NOTE ON A FORMULA FOR THE MULTIPLE CORRELATION COEFFICIENT

By H. M. Bacon

There are many useful formulas available for the calculation of the multiple correlation coefficient in a k variable problem.[1] Since it frequently happens that the regression equation is the primary object of the statistical analysis, the well known formula

$$r^2_{1\cdot 23\cdots k} = \beta_{12\cdot 34\cdots k}\, r_{12} + \beta_{13\cdot 24\cdots k}\, r_{13} + \cdots + \beta_{1k\cdot 23\cdots(k-1)}\, r_{1k}$$

can be used to considerable advantage. While many different demonstrations of this formula are perfectly familiar, the one given in this note may prove of some interest.

First let us recapitulate briefly certain facts about the regression coefficients and the multiple correlation coefficient. Suppose we have k sets of N numbers each:

$$\begin{matrix} X_{11} & X_{12} & \cdot & \cdot & X_{1N} \\ X_{21} & X_{22} & \cdot & \cdot & X_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{k1} & X_{k2} & \cdot & \cdot & X_{kN}. \end{matrix}$$

Let $\bar{x}_j$ be the mean of the j-th set, and let $x_{ji} = X_{ji} - \bar{x}_j$. We then have k sets of N deviations from means, and we shall suppose the following k sets to be linearly independent:

$$\begin{matrix} x_{11} & x_{12} & \cdot & \cdot & x_{1N} \\ x_{21} & x_{22} & \cdot & \cdot & x_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{k1} & x_{k2} & \cdot & \cdot & x_{kN}. \end{matrix}$$

We shall consider only the regression of the "variable" x_1 upon $x_2, x_3, \cdots, x_k$. Clearly the results obtained can be made to describe the regression of any one of the variables upon the other $k - 1$ variables by rearranging the subscripts. As usual let $\lambda_2, \lambda_3, \cdots, \lambda_k$ have values which will make the sum of squares

$$F(\lambda_2, \lambda_3, \cdots, \lambda_k) = \Sigma(x_{1i} - \lambda_2 x_{2i} - \lambda_3 x_{3i} - \cdots - \lambda_k x_{ki})^2$$

a minimum. For simplicity we shall omit stating limits of summation and understand hereafter that Σ means "sum for i from $i = 1$ to $i = N$." Neces-

[1] For example, see W. J. Kirkham, "Note on the Derivation of the Multiple Correlation Coefficient", *The Annals of Mathematical Statistics*, Volume VIII (1937), pp. 68–71.