

QQQ – Further Statistics 2 – Chapter 2 – Correlation (Version 2)

Total Marks: 25

(25 = Platinum, 23 = Gold, 20 = Silver, 17 = Bronze)

1. An educational expert found that the correlation coefficient between the hours of revision and the scores achieved by 25 students in their A-level exams was 0.379

Her data came from a bivariate normal distribution.

Carry out a hypothesis test at the 1% significance level to determine if there is a positive correlation between the hours of revision and the scores achieved by students in their A-level exams.

The critical value of the correlation coefficient is 0.4622

[4 marks]

(Total 4 marks)

AQA June 2020 Paper 3 Q16

2. In a singing competition, two judges ranked the vocal **quality** of each of 8 singers.

The singers are labelled *A* to *H* and the table below shows the ranks given by each judge.

Rank	1	2	3	4	5	6	7	8
Judge 1	<i>C</i>	<i>B</i>	<i>A</i>	<i>G</i>	<i>E</i>	<i>F</i>	<i>H</i>	<i>D</i>
Judge 2	<i>B</i>	<i>A</i>	<i>C</i>	<i>G</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>H</i>

(a) Calculate the Spearman's rank correlation coefficient between the judges' ranks. **(4)**

Each judge also gave a mark out of 20 for each singer's vocal **range**.

Singer	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
Judge 1	18	13	16	12	11	7	13	6
Judge 2	15	17	14	14	9	8	12	8

(b) Without carrying out any further calculations, explain how the Spearman's rank correlation coefficient between the judges' marks for vocal **range** should be calculated.

(2)

(Total 6 marks)

Edexcel AS June 2025 Q1

3. Anisa is investigating the relationship between marks on a History test and marks on a Geography test. She collects information from 7 students. She wants to calculate the Spearman's rank correlation coefficient for the 7 students so she ranks their performance on each test.

Student	History mark	Geography mark	History rank	Geography rank
<i>A</i>	76	58	1	3
<i>B</i>	70	60	2	2
<i>C</i>	64	57	<i>s</i>	<i>t</i>
<i>D</i>	64	63	<i>s</i>	1
<i>E</i>	64	57	<i>s</i>	<i>t</i>
<i>F</i>	59	50	6	7
<i>G</i>	55	52	7	6

(a) Write down the value of *s* and the value of *t* **(2)**

The full product moment correlation coefficient (pmcc) formula is used with the ranks to calculate the Spearman's rank correlation coefficient instead of $r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$ and the value obtained is 0.7106 to 4 significant figures.

(b) Explain why the full pmcc formula is used to carry out the calculation. **(1)**

(c) Stating your hypotheses clearly, test whether or not there is evidence to suggest that the higher a student ranks in the History test, the higher the student ranks in the Geography test. Use a 5% level of significance. **(4)**

(Total 7 marks)

4. A random sample of 24 adults is taken. The height, h metres, and the arm span, s metres, for each adult are recorded.

These data are summarised below.

$$S_{hh} = 0.377 \quad S_{sh} = 0.352 \quad \bar{s} = 1.70 \quad \bar{h} = 1.68$$

The least squares regression line of h on s is

$$h = a + 0.919s$$

where a is a constant.

(a) Calculate the product moment correlation coefficient. **(3)**

A doctor uses the least squares regression line of h on s as a model to predict a person's height based on their arm span.

(b) Use the model to predict the height of an adult with arm span 1.79 metres. **(2)**

Ewan has an arm span of 1.70 metres and a height of 1.75 metres. His information is added to the sample as the 25th adult.

(c) Explain how the gradient of the regression line for the sample of 25 adults compares with the gradient of the regression line for the original sample of 24 adults.
Give a reason for your answer. **(3)**

(Total 8 marks)

Solutions

1.

Scheme	Marks	A0
States both hypotheses correctly for one-tailed test Do not accept $r = 0$ and $r > 0$ Accept ρ in words eg. H_0 :No correlation/coefficient is 0 H_1 :Positive correlation/coefficient > 0	B1	2.5
Compares given critical value 0.4622 with the given correlation 0.379 coefficient Condone 0.462 or 0.46 or 0.5 for 0.4622 Condone 0.38 or 0.4 for 0.379	M1	3.5a
Compares given critical value 0.4622 with the given correlation 0.379 coefficient correctly and infers H_0 is not rejected Allow reference to H_1	A1	2.2b
Concludes correctly in context that there is insufficient evidence to suggest that there is positive correlation between the hours of revision and the scores achieved by them in their A level exams	R1	3.2a
	(4 marks)	

Typical solution

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

$$0.379 < 0.4622$$

Accept H_0

There is insufficient evidence to suggest that there is a positive correlation between the hours of revision and the scores achieved by them in their A level exams

2.

Mark scheme (a)

Scheme	Marks	AO																											
<table><tr><td></td><td><i>A</i></td><td><i>B</i></td><td><i>C</i></td><td><i>D</i></td><td><i>E</i></td><td><i>F</i></td><td><i>G</i></td><td><i>H</i></td></tr><tr><td>Judge 1</td><td>3</td><td>2</td><td>1</td><td>8</td><td>5</td><td>6</td><td>4</td><td>7</td></tr><tr><td>Judge 2</td><td>2</td><td>1</td><td>3</td><td>5</td><td>6</td><td>7</td><td>4</td><td>8</td></tr></table>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	Judge 1	3	2	1	8	5	6	4	7	Judge 2	2	1	3	5	6	7	4	8	M1	1.1b
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>																					
Judge 1	3	2	1	8	5	6	4	7																					
Judge 2	2	1	3	5	6	7	4	8																					
$\sum d^2 = 1 + 1 + 4 + 9 + 1 + 1 + 0 + 1 \quad [= 18]$	M1	1.1b																											
$r_s = 1 - \frac{6 \times \text{“18”}}{8 \times 63}$	M1	1.1b																											
$= 0.7857 \quad = \text{awrt } \underline{0.786}$	A1	1.1b																											
	(4)																												

Notes

M1: attempt to pair ranking by letter – at least 4 correct for each Judge

M1: attempt to find $\sum d^2$ (some correct d values found and sum attempted)

M1: for using their $\sum d^2$ in formula for r_s with $n = 8$ (Independent of ranking)

A1: awrt 0.786 or exact fraction e.g. $\frac{11}{14}$

Mark scheme (b)

Scheme	Marks	AO
(Rank the marks and) for ties, use average ranks	B1	2.4
Calculate product moment correlation coefficient (using average ranks)	B1	2.4
	(2)	
	(6 marks)	

Notes

B1: explaining that average ranks need to be used. Must see examples or clear statement.

B1: explaining that the pmcc formula must be used on ranks. The “on ranks” may be implied or directly stated. Do not award if suggestion is simply to use pmcc on data.

Mark scheme (a)

Scheme	Marks	AO
$s = 4$	B1	1.1b
$t = 4.5$	B1	1.1b
	(2)	

Notes**B1:** cao**B1:** cao**Mark scheme (b)**

Scheme	Marks	AO
Because there are <u>tied ranks</u> .	B1	2.4
	(1)	

Notes**B1:** Correct explanation**Mark scheme (c)**

Scheme	Marks	AO
$H_0 : \rho_s = 0 \quad H_1 : \rho_s > 0$	B1	2.5
$CV = 0.7143$	B1	1.1b
$r_s = 0.7106$ does not lie in the critical region.	M1	2.1
There is insufficient evidence to suggest that the higher the rank in the History test, the higher the rank in the Geography test (oe).	A1	2.2b
	(4)	
	(7 marks)	

Notes**B1:** Both hypotheses correct with correct notation (must use ρ_s or ρ)**B1:** Correct critical value 0.7143 or better

M1: Drawing a correct inference using their CV and 0.7106

A1: Drawing a correct inference (condone “marks” instead of ranks) in context using their CV and 0.7106

4.

Mark scheme (a)

Scheme	Marks	AO
$S_{ss} = \frac{S_{sh}}{b} = \frac{0.352}{0.919} [= 0.383 \dots] \text{ or } \frac{352}{919}$	M1	2.1
$r = \frac{0.352}{\sqrt{0.377 \times \text{“0.383”}}}$	M1	1.1b
$r = 0.9263 \dots$ awrt <u>0.926</u>	A1	1.1b
	(3)	

Notes

1st M1: realising the need to find S_{ss} Must have $0.919 = \frac{0.352}{S_{ss}}$ or better

2nd M1: attempt to find r ft their S_{ss} provided $S_{ss} > 0.33$

A1: awrt 0.926

Mark scheme (b)

Scheme	Marks	AO
$h - 1.68 = 0.919(1.79 - 1.70)$	M1	3.4
$h = 1.76271$ awrt <u>1.76</u>	A1	1.1b
	(2)	

Notes

M1: using model with means and 1.79 to find h . May see: $0.1177 + 0.919 \times 1.79$

Allow $0.1177 + 0.919 \times 1.75$ as MR

A1: awrt 1.76

Mark scheme (c)

Scheme	Marks	AO
--------	-------	----

S_{ss} would remain the same since $s_{25} = \bar{s}$	M1	2.4
S_{sh} would remain the same additional $(s - \bar{s})(h - \bar{h}) = 0$	M1	1.1b
So the gradient would be unchanged.	A1	2.2a
	(3)	
	(8 marks)	

Notes

1st M1: explaining the effect on S_{ss}

2nd M1: considering the effect on S_{sh}

A1: correct deduction

Alternative

1st M1 explaining that $s_{25} = \bar{s}$ so point lies on $s = \bar{s}$ Must be explicit

2nd M1 $h_{25} > \bar{h}$ so the line would move upwards (parallel to original regression line)

A1 correct deduction