Subject: Calculating Age structure explanatory variables on Household Wealth
Posted by xrl1g11 on Wed, 16 Nov 2016 14:44:32 GMT
View Forum Message <> Reply to Message

Dear DHS Staff,

I am using the 2011 Ethiopia DHS data set on SPSS. My thesis is about examining the role of age structure on household wealth. I am finding it difficult to construct the variables I need for my multinomial logistic regression analysis (household wealth is a multicategorical dependent variable).
The variables I need are:
youth dependency ratio (household members aged 0-14 and 65+ divided by household members aged 15-64)
number of young dependents aged under 15 per household
number of working age members (15-64) per household
(Other variables are included but I'm stuck with the demographic variables needed).

The information needed, for example, to create the new variable Number of Young Dependents per household, requires information on the age of household members and the household size (which includes information on the no. of households as well). I've recoded the variable age of household members into three distinct groups: young dependents (aged 0-14), working ages (15-64) and old dependents (65+). I've cross tabulated this new variable with the number of the de jure household members. The table produced shows me the information I need to now calculate the number of young dependents per household as it gives me information on how many household members aged 0-14, 15-64, 65+ live in a certain household size. However, I just can't seem to extract the information from the cross tab, and divide the number of household members under 15 by the number of households.

I attempted to export the cross tabulation to excel, and then manually do the calculation to calculate average number of children per household by household wealth. But I do not know if that's the appropriate way (I was thinking to calculate it and then import the results back into spss, but I'm unsure if that will work).

I'm certain this should be relatively easy to perform on SPSS, but I just can't seem to figure out how to do it on SPSS. I hope it is possible on SPSS, as I don't think what I want to obtain is complex. I want to say for example, that households with a greater number of dependent children are more likely to be from poorer households than wealthier households. And that households with a higher youth dependency ratio are more likely to be poorer than richer.

I hope this made sense, I just need some solid guidance as to how to create the variables I need, using information from the household recode file, or the household member file. My unit of analysis is the household and ideally I want to use the household recode file, but I'm having even more trouble obtaining the information I need, as the age of household members has 22 response variables (HV105$01-$22) which makes things more tougher for me. I'd be happy to give any more information.

In short, I need to know how to create a youth dependency ratio, number of young dependents per household and number of working ages per household from the 2011 EDHS data set.

Regards,

Xavier

Subject: Re: Calculating Age structure explanatory variables on Household Wealth
Posted by Bridgette-DHS on Wed, 16 Nov 2016 15:31:31 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

Quote:I do not use SPSS. To do this in Stata you could just use the following lines. I am in effect omitting people whose age is not given.

```
set more off
use e:\DHS\DHS_data\PR_files\ETPR61FL.dta, clear
sort hhid
save e:\DHS\DHS_data\scratch\ETtemp.dta

gen age=.
replace age=3 if hv105<98
replace age=2 if hv105<65
replace age=1 if hv105<15

gen age1=0
gen age2=0
gen age3=0

replace age1=1 if age==1
replace age2=1 if age==2
replace age3=1 if age==3

tab1 age*

gen hhsize=1
replace hhsize=0 if hv105>=98

collapse (sum) hhsize age1 age2 age3, by(hhid)
gen prop_age1=age1/hhsize
gen prop_age2=age2/hhsize
gen prop_age3=age3/hhsize
gen youth_dep_ratio=age1/age2
gen oldage_dep_ratio=age3/age2
gen dep_ratio=(age1+age3)/age2

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

drop _merge

* then save; you could also merge with the HR file

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 598 times

---

Subject: Re: Calculating Age structure explanatory variables on Household Wealth
Posted by Bridgette-DHS on Thu, 17 Nov 2016 18:52:09 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

Quote:Hi Xavier---Glad it worked. Variables such as prop_age1 are household-level variables that after the merge are coded on to the records for each person in the household. It could be good to have them there for some kinds of analyses. For example, you might have a hypothesis that children who live in a household with a high child dependency ratio are less likely to be in school. This would be a kind of contextual effects model.

If you really want just one record per household, the easiest way to do that after the merge would be with the line "keep if hvidx==1" or "keep if hv101==1". The latter will just keep the line for the household head (who is usually, but not always, the person with hvidx=1). Hope that will give what you want. Otherwise, to get the mean value of prop_age1 from the merged file you could do something such as "summarize prop_age1 if hvidx==1" to get just one value per household.

Note that the summarize commands should include weighting by hv005, or you should use svyset and svy. I did not need to use weights when constructing prop_age1, etc., because everyone in the same household has the same weight. Good luck.

Subject: Re: Calculating Age structure explanatory variables on Household Wealth
Posted by xrl1g11 on Fri, 18 Nov 2016 01:34:51 GMT
View Forum Message <> Reply to Message

Dear Tom,

Thank you very much! Everything worked out fine.

I have one last question, I would like to include a variable in my regression analysis that is a measure of household health or household member health from the PR file. I created child stunting as a health variable (from a coding that was listed on this forum), which worked, but it has reduced the sample size from 77,727 to only 10,000 (as it reflects only children under 5). Also, at the household level, the variable cannot be created as the head of household cannot be a child under 5. One variable from the PR file that I feel may suffice, would be the BMI variable (I believe HA40 is BMI for women and HB40 for men).

Is it possible to create a separate BMI variable (one for male, one for female, and possibly one for child) which is categorized into under 18.5 BMI (indicator of chronic malnutrition) and normal BMI range? If so, what would be the coding for it?

I'm looking for a good general indicator for health and aim to include it in the regression analysis (along with the other variables which have all been successfully created); I would like to examine how household health or household member's health is related to their wealth level. I'd imagine if BMI was the variable, the odds of those with a BMI less than 18.5 are more likely to be from poorer households as opposed to richer households (something along those lines).

If I could get help with creating this variable or another potential health indicator that is commonly used by the DHS, that would be great!

P.S. I have also merged IR into PR to obtain some health variables from the woman's recode but ideally i'd like to just work with the PR file.

Any guidance would be greatly appreciated! Thanks for your time.

Regards,

Xavier

Subject: Re: Calculating Age structure explanatory variables on Household Wealth
Posted by Bridgette-DHS on Fri, 18 Nov 2016 15:39:13 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

Quote:Hi Xavier--I have modified what I sent yesterday to include the means and minima of the household's BMI scores as household variables. Please note that the BMI scores indicate nutritional status, which is not the same as health. Stunting, underweight, and wasting are generally preferred over BMI as indicators of poor child nutrition. Also, for children, BMI (hc73) is given as a z score. Critically low values would be those that are <-200. I would suggest using cutoffs such as 18.5 (or -200) AFTER transferring the household values to a household file, rather than before, but you could certainly apply the cutoffs before.

This procedure, using collapse and merge, provides a general strategy for constructing household-level variables, cluster-level variables, etc. You may find other indicators that are better than BMI. Note that in the final line I did not use weights. For substantive conclusions you would use weights. Good luck!

```
set more off
use e:\DHS\DHS_data\PR_files\ETPR61FL.dta, clear
sort hhid
save e:\DHS\DHS_data\scratch\ETtemp.dta

gen age=.
replace age=3 if hv105<98
replace age=2 if hv105<65
replace age=1 if hv105<15

gen age1=0
gen age2=0
gen age3=0
replace age1=1 if age==1
replace age2=1 if age==2
replace age3=1 if age==3

tab1 age*
gen hhsize=1
replace hhsize=0 if hv105>=98

* construct bmi for women; remove tail; same for men and children
gen bmi_women=ha40
```

```
replace bmi_women=. if ha40>4000

gen bmi_men=hb40
replace bmi_men=. if hb40>4000

* bmi for children is coded as a z score
gen bmi_children=hc73
replace bmi_children=. if hc73>500

* now construct the household means and minima of these bmi scores

gen bmi_women_mean=bmi_women
gen bmi_women_min=bmi_women

gen bmi_men_mean=bmi_men
gen bmi_men_min=bmi_men

gen bmi_children_mean=bmi_children
gen bmi_children_min=bmi_children

drop bmi_women bmi_women bmi_children

collapse (sum) hhsize age1 age2 age3 (mean) *mean (min) *min, by(hhid)
gen prop_age1=age1/hhsize
gen prop_age2=age2/hhsize
gen prop_age3=age3/hhsize
gen youth_dep_ratio=age1/age2
gen oldage_dep_ratio=age3/age2
gen dep_ratio=(age1+age3)/age2

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

drop _merge
* then save; you could also merge with the HR file

summarize prop* *mean *min if hvidx==1
```