8 Statistical Diagrams

8.1 Data Collection, Extraction and Presentation

In this section we look at collecting and presenting data. We also look at extracting data from tables and timetables.

Example 1

The table gives the distances, in miles, between some towns and cities. Use this table to answer the questions below.

DERBY	DONCASTER	HULL	LEICESTER	LINCOLN	NOTTINGHAM	SHEFFIELD
54						
88	42					
28	69	87				
52	39	37	51			
16	44	73	25	36		
37	18	61	62	46	37	

(a) How far is it from Hull to Nottingham?

(b) Andy drives from Leicester to Lincoln and then on from Lincoln to Doncaster. How far does he drive altogether?

(c) Ian drives from Hull to Sheffield and then back to Hull. How many miles, in total, does he drive?

Solution

(a) From the table, Hull to Nottingham is 73 miles.

(b) From the table, Leicester to Lincoln is 51 miles
and Lincoln to Doncaster is 39 miles.

Total distance $= 51 + 39$

$= 90$ miles

(c) From the table, Hull to Sheffield is 61 miles.

Total distance $= 61 \times 2$

$= 122$ miles

Example 2

Use the following timetable to answer the questions below.

SATURDAY													
Letchworth	d	.....	1955	.....	.....	1955	2030	.....	2055	.....	.....	2055	2130
Hitchin	d	.....	2004	.....	.....	2004	2034	.....	2104	.....	.....	2104	2134
Stevenage	d	.....	2009	.....	.....	2030	2039	.....	2109	.....	.....	2130	2139
Watton-at-Stone	d	.....	.....	.....	.....	2037	.....	.....	.....	.....	.....	2137	.....
Hertford North	d	.....	.....	.....	.....	2043	.....	.....	.....	2113	.....	2143	.....
Bayford	d	.....	.....	2013	.....	2047	.....	.....	.....	2117	.....	2147	.....
Cuffley	d	.....	.....	2017	.....	2052	.....	.....	.....	2122	.....	2152	.....
Crews Hill	d	.....	.....	2022	.....	2055	.....	.....	.....	2125	.....	2155	.....
Gordon Hill	d	.....	.....	2025	.....	2058	.....	.....	.....	2128	.....	2158	.....
Enfield Chase	d	.....	.....	2028	.....	2100	.....	.....	.....	2130	.....	2200	.....
Grange Park	d	.....	.....	2030	.....	2102	.....	.....	.....	2132	.....	2202	.....
Winchmore Hill	d	.....	.....	2032	.....	2103	.....	.....	.....	2133	.....	2203	.....
Palmers Green	d	.....	.....	2036	.....	2106	.....	.....	.....	2136	.....	2206	.....
Bowes Park	d	.....	.....	2038	.....	2108	.....	.....	.....	2138	.....	2208	.....
Knebworth	d	.....	2013	.....	.....	.....	2043	.....	2113	.....	.....	.....	2143
Welwyn North	d	.....	2017	.....	.....	.....	2047	.....	2117	.....	.....	.....	2147
Welwyn Garden City	d	1958	2020	.....	2028	.....	2050	2058	2120	.....	2128	.....	2150
Hatfield	d	2002	2023	.....	2032	.....	2053	2102	2123	.....	2132	.....	2153
Welham Green	d	2006	.....	.....	2036	.....	.....	2106	.....	.....	2136	.....	.....
Brookmans Park	d	2008	.....	.....	2038	.....	.....	2108	.....	.....	2138	.....	.....
Potters Bar	d	2011	2029	.....	2041	.....	2059	2111	2129	.....	2141	.....	2159
Hadley Wood	d	2015	.....	.....	2045	.....	.....	2115	.....	.....	2145	.....	.....
New Barnet	d	2017	.....	.....	2047	.....	.....	2117	.....	.....	2147	.....	.....
Oakleigh Park	d	2019	.....	.....	2049	.....	.....	2119	.....	.....	2149	.....	.....
New Southgate	d	2022	.....	.....	2052	.....	.....	2122	.....	.....	2152	.....	.....
Alexandra Palace	d	2025	.....	2040	2055	2110	.....	2125	.....	2140	2155	2210	.....
Hornsey	d	2027	.....	2042	2057	2112	.....	2127	.....	2142	2157	2212	.....
Harringay	d	2029	.....	2044	2059	2114	.....	2129	.....	2144	2159	2214	.....
Finsbury Park	d	2032	2041	2047	2102	2117	2111	2132	2141	2147	2202	2217	2211
Drayton Park	d	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Highbury and Islington	d	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Essex Road	d	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Old Street	d	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Moorgate	a	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
London Kings Cross	a	.....	2048	2055	2110	2125	2118	2140	2148	2155	2210	2225	2218

(a) Alan catches the 2017 train at Cuffley. When does he arrive at Hornsey?

(b) Julie arrives at Hornsey at 2212. When did she leave Palmers Green?

Solution

(a) Alan arrives at 2042.

(b) Julie left at 2206.

Example 3

A class of pupils take a test. Their scores are listed below:

17	23	46	31	17	19	26	31	42	5
21	32	36	37	37	38	41	40	19	12
7	48	29	39	42	38	41	32	36	35

Draw a *stem and leaf diagram* for this data.

Solution

In this stem and leaf diagram we treat the numbers of 10s as the stem and the numbers of units as the leaves.

In the following stem and leaf plot the data has not been put into order;

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

The leaves can now be ordered as shown to produce the final diagram:

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

Example 4

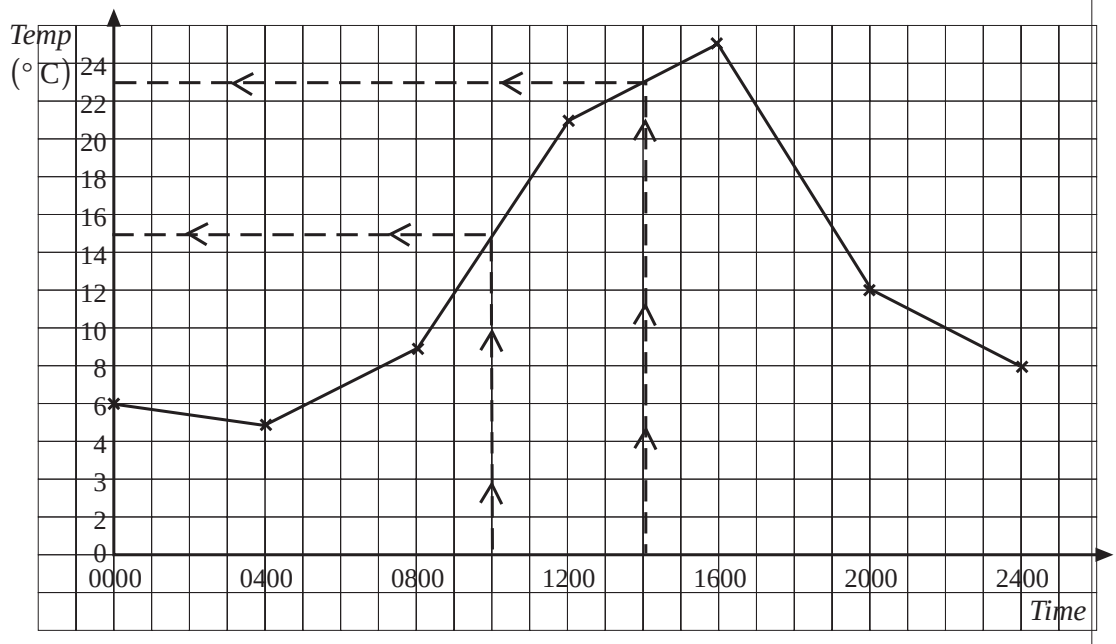
A student records the temperature in a greenhouse every 4 hours during 1 day. The results are listed below:

Time	0000	0400	0800	1200	1600	2000	2400
Temperature (° C)	6	5	9	21	25	12	8

Draw a line graph and use it to estimate the temperature at 1000 and 1400.

Solution

The line graph is shown below:



The dotted lines show how to estimate the temperatures at 1000 and 1400. These estimates are:

15 ° C at 1000

and 23 ° C at 1400.

Example 5

Throughout a 4-week period a class recorded the number of children absent each day. Their results are listed below:

1	0	4	3	1	2	1	3	4	5
7	1	2	2	3	3	1	3	1	0

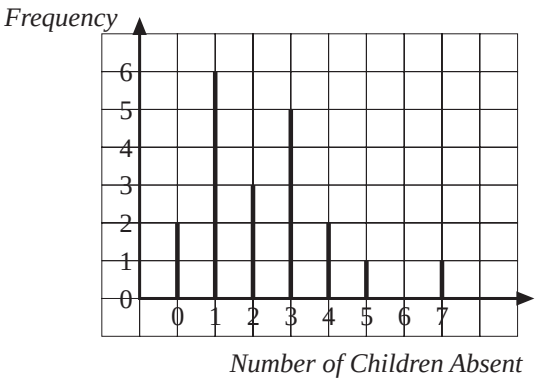
Collate this data using a tally chart and draw a vertical line graph to illustrate the data.

Solution

The tally chart is shown below:

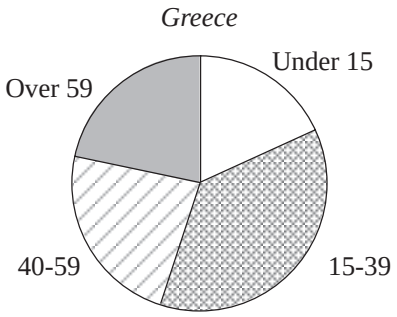
Number of Children Absent	*Tally*	*Frequency*
0	II	2
1	𝍸 I	6
2	III	3
3	𝍸	5
4	II	2
5	I	1
6		0
7	I	1

The vertical line graph is shown below:

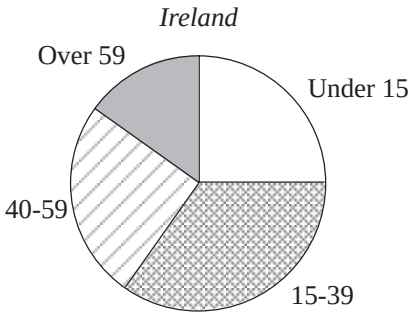


Example 6

These pie charts show some information about the ages of people in Greece and in Ireland. There are about 10 million people in Greece, and there are about 3.5 million people in Ireland.



10 million people



3.5 million people

(a) Roughly what *percentage* of people in Greece are aged 40 - 59 ?

(b) There are about 10 million people in Greece. Use your percentage from part (a) to work out roughly *how many* people in Greece are aged 40 - 59.

(c) Dewi says that these charts show that there are *more* people *under 15* in *Ireland* than in *Greece.*

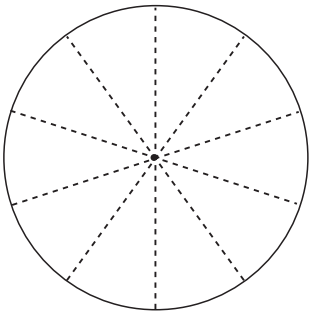
Dewi is wrong. Explain why the charts do *not* show this.

(d) There are about 60 million people in the UK. The table shows roughly what percentage of people in the UK are of different ages.

Under 15	*15-39*	*40-59*	*over 59*
20%	35%	25%	20%

Copy and complete the pie chart below to show the information in the table. Label each section of your pie chart clearly with the *ages.*

UK



60 million people

(KS3/98/Ma/Tier 5-7/P2)

Solution

(a) The angle for 40-59 is about $90\,°$; the fraction of the total is $\frac{90}{360} = \frac{1}{4}$, or 25%.

(b) 25% of Greece's population $= \frac{1}{4} \times 10 \text{ million } = 2.5 \text{ million}$.

(c) This is not true; the percentage of people under 15 is higher in Ireland than in Greece, but Greece has a far larger population than Ireland. The actual numbers are:

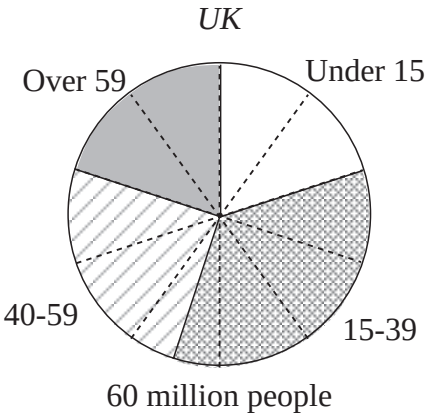
Ireland : $\frac{1}{4} \times 3.5 \text{ million } \approx 0.875 \text{ million}$

Greece : $\frac{60}{360} \times 10 \text{ million } = \frac{1}{6} \times 10 \text{ million} \approx 1.67 \text{ million}$

(d) Since there are 10 equal sectors in the pie chart, each sector is $\frac{360}{10}\,° = 36\,°$, and each sector represents 10% of the people in the UK.

Sectors are:

under 15	*15-39*	*40-59*	*over 59*
2 sectors	$3\frac{1}{2}$ sectors	$2\frac{1}{2}$ sectors	2 sectors



Exercises

1. Use this mileage chart to answer the following questions:

BEDFORD	CAMBRIDGE	HITCHIN	HUNTINGDON	NORTHAMPTON	ROYSTON	WELLINGBOROUGH
29						
16	27					
20	15	30				
21	50	37	43			
24	13	14	21	45		
20	44	36	26	10	44	

   (a) How far is it from Hitchin to Royston?

   (b) Alan drives from Bedford to Royston and then back again. How far does he travel in total?

   (c) David cycles from Bedford to Hitchin, then from Hitchin to Royston and from Royston back to Bedford. How far does he cycle altogether?

   (d) A lorry is driven from Cambridge to Northampton, then from Northampton to Hitchin and from Hitchin back to Cambridge. How far does the lorry travel altogether?

   (e) Is the journey from Cambridge to Northampton *shorter* than the journey from Cambridge to Wellingborough?

2. The table gives the distances, in kilometres, between some European cities. Use the table to answer the following questions:

BRUSSELS	EINDHOVEN	LUXEMBOURG	PARIS	ROTTERDAM	TROYES
136					
212	282				
307	435	376			
182	113	354	452		
401	537	296	173	552	

   (a) Jai drives from Paris to Eindhoven and then drives back to Paris. How far does he travel?

   (b) Harry leaves Rotterdam and travels to Brussels and on to Troyes before returning to Rotterdam. How far does he travel altogether?

   (c) Andrea leaves Paris, drives to Troyes and from there on to Brussels. How far does she travel?

   (d) A driver has to travel from Luxembourg to Brussels, calling at Eindhoven and Rotterdam on the way. Calculate the shortest length of this journey.

3. Use the following timetable to answer these questions:

   (a) Which train should you catch at Birmingham Moor Street to arrive in Wilmcote before 1700 ?

   (b) Nick arrives in Yardley Wood at 1603. At what time did he leave Jewellery Quarter?

   (c) Ali leaves Spring Road at 1534. At what time will he arrive at Earlswood?

(d) Michaela arrives at Hall Green station at 1445. She wants to travel to Henley-in-Arden. What is the earliest time that she could arrive there?

(e) Denise wants to travel from Bordesley to Yardley Wood. At what time must she leave Bordesley?

(f) Johnny wants to travel from Small Heath to Earlswood. He arrives at Small Heath station at 1500. Describe how he can get to Earlswood.

Jewellery Quarter	1350	1420	1440	1450	1520	1540	1550	1620	1640	1650	1709
Birmingham Snow Hill	1405	1425	1445	1505	1525	1545	1605	1625	1645	1704	1716
Birmingham Moor Street	1408	1448	1448	1508	1528	1548	1608	1628	1648	1707	1718
Bordesley	F	F	--	--	--	--	--	--	--	1709	--
Small Heath	1412	--	1452	1512	--	1552	1612	--	1652	1712	--
Tyseley	1414	--	1454	1514	--	1554	1614	1632	1654	1714	--
Spring Road	1417	1434	1457	1517	1534	1557	1617	1635	1657	1717	1725
Hall Green	1420	1437	1500	1520	1537	1600	1620	1638	1700	1723	1728
Yardley Wood	1423	1440	1503	1523	1540	1603	1623	1641	1703	1727	1731
Shirley	1426	1443	1506	1526	1543	1606	1626	1644	1706	--	1734
Whitlocks End	--	1446	--	--	1546	--	--	1647	--	--	1737
Wythall	--	--	--	--	1548	--	--	1649	--	--	1739
Earlswood	--	--	--	--	1551	--	--	1652	--	--	1742
The Lakes	--	--	--	--	1553	--	--	1654	--	--	1744
Wood End	--	--	--	--	1555	--	--	1656	--	--	1746
Danzey	--	--	--	--	1558	--	--	1659	--	--	1749
Henley-in-Arden	--	--	--	--	1603	--	--	1704	--	--	1754
Wootton Wawen	--	--	--	--	1605	--	--	1706	--	--	1756
Wilmcote	--	--	--	--	1611	--	--	1714	--	--	1802
Stratford-upon-Avon	--	--	--	--	1615	--	--	1718	--	--	1807

4. Use the following timetable to answer the questions below:

(a) Jack catches the 0933 train from Manchester. At what time would he arrive in Weston-super-Mare?

(b) Josh wants to travel to Camborne. What is the latest time that hecould leave Manchester Piccadilly?

(c) Kate catches the 1026 at Stafford. At what time will she arrive in Torquay?

(d) Hannah leaves Wolverhampton and arrives in Weston-super-Mare at 1522. At what time did she leave Wolverhampton?

(e) Matthew leaves Taunton at 1405. At what time does he arrive in Penzance?

(f) Serena catches the 1641 at St Austell. At what time does she arrive in St Erth?

Aberdeen	d							
Dundee	d							0620
Edinburgh	d					**0710**		**0815**
Haymarket	d					--		--
Glasgow Central	d					--	0820	--
Motherwell	d					--	0836	--
Lockerbie	d					--	--	--
Carlisle	d					--	0943	--
Penrith	d					--	1000	--
Oxenholme Lake District	d					--	1027	--
Lancaster	d					--	1045	--
Preston	d					--	1109	--
Wigan North Western	d					--	1121	--
Warrington Bank Quay	d					--	1133	--
Liverpool Lime Street	d		0910		0949	--	1049	--
Runcorn	d		0927		1005	--	1105	--
Hartford	d		0940		1017	--	1119	--
Bolton	d		0823		0936	--	1036	--
Manchester Piccadilly	d		**0917**	**0933**	**1017**	--	**1117**	--
Stockport	d		0905	--	1005	--	1105	--
Wilmslow	d		**0936**	**0949**	**1036**	--	**1136**	--
Crewe	d		0955	**1013**	1050	--	1157	--
Macclesfield	d		1914	--	1021	--	1116	--
Congleton	d		--	--	1030	--	--	--
Stoke-on-Trent	d		0936	--	1043	--	1138	--
Stafford	d		**1026**	--	**1119**	--	**1225**	--
Wolverhampton	d		**1042**	--	**1135**	1152	**1244**	1252
Birmingham New Street	a		**1100**	--	**1151**	**1225**	**1303**	**1321**
Birmingham New Street	d	0948	**1111**	--	**1202**	**1230**	**1307**	**1330**
Cheltenham Spa	a	**1110**	**1151**	--	**1242**	**1310**	**1355**	--
Gloucester	a	**1118**	1224	--	1312	**1318**		--
Bristol Parkway	a	**1152**	**1231**	--	--	**1352**	**1430**	--
Bristol Temple Meads	a	**1204**	**1243**	**1254**	**1328**	**1404**	**1445**	**1451**
Weston-super-Mare	a	1257	1317	**1320**	1423	**1424**	1522	1556
Taunton	a	**1241**	1319	**1346**	**1405**	**1447**		1630
Tiverton Parkway	a	1310	1349	--	--	**1500**		--
Exeter St Davids	a	**1307**	**1349**	**1414**	**1434**	**1517**		**1547**
Newton Abbot	a	1348	**1423**	**1442**	1511	**1549**		1621
Torquay	a	1425	**1435**	--	1522	1616		1707
Paignton	a	1432	**1445**	--	1530	1623		1714
Totnes	a	1404	1455	**1455**	**1511**	**1603**		1635
Plymouth	a	**1409**	1523	**1523**	**1541**	**1632**		**1645**
Liskeard	a	**1432**	1552	**1551**	**1609**			**1713**
Bodmin Parkway	a	**1445**	1604	**1603**	**1622**			**1726**
Par	a	**1456**	1615	**1615**	1651			**1740**
St Austell	a	**1503**	1623	**1622**	**1641**			1757
Truro	a	**1521**	1641	**1640**	**1700**			1821
Redruth	a	**1533**	1653	**1653**	**1715**			1830
Camborne	a	**1541**	1700	**1700**	--			1839
St Erth	a	**1551**	1711	**1710**	**1730**			1850
Penzance	a	**1605**	1723	**1723**	**1745**			1903

5. As part of a science project, the height of a plant is measured every 3 days. The readings are listed in the following table:

Day	0	3	6	9	12	15	18
Height (cm)	4	6	9	14	16	19	24

(a) Draw a *line graph* to show how the height of the plant varies with time.

(b) Estimate the height of the plant after 14 days.

(c) Estimate the age of the plant when the height was 8 cm.

6. Records were kept of the mass of a baby for the first few days of its life. The information is listed in the table below:

Day	0	2	4	6	8	10	12	14	16
Mass (kg)	3.7	3.6	3.3	3.5	3.7	3.8	4.0	4.2	4.3

(a) Draw a *line graph* to show how the mass of the baby changes.

(b) Use the line graph to estimate the mass on:

(i) day 1, (ii) day 7, (iii) day 15.

7. Jane measured the height of her son, Chris, every two years and kept a record of the heights.

Chris' Age	1	3	5	7	9	11
Height (cm)	59	81	102	110	131	156

(a) Draw a *line graph* using this data.

(b) Estimate Chris' height when he was:

(i) 2 years old, (ii) 10 years old.

8. The results of a maths test for one class are listed below:

42 31 29 38 24 17 9 18 28 27
34 35 38 40 40 19 32 39 22 11
11 9 2 17 32 19 22 29 31 33

Illustrate this data using an *ordered stem and leaf diagram* using stems of 0, 10, 20, 30 and 40.

9. The data collected in a survey on the number of children in each family is listed below:

2 3 1 2 1 2 3 1 2 6 1 2 3 3 4 1 5 2 3 2
1 3 1 2 4 5 2 2 2 3 1 1 3 1 1 2 2 3 4 2

(a) Draw up a tally and frequency table for this data.

(b) Illustrate this data using a *pictogram.*

(c) Illustrate this data using a *vertical line graph.*

10. Data was collected on the amount, in pence, that children spent in a tuckshop in one session. This data is illustrated in the following stem and leaf diagram.

Stem	*Leaf*
20	7 7 8 8 8 9 9 9 9
30	0 0 0 0 1 2 2 2 5 5 5 6 6 8 8 9
40	0 0 1 1 1 2 3 3 3 4 4 5 5 5 7 7
50	0 0 0

Use a *vertical line diagram* to illustrate the data.

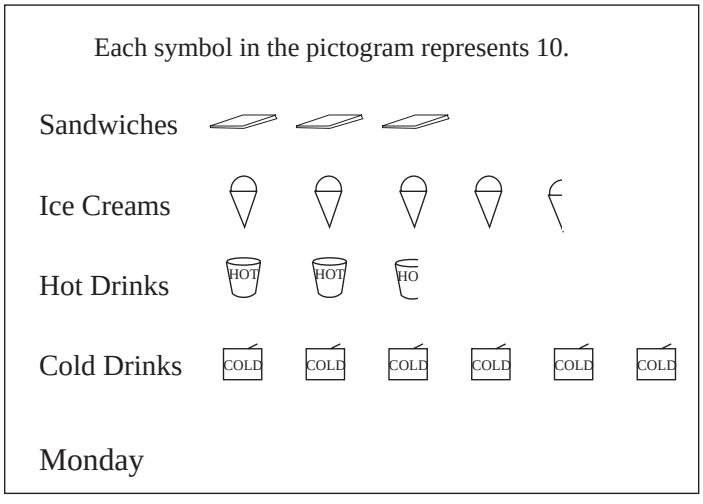
11. A survey into the types of cars in a car park collected data listed below:

F	P	F	B	P	Re	C	M	C	Re	F	V
Ro	Ro	Fi	F	Fi	Fi	B	P	C	Re	M	Re
F	Fi	M	Ro	F	F	F	P	Re	Ro	P	C
M	F	F	Re	Ro	C	Ro	F	M			

Key: F Ford, P Peugeot, B BMW, C Citroen, M Mazda
Fi Fiat, Ro Rover, Re Renault, V Vauxhall

Illustrate this data with a *bar chart.*

12. A small cafe sells sandwiches, ice creams, hot drinks and cold drinks. The pictogram shows what they sold on *Monday.*



(a) How many *cold drinks* did they sell?

(b) How many *ice creams* did they sell?

(c) How many *hot drinks* did they sell?

The pictogram below shows how many sandwiches and ice creams the cafe sold on *Tuesday.*

(d) The cafe also sold *40 hot drinks* on Tuesday. Show this number on a copy of the pictogram below.

Each symbol in the pictogram represents 10.

Sandwiches 

Ice Creams 

Hot Drinks

Cold Drinks

Tuesday

(e) The cafe also sold *12 cold drinks* on Tuesday. Show this number of cold drinks on the pictogram you have drawn.

(f) Look at both the pictograms. What can you tell about the weather on each day?

(KS3/96/Ma/Tier 3-5/P1)

13. Look at this bus timetable, from Highbury to Colton:

Bus Timetable: Highbury to Colton					
Highbury *depart:*	07:45	08:30	09:30	10:45	11:30
Colton *arrive*:	08:30	09:15	10:15	11:30	12:15

(a) A bus leaves Highbury at 08:30.

(i) What time does it arrive in Colton?

(ii) How much time does the bus journey take?

(b) 5 friends are going from Highbury to Colton by bus. They want to arrive by 10:30. Which is the *latest* bus they can catch from Highbury?

(c) Each bus ticket costs £2.20. How much do the 5 bus tickets cost altogether?

(KS3/98/Ma/Tier 3-5/P1)

14. (a) Lisa works in a shoe shop. She recorded the size of each pair of trainers that she sold during a week. This is what she wrote down:

	Sizes of Trainers Sold						
Monday	7	7	5	6			
Tuesday	6	4	4	8			
Wednesday	5	8	6	7	5		
Thursday	7	4	5				
Friday	7	4	9	5	7	8	
Saturday	6	5	7	6	9	4	7

Use a *tallying method* to make a table showing how many pairs of trainers of each size were sold during the whole week.

(b) Which size of trainer did Lisa sell most of?

(c) Lisa said that most of the trainers sold were bigger than size 6. How can you tell from the table that Lisa is *wrong*?

(KS3/95/Ma/Levels 4-6/P1)

15. This chart shows the distances in miles between six towns.

Cardiff					
394	Edinburgh				
198	221	Liverpool			
152	380	203	London		
318	107	157	280	Newcastle	
166	496	300	215	410	Plymouth

Example: Cardiff and London are 152 miles apart.

(a) How far apart are Cardiff and Newcastle?

(b) How far apart are London and Edinburgh?

(c) Which town is 198 miles from Cardiff?

(d) Which two towns are exactly 300 miles apart?

(e) Which town is the greatest distance from Plymouth?

(f) Which town is the smallest distance from Cardiff?

(g) Gwen is a lorry driver. She drove from London to Newcastle, then from Newcastle to Edinburgh. She filled in her job sheet.

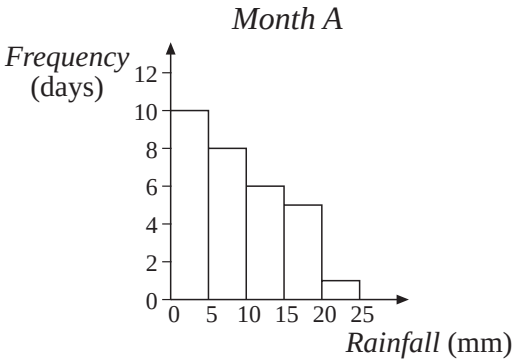
From	*To*	*Distance*
London	Newcastle	280
Newcastle	Edinburgh	107

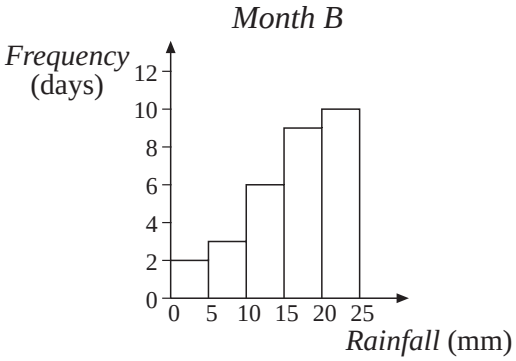
She drove back using the same route. Copy and complete her job sheet.

From	*To*	*Distance*
Edinburgh		

(KS3/94/Ma/3-5/P1)

16. The two frequency diagrams below show the amount of rain that fell in two different months.





(a) Kath says that there are 30 days in month A. Explain how you know she is *right.*

(b) Carl asks 5 friends how much rain fell during month A. They said:

Jon: 5 mm, *Dipta:* 25 mm, *Ian:* 30 mm, *Nerys:* 75 mm, *Sue:* 250 mm

Only *one* friend could have been right. You can tell who it is *without* trying to work out the *total* rainfall.

Which *one* of Carl's friends could have been right? Explain how you know.

(c) Sudi said:

"The diagram for month B shows that it rained more at the end of the month."

Sudi is wrong. Explain why the diagram does *not* show this.

(KS3/94/Ma/5-7/P1)

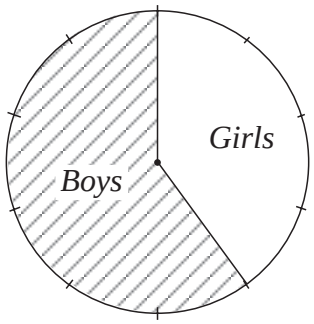
17. There are 50 children altogether in a playgroup.

(a) (i) *How many* of the children are girls?

(ii) What *percentage* of the children are girls?

(b) 25 of the children are 4 years old.
20 of the children are 3 years old.
5 of the children are 2 years old.

Show this information on a pie diagram.



(KS3/97/Ma/Tier 4-6/P2)

8.2 Statistical Measures

In this section we recap the statistical measures *mean, median, mode* and *range.* The *mean, median* and *mode* give an indication of the 'average' value of a set of data, i.e. some idea of a typical value. The *range,* however, provides information on how spread out the data is, i.e. how varied it is.

Definition	*Example*
$\textit{Mean} = \dfrac{\text{sum of all data}}{\text{number of values}}$	For 1, 2, 2, 3, 4 $\textit{Mean} = \dfrac{1+2+2+3+4}{5}$ $= \dfrac{12}{5}$ $= 2.4$
Mode = most common value	For 1, 2, 2, 3, 4 *Mode* = 2 For 1, 2, 2, 3, 4, 4, 5 *Mode* = 2 and 4

Definition	*Example*
Median = middle value when data is arranged in order	For 1, 2, (2,) 3, 4 *Median* = 2 For 1, 2, (2, 3,) 4, 4 $Median = \frac{2+3}{2}$ $= 2.5$
Range = largest value – smallest value	For 1, 2, 2, 3, 4 *Range* = 4 – 1 = 3

In this section, we extend these basic ideas to grouped data.

Example 1

The shoe sizes for a class are summarised in the table shown.

Calculate:

(a) the *mode*, (b) the *median*

(c) the *mean* (d) the *range*

for this data

Shoe Size	*Frequency*
4	2
5	4
6	7
7	5
8	6
9	3
10	3

Solution

(a) The *mode* = 6 (i.e. the size with highest frequency)

(b) There are 30 values altogether. Since 30 is even, there will be two central values. These will be the 15th and 16th values. From the frequency table, these are both 7. (You could list them all in order, but it is easy to see from the table that there are 13 values before the five '7' values are reached.)

So the median $= \frac{7+7}{2} = 7.$

(c) The *mean* is the sum of all the data values divided by the total number of values, and is better calculated from the table by adding an extra

'*frequency* × *size*'

column, as shown in the following table:

(x) *Size*	(f) *Frequency*	(fx) *Frequency* × *Size*
4	2	$2 \times 4 = 8$
5	4	$4 \times 5 = 20$
6	7	$7 \times 6 = 42$
7	5	$5 \times 7 = 35$
8	6	$6 \times 8 = 48$
9	3	$3 \times 9 = 27$
10	3	$3 \times 10 = 30$
Total	30	210

The mean $= \dfrac{\sum fx}{\sum f} = \dfrac{210}{30} = 7$

(d) The *range* = highest value – lowest value

$= 10 - 4$

$= 6$

Note

If a data set contains n values then the median can be obtained as the $\left(\dfrac{n+1}{2}\right)$th value If n is odd, this formula will pick out the value that you need. For example, if there are 157 data values then the median will be the $\left(\dfrac{157+1}{2}\right)$th value, i.e. the 79th value. If n, the number of data values, is even, then the formula will pick out the two values that you need to average to obtain the median. In Example 1, we had $n = 30$ data values, so the median is the $\left(\dfrac{30+1}{2}\right)$th value, i.e. the 15.5th value. The ' .5' tells us we need to average the 15th and 16th values, which is what we did to get the median 7.

Example 2

The table shows the Morse code for 26 letters and how long it takes to send each letter.

Letter	Code	Sending Time	Letter	Code	Sending Time	Letter	Code	Sending Time
A	• —	5	J	— — — •	13	S	• • •	5
B	— • • •	9	K	— • —	9	T	—	3
C	— • — •	11	L	• — • •	9	U	• • —	7
D	— • •	7	M	— —	7	V	• • • —	9
E	•	1	N	— •	5	W	• — —	9
F	• • — •	9	O	— — —	11	X	— • • —	11
G	— — •	9	P	• — — •	11	Y	— • — —	13
H	• • • •	7	Q	— — • —	13	Z	— — •	11
I	• •	3	R	• — •	7			

If a letter is frequent we want to be able to send it quickly. The following table shows the 6 most frequent letters in 4 languages:

	English	*French*	*Italian*	*Spanish*
1	E	E	E	E
2	T	A	O	A
3	A	S	A	O
4	O	I	I	S
5	N	R	N	R
6	I	N	R	I

(a) Copy and complete the following table of the mean, median and modal sending times for the 6 most frequent letters in each language.

	English	*French*	*Italian*	*Spanish*
Mean			5.3	5.3
Median			5	5
Mode	3 and 5	5		

(b) Use your table in part (a) to decide which *two* languages are likely to send the quickest messages in Morse. Explain how you decided.

(c) Samuel Morse invented the code. Messages in his own language are quick to send. Look at the table of the 6 most frequent letters in each language.

Which *one* of these letters has a code which suggests that Samuel Morse's own language was English? Explain how you decided.

(KS3/94/Ma/5-7/P2)

Solution

(a) *English* : mean time $= \dfrac{1+3+5+11+5+3}{6} = \dfrac{28}{6}$

$= 4\dfrac{2}{3} = 4.7$ (to 1 decimal place)

median time: in order 1, 3, 3, 5, 5, 11

median $= \dfrac{3+5}{2} = 4$

French : mean time $= \dfrac{1+5+5+3+7+5}{6} = \dfrac{26}{6}$

$= 4\dfrac{1}{3} = 4.3$ (to 1 decimal place)

median time: in order 1, 3, 5, 5, 5, 7

median $= \dfrac{5+5}{2} = 5$

Italian times are 1, 11, 5, 3, 5, 7, so modal time is 5.

Spanish times are 1, 5, 11, 5, 7, 3, so modal time is 5.

These values can now be put in the table, as below.

	English	*French*	*Italian*	*Spanish*
Mean	4.7	4.3	5.3	5.3
Median	4	5	5	5
Mode	3 and 5	5	5	5

(b) *English* and *French* , since their mean values are significantly lower than the mean values for *Italian* and *Spanish.*

(c) The letter T, which is the 2nd most frequently used letter in English, has a very short sending time, but is not in the top 6 for French, Italian or Spanish.

Note

Often data is provided in summary form, so that estimates have to be made to find the mean value.

Example 3

Data on the number of minutes that a particular train service was late have been summarised in the table. (Times are given to the nearest minute.)

Minutes Late	*Frequency*
on time	19
1-5	12
6-10	9
11-20	4
21-40	4
41-60	2
over 60	0

(a) How many journeys have been included?

(b) What is the modal group?

(c) Estimate the mean number of minutes the train is late for these journeys.

(d) Which of the two averages, *mode* and *mean*, would the train company like to use in advertising its service? Why does this give a false impression of the likelihood of being late?

(e) Estimate the probability of a train being more than 20 minutes late on this service.

Solution

(a) Total number of journeys $= 19 + 12 + 9 + 4 + 4 + 2 = 50$

(b) 'On time'

(c) It is more convenient to use a table for this calculation; for each 'group', the midpoint is used for the calculation (this is why it is an estimate and not an exact value).

Minutes Late	*Midpoint* (x)	*Frequency* (f)	fx
On time	0	19	0
1-5	3	12	36
6-10	8	9	72
11-20	15.5	4	62
21-40	30.5	4	122
41-60	50.5	2	101
	Total	50	393

(Note that, because the times in the table are given to the nearest minute, the class described as '11-20' actually means $10.5 \le T < 20.5$. This class has width 10 minutes, so half way will be 5 minutes after the start point 10.5, so the midpoint $= 10.5 + 5 = 15.5$.)

Mean value $\approx \dfrac{393}{50} \approx 7.86$ minutes

(d) Clearly 'on time'; the modal average, would give a better impression, but it would be giving a false impression as over 50% of trains were in fact late!

(e) Estimate $= \dfrac{6}{50} = 0.12 = 12\%$

Exercises

1. The number of days absence for each pupil in a class is summarised in the table.

No. of Days Absent	*Frequency*
0	10
1	11
2	5
3	0
4	2
5	1
6	1
more than 6	0

Calculate:

(a) the *mode,*

(b) the *median,*

(c) the *mean,*

(d) the *range,*

for the data.

2. A new minibus service to the nearest town is provided for an isolated village. The number of people using the service during the first month of operation is summarised in the table.

No. of People Using the Service	*Frequency*
0	5
1	4
2	3
3	4
4	6
5	2
6	2
7	3
8	1
9	1
10	0

(a) Calculate:

(i) the *mode,*

(ii) the *median,*

(iii) the *mean.*

(b) Which of these average values give the best justification for continuing the service? Why could it be criticised as not giving a fair representation of the use of the minibus service?

3. A machine in a youth club sells snacks as listed in the following table.

Crisps	*20p*
Chocolate bars	*35p*
Drinks	*40p*
Rolls	*75p*
Sandwiches	*£1.00*

Len writes down the amounts of money which different people spend one evening during each hour that the club is open:

Amounts of Money Spent During Each Hour		
5 p.m. to 6 p.m.	6 p.m. to 7 p.m.	7 p.m. to 8 p.m.
40p	75p	£1.75
60p	55p	£1.40
55p	60p	£1.60
20p	40p	75p
40p	£1.15	£1.40
60p	40p	£1.10
55p	75p	60p
40p	40p	£1.50

(a) Explain why Len is correct when he says that the *mode* of the amounts of money spent is 40p.

(b) Copy the chart below and fill in the column for 7 p.m. to 8 p.m. Then fill in the column for the *total* number of people who spent each amount.

Amount of Money Spent	*Time*			*Total Number of People Who Spent Each Amount*
	5 p.m. to 6 p.m.	*6 p.m. to 7 p.m.*	*7 p.m. to 8 p.m.*	
Under 50p	𝍸 I	𝍸		7
50p to 99p	𝍸 I	𝍸 I		
£1.00 to £1.49		I		
Over £1.49				

(c) Len says: "*Now 50p to 99p is the mode.*"
Is Len right? Explain your answer.

(d) Look at where the tally marks are on the chart. What do you notice about the amounts of money people spent at different times in the evening? Give a reason which could explain the difference you notice.

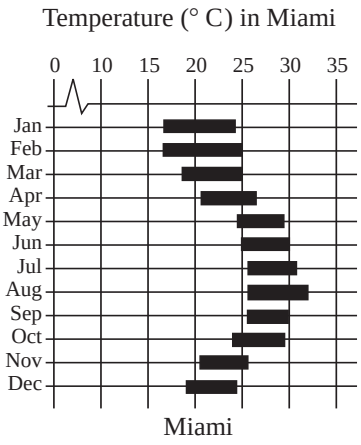
(KS3/96/Ma/Tier 4-6/P2)

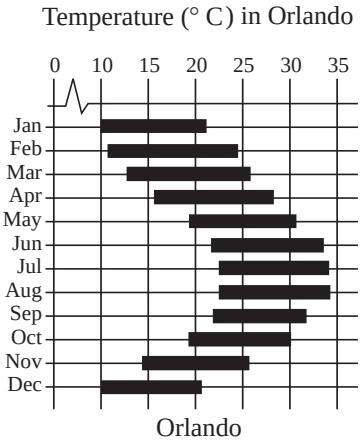
4. This graph shows the *range* of *temperature* in Miami each month. For example, in January the temperature ranges from 17 °C to 24 °C.

(a) In which month does Miami have the *smallest range*?

(b) In July, the *range* in the temperature in Miami is 5 °C. There are five other months in which the range in the temperature is 5 °C. Which five months are they?

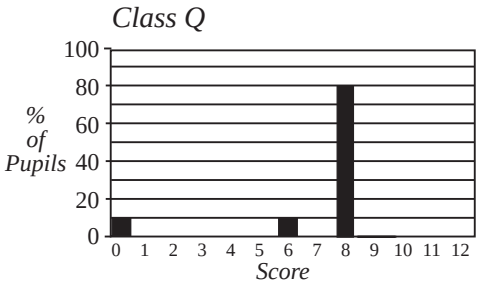
(c) This graph shows the range in the temperature in Orlando each month. In which three months is the maximum temperature in Miami greater than the maximum temperature in Orlando?

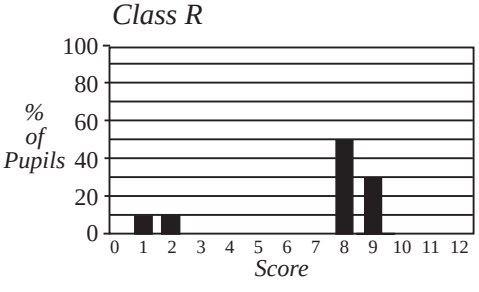


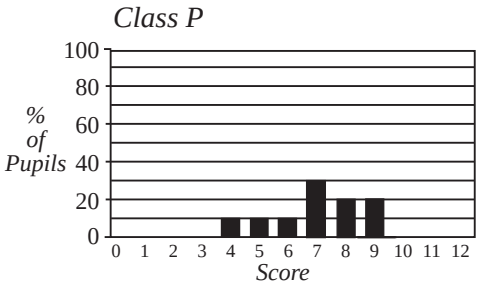


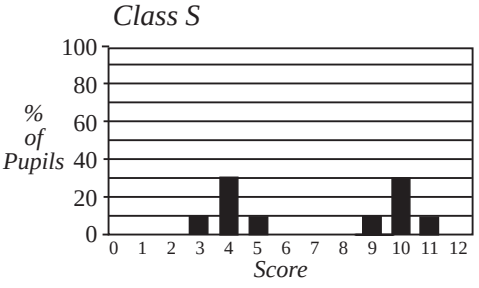
(KS3/99/Ma/Tier 4-6/P1)

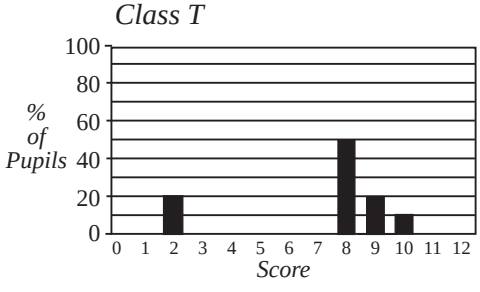
5. The pupils in five classes did a quiz. The graphs show the scores in each class. Each class had a mean score of 7. In three of the classes, 80% of the pupils got more than the mean score.











(a) In which *three* classes did 80% of the pupils score *more than 7* ?

(b) Look at the graphs which show that 80% of the pupils scored more than 7. Some of the statements below are true when 80% of the pupils scored more than 7.

Write down the letter for each of the statements below which is true.

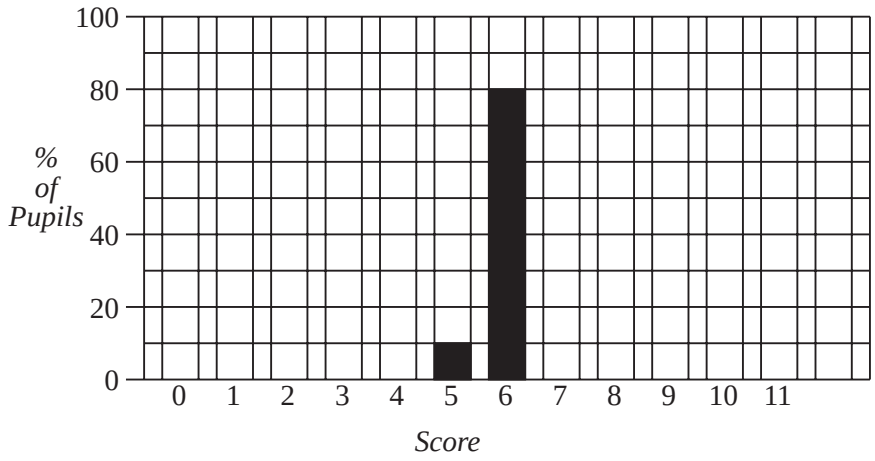
A : All of the pupils scored at least 2.

B : Most of the pupils scored at least 8.

C : Most of the pupils scored at least 10.

D : Some of the pupils scored less than 6.

(c) In another quiz the *mean score* was 6. Copy and complete the following graph to show a mean score of 6.

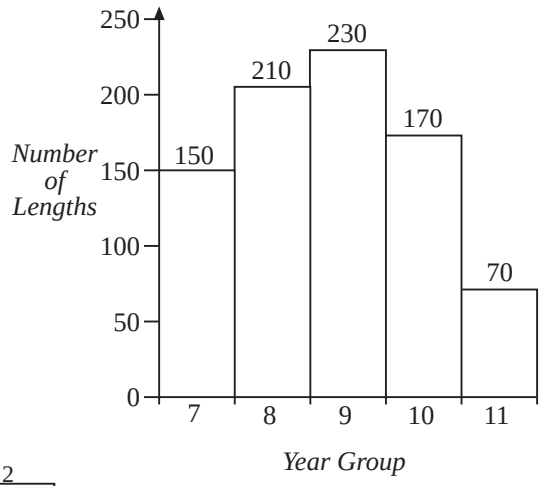


(KS3/98/Ma/Tier 5-7/P1)

6. A school has 5 Year groups. 80 pupils from the school took part in a sponsored swim. Lara and Jack drew these graphs.

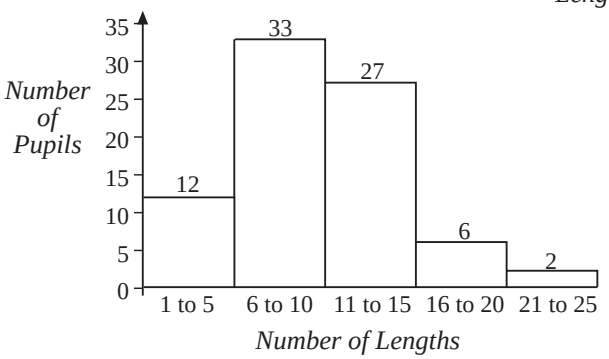
Lara's graph:

Number of Lengths Swum by Each Year Group



Jack's graph:

Number of Pupils Who Swam Different Numbers of Lengths



(a) Look at Lara's graph. Did Year 10 have *fewer* pupils taking part in the swim than *Year 7* ? Write down one of the following as your answer:

Yes, *No* or *Cannot tell*

Explain your answer.

(b) Use Lara's graph to work out the mean number of lengths swum by each of the 80 pupils. Show your working.

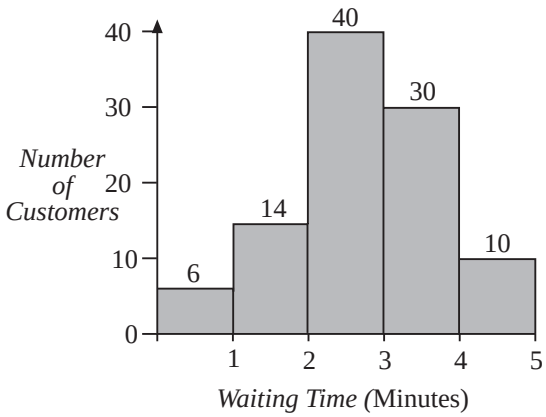
(c) Use Jack's graph to work out the mean number of lengths swum by each of the 80 pupils. Show your working.

(d) Explain why the means calculated from Lara's graph and Jack's graph are different.

(KS3/96/Ma/Tier 6-8/P2)

7. A customer at a supermarket complains to the manager about the waiting times at the check-outs. The manager records the waiting times of 100 customers at check-out 1.

Results



(a) Use the graph to estimate the probability that a customer chosen at random will wait for 2 minutes or longer.

(b) Use the graph to estimate the probability that a customer chosen at random will wait for 2.5 minutes or longer.

(c) Calculate an estimate of the *mean* waiting time per customer. Show your working. You may complete a copy of the table below to help you with the calculation.

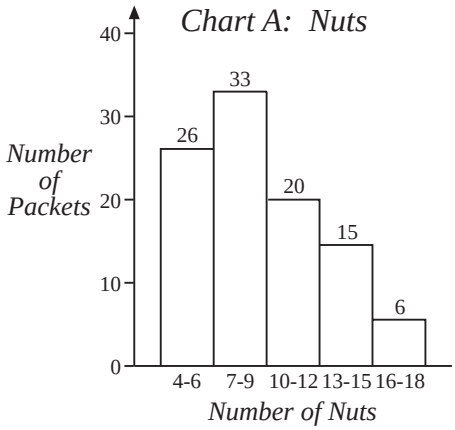
Waiting Time (minutes)	*Mid-point of bar* (x)	*Number of Customers* (f)	(fx)
0 -	0.5	6	3
1 -	1.5	14	
2 -	2.5	40	
3 -	3.5	30	
4 - 5	4.5	10	
		100	

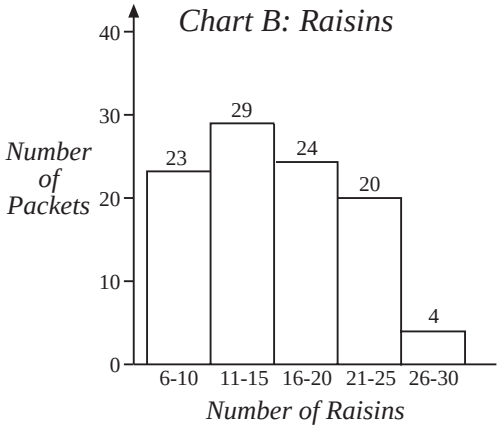
(d) The manager wants to improve the survey. She records the waiting times of more customers. Give a *different* way the manager could improve the survey.

(KS3/98/Ma/Tier 6-8/P1)

8. A company makes breakfast cereal containing nuts and raisins. They counted the number of nuts and raisins in 100 small packets.

Results:





(a) Calculate an estimate of the *mean* number of *nuts* in a packet. Show your working. You may complete a copy of the table below to help you with the calculation.

Number of Nuts	*Mid-point of Bar* (x)	*Number of Packets* (f)	fx
4 - 6	5	26	130
7 - 9	8	33	
10 -12	11	20	
13 - 15	14	15	
16 - 18	17	6	
		100	

(b) Calculate an estimate of the number of packets that contain 24 or more raisins.

(c) Which of the two charts shows the *greater range*? Explain your answer.

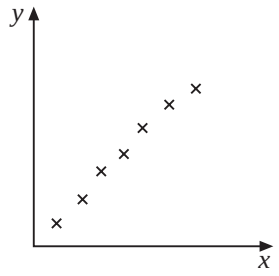
(d) A packet is chosen at random. Calculate the probability that it contains 9 nuts or fewer.

(e) The number of raisins in a packet is independent of the number of nuts. A packet is chosen at random. Calculate the probability that it contains 16 to 18 nuts and 6 to 10 raisins. Show your working.

(KS3/97/Ma/Tier 6-8/P2)

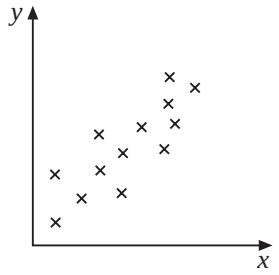
8.3 Plotting Scatter Diagrams

In this section we review plotting scatter diagrams and discuss the different types of correlation that you can expect to see on these diagrams.



Strong positive correlation between x and y. The points lie close to a straight line with

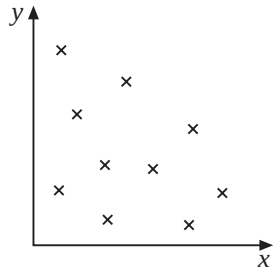
y increasing as x increases.



Weak, positive correlation between x and y. The trend shown is that

y increases as x increases

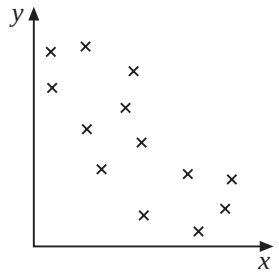
but the points are not close to a straight line.



No correlation between x and y; the points are

distributed randomly

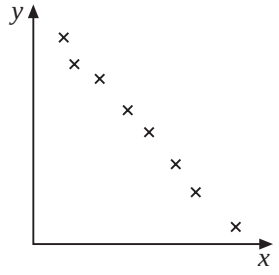
on the graph.



Weak, negative correlation between x and y. The trend shown is that

y decreases as x increases

but the points do not lie close to a straight line.



Strong, negative correlation. The points lie close to a straight line, with

y decreasing as x increases.

If the points plotted were all on a straight line we would have *perfect correlation*, but it could be positive or negative as shown in the diagrams above.

Example 1

The following table lists values of x and y.

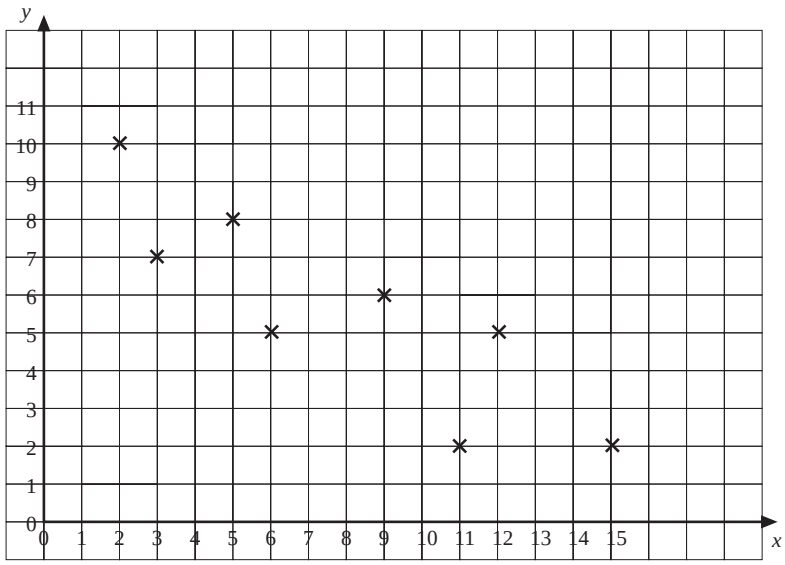
x	2	3	5	6	9	11	12	15
y	10	7	8	5	6	2	5	2

(a) Use the data to draw a scatter graph.

(b) Describe the type of correlation that you observe.

Solution

(a) The scatter graph is shown below.



(b) It shows weak, negative correlation.

Example 2

What sort of correlation would you expect to find between:

(a) a person's age and their house number,

(b) a child's age and their height,

(c) an adult's age and their height ?

Solution

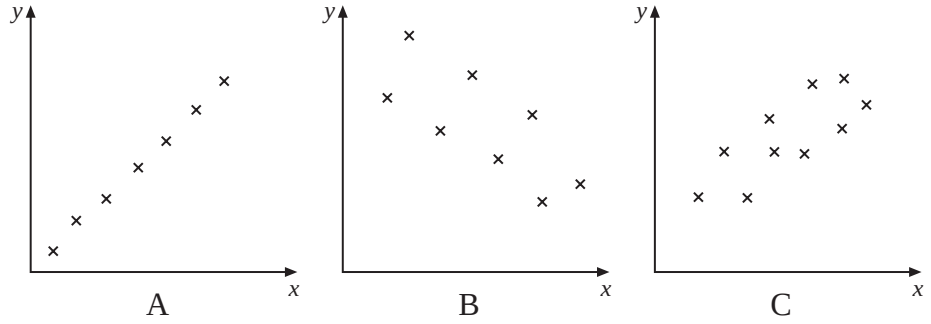
(a) No correlation, because these two quantities are not linked in any way.

(b) Positive correlation, because children get taller as they get older.

(c) No correlation, because the height of adults does not change with their age.

Exercises

1. Consider the following scatter graphs:



(a) Which graph shows *strong* correlation?

(b) Which graphs show *positive* correlation?

(c) Which graph shows *negative* correlation?

(d) Which graph shows a *weak, positive* correlation?

2. The following table lists values of x and y.

x	2	4	6	7	8	9	10	11	12
y	3	5	8	5	9	6	9	9	11

(a) Plot a scatter graph for this data.

(b) Describe the correlation between x and y.

3. Copy and complete the table below for 10 people in your class.

House Number										
Day of Month of Birthday										

(a) Plot a scatter graph for your data.

(b) Describe the type of correlation that there is between these two quantities.

4. A driver keeps a record of the distance travelled and the amount of fuel in his tank on a long journey.

Distance Travelled (km)	0	50	100	150	200	250	300
Fuel in Tank (litres)	80	73	67	61	52	46	37

(a) Illustrate this data with a scatter plot.

(b) Describe the type of correlation that is present.

5. What type of correlation would you expect to find between each of the following quantities:

   (a) *Age* and *pocket money,*

   (b) *IQ* and *height,*

   (c) *Price of house* and *number of bedrooms,*

   (d) *Person's height* and *shoe size ?*

6. In a class 10 pupils took a Science test and an English test. Their scores are listed in the following table:

Pupil	A	B	C	D	E	F	G	H	I	J
English Score	2	10	18	4	9	7	18	19	3	10
Science Score	18	12	6	3	11	20	4	17	7	2

   (a) Draw a scatter graph for this data.

   (b) Describe the correlation between the two scores.

7. Chris carries out an experiment in which he measure the extension of a spring when he hangs different masses on it. The following table lists his results:

Mass (grams)	20	50	100	120	200
Extension (cm)	1.2	3.0	6	7.2	12

   (a) Draw a scatter graph for this data.

   (b) Describe the correlation between the mass and the extension.

8. Every day Peter picks the ripe tomatoes in his greenhouse. He keeps a record of their mass and the number that he picks. His results are listed in the following table:

Number of Tomatoes Picked	1	3	2	5	8	6	7	4
Total Mass (grams)	40	180	60	270	390	220	420	210

   (a) Draw a scatter graph for this data.

   (b) Describe the correlation between the number of tomatoes picked and their total mass.

9. A competition has 3 different games.

(a) Jeff plays 2 of the games.

	Game A	*Game B*	*Game C*
Score	62	53	

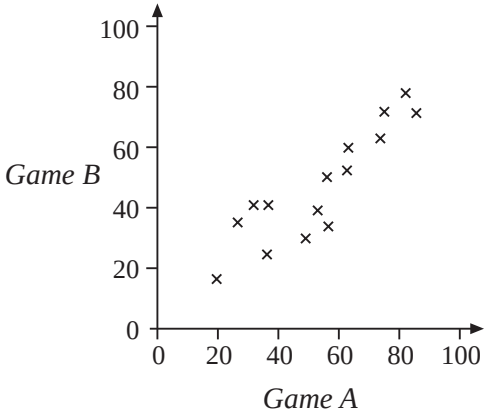
To win, Jeff needs a *mean* score of 60. How many points does he need to score in Game C? Show your working.

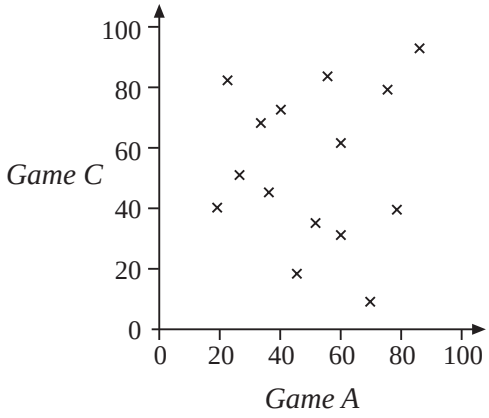
(b) Imran and Nia play the 3 games. Their scores have the *same mean.* The *range* of Imran's score is *twice* the range of Nia's scores.

Copy the following table and fill in the missing scores.

Imran's Scores		40	
Nia's Scores	35	40	45

The scatter diagrams show the scores of everyone who plays all 3 games.





(c) Look at the scatter diagrams. Write down a statement from the table below which most closely describes the *relationship* between the games.

Game A and Game B				
perfect negative relationship	negative relationship	no relationship	positive relationship	perfect positive relationship

Game A and Game C				
perfect negative relationship	negative relationship	no relationship	positive relationship	perfect positive relationship

(d) What can you tell about the *relationship* between the scores on Game B and the scores on Game C? Write down the statement below which most closely describes the relationship.

Game B and Game C				
perfect negative relationship	negative relationship	no relationship	positive relationship	perfect positive relationship

(KS3/98/Ma/Tier 6-8/P2)

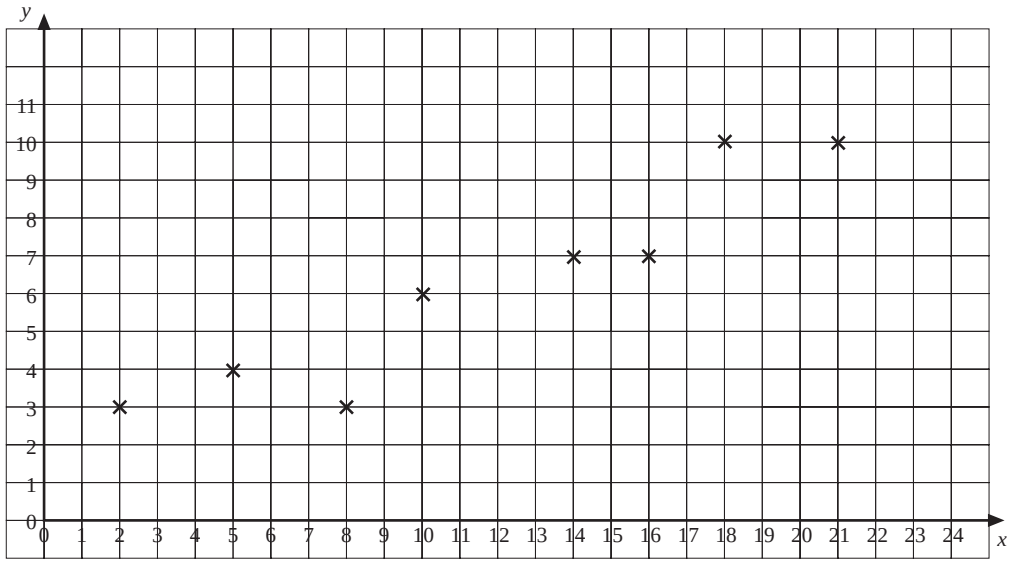
8.4 Lines of Best Fit

When there reasonable correlation between two variables on a scatter plot, it is possible to draw a *line of best fit.* This line represents the underlying relationship between the two quantities. When drawing a line of best fit the aim is to keep the distances of all the points from the line to a minimum. Sometimes it is helpful to try to keep the number of points above the line the same as the number of points below the line.

Lines of best fit can be used to make predictions. The accuracy and reliability of those predictions will depend on the strength of the correlation between the two variables.

Example 1

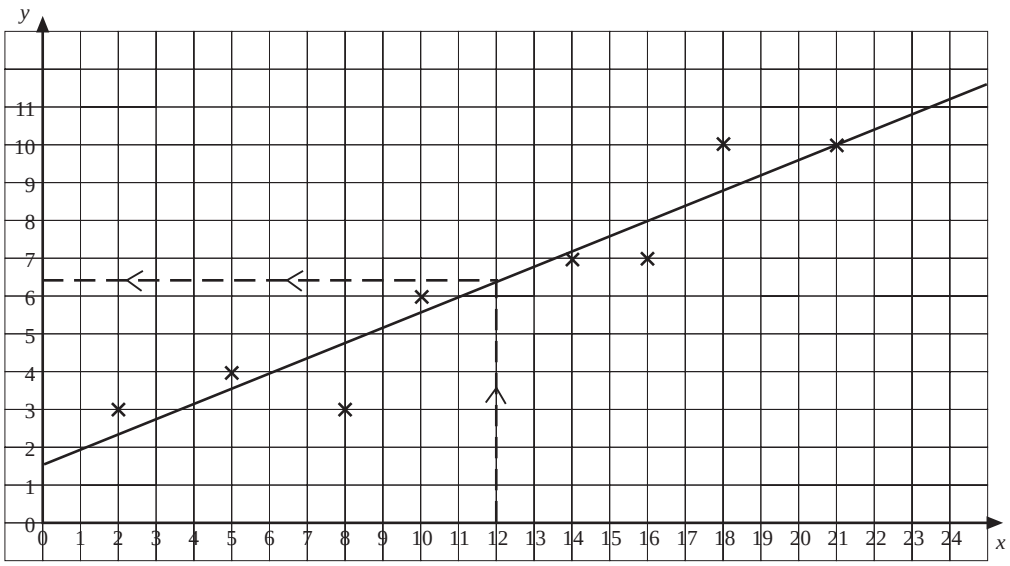
(a) Draw a line of best fit for the points in the following scatter graph:



(b) Use the line to predict the value of y when $x = 12$.

Solution

(a) The line of best fit has been drawn on the following scatter graph:



Note that there are 3 points above the line and 3 below. The total distances to the points above the line is similar to the total distance to the points below the line.

(b) Using the dotted line, we have $y = 6.4$ when $x = 12$.

Example 2

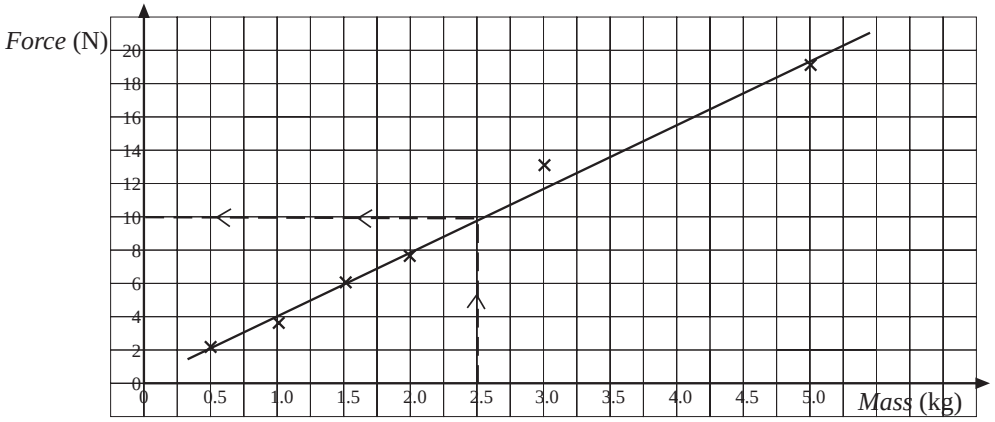
The following data was collected from an experiment. In the experiment, objects of different masses were placed on a horizontal surface and the force needed to make them start to move was recorded.

Mass (kg)	0.5	1.0	1.5	2.0	3.0	5.0
Force (N)	2.1	3.8	6.1	7.9	13.2	19.1

Use a scatter graph to estimate the force needed for a 2.5 kg mass.

Solution

The scatter graph and line of best fit are shown below.



The graph also shows that the estimated force for a 2.5 kg mass is 10 N.

Exercises

1. (a) Use the data shown to draw a scatter plot.

   (b) Draw a line of best fit for the data.

x	1	2	3	4	5	6
y	7	10	12	15	19	21

   (c) Estimate the value of y when $x = 0$.

2. The Maths and Science test results for 10 pupils are listed below:

Pupil	A	B	C	D	E	F	G	H	I	J
Maths Score	45	83	65	62	71	52	69	72	58	64
Science Score	39	80	59	60	65	54	65	67	56	64

   (a) Draw a scatter graph for this data and then draw a line of best fit.

   (b) Estimate the score on the Science test for pupils who scored:
   (i) 73 (ii) 40
   on the Maths test.

3. The following data was collected by a lorry driver who was interested in how much fuel he used on different journeys.

Length of Journey (miles)	100	250	150	180	220	300
Fuel Used (litres)	24	59	44	50	59	97

   (a) Draw a scatter graph for this data.

   (b) Draw a line of best fit.

   (c) Estimate how much fuel would be needed for a 200 mile journey.

4. A pupil carried out an experiment where he recorded the length of a spring when various masses were hung from it.

Mass (grams)	50	80	100	150	200	300
Length (cm)	6.0	6.6	6.9	8.0	9.1	11.1

   Use a scatter graph and a line of best fit to estimate the length of the spring when:

   (a) no mass is hung from it,

   (b) a mass of 250 grams is hung from it.

5. Rafiq collected the following data on the height and shoe size of some pupils in his class:

Shoe Size	6	4	8	5	9	10	4	5.5
Height (cm)	143	150	172	146	165	177	141	156

(a) Draw a scatter plot and a line of best fit for the data.

(b) Estimate the height of a person with a shoe size of 7.5.

(c) Ian has a height of 170 cm. Estimate his shoe size.

6. A garage owner keeps a record of the age and price of the small family cars that the garage sells. Some of these records are given in the following table:

Age (years)	6	5	7	3	1	2	3	7	9	10
Price (£)	5700	6800	5300	7700	8500	7900	7800	5700	3700	3600

(a) Draw a scatter graph and a line of best fit for this data.

(b) Estimate the price of a 4-year-old car and a 12-year-old car.

7. An electric heater was turned on in a cold room. The temperature was recorded every 2 minutes.

Time (minutes)	0	2	4	6	8	10	12	14	16	18	20
Temperature (°C)	8.0	9.3	10.4	11.5	12.7	13.9	15.0	16.0	17.0	18.2	19.4

(a) Estimate the temperature after 15 minutes.

(b) Estimate when the temperature will reach 22 °C.

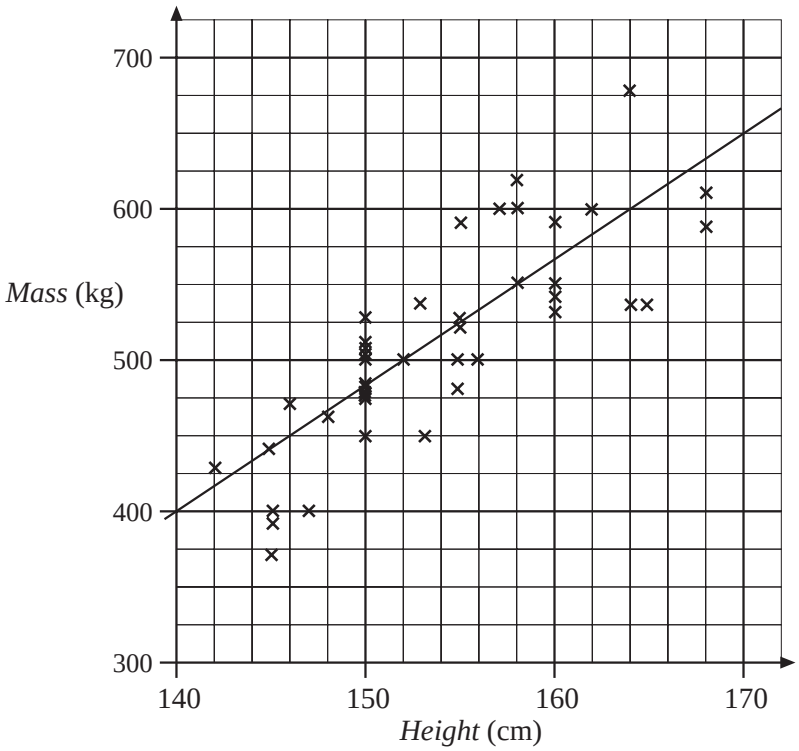
8. A biology student measured the height of a small plant at weekly intervals. The results obtained are listed in the following table:

Time (weeks)	0	1	2	3	4	5	6	7
Height (cm)	1.2	2.5	3.6	4.5	5.3	6.4	7.2	8.3

(a) Estimate the height of the plant after $3\frac{1}{2}$ weeks.

(b) Estimate when the height of the plant will be 10 cm.

9. The scatter diagram shows the heights and masses of some horses. The scatter diagram also shows a line of best fit.



(a) What does the scatter diagram show about the *relationship* between the height and mass of the horses?

(b) The *height* of a horse is 163 cm. Use the line of best fit to estimate the mass of the horse.

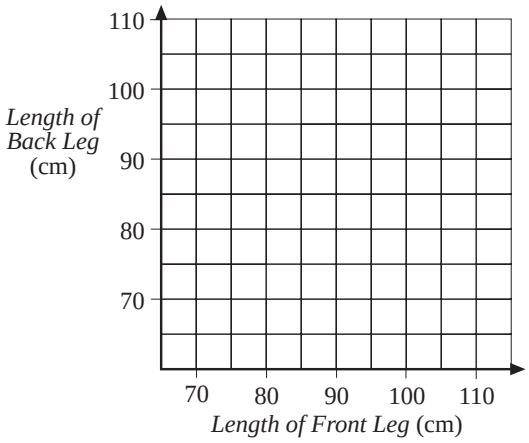
(c) A different horse has a mass of 625 kg. Use the line of best fit to estimate the height of the horse.

(d) A teacher asks his class to investigate this statement:

"The length of the *back leg* of a horse is *always less than* the length of the *front leg* of a horse."

What might a scatter graph look like if this statement is correct?

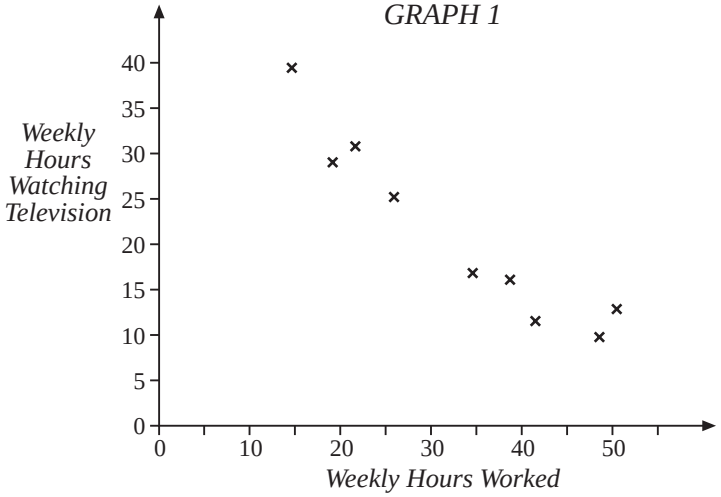
Show your answer on a copy of the axes below.



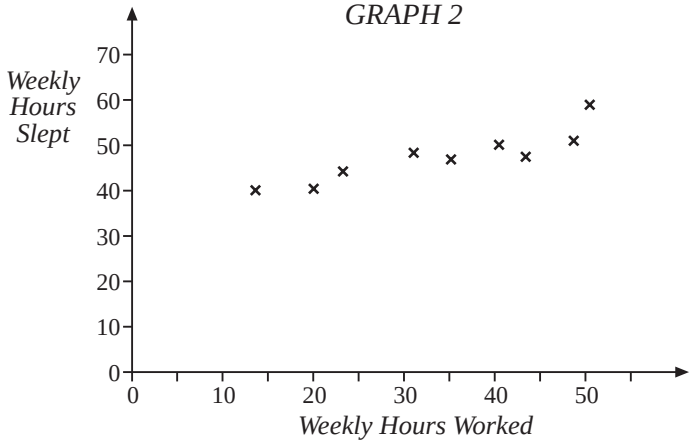
(KS3/99/Ma/Tier 6-8/P1)

10. Nine students were discussing their holiday jobs working on a local farm. They decided to find out if there were any relationships between the time they spent working, sleeping, watching television and the distance they had to travel to work. The students plotted three scatter graphs.

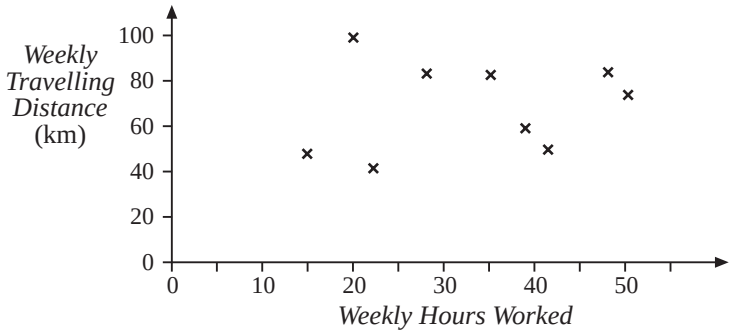
GRAPH 1



GRAPH 2



GRAPH 3



(a) What does Graph 1 show about the relationship between the weekly hours spent watching television and the weekly hours worked?

(b) What does Graph 2 show about the relationship between the weekly hours slept and the weekly hours worked?

(c) What does Graph 3 show about the relationship between the weekly travelling distance and the weekly hours worked?

(d) Another student works 30 hours per week. Use Graph 1 to estimate the weekly hours spent watching television by this student. Explain how you decided on your estimate.

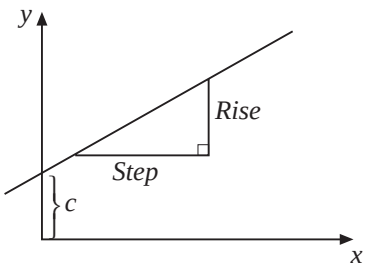
(KS3/95/Ma/Levels 6-8/P2)

8.5 Equation of the Line of Best Fit

If you draw a line of best fit, it is possible to determine the *equation* of the line of best fit. You will remember that the equation of a straight line is given by

$$y = mx + c$$

where m is the gradient and c is the intercept.



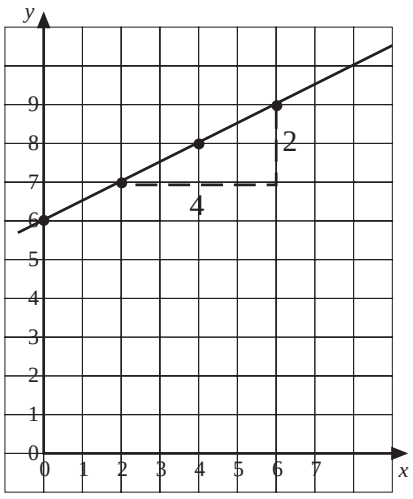
$$m = \frac{Rise}{Step}$$

Example 1

The points with coordinates (0, 6), (2, 7), (4, 8) and (6, 9) lie on a straight line. Draw the line and determine its equation.

Solution

The points and the line are shown on the graph.

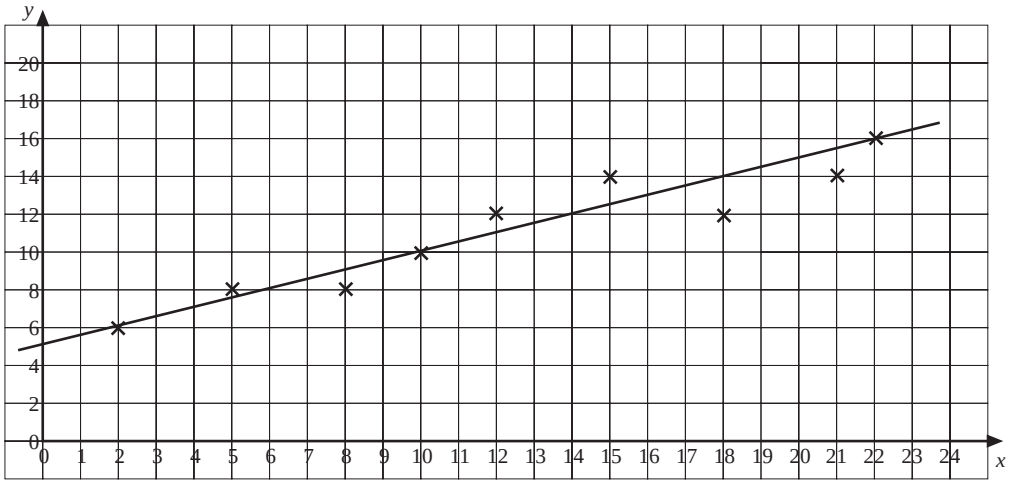


The intercept is 6. The gradient $= \frac{2}{4} = \frac{1}{2}$, so the equation of the line is

$$y = \frac{1}{2}x + 6$$

Example 2

The following graph shows a scatter plot and a line of best fit:



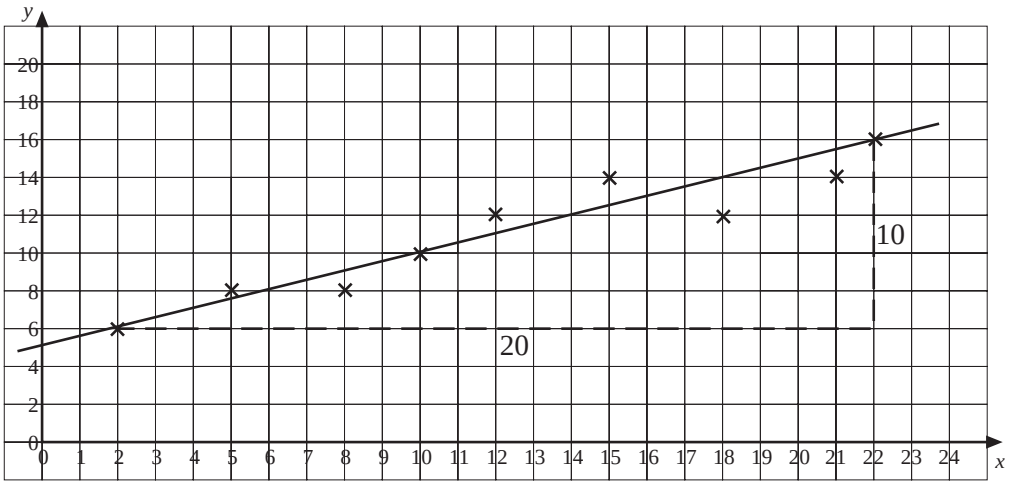
(a) Determine the equation of the line of best fit.

(b) Use the equation to estimate y when $x = 4$.

(c) Use the equation to estimate x when $y = 18$.

Solution

(a) The intercept and the gradient can be found from the graph, as shown on the following diagram. (Note that the scales on the vertical and horizontal axes are not the same.)



$$c = 5, \qquad m = \frac{10}{20}$$
$$= \frac{1}{2}$$

so the line of best fit has equation $y = \frac{1}{2}x + 5$.

(b) Substitute $x = 4$ into the equation.

$$y = \frac{1}{2} \times 4 + 5$$
$$= 2 + 5$$
$$= 7$$

(c) Substitute $y = 18$ into the equation for the line of best fit and solve the equation this gives.

$$18 = \frac{1}{2}x + 5$$

$$13 = \frac{1}{2}x \qquad \textit{(subtracting 5 from both sides)}$$

$$x = 2 \times 13 \qquad \textit{(multiplying both sides by 2)}$$
$$= 26$$

Note of Warning!

In (b) above, the value of x used was within the range of values of x provided by the original data. We can be confident that the estimate we obtain is reasonable. This process is called *interpolation*.

In (c) above, the value of x we obtain is well outside the range of values of x provided by the original data. This process is called *extrapolation* and the results must be *treated with caution* as they may be very unreliable. In some cases, extrapolation can generate bogus predictions.

Exercises

1. Each set of points below lies on a straight line. Determine the equation of each line.

   (a) (0, 3), (5, 5), (10, 7) and (15, 9)

   (b) (1, 5.3), (3, 5.5), (5, 5.7) and (7, 5.9)

   (c) (0, 6), (3, 5.4), (5, 5) and (8, 4.4)

2. The relationship between two quantities L and x is to be investigated using the data shown.

x	0	100	200	300	400
L	6	6.4	6.9	7.3	7.6

   (a) Draw a scatter graph with x on the horizontal axis and draw a line of best fit.

   (b) Determine the equation of the line of best fit.

(c) Use the equation to estimate L when $x = 250$ and 500. Comment on how reliable your estimates are.

3. In the calibration of a thermometer, the height, H cm, of the mercury is recorded at different temperatures. The results are listed below.

Temperature (°C)	5	20	35	50	80
H (cm)	4.5	21.0	35.2	51.2	78.6

(a) Draw a scatter graph and a line of best fit.

(b) Determine the equation of the line of best fit.

(c) Estimate H when the temperature is $60\ ^\circ\text{C}$ and $120\ ^\circ\text{C}$.

(d) Which of your estimates is the more reliable? Explain why.

4. Refer back to the scatter graphs and lines of best fit you used each of the questions 1 to 8 in the Exercises in section 8.4. Determine the equation of the line of best fit for each question.

5. A long distance lorry driver records the times it takes to make journeys of different lengths. This information is recorded below:

Length of Journey (miles)	150	229	260	290	320
Time Taken (hours)	$3\frac{1}{4}$	$4\frac{1}{2}$	$6\frac{1}{4}$	$6\frac{1}{2}$	$7\frac{3}{4}$

(a) Comment on the way that the driver records the time taken.

(b) Plot the data and draw a line of best fit.

(c) Determine the equation of the line of best fit.

6. In an experiment a flask of water is heated. The temperature of the water is recorded at two minute intervals. The results are recorded in the following table:

Time (minutes)	0	2	4	6	8	10
Temperature (°C)	18	30	42	56	71	84

(a) Plot the data on a graph and determine the equation of the line of best fit.

(b) Use the equation to predict the temperature after 11 minutes.

(c) Why would it not be wise to use the line of best fit to predict temperatures for later times than 11 minutes?

7. A driver records the petrol consumed on a number of journeys of different lengths. The data is presented in the table below:

Journey Length (miles)	100	180	250	300	320	350
Petrol Consumption (gallons)	3.5	5.6	7.9	8.4	9.3	10.9

Plot a graph of petrol consumed (vertical axis) against journey length (horizontal axis) and determine the equation of the line of best fit. Use this to predict the petrol needed for a journey of 280 miles.

8. The number of triplets and higher order births per 100 000 of the population, as recorded for various years between 1984 and 1994, is given in the following table:

Year	1984	1987	1988	1989	1991	1992	1994
No. of Triplets and Higher Order Births per 100 000 of the Population	13	21	20	29	32	31	40

(a) Plot a graph to illustrate this data and draw a line of best fit.

(b) Determine the equation of the line of best fit.

(c) Estimate the number of triplets and higher order births per 100 000 of the population in the year 2020. Comment on the reliability of your estimate.