Correlation 2B

1 a The data in the scatter graph clearly follows a linear trend, and the product moment correlation coefficient is more suitable for linear correlation.

b Spearman's rank correlation coefficient is easier to calculate.

2 The data is non-linear.

3 The number of attempts taken to score a free throw is not normally distributed (it is geometric), so the researcher should use Spearman's rank correlation coefficient.

4 a The data are ranked. There are no tied ranks. The table shows d and d^2 for each pair of ranks:

r_x	r_y	d	d^2
1	3	–2	4
2	2	0	0
3	1	2	4
4	5	–1	1
5	4	1	1
6	6	0	0
		Total	10

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 10}{6(6^2-1)} = 1 - 0.28571\ldots = 0.714 \text{ (3 s.f.)}$$

There is limited evidence of positive correlation between the pairs of ranks. This value is between weak and strong positive correlation.

b There are no tied ranks. The table shows d and d^2 for each pair of ranks:

r_x	r_y	d	d^2
1	2	–1	1
2	1	1	1
3	4	–1	1
4	3	1	1
5	5	0	0
6	8	–2	4
7	7	0	0
8	9	–1	1
9	6	3	9
10	10	0	0
		Total	18

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 18}{10(10^2-1)} = 1 - 0.10909\ldots = 0.891 \text{ (3 s.f.)}$$

There is fairly strong positive correlation between the pairs of ranks.

4 c There are no tied ranks. The table shows d and d^2 for each pair of ranks:

r_x	r_y	d	d^2
5	5	0	0
2	6	–4	16
6	3	3	9
1	8	–7	49
4	7	–3	9
3	4	–1	1
7	2	5	25
8	1	7	49
		Total	158

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 158}{8(8^2-1)} = 1 - 1.88095\ldots = -0.881 \text{ (3 s.f.)}$$

There is fairly strong negative correlation between the pairs of ranks.

5 a The data is positively correlated and ranked, therefore $r_s = 1$.

b The data is clearly correlated, and has a negative trend, therefore this case corresponds to the only negative value, $r_s = -1$.

c The data is strongly correlated, and ranked with only one outlier, so r_s is close to 1, i.e. $r_s = 0.9$.

d This data set is more scattered than the others and there is no clear trend, so $r_s = 0.5$.

6 a The table shows the ranking for goals scored (r_g) (the league position is the ranking in the league r_l) and then d and d^2 for each pair of ranks:

Goals	r_g	r_l	d	d^2
49	1	1	0	
44	2	2	0	0
43	3	3	0	0
36	6	4	2	4
40	4	5	–1	1
39	5	6	–1	1
29	9	7	2	4
21	12	8	4	16
28	10	9	1	1
30	8	10	–2	4
33	7	11	4	16
26	11	12	1	1
			Total	48

6 b There are no tied ranks, and $d^2 = 48$, so:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times48}{12(12^2-1)} = 1 - 0.16783\ldots = 0.832 \text{ (3 s.f.)}$$

This shows fairly strong positive correlation between the pairs of ranks. This suggests that the more goals a team scores, the higher its league position is likely to be.

7 There are no tied ranks. The table shows d and d^2 for each pair of ranks:

r_Q	r_T	d	d^2
1	1	0	0
2	2	0	0
3	5	–2	4
4	6	–2	4
5	4	1	1
6	3	3	9
7	8	–1	1
8	7	1	1
		Total	20

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times20}{8(8^2-1)} = 1 - 0.23809\ldots = 0.762 \text{ (3 s.f.)}$$

There is fairly strong positive correlation between the pairs of ranks. This suggests the trainee vet is rating the cows for quality in a similar way to the qualified vet.

8 a The marks are discrete values drawn from a specified scale in order to rank the competitors.

b The table shows the ranks of each judge and d and d^2 for each pair of ranks:

J_1	J_2	r_{J1}	r_{J2}	d	d^2
7.8	8.1	4	4	0	0
6.6	6.8	9	8	1	1
7.3	8.2	7	3	4	16
7.4	7.5	6	7	–1	1
8.4	8.0	3	5	–2	4
6.5	6.7	10	9	1	1
8.9	8.5	1	1	0	0
8.5	8.3	2	2	0	0
6.7	6.6	8	10	–2	4
7.7	7.8	5	6	–1	1
				Total	28

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times28}{10(10^2-1)} = 1 - 0.16969\ldots = 0.830 \text{ (3 s.f.)}$$

There is a strong positive correlation between the marks, hence the two judges agree well.

8 c Now there is a tied rank for the value 7.7 for competitors *A* and *J*, and we should give each of the equal values a rank equal to the average of their ranks, which would be 4.5.

9 a The emphasis here is on ranks/marks so the data sets are unlikely to be from a bivariate normal distribution.

b The table shows the ranks of each judge (using averages where scores are tied in rank) and d and d^2 for each pair of ranks:

J_1	J_2	r_{J1}	r_{J2}	d	d^2
4.5	5.2	1	5	–4	16
5.1	4.8	2	1	1	1
5.2	4.9	3.5	2	1.5	2.25
5.2	5.1	3.5	4	–0.5	0.25
5.4	5.0	5	3	2	4
5.7	5.3	6	6	0	0
5.8	5.4	7	7	0	0
				Total	23.5

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 23.5}{7(7^2-1)} = 1 - 0.41964\ldots = 0.580 \text{ (3 s.f.)}$$

c Both show positive correlation, but the judges agree more on the second dive.

10 a The table shows the ranks of each taster (using averages where scores are tied in rank) and d and d^2 for each pair of ranks:

T_1	T_2	r_{T1}	r_{T2}	d	d^2
7	8	4.5	4	0.5	0.25
8	9	3	3	0	0
6	5	6	6	0	0
7	7	4.5	5	–0.5	0.25
9	10	2	1.5	0.5	0.25
10	10	1	1.5	–0.5	0.25
				Total	1

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 1}{6(6^2-1)} = 1 - 0.02857\ldots = 0.971 \text{ (3 s.f.)}$$

b i The rank doesn't change, so the coefficient also doesn't change.

ii If both tasters scored the additional tea the same, then $d = 0$ and Σd^2 doesn't change. However, now $n = 7$ so overall r_s increases.

c As there are many tied ranks, Spearman's rank correlation coefficient cannot be used. Instead, calculate the product moment correlation coefficient directly from the ranked data using the mean of the tied ranks.