Subject: Re: Link the DHS Individuals (IR) with DHS HIV (AR)
Posted by Bridgette-DHS on Wed, 18 Mar 2015 11:56:06 GMT
View Forum Message <> Reply to Message

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

The answer you refer to is about six months old, not several years.... Like many Stata users, I have stayed with old syntax for the merge command, rather than the one that was introduced with Stata 11 or 12. Both the old and the new versions produce a diagnostic variable called "_merge". The variable has three values of most interest, 1, 2, and 3. "1" means the case was only in the "master" file, i.e. the first one in the sequence. "2" means the case was only in the "using" file, i.e. the second one in the sequence. "3" means the case was in both files.

Usually, when I do a merge, I want just the cases with _merge==3, but that is definitely not always the case. Sometimes I also want the cases with 1 or 2 (almost never 1, 2, and 3, but even that could happen). You have to think about what makes sense. Say, for example, that you wanted to merge the children under 5 in the household survey (the PR file) with the children in the KR file. Say you started with the KR file and then merged the PR data (again for children under 5) with the KR file. The PR and KR files would be the master and using files, respectively. The PR file includes all children in the household, including children whose mother is not a resident of the household. Those children will get _merge=1. The KR file includes children who do not live in the same household as the mother, and they will get _merge=2. If the mother and child both live in the household, the child will get _merge=3. They are the only ones for whom you will get both PR and KR variables, so you probably would want to keep just them--but that's not necessarily the case.