
Subject: Re: pooling countries to run fixed effect
Posted by [JaneQuan](#) on Thu, 17 Jun 2021 12:13:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Bridgette-DHS wrote on Mon, 14 December 2020 10:26
Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

You probably have a variable in the pooled file that is called "survey" that takes the values 1 through 10. If not, I recommend that you construct such a variable.

Fixed effects for survey just means that you include "survey" as a categorical variable in the model. That is, using the full file, you include "i.survey" as a covariate on the right hand side of the regression. I agree with your advisor on including such effects. This gives a different intercept for each survey.

When you pool the surveys like this you need to construct new cluster and stratum variables and you may want to redefine the weights. These components all go into an svyset command and "svy:" is included in front of the estimation commands. You should find several forum postings on how to do that.

Hi Bridgette!

I am also using the pooled DHS data for a pooled logit model, and I need to specify the "cluster" to use cluster-robust standard error, since the disturbance of the same individual in different periods may have autocorrelation.

Because I pooled data, so I should reconstruct the cluster (this part is not a problem to me), but when I check the description of the variable cluster(v001), it recommends that I should use it with the variable STRATA(V022).

So I also checked the variable STRATA(V022), and then it says "The DHS Program recommends using STRATA along with the variable PSU (V021) to account for the impact of the sample design clustering on the estimates of variance and standard errors. ". --To here, I am confused. And I checked V021, V022, V001 from the data, it seems there is no difference among these three variables. So my questions are:

1. what's the difference among those three variables, especially between variables V021 and V001?
2. Should I manipulate or weight the variable "cluster(V001)" in order to use it in the logit model? How?
3. If I need to construct a new STRATA variable, then I can use the do_file from this link, right?
4. I checked the "Guide to the DHS Statistics", and it seems the variables that I am using in my analysis has no need to use the command "svyset". But there is one variable-"HV245 (hectares of

all need to use the command "svyset" no matter what variables we are using?

Thank you in advance!
Regards.

