Subject: Dealing with country-specific codes with panel data across 10 countries
Posted by jane_cheatley on Tue, 26 Jul 2016 13:52:53 GMT
View Forum Message <> Reply to Message

Hi,

I have recently appended DHS surveys undertaken between 2000-2015 in 10 Sub-Saharan African countries (i.e. Uganda, Tanzania, Rwanda, Malawi, Mozambique, Namibia, Zambia, Ethiopia, Kenya and Madagasar). I am going to run a logistic regression to understand key variables that have a significant impact on under-5 mortality. In the literature, it is evident that type of toilet facility, ethnicity and place of delivery are key factors. A review of past articles has also provided ideas on how responses could be grouped to make interpretation of results easier (e.g. For type of toilet facility - flush, other improved facility, unimproved facility and no facility).

In my appended data set when I tab, for example, v116 (type of toilet facility) I have a few labelled variables and a range of country specific codes. To understand what each code meant for each country and year I went through the Final Reports, which had questionnaires at the bottom. It appears that the same response has a different numeric code across countries and sometimes across years. To overcome this, I went into each country's specific Stata file which I have also created (i.e. In addition to the Master file with all years and countries, I created country-specific Stata files with all years - e.g. Ethiopia_Master with DHS 2000,2005 and 2011 data appended) and created dummies for type of toilet facility by year (e.g. using code such as gen flush_toilet=1 if v116==11&year==2000, replace flush_toilet=0 if v116!=11&year==2000).

I am still having the following issues:

- even if the above method were to work, I am not sure how to combine the dummies across the years in a way that is meaningful (e.g. summing does not make sense for dummies)

- often the codes in the questionnaire do not match those in the data and thus certain codes are still missing.

In academic papers, it seems that this has been done quite often, however, the way I am going about it seems very time-consuming and prone to error.

I would be very grateful if anyone would be able to steer me in the right direction.

Best,

Jane