Subject: Selecting only independent data
Posted by 77rana on Fri, 19 Oct 2018 22:21:42 GMT
View Forum Message <> Reply to Message

I am working on a data file from EDHS2014. The way data is collected, it seems like it contains dependent data. Which means that more than one member of a household were interviewed. I am trying to select only one person per household to make it possible for me to run analysis I need. My questions:

1- Am I correct in assuming that there might be more than one household member in the dataset? for example, father and son would show up as separate rows.
2- What variables should I use to randomly select a sample of one person per household. I am potentially looking at PSU and hhid combined, but I see that hhid's often appear more than once (range 1-3).

I am really confused. please help.