

1. Sammy is studying the number of units of gas, g , and the number of units of electricity, e , used in her house each week. A random sample of 10 weeks use was recorded and the data for each week were coded so that $x = \frac{g-60}{4}$ and $y = \frac{e}{10}$. The results for the coded data are summarised below

$$\sum x = 48.0, \quad \sum y = 58.0, \quad S_{xx} = 312.1, \quad S_{yy} = 2.10, \quad S_{xy} = 18.35$$

- (a) Find the equation of the regression line of y on x in the form $y = a + bx$.

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

(4)

- (b) Hence find the equation of the regression line of e on g in the form $e = c + dg$.

Give the values of c and d correct to 2 significant figures.

(4)

- (c) Use your regression equation to estimate the number of units of electricity used in a week when 100 units of gas were used.

(2)

(Total 10 marks)

2. A biologist is comparing the intervals (m seconds) between the mating calls of a certain species of tree frog and the surrounding temperature (t °C). The following results were obtained.

t °C	8	13	14	15	15	20	25	30
m secs	6.5	4.5	6	5	4	3	2	1

(You may use $\sum tm = 469.5$, $S_{tt} = 354$, $S_{mm} = 25.5$)

- (a) Show that $S_{tm} = -90.5$. (4)
- (b) Find the equation of the regression line of m on t giving your answer in the form $m = a + bt$. (4)
- (c) Use your regression line to estimate the time interval between mating calls when the surrounding temperature is 10 °C. (1)
- (d) Comment on the reliability of this estimate, giving a reason for your answer. (1)

(Total 10 marks)

3. A scientist is researching whether or not birds of prey exposed to pollutants lay eggs with thinner shells. He collects a random sample of egg shells from each of 6 different nests and tests for pollutant level, p , and measures the thinning of the shell, t . The results are shown in the table below.

p	3	8	30	25	15	12
t	1	3	9	10	5	6

[You may use $\sum p^2 = 1967$ and $\sum pt = 694$]

- (a) On graph paper, draw a scatter diagram to represent these data. (2)
- (b) Explain why a linear regression model may be appropriate to describe the relationship between p and t . (1)
- (c) Calculate the value of S_{pt} and the value of S_{pp} . (4)
- (d) Find the equation of the regression line of t on p , giving your answer in the form $t = a + bp$. (4)
- (e) Plot the point (\bar{p}, \bar{t}) and draw the regression line on your scatter diagram. (2)
- The scientist reviews similar studies and finds that pollutant levels above 16 are likely to result in the death of a chick soon after hatching.
- (f) Estimate the minimum thinning of the shell that is likely to result in the death of a chick. (2)

(Total 15 marks)

4. Statistical models can provide a cheap and quick way to describe a real world situation.

(a) Give two other reasons why statistical models are used.

(2)

A scientist wants to develop a model to describe the relationship between the average daily temperature, x °C, and her household's daily energy consumption, y kWh, in winter.

A random sample of the average daily temperature and her household's daily energy consumption are taken from 10 winter days and shown in the table.

x	-0.4	-0.2	0.3	0.8	1.1	1.4	1.8	2.1	2.5	2.6
y	28	30	26	25	26	27	26	24	22	21

[You may use $\sum x^2 = 24.76$ $\sum y = 255$ $\sum \sum xy = 283.8$ $S_{xx} = 10.36$]

(b) Find S_{xy} for these data.

(3)

(c) Find 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)

(d) Give an interpretation of the value of a .

(1)

(e) Estimate her household's daily energy consumption when the average daily temperature is 2°C.

(2)

The scientist wants to use the linear regression model to predict her household's energy consumption in the summer.

(f) Discuss the reliability of using this model to predict her household's energy consumption in the summer.

(2)

(Total 14 marks)

5. A farmer collected data on the annual rainfall, x cm, and the annual yield of peas, p tonnes per acre.

The data for annual rainfall was coded using $v = \frac{x-5}{10}$ and the following statistics were found.

$$S_{vv} = 5.753 \quad S_{pv} = 1.688 \quad S_{pp} = 1.168 \quad \bar{p} = 3.22 \quad \bar{v} = 4.42$$

- (a) Find the equation of the regression line of p on v in the form $p = a + bv$.

(4)

- (b) Using your regression line estimate the annual yield of peas per acre when the annual rainfall is 85 cm.

(2)

(Total 6 marks)

6. A teacher took a random sample of 8 children from a class. For each child the teacher recorded the length of their left foot, f cm, and their height, h cm. The results are given in the table below.

f	23	26	23	22	27	24	20	21
h	135	144	134	136	140	134	130	132

(You may use $\sum f = 186$ $\sum h = 1085$ $S_{ff} = 39.5$ $S_{hh} = 139.875$ $\sum fh = 25\,291$)

- (a) Calculate S_{fh} . (2)
- (b) Find the equation of the regression line of h on f in the form $h = a + bf$.
Give the value of a and the value of b correct to 3 significant figures. (5)
- (c) Use your equation to estimate the height of a child with a left foot length of 25 cm. (2)
- (d) Comment on the reliability of your estimate in part (c), giving a reason for your answer. (2)

The left foot length of the teacher is 25 cm.

- (e) Give a reason why the equation in part (b) should not be used to estimate the teacher's height. (1)

(Total 12 marks)

TOTAL FOR PAPER: 67 MARKS