IT195: SPSS – Using Syntax Commands and Data management

These notes provide an introduction to using the PC (Windows) version of SPSS release 19, which is accessible from the Networked PC service. This Guide makes use of example files which are installed on the Networked PC service. If you do not use the Networked PC service then you can obtain the files from the ITS WWW pages.
Conventions:

In this document, the following conventions are used:

- A **bold typewriter font** is used to represent the actual characters you type at the keyboard.
- A **slanted typewriter font** is used for items such as filenames which you should replace with particular instances.
- A typewriter font is used for what you see on the screen.
- A **bold font** is used to indicate named keys on the keyboard, for example, Esc and Enter, represent the keys marked Esc and Enter, respectively.
- Where two keys are separated by a forward slash (as in Ctrl/B, for example), press and hold down the first key (Ctrl), tap the second (B), and then release the first key.
- A **bold font** is also used where a technical term or command name is used in the text.
Contents

1 Using syntax commands ........................................................................................................3
  1.1 The Open Syntax command .........................................................................................4
  1.2 SPSS Syntax commands .............................................................................................4
  1.3 Running syntax commands .........................................................................................5
  1.4 Making global changes in the Syntax Editor ..............................................................6
  1.5 Saving the contents of the Syntax Editor .....................................................................6

2 Further syntax ..................................................................................................................7
  2.1 Selecting and running several syntax commands at once ............................................7
  2.2 Mixing menu commands with syntax commands .........................................................7
  2.3 Pasting menu commands into the Syntax Editor .........................................................8
  2.4 Getting help with syntax commands .........................................................................9
  2.5 Entering commands into a Syntax Editor ................................................................10
  2.6 Creating an SPSS program ....................................................................................10

3 More on data management ...............................................................................................11
  3.1 Data selection .............................................................................................................12
  3.2 Splitting files ............................................................................................................13
  3.3 Sorting the working data file ....................................................................................14
  3.4 Saving selected variables only ................................................................................15
    3.4.1 Open a file with selected variables only .............................................................15
    3.4.2 Saving selected cases only ................................................................................16
  3.5 Combining files .........................................................................................................17
    3.5.1 Same cases — different variables .....................................................................17
    3.5.2 Same variables — different cases .....................................................................19

4 Being in control — SPSS options ..................................................................................20
  4.1 Command recording ..................................................................................................21
  4.2 Chart options ............................................................................................................22
  4.3 Viewer options ..........................................................................................................22
  4.4 Saving Options settings ............................................................................................22

5 Portability ........................................................................................................................22
  5.1 Command files ..........................................................................................................22
  5.2 SPSS data files ..........................................................................................................22
  5.3 Chart files ................................................................................................................23

6 Further information ........................................................................................................23
  6.1 CIS documents ..........................................................................................................23
  6.2 SPSS manuals ..........................................................................................................23
  6.3 SPSS newsgroups and internet links .........................................................................24
  6.4 Help and advice ........................................................................................................24
1 Using syntax commands

The process of running SPSS by selecting commands with the mouse through on-screen menus and selecting options from dialog boxes is fine when you are exploring your data or experimenting with SPSS. However there soon comes a time when you want to repeat a sequence of commands, perhaps with a new or enlarged data file, and then you may prefer to call up a prepared sequence of syntax commands which can be run with a single mouse click. You will also be able to make minor modifications to commands in order to run a slightly different command.

There are other occasions when the use of syntax commands can be preferable to selecting commands from the menus, specially when you may need to apply the same operations to many variables.

SPSS assists you in preparing such syntax commands in a number of ways.

- SPSS keeps a log of all the commands you run, in the form of a file of syntax commands. This file is called statistics.jnl. On the Networked PC service this file will be located at J:\My_Documents\SPSSInc\IBMStatistics19\statistics.jnl. If you have installed SPSS on a standalone PC this file will probably be in the directory C:\windows\temp\.
- You can build up a file of syntax commands as you run SPSS from the on-screen menus, by pasting the commands you want into a Syntax Editor.
- The online help provides details of all the syntax commands which can be useful if you want to enter new syntax commands, or modify existing ones.

You will learn how to make use of all of these during this session.

1 Use Notepad (under Start | All Programs | Accessories) to look at what is in the file spss.jnl.

You should see a list of the syntax command equivalents of all the SPSS commands you have used! You should see commands such as:

```
DATA LIST     SAVE OUTFILE      MISSING VALUES
VARIABLE LABELS     VALUE LABELS    DESCRIPTIVES
FREQUENCIES     COMPUTE          RECODE
```

2 Close Notepad: select File | Exit.

For this session you will be running syntax commands from a sample file that has been prepared for you. This file, called survey1.sps, is similar to your own statistics.jnl file. You will also read some data from another sample text data file that has been prepared for you. This file, called survey2.dat is similar to survey1.dat — it has the same variables but many more cases. Rather than having commas separating each variable it is fixed width format i.e. variables are aligned in columns separated by spaces.
You are now going to get SPSS to run selected commands from the syntax file `survey1.sps`.

### 1.1 The Open Syntax command

1. Start up **SPSS** in the usual way.
2. Select **File | Open | Syntax**
3. Change the drive and directory to where the sample syntax file is (e.g. J:\stats\).
4. Click on the name of the SPSS syntax file (**survey1.sps**), and click on **Open**.

This will open up a new window — the Syntax Editor — in which you will see all the commands that were in the file **survey1.sps**. The name of the file appears in the title bar.

### 1.2 SPSS Syntax commands

Take a look at the commands in the Syntax Editor. The beginning of the file looks like this:

```plaintext
DATA LIST
   FILE='survey1.dat' FIXED RECORDS=1 TABLE /1
   person 1-3 sex 5(A) age 8-9 marstat 11(A) nkids 14
   income 17-21 height 23-25(1) sheight 27-29(1).
EXECUTE.
SAVE OUTFILE='survey1.sav'
   /COMPRESSED.
MISSING VALUES sex ("X").
MISSING VALUES age (99).
MISSING VALUES nkids (9).
MISSING VALUES income (99999).
MISSING VALUES height (99.9).
MISSING VALUES sheight (99.8, 99.9).
FREQUENCIES
   VARIABLES=marstat sex nkids.
DESCRIPTIVES
   VARIABLES=age height income nkids sheight
   /FORMAT=LABELS NOINDEX
   /STATISTICS=MEAN STDDEV MIN MAX
   /SORT=MEAN (A).
```

The commands obey the following rules:

- Each command starts with a command name at the beginning of a line (e.g. **DATA LIST, EXECUTE, SAVE** etc.).
There are at most 80 characters on a line.
For commands that have continuation lines each continuation line is indented. (It starts with at least one space).
Some commands have subcommands which are separated from one another by forward slashes '/'.
Each command is terminated by a period '.'— the command terminator.
Note also that commands and variable names can be typed in upper or lower case.
File names are enclosed in quotes.
Blank lines can be used to separate commands and improve readability.
You will now learn how to select and run commands from the Syntax Editor.

1.3 Running syntax commands
You can select individual commands to run, and you can run them in any order. You can also select a group of commands to run, or you can select all the commands in the window and run them. You will start with running single commands.

The first command in the file is the DATA LIST command. This is similar to the File | Read Text Data command that you did during an earlier section of this guide although in this case it reads data in fixed width column format (to read comma-separated data use the GET DATA command). It specifies the name of the file to be read, it lists the names of the variables and where in the file the scores for each variable are to be found. It also allows decimal points to be defined (for numeric variables), and which variables are string variables.

You are going to run this command from the Syntax Editor, but before you do this you need to change the reference to the data file from survey1.dat to survey2.dat so that the correct file is read in. You will also need to prefix the filename with its full path.

1 In the DATA LIST command change

    survey1.dat

into

    J:\stats\survey2.dat

Now you are ready to run the DATA LIST command.

2 Make sure the text insertion point is still somewhere within the DATA LIST command.

3 Click on the Run button: 

In a few moments you should see the variable names appearing in the Data Editor, but no data! The Output Viewer will appear with information on the variables to be imported (also check this viewer for error messages).

To get the data to appear you have to run the EXECUTE command.

4 Bring the Syntax Editor to the front.

5 Select the EXECUTE command (point somewhere within it and click).
Click on Run:  

The data from the file survey2.dat should now appear in the Data Editor.

The next command in the file is the SAVE OUTFILE command. This is the equivalent of the menu command File | Save that you have used before. It specifies the name of the file to be written to. You are going to run this command from the Syntax Editor, but before you do this you need to change the reference to the data file from survey1.sav to survey2.sav, and include the full path to the directory where you want to save your file.

1. In the SAVE OUTFILE command change

\texttt{survey1.sav}

into

\texttt{..path..\survey2.sav}

where \texttt{..path..} should be replaced with the full path to the directory where you want to save your file (e.g. type \texttt{J:\stats\survey2.sav} if you are using the Networked PC service).

2. Run the SAVE OUTFILE command.

You should see the new file name appearing in the title bar of the Data Editor.

1.4 Making global changes in the Syntax Editor

You have already seen how to make simple changes to the commands in the Syntax Editor. In this file there are several occurrences of \texttt{survey1} that need to be changed to \texttt{..path..\survey2} (where \texttt{..path..} should be replaced with the full path to the directory where you want to save your files). Instead of finding and changing each one separately, you can change them all at once.

1. Place the text insertion point at the beginning of the first command in the Syntax Editor.
2. Select Edit | Replace.
3. In the Replace dialog box, type

\texttt{survey1}

as the text to Find what, and

\texttt{..path..\survey2}

as the Replace with text where \texttt{..path..} should be replaced with the full path to the directory where you want to save your files (e.g. \texttt{J:\stats\survey2}).

4. Click on Replace All.
5. Click on Cancel as you have no further global changes to make.

1.5 Saving the contents of the Syntax Editor

Now that you have made a number of changes to the content of the Syntax Editor you should save it.
1 Make sure the Syntax Editor is the active window.
2 Select File | Save.

2 Further syntax
Syntax commands are particularly useful when running several commands on data, so rather than running one command, waiting for it to complete then setting off the next command they can be combined into one syntax file to run automatically one after the other.

1 Open J:\stats\survey1.sps
2 Open J:\stats\survey2.sav

2.1 Selecting and running several syntax commands at once
You can select several commands to run one after the other, rather than having to select each one individually. You need to run a few groups of commands now.
Run the MISSING VALUES commands now. To select these commands:
1 Drag the mouse over all the required commands (whilst pressing down the left mouse button) until you have highlighted at least part of each command.
2 Click on Run.
No new output will be generated, but missing values will have been added.
Now run the VARIABLE LABELS and the VALUE LABELS commands for the marstat variable.
1 Select the appropriate commands.
2 Click on Run.
Again this should produce no new output.

2.2 Mixing menu commands with syntax commands
It is not necessary to continue to use syntax commands once you have started to use them. You may go back to using menu commands whenever you wish.
For instance you could now use a menu command to save the working data file. This will save the changes you have just made to the variable and value labels, and the missing values declarations.
1 Make the Data Editor the active window.
2 Use the File | Save command to save the working data file.
Now go back to running further syntax commands.
3 Make the Syntax Editor active again.
4 Run the commands to create and label the new variables status, and incomcat. (You will have to scroll down the Syntax Editor to find
these commands.) Take care that you run the command to create each variable before you create its labels.

5. Run the commands to create the new variable `tall`. Note that this requires a sequence of commands beginning with `DO IF` and ending with `END IF`. It is important that you select and run the full sequence of commands. Do not attempt to run a `DO IF` without running the rest of the sequence up to and including the `END IF`.

6. Look at the new scores in the Data Editor and verify that they are correct.

Note that the above data transformation commands only complete if you run an `EXECUTE` command (or select `Transform | Run Pending Transforms`). Incomplete operations are indicated by the words `Transformations pending` appearing in the status bar.

Save the data file using the method of your choice:

7. *Either*
   - select one of the `SAVE OUTFILE` commands in the Syntax Editor and click on `Run`,
   - or
   - make the Data Editor active and select `File | Save`.

### 2.3 Pasting menu commands into the Syntax Editor

It is possible to select commands from the on-screen menus in the usual way, and then Paste them into the Syntax Editor. In this way you can build up new commands in the Syntax Editor without knowing the correct syntax to use, or even the correct command name.

You will use this to create and run a command for generating a two way frequency table to look at the relationship between income and height using the two categorical variables `incomcat` and `tall`.

1. Select `Analyze | Descriptive Statistics | Crosstabs`.
2. In the `Crosstabs` dialog box, select `tall` as the row variable, and `incomcat` as the column variable.
Click on Paste. Instead of running the command, SPSS puts the corresponding syntax command into the Syntax Editor — at the end.

In the Syntax Editor, select the new CROSSTABS command and run it.

The output consists of a table in which the cells show the counts of each combination of the values of the variables incomcat and tall.

2.4 **Getting help with syntax commands**

Sometimes you know the name of a command but are not sure of the exact syntax, or of what options are available. Or you may want to modify an existing command and need some help with it. There is a quick way to access the online Help system to find information about the syntax of a command in the Syntax Editor.

1. In the Syntax Editor, select the CROSSTABS command.

2. Click on the Syntax Help button in the tool bar.

SPSS will display details of the CROSSTABS command. Anything in square brackets [ ] is optional (almost everything!), anything in upper case you type as it is (e.g., the words CROSSTABS, and BY), anything in lower case you have to replace with your own requirements (e.g., varlist). You choose between items enclosed within braces { }, and ... means more of
the same (e.g. varlist ... is used to indicate a list of variable names separated by commas or spaces).

2.5 Entering commands into a Syntax Editor
The variable tall has no value labels assigned to it. You are now going to enter and run a new syntax command to provide appropriate value labels. You can find out what to do by looking at other VALUE LABELS commands, or you could use the syntax help.

1. In the Syntax Editor look at the commands for creating value labels.
2. Type

   VALUE LABELS

   as a new command at the end of the Syntax Editor.
3. With the text insertion point still within this command click on the Syntax Help button:

   SPSS will display details of the Value labels command.
4. Using the help information complete the syntax command to provide appropriate value labels for tall.
   (Don’t forget the command terminator(,))
5. Run your new VALUE LABELS command.
6. Re-run the CROSSTABS command and see how the use of value labels improves the information content of your output.
7. When you are done with the Help, close the Help window.
8. Save the working data file.
9. Save the revised Syntax Editor.
10. Save and/or print any of the output you wish.

2.6 Creating an SPSS program
The commands in a Syntax Editor can be used as the basis for an SPSS program. Such a program would consist of a set of commands that you would want to run sequentially. In future you would be able to run the whole program by selecting all the commands in the Syntax Editor and clicking on Run.

The commands in your Syntax Editor could now be modified to form a program that could be used in the initial stages of a project based on the survey data. You would need to keep those commands that read the data from the ASCII data file and provide all the data definition information (missing values, variable labels and value labels). You would also need to keep the commands to create the new variables that you would need for further analysis. You would also need the commands to perform the exploratory data analysis so that you could validate your data before proceeding with any further analysis. And finally you would need the command to save the data as an SPSS data file.
In addition you can add explanatory comments to an SPSS program. Comments can contain any text. They begin with the word `comment` (or an asterisk `*`), and should end with the usual command terminator. Examples of comment lines are:

```
COMMENT This is an example of an SPSS program.
Comment The file survey2.dat contains some hypothetical data.
*Now save the working data file as an SPSS data file.
```

1. Modify the commands in your Syntax Editor to create a program that could be used in this way. Include some appropriate comments. Take care that you observe the rules for syntax commands (see the start of this section). Remember that you can copy or move blocks of text if you wish using the `Edit | Cut`, `Copy` and `Paste` commands.

2. Save the contents of the Syntax Editor in a new file. Choose a suitable name (e.g. `newsurvey2`).

Before you run your program, clear out the Output Viewer:

3. Make the Output Viewer active, then select `Edit | Select All`.
4. Select `Edit | Delete`.

Now run your complete program:

5. Select `Edit | Select All`.
6. Click on `Run`.

Note that if you have any unsaved changes in the Data Editor, SPSS will ask you if you want to save them. This is because the `DATA LIST` command creates a new active data file, and SPSS gives you the chance to save the previous one before it is discarded.

7. Look at the results in the Output Viewer and check for errors.
8. Use the `Edit | Find` command to look for words such as `error` or `warning` (double-click on the output before selecting `Edit | Find`). This is particularly useful when you have programs that produce a lot of output.

When you are satisfied that your program is doing what you want:

9. Save it again.

You may also wish to print it. The procedure for printing the Syntax Editor is the same as for the Output Viewer.

Before moving on to the next section, edit, save and print what you want from the Output Viewer and Syntax Editor then exit from these windows.

### 3 More on data management

SPSS is not only a statistical analysis package, it is also a data management system. This section describes some of the data management facilities provided by SPSS:
<table>
<thead>
<tr>
<th></th>
<th>data selection</th>
<th>Allowing you to perform commands on only selected cases in the working file</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>sorting and splitting files</td>
<td>Allowing you to apply the same commands to each of several groups of cases in the file</td>
</tr>
<tr>
<td>3</td>
<td>saving subsets of files</td>
<td>allowing you to save only certain cases or certain variables</td>
</tr>
<tr>
<td>4</td>
<td>merging files</td>
<td>allowing you to combine data from two or more files</td>
</tr>
</tbody>
</table>

Note that in some of the following examples you may select the required commands by clicking on the menus, making the appropriate selections in the dialog boxes and clicking on OK. However there are some occasions when you will need to use syntax commands as there is no equivalent menu command.

### 3.1 Data selection

Sometimes you want to perform some analyses, or create some charts, based on the data for a selection of the cases in your data file — perhaps just the females, or just those that are married. You first make a selection to filter out those cases not required. Then you run the required analyses. The filter remains in effect until you turn it off. When you have finished working on the subset of your data you can remove the filter.

Use this method to obtain descriptive statistics on income and height for just the females in the file survey2.sav.

1. Select **Data | Select Cases**.
   
   This will open up the **Select Cases** dialog box.

   In the **Select Cases** dialog box you should find that at present **All cases** are selected. Notice the different possibilities for selecting cases — you can select a random sample, you can use an existing variable as a filter variable (more about filter variables later), or, as in this case, you can provide the selection criteria.

2. Click on **If condition is satisfied**, then click on the **If** button.

3. In the **Select Cases: If** dialog box, enter the condition
   
   `sex='F'`

4. Click on **Continue**.

5. **Either**
   
   - click on **OK** to run the command immediately,
   - or
   - click on **Paste** to put the syntax commands into the Syntax Editor, then select the commands in the Syntax Editor and **Run** them.
Notice the words Filter On in the status bar at the bottom of the Data Editor. This reminds you that the next commands to be run will use only the selected cases.

1. Now run the Descriptives command on the variables height and income.
2. Look at the output.
   (It's surprising how many people don't look at their output!) You should find that the results are based on the data for the females only. If you look at the data in the Data Editor you will see that you have a new variable called filter_. This is the filter variable. It has the value Selected for each case selected, and the value Not Selected for those cases not selected. Notice also that each case not selected has a line drawn through the case number.
3. Now turn the filtering off.
   You should be able to find out how to do this yourself. Use the online help if necessary! NB. The filter variable filter_ will still be present when filtering is turned off — you can just delete this variable.

If you wanted to repeat the analysis on the males only instead of the females, you could repeat the above sequence of commands with sex='M' instead of sex='F' as the condition.

It is not always necessary to create a special variable when you want to filter cases. You may already have a suitable variable. You may specify any existing numeric variable as a filter variable. Cases that have 0 as the value of the filter variable (or missing) will be excluded from subsequent calculations.

### 3.2 Splitting files

Now consider the situation where you have more than two groups of cases that you want to analyse — say for instance you wanted to calculate some statistics separately on each of the four marital status groups (married, single, widowed, divorced). Clearly the above method of selecting each group in turn, then performing the required analyses could become rather tedious, even with the use of syntax commands. SPSS provides a mechanism for splitting the file according to the values in a particular variable (or variables), and then applying the analysis commands separately to each part of the split file. The split remains in effect until you turn it off.

Use this method to obtain separate histograms of the numbers of children for each of the four marital status groups in the data file.

1. Make sure that filtering is off, so that you will use all cases in the data file.
2. Select Data | Split File.
3. In the Split File dialog box, select Organize output by groups.
4. Select marstat as the Groups Based on: variable.
5 Make sure that **Sort the file by grouping variables** is selected.

6 *Either*
   - click on **OK** to run the command immediately,
   - or
   - click on **Paste** to put the syntax commands (in this case there will be two) into the Syntax Editor, then select the commands in the Syntax Editor and **Run** them.

Notice the words **Split by Marstat** in the status bar at the bottom right of the Data Editor. This reminds you that the next commands will be performed separately on each subgroup of Marstat in the file. If you look at the data in the Data Editor you will see that the cases in the file have been sorted by the value of the **marstat** variable — all the **Ds** come first, then the **Ms**, then the **Ss** and finally the **Ws**.

When SPSS processes the subsequent commands, it halts whenever the **marstat** variable changes and produces a set of results, then carries on with the next group of cases. It is therefore vital that the cases are sorted by the splitting variable. If you look at the command syntax you will see that the appropriate commands have been prepared for you.

7 Now use the **Graph** command to create a **Histogram** of the variable **nkids**.

Notice that although only one Graph command was run, four charts are produced, one based on each section of the file as defined by the **SPLIT FILE** command.

8 Turn split file processing off.

Note that if the data were already sorted by the variable specified for the grouping, then you could have selected **File is already sorted** instead of **Sort the file by grouping variable** in the **Split File** dialog box. This can save processing time when you have a large datafile. The equivalent in the syntax would have been to omit the **Sort** command.

### 3.3 Sorting the working data file

There may be times when you wish to sort the data in the file without necessarily combining it with a split file command. For instance it may be easier to check some data values if the data are arranged in a particular way.

Sort the data now by the **person** variable.

1 *Either*
   - Select **Data | Sort Cases**.

Look at the data in the Data Editor and check that the data are indeed now sorted by the **person** scores.

Note that you can choose the sort order — ascending or descending.
3.4 Saving selected variables only

After working with a file for sometime you may wish to save the working data file, but not necessarily save all the variables. Instead of deleting those you don’t want to save, you could just save those variables you wish to keep. The unsaved variables will still remain in the working data file, but will not be copied to the SPSS data file.

Use this method now to make a new SPSS data file with just the data for the variables person, sex, age, marstat and nkids.

Make sure that the data are sorted by person before you do this — see previous section.

You will need a Syntax Editor window to save selected variables.

If you haven’t been using syntax commands in this session you won’t have the Syntax Editor open. You can open one by selecting Paste from any dialog box, or you can select File | New | Syntax.

1 In the Syntax Editor, type the following command:

```
SAVE OUTFILE='J:\stats\survey3a.sav'
/KEEP=person, sex, age, marstat, nkids.
```

Note that when you specify a list of variables that are next to one another in the working data file, you can abbreviate the list using the word to, e.g.

```
SAVE OUTFILE='J:\stats\survey3a.sav'
/KEEP=person to nkids.
```

Don’t forget the command terminator.

2 Run the completed SAVE command.

Note that the working data file is unaffected by this save command. The title in the Data Editor remains unchanged, and all the original variables are still present.

3.4.1 Open a file with selected variables only

Sometimes you may have a data file with a large number of variables in it, but you want to perform some analyses based on just a few of those variables. Instead of getting SPSS to read all the data, you could read in just those variables you want to work with. It will be easier to work with the file without the dialog boxes listing variables you will not be using.

The option of specifying which variables you want to read is only available using a syntax command. You cannot do this using the menu commands.

Use this now to open the SPSS data file survey3.sav with just the data for the variables person, income, height and sheerht.

1 In the Syntax Editor, type the following command:

```
GET
```

2 Click on the Syntax help button for information about the subcommand you will need.
3 On the basis of the online help, complete the command specifying `survey3.sav` as the name for the SPSS data file that you will read, and `person`, `income`, `height` and `sheight` for the variables you want to keep.

Don’t forget the command terminator.

4 Run the completed GET command.

Save this reduced data file as a new SPSS data file using the filename `survey3b.sav`.

Either

1 Use the **File** | **Save As** (not **File** | **Save Data**) command

or

1 Use the syntax command

   `SAVE OUTFILE='J:\stats\survey3b.sav'.`

You should now have two new SPSS data files, each containing a subset of the variables from the original file `survey3.sav`. You will discover shortly how to re-combine these files using a procedure that merges files holding data for the same cases but different variables.

2 Open each of these two files (`survey3a.sav` and `survey3b.sav`) in turn to confirm that each contains all the cases, but a different selection of variables.

As you can only have one data file open at once, SPSS will close the file that’s open before opening another one.

3.4.2 **Saving selected cases only**

Sometimes you may wish to create separate SPSS data files for different groups of cases — say the males and the females.

For each subset of cases you require as a separate file, you will need to read the complete file, delete those cases not required, then save what is left as a new file with a different name. The easiest way to do this is to use syntax commands.

For example, create two new SPSS data files based on the data in `survey3.sav`, one for the males in the file and the other for the females.

Call them `survey3m.sav` and `survey3f.sav`.

1 Open the file `survey3.sav`.

2 In the Syntax Editor, type and run the commands:

   `SELECT IF (SEX='M').
   EXECUTE.`

This will delete from the working data file those cases that do not satisfy the selection criterion specified.

3 Use **File** | **Save As** (not **File** | **Save**) to save this reduced data file as a new SPSS data file using the filename `survey3m.sav`. 
4 Repeat these last three steps with suitable modifications (starting from the command to open survey3.sav) to create a file called survey3f.sav for the data for the females only, and another called survey3x.sav for those with the missing value X for sex.

5 Open each of the three files survey3m.sav, survey3f.sav and survey3x.sav, in turn, to confirm that each contains all the variables, but a different selection of cases.

3.5 Combining files

SPSS provides a mechanism for combining data from two or more files. You can merge several files containing data for the same cases but different variables, and you can merge several files containing data for the same variables but different cases. The data to be merged must be in SPSS data files.

3.5.1 Same cases — different variables

There may be occasions when you have the data for the same cases in several files, and may wish to bring them together in order to look for relationships between the variables in the different files. For instance you may be conducting a longitudinal study in which you follow up the same cases over a period of time and collect data from them on a number of occasions, or you may be collecting data on your cases from a number of different sources.

You would need to be able to match the cases in one file with those in the other file. This is usually done by having a matching case identifier in each file, and having the data in each file sorted by the identifying variable.

If you have been following all the exercises in these notes then you should have two files survey3a.sav and survey3b.sav that satisfy these criteria. Copies of these files are also available with the other sample files in T:\its\spss.

If you use syntax commands, you can combine variables from several files at once. If you use menu commands you can only add data from two files at a time, so you must first open one of the files, then add to it variables from one other file, and repeat this for each additional file. If you want to merge data from several files, then it is better to use syntax commands.

Whichever method you use, you can specify, from each file, which variables to keep (or which variables to drop). You can also specify which variable in each file to use as the key variable for matching the cases from the several files. Whether you use menus or syntax, the new composite file becomes the new working data file.

You are now going to merge the data from the two files survey3a.sav and survey3b.sav. Try this first using menu commands and dialog boxes, but Paste the commands into a Syntax Editor and Run them from there. This way you will see the equivalent syntax commands.

1 Open the file survey3a.sav.
2. Select **Data | Merge Files | Add Variables**. This will open up the **Add Variables: Read File** dialog box.

3. Browse to the file **survey3b.sav** and click on **Continue**. This will open up the **Add Variables** dialog box.

4. Choose which variables you want to include in the new composite file. Notice that you can exclude variables from either the working data file, or the new file being added.

5. Click on **Match cases on key variable in sorted files**, and specify person as the **Key Variable**.

6. Click on **Paste**. SPSS will remind you that the data in each file need to be sorted by the key variable.

7. Click on **OK** to close the message box.

8. In the Syntax Editor, select the **MATCH FILES** command and the **EXECUTE** command that follows it, and **Run** them.
Look at the Data Editor and check that you have all the variables and all the cases that you expect.

If you have chosen to keep all the variables from both files, your syntax commands will look like:

```
MATCH FILES /FILE=*  
  /FILE='J:\stats\survey3b.sav' 
BY person.
```

It is not difficult to see what is happening here. Each `/FILE` subcommand specifies a file to be involved in the merge, `/FILE=*` is used to indicate the present working data file.

An alternative way of merging the two files `survey3a.sav` and `survey3b.sav` which does not require you to read in one of the files first, is to use a modified form of the MATCH FILES syntax command.

1. Modify the syntax commands as follows, and run them:

```
MATCH FILES /FILE='J:\stats\survey3a.sav'  
  /FILE='J:\stats\survey3b.sav' 
BY person.
EXECUTE.
```

Note that using syntax you can add variables from several files at once by including several `/FILE` subcommands. Up to 50 filenames can be specified on one MATCH FILES command!

It is not necessary to include all the variables from each file when you are combining files with the MATCH FILES command.

1. Create a new file containing just the variables `person`, `sex`, `age` and `income` by combining just the chosen variables from the files `survey3a.sav` and `survey3b.sav`.

You may need to look at the online help to get further details of the MATCH FILES command (or experiment with the Add Variables dialog box), in order to find out about how to keep selected variables only.

### 3.5.2 Same variables — different cases

There may be occasions when you have the data for some cases in one file, and data for further cases, but the same variables, in another file, and you may wish to bring them together in order to analyse the larger data set. For instance you may have started experimenting with the analysis of your data when only a small sample of the full data set was available. Or you may have collected data for different cases on different occasions and have them entered initially into different files.

You would normally have the same variables in each file, but different cases. If you have been following all the exercises in these notes then you should have three files `survey3f.sav`, `survey3m.sav` and `survey3x.sav` that satisfy these criteria. Copies of these files are also available with the other sample files in `T:\its\spss`.

If you use syntax commands, you can combine cases from several files at once. If you use menu commands you can only add data from two files at a
time, so you must first open one of the files, then add to it the cases from one other file, and repeat this for each additional file. If you want to merge data from several files, then it is better to use syntax commands.

Whichever method you use, the new composite file becomes the new working data file.

Merge the data from the two files `survey3f.sav` and `survey3m.sav`. Try this first using menu commands and dialog boxes, but Paste the commands into a Syntax Editor and Run them from there. This way you will see the equivalent syntax commands.

1. Open the file `survey3f.sav`.
2. Select *Data* | *Merge Files* | *Add Cases*. This will open up the *Add Cases: Read File* dialog box.
3. Browse to the file `survey3m.sav` and click on *Continue*. This will open up the *Add Cases* dialog box.
4. Choose which variables you want to include in the new composite file. Notice that you can exclude variables if you wish.
5. Click on *Paste*.
6. In the Syntax Editor, run the ADD FILES command and the EXECUTE command that follows it.
7. Look at the Data Editor and check that you have all the variables and all the cases that you expect.

If you have chosen to keep all the variables from both files, your syntax commands will look like:

```
ADD FILES /FILE=*
   /FILE='J:\stats\survey3m.sav'.
```

Note that using syntax you can add cases from several files at once by including several /FILE subcommands — up to 50 filenames can be specified on one ADD FILES command!

8. Write and Run a single ADD FILES command to merge the data from the three files `survey3f.sav`, `survey3m.sav` and `survey3x.sav`.
9. Look at the Data Editor and check that you have all the variables and all the cases that you expect.

### 4 Being in control — SPSS options

There are several ways in which you can customise SPSS to suit your way of working. For instance you can specify whether or not you want to record all your commands in a journal file, whether or not you want the commands you use to be included in the output, whether you want any areas in your charts to be filled with colours or patterns, and so on.
Any changes to your options will also effect subsequent sessions. On the Networked PC Service they will revert to the standard defaults next time you login.

1. Select **Edit | Options**.

This will open up the **Options** dialog box on the **General** tab:

![Options dialog box](image)

### 4.1 Command recording

Unless you have already made changes to the default settings, you will see that command recording is **on**, and that new commands are **appended** to the journal file. If you click on the **Browse** button in the **Session Journal** box, you will be able to change the name and location of the file to use as the journal file. You can also change the other settings. The following table indicates what effect changing these settings will have:

<table>
<thead>
<tr>
<th>Journal Status</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>off</td>
<td>No commands are recorded.</td>
</tr>
</tbody>
</table>
| on overwrite   | Commands are always recorded.  
                 A new journal file is created for each SPSS session. 
                 Any previous journal file is deleted. |
| on append      | Commands are recorded only if journal file exists. 
                 Commands added to the end of the journal file. 
                 If journal file doesn't exist, commands are not recorded. |
Click on the appropriate buttons in the **Session Journal** box to make whatever changes you wish.

### 4.2 Chart options
Click on the **Charts** tab to display the chart options. This will allow you to alter some of the defaults that are used when you create charts.

Make any changes you wish, then click on **OK**.

### 4.3 Viewer options
1. Click on the **Viewer** tab to display the viewer options. This will allow you some control over the layout and content of the output produced by SPSS.
2. Make any changes you wish, then click on **OK**.

### 4.4 Saving Options settings
1. When you are finished making changes, click on **OK** in the **Options** dialog box.

The changes you have made will be effective throughout the remainder of the current SPSS session, and for subsequent sessions. On the Networked PC service they will revert to the standard defaults next time you login.

### 5 Portability
SPSS is available for many different types of computer. At Durham University, both **UNIX** and PC (Microsoft Windows) versions are available.

This section summarises what you have to do if you want to transfer your work from one system to another.

#### 5.1 Command files
Although there are some differences in the way SPSS commands are run on different systems, the syntax of the commands is essentially the same for all systems. This means that an SPSS program written for one system should transport with little difficulty to another system. Some details may need to be changed. For instance the rules for the naming of files vary between systems, and so any file names referenced in a program may need to be changed. Also there are some commands which are available on some systems but not others.

#### 5.2 SPSS data files
Most SPSS data files are portable. This means that an SPSS data file written on one system can be transferred to another system. For instance, SPSS data files created by SPSS for Windows can be copied to UNIX and read using the Unix version of SPSS, and vice versa. Note that SPSS data files are not text (ASCII) files and must be transferred from one system to the other in binary mode (**not** text mode).

It is also possible to transfer SPSS files from one system to another using **SPSS portable** files. SPSS portable files are text files, and should be transferred from one system to another in the same way as you would transfer any other text file. All versions of SPSS have the ability to save the
working data file as an SPSS portable file, and all versions of SPSS have
the ability to **import** an SPSS portable file.

5.3 **Chart files**
Chart files are system specific. This means that charts created on one
system cannot be read by SPSS running on a different system.

6 **Further information**

6.1 **CIS documents**
The following documents are available from the IT Service Desk.

*Guide 167: Copying material from SPSS to Microsoft Word 2002 (Office
XP)*

This guide explains how you can copy data and output (including charts)
from SPSS for Windows into a Word for Windows document.

*Guide 86: SPSS for Windows - Importing and Exporting Data*

This guide explains in detail how to copy data from Excel, Word and
Access into SPSS, and vice versa.

*Guide 27: SPSS for Unix*

These notes provide an introduction to the Unix version of SPSS. It is very
similar in content to this document.

6.2 **SPSS manuals**
The following SPSS manuals give full details of the Windows version of
SPSS.

*SPSS 10 for Windows Base System User’s Guide*

This manual provides detailed information on how to run SPSS 9.0 for
Windows using the menus and dialog boxes. It also explains many of the
statistical concepts involved, how to use the procedures correctly, and how
to interpret the results.

*SPSS 10 Base System Syntax Reference Guide*

This manual is a reference to the command syntax for the Base system.
Although most of the facilities of the Base system are available using the
menus, there are some that can only be obtained by entering and running
syntax commands.

*SPSS 10 Base Applications Guide*

This manual covers the statistical procedures in SPSS Base 9.0 and
contains examples for each procedure, with detailed descriptions of the
output.

*SPSS 10 Interactive Graphics*

This manual documents in detail interactive charts.

*SPSS 10 Regression Models*

This manual provides a guide to the statistical techniques in SPSS
Regression Models and how to perform statistical analyses with the dialog
box interface and syntax. Statistical procedures in this module include: multinomial logistic regression, binary logistic regression, nonlinear regression, two-stage least squares, probit, and weighted least squares.

**SPSS 10 Advanced Models**

This manual documents the procedures available in SPSS Advanced Models. It shows how to use dialog boxes and syntax. Statistical procedures include: General linear model, loglinear, hiloglinear, GENLOG, survival analysis Kaplan-Meier, Variance component estimation and Cox regression.

**SPSS 10.0 Guide to Data Analysis**

This includes sections on preparing data for analysis, describing data, hypothesis testing and examining relationships in data.

SPSS manuals are available in the University Library, and, for reference only, in the IT Service Desk, Computer Centre.

6.3 **SPSS newsgroups and internet links**

The CIS maintain web-pages containing information of use to SPSS users at Durham. They include a list of relevant discussion lists and newsgroups, and an up-to-date list of SPSS software available at Durham. They are at: [http://www.dur.ac.uk/cis/spss/](http://www.dur.ac.uk/cis/spss/)

6.4 **Help and advice**

For questions and advice about all computing matters, IT Service Desk is available:

- by personal contact in the IT Service Desk in the Computer Centre
- by telephone (41515) Monday to Friday.
- by email to itservicedesk@durham.ac.uk. Messages are normally answered during the same day.
Appendix A: Sample files

<table>
<thead>
<tr>
<th>File type</th>
<th>File names</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCII data files</td>
<td>survey1.dat, survey2.dat</td>
</tr>
<tr>
<td>SPSS syntax file</td>
<td>survey1.sps</td>
</tr>
<tr>
<td>SPSS data files</td>
<td>Survey3.sav</td>
</tr>
</tbody>
</table>

Appendix B: Recommended naming conventions for SPSS file types

<table>
<thead>
<tr>
<th>extension</th>
<th>used for</th>
<th>text file?</th>
</tr>
</thead>
<tbody>
<tr>
<td>.dat</td>
<td>ASCII data file</td>
<td>yes</td>
</tr>
<tr>
<td>.sav</td>
<td>SPSS data file</td>
<td>no</td>
</tr>
<tr>
<td>.sps</td>
<td>SPSS command file</td>
<td>yes</td>
</tr>
<tr>
<td>.spv</td>
<td>SPSS output listing</td>
<td>No</td>
</tr>
<tr>
<td>.sct</td>
<td>SPSS chart file</td>
<td>No</td>
</tr>
</tbody>
</table>

Note that you can view and edit the text files using Word, Notepad or any other text editor. You should not attempt to view or edit the data in an SPSS data file except by opening the file from within an SPSS session. SPSS data files (.sav), SPSS output files (.spo) and SPSS chart template files (.sct) are in a special format that only SPSS can read.

Appendix C: Operators

<table>
<thead>
<tr>
<th>Arithmetic operators</th>
<th>Relational operators</th>
<th>Logical operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>&lt;</td>
<td>&amp;</td>
</tr>
<tr>
<td>-</td>
<td>&gt;</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>&lt;=</td>
<td></td>
</tr>
<tr>
<td>/</td>
<td>&gt;=</td>
<td></td>
</tr>
<tr>
<td>**</td>
<td>=</td>
<td>~</td>
</tr>
<tr>
<td>()</td>
<td>~=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>order of operations</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D: Answers to selected exercises

The following solutions are suggested for some of the exercises. Note that there will often be several ways of achieving the desired result, all of which may be equally valid. It does not matter if your solution is different from the one offered here so long as it works!

Section 2.1 (turning filtering off)
1. Select Data | Select Cases, then
2. in the Select Cases dialog box, choose All cases, then
3. click on OK.

or
1. Type the command FILTER OFF. in the Syntax Editor, and run it.

Section 2.2 (cancelling split file processing)
1. Select Data | Split File, then
2. in the Split File dialog box, choose Analyze all cases, then
3. click on OK.

or
1. Type the command SPLIT FILE OFF. in the Syntax Editor, and run it.

Section 3.4.1 (opening a file with selected variables only)
GET FILE='survey3.sav'
/KEEP=person, income, height, sheight.
EXECUTE.

Section 3.5.1 (selecting variables when combining files with match files)
MATCH FILES /FILE='survey2a.sav'
/FILE='survey2b.sav'
/FILE='survey2f.sav'
/KEEP=person, sex, age, income
/BY person.
EXECUTE.

Section 3.5.2 (combining cases from three files)
ADD FILES /FILE='survey2f.sav'
/FILE='survey2m.sav'
/FILE='survey2x.sav'.
EXECUTE.
Index

C
Combining files 15
Crosstabs 6

D
Data List 3

E
Execute 3

G
Get 13

H
Help
Syntax commands 7

O
Open
File with selected variables 13
Syntax command 2
Options
Chart 20
Command recording 19
Saving 20
Viewer 20

P
Paste Menu commands 6

R
Run button 3

S
Save
Outfile 2
Selected cases 14
Selected variables 13
Syntax window 4
Select cases
Data selection 10
Sort cases 12
Splitting files 11
Spss.jnl 1
Syntax
Help with commands 7
Introduction 1
Rules 2

T
Transforms
Run pending 6