
Subject: Using community-level variables in regression models

Posted by [Lizzynaija](#) on Thu, 16 Jul 2015 14:25:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear DHS researchers,

I am analyzing the association of community-level variables with my outcome, neonatal death, in the 2013 Nigeria DHS. Most of these variables do not already exist within the dataset, so I created them using the `collapse` command to obtain the means/aggregates of the individual level variables at cluster level. I am now trying to work with them in logistic regression models, and I am not sure if I am using them in the correct way.

For example, I created a variable to represent the proportion of people that are uneducated within in a cluster: by creating a 0/1 variable (where 1 = uneducated). Collapsing on this variable gave me the mean of the 0/1 variable, which is the proportion of people within each cluster that are uneducated. And so on for the other variables.

I would now like to use these within my regressions, but not sure about whether to use as a continuous var., or whether to categorize? I tried using the community-level variables as continuous variables, but was not too sure about the interpretation. However, if yes to categorizing, should I use a median split vs. tertiles vs. quartiles? And also how to create these categories correctly - I tried using the `xtile` command, but I am not sure if this is doing what I need it to do.

Also, I would like to ask if it is mandatory to use the `svy: logit` for my regression analyses?

Finally, could you help me with the correct commands to turn off the Stata scientific notation? I keep getting output like "1.2e+04" which is making it difficult to properly calculate my rates.

Thank you in advance for your help,
Elizabeth
