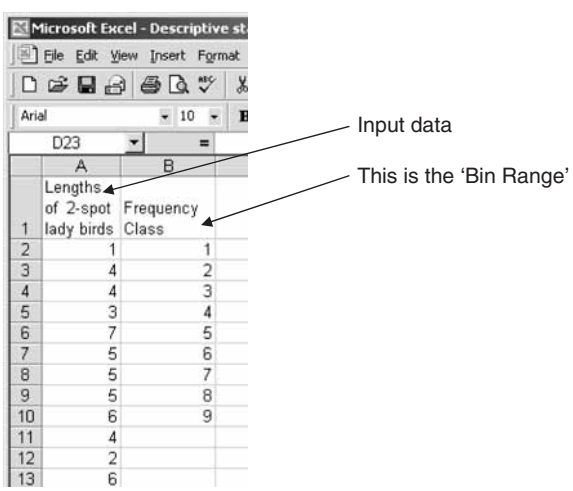
3.8.1. Are my data parametric?

EXAMPLE 3.2. Length (mm) of two spot ladybirds (*Adalia bipunctata*)

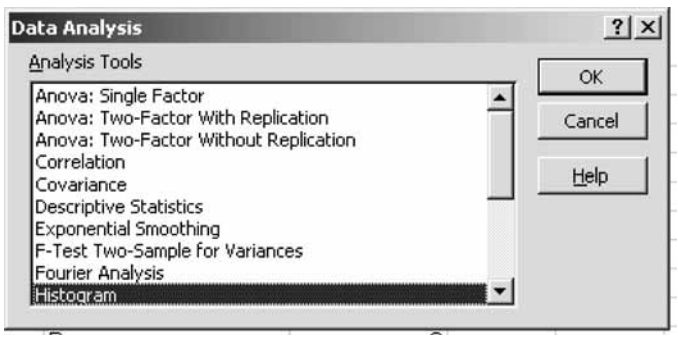
BOX 3.2. How to check if your data are normally distributed (parametric)

Step 1. Open a new worksheet by selecting 'File' from the top tool, then, 'New'. Enter the data into the worksheet, either in an array or in one long column. The latter is probably more convenient and a label can be put in the top cell.

In the next column the 'bin range' must be entered. This represents the frequency classes into which you wish to organise the data. For example, here we want to know how many of the data have a value of 1, of 2, of 3, etc. In Excel, these are known as a 'set of boundary values that define bin ranges'. These values are arranged in ascending order.



Step 2. From the top tool bar, select 'Tools', then 'Data Analysis' from the drop down menu. The 'Data Analysis' box will open. Select 'Histogram'. Click 'OK'.



Step 3. In the 'Histogram' dialogue box, enter the input data.
Ensure that the Input Range box in the Input box is highlighted. Input the cell references of the data by clicking on cell A1 (this includes the data or sample label) and dragging down the column to the last cell which contains data. The area on the spreadsheet will now be highlighted and the cell references shown in the input box.

Step 4. Click in the box marked Bin Range: and repeat the process with the second column of data containing the frequency ranges.

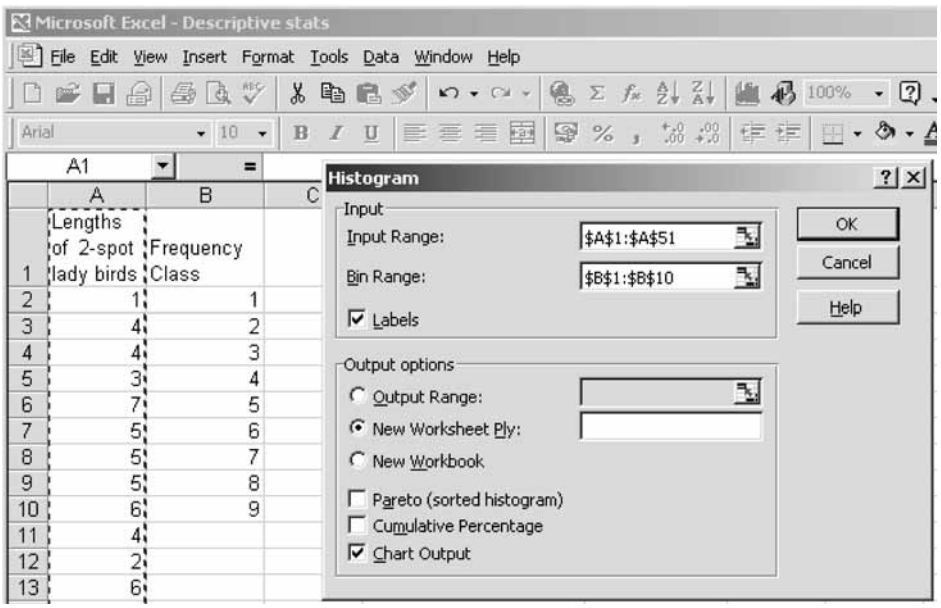
Step 5. The box marked Labels should be clicked, this will put a tick in the box which shows that the first cell for each data set contains a label and not data. If this box is not ticked, Excel will treat the material in the first cell as data and will not be able to complete the calculation. Note that it is useful to use the labels to identify your data. These labels are used by Excel to identify the output data.

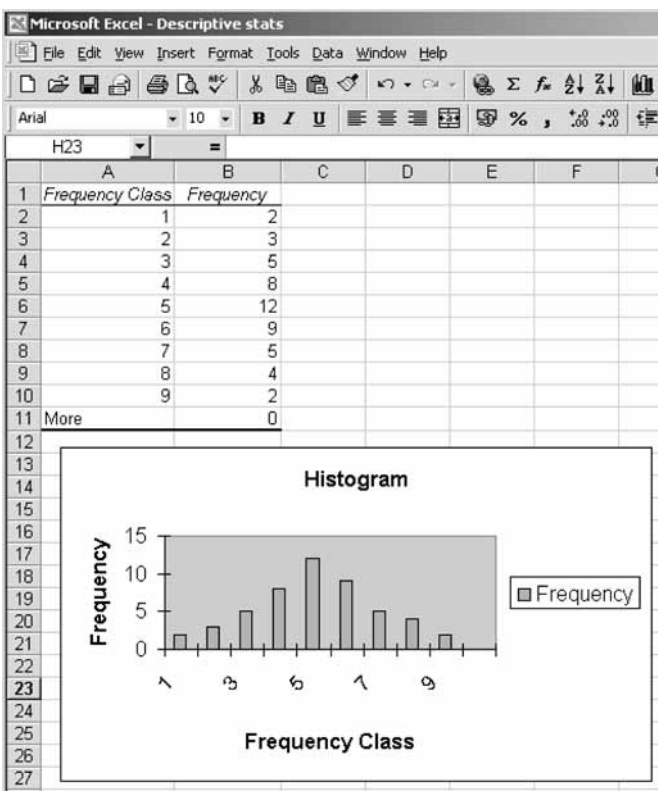
Step 6. Next, select the output options. To return the output data below the input data, select 'Output Range:' and then click in the box (the cursor will now flash in the box). Scroll over an area where you want the results to be displayed. Note that you could just select a couple of cells—Excel will determine the actual size that it requires for the results table. Note too, that it is essential to click in the box as well as selecting the 'Output Range' button. If this is not done the location is entered into the 'Input Range' box and the analysis cannot be completed.

Step 7. You can choose to have the results entered on to a 'New Worksheet Ply', in which case the results will be given on a fresh sheet, accessed by the tabs at the bottom of the current sheet. Alternatively a 'New Workbook' can be selected. We have opted to put the results table beneath the original data on the spreadsheet.

Step 8. Tick the box for Chart Output.

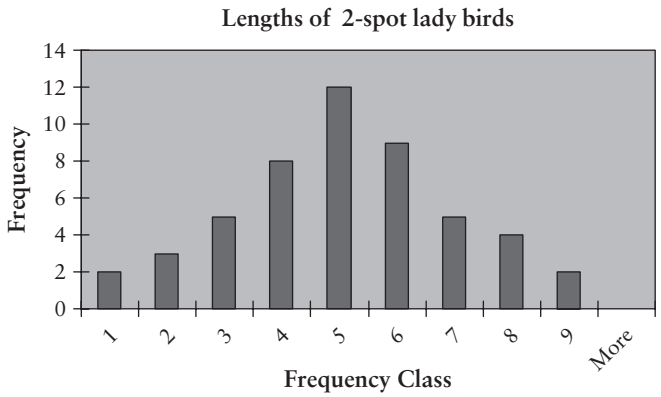
Step 9. Click OK and the Results table will be returned together with a chart.


Microsoft Excel - Descriptive stats
File Edit View Insert Format Tools Data Window Help
Arial 10
A1
Lengths of 2-spot lady birds
Frequency Class
Histogram
Input
Input Range: $A$1:$A$51
Bin Range: $B$1:$B$10
Labels
OK
Cancel
Help
Output options
Output Range:
New Worksheet Ply:
New Workbook
Pareto (sorted histogram)
Cumulative Percentage
Chart Output


Microsoft Excel - Descriptive stats
File Edit View Insert Format Tools Data Window Help
Arial 10
H23
Frequency Class
Frequency
More
Histogram
Frequency
Frequency Class
Frequency

Step 10. The output data shows the frequency classes and the frequency of observations relating to each class. 'More' indicates possible additional frequency classes beyond those given, but here the frequency in this class is zero, i.e. there is no additional data beyond the given frequency classes.

The chart is returned in a compressed form and needs to be enlarged. Click on the chart to highlight it and drag it to a suitable location. Click on a highlighted black square in one of the corners and drag to expand the chart. Any part of the chart may be edited by clicking on it to highlight it. Then click the right mouse button to bring up a menu.



Step 11. From the descriptive statistics (see Box 3.1), we note that the mean is 5.08 mm, the median and mode are both 5.0 mm, and the standard deviation is 1.936 mm.

Lengths of 2-spot lady birds	
Mean	5.08
Standard Error	0.27380911
Median	5
Mode	5
Standard Deviation	1.936122782
Sample Variance	3.748571429
Kurtosis	-0.293549691
Skewness	-0.047184265
Range	8
Minimum	1
Maximum	9
Sum	254
Count	50
Confidence Level(95.0%)	0.550239673

The values for central tendency are confirmed by the chart.

Step 12. Go through the individual tests to test for 'normality'.

(a) Is the data on an **interval** scale? Yes—it is measured in mm.
(b) Does the distribution appear to be a 'bell' shaped curve? Looking at the graph, the answer is 'yes'.
(c) Do about 68% of your observations fall within the range $\bar{x} \pm 1$ s?

The mean is 5.080 mm, and the standard deviation is 1.936 mm.

$\bar{x} + 1 \text{ s} = 5.080 + 1.936 = 7.016$

$\bar{x} - 1 \text{ s} = 5.080 - 1.936 = 3.144$

The measurements between these are 4, 5, 6 and 7 mm. From the frequency table relating to these frequency classes note that we have frequencies of 8, 12, 9, and 5 respectively. This represents a total of 34 ladybirds. This is $(34/50) \times 100 = 68\%$ of the total. Thus about 68% of the observations lie in the range $\bar{x} \pm 1$ s.

(d) Does **median** = **mode** = **mean**?
Median = 5.0 mm, mode = 5.0 mm (from the graph), and mean = 5.080 mm. Thus they are all close together.

From the information we can obtain at this stage, it looks as if the data is Normally distributed.