Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 1

Question:

An NHS trust has the following data on operations.

Number of operating theatres	5	6	7	8
Number of operations carried out per day	25	30	35	40

Which is the independent and which is the dependent variable?

Solution:

number of operating theatres – independent variable

number of operations – dependent variable

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 2

Question:

A park ranger collects data on the number of species of bats in a particular area.

Number of suitable habitats	10	24	28
Number of species	1	2	3

Which is the independent and which is the dependent variable?

Solution:

number of suitable habitats – independent variable

number of species – dependent variable

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 3

Question:

The equation of a regression line in the form $y = a + bx$ is to be found. Given that $S_{xx} = 15$, $S_{xy} = 90$, $\bar{x} = 3$ and $\bar{y} = 15$ work out the values of a and b.

Solution:

$b = \frac{90}{15} = \underline{\mathbf{6}}$

$a = 15 - (6 \times 3) = \underline{\mathbf{-3}}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 4

Question:

Given $S_{xx} = 30$, $S_{xy} = 165$, $\bar{x} = 4$ and $\bar{y} = 8$ find the equation of the regression line of y on x.

Solution:

$b = \frac{165}{30} = 5.5$

$a = 8 - (5.5 \times 4) = 8 - 22 = -14$

Equation is: $\underline{\boldsymbol{y = -14 + 5.5x}}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 5

Question:

The equation of a regression line is to be found. The following summary data is given.

$S_{xx} = 40$, $S_{xy} = 80$, $\bar{x} = 6$, $\bar{y} = 12$.

Find the equation of the regression line in the form $y = a + bx$.

Solution:

$b = \frac{80}{40} = 2$

$a = 12 - (2 \times 6) = 0$

Equation is: **$y = 2x$ or $y = 0 + 2x$**

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 6

Question:

Data is collected and summarised as follows.

$\Sigma x = 10$ $\Sigma x^2 = 30$ $\Sigma y = 48$ $\Sigma xy = 140$ $n = 4.$

a Work out $\bar{x}, \bar{y}$, S_{xx} and S_{xy}.

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

Solution:

a $\bar{x} = \frac{10}{4} = 2.5$

$\bar{y} = \frac{48}{4} = 12$

$S_{xx} = \Sigma x^2 - \frac{\left(\sum x\right)^2}{n}$

$S_{xx} = 30 - \frac{10 \times 10}{4} = 30 - 25 = \underline{\mathbf{5}}$

$S_{xy} = \Sigma xy - \frac{\sum x \sum y}{n}$

$S_{xy} = 140 - \frac{10 \times 48}{4} = 140 - 120 = \underline{\mathbf{20}}$

b

$b = \frac{20}{5} = 4$

$a = 12 - (4 \times 2.5) = 12 - 10 = 2$

Equation is: $\underline{\boldsymbol{y = 2 + 4x}}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 7

Question:

For the data in the table:

x	2	4	5	8	10
y	3	7	8	13	17

a calculate S_{xx} and S_{xy},

b find the equation of the regression line of y on x in the form $y = a + bx$.

Solution:

a $\Sigma x = 29$ $\Sigma x^2 = 209$ $\Sigma y = 48$ $\Sigma xy = 348$

$\bar{x} = 5.8$ $\bar{y} = 9.6$ $n = 5$

$$S_{xx} = \Sigma x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$S_{xx} = 209 - \frac{29 \times 29}{5} = 209 - 168.2 = \underline{\mathbf{40.8}}$$

$$S_{xy} = \Sigma xy - \frac{\sum x \sum y}{n}$$

$$S_{xy} = 348 - \frac{29 \times 48}{5} = 348 - 278.4 = \underline{\mathbf{69.6}}$$

b $\bar{x} = \frac{29}{5} = 5.8$ $\bar{y} = \frac{48}{5} = 9.6$

$$b = \frac{69.6}{40.8} = 1.70(58823)$$

$$a = \bar{y} - b\bar{x} = 9.6 - (1.7058823 \times 5.8) = -0.29(41173)$$

Equation is: $\underline{\mathbf{y = -0.294 + 1.71x}}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 8

Question:

A field was divided into 12 plots of equal area. Each plot was fertilised with a different amount of fertilizer (h). The yield of grain (g) was measured for each plot. Find the equation of the regression line of g on h in the form $g = a + bh$ given the following summary data.

$$\Sigma h = 22.09 \quad \Sigma g = 49.7 \quad \Sigma h^2 = 45.04 \quad \Sigma g^2 = 244.83 \quad \Sigma hg = 97.778 \quad n = 12$$

Solution:

$$S_{hh} = 45.04 - \frac{22.09 \times 22.09}{12} = 45.04 - 40.66(4008) = 4.37(5992)$$

$$S_{hg} = 97.778 - \frac{22.09 \times 49.7}{12} = 97.778 - 91.48(9416) = 6.28(8583)$$

$$\bar{h} = \frac{22.09}{12} = 1.84(08333) \quad \bar{g} = \frac{49.7}{12} = 4.14(16666)$$

$$b = \frac{6.288583}{4.375992} = 1.43(70647)$$

$a = 4.1416666 - (1.4370647 \times 1.8408333) = 1.49(627)$

Equation is: $\underline{g = 1.50 + 1.44h}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 9

Question:

An accountant monitors the number of items produced per month by a company (n) together with the total production costs (p). The table shows these data.

Number of items, *n*, (1000s)	21	39	48	24	72	75	15	35	62	81	12	56
Production costs, *p*, (£1000s)	40	58	67	45	89	96	37	53	83	102	35	75

(You may use $\Sigma n = 540$ $\Sigma n^2 = 30\ 786$ $\Sigma p = 780$ $\Sigma p^2 = 56\ 936$ $\Sigma np = 41\ 444$)

a Calculate S_{nn} and S_{np}.

b Find the equation of the regression line of p on n in the form $p = a + bn$.

Solution:

a) $\Sigma n = 540$ $\Sigma n^2 = 30786$ $\Sigma p = 780$ $\Sigma np = 41444$

$S_{nn} = 30786 - \frac{540 \times 540}{12} = 30786 - 24300 = \underline{\mathbf{6486}}$

$S_{np} = 41444 - \frac{540 \times 780}{12} = 41444 - 35100 = \underline{\mathbf{6344}}$

b) $\bar{n} = 45$ $\bar{p} = 65$

$b = \frac{6344}{6486} = 0.97(81066)$

$a = 65 - (0.9781066 \times 45) = 65 - 44.0148 = 20.98(52)$

Equation is: $\underline{\mathbf{y = 20.98 + 0.978x}}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise A, Question 10

Question:

The relationship between the number of coats of paint applied to a boat and the resulting weather resistance was tested in a laboratory. The data collected are shown in the table.

Coats of paint (x)	1	2	3	4	5
Protection (years) (y)	1.4	2.9	4.1	5.8	7.2

a Calculate S_{xx} and S_{xy}.

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

Solution:

a) $\Sigma x = 15$ $\Sigma x^2 = 55$ $\Sigma y = 21.4$ $\Sigma xy = 78.7$ n = 5

$$S_{xx} = \Sigma x^2 - \frac{(\sum x)^2}{n}$$

$$S_{xx} = 55 - \frac{15 \times 15}{5} = \underline{\mathbf{10}}$$

$$S_{xy} = \Sigma xy - \frac{\sum x \sum y}{n}$$

$$S_{xy} = 78.7 - \frac{15 \times 21.4}{5} = 78.7 - 64.2 = \underline{\mathbf{14.5}}$$

b

$\bar{x} = 3$ $\bar{y} = 4.28$

$$b = \frac{14.5}{10} = 1.45$$

$a = 4.28 - (1.45 \times 3) = 4.28 - 4.35 = -0.07$

Equation is: $\underline{\boldsymbol{y = -0.07 + 1.45x}}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise B, Question 1

Question:

Given that the coding $p = x + 2$ and $q = y - 3$ has been used to get the regression equation $p + q = 5$ find the equation of the regression line of y on x in the form $y = a + bx$.

Solution:

$(x + 2) + (y - 3) = 5$

$x + y - 1 = 5$

$\boldsymbol{y = 6 - x}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise B, Question 2

Question:

Given the coding $x = p - 10$ and $y = s - 100$ and the regression equation $x = y + 2$ work out the equation of the regression line of s on p.

Solution:

$p - 10 = s - 100 + 2$

$\boldsymbol{s = p + 88}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise B, Question 3

Question:

Given that the coding $g = \frac{x}{3}$ and $h = \frac{y}{4} - 2$ has been used to get the regression equation $h = 6 - 4g$ find the equation of the regression line of y on x.

Solution:

$\frac{y}{4} - 2 = 6 - 4\left(\frac{x}{3}\right)$

$3y - 24 = 72 - 16x$ (multiply through by 12)

$3y = 96 - 16x$ so $y = 32 - \frac{16}{3}x$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise B, Question 4

Question:

The regression line of t on s is found by using the coding $x = s - 5$ and $y = t - 10$.

The regression equation of y on x is $y = 14 + 3x$.

Work out the regression line of t on s.

Solution:

$t - 10 = 14 + 3(s - 5)$

$t = 24 + 3s - 15$

$\mathbf{t = 9 + 3s}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise B, Question 5

Question:

A regression line of c on d is worked out using the coding $x = \frac{c}{2}$ and $y = \frac{d}{10}$.

a Given $S_{xy} = 120$, $S_{xx} = 240$, the mean of $x(\bar{x})$ is 5 and the mean of $y(\bar{y})$ is 6, calculate the regression line of y on x.

b Find the regression line of d on c.

Solution:

a $b = \frac{120}{240} = 0.5$

$a = 6 - 0.5 \times 5 = 3.5$

$\boldsymbol{y = 3.5 + 0.5x}$

b $\frac{d}{10} = 3.5 + 0.5 \times \frac{c}{2}$ (multiply by 10)

$\boldsymbol{d = 35 + 2.5c}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise B, Question 6

Question:

Some data on heights (h) and weights (w) were collected. The results were coded such that $x = \frac{h-8}{2}$ and $y = \frac{w}{5}$. The coded results are shown in the table.

x	1	5	10	16	17
y	9	12	16	21	23

a Calculate S_{xy} and S_{xx} and use them to find the equation of the regression line of y on x.

b Find the equation of the regression line of w on h.

Solution:

a $\Sigma x = 49$ $\Sigma x^2 = 671$ $\Sigma y = 81$ $\Sigma xy = 956$

$S_{xy} = 956 - \frac{49 \times 81}{5} = 956 - 793.8 = \mathbf{162.2}$

$S_{xx} = 671 - \frac{\left(\sum x\right)^2}{5} = 671 - \frac{2401}{5} = 671 - 480.2 = \mathbf{190.8}$

$\bar{y} = 16.2$ $\bar{x} = 9.8$

$b = \frac{162.2}{190.8} = 0.85 \dots \dots ..$

$a = 16.2 - (0.85 \times 9.8) = 16.2 - 8.33 \dots.. = 7.87$

Equation of y on x is: $\mathbf{y = 7.87 + 0.85x}$

b $\frac{w}{5} = 7.87 + 0.85\left(\frac{h-8}{2}\right)$ (multiply by 5)

$w = 39.35 + 2.125(h - 8)$

$w = 39.35 + 2.125h - 17$

$\mathbf{w = 22.35 + 2.125h}$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 1

Question:

Given the regression line $y = 24 - 3x$ find the value of y when x is 6.

Solution:

$y = 24 - (3 \times 6) = 6$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 2

Question:

The regression line for the weight (w) in grams on the volume (v) in cm^3 for a sample of small marbles is $w = 300 + 12v$.

Calculate the weight when the volume is 7 cm^3.

Solution:

$w = 300 + (12 \times 7) = 384$

Weight is **384** grams

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 3

Question:

a State what is meant by extrapolation.

b State what is meant by interpolation.

Solution:

a Extrapolation means using the regression line to estimate outside the range of the data collected. It can be unreliable.

b Interpolation means using the regression line to estimate within the range of the data collected. It is usually reasonably reliable.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 4

Question:

12 children between the ages (x) of five and 11 years were asked how much pocket money (y) they were given each week. The equation for the regression line of y on x was found to be $y = 2x - 8$.

a Use the equation to estimate the amount of money a seven year old would get. State, with a reason, whether or not this is a reliable estimate.

b One of the children suggested that this equation must be wrong since it showed that a three year old would get a negative amount of pocket money. Explain why this has happened.

Solution:

a $y = (2 \times 7) - 8 = 6$

A 7 year old would get **£6.00**. This is a reasonable estimate as 7 years is within the range of ages asked. It is interpolation.

b This would involve extrapolation, which may not be reliable. Three years old is outside the range of ages asked. A three year old is probably not given pocket money.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 5

Question:

The pulse rates (y) of 10 people were measured after doing different amounts of exercise (x) for between two and 10 minutes. The regression equation $y = 0.75x + 72$ refers to these data. The equation seems to suggest that someone doing 30 minutes of exercise would have a pulse rate of 94.5. State whether or not this is sensible. Give a reason for your answer.

Solution:

This is not a sensible estimate since the data collected only covers 2 to 10 minutes. The answer 94.5 involves extrapolation. In fact the pulse rate can not keep rising. 30 minutes is a long way outside the limits of the data.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 6

Question:

Over a period of time the sales, (y) in thousands, of 10 similar text books and the amount, (x) in £ thousands, spent on advertising each book were recorded. The greatest amount spent on advertising was £4.4 thousand, and the least amount was £0.75 thousand.

An equation of the regression line for y on x was worked out for the data.

The equation was $y = 0.93 + 1.1x$.

a Use the equation to estimate the sales of a text book if the amount spent on advertising is to be set at £2.65 thousand. State, with a reason, whether or not this is a reliable estimate.

b Use the equation to estimate the sales of the book if the amount to be spent on advertising is £8000. State, with a reason, whether or not this is a reliable estimate.

c Explain what the value 1.1 tells you about the relationship between the sales of books and the amount spent on advertising.

d Interpret the meaning of the figure 0.93.

Solution:

a $y = 0.93 + (1.1 \times 2.65) =$ **3.845** thousands. This is reasonably reliable since £2.65 thousand is within the range of the collected data. It involves interpolation.

b $y = 0.93 + (1.1 \times 8) = 9.73$ thousands. This may be unreliable since £8 thousand is outside the range of the data collected. It involves extrapolation.

c 1.1 thousand is the number of extra books sold for each £1 thousand spent on advertising.

d 0.93 thousand is the number of books likely to be sold if there is no money spent on advertising. This is only just outside the range of values so it is a reasonably reliable estimate.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 7

Question:

Research was done to see if there is a relationship between finger dexterity and the ability to do work on a production line. The data is shown in the table.

Dexterity Score (x)	2.5	3	3.5	4	5	5	5.5	6.5	7	8
Productivity (y)	80	130	100	220	190	210	270	290	350	400

The equation of the regression line for these data is $y = -59 + 57x$

a Use the equation to estimate the productivity of someone with a dexterity of 6.

b State the contextual meaning of the figure 57.

c State, giving in each case a reason, whether or not it would be reasonable to use this equation to work out the productivity of someone with dexterity of:

i2 **ii** 14.

Solution:

a $y = -59 + (57 \times 6) = 283$

b 57 is the gradient. For every one rise in the dexterity production rises by 57.

c i) 2 may be a little unreliable since it lies just outside the values in the table. It would involve extrapolation.

ii) 14 could be very unreliable as it lies well outside the range of the values in the table. It would involve extrapolation.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise C, Question 8

Question:

A regression line of $y = 5 + 3x$ is found using 10 data sets. Another piece of data is recorded which when put on a scatter diagram with the original data proves to be well above the regression line for all the other data. Write down whether or not the regression equation would change if this piece of data were included in the calculation.

Solution:

The equation of the regression line would change. The line is likely to tilt.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 1

Question:

A metal rod was found to increase in length as it was heated. The temperature (t) and the increase in length (l mm) were measured at intervals between 30°C and 400°C degrees. The regression line of l on t was found to be $l = 0.009t - 0.25$.

a Find the increase in length for a temperature of 300°C.

b Find the increase in length for a temperature of 530°C.

c Write down, with reasons, why the answer to **a** might be reliable and the answer **b** unreliable.

Solution:

a $l = (0.009 \times 300) - 0.25.$ = **2.45mm**

b $l = (0.009 \times 530) - 0.25.$ = **4.52mm**

c a is likely to be a reasonable estimate since it involves interpolation. 300°C is within the range of the data covered.

b involves extrapolation so is likely to be unreliable. 530°C is outside the range of the data used. The metal might well melt before this temperature is reached.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 2

Question:

Two variables s and t are thought to be connected by a law of the form $t = a + bs$, where a and b are constants.

a Use the summary data:

$\Sigma s = 553$ $\Sigma t = 549$ $\Sigma st = 31\,185$ $n = 12$ $\bar{s} = 46.0833$
$\bar{t} = 45.75$ $S_{ss} = 6193$

to work out the regression line of t on s.

b Find the value of t when s is 50

Solution:

a $S_{st} = 31185 - \frac{553 \times 549}{12} = 31185 - 25299.75 = 5885.25$

$b = \frac{5885.25}{6193} = 0.959(3067)$

$a = 45.75 - (0.950\ldots \times 46.083)$

$a = 45.75 - 43.792(988)$

$a = 1.957$

$t = 1.96 + 0.95s$

b $t = 1.96 + (0.95 \times 50)$

$t = 49.46$

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 3

Question:

A biologist recorded the breadth (x cm) and the length (y cm) of 12 beech leaves. The data collected can be summarised as follows.

$\Sigma x^2 = 97.73$ $\Sigma x = 33.1$ $\Sigma y = 66.8$ $\Sigma xy = 195.94$

a Calculate S_{xx} and S_{xy}.

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

c Predict the length of a beech leaf that has a breadth of 3.0 cm.

Solution:

a $S_{xx} = 97.73 - \frac{33.1 \times 33.1}{12} = 97.73 - 91.30 = \mathbf{6.43}$

$S_{xy} = 195.94 - \frac{33.1 \times 66.8}{12} = 195.94 - 184.26 = \mathbf{11.68}$

b $\bar{x} = \frac{33.1}{12} = 2.76 \bar{y} = \frac{66.8}{12} = 5.57$

$b = \frac{11.68}{6.43} = 1.82$

$a = 5.57 - (1.82 \times 2.76) = 5.57 - 5.02 = 0.55$

Equation is: $\boldsymbol{y = 0.55 + 1.82x}$

c length = 0.55 + (1.82 × 3) = **6.01** cm

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 4

Question:

Energy consumption is claimed to be a good predictor of Gross National Product. An economist recorded the energy consumption (x) and the Gross National Product (y) for eight countries. The data are shown in the table.

Energy Consumption *x*	3.4	7.7	12.0	75	58	67	113	131
Gross National Product *y*	55	240	390	1100	1390	1330	1400	1900

a Calculate S_{xy} and S_{xx}.

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

c Estimate the Gross National Product of a country that has an energy consumption of 100.

d Estimate the energy consumption of a country that has a Gross National Product of 3500.

e Comment on the reliability of your answer to **d**.

Solution:

a

$\Sigma x^2 = 43622.85$ $\Sigma x = 467.1$ $\Sigma y = 7805$ $\Sigma xy = 666045$

$S_{xx} = 43622.85 - \frac{467.1 \times 467.1}{8} = 43622.85 - 27272.80 \ldots. = \mathbf{16350.048\ldots}$

$S_{xy} = 666045 - \frac{467.1 \times 7805}{8} = 666045 - 455714.43 = \mathbf{210330.56\ldots}$

b $\bar{x} = 58.3875$ $\bar{y} = 975.625$

$b = \frac{210330.56\ldots}{16350.04\ldots} = 12.864\ldots\ldots$

$a = 975.625 - (12.86\ldots \times 58.3875) = 975.625 - 751.10939 = 224.515\ldots\ldots$

Equation is: $\boldsymbol{y = 224.52 + 12.86x}$

c Gross National product = 224.515.. + (12.86… × 100) = **1510.5**

d $3500 = 224.515\ldots + 12.864..x$

$12.864..x = 3500 - 224.515\ldots$

$x = \frac{3275.48..}{12.864..} = 254.6$

Energy consumption = **254.6**

e This answer is likely to be unreliable as it involves extrapolation. 3500 is well outside the limits of the data set used.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 5

Question:

In an environmental survey on the survival of mammals the tail length t (cm) and body length m (cm) of a random sample of six small mammals of the same species were measured.

These data are coded such that $x = \frac{m}{2}$ and $y = t - 2$.

The data from the coded records are summarised below.

$\Sigma y = 13.5$ $\Sigma x = 25.5$ $\Sigma xy = 84.25$ $S_{xx} = 59.88$

a Find the equation of the regression line of y on x in the form $y = ax + b$.

b Hence find the equation of the regression line of t on m.

c Predict the tail length of a mammal that has a body length of 10 cm.

Solution:

a $S_{xy} = 84.25 - \frac{25.5 \times 13.5}{6} = 84.25 - 57.375 = 26.875$

$\bar{x} = 4.25 \quad \bar{y} = 2.25$

$b = \frac{26.875}{59.88} = 0.4488 \ldots ..$

$a = 2.25 - (0.4488\ldots \times 4.25) = 2.25 - 1.9074 = 0.3425\ldots$

Equation is $\boldsymbol{y = 0.343 + 0.449x}$

b $t - 2 = 0.343.. + 0.448..\left(\frac{m}{2}\right)$

$\boldsymbol{t = 2.343 + 0.224m}$

c tail length = 2.343 + (0.224 × 10) = **4.58 cm**

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 6

Question:

A health clinic counted the number of breaths per minute (r) and the number of pulse beats (p) per minute for 10 people doing various activities. The data are shown in the table.

The data are coded such that $x = \frac{r-10}{2}$ and $y = \frac{p-50}{2}$.

x	3	5	5	7	8	9	9	10	12	13
y	4	9	10	11	17	15	17	19	22	27

(You may use $\Sigma x = 81$ $\Sigma x^2 = 747$ $\Sigma y = 151$ $\Sigma y^2 = 2695$ $\Sigma xy = 1413$.)

a Calculate S_{xy} and S_{xx}.

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

c Find the equation of the regression line for p on r.

d Estimate the number of pulse beats per minute for someone who is taking 22 breaths per minute.

e Comment on the reliability of your answer to **e**.

Solution:

a

$\Sigma x = 81$ $\Sigma y = 151$ $\Sigma x^2 = 747$ $\Sigma xy = 1413$ $\bar{x} = 8.1$ $\bar{y} = 15.1$

$S_{xx} = 747 - \frac{81 \times 81}{10} = 747 - 656.1 = \mathbf{90.9}$

$S_{xy} = 1413 - \frac{81 \times 151}{10} = 1413 - 1223.1 = \mathbf{189.9}$

b $b = \frac{189.9}{90.9} = 2.089 \ldots..$

$\bar{x} = 8.1$ $\bar{y} = 15.1$

$\boldsymbol{a} = 15.1 - (2.089\ldots \times 8.1) = 15.1 - 16.9217\ldots. = -1.82\ldots$

Equation is: $\mathbf{y = -1.82 + 2.09x}$

c $\frac{p-50}{2} = -1.82 + 2.09\left(\frac{r-10}{2}\right)$ (multiply by 2)

$p - 50 = -3.64 + 2.09r - 20.9$

$\mathbf{p = 25.46 + 2.09r}$

d Pulse Beats = 25.46 + (2.09 × 22) = **71.44**

e The answer to d is reasonably reliable since it involves interpolation. 22 is within the range of the data set used.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 7

Question:

A farm food supplier monitors the number of hens kept (x) against the weekly consumption of hen food (y kg) for a sample of 10 small holders. He records the data and works out the regression line for y on x to be $y = 0.16 + 0.79x$.

a Write down a practical interpretation of the figure 0.79.

b Estimate the amount of food that is likely to be needed by a small holder who has 30 hens.

c If food costs £12 for a 10 kg bag estimate the weekly cost of feeding 50 hens.

Solution:

a 0.79 kg is the average amount of food consumed in 1 week by 1 hen.

b $y = 0.16 + 0.79 \times 30 =$ **23.86 kg**

c Cost = $(0.16 + 0.79 \times 50)$£12 = **£475.92**

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 8

Question:

Water voles are becoming very rare; they are often confused with water rats. A naturalist society decided to record details of the water voles in their area. The members measured the weight (y) to the nearest 10 grams, and the body length (x) to the nearest millimetre, of eight active healthy water voles. The data they collected are in the table.

Body Length (x) mm	140	150	170	180	180	200	220	220
Weight (y) grams	150	180	190	220	240	290	300	310

a Draw a scatter diagram of these data.

b Give a reason to support the calculation of a regression line for these data.

c Use the coding $l = \frac{x}{10}$ and $w = \frac{y}{10}$ to work out the regression line of w on l.

d Find the equation of the regression line for y on x.

e Draw the regression line on the scatter diagram.

f Use your regression line to calculate an estimate for the weight of a water vole that has a body length of 210 mm. Write down, with a reason, whether or not this is a reliable estimate.

The members of the society remove any water voles that seem unhealthy from the river and take them into care until they are fit to be returned.

They find three water voles on one stretch of river which have the following measurements.

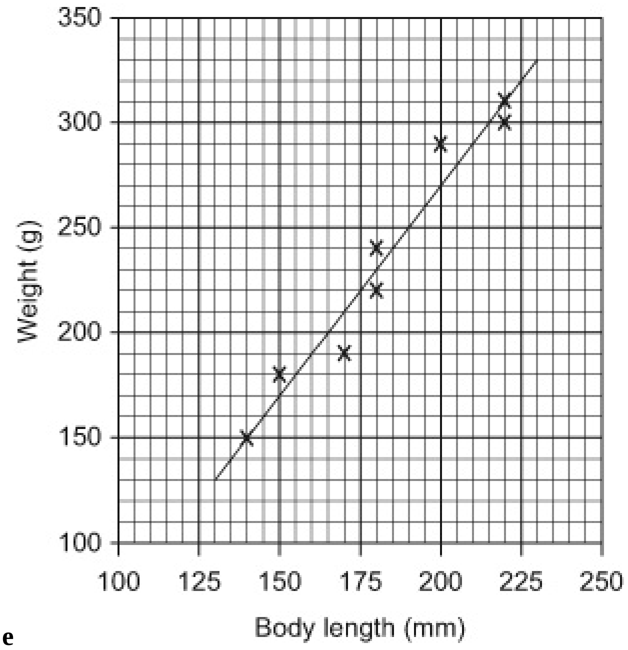
A: Weight 235 gm and body length 180 mm

B: Weight 180 gm and body length 200 mm

C: Weight 195 gm and body length 220 mm

g Write down, with a reason, which of these water voles were removed from the river.

Solution:



a & e

b There appears to be a linear relationship between body length and body weight.

c

l	14	15	17	18	18	20	22	22
w	15	18	19	22	24	29	30	31

$\Sigma l^2 = 2726$ $\Sigma l = 146$ $\Sigma w = 188$ $\Sigma lw = 3553$

$\bar{l} = 18.25$ $\bar{w} = 23.5$

$S_{ll} = 2726 - \frac{146 \times 146}{8} = 2726 - 2664.5 = 61.5$

$S_{lw} = 3553 - \frac{146 \times 188}{8} = 3553 - 3431 = 122$

$b = \frac{122}{61.5} = 1.9837 \ldots \ldots$

$a = 23.5 - (1.9837 \ldots \times 18.25) = 23.5 - 36.2032\ldots.. = -12.70\ldots$

Equation is: $\boldsymbol{w = -12.7 + 1.98l}$

d $\frac{y}{10} = -12.7 + \left(1.98 \times \frac{x}{10}\right)$ (multiply through by 10)

$\boldsymbol{y = -127 + 1.98x}$

e *See diagram*

f Tail length = − 127 + 1.98 × 210 = **288.8mm**

This is a reliable estimate since it involves interpolation. 210 is within the range of the data.

g B and C are both underweight so were probably removed from the river. A is slightly overweight so was probably left in the river.

Solutionbank S1
Edexcel AS and A Level Modular Mathematics

Regression
Exercise D, Question 9

Question:

A mail order company pays for postage of its goods partly by destination and partly by total weight sent out on a particular day. The number of items sent out and the total weights were recorded over a seven day period. The data are shown in the table.

Number of items (n)	10	13	22	15	24	16	19
Weight in kg (w)	2800	3600	6000	3600	5200	4400	5200

a Use the coding $x = n - 10$ and $y = \frac{w}{400}$ to work out S_{xy} and S_{xx}.

b Work out the equation of the regression line for y on x.

c Work out the equation of the regression line for w on n.

d Use your regression equation to estimate the weight of 20 items.

e State why it would be unwise to use the regression equation to estimate the weight of 100 items.

Solution:

a

Coded number x	0	3	12	5	14	6	9
Coded weight y	7	9	15	9	13	11	13

$\Sigma x = 49$ $\Sigma x^2 = 491$ $\Sigma y = 77$ $\Sigma xy = 617$

$S_{xy} = 617 - \frac{49 \times 77}{7} = 617 - 539 = \mathbf{78}$

$S_{xx} = 491 - \frac{49^2}{7} = 491 - 343 = \mathbf{148}$

b $\bar{y} = 11$ $\bar{x} = 7$

$b = \frac{78}{148} = 0.5270 \ldots \ldots$

$a = 11 - (0.5270\ldots \times 7) = 11 - 3.6891\ldots\ldots = 7.3108\ldots$

Equation is: $\boldsymbol{y = 7.31 + 0.53x}$ (to two decimal places)

c $\frac{w}{400} = 7.31 + 0.527(n - 10)$ (multiply by 400)

$w = 2924 + 210.8(n - 10)$

$w = 2924 + 210.8n - 2108$

Equation is: $\boldsymbol{w = 816 + 211n}$

d $w = 816 + 211 \times 20 =$ **5036**

e This is far outside the range of values. This is extrapolation.