A NOTE ON THE DERIVATION OF FORMULAE FOR MULTIPLE AND PARTIAL CORRELATION*

By Louis Guttman

1. **Multiple Correlation.** Let the measurements of N individuals on each of the n variables $x_1, x_2, \cdots, x_k, \cdots, x_n$, be expressed as relative deviates; that is, such that

$$\Sigma x_k = 0, \qquad \Sigma x_k^2 = N, \qquad k = 1, 2, 3, \cdots, n,$$

where the summations extend over the N individuals.

If values of λ_k are determined so that

$$\Sigma(x_1 - \lambda_2 x_2 - \lambda_3 x_3 - \cdots - \lambda_n x_n)^2 \text{ is a minimum,}$$

and if we let

$$X_1 = \lambda_2 x_2 + \lambda_3 x_3 + \cdots + \lambda_n x_n, \tag{1}$$

then the multiple correlation coefficient, obtained from the regression of x_1 on the remaining $n - 1$ variables, is defined as

$$r_{1.234\cdots n} = r_{x_1 X_1}.$$

The square of the standard error of estimate of x_1 on the remaining $n - 1$ variables is defined as

$$\sigma^2_{1.234\cdots n} = \frac{1}{N}\Sigma(x_1 - X_1)^2.$$

The minimizing values for λ_k are obtained from the normal equations

$$\Sigma(x_1 - X_1)x_k = 0, \qquad k = 2, 3, \cdots, n. \tag{2}$$

which may be written in expanded notation as,

$$\begin{aligned}
\lambda_2 + r_{23}\lambda_3 + r_{24}\lambda_4 + \cdots + r_{2n}\lambda_n &= r_{12}\\
r_{32}\lambda_2 + \lambda_3 + r_{34}\lambda_4 + \cdots + r_{3n}\lambda_n &= r_{13}\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\cdots\\
r_{n2}\lambda_2 + r_{n3}\lambda_3 + r_{n4}\lambda_4 + \cdots + \lambda_n &= r_{1n}
\end{aligned}$$

where $r_{jk} = \frac{1}{N}\Sigma x_j x_k = r_{kj}$, $r_{jj} = 1$.

* The notions involved in this demonstration are certainly well-known. However, the directness and simplicity of the derivations may lend some merit to their exhibition. The writer is indebted to Professor Dunham Jackson for useful advice.

From Cramer's rule it is seen that

$$\lambda_k = -\frac{R_{1k}}{R_{11}}, \quad \text{if} \quad k \neq 1, \ R_{11} \neq 0,$$

where R_{jk} is the cofactor of r_{jk} (or of r_{kj}) in the symmetric determinant

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & r_{jk} & \cdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} \end{vmatrix}.$$

Summing both sides of (1) over the N individuals shows that $\Sigma X_1 = 0$, so that the variance of X_1 is

$$\sigma^2_{X_1} = \frac{1}{N} \Sigma X_1^2.$$

From (2), the residual $(x_1 - X_1)$ is orthogonal to each of the x_k except x_1; therefore the residual is orthogonal to any linear combination of these x_k and in particular to X_1; that is,

$$\Sigma(x_1 - X_1)X_1 = 0, \tag{3}$$

or

$$\sigma_{X_1} r_{x_1 X_1} = \sigma^2_{X_1}$$

and therefore

$$r_{x_1 X_1} = \sigma_{X_1}. \tag{4}$$

Multiplying both sides of (1) by $\frac{x_1}{N}$ and summing over the individuals, we get:

$$\begin{aligned} \sigma_{X_1} r_{x_1 X_1} &= r_{12}\lambda_2 + r_{13}\lambda_3 + \cdots + r_{1n}\lambda_n \\ &= -\frac{1}{R_{11}} (r_{12}R_{12} + r_{13}R_{13} + \cdots + r_{1n}R_{1n}) \\ &= 1 - \frac{R}{R_{11}}. \end{aligned}$$

From (4) then,

$$r^2_{1.234\cdots n} = 1 - \frac{R}{R_{11}}.$$

It is clear that in general

$$r^2_{k.123\cdots,k-1,k+1,\cdots n} = 1 - \frac{R}{R_{kk}}.$$

To find the standard error of estimate, expand

$$\frac{1}{N}\Sigma(x_1 - X_1)^2 = 1 - 2\sigma_{X_1} r_{x_1 X_1} + \sigma_{X_1}^2$$
$$= 1 - r_{x_1 X_1}^2$$
$$= \frac{R}{R_{11}}.$$

In general, when $\sigma_k = 1$,

$$\sigma_{k.123\cdots,k-1,k+1,\cdots n}^2 = \frac{R}{R_{kk}}. \tag{5}$$

2. **Partial Correlation.** If values of μ_k and ν_k are determined so that

$$\Sigma(x_1 - \mu_3 x_3 - \mu_4 x_4 - \cdots - \mu_n x_n)^2 \text{ is a minimum}$$

and

$$\Sigma(x_2 - \nu_3 x_3 - \nu_4 x_4 - \cdots - \nu_n x_n)^2 \text{ is a minimum,}$$

and if we let

$$\begin{aligned} Y_1 &= \mu_3 x_3 + \mu_4 x_4 + \cdots + \mu_n x_n \\ Y_2 &= \nu_3 x_3 + \nu_4 x_4 + \cdots + \nu_n x_n, \end{aligned} \tag{6}$$

then the partial correlation coefficient between x_1 and x_2, holding the remaining $n - 2$ variables constant, is defined as

$$r_{12.34\cdots n} = r_{(x_1 - Y_1)(x_2 - Y_2)};$$

and since $\Sigma(x_k - Y_k) = 0$,

$$r_{12.34\cdots n} = \frac{\frac{1}{N}\Sigma(x_1 - Y_1)(x_2 - Y_2)}{\sigma_{1.34\cdots n}\sigma_{2.34\cdots n}}. \tag{7}$$

Each μ_k is the negative of the ratio of the cofactor of r_{1k} to the cofactor of r_{11} in the determinant obtained by striking out the second row and the second column from R. We shall use the notation $R_{hi\text{-}jk}$ to mean the algebraic complement of the second order minor in R, whose complement is obtained by striking out row h and column i and then row j and column k. Then

$$\mu_k = \frac{R_{22\text{-}1k}}{R_{22\text{-}11}}.$$

By argument similar to that used in (3),

$$\Sigma(x_1 - Y_1)Y_2 = 0,$$

or

$$\Sigma x_1 Y_2 = \Sigma Y_1 Y_2.$$

Similarly,

$$\Sigma x_2 Y_1 = \Sigma Y_1 Y_2 .$$

Then the numerator of the right member of (7) becomes, after expanding and collecting terms,

$$r_{12} - \sigma_{Y_1} r_{x_2 Y_1} . \tag{8}$$

Multiplying both sides of (6) by $\frac{x_2}{N}$ and summing over the N individuals, we have,

$$\begin{aligned} \sigma_{Y_1} r_{x_2 Y_1} &= r_{23}\mu_3 + r_{24}\mu_4 + \cdots + r_{2n}\mu_n \\ &= \frac{1}{R_{22\text{-}11}} (r_{23} R_{22\text{-}13} + r_{24} R_{22\text{-}14} + \cdots + r_{2n} R_{22\text{-}1n}) \\ &= r_{12} + \frac{R_{12}}{R_{22\text{-}11}} . \end{aligned} \tag{9}$$

Analogous to (5), we have,

$$\sigma^2_{1.34\cdots n} = \frac{R_{22}}{R_{22\text{-}11}}, \qquad \sigma^2_{2.34\cdots n} = \frac{R_{11}}{R_{11\text{-}22}} . \tag{10}$$

From (8), (9), and (10) the right member of (7) becomes

$$\frac{-R_{12}}{\sqrt{R_{11} R_{22}}} .$$

It is seen that in general

$$r_{jk.12\cdots j-1,j+1,\cdots,k-1,k+1,\cdots n} = \frac{-R_{jk}}{\sqrt{R_{jj} R_{kk}}} .$$

UNIVERSITY OF MINNESOTA.