
Subject: How to add village level characteristics
Posted by [mdixit](#) on Wed, 27 Jul 2022 05:43:17 GMT
[View Forum Message](#) <> [Reply to Message](#)

I am using DHS 2015-16 and 2019-2020 rounds for studying the impact of a policy in one of the states in India. I am using a DID framework for estimating the impact. though I have been able to add number of individual level control variables but I am unable to add village level control variables. I am trying to use `i.clusternumber(v001)` to address the same issue but the regression takes a very long time to execute since there are more than 5000 clusters in my analysis. What needs to be done to address this issue? and whether cluster in DHS depicts a village or not?

Subject: Re: How to add village level characteristics
Posted by [Janet-DHS](#) on Mon, 01 Aug 2022 16:27:14 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Cluster-level control variables would usually be interpreted as something like the proportion of the households in the bottom two wealth quintiles or the proportion of women with no schooling, etc. You can also attach cluster-level variables using the geographic covariates data file.

Note that the cluster id numbers are nested within states (v024). A unique cluster-level ID could be constructed with `"egen cluster_ID=group(v024 v001)"` and then a "fixed effects" model would include the term `"i.cluster_ID"`. You definitely should not use fixed effects for clusters. Apart from the time required to run a model with 5000 clusters (if it would run at all), the model would be seriously over-fitted or over-determined and your substantive covariates would become insignificant. I recommend cluster-level variables as described above and/or a multi-level model, which will indicate how much of the total explanatory power is individual-level and how much is cluster-level.

The correspondence between clusters and villages (or neighborhoods, in urban areas) is very loose. Some clusters actually include more than one village, for example. Such a correspondence is often assumed but you can't rely on it.
