Chapter Two.

Summarising data and describing distributions.

Situation: COMPANY SALARIES.

A particular company employs thirteen people:

1 managing director,

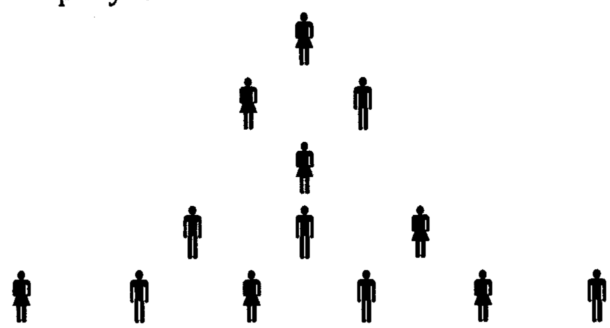
2 managers,

1 accountant,

3 chargehands,

6 machine operators.

The salary structure in the company is shown below.



Managing director: $115000

Managers: $90000 each

Accountant $78000

Chargehands $62500 each

Machine operators $55000 each

- Find the company's mean salary, median salary and modal salary.
- If a group were arguing that higher salaries should be awarded to the people working for this company which of the previous three answers would they claim to be the average to best suit their argument?
- If the managing director wished to quote an average salary for this company, in support of her claim that the average salary was already high, which average best suits her argument?
- If this company took over another company, retaining all 7 employees of the other company on their existing salaries, which had a mean of $66000, what would be the mean salary of the new "20 employee" company?

In the situation on the previous page you had to summarise information by determining the mean, median and mode, concepts that you were reminded of in the *Preliminary work* section at the beginning of this book and that some of the questions in Miscellaneous Exercise One required you to find. The mean, median and mode summarise the location of a set of scores. They are *summary statistics* and are *measures of location*. The mean and the median give an indication of *central tendency*.
The next three examples remind you how to determine these quantities from a frequency table (example 1), a dot frequency graph (example 2) and a stem and leaf plot (example 3).

Example 1
The fifty scores shown below

14	17	15	17	12	16	19	16	17	16
16	10	15	17	18	17	14	16	17	16
18	19	20	14	15	18	18	18	15	17
15	16	17	18	16	16	13	15	18	15
17	17	15	16	19	15	17	18	13	14

can be neatly displayed in the form of a frequency table, as shown below:

Score	10	11	12	13	14	15	16	17	18	19	20
Frequency	1	0	1	2	4	9	10	11	8	3	1

Determine the mode, the mean and the median of the fifty scores.

The mode is readily determined from this table by seeing which score has the greatest frequency. The mode of the set of scores is 17.

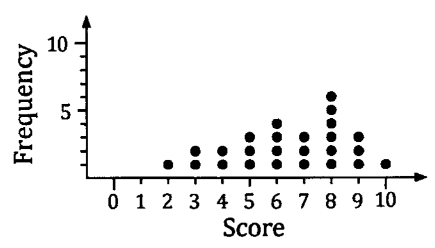
The table tells us that we have one 10, zero 11s, one 12, two 13s etc, Hence, to determine the total of the fifty scores we calculate:

$$1\times10 + 1\times12 + 2\times13 + 4\times14 + 9\times15 + 10\times16 + 11\times17 + 8\times18 + 3\times19 + 1\times20$$
$$= 10 + 12 + 26 + 56 + 135 + 160 + 187 + 144 + 57 + 20$$
$$= 807$$

Hence the mean will be given by $\frac{807}{50}$ i.e. 16·14.

The median of fifty scores will be the mean of the 25th and 26th scores once the scores have been written in order. Summing the frequencies in the above table from the left end we see that the 25th and 26th scores will both be 16. Thus the median score is 16.

Example 2
Determine the mode, mean, median and range for the set of scores shown in the dot frequency graph on the right.



The score which occurs more frequently than any other is 8.
Hence the mode = 8

The dot frequency graph tells us that we have one score of 2, two scores of 3 etc.

$$\text{Hence the mean} = \frac{1\times2+2\times3+2\times4+3\times5+4\times6+3\times7+6\times8+3\times9+1\times10}{25}$$
$$= 6{\cdot}44$$

The median of 25 numbers will have twelve numbers either side of it, i.e. it will be the thirteenth number. Counting along the dot frequency graph we see that the thirteenth number will be a seven. Hence the median is 7.

Remember that the range is the difference between the highest score (10) and the lowest score (2). Hence the range is 8.

Example 3

Determine the mode, median, mean and range for the set of twenty scores shown in the stem and leaf diagram on the right.

12	7
13	9 3 1
14	2 0 3 2 7
15	2 5 5 1 9 5
16	8 3 8
17	3 0

By inspection the mode is 155.

With twenty scores the median will be between the 10th and 11th scores, when the scores are considered in order of size (see right). Hence the median will be the mean of 151 and 152, i.e. 151·5.

By calculation the mean is 150·65 and the range is 173 – 127, i.e. 46.

12	7
13	1 3 9
14	0 2 2 3 7
15	1 2 5 5 5 9
16	3 8 8
17	0 3

Note: If you look for a formula for the mean of a set of numbers in books containing mathematical formulae you may not find the rule stated in the form:

$$\text{Mean} = \frac{\text{The sum of the scores.}}{\text{The number of scores there are.}}$$

Instead the formula may involve symbols like $\bar{x}$ and $\sum$, as explained below.

- As mentioned in the Preliminary Work, we use the symbol $\bar{x}$ to indicate the mean of a set of scores. For the twenty numbers of example 3, $\bar{x} = 150{\cdot}65$.
- The Greek letter $\sum$, pronounced sigma, is used in Mathematics to indicate that numbers are being added together, i.e. a *summation* is being determined. Thus if we consider the numbers 8, 7, 6, 11 to be values of x then
$$\sum x = 8 + 7 + 6 + 11 \ (= 32)$$
- Putting the above ideas together, for a set of n scores it follows that
$$\bar{x} = \frac{\sum x}{n}$$
If we use f to indicate the frequency with which each score occurs it further follows that for data given as a frequency table
$$\bar{x} = \frac{\sum fx}{\sum f}$$

Combining groups.
Knowing the number of scores and their sum we can calculate the mean of the scores. For example if 15 scores have a sum of 108 the scores have a mean of

$$\frac{108}{15} = 7{\cdot}2$$

It also follows that if we know the number of scores and their mean we can find the sum of the scores.
For example, knowing that 15 scores have a mean of 7·2 the scores must have a sum of

$$15 \times 7{\cdot}2 = 108$$

Indeed you probably used this idea if you managed the last dot point of the *Situation* at the beginning of the chapter. In that situation we were told that the 7 employees of a company had a mean salary of \$66 000 from which we could calculate the total amount paid out for these salaries as

$$7 \times \$66\,000 = \$462\,000$$

This idea of determining the total of a set of scores, knowing the mean and the number of scores can be useful when solving some problems as examples 4, 5 and 6 show.

Example 4
The mean of six scores is 23·5. If five of the scores were 17, 20, 19, 25 and 30 find the sixth score.

If six scores have a mean of 23·5 then these six scores have a sum of $23{\cdot}5 \times 6$
$= 141$
The five given scores have a sum of $17 + 20 + 19 + 25 + 30 = 111$
Thus the sixth score must be $141 - 111 = 30$
The sixth score is 30.

Example 5
To pass a particular course a student needs to gain a mean of at least 55% in the five tests that form the course assessment. In the first four tests the student achieves marks of 46%, 57%, 54% and 57%. What percentage mark must the student gain in test five if they are to pass the course?

To gain a 55% average in 5 tests the total marks in the 5 tests must be 55×5
$= 275$
The first 4 tests have a sum of $46 + 57 + 54 + 57 = 214$
Thus in the fifth test the student needs $275 - 214 = 61$
The student needs to score at least 61% in the fifth test in order to pass the course.

Example 6

In a test the 15 girls in a class score a mean mark of 21·2 and the ten boys score a mean mark of 22·4. Calculate the mean for the whole group of 25 students.

The 15 girls achieved a mean of 21·2 thus they gained a total mark of $15 \times 21{\cdot}2$
$= 318$

The 10 boys achieved a mean of 22·4 thus they gained a total mark of $10 \times 22{\cdot}4$
$= 224$

Thus the 25 students gained a total mark of $318 + 224 = 542$

The 25 students achieved a mean mark of $\frac{542}{25} = 21{\cdot}68$

Exercise 2A.

Find the mean, median, mode and range of each of the distributions shown in questions 1 to 6 (correct to one decimal place if necessary).

1.

Score	0	1	2	3	4	5
Frequency	1	2	2	3	4	7

2.

Score	0	1	2	3	4	5	6	7	8	9	10
Frequency	1	0	2	3	2	4	7	6	2	3	1

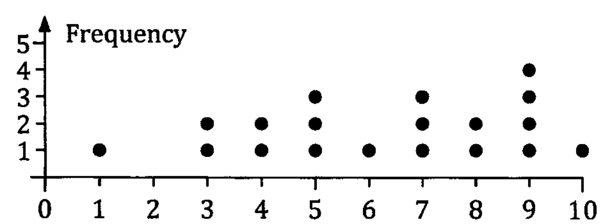
3.

Score	15	16	17	18	19	20
Frequency	1	2	7	8	3	1

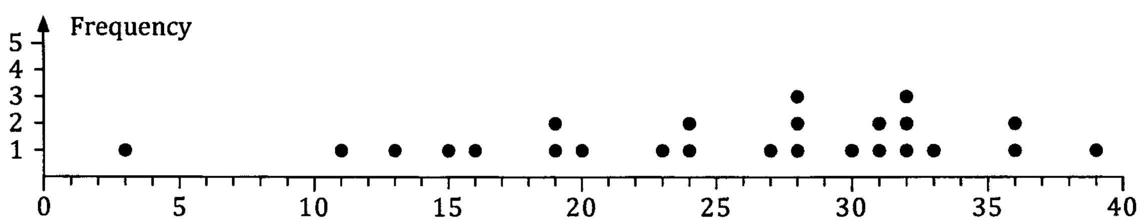
4.

Score	98	100	101	103	104	105
Frequency	1	1	3	3	1	1

5.



6.



7. The stem and leaf diagram on the right can accommodate numbers from 110 to 169, the first two digits being the stem and the last digit the leaf. The diagram shows the scores recorded by 25 competitors in a shooting competition. For the scores recorded determine:

16	4 4
15	3 1 8 7
14	0 1 7 9 5 3
13	0 9 7 4 2 5 6 2
12	0 6 5
11	9 8

(a) the lowest score, (b) the highest score,
(c) the median score, (d) the mean score.

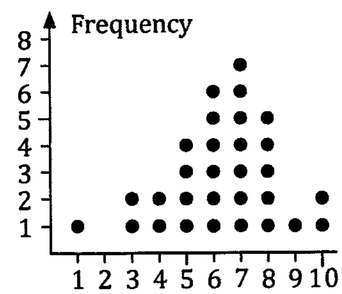
8. The stem and leaf diagram on the right shows scores achieved by 24 students in an exam that was marked out of 120. The stem is the tens digit and the leaf is the units digit. Determine

9	4 0 2
8	2 5 8 1 2 5 4 3
7	5 7 4 4 7
6	1 7 9 8 4 8 6
5	
4	8

(a) the lowest score,
(b) the highest score,
(c) the median score,
(d) the mean score (correct to 1 d.p.).

9. Estimate the mean number of thumbs per Australian adult.

10. The dot frequency diagram on the right shows the marks obtained by 30 year 8 students in a mental arithmetic test.
Determine the mean, median and mode for this distribution of marks.



11. A class of school students calculated their mean age as 17·2 years, correct to one decimal place. This average did not include the teacher. If the teacher were to be included would the mean be increased or decreased?

12. The mean of eight scores is 54·25. If seven of the scores were 60, 50, 37, 60, 55, 32 and 65 find the eighth score.

13. Three mathematics classes sat the same exam. The mean marks for the classes were 55%, 62% and 56% and the number of students in each class were 24, 15 and 21 respectively. Find the mean for the three groups put together.

14. To pass a particular course a student has to achieve a mean of at least 50% in the ten pieces of work that form the assessment items. In the first nine of these pieces of work the student achieves a mean of 46%. What percentage mark must the student achieve in the tenth item if he is to pass the course?

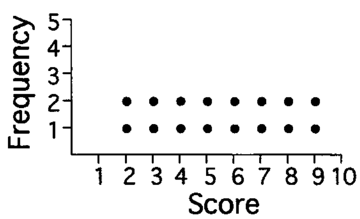
15. The mean of 25 scores is 54. If 20 of the scores had a mean of 55 find the mean of the other five scores.

16. The 13 boys in a class gained a mean mark of 57% in a test in which the class mean was 59%. If the class consisted of 20 students altogether find the mean achieved by the girls in the class. (Give your answer correct to one decimal place.)

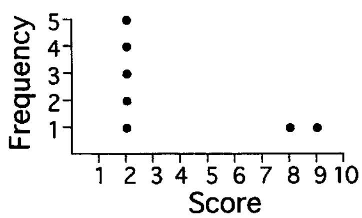
For each of the dot frequency graphs shown in questions 17 to 32, without actually calculating the mean and the median, state which one of the following statements apply:

- ☞ The mean is the same as the median.
- ☞ The mean is greater than the median.
- ☞ The mean is less than the median.

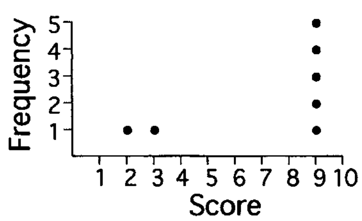
17.



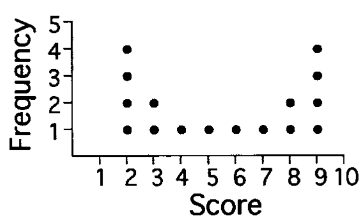
18.



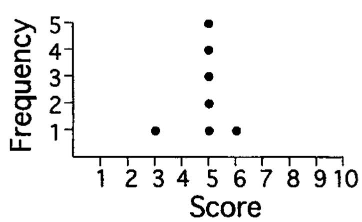
19.



20.

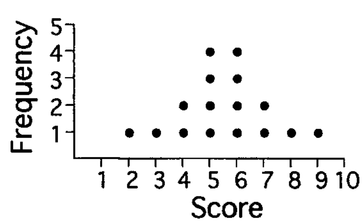


21.

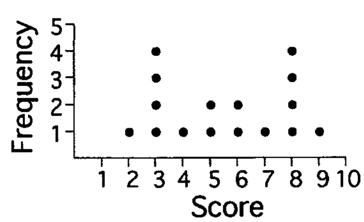


22.

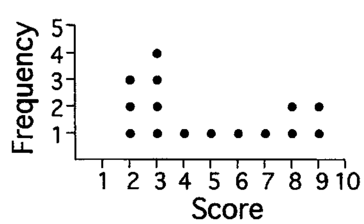
23.



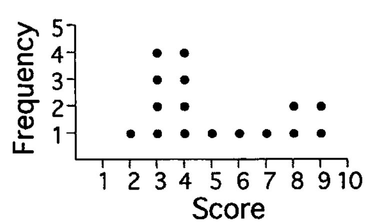
24.



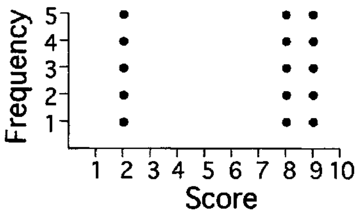
25.



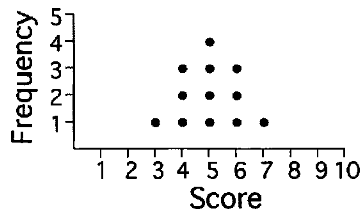
26.



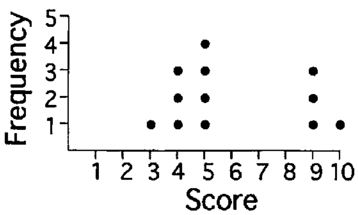
27.



28.

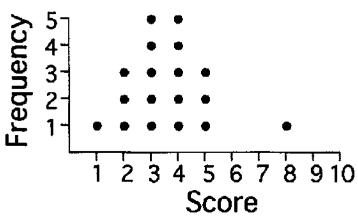


29.

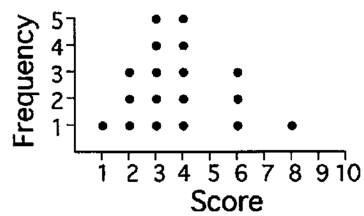


30.

31.



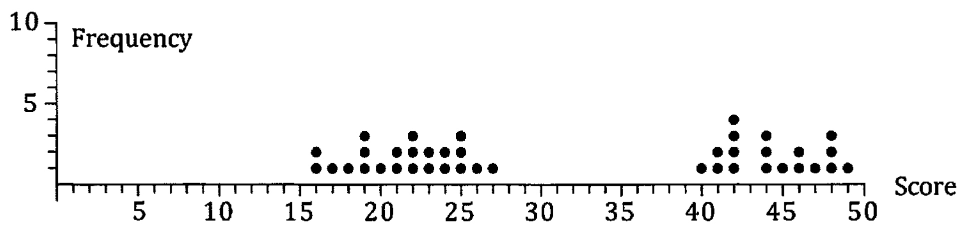
32.



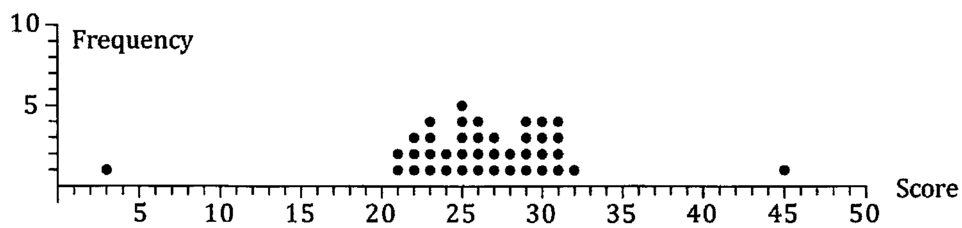
33. Whilst each of the dot frequency diagrams shown in this question feature 40 data points they show quite different distributions of the 40 scores.
Write a few sentences describing each distribution of 40 scores.
Some useful words and phrases that you might consider using in your descriptions could include:

lowest score	highest score	tightly packed	spread out
clusters	gaps	outliers	uniform
groups	dense regions	symmetrical	

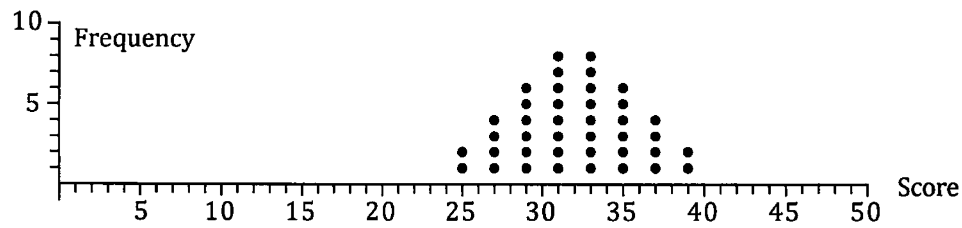
(a)



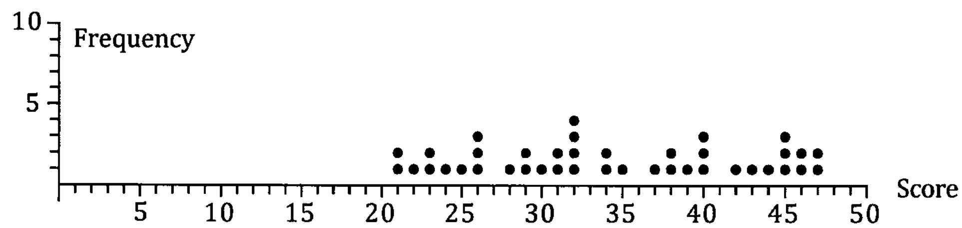
(b)



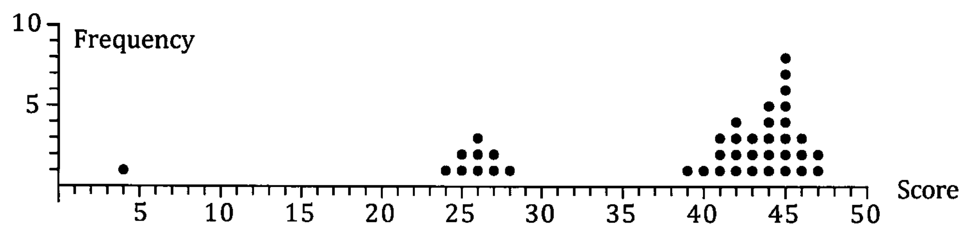
(c)



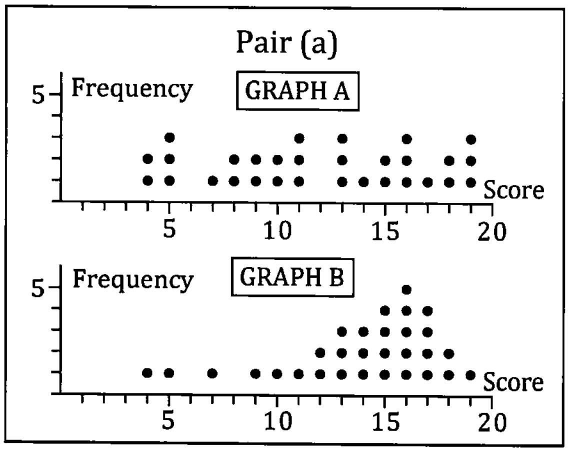
(d)

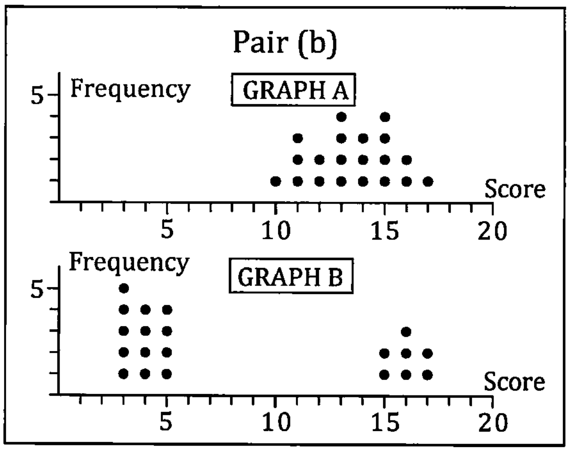


(e)



34. In each part of this question two dot frequency graphs are given, Graph A and Graph B. In each pair the two graphs involve sets of data with the same number of data points. However, in each case the two graphs are quite different. Write a few sentences comparing graphs A and B in each case. Whilst you do not need to determine the mean and the median exactly for any of the graphs your answers should include comparisons of means, medians and spread.





Use of statistical functions on a calculator.

When asked to determine the mean of a set of scores are you adding up all of the scores and then dividing by the number of scores or are you putting the scores into a calculator and using its ability to determine the mean?

Many calculators have an inbuilt ability to output various summary statistics for data that is put into the calculator.

This usually involves the following steps:

1. Setting or selecting the statistics facility of your calculator.
2. Clearing any statistical data already in your calculator.
3. Inputting the data.
4. Outputting the required data.

Your calculator is able to output statistical information for the set of data you put in. The calculator may display more statistical information than you are familiar with at present but somewhere in the list it may well display the number of scores in the set, often listed as "n", the mean of the scores, often listed as $\bar{x}$, the median, the mode and various other things, some of which you will encounter in later mathematical studies.

For example below left shows such a display for the data set:

3, 5, 2, 0, 2

Scrolling down such a display shows more information as shown below right.

$\bar{x}$	= 2.4
Σx	= 12
Σx^2	= 42
$x\sigma_n$	= 1.62480768
$x\sigma_{n-1}$	= 1.81659021
n	= 5
minX	= 0
Q1	= 1

Scroll down

Med	= 2
Q3	= 4
maxX	= 5
Mode	= 2
ModeN	= 1
ModeF	= 2

Can you locate the mean,
the number of scores,
the median,
the mode?

Get to know **your** calculator with regard to inputting data and getting it to display summary statistics for the data.

If the data is presented as a frequency table the information can be put into many calculators in this frequency form, i.e. we do not need to input all of the scores individually. Usually one column is used for the scores and another for the frequencies. We then set the calculator to read the information in each column appropriately.
For example, the following table

Score	10	11	12	13	14	15	16	17	18	19	20
Frequency	1	0	1	2	4	9	10	11	8	3	1

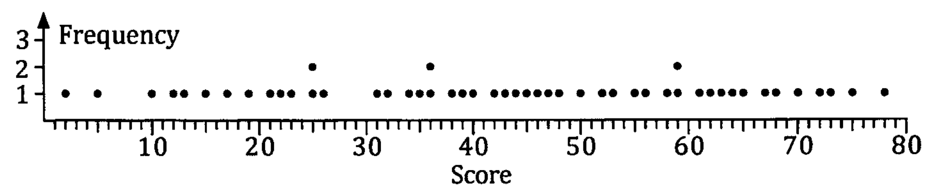
could be put into two columns, see display below left, and the summary statistics displayed, see below right (more statistics being available if we were to scroll down).

	list 1	list 2	list 3	list 4
1	10	1		
2	11	0		
3	12	1		
4	13	2		
5	14	4		
6	15	9		

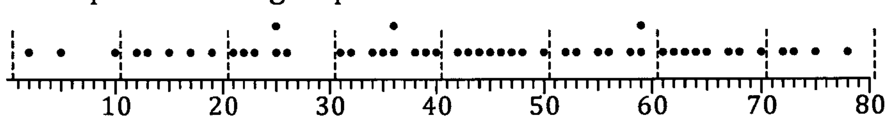
$\bar{x}$ = 16.14
Σx = 807
Σx^2 = 13205
$x\sigma_n$ = 1.897472
$x\sigma_{n-1}$ = 1.91673617
n = 50
minX = 10
Q1 = 15

Once again get to know how to put information from a frequency table into your calculator and how to output the mean etc.

Grouped data.
As was mentioned in chapter one, continuous numerical data is often grouped because of rounding and sometimes discrete numerical data is grouped for convenience. For example suppose we are given a set of fifty numbers with very few repeats. If we were to display the scores as a dot frequency graph table we would have almost as many "columns" to our graph as we have scores as most of the scores occur just once:



The data may be better presented in groups or class intervals:



Giving the grouped frequency table:

Score	1→10	11→20	21→30	31→40	41→50	51→60	61→70	71→80
Frequency	3	5	6	9	8	7	8	4

If we only have the grouped data and do not know the original scores we talk of the **modal group** or **modal class**, rather than referring to a mode. For the frequency table just encountered, and shown again below, the modal class is the 31 → 40 class.

Score	1→10	11→20	21→30	31→40	41→50	51→60	61→70	71→80
Frequency	3	5	6	9	8	7	8	4

Similarly we find the class in which the median lies, called the **median group** or **median class**, in this the case the 41 → 50 class.

To determine a mean we assume all the scores in an interval are at the **midpoint** of the interval. Clearly this is unlikely to be the case but it will give a reasonable estimate for the mean when a large number of scores are involved. Thus for the above table we calculate the mean based on three scores of 5·5, five scores of 15·5, six scores of 25·5 etc.

Check that you agree that applying this idea to the above table of grouped data gives an estimated mean of 42·9.

Exercise 2B (Use the statistical capability of your calculator.)

Find the mean, median, mode and range of each of the sets of data given in questions 1 to 7 (correct to one decimal place if necessary).

1. 125, 137, 137, 143, 153, 162, 165.

2. 85, 85, 78, 72, 83, 78, 90, 89, 78.

3. 8, 34, 19, 14, 25, 15, 40, 26, 17, 30.

4. 55, 42, 36, 63, 45, 35, 76, 50, 50, 58,
40, 72, 35, 80, 75, 66, 48, 35, 62, 35,
66, 40, 56, 52, 38.

5.

Score	0	1	2	3	4	5	6	7
Frequency	8	12	18	20	9	3	0	1

6.

Score	5	6	7	8	9	10	11	12
Frequency	24	35	17	28	33	31	27	19

7.

Score	15	16	17	18	19	20
Frequency	1	3	5	13	16	22

In the next four questions use the midpoint of each class interval to determine the mean of each of the following distributions, correct to one decimal place.

8.

Score	Frequency
1 → 5	15
6 → 10	28
11 → 15	7
16 → 20	3
21 → 25	1
26 → 30	1

9.

Score	Frequency
1 → 4	3
5 → 8	8
9 → 12	15
13 → 16	8
17 → 20	3

10.

Score	Frequency
20 → 24	6
25 → 29	10
30 → 34	17
35 → 39	7
40 → 44	5
45 → 49	4
50 → 54	1

11.

Score (x)	Frequency
$0 \leq x < 20$	5
$20 \leq x < 40$	13
$40 \leq x < 60$	21
$60 \leq x < 80$	72
$80 \leq x < 100$	54

12. As part of the process of assessing the value of a property a real estate agent considers the prices of other properties recently sold in the same area. The selling prices of ten such properties were as follows:

\$437 000	\$425 000	\$456 000	\$421 000	\$442 000
\$445 000	\$441 000	\$437 000	\$432 000	\$540 000

Find the mean and median of these prices.
How many of the ten prices are lower than the mean?
How many of the ten prices are lower than the median?

13. A real estate survey investigated the number of bedrooms in each of 100 houses in a particular area. The results were as shown below:

Number of bedrooms	1	2	3	4	5	6	7
Frequency	2	3	37	49	5	3	1

Find the mean number of bedrooms per house for these houses.

14. A company employs 25 people and has seven salary levels. The number of employees on each level are shown in the following table.

Salary	\$62 000	\$68 000	\$71 000	\$78 000	\$85 000	\$100 000	\$110 000
N°. of employees	4	10	5	1	2	2	1

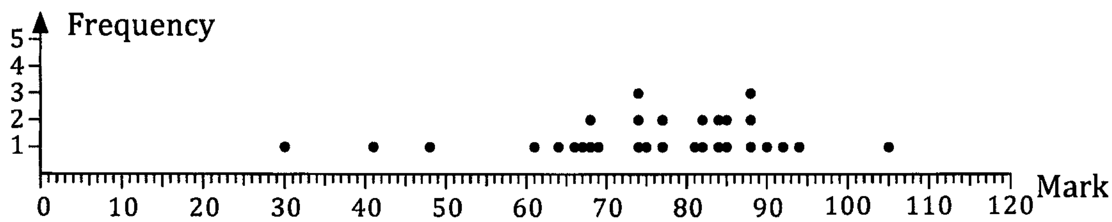
Calculate (a) the modal salary,
(b) the median salary,
(c) the mean salary.

15. In one particular year, the number of hours of sunlight recorded each day in December at a particular weather recording location were as follows.

			1st Dec 11·2 hrs	2nd Dec 9·4 hrs	3rd Dec 8·5 hrs	4th Dec 12·2 hrs
5th Dec 19·2 hrs	6th Dec 11·2 hrs	7th Dec 11·8 hrs	8th Dec 10·4 hrs	9th Dec 8·7 hrs	10th Dec 10·3 hrs	11th Dec 9·1 hrs
12th Dec 11·1 hrs	13th Dec 10·0 hrs	14th Dec 12·0 hrs	15th Dec 11·2 hrs	16th Dec 12·9 hrs	17th Dec 13·1 hrs	18th Dec 12·0 hrs
19th Dec 11·7 hrs	20th Dec 9·3 hrs	21st Dec 9·0 hrs	22nd Dec 11·1 hrs	23rd Dec 12·8 hrs	24th Dec 10·3 hrs	25th Dec 13·2 hrs
26th Dec 12·1 hrs	27th Dec 7·3 hrs	28th Dec 5·2 hrs	29th Dec 9·9 hrs	30th Dec 10·1 hrs	31st Dec 11·3 hrs	

Calculate the mean and the median number of hours of sunlight per day for the December of this particular year at this location.

16. The scores obtained by 30 students in an exam are shown in the dot frequency diagram below. The exam was marked out of 120.



Calculate
(a) the mean of these scores (correct to one decimal place),
(b) the median of these scores.
(c) the number of students scoring less than 60%
(d) the percentage of students scoring greater than 75%.

17. The heights of a group of 29 students, measured to the nearest centimetre, were as shown in the stem and leaf diagram below.

Girls						Boys						
					18	6						
				5	17	9	3	0	5	2		
	8	3	5	0	16	9	6	3	2	8	5	9
7	9	5	8	5	15	9	5	9				
			9	8	14	9						

(a) How many girls were there in the group?
(b) How tall was the shortest girl?
(c) Find the mean height for the girls (to nearest cm).
(d) Find the mean height for the boys (to nearest cm).
(e) Find the mean height for the group of 29 students (to nearest cm).
(f) Display the heights of the 29 students as a dot frequency diagram, making some distinction on your diagram between boys and girls.

18. (a) Calculate the mean of the 50 scores shown below.

23	42	40	47	31	42	39	31	43	26
33	50	23	40	49	30	42	40	29	37
44	48	43	26	36	43	50	32	31	44
47	31	45	37	48	32	42	41	43	23
28	47	36	26	36	45	23	49	29	48

(b) Rearrange the above 50 scores as grouped data using the class intervals:
21 - 25, 26 - 30, 31 - 35, 36 - 40, 41 - 45 and 46 - 50.
Use the midpoints of the intervals to determine the mean for this grouped data.

19. One hundred students were asked to note the number of hours they study in the "study week" they are given prior to an examination period. The frequency table below shows the results of this survey.

Number of hours (h)	Number of students
$0 \le h < 10$	3
$10 \le h < 20$	4
$20 \le h < 30$	10
$30 \le h < 40$	20
$40 \le h < 50$	29
$50 \le h < 60$	18
$60 \le h < 70$	9
$70 \le h < 80$	4
$80 \le h < 90$	2
$90 \le h < 100$	1

(a) Find the modal class for the number of hours spent studying in the week.
(b) Use the class midpoints to determine the mean for the distribution.

20. As part of the process of assessing the value of a block of land a real estate agent considers other blocks recently sold in the area. The agent is able to access such information from data held in his computer. For 68 recent sales the information was as follows:

Price (\$$C$)	Midpoint of interval	Frequency
$200\,000 \le C < 210\,000$	205 000	2
$210\,000 \le C < 220\,000$	215 000	13
$220\,000 \le C < 230\,000$	225 000	10
$230\,000 \le C < 240\,000$	235 000	15
$240\,000 \le C < 250\,000$	245 000	7
$250\,000 \le C < 260\,000$	255 000	7
$260\,000 \le C < 270\,000$	265 000	4
$270\,000 \le C < 280\,000$	275 000	6
$280\,000 \le C < 290\,000$	285 000	3
$290\,000 \le C < 300\,000$	295 000	1

(a) In which class interval does the median price of the 68 blocks lie?
(b) Use the interval midpoints to calculate the mean price (to nearest $1 000).

Describing a distribution of scores.

The last two questions of the first exercise in this chapter, Exercise 2A, required you to write some sentences describing some distributions. Indeed one of those questions suggested the following useful words that you could consider using:

lowest score	highest score	tightly packed	spread out
clusters	gaps	outliers	uniform
groups	dense regions	symmetrical	

Let us now formalize what aspects we should consider when asked to describe a data distribution.

If asked to write a description of a set of scores we should comment on such things as

- the location of the scores
- how spread out they are
- the "shape" of the distribution

and • anything else of relevance.

☛ Location: The mean and the median of the scores give information about location.

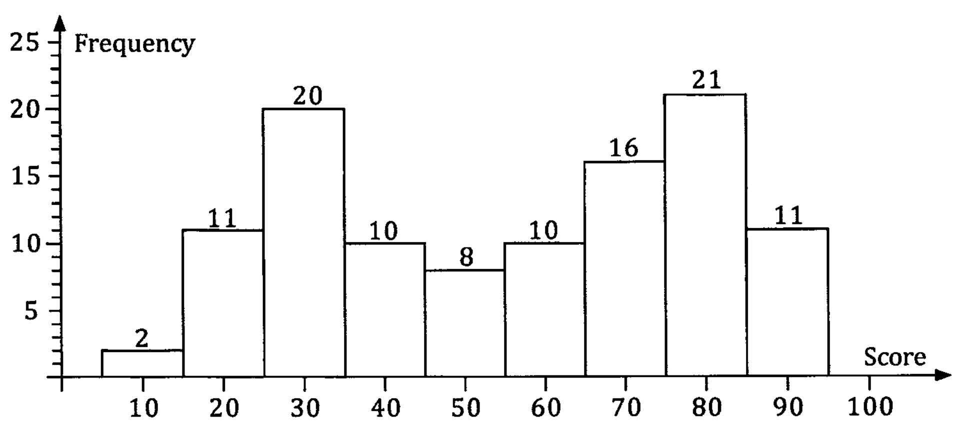
☛ Spread: The range of the scores gives information about their spread.

☛ Shape: Symmetry, gaps, clustering, more dense/less dense regions, outliers, modality (does the data have one modal class or is it perhaps bimodal with two modal classes) all convey information about the shape of a distribution.

Note: From work of earlier years some readers may also be familiar with *standard deviation* as a measure of spread, and with the idea that a distribution can be *skewed*. These terms are not included in the descriptions of distributions given in this section because later chapters cover these concepts. (Standard deviation is covered in chapter 3 and skewness in chapter 4.)

Example 7

The histogram below shows the distribution of scores achieved by the students of a school in a mathematics exam.



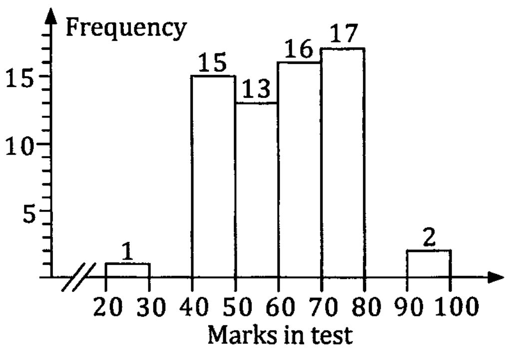
Describe the distribution of marks in this exam.

One hundred and nine students from the school sat the exam.	← Any relevant information perhaps not covered in location, spread and shape?
Using the mid-point of each class interval gives an estimated mean of 55·3 and the middle ranked student, i.e. the 55th student, achieved a mark between 55 and 65.	← Comment about location.
The scores were well spread out from about 5 to 95, i.e. the range was about 90.	← Comment about spread.
The scores were reasonably symmetrically spread about a mid point of about 55. The distribution of scores was bimodal peaking around 30 and again around 80.	← Comment on shape.
Approximately 30% of the students scored less than 35 and approximately 30% scored more than 75.	← Anything else you notice of relevance.

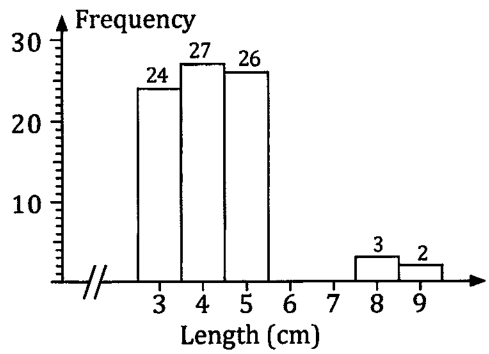
Exercise 2C

Describe each of the distributions shown in questions 1 to 6.

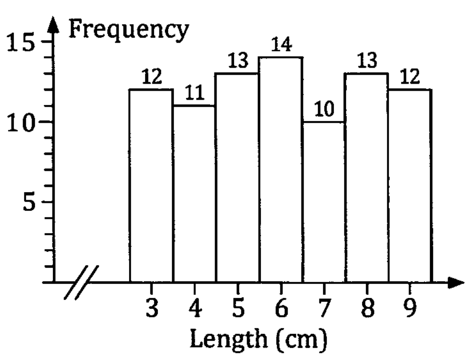
1.



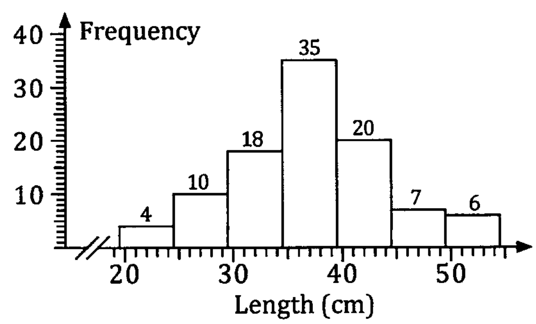
2.



3.



4.



5.

Score	21-25	26-30	31-35	36-40	41-45	46-50	51-55
Frequency	35	21	17	12	7	3	1

6.

Score (x)	$0 \le x < 10$	$10 \le x < 20$	$20 \le x < 30$	$30 \le x < 40$	$40 \le x < 50$	$50 \le x < 60$	$60 \le x < 70$
Frequency	28	15	7	5	9	17	25

Organising, describing and interpreting data.

Lizards.

A survey of a particular species of lizard involved capturing about one hundred different lizards of the species, measuring their length, marking them with an identifiable tag and then releasing them back into the wild. (The tagging was to ensure that the same lizard did not feature more than once in the measuring process and also to allow for follow up studies to be carried out later.)

The lizard lengths, recorded in millimetres, are shown below (the shortest and longest lengths in the list are shown in bold):

114	140	161	110	88	153	112	163	107	151
54	113	39	115	116	106	69	165	103	104
100	110	173	93	158	109	160	60	175	112
160	155	117	162	**33**	168	152	106	156	159
90	116	104	154	161	101	164	111	107	103
167	110	158	109	115	154	42	163	158	106
105	104	47	113	163	77	158	167	**197**	119
162	156	147	160	105	163	109	108	162	111
156	63	104	104	170	131	108	115	111	153
107	122	164	116	102	162	57	160	103	115
158	105								

Organise the data into a grouped frequency table involving what you consider to be an appropriate number of equal width intervals.

Display the data graphically and write some sentences describing the distribution of lengths using suitable statistical vocabulary, e.g. range, outliers, clusters, modal class, mean length, etc.

If the distribution of lengths shows any particularly notable features suggest some possible reasons why this might be the case. (The scientific team can then consider exploring such suggestions in further surveys.)

Miscellaneous Exercise Two.

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the preliminary work section at the beginning of the book.

1. Expand each of the following and simplify where possible.

(a)	$3(2x + 5)$	(b)	$5(7x - 3)$
(c)	$-2(1 - 5x)$	(d)	$6(2x + 1) + 5(2x - 3)$
(e)	$2(2x + 1) - 3(2x - 3)$	(f)	$3(1 - 2x) + 2(5x + 3)$
(g)	$2(2x + 3) - (5x + 1) + 2x$	(h)	$5(1 + 2x) - 2(3 - 2x)$

2. In a National Heart Foundation survey 1915 women and 1863 men, all aged between 40 and 60, ticked either Yes or No when asked the question:

   In the past 2 weeks, did you walk for recreation or exercise? No ☐ Yes ☐

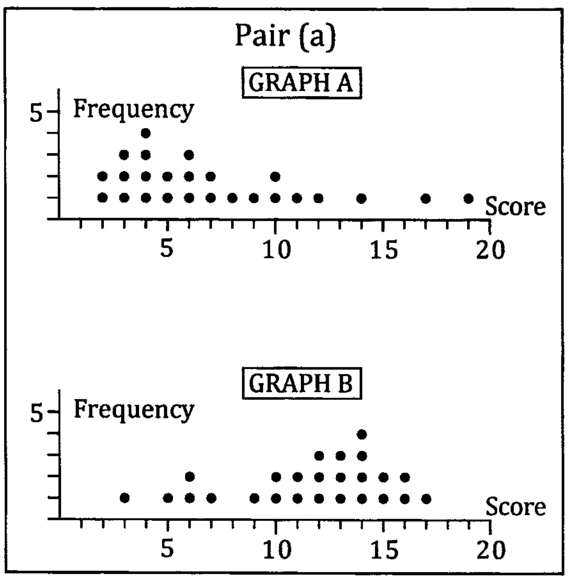
   What type of variable is involved here?

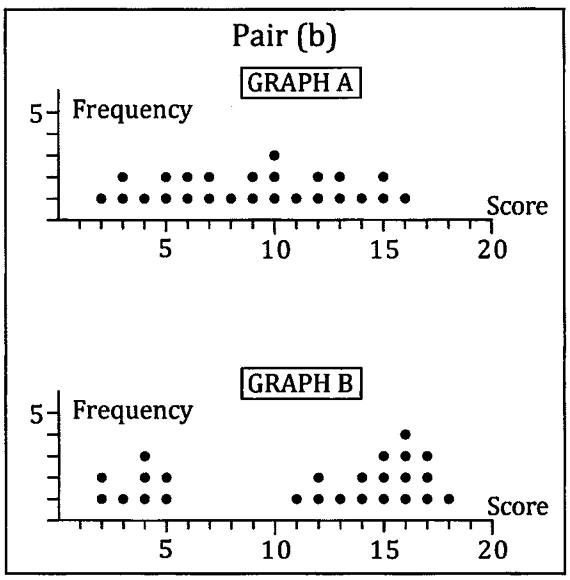
3. List advantages and disadvantages of using the mean as the representative score in a set of scores.

   Do the same for using the median in this way and then for using the mode in this way.

4. To gain a pass a student needs to achieve a mean of at least 60% in eight tests. In the first seven tests the student achieved a mean of 54%. What percentage must the student achieve in test eight if they are to pass the course?

5. In each part of this question two dot frequency graphs are given, Graph A and Graph B. In each pair the two graphs involve sets of data with the same number of data points. However, in each case the two graphs are quite different. Write some sentences comparing datasets A and B in each case and mentioning such things as means, medians, lowest and highest scores, range, gaps, spread out, clusters, percentages of scores, outliers etc.





6. List six prices for second hand cars with a mean of $28 000 but for which this mean value is not a particularly good choice to represent a price that is centrally located with regard to the six prices.

7. HEALTH SURVEY.
One hundred men and one hundred women took part in a health survey. Fifty of the men and fifty of the women were aged between 20 and 25 and the remaining fifty men and fifty women were aged between 65 and 70.

One of the variables measured was the diastolic blood pressure of these two hundred people. The results of such measurements are shown below with the blood pressures given to the nearest millimetre of mercury (mmHg).

MALES (aged 20 to 25)	
B. P.	Frequency
50 - 59	1
60 - 69	10
70 - 79	21
80 - 89	14
90 - 99	3
100 - 109	1
110 - 119	0
Total	50

MALES (aged 65 to 70)	
B.P.	Frequency
50 - 59	0
60 - 69	3
70 - 79	12
80 - 89	17
90 - 99	13
100 - 109	4
110 - 119	1
Total	50

FEMALES (aged 20 to 25)	
B. P.	Frequency
50 - 59	4
60 - 69	22
70 - 79	19
80 - 89	5
90 - 99	0
100 - 109	0
110 - 119	0
Total	50

FEMALES (aged 65 to 70)	
B. P.	Frequency
50 - 59	1
60 - 69	3
70 - 79	13
80 - 89	20
90 - 99	10
100 - 109	2
110 - 119	1
Total	50

(a) Use the centre of each class interval to determine a mean blood pressure for each of these four groups.
(b) Draw frequency histograms for the two male groups.
(c) Draw frequency histograms for the two female groups.
(d) Comment on any trends suggested from parts (a) (b) and (c).