LETTERS TO THE EDITOR.

A General Coefficient of Statistical Relationship.

Suppose two variables x and y are related by the equation $x = ky + z$, where k is a constant and z another variable which is statistically independent of y. Then the relationship between x and y is usually expressed by means of the correlation coefficient between them. This is not a general coefficient in that the test of significance of the correlation coefficient calculated from a sample is not accurate unless both x and y are normally distributed. The coefficient described below is quite general in this sense and hence is better than the correlation coefficient.

Relation-Coefficient.—In the x, y population let us consider first the values of x only. Let the median for the x population be denoted by $म_x$, see (1). We subtract this value from each of the x values, that is to say, we replace the x values by their departures from $म_x$. Similarly we replace the y values by their departures from $म_y$. Each individual in the x, y population will now consist of two departures. We count the total number, n_1, of individuals having departures of the same sign, each individual containing one or both departures zero being counted as ½. Similarly we count the total number, n_2, of individuals having departures of opposite sign. If n be the total number of individuals in the whole population, then a general coefficient of statistical relationship, which may be termed "relation-coefficient" and denoted by ऋ,* is given by

$$ऋ = \frac{n_1 - n_2}{n} \qquad (1)$$

n is, of course, equal to $n_1 + n_2$.

It is easy to see that ऋ may have any value from $+1$ to -1, and is zero when the variables are statistically independent of each other. In this respect it resembles the correlation coefficient.

In the case of a sample, we use $म_x'$ and $म_y'$ the medians of the x and y values in the sample, see (1), and proceed as above. We shall denote the relation-coefficient in a sample by ऋ'.

A Test of Significance for ऋ.—Suppose our sample has n pairs of values $x_1y_1, x_2y_2, \cdots$ and x_ny_n. ऋ' is easily determined. To test the significance of ऋ' we proceed as follows. Let us consider the values $x_1, x_2 \cdots x_n$. Suppose we use P as our limit for random chance. We now find the limits for $म_x$ using $\sqrt{P}$ as the limit for random chance as explained in (1). Let the limits be $म_x^{(1)}$ and $म_x^{(2)}$ in the order of ascending magnitude. Similarly, on the same limit $\sqrt{P}$ we find the limits $म_y^{(1)}$ and $म_y^{(2)}$ for $म_y$. Let $म_y^{(1)} < म_y^{(2)}$.

Arguing in a manner similar to that given in (1) we see that when x and y are statistically independent of each other and only when $म_x^{(1)} \leqslant म_x \leqslant म_x^{(2)}$ and $म_y^{(1)} \leqslant म_y \leqslant म_y^{(2)}$ our sample could have been obtained from the x, y population by random chance for which the limit is P.

We use $म_x^{(1)}$ and $म_y^{(1)}$ and find the departures in our sample. It is sufficient if

* ऋ is a Sanskrit letter pronounced like "ru" in "run", but with the vowel sound slightly more prolonged.

only the signs are given. We now calculate र from equation I. Using $\bar{म}_y^{(2)}$ instead of $म_y^{(1)}$ we get another value for र. Similarly with $म_x^{(2)}$ and $म_y^{(1)}$ and $म_y^{(2)}$ separately we get two more values of र. We now take the highest and the lowest of these four values. Let $र_1$ and $र_2$ be these values. We will call the interval from $र_1$ to $र_2$ (including the end values) the P interval for र. In a similar manner we can obtain an interval for र on any other limit for random chance.

A General Test of Significance of र′.—Our test of significance may now be stated thus:

Using some limit for random chance we calculate the interval for र as explained above. If this interval does not contain zero, र′ is significant, and if zero be an end value of this interval र′ may be considered to be just significant.

It is easy to see that this test is quite general, that is, it is applicable to all samples irrespective of the frequency distributions in the populations, from which the samples were obtained. Details will be published elsewhere.

Poona 5,
July 8, 1937.

S. R. SAVUR.

[1] S. R. Savur, *Proc. Ind. Acad. Sci.*, **1937, 5, 564.**

The Biogenesis of the Terpenes.

ADOPTING the view that, in general, the widely distributed terpenes are only indicative of the ease with which the biological