Introduction to IBM SPSS Statistics 22

Section 1 – COMMON TASKS

Creating & Working with SPSS Data Files

Opening SPSS (in CSCE Computer Clusters)

From the **Start** button, select:

or   **All Programs** > **IBM SPSS Statistics** > **IBM SPSS Statistics 22**

Opening a Data File

To create a new data file at the opening SPSS screen menu:

> Select **New Files: New Dataset** then click on **OK**.

To create a new data file when you have been working on another data file:

> From the **File** menu, choose **New** then **Data**. (The current data file will be closed.)

To open an existing file:

> From **File** menu, choose **Open** then **Data**.
>
> Selecting **Open another file…** at the opening SPSS screen menu has the same effect.
>
> If you can see the name of the file, click on the name and then click the **Open** button (or alternatively, double-click on the file name).
>
> If the file is on another drive or in another folder, first use the arrow to the right of the **Look in:** box to move to the relevant drive/folder.

Key Concepts

SPSS works with at least two windows:

> **Data Editor**   – where you can view and change the variables and the data.
> (you can have only one data file open at a time);
>
> **Output Viewer** – where the results of any procedures are displayed.
>
> Although it is possible to undertake most tasks in SPSS simply by using the various menus, it is also possible to undertake these tasks by typing a series of commands. If this method is used, then there will be another window displaying the lines of syntax – the **Syntax Editor**.

In the Data View in Data Editor, which looks similar to a spreadsheet:

> each **column** = a **variable** (or **field**);
>
> each **row** = a **case** (or **record** or **observation**);
>
> a **cell** is where a row and a column intersect;
>
> the content of a cell indicates the **attribute** (or **value**) recorded
> for that particular variable for that particular case.

The first stage in the creation of a new data file using information collected from,
for example, a questionnaire survey, is to *define the variables*.

The second stage is to *enter the data*.

Creating a New Variable

A new variable can be created by by defining and typing in the variable in the Data Editor or by deriving it from existing variables (see "Compute a Variable", p.8).

In a new Data file, the new variable will be created in the first column. In a file where there are already other variables, the new variable will be added to the right of the existing variables. Alternatively, it can be inserted to the left of an existing variable by clicking on the existing variable's name and selecting **Edit > Insert Variable** from the menu, or by *right*-clicking on the variable's name and choosing **Insert Variable**.

The variable will initially be assigned a name of the form VAR00001. It will be defined as numeric of width 8 with 2 places after the decimal point (e.g. 12345.78).

To change the characteristics of the new variable, you have to define the variable.

Defining a Variable

Variables (including any that you insert into the data file in **Data View**) are defined in the **Variable View** within the **Data Editor**.

Within **Variable View**, rows correspond to the variables in the data file, and columns correspond to their defining characteristics. Selecting a cell allows the corresponding characteristic to be specfied or changed, either by over-typing or via a scrollable list.

For some attributes, when you click on a cell, a small grey box **...** will appear. Clicking on the grey box leads to a pop-up window where further changes can be made.

The column headings are:

Name   *variable name*

This should begin with a letter and be no more than 64 characters (letters and/or numbers – no spaces). Early versions of SPSS only allowed 8 characters. The 8-character limit may still apply when importing from or exporting to files used by some other software. Variable names must be unique.

Type   *variable type* (e.g. numeric, string)

If the data for the variable consists of text, select **String**. which can be upto

Width   *number of digits or characters*

This should be great enough to display the largest number or string.

Short string variables (up to 8 characters in length) can be used in some statistical procedures such as crosstabs, but very few statistical procedures in SPSS can use long string variables of more than 8 characters. String variables can contain up to 32,767 characters.

While numbers larger than that specfied in "width" can still be entered and will be recorded correctly, it is not possible to input a string longer than that specified in "width".

Decimals   *number of decimal places* (for numeric variables)

Label   *variable label*

As the name of the variable is restricted in terms of length and characters it can contain, it is often useful to have a longer, more meaningful label that will appear in the tables and graphs that you create.

The label can be up to 256 characters, but some statistical procedures will only display the first part of a very long label, so it is desirable to keep the first 20 or so characters unique.

Values *value labels*

While it is possible to record information as text (as a string variable), it is often easier and more efficient to enter information as a numeric code. In addition, in an analysis of string variables (e.g. frequency table), the values will be recorded in strict alphabetical order, which may not be particularly intuitive, particularly if the data is nominal rather than ordinal in nature. However, we would still want the information to be displayed with meaningful labels in any tables or graphs. To do this we must define the value labels.

Value labels can be up to 120 characters, but some statistical procedures will only display the first part of any very long label – so try to keep the first 16 or so characters unique.

Click on the appropriate cell, then click on the grey **...** box

In the **Value:** box, enter the relevant value code, and

in the **Label:** box enter the corresponding label.

Click the **Add** button.

Repeat the process for other values.

Click **OK** button when finished.

Tip: In the Data Editor window, you can switch back and forward between viewing the raw data and the assigned value labels, by selecting/deselecting **Value Labels** under the **View** menu (or clicking on the value labels icon in the toolbar).

Missing *missing values*

In many data collection exercises, some information may be unavailable. This can be for a variety of reasons (e.g. refusal, not collected, unknown). In the Data Editor, leaving a cell blank will be interpreted as meaning that the information for that variable is missing for that particular case. These cases are termed 'system missing'. Other values that we want to define as missing are termed 'user-defined missing'.

For string variables, a blank cell is treated in the same way as any other valid value and is not regarded as system missing. However, a blank space can be defined as user missing, provided that it is a 'short-string' variable i.e. one which has been defined as 8 characters or less.

Click on the appropriate cell, then click on the grey **...** box

Click on **Missing Values…** button to define or change the values to be regarded as missing in statistical procedures.

Click **OK** button when finished.

Columns *width of column as displayed in* Data View

Align *alignment within column as displayed in* Data View

Columns and Align only affect the way variables are displayed in Data View. It does not affect how they are displayed in the output from any data analysis.

Measure

Nominal - data values have no intrinsic ranking.
Ordinal - data values have some intrinsic ranking but intervals are not constant.
Scale - data is on a continuous, measurable metric; can only ever be numeric.

String (alphanumeric) variables will usually be Nominal.

Role

Can be used to pre-select variables for analysis. Default is '**Input**'.

Input   - a predictor/independent variable
Target   - a dependent variable
Both   - may be used as independent or dependent
None   - no role assigned
Partition - variable used to partition the data into separate samples

(**Split** is not used in SPSS Statistics: it is only used in SPSS Modeler).

Deleting a Variable

If in **Data View**

click on the variable name at the top of the column,
    press the **Delete** key
or right-click and select **Cut** or **Clear**
or from menu, select **Edit** > **Clear**.

If in **Variable View**

click on the row number to the left of the variable's name,
    press the **Delete** key
or right-click and select **Clear**
or from menu, select **Edit** > **Clear**.

Tip: If you accidentally delete the wrong variable,
    press **Ctrl**+**Z** keys at the same time
or from menu, select **Edit** > **Undo**

Entering Data Using the Data Editor

In Data View, the currently selected cell is shaded.

Use the arrow keys to move to the desired cell.

Alternatively, use the scroll bars to the left and foot of the Data View window or the **P**age **Up** and **P**age **Down** keys to move to another part of the data file. Click on the required cell to select it.

Type the required value. This will replace any existing value in that cell. If you want to delete an existing value without entering a new value, press the **Delete** key.

> Note: String variables are case sensitive (upper and lower case letters are not the same).

To move up or down to another case for the same variable, press the up or down arrow key. Alternatively, pressing the **Enter**/Return key will move to the cell below.

To move horizontally to another variable for the same case, press the left or right arrow key. Alternatively, pressing the **Tab** key will move to the cell to the right. If it is the last variable for that case, it will jump to the first variable of the next case.

Inserting a Case

To insert a new case, select a cell on the row *below* where the case is to be inserted.

From menu, select **Edit** > **Insert Cases** *or* click on **Insert Case** icon on the toolbar.

If you are not concerned with the order of the cases, move to the blank row below the last case. As soon as a value is entered for any of the variables, a new case will be created automatically.

Deleting a Case

Click on the row number at the left-hand end of the row, right-click and select **Cut**.

Saving Your Work

You will have (at least) two files you can save – the Data (**.sav**) file and the Viewer output (**.spv**) file. *Each file has to be saved separately* if you want to keep its contents.

If you want to save any changes to the current file:

> From the **File** menu, select **Save**.

*** Warning ***
SPSS does not create a backup or versioned copy to which you can revert to if things go wrong. Unless you are certain you want to overwrite an existing file, save the file with a new name.

If you want to save the file with a new name:

> From the **File** menu, select **Save As…**.
>
> In the **File name:** box type the name you would like to give to the new file.
>
> If you want to save the file in a different folder or on a different drive,
> use the arrow to the right of the **Look in:** box to change drive/folder.

Help Facilities

Use the **Help** menu to look up information in the index about particular tasks in SPSS. Many dialogue boxes also have their own Help button.

For an overall view of SPSS, a useful starting point is the tutorial. From the **Help** menu, choose **Tutorial**. This can take a few minutes to load.

You can also select a tutorial at the opening SPSS screen menu.

From the **Help** menu, you can also access the **Statistics Coach**. By answering a few questions about the type of analysis you wish to undertake, the Statistics Coach will suggest the most appropriate techniques to use. However, this should be used with caution – you should only use statistical methods that you really understand!

Useful Shortcut Keys

Ctrl+**Home** will take you to the top of the file. In a Data file this will be the first case and the first variable.

Ctrl+**End** will take you to the end of the file. In a Data file this will be the last case and the last variable.

Page Up and **Page Down** will scroll up / down, one page (screen) at a time.

Ctrl+**Page Up** and **Ctrl**+**Page Down** will scroll right / left, one page (screen) at a time.

In a Data file, **Home** will take you to the left-hand end of the current row, i.e. to the first variable and the current case.

In a Data file, **End** will take you to the right-hand end of the current row, i.e. to the last variable and the current case.

In an Output file, **Ctrl**+**A** will select the entire output (tables and graphs).

At many stages (in the Data Editor or in the Output Viewer), clicking the right-hand mouse button will bring up a menu of the tasks that you are most likely to want to undertake with the particular item which you have selected.

Manipulating Data in SPSS

Computing a Variable

It is possible to derive a new variable from existing variables, e.g. by summing the values for several variables together. It is also possible to change an existing variable, e.g. by multiplying it by another variable. In both cases, the compute procedure is used.

From the **Transform** menu, select **Compute Variable…**

In the **Target Variable:** box enter the name of the new variable to be created, or the name of the existing variable whose values you want to change.

If it is a new variable, the **Type&Label…** button leads to another dialogue box where you can make some changes to the definition of the variable you are about to create.

Click the **Numeric Expression:** box and either type or use the various buttons to enter the desired formula. Existing variable names can be pasted into the box by using the
▸ button.

As well as conventional arithmetic and algebraic symbols, there are also several special functions that can be used, e.g. SUM(Q1,Q4,ABC) will add the values of variables Q1, Q4 and ABC together.

If the value of any of the variables used in the formula is missing for a particular case, the value of the new variable will be defined as system missing for that case.

When the **OK** button is clicked, the new variable will be computed (or the existing variable will be over-written with its new values).

Recoding Data

Note: You must be in the data editor window (rather than the output viewer) to recode a variable.

From **Transform** menu, select **Recode…**
into Same Variables… if you want to replace the existing values of a variable,
or **into Different Variables…** if you want to create a new variable, keeping the existing variable unchanged.

Select the variable(s) to be recoded in the left-hand box. Click the ▸ button to copy these to the right-hand **InputVariable->OutputVariable** box.

If you are recoding into a different variable, there will also be an **Output Variable** pane, where you should enter the **Name:** and, optionally, a **Label:** for the new variable. Clicking the **Change** button will cause the **?** to be replaced with the new variable name.

Click the **Old and New Values…** button. A new dialogue box will appear.

On the left-hand side, enter the old **Value** that you want to recode, and on the right-hand side, enter the new **Value** that you want to change it to. Click the **Add** button. Repeat the process for any other values you want to change.

You can also recode a **Range** of values into a single new value.

If there are values you wish to leave unchanged, select **All other values** on the left and **Copy old value(s)** on the right. Again click the **Add** button.

When complete, click **Continue** button. At the main Recode dialogue screen, click **OK**.

Automatic Recoding

In many statistical procedures within SPSS, when category codes are not sequential, the resulting empty cells can substantially reduce performance and increase memory requirements. In addition, some procedures cannot use string variables, and some require consecutive integer values for factor levels.

Automatic Recode converts string and numeric values into consecutive integers.

String values are recoded in strict alphabetical order, with uppercase letters preceding lowercase.

The resulting new variable retains the variable and value labels of the old variable from which it has been recoded. For any values without a defined value label, the original raw value is used as the label for the recoded value.

Missing values are recoded into missing values higher than any non-missing values, with their order preserved. For example, if the original variable has 4 non-missing values, the lowest missing value would be recoded to 5, and the value 5 would automatically be defined as a missing value for the new variable.

From **Transform** menu, select **Automatic Recode…** then

Select the variable to be recoded in the left-hand box.

Click the ▶ button to copy this variable to the right-hand **Variable->New Name** box.

The variable name will appear followed by ->????????

Type a name for the new variable in the **New Name** box.

Click on the **Add New Name** button. The questions marked will be replaced.

Click on the **OK** button to run the recode.

Selecting Cases

Sometimes we want to analyse data for only a sub-set of the cases (perhaps to exclude cases that have unusual values, or to restrict the analysis to cases that meet certain criteria e.g. women aged 60 or over).

From the **Data** menu, select **Select Cases…**

Click on **If condition is satisfied** then click **If…** button. A new dialogue box appears.

In the right-hand box type the criteria for selection. Existing variable names can be pasted into the box by using the ▶ button. Alternatively, you can simply type in the name of the variable. Similarly, you can type in the mathematical symbols, or you can press the appropriate mathematical symbol key on the calculator-style keyboard area.

> ampersand **&** key is the equivalent of logical 'AND'
> vertical bar **|** key is the equivalent of logical 'OR'
> tilde **~** key is the equivalent of logical 'NOT'.

As well as conventional arithmetic and algebraic symbols, there are also several special functions that can be used.

Examples:

Age>59 *selects cases where* value for variable 'Age' was greater than 59.

Age<=21 & Year<>4 *selects cases where* 'Age' is less than or equal to 21
and 'Year' is not equal to 4.

Once you have typed the selection criteria, click **Continue**.

Ensure that **Filter Out Unselected Cases** is selected. Otherwise, cases that do not meet the criteria will be permanently deleted from the file.

Click **OK** to proceed with the selection.

In Data View, cases that have <u>not</u> been selected will appear with a stroke through their case number (down the left-hand side).

There will also be a new variable called **filter_$** which has a value of 0 if the case has not been selected and 1 if it has. It will have a variable label that specifies the criteria by which the cases have been filtered.

To remove the filter and reselect all cases:

From the **Data** menu, select **Select Cases…** Click on **All cases** then on **OK**.

In Data View, there will no longer be strokes through any case numbers. The **filter_$** variable, however, will still be there. If you were to run another Select Cases, the values for this variable would change but not the variable name. The variable can be safely deleted if desired.

Sorting Cases

It is sometimes useful to be able to sort cases, e.g. to group cases with the same values for a variable together to permit a visual check of the data.

From the **Data** menu, select **Sort Cases…**

In the left-hand box, select the variable by which you wish to sort the cases.

Click the ▶ button to copy the variable to the right-hand **Sort by:** box.

You can choose to sort in either **Ascending** or **Descending** order.

It is possible to sort by more than one variable. Cases are ordered firstly by the first variable in the Sort by: box. If there is more than one case with the same value, then those cases will be sorted by the second variable specified, and so on. You can select ascending or descending sort for each of the variables.

Click **OK** to run the sorting procedure.

Note: When you sort cases, the case numbers that appear at the left-hand side in Data View will no longer correspond to the same cases as they did before the sort. If you want to always be able to uniquely identify cases, you should create your own variable for this purpose.

Output from SPSS

Note: It is possible to have more than one Output file open simultaneously, each of which has its own Output Viewer window. (The output does not need to have been created from the active data file!) Any new output will be appended to the Output file which has its "Designate Window" icon (an explanation-mark) selected – it will be red.

Using the output Viewer

The viewer consists of two panes – the **Outline** pane on the left side and the **Results** pane on the right. The Outline pane is the equivalent of chapter or section headings in a book, while the Results pane would be the actual text of the book.

The red arrow in the Outline or Results pane indicates which element in the viewer is active. Also, if you select an element in either pane, a frame will appear around that element in the Results pane.

It can be useful to sometimes hide some of the results so that you can focus on others.

To **HIDE** a table or chart (without deleting it):
Click on the item. From the **View** menu, select **Hide**.

To **SHOW** a table or chart once it has been hidden:
Click on the item in the Outline pane. From the **View** menu, select **Show**.

To **DELETE** a table or chart, click on that element (either in Outline or Results pane) and press the Delete key. Alternatively, from the **Edit** menu, select **Delete**.

If you delete an element accidentally: from the **Edit** menu, choose **Undo**.

To **PRINT** the complete contents of the Viewer:
from the **File** menu, choose **Print…** Click **OK**.

To print only a selection:
Click on the element that you want to print
from the **File** menu, choose **Print…**. Ensure **Selection** is selected. Click **OK**.

To select more than one item, hold down the Ctrl key while you click on the elements.

Editing Tables

To edit a table, double-click anywhere on the table.

The menu at the top of the output window will change. A new toolbar will appear together with a Pivot Tray window.

The toolbar includes tools to change the appearance of any text that you select.

The Pivot Tray window allows you to alter where specific information should appear in the table. Altering the order and position of the various coloured icons around the edge of the example pivot table, alters the position of the statistics in your own table. To move an icon, click on it and drag it to the desired location.

To reverse the column and rows in the table:
from **Pivot** menu, select **Transpose Rows and Columns**

To change the appearance of the table, from **Format** menu:
select **Autofit** to adjust row and column widths and font sizes to fit the contents; or
select **TableLooks…** and choose an alternative style to change the overall design.

Editing Graphs

To edit a graph, double-click anywhere on the graph.

A copy of the chart will appear in a new windows - the Chart Editor. There should also be a **Properties** window. If it is not displayed, from the **Edit** menu, select **Properties**.

Once in the Chart Editor, clicking on a particular element (text, bar, line etc.) within the chart will normally open up the **Properties** window and display the most likely attributes that you will want to change.

There are many things that can be done to customise the chart. A few of the more common options are covered below.

Title

To add a title to appear above the chart, click on the **Options** menu, then select **Title**. This will insert a text box on the graph, which you can edit.

Similarly, you can add a footnote that will appear below the chart: from **Options** menu, select **Footnote**.

Text Format

Click on the text you wish to change. In the **Properties** window, under the **Text Style** tab, you can change the font, size and colour as desired.

Format of Bars and Segments (in charts)

Click on a bar/segment. In the **Properties** window, under the **Fill & Border** tab, you can alter the colours and patterns used for that bar/segment.

When you have finished making changes to the graph, close the Chart Editor window (from **File** menu, select **Close**).

Exporting the Output

[See also p.6 "Saving Your Work"]

The viewer output (**.spv**) files can only be accessed using SPSS version 16 or later. Earlier versions of SPSS have used different formats of output file (**.spo**). SPSS 22 can not read these earlier .spo output files, nor can SPSS versions 15 and earlier read .spv output files.

In SPSS it is possible to "export" the output into other file formats that can be accessed by a greater number of software packages.

From the **File** menu select **Export…**

In the **Type:** box, select the type of file that you want to create. The most useful is likely to be **Word/RTF** file (*.doc) that can be opened using Microsoft Word. Tables are exported with all the formatting attributes retained, e.g. cell borders, font styles and background colours. Charts are exported in Windows metafile format.

In early versions of SPSS for Windows, the only way of accessing the output in other software packages was to select a table or graph then to choose copy, switch to the other software and then paste in the selected item. How to do this to MS Word are detailed below.

Copying a Table into a Word document

In the Output Viewer, select the desired table in the output listing by clicking on it once.

Method 1
– for when you want to be able to edit the table or change its appearance in Word.

In SPSS, from the **Edit** menu select **Copy** (*or*, alternatively, press **Ctrl+C**).
Switch to your Word document.

If using Word 2007:   On the **Home** tab select **Paste** then **Paste**
                      (alternatively press **Ctrl+V**)

If using Word 2010:

    On the **Home** tab select **Paste**
    then **Keep Source Formatting** *(as close as possible to what it looked like in SPSS)*
    or **Merge Formatting** *(makes it more like the rest of your Word document)*
    or **Keep Text Only**
    or press **Ctrl+V** *(pastes using whichever of the above is your default formatting)*

Method 2
– for when you want to retain the table in Word *exactly* as it appeared in SPSS.

In SPSS, from the **Edit** menu select **Copy Special…**
Ensure that the **Image** option is ticked. (There is no need to deselect the others.)

Switch to your Word document.

If using Word 2007   On the **Home** tab select **Paste Special…**
  or Word 2010:   select **Picture** click **OK**

Copying a Graph into a Word document

In the Output Viewer select the desired graph in the output listing.
From the **Edit** menu select **Copy**. (alternatively press **Ctrl+C**)
Switch to Word.

If using Word 2007:

    On the **Home** tab select **Paste** then **Paste**
    or, alternatively, press **Ctrl+V**

If using Word 2010:

    On the **Home** tab select **Paste** then **Keep Source Formatting** (the default)
    or, alternatively, press **Ctrl+V**

Syntax Commands

One of the advantages of the Windows version of SPSS is that most (but not all) tasks can be undertaken by selecting items from menus. In older versions of SPSS, it was necessary to type in all the commands or syntax.

Within SPSS for Windows, the syntax commands are hidden but are still available. Examine any output file in the Output Viewer. In the left-hand Outline frame, double-click on any Notes item to "unhide" the details of the analysis that was undertaken, including the syntax that was used.

When you run an anlysis using the normal Windows menu selection system, instead of clicking on the final **OK** button to run the procedure, press the **Paste** button instead. The syntax commands will be pasted into a file in the Syntax Editor window. A new syntax window will be created if one is not already open.

Any text within the Syntax Editor window can be copied, deleted or amended.

To run (execute) commands:

Highlight the command(s)

From the **Run** menu choose **Selection**.

Note: Unlike SPSS Data (.sav) and SPSS Viewer document (.spv) files, SPSS Syntax (.sps) files can be read and edited in any other software package that can edit a plain text file. If you create a syntax file in Microsoft Word, you should save it as a "Text only with line breaks" (.txt) file but should name it with a .sps extension.

The following points should be noted when creating/editing a syntax file:

− Each command must begin on a new line and end with a period (.) i.e. a full-stop.

− Most subcommands are separated by slashes (/).

− Variable names must be spelled out fully.
Most built-in SPSS commands can be abbreviated to the first three characters.

− Each line of command syntax cannot exceed 80 characters.

− Text within apostrophes or quotation marks must be contained on the same line.

− You can add space or break lines at almost any point where a single blank is allowed, such as around slashes, parentheses, arithmetic operators, or between variable names.

− Command syntax is case insensitive (apart from the the values of string variables).

Section 2 – STATISTICAL PROCEDURES

Tips on Analysing a Data Set in SPSS

1. Before starting on the detailed analysis of a data set, make a preliminary inspection of all the data. As well as highlighting possible errors or spurious values in the data, this will give you a flavour of the types of data and the range of values. It will also show whether there are sufficient numbers of observations for some of the analysis that you may have been considering.

   - Use the **Frequencies** procedure [p.17], selecting all the variables, remembering the number of values to include in tabulations but including a few statistical measures of central tendency and dispersion.

2. Having undertaken the preliminary inspection, decide on which variables will be the most suitable to summarise the general characteristics of the observations (cases) that are included in the data set. Take into consideration whether you will be wanting to compare the results with known characteristics of the statistical population or with those cases included in other research studies.

   - For *nominal* or *ordinal* data, the mode is usually the only sensible measure of central tendency, although the median may be appropriate for some ordinal variables. The data would probably be best summarised by a bar- or pie-chart.

   - For *interval* or *ratio* data (scale variables in SPSS), there will usually be a large number of values so that it is not sensible to produce contingency (frequency) tables or bar- and pie-charts. Instead, use the descriptive statistics that summarise central tendency and dispersion. Histograms and box-plots will probably be the best choice for summarising the data graphically.

3. The same procedures outlined in 2 above can be used to summarise the results for other variables on the data file (e.g. the response to a question about a specific issue under investigation).

4. If there are too many discrete values for a variable, it is sometimes useful to **Recode** [p.8] the variable into a new variable (for safety!) with a smaller number of categories. If the original was of interval or ratio type, the new variable will be ordinal. You will need to **define** the variable you have created and add value labels [p.3].

5. We usually also want to try and explain why a particular pattern is observed, and this will usually mean testing for relationships between two or more variables. Remember, a statistical association between variables does not necessarily mean that there is a causal relationship.

   To depict in tabular form the association between two *nominal* or *ordinal* variables, use the **Crosstabs** procedure [p.18] to produce a two-way contingency table. A clustered or stacked bar-chart would be a good choice for a graph.

   - To depict in tabular form the association between an interval (or ratio) variable and a nominal (or ordinal) variable, use either the **Explore** procedure [p.18] or **Compare Means** [p.18] procedure. A box-plot will be a good choice of graph.

   - To depict in tabular form the association between two *interval* or *ratio* variables, use the bivariate **Correlate** or **Regression** procedures [p.20]. A scatter-plot (perhaps including a best-fit trend line) might be a good choice of graph [p.19].

Which Procedure to select?

If you are not sure which statistics you should use, from the **Help** menu, select **Statistics Coach**. By answering a short series of questions about the type of data you are analysing and the type of analysis you wish to undertake, you will be guided to the most appropriate procedure to use.

Only use statistical procedures that you fully understand. Read the texts on the course reading list to remind yourself of the advantages/disadvantages and uses of the various statistics. SPSS can generate lots of statistics, but they may not always be appropriate.

Which Graph to select?

A wide selection of statistical charts can be produced in SPSS.

Many statistical procedures include an option to generate charts. In addition, you can chose from the **Graphs** menu.

Legacy Dialogs (accessed from the main **Graphs** menu) leads to a series of different graph/chart types from which to select. In all cases, you will then be asked to **Define** the graph by specifying the variables to be displayed and how they are to be plotted. The precise information required and the format of the dialogue boxes will vary according to the type of graph you select.

If you are not sure what each chart type looks like, select **Graphboard Template Chooser** (from the main **Graphs** menu). You have to ensure that the variable has the correct Measure (nominal, ordinal or scale) defined in Variable View.

Univariate Descriptive Statistics

Frequency Distributions

Frequency distributions are tabular or graphical presentations of data that show each category for a variable and the frequency of the category's occurrence in the data set.

In SPSS, the "Frequencies" procedure is usually used for nominal and ordinal types of variables, e.g. gender, political party.

From **Analyze** menu, select **Descriptive Statistics**, then **Frequencies…**
In the left-hand box, select the variable(s) for which you wish to produce a frequency distribution. Click the ▶ button to copy the variable to the right-hand **Variable(s):** box.
Note that the frequency table will only be produced if the **Display frequency tables** box is ticked.

It is also possible to produce a graph of the frequency distribution.
Click the **Charts…** button and select **Bar charts**, **Pie charts** or **Histograms** as desired.
If you choose, a bar chart you can also specify whether to display the number of cases (**Frequencies**) or the percentage of cases (**Percentages**) on the scale axis.
Click **Continue** to return to the main dialogue box.

The **Statistics…** button opens a dialogue box where you can select various measures of Central tendency, Dispersion, Distribution and Percentiles. In selecting a specific statistic, you should ensure that it is an appropriate measure for the type of variable being analysed. For scale variables, you will usually only want statistics, without a frequency table or chart.

Click **OK** to run the procedure and produce the requested tables, charts or statistics.

Tip: clicking on the Help button within a pop-up window will provide further information about the use or meaning of the various items and options.

Descriptives

In SPSS, the "Descriptives" procedure is used for summary descriptive statistics for quantitative variables, especially where a variable has a large number of values, e.g. age, income. The results are similar to that for "Frequencies" but without a frequency table or chart.

From **Analyze** menu, select **Descriptive Statistics**, then **Descriptives…**
In the left-hand box, select the variable(s) for which you wish to produce statistics. Click the ▶ button to copy the variable to the right-hand **Variable(s):** box.

Click on the **Options…** button to alter the specific statistics you want produced.

Examining Associations between Variables

Cross-tabulation   (categorical variable by categorical variable)

Contingency tables (cross-tabulations) provide a summary of the number of cases (observations) which have particular combinations of values for two or more variables. The "Crosstabs" procedure is usually used when the variables are nominal or ordinal.

From the **Analyze** menu, select **Descriptive Statistics**, then **Crosstabs…**

In the left-hand box, select the variables which you wish to appear in the cross-tabulation tables. Click the ▸ button to copy one variable to the right-hand **Row(s):** box, and one variable to the **Column(s):** box.

Clicking the **Cells…** button opens a dialogue box where you can specify what is to appear in the tables.

Recommended settings are **Observed** Counts and both **Row** and **Column** Percentages.

Click **Continue** to exit Cells… dialogue box, then **OK** to run the procedure.

It is also possible to produce 3-way cross-tabulations where separate 2-way tables are produced for each value of a third variable. The procedure is the same as above, except that the third variable is added to the **Layer** box.

If you select **Display clustered bar charts** a bar chart will be produced with separate 'bars' for each value of the column variable against each value of the row variable. If a third 'layer' variable is included, there will be several bar charts, one for each value of the layer variable.

Click on the **Statistics…** button if you want to include any statistical tests.

Explore Data   (scale variable by categorical variable)

If you want to produce statistics about the dispersion and distribution of a scale (ratio or interval) variable, as in the "Frequencies" procedure (p.17 above), but controlling for a particular categorical (ordinal or nominal) variable, the Explore procedure can be used.

From **Analyze** menu, select **Descriptive Statistics**, then **Explore…**

In the left-hand box, select the variable for which you wish to produce the statistics. Click the ▸ button to copy the variable to the right-hand **Dependent List:** box.

In the left-hand box, select the independent variable. Click the ▸ button to copy the variable to the right-hand **Factor List:** box.

Select which information you want to Display: **Statistics**, **Plots** (i.e. graphs) or **Both**.

Click on the **Statistics…** and **Charts…** buttons to open dialogue boxes if you want to change the default settings.

> One of the available charts is a boxplot (also known as a box-and-whiskers plot). The box represents the inter-quartile range that contains the middle 50% of values. The thick line across the box indicates the median. The whiskers are lines that extend from the box to the highest and lowest values, excluding any outliers. Oultiers are those values that lie at least 1.5 box lengths above or below the box.

Click **OK** to run the procedure and produce the requested plots and statistics.

Compare Means (scale variable by categorical variable)

Sometimes, we want to produce a table that summarises how a scale (ratio or interval) variable varies between different values of one or more categorical (ordinal or nominal) variables, e.g. how income varies by gender and ethnic group.

From **Analyze** menu, select **Compare Means**, then **Means…**

In the left-hand box, select the variable for which you wish to calculate the mean values. Click the ▸ button to copy the variable to the right-hand **Dependent List:** box.

In the left-hand box, select the independent variable. Click the ▸ button to copy the variable to the right-hand **Independent List:** box.

Using one independent variable will generate means for the dependent variable for each value of the independent variable.

We can add a second independent variable (in SPSS terminology, a layer) by clicking on the **Next** button and then selecting another independent variable as before.

> Adding a second layer, allows us to produce a table similar to a cross-tabulation but displaying means (rather than percentage of cases).

Select which information you want to display: Click on **Options…** button.
Default is mean, number of cases and standard deviation.

Click **OK** to run the procedure and produce the output.

Scatterplot (scale variable by scale variable)

Scatteplot provides a graphical depiction of how the distribution of one continuous scale variable varies by another continuous scale variable, e.g. income by length of employment.

From **Graphs** menu, select **Legacy Dialogs**, then **Scatter/Dot…**

Click on the icon representing a **Simple Scatter**, then click on the **Define** button.

In the left-hand box, select the variable that you wish to place on the vertical axis of the graph.
Click on the ▸ button to copy the variable to the right-hand **Y Axis:** box.

Similarly, select the variable for the horizontal axis and copy to the **X Axis:** box.

Click **OK** to produce the output.

If you wish to add a best-fit regression/trend line between the two variables, double-click on the graph to edit it (see p.12 above).

From the **Elements** menu, select **Fit Line at Total**.

(If you want to fit separate lines for each subgroup that has been selected in "Set Markers by", select **Fit Line at Subgroups** instead.)

When you have finished making changes, close the Chart Editor window to return to the normal output viewer.

Correlation (ordinal/scale variable by ordinal/scale variable)

The "Correlation" procedure provides a statistical summary of the association between two variables that are at least ordinal (more ideally, continuous scale).

From **Analyze** menu, select **Correlate**, then **Bivariate…**

Using the ▸ button copy the two variables you wish to correlate to the right-hand **Variables:** box.

Select the appropriate Correlation Coefficients, according to the type of variables.

Click **OK** to produce the output.

Regression

The "Regression" procedure provides a statistical summary of the extent to which one dependent variable is determined by the value of one or more independent variables. Whereas correlation indicates some association between variables, regression is used when we believe there to be a causal relationship.

From **Analyze** menu, select **Regression**, then **Linear…**

Using the ▸ button copy the dependent variable to the right-hand **Dependent:** box.

Similarly, copy the independent variable to the **Independen(s):** box.

Click **OK** to produce the output.

The above Regression is used when we expect a straight-line linear relationship between the variables. Much more advanced methods are also available within other SPSS statistical procedures.

Analysing Multiple-Response Variables

Defining Multiple-Response Sets

The **Define Multiple Response Sets** procedure groups elementary variables into multiple dichotomy and multiple category sets, for which you can obtain frequency tables and crosstabulations.

To Define Multiple Response Sets:

From the **Analyze** menu choose **Multiple Response** then **Define Variable Sets…**

> Select two or more variables into the **Variables in Set:** box
>
> > If your variables are coded as dichotomies, indicate which value you want to have counted. If your variables are coded as categories, define the range of the categories.
>
> Enter a unique **Name** for each multiple response set.
> Assign the set a **Label** (similar to a variable label but only up to 40 characters).
>
> Click **Add** to add the multiple response set to the list of defined sets.

You can define up to 20 multiple response sets. Each set must have a unique name. To remove a set, highlight it on the list of multiple response sets and click **Remove**. To change a set, highlight it on the list, modify any set definition characteristics, and click **Change**.

You can code your elementary variables as dichotomies or categories. To use dichotomous variables, select Dichotomies to create a multiple dichotomy set. Enter an integer value for Counted value. Each variable having at least one occurrence of the counted value becomes a category of the multiple dichotomy set. Select Categories to create a multiple category set having the same range of values as the component variables. Enter integer values for the minimum and maximum values of the range for categories of the multiple category set. The procedure totals each distinct integer value in the inclusive range across all component variables. Empty categories are not tabulated.

Each multiple response set must be assigned a unique name of up to seven characters. The procedure prefixes a dollar sign ($) to the name you assign. You cannot use the following reserved names: CASENUM, SYSMIS, JDATE, DATE, TIME, LENGTH, and WIDTH.

The name of the multiple response set exists only for use in multiple response frequencies and crosstabs procedures (as described below). You cannot refer to multiple response set names in other procedures.

Multiple-Response Frequencies

The **Multiple Response Frequencies** procedure produces frequency tables for multiple response sets. You must first define one or more multiple response sets.

For multiple dichotomy sets, category names shown in the output come from variable labels defined for elementary variables in the group. If the variable labels are not defined, variable names are used as labels. For multiple category sets, category labels come from the value labels of the first variable in the group. If categories missing for the first variable are present for other variables in the group, define a value label for the missing categories.

Missing Values. Cases with missing values are excluded on a table-by-table basis. Alternatively, you can choose one or both of the following:

Exclude cases listwise within dichotomies. Excludes cases with missing values for any variable from the tabulation of the multiple dichotomy set. This applies only to multiple response sets defined as dichotomy sets. By default, a case is considered missing for a multiple dichotomy set if none of its component variables contains the counted value. Cases with missing values for some (but not all variables) are included in the tabulations of the group if at least one variable contains the counted value.

Exclude cases listwise within categories. Excludes cases with missing values for any variable from tabulation of the multiple category set. This applies only to multiple response sets defined as category sets. By default, a case is considered missing for a multiple category set only if none of its components has valid values within the defined range.

Example. Each variable created from a survey question is an elementary variable. To analyze a multiple response item, you must combine the variables into one of two types of multiple response sets: a multiple dichotomy set or a multiple category set. For example, if a consumer survey asked which of four newspapers (Herald, Gazette, Record, Times) respondents had read in the previous seven days and you used dichotomous variables and defined a multiple dichotomy set, each of the four variables in the set would become a category of the group variable. The counts and percentages for the four papers are displayed in one frequency table.

If you discover that no respondent mentioned more than two of the newspapers, you could create two variables ("paper1" and "paper2"), each having four codes, one for each paper. If you define a multiple category set, the values are tabulated by adding the same codes in the elementary variables together. The resulting set of values is the same as those for each of the elementary variables. For example, 25 responses for Herald are the sum of the 14 Herald responses for variable "paper1" and the 11 Herald responses for variable "paper2". The counts and percentages for the newspapers are displayed in one frequency table.

Statistics. Frequency tables displaying counts, percentages of responses, percentages of cases, number of valid cases, and number of missing cases.

Multiple-Response Crosstabulations

The **Multiple Response Crosstabs** procedure crosstabulates defined multiple response sets, elementary variables, or a combination. You can also obtain cell percentages based on cases or responses, modify the handling of missing values, or get paired crosstabulations. You must first define one or more multiple response sets.

For multiple dichotomy sets, category names shown in the output come from variable labels defined for elementary variables in the group. If the variable labels are not defined, variable names are used as labels. For multiple category sets, category labels come from the value labels of the first variable in the group. If categories missing for the first variable are present for other variables in the group, define a value label for the missing categories. The procedure displays category labels for columns on three lines, with up to eight characters per line. To avoid splitting words, you can reverse row and column items or redefine labels.

Statistics. Crosstabulation with cell, row, column, and total counts, and cell, row, column, and total percentages. The cell percentages can be based on cases or responses.