Here the highest mark represents the best student in a subject. But notice that some subjects have higher marks than others.

The second set have been adjusted – *scaled* scores. Here the highest mark gets 100 and the lowest 0 – a line is drawn and the other marks are read from the line. This is quite an acceptable technique. Now we notice that the total scores are the reverse of the raw scores.

The third set are based on *position* – with the highest mark getting 1 and the lowest 12. Now all the students have a score of 52!

"Lies, damned lies, and statistics"
Benjamin Disraeli

There are many other ways to "lie with statistics" such as *false percentages, misleading graphs, comparing unlike data, etc.*

So, we must be careful when presented with statistical data.

Some questions we can ask are:

- Who is presenting the data? Are they knowledgeable in the area? Are they reliable and unbiased?
- How did they get their data? Who filled in the questionnaire or survey? Was it a random sample? Were the questions written in such a way that the answers were likely to be true?
- Is any relevant information missing? Are we only seeing what the person wants us to see? Is there another side to the story?
- Is the finding consistent with previous reports? Can we believe it?

1 What can you do to prove that a certain statistical result is trustworthy?

2 Are statistical results more likely to be true in certain areas of knowledge than in others?

Past examination questions for topic 6

Paper 1

1 This table shows the age distribution of the teachers who smoke at Laughlin High School.

Ages	Number of smokers
$20 \leq x < 30$	5
$30 \leq x < 40$	4
$40 \leq x < 50$	3
$50 \leq x < 60$	2
$60 \leq x < 70$	3

a Calculate an estimate of the mean smoking age.
b Draw a histogram to represent the data. *M00q9*

2 This stem and leaf diagram gives the heights in cm of 39 school children.

Stem	Leaf
13	2, 3, 3, 5, 8
14	1, 1, 1, 4, 5, 5, 9
15	3, 4, 4, 6, 6, 7, 7, 7, 8, 9, 9
16	1, 2, 2, 5, 6, 6, 7, 8, 8
17	4, 4, 4, 5, 6, 6
18	0

Key 13 | 2 represents 132 cm

a i State the lower quartile height.
ii State the median height.
iii State the upper quartile height.
b Draw a box and whisker plot of the data. *Spec05q7*

3 These are the heights of some sunflowers in cm.

180 184 195 177 175 173 169 167 197 173 166
183 161 195 177 192 161 165

Represent the data by a stem and leaf diagram. *Spec05q29*

4 The table shows the number of children in 50 families.

Number of children	Frequency	Cumulative frequency
1	3	3
2	*m*	22
3	12	34
4	*p*	*q*
5	5	48
6	2	50
	Total, *T*	

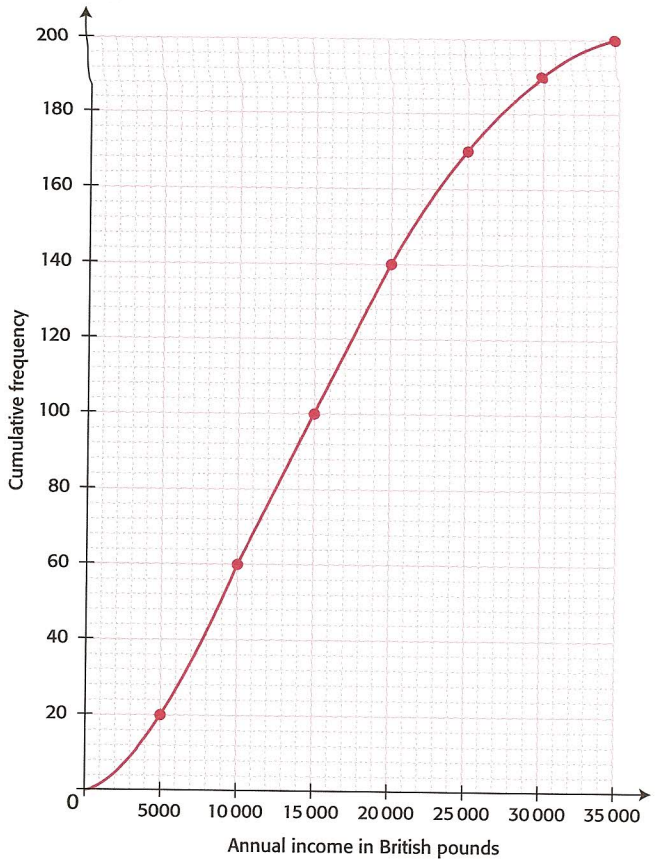
a Write down the value of *T*.
b Find the values of *m*, *p* and *q*. *N99q2*

5 The mean of the ten numbers listed here is 5.5.
4, 3, *a*, 8, 7, 3, 9, 5, 8, 3
a Find the value of *a*.
b Find the median of these numbers. *M99q3*

6 Peter marked 80 exam scripts. He calculated the mean mark for the scripts to be 62.1.
Maria marked 60 scripts with a mean of 56.8.
Peter discovers an error in his marking. He gives two extra marks each to eleven of the scripts.

a Calculate the new value of the mean for Peter's scripts.
After the corrections have been made Peter and Maria put all their scripts together.
b Calculate the value of the mean for all the scripts. *M05q13*

7 The graph shows the cumulative frequency for the yearly incomes of 200 people.



Use the graph to estimate

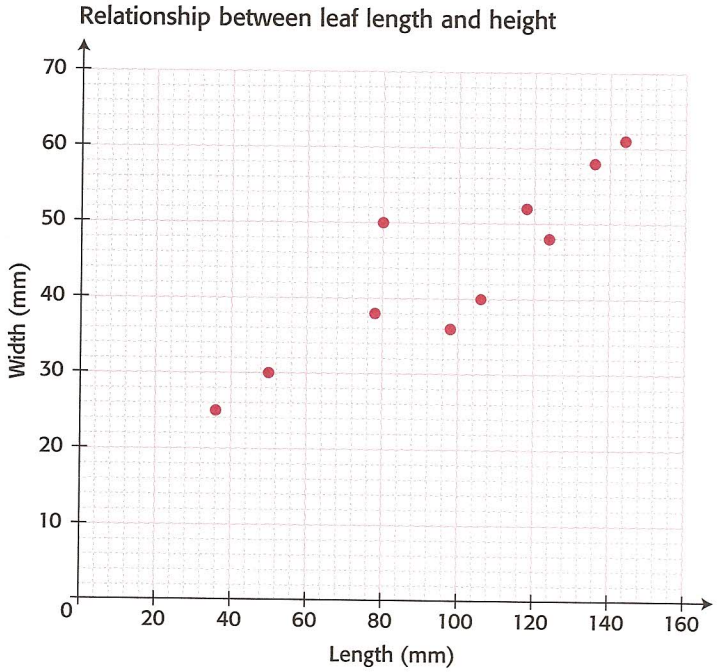
a the number of people who earn less than 5000 British pounds per year.

b The median salary of the group of 200 people.

c The lowest income of the richest 20% of this group.

M01q8

8 The lengths and widths of ten leaves are shown on this scatter diagram.



a Plot the point $M(97, 43)$ which represents the mean length and mean width.
b Draw a suitable line of best fit.
c Write a sentence describing the relationship between leaf length and leaf width for this sample. *M09q13*

9 Tom performs a chi-squared test to see if there is any association between the time to prepare for a penalty kick (short time, medium time, long time) and the outcome (scores a goal, does not score a goal). Tom performs this test at the 10% level.

a Write down the null hypothesis.
b Write down the number of degrees of freedom.
c The p-value for this test is 0.073. What conclusion can Tom make? Justify your answer. *Spec05q30*

Paper 2

1 The table shows the times, to the nearest minute, taken by 100 students to complete a mathematics task.

Time (t) minutes	11–15	16–20	21–25	26–30	31–35	36–40
Number of students	7	13	25	28	20	7

a Construct a cumulative frequency table with upper class boundaries 15.5, 20.5 etc.
b On graph paper, draw a cumulative frequency graph, using a scale of 2 cm to represent 5 minutes on the horizontal axis and 1 cm to represent 10 students on the vertical axis.
c Use your graph to estimate
  i the number of students that completed the task in less than 17.5 minutes
  ii the time it will take for $\frac{3}{4}$ of the students to complete the task. *N00q3*

2 The heights of 200 students are recorded in this table.

Height (h) in cm	**Frequency**
$140 \leq h < 150$	2
$150 \leq h < 160$	28
$160 \leq h < 170$	63
$170 \leq h < 180$	74
$180 \leq h < 190$	20
$190 \leq h < 200$	11
$200 \leq h < 210$	2
Total	200

a Write down the modal group.
b Calculate an estimate of the mean and standard deviation of the heights.