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

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

Subject: Re: Dealing with country-specific codes with panel data across 10 countries
Posted by Bridgette-DHS on Tue, 26 Jul 2016 14:46:21 GMT

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

What you are describing is indeed very time consuming. We had to do much the same thing for Methodological Report 15 in 2014 (https://www.dhsprogram.com/pubs/pdf/MR15/MR15.pdf). There are probably many ways to approach it, but we did it as follows.

The first attachment is an excel file with one sheet for each country. The sheets have columns for the successive surveys and panels for the variables. The final column of the sheet is a "harmonized" coding. We were working with the HR files, with one record per household. You would probably be using the IR or KR files, but that would make little difference in the steps..

The second attachment is a Stata program I wrote to do the recoding. Basically it converts the excel sheets into recode instructions for every variable and every file. It's not a simple program but there are lots of comments. Hope you can figure out the logic--I can't provide much support beyond this. Good luck!

File Attachments
1) Copy of Harmonizing data-20140424.xlsx, downloaded 921 times
2) CWI_recoding_do_3June2014.txt, downloaded 1190 times

Subject: Re: Dealing with country-specific codes with panel data across 10 countries
Posted by jane_cheatley on Wed, 27 Jul 2016 12:35:07 GMT
View Forum Message <> Reply to Message

Thanks Bridgette. I'll give it a go.

Subject: Re: Dealing with country-specific codes with panel data across 10 countries
Posted by jane_cheatley on Mon, 01 Aug 2016 09:00:35 GMT
View Forum Message <> Reply to Message

Hi Bridgette,

Thank you again for passing on those documents. In regards to the Excel document you sent, do you have similar data for Ethiopia and Rwanda?

Best,

Jane

Subject: Re: Dealing with country-specific codes with panel data across 10 countries
Posted by Bridgette-DHS on Mon, 01 Aug 2016 12:08:25 GMT
View Forum Message <> Reply to Message

Following is a response from Tom Pullum:

No, what I sent you includes all of the countries that were in that analysis.

Subject: Re: Dealing with country-specific codes with panel data across 10 countries
Posted by kingx025 on Mon, 09 Jul 2018 00:15:23 GMT
View Forum Message <> Reply to Message

Because the responses and codes for type of toilet facility are country (and sample) specific, anyone working with this variable across multiple surveys may wish to work with the variable in IPUMS-DHS. We attach the household file to other file types (women, children, births) and consistently use the HV205 DHS variable for the IPUMS-DHS variable TOILETTYPE (also available as HV205). To retain all detail across all samples, the IPUMS-DHS variable is a four digit variable. The first digit groups responses into broad categories such as "No facility," "Flush toilet," "Non-Flushing Toilet," "Pit Toilet Latrine," and "Unimproved Toilet." Additional detail in the second, third, and fourth digit allows the researcher to distinguish between, for example, "Pit latrine without slab or open pit" and "Ventilated improved pit latrine."
You can see the codes and variable labels for the variable on type of toilet facility in IPUMS-DHS here:
 https://www.idhsdata.org/idhs-action/variables/TOILETTYPE#co des_section
An X means that value is available in a given sample, listed in the columns at the top of the table.

Using the pre-harmonized data on toilet facilities (and other variables with country specific responses, such as on source of drinking water) from IPUMS-DHS can save time and prevent inadvertent errors. Nine of the ten African countries mentioned by the original poster are currently included in IPUMS-DHS, and the tenth, Namibia, will be available in IPUMS-DHS by Fall 2018.

Miriam King