Correlation 2C

1 a $S_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} == 131 - \frac{29^2}{7} = 10.8571\ldots$

$$S_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 140 - \frac{28^2}{7} = 28$$

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 99 - \frac{28 \times 29}{7} = -17$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-17}{\sqrt{10.8571\ldots \times 28}} = -0.975 \text{ (3 s.f.)}$$

b Assume that both sets of data are normally distributed.

$H_0: \rho = 0,\ H_1: \rho \neq 0$

Sample size = 7

Significance level in each tail = 0.005

From the table on page 216 of the textbook, critical values for r for a 0.005 significance level with a sample size of 7 are $r = \pm 0.8745$, so the critical region is $r < -0.8745$ and $r > 0.8745$.

The value found in part **a** is $-0.975 < -0.8343$. It lies within the critical region, so reject H_0. There is evidence at the 1% level of significance that the data is correlated.

2 a $r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right)}}$

$$= \frac{\left(20\,704 - \frac{168 \times 1275}{11}\right)}{\sqrt{\left(2585 - \frac{168^2}{11}\right)\left(320\,019 - \frac{1275^2}{11}\right)}} = 0.677 \text{ (3 s.f.)}$$

b $H_0: \rho = 0,\ H_1: \rho > 0$

Sample size = 11

Significance level = 0.05

From the table, the critical value for r for a 0.05 significance level with a sample size of 7 is $r = 0.5214$, so the critical region is $r > 0.5214$.

As $0.677 > 0.5214$, r lies within the critical region, so reject H_0. There is evidence at the 5% level of significance that the data is correlated.

There is evidence of positive correlation between the age and height of members of the athletics club – the older the player, the taller they tend to be.

Assumption: both the ages and the heights of the players are normally distributed.

3 a $H_0: \rho = 0,\ H_1: \rho \neq 0$

Sample size = 30

Significance level in each tail = 0.025

From the table, critical values for Spearman's rank correlation coefficient r_s for a 0.025 significance level with a sample size of 30 are $r_s = \pm 0.3624$.

So the critical region is $r_s < -0.3624$ and $r_s > 0.3624$.

b If $r_s = 0.5321$, the coefficient falls in the critical region. So reject the null hypothesis. There is evidence to suggest that engine size and fuel consumption are related.

4 a $H_0: \rho = 0,\ H_1: \rho > 0$

Sample size = 8

Significance level = 0.01

From the table, the critical value for r for a 0.01 significance level with a sample size of 8 is $r = 0.7887$, so the critical region is $r > 0.7887$.

As $0.774 < 0.7887$, accept H_0. There is insufficient evidence of positive correlation between technical ability and artistic performance at the 1% significance level.

b The table shows the ranks for technical ability and artistic performance (there are no tied ranks) and d and d^2 for each pair of ranks:

Gymnast	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
Technical ability	8.5	8.6	9.5	7.5	6.8	9.1	9.4	9.2
Artistic performance	6.2	7.5	8.2	6.7	6.0	7.2	8.0	9.1
r_T	6	5	1	7	8	4	2	3
r_A	7	4	2	6	8	5	3	1
d	–1	1	–1	1	0	–1	–1	2
d^2	1	1	1	1	0	1	1	4

$$\sum d^2 = 10$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 10}{8(8^2-1)} = 0.881 \text{ (3 s.f.)}$$

c The scores are used to rank the gymnasts and are unlikely to be normally distributed.

d $H_0: \rho = 0,\ H_1: \rho > 0$

Sample size = 8

Significance level = 0.01

The critical value for r_s for a 0.01 significance level with a sample size of 8 is $r_s = 0.8333$.

As $0.881 > 0.8333$, r_s lies within the critical region, so reject H_0. There is sufficient evidence at the 1% significance level that there is a positive correlation between technical ability and artistic performance.

5 a The data is given in the form of ranks.

b The table shows d and d^2 for each pair of ranks (there are no tied ranks):

Skater	i	ii	iii	iv	v	vi	vii	viii
r_A	2	5	3	7	8	1	4	6
r_B	3	2	6	5	7	4	1	8
d	–1	3	–3	2	1	–3	3	–2
d^2	1	9	9	4	1	9	9	4

$$\sum d^2 = 46$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 46}{8(8^2-1)} = 0.452 \text{ (3 s.f.)}$$

c $H_0: \rho = 0,\ H_1: \rho > 0$

Sample size = 8

Significance level = 0.05

The critical value for r_s for a 0.05 significance level with a sample size of 8 is $r_s = 0.6429$.

As $0.452 < 0.6429$, there is no reason to reject H_0. There is insufficient evidence of a positive association between the rankings of the judges.

6 The table shows the respective ranks of each team for goals scored and goals conceded (there are no tied ranks) and d and d^2 for each pair of ranks:

Team	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Goals for	39	40	28	27	26	30	42
Goals against	22	28	27	42	24	38	23
r_F	3	2	5	6	7	4	1
r_A	7	3	4	1	5	2	6
d	–4	–1	1	5	2	2	–5
d^2	16	1	1	25	4	4	25

$$\sum d^2 = 76$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 76}{7(7^2-1)} = -0.357 \text{ (3 s.f.)}$$

Testing for negative association between ‘goals for’ and ‘goals against’, so the hypotheses are:

$H_0: \rho = 0,\ H_1: \rho < 0$

Sample size = 7

Significance level = 0.01

The critical value for r_s for a 0.01 significance level with a sample size of 7 is $r_s = -0.8929$.

As $-0.357 > -0.8929$, there is no reason to reject H_0. There is insufficient evidence to show that a team that scores a lot of goals concedes very few goals.

7 a The table shows the respective ranks for takings and profits (there are no tied ranks) and d and d^2 for each pair of ranks:

Shop	**1**	**2**	**3**	**4**	**5**	**6**
Takings	400	6200	3600	5100	5000	3800
Profits	400	1100	450	750	800	500
r_T	6	1	5	2	3	4
r_P	6	1	5	3	2	4
d	0	0	0	–1	1	0
d^2	0	0	0	1	1	0

$$\sum d^2 = 2$$

$$r_s = 1-\frac{6\sum d^2}{n(n^2-1)} = 1-\frac{6\times 2}{6(6^2-1)} = 0.943 \text{ (3 s.f.)}$$

b $H_0: \rho = 0,\ H_1: \rho > 0$

Sample size = 6

Significance level = 0.05

The critical value for r_s for a 0.05 significance level with a sample size of 6 is $r_s = 0.8286$.

As $0.943 > 0.8286$, r_s lies within the critical region, so reject H_0. There is sufficient evidence at the 5% significance level that profits and takings are positively correlated.

8 a The table shows d and d^2 for each pair of ranks:

r_{maths}	1	2	3	4	5	6	7	8	9	10	11	12
r_{music}	6	4	2	3	1	7	5	9	10	8	11	12
d	–5	–2	1	1	4	–1	2	–1	–1	2	0	0
d^2	25	4	1	1	16	1	4	1	1	4	0	0

$$\sum d^2 = 58$$

$$r_s = 1-\frac{6\sum d^2}{n(n^2-1)} = 1-\frac{6\times 58}{12(12^2-1)} = 0.797 \text{ (3 s.f.)}$$

b $H_0: \rho = 0,\ H_1: \rho \neq 0$

Sample size = 12

Significance level in each tail = 0.025

From the table, critical values for Spearman's rank correlation coefficient r_s for a 0.025 significance level with a sample size of 12 are $r_s = \pm 0.5874$.

As $0.797 > 0.5874$, r_s lies within the critical region, reject H_0. There is sufficient evidence at the 5% significance level that there is a correlation between the results in mathematics and music. It seems that students that do well in mathematics also do well in music.

9 The table shows the respective ranks given by the child and the correct ordering and d and d^2 for each pair of ranks.

Rank, given	1	3	8	6	2	4	7	5	10	9
Rank, correct	1	2	3	4	5	6	7	8	9	10
d	0	1	5	2	–3	–2	0	–3	1	–1
d^2	0	1	25	4	9	4	0	9	1	1

$$\sum d^2 = 54$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 54}{10(10^2-1)} = 0.673 \text{ (3 s.f.)}$$

$\mathrm{H}_0: \rho = 0,\ \mathrm{H}_1: \rho > 0$

Sample size = 10

Significance level = 0.05

The critical value for r_s for a 0.05 significance level with a sample size of 10 is $r_s = 0.5636$.

As $0.673 > 0.5636$, r_s lies within the critical region, so reject H_0. There is sufficient evidence at the 5% significance level that there is a positive association between the child's order and the correct ordering.

10 Use the Spearman's rank correlation coefficient as the data is given in the form of ranks.
The table shows d and d^2 for each pair of ranks:

Year	**1**	**2**	**3**	**4**	**5**	**6**
Crop	62	73	52	77	63	61
rank, crop	4	2	6	1	3	5
rank, wetness	5	4	1	6	3	2
d	–1	–2	5	–5	0	3
d^2	1	4	25	25	0	9

$$\sum d^2 = 64$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 64}{6(6^2-1)} = -0.829 \text{ (3 s.f.)}$$

$\mathrm{H}_0: \rho = 0,\ \mathrm{H}_1: \rho \neq 0$

Sample size = 6

Significance level in each tail = 0.025

From the table, critical values for Spearman's rank correlation coefficient r_s for a 0.025 significance level with a sample size of 6 are $r_s = \pm 0.8857$.

As $-0.829 > -0.8857$, there is no reason to reject H_0. There is insufficient evidence of a correlation between crop yield and wetness.

11 a The researcher is likely to use the product moment correlation coefficient since the data is continuous and both height and mass are likely to be normally distributed.

b As the alternative hypothesis was accepted, $r >$ critical value for a sample size of 14. Using the table of critical values for correlation coefficients, for a significance level of 1%, the critical value is 0.6120 and 0.546 < 0.6129. For a significance level of 2.5%, the critical value is 0.5324 and 0.546 > 0.5324. So the smallest possible significance level is 2.5%.

c If the test is carried out at the 5% level of significance, for a sample of 10 the critical value is 0.5494 and 0.546 < 0.5494. For a sample of 11 the critical value is 0.5214 and 0.546 > 0.5214. So the smallest possible sample is 11.

