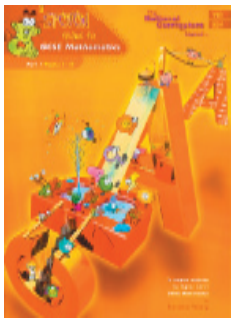


ISBN-10: 1-874428-91-3
ISBN-13: 978-1-874428-91-6

Sample Topic
from the
Y10 Higher GCSE text

You have permission to print this topic and try it with your students.

This document contains the complete topic (p01–43) and answers (p44-47)

Teachers' notes, worksheets, revision and assessment for this topic can be downloaded and printed off in the Y10 Spoton Resource Sample document.

Spoton

ENJOYMENT

The 'Maths is …' Jugglers

Knowledge Skills Understanding

SPOTON GUIDE
Y10

Topic 7

Analysing Statistics

CONTENTS

Section 1: Frequency diagrams
Section 2: Data calculations
Section 3: Tables of ungrouped data
Section 4: Tables of grouped data
Section 5: Tables of continuous data
Section 6: Questionnaires and bias
Section 7: Testing hypotheses
Section 8: Scatter graphs
Section 9: Controlled experiments
Section 10: Time series graphs (and moving averages)
Section 11: Cumulative frequency
Section 12: A simple survey
Section 13: Sampling
Section 14: Histograms

New Higher Syllabus material at D grade level

The Spoton texts already cover much of the material at D grade level that is required for the new Higher GCSE Syllabus.

However, each school purchasing Spoton texts will be provided with a small package that contains:

- a list of all the D grade items
- a match of most of these items to the Spoton texts
- photocopiable material from our Intermediate texts for the few items that are not covered, for teachers to use as they wish.

Also, any school can download this package free from our website .

For ALL numerical calculations:

- mental arithmetic or pencil-and-paper techniques should be used wherever possible
- calculators should only be used if the arithmetic is very complex

Key: means that a calculator is essential for one or more pieces in this section.

NOTE: Able students ought to tackle as many EXTENSION problems as they can find time for. These develop high level problem solving skills and, if sufficient are done, will raise the students problem-solving capabilities well above basic GCSE questions. They will put able students on the path towards A* at GCSE and a high grade at A-Level.

Spoton

Contents
of the
Higher GCSE Course

ISBN-10: 1-874428-91-3
ISBN-13: 978-1-874428-91-6

THE SPOTON GUIDE PART ONE (Y10)

Topic 1:	Number Techniques — *Covers the techniques that will be assumed in the other topics*	5 –	37
Topic 2:	Sequences and Proof	38 –	74
Topic 3:	Expressions and Equations	75 –	110
Topic 4:	Pythagoras and Trigonometry	111 –	145
Topic 5:	Fractions, Decimals, Percentages, Ratios	146 –	178
Topic 6:	Transformations, Shapes, Areas and Volumes	179 –	204
Topic 7:	**Analysing Statistics** **SAMPLE TOPIC**	**205 –**	**248**
Topic 8:	Further Number Techniques — *These topics are in both parts of the course to provide flexibility.*	249 –	282
Topic 9:	Graphs	283 –	304
Topic 10:	Measuring Likelihood	335 –	336
	ANSWERS	337 –	360

ISBN-10: 1-874428-92-1
ISBN-13: 978-1-874428-92-3

THE SPOTON GUIDE PART TWO (Y11)

Topic 8:	Further Number Techniques — *These topics are in both parts of the course to provide flexibility.*	5 –	38
Topic 9:	Graphs	39 –	60
Topic 10:	Measuring Likelihood	61 –	92
Topic 11:	Arcs, Areas, Volumes and Formulae	93 –	119
Topic 12:	**Algebra – towards A-Level** **SAMPLE TOPIC**	**120 –**	**170**
Topic 13:	Graphs and Functions	171 –	218
Topic 14:	Further Geometric Techniques	219 –	251
Topic 15:	3D-Trigonometry and Circle Theorems	252 –	290

The emphasis here is on non-calculator skills, with a particular stress on mental agility. The Teachers' Guide for each topic (in the Resource Pack) includes a list of items that could be used as mental/oral starters and the techniques taught/reviewed here should be repeated regularly over the weeks following their introduction. For details of Teachers' Resource Packs, see the next page.

The course should start with Topic 1. This contains the number techniques that will be assumed thereafter throughout the course.
The rest of the topics are independent and can be done in any order.
Any techniques required within a topic that are taught elsewhere, will be repeated at the point where they are required.

Spoton

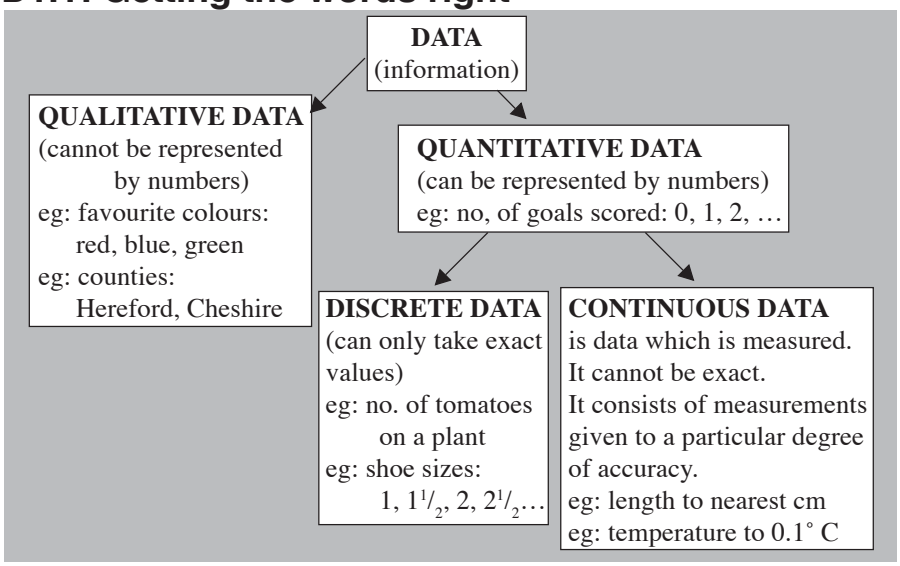
Analysing Statistics

Section 1: Frequency diagrams

In this section you will:
- review some statistical terms
- read and compare data from frequency diagrams for ungrouped and grouped data

DEVELOPMENT

D1.1: Getting the words right



State whether each of these sets of data is qualitative (QL) or quantitative (QN) data:

1. ages of children in a family
2. favourite football teams
3. suits of cards
4. heights of Y7 pupils
5. pets
6. number of pets

State whether each of these sets of data is discrete (D) or continuous (C) data:

7. numbers of children in families
8. heights of children in families
9. scores in game of darts
10. numbers of books in pupils' bags
11. lengths of handles on pupils' bags
12. weight of pupils' bags
13. heights of Y7 pupils
14. numbers of pets
15. number of golf shots for each hole
16. length of golf shots
17. temperature in the classroom
18. ages of each of the teachers

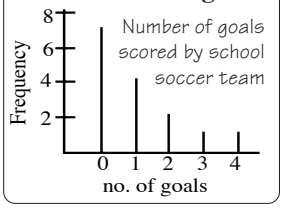
• Check your answers.

D1.2: Frequency diagrams

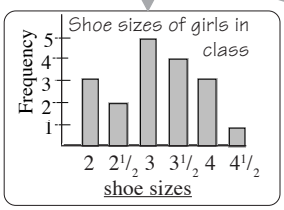
Data source: Social Trends

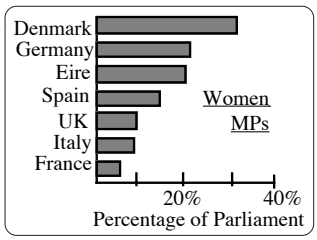
Ungrouped data (qualitative or discrete)

Vertical line diagram



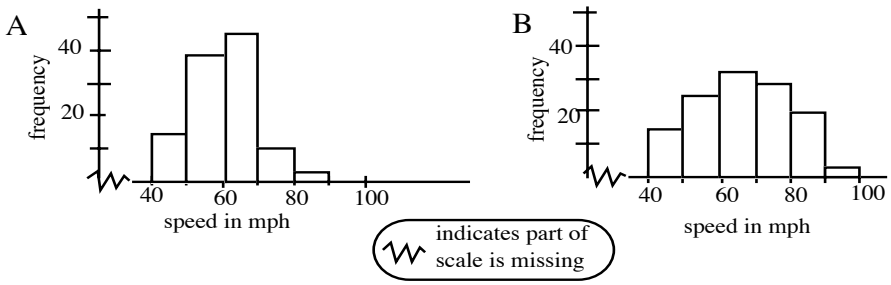
Bar charts or bar graphs





1. What is the modal (most common) number of goals ?
2. How many matches did the team play altogether ?
3. In how many matches did they score more than 1 goal ?
4. How many girls were asked their shoe sizes ?
5. How many girls take size 3 or smaller ?
6. Which of these countries has the largest percentage of women MPs ?
7. Which of these countries has a smaller percentage of women MPs than the UK ?

Grouped data (discrete or continuous)



8. Both of these frequency diagrams shows the speeds of cars on a motorway.
   One of them shows the speeds when a police car is also on the motorway.
   One of them shows the speeds of cars when there is no police car.

   Which is the one that shows the speeds with the police car present ?
   Explain how you can tell.

• *Check your answers.*

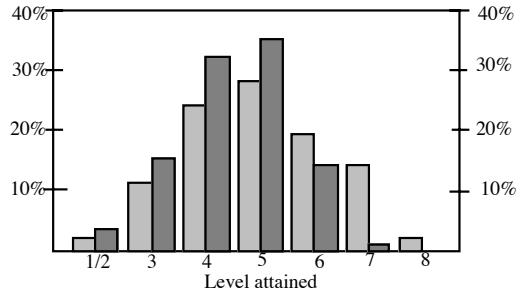
Note: when data is grouped:
- the scale is continuous
- there are no gaps between the bars

Note: when data is not grouped:
- the scale is not continuous
- it is usual but not essential to have gaps between the bars. However if there are gaps, then it is obvious that the data is not grouped

D1.3: Comparing frequencies with ungrouped data

Ungrouped data – multiple bar charts

Data source: Social Trends



Say whether each of these statements is true (T) or false (F):

A: Few students attained Levels 7 and 8 in Science.

B: The most common level attained in both Maths and Science was Level 4.

C: Around 30% of students attained less than Level 4 in Maths.

D: Around 50% of students attained less than Level 4 in Science.

E: More students achieved higher levels in Maths than in Science.

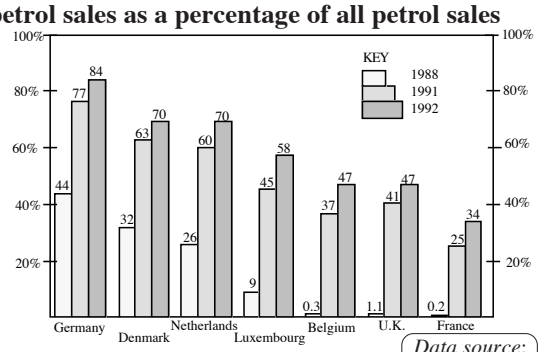
F: The Science tests appear to have been harder than the Maths tests.

• *Check your answers.*

EXTENSION

E1.4: *The changeover to unleaded petrol

Unleaded petrol sales as a percentage of all petrol sales



Data source: Social Trends

Journalists can often select data to fit the view that they want to deliver.

Imagine you are a free-lance journalist. You write articles for several different newspapers.

1. Write an paragraph for one newspaper to go under this headline.
   You must quote figures to back up your argument

   "EUROPE CHANGING RAPIDLY TO LEAD-FREE PETROL"

2. Write a paragraph for another newspaper to go under this headline.
   You must quote figures to back up your argument

   "CHANGE TO LEAD-FREE PETROL IS RUNNING OUT OF STEAM"

3. Which is the fairer picture? • *Show these to your teacher.*

DEVELOPMENT

D1.5: Comparing frequencies with grouped data

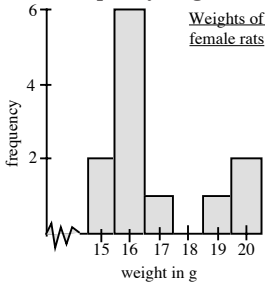
A scientist was investigating a group of rats that she caught on derelict building site. She recorded the weight of each rat to the nearest gram. She wanted to compare the weights of male and female rats.

Weights of female rats (g)					
16	15	20	17	16	19
15	16	16	20	16	16

Frequency table

Weight	frequency
14.5–>15.5	
15.5–>16.5	
16.5–>17.5	
17.5–>18.5	
18.5–>19.5	
19.5–>20.5	

Frequency diagram



1. *Complete the frequency table for the weights of the female rats.*

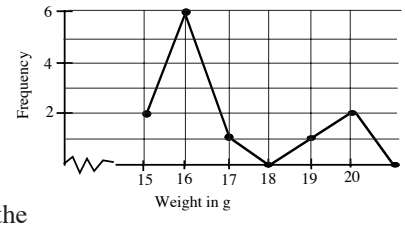
The scale for grouped data is continuous, so she could not draw a multiple bar chart. She drew a frequency diagram for females.

Weights of male rats (g)					
17	18	20	17	19	19
19	20	19	21	20	18

3. *Make a frequency table for the weights of the male rats.*

4. *Copy the frequency diagram for female rats. On the same diagram, make a frequency diagram for the weights of male rats. Use a different colour.*

She found that the two frequency diagrams were not very clear. So she decided draw two frequency polygons on one set of axes instead. The frequency polygon for the weights of the female rats has already been drawn.



5. For the frequency polygon, do you plot the middle values of each interval, or the highest value ?
6. Copy the frequency polygon for female rats.
   On the same diagram, draw a frequency polygon for the male rats.
   [Unless you fall off the scale, you should always extend the polygon down to a zero value on each end. In this case the zero value was put in on the right of the scale, but there was no room to do so on the left.]
7. Male rats usually weigh more than female rats, except when the females are pregnant.
   How many female rats do you think are pregnant ?
   Explain how you get your answer.
8. What are the usual weights of male and female (not pregnant) rats ?
9. The modal (most common) weight of female rats is higher than the modal weight of male rats. True or false ?
10. The range of weights was the same for both males and females. True or false ?

• Check your answers.

Section 2: Data calculations

In this section you will:
- work with three kinds of average and the range of a set of data
- use the statistical functions of a calculator to work out the mean

DEVELOPMENT

D2.1: Measures of data

The **mode, median** and **mean** are three different types of **average**.
Each one is a way of telling us where the middle of the data is.
The **range** is a measure of the **spread of the data**.

0 1 3 4 5 5

Mode = most common value
Here, the mode is 5
The mode is also called the **modal value.**
There may be more than one mode.
There may be no mode.

Median = middle value
(when the values are placed in order)
or halfway between the two middle values.
Here, median = halfway between 3 and 4.
Here, the median = 3.5
The median is also called the **median value.**

Mean = $\frac{\text{sum of all values}}{\text{number of values}}$
Here, the mean = 18 ÷ 6 = 3
The mean is also called the **mean average** or the **arithmetic mean**

Range = difference between largest and smallest values
Here, the range = 5 – 0
range = 5

Find the mean, mode, median and range of each of these sets of numbers.
Start by writing each set in order of size.

1. 3 6 10 5 3
2. 26 24 24 26 24
3. 13 14 13 14 13 15 13 14
4. 6 6 2 10 2 9 10 2 6 11 2
5. 4 4 3 4 2 2 1 0 1 0 2 1 3 0 2 3
6. 2.5 1.5 1.5 2.05 2.5 2.5 2.01
7. 507 506 506 522 540
8. 0.5 0.38 0.55 0.55 0.42
9. –1 2 4 –1 2 –5 –1
10. 2 13 21 5 11 13 8 13 17 9 9

• Check your answers.

D2.2: Using calculator functions to find the mean

1. Find out how to put your calculator into statistical mode.

EXAMPLE For 2 4 6 7 9 use your calculator to:
(a) work out the mean
(b) give you the number of items entered
(c) give you the sum of the data list

On some calculators, RCL is MR

Put the data in : *either* 2 M+ 4 M+ 6 M+ 7 M+ 9 M+
or 2 DATA 4 DATA 6 DATA 7 DATA 9 DATA

(a) RCL $\bar{x}$ gives the mean = 5.6

(b) RCL n gives the number of items entered = 5 (useful as a check against errors)

(c) RCL Σx gives the sum of the data list = 28

Σx is read as 'sigma x'
Σ means 'the sum of'

For each of these data lists, use your calculator to find
(a) the mean to 2 d.p.
(b) the number of items in the list
*(c) the sum of the data lis*t

Use CA or AC to clear data between questions.

2. 37 42 43 51 63 72 67 55 43 46
3. 125 134 141 132 134 135 142 147 143 149 127 164
4. 4.5 6.2 7.3 7.1 6.2 8.9 8.2 7.4 3.7
5. The ages of the members of my family in my generation (me, my brothers, sisters and cousins) are given here:
   47 53 58 51 59 63 72 44 68 70 56 50 46
   (a) Find the average age of my generation in our family.
   (b) Find the sum of our ages.

• *Check your answers.*

D2.3: The maths test

Marks for 8FD

7	5	10	4	6	4	7	9	6	8
10	7	7	5	8	3	6	1	4	7
6	7	0	5	7	3	7	8	4	6

Results for 8MB
Mean = 6.5
Number of pupils in 8MB = 20
Range = 7

Calculate …
1. …the mean, mode, median and range for 8FD's marks.
2. …the total number of marks for 8MB.
3. …the total number of marks for both groups
4. …the mean mark for both groups together.

State …
5. … which group did best.
6. … which group had the more consistent set of marks.
7. Two pupils in 8MB did the test late. They got 2 and 5
   Calculate the new mean mark for 8MB, to 2 d.p.

• *Check your answers.*

Section 3: Tables of ungrouped data

In this section you will:
- work out the mean, median and mode of ungrouped data in tables
- use the statistical functions of a calculator to work out the mean

DEVELOPMENT

D3.1: Calculations using tables of ungrouped data

EXAMPLE: This table shows the number of goals scored in football matches by Wurling Wanderers in 2000-1.
Calculate the mean, mode, median and range of the number of goals.

No. of goals (N)	Frequency (F)	N x F
0	2	0
1	3	3
2	7	14
3	5	15
4	2	8
5	0	0
6	1	6
Totals	20	46

total number of matches

total number of goals

Mean = $\frac{\text{total no. of goals}}{\text{no. of matches}} = \frac{46}{20} = \boxed{2.3}$

Mode = [2] [more matches had 2 goals than any other number]

Median = [2] [median score = 10 1/2th score. If you add up the frequency column as a running total, you get 10th and 11th score both as 2]

Range = [6] [6 – 0]

Number of people living in each house on the Meadowlands Estate

No. of people (N)	Freq. (F)	N x F
1	6	6
2	7	
3	8	
4	10	
5	5	
6	3	18
7	0	
8	1	
Totals		

1. Copy and complete this table.
2. How many houses have three people living in them ?
3. How many houses have six people living in them ?
4. How many houses are there on the estate ?
5. What is the total number of people living in the houses ?
6. What is the range of the number of people ?
7. What is the modal number of people ?
8. Calculate the mean number of people. Show how you work it out.
9. Work out the median number of people

Find the mean, median and mode for each frequency table:

10. Number of goals scored by Hobblers F.C.

Score	Frequency
0	6
1	8
2	5
3	3
4	2
5	0
6	1

11. Weights of eggs laid on 1st June

Weight in g	Frequency
65g	3
66g	6
67g	10
68g	10
69g	5
70g	3
71g	2
72g	1

• *Check your answers.*

D3.2: Using calculator functions to find the mean

EXAMPLE Work out the mean of 3 4 4 4 5 8 8 8

To enter 3 lost of 4: *either* key in 4 x 3 and then M+ or DATA (frequency x data)
or key in 4,3 DATA (frequency, data)

Hint: when you have keyed in all the data, call up n (the number of items entered). If n is not the same as the number of items, then you have probably entered frequency, data instead of data, frequency.

RCL $\bar{x}$ gives the mean = 5.5

Use a calculator to work out the mean for each data list or table (to 2 d.p.):

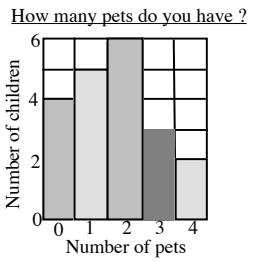
1. 99 99 99 102 104 104 109
2. 51 54 58 58 60 57 43 45 45 45 50
3.

x	23	24	25	26	27	28
frequency (f)	1	3	2	4	5	1

4. Use a calculator to work out the mean of the data in question 10 of D7.1
5. Use a calculator to work out the mean of the data in question 11 of D7.1

• *Check your answers.*

D3.3: How many pets ?



1. Complete this table :

Number of pets	0	1	2	3	4
Number of children	4				

2. How many children were asked about their pets ?
3. What is the modal number of pets ?
4. Calculate the mean number of pets.
5. The median number of pets is 2. Show how you work it out.
6. Three quarters of the children have less than n pets. What is the value of n ?

Section 4: Tables of grouped data

In this section you will:
- estimate the mean of grouped data in tables
- use average values and frequency polygons to compare distributions of discrete data
- draw frequency diagrams for the grouped data

DEVELOPMENT

D4.1: Calculations using tables of grouped data

The method for estimating the mean is the same as that for calculating the mean for ungrouped data – EXCEPT that you use the midpoint as the representative value of each group.

RULE: midpoint = (bottom value of group + top value of group) ÷ 2

Y10 Summer exam: Boys' Marks

Marks	Freq. (F)	Midpoint (M)	F x M
11–20	4	15.5	
21–30	6		
31–40	7		
41–50	7		
51–60	13		
61–70	16		
71–80	7		
Totals			

number of boys

estimated total mark

1. Copy and complete this table.
2. Estimated mean = $\frac{\text{sum of (M x F)}}{\text{total frequency}}$

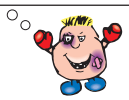
Work out the estimated mean. Show how you do it.

Y10 Summer exam: Girls' Marks

62	21	50	69	51	63	80	34	65	41
60	68	70	23	72	49	66	68	52	62
31	61	47	66	24	70	55	77	28	64
70	54	79	40	56	65	38	63	46	59
65	38	66	78	43	69	69	24	58	67

Why 'estimate' in stead of calculate ?

This is because we are using the midpoint of each group, rather than the exact values.



Ruff

3. Put the girls' marks onto a grouped frequency table. Use the same groups as on the boys' table. Complete the table.
4. The modal group for boys is 61–70. What is the modal group for girls ?
5. How many girls took the exam ?
6. Estimate the mean value of the girls' marks . Show your working.
7. *Complete this table:*

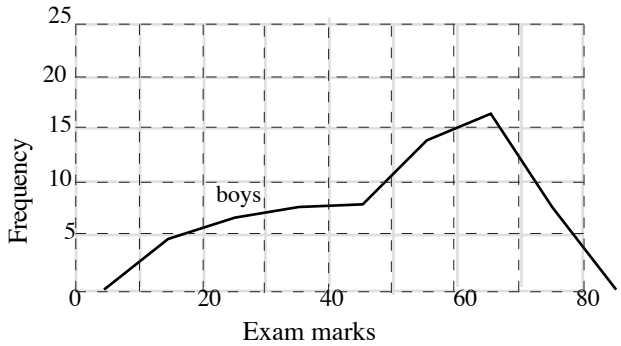
	Boys	Girls
Estimated mean		
Modal group	61–70	
Median group		61–70
Estimated range	69	

8. "On average, the girls did better than the boys."
   Explain why this statement is true.
9. The girls were more consistent than the boys.
   Explain how you can tell this from the table.
10. Which of these three averages best represents this data ?
    Explain why you think it does.
11. Draw a frequency diagram for the girls' marks.
    Remember that the scale across the page must be continuous.

*12. The mean of the whole of Y10's marks (boys and girls) is 52.95, but if you find the mean of the boys' and girls' means, you get 53.87
    Explain why the two values are not the same.

• Check your answers. Make sure that your frequency diagram is correct before you do the frequency polygon in D8.2

D4.2: Frequency polygons



This is a frequency polygon for the Y10 boys' exam marks. The intervals below the first value and above the last value are drawn as having zero frequency.

1. Copy this frequency polygon.
   On the same diagram, draw the frequency polygon for the girls' marks.
2. <u>Use the frequency polygons only,</u> to say whether each of the following statements is true (T), false (F), or impossible to say from the diagram (I):
   - A: The range of marks for the boys is the same as the range for the girls.
   - B: The lowest few marks are all boys' marks.
   - C: More boys than girls scored over 70.
   - D: The modal class is the class with the highest frequency peak.
   - E: The girls did better, on average, than the boys.

• Check your answers.

Section 5: Tables of continuous data

In this section you will:
- do calculations with tables of continuous data
- draw frequency diagrams for continuous data (histograms with equal intervals)
- choose and label correctly groups of continuous data

DEVELOPMENT

D5.1: Calculations using continuous data

Weights of parcels

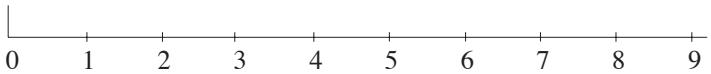
5.3 kg 3.4 kg 3.1 kg 5 kg 6.9 kg 6.1 kg 3.9 kg 4 kg 6.2 kg 7.4 kg
5.9 kg 6.7 kg 7.8 kg 6 kg 5.7 kg 6.4 kg 6.6 kg 4.8 kg 3.3 kg 5 kg

1. Copy and complete this grouped frequency table for the weights of parcels.

Groups	Tally	Frequency	Midpoint	M x F
3 kg ≤ W < 4 kg				
4 kg ≤ W < 5 kg				
5 kg ≤ W < 6 kg				
6 kg ≤ W < 7 kg				
7 kg ≤ W < 8 kg				
Totals				

2. Calculate an estimate for the mean of this data.

3. Draw a frequency diagram (a histogram) for the data in your table.
Use this scale across the page.



• *Check your answers.*

D5.2: Choosing and labelling groups

Points to remember when choosing groups for continuous data

• there must be no gaps between groups • groups must not overlap

$1 \le x < 2$ $2 \le x < 3$	$1 < x \le 2$ $2 < x \le 3$	*but*	$1 \le x \le 2$ $2 \le x \le 3$	*and*	$1 < x < 2$ $2 < x < 3$
is OK	is OK		is not OK		is not OK

Look at $1 \le x \le 2$, $2 \le x \le 3$

1. In which of these two groups would you put (a) 1.9 (b) 2.001 (c) 2 ?
2. Why is this grouping not OK ?

Look at $1 \le x < 2$, $3 \le x < 4$

3. In which of these two groups would you put (a) 1.8 (b) 3.005 (c) 2.7 ?
4. Why is this grouping not OK ?

Look at $1 < x < 2$, $2 < x < 3$

5. In which of these two groups would you put (a) 1.7 (b) 2.2 (c) 2 ?
6. Why is this grouping not OK ?

When doing a survey where data will be grouped:

- decide whether the data is discrete or continuous
- decide on a suitable number of groups [5 – 10 groups is sensible]
- choose your groups [should be of equal width]
- label your groups clearly [top and bottom values must be clearly shown]

DISCRETE DATA:

No. of coins in pocket	Shoe sizes
1 – 5	$1\frac{1}{2}$ – 4
6 – 10	$4\frac{1}{2}$ – 7
11 – 15	$7\frac{1}{2}$ – 10

CONTINUOUS DATA:

Length of pencil in cm	Age
$5 \le x < 10$	$1 \le x < 4$
$10 \le x < 15$	$4 \le x < 7$

It is more usual to include the bottom value and not the top value.

7. In a Youth Club survey, members were asked
" How old are you to the nearest year ?"
Some of the answers given were: 14, 13, 13, 12, 11, 12, 13, 14, …

Which two of these tables would it be correct to use for this survey ?

Table A

Answer	Group	Freq.
11	10.5 ≤ Age < 11.5	
12	11.5 ≤ Age < 12.5	
…	……	

Table B

Answer	Group	Freq.
11	10.5 – 11.5	
12	11.5 – 12.5	
…	……	

Table C

Answer	Group	Freq.
11	11 ≤ Age < 12	
12	12 ≤ Age < 13	
…	……	

Table D

Answer	Group	Freq.
11	11 – 12	
12	12 – 13	
…	……	

Table E

Answer	Group	Freq.
11	At least 10.5 but less than 11.5	
12	At least 11.5 but less than 12.5	
…	……	

Table F

Answer	Group	Freq.
11	At least 11 but less than 12	
12	At least 12 but less than 13	
…	……	

8. The question is changed to "How old are you in completed years ?"
Which two of the tables would it be correct to use for this survey ?

• Check your answers.

PRACTICE

P5.3: Cormorant watching

Becky was sitting on a rock at the edge of the sea watching a cormorant diving for fish. Her watch could be used as a stopwatch, so she decided to measure the length of time that the cormorant was underwater for each dive.

Length of time of dive				
7.1 s	31.1 s	18.3 s	43 s	25.3 s
48.3 s	8.8 s	60 s	40 s	47.1 s
56 s	36 s	63.2 s	7.2 s	23 s
24.2 s	21.5 s	39.7 s	32 s	50 s
11.3 s	45.4 s	30.2 s	47.7 s	13.8 s
38.6 s	18.2 s	26.4 s	45.1 s	40 s
30 s	50.1 s	25 s	15.8 s	41.4 s
22.9 s	33 s	12.6 s	27 s	23 s

1. *Copy and complete this grouped frequency table for the diving times.*

Groups	Tally	Frequency	Midpoint	M x F
0 s ≤ t < 10 s 10 s ≤ t < 20 s …… ……				
Totals				

2. What is the modal group of diving times ?
3. Calculate an estimate for the mean time. Show your working.
4. In which group does the median time lie ?
5. Give an estimate for the range of times.

• *Check your answers.*

Alternative notation

Some questions may ask you to put data into **groups**.
Some questions may ask you to put data into **classes**.
They are the same instruction.
Data groups and data classes are the same.

EXTENSION

Extension problems put able students on the path towards A* at GCSE and a high grade at A-Level.

E5.4: *Bat survey

Naturalists did a survey of two colonies of bats in East Sussex.

The **pipistrelle** is the smallest British bat. Its head and body together measure only around 4 cm. In contrast, its wingspan is 18 – 23 cm. It is often found roosting inside roof cavities of houses, particularly the more modern ones, as it prefers the warmth of these.

The **common long-eared bat** is a little larger than the pipistrelle. Its head and body measure 3.7 – 4.8 cm. However, its name is an understatement. Its ears are almost as long as the rest of its body. It has the most sensitive hearing system of all British bats and likes to forage in woodland areas.

The naturalists watched the bats, observing their behaviour. They also captured 100 bats in each colony and took various measurements before releasing them. Back in the office, a clerk collated all the data that had been collected. Amongst the data were these two tables of weights.

Table A

Weights in g	Frequency
5.0 ≤ W < 5.5	3
5.5 ≤ W < 6.0	10
6.0 ≤ W < 6.5	21
6.5 ≤ W < 7.0	35
7.0 ≤ W < 7.5	19
7.5 ≤ W < 8.0	5
8.0 ≤ W < 8.5	5
8.5 ≤ W < 9.0	3
9.0 ≤ W < 9.5	1
9.5 ≤ W < 10.0	0

Table B

Weights in g	Frequency
5.0 ≤ W < 5.5	5
5.5 ≤ W < 6.0	11
6.0 ≤ W < 6.5	10
6.5 ≤ W < 7.0	18
7.0 ≤ W < 7.5	16
7.5 ≤ W < 8.0	15
8.0 ≤ W < 8.5	10
8.5 ≤ W < 9.0	7
9.0 ≤ W < 9.5	4
9.5 ≤ W < 10.0	1

Unfortunately, there was no label on either table to say which set of bats it referred to.

Your task: You are the clerk. You want to get all this paper work finished today, as you go on holiday tomorrow. You can't contact the naturalists so you have to sort it out yourself. In case you are wrong, you need to be able to quote figures to justify your decision.

You know that the pipistrelle is slightly smaller than the long-eared bat. Work out which table is which. Write a note explaining to the naturalists how you worked it out. This will be attached to the sets of data – just in case you are wrong.

• *Your teacher has the answers to this.*

Section 6: Questionnaires and bias

In this section you will:
- think about what is meant by 'bias'
- think about how to make unbiased statements

DEVELOPMENT

D6.1: A question of bias

Ideal for small group work

The way a question is asked can affect the answers given to it.

A question that has no effect on the answer is called **a neutral question**.

A question that makes the interviewee more likely to answer in a particular way is called **a biased question.**

Here we look at two major causes of bias:

Leading questions
"Do you think there is too much violence on TV ?"
This is more likely to get the response "Yes" than "No".

Emotive questions
"What do you think about people who kill defenceless animals to make fur coats?"

1.

Q	Leading	Emotive	Leading and Emotive	Neutral
A B :				

Draw up a table like this.
For each question below, put a tick under the heading that best describes it:

A: Do you think that there are too many repeats on TV ?

B: What do you think about the number of TV repeats? Too many or about right?

C: It has been suggested that all 18 year olds should serve in the army for two years. Do you agree ?

D: Young people today have no sense of discipline. Don't you agree that two years in the army for every 18 year old would make them much better citizens ?

E: CFCs cause damage to the ozone layer. Don't you think they should be banned ?

F: It is thought that CFCs may cause damage to the ozone layer.
Do you think that:
we should stop using CFCs as soon as possible ☐
or more research should be done to find out how harmful they are ☐
or stop using CFCs *and* do more research ☐

2. *Rewrite each of these question as a neutral question:*

   P: Do you think that school uniforms ought to be abolished ?

   Q: All murderers deserve to hang. Do you agree ?

   R: Cannabis is less addictive than tobacco and ought to be legalised. Do you agree ?

Another form of bias is to ask people who are more likely to give just one of the answers.

3. *Explain why each of the following is likely to result in a biased survey:*

Survey W:
" What do you think about the amount of sport on TV ?
too little ☐ too much ☐ about right ☐ ?"
People asked: crowd coming out of football match.

Survey X:
" What do you think about the amount of sport on TV ?
too little ☐ too much ☐ about right ☐ ?"
People asked: people shopping on a Saturday afternoon.

Survey Y:
" Is the state pension enough to live on ?
too little ☐ too much ☐ about right ☐ ?"
People asked: pensioners.

Survey Z:
" Is the state pension enough to live on ?
too little ☐ too much ☐ about right ☐ ?"
People asked: diners at the Dorchester Hotel in Park Lane.

4. Izzy asked the head if she could do a survey in school on fox-hunting.
Her question was:

"Don't you think that fox-hunting should be banned ?"

The head said that she could not use this question, as it was a leading question.
If she could prepare a neutral question to ask, with at least three possible responses, then she could have permission to do the survey.

Your task: design a question that will be acceptable to the head.

5. Design a neutral question, with at least three responses, for a survey on whether P.E. ought to be compulsory.

• *Check your answers. Your teacher will need to check Tasks 4 & 5.*

Section 7: Testing hypotheses

In this section you will:

- investigate different ways of testing hypotheses
- look at possible causes of bias and how to avoid them

DEVELOPMENT

D7.1: Methods for testing hypotheses

A hypothesis is a statement. It is either true or false.

In order to find out which it is, we **test the hypothesis**.

There are several methods of testing hypotheses.

Any method of testing must be unbiased.

Method 1: Questionnaire

This is used to test a hypothesis that is based on opinion, rather than fact.
To prevent bias: the question asked must be neutral and care must be taken over who is asked.

1. **Hypothesis**: "Most people in this school think stag-hunting is cruel"

(a) Design a neutral question that will test this hypothesis.
It must have at least three possible responses.

(b) It is not possible to ask everyone in the school. Should you:
A: ask everyone in one class ?
or B: ask some from each year group ?

(c) How would make sure that the sample you asked was representative of all the students in the school ?

Method 2: Survey using observation sheet

This is used to test a hypothesis that is based on fact, rather than opinion.
To prevent bias: care must be taken over who is asked.

2. **Hypothesis**: Older pupils in the school watch more TV than younger pupils

(a) Design an observation sheet that will test this hypothesis.

(b) It is not possible to ask everyone in the school. Should you:
A: ask everyone in one class ?
or B: ask some from each year group ?
or C: ask some from the oldest and youngest year groups ?

Do you know what is meant by primary data or secondary data ?
Primary data comes directly from the original source. If it is edited, commented on, added to, summarised or … it becomes **secondary data.**
A student who fills in an observation sheet on the kinds of cars passing the school gate is collecting primary data. When the data is organised or displayed in a diagram, it becomes secondary data.

Icee

Method 3: Using reliable sources of secondary data

These are used to test a hypothesis that is based on facts about people in general, rather than a small group of people.

To prevent bias: care must be taken to use all relevant available statistical facts to back up any assertions.

For reliable national data, suggested sources are "Regional Trends", "Social Trends" and the "Annual Abstract of Statistics" (all produced by the Office for National Statistics and available in the reference section of most libraries). The Office for National Statistics' website (www.ons.gov.uk) is just one of many internet sites where useful statistical data may be found.

3. **Hypothesis**: The Midlands is the most densely populated region in the UK, and Wales is the least densely populated region.

Use the data from "Regional Trends" on the next page to test this hypothesis.
Say whether you think the hypothesis is true or false.
Explain why, using figures from the data.

Method 4: Experiment

This is used to test a hypothesis about experimental outcomes.

To prevent bias: care must be taken to make the experiment unbiased.

4. **Hypothesis**:If there are 35 red counters and 30 black counters in a box and one is taken out unseen and then replaced, you are more likely to get a red counter than a black each time.

Here are three ways of conducting this experiment. One should give a set of unbiased results. The other two ways are likely to produce experimental bias.

Explain which two are likely to be biased – and why ?

Expt 1: the box of counters is shaken between each selection and the person making the selection does not look into the box when making the selection.

Expt 2: the person who is making the selection looks at the counters as they are choosing them.

Expt 3: the box of counters is not shaken between each selection

5. A spinner made of a hexagonal piece of card and a match is spun repeatedly. The six edges are numbered 1, 2, 3, 4, 5, 6.
Hypothesis: You are more likely to get an odd number than an even number.

Bob took care that the match was always vertical and the spinner horizontal, before each spin. He spun his spinner 100 times.

Ellen spun her spinner 20 times and then made a new one. So, she used three spinners to do 60 spins.

Explain why Bob's results would be biased and Ellen's results less likely to be biased.

[If you don't know, make a spinner and watch what happens to it after repeated spins]

• *Check your answers.*

EXTENSIONS

E7.2: National hypotheses

Here are three hypotheses that you can test using the data from "Regional Trends" below. Test one or more of these hypotheses. Give clear explanations backed by data.

Hypothesis A: People earn more in the south than they do in the north.

Hypothesis B: Unemployment is higher in the north than in the south

Hypothesis C: People who earn the most, spend the most.

Data source: Regional Trends

Regional Variations in the UK (1992)

For purposes of economic comparison, the UK is split up into 11 regions: North, Yorkshire & Humberside, East Midlands, East Anglia, South East, South West, West Midlands, North West, Wales, Scotland, Northern Ireland.

In the table below, all figures represent thousands, unless you are told otherwise.

	North	Y & H	E. Mids	E. Anglia	S.E.	S.W.
Area in 1000 km^2	15.40	15.42	15.63	12.57	27.22	23.84
Population	3 099	5 002	4 062	2 087	17 703	4 746
Total work force	1 500	2 400	2 000	1 000	9 000	2 300
Unemployment	11.2%	9.9%	8.7%	7.1%	9.4%	9.1%
Household income	£321	£316	£340	£374	£455	£373
Household expenditure	£236	£241	£273	£277	£320	£270
Offences recorded	13.63	12.57	11.38	8.55	10.59	9.44
Clear up rate	24%	26%	28%	32%	20%	22%
Number of police	7.65	11.51	8.37	3.93	49.54	9.71

	W. Mids.	N.W.	Wales	Scotland	N.I.	
Area in 1000 km^2	13.01	7.33	20.77	77.08	13.48	
Population	5 278	6 400	2 899	5 111	1 610	
Total work force	2 600	3 200	1 300	2 500	700	
Unemployment (%)	10.7%	10.1%	8.9%	9.5%	12.1%	
Household income (£)	£335	£328	£311	£325	£277	
Household expenditure	£237	£253	£251	£238	£257	
Offences recorded	10.46	11.83	9.94	11.54	4.19	
Clear up rate	28%	36%	35%	32%	34%	
Number of police	12.19	16.82	6.48	14.09	11.64	

Section 8: Scatter Graphs

In this section you will:
- interpret scatter graphs
- develop an understanding of correlation
- use a 'line of best fit' to estimate missing values

DEVELOPMENT

D8.1: Is there a relationship ?

Small group or class discussion

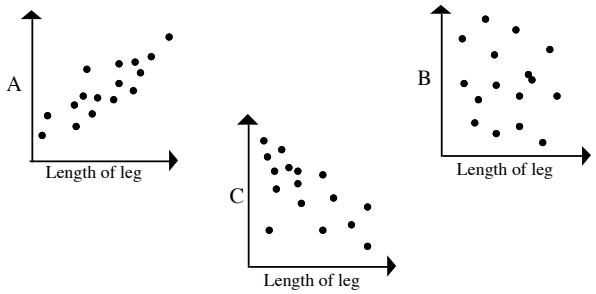
You have worked with graphs of two related variables. You have used the relationship between the variables to draw a graph and used a graph to work out the relationship between two variables.

Here we plot values of two variables and use the resultant scatter graph to decide whether there is a relationship between them.

A group of 14–year–olds collected four sets of data for their class:

(1) length of foot (2) length of leg
(3) time taken to run 100 m (4) time taken to get to school.

They plotted this data onto these three graphs:



The missing labels (A, B and C) are each one of these:

length of foot | time taken to run 100m | time taken to get to school

Discussion points

Which label goes on which graph ? How can you tell ?

From the scatter graphs, is each statement true or false ?

- people with long legs usually have long feet
- people who have long legs can run faster
- there is no correlation between the length of leg and the time taken to get to school.

D8.2: Scatter graphs and correlation

Types of correlation

perfect positive correlation

strong/good positive correlation

weak/poor positive correlation

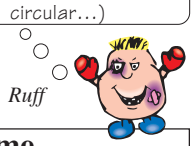
no apparent linear correlation

perfect negative correlation

strong/good negative correlation

weak/poor negative correlation

In each case, we are talking about linear correlation. It is quite possible that there could be another form of correlation (quadratic, circular…)



Plot scatter graphs for each of the following sets of data. Decide what kind of correlation you have in each case. Write a sentence saying what this tells you about the relationship between the data.

Case 1: Average number of eggs and incubation time

Bird	No. of eggs	Incubation time	Bird	No. of eggs	Incubation time
Blackbird	4.5	15	Falcon	3.5	30
Jackdaw	5	18	Mallard	10.5	26
Puffin	1.5	42	Thrush	4.5	14
Vulture	1	51	Woodpecker	5	11
Blue tit	10	15	Robin	6	14
Golden eagle	1.5	43	Stork	4	30
Gannet	1	44	Razorbill	1	36

Case 2: Football: goal difference and points

Team	Goal difference	Points	Team	Goal difference	Points
Man Utd	5	21	Arsenal	15	26
Ipswich	-10	9	Leeds	8	24
Liverpool	12	29	Chelsea	8	23
Sunderland	-2	19	Aston Villa	4	23
Charlton	-2	15	Newcastle	6	24
Tottenham	3	21	Southampton	-11	10

Case 3: Gestation period (days) and average life span (years)

Animal	Gestation	Av. life span	Animal	Gestation	Av. life span
Antelope	245	17	Badger	180	12
Horse	325	50	Hare	42	8
Giraffe	440	28	Camel	380	35
Cat	60	17	Lion	106	35
Mouse	21	3	Donkey	375	45
Elephant	640	65	Orang-utan	260	50

D8.3: Using a line of best–fit

Score on Paper 1	45	51	67	34	65	60	56	73	34
Score on Paper 2	38	50	61	33	58	55	50	74	30

1. Draw a scatter graph of these two sets of exam scores.
   Draw a line of best–fit through the points.
   [Try to get as many points above as below the line.]
2. Farzad got 50% on Paper 1. On the day of Paper 2, he was involved in a car accident on the way to school. The exam board used a line of best–fit to estimate the score he would have got on Paper 2.
   Use the line of best–fit to estimate the score you think he should get.
   Explain how you do it.
3. Ann scored 41% on Paper 2. Estimate the mark you think she would have got on Paper 1.
4. Paul got 60% on Paper 1. Unfortunately, he had migraine during Paper 2. He scored 51%.
   What mark do you think he might have got, if he had been well ?
   Is it worth appealing to the Exam Board ?
5. Use the scatter graph you drew for "Gestation period and average life span"
   Draw a line of best fit on it.
   A human has a gestation period of 275 days. Use the line of best fit to estimate the average life span of a human.

• *Check your answers.*

EXTENSIONS

E8.4: Testing hypotheses using scatter graphs

Type 1: Using data in "Regional Trends" on page 224

Hypothesis 1: The crime rate is higher in areas of high unemployment.
[Plot "Offences recorded" against "Unemployment".]

Hypothesis 2: A large police force acts as a deterrent to potential criminals

Hypothesis 3: A larger police force would give a better crime clear-up rate.

Type 2: Collect your own data

Hypothesis 4: Taller people have bigger feet.
[Plot "height" against "area of foot".]

Type 3: All your own work

Hypothesis 5: ?? [Choose your own hypothesis and collect your own data]

Choose one or more hypotheses.
Find the data.
Use a scatter graph to test the hypothesis.
Optional : you could use a line of best-fit and invent your own questions to answer with it.
Write a short paragraphs summarising what you have deduced.

Section 9: Controlled experiments

In this section you will:
- learn the terminology of controlled experiments
- analyse the results statistically of one or more of these controlled experiments

DEVELOPMENT

D9.1: Controlled experiments

Terminology

In a **controlled experiment**, one of the variables being considered is dependent on the the other variable. The independent variable is called the **controlled variable**. Its values are chosen (or controlled). The other variable is the **dependent variable**.

In a scatter diagram, the controlled variable is always along the bottom of the diagram.

Perform one or more of the following experiments:

Experiment 1: Estimating time intervals

Seconds are counted as "one plus, two plus, three plus …"

The distance of a storm from an observer can be estimated by counting the seconds between the flash of lightning and the thunder clap. It takes approximately 5 seconds for the sound to cover 1 mile. The teacher signals when the lightning flashes and bangs when the thunder is heard. The student estimates how far away the storm is.

There are two variables : the time between the 'lightning' and the 'thunder' as measured by the teacher and the time between the 'lightning' and the 'thunder' as estimated by the student.

(a) Which of these variables is the controlled variable ?

(b) Plot these two variables on a scatter graph. Draw a line-of-best-fit.

(c) Invent some questions and answer them using the scatter graph.

Experiment 2: Finding the balance point

A mass is hung from one end of a ruler.

The balance point is found and its distance measured from the other end of the ruler.

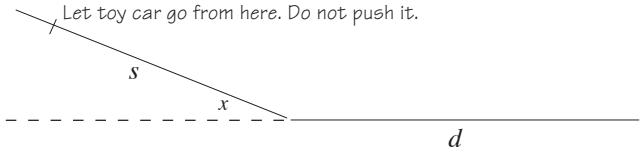
There are two variables: the mass and the distance of the balance point from the end of the ruler.

(a) Which of these variables is the controlled variable ?

(b) Plot these two variables on a scatter graph. Draw a line-of-best-fit.

(c) Invent some questions and answer them using the scatter graph.

Experiment 3: How far does it go ?



A toy car is allowed to roll from a point on a ramp inclined at $x°$ to the horizontal.
The distance it travels along the ground at the bottom of the ramp is measured.
There are three variables: the distance of the start from the bottom of the ramp
the distance along the ground
the angle of the ramp to the horizontal.

(a) Which of these three variables is the dependent variable ?

(b) One of the other two variables can be the controlled variable.
The other measurement must be kept fixed.

(c) Invent your own experiment(s) using the car and the ramp. Display and analyse your results statistically.

Experiment 4: The journey to work

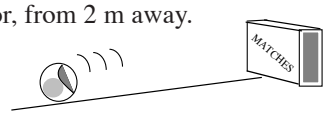
Choose someone in your family that goes to work every day in a car. Record the time that they leave home every day for at least 10 working days, and the corresponding time that they reach work.

Investigate the connection between the two times.
Display and analyse your results statistically.

EXTENSION *Groups of 3-6*

E9.2: And now for something different !

Task 1: A matchbox is placed on the floor against a wall.
Shots are aimed at it, rolled along the floor, from 2 m away.
Each member of the group :
- has one set of four shots;
- records how many hits out of 4;
- repeats this set of four shots 25 times.



Task 2: Calculate the mean, median, mode and range for each member of the group.

Task 3: Draw a compound bar chart to illustrate the frequency of each score for each member of the group.

Task 4: Decide who is the best shot and who is the worst shot.
Explain what you used to make the decision.

Task 5: How could you make this into a controlled experiment ?

Section 10: Time series graphs

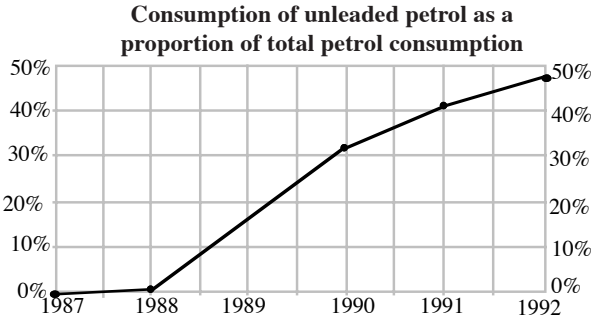
In this section you will:
- work with time series, trends and seasonality
- learn how to use moving averages to smooth time series

DEVELOPMENT

D10.1: Trends in use of unleaded petrol

The • give the percentage of unleaded petrol used in the previous year.

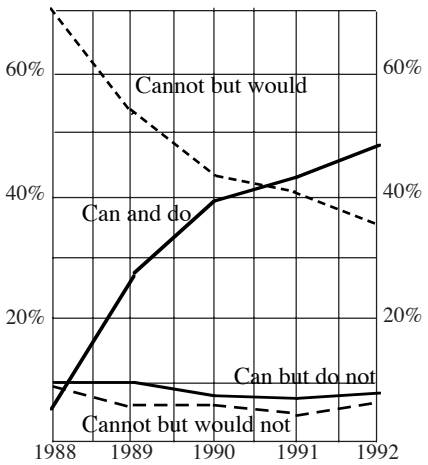
The values in between these data points have no meaning. The lines are there to draw your eye from one point to the next. They help you to visualise the trend in the data.



– – – – – are often used instead of ———— to indicate links rather than data.

1. You will notice that no value has been put in for 1989.
   Estimate the consumption of unleaded petrol in 1989.
2. Assuming the present trend continues, estimate what the percentage consumption of unleaded petrol in (a) 1993 (b) 1994.
3. Estimating values between two given data points is called "interpolation" [see Q1]
   Estimating values beyond the given data points is called "extrapolation" [see Q2]
   Interpolation is usually more accurate than extrapolation.
   Explain why you think this is so ?

   This multiple line graph shows the trends for four different groups of people.



4. Add together the data for "Can & do" and "Cannot but would do" to produce a single line graph for "Are willing to" for 1988 – 1992. Leave room to extrapolate the data to 1993 and 1994.
5. Unless there is a change in outlook, what looks like being the maximum percentage use of unleaded petrol ?
6. Suggest ways in which the UK could eventually get 100% use of unleaded petrol.

• *Check your answers.*

D10.2: Analysis of time series

A **time series** is a set of observations taken at specified times, usually at equal intervals. Examples of time series are :

- the amount of electricity used every quarter
- the temperatures taken hourly for one place
- the sales figures for a department store
- …

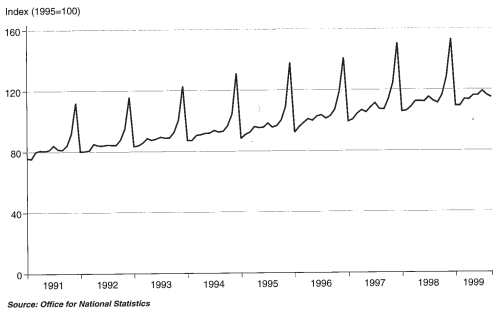
The graph of a time series will probably have several underlying features:

- the long term trend
- the long term trend with cyclical movement

[Business cycles can represent intervals of prosperity, recession, depression and recovery. Cycles are longer than one year.] We are not going to consider these here.

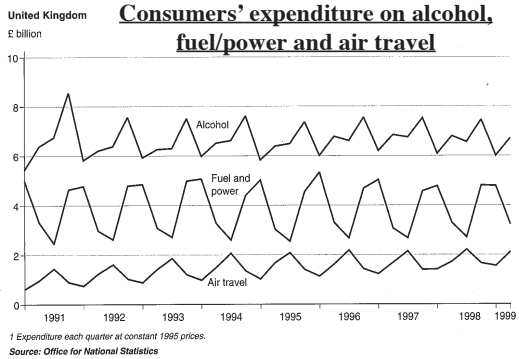
- the long term trend with seasonality [see the diagram below]

Volume of retail sales in Great Britain



1. Is the long term trend shown here increasing, decreasing or fluctuating ?
2. After what period of time does the pattern almost repeat ?
   This time period is the seasonal variation (or seasonality).
3. The most obvious seasonal feature here is the repeating peak.
   At what time of the year does it always occur ? What do you think causes it ?
4. What economic condition could cause a change in the long term trend ?
5. If there was a downturn in the long term trend, what do you think would happen to the peaks ?

Consumers' expenditure on alcohol, fuel/power and air travel



6. For each of these three time series state:

(a) how long the seasonal variation is.

(b) when in the year the peaks occur and why they occur at that time.

(c) when in the year the troughs occur and why they occur at that time.

• *Check your answers.*

D10.3: Moving averages

Moving averages are used to smooth time series.
They tend to reduce the amount of variation, enhancing the underlying trend.

EXAMPLE Find the 3 point moving averages for this set of data.

	7	6	7	8	9	7	8	8	9	8	10
3 point averages are:		6.7	7	8	8	8	7.7	8.3	8.3	9	

$\frac{7+6+7}{3}$ $\frac{8+9+7}{3}$

Each *n*-point moving average should be written, (or plotted on a time series graph, if required) under the middle of the set of *n* points.
If *n* is odd, the average goes under the middle value.
If *n* is even, the average goes in between the two middle values.

1. (a) Find the 4 point moving averages for the set of data in the example.
   (b) On one diagram, draw time series graphs in two different colours for the set of data and the moving averages.

2. The number of absentees in 10BY tutor group for the last three weeks were:

   5 4 4 3 2 | 4 2 2 1 3 | 6 6 4 2 0

   (a) Work out the 5 point moving averages for this data.
   (b) Draw a time series graph for the number of absentees.
   (c) On the same diagram, draw a time series graph for the moving averages. Use a different colour.

3. Schools are obliged to publish, each year, the percentage of A*-C passes. Originally it was suggested that this percentage should increase year on year. However, the standard of students varies from year group to to year group, with some year groups much better or worse than other year groups. So, it was suggested that inspectors, or other interested parties, should actually look at the 3 point moving average for the annual results.

   (a) Find the set of 3 point moving averages for this set of A*-C percentages:

   48 49 55 70 75 67 65 68 71 70

   (b) Looking at these moving averages, would you think that an inspector would feel that standards were being maintained.

4. In D15.1, there was a time series graph showing consumers expenditure on airtravel. The quarterly figures for 1995-1998 are (in £ billion):

   1 1.85 2.1 1.4 | 1.1 1.8 1.4 1.2 | 1.75 2.1 1.4 1.4 | 1.85 2.2 1.7 1.55

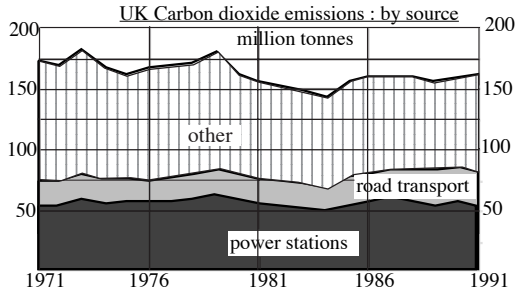
   (a) Find the set of 4 point moving averages for this data.
   (b) How would you describe the underlying trend in expenditure on air travel shown by the moving averages ?

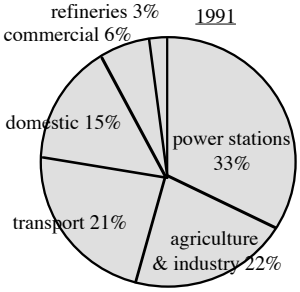
• *Check your answers.*

EXTENSIONS

Extension problems put able students on the path towards A* at GCSE and a high grade at A-Level.

E10.4: *The greenhouse effect





1. The largest contribution to carbon-dioxide emissions is labelled 'Other' on the compound line graph. List the items that are included in Other'.
2. *Copy this paragraph. Use the information given to fill in the missing figures.*

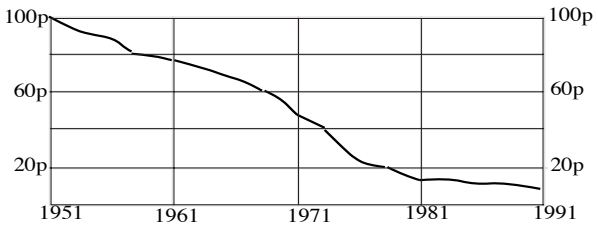
   The increase in carbon dioxide (CO_2) in the atmosphere is largely due to burning fossil fuels and destroying forests. Forests remove CO_2 from the atmosphere. CO_2 is a major contributor to the greenhouse effect. It is one of the gases that allows sunlight through the atmosphere to the earth, but it then prevents the heat escaping. The largest source of CO_2 in 1991 was fossil fuel burning power stations. They produced around .....million tons of CO_2, one third of all emissions. Between 1971 and 1991, overall emissions of CO_2 have fallen by .....%. Emission from motor vehicles have doubled to ..... milion tonnes from 1971 to 1991. This was ..... of all emissions in 1991. Domestic use of fuel put ..... million tonnes of CO_2 into the atmosphere in 1991.

• Check your answers.

E10.5: *Purchasing power of a 1951 pound

High inflation causes the purchasing power of the pound to drop rapidly.

Say whether each of these statements is TRUE or FALSE:



A: The purchasing power of the pound had halved between 1951 and 1971.

B: By 1971 it cost £2 to buy what £1 had bought in 1951.

C: The purchasing power of the 1971 pound halved by 1979.

D: In 1992, it cost £14 to buy what £1 had bought in 1951

E: Between 1981 and 1992 the UK had had high inflation.

• Check answers.

Section 11: Cumulative frequency

In this section you will:
- meet / review cumulative frequency
- draw and read information from cumulative frequency diagrams
- display information using box-and-whisker diagrams

DEVELOPMENT

D11.1: Running totals

1. This table shows the marks the Pan–Galactic trainees got in their Space Navigation exam.

   Copy the table and complete the running totals column.

Space Navigation Exam

marks		frequency	running totals
at least	below		
10	20	3	3
20	30	12	15
30	40	22	37
40	50	36	
50	60	38	
60	70	20	
70	80	7	138

2. How many trainees took the exam ?

3. 

   Optymistic

   Optymistic said that 12 trainees had got less than 30 marks.
   Pesymistic said that 15 had got less than 30 marks.
   Pesymistic was right.
   Explain why.

   Pesymistic

4. How many trainees got less than (a) 50 marks (b) 70 marks ?
5. The pass mark was 40 marks. How many trainees failed the exam ?

The table below describes the results of the Galactospeak exam.

Galactospeak Exam

marks		running totals
at least	below	
10	20	2
20	30	26
30	40	47
40	50	86
50	60	118
60	70	134

6. How many trainees sat this exam ?
7. How many trainees got less than 40 marks ?
8. How many trainees got between 30 & 40 marks
9. How many got 60 marks or more ?
10. How many got 40 marks or more ?
11. How many trainees got between 20 & 40 marks ?
12. Between which two values does the median mark lie ?

• *Check your answers.*

The correct mathematical term for a set of running totals is
cumulative frequency
A lot more information is available from cumulative frequencies, if they are put onto a cumulative frequency diagram.
A **cumulative frequency diagram** is obtained by plotting (the upper class boundary) against (the cumulative frequency.)
The points are usually joined with straight lines.

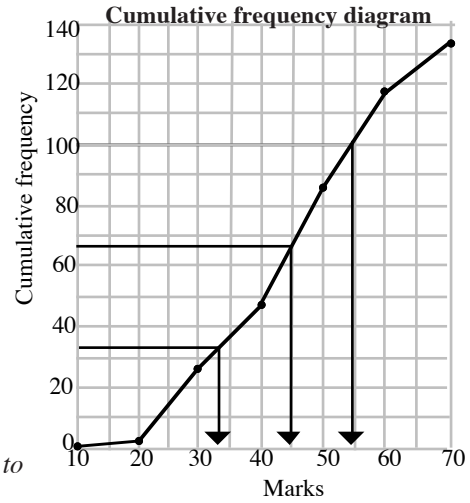
D11.2: Cumulative frequency diagrams

Galactospeak Exam

marks		cumulative
at least	below	frequency
10	20	2
20	30	26
30	40	47
40	50	86
50	60	118
60	70	134

upper class boundaries

The cumulative frequencies are plotted against the upper class boundaries.

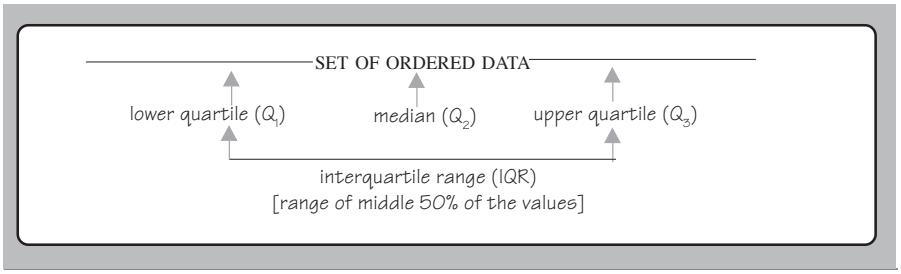


Use the cumulative frequency diagram to answer each of these questions.

1. From the graph, we can estimate that 102 students got less than 55 marks. Explain how it is done.
2. Estimate how many trainees got less than 50 marks.
3. Estimate how many trainees got 50 marks or more.
4. Estimate what mark the 120th trainee from the bottom got.

From the cumulative frequency table, you can work out which class the median is in. A more accurate estimate for the median can be found from the **cumulative frequency diagram**. There were 134 examinees. The middle student is approximately the 67th.

5. Estimate the median mark (the mark of the 67th student).
6. The value one quarter from the bottom is the **lower quartile**. Estimate the value of the lower quartile (the mark of the 33.5th student).
7. The value one quarter from the top is the **upper quartile**. Estimate the value of the upper quartile (the mark of the 100.5th student).
8. The **interquartile range** is the range between the upper and lower quartiles.



9. Draw a cumulative frequency diagram for the results of the Space Navigation exam. Use the upper class boundary and the cumulative frequency for each class of marks.
10. How many trainees took this Space Navigation exam ?

Space Navigation Exam

marks		frequency	cumulative frequency
at least	below		
10	20	3	3
20	30	12	15
30	40	22	37
40	50	36	73
50	60	38	111
60	70	20	131
70	80	7	138

11. Use the cumulative frequency diagram to estimate values for the median mark, the lower quartile (Q_1) mark and the upper quartile (Q_3) mark.
Draw arrows on the diagram to show your working out.
12. Work out an estimate for the interquartile range (IQR).
13. (a) Draw a cumulative frequency diagram for the heights of students in Y10.
(b) Use the graph to estimate values for the median and the upper and lower quartiles. Show clearly on your graph how you work them out.
(c) Work out an estimate for the interquartile range.

Heights of Y10 students in cm

Heights	Freq.	Cu. Freq.
$120 \le h < 130$	7	7
$130 \le h < 140$	20	27
$140 \le h < 150$	27	54
$150 \le h < 160$	42	96
$160 \le h < 170$	51	147
$170 \le h < 180$	38	185
$180 \le h < 190$	11	196
$190 \le h < 200$	4	200

• *Check your answers.*

D11.3: But there are gaps between these groups !!

A group of sixth formers each measured their temperature to the nearest 0.1°C.
The results that they recorded are given in the first two columns of this table.
But, the upper class boundary of the first class is NOT 36.1
The upper class boundary of the first class is 36.15.
In cases like this, it is helpful to make a column that shows all
the members of each class. We call this the true class interval.

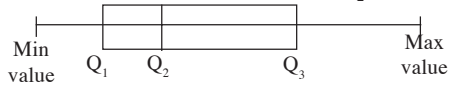
1. Copy and complete this table.
2. Draw a cumulative frequency diagram for the temperatures. Remember that the upper class boundaries are the upper values of the true class intervals.
3. Calculate values for the median, lower and upper quartiles and the interquartile range, showing all working on the diagram.

body temp.(°C)	freq	cumulative frequency	true class interval
36.0 – 36.1	3	3	$35.95 \le T < 36.15$
36.2 – 36.3	2		
36.4 – 36.5	8		
36.6 – 36.7	14		
36.8 – 36.9	29		
37.0 – 37.1	18		
37.2 – 37.3	5		
37.4 – 37.5	1		

• *Check your answers.*

D11.4: Box-and-whisker plots (a.k.a box plots)

A box-and-whisker plot is a diagram which shows the relative positions of the extreme values of the data and the median, Q_2, and quartiles, Q_1 & Q_3.



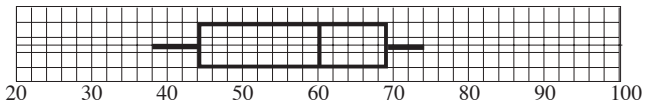
EXAMPLE In the Y10 exams, Ellie's results were:

38% 55% 60% 62% 74% 42% 68% 46% 70%

Draw a box-and-whisker plot for these results.

Ordered results: 38% 42% 46% 55% 60% 62% 68% 70% 74%

Min value = 38 Q_1 = 44 Q_2(median) = 60 Q_3 = 69 Max value = 74



1. (a) Copy the box-and-whisker plot for Ellie's results (above).
   Below it, draw a box-and-whisker plot for Aftab's results in the same exams:
   46% 50% 68% 38% 95% 76% 50% 60% 52% 56%

   (b) Using the two box-plots only, say whether each of the following statements is true (T) or false (F):
   A: The range of Aftab's results is greater than the range of Ellie's results.
   B: The IQR of Aftab's results is greater than the IQR of Ellie's results.
   C: The middle 50% of Ellie's results is evenly spaced about the median.
   D: The middle 50% of Aftab's results is evenly spaced about the median.
   E: The median for Aftab's results is skewed towards the minimum value.
   F: Overall, Ellie did better than Aftab.

2. A class of Y2 children were timed running over 40m.
   The times were measured in seconds.

	Boys	Girls
Shortest time	15	18
Longest time	33	30
Q_1	20	22
Q_2	23	24
Q_3	26	27

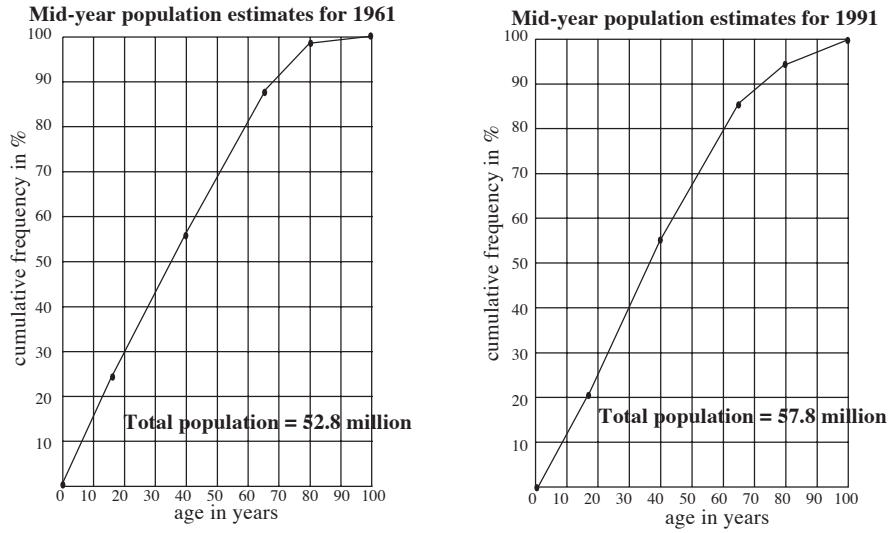
   (a) On squared paper and using the same scale for both, draw box plots to represent this data.
   (b) Compare the results for the boys and the girls.

• *Check your answers.*

EXTENSION

E11.5: Trends in Britain's population

The two cumulative frequency diagrams show the mid-year population estimates for 1961 and 1991. It is on figures like these that the government bases its planning for such services as education, hospitals, care for the elderly …



Mid-year population predictions for 2021						
	less than 16	at least 16 & under 40	at least 40 & under 65	at least 65 & under 80	at least 80 & under 95	pop. in millions
2021	18.5%	30.0%	32.3%	14%	5.2%	62.0

1. Draw a cumulative frequency diagram for the predicted population in 2021.
2. Copy the box-and-whisker diagram for 1961. Alongside it draw two box-and-whisker diagrams for 1991 and 2021.
3. *Use the diagrams to say whether each statement is true or false*:

A: The spread of the middle 50% of the population remains fairly constant.

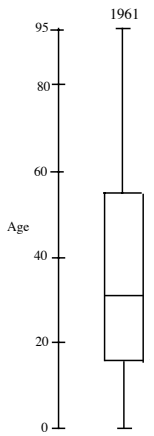
B: By 2021, a quarter of the population will be over 60.

C: The age range of the top quarter of the population is steadily decreasing.

D: The number of elderly people is steadily increasing.

E: The age range of the bottom quarter of the population is increasing.

F: The number of young people is decreasing.



• *Check your answers.*

Section 12: A simple survey

In this section you will design and carry out a simple survey.

DEVELOPMENT

D12.1: Design a simple survey

Imagine that you and your friends are about to launch themselves as a rock band.

Devise four possible names for the band.

You are going to design a survey of students to find which name they think would be the most effective.

The task for your group : You are going to do surveys of two distinct groups of students in your year (boys and girls or musicians and non-musicians or…?) *or* one group from your year and one group from the year above or below.

Points you need to consider before you start:

- why is it not a good idea to survey everyone in Y10 ?
- how big a sample should you choose ?
- how would you choose the sample so it will be representative of the whole group ?
- how should you choose who will be in the sample ?
- should both samples be the same size ?
- what will you do if someone says "I don't like any of them." or "I can't choose." ?

After you have done the survey, each member of the group should write his/her own account of the survey.

Your account should:

- explain in detail what you decided to do
- explain the reasons behind each of the decisions you had to make
- explain how you carried the survey out
- show the results of the survey clearly
- display the data using diagrams that make it easy to compare what the members of each of the two groups liked
- display the data using diagrams that show what proportion of each group liked each name
- include a short paragraph summarising what you found out

• *Show your account to your teacher.*

Section 13: Sampling

In this section you will:
- meet several methods of sampling
- understand the effects different sample methods and sample sizes have on the reliability of conclusions
- justify sampling methods

DEVELOPMENT

D13.1: Simple samples

Y10 students at Whynot school were instructed to design, and carry out, a statistical project on homework.

Whynot School	Y7	Y8	Y9	Y10	Y11	Y12	Y13	
Boys	90	82	63	64	52	23	18	
Girls	71	65	62	60	57	32	20	Total 759

Jes and Alex decide to ask everyone in the school this question
" Is the amount of homework that you are given
too little ☐ about right ☐ too much ☐ ?"

1. Is the question biased ?
2. After a little thought, Jes decides it is not sensible to ask everyone in the school. Explain why it is not sensible ?

Jes decides to ask everyone in the top two Y10 maths sets.
This is called **a simple sample.**

3. Alex says that this sample would not be representative of the whole school. Do you agree ? Explain why or why not.

Alex decides to get to school early and ask the first 50 students coming into school.
This is also **a simple sample.**

4. How representative do you think this example would be of the whole school ? Would you describe it as poor, reasonable or good ?
5. Is 50 a big enough sample size ?
6. Would asking more students give a more reliable result ?
7. Would 10 be a big enough sample size ?
8. If they asked the first ten people in each Y10 tutor group, would that be a good representative sample for the school ?
9. If they asked the first ten people in one tutor group in each year, would that be a good representative sample for the school ?

• Check your answers.

D13.2: Random sampling

Lucy and Zoe have written down these three questions:
A: How much homework do you do ?
B: How much homework did you do last week ?
C: How many minutes homework did you do last night ?

1. Why is A not a good question ?
2. Why is C a better question than B ?

A random sample is one where every member of the target population is equally likely to be chosen for the sample.

Lucy and Zoe want a random sample of 50.
Lucy's first thought is to give everyone in the school a number, put each number onto a piece of paper and put all the pieces of paper in a box.
Then she would mix them all up and draw one out, 50 times.

3. Would this give a random sample ?
4. Lucy decided that this was not a good idea. Why is it not ?

Zoe used the school registers to give everyone a number.
To make things easier, she just wrote down
7AB: 1 – 28, 7MC: 29 – 55, 7BJ: 56 – … … … 13 PJ: 736 – 759
She used the random number generator on her calculator to generate the fifty members of her random sample. (RAN# on most calculators)

5. The first number Zoe's calculator gave her was 0.036

   Zoe just wrote down 036 036 would be the 7th person in 7MC

   The next two numbers her calculator gave were 0.058 and 0.737 Where will she find the two people corresponding to these numbers?

6. Zoe wrote down: 036 058 737 671 490 ~~943~~ 033 ~~761~~ ~~938~~ (671) …
   (a) Why did she cross out 943 761 and 938 ?
   (b) She didn't count the number with a circle round it either. Why not ?

Lucy thought of another way of getting a random sample. She would choose the first number at random from the first register (either with a pin or just asking someone to choose a number between 1 and 28). Then she would take every 15th number after that – and, if possible, before that.

This is a systematic random sample

7. Why did she choose every fifteenth number to get a sample of around 50 ?
8. This is a random sample because the random choice of the first number means that all numbers in the target population are equally likely to be selected. Why could she not toss a dice to give the first number ?

• *Check your answers.*

D13.3: Stratified sampling

Hamid and Sara decide to test the hypothesis:
"Girls spend more time on homework than boys."

They decide that they will take a 10% stratified sample of the target population.

Hamid suggests that, as there are 392 boys and 367 girls, they take a random sample of 39 boys and 37 girls.

Sara suggests that they take 10% of the boys and the girls in each year, chosen at random.

Both of these are **stratified samples**

1. Both methods will give a good representative sample to test the hypothesis. But Sara's method will give more information. What additional information will it give ?

Here are three hypotheses that other groups of students want to test.
A: Y7 pupils spend more time on homework than Y8 pupils.
B: Y11 students do more work at home than Y12
C: The further you move up the school the more homework you do.

2. Suggest a method of stratified sampling that could be used to test all three hypotheses.

• *Check your answers.*

D13.4: Quota sampling

This is a method of stratified sampling in which the selection of members in each stratum is non-random. The population is divided into groups in terms of known characteristics (age, sex, social class …). The sample to be interviewed has the same groups in the same proportion. But, from then on, the choice of the sample members is left to the interviewer. This type of sampling is used by market research organisations and some national opinion polls (particularly in respect to voting intentions), mainly because it is cheap to administer. Any sample member can be replaced by another with the same basic characteristics. This is not so in random sampling, where the interviewer must go back to sample members who were not available when first sought.

1. In your school there are 500 boys, 450 girls and 50 staff. You want to test how many people in the school watched the athletics on TV last night. You ask the first member of staff that you meet and the first nine girls that you meet. How many boys do you ask ?
2. 1 member of staff, 5 girls and 7 boys in your quota sample, watched the athletics. Estimate how many people in the school watched the athletics.
3.

Meriden	pensioners	children	working	unemployed
male	168	278	770	86
female	222	232	670	73

You are designing an opinion poll on the uses of the proposed community centre. You are using a quota sample based on the population of Meriden. Your sample includes four unemployed men. How many of each of the other groups should be in the sample ?

• *Check your answers.*

D13.5: Other forms of sampling

Repeated sampling : quality control
A light bulb manufacturer tests a sample of 25 bulbs every four hours.

1. He needs to be sure that at least 95% of the bulbs work perfectly. How many of the sample need to fail before he checks on the manufacturing standards or raw materials ?

Cluster sampling : samples of samples
Every tenth authority is chosen from an alphabetical list of education authorities.
Every tenth secondary school is chosen from each authority's alphabetical list.
The GCSE results of each of these schools is published.

2. Give two reasons why this sample may be biased.

• *Check your answers.*

D13.6: Choosing sampling methods

1. One quarter of the school computer club members are girls. Ellen thinks that it would be better if more girls joined. She does a survey of all the members of the club to see if they think so too.
   (a) Why is this not a good group of people to do the survey with ?
   (b) Describe a better sample for her to survey.
2. Paul wanted to estimate what percentage of the school owned a computer. Design a sampling method that would give him a good estimate.

A healthfood chain is thinking of opening a vegetarian restaurant in either Millam or neighbouring Norbury. Sally lives in Millam and would like to see a vegetarian restaurant open there. She decides to do a survey and use it to persuade them to choose Millam.

3. Sally's school goes from Y8 to Y11. She asks everyone in Y10 whether they were vegetarians and multiplies the answer by 4, to estimate how many vegetarians are in her school. Would this be a representative sample ?
   How could she have made it a better sample ?
4. There are around 800 students in Sally's school. There are 40,000 people in Millam. She wants to estimate the number of vegetarians in Millam. She multiplies the number of vegetarians in the school by 50.
   Why will this not give a reasonable estimate?
5. Describe a method of sampling that will give a reasonable estimate of the number of vegetarians in the town.

• *Check your answers.*

EXTENSION

E13.7: Using random numbers to make a sample estimate

Random numbers are produced by computer. There is no significance whatsoever in how the digits are arranged. Some tables may place them in pairs, some in threes, some in blocks of 5 … You choose your own starting point at random on the table and move left, or right, or up, or down, or diagonally … to the next digit. It is usual to continue moving in the same way as you started.

Here you are going to make a random sample of the points on the map in order to estimate what proportion of the area is woodland, orchard,…. Every point on the map must have an equal chance of being selected. A point is fixed by a 6-figure grid reference.

255 632 gives the position of the church

Accuracy increases with the size of the sample, but at the cost of increased effort. Thirty points is regarded as the minimum.

Use map on next page.

Task 1: Method 1: Starting anywhere on the random number table, take the first 6 digits. Record on a table like this whether the point is woodland, orchard, rough grazing or other. If the point is not on the table, discard it. Repeat until you have recorded three points on the map. Stop – explain why this method is going to take a long time.

Rough gr.		Woodland		Orchard		Other	
Tally	%	Tally	%	Tally	%	Tally	%

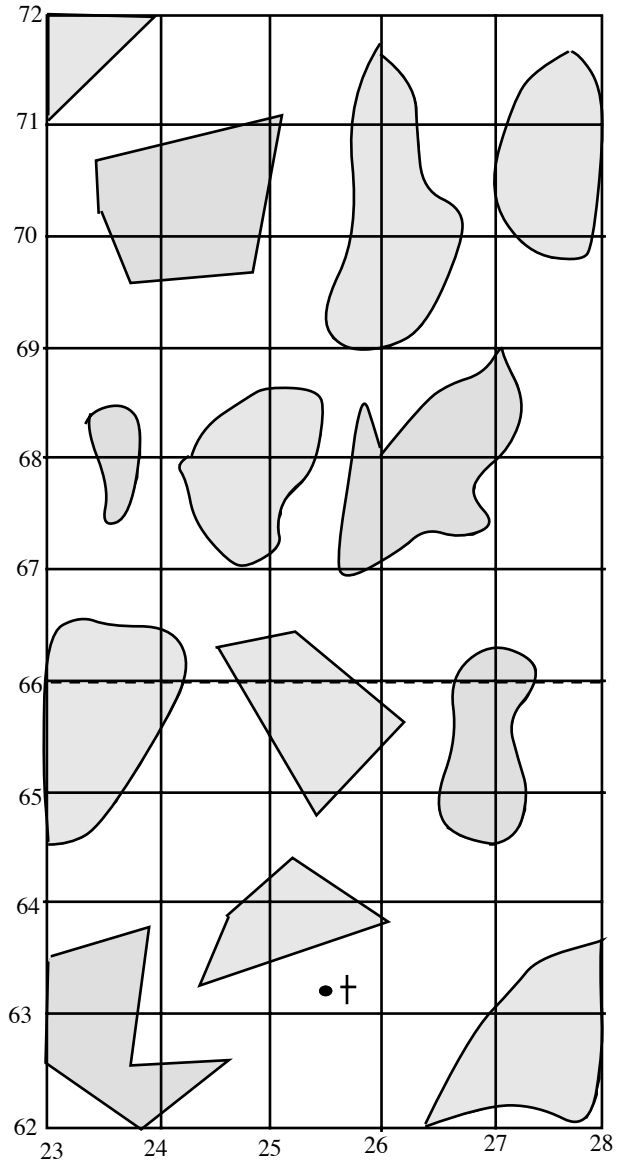
Task 2: Method 2: Since all the eastings (first three digits) start with 2, take the first two random digits as the second and third digits of the grid reference. Use the next two random digits as the fifth and sixth digits of the reference. Record the results as in Task 1. When you have enough points, work out the percentage of the sample points in each category.

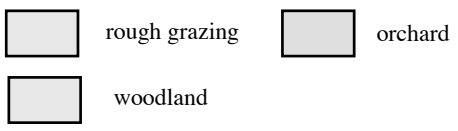
Task 3: Method 2 but modified to produce stratified sample

You now have an additional piece of information: the southern most 40% of the map (below the dotted line) is sandstone. The rest is clay. Split your table into two parts so you can record whether each sample point is sandstone or clay, as well as the original categories. Make a random sample of 50 points, using method 2. Sandstone is 40% of the map so must have 40% of the points (that is 20). If more than 20 points are generated in the sandstone part of the map, they are ignored.

Work out the percentage of the sample points in each category.

• *Your teacher has the answers to this.*





Random number table

470	974	976	476	125
505	761	885	675	199
938	851	293	574	392
146	212	772	131	970
879	643	925	536	227
865	868	330	052	224
820	739	715	597	383
847	008	444	448	616
084	136	342	385	861
888	532	085	812	791
656	920	744	348	671
270	780	929	208	763
935	588	726	029	192
418	473	896	165	058
334	701	544	399	054
489	731	992	120	979
867	968	194	870	299
506	549	083	556	744
907	254	548	881	682
264	604	165	977	019
109	679	141	160	746
324	045	030	174	746
493	740	028	084	308
031	547	679	094	598
923	254	137	653	505
429	526	116	840	763
917	661	796	096	048
460	785	819	896	317
600	515	071	307	605
282	462	473	160	025
385	320	897	138	629
490	328	465	759	214
148	037	636	649	578

Section 14: Histograms

In this section you will:
- meet and work with histograms
- calculate and use frequency densities

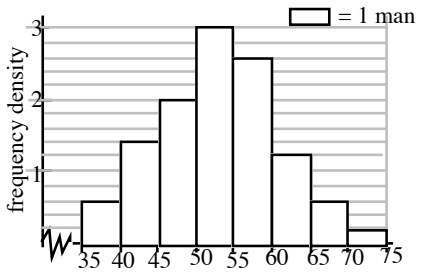
DEVELOPMENT

D14.1: Introducing histograms

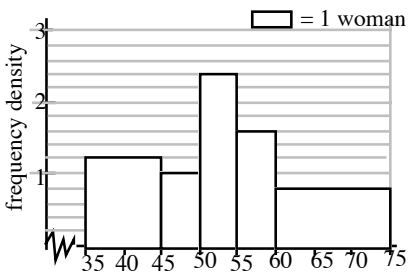
A histogram is used to illustrate continuous data or discrete data which has been grouped, and is therefore treated as if it were continuous data.

The area of each block represents the frequency.

Ages of competitors in Senior Men's Golf Competition



Ages of competitors in Senior Women's Golf Competition

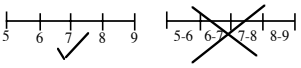


1. *Complete these frequency tables for men and women using the information given in the histograms*:

Men's ages	Frequency density	Frequency
35 —>40	0.6	3
40 —>45	1.4	7
……		

Women's ages	Frequency density	Frequency
35 —>45		
……		

Points to note:



- the scale across the page is continuous
- the bocks take the widths of the true classes
- the boundaries of the blocks are the LCBs and UCBs of the classes
- the classes may, or may not, have equal class widths
- the area of each block represents the frequency
- frequency density x class width = frequency
- the scale up the page is essential for calculating the height of each block, when drawing a histogram

BUT

for reading values from the histogram, the key [eg ☐ = 1 man], which is optional, is far more useful.

2. (a) *Copy and complete this table:*

Annual income of part-time workers at the Skycom Factory					
Pay (£,1000)	less than 2	2 —>4	4 —>5	5—>10	10 —>20
class width		2			
frequency	20	33	12	15	6
freq. density		16.5			

(b) *Draw the histogram for this distribution.*

Points to remember:
- check if the classes given are the true classes
- if they are not the true classes, add a column for true classes to your table
- boundaries of blocks are boundaries of true classes.

3. (a) *Copy and complete this table:*

Lengths of slugs measured to the nearest 0.1 cm				
Length	frequency	true classes	class width	freq. density
3.0 – 3.4	8	$2.95 \le l < 3.45$	0.5	16
3.5 – 3.8	6			
3.9 – 4.2	7			
4.3 – 4.5	9			
4.6 – 5.0	20			
5.1 – 5.5	3			

(b) *Draw the histogram for this distribution.*

4.

Body temperatures of 50 15-year olds, measured to nearest 0.1°C						
Temp.	36.0 – 36.3	36.4 – 36.5	36.6 – 36.7	36.8 – 36.9	37.0 – 37.1	37.2 – 38.0
Freq.	2	4	8	20	14	2

(a) *Make a frequency density table for this data.*
(b) *Draw the histogram for this distribution.*

Grouped discrete data
Histograms always stand on a continuous scale.
So, grouped discrete data is treated exactly as if it were continuous.

5.

Books returned by Y11 students after GCSE exams					
Number	1 – 5	6 – 8	9 – 10	11 – 15	16 – 20
Freq.	8	32	40	15	10

(a) *Make a frequency density table for this data.*
(b) *Draw the histogram for this distribution.*

• *Check your answers.*

D14.2: Extracting information from histograms

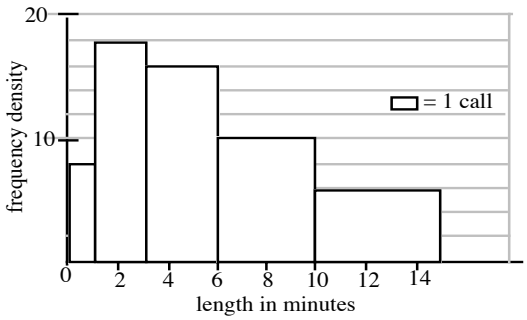
1. The histogram shows the lengths of telephone calls made from an office in one day.

(a) What was the modal class of telephone call ?

(b) How many telephone calls were made that day ?

(c) The mean length of call was roughly 6 minutes.

Estimate the percentage of calls which were longer than the mean.



2. In a survey, the employees of a company were asked how far they lived from their workplace. Each employee ticked one of the following responses:

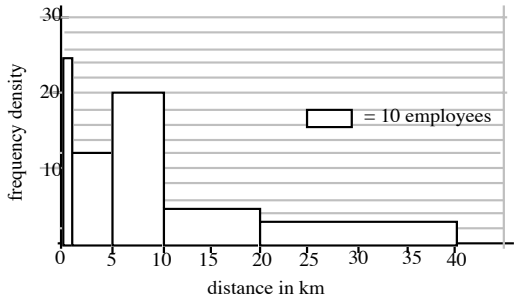
less than 1 km	☐
at least 1 km but less than 5 km	☐
at least 5 km but less than 10 km	☐
at least 10 km but less than 20 km	☐
at least 20 km but less than 40 km	☐
40 km or more	☐

100 employees ticked the third box.

Nobody ticked the last box.

The histogram shows the results of the survey.

Work out how many employees filled in the response form.

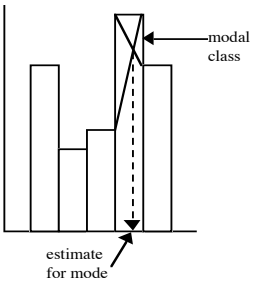


• *Check your answers.*

EXTENSION

E14.3: Estimating the modal value

Here is a way to estimate the value of the mode for a grouped distribution.



Estimate the mode for each of the distributions that you drew histograms for in questions 2, 3, 4 of D14.1.

• *Check your answers.*

Topic 7: Analysing Statistics

Section 1: Frequency diagrams p 206

D1.1 : Getting the words right

1. QN 2. QL 3. QL 4. QN 5. QL
6. QN 7. D 8. C 9. D 10. D
11. C 12. C 13. C 14. D 15. D
16. C 17. C 18. C

D1.2 : Frequency diagrams

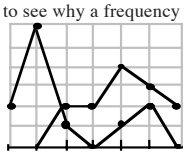
1. 0 2. 15 3. 4 4. 18 5. 10
6. Denmark 7. Italy & France
8. Diagram A: very few cars were travelling over the speed limit of 70 mph

D1.3:Comparing frequencies with ungrouped data

A is T B is F C is T D is F E is T F is T

D1.5: Comparing frequencies with grouped data

1. Frequencies are 2, 6, 1, 0, 1, 2
2. Frequencies are 2, 2, 4, 3, 1
3. The student is only required to see why a frequency polygon is preferable to this
4. Middle value 5.
6. 3 – there are 3 unusually large females
7. M 19 g F 16 g
8. F 9. F



Section 2: Data calculations p 210

D2.1 : Measures of data

1. 5.4; 3, 5; 7
2. 24.8; 24; 24; 2
3. 13.625; 13, 13.5; 2
4. 6; 2; 6; 9
5. 2; 2; 2; 4
6. 2.08; 2.5; 2.05; 1
7. 516.2; 506; 507; 34
8. 0.48; 0.55; 0.5; 0.17
9. 0; –1, –1; 7
10. 11; 13, 11; 19

D2.2 : Using calculator functions to find the mean

2. (a) 51.9 (b) 10 (c) 519
3. (a) 139.42 (b) 12 (c) 1673
4. (a) 6.61 (b) 9 (c) 59.5
5. (a) 53.46 (b) 695

D2.3 : The maths test

1. 5.9, 7, 7, 10 2. 130 3. 307 4. 6.14
5. 8MB 6. 8MB 7. 6.23

Section 3: Tables of ungrouped data p 212

D3.1: Calculations using tables of ungrouped data

1. NxF column is 6 14 24 40 25 18 0 8
   Totals are 40 & 135
2. 8 3. 3 4. 40 5. 135 6. 7
7. 4 8. 135 ÷ 40 = 3.375 9. 3
10. 1.64; 1; 1 11. 67.75; 67; 67 & 68

D3.2: Using calculator functions to find the mean

1. 102.29 2. 51.45 3. 25.75
4. 1.64 5. 67.75

D3.3: How many pets

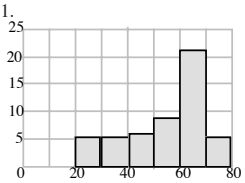
1. no. of children 4 5 6 3 2
2. 20 3. 2 4. 6.8 5. both the 10th and 11th child had two pets, so the median is 2 6. $n = 3$

Section 4: Tables of grouped data p 214

D4.1: Calculations using tables of grouped data

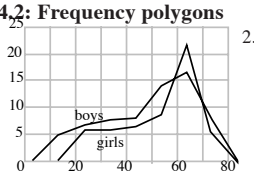
1. M column is 15.5, 25.5, 35.5, 45.5, 55.5, 65.5, 75.5
FxM column is 62 153 248.5 318.5 721.5 1048 528.5
Totals are 60 & 3080
2. $3080 \div 60 \approx 51.33$
3. F column is 0, 5, 5, 6, 8, 21, 5
FxM column is 0 127.5 177.5 273 444 1375.5 377.5
Totals are 50 & 2775
4. 61-70 5. 50 6. $2775 \div 50 = 55.5$
7.

51.33	55.5
61-70	61-70
61-70	61-70
69	59

8. All three average estimates are the same or higher for the girls.
9. The range for the girls' marks is smaller than the boys'
10. mean : it is more precise. Also, there are no extremely small or large values to distort the mean.
11. 

12. There are 60 boys and 50 girls, so the means are not equally weighted.

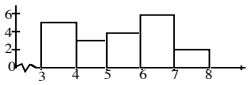
D4.2: Frequency polygons

1. 

2. A is F
B, C, D are T
E is I

Section 5: Tables of continuous data p 216

D5.1: Calculations using continuous data

1. Midpoints are 3.5, 4.5, 5.5, 6.5, 7.5
Freq are 4, 2, 5, 7, 2
MxF are 14, 9, 27.5, 45.5, 15
Totals are 20 and 111 2. 5.55
3. 

D5.2: Choosing and labelling groups

1. (a) first (b) second (c) both
2. the groups overlap
3. (a) first (b) second (c) neither
4. gap between groups
5. (a) first (b) second (c) neither
6. gap between groups 7. A & E 8. C and F

P5.3: Cormorant watching

1. Freq are 3 6 9 8 9 3 2
Midpts are 5 15 25 35 45 55 65
MxF are 15, 90, 225, 280, 405, 165, 130
Totals are 40 and 1310
2. $20 \le t < 30$ and $40 \le t < 50$
3. $1310 \div 40 = 32.75$ 4. $30 \le t < 40$

Section 6: Questionnaires p 220

D6.1: A question of bias

1.

Q	L	E	L+E	N
A	√			
B				√
C				√
D			√	
E	√			
F				√

2. P: Should we have a school uniform ?
Yes ? No ? Don't know/care ?
Q: Should hanging be a possible punishment for murder? Yes ? No ? Don't know/care ?
R: Should the using of cannabis be legalised ?
Yes ? No ? Don't know/care ?
3. W: They will be biased towards watching sport on TV
X: They will probably be biased against sport on TV – gone shopping to avoid it.
Y: They will be biased towards saying "too little".
Z: They will be unlikely to know the facts.
4. *Your teacher will need to mark Q 4 & 5*

Section 7: Testing hypotheses p 222

D7.1: Methods for testing hypotheses

1. (a) Do you think stag hunting is
cruel ☐ necessary to keep down the numbers ☐
don't know ☐
or Should stag hunting be legal ☐ illegal ☐
don't know ☐ (b) B
(c) sample an equal number from each year
2. (a) How many hours TV did you watch last week ? (or last night? or …)

Time	Tally	Frequency
$0 \le$ time < 2 h		
$2 \le$ time < 4 h		
…		

(b) B or C
3. (a) False: For each region, state population per km^2, and note that Scotland is the least populated, (66 people /km^2) & the NW is the most densely populated (873/km^2)
4. Expt 2 is biased, because the act of looking at the counters influences the choice of counters.
Expt 3 is biased, because if the counters are not shaken, then the counters on the top are more likely to be chosen.
5. After a few spins, the hole in the spinner gets larger, the point on the match gets worn and the card bends, so the spinner leans to one side.

Section 8: Scatter graphs p 225

D8.1: Is there a relationship ?

A is length of foot B is time taken to get to school
C is time taken to run 100 m All true

D8.2:Scatter graphs and correlation

Case 1 : negative correlation - the eggs of birds who have the smallest clutches need a longer incubation time (the eggs are probably larger)

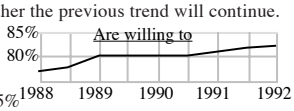
Case 2: good positive correlation - the better teams have both a high point score and a good goal difference
Case 3: positive correlation - animals with longer gestation periods live longer

D8.3:Using a line of best fit

2. 46-47% 3. 43-44%
4. Yes, he could have expected to get around 56%
5. A line of best fit should give a life span for humans around 30-45 years - which is not true !

Section 10: Time series graphs p 230

D10.1: Trends in use of unleaded petrol

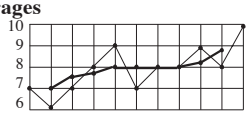
1. 17% 2. (a) around 55% (b) around 60%
3. Extrapolation is more risky because no-one can tell whether the previous trend will continue.
4. 

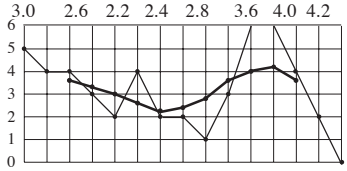
5. 84-85%
6. Legislation to stop production of cars that use leaded petrol (which is what happened)

D10.2: Analysis of time series

1. increasing 2. one year
3. It repeats at the end of the year because of the Christmas shopping boom.
4. A recession
5. The peaks would be in the same position but would be smaller.
6. Alcohol : (a) one year
   (b) around Christmas and New Year - celebrations
   (c) just after New Year - people are getting over the earlier celebrations and have less money
   Fuel and power : (a) one year
   (b) winter months - when it is colder
   (c) summer - when it is warmer
   Air travel (a) one year
   (b) summer months - more people go on holiday
   (c) winter - most people stay at home

D10.3: Moving averages

1. (a) (b) 

   7 7.5 7.75 8
   8 8 8.25 8.75
2. (a) 3.6 3.4
   3.0 2.6 2.2 2.4 2.8 3.6 4.0 4.2 3.6


3. (a) 51 58 67 71 69 67 68 70
   (b) standards are being maintained
4. (a) 1.59 1.61 1.60 1.43 1.38 1.54
   1.61 1.61 1.66 1.69 1.71 1.79 1.83
   (b) A steady increase

E10.4: The greenhouse effect

1. agriculture and industry, domestic, commercial, refineries
2.50; 10, 30, 21, 20

E10.5: *Purchasing power … All true

Section 11: Cumulative frequency p234

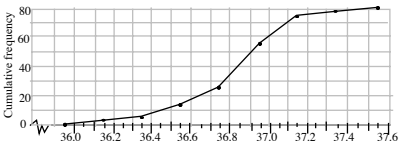
D11.1: Running totals

1. RTs are 3 15 37 73 111 131 138
2. 138
3. 15 is the sum of the first two groups, all of whom got less than 30.
4. (a) 73 (b) 131 5. 37 6. 134
7. 47 8. 21 9. 16 10. 87
11. 45 12. 40&50 13. 40&50

D11.2: Cumulative frequency diagrams

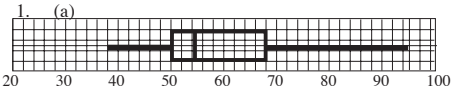
1. Start at the 55 marks on the axis. Move up this line to the cumulative frequency line. Go across to the other axis – 105 people. [Can be done in reverse]
2. 86 3. 48 4. 62 5. 45 6. 32 7. 52
8. 20
9. If answers to 11 & 12 are correct, then your diagram will be OK. If not, show it to your teacher.
10. 138 11. $M \approx 49$ $Q_1 \approx 39$ $Q_3 \approx 58$
12. IQR ≈ 19
13. (a) All point should be plotted at the top of each interval – not the middle. If your answers to (b) & (c) are accurate, then your graph is likely to be correct.
    (b) median ≈ 161 $Q_1 \approx 158$ $Q_3 \approx 171$
    (c) IQR ≈ 13 (or your $Q_3 - Q_1$)

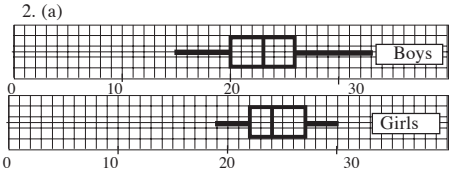
D11.3: But there are gaps between these groups

1. Cu Freq are 3, 5, 13, 27, 56, 74, 79, 80
top of true intervals are 36.15, 36.35, 36.55, 36.75, 36.95, 37.15, 37.35, 37.55
2. 

3. M ≈ 36.85 LQ ≈ 36.65 UQ ≈ 37.00 IQR ≈ 0.35

D11.4: Box and whisker plots (aka box plots)

1. (a) 

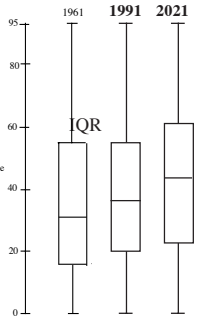
   (b) A: True B: False C: False
   D: False E: True F: True
2. (a) 

   (b) The fastest and the slowest were both boys. On average, the girls were fastest. The boys results were more symmetrical about the median.

E11.5: Trends in Britain's population

1. If $Q_1 \approx 22$, $Q_2 \approx 43$, and $Q_3 \approx 61$, then your diagram is probably correct.
2. data required to draw box plots

		Q_1	Q_2	Q_3
1961	16	35.5	55	39
1991	19.5	36.5	57.5	38
2021	22	43	61	39



3. All true

Section 13: Sampling p 240

D13.1: Simple samples

1. No 2. Too many people *or* would take too long
3. Agree. Does not represent all abilities or all ages.
4. Reasonable 5. Y 6. Y 7. N 8. Y 9. N

D13.2:Random sampling

1.Too vague: Q needs to be more specific
2. Precise and easy to work out answer– so, more likely to get correct, rather than guessed answer
3. Y 4. Take too long
5. 3rd person in 7BJ; 2nd person in 13PJ
6.(a) out of range (b) already got it
7. 15 x 50 = 750 & 759 in school
8. All numbers must be equally likely to be chosen – 7-15 cannot be chosen if you toss a dice

D13.3:Stratified sampling

1. Will also tell you if hypothesis is true, for each year group
2. Random samples of eah year [Plus, if you record boys and girls separately, you can also test H and S's hypothesis]

D13.4: Quota sampling

1. 10 boys 2. 650
3.

	P	Ch	Work	Unemp
M	8	14	39	4
F	11	12	34	18

D13.5:Other forms of sampling

1. 2
2. May not be representative of different parts of the country. May not be representative of each kind of school.

D13.6: Choosing sampling methods

1.(a) too few girls in this group(b) tutor group *or* year group *or* random sample
2. Stratified random sampling – random samples from each year
3. Possibly, if we assume vegetarians are in the same proportion in each year. Better to do random sampling from each year.
4. School population not representative of town as a whole 5. Quota sampling

Section 14: Histograms p 246

D14.1:Introducing histograms

1. M

45 —>50	2.0	10	
50 —>55	3.0	15	
55 —>60	2.6	13	
60 —>65	1.2	6	
65 —>70	0.6	3	
70 —>75	0.2	1	

Women

35 —>45	1.2	12	
45 —>50	1.0	5	
50 —>55	2.4	12	
55 —>60	1.6	8	
60 —>75	0.8	12	

2.

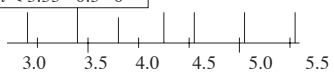
(a)						
cw	2	2	1	5	10	
fr. d.	10	16.5	12	3	0.6	

(b) Histogram must have
• continuous scale 0…20
• block heights = fr. d. in (a)

3.

$3.45 \le l < 3.85$	0.4	15	
$3.85 \le l < 4.25$	0.4	17.5	
$4.25 \le l < 4.55$	0.3	30	block heights = fr. d. in (a)
$4.55 \le l < 5.05$	0.5	40	
$5.05 \le l < 5.55$	0.5	6	

(b)



4. tc 35.95–36.35 36.35–36.55 36.55–36.75
36.75–36.95 36.95–37.15 37.15–38.05

cw	0.4	0.2	0.2	0.2	0.2	0.9
fr.d.	5	20	40	100	2	

(b) Histogram must have • continuous scale
• block heights = fr. d. in (a)

5.

tc	0.5–5.5	5.5–8.5	8.5–10.5	10.5–15.5	15.5–20.5
cw	5	3	2	5	5
fr.d.	3.6	11	20	3	2

(b) Histogram must have • continuous scale
• block heights = fr. d. in (a)

D14.2:Extracting information from histograms

1. (a) 3 – 6 min (b) 163 (c) 43% 2. 283

E14.3:Estimating the modal value

2. ≈ £2,600 3. ≈ 4.67 4. ≈ 36.885