
Subject: DHS CLUSTERS

Posted by [Magashi](#) on Wed, 24 Jul 2024 05:50:18 GMT

[View Forum Message](#) <> [Reply to Message](#)

I am planning to pool Tanzania DHS data for 2004/05, 2009/10, 2015/16, and 2022 for a multilevel analysis of my variables of interest. The TDHS datasets have different numbers of clusters: 475 clusters for 2004/05 and 2009/10 (based on the 2002 Population and Housing Census), 608 clusters for 2015/16 (based on the 2012 Census), and 629 clusters for 2022 (also based on the 2012 Census).

When conducting a multilevel analysis at the individual level (level 1) and cluster level (level 2) for the pooled data, is it acceptable to use the cluster numbers (v001) as they are after appending the data? For example, my Stata command for the multilevel analysis is:

```
```stata
mixed depvar indepvar || v001:
```
```

However, I am concerned that the same cluster number (e.g., cluster 1) could refer to different clusters in different survey rounds. Should I transform the cluster variable to ensure unique identification across survey rounds? Your insights would be greatly appreciated.

Subject: Re: DHS CLUSTERS

Posted by [Janet-DHS](#) on Wed, 24 Jul 2024 19:27:37 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS staff member, Tom Pullum:

You are right. It is necessary to assign unique ID codes to the clusters--and also to the strata. If you are using Stata you can do this with the "egen group" command. First assign the numbers 1, 2, 3, 4 to the respective surveys and call that code "survey". Then in the pooled file enter a line like "egen clusterID=group(survey v001)". Also enter "egen stratumID=group(survey v023)". Then in svyset you would use clusterID and stratumID as the cluster and stratum variables. If you were not using Stata, you would do something similar. For most purposes, you can leave v005 alone.
