

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	
TOTAL	



General Certificate of Secondary Education
Higher Tier
June 2014

Statistics
Unit 1 Written Paper

43101H

H

Monday 23 June 2014 1.30 pm to 3.30 pm

<p>For this paper you must have:</p> <ul style="list-style-type: none"> • 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 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 4 4 3 1 0 1 H 0 1

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

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

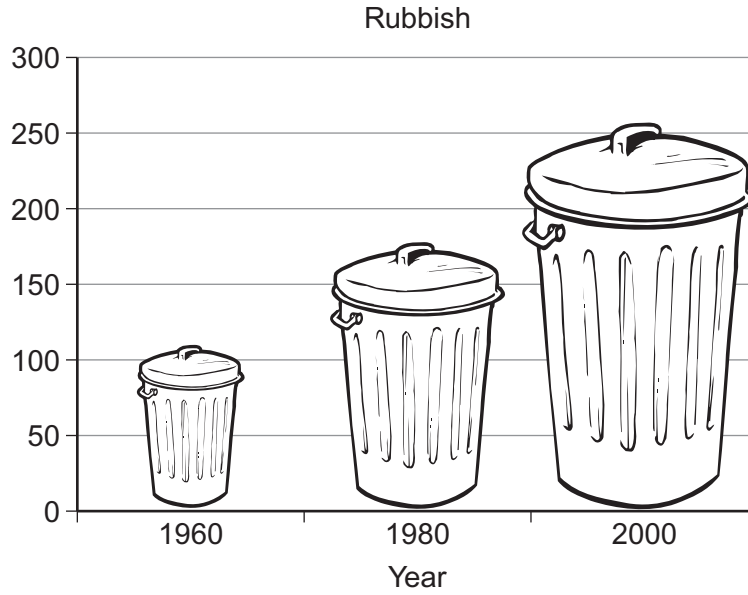
The same formulae apply to the standard deviation of a grouped frequency distribution where x is the mid-interval value.

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



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

1 The graph was used to show how much more rubbish has been produced since 1960.



Give **two** different criticisms of the graph.

[2 marks]

Criticism 1

Criticism 2

Turn over for the next question

Turn over ►



2 John collects stamps from all over the world.

Decide whether each of the following variables is discrete, continuous or qualitative.
In each case, circle the correct word.

2 (a) The length of a stamp.

[1 mark]

Discrete

Continuous

Qualitative

2 (b) The number of stamps in his collection.

[1 mark]

Discrete

Continuous

Qualitative

2 (c) The country the stamp is from.

[1 mark]

Discrete

Continuous

Qualitative



3 The table shows the age distribution and the number of deaths for the village of Mornington in 2012.

Age, in years	Population	Number of deaths
under 10	1500	3
10 – 24	2100	8
25 – 44	3600	12
45 – 64	1700	9
65 and over	3100	109

3 (a) Calculate the crude death rate for this village.

[3 marks]

.....

.....

Answer per thousand

3 (b) The crude death rate for the nearby village of Hopewell is 18.7 per thousand.
How might the age distribution of Hopewell differ from that of Mornington?

[1 mark]

.....

.....

3 (c) Explain why the standardised death rate is a better measure to use than the crude death rate.

[1 mark]

.....

.....



4 Heidi is planning to hold a charity collection at a local restaurant.
She will do this on either a Saturday evening or a Sunday evening.

To help her decide which evening to choose she finds out the number of people who visited the restaurant on these days.

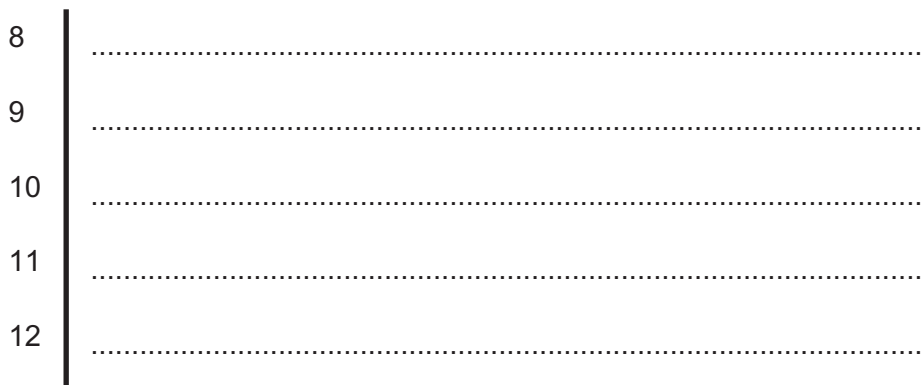
Here are her results for the last 15 Saturdays.

113	101	87	94	126	128	111	96
89	102	106	89	105	119	88	

4 (a) Show the data in an ordered stem-and-leaf diagram.

[3 marks]

Key | represents people



4 (b) Work out the median for the data.

[2 marks]

.....
.....

Median = people



4 (c) Show that the interquartile range for the data is 24 people.

[3 marks]

.....

.....

.....

.....

.....

.....

4 (d) Data for the number of people visiting the restaurant over the last 15 **Sundays** is summarised below.

Median = 110

Interquartile range = 17

Give **two** reasons why Heidi should hold the collection on a Sunday rather than a Saturday.

[2 marks]

Reason 1

.....

Reason 2

.....

Turn over for the next question

10

Turn over ►



5 A teaching agency employs 400 tutors.
Each tutor teaches one of three subjects: English, Maths or Science.
The number of tutors of each subject is shown in the table.

Subject	Number of tutors
English	200
Maths	185
Science	15

The manager wishes to survey a sample of tutors.
She decides to sample 50 of the 400 tutors.

5 (a) Give a reason why a random sampling method might **not** be appropriate in this case. **[1 mark]**

.....
.....

5 (b) The manager suggests selecting the first 50 tutors from an alphabetical list of the 400.
Name this method of sampling. **[1 mark]**

Answer

5 (c) The assistant manager suggests selecting a sample of 50, stratified by subject.
How many Maths tutors would there be in this sample? **[3 marks]**

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

Answer



5 (d) The survey will be carried out online.

Give **one** advantage and **one** disadvantage of using this method of data collection.

[2 marks]

Advantage

.....

Disadvantage

.....

5 (e) One of the questions on the survey will be:

How long have you worked as a tutor?	
Tick a box.	
2 – 4 years	<input type="checkbox"/>
5 – 7 years	<input type="checkbox"/>
8 – 10 years	<input type="checkbox"/>
10 – 12 years	<input type="checkbox"/>
over 12 years	<input type="checkbox"/>

Give **two** criticisms of the response choices provided.

[2 marks]

Criticism 1

.....

Criticism 2

.....



6 The table is a summary of key statistics relating to Housing, Traffic, Recycling and Crime in the UK in 2009.

	Number of houses built (thousands)	Median house price (£ thousands)	Percentage increase in traffic between 1999 and 2009	Percentage of household waste recycled	Number of recorded crimes per 100,000 population
English Regions					
North East	3.9	120	9.1	34.7	6 900
North West	12.9	130	9.2	38.5	8 101
Yorkshire and the Humber	10.8	130	8.4	36.8	8 304
East Midlands	9.0	135	7.3	45.6	7 894
West Midlands	9.1	142	7.8	40.0	7 260
East	13.5	175	6.8	46.1	6 437
London	12.3	250	-6.3	31.8	10 893
South East	22.0	203	4.2	40.0	7 172
South West	13.8	175	11.6	43.5	6 598
England	107.3	170	6.5	37.8	7 883
Wales	6.4	133	12.7	40.4	7 404
Scotland	16.1	-	10.6	37.7	6 508
Northern Ireland	8.7	-	20.3	35.6	6 097

Source: Adapted from Social Trends (41)

6 (a) Which **English Region** had the lowest level of 'recorded crime' in 2009? **[1 mark]**

Answer

6 (b) Look at the values in the column headed 'Percentage increase in traffic between 1999 and 2009'.

Find the range of these percentages for the **English Regions**.

[2 marks]

.....

Answer %



6 (c) Calculate, as a percentage, how much more expensive median house prices are in England compared to Wales in 2009.

[3 marks]

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

Answer %

6 (d) A statistician plans to draw a pie chart to compare the 'number of houses built' in 2009 for the four countries of the UK.

Calculate the angle for **Scotland**.
Give your angle to the nearest degree.

[4 marks]

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

Answer degrees

6 (e) London claims to be the most environmentally friendly English Region.

6 (e) (i) Give one reason why the data support this claim.

[1 mark]

.....

6 (e) (ii) Give one reason why the data do **not** support this claim.

[1 mark]

.....



7 (a) Kiran has three different breeds of cow.

He has 280 cows altogether.
One-quarter of his cows are Angus breed, 125 are Hereford and the rest are Ayrshire.

Kiran needs to test all his cows for a disease but has so far only been able to test 40% of each breed.

Complete the table to show this information.

[5 marks]

	Angus	Hereford	Ayrshire	Total
Tested				
Not tested				
Total		125		280

7 (b) One of the cows is selected at random.

Work out the probability that the cow chosen is

7 (b) (i) a 'Hereford' and 'not tested'

[1 mark]

.....

Answer

7 (b) (ii) not an 'Ayrshire'

[2 marks]

.....

.....

.....

Answer



7 (b) (iii) either an 'Angus' or 'tested' or both.

[3 marks]

.....

.....

.....

.....

Answer

7 (c) Kiran now chooses **three** of his 280 cows at random.

Calculate the probability that exactly **one** of these three cows has been tested for the disease.

[4 marks]

.....

.....

.....

.....

Answer

Turn over for the next question



8 The table shows information about the distance travelled (in kilometres) by a sample of 132 visitors to a theme park.

Distance travelled, x (km)	$0 < x \leq 5$	$5 < x \leq 10$	$10 < x \leq 20$	$20 < x \leq 35$	$35 < x \leq 60$
Frequency	18	30	48	21	15

8 (a) Estimate how many of these visitors travelled 7 km or more.

[3 marks]

.....

.....

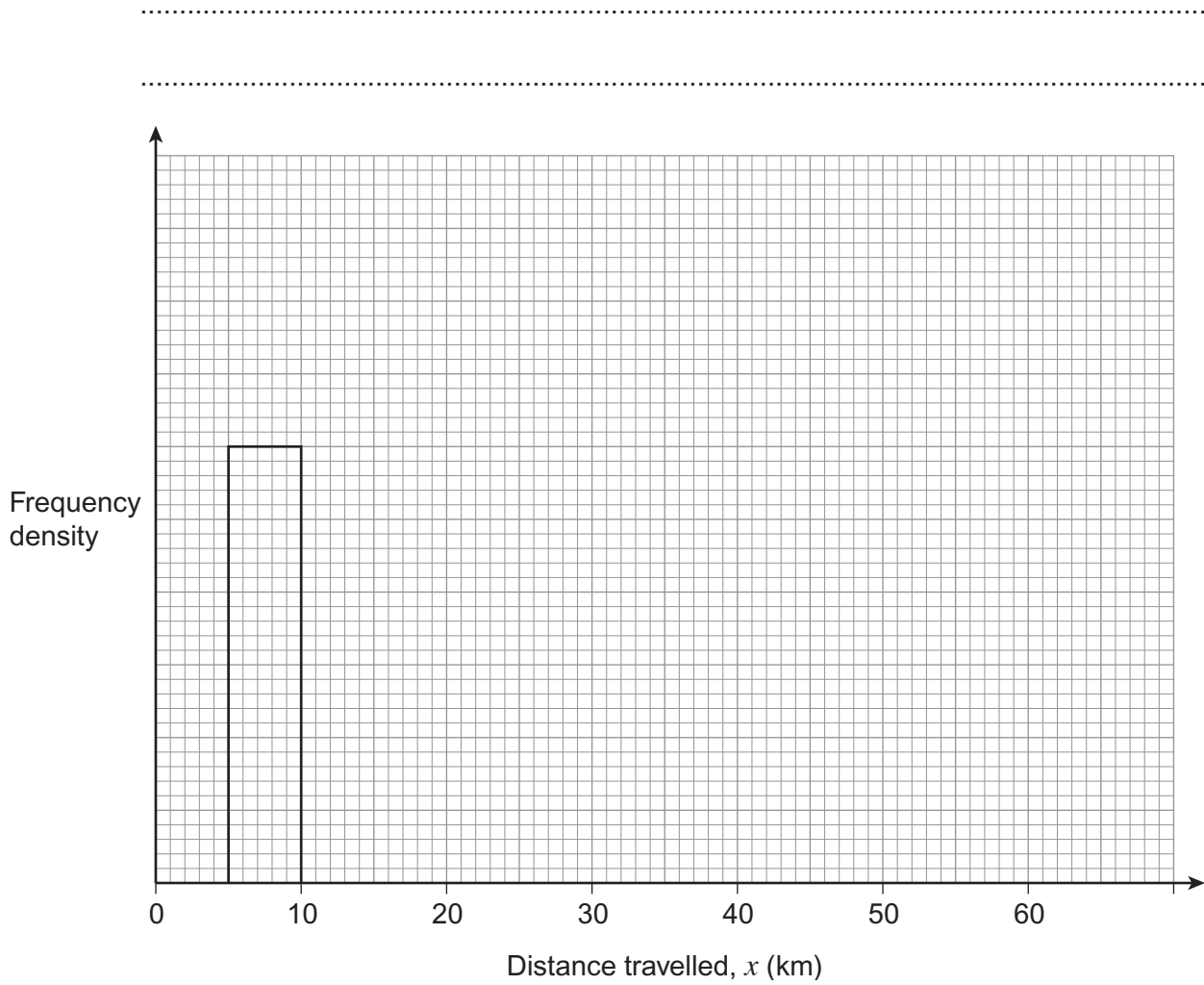
.....

Answer



8 (b) Complete the histogram.

[4 marks]



8 (c) Describe the skewness of the distribution.

[1 mark]

.....

8 (d) Later the data were regrouped using the following classes:

$$0 < x \leq 20$$

$$20 < x \leq 60$$

Give **one** disadvantage of grouping the data using these classes.

[1 mark]

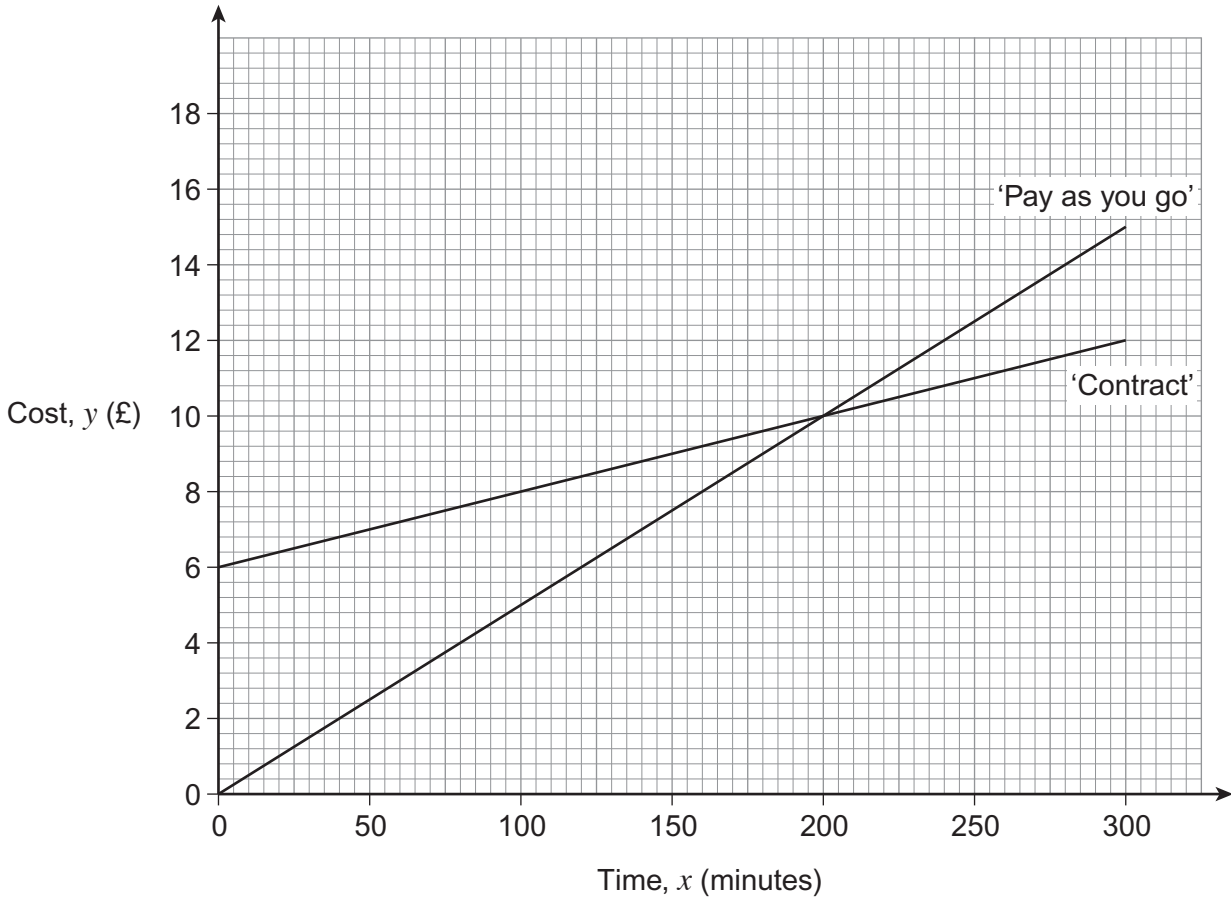
.....

.....

Turn over ►



9 A mobile phone company sells both 'Pay as you go' and 'Contract' phones.
The graph shows how the cost, y (£), of using both types of phone varies with the time, x (minutes), taken on calls each month.



9 (a) Show that the equation of the line for the 'Pay as you go' phone option is $y = \frac{1}{20}x$
[1 mark]

.....

.....



9 (b) (i) Find the equation of the line for the 'Contract' phone option.

Give your answer in the form $y = mx + c$

[2 marks]

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

Answer $y =$

9 (b) (ii) Interpret, in context, the values for the intercept, c , and the gradient, m , of the line.

[2 marks]

The intercept, c , is

The gradient, m , is

9 (c) In the last three months Kerri used her phone for 180, 250 and 230 minutes.

Would she be better off financially with a 'Contract' or 'Pay as you go' phone?

Give a reason for your answer.

[2 marks]

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

7

Turn over ►



10 A manufacturing firm regularly uses a mail distribution company, 'Expo', to send documents from its factory to its Head Office.

Management at the manufacturing firm decide to check the journey times (in minutes) of a sample of 36 deliveries made by 'Expo' drivers.

10 (a) A summary of the times is

$$\sum x = 3636$$

$$\sum x^2 = 367\ 677$$

10 (a) (i) Show that the mean time is 101 minutes.

[1 mark]

.....
.....

10 (a) (ii) Show that the standard deviation of the times is 3.5 minutes.

[2 marks]

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

10 (b) Assume that the delivery times are Normally distributed.

10 (b) (i) One of the delivery times was recorded as 115 minutes.

Use the values of the mean and standard deviation to calculate the standardised score for this delivery time.

[2 marks]

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

Answer



10 (b) (ii) It is later claimed that the time of 115 minutes had been incorrectly recorded.

Based on your answer to part (b)(i), explain why this claim is likely to be correct.

[1 mark]

.....
.....

10 (c) A rival delivery firm, 'Rapid', claim they can make the deliveries in a shorter time than 'Expo'.

As a check on this claim 'Rapid' were given a number of trial deliveries to make.

Give a reason why the results of the trial deliveries by 'Rapid' may **not** be a suitable basis to use in comparison with the delivery times for 'Expo'.

[1 mark]

.....
.....

10 (d) A random sample of 28 of the delivery times recorded by 'Rapid' showed that 21 deliveries were made in under 102 minutes.

10 (d) (i) Estimate the percentage of **all** 'Rapid' delivery times that would be completed in under 102 minutes.

[1 mark]

.....

Answer %

10 (d) (ii) To improve the accuracy of this estimate it is suggested that the variability should be reduced by one half.

What size sample would be needed to achieve this level of improvement?

[1 mark]

.....

Answer



11 Hannah is carrying out a study of the railway industry.

Part of her work involves investigating the cost of travelling to London by train. She collects data from 9 towns on annual season ticket prices and journey times to London.

Town	Ticket price (£)	Journey time (mins)	Rank ticket price	Rank journey time	d	d^2
A	3523	21				
B	2145	58				
C	3460	42				
D	3712	75		1		
E	3231	56				
F	6400	59	1	2		
G	3251	53				
H	4300	56	2			
I	3312	40				
Total =						

11 (a) (i) Complete the ranking column for each set of data.

[2 marks]

11 (a) (ii) Calculate the value of Spearman's rank correlation coefficient for the data.

[4 marks]

.....

.....

Answer

11 (b) Comment on the claim that the cost of an annual season ticket is **strongly** linked to the average journey time.

[1 mark]

.....

.....



11 (c) Calculate the maximum amount that the annual fare for town F can be reduced by before it changes the value of Spearman’s rank correlation coefficient. **[2 marks]**

.....

Answer £

11 (d) Hannah’s manager suggested she might want to use the Product Moment correlation coefficient rather than Spearman’s rank correlation coefficient for some of her analysis. She calculates values for the Product Moment correlation coefficient for 5 sets of data.

0.84 0.08 -0.57 0.01 -0.85

Put these values in order, starting with the least correlated.

[2 marks]

least correlated

most correlated

.....

11 (e) As part of her work Hannah calculated the Product Moment correlation coefficient between the number of first-class passengers and journey time.

She obtained the value -0.85

She concluded,

‘this indicates that an increase in the number of first-class passengers on the train will cause an increase in journey time’.

Give **two** reasons why her conclusion is incorrect.

[2 marks]

Reason 1

.....

Reason 2

.....



12 A gift shop recorded their monthly sales in hundreds of pounds, from the start of January 2012 to the end of December 2013.

The results are summarised in the table.

Month	Sales in 2012 (hundreds of pounds)	Sales in 2013 (hundreds of pounds)	Moving annual total (hundreds of pounds)
January	150	167	1658
February	164	175	1669
March	148	162	1683
April	132	156	1707
May	121	146	1732
June	110	134	1756
July	108	130	1778
August	96	119	1801
September	118	136	
October	148	170	
November	165	192	1868
December	181	220	1907

12 (a) The first value in the 'Moving annual total' column gives the total sales for the 12 months from **February 2012 to the end of January 2013**.

Calculate the two missing values in the 'Moving annual total' column.
Write your answers in the table.

[3 marks]

.....

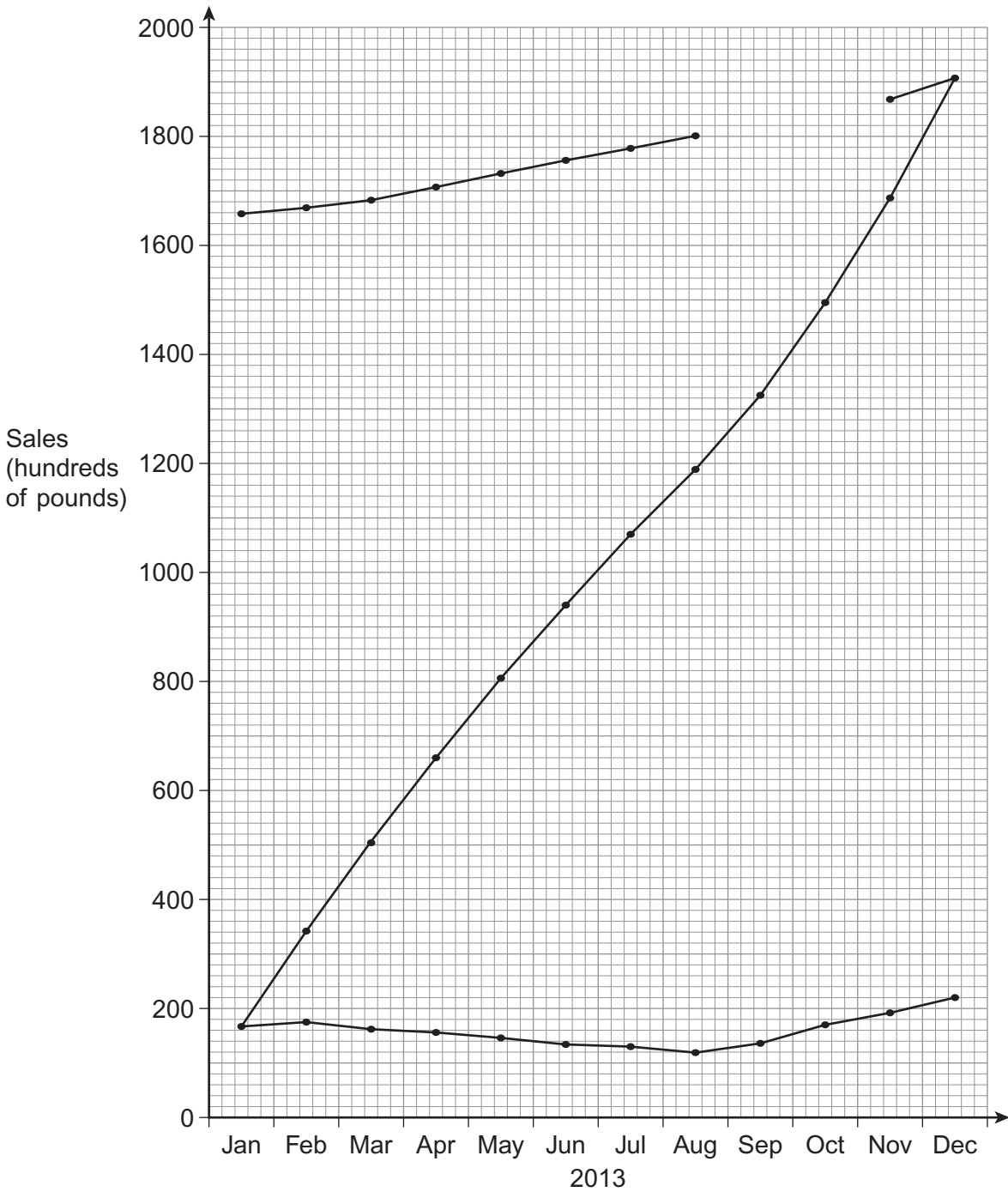
.....

.....

12 (b) Plot the values for September and October to complete the 'Z' chart opposite.

[2 marks]





12 (c) Give one comment on the pattern of sales shown in the 'Z' chart.

[1 mark]

.....
.....

END OF QUESTIONS

6



There are no questions printed on this page

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

