Subject: Creating an unbalanced panel
Posted by Niranjana on Tue, 26 Nov 2019 14:35:03 GMT
View Forum Message <> Reply to Message

Hi,

I am working with the Individual modules of NFHS I-IV. I want to create an unbalanced panel where the woman enters the panel at the time of her marriage and exits at the time of interview also containing birth history data among other variables. Basically, I want the following

indvID Year Ageatbirth Child_id Date of Birth Birth_order
1 1981 25 1 4/3/1981 1
1 1983 27 2 12/15/1983 2
1 1985 29 3 11/1/1985 3
2 1980 22 1 1/1/1980 1
2 1984 26 2 3/1/1984 2

Do I simply reshape the data to get the result? Thanks!

Subject: Re: Creating an unbalanced panel
Posted by Bridgette-DHS on Wed, 11 Dec 2019 14:55:38 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I don't understand why you would want to reshape the data. The information you want is on each woman's record in the IR file. If you want to apply survival methods, for example, you have everything you need--the date of birth, date of marriage, date of each birth (including birth order) and date of interview, on the woman's record.

Subject: Re: Creating an unbalanced panel
Posted by Niranjana on Wed, 11 Dec 2019 15:12:12 GMT
View Forum Message <> Reply to Message

Thank you Tom and Bridgette.

I know that IR file contains all necessary information.

This is the code I used:

```
gen id =_n
sort v007 caseid
by year caseid: gen newid = 1 if _n==1
replace newid = sum(newid)
replace newid = . if missing(caseid)


duplicates report newid
rename b*_0* b*_*
keep caseid-v458
reshape long bidx_ bord_ b0_ b1_ b2_ b3_ b4_ b5_ b6_ b7_ b8_  b10_ b11_ b12_ b13_ b15_
b16_, i(newid) j(bindex)
drop if bidx==.
```

I was able to create a panel to analyse the impact of historical policies. Would this be right?

Subject: Re: Creating an unbalanced panel
Posted by Bridgette-DHS on Fri, 13 Dec 2019 18:52:53 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The file constructed with these lines has one record for every birth and is equivalent to the BR file, which is available on the website. I really don't know what you mean by "unbalanced panel". DHS surveys do not have a panel design. Are you trying to simulate a panel design, in which the same women would have been re-interviewed at, say, five-year intervals? You could definitely do that, by truncating the birth histories and then re-assembling them, but what would be the gain in doing that, when you have complete retrospective birth histories? I also don't know what "unbalanced" means in this context.

I believe you are trying to transform the data into an appropriate format for some specific analytical method. What is that method?

Subject: Re: Creating an unbalanced panel
Posted by Niranjana on Mon, 16 Dec 2019 15:08:50 GMT
View Forum Message <> Reply to Message

Thank you for the response Tom.

I was looking to create a dataset with multiple entries per woman depending on the number of children she has with the exact same variables as would be present in the IR file for the year. Except now with instead of multiple columns for each child (as in IR), the dataset will be long with multiple rows per woman and her birth history.

This is what I meant by an unbalanced panel -- the woman enters the panel at the time of marriage with year entries for each birth and exits the panel on the year of interview. I reviewed BR files and the data seems to represent exactly what I was looking for.

Do you think it would be wise if I were to append multiple waves of the NFHS BR datasets? Are the caseid variables unique across each wave? using -duplicates report caseid- shows that some caseids are reported over 20 times even though the max number of children recorded is 18 for NFHS I AND II.

Thanks so much for the help, Bridgette and Tom!

Subject: Re: Creating an unbalanced panel
Posted by Bridgette-DHS on Fri, 20 Dec 2019 14:35:23 GMT
View Forum Message <> Reply to Message

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

I'm glad that the BR file will meet your needs.

Successive DHS surveys in the same country never include interviews with the same households or respondents. I don't believe that even the same cluster (enumeration area) ever re-appears. There is no sense in which the ID codes for individuals, households, or clusters in one survey can be linked with the ID codes in the next survey. (The continuous surveys in Peru and Senegal have a rolling design in which some clusters will re-appear, but not households or individuals, and over a five-year interval there is complete turnover.)

It is possible to examine birth cohorts, defined by calendar year of birth, in successive surveys. Qingfeng Li and Amy Tsui have worked on this approach; see, for example, https://www.cpc.unc.edu/projects/china/publications/1955.
