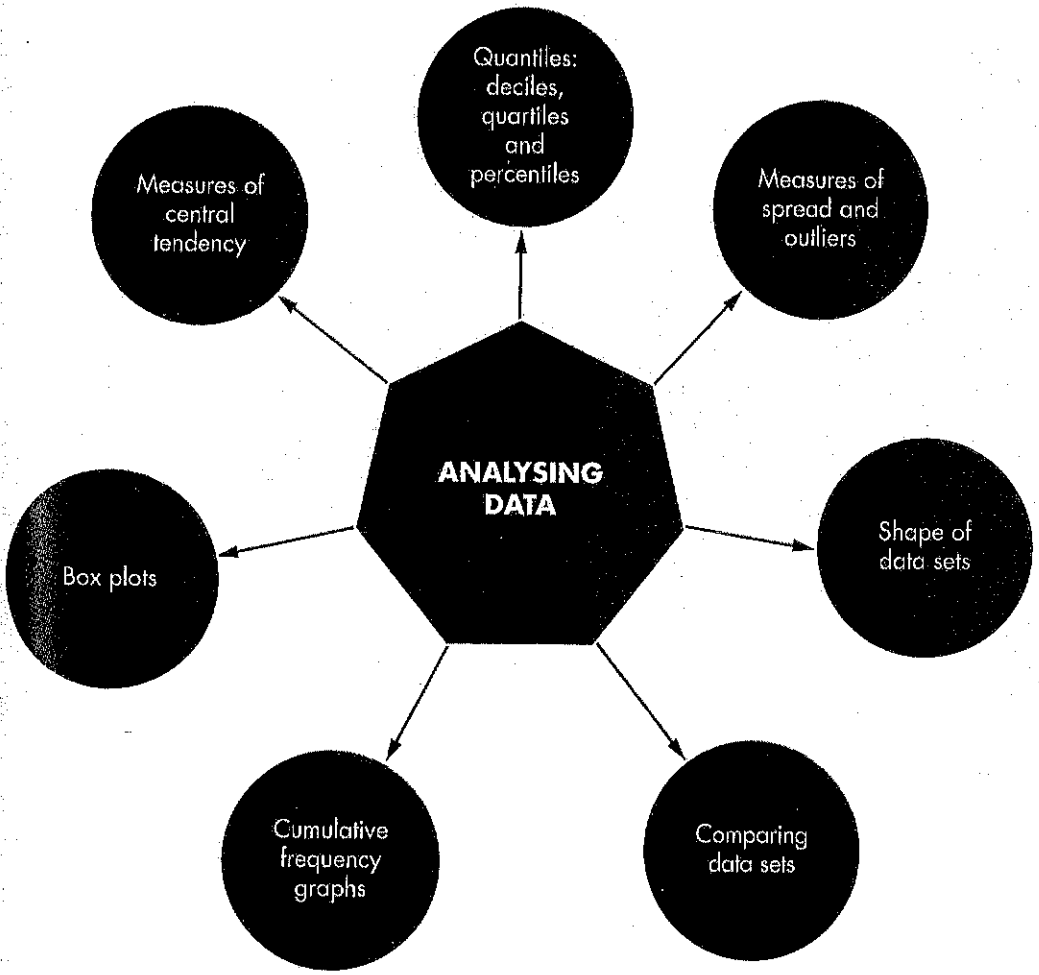
This chapter, Analysing data, examined the statistical measures of central tendency (mean, median, mode) and spread (range, interquartile range, standard deviation). You should be competent at making statistical calculations on sets of numerical data, including those represented in frequency tables, class intervals (grouped data), dot plots and stem-and-leaf plots. Make sure you know how to use the statistical functions of your calculator. You should understand the new concepts of quantiles (quartiles, deciles and percentiles), be able to interpret cumulative frequency graphs and construct box plots using a five-number summary. You must also be able to describe, compare and interpret data sets in terms of modality, shape (symmetrical and skewness), measures of central tendency and spread and also look at the effect of outliers.

Statistics review

Statistics crossword

Make a summary of this topic. Use the outline at the start of this chapter as a guide. An incomplete mind map is shown below. Use your own words, symbols, diagrams, boxes and reminders. Gain a 'whole picture' view of the topic and identify any weak areas.



ISBN 9780170413565

Exercise 10.01

1 The heights (in centimetres) of a group of ballet dancers are:

165	183	170	168	175	179	168	170
181	168	172	177	171	170	175	179

a Calculate the mean, correct to one decimal place.

b Find the median height.

c What is the mode?

Exercise 10.01

2 Motor vehicles were clocked, by police radar, travelling at the following speeds (in km/h):

78	95	64	77	81	84	77	89	90	78
79	80	82	84	80	79	95	86	84	70
78	65	82	91	89	60	85	81	78	68
90	84	69	70	80	91	85	84	80	76
68	65	85	76	79	83	82	91	84	80

a Sort the data in a frequency table using classes of 60–< 70, 70–< 80, and so on, and include a column of class centres.

b Calculate an estimate for the mean speed.

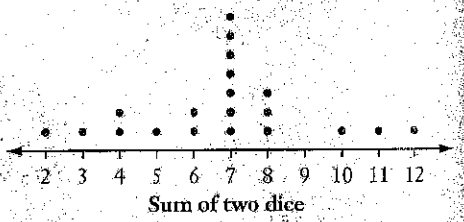
c Find the median class of speeds.

d What is the modal class?

Exercise 10.01

3 The dot plot represents the sum of two dice rolled 20 times.

Find the mean, median and mode of this data.



Exercise 10.01

4 The house prices realised at auction one Saturday in Vincentia were:

$642 000	$585 000	$352 000	$1 480 000
$705 000	$415 000	$680 000	$740 000

a Calculate the mean price.

b Calculate the median price.

c Is the mean or the median the better measure to use as the average price of the houses? Why?

Exercise 10.01

5 Which measure of central tendency is most appropriate for describing each average below? Give a reason for each answer.

a The average men's shoe size

b The average height of Year 11 students

c The average starting salary of an Australian worker

ISBN 9780170413565

6 A grouped data frequency table is shown.

Class interval	Frequency
11–15	4
16–20	7
21–25	12
26–30	24
31–35	15

What is the mean? Select **A, B, C** or **D.**

A 24.1 **B** 25.3

C 26.1 **D** 28.1

7 In a national mathematics test, Simone scored 84. Exercise 10.02

a This score was above the 7th decile, D_7. Approximately what percentage of students taking the test scored lower than her?

b More specifically, Simone's score was at the 78th percentile, P_{78}. What percentage of students scored higher than her?

8 **a** What is the meaning of 'interquartile range'? Exercise 10.03

b A random sample of 15 packets of corn chips had the following masses in grams. Find the range and interquartile range of these masses.

52 51 50 49 50 50 48 51

51 50 49 53 50 49 51

9 This stem-and-leaf plot on the right represents the number of points per match scored by the Sharks in a football season. For this data, find: Exercise 10.03

a the range

b the interquartile range.

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

10 In a small business, eight employees earn the following wages per week. Exercise 10.04

$1026 $874 $950 $950 $980 $1140 $1216 $1710

Is the wage of $1710 an outlier for this set of data? Justify your answer with calculation.

11 Consider the set of scores: Exercise 10.04

4 7 8 8 12 15 19 20

a What is the effect on the mean and median if an outlier of 40 is added to this data set.

b Is the mean or median a better measure of central tendency when there is an outlier in the data set?

ISBN 9780170413565

ISBN 9780170413565

Exercise 10.05

12 Students were surveyed about the number of pairs of shoes they owned, and the results are shown in the table on the right.

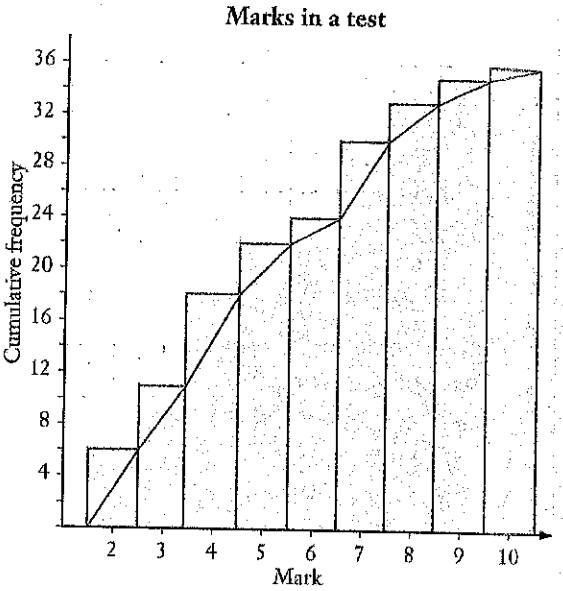
Pairs of shoes	Frequency
5	8
6	11
7	10
8	6
9	5

a Copy the table, adding a cumulative frequency column. Then draw a cumulative frequency histogram and polygon.

b Use your polygon to calculate:
- **i** the median
- **ii** the interquartile range
- **iii** the 3rd decile.

Exercise 10.05

13 The cumulative frequency graph shows the results of an assignment marked out of 10.

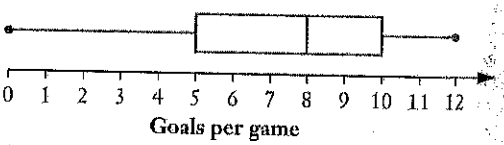


a How many students completed the assignment?

b Use the graph to estimate:
- **i** the median
- **ii** the interquartile range
- **iii** the 6th decile
- **iv** the 45th percentile.

Exercise 10.06

14 This box plot represents the number of goals scored per game by a hockey team over a season.



a What was the lowest score?

b Find the interquartile range.

c In what fraction of games were more than 8 goals scored?

d In what percentage of games were fewer than 5 goals scored?

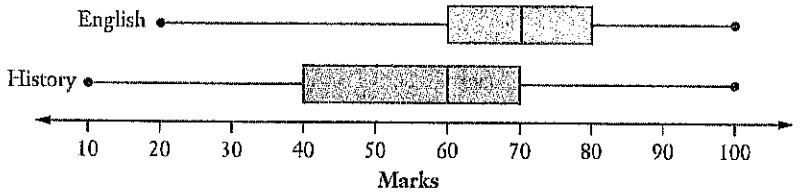
Exercise 10.06

15 **a** Create a five-number summary for the corn chip packet masses in Question **8b**.

b Represent the mass data on a box plot.

ISBN 9780170413565

16 The parallel box plots show the distribution of marks for exams in English and History.



a Which subject has the smaller spread of marks? Give reasons.

b The number of students who scored 70 or less is the same for both subjects. If 144 students did the English exam, how many students did the History exam?

17 For quality testing, a manufacturer takes a random sample of 10 screws, each designed to have a length of 2 cm. The actual lengths of the screws, in centimetres, are:

2.00 1.99 1.98 2.01 2.01 1.97 2.03 1.98 2.01 2.00

Exercise 10.07

a Find the mean screw length.

b Find the standard deviation, correct to two decimal places.

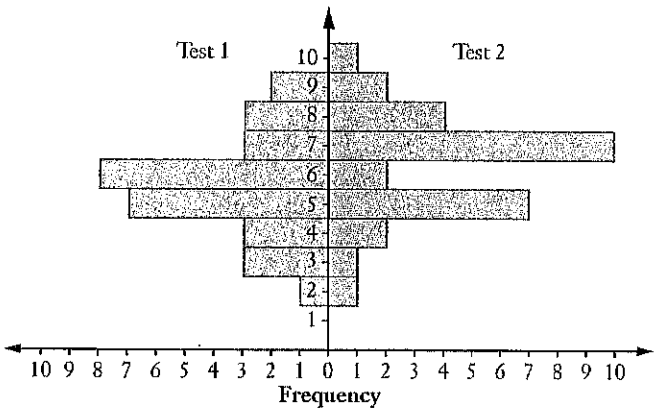
18 For the shoe data from Question **12**, calculate (correct to one decimal place):

Exercise 10.07

a the mean **b** the standard deviation.

19 The results for the multiple-choice section in two tests taken by a Year 11 Mathematics class are shown below.

Exercise 10.08



a Find the mean, median and mode for each test.

b Describe the shape of the data set for each test.

c For each test, find:

i the range **ii** the interquartile range **iii** the standard deviation.

d Are there any significant differences in the results of the two tests? Justify your answer by referring to the measures of central tendency and spread of the tests.

Chapter quiz

ISBN 9780170413565

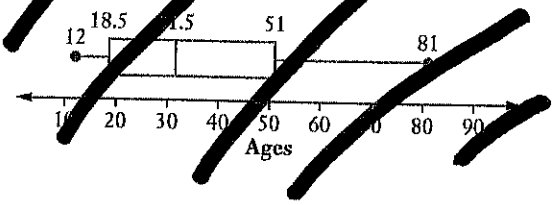
SOLUTIONS

Sample HSC problem

a 39.25 **b** 31.5

c Mean involves every score and there are no outliers.

d 51 − 18.5 = 32.5



Test yourself 10

1 **a** 173.2 **b** 171.5 **c** 168, 170

2 **a**

Speed (km/h)	Class centre	Frequency
60 – < 70	65	7
70 – < 80	75	13
80 – < 90	85	23
90 – < 100	95	7
		50

b 81 km/h **c** 80–90 **d** 80–90

3 mean = 6.8, median = 7, mode = 7

4 **a** mean $699 875 **b** median $661 000

c Median, the outlier of 1 480 000 does not affect its value as it does the mean.

5 **a** mode **b** mean **c** median

6 C

7 **a** 70% **b** 22%

8 **a** Range of middle 50% of scores

b Range = 53 − 48 = 5, IQR = 51 − 49 = 2

9 **a** 56 **b** 19

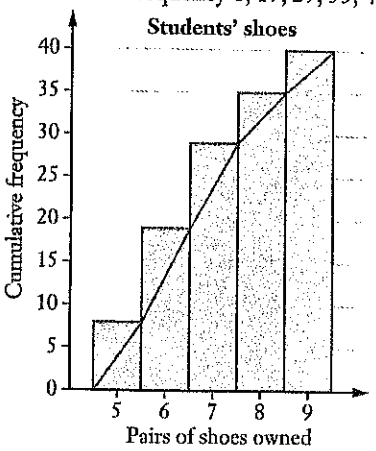
10 $Q_3 + 1.5 \times \text{IQR} = 1178 + 1.5 \times 228 = 1520$

So, 1710 is an outlier.

11 **a** Increases the mean from 11.625 to 14.778 and also increases the median from 10 to 12.

b The median – less affected by the outlier.

12 **a** Cumulative frequency 8, 19, 29, 35, 40



b **i** 7 **ii** 8 − 6 = 2 **iii** 6

13 **a** 36

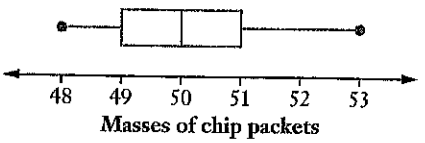
b **i** 4.5 **ii** 7 − 3 = 4

iii 5 **iv** 4

14 **a** 0 **b** 5 **c** $\frac{1}{2}$ **d** 25%

15 **a** 48, 49, 50, 51, 53

b



16 **a** English – smaller range and IQR.

b 96

17 **a** 1.998 cm **b** 0.02 cm

18 **a** 6.7 **b** 1.3

19 **a** Test 1: mean = 5.6, median = 6, mode = 6

Test 2: mean = 6.3, median = 7, mode = 7

b Test 1: Results are symmetrical.

Test 2: Results are negatively skewed.

c Test 1: **i** 7 **ii** 2 **iii** 1.74

Test 2: **i** 8 **ii** 2 **iii** 1.80

d Results of Test 2 are just better than Test 1 as mean, mode and median of Test 2 are higher than for Test 1. The spread for both tests are similar as there is only a difference of 1 between ranges, the IQRs are equal and the standard deviations are approximately equal.

ISBN 9780170413565