1. A medical researcher is studying the relationship between age (x years) and volume of blood (y ml) pumped by each contraction of the heart. The researcher obtained the following data from a random sample of 8 patients.

Age (x)	20	25	30	45	55	60	65	70
Volume (y)	74	76	77	72	68	67	64	62

[You may use $\sum x = 370$, $S_{xx} = 2587.5$, $\sum y = 560$, $\sum y^2 = 39\,418$, $S_{xy} = -710$]

(a) Calculate S_{yy}

(2)

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

(2)

(c) Interpret your value of the correlation coefficient.

(1)

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

(d) State, giving a reason, whether or not your value of the correlation coefficient supports the researcher's belief.

(1)

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

(4)

Jack is a 40-year-old patient.

(f) (i) Use your regression line to estimate the volume of blood pumped by each contraction of Jack's heart.

(ii) Comment, giving a reason, on the reliability of your estimate.

(2)

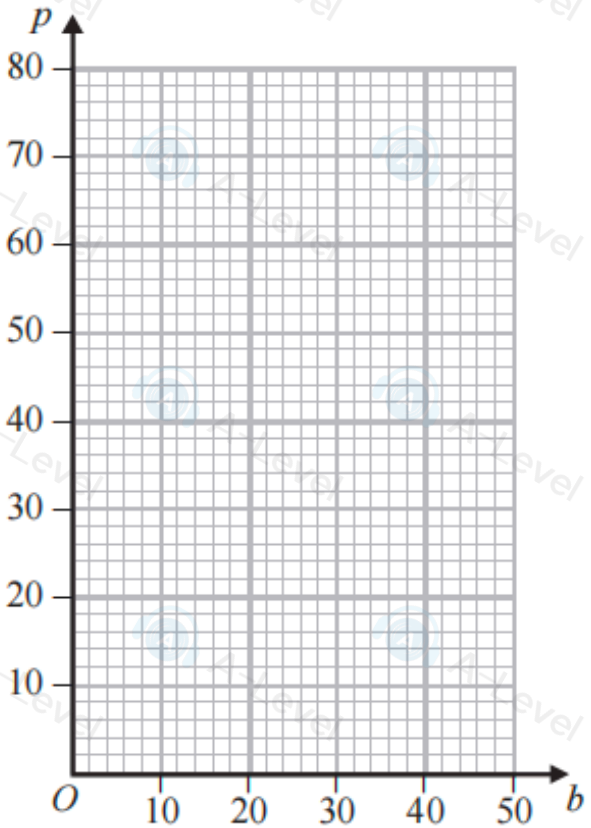
7. A doctor is investigating the correlation between blood protein, p, and body mass index, b.

He takes a random sample of 8 patients and the data are shown in the table below.

Patient	A	B	C	D	E	F	G	H
b	32	36	40	44	42	21	27	37
p	18	21	31	39	21	12	19	70

(a) Draw a scatter diagram of these data on the axes provided.

(2)



The doctor decides to leave out patient H from his calculations.

(b) Give a reason for the doctor's decision.

(1)

For the 7 patients A, B, C, D, E, F and G,

$$S_{bp} = 369, \quad S_{pp} = 490 \quad \text{and} \quad S_{bb} = 423\frac{5}{7}$$

(c) Find the product moment correlation coefficient, r, for these 7 patients.

(2)

(d) Without any further calculations, state how r would differ from your answer in part (c) if it was calculated for all 8 patients.

(1)

5. Franca is the manager of an accountancy firm. She is investigating the relationship between the salary, £x, and the length of commute, y minutes, for employees at the firm. She collected this information from 9 randomly selected employees.

The salary of each employee was then coded using $w = \dfrac{x - 20\,000}{1000}$

The table shows the values of w and y for the 9 employees.

w	6	8	8	–1	25	15	3	–2	19
y	45	50	35	65	25	40	50	75	20

(You may use $\sum w = 81$ $\sum y = 405$ $\sum wy = 2490$ $S_{ww} = 660$ $S_{yy} = 2500$)

(a) Calculate the salary of the employee with $w = -2$

(1)

(b) Show that, to 3 significant figures, the value of the product moment correlation coefficient between w and y is –0.899

(3)

(c) State, giving a reason, the value of the product moment correlation coefficient between x and y

(1)

The least squares regression line of y on w is $y = 60.75 - 1.75w$

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

(3)

(e) Estimate the length of commute for an employee with a salary of £21 000

(2)

Franca uses the regression line to estimate the length of commute for employees with salaries between £25 000 and £40 000

(f) State, giving a reason, whether or not these estimates are reliable.

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO N

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.

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

(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)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

4. A doctor is studying the scans of 30-week old foetuses. She takes a random sample of 8 scans and measures the length, f mm, of the leg bone called the femur. She obtains the following results.

52 53 56 57 57 59 60 62

(a) Show that $S_{ff} = 80$

(3)

The doctor also measures the head circumference, h mm, of each foetus and her results are summarised as

$$\sum h = 2209 \qquad \sum h^2 = 610\,463 \qquad S_{fh} = 182$$

(b) Find S_{hh}

(2)

(c) Calculate the product moment correlation coefficient between the length of the femur and the head circumference for these data.

(2)

The doctor believes that there is a linear relationship between the length of the femur and the head circumference of 30-week old foetuses.

(d) State, giving a reason, whether or not your calculation in part (c) supports the doctor's belief.

(1)

(e) Find an equation of the regression line of h on f.

(4)

The doctor plans in future to measure the femur length, f, and then use the regression line to estimate the corresponding head circumference, h.

A statistician points out that there will always be the chance of an error between the true head circumference and the estimated value of the head circumference.

Given that the error, E mm, has the normal distribution $N(0,4^2)$

(f) find the probability that the estimate is within 3 mm of the true value.

(3)

Leave blank

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)

DO NOT WRITE IN THIS AREA DO NOT W

---

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)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THI