edexcel

GCSE (9-1) Statistics

Sample Assessment Materials

Pearson Edexcel Level 1/Level 2 GCSE (9-1) in Statistics (1ST0)

First teaching from September 2017

First certification from June 2019 | Issue 3

Pearson

LCCI qualifications

Edexcel, BTEC and LCCI qualifications are awarded by Pearson, the UK's largest awarding body offering academic and vocational qualifications that are globally recognised and benchmarked. For further information, please visit our qualification website at qualifications.pearson.com. Alternatively, you can get in touch with us using the details on our contact us page at qualifications.pearson.com/contactus

About Pearson

Pearson is the world's leading learning company, with 35,000 employees in more than 70 countries working to help people of all ages to make measurable progress in their lives through learning. We put the learner at the centre of everything we do, because wherever learning flourishes, so do people. Find out more about how we can help you and your learners at qualifications.pearson.com

Original origami artwork: Mark Bolitho
Origami photography: Pearson Education Ltd/Naki Kouyioumtzis

ISBN 978 1 4469 5267 2

Summary of Pearson Edexcel GCSE in Statistics Sample Assessment Materials

Issue 2 changes

Summary of changes made between issue 1 and issue 2	Page number
Paper 1F The mark schemes have been amended for questions 9, 10 and 14, to improve clarity for centres, as to where marks are awarded.	35, 36, 39
Paper 2F The mark schemes have been amended for question 14, to improve clarity for centres, as to where marks are awarded.	74, 75
Paper 1H The mark schemes have been amended for questions 1, 5 and 7, to improve clarity for centres, as to where marks are awarded.	101, 103, 104
Paper 2H The mark schemes have been amended for questions 5, 9(a), (b), 11(b) and 15, to improve clarity for centres, as to where marks are awarded.	136, 137, 140, 141, 144

Issue 3 changes

Summary of changes made between issue 2 and issue 3	Page number
The general marking guidance has been updated to show the same information that is used in the live assessments.	3, 4
Paper 1F The table for question 11, row 8, has been amended to show '48' in the 'gender' column and 'boy' in the 'length of text message' column. This is a deliberate error in the table which candidates are expected to notice.	18
Paper 1H The table for question 2, row 8, has been amended to show '48' in the 'gender' column and 'boy' in the 'length of text message' column. This is a deliberate error in the table which candidates are expected to notice.	80
Mark scheme for paper 1H For question 8(a) in the answer column the final answer has been corrected to A1 for 98.7 and 101.7	105
Paper 2H For question 10 the comparative pie chart for Snowdonia has been corrected to a reduced size.	124
Mark scheme for paper 2H For question 13(c) in the answer column the spacing has been amended to show the answers clearly.	143

If you need further information on these changes or what they mean, contact us via our website at: qualifications.pearson.com/en/support/contact-us.html.

Contents

Introduction	1
General marking guidance	3
Paper 1F	5
Paper 1F: Mark scheme	29
Paper 2F	41
Paper 2F: Mark scheme	65
Paper 1H	77
Paper 1H: Mark scheme	101
Paper 2H	111
Paper 2H: Mark scheme	135

Introduction

The Pearson Edexcel Level 1/Level 2 GCSE (9–1) in Statistics is designed for use in schools and colleges. It is part of a suite of GCSE qualifications offered by Pearson.

These sample assessment materials have been developed to support this qualification and will be used as the benchmark to develop the assessment students will take.

General marking guidance

These notes offer general guidance, but the specific notes for examiners appertaining to individual questions take precedence.

1 All candidates must receive the same treatment. Examiners must mark the last candidate in exactly the same way as they mark the first.

Where some judgement is required, mark schemes will provide the principles by which marks will be awarded; exemplification/indicative content will not be exhaustive. When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the response should be sent to review.

2 All the marks on the mark scheme are designed to be awarded; mark schemes should be applied positively. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme. If there is a wrong answer (or no answer) indicated on the answer line always check the working in the body of the script (and on any diagrams), and award any marks appropriate from the mark scheme.

Questions where working is not required: In general, the correct answer should be given full marks.

Questions that specifically require working: In general, candidates who do not show working on this type of question will get no marks – full details will be given in the mark scheme for each individual question.

3 Crossed out work

This should be marked **unless** the candidate has replaced it with an alternative response.

4 Choice of method

If there is a choice of methods shown, mark the method that leads to the answer given on the answer line.

If no answer appears on the answer line, then mark both methods **as far as they are identical** and award these marks.

5 Incorrect method

If it is clear from the working that the "correct" answer has been obtained from incorrect working, award 0 marks.

6 Follow through marks

Follow through marks which involve a single stage calculation can be awarded without working as you can check the answer, but if ambiguous do not award.

Follow through marks which involve more than one stage of calculation can only be awarded on sight of the relevant working, even if it appears obvious that there is only one way you could get the answer given.

7 Ignoring subsequent work

It is appropriate to ignore subsequent work when the additional work does not change the answer in a way that is inappropriate for the question or its context. (e.g. an incorrectly cancelled fraction when the unsimplified fraction would gain full marks).

It is not appropriate to ignore subsequent work when the additional work essentially makes the answer incorrect (e.g. incorrect algebraic simplification).

8 Probability

Probability answers must be given as a fraction, percentage or decimal. If a candidate gives a decimal equivalent to a probability, this should be written to at least 2 decimal places (unless tenths).

Incorrect notation should lose the accuracy marks, but be awarded any implied method marks.

If a probability fraction is given then cancelled incorrectly, ignore the incorrectly cancelled answer.

9 Range of answers

Unless otherwise stated, when an answer is given as a range (e.g. 3.5 – 4.2) then this is inclusive of the end points (e.g. 3.5, 4.2) and all numbers within the range.

Guidance on the use of abbreviations within this mark scheme

M method mark awarded for a correct method or partial method

A accuracy mark (awarded after a correct method; if no method or process is seen then full marks for the question are implied but see individual mark schemes for more details)

B unconditional accuracy mark (no method needed)

oe or equivalent

cao correct answer only

ft follow through (when appropriate as per mark scheme)

sc special case

dep dependent (on a previous mark)

indep independent

awrt answer which rounds to

isw ignore subsequent working

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

Centre Number | | | | | |

Candidate Number | | | | |

Sample assessment material for first teaching September 2017

Morning (Time: 1 hour 30 minutes) | Paper Reference **1ST0/1F**

Statistics

Paper 1
Foundation Tier

You must have:
Ruler graduated in centimetres and millimetres, protractor, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.

Information

- The total mark for this paper is 80.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

S54190A

1/1/1/1/1/

S 5 4 1 9 0 A 0 1 2 4

Pearson

Answer ALL the questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The incomplete pictogram shows the numbers of hours that Milly spent studying Biology, English and French.

Biology	▭ ▭
English	▭ ▭ ▭ ▭ ▭
French	▭
Maths	

Key: ▭ represents 2 hours

Milly spent 6 hours studying Maths.

(a) Complete the pictogram for Maths.

(1)

Milly spent more hours studying English than she spent studying Biology.

(b) Work out how many more hours.

..... hours

(2)

Milly's teacher wants her to spend a total of more than 20 hours studying.

(c) Has Milly spent a total of more than 20 hours studying?
Give a reason for your answer.

.....

.....

.....

(2)

(Total for Question 1 is 5 marks)

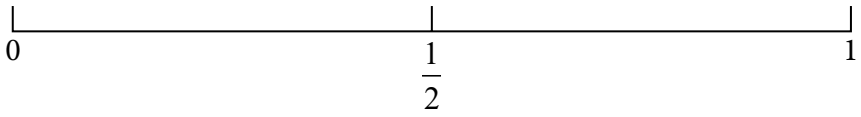
DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

2 Ramon uses 10 letter cards to spell the word **STATISTICS**.

S	**T**	**A**	**T**	**I**	**S**	**T**	**I**	**C**	**S**

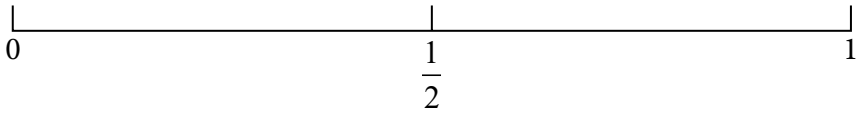
Ramon takes one of these cards at random.

(a) On the probability scale, mark with a cross (×) the probability that the letter on the card is **S**.



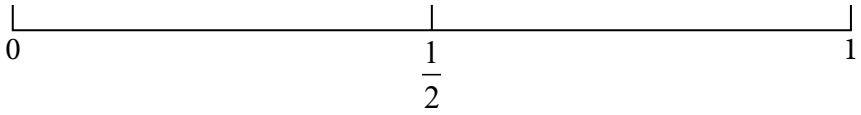
(1)

(b) On the probability scale, mark with a cross (×) the probability that the letter on the card is **Z**.



(1)

(c) On the probability scale, mark with a cross (×) the probability that the letter on the card is **not C**.



(1)

(Total for Question 2 is 3 marks)

3 Leyla wants to find out how often people in her town eat in a restaurant.

She asked a sample of 30 people how many times they had eaten in a restaurant during the last week.

Here are Leyla's results.

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

(a) Fill in the tally chart for this information **and** complete the frequency column.

Number of times	Tally	Frequency
1		
2		
3		
4		
5		

(2)

(b) Write down the mode.

.....

(1)

(c) Work out the number of people in Leyla's sample who had eaten in a restaurant fewer than 4 times during the last week.

.....

(2)

(d) Suggest a suitable diagram that Leyla could use to represent her data.

.....

(1)

(Total for Question 3 is 6 marks)

4 A researcher is investigating how much the employees at a large company are paid.

One hypothesis she investigates is

"Men are paid more than women".

The researcher could find it difficult to collect information to test her hypothesis.

(a) Give one difficulty the researcher could have when trying to find out how much each employee is paid.

.....

.....

(1)

(b) State the population for this investigation.

.....

(1)

(c) (i) Explain the difference between primary data and secondary data.

.....

.....

.....

(2)

(ii) The researcher plans to collect primary data. Give a reason why.

.....

(1)

The researcher plans to give a questionnaire to 60 employees of the company.

She asks the first 30 males and the first 30 females who come into work one morning to complete her questionnaire.

(d) Give one advantage and one disadvantage of this sampling method.

Advantage .....

.....

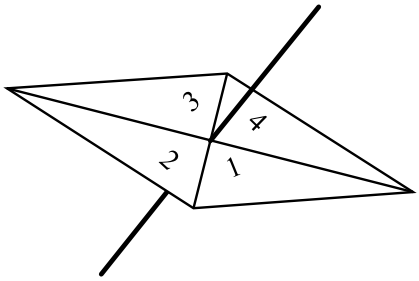
Disadvantage .....

.....

(2)

(Total for Question 4 is 7 marks)

5 Zander spins a fair 4-sided spinner.



(a) Write down the probability that the spinner lands on 1

.....

(1)

Zander now spins the spinner 60 times.

(b) Work out the expected number of times the spinner will land on 1

.....

(2)

Helen spins a biased 3-sided spinner 45 times.

On each spin, the spinner can land on 1 or on 2 or on 3

Here are her results.

Number	Frequency
1	7
2	9
3	29

Helen is going to spin the 3-sided spinner again.

(c) Use the results in the table to find an estimate that the spinner will land on 2 or on 3

.....

(2)

Helen concludes that the 3-sided spinner is most likely to land on a score of 3 the next time it is spun.

(d) Comment on the reliability of Helen's conclusion.

.....

.....

(2)

(Total for Question 5 is 7 marks)

6 Richard works in an animal rescue centre.

Richard has collected data on the weights, in kilograms, of 10 male cats and the weights, in kilograms, of 10 female cats at the centre.

Male	3.0	3.2	3.3	3.5	3.6	3.8	3.9	4.2	4.4	4.9
Female	3.0	3.1	3.1	3.2	3.3	3.3	3.5	3.7	3.9	9.5

Richard wants to compare the average weight of the male cats with the average weight of the female cats.

Richard thinks that he should use either the mean or the median.

(a) Which one of the mean or the median do you think he should use?
Give a reason for your answer.

.....

.....

.....

(2)

Richard plans to use a scatter diagram in order to compare the weights of the male cats with the weights of the female cats.

(b) Discuss whether or not a scatter diagram would be a suitable diagram to use.

.....

.....

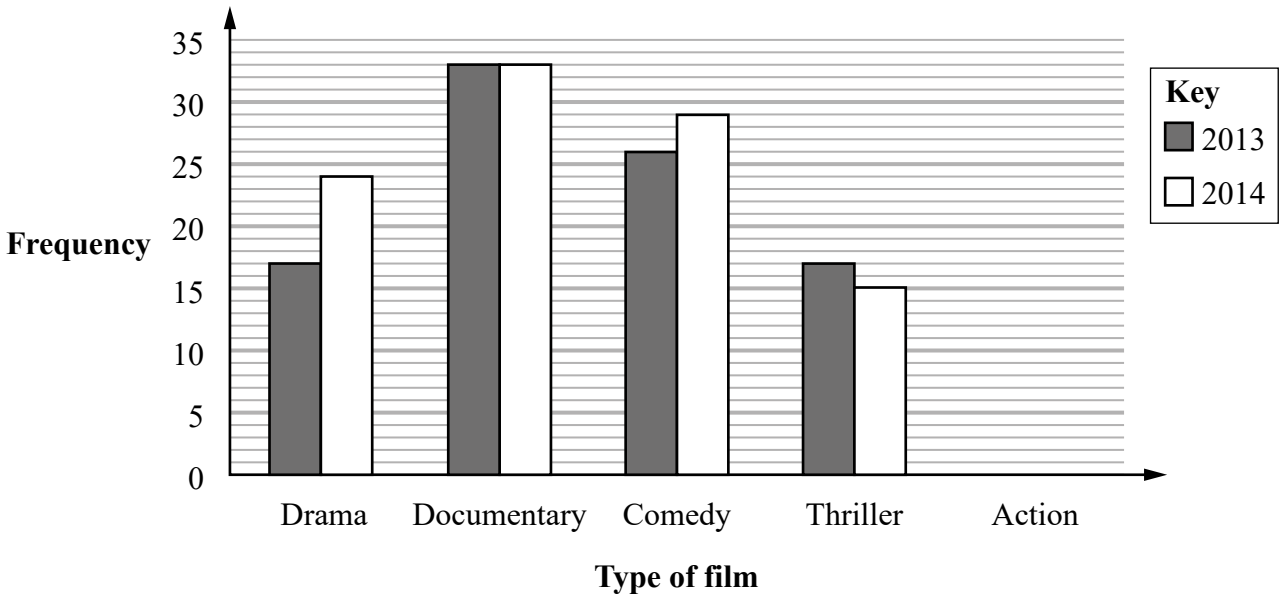
.....

.....

(2)

(Total for Question 6 is 4 marks)

7 The incomplete multiple bar chart shows information about the numbers of UK films first shown in 2013 and in 2014 for some types of film.



(Source: *BFI Statistical Yearbooks*)

9 UK action films were first shown in 2013

7 UK action films were first shown in 2014

(a) Complete the multiple bar chart.

(2)

The number of UK films first shown in 2013 was 17 for two types of film.

(b) Which two types of film?

..... and .....

(1)

(c) Work out the total number of drama films that were first shown in these two years.

.....

(1)

(d) Compare the numbers of UK films of the different types that were first shown in 2013 with those first shown in 2014

.....

.....

.....

.....

(2)

(e) Explain why it would not be appropriate to display the information from the multiple bar chart in a time series graph.

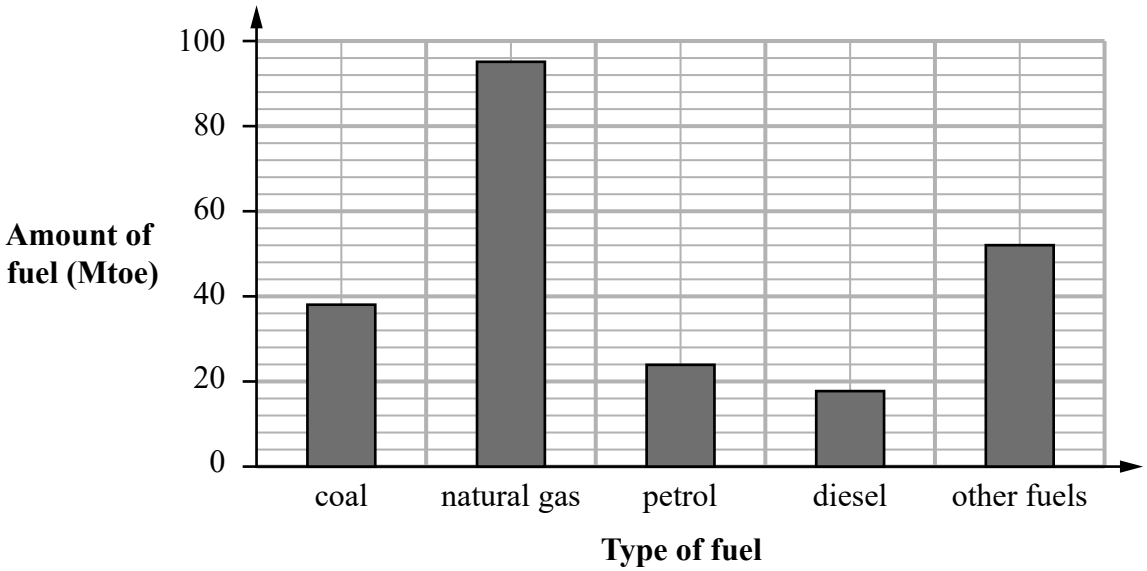
.....

.....

(1)

(Total for Question 7 is 7 marks)

8 The bar chart gives information about the amounts of different types of fuel, in Mtoe (millions of tonnes of oil equivalent), that were used in the United Kingdom in the year 2000



(Source: *Department of Energy and Climate Change*)

Ignoring 'other fuels',

(a) write down the type of fuel that used the least amount of Mtoe in the year 2000

.....

(1)

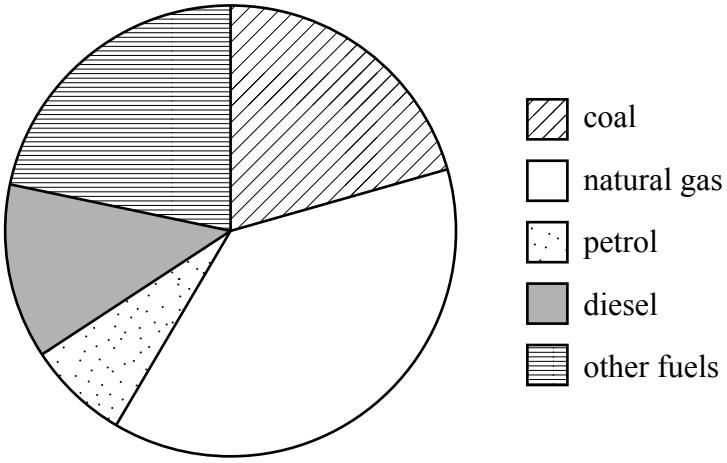
One type of fuel used one quarter of the amount of Mtoe used by natural gas in the year 2000

(b) Which type of fuel was this?

.....

(1)

The pie chart shows information about the amounts of different types of fuel, in Mtoe, that were used in the United Kingdom in the year 2013



The amount of natural gas used, in Mtoe, was the greatest in both 2000 and 2013

(c) Explain how the bar chart and the pie chart show this.

.....

.....

.....

(1)

Juanita says that more diesel was used in 2013 than in 2000

(d) Explain whether or not the information given in the two diagrams can be used to support this statement.

.....

.....

.....

.....

.....

.....

(2)

(Total for Question 8 is 5 marks)

9 The directors of a company want to make changes to the company's pension scheme.

The directors want to find out what the employees think about the proposed changes to the pension scheme.

The directors will collect the information by using one of two data collection methods.

Method 1: each employee will be interviewed by one of the directors.

Method 2: each employee will complete a questionnaire without filling in their name.

There are 100 employees in the company.

Discuss how appropriate each of these two data collection methods are.

(Total for Question 9 is 4 marks)

10 The table shows information about the ages, when elected, of French presidents and UK prime ministers for the years 1850 to 2015

	Lowest value	Lower quartile	Median	Upper quartile	Highest value
French presidents	40	53	60.5	65	74
UK prime ministers	43	53	56.5	63	70

(Source: *Wikipedia*)

Compare and interpret the spread of ages of French presidents with UK prime ministers for the years 1850 to 2015

(Total for Question 10 is 5 marks)

11 Kerry is investigating whether there is a difference in the lengths of the text messages sent by boys and sent by girls at her school.

She writes the following hypothesis for the investigation.

"The length of text messages sent by girls is greater than the length of text messages sent by boys".

Kerry decides to use a census of the 800 students in her school.
She is going to ask each student to record the number of characters in their last text message.

Kerry then collects this information from each student through an online database.

Part of the database is shown below.

	Gender	**Length of text message**
1	male	73
2	F	68
3	girl	thirty five
4	boy	114,
5	boy	85
6	girl	
7	M	56
8	48	boy
9	girl	5
10	G	75
11	B	41
12	girl	28

(a) Give **two** reasons why Kerry must clean the data before processing it.

Reason 1: .....

.....

Reason 2: .....

.....

(2)

(b) Discuss how Kerry's data collection plan could affect the reliability of her conclusions.

.....

.....

.....

.....

.....

.....

.....

.....

(2)

(Total for Question 11 is 4 marks)

12 The table shows information about houses for sale in Oxford.

Number of bedrooms	1	2	3	4	5 or more	Total
Number of houses for sale	140	300	420	240	100	1200

(Source: *adapted from rightmove.co.uk*)

An estate agent says the mode of the number of bedrooms for these houses is 3

(a) Explain how she knows this.

.....

.....

(1)

The estate agent wants to investigate the prices of these houses.

She takes a stratified sample of 60 houses according to the number of bedrooms.

(b) Work out the number of houses in her sample for each number of bedrooms.

Number of bedrooms	1	2	3	4	5 or more
Number of houses in the sample					

(3)

(c) Describe how to select the 60 houses in the sample.

.....

.....

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 12 is 7 marks)

13 Morgan is investigating the 180 Year 11 students in his school.

He collected information from the 30 students in his class.

Part of the spreadsheet he used to collect the information is shown below.

Student number	Name	Left-handed (1 = yes, 0 = no)	Height (cm)
1	Jason	0	169
2	Rami	1	165
29	Youen	0	164
30	Elena	0	162
Total		4	5031

Morgan uses these results to find estimates for all Year 11 students.

(a) Find his estimate for the number of left-handed students in Year 11

.....

(2)

(b) Explain how Morgan can use the information in the spreadsheet to estimate the mean height of all the students in Year 11

.....

.....

.....

(1)

(Total for Question 13 is 3 marks)

14 The table shows information about the retail price index (RPI) and the price of a second-class stamp (in pence) in the United Kingdom for January 1996, January 2006 and January 2016

	Jan 1996	**Jan 2006**	**Jan 2016**
Retail price index (RPI)	100	129	172
Price of second-class stamp (pence)	20	23	54

(Sources: *ons.gov.uk and royalmail.com*)

Describe how the increase in the price of a second-class stamp compares with the RPI over the ten years to January 2006 **and** over the twenty years to January 2016

(Total for Question 14 is 5 marks)

15 The table and the time series graph give some information about the numbers of visits abroad (in millions) made by UK residents for the years 2013 to 2015

Year	Quarter	Number of visits (millions)	4-point moving average (millions)
2013	1	10.2	
	2	15.9	14.45
	3	20.1	14.55
	4	11.6	14.75
2014	1	10.6	14.85
	2	16.7	15.00
	3	20.5	15.225
	4	12.2	15.55
2015	1	11.5	16.00
	2	18.0	
	3	22.3	
	4	13.4	

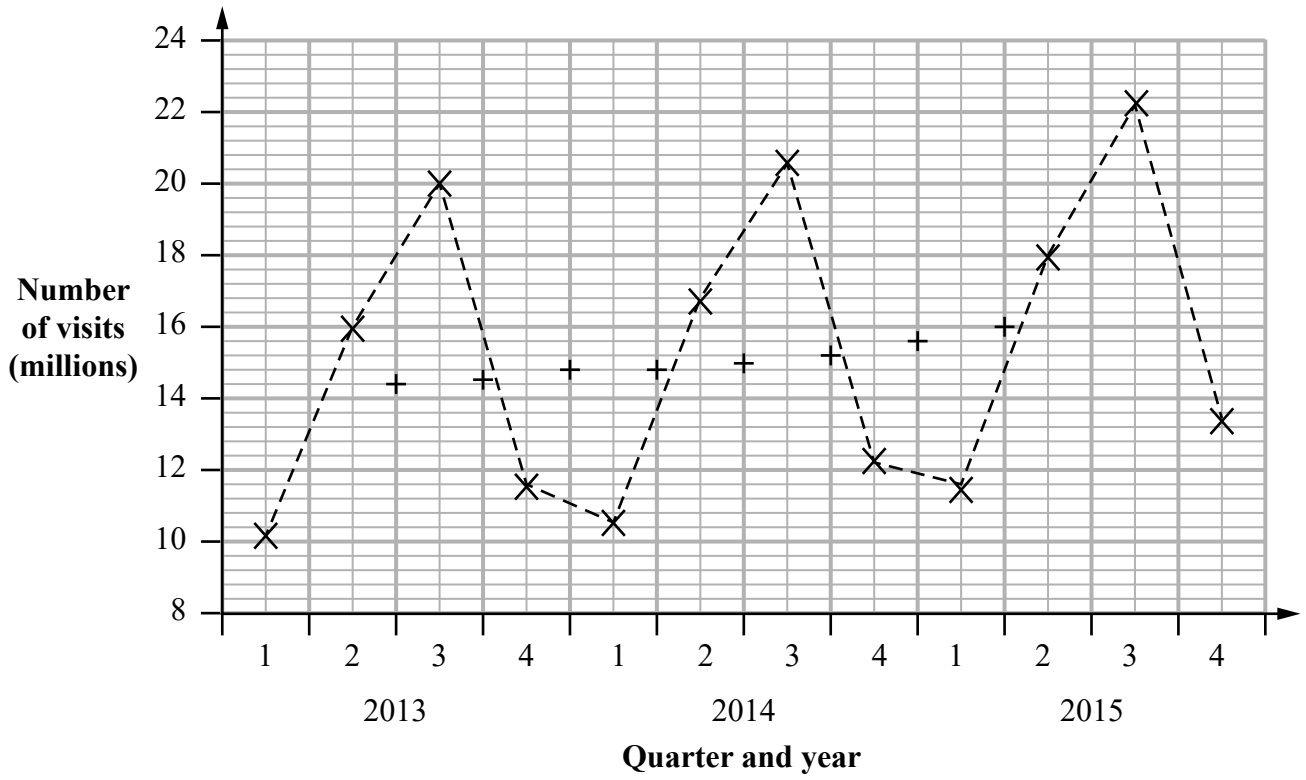
(Source: *Office for National Statistics*)

The last 4-point moving average is missing from the table and from the time series graph.

(a) (i) Calculate this 4-point moving average.
Write your answer in the table.

(ii) Plot this 4-point moving average on the time series graph.

(3)



(b) For which quarter was the number of visits abroad by UK residents the greatest?

.....

(1)

(c) Describe and interpret the trend in the number of visits abroad made by UK residents for the years 2013 to 2015

.....

.....

.....

.....

(2)

(d) Compare and interpret the seasonal variation for each quarter in the number of visits abroad made by UK residents for the years 2013 to 2015

.....

.....

.....

.....

(2)

(Total for Question 15 is 8 marks)

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

BLANK PAGE

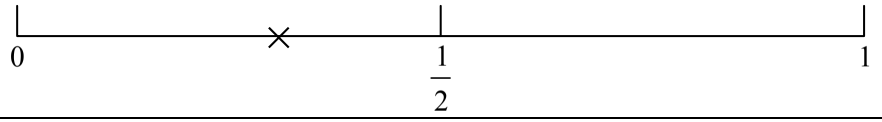
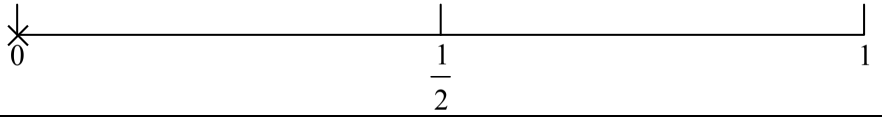
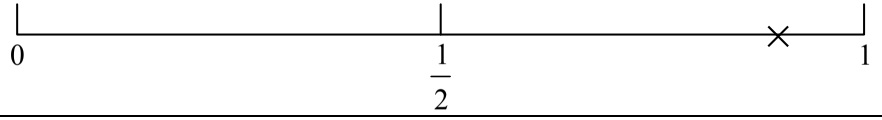
BLANK PAGE

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Paper 1 Foundation tier mark scheme

Question number	Answer	Additional guidance	Mark
1(a)	B1 Maths 		**(1)**
1(b)	M1 10 – 4 or 3×2 A1 6		**(2)**
1(c)	B1 4 + 10 + 2 + 6 [= 22] or 8×'2' or '11'×2 B1 22 > 20 so yes, Milly has spent a total of more than 20 hours studying.	B1 for calculating statistic B1 for conclusion in context based on correct statistical reasoning	**(2)**
2(a)	B1 	B1 for × indicated between 0 and ½ (but closer to ½).	**(1)**
2(b)	B1 	B1 for × indicated at 0	**(1)**
2(c)	B1 	B1 for × indicated between ½ and 1 (but closer to 1).	**(1)**

<table>
<tr><th>Question number</th><th>Answer</th><th>Additional guidance</th><th>Mark</th></tr>
<tr><td>3(a)</td><td>B1B1
<table>
<tr><th>Number of times</th><th>Tally</th><th>Frequency</th></tr>
<tr><td>1</td><td>𝍸 𝍸 ||||</td><td>14</td></tr>
<tr><td>2</td><td>𝍸 ||</td><td>7</td></tr>
<tr><td>3</td><td>||||</td><td>4</td></tr>
<tr><td>4</td><td>|||</td><td>3</td></tr>
<tr><td>5</td><td>||</td><td>2</td></tr>
</table></td><td>1st B1 any one row or one column correct
2nd B1 all correct</td><td>(2)</td></tr>
<tr><td>3(b)</td><td>B1ft 1</td><td></td><td>(1)</td></tr>
<tr><td>3(c)</td><td>M1 ‘14’ + ‘7’ + ‘4’
A1ft 25</td><td>M1 for the addition of correct 3 numbers from their tally chart
A1 ft their tally chart</td><td>(2)</td></tr>
<tr><td>3(d)</td><td>B1 e.g. bar chart</td><td>B1 for any suitable diagram, e.g. pie chart, bar line chart</td><td>(1)</td></tr>
</table>

Question number	Answer	Additional guidance	Mark
4(a)	B1 e.g. Some employees may not wish to give their salary (confidential/personal).		**(1)**
4(b)	B1 All of the employees at the company.		**(1)**
4(c)(i) **4(c)(ii)**	B1 e.g. primary data is data you collect yourself. B1 e.g. secondary data is data collected by someone else. B1 e.g. known reliability, no secondary data available		**(2)** **(1)**
4(d)	B1 Advantage: • e.g. Convenient • e.g. Equal number of males and females represented B1 Disadvantage: • e.g. Not representative • e.g. Biased • e.g. Employees coming early may have similar pay	1st B1 any one correct advantage 2nd B1 any one correct disadvantage	**(2)**

Question number	Answer	Additional guidance	Mark
5(a)	B1 $\frac{1}{4}$ oe	For probability answers accept equivalent fractions, decimals or percentages	**(1)**
5(b)	M1 $\frac{1}{4} \times 60$ A1 15		**(2)**
5(c)	M1 $\frac{9+29}{45}$ or $1-'\frac{7}{45}'$ A1 $\frac{38}{45}$ oe	For probability answers accept equivalent fractions, decimals or percentages	**(2)**
5(d)	B2 e.g. 'reliable since the conclusion is based on a large number of spins' or e.g. 3 is scored many more times than 1 or 2	B2 for 'reliable' with supporting reason based on collected data OR B1 reliable with incomplete reasoning	**(2)**

Question number	Answer	Additional guidance	Mark
6(a)	B2 for median AND a correct reason, e.g. • reference to the extreme value in the female cat weights • reference to the median not being affected by extreme values OR if B2 not earned... B1 for an incomplete answer e.g. median with an attempt at a reason, OR correct reason without a conclusion OR if B1 not earned... SC B1 for mean AND uses all of the data	B2 for a complete answer assessing the appropriateness of the choice of average. OR if B2 not earned... B1 for an incomplete answer assessing the appropriateness of the choice of average.	**(2)**
6(b)	B2 Not suitable as the data is not bivariate/in related pairs oe OR if B2 not earned... B1 Not suitable, with attempt at a reason	B2 for a complete answer assessing the appropriateness of the choice of diagram OR if B2 not earned... B1 for an incomplete answer assessing the appropriateness of the choice of diagram	**(2)**

Question number	Answer	Additional guidance	Mark
7(a)	B1 Bar heights of 9 and 7 drawn for action B1 Correct shading		**(2)**
7(b)	B1 Drama and Thriller		**(1)**
7(c)	B1 41		**(1)**
7(d)	B1B1 Any two from: • More drama films in 2014 • More comedy films in 2014 • Same number of documentary films in both years • Fewer thriller films in 2014 • Fewer action films in 2014 • More total films in 2014	B1 for each statistical conclusion. Allow equivalent converse statements.	**(2)**
7(e)	B1 for e.g. 'There are only two years worth of data'/'A time series graph is better for data over a longer period of time'.	B1 for appropriate reason why a time series graph would not be suitable for this data.	**(1)**
8(a)	B1 Diesel		**(1)**
8(b)	B1 Petrol		**(1)**
8(c)	B1 Natural gas is tallest bar on bar chart and largest sector on pie chart.		**(1)**
8(d)	B1 The diagrams cannot be used to support the statement. B1 The pie chart shows proportions (not amounts) oe		**(2)**

Question number	Answer	Additional Guidance	Mark
9	B1 B1 B1 B1 for each of four aspects from: • Understanding e.g. Questions/responses can be explained in an interview or may not be understood in a questionnaire • Candour e.g. Employee may be less open/honest in an interview or questionnaire can be done without pressure • Resources e.g. Interviewing can be time consuming/expensive or questionnaire can be done by all at the same time (or in their own time, or more cheaply) • Inclusivity e.g. interviews more likely to include all employees or questionnaires might not be returned or directors' views are not included by interview • Interviewer bias e.g. Possible bias from director (in interview)/ no interviewer bias with questionnaire	B1 for each correct comment assessing the appropriateness of the data collection methods.	**(4)**

Question number	Answer	Additional guidance	Mark
10	M1A1A1 for calculation of range or IQR OR for a pair of box plots drawn **Range** M1 74 – 40 or 70 – 43 A1 34 A1 27 OR **IQR** M1 65 – 53 or 63 – 53 A1 12 A1 10 B1 for e.g. The range/IQR for French presidents is greater than the range/IQR for UK prime ministers oe B1 for e.g. The ages of French presidents are more varied than the ages of UK prime ministers	M1 for a calculation of range or IQR OR for a pair of box plots drawn A1 for one correct value OR for one correct box plot A1 for second correct value to allow a comparison to be made OR for second correct box plot (same scale) to allow comparisons to be made B1 for a correct comparison of ranges or IQRs B1 for a correct contextual interpretation of comparison of spread	**(5)**

Question number	Answer	Additional guidance	Mark
11(a)	B1B1 for two correct reasons • e.g. data given in different formats • e.g. remove extraneous symbols • e.g. remove anomalies/outliers • e.g. data given in wrong order	B1 for each correct reason for the need to clean data on the database prior to processing it	**(2)**
11(b)	B1B1 for two correct aspects • e.g. large sample size increases reliability • e.g. issues due to how the data collection is carried out may decrease reliability (recorded by students and not by Kerry/students may type in information wrong) • e.g. for an example of factors that might not be consistent in the data collection which may decrease reliability (last message might not represent all text messages sent) • e.g. non-response decreases reliability	B1 for each correct comparison assessing the reliability of the conclusions drawn	**(2)**

<table>
<tr><th>Question number</th><th>Answer</th><th>Additional guidance</th><th>Mark</th></tr>
<tr><td>12(a)</td><td>B1 3 bedrooms has the highest frequency (for any individual number of bedrooms)</td><td>Allow equivalent statistical reasoning based on the table indicating why 3 is the mode.</td><td>(1)</td></tr>
<tr><td>12(b)</td><td>M1 $\frac{140}{1200}\times 60$ o.e.

A1A1
<table><tr><td>Bedrooms</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5+</td></tr><tr><td>Houses in sample</td><td>7</td><td>15</td><td>21</td><td>12</td><td>5</td></tr></table></td><td>Accept a correct equivalent calculation shown for any one class
M1 implied by one correct answer OR an indication they need 1 in 20

1st A1 for any one value correct
2nd A1 for all correct</td><td>(3)</td></tr>
<tr><td>12(c)</td><td>B1 Use a sampling frame for each strata

B1 Select houses randomly or generate random numbers

B1 For an aspect of detail</td><td>Each category/strata to be considered separately

Samples have to be random

e.g. How the random numbers are obtained and used</td><td>(3)</td></tr>
</table>

Question number	Answer	Additional guidance	Mark
13(a)(i)	M1 $\frac{4}{30}\times 180$ A1 $=24$	Accept any equivalent calculation	(2)
13(b)	B1 e.g. 'use the mean of the sample of an estimate''	B1 for a correct explanation showing understanding that the mean of the sample represents the mean of the population	(1)
14	M1A1A1 for appropriate calculations to compare 1996 with 2006 or 1996 with 2016 e.g. **Comparing index numbers** M1 $\frac{23}{20}\times 100$ or $\frac{54}{20}\times 100$ A1 115 A1 270 OR **Comparing prices** M1 $20\times\frac{129}{100}$ or $20\times\frac{172}{100}$ A1 25.8 A1 34.4 B1 for e.g. from 1996 to 2006 the change/increase in price was less than the RPI B1 for e.g. from 1996 to 2016 the change/increase in price was higher than the RPI	M1 for a calculation that could be used to compare 1996 with 2006 or 2016 A1 for one correct value A1 for second correct value that allows a comparison to be made B1 for a correct contextual statement for 2006 or 2016 B1 for a correct contextual statement for 2006 and 2016	(5)

Question number	Answer	Additional guidance	Mark
15(a)(i)	M1 $\frac{11.5+18+22.3+13.4}{4}$		**(3)**
	A1 16.3		
(a)(ii)	B1ft Point plotted at height '16.3' between Q2 and Q3 2015		
15(b)	B1 Quarter 3 (2015)		**(1)**
15(c)	B1 Upwards B1 The number of visits abroad is increasing as time goes by.	B1 for a correct description of the trend B1 for a contextualised interpretation	**(2)**
15(d)	B1 e.g. Seasonal variation is highest in Q3/lowest in Q1 B1 e.g. More visits abroad in Q3/Fewer visits abroad in Q1	B1 for a correct comparison of one of the quarters with the others B1 for a contextualised interpretation	**(2)**

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

Centre Number | | | | | |

Candidate Number | | | | |

Sample assessment material for first teaching September 2017

Morning (Time: 1 hour 30 minutes) | Paper Reference **1ST0/2F**

Statistics

Paper 2
Foundation Tier

You must have:
Ruler graduated in centimetres and millimetres, protractor, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided
  – *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified at the end of your solution**.

Information

- The total mark for this paper is 80.
- The marks for **each** question are shown in brackets
  – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

S54192A

1/1/1/1/

S 5 4 1 9 2 A 0 1 2 4

Pearson

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 Jenny wants to find out what students at her school think about the after-school clubs.

Jenny is going to use a questionnaire.

Here is one of the questions she wants to put on the questionnaire.

It is great that we have a range of clubs at school, isn't it? Yes ☐ No ☐ Don't know ☐

(a) This is not a suitable question.
Explain why.

.....

(1)

Here is another of the questions that Jenny wants to put on the questionnaire.

How many times a week do you go to an after-school club? 1–2 ☐ 2–3 ☐ 4–5 ☐

(b) Discuss whether or not this is a suitable question for the questionnaire.

.....

.....

.....

(2)

(Total for Question 1 is 3 marks)

2 Tom and Samira want to collect data on the numbers of hours students at their school spend on homework.

There are 1100 students at their school.

Tom is planning to use a random sample of 50 students.

(a) Explain what is meant by a random sample.

.....

.....

.....

(1)

(b) Describe how Tom could use random numbers to take a random sample of the students at his school.

.....

.....

.....

.....

(3)

Samira is planning to use a stratified sample that is stratified by school year.

(c) Comment on whether Samira's plan is appropriate.

.....

.....

.....

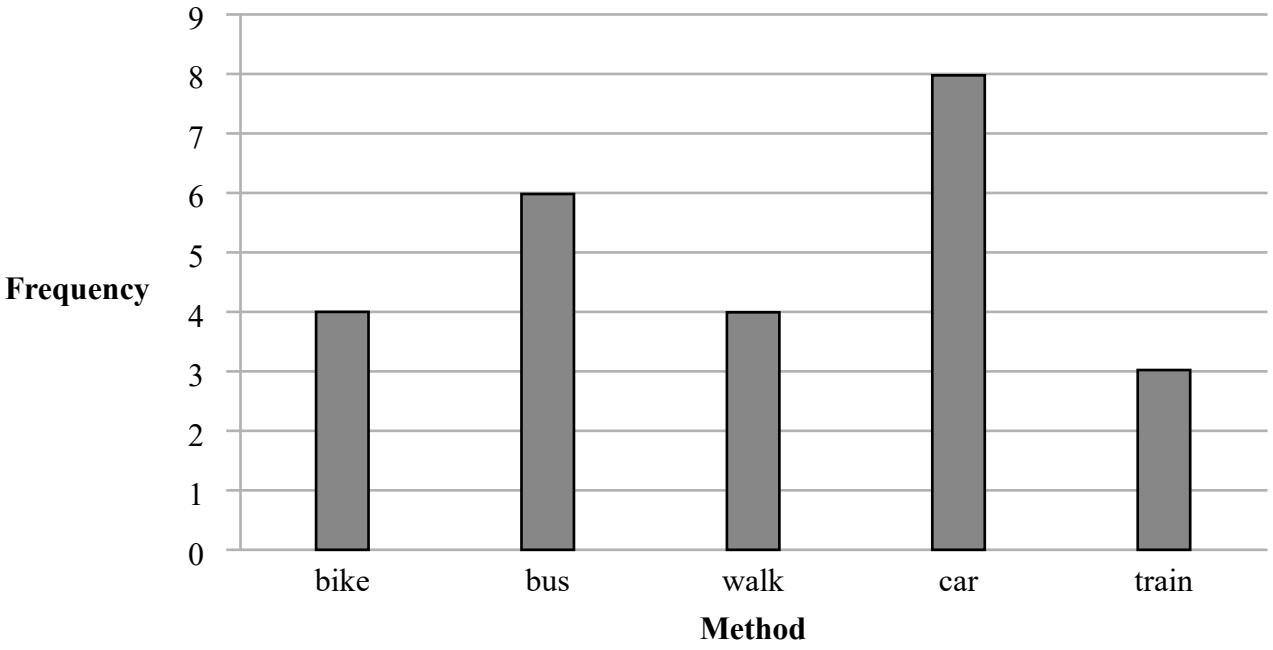
.....

(2)

(Total for Question 2 is 6 marks)

3 Razwan collected data about the methods used to get to work that morning by the 25 people who work at his company.

The bar chart shows information about the methods used.



(a) Which method was used by the greatest number of people?

.....

(1)

(b) Which method was used by half as many of the people who got to work by bus?

.....

(1)

Razwan concludes that travelling to work by car is the most popular method used to get to work in his city.

(c) Give one limitation of Razwan's conclusion.

.....

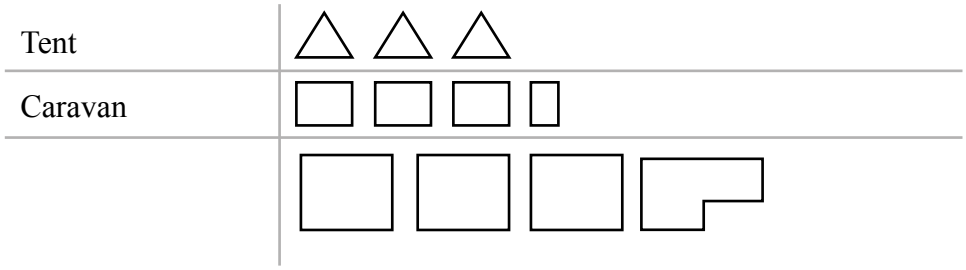
.....

(1)

(Total for Question 3 is 3 marks)

4 Rebecca collected information about the type of accommodation that 30 people stayed in the last time they went on holiday.

She drew this diagram to show her results.



Write down three things that could be misleading or that are wrong in Rebecca's diagram.

1 .....

2 .....

3 .....

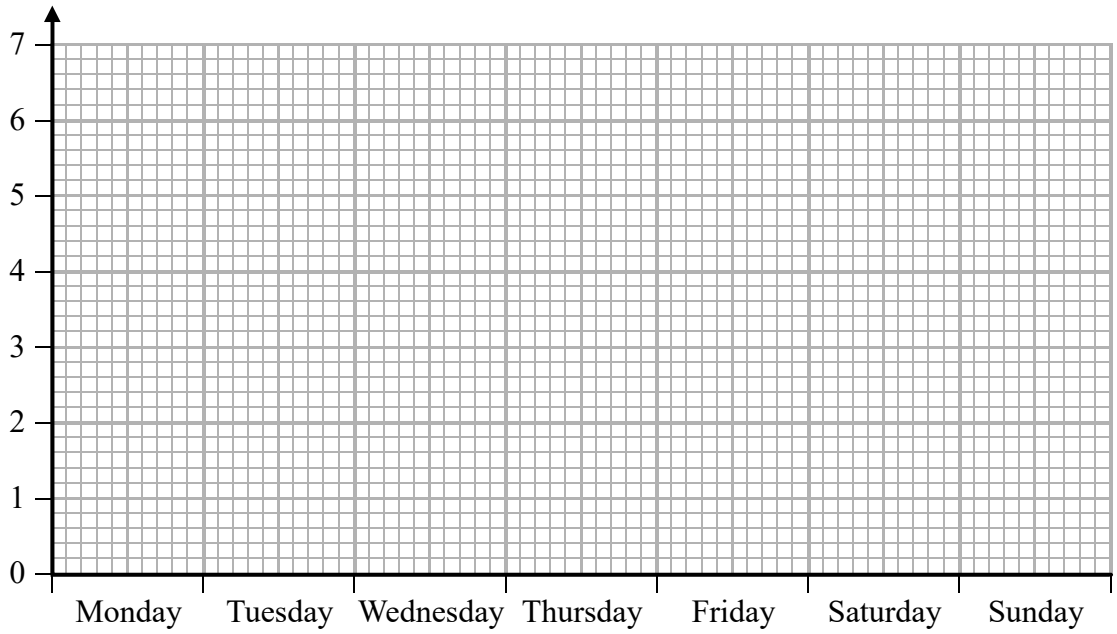
(Total for Question 4 is 3 marks)

5 Diane recorded the number of hours that she watched television each day last week.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Number of hours	2	1	2	0	4	6	6

(a) Draw a line graph for this data.

Label each axis.



(3)

(b) Calculate the mean number of hours.

..... hours

(2)

(c) Find the median.

..... hours

(2)

Noah recorded the number of hours that he watched television each day last week.

He calculated the mean and the median of his results.

	Mean	Median
Number of hours	4	1

(d) Use your answers to part (b) and part (c) to compare the average amount of television watched by Diane and by Noah last week.

.....

.....

.....

(2)

(Total for Question 5 is 9 marks)

6 Claire collected data on the weights of the England football team and the weights of the England rugby team from the internet.

She calculated the mean and range of the weights of each team.
Her results are shown in this table.

	Mean	**Range**
Football team	77.0 kg	30 kg
Rugby team	104.6 kg	42 kg

(Sources: *thefa.com and englandrugby.com*)

(a) State **two** possible problems with obtaining data from the internet.

.....

.....

.....

.....

(2)

(b) Use the information in the table to compare the distribution of weights of the England football team with the weights of the England rugby team.
Interpret your comparison.

.....

.....

.....

.....

.....

.....

(3)

(c) Suggest a possible problem with collecting primary data in this situation.

.....

.....

(1)

(Total for Question 6 is 6 marks)

This question must be answered with a cross in a box ☒. If you change your mind about an answer, put a line through the box ☒ and then mark your new answer with a cross ☒.

7 Raina has been watching the judging of a cake baking competition.

Two judges ranked the 10 bakers for their sponge cakes.

Raina calculated the Spearman's rank correlation coefficient for the ranks given by the judges.

She got a value of 0.8

(a) (i) What type of correlation is shown by the value 0.8?
Put a cross in one of the boxes below.

Negative correlation ☒ **No correlation** ☒ **Positive correlation** ☒

(ii) Interpret Raina's value.

.....

.....

(2)

The same two judges will also be judging a flower-arranging competition.

(b) Is it possible to say anything about the ranks they are likely to give for flower arranging based on the value of Spearman's rank correlation coefficient that Raina calculated?

Give a reason for your answer.

.....

.....

.....

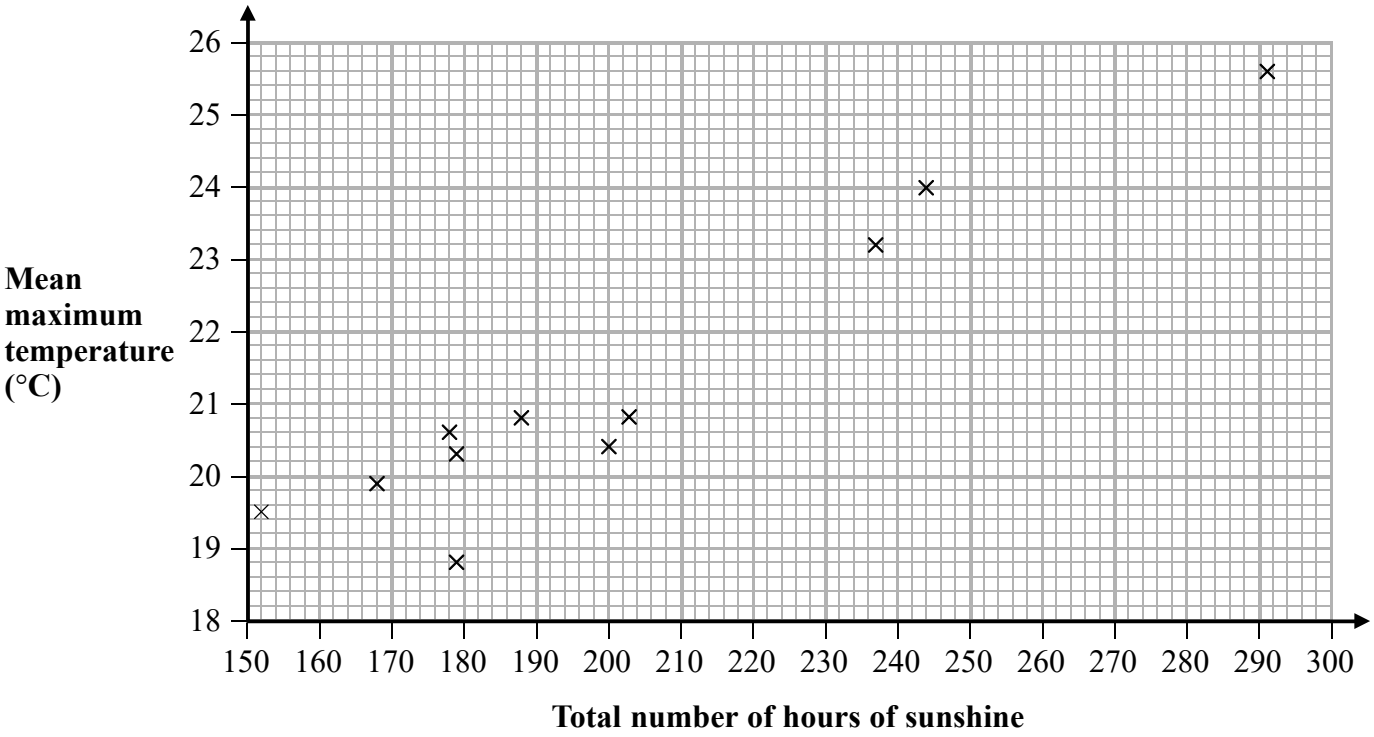
(1)

(Total for Question 7 is 3 marks)

8 The maximum temperature (°C) and the number of hours of sunshine were recorded in Sheffield each day in July for 11 successive years.

For each July, the mean maximum temperature and the total number of hours of sunshine were calculated.

The scatter diagram shows this information.



(Source: *Metoffice*)

In one of these years there was a total of 244 hours of sunshine in July.

(a) For this year, write down the mean maximum temperature.

..... °C

(1)

(b) For the year with the lowest total number of hours of sunshine in July, write down the mean maximum temperature.

..... °C

(1)

(c) Draw a line of best fit on the scatter diagram.

(1)

(d) Describe and interpret the type of correlation shown by the scatter diagram.

.....

.....

.....

.....

(3)

For a different year in Sheffield, there was a total of 220 hours of sunshine in July.

(e) (i) Estimate the mean maximum temperature for July that year.

..... °C

(ii) Give a reason why your answer to part (e)(i) should be reliable.

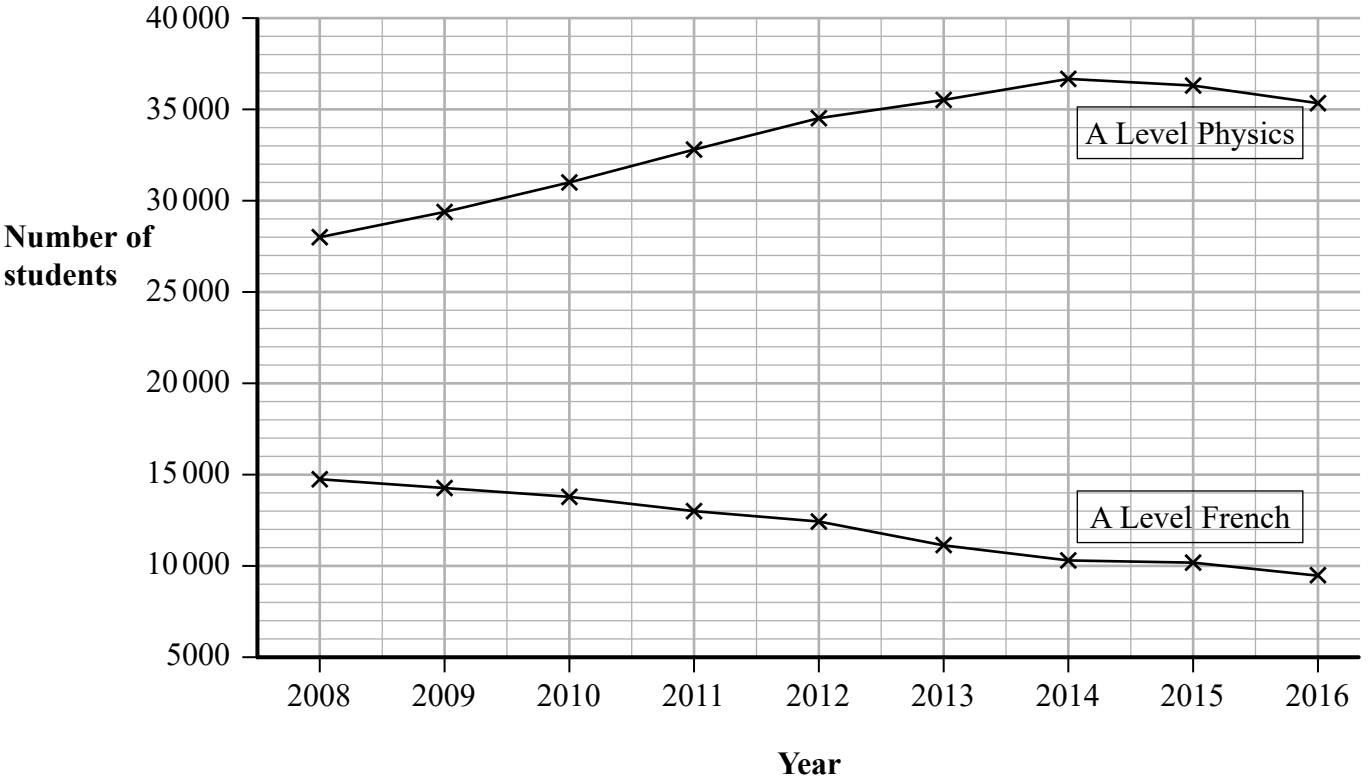
.....

.....

(2)

(Total for Question 8 is 8 marks)

9 The time series graphs show information about the numbers of students taking A Level Physics and A Level French in each of the years from 2008 to 2016



(Source: *JCQ*)

(a) Work out an estimate for the difference in the number of students taking A Level Physics and the number of students taking A Level French in the year 2008

.....

(2)

(b) Explain why the answer to part (a) is an estimate.

.....

.....

(1)

(c) Describe the trend in the number of students taking A Level French from 2008 to 2016

.....

.....

(1)

Alizee is going to give a presentation to students at a school. She wants to show her target audience that Physics is the most popular subject at A Level.

(d) Explain whether or not it would be appropriate for Alizee to use the time series graph to show this.

.....

.....

(1)

Ivan is going to give a presentation to students at a school. He wants to show his target audience that the number of students taking A Level Physics is decreasing.

(e) Explain how Ivan can use information from the time series graph to show this.

.....

.....

(1)

Lionel wants to use the time series graphs to predict the number of students taking A Level Physics and the number of students taking A Level French in 2020

(f) Discuss whether or not it would be appropriate to do so.

.....

.....

.....

.....

.....

(2)

(Total for Question 9 is 8 marks)

10 27 adults were each asked to count the number of times they could bounce a ball on a bat.

Here are the results.

5	8	13	5	7	23	30	6	21
24	23	22	13	9	12	6	12	34
22	20	35	22	12	16	24	13	12

(a) Complete the stem and leaf diagram for this information.

0	
1	
2	
3	

Key: 0 | 5 represents 5 bounces

(2)

(b) Work out the interquartile range of the results.

.....

(2)

The median of the results for 27 children is 9

The interquartile range of the results for these children is 6

Alex thinks these results show that adults are better than children at bouncing a ball on a bat.

(c) Do you agree with Alex?
You must give reasons for your answer.

.....

.....

.....

.....

.....

.....

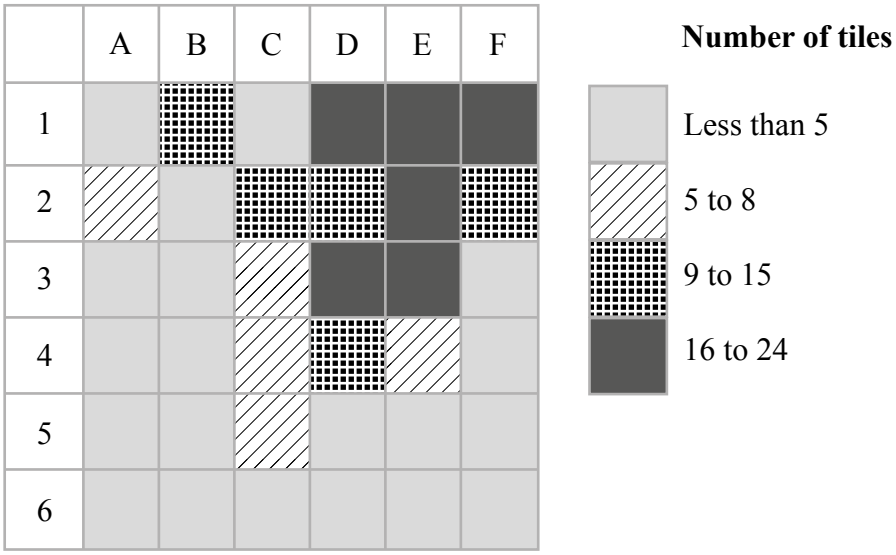
(3)

(Total for Question 10 is 7 marks)

11 Archaeologists divided a field into 36 squares of equal size.

The number of Roman roof tiles found in each square was recorded.

The choropleth map below was drawn using this information.



Use the choropleth map to describe where in the field the greatest number of roof tiles was found.

Give a statistical reason for your answer.

.....

.....

.....

.....

(Total for Question 11 is 2 marks)

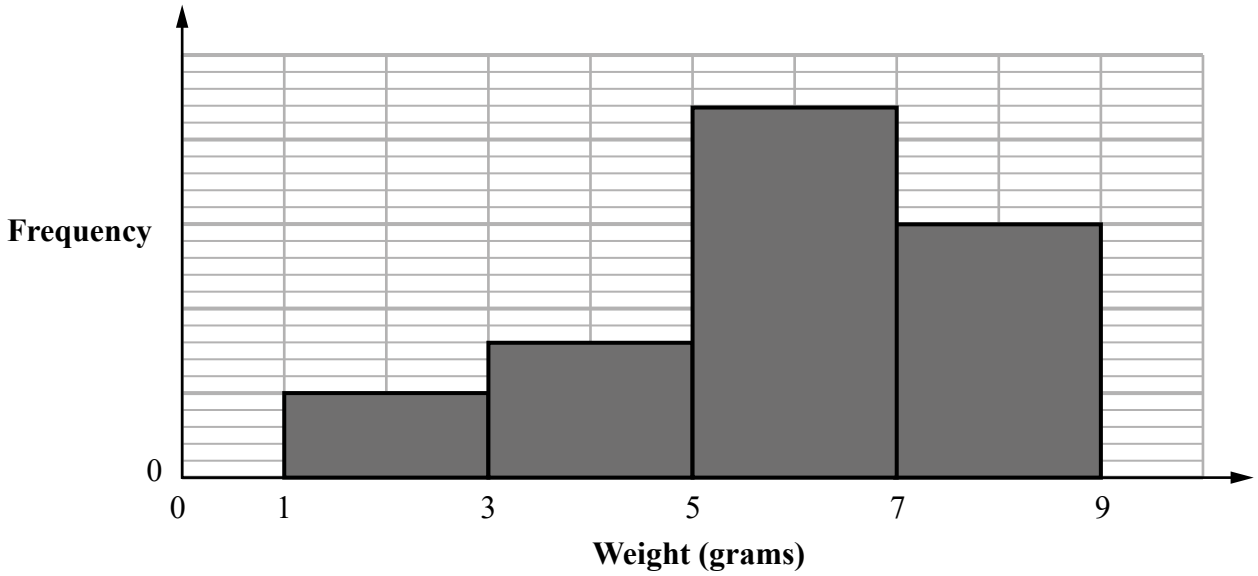
12 Tomoyo found the weight, in grams, of each of 100 cherries.

(a) Circle the **two** words from the list that best describe the data Tomoyo found.

quantitative **qualitative** **discrete** **continuous** **bivariate** **ordinal** **categorical**

(2)

Tomoyo grouped the weights and she then drew this diagram for her results.



The incomplete frequency table shows some information about her results.

Weight (w grams)	Frequency
$1 \leqslant w < 3$	10
$3 \leqslant w < 5$	
$5 \leqslant w < 7$	
$7 \leqslant w < 9$	

(b) (i) Complete the frequency column in the table.

(2)

(ii) Calculate an estimate of the mean weight of the 100 cherries.

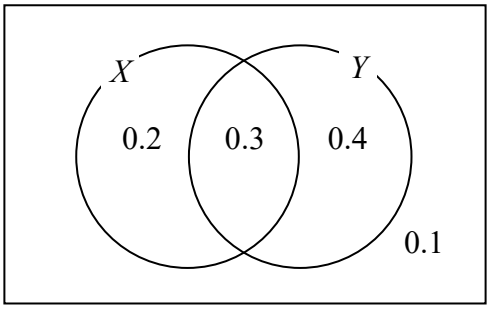
..... g

(3)

(Total for Question 12 is 7 marks)

13 X and Y are two events.

The Venn diagram shows information about the probabilities of events related to X and Y happening.



(a) Find

(i) the probability of event Y happening

.....

(1)

(ii) P(X and Y)

.....

(1)

(iii) P($Y \mid X$)

.....

(2)

Two different events A and B are independent

P(A) = 0.8 and P(B) = 0.5

(b) Find P(A and B)

.....

(2)

(Total for Question 13 is 6 marks)

14 Gary is going to investigate the amounts of time students spend watching TV.

He is going to write a plan for this investigation.

His hypothesis is

> "The amount of time that boys spend watching TV is greater than the amount of time that girls spend watching TV".

Write down three other things he should include in his plan.

Explain why each of these things is appropriate.

You must refer to more than one stage of the statistical enquiry cycle.

(Total for Question 14 is 6 marks)

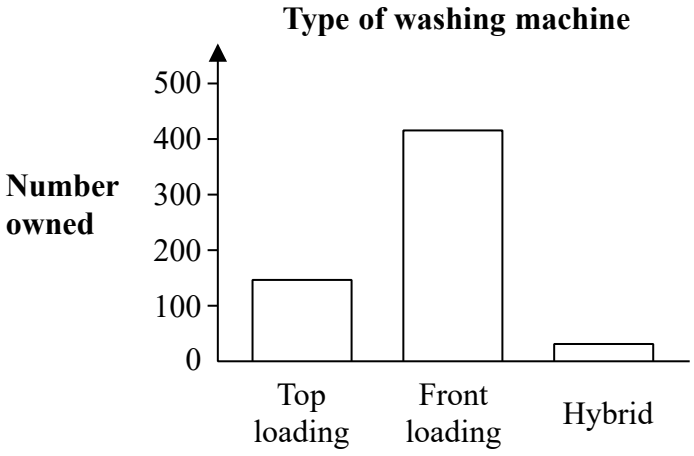
15 The editor of a home magazine collects information about the types of washing machines people use.

She uses statistical software to represent the results in three different ways.

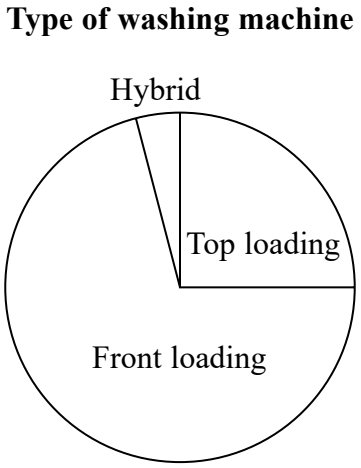
Way 1: Table.

Type of washing machine	**Top loading**	**Front loading**	**Hybrid**
Number owned	150	425	25

Way 2: Bar chart.



Way 3: Pie chart.



For each of the ways, comment on why the editor might choose this way to represent the results in the magazine.

Way 1

.....

.....

Way 2

.....

.....

Way 3

.....

.....

(Total for Question 15 is 3 marks)

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

BLANK PAGE

BLANK PAGE

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

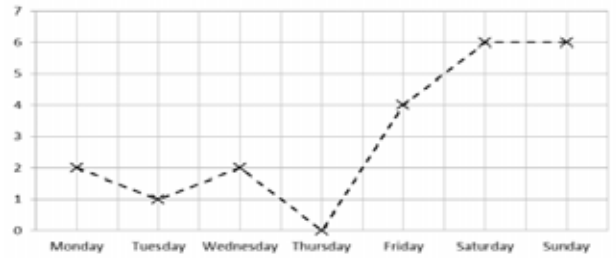
DO NOT WRITE IN THIS AREA

Paper 2 Foundation tier mark scheme

Question number	Answer	Additional guidance	Mark
1(a)	B1 It is a leading question / it is biased	B1 for assessing the appropriateness of the given question	**(1)**
1(b)	B2 for a complete answer, e.g. • Not suitable AND the response boxes overlap • Not suitable AND no option for never / cannot say 6 or more OR if B2 not earned... B1 for an incomplete answer e.g. giving reasons but no conclusion OR B1 for one of • Contains a time frame • It is a closed question which is better than an open question	B2 for assessing the appropriateness of the given question and reaching the correct conclusion OR if B2 not earned... B1 for an incomplete attempt at assessing the appropriateness of the given question	**(2)**

Question number	Answer	Additional guidance	Mark
2(a)	B1 All people/items have the same/equal chance of being chosen.	B1 for demonstrating understanding of a random sample	**(1)**
2(b)	B1B1B1 for each of three aspects from: • Get a list / register (as the sampling frame) of all the students in the school oe • Number the students in the (sampling frame) list / register • Generate random numbers using a calculator / computer / random number table • Students with the matching number are selected	B1×3 for demonstrating understanding of how to select a random sample	**(3)**
2(c)	B2 for a complete answer e.g. the plan is appropriate **AND** e.g. the number of hours spent on homework might differ between school years OR if B2 not earned... B1 for an incomplete answer e.g. the plan is appropriate, with an attempt at a reason OR for correct reason without conclusion	B2 for complete answer assessing the appropriateness of the suggested plan OR if B2 not earned... B1 for an attempt at assessing the appropriateness of the suggested plan	**(2)**

Question number	Answer	Additional guidance	Mark
3(a)	B1 car (cao)		**(1)**
3(b)	B1 train (cao)		**(1)**
3(c)	B1 for any one limitation of Razwan's conclusion e.g. • The data only relates to people working at one company • The data only relates to one morning / people might use different transport on other days • This is only a small sample	B1 for a comment assessing the stated conclusion	**(1)**
4	B1B1B1 for three correct things identified. e.g. • No label for the third row • Different shaped symbols (triangles and rectangles) • Different sized symbols / symbols not aligned • No key • No title	B1 ×3 for demonstrating understanding of key features of a pictogram	**(3)**

Question number	Answer	Additional guidance	Mark
5(a)	B1B1B1  	1st B1 for plotting points correctly 2nd B1 for correctly connecting points with straight lines 3rd B1 for labelling both axes (***hours*** for vertical axis and ***days of week*** for horizontal axis)	**(3)**
5(b)	M1 $\frac{2+1+2+0+4+6+6}{7}$ A1 3	M1 for appropriate addition and division by 7 (condone one error) A1 cao	**(2)**
5(c)	M1 0 1 2 2 4 6 6 A1 2	M1 for ordering or using $(n+1)/2$ A1 cao	**(2)**
5(d)	B1ft Diane has a lower mean than Noah. B1ft Diane has a higher median than Noah.	B1 for each correct statistical conclusion (follow through their values in (b) and (c)	**(2)**

Question number	Answer	Additional guidance	Mark
6(a)	B1B1 for two correct problems e.g. • Data may not be up to date • Data may not be in the required format • Some data may be missing • Reliability of website may not be known		**(2)**
6(b)	B1 • Rugby team mean > football team mean B1 • Rugby team range > football team range B1 e.g. • the rugby team are heavier / football team are lighter • the rugby team have less consistent weights / the football team have more consistent weights	B1B1 for correct statistical reasoning comparing means and ranges B1 for contextual interpretation of a comparison of means or ranges	**(3)**
6(c)	E.g. not able to access all of the players	B1 for any suitable appreciation that secondary data is appropriate	**(1)**

Question number	Answer	Additional guidance	Mark
7(a)(i)	B1 Positive correlation	B1 for correct statistical conclusion	**(1)**
7(a)(ii)	B1 The judges were in reasonable/good agreement oe	B1 for contextual interpretation of conclusion	**(1)**
7(b)	B1 for answer with reason, e.g. • No AND reference to the change in context (sponge cakes versus flower arranging) • Yes, likely to give similar ranks AND reference to the fact that the judges have similar tastes		**(1)**
8(a)	B1 24		**(1)**
8(b)	B1 19.5 (allow answers in the range 19.4-19.6)		**(1)**
8(c)	B1 Line drawn between (170, 19)/(170, 20) and (270, 24)/(270, 25)		**(1)**
8(d)	B1 Positive B1 Strong B1 As the total number of hours of sunshine increases, the mean maximum temperature increases.	A response that covers all 3 aspects can score 3 marks	**(3)**
8(e)(i)	B1ft 21.5 - 22.5	B1 for answer in range 21.5 - 22.5 or follow through value read off their line of best fit with positive gradient.	**(2)**
8(e)(ii)	B1 e.g. 'The result has been interpolated'	B1 for understanding that the estimate is made within the range of given x – values	

Question number	Answer	Additional guidance	Mark
9(a)	M1 28000 – 15000 A1 13000	M1 for use of 28000 and 15000	**(2)**
9(b)	B1 for e.g. 'Exact values can't be read off graph' or 'As values came from a possibly unreliable source (secondary data) they could be inexact'	B1 for reason relating to actual values are unknown/source of data may be unreliable	**(1)**
9(c)	B1 for downwards	B1 for a correct description of the trend	**(1)**
9(d)	B1 for e.g. 'It would not be appropriate to do this as the time series graph only shows information for 2 subjects (and not all subjects)'	B1 for not appropriate with correct supporting reason	**(1)**
9(e)	B1 for e.g. 'By only showing the data from 2014 onwards'	B1 for selecting the part of the data set which shows this information	**(1)**
9(f)	B2 for e.g. 'Not appropriate since trend may not continue'	B2 for assessing the use of the time series graph to predict future with appropriate conclusion OR B1 for 'Not appropriate' with incomplete reasoning	**(2)**

Question number	Answer	Additional guidance	Mark
10(a)	B2 0 | 5566789 1 | 22223336 2 | 012223344 3 | 045	B2 cao OR if B2 not earned B1 for unordered diagram or ordered diagram with at most 2 errors	**(2)**
10(b)	M1 23 – 9 A1 = 14	M1 for attempt at IQR with at least one quartile (23 or 9) correct A1 for 14 cao	**(2)**
10(c)	B1 for adult median is 13 B1ft for the median for the adults is higher than the median of the children B1ft (dep) so yes Alex's conclusion is supported by the data.	B1 adult median is 13 B1 for correct comment assessing the appropriateness of the conclusion 3rd B1 is dependent upon the 2nd B1	**(3)**
11	B1 Top right of grid B1 Squares are shaded darkest in this region.	B1 for a correctly identifying the region (e.g. NE corner) B1 for a statistical reason relating to use of the key.	**(2)**
12(a)	B1 quantitative B1 continuous	B1 for one identified B1 for both identified with no extras	**(2)**
12(b)(i)	M1 16 44 A1 30	M1 for correctly applying a scale to find a second value (implied either by labelling a scale or correctly finding one other frequency). A1 for all values correct.	**(2)**
12(b)(ii)	M1M1A1 $\frac{2\times10+4\times16+6\times44+8\times30}{100}=5.88$	M1 for consistent use of fx with x within interval M1 for correct use of Σfx with x the mid-interval value A1 for 5.88	**(3)**

Question number	Answer	Additional guidance	Mark
13(a)(i)	B1 (0.3+0.4=) 0.7	For probability answers accept equivalent fractions, decimals or percentages	**(1)**
13(a)(ii)	B1 0.3		**(1)**
13(a)(iii)	M1 $\frac{0.3}{0.5}$ A1 0.6		**(2)**
13(b)	M1 0.8×0.5 A1 0.4		**(2)**

Question number	Answer	Additional guidance	Mark
14	**Collecting data** B1 for identifying one appropriate thing that should be included in the plan for collecting data **and** B1 for explaining why this aspect is appropriate OR B1 for deciding what data to collect and/or how to collect and record it **and** B1 for an appropriate reason OR B1 for a strategy to process data **and** B1 for an appropriate reason OR B1 for designing a collection method for primary/secondary data **and** B1 for an appropriate reason OR B1 for appreciating the importance of acknowledging sources **and** B1 for an appropriate reason OR B1 for recognising where issues of sensitivity may influence data availability **and** B1 for an appropriate reason **Processing and presenting** B1 for planning to organise and/or process data **and** B1 for an appropriate reason OR	B1B1B1 for each of three planned elements and B1B1B1 for each of three appropriate reasons from their three things in the statistical enquiry cycle. Maximum 4 marks if only one aspect (from Collecting data, Processing and presenting, Interpreting and Evaluating) is referenced. B1 for e.g. use amount of time measured to the nearest minute **and** B1 for e.g. this is sufficient as there will be a large range of times B1 for e.g. collect data for 23 boys and 23 girls **and** B1 for e.g. this will make the calculation of quartiles easier B1 for e.g. use random sampling **and** B1 for e.g. this reduces bias as Gary's friends/class/peers may generally watch the same programs B1 for e.g. use primary data **and** B1 for e.g. this increases reliability as Gary will know how the data was collected B1 for e.g. A student (Gary) should collect the data **and** B1 for e.g. students are more likely to give an honest answer to a fellow student (less threatening) B1 for e.g. use a grouped frequencies table for the data **and** B1 for e.g. as this will enable a quick way of estimating the mean or this can be used to draw a histogram	**(6)**

Question number	Answer	Additional guidance	Mark
14 ***continued***	B1 for planning to generate diagrams and/or visualisations to represent the data **and** B1 for an appropriate reason OR B1 for planning to generate statistical measures to compare data **and** B1 for an appropriate reason **Interpreting** B1 for planning to interpret diagrams and/or calculations/measures **and** B1 for an appropriate reason OR B1 for planning to make an inference and/or prediction **and** B1 for an appropriate reason **Evaluating** B1 for planning to identify weaknesses in approach or representation **and** B1 for an appropriate reason OR B1 for planning to refine the processes to elicit further clarification of the hypothesis **and** B1 for an appropriate reason	B1 for e.g. use box plots **and** B1 for e.g. these will enable the comparison of both the medians and the IQRs (i.e. the distributions) of the data B1 for e.g. interpret results for each individual school year **and** B1 for e.g. as different years could have different watching habits B1 for e.g. by comparing means **and** B1 for e.g. you can see whether the amount of time that boys spend watching TV is greater, in general, than the amount of time that girls spend watching TV B1 for e.g. use the results from the school to predict the results nationally **and** B1 for e.g. as students in different parts of the country are likely to have the same watching habits B1 for e.g. choose not to display the information in histograms **and** B1 for e.g. as "the target audience" may not know how to interpret them B1 for e.g. consider using more than one type of visual representation (for the same information) **and** B1 for e.g. as different representations focus on different aspects of the data	

Question number	Answer	Additional guidance	Mark
15	B1 for one correct comment for **each** way e.g. Way 1: Shows exact values. Way 2: Shows values/differences clearly or has visual impact. Way 3: Shows proportions clearly or has visual impact.	B1 for **each** correct comment assessing the appropriateness of the ways of representing the results A mark for 'visual impact' can only be awarded once, either as the alternative for Way 2 or as the alternative for Way 3.	**(3)**

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

Centre Number | | | | | |

Candidate Number | | | | |

Sample assessment material for first teaching September 2017

(Time: 1 hour 30 minutes) | Paper Reference **1ST0/1H**

Statistics

Paper 1
Higher Tier

You must have:
Ruler graduated in centimetres and millimetres, protractor, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.

Information

- The total mark for this paper is 80.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

S54191A

1/1/1/1/1/

S 5 4 1 9 1 A 0 1 2 4

Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n}\sum(x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Rates of change (e.g. Birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

Answer ALL the questions.

Write your answers in the spaces provided.

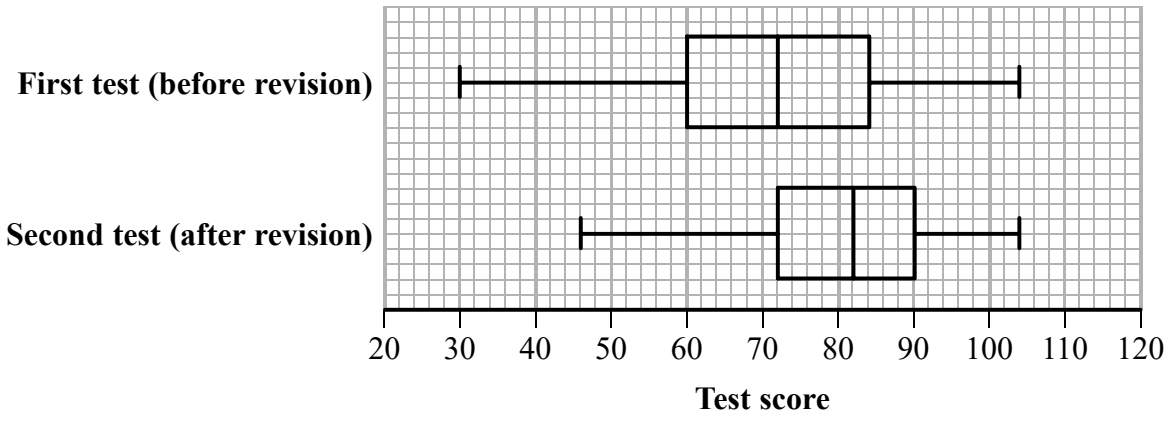
You must write down all the stages in your working.

1 A teacher gave her students two mathematics tests on the same topics.

The first test was given before revision of the topics and the second test was given after revision of the topics.

The two tests had the same maximum mark and the same difficulty.

The box plots give information about the students' scores in these tests.



Describe what effect the revision had on the test scores of these students.

You must explain how you reach your conclusions.

(Total for Question 1 is 4 marks)

2 Kerry is investigating whether there is a difference in the lengths of the text messages sent by boys and sent by girls at her school.

She writes the following hypothesis for the investigation.

"The length of text messages sent by girls is greater than the length of text messages sent by boys".

Kerry decides to use a census of the 800 students in her school.
She is going to ask each student to record the number of characters in their last text message.

Kerry then collects this information from each student through an online database.

Some of the database is shown below.

	Gender	**Length of text message**
1	male	73
2	F	68
3	girl	thirty five
4	boy	114,
5	boy	85
6	girl	
7	M	56
8	48	boy
9	girl	5
10	G	75
11	B	41
12	girl	28

(a) Give **two** reasons why Kerry must clean the data before processing it.

Reason 1: .....

.....

Reason 2: .....

.....

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

(b) Discuss how Kerry's data collection plan could affect the reliability of her conclusions.

.....

.....

.....

.....

.....

.....

.....

.....

(2)

(Total for Question 2 is 4 marks)

3 The table shows information about houses for sale in Oxford.

Number of bedrooms	1	2	3	4	5 or more	Total
Number of houses for sale	140	300	420	240	100	1200

(Source: *adapted from rightmove.co.uk*)

An estate agent says the mode of the number of bedrooms for these houses is 3

(a) Explain how she knows this.

.....

.....

(1)

The estate agent wants to investigate the prices of these houses.

She takes a stratified sample of 60 houses according to the number of bedrooms.

(b) Work out the number of houses in her sample for each number of bedrooms.

Number of bedrooms	1	2	3	4	5 or more
Number of houses in the sample					

(3)

(c) Describe how to select the 60 houses in the sample.

.....

.....

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 3 is 7 marks)

4 Morgan is investigating the 180 Year 11 students in his school.

He collected information from the 30 students in his class.

Part of the spreadsheet he used to collect the information is shown below.

Student number	Name	Left-handed (1 = yes, 0 = no)	Height (cm)
1	Jason	0	169
2	Rami	1	165
29	Youen	0	164
30	Elena	0	162
Total		4	5031

Morgan uses these results to find estimates for all Year 11 students.

(a) Find his estimate for the number of left-handed students in Year 11

.....

(2)

(b) Explain how Morgan can use the information in the spreadsheet to estimate the mean height of all the students in Year 1

.....

.....

.....

(1)

(Total for Question 4 is 3 marks)

5 The table shows information about the retail price index (RPI) and the price of a second-class stamp (in pence) in the United Kingdom for January 1996, January 2006 and January 2016

	Jan 1996	**Jan 2006**	**Jan 2016**
Retail price index (RPI)	100	129	172
Price of second-class stamp (pence)	20	23	54

(Sources: *ons.gov.uk and royalmail.com*)

Describe how the increase in the price of a second-class stamp compares with the RPI over the ten years to January 2006 **and** over the twenty years to January 2016

(Total for Question 5 is 5 marks)

6 Rahul, Lisa and Paul are investigating how much the workers in a company earn.

They have been told that in a week the workers earn £260 or £370 or £510

Last week
20% of the workers earned £260
35% of the workers earned £370
45% of the workers earned £510

Rahul, Lisa and Paul want to work out the average earnings for these workers last week.

Rahul thinks that they should find the mean of £260, £370 and £510
Lisa thinks that they should find the median of £260, £370 and £510
Paul thinks that they should find the weighted mean of the earnings.

(a) Which one of these three averages should they use?
Give a reason for your answer.

.....

.....

.....

.....

(2)

Rahul works out that the mean of the earnings is £380
Lisa finds that the median of the earnings is £370

(b) Work out the weighted mean of the earnings for Paul.

£.....

(2)

(Total for Question 6 is 4 marks)

7 A fitness company is planning to open a new gym in a town.

The company wants to collect information about the health of people in the town.

The company plans to interview people in the town centre and ask them questions from a questionnaire.

Here are two of the questions from the questionnaire.

A How old are you?

B What is your weight?

Discuss whether using these questions in an interview is an appropriate way to collect this information.

(Total for Question 7 is 3 marks)

8 The table gives information about the monthly average price per litre, in pence, of diesel over a period of five months.

The table also gives some of the chain base index numbers, correct to one decimal place, for this information.

	May	June	July	August	September
Monthly average price (p)	109.1	111.8	112.7	111.2	113.1
Chain base index number		102.5	100.8		

(Source: *theaa.com*)

(a) Calculate the chain base index numbers for August and September and write them in the table.
Give each value correct to one decimal place.

(2)

(b) (i) Calculate the geometric mean of the four chain base index numbers.
You must show your working.
Give your answer correct to one decimal place.

.....

(2)

(ii) Interpret your answer.

.....

.....

.....

(2)

(Total for Question 8 is 6 marks)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

9 Tania wants to estimate the number of snails in a pond.
She takes a sample of 10 snails from the pond.
She marks each snail with a waterproof dye and then puts the snail back in the pond.

Two weeks later, Tania takes another sample of 10 snails from the pond.
She finds that only one of the snails is marked with the dye.

Tania says,

"I estimate there are 100 snails in the pond."

How reliable is Tania's estimate?
Give reasons for your answer.
You are not required to check Tania's calculation.

.....

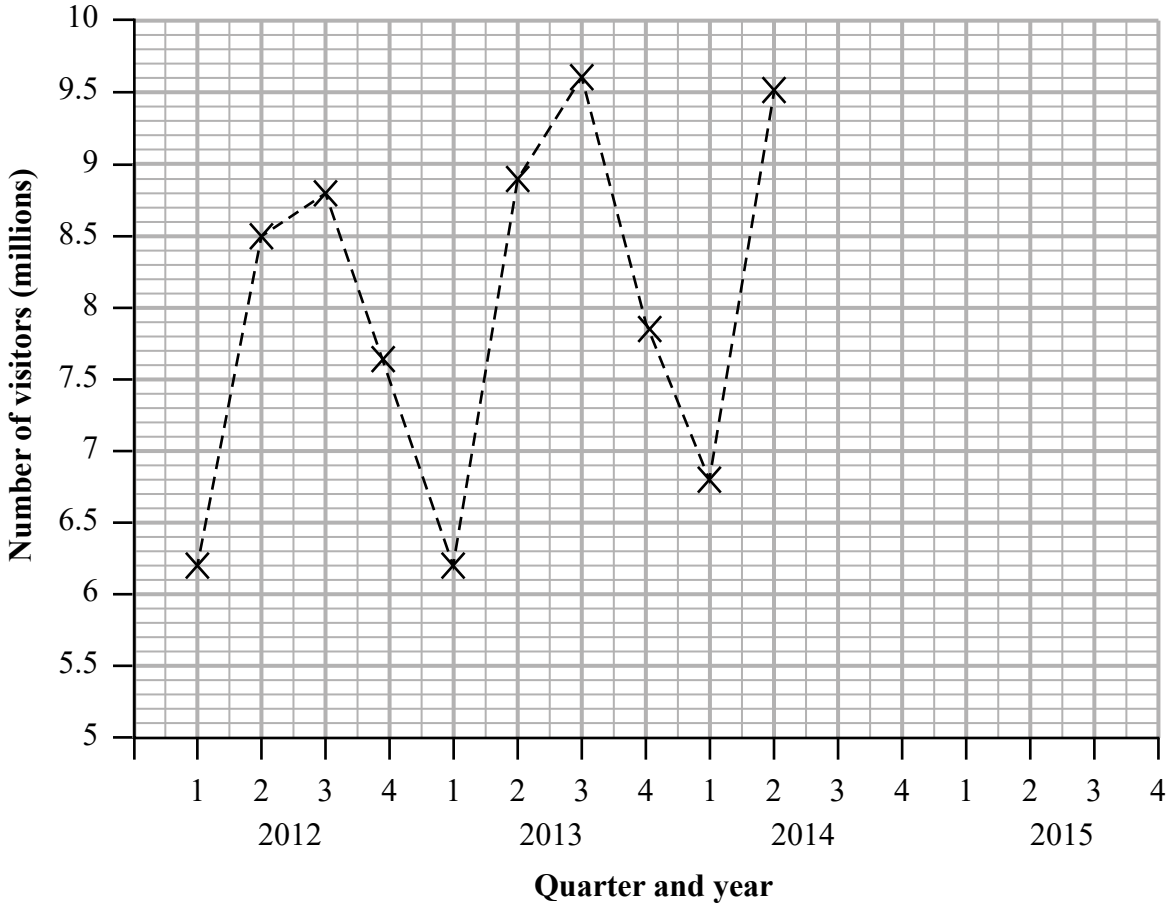
.....

.....

.....

(Total for Question 9 is 2 marks)

10 The time series graph shows some information about the numbers of visitors to the UK from 2012 to 2014



(Source: *visitbritain.org*)

(a) Identify **and** interpret in context **one** example of the seasonality shown by the time series graph.

.....

.....

.....

.....

.....

(2)

Tony calculates the 4-point moving averages for the information shown in the time series graph.

Here are his results.

7.77 7.79 7.90 8.11 8.17 8.29 8.45

(b) Plot these moving averages on the time series graph and hence draw a trend line for the number of visitors to the UK from 2012 to 2014

(3)

(c) Describe the trend.

.....

(1)

Tony says,

"Using the time series graph I estimate that there were approximately 7.35 million visitors to the UK in Quarter 1 of 2015".

(d) Use the average seasonal effect for Quarter 1 to show that Tony's estimate is reasonable.

(3)

Tony calculated 4-point moving averages for the information shown in the time series graph.

(e) Explain why this is appropriate.

.....

.....

(1)

(Total for Question 10 is 10 marks)

11 In a television talent contest, 9 acts were given the following ranks by the judges and by a public telephone vote.

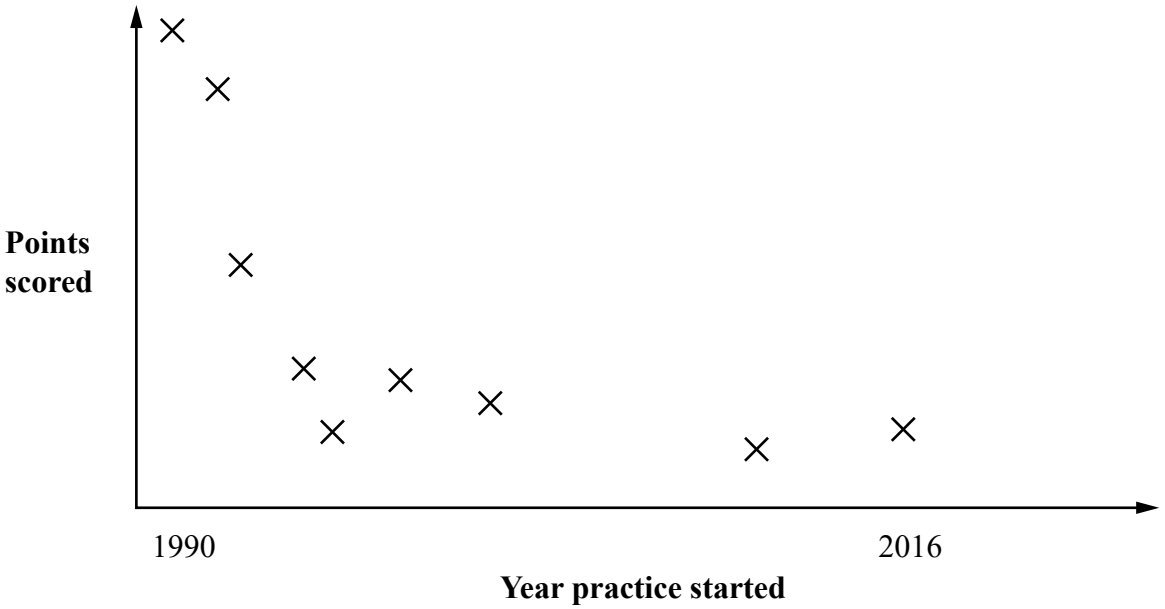
Act	Judges' rank	Public vote rank		
A	9	9		
B	3	2		
C	1	3		
D	4	5		
E	5	8		
F	8	6		
G	6	4		
H	2	1		
I	7	7		

(a) Use calculations to determine how much agreement there is between the judges and the public.

(5)

Gurdeep was investigating the relationship between the number of points scored and the year practice started for each of 9 acts.

This is the scatter diagram he obtained using statistical software.



The statistical software also calculated two correlation coefficients.

Spearman's rank correlation coefficient
Pearson's product moment correlation coefficient

(b) (i) Circle **one** value in each row below to show the most likely **pair** of correlation coefficients for this data.

Spearman's rank correlation coefficient:	−0.9	−0.7	0	0.7	0.9
Pearson's product moment correlation coefficient:	−0.9	−0.7	0	0.7	0.9

(2)

(ii) Explain your choice of answers in part (i).

.....

.....

.....

(1)

(Total for Question 11 is 8 marks)

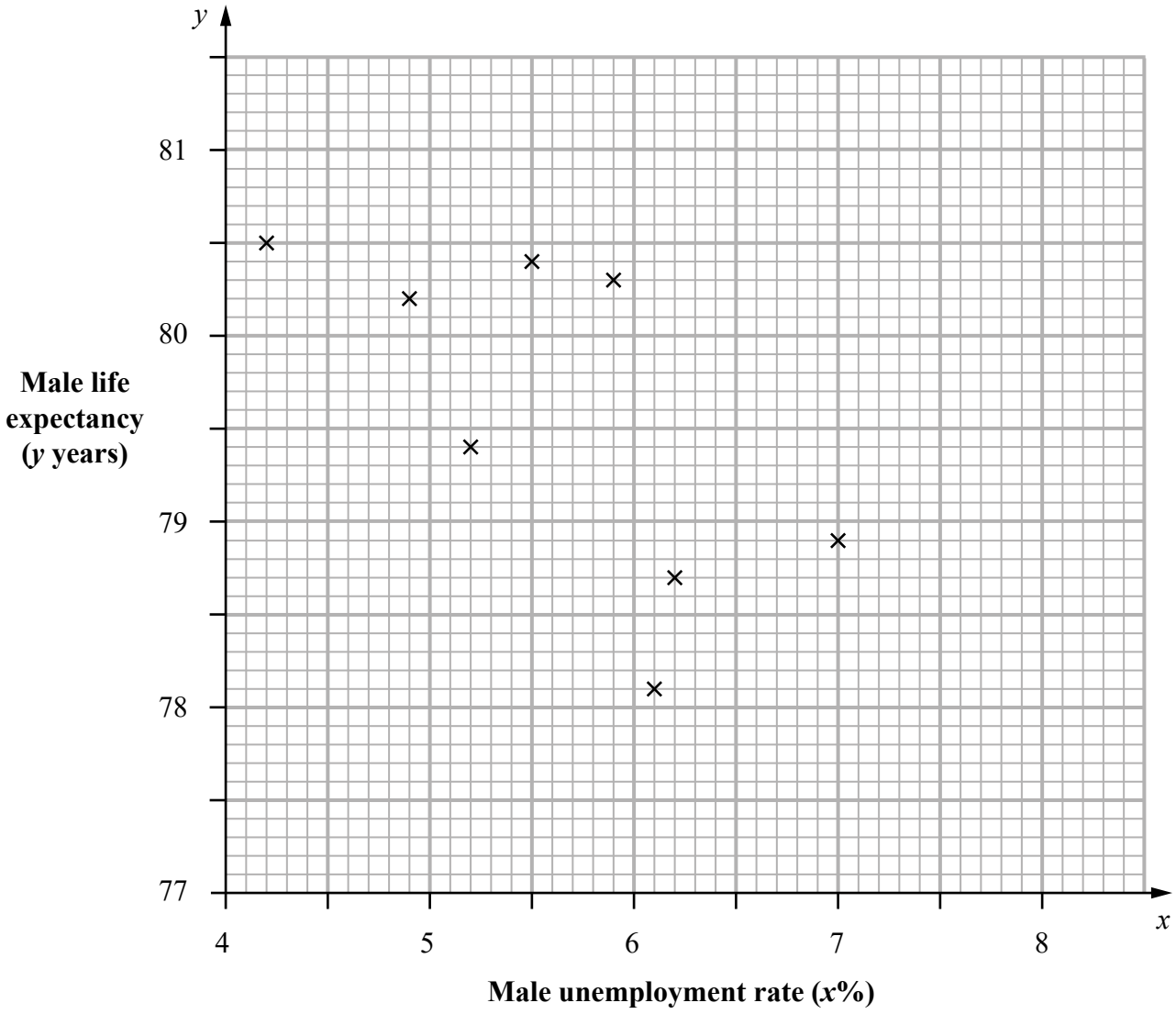
12 Rajesh is investigating to see if there is an association between the male unemployment rate and male life expectancy in the regions of England for 2014

His hypothesis is

"Where the male unemployment rate is high, the male life expectancy is low."

Rajesh finds the male unemployment rate (x%) and the male life expectancy (y years) for each of eight of the nine regions of England for 2014

This information is shown on the scatter diagram below.



(Source: *ons.gov.uk*)

(a) Explain, giving a statistical reason, whether or not this scatter diagram supports Rajesh's hypothesis.

.....

.....

.....

(2)

The equation of the regression line for the data in the scatter diagram is $y = 83.5 - 0.7x$

(b) Draw this line on the scatter diagram.

(2)

The male unemployment rates for these eight regions have a mean of 5.6%

(c) Use this information to find the mean of the male life expectancies for these regions.

..... years

(1)

(d) Interpret the value of the gradient of this regression line.

.....

.....

.....

.....

(2)

Rajesh now finds that the male unemployment rate for the missing ninth region is 8.0%

(e) Give **two** reasons why Rajesh should not use the regression line to predict the male life expectancy in this region.

.....

.....

.....

.....

.....

(2)

(Total for Question 12 is 9 marks)

13 A film company employs Gary to investigate the film-watching habits of people living in the UK.

Gary is going to use a questionnaire.

Here is Question 1 on Gary's questionnaire.

> **Question 1**
>
> Spin a fair coin.
>
> If you get **Heads**, tick box A.
> If you get **Tails**, answer this question.
>
> **Have you downloaded a film illegally during the last month?**
>
> If **yes**, tick box A. If **no**, tick box B.
>
> A ❑ B ❑

The method used to decide whether or not to answer a question by spinning a coin is called the random response technique.

(a) Explain why this method is used.

.....

.....

.....

.....

(1)

Gary sends the questionnaire to a sample of people living in a town.

He uses a telephone directory as the sample frame.

For Question 1

743 people ticked box A

679 people ticked box B

(b) Calculate an estimate of the proportion of the people in the sample who had downloaded a film illegally during the last month.

.....

(3)

Gary is going to write a report on the outcome of Question 1.

He is going to use the answer to part (b) as an estimate of the proportion of all the people living in the UK who had downloaded a film illegally during the last month.

(c) Is it appropriate for Gary to use the answer to part (b)?
Give **two** reasons for your answer.

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 13 is 7 marks)

14 The table gives information about blood donations in the UK.

It shows the blood groups O, A, B and AB and the number of donations for each blood group expressed as a percentage of the total number of all donations.

Blood group	**O**	**A**	**B**	**AB**
Percentage of all donations	48%	38%	10%	4%

(Source: *blood.co.uk*)

6 people attend a clinic on Monday to donate blood.

(a) (i) Name the probability distribution that can be used to model the number of people from these 6 people who will have blood group AB.

.....

(ii) Write down one condition needed so that this distribution is a suitable model.

.....

.....

(2)

(b) Work out the probability that exactly one of these 6 people will have blood group AB. Give your answer correct to 3 decimal places.

.....

(3)

On Tuesday n people attend the clinic to donate blood.

The probability that at least one of these n people will have blood group AB is greater than 0.5

(c) What can you conclude, if anything, about the value of n?
You must show your working.

.....

.....

.....

(3)

(Total for Question 14 is 8 marks)

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Paper 1 Higher tier mark scheme

Question number	Answer	Additional guidance	Mark
1	B1 for e.g. median greater after revision B1 for e.g. IQR/range smaller after revision B1 for e.g. negative skew after revision, symmetrical before revision B1 for e.g. • They did better after revision • They were more consistent after revision	B1 for a correct statistical statement comparing the medians B1 for a correct comparison of the IQRs or ranges B1 for a correct comparison of the skews B1 for a correct contextual interpretation comparing medians or IQR/ranges	**(4)**
2(a)	B1B1 for two correct reasons • e.g. data given in different formats • e.g. remove extraneous symbols • e.g. remove anomalies/outliers • e.g. data given in wrong order	B1 for each correct reason for the need to clean data on the database prior to processing it	**(2)**
2(b)	B1B1 for two correct aspects • e.g. large sample size increases reliability • e.g. issues due to how the data collection is carried out may decrease reliability (recorded by students and not by Kerry/students may type in information wrong) • e.g. for an example of factors that might not be consistent in the data collection which may decrease reliability (last message might not represent all text messages sent) • e.g. non-response decreases reliability	B1B1 for two correct comparison assessing the reliability of the conclusions drawn	**(2)**

<table>
<tr><th>Question number</th><th>Answer</th><th>Additional guidance</th><th>Mark</th></tr>
<tr><td>3(a)</td><td>B1 3 bedrooms has the highest frequency (for any individual number of bedrooms)</td><td>Allow equivalent statistical reasoning based on the table indicating why 3 is the mode.</td><td>(1)</td></tr>
<tr><td>3(b)</td><td>M1 $\frac{140}{1200}\times 60$ oe

A1A1<table><tr><td>Bedrooms</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5+</td></tr><tr><td>Houses in sample</td><td>7</td><td>15</td><td>21</td><td>12</td><td>5</td></tr></table></td><td>Accept a correct equivalent calculation shown for any one class
M1 implied by one correct answer OR an indication they need 1 in 20

1st A1 for any one value correct
2nd A1 for all correct</td><td>(3)</td></tr>
<tr><td>3(c)</td><td>B1 Use a sampling frame for each strata

B1 Select houses randomly or generate random numbers

B1 For an aspect of detail</td><td>Each category/strata to be considered separately

Samples have to be random

e.g. How the random numbers are obtained and used</td><td>(3)</td></tr>
<tr><td>4(a)</td><td>M1 $\frac{4}{30}\times 180$
A1 $= 24$</td><td>Accept any equivalent calculation</td><td>(2)</td></tr>
<tr><td>4(b)</td><td>B1 e.g. ‘use the mean of the sample of an estimate’</td><td>B1 for a correct explanation showing understanding that the mean of the sample represents the mean of the population</td><td>(1)</td></tr>
</table>

Question number	Answer	Additional guidance	Mark
5	M1A1A1 for appropriate calculations to compare 1996 with 2006 or 1996 with 2016 e.g. **Comparing index numbers** M1 $\frac{23}{20}\times 100$ or $\frac{54}{20}\times 100$ A1 115 A1 270 OR **Comparing prices** M1 $20\times\frac{129}{100}$ or $20\times\frac{172}{100}$ A1 25.8 A1 34.4 B1 for e.g. from 1996 to 2006 the change/increase in price was less than the RPI B1 for e.g. from 1996 to 2016 the change/increase in price was higher than the RPI	M1 for a calculation that could be used to compare 1996 with 2006 or 2016 A1 for one correct value A1 for second correct value that allows a comparison to be made B1 for a correct contextual statement for 2006 or 2016 B1 for a correct contextual statement for 2006 and 2016	**(5)**

Question number	Answer	Additional guidance	Mark
6(a)	B2 for weighted mean (or Paul) AND reference to there being different proportions/percentages of people earning each weekly amount OR B1 for weighted mean (or Paul) with attempt at reason	B2 for a complete assessment of the appropriate choice with reason OR B1 for an incomplete assessment of the appropriate choice	**(2)**
6(b)	M1 for $0.2\times260 + 0.35\times370 + 0.45\times510$ A1 for 411	M1 for a correct method to find the weighted mean A1 for the correct answer	**(2)**
7	B1 for the questions are personal/people may be embarrassed AND B2 for a correct comment assessing the appropriateness of the interview and a reason e.g. • appropriate and e.g. an interview will have a good response rate (or higher response rate than e.g. postal survey) • appropriate and e.g. the interviewer could be trained to put people at their ease when answering the personal questions • not appropriate and e.g. people may not feel comfortable talking about their health/fitness with the interviewer OR (if B2 not earned) B1 for a correct commenting relating to the appropriateness of the interview without a decision	B1 for assessing appropriateness of questions B2 for a correct comment assessing the appropriateness of the interview and a reason OR (if B2 not earned) B1 for an incomplete assessment of the appropriateness of the interview	**(3)**

Question number	Answer	Additional guidance	Mark
8(a)	M1 for either 111.2/112.7×100 (=98.7) OR 113.1/111.2×100 (=101.7) A1 for 98.7 and 101.7	M1 for correct calculation of chain base index number. May be implied by one correct answer A1 both correct	**(2)**
8(b)(i)	M1ft for $\sqrt[4]{}$(102.5×100.8×'98.7'×'101.7') A1ft for100.9	M1ft for correct calculation of the geometric mean of the four chain base index numbers A1ft correct answer ft their answers in (a)	**(2)**
8(b)(ii)	B1ft for (average) rate of 'increase' ... B1ft ... is '0.9'% per month	B1ft for correct contextual interpretation as rate of increasing B1ft for complete correct contextual interpretation of their value for geometric mean	**(2)**
9	B2 for not reliable with a correct reason e.g. samples too small or time interval between samples too long (as population may have changed between samples)	B2 for a correct comment assessing the appropriateness of the conclusion OR if B2 not earned B1 for an incomplete attempt to assess the appropriateness of the conclusion	**(2)**

Question number	Answer	Additional guidance	Mark
10(a)	B1 for a correct statement identifying any seasonality, e.g. the greatest/least values are in quarter 3/ quarter 1 B1 for a correct interpretation in context for the identified seasonality, e.g. more/less overseas visitors in summer/winter	B2 for a correct statement identifying and interpreting an example of seasonality (B1 for one of these features)	**(2)**
10(b)	M1 Correct horizontal plotting, first between Q2/Q3 of 2012 A1 All points correct B1 Straight trend line through moving averages within tolerance	M1 for recognising correct horizontal position for plotting moving averages (at least five correct) A1 for accurately plotting all moving averages (Allow within one small square tolerance) B1 Their line should extend horizontally at least from 2012 Q3 to 2013 Q4 and be vertically within one square of 7.8 at 2012 Q3 and one square of 8.4 at 2013 Q4	**(3)**
10(c)	B1 Upward/rising trend	Accept equivalent wording demonstrating statistical reasoning	**(1)**
10(d)	B1ft trend line value 8.9 (±0.1) M1 for $8.9 - \frac{(7.55-6.25)+(8-6.3)+(8.45-6.8)}{3}$ A1 for 7.35	B1ft for value which follows from their line at 2015 Q1 M1 for complete method to find seasonal variation for their trend line A1 for answer in range [7.2−7.5]	**(3)**
10(e)	B1 for e.g. the pattern in the data repeats after four quarters	B1 for a correct statement assessing the appropriateness of using 4-point moving averages	**(1)**

Question number	Answer	Additional guidance	Mark
11(a)	M1 difference in ranks: 0, 1, −2, −1, −3, 2, 2, 1, 0 M1 $(r_s =)\ 1-\frac{6\times 24}{9\times(9^2-1)}$ A1 0.8 B1ft Positive (rank) correlation B1ft Judges were in agreement with the public	M1 for difference in ranks (condone one slip and allow ±). Can be implied by $\Sigma d^2 = 24$ M1 for demonstrating correct use of Spearman's formula A1 cao B1ft for statistical interpretation of their '0.8' B1ft for correct contextual conclusion from their '0.8'	**(5)**
11(b)(i) **11(b)(ii)**	B1 **both negative** values identified for the graph B1 Spearman's = −0.9 AND pmcc = −0.7 B1 pmcc is less strong correlation as it measures closeness to a **linear** model	B1 for recognising graph will give negative correlation B1 cao, for recognising pmcc calculation will be closer to zero when correlation is non-linear B1 for equivalent statistical reasoning that pmcc will be closer to 0 as graph does not suggest a straight line.	**(3)**

Question number	Answer	Additional guidance	Mark
12(a)	B1 Graph supports the hypothesis + reason B1 Scatter shows **negative correlation**	B1 for conclusion supported by sensible reason B1 for statistical reasoning using words in bold	**(2)**
12(b)	M1 e.g. $83.5 - 0.7 \times 4 = 80.7$ or $83.5 - 0.7 \times 8 = 77.9$ or sensible straight line with correct gradient A1 correct straight line within tolerance	M1 for one pair of coordinates correctly identified, or for a sensible attempt at straight line with correct gradient. A1 Their line should extend horizontally at least from 4.2 to 7 and (if extended) be vertically within one square of 80.7 at $x = 4$ and one square of 77.9 at $x = 8$	**(2)**
12(c)	B1 79.6 from graph	B1 for answer between 79.5 and 79.7 (may use equation OR vertical line drawn from $x = 5.6$)	**(1)**
12(d)	B1 e.g. change in life expectancy as unemployment increases. B1 **0.7 years fall** in life expectancy (per 1% increase in unemployment)	B1 for recognising in context that gradient indicates a **rate**. Accept equivalent wording. No figures needed for 1st B1 B1 for interpreting in context the value. Mark may be gained for correct equivalent figures used within their comment. (e.g. 1.4% more unemployment results in 1 year reduction in life expectancy)	**(2)**
12(e)	B1 B1 for any two comments from • Involves extrapolation/8% is outside of range • It is a for a different region so may be affected by other factors • Correlation does not look very strong	Accept equivalent statistical reasoning. B1 for each correct point but allow each bullet point once only.	**(2)**

Question number	Answer	Additional guidance	Mark
13(a)	B1 for e.g. this is a sensitive question or people may not want to answer it otherwise	B1 for a correct response referring to the sensitivity of the question	**(1)**
13(b)	e.g. M1 for 0.5 × (743 + 679) (= 711) M1 for $\frac{743-"711"}{"711"}$ A1 for 0.045(007…)	M1 for method to estimate the number of people who answered yes because they got Heads M1 for method to estimate the proportion of people who have downloaded illegally A1 for a correct proportion, e.g. 0.045 or 4.5% or better	**(3)**
13(c)	B1 for not appropriate B1 for a correct reason, e.g. the town may not be representative of the UK B1 for a different correct reason, e.g. the telephone directory may not include everyone in the town	B1 B1 B1 for assessing the appropriateness of the statistical methodology with correct reasons	**(3)**

Question number	Answer	Additional guidance	Mark
14(a)(i)	B1 for binomial (distribution)	B1 for correctly identifying the binomial distribution	**(2)**
14(a)(ii)	B1 for a correct property, e.g. independent trials or the 6 people are not related OR only two possible outcomes, i.e. success/failure OR fixed number of trials, i.e. $n = 6$, etc	B1 for a correct property needed for the binomial distribution	
14(b)	M1 for $1 - 0.04$ ($= 0.96$) M1 for $6 \times 0.04 \times 0.96^5$ A1 for 0.196	M1 for finding the probability of not AO M1 for a complete method A1 for answer in range 0.19 – 0.20	**(3)**
14(c)	M1 for $1 - 0.96n$, for any $n \geq 2$ A1 for ($n = 17$) B1 for a correct conclusion, e.g. the smallest value of n is 17	M1 for a correct method to work out at least one probability A1 for the correct answer (17) associated with the correct probability (0.5004…) B1 for a correct interpretation of the answer (17) in the context of the problem	**(3)**

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

Centre Number | | | | | |

Candidate Number | | | | |

Sample assessment material for first teaching September 2017

(Time: 1 hour 30 minutes) | Paper Reference **1ST0/2H**

Statistics

Paper 2
Higher Tier

You must have:
Ruler graduated in centimetres and millimetres, protractor, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.

Information

- The total mark for this paper is 80.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

S54193A

1/1/1/1/1/1/

S 5 4 1 9 3 A 0 1 2 4

Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n}\sum(x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Rates of change (e.g. Birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

Answer ALL the questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The table gives information about the public spending (£ billion) by Central Government from 1995 to 2000

Public spending (£ billion)	**1995**	**2000**	**2005**	**2010**	**2015**
Pensions	41.4	65.7	86.4	116.4	149.8
Healthcare	40.0	49.2	82.6	116.8	131.3
Education	11.6	14.9	23.0	33.4	39.7
Defence	25.6	27.8	33.4	42.5	45.2
Welfare	40.9	35.6	42.4	60.9	58.3
Protection	6.1	7.3	14.0	16.9	15.3
Transport	7.3	4.7	8.6	13.3	12.2
General government	5.0	6.6	9.4	10.5	9.4
Other spending	12.4	17.1	36.0	57.8	77.1
Interest and balancing	26.6	25.6	23.8	30.5	45.2
Total public spending	216.7	254.4	359.6	498.9	583.5

(Source: *ukpublicspending.co.uk*)

(a) Write down the public spending category in 2000 with the lowest public spending.

.....

(1)

(b) Describe the trend in the total public spending from 1995 to 2015

.....

(1)

The public spending column for 1995 adds up to 216.9 but the total is given as 216.7

(c) Give a possible reason why.

.....

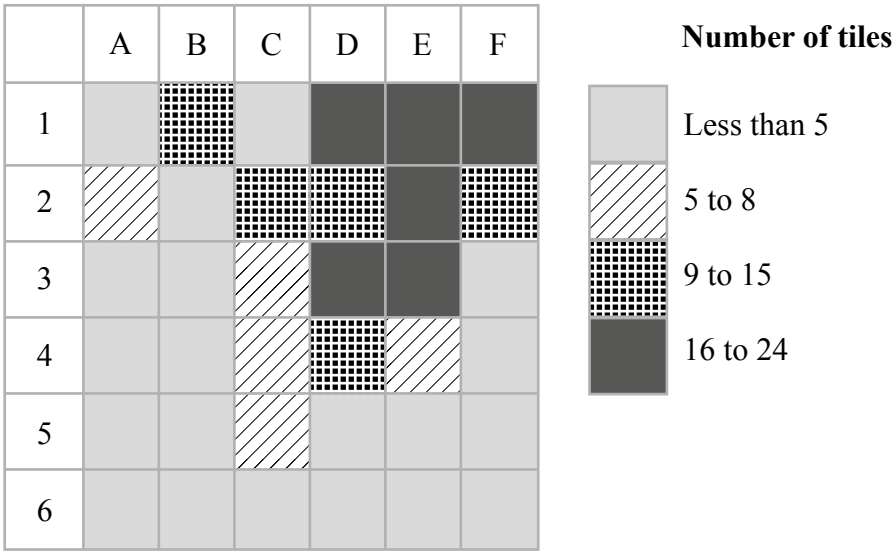
(1)

(Total for Question 1 is 3 marks)

2 Archaeologists divided a field into 36 squares of equal size.

The number of Roman roof tiles found in each square was recorded.

The choropleth map below was drawn using this information.



Use the choropleth map to describe where in the field the greatest number of roof tiles was found.

Give a statistical reason for your answer.

.....

.....

.....

.....

.....

(Total for Question 2 is 2 marks)

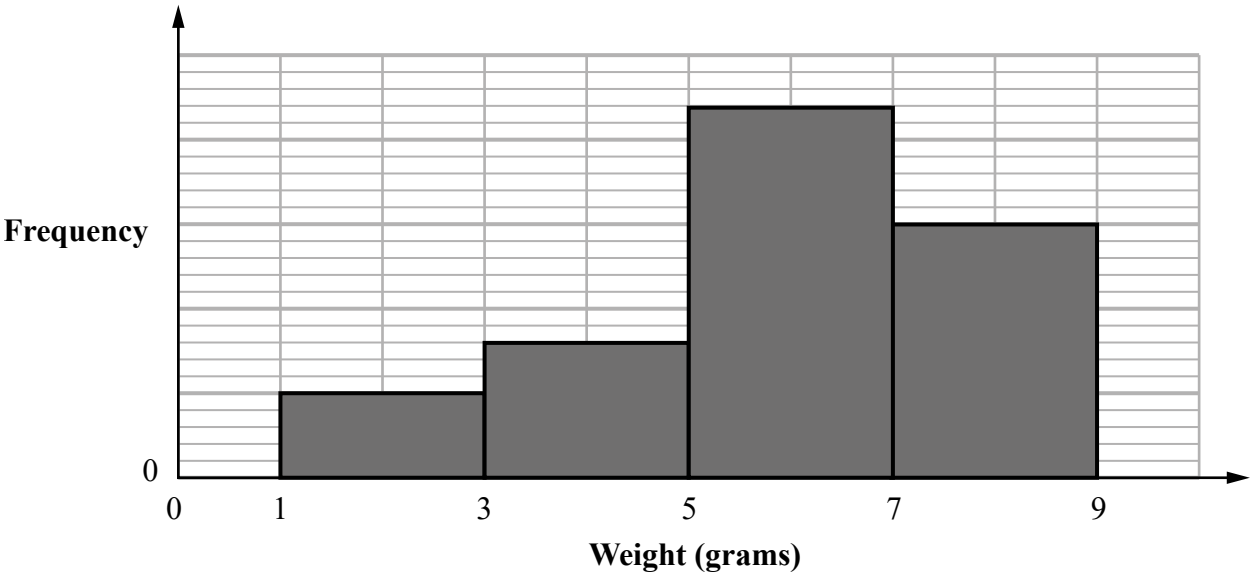
3 Tomoyo found the weight, in grams, of each of 100 cherries.

(a) Circle the **two** words from the list that best describe the data Tomoyo found.

quantitative **qualitative** **discrete** **continuous** **bivariate** **ordinal** **categorical**

(2)

Tomoyo grouped the weights and she then drew this diagram for her results.



The incomplete frequency table shows some information about her results.

Weight (w grams)	Frequency
$1 \leqslant w < 3$	10
$3 \leqslant w < 5$	
$5 \leqslant w < 7$	
$7 \leqslant w < 9$	

(b) (i) Complete the frequency column in the table.

(ii) Calculate an estimate of the mean weight of the 100 cherries.

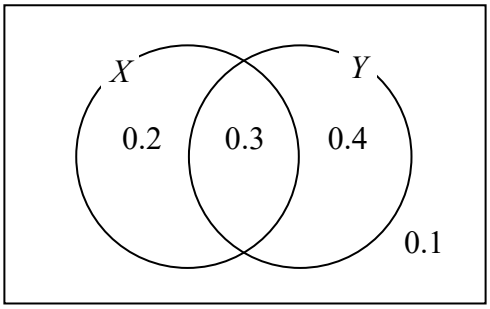
..... g

(5)

(Total for Question 3 is 7 marks)

4 X and Y are two events.

The Venn diagram shows information about the probabilities of events related to X and Y happening.



(a) Find

(i) the probability of event Y happening.

.....

(ii) P(X and Y),

.....

(iii) P($Y \mid X$).

.....

(4)

Two different events A and B are independent.

P(A) = 0.8 and P(B) = 0.5

(b) Find P(A and B).

.....

(2)

(Total for Question 4 is 6 marks)

5 Gary is going to investigate the amounts of time students spend watching TV.

He is going to write a plan for this investigation.

His hypothesis is

> "The amount of time that boys spend watching TV is greater than the amount of time that girls spend watching TV".

Write down three other things he should include in his plan.

Explain why each of these things is appropriate.

You must refer to more than one stage of the statistical enquiry cycle.

(Total for Question 5 is 6 marks)

6 Richard works in an animal rescue centre.

He wants to compare the weights of the male cats and the weights of the female cats.

The table shows information about the weights, in kg, of a sample of male cats and the weights, in kg, of a sample of female cats.

	Least	Lower quartile	Median	Upper quartile	Greatest
Male	3.0	3.5	3.7	4.2	5.4
Female	3.0	3.2	3.3	3.8	4.6

(a) Use the information in the table to compare the distribution of the weights of male cats with the distribution of the weights of female cats.
Interpret your comparisons.

.....

.....

.....

.....

.....

.....

(2)

The information for female cats is based on data collected from 47 cats at the centre.

(b) Work out the number of these female cats with a weight greater than or equal to 3.8 kg.

.....

(2)

(Total for Question 6 is 4 marks)

7 Lata is investigating whether there are relationships between the test scores in different school subjects.

Lata has collected the test scores in English and the test scores in Maths for 15 students.

She decides to plot the data on a scatter diagram.

(a) Explain whether or not this is a good choice of diagram for her investigation.
You should refer to the type of data in your answer.

.....

.....

.....

(2)

Lata worked out Pearson's product moment correlation coefficient for the English test scores and the Maths test scores for the 15 students.

She got a value of 0.65

Lata thinks that this means that if she improves her English test score then her Maths test score will improve.

(b) Is Lata right?
Explain your answer.

.....

.....

(1)

Lata also collects the Science test scores for the 15 students.
Pearson's product moment correlation coefficient for the Maths test scores and the Science test scores is 0.75

(c) Compare the two correlation coefficients 0.65 and 0.75
Interpret your answer in context.

.....

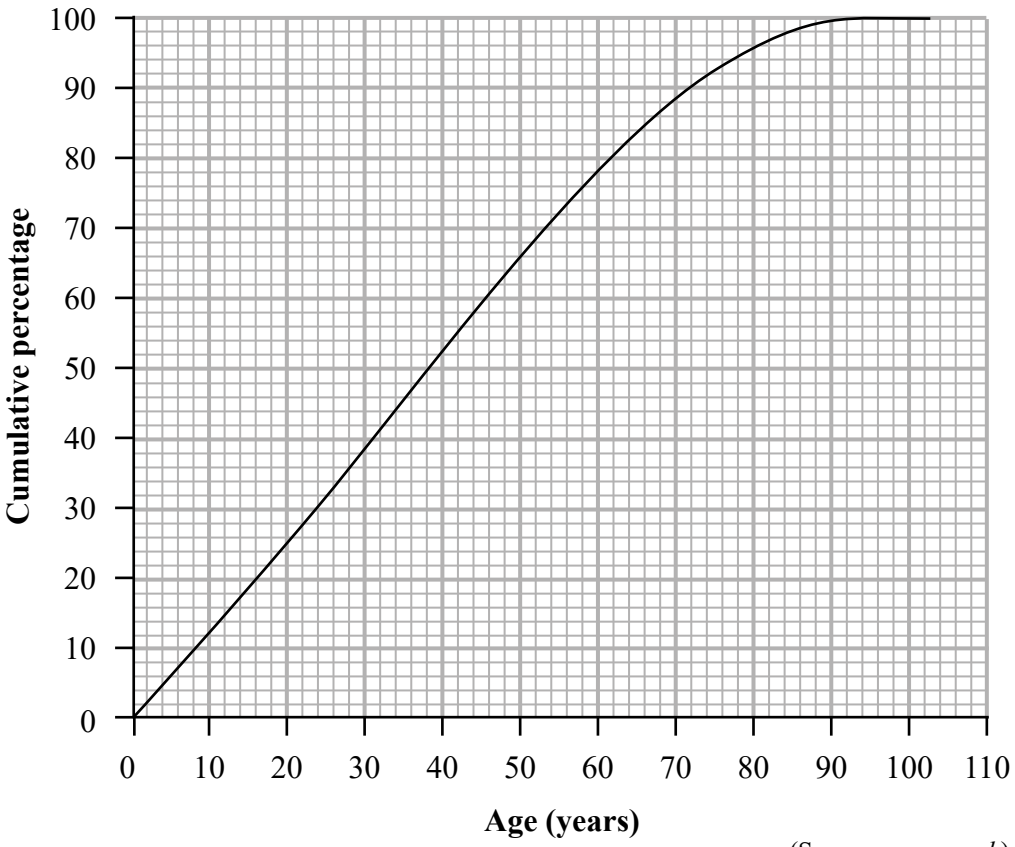
.....

.....

(1)

(Total for Question 7 is 4 marks)

8 The cumulative percentage graph shows information about the ages of the people living in the UK in 2014



(Source: *ons.gov.uk*)

(a) Find the 10th to 90th interpercentile range for this information.

..... years

(2)

Here are some statistics about the ages of the people living in Manchester in 2014

Median	29 years
10th to 90th interpercentile range	53 years

(Source: *ons.gov.uk*)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Compare the distribution of the ages of people living in Manchester in 2014 with that for the UK in 2014

.....

.....

.....

(3)

The table below gives more information about the ages of people living in Manchester in 2014

Mean	32 years
Standard deviation	19.3 years

(Source: *ons.gov.uk*)

(c) (i) Calculate the skew for the ages of people living in Manchester in 2014

.....

(ii) Interpret the skew in context.

.....

.....

.....

(3)

(Total for Question 8 is 8 marks)

9 The table shows the results in two events of the women's heptathlon for Jessica Ennis-Hill in the 2015 World Championships.

The mean and standard deviation for each of these two events for all the athletes who completed the heptathlon are also given.

	Jessica Ennis-Hill	**Mean**	**Standard deviation**
Long Jump (metres)	6.43	6.10	0.26
High Jump (metres)	1.86	1.79	0.066

(Source: *iaaf.org*)

(a) Use standardised scores for this information to compare Jessica Ennis-Hill's performance in the Long Jump with her performance in the High Jump.

Explain how you reach your conclusion.

(5)

The table below shows Jessica Ennis-Hill's result for the Javelin Throw.

It also shows information about the mean and the standard deviation for this event of all the athletes who completed the heptathlon.

	Jessica Ennis-Hill	**Mean**	**Standard deviation**
Javelin Throw (metres)	42.51	x	5.85

(Source: *iaaf.org*)

Jessica Ennis-Hill's standardised score for the Javelin Throw was -0.32

(b) Work out the value of x.

.....

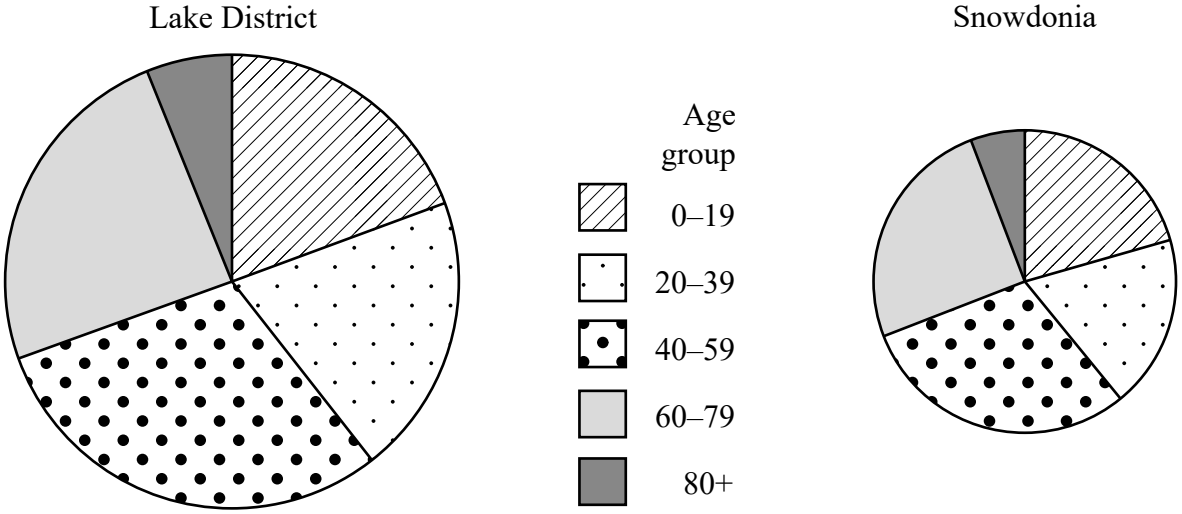
(2)

(Total for Question 9 is 7 marks)

10 The comparative pie charts give information about the numbers of people living in two national parks.

They also show the age groups of the people.

The numbers of people living in two national parks



(Source: *ons.gov.uk*)

(a) Compare the total number of people living in the Lake District with the total number of people living in Snowdonia.

Explain how you reach your conclusion.

.....

.....

.....

.....

.....

(2)

The angles at the centres of the circles for the 40–59 age group sectors are the same for both pie charts.

(b) Describe how the number of people aged 40–59 years in the Lake District compares with the number of people aged 40–59 years in Snowdonia.

Explain how you reach your conclusion.

.....

.....

.....

.....

.....

(2)

Gill is going to display these comparative pie charts in a report. The members of her target audience are not statisticians.

(c) What additional information could Gill include with the pie charts to help her target audience interpret the pie charts?

.....

.....

(1)

Gill wants to draw a third comparative pie chart for Exmoor.

The number of people living in the Lake District is 3.9 times the number of people living in Exmoor.

The radius of the pie chart for the Lake District is 3 cm.

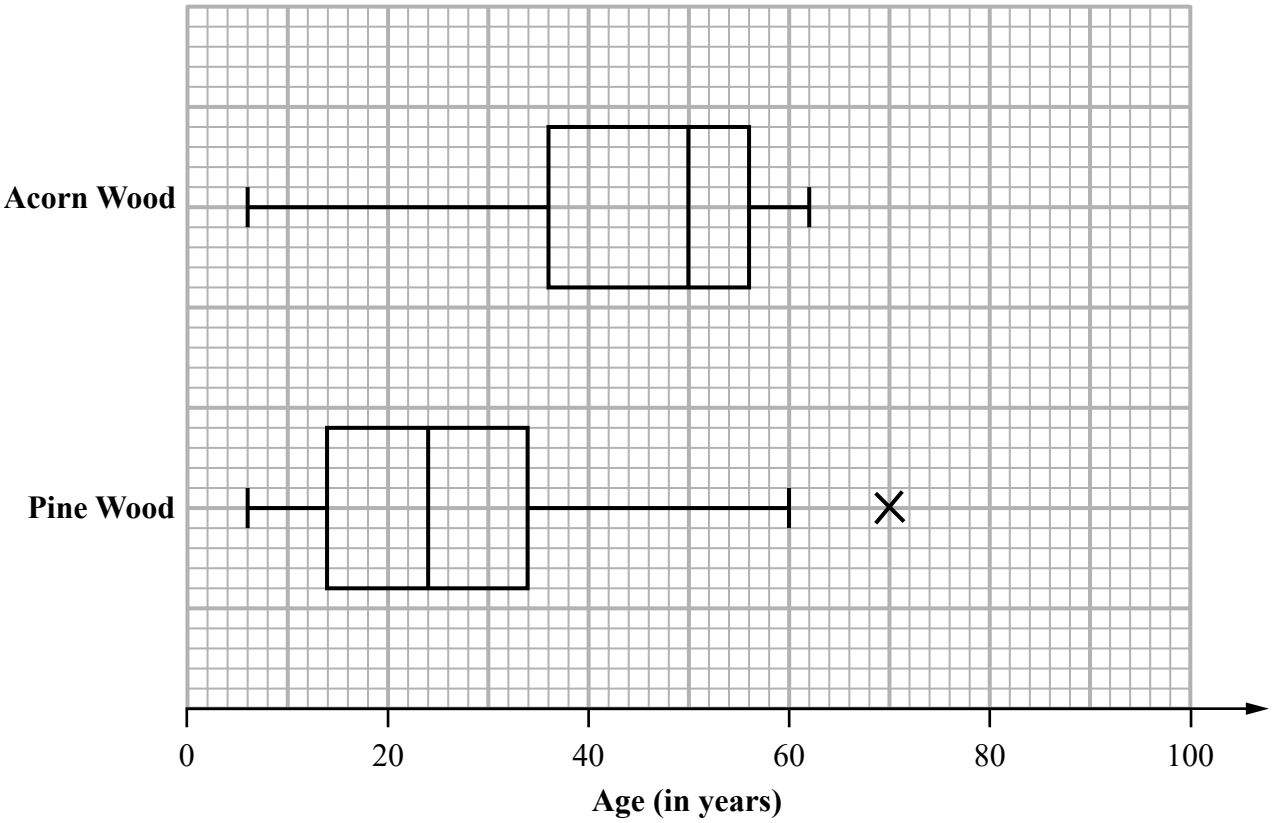
(d) Show that the radius of the pie chart for Exmoor is 1.52 cm, correct to 2 decimal places.

..... cm

(1)

(Total for Question 10 is 6 marks)

11 The box plots give information about the distributions of the ages of the trees in Acorn Wood and in Pine Wood.



(a) Justify, by calculation, that 70 is an outlier for Pine Wood.

(3)

Simon uses the information in the box plots to conclude that

"The average age of the trees in Acorn Wood is greater than the average age of the trees in Pine Wood.
Both distributions have the same spread.
Both distributions have a positive skew."

(b) Comment on Simon's conclusions with reference to his use of statistical words and the accuracy of his statements.
Give reasons for your answer.

(5)

(Total for Question 11 is 8 marks)

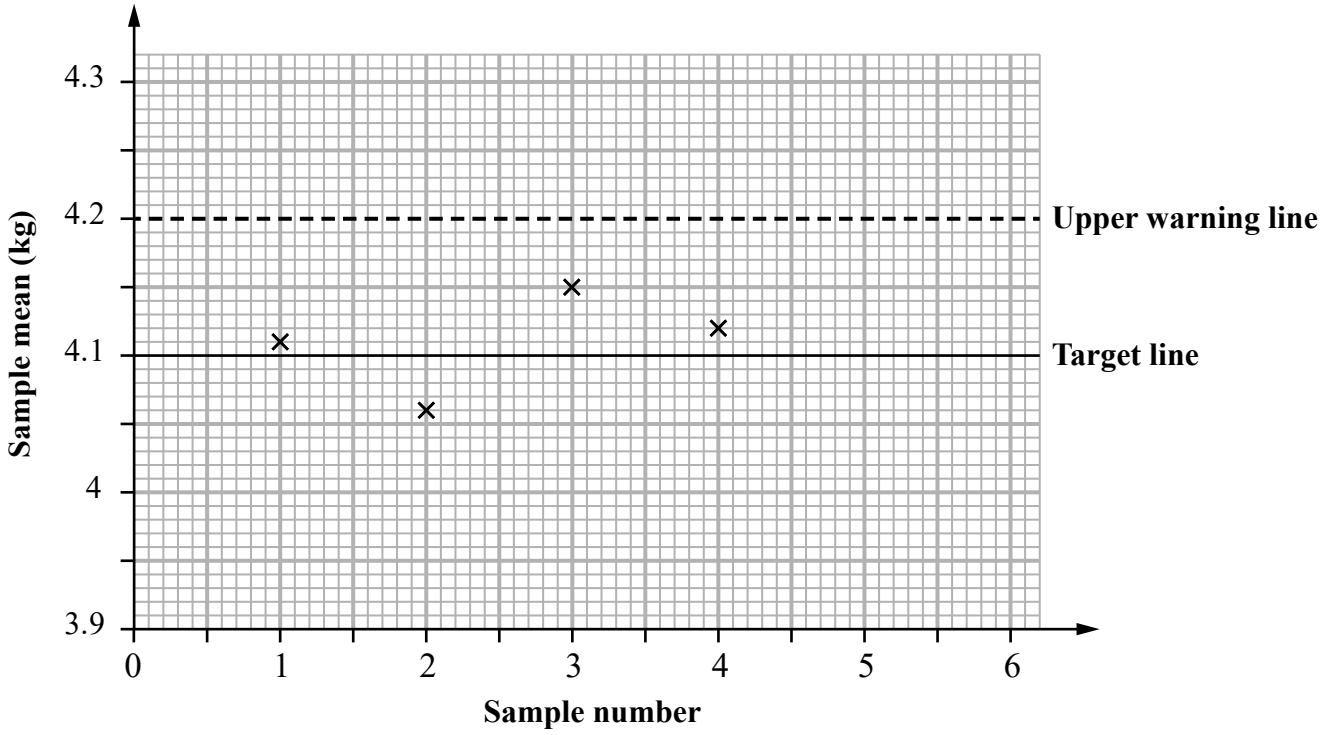
12 Glowbright Garden Products produce bags of charcoal for use on barbecues.

For quality control, a sample of 5 bags of charcoal is taken at regular intervals and the mean weight of the bags in the sample calculated.

The sample means should be normally distributed with a mean of 4.1 kg and a standard deviation of 0.05 kg.

A quality control chart for the sample means is drawn.

Four sample means have been plotted.



(a) Complete the control chart by drawing the upper action line, the lower action line **and** the lower warning line.
Label your lines.

(2)

The sample mean of sample 5 is 4.22 kg.

(b) Determine any actions that need to be taken.

.....

.....

.....

.....

.....

.....

.....

(2)

David says that the standard deviation of the weights of bags of charcoal produced by Glowbright Garden Products should be 0.05 kg.

(c) Is David correct?

.....

.....

.....

.....

(1)

(Total for Question 12 is 5 marks)

13 Peter thinks that the ages at inauguration of the presidents of the USA are normally distributed.

He collects information about the ages at inauguration, in years, of 43 presidents of the USA from the internet.

The grouped frequency table gives information about his results.

Age at inauguration (x years)	**Frequency (f)**
$42 \leqslant x < 47$	4
$47 \leqslant x < 52$	11
$52 \leqslant x < 57$	13
$57 \leqslant x < 62$	10
$62 \leqslant x < 72$	5

(Source: *robinsonlibrary.com*)

(a) Write down one disadvantage of collecting information from the internet.

.....

(1)

Peter uses a spreadsheet to calculate the following summary statistics for the information in the table.

$$\sum fx = 2361 \qquad \sum fx^2 = 131\ 334.5$$

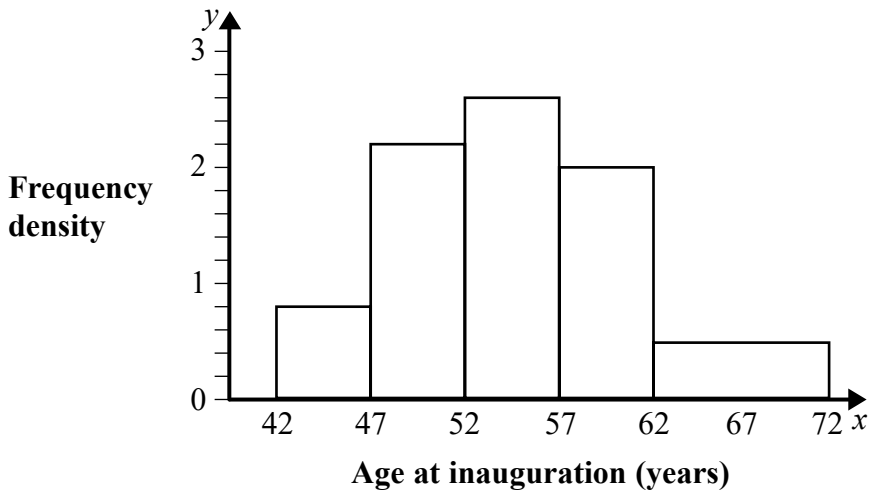
where the values of x are the class midpoints.

(b) Show that an estimate of the standard deviation of the ages at inauguration is 6.29

You may use Peter's summary statistics.

(2)

Peter now draws this histogram to show the information in the grouped frequency table.



(c) Calculate an estimate for the proportion of these 43 presidents whose age at inauguration was within 1 standard deviation of the mean.
Give your answer correct to 2 significant figures.

.....

(5)

Peter states that the age at inauguration of these presidents is normally distributed.

(d) Do you agree?

Use your answer to part (c) to justify your answer.

.....

.....

(1)

(Total for Question 13 is 9 marks)

14 X and Y are two events such that

$$P(X \text{ and } Y) = 0.2 \qquad P(X \mid Y) = 0.4 \qquad P(Y \mid X) = 0.5$$

Hiki says that

"X and Y are independent events".

Do you agree?
Explain why.

.....

.....

.....

.....

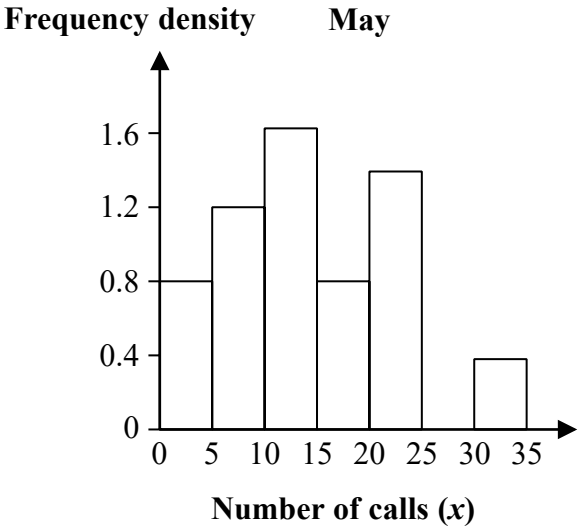
.....

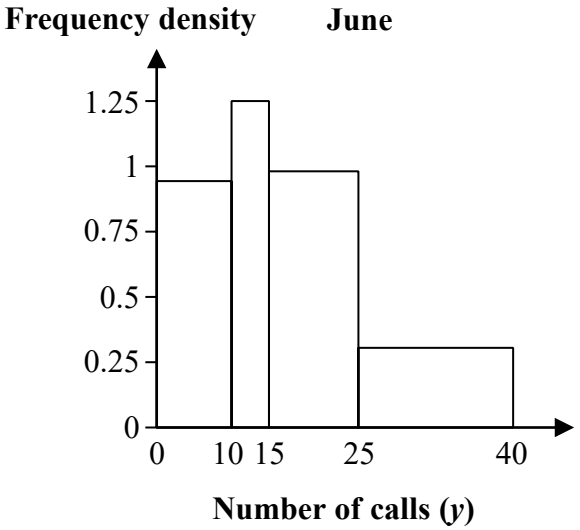
(Total for Question 14 is 2 marks)

15 Jai is investigating the numbers of calls received by a helpline.

He recorded the number of calls received by the helpline on each day in May and on each day in June.

Jai wanted to compare the results by drawing histograms.
He drew the following diagrams.





Discuss whether these diagrams are appropriate in order to compare the results.

(Total for Question 15 is 3 marks)

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Paper 2 Higher tier mark scheme

Question number	Answer	Additional guidance	Mark
1(a)	B1 for Transport	B1 cao	**(1)**
1(b)	B1 for upwards	B1 for upwards oe, eg increasing	**(1)**
1(c)	B1 for a correct answer, e.g. rounding error	B1 for a correct answer	**(1)**
2	B1 Top right of grid B1 Squares are shaded darkest in this region.	B1 for a correctly identifying the region (e.g. NE corner) (But simply listing individual squares is B0) B1 for a statistical reason relating to use of the key.	**(2)**
3(a)	B1 quantitative B1 continuous		**(1)** **(1)**
3(b)(i) **3(b)(ii)**	M1 16 44 A1 30 M1M1A1 $\frac{2\times10+4\times16+6\times44+8\times30}{100}=5.88$	M1 for correctly applying a scale to find a second value (implied either by labelling a scale or correctly finding one other frequency). A1 for all values correct. M1 for consistent use of fx with x within interval M1 for correct use of Σfx with x the mid-interval value A1 for 5.88	**(5)**
4(a)(i) **4(a)(ii)** **4(a)(iii)**	B1 (0.3 + 0.4=) 0.7 B1 0.3 M1 $\frac{0.3}{0.5}$ A1 = 0.6	For probability answers accept equivalent fractions, decimals or percentages	**(4)**
4(b)	M1 0.8×0.5 A1 = 0.4		**(2)**

Question number	Answer	Additional guidance	Mark
5	**Collecting data** B1 for identifying one appropriate thing that should be included in the plan for collecting data **and** B1 for explaining why this aspect is appropriate OR B1 for deciding what data to collect and/or how to collect and record it **and** B1 for an appropriate reason OR B1 for a strategy to process data **and** B1 for an appropriate reason OR B1 for designing a collection method for primary/secondary data **and** B1 for an appropriate reason OR B1 for appreciating the importance of acknowledging sources **and** B1 for an appropriate reason OR B1 for recognising where issues of sensitivity may influence data availability **and** B1 for an appropriate reason **Processing and presenting** B1 for planning to organise and/or process data **and** B1 for an appropriate reason OR	B1B1B1 for each of three planned elements and B1B1B1 for each of three appropriate reasons from their three things in the statistical enquiry cycle. Maximum 4 marks if only one aspect (from Collecting data, Processing and presenting, Interpreting and Evaluating) is referenced. B1 for e.g. use amount of time measured to the nearest minute **and** B1 for e.g. this is sufficient as there will be a large range of times B1 for e.g. collect data for 23 boys and 23 girls **and** B1 for e.g. this will make the calculation of quartiles easier B1 for e.g. use random sampling **and** B1 for e.g. this reduces bias as Gary's friends/class/peers may generally watch the same programs B1 for e.g. use primary data **and** B1 for e.g. this increases reliability as Gary will know how the data was collected B1 for e.g. A student (Gary) should collect the data **and** B1 for e.g. students are more likely to give an honest answer to a fellow student (less threatening) B1 for e.g. use a grouped frequencies table for the data **and** B1 for e.g. as this will enable a quick way of estimating the mean or this can be used to draw a histogram	**(6)**

Question number	Answer	Additional guidance	Mark
5 ***continued***	B1 for planning to generate diagrams and/or visualisations to represent the data **and** B1 for an appropriate reason OR B1 for planning to generate statistical measures to compare data **and** B1 for an appropriate reason **Interpreting** B1 for planning to interpret diagrams and/or calculations/measures **and** B1 for an appropriate reason OR B1 for planning to make an inference and/or prediction **and** B1 for an appropriate reason **Evaluating** B1 for planning to identify weaknesses in approach or representation **and** B1 for an appropriate reason OR B1 for planning to refine the processes to elicit further clarification of the hypothesis **and** B1 for an appropriate reason	B1 for e.g. use box plots **and** B1 for e.g. these will enable the comparison of both the medians and the IQRs (i.e. the distributions) of the data B1 for e.g. interpret results for each individual school year **and** B1 for e.g. as different years could have different watching habits B1 for e.g. by comparing means **and** B1 for e.g. you can see whether the amount of time that boys spend watching TV is greater, in general, than the amount of time that girls spend watching TV B1 for e.g. use the results from the school to predict the results nationally **and** B1 for e.g. as students in different parts of the country are likely to have the same watching habits B1 for e.g. choose not to display the information in histograms **and** B1 for e.g. as "the target audience" may not know how to interpret them B1 for e.g. consider using more than one type of visual representation (for the same information) **and** B1 for e.g. as different representations focus on different aspects of the data	

Question number	Answer	Additional guidance	Mark
6(a)	B1 Male median > female median OR Male IQR > female IQR OR Male range > female range B1 Male cats weigh more than female cats OR there is a greater variation in the weights of male cats compared with female cats	B1 for comparison of median or range/IQR values B1 for contextual interpretation of comparison of median or range/IQR values	**(2)**
6(b)	M1 for $\frac{3(47+1)}{4}$ $(= 36)$ accept 47×0.25 $(= 11.75)$ A1 =12	M1 for $\frac{3(47+1)}{4}$ $(= 36)$ A1 for 12	**(2)**
7(a)	B2 for Yes / good choice of diagram AND reference to the data being bivariate OR if B2 not earned B1 for Yes / good choice of diagram and reason that does not refer to the type of data OR B1 for referring to the data as bivariate but without commenting on whether the diagram is appropriate or not	B2 for complete assessment of the appropriateness of the diagram with a reason OR if B2 not earned B1 for an incomplete assessment of the appropriateness of the diagram	**(4)**
7(b)	B1 for Lata is not right – (the test results are correlated, but) the English test mark does not cause the Maths test mark oe	B1 for assessment of the given conclusion, including reference to correlation not implying causation	**(1)**
7(c)	B1 for correct interpretation in context e.g. • Correlation for Maths test scores and Science test scores is stronger than the correlation between Maths test score and English test score (or second correlation is stronger) oe • There is a greater association between Maths test score and Science test score than between Maths test score and English test score	B1 for statistical interpretation in context	**(1)**

Question number	Answer	Additional guidance	Mark
8(a)	M1 $(P_{90} - P_{10} =) 72 - 8$ A1 64	M1 for subtracting two percentiles identified from graph, at least one correct. May be seen on graph. A1 Accept [63~65] if working shown and P_{10} or P_{90} correct.	**(2)**
8(b)	B1 UK median is 38 (from graph) B1ft Manchester has a smaller 10 to 90 percentile range B1ft for either • Manchester population (9 years) younger on average • Manchester population has less variation of ages	B1 for stating UK median is 38 (or 9 years higher than Manchester). B1ft for correct statistical reasoning. (Follow through their answer to (a) for 2nd and 3rd B1 mark) B1ft for one contextual interpretation of average or spread comparisons. Accept equivalent/converse statements about UK population.	**(3)**
8(c)(i) **8(c)(ii)**	M1 (Skew =) $\frac{3(32-29)}{19.3}$ A1 $= 0.4663...$ B1ft (Positive skew means that) the majority in the population are younger with fewer older people, OR there is a greater spread of ages at the upper end	M1 for demonstrating correct use of formula A1 for 0.47 or better B1ft allow any equivalent wording for a correct contextual interpretation of their skew value. (This mark may be gained independently following comparison of mean/median)	**(3)**

Question number	Answer	Additional guidance	Mark
9(a)	M1A1A1 for calculating standardised scores M1 Long jump: $\frac{6.43-6.10}{0.26}$, High jump: $\frac{1.86-1.79}{0.066}$ A1 Long jump: 1.3 A1 High jump: 1.1 B2 for e.g. Better performance in long jump (relative to their competitors) as there is a higher standardised score in long jump OR (if B2 not scored) B1 for e.g. Better performance in long jump (relative to their competitors) with an attempt at a reason	M1 for either correct calculation A1 for awrt 1.3 A1 for awrt 1.1 B2 for a correct contextual interpretation of results with a correct reason using standardised scores OR (if B2 not scored) B1 for a correct contextual interpretation of results with an attempt at a reason	**(5)**
9(b)	M1 $-0.32=\frac{42.51-x}{5.85}$ A1 (x =) 44.382	M1 for demonstrating correct use of formula. A1 for awrt 44.38	**(2)**
10(a)	B1 More people live in the Lake District B1 Larger pie chart. o.e.	B1 Accept converse contextual statements about Snowdonia. B1 for any statistical reasoning implying larger chart (may refer to radius/diameter/area/size, etc)	**(2)**
10(b)	B1 More (people aged 40-59) live in the Lake District + reason B1 Larger area.	B1 Accept converse contextual statements about Snowdonia. Answer without reason scores B0 B1 for statistical reasoning with reference to area clearly implied. (Accept complete clear argument based on proportions of different sized populations.)	**(2)**
10(c)	B1 for a correct answer, e.g. populations (of regions), angles (of sectors), numbers of people (in each sector)	B1 for a correct answer	**(1)**
10(d)	B1 for $3 \div \sqrt{3.9}$ (=1.519…)	B1 for a complete method to find radius of circle for Exmoor	**(1)**

Question number	Answer	Additional guidance	Mark
11(a)	B1 for LQ = 14, UQ = 34 M1 for 34 + 1.5 × ('34' – '14') (= 64) 70 is an outlier as 70 > 64	B1 for 14 or 34 M1 for 34 + 1.5 × ('34' – '14') with their quartiles A1 for demonstrating understanding of calculations for outliers	**(3)**
11(b)	B1 B1 B1 B1 B1 for each of five correct aspects e.g. • Use of statistical words, e.g. average/spread is too vague in this context (conclusion not appropriate) • Comparing medians (conclusion appropriate) • Comparing ranges or IQRs, e.g. IQRs are the same (conclusion appropriate) or Pine Wood range greater than Acorn Wood range (conclusion not appropriate) • Identifying Acorn Wood as having a negative skew (conclusion not appropriate) • Identifying Pine Wood as having no skew or is symmetrical (conclusion not appropriate)	B1 B1 B1 B1 B1 for each of five correct comments assessing the appropriateness of the conclusions	**(5)**

Question number	Answer	Additional guidance	Mark
12(a)	B1 lower warning line added at 4.0 and labelled. B1 both action lines correct and labelled (Upper line at 4.25, lower line at 3.95)	B1 for knowing warning lines should be symmetrical. B1 for knowing that action lines should be 3 × s.d. from the mean. Both lines correct with correct labels.	**(2)**
12(b)	B1ft Another sample should be taken immediately B1 If further sample is also beyond warning line then production should be stopped. (If within warning lines then production can continue.)	B1ft for attempting correct practical decision based on position of their value relative to warning/action lines. B1 (cao) for complete description of actions needed including criteria for stopping after a second sample.	**(2)**
12(c)	B1 David is incorrect because sample means will be more closely grouped than individual values OR standard deviation will be >0.05	B1 for demonstrating understanding that sample means will have a closer grouping.	**(1)**

Question number	Answer	Additional guidance	Mark
13(a)	M1 for correct answer, e.g. may be out of date, sources may be unreliable, etc.	B1 for a correct disadvantage of collecting data using technology	**(1)**
13(b)	B1 for $\frac{131334.5}{43}$ or $\frac{2361}{43}$ B1 for $\sqrt{\frac{131334.5}{43} - \left(\frac{2361}{43}\right)^2}$ or $\frac{131334.5}{43} - \left(\frac{2361}{43}\right)^2$ leading to correct answer	B1 for a correct use of 43 in a calculation B1 for showing a correct complete calculation	**(2)**
13(c)	e.g. M1 LB = mean – 6.29 (= 48.62) M1 UB = mean + 6.29 (= 61.2) M1 (mean – 52) × 2.6 + (52 – LB) × 2.2 (= 15.002) (57 – mean) × 2.6 + (61.2 – 57) × 2 (= 13.834) '15.002' + '13.834' (= 28.836) M1 28.836 ÷ 43 A1 = 0.671	M1 for correct method to find number of presidents 1 sd below mean M1 for correct method to find number of presidents 1 sd above mean M1 for correct method to find total number of presidents within 1 sd of mean M1 for correct method to find proportion of presidents within 1 sd of mean A1 for answers rounding to 0.67 or 67%	**(5)**
13(d)	B1 for e.g. yes, close to 68%	B1 for referring to 68%	**(1)**

Question number	Answer	Additional guidance	Mark
14	B2 for yes with a correct complete reason, e.g. $P(X \mid Y) = 0.4$ implies $P(Y) = 0.5$, and as $P(Y \mid X) = 0.5$, so X and Y are independent events OR if B2 not earned B1 for using $P(X \text{ and } Y) = P(X) \times P(Y)$ or $P(Y \mid X) = P(Y)$ or $P(X \mid Y) = P(X)$	B2 for a correct comment assessing the appropriateness of the conclusion OR if B2 not earned B1 for an incomplete attempt to assess the appropriateness of the conclusion	**(2)**
15	B1 B1 B1 for each of three correct aspects • e.g. diagrams are not appropriate for discrete data • e.g. amount of data is not appropriate as too small for the number of intervals used • e.g. it is not appropriate to compare the two diagrams as different class intervals / different frequency density scales are used	B1 B1 B1 for each of three correct comments assessing the appropriateness of the diagrams used	**(3)**

edexcel

For information about Edexcel, BTEC or LCCI qualifications
visit qualifications.pearson.com

Pearson Education Limited. Registered in England and Wales No. 872828
Registered Office: 80 Strand, London WC2R 0RL
VAT Reg No GB 278 537121

ISBN 978-1-4469-5267-2