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

Subject: Re: Using community-level variables in regression models
Posted by Reduced-For(u)m on Thu, 16 Jul 2015 21:31:11 GMT
View Forum Message <> Reply to Message

It seems like investigating the determinants of neonatal mortality in Nigeria has been a popular thing to do lately. Some comments on your analysis plan:

1) Continuous/Categorical: this is up to you. In general, I don't think there is much to gain from turning a perfectly good continuous variable into a categorical one, but that is just my opinion. If you decide to go categorical, you should probably use no more than a few categories, otherwise you are likely to lose a lot of power. Plus, at the cluster level, there will be a lot of noise in your community estimates, and categorizing them across lots of bins is probably not helpful, because

the more bins you have the more likely any given cluster is placed in the wrong bin.

2) You need to use the "svy" prefix for two reasons: in part to get the proper weights so that over-sampled populations aren't overly influential in your regressions, but more than that (in this context) to get standard error estimates that are appropriate (without accounting for the clustering, your standard errors and p-values will be too small.

3) You usually can't directly interpret the results from a logistic regression on either a continuos OR categorical variable without transforming them in some way. You need to turn them into something like marginal effects or relative risk ratios or something like that. I like marginal effects, but that is a preference and not universal. Stata can do this if you use the mfx command* or some other options. If you don't know how to interpret these, you will need to do some background reading. If you are getting coefficients in the tens of thousands range, you are likely either mis-specifying something or looking at a coefficient that still needs to be transformed in some way to be interpretable.

 *http://www.stata.com/support/faqs/statistics/marginal-effec ts-methods/

Subject: Re: Using community-level variables in regression models
Posted by Lizzynaija on Sun, 19 Jul 2015 21:02:17 GMT
View Forum Message <> Reply to Message

Thank you very much Reduced-For(u)m - I think these explanations help with thinking through my analysis.

If I may ask another question, which is best to use for multilevel modeling in Stata - xtmixed or gllamm? I have been reading some resources online, but am still unsure as to the correct command, especially with the need to apply survey weights in the regressions.

I would appreciate insight on this, and if possible some sample code that shows how this is done.

Best regards

Subject: Re: Using community-level variables in regression models
Posted by Reduced-For(u)m on Mon, 20 Jul 2015 20:24:51 GMT
View Forum Message <> Reply to Message

I think that is a question for the Stata forum. My understanding is that the "mixed effects" command suite in Stata 14 replace xtmixed (though xtmixed still works).

Here are some resources, but further questions should go to the Statalist and not here.

http://www.stata.com/help.cgi?xtmixed

---

http://www.stata.com/manuals13/meme.pdf#memeRemarksandexampl es

---

Subject: Re: Using community-level variables in regression models
Posted by Krishna on Mon, 06 Feb 2017 07:04:00 GMT
View Forum Message <> Reply to Message

Dear Liz..
Greeting. Please tell me the syntax of community level variable.

Subject: Re: Using community-level variables in regression models
Posted by Krishna on Mon, 06 Feb 2017 07:07:10 GMT
View Forum Message <> Reply to Message

Please tell me how to create community level variable in SPSS kids file

---