1. The product moment correlation coefficient is denoted by r and Spearman's rank correlation coefficient is denoted by r_s.

(a) Sketch separate scatter diagrams, with five points on each diagram, to show

   (i) $r = 1$,

   (ii) $r_s = -1$ but $r > -1$.

(3)

Two judges rank seven collie dogs in a competition. The collie dogs are labelled A to G and the rankings are as follows

Rank	1	2	3	4	5	6	7
Judge 1	A	C	D	B	E	F	G
Judge 2	A	B	D	C	E	G	F

(b) (i)   Calculate Spearman's rank correlation coefficient for these data.

(6)

   (ii)   Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the judges are generally in agreement.

(5)
(Total 14 marks)

2. Over a period of time, researchers took 10 blood samples from one patient with a blood disease. For each sample, they measured the levels of serum magnesium, s mg/dl, in the blood and the corresponding level of the disease protein, d mg/dl. The results are shown in the table.

s	1.2	1.9	3.2	3.9	2.5	4.5	5.7	4.0	1.1	5.9
d	3.8	7.0	11.0	12.0	9.0	12.0	13.5	12.2	2.0	13.9

[Use $\sum s^2 = 141.51$, $\sum d^2 = 1081.74$ and $\sum sd = 386.32$]

(a) Draw a scatter diagram to represent these data.

(3)

(b)   State what is measured by the product moment correlation coefficient.

(1)

(c)   Calculate S_{ss}, S_{dd} and S_{sd}.

(3)

(d)   Calculate the value of the product moment correlation coefficient r between s and d.

(2)

(e)   Stating your hypotheses clearly, test, at the 1% significance level, whether or not the correlation coefficient is greater than zero.

(3)

(f)   With reference to your scatter diagram, comment on your result in part (e).

(1)
(Total 13 marks)
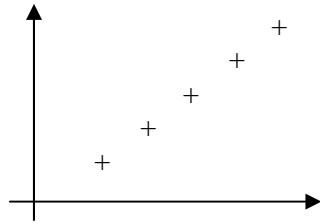
1. (a) (i)   (ii)
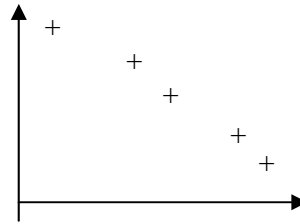
(i) B1
(ii) B1B1   3

(i)   1st B1 for 5 or more points on a straight line of positive gradient

(ii)   2nd B1 for 4 or more points satisfying $-1 < r < 0$
3rd B1 for 5 or more points of decreasing ranks not on
a straight line

(b) (i)   M1M1

	A	B	C	D	E	F	G
Rank (Judge 1)	1	4	2	3	5	6	7
Rank (Judge 2)	1	2	4	3	5	7	6
d^2	0	4	4	0	0	1	1

$\sum d^2 = 10$ M1A1

$r_s = 1 - \dfrac{6 \times 10}{7 \times (49-1)} = 1 - \dfrac{5}{28} = \dfrac{23}{28}$ or awrt **0.821**   M1A1   6

1st M1 for attempting to rank one of the judges (at least 2
correct rankings)

2nd M1 for ranking both (may be reversed) (at least 2
correct rankings)

3rd M1 for attempting d^2.

1st A1 for $\sum d^2 = 10$

4th M1 for correct use of the r_s formula

(ii) $H_0 : \rho = 0$ $H_1: \rho > 0$ (Allow ρ_S) ($H_1 : \rho \neq 0$ scores B0)   B1, B1
r_s 5% one tail critical value is **0.7143**   B1
Significant result or reject null hypothesis   M1
There is evidence of a (positive) correlation between the
judges <u>or</u> the judges agree   A1ft   5

3rd B1 for the correct critical value – depends upon their

        H_1: $\rho > 0$ needs 0.7143, $\rho \neq 0$, 0.7857
        The H_1 may be in words so B0B1 is possible.
        If no H_1 award for 0.7143 only.

    5th M1 for a correct statement relating their r_s and their cv
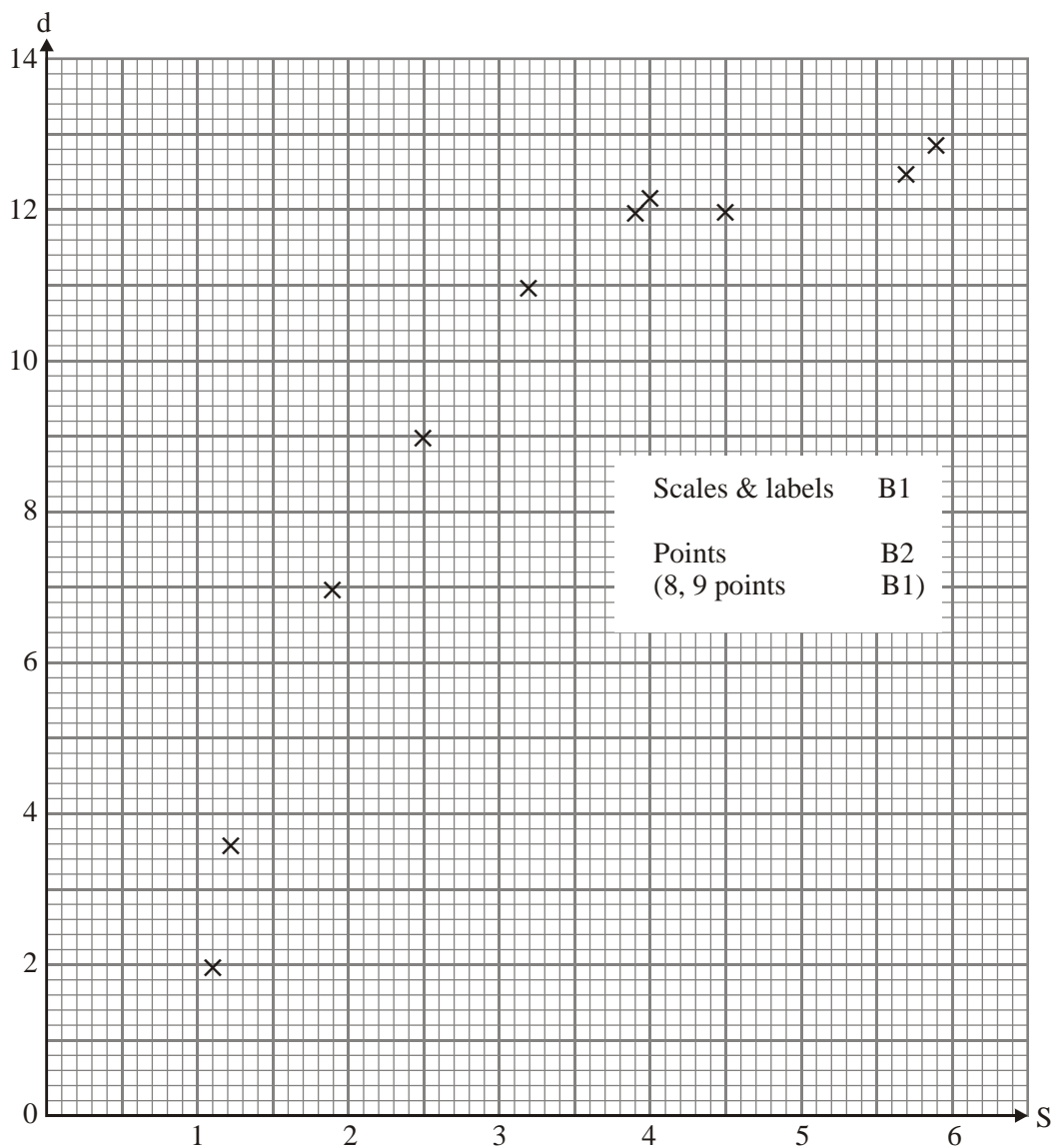   (may be implied by correct comment)

    3rd A1ft follow through their r_s and their cv.
          Comment in context. Must mention judges.

          Don't insist on "positive" and condone it if they are using $\rho \neq 0$.

[14]

2.   (a)   <u>NB</u> No graph paper $\Rightarrow$ $0/_3$ 3

Scales & labels   B1

Points   B2
(8, 9 points   B1)

    (b)   <u>Linear association</u> between *s* and *d*   B1   1

(c) $S_{ss} = 141.51 - \dfrac{33.9^2}{10} = \underline{26.589}$; $S_{dd} = \underline{152.444}$; $S_{sd} = \underline{59.524}$ B1; B1; B1   3

(d) $r = \dfrac{59.524}{\sqrt{152.444 \times 26.589}}$ M1

       $= \underline{0.93494\ldots}$ A1   2

            awrt 0.935

(e) $H_0: \rho = 0$; $H_1: \rho > 0$ B1
Critical value at 1% = 0.7155   B1
Reject H_0; levels of serum & disease are positively correlated   B1   3

(f)   Linear correlation significant but scatter diagram looks non-linear.   B1   1

[13]

1. The majority of candidates could draw a correct diagram for $r = 1$ but far fewer managed to do so for (a)(ii), a set of points lying on a line of negative gradient was a common error. In part (b) the ranking caused some to stumble and a value of 4 for $\sum d^2$ was fairly common but most could use the formula for r_s correctly. The hypotheses in part (c) were sometimes given in words or in terms of r_s rather than ρ and a number of candidates used a two-tailed test. The appropriate critical value was usually given and the conclusions were nearly always correct and in context.

2. The scatter diagram was usually well drawn but the obvious curve in the data was rarely commented on in the final part. Few candidates knew what was measured by the product moment correlation coefficient. The numerical parts of the question and the significance test were well answered.