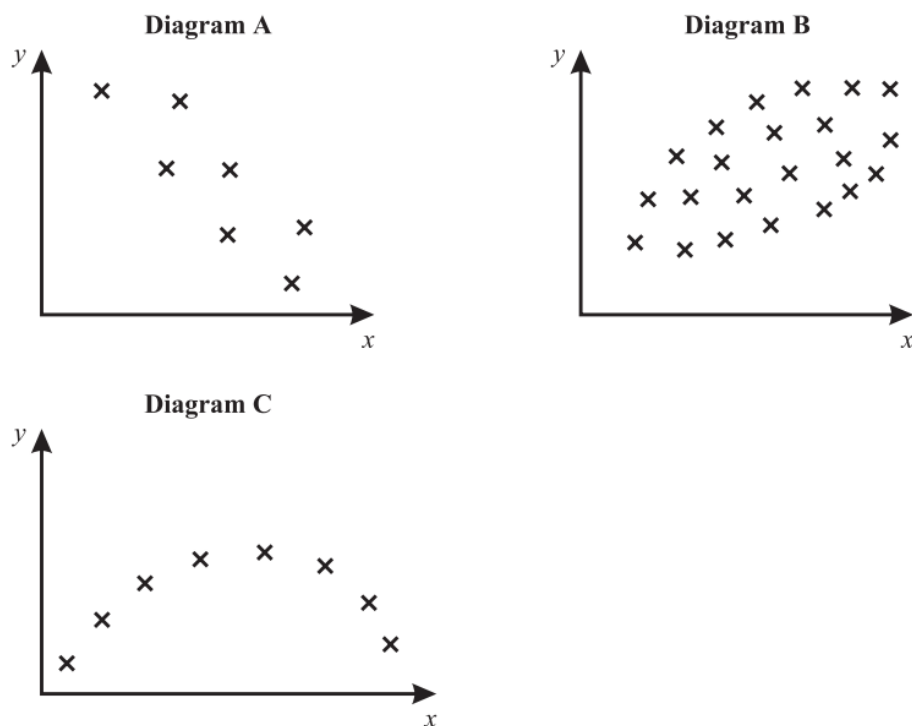**1** (a) Estimate, without undertaking any calculations, the value of the product moment correlation coefficient in each of the scatter diagrams below.

Diagram A

Diagram B

Diagram C

(5 marks)

(b) The product moment correlation coefficient is an unsuitable measure of association for the data illustrated in one of the diagrams in part (a). State, giving a reason, which diagram this is.
(2 marks)

Question	Solution	Marks	Total Marks	Comments
1 (a)(i)	A − 0.8	B1 B1		
(ii)	B 0.5	B1 B1		
(iii)	C 0.0	B1	(5)	
(b)	C is unsuitable, since relationship is clearly non-linear.	M1		
	PMCC measures strength of a linear relationship	A1	(2)	
		TOTAL	7	

(a)(i)	A	B1 negative (disallow if <-1)		
		B1 magnitude 0.8 (0.6 - 0.98)		
(ii)	B	B1 positive (disallow if >1)		
		B1 magnitude 0.5 (0.1 - 0.8 must be less than their A)		
(iii)	C	B1 0 (− 0.2 - 0.2)		(5)
(b)	M1	C cao		
	A1	not linear		(2)

6 [A sheet of graph paper is provided for use in this question.]

The following table gives information on the engine capacity and price of models of car in January 2000.

Model	A	B	C	D	E	F	G	H	I	J
Engine Capacity, x **(cc)**	1 000	1 270	1 750	2 230	1 990	600	650	1 500	1 450	1 650
Price, y **(£'s)**	8 800	9 400	11 320	14 960	14 860	5 980	6 400	9 320	7 100	13 300

(a) Plot a scatter diagram of these data. *(3 marks)*

(b) Find the equation of the regression line of price on engine capacity. Draw the line on your scatter diagram. *(7 marks)*

(c) Calculate the residuals for models **I** and **J**. *(3 marks)*

(d) Interpret, briefly, the residuals you have calculated for a customer who requires a model with a low price relative to its engine capacity. *(2 marks)*

Question	Solution	Marks	Total Marks	Comments
6 (a)	On graph paper	M1 B1 A1	(3)	
(b)	$y = 2315 + 5.556x$	B2 B2		
	$\left[\begin{array}{ll}\sum x^2 = 22\,465\,900 & \sum x = 14\,090 \\ \sum y = 101\,440 & \sum xy = 157\,448\,200\end{array}\right]$			
	$x = 0 \qquad y = 2315$	M1		
	$x = 2000 \qquad y = 13428 \qquad +$ line	M1 A1	(7)	
(c)	Residual for **I** $= 7100 - 2315 - 5.556 \times 1450 = -3270$	M1 A1 ✓		
	Residual for **J** $= 13\,700 - 2315 - 5.556 \times 1650 = 1817$	A1	(3)	
(d)	**I** has a large negative residual indicating price much lower than expected for a car with this engine capacity.	B1		
	J has fairly large residual indicating price higher than expected for this engine capacity.			
	For this customer **I** is a good buy, **J** is a bad buy.	B1	(2)	
		TOTAL	15	

(a)	M1	method	
	B1	scales shown and axes labelled	
	A1	accurate plot by eye – allow one small slip	**(3)**
(b)	B2	2315 (2310 - 2320)	
	B2	5.556 (5.55 - 5.56)	
		(if method shown use M1 A1 for each B2)	
	M1	two points on their line evaluated	
	M1	method for line	
	A1	accurate plot by eye	**(7)**
(c)	M1	method for residuals – previous M marks **not** required, ignore sign	
	A1 √	correct substitution their numbers	
	A1	− 3270 (− 3200 - − 3300) and 1817 (1770 - 1870)	**(3)**
(d)	B1	**I** – good buy	
	B1	**J** – bad buy	**(2)**

Question 6 (a)

M1 A1 B1 **(3)**
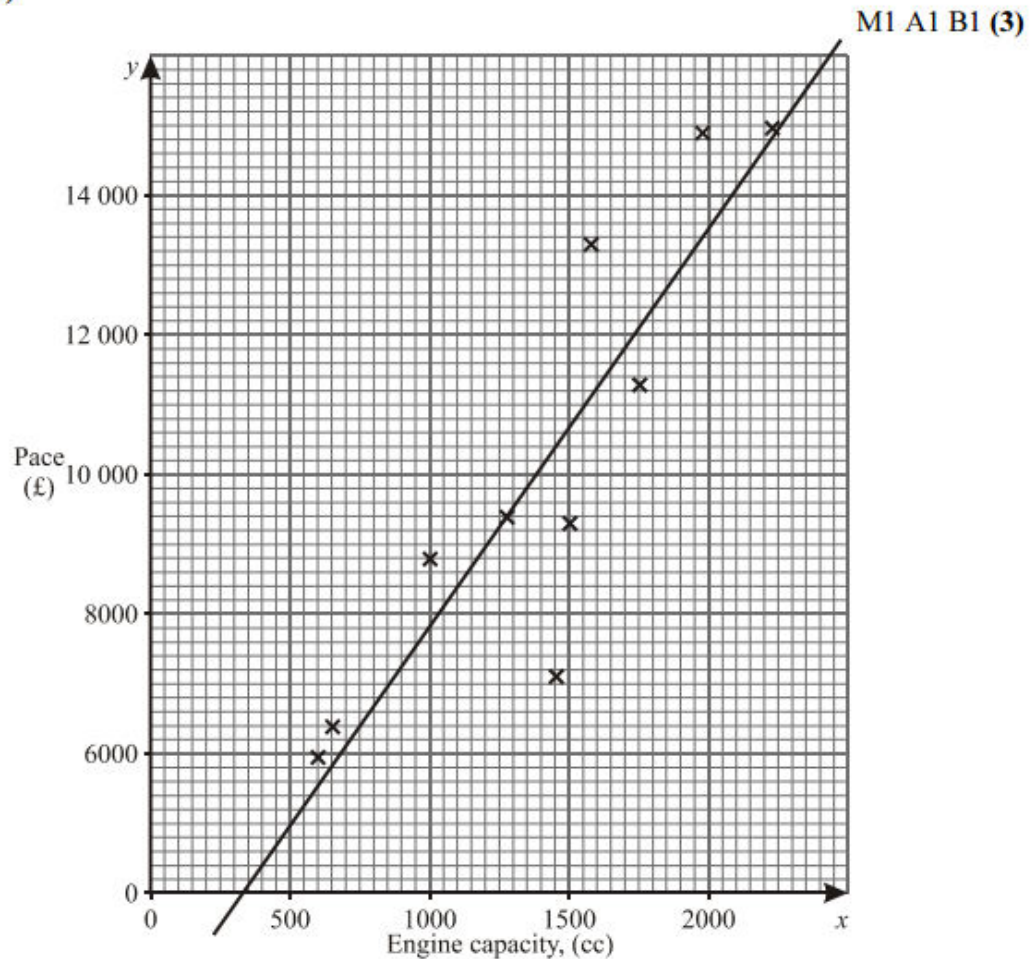
1 The following table shows the weekly gas and electricity consumption for a house in Manchester for a sample of nine weeks in 2001.

Week	1	2	3	4	5	6	7	8	9
Gas consumption, x kWh	312	46	23	406	350	67	295	247	110
Electricity consumption, y kWh	84	57	54	96	82	63	59	73	60

(a) Calculate the value of the product moment correlation coefficient. *(3 marks)*

(b) The householder expects that, for weeks when the gas consumption is high, the electricity consumption will be low and vice versa.

    (i) State to what extent the value you calculated in part (a) confirms or denies the householder's expectation. *(2 marks)*

    (ii) Give a reason why the value you calculated in part (a) is plausible. *(2 marks)*

Question Number and Part	Solution	Marks	Total Marks	Comments
1(a)	0.850	B3	3	0.850 $(0.849 - 0.851)$ allow M2 A1 if method shown.
(b)	high gas consumption associated with high electricity consumption, contrary to householder's expectation	E1 E1	2	inconsistent with expectations high gas associated with high electricity
(c)	Cold weather likely to lead to high consumption of both gas and electricity. Hot weather to low consumption	E2,1	2	Adequate reason, properly explained gets both marks. (E1 for comment on data)
	Total		7	

8 [A sheet of graph paper is provided for use in this question.]

Hennie, a statistics teacher, is an unenthusiastic gardener. During the summer months she records the time, y minutes, it takes her to cut her lawn together with the time, x days, since she last cut it.

x	5	13	10	19	7	12	21	18	24	12
y	23	32	28	48	21	35	44	39	50	39

(a) Plot a scatter diagram of the data. *(3 marks)*

(b) Find the equation of the regression line of y on x and plot it on your scatter diagram.
(6 marks)

(c) Give an interpretation, in the context of this question, of:

  (i) the gradient of the regression line;

  (ii) the intercept of the regression line with the y-axis. *(3 marks)*

(d) Hennie's brother Ludwig suggests to her that she could save time by cutting her lawn in spring and then waiting until the autumn before cutting it again. Give one statistical reason **and** one reason specific to this context why it would be unwise to use your regression equation to estimate the time it would take Hennie to cut her lawn 150 days since she last cut it. *(2 marks)*

(e) Explain why it was appropriate in this question to calculate the regression equation of y on x rather than that of x on y. *(2 marks)*

8(a)	**See attached diagram**	M1 B1 A1	3	Method for scatter diagram Scales and labels Reasonably accurate plot
(b)	$y = 14.6 + 1.51x$	B2		$14.6\ (14.5 - 14.6)$ allow M1 A1 if method shown
		B2		$1.51(1.5 - 1.52)$
	$x = 5 \quad y = 22.1 \qquad x = 25 \quad y = 52.4$	M1		method – their line
		A1	6	accurate line
(c)(i)	Estimate of additional time, min. to cut lawn for each additional day since it was last cut	E1 E1		Estimate/average Additional time per additional day
(ii)	Estimate of time to cut lawn a second time immediately after first cut	E1	3	Time for second cut
(d)	Extrapolation unwise. Might not be possible to use lawnmower on 150 days growth	E1 E1	2	Extrapolation Lawnmower wouldn't cut/lawn unuseable etc.
(e)	Time to cut lawn may be affected by time since last cut. Reverse cannot be true	E2(1)	2	Time to cut depends on time since last cut not vice versa
	Total		16	

Time to cut lawn (minutes)

Time since lawn last cut (days)