

Statistics 1 Correlation and Regression Questions

- 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)
-

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

The table shows the times, in seconds, taken by a random sample of 10 boys from a junior swimming club to swim 50 metres freestyle and 50 metres backstroke.

Boy	A	B	C	D	E	F	G	H	I	J
Freestyle (x seconds)	30.2	32.8	25.1	31.8	31.2	35.6	32.4	38.0	36.1	34.1
Backstroke (y seconds)	33.5	35.4	37.4	27.2	34.7	38.2	37.7	41.4	42.3	38.4

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

(b) Hence:

(i) give **two** distinct comments on what your scatter diagram reveals; (2 marks)

(ii) state, **without calculation**, which of the following 3 values is most likely to be the value of the product moment correlation coefficient for the data in your scatter diagram.

0.912 0.088 0.462 (1 mark)

(c) In the sample of 10 boys, one boy is a junior-champion freestyle swimmer and one boy is a junior-champion backstroke swimmer.

Identify the **two** most likely boys. (2 marks)

(d) **Removing the data for the two boys whom you identified in part (c):**

(i) calculate the value of the product moment correlation coefficient for the remaining 8 pairs of values of x and y ; (3 marks)

(ii) comment, in context, on the value that you obtain. (1 mark)

- 1 The table shows, for each of a random sample of 8 paperback fiction books, the number of pages, x , and the recommended retail price, $\pounds y$, to the nearest 10p.

x	223	276	374	433	564	612	704	766
y	6.50	4.00	5.50	8.00	4.50	5.00	8.00	5.50

- (a) (i) Calculate the value of the product moment correlation coefficient between x and y .
(3 marks)
- (ii) Interpret your value in the context of this question.
(2 marks)
- (iii) Suggest one other variable, in addition to the number of pages, which may affect the recommended retail price of a paperback fiction book.
(1 mark)
- (b) The same 8 books were later included in a book sale. The value of the product moment correlation coefficient between the number of pages and the sale price was 0.959, correct to three decimal places.

What can be concluded from this value? (2 marks)

- 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

$$y = 265 + bx$$

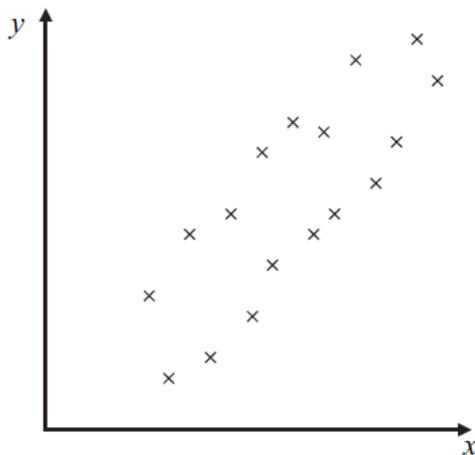
The manufacturer also states that, as the car is used, the tyre pressure will decrease at twice the rate of that found in part (a).

- (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.

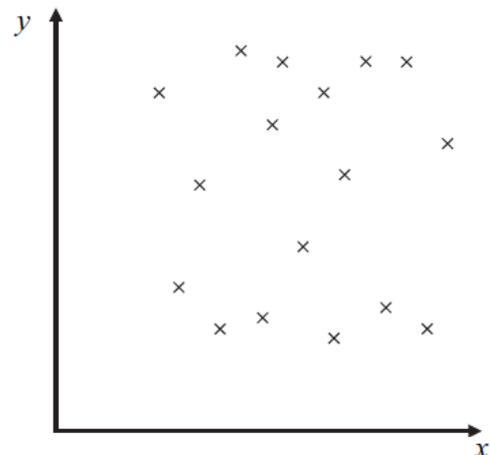
Show that, accepting the manufacturer's statements, the tyre pressure can be expected to have fallen below its minimum safety value of 220 kPa. (2 marks)

- 3 Estimate, **without undertaking any calculations**, the value of the product moment correlation coefficient between the variables x and y in each of the three scatter diagrams.

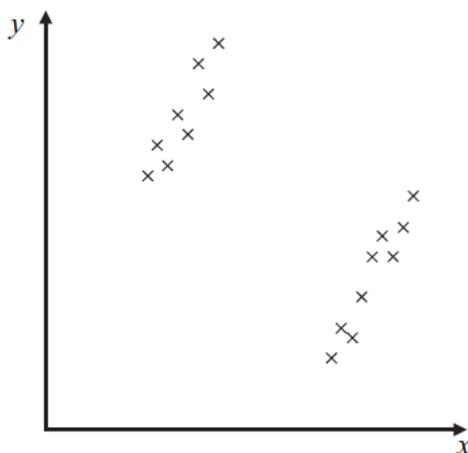
(a)



(b)



(c)



(5 marks)

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 m^2 . *(4 marks)*
-

1 The table shows the length, in centimetres, and maximum diameter, in centimetres, of each of 10 honeydew melons selected at random from those on display at a market stall.

Length	24	25	19	28	27	21	35	23	32	26
Maximum diameter	18	14	16	11	13	14	12	16	15	14

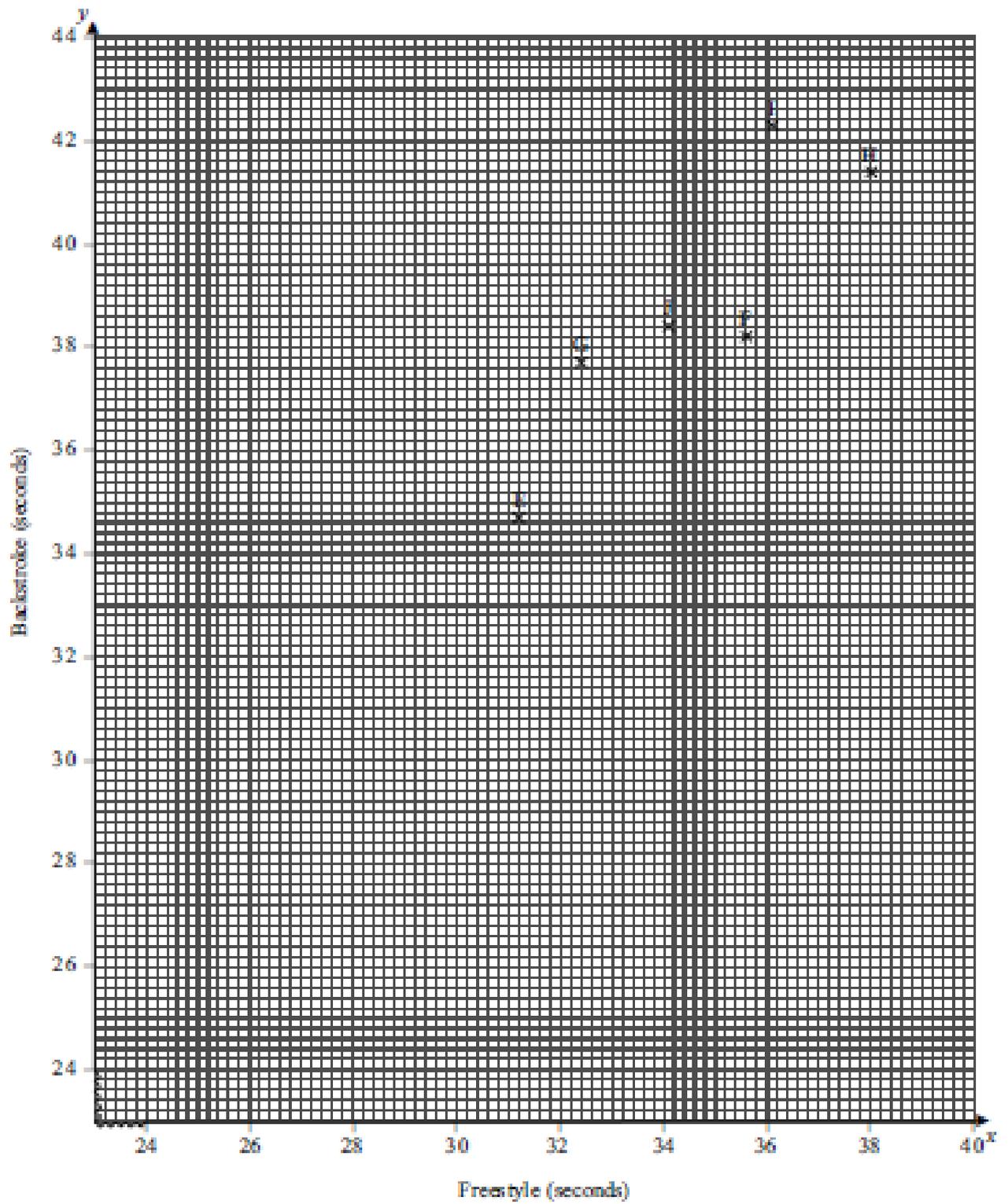
- (a) Calculate the value of the product moment correlation coefficient. *(3 marks)*
- (b) Interpret your value in the context of this question. *(2 marks)*
-

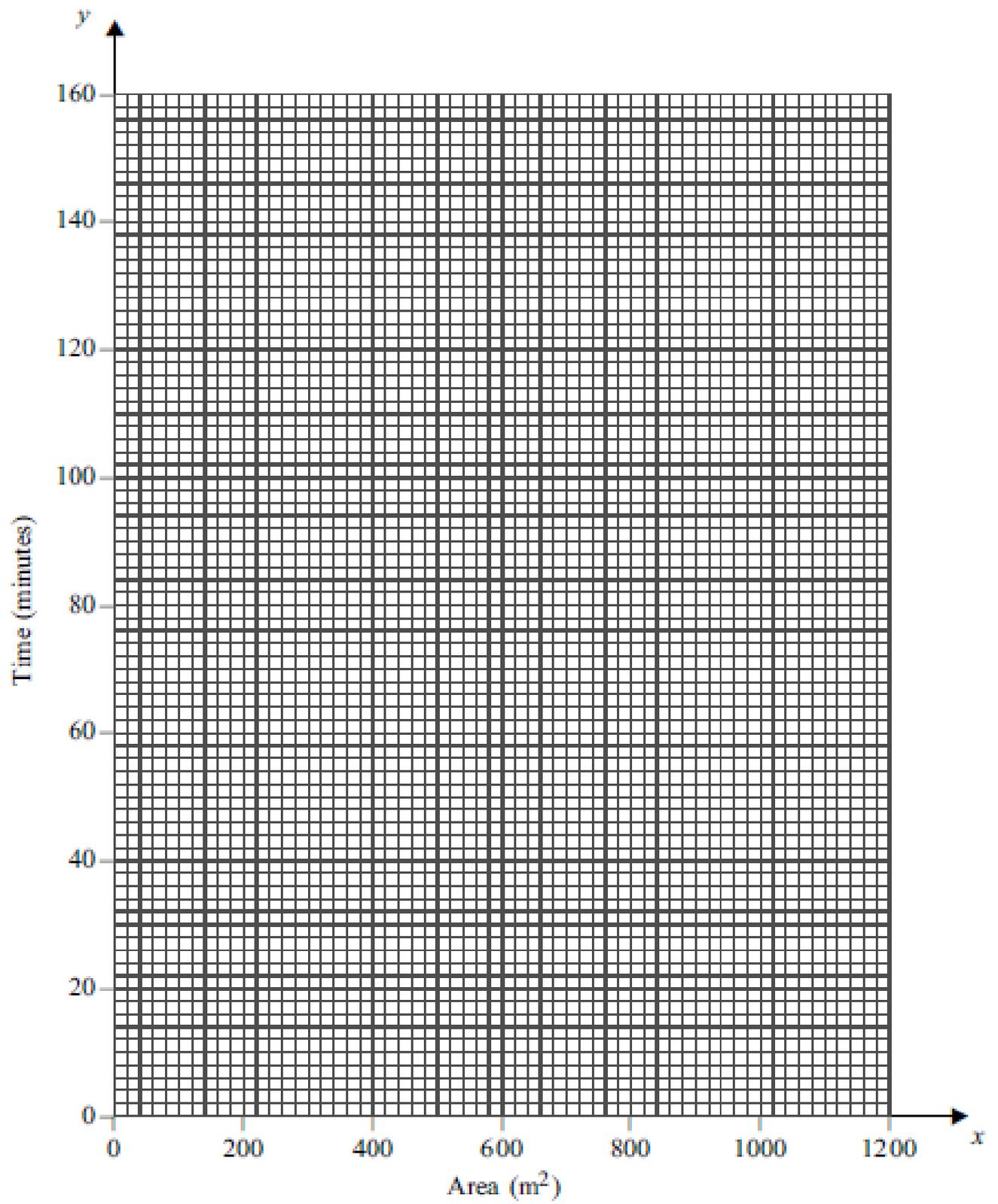
- 5 Bob, a gardener, measures the time taken, y minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, x °C. His results are shown in the table.

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)*
-

Figure 1 (for use in Question 5)
Scatter Diagram for Freestyle and Backstroke Swimming Times



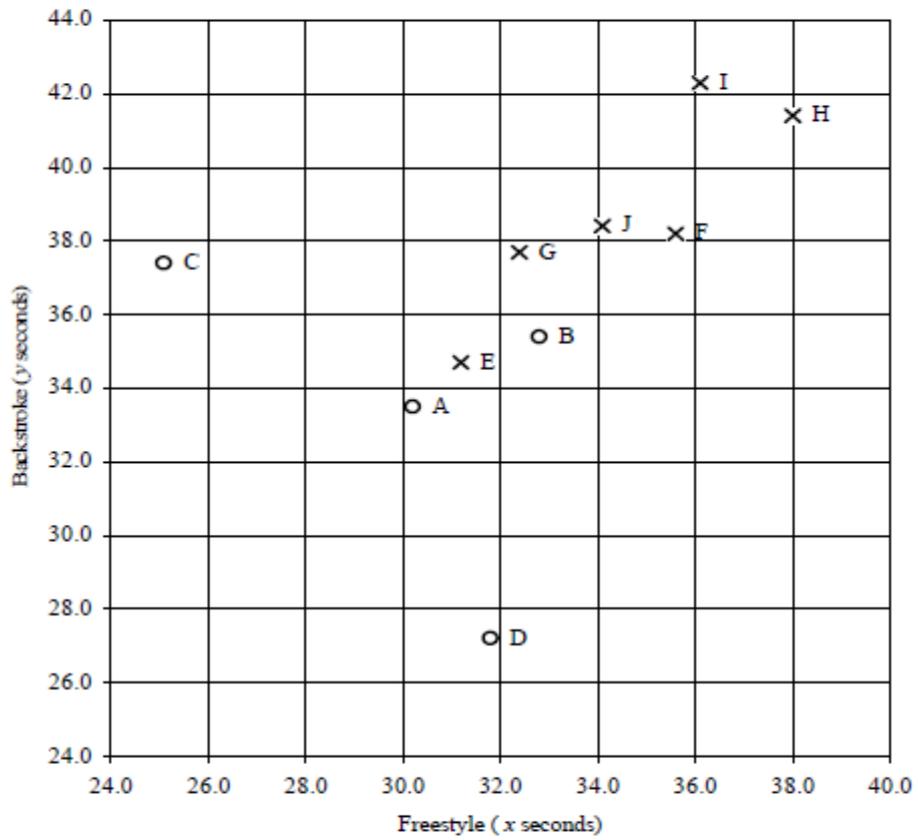


Statistics 1 Correlation and Regression Answers

1(a)	Gradient, $b = 0.886$ to 0.887 $b = 0.88$ to 0.89 Intercept, $a = 2.31$ to 2.33 $a = 2.3$ Attempt at Σx Σx^2 Σy Σxy or Attempt at S_{xx} S_{xy} Attempt at a correct formula for b $b = 0.886$ to 0.887 $a = 2.31$ to 2.33 Accept a & b interchanged only if $y = ax + b$ stated or subsequently used correctly in either (b) or (c)	B2 (B1) B2 (B1) (M1) (m1) (A1) (A1)		AFWW AFWW AFWW AWRW 72, 624, 87, 720 105.6, 93.6 AFWW AFWW 4
(b)	a: average waiting time of 2.32 minutes (139 seconds) when entering empty restaurant b: average increase in waiting time of 0.886 minutes (53 seconds) for each customer in restaurant on entry	B1 B1	2	OE; accept minimum waiting time OE
(c)	Use of $y = a + 5b$ or $y = a + 25b$	M1		
(i)	For $x = 5$ $y = 6.6$ to 6.8			
(ii)	For $x = 25$ $y = 24.3$ to 24.6	A1	2	Both; AFWW
(d)(i)	Reliable as interpolation and small residuals or Reliable as interpolation but large percentage residuals so inconclusive or Large percentage residuals so unreliable	B1 B1 (B1) (B1) (B1)		Within range OE OE
(ii)	Unreliable as extrapolation	B1	3	Outside range OE
Total			11	

5(a)	Scatter Diagram or or	B2 (B1) (B1)	2	4 labelled points plotted 3 labelled points plotted 4 unlabelled points plotted
(b)(i)	Positive/linear correlation/relationship except for two unusual values/results	B1 B1	 2	OE OE
(ii)	0.462	B1	1	CAO; accept 3 rd /final/last value
(c)	C and D C is likely freestyle champion D is likely backstroke champion or C is likely freestyle champion D is likely backstroke champion	B1 B1 (B1) (B1)	 2	CAO Style identified
(d)(i)	 or or Attempt at or Attempt at Attempt at a correct formula for r $r = 0.912$ to 0.913	 B3 B2 B1 (M1) (m1) A1	 3	AWFW AWFW AWRT 270.4, 9188.46 301.6, 11437.84 10246.53 48.94, 67.52, 52.45 AWFW
(ii)	Boys are faster/slower at both strokes or Boys are equally good at both strokes	B1	1	OE;do not accept freestyle times are proportional to backstroke times
Total			11	

Swimming Times



(a) Scatter Diagram

4 labelled points plotted	B2
3 labelled points plotted	(B1)
4 unlabelled points plotted	(B1)

Graph = 2

1(a)(i)	$r = 0.143$ to 0.1432	B3		AWFW
	or			
	$r = 0.142$ to 0.144	B2		AWFW
	or			
	$r = 0.1$ to 0.2	B1		AWRT
Attempt at	Σx Σx^2 Σy Σy^2 Σxy			3952, 2228282 47.00, 292.0000 23517.50
or				
Attempt at	S_{xx} S_{yy} S_{xy}	M1		275994, 15.875, 299.5
Attempt at a correct formula for r		m1		
$r = 0.143$ to 0.1432		A1	3	AWFW
(ii)	Little/weak/no correlation/relationship/association between number of pages and (retail) price	B1		or equivalent; but not poor
		B1	2	context
(iii)	Size (page, thickness), author, ranking, publicity/marketing, cover design, recommendations on back, publisher, font, popularity, quality, print-run, etc	B1	1	or any sensible variable but not pictures, coloured pictures, age, words, weight, mass
(b)	(Very) strong/almost exact positive/perfect correlation/relationship/association between number of pages and sale/new price Sale price appears to be determined by number of pages	B1		or equivalent
		B1	2	context
		B2		or equivalent
Total			8	

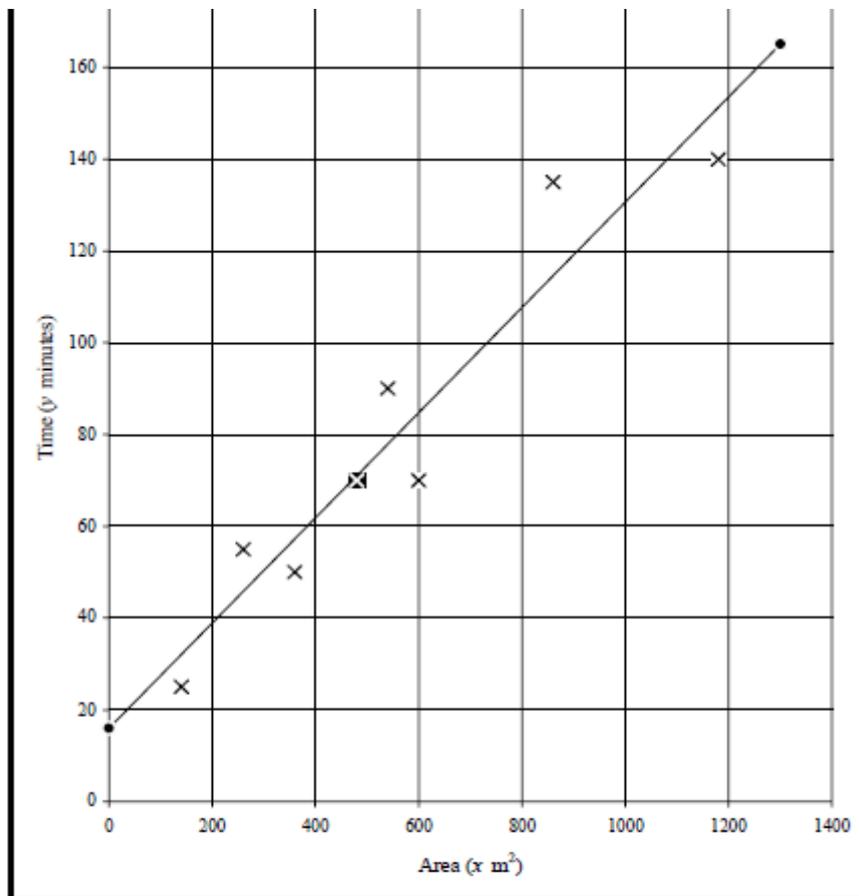
3(a)(i)	Gradient, $b = -3.24$ to -3.26 $b = -3.2$ to -3.3	B2		AWFW	(-3.25)
		B1		AWFW	
	Intercept, $a = 262$ to 264 $a = 260$ to 270	B2		AWFW	(262.88)
		B1		AWFW	
	Attempt at Σx Σx^2 Σy Σxy or Attempt at S_{xx} S_{yy} Attempt at a correct formula for b $b = -3.24$ to -3.26 $a = 262$ to 264				108, 1836, 2015, 22425
	M1 m1 A1 A1			540, -1755	
			4	AWRT AWFW	
	Accept a & b interchanged only if identified correctly in (b) and (c)				
(ii)	Gradient, b : Decrease in pressure per month Change in pressure	B2		or equivalent	
		B1	2	or better	
(iii)	Intercept, a : Initial pressure or pressure at $x = 0$ Reference to 265, actual or expected value	B1		or equivalent; not y-intercept	
		B1	2		

(b)(i)	Value for $b = 2 \times$ [gradient or b from (a)(i)]	M1		accept $2b$; ignore sign
	$= -6.4$ to -6.6	A1✓	2	AWFW (-6.5) ✓ from (a)(i) but must be < 0
(ii)	$P_8 = 265 - 6.5 \times 8$	M1		must use 265 and $x = 8$ and $2 \times [b (< 0)$ from (a)(i)] or [from (b)(i) (< 0)]
	$= 212$ to 214 (< 220)	A1	2	AWFW AG
Total			12	

3(a)	$0.5 \leq \text{Value} \leq 0.95$ Positive value < 1 (and > 0)	B2 (B1)		Value is actually 0.8
(b)	$-0.2 \leq \text{Value} \leq +0.2$	B1		Value is actually 0.0
(c)	$-0.95 \leq \text{Value} \leq -0.5$ Negative value > -1 (and < 0)	B2 (B1)	5	Value is actually -0.7
Total			5	

7(a)	8 or 7 points plotted accurately (6 or 5 points plotted accurately)	B2 (B1)	2	
(b)	Gradient, $b = 0.114$ to 0.115 ($b = 0.11$ to 0.12)	B2 (B1)		AWFW (0.11469)
	Intercept, $a = 15.9$ to 16.1 ($a = 13$ to 19)	B2 (B1)		AWFW (16.00824)
	Attempt at $\sum x$, $\sum x^2$, $\sum y$ and $\sum xy$ or Attempt at S_{xx} and S_{xy}	(M1)		4420, 3230800, 635 and 441300 788750 and 90462.5
	Attempt at correct formula for b $b = 0.114$ to 0.115 $a = 15.9$ to 16.1	(m1) (A1) (A1)		AWFW AWFW
	Accept a and b interchanged only if then identified correctly later in question			
Line plotted accurately (Evidence of correct method for ≥ 2 points)	B2 (M1)	6	At least from $x = 200$ to 1000	

(c)	$Res_H = y_H - Y_H = 70 - (a + b \times 480)$	M1		Used; or implied by correct answer; allow for $Y_H - y_H$ shown
	$= -1.5$ to -0.5	A1		AWFW (-1.06)
	Point H is (almost) on / just below the line	B1	3	Accept near / close / just above or equivalent
(d)	$Y = a + b \times 560$ or reading from scatter diagram	M1		Used
	$= 79$ to 81	A1		AWFW (80.2)
	Cost = $Y \times \frac{12}{60}$ or $\frac{Y}{5}$	M1		Used
	$= \pounds 15.8$ to $\pounds 16.2$	A1	4	AWFW; ignore units (£16.05)
Total			15	



- | | | |
|--------------------|--|-----|
| (a) | 8 or 7 points plotted accurately | B2 |
| | (6 or 5 points plotted accurately) | B1) |
| (b) | Line plotted accurately | B2 |
| | (Evidence of correct method for ≥ 2 points) | M1) |
| (Graph = 4) | | |

1(a)	$r = -0.526$ to -0.525	B3	AWFW
	or $r = -0.53$ to -0.52	(B2)	AWFW; ignore sign
	or $r = -0.6$ to -0.4	(B1)	AWFW; ignore sign
	OR Attempt at $\sum x$, $\sum x^2$, $\sum y$, $\sum y^2$ and $\sum xy$	(M1)	260, 6970, 143, 2083 and 3671
	or Attempt at S_{xx} , S_{yy} and S_{xy}	(m1)	210, 38.1 and -47
Attempt at a correct formula for r	(m1)		
	$r = -0.526$ to -0.525	(A1)	3 AFWW
(b)	Weak/some/moderate negative correlation (relationship/association)	B1	OE; must qualify strength and indicate negative B0 for strong/poor/reasonable/average B0 if $r > 0$ or $r < -1$ B0 if contradictory statements
	between		
	length and (maximum) diameter	B1	Context
	Ignore subsequent comments (as below) only if B1 B1 already scored		
	OR		
Some evidence that large lengths are associated with small diameters	(B1) (B1)		OE; must qualify strength and indicate negative
OR			
Longer melons tend to have smaller diameters / be thinner	(B1) (B1)	2	OE; must qualify strength and indicate negative
	Total		5

5(a)	Time taken depends upon temperature	B1	1	OE; not x set values
(b)	b (gradient) = -0.0873 to -0.087	B2		AWFW $(-0.087\dot{2}\dot{7})$
	b (gradient) = -0.09 to -0.08	(B1)		AWFW; $-8.73^{-02} \Rightarrow B0$
	a (intercept) = 5.94 to 5.96	B2		AWFW $(5.95\dot{0}\dot{9})$
	a (intercept) = 5.6 to 6.1	(B1)		AWFW
	Attempt at $\sum x$, $\sum x^2$, $\sum y$ and $\sum xy$			396, 16016, 30.9 and 958.8
	or	(M1)		
	Attempt at S_{xx} and S_{xy}			1760 and -153.6
	Attempt at correct formula for b	(m1)		
	$b = -0.0873$ to -0.087	(A1)		AWFW
	$a = 5.94$ to 5.96	(A1)	4	AWFW
	Accept a and b interchanged only if then identified correctly later in question			
(c)(i)	Each 1°C rise in temperature results in an (average) decrease of 0.087 m (5 s) in time taken for pellets to dissolve	B1		Quantified rise in x (results in)
		B1	2	Decrease in y OE
(ii)	a is y -value at $x = 0$ at which water is solid/ice/frozen so pellets cannot dissolve	B1		Indication that it is y at $x = 0$
		B1	2	Mention of solid or ice or frozen
(ii)	a is y -value at $x = 0$ at which water is solid/ice/frozen so pellets cannot dissolve	B1		Indication that it is y at $x = 0$
		B1	2	Mention of solid or ice or frozen
(d)(i)	When $x = 30$			
	$y = 3.3$ to 3.4 $y = 2.9$ to 3.7	B2 (B1)		AWFW $(3.33\dot{2}\dot{7})$ AWFW
	If B0, use of their equation with $x = 30$	(M1)	2	
(ii)	When $x = 75$			
	$y < 0$ or negative which is impossible	B1 \uparrow Dep \uparrow B1		OE
		B1	2	OE; not extrapolation
	Total		13	