Check-Up

5

Bar charts for grouped discrete data

A teacher recorded the results of a science test (marked out of 80), for a year group of 84 pupils, in an Excel worksheet. This generated the bar chart shown here. Are these statements true or false?

a) The graph shows that the highest mark obtained in the test was 80.

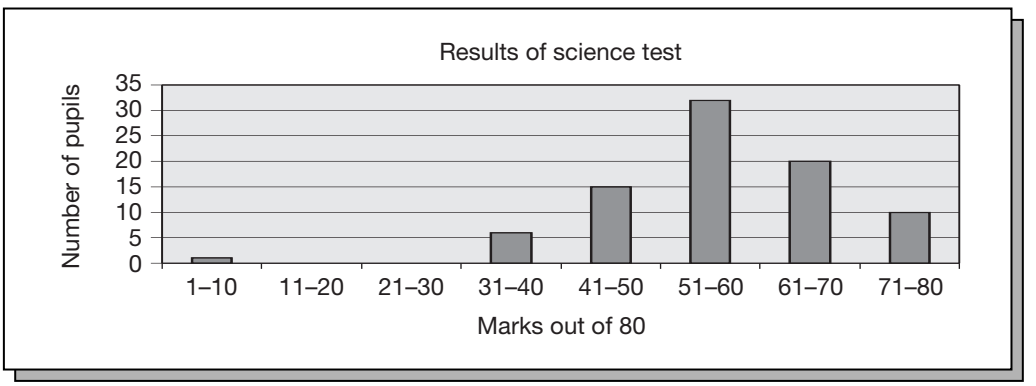
b) 15 pupils achieved marks in the range 41 to 50 inclusive.

c) No pupils scored marks in the range 11 to 30 inclusive.

d) 30 pupils scored more than 60 marks out of 80.

e) Fewer than 10% of the pupils scored less than half marks.

f) One pupil did very much worse than all the rest.



Answer to check-up 5

a) False. b) True. c) True. d) True. e) True. f) True.

Discussion and explanation of check-up 5

The variable in this example is simply the mark out of 80 achieved by pupils in the science test. This is a discrete variable, since it separates the set of pupils into separate subsets, i.e. those with scores of 0, 1, 2, 3, 4, 5, 6, and so on, up to 80. But clearly there are too many of these subsets to represent in a bar chart. The resulting diagram would convey nothing that you could not see better by just looking at the marks themselves. In a case like this, where the variable takes numerous values, it is usual to group the values into intervals covering a range of values. The width of each interval should normally be the same. In this example it was decided to go for intervals covering a range of 10 marks, thus producing eight possible subsets across the whole range of marks obtained by the pupils: 1–10, 11–20, 21–30, and so on. The size of the intervals should be chosen to produce from five to twelve subsets: fewer than five and too much information is lost, more than twelve and too much detail is retained for the graph to be of any use.

So in the graph, each column represents the number of pupils whose marks fall within the given range. The graph loses the precise details of the individual marks, so, for example, we can read off that 10 pupils scored from 71 to 80, but we do not know if anyone actually scored 80. Notice that it is important in covering the whole range to include in the graph any intervals in which there were no pupils. The absence of pupils scoring in the intervals 11–20 and 21–30 is significant information, given the fact that one pupil scored a mark in the range 1–10. This pupil's poor result compared with the rest is immediately communicated by the pictorial representation of the data in the bar chart. You can also see at a glance that very few pupils scored 40 marks or less, with only a small part of the population appearing in the left-hand half of the graph. (In fact, the total here is only 7, which is less than 10% of the whole set of 84 pupils.) The whole purpose of putting numerical data into graphical form is to make it possible to take in at a glance the way the data is distributed across the range of possible values. If the teacher looking at this graph can get an immediate impression of how the pupils have done on the test, then the bar chart has fulfilled its purpose.

Summary of key ideas

- If a set of numerical data contains numerous different values of a discrete variable then it is often useful to group the data into intervals before representing it in a bar chart.
- The intervals should be of equal width, should be chosen to cover the whole range, and should produce from five to twelve subsets.
- In a bar chart for grouped discrete data, the height of a column indicates the frequency of values that fall within the interval shown.

Further practice

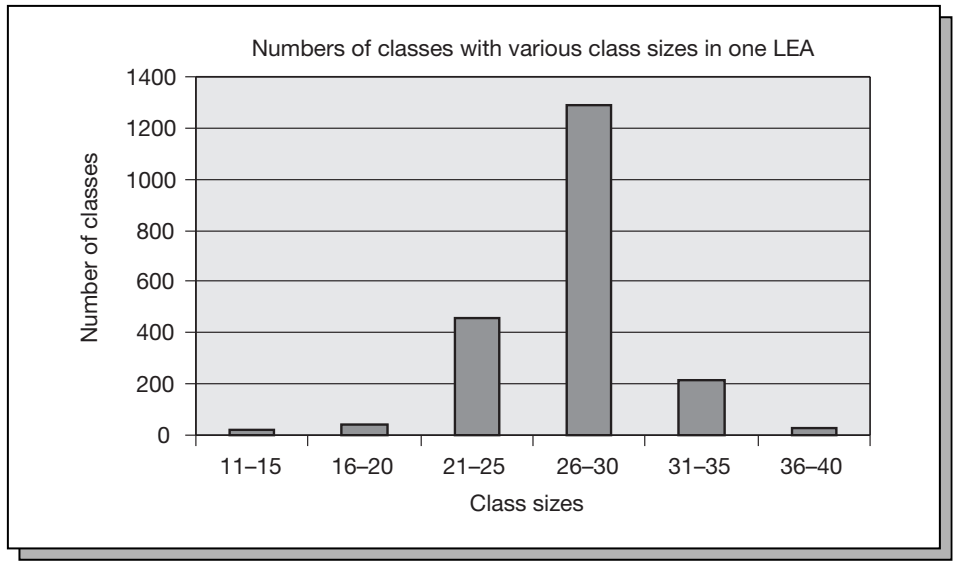
5.1 Which of the following sets of data would probably need to be grouped into intervals before being represented in a bar chart:

a) the reading ages of the pupils in a Year 2 class, expressed in years and months (e.g. 5.2, 5.11, 7.2)

b) the number of GCSE subjects passed at grade C or above by pupils in a Year 11 class

c) the total points score for pupils in a Year 11 class based on their GCSE/GNVQ grades for all subjects (where grade G = 1 point, F = 2 points, and so on).

5.2 A local education authority, concerned about the variation in class size across the LEA, produced this bar chart.



a) About how many classes have over 35 pupils?

b) About how many classes have 20 or fewer pupils?

c) Approximately how many classes in total are represented here?

d) The LEA's target is to have from 26 to 30 pupils in 75% or more of their classes; have they achieved this?