Chapter 5: Correlation and Regression

Mr Faruk
Teacher of Mathematics
BSc/MSc/PGCE Mathematics
cieigcsesolutions@gmail.com

1. The percentage oil content, p, and the weight, w milligrams, of each of 10 randomly selected sunflower seeds were recorded. These data are summarised below.

$$\sum w^2 = 41252 \qquad \sum wp = 27557.8 \qquad \sum w = 640 \qquad \sum p = 431 \qquad S_{pp} = 2.72$$

(a) Find the value of S_{ww} and the value of S_{wp}

(3)

(b) Calculate the product moment correlation coefficient between p and w

(2)

(c) Give an interpretation of your product moment correlation coefficient.

(1)

The equation of the regression line of p on w is given in the form $p = a + bw$

(d) Find the equation of the regression line of p on w

(4)

(e) Hence estimate the percentage oil content of a sunflower seed which weighs 60 milligrams.

(2)

6. Following some school examinations, Chetna is studying the results of the 16 students in her class. The mark for paper 1, x, and the mark for paper 2, y, for each student are summarised in the following statistics.

$$\bar{x} = 35.75 \qquad \bar{y} = 25.75 \qquad \sigma_x = 7.79 \qquad \sigma_y = 11.91 \qquad \sum xy = 15\,837$$

(a) Comment on the differences between the marks of the students on paper 1 and paper 2

(2)

Chetna decides to examine these data in more detail and plots the marks for each of the 16 students on the scatter diagram opposite.

(b) (i) Explain why the circled point (38, 0) is possibly an outlier.

(ii) Suggest a possible reason for this result.

(2)

Chetna decides to omit the data point (38, 0) and examine the other 15 students' marks.

(c) Find the value of $\bar{x}$ and the value of $\bar{y}$ for these 15 students.

(3)

For these 15 students

(d) (i) explain why $\sum xy$ is still 15 837

(ii) show that $S_{xy} = 1169.8$

(3)

For these 15 students, Chetna calculates $S_{xx} = 965.6$ and $S_{yy} = 1561.7$ correct to 1 decimal place.

(e) Calculate the product moment correlation coefficient for these 15 students.

(2)

(f) Calculate the equation of the line of regression of y on x for these 15 students, giving your answer in the form $y = a + bx$

(4)

The product moment correlation coefficient between x and y for all 16 students is 0.746

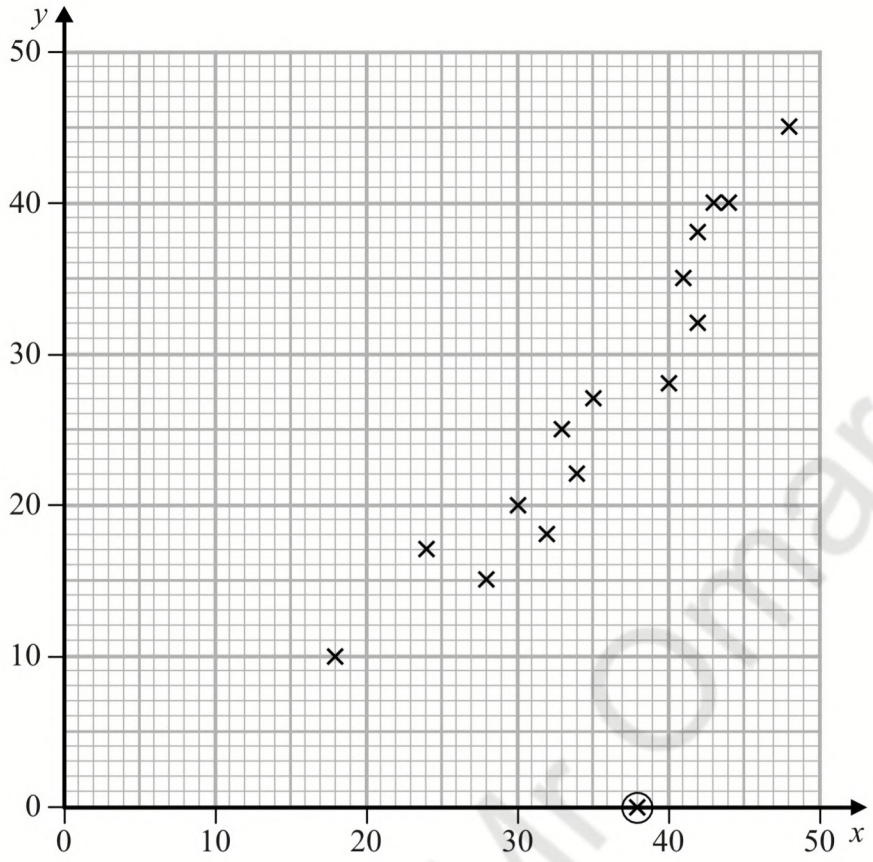
(g) Explain how your calculation in part (e) supports Chetna's decision to omit the point (38, 0) before calculating the equation of the linear regression line.

(1)

(h) Estimate the mark in the second paper for a student who scored 38 marks in the first paper.

(1)

Question 6 continued



6. *Ranpose* hospital offers services to a large number of clinics that refer patients to a range of hospitals.

The manager at *Ranpose* hospital took a random sample of 16 clinics and recorded

- the distance, x km, of the clinic from *Ranpose* hospital
- the percentage, y %, of the referrals from the clinic who attend *Ranpose* hospital.

The data are summarised as

$$\bar{x} = 8.1 \qquad \bar{y} = 20.5 \qquad \sum y^2 = 8266 \qquad S_{xx} = 368.16 \qquad S_{xy} = -630.9$$

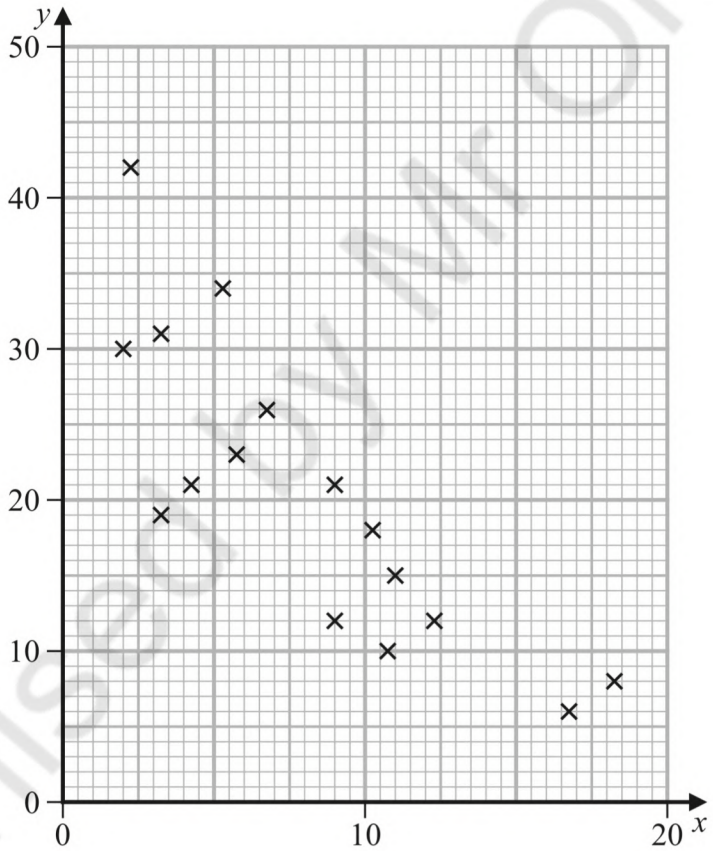
(a) Find the product moment correlation coefficient for these data.

(4)

(b) Give an interpretation of your correlation coefficient.

(1)

The manager at *Ranpose* hospital believes that there may be a linear relationship between the distance of a clinic from the hospital and the percentage of the referrals who attend the hospital. She drew the following scatter diagram for these data.



(c) State, giving a reason, whether or not these data support the manager's belief.

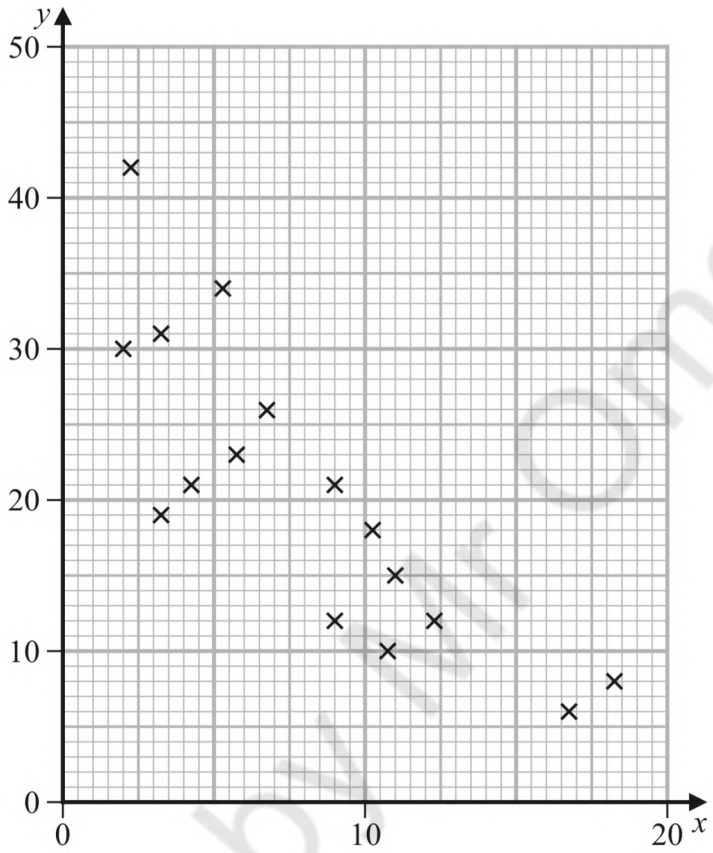
(1)

Question 6 continued

[The summary data and the scatter diagram are repeated below.]

The data are summarised as

$\bar{x} = 8.1$ $\bar{y} = 20.5$ $\sum y^2 = 8266$ $S_{xx} = 368.16$ $S_{xy} = -630.9$



(d) Find the equation of the regression line of y on x, giving your answer in the form $y = a + bx$

(4)

(e) Give an interpretation of the gradient of your regression line.

(1)

(f) Draw your regression line on the scatter diagram.

(1)

The manager believes that *Ranpose* hospital should be attracting an "above average" percentage of referrals from clinics that are less than 5 km from the hospital. She proposes to target one clinic with some extra publicity about the services *Ranpose* offers.

(g) On the scatter diagram circle the point representing the clinic she should target.

(1)

1. Jeremiah is investigating the relationship between the annual heating bill, h dollars, and the total floor area, f square metres, of buildings.

A random sample of 8 buildings is taken and the data for each building are coded using

$$x = \frac{f - 3500}{80} \text{ and } y = \frac{h - 4500}{80}$$

The results for the coded data are summarised below

$$\sum x = 5 \qquad \sum y = 0 \qquad \sum xy = 1818 \qquad S_{xx} = 1754$$

(a) Calculate S_{xy}

(1)

(b) Find the equation of the regression line of h on f in the form $h = a + bf$

(6)

(c) Give an interpretation of (i) the value of b in your regression line,

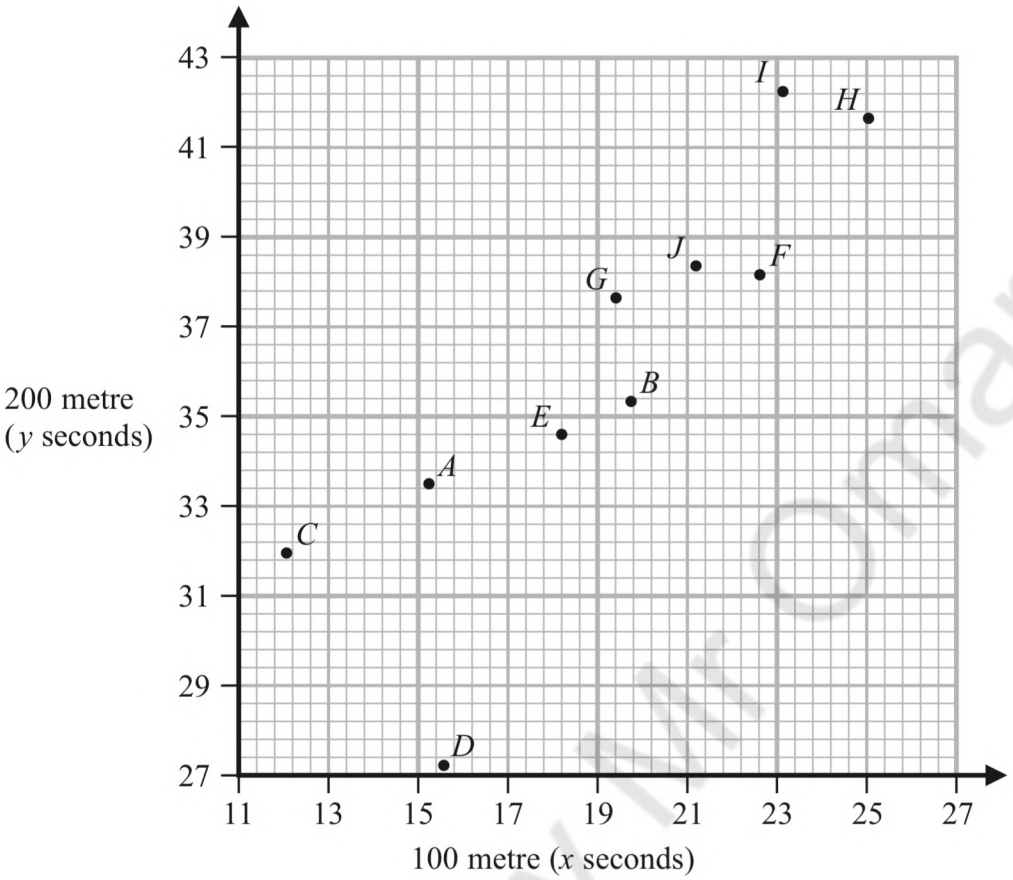
(ii) the value of a in your regression line.

(2)

(d) Estimate the annual heating bill for a building with a total floor area of 4600 square metres.

(2)

4. A random sample of 10 boys A, B, C, D, E, F, G, H, I and J is taken from a junior athletics club. Each boy selected is asked to run a 100 metre race and a 200 metre race. The time taken, x seconds, by each boy to run the 100 metre race is recorded and the time taken, y seconds, by each boy to run the 200 metre race is recorded. The results are plotted on the scatter diagram below.



(a) State, without calculation, which of the 3 values below is most likely to be a value of the product moment correlation coefficient for the data in the scatter diagram.

0.72 0.05 0.95

(1)

In the sample of 10 boys, one is a junior champion 100 metre runner and one is a junior champion 200 metre runner.

(b) Write down the boy who is most likely to be the 100 metre junior champion.

(1)

The data for the two junior champions are removed and the remaining data are summarised below

$\sum x^2 = 3445.26$ $\sum x = 164.4$ $S_{yy} = 67.52$ $S_{xy} = 60.85$

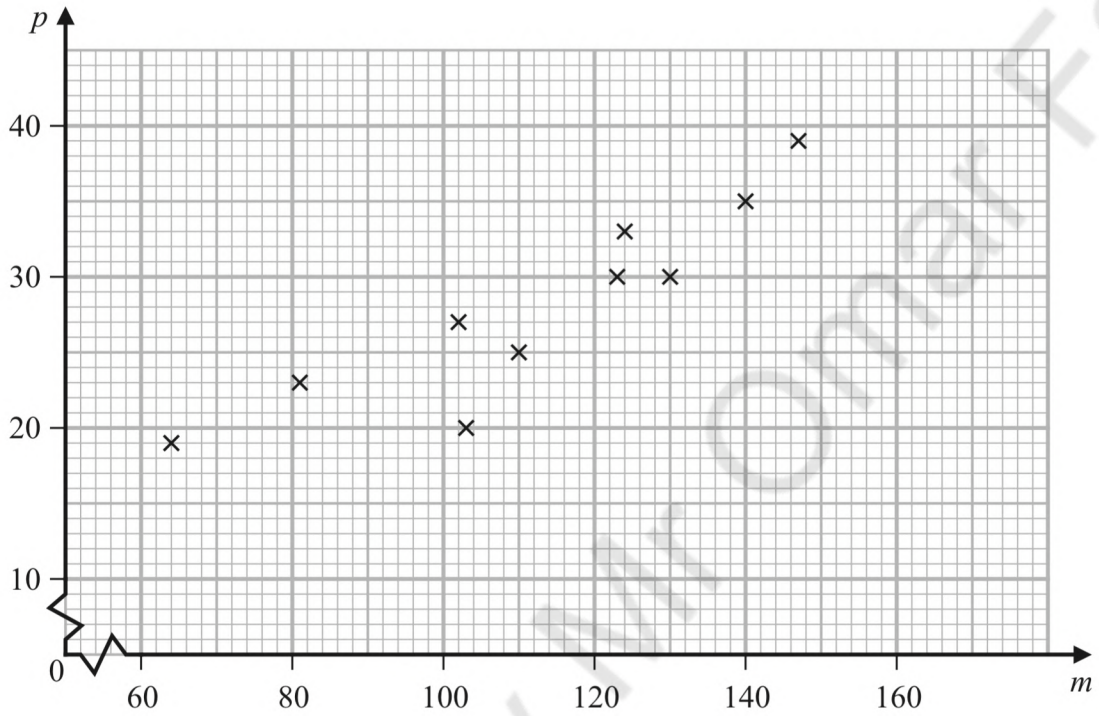
(c) (i) Calculate the value of the product moment correlation coefficient for the remaining data.

(3)

(ii) Comment, in context, on the value of the product moment correlation coefficient that you obtained in part (i).

(1)

3. *Soapern* sells washing machines. When a customer buys a washing machine from *Soapern*, the customer is also invited to buy a guarantee policy to cover breakdowns and repairs for the next three years.
The manager of *Soapern* believes that the relationship between the number of washing machines sold (m) and the number of guarantee policies sold (p) can be modelled by a straight line.
She collected data each month for 10 months. The scatter diagram below illustrates these data.



The data are summarised by the following statistics.

$$\sum m = 1124 \quad \sum p = 281 \quad \sum mp = 32\,958 \quad S_{mm} = 6046.4 \quad S_{pp} = 382.9$$

(a) Show that $S_{mp} = 1373.6$

(1)

(b) Find the value of the product moment correlation coefficient for these data.

(2)

(c) State, giving a reason, whether or not the data are consistent with the manager's belief.

(1)

The manager noticed that the total number of washing machines sold was k times the total number of guarantee policies sold and suggests a model of the form $p = \frac{1}{k}m$, where k is an integer.

(d) Find the value of k.

(2)

Jiang works for *Soapern* and thought that this model oversimplified the situation and suggested that a linear regression of p on m may be more appropriate.

(e) Find the equation of the linear regression of p on m, giving your answer in the form $p = a + bm$, where a and b should be given to 3 significant figures.

(4)

(f) Use Jiang's model to estimate the number of guarantee policies sold when 70 washing machines are sold in a month.

(1)

Usually about 70 washing machines are sold in January. *Soapern* decides to offer a bonus to staff during January based on the number of guarantee policies sold. If the number of guarantee policies sold is greater than the number estimated by the model, the bonus will be paid.

(g) State, giving your reasons, whether you would recommend that the staff use the manager's model or Jiang's model.

(2)

5. A large company rents shops in different parts of the country. A random sample of 10 shops was taken and the floor area, x in $10\,m^2$, and the annual rent, y in thousands of dollars, were recorded.
The data are summarised by the following statistics

$$\sum x = 900 \quad \sum x^2 = 84\,818 \quad \sum y = 183 \quad \sum y^2 = 3434$$

and the regression line of y on x has equation $y = 6.066 + 0.136x$

(a) Use the regression line to estimate the annual rent in dollars for a shop with a floor area of $800\,m^2$

(2)

(b) Find S_{yy} and S_{xx}

(3)

(c) Find the product moment correlation coefficient between y and x.

(4)

An 11th shop is added to the sample. The floor area is $900\,m^2$ and the annual rent is 15 000 dollars.

(d) Use the formula $S_{xy} = \sum(x - \bar{x})(y - \bar{y})$ to show that the value of S_{xy} for the 11 shops will be the same as it was for the original 10 shops.

(2)

(e) Find the new equation of the regression line of y on x for the 11 shops.

(3)

The company is considering renting a larger shop with area of $3000\,m^2$

(f) Comment on the suitability of using the new regression line to estimate the annual rent. Give a reason for your answer.

(1)

5. A company director wants to introduce a performance-related pay structure for her managers. A random sample of 15 managers is taken and the annual salary, y in £1000, was recorded for each manager. The director then calculated a performance score, x, for each of these managers.
The results are shown on the scatter diagram in Figure 1 on the next page.

(a) Describe the correlation between performance score and annual salary.

(1)

The results are also summarised in the following statistics.

$$\sum x = 465 \quad \sum y = 562 \quad S_{xx} = 2492 \quad \sum y^2 = 23140 \quad \sum xy = 19428$$

(b) (i) Show that $S_{xy} = 2006$

(1)

(ii) Find S_{yy}

(2)

(c) Find the product moment correlation coefficient between performance score and annual salary.

(2)

The director believes that there is a linear relationship between performance score and annual salary.

(d) State, giving a reason, whether or not these data are consistent with the director's belief.

(1)

(e) Calculate the equation of the regression line of y on x, in the form $y = a + bx$
Give the value of a and the value of b to 3 significant figures.

(4)

(f) Give an interpretation of the value of b.

(1)

(g) Plot your regression line on the scatter diagram in Figure 1

(2)

The director hears that one of the managers in the sample seems to be underperforming.

(h) On the scatter diagram, circle the point that best identifies this manager.

(1)

The director decides to use this regression line for the new performance related pay structure.

(i) Estimate, to 3 significant figures, the new salary of a manager with a performance score of 30

(2)

Question 5 continued

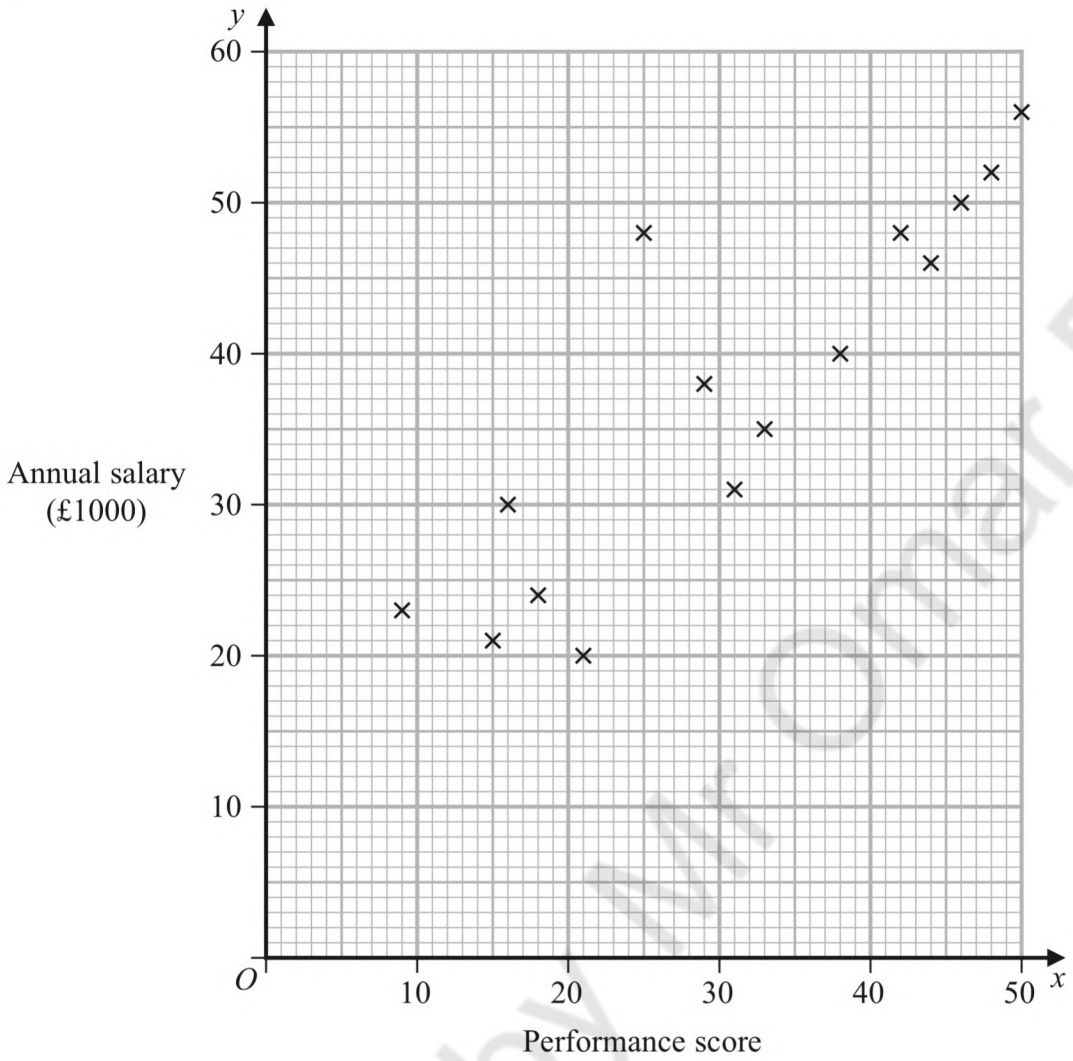


Figure 1

6. Two economics students, Andi and Behrouz, are studying some data relating to unemployment, x%, and increase in wages, y%, for a European country. The least squares regression line of y on x has equation

$$y = 3.684 - 0.3242x$$

and $\sum y = 23.7$ $\sum y^2 = 42.63$ $\sum x^2 = 756.81$ $n = 16$

(a) Show that $S_{yy} = 7.524375$

(1)

(b) Find S_{xx}

(4)

(c) Find the product moment correlation coefficient between x and y.

(3)

Behrouz claims that, assuming the model is valid, the data show that when unemployment is 2% wages increase at over 3%

(d) Explain how Behrouz could have come to this conclusion.

(1)

Andi uses the formula

$$\text{range} = \text{mean} \pm 3 \times \text{standard deviation}$$

to estimate the range of values for x.

(e) Find estimates of the minimum value and the maximum value of x in these data using Andi's formula.

(3)

(f) Comment, giving a reason, on the reliability of Behrouz's claim.

(2)

Andi suggests using the regression line with equation $y = 3.684 - 0.3242x$ to estimate unemployment when wages are increasing at 2%

(g) Comment, giving a reason, on Andi's suggestion.

(2)

2. A large company is analysing how much money it spends on paper in its offices each year. The number of employees in the office, x, and the amount spent on paper in a year, p (\$ hundreds), in each of 12 randomly selected offices were recorded.

The results are summarised in the following statistics.

$$\sum x = 93 \qquad S_{xx} = 148.25 \qquad \sum p = 273 \qquad \sum p^2 = 6602.72 \qquad \sum xp = 2347$$

(a) Show that $S_{xp} = 231.25$

(1)

(b) Find the product moment correlation coefficient for these data.

(3)

(c) Find the equation of the regression line of p on x in the form $p = a + bx$

(4)

(d) Give an interpretation of the gradient of your regression line.

(1)

The director of the company wants to reduce the amount spent on paper each year.

He wants each office to aim for a model of the form $p = \frac{4}{5}a + \frac{1}{2}bx$, where a and b are the values found in part (c).

Using the data for the 93 employees from the 12 offices,

(e) estimate the percentage saving in the amount spent on paper each year by the company using the director's model.

2. Tom's car holds 50 litres of petrol when the fuel tank is full.

For each of 10 journeys, each starting with 50 litres of petrol in the fuel tank, Tom records the distance travelled, d kilometres, and the amount of petrol used, p litres.

The summary statistics for the 10 journeys are given below.

$$\sum d = 1029 \quad \sum p = 50.8 \quad \sum dp = 5240.8 \quad S_{dd} = 344.9 \quad S_{pp} = 0.576$$

(a) Calculate the product moment correlation coefficient between d and p

(3)

The amount of petrol remaining in the fuel tank for each journey, w litres, is recorded.

(b) (i) Write down an equation for w in terms of p

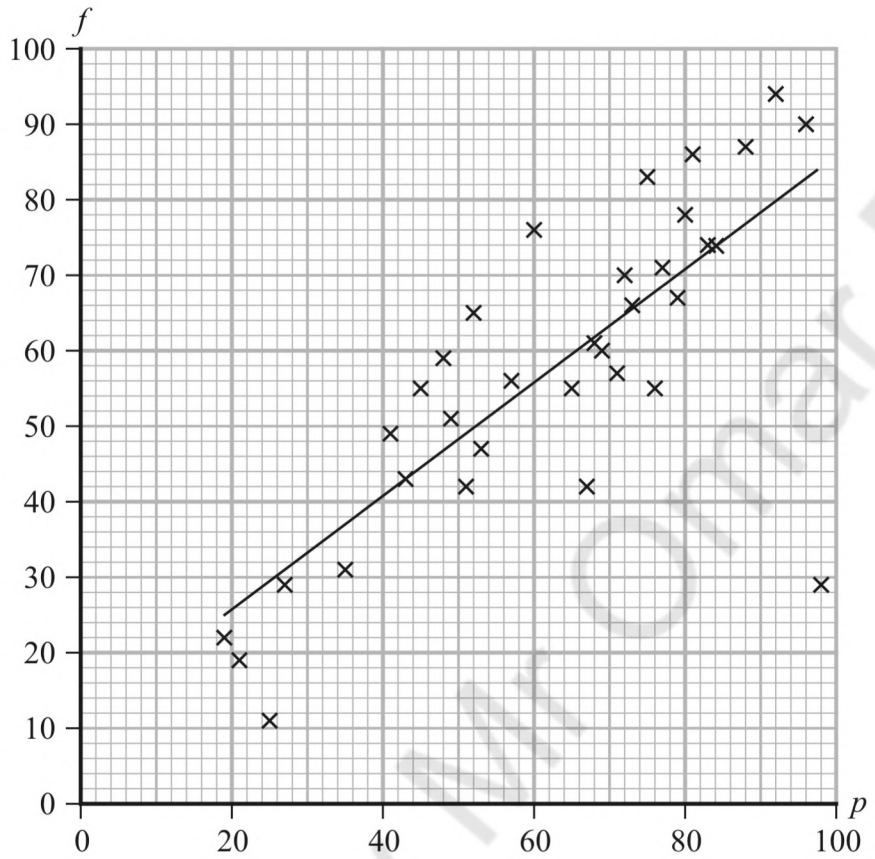
(ii) Hence, write down the value of the product moment correlation coefficient between w and p

(2)

(c) Write down the value of the product moment correlation coefficient between d and w

(1)

6. Students on a psychology course were given a pre-test at the start of the course and a final exam at the end of the course. The teacher recorded the number of marks achieved on the pre-test, p, and the number of marks achieved on the final exam, f, for 34 students and displayed them on the scatter diagram.



The equation of the least squares regression line for these data is found to be

$$f = 10.8 + 0.748p$$

For these students, the mean number of marks on the pre-test is 62.4

(a) Use the regression model to find the mean number of marks on the final exam.

(2)

(b) Give an interpretation of the gradient of the regression line.

(1)

Considering the equation of the regression line, Priya says that she would expect someone who scored 0 marks on the pre-test to score 10.8 marks on the final exam.

(c) Comment on the reliability of Priya's statement.

(1)

(d) Write down the number of marks achieved on the final exam for the student who exceeded the expectation of the regression model by the largest number of marks.

(1)

Question 6 continued

(e) Find the range of values of p for which this regression model, $f = 10.8 + 0.748p$, predicts a greater number of marks on the final exam than on the pre-test.

(3)

Later the teacher discovers an error in the recorded data. The student who achieved a score of 98 on the pre-test, scored 92 not 29 on the final exam.

The summary statistics used for the model $f = 10.8 + 0.748p$ are corrected to include this information and a new least squares regression line is found.

Given the **original** summary statistics were,

$$n = 34 \quad \sum p = 2120 \quad \sum pf = 133\,486 \quad S_{pp} = 15\,573.76 \quad S_{pf} = 11\,648.35$$

(f) calculate the gradient of the new regression line. Show your working clearly.

(5)

2. Stuart is investigating the relationship between Gross Domestic Product (GDP) and the size of the population for a particular country.
He takes a random sample of 9 years and records the size of the population, t millions, and the GDP, g billion dollars for each of these years.

The data are summarised as

$$n = 9 \quad \sum t = 7.87 \quad \sum g = 144.84 \quad \sum g^2 = 3624.41 \quad S_{tt} = 1.29 \quad S_{tg} = 40.25$$

(a) Calculate the product moment correlation coefficient between t and g

(3)

(b) Give an interpretation of your product moment correlation coefficient.

(1)

(c) Find the equation of the least squares regression line of g on t in the form $g = a + bt$

(4)

(d) Give an interpretation of the value of b in your regression line.

(1)

(e) (i) Use the regression line from part (c) to estimate the GDP, in billions of dollars, for a population of 7 000 000

(2)

(ii) Comment on the reliability of your answer in part (i). Give a reason, in context, for your answer.

(1)

Using the regression line from part (c), Stuart estimates that for a population increase of x million there will be an increase of 0.1 billion dollars in GDP.

(f) Find the value of x

(2)

2. The production cost, £c million, of a film and the total ticket sales, £t million, earned by the film are recorded for a sample of 40 films.

Some summary statistics are given below.

$$\sum c = 1634 \qquad \sum t = 1361 \qquad \sum t^2 = 82\,873 \qquad \sum ct = 83\,634 \qquad S_{cc} = 28\,732.1$$

(a) Find the exact value of S_{tt} and the exact value of S_{ct}

(3)

(b) Calculate the value of the product moment correlation coefficient for these data.

(2)

(c) Give an interpretation of your answer to part (b)

(1)

(d) Show that the equation of the linear regression line of t on c can be written as

$$t = -5.84 + 0.976c$$

where the values of the intercept and gradient are given to 3 significant figures.

(3)

(e) Find the expected total ticket sales for a film with a production cost of £90 million.

(2)

Using the regression line in part (d)

(f) find the range of values of the production cost of a film for which the total ticket sales are less than 80% of its production cost.

(2)

6. A research student is investigating the maximum weight, y grams, of sugar that will dissolve in 100 grams of water at various temperatures, x °C, where $10 \leqslant x \leqslant 80$

The research student calculated the regression line of y on x and found it to be

$$y = 151.2 + 2.72x$$

(a) Give an interpretation of the gradient of the regression line.

(1)

(b) Use the regression line to estimate the maximum weight of sugar that will dissolve in 100 grams of water when the temperature is 90 °C.

(2)

(c) Comment on the reliability of your estimate, giving a reason for your answer.

(2)

Using the regression line of y on x and the following summary statistics

$$\sum y = 3119 \qquad \sum y^2 = 851\,093 \qquad \sum x^2 = 24\,500 \qquad n = 12$$

(d) show that the product moment correlation coefficient for these data is 0.988 to 3 decimal places.

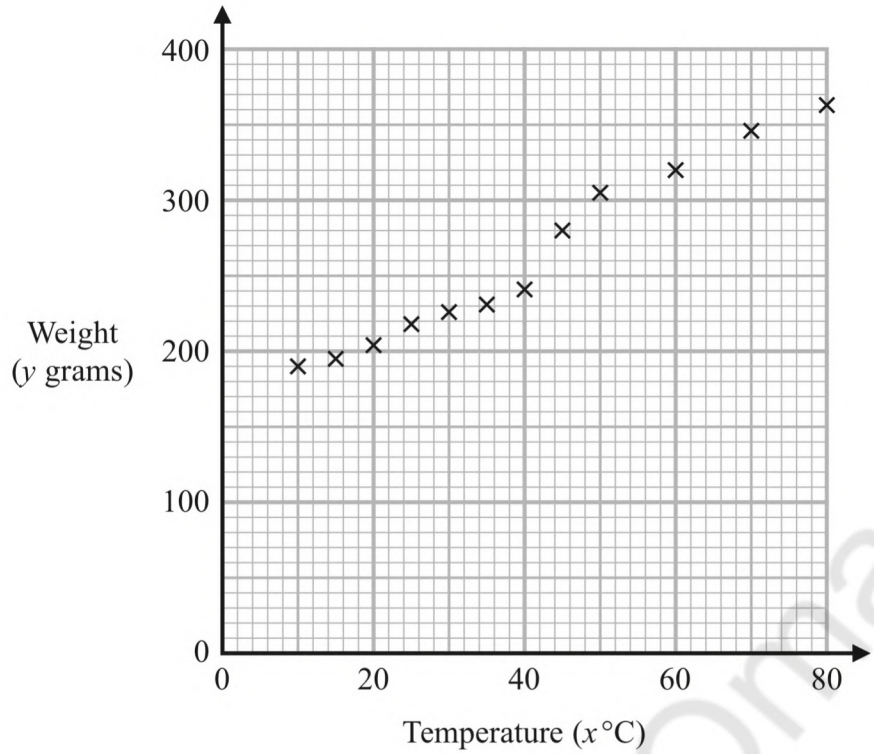
(7)

The research student's supervisor plotted the original data on a scatter diagram, shown on page 23

With reference to both the scatter diagram and the correlation coefficient,

(e) discuss the suitability of a linear regression model to describe the relationship between x and y.

(2)



2. Two students, Olive and Shan, collect data on the weight, w grams, and the tail length, t cm, of 15 mice.

Olive summarised the data as follows

$S_{tt} = 5.3173$ $\quad \sum w^2 = 6089.12$ $\quad \sum tw = 2304.53$ $\quad \sum w = 297.8$ $\quad \sum t = 114.8$

(a) Calculate the value of S_{tw} and the value of S_{ww}

(3)

(b) Calculate the value of the product moment correlation coefficient between w and t

(2)

(c) Show that the equation of the regression line of w on t can be written as

$$w = -16.7 + 4.77t$$

(3)

(d) Give an interpretation of the gradient of the regression line.

(1)

(e) Explain why it would not be appropriate to use the regression line in part (c) to estimate the weight of a mouse with a tail length of 2 cm.

(2)

Shan decided to code the data using $x = t - 6$ and $y = \frac{w}{2} - 5$

(f) Write down the value of the product moment correlation coefficient between x and y

(1)

(g) Write down an equation of the regression line of y on x
You do not need to simplify your equation.

(1)

6. The variables x and y have the following regression equations based on the same 12 observations.

	Regression equation
y on x	$y = 1.4x + 1.5$
x on y	$x = 1.2 + 0.2y$

(a) (i) Find the point of intersection of these lines.

(ii) Hence show that $\sum x = 25$

(4)

Given that

$$\sum xy = \frac{6961}{60}$$

(b) Find S_{xy}

(4)

(c) Find the product moment correlation coefficient between x and y

(4)

2. The average minimum monthly temperature, x degrees Fahrenheit (°F), and the average maximum monthly temperature, y degrees Fahrenheit (°F), in Kolkata were recorded for 12 months.

Some of the summary statistics are given below.

$$\sum x = 862 \quad \sum x^2 = 62\,802 \quad S_{yy} = 413.67 \quad S_{xy} = 512.67 \quad n = 12$$

(a) (i) Calculate the mean of the 12 values of the average **minimum** monthly temperature.

(ii) Show that the standard deviation of the 12 values of the average **minimum** monthly temperature is 8.57°F to 3 significant figures.

(3)

(b) Calculate the product moment correlation coefficient between x and y

(3)

For comparative purposes with a UK city, it was necessary to convert the temperatures from degrees Fahrenheit (°F) to degrees Celsius (°C).

The formula used was

$$c = \frac{5}{9}(f - 32)$$

where f is the temperature in °F and c is the temperature in °C

(c) Use this formula and the values from part (a) to calculate, in °C, the mean and the standard deviation of the 12 values of the average **minimum** monthly temperature in Kolkata.
Give your answers to 3 significant figures.

(4)

Given that

- u is the equivalent temperature in °C of x
- v is the equivalent temperature in °C of y

(d) state, giving a reason, the product moment correlation coefficient between u and v

(2)

4. A French test and a Spanish test were sat by 11 students.

The table below shows their marks.

Student	A	B	C	D	E	F	G	H	I	J	K
French mark (f)	24	30	32	32	36	36	40	44	50	60	68
Spanish mark (s)	16	90	24	28	32	36	38	44	48	48	68

Ignoring the point (30, 90), Greg calculated the following summary statistics.

$$\sum f = 422 \quad \sum s = 382 \quad S_{ff} = 1667.6 \quad S_{fs} = 1735.6$$

(b) Use these summary statistics to show that the equation of the least squares regression line of s on f for the remaining 10 students is

$$s = -5.72 + 1.04f$$

where the values of the intercept and gradient are given to 3 significant figures. You must show your working.

(3)

(c) Give an interpretation of the gradient of the regression line.

(1)

Two further students sat the French test but missed the Spanish test.

(d) Using the equation given in part (b), estimate

(i) a Spanish mark for the student who scored 55 marks in their French test,

(ii) a Spanish mark for the student who scored 18 marks in their French test.

(3)

(e) State, giving a reason, which of the two estimates found in part (d) would be the more reliable estimate.

(2)

4. A biologist is studying bears. The biologist records the length, d cm, and the girth, g cm, of 8 bears. The biologist summarises the data as follows

$$\sum d = 1456.8 \qquad \sum g = 713.2 \qquad \sum dg = 141\,978.84 \qquad \sum g^2 = 72\,675.98$$

$$S_{dd} = 16\,769.78$$

(a) Calculate the exact value of S_{dg} and the exact value of S_{gg}

(3)

(b) Calculate the value of the product moment correlation coefficient between d and g

(2)

(c) Show that the equation of the regression line of g on d can be written as

$$g = -42.3 + 0.722d$$

where the values of the intercept and gradient are given to 3 significant figures.

(3)

(d) Give an interpretation, in context, of the gradient of the regression line.

(1)

Using the equation of the regression line given in part (c)

(e) (i) estimate the girth of a bear with a length of 2.5 metres,

(ii) explain why an estimate for the girth of a bear with a length of 0.5 metres is not reliable.

(2)

Using the regression line from part (c), the biologist estimates that for each x cm increase in the length of a bear there will be a 17.3 cm increase in the girth.

(f) Find the value of x

(2)

2. A biologist records the length, y cm, and the weight, w kg, of 50 rabbits.
The following summary statistics are calculated from these data.

$$\sum y = 2015 \quad \sum y^2 = 81938.5 \quad \sum w = 125 \quad S_{ww} = 72.25 \quad S_{yw} = 219.55$$

(a) (i) Show that $S_{yy} = 734$

(ii) Calculate the product moment correlation coefficient for these data.
Give your answer to 3 decimal places.

(3)

(b) Interpret your value of the product moment correlation coefficient.

(1)

The biologist believes that a linear regression model may be appropriate to describe these data.

(c) State, with a reason, whether or not your value of the product moment correlation coefficient is consistent with the biologist's belief.

(1)

(d) Find the equation of the regression line of w on y, giving your answer in the form $w = a + by$

(4)

Jeff has a pet rabbit of length 45 cm.

(e) Use your regression equation to estimate the weight of Jeff's rabbit.

(2)

6. As a part of a selection process, applicants for a television game show have to take a test, and then complete a task as quickly as possible. The test score, w, and the time taken to complete the task, t minutes, are recorded for each applicant.

The summary statistics below represent the data for a random sample of 30 applicants.

$$S_{wt} = -1648.83 \qquad S_{ww} = 2396.97 \qquad \sum w = 839 \qquad \sum t = 635 \qquad \sum t^2 = 14\,837$$

(a) Show that the product moment correlation coefficient for these data is −0.901 to 3 significant figures.

(2)

A scatter diagram of t against w is plotted for these data.

(b) State **two** features of the graph that you would expect to see, given the correlation coefficient in part (a)

(2)

(c) Calculate the equation of the regression line of t on w in the form

$$t = a + bw$$

Give the values of the constants a and b to 3 significant figures.

(4)

(d) Give an interpretation of the gradient of this regression line.

(1)

The test score, w, was a score out of 50
The manager of the selection process now decides to double all the values of w to make them into percentages.
The manager then recalculates the product moment correlation coefficient and the equation of the regression line.

(e) State, for **each** of the following, whether the value would increase **or** decrease **or** stay the same as a result of applying this change.

(i) The product moment correlation coefficient

(ii) The magnitude of the gradient of the regression line

(iii) The t intercept of the regression line

(3)