Considering Correlation Coefficients: The Meaning of Zero Correlation

Bonnie H. Litwiller and David R. Duncan

Teachers often search for situations in which correlation coefficients can be computed and interpreted. We will discuss three examples.

All students who have studied statistical concepts are familiar with the formula for correlation coefficients. If X and Y are two paired variables, the correlation coefficient describing the strength of the relationship between X and Y is:

$$r = \frac{n\sum(XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum(X^2) - (\sum X)^2][n\sum(Y^2) - (\sum Y)^2]}}$$

Most students are aware that $-1 \leq r \leq 1$ and that if r is close to either -1 or $+1$, the two variables are strongly related. If r is close to $+1$, the relationship is direct (large values of X are typically associated with large values of Y and small values of X with small values of Y); if r is close to -1, the relationship is inverse (large Xs are typically associated with small Ys and small Xs with large Ys). What happens if the correlation coefficient is at or near zero? What does a coefficient of zero mean?

Example 1

Suppose that these are pairs of test scores for five students. The X value is the student's score on test 1 and the Y value is the student's score on test 2. Using the data of Table 1, the correlation coefficient can be computed.

Data Table 1

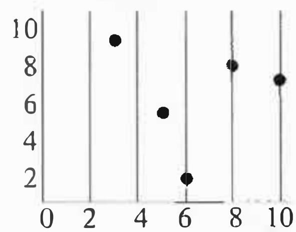
X	Y	X^2	XY	Y^2
3	10	9	30	100
5	6	25	30	36
6	2	36	12	4
8	9	64	72	81
10	8	100	80	64
32	35	234	224	285

$$r = \frac{5(224) - 32(35)}{\sqrt{[5(234) - (32)^2][5(285) - (35)^2]}}$$

$$= \frac{0}{\sqrt{(146)(200)}}$$

$$= 0$$

Graph 1



The zero coefficient suggests the absence of a relationship between the X and Y variables. In the graph, the five pairs are depicted; on inspection, no obvious relationship appears.

Example 2

Again five (X, Y) pairs of test scores are reported and the coefficient is computed.

Data Table 2

X	Y	X^2	XY	Y^2
0	4	0	0	16
1	1	1	1	1
2	0	4	0	0
3	1	9	3	1
4	4	16	16	16
10	10	30	20	34

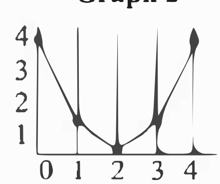
$$r = \frac{5(20) - 10(10)}{\sqrt{[5(30) - (10)^2]\cdot[5(34) - (10)^2]}}$$

$$= \frac{0}{\sqrt{(50)(70)}}$$

$$= 0$$

The zero correlation again suggests the absence of a relationship between the (X, Y) pairs; however, an examination of Graph 2 reveals that the parabola $Y = (X-2)^2$ perfectly describes the relationship between X and Y.

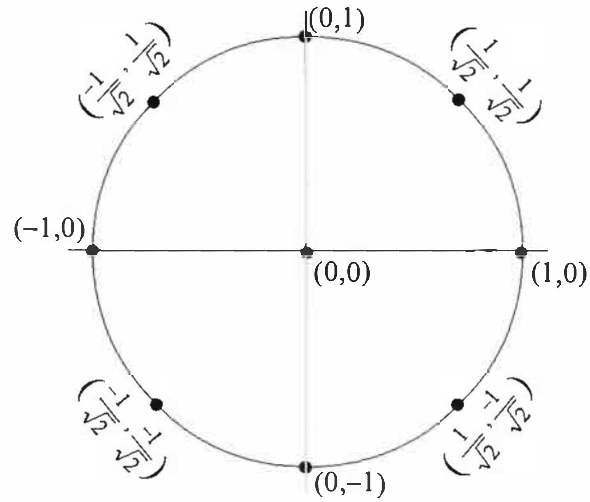
Graph 2



Why was the correlation coefficient unaware of this relationship? This anomaly occurs because the coefficient is looking only for a linear relationship between X and Y. The absence of a linear relationship on Graph 2 causes the coefficient to be 0, even though a quadratic relationship occurred.

Example 3

Graph 3



Graph 3 displays eight (X, Y) pairs on the unit circle ($X^2 + Y^2 = 1$). We note that there is a geometric relationship among these (X, Y) pairs. Is the correlation coefficient aware of this relationship?

Data Set 3

X	Y	X^2	XY	Y^2
1	0	1	0	0
$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	1	0	0	1
$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$
−1	0	1	0	0
$\frac{-1}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	−1	0	0	1
$\frac{1}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{-1}{2}$	$\frac{1}{2}$
0	0	4	0	4

$$r = \frac{8(0) - 0(0)}{\sqrt{[8(4) - (0)^2] \cdot [8(4) - (0)^2]}}$$

$$= 0$$

The correlation coefficient was totally unaware of the circular relationship that is evident when Graph 3 is inspected.

The key idea behind these examples is that the correlation coefficient measures the strength of the linear relationship between X and Y. Our original wording preceding the formula at the beginning of this article was, thus, incomplete.

Challenge for the Readers and Their Students

Display and calculate the correlation coefficient for other sets of (X, Y) pairs in which a nonlinear relationship is present.