A THEOREM ON THE CORRELATION COEFFICIENT FOR SAMPLES OF THREE WHEN THE VARIABLES ARE INDEPENDENT

By C. Chandra Sekar[1]

All India Institute of Hygiene and Public Health, Calcutta

In this note the following theorem will be established:

Theorem. *If* (x_i, y_i) *for* $i = 1, 2$ *and* 3 *denote three pairs of random values of two independent continuous stochastic variables* x *and* y, r, *their correlation coefficient, is given by*

$$(1) \qquad r = \frac{1}{3s_x s_y}\sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y}),$$

where

$$(2) \qquad \bar{x} = \frac{1}{3}\sum_{i=1}^{3}x_i, \qquad \bar{y} = \frac{1}{3}\sum_{i=1}^{3}y_i,$$
$$s_x^2 = \frac{1}{3}\sum_{i=1}^{3}(x_i - \bar{x})^2, \qquad s_y^2 = \frac{1}{3}\sum_{i=1}^{3}(y_i - \bar{y})^2,$$

and $P(a \leq r \leq b)$ *denotes the probability of* r *taking values in the range* $a \leq r \leq b$, *then*

$$(3) \qquad P\left(-1 \leq r \leq -\frac{1}{2}\right) = P\left(-\frac{1}{2} \leq r \leq \frac{1}{2}\right) = P\left(\frac{1}{2} \leq r \leq 1\right) = \frac{1}{3}.$$

Proof. If τ_i is defined by

$$(4) \qquad \tau_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, 3,$$

it is readily seen that the three values of τ are connected by the two relations

$$(5) \qquad \sum_{i=1}^{3}\tau_i = 0, \qquad \sum_{i=1}^{3}\tau_i^2 = 3.$$

Similar conditions exist between the three t_i's defined by

$$(6) \qquad t_i = \frac{y_i - \bar{y}}{s_y}, \qquad i = 1, 2, 3.$$

The set (τ_1, τ_2, τ_3) can be considered as the Cartesian coordinates of a point in three dimensional space. The conditions (5) restrict the point to a circle. The set (t_1, t_2, t_3) defined by (6) represents a point on the same circle. The correlation coefficient, r, defined in (1) and also given by

$$(7) \qquad r = \tfrac{1}{3}\sum_{i=1}^{3}\tau_i t_i$$

[1] On loan to Population Division, United Nations.