ON MULTIPLE AND PARTIAL CORRELATION COEFFICIENTS OF A CERTAIN SEQUENCE OF SUMS

By
CARL H. FISCHER

In a recent paper* the writer considered a sequence of q variables defined as follows: The first variable, x_1, is defined as the sum of n_1 values of a variable, t, drawn at random from a population characterized by a rather arbitrary continuous probability function, $f(t)$. Each succeeding variable, x_i, $(i > 1)$, is defined as the sum of $k_{i-1,i}$ values of t drawn at random from the n_{i-1} values composing x_{i-1}, plus the sum of $n_i - k_{i-1,i}$ values of t drawn at random from the parent population.

For variables thus defined, it was proved that the correlation coefficient between any two consecutive sums, x_i and x_{i+1}, is independent of the probability function, $f(t)$, and is given by

$$(1) \qquad r_{x_i x_{i+1}} = \frac{k_{i,i+1}}{(n_i n_{i+1})^{1/2}}$$

It was further shown that the correlation coefficient between two variables not consecutive in the sequence is equal to the product of the respective coefficients of correlation between all intermediate pairs of consecutive variables. Thus, the coefficient of correlation between x_j and x_p, $(j < p)$, is

$$r_{x_j x_p} = r_{x_j x_{j+1}} \cdot r_{x_{j+1}, j+2} \cdots r_{x_{p-2}, p-1} \cdot r_{x_{p-1} x_p} ;$$

or, in a simpler notation,

$$(2) \qquad r_{jp} = r_{j,\, j+1} \cdot r_{j+1,\, j+2} \cdots r_{p-2,\, p-1} \cdot r_{p-1,\, p} \, .$$

---

* On Correlation Surfaces of Sums with a Certain Number of Random Elements in Common. Annals of Mathematical Statistics, Vol. IV, pp. 103-126. May, 1933.

Let us now determine the multiple and partial correlation coefficients existing among a sequence of variables thus defined.

Consider the fundamental symmetric determinant, R, which, with its various co-factors, appears in the standard formulas for multiple and partial correlation coefficients.[1] If we substitute for each r_{ij}, $(j \neq i-1, i, i+1)$, in R, its equivalent from equation (2), we have

$$(3)\ R = \begin{vmatrix} 1 & r_{12} & r_{12}r_{23} \cdots & r_{12}r_{23}\cdots r_{q-1,q} \\ r_{12} & 1 & r_{23} \cdots\cdots & r_{23}r_{34}\cdots r_{q-1,q} \\ r_{12}r_{23} & r_{23} & 1 \cdots\cdots & r_{34}\cdots r_{q-1,q} \\ r_{12}r_{23}r_{34} & r_{23}r_{34} & r_{34} \cdots\cdots & r_{45}\cdots r_{q-1,q} \\ \vdots & \vdots & \vdots & \vdots \\ r_{12}r_{23}\cdots r_{q-2,q-1} & r_{23}\cdots r_{q-2,q-1} & r_{34}\cdots r_{q-2,q-1} \cdots & r_{q-1,q} \\ r_{12}r_{23}\cdots r_{q-1,q} & r_{23}\cdots r_{q-1,q} & r_{34}\cdots r_{q-1,q} \cdots & 1 \end{vmatrix}$$

Multiply the second row of R by r_{12} and subtract it from the first row. Now multiply the third row by r_{23} and subtract it from the second row. Continue this process, multiplying the j-th row by $r_{j-1,j}$ and subtracting this row from the $(j-1)$st row, until all possible rows have been so treated. Equation (3) may now be written

[1] H. L. Rietz, "Mathematical Statistics", Carus Monograph No. 4, pp. 94-100.

$$(4)\ R = \begin{vmatrix} (1-r_{12}^2) & 0 & 0 \cdots\cdots 0 & 0 \\ r_{12}(1-r_{23}^2) & (1-r_{23}^2) & 0 \cdots\cdots 0 & 0 \\ r_{12}r_{23}(1-r_{34}^2) & r_{23}(1-r_{34}^2) & (1-r_{34}^2) \cdots\cdots 0 & 0 \\ r_{12}r_{23}r_{34}(1-r_{45}^2) & r_{23}r_{34}(1-r_{45}^2) & r_{34}(1-r_{45}^2) \cdots\cdots 0 & 0 \\ \vdots & \vdots & \vdots \qquad \vdots & \vdots \\ r_{12}r_{23}\cdots r_{q-2,q-1}(1-r_{q-1,q}^2) & r_{23}\cdots r_{q-2,q-1}(1-r_{q-1,q}^2) & r_{34}\cdots r_{q-2,q-1}(1-r_{q-1,q}^2)\cdots(1-r_{q-1,q}^2) & 0 \\ r_{12}r_{23}\cdots r_{q-1,q} & r_{23}\cdots r_{q-1,q} & r_{34}\cdots r_{q-1,q} \cdots\cdots r_{q-1,q} & 1 \end{vmatrix}$$

The expansion of R may now be readily accomplished. The application of this same method of procedure to each of the various R_{ij}, (where R_{ij} is the co-factor of the element r_{ij}), yields without difficulty the results made use of in the remainder of this paper.

A. Multiple Correlation Coefficients.

The formula for the multiple correlation coefficient of one variable on the remaining $q-1$ variables is

$$(5) \qquad r_{j\cdot 123\cdots j-1,\, j+1,\cdots q} = \left(\frac{1-R}{R_{jj}}\right)^{\frac{1}{2}} .$$

From equations (3) and (4) we derive the following expressions for the necessary R and R_{jj}, $(j = 1, 2, 3, \cdots q)$.

$$R = (1-r_{12}^2)(1-r_{23}^2)(1-r_{34}^2)(1-r_{45}^2) \cdots\cdots (1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$R_{11} = (1-r_{23}^2)(1-r_{34}^2)(1-r_{45}^2) \cdots\cdots (1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$R_{22} = (1-r_{12}^2 r_{23}^2)(1-r_{34}^2)(1-r_{45}^2) \cdots (1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$\text{(6)}\quad R_{33} = (1-r_{12}^2)(1-r_{23}^2 r_{34}^2)(1-r_{45}^2) \cdots\cdots (1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$\vdots$$

$$R_{jj} = (1-r_{12}^2)(1-r_{23}^2) \cdots (1-r^2_{j-2,j-1})(1-r^2_{j-1,j}\, r^2_{j,j+1})(1-r^2_{j+1,j+2}) \cdots$$

$$\cdots\cdots (1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$\vdots$$

$$R_{q-1,q-1} = (1-r_{12}^2)(1-r_{23}^2) \cdots\cdots (1-r^2_{q-3,q-2})(1-r^2_{q-2,q-1}\, r^2_{q-1,q});$$

$$R_{qq} = (1-r_{12}^2)(1-r_{23}^2) \cdots\cdots (1-r^2_{q-3,q-2})(1-r^2_{q-2,q-1}).$$

Upon substituting the proper values from (6) in formula (5) for the multiple correlation coefficients of the first and last variables, respectively, on the others in the sequence, we find

$$\text{(7)}\qquad r_{1\cdot 234\cdots q} = r_{12};$$

$$\text{(8)}\qquad r_{q\cdot 123\cdots q-1} = r_{q-1,q}.$$

The multiple correlation coefficient of any other variable, x_j , on the remaining $q-1$ variables is given by

$$(9) \qquad r_{j\cdot 123\cdots q} = \left[\frac{1-(1-r^2_{j-1,j})(1-r^2_{j,j+1})}{(1-r^2_{j-1,j}r^2_{j,j+1})}\right]^{\frac{1}{2}} = \left[\frac{(r^2_{j-1,j}+r^2_{j,j+1}-2r^2_{j-1,j}\,r^2_{j,j+1})}{(1-r^2_{j-1,j}\,r^2_{j,j+1})}\right]^{\frac{1}{2}}.$$

It is to be noted that the right member of (9) is independent of all of the simple correlation coefficients except $r_{j-1,j}$ and $r_{j,j+1}$.

B. Partial Correlation Coefficients.

The formula for the partial correlation coefficient between any two variables is

$$(10) \qquad r_{ij\cdot 1234\cdots q} = \frac{-R_{ij}}{(R_{ii}\,R_{jj})^{1/2}} .$$

From equation (3) we derive the following expressions for the co-factors of elements other than those of the principal diagonal of the determinant. Because of the symmetry of the fundamental determinant, we know that $R_{ij} = R_{ji}$; hence in expanding the co-factors of the elements of each row, we shall consider only the R_{ij} where $i \leqq j$.

1. The co-factors of the elements of the first row are

$$(11) \qquad \begin{aligned} -R_{12} &= r_{12}(1-r^2_{23})(1-r^2_{34})\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q}); \\ R_{1i} &= 0, \ (i = 3,4,5,\cdots q). \end{aligned}$$

2. The co-factors of the elements of the second row are

$$(12) \qquad \begin{aligned} -R_{23} &= (1-r^2_{12})r_{23}(1-r^2_{34})\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q}); \\ R_{2i} &= 0, \ \ (i = 4,5,6,\cdots q). \end{aligned}$$

3. The co-factors of the elements of the j-th row are

$$(13) \quad -R_{j,j+1} = (1-r_{12}^2)(1-r_{23}^2)\cdots(1-r_{j-1,j}^2)r_{j,j+1}(1-r_{j+1,j+2}^2)\cdots(1-r_{q-1,q}^2);$$

$$R_{ji} = 0, \ (i = j+2, j+3, \cdots\cdots q).$$

4. The co-factor of the last element of the q-th row is

$$(14) \qquad -R_{q-1,q} = (1-r_{12}^2)(1-r_{23}^2)\cdots\cdots(1-r_{q-2,q-1}^2)r_{q-1,q}.$$

We see at once that all partial correlation coefficients between non-consecutive variables vanish, as each co-factor $R_{ij} = 0$ if $i \neq j-1, j, j+1$. The non-vanishing coefficients, those between consecutive variables, are given below.

$$(15) \qquad r_{12\cdot 345\cdots q} = r_{12}\left[\frac{(1-r_{23}^2)}{(1-r_{12}^2\, r_{23}^2)}\right]^{\frac{1}{2}};$$

$$(16) \qquad r_{j,j+1\cdot 1234\cdots q} = r_{j,j+1}\left[\frac{(1-r_{j-1,j}^2)(1-r_{j+1,j+2}^2)}{(1-r_{j-1,j}^2\, r_{j,j+1}^2)(1-r_{j,j+1}^2\, r_{j+1,j+2}^2)}\right]^{\frac{1}{2}};$$

$$(17) \qquad r_{q-1,q\cdot 1234\cdots q-2} = r_{q-1,q}\left[\frac{(1-r_{q-2,q-1}^2)}{(1-r_{q-2,q-1}^2\, r_{q-1,q}^2)}\right]^{\frac{1}{2}}$$

From (16) we can state that in general the partial correlation coefficient of consecutive sums x_j and x_{j+1} is independent of all simple correlation coefficients except $r_{j-1,j}$, $r_{j,j+1}$, and $r_{j+1,j+2}$.

C. *Summary.*

To summarize, we have shown that

1. The multiple correlation coefficient of a variable x_j on the remaining variables of our sequence is independent of all simple correlation coefficients except those between x_j and the immediately preceding and the immediately following variables, respectively.
2. The partial correlation coefficients between all pairs of non-consecutive variables in our sequence are zero; a result that appeals to the intuition when it is recalled that we are eliminating the effect of the variables that form the connecting links between the two under consideration.
3. The partial correlation coefficient of any pair of consecutive variables, x_j and x_{j+1}, is independent of all simple correlation coefficients except those between the two consecutive variables in question, between the first of these and the variable immediately preceding it, and between the second of the pair and the variable immediately following it.

Carl H. Fischer