Some Remarks on Teaching the Correlation Coefficient

Kameo Matusita

5-28-7 Kamiuma, Setagaya-ku,
Tokyo, Japan 154-0011

In this paper the author would like to state, by giving examples, some matters to be attended to in teaching the correlation coefficient. The matters are not new especially, but seem to be often ignored or only briefly touched on. Further, the relation between the correlation coefficient and the affinity of the distributions concerned is referred to in the Gaussian case.

1. In teaching multivariate analysis the correlation coefficient is usually treated. Let r stand for the correlation coefficient of two variates X, Y. In school students normally learn that

(i) r is conecerned with linear relationship between X and Y,
(ii) $-1 \leqq r \leqq 1$
(iii) $r = \pm 1$ when and only when $Y = aX + b$ *w. p. 1*, a, b being constants,
etc.

Further, students learn that

(iv) when $r = 0$, X and Y are said to be uncorrelated.

Now, we often notice that students tend to consider that:

(a) when the value of the correlation coefficient is large (small), the relation between the two variates is close to the linear, thus, when $r = 0.9$ or 0.95 the relation is nearly linear;

(b) when the value of the correlation coefficient is zero or near zero the two variates have no or almost no functional relation;

or

(c) when the value of the correlation coefficient is positive(negative) , the value of Y becomes larger (smaller) as a whole, as the value of X becomes large.

Of course, these are not always correct, but students who hold these (a), (b), (c) to be correct are liable to make wrong interpretations of the results they obtain in their research of real problems. Therefore, it is necessary to call students' attention to the matters. The best way for that will be to show them counter-examples to each of (a), (b), (c) .

In the following, we will give counter-examples. Similar examples may be found in the literature, but ours are a little more detailed or different.

2. Example 1 (to (a)).

Let X be a random variable uniformly distributed over $[\, 0\, , G\,]$, G being any positive number, and let $Y = X^a$, a being a positive number. Then, the correlation coefficient of X and Y is independent of G, and for varying a, we have:

a	Cor (X, Y)
0.1	0.9035
0.5	0.9798
0.8	0.9974
1	1.0000
1.2	0.9980
1.5	0.9897
2	0.9682

(see (5)). It is easy to see, especially when G is large, that for $a \neq 1$ the relation between X and Y is not close to linearity. Here "being close to" is a somewhat subjective notion. However, everyone will agree that $y = x^{0.5}$ or $y = x^{1.5}$ is not close to $y = x$. As to $y = x^{0.8}$, or $y = x^{1.2}$ (corresponding to $r = 0.9974$, 0.9981, respectively), we have, for instance, $20^{0.8} - 20 = 10.98 - 20 = -9.02$; $20^{1.2} - 20 = 36.41 - 20 = 16.41$. As we can take as large G as we like, we can obtain much larger difference values. This means that $y = x^{0.8}$ and $y = x^{1.2}$ are distant from linearity when the bound G is large. The same can be said for $y = x^{a}$ with any positive $a (\neq 1)$.

This example shows that even for a large value of the correlation coefficient of X and Y we can not consider *unconditionally* that the relation between X and Y is close to linearity.

Example 2 (to (b)).

Let $X = \cos\theta$, $Y = \sin\theta$, θ being uniformly distributed over $[\varepsilon, 2\pi]$, $0 \leqq \varepsilon < 2\pi$. Then we have $X^2 + Y^2 = 1$, and Cor $(X, Y) \to 0$ $(\varepsilon \to 0)$. Especially, when $\varepsilon = 0$, Cor $(X, Y) = 0$.

This example shows that even when the correlation coefficient between X and Y is zero or small there can be a relation between X and Y.

Example 3 (to (c))

Let (X, Y) take $(1, -1)$, $(2, -2)$, $(3, -3)$, $(4, -4)$, $(5, 20)$ each with probability 1/5. Then we have

Cor $(X, Y) = 0.62$

Concerning the first four points Y decreases as X increases.

This example shows that even when the correlation coefficient between X and Y is positive Y does not always increase *as a whole* as X increases.

3. As is seen from the above mentioned,

> *it is risky to presume the relation between two variates only from the numerical value of the correlation coefficient. Therefore, it is still more risky when the value of the correlation coefficient is calculated from a sample.*

Students ought to memorize this to avoid making errors. It is often noticed that students see only numerical values obtained through computers.

4. Then the problem will arise: how does the correlation coefficient work in general? In Gaussian cases, the correlation coefficient seems to express the situation of the two variables concerned between independence and linearity. In fact, we can show it by means of the affinity of the distributions related to the joint distribution of the two variables (see, for example, (2)). The affinity is defined as follows.

Let F and G be distributions in the space R with measure m , and let f and g be their density functions $w.\ r.\ t.$ measure m , respectively. Then the affinity of F and G is defined as

$$\rho(F,\ G\ )=\ \int_R (\ f\ g\ )^{\frac{1}{2}}dm\ \ .$$

For this ρ we have $0 \leqq \rho \leqq 1$ and when and only when $F = G,\ \rho = 1$.
Further, when we consider the distance between F and G,

$$d_2(F,G)=\sqrt{\int_R (f^{\frac{1}{2}}-g^{\frac{1}{2}})^2 dm}$$

we have

$$d_2^2(F,G) = 2(1-\rho)\ .$$

Thus we can see that ρ represents closeness between F and G.

Let (X_1, X_2) be distributed as the Gaussian $F = N\left(0,\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}\right)$, ($\sigma_{12}= \sigma_{21}$), and let F_1 , F_2 be the marginal distributions of X_1 , X_2 , respectively. Further, let G be the distribution $N\left(0,\begin{pmatrix}\sigma_{11} & 0\\ 0 & \sigma_{22}\end{pmatrix}\right)$, that is, the direct product of F_1 and F_2 . Let ρ denote the affinity of F and G, and let r denote the correlation coefficient between X_1 and X_2 . Then as a relation between ρ and r we have

$$\rho\ =\ \frac{2(1-r^2)^{\frac{1}{4}}}{(4-r^2)^{\frac{1}{2}}}$$

(see (4)), Through this relation we can see that the correlation coefficient r is connected to the distance between F and G. As to this relation we have:

r	ρ	$-\log_2(1-\rho)$
0.0	1.0000	
0.1	0.9987	9.5873
0.3	0.9879	6.3688
0.5	0.9611	4.6841
0.7	0.9021	3.3540
0.9	0.7393	1.9395
1.0	0.0000	0.0000

5. In the above, we have mentioned about the Gaussian distribution. However, we can hardly expect that real phenomena are *fully* described by the Gaussian distributions. When dealing with real phenomena we often observe outliers. It is advised to let students know about those matters. On the other hand, we have to consider methods of treating the problem which reduce the influence of outliers.

6. When we treat the correlation coefficient we have to consider in what situation X and Y

appear. It is very important to make students grasp the correlation graphically. That is, let $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ be observed values of (X, Y). Make students plot them on the plane and look at the scatter of the points carefully. Then they will obtain some idea of the correlation of X and Y. They will notice the existence of outliers, too. As a smmary of these points, we can consider a (resistant) fitted line to the scatter of those points. (See, for example, (1)) . By a fitted line we can roughly grasp how X and Y correlate.

Now, when a fitted line *well* represent the whole points of a data set, that is, the plot of the data set shows basically the straight line trend except for some randomness, we say, as is known, that the data set has a linear structure. Then, as to the correlation coefficient, it will be recommended to consider it in cases of data sets with linear structure or of data from distributions like the Gaussian. Therefore, when we consider the correlation coefficient it is important to plot the data and regard the figure at the same time. We should make students bear the matter in mind. Further, it will be better to make them understand that the correlation coefficeient of two variables is their dimensionless covariance and this alone will not give much information about their relationship or their (joint) distribution (except for the Gaussian). (Of course, the correlation coefficient with the value 1 tells us linearity.)

REFERENCES

(1) Emerson,J.D. and Hoaglin,D.C. (1983): Resistant Lines for y versus x, In: *Understanding Robust and Exploratory Data Analysis*, Hoaglin,D.C., Mosteller,F. and Tukey, J.(ed.), John Wiley. & Sons,

(2) Matusita,K. (1966): A Distance and Related Statistics in Multivaliate Analysis, In: *Multivariate Analysis*, Krishnaiah,P.R.(ed.), Academic Press.

(3) Matusita,K. (1965): A Classification Based on Distance in Multivariate Gaussian Cases, In: *Proc. Fifth Berkeley Symp. on Math. Stat. And Prob.*, Univ. Calif. Press.

(4) Matusita,K. (1973): Correlation and Affinity in Gaussian Cases, In: *Multivariate Analysis III*, Krishnaiah, P.R.(ed.), Academic Press.

(5) Mosteller,F. and Tukey,J. (1997): *Data Analysis and Regression*, Addison Wesley.

RÉSMÉ

Dans cet article l'auteur présente quelques sujets à remarquer en enseignement du coefficient de corrélation. En plus, il donne une relation entre le coefficient de corrélation et l'affinité des distributions concernées dans le cas gaussien.