A NOTE ON THE MATHEMATICAL EXPECTATION OF THE VARIANCE OF THE REGRESSION COEFFICIENT

By SUBHENDUSEKHAR BOSE,

STATISTICAL LABORATORY, PRESIDENCY COLLEGE, CALCUTTA.

The following expression* for the variance of regression coefficient ($b = r \cdot s_1/s_2$) was given by R. A. Fisher ("The Goodness of Fit of Regression Formulæ and the Distribution of Regression Coefficients," *Jour. Roy. Stat. Soc.*, Vol. 75, Part IV, July, 1922, p. 608).

$$s_b^2 = \frac{S(y-Y)^2}{(n-2)} \cdot \frac{1}{S(x-\bar{x})^2} \quad \ldots \quad \ldots \quad (1)$$

where y, x, are the correlated variates, $\bar{x}$ the mean value of x; Y the value of y estimated from the regression equation $y = a + b(x - \bar{x})$ and n is the size of the sample.

$$S(y-Y)^2 = (1-r^2) . S(y-\bar{y})^2 \quad \ldots \quad \ldots \quad (1.1)$$

$$\frac{S(y-\bar{y})^2}{S(x-\bar{x})^2} = \frac{s_1^2}{s_2^2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (1.2)$$

$$s_b^2 = \frac{(1-r^2)}{(n-2)} \frac{s_1^2}{s_2^2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (1.3)$$

where s_1, s_2, are the sample values of the standard deviations of y and x respectively.

Karl Pearson gave the population value of the variance of the regression coefficient in terms of the population parameters ("Further Contributions to the Theory of Small Samples," *Biometrika*, Vol. 17, 1925, p. 196).

$$\sigma_b^2 = \frac{(1-\rho^2)}{(n-3)} \frac{\sigma_1^2}{\sigma_2^2} \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

where σ_1^2, σ_2^2 are the population values of the variances of y and x, and ρ is the population value of the coefficient of correlation.

P. C. Mahalanobis in his Editorial Note on my "Tables for Testing the Significance of Linear Regression in the case of Time Series and other Single-Valued Samples" (*Sankhyā*, Vol. 1, Parts 2 and 3, August, 1934, p. 284) had stated that σ_b^2 (Pearson's

*The notations used in the original papers cited here have been changed considerably in many cases.

expression) is the mathematical expectation of s_b^2 (Fisher's expression). That is, using E as a symbol for mathematical expectation,

$$E\left(\frac{(1-r^2)\, s_1^2}{(n-2)\, s_2^2} \right) = \frac{(1-\rho^2)\, \sigma_1^2}{(n-3)\, \sigma_2^2} \quad \dots \quad \dots \quad \dots \quad (3)$$

A proof of this result is given below. The mathematical expectation of s_b^2 can be directly evaluated:—

$$E\left(\frac{(1-r^2)\, s_1^2}{(n-2)\, s_2^2} \right) = \frac{\int_0^\infty \int_0^\infty \int_{-1}^{+1} \frac{(1-r^2).\, s_1^2}{(n-2).\, s_2^2} .\ \phi(r, s_1, s_2)\, dr.\, ds_1.\, ds_2}{\int_0^\infty \int_0^\infty \int_{-1}^{+1} \phi(r, s_1, s_2).\, dr.\, ds_1.\, ds_2} \quad (4)$$

where $\phi(r, s_1, s_2)$ is the joint distribution of the two standard deviations and the coefficient of correlation.

For a bi-variate normal population wth standard deviations σ_1, σ_2 and coefficient of correlation ρ. R. A. Fisher ("Frequency Distributions of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," *Biometrika*, Vol. 10, 1915, p. 510) has shown this distribution to be

$$\phi(r, s_1, s_2) = Z_0.\ e^{-\frac{s_1^2}{g_1^2}} .\ e^{-\frac{s_2^2}{g_2^2}} .\left(\frac{s_1\ s_2}{g_1\, g_2} \right)^{n-2} .\ e^{hr} .\ (1-r^2)^{\frac{n-4}{2}} \quad \dots \quad (5)$$

where z_0 is a constant, and

$$g_1^2 = \sigma_1^2\,(1-\rho^2)/n \ \dots \quad \dots \quad \dots \quad \dots \quad (5.1)$$

$$g_2^2 = \sigma_2^2\,(1-\rho^2)/n \ \dots \quad \dots \quad \dots \quad \dots \quad (5.2)$$

$$h = \rho.\, s_1\, s_2 / g_1\, g_2 \dots \quad \dots \quad \dots \quad \dots \quad (5.3)$$

we first require

$$I_1 = \int_0^\infty \int_0^\infty \int_{-1}^{+1} (1-r^2)\, s_1^2\, s_2^{-2}\, \phi(r, s_1, s_2)\, dr.\, ds_1.\, ds_2 \quad \dots \quad \dots \quad (6)$$

Pearson has given in the paper already cited (*Biometrika*, Vol. 17, p. 195) a very general expression for integrals of the following form:—

$$\int_0^\infty \int_0^\infty \int_{-1}^{+1} r^p .\ (1-r^2)^{t/2} .\ s_1^{q_1} .\ s_2^{q_2} .\ \phi(r, s_1, s_2)\, dr.\, ds_1.\, ds_2 \quad \dots \quad (6.1)$$

Comparing equations (6) and (6·1), in Pearson's notation we have $p=0$, $l=2$, $q_1=2$, $q_2=-2$, Substituting these values in Pearson's equation on p. 195 we get

$$I_1 = Z_0 . \frac{'g_1^2}{g_2} \sqrt{\pi} . \Gamma \frac{n}{2} . 2^{n-3} . \Gamma\left(\frac{n-3}{2}\right) . \left(1 + \frac{n-3}{2} . \frac{\rho^2}{1!} + \frac{n-3}{2} \frac{n-1}{2} . \frac{\rho^4}{2!} \right)$$

$$= Z_0 . \frac{'g_1^2}{g_2} . \sqrt{\pi} . \Gamma\left(\frac{n}{2}\right) . 2^{n-3} . \Gamma\left(\frac{n-3}{2}\right) . (1-\rho^2)^{-(n-1)/2} \qquad (6{\cdot}2)$$

Pearson has also given the value of the above integral in equation (xx), p. 184 of the paper already cited. In our notation,

$$I_2 = \int_0^\infty \int_0^\infty \int_{-1}^{+1} \phi (r, s_1, s_2) \; dr . ds_1 . ds_2 \quad \ldots \quad \ldots \quad \ldots \qquad (7)$$

$$= Z_0 . 'g_1 \, 'g_2 \sqrt{\pi} . 2^{n-3} . \Gamma\left(\frac{n-2}{2}\right) . \Gamma\left(\frac{n-1}{2}\right) . (1-\rho^2)^{-(n-1)/2} \quad \ldots \qquad (7.1)$$

We thus find

$$E\left(\frac{(1-r^2)\, s_1^2}{(n-2)\; s_1^2}\right) = \frac{I_1}{n-2} . \frac{1}{I_2}$$

$$= \frac{Z_0 . 'g_1^2 . 'g_2^{-1} . \sqrt{\pi} . \Gamma \frac{n}{2} . \Gamma\left(\frac{n-3}{2}\right) . 2^{n-3} (1-\rho^2)^{-(n-1)/2}}{(n-2) . Z_0 . 'g_1 . 'g_2 . \sqrt{\pi} . 2^{n-3} . \Gamma\left(\frac{n-2}{2}\right) . \Gamma \frac{n-1}{2} . (1-\rho^2)^{-(n-1)/2}}$$

$$= \frac{'g_1^2}{'g_2^2} . \frac{(1-\rho^2)}{(n-3)} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (8)$$

Since $\quad 'g_1^2/g_2^2 = \sigma_1^2/\sigma_1^2 \quad \ldots \quad \ldots \quad \ldots$ from (5·1) and (5·2)

we get finally

$$E \frac{(1-r^2)s_1^2}{(n-2)s_2^2} = \frac{(1-\rho^2)\sigma_1^2}{(n-3)\; \sigma_2^2} \quad \ldots \quad \ldots \quad \ldots \qquad (9)$$

which is the required result.

(Calcutta, July, 1934.)