NOTE ON A REMARK OF D. J. STRUIK ON CORRELATION COEFFICIENTS

BY Y. B. D. DERKSEN

In his paper on *Correlation and group theory*,* D. J. Struik points to the dualism between the total and partial correlation coefficients.

Let there be given any scatter diagram in n-dimensional space, consisting of N points determined by the coordinates

$$x_1^{(i)}, \quad x_2^{(i)}, \quad \cdots, \quad x_n^{(i)}, \qquad (i = 1, 2, \cdots, N).$$

For ease of computation suppose that

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} = 0, \qquad (j = 1, 2, \cdots, n).$$

Then, if

$$\sigma_{jk} = \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} \cdot x_k^{(i)},$$

the Gram determinant of the system is

$$G \equiv \begin{vmatrix} \sigma_{11} & \sigma_{12} \cdots \sigma_{1n} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \sigma_{n1} & \cdots \quad \sigma_{nn} \end{vmatrix}.$$

Now let the minor of the element σ_{jk} be denoted by Σ_{jk}; then the equation of the so-called correlation hyperellipsoid may be written either in the form

(1) $\qquad \Sigma_{11}x_1^2 + 2\Sigma_{12}x_1x_2 + \cdots + \Sigma_{nn}x_n^2 = \text{const.},$

or, with tangential coordinates,

(2) $\qquad \sigma_{11}u_1^2 + 2\sigma_{12}u_1u_2 + \cdots + \sigma_{nn}u_n^2 = \text{const.}$

The total correlation coefficients follow from

(3) $\qquad\qquad r_{jk} = \frac{\sigma_{jk}}{\left(\sigma_{jj} \cdot \sigma_{kk}\right)^{1/2}},$

and the partial correlation coefficients are given by

---

* This Bulletin, vol. 36 (1930), pp. 869–878.

$$(4) \qquad r_{jk \cdot 12 \cdots (j) \cdots (k) \cdots n} = \frac{- \Sigma_{jk}}{(\Sigma_{jj} \Sigma_{kk})^{1/2}}.$$

The equations (1), (2), (3), and (4) show, according to Struik, that the total and partial correlation coefficients may be considered as the duals of each other.

We may ask if any statistical meaning can be attached to this view. This is answered by the following theorem.

Given any scatter diagram in n-dimensional space, it is possible to indicate a second, also in n-dimensional space, whose total correlation coefficients equal the opposites of the partial correlation coefficients of the first, and vice versa.

PROOF. Let the n coordinates of a point of a second scatter diagram be defined by

$$(5) \quad y_p^{(i_1, i_2, \cdots, i_{n-1})} = (-1)^p \begin{vmatrix} x_1^{(i_1)} & x_1^{(i_2)} & \cdots & x_1^{(i_{n-1})} \\ x_2^{(i_1)} & x_2^{(i_2)} & \cdots & x_2^{(i_{n-1})} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{p-1}^{(i_1)} & x_{p-1}^{(i_2)} & \cdots & x_{p-1}^{(i_{n-1})} \\ x_{p+1}^{(i_1)} & x_{p+1}^{(i_2)} & \cdots & x_{p+1}^{(i_{n-1})} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_n^{(i_1)} & x_n^{(i_2)} & \cdots & x_n^{(i_{n-1})} \end{vmatrix},$$

$(p = 1, 2, \cdots, n)$, where $(i_1, i_2, \cdots, i_{n-1})$ is any combination of $n-1$ numbers from the N integers.

If the total and partial correlation coefficients of this second scatter diagram are denoted by ρ_{jk} and $\rho_{jk \cdot 12 \cdots (j) \cdots (k) \cdots n}$, then we shall prove

$$(6) \qquad \rho_{jk} = - r_{jk \cdot 12 \cdots (j) \cdots (k) \cdots n},$$

$$(7) \qquad \rho_{jk \cdot 12 \cdots (j) \cdots (k) \cdots n} = - r_{jk}.$$

Consider now a well known theorem from the theory of determinants:

$$(8) \qquad \sum_{jk} = \frac{(-1)^{j+k}}{N^{n-1}}$$

$$\begin{vmatrix} \sum_1^N (x_1^{(i)})^2 & \sum_1^N x_1^{(i)} x_2^{(i)} & \cdots & \sum_1^N x_1^{(i)} x_{k-1}^{(i)} & \sum_1^N x_1^{(i)} x_{k+1}^{(i)} & \cdots & \sum_1^N x_1^{(i)} x_n^{(i)} \\ & & & & & & \\ \sum_1^N x_{j-1}^{(i)} x_1^{(i)} & \cdots & & & & & \\ & & & & & & \\ \sum_1^N x_{j+1}^{(i)} x_1^{(i)} & \cdots & & & & & \\ & & & & & & \\ \sum_1^N x_n^{(i)} x_1^{(i)} & \cdots & & & & & \end{vmatrix}$$

$$= \frac{1}{N^{n-1}} \sum_{i_1, i_2, \cdots, i_{n-1}} y_j^{(i_1, \cdots, i_{n-1})} y_k^{(i_1, \cdots, i_{n-1})},$$
$$(i_1 < i_2 < i_3 < \cdots < i_{n-1}).$$

Omitting the conditions $i_1 < i_2 < \cdots < i_{n-1}$, (8) becomes

$$9) \quad \sum_{jk} = \frac{1}{N^{n-1}(n-1)!} \sum_{\substack{i_1, i_2, \cdots, i_{n-1} \\ (i_p \neq i_q)}} y_j^{(i_1, i_2, \cdots, i_{n-1})} y_k^{(i_1, \cdots, i_{n-1})},$$

for the number of permutations of a set of $n-1$ fixed values $i_1, i_2, \cdots, i_{n-1}$ is $(n-1)!$, and although the individual quantities $y_p^{(i_1, \cdots, i_{n-1})}$ may change sign by such a permutation, the product $y_j^{(i_1, \cdots, i_{n-1})} y_k^{(i_1, \cdots, i_{n-1})}$ never changes. The right member of (9) contains $N(N-1) \cdots (N-n+2)$ terms.

Now developing the determinant $y_p^{(i_1, \cdots, i_{n-1})}$ by the elements of the first column and summing with respect to i_1 only, we have $\sum_{i_1} y_p^{(i_1, \cdots, i_{n-1})} = 0$, as $\sum_{i_1=1}^N x_j^{(i_1)} = 0$, $(j = 1, 2, \cdots, n)$. Therefore $\sum_{i_1, \cdots, i_{n-1}} y_p^{(i_1, \cdots, i_{n-1})} = 0$. Let

$$s_{jk} = \frac{1}{N(N-1) \cdots (N-n+2)} \sum_{i_1, \cdots, i_{n-1}} y_j^{(i_1, \cdots, i_{n-1})} y_k^{(i_1, \cdots, i_{n-1})},$$

let G' denote the determinant obtained from G by replacing σ_{jk} by s_{jk}, and let S_{jk} be the minor of the element s_{jk} of G'. From the definition of the total correlation coefficient and (9), it follows that

$$(6) \qquad \rho_{jk} = \frac{s_{jk}}{(s_{jj}s_{kk})^{1/2}} = \frac{\Sigma_{jk}}{(\Sigma_{jj}\Sigma_{kk})^{1/2}} = -\, r_{jk \cdot 12 \cdots (j) \cdots (k) \cdots n}.$$

From the formula for the partial correlation coefficient and (9) we get, using a theorem on reciprocal determinants,

$$(7) \qquad \rho_{jk \cdot 12 \cdots (j) \cdots (k) \cdots n} = \frac{-\, S_{jk}}{(S_{jj}S_{kk})^{1/2}} = \frac{-\, \sigma_{jk}}{(\sigma_{jj}\sigma_{kk})^{1/2}} = -\, r_{jk}.$$

An interesting result can be obtained by a generalization of Struik's remark: *A partial correlation coefficient of the pth order may be considered as the dual of another partial correlation co-efficient of the $(n-p-2)$th order.*

The statistical meaning of this statement is as follows.

If $r_{12 \cdot (i+1)(i+2) \cdots n}$ is a partial correlation coefficient of the $(n-i)$th order of the first scatter diagram and $\rho_{12 \cdot 3 \cdots i}$ a like coefficient of the $(i-2)$th order of the system defined by (5), then making use of a theorem on reciprocal determinants,

$$\rho_{12 \cdot 3 \cdots i} = \frac{\begin{vmatrix} \Sigma_{12} \cdots \Sigma_{1i} \\ \Sigma_{32} \qquad \cdot \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \Sigma_{i2} \cdots \Sigma_{ii} \end{vmatrix}}{\begin{vmatrix} \Sigma_{22} \cdots \Sigma_{2i} \\ \cdot \qquad \cdot \\ \cdot \qquad \cdot \\ \Sigma_{i2} \cdots \Sigma_{ii} \end{vmatrix}^{1/2} \begin{vmatrix} \Sigma_{11} \; \Sigma_{13} \cdots \Sigma_{1i} \\ \Sigma_{31} \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \Sigma_{i1} \cdots\cdots \Sigma_{ii} \end{vmatrix}^{1/2}}$$

$$= \frac{- \begin{vmatrix} \sigma_{21} \; \sigma_{2\,i+1} \cdots \sigma_{2n} \\ \sigma_{i+1\,1} \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \sigma_{n\,1} \cdots\cdots \sigma_{nn} \end{vmatrix}}{\begin{vmatrix} \sigma_{11} \; \sigma_{1\,i+1} \cdots \sigma_{1n} \\ \sigma_{i+1\,1} \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \sigma_{n\,1} \cdots\cdots \sigma_{nn} \end{vmatrix}^{1/2} \begin{vmatrix} \sigma_{22} \; \sigma_{2\,i+1} \cdots \sigma_{2n} \\ \sigma_{i+1\,2} \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \cdot \qquad\quad \cdot \\ \sigma_{n2} \cdots\cdots \sigma_{nn} \end{vmatrix}^{1/2}}.$$

Therefore

(10) $\rho_{12\cdot3\cdots i} = - r_{12\cdot(i+1)(i+2)\cdots n}.$

The numbering of the variables is immaterial; therefore generally,

(11) $\rho_{jk\cdot\{p\}} = - r_{jk\cdot\{n-p-2\}},$

where $\{p\}$ denotes any groups of p subcripts not including j or k, and $\{n-p-2\}$ the remaining ones. It is clear that (6) and (7) are particular cases of (11).

From the theory of partial correlation we know how to express partial correlations of order $n-2$ in terms of those of order $n-3$ and vice versa:

(12) $r_{12\cdot34\cdots n} = \dfrac{r_{12\cdot34\cdots(n-1)} - r_{1n\cdot34\cdots(n-1)}\cdot r_{2n\cdot34\cdots(n-1)}}{(1 - r_{1n\cdot34\cdots(n-1)}^2)^{1/2}\cdot(1 - r_{2n\cdot34\cdots(n-1)}^2)^{1/2}},$

(13) $r_{12\cdot34\cdots(n-1)} = \dfrac{r_{12\cdot34\cdots n} + r_{1n\cdot23\cdots(n-1)}\cdot r_{2n\cdot13\cdots(n-1)}}{(1 - r_{1n\cdot23\cdots(n-1)}^2)^{1/2}\cdot(1 - r_{2n\cdot13\cdots(n-1)}^2)^{1/2}}.$

These formulas resemble the well known formulas

(14) $\rho_{12\cdot3} = \dfrac{\rho_{12} - \rho_{13}\cdot\rho_{23}}{(1 - \rho_{13}^2)^{1/2}\cdot(1 - \rho_{23}^2)^{1/2}},$

(15) $\rho_{12} = \dfrac{\rho_{12\cdot3} + \rho_{13\cdot2}\cdot\rho_{23\cdot1}}{(1 - \rho_{13\cdot2}^2)^{1/2}\cdot(1 - \rho_{23\cdot1}^2)^{1/2}},$

though (12) and (13) are usually derived in a rather complicated way (see Yule, *Theory of Statistics*, Chapter 12) that does not show this similarity. Equation (14) follows from (4) if we take $n=3$, $j=1$, $k=2$.

Now if we apply (10) (for $n=3$) to (14), we immediately get (15) and then replacing the subscript 3 in (14) and (15) by the subscript n and again applying (11), we obtain (12) and (13).

DEN HAAG, HOLLAND