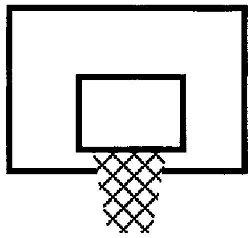
Chapter Three.
Measures of dispersion or spread.

Situation.

The heights of the "starting 5" players of two basketball teams are given below.

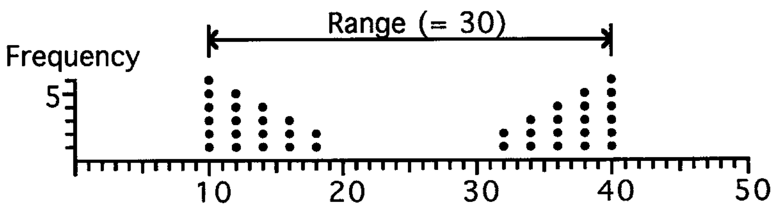


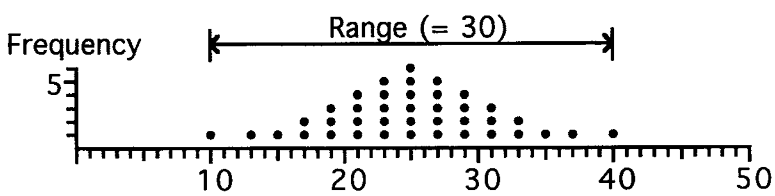
Team A		Team B	
Player 1	211 cm	Player 1	186 cm
Player 2	184 cm	Player 2	184 cm
Player 3	184 cm	Player 3	184 cm
Player 4	172 cm	Player 4	184 cm
Player 5	169 cm	Player 5	182 cm
Mean height.	184 cm	Mean height.	184 cm
Median height.	184 cm	Median height.	184 cm
Modal height.	184 cm	Modal height.	184 cm

Note that the two teams have the same mean as each other, the same median as each other and the same mode as each other. Can we conclude that, with regard to heights, the two teams are similar?

The situation above shows that whilst averages can be very useful in summarising data they do not tell the whole story. We also need to consider how widely the data is **spread** or **dispersed**. Thus as well as being able to summarise data using means and medians as *measures of central tendency* we also need some *measures of dispersion*. Of course we do already have one such measure of dispersion - the range of the scores. However, as was stated in the Preliminary work section at the beginning of this text "*Whilst the range is easy to calculate, it is determined using just two of the scores and does not take any of the other scores into account*". Hence the range is not that useful for comparing the spread of distributions.

For example notice that the two distributions shown below, each involving 40 scores, have the same range but show very different spread patterns.





This reliance on just the lowest and highest scores makes the range of limited use and so we need to consider other measures of spread.

We will now consider two other ways of quantifying spread, namely

the **mean deviation**,

and the **standard deviation**.

Each of these considers how much each score deviates from the mean score. In this way we can obtain measures that will tell us how concentrated the scores are about the mean value and that use each and every score in the data set in their determination.

Consider again the heights of team A from the previous page and listed again below:

Team A (mean 184 cm)

Player	Height	Deviation from the mean	
1	211 cm	211 cm − 184 cm =	+27 cm
2	184 cm	184 cm − 184 cm =	0 cm
3	184 cm	184 cm − 184 cm =	0 cm
4	172 cm	172 cm − 184 cm =	−12 cm
5	169 cm	169 cm − 184 cm =	−15 cm

If we sum the deviations from the mean the answer is zero (as you may have expected from your understanding of the mean). Thus we cannot find the average of these deviations as they are. To avoid this problem we could:

- Ignore the negative signs and find the average of the absolute values of the deviations. This technique gives the **mean deviation** of the heights.

  For team A, mean deviation of heights $= \dfrac{27 + 0 + 0 + 12 + 15}{5}$

  $= 10{\cdot}8$

Alternatively we could:

- Square the deviations, find the average of these square deviations (this is called the **variance** of the scores) and then square root this variance. (This final step of finding the square root is to give a measure that has the same units as the original deviations that had been squared.)
  This technique gives the **standard deviation** of the scores.

  For team A, variance of heights $= \dfrac{(27)^2 + (0)^2 + (0)^2 + (-12)^2 + (-15)^2}{5}$

  $= 219{\cdot}6$

  Thus the standard deviation $= \sqrt{219{\cdot}6}$

  $= 14{\cdot}8$, correct to 1 decimal place.

Note: Finding the standard deviation may seem a more complicated process than that of finding the mean deviation. However it is the standard deviation that is the more commonly used measure of dispersion in data analysis.
For this reason we will focus our attention on the standard deviation and will not pursue the concept of a mean deviation.

Standard deviation.

Example 1

Find the mean, the variance and the standard deviation of the set of scores:

4, 7, 10, 13, 21.

$$
\begin{aligned}
\text{Mean} &= \frac{4+7+10+13+21}{5} \\
&= 11 \\
\text{Variance} &= \frac{(4-11)^2+(7-11)^2+(10-11)^2+(13-11)^2+(21-11)^2}{5} \\
&= 34 \\
\text{Standard deviation} &= \sqrt{34} \\
&= 5{\cdot}83 \text{, correct to two decimal places.}
\end{aligned}
$$

The scores have a mean of 11, a variance of 34 and a standard deviation of 5·83 (correct to 2 d.p.).

Exercise 3A

Just as the various measures of central tendency can be obtained using the statistical capabilities of many calculators so too can some of the measures of dispersion. You will be encouraged to use your calculator to obtain such measures soon but for this exercise obtain the variance and standard deviation "the long way", as in the above example, to gain understanding of the concepts.

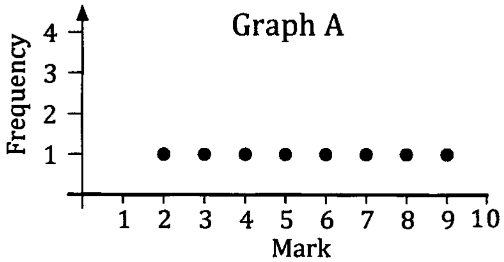
1. Find the range of each of the following sets of scores.
   (a) 5, 7, 11, 12, 17, 19, 21, 36.
   (b) 104, 115, 117, 117, 118, 121, 122, 125, 125, 146.
   (c) 121 000, 109 000, 128 000, 90 000, 110 000, 95 000, 112 000, 107 000.

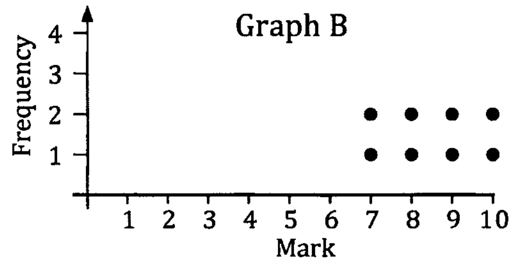
2. Find the variance of each of the following sets of scores.
   (a) 5, 5, 6, 8, 10, 12, 12, 13, 14, 15.
   (b) 7, 9, 10, 12, 15, 18, 18, 19.

3. Find the standard deviation of each of the following sets of scores.
   (Give answers correct to 2 decimal places when rounding is necessary.)
   (a) 15, 15, 15, 15, 15, 15, 15.
   (b) 5, 7, 10, 11, 13, 13, 16, 21.
   (c) 104, 115, 117, 117, 118, 121, 122, 125, 125, 146.
   (d) 65, 67, 72, 83, 84.

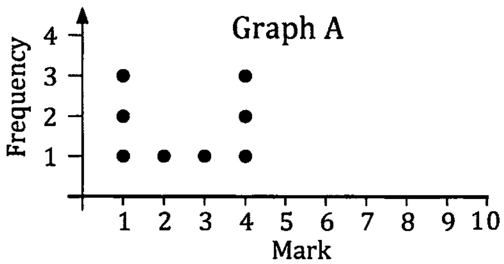
For questions 4 to 7 state which of the two diagrams shows the set of scores with
(a) the greater mean, (b) the greater standard deviation.
(You should **not** need to calculate the means and standard deviations.)

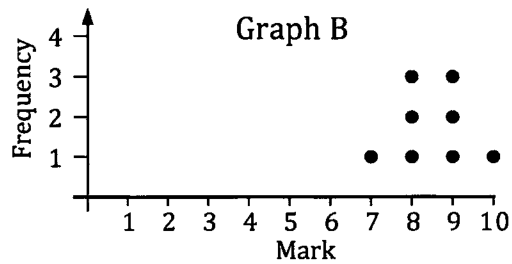
4.



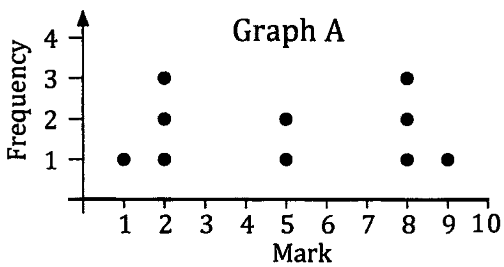


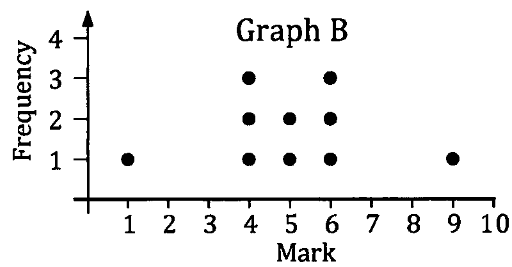
5.



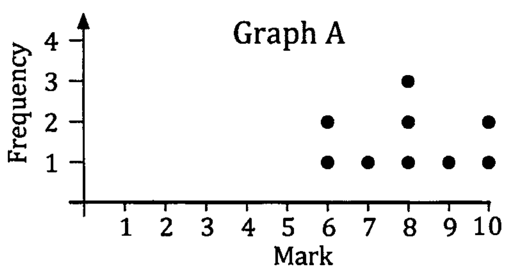


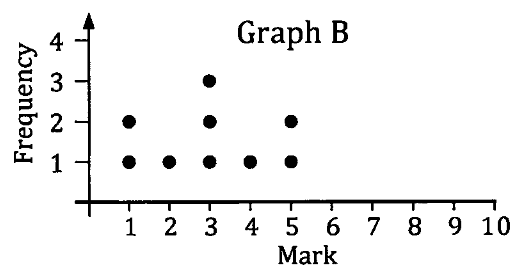
6.





7.





8. For each of parts (a) to (d) given below, two sets of scores are given. For each part state which set, I or II, has the greater standard deviation (you should **not** need to calculate the standard deviations).

(a)	Set I:	6	6	6	6	6	6	6			mean = 6
	Set II:	4	5	6	6	6	7	8			mean = 6
(b)	Set I:	6	8	9	9	10	11	11	12	14	mean = 10
	Set II:	6	6	6	9	10	11	14	14	14	mean = 10
(c)	Set I:	20	21	21	22	23	24	25	25	26	mean = 23
	Set II:	13	15	21	22	23	24	25	31	33	mean = 23
(d)	Set I:	1	1	2	2	10	18	18	19	19	mean = 10
	Set II:	1	1	9	9	10	11	11	19	19	mean = 10

9. Find the range, the mean and the standard deviation of the following set of scores. (Give answers correct to two decimal places if rounding is necessary.)

   8, 8, 9, 11, 11, 15, 16, 19, 19, 20, 20, 21, 21, 25, 32.

Use of statistical functions on a calculator.

As was stated at the beginning of the previous exercise, just as the various measures of central tendency can be obtained using the statistical capabilities of many calculators so too can some of the measures of dispersion. Also, you should already know how to put data into a calculator and output statistical measures such as the mean and the median.

The standard deviation of a set of scores can similarly be obtained using your calculator.

We tend to use either s or σ as symbols for standard deviation. σ is a letter from the Greek alphabet and is pronounced sigma. It is a "lower case" sigma, capital sigma is written $\sum$.

The diagram below shows a typical graphic calculator display for the data set:

3 5 2 0 2

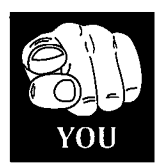
1-Variable		
$\bar{x}$	= 2.4	⟵ The mean of the scores
$\sum x$	= 12	⟵ The sum of the scores
$\sum x^2$	= 42	⟵ The sum of the squares of the scores
$x\sigma n$	= 1.62480768	⟵ The standard deviation of the scores
$x\sigma n\text{-}1$	= 1.81659021	⟵ The different standard deviation - see note ② below
n	= 5	⟵ The number of scores

Scrolling down such a display would allow further statistical information for this set of scores to be viewed.

For this set of scores: Mean = 2·4

Standard deviation = 1·625 (correct to 3 decimal places)

Make sure that **you** can obtain these two values from **your** calculator.



Note ① The standard deviation is a measure of spread. For most distributions very few, if any, of the scores would be more than three standard deviations from the mean, i.e. the vast majority of the scores (and probably all of them) would lie between

$$(\bar{x} - 3\sigma) \text{ and } (\bar{x} + 3\sigma).$$

Indeed we would frequently find that about two thirds of the scores would lie within 1 standard deviation of the mean, i.e. between $(\bar{x} - \sigma)$ and $(\bar{x} + \sigma)$.

Note ② The calculator display shown has two different standard deviations,

$$\sigma_n \text{ and } \sigma_{n-1}$$

σ_n is the standard deviation of the five scores.

σ_{n-1}, sometimes shown as s_x, gives an answer a little bigger than σ_n by dividing the sum of the squared deviations by (n − 1) rather than n. This would be used if the five scores were a sample taken from a larger population and we wanted to use the deviation of the sample to estimate the standard deviation of the whole population, as is often the case in real life. Division by (n − 1) rather than n compensates for the fact that there is usually less variation in a *small* sample than there is in the entire population. (This is known as Bessel's correction.) If we make the sample large then n will be large and there will be little difference between σ_n and σ_{n-1}. This book makes the distinction between σ_n and σ_{n-1} and if a question asks for the standard deviation of a set of scores to be determined σ_n is given, unless the question *specifically* states that the task is to use the small sample data to estimate the standard deviation of the population of which the sample is a part. However in some States of Australia this distinction may not be part of the course and instead you might simply be expected to give σ_{n-1}, or s_x, whenever standard deviation is requested. Hence it is important that you make sure you know which standard deviation, σ_n or σ_{n-1}, you are expected to give in examinations when a question simply asks you to determine a standard deviation. To assist readers in states that only require σ_{n-1} to be determined answers in the back of this text will tend to give both values for those cases where, to the given accuracy, the values differ.

Example 2

Find the mean and standard deviation for the following set of scores.

9, 10, 11, 13, 19, 20, 21, 21, 22, 24, 25, 31, 31, 32, 44.

(a) How many of the scores are such that

$$\bar{x} - 1 \text{ standard deviation} < \text{score} < \bar{x} + 1 \text{ standard deviation} \quad ?$$

(b) How many of the scores are such that

$$\bar{x} - 2 \text{ standard deviations} < \text{score} < \bar{x} + 2 \text{ standard deviations} \quad ?$$

Using a calculator $\bar{x} = 22{\cdot}2$, standard deviation $(\sigma_n) = 9{\cdot}27$ (2 d.p.)

(a) $\bar{x}$ − standard deviation $\approx 12{\cdot}93$ and $\bar{x}$ + standard deviation $\approx 31{\cdot}47$.

10 of the 15 scores lie between $\bar{x} - \sigma$ and $\bar{x} + \sigma$.

(b) $\bar{x}$ − 2 standard deviations $\approx 3{\cdot}66$ and $\bar{x}$ + 2 standard deviations $\approx 40{\cdot}74$.

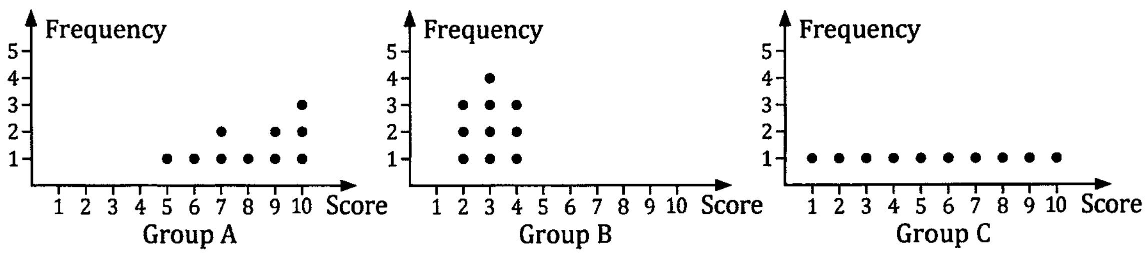
14 of the 15 scores lie between $\bar{x} - 2\sigma$ and $\bar{x} + 2\sigma$.

Exercise 3B

Find the mean and the standard deviation for each of the sets of scores given in numbers 1 to 5. (Give answers correct to one decimal place when rounding is necessary unless stated otherwise.)

1. 10, 11, 12, 13, 14, 15.
2. 15, 26, 47, 16, 33, 49, 8, 11, 41, 26, 19, 14.
3. 31, 29, 33, 32, 34, 29, 30, 30.
4. 6·6, 6·2, 7·3, 8·1, 6·8, 7·0, 6·9, 7·1, 6·9, 7·0. (Answers correct to 2 d.p.)
5. 30, 29, 34, 27, 26, 25, 26, 38, 38, 23, 39, 35, 26, 27, 29, 32, 29, 31, 32, 30, 31, 27, 29, 32, 30, 32, 31, 28, 32, 30, 29, 30.

6. Three groups of 10 students do a spelling test marked out of 10. The scores achieved by each group are shown in the dot frequency graphs below.



   (a) Without calculating values but just by looking at the graphs state which of the three groups have scores with
      (i) the greatest standard deviation, (ii) the smallest standard deviation,
      (iii) the greatest mean, (iv) the smallest mean.
   (b) Calculate the mean and standard deviation for each group.

7. In a particular sporting competition contestants are awarded a score by each of eight judges. The eight scores for one competitor were:
   5·9 5·9 5·7 6·0 5·1 5·8 5·8 5·9
   (a) Find the mean and standard deviation of these scores. (To 2 d.p.)
   (b) If the highest score and the lowest score are discarded find the mean and standard deviation of the remaining scores. (To 2 d.p.)

8. An engineering company makes a particular component that theoretically is to be of length 63 cm. Quality control imposes certain restrictions that any randomly selected sample of ten of these components must satisfy. One such random sample has lengths (in cm):
   63·0, 62·9, 63·0, 63·2, 63·0, 63·1, 63·0, 62·9, 63·1, 63·1.
   (a) Does this sample satisfy the restriction: 62·95 cm < mean < 63·05 cm?
   (b) Does this sample satisfy the restriction: standard deviation < 0·1 cm?

9. The 25 students in a year 8 class have the following heights (nearest cm).

   181 145 162 158 165 150 164 155 173 160 164 154 161
   152 167 169 148 163 175 153 166 153 166 147 160

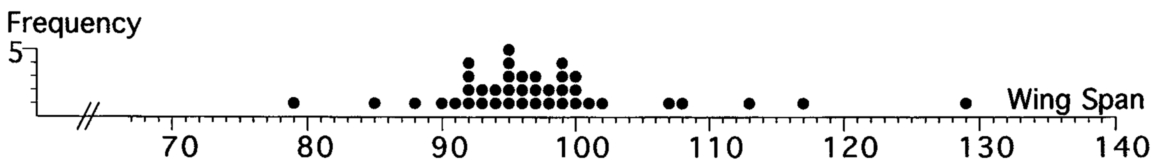
   (a) Find the mean and standard deviation of these 25 heights (2 dp).
   (b) These 25 heights are to be used to estimate the standard deviation of the 231 year 8 students in this school. What would this estimated standard deviation be (2 dp)?

10. An entomologist catches an adult moth that he is sure is one belonging to a particular species but he is not sure whether it is type A of the species or type B. Some of the smaller examples of type B moths can easily be mistaken for a type A moth and positive examination then requires analysis of body tissue. The entomologist refers to a reference book which states that in an extensive survey involving thousands of these adult moths it was found that the body length of the two types were such that:

    For the type A moths surveyed. Mean 15 mm, Standard deviation 1mm.
    For the type B moths surveyed. Mean 22 mm, Standard deviation 3mm.

    The entomologist measures the body length of "his" moth as 18 mm. Decide which it is more likely to be, a large type A or a small type B, and explain your choice.

11. A scientist collects 40 butterflies of a particular species and measures the lengths of the wing span of each one. The lengths, to the nearest millimetre, are shown in the dot frequency diagram below.



    (a) Calculate the mean and standard deviation (st. dev^n) for this set of lengths, giving the standard deviation correct to 3 decimal places.
    (b) What percentage of the 40 lengths lie within one standard deviation of the mean? i.e. What percentage of the lengths are such that:
    (mean – 1 st. dev^n) < length < (mean + 1 st. dev^n).
    (c) What percentage of the 40 lengths lie within two st. dev^{ns} of the mean?
    (d) What percentage of the 40 lengths lie within three st. dev^{ns} of the mean?
    (e) If the scientist wanted to use these 40 lengths to estimate the standard deviation of the entire butterfly population of this species what would this estimated standard deviation be (2 dp)?

12. A scientific experiment involved students determining the temperature at which a particular chemical reaction took place. Ten groups carried out the experiment and obtained the following answers:

    156°C, 163°C, 154°C, 158°C, 159°C, 161°C, 121°C, 163°C, 159°C, 159°C.

    One of the groups discovered that they had made a number of errors in carrying out the experiment and in calculating the answer.

    Determine the mean and standard deviation of the results if

    (a) the answer likely to be from the group making errors **is** included,
    (b) the answer likely to be from the group making errors is **not** included.

13. Twenty four of the twenty five students in a class sat a maths test that was marked out of 40. The marks obtained were as shown below:

22	25	21	18	25	32	30	40	28	16	31	21
24	14	25	34	37	27	18	27	39	35	28	35

The twenty fifth student was absent for the test due to hospitalisation and so the teacher had to estimate a mark for this student in this test. Noticing that on previous tests this student usually performed above the class mean the teacher awards an estimated mark of ($\bar{x}$ + 0·6s) where $\bar{x}$ and s are respectively the mean and standard deviation of the 24 marks. To the nearest 0·5 of a mark what was the student's estimated mark.

14. The maximum and minimum temperatures, as recorded at Perth airport, for each day of December in a particular year were as follows:

Day	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Max (°C)	27·3	27·9	29·4	29·5	30·2	31·4	33·9	34·4	21·9	23·3
Min (°C)	13·9	14·0	15·7	14·6	14·6	15·0	17·0	24·9	16·1	12·1

Day	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
Max (°C)	27·0	27·2	24·7	26·2	34·5	39·2	41·2	36·2	28·7	24·1
Min (°C)	13·0	13·0	12·5	12·9	13·6	19·8	20·8	15·2	18·6	15·0

Day	21st	22nd	23rd	24th	25th	26th	27th	28th	29th	30th	31st
Max (°C)	26·0	29·1	29·2	30·0	33·0	37·4	38·6	24·7	28·7	36·0	31·9
Min (°C)	10·5	13·1	13·6	15·4	16·0	16·7	23·3	19·4	14·0	16·6	19·8

[Source of data: Bureau of Meteorology.]

(a) Determine the mean, range and standard deviation of the daily maximum temperatures featuring in the above table, giving your answers correct to one decimal place.

(b) Data collected over a period of more than 100 years, prior to the year featured above, gave Perth's mean maximum daily temperature for December as 27·4°C. Compare your mean daily maximum temperature for the above table with this long term mean.

(c) Determine the mean, range and standard deviation of the daily minimum temperatures featuring in the above table, giving your answers correct to one decimal place.

(d) Data collected over a period of more than 100 years, prior to the year featured in the table, gave Perth's mean minimum daily temperature for December as 16·3°C. Compare your mean daily minimum temperature for the above table with this long term mean.

15. The marks achieved in an exam sat by 100 candidates are shown below.

53	58	45	61	89	55	60	49	62	26	65	92	51	59	40
56	21	61	65	56	80	40	69	54	83	59	83	47	77	36
46	58	62	52	69	97	66	64	75	14	77	62	56	58	81
64	36	69	64	66	72	47	80	50	70	56	43	68	52	28
40	69	52	78	32	78	67	56	83	62	43	67	64	56	85
62	49	59	89	62	66	42	62	53	47	85	74	86	79	75
70	75	53	72	23	70	77	74	80	71					

(a) Calculate the mean and standard deviation for this set of marks (correct to 1dp).
(b) Display the data as a dot frequency graph.
(c) Show on your graph the "grade borderlines" and state the number of candidates awarded each grade given that grades were awarded as follows:

A:			exam mark	$\geq$	(mean + 1·25 × st. dev^{n}.)
B:	(mean + 0·5 × st. dev^{n}.)	$\leq$	exam mark	$<$	(mean + 1·25 × st. dev^{n}.)
C:	(mean − 0·5 × st. dev^{n}.)	$\leq$	exam mark	$<$	(mean + 0·5 × st. dev^{n}.)
D:	(mean − 1·5 × st. dev^{n}.)	$\leq$	exam mark	$<$	(mean − 0·5 × st. dev^{n}.)
F:			exam mark	$<$	(mean − 1·5 × st. dev^{n}.)

16. A company manufactures components for aircraft engines. The quality control for one particular component involves 25 of the components being randomly selected and measured from every batch of 500. If **any** one of the conditions stated below is found to apply to the sample then production is halted, each of the other 475 components in the batch is checked and the machine is re-set.

Sample reject, condition ①: mean < 64·8 cm
Sample reject, condition ②: mean > 65·2 cm
Sample reject, condition ③: standard deviation > 0·15 cm
Sample reject, condition ④: any one component > 65·3 cm
Sample reject, condition ⑤: any one component < 64·7 cm

For each of the following samples of 25 determine whether the sample passes these checks and, for any sample that does not pass, state the reason for it not passing. (All measurements are in centimetres.)

Sample A				
65·0	65·0	64·7	64·9	65·1
65·1	65·0	65·0	65·1	65·0
65·0	64·8	64·9	65·0	64·8
64·8	65·0	65·0	65·1	65·0
65·0	64·9	65·1	65·0	65·2

Sample B				
65·0	65·0	65·0	65·0	65·0
65·1	65·1	65·1	64·9	65·0
65·0	65·0	64·8	64·6	64·9
64·9	65·1	65·0	64·9	65·0
65·0	65·0	64·8	65·0	65·0

Sample C				
64·9	64·9	64·8	65·1	65·0
64·9	65·0	65·1	65·0	65·0
65·0	65·1	65·0	65·1	65·0
65·1	65·1	64·9	65·0	64·8
65·1	65·0	65·1	65·0	65·0

Sample D				
64·8	65·2	65·0	65·0	65·2
65·0	64·8	65·1	65·2	64·8
65·0	65·1	65·1	64·8	65·2
64·7	64·8	64·8	65·2	65·0
64·9	65·1	65·0	65·2	64·8

Frequency tables.

Remember that if data is given in the form of a frequency table it can be entered into many calculators in this frequency form.
For example, given 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

the 50 scores can be keyed into a calculator in this frequency form. We do not need to key in the 50 scores separately.

	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		

1-Variable
$\bar{x}$ = 16.14
Σx = 807
Σx^2 = 13205
$x\sigma n$ = 1.897472
$x\sigma n\text{-}1$ = 1.91673617
n = 50

Outliers.

Any data values that are unusually far away from the others are known as **outliers**. These extreme values can have a big effect on the standard deviation.
Suppose for example that the frequency table shown above were also to include one score of 51.

Score	10	11	12	13	14	15	16	17	18	19	20	51
Frequency	1	0	1	2	4	9	10	11	8	3	1	1

Now the standard deviation is
5·2 (to 1dp),
as shown on the right, compared to the previous value of 1·9 without the outlier.

1-Variable
$\bar{x}$ = 16.8235294
Σx = 858
Σx^2 = 15806
$x\sigma n$ = 5.18559801
$x\sigma n\text{-}1$ = 5.23719727
n = 51

For this reason outliers need careful consideration. Perhaps in this case it was simply an error involving a score of 15 being written as 51.
Sometimes we might determine the standard deviation without the outlier included and then comment on the presence of the outlier.

Do not forget though that sometimes the one result that is different to the rest is the really interesting one. Suppose for example we were testing to see the effect various drugs have on a particular disease. The drug that produces results significantly different to the other drugs could well be the important one!

Grouped data.

Just as we use the mid-point of each interval to estimate the mean for grouped data, we do the same thing to obtain an estimate for the standard deviation.

Thus for the grouped data given below:

Score	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89
Frequency	3	11	18	28	24	10	6

$n = 100$, $\bar{x} = 55{\cdot}8$, $\sigma_n = 14{\cdot}3$ (1 d.p.).

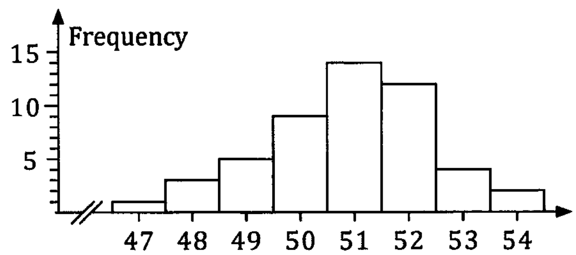
L1	L2	L3
24.5	3	
34.5	11	
44.5	18	
54.5	28	
64.5	24	
74.5	10	
84.5	6	

1-Var Stats
$\bar{x} = 55.8$
$\Sigma x = 5580$
$\Sigma x^2 = 331895$
$Sx = 14.4008277$
$\sigma x = 14.32864264$
$n = 100$

Remember: Many measured quantities are naturally grouped by rounding and the data is often displayed as a histogram.

Confirm that for the histogram shown on the right:

Mean = $50{\cdot}88$,
Standard deviation = $1{\cdot}5$ (1 d.p.).



Exercise 3C

Find the mean and standard deviation of each of the distributions shown in questions 1 to 5. (Give answers correct to one decimal place.)

1.

Score	0	1	2	3	4	5
Frequency	3	7	15	24	19	12

2.

Score	15	20	25	30	35	40	45	50
Frequency	1	3	4	8	8	4	3	1

3.

Score	15	20	25	30	35	40	45	50
Frequency	8	4	3	1	1	3	4	8

4.

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

5.

Score	10	20	30	40	50	60	70	80	90	100
Frequency	1	2	4	3	5	7	15	19	12	7

Use the midpoint of each class interval to determine the mean and standard deviation of the following distributions shown in questions 6 to 10. (Give answers correct to one decimal place if rounding is necessary.)

6.

Score	Frequency
20 → 24	7
25 → 29	12
30 → 34	18
35 → 39	20
40 → 44	24
45 → 49	13
50 → 54	6

7.

Score (x)	Frequency
$0 \le x < 10$	17
$10 \le x < 20$	13
$20 \le x < 30$	9
$30 \le x < 40$	7
$40 \le x < 50$	4

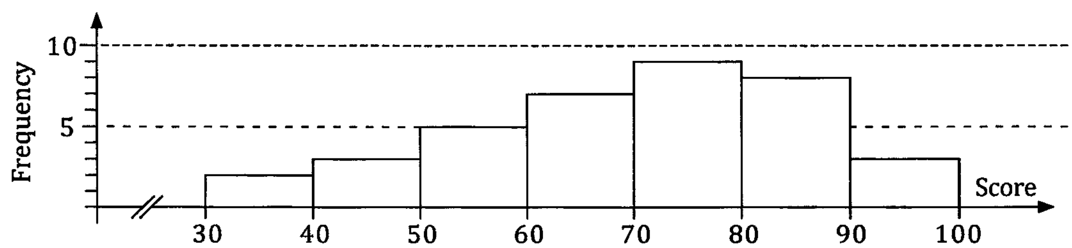
8.

Score	Frequency
0 → 9	3
10 → 19	8
20 → 29	15
30 → 39	24
40 → 49	34
50 → 59	16

9.

Score (x)	Frequency
$0 \le x < 20$	1
$20 \le x < 40$	3
$40 \le x < 60$	10
$60 \le x < 80$	9
$80 \le x < 100$	7

10.



11. A golf club organises a "club members championship" each year for the top 35 ranked players in the club. In a particular year the scores achieved by these players in the championship round were as shown below.

Score	67	69	71	72	73	74	75	76	78	82	85	91
Number of players	1	1	2	4	7	5	3	5	3	2	1	1

(a) How many standard deviations from the mean was the best (i.e. lowest) score? (Answer correct to one decimal place.)

(b) How many standard deviations from the mean was the worst (i.e. highest) score? (Answer correct to one decimal place.)

12. A company wishes to test a coating it is developing for seeds. The coating is designed to provide the seeds with essential nutrients and to stimulate germination and growth. The company arranges 200 trays each containing identical potting mix. In each of 100 of these trays 50 coated seeds are planted and in each of the other 100 trays 50 uncoated seeds are planted, all the seeds being of the same quality and type. After a certain number of weeks the company counts the number of successful germinations in each tray - a "success" being a healthy seedling at least 6 cm in height. The results were as follows:

Uncoated Seeds	
N°· of successes in tray.	N°· of trays.
21 - 25	7
26 - 30	21
31 - 35	28
36 - 40	23
41 - 45	15
46 - 50	6

Coated Seeds	
N°· of successes in tray.	N°· of trays.
21 - 25	1
26 - 30	2
31 - 35	10
36 - 40	17
41 - 45	42
46 - 50	28

Calculate the mean and standard deviation for each set of 100 trays and comment on your results.

13. The time that thirty patients had to wait beyond their allotted appointment time at a particular health centre was noted. The results are shown tabulated below.

Time (t mins)	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$	$40 \le t < 50$	$50 \le t < 60$
N°· of patients	8	15	4	2	0	1

Find the mean and standard deviation of this distribution of waiting times both with and without the outlier included. (Round to one decimal place if necessary.)

14. One hundred primary schools are surveyed regarding the number of students on the roll of each school. The information collected is shown tabulated below:

N° of students	1 to 50	51 to 100	101 to 150	151 to 200	201 to 250
N° of schools	5	5	10	9	18

N° of students	251 to 300	301 to 350	351 to 400	401 to 450	451 to 500
N° of schools	22	15	6	7	2

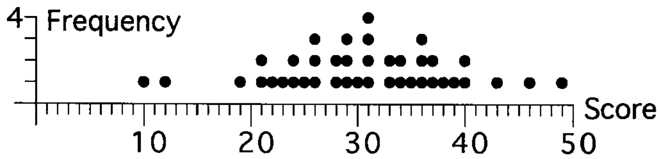
N° of students	501 to 550	551 to 600	601 to 650	651 to 700	701 to 750
N° of schools	0	0	0	0	1

(a) By taking the centre of each interval as the number of students in each school in that interval, determine estimates for the mean number of students per school and the standard deviation of the distribution.

(b) Determine the mean and standard deviation of the data once the outlier in the data is removed.

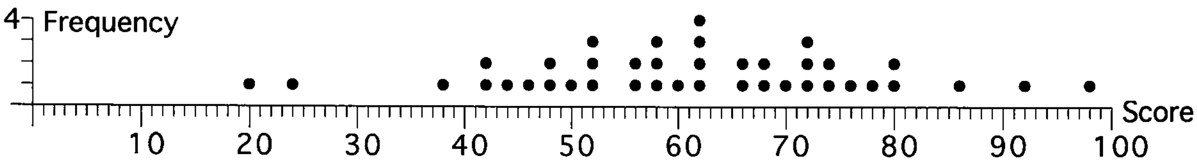
Central tendency and spread - An investigation.

Forty students sat a maths test marked out of 50. The scores they achieved are shown in the dot frequency graph below.



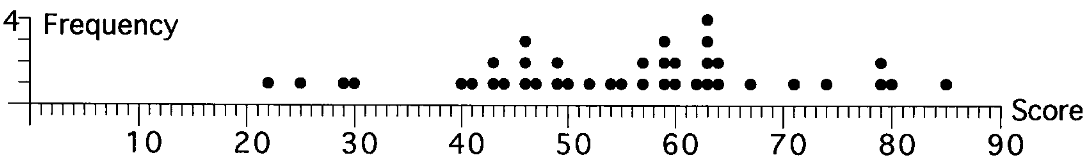
Find the mean (to 2 d.p.) and standard deviation (to 1 d.p.) of the forty scores.

The teacher wanted these scores as percentages so she multiplied each score by 2. The dot frequency graph of the percentage scores is given below.



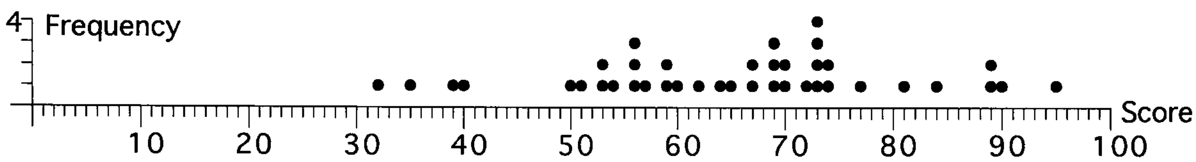
Find the mean and standard deviation of these forty percentage scores.

A second test, given to the same forty students, was in two parts. Part A consisted of 10 mental questions and was marked out of 10. Part B was more involved and was marked out of 90. The dot frequency for the scores achieved in part B is shown below.



Find the mean and standard deviation of these forty scores.

Part A proved to be very straightforward and all of the students scored 10 out of 10! The dot frequency for the combined score out of 100 is given below.



Find the mean and standard deviation of these forty scores.

Suggest how the mean and the standard deviation of a set of scores are affected

if
- the scores have the same number added to each of them (or subtracted from each them),

and/or
- each of the scores is multiplied by (or divided by) the same number.

Test your suggestions (conjectures) by putting a set of scores into your graphic calculator, creating a new set of scores by adding a number to each of the scores in the first set, and/or by multiplying each set of scores in the first set by a number, and then comparing the mean and standard deviation of the two sets of scores.

Miscellaneous Exercise Three.

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. Write the following numbers in order of size, smallest first.

   0·201 0·12 0·102 0·21 0·1 0·021 0·012 0·2

2. Without the assistance of a calculator, write the following numbers in order of size, smallest first.

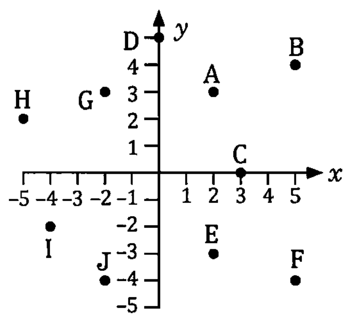
   $\frac{1}{2}$ $\frac{1}{5}$ $\frac{1}{3}$ $\frac{2}{3}$ $\frac{3}{4}$ $\frac{7}{10}$ $\frac{1}{100}$ $\frac{3}{5}$

3. Evaluate each of the following.

   (a) $3 + 2 \times 4$ (b) $5 \times 4 + 6$ (c) $3 + 5^2$

   (d) $32 \div 4 \div 2$ (e) $32 \div (4 \div 2)$ (f) $(5 + 3)^2 - 8 \times 3$

4. State the coordinates of each of the points A to J shown on the right. (All coordinates involve integer values only.)



5. The statistics regarding commercial orchard fruit in Western Australia for a particular year are shown in the following table:

Fruit	Number of trees (1000's)	Production (tonnes)	Gross value of production ($1000's)
Apples	778	37 418	19 497
Pears	175	8 399	4 886
Lemons & Limes	18	1 125	738
Mandarins	55	1 315	1 830
Oranges	183	5 304	1 830
Nectarines	147	2 333	2 333
Peaches	126	2 507	4 070
Plums & Prunes	190	3 494	4 392

[Source of data: Australian Bureau of Statistics.]

(a) What type of variable does the first column of the table involve?

For this particular year:

(b) How many commercial apple trees were in Western Australia?

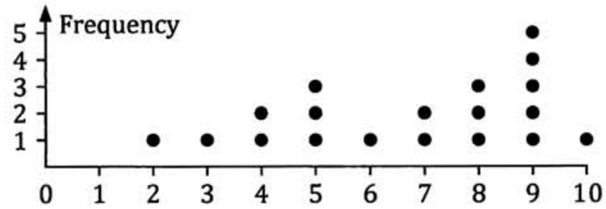
(c) How many tonnes of peaches were produced commercially?

(d) Find the gross value per tonne for (i) oranges, (ii) nectarines?

(e) On average, how many kilograms of apples did each commercial apple tree yield?

6. List one advantage and one disadvantage of using the range as an indicator of the spread or variability in a set of scores.

7. The dot frequency graph on the right shows the scores achieved by 19 students in a spelling test. Find the mean, median, mode and standard deviation of the distribution. (Round to two decimal places if rounding is necessary.)



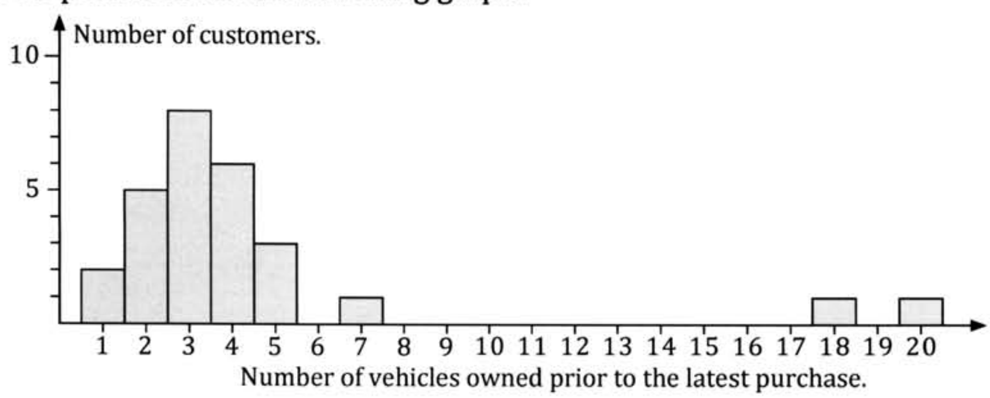
8. Ten scores have a mean of 85·4. However it was later found that one of the scores had been recorded incorrectly as 75 when it should have been 57. If this error is corrected what is the new mean of the ten scores?

9. Four maternity hospitals, A, B, C and D report to the regional health authority and state the number of live births that occurred in a particular month and the mean birth weight of the babies. The data was as follows:

Hospital A:	84 live births, mean weight 3·025kg.
Hospital B:	27 live births, mean weight 3·140kg.
Hospital C:	53 live births, mean weight 2·935kg.
Hospital D:	17 live births, mean weight 2·855kg.

Calculate the mean birth weight for all the live births from these maternity hospitals in the month that the above data applies to.

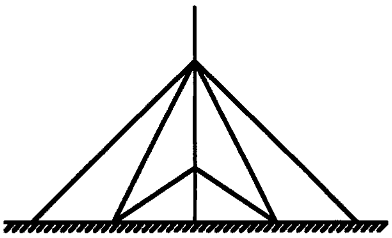
10. For a period of time, a car salesman asked customers trading in their old vehicle for a new one, how many vehicles they had owned in their lives prior to the purchase of their new one.
The responses led to the following graph:



(a) According to the graph none of the people involved in the survey said they had owned no vehicles prior to the latest purchase. Why is this?

(b) The salesman concluded that, on average, people purchasing a new vehicle from him had owned approximately 4·4 vehicles before this latest purchase. Comment on this conclusion.

11. (Pythagoras)
A vertical mast is 80 metres tall and is to be held in place by a number of wires. These wires are each one of three different lengths and are classified as short, medium or long.



Each short wire is to have one end attached to a point one quarter of the way up the mast and the other to a point on the ground, level with the base of the mast and thirty metres from it.
Each medium wire is to have one end attached to a point twenty metres from the top of the mast and the other to a point on the ground, level with the base of the mast and thirty metres from it.
Each long wire is to have one end attached to a point twenty metres from the top of the mast and the other to a point on the ground, level with the base of the mast and sixty metres from it.
The wires are to be made a little longer than is required, brought to the site, and then suitably adjusted. Each wire is to be made to "the accurate length plus 50 cm then round up to the next 10 cm".
Find the length that each classification of wire should be made to.

12. The assessment of her college course involves Suzanne sitting five exams, one in each of the units A, B, C, D and E. The twenty five students following this course all took the five exams and their results in the exam for unit A were as follows:

55	50	54	49	14	53*	50	37	37	48
40	71	20	57	61	55	9	46	30	44
50	43	48	34	43			* Suzanne's result.		

(a) Determine the mean and standard deviation for these scores.

The mean and standard deviation for the marks obtained by these students in the other four exams are shown below, together with Suzanne's score.

Unit B:	Mean	67	Standard deviation	12	Suzanne's score	61
Unit C:	Mean	37	Standard deviation	8	Suzanne's score	49
Unit D:	Mean	83	Standard deviation	5	Suzanne's score	80
Unit E:	Mean	72	Standard deviation	10	Suzanne's score	79

(b) In order to compare her marks in the five exams Suzanne decides to standardise the marks by expressing each mark in terms of the number of standard deviations the mark is above or below the mean. For example in a course having a mean of 55 and a standard deviation of 12 then a score of 67 (= 55 + 1(12)) becomes 1, a score of 79 (= 55 + 2(12)) becomes 2, a score of 43 (= 55 - 1(12)) becomes -1 etc.
List the units in order from the one that Suzanne achieved her highest standardised score to the one with her lowest standardised score and state the standardised score for each unit.