

Write your name here

Surname

Other names

Pearson Edexcel

Level 1/Level 2 GCSE (9 - 1)

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--

Statistics

Paper 2

Higher Tier

Specimen Papers Set 1

Time: 1 hour 30 minutes

Paper Reference

1ST0/2H

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 ►

S59309A

©2018 Pearson Education Ltd.

6/6/4/6/1/1/



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

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

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

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1** Tracy wants to find out what improvements to the youth club are wanted by the members of the club.

She plans to give each member of the club a questionnaire.

Here is part of the questionnaire.

1. Name:

2. Age: 10 to 12 years ☐ 12 to 14 years ☐ 14 to 16 years ☐

3. Do you agree that the condition of the club needs improving?
Yes ☐ No ☐ Don't know ☐

4. How much time do you spend in the youth club?
.....

5. Do you know someone who has caused damage in the youth club?
Yes ☐ No ☐

- (a) Explain why Question 2 will not give reliable results.

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

- (b) Explain why Question 3 is not a good question.

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

- (c) Explain why there might be a problem with the answers given to Question 4

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

- (d) Explain why the answers given to Question 5 may not be valid.

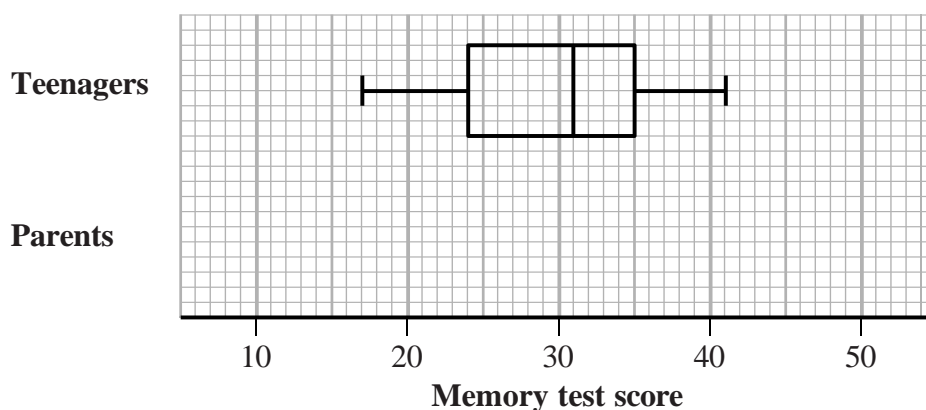
.....
.....
(1)

(Total for Question 1 is 4 marks)



- 2 Matthew recorded the scores in a memory test taken by some teenagers and by their parents. The teenagers took the same test as their parents.

The box plot shows information about the recorded test scores for the teenagers.



The table gives information about the recorded test scores for the parents.

Least score	Lower quartile	Median	Upper quartile	Greatest score
12	21	29	35	39

- (a) On the grid above, draw a box plot for the recorded test scores for the parents.

(2)

- (b) Compare the distributions.

Give three comparisons and interpret one of your comparisons.

.....

.....

.....

.....

.....

.....

.....

.....

(4)



The information for the teenagers is based on data collected from 56 teenagers.

- (c) Estimate the number of these teenagers whose recorded test score was greater than 24

.....
(2)

(Total for Question 2 is 8 marks)

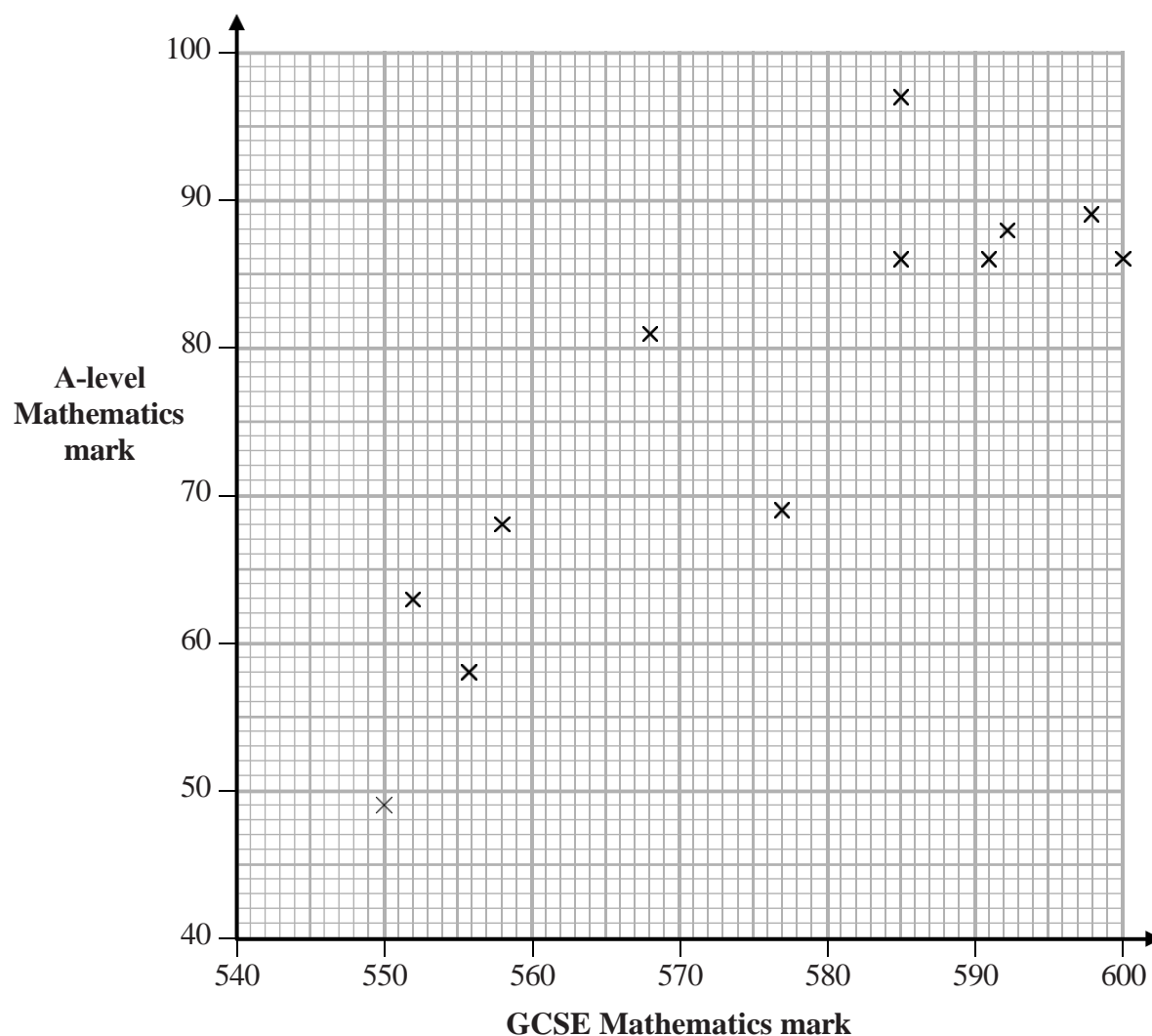


- 3 Julie was investigating the relationship between the marks gained by students in their GCSE Mathematics exam and the marks gained by the same students in an A-level Mathematics exam.

(a) Suggest a hypothesis Julie could use.

(1)

Julie drew a scatter diagram using the marks gained in GCSE Mathematics and the marks gained in A-level Mathematics by each of 12 students.



- (b) For this scatter diagram, explain why the GCSE Mathematics mark is the explanatory variable.

(1)

(c) Explain, giving a statistical reason, whether or not the scatter diagram supports your hypothesis in part (a).

(2)

Using statistical software, Julie obtained the following information for her 12 students.

Mean GCSE Mathematics mark	578
Mean A-level Mathematics mark	78
Gradient of line of best fit	0.7

(d) Using this information, draw a line of best fit on the scatter diagram.

(2)

(e) Interpret the gradient of the line of best fit.

(1)

Julie wants to use the line of best fit to predict the A-Level Mathematics mark for a 13th student. For GCSE Mathematics this student gained a mark of 540

(f) Explain whether or not it would be appropriate to use the line of best fit to make her prediction.

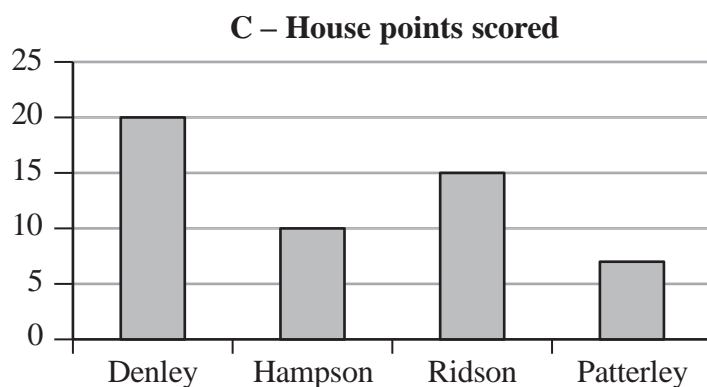
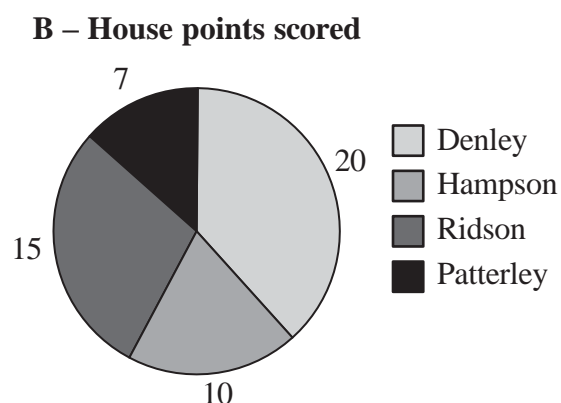
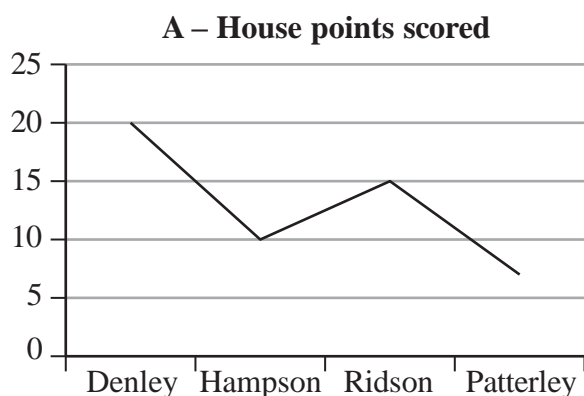
(1)

(Total for Question 3 is 8 marks)



- 4 For each of four school houses at his school, Lee is analysing the points scored in a sports competition.

Using statistical software, he produced the three diagrams below.



- (a) Explain why diagram A is not appropriate for the data.

(1)

Lee wants to show clearly the proportion of the total points scored by each house.

- (b) Explain which of B or C is the best diagram for Lee to use and why.

(2)

(Total for Question 4 is 3 marks)



- 5 Richard wants to find an estimate for the number of rabbits living in a section of woodland.

He caught a sample of 25 rabbits, attached a tag to each rabbit, and then released the 25 rabbits back into the same section of the woodland.

Four weeks later, Richard returned to the same section of woodland and caught a sample of 10 rabbits.

Two of these rabbits were tagged.

Using these results, Richard estimated that there are 125 rabbits in the section of woodland.

- (a) Show how Richard worked out his estimate of 125 rabbits.

(2)

- (b) Considering Richard's statistical method, discuss the reliability of his estimate.

(3)

(Total for Question 5 is 5 marks)



- 6 Barbara found the following information about the average gross pay, in euros (€), for males and for females in ten countries in 2010

Country	Average gross pay (€)	
	Male	Female
Denmark	62 120	49 254
Germany	44 465	34 740
Ireland	48 459	19 177
Spain	29 009	25 101
France	37 627	30 406
Netherlands	47 373	36 696
Portugal	19 424	15 299
Sweden	41 311	33 305
United Kingdom	42 710	31 115
Switzerland	60 135	48 972
Mean	43 263	32 407
Product moment correlation coefficient	0.832	

(Source: Eurostat)

Barbara investigates this information using a suitable graph.

Using the information in the table and her graph, she reaches the following conclusion:

There is strong evidence that, for every €10 000 increase in gross pay for males, gross pay for females increases by €8000



- (a) Explain, giving reasons, what Barbara is most likely to have done to reach this conclusion.

You should consider

- which type of graph she will have used and why
- how she will have used the graph and the information in the table

(5)

- (b) Explain why Barbara's conclusion, based on using the data in the table, may be unreliable.

(1)

(Total for Question 6 is 6 marks)

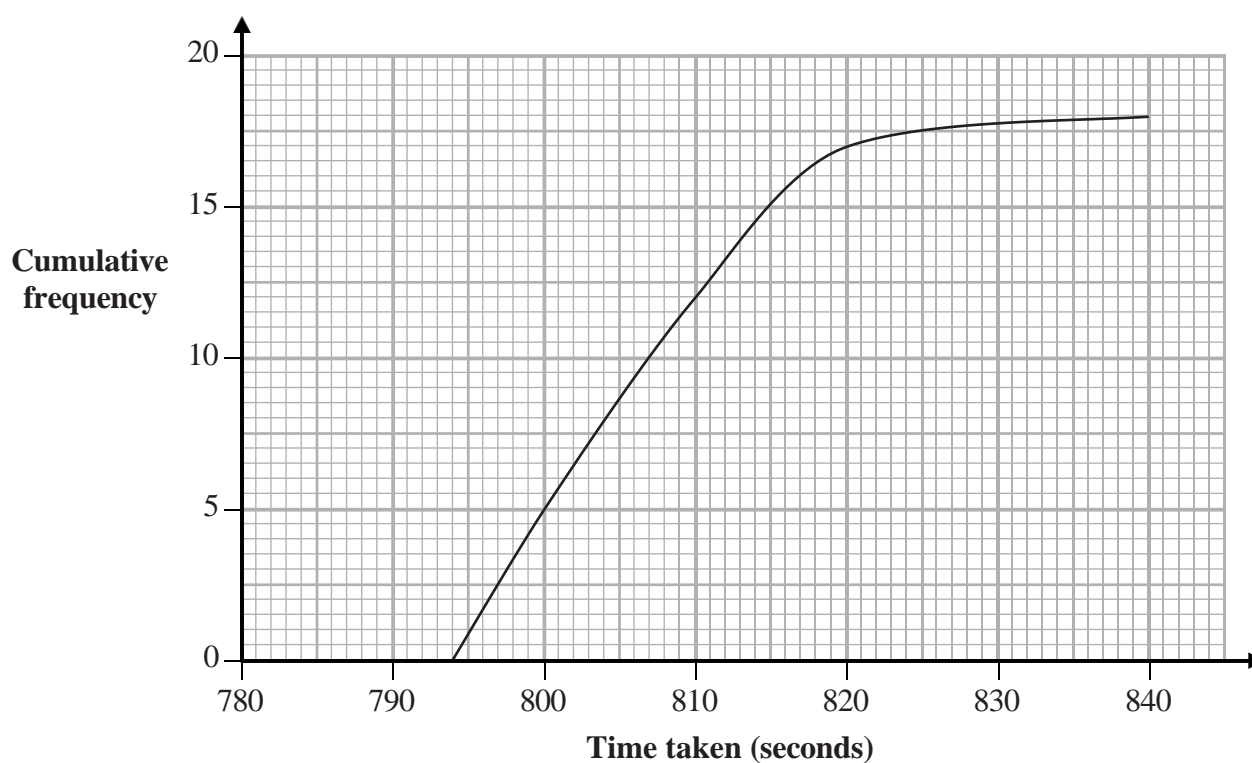


- 7 The table gives information about the times taken, in seconds, by the athletes in the 5000 m Diamond League race held in Oslo in 2016

20th percentile	Median	80th percentile
789	794	808

(Source: adapted from *diamondleague.com*)

The cumulative frequency graph shows information about the times taken, in seconds, by the 18 athletes in the 5000 m Diamond League race held in Zurich in 2016



(Source: adapted from *diamondleague.com*)



- (a) Use statistical calculations to determine in which of the two races were the times taken by the athletes the least **and** in which of the two races were the times taken by the athletes the more consistent.

You must explain how you reach your conclusions.

(5)

The interquartile range of the race time in Zurich is 13 seconds.

The greatest time taken by an athlete in the 5000m race in Zurich is 840 seconds.

Mo thinks that this time is an outlier.

- (b) Determine whether or not Mo is correct.

(2)

(Total for Question 7 is 7 marks)



- 8 Owen collected data about the number of people per km^2 living in each of the 56 counties of England and Wales in 2015

Some information about Owen's data is shown in this incomplete table.

Number of people per km^2 (x)	Number of counties
$0 \leq x < 200$	15
$200 \leq x < 400$	14
$400 \leq x < 600$	14
$600 \leq x < 1000$	r
$1000 \leq x < 6000$	s

(Source: Office for National Statistics)

- (a) Work out the number of counties that have at least 600 people per km^2 .

(1)

- (b) Use linear interpolation and the information in the table to find an estimate of the median number of people per km^2 .

.....people per km^2
(3)

- (c) (i) Using the information in the table, explain how the mean of Owen's data would compare with the median of Owen's data.

- (ii) Explain whether the mean or the median is the most appropriate measure of central tendency to use for Owen's data.

(3)



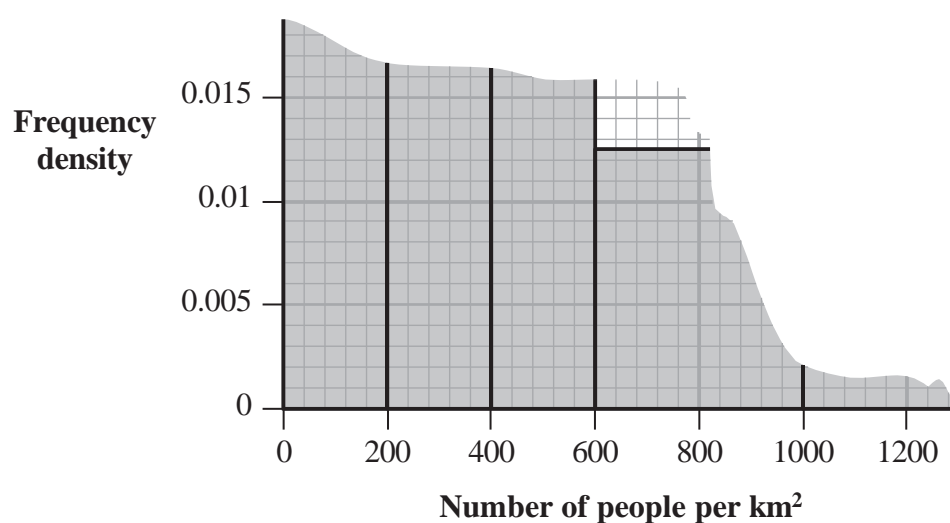
Owen uses statistical software to draw a histogram to show the data he collected.

(d) Explain whether or not a histogram is appropriate.

(1)

Owen prints his histogram but the page gets torn.

Here is what remains.



(e) Work out the value of r and the value of s in the table.

$r = \dots\dots\dots$ $s = \dots\dots\dots$

(2)

(Total for Question 8 is 10 marks)



- 9 The ambulance service has a target of responding to 75% of life threatening calls in 8 minutes or less.

The table shows the actual response rates for two locations in the UK in 2014

Location	Percentage of life threatening calls responded to in 8 minutes or less
Littlehampton	75%
York	93%

(Source: *ambulanceresponsetimes.co.uk*)

For a life threatening call made in 2014

- (i) write down the absolute risk that an ambulance takes more than 8 minutes to respond in Littlehampton,

.....
(1)

- (ii) work out the **relative** risk that an ambulance takes more than 8 minutes to respond in York, compared with the target.

.....
(2)

(Total for Question 9 is 3 marks)



- 10** Adila and Brandon both travel to work each day.
Information about their past journey times, in minutes, is given in the table.

	Mean	Standard deviation
Adila	75	10
Brandon	65	4

On Monday, they each took 60 minutes to travel to work.

- (a) Calculate the standardised scores for the two journey times on Monday.

Adila.....

Brandon.....
(3)

- (b) Using your answers to part (a), explain which of Adila or Brandon had the better journey time on Monday.

.....

.....

.....

.....
(2)

- (c) Explain why standardised scores are appropriate to use in this context.

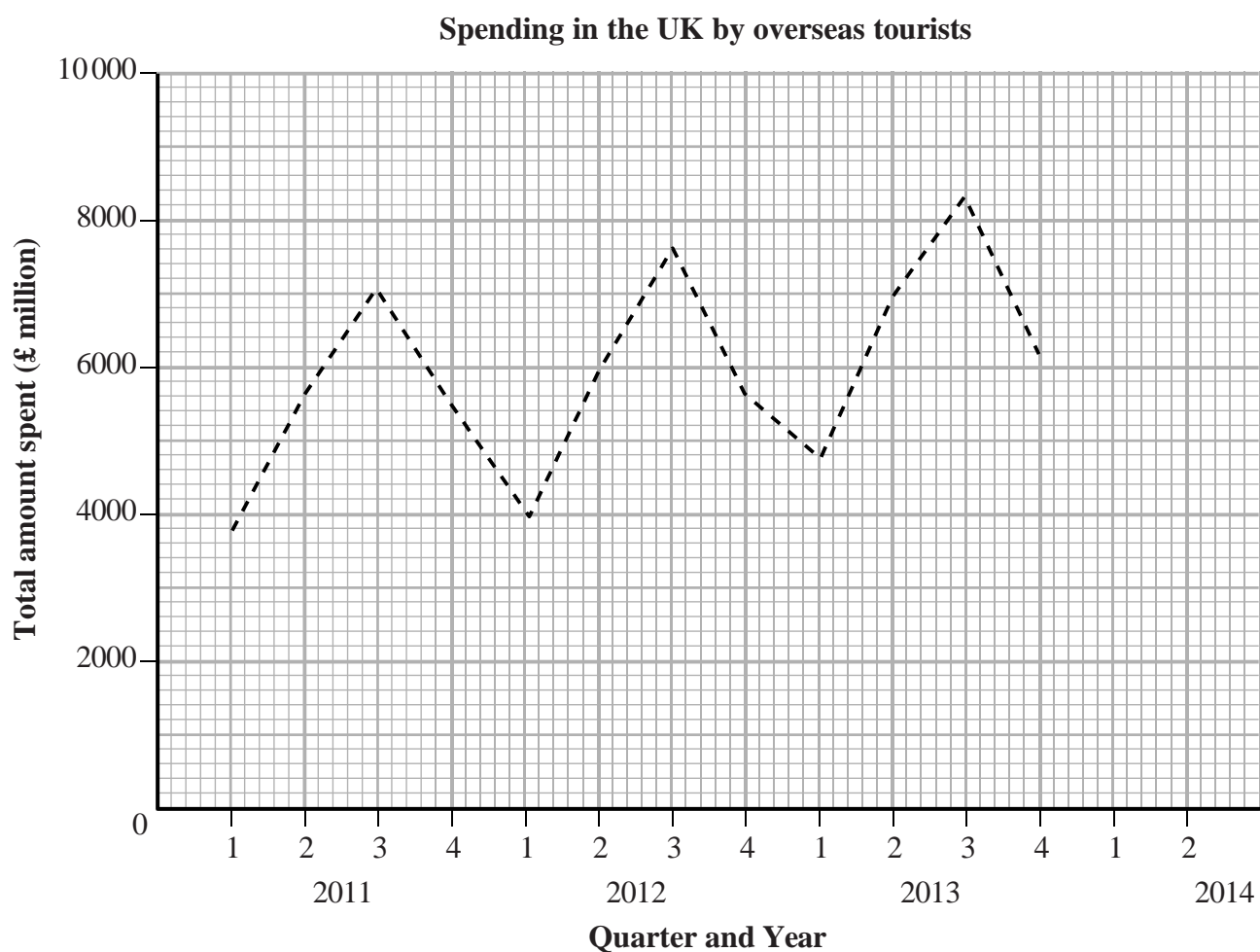
.....

.....
(1)

(Total for Question 10 is 6 marks)



- 11 The time series graph shows information about the total amount of money, in millions of pounds, spent in the UK by overseas tourists for each quarter between 2011 and 2013



(Source: *Office for National Statistics*)

Divya says that the time series graph shows seasonal variation.

- (a) Explain whether or not she is correct.

(1)



Divya considers finding 3-point or 4-point or 5-point moving averages for the data in the time series graph.

- (b) Explain which of these moving averages is the most appropriate to find.

(1)

Divya wants to calculate an estimate for the spending in the UK by overseas tourists in Quarter 1 of 2014

- (c) Describe how Divya can use moving averages, mean seasonal effect and the time series graph to do this.

(5)

(Total for Question 11 is 7 marks)



- 12 The incomplete table gives some information about the number of births in the USA, the total population of the USA and the crude birth rate for the USA in 1960 and in 2021.

Year	Number of births	Total population	Crude birth rate
1960	179.3 million	23.7
2021	3,659,289	331.9 million

(Source: *data.worldbank.org*, *www.census.gov*, *www.cdc.gov*)

- (a) Using the formula below, complete the table.

$$\text{crude birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}}$$

(4)

The table gives information about the crude birth rate for France and for Germany in 2021.

Country	Crude birth rate
France	10.9
Germany	9.6

Jan concludes that the population of France increased more than the population of Germany in 2021

- (b) Explain why Jan may **not** be correct.

.....

.....

..... (1)

(Total for Question 12 is 5 marks)



- 13 A biased coin is twice as likely to land Heads than Tails.

An experiment involves flipping 9 identical biased coins.

X represents the number of Heads obtained in the experiment.

Michael says that the values of X can be modelled by the distribution $B\left(9, \frac{2}{3}\right)$

- (a) Write down the name of this distribution model.

.....
(1)

- (b) Using Michael's model, show why the mean of X is 6

(1)

The experiment is carried out 15 times and the number of Heads obtained in each experiment is recorded.

Michael uses his model to find an estimate for the number of experiments in which the number of Heads recorded is exactly 6

- (c) Find Michael's estimate.

.....
(4)



- (d) Write down **two** assumptions which Michael needs to make about the experiment so that the distribution $B\left(9, \frac{2}{3}\right)$ is an appropriate model to use.

$B\left(9, \frac{2}{3}\right)$

(2)

(Total for Question 13 is 8 marks)

TOTAL FOR PAPER IS 80 MARKS



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

