

## EXERCISE 3

## Merging files – Adding cases & variables

---

This Exercise shows you how to open a saved file and how to merge your data with other files. The relevant sections are 2.6 (resuming work on a saved data set), 3.1.2 (reading in SPSS files) and 3.3.4 (merging files).

### Opening your saved file in SPSS

Log in to SPSS as described in Section 2.2. Open the file that you saved under the name *Ex2 Questionnaire Data* from the previous Exercise. Select the radio button **Open an existing data source** in the SPSS opening window and highlight the filename. It may be necessary, especially in a networked computer, to select the **More Files...** option and select another disk drive or folder. Alternatively, select the **Type in data** radio button to open **Data View**, choose **File**, click **Open**, and finally **Data** to open the **Open File** selection box. Change the **Look in:** selection to the folder and/or disk drive where the file has been stored, click the filename so that it appears in the **Filename** box, and then **Open**. Your file should appear in **Data View**, where part of your own data might appear as follows:

	CaseNo	MyName	Age	Sex	Faculty	Status	Stones	Pounds	Kilo
1	1	Alan Brown	25	1	1	1	14	0	

### Locating the larger data set

The file containing the large data set with which you are going to merge your own data is to be found at the following WWW address:

<http://www.psypress.com/spss-made-simple/datasets.asp>

in the file labelled *Ex3 Questionnaire data* within the section *Release 16*. If this file has not already been downloaded on to a file server or hard disk drive for easier access, enter WWW and save the file to a more convenient place such as your hard disk drive. (In order to see the appropriate icons for accessing WWW, reduce the size of the SPSS window.)

### Merging your data with the larger data set

To carry out the file merge, select **Data** → **Merge Files** → **Add Cases...** to obtain the **Add Cases to Ex2 Questionnaire Data.sav** selection box, which prompts you to specify the file (the large data set) from which you want to merge other cases with your own. Click the **An external SPSS data file** radio button and then click **Browse...** to locate the file *Ex3 Questionnaire Data* and click the file name so that it appears in the panel. Click **Continue** to open the **Add Cases From Ex3 Questionnaire Data** dialog box (see Figure 1).

If you have entered the correct variable names in your own data file, the **Add Cases From** dialog box should show just one variable name (*MyName*) in the **Unpaired Variables:** box and all the other variable names should be in the **Variables in New Working Data File:** box. A match will not be found for your variable *MyName*, because the large data set does not contain the names of the participants.

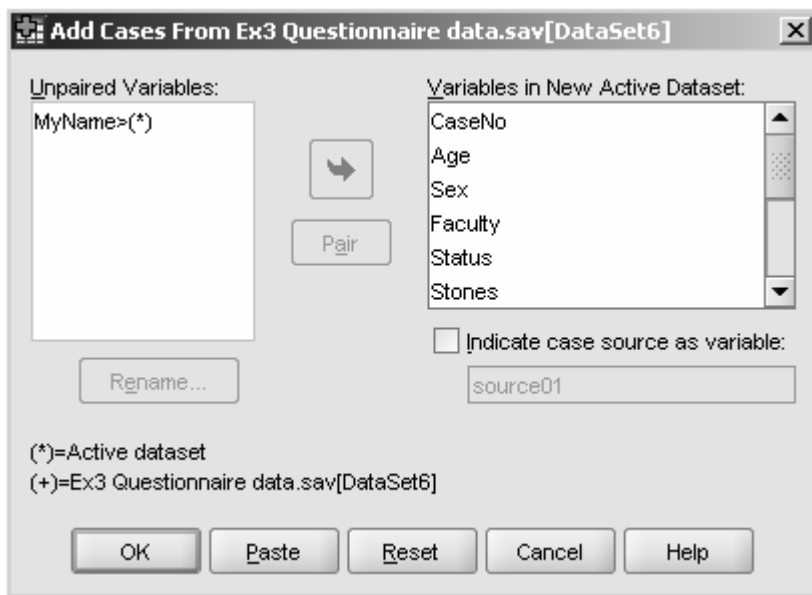


Figure 1. The **Add Cases From** dialog box showing the names of Unpaired Variables (here there is just one, *MyName*) and corresponding names from both files (Variables in New Active Dataset)

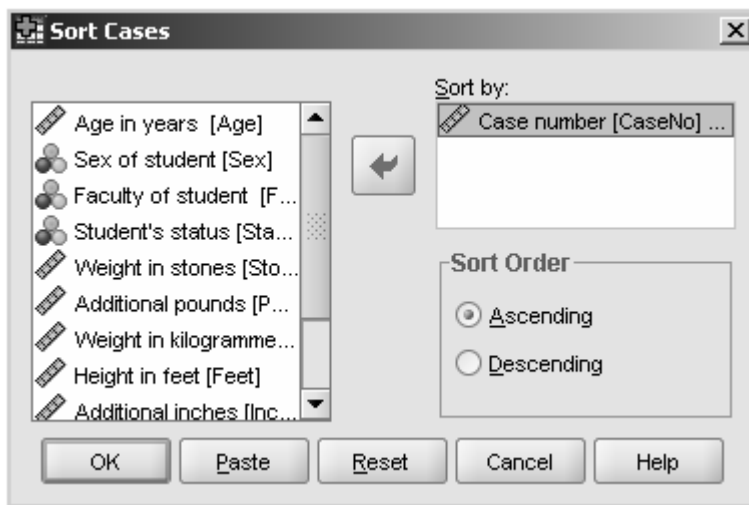
Suppose, however, there had been a mismatch between one of the variable names you had typed into **Data View** and the name of the corresponding variable in the large data set. Suppose that, when you were building *Ex2 Questionnaire Data*, you had typed *Ages* instead of *Age*. The variables *Ages* and *Age* would both have appeared in the **Unpaired Variables:** box. You would have then had to select *Ages* and click **Rename** to obtain another dialog box, allowing you to rename it as *Age*. The correct variable name *Age* would then be transferred to the list in **Variables in the New Working Data File**.

Click **OK** to merge the files. The first few variables in the first two lines of the merged data set might appear as follows (click on the labels icon to see the labels for *Sex*, *Faculty* etc.):

	CaseNo	Age	Sex	Faculty	Status	Stones	Pounds
1	1	25	1	1	1	14	0
2	1	23	1	1	1	11	0

Notice that the same case number appears in the first two rows. Since your own case is being added to the 334 cases already in the data set, change your case number to 335.

Clearly, since yours is the 335<sup>th</sup> case, you want your own data to appear below the others in **Data View**. To rearrange the cases in ascending order of magnitude, choose **Data** → **Sort Cases...** to see the **Sort Cases** dialog box. Transfer *CaseNo* to the **Sort By** panel on the right and click **OK**.



You will now find that your own data occupy the bottom row in **Data View**. Save the merged file as *Merged Questionnaire Data*.

### A warning

In the large data set, the *Faculty* variable was of the numeric type, with values assigned as follows: 1 = *Arts*, 2 = *Science*, 3 = *Medicine*, 4 = *Other*.

Suppose that you had inadvertently assigned 1 to *Science* and 2 to *Arts*, instead of the other way round, and that as a science student, you had recorded a 1 in your own data set. SPSS will not warn you of the discrepancy. Instead, it will adopt your convention throughout the merged data set and all those people who recorded 1s in the larger original data set will now be recorded as scientists, not arts students. You can confirm this by choosing Value Labels from the View menu. All those previously recorded as science students will now be recorded as arts students and vice versa.

When two files are being merged, it is the value assignments **in the first file** that determine those for the entire merged file, even when, as in the present example, the former contains only a single case.

### Another file-merging exercise: Adding extra variables to a file

In Section 3.3.2, we described the use of the **Aggregate** procedure to create a file showing the average marks of students who had taken various courses at a business school. Open the *Students marks* file from the website specified earlier or enter it from the appendix to this Exercise. Follow the **Aggregate** procedure described in Section 3.3.2 to obtain the course averages and the numbers of students taking the courses. Save the new information to a file named *Course average marks*.

Also available from the website or from the appendix to this Exercise is the *Assignment* file which contains the courses to which the lecturers at the business school were assigned. The aim of this Exercise is to add, to the file containing the course averages, the names of the lecturers who were assigned to those courses. The link between the course averages and the lecturers can be made with the **Merge Files** procedure, using *Assignment* as a lookup file to obtain the names of the lecturers who gave the courses listed in *Course average marks*.

- Open *Course average marks*.
- Choose **Data** → **Sort Cases**.

- Fill in the **Sort Cases** dialog box, specifying that you are sorting by *Course*. Keep the sorting order as Ascending. Click **OK**.

Save the sorted file as *Course average marks*.

(It would normally be necessary to check that the other file being used for the merging of data also had its entries in the variable *Course* arranged alphabetically and if not, to use a **Sort Cases...** procedure to do so. In this case, the file *Assignment* already has its *Course* entries arranged alphabetically.)

- Choose **Data** → **Merge Files** → **Add Variables...** (see Section 3.3.4).
- In the **Add Variables to Course average marks.sav** dialog box (see Figure 29), ensure that the radio button for **An external SPSS data file** is active, click **Browse...**, select *Assignment* and click **Continue**.
- In the **Add Variables from Assignment.sav** dialog box, click the selection box **Match cases on key variables in sorted files** and then click the **Active data set is keyed table** radio button.
- Highlight *Course* from the **Excluded Variables** box and transfer it to the **Key Variables** box.
- Click **OK**.

Look at the resulting data file and notice the new column headed *Lecturer*. This shows that the lecturer's name has been paired with each course and the average mark that those students taking the course received.

- **Which of Tom Fielding's courses has the highest average mark and what is the average mark for that course?**

### Finishing the session

Close down SPSS and any other windows before logging out.

### Appendix to Exercise 3 - Student Marks

Student	Course	Mark	Student	Course	Mark
Anne	Accountancy	80	Rebecca	Law	85
Rebecca	Accountancy	78	Susan	Law	91
Susan	Accountancy	87	Fred	Management	70
Anne	Computing	49	Jim	Management	57
Fred	Computing	55	Joe	Management	53
Rebecca	Computing	65	John	Management	41
Susan	Computing	56	Kevin	Management	45
Anne	German	40	Rebecca	Management	43
Fred	German	72	Sebastian	Management	44
Jim	German	73	Jim	Mathematics	66
Susan	German	85	Joe	Mathematics	60
Anne	Graphics	58	John	Mathematics	68
Fred	Graphics	54	Kevin	Mathematics	71
Jim	Graphics	65	Mary	Mathematics	56
Joe	Graphics	55	Sebastian	Mathematics	61
Kevin	Graphics	58	Joe	Politics	62
Mary	Graphics	50	John	Politics	45
Rebecca	Graphics	62	Kevin	Politics	49
Anne	Law	93	Mary	Politics	56
Fred	Law	88	Sebastian	Politics	43
Jim	Law	89	John	Spanish	62
Joe	Law	91	Mary	Spanish	57
John	Law	43	Sebastian	Spanish	53
Kevin	Law	83	Mary	Statistics	70
Rebecca	Law	85	Sebastian	Statistics	62

### Appendix to Exercise 3 - Assignment data file

Lecturer	Course
EvelynBrown	Accountancy
TimRice	Book Keeping
DavidJones	Computing
EvelynBrown	German
JoanSmith	Graphics
TimRice	Investment Management
SarahAlbert	Law
TomFielding	Management
EvelynBrown	Mathematics
TomFielding	Politics
TomFielding	Spanish
TimRice	Spreadsheets
EvelynBrown	Statistics
JoanSmith	The Internet
JoanSmith	Web Management