
Subject: pooling countries to run fixed effect
Posted by [lekaicasi46](#) on Sat, 12 Dec 2020 18:16:38 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello! I have an appended dataset that contains 10 countries that have the domestic violence module. In order to study the determinants of seeking help, i need to run a regression with seeking help as the dependant variable and some explanatory variables. I have 10 countries and for each country there is 2 years of data, however its not panel data its cross sectional data. My advisor told me i should pool the countries together and run fixed effects but i only know how to run separate regression for each country. How can i do that? how will my regression equation be?

Subject: Re: pooling countries to run fixed effect
Posted by [Bridgette-DHS](#) on Mon, 14 Dec 2020 15:26:02 GMT
[View Forum Message](#) <> [Reply to Message](#)

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.

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.

Subject: Re: pooling countries to run fixed effect
Posted by [Bridgette-DHS](#) on Thu, 17 Jun 2021 14:36:02 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The variables v001 and v021 are exactly the same in virtually all surveys. There are a handful of old surveys in which one of them is missing, in which case you have to use the other. (For example, if v021 is empty, you would have to use v001.) I believe there is one old survey from Egypt in which v001 and v021 differ, and priority should be given to v021. My general rule would be this: use v021 when it is present, and when it is not, use v001. That will cover all surveys. However, I believe v001 is safe for all surveys except that old one in Egypt.... Similarly, in most recent surveys v022 and v023 are identical and are the stratum. Either can be used. However, for some surveys the stratum variable is different. There is a file in our GitHub site that gives the strata for all surveys.

If "stratumid" and "clusterid" are the correct variables in each survey, then you can use "egen group" to construct the combined ID's as "egen clusterid_all=group(clusterid survey)" and "egen stratumid_all=group(stratumid survey)". Then construct svyset. These steps have appeared on the forum several times. Adjustments to the weights have been discussed on the forum many times, along with cautions about pooling surveys. Within DHS, we pool surveys when analyzing a variable for which there are very few respondents in a single survey, or when analyzing trends within a single country, or when analyzing differences between surveys or countries.

You do not need to combine the clusters and strata into some kind of new variable, if that's what you were thinking. Svset and svy will properly nest the clusters within the strata, and should be used for any estimation command regardless of what variables are in the model. The weights, clusters, and strata are characteristics of the cases and are determined by the sample design. They have nothing to do with any specific variables. Hope this is helpful.

Subject: Re: pooling countries to run fixed effect
Posted by [JaneQuan](#) on Fri, 18 Jun 2021 10:31:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

This is very helpful!
Thank you so much:)
