

# THE PSYCHOLOGICAL BULLETIN

---

## A METHOD OF CALCULATING THE PEARSON COEFFICIENT OF CORRELATION WITHOUT THE USE OF DEVIATIONS OR CROSS MULTIPLYING<sup>1</sup>

BY J. CROSBY CHAPMAN

*Yale University*

The cross multiplication of variables in determining the coefficient of correlation is a process in which owing to the variability of sign there is great chance for error. In addition the tables for calculating products of two different numbers are much more laborious to use and more difficult to procure than the tables giving squares. In the following method of determining the correlation the product sum process is avoided by substituting a process in which the tables giving squares can be employed.

Suppose the two variables to be  $X_2$  and  $X_1$ , and let their deviations be as usual  $x_2$  and  $x_1$ ; also let it be agreed to write the sum of the variables ( $x_1 + x_2$ ) as  $x_{1+2}$ .

Now

$$\Sigma x_{1+2}^2 = \Sigma x_1^2 + \Sigma x_2^2 + 2\Sigma x_1x_2.$$

Substituting for  $\Sigma x_1x_2$

$$r = \frac{\Sigma x_{1+2}^2 - \Sigma x_1^2 - \Sigma x_2^2}{2 \sqrt{\Sigma x_1^2} \sqrt{\Sigma x_2^2}}.$$

Replacing deviations from means by deviations from zero.

$$r = \frac{\Sigma X_{1+2}^2 - \Sigma X_1^2 - \Sigma X_2^2 - \frac{1}{n} \{ (\Sigma X_{1+2})^2 - (\Sigma X_1)^2 - (\Sigma X_2)^2 \}}{2 \sqrt{\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n}} \sqrt{\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n}}}.$$

<sup>1</sup> See *A Method of Calculating the Pearson Coefficient of Correlation Without the Use of Deviations*, by L. L. THURSTONE, PSYCHOL. BULL., June 15, 1917.

To determine  $r$  the following terms must be calculated?—

$$\begin{array}{lll} \Sigma X_1, & \Sigma X_2, & \Sigma X_{1+2}, \\ (\Sigma X_1)^2, & (\Sigma X_2)^2, & (\Sigma X_{1+2})^2, \\ \Sigma X_1^2, & \Sigma X_2^2, & \Sigma X_{1+2}^2. \end{array}$$

Even though the elimination of the cross multiplication introduces higher figures in the terms  $X_{1+2}$  in many cases where the variables are reasonably small and the number of cases are not great, this is more than compensated for by the ease with which squares can be obtained from tables and the simplicity of adding operations with the help of a machine. Furthermore in constructing a machine for correlation the absence of cross multiplication is most desirable.