Note by the Editor

One word of caution is necessary. In testing the significance of the regression coefficient we are substituting an estimate of the variance of y based on the sample itself for the population value of the variance. This introduces an uncertainty which increases rapidly as the size of the sample is decreased. It is desirable to remember that the use of the exact expressions developed in recent years for small samples cannot get over this particular difficulty. Great caution is therefore needed in interpreting the significance of regression coefficients based on small samples.

The situation of course is entirely different when the population values of the standard deviations and the correlation coefficient may be considered as known. In this case the exact distribution of the regression coefficient '*b*' may be used with safety to test whether the sample values are in reasonable agreement with the population value.

If ρ is the correlation, and σ_y and σ_x the standard deviations in the parent population, then the population value of the regression is $\beta=\rho\cdot\sigma_y/\sigma_x$, the variance of '*b*' the regression in samples of size N is $\sigma_b^2=\sigma_y^2(1-\rho^2/\sigma_x^2\ (N-3))$, and the distribution of '*b*' is given by

$$z=\frac{z_0}{\left\{\frac{\sigma_y^2(1-\rho^2)}{\sigma_x^2}+(b-\beta)^2\right\}^{+\frac{1}{2}N}}$$

The use of the above formula has been fully explained by Prof. Karl Pearson in Biometric Tables, Part II, Table XXV, pp. cxxvi-cxxxi. It is useful to refer to the following papers in this connexion.

Karl Pearson: "Further Contributions to the Theory of Small Samples (*Biometrika*, Vol. 17, 1925, pp. 176-199).

U. Romanowaky: "On the moments of standard derivations and correlation coefficients in samples from normal population" (*Metron*, Vol. 5, 1925, pp. 3-46).

R. A. Fisher: "The Goodness of Ft of Regression Formulæ and the Distribution of Regression Co-efficients" (*Jour. Roy. Stat. Soc.* Vol. LXXXV, July, 1922, pp. 597-612).

For large samples the relevant formulæ are given by Karl Pearson. "On the Probable Error of Frequency Constants, Part II" (*Biometrika*, Vol. 9, 1913, pp. 1-10).

It is worth pointing out that in Pearson's formula quoted in this note, σ_x^2, σ_y^2 and ρ represent population values of the two variances and the coefficient of correlation, while in Fisher's formula used in the text of the paper, s_x^2, s_y^2 and r represent the observed values in the sample. In fact Pearson's value $\sigma_y^2(1-\rho^2)/\sigma_x^2(N-3)$ is equal to the mathematical expectation or population value of Fisher's expression $s_y^2(1-r^2)/s_x^2(N-2)$.

P. C. M.

(March, 1934).