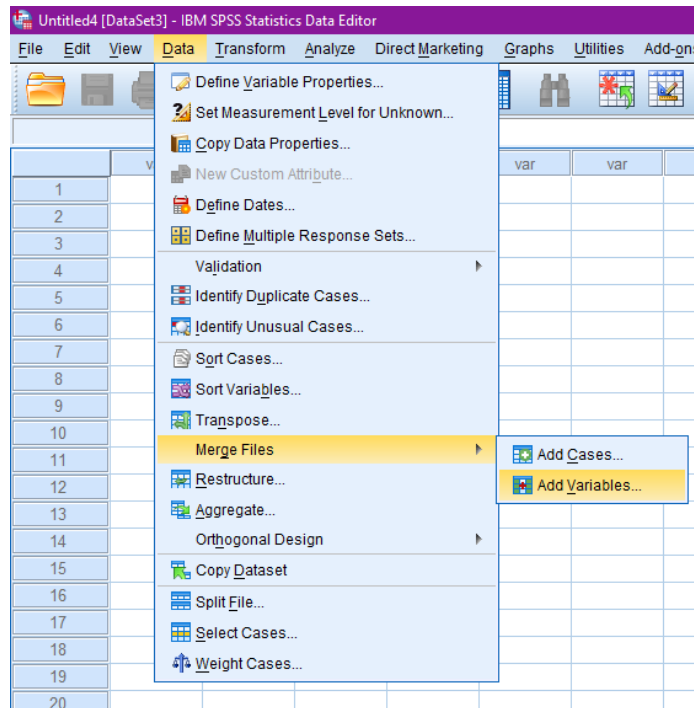


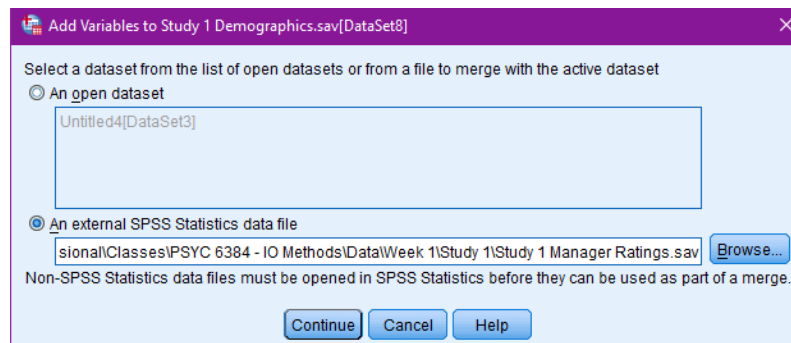
Merging Datasets in SPSS

Sometimes you'll need to merge datasets together in SPSS. **In order to merge two data files, both files must be saved as .sav (SPSS data) format.**

1. You will need to pick a "key variable" on which the datasets will be matched. ***You must sort both datafiles ascending by the key variable and then save them.***
2. Select "Data > Merge Files > Add Variables" from SPSS's main menu.



3. Tell SPSS that you want to merge variables from an external file:



4. Set the merge options appropriately:

The screenshot shows the 'Add Variables from' dialog box in SPSS. The 'Excluded Variables' list is empty. The 'New Active Dataset' list contains variables: gender_Male(*), gender_Female(*), gender_Trans(*), gender_Other(*), age(*), race_Asian(*), race_Black(*), race_Caucasian_Whit..., race_Hispanic_Latino(*), and race_Indian_Asian(*). The 'Key Variables' list contains 'employee_id'. The 'Match cases on key variables in sorted files' checkbox is checked. The 'Non-active dataset is keyed table' radio button is selected. A 'Rename...' button is visible. At the bottom are 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' buttons.

Check this box

This option means that the file you're merging can have *only one row per key* (matching) variable

This is the variable the files will match on

When matching two datasets, one file is called the *keyed table*. The primary dataset can have multiple rows per participant (e.g., five rows for participant #3). However, the *keyed table* can have only one row per participant (e.g., one row for participant #3). All of the variables in the *keyed table* will be appended to the matching rows in the primary dataset (e.g., all of the variables in the row for participant #3 in the *keyed table* will be appended to the end of all five rows for participant #3 in the primary dataset).

5. Click "Okay" and enjoy the merged dataset!

- Note that the merge **will fail** if both datasets are not sorted ascending on the key (see step #1) or if the *keyed table* has more than one row for any participant.