

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

--	--	--	--

Time 1 hour 30 minutes

**Paper
reference**

1ST0/1H

Statistics
PAPER 1
Higher Tier

You must have:

Ruler graduated in centimetres and millimetres, protractor,
pair of compasses, 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.
- Good luck with your examination.

Turn over ►

P65514A

©2021 Pearson Education Ltd.

11/11/11 1stat-nov-21



552 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

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

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

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1** Emily is designing a questionnaire to obtain information about how people in Manchester travel to work.

Before she gives out the questionnaire, she is going to do a pilot study.

- (a) Give a reason why a pilot study is a suitable thing for her to do.

.....
.....
(1)

- (b) Discuss whether or not it would be appropriate to carry out the pilot study on the entire population.

.....
.....
(1)

One of the questions on Emily's questionnaire is designed to collect qualitative data. Jane suggests to Emily that she should draw a histogram for the data collected from this question.

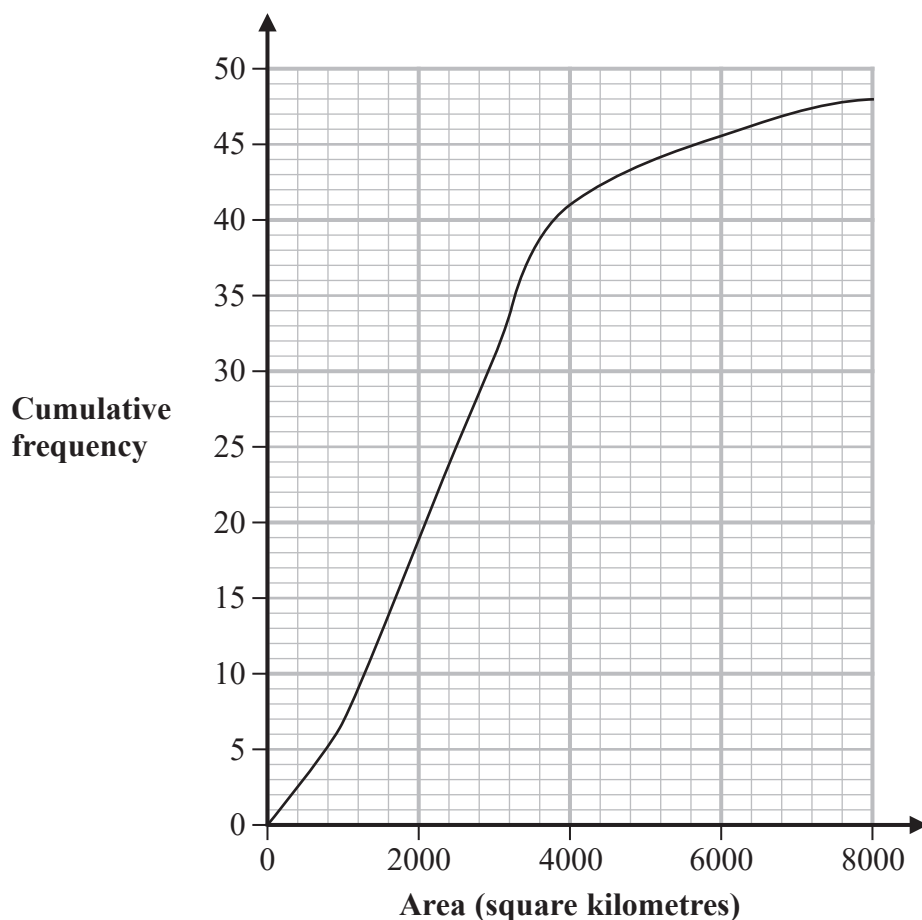
- (c) Comment on Jane's suggestion.

.....
.....
(1)

(Total for Question 1 is 3 marks)



2 The cumulative frequency graph gives information about the area, in square kilometres, of the 48 counties in England.



(Source: https://en.wikipedia.org/wiki/List_of_ceremonial_counties_of_England)

(a) (i) Find an estimate of the 75th percentile of this information.

..... square kilometres
(2)

(ii) Interpret this value in context.

.....
.....
(1)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Half of the counties in England have an area between 2000 square kilometres and k square kilometres.

(b) Work out an estimate for the value of k .

.....
(3)

(Total for Question 2 is 6 marks)



3 John and Susan are investigating films.

John wants to find out the favourite type of film of the students at his school.

Susan wants to find out which type of film made the most money in UK cinemas last year.

(a) State the population for

(i) John's investigation,

(1)

(ii) Susan's investigation.

(1)

Susan plans to collect the information for her investigation from the internet.

(b) Give advice to Susan so that she can ensure that the information she collects is reliable.

(1)

John is deciding between two different sampling methods for his investigation.

Method A: Quota sampling by sampling 20 students from each year group.

Method B: Opportunity sampling by selecting the first 50 students he sees one day.

(c) For each method, give one reason why the method is **not** random.

Method A:

Method B:

(2)



(d) Explain which of the two methods John should use in order to minimise bias.
Give a reason for your answer.

.....

.....

.....

.....

(2)

(Total for Question 3 is 7 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

- 4 The table shows the value of the gross domestic product (£ million) of the UK for each quarter from 2007 to 2009

Year	Quarter	Gross domestic product (£ million)
2007	1	444 292
	2	447 498
	3	451 288
	4	455 043
2008	1	456 663
	2	453 283
	3	445 818
	4	436 137
2009	1	428 886
	2	428 073
	3	428 682
	4	430 166

(Source: <https://www.ons.gov.uk/economy/grossdomesticproductgdp>)

A country is in recession when its gross domestic product falls in two or more consecutive quarters.

The UK went into recession in 2008 Quarter 3

A country comes out of recession in the quarter in which its gross domestic product rises.

- (a) According to the table, in which year and quarter did the UK come out of this recession?

.....
(1)

- (b) Using 2007 Quarter 1 as the base, work out the simple index number for the gross domestic product of the UK in 2008 Quarter 1

Give your answer correct to the nearest whole number.

.....
(2)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

The table below shows the simple index number for the gross domestic product of the UK for each quarter in 2010 using 2007 Quarter 1 as the base.

Year	Quarter	Gross domestic product simple index number
2010	1	97.3
	2	98.1
	3	98.7
	4	98.8

(c) Calculate the value of the gross domestic product of the UK in 2010 Quarter 1

£..... million
(2)

Marc says that the UK was in recession throughout 2010 because all of the simple index numbers are less than 100

(d) Explain whether or not Marc is correct.

.....

.....

.....

(2)

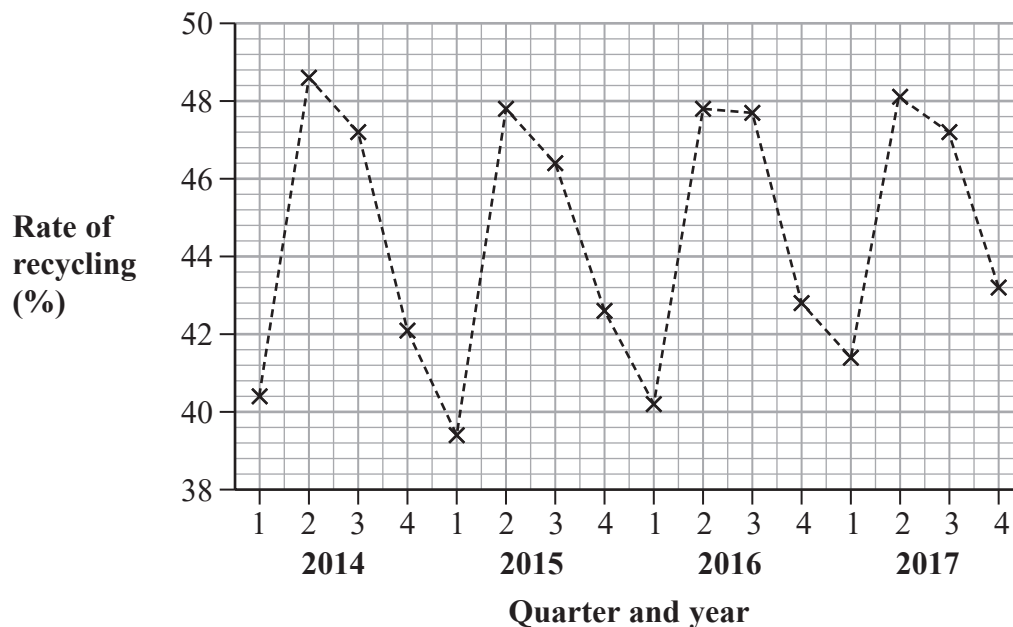
(Total for Question 4 is 7 marks)



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 .

- 5 The percentage of household waste that is recycled is called the rate of recycling.

The time series graph shows information about the rate of recycling for households in the UK for each quarter from 2014 to 2017



(Source: Department for Environment, Food and Rural Affairs)

- (a) For which quarter each year from 2014 to 2017 was the rate of recycling between 42% and 44%?

1 2 3 4

(1)

- (b) Describe the seasonal trend shown by the time series graph.

(2)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Andrea uses n -point moving averages for the information shown in the time series graph in order to help determine the trend.

- (c) Write down an appropriate value of n .
Give a reason for your answer.

(2)

Andrea uses the time series graph to predict the rate of recycling in the UK for 2018 Quarter 3

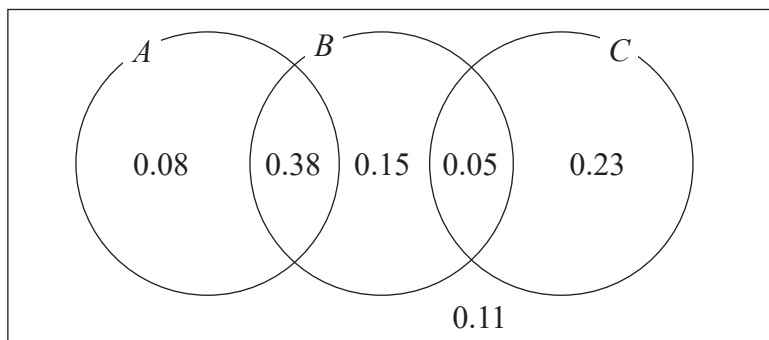
- (d) Discuss the reliability of Andrea's prediction.

(1)

(Total for Question 5 is 6 marks)



6 The Venn diagram shows probabilities relating to the three events A , B and C .



(a) Write down which two events from A , B and C are mutually exclusive. Give a reason for your answer.

.....

.....

.....

(2)

(b) Find $P(B)$.

.....

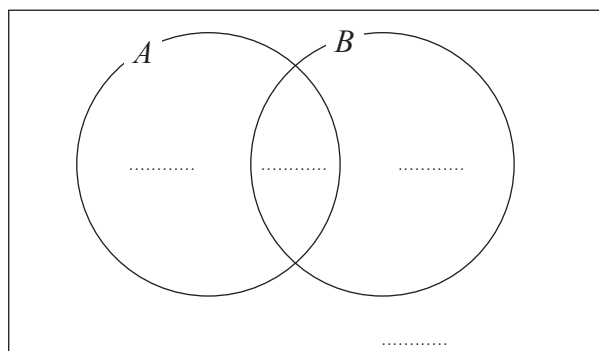
(1)

(c) Find $P(A \text{ or } C)$.

.....

(2)

(d) Complete the Venn diagram below to show the probabilities relating to the two events A and B .



(2)

(Total for Question 6 is 7 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

7 The table shows the number, in millions, of visitors to each of four tourist attractions in 2005 and in 2018

Attraction	Number of visitors in 2005 (millions)	Number of visitors in 2018 (millions)
British Museum	4.5	5.8
Tower of London	1.9	2.9
Edinburgh Castle	1.2	2.1
St Paul's Cathedral	0.7	1.7
Total	8.3	12.5

(Source: *www.alva.org.uk*)

Jill draws two comparative pie charts for the information in the table.

(a) Describe an advantage of using comparative pie charts for this information rather than using ordinary pie charts.

(1)

(b) (i) Compare the angles for the sectors in the pie charts representing the British Museum in Jill's comparative pie charts for 2005 and for 2018

(1)

(ii) Compare the areas for the sectors in the pie charts representing the British Museum in Jill's comparative pie charts for 2005 and for 2018

(1)

Jill's comparative pie chart for 2005 has radius 5 cm.

(c) Calculate the radius of Jill's comparative pie chart for 2018
Give your answer correct to one decimal place.

..... cm

(2)

(Total for Question 7 is 5 marks)



- 8 Rachel has collected information on the minimum wage, in US dollars per hour, for each of 50 US regions.

She uses a spreadsheet to sort the data into ascending order of size.

Part of her spreadsheet is shown below.

	A	B
1	5.15	
2	5.15	
3	7.25	
4	7.25	
~~~~~		
46	9.75	
47	9.75	
48	10	
49	10	
50	11.5	

(Source: <https://www.dol.gov/whd/state/stateMinWageHis.htm>)

Rachel uses statistical software to summarise the data.

The output is shown below.

n	50
Mean	8.1192
Minimum	5.15
Lower quartile	7.25
Median	8.05
Upper quartile	8.75
Maximum	11.5



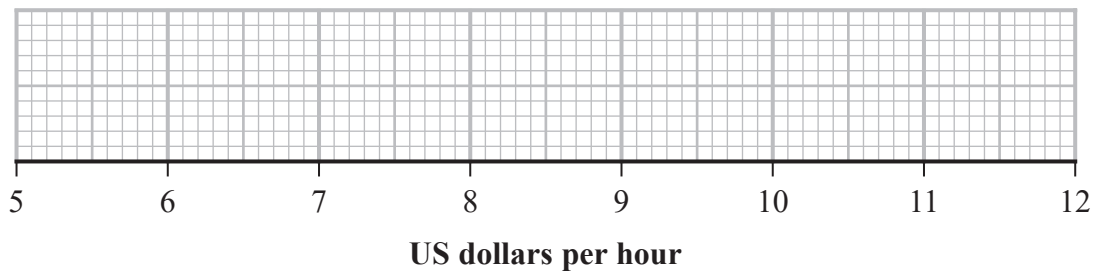
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Using the information in the table and the spreadsheet, draw a box plot with any outliers clearly marked.

You must show any calculations you use.



(Total for Question 8 is 6 marks)



9 A professor uses a weighted mean in order to calculate each student's mark for the term.

Here is information about the weighting of each task.

<b>Task</b>	Assignment 1	Assignment 2	Assignment 3	Final exam
<b>Weighting</b>	2	3	5	8

Each task is marked out of 100 marks and it is not possible to achieve a score greater than 100 marks for a task.

The table below gives the marks achieved by Anders on his three assignments.

Assignment 1	Assignment 2	Assignment 3
72	84	88

To achieve a grade A, the weighted mean for the term must be at least 90 marks.

Anders is going to take his final exam next week.

Determine whether or not Anders can achieve a grade A.

(Total for Question 9 is 4 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA





- 10 Amelia wants to see if there is a relationship between the height of a female athlete and the time it takes her to run 100 metres.

The table gives information collected by Amelia from the internet on the heights of 8 female athletes and the best time taken by each athlete to run 100 metres.

Athlete	Height (metres)	Time (seconds)	Height ranks	Time ranks		
A	1.79	10.90	1			
B	1.75	10.83	2			
C	1.68	10.94	3			
D	1.67	10.71	4	1		
E	1.65	11.80	5			
F	1.60	10.92	6			
G	1.59	10.87	7			
H	1.52	10.86	8			

(Source: *www.olympic.org*)

Amelia's hypothesis is that the taller the female athlete, the less time it takes her to run 100 metres.

- (a) Is Amelia's hypothesis supported by the data?  
You must justify your answer.

(5)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Write down one thing that Amelia could do to improve the reliability of her conclusion.

.....

.....

.....

(1)

**(Total for Question 10 is 6 marks)**



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

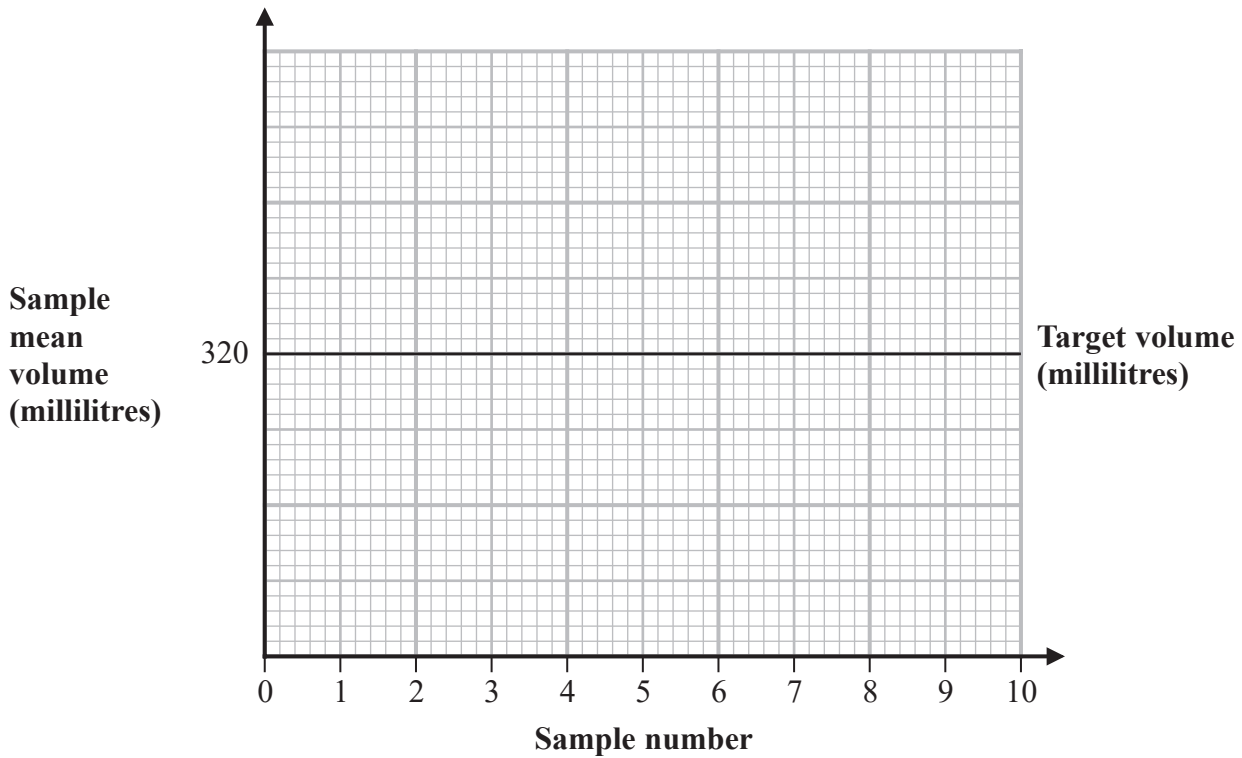
11 A company uses a machine to fill cans with orange juice.

For quality assurance, the company takes samples of 4 cans of orange juice at regular intervals in order to check the mean amount of orange juice in the 4 cans.

When the machine is working properly, the mean volume per can of orange juice, in millilitres, in a sample of 4 cans is modelled by the normal distribution  $N(320, 2^2)$ .

The company uses quality control charts in order to monitor the production process.

(a) Complete the quality control chart below for the company to use.



(5)

Ben is in charge of quality control.

He takes samples of 4 cans at regular intervals throughout the day. In 20 samples, Ben finds that there are 2 samples for which the sample mean exceeds the action limit.

Ben decides to allow the machine to continue filling cans, and that the next time a sample mean exceeds the action limit, he will shut down and reset the machine.

(b) Explain whether or not Ben's actions are appropriate with respect to quality control.

.....

.....

.....

(1)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Kohan claims that the volume of orange juice in individual cans should be modelled by the same normal distribution as the mean volume per can of orange juice in a sample of 4 cans.

- (c) Explain whether or not Kohan's claim is justified.  
You must give reasons for your answer.

.....

.....

.....

.....

.....

(3)

**(Total for Question 11 is 9 marks)**

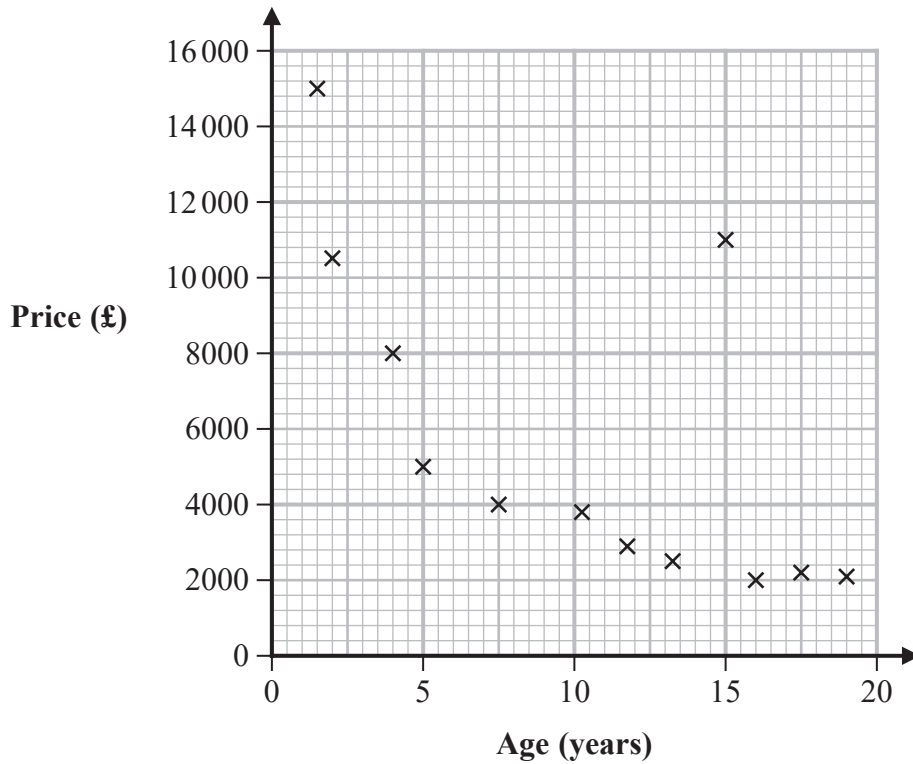


DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

12 The age and price of 12 used cars of the same type are plotted on the scatter diagram below.



One of the points on the scatter diagram represents an outlier.

(a) Draw a circle around the outlier. (1)

(b) Comment on the price of the car represented by this point. (1)



Pearson's product moment correlation coefficient for the 12 cars, including the outlier, is calculated as  $-0.662$

With the outlier removed, Pearson's product moment correlation coefficient for the 11 cars is calculated as  $-0.842$

(c) (i) Give a reason why it might be appropriate to remove the outlier.

.....  
.....

(ii) Give a reason why it might not be appropriate to remove the outlier.

.....  
.....

(2)

With the outlier removed, Spearman's rank correlation coefficient for the data in the scatter diagram is calculated as  $r_s$

(d) Describe fully how the value of  $r_s$  should compare with  $-0.842$   
Give a reason for your answer.

.....  
.....  
.....  
.....

(3)

**(Total for Question 12 is 7 marks)**



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

13 Tomi has collected data about the weights, in kg, of cats of a particular breed.

The table gives some of the percentiles of his data.

Percentile	Weight (kg)
97.5th	7.0
80th	6.1
60th	5.7
40th	5.3
20th	4.9
5th	4.3
2.5th	4.0

(a) Find the 2.5th to 97.5th interpercentile range.

..... kg  
(1)

One of the cats that Tomi collected data about is selected at random.

(b) Find the probability that the weight of this cat is between 4.3 kg and 5.7 kg.

.....  
(1)

Tomi plans to use the mean and the standard deviation of the weights of these cats to summarise his data.

(c) Give a reason why Tomi's plans are appropriate.

.....  
.....  
(1)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Tomi claims that the weights of these cats can be modelled using a normal distribution with mean 5.5 kg and standard deviation 0.75 kg.

(d) Assess whether or not the data support Tomi's claim.

.....

.....

.....

.....

(4)

**(Total for Question 13 is 7 marks)**

**TOTAL FOR PAPER IS 80 MARKS**

