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 = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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.

.....

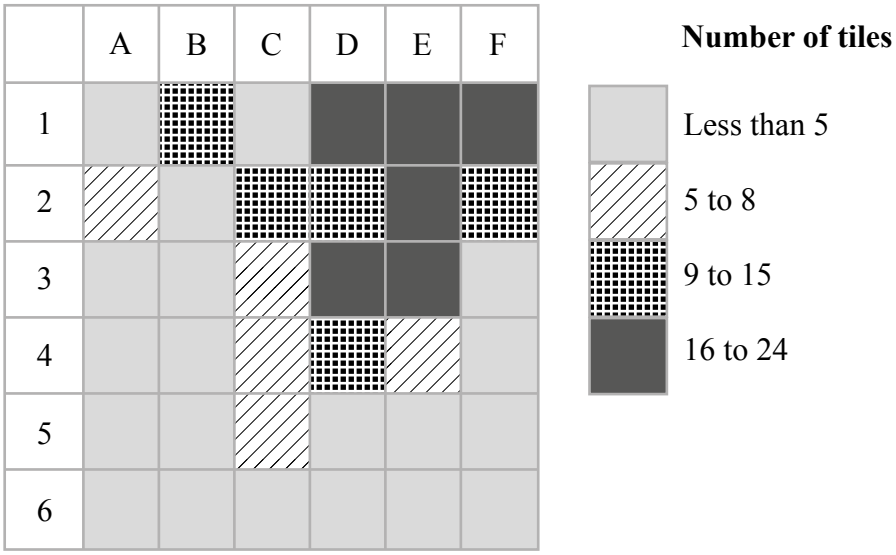
(1)

(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)

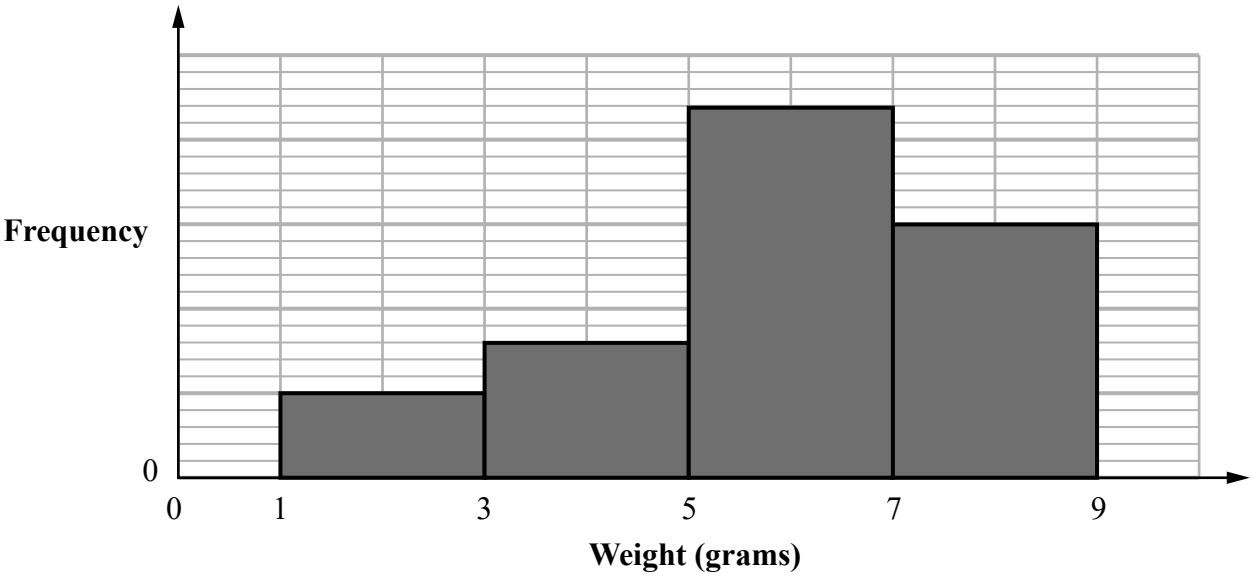
3 Tomoyo found the weight, in grams, of each of 100 cherries.

(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.

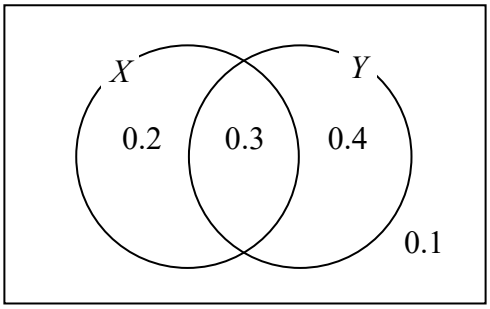
..... g

(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)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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.

.....

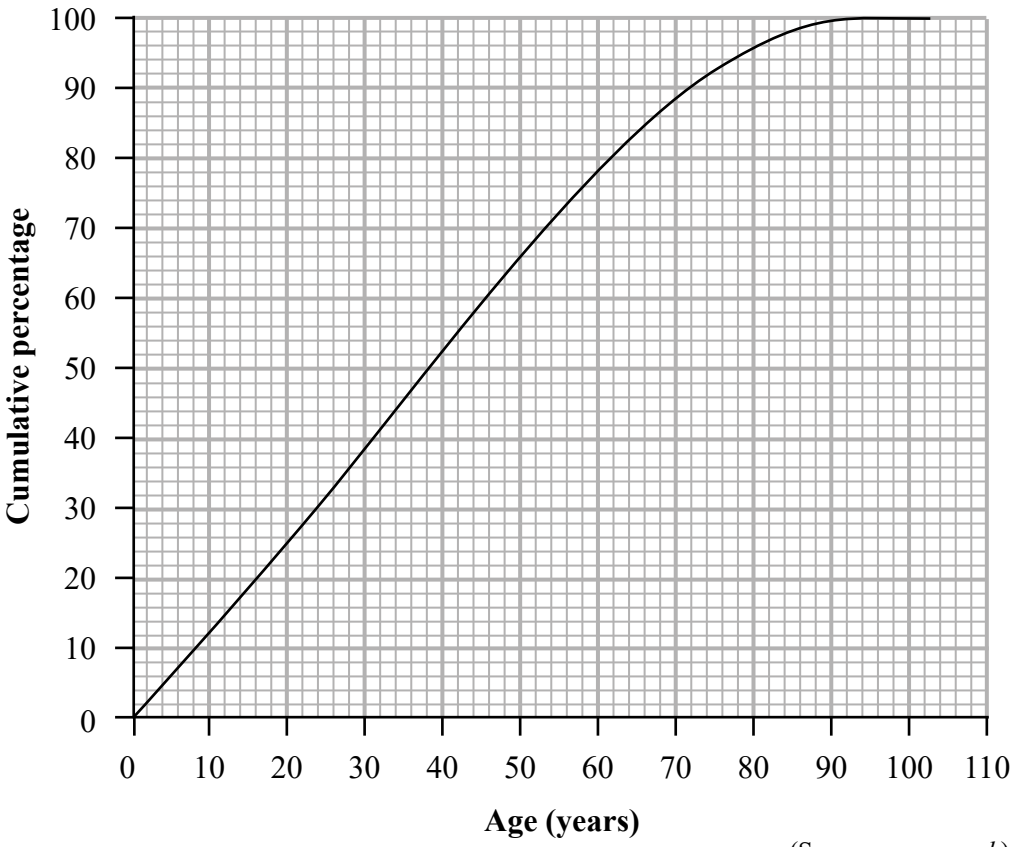
.....

.....

(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.

.....

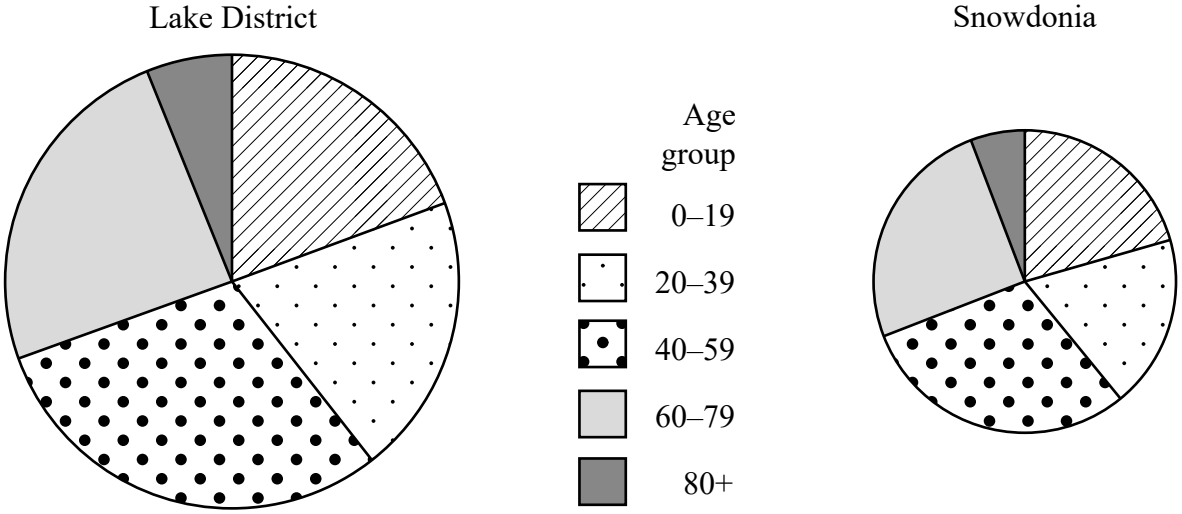
(2)

(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.

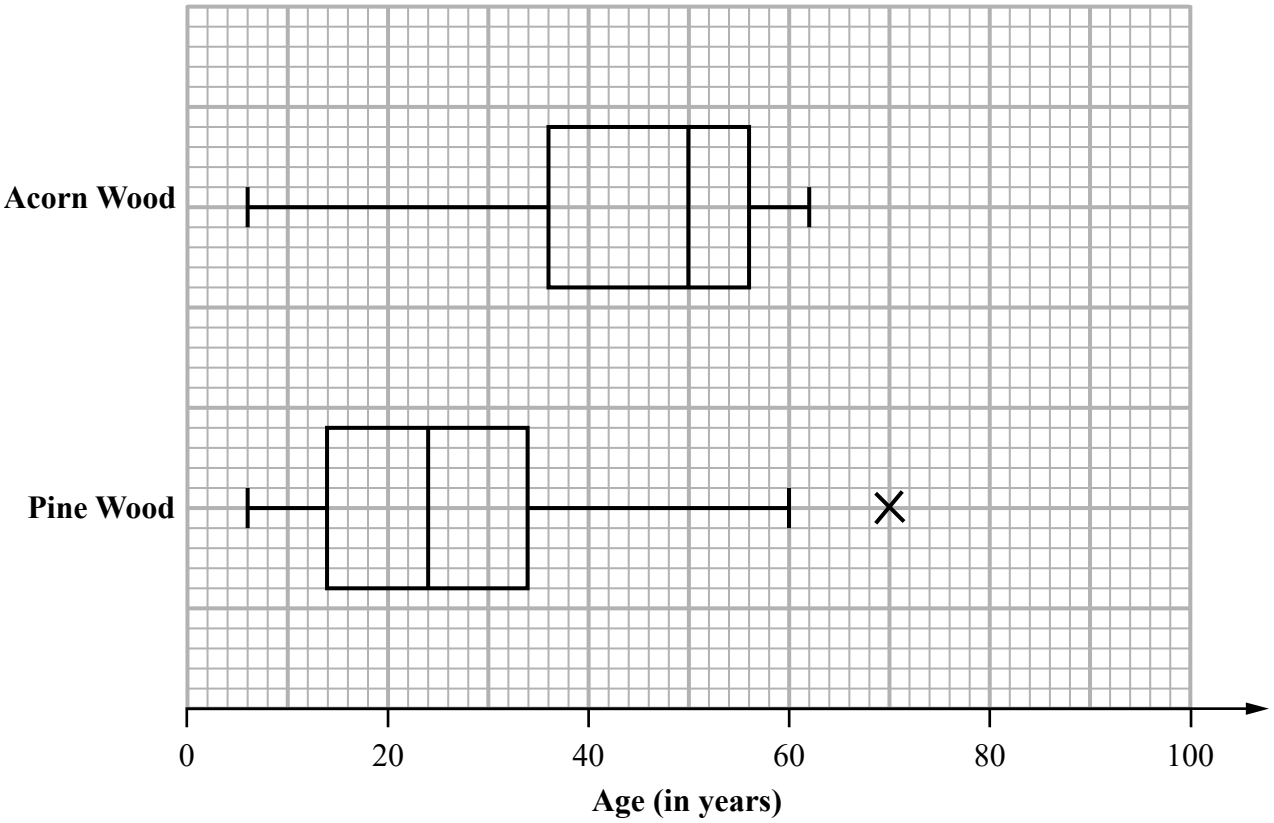
(d) Show that the radius of the pie chart for Exmoor is 1.52 cm, correct to 2 decimal places.

..... 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)

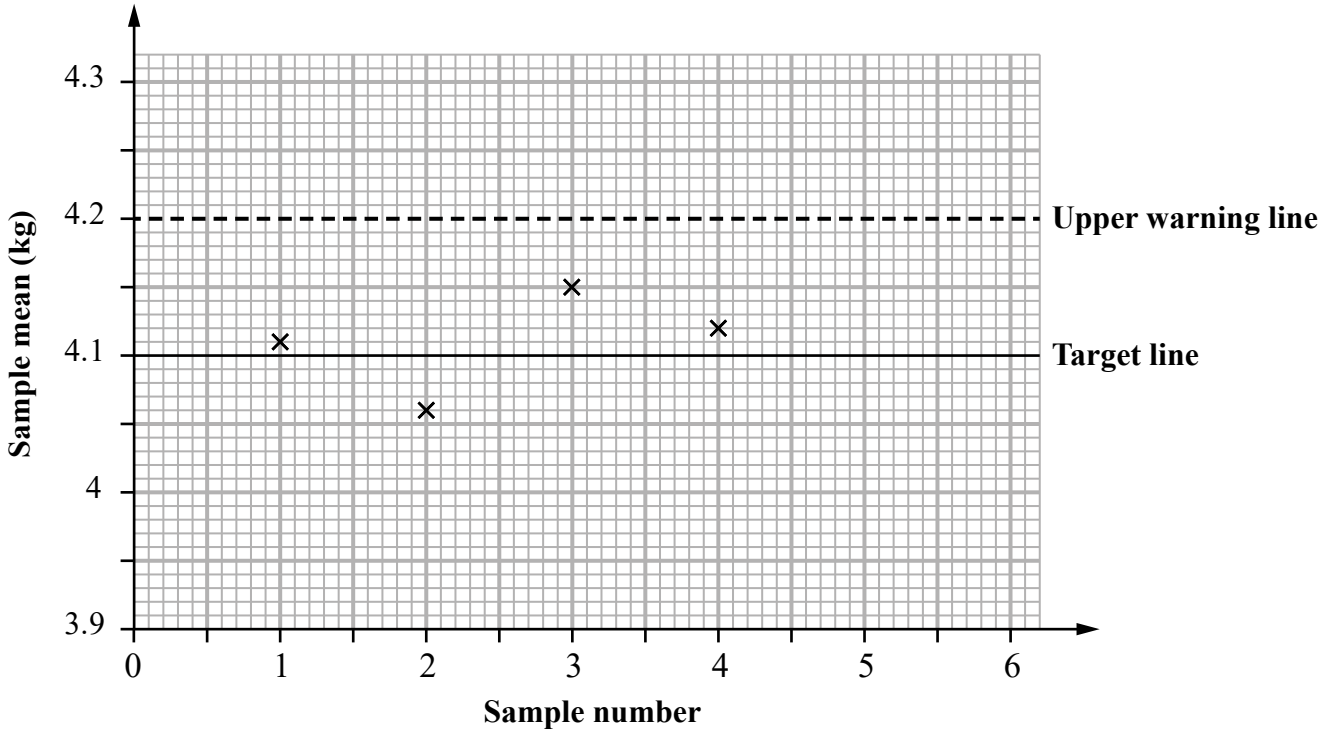
12 Glowbright Garden Products produce bags of charcoal for use on barbecues.

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.

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

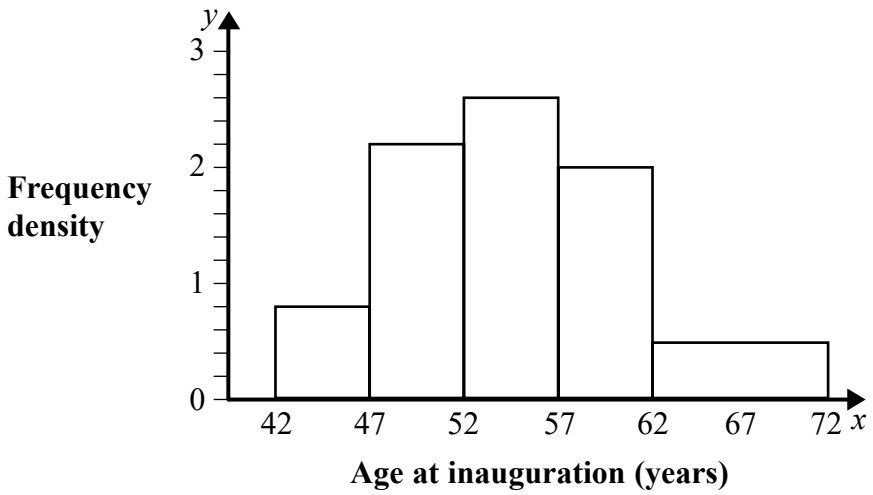
where the values of x are the class midpoints.

(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.

.....

.....

.....

.....

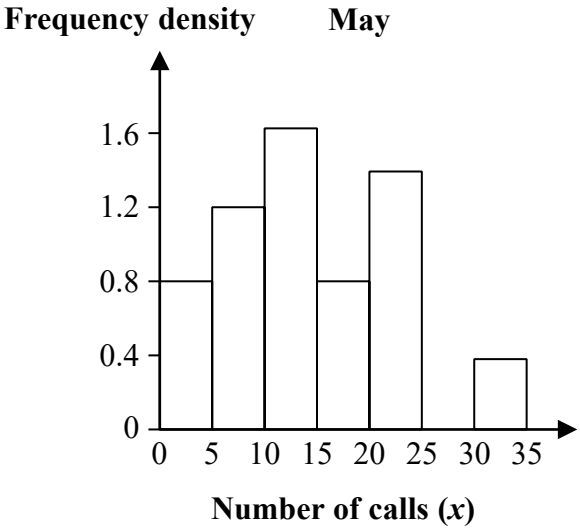
.....

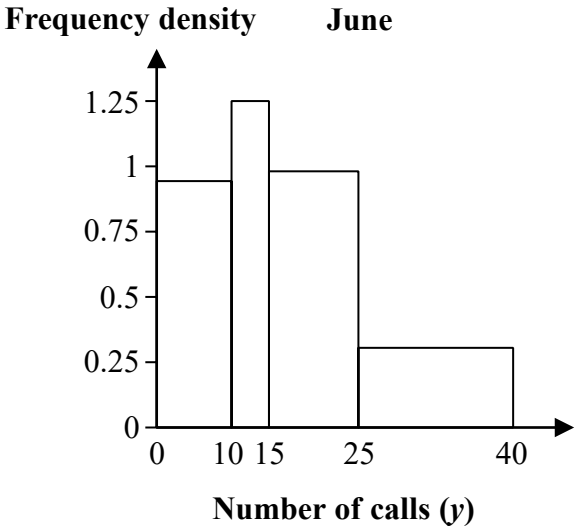
(Total for Question 14 is 2 marks)

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