Also

$$\frac{1}{\sqrt{2\pi}}\int_0^k e^{-\frac{1}{2}y^2}\,dy = \frac{(a+b)-(c+d)}{2N} = .187, \text{ and } k = .4874.$$

Then,

$$H = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}h^2} = .3635, \text{ and } K = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}k^2} = .3543.$$

All the quantities except $r$ in the following approximate equation are known:

$$\frac{ad-bc}{N^2HK} = r + \frac{r^2}{2}hk + \frac{r^3}{6}(h^2-1)(k^2-1)$$
$$+ \frac{r^4}{24}h(h^2-3)k(k^2-3) + \frac{r^5}{125}(h^4-6h^2+3)(k^4-6k^2+3).$$

Therefore,

$$.0261r^5 + .0681r^4 + .1034r^3 + .1052r^2 + r - .4314 = 0.$$

Then, $r$ is approximately equal to .4051. Consequently, for practical purposes we can assume that $r = .4$.

28 Boody Street
Brunswick, Maine

John L. Roberts

## NOTE ON THE DERIVATION OF THE MULTIPLE CORRELATION COEFFICIENT

Consider $N$ observed values of each of $n$ variables. These $n \cdot N$ values may be tabulated in a double-entry table as follows:

$$\begin{matrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1N} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nN} \end{matrix}$$

where $X_{ik}$ is the $k^{\text{th}}$ value of the $i^{\text{th}}$ variable.

Using the $i^{th}$ variable as the dependent variable, the general linear relationship between the $n$ variables may be expressed by

$$x_i = {}_ia_1\,x_1 + {}_ia_2\,x_2 + \cdots + {}_ia_{i-1}\,x_{i-1} + {}_ia_{i+1}\,x_{i+1} + \cdots + {}_ia_n\,x_n \tag{1}$$

where

${}_ia_j$ is the general parameter which is to be determined empirically;

$x_j = X_j - M_j$;

$M_j$ is the arithmetic mean of the $j^{\text{th}}$ variable.

By the method of least squares, the constants of (1) must satisfy the normal equations:

$$(\Sigma x_1^2)_i a_1 + (\Sigma x_1 x_2)_i a_2 + \cdots + (\Sigma x_1 x_{i-1})_i a_{i-1}$$
$$+ (\Sigma x_1 x_{i+1})_i a_{i+1} + \cdots + (\Sigma x_1 x_n)_i a_n = \Sigma x_1 x_i$$
$$(\Sigma x_2 x_1)_i a_1 + (\Sigma x_2^2)_i a_2 + \cdots + (\Sigma x_2 x_{i-1})_i a_{i-1}$$
$$+ (\Sigma x_2 x_{i+1})_i a_{i+1} + \cdots + (\Sigma x_2 x_n)_i a_n = \Sigma x_2 x_i$$

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

$$(\Sigma x_{i-1} x_1)_i a_1 + (\Sigma x_{i-1} x_2)_i a_2 + \cdots + (\Sigma x_{i-1} x_n)_i a_n = \Sigma x_{i-1} x_i$$
$$(\Sigma x_{i+1} x_1)_i a_1 + (\Sigma x_{i+1} x_2)_i a_2 + \cdots + (\Sigma x_{i+1} x_n)_i a_n = \Sigma x_{i+1} x_i$$

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

$$(\Sigma x_n x_1)_i a_1 + (\Sigma x_n x_2)_i a_2 + \cdots + (\Sigma x_n^2)_i a_n = \Sigma x_n x_i$$

where

$$\left(\sum x_i x_j\right) = \sum_{k=1}^{N} (X_{ik} - M_i)(X_{jk} - M_j).$$

But

$$(\Sigma x_i x_j) = N r_{ij} \sigma_i \sigma_j,$$
$$(\Sigma x_i^2) = N \sigma_i^2 = N r_{ii} \sigma_i \sigma_i \qquad (2)$$

where

$r_{ij}$ is the Pearsonian coefficient of correlation between the $i^{\text{th}}$ and $j^{\text{th}}$ variables,

$\sigma_i$, the standard deviation of the $i^{\text{th}}$ variable.

Substituting the right members of (2) in the normal equations, we obtain the system:

$$\sum_{k=1}^{n} r_{1k} \sigma_1 \sigma_k \, {}_i a_k = 0$$
$$\sum_{k=1}^{n} r_{2k} \sigma_2 \sigma_k \, {}_i a_k = 0$$
$$\vdots$$
$$\sum_{k=1}^{n} r_{i-1,k} \, \sigma_{i-1} \sigma_k \, {}_i a_k = 0 \qquad (3)$$
$$\sum_{k=1}^{n} r_{i+1,k} \, \sigma_{i+1} \sigma_k \, {}_i a_k = 0$$
$$\vdots$$
$$\sum_{k=1}^{n} r_{nk} \sigma_n \sigma \, {}_i a_k = 0$$

where

$$ {}_i a_i = -1. $$

Let

$$ A = \begin{vmatrix} r_{11}\sigma_1\sigma_1 & \cdots & r_{n1}\sigma_n\sigma_1 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ r_{1n}\sigma_1\sigma_n & & r_{nn}\sigma_n\sigma_n \end{vmatrix}, \tag{4} $$

$A_{ij}$ be the first minor of the element $r_{ij}\sigma_i\sigma_j$ in $A$, ${}_{ik}A$ be $A$ with the $i^{\text{th}}$ and $k^{\text{th}}$ columns interchanged, and ${}_{ik}A_{ii}$ be the first minor of the element in the $i^{\text{th}}$ column and $i^{\text{th}}$ row of ${}_{ik}A$.

Solving (3) for ${}_i a_k$ by Cramer's rule, we find

$$ {}_i a_k = \frac{{}_{ik}A_{ii}}{A_{ii}}. $$

But it can easily be proved that

$$ {}_{ik}A_{ii} = (-1)^{i-k+1} A_{ik}; $$

hence

$$ {}_i a_k = (-1)^{i-k+1} \frac{A_{ik}}{A_{ii}}. $$

Using cofactors of $A$ instead of minors, we have

$$ {}_i a_k = (-1)^{i-k+1} \frac{(-1)^{i+k} D_{ik}}{D_{ii}} = -\frac{D_{ik}}{D_{ii}}. $$

Without writing the determinant out in full, we notice that the $\sigma$'s can be factored out. Hence

$$ \begin{aligned} {}_i a_k &= -\frac{\sigma_1^2\sigma_2^2 \cdots \sigma_{k-1}^2\sigma_k\sigma_{k+1}^2 \cdots \sigma_{i-1}^2\sigma_i\sigma_{i+1}^2 \cdots \sigma_n^2 K_{ik}}{\sigma_1^2\sigma_2^2 \cdots \sigma_{i-1}^2\sigma_{i+1}^2 \cdots \sigma_n^2 K_{ii}} \\ &= -\frac{\sigma_i K_{ik}}{\sigma_k K_{ii}}, \end{aligned} \tag{5} $$

where

$$ K = \begin{vmatrix} r_{11} & \cdots & r_{1n} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ r_{n1} & & r_{nn} \end{vmatrix}. $$

Using these derived values for the coefficients, we may write (1) in the symmetric form:

$$\frac{K_{i1}}{\sigma_1}(X_1 - M_1) + \frac{K_{i2}}{\sigma_2}(X_2 - M_2) + \cdots + \frac{K_{in}}{\sigma_n}(X_n - M_n) = 0,$$

or

$$\sum_{j=1}^{n} \frac{K_{ij}x_j}{\sigma_j} = 0. \tag{6}$$

For a multiple correlation coefficient, we use the formula

$$R_i^2 = 1 - \frac{\sum_{j=1}^{N}\left[x_{ij} - \left(\sum_{k=1}^{i-1} {}_i a_k x_{kj} + \sum_{k=i+1}^{n} {}_i a_k x_{kj}\right)\right]^2}{N\,\sigma_i^2}$$

which measures the amount of observed dispersion from the regression plane in which $X_i$ is the dependent variable.

Substituting the values for the $a$'s, we find

$$R_i^2 = 1 - \frac{\sum_{j=1}^{N}\left(\frac{K_{i1}x_{1j}}{\sigma_1} + \frac{K_{i2}x_{2j}}{\sigma_2} + \cdots + \frac{K_{in}x_{nj}}{\sigma_n}\right)^2}{K_{ii}^2\,N}.$$

Squaring the bracket expression and using (2) we obtain

$$R_i^2 = 1 - \frac{1}{K_{ii}^2}\left[\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\frac{K_{ik}K_{il}\sum_{j=1}^{N} x_{kj}x_{lj}}{N\sigma_k\sigma_l}\right)\right]$$

$$= 1 - \frac{1}{K_{ii}^2}\left[\sum_{k=1}^{n}\sum_{l=1}^{n} K_{ik}K_{il}r_{kl}\right]$$

$$= 1 - \frac{1}{K_{ii}^2}\left[\sum_{k=1}^{n}\left(K_{ik}\sum_{l=1}^{n} K_{il}r_{kl}\right)\right].$$

The second sum is the sum of the products of the elements in the $k^{\text{th}}$ row by the cofactors of the elements in the $i^{\text{th}}$ row. This sum is necessarily zero unless $k = i$; but if $k = i$, this sum is equal to $K$.

$$R_i^2 = 1 - \frac{1}{K_{ii}^2}(K_{ii}\,K) = 1 - \frac{K}{K_{ii}}.$$

Oregon State Agriculture College
School of Science
Corvallis, Oregon

William J. Kirkham