inkMaths

ANSWER BOOK

Correlation coefficients: product moment and Spearman

Further Statistics 2 · Year FM

6 questions · 22 marks

NAVY EDITION

inkphysics.co.uk/maths/ · youtube.com/@InkPhysics

SECTION A · Warm-up

A1 It measures the strength and direction of the linear association between two variables, and always lies between −1 and 1 inclusive. B1 for the strength and direction of linear association, B1 for the range −1 to 1. Values at the ends mean the points lie exactly on a straight line, falling or rising; zero means no linear association, which is not the same as no relationship. [2]

A2 $r = 58/\sqrt{(40 \times 86)} = 58/\sqrt{3440} = 58/58.65 =$ **0.989** to three decimal places. That is very strong positive linear correlation, so the points lie almost exactly on a rising straight line. M1 for the formula, A1 for 0.989, B1 for the interpretation. An interpretation in words is a mark of its own, and a bare number does not earn it. [3]

SECTION B · Standard

B1 Coding by a shift and a positive scale factor leaves r unchanged, because the scatter is only translated and stretched, so its straightness is untouched. The gradient does change. If the regression of v on u has gradient B, then $b = B \times (10/5) = 2B$, since y varies ten units for every one of v and x five for every one of u. B1 for r unchanged, B1 for the reason, M1 for relating the two gradients, A1 for $b = 2B$. A negative scale factor on exactly one variable would flip the sign of r. [4]

B2 The differences are 0, −1, 1, −1, 1, −1, 1, so $\Sigma d^2 = 0 + 1 + 1 + 1 + 1 + 1 + 1 = 6$. With $n = 7$, $n(n^2 - 1) = 7 \times 48 = 336$, so $r_s = 1 - 36/336 = 1 - 0.1071 =$ **0.893**. M1 for the differences, A1 for $\Sigma d^2 = 6$, M1 for the formula, A1 for 0.893. The two rankings agree closely. [4]

B3 Tied items share the average of the ranks they would have occupied, so two tying for third and fourth both take rank 3.5, and the next item takes rank 5. With $n = 6$, $n(n^2 - 1) = 6 \times 35 = 210$, so $r_s = 1 - 48/210 =$ **0.771**. B1 for the shared rank of 3.5, M1 for the formula, A1 for 0.771. With ties present the formula is an approximation; the product moment coefficient applied to the ranks is exact. [3]

SECTION C · Stretch

C1 **Product** moment. With $n = 6$, $\Sigma x = 21$, $\Sigma y = 69$, $\Sigma xy = 342$, $\Sigma x^2 = 91$ and $\Sigma y^2 = 1497$. [6]
$S_{xy} = 342 - (21)(69)/6 = 342 - 241.5 = 100.5$. $S_{xx} = 91 - 441/6 = 17.5$. $S_{yy} = 1497 - 4761/6 = 703.5$.
$r = 100.5/\sqrt{(17.5 \times 703.5)} = 100.5/110.96 =$ **0.906**. Set the five sums out before substituting; the method marks are for those, and a single wrong total sinks the whole answer.
Spearman. The counts increase throughout, so ranking them 1 to 6 gives exactly the same order as the times. Every d is zero, so $\Sigma d^2 = 0$ and $r_s = 1 - 0 =$ **1**.
Why they differ. The relationship is a perfect increasing one, which is all that Spearman's coefficient asks about, so it returns its maximum. The product moment coefficient asks whether the points lie on a straight line, and these lie on a curve that steepens sharply, so it falls short of 1. M1 for S_{xy}, S_{xx} and S_{yy}, A1 for $r = 0.906$, M1 for ranking the counts, A1 for the Spearman value of 1, B1 for Spearman needing only a monotonic relationship, B1 for the product moment coefficient measuring straightness.
Spearman is the sensible choice when the relationship is monotonic but curved, or when the data are already ranks. The product moment coefficient is for linear association in genuinely numerical data. A value of r near zero would not even rule out a relationship; the points $(-2, 4)$, $(-1, 1)$, $(0, 0)$, $(1, 1)$, $(2, 4)$ give $r = 0$ while y is exactly x^2.