CORRELATION SURFACES OF TWO OR MORE INDICES WHEN THE COMPONENTS OF THE INDICES ARE NORMALLY DISTRIBUTED

By George A. Baker

Indices are widely used in statistical analyses.[1] In many cases incorrect conclusions are drawn because indices are not uncorrelated or independent even though all of the component variables are independent. In a previous paper[2] the distribution of an index both of whose components follow the normal law was given exactly i.e. without approximation. The purpose of the present paper is to give the simultaneous distribution of two or more indices when each of the components follow the normal law. The case for two indices will be discussed in detail and the extension to more indices will be indicated.

Let x_1, x_2, and x_3, be correlated variables each being normally distributed about their respective means m_1, m_2, m_3, with standard deviations σ_1, σ_2, σ_3, and let the correlations between the variables in pairs be represented by r_{12}, r_{13}, r_{23}. Then the simultaneous distribution of these three variables will be

$$\begin{aligned} \frac{1}{(2\pi)^{3/2}R^{1/2}\sigma_1\sigma_2\sigma_3}\,\text{exp.}\, -\frac{1}{2}\frac{1}{R}\Bigg[&\frac{R_{11}(x_1-m_1)^2}{\sigma_1^2}+\frac{R_{22}(x_2-m_2)^2}{\sigma_2^2}+\frac{R_{33}(x_3-m_3)^2}{\sigma_3^2} \\ &+2R_{12}\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}+2R_{13}\frac{(x_1-m_1)(x_3-m_3)}{\sigma_1\sigma_3} \\ &+2R_{23}\frac{(x_2-m_2)(x_3-m_3)}{\sigma_2\sigma_3}\Bigg]dx_1\,dx_2\,dx_3 \end{aligned} \tag{1}$$

where

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

and R_{ij} are the respective second order minors of R.

---

[1] Rietz, H. L. "On the Frequency Distribution of Certain Ratios," Annals of Mathematical Statistics, Vol. VII, No. 3, Sept. 1936, pp. 145–153.

[2] Baker, G. A., "Distribution of the Means Divided by the Standard Deviations of Samples From Non-homogeneous Populations," Annals of Mathematical Statistics, Feb. 1932, pp. 3–5.

If we make the transformation

$$z_1 = \frac{x_1}{x_3}, \qquad x_1 = z_1 z_3$$

$$z_2 = \frac{x_2}{x_3}, \qquad x_2 = z_2 z_3$$

$$z_3 = x_3, \qquad x_3 = z_3$$

$$dx_1\, dx_2\, dx_3 = z_3^2\, dz_1\, dz_2\, dz_3$$

which is certainly valid if x_1, x_2, x_3, are all positive, then (1) becomes

$$\frac{1}{(2\pi)^{3/2} R^{1/2} \sigma_1 \sigma_2 \sigma_3} \exp. -\frac{1}{2}\frac{1}{R}\left[\frac{R_{11}(z_1 z_3 - m_1)^2}{\sigma_1^2} + \frac{R_{22}(z_2 z_3 - m_2)^2}{\sigma_2^2}\right.$$

$$(2) \qquad + \frac{R_{33}(z_3 - m_3)^2}{\sigma_3^2} + 2R_{12}\frac{(z_1 z_3 - m_1)(z_2 z_3 - m_2)}{\sigma_1 \sigma_2} + 2R_{13}\frac{(z_1 z_3 - m_1)(z_3 - m_3)}{\sigma_1 \sigma_3}$$

$$\left. + 2R_{23}\frac{(z_2 z_3 - m_2)(z_3 - m_3)}{\sigma_2 \sigma_3}\right] z_3^2\, dz_1\, dz_2\, dz_3 .$$

If x_1, x_2, x_3 are all positive the corresponding distribution of z_1 and z_2 can be obtained by integrating (2) between the limits 0 and ∞ with respect to z_3. If x_1, x_2, x_3 are all negative z_1 and z_2 are again both positive so that in order to get the total distribution for z_1 and z_2 it is necessary to add to the integral of (2) between the limits 0 and ∞ with respect to z_3 the similar integral of (2) with z_3 replaced by $-z_3$. The result is

$$(3) \qquad \frac{2e^{-\frac{1}{2}\frac{c}{R}} e^{\frac{1}{2}\frac{b^2}{Ra}}}{(2\pi)^{3/2} R^{1/2} \sigma_1 \sigma_2 \sigma_3}\left[\frac{\sqrt{\pi}}{\sqrt{2}}\frac{R^{3/2}}{a^{3/2}} - \frac{b^2}{a^2}\int_0^{\frac{b}{\sqrt{R}\sqrt{a}}} e^{-\frac{1}{2}z^2}\, dz + \frac{R^{1/2} b^2}{a^{5/2}}\frac{\sqrt{\pi}}{\sqrt{2}}\right]$$

where

$$a = \frac{R_{11}}{\sigma_1^2} z_1^2 + \frac{R_{22}}{\sigma_2^2} z_2^2 + \frac{R_{33}}{\sigma_3^2} + \frac{2R_{12}}{\sigma_1 \sigma_2} z_1 z_2 + \frac{2R_{13}}{\sigma_1 \sigma_3} z_1 + \frac{2R_{23}}{\sigma_2 \sigma_3} z_2$$

$$b = \frac{R_{11}}{\sigma_1^2} m_1 z_1 + \frac{R_{22}}{\sigma_2^2} m_2 z_2 + \frac{R_{33}}{\sigma_3^2} m_3 + \frac{R_{12}}{\sigma_1 \sigma_2} z_1 m_2 + \frac{R_{12}}{\sigma_1 \sigma_2} m_1 z_2 + \frac{R_{13}}{\sigma_1 \sigma_3} m_3 z_1$$

$$+ \frac{R_{13}}{\sigma_1 \sigma_3} m_1 + \frac{R_{23}}{\sigma_2 \sigma_3} m_3 z_2 + \frac{R_{23}}{\sigma_2 \sigma_3} m_2$$

$$c = \frac{R_{11}}{\sigma_1^2} m_1^2 + \frac{R_{22}}{\sigma_2^2} m_2^2 + \frac{R_{33}}{\sigma_3^2} m_3^2 + \frac{2R_{12}}{\sigma_1 \sigma_2} m_1 m_2 + \frac{2R_{13}}{\sigma_1 \sigma_3} m_1 m_3 + \frac{2R_{23}}{\sigma_2 \sigma_3} m_2 m_3 .$$

The same result (3) is obtained for z_1, and z_2 negative, z_1 positive and z_2 negative, z_1 negative and z_2 positive. That is (3) is the simultaneous distribution of z_1 and z_2. The extension to more than 2 indices is immediate. The form of the distribution of the indices and the denominator variable is the same as (2) except that a, b, and c, the coefficients of z_3^2, z_3 and the constant term respectively in the exponent of e, will be different in that they will include the new indices and the exponent on the denominator variable will be the same as the number of indices involved. The distribution of the indices will again be obtained by integrating from 0 to ∞ with respect to the denominator variable.

The case when all of the variables x_1, x_2, x_3 are independent is especially interesting. If r_{12}, r_{13}, r_{23} are all zero then $R = R_{11} = R_{22} = R_{33} = 1$, $R_{12} = R_{13} = R_{23} = 0$ and a, b, c, become a', b', c', respectively.

$$a' = \frac{z_1^2}{\sigma_1^2} + \frac{z_2^2}{\sigma_2^2} + \frac{1}{\sigma_3^2}$$

$$b' = \frac{m_1 z_1}{\sigma_1^2} + \frac{m_2 z_2}{\sigma_2^2} + \frac{m_3}{\sigma_3^2}$$

$$c' = \frac{m_1^2}{\sigma_1^2} + \frac{m_2^2}{\sigma_2^2}\frac{m_3^2}{\sigma_3^2}$$

Under these conditions and the further condition that m_1, m_2, m_3 are large with respect to σ_1, σ_2, σ_3 respectively so that the integral term of (3) maybe neglected (3) becomes

$$\text{(4)} \qquad \frac{e^{-\frac{1}{2}\left(\frac{m_1^2}{\sigma_1^2}+\frac{m_2^2}{\sigma_2^2}+\frac{m_3^2}{\sigma_3^2}\right)}}{2\pi\sigma_1\sigma_2\sigma_3} \frac{e^{\frac{1}{2}\frac{\left(\frac{m_1 z_1}{\sigma_1^2}+\frac{m_2 z_2}{\sigma_2^2}+\frac{m_3}{\sigma_3^2}\right)^2}{\left(\frac{z_1^2}{\sigma_1^2}+\frac{z_2^2}{\sigma_2^2}+\frac{1}{\sigma_3^2}\right)}}}{\left(\frac{z_1^2}{\sigma_1^2} + \frac{z_2^2}{\sigma_2^2} + \frac{1}{\sigma_3^2}\right)^{\frac{3}{2}}} \left[1 + \frac{\left(\frac{m_1 z_1}{\sigma_1^2} + \frac{m_2 z_2}{\sigma_2^2} + \frac{m_3}{\sigma_3^2}\right)^2}{\left(\frac{z_1^2}{\sigma_1^2} + \frac{z_2^2}{\sigma_2^2} + \frac{1}{\sigma_3^2}\right)}\right]$$

It is clear that z_1 and z_2 are not independent in the probability sense for distribution (4).

The question as to the possibility of having the variables independent and the indices independent at the same time arises. Denote the distribution functions of x_1, x_2, x_3, by $X_1(x_1)$, $X_2(x_2)$, $X_3(x_3)$ and of z_1, z_2 by $Z_1(z_1)$, $Z_2(z_2)$. Then, if $x_i \geq 0$, $i = 1, 2, 3$ it is necessary that

$$\text{(5)} \qquad \int_a^b X_1(z_3 z_1) X_2(z_3 z_2) X_3(z_3) z_3^2 \, dz_3 = Z_1(z_1) Z_2(z_2)$$

a and b being suitable limits.

For instance, let

$$X_1(x_1) = \frac{1}{x_1^2}, \qquad 1 \leq x_1 \leq 3$$

$$X_2(x_2) = \frac{1}{x_2^2}, \qquad 1 \leq x_2 \leq 3$$

$$X_3(x_3^2) = x_3^2, \qquad 1 \leq x_3 \leq 2$$

then

$$Z_1(z_1) = \frac{c_1}{z_1^2}$$

$$Z_2(z_2) = \frac{c_2}{z_2^2}$$

for value of z_1 and z_2 within a straight line sided area the corners of which are $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 1)$, $(1, 1)$ and $(1, 2)$. z_1, and z_2 are not uncorrelated throughout their entire set of values but are for this particular set of values. Thus is appears that it is possible that the indices may be independent when the variables are, but not necessarily so.

Indices should be used with care since it is very easy to draw invalid conclusions from the consideration of them. Usually it is better to use partial correlation analysis to remove the influence of a third factor than to calculate indices.