

Topic Test: Correlation

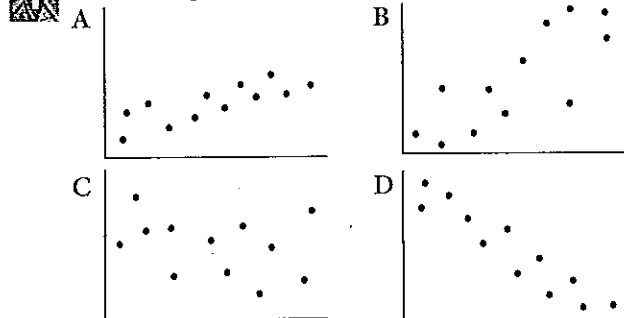
Remember: these are HSC-type questions.

Time allowed: 40 minutes **Total marks: 25**

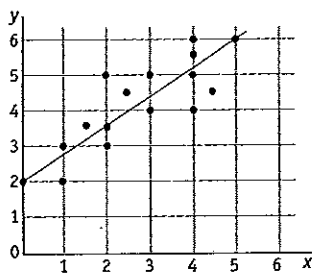
Part A (Suggested time: 15 minutes)

Choose the correct answer (A, B, C or D) for each question. One mark each

1 Which diagram shows a strong negative correlation?



2 A line of fit has been drawn on the scatterplot as shown.



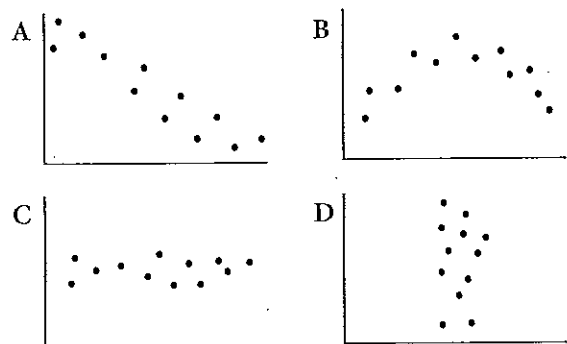
What is the equation of this line of fit?

- A $y = 4x + 2$ B $y = \frac{4}{5}x + 2$
 C $y = 5x + 2$ D $y = \frac{5}{4}x + 2$

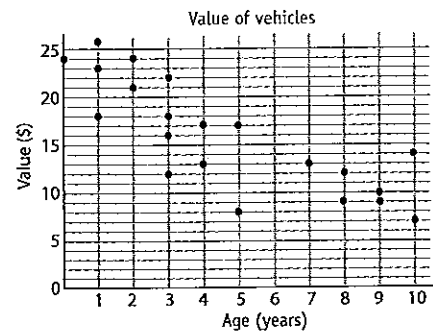
3 There is a moderate correlation between two sets of data. What might the correlation coefficient be?

- A 0 B 0.2
 C 0.5 D 0.8

4 In which scatterplot does the data NOT form a linear pattern?



5 A scatterplot has been drawn to show the age of vehicles traded-in at a car dealer's and the price paid for those vehicles. What is the difference in age of the two vehicles that received \$13 000 as a trade-in?



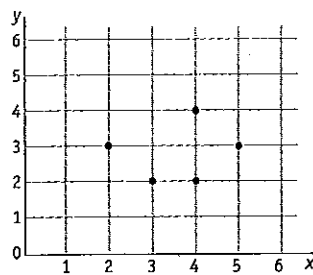
- A 2 years B 3 years
 C 4 years D 5 years

6 There is a very strong positive correlation between musical ability and mathematical ability among a group of students. Which statement is NOT correct?

- A A student from the group who is good at music is also likely to be good at maths.
 B A student from the group who has little mathematical ability is also likely to have little musical ability.
 C If a student from the group has average musical ability they must also have average mathematical ability.
 D The correlation coefficient between musical and mathematical ability for the group would be about 0.9.

7 For the points shown on the scatterplot, what is the median point?

- A (3, 4)
- B (4, 3)
- C (3.5, 3.5)
- D none of these



8 A study of particular plants has shown that the older the plant, the less it grows. How could the correlation between the age of the plant and the growth of the plant be described?

- A positive
- B negative
- C zero
- D there is not enough information to make such a description

9 The correlation coefficient between two sets of data is -0.3 . How could the correlation be described?

- A very strong negative
- B moderately strong negative
- C moderately weak negative
- D very weak negative

10 Which is NOT a step necessary in drawing a median regression line?

- A dividing the data into five data sets with vertical lines
- B finding the median of each data set
- C joining the medians of the first and last data sets
- D sliding the line $\frac{1}{3}$ of the way towards the middle median point

Part B

(Suggested time: 25 minutes)

Show all working.

15 marks

11 A survey was taken of the length of time (in months) employees were employed and the time (in minutes) each took to complete a particular task. The results are shown in the table.

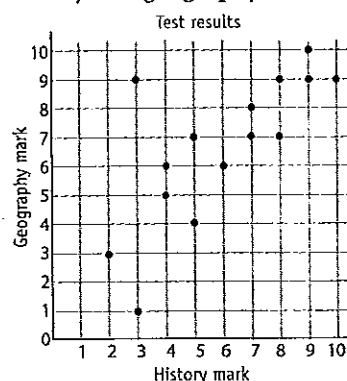
Time employed (months)	Time to complete task (min)
3	37
10	29
18	22
7	32
5	35
12	30
9	30
15	25
4	35
14	27

- a Draw a scatterplot and plot the information. (Show the time employed on the x -axis and the time to complete the task on the y -axis.) 2 marks
- b Draw in a line of fit. 1 mark
- c Find the equation of the line of fit. 2 marks
- d Use the equation of the line of fit to predict the length of time it would take someone who has been employed for 2 years to complete the task. 1 mark
- e If an employee takes 20 minutes to complete the task, for how long would you expect they have been employed? 1 mark
- f Briefly explain why the equation of the line of fit could not be used to predict the amount of time needed to complete the task by an employee of four years standing. 1 mark

12 The correlation coefficient between a high level of a particular protein and the incidence of a particular disease is -0.7 .

- a Briefly explain what this means. 1 mark
- b Angus reads this information and immediately begins to take tablets to increase his level of the protein so that he won't get the disease. Explain why Angus' reasoning is flawed. 1 mark

13 A scatterplot has been drawn to show the marks gained (out of 10) by a group of students in both history and geography.



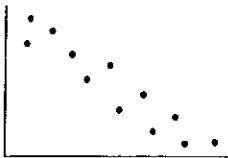
- a One of the students commented: 'There is a strong positive correlation between the two results. Obviously if you do well in one subject, you must do well in the other.' Give an example of why this student is wrong. 1 mark
- b Draw in the median regression line. 4 marks

Go to p 288 for **Quick Answers**
or to p 322 for **Worked Solutions**

Solutions

Topic Test p112

- 1 Diagram D shows a strong negative correlation. D



2
$$\text{gradient} = \frac{\text{vertical change in position}}{\text{horizontal change in position}}$$

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

Vertical intercept = 2

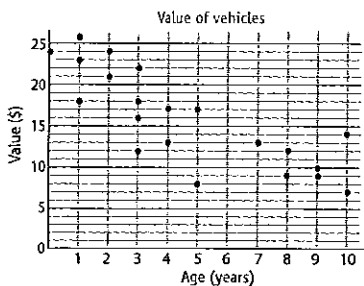
Equation is $y = \frac{4}{5}x + 2$ B

- 3 0.5 is moderate correlation C

- 4 B is not linear. B



- 5 The two vehicles that received \$13 000 as a trade-in are 4 years old and 7 years old. Difference = 3 years. B



- 6 C is wrong. A student from the group with average musical ability is also likely to have average mathematical ability, but it is not certain. C

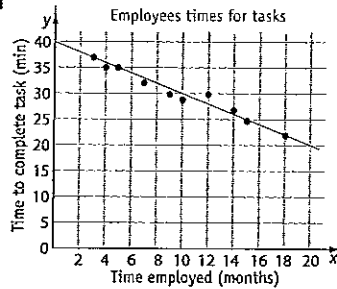
- 7 x-values: 2, 3, 4, 4, 5
Median is 4
y-values: 2, 2, 3, 3, 4
Median is 3
Median point is (4, 3). B

- 8 Negative correlation. B

- 9 A correlation coefficient of -0.3 shows moderately weak correlation. C

- 10 A is not a step.
[The data is divided into 3 data sets not 5.] A

- 11 a, b



c
$$\text{gradient} = \frac{\text{vertical change}}{\text{horizontal change}}$$

$$= \frac{-3}{3}$$

$$= -1$$

Vertical intercept = 40

Equation is $y = -x + 40$ ✓

[A different line of fit will have a different equation.]

- d When $x = 24$,

$$y = -24 + 40$$

$$= 16$$

It would take someone who has been employed for two years about 16 minutes to complete the task. ✓

- e When $y = 20$,

$$20 = -x + 40$$

$$-20 = -x$$

$$x = 20$$

If an employee takes 20 minutes to complete the task, they would probably have been employed for 20 months. ✓

- f When $x = 48$,

$$y = -48 + 40$$

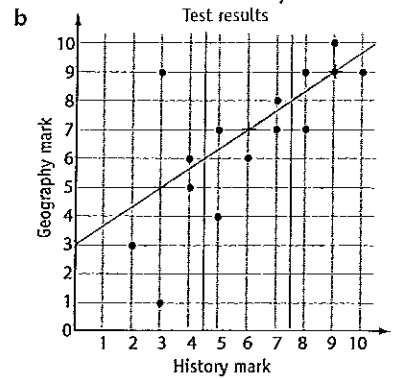
$$= -8$$

Because the amount of time cannot be negative, the equation makes no sense for an employee of four years standing. ✓

- 12 a There is a moderately strong negative correlation. In general, when there is a high level of the protein there is a low incidence of the disease. ✓

- b There is no evidence that high levels of the protein cause the incidence of the disease to be low. ✓

- 13 a One student who scored 9 in geography only scored 3 in history. Although in general, those who do well in one subject also do well in the other, there is no certainty. ✓



✓✓✓✓