Subject: Re: Creating Groups within datasets
Posted by Bridgette-DHS on Mon, 22 Jun 2015 16:27:44 GMT

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

I will illustrate how to do this in Stata with the IR file from the Ethiopia 2011 survey, ETIR61FL.dta. Say you want to use three variables, which in your case would be age in five year groups (v013), region (v024), and highest educational level (v106). There would be 7x11*4 = 308 combinations. This is more than the 80 you mentioned but will serve to illustrate.

The command to construct the joint variable, which I would call "age_region_ed" would be "egen age_region_ed=group(v013 v024 v106), lname(age_region_ed)". The "lname()" option will construct category labels, in this case with the name "age_region_ed". (It is possible, and convenient, to have the same name for the label as for the variable.) Women 15-19 in tigray with primary education would be category 2 of the joint variable. To do a regression, for example, limited to those women, you would include "if age_region_ed==2". There are 326 such women. The number of women in some groups is very small. 16 groups, in fact, have NO women. For that reason there are 292 categories, rather than 308.

You can recode the separate variables before the "egen group" command or you can combine categories of age_region_ed after it is constructed, in either case with usual recode commands.

The five-year age interval that I believe you want would be 15-19, which includes completed years of age 15, 16, 17, 18, and 19, rather than 15-20.

Let me know if you have other questions.