

School of Medicine

GUIDE TO USING SPSS (IBM SPSS version 24)

Contents

1.	Starting out.....	3
1.1	What is SPSS?	3
1.2	The Data Editor window	3
1.3	Entering your own data into SPSS.....	4
1.4	Opening an existing data file	7
1.5	Viewing data.....	8
1.6	Carrying out procedures: the dialog box	9
1.7	The Output (Viewer) window	10
2.	Exploring variables.....	11
2.1	Exploring categorical variables	11
2.1.1	Frequency tables	11
2.1.2	Bar charts	13
2.1.3	Pie charts.....	14
2.2	Exploring continuous variables	16
2.2.1	Summary statistics.....	16
2.2.2	Histogram	18
2.2.3	Box plot.....	21
2.2.4	Stem and leaf plot.....	22
2.3	Advanced graphs	23
3.	Creating new variables.....	24
3.1	Compute command	24
3.2	Recode command	25
3.3	Visual binning command	27
4.	Looking at different subgroups	31
4.1	Using the Select Cases command.....	31
4.2	Exploring a categorical variable in different groups.....	32
4.2.1	Crosstabs command.....	32
4.2.2	Clustered bar chart	34
4.3	Exploring a continuous variable in different groups.....	36
5.	Saving, editing and printing	37
5.1	Saving data	37
5.2	Editing items in output window	38
5.3	Saving output	38
5.4	Printing output from SPSS.....	39
5.5	Copying output to Microsoft Word.....	39
5.6	Exiting SPSS	41

1. STARTING OUT

1.1 What is SPSS?

SPSS is a windows based computer package used for statistical analyses. This guide is for version 24 of the software, although other versions are fairly similar. SPSS has menus and is therefore user friendly.

To run SPSS on University Computer Room and office desktops, go to:
Start > All Programs > UoN Applications > Statistical & Mathematical > SPSS

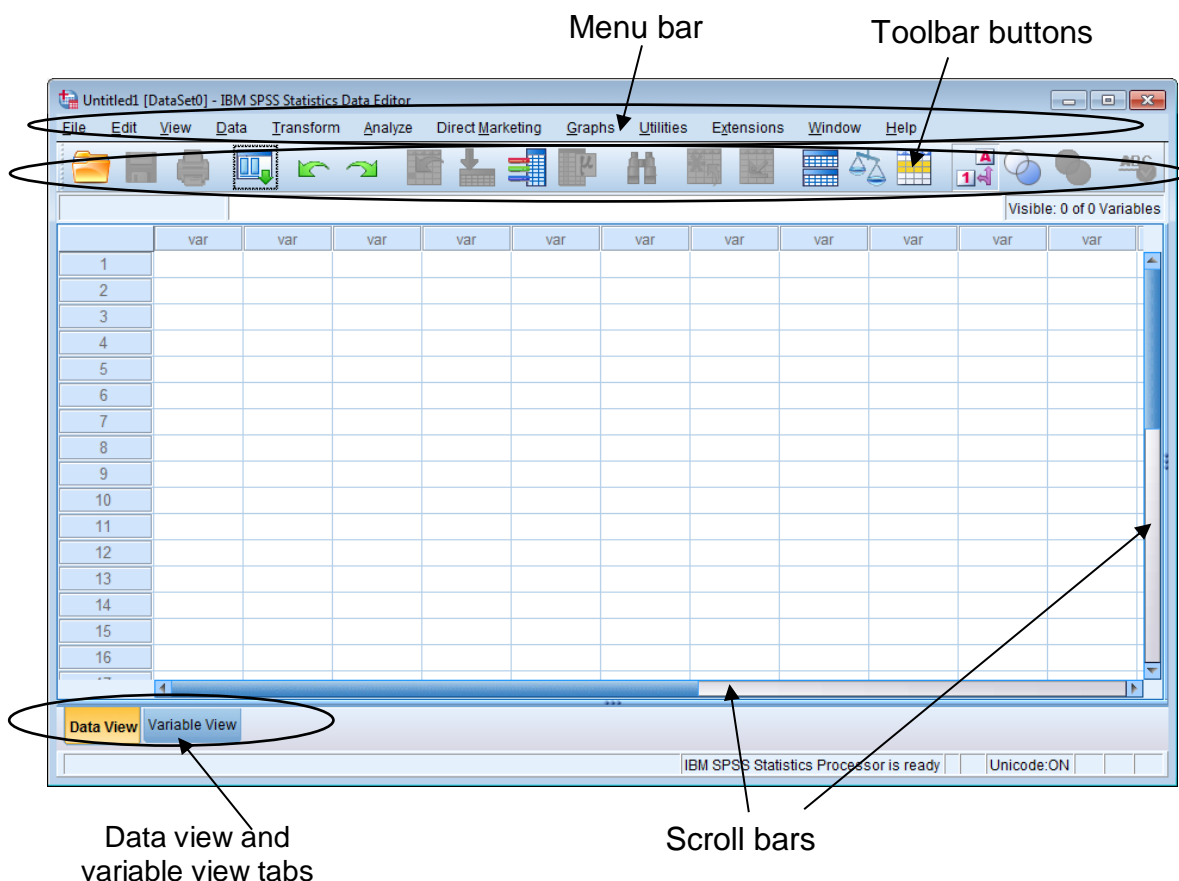
SPSS has two main windows: the Data Editor window and the Output window.

1.2 The Data Editor window

When you first start the SPSS package, a **Data Editor** window opens. This displays the contents of the working data file. This window has two different views: Data View and Variable View and you can switch between the two using the tabs at the bottom.

- **Data View** displays the actual raw data in spreadsheet format.
- **Variable View** displays a list of the variables in the dataset.

Initially the Data Editor is empty ('Untitled1' appears as the file name at the top).



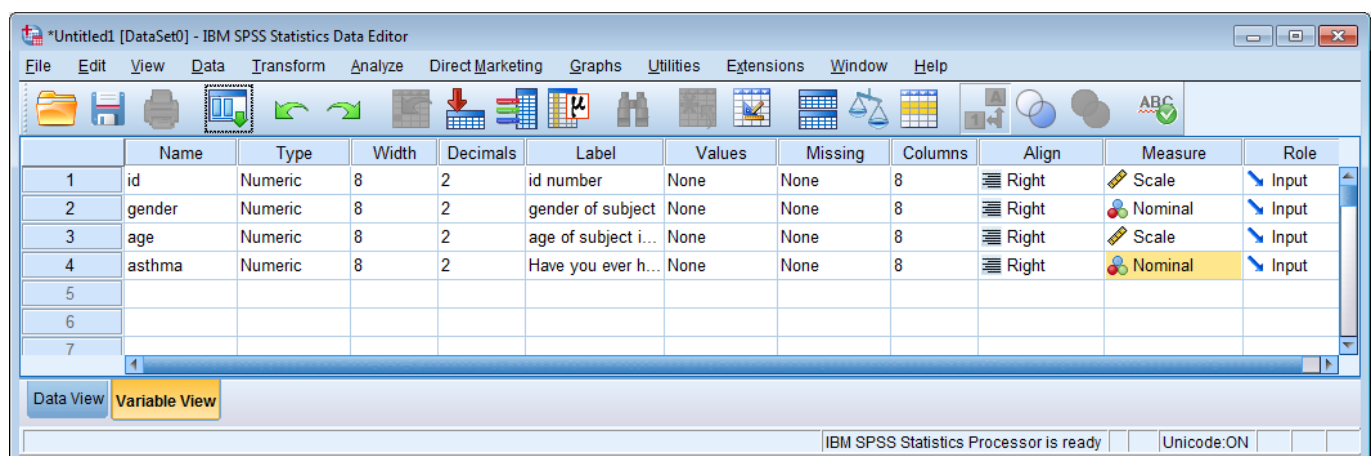
To carry out procedures in Data View, use the **menus** along the top, or the **toolbar buttons** which are shortcuts of the most popular commands. Position the mouse over any toolbar button for a description of the command.

1.3 Entering your own data into SPSS

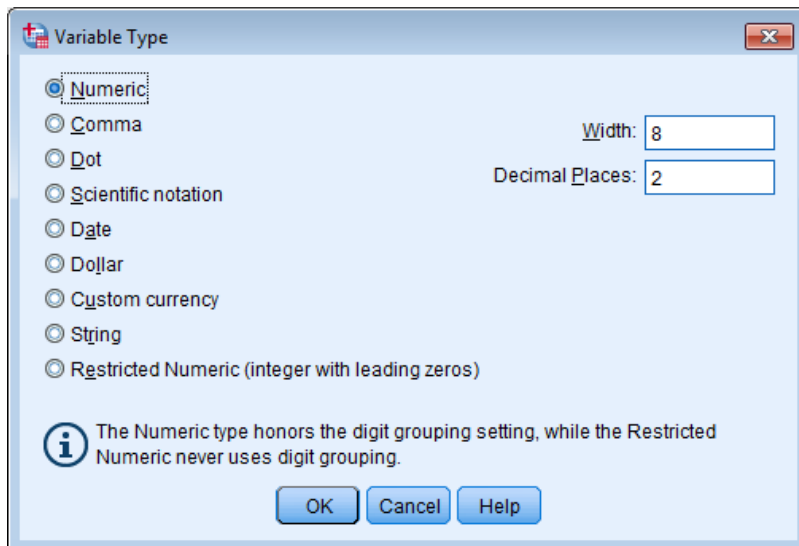
If you have data that have not yet been entered into a package, you can choose to enter your data directly into the data editor in SPSS. Data in SPSS are generally entered such that each row is an individual or subject and each column is a variable.

Step 1: Defining each variable

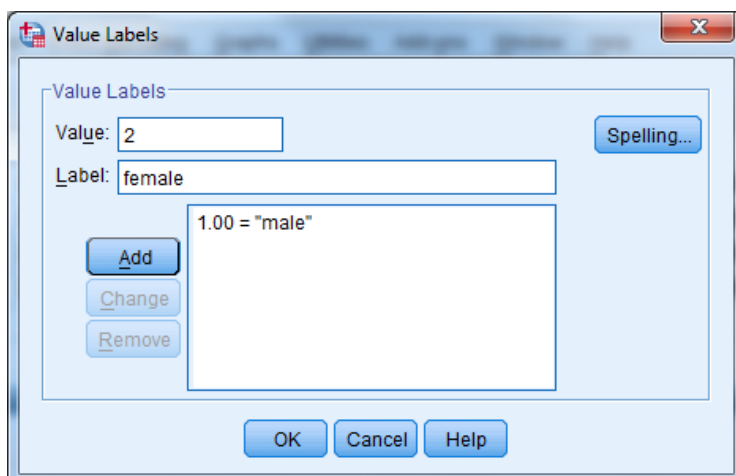
Before entering the data into the Data View spreadsheet, you need to **define** the variables. Go to Variable View. Here each row represents a variable. For each of your variables, complete the boxes:



- **Name:** in this box give the variable a name (this must be no more than 8 characters with no spaces and not starting with a number), eg gender.
- **Type:** most variables will be numeric (the actual value if continuous, or a code if categorical). If you have free text or something that cannot be coded eg name, this will be a string variable, or you may have variable that is a date. To change, click on the box, and a grey button with 3 dots on it will appear. Click on this, choose the type of variable you want to define it as, and then press OK.



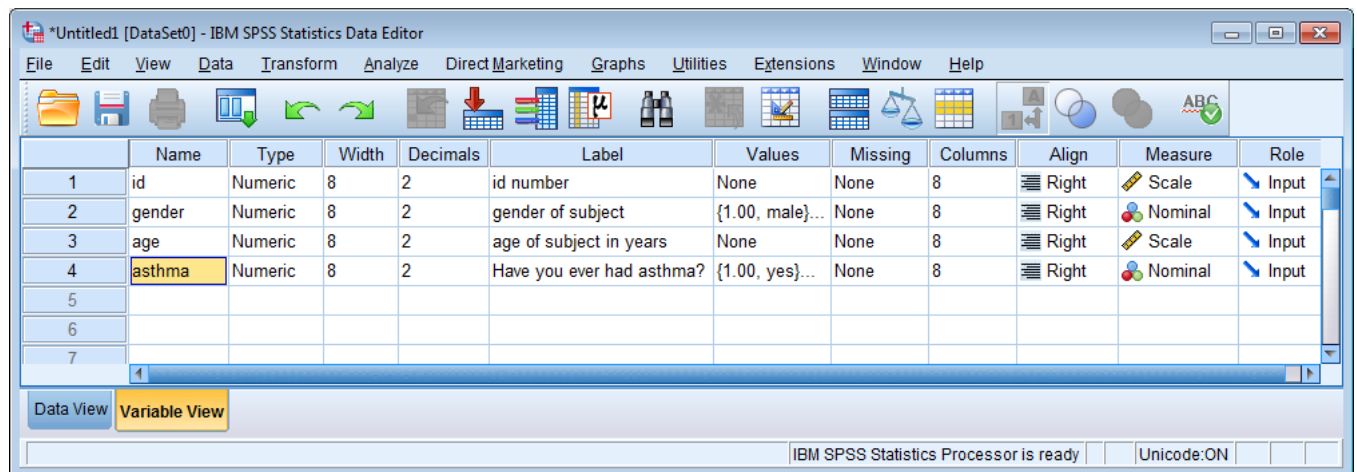
- **Width:** defines the actual width of the variable (ie number of characters). String variables usually need more than numeric variables. Use the arrows to change (or change using the variable type window above).
- **Decimals:** if your variable has any decimal places, this can be changed from 0. Again use the arrows to change, or change in the variables type window.
- **Label:** type in a longer description of the variable here.
- **Values:** categorical variables only. Values of these variables are entered as codes so you need to define what the codes represent (eg for gender 1=male, 2=female). Click on the grey box and type in the value (eg '1') and the value label (eg 'male') and click *add*. Repeat for all categories of the variable and press *continue* when finished.



- **Missing:** if you have a code that represents missing values (eg 999), it is important to define such entries as missing so that they are treated as such in the analyses. Click on the grey box and put in the value(s) under discrete missing values.
- **Column:** how wide the column is as it appears in data view.
- **Align:** to set whether entries are right, left or centre aligned.

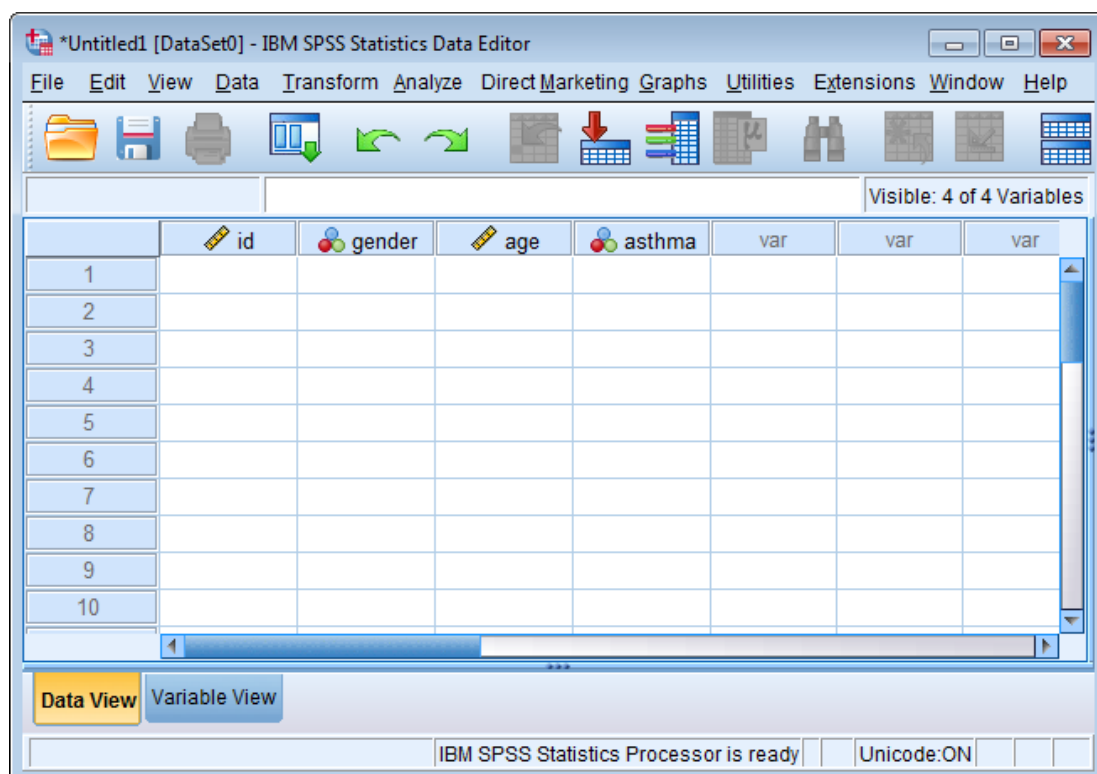
- **Measure:** to define the level of measurement of the variable as either scale (numeric values on a continuous scale), ordinal (ordered categories) or nominal (unordered categories).
- **Role:** you can define variables here according to their role: input (if exposure or independent variable), target (if outcome or dependent variable), both (if will be treating as both) or none. Input is the default. Generally you do not need to worry about this field as not many SPSS procedures use it.

Repeat for all your variables:

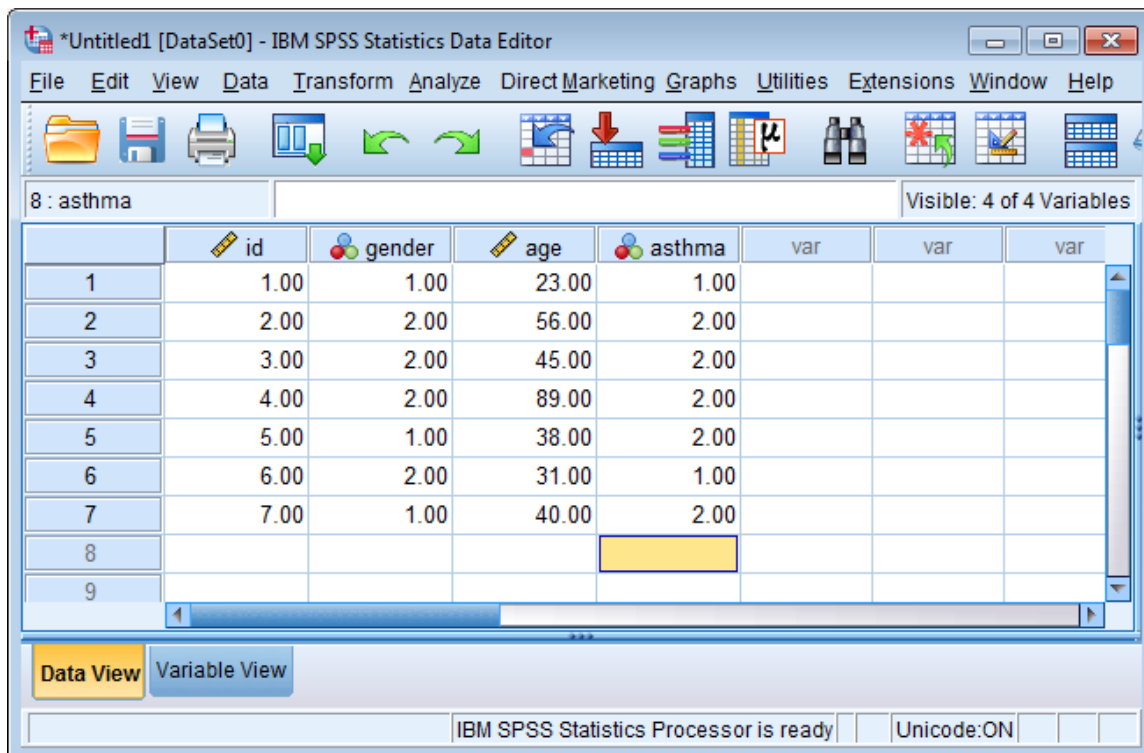


Step 2: entering actual values

- Go to Data View (note the columns now have headings):



- Enter all data for subject 1 (row 1) first. Position the mouse over the first box in row 1 representing the first variable and type in the appropriate value.
- Use the right arrow key to move across to the next box and type in the value of variable 2 for that subject. Continue for all variables.
- Then move on to the next row for subject 2's data, etc.

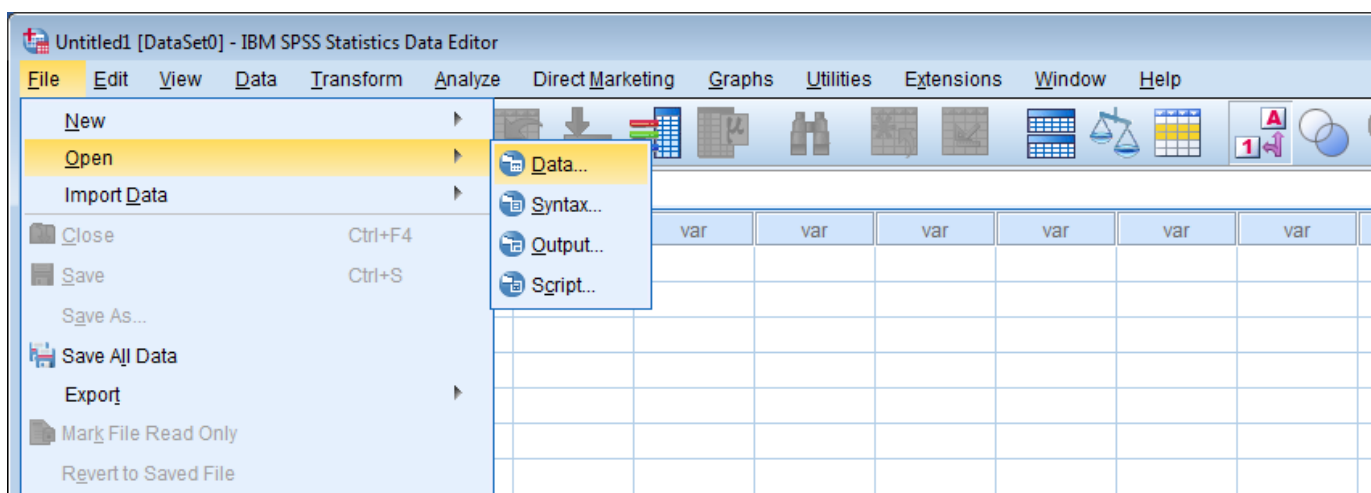


Remember to save your data file (see section 5.1).

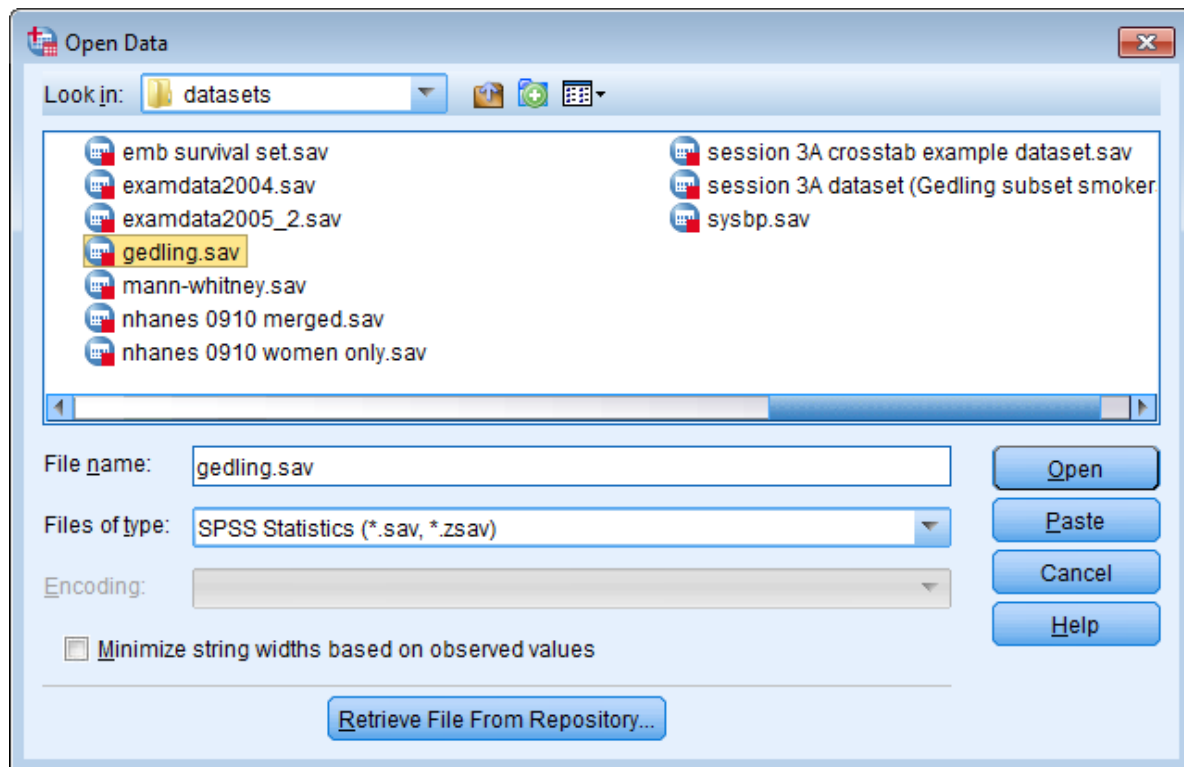
1.4 Opening an existing data file

Your data may already have been entered and saved as a data file in SPSS or a different package such as Excel, SAS, Access etc. To open an existing data file:

- In Data Editor window, select *File, Open, Data*.



- In the window that appears, choose the file type (eg *.sav for SPSS file, *.xls for Excel etc), and then browse and select your file so it appears by 'File name'.




- Click on the *Open* button.

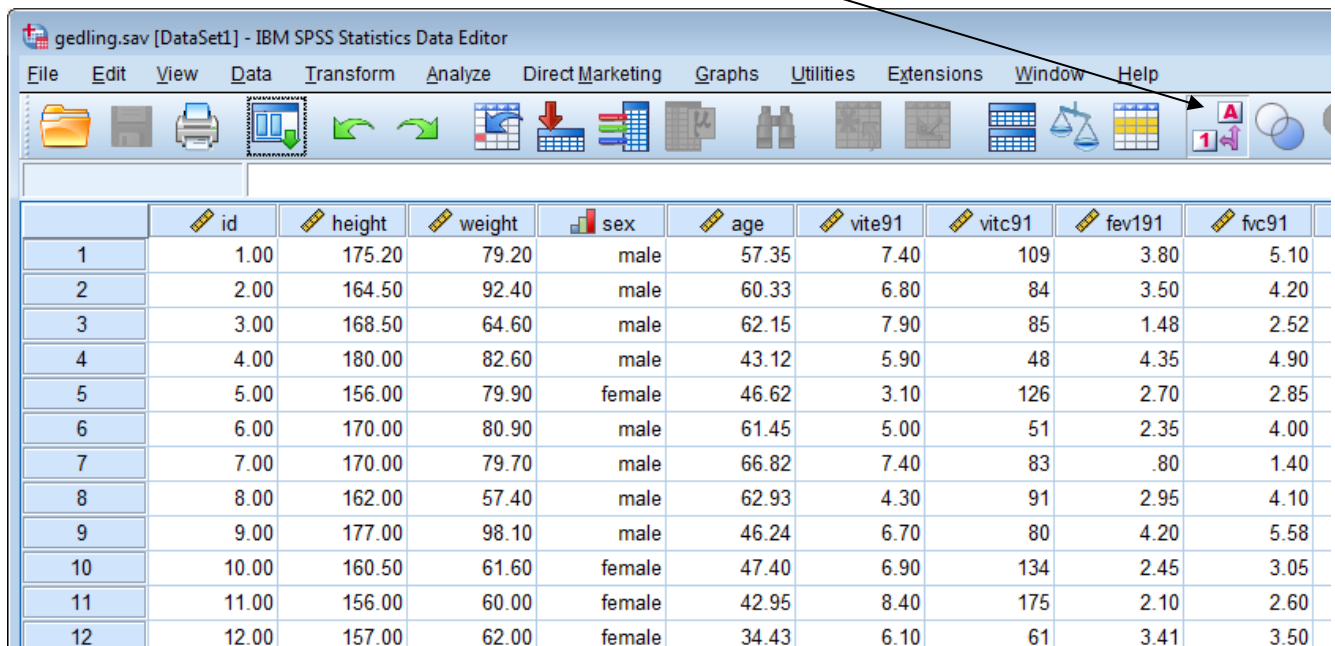
The name of the file will appear at the top of the Data Editor window in place of 'untitled1'. You can have more than one data file open in SPSS at the same time.

1.5 Viewing data

To get a feel for your data, go to Data View.

- Use the scroll bars to move up and down, and from right to left, to view the data.
- Use the arrow keys on your keyboard, or the mouse, to move from one cell to another in the spreadsheet.
- To move to the top, bottom, far right or far left of the spreadsheet, press control and the appropriate arrow key together. This is an easy way of seeing how many people are in your dataset.
- You can display the values of your categorical variables as the numeric codes entered (eg 1's and 2's for gender), or to view the value labels which you have defined in Variable View (eg male and female; see 1.3) go to *View* on the menubar, and choose *value labels*.

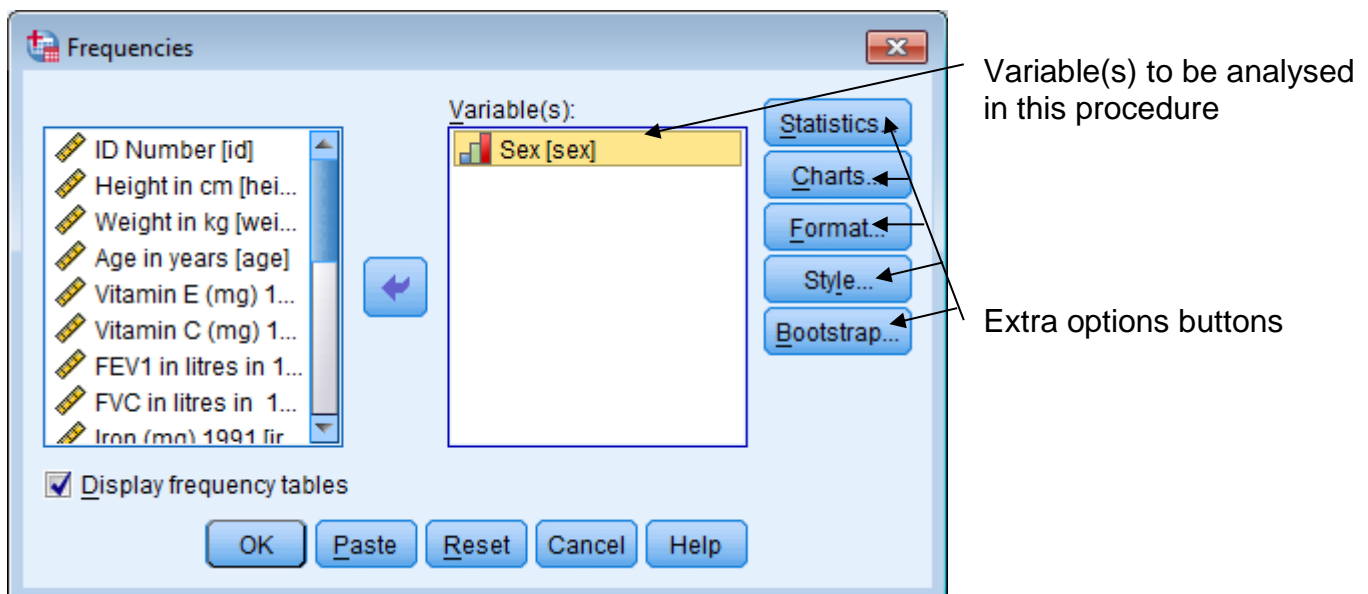
Alternatively use the value labels button  on the toolbar to swap between the two display options.



	id	height	weight	sex	age	vite91	vitc91	fev191	fcc91
1	1.00	175.20	79.20	male	57.35	7.40	109	3.80	5.10
2	2.00	164.50	92.40	male	60.33	6.80	84	3.50	4.20
3	3.00	168.50	64.60	male	62.15	7.90	85	1.48	2.52
4	4.00	180.00	82.60	male	43.12	5.90	48	4.35	4.90
5	5.00	156.00	79.90	female	46.62	3.10	126	2.70	2.85
6	6.00	170.00	80.90	male	61.45	5.00	51	2.35	4.00
7	7.00	170.00	79.70	male	66.82	7.40	83	.80	1.40
8	8.00	162.00	57.40	male	62.93	4.30	91	2.95	4.10
9	9.00	177.00	98.10	male	46.24	6.70	80	4.20	5.58
10	10.00	160.50	61.60	female	47.40	6.90	134	2.45	3.05
11	11.00	156.00	60.00	female	42.95	8.40	175	2.10	2.60
12	12.00	157.00	62.00	female	34.43	6.10	61	3.41	3.50

1.6 Carrying out procedures: the dialog box

When a particular statistical procedure is selected, a **dialog box** appears. This box is used for specify the details of the procedure, in particular, which variable(s) you want to analyse. For example below is the dialog box for the 'frequencies' command:



- In most dialog boxes, all variables are listed on the left and to choose the variable of interest, highlight it and click on the arrow to move across to the empty box.

- There may be other buttons that allow you to select specific **options**, eg which particular graphs/statistics to display, the format of the output, how missing values are to be treated. When one of these buttons is pressed, a separate box will open in which options can be specified. To close this box and return to the dialog box, select either *continue* (accepts any changes made) or *cancel* (rejects any changes made).
- Press OK to run the command.
- Next time the same command is chosen (within the same session) the variables and options chosen previously will still be there. To remove a variable, highlight and click on the arrow to take it back to the left-hand list. Click 'Reset' to reset completely.

1.7 The Output (Viewer) window

After carrying out your first procedure, a new window will open. This is the **Output** or **Viewer** window where all the output is displayed (tables, statistics and charts).

- As each subsequent procedure is performed, the new output is simply appended to the existing output.
- The output appears in **outline** down the left hand side section of the viewer window to make it easy to jump to different parts of the output.
- Click on any procedure that has been performed in the outline section and the detailed output from it will appear in the right hand section.

The screenshot shows the IBM SPSS Statistics Viewer window. The left pane, titled 'Output', contains an outline of the current session. Under 'Frequencies', the 'Statistics' and 'Sex' items are listed. The right pane displays the detailed output for the 'Frequencies' procedure. It includes a 'Statistics' table for 'Sex' and a detailed table of frequencies and percentages.

		Sex			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	472	48.5	48.5	48.5
	female	501	51.5	51.5	100.0
Total		973	100.0	100.0	

The Output window has its own menu bar and toolbar buttons, although many appear in both the Data Editor and Output windows.

2. EXPLORING VARIABLES

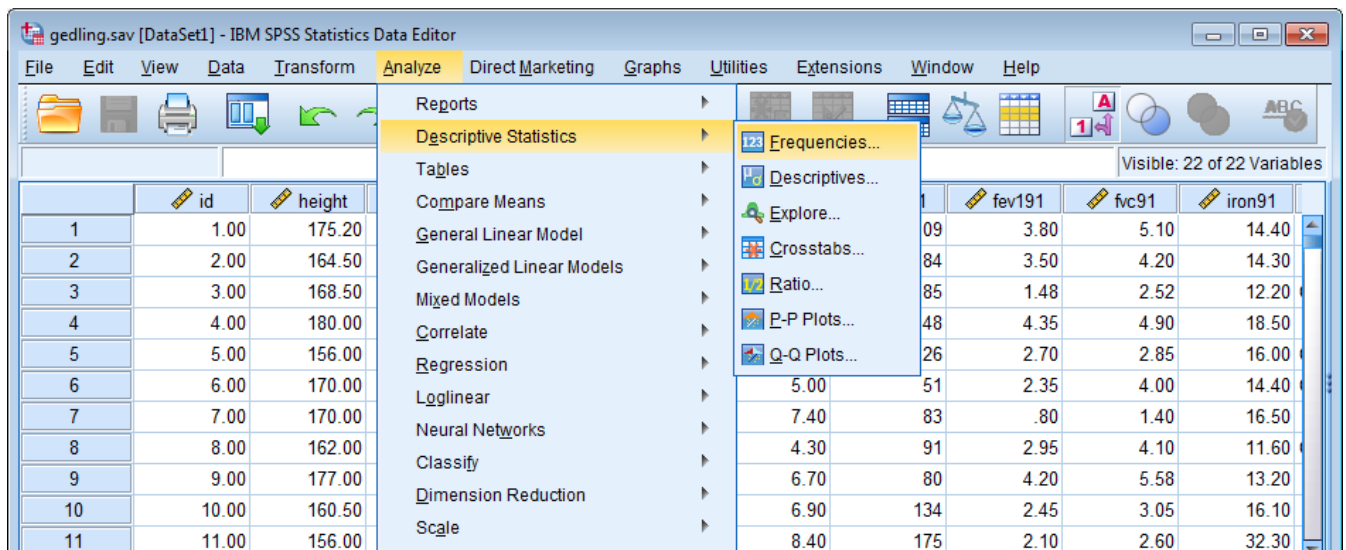
An important step of any data analysis, before any formal tests are carried out, is to explore or summarise the variables of interest.

2.1 Exploring categorical variables

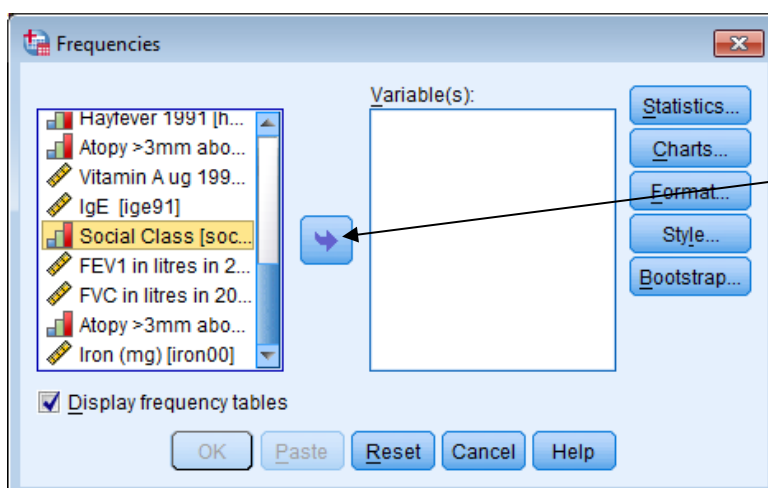
For categorical variables, remember the numeric values in the spreadsheet are just **codes** representing the different categories. To summarise such variables, we simply count the number of observations in each category (called **frequencies**) and display these either in a **frequency table** or **graphically (bar or pie chart)**.

2.1.1 *Frequency tables*

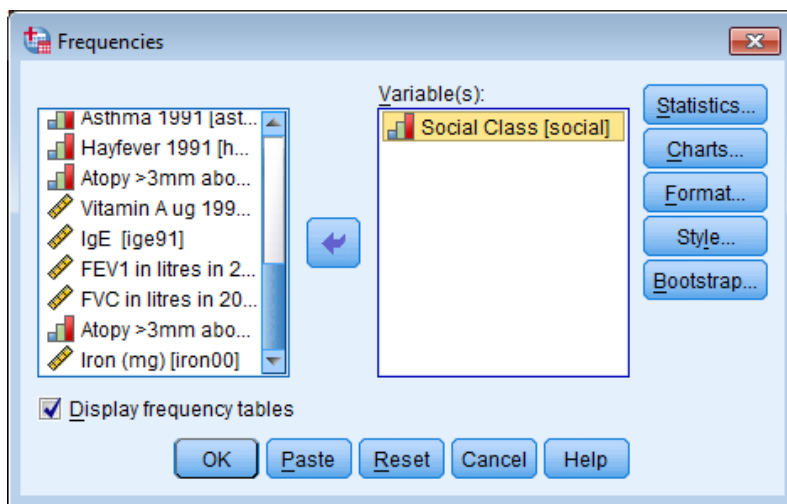
➤ Select *Analyse, Descriptive Statistics, Frequencies*:



➤ In the dialog box, select the variable of interest:



Click on the arrow and it will move into the box titled 'Variable(s)':



➤ Press OK. In the output window, a frequency table will be displayed as follows:

The screenshot shows the IBM SPSS Statistics Viewer window. The 'Output' pane on the left shows a tree structure with 'Frequencies' expanded, showing 'Title', 'Notes', 'Statistics', and 'Social Class'. The main window displays the 'Frequencies' output. At the top, the variable 'Social Class' is listed. Below it is a summary table:

Social Class		
N	Valid	963
	Missing	10

Below this is a detailed table for 'Social Class' with columns: Frequency, Percent, Valid Percent, and Cumulative Percent.

Social Class		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I	132	13.6	13.7	13.7
	II	342	35.1	35.5	49.2
	III NM	264	27.1	27.4	76.6
	III M	197	20.2	20.5	97.1
	IV-V	28	2.9	2.9	100.0
	Total	963	99.0	100.0	
Missing	System	10	1.0		
	Total	973	100.0		

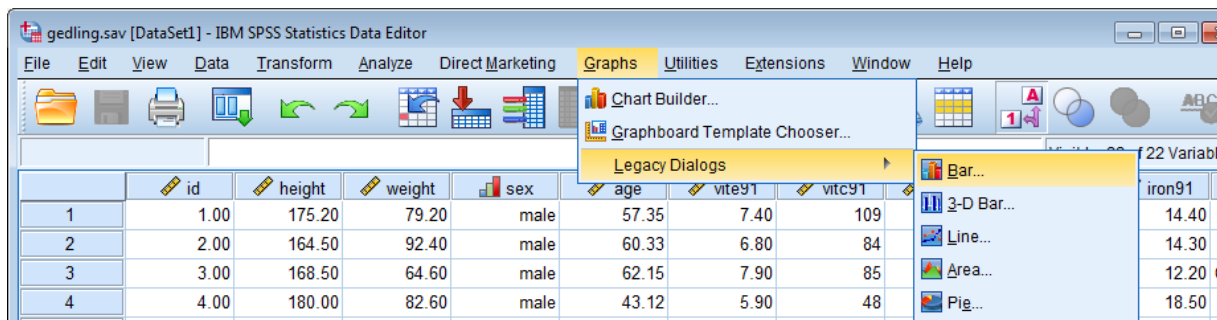
Annotations with arrows point to specific parts of the output:

- 'Name of variable' points to 'Social Class'.
- 'Number of subjects with a valid value' points to '963' in the summary table.
- 'Number of subjects with missing values' points to '10' in the summary table.
- 'Number in each category' points to the 'Frequency' column in the detailed table.
- '% in each category' points to the 'Percent' column in the detailed table.
- '% in each category (excluding missing)' points to the 'Valid Percent' column in the detailed table.

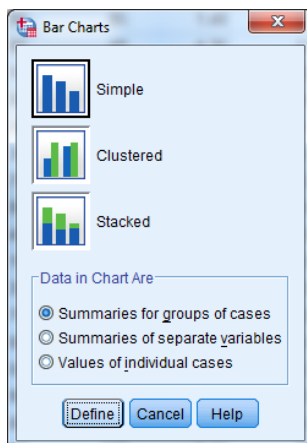
The status bar at the bottom right says 'IBM SPSS Statistics Processor is ready'.

2.1.2 Bar charts

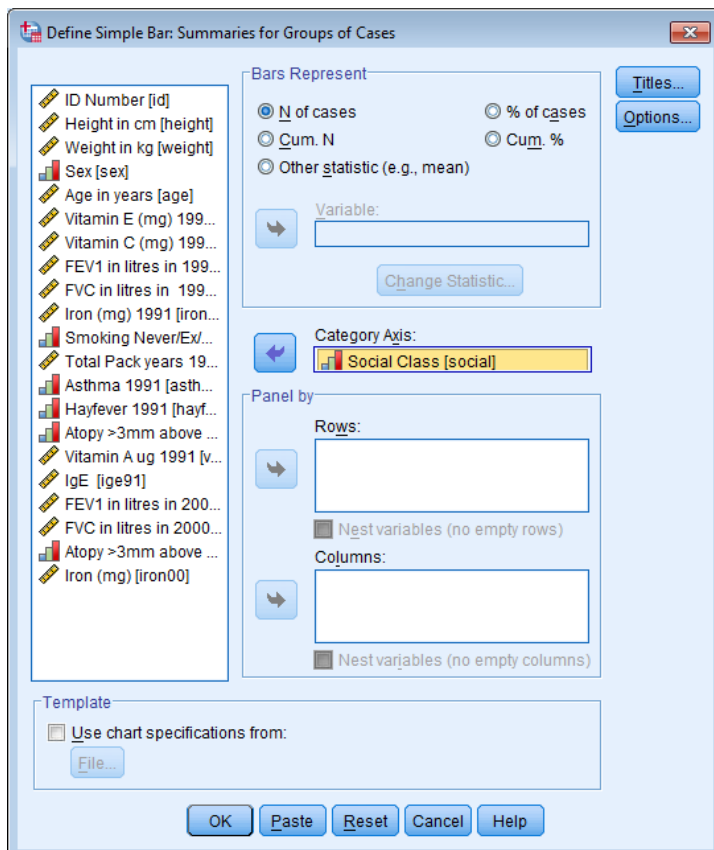
➤ Select *Graphs, Legacy Dialogs, Bar*:



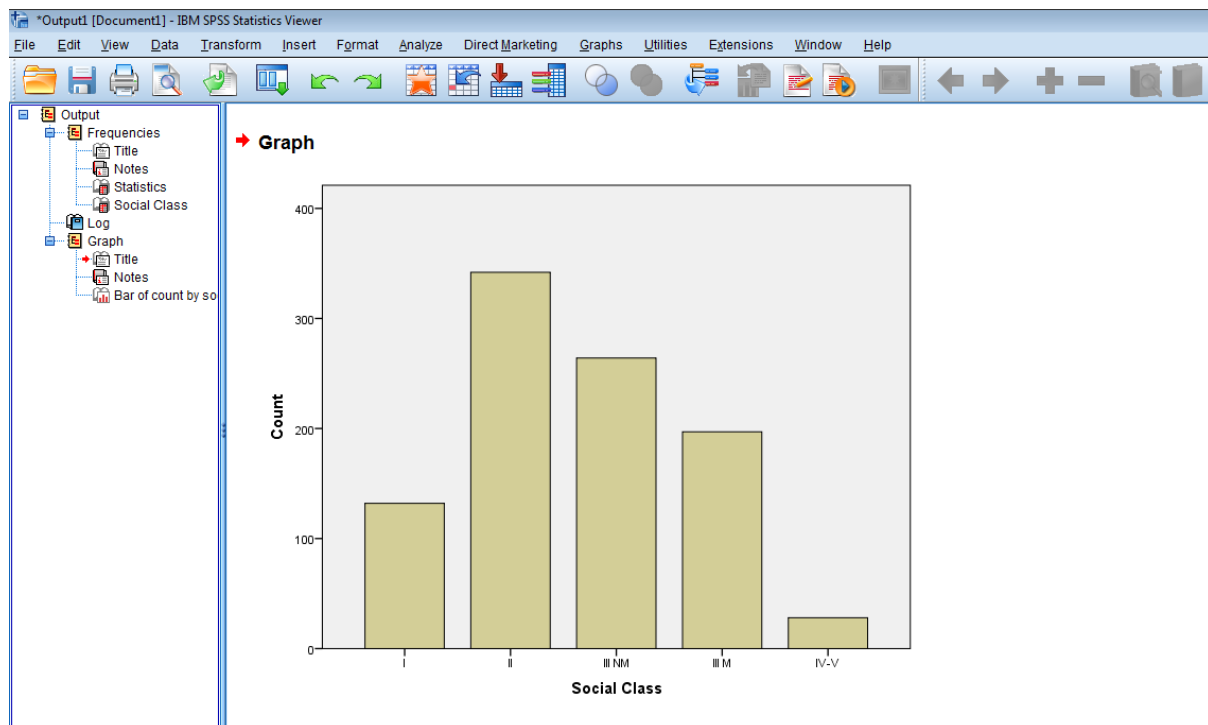
➤ Select *Simple* (should be already selected by default) and click on 'Define':



➤ Choose the variable of interest from the list on the left, and click on the arrow button to move it into the 'Category Axis' box.



- Click OK. The output will look like this:

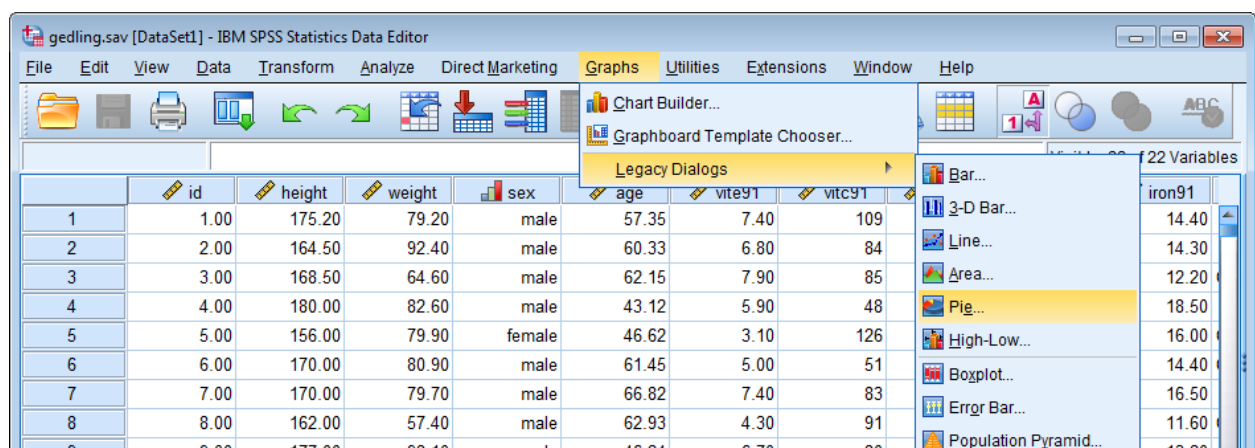


Note: Sometimes the y-axis of the graph produced in SPSS will not start from 0 and this can be misleading. This can be changed by double-clicking on the graph, then double-clicking on the y-axis to edit. Change the minimum of the range to 0.

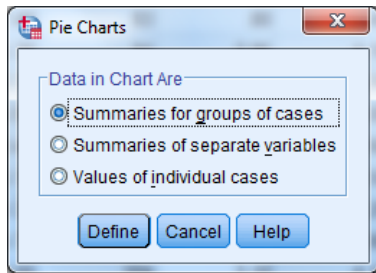
- To construct a bar chart where the bars represent the **percentage** frequencies, follow the same instructions, but in the dialog box under 'bars represent', chose % of cases instead of N of cases.
- To include missing values in the graph, chose *Options* from the dialog box and tick 'display groups defined by missing values'.

2.1.3 Pie charts

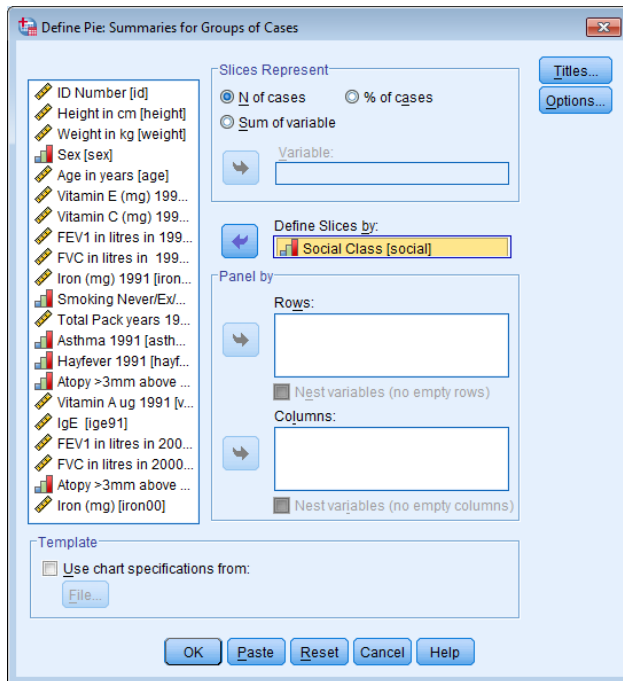
- Select *Graphs, Legacy Dialogs, Pie:*



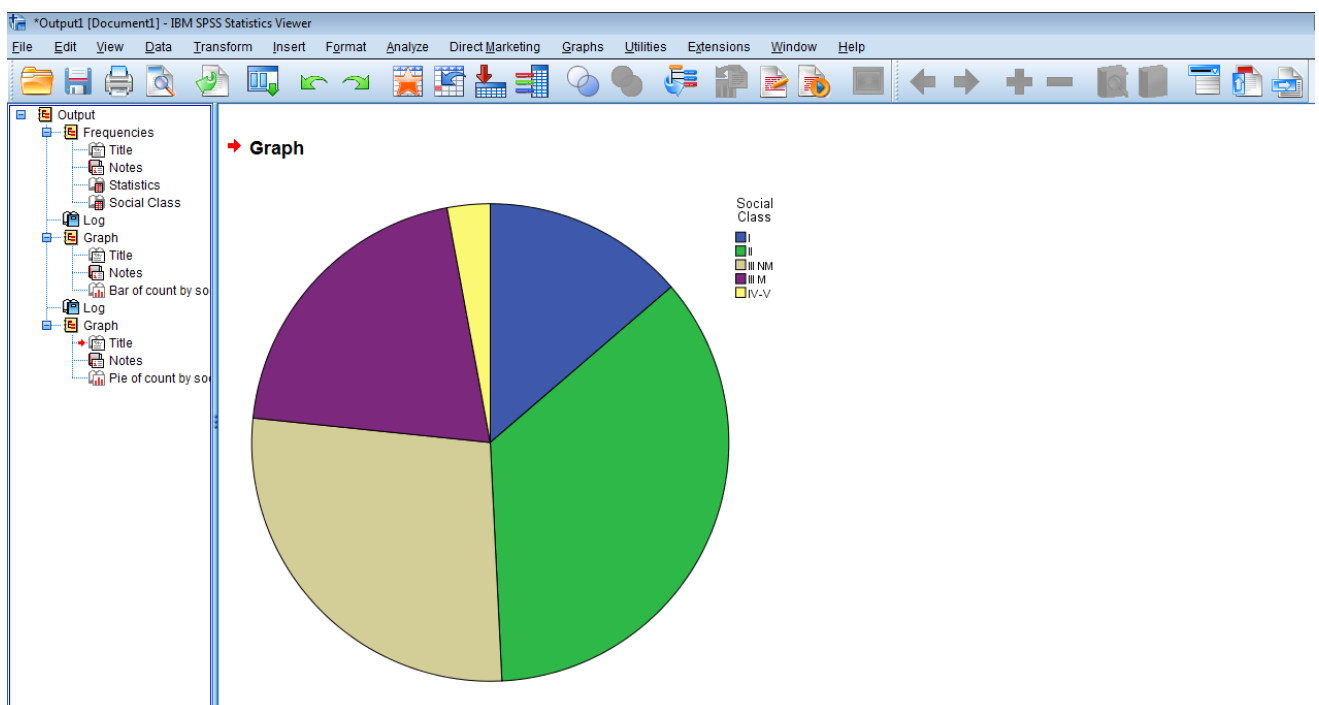
- Select *summaries for groups of cases* and click on the 'Define' button:



- Find variable of interest from the left column and move to the 'Define Slices by' box:



- Under 'Slices Represent' leave as the default 'N of cases' (choosing % of cases will result in the same pie chart). The following output is produced:



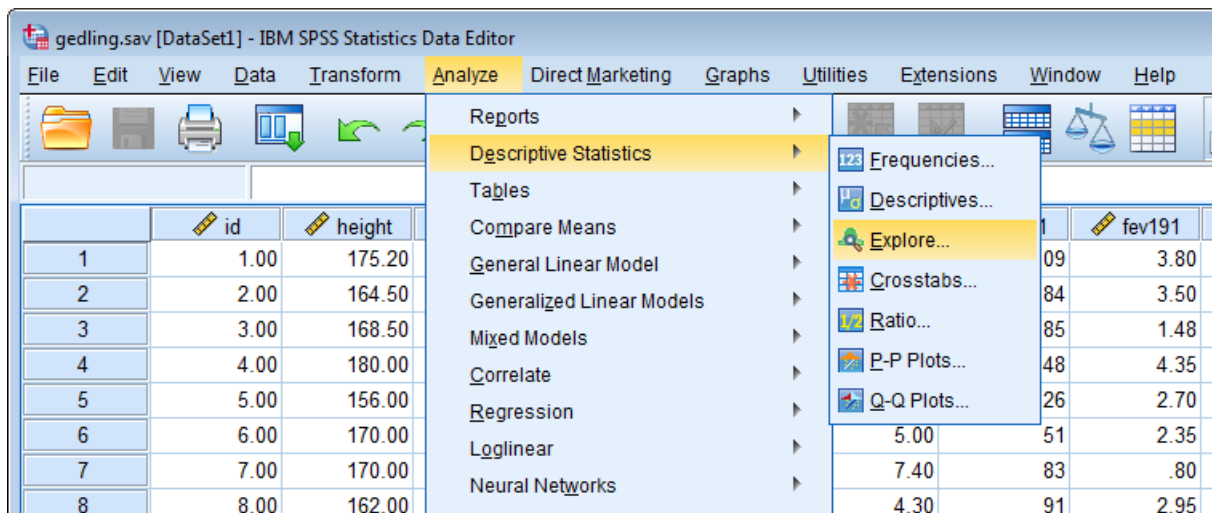
Missing values can be included in the same way as for the bar chart.

2.2 Exploring continuous variables

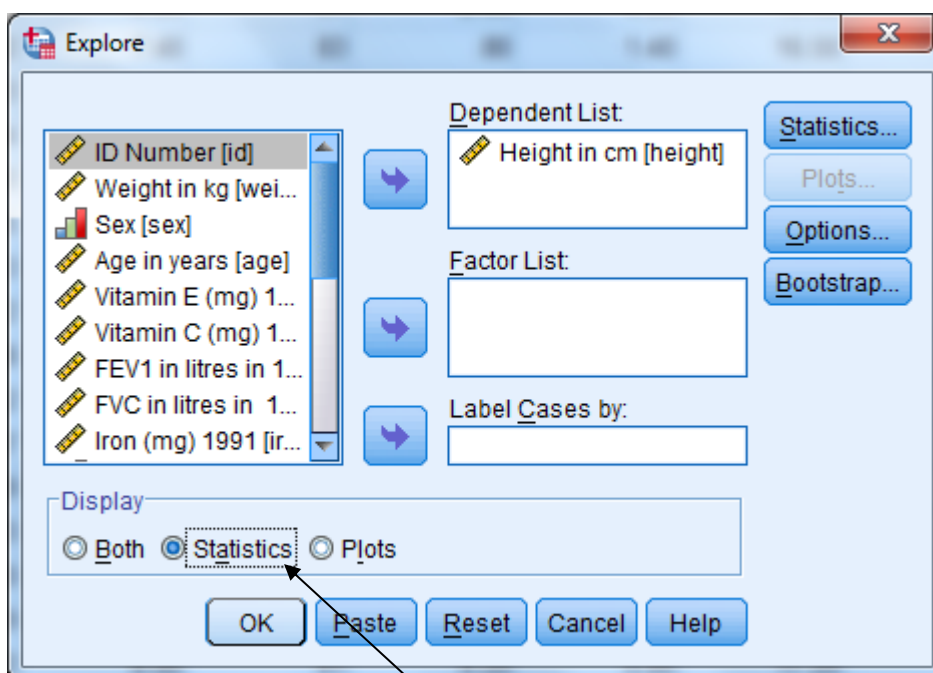
We can summarise continuous variables using **summary statistics** (mean, median, standard deviation, range etc) and **graphs** (histogram, box plot, stem and leaf plot).

2.2.1 Summary statistics

➤ Go to *Analyze, Descriptive Statistics, Explore*:



➤ Select the variable of interest and move into the 'Dependent List' box:



- Under display, chose 'Statistics' to display summary statistics only.
- Click OK. Here is the output produced for 'height':

Case Processing Summary

	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Height in cm	973	100.0%	0	0.0%	973	100.0%

Descriptives

Height in cm	Statistic	Std. Error
95% Confidence Interval for Mean	Lower Bound	166.9456
	Upper Bound	168.1206
5% Trimmed Mean		167.3974
Median		167.0000
Variance		87.206
Std. Deviation		9.33841
Minimum		145.00
Maximum		197.50
Range		52.50
Interquartile Range		13.00
Skewness	.208	.078
Kurtosis	-.253	.157

Annotations in the image:

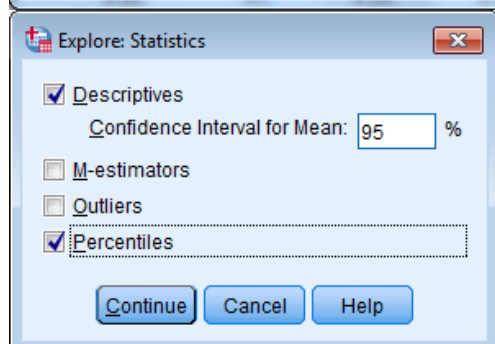
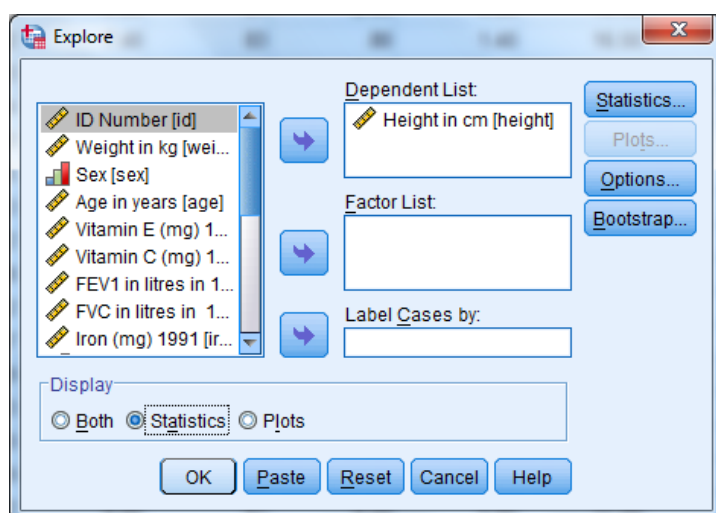
- Number with valid value (points to 973 in Case Processing Summary)
- Number excluded from analysis (missing values) (points to 0 in Case Processing Summary)
- Table of summary statistics (points to the Descriptives table)

In the 'Statistic' column of the table, the mean and its 95% confidence interval, the median, variance, standard deviation, and the minimum and maximum values (along with their difference, the 'range') are displayed. Also, the standard error of the mean is given in the last column, on the first line. The values for skewness and kurtosis aren't commonly used. For the value of the interquartile range, see below.

Interquartile range

Note the value displayed next to 'Interquartile range' using the *Explore* procedure is not the values we want: it is the *difference* between the 75th percentile and 25th percentile rather than the actual values. To display the actual values:

- Click on *Statistics* button in the Explore dialog box
- Then tick the box next to the word 'Percentiles'



In the output, underneath the main descriptives table of summary statistics, you will see an additional table titled Percentiles, like below. Reading the top line of output, we get the interquartile range (25th to 75th percentiles) which here for height is 161 to 174 cm.

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays a table titled 'Percentiles'. The table has columns for various percentiles (5, 10, 25, 50, 75, 90, 95) and rows for 'Weighted Average (Definition 1)' and 'Tukey's Hinges'. The variable being analyzed is 'Height in cm'. The 'Weighted Average' row shows values: 153.0000, 155.7600, 161.0000, 167.0000, 174.0000, 180.0000, 183.0000. The 'Tukey's Hinges' row shows values: 161.0000, 167.0000, 174.0000. Two arrows from the text above point to the 25th and 75th percentile values in the 'Weighted Average' row.

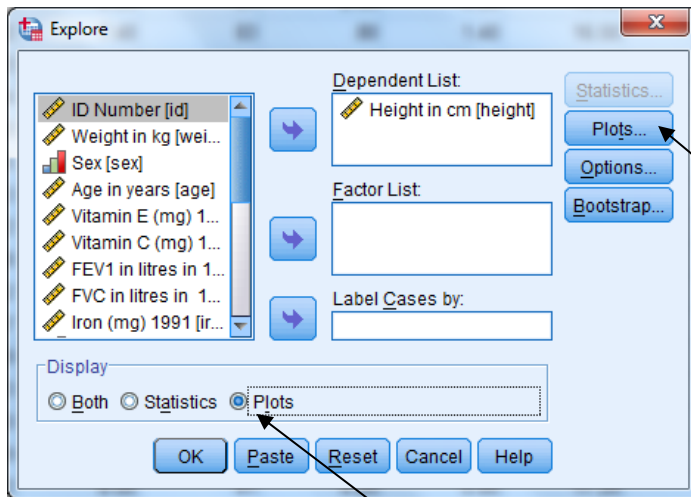
		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	Height in cm	153.0000	155.7600	161.0000	167.0000	174.0000	180.0000	183.0000
Tukey's Hinges	Height in cm			161.0000	167.0000	174.0000		

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 142, W: 758 pt

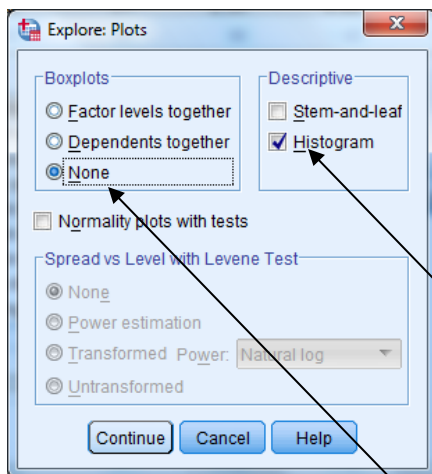
2.2.2 Histogram

There are two ways to produce a histogram in SPSS. The first uses the 'Explore' procedure as above:

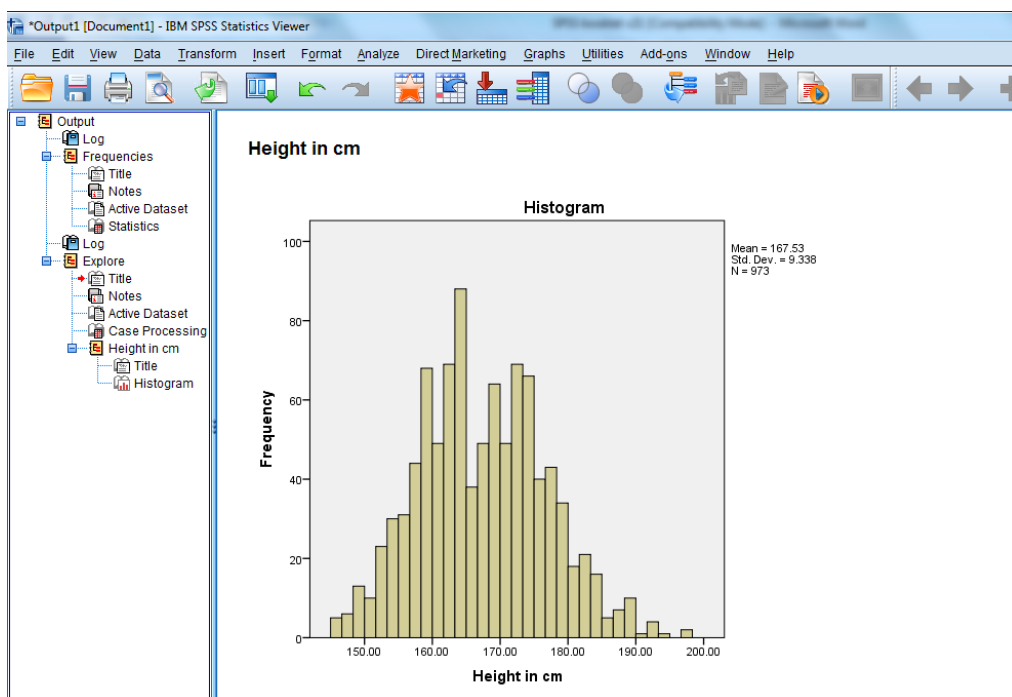
- *Analyse, Descriptive Statistics, Explore...*
- Select the variable of interest and move into the 'Dependent List' box:



- Under 'Display', choose 'Plots' instead of 'Statistics', and then click on the 'Plots...' button. The following window will open:



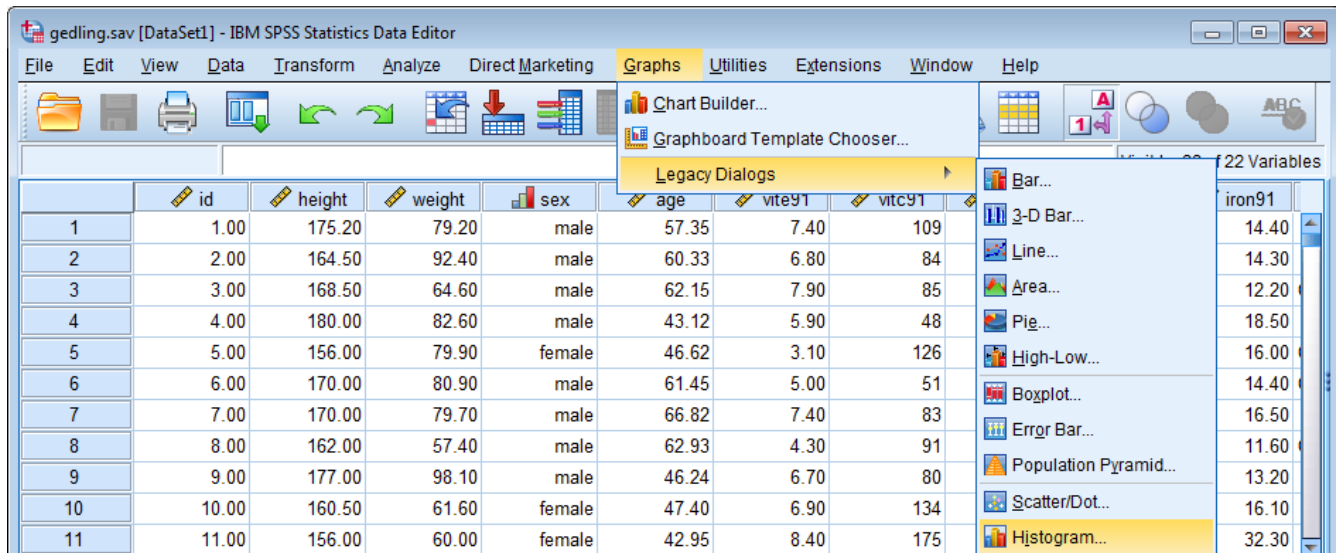
- Under Boxplots select 'None', then tick 'Histogram'. Click *Continue*, and then *OK*. In the output window a histogram will be displayed.



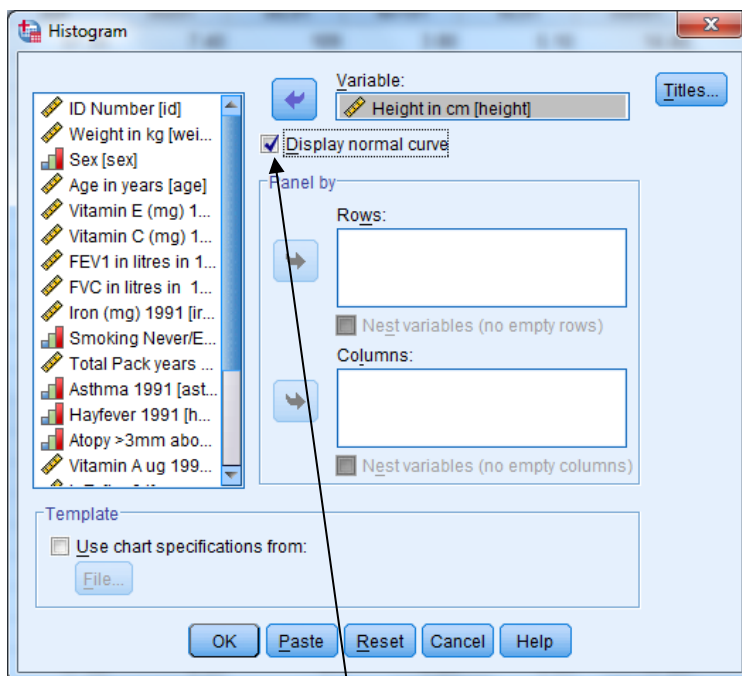
Next to the histogram, the standard deviation and mean are displayed (remember only to be used if graph is roughly symmetric, ie Normal distribution), along with the number of subjects included in the analysis (N).

The alternative way of producing a histogram is using the 'Graph' menu:

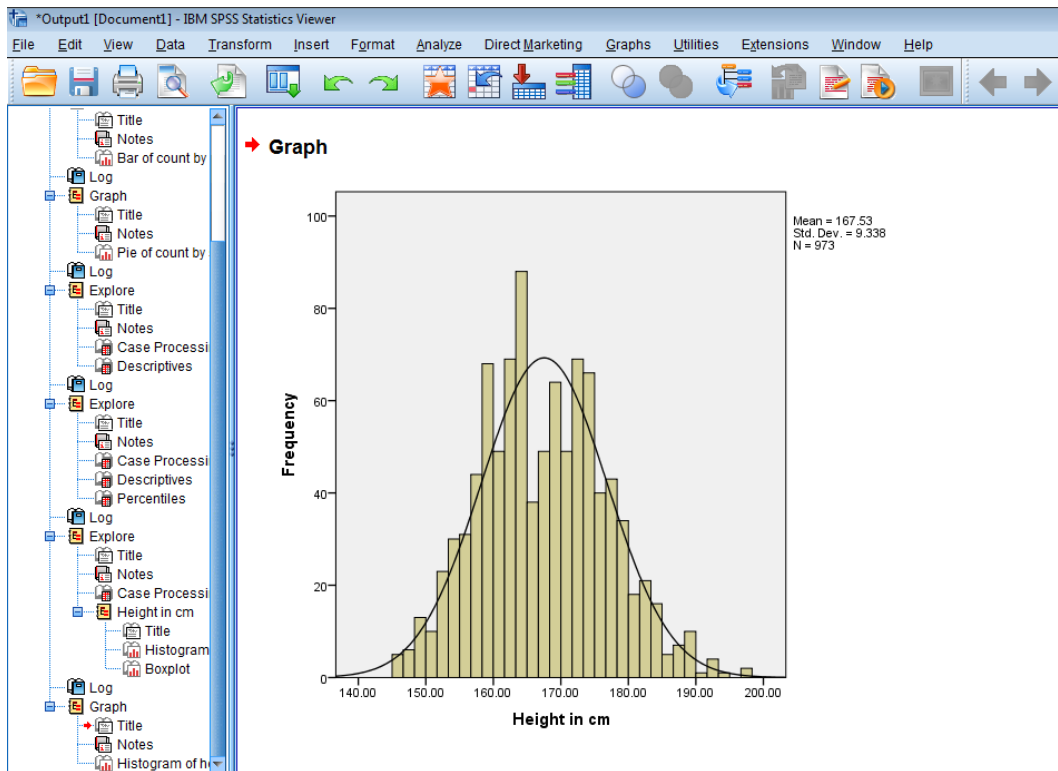
➤ *Graphs, Legacy Dialogs, Histogram...*



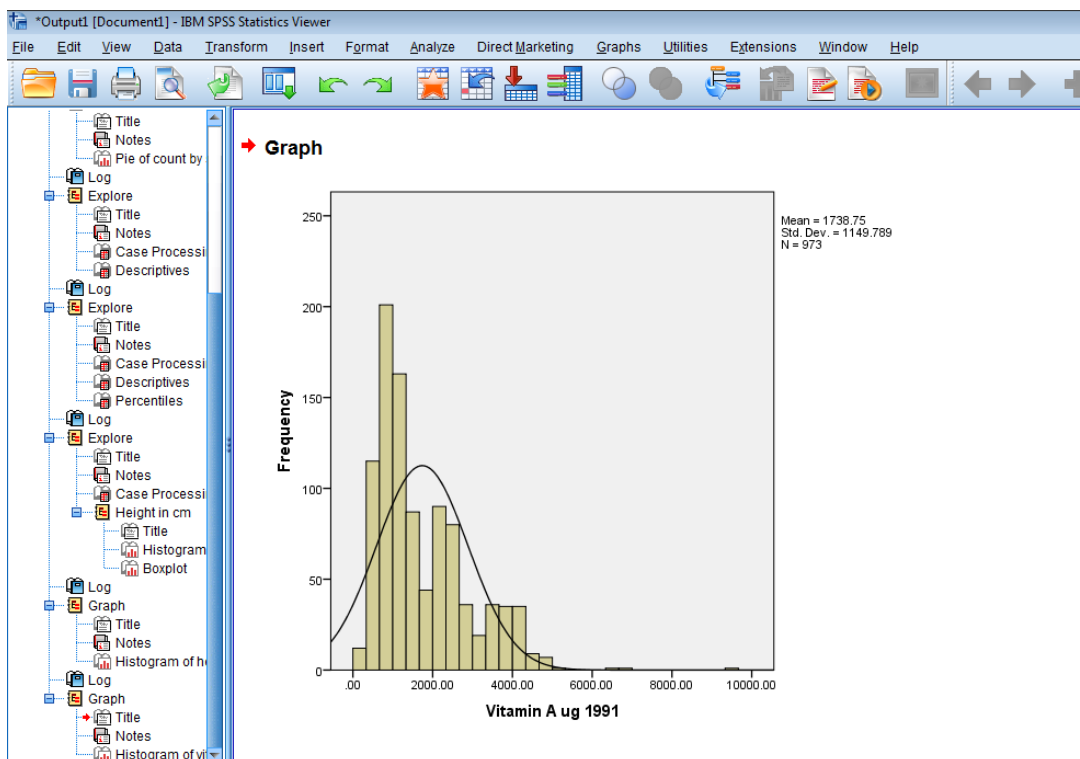
➤ Move the variable of interest into the 'Variable' box:



- Tick 'display normal curve' if you want a Normal curve to appear on the histogram to help you decide whether the distribution is Normal or not.
- Click OK. The output will look like this. You can see the histogram follows the curve fairly well.....

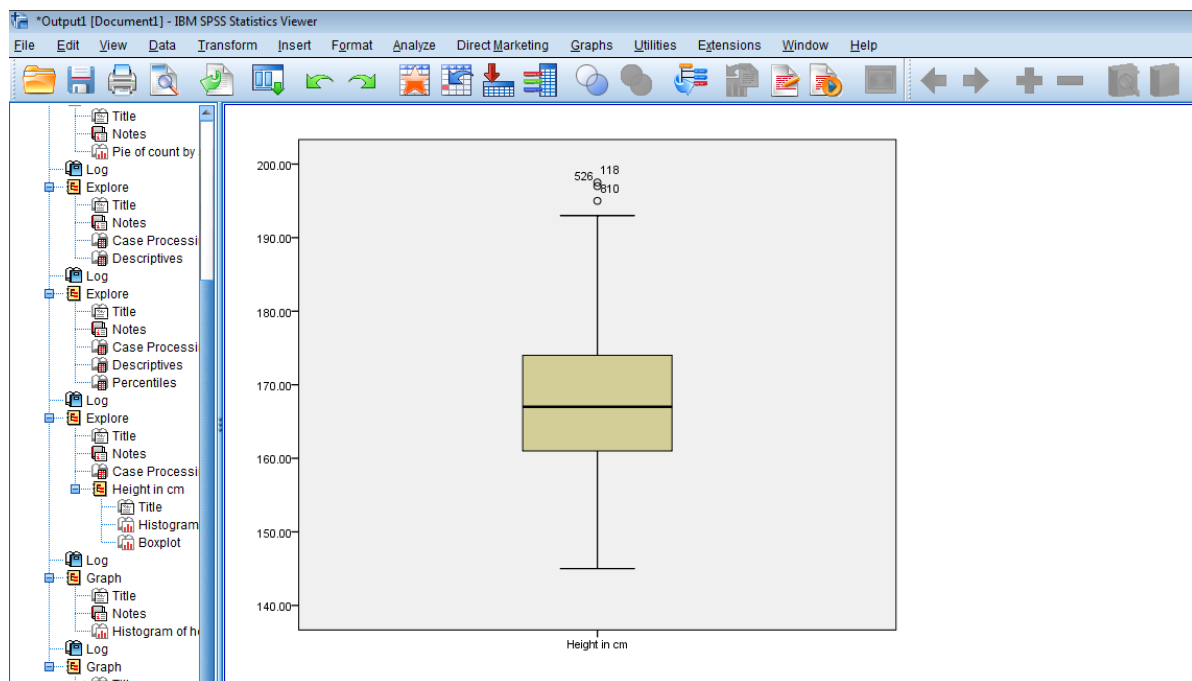


.....but less so in this example below:



2.2.3 Box plot

When you use the 'Explore' procedure in SPSS and select plots, a graph called a **box plot** (sometimes called box-and-whisker plot) will be automatically displayed, unless you select none:



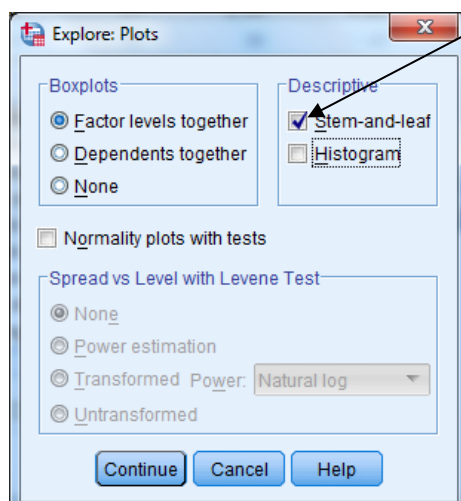
This is a summary plot based on the median, quartiles, and extreme values.

- The box represents the interquartile range which contains the middle 50% of values.
- The line across the box indicates the median.
- The whiskers are lines that extend from the box to the highest and lowest values, excluding outliers.
- Outliers are plotted individually, labelled with their case number.

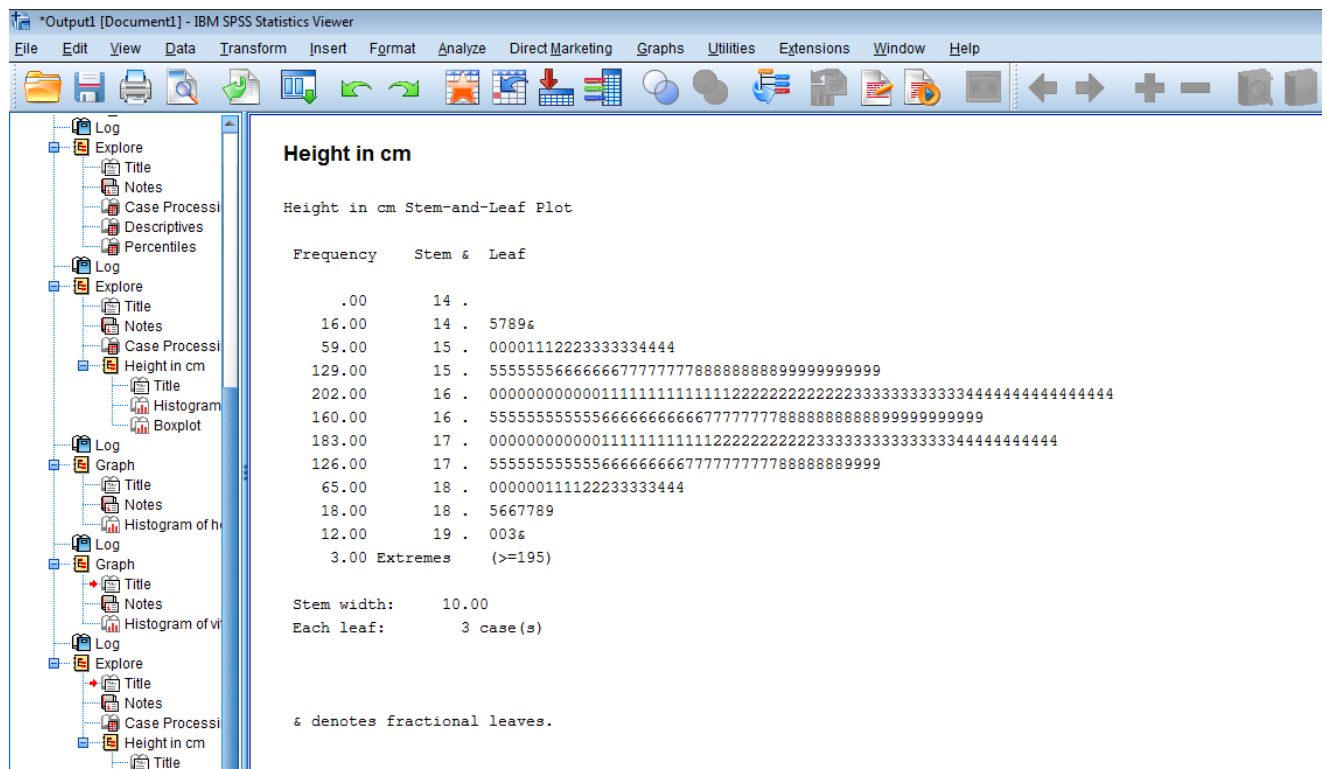
2.2.4 Stem and leaf plot

Another graph that shows the frequency distribution is the **stem and leaf plot**, but this is not commonly used.

- The 'stem and leaf' option is already ticked by default under the Plots option of Explore. Untick it if you don't want it displayed.



➤ Press continue and OK. The output will look like this:



The stem-and-leaf plot is like a histogram on its side and it shows all the actual observations. Values are obtained by multiplying the value on the stem, by the stem width and adding on the leaf value.

2.3 Advanced graphs

SPSS does have alternative ways of creating graphs which can be useful for creating more complicated or advanced plots. Choose either

- *Graphs, Chart Builder...* or
- *Graphs, Graphboard Template Chooser...*

3. CREATING NEW VARIABLES

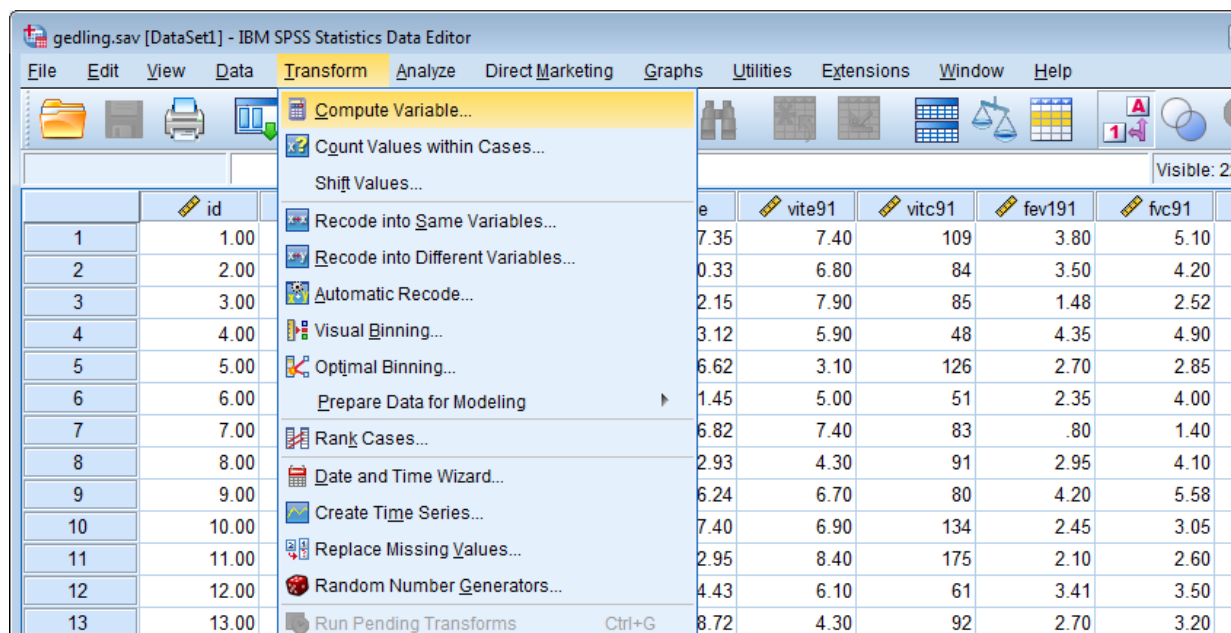
Sometimes you may wish to create a **new variable** from an existing variable. For example you may:

- have a skewed variable and want to make a new variable which is the transformed values (eg logged values) of the original variable,
- want to group a continuous variable into a categorical variable, where the categories represent bands, quartiles etc (eg age into 10 year age bands), or
- want to regroup a categorical variable, for example, you may have small numbers in some groups and need to merge them together.

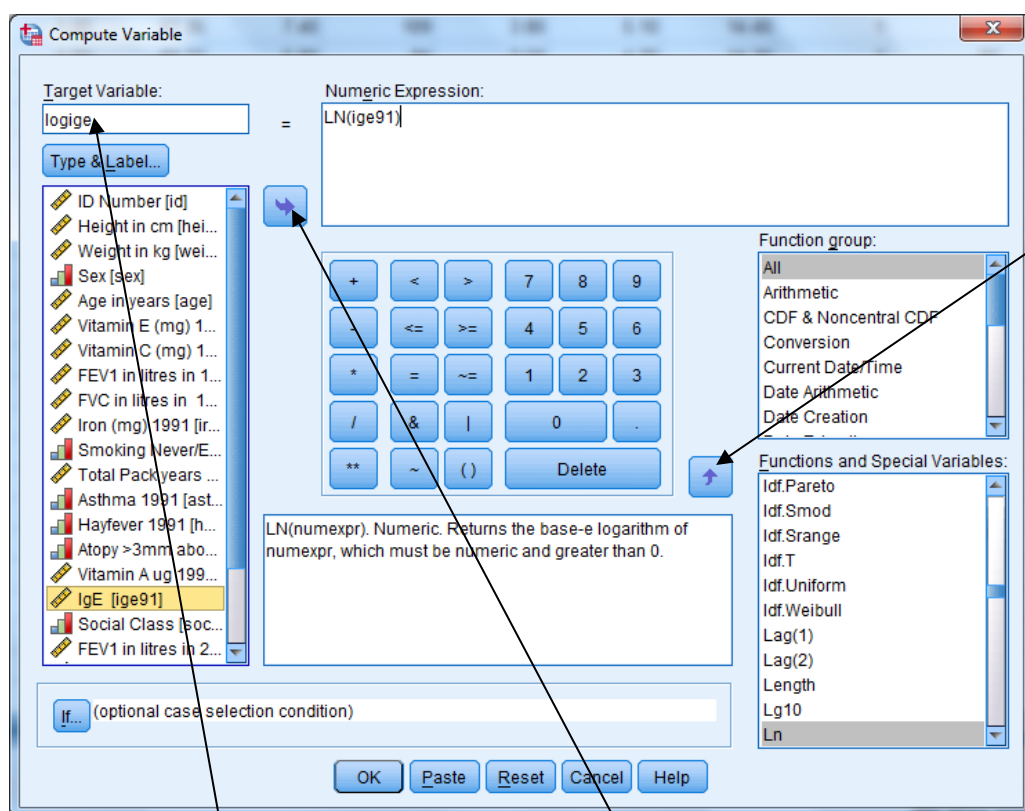
3.1 Compute command

Use this command if you want to compute a new variable by applying some kind of **formula** or **function** (eg logging, adding, square rooting etc) to an existing variable(s):

➤ Select *Transform, Compute Variable*



- Give your new variable a name by typing a name for the variable in the 'Target Variable' box.
- Click on 'All' under 'Function group' and then find the specific function you want to apply from the list of functions that appear underneath (under 'Functions and Special Variables'). Click on the up arrow to move it into the numeric expression box.
- From the list of variables on the left, find the original variable you want to apply the function to. Click on the arrow to the right to complete the numeric expression.



Click here to apply chosen function, ie to move function up to Numeric Expression box

Type a name for your new variable here

Click here to move selected original variable to Numeric Expression box

- Press OK. The new variable will now appear in the data view window at the far right of the spreadsheet. Remember to label your new variable by going to Variable View.

Functions you may want to use are as follows (var represents the original variable):

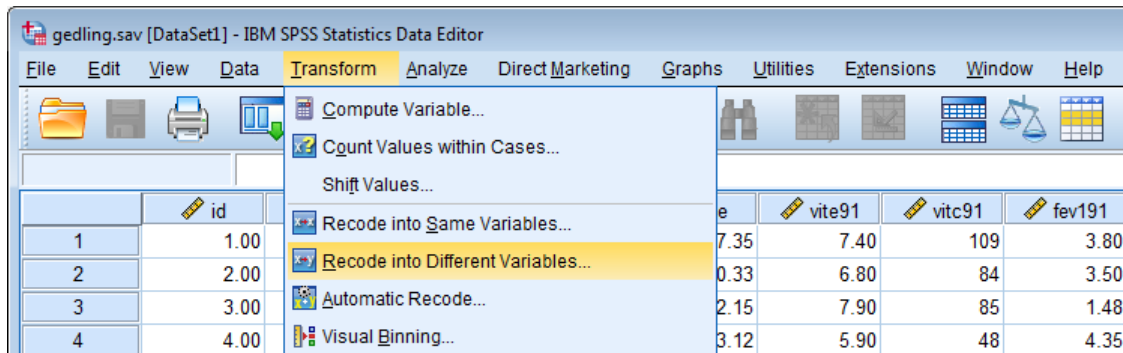
LN(var)	natural logarithm
LG10(var)	logarithm to base 10
SQRT(var)	square root
1/var	reciprocal, ie 1 divided by value of variable
var**2	squared

Note: It is not possible to take the logarithm or reciprocal of a zero value; therefore if your original variable has any zero values, you will need to add a small value before applying the log/reciprocal, eg LN(ige + 0.1) rather than LN(ige).

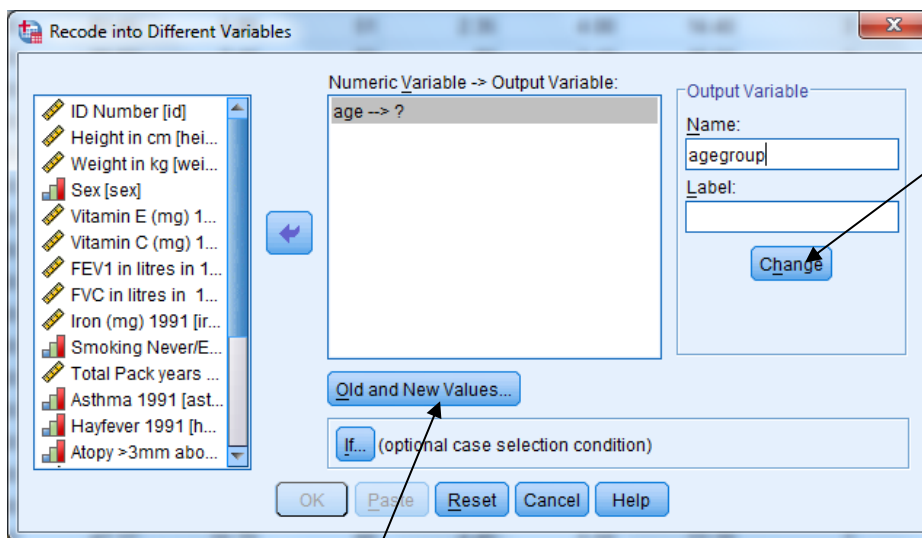
3.2 Recode command

Use this command if you want to create a new variable by **reassigning** the values of an existing variable, eg to categorise an existing continuous variable, or to merge categories of a categorical variable together.

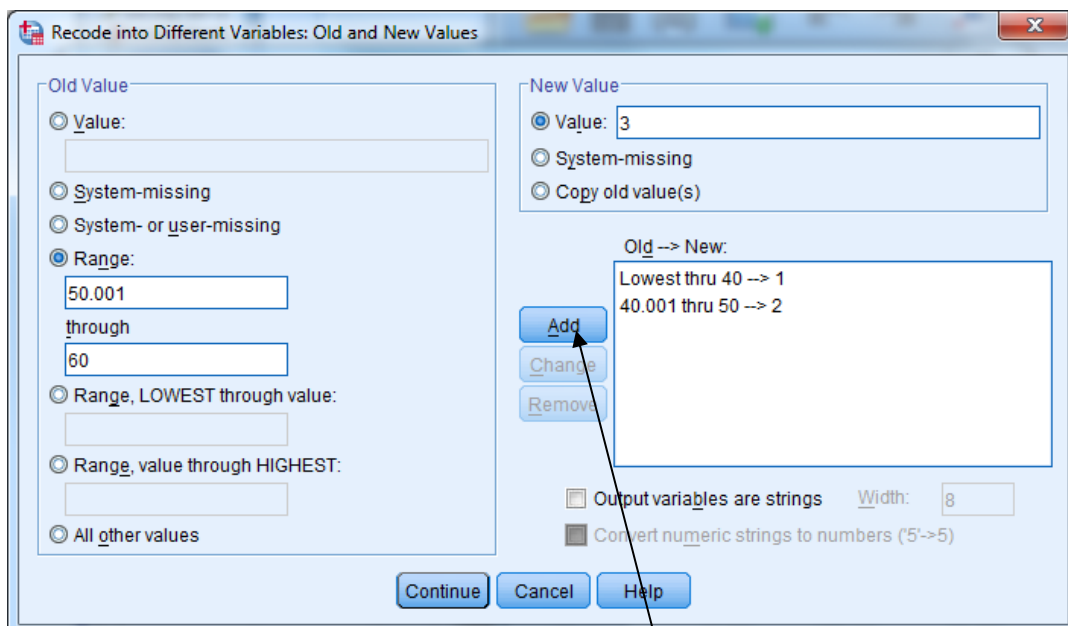
- Select *Transform, Recode into Different Variables*:



- Click across the old variable into 'Numeric Variable -> Output Variable' box and type the name you want to call the new variable in the box titled 'Name'. Press *Change*.



- Click on 'Old and New Values'. In the next box, you need to define the recoding.



- For each value or range of values of the old variable entered into the left side, type in the new value in the right side. Press 'Add' after each change.

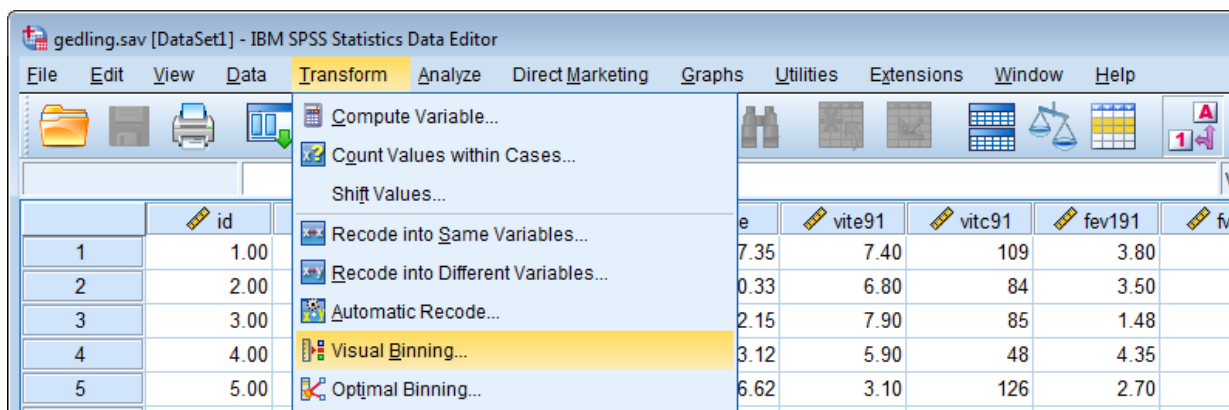
Examples

1. You want to recode age into age categories such that those aged 40 or younger are coded as 1, those aged 40-50 coded as 2, aged 50-60 coded as 3 and 60+ as 4.
 - Under 'Old value' choose the fifth option and by 'Range, LOWEST through value' type 40. Type 1 under new value and press add.
 - Choose the fourth option of old values (Range), and type 40.0001 through 50. Type 2 under new value and press add,
 - Do the same for 50.001 through 60 but type 3 under new value.
 - Choose the sixth option (Range, value through HIGHEST) and type 60.001, and then type 4 under new value. Press add.
 2. You want to recode social class (currently coded 1 to 6) so that the 5 and 6 categories are merged together, ie 6's are recoded as 5's.
 - Under old values choose the first option and type 6. Type 5 in the new value box and press add.
 - Choose the bottom option under old values 'All other values' and choose 'copy old values' under new values. Press add. This will leave the rest of the coding the same.
- Press *continue* when all 'old to new' changes appear in the box
- Press OK. The new variable will appear at the far right of the spreadsheet.
- Go to variable view to complete the 'label' box, and define the 'values' (ie codes) of the new variable.

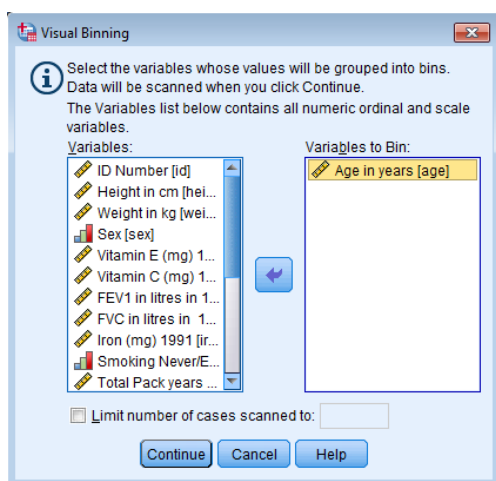
3.3 Visual binning command

An alternative way to recode a continuous variable into a categorical variable is to use the visual binning command. This is the appropriate command to use when you want your categories to have:

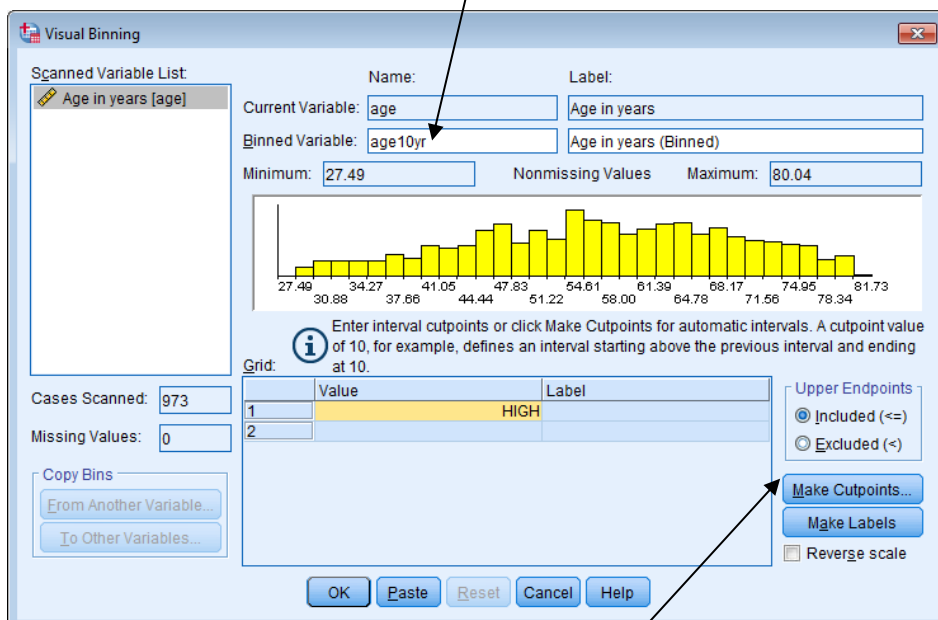
- equal width intervals (eg age 10.01-20, 20.01-30, 30.01-40 etc)
 - intervals with the same number of cases in each (called percentiles, eg tertiles if you want 3 equal groups, quartiles for 4, quintiles for 5 etc)
 - intervals based on standard deviations (eg 3 cutpoints: mean, mean-1sd & mean+1sd)
- Select *Transform, Visual Binning...*



- Move the continuous variable you want to categorise into the box on the right:



- Press *Continue*. In the dialog box that appears, click on the variable you want to categorise on the left. A graph of its distribution will appear as below.
- In the box next to 'Binned Variable:' type in the name you want to give your new variable ('age10yr' here):



- Looking at the graph will help you decide how to categorise the data. To define your categories, click on 'Make Cutpoints'. The following window will open:

Make Cutpoints

☒ **Equal Width Intervals**

Intervals - fill in at least two fields

First Cutpoint Location: 30.00

Number of Cutpoints: 6

Width: 10

Last Cutpoint Location: 80.00

☐ **Equal Percentiles Based on Scanned Cases**

Intervals - fill in either field

Number of Cutpoints:

Width(%):

☒ **Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases**

☐ +/- 1 Std. Deviation

☐ +/- 2 Std. Deviation

☐ +/- 3 Std. Deviation

i Apply will replace the current cutpoint definitions with this specification.
A final interval will include all remaining values: N cutpoints produce N+1 intervals.

Apply Cancel Help

For equal width intervals, choose the top option and fill in at least 2 fields.

For example, start by specifying the first cutpoint location here.

Then specify the width of interval you want here.

The number of cutpoints will then be calculated and filled in automatically.

- When finished specifying your cutpoints, click on *Apply*.

Visual Binning

Scanned Variable List:

Age in years [age]

Name: age Label: Age in years

Binned Variable: age10yr Label: Age in years (Binned)

Minimum: 27.49 Nonmissing Values Maximum: 80.04

Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Grid:

	Value	Label
1	30.000	<= 30.00
2	40.000	30.01 - 40.00
3	50.000	40.01 - 50.00
4	60.000	50.01 - 60.00
5	70.000	60.01 - 70.00
6	80.000	70.01 - 80.00
7		HIGH 80.01+
8		

Upper Endpoints

☒ Included (<=)

☐ Excluded (<)

Make Cutpoints...

Make Labels

☐ Reverse scale

OK Paste Reset Cancel Help

- Lines will now be shown on the graph to show how the categories will be made. By clicking on 'Make labels', labels will be created automatically for the new categories (shown in the label column). This means there is no need to go to variable view later and label them.
- Click OK and the new variable will be made.

- To create a variable based on groups with equal numbers of observations in each (percentiles), choose the second option of the 'make cutpoints' box.

The 'Make Cutpoints' dialog box is shown with the following settings:

- Equal Width Intervals:** Not selected.
- Equal Percentiles Based on Scanned Cases:** Selected.
 - Intervals - fill in either field:
 - Number of Cutpoints: 3
 - Width(%): 25.00
- Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases:** Not selected.

Buttons at the bottom: Apply, Cancel, Help.

Type in the number of cutpoints you want here (remember this will be one less than the number of groups you want. So type 3 for quartiles, 4 for quintiles etc)

- Similarly, to create groups based on the mean and standard deviations choose the third option.

The 'Make Cutpoints' dialog box is shown with the following settings:

- Equal Width Intervals:** Not selected.
- Equal Percentiles Based on Scanned Cases:** Not selected.
- Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases:** Selected.
 - ☒ +/- 1 Std. Deviation
 - ☐ +/- 2 Std. Deviation
 - ☐ +/- 3 Std. Deviation

Buttons at the bottom: Apply, Cancel, Help.

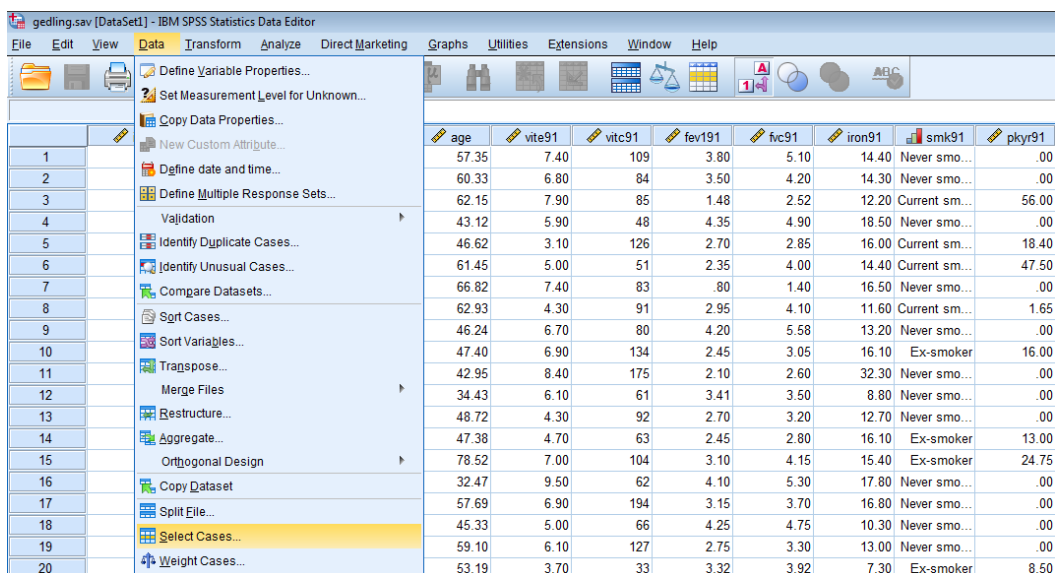
Selecting this option will create 3 cut points: the mean, the mean + 1sd and the mean - 1sd

4. LOOKING AT DIFFERENT SUBGROUPS

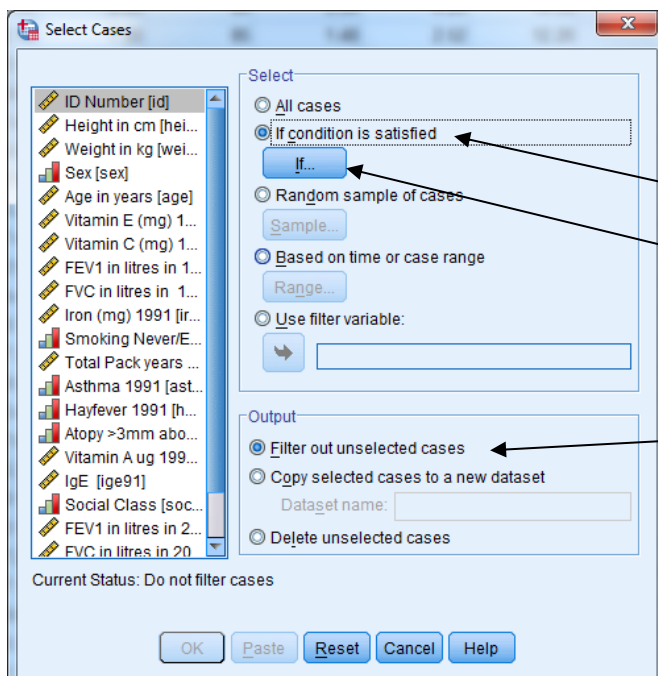
Having explored and summarised an outcome variable of interest in your whole dataset, you may wish to look at this variable in different subgroups, eg in men and women separately. One way of doing this is to select those members of your dataset you are interested in, eg men, and then carrying out the relevant procedure from section 2 on that subgroup. Alternatively, you can specify in SPSS that want to look at different levels of some factor or 'grouping variable', eg gender.

4.1 Using the Select Cases command

➤ To select out certain people for subsequent analysis, choose *Data, Select Cases...*



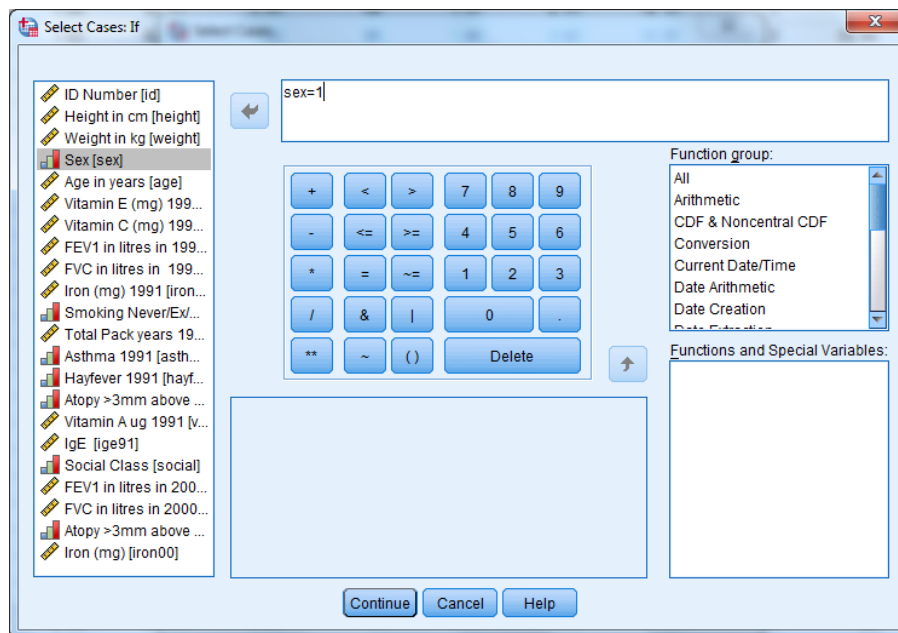
➤ The following box will appear.



Select 'If condition is satisfied',
then click on the 'If....' button

Make sure 'Filter out unselected
cases' is selected

- When you click on the 'If' button, the following box will appear. Here you need to type in the condition or function that defines the cases you want to include, eg if you just want males (coded 1), type `sex=1` in the top box on the right:



- If you want to select more than one category you need to use 'or' to select them, eg to select social class groups 1 and 2, type: `social =1 or social = 2`.
- Press continue and OK.

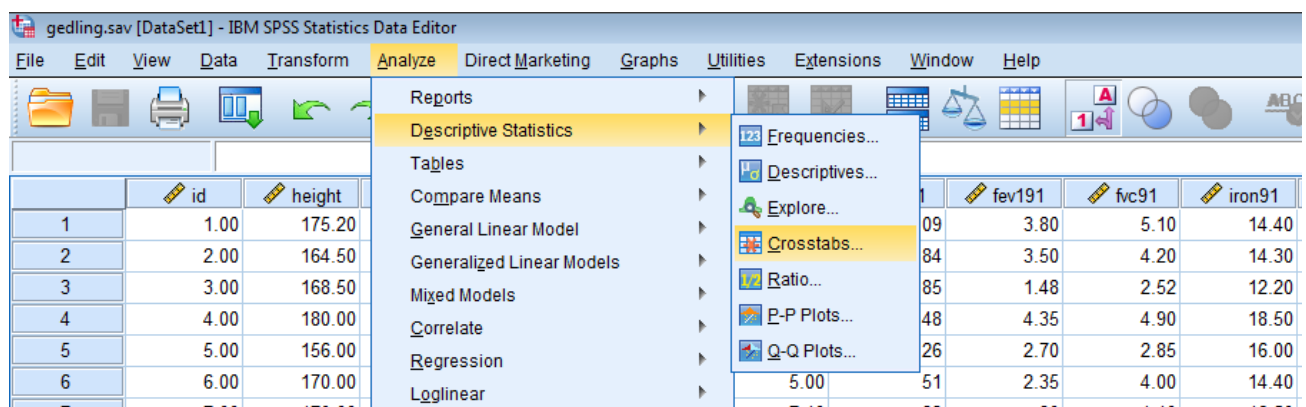
All subsequent analyses will then be carried out on this selected group. To reset back to the whole dataset, choose *Data, Select Cases..., All Cases, OK*.

4.2 Exploring a categorical variable in different groups

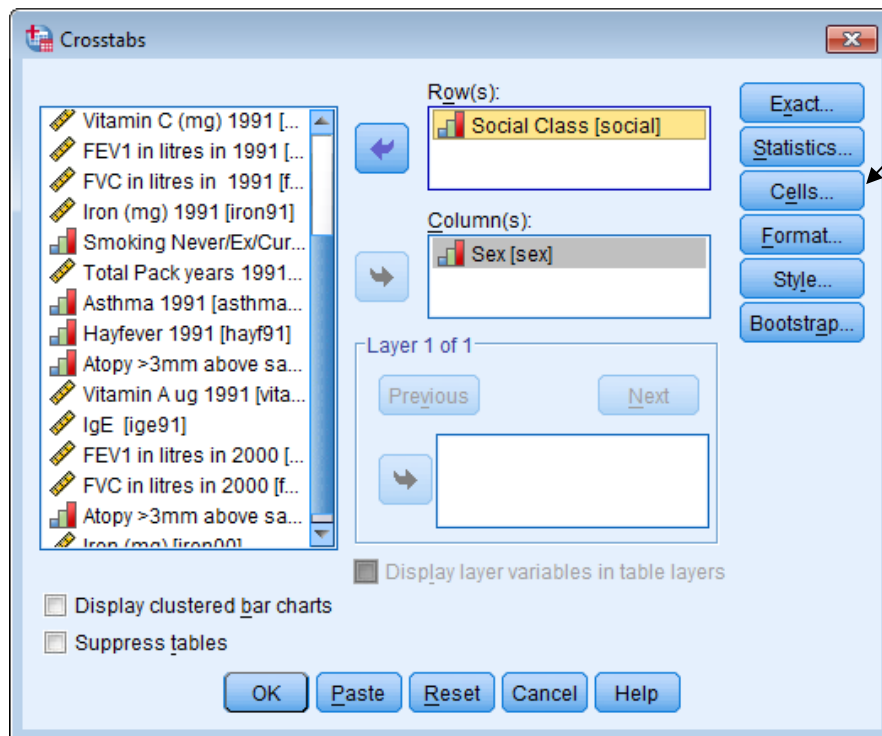
4.2.1 *Crosstabs command*

To look at the numbers (frequencies) and % of a categorical variable by each level of some grouping variable, eg gender:

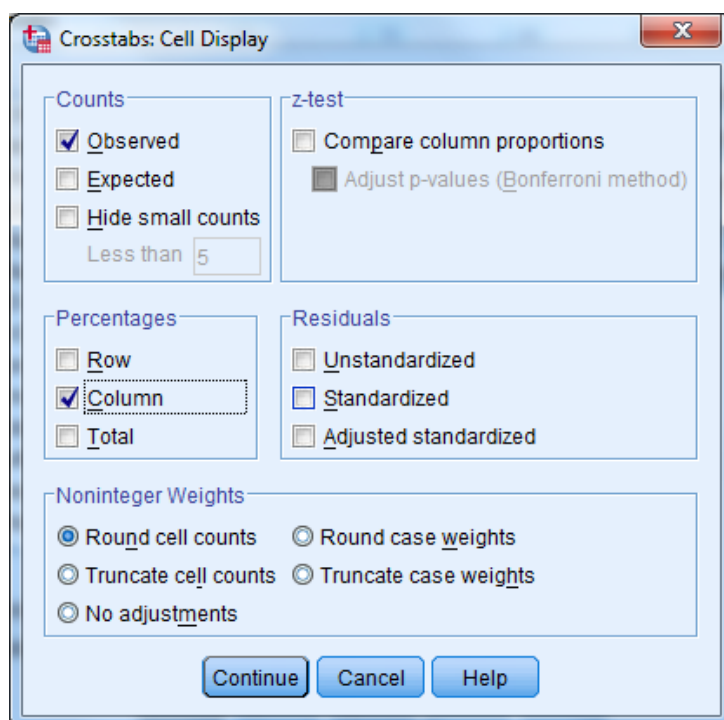
- *Analyse, Descriptive Statistics, Crosstabs...*



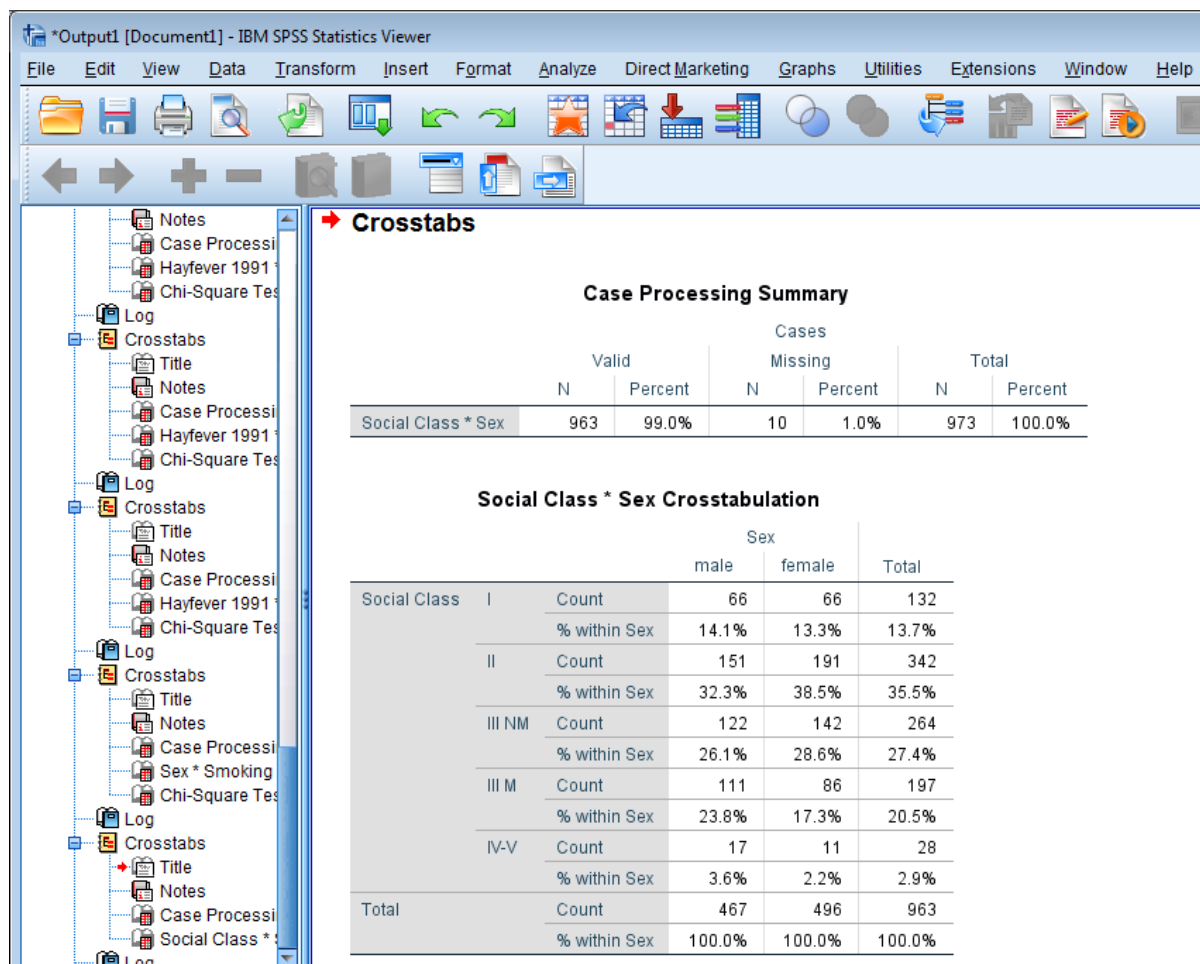
- Put one variable in the Row(s): box, and the other in Column(s):. Then click on 'Cells...'. .



- In the window that appears, tick 'Row' or 'Column' (as appropriate) under Percentages. Here sex is the grouping (exposure) variable, which we have put in the column box, so we tick column percentages.



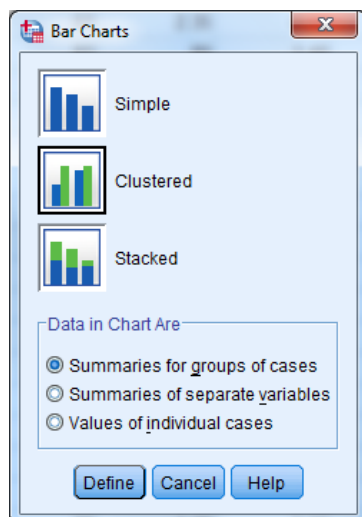
- Press Continue and OK. The table produced has the number and % falling into each level of the outcome (social class) by each level of the grouping variable (sex):



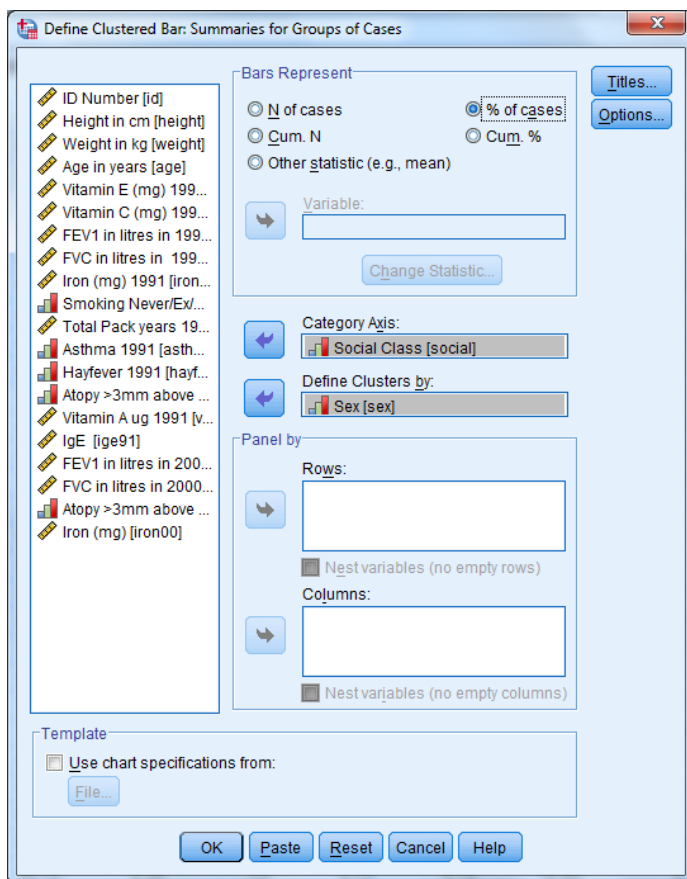
4.2.2 Clustered bar chart

To create a bar chart which displays the data by level of a grouping variable:

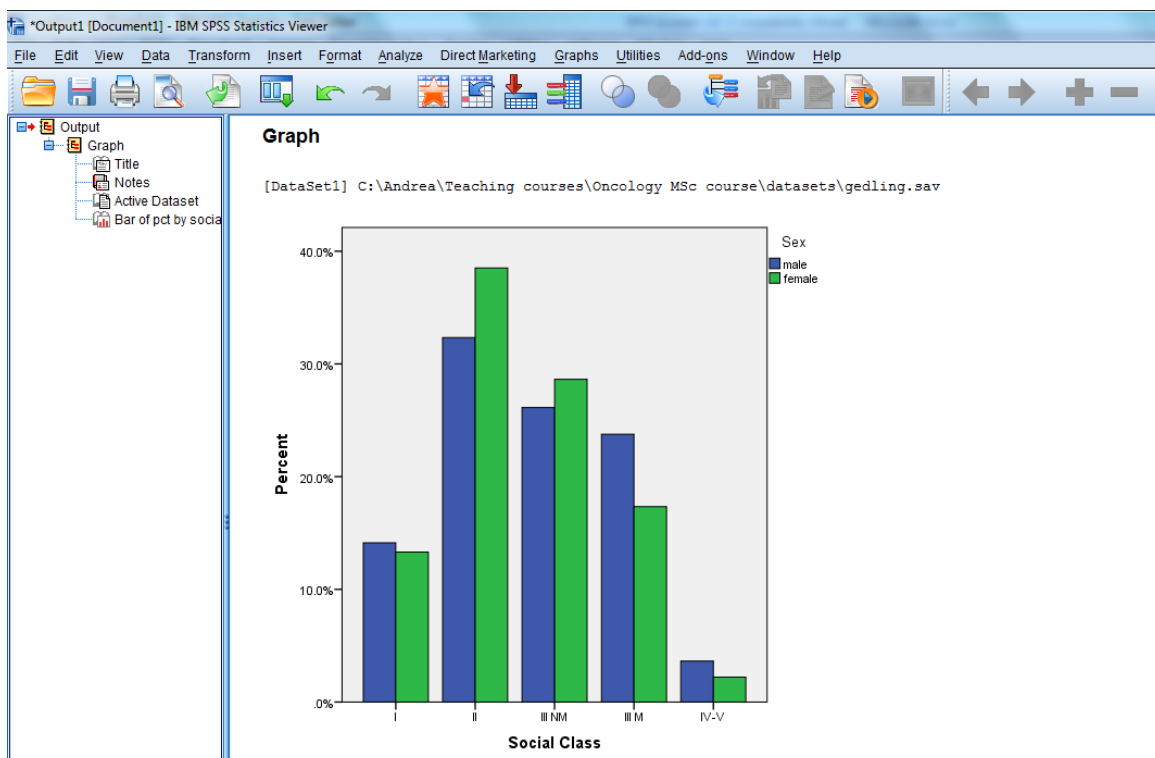
➤ *Graphs, Legacy dialogs, Bar...*



➤ Select 'Clustered', then press 'Define'. The following window will open:



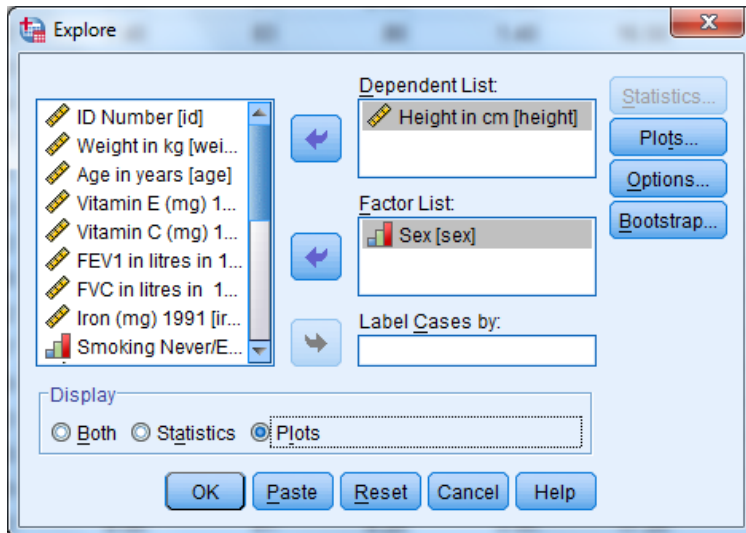
- Select '% of cases' under Bars Represent (because if you choose N of cases, it is difficult to make a direct comparison between the two distributions). Move the variable of interest (eg social class) into the 'Category Axis:' box, and move the grouping variable (eg sex) into the 'Define Clusters by:' box.
- Press OK. The following graph will be displayed:



4.3 Exploring a continuous variable in different groups

To explore the distribution of some continuous outcome variable in different groups, we use the explore command as before, but specify the grouping variable (factor).

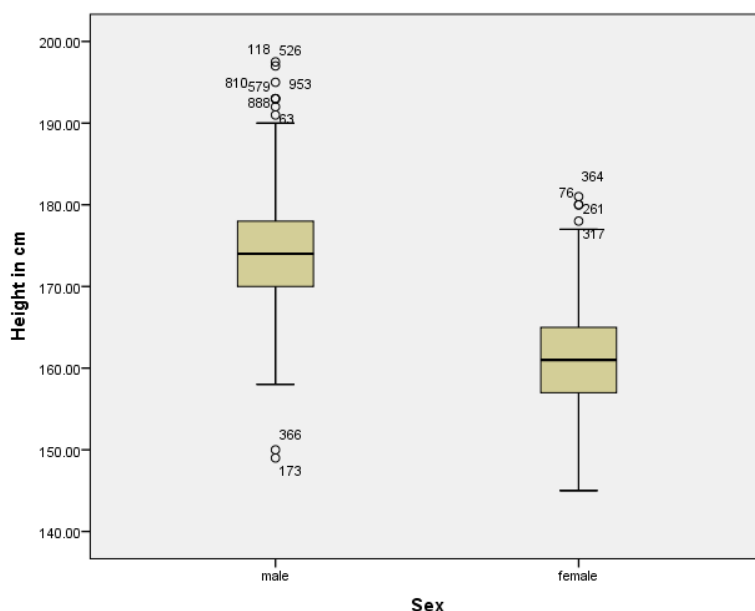
- *Analyse, Descriptive Statistics, Explore...*
- Move your continuous outcome variable into the 'Dependent List:' box and your grouping variable into the 'Factor List:' box



- Choose whether you want to display plots, statistics or both. To display a histogram, remember to also click on 'Plots..' and select histogram.
- Press OK.

The tables and graphs produced are the same as described in section 2.2, but are displayed for different levels of the grouping variable (eg men and women) separately.

The box plot produced shows the different groups on the same plot:

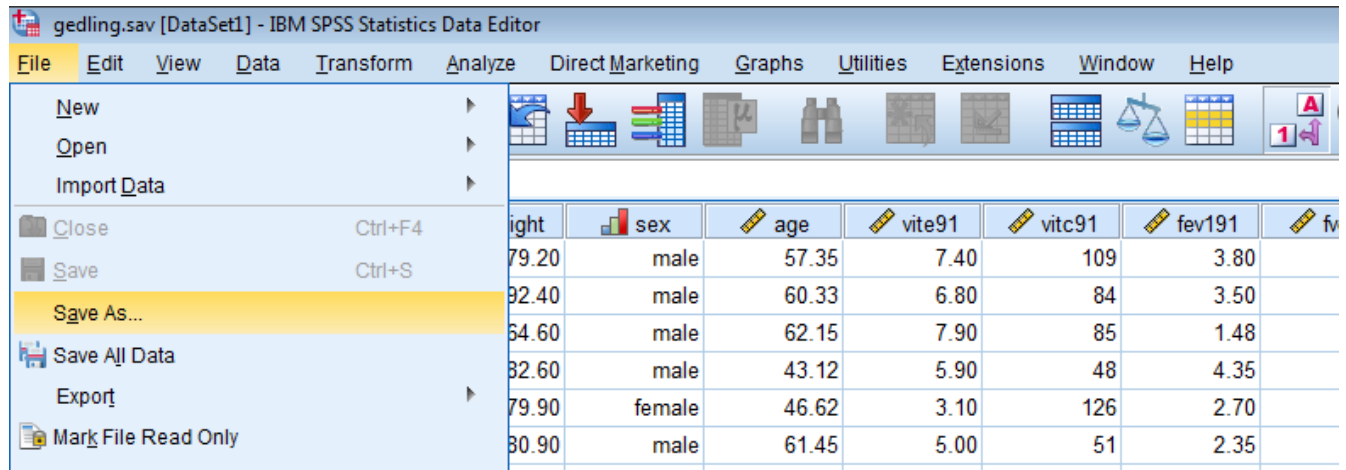


5. SAVING, EDITING AND PRINTING

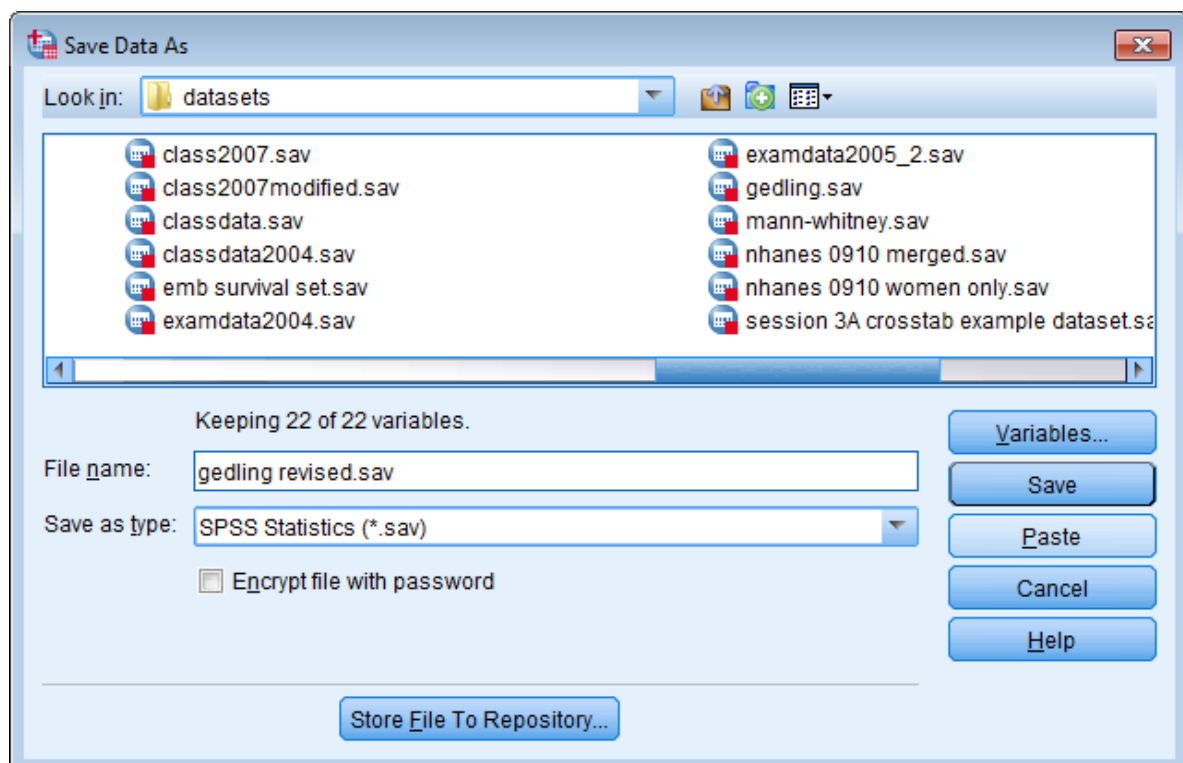
5.1 Saving data

If you enter data yourself, or make changes to an existing dataset (eg create new variables), you will need to **save** the new **data file**:

- Go to Data Editor window and select *File, Save As* from the menu:



- Type in a new file name and chose the directory where you want it to be saved:

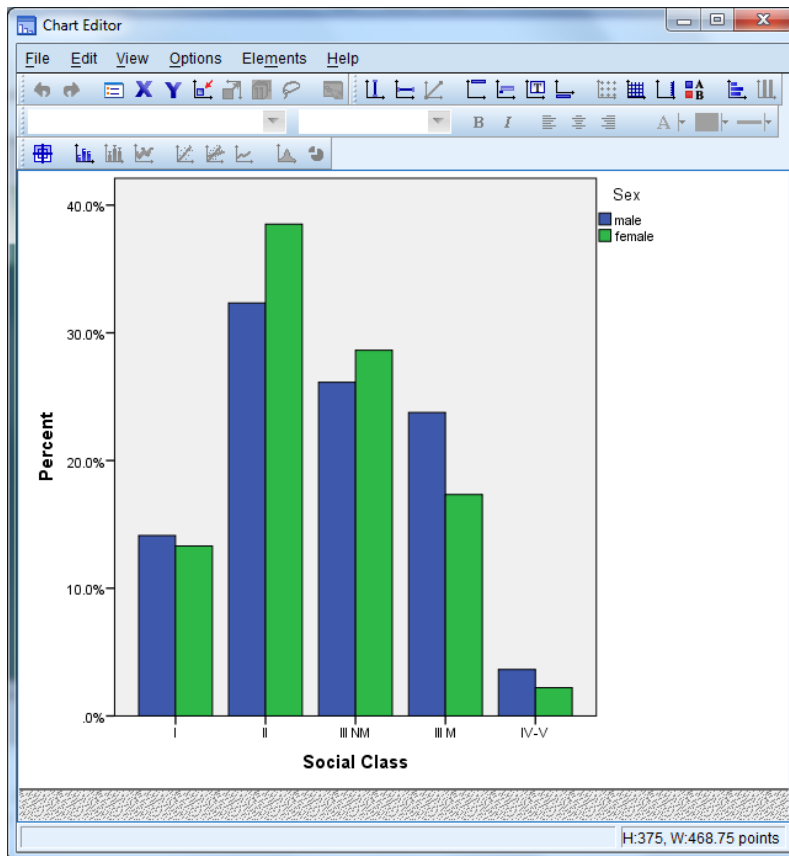


- Under 'Save as type:', select SPSS (*.sav) to save as an SPSS data file. It is also possible to save the data as a different file format (eg Excel or Stata) by selecting the desired file type.
- Click on *Save*.

5.2 Editing items in output window

You can change or edit any of the output in SPSS.

- Double click on the table or graph you want to edit and the object will open in a separate Chart Editor window.
- You can make changes by double clicking on the bit you want to change. Or in Chart Editor you can use the menus. Eg to change the axis scale of a graph, select *Edit* and then *Select Y axis*. Then make the changes you want to.



- When you have finished making changes to Chart Editor, select *File, Close*.

5.3 Saving output

Having finished a session in SPSS, you may want to save your output.

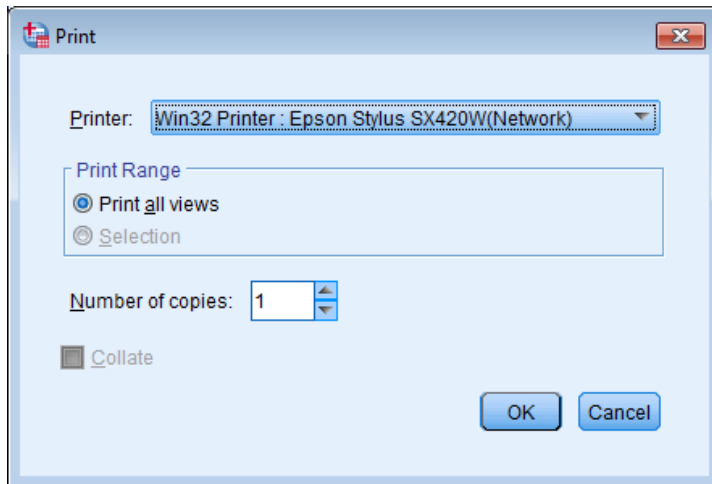
- Go to the output/viewer window
- Using the outline section, delete any output you do not wish to keep
- Select *File*, then *Save As...*
- Choose a directory, type in a file name, and save as .spv file.
- Click *Save*.

This file can then be opened in the output window at a later session.

5.4 Printing output from SPSS

To print your output from SPSS:

- Go to the Output window.
- Select *File, Print...*
- Under Print Range, choose 'Print all views' to print all output. Alternatively, highlight selected output and choose 'Selection'. Press OK.



5.5 Copying output to Microsoft Word

Instead of saving your output as an SPSS viewer file, you can choose to

- a) copy objects from the output and paste into Word, and then save as a Word file, or
- b) export the output as a Word file.

Once in Word, you can add text, print from word etc. Note that it is advisable to edit the output in SPSS (see section 5.2) before exporting/copying to Word. Once in Word, only limited editing is possible.

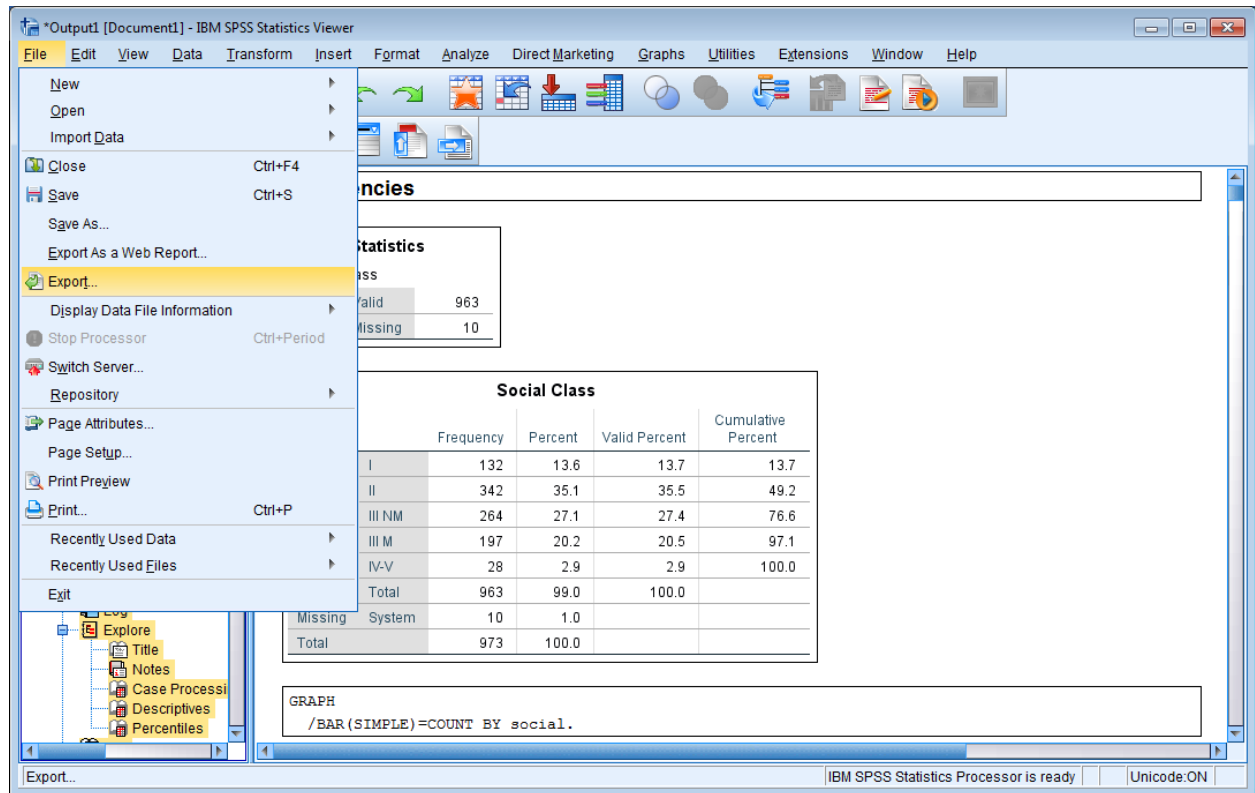
To copy individual objects from the output:

- Go to the Output window.
- Right click on the output you want to copy, either from the main output display on the right, or from the outline list on the left.
- Select *Copy* (or *Copy special* and choose the format you would like)
- Open a Word document and select *Paste*

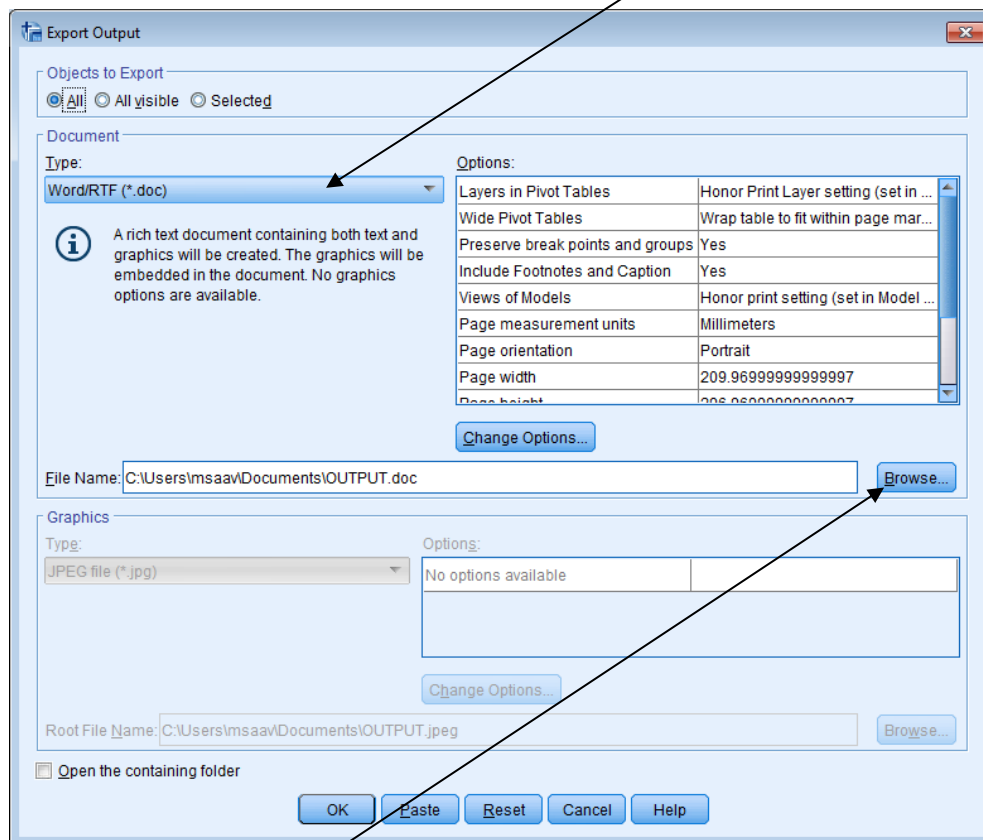
To copy all your output, or large chunks of it:

- Click on 'output' at the top of the outline list on the left to select all output. Or select parts of the output from the outline list by clicking on them whilst the 'Ctrl' key is pressed down.

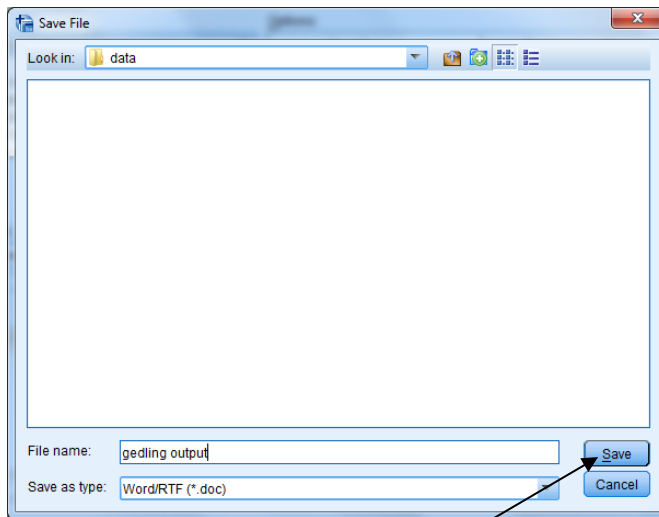
- Choose *File, Export...* from menu bar:



- In the window that appears, make sure Word/RTF is selected under Type.



- Then click on *Browse...* to find the directory you want your export file to go, and give the file a name.

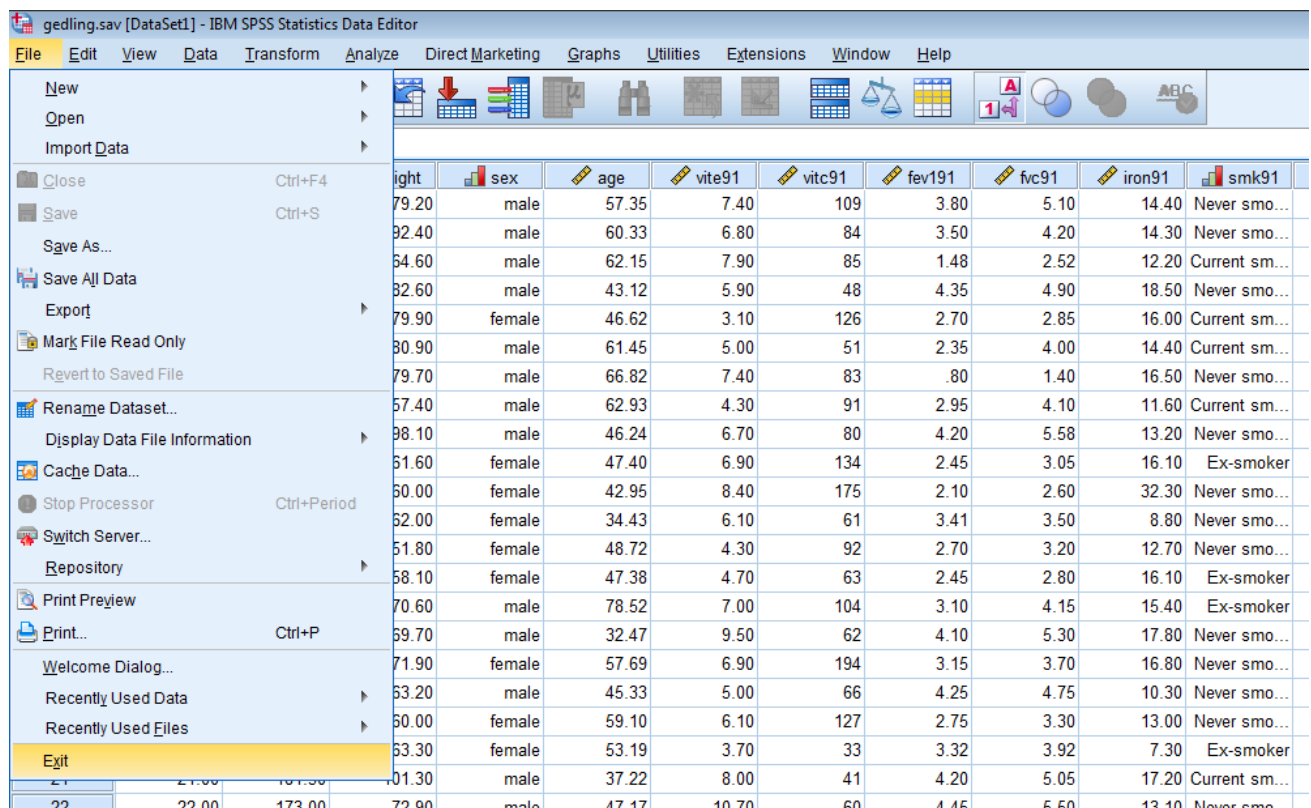


➤ Click Save, then OK.

This new file can then be opened in Word.

5.6 Exiting SPSS

To exit SPSS, choose *File, Exit*.



If you have made changes to your data or not saved your output before exiting, SPSS will ask you if you want to. Be careful when pressing Yes as this may overwrite the original dataset. Instead, press cancel and save as a different name (see 5.1).