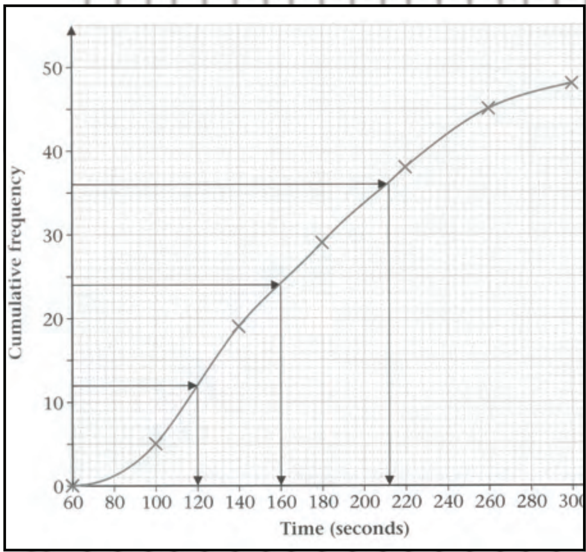
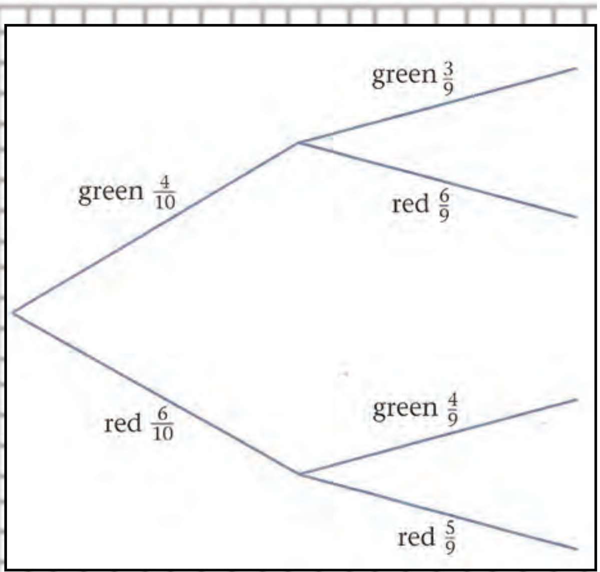
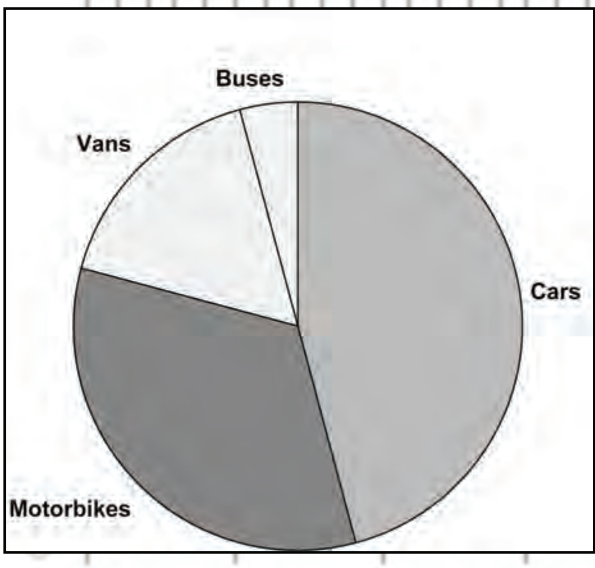
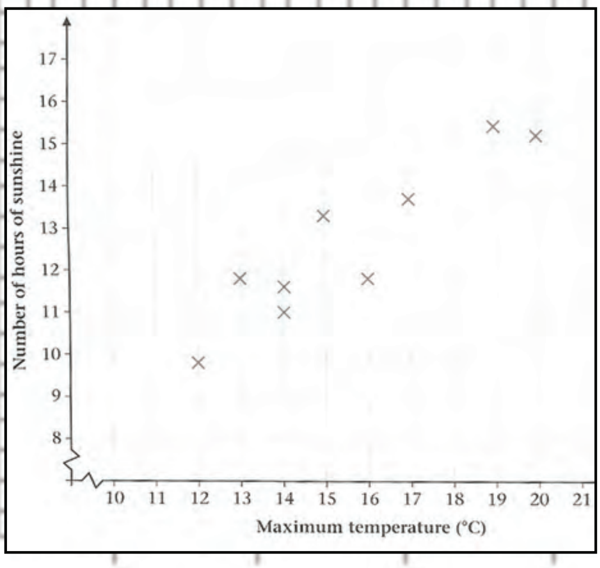




Maths Module 3: Statistics

Student's Book





Contents

1. Collecting Data

1.1 Qualitative and quantitative data

Data can be either **qualitative** or **quantitative**

- Qualitative data is described using words
- Quantitative data consists of numbers

Quantitative data can be either **discrete** or **continuous**

- Discrete data takes only certain values and has an exact value. For example, shoes can be bought in these sizes: 6 6 1/2 7 7 1/2 8 8 1/2 9

  The above values are discrete. There are no values between them. They are separate.
- Continuous data cannot be measured exactly. The accuracy of the measurement depends on the accuracy of the measuring device.

EXAMPLE: Ma Thandar works as a dressmaker. She makes a wedding dress for her friend. Write down two variables associated with the dress that are:

a. qualitative **b.** discrete **c.** continuous

Answer:

a. The colour of the dress is qualitative. We use words to describe colour.
b. The number of buttons is discrete. It is described by whole numbers.
c. The length of a dress is continuous. A measurement in centimetres or inches can take any value.

Practice

i. In the following questions circle whether the data is discrete or continuous:

a. Shoe Size	Discrete	Continuous
b. Number of children in a school	Discrete	Continuous
c. The length of someone's arm	Discrete	Continuous
d. Number of songs on a CD	Discrete	Continuous
e. The temperature today	Discrete	Continuous
f. The number of teeth in your mouth	Discrete	Continuous

ii. Make a list of all the characteristics you could use to describe a person - hair colour, sex, age etc. Catagorise your list as discrete or continuous.

iii. For each type of data below, write down whether it is qualitative or quantitative

a. The names of students in a class
b. The heights of students in a class
c. The most popular football team in Asia
d. The number of children in a family
e. The native countries of migrants in Thailand
f. The amount of rainfall in a day in Chin state
g. The gender of school teachers in Mandalay
h. The number of people who die from malaria each year in Myanmar
i. The weight of rice bought in a month by a family in Yangon

iv. For each type of quantitative data given above, state whether it is discrete or continuous.

v. Give an example of a qualitative and quantitative measure associated with:

- **a.** A motorbike
- **b.** A herd of elephants
- **c.** A person from Northern Myanmar

1.2 Sampling

To collect data we need to do a survey. Who we survey depends on the population we want to collect data from. If we want to collect data from everybody in the population then we need to do a census.

Often it is not possible to do a census so we take a **sample** of the population. Taking a sample involves choosing part of the population and collecting data only from that part.

EXAMPLE: Would a census or a sample be used to investigate the following?

- **a.** People's opinions of Barack Obama in America which has a population of around 300,000,000.
- **b.** People's opinions of Barack Obama in smalltown which has a population of 105

Answer:

- **a.** The population in this investigation is everybody who is American. It would take too long to survey everybody so a sample would be chosen.
- **b.** The population in this investigation is everybody who lives in Smalltown. As this population is quite small (105 people) then a census could be taken.

Practice

For each survey in the list below, identify the population being surveyed and state whether a census or a sample would be used.

- **a.** The names of students in a class
- **b.** The heights of students in a class
- **c.** The number of children in families in a small village
- **d.** The number of children in families in China
- **e.** The nationalities of migrants in Thailand
- **f.** The gender of school teachers in Mandalay
- **g.** The number of people who die from malaria each year in Myanmar - myriad pro

1.3 Primary and secondary data

Data that we collect ourselves is called **primary data**. Primary data can be collected by using questionnaires, interviews, experiments and observatio ns.

Data that already exists is called **secondary data**. We can use it but it was collected by another person or organisation. The place we find the data is called the **source**. Secondary data can found by searching the internet, reading reports or looking at statistics from the government, United Nations, etc.

Practice

i. State whether you would collect primary or secondary data for the research given below. Explain your answers.

- **a.** A survey of student attendance at school
- **b.** A report about daily sales in your local teashop
- **c.** A report about tourist places in Myanmar
- **d.** A survey about local people's opinions of the United Nations
- **e.** A report comparing levels of poverty in African countries.

ii. Explain one possible source of secondary data for questions **c** and **e** above.

iii. Complete the table below to catagorise the statements as advantages or disadvantages of primary and secondary data.

Cheap to collect
Takes a long time to collect
You know how it was collected
Data may be old
Expensive to collect
The data may be inaccurate
Easy to collect
Can choose who to collect data from

Data Type	Advantages	Disadvantages
Secondary		
Primary		

iv. If you have access to the internet use it to find the following secondary data. Give two sources for each answer.

a. The member countries of the European Union
b. The five largest countries in the world by Population
c. The five poorest countries in the world
d. The ten most recent presidents of the United States of America
e. Five of the rights listed in the Universal Declaration of Human Rights

1.4 Methods for collecting primary data

There are a variety of different ways to collect data once the sample has been chosen. These include:

Questionnaires

- We write a set of questions and give them to people to answer.
- The questions can ask for facts or opinions.

Interviews

- We prepare a set of questions to ask people and record what they say.
- Each interview can involve just one person or a group of people.

Observations

- We collect the data we need by going and watching what is happening.
- For example we could observe the number of people going to a shop in one day.

Experiments

- Useful for collecting scientific data.
- Experiments are used to test if an idea is true or not.

In this section we will focus on questionnaires.

EXAMPLE: Design a short questionnaire to survey the kind of music people in Smalltown like.

Are you (tick the correct response)?							
Male		Female					
How old are you?							
Under 17		17-20		21-30		Over 30	
What kind of music do you like the most?							
Rock		Metal		Traditional		None	

The example above shows a good questionnaire because:

- The questions are short
- The questions are easy to answer
- The language is simple
- It is clear how to answer the questions

Practice

i. Look at the 3 questions below, for each one

a. Give at least one reason why it is not a good question

b. Suggest a way to improve the question

How old are you?					
Young		Middle aged		Old	

How tall are you?					
Under 1 metre		Between 1 and 2 metres		Over 2 metres	

Do you agree that the teachers in your school are amazing?			
Yes		No	

ii. Write a questionnaire of no more than 5 questions to survey students' expectations and ambitions when they finish school.

1.5 Recording data in tables

EXAMPLE: The table below shows the days of the week on which some students were born.

Day	Tally	Frequency
Monday	卌 ||	
Tuesday	卌	
Wednesday	||||	
Thursday	卌 卌	
Friday	卌 |||	
Saturday	卌 ||	
Sunday	卌 卌 |	
	TOTAL	

Once we have completed a survey by questioning, interviewing or observing, we need a way of recording the data so we can analyse the results. The easiest way to do this is by using tables. Tables give us a way to understand what data means.

Each person is recorded by a single line in the tally column. Groups of 5 are recorded using 卌.

The total number of students born on each day is shown in the frequency column.

The completed table is a frequency distribution.

Think

The table in the example is not complete. Read the questions and write your answers in the table:

a. How many students were born on Wednesday?
b. On which day were the most students born?
c. On which two days were the same number of students born?
d. What is the total number of students surveyed?

The data in the table above is discrete. For continuous data, the data is organised in groups. The groups are called **class intervals**. The data is in the table below is grouped in equal class intervals **of 10 years.**

EXAMPLE: The table below shows population data for the United Kingdom in 1999.

Ages	**Frequency**
0 - 9	3,890,782
10 - 19	4,088,469
20 - 29	4,172,971
30 - 39	5,042,082
40 - 49	4,818,389
50 - 59	3,454,277
60 - 69	2,374,917
70 - 79	1,777,692
80 - 89	744,780
90+	126,935

Think

Answer the questions about the data:

a. Which class interval has the largest population?
b. Which class interval has the smallest population?
c. How many people were aged between 10 and 29 in the United Kingdom in 1999?
d. Why do you think the final class interval is different from the others?

Practice

i. A class of students was asked what job they would like to do when they graduate from school. The answers to the survey are shown below:

Teacher	Doctor	Translator	Doctor	Teacher
Teacher	Doctor	Musician	Soldier	Musician
Translator	Translator	Soldier	Teacher	Musician
Nurse	Teacher	Nurse	Teacher	Doctor
Doctor	Soldier	Nurse	Teacher	

a. Draw a frequency table for the data
b. How many students were surveyed?
c. Which job was the most popular?
d. Why is it useful to show this data in a table?

ii. The ages in years of 40 people are shown below

27	**34**	**54**	**57**	**3**	**12**
15	**19**	**29**	**30**	**33**	**42**
35	**20**	**29**	**28**	**9**	**11**
26	**42**	**50**	**26**	**10**	**7**
33	**49**	**21**	**18**	**1**	**25**
24	**34**	**19**	**20**	**27**	**37**
43	**56**	**37**	**14**		

a. Use the data to complete the table below.
b. What is the width in years of each class interval?
c. How many people are in the interval 30-39?
d. How many people are less than 20 years old?

Age	Tally	Frequency
0-9		
10-19		
20-29		
	TOTAL	

2. Analysing Data

2.1 Mean, mode and median

In this section we will learn about three different methods for calculating averages of data. These methods are called mean, mode and median. Knowing these averages helps to understand the data.

EXAMPLE: The table below shows the population of 7 Asian countries. The populations are given to the nearest million. Calculate

a. The mean **b.** The median **c.** The mode

Country	Population (millions)
Malaysia	26
Vietnam	84
Bangladesh	147
South Korea	48
Phillipines	86
Thailand	63
Myanmar	48

Answer:

a. We calculate the mean by adding all the populations and dividing by the number of countries

$$\text{mean} = \frac{26 + 84 + 147 + 48 + 86 + 63 + 48}{7} = \frac{502}{7} = 72$$

To the nearest million the mean population of the 7 countries is 72,000,000.

b. To find the median we put the populations in order: 26, 48, 48, 63, 84, 86, 147. The median is the middle value. In this example the middle value is the 4th value. So, median = 63,000,000.

c. The modal population is the population which occurs the most often. Here, the mode = 48,000,000.

Think

Look at the example above and complete the sentences using the words in the box:

middle	sum	median	mode	divided	most

The mean of a set of data is the ____________ of all the values ______________ by the total number of values.

The ______________ is the ______________ value when the data is arranged in order of size.

The _____________ of a set of data is the value which occurs the ____________ often.

Practice

i. Ten students got the following marks out of 40 in their English exam:

37 34 34 34 29 27 27 10 4 28

Calculate: **a.** the mode **b.** the median **c.** the mean

ii. The table below shows the population of the capital cities of the 7 Asian countries on the previous page.

Calculate: **a.** the mode **b.** the median **c.** the mean

City	Population
Kuala Lumpur	1,145,000
Hanoi	3,083,000
Dakar	3,839,000
Seoul	10,231,217
Manila	1,581,082
Bangkok	5,882,000
Yangon	4,082,000

iii. The graph below shows the ten countries in the world with the highest number of migrants from other countries.

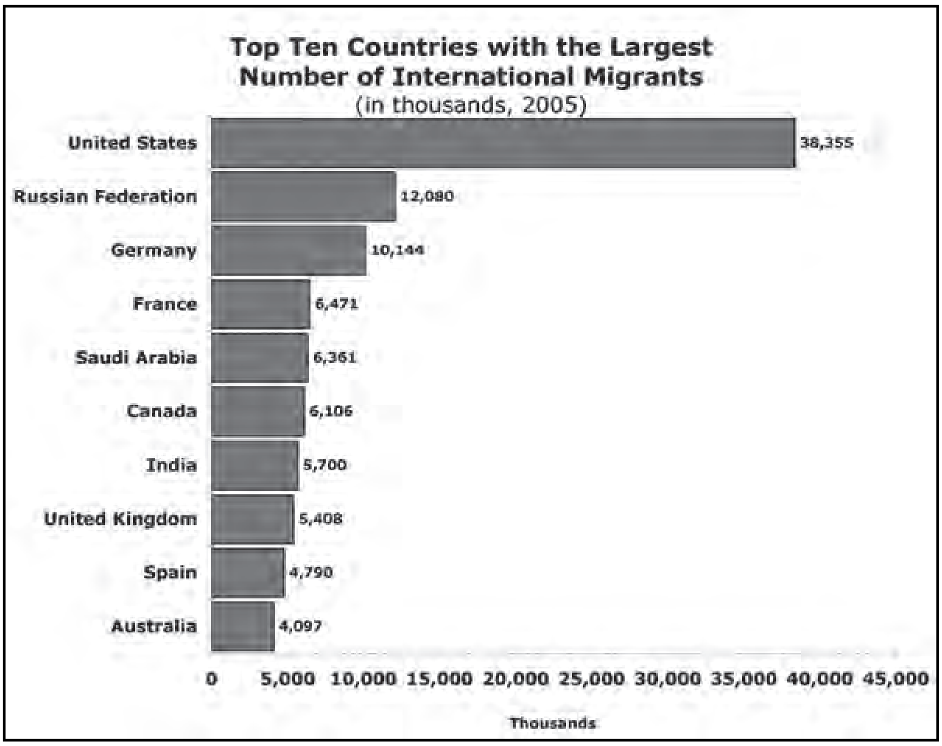
a. Circle the correct migrant population of France.

6,471 647,100 6,471,000 64,710,000

b. Circle the correct migrant population of Russia.

twelve million and eighty thousand one million two hundred thousand eight hundred

twelve million and eighty



iv. Use the graph above to calculate the mean. There is also a formula for finding the mean:

Σ means 'the sum of.....'.

It is pronounced 'sigma'

x is each data value

$$\text{mean} = \frac{\Sigma x}{n}$$

n is the total number of data values

Think

The total of a set of 12 numbers is 36. Calculate the mean.

In this example Σx = _____? n = _____?

Once we know Σx and n we can calculate the mean. In this example mean Σx/n = _____?

2.2 Choosing an appropriate average

The mode is useful when we want to know the most common value for the data, e.g.

- Which job is the most common choice for graduating students?
- Which day of the week is the most common birthday? The mean and median are useful when we want to give a typical value for the data. However, the mean is influenced if one value in the data is a lot bigger or smaller than the other values.

EXAMPLE: Use the table from page 7, adding to it a row to include China (below).

a. Calculate **(i)** the mean and **(ii)** the median.

b. Explain why we would not use the mean in this example.

Country	**Population (millions)**
China	1313

Answer:

$$\text{mean} = \frac{26 + 84 + 147 + 48 + 86 + 63 + 48 + 1313}{8} = \frac{1815}{8} = 227$$

a. (i) The mean is

To the nearest million the mean population of the 8 countries is 227,000,000.

(ii) 26, 48, 48, 63, 84, 86, 147, 1313. The median is the middle value. In this example there are an even number of data values so the median is the mean of the middle two data values.

$$\text{median} = \frac{63 + 84}{2} = \frac{147}{2} = 73.5 = 74$$

The median population is 74,000,000.

c. In the example in 2.1 the mean population was 72,000,000. In this example we added China to the data. China has a very large population. The mean has increased to 227,000,000. This number is much higher than 7 of the countries' populations, so we would not use the mean in this example. It would be better to use the median value.

2.3 The quartiles

The median divides a set of data into two. It is the value half way into the data. We can also divide data into four quarters. When the data is arranged from the smallest to the largest:

- The **lower quartile** is the value one quarter of the way into the data.
- The **upper quartile** is the value three quarters of the way into the data

EXAMPLE: 15 people were asked 'How many cousins do you have?' The data is given below in order from smallest to largest. Find

a. The lower quartile **b.** The median **c.** The upper quartile

Number of cousins	0	2	4	5	7	8	10	12	12	12	13	14	14	15	16

Answer:

a. The lower quartile is the 4th value. The lower quartile is 5.

b. The median is the 8th value. The median is 12.

c. The upper quartile is the 12th value. The upper quartile is 14.

Think

In the example above n = 15. Use this information to match the words on the left with the correct formula on the right.

In the example above n = 15. Use this information to match the words on the left with the correct formula on the right.

Lower quartile	$\left(\frac{n+1}{2}\right)$ th Value
Median	$\frac{3(n+1)}{4}$ th Value
Upper quartile	$\left(\frac{n+1}{4}\right)$ th Value

EXAMPLE: The table below shows the life expectancy of people in South America. Calculate:

a. The median **b.** The lower quartile **c.** The upper quartile

Country	**Argentina**	**Bolivia**	**Brazil**	**Chile**	**Colombia**	**Ecuador**
Life expectancy	75.5	71.1	71.1	76.8	71.1	71.9

F. Guiana	**Guyana**	**Surinam**	**Paraguay**	**Peru**	**Uruguay**	**Venezuala**
76.7	63.1	69.2	74.4	70.9	75.9	73.8

Answer:

Before answering any of the questions we need to arrange the numbers from smallest to largest.

63.1 69.2 70.9 71.1 71.1 71.1 71.9 73.8 74.4 75.5 75.9 76.7 76.8

a. The median is the $\frac{13+1}{2} = 7\text{th value}$. The median is 71.9 years.

b. The lower quartile is the $\frac{13+1}{4} = 3\frac{1}{2}\text{th value}$. The lower quartile is halfway between the 3rd value and the 4th value, (70.9 + 71.1)/2 = 71 years.

c. The upper quartile is the $\frac{3(13+1)}{4} = 10\frac{1}{2}\text{th value}$. The upper quartile is halfway between the 10th and 11th value, (75.5 + 75.9)/2 = 75.7 years.

Practice

Below is the table from page 8 showing capital city populations of 7 Asian countries.

Calculate:

a. The lower quartile

b. The upper quartile

City	Population
Kuala Lumpur	1,145,000
Hanoi	3,083,000
Dakar	3,839,000
Seoul	10,231,217
Manila	1,581,082
Bangkok	5,882,000
Yangon	4,082,000

2.4 The range and interquartile range

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

Range = highest value - lowest value

The interquartile range is the difference between the upper and lower quartiles.

Interquartile range = upper quartile - lower quartile

EXAMPLE: Use the table to find:

a. The range **b.** The interquartile range

Number of children (x)	Frequency (f)
0	23
1	30
2	28
3	17
4	9
Total (Σf)	**107**

Answers:

a. Range = highest value - lowest value = 4 - 0 = 4

b. Lower quartile is the

$$\frac{n+1}{4} = \frac{107+1}{4} = \frac{108}{4} = 27\text{th value}$$

The 27th value is in the catagory of families with 1 child.

The lower quartile is 1 child.

Upper quartile is $\frac{3(n+1)}{4} = \frac{3(107+1)}{4} = \frac{324}{4} = 81\text{st value}$

The 81st value is in the catagory of families with 2 children. The upper quartile is 2 children.

The interquartile range = upper quartile - lower quartile = 2 - 1 = 1 child.

Practice

i. Use this table to find the range and the interquartile range for the life expectancies of people in South America.

Country	Argentina	Bolivia	Brazil	Chile	Colombia	Ecuador
Life expectancy	75.5	71.1	71.1	76.8	71.1	71.9

F. Guiana	Guyana	Surinam	Paraguay	Peru	Uruguay	Venezuala
76.7	63.1	69.2	74.4	70.9	75.9	73.8

ii. Use the data for the practice question above showing the capital city populations of seven Asian countries to calculate the range and interquartile range of the populations.

2.5 Averages from frequency distributions

In this section we will learn how to find averages from frequency distributions.

There is also a formula for finding the mean of a frequency distribution:

Σ means 'the sum of.....'.
It is pronounced 'sigma'

x is each data value

$$\text{mean} = \frac{\Sigma fx}{\Sigma f}$$

f is the frequency for each data value

EXAMPLE: Use the formula to calculate the mean for the table in the previous example. To do this we add a column by calculating the values fx:

Number of children (*x*)	**Frequency (*f*)**	***fx***
0	23	0
1	30	30
2	28	56
3	17	51
4	9	36
	Σf = 107	**$\Sigma f x$ = 173**

When we have the values for Σf and Σfx we can calculate the mean:
$\Sigma fx/\Sigma f = 173/107 = 1.62$

Practice

i. The table below shows the number of goals scored per game in 31 matches.

Number of goals	0	1	2	3	4	5	6	7
Frequency	4	11	8	6	1	0	0	1

a. Use the table above to complete the one to the right

b. Use the formula to calculate the mean of the data

Number of goals (___)	**Frequency (___)**	***fx***
0		
1		
2		
3		
4		
5		
6		
7		
	Σf = ______	**Σf x = ___**

ii. The table below shows the results of a survey into the number of people in different households. Use the method in the example to find the mean of the data.

Number of people	2	3	4	5	6	7	8	9
Frequency	4	11	8	6	3	2	0	2

2.6 Averages from grouped data

Practice

EXAMPLE: The completed table from exercise ii on page 6 is shown below. Calculate:

a. the mean **b.** the median **c.** the mode

Age	Frequency (f)	Middle value (x)	*fx*
0 - 9	4	4.5	18
10 - 19	8	14.5	116
20 - 29	12	24.5	294
30 - 39	8	34.5	276
40 - 49	4	44.5	178
50 - 59	4	55.5	222
Total (Σ*f*)	**40**	**Total (Σ*fx*)**	**1104**

The values in the third column are the middle values of the class intervals.

For example (0 + 9)/2 = 4.5

We need these values to calculate the mean.

Answer:

The formula for calculating the mean is

$$\frac{\text{sum of (middle values x frequencies)}}{\text{sum of frequencies}} = \frac{\Sigma fx}{\Sigma f} = \frac{1104}{40} = 27.6$$

a. There were 40 people so the median is given by the (40 + 1)/2 = 20.5 th value. This values lies in the class interval 20 - 29 years. We cannot give an exact value for the median.

b. We do not find the mode for grouped data. We find the modal class. The modal class is the class interval with the highest frequency. In this example the modal class is 20 - 29 years.

Note: We only use the modal class if the class intervals are the same.

i. The table below shows the time taken by a class of students to complete their maths homework.

Time taken (minutes)	0-9	10-19	20-29	30-39	40-49
Number of students (*f*)	2	4	12	5	2

a. Use the table above to complete this table:

Age	Frequency (*f*)	Middle value (x)	*fx*
0 - 9		4.5	9
10 - 19	4		
20 - 29			294
30 - 39	5		
40 - 49		44.5	
Total (Σf)		**Total (Σfx)**	

b. Find the mean for the data using the method in the example.

ii. For each table below, draw a new table similar to the one in the example and find the mean of the data.

a.

Class	1-5	6-10	11-15	16-20
Frequency	2	9	3	1

b.

Class	10-19	20-29	30-39	40-49	50-59
Frequency	8	11	13	9	7

c.

Class	10-12	12-14	14-16	16-18	18-20
Frequency	1	5	12	3	0

2.7 Scatter diagrams

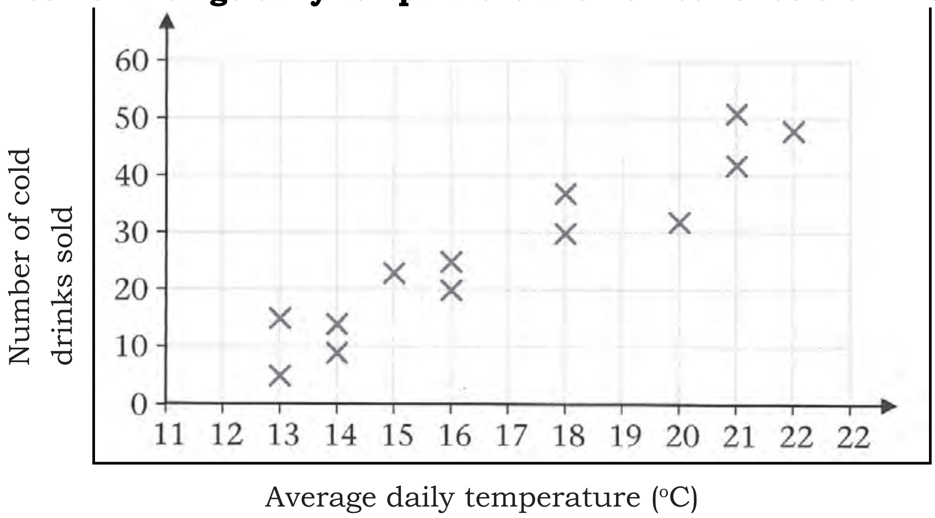
We can use scatter diagrams to show if two sets of data are related.

EXAMPLE: Kyi Phyu wanted to know if there was a relationship between the number of cold drinks sold and the average daily temperature. To do this she did a survey for 13 days. The results of the survey were:

Average Temperature (°C)	13	14	21	22	16	18	13	20	21	18	15	16	14
Cold drinks sold	5	9	51	48	20	30	15	32	42	37	23	25	14

We can plot each point (13,5), (14,9) etc. on a graph. This graph is called scatter diagram or scatter graph.

Comparison of average daily temparature and number of cold drinks sold



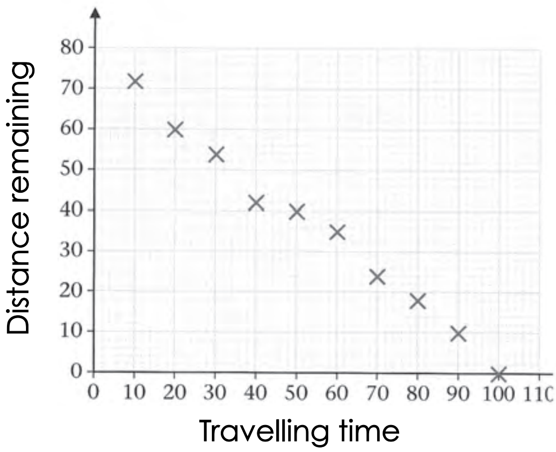
The graph shows that there may be a relationship between the number of drinks sold and the average daily temperature. As the temperature increases the number of drinks sold increases.

Practice

i. In the example above, why do you think more drinks are sold as the temperature increases?

ii. While driving to visit her sister, Chandra wrote down the distance remaining every 10 minutes. The scatter diagram below shows the results. Is there a relationship between the time she has been driving and the distance remaining?

Relationship between travelling time and remaining distance



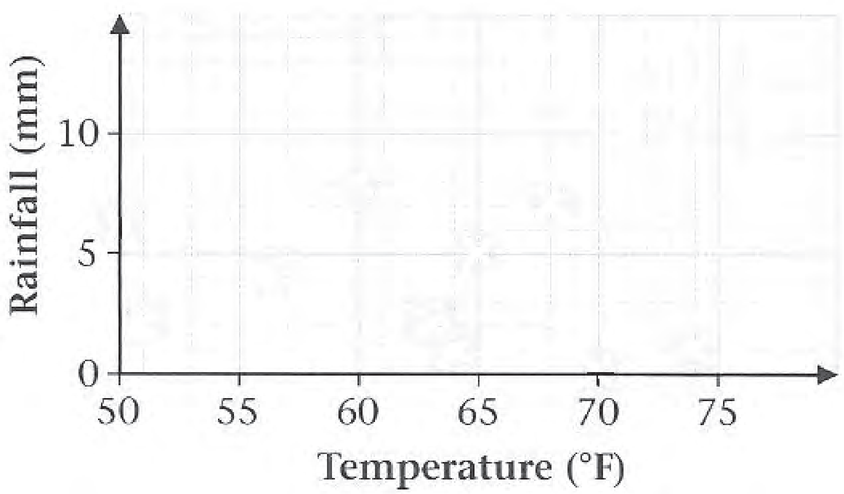
iii. Every Monday for 11 weeks, Aung Mon recorded the temperature in °F and the rainfall in mm. The results were:

Temperature (°F)	74	70	63	68	65	64	60	51	54	56	50
Rainfall (mm)	1	0	2	7	5	1	8	2	7	4	6

a. Use the table to complete the scatter diagram below

b. Use the scatter diagram to say if there is a relationship between the temperature and rainfall.

Comparison of daily temperature and amount of rainfall



These examples show that scatter diagrams can be used to show if there is a relationship between two sets of data. A relationship between two sets of data is called a correlation.

Positive Correlation

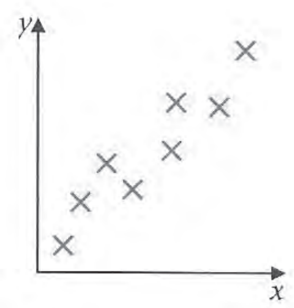
When the values in two sets of data increase or decrease at the same time, they have a positive correlation.

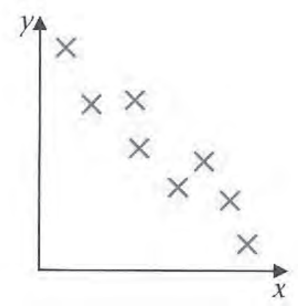
Negative Correlation

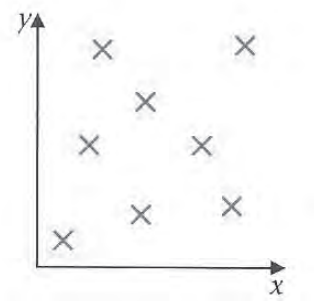
When the values in one data set decrease as the values in the other set increase, there is a negative correlation.

No Correlation

When there is no relationship between the two sets of values, there is no correlation.







Think

i. Look at the example and question **i.** on the previous page and complete the sentences

There is a ________________ correlation between the average daily temperature and the number of cold drinks sold, because as the temperature __________________ the number of cold drinks sold ______________.There is a ________________ correlation between the time spent driving and the distance remaining, because as the time ____________________ the distance remaining _________________.

Practice

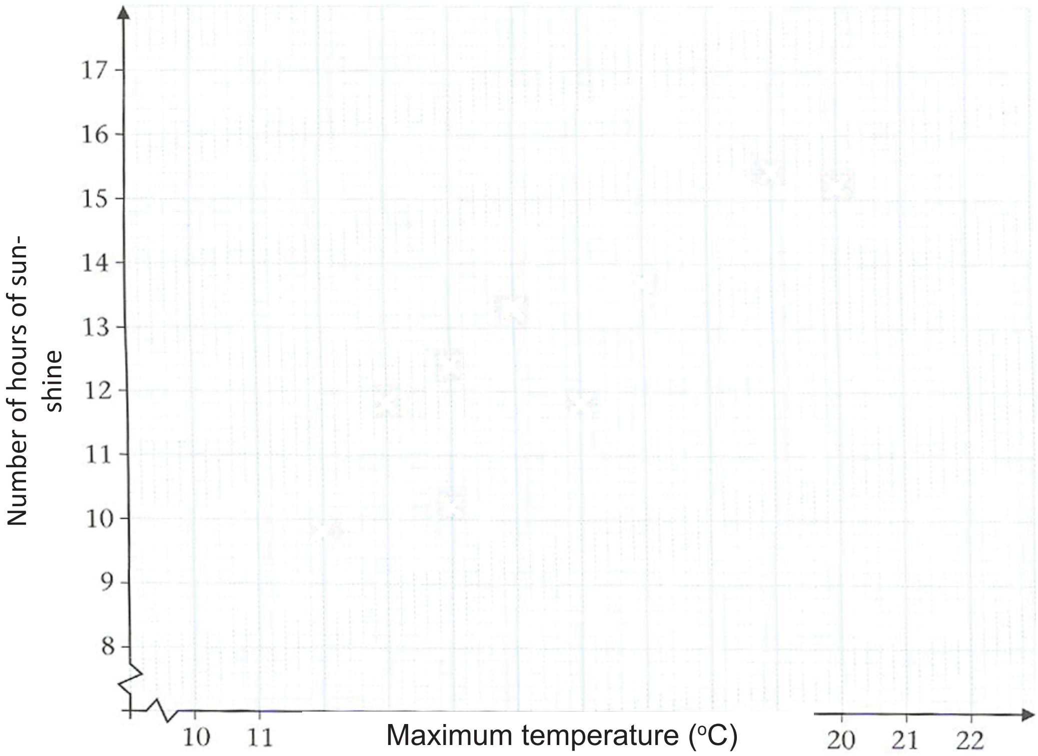
i. The table below shows the number of hours of sunshine and the maximum temperature in nine British towns on one day.

Hours of sunshine	13.5	15	10	11.5	12	15	12	11	14
Maximum temperature (°C)	15	20	12	14	13	19	16	14	17

a. Complete the scatter diagram on the next page by plotting the points and giving it a title

b. Use the scatter diagram to complete the sentence below:

There is a ________________ correlation between the hours of sunshine and the maximum temperature, because as the hours ____________________ the temperature _________________.



ii. Min Tin wanted to answer this question:

'Is there a relationship between the area of a country and its population?'

To answer the question, he searched the internet to find the area and population of 10 countries. The results are shown below:

Area (000s)	250	184	292	387	435	211	169	146	262	95
Population (millions)	31	17	16	79	75	61	27	127	47	61

a. Draw a scatter diagram of the data

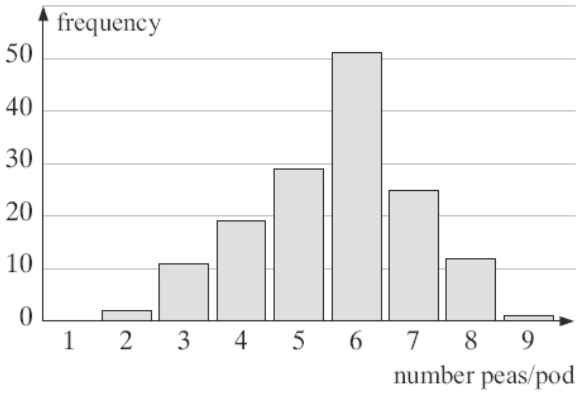
b. Comment on the relationship between area and population of a country.

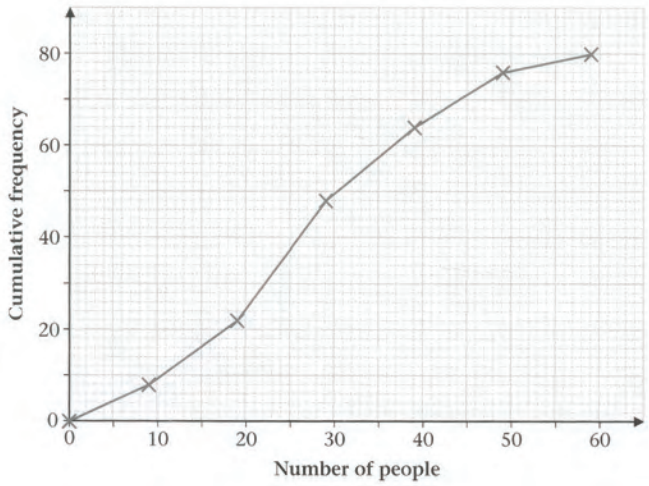
3. Presenting Data

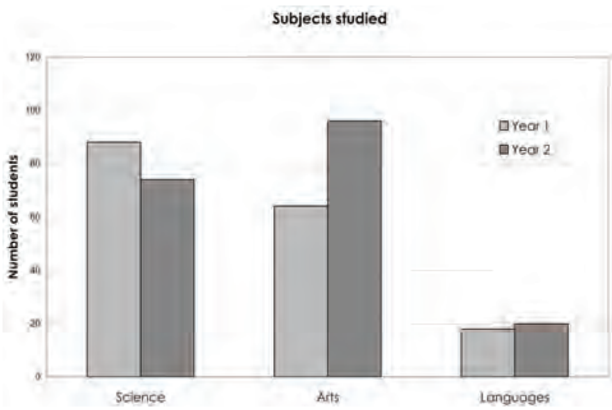
3.1 Introduction

Diagrams are often used to present data. There are many different kinds of diagrams. In this chapter we will learn about 4 different types:

- **Pie charts**, **bar graphs and cumulative frequency graphs** for presenting discrete data.
- **Histograms** and **cumulative frequency graphs** for continuous data.







Think

Look at the diagrams shown. You may have seen diagrams similar to this before.

a. Make a list of the places where you have seen diagrams for presenting data either at home, at school or in your community. Think about what these diagrams were presenting.

b. Why do you think people use diagrams to present data?

3.2 Pie Charts

Pie charts are mostly used when the data can be organised into categories, such as different colours or types of transport.

EXAMPLE: Some students did a survey of the types of vehicles passing in front of their school in one hour. They recorded the results in a table. Draw a pie chart of the data.

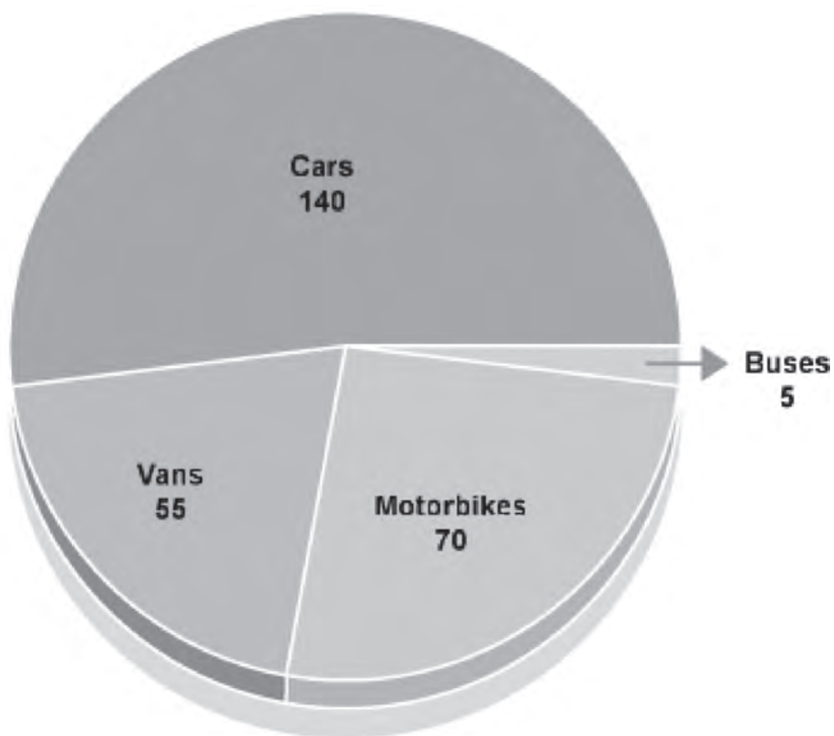
Type of vehicle	Number of vehicles
Cars	140
Motorbikes	70
Vans	55
Buses	5
Total vehicles	**270**

Answer: A pie chart is a circular diagram so we need to represent each part of the data as a proportion of 360 degrees. Look at the table to see how we do this:

Type of vehicle	Number of vehicles	Calculation	Degrees of circle
Cars	140	(140/270)*360	187
Motorbikes	70	(70/270)*360	93
Vans	55	(55/270)*360	73
Buses	5	(5/270)*360	7

We then need to use a protractor and a compass to draw the pie chart:

Types of vehicles passing in front of our school



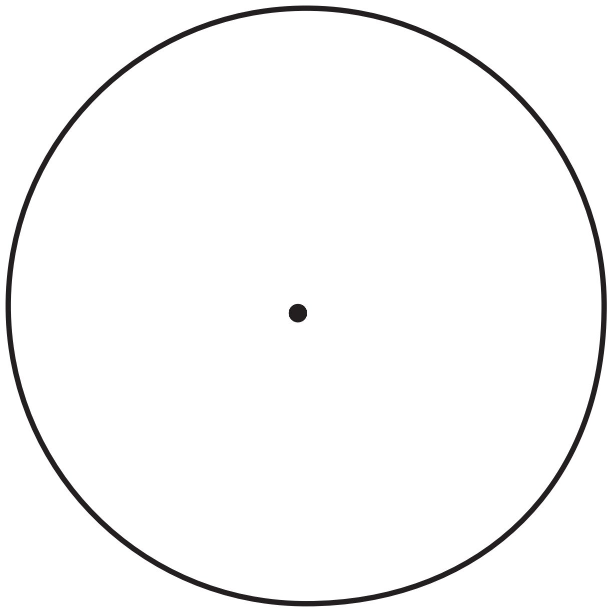
Practice

i. The following day the students did another survey of the types of vehicles passing the town hall.

a. Complete the table of their survey:

Type of vehicle	Number of vehicles	Calculation	Degrees of circle
Cars	110	(110/240)*360	
Motorbikes	80		
Vans	40		
Buses	10		

b. Use a protractor to complete the pie chart of the data:



ii. The table shows the grades achieved by 30 students in their final exams. Draw a pie chart of the data.

Grade	A	B	C	D	E
Frequency	7	11	6	4	2

iii. The pie chart below shows how a group of scholarship students in America travel to university. Use the chart to answer the questions.

a. What is the most popular method of transport?

b. What fraction of the students travel by car?

c. Six students travel by car. What is the total number of students?

d. How many students travel by taxi?

e. How many students cycle?

3.3 Bar Graphs

EXAMPLE: A farmer wanted to know the effects of using fertiliser on his crop of peas. He set up two pieces of land of equal size and planted many peas in both. He used fertiliser on one piece of land but not the other. All the other factors were the same.

At harvest time he selected 150 pea pods from each piece of land and counted the number of peas in each pod. The results were:

Without fertiliser

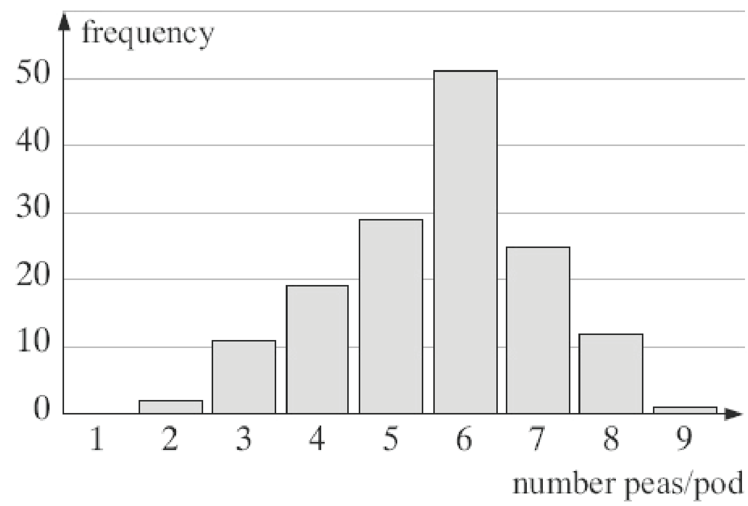
4 6 5 6 5 6 4 6 4 9 5 3 6 8 5 4 6 8 6 5 6 7 4 6 5 2 8 6 5 6 5 5 5 4 4 4 6 7 5 6
7 5 5 6 4 8 5 3 7 5 3 6 4 7 5 6 5 7 5 7 6 7 5 4 7 5 5 5 6 6 5 6 7 5 8 6 8 6 7 6
6 3 7 6 8 3 3 4 4 7 6 5 6 4 5 7 3 7 7 6 7 7 4 6 6 5 6 7 6 3 4 6 6 3 7 6 7 6 8 6
6 6 6 4 7 6 6 5 3 8 6 7 6 8 6 7 6 6 6 8 4 4 8 6 6 2 6 5 7 3

With fertiliser

6 7 7 4 9 5 5 5 8 9 8 9 7 7 5 8 7 6 6 7 9 7 7 7 8 9 3 7 4 8 5 10 8 6 7 6 7 5 6 8
7 9 4 4 9 6 8 5 8 7 7 4 7 8 10 6 10 7 7 7 9 7 7 8 6 8 6 8 7 4 8 6 8 7 3 8 7 6 9 7
6 9 7 6 8 3 9 5 7 6 8 7 9 7 8 4 8 7 7 7 6 6 8 6 3 8 5 8 7 6 7 4 9 6 6 6 8 4 7 8
9 7 7 4 7 5 7 4 7 6 4 6 7 7 6 7 8 7 6 6 7 8 6 7 10 5 10 4 7 7

In this form the data is not much use as we cannot compare the two sets. Before we make a bar chart we represent the data in a tally chart. For the 'without fertiliser' data we have:

Number of peas in a pod	Tally	Frequency
1		0
2	||	2
3	𝍸 𝍸 |	11
4	𝍸 𝍸 𝍸 ||||	19
5	𝍸 𝍸 𝍸 𝍸 𝍸 ||||	29
6	𝍸𝍸𝍸𝍸𝍸𝍸𝍸𝍸𝍸𝍸 |	51
7	𝍸𝍸𝍸𝍸𝍸	25
8	𝍸𝍸 ||	12
9	|	1



If we plot the number of peas in a pod on the horizontal axis and the frequency on the vertical axis then we can draw the bar graph, as shown.

The data is discrete so the bars in the chart do not touch.

Practice

i.

a. All graphs should have a title. What title would you give to the graph in the previous example?

b. What is the modal value for the data set in the example?

ii. Look at the table below, showing the number of goals scored in 31 matches.

Number of goals	0	1	2	3	4	5	6	7
Frequency	4	11	8	6	1	0	0	1

Draw a bar graph to represent the data. Remember to label the axes and give the graph a title.

iii. A class of 20 students was asked 'How many pets live in your house?' The following data was collected:

0 1 2 2 1 3 4 3 1 2 0 0 1 0 2 1 0 1 0 1

a. Is this data discrete or continuous. Why?

b. Draw a frequency table for the data.

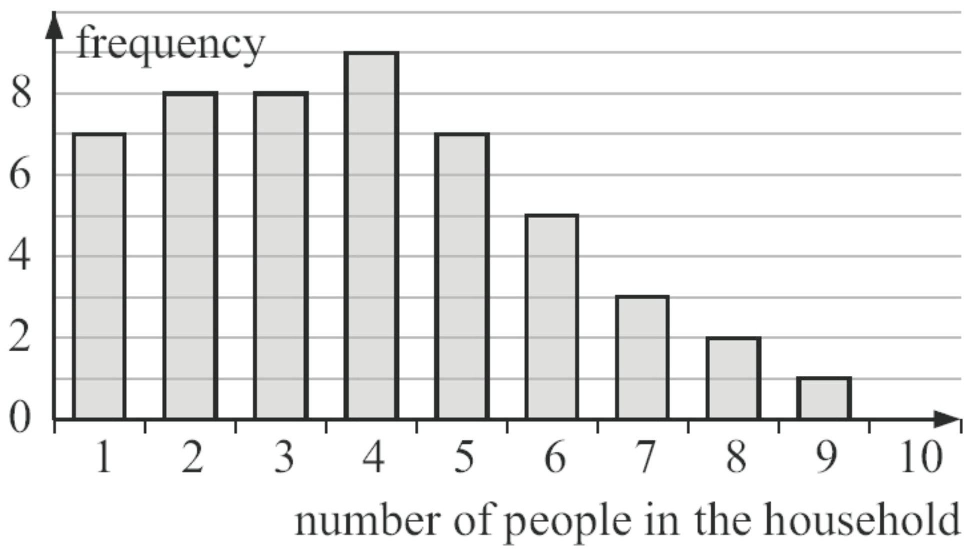
c. Use the frequency table to draw a bar graph Remember to label the axes and give the graph a title.

d. Look at the graph and answer the questions:

What percentage of the households had no pets?

What percentage of the households had 3 or more pets?

iv. All the households in a small village were asked 'How many people live in your household?' A bar graph of the results is shown below.



a. How many households gave data in the survey?

b. How many households had only one or two occupants?

c. What percentage of households had five or more occupants?

v. Look back at the example at the beginning of section 3.3. For the 'with fertiliser' data:

a. Organise the data in a tally-frequency table.

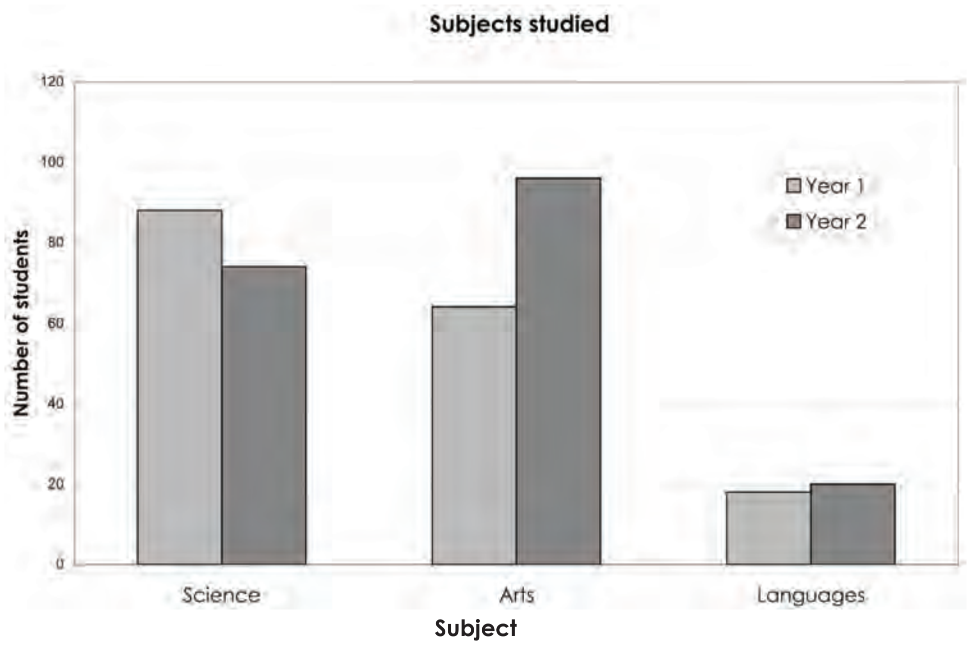
b. Draw a column graph of the data.

c. What evidence is there that fertiliser increases the number of peas in a pod?

3.4 Multiple bar graphs

We can display two sets of data side by side on a bar graph. This kind of bar graph is called a multiple bar graph.

EXAMPLE: The multiple bar graph below shows how many students studied which subjects in different year groups.



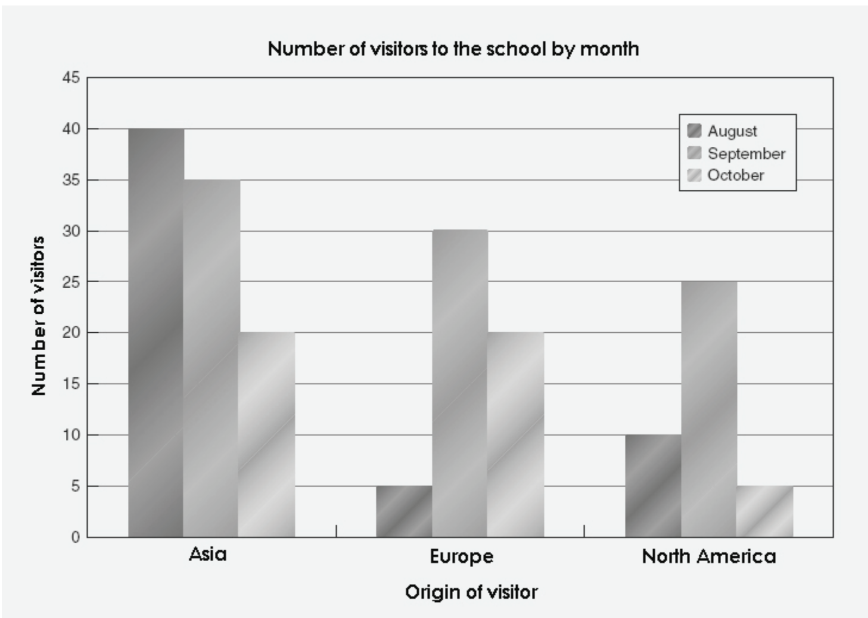
Think

Look at the bar graph above and answer these questions (approximate answers are acceptable):

- **a.** What do the first columns for each subject show?
- **b.** How many students studied Arts in year 1?
- **c.** How many year 2 students studied Science?
- **d.** How many language students were there in total?
- **e.** Which subjects were the most popular in year 2?
- **f.** How many students were there in year 1 and year 2?

Practice

i. Mandalay ELT College has a lot of visitors from different continents. The bar chart below shows the number of visitors from Asia, Europe and North America over a period of 3 months.



Use the bar graph to complete the statements.

a. The darkest columns represent the number of visitors in ______________________.

b. The school had the most visitors from Europe in _______________________.

c. The school had the fewest visitors from Asia in ___________________.

d. The school had ________ visitors from Asia in September.

e. The school had _________ visitors from America during the three months shown.

f. The month with the most visitors was _____________.

g. The total number of visitors for the 3 months was ________________.

ii. Myitkyina Health clinic records all the child patients it treats in one day. The table shows the numbers for last Saturday. M means male, F means female. Use the table to answer the questions below. Work in pairs.

Age	Sex M/F		Age	Sex M/F		Age	Sex M/F		Age	Sex M/F
1	F		3	M		5	M		2	F
3	M		2	M		3	M		5	M
1	M		2	F		1	F		3	M
3	M		1	F		2	M		1	M
4	F		1	F		5	M		2	M
3	M		4	F		2	M		2	M
1	F		1	F		5	M		2	F
5	F		5	F		3	M		2	F

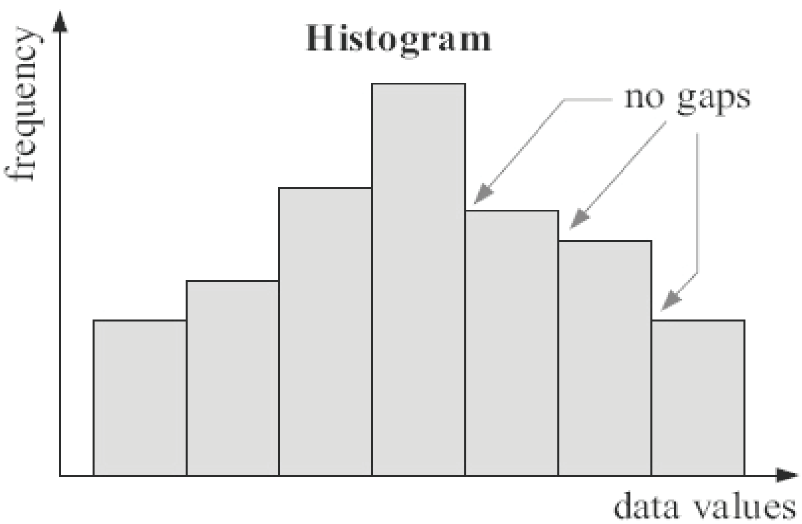
a. Draw two frequency tables - one for males, one for females - which show the frequencies by age.

b. Use the tables to draw a multiple bar graph to compare children by ages and sex.

c. Write five fill in the blank statements about the bar graph similar to those in question **i.** When you have finished swap your statements with another pair and answer their statements.

3.5 Histograms

We use bar charts when we are presenting discrete data. To present continuous data we organise the data into class intervals and draw a histogram. In a histogram the bars are connected to show that the data is continuous. If the bars are the same width then the frequency is given by the height of the bar. The diagram is an example of a histogram.



EXAMPLE: The people of Verti village are very tall. A sample of 30 people in the village was measured. The results are shown below in centimetres.

244.6	245.1	248	248.8	250	251.1	251.2	253.9	254.5	254.6
255.9	257	260.6	262.8	262.9	263.1	263.2	264.3	164.4	265.0
265.5	265.6	266.5	267.4	269.7	270.5	270.7	272.9	275.6	277 5

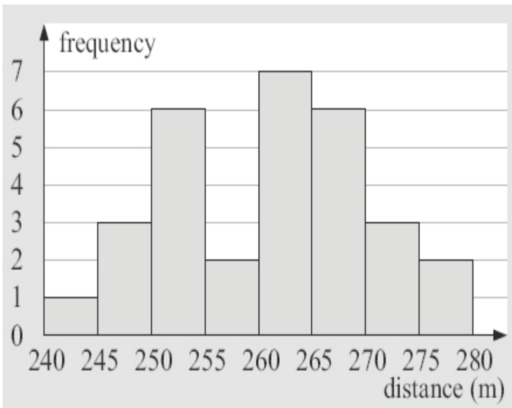
a. Group the data into class intervals **b.** Draw a histogram of the data

Answer:

a. The lowest value in the set is 244.6. The highest is 277.5. This gives a range of about 35. So we can use a group width of 5 to get 8 groups.

Height in cm (d)	Tally	Frequency
$240 \le d < 245$	|	1
$245 \le d < 250$	|||	3
$250 \le d < 255$	𝍸 |	6
$255 \le d < 260$	||	2
$260 \le d < 265$	𝍸 ||	7
$265 \le d < 270$	𝍸 |	6
$270 \le d < 275$	|||	3
$275 \le d < 280$	|	2

b. We can draw the histogram using this table, by putting the frequency on the vertical axis and the heights on the horizontal axis.



Practice

i. The histogram in the example needs a title. Write one.

ii. The table below shows the heights of a squad of basketball players. Draw a histogram to represent the data. Remember to label the axes and give the graph a title.

Height (cm)	Frequency
$170 \le H < 175$	1
$175 \le H < 180$	8
$180 \le H < 185$	9
$185 \le H < 190$	11
$190 \le H < 195$	9

iii. The numbers below show the weights of 17 parcels sent from relatives in America to their families in Yangon.

1.2 kg   1.8 kg   250 g   2.34 kg   2.99 kg   750 g   3.4 kg   3.85 kg   4.6 kg
2.12 kg   1.11 kg   1.67 kg   4.9 kg   4.12 kg   2.31 kg   1.75 kg   4.23 kg

a. Explain why this data is continuous.

b. Complete the grouped frequency table below. Each class interval has size 1 kg.

Weight (kg) (W)	Frequency
0 ≤ W < 1	2
1 ≤ W < 2	
2 ≤ W ___	4

c. Draw a histogram of the data. Label the axes and give the graph a title.

iv. In Oompa Loompa Land, the people are very short. The heights of the 30 shortest people in Oompa Loompa Land are:

115.7   122.1   110.2   129.7   130.5   122.9   133.5   113.7   120.7   115.2   125.9 126.2
128.0   134.7   131.1   118.3   112.5   128.3   132.3   133.5   133.6   119.7   123.4
124.1   134.0   130.2   134.4   126.8   129.1

a. Group the data into class intervals with width of 5 cm and create a frequency table

b. Draw a histogram of the data

3.6 Cumulative frequency

The cumulative frequency is the total frequency up to a particular upper class boundary.

EXAMPLE: The number of points scored in 40 games played between 2 people was recorded in a table. Use the table to draw a cumulative frequency table.

Number of points	Frequency
0 - 10	2
11 - 20	5
21 - 30	9
31 - 40	12
41 - 50	8
51 - 60	4

The number of points scored is discrete so the upper class boundaries are 10, 20, 30, ...

Number of points	Cumulative Frequency
0 - 10	2
0 - 20	2 + 5 = 7
0 - 30	2 + 5 + 9 = 16
0 - 40	2 + 5 + 9 + 12 = 28
0 - 50	2 + 5 + 9 + 12 + 8 = 36
0 - 60	2 + 5 + 9 + 12 + 8 + 4 = 40

EXAMPLE: The length of 50 books was recorded in a table. Use the table to draw a cumulative frequency table.

Length (cm)	0 - 8	9 - 13	14 - 18	19 - 23	24 - 28	29 - 33
Frequency	5	10	16	9	7	3

The length is continuous data so the upper class boundaries are 8.5, 13.5, 18.5, …

Length (cm)	Cumulative Frequency
$0 \le l < 8.5$	5
$0 \le l < 13.5$	15
$0 \le l < 18.5$	31
$0 \le l < 23.5$	40
$0 \le l < 28.5$	47
$0 \le l < 33.5$	50

Practice

Draw a cumulative frequency table for each table below.

a.

Time listening to the radio (hours)	Frequency
0 - 3	3
4 - 7	5
8 - 11	8
12 - 15	3
16 - 18	1

b.

Number of students in the class	Frequency
0 - 5	8
6 - 10	7
11 - 15	9
16 - 20	7
21 - 25	9

c.

Age of mother at birth of baby (years)	Frequency
16 - 20	3
21 - 25	6
26 - 30	17
31 - 35	26
36 - 40	11
41 - 50	2

d.

Daily temperature (°C)	Frequency
$-10 \le t < 0$	12
$0 \le t < 10$	86
$10 \le t < 20$	185
$20 \le t < 30$	79
$30 \le t < 40$	3

3.7 Cumulative frequency graphs

You can display data in a **cumulative frequency graph** by plotting the cumulative frequency against the upper class boundary for each class interval.

Practice

Draw a cumulative frequency graph for each table in the practice section on the previous page.

EXAMPLE: The number of people queuing to buy petrol was recorded at 10 minute intervals for one hour. The table shows the frequency distribution.

Draw the cumulative frequency graph for the data.

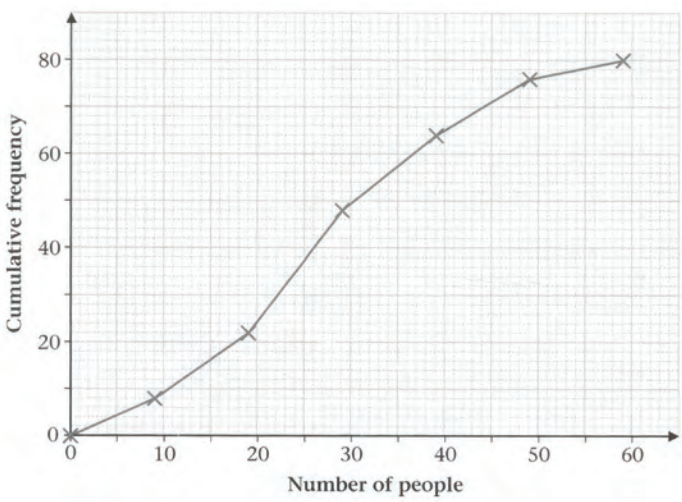
Number in queue	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59
Frequency	8	14	26	16	12	4

To draw the cumulative frequency graph you must first draw a cumulative frequency table. The numbers of people are discrete, so the upper class boundaries are 9, 19, 29, 39, ...

Number of people	**Cumulative Frequency**
$0 \le m \le 9$	8
$0 \le m \le 19$	22
$0 \le m \le 29$	48
$0 \le m \le 39$	64
$0 \le m \le 49$	76
$0 \le m \le 59$	80

We draw the cumulative frequency graph by plotting (9,8), (19,22), (29, 48)...

Number of people queuing for petrol against time



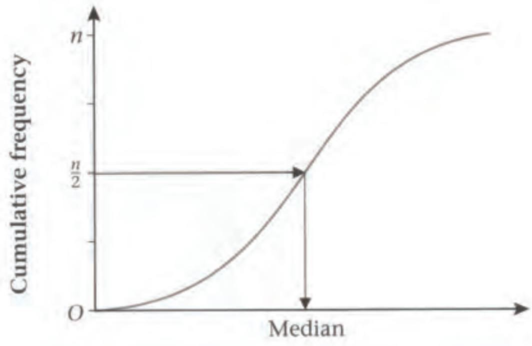
Note that the points are connected using straight lines because the data is discrete.

The cumulative frequency is always plotted on the vertical axis

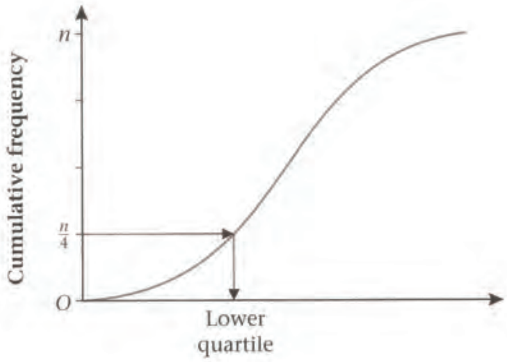
3.8 Spread from cumulative frequency graphs

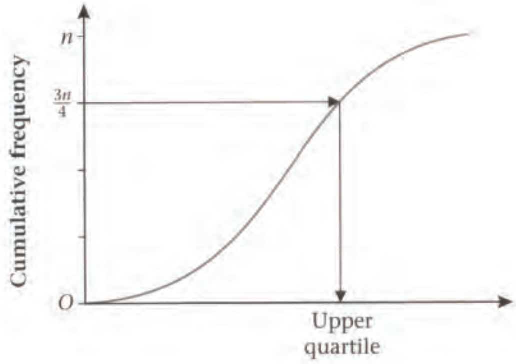
A cumulative frequency graph can be used to estimate the median, the upper quartile and the lower quartile of a distribution. The graphs below demonstrate how to do this.

Estimating the median - To estimate the median from a cumulative frequency graph we find the n/2 th value in the distribution.

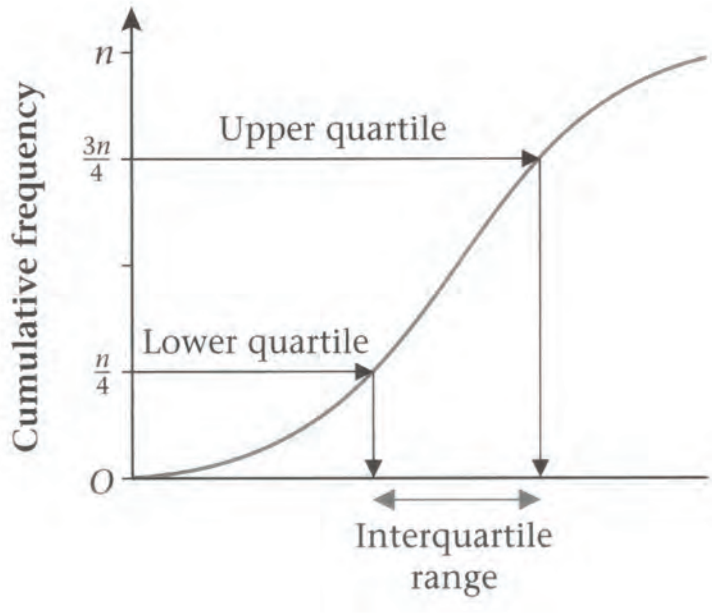


Estimating the quartiles - The lower quartile is the n/4 th value in the distribution. The upper quartile is the 3n/4 th value in the distribution.





The interquartile range - The interquartile range = upper quartile - lower quartile.



EXAMPLE: The table gives information about the time (in seconds) between planes landing at an airport. Find an estimate for:

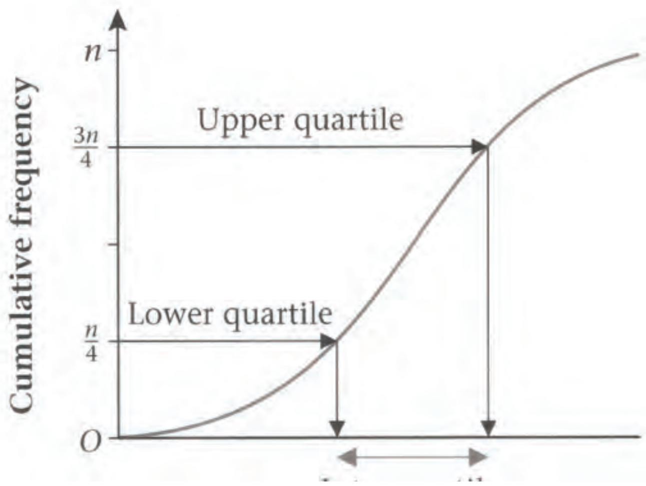
a. Median **b.** The lower and upper quartiles **c.** The interquartile range

First we need the cumulative frequency table and the graph.

Time (seconds)	Cumulative Frequency
60 ≤ t < 100	5
60 ≤ t < 140	19
60 ≤ t < 180	29
60 ≤ t < 220	38
60 ≤ t < 260	45
60 ≤ t < 300	48

Time (seconds)	Frequency
60 ≤ t < 100	5
100 ≤ t < 140	14
140 ≤ t < 180	10
180 ≤ t < 220	9
220 ≤ t < 260	7
260 ≤ t < 300	3

Time taken between planes landing at an airport



We can read the estimates from the graph

a. The median is the 48/2 th value = 24th value = 160 seconds

b. The lower quartile is the 48/4 th value = 12th value = 120 seconds. The upper quartile is the 3(48)/4 th value = 36th value = 212 seconds

c. The interquartile range is upper quartile - lower quartile = 212-120 = 92

Practice

v. For tables 1 and 2.

a. Draw a cumulative frequency table

b. Draw a cumulative frequency graph

c. Find an estimate for the median

d. Find an estimate for the lower and upper quartiles

e. Find an estimate the interquartile range

1

Number of particles	Frequency
0 - 50	10
51 - 100	16
101 - 150	13
151 - 200	11
201 - 250	7
251 - 300	3

2

Age of company employee (years)	Frequency
16 < a ≤ 20	6
21 < a ≤ 25	9
26 < a ≤ 30	14
31 < a ≤ 35	4
36 < a ≤ 40	2
41 < a ≤ 45	1

ii. The table gives information about the body temperatures of a random sample of people. Find:

a. The median

b. The interquartile range

c. The number of people with a body temperature of less than 37 °C

Body temperature (°C)	Frequency
< 36.0	8
< 36.3	23
< 36.6	44
< 36.9	78
< 37.2	101
< 37.5	115
< 37.8	120

4. Probability

4.1 Finding probabilities

Probability is about calculating or estimating what might happen in the future. The probability of something happening, an event, is expressed as a number between 0 and 1.

If an event is impossible its probability is 0.   If an event is certain its probability is 1.

We often use words such as **impossible**, **certain**, **likely, unlikely** to describe probabilities.

Practice

i. Complete the table below by thinking of events that are either impossible, certain or in between. An example is given for each case.

Impossible	In between	Certain
A man will get pregnant	I will eat chocolate today	This lesson will end

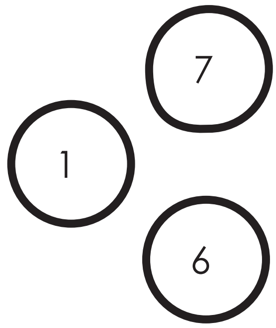
ii. Seven counters numbered 1 to 7 are placed in a bag. One counter is taken out and you have to guess whether the next counter will be higher or lower.

In each case below use one of the words *certain, likely, unlikely, impossible* to complete the sentence.

a. If the first counter taken is 7, then the probability that the second counter is lower than 7 is _______________.

b. If the second counter taken is 1, then the probability that the third counter is lower than 2 is _______________.

c. If the third counter taken is 6, then the probability that the fourth counter is greater than or equal to 5 is _______________.



If we want to be more accurate than using words, we can calculate probabilities using the formula:

$$P(\text{event}) = \frac{\text{the number of ways the event can occur}}{\text{the total number of possible outcomes}}$$

P(event) means the probability of an event. The answer can be written as a fraction, decimal or a percentage.

EXAMPLE: A normal die is numbered 1 to 6. If we throw the die:

a. What is the total number of possible outcomes?
b. What is the chance of throwing a 6?
c. What is the chance of throwing an odd number?

Answer:

a. The possible outcomes are 1, 2, 3, 4, 5, 6. So, the total number of possible outcomes is 6.
b. There is only one 6 on a die so number of ways of throwing a six 1. Using the formula gives:

$$P(\text{throwing a six}) = \frac{\text{the number of ways the event can occur}}{\text{the total number of possible outcomes}} = \frac{1}{6}$$

c. There are 3 ways of getting an odd number - 1, 3, 5. Using the formula gives:

$$P(\text{throwing an odd number}) = \frac{\text{the number of ways the event can occur}}{\text{the total number of possible outcomes}} = \frac{3}{6}$$

Practice

i. A normal die is rolled. What is the probability of throwing:

a. 5?
b. an even number?
c. a multiple of 3?
d. not 6?

ii. Five strawberry, two orange and three blackcurrant flavoured sweets are placed in a box. A sweet is taken from the box. Find the probability that the sweet is:

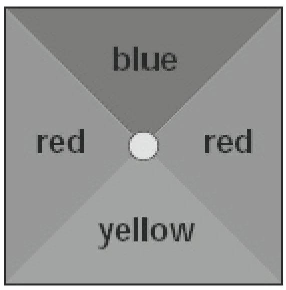
a. blackcurrant flavoured
b. not orange flavoured

iii. A normal pack of cards contains 52 cards. There are 4 suits - hearts, spades, diamonds and clubs. The hearts and diamonds are red. The spades and clubs are black. Each suit contains 13 cards - an ace, a king, a queen and a jack and nine cards labelled 2 to 10. This is shown in the table:

Black	Spades	King	Queen	Jack	10	9	8	7	6	5	4	3	2	Ace
	Clubs	King	Queen	Jack	10	9	8	7	6	5	4	3	2	Ace
Red	Diamonds	King	Queen	Jack	10	9	8	7	6	5	4	3	2	Ace
	Hearts	King	Queen	Jack	10	9	8	7	6	5	4	3	2	Ace

Use the table to answer the questions. If you take a card from a pack what is the probability that:

a. The card will be red?
b. The card will be black?
c. The card will be a heart?
d. The card will be a jack?
e. The card will be a black king?



iv. Some students play a game with the spinner shown.

a. How many outcomes are there? (Hint: the number of outcomes is not the same as the number of squares.)

b. Which colour should I choose if I want to win?

c. Use probability to explain your answers to b?

v. Complete the table below to show the probability of the different events in fractions, decimals and percentages. An example is given for you.

Event	**Probability**		
	Fraction	Decimal	Percentage
A newborn baby is a boy	1/2	0.5	50 %
Rolling a die and getting an even number			
Spinning the spinner in iv. and getting blue			
Pulling a red card from a pack of cards			

4.2 More than one event

In section 4.1 we only considered single event probabilities. Sometimes we would like to know the outcome of more than one event.

EXAMPLE: A card is draw from a normal pack of cards. What is the probability that the card is a 'king' or a 'ten'?

There are 4 kings in a pack of 52, so P(king) = 4/52.

There are also 4 tens so P(ten) = 4/52.

To find P(king or ten) we add the two probabilities; P(king or ten) = 4/52 + 4/52 =8/52.

Note: In this example choosing a king and choosing a ten cannot happen at the same time. We say they are mutually exclusive. For two mutually exclusive events A, B: P(A or B) = P(A) + P(B).

EXAMPLE: There are 5 red, 3 green and 2 yellow counters in a bag. A counter is taken from the bag. Calculate the probability that the counter will be:

a. red **b.** green **c.** yellow **d.** red or green **e.** not yellow

a. P(red) = 5/10 because there are 5 red counters and 10 counters in total

d. P(green) = 3/10

e. P(yellow) = 2/10

f. The two events are mutually exclusive so

P(red or green) = P(red) + P(green) = 5/10 + 3/10 = 8/10

g. The probability of getting red, green or yellow is 1 because it is a certain event:

P(red) + P(green) + P(yellow) = 1

So, P(red or green) = 1 - P(yellow).

Now, P(red or green) is the same as P(not yellow) so:

P(not yellow) = 1 - P(yellow) = 1 - 2/10 = 8/10

Think

Look at the previous example:

a. Explain why P(green) = 3/10
b. Explain why the events P(green) and P(red) are mutually exclusive
c. Explain why the event P(red or green) is the same as P(not yellow)
d. Explain why the equation P(not yellow) = 1 - P(yellow) is true

Practice

A card is taken from a normal pack of cards. Find the probability that:

a. The card is an ace or a ten
b. The card is black or red
c. The card is an ace or a ten or a nine
d. The card is a black king or a red jack

If there is more than one object used to generate outcomes then we can draw a table to map out all the possible outcomes. The table is called the **sample space**.

EXAMPLE: Two normal dice are rolled and the numbers shown added together. Calculate the probability that the sum will be:

a. 10 **b.** a multiple of 5 **c.** not 7

The sample space can be drawn like this:

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The table tells us that the the total number of possible outcomes is 36 because the total number of squares is 36.

a. We can see in the table that there are 3 ways to make 10 from rolling 2 dice: 6 and 4, 5 and 5, 4 and 6. Using the formula we have

$$P(10) = \frac{\text{the number of ways the event can occur}}{\text{the total number of possible outcomes}} = \frac{3}{36}$$

b. The only numbers in the table that are a multiple of 5 are 5 and 10. There are 4 ways to make 5 so P(5) = 4/36. We know that the probability of making 10 is P(10) = 3/36. So,

P(multiple of 5) = P(5) + P(10) = 4/36 + 3/36 = 7/36

c. The table shows that there are 6 ways to make 7 when rolling two dice, so P(7) = 6/36.

P(not 7) = 1 - P(7) = 1 - 6/36 = 30/36

Practice

i. The sample space shows the outcomes of the sex of twins. The top is the outcomes for the first twin. The side is the outcomes for the second twin.

G means girl, B means boy.

a. How many outcomes are there?

b. There is one outcome both twins will be girls. Shown by GG. Complete the statement:

P(2 girls) = 0._____

c. P(2 boys) =

d. P(1 girl and 1 boy) =

e. Complete the sentence:

A woman is more likely to have _______________ than ________________ or _________________.

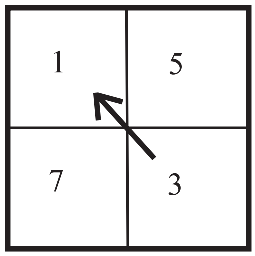
f. What is the probability that a woman will have twins of the same sex?

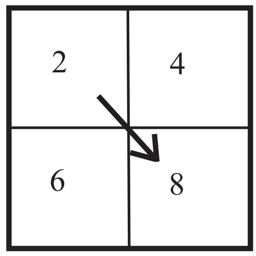
	sex of first twin	
sex of second twin	G	B
G	GG	BG
B	GB	BB

ii. A game is played using the two spinners shown. In one 'turn' a player spins both spinners. The sum of the two numbers is the score for that 'turn'. The score shown is 9.

a. Complete the table showing all possible scores for one 'turn'

+	1	3	5	7
2		5	7	
4	5			
6		9	11	
8	9		13	





b. Write the probability that a player will score 11 in one 'turn'.

c. Write the probability that a player will score more than 10.

d. Write the probability that a player will score a prime number.

e. Write the probability that a player will score a multiple of 3.

4.3 Tree diagrams

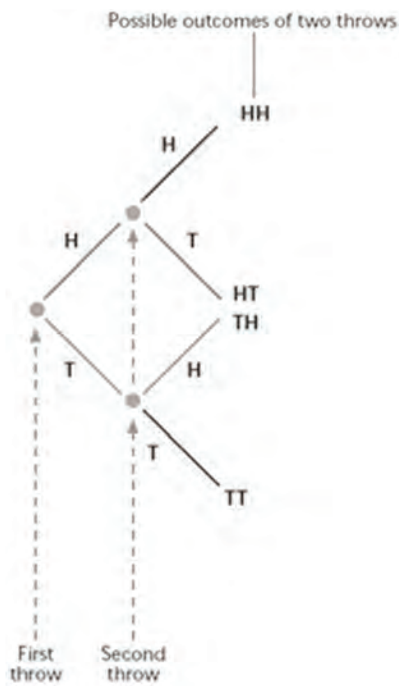
EXAMPLE: A British one pound coin has two sides - head and tail. If we flip the coin there are two possible outcomes - head or tail.

The outcome of flipping coin can be represented on a tree diagram. The one here shows the outcomes of flipping a coin twice.

The dots show the event of flipping a coin.

If we follow the line along the top then we see that the first flip is a head (H) and the second flip is also a head (H).

The outcome is HH, 2 heads.



Practice

i. Look at the tree diagram in the example and answer the questions.

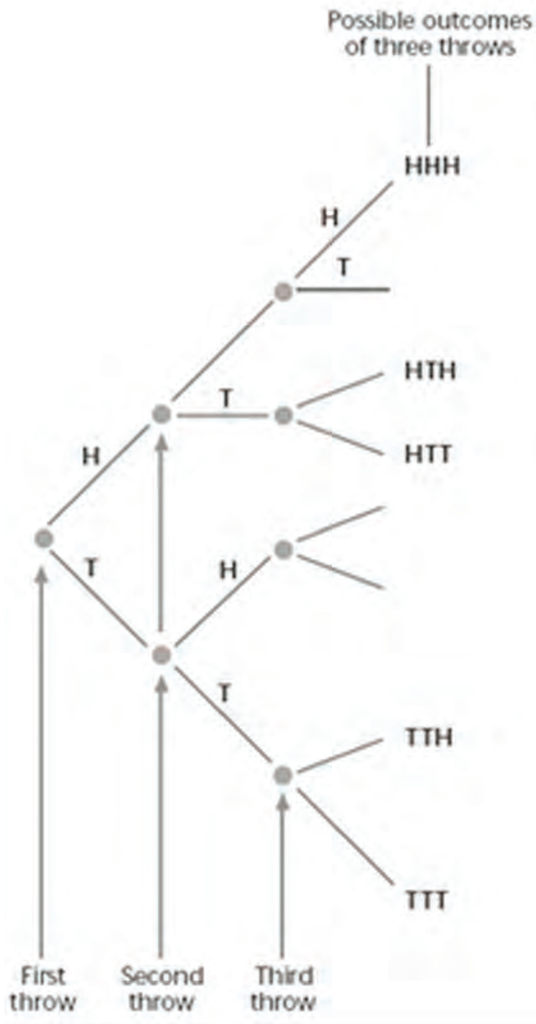
a. How many possible outcomes are there of flipping a coin twice?
b. What is the probability of getting two heads? Write the answer as a fraction.
c. What is the probability of getting two tails? Write the answer as a fraction.
d. What is the probability of getting one head and one tail? Write the answer as a fraction.

ii. Draw a sample space to show the possible outcomes of flipping a coin twice.

iii. Draw a probability tree to show the possible outcomes of the sex of twins.

iv. The probability tree to the right shows all the possible outcomes of flipping a coin three times.

a. Complete the tree by filling in the missing spaces.
b. How many possible outcomes of flipping a coin three times are there?
c. How many ways can I get three heads?
d. What is the probability of getting three heads?
e. What is the probability of getting two tails and one head?
f. What is the probability of getting one tail and two heads?
g. What is the probability of getting one head with one flip of a coin?
h. What is the probability of getting two heads with two flips of a coin?
i. What is the probability of getting three heads with three flips of a coin?
j. Look at your answers to g, h and i. Can you see a pattern? What do you think is the probability of getting four heads with fours flips of a coin?



4.4 Dependent and independent events

Imagine a bag contains eight balls. Five are red and three are not red. If we take out two balls, *what is the probability that both balls will be red?*

The answer to this question depends on whether we replace the first ball or not.

If we replace the first ball the number of balls the second time will be the same as the first time. The outcome of choosing the second ball does not depend on the outcome of the first pick.

The two events are independent.

If we do not replace the first ball then the number of balls on the second pick will be different. The second pick depends on the result of the first pick.

The two events are dependent.

The two tree diagrams on the next page show the probability of getting two reds for the two situations:

Independent events - first ball is replaced before the second one is picked.

The probability of getting two red balls is:

P(red) x P(red) = P(red and red) =

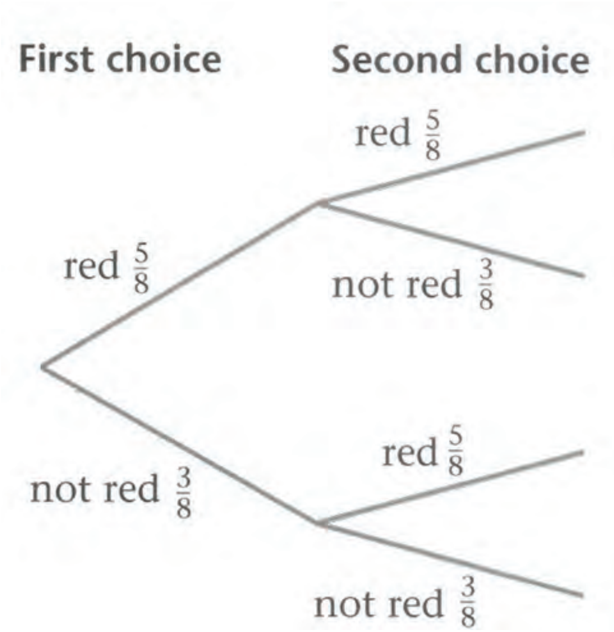
$$\frac{5}{8} \times \frac{5}{8} = \frac{25}{64}$$

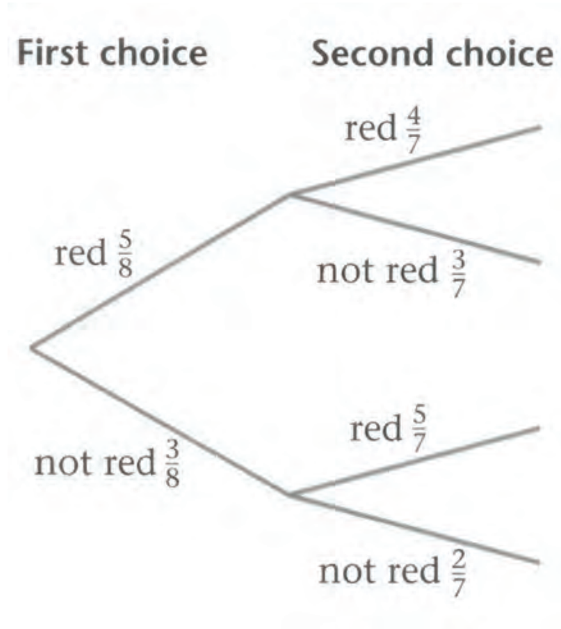
Dependent events - first ball is not replaced before the second one is picked.

The probability of getting two red balls is:

P(red) x P(red) = P(red and red) =

$$\frac{5}{8} \times \frac{4}{7} = \frac{20}{56} = \frac{5}{14}$$





Practice

i. Look back at the example on the previous page. What is the probability of making two picks and getting no reds if:

- **a.** The balls are replaced after every pick
- **b.** The balls are not replaced after they are picked

ii. Imagine a bag containing 10 balls. Six balls are red and four are green. A ball is taken from the bag. A second ball is taken without replacing the first.

- **a.** Draw a tree diagram to show all the possible outcomes of these events.
- **b.** Calculate the probability that both balls will be red.

iii. The CEC college cat has a litter of kittens: five female and two male. The school decides to give the kittens to people in the community. People come and choose the cats at random. Draw a tree diagram and use it to calculate the probability that the first three kittens chosen will be:

- **a.** all male
- **b.** all female
- **c.** will include two male kittens

Glossary of Keywords

Here is a list of mathematical words from this module. The section where the word appears is given in brackets. Find the words and what they mean - your teacher will test your memory soon!

Qualitative	(1.1)
Quantitative	(1.1)
Discrete	(1.1)
Continuous	(1.1)
Variable	(1.1)
Survey	(1.2)
Population	(1.2)
Census	(1.2)
Sample	(1.2)
Primary data	(1.3)
Secondary data	(1.3)
Source	(1.3)
Questionnaire	(1.4)
Interview	(1.4)
Observation	(1.4)
Experiment	(1.4)
Table	(1.5)
Tally	(1.5)
Frequency	(1.5)
Frequency distribution	(1.5)
Class interval	(1.5)
Average	(2.1)
Mean	(2.1)
Mode	(2.1)
Median	(2.1)
Quartiles	(2.3)
Lower quartile	(2.3)
Upper quartile	(2.3)
Life expectancy	(2.3)
Range	(2.4)
Interquartile range	(2.4)
Scatter diagram	(2.8)
Correlation	(2.8)
Positive correlation	(2.8)
Negative correlation	(2.8)
Diagrams	(3.1)
Pie charts	(3.1)
Bar graph	(3.1)
Histogram	(3.1)
Cumulative frequency	(3.1)
Cumulative frequency graph	(3.1)
Multiple bar graph	(3.4)
Probability	(4.1)
Event	(4.1)
Impossible	(4.1)
Certain	(4.1)
Likely	(4.1)
Unlikely	(4.1)
Sample space	(4.2)
Tree diagrams	(4.3)
Dependent event	(4.4)
Independent event	(4.4)

Assessment

This assessment is written to test your understanding of the module. Review the work you have done before taking the test. Good luck!

Part 1 - Vocabulary

These questions test your knowledge of the keywords from this module.
Complete the gaps in each sentence by using the words in the box. Be careful, there are 20 words but only 15 questions.

discrete	continuous	secondary	certain	probability
mode	scatter diagram	median	independent	correlation

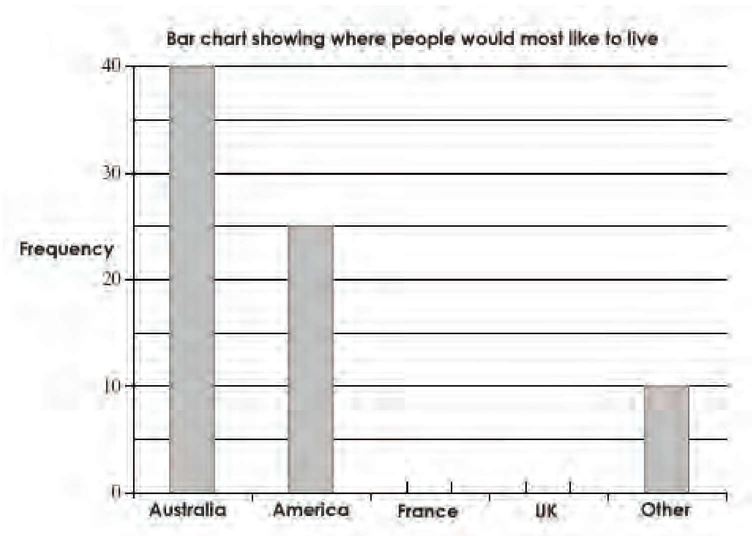
a. Data which can take any value is called ____________________ data.
b. If we write a report using data we found on the internet then this data is called __________________ data.
c. The ____________________ is the middle value of a data set when the data is arranged in order.
d. The value which occurs most often in a data set is the ________________ .
e. We can see if two sets of data are related by drawing a ___________________ .
f. The relationship between two sets of data is called a ___________________ .
g. Pie charts are used to present __________________ data.
h. If an event is impossible then the ___________________ is 0.
i. If an event is _________________ then the probability is 1.
j. If the outcome of an event is not affected by previous events then these events are ___________________ .

Part 2 - Mathematics

These questions test your understanding of the Mathematics in this module. Try to answer all the questions. Write your calculations and answers on separate paper.

1. 100 people were asked where they would most like to live. The results are shown in the bar chart.

a. How many people said they would like to live in Australia?
b. How many people said they would like to live in America?
c. Five out of the 100 people said France. How many people said the UK?
d. Complete the bar chart.



2. Below are some probability values

0.75 1 0.5 0 0.25

Match the values with the words in the table

Word	Probability
Impossible	
Likely	
Certain	
Unlikely	
Even chance	

3. Chit Aung went fishing six times. The number of fish he caught was:

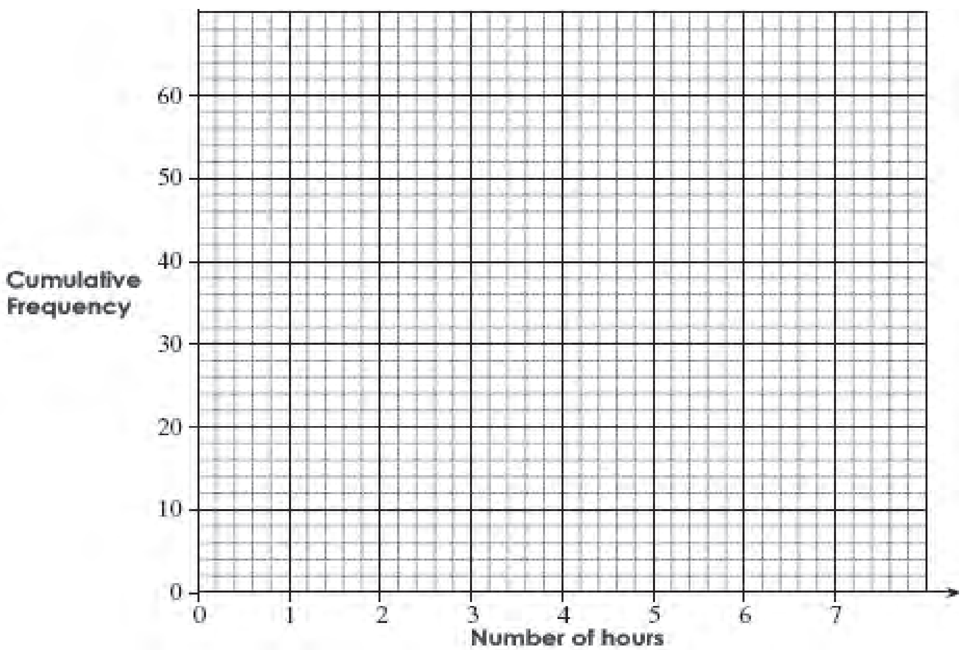
3 5 6 8 11 x

a. The range of the number of fish caught is 10. Work out x, the largest number
b. Find the median number of fish caught
c. Find the mean number of fish caught
d. Do these data have a mode? Give a reason for your answer.

4. Kyaw Kyaw is a teacher. The table shows the cumulative frequency of the number of hours extra he has worked each week for the past 60 weeks.

Number of hours	Cumulative Frequency
less than 1	2
less than 2	7
less than 3	10
less than 4	22
less than 5	41
less than 6	58
less than 7	60

Draw a cumulative frequency polygon using the axes below



5. A six-sided die is rolled. The die has the numbers 1, 2, 3, 4, 5, 6 on its faces.

a. Find the probability that the die shows a 4

b. Find the probability that the die shows an even number.

A second fair six-sided die is also rolled. This die has numbers 0, 1, 2, 3, 4, 5 on its faces.

c. The numbers obtained from the two dice are multiplied to give a score. Complete the table to show all the possible scores

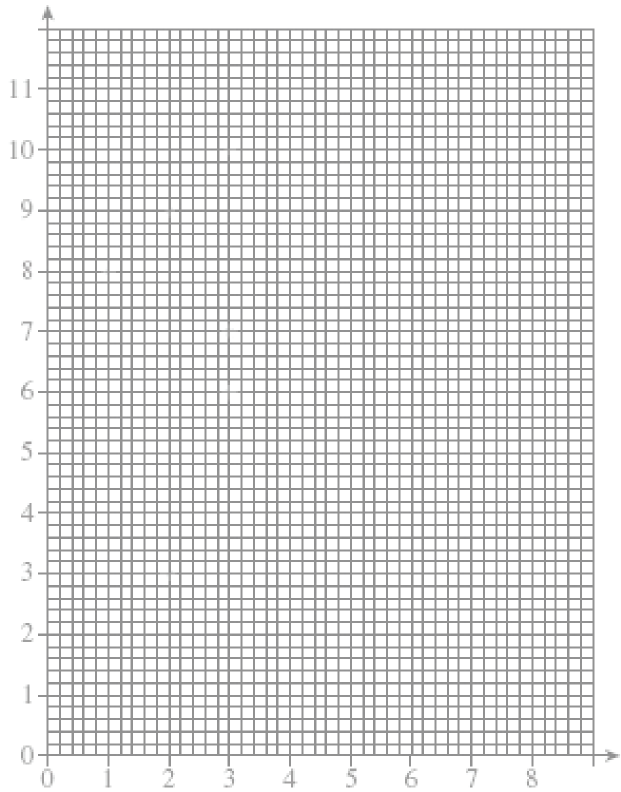
	1	**2**	**3**	**4**	**5**	**6**
0					0	
1			3			
2		4				
3	3					
4						24
5				20		

d. Explain why the probability of getting 15 or more is 1/4.

6. The table shows the number of people in 12 households and the number of letters sent to each household in a month.

Number in household	1	2	2	3	3	3	4	5	5	5	7	8
Number of letters	8	3	9	7	6	10	3	9	8	11	4	7

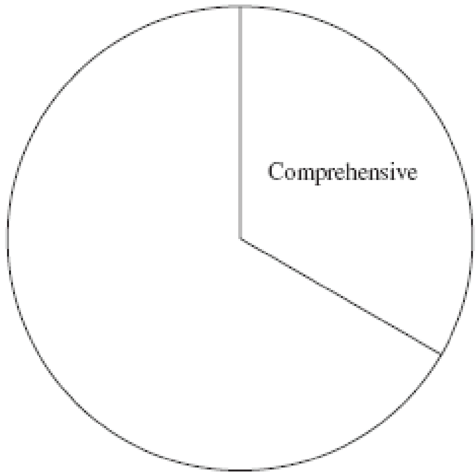
a. Complete the scatter diagram below using the data



b. Is there a correlation between the two sets of data? Explain.

7. The table shows the number of schools by type of school in a city in the UK.

Type of school	Number	Angle on Pie Chart
Primary	63	
Comprehensive	45	120
Grammar		48
Others		
Total	135	360



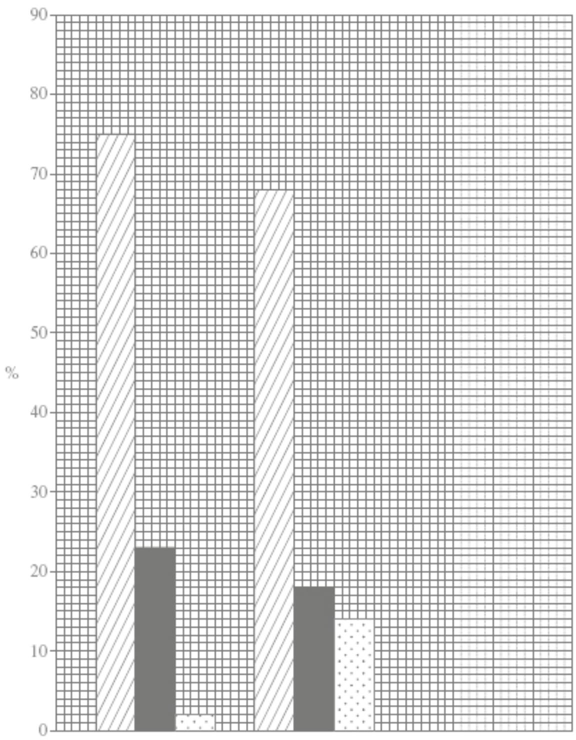
a. Complete the table

b. Use the table to complete the pie chart.

8. The table shows length of time people in an office have been doing their current jobs in 3 companies.

	Less than 6 months (%)	6 to 12 months (%)	Longer than 12 months (%)
Yadana Inc.	75	23	2
Shwe Taw	68	18	14
Taungyi Ltd.	80	17	3

Complete the bar graph shown below. Label the axes and give the graph a title.



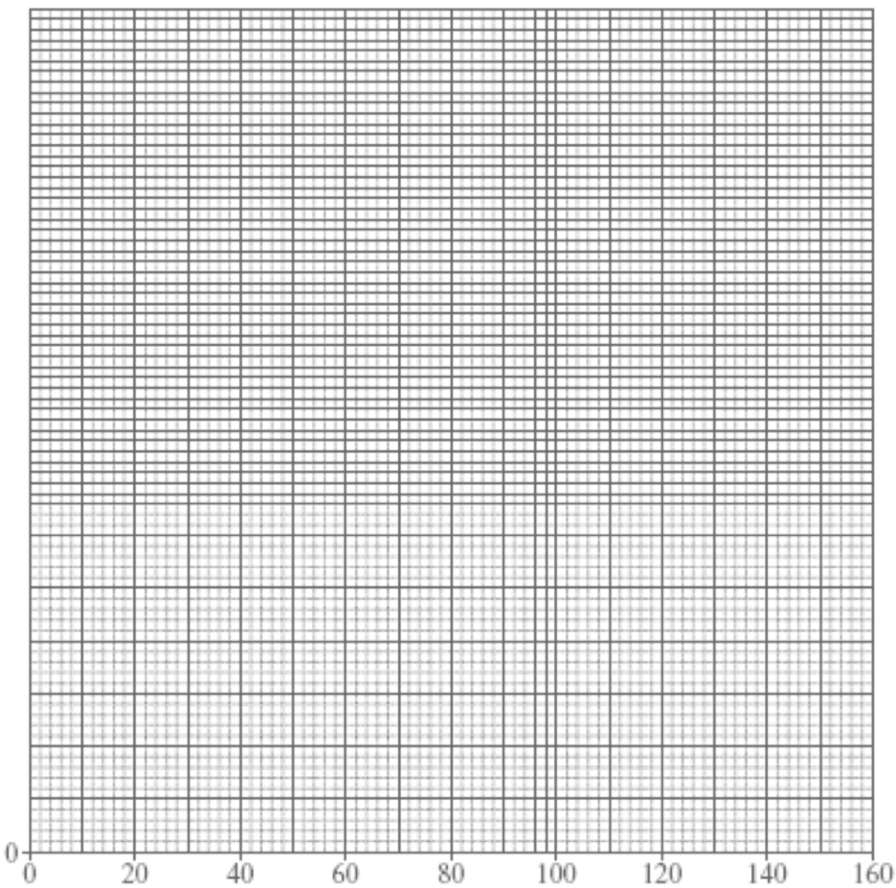
9. The table below shows the length of time people waited in a health clinic to see a doctor.

Time, t (minutes)	Frequency
$0 < t \le 20$	6
$20 < t \le 40$	18
$40 < t \le 60$	30
$60 < t \le 80$	9
$80 < t \le 100$	12

a. Use the table to calculate the mean value of the data

Time	Frequency (f)	Middle value (x)	fx
$0 < t \le 20$	6		
$20 < t \le 40$	18		
$40 < t \le 60$	30		
$60 < t \le 80$	9		
$80 < t \le 100$	12		
Total (Σf)		**Total (Σfx)**	

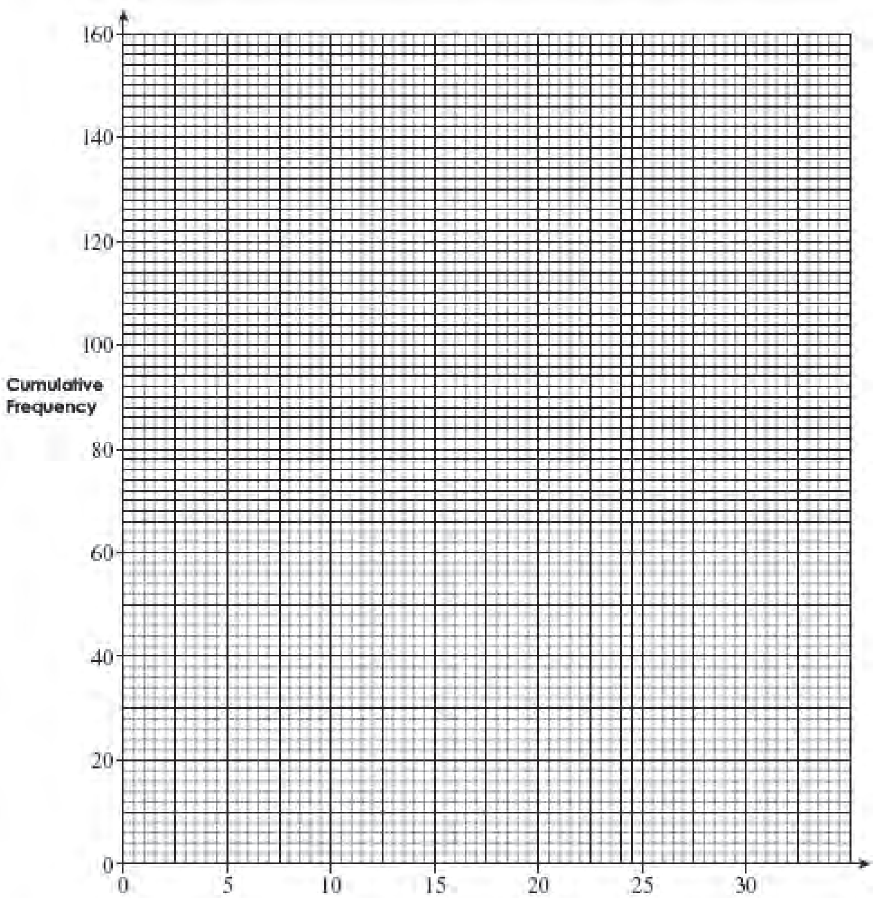
b. Draw a histogram of the data.



10. The table below shows the distance travelled by 160 people to return to their villages.

Distance, m, (miles)	Frequency
$0 < m \leq 5$	10
$5 < m \leq 8$	29
$8 < m \leq 12$	29
$12 < m \leq 16$	44
$16 < m \leq 21$	27
$21 < m \leq 30$	21

a. Draw a cumulative frequency polygon of the data.

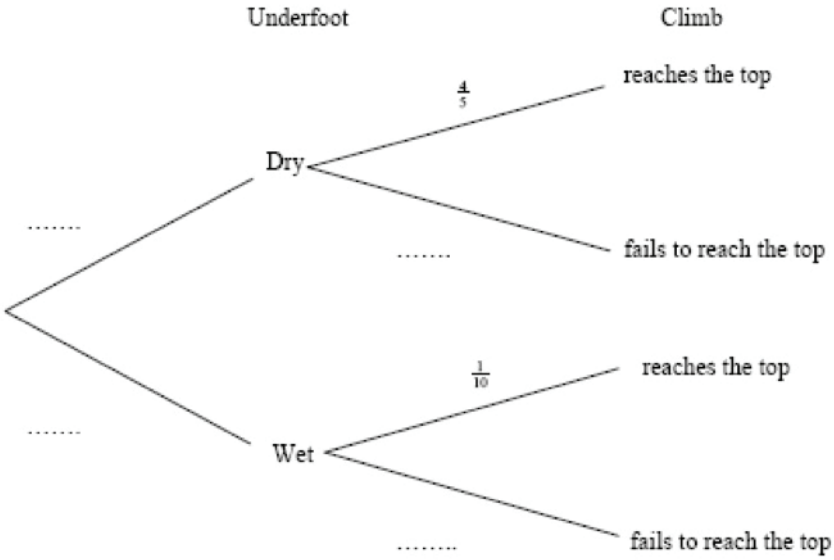


b. Use your graph to estimate the median distance

c. Use your graph to find the interquartile range of the data

11. The probability that a villager reaches the top of a mountain on a dry day is 4/5. On a rainy day the probability is 1/10. The probability that it will rain is 1/4.

a. Complete the tree diagram



b. Find the probability that a climber makes it to the top of the mountain on a randomly chosen day.

12. The bar chart shows the number of goals scored by position by Yadanarbon FC. Use the bar chart to draw a pie chart of the data.