

Regression

1

The table below shows the packing times for a particular employee for a random sample of orders in a mail order company.

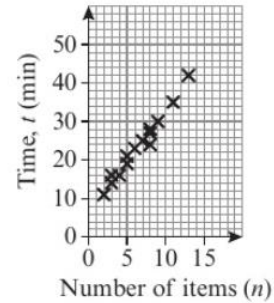
Number of items (n)	2	3	3	4	5	5	6	7	8	8	8	9	11	13
Time (t min)	11	14	16	16	19	21	23	25	24	27	28	30	35	42

A scatter diagram was drawn to represent the data.

a Describe the correlation between number of items packed and time taken. **(1 mark)**

The equation of the regression line of t on n is $t = 6.3 + 2.64n$.

b Give an interpretation of the value 2.64. **(1 mark)**



2 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 is shown in the table.

a Draw a scatter diagram to represent this data.

The equation of the regression line is $y = 2.93 + 1.45x$.

Helen says that a gradient of 1.45 means that if 10 coats of paint are applied the protection will last 14.5 years.

b Comment on Helen's statement.

Coats of paint (x)	Protection (years) (y)
1	4.4
2	5.9
3	7.1
4	8.8
5	10.2

3

James placed different masses (m) on a spring and measured the resulting length of the spring (s) in centimetres. The smallest mass was 20 g and the largest mass was 100 g.

He found the equation of the regression line of s on m to be $s = 44 + 0.2m$.

a Interpret the values 44 and 0.2 in this context. **(2 marks)**

b Explain why it would not be sensible to use the regression equation to work out:

i the value of s when $m = 150$ **ii** the value of m when $s = 60$. **(2 marks)**

4

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 is 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

The equation of the regression line of y on x is $y = 225 + 12.9x$.

The economist uses this regression equation to estimate the energy consumption of a country with a Gross National Product of 3500.

Give two reasons why this may not be a valid estimate. **(2 marks)**

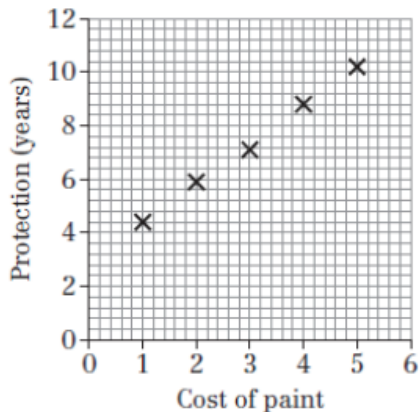
Answers to regression

1

- a There is strong positive correlation.
- b If the number of items increases by 1, the time taken increases by approximately 2.64 minutes.

2

a



- b There are two key problems with Helen's statement:

First, 10 coats of paint is very far outside our range of given data, and we cannot assume that this linear relationship continues as we extrapolate, so using the regression line is not necessarily valid.

Second, even if we accept the extrapolation as valid, a gradient of 1.45 means that, for every extra coat of paint, the protection will increase by 1.45 years. Therefore, if 10 coats of paint are applied, the protection will be 14.5 years longer than if no paint were applied. Helen has, however, forgotten to include the constant 2.93 years, which is the weather resistance if no paint were applied. After 10 coats of paint the protection will last approximately $2.93 + 14.5 = 17.43$ years.

3

- a 44 cm is the length of the spring with no mass attached. If a mass of 1 g is attached, the spring will increase in length by approximately 0.2 cm.
- b i $m = 150$ is outside the range of the data (extrapolation) so is less likely to be accurate.
ii This particular regression equation should only be used to predict a value of s given m . To predict a value of m given s , you should use the regression equation of m on s .

4

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

The economist used the regression line of y on x . Estimating energy consumption (x) from Gross National Product (y) would need the regression line of x on y .