Statistics 1

Student Name	Class	Score
Parent Signature	Date	

14:01 | Collecting Data

Outcome DS 4.2

To collect useful statistical information, you should:
* ask suitable questions
* measure items or ask people who are representative.

Often, it is more convenient to take a sample. A sample is part of the whole group we want information on.

A sample should be chosen at random, so that each item or person has an equal chance of being chosen.

Example: John wants to find out about young people in his town and their favourite form of entertainment. He asks every fifth person leaving a cinema the question 'Did you enjoy this movie?'
a Is this sample representative? Explain.
b Is the question he asked suitable?
c Suggest how he could improve on this survey.

Answer:
a No. He is more likely to be asking people who do enjoy going to the movies compared with other forms of entertainment. He also hasn't defined what a 'young' person is, and will be surveying from all age groups at the cinema, not just young people. However, it is a good idea to ask every fifth person (cuts down on the work, and won't significantly affect the result).
b No—not suitable. They may have enjoyed this particular movie, but going to the cinema may not be their favourite form of entertainment.
c He should define young people as 'students at high school', for example. If his school is the only one in his town, he could then choose, say, 100 students at random from the school roll. The question he should ask is 'What is your favourite form of entertainment?', or he could give them a list of activities and ask them to rate them 1, 2, 3, etc.

1 Give reasons why these surveys are not representative.

a To find out whether the phones in a street are working correctly, a technician checks the first ten phone lines.

_____

_____

_____

b To survey whether people can afford to go to the dentist, Kaye hands out a questionnaire at her local dentist.

_____

_____

_____

c To survey whether people in NSW are happy with public transport systems, a market research company chooses 1000 people in Newcastle.

_____

_____

_____

d Keith does household employment surveys. To find out the percentage of people in Sydney in full-time work, he chooses names at random from the Sydney phone directory and phones them between 9 a.m. and 5 p.m.

_____

_____

_____

2 A market-research firm does a door-to-door survey in a small town. The survey will measure attitudes about people with disabilities. Explain which of the following methods would give the most accurate information.

A asking the person who opens the door for their opinion

B asking to speak to the person who had the most recent birthday

_____

_____

_____

3 Lee is carrying out a survey to find out the weights of bags students bring to her school. Here are some possible ways she could take a sample. Write down whether each method will be representative. If it is unrepresentative, explain why.

a She could survey all students who cycle to school.

_____

_____

_____

b She could get a copy of the school roll and choose 100 students at random from it.

_____

_____

_____

c She could survey all Year 8 students.

_____

_____

_____

4 At the beginning of a sports program on TV, the presenter asked people to phone in during the program and answer this question:

Should the coach of the NSW cricket team resign?

Phone 190-566-566

Press 1 for 'Yes'
Press 2 for 'No'
Press 3 for 'Don't know/care'

Calls cost 99 cents a minute.
Children—ask your parents first.

At the end of the program the presenter gives the results:

'Yes'	6780	63%
'No'	3954	37%
'Don't know/care'	4	0%

There are several faults with a survey like this. Which of these are good reasons why the survey is faulty? Answer true or false for each one.

a The survey is inaccurate because not enough people replied. _____

b The survey is unrepresentative because only people watching the program knew it was taking place. _____

c The survey is inaccurate because people might have telephoned more than once. _____

d The results are misleading because people who did not care would not bother spending money on replying to it. _____

e The results are not accurate because some people may not have been able to get through on the phone number. _____

5 Caterina is planning a survey on television-watching habits of her fellow students. Explain why these questions are not suitable.

a 'Did you watch Channel 7 or Channel 9 last night?'

_____

_____

b 'What channels did you watch last night? Tick all boxes that apply.

ABC ☐   7 ☐   9 ☐   10 ☐

_____

_____

c How many hours of television did you watch last month?

_____

_____

6 Terri is investigating whether people are satisfied with the service at their local bank. She only asks one question of each person.

Terri: 'Hello Mrs A! Did you have to wait too long in a queue the last time you visited the bank?'

Mrs A: 'No—I was second in line.'

Terri: 'Excuse me, Mr B. I'd like to ask you about your impressions of the service at the bank. How long did you have to wait in line last time you were there?'

Mr B: 'Five minutes, and it seemed like ages!'

Terri: 'Hi Glen! Whadd'ya reckon about the service at the bank here?'

Glen: 'Gidday Terri! It's a cool bank—no worries!'

a Give a reason why Terri's questions are unlikely to provide useful information.

_____

_____

b Design a questionnaire that would give more useful information about people's attitudes to the service at their bank. Use your own paper.

Statistics 2

Student Name		Class	Score
Parent Signature		Date	

14:02 | Sorting Data

Outcome DS 4.1

A **frequency distribution table** shows how many there are of each item. It saves having to list each value individually.

Example: Here are the ages of the players in the school orchestra:

12, 12, 12, 13, 13, 13, 13, 13, 13, 13, 13, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 18, 18

Show this information in a frequency distribution table.

Age of student	Tally	Frequency
12					3
13	⊬⊬ |||	8
14						4
15	⊬⊬ ||	7
16	⊬⊬	5
17		0
18				2

1 This is a list of the number of 'Scratch and Win' tickets each customer buys at a Tattslotto shop.

0, 2, 1, 0, 1, 4, 3, 1, 1, 0, 2, 0, 1, 6, 0, 2, 0, 1, 3, 1, 0, 2, 1, 3, 4, 1, 0, 0, 1, 2, 1, 0, 1

a Complete this frequency distribution table to summarise the information.

Number of tickets	Tally	Frequency
0		
1		
2	⊬⊬	5
3		
4		
5		
6		

b What was the highest number of tickets bought by any of these customers? _____

c A customer walks into the shop. What is the most likely number of tickets they will buy? Explain

_____

_____

2 This frequency distribution table shows the number of people who live in each of the houses in Algebra Avenue.

Number of people	Frequency
0	2
1	5
2	12
3	8
4	14
5	11
6	3
7	1
Total	

a How many houses have three people living in them? _____

b How many houses have no one living in them at present? _____

c Add up the numbers in the frequency column, and write the sum in the empty cell. Does this give the total number of *houses* or the total number of *people*? _____

3 This frequency distribution table gives the number of each type of Australian coin collected from a coin phone.

Coin	Frequency	Value
5c	33	
10c	49	$4.90
20c	80	
50c	35	
$1	71	
$2	23	
Totals		

a Complete the entries in the table by calculating the value for each type of coin, and working out the two totals.

b How many coins were in the coin phone? _____

c How much had been spent on calls from the phone? _____

14:03 | Analysing Data (Part 1)

The **range** of a set of values is the difference between the highest value and the lowest value.

Example 1: Calculate the range of these temperatures:

18.9, 15.6, 22.1, 20.3, 17.0, 16.9, 24.6, 17.1

Answer: The range of the temperatures is $24.6 - 15.6 = 9°C$

The **median** is the middle value when all the values are placed in order.

Example 2: The median of the numbers {0, 9, **31**, 38, 42} is 31.

Calculate the median of the eight temperatures above.

Answer: Arrange the values in order first!

15.6, 16.9, 17.0, **17.1**, **18.9**, 20.3, 22.1, 24.6

There are two middle values (take half-way between these):

$$\text{Median} = \frac{17.1 + 18.9}{2} = \frac{36}{2} = 18°C$$

The **mean** of a set of numbers is calculated by dividing the total of the numbers by how many numbers there are.

$$\text{Mean} = \frac{\text{sum of the values}}{\text{number of values}}$$

Example 3: Calculate the mean of the numbers 3, 8, 9, 11, 0, 11.

Answer: There are six values altogether.

$$\text{Mean} = \frac{3 + 8 + 9 + 11 + 0 + 11}{6}$$
$$= \frac{42}{6}$$
$$= 7$$

(Note that 0 was included as one of the values.)
The mean gives us information about the sum of the numbers.

Example 4: The mean weight of the eight forwards in the school rugby team is 71.5 kg. What is their total weight?

$$\text{Total weight} = \text{number of values} \times \text{mean}$$
$$= 8 \times 71.5$$
$$= 572 \text{ kg}$$

The mean can also be calculated from a frequency distribution table.

To calculate the mean from a frequency distribution table you add a third column. This is headed 'number × frequency' or $x \times f$.

Example 5: This table shows the number of hires each day for a taxi one month.

Number of hires, x	Frequency, f	$x \times f$
3	1	3
4	6	24
5	5	25
6	7	42
7	9	63
8	0	0
9	2	18
Total	30	175

There were nine days on which there were seven hires, and this contributes 63 towards the overall total.

$$\text{Mean number of hires}$$
$$= \frac{3 + 24 + 25 + 42 + 63 + 0 + 18}{30}$$
$$= \frac{175}{30} = 5.83 \text{ (2 dec. pl.)}$$

The **mode** can also be read off the table. It is the value with the largest frequency. In the above example, the mode is 7, because it occurs nine times (more than any other value).

1 Write down the range and median of each of these sets of numbers.

a {12, 19, 22, 28, 31} _____ _____

b {0, 6, 9, 11, 19, 20} _____ _____

2 Write these sets of numbers in order first. Then write down the range and median of each set.

a {49, 62, 31} _____

b {7, 5, 9, 8, 3, 1, 8, 6, 5} _____

c {31, 38, 27, 59, 41, 30} _____

Statistics 3

Student Name		Class	Score
Parent Signature		Date	

14:03 | Analysing Data (Part 2)

Outcome DS 4.2

1 Calculate the mean for each of these sets of numbers.

   a {4, 8, 12, 4, 1, 1}　　_____

   b {40, 50}　　_____

   c {21, 0, 19, 20}　　_____

2 Ten numbers add up to 89. What is their mean?

_____

3 Calculate the mean for each of these sets of numbers. Give each answer correct to two decimal places.

   a {84, 31, 101, 6, 47, 89, 49, 55, 111, 39, 98}

_____

   b {1083, 417, 37.8, 946} _____

4 A parking attendant empties out the contents of a parking meter. These are the coins in there.

20c, $2, 50c, 50c, 10c, $1, 50c, 20c, $1, 20c

Calculate the mean value of the coins. _____

5 A rowing 'eight' has a mean weight of 86.375 kg. Calculate their combined weight. _____

6 A rugby pack of eight schoolboy players with a mean weight of 62 kg are pushing against a pack of six adult players with a mean weight of 81 kg. Which pack is heavier? Explain.

_____

_____

7 Kevin has an after-school job delivering advertising. After nine weeks he has earned a mean of $24 per week. How much does he need to earn in the tenth week to increase his mean weekly earnings to $25? _____

_____

8 Here are the times for ten swimmers to complete one length of a swimming pool.

Swimmer	Time	Swimmer	Time
Desley	1 min 12 s	Anita	59 s
Frank	49 s	Julie	1 min 17 s
Gary	50 s	Sunil	1 min 5 s
Maria	49 s	Colin	51 s
Barry	1 min 17 s	Eun-Wah	1 min 3 s

   a Calculate the mean time (remember to change minutes to seconds if necessary). _____

_____

   b Which swimmer swam closest to the mean time? _____

9 Jelena and her brother have a mean height of 165 cm. Her brother is 172 cm tall. How tall is Jelena? _____

_____

10 A roadside stall has some avocados for sale at $2 a bag. These are the coins in the 'honesty' box on Tuesday.

   5 × 20c coins　　2 × 50c coins
   2 × $1 coins　　1 × $2 coin

   a What is the median value of all the coins?

_____

_____

   b Calculate the range. _____

   c On Wednesday there were 24 coins in the box. The mean value of the coins was 25 cents, and the median value was 20 cents. Which gives better information about the number of bags sold—the mean or the median? _____

_____

11 Here is a summary of Derryn's toll bill for her telephone:

> Number of calls = 23
> Total amount = $63.75
> Mean time per call = 2.35 minutes

Use this information to estimate, rounding your answers sensibly:

a the total time she spent on the phone making toll calls _____

b the mean price per call _____

12 Ingrid surveyed her classmates about the number of siblings (brothers and sisters) each one had. The results are in the table.

Number of siblings	Frequency	$x \times f$
0	2	
1	7	
2	8	
3	7	
4	0	
5	2	
Totals		

a How many 'only' children (no brothers or sisters) were in the class? _____

b How many students were surveyed?

_____

c Complete the third column in the table and write in the totals.

d Write down and complete the calculation that gives the mean number of siblings per student.

_____

e What is the mode number of siblings per student? _____

13 *'The average politician is a middle-aged male.'*

Which of the three terms for average (i.e. mean, median, mode) is the statement referring to?

14 The Fiorentina (an Italian restaurant) takes bookings for Saturday nights. Here are the numbers booked per table last Saturday:

2 2 4 2 4 6 4 4 4 2 2 6 2 2 4 5 2
2 4 3 4 6 2 8 2 4 2 4 2 2 2 2 4

a Calculate the mode. _____

b Calculate the median. _____

c Calculate the mean. _____

d The chef needs to decide how many vegetables to buy on Saturday morning. Which average would give the most useful information for him? Explain.

_____
_____
_____

e Which average would give the most useful information for the head waiter when she decides to buy some more tables? Explain.

_____
_____
_____

15 Garry works for the council. His job is to empty coins (ranging from 5c to $2) from parking meters and count the takings.

a Which average of the coins' values (mean, median or mode) must be a whole number?

_____

b Which average can be used to calculate the total amount of money? Which extra piece of information would be needed to do this?

_____

16 Write down a set of five numbers (not all the same) that have a mean = mode = median = 8.

_____

Fun Spot

A pair of children in a family with four children are twins. The other two children are 19 and 16. The median age of all four children is 14. How old are the twins?

_____

Statistics 4

Student Name	Class	Score
Parent Signature	Date	

14:04 | Grouped Data

Outcome DS 4.1

When data is measured, all or nearly all of the values are different.

You can **group** similar items together to summarise them conveniently.

Example 1: This frequency distribution table summarises the odometer readings on 45 cars for sale at a used-car dealership.

Odometer reading (km)	Frequency (f)
20 000 –	2
30 000 –	11
40 000 –	24
50 000 –	5
60 000 – 70 000	3

Grouped continuous data is displayed in a graph called a frequency histogram.

Each group is represented by a wide column. The columns are joined together.
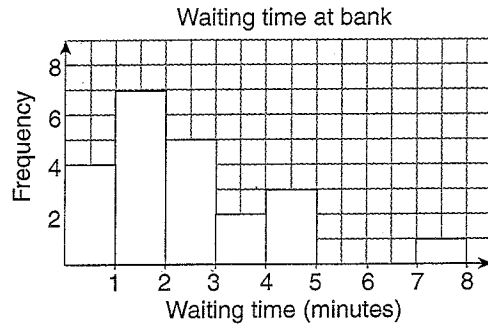
The frequency is shown on the vertical axis.

The horizontal axis shows what is being measured.

Example 2:

Odometer reading (km)
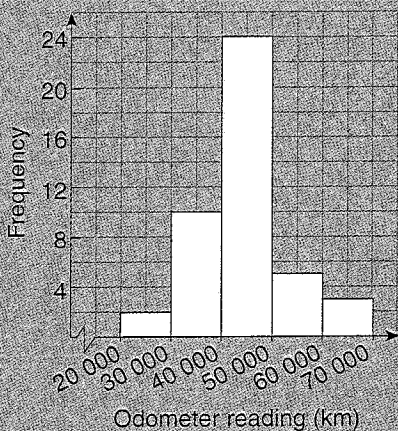
1 Peta carried out a survey to find out how long some customers had to wait in a queue at a bank. This frequency histogram shows the results.

Waiting time at bank

a How many customers had to wait less than a minute? _____

b How many customers had to wait between 3 and 6 minutes? _____

c How many customers were surveyed altogether? _____

d Explain why you cannot work out how many customers waited more than $2\frac{1}{2}$ minutes from this information.

_____

_____

2 Here are 25 consecutive docket totals from the cash register in the 'Express Lane' at a supermarket checkout.

$42.15	$3.21	$17.95	$26.62	$3.70
$12.30	$53.00	$39.04	$18.09	$16.26
$24.06	$2.35	$29.39	$16.30	$8.79
$12.31	$13.34	$6.90	$39.23	$22.31
$20.18	$4.55	$4.95	$16.10	$10.50

a Copy and complete this frequency distribution table to summarise the data.

Total of docket	Frequency
$0–$9.99	
$10–$19.99	
$20–$29.99	
$30–$39.99	
$40–$49.99	
$50–$59.99	

b How many dockets had totals of less than $20?

_____

3 Twenty pumpkins were weighed in a pumpkin-growing competition in Dubbo. Their weights to the nearest kilogram were:

25 43 27 19 30 31 37 21 33 45
39 27 18 31 46 37 32 34 20 28

a Copy and complete this table to show the weights.

Weight interval (kg)	Tally	Frequency
18–22	| | | |	4
23–27		
28–32		
33–37		
38–42		
43–47		

b How many of the pumpkins weighed less than 28 kg? _____

c Display this data in a frequency histogram using 20, 25, 30, 35, 40, 45 as the midpoints of the intervals on the horizontal axis.

14:05A | The Dot Plot

- A dot plot uses a marked scale.
- Each time an item is counted, it is marked by a dot.

Example: This graph shows the number of passengers in the school minibus for all the journeys one week.

The journeys in a week
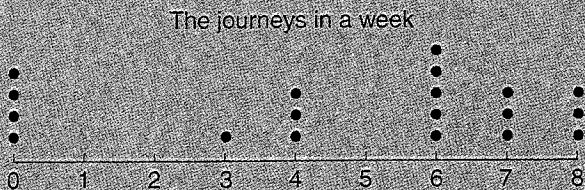
It shows that there were four journeys with no passengers, and that the highest number of passengers carried was eight.

1 Use the minibus dot plot above.

a How many journeys were made altogether?

_____

b What was the most common number of passengers carried? _____
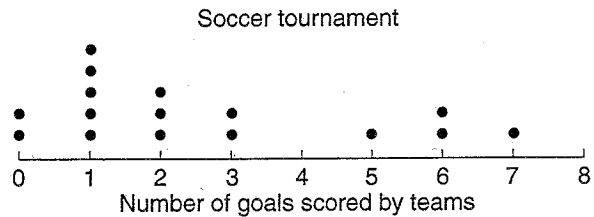
2 Peter visited every petrol station in his town to find out how many pumps each one had. Here are the results:

8, 7, 8, 9, 4, 8, 2, 6, 8, 7, 10, 8

Display this information in a dot plot.

3 Explain how you would change the dot plot above for the school minibus to show the number of *occupants* instead of the number of passengers.

_____

_____

4 This dot plot shows the number of goals scored by each of the 16 teams at a one-day soccer tournament.

Soccer tournament

Number of goals scored by teams

a What was the most number of goals scored by a team? _____

b What was the most common number of goals scored by these teams? _____

c How many goals were scored altogether? ____

5 Here is a list of the number of students in the ten classes in Year 8 at Whitlam College.

Class	Number of students	Class	Number of Students
8Ar	33	8Mn	32
8Bh	30	8Re	28
8Ct	29	8Sa	30
8Ew	30	8St	31
8Lk	31	8Wo	30

Display the information on a dot plot.

Chapter 14

Student Name		Class	Score
Parent Signature		Date	

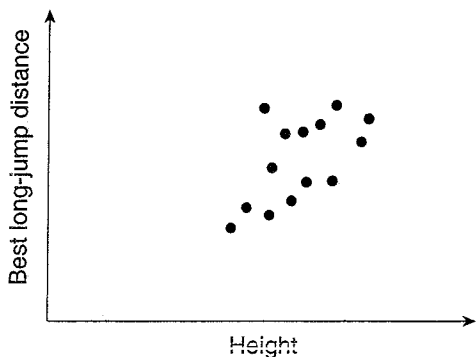## 14:05B | The Scatter Diagram

Outcomes DS 4.1–4.2

> A scatter diagram shows the relationship between two quantities. It has two axes, drawn at right-angles to each other. Each axis shows a different quantity and usually shows a scale.

1 This scatter diagram shows the heights and ages of five children in a family.
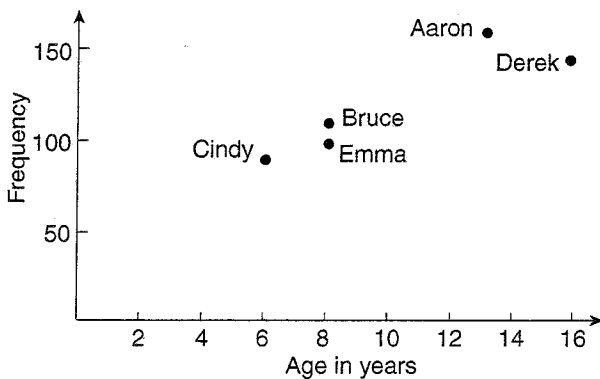
a Who is the youngest of the five children?

_____

b Who is the tallest of the five children?

_____

c How does the graph show that Bruce and Emma are twins?

_____

_____

d Describe what the graph shows about the relationship between age and height for these children.

_____

_____

2 This is a scatter diagram which shows the height and best long-jump distance for 14 Year 8 students.

Explain what the scatter diagram tells you about the relationship between the height and the distance a student can jump.

_____

_____

3 This graph shows the relationship between health expenditure per person and the GDP per person for a number of countries. GDP is a measure of the value of goods and services produced in a country.

Relationship between health expenditure and GDP in the OECD
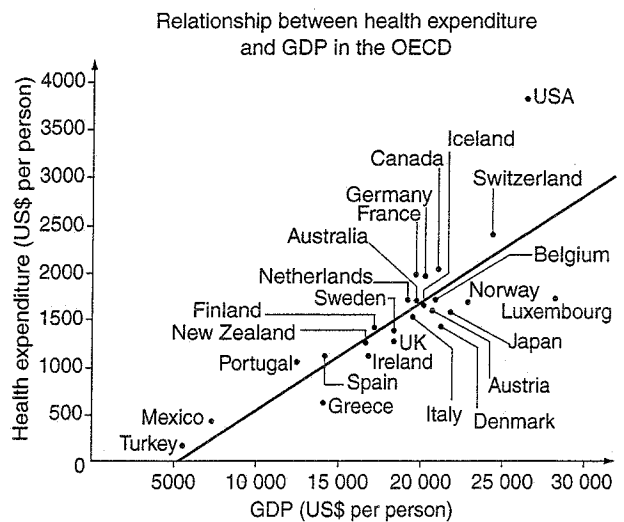
a Use the graph to estimate how much was spent on health *per person* in Australia. _____

b Which country spent more than twice the amount per person on health than Australia did?

_____

c Which country had the closest GDP per person to Australia? _____

d Which countries spent about the same on health expenditure per person as Australia?

_____

e Explain what the graph shows about the *relationship* between GDP and health expenditure per person for these countries.

_____

_____

_____

14:06 | Stem-and-leaf Plots

A stem-and-leaf plot is used to record and organise data. The most significant digit(s) go in a vertical *stem*. The last digits of the data are placed alongside in a *leaf*.

Example: This stem-and-leaf plot shows passenger loads on the 31 days of August last year for flight QF081 from Sydney to Adelaide. Two types of aircraft were used—a Boeing 737 and a Boeing 767 (on Fridays).

```
20 | 0
19 | 2
18 | 1 3
17 |
16 | 6
15 |
14 |
13 |
12 |
11 | 0 0 1 1 2 4 4 4 4 4 4
10 | 0 1 1 1 2 2 5 5 6 7 8
 9 | 4 7 7 9
```

The graph shows the most number of passengers carried was 200.

1 Construct a stem-and-leaf plot for these scores in a golf tournament.

77, 82, 71, 69, 80, 105, 91, 84, 88, 73, 71, 77, 81, 85, 92, 85, 76, 68, 92, 83

2 This stem-and-leaf plot shows a set of pulse rates taken from a class of 32 students after a five-minute 'step' test. The figures were taken over a one-minute period as soon as the test finished.

```
 7 | 8
 8 | 2 5
 9 | 6 7
10 | 0 4 6 6 9
11 | 1 2 2 8
12 | 2 5 6 7 7 9
13 | 0 0 2 5 5 5 6
14 | 3 4 7
15 |
16 | 2 4
```

a What was the mode pulse rate? _____

_____

b Work out the median pulse rate. _____

_____

c Work out the range. _____

_____

3 Refer to the stem-and-leaf plot for flight QF081 in the worked example.

a What was the least number of passengers on these flights? _____

b Work out the median. _____

c How many Fridays were there last August?

_____

d What do you think is the passenger load of a 737 when full? _____

_____

4 This 'back-to-back' stem-and-leaf plot shows the number of goals scored and conceded by the school's basketball team in its 12 games last season.

```
   Scored  |   | Conceded
           | 7 | 3
           | 6 | 2 2 5
   9 8 2 2 1 | 5 | 1 5 7 7
 8 7 7 3 1 0 | 4 | 8 9
         5 | 3 | 6
           | 2 | 9
```

a Calculate the median score for and the median score against. _____

_____

_____

b The diagram gives very little information about the number of games won and lost.

  i Explain why the team must have won at least one game.

_____

_____

_____

  ii What is the most games the team could have won? _____

_____