

Write your name here

Surname

Other Names

AS/A Level Mathematics

Correlation and Regression

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled..
- Answer the questions in the spaces provided
– there may be more space than you need.
- You should show sufficient working to make your methods clear.
Answers without working may not gain full credit.
- Answers should be given to three significant figures unless otherwise stated.

Information

- The marks for **each** question are shown in brackets
– use this as a guide as to how much time to spend on each question.

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

- 1 Using the large data set Colin studied the relationship between rainfall (r) and temperature (t) in Camborne. He took a sample of 12 days from May and June 1987 and obtained the following results.

| | | | | | | | | | | | | |
|-------------------------|------|-----|-----|-----|------|------|------|------|------|------|------|------|
| Rainfall (cm) | 3.1 | 0.1 | 6 | 2.2 | 0.3 | 4.2 | 1.7 | 7.5 | 0.1 | 7.1 | 3.9 | 3.1 |
| Temperature (°C) | 10.7 | 8.9 | 8.8 | 9.2 | 11.1 | 10.2 | 12.6 | 10.4 | 11.3 | 11.6 | 12.8 | 13.5 |

- (a) Use your knowledge of the large data set to explain why it is unlikely that Colin used a random sample.

Colin used a computer program to obtain the following statistics for Rainfall

$$Q_1 = 1.35$$

$$Q_2 = 3.1$$

$$Q_3 = 4.65$$

- (b) Determine whether there are any outliers in the sample.
- (c) Plot the information on a scatter graph
- (d) Explain why a linear regression model may not be suitable for this data.

(Total for question 1 is 4 marks)

- 2 A football coach measured the heights and weights of 12 players, The data is shown below.

| | | | | | | | | | | | | |
|--------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Height (cm) | 188 | 194 | 178 | 175 | 185 | 175 | 188 | 193 | 180 | 190 | 181 | 169 |
| Weight (kg) | 70 | 100 | 83 | 69 | 77 | 58 | 90 | 86 | 71 | 94 | 68 | 61 |

- (a) Draw a scatter graph for this information.
- (b) Give an interpretation of the correlation between the height and weight of the footballers.
- The equation of the regression line is $w = 1.37h - 173$
- (c) Give an interpretation of the gradient of this regression line.
- (d) Use the equation of the regression line to estimate the weight of a player who is 170cm tall.
- (e) Comment on the reliability of your estimate in part (d), giving a reason for your answer.

(Total for question 1 is 4 marks)

- 3 Ross did an investigation into the relationship between study into temperature (t) and total hours of sunshine (s) at Heathrow. He finds that the equation of the regression line is $s = 0.25t + 1.3$.

Give an interpretation of the figure 0.25 figure in this regression line.

(Total for question 1 is 4 marks)