Subject: Re: Calculating Age structure explanatory variables on Household Wealth
Posted by xrl1g11 on Thu, 17 Nov 2016 17:16:52 GMT
View Forum Message <> Reply to Message

Dear Tom,

Thank you so much! It worked; for the most part. I've decided to use Stata for my analysis, but have come across another problem. When using your coding, all the needed demographic variables were created perfectly. However, upon using the last bit of your coding:

"sort hhid

merge hhid using e:\DHS\DHS_data\scratch\ETtemp.dta

drop _merge"

the values for youth_dep_ratio, dep_ratio, prop_age1, prop_age2, prop_age1, all changed. I have noticed, that after the use of the "collapse (sum) hhsize age1 age2 age3, by(hhid)" command (before doing the merge as instructed), the original number of observations dropped from 77,744 to 16,702. Which makes sense, as the observation changed from number of household members to number of households, which allows for the calculation of my demographic variables needed.

However, when I carry out my multinomial logistic regression after using that merge command (outlined at the top), the observation count drops back to 77,744 and the definition for say prop_age1 (originally proportion of 0-14 per household) changes from per household to per household member (I hope i'm interpreting this correctly). The values of my created demographic variables have also all changed or am I doing something wrong? For example, after the collapse command, the results of "sum prop_age1" is:

Variable	Obs	Mean	Std. Dev.	Min	Max
prop_age1	16,702	.371195	.254877	0	1

I'm interpreting this as the average proportion of children aged 0-14 per household is 0.37.

After using the merge command as written above, the results of the "sum prop_age1" is

Variable	Obs	Mean	Std. Dev.	Min	Max
prop_age1	77,744	.44938	.2244921	0	1

I'm interpreting this as the average proportion of children aged 0-14 per household member (which I think is correct as there are 34,929 children divided by 77,727 total = 0.449). But these two values are inherently different. The meaning seems to have changed and therefore when running the multinomial logistic regression, I may get the wrong results.

I want to carry out a multinomial logistic regression with Household Wealth as the dependent

variable, and have these as demographic and socio-economic independent variables: youth_dep_ratio; prop_age1 (0-14); prop_age2 (15-64); Household Size (grouped 1,2,3-4,5-6,7+); Type of residence; Sex of Household Head; highest Educational level attained.

All the other variables listed I have been able to obtain and construct from the PR file. My issue is, if I carry out the multinomial logistic regression with Wealth from the PR file and use the created demographic variables, is the interpretation changed from the odds of a household being from the poorest wealth quintile, to the odds of a household member being from the poorest wealth quintile?

Is there a way to make my unit of analysis the household as opposed to household member by using the PR file? You mentioned I could merge with the HR file, but my version of Stata says there is no room to add more variables. Basically, all the variables I need are in the PR file, but my (preferred) unit of analysis is the household. I will try to merge the PR file onto the HR file, once I figure out how to open the HR file with my current version. If there is any other way around, please let me know!

Apologies for the lengthy post, help would be very much appreciated!

P.S. I've attached a file with the commands I've used for reference.

Kindest regards,

Xavier

File Attachments

1) Ado of Construction of Variables.txt, downloaded 599 times