Subject: v001 and DHSCLUST
Posted by cansulancaster on Wed, 14 Aug 2024 08:28:00 GMT
View Forum Message <> Reply to Message

I have appended Benin DHS 6 Women's file (BJIR61FL) and Benin DHS 7 Women's file (BJIR71FL) . That is, I have appended cross-sectional dataset for Benin.
My question is the description of my problem: I will merge appended Benin DHS with another dataset, which is geo-coded and matches with Benin's survey years. For this dataset, I have longitude and latitude information. Using long and lat information, I have created cluster_id on STATA such that egen cluster_id=grouo(longitude latitude). This code automatically generates clusters (let's call it cluster_id). My intention is to merge this dataset with Benin DHS using clusters and relevant years because I will work at the cluster level. I have previously merged this dataset, for example with geo-coded Afrobarometer, by generating the cluster_ids in exactly the same way. So, I had no problems. However, in DHS, there is v001 (cluster number) variable in the IR files. In Benin DHS 6, there are 750 clusters from 1 to 750 (on STATA: contract v001) . Similarly, in benin DHS 7, there are 555 clusters from 1 to 555. When I append the DHS 6 and DHS 7, there are naturally 750 clusters from 1 to 750. However, as you can notice, v001 is 1,2,3,... in DHS 6 and 1,2,3,4... in DHS 7. They look like the same, even if they might be entirely different in location-wise. My question is that, can I use v001 to such a merge after appending? Or do you suggest me to merge IR datasets with Benin DHS 6 geographic data (BJGE61FL.shp) Benin geographic data (BJGE71FL.shp) using the following code (I provide the code for only DHS 6, not to get confused)
* Convert shapefile in stata files
* First change the working directory where you store your shapefile
cd "C:\Users\cansu\Dropbox\Data_Cansu\JMP\DHS Datasets- JMP\Benin"
shp2dta using "BJGE61FL.shp", data("Benin_Data.dta") coor("Benin_Coor.dta") replace
*Use Benin_Data generated by the previous line of the code, and rename cluster number to merge it with Women's file:
ren DHSCLUST v001
*Use Benin DHS 6 Women's file
merge m:1 v001 using "C:\Users\cansu\Dropbox\Data_Cansu\JMP\DHS Datasets-JMP\Benin\Benin_Data.dta"
*Since now I have LONGNUM and LATNUM coming from Benin_Data File, I can proceed with generating cluster_id such that:
egen cluster_id=group(LONGNUM LATNUM)
sum v001 cluster_id

Variable	Obs	Mean	Std. Dev.	Min	Max
v001	13,407	359.1057	223.2945	1	750
cluster_id	13,407	370.0516	229.8351	1	747

As you may notice, there is a slight difference between two variables. I do not understand why. Overall, after appending DHS 6 and 7, should I proceed with v001 (I suspect that that would be wrong because there are 1s in round 6 and 7, they cannot be the same?). Thanks a lot!!

Subject: Re: v001 and DHSCLUST
Posted by Anonymous on Mon, 21 Apr 2025 07:34:51 GMT

View Forum Message <> Reply to Message

Your approach of merging the DHS datasets with their respective geographic datasets (BJGE61FL for DHS 6 and BJGE71FL for DHS 7) is more advisable than using v001 for merging the appended dataset.