**Pearson
Edexcel GCSE**

Statistics
Paper 1H

Higher Tier

Thursday 18 June 2015 – Afternoon	Paper Reference
Time: 2 hours	**5ST1H/01**

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

Information

- The total mark for this paper is 100.
- The marks for **each** question are shown in brackets
  – *use this as a guide as to how much time to spend on each question.*
- Questions labelled with an **asterisk** (*****) are ones where the quality of your
  written communication will be assessed
  – *you should take particular care on these questions with your spelling, punctuation
    and grammar, as well as the clarity of expression.*

Advice

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

Turn over ▶

PEARSON

Higher Tier Formulae

You must not write on this page.
Anything you write on this page will gain NO credit.

Mean of a frequency distribution $= \dfrac{\sum fx}{\sum f}$

Mean of a grouped frequency distribution $= \dfrac{\sum fx}{\sum f}$, where x is the mid-interval value.

Variance $= \dfrac{\sum (x - \bar{x})^2}{n}$

Standard deviation (set of numbers) $\sqrt{\left[\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n} \right)^2 \right]}$

or $\sqrt{\left[\dfrac{\sum (x - \bar{x})^2}{n} \right]}$

where $\bar{x}$ is the mean set of values.

Standard deviation (discrete frequency distribution) $\sqrt{\left[\dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f} \right)^2 \right]}$

or $\sqrt{\left[\dfrac{\sum f(x - \bar{x})^2}{\sum f} \right]}$

Spearman's Rank Correlation Coefficient $1 - \dfrac{6 \sum d^2}{n(n^2 - 1)}$

1   The composite bar chart shows information about the area of land with forest and the area of land with no forest in some countries.

Source: fao.org
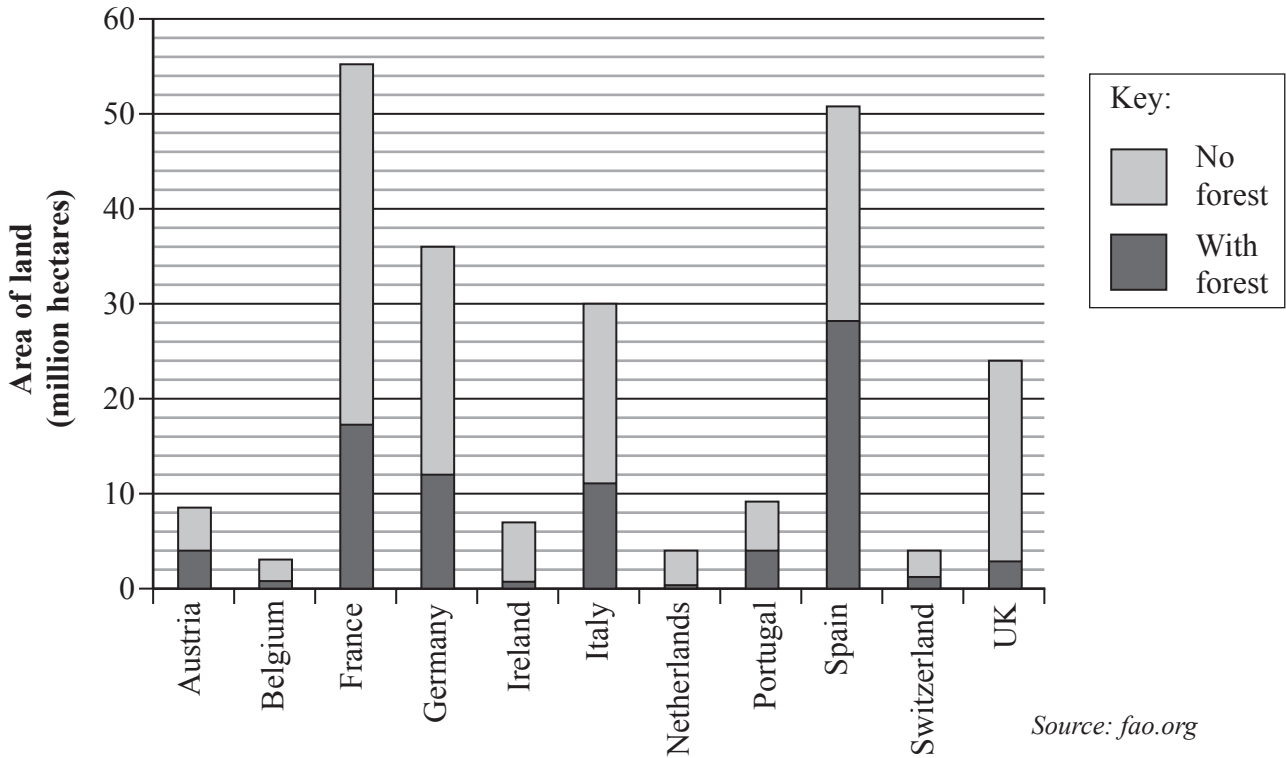
(a) Write down the country

(i) with the largest total area of land,

....

(ii) with the largest area of forest.

....

(2)

(b) Work out the area of land with no forest in the UK.

.... million hectares

(2)

(c) Estimate the area of land in Germany with forest as a percentage of the total area of land in Germany.

.... %

(2)

(Total for Question 1 is 6 marks)

2 The Bolton Boys are a basketball team.

Here is a list of the points scored by the team in 16 games this season.

| 97 | 74 | 89 | 81 | 78 | 85 | 102 | 100 |

| 76 | 103 | 87 | 101 | 76 | 82 | 98 | 95 |

(a) Complete an ordered stem and leaf diagram for these points.
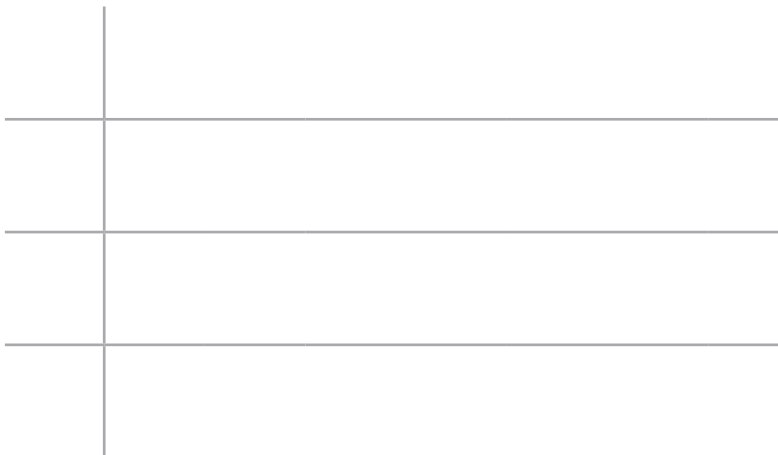
(3)

(b) Find the median.

.... points

(2)

(c) Work out the range.

.... points

(2)

The Durham Dribblers are also a basketball team.
They have a median score of 96 points and a range of 18 points.

(d) Compare the distributions of points scored by the Bolton Boys and the Durham Dribblers.

....

....

....

....

(2)

The Bolton Boys and Durham Dribblers each have a basketball game next week.

(e) Discuss which team is likely to score most points in their game.

....

....

....

....

(2)

(Total for Question 2 is 11 marks)

3 The table shows information about the weights (w grams) of 80 peaches.

Weight (w grams)	Frequency
$120 < w \leqslant 130$	7
$130 < w \leqslant 140$	13
$140 < w \leqslant 150$	18
$150 < w \leqslant 160$	28
$160 < w \leqslant 170$	14

(a) Draw a frequency polygon for this information.
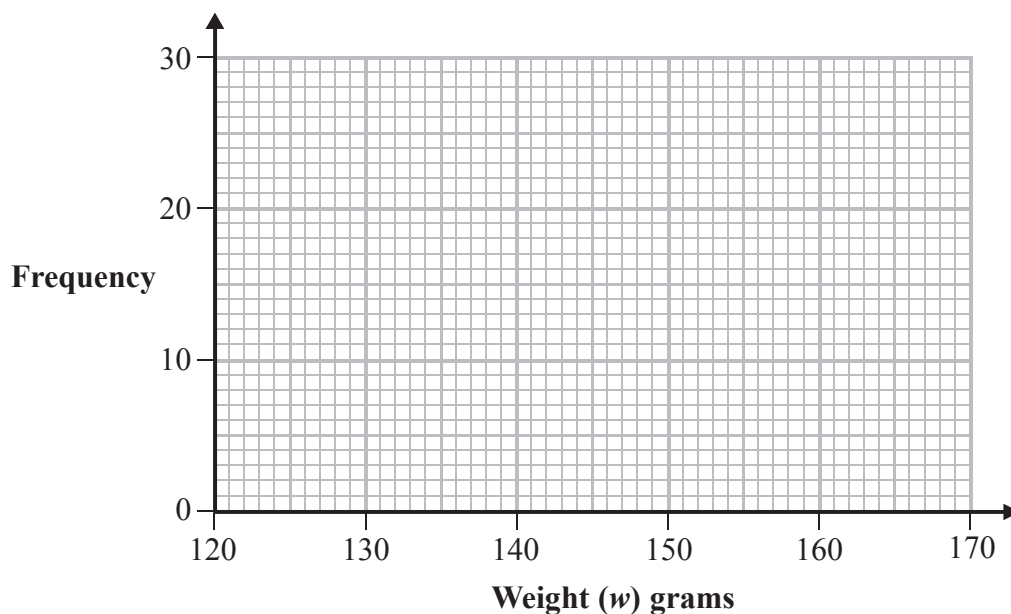
(2)

Marc says that the median weight of these peaches is in the modal class interval.

(b) (i) Write down the modal class interval.

....

(ii) Does the information support what Marc says?

You must explain your answer.

....

....

....

....

(2)

(Total for Question 3 is 4 marks)

4 The Daily Dispatch and The Morning Chronicle are two newspapers in a town.

Each newspaper conducted a survey in the town to see who the voters will vote for in a local election.

The population for each survey is all the voters in the town.

Complete the sentence below with the correct statistical term.

(a) A **list** of all voters in the town is known as a ....

(1)

Each newspaper conducted a random sample of all the voters in the town.

(b) Explain what is meant by the word 'random'.

....

....

(1)

One of the candidates is Mr Lopez.
The table shows some information about the results of each newspaper's survey.

	Number of people who will vote for Mr Lopez	Total number of people surveyed
The Daily Dispatch	54	120
The Morning Chronicle	38	80

There is a total of 1000 voters in the town.

The Daily Dispatch estimates that 450 of these voters will vote for Mr Lopez.
The Morning Chronicle estimates that 475 of these voters will vote for Mr Lopez.

(c) Discuss which newspaper's estimate should be more reliable.

....

....

....

(2)

(Total for Question 4 is 4 marks)

5 Estelle is the manager of a company with 50 employees.

She is going to give a questionnaire to each employee to get information about the food served in the company's canteen.

(a) Give one advantage of using a census rather than a sample.

....

....

(1)

Estelle will be collecting primary data.

(b) (i) Explain what is meant by primary data.

....

....

(ii) Give one advantage of using primary data.

....

....

(2)

Here is one of Estelle's hypotheses.

> The amount of money male employees spend on food in the canteen varies more than the amount of money female employees spend in the canteen.

(c) Write down the name of a statistic that Estelle could calculate to investigate this hypothesis.

....

(1)

*(d) Describe what makes a good question on a questionnaire.

....

....

....

....

....

....

....

....

....

(3)

Estelle may have problems analysing the data from her questionnaire.

(e) Give **two** of these problems.

....

....

....

....

....

(2)

(Total for Question 5 is 9 marks)

6 The cumulative frequency diagram shows some information about the heights, in metres, of a random sample of 50 of the tallest roller coasters in the world.
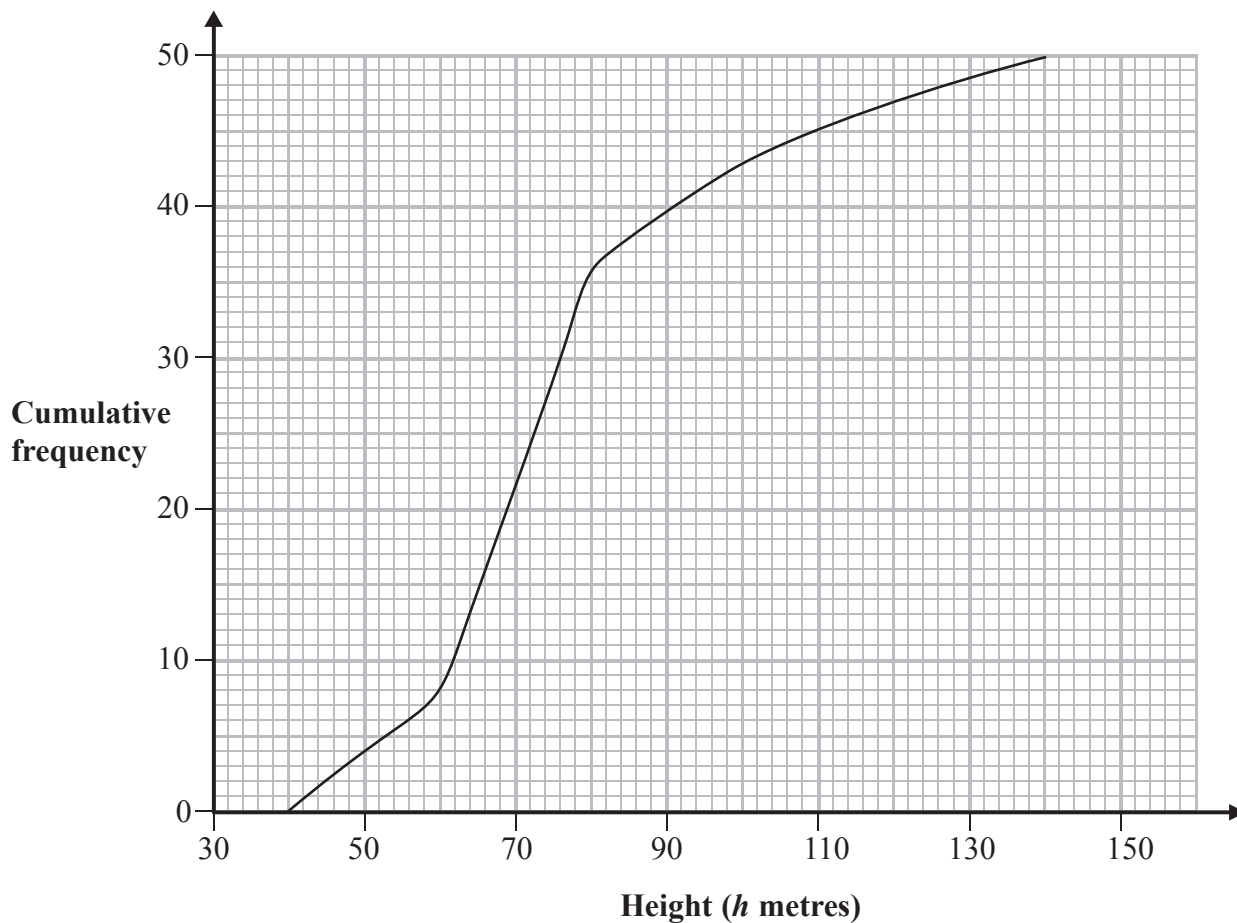
Source: Roller Coaster Database

(a) Write down the number of roller coasters with a height of 80 metres or less.

....

(1)

(b) Work out an estimate of the number of these roller coasters with a height between 60 metres and 110 metres.

....

(2)

A safety company is going to inspect roller coasters with a height greater than 86 metres.

(c) Calculate an estimate of the percentage of the 50 roller coasters that the safety company is going to inspect.

.... %

(3)

(Total for Question 6 is 6 marks)

7 Alexander measured the reaction times of Year 7 and Year 9 students.

The box plot shows information about the reaction times of the **Year 7** students.
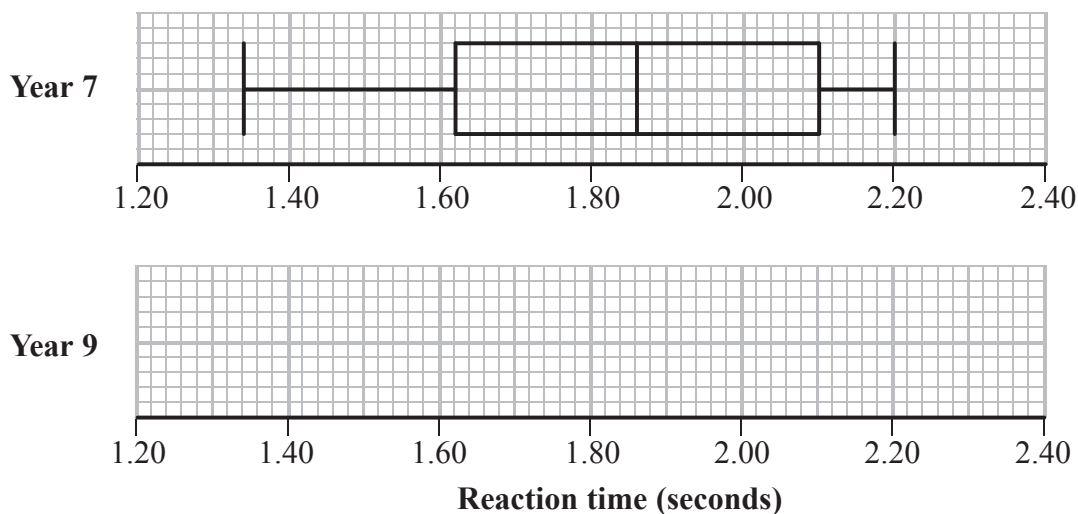
Year 7

1.20   1.40   1.60   1.80   2.00   2.20   2.40

Year 9

1.20   1.40   1.60   1.80   2.00   2.20   2.40

Reaction time (seconds)

The incomplete table shows some information about the reaction times of the **Year 9** students.

Minimum	Lower quartile	Median	Upper quartile	Maximum
1.40	1.68			2.26

The Year 7 and Year 9 distributions have the same median.

The Year 7 and Year 9 distributions have the same interquartile range.

(a) On the grid above, use this information to draw the box plot for the reaction times of the Year 9 students.

(3)

(b) Describe the skew of each distribution.

....

....

....

....

(2)

The box plot for Year 7 was based on data collected from 76 students in Year 7

(c) Estimate the number of these Year 7 students with a reaction time less than 2.10 seconds.

....

(2)

(Total for Question 7 is 7 marks)

8 Rudra is carrying out an investigation about the amount of money teenagers spend.

He is going to ask 50 teenagers some questions.

He asks teenagers going into a shopping centre until he has asked 25 boys and 25 girls.

(a) Write down the name of this sampling method.

....

(1)

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

Advantage ....

....

....

Disadvantage ....

....

....

(2)

Kate is also investigating the amount of money teenagers spend.
She asks students at a college.

Kate is going to take a sample stratified by gender.

*(c) Describe how she can select this sample.
Suggest how using a computer might help her.

....

....

....

....

....

....

....

(3)

(Total for Question 8 is 6 marks)

9 There were 8 divers in a diving competition.

Two judges ranked the 8 divers on their first dive.

Both Rosie and Stefan calculated the Spearman's rank correlation coefficient for the Judges' ranks.

Rosie got a value of 0.952
Stefan got a value of −1.25

Stefan did **not** calculate the value correctly.

(a) Explain how you know this.

....

....

(1)

Rosie's value is correct.

(b) Describe and interpret Rosie's value.

....

....

....

(2)

Each diver had a second dive.

Rosie also calculated the Spearman's rank correlation coefficient for the judges' ranks of the second dive.

She got a value of 0.643

(c) Compare the two correlation coefficients Rosie calculated.

....

....

....

(2)

(Total for Question 9 is 5 marks)

10 The table shows information about the average price of milk per year, in pence per litre, for the years 2008 to 2012

Year	2008	2009	2010	2011	2012
Price (pence per litre)	26	24	25	27	28

Source: Office for National Statistics

The incomplete table shows some of the chain base index numbers for these prices.

Year	2009	2010	2011	2012
Chain base index number	92	104		

(a) Interpret the value 92 in this table.

....

....

....

(2)

(b) Work out the chain base index numbers for the years 2011 and 2012
You must show your working.

2011 ....

2012 ....
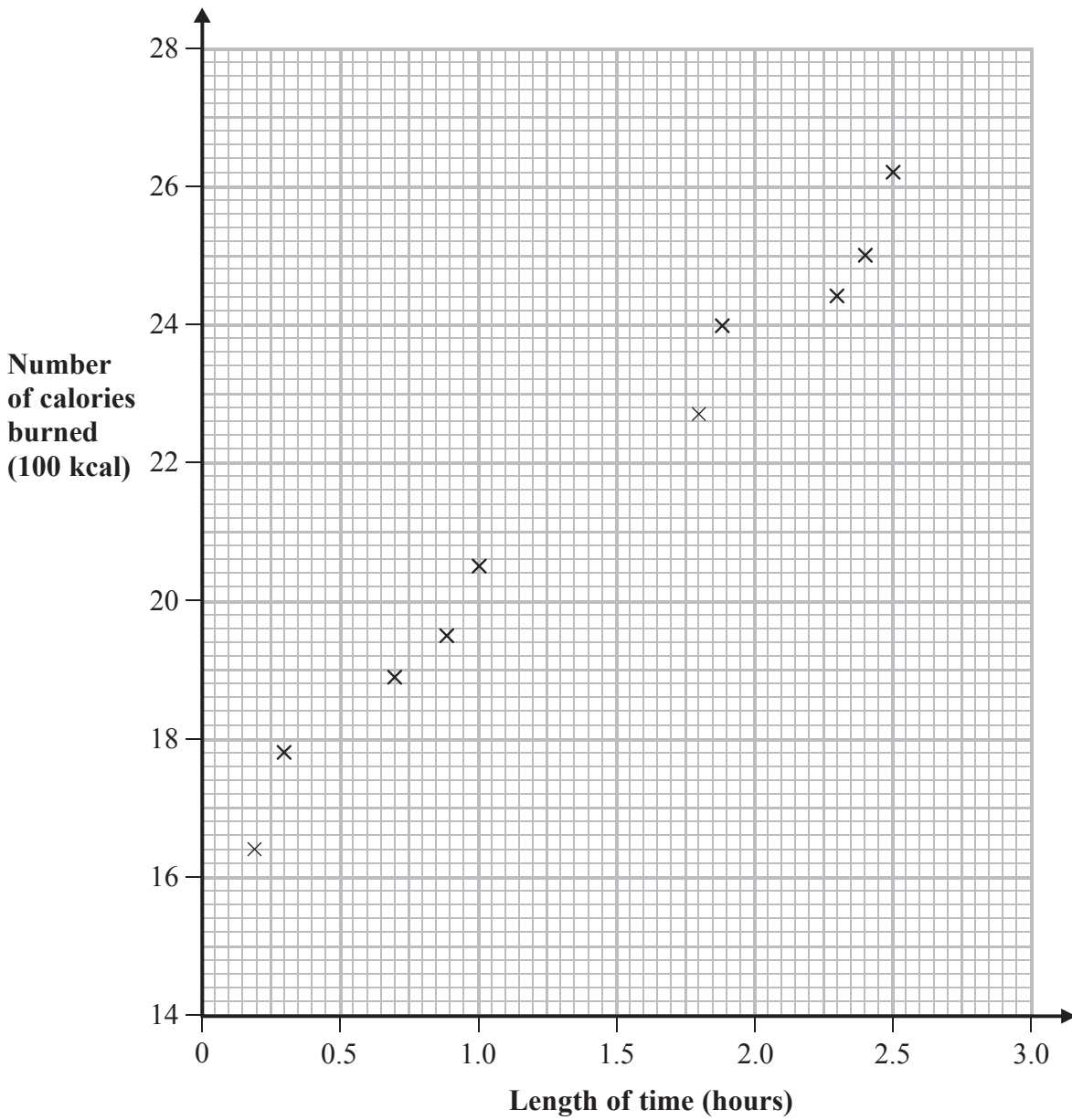
(3)

(Total for Question 10 is 5 marks)

BLANK PAGE

11 A sample of 10 people did an exercise for different lengths of time.

The scatter graph shows the number of calories burned, in 100 kcal, and the length of time, in hours, for each person.

The number of calories burned is the response variable.

(a) Explain what is meant by the term 'response variable'.

....

....

(1)

For this sample

the mean length of time is 1.4 hours,
the mean number of calories burned is 2150 kcal.

(b) On the scatter diagram

  (i) plot the mean point,

  (ii) draw a line of best fit.

(2)

(c) (i) Calculate the gradient of your line of best fit.

....

  (ii) Interpret the value of this gradient.

....

....

(3)

(Total for Question 11 is 6 marks)

12 The table shows information about the political parties and genders of the Members of Parliament in the UK in July 2013

Political party	Male	Female	Total
Liberal Democrat	48	7	55
Labour	171	86	257
Conservative	256	48	304
Other	28	6	34
Total	503	147	650

Source: www.parliament.uk

One Member of Parliament is chosen at random.

(a) Write down the probability that this Member of Parliament

(i) is female,

....

(ii) is male and is in the Labour party.

....

(2)

(b) Find the probability that this Member of Parliament

(i) is **not** in the Liberal Democrat party,

....

(ii) is female or is in the Conservative party.

....

(4)

The table below shows the number of males and the number of females in the parliaments of the UK and Argentina in July 2013

Country	Male	Female	Total
UK	503	147	650
Argentina	160	97	257

(c) Compare the proportions of females in the parliaments of the UK and Argentina.

....

....

....

....

....
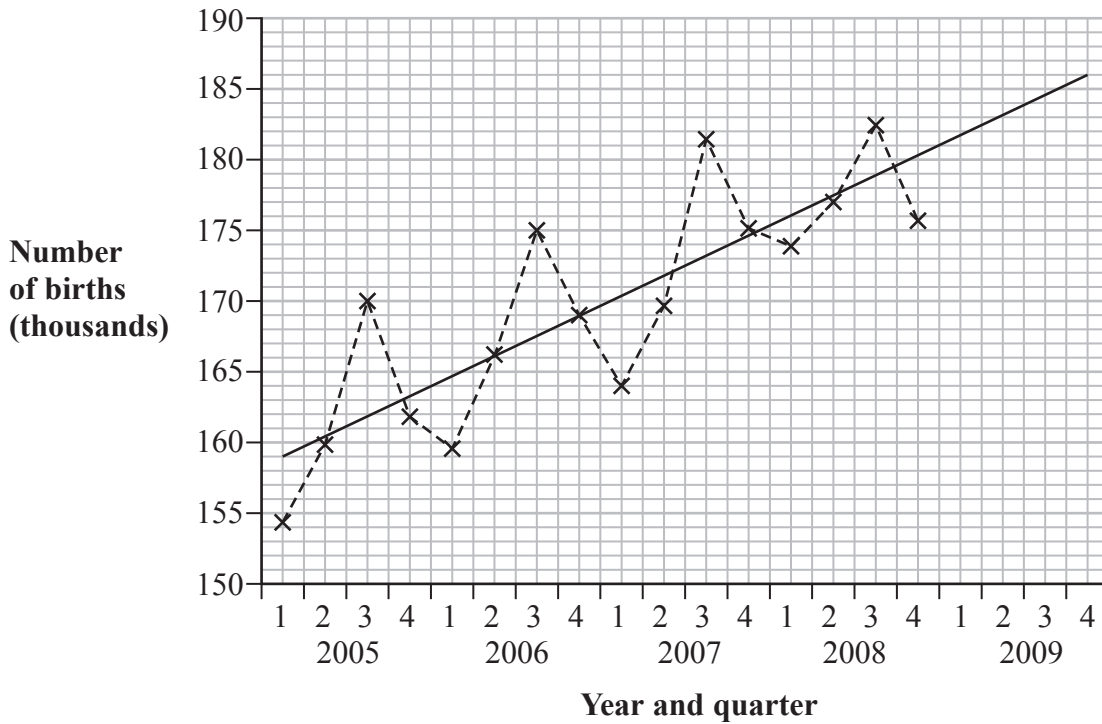
(2)

(Total for Question 12 is 8 marks)

13 The time series graph shows the number of births, in thousands, for each quarter in England and Wales for the years 2005 to 2008

Source: Office for National Statistics

(a) Write down the number of births for quarter 3, 2005

.... thousand

(1)

A trend line has been drawn on the graph.

(b) Work out the seasonal variation for quarter 3, 2005

.... thousand

(1)

Emily wants to predict the number of births for quarter 3, 2009 using information from the time series graph.

(c) Explain how Emily can do this.

....

....

....

....

....

....

(2)

(d) Discuss the reliability of this prediction.

Give a reason for your answer.

....

....

....

(1)

The actual number of births for quarter 3, 2009 was 181.3 thousand.

(e) Does this value fit with the trend in the number of births?
Explain your answer.

....

....

....

(1)

(Total for Question 13 is 6 marks)

14 The Norway women's handball team played 8 games at the 2012 Olympics.

Here are the number of points, x, they scored in each game.

$$23 \quad 27 \quad 24 \quad 24 \quad 20 \quad 21 \quad 31 \quad 26$$

Source: www.olympic.org

(a) Show that $\sum x^2 = 4888$

(1)

(b) Find the standard deviation of the points scored.
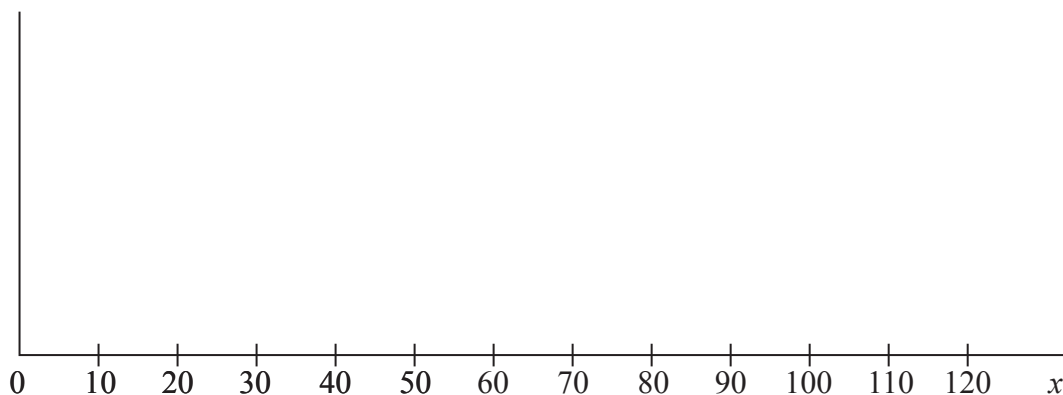   Give your answer to 1 decimal place.

(3)

(Total for Question 14 is 4 marks)

15 A variable, X, has a normal distribution with mean 60

95% of the values of X lie between 45 and 75

(a) Sketch a diagram to show this distribution.

```
|
|
|
|
|
|
|
|
|
|___|___|___|___|___|___|___|___|___|___|___|___|___|___
0   10  20  30  40  50  60  70  80  90  100 110 120   x
```

(3)

Ross wants to find out how long it takes boys and girls in different year groups to complete jigsaw puzzles.

He collects the following information for each child.

Gender
Year group
Number of pieces in the jigsaw puzzle
Time taken to complete the puzzle
Favourite subject

(b) Which of these variables is most likely to be modelled by a normal distribution? Give a reason for your answer.

....

....

....

(2)

(Total for Question 15 is 5 marks)

16 Julia and Alan are planting some seeds.
Each seed will either produce a flower or not produce a flower.

Each of these seeds is three times as likely to produce a flower as not to produce a flower.

(a) Show that the probability that a particular seed will produce a flower is $\dfrac{3}{4}$

(1)

Julia plants two seeds.

(b) Work out the probability that exactly one of the two seeds will produce a flower.

....

(2)

Alan plants four seeds.

(c) (i) Name the probability distribution that may be used to model the number of seeds that will produce flowers.

....

(ii) Write down one property needed so that this distribution will be a suitable model.

....

....

(2)

(d) Work out the probability that at least 3 of these 4 seeds will produce flowers.

You may use $(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

....

(3)

(Total for Question 16 is 8 marks)

TOTAL FOR PAPER IS 100 MARKS

BLANK PAGE