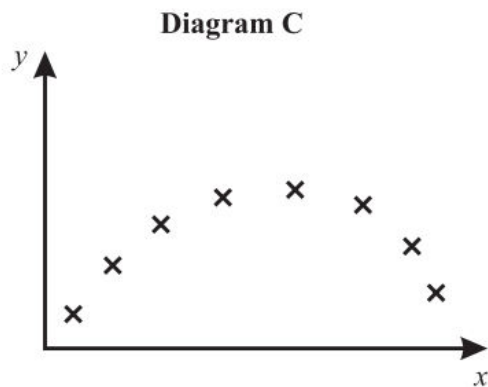
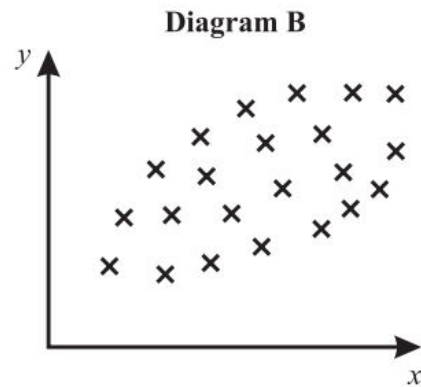
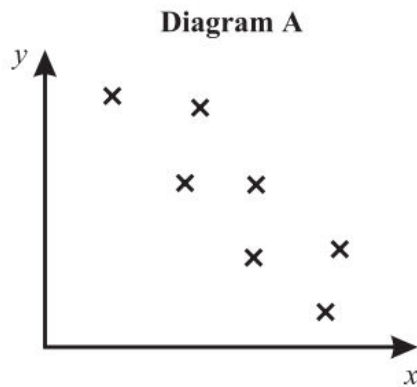


S1 Correlation Challenge

Challenge 1

- (a) Estimate, without undertaking any calculations, the value of the product moment correlation coefficient in each of the scatter diagrams below.



(5 marks)

- (b) The product moment correlation coefficient is an unsuitable measure of association for the data illustrated in one of the diagrams in part (a). State, giving a reason, which diagram this is. (2 marks)



Challenge 2

[A sheet of graph paper is provided for use in this question.]

The following table gives information on the engine capacity and price of models of car in January 2000.

| Model | A | B | C | D | E | F | G | H | I | J |
|---------------------------|-------|-------|--------|--------|--------|-------|-------|-------|-------|--------|
| Engine Capacity, x (cc) | 1 000 | 1 270 | 1 750 | 2 230 | 1 990 | 600 | 650 | 1 500 | 1 450 | 1 650 |
| Price, y (£'s) | 8 800 | 9 400 | 11 320 | 14 960 | 14 860 | 5 980 | 6 400 | 9 320 | 7 100 | 13 300 |

- (a) Plot a scatter diagram of these data. (3 marks)
- (b) Find the equation of the regression line of price on engine capacity. Draw the line on your scatter diagram. (7 marks)
- (c) Calculate the residuals for models **I** and **J**. (3 marks)
- (d) Interpret, briefly, the residuals you have calculated for a customer who requires a model with a low price relative to its engine capacity. (2 marks)



Challenge 3

The following table shows the weekly gas and electricity consumption for a house in Manchester for a sample of nine weeks in 2001.

| Week | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|----------------------------------|-----|----|----|-----|-----|----|-----|-----|-----|
| Gas consumption, x kWh | 312 | 46 | 23 | 406 | 350 | 67 | 295 | 247 | 110 |
| Electricity consumption, y kWh | 84 | 57 | 54 | 96 | 82 | 63 | 59 | 73 | 60 |

- (a) Calculate the value of the product moment correlation coefficient. *(3 marks)*
- (b) The householder expects that, for weeks when the gas consumption is high, the electricity consumption will be low and vice versa.
- (i) State to what extent the value you calculated in part (a) confirms or denies the householder's expectation. *(2 marks)*
- (ii) Give a reason why the value you calculated in part (a) is plausible. *(2 marks)*



Final Challenge

[A sheet of graph paper is provided for use in this question.]

Hennie, a statistics teacher, is an unenthusiastic gardener. During the summer months she records the time, y minutes, it takes her to cut her lawn together with the time, x days, since she last cut it.

| | | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|----|
| x | 5 | 13 | 10 | 19 | 7 | 12 | 21 | 18 | 24 | 12 |
| y | 23 | 32 | 28 | 48 | 21 | 35 | 44 | 39 | 50 | 39 |

- (a) Plot a scatter diagram of the data. *(3 marks)*
- (b) Find the equation of the regression line of y on x and plot it on your scatter diagram. *(6 marks)*
- (c) Give an interpretation, in the context of this question, of:
- (i) the gradient of the regression line;
 - (ii) the intercept of the regression line with the y -axis. *(3 marks)*
- (d) Hennie's brother Ludwig suggests to her that she could save time by cutting her lawn in spring and then waiting until the autumn before cutting it again. Give one statistical reason **and** one reason specific to this context why it would be unwise to use your regression equation to estimate the time it would take Hennie to cut her lawn 150 days since she last cut it. *(2 marks)*
- (e) Explain why it was appropriate in this question to calculate the regression equation of y on x rather than that of x on y . *(2 marks)*

