1. A personnel manager wants to find out if a test carried out during an employee's interview and a skills assessment at the end of basic training is a guide to performance after working for the company for one year.

The table below shows the results of the interview test of 10 employees and their performance after one year.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Interview test, *x* %.	65	71	79	77	85	78	85	90	81	62
Performance after one year, *y* %.	65	74	82	64	87	78	61	65	79	69

[You may use $\sum x^2 = 60475, \sum y^2 = 53122, \sum xy = 56076$]

(a) Showing your working clearly, calculate the product moment correlation coefficient between the interview test and the performance after one year.

(5)

The product moment correlation coefficient between the skills assessment and the performance after one year is –0.156 to 3 significant figures.

(b) Use your answer to part (a) to comment on whether or not the interview test and skills assessment are a guide to the performance after one year. Give clear reasons for your answers.

(2)
(Total 7 marks)

2. A second hand car dealer has 10 cars for sale. She decides to investigate the link between the age of the cars, *x* years, and the mileage, *y* thousand miles. The data collected from the cars are shown in the table below.

Age, *x* (years)	2	2.5	3	4	4.5	4.5	5	3	6	6.5
Mileage, *y* (thousands)	22	34	33	37	40	45	49	30	58	58

[You may assume that $\sum x = 41, \sum y = 406, \sum x^2 = 188, \sum xy = 1818.5$]

(a) Find S_{xx} and S_{xy}.

(3)

(b) Find the equation of the least squares regression line in the form $y = a + bx$. Give the values of a and b to 2 decimal places.

(4)

(c) Give a practical interpretation of the slope b

(1)

(d) Using your answer to part (b), find the mileage predicted by the regression line for a 5 year old car.

(2)
(Total 10 marks)

3. A young family were looking for a new 3 bedroom semi-detached house. A local survey recorded the price x, in £1000, and the distance y, in miles, from the station of such houses. The following summary statistics were provided

$$S_{xx} = 113\,573,\ S_{yy} = 8\,657,\ S_{xy} = -808\,917$$

(a) Use these values to calculate the product moment correlation coefficient.

(2)

(b) Give an interpretation of your answer to part (a).

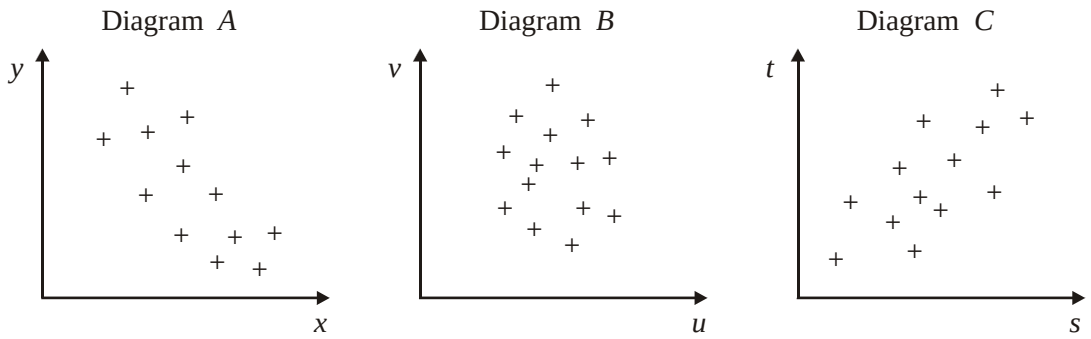
(1)

Another family asked for the distances to be measured in km rather than miles.

(c) State the value of the product moment correlation coefficient in this case.

(1)
(Total 4 marks)

4. The scatter diagrams below were drawn by a student.



The student calculated the value of the product moment correlation coefficient for each of the sets of data.

The values were

0.68 –0.79 0.08

Write down, with a reason, which value corresponds to which scatter diagram.

(Total 6 marks)

1. (a) $\Sigma x = 773$, $\Sigma y = 724$ B1, B1

$$r = \frac{10\times 56076 - 773\times 724}{\sqrt{(10\times 60475 - 773^2)(10\times 53122 - 724^2)}} \text{ o.e.}$$ M1A1ft

$r = 0.155357...$ A1 5

1st B1 for $\sum x$ and 2nd B1 for $\sum y$, should be seen or implied.

M1 for at least one correct attempt at one of S_{xx}, S_{yy} or S_{xy} and then using in the correct formula

1st A1ft for a fully correct expression. (ft their $\bar{x}$ and their $\bar{y}$) or 3 correct expressions for S_{xx}, S_{xy}, and S_{yy} but possibly incorrect values for these placed correctly in r.

2nd A1 for awrt 0.155

(b) Both weak correlation B1gB1h 2
Neither score is a good indication of future performance
Interview test is slightly better since correlation is positive

If $|r| > 0.5$ they can score B1g in (b) for saying that it (skills test) is not a good guide to performance but B0h since a second acceptable comment about both tests is not possible.

Give B1 for one correct line, B1B1 for any 2.
If the only comment is the test(s) are a good guide: scores B0B0
If the only comment is the tests are not good: scores B1B0 (second line)

The third line is for a comment that suggests that the interview test is OK but the skills test is not since one is positive and the other is negative.

Treat 1st B1 as B1g and 2nd as B1h

An answer of "no" alone scores B0B0

NB

$$S_{xx} = 60475 - \frac{(773)^2}{10} = 722.1, S_{yy} = 53122 - \frac{(724)^2}{10} = 704.4, S_{xy} = 56076 - \frac{773\times 724}{10} = 110.8$$

[7]

2. (a) $S_{xy} = 1818.5 - \frac{41\times 406}{10}$, = 153.9 (could be seen in (b)) AWRT 154 M1, A1

$S_{xx} = 188 - \frac{41^2}{10} = 19.9$ (could be seen in (b)) A1 3

M1 for correct attempt or expression for either

1st A1 for one correct

2nd A1 for both correct

(b) $b = \frac{153.9}{19.9}$, = 7.733668.... AWRT 7.73 M1, A1

$a = 40.6 - b \times 4.1 (= 8.89796....)$ M1

$y = 8.89 + 7.73x$ A1 4

Ignore the epen marks for part (b) they should be awarded as per this scheme

1st M1 for $\frac{\text{their } S_{xy}}{\text{their } S_{xx}}$

1st A1 for AWRT 7.73

2nd M1 for attempt at correct formula for a (minus required).
Ft their b.
Quoting a correct formula but making one slip in sub.
eg. $\bar{y}$ = 406 is OK

2nd A1 for correct equation with 2dp accuracy.
Accept a = 8.89, and b = 7.73 even if not written as final equation.

Correct answers only (from calc) score 4/4 if correct to 2dp or 3/4 if AWRT 2dp

(c) A typical car will travel 7700 miles every year B1ft 1

B1ft for their $b \times 1000$ to at least 2 sf.
Accept "7.7 thousand" but value is needed

(d) $x = 5$, $y = 8.89 + 7.73 \times 5 (= 47.5 - 47.6)$ M1

So mileage predicted is AWRT 48000 A1 2

M1 for substituting $x = 5$ into their final answer to (b).

A1 for AWRT 48000 (Accept "48 thousands")

Accept calculations for S_{xx} and S_{xy} in (a) or (b)

[10]

3. (a) $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{-808.917}{\sqrt{113573 \times 8.657}}$ M1

= –0.81579... A1 2

M1 for knowing formula and clear attempt to sub in correct values from question.
Root required for method.
Anything that rounds to –0.82 for A1.
Correct answer with no working award 2/2

(b) Houses are cheaper further away from the station or equivalent statement — B1 — 1

Context based on negative correlation only required.
Accept Houses are more expensive closer to the station or equivalent statement.
Require 'house prices' or 'station' and clear correct comparison.

(c) –0.816 — B1ft — 1

Accept anything that rounds to –0.82 or 'the same' or 'unchanged' or equivalent.
Award B1 if value quoted same as answer to (a).

[4]

4. Diagram A : y & x : r = – 0.79; As x increases, y decreases or most points lie in the 2nd and 4th quadrant. — B1;B1dep

Diagram B : v & u : r = 0.08; No real pattern. Several values of v for one value of u or points lie in all four quadrants, randomly scattered. — B1; B1dep

Diagram C: t & s : r = 0.68; As s increases, t increases or most points lie in the 1st and 3rd quadrants — B1;B1dep

[6]

1. Part (a) was answered very well and most candidates scored full marks here but responses to part (b) were mixed. Some thought that because both values were similar, but one positive and one negative, they "cancelled out" and others only commented on one of the tests or thought that the correlation coefficients were between the two tests. However a number of fully correct solutions were seen.

2. The first two parts of this question were answered very well. There were few problems encountered in parts (a) and (b) although $a = 8.91$ was a common error caused by using the rounded value of b not a more accurate version. Most candidates adhered to the instruction to give their answers to 2 decimal places. Problems started though when the candidates were asked to interpret the equation. Many candidates simply said that mileage increases with age and few who mentioned the 7.7 value remembered the thousands. A simple response such as "the annual mileage is 7700 miles" or "each year a car travels 7700 miles" was rarely seen.

In part (d) most could substitute $x = 5$ into their equation but once again the "thousand" was forgotten and the unlikely figure of 48 miles for a 5 year old car was all too common.

3. Most candidates had little trouble with part (a). In part (b) a sizeable minority correctly identified negative correlation, but failed to put it into the context of the question. In part (c) again a sizeable minority did not attempt this part or multiplied their answer for part (a) by some arbitrary factor.

4. Most candidates were able to match the given values of the product moment correlation coefficient with the correct diagram. However, relatively few were able to give acceptable reasons based upon correlation rather than regression considerations. Even those candidates who correctly referred to variables increasing and decreasing missed identifying u, v and s, t or found it difficult to describe diagram B. There was a tendency to write about the points being in the middle of the diagram without reference to their randomness or scattering.