Matching Files with SPSS (Adding Variables)[1]

One case per ID in each file.
Sometimes you will have variables in different data sets that you want to put together. For example, you have interview data for the same set of people from two time periods, and you want be able to use answers from interview 1 and interview 2 in the same analysis. This requires that you merge the files. In SPSS, the command required for this is "Match Files". Merging is easiest if there is a variable, or set of variables that uniquely identifies each case. In the following example, "ID" is used as that variable. Each file that is to be merged should be sorted in the order of the identifying variable(s). **The "BY" statement is extremely important.** Without it, cases will not be match properly if there are any missing cases in either file. It is always wise to check to make sure that the merge happened properly. Examine variables and cases from the files before and after the merge, to make sure that it happened as you expected and hoped.

```
** Match.sps .
** Substitute 'c:\full\path\filename.sav' with information about where;
** your files are stored, or where you want them to be stored.

get file = 'c:\full\path\interview1.sav' .
sort cases by id .
save outfile = 'c:\full\path\tmp1.sav' .

get file = 'c:\full\path\interview2.sav' .
sort cases by id .
save outfile = 'c:\full\path\tmp2.sav' .

MATCH FILES  FILE='c:\full\path\tmp1.sav'   /
             FILE='c:\full\path\tmp2.sav'   / BY id / MAP .
list var = all .
```

One case per ID in one file, multiple cases per ID in another file.
Suppose you want to merge contextual data with your interview data. The state, county and census tract of where each person lived is coded in your data file. You have used 1990 STF files to get tract level data, and you want to be able to examine this contextual data along with the interview data. There may be more than one person in some tracts, and none in others. In SPSS, this circumstance requires a special kind of merge. The file with ONE case per unique ID is preceded by "table =". The file with varying number of cases per ID is preceded by "file =". It is always wise to check to make sure that the merge happened properly. Examine variables and cases from the files before and after the merge, to make sure that it happened as you expected and hoped.

```
** Mattab.sps .
** Substitute 'c:\full\path\filename.sav' with information about where;
** your files are stored, or where you want them to be stored.
get file = 'c:\full\path\tract.sav' .
sort cases by state county tract .
save outfile = 'c:\full\path\tmpone.sav' .

get file = 'c:\full\path\interview.sav' .
** create a variable so that tracts with no cases can be eliminated after merge .
sort cases by state county tract .
save outfile = 'c:\full\path\tmpmany.sav' .

MATCH FILES  TABLE= 'c:\full\path\tmpone.sav'  /
             FILE=  'c:\full\path\tmpmany.sav  / BY state county tract / MAP .
* Important note: any cases in the "table" file that do not have corresponding
* in the "file" file the will be deleted.  In this example, any tract in.
* "tmpone.sav" that is not represented in 'tmpmany.sav' will not be included in .
* the resulting file .
```