

# *Data*

## **Data**

Data is collected for many reasons.

- A clothes shop may want to find the types of clothes their customers would like stocked.
- A local government may want to find if a pedestrian crossing is needed in a certain street.
- A farmer may need to find whether to water more regularly.
- A company producing string may need to find the strength of their product.

Many organisations or groups collect data. For example:

- All levels of government
- Retailers
- Companies
- Scientific and research organisations
- Educational organisations
- Community groups
- Sporting and recreational organisations
- Workplaces

When collecting data the following steps need to be followed:

- What data is needed?
- What is the best way to collect the data?
- What is the best way to present the data?
- How is the data to be interpreted?

The Australian census is conducted every five years and is completed by every Australian. People answer questions about age, where they live, religion, etc. This information is available from the Australian Bureau of Statistics (ABS) and is used for many reasons such as planning for the future.

## Collecting Data

There are several ways that data can be collected.

- From the many ***data collecting organisations*** (for example: the Australian Bureau of Statistics).
- ***Observation*** (for example: standing near a road and counting the number of cars)
- ***Questionnaire*** (for example: asking a number of people to answer a set of questions) - this can be achieved face-to-face, or by mail, email, telephone, etc.
- ***Experiment*** (for example: testing a number of car windows to see if they break under certain conditions)

Collecting data can be a time consuming and expensive process so it is important that sufficient planning be put into preparing for the collection of data.

If a ***questionnaire*** is being designed it is important that the questions are simple, clear and relevant. It is also important that the information can be easily collated such as ***tick the box*** or a ***rating scale***.

### *Examples*

Do you think the main street needs more rubbish bins? *Yes*  *No*

How is the quality of your local street? *poor*  *good*  *excellent*

How well is the local radio station meeting the needs of the community? Use the rating scale of 0 - 10 (0 for not meeting the needs to 10 for meeting all the needs).

Where would you like to take your next holiday?

*beach*  *snow*  *outback*  *city*  *overseas*  *other*

**EXERCISE 16A**

1. Choose which method (*observation* or *questionnaire*) would be best to use for the collection of the following data.
  - (a) How many students at a secondary school want to proceed to tertiary studies?
  - (b) How many students at a school eat at least two pieces of fruit each day?
  - (c) How many cars are in a supermarket car park?
  - (d) A local government wants to find out if the residents want a swimming pool?
  - (e) How many houses in a city have solar water heaters?
  - (f) How many families in a city own more than one television?
  - (g) How many supporters at a sporting event are wearing team colours?
  - (h) How many people in a city follow the local sporting team?
  - (i) How many people in a small town do their shopping in the town and how many shop in the major city nearby?
  - (j) How many families have more than one car?
  - (k) How many people have been overseas on a holiday?
2. Think of an experiment that could be used in the following situations.
  - (a) A producer of light globes wants to find the average life of their light globes.
  - (b) The manufacturer of a hay baling machine wants to find the average mass of the hay bales.

- (c) A producer of balloons wants to find out the average diameter of the balloons before they burst.



3. For the following situations of data collection, suggest questions for a questionnaire.
  - (a) A local council is considering whether to build a swimming pool.
  - (b) A radio station wants to find the best music to play.

## Samples and Population

When collecting data, it is often not possible to include every person or object that is affected. In these situations a smaller representative sample is used.

In statistics, the total number of people or objects affected is called the *population*. A smaller representative group is called a *sample*.

*For example*, a local council wants to find out if the residents want a library or a swimming pool. 500 residents are surveyed to get an idea of what they prefer.

In this example all the residents in the area would be the *population*. The 500 that were surveyed is the *sample*.

To make the results of a survey as accurate as possible:

1. The sample must be as large as possible.
2. The people or items surveyed must be a true representation of the population.

It has been found that the minimum acceptable size for the sample is the square root of the population.

$$\text{sample} = \sqrt{\text{population}}$$

### *Example*

For a town that has 4500 people, an acceptable sample to survey would be  $\sqrt{4500} = 67$  (Round to the nearest whole number)

The data found from the sample can then be converted to percentages and applied to the entire population.

### *Example*

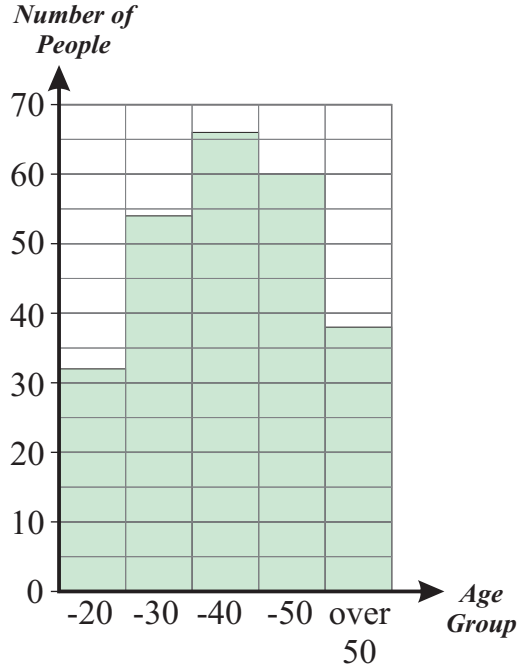
There were 850 students at a school. 60 students were asked if they played a musical instrument. To get a true representation it is important that a cross-section of the student population is involved in the sample (boys/girls, students from each year level). If 45 out of the 60 students played a musical instrument this is 75%. It can be assumed that 75% of the students in the school play musical instruments.

**EXERCISE 16B**

1. What is the minimum sample size that should be surveyed for the following populations. Round to the nearest whole number.
    - (a) A school of 880 students.
    - (b) A city with 260 000 people.
    - (c) A state with 2 500 000 people.
    - (d) A penguin colony of 8000.
  
  2. In a school of 900 students a survey was to be conducted using a questionnaire. All classes in the school had 25 students. It was decided to give the questionnaire to two students in each class. Is this an adequate number of students for the questionnaire to be considered acceptable?
  
  3. A farmer had 3500 sheep. He wanted to weigh a sample number of the sheep to check their weight.
    - (a) What is the minimum number of sheep he should sample?
    - (b) He decided to weigh 80 sheep and 12 were underweight. What percentage of the sample was underweight?
    - (c) Based on this sample, how many of the total flock would the farmer expect to be underweight?
- 
4. A football club had 25 400 members. The club wanted to survey the members to see how many wanted a monthly newsletter.
    - (a) What is the minimum number of members that should be surveyed?
    - (b) The club surveyed 400 members and 328 wanted a newsletter. What percentage of the members surveyed wanted a newsletter?
    - (c) Based on this sample, how many of the members wanted a newsletter?
  
  5. 6000 cars used a major road between 8 am and 9 am on a weekday. The number of people in 200 of the cars was noted. 30 had four people, 50 had three people, 75 had two people and the remainder had one person. Based on these figures, how many people used the road between 8 am and 9 am?

6. 5500 people attended a music festival. 250 of them were asked their age and the results are shown in this column graph.

- (a) Based on these figures, what percentage of the people at the festival were in each age group? Give answers correct to one decimal place.
- (b) Based on these figures, how many people at the festival were in each age group?



7. The tally sheet below shows the results of observing children as they left a show and recording how many show bags they had bought.

<i>Number of Show Bags</i>	<i>Tally</i>
0	
1	
2	
3	
4	
5	
6	

There were 8500 children who visited the show. Based on the above sample, find the following. Round answers to the nearest whole number.

- (a) How many children who visited the show bought:
- no show bags?
  - less than three show bags?
  - more than four show bags?
- (b) What was the total number of show bags purchased?

## Types of Data

There are three main types of data:

1. **Continuous numerical** - the data is numerical and can take any value (within a practical range).
2. **Discrete numerical** - the data is numerical but can take only certain values.
3. **Categorical** - the data is not numerical but falls into different categories.

1. **Examples of continuous numerical** data:

- height of people
- length of someone's hair

2. **Examples of discrete numerical** data:

- number of people at a theatre
- goals kicked by a footballer

3. **Examples of categorical** data:

- colour of people's eyes
- student's favourite sport

### EXERCISE 16C

Classify the following data as *continuous numerical*, *discrete numerical* or *categorical*.

1. Number of people in the audience of a concert.
2. Favourite holiday destinations.
3. Volume of water in a reservoir.
4. The number of bricks needed to build a house.
5. Price of a car.
6. Time taken to swim 100 metres.
7. Height jumped by a pole vaulter.
8. Level of satisfaction with the performance of a car - poor, good, very good, excellent.
9. Number of passengers on an aeroplane.
10. Number of points scored by a football team in a season.
11. Amount of ink remaining in an ink cartridge.
12. Length of a telephone conversation.

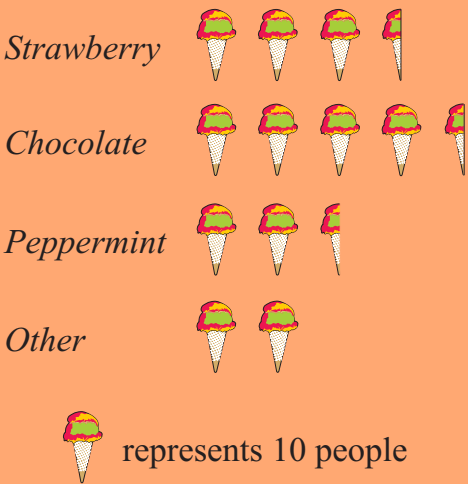
# Presenting Data

Popular methods of presenting discrete numerical and categorical data are:  
 Pictographs, Column Graphs, Bar Graphs and Pie Charts

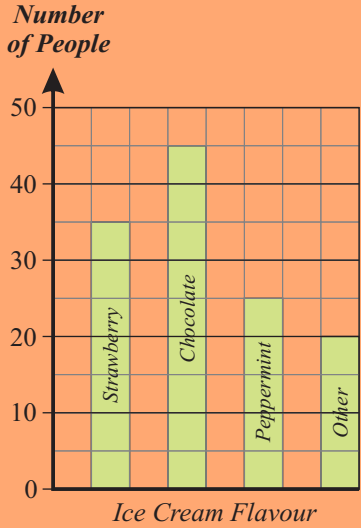
*Example*

<i>Favourite Ice Cream Flavour</i>	<i>Number of People</i>
Strawberry	35
Chocolate	45
Peppermint	25
Other	20

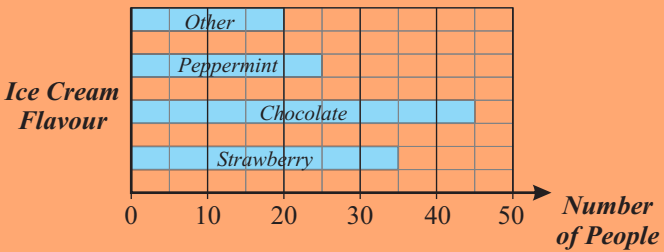
*Pictograph*



*Column Graph*



*Bar Graph*





**Pie Chart**

The angles in a pie chart can be calculated by using one of these formulae.

$$\frac{\text{number}}{\text{total number}} \times 360$$

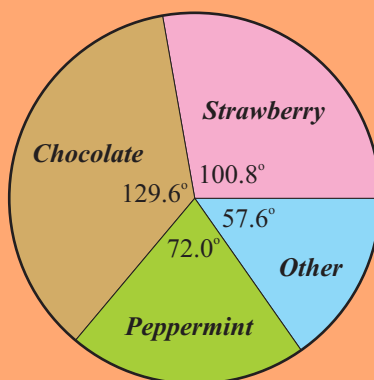
or

$$\frac{\text{percentage}}{100} \times 360$$

**Example** the angle for strawberry =  $\frac{35}{125} \times 360$   
=  $100.8^\circ$

<i>Favourite Ice Cream Flavour</i>	<i>Number of People</i>	<i>Angle</i>
Strawberry	35	$100.8^\circ$
Chocolate	45	$129.6^\circ$
Peppermint	25	$72.0^\circ$
Other	20	$57.6^\circ$
total = 125		

The pie chart can be drawn accurately using the calculated angles.

**EXERCISE 16D**









- A survey was conducted to find what colour car people would prefer. The results were:  
**white** - 452, **red** - 224, **blue** - 315, **green** - 96, **black** - 62, **other** - 89  
What would be the angle for each colour on a pie chart?  
Give answers correct to one decimal place.

2. The students in year 10 at a school were asked what they planned to do at the end of secondary school. Their responses are shown in the tally sheet below.

<i>Plans for year after secondary school</i>	<i>Tally</i>
Tertiary Studies	
TAFE	
Apprenticeship	
Work	
Travel	
Other	

- (a) How many students are in year 10 at the school?  
 (b) Construct a column graph displaying this data.  
 (c) If a pie chart was constructed to display this data, what would be the angle for each of the activities?  
 Give answers correct to one decimal place)

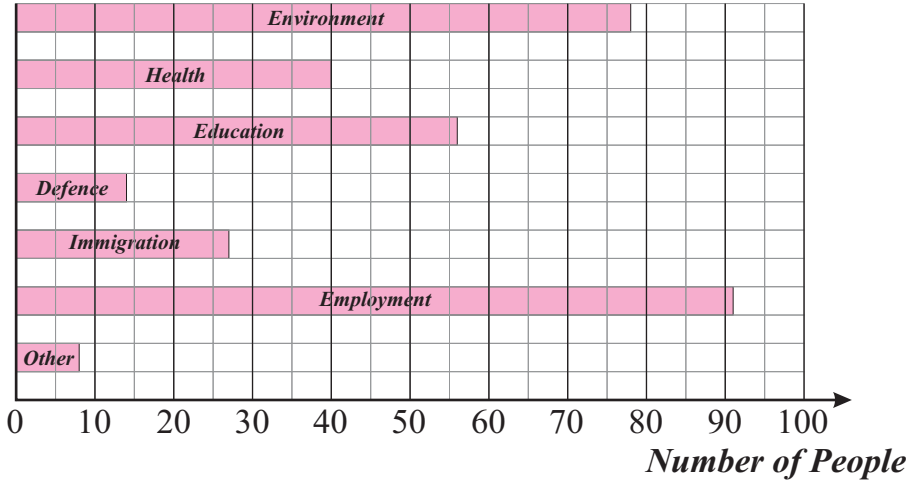
3. This pictograph shows the results of a survey conducted to find the activity preferred by a number of people to keep fit.

Running	
Swimming	
Walking	
Dancing	
Gym	
Bicycle riding	
Team Sport	
Other	

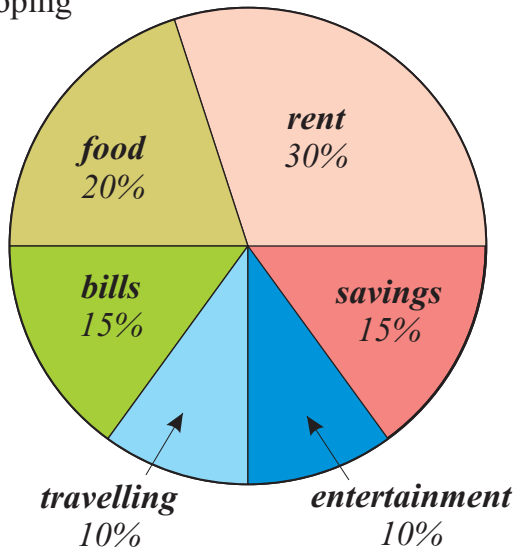
 represents 10 people

- (a) How many people were surveyed?  
 (b) What percentage of the people surveyed participated in each of the activities?  
 Give answers to correct to one decimal place.  
 (c) Based on these figures, how many people in a town of 6000 would choose swimming?

4. The Federal Government conducted a survey to find out which portfolio people considered to be the most important. The results are shown on the bar graph below.



- (a) Based on these figures, what percentage of the people selected each of the portfolios as the most important?  
Give answers correct to one decimal place.
- (b) If this data was to be displayed on a pie chart, what would be the angle for each portfolio?  
Give answers correct to one decimal place.
5. The pie chart shown here represents the percentage of weekly income that Louis and Sam were hoping to spend on their weekly expenses. Their combined nett income each week is \$1300. How much should they spend on each of their weekly expenses?

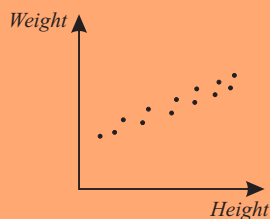


## Bivariate Data

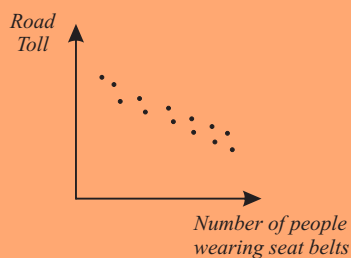
The shape of a bivariate data graph gives an indication of the relationship between the two variables.

***Bivariate means two variables.***

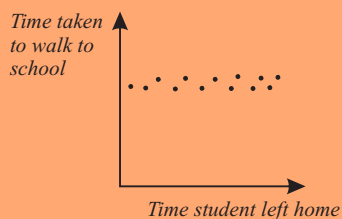
### ***Examples***



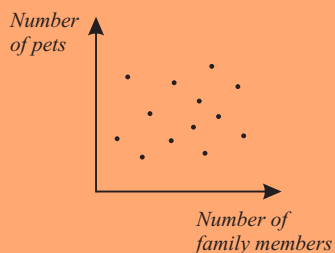
This graph indicates that as the height ***increases*** the weight also ***increases***.



This graph indicates that as the number of people wearing seat belts ***increases*** the road toll ***decreases***.



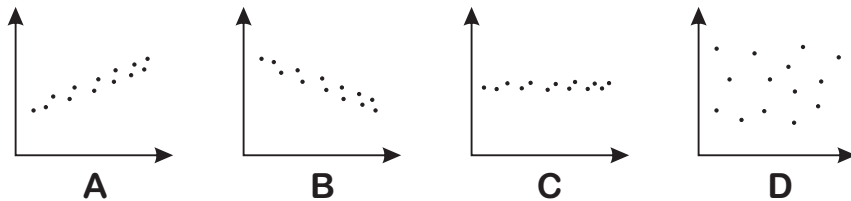
This graph indicates that the time that this student leaves home does not affect the time taken to walk to school. The time taken to walk to school is always about the same.



This graph indicates that the number of pets in a family is not affected by the number of family members.

## EXERCISE 16E

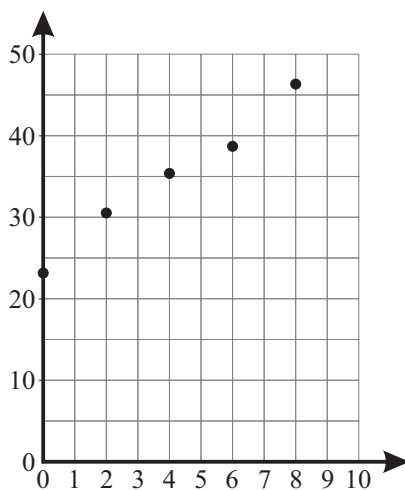
1. Select which of these four graphs would best illustrate the shape of the graph resulting from plotting data obtained from each of the bivariate studies below.



- The hours spent on homework versus the results achieved on tests.
- The number of police patrol cars versus the number of car accidents.
- The age of a person versus their body temperature.
- The price of a item versus the number sold.
- The time spent watching television versus exam results.
- The time taken to run 100 metres versus mark on mathematics exam.
- Age of a car versus its value.
- The number of spectators attending a football game versus the price of petrol.
- The time taken for a particular student to walk to school versus the daily temperature.

2. The *monthly profit* of a company versus the *monthly advertising costs* is shown on this graph.
- From this graph does increasing advertising costs *increase* or *decrease* the profit.
  - Based on this graph, what would be the approximate expected profit if the following amounts were spent on advertising?
    - \$5000
    - \$10 000

*Monthly Profit*  
( $\times \$1000$ )



*Monthly Advertising Costs* ( $\times \$1000$ )

3. The *Sticky B* honey company found that changing the temperature of the honey in a vat changed the time it took to fill a jar. The data below shows the time to fill a honey jar from a vat for different temperatures of the honey.

<b>Temperature (°C)</b>	6	8	10	12	14	16	18	20	22
<b>Time to fill jar (sec)</b>	9.8	9.2	8.2	7.5	7.2	6.2	5.6	4.7	4.1

- (a) Draw a graph displaying this information using scales of **Temperature** (0°C - 26°C) on the horizontal axis and **Time to fill jar** (0 - 12 seconds) on the vertical axis.

- (b) Does increasing the temperature **increase** or **decrease** the time taken to fill a honey jar?

- (c) Based on this information, approximately how long would it take to fill a honey jar at the following temperatures? A straight line of best fit through the points will help.

(i) 15°C (ii) 25°C (iii) 3°C



4. The amount of fuel needed by a ferry to travel across a bay was dependant on the number of cars it was carrying. The volume of fuel used (litres) and the number of cars on board were recorded on eight days. The results are shown below. The weather conditions were similar on the eight days.

<b>Number of cars</b>	5	8	6	2	12	7	10	9
<b>Volume of fuel (L)</b>	38.1	42.4	39.6	33.0	50.2	41.6	46.3	45.9

- (a) Draw a graph displaying this information using scales of **Number of cars** (0 - 15) on the horizontal axis and **Volume of fuel** (30 - 56 L) on the vertical axis.

- (b) Based on this information, find the approximate amount of fuel used with the following number of cars on board.

A straight line of best fit through the points will help.

(i) no cars (ii) 3 cars (iii) 15 cars

## Mean, Mode and Median - Discrete Ungrouped Data

The *mean*, *mode* and *median* are statistical measures used to summarise data and give a measure of the *central tendency* of a set of numerical data.

The *mean* of a set of ungrouped values can be found by adding all the values and dividing by the total number of values.

$\bar{x}$  is the mathematical symbol for the mean of the  $x$  values

$\Sigma$  is the Greek letter sigma which is used in mathematics to represent *the sum of*. Therefore:

$\Sigma x$  represents the sum of the  $x$  values

$N$  represents the total number of  $x$  values

$$\text{mean} = \bar{x} = \frac{\text{sum of all the values}}{\text{number of values}}$$

$$\bar{x} = \frac{\Sigma x}{N}$$

*mode* = the value that occurs the most (there may be no mode, one mode or more than one mode)

*median* = the middle value when the values are arranged in numerical order (if there is an even number of values the median is the average of the two middle values)

**Example 1** Find the mean, mode and median of the following set of values:

6 8 8 15 15 15 19 29 38

$$\text{mean} = \bar{x} = \frac{\Sigma x}{N} = \frac{6+8+8+15+15+15+19+29+38}{9} = 17$$

*mode* = 15 (the value that occurs the most)

*median* = 15 (middle value: 6 8 8 15 (15) 15 19 29 38)

**Example 2** Find the mean, mode and median of the following set of values:

**8 24 19 11 3 37 15**

$$\text{mean} = \bar{x} = \frac{\Sigma x}{N} = \frac{8+24+19+11+3+37+15}{7} = 16.7 \text{ (one dec. pl.)}$$

*There is no mode (only one of each value)*

*To find the median the values need to be arranged in numerical order:*

3 8 11 15 19 24 37

**median = 15** (middle value: 3 8 11 (15) 19 24 37)

**Example 3** Find the mean, mode and median of the following set of values:

**5 17 11 16 17 16 21 14 14 19 16 14 8 11**

$$\begin{aligned} \text{mean} &= \bar{x} = \frac{\Sigma x}{N} \\ &= \frac{5+17+11+16+17+16+21+14+14+19+16+14+8+11}{14} \\ &= 14.2 \text{ (one decimal place)} \end{aligned}$$

Arranging the values in numerical order will make it easier to find the mode and the median.

5 8 11 11 14 14 14 16 16 16 17 17 21 19

**mode = 14 and 16**

*There is an even number of values therefore the median is the average of the middle two values.*

5 8 11 11 14 14 (14) (16) 16 16 17 17 21 19

$$\text{median} = 15 \left( \frac{14+16}{2} \right)$$



### EXERCISE 16F

1. For each of the following sets of values find the:

(i) mean (ii) mode (iii) median

Give answers correct to one decimal place where appropriate.

(a) 6 9 10 10 12 14 17 19 29

(b) 7 8 12 16 23 28

(c) 25 28 33 37

(d) 28 17 21 15 14 32 21

(e) 8 10 9 9 11 12 11 13 9 10 14 10

(f) 24 36 18 44 56 21 38 58

2. The ages of 11 people at a library are shown below.

28, 38, 43, 61, 55, 29, 11, 39, 50, 62, 44

(a) What is the mean age of the people in the library?

Give answer correct to one decimal place.

(b) What is the median age of the people in the library?

3. This table shows the scores *for* and *against* a football team for the first eight rounds of a competition.

Find the mean and median scores for and against the team for these eight rounds.

Give answers correct to one decimal place.

Round	score for	score against
1	128	113
2	97	86
3	133	90
4	102	122
5	87	75
6	73	98
7	88	87
8	110	62

4. The prices of six houses sold in a suburb are shown below.

\$200 000, \$156 000, \$278 000, \$223 000, \$1 504 000, \$189 000

(a) Find the mean and median price of the houses sold.

(b) Which of these values give a better indication of the value of the houses in this suburb? Why?

## Mean, Mode and Median - Discrete Grouped Data

When there is a large amount of data with similar values it is convenient to group the data and display it in a frequency table.

### Example 1

15 16 15 14 18 15 16 14 15 13 17 18 13  
14 13 14 16 18 17 16 14 15 14 13 14 15

To group the data arrange it in numerical order or construct a tally sheet.

13 13 13 13 14 14 14 14 14 14 14 15 15  
15 15 15 15 16 16 16 16 17 17 18 18 18

Value	Tally
13	
14	
15	
16	
17	
18	

Value $x$	Frequency $f$
13	4
14	7
15	6
16	4
17	2
18	3
<b>total (N)</b>	<b>26</b>

A frequency table can then be drawn.

The frequency is how many times each value occurs.

$$\begin{aligned} \text{mean} &= \bar{x} = \frac{\Sigma(x \times f)}{N} \\ &= \frac{(13 \times 4) + (14 \times 7) + (15 \times 6) + (16 \times 4) + (17 \times 2) + (18 \times 3)}{26} \end{aligned}$$

$$\text{mean} = 15.1 \text{ (one decimal place)}$$

$$\text{mode} = 14$$

$$\text{median} = 15 \text{ (the two middle numbers are 15 and 15)}$$

**Example 2**

Find the mean, mode and median of the data represented in this frequency table.

<i>Value</i> <i>x</i>	<i>Frequency</i> <i>f</i>
25	6
26	8
27	10
28	10
29	7

**Note:**  $N = \Sigma f = 41$

$$\begin{aligned} \text{mean} = \bar{x} &= \frac{\Sigma(x \times f)}{N} \\ &= \frac{(25 \times 6) + (26 \times 8) + (27 \times 10) + (28 \times 10) + (29 \times 7)}{41} \end{aligned}$$

**mean** = 27.1 (one decimal place)

**mode** = 27 and 28

**median** = 27 (there are 41 values therefore the median will be the 21<sup>st</sup> value)

$$\text{median} = \text{value number} \left( \frac{N+1}{2} \right)$$

To find the median in a frequency table use the 'running total' of the frequency called the **cumulative frequency**.

<i>Value</i> <i>x</i>	<i>Frequency</i> <i>f</i>	<i>Cumulative</i> <i>Frequency</i>
25	6	6
26	8	14
27	10	24
28	10	34
29	7	41

← 21<sup>st</sup> value  
in this group

**EXERCISE 16G**

- Construct a frequency table for the following sets of data.
  - 8 8 9 9 9 9 10 10 10 11 11 11 11 12 12 12 13 14
  - 23 24 21 20 21 22 24 25 21 23 20 20 22 24 23 23 22
  - 18 16 17 19 20 16 17 18 16 15 16 17 18 20 17 17 16  
18 17 15 16 20 18 17 15 17 19 20 19 17 16 18 15 20
- Find the mean, mode and median for the sets the data in question 1. Give answers to one decimal place where appropriate.
- Jayne plays for the Lizards futsal team. She kicked the following number of goals in each of the games in one season.

**3, 6, 4, 2, 3, 4, 5, 1, 2, 4, 3, 3, 3, 2, 5, 7, 2, 3, 1, 0, 5, 2, 4, 3**

- Construct a frequency table for this information.
- Calculate the mean number of goals kicked by Jayme each game. Give answer to one decimal place.
- Find the median number of goals kicked by Jayme each game.



- Tamazyn is the quality controller at a biscuit company. She wanted to check how many biscuits were in the packets of *Fruity Bites*. The company states there is an average of at least 35 biscuits in each packet. Tamazyn counted the number of biscuits in 40 packets and the data is given below.

**36 35 34 37 35 34 38 33 35 36 37 35 34 36 37 38 32 35 37 36  
35 36 37 38 33 34 35 36 36 37 37 36 35 34 35 36 33 34 35 36**

- Construct a frequency table to display this information.
- Calculate the mean number of biscuits per packet. Give answer correct to one decimal place.
- Based on this data, is the company's claim correct?

## Mean, Mode and Median - Continuous Data

*Continuous numerical data* is often grouped in tables such as the one shown here representing the height of a number of people.

The class interval of 140-<150 represents all heights between 140 cm and 150 cm, including 140 cm but less than 150 cm.

*Height of people*

<i>Class Interval (cm)</i>	<i>Frequency f</i>
140-<150	8
150-<160	14
160-<170	27
170-<180	35
180-<190	25

The *mean*,  $\bar{x}$ , can be calculated by using the midpoints of the class intervals as the  $x$  values.

*Height of people*

<i>Class Interval (cm)</i>	<i>Midpoint x</i>	<i>Frequency f</i>
140-<150	145	8
150-<160	155	14
160-<170	165	27
170-<180	175	35
180-<190	185	25
<i>total (N)</i>		109

$$\begin{aligned} \text{mean} = \bar{x} &= \frac{\Sigma(x \times f)}{N} \\ &= \frac{(145 \times 8) + (155 \times 14) + (165 \times 27) + (175 \times 35) + (185 \times 25)}{109} \end{aligned}$$

**mean**    **170.0 cm** (one decimal place)

**EXERCISE 16H**

1. Calculate the mean of the following sets of data.  
Give answers correct to one decimal place.

(a)

*Mass*

<i>Class Interval (kg)</i>	<i>Frequency <math>f</math></i>
50-<60	13
60-<70	19
70-<80	33
80-<90	28
90-<100	17

(b)

*Length*

<i>Class Interval (cm)</i>	<i>Frequency <math>f</math></i>
40-<60	20
60-<80	33
80-<100	57
100-<120	44
120-<140	23

(c)

*Height*

<i>Class Interval (cm)</i>	<i>Frequency <math>f</math></i>
0-<100	22
100-<200	57
200-<300	143
300-<400	127
400-<500	98

(d)

*Age*

<i>Class Interval (years)</i>	<i>Frequency <math>f</math></i>
0-<5	43
5-<10	50
10-<15	59
15-<20	68
20-<25	30

2. A number of seals were caught, tagged and released. Their masses, in kg, are shown below.

27 38 16 42 57 9 35 47 31 29 42 55 28 30 45 19 61 58  
29 37 33 41 18 60 34 23 19 8 46 59 62 41 27 37 26 32

- (a) Construct a frequency table to display this information using class intervals 0-<10, 10-<20, 20-<30, etc.  
(b) Use the frequency table to calculate the mean mass of the seals.  
Give answer correct to one decimal place.  
(c) Calculate the mean mass of the seals using the individual masses of the seals. Give answer correct to one decimal place.

The **median** of continuous numerical data can be found by constructing a **cumulative frequency graph**. This can be done after completing the **cumulative frequency** ('running total') column in the frequency table. See example below.

### Height of people

Class Interval (cm)	Frequency $f$	Cumulative Frequency
140-<150	8	8
150-<160	14	22
160-<170	27	49
170-<180	35	84
180-<190	25	109

The median is the item number  $(\frac{N+1}{2})$ .

In this example  $N = 109$  therefore the median will be item number 55 ( $\frac{109+1}{2}$ ).

To find the median follow these steps:

**Step 1** - Draw the cumulative frequency graph.

**Step 2** - On the cumulative frequency axis locate  $(\frac{N+1}{2})$ , 55 in this example.

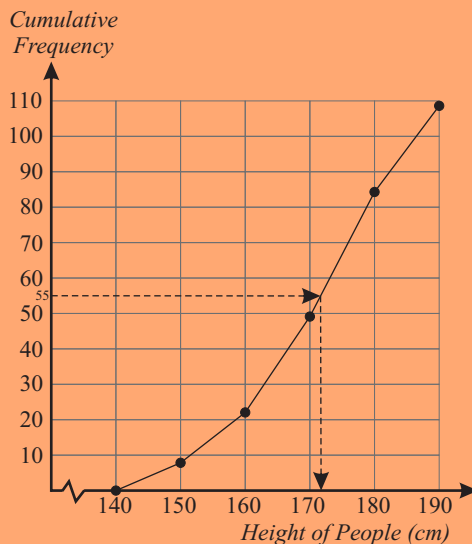
**Step 3** - Draw a horizontal line across to the curve.

**Step 4** - Draw a vertical line from this point to the data axis. This point will be the approximate value of the median.

In this example:

**median  $\approx 172$  cm**

### Cumulative Frequency Graph



## EXERCISE 16I

1. For each of the following sets of data draw a cumulative frequency graph and hence find an approximate value for the median.

(a) *Mass*

<i>Class Interval (kg)</i>	<i>Frequency <math>f</math></i>
50-<60	5
60-<70	12
70-<80	18
80-<90	13
90-<100	7

(b) *Length*

<i>Class Interval (cm)</i>	<i>Frequency <math>f</math></i>
40-<60	12
60-<80	23
80-<100	28
100-<120	29
120-<140	20

(c) *Height*

<i>Class Interval (cm)</i>	<i>Frequency <math>f</math></i>
0-<100	11
100-<200	26
200-<300	39
300-<400	19
400-<500	7

(d) *Age*

<i>Class Interval (years)</i>	<i>Frequency <math>f</math></i>
0-<5	22
5-<10	37
10-<15	47
15-<20	42
20-<25	29

2. The ages of a number of people who attended a festival are shown below.

27 21 38 45 16 28 42 10 57 51 9 27 35 47 31 29 42 55  
28 30 45 19 61 58 29 22 9 37 41 33 41 18 60 34 23 19  
8 46 35 59 19 62 16 41 27 27 37 52 26 50 32 29 15 48

- Construct a frequency table to display this information using class intervals 0-<10, 10-<20, 20-<30, etc.
- Use the frequency table to draw a cumulative frequency graph.
- Use this graph to find the approximate value for the median age.
- Arrange the ages in numerical order and find the exact value of the median.



The mean, mode (modal class) and median can be found from a frequency table of continuous numerical data by including the following columns to the table:

*midpoint (x), (x × f), cumulative frequency*

**Example**

**Length**

<b>Class Interval (cm)</b>	<b>Frequency f</b>
0-<20	8
20-<40	15
40-<60	23
60-<80	19
80-<100	11

Include *midpoint (x), (x × f), cumulative frequency* columns.

**Length**

<b>Class Interval (cm)</b>	<b>Midpoint x</b>	<b>Frequency f</b>	<b><math>x \times f</math></b>	<b>Cumulative Frequency</b>
0-<20	10	8	80	8
20-<40	30	15	450	23
40-<60	50	23	1150	46
60-<80	70	19	1330	65
80-<100	90	11	990	76
			4000	

$\Sigma(x \times f)$

$N$

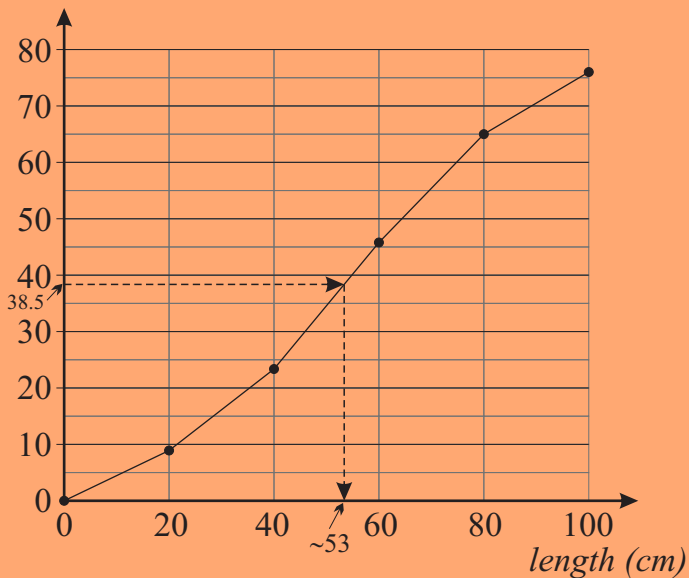
$$\begin{aligned} \text{mean} &= \frac{\Sigma(x \times f)}{N} \\ &= \frac{4000}{76} \end{aligned}$$

$$\text{mean} = 52.6 \text{ cm (one decimal place)}$$

The median can be found by drawing the cumulative frequency curve. Find  $(\frac{N+1}{2})$  on the cumulative frequency axis.

$$\frac{N+1}{2} = \frac{76+1}{2} = \frac{77}{2} = 38.5$$

*cumulative  
frequency*



***median*  $\approx$  53 cm**

The ***modal class*** is the class that has the highest frequency = **40 cm - <60 cm**

***Answers:***

***mean* = 52.6 cm (one decimal place)**

***modal class* = 40 cm - <60 cm**

***median*  $\approx$  53 cm**

## EXERCISE 16J

1. For the following sets of data:

- redraw the frequency table including *midpoint*,  $(x \times f)$  and *cumulative frequency* columns.
- calculate the mean (correct to one decimal place)
- find the modal class
- draw a cumulative frequency curve
- find an approximate value of the median

(a) *Height*

<i>Class Interval (m)</i>	<i>Frequency f</i>
0-<10	12
10-<20	23
20-<30	45
30-<40	32
40-<50	19

(b) *Time*

<i>Class Interval (min)</i>	<i>Frequency f</i>
30-<35	15
35-<40	23
40-<45	28
45-<50	33
50-<55	27

2. The times taken (in seconds) for a number of students to solve a puzzle are shown below.

35 78 48 83 51 50 42 66 72 61 58 67 62 80 71 56  
 64 73 68 60 59 44 69 52 38 47 56 63 71 82 65 47  
 56 37 80 77 63 57 43 48 58 62 59 66 73 47 64 61

- Using class intervals of 30-<40, 40-<50, etc construct a frequency table that includes *class interval*, *midpoint*, *frequency*,  $(x \times f)$  and *cumulative frequency*. A tally sheet may assist in finding the frequency in each class interval.
- Calculate the mean from this frequency table (one decimal place).
- Find the modal class.
- Draw a cumulative frequency graph.
- Find the approximate value for the median from this graph.
- Use the raw data to calculate the mean and median.
- See how long it takes you to solve the puzzle:

*Tom, Tim and Tam are brothers. Tom, the eldest, is ten years older than Tim, the youngest. The mean of their ages is 10 and the median age is 8. Find the ages of the brothers.*

## Stemplots (Stem-and-Leaf Plots)

### EXERCISE 16K

#### Example

1. Construct a non-ordered stemplot for the numbers below.
2. Construct an ordered stemplot for the numbers below.
3. Find the median.

24, 35, 26, 48, 51, 17, 34, 26, 28, 40, 20, 19, 22, 37, 41, 49, 51, 53

#### 1. non-ordered stemplot

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

$$\overline{1|3} = 13$$

#### 2. ordered stemplot

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

$$\overline{1|3} = 13$$

3. The *median* is item number  $(\frac{N+1}{2})$ .

$N = 18$  so the median will be item number 9.5 - average of items 9 and 10 (34 and 35).

$$\text{median} = 34.5$$

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

$$\overline{1|3} = 13$$

1. For the following sets of data:

- (i) construct a non-ordered stemplot
- (ii) construct an ordered stemplot
- (iii) find the median

(a) 23, 36, 67, 35, 37, 49, 51, 28, 33, 36, 40, 38, 57, 52, 29, 37

(b) 8, 16, 7, 32, 25, 26, 19, 18, 16, 18, 25, 33, 42, 19, 41, 39, 30

2. Two athletes each threw a javelin 10 times and the distances, to the nearest metre, are shown below.

**Spike: 36, 44, 52, 46, 36, 28, 33, 37, 40, 48**  
**Lance: 36, 35, 59, 33, 29, 32, 34, 37, 41, 36**

For these athlete's results complete the following.

- Construct non-ordered stemplots.
- Construct ordered stemplots.
- Find the furthest distance each athlete threw the javelin.
- Find the median of each set of throws.
- Find the mean distance thrown by each athlete.



3. Jacqui, a forestry researcher, wanted to investigate how the population of honeyeaters was recovering in a national park a year after a bush fire. The national park was 5000 hectares. Each day she covered an area of 10 hectares counting the number of honeyeaters. She did this for twenty-four days and the number of honeyeaters she counted each day are shown below.

**58, 36, 27, 98, 26, 38, 50, 42, 62, 36, 48, 50,  
25, 16, 30, 42, 38, 55, 46, 61, 45, 74, 39, 40**

- Construct a non-ordered stemplot.
- Construct an ordered stemplot.
- What is the median number of honeyeaters in a 10 hectare area?
- What is the mean number of honeyeaters in a 10 hectare area?

Give answer correct to one decimal place.

- Based on these figures, calculate the number of honeyeaters in the national park using:
  - the median
  - the mean (nearest whole number)



## Boxplots (Box-and-Whisker Plots)

**Examples** For each of the following sets of data:

- (i) find the *interquartile range*.
- (ii) construct a *boxplot*.

1.

19, 18, 12, 28, 26, 13, 15, 24

**Step 1** Arrange in numerical order from smallest to largest.

12 13 15 18 19 24 26 28

**Step 2** Find the median.

12 13 15 18 19 24 26 28

↖ Median = 18.5

**Step 3** Find the median of the bottom half of the numbers.

This is called the *lower quartile* ( $Q_1$ ).

12 13 15 18 19 24 26 28

$Q_1 = 14$

**Step 4** Find the median of the top half of the numbers.

This is called the *upper quartile* ( $Q_3$ ).

12 13 15 18 19 24 26 28

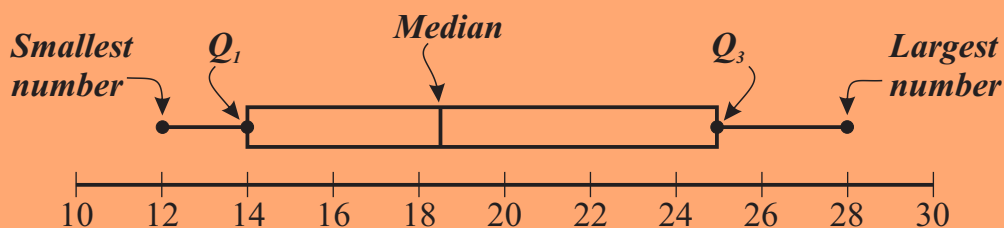
↖  $Q_3 = 25$

**Step 5** *Interquartile range* =  $Q_3 - Q_1$

$$= 25 - 14$$

$$= 11$$

**Step 6** Draw the boxplot.



2.

40, 33, 27, 25, 34, 29, 35, 31, 37, 23, 38

**Step 1** 23 25 27 29 31 33 34 35 37 38 40

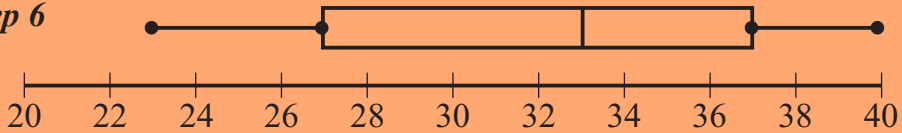
**Step 2** 23 25 27 29 31 (33) 34 35 37 38 40 **Median = 33**

**Step 3** 23 25 (27) 29 31 (33) 34 35 37 38 40 **Q<sub>1</sub> = 27**

**Step 4** 23 25 (27) 29 31 (33) 34 35 (37) 38 40 **Q<sub>3</sub> = 37**

**Step 5** **Interquartile range = 37 - 27 = 10**

**Step 6**



### EXERCISE 16L

1. For each of the following sets of data:

(i) find the interquartile range.

(ii) construct a boxplot.

(a) 9, 12, 13, 10, 8, 17, 15

(b) 17, 21, 28, 37, 40, 13, 42, 33, 24, 15

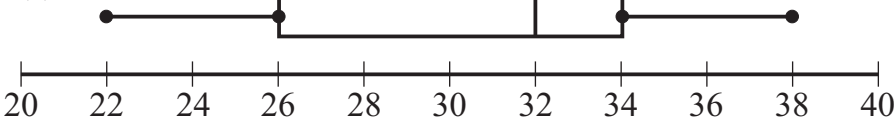
(c) 40, 38, 47, 51, 55, 43, 36, 40, 37

(d) 12, 19, 18, 28, 29, 30, 27, 12

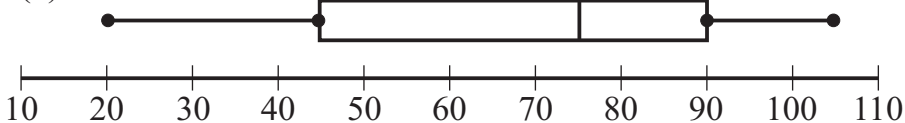
2. For the following boxplots find the:

(i) median (ii) interquartile range

(a)



(b)



## Outliers

As described earlier, the *mean*, *mode* and *median* are statistical measures that are used to summarise data and give a measure of the '*centre*' of a set of data.

Sometimes the *mean* can be affected disproportionately by items of data that are considerably larger or smaller than the rest of the data. These items of data are called *outliers*.

Outliers must be included when calculating the mean, mode and median. They will not affect the mode or median but could distort the mean of a set of data.

### Examples

Calculate the mean, mode and median of the following sets of data.

1. 5, 56, 58, 60, 61, 61, 61, 62, 62

*mean* = 54, *mode* = 61, *median* = 61

2. 25, 27, 27, 28, 28, 29, 29, 30, 32, 157

*mean* = 41.2, *mode* = 27, 28, 29, *median* = 28.5

In both these examples it can be seen that the mean has been affected by the extremely low or high values.

When analysing data, outliers need to be recognised as factors that could distort the summary of the data particularly the mean. Sometimes statistics are deliberately distorted by outliers.

### Example

Six houses sell in a suburb. The prices are:

**\$200 000, \$178 000, \$210 000, \$951 000, \$206 000, \$235 000**

The *median* price = **\$208 000**

The *mean* price = **\$330 000**

The price of \$951 000 has shifted the mean house price to well above the price of the other five houses. This may have been an unusual house not representative of the area. The median price is probably a better indication of the house prices in this suburb. A real estate agent in the area could use the mean price as a way of increasing price expectations.



An outlier is defined as any item of numerical data that is:

- more than  $1.5 \times$  the interquartile range (IQR) above the upper quartile value ( $Q_3$ ) or
- more than  $1.5 \times$  the interquartile range (IQR) below the lower quartile value ( $Q_1$ )

Outliers are shown as dots on a boxplot but not included in the 'whiskers'.

### ***Example***

Draw a boxplot of the following data indicating all outliers as dots.

**45, 21, 33, 78, 53, 36, 29, 3, 31, 40**

**Step 1** Arrange data in ascending numerical order.

3, 21, 29, 31, 33, 36, 40, 45, 53, 78

**Step 2** Find the median

3, 21, 29, 31, 33,  $\uparrow$  36, 40, 45, 53, 78

**median** = 34.5

**Step 3** Find  $Q_1$  and  $Q_3$

3, 21, (29), 31, 33,  $\uparrow$  36, 40, (45), 53, 78  
 $Q_1$  **median** = 34.5  $Q_3$

**Step 4** Find the interquartile range (IQR)

$$\text{IQR} = Q_3 - Q_1 = 45 - 29 = 16$$

**Step 5** Find  $1.5 \times \text{IQR} = 1.5 \times 16 = 24$

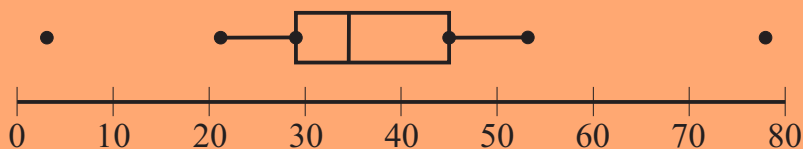
**Step 6** Find  $Q_1 - (1.5 \times \text{IQR})$  and  $Q_3 + (1.5 \times \text{IQR})$

$$\begin{array}{lcl} Q_1 - 24 = 29 - 24 & & Q_3 + 24 = 45 + 24 \\ = 5 & & = 69 \end{array}$$

Therefore any values less than 5 or greater than 69 are classified as outliers.

(3), 21, 29, 31, 33, 36, 40, 45, 53, (78)  
 $\longleftarrow$  **outliers**  $\longrightarrow$

**Step 7** Draw the boxplot



### EXERCISE 16M

- Construct boxplots for the following sets of data, clearly showing any outliers.
  - 29, 15, 25, 48, 69, 95, 46, 13, 18, 28, 33
  - 50, 43, 53, 57, 63, 48, 11, 91, 37
  - 55, 52, 35, 39, 40, 10, 71, 53
- The scores made by two batters in cricket for 10 matches are shown below.

**Edgey Ed:** 10, 6, 0, 21, 389, 3, 8, 0, 79, 0

**Driver Dan:** 45, 36, 58, 55, 87, 22, 48, 55, 40, 60

- Find the median score by each batter.
- Find the mean score by each batter.
- Who is the more consistent batter?
- Does the median or the mean give a better indication of who is the more consistent batter?
- Find the mean of each batter's scores if the top score for each is ignored (correct to one decimal place).



- The time taken for a tram to complete a city route was recorded each day for 10 week days. These times (to the nearest minute) are shown here.

43, 39, 41, 44, 40, 67, 42, 43, 39, 40

- Find the median time.
- Calculate the mean time.
- Calculate the mean time if outliers are ignored (correct to one decimal place).

## Normal Distribution and Standard Deviation

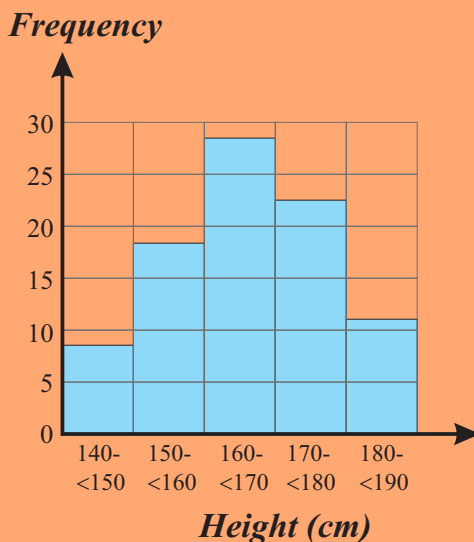
If the frequency of numerical data is displayed on a graph with the columns joined this is called a histogram.

### Example

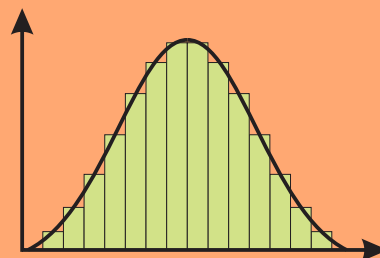
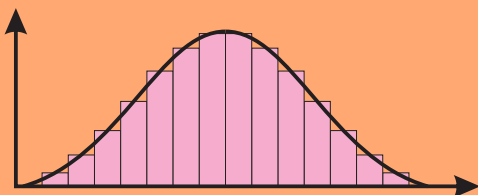
The height of a number of students is shown in the frequency table and frequency graph (histogram) below.

*Height of students*

<i>Height Interval (cm)</i>	<i>Frequency <math>f</math></i>
140-<150	8
150-<160	19
160-<170	28
170-<180	22
180-<190	11



The shape of this histogram is very similar to the shape of histograms formed by displaying the frequency of many types of data: height, weight, growth rates, times taken, heart rates, etc. The more data that is collected, in many cases, the closer the histogram will look like the graphs shown here.



This curve is normal for many types of 'real life' data and is called a ***Normal Distribution***.

The curve is also called a ***Bell Curve*** because of its shape.

The 'perfect' Bell Curve is symmetrical.

The centre of the perfect curve is the mean and median and more data is in this section of the graph. The number of items of data decreases the further they are from the centre.

This is to be expected. For example the height of people. There is an average (mean) height and there less people the further you move from this mean.

One measure of the spread of data in a Normal Distribution curve is the ***Standard Deviation***. It is a measurement that indicates the deviation from the standard measurement (mean).

The symbol for standard deviation is  $\sigma$  (the Greek letter sigma).

We will study how to calculate standard deviation later but in terms of a normal distribution graph the standard deviation is a measure of the spread of the data from the mean.

***68% of the data will lie between one standard deviation either side of the mean.***

***95% of the data will lie between two standard deviations either side of the mean.***

***99.7% of the data will lie between three standard deviations either side of the mean.***

Remember the symbol for mean is  $\bar{x}$ .

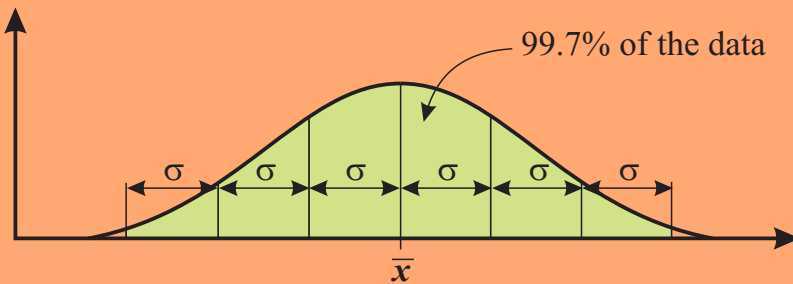
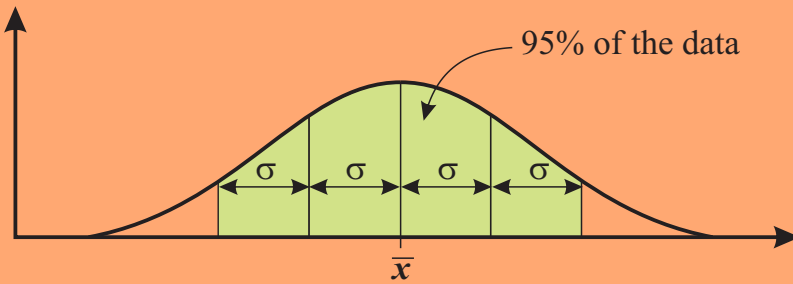
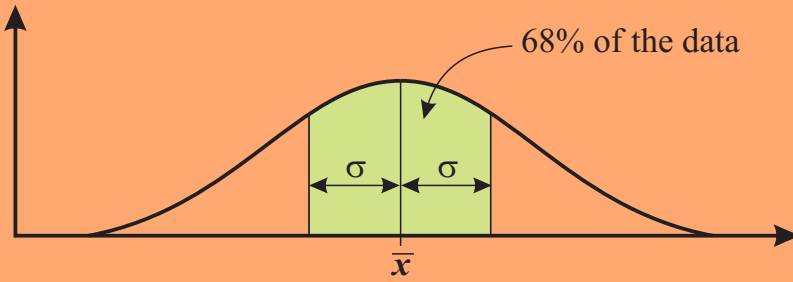
Using symbols:

$\bar{x} \pm \sigma$  represents 68% of the data.

$\bar{x} \pm 2\sigma$  represents 95% of the data.

$\bar{x} \pm 3\sigma$  represents 99.7% of the data.

Looking at the normal distribution graph:



**Example 1**

A number of people donated money to a charity.  
The mean amount donated was \$30 per person and the standard deviation of the amounts donated was \$5 per person.

This means that:

68% of the people donated between \$25 and \$35 (one standard deviation either side of the mean)

95% of the people donated between \$20 and \$40 (two standard deviations either side of the mean)

99.7% of the people donated between \$15 and \$45 (three standard deviations either side of the mean)

**Example 2**

A group of students were given a spelling test of 45 words.

The mean number of words spelt correctly was 32.

68% of the students spelt between 28 and 36 words correctly.

1. Calculate the standard deviation of the results.
2. Complete the following sentence.

*95% of the students correctly spelt between  and  words.*

**Answers**

1. 68% of the students is one standard deviation either side of the mean.  
The mean is 32 which is midway between 28 and 36.  
The standard deviation is therefore **4**.
2. 95% of the students is two standard deviations either side of the mean. Two standard deviations = 8.

$$\begin{aligned}\bar{x} \pm 2\sigma \\ &= 32 \pm 8 \\ &= 24, 40\end{aligned}$$

*95% of the students correctly spelt between 24 and 40 words.*

**Example 3**

Sunny and Manger are mango farmers. They each harvest 1000 mango trees.

The average (mean) yield of mangoes per tree for each farmer was 80 kg.

The standard deviation for Sunny's yield was 10 kg.

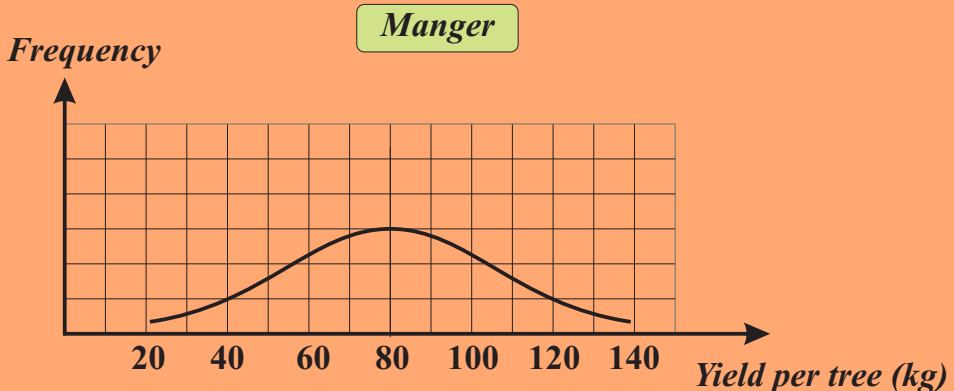
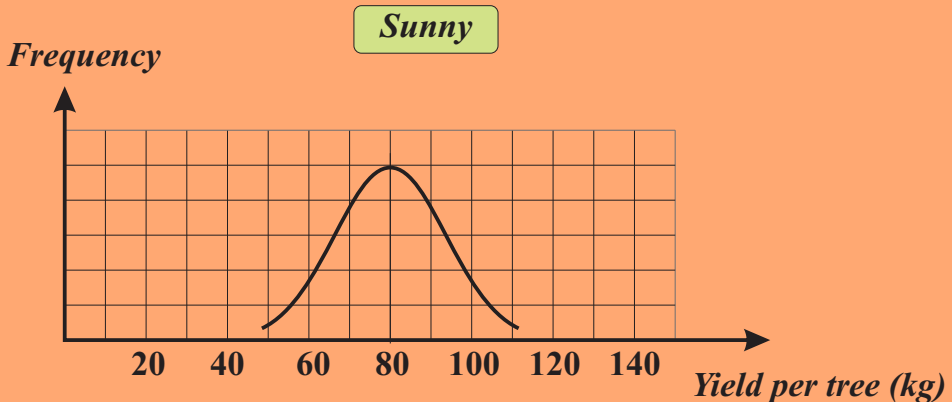
The standard deviation for Manger's yield was 20 kg.

Assume the data is a normal distribution.

1. Draw a curve showing the shape of the frequency distribution histogram for each, showing the main differences.
2. What effect does the size of the standard deviation have on the frequency distribution histogram and normal distribution curve.
3. Describe the differences between Sunny's and Manger's mango trees.

**Answers**

1.



2. A larger standard deviation means a larger spread (variation) of the data.

For *Sunny's farm* the mean is 80 and standard deviation is 10. This means 68% of the trees will be between 70 and 90 (one standard deviation either side of the mean).

95% of the trees will be between 60 and 100 ( $\bar{x} \pm 2\sigma$ ).

99.7% of the trees will be between 50 and 110 ( $\bar{x} \pm 3\sigma$ ).

For *Manger's farm* the mean is 80 and standard deviation is 20.

68% of the trees will be between 60 and 100 ( $\bar{x} \pm \sigma$ ).

95% of the trees will be between 40 and 120 ( $\bar{x} \pm 2\sigma$ ).

99.7% of the trees will be between 20 and 140 ( $\bar{x} \pm 3\sigma$ ).

In this example because the overall number of trees is the same for both farmers the number of trees around the mean on Manger's farm will be less because there are more 'further away' from the mean.

3. There is a larger variation in the yield of Manger's mangoes. Sunny's farm has more trees that produce around the average (mean) yield.



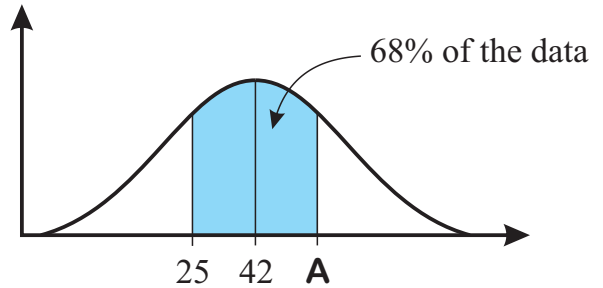


*Assume all the questions in this exercise follow a normal distribution curve.*

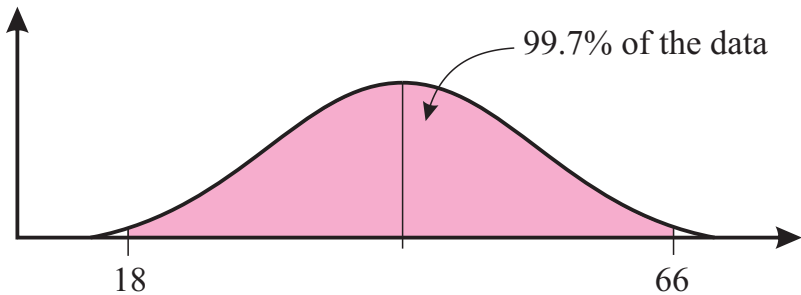
### EXERCISE 16N

- The *mean* of a set of data is 50 and the *standard deviation* is 6. Complete the following sentences.
  - 68% of the data will occur between  and .
  - 95% of the data will occur between  and .
  - 99.7% of the data will occur between  and .
- The central 68% of a set of data lies between 72 and 88.
    - What is the standard deviation of this set of data?
    - What is the mean of this set of data?
  - The central 95% of a set of data lies between 96 and 156.
    - What is the standard deviation of this set of data?
    - What is the mean of this set of data?
  - The central 99.7% of a set of data lies between 45 and 117.
    - What is the standard deviation of this set of data?
    - What is the mean of this set of data?
  - The central 68% of a set of data lies between 17.2 and 34.6.
    - What is the standard deviation of this set of data?
    - What is the mean of this set of data?
- The mean of a set of data is 84 and 34% of the data lies between 68 and 84. Complete the following sentences.
  - 68% of the data will occur between  and .
  - 95% of the data will occur between  and .
  - 99.7% of the data will occur between  and .
- The mean of a set of data is 56 and the standard deviation is 12.
  - What percentage of the data is greater than 68?
  - What percentage of the data is less than 32?
  - What percentage of the data is greater than 44?
  - What percentage of the data is less than 92?

5. For the data shown on this graph find:  
 (a) the standard deviation.  
 (b) **A**.



6. For the data shown on the graph below find:  
 (a) the standard deviation  
 (b) the mean



7. Scientists measured and recorded the age of a number of koalas in a state park. They found that 68% of the koalas were between 5 and 9 years old.  
 (a) What is the mean age of the koalas?  
 (b) What percentage of the koalas would be older than 3?  
 (c) What percentage of the koalas would be older than 13?



8. Mason is a cricketer who calculated the mean of his batting scores after 100 innings to be 43 and 68% of his scores were between 31 and 55.
- What is the standard deviation of his scores?
  - What percentage of his scores were over 67?
9. The average (mean) rainfall in July for a town was 83 mm and 16% of rainfall totals in July were more than 98 mm.  
What is the standard deviation of the rainfall totals?
10. Records showed that 95% of the maximum temperatures in January in an outback town were between  $28^{\circ}\text{C}$  and  $42^{\circ}\text{C}$ .
- What is the mean maximum temperature?
  - What is the standard deviation of the maximum temperatures?
11. Dahna and Leena were discus throwers. They recorded the distances of 120 of each of their throws.  
The average (mean) distance thrown by Dahna was 45 m.  
The standard deviation of Dahna's throws was 2 m.  
The average of Leena's throws was 42 m and the standard deviation was 4 m.
- Draw a curve showing the shape of the frequency distribution histogram for each of the discus throwers.
  - Describe the differences in the distances thrown by the two throwers.



## Calculating Standard Deviation

The value of the standard deviation for a set of data can be calculated using most calculators.

However it is worthwhile to know the formula used to calculate standard deviation.

There are two versions of the formula:

- one is used when calculating the standard deviation of a **population**.
- the other is used when calculating the standard deviation of a **sample**.

For a **population** the standard deviation is calculated by using this formula:

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}}$$

$\bar{x}$  is the mean

$N$  is the number in the population

$\Sigma(x - \bar{x})^2$  means the sum of all the  $(x - \bar{x})^2$  values.

For a **sample** the standard deviation is calculated by using this formula:

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N - 1}}$$

$\bar{x}$  is the mean

$N$  is the number in the sample

$\Sigma(x - \bar{x})^2$  means the sum of all the  $(x - \bar{x})^2$  values.

The easiest way to calculate standard deviation using these formulae is to set up a table of values as shown in the following examples. A spreadsheet could be used.

**Example 1**

The ages of the 12 teachers at a small school are shown below. Calculate the mean and standard deviation of the ages. Give answers correct to one decimal place.

**28, 45, 31, 24, 33, 29, 25, 36, 30, 27, 26, 31**

With this example we are dealing with all the teachers at the school so the *population* version of the standard deviation formula is used.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}}$$

**Step 1** Calculate the mean ( $\bar{x}$ ).

$$\bar{x} = \frac{\Sigma x}{N}$$

$$\bar{x} = 30.4 \text{ (one decimal place)}$$

**Step 2** Draw the following table.

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
28		
45		
31		
24		
33		
29		
25		
36		
30		
27		
26		
31		

**Step 3** Complete the table.

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
28	-2.4	5.76
45	14.6	213.16
31	-0.6	0.36
24	-6.4	40.96
33	2.6	6.76
29	-1.4	1.96
25	-5.4	29.16
36	5.6	31.36
30	-0.4	0.16
27	-3.4	11.56
26	-4.4	19.36
31	0.6	0.36
<b>Total</b>		<b>360.92</b>

**Step 4** Calculate the standard deviation.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}}$$

$$\Sigma(x - \bar{x})^2 = 360.92$$

$$N = 12$$

$$\sigma = \sqrt{\frac{360.92}{12}}$$

$$= \sqrt{30.0766..}$$

$$\sigma = \mathbf{5.5} \text{ (one decimal place)}$$

**Example 2**

Ten bilbies were tagged in a national park and each fitted with a GPS to study their movement.

The distance travelled by each bilby, in metres, in one night was recorded and these distances are shown below.

68, 130, 96, 121, 87, 118, 104, 92, 88, 107

1. Calculate the mean and standard deviation of these distances. Give answers correct to one decimal place.
2. Calculate the distances between which 68% of these bilbies travelled each night.

With this example we are dealing with a sample of the bilbies so the *sample* version of the standard deviation formula is used.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N - 1}}$$

**Answers**

1.

**Step 1** Calculate the mean ( $\bar{x}$ ).

$$\bar{x} = \frac{\Sigma x}{N}$$

$$\bar{x} = 101.1 \text{ (one decimal place)}$$

**Step 2** Draw the following table.

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
68		
130		
96		
121		
87		
118		
104		
92		
88		
107		

**Step 3** Complete the table.

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
68	-33.1	1095.61
130	28.9	835.21
96	-5.1	26.01
121	19.9	396.01
87	-14.1	198.81
118	16.9	285.61
104	2.9	8.41
92	-9.1	82.81
88	-13.1	171.61
107	5.9	34.81
<b>Total</b>		<b>3134.90</b>

**Step 4** Calculate the standard deviation.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N - 1}}$$

$$\Sigma(x - \bar{x})^2 = 3134.90$$

$$N - 1 = 9$$

$$\sigma = \sqrt{\frac{3134.90}{9}}$$

$$= \sqrt{348.322..}$$

$$\sigma = \mathbf{18.7} \text{ (one decimal place)}$$

**2.** 68% of the bilbies travelled a distance of  $\bar{x} \pm \sigma$ .

$$= 101.1 \pm 18.7$$

$$= \mathbf{82.4 \text{ m and } 119.8 \text{ m}}$$



**EXERCISE 160**

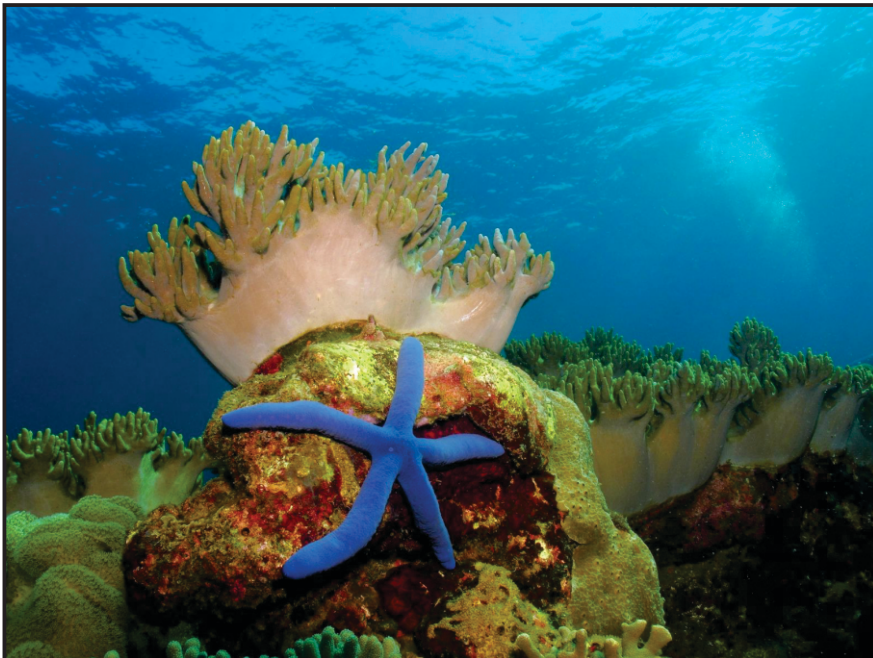
1. There were 14 students in a Year 10 mathematics class. For a test on statistics the marks achieved, as a percentage, by the students are shown below. Calculate the mean and standard deviation of these marks. Give answers correct to one decimal place.

68, 79, 93, 80, 82, 77, 95, 69, 73, 80, 88, 90, 66, 100

2. Marine scientists were studying a certain type of starfish in an area of the Great Barrier Reef. They divided this small area of the reef into 12 squares each with an area of  $200 \text{ m}^2$  and counted the number of starfish in each square. The number of starfish in each  $200 \text{ m}^2$  area is shown below.

50, 28, 22, 62, 33, 45, 27, 36, 31, 39, 28, 41

- (a) Calculate the mean and standard deviation for this data. Give answers correct to one decimal place.  
(b) Between which two amounts would 95% of the data lie?



## PROBLEM SOLVING

1. A die was rolled five times. The mode of the numbers that appeared on the top face was one more than the median which was one more than the mean.  
What were the five numbers that were rolled?



2. A basketball team is about to play in the 22<sup>nd</sup> game of the season. It needs to score 111 points to have an average of 100 points per game.  
What is the average points per game before this game?  
Give answer to two decimal places.



3. Zac has sat four tests each with 10 questions. The median of his test results is six correct answers and the mean is five correct answers. Zac completes another 10 question test and the median of his five results is seven correct answers and the mean is six.  
Each time Zac sat the test he answered more questions correctly than he did on the previous test.  
How many questions did he answer correctly on each of the five tests?

## PUZZLE

The answers to this puzzle are the same horizontally and vertically down.

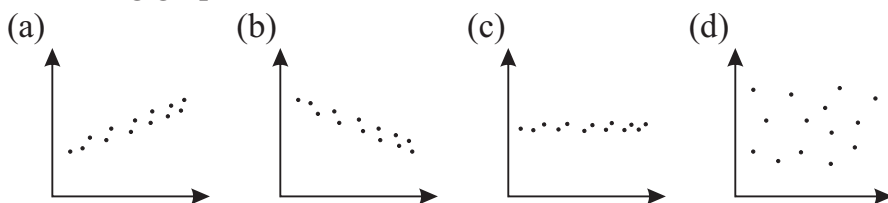
1. Unit of mass
2. Uncommon
3. Region
4. Average

1.	2.	3.	4.
2.			
3.			
4.			

## CHAPTER REVIEW

1. (a) There are 32 840 spectators at a game between the Rebels and Falcons. What is the minimum number of people that should be surveyed to find out who they barrack for?  
 (b) 200 people are surveyed and 125 barrack for the Rebels. Based on these figures, how many of the spectators barrack for the Rebels?
2. 400 people were asked what they ate for breakfast. The results were: cereal - 178, toast - 115, eggs - 59, fruit/yoghurt - 28, other - 20. If this data was displayed on a pie chart what would be the angle for each type of breakfast?

3. Describe the relationship between the data represented on the following graphs.



4. Find the mean, mode and median of the following sets of data.

- (a) 6, 15, 7, 8, 11, 10, 11, 12, 13, 11, 10  
 (b) 25, 19, 24, 27, 31, 30

5. Find the mean, mode and median of the data displayed in this frequency table.  
 Give the mean correct to one decimal place.

<i>Value</i> $x$	<i>Frequency</i> $f$
30	6
31	9
32	13
33	15
34	4

6. Construct a frequency table for the times (in seconds) given below and find the mean and median times.

56 58 65 56 57 58 63 62 61 59 58 60 64 63 61 66  
 59 60 61 63 64 59 58 59 60 61 62 61 60 59 56 59

7. For the following set of data:
- calculate the mean (correct to one decimal place)
  - find the modal class
  - draw a cumulative frequency curve
  - find the median

<i>Class Interval (m)</i>	<i>Frequency f</i>
0-<10	12
10-<20	23
20-<30	45
30-<40	32
40-<50	19

8. For the masses (in kg) below:
- construct a non-ordered stemplot
  - construct an ordered stemplot
  - find the median
  - construct a frequency table that includes *class interval (30-<40, 40-<50, etc)*, *midpoint*, *frequency*,  $(x \times f)$  and *cumulative frequency*.
  - calculate the mean from this table correct to one decimal place
  - draw a cumulative frequency graph
  - find the approximate value of the median from this graph

45 68 48 83 51 50 42 66 72 61 58 67 62 70 71 56  
 64 73 68 60 59 44 59 52 38 47 56 63 71 82 65 47  
 56 37 80 77 63 57 43 48 58 62 59 66 73 47 64 61

9. Construct a boxplot for the following set of data clearly showing any outliers.

69 71 53 11 45 115 58 39 56

10. Robin is an archer who calculated the mean of his archery scores after 100 events to be 152 and 68% of his scores were between 125 and 179.
- What is the standard deviation of his scores?
  - What percentage of his scores were over 125?