

Analysing your Questionnaire using SPSS (v. 18)

SPSS (originally, Statistical Package for the Social Sciences) is a computer program used for statistical analysis. Between 2009 and 2010 the premier vendor for SPSS was called PASW (Predictive Analytics SoftWare) Statistics, but is now owned by IBM and the name has reverted to SPSS. This handout is an introductory guide to the *use of the software* and it describes how to conduct some of the basic statistical tests. It is not a guide to the selection or interpretation of the statistical tests; this needs to be discussed and agreed with your supervisor.

SPSS step-by-step

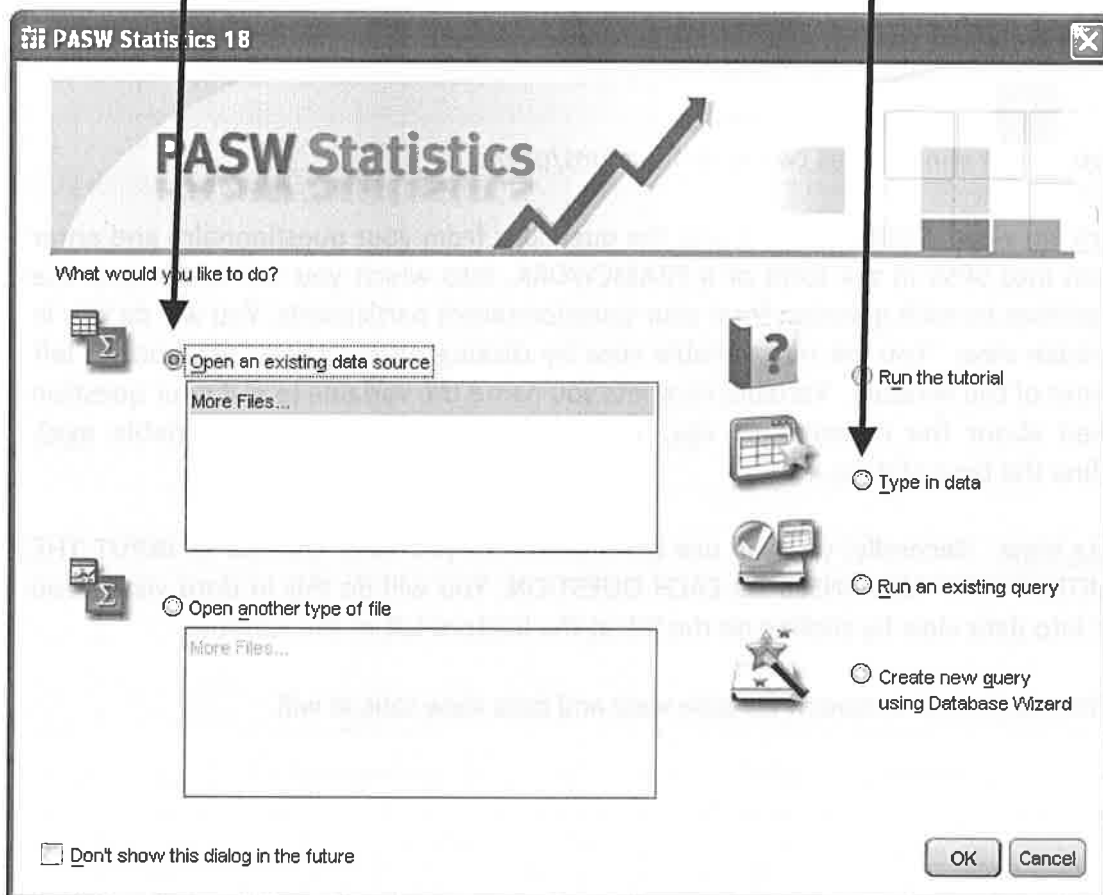
1. Select (left-click on) **Start | All Programs | SPSS Inc | PASW Statistics 18.**

SPSS will load and a dialogue box will appear, asking 'What would you like to do?'

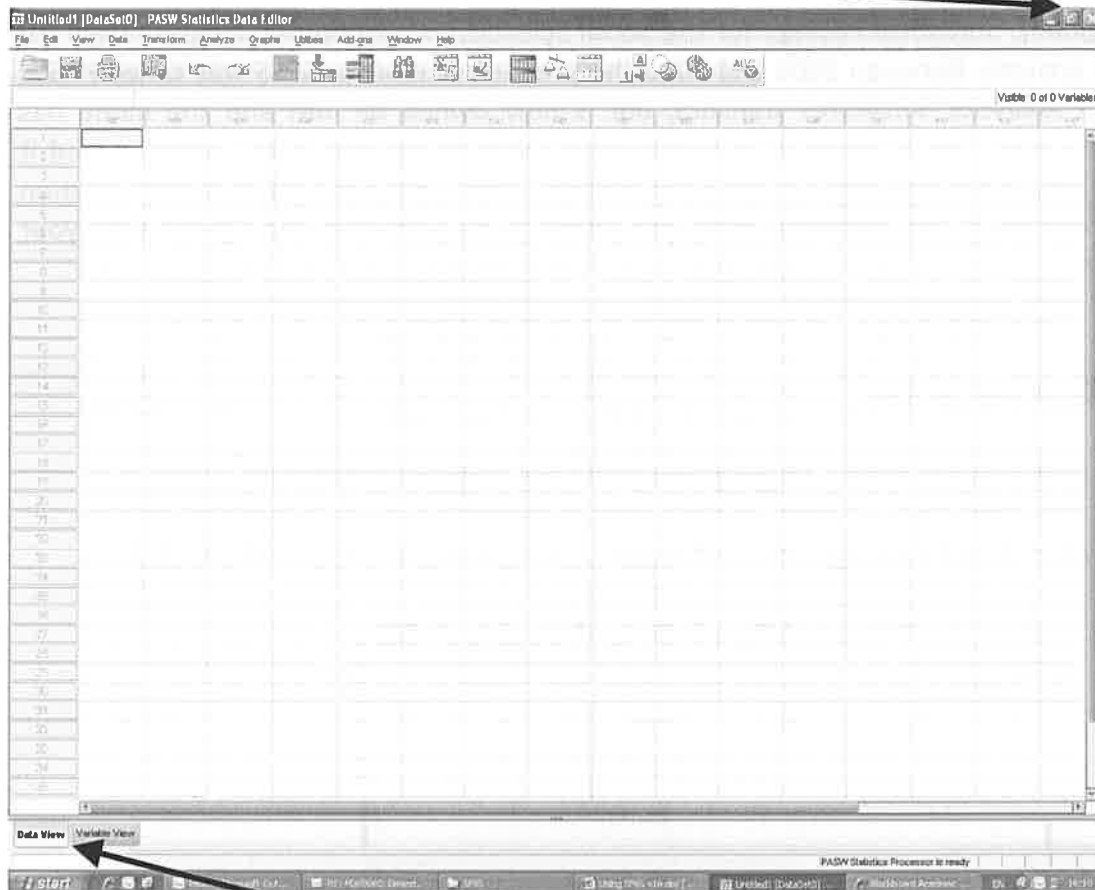
- a) To create a new file: select **Type in Data**, then click on **OK**.

OR

- b) To open a previously created file: select **Open an existing data source**, then **OK**.



2. The *Data Editor* window will appear. Maximise the window, if necessary, by left-clicking on the maximise icon in the upper right hand corner of the window.



3. The Data Editor window has two main functions/pages.

- a) Variable view. Firstly, you will take the questions from your questionnaire and enter them into SPSS in the form of a FRAMEWORK, into which you will *later* input the responses to each question from your questionnaire's participants. You will do this in *variable view*. You get into variable view by clicking on the tab at the bottom left corner of the window. Variable view lets you name the variable (e.g. if your question asked about the respondents age, you would probably name the variable **age**), define the type of data, etc.
- b) Data view. Secondly, you will use the framework you have created to INPUT THE PARTICIPANTS' RESPONSES TO EACH QUESTION. You will do this in *data view*. You get into data view by clicking on the tab at the bottom left of the window.

Note: You can *switch* between variable view and data view tabs at will.

4. Select the **Data View** tab. You will see that the data view is made up of a number of columns and rows.

- a) Each *column* represents a single variable, usually one question from your questionnaire.
- b) Each *row* represents the answers to your questionnaire from one respondent. In other words, row 1 gives the responses to each question in your questionnaire from respondent 1 (each question response being placed in a new column), row 2 gives the responses from respondent 2, etc.

Column (question)

Row (answer)

5. Select the **Variable View** tab. This view lets you define your variables. Each row represents a variable (or question). Each column sets the definition of the variable.

VARIABLE VIEW: **creating a framework (or structure) for later data input**

Assume that the following is an extract from your questionnaire. Having already designed the questionnaire, you now need to create the framework in SPSS that will allow you to input respondents' answers to the questions. So, we must first enter the questions themselves in Variable View (sometimes called design view). Later we will switch to Data View to input the participant responses.

Each completed questionnaire represents all of the different answers to different questions (variables) from one respondent (case)

Although question three seems to be one question, it is in fact five separate questions

Teacher Stress Questionnaire

In each question please TICK or CIRCLE your preferred response.

QUESTION ONE

Gender

Male ☐1
Female ☐2

QUESTION TWO

Age

18-25 ☐1
26-35 ☐2
36-45 ☐3
46-55 ☐4
55+ ☐5

QUESTION THREE

For each of the following items please circle the number appropriate to the degree of stress you experience.

	No stress	Mild stress	Appreciable stress	Extreme stress
a) Noisy pupils	1	2	3	4
b) Difficult classes	1	2	3	4
c) Pupils' poor attitude to work	1	2	3	4
d) Maintaining classroom discipline	1	2	3	4
e) Pupils range of ability too wide	1	2	3	4

Each question represents a variable – in this instance, it's Gender

Each multiple-choice question has a limited number of possible responses. We can add codes (1,2,3,4, etc.) to the responses to simplify data entry

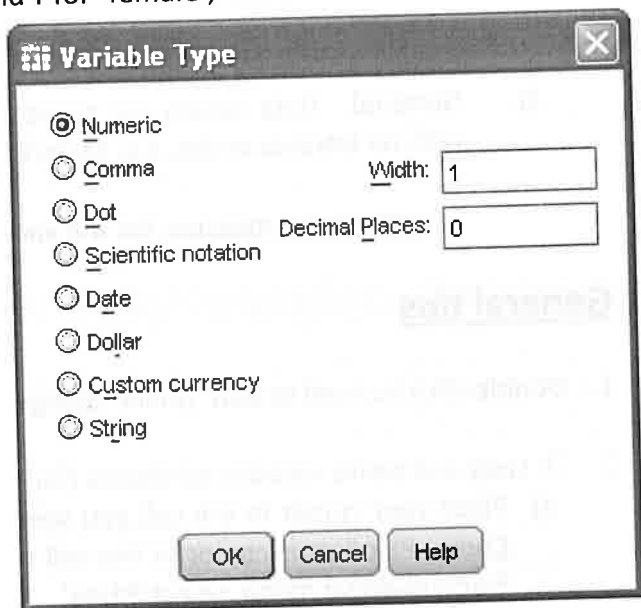
Make sure you are in *variable view*. Remember, that when we are in *variable view* each variable is represented by a row. So row 1 will represent question 1, row 2 will represent question 2, etc.

- Our first variable will be QUESTION 1, which relates to the gender of the respondents. Firstly, we need to name the variable. Place your cursor in the cell that you want to define. In this case it will be row 1, column heading 'name'. Type in **Gender**.

2. Next we need to specify the variable type. This means the format in which you will input the respondent's answer (data). For this question (gender), there can only be two responses, that is, 'male' or female'. However, it would be time- and labour-intensive to type the responses in full. Instead we will assign a code for each response category. We could choose to assign, say, m for 'male', and f for 'female', but in this case, we will use the number 1 for 'male' and the number 2 for 'female'. Therefore our data input will be in the form of a number (rather than alphanumeric letters, or 'string').

a) Place your cursor in the cell in row 1 under the column heading, *Type*. Click on the button that appears in the cell. Select the data type in the *Define Variable Type* dialogue box. For this variable, the data type will be **Numeric**.

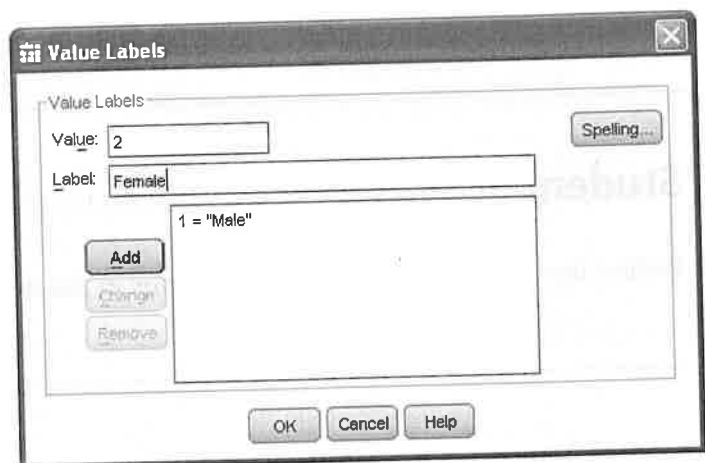
b) Set the width to **1** character, and the decimal places to **0**. Click on **OK**.



3. We will now define descriptive variable and value labels for data values. This is important if, at a later stage, you will want to generate graphs or charts from your data. Although we have already named the variable, we were restricted to a limited number of characters, which sometimes does not fully describe the variable. Value labels represent the different categories from which the respondent could choose, e.g. male or female. Unless we assign value labels, our charts will be very difficult, or even impossible, to interpret.

a) Place your cursor in the cell in row 1 under the column heading, *Label*. Type in **Gender**. If we wished, we could be more descriptive, e.g. 'Gender of respondents', but in this case, 'Gender' is adequate.

b) Place your cursor in the cell in row 1 under the column heading, *Values*. Click on the button that appears in the cell. Enter the data *value* as **1**, then enter the *value label* **Male**. Click on **Add**. Repeat the process for all other categories available. So enter the data value **2**, then enter the descriptive label **Female**. Click on **Add**. When you have finished, click on **OK**.



4. Next we need to define how to measure the data. Place your cursor in the cell in row 1 under the column heading, *Measure*. Click on the button that appears in the cell. You may select one of three measurement levels:
- Scale.** Data values are numeric on an interval or ratio scale (e.g. age, income).
 - Ordinal.** Data values are numeric or alphanumeric, and represent categories with some intrinsic order (e.g. strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).
 - Nominal.** Data values are numeric or alphanumeric, and represent categories with no intrinsic order, e.g. Male/Female, Yes/ No.

As we are setting the measure for the variable (question) *gender*, select **Nominal**.

General tips

1. Double-click in a cell to **edit** (make changes).
2. To **copy** and **paste** variable attributes (for one variable definition/element only).
 - a) Place your cursor in the cell you wish to copy. In the Edit pull-down menu, select **C**opy. Place your cursor in the cell where you wish to paste the attributes, In the Edit pull-down menu, select **P**aste.
OR
 - b) Point your cursor at the cell you wish to copy and right-click. Select **C**opy from the shortcut menu. Point your cursor at the cell where you wish to paste the attributes and right-click. Select **P**aste from the shortcut menu.
3. To copy *all* the attributes from one variable to another:
 - a) Click on the row number of the variable you want to copy. In the Edit pull-down menu, select **C**opy. Click the row of the variable to which you want to paste the attributes. In the Edit pull-down menu, select **P**aste.
OR
 - b) Point your cursor at the row you wish to copy and right-click. Select **C**opy from the shortcut menu. Point your cursor at the row where you wish to paste the attributes and right-click. Select **P**aste from the shortcut menu.
4. If necessary, **save** your work by clicking on the 'save' icon on the toolbar.

Student activity

Define the variables for all other questions in the questionnaire above.

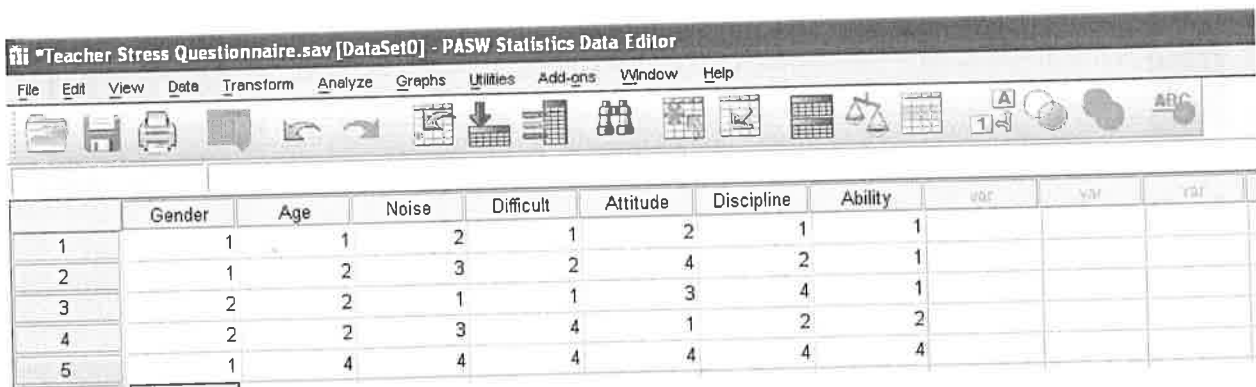
DATA VIEW:

Entering data (i.e. respondents' answers to your questions)

Select the *Data View* tab. Remember, that in data view, each **row** represents the answers to your questionnaire from one respondent.

There are **two** main ways to enter data:

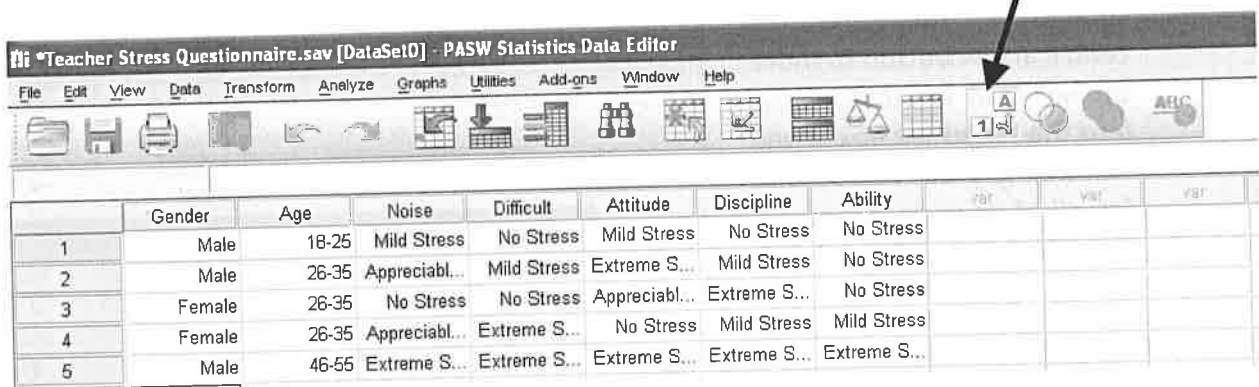
1. Simply type in the code that the respondent has circled for each question within the questionnaire, i.e. in response to question 1 (gender), type in either 1 or 2.



	Gender	Age	Noise	Difficult	Attitude	Discipline	Ability	YBI	YBI	YBI
1	1	1	2	1	2	1	1			
2	1	2	3	2	4	2	1			
3	2	2	1	1	3	4	1			
4	2	2	3	4	1	2	2			
5	1	4	4	4	4	4	4			

Or,

2. Select the appropriate response from the pull-down list in each cell. Note, if value labels are not currently displayed in the *Data Editor* window, select the *Value labels* icon on the toolbar (or select *Value labels* from the *View* pull-down menu).

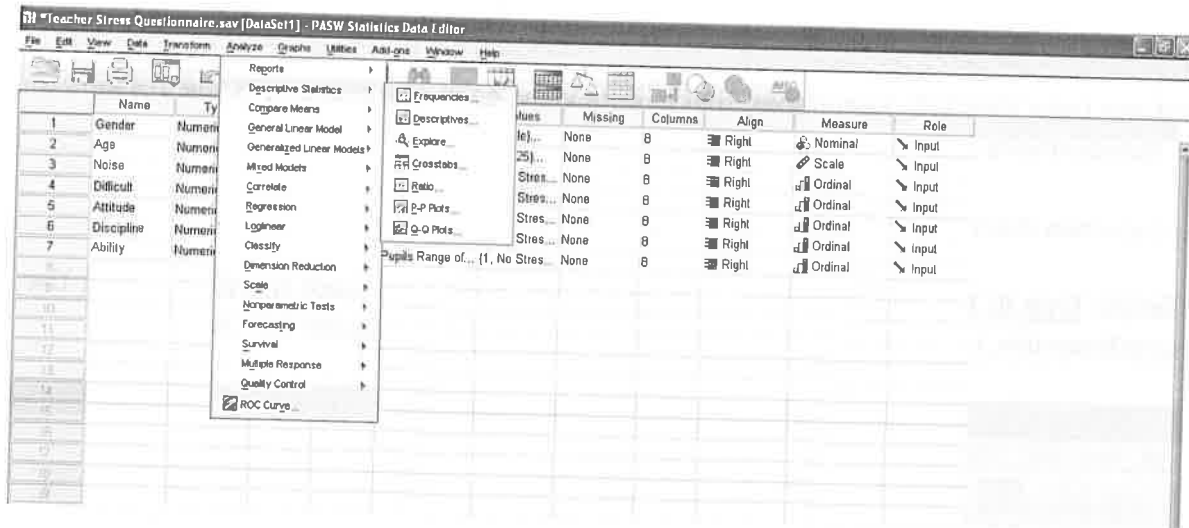


	Gender	Age	Noise	Difficult	Attitude	Discipline	Ability	YBI	YBI	YBI
1	Male	18-25	Mild Stress	No Stress	Mild Stress	No Stress	No Stress			
2	Male	26-35	Appreciabl...	Mild Stress	Extreme S...	Mild Stress	No Stress			
3	Female	26-35	No Stress	No Stress	Appreciabl...	Extreme S...	No Stress			
4	Female	26-35	Appreciabl...	Extreme S...	No Stress	Mild Stress	Mild Stress			
5	Male	46-55	Extreme S...	Extreme S...	Extreme S...	Extreme S...	Extreme S...			

Student activity

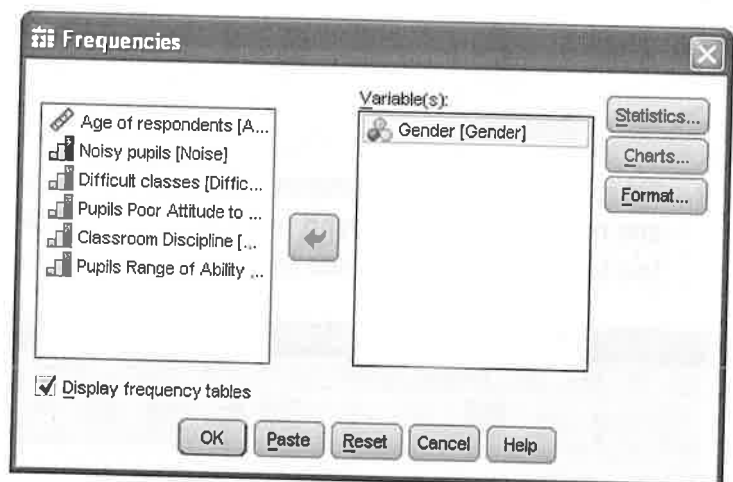
Enter the responses to your questionnaire from five respondents (make them up, it's just for practice).

Using SPSS to analyse data already entered

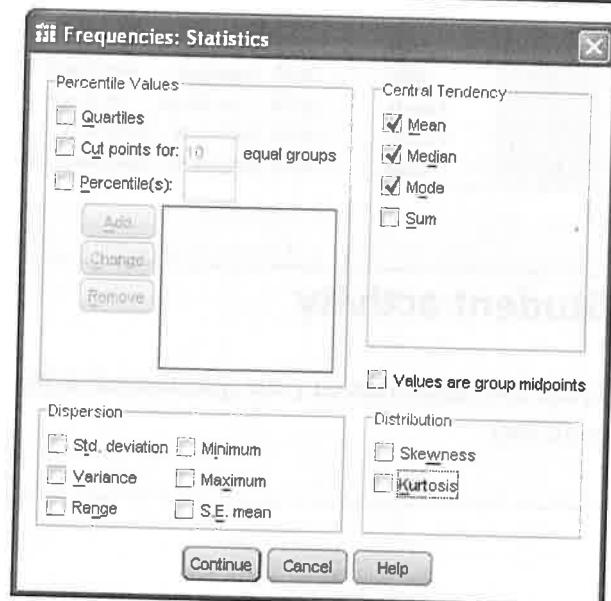


1. Select **Analyze** from the drop down menu bar.
2. Select **Descriptive Statistics**.

3. Select **Frequencies** (to obtain a count/tally of your data and/or to obtain the mean, median or mode).



1. Select the variable(s) that you want to analyse from the source list then use the centre arrow button to move the selected variables across into the target list box. If you want to get the frequencies for *all* of your questions, hold down the shift button, then click on each of the variables that you want (to highlight them, just as you do in Microsoft Word) – then use the arrow button to move them across.



2. Click on the **Statistics** button at the top right of the dialogue box. Choose the procedure you wish to run, e.g. **mean, median** and/or **mode**, then click the '**Continue**' button. Click on '**OK**' to run the procedure. The results (output) are displayed in the viewer (and reproduced below).

Frequencies

Statistics

Gender

N	Valid	5
	Missing	0
Mean		1.40
Median		1.00
Mode		1

This first *output table* shows us the statistics for the Variable **Gender**. The results show that there were 5 respondents (of course, we are just practicing – in reality you may have hundreds of respondents).

We chose to run tests for the mean, median and mode so they are also specified in the table (all measures of average, otherwise known as measures of central tendency). This is a good opportunity for us to consider if the test is appropriate. For example, would we be interested in the *mean (average) gender*? In reality, a

mean of 1.4 is somewhere between male (1) and female (2)...it's meaningless because gender is nominal data (the number is just a code for data entry or label). The table below indicates when to use the mean, median and mode

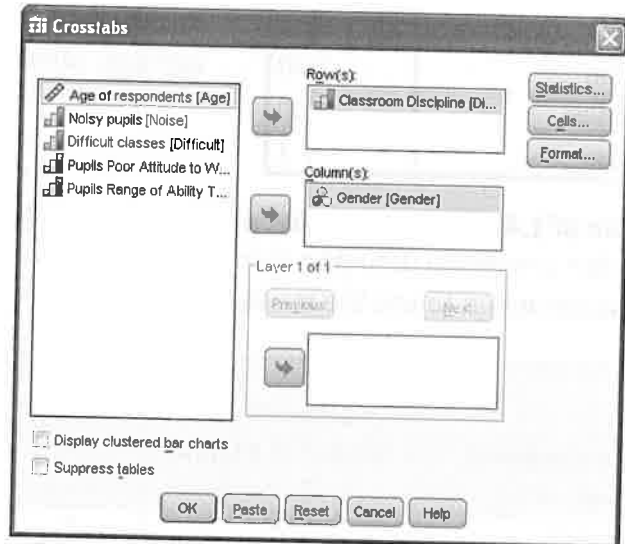
Measures of Central Tendency				
Adapted from http://hsc.uwe.ac.uk/dataanalysis/quantDescCentral.asp				
Measurement Scale	Mean	Median	Mode	Best measure of the 'middle'
Nominal (categorical)	×	×	✓	Mode
Ordinal	×	✓	✓	Median
Scale (interval or ratio)	✓	✓	✓	Symmetrical data (normal distribution): Mean Skewed data: median

The *frequency table* below shows us the number of respondents (absolute frequencies) who replied that they were male (3) or female (2). These frequencies are also shown in percentage format (relative frequencies) as it's easier to understand and compare numbers when using percentages (relative to each other).

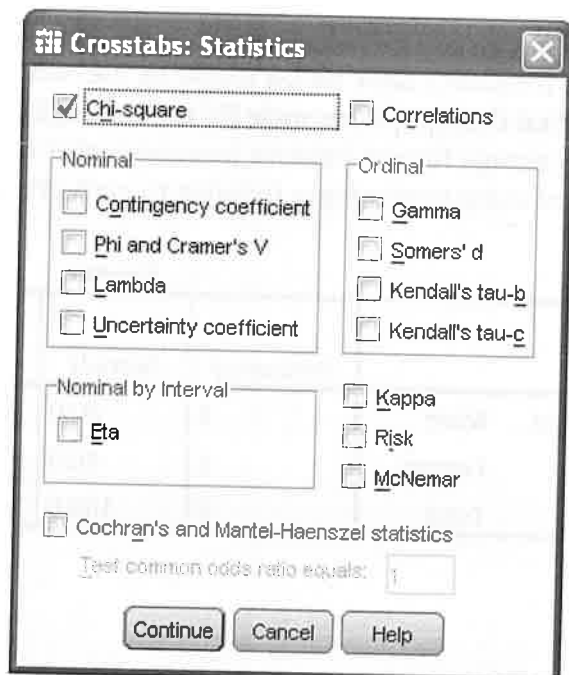
Gender		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	3	60.0	60.0	60.0
	Female	2	40.0	40.0	100.0
	Total	5	100.0	100.0	

4. A common question is whether the behaviour of a variable differs between groups within the population. For example, in terms of the variable *Classroom discipline* you might want to identify if males answer differently to females (using the *Gender* variable). One of the simplest ways of linking two variables for comparison is to get SPSS to run a cross-tabulation (crosstab), which produces responses as a grid/matrix. To do this, select **Analyse | Descriptive Statistics | Crosstabs**.

- a) Select the variable(s) that you want to analyse from the source list – use the arrow button to move the selected variables into the target row and column list boxes. It is conventional to take the *dependant* variable as the **row** and the *independent* variable as the **column**. The independent variable is that which we expect to be influencing the other. The dependent variable varies with, or depends upon, the independent variable. In this case, classroom discipline cannot influence a person's gender, but gender can influence classroom discipline. Thus, classroom discipline is the dependent variable and is taken as the row in the table.



- b) If you want to run a statistical test (on nominal or ordinal data) to try to identify if there is a significant relationship between variables, click on the **Statistics** button at the top, right of the box. Select the **Chi-square** option then click on the **Continue** button. The results from the chi-square independence test will tell us if the relationship between gender and classroom discipline has arisen by chance or if there is a statistical relationship between the two variables.



3. If you wish your results to also be displayed as percentages, click on the **Cells** button at top right of the crosstabs box. Select row and column percentages by ticking in the appropriate boxes, then click on the **Continue** button. Click on 'OK' to run the procedure. The results are displayed in the viewer (and reproduced below).

REMEMBER THAT THE HELP MENU IS AVAILABLE AT ALL TIMES, AS ARE TUTORIALS FOR PARTICULAR FUNCTIONS.

Crosstabs: Cell Display

Counts

☒ Observed
☐ Expected

Percentages

☒ Row
☒ Column
☒ Total

Residuals

☐ Unstandardized
☐ Standardized
☐ Adjusted standardized

Noninteger Weights

☒ Round cell counts ☐ Round case weights
☐ Truncate cell counts ☐ Truncate case weights
☐ No adjustments

Continue Cancel Help

Classroom Discipline * Gender Crosstabulation

			Gender		Total
			Male	Female	
Classroom Discipline	No Stress	Count	1	0	1
		% within Classroom Discipline	100.0%	.0%	100.0%
		% within Gender	33.3%	.0%	20.0%
		% of Total	20.0%	.0%	20.0%
	Mild Stress	Count	1	1	2
		% within Classroom Discipline	50.0%	50.0%	100.0%
		% within Gender	33.3%	50.0%	40.0%
		% of Total	20.0%	20.0%	40.0%
	Extreme Stress	Count	1	1	2
		% within Classroom Discipline	50.0%	50.0%	100.0%
		% within Gender	33.3%	50.0%	40.0%
		% of Total	20.0%	20.0%	40.0%
Total	Count		3	2	5
	% within Classroom Discipline		60.0%	40.0%	100.0%
	% within Gender		100.0%	100.0%	100.0%
	% of Total		60.0%	40.0%	100.0%

The chi-square significance test in the far-right column measures the likelihood that the observed association between the independent variable (in this case, gender) and the dependent variable (in this case, classroom discipline) is caused by chance. Among statisticians a chi square of .05 is a conventionally accepted threshold of statistical significance; values of less than .05 are commonly referred to as 'statistically significant'. In practical terms, a chi square of less than .05 means that if, in fact, there was no association in the population between the independent and dependent variables, the observed association would be expected to occur by chance less than 5 times in 100 samples of the type we used. Thus, when the chi-square is less than .05, we can be confident in rejecting the possibility that no association exists between the independent and dependent variables. As the chi-square increases above .05 the likelihood that the observed association occurred by chance increases.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.833 ^a	2	.659
Likelihood Ratio	1.185	2	.553
Linear-by-Linear Association	.296	1	.586
N of Valid Cases	5		

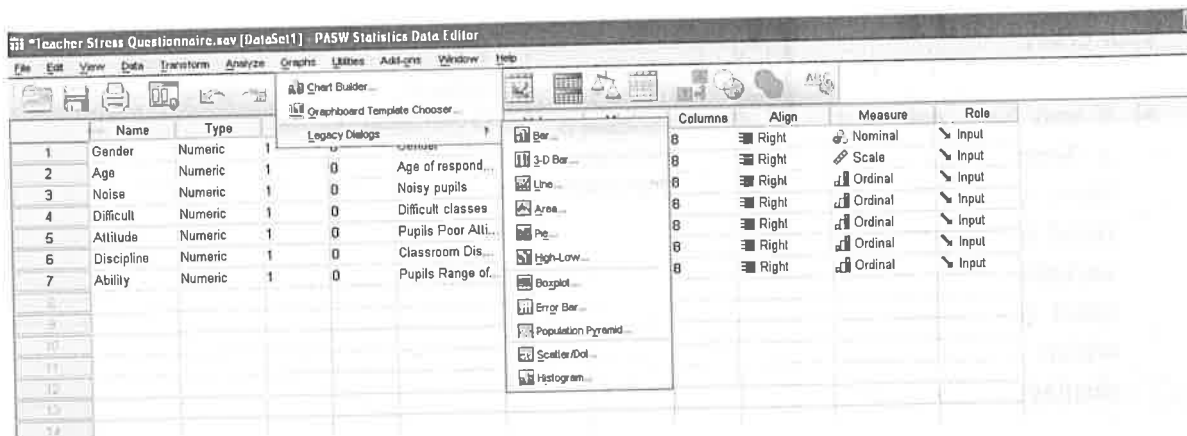
a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is .40.

However, the chi-square test can only be used if the following assumptions are met:

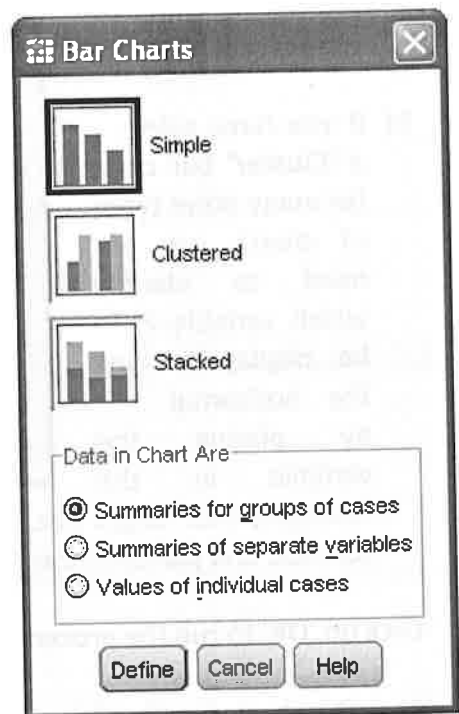
- All expected frequencies at least equal 1, and
- No more than 20% of the expected frequencies are lower than 5.

Look at the last two lines of the output table and check whether any of the cells have an expected count less than 5. In this case, all cells have expected counts less than 5 so the test result is unreliable.

Creating Charts/Graphs



1. Select **Graphs | Legacy Dialogs** from the drop down menu.
2. Select the type of chart that you want, e.g. **Bar**.
3. The Bar Charts dialogue box appears. Choose the icon for the specific type of chart that you want, e.g. **Simple** or **Clustered**.
4. You also need to indicate how your data are organised – usually **summaries for groups of cases**. Click on the **Define** button.

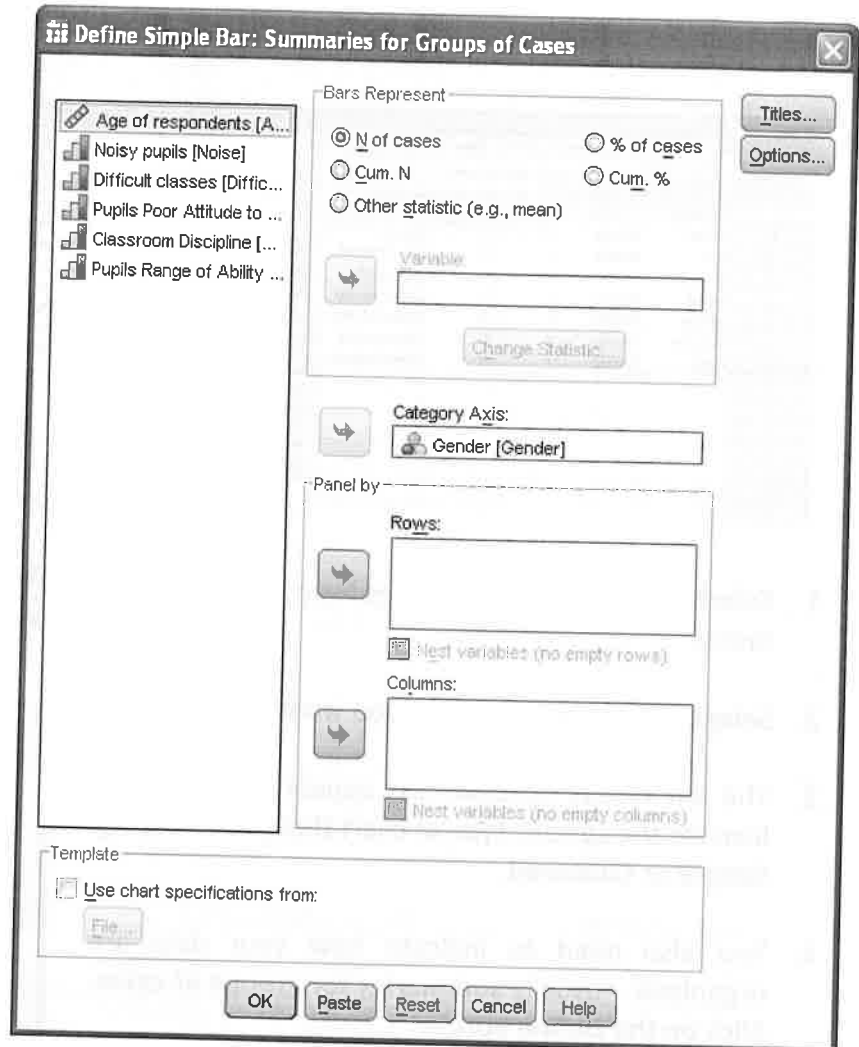


5. You then need to identify which variables you wish to display in your chart.

a) If you have selected a **'Simple'** bar chart then you will only need to identify one variable (the value label (i.e. name) of which is then displayed along the horizontal axis, with the values rising on the vertical axis).

b) If you have selected a **'Cluster'** bar chart (or many other types of chart), you will need to identify which variable is to be displayed along the horizontal axis by placing the variable in the

'Category axis' target box. The variable to be displayed on the vertical axis is then selected and placed in the 'Define clusters by' target box.



6. Click on **'OK'** to run the procedure. The results are displayed in the viewer.

7. To edit a chart, double-click on the chart as you see it in the viewer. To select particular objects you want to change on the chart, just click on them.

REMEMBER THAT THE HELP MENU IS AVAILABLE AT ALL TIMES, AS ARE TUTORIALS FOR PARTICULAR FUNCTIONS.