hope you will consider them.

Subject: Re: Dropping mothers' records without dropping the child's
Posted by Alinoor on Fri, 18 Mar 2016 13:37:51 GMT
View Forum Message <> Reply to Message

Thank you Tom. I found your suggestions very useful in my analysis.