Statistics 1 – Correlation and Regression Exam Questions

5   [One sheet of graph paper is provided for use in answering this question.]

A biologist assumes that there is a linear relationship between the amount of fertilizer supplied to tomato plants and the subsequent yield of tomatoes obtained.

Eight tomato plants, of the same variety, were selected at random and treated, weekly, with a solution in which x grams of fertilizer was dissolved in a fixed quantity of water. The yield, y kilograms, of tomatoes was recorded.

Plant	A	B	C	D	E	F	G	H
x	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
y	3.9	4.4	5.8	6.6	7.0	7.1	7.3	7.7

(a)   Plot a scatter diagram of yield, y, against amount of fertilizer, x.   *(3 marks)*

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

(c)   Estimate the yield of a plant treated, weekly, with 3.2 grams of fertilizer.   *(2 marks)*

(d)   Indicate why it may **not** be appropriate to use your equation to predict the yield of a plant treated, weekly, with 20 grams of fertilizer.   *(1 mark)*

3   Over a period of one year, a greengrocer sells tomatoes at six different prices (x pence per kilogram). He calculates the average number of kilograms, y, sold per day at each of the six different prices. From these data the following are calculated.

$$\Sigma x = 200 \qquad \Sigma y = 436 \qquad \Sigma xy = 12\,515$$

$$\Sigma x^2 = 7250 \qquad \Sigma y^2 = 39\,234 \qquad n = 6$$

(a)   Calculate the value of the product moment correlation coefficient.   *(4 marks)*

1   The table shows a Verbal Reasoning test score, x, and an English test score, y, for each of a random sample of 8 children who took both tests.

Child	A	B	C	D	E	F	G	H
x	112	113	110	113	112	114	109	113
y	69	65	75	70	70	75	68	76

(a)   Calculate the value of the product moment correlation coefficient between the scores in Verbal Reasoning and English.   *(5 marks)*

(b)   Comment briefly, in context, on the result obtained in part (a).   *(2 marks)*

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

One end A of an elastic string was attached to a horizontal bar and a mass, m grams, was attached to the other end B. The mass was suspended freely and allowed to settle vertically below A. The length AB, l mm, was recorded, for various masses as follows.

m	100	200	300	400	500	600
l	228	236	256	278	285	301

(a) Draw a scatter diagram to illustrate the above information.   *(3 marks)*

(b) Calculate the least squares line of regression of l on m, and plot this line on your scatter diagram.   *(7 marks)*

(c) Give, in context, interpretations for:

    (i)   the gradient of the line;

    (ii) the intercept of the line on the l-axis.   *(2 marks)*

(d) Estimate the length of the string when a mass of 360 grams is attached at B.   *(1 mark)*

(e) State a physical limitation that there might be in using your equation to estimate the length of the string when a mass of 1200 grams is attached at B.   *(1 mark)*

1 A market trader sells ball-point pens on his stall. He sells the pens for a different fixed price, x pence, in each of six weeks. He notes the number of pens, y, that he sells in each of these six weeks. The results are shown in the following table.

x	10	15	20	25	30	35
y	68	60	55	48	38	32

Calculate the equation of the least squares regression line of y on x.   *(5 marks)*

2 A police authority conducts an eight week experiment. In each week it records the number of foot patrols, x, made in a small town and the number of reported crimes, y, in that town. The data are summarised as follows.

$$\sum x = 52 \qquad \sum x^2 = 380 \qquad \sum xy = 1335$$

$$\sum y = 225 \qquad \sum y^2 = 7007 \qquad n = 8$$

(a) Calculate the value of the product moment correlation coefficient for these data.
(5 marks)

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

A mathematics teacher recorded the length of time, y minutes, taken to travel to school when leaving home x minutes after 7 am on seven selected mornings. The results are as follows.

x	0	10	20	30	40	50	60
y	16	27	28	39	39	48	51

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

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

    (ii) Draw the regression line on your scatter diagram. *(1 mark)*

(c) The mathematics teacher needs to arrive at school no later than 8.40 am.

The number of minutes by which the mathematics teacher arrives early at school, when leaving home x minutes after 7 am, is denoted by z.

  (i) Deduce that
$$z = (100 - a) - (1 + b)x.$$
(3 marks)

  (ii) Hence estimate, to the nearest minute, the latest time that the mathematics teacher can leave home without then arriving late at school. *(3 marks)*

2 The following table shows, for a sample of towns in Great Britain, the number of solicitors, x, and the number of cars stolen last week, y.

x	12	7	11	19	5	21	3	4	17
y	14	3	21	28	6	43	1	12	30

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

    (ii) Interpret your result from part (a)(i) in the context of the question. *(1 mark)*

(b) Comment on the suggestion that most car thieves are solicitors. *(2 marks)*

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

The following table shows the hours of sunshine, *x*, during nine days in August and the number of ice creams, *y*, sold by a beach shop in Cornwall.

x	4.3	6.9	0.0	10.4	5.2	1.8	8.0	9.2	2.1
y	224	208	123	419	230	184	362	351	196

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

(b) Calculate the equation of the regression line of *y* on *x* and draw the line on your scatter diagram.
                                                                       (6 marks)

(c) Calculate the residuals for the days when the number of hours of sunshine was:

   (i) 8.0;

   (ii) 6.9.   *(3 marks)*

(d) On one of the days the shop closed early to allow the owner to attend a birthday party.

Suggest, giving a reason, which day this was.   *(2 marks)*

(e) The owner asks you to use the regression equation to forecast the daily sales if there were 20 hours of sunshine. Give **two** reasons why it would be inappropriate to do this.   *(2 marks)*

2 Henri and Michelle are two journalists who write regular newspaper columns advising readers which wines offer good value for money. They taste a number of wines and then estimate the retail prices of the corresponding bottles of wine. Their estimates are shown in the following table.

Wine	A	B	C	D	E	F	G	H	I
Henri's estimate (£)	7	3	25	50	2	5	17	80	12
Michelle's estimate (£)	9	19	20	25	9	10	4	26	60

(a) Calculate the value of the product moment correlation coefficient between Henri's and Michelle's estimates.   *(3 marks)*

(b) Interpret, briefly, your value of the correlation coefficient.   *(2 marks)*

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

Nasser organises a street collection for a mental health charity. The collection takes place in a large city on a particular Saturday. Volunteers, with collecting tins, stand in busy places and ask passers-by for donations. The following table shows, for ten volunteers, the times, x minutes, they spent collecting together with the amounts, to the nearest pound, y, they collected.

Collector	A	B	C	D	E	F	G	H	I	J
x	65	187	126	52	143	90	157	74	88	195
y	21	55	23	8	28	27	44	19	17	47

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

(b) Calculate the equation of the regression line of y on x and draw the line on your scatter diagram. *(6 marks)*

(c) The following table shows the residuals for some of the collectors.

Collector	A	B	C	D	E	F	G	H	I	J
Residual	6.25	7.50	−8.13	−3.26	−7.69	5.54	4.55	1.83		

    (i) Calculate the residuals for collectors I and J. *(3 marks)*

    (ii) Calculate the mean **magnitude** of the ten residuals. *(2 marks)*

(d) Karen, a new volunteer, collected for 110 minutes.

    (i) Use your regression equation to estimate the amount Karen collected. *(1 mark)*

    (ii) In fact Karen collected £22. Use the results of your calculations to advise Nasser on whether or not this suggests that Karen should have been supervised when collecting. Explain your answer. *(2 marks)*

1 Estimate, **without undertaking any calculation**, the value of the product moment correlation coefficient between the variables x and y in **each** of the scatter diagrams below.

(a)

(b)

(c)
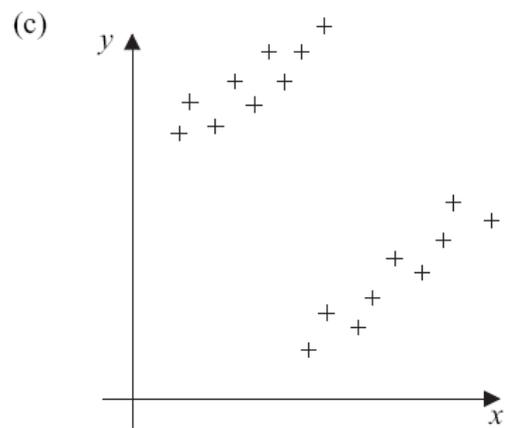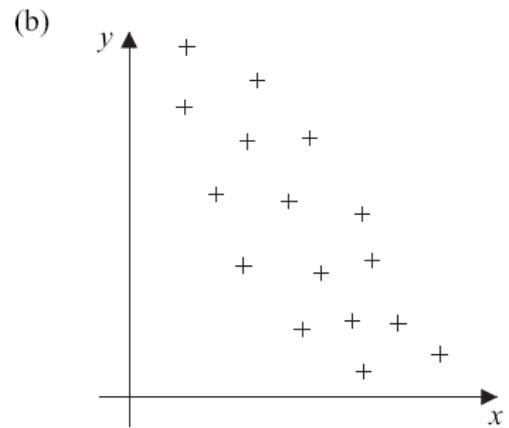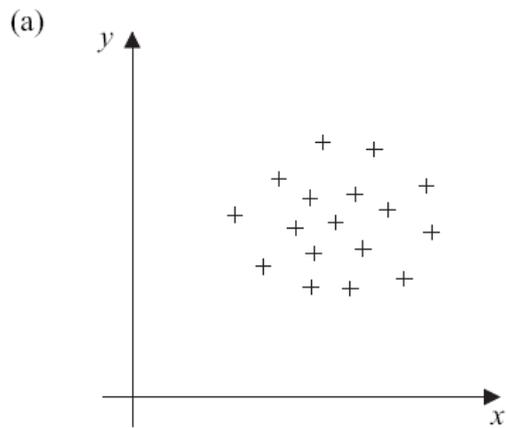
(5 marks)

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

Carina obtains cash from an ATM (cash machine). She suspects that the rate at which she spends cash is affected by the amount of cash she withdrew at her previous visit to an ATM. To investigate this she deliberately varies the amounts she withdraws. She records, for each visit to an ATM, the amount, £x, withdrawn, and the number of hours, y, until her next visit to an ATM.

Withdrawal	1	2	3	4	5	6	7	8	9	10
x	40	10	100	110	120	150	20	90	80	130
y	56	62	195	330	94	270	48	196	214	286

(a) Draw a scatter diagram of the data. *(2 marks)*

(b) Calculate the equation of the regression line of y on x and draw it on your scatter diagram. *(6 marks)*

(c) (i) Carina made one withdrawal immediately before going on a weekend visit to Edinburgh. Identify the most likely withdrawal, giving a reason. *(2 marks)*

(ii) Following another withdrawal, Carina was confined to bed for several days with a heavy cold. Identify the most likely withdrawal, giving a reason. *(2 marks)*

(d) (i) Interpret, in context, the gradient of the regression line.

(ii) Comment on the evidence, if any, that Carina's rate of spending cash is affected by the amount she withdraws. *(3 marks)*

5 (a)		B1 B2	3	Scales Plots $(-1EE)$
(b)	$\bar{x} = 2.75$ $\bar{y} = 6.225$	B1		Both
	$b = \dfrac{8 \times 148.3 - 22 \times 49.8}{8 \times 71 - 22^2}$	M1		
	$= 1.08095...$	A1		
	$a = 6.225 - 1.08... \times 2.75$	M1		
	$= 3.252$	A1		
	$y = 1.08x + 3.25$	B1√	6	
(c)	$y = 1.08 \times 3.2 + 3.25$	M1		
	$= 6.71$	A1	2	Accept 6.7 $\begin{cases} \text{or line put on scatter diagram} & \text{M1} \\ \text{answer visibly read from line} & \text{A1} \end{cases}$
(d)	20 is well beyond the range of the data	E1	1	
	Total		12	

3	(a)	$S_{xx} = 7250 - \dfrac{(200)^2}{6} = 583\dfrac{1}{3}$	B1		Both
		$S_{yy} = 39234 - \dfrac{(436)^2}{6} = 7551\dfrac{1}{3}$			
		$S_{xy} = 12515 - \dfrac{200 \times 436}{6} = -2018\dfrac{1}{3}$	B1		
		$r = \dfrac{-2018\dfrac{1}{3}}{\sqrt{588\dfrac{1}{3} \times 7551\dfrac{1}{3}}} = -0.962$	M1 A1√	4	CAO

1(a)	$S_{xx} = 100372 - \dfrac{896^2}{8} = 20$	B1		or $8 \times 100372 - 896^2 = 160$
	$S_{yy} = 40436 - \dfrac{568^2}{8} = 108$	B1		or $8 \times 40436 - 568^2 = 864$
	$S_{xy} = 63623 - \dfrac{896 \times 568}{8} = 7$	B1		or $8 \times 63623 - 896 \times 568 = 56$
	$r = \dfrac{7}{\sqrt{20 \times 108}} = 0.151$	M1 A1	5	$r = \dfrac{56}{\sqrt{160 \times 864}} = 0.151$
(b)	The PMCC is near zero	E1√		
	Little or no association between VR and English scores	E1√	2	
	Total		**7**	

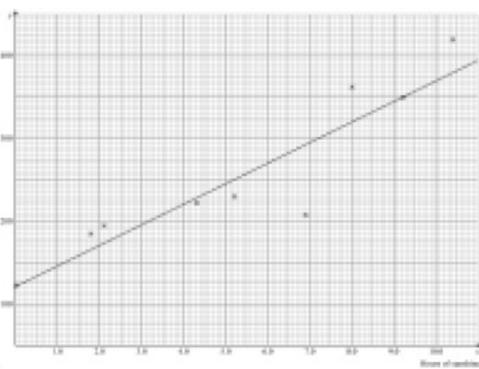|					
6	(a)		B1		Scales and axes
			B2	3	Plots (-1EE)

(b)

$$S_{xy} = 581100 - \frac{2100 \times 1584}{6}$$

$$= 26700$$

$$S_{xx} = 910000 - \frac{2100^2}{6}$$

$$= 175000 \qquad \text{M1}$$

$$b = \frac{26700}{175000} = 0.1526 \qquad \text{A1}$$

$$\bar{x} = 350 \qquad \bar{y} = 264 \qquad \text{B1} \qquad \text{Both}$$

$$a = 264 - 0.1526 \times 350 = 210.6 \qquad \text{M1}$$

Accept more accurate
answers sensibly rounded.

$$l = 211 + 0.153m \qquad \text{A1}$$

Accept also $y = 211 + 0.153x$

Draws line on graph   M1A1√   7

	(c)(i)	The increase in length per gram added	E1		
	(ii)	The length with no mass	E1	2	
	(d)	265 or 266 (mm)	B1√	1	
	(e)	The string will break if too much weight is added	E1	1	
		Total		**14**	

1	$S_{xy} = 6140 - \dfrac{135 \times 301}{6} = -632.5$			
	$S_{xx} = 3475 - \dfrac{135^2}{6} = 437.5$	M1		
	$b = -\dfrac{632.5}{437.5} = -1.446$	A1		
	$\bar{x} = \dfrac{135}{6} = 22.5 \quad \bar{y} = \dfrac{301}{6} = 50.1\dot{6}$	B1		Both
	$a = 50.1\dot{6} - (-1.446) \times 22.5 = 82.70$	M1		
	$y = 82.7 - 1.45x$	A1	5	AWRT
	Total		**5**	

2(a)	$Sxy = 1335 - \dfrac{52 \times 225}{8} = -127.5$	B1		
	$Sxx = 380 - \dfrac{52^2}{8} = 42$	B1		
	$Syy = 7007 - \dfrac{225^2}{8} = 678.875$	B1		
	$r = \dfrac{-127.5}{\sqrt{42 \times 678.875}} = -0.755$	M1 A1	5	

4(a)

4(a)		B3	3	Plots B2 Axes and scale B1
(b)(i)	$S_{xy} = 9020 - \dfrac{210 \times 248}{7} = 1580$			
	$Sxx = 9100 - \dfrac{210^2}{7} = 2800$	M1		
	$b = \dfrac{1580}{2800} = 0.564$	A1		
	$\bar{x} = \dfrac{210}{7} = 30, \ \bar{y} = \dfrac{248}{7}$	B1		Both
	$a = \dfrac{248}{7} - \left(\dfrac{158}{280}\right) \times 30 = 18.5$	M1		
	$y = 18.5 + 0.564x$	A1	5	
(ii)	Draws a line (0, 18.5, 60, 52.4)	B1	1	
(c)(i)	$z =$ Time available – Time taken	M1		
	$= (100 - x) - (a + bx)$	A1		
	$= (100 - a) - (1 + b)x$	A1	3	
(ii)	$z = 81.5 - 1.564x$	B1√		
	For $z = 0$			
	$x = \dfrac{81.5}{1.564} \Rightarrow 52$	M1		
	The latest is 7.52 a.m.	A1	3	CAO
Total			15	

2(a) (i)	0.929	B3	3	(0.928 – 0.93) Allow M2 A1 if method shown
(ii)	Strong association - towns with large number of solicitors have large number of stolen cars.	E1	1	Strong association Large x associated with large y or equivalent Linear Not cause and effect
(b)	Not cause and effect - both depend on size of town	E1 E1	2	Both depend on size of town May be scored in either part (a)(ii) or (b) Combined maximum 3. Up to 2 may be scored in (a)(ii)
	Total		**6**	

8(a)		M1 B1 A1	3	Method Scales and Labels Accurate plot by eye - allow one small slip. Allow x vertical y horizontal
(b)	$y = 121 + 25.3\,x$	B2 B2		121 (120 – 121) 25.3 (25.2 – 25.3) Allow M1 A1 for (a) and (b) if method shown
	$x=0$ $y=121$ $x=10$ $y=373$ (or 374)	M1 A1	6	Method for line Accurate line by eye
(c)(i)	$362 - 120.69 - 25.277{\times}8 = 39.1$	M1 m1		Method their line- ignore sign - allow graphical method Method requires consistent signs
(ii)	$208 - 120.69 - 25.277{\times}6.9 = \text{-}87.1$	A1	3	39.1 (38.7 – 39.5) and -87.1 ($-86.5 - -87.5$)
(d)	Day with 6.9 hour of sunshine	B1		Day with $x = 6.9$
	Sales long way below those predicted by regression line	E1	2	Well below regression line
(e)	20 hours of sunshine not possible in Cornwall.	E1		20 hours of sunshine not possible
	Extrapolation unwise	E1	2	Requires extrapolation
	Total		**16**	

2 (a)	0.209	B3	3	0.209 – 0.21 allow M2A1 if method shown
(b)	Little evidence of linear association between Henri's and Michelle's estimates. Such slight evidence as there is suggests some agreement.	E1✓ E1✓	2	Small/weak/no Some evidence of agreement Allow a mark for appropriate mention of 'linear' (not for 'positive')
	Total		**5**	

8(a)	(See graph on next page)	M1 B1 A1	3	method for scatter diagram Scales and labels Accurate plot (by eye) allow one small slip
(b)	$y = -2.70 + 0.268x$ $x = 20\ \ y = 2.67\ \ \ x = 200\ \ y = 50.99$ $+$ line	B2 B2 M1 A1	6	-2.70 $(-2.69\ ,\ -2.7)$ **sc** B1 2.70 0.268 $(0.268\ ,\ 0.269)$ Allow M1 A1 M1 A1 if method shown method for line Accurate line
(c)(i)	I $\ 17 - (-2.6951) - 0.268437 \times 88$ $= -3.93$ J $\ 47 - (-2.6951) - 0.268437 \times 195$ $= -2.65$	M1 m1 A1	3	method their line – ignore sign method needs all previous M marks- ignore sign -3.93 $(-3.8\ ,\ -4)$ and -2.65 $(-2.5\ ,\ -2.7)$ allow read from graph, allow -3
(ii)	5.13	M1 A1	2	Method 5.13 $(5.1\ ,\ 5.2)$
(d)(i)	26.8	B1	1	26.8 $(26.7\ ,\ 27)$
(ii)	£22 about £5 below amount predicted by regression equation. Similar to mean residual. No reason to say Karen should have been supervised.	E1✓ E1✓	2	Below predicted amount No reason to say she should have been supervised, with references to residuals implied
	Total		**17**	
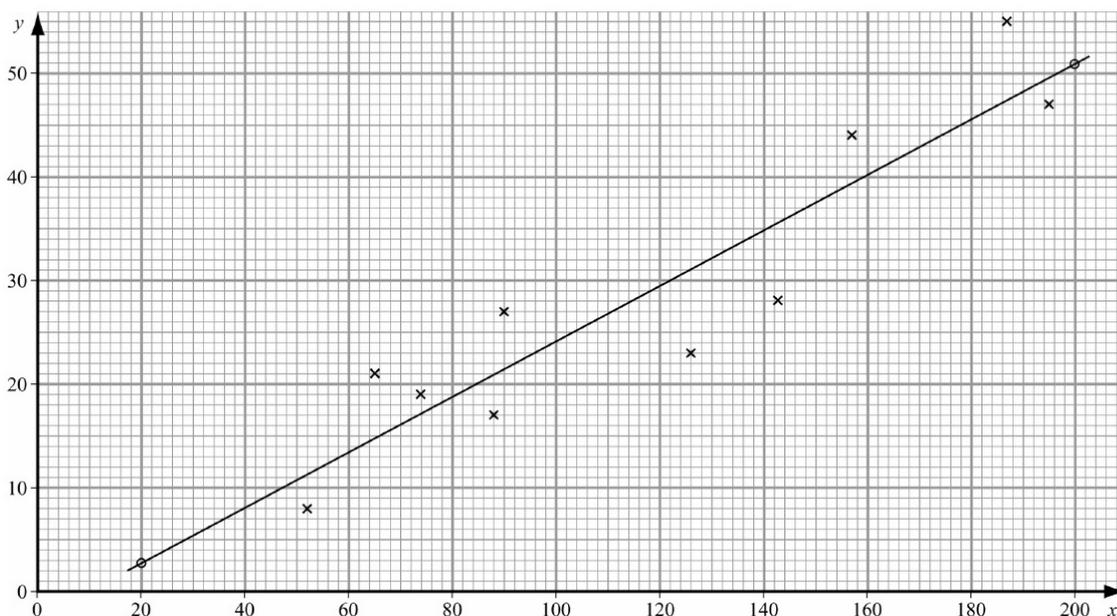
1(a)	0.0		B1		0.0 (−0.2 ~ 0.2)	
(b)	−0.8		B1		negative and > −1	
			B1		magnitude 0.6 ~ 0.98	
(c)	−0.8		M1		negative and > −1	
			A1	5	magnitude 0.4 ~ 0.98	
		Total		**5**		

7(a)	see graph on next page	M1		method for scatter diagram	
		A1	2	reasonably accurate plot, by eye, allow one small slip, disallow for joined up points	
(b)	$y = 30.3 + 1.70x$	B2		30.3 (30.2 ~ 30.3),allow M1A1	
		B2		1.70 (1.70 ~ 1.71),allow M1A1	
	$x = 0 \ y = 30.3 \qquad x = 150 \ y = 285.9$	M1		method for line	
		A1	6	accurate line	
(c)(i)	5 (120,94)	B1		5	
	Spent a lot of cash in a relatively short time	E1	2	reason/point below line	
(ii)	4 (110,330)	B1		4	
	Spent a small amount of cash in a relatively long time	E1	2	reason/point above line	
(d)(i)	Estimate of hours per pound spent	E1			
(ii)	Ignoring exceptional points graph is approximately linear, No substantial evidence of change in rate of spending cash.	E1		graph approximately linear	
		E1	3	no evidence of change in rate - needs attempt at a reason.	
		Total		**15**	