An Introduction to
IBM SPSS v22

Using a Computer to Analyse Data – IBM SPSS (v22)

Contents:

As a student or in work, you may be asked to analyse data. You can use a computer to help you do this. One application used for this is **IBM SPSS** (Statistical Package for the Social Sciences) which is available on all university computers. The following notes will describe how to work with IBM SPSS and will introduce some of the most useful features of the program.

Note: Prior to using IBM SPSS you should have an understanding of basic statistics and the outcomes for the data set you are working with.

When/if Things Go Wrong

Remember that when/if an error occurs try the following to rectify the situation:

Use the **Undo** option by either selecting the **Edit** menu and choosing **Undo**, or clicking on the **Undo** icon .

Help

If there is a feature of IBM SPSS you would like to use but do not know how, use the help facility within the software. This provides instructions on using all features of the software. To use help click on the **Help** menu and from the drop-down list choose **Topics**. A window will open displaying descriptions of the other help options and a search engine allowing you to search for specific help topics.

What is IBM SPSS?

IBM SPSS is a data analysis program mostly used for the analysis of survey and questionnaire data. Using it will enable you to analyse data and present results from this analysis in text and graphical format.

Starting IBM SPSS

To open **IBM SPSS** click on the **Start** button and from the **Start** menu select **All Programs.**

Then click on the **SPSS Statistics** folder, then Σ **IBM SPSS Statistics 22.** The IBM SPSS Licence window will be displayed.

The IBM SPSS Statistics "What would you like to do?" window will then be displayed, with the IBM SPSS Statistics Data Editor window showing behind, **Figure 1**.
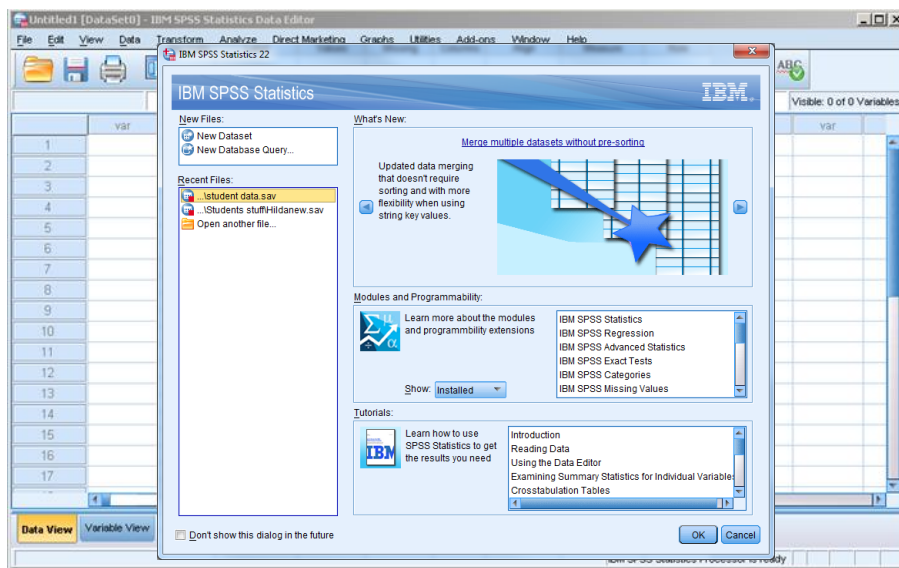
Figure 1: IBM SPSS "What would you like to do?" window.

If this is the first time you have used this, or you are creating a new data file, choose the **New Dataset** option from the **New Files:** field by clicking on it and then clicking **OK**. If you have used this before you can open an existing data source from the **Recent Files**: field. The **IBM SPSS Data Editor** window will now be visible, ready for the insertion of data, **Figure 2**.

Task 1

Open the **IBM SPSS** program. Choose Type in data.
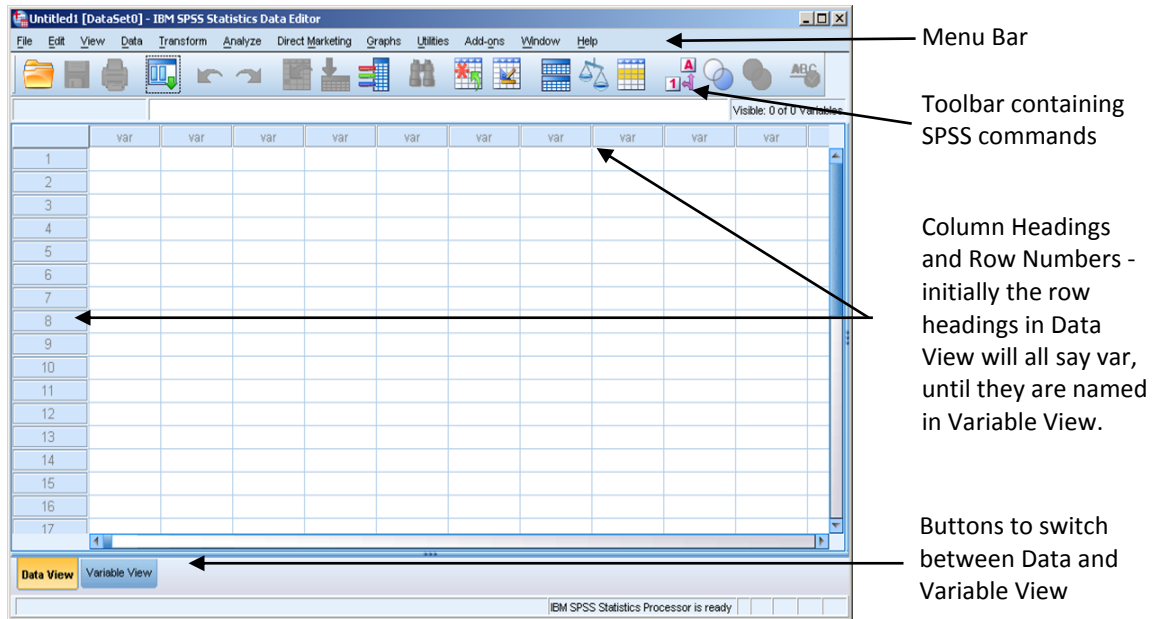
Figure 2: IBM SPSS Statistics 22 Data Editor Window in Data View

When **IBM SPSS** opens, the **Data Editor** window is displayed ready for the insertion of data. If you have used this before you can open an existing data source from the **File** menu.

Switching between Data/Variable View

In **IBM SPSS** there are two views which can be used **Data View** and **Variable View**. By default the **Data View** will be the one which is showing. To switch between the views click on the buttons at the bottom left hand side of the **Data Editor** window, **Figure 3**.
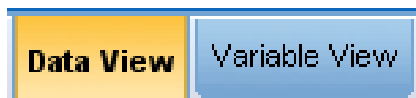
Figure 3: Data/Variable View buttons

- The **Data View**, **Figure 2** allows you to enter and edit data (values) in the **Data Editor** window. All column headings show the default label "**var**".
- The **Variable View, Figure 4** allows you to define the names and properties/characteristics of the variables in the dataset.
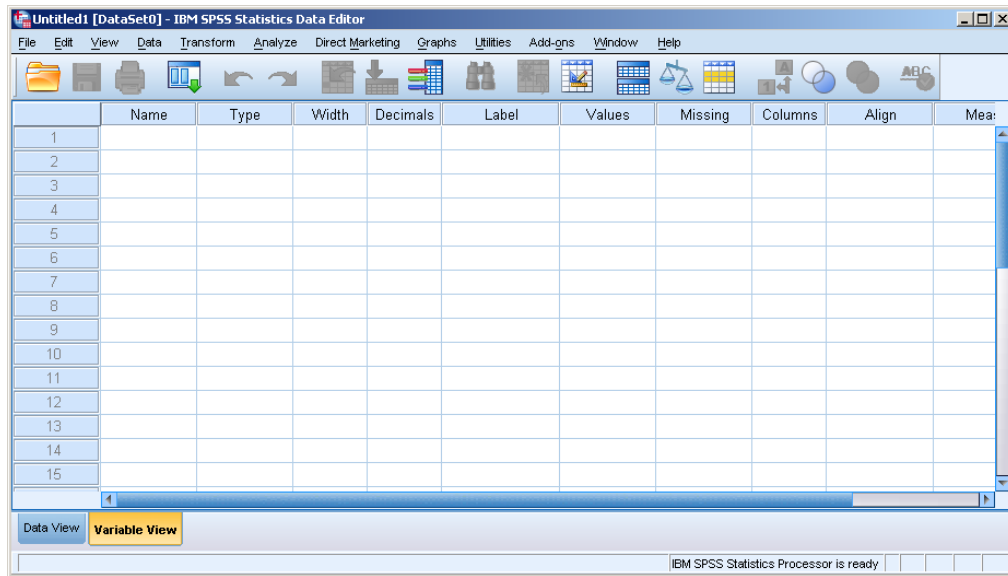
Figure 4: Variable View

 Task 2

Switch between the two different views in the Data Editor window.

 Note: When switching between views note the change in the column headings.

Creating a Data File

When creating a data file you can enter the data directly in **Data View** and let the computer configure the variable properties, or choose **Variable View** to name the variables and set properties before entering the data. It is easier to name the variables and enter the properties for them before entering the actual data values, by doing this you can ensure that the variable properties are configured correctly.

Setting up the Data Variables

To create the data file select V**ariable** V**iew** by clicking on the **Variable View** button. In **Variable View**, each row in the **Data Editor** represents a variable from your data set, for example, gender. Each column represents a property of that variable, see **Figure 4**.

 Task 3

Switch to Variable View if you are in Data View.

When the variable name is typed in the **Name** column the other columns in the row will be filled with default properties. These properties can be changed to best reflect the data being entered. The properties and their functions are listed below in **Table 1**.

**Note:** You cannot leave spaces, or insert a hyphen in the Name field**.** Ensure you choose variable names which describe the data to be entered as this will make it easier to interpret and analyse later.

Variable properties	Description
Name	A name of your choosing for the variable, it should describe the variable for example Age.
Type	The type of data to be recorded. By default this column will show as numeric for the entry of numbers. You can change this to String for characters, Scientific Notation or Date.
Width	The maximum number of characters/numbers required to display the data, by default this will be 8. If you require more or less you must change this.
Decimals	The number of decimal places which will be displayed, by default this will be 2. If you wish more or less you must change this.
Label	A text entry to describe the data provided by the variable. This can be longer than the variable name and may include spaces. For example our variable TrainReq can be entered as Training Required.
Values	Should only be used for variables which are categorical or ordinal enabling you to specify what a value actually means for example, 1 = male and 2 = female.
Missing	Indicates whether a value is present or not.
Columns	The width of the variable column in Data View
Align	The alignment of the data can be set to Right/Left/Centre
Measure	Three option available: nominal (discrete values with no intrinsic ranking), ordinal (discrete values with intrinsic ranking e.g. satisfaction from 1-5) and scale (continuous values, e.g. number of students in a class, height, weight).
Role	Indicates the role the variable will play in the data.

Table 1: Variable properties and descriptions

For example a variable name of Organisation could be entered. When this is entered the **Type** column will automatically show as **Numeric**. If you wish to enter the name of an organisation in this field for example Glasgow Caledonian University, then it is not numeric data and should be changed. To change the data type click in the cell (for our example this states **Numeric)**. A small grey box will appear at the right hand side of the cell with leader dots **…,** **Figure 5**.

Type: property.

menu box
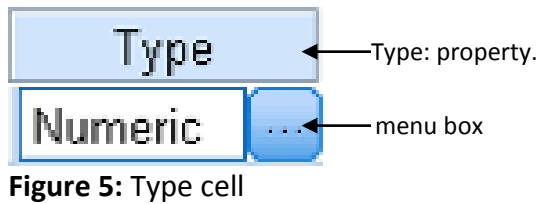
Figure 5: Type cell
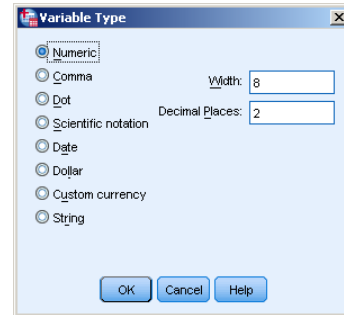
Figure 6: Variable Type dialogue box

Click on the grey box and the **Variable Type** dialogue box will open presenting options to choose from.

For Organisation the Type would be String. Click the radio button beside this option. Then click **OK**.

As each variable has its own specific properties, in some cases you may wish to assign a value rather than a descriptive label for analysis purposes. For example, if one of the data variables is eye colour (Blue, Brown, Green, or Grey), you may wish to enter a numeric value for each option 1; 2; 3 etc. This numeric value will not intuitively tell the user what the data means, so a descriptive label should be assigned for example, Blue=1, Brown=2 etc. To do this, click in the **Values** cell of the variable row you are entering. The grey box appears at the side, click on it. The **Value Labels** dialogue box will open.
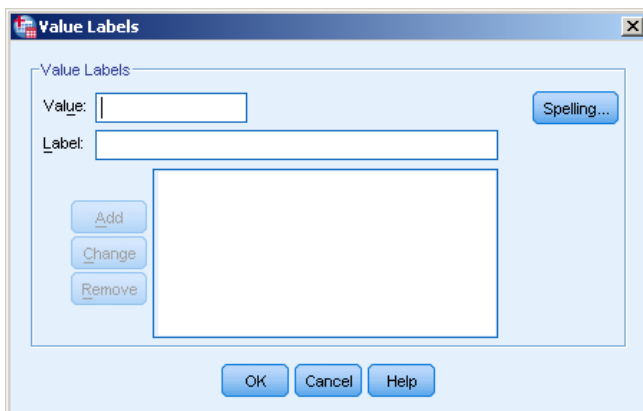
Figure 7: Value Labels dialogue box

To enter values for a label, type the value number in the **Value** field. Next enter the descriptive label for the value in the **Label** field. The **Add** button will now become active. Click **Add**. The value and corresponding label will appear in the large white field below. Once you have added all of the values and labels for that variable click **OK**. The dialogue box will close and the options will be displayed in the cell.

Task 4

Enter the following data in IBM SPSS in Variable View.

| Name	Type	Width	Decimal	Label	Value	Missing	Column	Align	Measure	Role
Course	**String**	**15**	**0**			**None**	**15**	**Left**	**Nominal**	**Input/ None**
Gender	**Numeric**	**1**	**0**		**0=Male 1=Female**				**Nominal**	**None**
Age										**None**
TrainReq										**None**
TrainReqHours										**None**

Choose appropriate properties for each column heading for each variable. Course has been shown. Choose properties for the other four variables. The data for Training Required is ranked 0 – 4 with 0 = No Training, 1 = Little Training, 2 = Some Training, 3= More Training, 4 = Extensive Training. Use the table to make choices for the other variables' properties.

Saving a Data File

Once you have entered the variables and defined their properties you should save the file with an appropriate name. To save a data file click on the **File** menu and choose **Save As**. The **Save Data As** dialogue box will open, **Figure 8**.

Look in: drop down menu

File name: field

Save as type: field
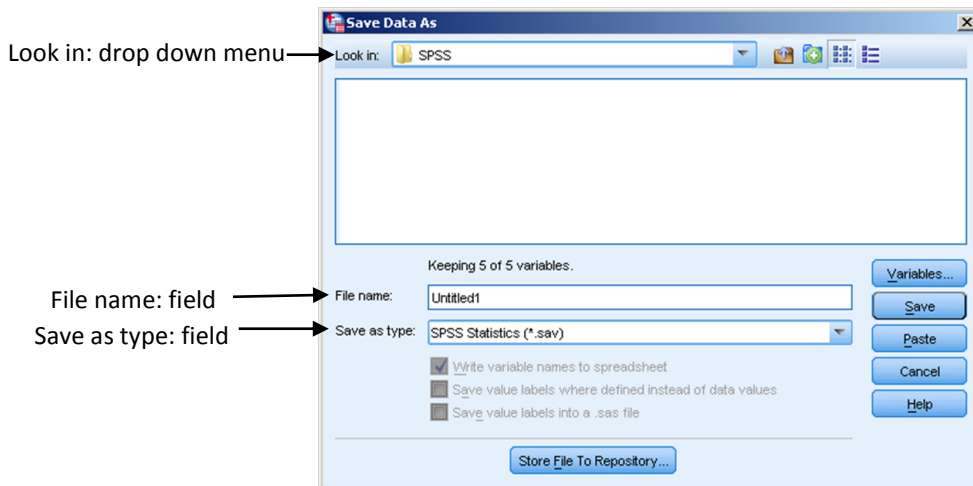
Figure 8: Save Data As dialogue box

Choose a location for your file from the **Look in:** drop-down menu (H: drive in University or USB drive or C: drive from home) Choose an appropriate file name and enter it in the **File Name** field. Click **Save**. IBM SPSS files are saved with the file extension **.sav**. When you click on **Save** the **IBM SPSS Statistics Viewer** window will open. This is known as the **Output** window.

> **Task 5**
>
> Save your data file with the name **Student Analysis**. The related Output file will now be displayed.

The Output Window

When you create a **IBM SPSS** data file you will also create an Output file. When you save your **IBM SPSS** data the **Output 1 [Document 1] - IBM SPSS Statistics Viewer** window will open. This Output window will show recent activities, i.e. information on the data file you have just saved, **Figure 9**. When you perform any statistical analysis or create graphs or charts they will also appear in the Output window.
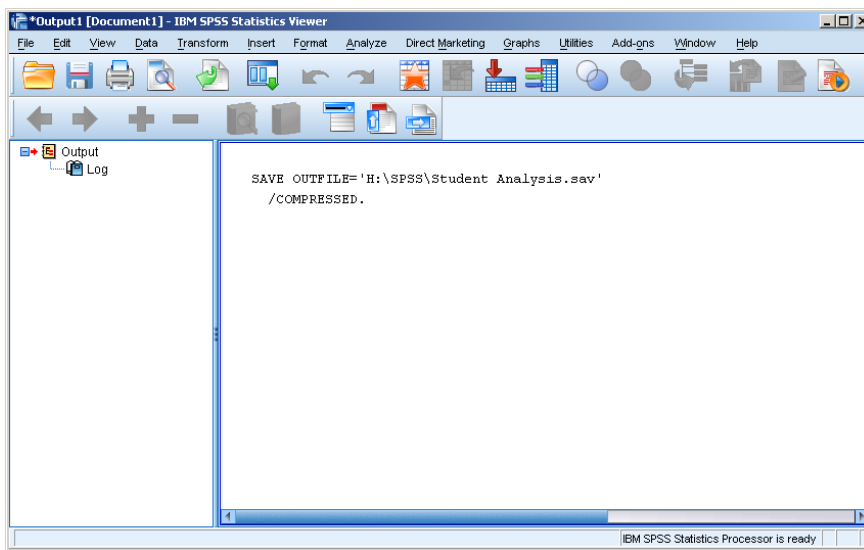
Figure 9: Output 1: IBM SPSS Statistics Viewer

 Note: If you have not yet saved the data file there may be no Output window showing.

If you close this window you will be given the option to save it. If it contains information you wish to keep you should save it with a name which reflects the information in it. Once saved, it will automatically close. Alternatively, save it by choosing the **File** menu, **Save As** and enter the information required.

Copying Data from the Output Window

To copy data from the Output window click on the data you would like to copy, a box will appear around it to show it has been selected. Click on the **Edit** menu and choose **Copy** or right mouse click on the box and choose **Copy** from the pop-up menu. The selection will be copied to the clipboard of the computer. Open the destination file and paste it either using the commands of the software or simply right mouse click at the entry point and choose **Paste** from the pop-up menu. The selection will be copied.

Removing Text from the Output Window

Click the text that you wish to remove and a box will appear around it. Press the **Delete** key on the keyboard. The selection will be removed from the Output window.

Printing Output

Highlight the section you wish to print by clicking on it. A box will appear around the selection. From the **File** menu choose **Print**. The **Print** dialogue box will open displaying the printing options. Click on the **Selected output** radio button to ensure that you print only the selection you want. Then click **OK**. You can also print the entire Output window by clicking on the **All visible outputs** radio button. Click on the **OK** button to print.

Note: If you wish to print the Output window you must ensure it is the active file by clicking in it. Printing can also be carried out by copying and pasting the data from the Output window into a Word document. This may be easier to manage and handle.

Closing a IBM SPSS Data File

To close the application, click on the close icon at the top right hand corner. The **IBM SPSS Statistics 19** dialogue box will open warning that closing the last **Data Editor** window will **Exit IBM SPSS Statistics**. If you are sure you wish to do this click on the **Yes** button. Alternately from the **File** menu choose **Exit.**

Task 6

Close the Ouptut file without saving, we do not require it at this stage. Close the **IBM SPSS** program.

Opening a Saved Data File in IBM SPSS

To re-open the **IBM SPSS** program, either double click on the **IBM SPSS 22** icon if you have one on the desktop, or click on the **Start** menu, **All Programs** and then the **SPSS Inc** folder, and **IBM SPSS Statistics 22** from the sub menu. Once the application is opened the **Data Editor** window will be displayed.

To open an existing data file, click on the File menu and choose **Open**. A sub menu will be displayed, choose the **Open Data** option. The **Open Data** dialogue box will open, **Figure 10**. Navigate to the drive and folder where you saved your file and double click on the file name or click once on the file name and click on the **Open** button. The data file will open in the **Data Editor** window. An Output window will also open showing your recent activity, i.e. that you have opened a specified data file from a specified location.
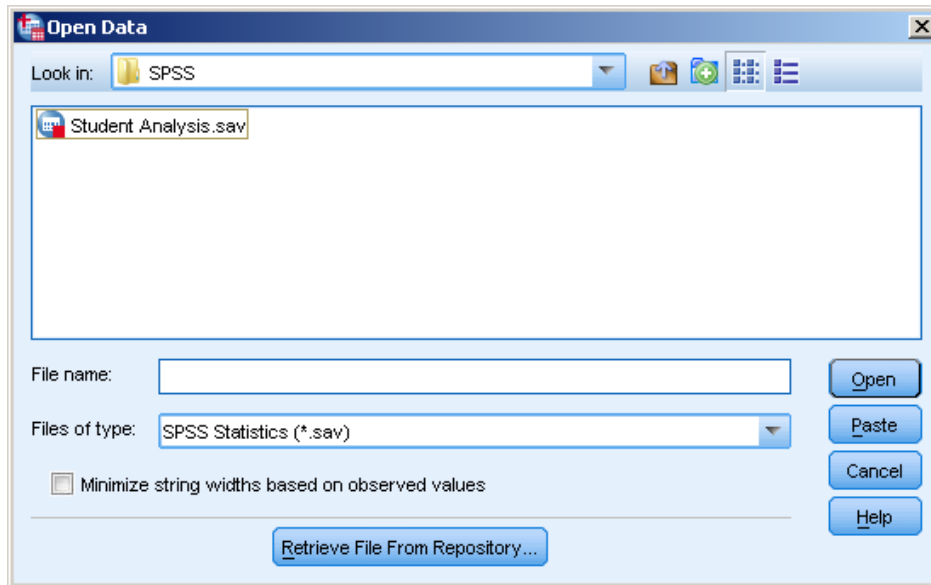
Figure 10: Open Data dialogue box

Note: The data file can be opened by double clicking the relevant data file name

(data file icons look like this,). **IBM SPSS** Statistics will open showing the **Data Editor** and file contents ready for use.

A saved output file can be opened by double clicking the relevant file name (output file icons look like that).

Task 7

Re-open the Student Analysis file in any of the ways listed above.

Entering Data Values

To enter the values, after you have defined the name and properties of the variables, switch to **Data View** by clicking on the button. Then enter the values for each of variable. If you have coded some of the variables, i.e. male=0, female=1, you need to enter the digit representing the value.

Each set of values entered for all variables (one row on the Data View table) is called a case.

Task 8

Enter the following values in the Data Editor (in Data View). If some variables are coded you have to enter the code representing the value.

Course	Gender	Age	TrainReq	TrainReqHours
Engineering	**Male**	**21**	**2**	**4**
Computing	**Female**	**25**	**3**	**6**
Business	**Male**	**18**	**0**	**0**
Fashion	**Male**	**22**	**1**	**2**
Fashion	**Female**	**45**	**0**	**0**
Health	**Female**	**22**	**0**	**0**
Health	**Female**	**30**	**0**	**0**
Computing	**Male**	**18**	**1**	**2**
Engineering	**Male**	**21**	**4**	**8**

Save the data file.

Adding Data from Other Files

More values can be added to the data file created by using Word or Excel files. This can be done, but care must be taken to ensure that the variables within both data files are the same or a mismatch will occur.

To add more values to your data file from Microsoft Word or Excel:

- Open the Word or Excel file containing the additional values.
- Open your IBM SPSS data file if it is not opened.
- Check if the values match the variables in your data file. Any non-compatible data will not be entered.
- From the Word or Excel file select all the values you want to put in your data file (make sure you do not select any headings) and copy them (click on the 🗎 icon on the **Home** tab, **Clipboard** group).
- Go to the IBM SPSS data file and make sure you are in Data View.
- Click in the first empty space after your current values and paste the data (click on the **Edit** menu and select **Paste**). If the data is not in exactly the same format an error warning will be given.

Task 9

Open the Additional Data File.docx which has been supplied, check that all of the information matches. If there are any mismatches fix them and then copy and paste the data at the end of the Student Analysis data file.

Task 10

Open the Further Data file.xlsx and add the data to the bottom of the Student Analysis data file.

Note: Remember to check that the data matches and correct any mismatch.

Selecting Parts of the Data File

When using IBM SPSS you do not need to use the whole of the data file that you have entered. Parts of the data file can be used by filtering them out. For example you may wish to analyse the data for the female respondents or for students over a particular age.

To select parts of a file you go to the **DATA** menu and choose **Select Cases**. The **Select Cases** dialogue box will open, **Figure 11**.
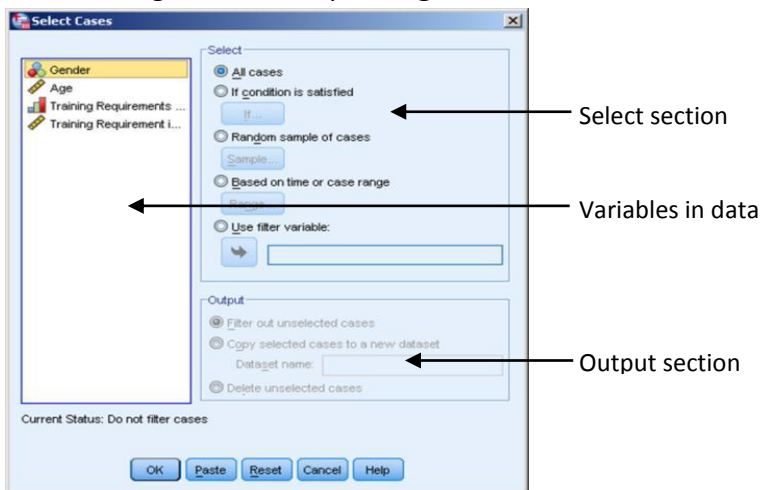
Figure 11: Select Cases dialogue box

From the **Select** section click on the **If condition is satisfied** radio button. The **If…** button will become active. Click on it and the **Select Cases: If** dialogue box will be displayed, **Figure 12**.
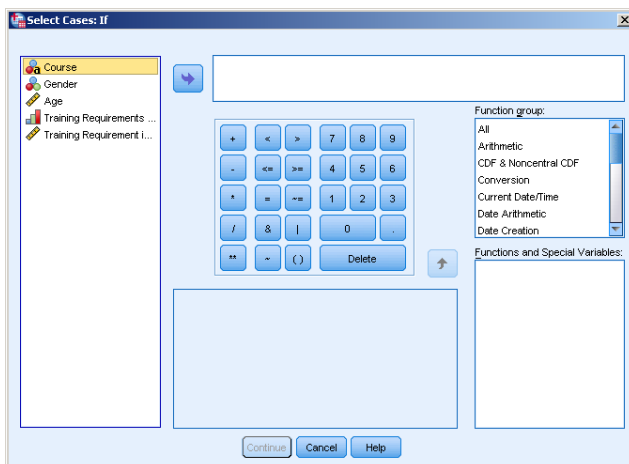
Figure 12: Select Cases: If dialogue box

Choose the variable you wish to select from the list on the left hand side by clicking on it and it will become highlighted. Then click on the arrow button to move it to the field at the right. Complete the expression which will filter the data you require by using the variables from your data set and the numeric key pad displayed (you can also use the keyboard to type in any characters if required). For example, to filter the female recipients it would be **GENDER=1**. To use a string variable you have to type in the filtering criteria in quotation marks, i.e. **COURSE="Computing"**. Use & (AND) if you need to add more than one criteria. Click on the **Continue** button at the bottom of the **Select Cases: If** dialogue box to finish constructing the filter.

The **Select Cases** dialogue box will still be open. You can then choose how you wish your selected data to be displayed from the **Output** section. By default the **Filter out unselected cases** option is selected, if you wish to use on of the other options click the radio button beside it. Once you have chosen your preferred options click on the **OK** button.

It make take a few minutes for the information to be processed, the processing will show in the Output window. When it has finished the filtered rows will have a score through their row number which is shown at the left hand side of the file and a new column will appear within the window. The new column will be named **filter_$** and a values of 1 or 0 will be showing if a case is selected or filtered in (1) or it has not been selected and has been filtered out (value 0).The filter will also appear as a variable in the Variable View of your data file.

	Course	Gender
1	Engineering	0
2	Computing	1
3	Business	0
4	Fashion	0
5	Fashion	1

Figure 13: Filter showing by score through on row number

filter_$
0
1
0
0
1
1

Figure 14: New Filter column

The words **Filter On** will appear at the bottom of the window in the information area as a reminder.

Filter On

Figure 15: Filter On reminder in task bar.

> **Task 11**
>
> Filter the data set to show students above the age of 30 who require training.

Removing a Filter

From the **Data** menu choose **Select Cases**, the **Select Cases** dialogue box will open. Click on the **All Cases** radio button in the **Select** section and then click on **OK**. The filter will be removed. If you do not plan to use the filter again delete the filter variable from the Variable View as well.

Task 12

Remove the filter. Remove the filter variable from the Variable View as well.

Graphs and Charts

Graphs and Charts are visual depictions of data and can be used to display facts and information about data and about relationships between variables. Displaying data in this way makes it easier to understand. There are many different kinds of graph or chart, for example line, pie and bar.

Note: You must decide which type of graph/chart is best suited to display your particular data set.

Bar Charts

These are probably the most common chart types and are used for comparing data. A bar chart displays vertical columns, where each column represents one of the values/categories being compared. A type of bar chart is the stacked bar chart and the main difference is that a stacked bar chart shows the amount each category contributes to the total, as in the diagram.

Line Charts

Line charts are most commonly used to display trends in data, and show continuous data over a set time period. This chart displays each value/category as a point, the points are connected by lines.

Pie Charts

Pie charts are often used to show values/categories as a percentage of the total. A pie chart is displayed as a circle which is broken into segments, the size of each segment is determined by its percentage of the total. You can determine the most important "value" within a data set using a pie chart.

XY Scatter Charts

XY scatter charts can be used to plot workbook data showing relationships/comparisons between the data values. An XY scatter chart can be used to display comparisons or relationships between two data values.

To create a chart in IBM SPSS it is easier to use the legacy charts available. To create a chart using the **Legacy** options, go to the **Graphs** menu and choose the **Legacy Dialogs** option. A sub menu will appear showing you the different chart types available, **Figure 16**. Scroll to the chart type you wish and click on it.
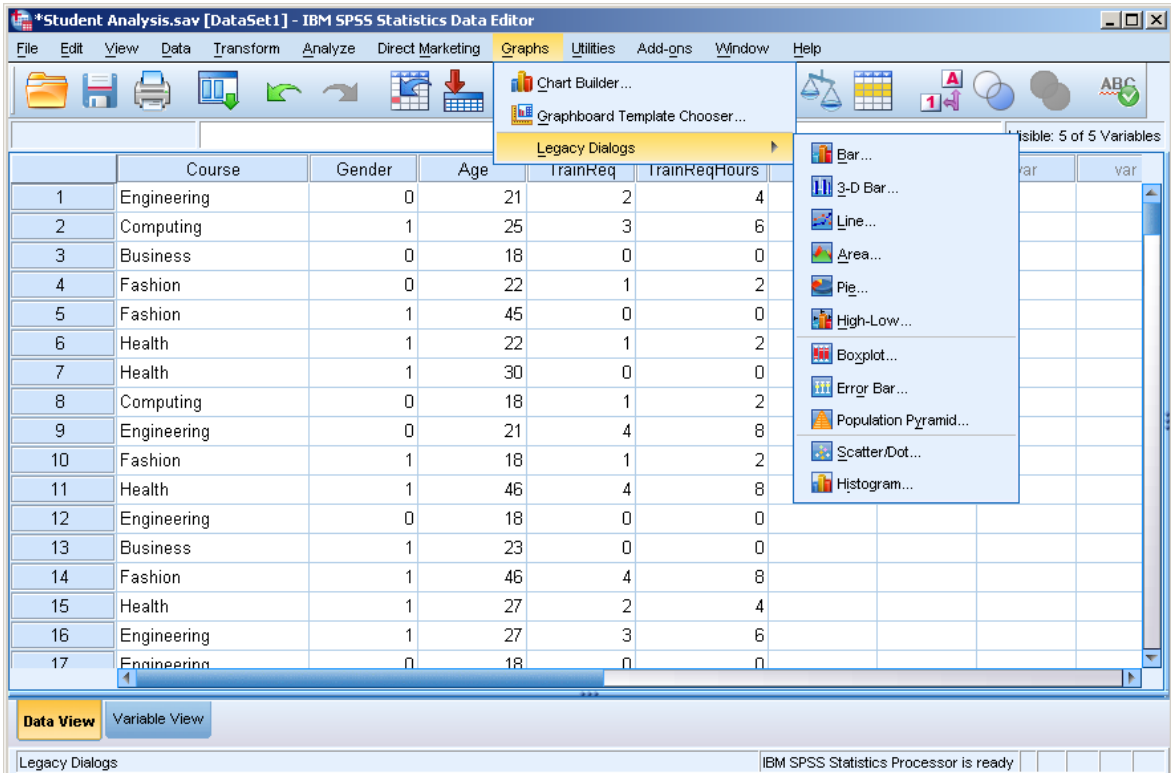
Figure 16: Graphs menu with Legacy Dialogs sub menu

Note: There is also an option within IBM SPSS which enables you to build a chart rather than using one of the legacy chart types. This is known as **Chart Builder**. To create a chart using **Chart Builder** click on the **Graphs** menu, and choose the **Chart Builder** option.

Creating a Bar Chart

If for example you wish to create a bar chart, the **Bar Charts** dialogue box will open giving you further bar chart options. Choose the appropriate type (Simple, Clustered or Stacked) and how you wish the data to be displayed in the chart by clicking on one of the radio buttons in the **Data in Chart Are** section, **Figure 17**.
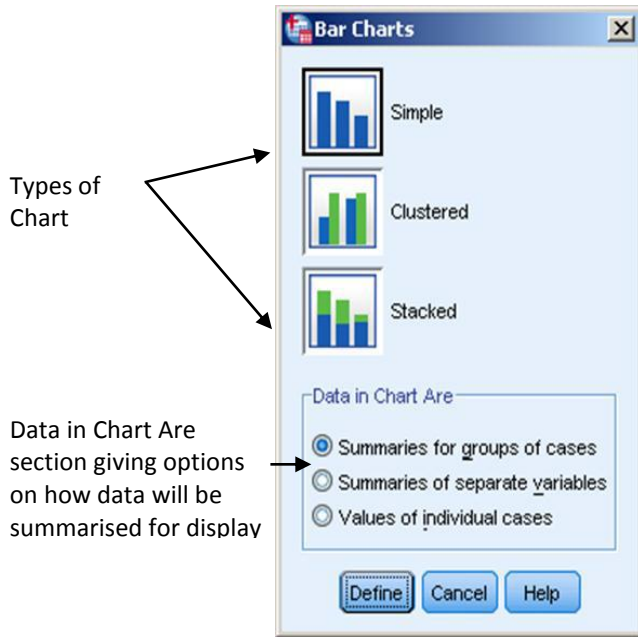
Figure 17: Bar Charts dialogue box

Task 13

From the Graphs menu click on Legacy Dialogues and choose Bar chart.

Next, click on the **Define** button. The **Define** dialogue box will open giving you options to further define the content of the chart you are creating. Depending on the chart type chosen you must select the variables you wish the chart to show.
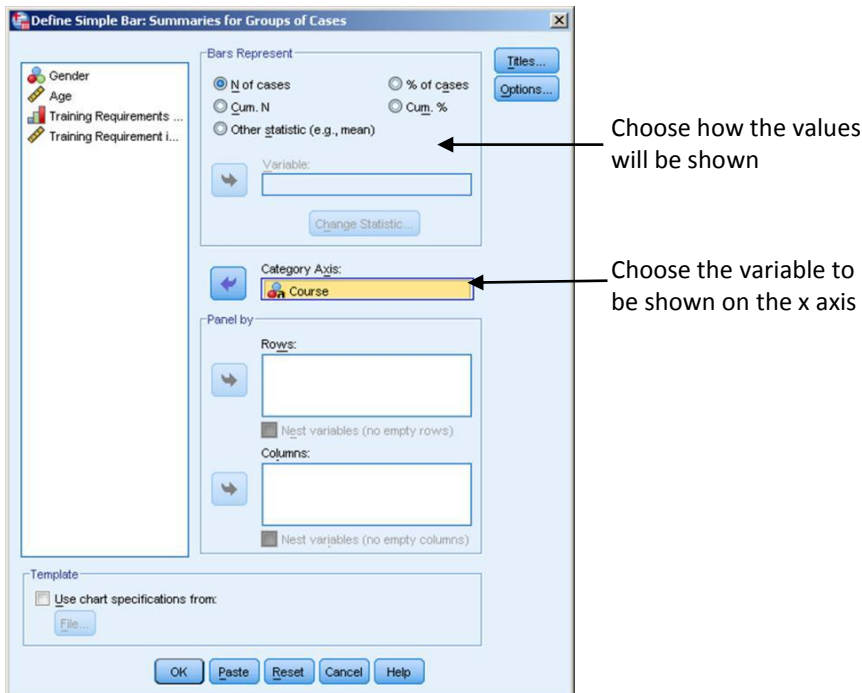

Figure 18: Define Simple Bar dialogue box

Move the variables by clicking on them and then clicking on the arrow [→] button to move them to the named area. Once you have chosen the variables and moved them to the named area click **OK**. The Output window will open and the chart will be displayed.


Task 14

Create a simple bar chart depicting the number of students on each course by percentage.

Creating a Pie Chart

To create a pie chart, from the **Graphs** menu choose the **Legacy dialogs** option and a sub menu of chart types will appear, choose **Pie**. The **Pie Charts** dialogue box will open, **Figure 19** allowing you to choose how the data will be summarised for display in the chart.
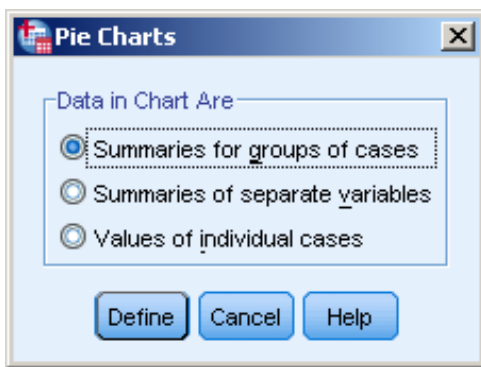
Figure 19: Pie Chart dialogue box

Choose how you would like the data to be displayed by checking one of the radio buttons, then click on the **Define** button, the **Define Pie: Summaries of (how the data will be grouped)** dialogue box will open presenting options, **Figure 20**.
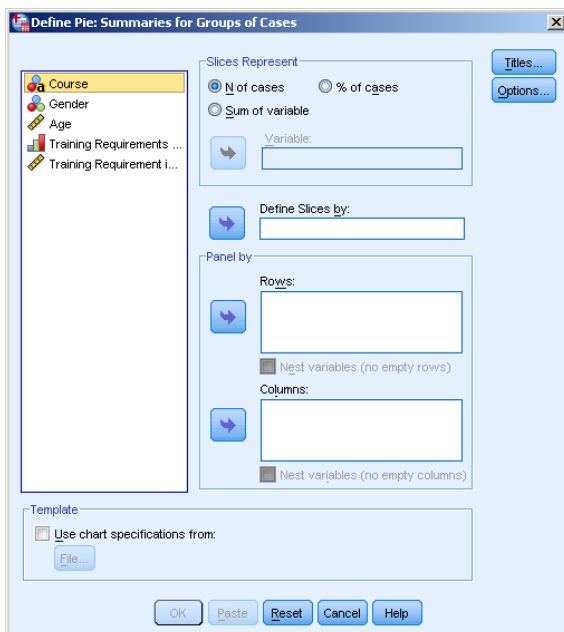
Figure 20: Define Pie: Summaries of … dialogue box

Move the variable(s) you wish to plot to the section required by clicking on it and then clicking on the arrow [→] button beside the section required.

The Output window will then open and display the chart.

Task 15

Create a pie chart displaying gender from the Student Analysis data file.


Editing a Chart

To edit a chart or graph, double click anywhere on the chart/graph area and the **Chart Editor** window will open displaying the chart and the editing toolbars. To modify a specific element of a chart double click on it and the **Properties** dialogue box will open, **Figure 21** showing relevant formatting settings in different tabs.

After you have completed all modifications of the chart simply close the Chart Editor from the close icon at the top right corner of the window and the Chart Editor will close and the modificatons of the chart will be displayed in the Output window.
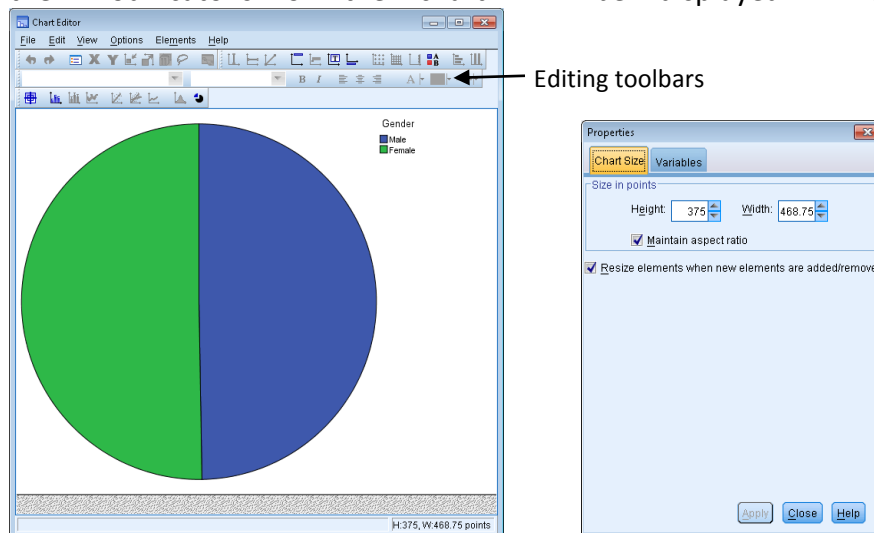
Figure 21: Chart Editor window and Properties dialogue box

Adding a Title to a Chart/Graph

To add a title to a chart, click on the **Insert a title** [icon] command from the toolbars at the top of the **Chart Editor** window. The title section of the chart will appear surrounded by a yellow border. A red flashing cursor will show in the word Title and the **Properties** dialogue box will open displaying the title options with the **Text Style** tab selected. Click in the Title area and enter the title for your chart. Then from the **Properties** dialogue box choose the font type, size and colour. Once you have done this click on the **Apply** button. The changes will be applied and will show in the Title area.
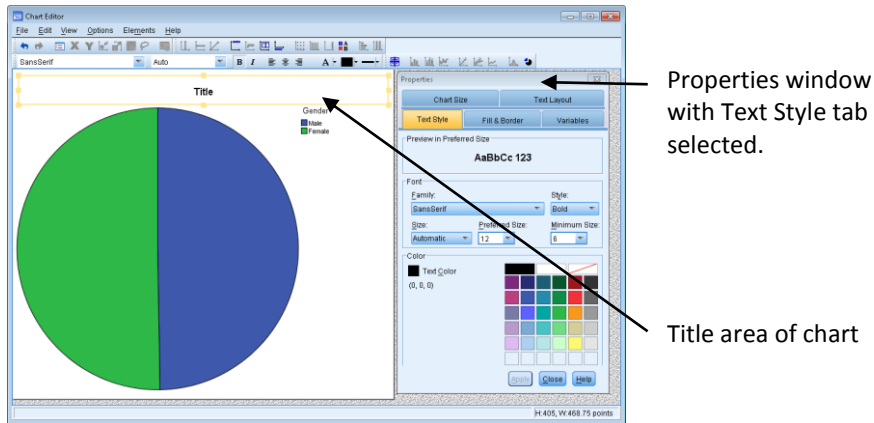
Figure 22: Title area and Properties dialogue box

Note: If the **Text Style** tab is not selected, select it. If the **Font** section of the **Text Style** tab is not active click outside the title area and then back inside.

Task 16

Enter the title **Student Gender Split** on the pie chart created in the previous task. Choose Calibri as the font type, size 28 and make it blue in colour.

Changing Colours within a Chart/Graph

Double click on the chart/graph in the Output window to open the **Chart Editor** window. Double click on the chart item that you wish to edit to open the **Properties** dialogue box. A yellow border will appear around your chosen item and the relevant tab will appear in the **Properties** dialogue box, **Figure 23**. If not selected, choose the **Fill & Border** tab from within this dialogue box to change the colour of the outline and content of the chart/graph shapes. Once you have chosen the colours click on the **Apply** button to apply the changes.
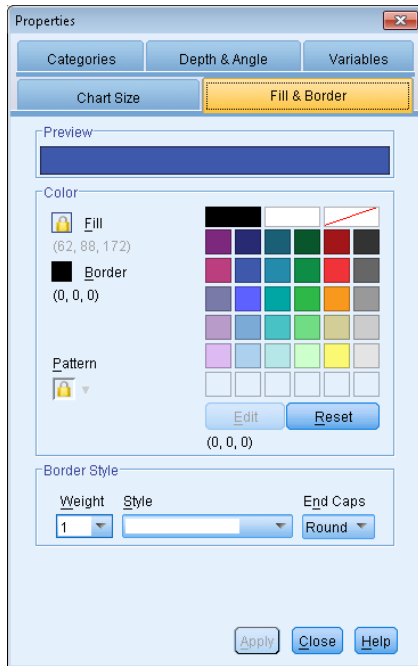
Figure 23: Properties dialogue box showing Fill & Border tab

Note: To colour the different sections of the pie chart you must separate them. To do this click on the **Explode Slice** command from the toolbars in the **Chart Editor** dialogue box then click on each individual section until the yellow border appears around it and then choose the colour from the palette.

Task 17

In the Student Gender Split pie chart re-colour the different sections of the pie chart.

Adding Data Labels

In order to view the data more clearly data labels can be added to the chart. To add **Data Labels** click on the chart within the Chart Editor until it becomes highlighted (it will have a yellow border around it). Then click on the **Add Data Label** command, data labels will be added to the chart and the **Data Value Labels** tab of the **Properties** dialogue box will open. Make any adjustments to the data labels required from the **Properties** dialogue box and click **Apply**.

Task 18

Add data labels to the portions of the pie chart Student Gender Split.

Changing the Labelling of Axes

It is also possible to change the default labels on the axes of your chart/graph. To do this, ensure the Chart Editor window is open and then double click the axis label that you wish

to change. A yellow border will appear around the label. Click inside the border and delete the label, enter the label of your choice.

Task 19

On the bar chart created in Task 14 change the Y axis to **Number of Students**.

Changing the Size of the Chart

You can also change the chart size from the Chart Editor window. To do this, select the **Chart Size** tab from the **Properties** dialogue box. The size in points can be chosen from here. Resize as appropriate and then click on the **Apply** button.

Task 20

Experiment with resizing the Course chart by changing the sizes. Close the Chart Editor window.

Statistical Scenarios

Before doing any statistical test, it is important that you understand the data set and the reasoning behind it. When choosing a test the following must be taken into consideration:

a. What type of data is available
b. What you wish to do with it
c. What questions you have for your data

In statistics a result is **Statistically Significant** if it is improbable to have been the result of chance.

Hypothesis testing is used when decisions are being made using a null hypothesis. The null hypothesis assumes truth until proven otherwise.

Recoding Variables

The data you have may not be suitable for the analysis or graphs/charts you want to do. If this is the case one option is to recode your variables. For example, if you want to show how training requirements change for different age groups but you have only the age, you can recode the age variable into age groups.

To recode a variable, go to the **Transform** menu and choose **Recode into Different Variables…** (it is recommended to use this option instead of Recode into Same Variables which will replace your original variables). The **Recode into Different Variables** dialogue box will open, **Figure 24**. From the list of available variables in your data file on the left select the one you want to recode and click on the arrow 🔁 button. The selected variable will appear in the **Input Variable->Output Variable:** box and the **Output Variable** options to the right will become active. You have to name your new variable and also create a label for it if you wish. Click on the **Change** button to specify that the old variable you have chosen is going to be recoded into the new one you have just named.
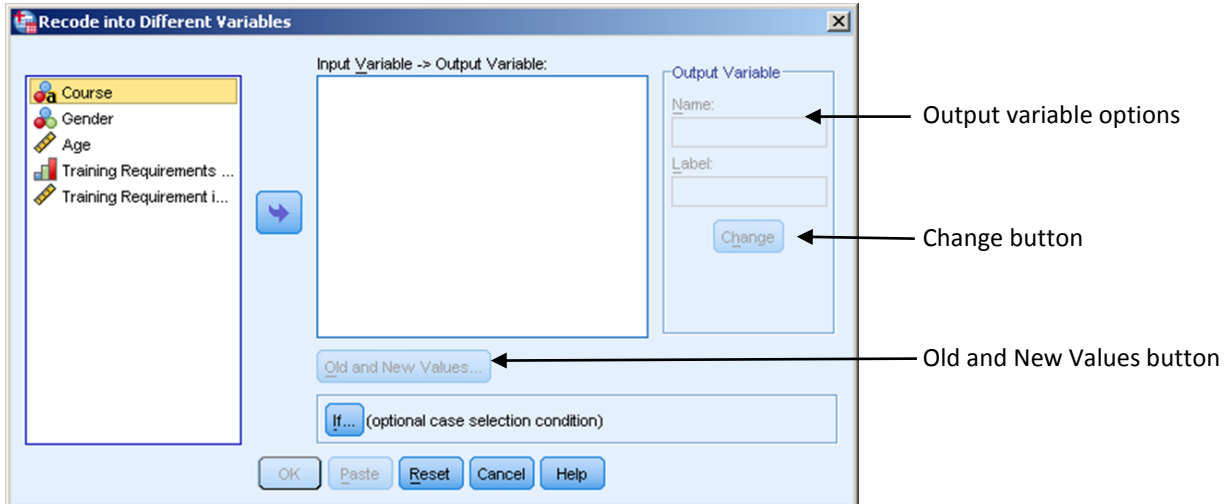
Figure 24: Recode into Different Variables dialogue box

Click on the **Old and New Values…** [Old and New Values] button to specify how you want to recode the old variable's values into the values of the new variable. The **Recode into Different Variables: Old and New Values** dialogue box will open, **Figure 25**. In the **Old Value** section you can specify the specific values or range of values of the original variable and in the **New Value** section you can specify what the new value of these is going to be. By default the new values are real numbers with 2 decimal spaces. If you want the new value to be a string, tick the **Output variables are strings** option at the bottom right hand side of the dialogue box. After you have made your choice for the original variable old values and the new variable new values the **Add** [Add] button will become active and you can click on it to add your choice to the list of **Old->New** values.

Note: When recoding, take care not to put the same original value into two different new values. For example, if you want to recode age into age groups – 20 and below, 21-30 and over 31 you should make sure that age 20 is only recoded into "below 20" and the next age group starts at 21 not 20.
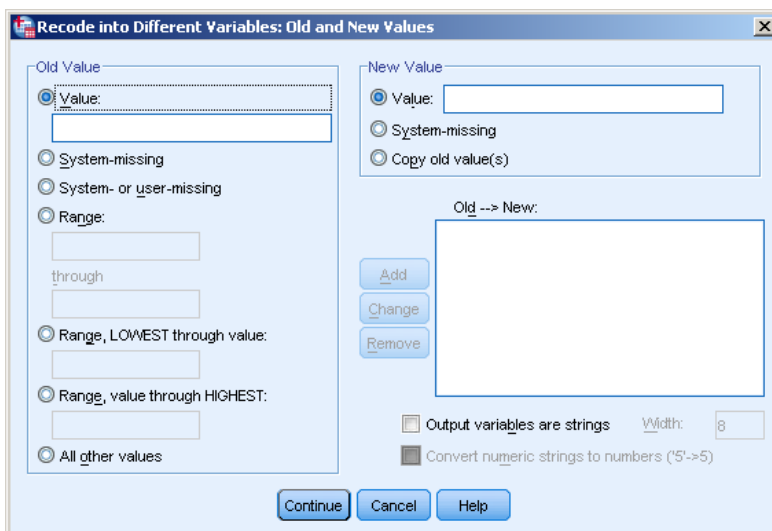

Figure 25: Recode into Different Variables: Old and New Values

After you have made all of your choices and added them click on the **Continue** Continue button. Click on the **OK** button in the **Recode into Different Variables** dialogue box and the new variable will be created and its values filled in according to your recoding instructions. The new variable with the name you have chosen will appear in the data file. Remember to save your data file if you want to keep this new variable.

Task 21

Recode the Age variable into Age Group: 20 and below, 21-25, 26-30, 31-35, 36-40, 41-45, and above 46.

Save your data file.

Frequency Tables

A frequency table displays the number of times (the **frequency** with which) a value appears within a variable set. For example, it may be useful to know the marks a group of students score in a class assessment.

This can be done automatically within **IBM SPSS** using the **Frequencies** command.

To create a frequency table from your data click on the **Analyze** menu. A drop-down menu will be displayed, choose **Descriptive Statistics** and from the sub menu **Frequencies**. The **Frequencies** dialogue box will open with all of the variables from your data set showing in the left hand pane. To choose a variable to be included in the frequency table click on it (it will become highlighted), then click the arrow ⬦ button to transfer it to the right hand side Variable(s) pane, **Figure 26**. Then click on the **OK** button.

Figure 26: Frequencies dialogue box

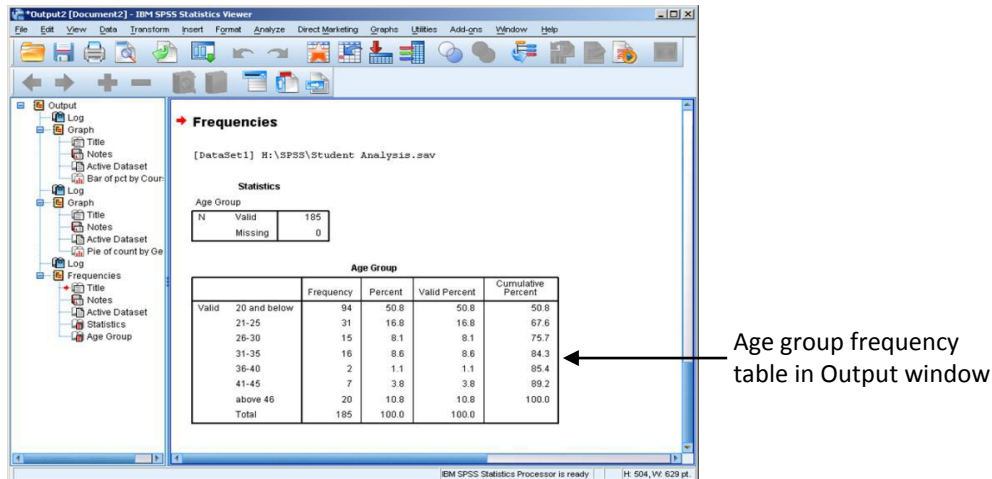The Output window will open displaying the output from your analysis,

Age group frequency table in Output window

Figure 27: Output IBM SPSS Statistics Viewer window

 Task 22

Create a frequency table from the Age variable.

This output data (analysis results) can be manipulated dependant on what you wish to do with it. You may wish to copy and paste it in another document (as explained on page 10), perhaps a report in Microsoft Word.

 Task 23

Copy the Age table from the Output window, open the Additional Data.docx file and copy this to the bottom of the file. Save and close the file.

Frequency Statistics

To get statistics on variable(s) chosen, from the **Analyze** menu choose **Descriptive Statistics** and then **Frequencies** from the sub menu. The **Frequencies** dialogue box will open, **Figure 28**. You can for example get the mean or standard deviation of a variable set from here.
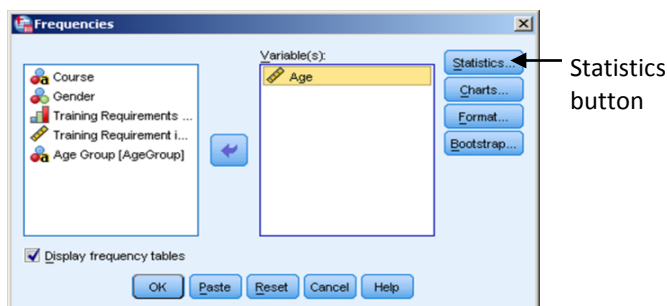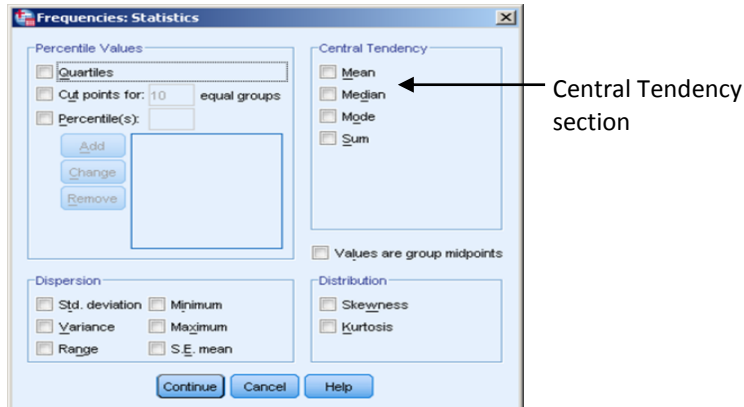

Statistics button

Figure 28: Frequencies dialogue box

To get the **Mean** (arithmetic average), choose an appropriate variable and add it to the Variable(s) pane then click on the **Statistics** button on the right. The **Frequencies:**

Statistics dialogue box opens with options. Click on the **Mean** by checking the box beside it in the **Central Tendency** section, **Figure 29**.

Central Tendency section

Figure 29: Frequencies: Statistics dialogue box

Then click on the **Continue** button which will take you back to the Frequencies dialogue box. Click the **OK** button. The Output window will now display the Mean for the ages in table format.

Mean Statistics

Figure 30: Mean showing in Output window

There are other statistics which can be processed from this area by checking the appropriate boxes to get the statistics you wish on a particular variable set. The results will be displayed again in the Output window in table format.

When you close the Output window you will be asked if you wish to save the output. If you wish to save the output you must save the file with an appropriate name.

Note: Saving the output file will not save the data file. Remember to save your data file as well if you have made changes to it.

Task 24

Show the Mean of the Age variable. Save the Output as Age Frequency and Mean.

Cross tabulation

Cross tabulation (known as Crosstabs) is used to generate a data table containing a side by side comparison of categorical variables which will help to identify if a relationship exists between them. Cross tabulation, however, does not highlight the causal relationship between variables. For example an academic institution may wish to know if there is a relationship between gender and course of study chosen. They may have a hypothesis that all courses are equally appealing to both men and women. They may

perform a crosstab to discover if this is true, but the reason behind course choice will not be shown.

To perform a cross tabulation, click on the **Analyze** menu and choose **Descriptive Statistics** and then **Crosstabs**. The **Crosstabs** dialogue box will open, **Figure 31**.
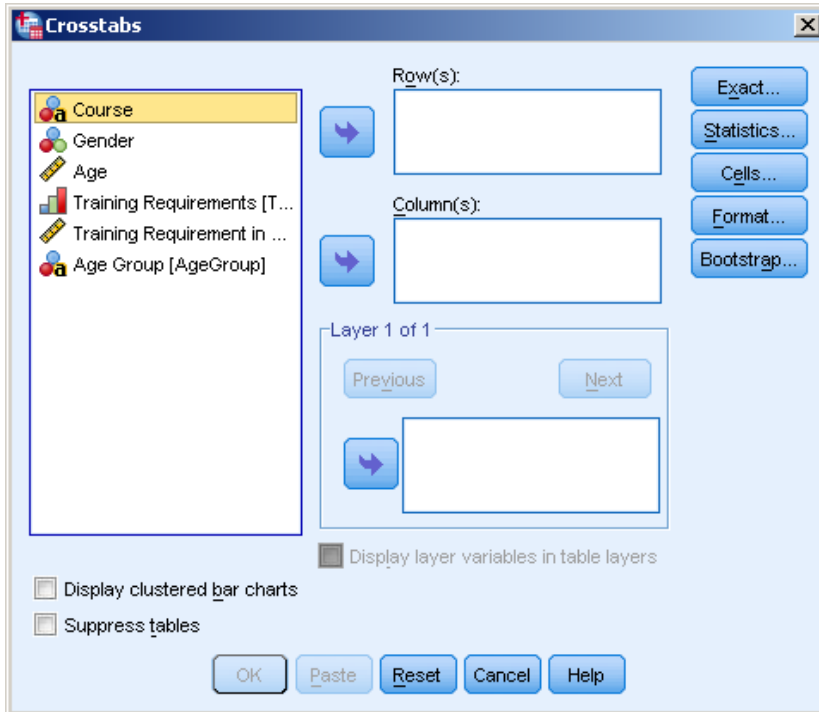
Figure 31: Crosstabs dialogue box

To create a crosstab to examine the above hypothesis that all courses are equally appealing to men and women, select the **Course** variable. Then move it to the **Row(s)** field using the arrow ➡ button. Select the **Gender** variable next by clicking on it and again using the arrow ➡ button move it to the **Column(s)** section. Click the **OK** button. The Output window will open displaying a table showing the results of the cross tabulation that you have performed, **Figure 32**.

Course * Gender Cross tabulation

Count

		Gender		Total
		Male	Female	
Course	Business	19	26	45
	Computing	24	11	35
	Engineering	36	9	45
	Fashion	9	24	33
	Health	4	23	27
Total		92	93	195

Figure 32: Gender * Course cross tabulation

From this table we can see that women seem to be predominant on certain courses and men on others but no reasons for these choices are shown.

Task 25

Perform a cross tabulation to determine if a connection exists between course and training requirement.

Chi-Square Test

Note: The variables used must be nominal or ordinal.

The Chi-Square Test is used to determine if a relationship exists between categorical variables within a dataset. There are a few different Chi-Square tests such as **Goodness of Fit** test which compares theoretical distribution with the observed sample from data. Another is the Chi-Square **Test of Association** which compares two variables in a sample of data to determine if a relationship exists between them.

In the previous example we performed a crosstab to try to identify if a relationship between gender and course choice existed. The resulting table suggests that there might be a relationship. However, this should be further tested statistically to fully confirm the results.

Chi-Square Test Setup

The hypothesis being tested is:

- Null hypothesis (H_0) assumes that there is no relationship between the variables; that is, that there is no relationship between gender and course choice.
- Experimental hypothesis (H_1) assumes there is a relationship between the variables; that is, that there is a relationship between gender and course choice.

To perform the Chi-square test, click on the **Analyze** menu, **Descriptive Statistics** and then **Crosstabs**. The **Crosstabs** dialogue box will open, **Figure 33**. Move the **Gender** variable by selecting it and then clicking on the arrow button to move it to the **Row(s)** pane on the right hand side. Click on **Course** and move it to the **Column(s)** pane. Click on the **Statistics** button at the right hand side of the dialogue box and the **Crosstabs: Statistics** dialogue box will open,   **Figure 34**. Check the **Chi-square** option and then click **Continue.** You will be taken back to the **Crosstabs** dialogue box. Click on the **OK** button to perform the chosen test. The Output window will open and display the results of your test.
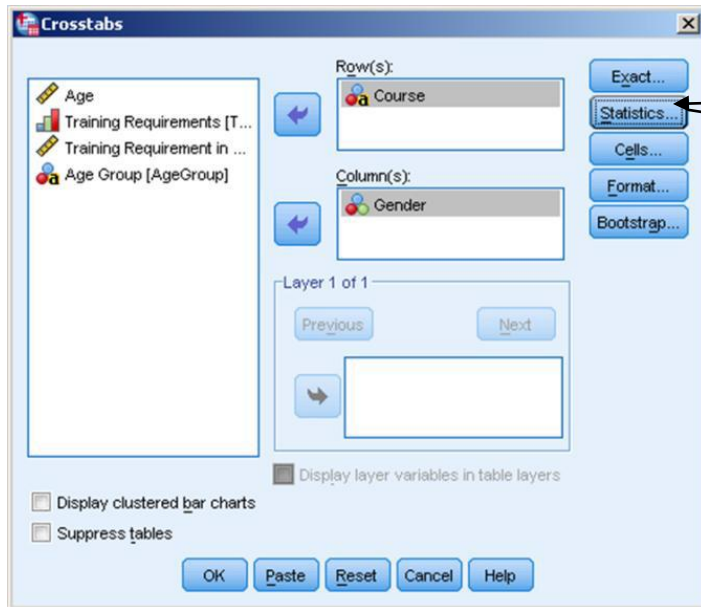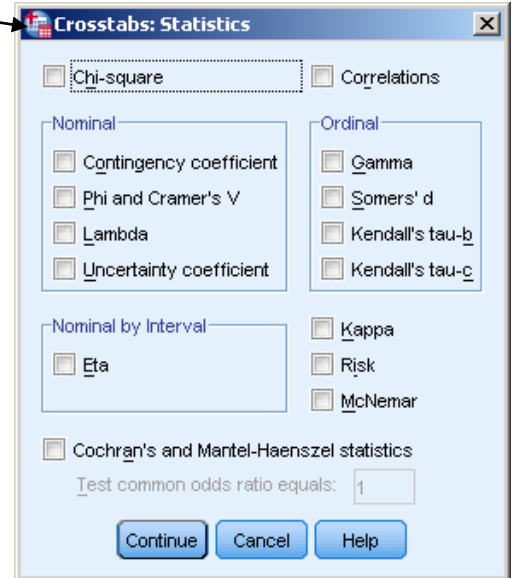
Figure 33: Crosstabs dialogue box   **Figure 34:** Crosstabs: Statistics dialogue box

Chi-Square Test Results from Output window

In statistics a significant result is one which is unlikely to have occurred by chance. The probability of this happening by chance is defined by the p-value. If a test of significance gives a p-value lower than 5% (0.05) then this result is said to be Statistically Significant at the 5% level, and the null hypothesis rejected.

 Note: Tests can also be done at 0.1%; 1% and 10% levels but justification is needed.

A table will appear in the Output window displaying the results of the cross tabulation, **Figure 35**.

Gender * Course Cross tabulation

Count

		Course					Total
		Business	Computing	Engineering	Fashion	Health	
Gender	Male	19	24	36	9	4	92
	Female	26	11	9	24	23	93
Total		45	35	45	33	27	195

Figure 35: Gender * Course cross tabulation

The Chi –Square Test results are also shown in the Output window in a table, **Figure 36**.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	42.302[a]	4	.000
Likelihood Ratio	45.234	4	.000
N of Valid Cases	195		

a. 0 cells (.0%) have expected count less than 5. The minimum
expected count is 13.43.

Figure 36: Chi-Square Tests

The Chi-Square Tests table show a result of Chi-square p=0.000, i.e. p<0.001 (<5% level of significance) for choice of course by gender, indicating that there is a significant relationship between course choice and gender, therefore there is evidence to reject the null hypothesis.


Task 26

Perform a Chi-Square test to test if a relationship exists between the course and training requirement where H_0 is that no relationship exists between course and training requirement

---

T-Test

The t-test is a hypothesis test, the most common version being the two sample t-test. This tests whether or not two independent samples have different mean (average) values.

For example, we may think that the average age of the students (men and women) is the same (the null hypothesis). The t-test allows us to test this by determining a probability (p-value) indicating how likely these results could have been returned by chance. Normally, if there is a less than 5% chance of these results being returned by chance, we would reject the null hypothesis stating that there is a Statistically Significant difference between the two genders.

To conduct a t-test click on the **Analyze** menu, **Compare Means** and choose the **Independent Sample T-Test** option. The **Independent Samples T Test** dialogue box opens, **Figure 37**.
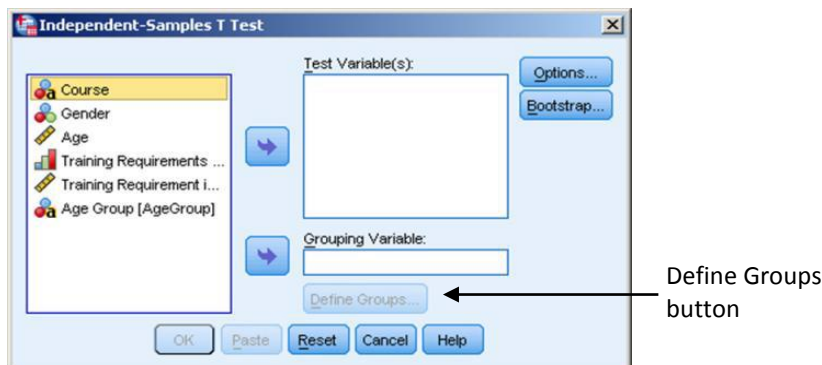
Figure 37: Independent Samples T Test dialogue box

As the test variable is Age, select **Age** by clicking on it and move it to the **Test Variable(s)** field by clicking on the arrow 🡆 button. Then add the **Gender** variable into the **Grouping Variable** field by clicking on the arrow 🡆 button. The **Define Groups** button will become active, click on it. The **Define Groups** dialogue box will open, **Figure 38**. Define the groups for the variable chosen, in this case group 1 would be Male so a 0 would be entered in the field as we have coded Males with the number 0. Female would be 1. Click on the **Continue** button you will be taken back to the **Independent Samples T Test** dialogue box.
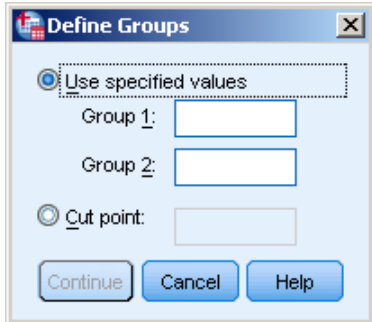
Figure 38: Define Groups dialogue box

Click **OK** and the results will appear in the Output window, **Figure 39** and **Figure 40**.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Age	Male	92	26.65	12.262	1.278
	Female	93	24.92	9.357	.970

Figure 39: Group Statistics table

The average (mean) age does not look to be significantly different, almost 2 years for these respondents.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Age	Equal variances assumed	10.042	.002	1.078	193	.283	1.727	1.603	-1.435	4.889
	Equal variances not assumed			1.076	170.176	.283	1.727	1.605	-1.441	4.896

Figure 40: Independent Samples t-test results table (grey cell shows the p-value)

From this we can see that p=0.283 and therefore p>0.05. As this is a non-significant difference, we can accept the Null Hypothesis stating that there is no significant difference in the average ages of men and women students.

Note: Levene's test for Equality of Variance is also shown in the results table in the Output window. Its p-value is 0.002 and therefore <0.05, therefore we must read the lower line for t-test result.

Task 27

Do a t-test to check the mean training hours of male and female respondents where it is assumed that both are the same.

Correlation

Correlation is the statistical relationship which exists between two variables. If there is a significant correlation between two variables a change in one will automatically produce a change in the other. There are many statistical ways to measure correlation, but you must know what type of correlation in your data you want to prove and choose the appropriate test. One example - **Pearsons** is outlined below.

Pearson's Coefficient Correlation

This is also known as the **Pearson Product Moment Correlation Coefficient (r)** and measures the strength of the linear relationship between two variables. This should be used for continuous variables of similar scale.

- A correlation of –1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down in a linear way.
- A zero 0 correlation indicates that there is no linear relationship between the variables.
- A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together in a linear way.

Prior to carrying out any correlation within **IBM SPSS** it is advisable to plot the data first. A scatter plot can be used for this. This will give a visual indication of the relationship between the variables. If the scatter plot distribution is in an elliptical pattern then it could be said that there is a strong linear relationship between the variables. The narrower the ellipse in shape, the stronger the relationship. If the points are distributed along a line a perfect line relationship is said to exist. If the points are more spread out in a circular pattern it could be assumed that there is a weak or no relationship between the variables.

For example, the null hypothesis can be: all students regardless of their training requirement will need the same number of hours of ICT training. A Pearson's correlation can be used to test this.

Firstly plot the data using a scatter plot. To do this from the **Graph** menu choose **Legacy Dialogs** and the **Scatter/Dot** option from the sub menu. The **Scatter/Dot** dialogue box

opens with chart sub types displayed. Choose a type (in most cases Simple Scatter will be sufficient) and then click on the **Define** button, **Figure 41**.
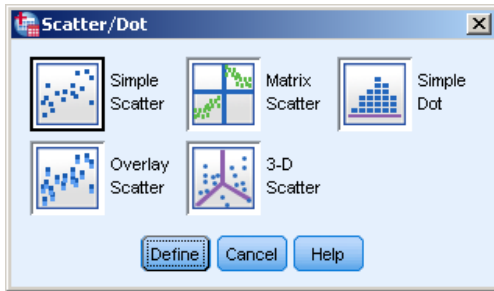
Figure 41: Scatter/Dot dialogue box

The **Scatter/Dot** dialogue box opens, move the **Training Requirement** variable to the Y axis field and the **Training Requirement in Hours** to the X axis field and then click the **OK** button. The Output window opens displaying the chart.
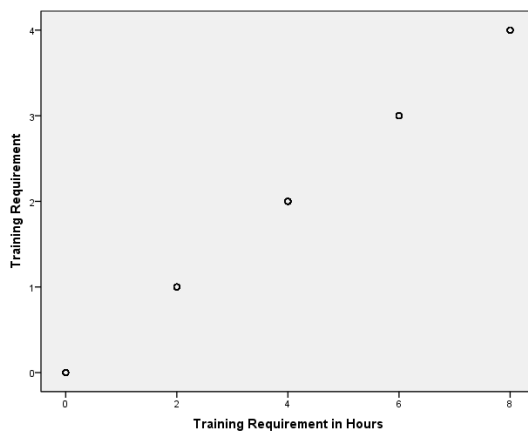
Figure 42: Scatter plot graph

From this we can see that the points are distributed along a line. This would indicate that a perfect relationship exists between the Training requirement and the hours of ICT training required.

 Note: This perfect line would be very unlikely in practice.

To carry out a **Pearson's** correlation on this same data, from the **Analyze** menu, choose **Correlate**. A sub menu appears, choose **Bivariate** and the **Bivariate Correlations** dialogue box will open. Click on the variables you want to test in the left section and then click the arrow button to move them to the right hand **Variables** section.
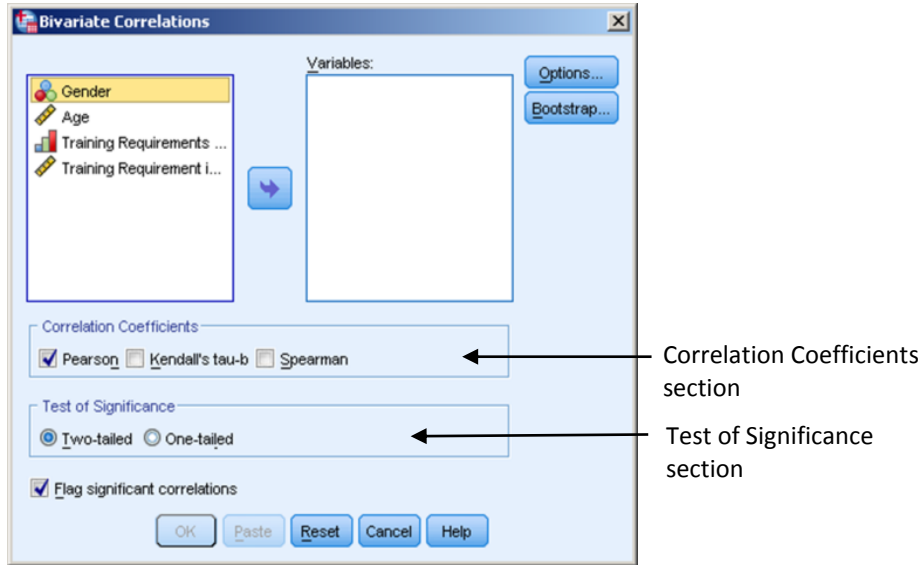
Figure 43: Bivariate Correlations dialogue box

By default **Pearson's** and the two-tailed test option will be selected in the **Correlation Coefficient** section of the dialogue box. The results will be displayed in the Output window, **Figure 44**.

Correlations

		Training Requirement	Training Requirement in Hours
Training Requirement	Pearson Correlation	1	1.000**
	Sig. (2-tailed)		.000
	N	195	195
Training Requirement in Hours	Pearson Correlation	1.000**	1
	Sig. (2-tailed)	.000	
	N	195	195

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 44: Correlations

It can be concluded that the results are very significant, the probability of this happening is p=0.000, i.e. p<0.001, and correlation is significant at 0.01 level where the Correlation =1. There is a strong correlation between training requirement and the number of hours of training required. Therefore the Null hypothesis would be rejected.

Note: The significance of the correlation depends on the sample size. For example 0.5 would not be significant for a small group, but could be very significant for a larger group.

Task 28

Create a scatter plot graph to see if a relationship exists between age and training requirement.

Then perform a Pearson's correlation to statistically evaluate the findings.

Data

Qualitative (assumed discrete)

Nominal

Ordinal

Quantitative (some count data) (assumed continuous)

Interval

Ratio

Is your data Normally Distributed?

No

Yes

Appropriate Descriptive Statistics

- Nominal: Mode; % frequencies
- Ordinal: Mode; Median; % frequencies
- Interval: Mean; Median; Mode; Standard deviation
- Ratio: Mean; Median; Mode; Standard deviation

Appropriate Graphs

- Nominal: Pie charts; Bar charts
- Ordinal: Pie charts; Bar charts
- Interval: Histograms; Polygons; Boxplots
- Ratio: Histograms; Polygons; Boxplots

Appropriate Inferential Statistics

- Nominal: Test of proportion(s); Chi-square test for Independence
- Ordinal: Test of proportion(s); Chi-square test for Independence; Nonparametric tests such as: Wilcoxon & Mann-Whitney
- No: Nonparametric tests such as: Wilcoxon & Mann-Whitney
- Yes: Parametric test such as t-test and ANOVA