

Using STATA to Match/Merge Two Files¹

Following is an example of matching two files with STATA. STATA commands are in bold. Variables and items that would change for your program are in lower case and not bold. Explanatory comments and documentation begin with asterisks. Please refer to the STATA manuals and on-line help for more information.

This example matches persons to their household. For this example Household ID (hh) is not unique, except within state, so both STATE and HH must be included in the merge statement.

For this example, a text (ASCII) file called "hhper.do" was created using WordPad or Notepad. STATA commands can be run interactively - typing and entering each command, or, programs can be written and run all at once. The second name - or extension of stata program files (which must be an ASCII file) is ".do". The advantage of creating program files is that is easier to save a record of your commands, and to document the file. This is especially important if you are creating a data file that you will use later for analysis. I recommend giving the .do file the same name as the data file that will be saved. This will help you locate the program file that created the data file. To run a .do program Windows in batch mode, use explorer to find the program (.do file), right-click on it, and click on open. This will cause STATA to be opened, and the program will be run. In UNIX, issuing the following command (substituting "filename" with the first name of the file to be submitted - in this case "hhper") **stata -b do filename &**

Stata commands are case sensitive on some platforms.

```
* hhper.do - an example of how to match/merge files in stata.
* The file "person.dta" has a varying number of cases per household.
* Each person is matched properly to his/her state and household.
* 12/22/2000, Patty Glynn. Include information about the project here.

* increase memory - the default for stata is insufficient for large data sets.
set memory 10000k

** get secondary data (multiple records per id), sort, save
** the source of the data should be documented.
use "d:\datadir\person.dta", clear
sort state hh
* save altered file file in different directory so original data are not overwritten.
save "d:\xxdir\person.dta", replace

** get master file, sort. The source of the data should be documented.
use "d:\datadir\house.dta", clear
sort state hh

*** the syntax for the merge command is:
** merge [variable list to merge by] using [file to merge]
** the [variable list to merge by] are variables common to each file.
** the file listed after "using" will be merged with the active file in memory -
** in the case, "d:\datadir\house.dta"
merge state hh using "d:\xxdir\person.dta"
sort state hh personid

** be sure to check to make sure the merge worked as you hoped.
** use the list command to list variables from both files to make sure
** that the records from the secondary file are matched to the mater file as expected.

list [vars to see-state hh personid "variable from hh file", "variable from person file" ]
save "d:\datadir\hhper.dta", replace
```

¹Prepared by Patty Glynn, University of Washington. 12/27/2000 modified 3/5/2001 C:\all\help\helpnew\statamrg.wpd