

Creating and Handling a Data File on SPSS

Index

	<u>Page</u>
<u>Creating and Handling a Data File on SPSS</u>	
1) Loading SPSS for Windows	2
2) Defining Variables	3
3) Entering the Data	7
4) Saving, Opening and Printing a Data File	8
5) Inserting or Deleting Variables	9
6) Searching for Data Items and Selecting Cases	9
 <u>Descriptive Statistics with SPSS</u>	
9) Frequency Tables and Percentages	11
10) Charts	12
11) Cross-tabulations	18
12) Histograms	19
13) Averages and Measures of Spread	21
14) Comparison of Means	22
15) Reporting a Statistical Output	24

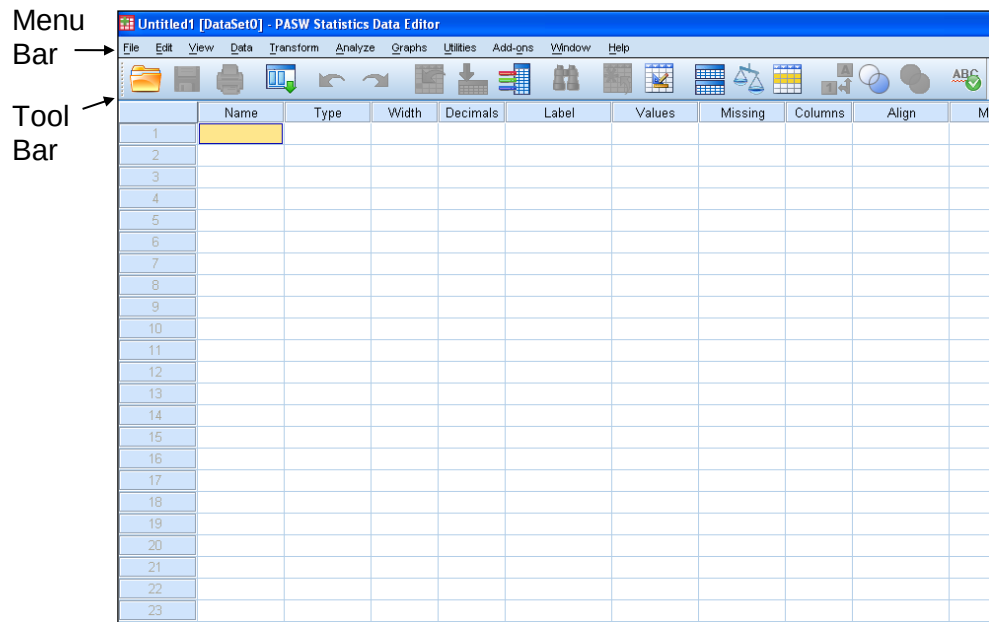
Creating a Data File

1. Loading SPSS

Load SPSS from the Start menu

Click on **Type in data** then **OK**.

The following screen display should be visible:

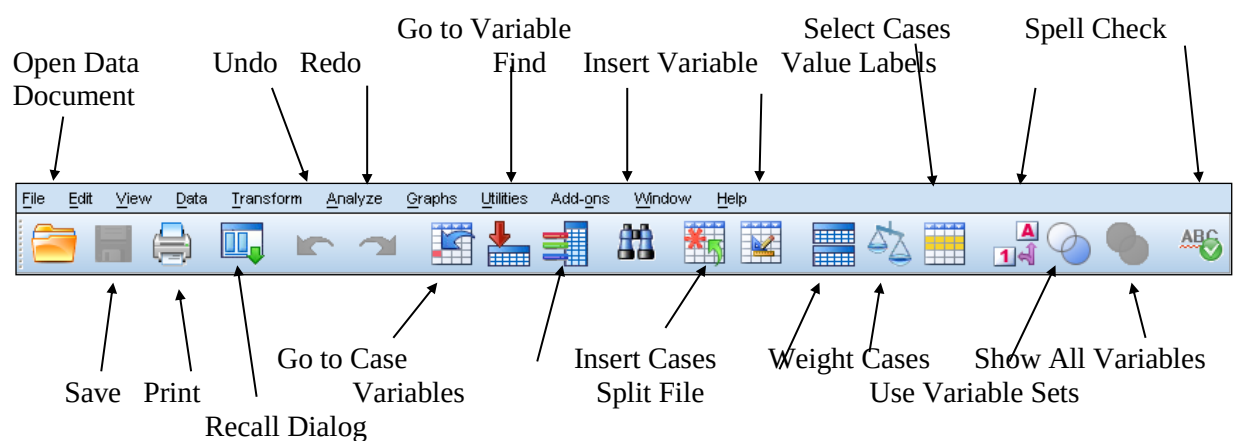


Menu bar

This is a drop-down menu system.

Tool Bar

These are short cuts for using the menu bar.



The purpose of each Tool icon is displayed when the mouse arrow is held for a few seconds above the icon.

2) Defining Variables

Download the data file “**week1_data.csv**” from the moodle. It contains data for a sample of 30 employees, 15 male and 15 female, of a large organisation, between the ages of 20 and 50. Open the data file to view the variables and cases (observations).

Refno	Gender	Jobhappy	Agegroup	Age	Salary	Prodbsal
002	1	2	1	20	18.8	19.4
004	1	1	3	48	27.2	26.4
006	1	3	3	45	30.8	30.2
011	2	3	2	36	27.2	27.0
016	2	1	1	25	19.8	20.2
019	1	1	2	39	22.3	22.7
022	1	1	2	39	19.3	19.4
023	2	3	3	42	25.3	25.6
026	1	2	1	27	17.1	17.6
035	2	2	3	41	32.3	33.4
037	2	3	3	47	21.3	21.4
044	2	2	1	28	21.7	22.0
045	1	2	2	37	22.1	22.1
050	2	2	2	34	19.2	19.2
055	1	1	1	23	17.3	18.1
057	2	3	2	30	20.9	20.9
061	2	3	2	34	19.4	19.4
076	2	3	1	22	14.9	15.6
082	1	1	2	31	23.3	23.6
088	2	3	2	38	16.7	17.2
090	1	1	2	32	22.3	22.4
097	1	2	2	34	22.2	21.8
099	2	2	3	40	25.1	25.9
101	1	3	1	21	24.9	25.2
103	2	3	1	21	19.4	19.8
132	1	1	3	44	21.2	21.1
148	1	2	3	41	29.4	30.3
164	1	3	1	29	14.9	15.1
171	2	3	3	47	27.2	28.1
172	2	3	1	22	21.7	22.6

Refno is a 3-digit reference code, unique for each employee.

Gender is categorical Nominal variables i.e. male or female are names of unordered categories and are coded: gender: 1 = male, 2 = female

Agegroup and **Jobhappy** are categorical Ordinal variables i.e. its values are categories which can be ordered: agegroup: 1 = 20 - 29 years old, 2 = 30 - 39 years old, 3 = 40 - 50 years old.

Jobhappy: 1 = no, 2 = not sure, 3 = yes

Age, **Salary** and **Prodbsal** are quantitative Scale variables i.e. their values are measurements on a proper scale. Salary is the actual salary in £1000's whereas

Prodbsal is the proposed salary in £1000's under a new productivity-based incentive scheme.

Entering Variable Descriptions

Variables are the column headings.

Each variable must be given a name, the format of the data described e.g. whole numbers or text, and any codes explained e.g. 1= 20-29 years old.

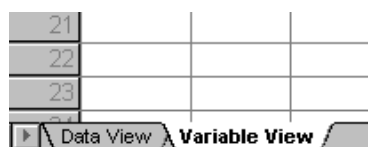
The case numbers 1, 2, 3, 4, 5, etc. are automatically inserted by SPSS.

a) Define Refno Variable

Cursor should be at the top of column 1 as shown:

	var	var	var
1			
2			
3			

Click on **Variable View** in the bottom left hand corner of the screen:



The following column headings now appear:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1										

Type **Refno** for Name then press the **Enter** key.

(Other columns will automatically be filled in but some will need to be changed.)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	refno	Num...	8	2		None	None	8	Right	Scale
2										

Under Type, click on then on  **String** (i.e. non-digits).

Note: Refno is defined a string variable so that the 0's of such values as 002 are displayed. Otherwise 002 would be displayed as 2.

Type **3** for Characters to accommodate the three digits of the reference number, then click on **OK**. This value appears under Width.

Under Label, type **Reference number** (shows the meaning of 'refno')

Under Columns, enter **5** for to accommodate the variable name refno. Note that the default column width is set at 8 and it is not essential to change this value to 5 except to make the data file smaller ie. Fit more on the screen.

The other columns should be correct.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	refno	String	3	0	Reference number	None	None	5	Left	Nominal

b) Define Gender Variable

Move to row 2 under Name using arrow keys or mouse.

Under Name, type **Gender**.

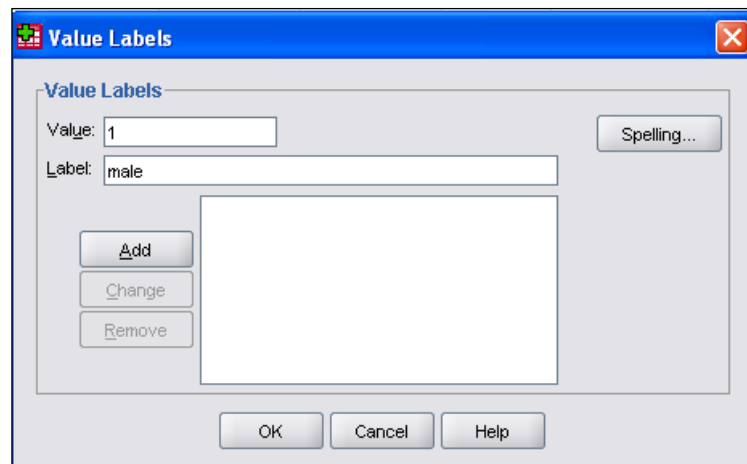
Under Type, select **Numeric**.

Type **1** for Width and **0** for decimal places, then click on **OK**.

There is no need to enter a Label.

Click in the **Values** column then on

For Value type **1**, click in the Label box,  type **male**, click on **Add**.



In the Value box, type **2**, in the Label box, type **female**, click on **Add**.

Click on **OK**.

Select **Left** under Align and select **Nominal** under Measure.

(Text eg. male/female is usually aligned to the left of a column, whereas scale data eg. 30.4 is usually aligned to the right of a column)

c) Define Jobhappy variable

Move to row 3 under Name and type **Jobhappy**.

Under Type, select **Numeric**.

Type **1** for Width and 0 for decimal places, then click on **OK**.

Under label, type **Are you happy in your job?**

Click in the **Values** column and then  on

In the Value box, type **1**, in the Label box, type **no**, click on **Add**.

In the Value box, type **2**, in the Label box, type **not sure**, click on **Add**.

In the Value box, type **3**, in the Label box, type **yes**, click on **Add**.

Click on **OK**.

Select **Left** under Align and select **Ordinal** under Measure.

d) Define Agegroup Variable

Move to row 4 under Name and type **Agegroup**.

Under Type, select **Numeric**.

Type **1** for Width and **0** for Decimal Places, and click on **OK**.

Click on **Values** and then on 

In the Value (Box), type **1**, click on Label (Box), type **20-29**, click on **Add**.

In the Value (Box), type **2**, click on Label (Box), type **30-39**, click on **Add**.

In the Value (Box), type **3**, click on Label (Box), type **40-50**, click on **Add**.

Select **Left** under Align and select **Ordinal** under Measure.

e) Define Age Variable

Move to row 5 under Name and type **Age**.

Under Type, select **Numeric**.

Type **2** for Width and **0** for Decimal Places, and click on **OK**.

The other columns should be correct ie. **Right** under Align and **Scale** under Measure.

f) Define Salary Variable

Move to row 6 under Name and type **Salary**.

Under Type, select **Numeric**.

Type **3** for Width and **1** for Decimal Places, and click on **OK**.

Under Label, type **Present salary (£000)**.

Under Columns, enter **6** to accommodate the variable name salary.

The other columns should be correct ie. **Right** under Align and **Scale** under Measure.

g) Define Prodbsal Variable

Move to row 7 under Name and type **Prodbsal**.

Under Type, select **Numeric**.

Type **3** for Width and **1** for Decimal Places, and click on **OK**.

Under Label, type **Productivity-based salary (£000)**.

Select **Right** under Align and **Scale** under Measure.

The other columns should be correct ie. **Right** under Align and **Scale** under Measure.

The completed Variable View table should appear as:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Refno	String	3	0	Reference number	None	None	5	Left	Nominal
2	Gender	Numeric	1	0		{1, male}	None	8	Left	Nominal
3	Jobhappy	Numeric	1	0	Are you happy in your job?	{1, No}...	None	8	Left	Ordinal
4	Agegroup	Numeric	1	0		{1, 20-29}	None	8	Left	Ordinal
5	Age	Numeric	2	0	Age	None	None	8	Right	Scale
6	Salary	Numeric	3	1	Present salary (£000)	None	None	8	Right	Scale
7	Prodsal	Numeric	3	1	Productivity-based salary (£000)	None	None	8	Right	Scale

3) Entering the Data

Click on **Data View**

a) Enter Data for Case 1

i.e. 002 1 2 1 20 18.8 19.4

Use arrow keys to move to the top of the refno column.

Type **002**, then use right arrow key to move across to the gender column.

Type **1** then move across to the jobhappy column etc.

Note: to correct errors, click on the cell and re-enter the value for that cell position.

b) Validation tests

After defining variables, SPSS will automatically detect some types of invalid data

e.g. i) **More characters than variable length**

If the data item for a variable has more characters than the stipulated variable length then only the stipulated number of characters will be accepted.

e.g. for case 2, try typing **1234** under refno, or **woman** under gender

ii) **Wrong type of data**

If a non-digit is entered for a **numeric** variable then a bleep will sound and the data rejected.

c) Load the Complete File (OR Enter Data for Cases 2 to 30)

Go to moodle and download the file “Week1_data.sav” and save the file in your desktop. In SPSS click **File, Open, Data**.

Alternatively, the data for cases 2 to 30 can be entered manually using arrow keys to move the cursor to correct cell positions and repeating the procedure as for case 1 using data from the sample data listing.

d) Display Value Labels

Value Labels for agegroup can be displayed as data on the screen instead of the codes 1, 2, 3 etc.

Click on the **Display value labels** icon  (or **Utilities, Value Labels**) to give:

<u>Case</u>	<u>Refno</u>	<u>Gender</u>	<u>Jobhappy</u>	<u>Agegroup</u>	<u>Age</u>
1	002	male	not sure	20 - 29	20
2	004	male	no	40 - 50	48
3	006	male	yes	40 - 50	45
4	011	female	yes	30 - 39	36
etc.					

Click on the Display value labels icon  to return to the codes.

4) Saving, Opening and Printing a Data File

a) Saving a file

If you wish to save the file or any other file in your own directory click on **File** and **Save As**

b) Opening a file

To open this file or any other you will need to use a similar procedure to that for saving as shown above. Click on **File, Open**.

Now click on the filename and on **Open**.

c) Printing a file

Click on **File** and **Print**.

Make sure that the selected printer is the one which you require then **OK**.

5) Inserting or Deleting Variables

To insert a new variable click in the column where the new variable is to be inserted

then click on the **Insert variable** icon  (or **Edit, Insert Variable**).

(The variable name and format can now be entered by clicking on the **Variable View** tab at the bottom of the screen.)

To delete a variable just click on the variable name then on **Edit** and **Cut**.

Now try i) inserting a new variable (blank column) at any position
 ii) then deleting it.

6) Searching for Data Items and Selecting Cases

a) Searching for a particular data item

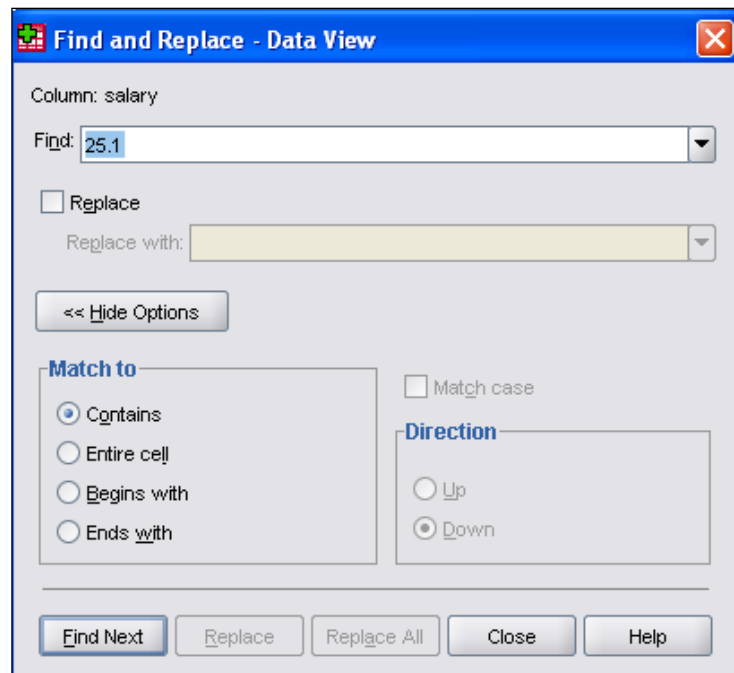
To search the file for a particular data item click on the variable column containing the required data item, then click on the **Find** icon (or **Edit, Find**).



To find a case with salary 25.1 click in the salary column then on Type **25.1** in the Find box.



The search window should appear as:



Click on **Find Next**. The search should stop at the case with salary 25.1. Now click on **Close** to view the case. If the data cannot be found, or the search is in the wrong direction then the message “Not Found” is displayed.

If a data item occurs more than once then you need to click on Find Next again after the first data item is found.

Now try searching for

- i) a 25 year old employee
- ii) any 21 year olds.

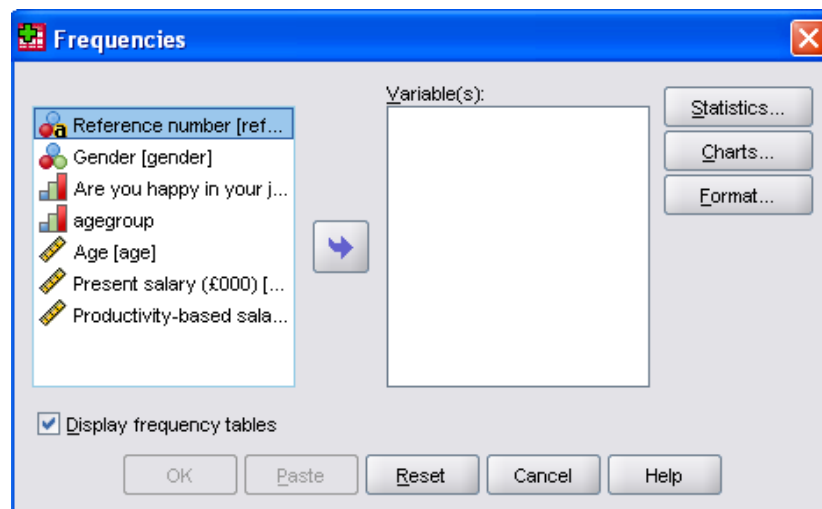
Descriptive Statistics with SPSS

We will now consider a selection of ways of analysing the data and displaying the results. This first section does not involve any significance testing.

7) Frequency table and Percentages

a) Obtain a frequency table for the agegroup variable.

Click on **Analyse**, **Descriptive Statistics** and **Frequencies** to give:



Click on **Agegroup** and the right-arrow .

Click on **Charts**, **Bar charts** and **Continue**, **OK**.

(Obtaining a bar-chart is normally optional)

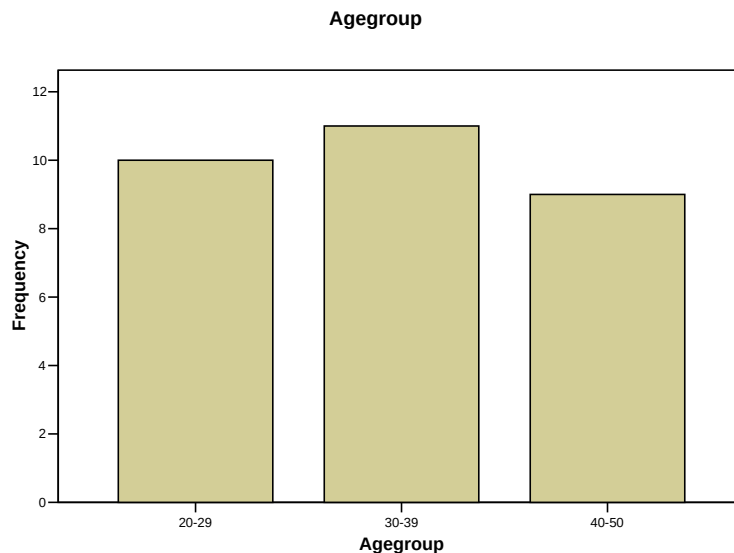
The following output is displayed in the output window:

- Note:
- i) The output shows value labels 20-29 etc. (if any).
 - ii) The valid percentage is the same as the percentage in this case.
However if there was a blank cell i.e. a value missing in the data, the valid percentages would be calculated out of 29 not 30.

- iii) The cumulative percentages show the percentage up to and including that value e.g. there are 70% up to 39 years old.

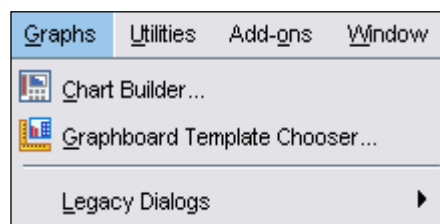
b) Bar-chart Option

The chart should appear as:



8) Charts (without obtaining frequencies)

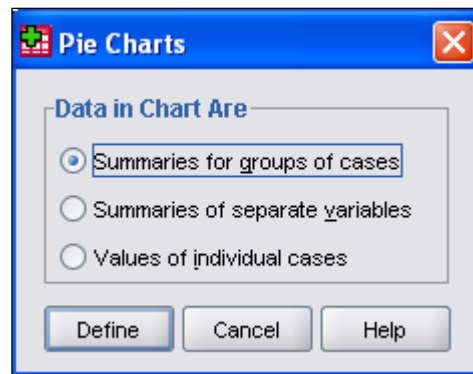
In SPSS v 17 there are a number of different ways to generate graphs which can be seen by clicking on the menu bar.



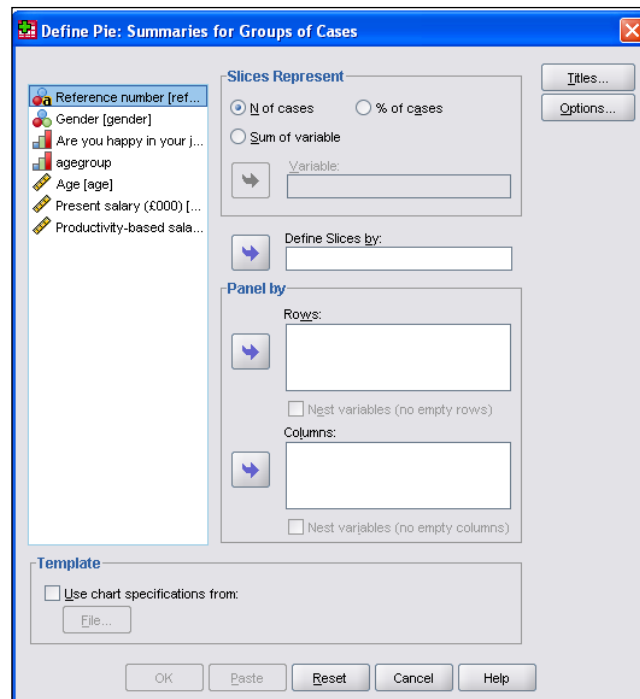
The option **Legacy Dialogs** is consistent with procedures available in earlier versions of SPSS, but more experienced users may choose to experiment with the alternative approaches to producing more interactive graphs.

a) **Obtain a pie-chart for the jobhappy variable**

Click on **Graphs, Legacy Dialogs and Pie.**



Select **Summaries for groups of cases** and click on **Define** to give:



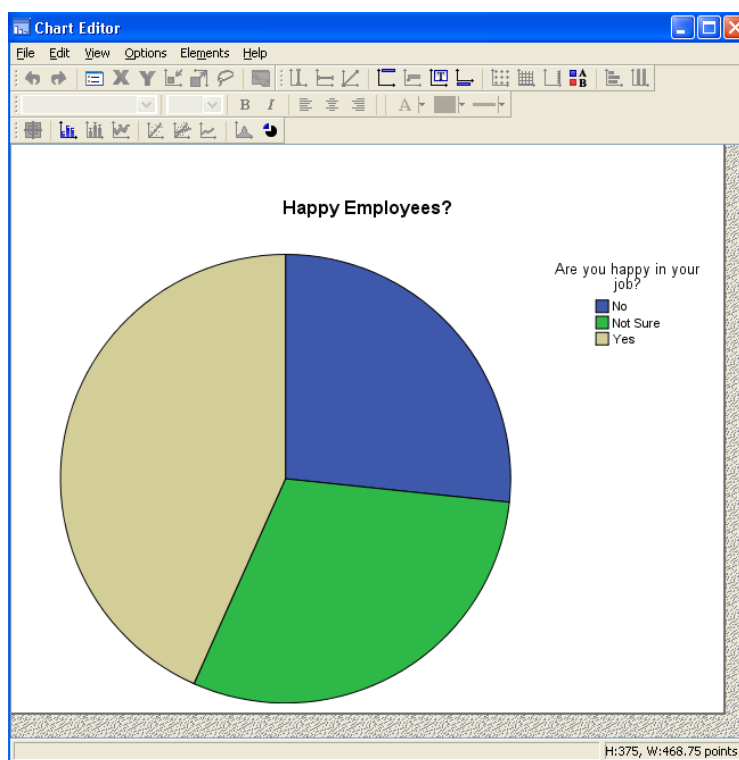
Click on **Jobhappy** and on the “Define Slices by” arrow  ..

N (umber) of cases should already be selected for Slices Represent.

Click on **Titles**, type in **Happy Employees?**, and click on **Continue, OK.**

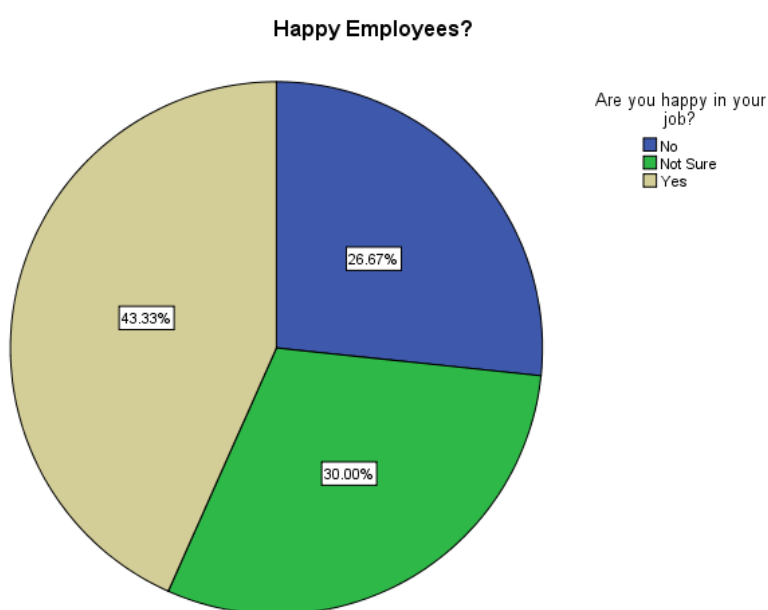
Editing the Chart (Add labels)

Click and then double-click on the **Chart** and the SPSS Chart Editor Window should appear.



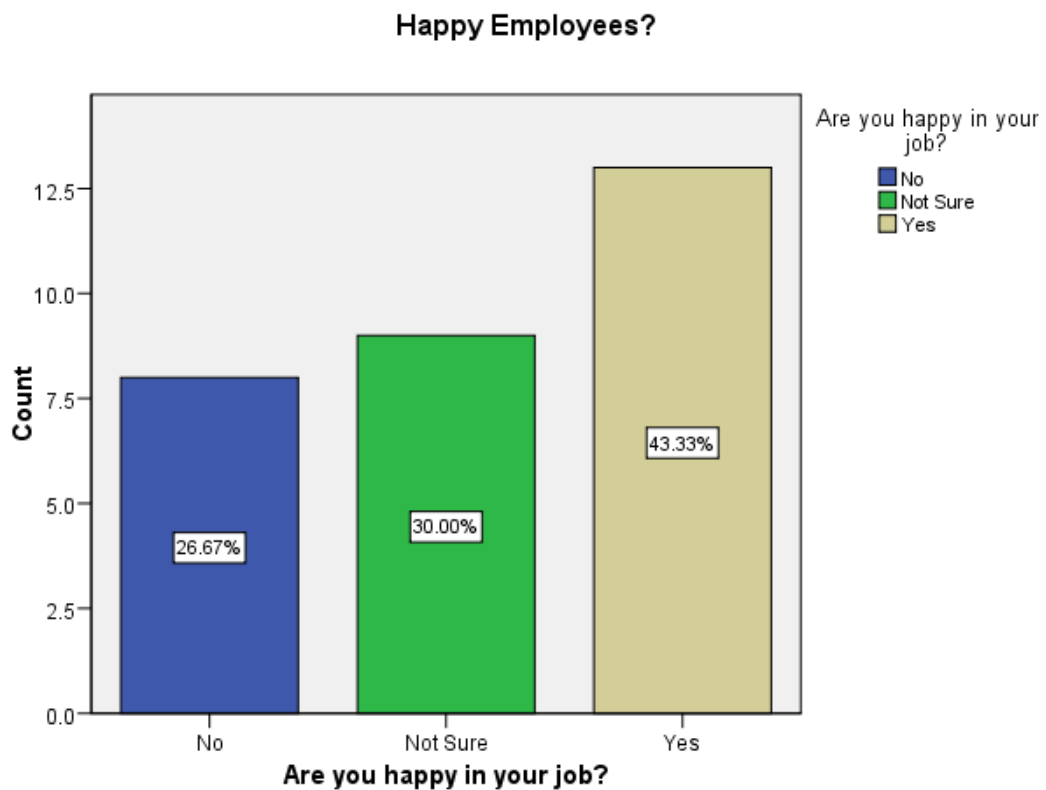
Click on the pie chart and it should then be highlighted with a blue border.

Click on **Elements**, **Show Data Labels** and the Properties window automatically appears for further editing. At this stage it is possible to change the labels, colours, titles or chart type. Close the Properties window to show the following chart.



Changing the chart type

Double click on the pie chart in the Chart Editor to highlight again with a blue border. In the Properties window, click on the **Variables** tab and change the Element type from **Pie** to **Bar** from the drop-down menu, **Apply**, **Close**.



Editing the Bar Chart

Change bar colour to white:

Double click on the bars and the Properties window should appear. Alternatively the Properties window can be opened with a right click over the selected bars.

Click on the **Fill & Border** tab, select each bar in turn, click on **Fill** and change the colour to **white** by clicking on the white rectangle, **Apply**.

Change bar style to drop shadow

Click on the **Depth & Angle** tab and click on **Shadow** under Effect, **Apply**, **Close**.

Remove the Legend

Double click on the legend and press delete on the keyboard

Change position of bar labels

Double click on the bar labels in the centre of the bars.

In the Properties window click on the **Data Value Labels** tab.

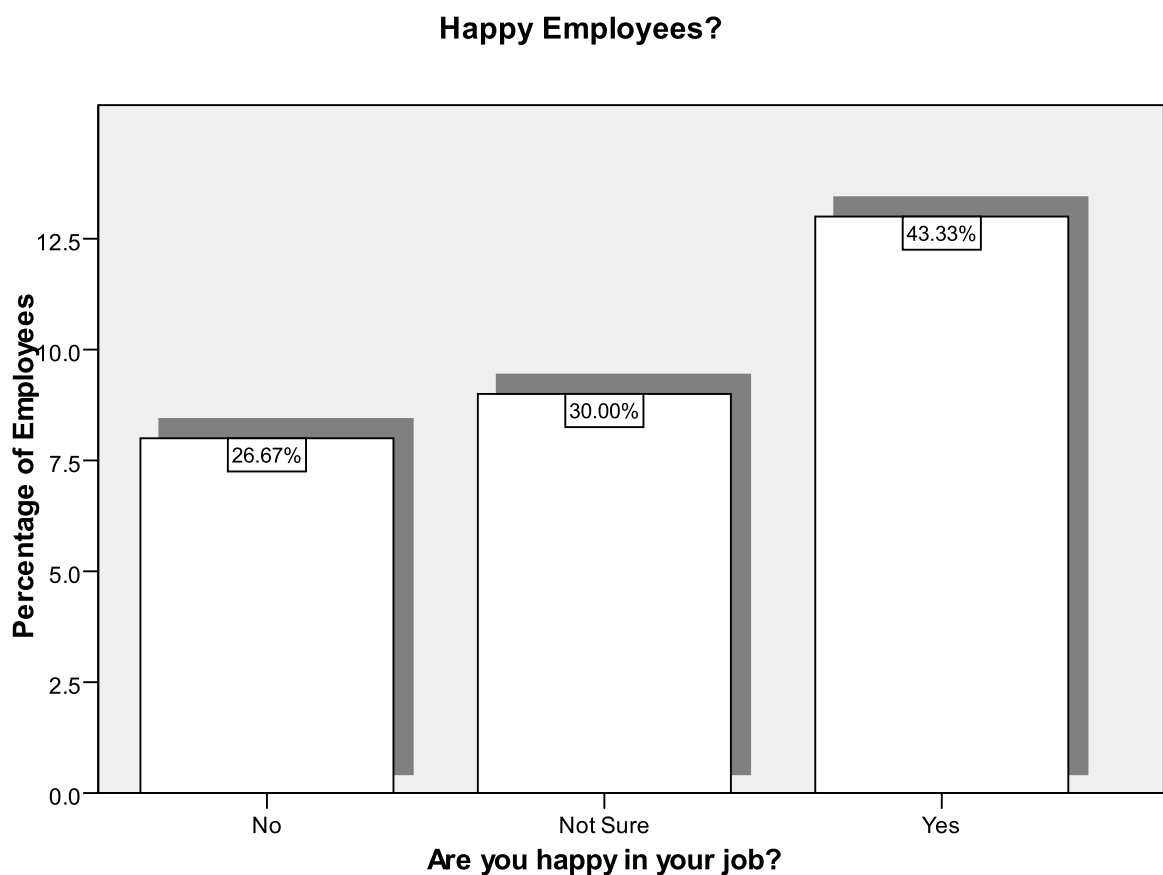
Click on **Custom** under Label Position and click on the **Above Center** icon.

Apply, Close.

Edit the Axis Label

Click on the vertical axis label and then click on it again to change it to **Percentage of Employees**.

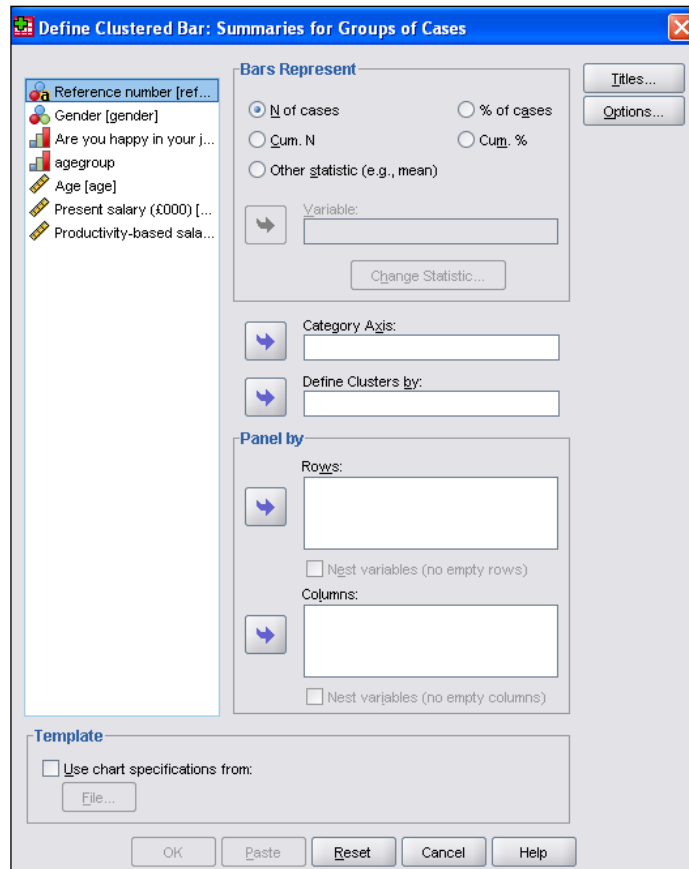
The edited chart should now appear as:



Close the Chart Editor window when editing has been completed

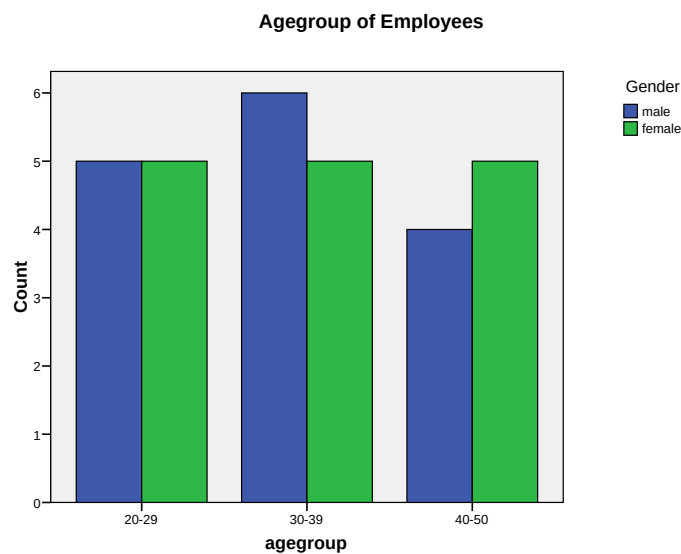
b) Cluster Bar Chart for gender and agegroup

Select **Graphs, Legacy Dialogs, Bar, Clustered, Define** to give:



Select **Agegroup** for the Category Axis, **Gender** for Define Clusters by.

Add a Title, Continue, **OK**. The following chart appears, displaying a bar for each combination of Agegroup and Gender:



The frequencies and percentages represented by each bar can be displayed using cross-tabulations as shown in the next examples.

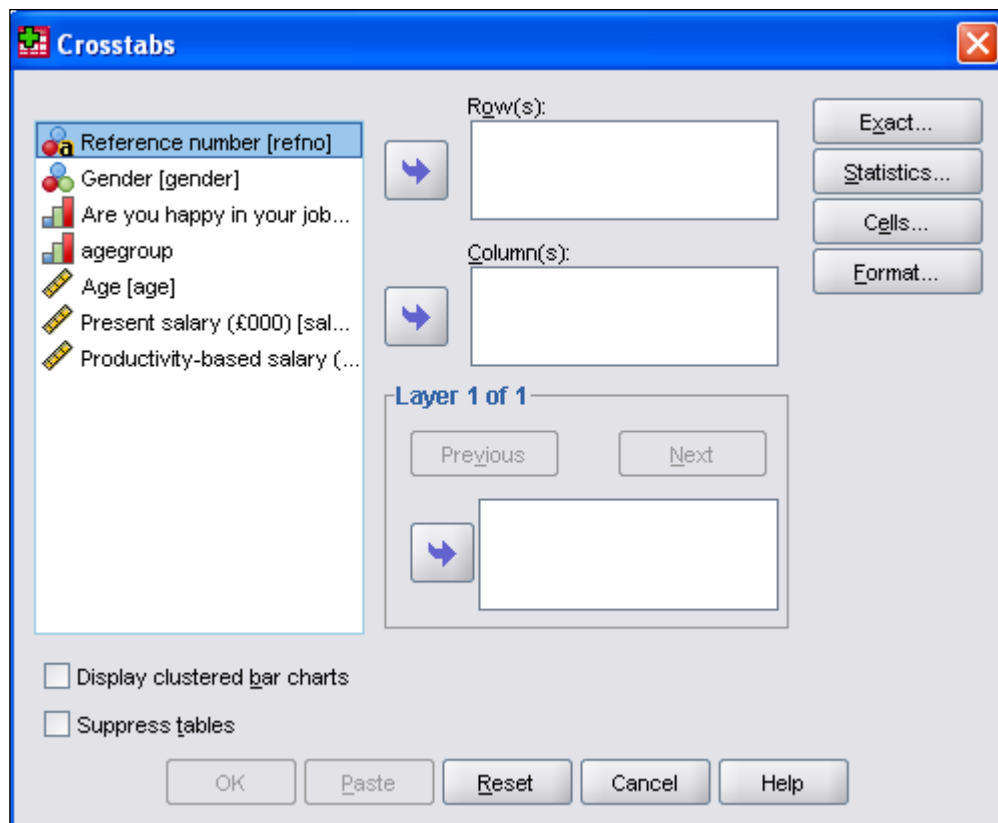
9) Cross-tabulations

a) Cross-tabulation of gender and agegroup

Produce a cross-tabulation between gender and agegroup to show the frequencies in each possible combination of the gender and agegroup values.

Click on **Analyse, Descriptive Statistics** and **Crosstabs**.

Select **Gender** for the row variable and **Agegroup** for the column variable, **OK**.



The output includes:

Gender * Agegroup Crosstabulation

Count		Agegroup			Total
		20-29	30-39	40-50	
Gender	male	5	6	4	15
	female	5	5	5	15
Total		10	11	9	30

Now click on the **Go to Data Editor** icon  to return to the data.

b) Cross-tabulation of Gender and Jobhappy

Row, column or total percentages can be obtained together with frequencies.

Click on **Analyse, Descriptive Statistics and Crosstabs**.

Select **Gender** for the Row variable and **Jobhappy** for the Column variable.

Click on **Cells** and select **Row** under Percentages, **Continue** and **OK**, to give the following:

Gender * Are you happy in your job? Crosstabulation						
			Are you happy in your job?			Total
			No	Not Sure	Yes	
Gender	male	Count	7	5	3	15
		% within Gender	46.7%	33.3%	20.0%	100.0%
	female	Count	1	4	10	15
		% within Gender	6.7%	26.7%	66.7%	100.0%
Total		Count	8	9	13	30
		% within Gender	26.7%	30.0%	43.3%	100.0%

This shows that 10 out of 15 or 66.7% of females are happy with their jobs, whereas 3 out of 15 or 20% of males are happy with their jobs.

10) Histograms

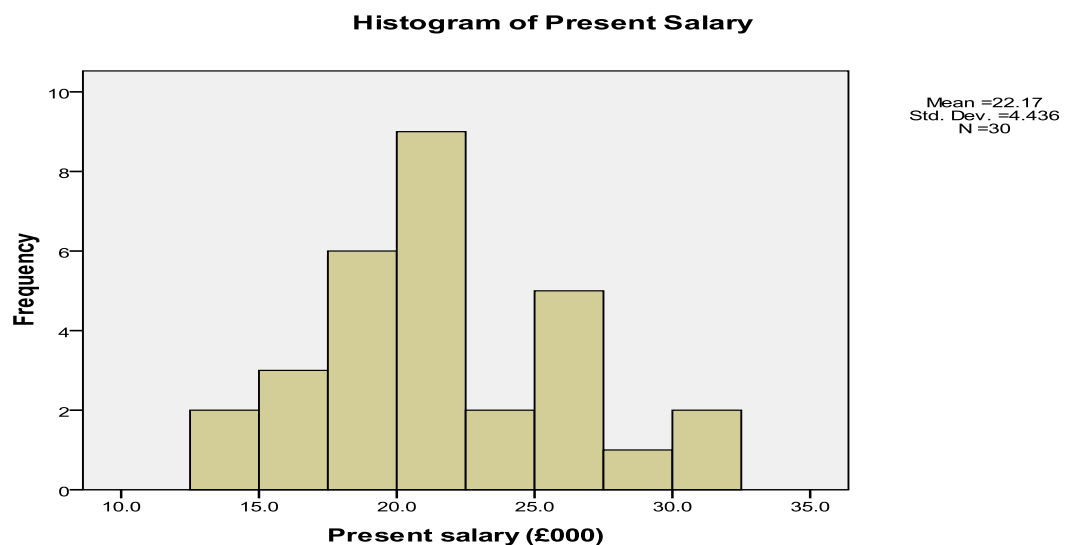
A histogram is used with quantitative variables to show the frequencies within a number of class intervals.

Histogram of salary

Click on **Graphs, Legacy Dialogs, Histogram**, select **Salary** for the variable.

Click on **Titles** and type **Histogram of Present Salary**

Click on **Continue, OK**, to give:



SPSS has used 8 intervals in the construction of the histogram.

The mean salary, standard deviation and number of values are also displayed.

Editing the Histogram

Double-click on the Histogram to view the SPSS Chart Editor.

To change the Bars to White

Double click on the bars and click on the Fill & Border tab in the Properties window. Click on **Fill** and then change the colour of the bars to white by clicking on the white rectangle, **Apply**, **Close**.

To improve the labelling

Double click on the Salary Axis to open the Properties window.

Click on the Scale tab.

For Major Increment deselect **Auto** and enter **2.5** under Custom, **Apply**, **Close**.

Click on the Frequency Axis Title, and then click again to edit the title. Change the title to **Number of Employees**.

Close the Chart Editor window.



Note that the boundary values of the intervals are labelled on the histogram.

In some cases SPSS labels the mid-points of the intervals.

This is dependent on the convenience of the data.

11) Averages and Measures of Spread

The mean, number of values and standard deviation are automatically displayed with a **Histogram**.

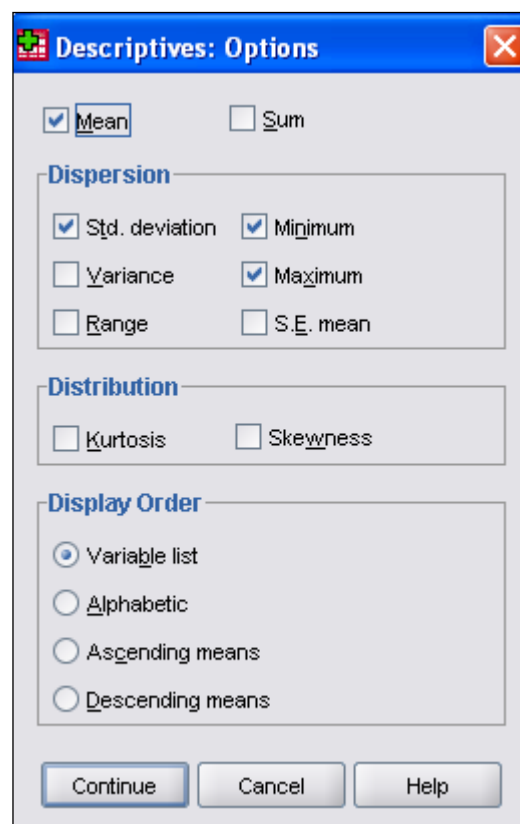
Frequencies (Analyse, Descriptive Statistics, Frequencies) has a Statistics option which will display such things as mean, median, mode, sum, standard deviation, maximum, minimum, range, variance, percentiles, etc.

Descriptives (Analyse, Descriptive Statistics, Descriptives) will give a number of measures of average and spread under Options. Although there are fewer options than with Frequencies, the Frequencies tool is not always appropriate.

Descriptives for salary and prodbsal

Return to the data editor then select **Analyse, Descriptive Statistics, Descriptives**.

Select both **Present salary** and **Productivity-based salary** for the variables and click on **Options** to give:



Select **Mean**, **Std. deviation**, **Maximum** and **Minimum**, **Continue** and **OK**.
The output should be:

Descriptive Statistics

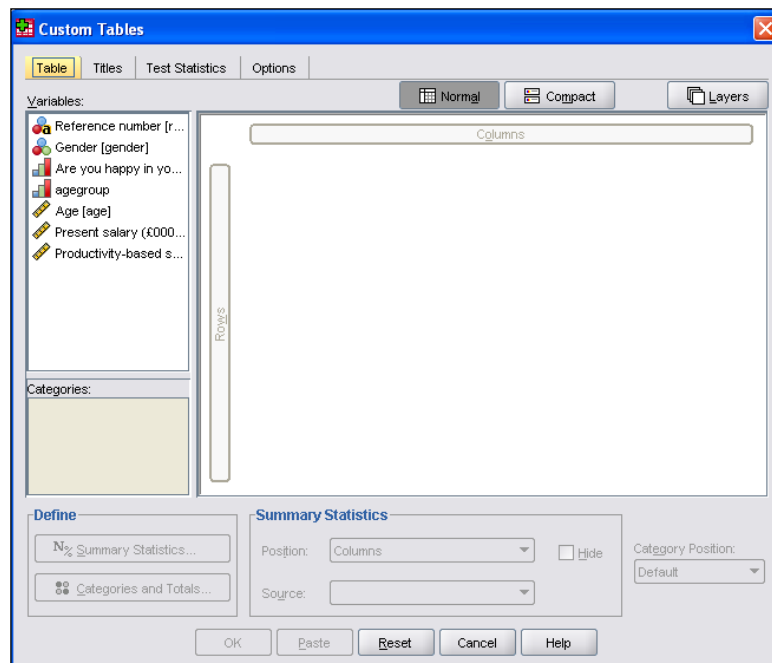
	N	Minimum	Maximum	Mean	Std. Deviation
Present salary (£000)	30	14.9	32.3	22.173	4.4357
Productivity-based salary (£000)	30	15.1	33.4	22.457	4.4264
Valid N (listwise)	30				

12) Comparison of Means

Summarising data for different groups of people can be displayed using a table of means and a bar chart of means.

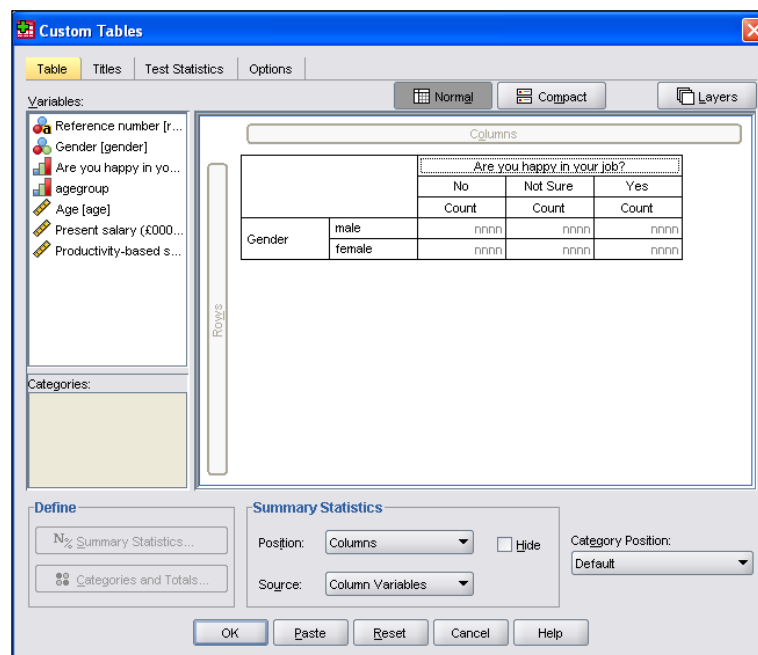
a) Table of Mean Salaries for Gender and Jobhappy

Select **Analyze, Tables, Custom Tables**. Click **OK** to the dialogue box which provides advices on the use of variables within Custom Tables.



‘Click and drag’ the variables with **Jobhappy** for Columns and **Gender** for Rows.

The following table will be generated.



Click and drag **Salary** to replace the **Count** for the cells of the table and the **Mean** will automatically be generated.

Click on the **Titles** tab and type **Mean Salaries of Employees** under Title. **Continue, OK.**

Mean Salary of Employees				
		Are you happy in your job?		
		No	Not Sure	Yes
		Present salary (£000)	Present salary (£000)	Present salary (£000)
		Mean	Mean	Mean
Gender	male	21.8	21.9	23.5
	female	19.8	24.6	21.4

b) Bar Chart of Mean Salaries for Gender and Jobhappy

Select **Graphs, Legacy Dialogs, Bar, Clustered, Define.**

Click on **Other statistic** under Bars Represent.

Select **Salary** for Variable.

Note that the mean score is selected as the default, but other statistics could be selected by clicking on **Change statistic**).

Select **Jobhappy** for the Category Axis and **Gender** for Define Clusters by.

Click on **Titles** and type **Mean Salaries of Employees, Continue, OK.**



13) Reporting a Statistical Output

It is useful to be able to insert relevant tables or charts into a word processed document as part of a report on statistical analysis.

To transfer SPSS output to a WORD document:

Use the output created from the Comparison of Means in the last analysis.

Click on the table of means from the output and this should now be surrounded by a border with a red arrow on the left-hand side.

Select **Edit, Copy.**

Minimise SPSS by clicking on the minus sign in the top right-hand corner of the screen. Similarly, it may be necessary to minimise the data editor, by clicking on the minus sign in the top right-hand corner of the screen.

Load **WORD** and type in:

The results below show the analysis of the mean salaries for gender and jobhappy.

Press the return key twice on the keyboard.

Then select **Edit, Paste Special**, click on **Picture, OK**.

(Note that **Edit, Paste** does not always produce the best results for inserting SPSS output into WORD documents and hence **Paste Special** is a preferable alternative, especially for tables).

The selected output should now be pasted into the WORD document.

Press the return key twice on the keyboard before minimising WORD and returning to the SPSS output (this should be seen as an icon at the bottom of the screen).

Click on the Bar Chart of Mean Salaries of Employees and repeat the process:

Edit, Copy, minimise SPSS.

Return to WORD and select **Edit, Paste, OK**.

(Note that this time there is no **Picture** option in the **Paste Special**, and **Paste** gives the best results).

Notes

- (i) The size of the output table can be changed by clicking and dragging the small squares on the sides of the border of a selected table.
- (ii) The titles of the tables have to be copied and pasted separately, but it may be preferable to type these directly into the WORD document as an alternative.
- (iii) **Insert, New Text** can be used as an alternative to add text to the SPSS output.