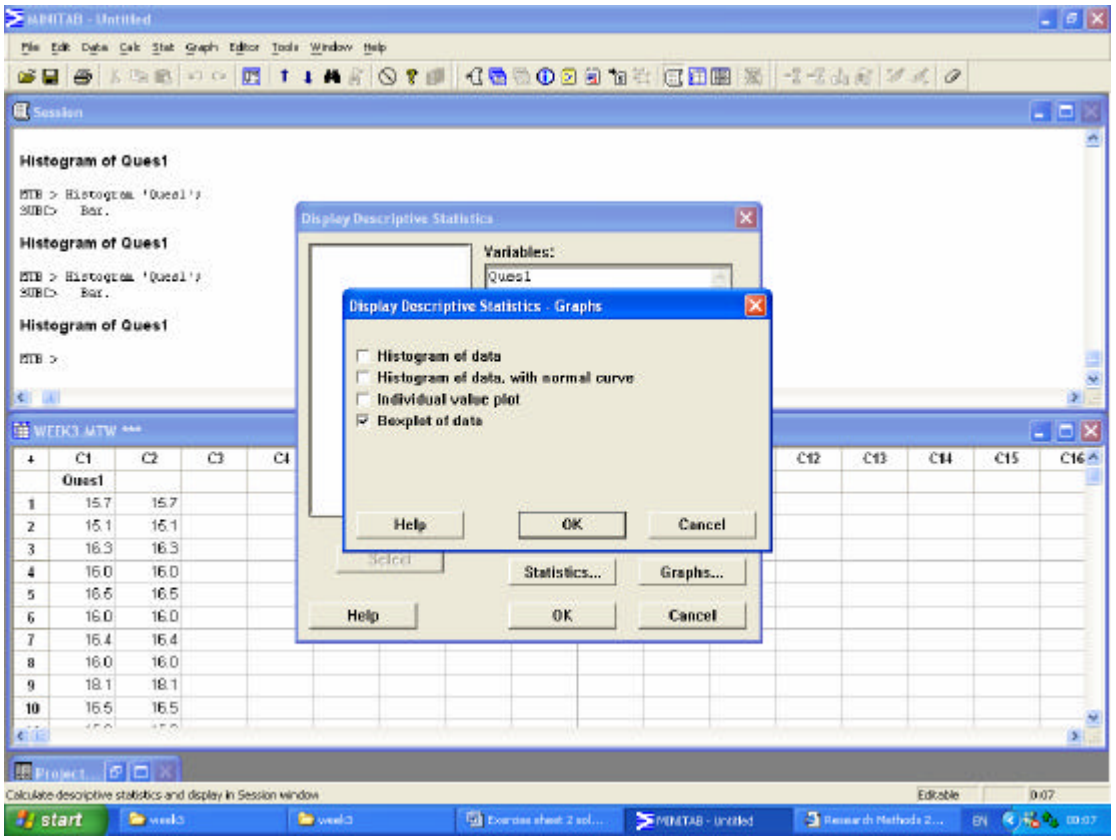
Research Methods 2

Week 3: Exercise Sheet 2, numerical summaries

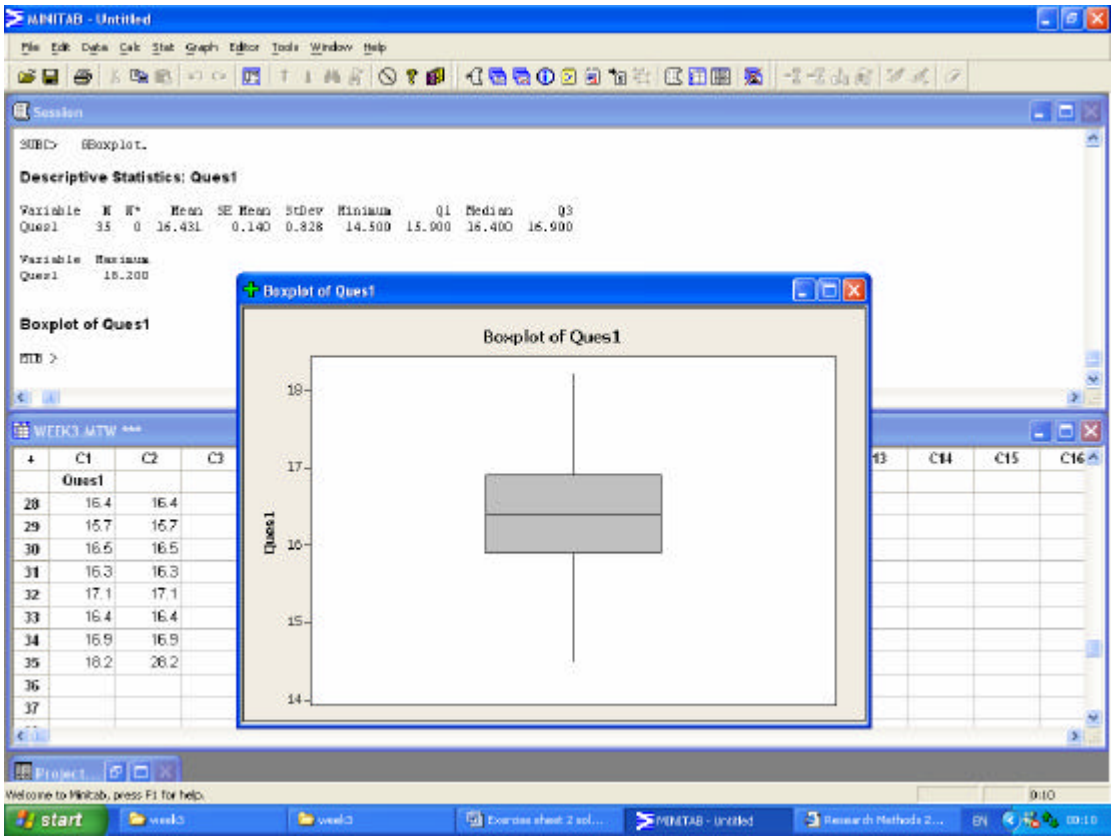
Solution sheet

Question 1.

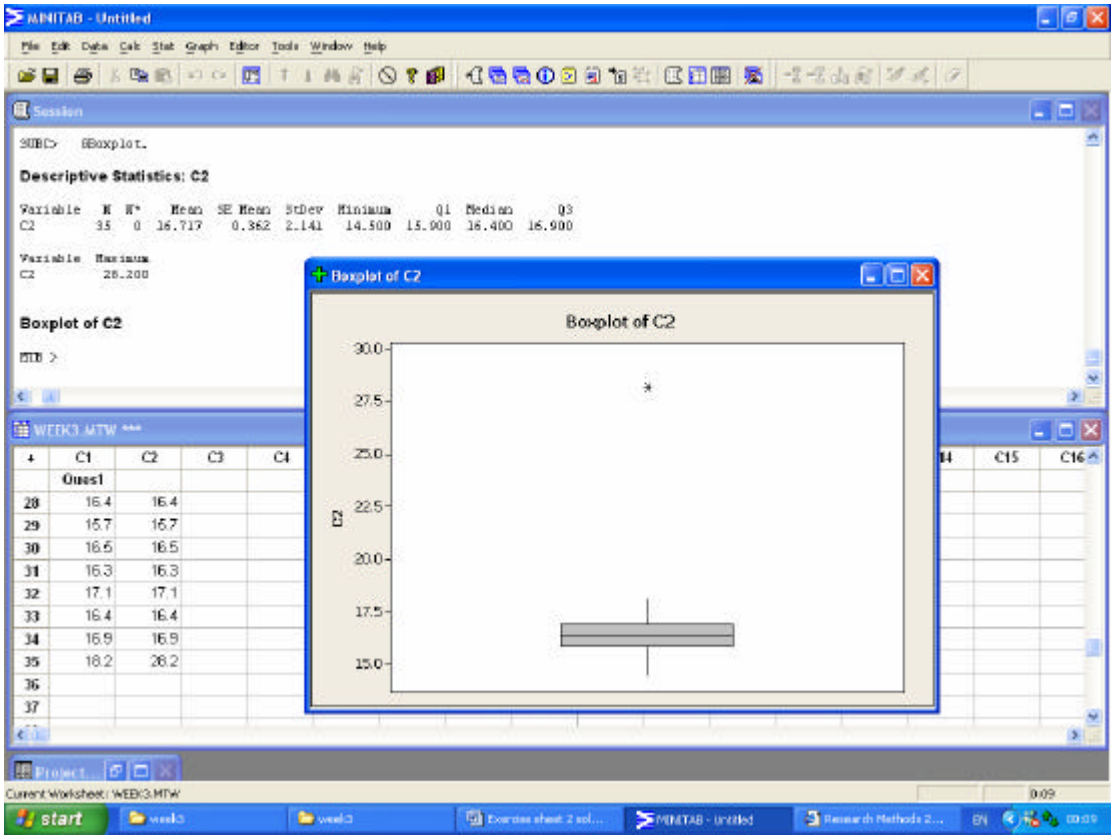
First select **Basic Statistics** and then **Display Descriptive Statistics...** from the **Stat** menu and double-click on 'Ques1' to place 'Ques1' in the **Variables:** box. Then select the **Graphs...** box and you will obtain the following screen.



As shown, the **Boxplot of data** item should be checked and **OK** clicked successively. This will give the result in the next screen. The line for the median is at 16.4 g/dl and the lower and upper quartiles are at 15.9 and 16.9 g/dl respectively No points are plotted individually, and the whiskers extend to the maximum and minimum values in the sample, namely 14.5 and 18.2 g/dl/



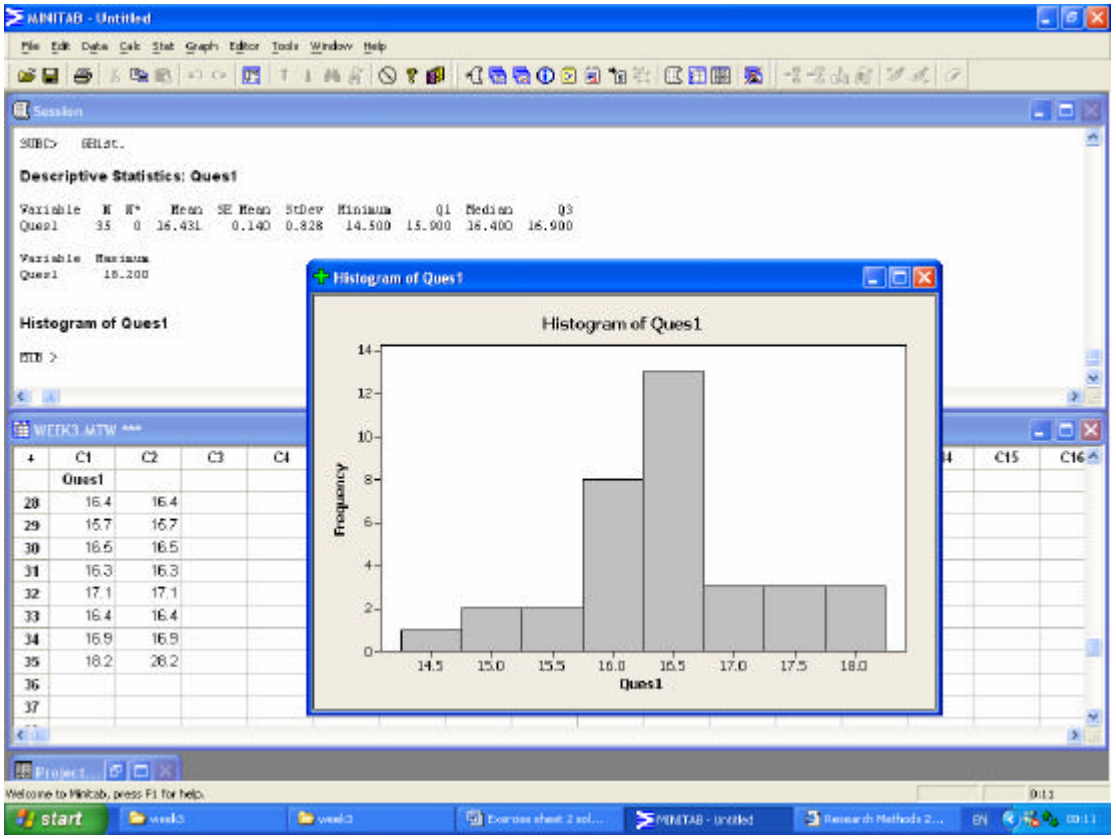
If 'Ques1' is copied to C2 and the largest value changed from 18.2 to 28.2 g/dl then the following screen is obtained.



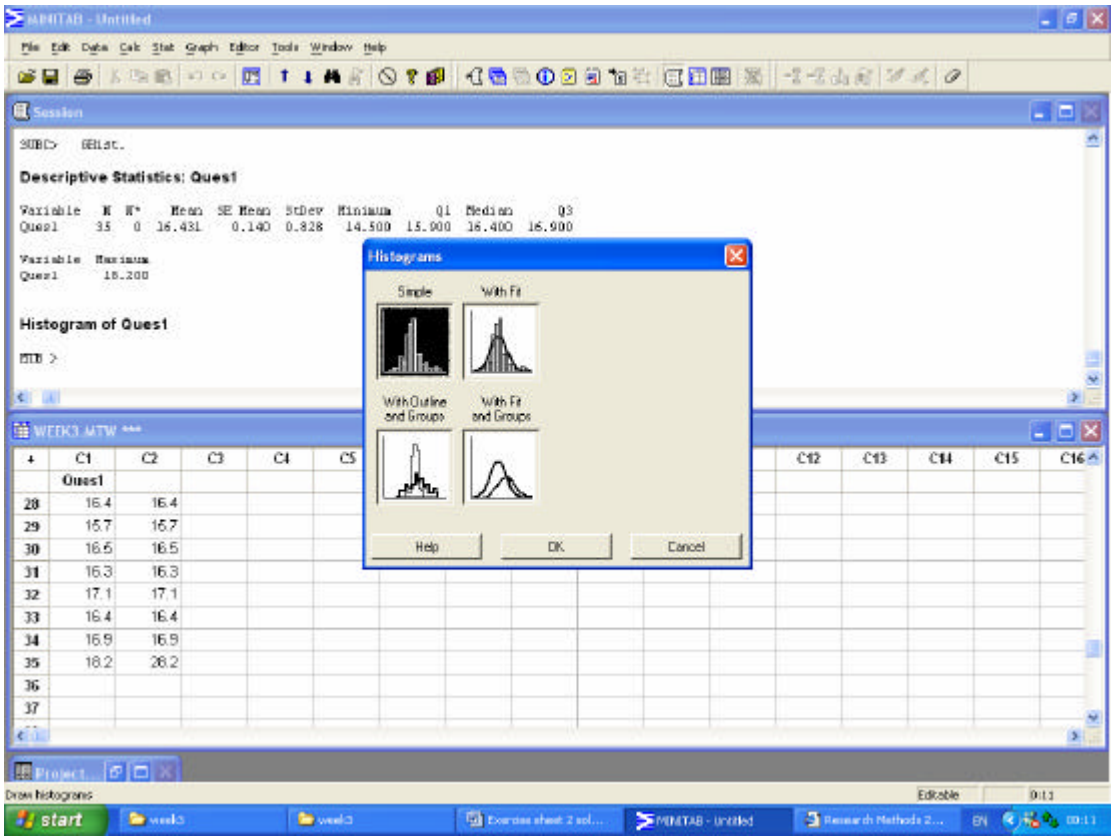
The box and the lower whisker remain unchanged but, as the largest point is now much further from the main part of the sample, the whisker stops at the second largest value and the largest value is plotted individually.

Question 2.

A histogram of the data in 'Ques1' can be obtained most simply by following the method that was used in question 1 to obtain a boxplot, but checking the item **Histogram of data**, rather than **Boxplot of data** in the first screen shown above. This will lead to the following display.

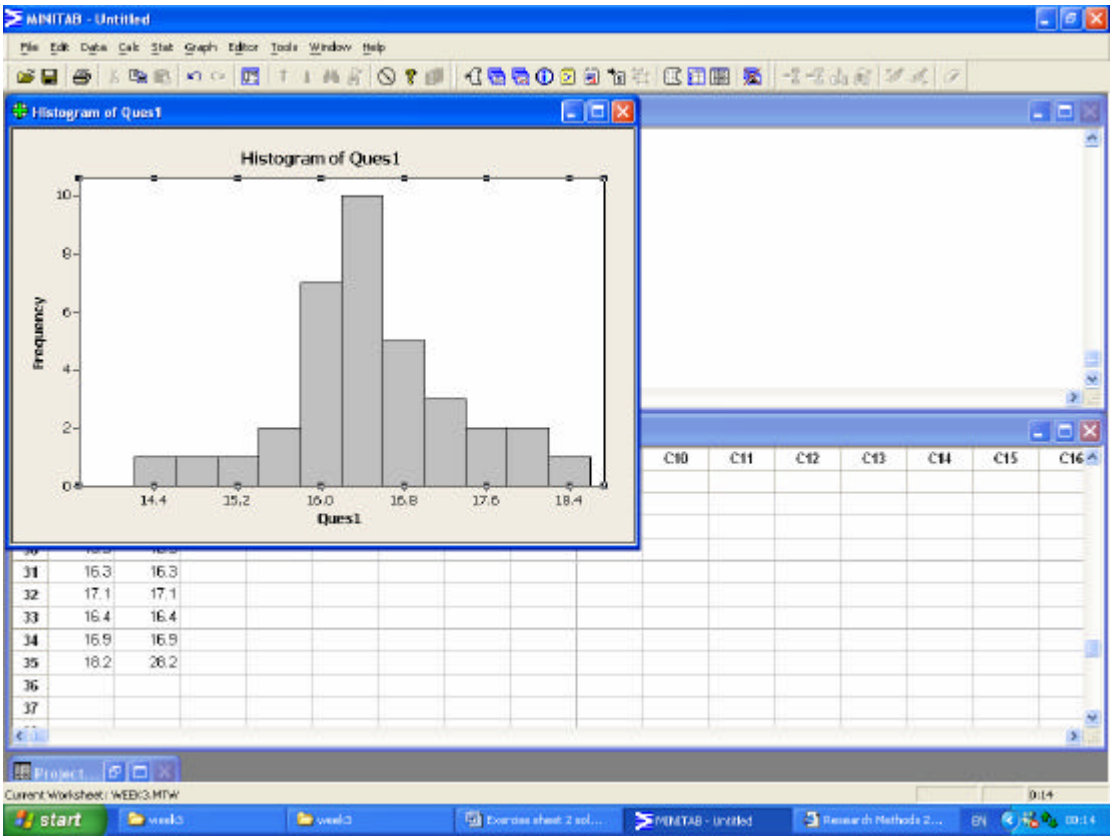


An alternative approach is to select **Histogram...** from the **Graph** menu, which would present you with the following screen.



This shows the screen with the thumbnail headed 'Simple' selected. Now click **OK** and then double-click on 'Ques1' in the left-hand box to move it to the **Graph variables:** box. Clicking on OK gives the same histogram as before.

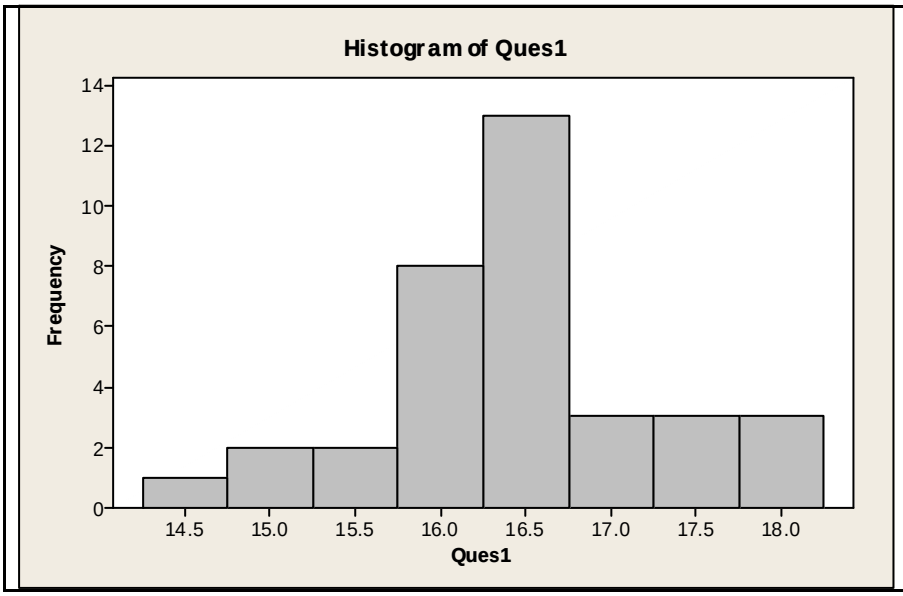
If this second approach is employed then the number of intervals used by Minitab to construct the histogram can be controlled by the user. To do this, double clock on the X-axis (i.e. the horizontal axis) of the histogram and select the Binning tab in the screen that ensues. The button for the number of intervals and a box into which you enter the number required is then obtained. Successively clicking on **OK** will lead to the following screen:



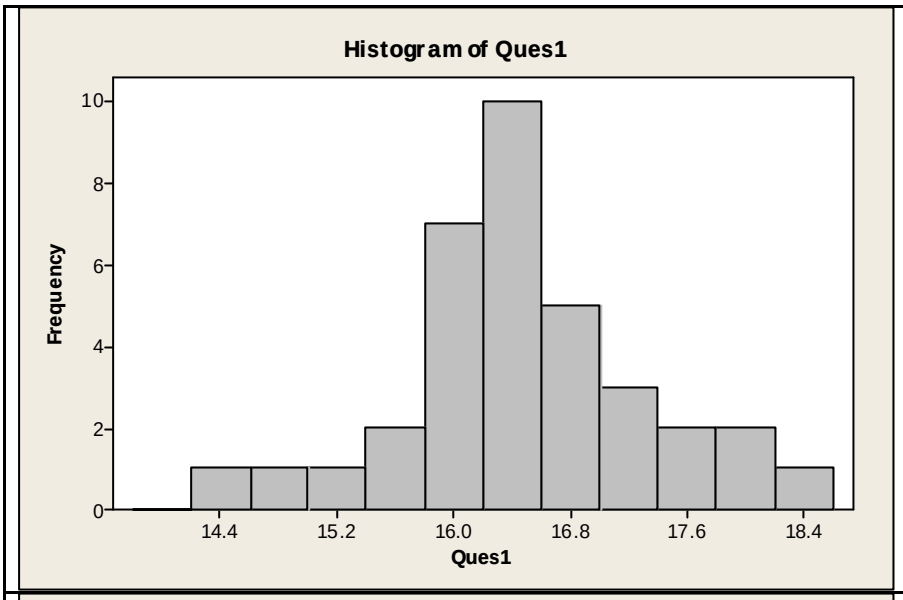
This shows the screen after the item **Number of intervals:** has been checked and then you have clicked in the box and typed the number 12.

The marks around the frame of the histogram can be removed by clicking anywhere in the region of window of the histogram outside the frame, e.g. near to the axis label.

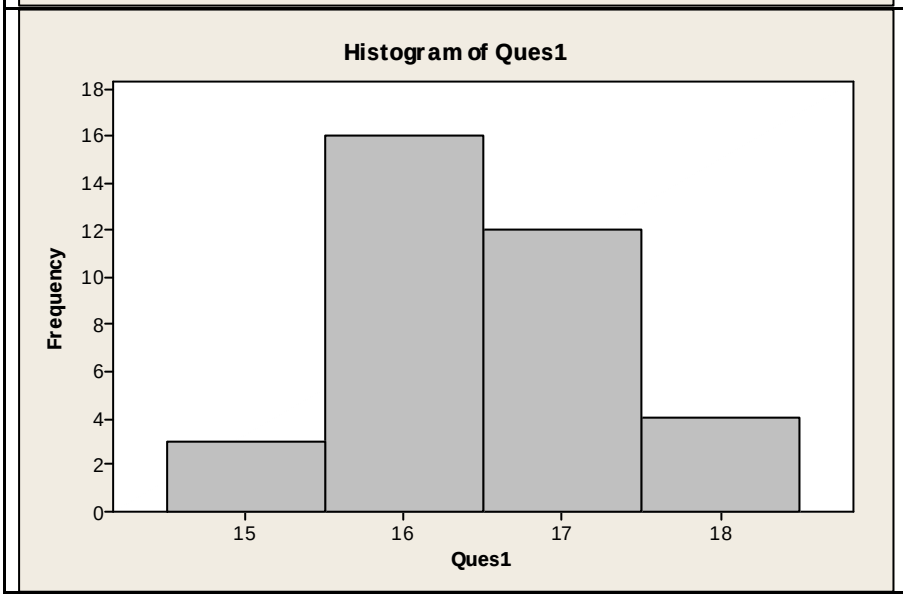
The results for the default (8 intervals) and 4 and 12 intervals.



8 intervals

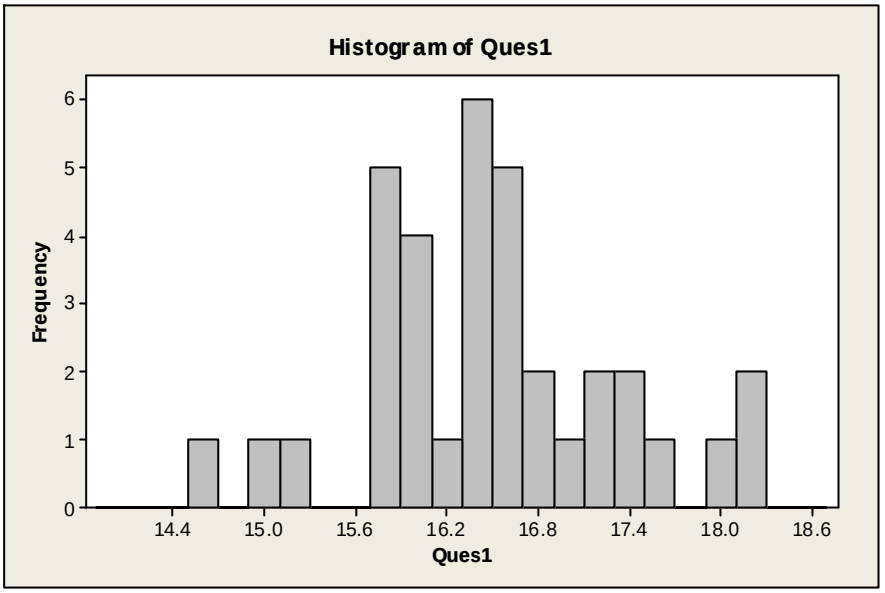


12 intervals



4 intervals

These histogram of the same sample, but using different numbers of intervals, show that as the number of intervals is increased a more detailed picture of the distribution of the data is obtained. If the number of intervals is too small, then the histogram gives a rather uninformative picture of the data. However, it should not be thought that the number of intervals should be made as large as possible. The effect of using too many intervals is shown below, where 24 intervals have been used.



This histogram has a more jagged outline, which does less to reveal the pattern of the distribution of the data than the histograms with 8 or 12 intervals.

In practice there is no 'right' answer to the question of how many intervals to use and a degree of judgment and common-sense must be applied.

End of solution sheet