Subject: Missing values for anemia
Posted by clarapd on Fri, 19 Feb 2016 21:36:28 GMT

Hi everyone, I'm new in this forum.

I'm trying to use the prevalence of anemia among women of reproductive age in several Sub-Saharan countries in a regression.

First, I merged the individual dataset with the household dataset with v001 and v002. Now, to match the anemia data (ha57_x, I understant that each one of these variables in linked to a woman in reproductive age of the household) with each woman I'm using the variable v003 (respondent's line number) and the ha0_x (Index to the household schedule):

gen anemia=ha57_1 if ha0_1=v003
replace anemia=ha57_2 if ha0_2=v003 ...

I'm doing something wrong because the number of missing values I get is always higher than in the report of the survey.

Can someone help me?

Many thanks,

Clara