| Centre Number | | | | | | Candidate Number | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Surname | | | | | | | | | | |
| Other Names | | | | | | | | | | |
| Candidate Signature | | | | | | | | | | |

| For Examiner's Use | |
|---|---|
| Examiner's Initials | |
| Pages | Mark |
| 3 | |
| 4 – 5 | |
| 6 – 7 | |
| 8 – 9 | |
| 10 – 11 | |
| 12 – 13 | |
| 14 – 15 | |
| 16 – 17 | |
| 18 – 19 | |
| 20 – 21 | |
| 22 – 23 | |
| 24 – 25 | |
| 26 – 27 | |
| TOTAL | |

AQA

General Certificate of Secondary Education
Higher Tier
June 2015

# Statistics

**43101H**

**Unit 1 Written Paper**

**H**

Thursday 18 June 2015 1.30 pm to 3.30 pm

**For this paper you must have:**
- a calculator
- mathematical instruments.

**Time allowed**
- 2 hours

**Instructions**
- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Do not write outside the box around each page or on blank pages.
- Do all rough work in this book. Cross through any work that you do not want to be marked.

**Information**
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 100.
- You may ask for more answer paper, graph paper and tracing paper. These must be tagged securely to this answer book.
- You are expected to use a calculator where appropriate.

**Advice**
- In all calculations, show clearly how you work out your answer.

You may need to use the following formulae:

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.

Standard deviation for a set of numbers $x_1, x_2, \ldots, x_n$ having a mean value of $\bar{x}$ is given by

$$\sqrt{\frac{\sum (x-\bar{x})^2}{n}} \quad \text{or} \quad \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Standard deviation for a frequency distribution

$$\sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}} \quad \text{or} \quad \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$

The same formulae apply to the standard deviation of a grouped frequency distribution where $x$ is the mid-interval value.

Spearman's rank correlation coefficient = $1 - \dfrac{6\sum d^2}{n(n^2-1)}$

Answer **all** questions in the spaces provided.

**1** Amy lives near a new airport.

She wants to investigate how the new airport has changed the house prices in the area where she lives.

She lists, in order, the five tasks that she plans to do to investigate this.

Two of the tasks are missing.

Fill in the missing tasks so that Amy has a suitable strategy for her investigation.

**[2 marks]**

First task Decide on a hypothesis

Second task .....

.....

Third task Collect the data

Fourth task Draw diagrams and make some calculations

Last task .....

.....

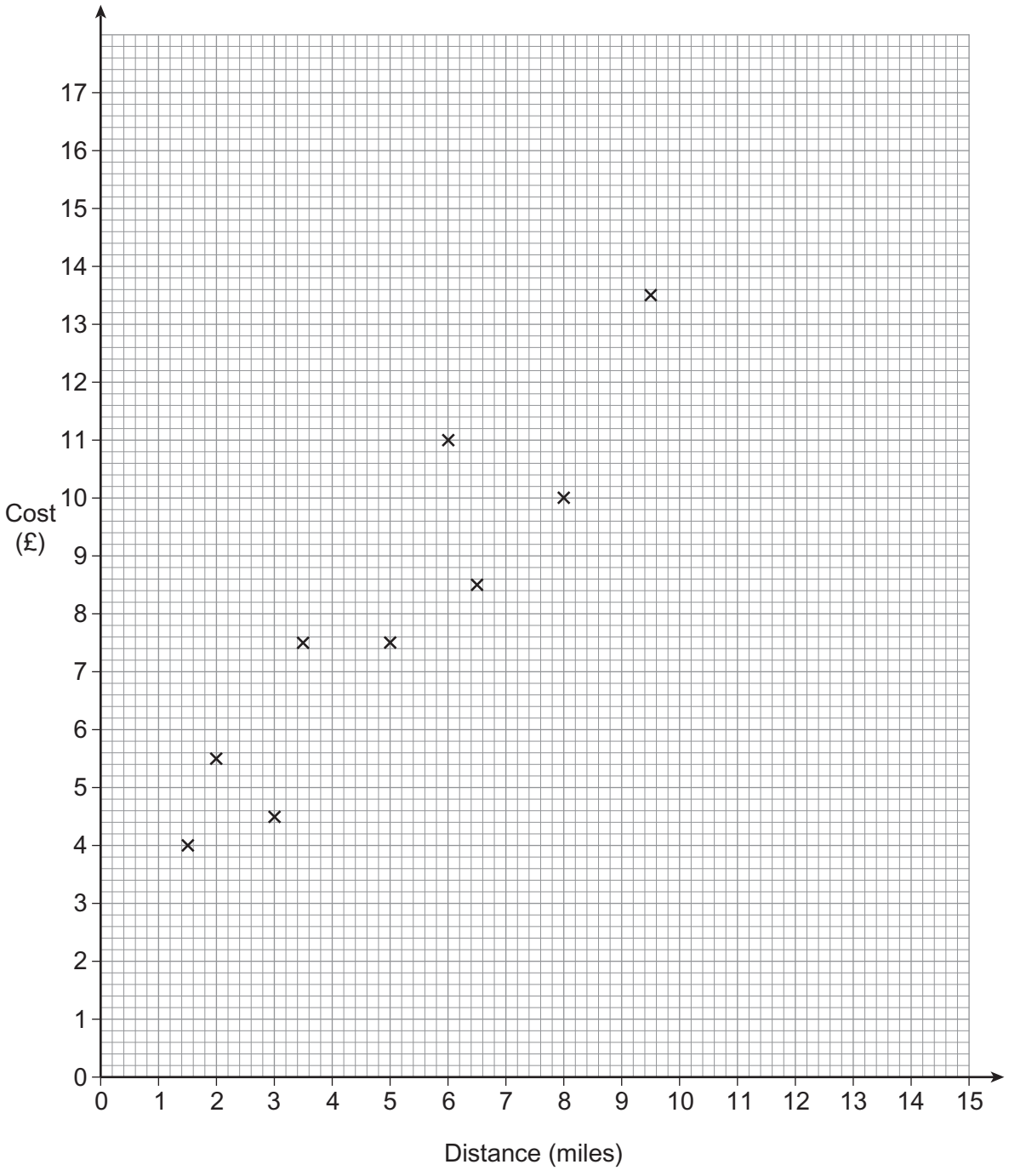
**Turn over for the next question**

2

**2** The table shows the distance travelled (miles) and the cost (£) of nine taxi journeys.

| **Distance (miles)** | 1.5 | 2 | 3 | 3.5 | 5 | 6 | 6.5 | 8 | 9.5 |
|---|---|---|---|---|---|---|---|---|---|
| **Cost (£)** | 4.00 | 5.50 | 4.50 | 7.50 | 7.50 | 11.00 | 8.50 | 10.00 | 13.50 |

The data are shown on the scatter diagram.



**2 (a)** Circle the most likely value of Spearman's rank correlation coefficient for the data.

**[1 mark]**

–0.87 0.08 0.93 8.45

**2 (b)** The mean distance travelled for these nine journeys is 5 miles and the mean cost is £8

Use these mean values to draw a line of best fit on the scatter diagram.

**[2 marks]**

**2 (c)** Use your line of best fit to estimate the cost of a 7 mile taxi journey.

**[1 mark]**

Answer £ .....

**2 (d)** Jack paid £15 for his taxi journey.

Use your line of best fit to estimate the distance he travelled.

**[1 mark]**

Answer ..... miles

**2 (e)** Which of the answers, 2(c) or 2(d), do you think is **more** reliable?

Tick a box.

**[1 mark]**

2(c) ☐ 2(d) ☐

Give a reason for your answer.

.....

.....

6

**3** Eight dancers take part in a competition.
Each dancer performs a dance which is marked by two judges.
The table shows the scores (out of 10) the judges gave to each dancer.

| | **Alex** | **Nina** | **Tanya** | **Rachel** | **Sam** | **Cruz** | **Jess** | **Mira** |
|---|---|---|---|---|---|---|---|---|
| **Judge A** | 6 | 5 | 10 | 3 | 9 | 3 | 7 | 7 |
| **Judge B** | 7 | 7 | 9 | 5 | 7 | 6 | 7 | 6 |

**3 (a)** Which judge gave higher marks on average?
Tick a box.

**[3 marks]**

Judge A ☐ Judge B ☐

You **must** show your working.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**3 (b)** How do you know there was some inter-observer bias?

**[1 mark]**

.....

.....

**3 (c)** Which dancer suffered the most from inter-observer bias?

**[1 mark]**

Answer .....

**3 (d)** Give **one** way of reducing or avoiding inter-observer bias in this situation.

**[1 mark]**

.....

.....

.....

.....

**Turn over for the next question**

6

**4** People who use mobile phones can choose different contracts for their phones. Each contract includes a number of free minutes of phone calls each month.

Sarah wants to find out if the number of minutes included in the contracts chosen by women is the same as the number of minutes included in the contracts chosen by men.

**4 (a)** Write a suitable hypothesis for Sarah to test.

**[1 mark]**

.....

.....

**4 (b)** Sarah decides to stand outside a mobile phone shop one lunchtime and survey everyone who goes inside the shop.

Describe **one** problem with this data selection method.

**[1 mark]**

.....

.....

**4 (c) (i)** Sarah collects information from mobile phone users with monthly contracts. She asks these users the following question:

How many free minutes are included in your monthly contract?

| | | | |
|---|---|---|---|
| 100 – 200 | ☐ | 200 – 400 | ☐ |
| 500+ | ☐ | Unlimited | ☐ |

Write down two criticisms of the **response** section.

**[2 marks]**

Criticism 1 .....

.....

Criticism 2 .....

.....

**4 (c) (ii)** For these mobile phone users, explain why it may not be possible to calculate an estimate of the mean number of free minutes per month.

**[1 mark]**

.....

.....

Sarah also wants to investigate how much people pay each month for their mobile phone contracts.

**4 (d)** Write a suitable question that Sarah can use.
You should include a response section.

**[3 marks]**

**4 (e)** Write down an extraneous variable, other than the number of free minutes, that could affect the cost of a mobile phone contract.

**[1 mark]**

.....

.....

9

**5** The table shows information about types of residential property in different regions in England.

**Type of properties in England by region**

| | House or Bungalow | | | Flat | | |
|---|---|---|---|---|---|---|
| | **Detached (%)** | **Semi-detached (%)** | **Terrace (%)** | **Purpose-built flats (%)** | **Converted flat (%)** | **All properties (thousands) (= 100%)** |
| **England** | **22.5** | **29.2** | **29.9** | **14.9** | **3.5** | **22 399** |
| North East | 12.6 | 37.3 | 34.6 | 13.8 | 1.7 | 1 152 |
| North West | 18.0 | 35.0 | 35.5 | 9.8 | 1.7 | 3 076 |
| Yorkshire | 20.9 | 34.4 | 32.3 | 9.8 | 2.6 | 2 258 |
| East Midlands | 33.3 | 35.1 | 23.1 | 7.1 | 1.4 | 1 923 |
| West Midlands | 25.3 | 32.1 | 30.0 | 11.2 | 1.4 | 2 325 |
| East | 28.3 | 29.6 | 28.3 | 12.1 | 1.7 | 2 466 |
| London | 4.5 | 15.1 | 32.3 | 38.7 | 9.4 | 3 248 |
| South East | 28.8 | 28.0 | 25.4 | 13.4 | 4.4 | 3 606 |
| South West | 31.9 | 24.8 | 28.1 | 10.7 | 4.5 | 2 345 |

Source: Adapted from 2008 English Housing Survey, Department for Communities and Local Government

**5 (a)** The data in this table was obtained from the Internet.

What is the name for data that has been collected by someone else?
Circle your answer.

**[1 mark]**

Secondary data  Primary data  Raw data  Grouped data

**5 (b)** Which region has the greatest proportion of **Detached** properties?

**[1 mark]**

Answer .....

**5 (c)** Which type of property is most common in England as a whole?
Tick a box.

**[1 mark]**

Detached ☐  Purpose-built flat ☐

Semi-detached ☐  Converted flat ☐

Terrace ☐

**5 (d)** Work out the total percentage of properties in the **South East** region that are Houses or Bungalows.

**[2 marks]**

.....

.....

Answer ..... %

**5 (e)** Write down one difference and one similarity between the types of properties in London and the North West region.

**[2 marks]**

Difference .....

.....

Similarity .....

.....

**5 (f)** Show that there is a smaller **number** of Semi-detached properties in the North East region than in the South West region.

**[3 marks]**

.....

.....

.....

.....

.....

.....

.....

.....

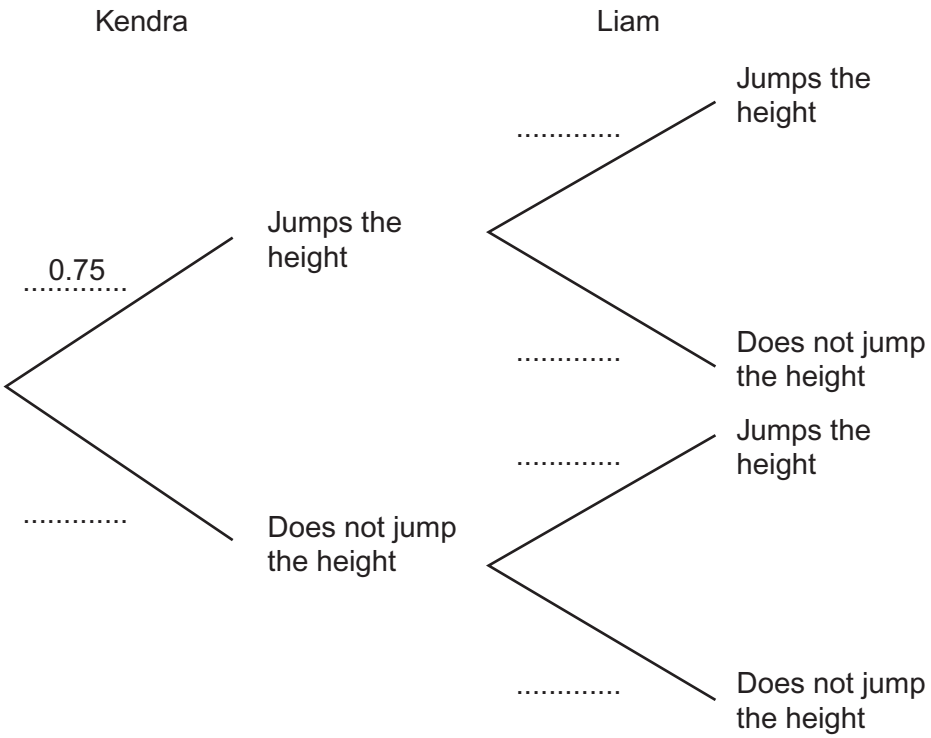
10

**6** Kendra and Liam are in a high jump competition.
They each have one chance to jump a height of 170 cm

The probability that Kendra can jump this height is 0.75
The probability that Liam can jump the height is 0.8

**6 (a)** Complete the tree diagram.
Assume that the jumps of Kendra and Liam are independent.

**[2 marks]**



**6 (b)** Work out the probability that **both** Kendra and Liam jump the height.

**[2 marks]**

.....

.....

.....

Answer .....

**6 (c)** Work out the probability that at least one of them jumps the height.

**[3 marks]**

.....

.....

.....

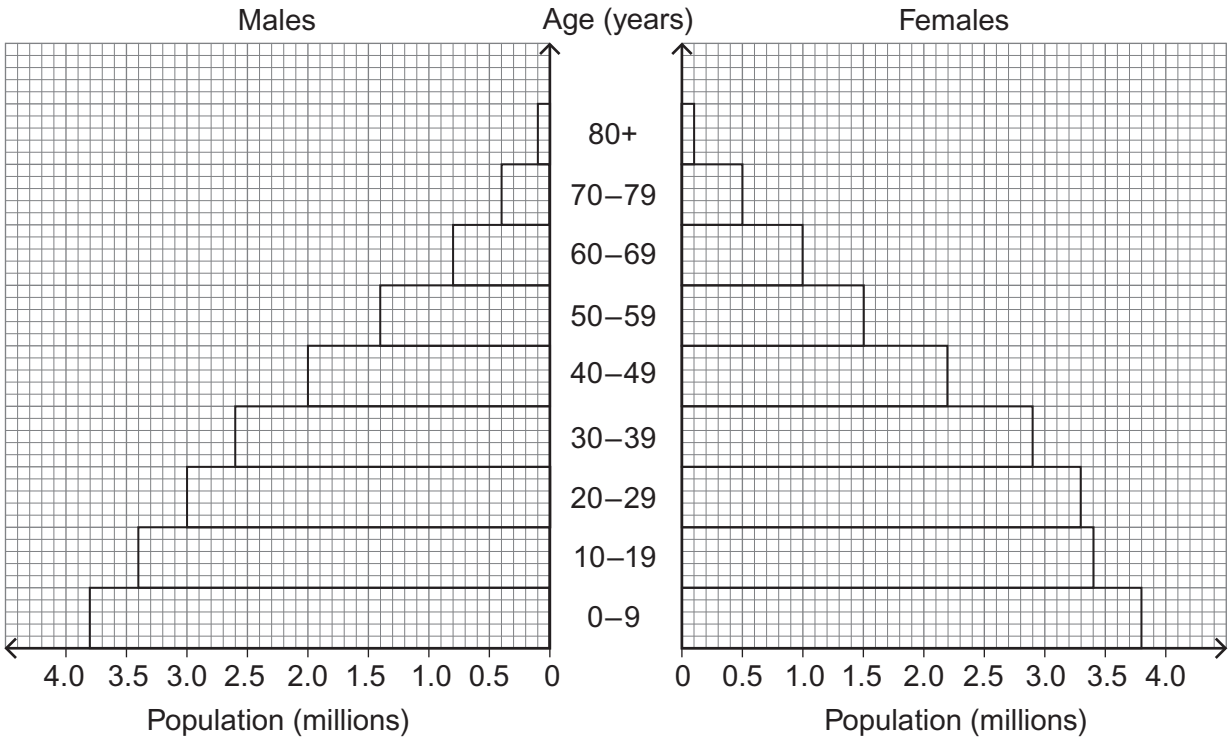
Answer .....

**Turn over for the next question**

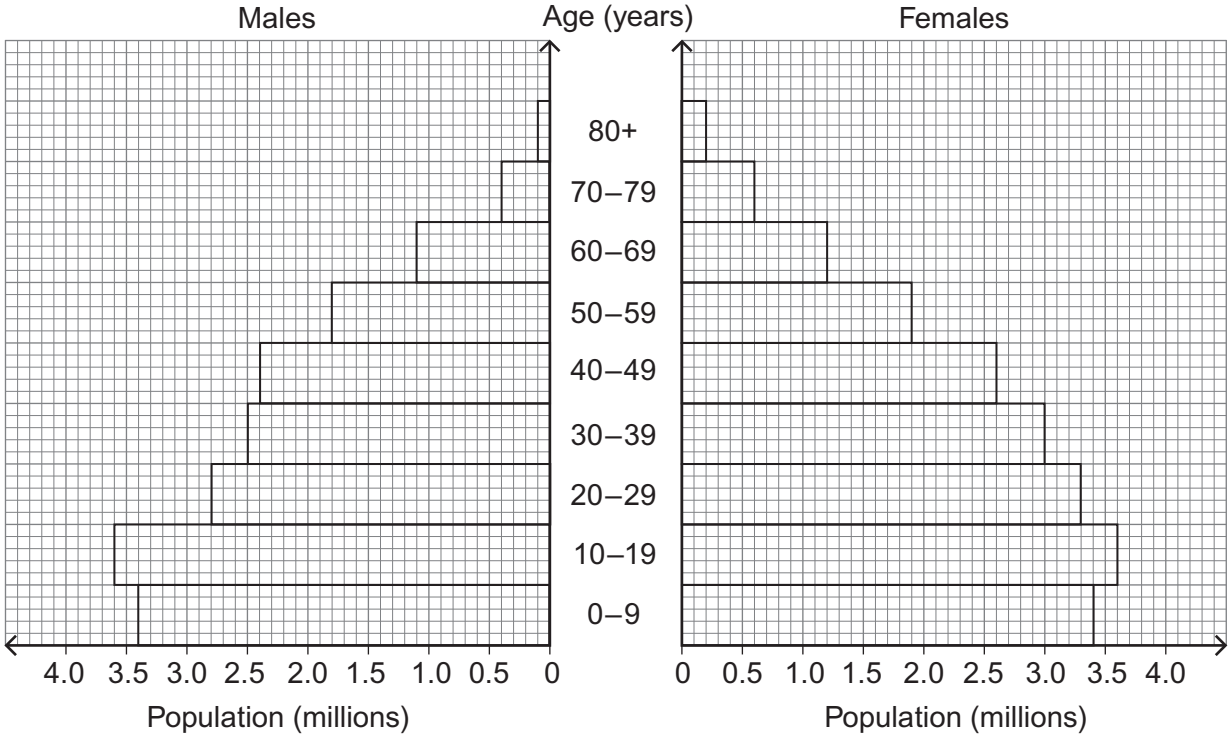
7

**7** The First World War began in 1914 and finished in 1918
The population pyramids show the number of people in England and Wales in different age groups.
The data were recorded in censuses carried out before and after the war.

**1911**




**1921**




Source: www.neighbourhood.statistics.gov.uk

**7 (a)** Write down the number of males aged 50–59 years in **1911**

**[1 mark]**

Answer ..... million

**7 (b)** Complete this sentence.

**[1 mark]**

In 1921, there were .....% **more** females aged 70–79 years than males aged 70–79

**7 (c)** Fewer babies were born in the years during the First World War.

Work out how many fewer children aged 0–9 years there were in 1921 compared to 1911

**[2 marks]**

.....

.....

Answer ..... million

**7 (d)** Explain how you can tell from the population pyramids that a large number of men died during the First World War.

**[1 mark]**

.....

.....

.....

5

**8** Matt owns a lorry company.
He regularly sends lorries from his depot to Dover.
He records the amount of fuel used (litres) for a sample of 120 of these journeys.

His results are shown in the table.

| **Amount of fuel used $x$ (litres)** | **Frequency** |
|---|---|
| $40 < x \leqslant 50$ | 8 |
| $50 < x \leqslant 60$ | 22 |
| $60 < x \leqslant 70$ | 50 |
| $70 < x \leqslant 80$ | 26 |
| $80 < x \leqslant 90$ | 10 |
| $90 < x \leqslant 100$ | 4 |

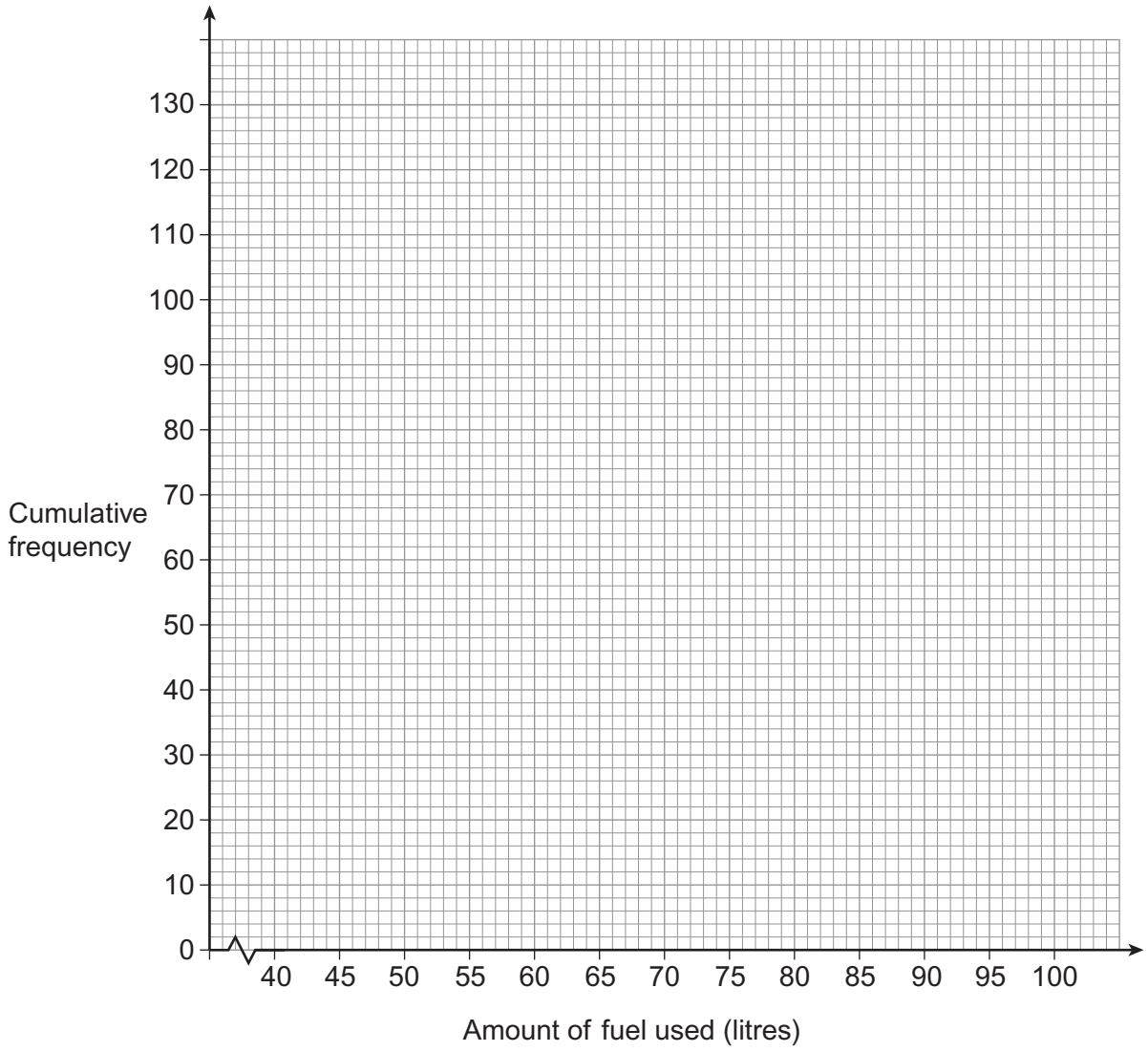
| **Cumulative Frequency** |
|---|
| 8 |
| 30 |
| |
| |
| |
| |

**8 (a)** Complete the cumulative frequency column above.

**[1 mark]**

**8 (b)** Draw a cumulative frequency graph to show the data.

**[3 marks]**



**8 (c)** The company has a target that at least three-quarters of the trips to Dover should use 75 litres of fuel or less.

Do the data suggest that the company's target is being met?
Tick a box.

Target met ☐ Target not met ☐

Show how you worked out your answer.

**[2 marks]**

.....

.....

.....

.....

**8 (d)** Work out the interpercentile range between the 10th and 90th percentiles.

**[3 marks]**

.....

.....

.....

.....

Answer ..... litres

**8 (e)** Write down **one** advantage of using the interpercentile range as a measure of spread rather than the range.

**[1 mark]**

.....

.....

.....

**10**

Matt thinks that the amount of fuel his lorries use might be affected by how fast they travel.
He plans a small experiment to test this.

He records how much fuel one of his lorries uses on different trips to Dover.
He sets a different maximum speed for each trip.
He makes sure that the weight of the cargo for each trip is the same.

**8 (f)** Explain why Matt makes sure that the weight of the cargo is the same each time.

**[1 mark]**

.....

.....

**8 (g)** Write down the response variable.

**[1 mark]**

.....

Matt's results are shown in the table.

| **Maximum speed (miles per hour)** | 60 | 59 | 58 | 57 | 56 | 55 | 54 | 53 | 52 |
|---|---|---|---|---|---|---|---|---|---|
| **Fuel used (litres)** | 68 | 69 | 64 | 61 | 63 | 59 | 62 | 56 | 58 |

**8 (h) (i)** Work out the value of Spearman's rank correlation coefficient for these data.

You are given that $\sum d^2 = 14$

**[3 marks]**

.....

.....

.....

Answer .....

**8 (h) (ii)** Explain, in context, the nature of the relationship between the maximum speed and the fuel used.

**[1 mark]**

.....

.....

6

**There are no questions printed on this page**

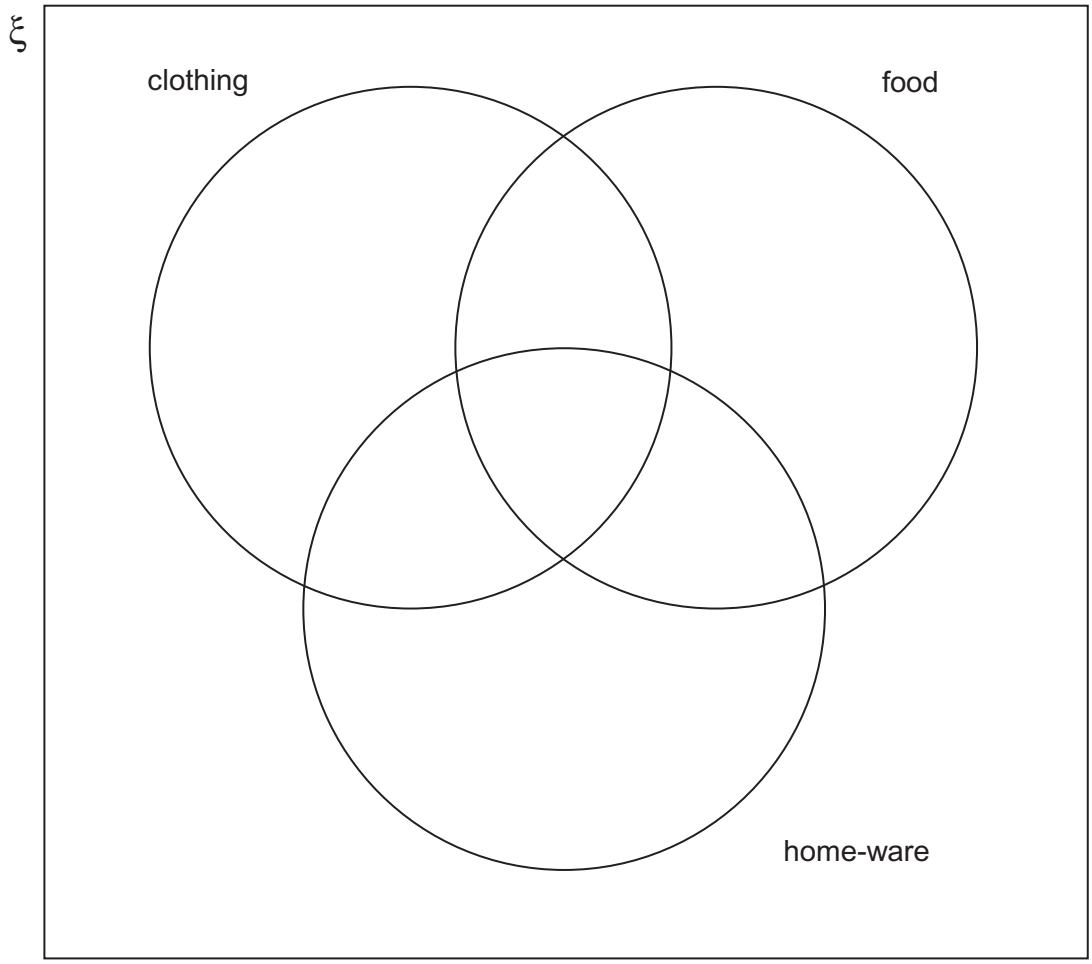
**Turn over for the next question**

**9** A department store manager surveys 120 people who visit the store.

- 9 buy clothing, food and home-ware
- 5 buy food and home-ware but **not** clothing
- 13 buy clothing and home-ware but **not** food
- 89 buy clothing, 55 buy food and 31 buy home-ware
- 40 buy clothing and food

**9 (a)** Use the data to help you complete the Venn diagram.

**[4 marks]**




**9 (b)** How many people did **not** buy clothing or food or home-ware?

**[1 mark]**

.....

Answer .....

**9 (c)** One of these 120 people is chosen at random.

Find the probability that the person bought

**9 (c) (i)** food **and** home-ware,

**[2 marks]**

.....

Answer .....

**9 (c) (ii)** clothing **or** home-ware,

**[2 marks]**

.....

Answer .....

**9 (c) (iii)** home-ware **given** that they bought clothing.

**[2 marks]**

.....

Answer .....

**Turn over for the next question**

**11**

**10** The personnel department at a company wants to find out the views of staff on certain issues.
It decides to carry out a survey on a sample of staff.

**10 (a)** Give **one** advantage of doing the survey on a sample of staff rather than doing a census.

**[1 mark]**

.....

.....

**10 (b)** Write down a possible sample frame that the personnel department could use when picking the sample.

**[1 mark]**

.....

The table shows information about the staff employed by the company.

| | | Number of male staff | Number of female staff |
|---|---|---|---|
| **Job title** | **Assistant** | 236 | 249 |
| | **Manager** | 383 | 492 |
| | **Senior Partner** | 75 | 65 |

**10 (c)** The personnel department decides to choose a sample of 160 staff, stratified by job title and gender.

Work out the number of male managers that it should choose.

**[3 marks]**

.....

.....

.....

Answer .....

One of the questions that the personnel department wants to ask in the survey is

> Have you ever taken a day off sick when you were not ill?
> Tick a box.
>
> Yes ☐ No ☐

**10 (d)** What problem might the personnel department have when it asks this question?

**[1 mark]**

.....

.....

The personnel manager suggests that data for this question are collected using the following random response technique.

> Each of the 160 staff in the sample secretly throws a coin.
> The question they then answer depends upon the result of the coin throw.

| Coin throw | Question to be answered |
|---|---|
| Heads | Did you throw Heads? |
| Tails | Have you ever taken a day off sick when you were not ill? |

104 staff gave the answer 'Yes'.

**10 (e)** Estimate the percentage of staff in the whole company that have taken a day off sick when they were not ill.

**[3 marks]**

.....

.....

.....

Answer ..... %

9

**11** The table shows the total number of people starting an apprenticeship in England each year between 2007/2008 and 2011/2012
It also contains some of the corresponding chain base index numbers.

| | **2007/2008** | **2008/2009** | **2009/2010** | **2010/2011** | **2011/2012** |
|---|---|---|---|---|---|
| **Number of apprenticeships (thousands)** | 225 | 240 | 280 | 457 | 520 |
| **Chain base index number** | | 106.7 | 116.7 | | |

Source: Department for Business, Innovation and Skills

**11 (a)** Calculate the chain base index numbers for 2010/2011 and 2011/2012
Write your answers in the table correct to one decimal place.

**[3 marks]**

.....

.....

.....

.....

**11 (b) (i)** Calculate the geometric mean of the four chain base index numbers.

**[2 marks]**

.....

.....

.....

Answer .....

**11 (b) (ii)** Interpret the value of your answer to (b)(i).

**[2 marks]**

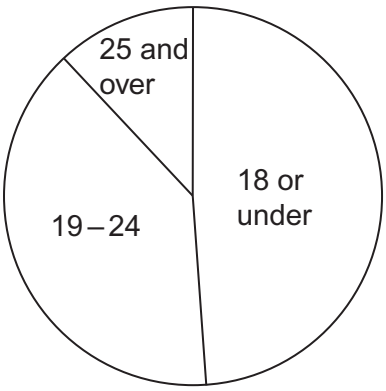
.....

.....

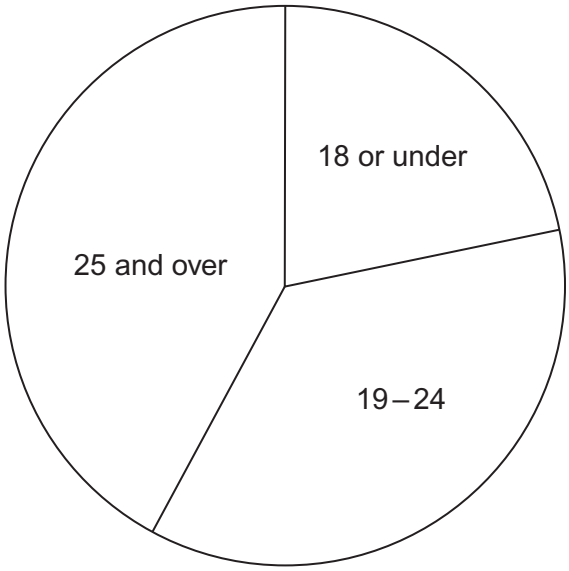
.....

The comparative pie charts show the ages of people starting an apprenticeship in the years 2007/2008 and 2012/2013

**2007/2008**




**2012/2013**




**11 (c)** Give **one** difference between the distribution of **ages** of people starting apprenticeships in 2007/2008 and 2012/2013

**[1 mark]**

.....

.....

**11 (d)** The radii of the two pie charts are 2.5 cm and 3.7 cm
The number of people starting an apprenticeship in 2007/2008 was 225 thousand.

Calculate the number of people starting an apprenticeship in 2012/2013
Give your answer in thousands.

**[3 marks]**

.....

.....

.....

.....

.....

Answer ..... thousand

11

**Turn over ▶**

**12** Laura works in a forest.

**12 (a)** There are some yew trees in the forest.
50% of the yew trees have a height of at least 15 metres.
95% of the yew trees have a height of between 8 metres and 22 metres.

Work out estimates of the mean and standard deviation of the heights of the yew trees.
Assume that the heights are normally distributed.

**[3 marks]**

.....

.....

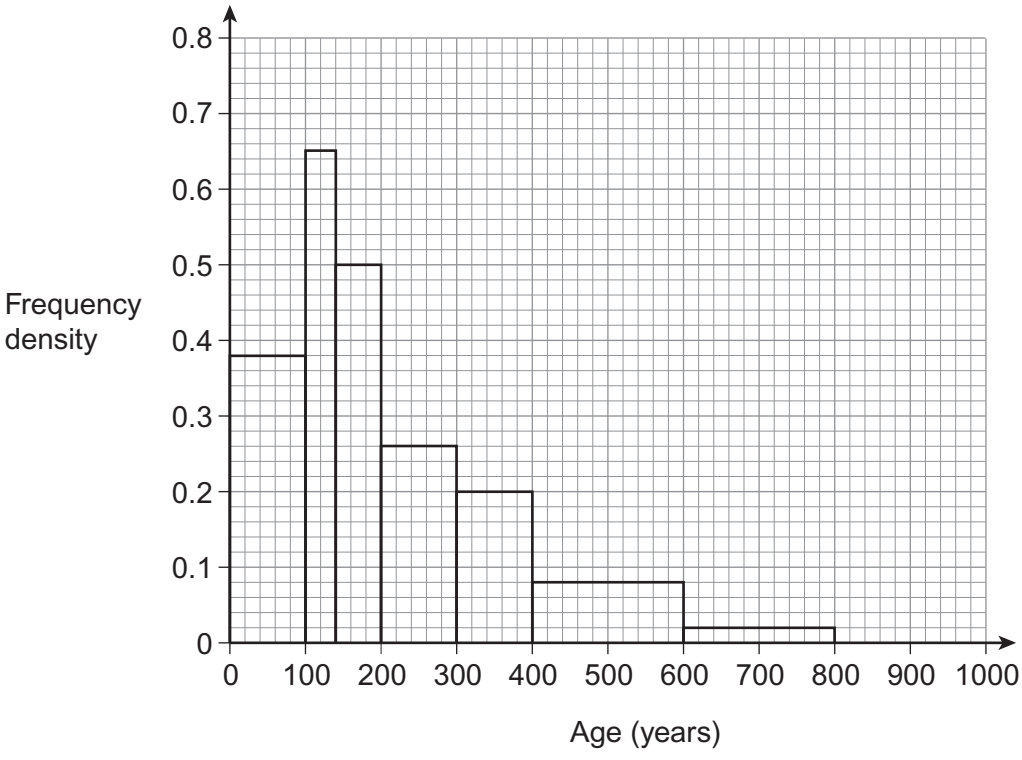
.....

mean = ..... metres

standard deviation = ..... metres

Laura's forest has 160 oak trees.

**12 (b)** The histogram shows the ages of the oak trees in the forest.



**12 (b) (i)** Explain how you can tell from the histogram that the ages of the oak trees are **not** normally distributed.

**[1 mark]**

.....

.....

**12 (b) (ii)** Estimate the number of oak trees in the forest that are **more** than 340 years old.

**[3 marks]**

.....

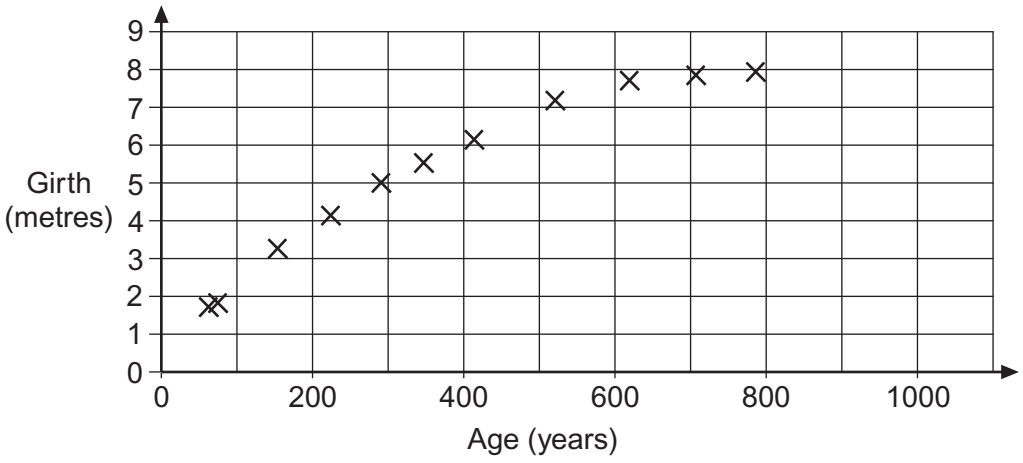
.....

.....

.....

Answer .....

**12 (c)** Laura measures the girth (circumference) of a sample of the oak trees.
The scatter graph shows the girth plotted against the age of the tree.




Laura suggests drawing a line of best fit on the scatter diagram.

Explain why a line of best fit is **not** appropriate.

**[1 mark]**

.....

.....

**END OF QUESTIONS**

8

**There are no questions printed on this page**

**DO NOT WRITE ON THIS PAGE**
**ANSWER IN THE SPACES PROVIDED**