

Regression Line and Estimates

Question Paper 3

Level	A Level
Subject	Maths
Exam Board	OCR
Module	Statistics 1
Topic	Bivariate Data
Sub Topic	Regression Line and Estimates
Booklet	Question Paper - 3

Time Allowed: 58 minutes

Score: /48

Percentage: /100

Grade Boundaries:

A*	A	B	C	D	E	U
>85%	77.5%	70%	62.5%	57.5%	45%	<45%

- 1 A city council attempted to reduce traffic congestion by introducing a congestion charge. The charge was set at £4.00 for the first year and was then increased by £2.00 each year. For each of the first eight years, the council recorded the average number of vehicles entering the city centre per day. The results are shown in the table.

Charge, £x	4	6	8	10	12	14	16	18
Average number of vehicles per day, y million	2.4	2.5	2.2	2.3	2.0	1.8	1.7	1.5

$$[n = 8, \Sigma x = 88, \Sigma y = 16.4, \Sigma x^2 = 1136, \Sigma y^2 = 34.52, \Sigma xy = 168.6.]$$

- (i) Calculate the product moment correlation coefficient for these data. [3]
- (ii) Explain why x is the independent variable. [1]
- (iii) Calculate the equation of the regression line of y on x . [4]
- (iv) (a) Use your equation to estimate the average number of vehicles which will enter the city centre per day when the congestion charge is raised to £20.00. [2]
- (b) Comment on the reliability of your estimate. [2]
- (v) The council wishes to estimate the congestion charge required to reduce the average number of vehicles entering the city per day to 1.0 million. Assuming that a reliable estimate can be made by extrapolation, state whether they should use the regression line of y on x or the regression line of x on y . Give a reason for your answer. [2]
- 2 (i) Some values, (x, y) , of a bivariate distribution are plotted on a scatter diagram and a regression line is to be drawn. Explain how to decide whether the regression line of y on x or the regression line of x on y is appropriate. [2]
- (ii) In an experiment the temperature, $x^\circ\text{C}$, of a rod was gradually increased from 0°C , and the extension, $y\text{ mm}$, was measured nine times at 50°C intervals. The results are summarised below.
- $$n = 9 \quad \Sigma x = 1800 \quad \Sigma y = 14.4 \quad \Sigma x^2 = 510\,000 \quad \Sigma y^2 = 32.6416 \quad \Sigma xy = 4080$$
- (a) Show that the gradient of the regression line of y on x is 0.008 and find the equation of this line. [4]
- (b) Use your equation to estimate the temperature when the extension is 2.5 mm. [1]
- (c) Use your equation to estimate the extension for a temperature of -50°C . [1]
- (d) Comment on the meaning and the reliability of your estimate in part (c). [2]

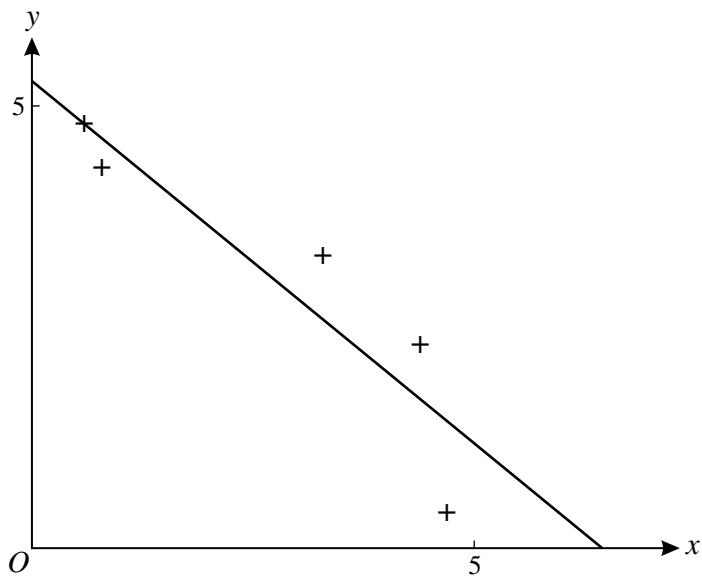
- 3 (i) Some values, (x, y) , of a bivariate distribution are plotted on a scatter diagram and a regression line is to be drawn. Explain how to decide whether the regression line of y on x or the regression line of x on y is appropriate. [2]

- (ii) In an experiment the temperature, x °C, of a rod was gradually increased from 0 °C, and the extension, y mm, was measured nine times at 50 °C intervals. The results are summarised below.

$$n = 9 \quad \Sigma x = 1800 \quad \Sigma y = 14.4 \quad \Sigma x^2 = 510\,000 \quad \Sigma y^2 = 32.6416 \quad \Sigma xy = 4080$$

- (a) Show that the gradient of the regression line of y on x is 0.008 and find the equation of this line. [4]
- (b) Use your equation to estimate the temperature when the extension is 2.5 mm. [1]
- (c) Use your equation to estimate the extension for a temperature of -50 °C. [1]
- (d) Comment on the meaning and the reliability of your estimate in part (c). [2]

- 4 The diagram shows the results of an experiment involving some bivariate data. The least squares regression line of y on x for these results is also shown.



- (i) Given that the least squares regression line of y on x is used for an estimation, state which of x or y is treated as the independent variable. [1]
- (ii) Use the diagram to explain what is meant by 'least squares'. [2]
- (iii) State, with a reason, the value of Spearman's rank correlation coefficient for these data. [2]
- (iv) What can be said about the value of the product moment correlation coefficient for these data? [1]

- 5 The table shows some of the values of the seasonally adjusted Unemployment Rate (UR), $x\%$, and the Consumer Price Index (CPI), $y\%$, in the United Kingdom from April 2008 to July 2010.

Date	April 2008	July 2008	October 2008	January 2009	April 2009	July 2009	October 2009	January 2010	April 2010	July 2010
UR, $x\%$	5.2	5.7	6.1	6.8	7.5	7.8	7.8	7.9	7.8	7.7
CPI, $y\%$	3.0	4.4	4.5	3.0	2.3	1.8	1.5	3.5	3.7	3.1

These data are summarised below.

$$n = 10 \quad \Sigma x = 70.3 \quad \Sigma x^2 = 503.45 \quad \Sigma y = 30.8 \quad \Sigma y^2 = 103.94 \quad \Sigma xy = 211.9$$

- (i) Calculate the product moment correlation coefficient, r , for the data, showing that $-0.6 < r < -0.5$. [3]
- (ii) Karen says “The negative value of r shows that when the Unemployment Rate increases, it causes the Consumer Price Index to decrease.” Give a criticism of this statement. [1]
- (iii) (a) Calculate the equation of the regression line of x on y . [3]
- (b) Use your equation to estimate the value of the Unemployment Rate in a month when the Consumer Price Index is 4.0%. [2]
- 6 The table shows the load a lorry was carrying, x tonnes, and the fuel economy, y km per litre, for 8 different journeys. You should assume that neither variable is controlled.

Load (x tonnes)	5.1	5.8	6.5	7.1	7.6	8.4	9.5	10.5
Fuel economy (y km per litre)	6.2	6.1	5.9	5.6	5.3	5.4	5.3	5.1

$$n = 8 \quad \Sigma x = 60.5 \quad \Sigma y = 44.9 \quad \Sigma x^2 = 481.13 \quad \Sigma y^2 = 253.17 \quad \Sigma xy = 334.65$$

- (i) Calculate the equation of the regression line of y on x . [4]
- (ii) Estimate the fuel economy for a load of 9.2 tonnes. [2]
- (iii) An analyst calculated the equation of the regression line of x on y . Without calculating this equation, state the coordinates of the point where the two regression lines intersect. [1]
- (iv) Describe briefly the method required to estimate the load when the fuel economy is 5.8 km per litre. [2]