NOTE ON CORRELATIONS

By D. B. De Lury

When the value of a correlation coefficient is to be estimated from a set of N pairs of observations, (x_i, y_i), $i = 1, 2, \cdots N$, the statistic ordinarily computed is, of course, the product-moment correlation coefficient,

$$r = s_{12}/(s_1 s_2), \quad \text{where}$$

$$ns_1^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2, \qquad ns_2^2 = \sum_{i=1}^{N} (y_i - \bar{y})^2, \qquad ns_{12} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}),$$

$$N\bar{x} = \sum_{i=1}^{N} x_i, \qquad N\bar{y} = \sum_{i=1}^{N} y_i, \qquad n = N - 1.$$

However, when x and y are known to have the same population mean and variance, the precision of the estimate may be improved slightly by using the intraclass correlation coefficient,

$$r' = \frac{2\sum_{1}^{N} (x_i - \xi)(y_i - \xi)}{\sum_{1}^{N} \{(x_i - \xi)^2 + (y_i - \xi)^2\}}, \qquad 2N\xi = \sum_{1}^{N} (x_i + y_i).$$

It may be of interest to inquire into the properties of an analogous coefficient, appropriate to the case of equal variances and different means. This coefficient would naturally be chosen to be

$$u = 2s_{12}/(s_1^2 + s_2^2) = \{2s_1 s_2/(s_1^2 + s_2^2)\}r.$$

Obviously, $| u | \leq | r |$.

The probability distribution of u is easily determined, under the assumption that x and y obey a bivariate normal distribution. If σ^2 is their common variance, no restriction is introduced by taking $\sigma = 1$. Then the probability element of s_1, s_2, r, is known to be[1]

$$\frac{n^n}{\pi(n-2)!(1-\rho^2)^{n/2}} (s_1 s_2)^{n-1} e^{-\frac{n}{2(1-\rho^2)}(s_1^2 - 2\rho r s_1 s_2 + s_2^2)} (1 - r^2)^{\frac{n-3}{2}} ds_1 ds_2 dr,$$

where ρ is the correlation of x and y. From this, the distribution of u can be obtained by making the transformation

$$u = \{2s_1 s_2/(s_1^2 + s_2^2)\}r, \qquad v = 2s_1 s_2/(s_1^2 + s_2^2), \qquad w = s_1^2 + s_2^2.$$

[1] R. A. Fisher, Biometrika, Vol. 10, p. 510.