UNBIASED ESTIMATION OF CERTAIN CORRELATION COEFFICIENTS[1]

By Ingram Olkin[2] and John W. Pratt[3]

University of Chicago

1. Summary and introduction. This paper deals with the unbiased estimation of the correlation of two variates having a bivariate normal distribution (Sec. 2), and of the intraclass correlation, i.e., the common correlation coefficient of a p-variate normal distribution with equal variances and equal covariances (Sec. 3).

In both cases, the estimator has the following properties. It is a function of a complete sufficient statistic and is therefore the unique (except for sets of probability zero) minimum variance unbiased estimator. Its range is the region of possible values of the estimated quantity. It is a strictly increasing function of the usual estimator differing from it only by terms of order $1/n$ and consequently having the same asymptotic distribution.

Since the unbiased estimators are cumbersome in form in that they are expressed as series or integrals, tables are included giving the unbiased estimators as functions of the usual estimators.

In Sec. 4 we give an unbiased estimator of the squared multiple correlation. It has the properties mentioned in the second paragraph except that it may be negative, which the squared multiple correlation cannot.

In each case the estimator is obtained by inverting a Laplace transform.

We are grateful to W. H. Kruskal and L. J. Savage for very helpful comments and suggestions, and to R. R. Blough for his able computations.

2. Correlation coefficient. Let $(x_1, y_1), \cdots, (x_N, y_N)$ be independently distributed, each bivariate normal with means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation ρ. The problem is to estimate ρ unbiasedly in the cases (i) μ_1, μ_2 known, $\sigma_1^2, \sigma_2^2, \rho$ unknown, and (ii) all parameters unknown.

Sufficiency and invariance suggest that we confine ourselves to odd functions of r, where r is the usual sample correlation coefficient in either case, namely,

$$r = \frac{\sum (x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sqrt{\sum (x_i - \hat{\mu}_1)^2 \sum (y_i - \hat{\mu}_2)^2}},$$

where $(\hat{\mu}_1, \hat{\mu}_2)$ equals (μ_1, μ_2) in (i) and $(\bar{x}, \bar{y})$ in (ii).

2.1. *Derivation of the unbiased estimator*. The density of r is

$$(2.1) \qquad p(r) = \frac{2^{n-2}}{\pi\Gamma(n-1)} (1 - \rho^2)^{n/2}(1 - r^2)^{(n-3)/2} \sum_{k=0}^{\infty} \Gamma^2\left(\frac{n+k}{2}\right) \frac{(2\rho r)^k}{k!},$$

Received August 20, 1956; revised August 15, 1957.

[1] This research was sponsored by the Office of Ordnance Research, U. S. Army, and the Office of Naval Research.

[2] Work done while on leave from Michigan State University.

[3] Now at Harvard University.

where the degrees of freedom are $n = N$ and $N - 1$ in cases (i) and (ii). (We assume $n \geqq 2$, the case $n = 1$ being degenerate.) The condition $E[G(r)] = \rho$, i.e., $G(r)$ is unbiased, is equivalent to

$$\frac{2^{n-2}}{\pi\Gamma(n-1)} \sum_{k=0}^{\infty} \Gamma^2\left(\frac{n+k}{2}\right) \frac{(2\rho)^k}{k!} \int_{-1}^{1} G(r)(1-r^2)^{(n-3)/2} r^k \, dr = (1-\rho^2)^{-n/2}\rho$$

$$= \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{n}{2}+j\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{\rho^{2j+1}}{j!}.$$

Comparing coefficients of powers of ρ, we find that $G(r)$ is indeed an odd function, and that

$$\int_0^1 G(r)(1-r^2)^{(n-3)/2} r^{2j+1} \, dr = \frac{\pi\Gamma(n-1)\Gamma(2j+2)}{2^{n+2j}\Gamma^2\left(\frac{n+2j+1}{2}\right)} \frac{\Gamma\left(\frac{n+2j}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma(j+1)}.$$

Using the identity (e.g., [3, 12.4.4])

$$\sqrt{\pi}\Gamma(2p) = 2^{2p-1}\Gamma(p)\Gamma(p + 1/2),$$

and making the substitution $r = \exp(-\frac{1}{2} y)$, we obtain

$$\int_0^{\infty} G(e^{-\frac{1}{2}y})(1-e^{-y})^{(n-3)/2} e^{-y} e^{-jy} \, dy = \Gamma\left(\frac{n-1}{2}\right) \frac{\Gamma\left(\frac{3}{2}+j\right)\Gamma\left(\frac{n}{2}+j\right)}{\Gamma^2\left(\frac{n+1}{2}+j\right)}.$$

As a function of j, for $n \geqq 2$, the right-hand side is the unilateral Laplace transform of

$$e^{-\frac{3}{2}y}(1-e^{-y})^{((n-1)/2)-1} F\left(\frac{1}{2}, \frac{1}{2}; \frac{n-1}{2}; 1-e^{-y}\right)$$

[1, p. 262 (7)], where F is the hypergeometric function

(2.2) $$F(\alpha, \beta; \gamma; x) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)\Gamma(\beta+k)\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma+k)} \frac{x^k}{k!}.$$

Therefore

(2.3) $$G(r) = rF(\tfrac{1}{2}, \tfrac{1}{2}; (n-1)/2; 1-r^2).$$

Some alternative representations of $G(r)$ are

(2.4) $$G(r) = r \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \int_0^1 \frac{t^{-1/2}(1-t)^{((n-2)/2)-1}}{[1-t(1-r^2)]^{1/2}} \, dt$$

and

$$(2.5)\qquad G(r) = r\,\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}\int_0^\infty \frac{t^{-1/2}(1+t)^{-(n-2)/2}}{(1+tr^2)^{1/2}}\,dt$$

[2, 2.12 (1) and (5)].

2.2 *Properties of the unbiased estimator.* $G(r)$ is an odd function of r by (2.3), and is strictly increasing since, in (2.5), $r(1 + tr^2)^{-1/2}$ is strictly increasing in r for each value of t, $0 < t < \infty$. For $\rho = \pm 1$, $G(r) = r = \pm 1$ with probability 1, and consequently, $-1 \leqq G(r) \leqq 1$, which is the range of ρ.

As remarked before, $G(r)$ is the unique minimum variance unbiased estimator of ρ.

To obtain the asymptotic distribution of $G(r)$, we note that, by (2.2),

$$F(\alpha, \beta; \gamma; x) = 1 + x\,O(1/\gamma)$$

as $\gamma \to \infty$ (uniformly in x for x in any bounded set), so that $G(r) = r + O_p(1/n)$. Therefore $\sqrt{n}[G(r) - \rho]$ has the same asymptotic distribution as $\sqrt{n}[r - \rho]$, which is $N(O, (1 - \rho^2)^2)$, [3, p. 366].

In order to facilitate the use of the unbiased estimator $G(r)$, Table 1 gives $G(r)$ and (for easier interpolation) $G(r)/r$ for $r = 0(.1)\,1$ and $n = 2(2)\,30$. The computation was carried out by means of the recursive relation

$$xF(\tfrac{1}{2}, \tfrac{1}{2}; \gamma + 1; x)$$
$$= \left[1 - \frac{1}{(2\gamma - 1)^2}\right][(2x - 1)F(\tfrac{1}{2}, \tfrac{1}{2}; \gamma; x) + (1 - x)F(\tfrac{1}{2}, \tfrac{1}{2}; \gamma - 1; x)],$$

[2, 2.8 (30)], together with the initial conditions

$$F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}; x) = 1/\sqrt{1 - x},$$
$$F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}; x) = \text{arc sin}\,\sqrt{x}/\sqrt{x},$$

[2, 2.8 (4) and (13)].

Approximations for $G(r)$ can be obtained from the expansion (2.2), which gives

$$(2.6)\qquad \frac{G(r)}{r} = 1 + \frac{1 - r^2}{2(n - 1)} + \frac{9(1 - r^2)^2}{8(n^2 - 1)} + O(n^{-3}).$$

(2.6) gives $G(r)/r$ within .01 for $n \geqq 14$ or .001 for $n \geqq 36$ if two terms are included, and within .01 for $n \geqq 10$ or .001 for $n \geqq 18$ if three terms are included. The neglected terms in the first line of (2.6) are all positive and decreasing in r^2 and n. Therefore, if $G(r)$ is estimated by cutting off this series, the estimate will be too small, by a percentage which decreases as r^2 and n increase.

The k that minimizes the maximum over r of the absolute difference between (2.6) and $1 + (1 - r^2)/2(n - k)$ is, for large n, $(-7 + 9\sqrt{2})/2 = 2.87$. This suggests the approximation

$$(2.7)\qquad \frac{G(r)}{r} = 1 + \frac{1 - r^2}{2(n - 3)}.$$

This is accurate within .01 for $n \geqq 8$, and within .001 for $n \geqq 18$.

TABLE 1

Ordinary bivariate correlation coefficient, n degrees of freedom

$$G(r) = rF\left(1/2, 1/2; \frac{n-1}{2}; 1 - r^2\right), \qquad r = \frac{s_{12}}{\sqrt{s_{11}\, s_{22}}}$$

1a. *Table of G(r)*

n	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	0	1	1	1	1	1	1	1	1	1	1
4	0	.148	.280	.398	.506	.605	.695	.780	.858	.931	1
6	0	.117	.232	.343	.450	.552	.650	.744	.833	.918	1
8	0	.110	.220	.327	.432	.534	.633	.730	.823	.913	1
10	0	.107	.214	.319	.423	.525	.625	.722	.817	.910	1
12	0	.106	.211	.315	.418	.520	.620	.718	.814	.908	1
14	0	.105	.209	.312	.415	.516	.616	.715	.812	.907	1
16	0	.104	.207	.311	.413	.514	.614	.713	.810	.906	1
18	0	.103	.206	.309	.411	.512	.612	.711	.809	.905	1
20	0	.103	.206	.308	.410	.511	.611	.710	.808	.905	1
22	0	.103	.205	.307	.409	.510	.610	.709	.807	.904	1
24	0	.102	.205	.307	.408	.509	.609	.708	.806	.904	1
26	0	.102	.204	.306	.407	.508	.608	.707	.806	.903	1
28	0	.102	.204	.305	.407	.507	.607	.707	.805	.903	1
30	0	.102	.204	.305	.406	.507	.607	.706	.805	.903	1
∞	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

1b. *Table of G(r)/r*

n	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	∞	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1
4	1.571	1.478	1.398	1.327	1.265	1.209	1.159	1.114	1.073	1.035	1
6	1.178	1.173	1.161	1.144	1.125	1.105	1.083	1.062	1.041	1.020	1
8	1.104	1.103	1.098	1.090	1.080	1.068	1.056	1.042	1.028	1.014	1
10	1.074	1.073	1.070	1.065	1.058	1.050	1.042	1.032	1.022	1.011	1
12	1.057	1.056	1.054	1.050	1.046	1.040	1.033	1.026	1.018	1.009	1
14	1.046	1.046	1.044	1.041	1.038	1.033	1.027	1.021	1.015	1.008	1
16	1.039	1.039	1.037	1.035	1.032	1.028	1.023	1.018	1.013	1.006	1
18	1.034	1.033	1.032	1.030	1.028	1.024	1.020	1.016	1.011	1.006	1
20	1.030	1.029	1.028	1.027	1.024	1.022	1.018	1.014	1.010	1.005	1
22	1.027	1.026	1.025	1.024	1.022	1.019	1.016	1.013	1.009	1.005	1
24	1.024	1.024	1.023	1.022	1.020	1.018	1.015	1.012	1.008	1.004	1
26	1.022	1.022	1.021	1.020	1.018	1.016	1.014	1.011	1.007	1.004	1
28	1.020	1.020	1.019	1.018	1.017	1.015	1.012	1.010	1.007	1.004	1
30	1.019	1.018	1.018	1.017	1.015	1.014	1.012	1.009	1.006	1.003	1
∞	1	1	1	1	1	1	1	1	1	1	1

By (2.2) and (2.3), $G(r)/r$ is larger than 1 and decreasing in r^2 and n, as Table 1b suggests.

2.3 *Partial correlation coefficient.* We observe that an unbiased estimator of the partial correlation coefficient can be immediately obtained from the preceding.

section. More precisely, suppose the columns of $X:p \times N$ are independently distributed each as p-variate normal with mean vector μ and covariance matrix Σ. We wish to give an unbiased estimator of the partial correlation coefficient $\rho_{12\cdot(q\cdots p)}$. The usual estimator, $r_{12\cdot(q\cdots p)}$, has the density (2.1) with $n = N - (p - q)$ if μ is known, and $n = N - 1 - (p - q)$ if μ is unknown. Therefore $G(r_{12\cdot(q\cdots p)})$ (with appropriate n) is the unique minimum variance unbiased estimator of $\rho_{12\cdot(q\cdots p)}$ and possesses the other properties of $G(r)$.

3. Intraclass correlation coefficient. Let the columns of X: $p \times N$ be independently distributed, each as $N(\mu, \Sigma^*)$, i.e., p-variate normal with mean vector μ and covariance matrix Σ^*. Suppose Σ^* is of the form $\sigma^2[(1 - \rho) I + \rho ee']$, where $e' = (1, \cdots, 1)$, i.e., $\sigma_{ii}^* = \sigma^2$, $\sigma_{ij}^* = \rho\sigma^2$ $(i \neq j)$, with ρ and σ^2 unknown. The problem is to estimate ρ unbiasedly.

We note that ρ is just the slope of the regression line of x_2 on x_1, and is therefore estimated unbiasedly by

$$\hat{\rho} = \frac{\sum_{\alpha=1}^{N} (x_{1\alpha} - x_{1\cdot})(x_{2\alpha} - x_{2\cdot})}{\sum_{\alpha=1}^{N} (x_{1\alpha} - x_{1\cdot})^2},$$

where a dot indicates an average over the omitted subscript. We will see presently that there is a complete sufficient statistic (u, v). $\hat{\rho}$ is not a function of (u, v), nor is it confined to the range of ρ, namely, $-1/(p - 1)$ to 1. However, by the Blackwell-Rao theorem, $E(\hat{\rho} \mid u, v)$ is the unique minimum variance unbiased estimator of ρ. Since $E(\hat{\rho} \mid u, v)$ is difficult to obtain, we shall use the joint distribution of u and v to obtain an unbiased estimator $h(u, v)$ of ρ, which, by completeness, must equal $E(\hat{\rho} \mid u, v)$.

As in the previous section, sufficiency and invariance suggest that we confine ourselves to functions of the conventional estimator r' of ρ. However, it is easier to deal with the density of (u, v), and it will turn out that the unbiased estimator $h(u, v)$ is a function $H(r')$ of r' alone.

3.1 *Reduction to canonical form.* Let $\Delta:p \times p$ be an orthogonal matrix with first row $p^{-1/2}e'$, and let $Y = \Delta X$. Then the columns of Y are independently distributed, each as $N(\Delta\mu, \Delta\Sigma^*\Delta')$. Now

$$\Delta\Sigma^*\Delta' = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 I \end{pmatrix},$$

where $\sigma_1^2 = \sigma^2 [1 + (p - 1)\rho]$, $\sigma_2^2 = \sigma^2 (1 - \rho)$. Because of the particular diagonal form of the covariance matrix, the $y_{i\alpha}(i = 1, \cdots, p; \alpha = 1, \cdots, N)$ are independent, and if we let $\eta = \Delta\mu = Ey$, then $y_{1\alpha}$ is $N(\eta_1, \sigma_1^2)$, $(\alpha = 1, \cdots, N)$ and $y_{i\alpha}$ is $N(\eta_i, \sigma_2^2)$, $(i = 2, \cdots, p; \alpha = 1, \cdots, N)$. We can therefore obtain two sums of squares, u and v, sufficient for σ_1^2 and σ_2^2 and distributed independently as $\sigma_1^2 \chi_a^2$ and $\sigma_2^2 \chi_b^2$ where the degrees of freedom a and b depend on our knowledge of μ. To write u and v conveniently, we first observe that

$$\sum_{i=1}^{p} \sum_{\alpha=1}^{N} y_{i\alpha}^2 = \operatorname{tr} YY' = \operatorname{tr} XX' = \sum_{i=1}^{p} \sum_{\alpha=1}^{N} x_{i\alpha}^2,$$

$$\sum_{\alpha=1}^{N} y_{1\alpha}^2 = p^{-1}e'XX'e = p \sum_{\alpha=1}^{N} x_{\cdot\alpha}^2,$$

$$\sum_{i=1}^{p} y_{i\cdot}^2 = N^{-2}e'Y'Ye = N^{-2}e'X'Xe = \sum_{i=1}^{p} x_{i\cdot}^2,$$

$$y_{1\cdot} = N^{-1}p^{-1/2}e'Xe = p^{1/2}x_{\cdot\cdot}.$$

Precisely, we consider the following three cases:

(i) $\mu = 0$ and hence $\eta = \Delta\mu = 0$. Let $a = N$, $b = (p-1)N$,

$$u = \sum_{\alpha=1}^{N} y_{1\alpha}^2 = p \sum_{\alpha=1}^{N} x_{\cdot\alpha}^2,$$

$$v = \sum_{i=2}^{p} \sum_{\alpha=1}^{N} y_{i\alpha}^2 = \sum_{\alpha=1}^{N} \sum_{i=1}^{p} (x_{i\alpha} - x_{\cdot\alpha})^2.$$

(ii) μ completely unknown and hence $\eta = \Delta\mu$ is also completely unknown. Let $a = N - 1$, $b = (p-1)(N-1)$,

$$u = \sum_{\alpha=1}^{N} (y_{1\alpha} - y_{1\cdot})^2 = p \sum_{\alpha=1}^{N} (x_{\cdot\alpha} - x_{\cdot\cdot})^2,$$

$$v = \sum_{i=2}^{p} \sum_{\alpha=1}^{N} (y_{i\alpha} - y_{i\cdot})^2 = \sum_{\alpha=1}^{N} \sum_{i=1}^{p} (x_{i\alpha} - x_{i\cdot} - x_{\cdot\alpha} + x_{\cdot\cdot})^2.$$

(iii) $\mu = \omega e$, where ω is an unknown scalar, and hence $\eta = \omega\Delta e = \omega\sqrt{p}(1, 0, \cdots, 0)'$. Let $a = N - 1$, $b = (p-1)N$,

$$u = \sum_{\alpha=1}^{N} (y_{1\alpha} - y_{1\cdot})^2 = p \sum_{\alpha=1}^{N} (x_{i\alpha} - x_{\cdot\cdot})^2,$$

$$v = \sum_{i=2}^{p} \sum_{\alpha=1}^{N} y_{i\alpha}^2 = \sum_{i=1}^{p} \sum_{\alpha=1}^{N} (x_{i\alpha} - x_{\cdot\alpha})^2.$$

In each case u/σ_1^2, and v/σ_2^2 are independently distributed as χ_a^2 and χ_b^2, and it is easily shown that (u, v) is a complete sufficient statistic for (σ_1^2, σ_2^2). The three cases can thus be treated simultaneously.

3.2 *Derivation of the unbiased estimator*. The condition that $h(u, v)$ be unbiased is

$$\int_0^\infty \int_0^\infty h(u, v) u^{a/2-1} v^{b/2-1} e^{-\theta u - \phi v}\, du\, dv = \Gamma\left(\frac{a}{2}\right) \Gamma\left(\frac{b}{2}\right) \frac{\phi - \theta}{\phi + (p-1)\theta} \cdot \frac{1}{\theta^{a/2}\phi^{b/2}}, \tag{3.1}$$

where $\theta = 1/(2\sigma_1^2)$, $\phi = 1/(2\sigma_2^2)$. The right-hand side is the bivariate Laplace transform of

$$\begin{aligned}(3.2)\qquad &\left(\frac{b}{2}-1\right)\int_0^L [u-(p-1)y]^{\frac{a}{2}-1}(v-y)^{\frac{b}{2}-2}\,dy \\ &\qquad -\left(\frac{a}{2}-1\right)\int_0^L [u-(p-1)t]^{\frac{a}{2}-1}(v-t)^{\frac{b}{2}-1}\,dt,\end{aligned}$$

where $L = \min[u/(p-1), v]$, [4, p. 36 (Satz 12), p. 208 (9), p. 236 (87)]. Integrating the first term of (3.2) by parts and letting $z = u/[(p-1)v]$, we obtain

$$\begin{aligned}(3.3)\qquad h(u, v) = h^*(z) &= 1-\left(\frac{a}{2}-1\right)\frac{p}{p-1}\int_0^1 (1-zw)^{\frac{b}{2}-1}(1-w)^{\frac{a}{2}-2}\,dw \\ &= 1-\frac{p}{p-1}F\left(1, 1-\frac{b}{2}; \frac{a}{2}; z\right) \qquad \text{for } 0 \leqq z \leqq 1,\end{aligned}$$

$$\begin{aligned}(3.4)\qquad h(u, v) = h^*(z) &= 1-\left(\frac{a}{2}-1\right)\frac{p}{p-1}\frac{1}{z}\int_0^1 \left(1-\frac{1}{z}w\right)^{\frac{a}{2}-2}(1-w)^{\frac{b}{2}-1}\,dw \\ &= 1-\frac{2}{b}\left(\frac{a}{2}-1\right)\frac{p}{p-1}\frac{1}{z}F\left(1, 2-\frac{a}{2}; \frac{b}{2}+1; \frac{1}{z}\right) \\ &\qquad \text{for } z \geqq 1,\end{aligned}$$

[2, 2.12 (1)]. Integrating the second term of (3.2) by parts we obtain the following alternative to (3.4):

$$\begin{aligned}(3.5)\qquad h^*(z) &= \left(\frac{b}{2}-1\right)\frac{p}{p-1}\int_0^1 \left(1-\frac{1}{z}w\right)^{\frac{a}{2}-1}(1-w)^{\frac{b}{2}-2}\,dw - \frac{1}{p-1} \\ &= \frac{p}{p-1}F\left(1, 1-\frac{a}{2}; \frac{b}{2}; \frac{1}{z}\right) - \frac{1}{p-1} \qquad \text{for } z \geqq 1,\end{aligned}$$

[2, 2.12 (1)].

The conventional estimate of ρ is (e.g., [5] and [6]),

$$\begin{aligned}(3.6)\qquad r' &= \frac{p}{p-1}\,\frac{\sum_\alpha \sum_{i \neq j} (x_{i\alpha}-\hat{\mu}_i)(x_{j\alpha}-\hat{\mu}_j)}{\sum_\alpha \sum_i (x_{i\alpha}-\hat{\mu}_i)^2} \\ &= \frac{1}{p-1}\left(\frac{pu}{u+v}-1\right) = \frac{pz}{1+(p-1)z} - \frac{1}{p-1},\end{aligned}$$

where $\hat{\mu}$ is the appropriate estimate of μ in (i), (ii), or (iii). Now

$$(3.7)\qquad z = \frac{(p-1)\,r'+1}{(p-1)^2(1-r')},$$

which is a strictly increasing function of r'. Thus $h^*(z)$ is a function $H(r')$ of r'.

TABLE 2

Intraclass correlation coefficient, bivariate case, n degrees of freedom

$$H(r') = r'F(1/2, 1; n/2; 1 - r'^2)$$

2a. *Table of $H(r')$*

n	r' 0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	0	1	1	1	1	1	1	1	1	1	1
4	0	.182	.333	.462	.571	.667	.750	.824	.889	.947	1
6	0	.132	.259	.379	.490	.593	.688	.775	.856	.931	1
8	0	.120	.237	.351	.459	.563	.661	.753	.841	.923	1
10	0	.114	.227	.337	.444	.547	.647	.741	.832	.918	1
12	0	.111	.221	.329	.435	.538	.638	.734	.826	.915	1
14	0	.109	.217	.324	.429	.532	.631	.728	.822	.913	1
16	0	.108	.215	.321	.425	.527	.627	.724	.819	.911	1
18	0	.107	.213	.318	.422	.524	.624	.722	.817	.910	1
20	0	.106	.211	.316	.419	.521	.621	.719	.815	.909	1
22	0	.105	.210	.314	.417	.519	.619	.717	.814	.908	1
24	0	.105	.209	.313	.416	.517	.617	.716	.813	.907	1
26	0	.104	.208	.312	.414	.516	.616	.715	.812	.907	1
28	0	.104	.208	.311	.413	.515	.615	.713	.811	.906	1
30	0	.104	.207	.310	.412	.513	.614	.713	.810	.906	1
∞	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

2b. *Table of $H(r')/r'$*

n	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	∞	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1
4	2.000	1.818	1.667	1.538	1.429	1.333	1.250	1.176	1.111	1.053	1
6	1.333	1.322	1.296	1.262	1.224	1.185	1.146	1.107	1.070	1.034	1
8	1.200	1.196	1.185	1.169	1.149	1.126	1.102	1.076	1.051	1.025	1
10	1.143	1.141	1.134	1.124	1.110	1.095	1.078	1.059	1.040	1.020	1
12	1.111	1.110	1.105	1.098	1.088	1.076	1.062	1.048	1.033	1.017	1
14	1.091	1.090	1.086	1.080	1.073	1.063	1.052	1.041	1.028	1.014	1
16	1.077	1.076	1.073	1.068	1.062	1.054	1.045	1.035	1.024	1.012	1
18	1.067	1.066	1.063	1.059	1.054	1.047	1.040	1.031	1.021	1.011	1
20	1.059	1.058	1.056	1.053	1.048	1.042	1.035	1.027	1.019	1.010	1
22	1.053	1.052	1.050	1.047	1.043	1.038	1.032	1.025	1.017	1.009	1
24	1.048	1.047	1.045	1.043	1.039	1.034	1.029	1.023	1.016	1.008	1
26	1.043	1.043	1.042	1.039	1.036	1.032	1.027	1.021	1.014	1.007	1
28	1.040	1.040	1.038	1.036	1.033	1.029	1.024	1.019	1.013	1.007	1
30	1.037	1.037	1.035	1.033	1.031	1.027	1.023	1.018	1.012	1.006	1
∞	1	1	1	1	1	1	1	1	1	1	1

3.3. *Properties of the unbiased estimator.* For $\rho = 1, z = \infty, r' = 1$ with probability 1, and $h^*(\infty) = H(1) = 1$; for $\rho = -1/(p - 1), z = 0, r' = -1/(p - 1)$ with probability 1, and $h^*(0) = H(-1/(p - 1)) = -1/(p - 1)$. Thus in the two cases when Σ^* is singular, $h^*(z) = H(r') = \rho$ with probability 1. Furthermore, $h^*(z)$ is a strictly increasing function of z, since the integrand of (3.3) for $0 \leqq z \leqq 1$ and of (3.5) for $z \geqq 1$ is strictly monotone for each value of w, $0 < w < 1$. Consequently, $H(r')$ is a strictly increasing function of r' and $-1/(p-1) \leqq h^*(z) = H(r') \leqq 1$, which is the range of ρ.

As remarked before, $h^*(z) = H(r')$ is the unique minimum variance unbiased estimator of ρ.

We will now obtain the asymptotic distribution of $h^*(z)$. Note that z is distributed as

$$\frac{1+(p-1)\rho}{1-\rho} \frac{a}{b(p-1)} F_{a,b},$$

and that $\sqrt{(p-1)N/p}\,(F_{a,b}-1)$ is asymptotically $N(0, 1)$. Therefore, letting $z_0 = [1+(p-1)\rho]/[(p-1)^2(1-\rho)]$, the quantity,

$$\sqrt{\frac{(p-1)N}{p}}\left[z\frac{(p-1)^2(1-\rho)}{1+(p-1)\rho} - 1\right] = \sqrt{\frac{N}{p}}\frac{(p-1)^{5/2}(1-\rho)}{1+(p-1)\rho}(z-z_0)$$

is asymptotically $N(0, 1)$. But, by (3.3), denoting $N^{-5/6}$ by ϵ, we have, for $z \leqq 1$,

$$\begin{aligned} 1 - h^*(z) &= \frac{p}{p-1}\frac{N}{2}\int_0^\epsilon [1 - w - (p-1)zw]^{N/2}\,dw \\ &\qquad \cdot [1 + O(\epsilon^2)]^{N/2}[1 + O(\epsilon)] + NO(1-\epsilon)^{N/2} \\ &= \frac{p}{p-1}\frac{1}{1+(p-1)z} + O\left(\frac{1}{N}\right), \end{aligned}$$

uniformly in z. We obtain the same result for $z \geqq 1$ from (3.4). Therefore

$$h^*(z) = \rho + (p-1)^2(1-\rho)^2(z-z_0)/p + O_p(1/N).$$

Therefore

$$\sqrt{N}\,[h^*(z)-\rho] = \sqrt{N}\,[H(r')-\rho] \quad \text{is asymptotically} \quad N(0, \sigma^2),$$

where $\sigma^2 = (1-\rho)^2[1+(p-1)\rho]^2/[p(p-1)]$.

Expanding r' about z_0 in (3.6) we find

$$r' = h^*(z) + O_p(1/N) = H(r') + O_p(1/N),$$

so that r' is asymptotically equivalent to $H(r')$. Incidentally, we find that $\sqrt{N}(r'-\rho)$ is asymptotically $N(0, \sigma^2)$, with the same σ^2.

In order to facilitate the use of the unbiased estimator in the bivariate case with n degrees of freedom, i.e., case (i) or (ii) with $p = 2$, Table 2 gives $H(r')$ and, (for easier interpolation), $H(r')/r'$ for $r' = 0(.1)1$ and $a = b = n = 2(2)\,30$. In this case, $H(r') = H_n(r')$ is an odd function of r'. The computation was carried out by means of the recursive relation

$$H_n(r') = \frac{n-2}{n-3}\left[\frac{r'}{1-r'^2} - \frac{r'^2}{1-r'^2}H_{n-2}(r')\right],$$

together with the initial conditions

$$H_2(r') = 1, \qquad H_4(r') = \frac{2r'}{1+r'}, \qquad (r' > 0).$$

$H_n(r')$ was derived for $n \geqq 3$. For $n = 2$, the inversion in (3.1) must be carried out separately, and the result agrees with the final form of $h^*(z)$ in (3.4) and (3.6). The recursive relation is obtained by application of the relations [2, 2.8 (36) and (39)]. The same formulas give recursive relations for any values of p, a, b.

For the bivariate case with n degrees of freedom,

$$H(r') = r'F(\tfrac{1}{2}, 1; n/2; 1 - r'^2), \tag{3.8}$$

which is obtained from [2, 2.8 (36) and 2.11 (34)].

Approximations for $H(r')$ can be obtained from the expansion (2.2) applied to (3.8), which gives

$$\frac{H(r')}{r'} = 1 + \frac{1 - r'^2}{n} + \frac{3(1 - r'^2)^2}{n(n+2)} + O(n^{-3}). \tag{3.9}$$

This gives $H(r')/r'$ within .01 for $n \geqq 19$ or .001 for $n \geqq 57$ if two terms are included, and within .01 for $n \geqq 12$ or .001 for $n \geqq 26$ if three terms are included. As in (2.6), the neglected terms in (3.9) are all positive and decreasing in r'^2 and n.

The k that minimizes the maximum over r of the absolute difference between $H(r')/r'$ and $1 + (1 - r'^2)/(n - k)$ is, for large n, $6(-1 + \sqrt{2}) = 2.48$. This suggests the approximation

$$\frac{H(r')}{r'} = 1 + \frac{1 - r'^2}{n - 5/2}. \tag{3.10}$$

This is accurate within .01 for $n \geqq 10$ or .001 for $n \geqq 26$.

4. Multiple correlation coefficient. Suppose we have N independent observations on a $p + 1$-variate normal distribution with mean vector μ and covariance matrix Σ, and we wish to give an unbiased estimator of the squared multiple correlation

$$\rho^2 = \rho^2_{0.(12\cdots p)} = 1 - \mathcal{R}/\mathcal{R}_{00},$$

where $\mathcal{R}$ is the determinant of the correlation matrix and $\mathcal{R}_{00}$ is its first cofactor. We are concerned with the cases (i) μ known, Σ unknown, and (ii) all parameters unknown.

As in 2.1, we confine ourselves to functions of

$$r^2 = r^2_{0.(12\cdots p)} = 1 - R/R_{00},$$

where R is the determinant of the appropriate (to (i) or (ii)) sample correlation matrix and R_{00} is its first cofactor.

The condition that $I(r^2)$ be unbiased is

$$\sum_{k=0}^{\infty} \frac{\Gamma^2\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{p}{2} + k\right)} \frac{\rho^{2k}}{k!} \int_0^1 I(r^2)(r^2)^{((p-2)/2)+k}(1 - r^2)^{(n-p-1)/2}\, dr^2$$

$$= \Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{n}{2}\right)(1 - \rho^2)^{-n/2}\rho^2,$$

where $n = N$ and $(N - 1)$ in cases (i) and (ii). Following the methods of Sec. 2, we obtain

$$I(r^2) = 1 - \frac{n-2}{n-p}(1 - r^2)F\left(1, 1; \frac{n-p+2}{2}; 1 - r^2\right).$$

As usual, $I(r^2)$ is strictly increasing in r^2, and differs from it only by terms of order $1/N$, and it is the unique minimum variance unbiased estimator of ρ^2. Also $I(1) = 1$. However, $I(0) = -p/(n - p - 2)$. We cannot hope for a non-negative unbiased estimator, since there is no region in the sample space having zero probability for $\rho^2 = 0$ and positive probability for $\rho^2 > 0$. For the same reason there can be no positive unbiased estimator of ρ either.

REFERENCES

[1] Bateman Manuscript Project, California Institute of Technology, *Tables of Integral Transforms*, Vol. 1, A. Erdelyi, editor, McGraw-Hill Book Co., New York, 1954.
[2] Bateman Manuscript Project, California Institute of Technology, *Higher Transcendental Functions*, Vol. 1, A. Erdelyi, editor, McGraw-Hill Book Co., New York, 1953.
[3] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.
[4] G. Doetsch and Dietrich Voelker, *Die Zweidimensionale Laplace-Transformation*, Birkhauser Basel, Zurich, 1950.
[5] D. B. DeLury, "Note on correlations," *Ann. Math. Stat.*, Vol. 9 (1933), pp. 149–151.
[6] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin, London, 1947.