1. A researcher claims that, at a river bend, the water gradually gets deeper as the distance from the inner bank increases. He measures the distance from the inner bank, *b* cm, and the depth of a river, *s* cm, at seven positions. The results are shown in the table below.

Position	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Distance from inner bank *b* cm	100	200	300	400	500	600	700
Depth *s* cm	60	75	85	76	110	120	104

(a) Calculate Spearman's rank correlation coefficient between *b* and *s*.

(6)

(b) Stating your hypotheses clearly, test whether or not the data provides support for the researcher's claim. Use a 1% level of significance.

(4)
(Total 10 marks)

2. A doctor is interested in the relationship between a person's Body Mass Index (BMI) and their level of fitness. She believes that a lower BMI leads to a greater level of fitness. She randomly selects 10 female 18 year-olds and calculates each individual's BMI. The females then run a race and the doctor records their finishing positions. The results are shown in the table.

Individual	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
BMI	17.4	21.4	18.9	24.4	19.4	20.1	22.6	18.4	25.8	28.1
Finishing position	3	5	1	9	6	4	10	2	7	8

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

(5)

(b) Stating your hypotheses clearly and using a one tailed test with a 5% level of significance, interpret your rank correlation coefficient.

(5)

(c) Give a reason to support the use of the rank correlation coefficient rather than the product moment correlation coefficient with these data.

(1)
(Total 11 marks)

3. 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)

4. The table below shows the price of an ice cream and the distance of the shop where it was purchased from a particular tourist attraction.

Shop	Distance from tourist attraction (m)	Price (£)
A	50	1.75
B	175	1.20
C	270	2.00
D	375	1.05
E	425	0.95
F	580	1.25
G	710	0.80
H	790	0.75
I	890	1.00
J	980	0.85

(a) Find, to 3 decimal places, the Spearman rank correlation coefficient between the distance of the shop from the tourist attraction and the price of an ice cream.

(5)

(b) Stating your hypotheses clearly and using a 5% one-tailed test, interpret your rank correlation coefficient.

(4)
(Total 9 marks)

5. The numbers of deaths from pneumoconiosis and lung cancer in a developing country are given in the table.

Age group (years)	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 and over
Deaths from pneumoconiosis (1000s)	12.5	5.9	18.5	19.4	31.2	31.0
Deaths from lung cancer (1000s)	3.7	9.0	10.2	19.0	13.0	18.0

The correlation between the number of deaths in the different age groups for each disease is to be investigated.

(a) Give **one** reason why Spearman's rank correlation coefficient should be used.

(1)

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

(6)

(c) Use a suitable test, at the 5% significance level, to interpret your result. State your hypotheses clearly.

(5)

(Total 12 marks)

6. 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)

7. A random sample of 8 students sat examinations in Geography and Statistics. The product moment correlation coefficient between their results was 0.572 and the Spearman rank correlation coefficient was 0.655.

(a) Test both of these values for positive correlation. Use a 5% level of significance.

(6)

(b) Comment on your results.

(2)
(Total 8 marks)

8. Two judges ranked 8 ice skaters in a competition according to the table below.

Skater / Judge	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
A	2	5	3	7	8	1	4	6
B	3	2	6	5	7	4	1	8

(a) Evaluate Spearman's rank correlation coefficient between the ranks of the two judges.

(6)

(b) Use a suitable test, at the 5% level of significance, to interpret this result.

(5)

(Total 11 marks)

1. (a)

Distance rank	1	2	3	4	5	6	7
Depth rank	1	2	4	3	6	7	5
$\lvert d \rvert$	0	0	1	1	1	1	2
d^2	0	0	1	1	1	1	4

M1

M1

$\Sigma d^2 = 8$ M1A1

$r_s = 1 - \dfrac{6 \times 8}{7 \times 48}$ M1

$= \dfrac{6}{7} = 0.\dot{8}57142\dot{}$ awrt **0.857** A1 6

Note

1st M1 for an attempt to rank the depths against the distances
2nd M1 for attempting *d* for their ranks. Must be using ranks.
3rd M1 for attempting $\sum d^2$ (must be using ranks)
1st A1 for sum of 8 (or 104 for reverse ranking)
4th M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.
2nd A1 for awrt (±) 0.857. Sign should correspond to ranking (so use of 104 should get –0.857)

(b) $H_0 : \rho = 0, H_1 : \rho > 0$ B1

Critical value at 1% level is 0.8929 B1

$r_s < 0.8929$ so not significant evidence to reject H_0, M1

The researcher's claim is not correct (at 1% level). A1ft 4
or insufficient evidence for researcher's claim
or there is insufficient evidence that water gets deeper further from inner bank.
or no (positive) correlation between depth of water and distance from inner bank

Note

1st B1 for both hypotheses in terms of ρ, H_1 must be one tail and compatible with their ranking
2nd B1 for cv of 0.8929 (accept ±)
M1 for a correct statement relating their r_s with their cv but cv must be such that |cv| < 1
A1ft for a correct contextualised comment. Must mention "researcher" and "claim" or "distance (from bank)" and

"depth (of water)"
Follow through their r_s and their cv (provided it is |cv| < 1)
Use of "association" is A0

[10]

2. (a)

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	
BMI	1	6	3	8	4	5	7	2	9	10	M1
or	10	5	8	3	7	6	4	9	2	1	
Finishing position	3	5	1	9	6	4	10	2	7	8	
d^2	4	1	4	1	4	1	9	0	4	4	

$\sum d^2 = 32$ (298) M1

$r_s = 1 - \frac{6 \times 32}{10 \times 99}$ M1 A1ft

= 0.80606… (–0.80606) accept $\pm\frac{133}{165}$ **awrt ± 0.806** A1 5

Note

1st M1 for attempt to rank BMI scores

2nd M1 for attempt at $\sum d^2$
(must be using ranks) No ranking can score 3rd

3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required. M1 only

1st A1ft for a correct expression. ft their $\sum d^2$ but only if all 3 Ms are scored

2nd A1 awrt ± 0.806 (but sign must be compatible with their $\sum d^2$)

(b) $H_0 : \rho = 0$, $H1 : \rho > 0$, B1 B1

Critical value is (±)0.5636 B1

(0.806 > 0.5636 therefore) in critical region/ reject H_0 M1

The lower the BMI the higher the position in the race./ support for doctors belief A1ft 5

Note

2nd B1 for $\rho > 0$ (or <0 but must be one tail and consistent with their ranking) NoH_1

3rd B1 for critical value that is compatible with theirH_1 . If one-tail must be ± 0.5636 if two-tail must be ± 0.6485 [Condone wrongsign] assume onetail for 3rd B1

M1 for a correct statement relating their r_s with their cv.

e.g. "reject H_0", "in critical region", "significant result"

May be implied by a correct comment

A1ft for correct comment in context. Must mention low/high BMI and race/fitness or doctor's belief. Comment should be one-tailed.

Allow positive correlation between… but NOT …positive relationship…

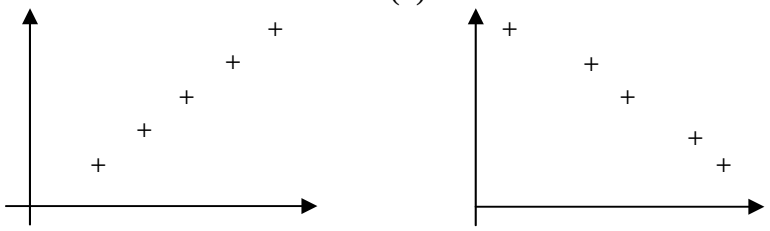
(c) The position is already ranked OR Position is not Normally distributed B1 1

Note

B1 for a correct and relevant comment either based on the fact that the data was originally partially ordered or on the underlying normal assumption "Quicker" or "easier" score B0

[11]

3. (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

$r_s = 1 - \frac{6 \times 10}{7 \times (49-1)} = 1 - \frac{5}{28} = \frac{23}{28}$ 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 or 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 A1ftfollow 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]

4. Rank:

Shop	Distance	Price	d	d^2		
A	1	9	8	64		
B	2	7	5	25		
C	3	10	7	49		
D	4	6	2	4		
E	5	4	1	1	*Ranking*	M1
F	6	8	2	4		
G	7	2	5	25		
H	8	1	7	49		
I	9	5	4	16		
J	10	3	7	49		

Reverse ranking or price $\sum d^2 = 44$ 286 $\sum d^2$ M1, A1

(a) $r_s = 1 - \frac{6 \times 286}{10(100-1)} = -0.73$ or $\frac{-11}{15}$ or -0.733awrt M1 A1

or 0.733awrt for $\sum d^2 = 44$ 5

(b) $H_0 : \rho = 0$ B1

$H_1 : \rho < 0$ ($H_1 : \rho > 0$ if reverse ranking) B1

cv = –0.5636 (–0.5636) B1

Reject H_0, evidence there is a significant negative correlation between the price of an ice cream and the distance from a tourist attraction. B1

(Ice cream gets cheaper further from the tourist attraction)
(– cv from correct table required) (position in context) 4

[9]

5. (a) The variables cannot be assumed to be normally distributed B1 1

(b)

	20–29	30–39	40–49	50–59	60–69	70+
Rank *x*	5	6	4	3	1	2
Rank *y*	6	5	4	1	3	2
d	1	1	0	2	2	0
d^2	1	1	0	4	4	0

M1 A1

dM1 (depends on ranking attempt)

$\sum d^2 = 10$ (follow through their rankings) A1ft

$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{60}{210} = 0.714$ ($\frac{5}{7}$ or awrt 0.714) M1 A1 6

(c) H_0: $\rho = 0$ B1

H_1: $\rho \neq 0$ (or $\rho > 0$) B1

$n = 6 \Rightarrow$ 5% critical value = 0.8857 (or 0.8286) B1ft

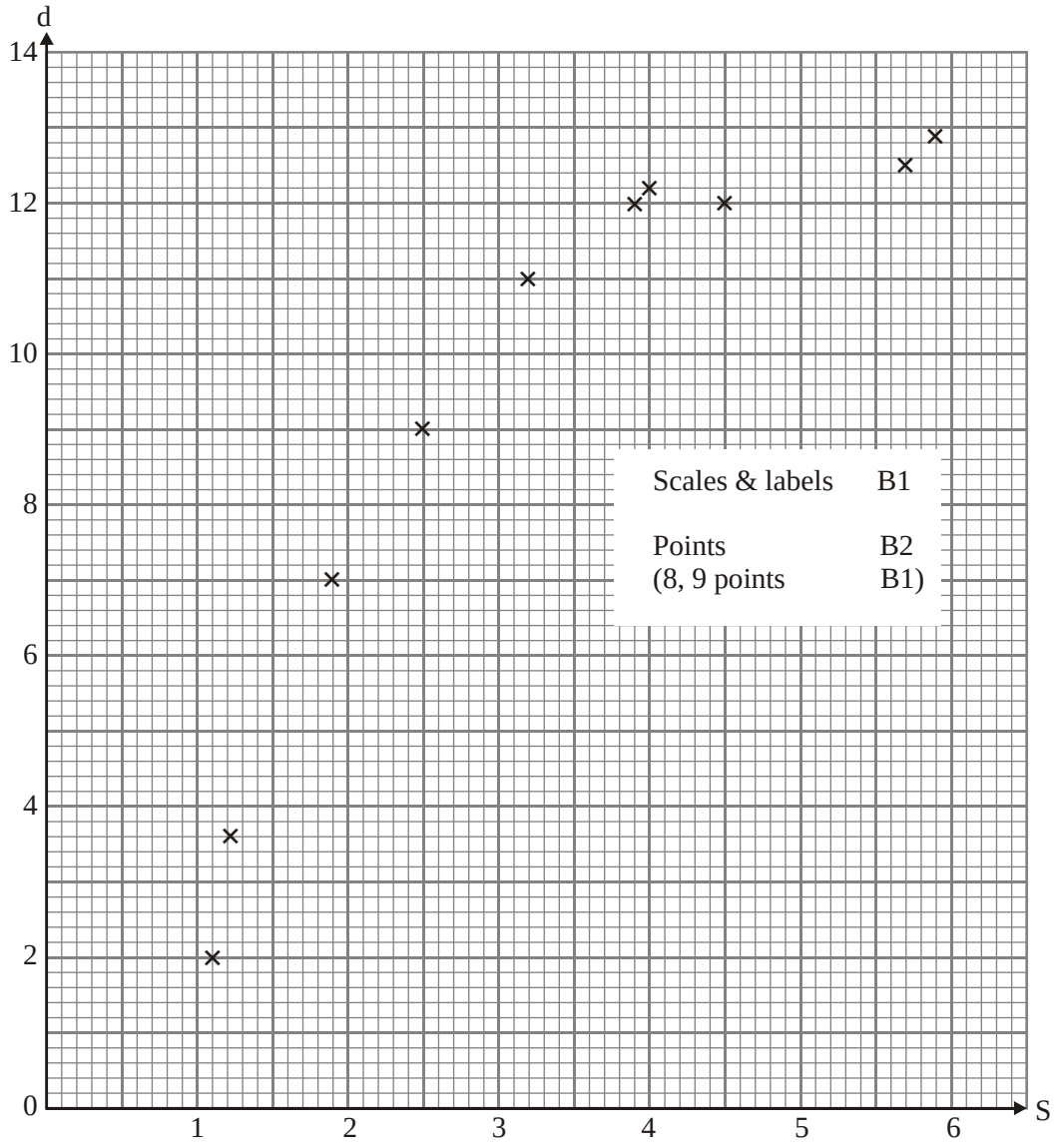
0.714 < 0.8857 M1

No evidence to reject H_0; A1 5

No evidence of correlation between deaths from pneumoconiosis and lung cancer.

[12]

6. (a) NB No graph paper $\Rightarrow {}^{0}/_{3}$ 3



(b) Linear association 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]

7. (a) $H_0 : \rho = 0$; $H_1: \rho > 0$ B1
both and ρ
5% CV – PMCC 0.6215 B1
0.572 < 0.6215 / not in critical region / not significant M1
No evidence of positive correlation A1
Spearman 0.6429 B1
Evidence of positive correlation B1 6

(b) No evidence to suggest that as Statistics marks increased Geography marks increased. B1
Context and not correlation
Evidence that students ranked highly in Statistics were also ranked highly in Geography B1 2
Ranked

[8]

8. (a)

A	2	5	3	7	8	1	4	6	
B	3	2	6	5	7	4	1	8	
$\lvert d \rvert$	1	3	3	2	1	3	3	2	
d^2	1	9	9	4	1	9	9	4	46

d M1
$\sum d^2$ M1 A1

$r_s = 1 - \frac{6 \times 46}{8 \times 63}$ M1 A1∫
$r_s = 0.452$ A1 6

(b)	$H_0 = \rho = 0$, $H_1 : \rho \neq 0$ ($\rho > 0$)		B1 B1	
	critical values are $\pm$ 0.7381 (0.6429)	0.7381 (0.6429)	B1	
	0.452<0.7381 (0.452<0.6429) or not sig			
	or Insufficient evidence to reject H_0		M1	
	No agreement between the two judges.	Context	A1$\int$	5

[11]

1. Most candidates secured full marks in part (a) with only a small number making arithmetic errors and a tiny minority failing to use ranks or using an incorrect formula. The hypothesis test in part (b) was often answered very well too. Some did not use ρ for the hypotheses and some failed to give a full conclusion in context but most had the correct critical value and gave a correct statement about H_0.

2. Part (a) was a fairly standard application of Spearman's rank correlation and it was answered very well with most ranking BMI from low to high. In part (b) there were fewer candidates losing marks for failing to give their hypotheses in terms of ρ and there were many good answers here too although sometimes it was difficult to interpret their conclusion: "there is positive correlation between BMI and finishing position" may be true but is not as clear as saying that "there is evidence to support the doctor's belief" or "there is evidence that a low BMI leads to a greater level of fitness". Many missed the point in part (c) and simply mentioned that Spearman's rank correlation was "easier" or that there were no tied ranks.

3. 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.

4. Part (a) was very well answered by the great majority of candidates; part (b) less so but still a very large number of fully correct answers were seen, with the final conclusion well stated in context. A typical error was to conclude that the correlation was positive without any further interpretation.

5. Part (a) was not answered well. Some candidates mentioned that the data was unlikely to be joint normally distributed but the usual offerings simply mentioned ease of use or that the question was concerned with ranks. The remainder of the question was answered well. Only a few candidates failed to use ranks in their Spearman's formula and most stated their hypotheses in part (b) in terms of ρ.

6. 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.

7. Part (a) was generally answered well. The most common errors were not putting their hypotheses in terms of rho (ρ) and not including the word positive in their conclusions. Few candidates gained any marks in part (b). Most candidates wrote a great deal about the relevant values of the two correlation coefficients but at no time related their answer to the context of the question.

8. This question also allowed candidates to score highly; indeed some otherwise poor scripts were redeemed by good marks here. However a mark was almost always lost for not giving *r* to 3 significant figures.