The Constancy of the Correlation Coefficient with Respect to Linear Transformations

By David R. Duncan and Bonnie H. Litwiller

Abstract. Teachers are often searching for situations in which correlation coefficients can be considered. Of particular interest is the case in which linear transformations are applied to the variables whose correlation is being investigated. We will apply several such transformations to an initial example.

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:

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

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" linearly 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 X's are typically associated with small Y's and small X's with large Y's). If the correlation coefficient is near zero, there is a minimal linear relationship.

Data Table 1 contains 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 this table, the correlation coefficient can be computed.

X	Y	X^2	XY	Y^2
3	7	9	21	49
5	8	25	40	64
6	10	36	60	100
8	12	64	96	144
10	10	100	100	100
32	47	234	317	457

Table 1

$$r = \frac{(5)(317) - (32)(47)}{\sqrt{\left((5)(234) - (32)^2\right)\left((5)(457) - (47)^2\right)}} \approx 0.77$$

The relatively high positive coefficient suggests a fairly strong direct linear relationship between X and Y.

What happens if linear transformations are applied to each of the variables? We will consider four sets of transformations.

(1) Suppose that a given constant is added to each X and another given constant is added to each Y. What happens to the coefficient of correlation? For instance, suppose that 3 is added to each X entry of Table 1 and (-2) is added to each Y entry of Table 1. Table 2 reports these results.

X	Y	X^2	XY	Y^2
6	5	36	30	25
8	6	64	48	36
9	8	81	72	64
11	10	121	110	100
13	8	169	104	64
47	37	471	364	289

Table 2

$$r = \frac{(5)(364) - (47)(37)}{\sqrt{\left((5)(471) - (47)^2\right)\left((5)(289) - (37)^2\right)}} \approx 0.77$$

The coefficient is unchanged by these linear transformations.

2. Suppose that each X (from Table 1) is multiplied by a given constant and each Y is multiplied by another given constant. What happens to the coefficient? For example, suppose that each X entry of Table 1 is multiplied by 3 and that each Y entry is multiplied by $\frac{1}{2}$. Table 3 displays this situation.

X	Y	X^2	XY	Y^2
9	3.5	81	31.5	12.25
15	4	225	60	16
18	5	324	90	25
24	6	576	144	36
30	5	900	150	25
96	23.5	2106	475.5	114.25

Table 3

$$r = \frac{(5)(475.5) - (96)(23.5)}{\sqrt{\left((5)(2106) - (96)^2\right)\left((5)(114.25) - (23.5)^2\right)}} \approx 0.77$$

Again, the coefficient is unaffected by these linear transformations.

3. Based upon the results of 1 and 2, it may be predicted that successive applications of multiplication and addition transformations to the data of Table 1 will leave the correlation coefficient invariant. Let us test this hypotheses with $X \to \frac{1}{2}X - 3$ and $Y \to 3Y + 4$. The results are shown in Table 4.

X	Y	X^2	XY	Y^2
-1.5	25	2.25	-37.5	625
-0.5	28	0.25	-14	784
0	34	0	0	1156
1	40	1	40	1600
2	34	4	68	1156
1	161	7.5	56.5	5321

Table 4

$$r = \frac{(5)\,(56.5) - (1)\,(161)}{\sqrt{\left((5)\,(7.5) - (1)^2\right)\left((5)\,(5321) - (161)^2\right)}} \approx 0.77$$

Again, the coefficient is unchanged!

4. What happens if the multiplicative factor of one of the variables is negative: $X \to 2X - 3$ and $Y \to -3Y + 4$. Table 5 reports the situation:

X	Y	X^2	XY	Y^2
3	-17	9	-51	289
7	-20	49	-140	400
9	-26	81	-234	676
13	-32	169	-416	1024
17	-26	289	-442	676
49	-121	597	-1283	3065

Table 5

$$r = \frac{(5)\,(-1283) - (49)\,(-121)}{\sqrt{\left((5)\,(597) - (49)^2\right)\left((5)\,(3065) - (-121)^2\right)}} \approx -0.77$$

The correlation coefficient is unchanged in absolute value, but, its sign is changed from that of the previous examples. A "direct" relationship is changed into an "inverse" relationship when the ordering of the Y's is essentially reversed by the negative factor.

In summary, the coefficient of correlation is impervious to change in absolute value when linear transformations are applied; but, the sign of the coefficient is reversed if one multiplication is positive and one is negative.

Challenges:

(1) What happens to the correlation coefficient if a negative multiplication is applied to both X and Y?

(2) Verify these relationships symbolically if $aX + b$ and $cY + d$ transformations are applied to $X_1, X_2, X_3, ..., Xn$ and $Y_{l1}Y_2, Y_3, ..., Yn$.

(3) Investigate other relationships involving coefficients of correlation.

(4) Investigate situations leading to a zero coefficient. Does this always signify the absence of any relationship between X and Y ?

Department of Mathematics
University of Northern Iowa
Cedar Falls, IA 50614-0506