S1: Linear Regression

Past Paper Questions
2006 - 2013

Name:

1 At a certain small restaurant, the waiting time is defined as the time between sitting down at a table and a waiter first arriving at the table. This waiting time is dependent upon the number of other customers already seated in the restaurant.

Alex is a customer who visited the restaurant on 10 separate days. The table shows, for each of these days, the number, x, of customers already seated and his waiting time, y minutes.

x	9	3	4	10	8	12	7	11	2	6
y	11	6	5	11	9	13	9	12	4	7

(a) Calculate the equation of the least squares regression line of y on x in the form $y = a + bx$. *(4 marks)*

(b) Give an interpretation, in context, for each of your values of a and b. *(2 marks)*

(c) Use your regression equation to estimate Alex's waiting time when the number of customers already seated in the restaurant is:

(i) 5;

(ii) 25. *(2 marks)*

(d) Comment on the likely reliability of **each** of your estimates in part (c), given that, for the regression line calculated in part (a), the values of the 10 residuals lie between +1.1 minutes and −1.1 minutes. *(3 marks)*

June 2006

3 A new car tyre is fitted to a wheel. The tyre is inflated to its recommended pressure of 265 kPa and the wheel left unused. At 3-month intervals thereafter, the tyre pressure is measured with the following results:

Time after fitting (x months)	0	3	6	9	12	15	18	21	24
Tyre pressure (y kPa)	265	250	240	235	225	215	210	195	180

(a) (i) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(ii) Interpret in context the value for the gradient of your line. *(2 marks)*

(iii) Comment on the value for the intercept with the y-axis of your line. *(2 marks)*

(b) The tyre manufacturer states that, when one of these new tyres is fitted to the wheel of a car and then inflated to 265 kPa, a suitable regression equation is of the form

$$y = 265 + bx$$

The manufacturer also states that, as the car is used, the tyre pressure will decrease at twice the rate of that found in part (a).

(i) Suggest a suitable value for b. *(2 marks)*

(ii) One of these new tyres is fitted to the wheel of a car and inflated to 265 kPa. The car is then used for 8 months, after which the tyre pressure is checked for the first time.

Show that, accepting the manufacturer's statements, the tyre pressure can be expected to have fallen below its minimum safety value of 220 kPa. *(2 marks)*

January 2007

7 [**Figure 1**, printed on the insert, is provided for use in this question.]

Stan is a retired academic who supplements his pension by mowing lawns for customers who live nearby.

As part of a review of his charges for this work, he measures the areas, $x\,\text{m}^2$, of a random sample of eight of his customers' lawns and notes the times, y minutes, that it takes him to mow these lawns. His results are shown in the table.

Customer	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
x	360	140	860	600	1180	540	260	480
y	50	25	135	70	140	90	55	70

(a) On **Figure 1**, plot a scatter diagram of these data. *(2 marks)*

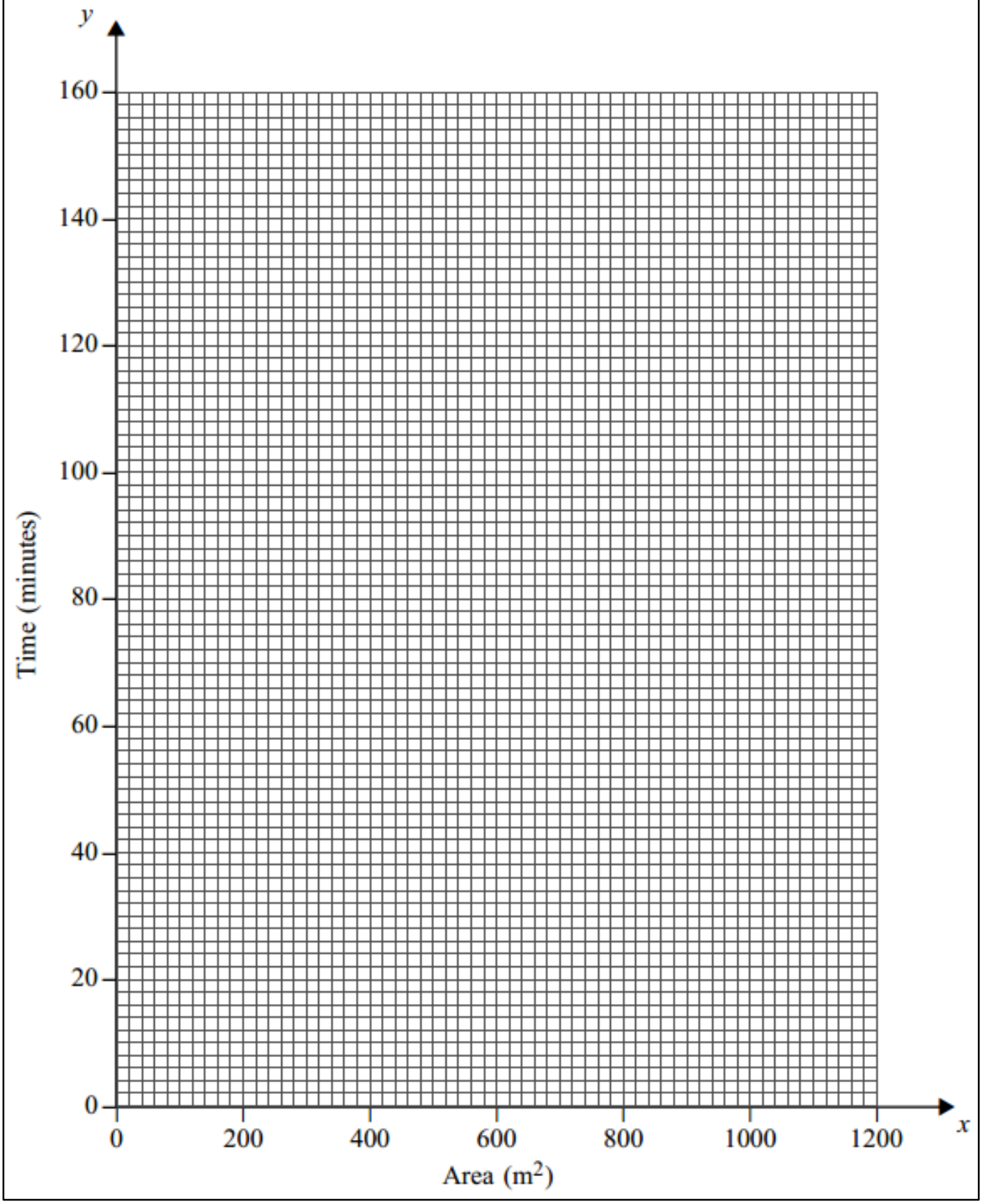
(b) Calculate the equation of the least squares regression line of y on x. Draw your line on **Figure 1**. *(6 marks)*

(c) Calculate the value of the residual for Customer H and indicate how your value is confirmed by your scatter diagram. *(3 marks)*

(d) Given that Stan charges £12 per hour, estimate the charge for mowing a customer's lawn that has an area of $560\,\text{m}^2$. *(4 marks)*

Figure 1 (for use in Question 7)

Lawn Areas and Mowing Times



June 2007

5 Bob, a gardener, measures the time taken, y minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, x °C. His results are shown in the table.

x	16	20	24	28	32	36	40	44	48	52	56
y	4.7	4.3	3.8	3.5	3.0	2.7	2.4	2.0	1.8	1.6	1.1

(a) State why the explanatory variable is temperature. *(1 mark)*

(b) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(c) (i) Interpret, in the context of this question, your value for b. *(2 marks)*

(ii) Explain why no sensible practical interpretation can be given for your value of a. *(2 marks)*

(d) (i) Estimate the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 30 °C. *(2 marks)*

(ii) Show why the equation cannot be used to make a valid estimate of the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 75 °C. *(2 marks)*

January 2008

4 [**Figure 1**, printed on the insert, is provided for use in this question.]

Roseen is a self-employed decorator who wishes to estimate the times that it will take her to decorate bedrooms based upon their floor areas. She records the floor area, $x\,\text{m}^2$, and the decorating time, y hours, for each of 10 bedrooms she has recently decorated.

x	11.0	22.0	7.5	21.0	13.0	16.5	14.0	16.0	18.5	20.5
y	15.0	35.0	16.0	23.5	24.0	17.5	14.5	27.5	22.5	34.5

(a) On **Figure 1**, plot a scatter diagram of these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

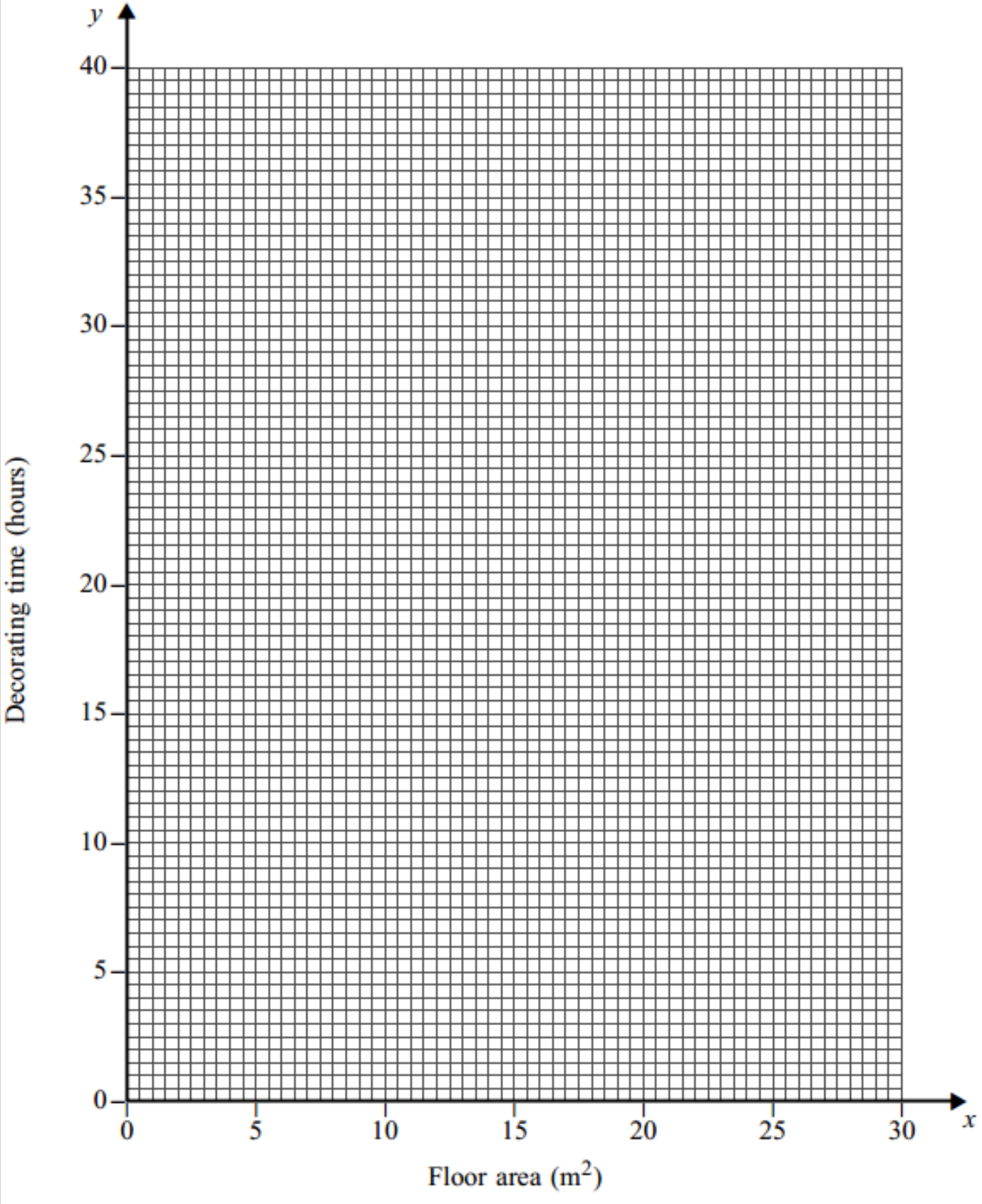
(c) Draw your regression line on **Figure 1**. *(2 marks)*

(d) (i) Use your regression equation to estimate the time that Roseen will take to decorate a bedroom with a floor area of $15\,\text{m}^2$. *(2 marks)*

(ii) Making reference to **Figure 1**, comment on the likely reliability of your estimate in part (d)(i). *(2 marks)*

Figure 1 (for use in Question 4)

Floor Areas and Decorating Times



June 2008

1 The table shows the times taken, y minutes, for a wood glue to dry at different air temperatures, x °C.

x	10	12	15	18	20	22	25	28	30
y	42.9	40.6	38.5	35.4	33.0	30.7	28.0	25.3	22.6

(a) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(b) Estimate the time taken for the glue to dry when the air temperature is 21 °C. *(2 marks)*

January 2009

6 [**Figure 1**, printed on the insert, is provided for use in this question.]

For a random sample of 10 patients who underwent hip-replacement operations, records were kept of their ages, x years, and of the number of days, y, following their operations before they were able to walk unaided safely.

Patient	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
x	55	51	62	66	72	59	78	55	62	70
y	34	33	39	49	48	43	51	41	46	51

(a) On **Figure 1**, complete the scatter diagram for these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(c) Draw your regression line on **Figure 1**. *(2 marks)*

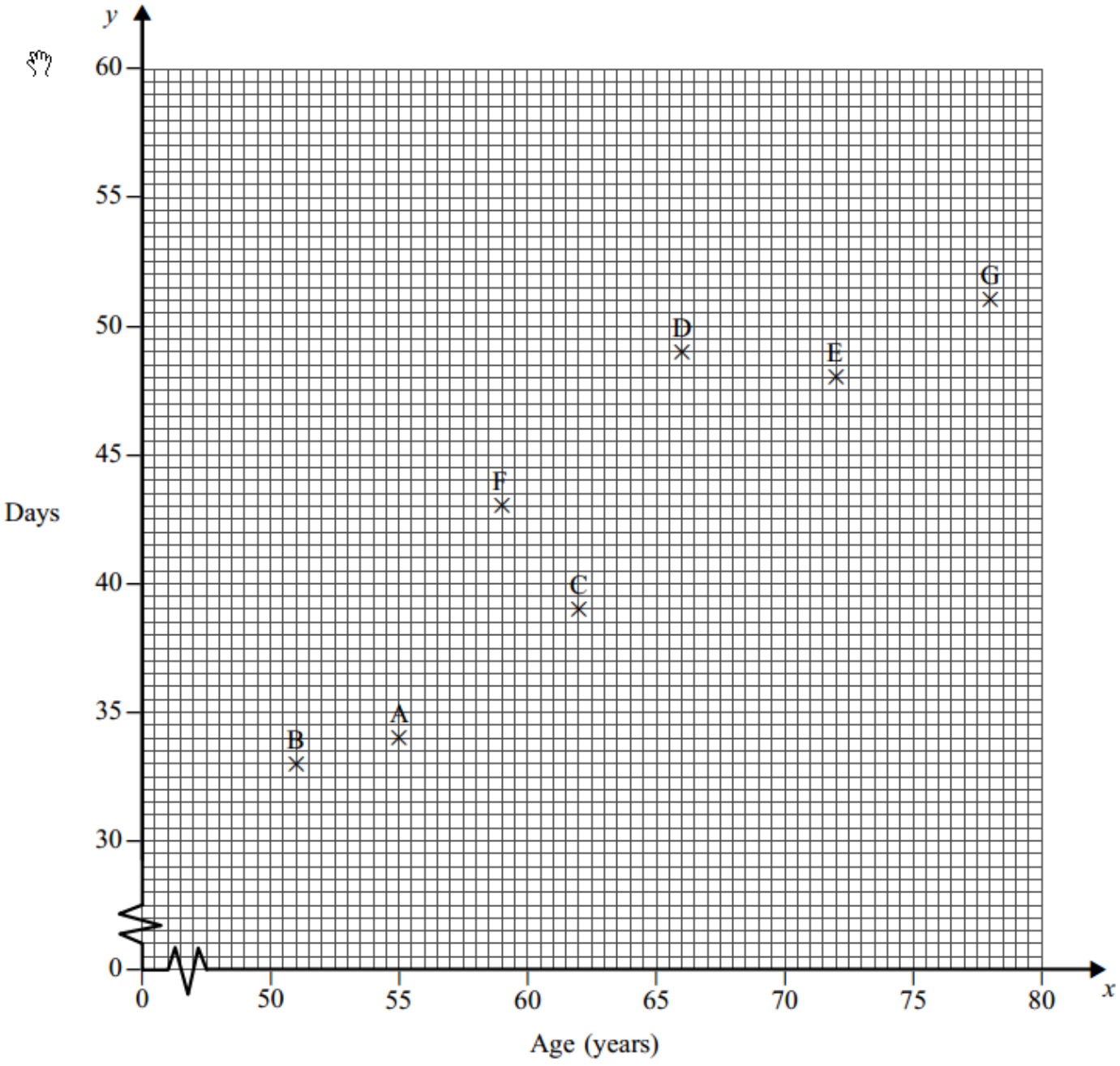
(d) In fact, patients H, I and J were males and the other 7 patients were females.

(i) Calculate the mean of the residuals for the 3 male patients. *(4 marks)*

(ii) Hence estimate, for a male patient aged 65 years, the number of days following his hip-replacement operation before he is able to walk unaided safely. *(3 marks)*

Figure 1 (for use in Question 6)

Hip-Replacement Operations



June 2009

4 As part of an investigation, a chlorine block is immersed in a large tank of water held at a constant temperature. The block slowly dissolves, and its weight, y grams, is noted x days after immersion. The results are shown in the table.

x days	5	10	15	20	30	40	50	60	75
y grams	47	44	42	38	35	27	23	16	9

(a) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(b) Hence estimate, to the nearest gram, the initial weight of the block. *(1 mark)*

(c) A company which markets the chlorine blocks claims that a block will usually dissolve completely after about 13 weeks. Comment, with justification, on this claim. *(3 marks)*

January 2010

3 The table shows, for each of a random sample of 7 weeks, the number of customers, x, who purchased fuel from a filling station, together with the total volume, y litres, of fuel purchased by these customers.

x	230	184	165	147	241	174	210
y	4551	3410	3252	3756	3787	4024	4254

(a) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(b) Estimate the volume of fuel sold during a week in which 200 customers purchase fuel. *(2 marks)*

(c) Comment on the likely reliability of your estimate in part (b), given that, for the regression line calculated in part (a), the values of the 7 residuals lie between approximately −415 litres and +430 litres. *(2 marks)*

6 During a study of reaction times, each of a random sample of 12 people, aged between 40 and 80 years, was asked to react as quickly as possible to a stimulus displayed on a computer screen.

Their ages, x years, and reaction times, y milliseconds, are shown in the table.

Person	Age (x years)	Reaction time (y ms)
A	41	520
B	54	750
C	66	650
D	72	920
E	71	280
F	57	620
G	60	740
H	47	950
I	77	970
J	65	780
K	51	550
L	59	730

(a) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(b) (i) Draw your regression line on the scatter diagram on page 16. *(2 marks)*

(ii) Comment on what this reveals. *(2 marks)*

(c) It was later discovered that the reaction times for persons E and H had been recorded incorrectly. The values should have been 820 and 590 respectively.

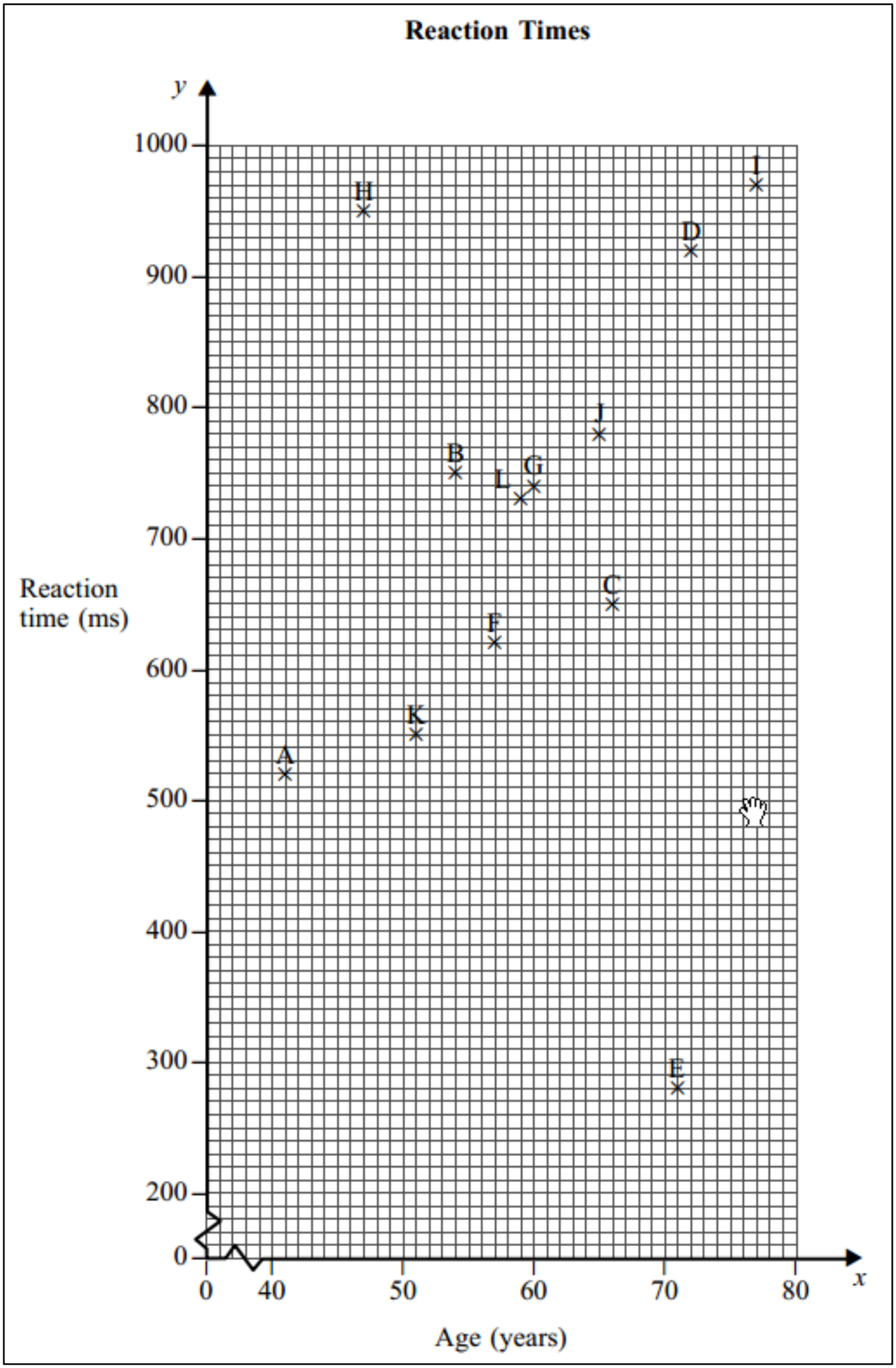
After making these corrections, computations gave

$$S_{xx} = 1272 \qquad S_{xy} = 14\,760 \qquad \bar{x} = 60 \qquad \bar{y} = 720$$

(i) Using the symbol ⊙, plot the correct values for persons E and H on the scatter diagram on page 16. *(1 mark)*

(ii) Recalculate the equation of the least squares regression line of y on x, and draw this regression line on the scatter diagram on page 16. *(3 marks)*

(iii) Hence revise as necessary your comments in part **(b)(ii)**. *(2 marks)*


Reaction Times
y
1000
900
800
700
600
500
400
300
200
0
Reaction time (ms)
H
I
D
J
B
G
L
C
F
K
A
E
0
40
50
60
70
80
x
Age (years)

January 2011

5 Craig uses his car to travel regularly from his home to the area hospital for treatment. He leaves home at x minutes after 7.30 am and then takes y minutes to arrive at the hospital's reception desk.

His results for 11 mornings are shown in the table.

x	0	5	10	15	20	25	30	35	40	45	50
y	31	42	32	58	47	56	79	68	89	95	85

(a) Explain why the time taken by Craig between leaving home and arriving at the hospital's reception desk is the response variable. *(1 mark)*

(b) Calculate the equation of the least squares regression line of y on x, writing your answer in the form $y = a + bx$. *(5 marks)*

(c) On a particular day, Craig needs to arrive at the hospital's reception desk no later than 9.00 am. He leaves home at 7.45 am.

Estimate the number of minutes **before** 9.00 am that Craig will arrive at the hospital's reception desk. Give your answer to the nearest minute. *(5 marks)*

(d) (i) Use your equation to estimate y when $x = 85$. *(1 mark)*

(ii) Give **one** statistical reason and **one** reason based on the context of this question as to why your estimate in part **(d)(i)** is unlikely to be realistic. *(2 marks)*

June 2011

3 (a) During a particular summer holiday, Rick worked in a fish and chip shop at a seaside resort.

He suspected that the shop's takings, £y, on a weekday were dependent upon the forecast of that day's maximum temperature, x °C, in the resort, made at 6.00 pm on the previous day.

To investigate this suspicion, he recorded values of x and y for a random sample of 7 weekdays during July.

x	23	18	27	19	25	20	22
y	4290	3188	5106	3829	5057	4264	4485

(i) Calculate the equation of the least squares regression line of y on x. *(4 marks)*

(ii) Estimate the shop's takings on a weekday during July when the maximum temperature was forecast to be 24 °C. *(2 marks)*

(iii) Explain why your equation may not be suitable for estimating the shop's takings on a weekday during February. *(1 mark)*

(iv) Describe, in the context of this question, a variable other than the maximum temperature, x, that may affect y. *(1 mark)*

3 (b) Seren, who also worked in the fish and chip shop, investigated the possible linear relationship between the shop's takings, £z, recorded in £000s, and **each** of two other explanatory variables, v and w.

(i) She calculated correctly that the regression line of z on v had a z-intercept of -1 and a gradient of 0.15.

Draw this line, for values of v from 0 to 40, on **Figure 1** below.

(ii) She also calculated correctly that the regression line of z on w had a z-intercept of 5 and a gradient of -0.40.

Draw this line, for values of w from 0 to 10, on **Figure 2** opposite. *(3 marks)*

Figure 1

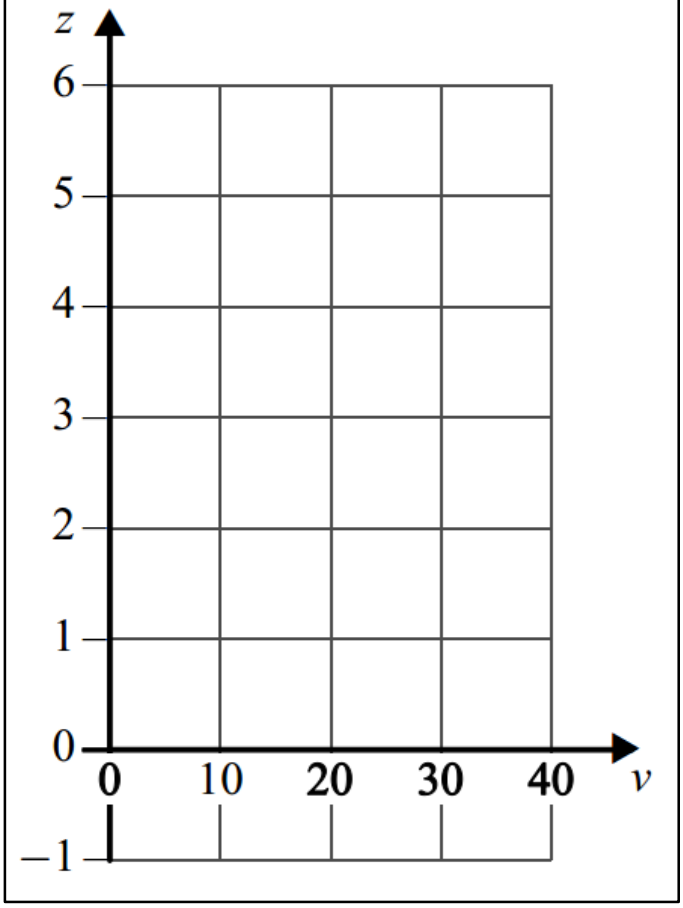
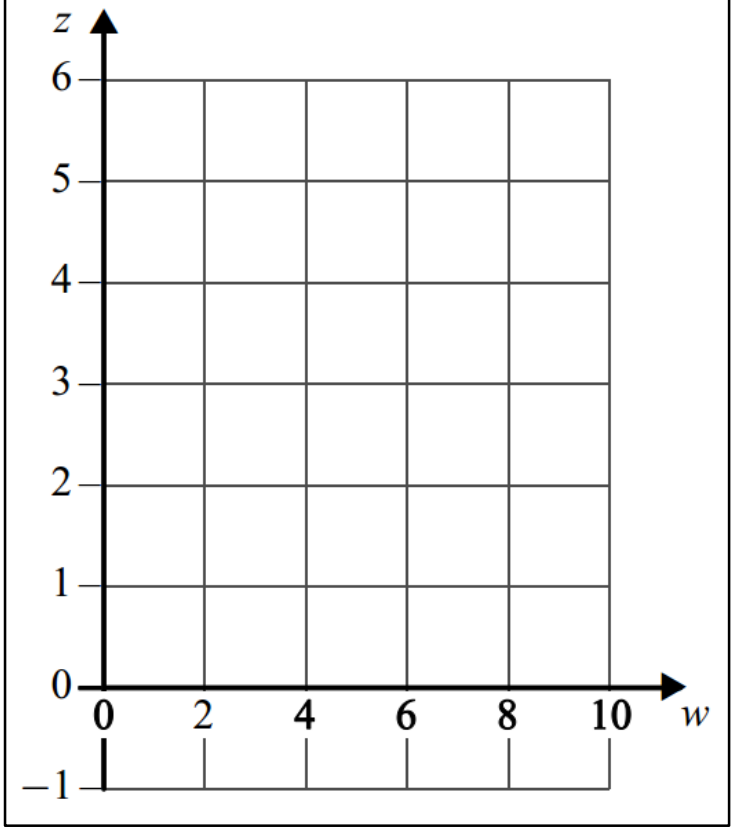


Figure 2



5 An experiment was undertaken to collect information on the burning of a specific type of wood as a source of energy. At given fixed levels of the wood's moisture content, x per cent, its corresponding calorific value, y MWh/tonne, on burning was determined. The results are shown in the table.

x	5	10	15	20	25	30	35	40	45	50	55	60	65
y	5.2	4.7	4.3	4.0	3.2	2.8	2.5	2.2	1.8	1.5	1.3	1.0	0.6

(a) Explain why calorific value is the response variable. *(1 mark)*

(b) Calculate the equation of the least squares regression line of y on x, giving your answer in the form $y = a + bx$. *(5 marks)*

(c) Interpret, in context, your values for a and b. *(3 marks)*

(d) Use your equation to estimate the wood's calorific value when it has a moisture content of 27 per cent. *(2 marks)*

(e) Calculate the value of the residual for the point $(35, 2.5)$. *(2 marks)*

(f) Given that the values of the 13 residuals lie between -0.28 and $+0.23$, comment on the likely accuracy of your estimate in part **(d)**. *(1 mark)*

(g) (i) Give a **general reason** why your equation should not be used to estimate the wood's calorific value when it has a moisture content of 80 per cent. *(1 mark)*

(ii) Give a **specific reason**, based on the context of this question and with numerical support, why your equation cannot be used to estimate the wood's calorific value when it has a moisture content of 80 per cent. *(2 marks)*

June 2012

3 The table shows the maximum weight, y_A grams, of *Salt A* that will dissolve in 100 grams of water at various temperatures, x °C.

x	10	15	20	25	30	35	40	45	50	60	70	80
y_A	20	35	48	57	77	92	101	111	121	137	159	182

(a) Calculate the equation of the least squares regression line of y_A on x. *(4 marks)*

(b) The data in the above table are plotted on the scatter diagram on the opposite page.

Draw your regression line on this scatter diagram. *(2 marks)*

(c) For water temperatures in the range 10 °C to 80 °C, the maximum weight, y_B grams, of *Salt B* that will dissolve in 100 grams of water is given by the equation

$$y_B = 60.1 + 0.255x$$

(i) Draw this line on the scatter diagram. *(2 marks)*

(ii) Estimate the water temperature at which the maximum weight of *Salt A* that will dissolve in 100 grams of water is the same as that of *Salt B*. *(1 mark)*

(iii) For *Salt A* and *Salt B*, compare the effects of water temperature on the maximum weight that will dissolve in 100 grams of water. Your answer should identify **two** distinct differences. *(2 marks)*

January 2013

1 Bob, a church warden, decides to investigate the lifetime of a particular manufacturer's brand of beeswax candle. Each candle is 30 cm in length.

From a box containing a large number of such candles, he selects one candle at random. He lights the candle and, after it has burned continuously for x hours, he records its length, y cm, to the nearest centimetre. His results are shown in the table.

x	5	10	15	20	25	30	35	40	45
y	27	25	21	19	16	11	9	5	2

(a) State the value that you would **expect** for a in the equation of the least squares regression line, $y = a + bx$. *(1 mark)*

(b) (i) Calculate the equation of the least squares regression line, $y = a + bx$. *(4 marks)*

(ii) Interpret the value that you obtain for b. *(2 marks)*

(iii) It is claimed by the candle manufacturer that the total length of time that such candles are likely to burn for is more than 50 hours.

Comment on this claim, giving a numerical justification for your answer. *(2 marks)*

4 The girth, g metres, the length, l metres, and the weight, y kilograms, of each of a sample of 20 pigs were measured.

The data collected is summarised as follows.

$$S_{gg} = 0.1196 \quad S_{ll} = 0.0436 \quad S_{yy} = 5880 \quad S_{gy} = 24.15 \quad S_{ly} = 10.25$$

(a) Calculate the value of the product moment correlation coefficient between:

(i) girth and weight;

(ii) length and weight. *(3 marks)*

(b) Interpret, in context, **each** of the values that you obtained in part **(a)**. *(3 marks)*

(c) Weighing pigs requires expensive equipment, whereas measuring their girths and lengths simply requires a tape measure. With this in mind, the following formula is proposed to make an estimate of a pig's weight, x kilograms, from its girth and length.

$$x = 69.3 \times g^2 \times l$$

Applying this formula to the relevant data on the 20 pigs resulted in

$$S_{xx} = 5656.15 \quad S_{xy} = 5662.97$$

(i) By calculating a third value of the product moment correlation coefficient, state which of g, l or x is the most strongly correlated with y, the weight. *(2 marks)*

(ii) Estimate the weight of a pig that has a girth of 1.25 metres and a length of 1.15 metres. *(2 marks)*

(iii) Given the additional information that $\bar{x} = 115.4$ and $\bar{y} = 116.0$, calculate the equation of the least squares regression line of y on x, in the form $y = a + bx$. *(3 marks)*

(iv) Comment on the likely accuracy of the estimated weight found in part **(c)(ii)**. Your answer should make reference to the value of the product moment correlation coefficient found in part **(c)(i)** and to the values of b and a found in part **(c)(iii)**. *(4 marks)*