

Thank you.

1a) -I am pooling the data sets because I will use difference in difference method with Pseudo(pooled) cross section data. I want to see the effect of an education reform. the policy change happened in 1997 and therefore my control and treatment group are not affected in 2003 data. and in 2008 data, my treatment group affected and control group not affected.

So the data set i want to create in the end will look like:

number of observation	ID variables	year	2003	2008	independent variables
1	2003	1	0	age sex
2	...	2003	1	0	age sex
3	...	2003	1	0
.....					
25	2008	0	1	age sex
.....					
45	2008	0	1
.....					

- I will add the data set to the post, it is not in the format i wanted. Could you please check it? the data set attached to the mail is not in the format i wanted. i do not have the dummy variables. and instead of year variable i have survey phase variable.

1b) Should i do 1b before pooling the data? By the way, v001 in my data set is corrected before pooling the data. So my PSU, primary sampling unit is v001? If you notice, in the normal data set it is V001 but in mine it is v001. what i mean is that some of the variables not standard in both data sets before pooling. and therefore before pooling the data set they were corrected. But how should i include v023 in this data set?

should i ignore the other variables such as v021, v022 and v004.

1d) v005 was also different in the real survey but we corrected it like:
v005= V005/1000000. this is done at cross section level. then two data set is combined.
should i change it at cross section level? or i can do it at pseudo panel level?

4) how do I understand strata is V022 format not in V023 format. Because i have both variable in the full data set of TDHS 2003 and 2008.

5) We combine the data set in spss. I will add the pooled format to the mail. Could you please check it?

File Attachments

1) [2003-2008.zip](#), downloaded 1079 times
