Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation

Exercise A, Question 1

Question:

For each of the data sets of ranks given below, calculate the Spearman's rank correlation coefficient and interpret the result.

a

r_x	1	2	3	4	5	6
r_y	3	2	1	5	4	6

b

r_x	1	2	3	4	5	6	7	8	9	10
r_y	2	1	4	3	5	8	7	9	6	10

c

r_x	5	2	6	1	4	3	7	8
r_y	5	6	3	8	7	4	2	1

Solution:

a

r_x	1	2	3	4	5	6
r_y	3	2	1	5	4	6
d	−2	0	2	−1	1	0
d^2	4	0	4	1	1	0

$\sum d^2 = 10$

$$r_s = 1 - \frac{6 \times 10}{6(6^2 - 1)}$$

$r_s = 0.714$

i.e. positive correlation between the pairs of ranks.

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

b

r_x	1	2	3	4	5	6	7	8	9	10
r_y	2	1	4	3	5	8	7	9	6	10
d	−1	1	−1	1	0	−2	0	−1	3	0
d^2	1	1	1	1	0	4	0	1	9	0

$\sum d^2 = 18$

$$r_s = 1 - \frac{6 \times 18}{10(10^2 - 1)}$$

$r_s = 0.891$

i.e. fairly strong positive correlation between the pairs of ranks of x and y.

c

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

$\sum d^2 = 158$

$$r_s = 1 - \frac{6 \times 158}{8(8^2 - 1)}$$

$r_s = 0.881$

i.e. fairly strong negative correlation between the pairs of ranks of x and y.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise A, Question 2

Question:

The number of goals scored by football teams and their positions in the league were recorded as follows for the top 12 teams.

Team	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Goals	49	44	43	36	40	39	29	21	28	30	33	26
League position	1	2	3	4	5	6	7	8	9	10	11	12

a Find Σd^2.
b Calculate Spearman's rank correlation coefficient for these data. What conclusions can be drawn from this result?

Solution:

Goals	49	44	43	36	40	39	29	21	28	30	33	26
League position	1	2	3	4	5	6	7	8	9	10	11	12
r_G	1	2	3	6	4	5	9	12	10	8	7	11
r_L	1	2	3	4	5	6	7	8	9	10	11	12
d	0	0	0	2	−1	−1	2	4	1	−2	−4	−1
d^2	0	0	0	4	1	1	4	16	1	4	16	1

The League position is their rank in the League (r_L).

$$\sum d^2 = 48$$

$$r_s = 1 - \frac{6 \times 48}{12(12^2 - 1)}$$

$$r_s = 0.832$$

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

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation

Exercise A, Question 3

Question:

A sample of a class's statistics projects was taken, and the projects were assessed by two teachers independently. Each teacher decided their rank order with the following results.

Project	A	B	C	D	E	F	G	H
Teacher A	5	8	1	6	2	7	3	4
Teacher B	7	4	3	1	6	8	2	5

a Find Σd^2.

b Calculate the rank correlation coefficient and state any conclusions you draw from it.

Solution:

a

r_A	5	8	1	6	2	7	3	4
r_B	7	4	3	1	6	8	2	5
d	−2	4	−2	5	−4	−1	1	−1
d^2	4	16	4	25	16	1	1	1

$$\sum d^2 = 68$$

b $r_s = 1 - \dfrac{6 \times 68}{8(8^2 - 1)}$

$r_s = 0.190$

There is virtually no correlation between the pairs of ranks awarded by the two teachers. They appear to be judging the projects using different criteria.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise A, Question 4

Question:

A veterinary surgeon and a trainee veterinary surgeon both rank a small herd of cows for quality. Their rankings are shown below.

Cow	A	D	F	E	B	C	H	J
Qualified vet	1	2	3	4	5	6	7	8
Trainee vet	1	2	5	6	4	3	8	7

Find the rank correlation coefficient for these data, and comment on the experience of the trainee vet.

Solution:

r_Q	1	2	3	4	5	6	7	8
r_T	1	2	5	6	4	3	8	7
d	0	0	−2	−2	1	3	−1	1
d^2	0	0	4	4	1	9	1	1

$$\sum d^2 = 20$$

$$r_s = 1 - \frac{6 \times 20}{8(8^2 - 1)}$$

$$r_s = 0.762$$

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.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation

Exercise A, Question 5

Question:

Two adjudicators at an ice dance skating competition award marks as follows.

Competitor	A	B	C	D	E	F	G	H	I	J
Judge 1	7.8	6.6	7.3	7.4	8.4	6.5	8.9	8.5	6.7	7.7
Judge 2	8.1	6.8	8.2	7.5	8.0	6.7	8.5	8.3	6.6	7.8

a Explain why you would use Spearman's rank correlation coefficient in this case.

b Calculate the rank correlation coefficient r_s, and comment on how well the judges agree.

Solution:

a The marks are discrete values within a very restricted scale.
They are also subjective judgements, not measurements.

b

J1	7.8	6.6	7.3	7.4	8.4	6.5	8.9	8.5	6.7	7.7
J2	8.1	6.8	8.2	7.5	8.0	6.7	8.5	8.3	6.6	7.8
r_1	4	9	7	6	3	10	1	2	8	5
r_2	4	8	3	7	5	9	1	2	10	6
d	0	1	4	−1	−2	1	0	0	−2	−1
d^2	0	1	16	1	4	1	0	0	4	1

$$\sum d^2 = 28$$

$$r_s = 1 - \frac{6 \times 28}{10(10^2 - 1)}$$

$$r_s = 0.830$$

This shows a fairly strong positive correlation between the pairs of ranks of the marks awarded by the two judges so it appears they are judging the ice dances using similar criteria and with similar standards.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation

Exercise A, Question 6

Question:

a A teacher believes that he can predict the positions in which his students will finish in an A-Level examination. When the results were out he wished to compare his predictions with the actual results. Which correlation test should he use and why?

b The table shows predicted and actual orders.

Student	A	B	C	D	E	F	G	H	I	J
Predicted, p	2	4	1	3	8	6	9	5	10	7
Actual, a	3	4	2	8	1	6	7	9	10	5

Calculate Spearman's rank correlation coefficient r_s between a and p. Comment on the result.

Solution:

a The teacher should use Spearman's rank correlation coefficient because the data being used are ranks (i.e. he is concerned with order).

b

p	2	4	1	3	8	6	9	5	10	7
a	3	4	2	8	1	6	7	9	10	5
d	−1	0	−1	−5	7	0	2	−4	0	2
d^2	1	0	1	25	49	0	4	16	0	4

$$\sum d^2 = 100$$

$$r_s = 1 - \frac{6 \times 100}{10(10^2 - 1)}$$

$$r_s = 0.394$$

This shows weak positive correlation between the orders. The teacher has very limited ability to predict the positions in which the students will finish. It doesn't appear to be very reliable.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise A, Question 7

Question:

A doctor assessed the lung damage suffered by a number of his patients who smoked, and asked each one 'For how many years have you smoked?' with the following results.

Patient	A	B	C	D	E	F	G
Number of years smoked	15	22	25	28	30	31	42
Lung damage grade	30	50	55	35	40	42	58

Calculate Spearman's rank correlation coefficient r_s and comment on the result. Give your value of Σd^2.

Solution:

Years smoked (y)	15	22	25	28	30	31	42
Lung damage grade (g)	30	50	55	35	40	42	58
r_y	1	2	3	4	5	6	7
r_g	1	5	6	2	3	4	7
d	0	−3	−3	2	2	2	0
d^2	0	9	9	4	4	4	0

$$\sum d^2 = 30$$

$$r_s = 1 - \frac{6 \times 30}{7(7^2 - 1)}$$

$$r_s = 0.464$$

This shows weak positive correlation between the pairs of ranks. There is some association between the lung damage grade and the number of years a person has smoked for – it is likely the lung damage grade is also likely to depend on other factors since the correlation is not very strong.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise B, Question 1

Question:

A product-moment correlation coefficient of 0.3275 was obtained from a sample of 40 pairs of values. Test whether or not this value shows evidence of correlation.

a at the 0.05 level (use a two-tailed test),

b at the 0.02 level (use a two-tailed test).

Solution:

a $\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{array}\right\}$ 2-tail $\alpha = 0.05$

test statistic $= 0.3275$

critical values $= \pm 0.3120$

t.s. > c.v. since $0.3275 > 0.3120$

reject H_0

Conclude there is evidence of correlation at the 5% used of significance.

b $\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{array}\right\}$ 2-tail $\alpha = 0.02$

test statistic $= 0.3275$

critical values $= \pm 0.3665$

t.s. < c.v. since $0.3275 < 0.3665$

accept H_0

Conclude no evidence of correlation at the 2% level of significance.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise B, Question 2

Question:

a Calculate the product-moment correlation coefficient for the following data, giving values for S_{xx}, S_{yy} and S_{xy}.

x	2	3	4	4	5	5	6
y	7	6	5	4	3	2	1

b Test, for these data, the null hypothesis that there is no correlation between x and y. Use a 1% significance level.
State any assumptions you have made.

Solution:

a

In an exam do not set up a table of values – you will run out of time!
Get the values directly from your calculator.

$\sum x = 29, \sum x^2 = 131, \sum y = 28, \sum y^2 = 140, \sum xy = 99, n = 7$

$$S_{xy} = 99 - \frac{29 \times 28}{7} = -17$$

$$S_{xx} = 131 - \frac{(29)^2}{7} = 10.857$$

$$S_{yy} = 140 - \frac{(28)^2}{7} = 28$$

$$\therefore r = \frac{-17}{\sqrt{(10.857 \times 28)}}$$

$$r = -0.975$$

b

$$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{array}\right\} \quad \text{2-tail } \alpha = 0.01$$

test statistic $= r = -0.975$

critical values $= \pm 0.8745$

t.s. < c.v. since $-0.975 < -0.8745$

so reject H_0

Conclude there is evidence of correlation between x and y.

Assumption: x and y are both normally distributed.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise B, Question 3

Question:

The ages X (years) and heights Y (cm) of 11 members of a football team were recorded and the following statistics were used to summarise the results.

$$\Sigma X = 168, \Sigma Y = 1275, \Sigma XY = 20\,704, \Sigma X^2 = 2585$$
$$\Sigma Y^2 = 320\,019$$

a Calculate the product-moment correlation coefficient for these data.

b Test the assertion that height and weight are positively correlated by using a suitable test. State your conclusion in words and any assumptions you have made. (Use a 5% level of significance.)

Solution:

a $$r = \frac{\left[20\,704 - \frac{168 \times 1275}{11}\right]}{\sqrt{\left[\left(2585 - \frac{168^2}{11}\right)\left(320\,019 - \frac{1275^2}{11}\right)\right]}}$$

$$\boxed{r = \frac{s_{xy}}{s_{xx}s_{yy}}}$$

$r = 0.677$

b $\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array}\right\}$ 1-tail $\alpha = 0.05$

test statistic $= 0.677$

critical values $= 0.5214$

t.s. > c.v. so reject H_0

Conclude there is evidence of positive correlation between the age and height of members of a football team – the older the player, the taller they tend to be.
Assumption: both the ages and the heights of the players are normally distributed.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise B, Question 4

Question:

a Explain briefly your understanding of the term 'correlation'. Describe how you used, or could have used, correlation in a project or in class work.

b Twelve students sat two Biology tests, one theoretical the other practical. Their marks are shown below.

Marks in theoretical test (t)	5	9	7	11	20	4	6	17	12	10	15	16
Marks in practical test (p)	6	8	9	13	20	9	8	17	14	8	17	18

Find to 3 significant figures,

i the value of S_{tp}

ii the product-moment correlation coefficient.

c Use a 0.05 significance level and a suitable test to check the statement that 'students who do well in theoretical Biology also do well in practical Biology tests'.

Solution:

a The product-moment coefficient of correlation is the measure of the strength of the linear link between two variables.
You could use it to investigate whether there is correlation between the age of a lichen and its diameter, for example.

b

t	5	9	7	11	20	4	6	17	12	10	15	16
p	6	8	9	13	20	9	8	17	14	8	17	18

Enter these data into your calculator to obtain the following results directly – *don't* complete a table of results in an exam as you will run out of time!

$$\sum t = 132, \sum t^2 = 1742, \sum p = 147, \sum p^2 = 2057, \sum tp = 1872, n = 12$$

i $S_{tp} = 1872 - \dfrac{132 \times 147}{12}$

$S_{tp} = 255$

ii $r = 0.935$

Get this value directly from your calculator.

c $H_0 : \rho = 0$
$H_1 : \rho > 0$ 1-tail $\alpha = 0.05$

test statistic $= 0.935$

critical values $= 0.4973$

t.s. > c.v. so reject H_0.

Conclude there is positive correlation between theoretical biology and practical biology marks – this implies that students who do well in theoretical biology tests also tend to do well in practical biology tests.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise B, Question 5

Question:

The following table shows the marks attained by 8 students in English and Mathematics tests.

Student	A	B	C	D	E	F	G	H
English	25	18	32	27	21	35	28	30
Mathematics	16	11	20	17	15	26	32	20

a Calculate the product-moment correlation coefficient.

A teacher thinks that the population correlation coefficient between the marks is likely to be zero.

b Test the teacher's idea at the 5% level of significance.

Solution:

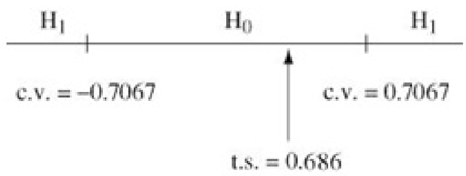
a $r = 0.68556\ldots$

In the exam get this directly from your calculator. If you set up a table of results you are likely to run out of time.

so $r = 0.686$ (3 s.f.)

b

$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{array}\right\}$ 2-tail $\alpha = 0.05$



test statistic $= 0.686$

critical values $= \pm 0.7067$

at upper tail, t.s. < c.v. since $0.686 < 0.7067$ so accept H_0.

Conclude there is evidence the p.m.c.c. could be zero so the teacher's theory is supported.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise B, Question 6

Question:

A small company decided to import fine Chinese porcelain. They believed that in the long term this would prove to be an increasingly profitable arrangement with profits increasing proportionally to sales. Over the next 6 years their sales and profits were as shown in the table below.

Year	1994	1995	1996	1997	1998	1999
Sales in thousands	165	165	170	178	178	175
Profits in £1000	65	72	75	76	80	83

Using a 1% significance level test to see if there is any evidence that the company's beliefs were correct, and that profit and sales were positively correlated.

Solution:

$r = 0.793$ ← Get this value directly from your calculator. *Don't* set up a table of values as you will be likely to run out of time in the exam if you do!

$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array}\right\}$ 1-tail $\alpha = 0.01$

test statistic $= 0.793$

critical values $= 0.8822$

t.s. < c.v. so accept H_0.

Conclude there is insufficient evidence at the 1% significance level to support the company's belief.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 1

Question:

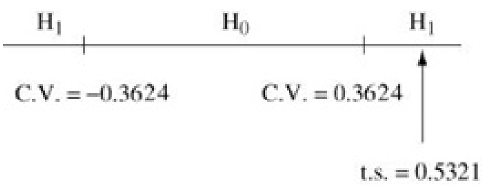
A Spearman's rank correlation obtained from the fuel consumption of a selection of 30 cars and their engine sizes gave a rank correlation coefficient $r_s = 0.5321$. Investigate whether or not the fuel consumption is related to the engine size. State your null and alternative hypotheses. (Use a 5% level of significance.)

Solution:

$H_0: \rho_s = 0$ There is no association between fuel consumption and engine size.
$H_1: \rho_s \neq 0$ There is an association between fuel consumption and engine size.
2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.5321$
$n = 30$ so critical value $= \pm 0.3624$
t.s. > c.v. in the upper tail
since $0.5321 > 0.3624$
so reject H_0.



Conclude there is evidence of an association between fuel consumption and engine size.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 2

Question:

For one of the activities at a gymnastics competition, 8 gymnasts were awarded marks out of 10 for each of artistic performance and technical ability. The results were as follows.

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

The value of the product-moment correlation coefficient for these data is 0.774.

a Stating your hypotheses clearly and using a 1% level of significance, test for evidence of a positive association between technical ability and artistic performance. Interpret this value.

b Calculate the value of the rank correlation coefficient for these data.

c Stating your hypotheses clearly and using a 1% level of significance, interpret this coefficient.

d Explain why the rank correlation coefficient might be the better one to use with these data.

Solution:

a $\left.\begin{array}{l}H_0 : \rho = 0 \\ H_1 : \rho > 0\end{array}\right\}$ $\boxed{\text{1-tail } \alpha = 0.01}$

Test statistic $= r = 0.774$
$n = 8$ critical value $= 0.7887$
t.s. < c.v. in upper tail test
so accept H_0.
Conclude there is insufficient evidence of positive correlation between technical ability and artistic performance at the 1% significance level.

b

T	8.5	8.6	9.5	7.5	6.8	9.1	9.4	9.2
A	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

$\sum d^2 = 10$

$$r_s = 1 - \frac{6 \times 10}{8(8^2 - 1)}$$

$r_s = 0.881$

c $\left.\begin{array}{l}H_0: \rho_s = 0 \text{ There is no association between technical ability and artistic performance.} \\ H_1: \rho_s > 0 \text{ There is positive association between technical ability and artistic performance.}\end{array}\right\}$ 1-tail $\alpha = 0.01$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.881$
$n = 8$ critical value $= 0.8333$
upper tail test, t.s. > c.v.
so reject H_0

Conclude there is evidence of a positive association between technical ability and artistic performance. Gymnasts who are better in their technical ability also appear to be better in their artistic performance.

d The data are discrete results in a limited range. They are judgments, not measurements. It is also unlikely that these scores will both be normally distributed.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 3

Question:

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.

b Use a suitable test, at the 5% level of significance. Interpret your findings to investigate for evidence of positive association between the rankings of the judges. ***E***

Solution:

a

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

$\sum d^2 = 46$

$$r_s = 1 - \frac{6 \times 46}{8(8^2 - 1)}$$

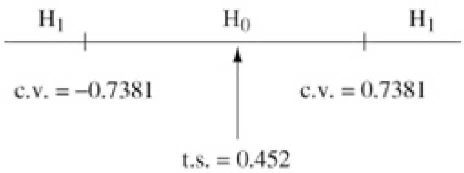
$r_s = 0.452$

b $H_0: \rho_s = 0$ There is no association between the rankings of the 2 judges.
$H_1: \rho_s \neq 0$ There is an association between the rankings of the 2 judges.

2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.452$
Critical values $= \pm 0.7381$
t.s. < c.v. in the upper tail
since $0.452 < 0.7381$
so accept H_0.



Conclude there is no association between the rankings awarded by the 2 judges. They appear to be using different criteria in their judgements.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 4

Question:

Each of the teams in a school hockey league had the total number of goals scored by them and against them recorded, with the following results.

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

Investigate whether there is any correlation between the goals for and those against by using Spearman's rank correlation coefficient. Use a suitable test at the 1% level to investigate the statement, ‘A team that scores a lot of goals concedes very few goals’.

Solution:

F	39	40	28	27	26	30	42
A	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

$\sum d^2 = 76$

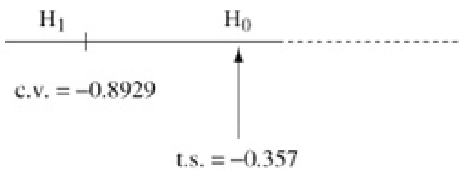
$$r_s = 1 - \frac{6 \times 76}{7(7^2-1)}$$

$r_s = -0.357$

$\left.\begin{array}{l}\text{H}_0: \rho_s = 0 \text{ There is no association between ‘goals for’} \\ \text{and ‘goals against’.} \\ \text{H}_1: \rho_s < 0 \text{ There is negative association between} \\ \text{‘goals for’ and ‘goals against’.}\end{array}\right\}$ 1-tail $\alpha = 0.01$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = -0.357$
critical value $= -0.8929$
Lower tail test where t.s. > c.v.
so accept H_0.



Conclude there is insufficient evidence to support the statement.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 5

Question:

The weekly takings and weekly profits for six branch shops of a company are set out below.

Shop	**1**	**2**	**3**	**4**	**5**	**6**
Takings (£)	400	6200	3600	5100	5000	3800
Profits (£)	400	1100	450	750	800	500

a Calculate the coefficient of rank correlation r_s between the takings and profit.

b It is assumed that profits and takings will be positively correlated. Using a suitable hypothesis test (stating the null and alternative hypotheses) test this assertion at the 5% level of significance.

Solution:

a

T	400	6200	3600	5100	5000	3800
P	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

$\sum d^2 = 2$

$r_s = 1 - \frac{6 \times 2}{6(6^2 - 1)}$

$r_s = 0.943$

b $H_0: \rho_s = 0$ There is no association between takings and profits.
$H_1: \rho_s > 0$ There is positive association between takings and profits.

1-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

upper tail test where t.s. > c.v.
so reject H_0.
Conclude there is evidence to support the statement.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 6

Question:

The rankings of 12 students in Mathematics and Music were as follows.

Mathematics	1	2	3	4	5	6	7	8	9	10	11	12
Music	6	4	2	3	1	7	5	9	10	8	11	12

a Calculate the coefficient of rank correlation r_s. [Show your value of Σd^2.]

b Test the assertion that there is no correlation between these subjects. State the null and alternative hypotheses used. Use a 5% significance level.

Solution:

a

r_{math}	1	2	3	4	5	6	7	8	9	10	11	12
r_{mus}	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

$$\sum d^2 = 58$$

$$r_s = 1 - \frac{6 \times 58}{12(12^2 - 1)}$$

$$r_s = 0.797$$

b $\left.\begin{array}{l}\text{H}_0: \rho_s = 0 \text{ There is no correlation between the rankings in Mathematics and Music.}\\ \text{H}_1: \rho_s \neq 0 \text{ There is correlation between the rankings in Mathematics and Music.}\end{array}\right\}$ 2-tail $\alpha = 0.05$

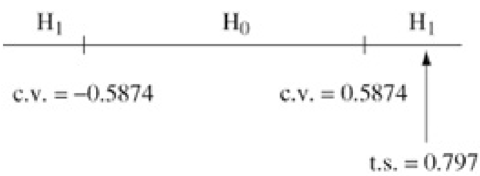
You should always put your hypotheses in terms of ρ first.

Test statistic $= 0.797$

critical values $= \pm 0.5874$

t.s. > c.v. in upper tail, since $0.797 > 0.5874$

so reject H_0.



Conclude there is evidence of an association between the rankings in Mathematics and Music.
Here it appears that students who do well in Mathematics are also likely to do well in Music.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 7

Question:

a A child is asked to place 10 objects in order and gives the ordering

A C H F B D G E J I

The correct ordering is

A B C D E F G H I J

Find a coefficient of rank correlation between the child's ordering and the correct ordering.

b Use a 5% significance level and test whether there is an association between the child's order and the correct ordering. Draw conclusions about this result. ***E***

Solution:

a

Child's order	A	C	H	F	B	D	G	E	J	I
Correct order	A	B	C	D	E	F	G	H	I	J
r_{child}	1	3	8	6	2	4	7	5	10	9
$r_{correct}$	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

$\sum d^2 = 54$

$$r_s = 1 - \frac{6 \times 54}{10(10^2 - 1)}$$

$$r_s = 0.673$$

b $H_0 : \rho_s = 0$ There is no association between the child's ordering and the correct ordering.
$H_1 : \rho_s \neq 0$ There is an association between the child's ordering and the correct ordering.

$\Bigg\}$ 2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= 0.673$

critical values $= 0.5636$

t.s. > c.v. so reject H_0.

Conclude there is evidence of a positive association between the child's ordering and the correct ordering so the child is showing some ability to perform the task.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise C, Question 8

Question:

The crop of a root vegetable was measured over six consecutive years, the years being ranked for wetness. The results are given in the table below.

Year	1	2	3	4	5	6
Crop (10 000 tons)	62	73	52	77	63	61
Rank of wetness	5	4	1	6	3	2

Calculate, to 3 decimal places, a Spearman's rank correlation coefficient for these data. Test the assertion that crop and wetness are not correlated. (Use a 5% level of significance).

Solution:

C	62	73	52	77	63	61
r_c	4	2	6	1	3	5
r_w	5	4	1	6	3	2
d	−1	−2	5	−5	0	3
d^2	1	4	25	25	0	9

$\sum d^2 = 64$

$$r_s = 1 - \frac{6 \times 64}{6(6^2 - 1)} = -0.82857\ldots$$

$r_s = -0.829$

$\left.\begin{array}{l} \mathrm{H}_0: \rho_s = 0 \text{ There is no correlation between the ranks of crop yield and wetness.} \\ \mathrm{H}_1: \rho_s \neq 0 \text{ There is correlation between the ranks of crop yield and wetness.} \end{array}\right\}$ 2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= -0.829$

critical value $= 0.8857$

In the lower tail, t.s. > c.v. since $-0.829 > 0.8857$

so accept H_0.

Conclude at the 1% significance level there is insufficient evidence of correlation between the ranks of crop yield and wetness. (More samples need to be taken as the coefficient is quite close to the critical value).

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation

Exercise D, Question 1

Question:

a Two judges at a cat show place the 10 entries in the following rank orders.

Cat	A	B	C	D	E	F	G	H	I	J
First judge	4	6	1	2	5	3	10	9	8	7
Second judge	2	9	3	1	7	4	6	8	5	10

Find a coefficient of rank correlation between the two rankings and, using the tables provided, comment on the extent of the agreement between the two judges.

b Explain briefly the role of the null and alternative hypotheses in a test of significance. ***E***

Solution:

a

r_F	4	6	1	2	5	3	10	9	8	7
r_S	2	9	3	1	7	4	6	8	5	10
d	2	−3	−2	1	−2	−1	4	1	3	−3
d^2	4	9	4	1	4	1	16	1	9	9

$$\sum d^2 = 58$$

$$r_s = 1 - \frac{6 \times 58}{10(10^2 - 1)} = 0.64848\ldots$$

$$r_s = 0.648$$

We are looking for *positive* correlation between the rankings awarded by the two judges. There is clear evidence of a positive correlation at the 10% and 5% level(s) of significance since our test statistic $r_s = 0.648$ is bigger than the critical value 0.5636. There is insufficient evidence of positive correlation at the 1% significance level where the critical value 0.7455 is greater than the test statistic.

b The null hypothesis is what we assume to be true unless proved otherwise and the alternative hypothesis is what we conclude is happening if we reject the null hypothesis. The null hypothesis is only rejected if it is true with a probability equal to the significance level of the test.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 2

Question:

a Explain briefly the conditions under which you would measure association using a rank correlation coefficient.

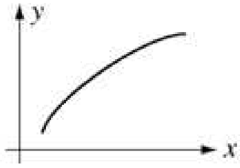
b Nine applicants for places at a college were interviewed by two tutors. Each tutor ranked the applicants in order of merit. The rankings are shown below.

Applicant	A	B	C	D	E	F	G	H	I
Tutor 1	1	2	3	4	5	6	7	8	9
Tutor 2	1	3	5	4	2	7	9	8	6

Investigate the extent of the agreement between the two tutors. ***E***

Solution:

a Where there is a link between the variables x and y but it isn't linear; it could be



for example.

Where the results you have are rankings already where items have been put in order of preference or judgements, or where alphabetical grades have been awarded – here the data won't be from a bivariate normal distribution.

b

r_{T_1}	1	2	3	4	5	6	7	8	9
r_{T_2}	1	3	5	4	2	7	9	8	6
d	0	−1	−2	0	3	−1	−2	0	3
d^2	0	1	4	0	9	1	4	0	9

$\sum d^2 = 28$

$$r_s = 1 - \frac{6 \times 28}{9(9^2 - 1)}$$

$$r_s = 0.7666\ldots = 0.767$$

$$\left.\begin{array}{l}\text{H}_0: \rho_s = 0 \text{ There is no correlation between the ranks awarded by the two tutors.}\\ \text{H}_1: \rho_s > 0 \text{ There is positive correlation between the ranks awarded by the two tutors.}\end{array}\right\} \text{1-tail}$$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.767$

This is greater than the critical value of 0.6833 at the $2\frac{1}{2}\%$ significance level so we would reject H_0 and we would conclude there is evidence of agreement between the two tutors at the $2\frac{1}{2}\%$ significance level.

At the 1% significance level the test statistic and critical value are very close so it is inconclusive at this level of significance.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 3

Question:

In a ski-jump contest each competitor made two jumps. The order of merit for the 10 competitors who completed both jumps are shown.

Ski-jumper	A	B	C	D	E	F	G	H	I	J
First jump	2	9	7	4	10	8	6	5	1	3
Second jump	4	10	5	1	8	9	2	7	3	6

a Calculate, to 2 decimal places, a rank correlation coefficient for the performance of the ski-jumpers in the two jumps.

b Using a 5% significance, and quoting from the table of critical values, investigate whether there is positive association between the two jumps. State your null and alternative hypotheses clearly. ***E***

Solution:

a

r_{1st}	2	9	7	4	10	8	6	5	1	3
r_{2nd}	4	10	5	1	8	9	2	7	3	6
d	−2	−1	2	3	2	−1	4	−2	−2	−3
d^2	4	1	4	9	4	1	16	4	4	9

$\sum d^2 = 56$

$$r_s = 1 - \frac{6 \times 56}{10(10^2 - 1)}$$

$r_s = 0.6606$

$r_s = 0.66$ (2 d.p.)

b $H_0: \rho_s = 0$ There is no correlation between the order of merit for the two jumps.
$H_1: \rho_s > 0$ There is positive correlation between the order of merit for the two jumps.
1-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= 0.66$

Critical value $= 0.5636$

t.s. > c.v. so reject H_0.

Conclude there is evidence of positive correlation between the order of merit for the two jumps – jumpers who did well in the first jump are also likely to do well in the second jump.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 4

Question:

An expert on porcelain is asked to place seven china bowls in date order of manufacture, assigning the rank 1 to the oldest bowl. The actual dates of manufacture and the order given by the expert are shown below.

Bowl	A	B	C	D	E	F	G
Date of manufacture	1920	1857	1710	1896	1810	1690	1780
Order given by expert	7	3	4	6	2	1	5

a Find, to 3 decimal places, the Spearman's rank correlation coefficient between the order of manufacture and the order given by the expert.

b Refer to the table of critical values to comment on your results. State clearly the null hypothesis being tested. ***E***

Solution:

a

Date of manufacture	1920	1857	1710	1896	1810	1690	1780
Date rank	7	5	2	6	4	1	3
Expert rank	7	3	4	6	2	1	5
d	0	2	−2	0	2	0	−2
d^2	0	4	4	0	4	0	4

$\sum d^2 = 16$

$$r_s = 1 - \frac{6 \times 16}{7(7^2 - 1)}$$

$r_s = 0.714$ (3 d.p.)

b $H_0: \rho_s = 0$ There is no association between the pairs of ranks.
$H_1: \rho_s > 0$ There is positive association between the pairs of ranks, i.e. they agree with each other. } 1-tail

You should always put your hypotheses in terms of ρ first.

Test statistic $= 0.714\ldots$
At the 5% significance level the critical value $= 0.7143$
t.s. > c.v. so reject H_0.
Conclude at the 5% level of significance there is minimal evidence that the expert is ranking the porcelain in date order correctly.
At the $2\frac{1}{2}\%$ significance level the critical value is 0.7857 which is bigger than the test statistic meaning H_0 would be accepted. So at the $2\frac{1}{2}\%$ significance level there is insufficient evidence of agreement.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 5

Question:

A small bus company provides a service for a small town and some neighbouring villages. In a study of their service a random sample of 20 journeys was taken and the distances x, in kilometres, and journey times t, in minutes, were recorded. The average distance was 4.535 km and the average journey time was 15.15 minutes.

a Using $\Sigma x^2 = 493.77, \Sigma t^2 = 4897, \Sigma xt = 1433.8$, calculate the product-moment correlation coefficient for these data.

b Stating your hypotheses clearly test, at the 5% level, whether or not there is evidence of a positive correlation between journey time and distance.

c State any assumptions that have to be made to justify the test in **b**. ***E***

Solution:

a $n = 20$

$\bar{x} = \frac{\sum x}{n}$ gives $4.535 = \frac{\sum x}{20}$

so $\sum x = 20 \times 4.535 = 90.7$

$\bar{t} = \frac{\sum t}{n}$ gives $15.15 = \frac{\sum t}{20}$

so $\sum t = 20 \times 15.15 = 303$

$$\therefore r = \frac{\left[1433.8 - \frac{(90.7)(303)}{20}\right]}{\sqrt{\left(493.77 - \frac{90.7^2}{20}\right)\left(4897 - \frac{303^2}{20}\right)}}$$

$r = 0.375$

b $\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array}\right\}$ 1-tail $\alpha = 0.05$

Test statistic $= 0.375$
critical value $= 0.3783$
t.s. < c.v. since $0.375 < 0.3783$
so accept H_0.

Conclude there is insufficient evidence of positive correlation between distance and time at the 5% level of significance.

c Both distance and journey time are normally distributed.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 6

Question:

A group of students scored the following marks in their Statistics and Geography examinations.

Student	A	B	C	D	E	F	G	H
Statistics	64	71	49	38	72	55	54	68
Geography	55	50	51	47	65	45	39	82

a Find the value of the Spearman's rank correlation coefficient between the marks of these students.

b Stating your hypotheses and using a 5% level of significance, test whether marks in Statistics and marks in Geography are associated.

Solution:

a

S	64	71	49	38	72	55	54	68
G	55	50	51	47	65	45	39	82
r_S	4	2	7	8	1	5	6	3
r_G	3	5	4	6	2	7	8	1
d	1	−3	3	2	−1	−2	−2	2
d^2	1	9	9	4	1	4	4	4

$$\sum d^2 = 36$$

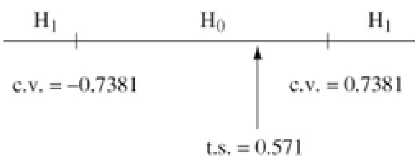
$$r_s = 1 - \frac{6 \times 36}{8(8^2 - 1)}$$

$$r_s = 0.571$$

b $\left.\begin{array}{l}\mathrm{H}_0: \rho_s = 0 \text{ There is no association between marks in Statistics and Geography.} \\ \mathrm{H}_1: \rho_s \neq 0 \text{ There is an association between marks in Statistics and Geography.}\end{array}\right\}$ 2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.571$
critical value $= \pm 0.7381$
In the upper tail t.s. < c.v.
so accept H_0.



Conclude there is no evidence of an association between marks students score in Statistics and Geography, i.e. students who are good at one of these subjects aren't necessarily going to do well in the other subject.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 7

Question:

An international study of female literacy investigated whether there was any correlation between the life expectancy of females and the percentage of adult females who were literate. A random sample of 8 countries was taken and the following data were collected.

Life expectancy (years)	49	76	69	71	50	64	78	74
Literacy (%)	25	88	80	62	37	86	89	67

a Evaluate Spearman's rank correlation coefficient for these data.

b Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence of a correlation between the rankings of literacy and life expectancy for females.

c Give one reason why Spearman's rank correlation coefficient and not the product-moment correlation coefficient has been used in this case. ***E***

Solution:

a

LE	49	76	69	71	50	64	78	74
Lit	25	88	80	62	37	86	89	67
r_{LE}	8	2	5	4	7	6	1	3
r_{LIT}	8	2	4	6	7	3	1	5
d	0	0	1	−2	0	3	0	−2
d^2	0	0	1	4	0	9	0	4

$\sum d^2 = 18$

$$r_s = 1 - \frac{6 \times 18}{8(8^2 - 1)}$$

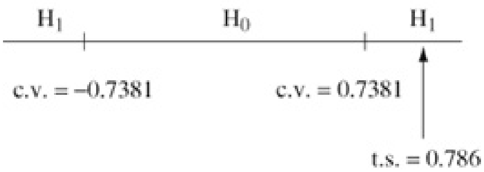
$r_s = 0.786$

b $H_0: \rho_s = 0$ There is no correlation between the rankings of literacy and life expectancy for females.
$H_1: \rho_s \neq 0$ There is correlation between the rankings of literacy and life expectancy for females.

2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.786$
critical value $= \pm 0.7381$
In upper tail t.s. > c.v.
so reject H_0.



Conclude there is evidence of correlation between the rankings of literacy and life expectancy for women. It appears that where a higher percentage of women are literate, they appear to have a higher life expectancy.

c We cannot assume that both life expectancy and percentage of literary are both normally distributed.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 8

Question:

Six friesian cows were ranked in order of merit at an agricultural show by the official judge and by a student vet.
The ranks were as follows:

Official judge	1	2	3	4	5	6
Student vet	1	5	4	2	6	3

a Calculate Spearman's rank correlation coefficient between these rankings.
b Investigate whether or not there was agreement between the rankings of the judge and the student.
State clearly your hypotheses, and carry out an appropriate one-tailed significance test at the 5% level. ***E***

Solution:

a

r_{OJ}	1	2	3	4	5	6
r_{SV}	1	5	4	2	6	3
d	0	−3	−1	2	−1	3
d^2	0	9	1	4	1	9

$\sum d^2 = 24$

$r_s = 1 - \dfrac{6 \times 24}{6(6^2 - 1)}$

$r_s = 0.314$

b $H_0: \rho_s = 0$ There is no correlation between the rankings of the official judge and student vet.
$H_1: \rho_s > 0$ There is positive correlation between the rankings of the official judge and student vet.
1-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.314$
critical value $= 0.8286$
t.s. < c.v. so accept H_0.

Conclude there is insufficient evidence of agreement between the rankings of the official judge and the student vet. They appear to be ranking using different criteria to each other.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 9

Question:

As part of a survey in a particular profession, age, x years, and salary, £ y thousands, were recorded.
The values of x and y for a randomly selected sample of ten members of the profession are as follows:

x	30	52	38	48	56	44	41	25	32	27
y	22	38	40	34	35	32	28	27	29	41

a Calculate, to 3 decimal places, the product-moment correlation coefficient between age and salary.
b State two conditions under which it might be appropriate to use Spearman's rank correlation coefficient.
c Calculate, to 3 decimal places, the Spearman's rank correlation coefficient between age and salary.
It is suggested that there is no correlation between age and salary.
d Set up appropriate null and alternative hypotheses and carry out an appropriate test. (Use a 5% significance level.) ***E***

Solution:

a $r = 0.340$

Get this value directly from your calculator in an examination. Don't set up a table of values – you will run out of time if you do.

b When both sets of data aren't from normal distributions. When at least one set of data is given as grades (letters) *or* ranking of preference or size.

c

x	30	52	38	48	56	44	41	25	32	27
y	22	38	40	34	35	32	28	27	29	41
r_x	8	2	6	3	1	4	5	10	7	9
r_y	10	3	2	5	4	6	8	9	7	1
d	−2	−1	4	−2	−3	−2	−3	1	0	8
d^2	4	1	16	4	9	4	9	1	0	64

$\sum d^2 = 112$

$$r_s = 1 - \frac{6 \times 112}{10(10^2 - 1)}$$

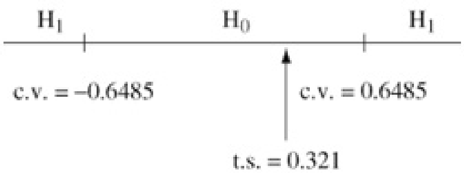
$r_s = 0.321$ (3 d.p.)

d $\left.\begin{array}{l} H_0: \rho_s = 0 \text{ There is no correlation between the rankings of age and salary.} \\ H_1: \rho_s \neq 0 \text{ There is correlation between the rankings of age and salary.} \end{array}\right\}$ 2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Here the test on Spearman's coefficient should be used because it is unlikely that salary and age are both normally distributed.

Test statistic $= r_s = 0.321$
critical value $= \pm 0.6485$
In upper tail, t.s. < c.v.
Since $0.321 < +0.6485$
accept H_0.



Conclude no evidence of correlation between the rankings of salary and age. This means that the older a person is, in the profession, it doesn't mean they will earn more.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 10

Question:

A machine hire company kept records of the age, x months, and the maintenance costs, £y, of one type of machine. The following table summarises the data for a random sample of 10 machines.

Machine	A	B	C	D	E	F	G	H	I	J
Age, x	63	12	34	81	51	14	45	74	24	89
Maintenance costs, y	111	25	41	181	64	21	51	145	43	241

a Calculate, to 3 decimal places, the product-moment correlation coefficient. (You may use $\sum x^2 = 30\,625, \sum y^2 = 135\,481, \sum xy = 62\,412$.)

b Calculate, to 3 decimal places, the Spearman's rank correlation coefficient.

c For a different type of machine similar data were collected. From a large population of such machines a random sample of 10 was taken and the Spearman's rank correlation coefficient, based on $\sum d^2 = 36$, was 0.782.

Using a 5% level of significance and quoting from the tables of critical values, interpret this rank correlation coefficient. Use a two-tailed test and state clearly your null and alternative hypotheses. ***E***

Solution:

a $r = 0.937$

Get this directly from your calculator in the examination. There is no need to set up a table of values; you will run out of time if you do.

b

x	63	12	34	81	51	14	45	74	24	89
y	111	25	41	181	64	21	51	145	43	241
r_x	4	10	7	2	5	9	6	3	8	1
r_y	4	9	8	2	5	10	6	3	7	1
d	0	1	-1	0	0	-1	0	0	1	0
d^2	0	1	1	0	0	1	0	0	1	0

$\sum d^2 = 4$

$$r_s = 1 - \frac{6 \times 4}{10(10^2 - 1)}$$

$$r_s = 0.976 \text{ (3 d.p.)}$$

c $H_0: \rho_s = 0$ There is no association between age and maintenance costs.
$H_1: \rho_s \neq 0$ There is an association between age and maintenance costs.
2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.782$
$n = 10$, critical value(s) $= \pm 0.6485$
In upper tail, t.s. > c.v. since $0.782 > +0.6485$ so reject H_0.

Conclude there is evidence of an association between age and maintenance costs.
It appears that older machines cost more to maintain.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 11

Question:

The data below show the height above sea level, x metres, and the temperature, $y°$ C, at 7.00 a.m., on the same day in summer at nine places in Europe.

Height, x (m)	1400	400	280	790	390	590	540	1250	680
Temperature, y (°C)	6	15	18	10	16	14	13	7	13

a The product-moment correlation coefficient is -0.975. Test this at the 5% significance level. Interpret your result in context.

On the same day the number of hours of sunshine was recorded and Spearman's rank correlation between hours of sunshine and temperature, based on $\sum d^2 = 28$, was 0.767.

b Stating your hypotheses and using a 5% two-tailed test, interpret this rank correlation coefficient. ***E***

Solution:

a $\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho < 0 \end{array}\right\}$ 1-tail $\alpha = 0.05$

Test statistic $= -0.975$

$n = 9$ critical value $= -0.5822$

Lower tail test, t.s. < c.v. since $-0.975 < -0.5822$ reject H_0.

Conclude there is evidence of negative correlation. There is evidence that the greater the height above sea level, the lower the temperature at 7.00 a.m. is likely to be.

b H_0: $\rho_s = 0$ There is no association between hours of sunshine and temperature.
H_1: $\rho_s \neq 0$ There is an association between hours of sunshine and temperature.
} 2-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.767$

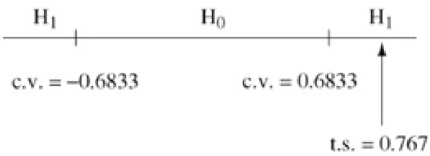
$n = 9$ critical value $= \pm 0.6833$

In upper tail

t.s. > c.v. since

$0.767 > +0.6833$

so reject H_0.



Conclude there is evidence of an association between hours of sunshine and temperature.

The more hours of sunshine the warmer the temperature.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 12

Question:

a Explain briefly, referring to your project work if you wish, the conditions under which you would measure association by using a rank correlation coefficient rather than a product-moment coefficient.

b At an agricultural show 10 Shetland sheep were ranked by a qualified judge and by a trainee judge. Their rankings are shown in the table.

Qualified judge	1	2	3	4	5	6	7	8	9	10
Trainee judge	1	2	5	6	7	8	10	4	3	9

Calculate a rank correlation coefficient for these data.

c Using a suitable table and a 5% significance level, state your conclusions as to whether there is some degree of agreement between the two sets of ranks. ***E***

Solution:

a You use a rank correlation coefficient if at least one of the sets of data isn't from a normal distribution, or if at least one of the sets of data is a letter grading or an order of preference. It is also used if there is a non-linear association between the variables.

b

r_Q	1	2	3	4	5	6	7	8	9	10
r_T	1	2	5	6	7	8	10	4	3	9
d	0	0	−2	−2	−2	−2	−3	4	6	1
d^2	0	0	4	4	4	4	9	16	36	1

$$\sum d^2 = 78$$

$$r_s = 1 - \frac{6 \times 78}{10(10^2 - 1)}$$

$$r_s = 0.527$$

c $H_0: \rho_s = 0$ There is no correlation between the rankings of the qualified and trainee judges.
$H_1: \rho_s > 0$ There is positive correlation between the rankings of the qualified and trainee judges.
1-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.527$

critical value $= 0.5636$

t.s. < c.v. since $0.527 < 0.5636$. These two values are very close.

Accept H_0.

Conclude there is insufficient evidence of agreement between the rankings awarded by the qualified and trainee judges at the 5% level of significance.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 13

Question:

a Explain briefly the use of a null hypothesis and a level of significance in statistical work.

b The positions in a league table of 8 rugby clubs at the end of a season are shown, together with the average attendance (in hundreds) at home matches during the season.

Club	A	B	C	D	E	F	G	H
Position	1	2	3	4	5	6	7	8
Average attendance	30	32	12	19	27	18	15	25

Calculate the coefficient of rank correlation between position in the league and home attendance. Comment on your results. ***E***

Solution:

a The null hypothesis is what is assumed to be true unless proved otherwise. (The alternative hypothesis tells you what is likely to be happening if the null hypothesis is rejected.) The level of significance tells you the probability of rejecting the null hypothesis if it is true. The null hypothesis is only rejected in favour of the alternative hypothesis if by doing so the probability of being wrong is less than or equal to the significance level.

b

Position	1	2	3	4	5	6	7	8
Attendance (hundreds)	30	32	12	19	27	18	15	25
r_p	1	2	3	4	5	6	7	8
r_a	2	1	8	5	3	6	7	4

Keep these rows together in the table to make it easy to find the d values.

d	−1	1	−5	−1	2	0	0	4
d^2	1	1	25	1	4	0	0	16

$$\sum d^2 = 48$$

$$r_s = 1 - \frac{6 \times 48}{8(8^2 - 1)}$$

$$r_s = 0.429$$

There is weak positive correlation between the ranking of attendance and position in the league. Being higher in the league doesn't necessarily mean the attendance will be higher – it may help though.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Regression and correlation
Exercise D, Question 14

Question:

The ages, in months, and the weights, in kg, of a random sample of nine babies are shown in the table below.

Baby	A	B	C	D	E	F	G	H	I
Age (x)	1	2	2	3	3	3	4	4	5
Weight (y)	4.4	5.2	5.8	6.4	6.7	7.2	7.6	7.9	8.4

a The product-moment correlation coefficient between weight and age for these babies was found to be 0.972. By testing for positive correlation at the 5% significance level interpret this value.

b A boy who does not know the weights or ages of these babies is asked to list them, by guesswork, in order of increasing weight. He puts them in the order

A C E B G D I F H

Obtain, to 3 decimal places, a rank correlation coefficient between the boy's order and the true weight order.

c Referring to the tables and using a 5% significance level, investigate for any agreement between the boy's order and the weight order. Discuss any conclusions you draw from you results.

Solution:

a $\left.\begin{array}{l} H_0: \rho = 0 \\ H_1: \rho > 0 \end{array}\right\}$ 1-tail $\alpha = 0.05$

Test statistic $= r = 0.972$

$n = 9$ critical value $= 0.5822$

upper tail t.s. > c.v. since $0.972 > 0.5822$

so reject H_0.

Conclude there is evidence of a positive association between age and weight. This means the older a baby is, the heavier it is likely to be.

b

Weight	4.4	5.2	5.8	6.4	6.7	7.2	7.6	7.9	8.4
Boy's order	A	C	E	B	G	D	I	F	H
r_w	1	2	3	4	5	6	7	8	9
r_b	1	3	5	2	7	4	9	6	8
d	0	−1	−2	2	−2	2	−2	2	1
d^2	0	1	4	4	4	4	4	4	1

$\sum d^2 = 26$

$$r_s = 1 - \frac{6 \times 26}{9(9^2 - 1)}$$

$r_s = 0.783$ (3 d.p.)

c H_0: $\rho_s = 0$ There is no association between the boy's order and the true weight order.

H_1: $\rho_s > 0$ There is positive association between the boy's order and the true weight order.

1-tail $\alpha = 0.05$

You should always put your hypotheses in terms of ρ first.

Test statistic $= r_s = 0.783$

critical value $= 0.6000$

upper tail test where t.s. > c.v. since $0.783 > 0.6000$ so reject H_0.

Conclude there is evidence of positive association between the boy's order and the actual weight of the babies.