



H09-30min Spearman's Rank Correlation: 30-minute Pack

EDEXCEL 1ST0 · Calculator allowed · about 30 minutes

Total Marks

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** Two food critics, Priya and Tom, independently score 5 new recipes out of 100. Neither critic sees the other's scores.

Recipe	Priya's score	Tom's score
Lentil dahl	82	78
Fish pie	75	80
Mushroom risotto	90	88
Bean chilli	68	65
Prawn curry	95	91

- (a) Rank Priya's scores and Tom's scores separately, giving rank 1 to the highest score in each column. (2)
- (b) Find the difference d between the two ranks for each recipe, and calculate sum of d^2 . (2)
- (c) Calculate r_s for this data. (2)
- (d) Priya and Tom's editor claims: 'The two critics broadly agree on which recipes are best.' Use your value of r_s to comment on this claim. (1)

(Total for Question 1 is 7 marks)

2 Spearman's rank correlation coefficient and the product moment correlation coefficient (PMCC) are both used to measure the strength of a relationship between two variables.

- (a) A dataset contains one extreme outlier that does not fit the general pattern of the rest of the data. Give one reason why Spearman's rank correlation coefficient is less affected by this outlier than the PMCC. (1)
- (b) A researcher only has data given as finishing positions in a race (1st, 2nd, 3rd, and so on), with no times recorded. Explain why Spearman's rank correlation coefficient can be calculated for this data but the PMCC cannot be calculated directly. (1)
- (c) A scientist measures the exact height (in cm) of 10 plants and the exact number of hours of sunlight they receive each day, and the scatter graph shows a clear straight-line pattern with no outliers. State which coefficient, Spearman's rank correlation coefficient or the PMCC, would normally be calculated directly from this raw data, and give a reason. (2)

(Total for Question 2 is 4 marks)

3 For each statement about Spearman's rank correlation coefficient, write down whether it is True or False.

(i) Spearman's rank correlation coefficient, r_s , can only take values between -1 and 1 inclusive. (1)

(ii) If two rankings share tied values, the ranks used in the formula $r_s = 1 - (6 \times \text{sum of } d^2) / (n(n^2-1))$ should be the mean of the positions the tied values would otherwise occupy. (1)

(iii) A value of $r_s = -0.9$ shows a weaker association than a value of $r_s = 0.4$. (1)

(iv) If n increases but sum of d^2 stays exactly the same, the value of r_s calculated from the formula must also stay exactly the same. (1)

(Total for Question 3 is 4 marks)

- 4 At a regional science fair, two moderators, Moderator Chen and Moderator Osei, independently score 8 projects out of 50. Ties occur in both moderators' scores.

Project	Moderator Chen score	Moderator Osei score
P1	45	39
P2	38	44
P3	45	33
P4	30	28
P5	27	30
P6	41	47
P7	33	28
P8	38	41

- (a) Chen's scores for P1 and P3 are tied, and so are the scores for P2 and P8. (2)
Complete the ranking of Chen's scores for all 8 projects, using mean ranks for the tied pairs.
- (b) Osei's scores for P4 and P7 are tied. Complete the ranking of Osei's scores for all 8 projects. (2)
- (c) Calculate d for each project and find sum of d^2 . (2)
- (d) Calculate r_s , giving your answer correct to 3 significant figures. (3)
- (e) Interpret this value of r_s in the context of the two moderators' scoring. (1)

(Total for Question 4 is 10 marks)

- 5** A student calculates Spearman's rank correlation coefficient for 5 pairs of rankings, using sum of $d^2 = 8$. Their working is shown below.

$$r_s = 1 - (6 \times 8) / 5^2 = 1 - 48/25 = 1 - 1.92 = -0.92$$

(a) Identify the mistake in the student's working. (2)

(b) Work out the correct value of r_s . (2)

(c) Explain how the student's mistake affected their conclusion about the two rankings. (1)

(Total for Question 5 is 5 marks)

Mark scheme · H09-30min Spearman's Rank Correlation: 30-minute Pack

Question 1

- (a) **B1** Priya's ranks 3, 4, 2, 5, 1 cao (Lentil dahl to Prawn curry)
- (a) **B1** Tom's ranks 4, 3, 2, 5, 1 cao (Lentil dahl to Prawn curry)
- (a) **Answer: Priya: 3, 4, 2, 5, 1. Tom: 4, 3, 2, 5, 1 (Lentil dahl, Fish pie, Mushroom risotto, Bean chilli, Prawn curry).**
- (b) **M1** d values -1, 1, 0, 0, 0 oe
- (b) **A1** sum of $d^2 = 2$ cao
- (b) **Answer: sum of $d^2 = 2$**
- (c) **M1** $r_s = 1 - (6 \times 2) / (5 \times 24)$, oe
- (c) **A1** 0.9 cao
- (c) **Answer: 0.9**
- (d) **B1** $r_s = 0.9$ is close to 1, supporting the claim that the two critics broadly agree oe
- (d) **Answer: The claim is supported - $r_s = 0.9$ is close to the maximum of 1, showing the critics broadly agree on the ranking of the recipes.**

Question 2

- (a) **B1** Spearman's rank correlation is calculated using the ranks (order) of the data rather than the actual values, so an extreme value only occupies the highest or lowest rank rather than distorting the calculation with its size oe
- (a) **Answer: Because Spearman's rank uses only the order (ranks) of the data, not the actual values, so one extreme value cannot distort the coefficient as much as it could distort the PMCC.**
- (b) **B1** the PMCC needs actual numerical data values to work with, but finishing positions are already ranks/ordinal data, so only Spearman's rank correlation coefficient can be applied directly oe
- (b) **Answer: The PMCC requires actual numerical measurements, but finishing positions are already just an order (ranks), so Spearman's rank correlation coefficient can be used directly, whereas the PMCC cannot.**
- (c) **B1** the PMCC
- (c) **B1** reason: the data is already precise numerical data with a linear pattern, so using the PMCC directly makes full use of the actual values rather than converting to ranks and losing information oe
- (c) **Answer: The PMCC - the data is precise numerical measurements with a linear pattern, so using the actual values (rather than converting to ranks) makes full use of the information available.**

Question 3

- (i) **B1** True cao
- (i) **Answer: True**
- (ii) **B1** True cao
- (ii) **Answer: True**
- (iii) **B1** False cao
- (iii) **Answer: False**
- (iv) **B1** False cao
- (iv) **Answer: False**

Question 4

- (a) **M1** P1 and P3 both share rank 1.5 (mean of 1 and 2); P2 and P8 both share rank 4.5 (mean of 4 and 5)
- (a) **A1** full set of Chen ranks: P1 1.5, P2 4.5, P3 1.5, P4 7, P5 8, P6 3, P7 6, P8 4.5
- (a) **Answer: Chen ranks: P1 1.5, P2 4.5, P3 1.5, P4 7, P5 8, P6 3, P7 6, P8 4.5**
- (b) **M1** P4 and P7 both share rank 7.5 (mean of 7 and 8)
- (b) **A1** full set of Osei ranks: P1 4, P2 2, P3 5, P4 7.5, P5 6, P6 1, P7 7.5, P8 3
- (b) **Answer: Osei ranks: P1 4, P2 2, P3 5, P4 7.5, P5 6, P6 1, P7 7.5, P8 3**
- (c) **M1** d values -2.5, 2.5, -3.5, -0.5, 2, 2, -1.5, 1.5 oe
- (c) **A1** sum of $d^2 = 37.5$ cao
- (c) **Answer: sum of $d^2 = 37.5$**
- (d) **M1** $n(n^2-1) = 8 \times 63 = 504$ oe
- (d) **M1** $r_s = 1 - (6 \times 37.5) / 504$, oe
- (d) **A1** awrt 0.554
- (d) **Answer: $r_s = 0.554$ (3sf)**
- (e) **B1** a moderate positive association - the two moderators show some agreement in how they rank the projects, but not especially strong agreement oe
- (e) **Answer: A moderate positive association - the moderators agree on the general ranking of the projects to some extent, but not very strongly.**

Question 5

- (a) **B1** the denominator used is n^2 instead of the correct $n(n^2-1)$
- (a) **B1** states the correct denominator: $n(n^2-1) = 5 \times 24 = 120$, not $n^2 = 25$
- (a) **Answer: The student used $n^2 (=25)$ instead of the correct denominator $n(n^2-1) (=120)$.**
- (b) **M1** $r_s = 1 - (6 \times 8) / 120$, oe
- (b) **A1** 0.6 cao
- (b) **Answer: 0.6**
- (c) **B1** the student's incorrect value (-0.92) suggested a strong negative association/disagreement, but the correct value (0.6) actually shows a moderate positive association/agreement - the error leads to the opposite conclusion oe
- (c) **Answer: The student's error changed both the size and the sign of the result: -0.92 wrongly suggests the two rankings strongly disagree, when the correct value of 0.6 actually shows a moderate positive agreement between them.**