

Analysing quantitative data using IBM SPSS Statistics (version 24)

1) INTRODUCTION

1.1) Aims

- ♦ To illustrate computational analysis of data.
- ♦ To introduce the basics of quantitative data analysis.
- ♦ To provide familiarisation with the IBM SPSS Statistics package so that the user can begin to assess its suitability for their own analysis.

2) BACKGROUND

2.1) IBM SPSS Statistics for Windows

IBM SPSS Statistics for Windows is a powerful computer package providing statistical analyses and data management. The IBM SPSS Statistics suite of programs is the most widely used statistical analysis package in the world.

2.2) Data types

Before data are analysed in IBM SPSS Statistics, it is necessary to understand the type of data as this will affect the analysis used.

- ♦ Nominal: nominal data consist of values which can be grouped into categories, for example gender which can be grouped into male and female; but cannot be expressed numerically
- ♦ Ordinal: ordinal data consist of values which can be placed in a definite order, for example agree, not sure, disagree; but cannot be expressed numerically
- ♦ Scale or quantifiable: scale data consists of values that can be expressed numerically as quantities; for example year of birth. Scale or quantifiable data can further be subdivided into two groups:
 - ♦ Discrete, where individual items of numeric data can have one of a finite number of values within a specified range; such as spinal column point for the variable salary scale. The value can usually be counted and it changes in discrete units, in this case whole numbers. In some instances discrete data may be rank data, for example the order a group of people finished in a race
 - ♦ Continuous, where numeric data is not restricted to specific values and is usually measured on a continuous scale; such as journey to work distance (in km).

Journey to work distance along here
0km 120km



With such data it is possible to tell the interval between the data values for different cases; for example the interval between a journey to work of 15 miles and another of 2 miles is 7 miles.

One potentially confusing aspect of IBM SPSS Statistics is that all data are usually coded numerically (e.g. 1 = male). Although it appears less meaningful to code such responses numerically, it is better from a data manipulation point of view since IBM SPSS Statistics allows only automatic recoding on codes, which are numeric.

2.3) The **TEACH1.SAV** data file

This data file consists of data about 347 people recruited to work for a UK local authority over a ten year period from the mid 1970s to mid 1980s and was obtained from personnel records. The data collection form is included as Appendix 2.

The vast majority of data relates to the time of their appointment and is taken from a range of secondary data sources such as their application form and the requisition orders to various media outlets for vacancy advertisements.

The data refer predominantly to non-manual employees, although there are a few manual employees. The data have been anonymised in a variety of ways and all locational data has been amended to preserve confidentiality. Permission was obtained to use these data in suitably anonymised format for teaching purposes.

The data file can best be thought of as a large spreadsheet with each column representing a variable for which data are available and each row representing that data collected for an individual or case:

	gender	born	marital	educate	profmemb	
1	2	67	1	5	3	
2	1	19	.	7	3	
3	2	24	2	7	3	

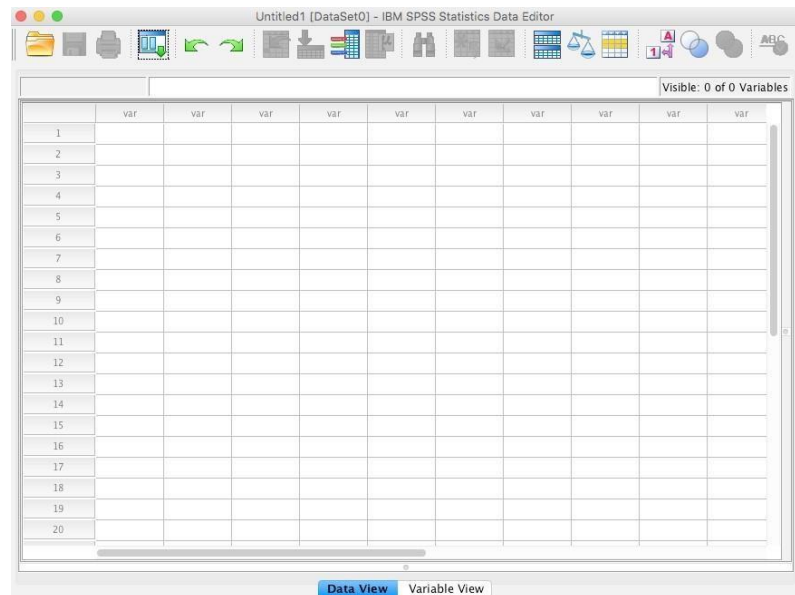
Thus for the table above, row 1 represents a person who has gender code 2 (female), was born in 1967, has marital status code 1 (single), was educated up to code 5 (O'level / GCSE grade C or above), and professional membership code 3 (none). The data then continues to the right for further variables. The symbol “.” A full list of variables and their codes are given in Appendix 1.

End of Exercise 1

3.1) The IBM SPSS Statistics Windows

When you load and run the IBM SPSS Statistics package it opens up a menu bar and two views. These are Data View (currently visible) and Variable View...

Data View This sheet will contain your data, each column representing a variable for which data are available and each row representing the data for an individual or case. At present this sheet should be blank. As



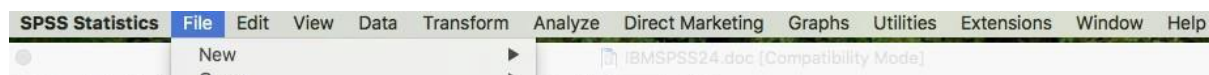
this sheet is currently selected, its name on the tab at the bottom is in bold.

Variable View

The structure of the data is maintained in the variable view tab;

At present this sheet is not visible as the view sheet is not active. Consequently the name is bold. Do not bother to click on the tab and this sheet yet, we will do that later.

Now move your mouse across the menu bar and click on the various drop down menus.



This provides a selection of options (File, Edit, View, Data...) which allow you for example to open files, edit data, generate graphs, create tables and perform statistical analyses. Selecting from this menu bar will, like in other windows packages, provide further pulldown menus and dialogue boxes.

The menu bar options are used as follows:

- ◆ **File** is used to access any files whether you want to **Open** an existing IBM SPSS Statistics file or read data in from another applications such as Excel of dBase, or start a **New** file. It is also the menu option you choose to **Save** files
- ◆ **Edit** can be used to alter data or text in the Data View or the Variable View and insert variables and cases
- ◆ **View** can be used to alter the way your screen looks. Please leave this on the default settings
- ◆ **Data** is used to define variables and make changes to the data file you are using
- ◆ **Transform** is used to make changes to selected variable(s) in the data file you are using. This can include **recode(ing)** existing variables and **compute(ing)** new variables
- ◆ **Analyze** is used to undertake a variety of analyses such as producing **Reports**, calculating **Descriptive Statistics** such as **Frequencies** and **Crosstabs** (crosstabulations) and associated summary statistics, as well as various statistical procedures such as **Regression** and **Correlation**
- ◆ **Direct Marketing** is a specialised option for use in direct marketing campaigns
- ◆ **Graphs** is used to build a variety of graphs and charts such as **Bar**, **Line** and **Pie** charts
- ◆ **Utilities** is for housekeeping such as displaying information on variables as well as more advanced manipulations
- ◆ **Extensions** is used to access online undertake other (more advanced) statistical analyses
- ◆ **Windows** operates in the same way as normal ◆ **Help** is a context sensitive help feature.

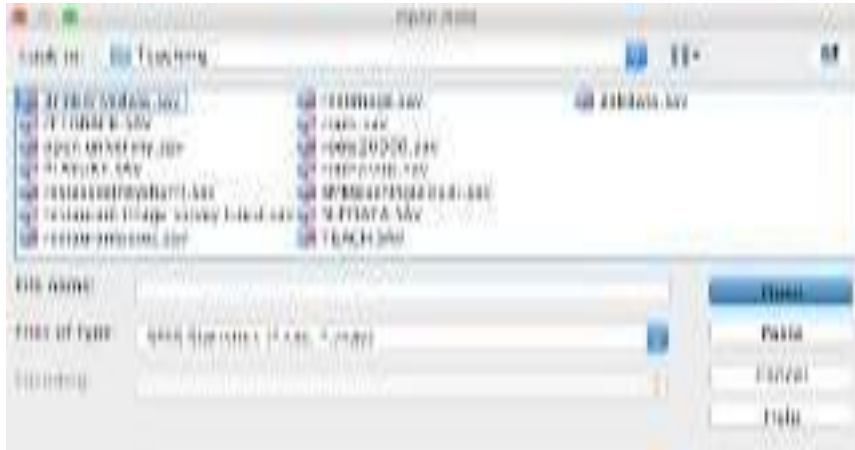
Exercise 2: To load a previously created IBM SPSS Statistics for Windows data file [teach1.sav](#)

Download the [teach1.sav](#) file to your OneDrive | U drive | Pen drive.

End of Exercise 2

Exercise 3: To check how variables have been coded

Open IBM SPSS Statistics, open the [teach1.sav](#)



You will now see data appear in the Data View window and the filename above the menu bar change. This may take some time so be patient!

To see the data view **Data View** at the bottom of

	gender	born	marital	educate	profme	prevemp	preveast	prevnor	appleast	applnor	occupat	empstat
1	2	60	2	5	3	3	530	179	620	128	49.20	4
2	2	47	2	.	3	5	.	.	620	128	46.30	4
3	2	57	2	5	3	3	576	156	626	126	46.30	4
4	2	67	1	5	3	4	.	.	608	109	33.10	3
5	1	19	.	7	3	9	.	.	608	115	13.10	4
6	2	24	2	7	3	9	.	.	380	107	13.10	4
7	2	58	2	7	3	5	.	.	626	126	46.30	4
8	1	69	1	4	3	4	.	.	614	151	6.10	3
9	.	60	2	4	3	5	.	.	620	127	49.20	4
10	1	42	2	5	1	3	566	149	626	126	13.30	1
11	1	67	1	4	3	4	.	.	620	128	46.30	4
12	2	58	2	7	3	5	.	.	620	128	46.30	4
13	2	64	2	3	1	1	531	105	630	79	6.10	4
14	2	67	2	2	3	5	.	.	593	135	8.00	4
15	2	63	1	2	3	5	.	.	342	62	33.10	3

To check what the column heading for each variable and the codes refer to:

sheet at the bottom of the window. You will now see:

The first column contains the variable **Name**, in the case of the first row “gender”. This is the column heading that appears in the **Data View**.

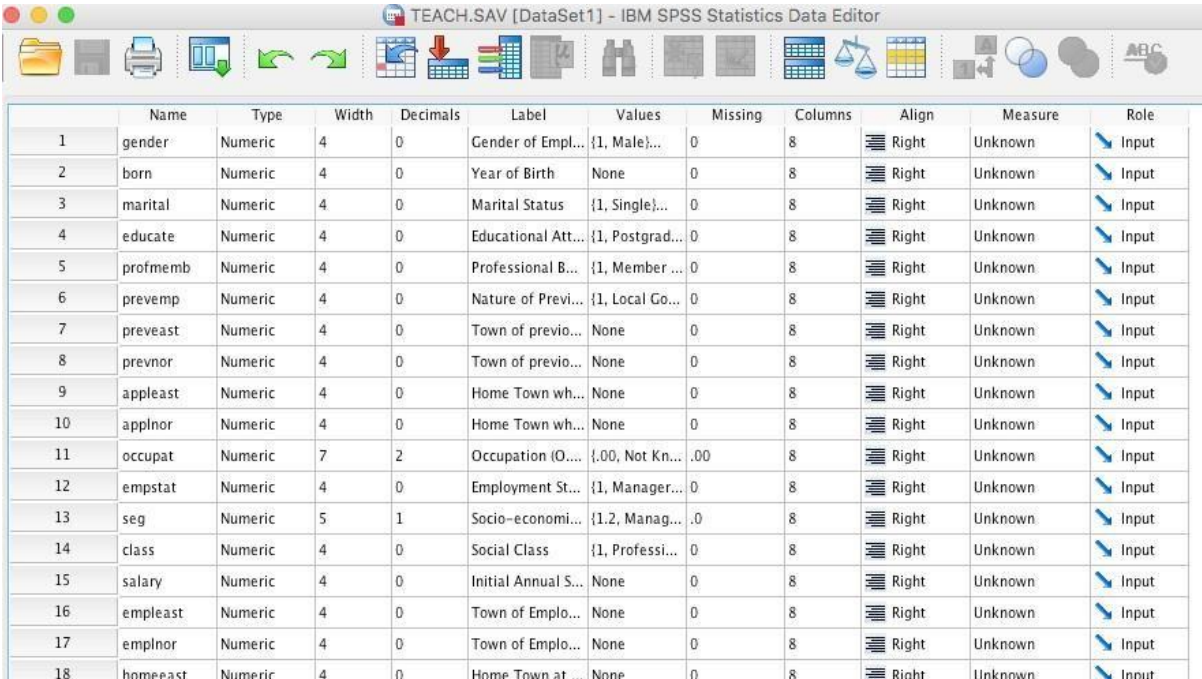
The second column refers to the **Type** of data. Although gender is categorical data, it is referred to as numeric because numeric code values have been used! The key to these code values is given in the column headed **Values**.

fifth column contains the variable’s **Label**. At present this is partially obscured by the subsequent column.

The To see the full value label:

Move your mouse pointer in-between the **Label** and the **Values** column headings until this  appears.

Click **Variable View**



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	gender	Numeric	4	0	Gender of Empl...	{1, Male}...	0	8	Right	Unknown	Input
2	born	Numeric	4	0	Year of Birth	None	0	8	Right	Unknown	Input
3	marital	Numeric	4	0	Marital Status	{1, Single}...	0	8	Right	Unknown	Input
4	educate	Numeric	4	0	Educational Att...	{1, Postgrad...	0	8	Right	Unknown	Input
5	profmemb	Numeric	4	0	Professional B...	{1, Member ...	0	8	Right	Unknown	Input
6	prevemp	Numeric	4	0	Nature of Previ...	{1, Local Co...	0	8	Right	Unknown	Input
7	preveast	Numeric	4	0	Town of previo...	None	0	8	Right	Unknown	Input
8	prevnor	Numeric	4	0	Town of previo...	None	0	8	Right	Unknown	Input
9	appleast	Numeric	4	0	Home Town wh...	None	0	8	Right	Unknown	Input
10	applnor	Numeric	4	0	Home Town wh...	None	0	8	Right	Unknown	Input
11	occupat	Numeric	7	2	Occupation (O...	{00, Not Kn...	.00	8	Right	Unknown	Input
12	empstat	Numeric	4	0	Employment St...	{1, Manager...	0	8	Right	Unknown	Input
13	seg	Numeric	5	1	Socio-economi...	{1,2, Manag...	.0	8	Right	Unknown	Input
14	class	Numeric	4	0	Social Class	{1, Professi...	0	8	Right	Unknown	Input
15	salary	Numeric	4	0	Initial Annual S...	None	0	8	Right	Unknown	Input
16	empleast	Numeric	4	0	Town of Emplo...	None	0	8	Right	Unknown	Input
17	emplnor	Numeric	4	0	Town of Emplo...	None	0	8	Right	Unknown	Input
18	homeeast	Numeric	4	0	Home Town at ...	None	0	8	Right	Unknown	Input

 Click and drag the column width to the right until the variable's label can be read.

Label	
Gender of Employee	{1,
Year of Birth	Mar

Value Labels

Value:

Label:

Add Change Remove

1 = "Male"
2 = "Female"

Spelling...

Help Cancel OK

Note: If you wish to edit a variable's label just retype the label in the appropriate cell.

The sixth column contains the key to the codes used for each variable. These are known as the **Values** Labels.

To see the Value Labels used:

 Click on the cell containing the first value for the variable gender {1, Male}...  Click on the  to the right of this cell.

The **Value Labels** dialogue box appears. It shows the current value labels for this variable.

Note: You can also use this option to change each value label for the codes or enter new value labels.

 Click  in the **Value Labels** dialogue box to return to the Variable View.

End of Exercise 3

Exercise 4: To undertake a frequency distribution



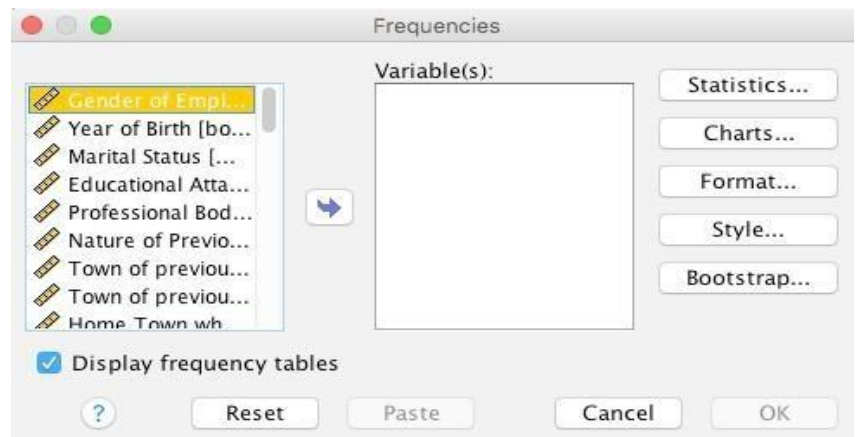
Return to

Data View.

Click **Analyse | Descriptive Statistics | Frequencies**

The **Frequencies** dialogue box appears.

variables are
alphabetically
downward



If not arranged alphabetically use the arrow on the left hand box to scroll down until **Gender** appears. Highlight **Gender** in the left hand box by clicking on it.



Click  to move gender into the **Variable(s)** box.

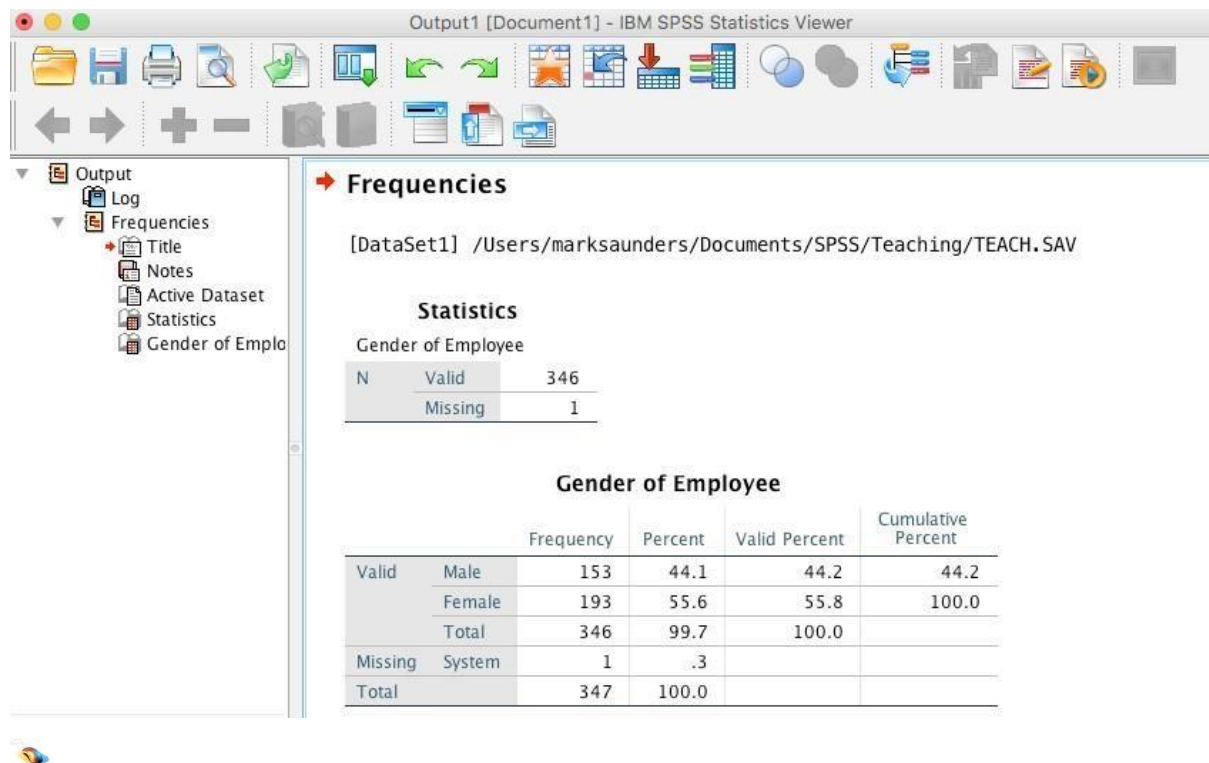


Note the arrow button changes direction and the cursor moves to the **Variable(s)** box. This is to allow you to reverse your decision if you wish.



Click .

You will see a series of tables displayed in the **IBM SPSS Statistics Viewer**. Note that IBM SPSS Statistics tells you if there are missing cases. In this instance, there is one missing case.



The screenshot shows the IBM SPSS Statistics Viewer window. On the left, the 'Output' pane lists 'Frequencies' under 'Log'. The main area displays the 'Frequencies' output for the variable 'Gender of Employee' from the dataset '[DataSet1] /Users/marksaunders/Documents/SPSS/Teaching/TEACH.SAV'.

Statistics
Gender of Employee

N	Valid	Missing
	346	1

Gender of Employee

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	153	44.1	44.2	44.2
	Female	193	55.6	55.8	100.0
	Total	346	99.7	100.0	
Missing	System	1	.3		
Total		347	100.0		

To copy any statistical output click on the output e.g. Gender of Employee tables summary), copy and paste into a word document, where you would then explain and interpret the result.

Repeat this process using **Analyze | Descriptive Statistics | Frequencies** for at least five other variables of your choice. You can do this by pointing and clicking on the menu commands which are visible at the top of your screen.

Whilst you are doing this explore the effect of the **Statistics...** **Charts...** **Format...** buttons on your output.

To remove the variables from the right **Variable(s)** box within the dialogue box either click **Reset** or highlight the variable in the right **Variable(s)** box and click .

To quit this analysis (for example if you make a mistake) click **Cancel**.

You may (or may not!) have noticed that each of the tasks you have performed in IBM SPSS Statistics have been automatically appended to the **IBM SPSS Statistics Output Viewer**. You can see this by scrolling through your output window using the up and down arrows on the right of the window.

You can edit the IBM SPSS Statistics Viewer and save it, or parts of it, to a file which can subsequently be read into a word processor. Alternatively you can print it out.

Help

Help 4.1: To delete output in the IBM SPSS Statistics output viewer

To delete some output in the IBM SPSS Statistics Output Viewer:

is

- Click the area you want to delete; a line will appear around it.

- Press **Delete**.



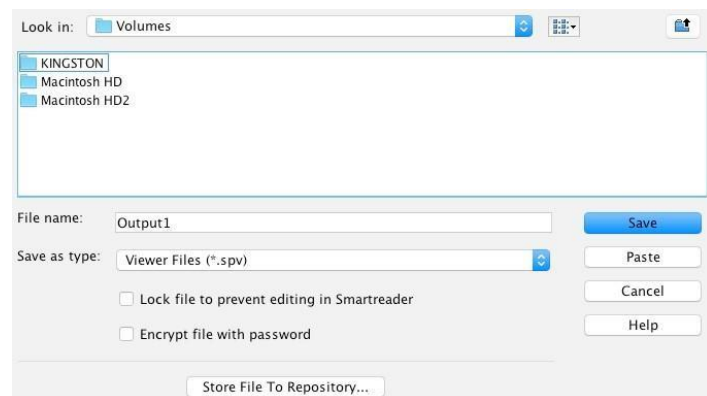
N	Valid	346
	Missing	1

To delete all the output in the IBM SPSS Statistics Output Viewer:

- Ensure that the IBM SPSS Statistics Output Viewer window maximised.

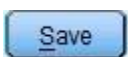
- In the IBM SPSS Statistics Output Viewer click **Edit | Select All**.

- Press **Delete**.



Help

Help 4.2: To save the contents of the IBM SPSS Statistics output viewer to a file

- Click **File | Save As...**
- Type in the filename you wish to save it to in the **File name** box, making sure the file type is ***.spv**.
- Ensure that the file is being saved to the correct drive and directory.
- Click .

End

of Exercise 4

Exercise 5: To calculate the arithmetic mean (average) and the standard deviation


Click **Analyze | DescriptiveStatistics | Descriptives**.

Scroll down to and select the **Journey to work** variable (it's at the end of the variable list), then put it in the **Variable(s)** box.

Click **OK**.

Note how the results are added to the end of your output in the IBM SPSS Statistics Viewer.

Note: you may need to maximise the window by clicking  to see all the statistics.



➔ **Descriptives**

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Journey to work (km) - new job	335	0	131	11.45	16.984
Valid N (listwise)	335				

We can therefore see that the mean journey to work is **11.45 km**.

Calculating a mean makes sense as we are working out the average for scale or quantitative data. However we have to be careful.

End

Exercise 6: To calculate the mode and other measures of central tendency

To calculate the mode for the variable gender:



Click [Analyze | DescriptiveStatistics | Frequencies](#).



Click  to empty the Variable(s) box.

Select **Gender**, then **Statistics...** to put it in the **Variable(s)**

Click **Statistics...** On the right



Select **Mode** in **Central Tendency** Dialogue box by clicking on it.

Click **Continue** to the **Frequencies** dialog box.

This time, we do not want an outputtable so click the box to the left ☐ **Display frequency tables**

To remove the tick.

Click **OK**.

The following will be the IBM SPSS Statistics

➔ Frequencies

Statistics		
Gender of Employee		
N	Valid	346
	Missing	1
Mode		2

We now know that the

most common gender is 2 female.

Repeat this process by calculating the most appropriate average for the following variables: **educate salary class others** of **your choice**

Your choices for the most appropriate average are:

- ♦ Mean: normally known as the average of the data values ♦
- Median: the midpoint once all the data values have been ranked
- ♦ Mode: the data value that occurs most often.

End

of Exercise 6

Exercise 7: To produce a bar chart

Click **Analyze | Descriptive Statistics | Frequencies**.

Deselect all variables by clicking the **Reset** button. Scroll down and select the variable (social class) **Class**.

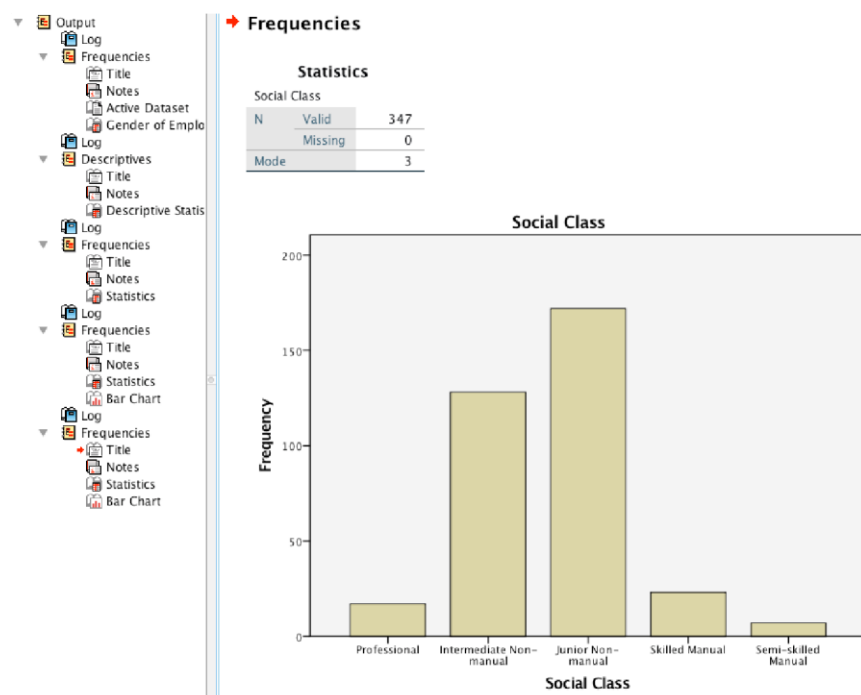
Click **Charts...**.

Use the radio button to select **Bar Charts | Continue**.

At the **Frequencies** dialogue box click **OK**.

The IBM SPSS Statistics Output Viewer will contain your bar chart.

Notice that missing data are automatically excluded from the chart. Notice also that you are presented with a different menu bar which allows you to edit the current chart and other options.



To the left of your bar chart are a series of icons. These provide an index to your output that is IBM SPSS Statistics Output Viewer.



Click the  Statistics

the left towards the  to see what happens.

icon on



Now practice your charting skills by creating another bar chart for the variable

bottom

(educational attainment) **Educate** but with the vertical axis displaying percentages rather than frequencies.

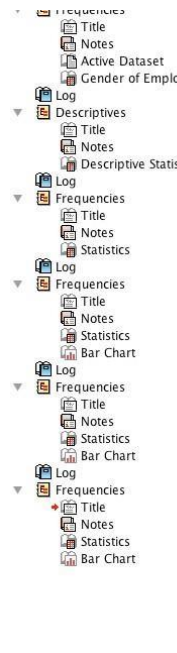
You will need to:

- ◆ Deselect the variable

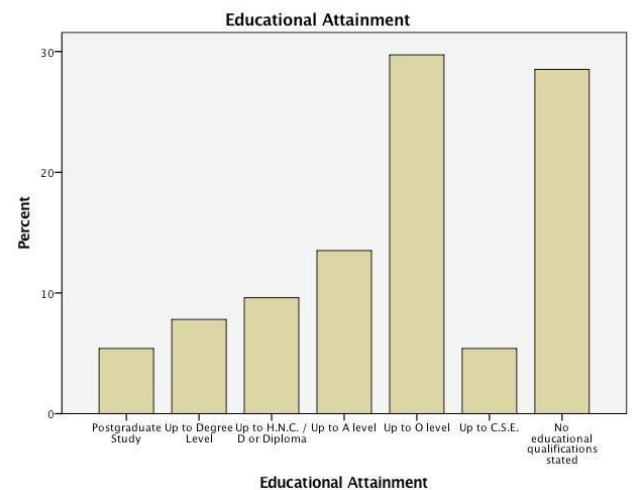
Class and select the variable **Educate**.

- ◆ Select **Percentages | Continue**.

This will give you a chart like the one on the right.



Statistics		
Educational Attainment		
N	Valid	333
	Missing	14
	Mode	5



End of Exercise 7

Exercise 8: To create a table (Crosstab) of one variable against another

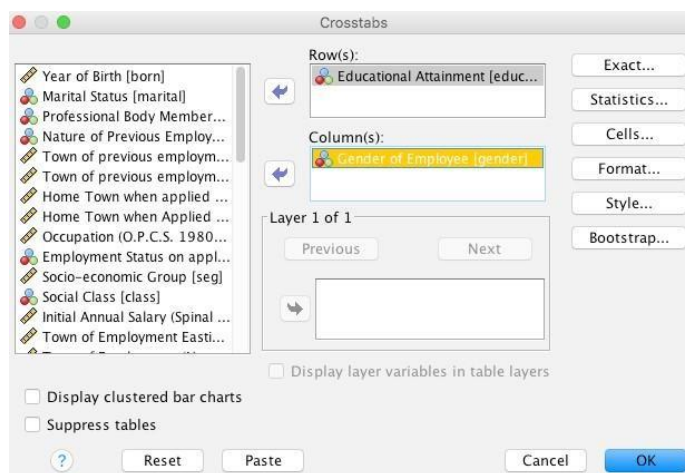
One of the most useful features of IBM SPSS Statistics is its ability to create crosstabulations of one variable against another. In this exercise, you will create a table of the variable

Educate (Educational attainment) by the variable **Gender**.

You will want your table to look like this.

To do this:

Click **Analyze | DescriptiveStatistics| Crosstabs**.



	male	female
postgrad		
up to degree		
up to HNC/D		
up to A'level		
up to O'level (GCSE C+)		
up to CSE (GCSE D-)		
No quals.		

Minimise the IBM SPSS Statistics Output Viewer.

This gives the **Crosstabs** dialogue box.

Select the **Row(s)** variable **Educational Attainment** and the **Column(s)** variable **Gender** using the same principles as when selecting frequencies.

Once you have selected row and column variables, you will be able to click **OK**.

Your table will appear in the IBM SPSS Statistics Output Viewer.

Log
Frequencies
Title
Notes
Statistics
Log
Frequencies
Title
Notes
Statistics
Bar Chart
Log
Frequencies
Title
Notes
Statistics
Bar Chart
Log
Frequencies
Title
Notes
Statistics
Bar Chart
Log
Crosstabs
Title
Notes
Case Processing
Educational Attai

➔ **Crosstabs**

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Educational Attainment * Gender of Employee	332	95.7%	15	4.3%	347	100.0%

Educational Attainment * Gender of Employee Crosstabulation

Count

		Gender of Employee		Total
		Male	Female	
Educational Attainment	Postgraduate Study	11	7	18
	Up to Degree Level	17	9	26
	Up to H.N.C. / D or Diploma	27	5	32
	Up to A level	23	21	44
	Up to O level	42	57	99
	Up to C.S.E.	2	16	18
	No educational qualifications stated	23	72	95
Total		145	187	332

End

of Exercise 8

Exercise 9: To calculate a Chi Square statistic for a table

Minimise the IBM SPSS Statistics Output Viewer.

Analyze Descriptive Statistics

Cross Tabs

Statistics

Select the **Row(s)** variable **Educational Attainment** and the **Column(s)** variable **Gender** using the same principles as when selecting frequencies.

Statistics...

Crosstab: Statistics dialogue box appears. **Chi Square**

Click **Continue**.
Results will be displayed in the IBM Statistics

The key elements of your output are in the row titled **Pearson Chi-Square** and the associated footnote.



Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	52.529 ^a	6	.000
Likelihood Ratio	55.996	6	.000
Linear-by-Linear Association	39.005	1	.000
N of Valid Cases	332		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.86.



The chi square statistic (the value for Pearson Chi Square),

is, in this case, 52.529 with 6 degrees of freedom (df). This is highly significant .000.

The footnote states that no cells have an expected count of less than 5 (so this test result can be used) and that the minimum expected frequency for each cell in the table is 7.86. This means the assumptions of the Chi Square test are satisfied.

So you are testing to see do the two groups behave differently - In this example where $p < 0.05$, this result is thought of as being “highly significant” meaning we think the variables are not independent, so they are associated.

In other words, because the p value .000 is $< .05$ we think that Gender is related to Educational Attainment (i.e. there is an association - Men and Women have different preferences for Educational levels).

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	52.529 ^a	6	.000
Likelihood Ratio	55.996	6	.000
Linear-by-Linear Association	39.005	1	.000
N of Valid Cases	332		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.86.

There are a few steps to interpret the output. So lets take this one step at a time.

In this example where $p < 0.05$, this result is thought of as being “highly significant” meaning we think the variables are related to each other (not independent).

In other words, because p value .000 $< .05$ we think that Gender is associated with Educational Attainment (Men and Women have different preferences for Educational levels).

Finding the P value

- ▶ We have one half of the information we need to find the P value at 0.05
- ▶ The second half is the degrees of freedom, which is the amount of groups present in the test - 1
 - ▶ In this case, $7-1 = 6$ degrees of freedom (df) (shown on the output – can you see it?)
- ▶ Using the 6 degrees of freedom column, we attempt to find the cell which contains our Chi Squared value of **52.529**
- ▶ We can then use the column to determine our P value
 - ▶ In the case of our example, the P value of **52.529** is much higher than **12.59** at 0.05 confidence level.
- ▶ You find these tables for each statistical test online

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	52.529 ^a	6	.000
Likelihood Ratio	55.996	6	.000
Linear-by-Linear Association	39.005	1	.000
N of Valid Cases	332		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.86.

Degrees of Freedom	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

To search for more specific tables, search for Chi Squared tables in your search engine. Links to tables will be provided on Moodle.

Look up the critical chi-square statistic value for $p = 0.05$ (95% confidence level) with 6 df (degree of freedom) shows **12.59**

So Is the result significant?

- ▶ How to check this? Here is the rule of thumb
- ▶ If the test value > table value then **REJECT H_0**
- ▶ In this example, our result showed that **52.529 > 12.59**
- ▶ (So our calculated test statistic is greater than the critical chi-square statistic value for $p = 0.05$ (95% confidence level) with 6 degree of freedom.
 - ▶ We **reject H_0** 'the educational attainment is the same for males and females'
 - ▶ Instead the **H_1 is true** that 'the educational attainment is different for males and females'

Exercise 10: To add row and column per cents to a table using crosstabs

Minimise the IBM SPSS Statistics Output Viewer.

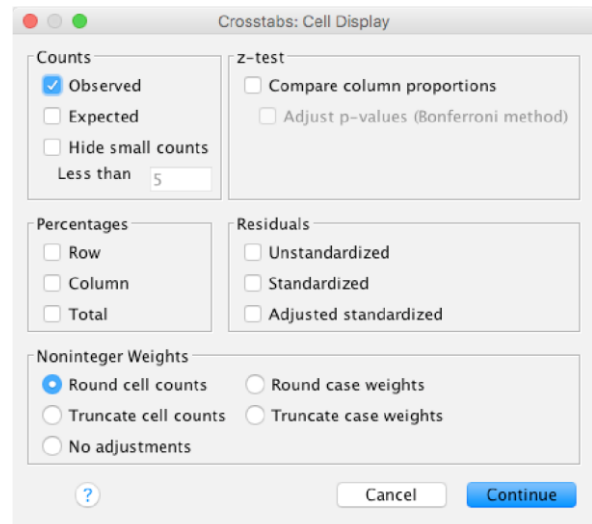
Select two variables you wish to use to create a table as outlined in Exercise 9.

Click **Cells...** near the top right of the crosstabs dialogue box.

Click on Row and/or Column and/or Total in the Cell Display dialogue box (see below) to obtain the desired percentages.

Click **Continue** | **OK**.

Use IBM SPSS Statistics to create further new tables from pairs of variables of your choice.



End of Exercise 10

Exercise 11: To Test for a Significant Relationship Between Two Variables (Correlation)

- Make sure you have loaded IBM SPSS Statistics and that the data file has already been opened.

Correlation analysis is used to measure the strength of the association (linear relationship) between two variables

- ▶ Positive – as x increases, y increases
 - ▶ As advertising dollars increase, sales increase
- ▶ Negative – as x increases, y decreases
 - ▶ As expenses increase, net income decreases

Correlation is only concerned with strength of the relationship

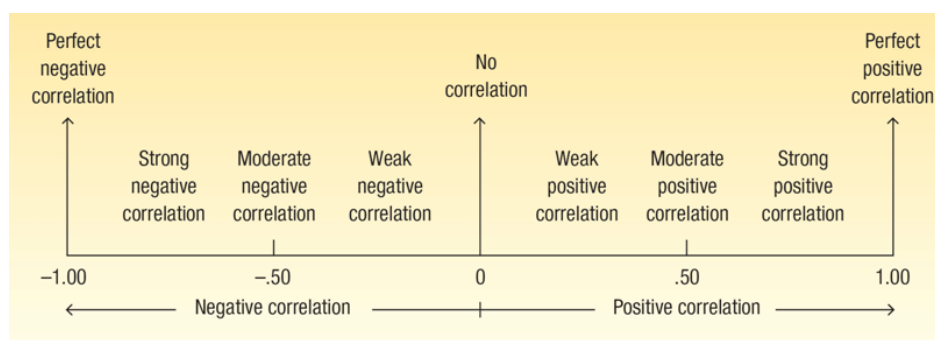
No causal effect is implied with correlation

A spurious correlation is when a correlation exists between two seemingly unrelated variables

- ▶ the correlation between company sales and the number of children employees have

The sample correlation coefficient r is a measure of the strength of the linear relationship between two variables, based on sample observations

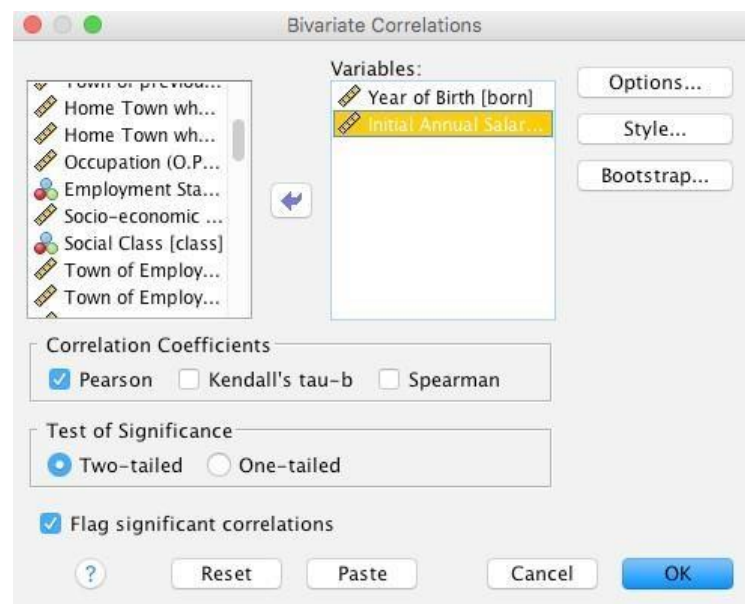
- ▶ Unit free
- ▶ Range between -1 and 1
- ▶ The closer to -1, the stronger the negative linear relationship
- ▶ The closer to 1, the stronger the positive linear relationship
- ▶ The closer to 0, the weaker the linear relationship
- ▶ +1 or -1 are perfect correlations where all data points fall on a straight line



Click **Analyze** | **Correlate** | **Bivariate**.
Click on the first variable you wish to obtain a correlation coefficient with another variable.

Click  to transfer the variable into the **Variables** box.

Repeat this procedure for the other variable(s) you wish to correlate with the first variable.



Choose the most appropriate **Correlation Coefficient** for your data and make sure there is a  in the box.

Choose the most appropriate **Test of Significance** and click on the radio button.

Note: Use a **Two-tailed** test when the direction of the relationship, positive or negative, cannot be specified in advance. Use a **One-tailed** test when it can be specified in advance.

Make sure there is a  in the box **Flag significant correlations**. This will ensure that the significance level is displayed. Click **OK**.

The results of the correlation will appear in the Output Window.

In this output window, the variables **Initial Annual Salary (Salary)** and **Year of Birth (Born)** from the teach.sav data set have been correlated.

Correlations			
		Year of Birth	Initial Annual Salary (Spinal Column Point)
Year of Birth	Pearson Correlation	1	-.013
	Sig. (2-tailed)		.820
	N	343	332
Initial Annual Salary (Spinal Column Point)	Pearson Correlation	-.013	1
	Sig. (2-tailed)	.820	
	N	332	335

IBM SPSS Statistics has produced a correlation matrix. Obviously there is a perfect correlation (1.0000) between the variable initial annual **Salary** and itself.

As the variable is correlated with itself it is impossible to calculate the significance ($p=.$). There is no correlation ($-.013$) between the variable **Salary** and the variable **Born** and this lack of correlation is significant at the 0.820 ($p=0.820$) level.

Stated another way -

Correlation

Using the variables - **Initial Annual Salary (Salary)** and **Year of Birth (Born)**.

Test if there is a relationship between these variables.

		Year of Birth	Initial Annual Salary (Spinal Column Point)
Year of Birth	Pearson Correlation	1	-.013
	Sig. (2-tailed)		.820
	N	343	332
Initial Annual Salary (Spinal Column Point)	Pearson Correlation	-.013	1
	Sig. (2-tailed)	.820	
	N	332	335

There is no correlation ($-.013$), in fact, it is a low negative correlation between the variable **Salary** and the variable **Born** and this lack of correlation is significant at the 0.820 ($p=0.820$) level. The **p value is high** indicating **strong variance**. This means that there is **82% variance** of the points along the line, so the points are not closely related, suggesting no relationship.

So you may have set up the following hypothesis

H_0 : There is no relationship between Initial Annual Salary (Salary) and Year of Birth (Born).

H_1 : There is a relationship between Initial Annual Salary (Salary) and Year of Birth (Born).

Accept the Null hypothesis - the analysis indicates no relationship.

Exercise 12: To run an Independent Samples t-Tests**T-test**

Independent t-test – look at the difference between two groups, say **Gender** where we have data for males and Females. Measure the difference against **Initial Annual Salary** or some variable that will vary!

Analyse | Compare Means | Independent Samples T-Test

To interpret the output

The descriptive statistics of interest are the mean for both groups for the Annual salary.

The main inferential statistics of interest are columns, *t*, *df* and Significance (two-sided *p*)

You can look up the critical *t* value in student *t* tables – find these tables online.

t Table

cum. prob one-tail	<i>t</i> _{.50}	<i>t</i> _{.25}	<i>t</i> _{.20}	<i>t</i> _{.15}	<i>t</i> _{.10}	<i>t</i> _{.05}	<i>t</i> _{.025}	<i>t</i> _{.01}	<i>t</i> _{.005}	<i>t</i> _{.001}	<i>t</i> _{.0005}
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	688.02
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.648	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

You find the critical *t* value on the table - The critical value on the *t* tables at 0.05 (95%) and 100 DF is **1.98**. so if we were able to read 333 DF then the critical value would be even lower than 1.98. The ***t* value calculated is 11.57**

- 1) When your *t* value is larger than the critical *t* value the means are statistically significantly different.

Look at the *p* value is it less than 0.05? Yes **it is .001**, a significant difference

We can look at the confidence interval to see if it crosses zero. If it crosses zero they are not different. In our case the values are 8.98 to 12.661, so **they do not cross zero, so they are different.**

We can see there are significant differences, and for clarity we can now go back to the means, where we can see that Males Initial Annual Salary is double that of females.

Useful also to calculate the effect size.

The Independent T test output

Group Statistics

	Gender of Employee	N	Mean	Std. Deviation	Std. Error Mean
Initial Annual Salary (Spinal Column Point)	Male	151	20.82	10.222	.832
	Female	184	10.00	6.808	.502

Independent Samples Test

Levene's Test for Equality of Variances				t-test for Equality of Means							
		F	Sig.	t	df	Significance One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Initial Annual Salary (Spinal Column Point)	Equal variances assumed	40.278	<.001	11.571	333	<.001	<.001	10.821	.935	8.982	12.661
	Equal variances not assumed			11.139	251.748	<.001	<.001	10.821	.972	8.908	12.735

Independent Samples Effect Sizes

		Standardizer ^a	Point Estimate	95% Confidence Interval	
			Lower	Upper	
Initial Annual Salary (Spinal Column Point)	Cohen's d	8.517	1.271	1.034	1.506
	Hedges' correction	8.536	1.268	1.032	1.502
	Glass's delta	6.808	1.589	1.318	1.858

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

Exercise 13: To load an Excel spreadsheet data file

 **Do not undertake this exercise until you need to load your own data from an Excel Spreadsheet.**

 Make sure that your Excel spreadsheet file is set out with one column per variable and one row for each individual (survey form).

Note: the first row should be the variable names. This is illustrated for the Excel equivalent of an extract from the teach.sav data file below:

	A	B	C	D	E	
1	gender	born	marital	educate	profmemb	
2	2	67	1	5	3	
3	1	19		7	3	
4	2	24	2	7	3	

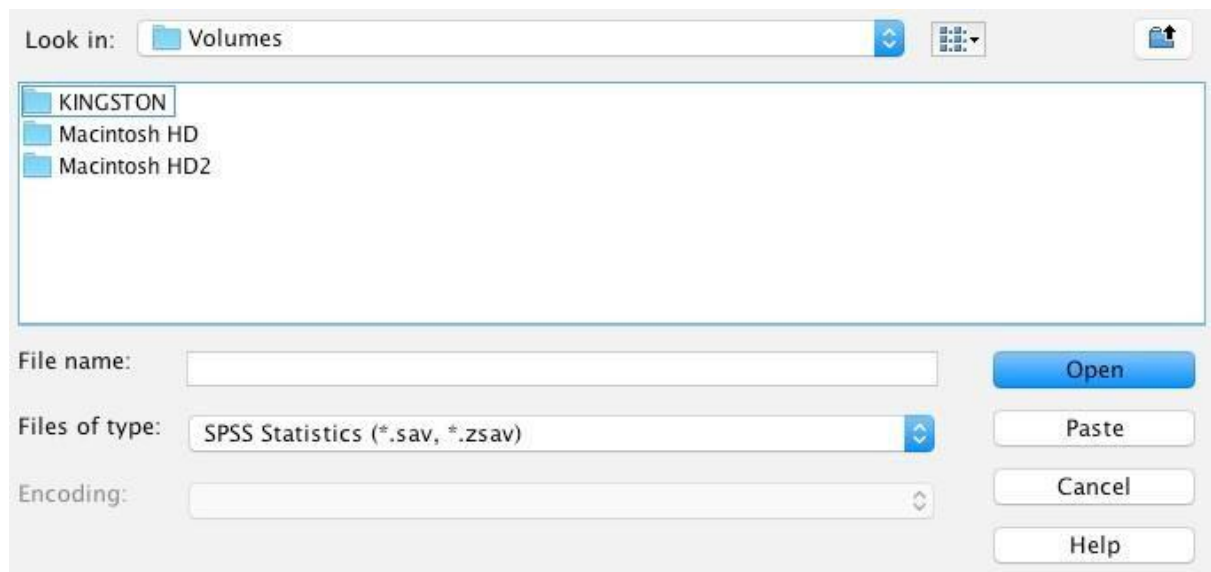
 Make sure you have loaded IBM SPSS Statistics (see Exercise 1).

 Insert your USB portable mass storage device.

 In the Menu Bar, click **File | Open | Data**. The **Open Data** dialogue box appears. Notice that IBM SPSS Statistics looks for data files in the most recently used sub directory. As you are going to load an Excel file from a USB portable mass storage device you need to insert this first.

 Click  in the **Look in:** box.

 Click on  and select **Volumes** to find the appropriate removable disk, for example: KINGSTON.



Click  in the **Files of type:** box and use the scroll arrows on the right of the dialogue box to find Excel.

Click **Excel (*.xls)**. You will see your Excel files displayed in the **Open Data** dialogue box.

Select the filename you want by clicking on it and then click on the **Open** button. The **OpeningFile Options** dialog box appears.

Make sure **names...** the **Data** change. Because you need to load variable SPSS Statistics and save Statistics data file (*.sav).  to the You will see the file appear in above the menu bar so be patient! The file from Excel you will still labels and value labels within IBM your data as an IBM SPSS

End of Exercise

5) FURTHER READING

The most useful book on IBM SPSS Statistics in my opinion is:

→ Field, A. (2023). *Discovering Statistics Using IBM SPSS Statistics* (4th Edn.). London: Sage.

This is both readable and easy to use! It also contains advice regarding when to use different statistical tests.

Another good book on IBM SPSS Statistics for beginners, which also clearly explains the statistics is: Pallant, J. (2013). *IBM SPSS Statistics Survival Manual* (5th edn).

Maidenhead: OUP.

These books offer a clear, non-technical approach to using IBM SPSS Statistics. They assume little familiarity with the data analysis software and cover both inputting data and how to generate and interpret a wide range of tables, diagrams and statistics.

A reasonably straight forward book on collecting your data and preparing it for quantitative analysis is:

→ Saunders, M.N.K., Lewis, P. and Thornhill, A. (2016). *Research Methods for Business Students* (7th edn.), Harlow: Pearson, Chapters 11 and 12.

If you need a statistics book that assumes virtually no statistical knowledge focussing upon which test or graph, when to use it and why. It is written for people who are fearful and anxious about statistics and do not think they can understand numbers then you may find the following helpful:

→ Berman Brown, R. and Saunders, M. (2008). *Dealing with statistics: What you need to know*. Maidenhead: McGraw Hill Open University Press.

6) APPENDIX 1: LIST OF VARIABLES AND THEIR CODES FOR DATA SET TEACH.SAV

Variable names are in capitals with the variable label on the same line. Codes and value labels appear on subsequent lines. **GENDER**

Gender of Employee

- 1) Male
- 2) Female

BORN

Year of Birth

19-- year

MARITAL

Marital Status

- 1) Single
- 2) Married
- 3) Widowed
- 4) Divorced

EDUCATE

Educational Attainment

- 1) Postgraduate Study
- 2) Up to Degree Level
- 3) Up to HNC/D or Diploma
- 4) Up to A level
- 5) Up to O level 6)
- Up to C.S.E.
- 7) No educational qualifications stated

PROFMEMB

Professional Body Membership

- 1) Member of Professional Body 2)
- Not a member of Professional Body

PREVEMP

Nature of Previous Employment

- 1) Local Government
- 2) Outside Local Government
- 3) Student
- 4) Unemployed
- 5) Self Employed
- 6) Youth Training Scheme
- 7) Retired

PREVEAST

Town of previous employment Eastings

PREVNOR

Town of previous employment Northings

APPLEAST

Home Town when applied Eastings

APPLNOR

Home Town when Applied Northings

OCCUPAT

Occupation (OPCS 1980 Classification)

- 1.00 Solicitor
- 2.10 Auditor
- 2.20 Accountant
- 2.50 Valuer
- 3.10 Personnel Officer

3.20	Work Study Officer	5.20	Advertising	Executive
4.20	Computer Programmer	6.10	E.H.	Officer
6.20	Building Inspector			
8.00	Admin Executive			
9.50	Legal Executive			
9.80	Curator (Museum)			
13.10	Warden (O.A.P.)			
13.20	Play Group Leader	13.30		
	Welfare Occupations			
	n.e.c.			
18.20	Vet			

22.20 Projectionist

25.00	Municipal Engineer	51.20	Switch Board Operator
31.10	Architect/Town Planner		
31.20	Quantity Surveyor	56.00	Meals on Wheels Operator
31.30	Building Surveyor		
33.10	Architect/Town Planner Technician	60.40	Estate Ranger
		60.60	Supervisor Security
		62.10	Park Keeper
33.20	Building/Engineering Technician	62.30	Art Gallery Attendant
33.40	Works Manager	63.30	Supervisor Bar
35.10	Maintenance Supervisor	63.40	Supervisor Catering
35.20	Clerk of Works	65.20	Bar person
36.20	Transport Manager	66.10	Counter Hand
36.30	Stores Controller	71.10	Supervisor Caretakers
37.20	Office Manager	71.40	Supervisor Car Parks
39.50	Entertainments/Sports Manager	72.10	Caretaker
	44.10 Caravan Site Manager	72.20	Cleaner
44.40	Managers n.e.c.	75.20	Car Park Attendant
45.20	Supervisor Stores Clerks	75.60	Service Worker n.e.c.
45.30	Supervisor Drawing Assistants	76.30	Foreman Gardeners
45.40	Supervisor Clerks	78.10	Horticulture Workers
45.50	Supervisor Cashiers	78.20	Gardener
46.10	Clerk Stores	83.00	Dog Warden
46.20	Tracer Assistant	100.30	Print Machine Operator
46.30	Clerk Non-retail		
47.00	Cashier Retail	105.10	Carpenter
48.20	Supervisor Machine Operators	109.30	Black Smith
49.10	Receptionist	111.00	Management Trainee
49.20	Typist	114.50	Foreman Fitters
50.00	Punch Card Operator	118.10	Fitter
51.10	Telephone Receptionist	125.00	Plumber
		133.40	Painter and Decorator
		139.10	Foreman Bricklayers

139.11	Foreman Workers	Sewage			manual
139.12	Foreman Workers	Construction	6.0	Junior Non	manual
139.80	Highways	Inspector	7.0	Personal Services	
140.10	Bricklayer		8.0	Foremen/Supervisors (manual)	
140.50	Building	Worker	9.0	Skilled	Manual
142.10	Sewage	Plant Attendant	10.0	Semi-skilled	Manual
152.20	Refuge Driver	Vehicle	11.0	Unskilled	Manual
152.30	Road Sweeper	Driver	15.0	Agricultural	Workers
156.10	Foreman	Storekeeper	156.40	CLASS	
157.10	Foreman	Refuse Collection		Social Class	
157.40	Refuse Collector			1) Professional	
159.80	Foreman	Labourers		2) Intermediate Non-manual	
160.80	Labourer			3) Junior Non-manual	
n.e.c.	EMPSTAT			4) Skilled Manual	
Employment Status on	application			5) Semi-skilled Manual	
1)	Manager			6) Unskilled Manual	
2)	Foreman /	Supervisor	3)	SALARY	
Apprentice /				Initial Annual Salary (Spinal Column Point)	
Trainee 4)	Employee			EMPLEAST	
n.e.c.				Town of Employment Eastings	
SEG				EMPLNOR	
Socio-economic	Group			Town of Employment (Northings)	
1.2	Managers in Government and Industry			HOME EAST	
4.0	Professional Employees			Home Town at start of	
5.1	Ancillary Workers			Employment	
5.2				(Eastings)	
s	Foremen/Supervisor			HOMENOR	
n	(no			Home Town at start of	
				Employment	
				(Northings)	

NOTIFY

- 1) Notification Outlet 1
- 2) Internal
- 3) Word of Mouth
- 4) Letter of Enquiry
- 5) Circular to adjacent Local
 Authorities
- 6) Job Centre
- 7) P.E.R. (Professional Executive
 Register)
- 8) Employment Agency
 (Private Sector)
- 9) Careers Office (local)

DAILY LOCAL NEWSPAPERS

- 10) Evening Post

WEEKLY LOCAL NEWSPAPERS

- 11) Two County Courier
- 12) Local Town Chronicle
- 13) County Messenger
- 14) County Express
- 15) Local Town News
- 16) East County Gazette

FREE WEEKLY LOCAL NEWSPAPERS

- 17) Local Town News in
 Focus
- 18) News in Focus 19) Local
 Town Times
- 20) Local Town Times and
 Gazette

NATIONAL DAILY

- NEWSPAPERS** 21) Daily Mail
- 22) Daily Telegraph
- 23) The Times
- 24) The Guardian
- NATIONAL SUNDAY NEWSPAPERS**
- 25) Sunday Times
- JOURNALS AND MAGAZINES (*PUBLIC SERVICE)**
- 26) Commercial Motor
- 27) Horticulture and Amenities Weekly
- 28) Army Appointments Service Magazine
- 29) Church Times
- 30) Parks and Recreation
- 31) Golf Illustrated
- 32) Groundsman
- 33) Health and Social Services Journal
- 34) Lady Magazine
- 35) Public Finance and Accounting*
- 36) Institute of Management Science Journal
- 37) UK Press Gazette
- 38) Building Trades Journal
- 39) Gardeners Chronicle
- 40) Solicitors Journal
- 41) Nursing Times
- 42) Nursing Mirror
- 43) Policy Holder Insurance Journal
- 44) Opportunities*
- 45) Estates Gazette
- 46) Clerk of Works Weekly
- 47) Computing
- 48) Association of Recreation Managers Appointments
- 49) Local Government Chronicle*
- 50) Municipal Journal*
- 51) Public Service and Local Government*
- 52) Municipal Engineer*
- 53) Baths Service Circular
- 54) Caterer and Hotel Supervisor
- 55) Accountancy Age
- 56) Management (Work Study) Services
- 57) Quantity Surveyors Weekly
- 58) Surveyor
- 59) Architects Journal
- 60) The Planner
- 61) Building
- 62) Planning
- 63) Law Society Gazette 64) Meat Trades Journal
- 65) Protection
- 66) Institute of Personnel Managers Digest
- 67) New Civil Engineer
- 68) Veterinary Record 69) Current Vacancies
- 70) The Stage
- 71) Regional Arts Magazine
- 72) Health and Safety at Work

NOTIFY2

Notification Outlet as for
 Notification Outlet as for
 NOTIFY1

NOTIFY4

Notification Outlet as for
 NOTIFY1

NOTIFY5

Notification Outlet as for
 NOTIFY1

NOTIFY6

Notification Outlet as for
 NOTIFY1

NOTIFY1**NOTIFY3****NOTIFY7**

Notification Outlet as for
 NOTIFY1

NOTIFY8

Notification Outlet as for
 NOTIFY1

NOTIFY9

Notification Outlet as for
 NOTIFY1

NOTIFY10

Notification Outlet as for
 NOTIFY1

YEAREMP

Year of Employment
 19-- Year of Employment

ASSISTAC

Assistance by Accommodation
 Provision

1) Permanent Accommodation
 Provided

2) Temporary Accommodation
 Provided

3) No assistance given

ASSISTMO

Assistance with Moving
 Expenses

1) Financial Assistance
 Given

2) No Assistance Given

19-- Year of termination

TERMWHY

1) Reason for Termination

2) New Job

3) Spouse's New Job

4) Going to College

TERMYEAR

Year of Termination of
Employment

- 5) Having a Baby
- 6) Early Retirement
- 7) Retirement
- 8) Redundancy
- 9) Dismissed
- 10) Death
- 11) Illness
- 12) Marriage
- 13) Family Commitments
- 14) First Job Commitments
- 15) Leaving the Country
- 16) Travel Costs
- 17) Fixed Term Contract
- 18) Other

FULLPART

Full or Part Time
Employment

- 1) Part time
- 2) Part time (second job)
- 3) Full time

HEARDN1

NOTIFY1 is how first
heard of vacancy

- 1) Yes
- 2) No

The following additional

variables have been created
using IBM SPSS Statistics

INTERNAL

Notified internally

1) used

2) not used

CIRCULAR

Notified by circulating other

Local

Authorities

1) used

2) not used

JOBCENTR

Notified at Job Centre

1) used

2) not used

LOCAL

Notified in Local Press

1) used

2) not used

REGIONAL

Notified in Regional Media

1) used

2) not used

NATIONAL

Notified in National Media

1) used

2) not used

DISTMIG

Distance Migrated (km)

-- straightline distance
migrated

JTOWORK

Journey to work (km) -new
job

-- straightline distance of
journey