THE STANDARD ERROR OF A MULTIPLE REGRESSION EQUATION[1]

By

JOHN RICE MINER

Since a multiple regression equation is essentially a hyperplane, fitted by the method of least squares, its standard error may be obtained from Gauss' *standard error of a function* recently discussed by Schultz (1930). Let the equation be

$$x_1 = b_{12.34\cdots m}x_2 + b_{13.24\cdots m}x_3 + \cdots + b_{1m.23\cdots(m-1)}x_m$$

where x_1 is the dependent variable, $x_2, x_3 \cdots\cdots x_m$ the independent variables, each measured from its respective mean, and $b_{12.34\cdots m}, \cdots, b_{1m.23\cdots(m-1)}$ the partial regression coefficients. Then the determinant of Schultz's equation (10) becomes

$$D = \begin{vmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & \Sigma x_2^2 & \Sigma x_2 x_3 & \cdots & \Sigma x_2 x_m \\ 0 & \Sigma x_2 x_3 & \Sigma x_3^2 & \cdots & \Sigma x_3 x_m \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & \Sigma x_2 x_m & \Sigma x_3 x_m & \cdots & \Sigma x_m^2 \end{vmatrix} = n^m \sigma_2^2 \sigma_3^2 \cdots \sigma_m^2$$

(1)

$$\begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & r_{23} & \cdots & r_{2m} \\ 0 & r_{23} & 1 & \cdots & r_{3m} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & r_{2m} & r_{3m} & \cdots & 1 \end{vmatrix} = n^m \sigma_2^2 \sigma_3^2 \cdots \sigma_m^2 \Delta, \text{sau}$$

[1]From the Department of Biology of the School of Hygiene and Public Health of the Johns Hopkins University.

Let Δ_{ij} be the cofactor of the element in the i 'th row and the j 'th column. Then

$$[\alpha\alpha] = D_{11}/D = \frac{1}{n},$$

$$[\beta\beta] = D_{22}/D = \Delta_{22}/n\sigma_2^2\Delta,$$

. . . .

$$[\mu\mu] = D_{mm}/D = \Delta_{mm}/n\sigma_m^2\Delta,$$

$$[\alpha\beta] = D_{12}/D = 0,$$

. . . .

$$[\alpha\mu] = D_{1m}/D = 0,$$

$$[\beta\gamma] = D_{23}/D = \Delta_{23}/n\sigma_2\sigma_3\Delta,$$

. . . .

$$[\beta\mu] = D_{2m}/D = \Delta_{2m}/n\sigma_2\sigma_m\Delta,$$

. . . .

$$\frac{\partial f}{\partial A} = 1, \ \frac{\partial f}{\partial B} = x_2, \ \frac{\partial f}{\partial C} = x_3, \ \ldots, \frac{\partial f}{\partial M} = x_m, \qquad \text{and}$$

$$\varepsilon^2 = \frac{n}{n-m}\,\sigma_1^2\left(1 - R_{1(23\ldots m)}^2\right).$$

Therefore, substituting these values in Schultz's equation (27.1), we have

$$\sigma_f = \frac{\sigma_1}{(n-m)^{\frac{1}{2}}}\left(1-R^2_{1(23\cdots m)}\right)^{\frac{1}{2}}\left\{1+\frac{\Delta_{22}}{\sigma_2^2\Delta}x_2^2+\frac{\Delta_{33}}{\sigma_3^2\Delta}x_3^2+\cdots+\frac{\Delta_{mm}}{\sigma_m^2\Delta}x_m^2\right.$$

(2)

$$\left.+2\frac{\Delta_{23}}{\sigma_2\sigma_3\Delta}x_2x_3+\cdots+2\frac{\Delta_{2m}}{\sigma_2\sigma_m\Delta}x_2x_m+\cdots\right\}^{\frac{1}{2}}.$$

For a simple regression equation this reduces to

$$(3)\quad \sigma_f = \frac{\sigma_1}{(n-2)^{\frac{1}{2}}}\left(1-r_{12}^2\right)^{\frac{1}{2}}\left\{1+\frac{x_2^2}{\sigma_2^2}\right\}^{\frac{1}{2}}.$$

This agrees with the expression given by Pearson (1913), if we remember that x_2 is measured from its mean and that Pearson does not correct for the number of parameters.

For a regression equation with two independent variables

$$\sigma_f = \frac{\sigma_1}{(n-3)^{\frac{1}{2}}}\left(1-R^2_{1(23)}\right)^{\frac{1}{2}}$$

$$(4)\quad \left\{1+\frac{x_2^2}{\sigma_2^2(1-r_{23}^2)}+\frac{x_3^2}{\sigma_3^2(1-r_{23}^2)}-\frac{2r_{23}x_2x_3}{\sigma_2\sigma_3(1-r_{23}^2)}\right\}^{\frac{1}{2}}$$

$$= \frac{\sigma_{1.23}}{(n-3)^{\frac{1}{2}}}\left\{1+\frac{x_2^2}{\sigma_{2.3}^2}+\frac{x_3^2}{\sigma_{3.2}^2}-\frac{2r_{23}x_2x_3}{\sigma_{2.3}\sigma_{3.2}}\right\}^{\frac{1}{2}}$$

As an example of the application of this formula we may calculate the standard error of the mean heart-weight (X_1) of the array of persons with an age (X_2) of 52.92 years and a body-weight (X_3) of 49.93 kilograms in a population of 213 persons characterized by the following biometric constants:

$$M_1 = 348.9 \text{ g}; \qquad \sigma_1 = 79.4 \text{ g}; \qquad r_{12} = +0.114$$
$$M_2 = 59.65 \text{ yrs.}; \qquad \sigma_2 = 17.54 \text{ yrs.}; \qquad r_{13} = +0.652$$
$$M_3 = 56.45 \text{ kg}; \qquad \sigma_3 = 14.38 \text{ kg}; \qquad r_{23} = -0.185\,.$$

From these data $r_{12.3} = +0.315$ and $r_{13.2} = +0.689$ and the regression equation of heart-weight on age and body-weight is

$$X_1 = 66.09 + 1.100 X_2 + 3.848 X_3$$

from which the mean heart-weight of persons aged 52.92 years and weighing 49.93 kg. is 316.4 g.

Substituting the appropriate values of the constants in (4) and remembering that $x_2 = X_2 - M_2 = -6.72$, $x_3 = X_3 - M_3 = -6.52$, and

$$(1-R^2_{1(23)})^{\frac{1}{2}} = (1-r^2_{12})^{\frac{1}{2}}(1-r^2_{13.2})^{\frac{1}{2}},$$

$$\sigma_f = \frac{79.4}{210^{\frac{1}{2}}} (0.993)(0.725)\left\{1 + \frac{(-6.72)^2}{(17.54)^2(0.966)} + \frac{(-6.52)^2}{(14.38)^2(0.966)} - \frac{2(-0.185)(-6.72)(-6.52)}{(17.54)(14.38)(0.966)}\right\}^{\frac{1}{2}} = 4.7 \text{ g}\,.$$

John Rice Miner

REFERENCES

Pearson, Karl. 1913. On the probable errors of frequency constants. Biometrika, 9:1-10.

Schultz, Henry. 1930. The standard error of a forecast from a curve. J. Amer. Stat. Assoc., 25:139-185.