Bivariate Data
Product moment correlation coefficient and Regression

4 A teacher was investigating the relationship between students' performances in mathematics and physics examinations. He took a sample of 8 students who studied both subjects and recorded each mathematics mark, x, and each physics mark, y. The results are given in the table below.

x	25	26	37	37	41	55	58	60
y	22	40	36	38	34	40	52	56

NOV 03

(i) Plot these data on a scatter diagram and describe the nature of the correlation shown by the diagram. [4]

In order to simplify the calculation of the product moment correlation coefficient for these data, the teacher transformed the data using the equations

$$u = x - 25, \qquad v = \tfrac{1}{2}(y - 22).$$

The results are summarised as

$$n = 8,\ \Sigma u = 139,\ \Sigma v = 71,\ \Sigma u^2 = 3759,\ \Sigma v^2 = 825,\ \Sigma uv = 1645.$$

(ii) Calculate the value of the product moment correlation coefficient between u and v, and hence write down the value of the product moment correlation coefficient between x and y. [3]

5 In a science experiment a student dropped a ball on to the ground from a height of x metres and measured the height, y metres, to which it bounced. The experiment was repeated to give 10 pairs of results. The student chose values of x which increased in steps of exactly 0.25 metres. The results of the experiment are given in the table below.

x	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
y	0.11	0.17	0.29	0.38	0.46	0.54	0.63	0.71	0.84	0.90

$[n = 10,\ \Sigma x = 13.75,\ \Sigma y = 5.03,\ \Sigma x^2 = 24.0625,\ \Sigma y^2 = 3.1893,\ \Sigma xy = 8.7575.]$

JUN 01

(i) Calculate the product moment correlation coefficient between x and y. [2]

(ii) By calculating the equation of the appropriate regression line, estimate the height to which the ball will bounce if it is dropped from a height of 2.4 metres. [5]

(iii) Comment on the reliability of

(a) the estimate found in part (ii), [1]

(b) the estimate obtained from the regression line of the height to which the ball will bounce if it is dropped from a height of 10.2 metres. [1]

PMCC + regression (cont 1)

2 A football was rolled in a straight line along the floor of a sports hall. The distance travelled, x metres, was recorded at a time, t seconds, after the ball was released, for values of t from $t = 1.0$ to $t = 4.5$. The results are given in the table below.

t	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
x	8.0	13.2	15.8	18.4	21.0	22.2	24.0	25.1

Jan 03

$[n = 8,\ \Sigma t = 22.0,\ \Sigma x = 147.7,\ \Sigma t^2 = 71.00,\ \Sigma x^2 = 2966.29,\ \Sigma tx = 455.05.]$

(i) On graph paper plot a scatter diagram of the data. [2]

(ii) Calculate the product moment correlation coefficient for the data. [3]

(iii) State the value of Spearman's rank correlation coefficient for the data. [1]

(iv) State with a reason whether a linear model would be appropriate for these data, referring to either your calculations or the scatter diagram. [2]

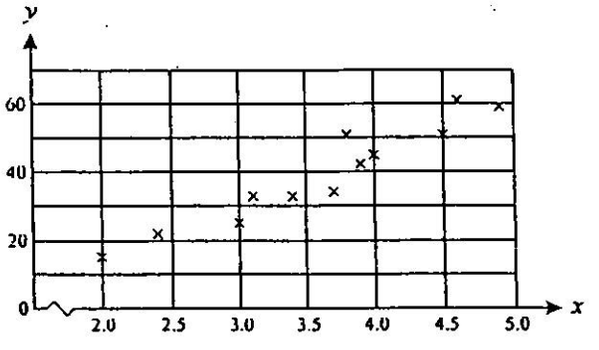
Spec

8 An experiment was conducted to see whether there was any relationship between the maximum tidal current, y cm s^{-1}, and the tidal range, x metres, at a particular marine location. [The *tidal range* is the difference between the height of high tide and the height of low tide.] Readings were taken over a period of 12 days, and the results are shown in the following table.

x	2.0	2.4	3.0	3.1	3.4	3.7	3.8	3.9	4.0	4.5	4.6	4.9
y	15.2	22.0	25.2	33.0	33.1	34.2	51.0	42.3	45.0	50.7	61.0	59.2

$[\Sigma x = 43.3, \Sigma y = 471.9, \Sigma x^2 = 164.69, \Sigma y^2 = 20\,915.75, \Sigma xy = 1837.78.]$

The scatter diagram below illustrates the data.



(i) Calculate the product moment correlation coefficient for the data, and comment briefly on your answer with reference to the appearance of the scatter diagram. [4]

(ii) Calculate the equation of the regression line of maximum tidal current on tidal range. [4]

(iii) Estimate the maximum tidal current on a day when the tidal range is 4.2 m, and comment briefly on how reliable you consider your estimate is likely to be. [3]

(iv) It is suggested that the equation found in part (ii) could be used to predict the maximum tidal current on a day when the tidal range is 15 m. Comment briefly on the validity of this suggestion. [2]

PMCC + regression (cont 2)

4 A student was experimenting with an electrical circuit in which the resistance of one component could be varied. The student increased the resistance in fixed steps from 10 units to 100 units and measured the voltage drop when a fixed current was passed through it. The table below gives the student's results.

Resistance, x units	10	20	30	40	50	60	70	80	90	100
Voltage drop, y units	149	202	253	307	353	407	443	451	550	602

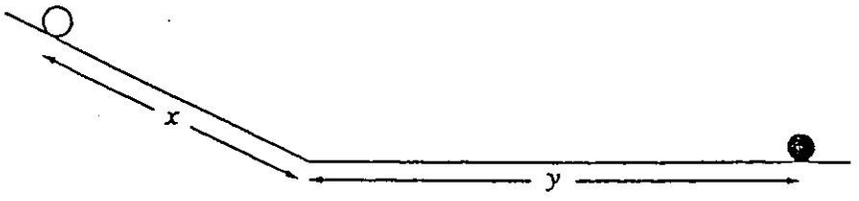
Jun 04

[$n = 10$, $\Sigma x = 550$, $\Sigma y = 3717$, $\Sigma x^2 = 38\,500$, $\Sigma y^2 = 1\,576\,075$, $\Sigma xy = 244\,260$.]

(i) The student wished to use his experimental data to estimate the resistance which would be needed for the voltage drop to be 220 units. Calculate the equation of the appropriate regression line and use it to estimate the resistance, x, which would correspond to $y = 220$. [6]

(ii) Calculate the product moment correlation coefficient for the data and use it to comment on the reliability of the estimate found in part (i). [3]

6



A student conducted an experiment in which she rolled a ball from a point on an inclined plane. The ball rolled down the inclined plane and then continued to roll along a horizontal plane (see diagram). The student measured the distance, x cm, up the inclined plane to the ball's initial position. She also measured the distance, y cm, travelled by the ball along the horizontal plane before it stopped. She repeated the experiment to give 10 pairs of readings altogether. The values of x were increased at regular intervals of 5 cm starting at 40 cm. For each repetition of the experiment she tried to roll the ball down the plane with the same initial speed. The results are given in the table below.

x	40	45	50	55	60	65	70	75	80	85
y	58	69	70	73	88	91	98	108	109	125

Jan 02

[$n = 10$, $\Sigma x = 625$, $\Sigma y = 889$, $\Sigma x^2 = 41\,125$, $\Sigma y^2 = 83\,153$, $\Sigma xy = 58\,440$.]

(i) Calculate the value of the product moment correlation coefficient for the data. [2]

(ii) The student's teacher suggested that, instead of using the original x and y values, she should transform the data using the equations $u = \frac{x - 40}{5}$ and $v = y - 50$, and then calculate the product moment correlation coefficient for the transformed data. Explain what relationship, if any, the product moment correlation coefficient for the transformed data would have with the product moment correlation coefficient calculated in part (i). [1]

The student wishes to estimate the value of x for a ball which rolls a horizontal distance of $y = 100$ cm.

(iii) Calculate the equation of an appropriate regression line and use it to estimate the value of x when $y = 100$. [5]

(iv) Give a reason for your choice of regression line. [1]

PMCC + regression (cont 3)

4 The table below shows the quantity, x units, of electricity produced by nuclear power, and the quantity, y units, of electricity produced by hydro-electric power, for the UK over a period of 12 months.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
x	1.80	1.60	2.00	1.60	1.70	2.00	1.51	1.65	1.97	1.55	1.68	2.11
y	0.05	0.04	0.07	0.05	0.04	0.04	0.02	0.02	0.02	0.02	0.04	0.06

[$n = 12$, $\Sigma x = 21.17$, $\Sigma y = 0.47$, $\Sigma x^2 = 37.8105$, $\Sigma y^2 = 0.0215$, $\Sigma xy = 0.8494$.]

Jan 01

A researcher believes that there may be a link between the variables x and y.

(i) Calculate the product moment correlation coefficient for the data. [2]

(ii) Calculate the equation of the regression line of y on x, giving your answer in the form $y = a + bx$. [3]

(iii) Use your regression line to estimate the amount of electricity produced by hydro-electric power in a month in which the quantity of electricity produced by nuclear power is 2.52. [2]

(iv) Comment on the reliability of the estimate found in part **(iii)**. [2]

7 A researcher took a random sample of 10 GCSE students who had just received their results. She recorded for each student the mean number of hours per week, x, that they spent watching television during their revision period and the number of GCSE 'points', y, that they obtained. The results are given in the table below.

x	15.5	16.9	18.5	19.2	20.3	22.0	25.0	27.5	30.0	30.9
y	49	46	43	40	39	46	25	25	20	18

[$n = 10$, $\Sigma x = 225.8$, $\Sigma y = 351$, $\Sigma x^2 = 5368.90$, $\Sigma y^2 = 13\,577$, $\Sigma xy = 7372.8$.]

(i) On graph paper draw a scatter diagram which illustrates the data. [3]

Jun 03

(ii) Calculate the product moment correlation coefficient for the data. [2]

(iii) Describe the relationship between the mean number of hours of television watched during the students' revision periods and the number of GCSE points that they achieve. State the evidence for your answer. [2]

(iv) Calculate the equation of an appropriate regression line to predict a student's GCSE points score from the mean number of hours of television which the student watched per week during their revision period. [3]

(v) Use the equation of the line found in part **(iv)** to predict the number of GCSE points which a student would achieve if the mean number of hours of television watched per week was 28.1 hours. [1]

(vi) A second researcher also used the set of data in the table but decided to transform the data before doing any calculations. She used $u = \frac{x - 15}{2}$ and $v = y - 50$. State the product moment correlation coefficient for the transformed data used by the second researcher. [1]

PMCC + regression (cont 4)

5 A set of bivariate data (x, y) was collected for an experiment. You are given that

$$n = 10,\ \Sigma x = 78.9,\ \Sigma y = 64.0,\ \Sigma x^2 = 743.45,\ \Sigma y^2 = 476.52,\ \Sigma xy = 575.81.$$

You are also given that the regression line of x on y has equation $x = 1.11 + 1.06y$, where the coefficients are given correct to 3 significant figures.

Jan 03

(i) Calculate the equation of the regression line of y on x, giving your answer in the form $y = a + bx$. [4]

(ii) State the coordinates of the point of intersection of the two lines. (You do not need to solve the simultaneous equations.) [2]

(iii) In the collection of the data neither variable was controlled. Use the appropriate line to estimate the value of x when $y = 3.4$. You may assume that $y = 3.4$ is within the range of the data collected. [2]

8 A student conducted a survey on a group of 5 families in which there was a father and an adult eldest son. The student recorded the height, x cm, of the father and the height, y cm, of the eldest son. The results are given below.

x	164	175	184	188	192
y	171	176	186	188	189

$[n = 5,\ \Sigma x = 903,\ \Sigma y = 910,\ \Sigma x^2 = 163\,585,\ \Sigma y^2 = 165\,878,\ \Sigma xy = 164\,700.]$

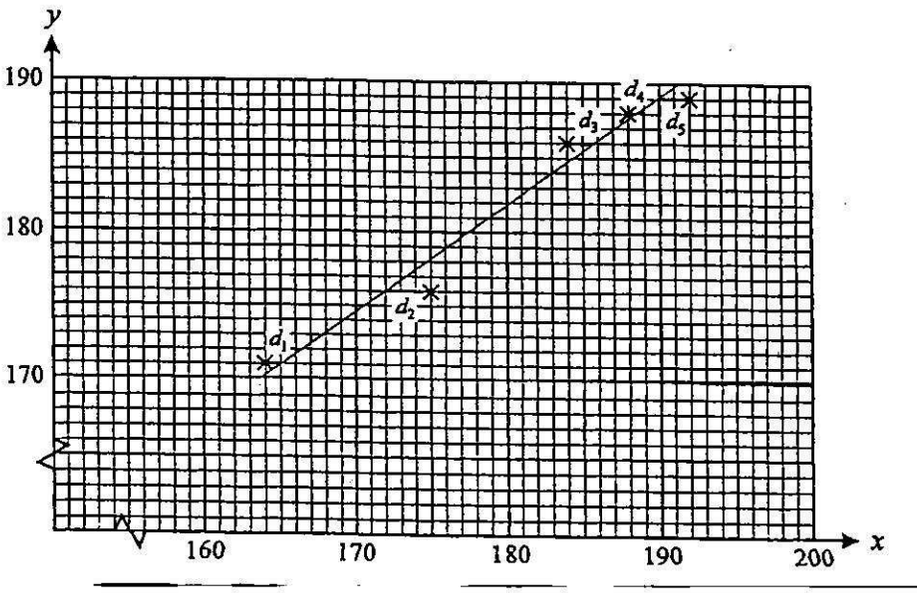
(i) The regression line of y on x and the regression line of x on y intersect at the point P. Find the coordinates of P. [3]

(ii) Calculate the equation of the regression line of y on x, giving your answer in the form $y = a + bx$, where a and b should be given correct to 3 significant figures. [3]

Nov 03

(iii) The equation of the regression line of x on y is $x = -69.1 + 1.37y$, where the coefficients are correct to 3 significant figures. Use the appropriate equation to estimate the height of a father whose son's height is 178 cm. [2]

(iv) The diagram below shows a sketch of a scatter diagram for the above data together with the regression line of x on y. Calculate Σd^2, the sum of the squares of the horizontal deviations of the five points from the regression line of x on y. [4]



PMCC + Regression (cont'd)

8 A teacher decided to investigate the connection between students' performances in History and Geography. She selected 5 students at random and recorded their scores, x and y, in History and Geography tests respectively. She found that, for this set of data, the regression line of y on x had equation $y = 18.5 + 0.1x$ and the regression line of x on y had equation $x = 16.6 + 0.4y$.

Jan 05

(i) Using these equations, calculate the values of the mean score in the History test and the mean score in the Geography test for the 5 students. [3]

(ii) Hence show that $\Sigma x = 125$ and $\Sigma y = 105$. [1]

It is given that $\Sigma x^2 = 3215$, $\Sigma y^2 = 2227.5$ and $\Sigma xy = 2634$.

(iii) Obtain the product moment correlation coefficient for the data. [3]

(iv) The Geography score of a sixth student was mislaid but his History score was known to be 26. Use one of the equations to estimate this student's Geography score, and give a reason for the use of the chosen equation. [3]

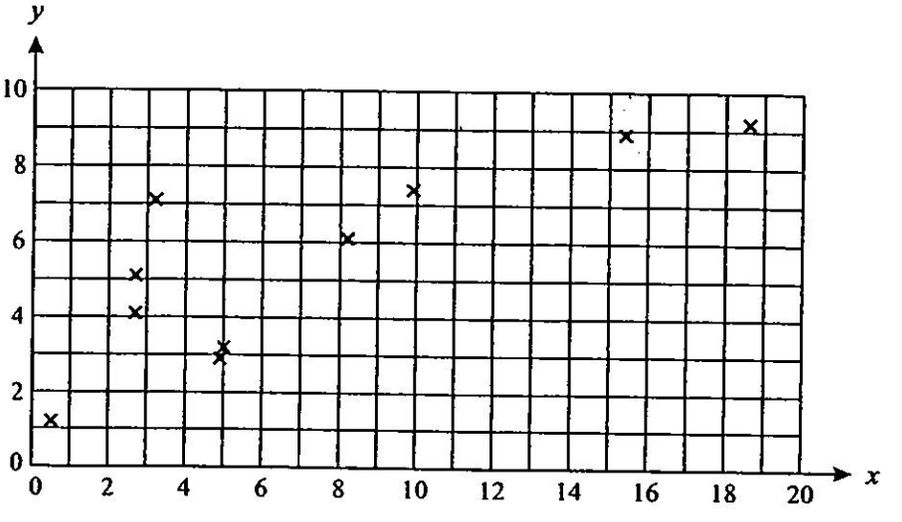
(v) Comment on the reliability of your estimate. [1]

4 An advertising company wished to assess the relationship between the amounts of money spent by 10 clients and the number of people who saw each client's advertisement. For each client the company recorded the amount spent, £x million, and the number of people, y million, who saw the client's advertisement. The results are given in the table and are illustrated by the scatter diagram.

x	5.0	9.9	8.2	18.6	4.9	0.5	15.4	3.2	2.7	2.7
y	3.2	7.4	6.1	9.2	2.9	1.2	8.9	7.1	5.1	4.1

[$n = 10$, $\Sigma x = 71.1$, $\Sigma y = 55.2$, $\Sigma x^2 = 822.45$, $\Sigma y^2 = 369.14$, $\Sigma xy = 509.83$.]

Nov 04



(i) Calculate the product moment correlation coefficient for the data. [2]

The company wishes to estimate the number of people who will see an advertisement for a client who has £6.4 million to spend.

(ii) Calculate the equation of an appropriate regression line. [3]

(iii) Use your answer to part (ii) to estimate the value of y when $x = 6.4$. [1]

(iv) Comment on the reliability of your estimate. [1]

PMCC + regression (cont 6)

2 A chemist recorded the mass, y grams, of a substance which dissolved in $100\,cm^3$ of water at a temperature of $x\,°C$, for 10 different values of x. The results are summarised below.

$$n = 10,\ \Sigma x = 550,\ \Sigma y = 803,\ \Sigma x^2 = 38\,500,\ \Sigma y^2 = 65\,848.9,\ \Sigma xy = 47\,520.$$

Jan 04

(i) Calculate the value of the product moment correlation coefficient. [2]

(ii) Calculate the equation of the regression line of y on x, in the form $y = a + bx$, giving a and b correct to 3 significant figures. [4]

(iii) Assuming that $x = 45$ is a temperature within the range used by the chemist for his experiment, use the equation you found in part (ii) to estimate the mass, in grams, of the substance which will dissolve in $100\,cm^3$ of water at a temperature of 45 °C. [2]

3 A set of bivariate data (x, y) was collected in an experiment in which x is the yield, in tonnes, of a particular crop in a given month and y is the number of centimetres of rainfall in that month. You are given that

$$n = 6,\ \Sigma x = 46.7,\ \Sigma y = 42.5,\ \Sigma x^2 = 400.65,\ \Sigma y^2 = 323.67,\ \Sigma xy = 346.37.$$

Nov 02

(i) Calculate the product moment correlation coefficient for the data. [3]

(ii) Calculate the equation of the regression line of x on y, giving your answer in the form $x = a + by$. [3]

(iii) Use the equation found in part (ii) to estimate the yield of the crop in a month in which the rainfall was 20.4 cm. [1]

(iv) State with a reason whether you think that the estimate found in part (iii) will be reliable. [1]

6 A set of kitchen scales was tested by weighing 8 objects whose exact weights were known. The results of the weighings, correct to the nearest gram, are given in the following table.

Known weight (x grams)	10	20	30	40	50	60	70	80
Measured weight (y grams)	9	19	28	41	52	62	73	83

$[n = 8,\ \Sigma x = 360,\ \Sigma x^2 = 20\,400,\ \Sigma y = 367,\ \Sigma y^2 = 21\,673,\ \Sigma xy = 21\,020.]$

Jun 05

(i) Calculate the product moment correlation coefficient for the data, giving your answer correct to 4 decimal places. [3]

(ii) State, giving a reason, which of x and y is the dependent variable. [1]

(iii) Calculate the equation of a suitable regression line for the data. [4]

(iv) Using the regression equation, estimate the true weights of objects whose measured weights are

(a) 56 grams,

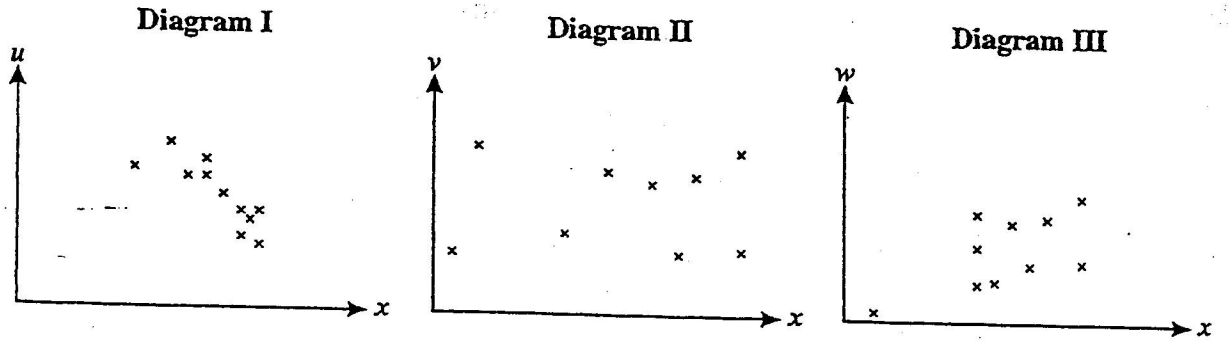
(b) 100 grams.

[2]

(v) Comment on the reliability of each of your two estimates. [2]

PMCC + regression (cont 7)

7 (a) The three scatter diagrams, I, II and III below, represent the way in which three variables, u, v and w, vary with respect to a fourth variable x.



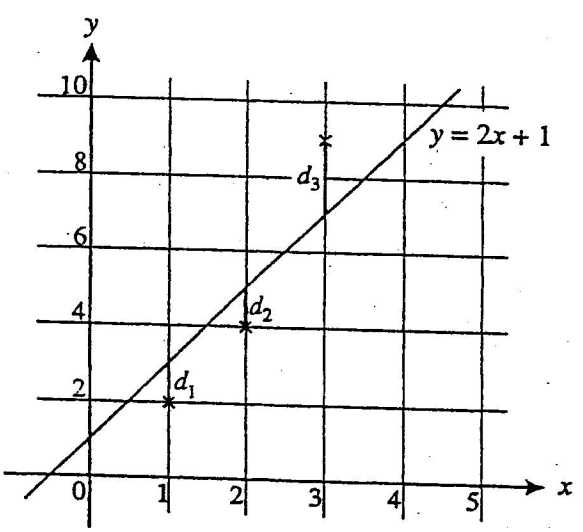
A student calculated the product moment correlation coefficient for each of the three sets of bivariate data and obtained the values

$$0.07, \qquad 0.62, \qquad -0.82.$$

State, with a reason in each case, which correlation coefficient corresponds to each of diagrams I, II and III. [4]

May 02

(b) To investigate the relationship between x and a new variable, y, the student collected three pairs of values of (x, y), namely (1, 2), (2, 4) and (3, 9). The three corresponding data points are shown on the scatter diagram below, together with the student's attempt at a regression line which he fitted by eye. The equation of the student's line is $y = 2x + 1$.



(i) Calculate the sum of squares of the deviations of the three data points from the student's line. That is, calculate $S^2 = d_1^2 + d_2^2 + d_3^2$. [3]

(ii) Calculate the equation of the least squares regression line of y on x for the student's data and calculate the sum of squares of the deviations of the three data points from this line. [5]