

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Pages	Mark
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	



General Certificate of Secondary Education
Higher Tier
June 2012

Statistics

Written Paper

43101H

Monday 18 June 2012 1.30 pm to 3.30 pm

H

For this paper you must have:

- mathematical instruments.

You may use a calculator



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 space 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 booklet.
- You are expected to use a calculator where appropriate.

Advice

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



J U N 1 2 4 3 1 0 1 H 0 1

WMP/Jun12/43101H

43101H

You may need to use the following formulae:

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

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

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

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

Standard deviation for a frequency distribution

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

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



Turn over for the first question

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

Turn over ►



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

1 Barbara is a nurse at a local hospital.

One of her duties is to record the number of patients admitted to the hospital unit and whether their injuries are Critical, Serious or Minor.

This table shows her results for the first 80 patients seen last Friday.

Type of injury	Number of patients
Critical (C)	40
Serious (S)	28
Minor (M)	12

To help future planning she decides to simulate the likely type of injuries for the next 20 patients.

She allocates 2-digit numbers to the three types of injury as follows.

Type of injury	Number
Critical (C)	00 – 49
Serious (S)	50 – 84
Minor (M)	85 – 99

1 (a) Explain why she allocated the numbers 00 – 49 to the Critical injury group.

.....

.....

.....

(2 marks)

1 (b) (i) The type of injury is simulated using the following 2-digit random numbers.

Use one of the three letters C, S, or M to complete the table.

82	05	72	56	14	90	81	62	50	87	54	96	21	32	78	07	61	53	82	57
S	C	S	S	C															

(2 marks)



1 (b) (ii) Fill in the table below to show the total results for each type of injury after the simulation has been carried out.

Type of injury	Number of patients
Critical (C)	
Serious (S)	
Minor (M)	

(1 mark)

1 (b) (iii) Make **two** comments comparing the simulated results with the expected results based on the first 80 patients seen last Friday.

Comment 1

.....

Comment 2

.....

(2 marks)

Turn over for the next question



2 A population pyramid is drawn to show the percentages of the UK population by gender and age in 2001.

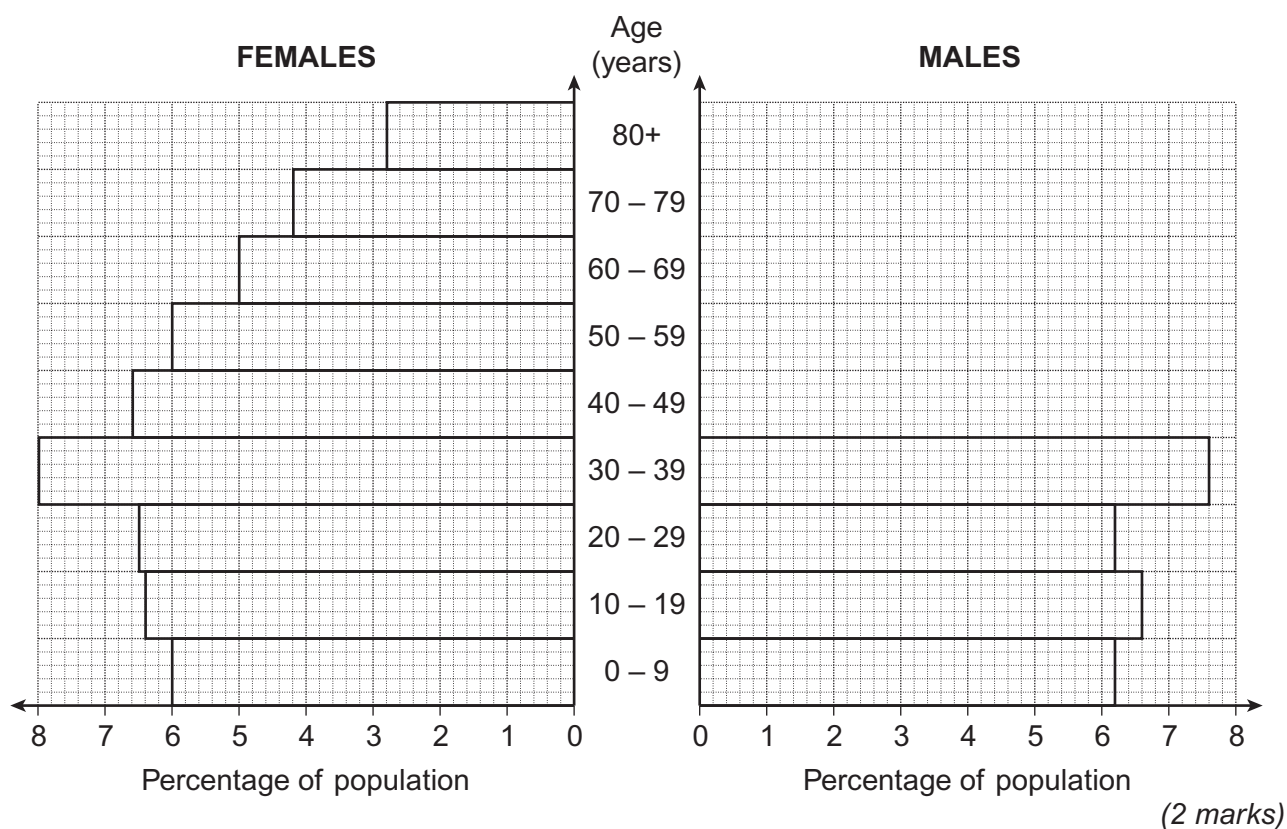
The data for females and for some of the male ages has already been drawn.

2 (a) Use the table to complete the drawing of the population pyramid for males.

Age (years)	Percentage of population (males)
0 – 9	6.2
10 – 19	6.6
20 – 29	6.2
30 – 39	7.6
40 – 49	6.8
50 – 59	6.0
60 – 69	4.5
70 – 79	3.2
80+	1.4

(Source: Adapted from Social Trends 2003, Number 33)

(You should assume that the age group 80+ covers the ages 80 – 89 years)



2 (b) What percentage of the UK population are females between the ages of 30 and 49?

.....
.....

Answer % (2 marks)

2 (c) Which age group has the same percentage of population for both males and females?

Answer (1 mark)

2 (d) Compare the population of males aged 70 and over with females aged 70 and over.

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

(1 mark)

Turn over for the next question



3 It is claimed that cleaning teeth with a new type of toothpaste, 'Wondershine', will reduce the number of fillings children need.

An experiment is to be set up to test this claim.

Here is a list of variables that may be connected to the experiment.

- A - How often the child cleans his/her teeth.
- B - The number of fillings the child needs during the experiment.
- C - How many sweets the child eats.
- D - The number of TVs in the child's house.
- E - Use of the toothpaste 'Wondershine'.

3 (a) For this experiment write down which of these variables is

3 (a) (i) the explanatory variable

Answer (1 mark)

3 (a) (ii) the response variable

Answer (1 mark)

3 (a) (iii) a possible extraneous variable

Answer (1 mark)

3 (a) (iv) a second possible extraneous variable.

Answer (1 mark)

3 (b) Tick a box to show whether variable D is an example of discrete, continuous or categorical data.

Discrete

Continuous

Categorical

(1 mark)



4 The table shows the detection rates (as percentages) for recorded crime, by the type of offence, for the year 2001/2.

Offence	England and Wales	Scotland	Northern Ireland
Violence	58	82	47
Sexual Offences	49	78	47
Drug Offences	94	99	77
Fraud	28	82	23
Robbery	17	37	14
Theft	17	34	13
Burglary	12	25	10
Criminal Damage	13	22	11
Other Offences	71	96	38
All recorded crime	23	45	20

(Source: Home Office Statistics June 2002)

4 (a) Which type of offence had the lowest rate of detection in Northern Ireland?

Answer (1 mark)

4 (b) Which type of offence was most likely to be detected?

Answer (1 mark)

4 (c) Which type of offence had exactly twice the detection rate in Scotland compared to England and Wales?

Answer (1 mark)

4 (d) Write down a possible reason why the detection rate for all offences was significantly higher in Scotland than in the other countries shown.

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

4 (e) Explain why the percentage for **all recorded crime** does not equal the average of the data in each column.

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

10

Turn over ►



5 The table shows the index numbers and weights for five different items of household expenditure in 2003, 2007 and 2009.

Item	Index Number			Weight
	2003	2007	2009	
Food	100	104	106	0.2
Heat and Light	100	105	108	0.2
Clothing	100	100	97	0.15
Mortgage	100	106	106	0.4
Other	100	105	109	0.05

5 (a) The base year is 2003.

How do you know this from the figures in the table?

..... (1 mark)

5 (b) What happened to household expenditure on mortgages from 2007 to 2009?

..... (1 mark)

5 (c) Which is the only item to show a reduction in expenditure from 2003 to 2009?

Answer (1 mark)

5 (d) Explain what the weight for Mortgage expenditure shows.

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

5 (e) (i) Heat and Light expenditure for a typical household in 2007 was £630.

How much would this have been in 2009?

.....
.....

Answer £ (2 marks)



5 (e) (ii) Find the actual increase in Heat and Light expenditure from 2003 to 2009.

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

Answer £ (3 marks)

Turn over for the next question



6 (a) Eight students study French and German.
The students take an oral test and a written test in each subject.
The marks are then ranked for each test.

6 (a) (i) The value of Spearman's rank correlation coefficient between the two oral tests is +0.76

Explain what this value shows.

.....
.....

(1 mark)

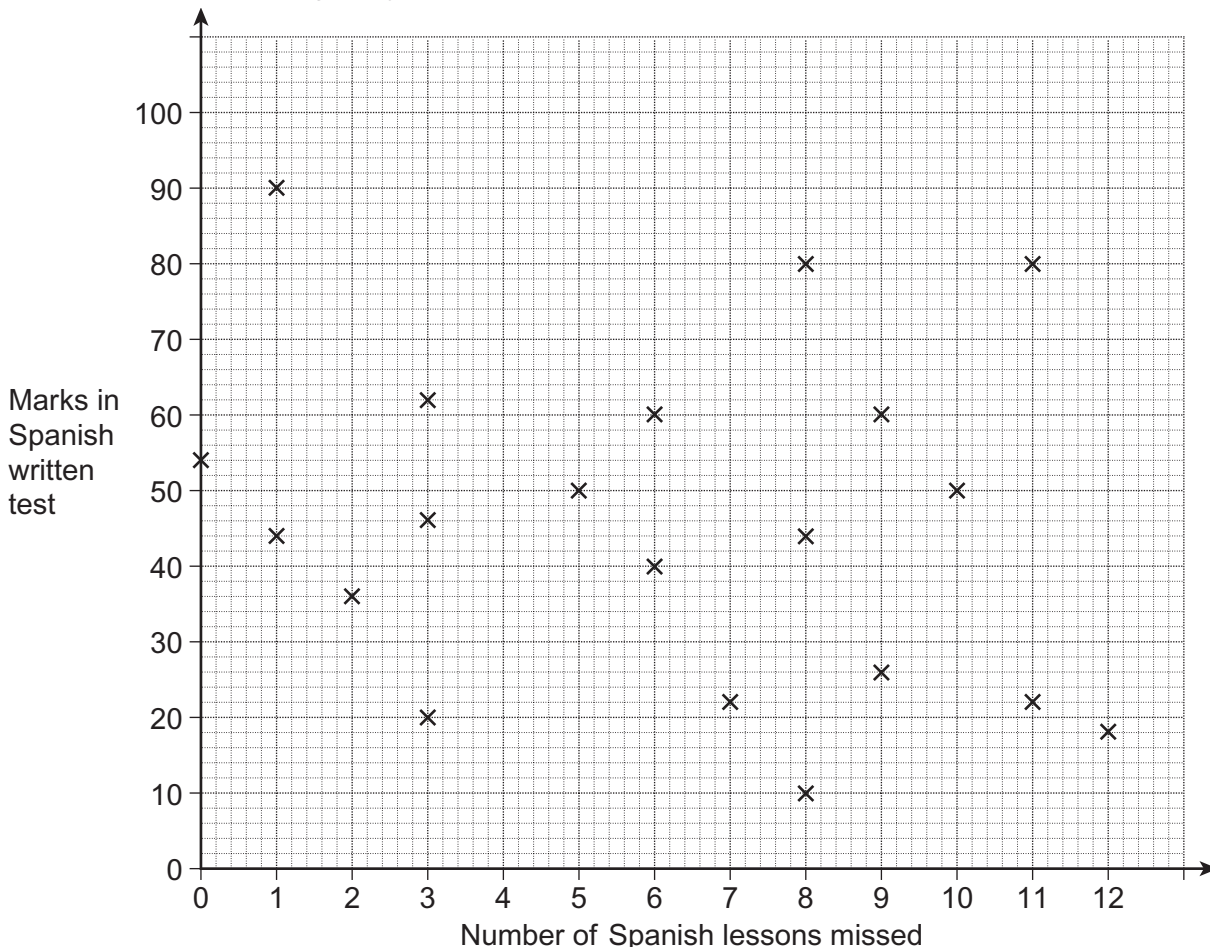
6 (a) (ii) The value of Spearman's rank correlation coefficient between the two written tests is -0.06

Explain what this value shows.

.....
.....

(1 mark)

6 (b) 20 students take a Spanish written test.
The scatter diagram shows their marks and the number of Spanish lessons they had missed during the year.



6 (b) (i) Write down the mark of the student who missed most lessons.

Answer (1 mark)

6 (b) (ii) Write down the number of lessons missed by the student having a mark of 36

Answer (1 mark)

6 (b) (iii) One student missed many lessons but still had a high mark in the test.

Write down the mark and number of lessons missed by this student.

Mark lessons missed (2 marks)

6 (b) (iv) The teacher looks at the scatter diagram and concludes:

'the more Spanish lessons a student attends, the higher the mark in the written test'

Does the information in the scatter diagram support this conclusion?
Give a reason for your answer.

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

(1 mark)

Turn over for the next question



7 A flight from Heathrow to Sydney has 400 passengers.

Seats on the plane are numbered

- 01 – 10 for first class
- 11 – 70 for business class
- and 71 – 400 for economy class passengers.

The airline needs to conduct an in-flight survey to find out what passengers think about the food served on the plane.

Interviews will be carried out when all seats are occupied and it is planned to conduct 80 interviews in total.

Three alternative sampling methods have been suggested.

Method A

Select at random

- 2 passengers from first class
- 12 passengers from business class
- 66 passengers from economy class.

Method B

- Choose a simple random sample of 80 different passengers from those on the plane.

Method C

- Select a random number between 11 and 70
- then take every 5th seat number until a seat within the range 396 – 400 is reached.

7 (a) For **Method A**.

7 (a) (i) Name the sampling method suggested.

Answer (1 mark)

7 (a) (ii) Show how the numbers 2, 12 and 66 were calculated for the sample sizes in this case.

.....

.....

.....

.....

(2 marks)



7 (b) For **Method B.**

Describe how random numbers could be used to select the sample.

.....

.....

.....

.....

.....

(3 marks)

7 (c) For **Method C.**

Give **two** reasons why this method should **not** be used for this survey.

Reason 1

.....

.....

Reason 2

.....

.....

(2 marks)

Turn over for the next question



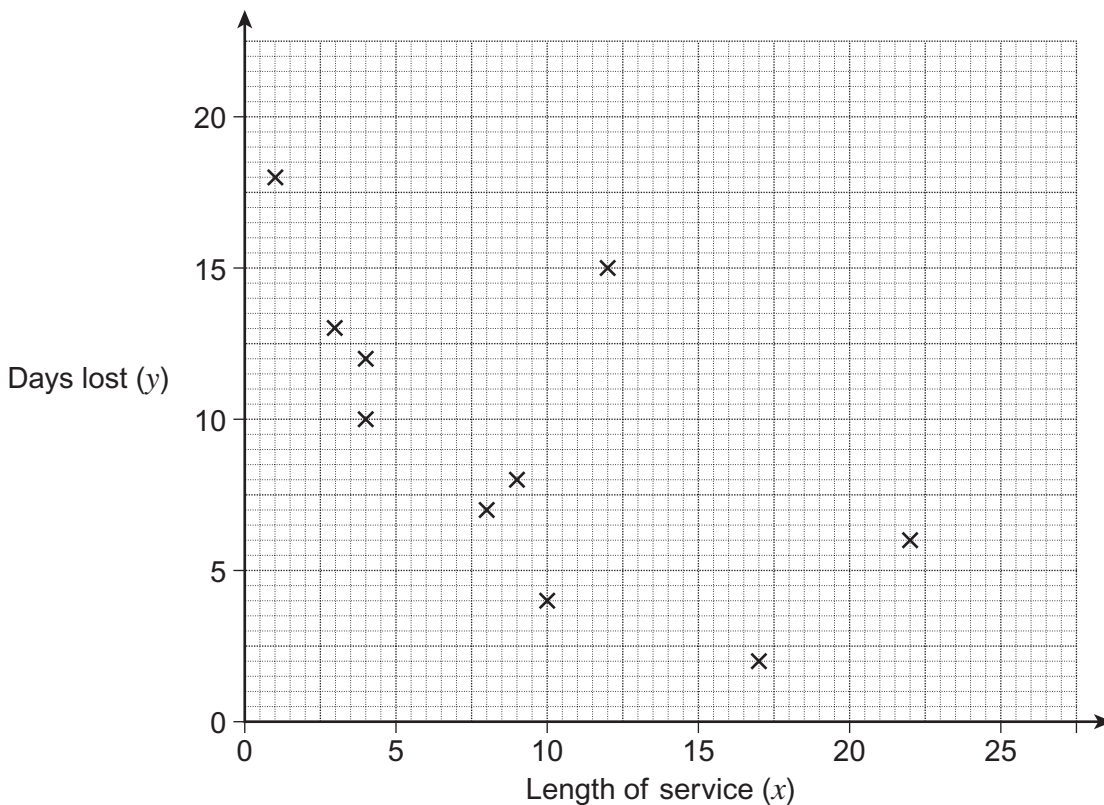
- 8 Management at a building company are worried about the number of working days lost, by employees, due to illness.

At the start of 2011 they select a random sample of ten employees.

For each employee they record the length of service (in years) with the company and the number of working days lost due to illness during 2011.

The results are summarised in the table and shown on the scatter diagram.

Employee	A	B	C	D	E	F	G	H	I	J
Length of service (x)	9	3	4	10	8	12	17	1	22	4
Days lost (y)	8	13	10	4	7	15	2	18	6	12



- 8 (a) The line of best fit for the data
- has a gradient of -0.5
 - passes through the point $(9, 9.5)$.

Draw this line on the scatter diagram.

(2 marks)



8 (b) (i) Write down the equation for the line of best fit in the form $y = mx + c$

Answer (1 mark)

8 (b) (ii) Explain why no sensible interpretation can be given for your value of c in the equation.

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

(1 mark)

8 (c) Estimate the number of days lost due to illness, during 2011, for an employee having 20 years' service with the company.

.....
.....

Answer days (1 mark)

8 (d) Give **two** reasons why the equation found in part (b) can **not** be used to estimate the number of days lost due to illness, for an employee having 30 years' service with the company.

Reason 1
.....

Reason 2
.....

(2 marks)

Turn over for the next question

7

Turn over ►



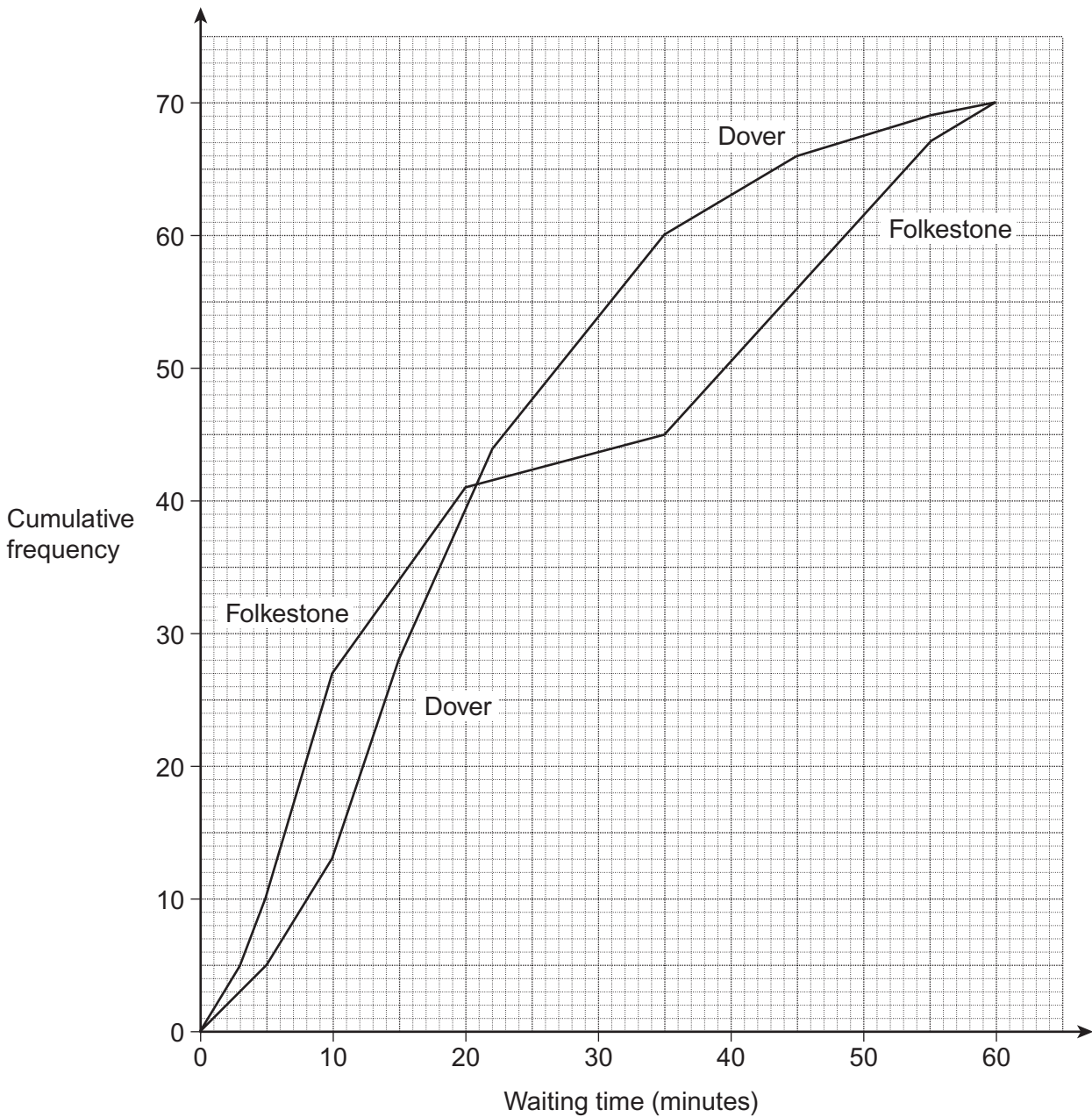
9 A ferry company has boats sailing daily to France from two UK ports, Dover and Folkestone.

The company has received complaints about the length of time passengers wait for their documents to be processed.

Last Friday the Manager conducted a survey of 70 passengers, selected at random, from each port.

He recorded the time they waited for their documents to be processed.

The times for each port are represented by the cumulative frequency graphs.



9 (a) Write down the median waiting time at Dover.

Answer minutes (1 mark)

9 (b) Estimate the range between the 1st and 9th deciles for the waiting time at Folkestone.

.....
.....

Answer minutes (2 marks)

9 (c) Find the **total** number of passengers at both ports who waited 45 minutes or more.

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

Answer (3 marks)

9 (d) Write down the waiting time where the cumulative frequency graphs are furthest apart.

Answer minutes (1 mark)

9 (e) Shakti can travel from either port to France.
She wants to minimise the chance of having to wait more than 30 minutes for her documents to be processed.

Assume the cumulative frequency graphs represent a typical day.

State, with a reason, which port she should use.

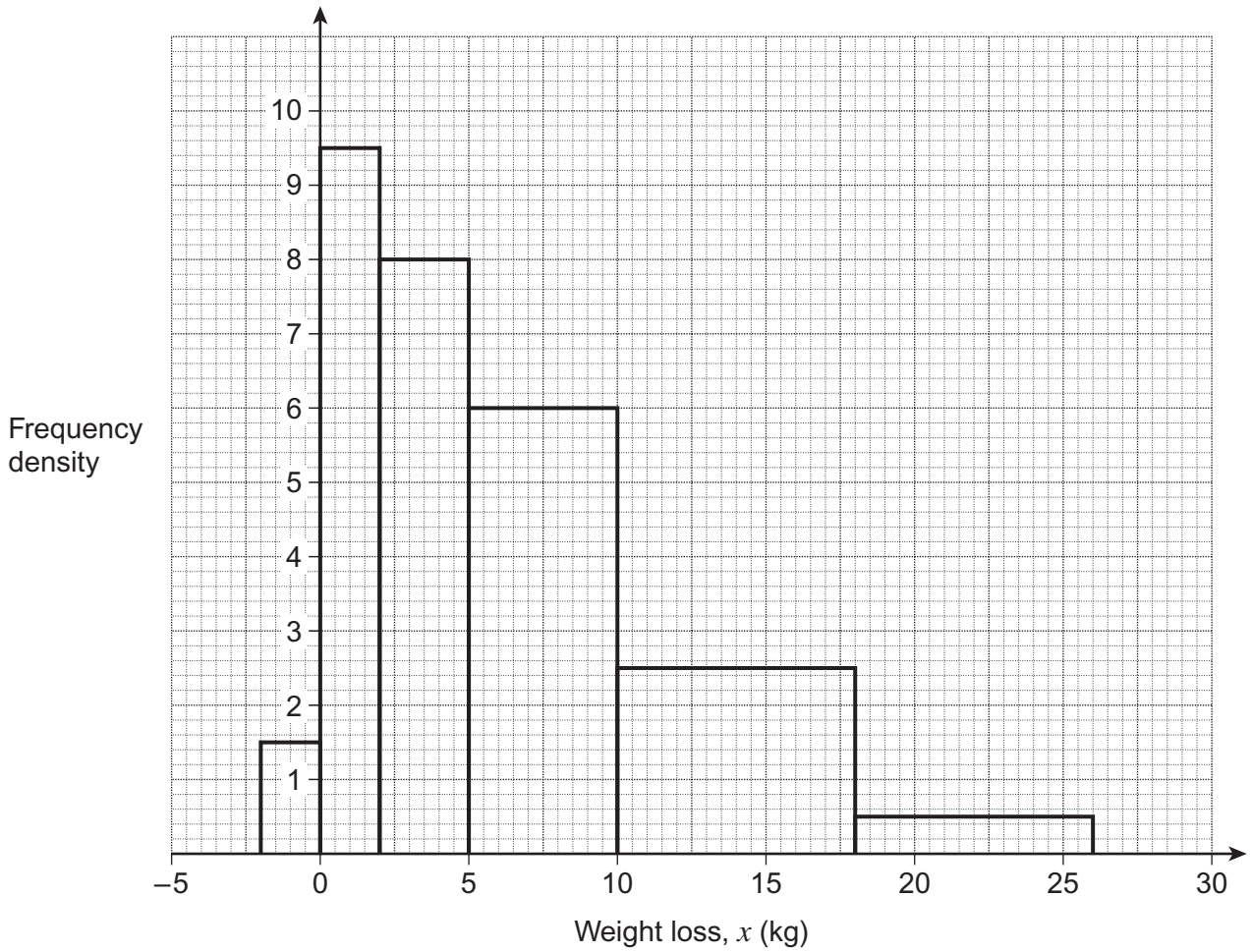
Port

Reason

..... (1 mark)



10 The histogram shows the weight loss, x kg, over a 20 week period for a sample of 100 males attending a health and fitness club.



10 (a) Use this histogram to complete the following frequency table for these data.

Weight loss, x (kg)	Frequency
$-2 \leq x < 0$	3
$0 \leq x < 2$	
$2 \leq x < 5$	
$5 \leq x < 10$	
$10 \leq x < 18$	
$18 \leq x < 26$	

.....

.....

.....

(3 marks)



10 (b) Data for a sample of 100 females was also collected.

You are given $\sum fx = 260$ and $\sum fx^2 = 2861.5$

Calculate estimates of the mean and standard deviation for the female weight loss over the period.

Write your answers in the table below.

.....

.....

.....

.....

	Mean (kg)	Median (kg)	Standard deviation (kg)
Male	6.93	5.67	5.53
Female		1.08	

(4 marks)

10 (c) Compare the weight losses for males and females.

.....

.....

.....

.....

(2 marks)

10 (d) One measure of skewness of a distribution can be calculated using the formula

$$\text{Skewness} = \frac{3 (\text{mean} - \text{median})}{\text{standard deviation}}$$

Use the results from the table in part (b) to calculate, to two decimal places, a measure of skewness for males and a measure of skewness for females.

.....

.....

.....

.....

Males

Females

(3 marks)

12

Turn over ►



10 (e) Make **two** distinct comments on what the measures of skewness show about the shapes of the two distributions.

Comment 1

.....

Comment 2

.....

(2 marks)



11 The weights of tins of tomatoes produced by Smiley plc are Normally distributed with mean, 157 g and standard deviation, 3.5 g.

11 (a) Within what weight limits would you expect almost all of the tins produced to fall?

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

Answer g to g (3 marks)

11 (b) The label on each tin states that the weight of tomatoes is 150g. Using the symmetry of the Normal distribution, estimate the percentage of these tins that are below 150g.

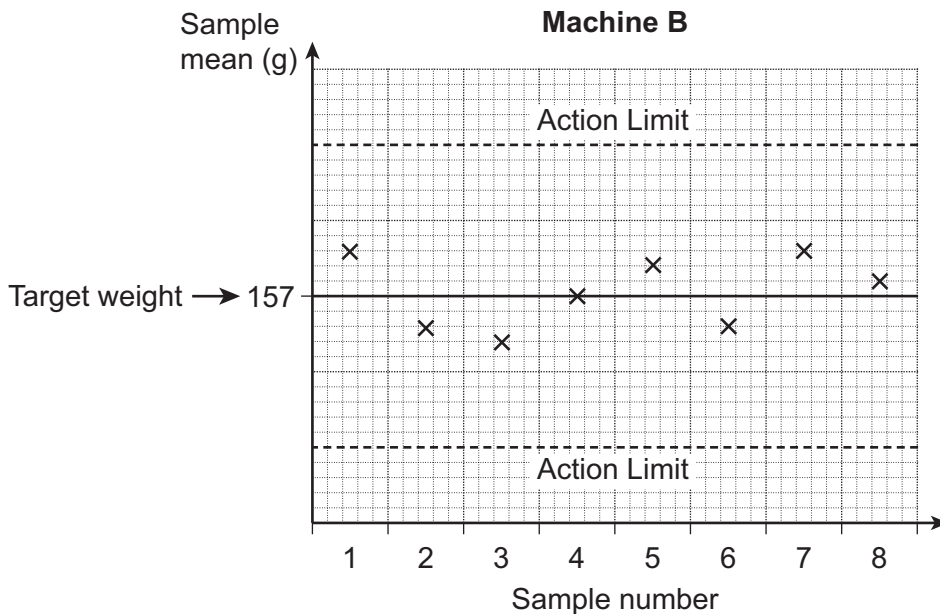
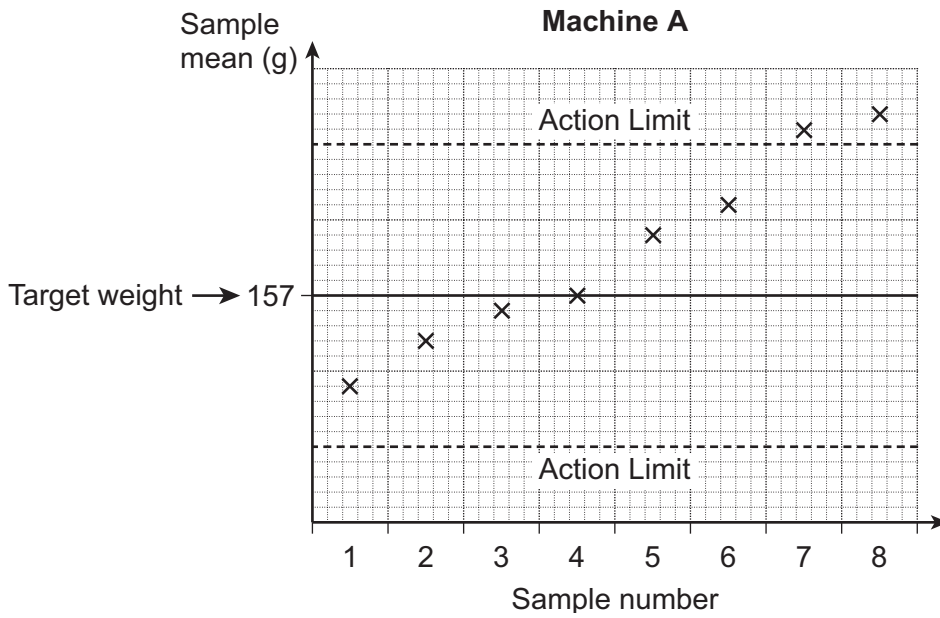
.....
.....
.....
.....

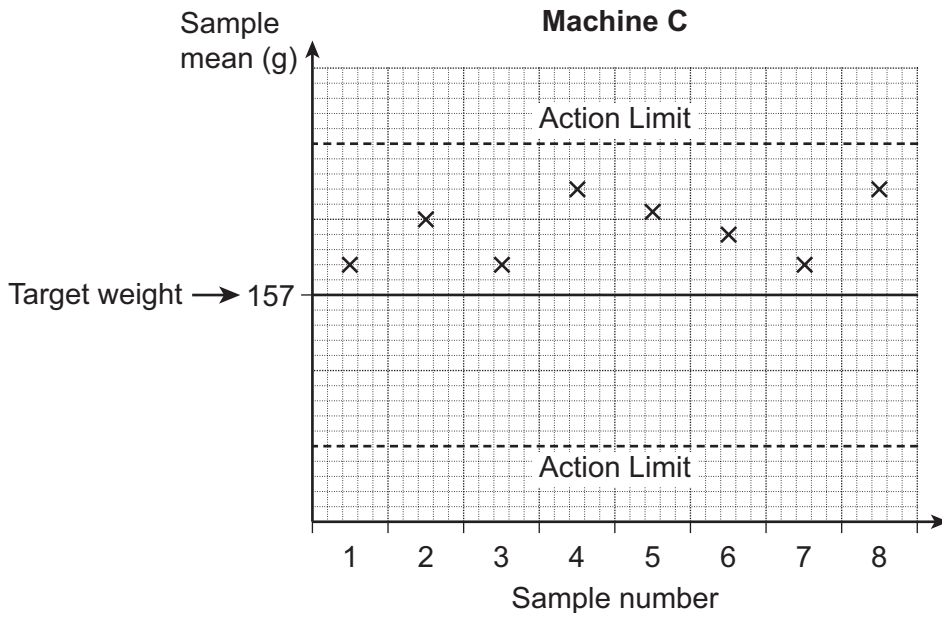
Answer % (3 marks)

Question 11 continues on the next page



- 11 (c)** The tins of tomatoes are produced on one of three machines A, B or C. Peter takes samples of tins at regular intervals from each machine. He plots the mean weight for each sample on a control chart. The results based on the first eight samples taken from each machine are shown.





11 (c) (i) For each machine, explain whether the chart shows it to be working satisfactorily or not.

Machine A

.....

Machine B

.....

Machine C

.....

(3 marks)

11 (c) (ii) Why is a chart for sample ranges usually drawn in addition to the chart for sample means?

.....

.....

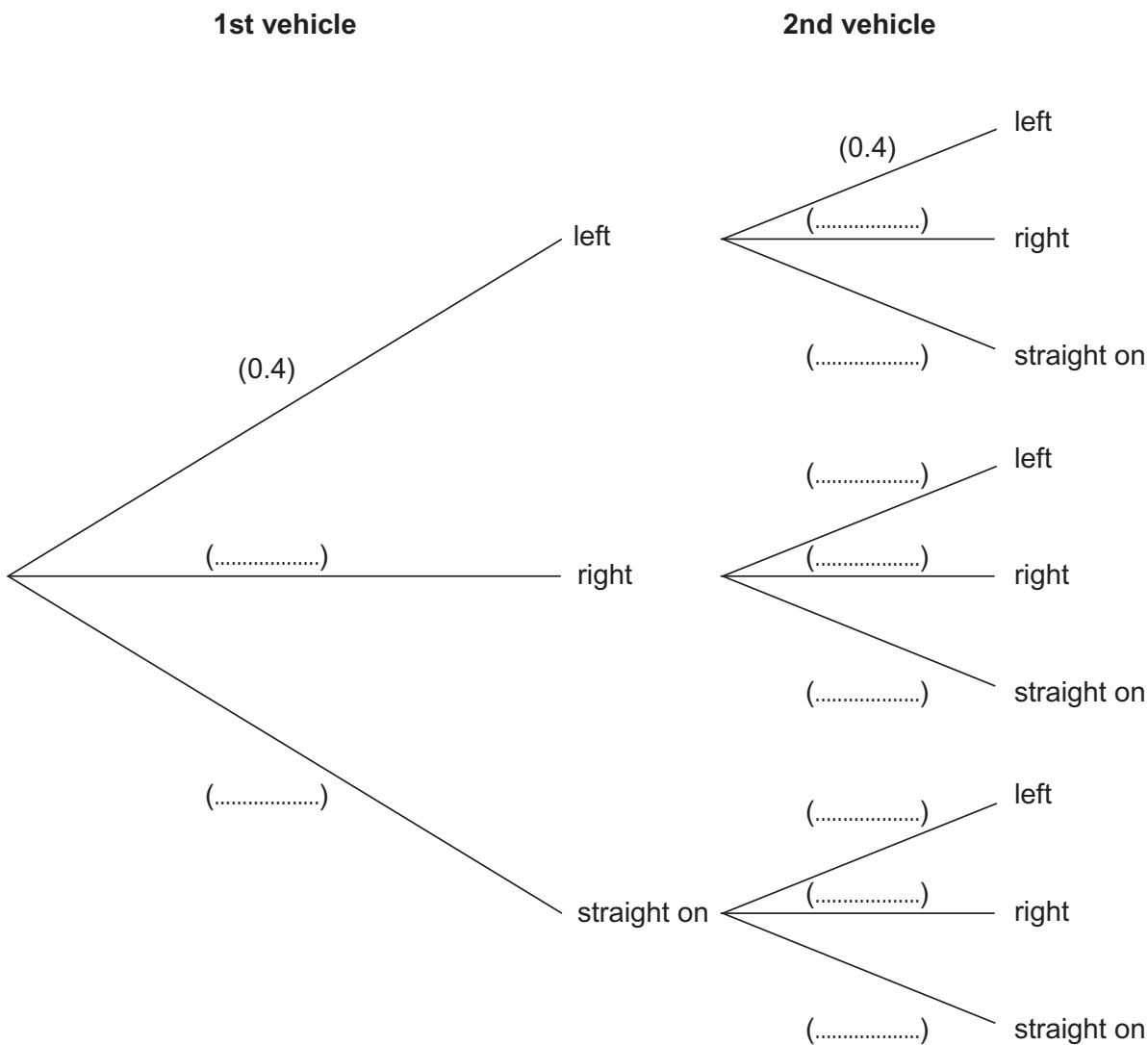
(1 mark)



- 12 Vehicles coming to a crossroads must go in one of three directions: left, right or straight on.
Traffic officers conducted a survey of vehicles coming from the south.
It showed that 40% turn left, 25% turn right and the rest go straight on.

12 (a) Assume the drivers of the vehicles choose direction independently of each other.

Complete the tree diagram to show the possible outcomes for the next two vehicles coming from the south.



12 (b) Use the tree diagram to find the probability that

12 (b) (i) both vehicles turn left

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

Answer (2 marks)

12 (b) (ii) one vehicle turns right and the other goes straight on

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

Answer (3 marks)

12 (b) (iii) both vehicles go in different directions.

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

Answer (4 marks)

12 (c) One day, 2800 vehicles come to the crossroads from the south.

How many of these would you expect to turn right?

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

Answer (2 marks)

END OF QUESTIONS



There are no questions printed on this page

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

