CORRELATION AND TIME SERIES

BY

D. D. KOSAMBI

(*Fergusson College, Poona*)

IN a recent attempt to examine the significance of reverse marks on Taxilan silver punch-marked coins, 753 square coins found in a pre-Mauryan hoard were tabulated by weight in arrays of 0, 1, ..., 10 reverse marks. The correlation r between reverse mark and weight was found to be $-\cdot 46$. Because the evidence pointed to the reverse marks being regularly placed in time, 3,000 current British Indian rupees were taken from active circulation, and their weights determined as a control measure. Discarding counterfeits, mint-defectives, and the (superseded) rare Victoria rupees, there remained 2,886 specimens in 18 arrays, one for each year from 1903 to 1920. The correlation between date of issue and weight was found to be ·43, which is compatible with the Taxilan value. The two values would actually be closer if Sheppard's corrections were applied, because the grouping unit for British coins is coarser, ·01 gm. (though the coins were weighed to ·0001 gm.) as against ·1 gr. for Taxilan specimens. The question as to this correlation value being a characteristic of all coinage regardless of period and denomination can only be settled by numerous observations on other currency. However, an affirmative result need not be taken as surprising because r measures the strength of association between time and weight in a manner that is independent of the unit of time, and of weight, therefore independent of the rate of wear. It is just possible that the extent to which date of issue is relevant information as regards the mean weight of the group remains the same over a large geographical region and great duration of time.

The next step is to examine whether the correlations give equally good linear regressions in both cases. The relevant information is summed up in Table I.

TABLE I

	D. F.	Sum Squares	Mean Squares
British			
Lin. regression ..	1	15423	
Deviations from regression	16	1130	70·63
Within arrays	2868	65563	22·86
TOTAL ..	2885	82116	
Taxila			
Lin. regression ..	1	20712	
Deviations from regression	9	2389	265·44
Within arrays	742	73502	99·58
TOTAL ..	752	96603	

(Figures rounded off from machine calculations.)

This increases the resemblance between the two coinages, as we see that the regression is very highly significant in both cases, and that the deviations from regression are just over the ·1 p. c. level of significance, when tested against the estimated variance within arrays, by Fisher's z test. But the story changes if we fit quadratic regressions. For British coins, the deviations from a quadratic regression have the sum-square 1092, which with the loss of one degree of freedom actually makes the deviations a little more significant; in the Taxilan case, we have the sum-square reduced to 938, which gives a mean square for deviations 117·31, quite insignificant. So, it is clear that no regression could fit appreciably better than the quadratic for Taxila, while the British deviations are due to other causes than a non-constant rate of wear with time. Both of these are to be expected, inasmuch as for all times, the tendency to get rid of a "bad" or worn coin heightens the rate of wear with age; and for Taxila ten reverse marks cover something like 120 years so that one need not expect the same rate of wear to apply throughout the period. For the British currency, the war period 1914–18 was one of absorption of coin, which flooded the market after the war and caused a stagnation of the 1918–20 issues. Besides, the time of sampling (August 1940) was one of extreme currency panic. Finally the sample taken at Poona can hardly be called truly representative of the vast numbers minted—at varying rates—and issued according to the needs of various parts of the country. That is, the British deviations are due in my opinion primarily to inefficient sampling and irregularity of the actual date of issue of the currency in question.

Perhaps the most curious feature of the investigation was the fact that mean values for both coinages lie practically on a straight line, when the graph is drawn. But the correlations are comparatively low. I mean to show here that this is a feature common to time series in general, where the correlation as calculated from the usual formula must necessarily be an underestimate of the population value. For the case in hand, theory proceeds on the assumption that the minted weights show a normal distribution, and that the loss due to wear is also normal. With these (or slightly more general) assumptions, we are led to what is known as the homogeneous random process, which is fundamental in the flow of heat, diffusion, the kinetic theory of gases, Brownian movements, the theory of speculation, and certain actuarial phenomena. For our purpose, it is enough to deduce that the means are regularly depressed with age, the variance increased; both obeying equations linear in time. There is an additional factor for absorption of currency, of type exp-bt, but this does not affect weight distribution within an array. It is also seen that the numbers in one array are independent of those in another unless the rate of issue is constant and that of absorption is known to be exact. All this does not give us a population in bivariate normal correlation. For a population of this latter type, one finds (in representative samples) the entries vanishing outside of an elliptical region of the tables, and the numbers thinning out towards the boundary of this ellipse; and this is theoretically to be expected. In a time series in general, there is no reason for this to happen. In fact, for a time series, the variance of the time is usually infinite, the ellipse of error being then drawn out into two straight lines. All the arrays, for such a time series, taken together only amount to a very thin slice taken from near the centre of a proper distribution in bivariate normal correlation.

One should not be surprised, therefore, if the calculation of r by the usual formula leads to something entirely different in the case in hand. To take a theoretically perfect example, let there be $m+1$ arrays labelled 0, 1, ..., m; in the pth array, let the number of specimens be n_p, their mean weight $a - bp$, the sum of squares of the deviations from the array mean $n_p(u^2 + pv^2)$. The average weights then lie exactly on a straight line, and the population correlation should be unity. But calculating by the usual formula we obtain

$$r^2 = \frac{b^2(\beta - \alpha^2)}{u^2 + \alpha v^2 + b^2(\beta - \alpha^2)}$$

where $\mathrm{N} = \sum_0^m n_p$; $\frac{1}{\mathrm{N}} \sum_0^m p n_p = \alpha$;

$$\frac{1}{\mathrm{N}} \sum_0^m p^2 \mathrm{N}_p = \beta.$$

Moreover, the sum of squares within arrays is $\mathrm{N}(u^2 + \alpha v^2)$, that between arrays being $\mathrm{N}b^2(\beta - \alpha^2)$; so, in place of the correlation coefficient r, we have actually obtained the correlation ratio η and unless the variation vanishes in each array (which is theoretically impossible) r^2 (here η^2) is always less than the population value (here unity), which can be approximated only by increasing the number of arrays indefinitely, not by merely taking more and more coins in a finite number of arrays. In fact that latter process leads to a quantity distributed not like the square of the correlation coefficient, but asymptotically like χ^2/N, which function has also been proposed as a measure of the correlation in place of r, or η.

In practice, the r^2 calculated as above amounts to taking the ratio of the sum of squares due to regression to the total sum of squares. In its place, *I suggest that for time series the sum of squares due to regression be divided by the total sum of squares between arrays for an estimate of ρ^2 which amounts to calculating the correlation coefficient from the weighted array means.* This is a better estimate of the "population value", and the degrees of freedom are now based on the number of arrays alone. These "adjusted" correlations are, for Taxila $\bar{r} = \cdot 946$; rupees $\bar{r} = \cdot 965$. Of course, this should be applied only to time series as such, in which it is known that the time variate has not a finite variance, and does not yield a population in bivariate normal correlation.

Tests of significance by analysis of variance might be justified in all cases. But those who insist upon the validity of the usual formula for r even in the time series would find it difficult to say just what population constant is estimated thereby. A population is said to be in bivariate normal correlation when its probability density is given by

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp-\frac{1}{2(1-\rho^2)}\times$$
$$\times\left\{\left(\frac{x}{\sigma_1}\right)^2+2\rho\frac{x}{\sigma_1}\frac{y}{\sigma_2}+\left(\frac{y}{\sigma_2}\right)^2\right\},$$

where s_1, s_2, r are estimates of σ_1, σ_2, ρ. Taking, for simplicity, $\sigma_1 = \sigma_2 = \sigma$, the axes of the error ellipse are found to be proportional to $\sigma\sqrt{(1+\rho)}$, $\sigma\sqrt{(1-\rho)}$. Making ρ tend to unity while σ approaches a finite limit means letting the ellipse shrink down to one of its axes as a line segment. The bivariate population then degenerates into one with a single variate whose variance is easily found from the corresponding axis, while the other axis tends to zero length. But in order to represent the usual time series, the ellipse must degenerate in other ways, a simple example being $\rho \rightarrow 1$, $\sigma\sqrt{(1-\rho)} \rightarrow a$. Here, one of the two axes becomes infinite, the other remaining finite. The ellipse is then stretched out into two parallel lines. Without an entirely new definition, the "population correlation" here can only be taken as unity. Attaching the usual meaning to the r formula is, therefore, now out of the question.

COMMENSALISM IN SPONGES*

BY

D. W. DEVANESEN AND P. I. CHACKO

(*Department of Industries, Madras*)

WELL-KNOWN examples of commensalism in the animal kingdom are found between crabs and sea-anemones. The sea-anemone *Adamsia* lives in association with a hermit-crab. The crab *Dorippe* carries a sea-anemone on the top of a bivalve shell which is mounted on its back and held in position by its hind pair of legs. But among the Krusadai littoral fauna are found instances of commensalism in siliceous sponges which being extraordinary deserve the special notice of naturalists.

(1) The sponge *Spirastrella inconstans* (Dendy) has imbedded in the outer portion of its body numerous cirripedes of the species *Balanus longirostrum* (Hoek). The sponge belongs to the family *Clavulidæ* of the order *Tetraxonida.* It is common all round the island, especially on the southwestern side. The sponge is composed of a bunch of stout, erect, digitate processes springing from a basal mass. Its colour is light brown; and it is often washed ashore. The cirripedes evidently draw their supply of food through the current of water set up by the choanoflagellate cells of the sponge. The cirripedes have therefore to expend little or no energy in producing the current. In return the sponge probably gains mechanical support by the inclusion of the exoskeleton of the cirripedes. The sponge may also help itself to surplus food-material broken by the cirripedes into finely divided grains. The number of barnacles in a sponge is very variable; but on an average there are fifteen barnacles to thirty-five grammes of the sponge, thus showing that the barnacles are rather sparsely distributed.

(2) The sponge *Adocia dendyi* (Burton)[1] is another example; but here the commensal is an alga, *Ceratodictyon spongiosum* (Zanard).[2] Further, as this is an intimate association between an animal and a plant, it is an example of symbiosis. For the symbiotic life sunlight is necessary. As the host occupies shallow flats, sunlight can reach the alga and photosynthesis is possible, the alga liberating oxygen for the choanoflagellates from the carbondioxide supplied by the latter. The sponge belongs to the family *Haplosceridæ*, of the order *Tetraxonida.* The sponge is found all round the island within the one-fathom zone, and is frequently washed ashore by waves. The alga belongs to the family *Gracilariaceæ* of the group *Rhodophyceæ.* The sponge when fresh is light green in colour. In this case also, the sponge does derive some rigidity by the presence of the branching alga.

---

* Published with the permission of the Director of Industries and Commerce, Madras.

[1] The sponge was identified Dr. by M. Burton, D.Sc., of the British Museum, London.

[2] The alga was identified by Prof. M. O. Parthasarathy Iyengar, University of Madras.