MULTIVARIATE CORRELATION MODELS WITH MIXED DISCRETE AND CONTINUOUS VARIABLES

By I. Olkin[1] and R. F. Tate[2]

Stanford University and Michigan State University; University of Washington

1. Introduction and summary. A model which frequently arises from experimentation in psychology is one which contains both discrete and continuous variables. The concern in such a model may be with finding measures of association or with problems of inference on some of the parameters.

In the simplest such model there is a discrete variable x which takes the values 0 or 1, and a continuous variable y. Such a random variable x is often used in psychology to denote the presence or absence of an attribute. Point-biserial correlation, which is the ordinary product-moment correlation between x and y, has been used as a measure of association. This model, when x has a binomial distribution, and the conditional distribution of y for fixed x is normal, was studied in some detail by Tate [13].

In the present paper, we consider a multivariate extension, in which $x = (x_0, x_1, \cdots, x_k)$ has a multinomial distribution, and the conditional distribution of $y = (y_1, \cdots, y_p)$ for fixed x is multivariate normal.

2. Outline. Consider a random sample of n independent vectors (x_α, y_α), $\alpha = 1, \cdots, n$, where x has a binomial distribution, $b(1, p)$. The conditional distributions of $(y \mid x = 1)$ and $(y \mid x = 0)$ are assumed to be $\mathfrak{N}(\mu_1, \sigma^2)$ and $\mathfrak{N}(\mu_0, \sigma^2)$, respectively. If we define $\Delta = (\mu_1 - \mu_0)/\sigma$, then

$$\rho_{x,y} = \Delta[pq/(1 + pq\Delta^2)]^{\frac{1}{2}}.$$

Thus, studying ρ involves studying induced relations between means; for example, $\mu_1 = \mu_0$ if and only if $\rho_{x,y} = 0$. The exact and asymptotic distributions of r_{xy} were obtained by Tate [13].

We are now concerned with a multivariate analog of this model. Let $(y_{1\alpha}, \cdots, y_{p\alpha}, x_{0\alpha}, \cdots, x_{k\alpha})$, $\alpha = 1, \cdots, n$, be a sequence of independent random vectors, where $(x_0, \cdots, x_k)$ has the multinomial distribution,

$$f(x_0, \cdots, x_k) = p_0^{x_0} p_1^{x_1} \cdots p_k^{x_k}; \qquad x_m = 0, 1;$$

$$\sum_0^k x_m = 1, \qquad 0 < p_m < 1, \qquad \sum p_m = 1.$$

The conditional distribution of $y = (y_1, \cdots, y_p)$ given $x_m = 1$ is assumed to be $\mathfrak{N}(\mu^{(m)}, \Sigma)$, that is, p-variate normal with mean vector $\mu^{(m)} = (\mu_{1m}, \cdots, \mu_{pm})$, $m = 0, 1, \cdots, k$, and positive definite covariance matrix Σ.

Received March 15, 1960; revised November 17, 1960.

[1] Research sponsored in part by the Office of Naval Research at Stanford University, and in part by the Office of Ordnance Research at Michigan State University.

[2] Research sponsored in part by the Office of Naval Research and in part by the National Science Foundation, Grant 14284, at the University of Washington.

As in the univariate case, the vanishing of various correlations, for example multiple correlation coefficients, induces certain constraints on the means. In Section 3, we give a number of relations between correlation coefficients and means. It will appear that the square of a correlation coefficient may act as a measure of dispersion among the possible multivariate normal conditional distributions.

For convenience of development as well as clarity, we consider separately the cases (i) $k = 1$, $p > 1$, (ii) $k > 1$, $p = 1$, (iii) $k > 1$, $p > 1$. In connection with Case (i), it will be shown that $\rho^2_{x_1(y_1,\cdots,y_p)}$ is closely related to the distance function of Mahalanobis [7]. Section 4, dealing with the relevant distribution theory for Case (i), will exhibit the relationship between $r^2_{x_1(y_1,\cdots,y_p)}$ and the T^2 statistic of Hotelling [5], and will contain the exact and asymptotic distributions for $r_{x_1(y_1,\cdots,y_p)}$. The method of derivation for the asymptotic distribution constitutes something of a departure from the usual approach, since the statistic involved is a function of sample means, but the classical method of Cramér ([2], Section 27.7) is not used, because it would involve too much calculation. The resulting distribution is formally identical to that obtained by Tate [13] for the ordinary correlation coefficient, which is an altogether surprising result.

Section 5 presents the distribution theory related to Case (ii), including derivations for the exact and asymptotic distributions of partial correlation coefficients, in addition to the main discussion of $r_{y_1(x_1,\cdots,x_k)}$. Unfortunately, the multiple correlation coefficient has a distribution which contains nuisance parameters, that is, parameters other than $p_0, p_1, \cdots, p_k$ of the x distribution, and the population multiple correlation coefficient. Moreover, this difficulty does not disappear in the limit. Cramér's method, referred to in the last paragraph, is used to advantage here.

Canonical correlations are introduced in Section 3, and serve to give a unified approach for our three cases. In the general case $k > 1$, $p > 1$, however, it is difficult to obtain results. An effort is made in Section 6 to indicate the problems involved. The vector correlation ρ_v between the vectors x and y, which is also introduced in Section 3, although theoretically inferior to canonical correlations has the property that more can be accomplished with the sampling theory for its estimate r_v. In Section 6 it is shown that r_v is essentially distributed as a U-statistic of Wilks [15]. A distribution of Rao [10] is important in this connection.

Throughout the paper estimates will be the natural sample counterparts of the parameters which they estimate. They can all be obtained by the method of maximum likelihood, and this is shown in Section 7.

Section 8 contains a summary of procedures developed throughout the paper, together with examples of situations in which they would be appropriate.

Moustafa [8] has made a detailed study of models employing a multivariate normal conditional distribution with one or more multinomial conditioning vectors. He considers cases more general than ours, and employs the asymptotic chi-square property of $-2 \log \lambda$ to perform his tests; correlation is not mentioned.

3. Relations between correlation coefficients and means.

3.1. *Model and preliminaries.* Consider the model[3]

$$(y_1, \cdots, y_p \mid x_m = 1, x_\nu = 0, \nu \neq m = 0, 1, \cdots, k) \sim \mathfrak{N}(\mu^{(m)}, \Sigma),$$

and suppose that the conditional means and covariances are given as follows:

Means

	x_0	x_1	$\cdots$	x_k
y_1	μ_{10}	μ_{11}	$\cdots$	μ_{1k}
$\vdots$		.	.	.
y_p	μ_{p0}	μ_{p1}	$\cdots$	μ_{pk}

Vectors

$$\mu^{(0)} = (\mu_{10}, \cdots, \mu_{p0})$$
$$\cdots \quad \cdots \quad \cdots$$
$$\mu^{(k)} = (\mu_{1k}, \cdots, \mu_{pk})$$

Covariances

	y_1	$\cdots$	y_p	x_0	x_1	$\cdots$	x_k
y_1	ψ_{11}	$\cdots$	ψ_{1p}	δ_{10}	δ_{11}	$\cdots$	δ_{1k}
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_p	ψ_{p1}	$\cdots$	ψ_{pp}	δ_{p0}	δ_{p1}	$\cdots$	δ_{pk}
x_0	δ_{10}	$\cdots$	δ_{p0}	γ_{00}	γ_{01}	$\cdots$	γ_{0k}
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	δ_{1k}	$\cdots$	δ_{pk}	γ_{k0}	γ_{k1}	$\cdots$	γ_{kk}

$$\equiv \begin{pmatrix} \Psi & \Delta \\ \Delta' & \Gamma \end{pmatrix}.$$

The unconditional moments are:

$$Ey_i = \sum_{m=0}^{k} E(y_i \mid x_m = 1)p_m = \sum_{m=0}^{k} \mu_{im}p_m \equiv \mu_{i\cdot}, \tag{3.1}$$

$$\begin{aligned} Ey_iy_j = \sum_{m=0}^{k} E(y_iy_j \mid x_m = 1)p_m &= \sum_{m} (\sigma_{ij} + \mu_{im}\mu_{jm})p_m \\ &= \sigma_{ij} + \sum_{m} p_m\mu_{im}\mu_{jm}. \end{aligned} \tag{3.2}$$

Hence,

$$\psi_{ij} = \sigma_{ij} + \sum_{m} p_m(\mu_{im} - \mu_{i\cdot})(\mu_{jm} - \mu_{j\cdot}), \tag{3.3}$$

$$\delta_{im} = p_m(\mu_{im} - \mu_{i\cdot}), \tag{3.4}$$

$$\gamma_{mm} = p_mq_m\,; \qquad \gamma_{m\nu} = -p_mp_\nu\,, \qquad (m \neq \nu), \qquad q_m = 1 - p_m\,. \tag{3.5}$$

If we let $U = (u_{im}) \equiv (\mu_{im} - \mu_{i\cdot})$, $i = 1, 2, \cdots, p$, $m = 0, 1, \cdots, k$; $p = (p_0, p_1, \cdots, p_k)$, $D_p = \operatorname{diag}(p_0, p_1, \cdots, p_k)$, then (3.3)-(3.5) can be written as

$$\Psi = \Sigma + UD_pU', \tag{3.6}$$

$$\Delta = UD_p\,, \tag{3.7}$$

$$\Gamma = D_p - p'p. \tag{3.8}$$

Note that $\Delta e' = \Gamma e' = 0$, where $e = (1, 1, \cdots, 1)$: $1 \times k + 1$. Moreover,

[3] $x \sim F(x)$ means that x is distributed according to the d.f. $F(x)$, and $x(n) \to F(x)$ means that the asymptotic d.f. of $x(n)$ is $F(x)$.

$\Gamma w' = 0$ if and only if w is a scalar multiple of e. Finally, $Up' = UD_p e' = \Delta e' = 0$, and from (3.7) and (3.8) we have

$$\Delta = U\Gamma. \tag{3.9}$$

3.2. *Relations between canonical correlations and means.* Consider the matrix

$$V_\lambda = \begin{pmatrix} -\lambda\Psi & \Delta \\ \Delta' & -\lambda\Gamma \end{pmatrix}.$$

The canonical correlations (introduced by Hotelling [6]) are defined as the numbers λ to each of which corresponds a non-trivial vector $\zeta = (\eta, \xi) = (\eta_1, \eta_2, \cdots, \eta_p; \xi_0, \xi_1, \cdots, \xi_k)$, with $V_\lambda \zeta' = 0$. Since $\Delta e' = \Gamma e' = 0$, will be trivial if $\eta = 0$ and ξ is a scalar multiple of e. Thus, if $\lambda \neq 0$, then λ is a canonical correlation if there exist vectors $\eta \neq 0$ and ξ such that

$$\Delta\xi' = \lambda\Psi\eta', \ \Delta'\eta' = \lambda\Gamma\xi'. \tag{3.10}$$

Note that if $\eta = 0$, then $\Gamma e' = 0$, in which case ξ would be a scalar multiple of e.

LEMMA 3.1. *The non-zero canonical correlations are precisely the non-zero roots of*

$$|UD_pU' - \theta\Sigma| = 0, \qquad \theta = \lambda^2/(1 - \lambda^2). \tag{3.11}$$

REMARK. Since $\Psi = UD_pU' + \Sigma$ is positive definite, $\theta \neq -1$, and $\lambda^2 = \theta/(1 + \theta)$ is well-defined.

PROOF. Given $|V_\lambda| = 0$, with $\lambda \neq 0$, $\eta \neq 0$, and using (3.6), (3.7), (3.9), and (3.10), we have

$$\lambda^{-1}UD_pU' = \lambda^{-1}U\Delta'\eta' = U\Gamma\xi' = \Delta\xi' = \lambda\Psi\eta' = \lambda(UD_pU' + \Sigma)\eta',$$

which implies (3.11).

Conversely, suppose that (3.11) holds with $\theta \neq 0$. Then there exists $\eta \neq 0$ such that

$$(1 - \lambda^2)UD_pU'\eta' = \lambda^2\Sigma\eta'.$$

It now suffices to prove that there exists a vector ξ satisfying

$$\Gamma\xi' = \lambda^{-1}\Delta'\eta',$$

for then $\Delta\xi' = \lambda\Psi\eta'$ by an argument similar to the above. But such a vector does exist, and is given by $\xi = e$; in which case $\Gamma e' = 0$ and $e\Delta'\eta' = 0$ for all η ||.

THEOREM 3.2. *The canonical correlations are zero if and only if*

$$\mu^{(0)} = \mu^{(1)} = \cdots = \mu^{(k)}.$$

PROOF. Clearly $\mu_{im} = \mu_{i\nu}$ for all i, m, and ν holds if and only if $U = 0$. If $U = 0$, then (3.11) implies that $\theta = 0$. Conversely, if $\theta = 0$, then $\Sigma^{-\frac{1}{2}}UD_pU'\Sigma^{-\frac{1}{2}} = 0$, so that $\Sigma^{-\frac{1}{2}}UD_p^{\frac{1}{2}} = 0$, which in turn implies that $U = 0$. ||.

If $k > 1$, $p = 1$, then $U = (\mu_{10} - \mu_{1\cdot}, \cdots, \mu_{1k} - \mu_{1\cdot})$, and there is only one non-zero root: namely, the multiple correlation coefficient. (The first subscript

is not needed in this discussion, and we omit it.) Hence,

$$\rho_0^2 \equiv \rho^2_{y(x_1,\cdots,x_k)} = \frac{\sum_{m=0}^{k} (\mu_m - \mu)^2 p_m/\sigma_{11}}{1 + \sum_{m=0}^{k} (\mu_m - \mu)^2 p_m/\sigma_{11}}. \tag{3.12}$$

If we define $\Delta_m = (\mu_m - \mu)(\sigma_{11})^{-\frac{1}{2}}$, and $\delta = \sum_0^k p_m \Delta_m^2$ then

$$\rho_0^2 = \sum_0^k p_m \Delta_m^2 \Big/ \left(1 + \sum_0^k p_m \Delta_m^2\right) = \delta/(1 + \delta). \tag{3.13}$$

Also,

$$\rho^2_{0m} \equiv \rho^2_{yx_m} = \frac{p_m \Delta_m^2}{q_m(1 + \delta)},$$

so that

$$\rho_0^2 = \sum_0^k q_m \rho^2_{0m}. \tag{3.14}$$

The multiple correlation between y and a subset $(x_1, \cdots, x_l)$ of $(x_l, \cdots, x_k)$ may also be computed: namely

$$\rho^2_{y(x_1,\cdots,x_l)} = \left(\sum_0^l p_m(\mu_m - \mu)^2 + \left[\sum_0^l p_m(\mu_m - \mu)\right]^2 \left[1 - \sum_0^l p_m\right]^{-1}\right) \Big/ \psi_{11}. \tag{3.15}$$

Now suppose $k = 1$, $p > 1$; then $U = (p_1 d', -p_0 d')$, where $d = (\mu^{(0)} - \mu^{(1)})$. Hence, $UD_pU' = p_0 p_1 d'd$, and (3.11) has one non-zero root, $\theta = p_0 p_1 d\Sigma^{-1}d'$, so that $\lambda^2 = \theta/(1 + \theta)$ is equal to

$$\rho^2_{x_1(y_1,\cdots y_p)} = \frac{p_0 p_1(\mu^{(0)} - \mu^{(1)})\Sigma^{-1}(\mu^{(0)} - \mu^{(1)})'}{1 + p_0 p_1(\mu^{(0)} - \mu^{(1)})\Sigma^{-1}(\mu^{(0)} - \mu^{(1)})'}. \tag{3.16}$$

We now find conditions for which the partial correlation coefficient vanishes.

THEOREM 3.3. *Let* $k > 1$, $p = 1$ *and* $\rho^2_{0(m+1)\cdot} \equiv \rho^2_{yx_{m+1}\cdot(x_1,\cdots,x_m)}$; *then* $\rho_{0(m+1)\cdot} = 0$ *if and only if*

$$\mu_{m+1} = \sum_{m+2}^{k} p_\nu \mu_\nu \Big/ \sum_{m+2}^{k} p_\nu. \tag{3.17}$$

PROOF. From

$$1 - \rho^2_{0(1,\cdots,m,m+1)} = [1 - \rho^2_{0(1,\cdots,m)}][1 - \rho^2_{0(m+1)\cdot}],$$

we have that $\rho^2_{0(m+1)\cdot} = 0$ if and only if $\rho^2_{0(1,\cdots,m+1)} = \rho^2_{0(1,\cdots,m)}$; from (3.19) this condition holds if and only if

$$p_{m+1}(\mu_{m+1} - \mu)^2 + \frac{\left[\sum_0^{m+1} p_\nu(\mu_\nu - \mu)\right]^2}{1 - \sum_0^{m+1} p_\nu} - \frac{\left[\sum_0^{m} p_\nu(\mu_\nu - \mu)\right]^2}{1 - \sum_0^{m} p_\nu} = 0.$$

Simplification yields

$$\left(\mu_{m+1} - \mu + \left[1 - \sum_0^m p_\nu\right] \sum_1^m p_\nu(\mu_\nu - \mu)\right) = 0$$

which is equivalent to (3.17). ||

3.3. *Vector correlation.* If we regard correlation in our model as merely a measure of dispersion for the various $\mathfrak{N}(\mu^{(m)}, \Sigma)$ distributions, then we are led quite naturally to a consideration of vector correlation. This concept is due to Wilks [15], and is an extension of the correlation ratio to the multivariate case via the use of generalized variance. In the notation of our model the coefficient of vector correlation ρ_v^2 can be expressed as

$$\rho_v^2 = 1 - \frac{|\Sigma|}{|\Sigma + UD_p U'|}. \tag{3.18}$$

It is easy to see that (3.18) reduces to (3.12) and (3.16) for $k > 1$, $p = 1$ and $k = 1$, $p > 1$, respectively.

4. Distribution theory for the case $k = 1, p > 1$. Let $(y_{1\alpha}, \cdots, y_{p\alpha}, x_{0\alpha}, x_{1\alpha})$, $\alpha = 1, \cdots, n$, be n independent random vectors, with conditional distribution $(y_{1\alpha}, \cdots, y_{p\alpha} | x_{m\alpha} = 1) \sim \mathfrak{N}(\mu^{(m)}, \Sigma)$, $m = 0, 1$. It will be convenient to define the following statistics:

$$n_0 = \sum_\alpha x_{0\alpha}, \qquad n_1 = \sum_\alpha x_{1\alpha}, \qquad n = n_0 + n_1,$$

$$\bar{y}_j = \sum_\alpha y_{j\alpha}/n, \qquad \bar{y}_j^{(0)} = \sum_\alpha y_{j\alpha}x_{0\alpha}/n_0, \qquad \bar{y}_j^{(1)} = \sum_\alpha y_{j\alpha}x_{1\alpha}/n_1,$$

$$\bar{y}^{(0)} = (\bar{y}_1^{(0)}, \cdots, \bar{y}_p^{(0)}), \qquad \bar{y}^{(1)} = (\bar{y}_1^{(1)}, \cdots, \bar{y}_p^{(1)}), \qquad S = (s_{ij}) : p \times p,$$

$$s_{ij} = \sum_{m=0}^{1} \sum_{\lambda=1}^{n_m} (y_{i\lambda}^{(m)} - \bar{y}_i^{(m)})(y_{j\lambda}^{(m)} - \bar{y}_j^{(m)})/(n-2), \qquad i, j = 1, \cdots, p,$$

where $\{y_{i\lambda}^{(m)}\}$ is the subset of $(y_{i1}, \cdots, y_{in})$ for which the corresponding elements of $(x_{m1}, \cdots, x_{mn})$ are equal to unity.

Corresponding to (3.16) we have as an estimate of $\rho^2 = \rho^2_{x_1(y_1, \cdots, y_p)}$,

$$\begin{aligned} r^2 \equiv r^2_{x_1(y_1, \cdots, y_p)} &= \frac{(n_0n_1/n(n-2))(\bar{y}^{(1)} - \bar{y}^{(0)})S^{-1}(\bar{y}^{(1)} - \bar{y}^{(0)})'}{1 + (n_0n_1/n(n-2))(\bar{y}^{(1)} - \bar{y}^{(0)})S^{-1}(\bar{y}^{(1)} - \bar{y}^{(0)})'} \\ &= T^2/(n - 2 + T^2) \end{aligned} \tag{4.1}$$

where

$$T^2 = (n_0n_1/n)(\bar{y}^{(1)} - \bar{y}^{(0)})S^{-1}(\bar{y}^{(1)} - \bar{y}^{(0)})'.$$

We can now state the following

THEOREM 4.1: $Q = [(n - p - 1)/p]\,[r^2/(1 - r^2)]$ *is distributed as a mixture of noncentral* $F_{p,n-p-1}(\tau^2)$ *distributions, with mixing coefficients*

$$\binom{n}{n_0} p_0^{n_0} p_1^{n_1},$$

and parameter

$$\tau^2 = \frac{n_0 n_1}{n p_0 p_1}\left(\frac{\rho^2}{1-\rho^2}\right).$$

PROOF: First note that from (4.1)

$$\left(\frac{r^2}{1-r^2}\right)\frac{n-p-1}{p} = \frac{T^2}{n-2}\left(\frac{n-p-1}{p}\right).$$

The conditional distribution of this statistic can be obtained immediately by applying a method of Bowker's (see Anderson [1], Theorem 5.2.2): If $T^2 = YS^{-1}Y'$, where $Y \sim \mathfrak{N}(\nu, \Sigma)$, Σ: $p \times p$, and $aS = \sum_1^a Z_\alpha' Z_\alpha$, where $Z_\alpha \sim \mathfrak{N}(0, \Sigma)$, then

$$\frac{T^2}{a}\left(\frac{a-p+1}{p}\right) \sim F_{p,a-p+1}(\tau^2),$$

where $\tau^2 = \nu\Sigma^{-1}\nu'$. Now, make the correspondence $a = n-2$, $Y = (n_0 n_1/n)^{\frac{1}{2}}(\bar{y}^{(1)} - \bar{y}^{0})$, $\nu = (n_0 n_1/n)^{\frac{1}{2}}(\mu^{(1)} - \mu^{(0)})$, and use (3.16) to compute τ^2. Forming the mixture, we have

$$f(Q) = \sum_{n_0=0}^{n}\sum_{h=0}^{\infty}\frac{e^{-\frac{1}{2}n_0 n_1 \rho^2/n p_0 p_1(1-\rho^2)}}{h!}\left(\frac{n_0 n_1}{2n p_0 p_1}\frac{\rho^2}{1-\rho^2}\right)^h f^{(Q)}_{p+2h,n-p-1}\binom{n}{n_0}p_0^{n_0}p_1^{n_1}. \|$$

Note that if $\rho^2 = 0$, T^2 has the T^2 distribution of Hotelling [5]; and, since the n_0 and n_1 sum out, Q has an ordinary $F_{p,n-p-1}$ distribution.

4.1. *Asymptotic distribution.* Define

$$g(r) = \frac{r^2}{1-r^2} \sim \frac{p}{n-p-1}F_{p,n-p-1}(\tau^2),$$

$h(r) = \{g(r)/[1 + g(r)]\}^{\frac{1}{2}} = r$. Then

$$h(r) \to \mathfrak{N}\left(h(\rho), \frac{1}{n}\left(\frac{dh}{dg}\bigg|_\rho\right)^2 \lim_{n\to\infty} nV(g(r))\right). \tag{4.2}$$

We now determine the various factors:

$$h(\rho) = \rho. \qquad g(\rho) = \rho^2/(1-\rho^2), \qquad \left(\frac{dh}{dg}\bigg|_\rho\right)^2$$
$$= [4g(\rho)(1+g(\rho))^3]^{-1} = (1-\rho^2)^4/4\rho^2,$$

$$EF_{c,d}(\tau^2) = \frac{d}{d-2}\left(1 + \frac{E\tau^2}{c}\right), \tag{4.3}$$

$$E[F_{c,d}(\tau^2)]^2 = \frac{d^2(c+2)}{c(d-2)(d-4)}\left[1 + \frac{E\tau^4 + 2(c+2)(E\tau^2)}{c(c+2)}\right], \tag{4.4}$$

$$E\tau^2 = [np_0 p_1(1-\rho^2)]^{-1}\rho^2 E n_0 n_1 = (n-1)g(\rho),$$

$$E\tau^4 = (np_0 p_1)^{-2}g^2(\rho)En_0^2 n_1^2$$
$$= (n-1)g^2(\rho)[(n-2)(n-3)p_0 p_1 + n - 1]/np_0 p_1.$$

With $c = p$, $d = n - p - 1$, after considerable calculation we obtain

$$\lim_{n\to\infty} nV(g(r)) = g^2(\rho)\,\frac{[1 - 2p_0\,p_1]}{p_0\,p_1} + 4g(\rho),$$

which upon substitution in (4.2) yields

THEOREM 4.2:

$$r \to \mathfrak{N}\left(\rho, \frac{4p_0\,p_1 - \rho^2(6p_0\,p_1 - 1)}{4np_0\,p_1}\,(1 - \rho^2)^2\right). \tag{4.5}$$

It is rather surprising that the asymptotic variance is independent of p, the number of variates in $(y_1, \cdots, y_p)$, except insofar as it affects $\rho^2 \equiv \rho^2_{x_1(y_1,\cdots,y_p)}$. As a consequence, (4.5) is identical in form with the result of Tate ([13], Th. 1) for $p = 1$. Thus, we can apply some of the results of that paper. In particular $V_\infty(r)$ has a minimum for each ρ when $p_0 = p_1 = \frac{1}{2}$, in which case $V_\infty(r) = (1 - \rho^2)^2(2 - \rho^2)/2n$. By a variance stabilizing transformation we obtain (when $p_0 = \frac{1}{2}$)

$$\tanh^{-1}[r^2(2 - r^2)]^{\frac{1}{4}} \sim \mathfrak{N}\{\tanh^{-1}[\rho^2(2 - \rho^2)]^{\frac{1}{4}}, 2/n\}.$$

In a recent paper, Hooper [4] considered the following model. Let $(y_\alpha, x_{1\alpha}, \cdots, x_{\Lambda\alpha})$, $\alpha = 1, \cdots, n$, be n independent random vectors, with

$$y_\alpha = \sum_\lambda \pi_\lambda x_{\lambda\alpha} + u_\alpha, \qquad x_{\lambda\alpha} = \xi_{\lambda\alpha} + \omega_{\lambda\alpha}, \qquad \lambda = 1, \cdots, \Lambda; \qquad \alpha = 1, \cdots, n,$$

where $\xi_{\lambda\alpha}$ are real numbers, $(\omega_{1\alpha}, \cdots, \omega_{\Lambda\alpha})$ are independent observations from a Λ-variate normal distribution with zero mean, independent of μ_α, $\alpha = 1, \cdots, n$, which are independent normal variates with zero means. If $\sum \xi^2_{\lambda\alpha} = 1$, then the asymptotic variance of the multiple correlation coefficient is $(1 - \rho^2)^2(2 - \rho^2)/2n$, which is the same as $V_\infty(r)$ when $p_0 = p_1 = \frac{1}{2}$. Although the results are the same, the connection, if any, between the two models is obscure.

5. Distribution theory for the Case $k > 1$, $p = 1$. Let $(y_\alpha, x_{0\alpha}, \cdots, x_{k\alpha})$, $\alpha = 1, \cdots, n$, be n independent random vectors, where the conditional distribution of $(y_\alpha | x_{m\alpha} = 1) \sim \mathfrak{N}(\mu_m, \sigma^2)$, $m = 0, 1, \cdots, k$. It will be convenient to define the following statistics:

$$n_m = \sum_\alpha x_{m\alpha}, \qquad n = \sum_0^k n_m, \qquad \bar{y} = \sum y_\alpha/n,$$

$$\bar{y}^{(m)} = \sum_\alpha y_\alpha x_{m\alpha}/n_m = \sum_{\lambda=1}^{n_m} y_\lambda^{(m)}/n_m,$$

where $\{y_\lambda^{(m)}\}$ is the subset of $(y_1, \cdots, y_n)$ for which the corresponding elements of $(x_{m1}, \cdots, x_{mn})$ are equal to unity.

$$\overline{x_m y} = \sum_\alpha y_\alpha x_{m\alpha}/n, \qquad \bar{x}_m = \sum x_{m\alpha}/n,$$

$$\mu = \sum_0^k p_m \mu_m, \qquad \Delta_m = (\mu_m - \mu)/\sigma, \qquad \delta = \sum_0^k p_m \Delta_m^2.$$

5.1. *Multiple Correlation Coefficient: Exact distribution.* Corresponding to (3.12), (3.13), (3.14), we have several equivalent forms for the estimator of ρ_0^2 :

$$r_0^2 \equiv r^2_{y(x_1,\cdots,x_k)} = n\sum_{m=0}^{k} s_m^2/n_m s^2, \tag{5.1}$$

where $s^2 = \sum_\alpha (y_\alpha - \bar{y})^2/n$, $s_m = (1/n)\sum_\alpha (y_\alpha - \bar{y})(x_{m\alpha} - n_m/n) = n_m(\bar{y}^{(m)} - \bar{y})/n$,

$$r_0^2 = (\overline{y^2} - \bar{y}^2)^{-1}\sum_0^k (\overline{x_m y} - \bar{x}_m\bar{y})^2/\bar{x}_m\,, \tag{5.2}$$

$$r_0^2 \equiv \sum_0^k (1 - \bar{x}_m) r^2_{0m}\,, \tag{5.3}$$

where $r^2_{0m} = r^2_{yx_m}$.

Using (5.1) we obtain

$$\frac{r_0^2}{1 - r_0^2} = \frac{\sum_0^k n_m(\bar{y}^{(m)} - \bar{y})^2}{\sum_1^n (y_\alpha - \bar{y})^2 - \sum_0^k n_m(\bar{y}^{(m)} - \bar{y})^2} = \frac{\sum_0^k n_m(\bar{y}^{(m)} - \bar{y})^2}{\sum_{m=0}^k \sum_{\lambda=1}^{n_m} (y_\lambda^{(m)} - \bar{y}^{(m)})^2}. \tag{5.4}$$

In view of the above we have, analogous to Theorem 4.1,

THEOREM 5.1. *The statistic*

$$Z = \left(\frac{n - k - 1}{k}\right)\frac{r_0^2}{1 - r_0^2}$$

is distributed as a mixture of noncentral $F_{k,n-k-1}(\tau^2)$ distributions, with mixing coefficients

$$\frac{n!}{n_0!n_1!\cdots n_k!}\, p_0^{n_0} p_1^{n_1}\cdots p_k^{n_k},$$

and parameter $\tau^2 = \sum_0^k n_m\Delta_m^2$.

PROOF. Follow the same type of argument as in the proof of Theorem 4.1. Note that again for this case we have $\tau^2 = 0$ whenever $\rho_0^2 = 0$, and hence $Z \sim F_{k,n-k-1}$, which is a well-known result (see Fisher [3]). ||

5.5.1 *Asymptotic Distribution.* We now need certain computations for moments. Using

$$Ex_m^a x_\nu^b y^c = \begin{cases} \sum_0^k p_m E(y^c|x_m = 1), & \text{if } a = b = 0, \\ p_m E(y^c|x_m = 1), & m = \nu \text{ or } b = 0, \end{cases}$$

we find that

$$V(x_m) = p_m q_m, \qquad \operatorname{Cov}(x_m, x_\nu) = -p_m p_\nu (m \neq \nu),$$
$$V(y) = 1 + \delta, \qquad V(y^2) = \sum p_m \Delta_m^4 - \delta^2 + 4\delta + 2,$$
$$\operatorname{Cov}(y, y^2) = \sum p_m \Delta_m^3, \qquad V(x_m y) = p_m(1 + q_m \Delta_m^2),$$
$$\operatorname{Cov}(x_m, x_m y) = p_m q_m \Delta_m, \qquad \operatorname{Cov}(x_m, x_\nu y) = -p_m p_\nu \Delta_m (m \neq \nu),$$
$$\operatorname{Cov}(x_m, y) = p_m \Delta_m, \qquad \operatorname{Cov}(x_m, y^2) = p_m(\Delta_m^2 - \delta),$$
$$\operatorname{Cov}(x_m y, y) = p_m(1 + \Delta_m^2), \qquad \operatorname{Cov}(x_m y, x_\nu y) = -p_m p_\nu \Delta_m \Delta_\nu (m \neq \nu),$$
$$\operatorname{Cov}(x_m y, y^2) = p_m \Delta_m (\Delta_m^2 - \delta + 2).$$

If we write r_0^2 as in (5.2), we have a function of sample moments, and we can expand about the population moments (see Cramér [2], p. 353). This leads to the following asymptotic result.

THEOREM 5.2.

$$r_0 \to \mathfrak{N}\left(\rho_0, \frac{1}{n} \frac{\sum p_m \Delta_m^4 + \delta^2 + 4\delta}{4\delta(1 + \delta)^3}\right).$$

Alternative forms for the asymptotic variance are

$$V_\infty(r) = \frac{(1 - \rho_0^2)^2}{n}\left[(1 - \rho_0^2)^2 \frac{\sum p_m \Delta_m^4}{4\rho_0^2} + 1 - \frac{3}{4}\rho_0^2\right],$$
$$V_\infty(r) = \frac{(1 - \rho_0^2)^2}{n}\left[\frac{\sum q_m^2 \rho_{0m}^4 / p_m}{4\rho_0^2} + 1 - \frac{3}{4}\rho_0^2\right],$$

since $\rho_{0m}^2 = p_m \Delta_m^2 (1 - \rho_0^2)/q_m$. The term $\sum q_m^2 \rho_{0m}^4 / p_m$ contains the nuisance parameters; we can, however, look at some bounds. We find that

$$\rho_0^4 \leqq p_m^{-1} \sum q_m^2 \rho_{0m}^4 \leqq \rho_0^4 (\sum p_m^{-1} - 2k - 1).$$

The right inequality follows from $\rho_{0m}^4 \leqq \rho_0^4$ and the left inequality follows from the Cauchy inequality

$$\sum [(q_m \rho_{0m}^2) p_m^{-1} (q_m \rho_{0m}^2)] \sum p_m \geqq (\sum q_m \rho_{0m}^2)^2 = \rho_0^4.$$

Thus,

$$\frac{(1 - \rho_0^2)^2 (2 - \rho_0^2)}{2n} \leqq V_\infty \leqq \frac{(1 - \rho_0^2)^2}{2n}(2 - \rho_0^2) \cdot \left[1 + \frac{\rho_0^2}{2(2 - \rho_0^2)} \sum_0^k \left(\frac{1}{p_m} - 2\right)\right].$$

For fixed ρ_0^2 and k the right-hand side takes on its minimum value of

$$\frac{(1 - \rho_0^2)^2}{2n}(2 - \rho_0^2)\left[1 + \frac{\rho_0^2(k^2 - 1)}{2(2 - \rho_0^2)}\right]$$

when $p_m \equiv 1/(k+1)$, which in turn of course reduces to the left-hand side, and the result of Tate [13], when $k = 1$.

5.2. *Partial correlation coefficient: Exact distribution.* We first find an expression for $r^2_{0(m+1)\cdot} \equiv r^2_{yx_{m+1}\cdot(x_0,\cdots,x_m)}$

THEOREM 5.3.

$$\frac{r^2_{0(m+1)\cdot}}{1-r^2_{0(m+1)\cdot}} = \frac{\left(n_{m+1}\sum_{m+2}^{k} n_\nu / \sum_{m+1}^{k} n_\nu\right)(\bar{y}^{(m+1)} - \bar{y}_0)^2}{\sum_{\nu=1}^{k}\sum_{\lambda=1}^{n_\nu}(y_\lambda^{(\nu)} - \bar{y}^{(\nu)})^2 + \sum_{\nu=m+2}^{k} n_\nu(\bar{y}^{(\nu)} - \bar{y}_0)^2},$$

where $\bar{y}_0 = \sum_{m+2}^{k} n_\nu \bar{y}^{(\nu)} / \sum_{m+2}^{k} n_\nu$.

PROOF. From $1 - r^2_{0(m+1)\cdot} = (1 - r^2_{0(0,1,\cdots,m+1)})/[1 - r^2_{0(0,1,\cdots,m)}]$, and the sample estimator of (3.15):

$$r^2_{0(0,1,\cdots,m)} = \left[\sum_0^m \hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y})^2 + \left(1 - \sum_0^m \hat{p}_\nu\right)^{-1}\sum_0^m \hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y})^2\right] \Big/ s^2$$

we obtain

$$\frac{r^2_{0(m+1)\cdot}}{1-r^2_{0(m+1)\cdot}} = \frac{1}{D}\left\{\hat{p}_{m+1}(\bar{y}^{(m+1)} - \bar{y})^2 + \frac{\left[\sum_0^{m+1}\hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y})\right]^2}{1 - \sum_0^{m+1}\hat{p}_\nu} - \frac{\left[\sum_0^{m}\hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y})\right]^2}{1 - \sum_0^{m}\hat{p}_\nu}\right\}$$

$$= \frac{1}{D}\frac{\hat{p}_{m+1}\left(1 - \sum_0^m \hat{p}_\nu\right)}{\left(1 - \sum_0^{m+1}\hat{p}_\nu\right)}\left\{\frac{(\bar{y}^{(m+1)} - \bar{y})\sum_{m+2}^{k}\hat{p}_\nu + \sum_0^{m+1}\hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y})}{\left(1 - \sum_0^m \hat{p}_\nu\right)}\right\}^2;$$

and finally

$$\frac{r^2_{0(m+1)\cdot}}{1-r^2_{0(m+1)\cdot}} = \frac{1}{D}\frac{\hat{p}_{m+1}\left[\sum_{m+2}^{k}\hat{p}_\nu(\bar{y}^{(m+1)} - \bar{y}^{(\nu)})\right]^2}{\sum_{m+2}^{k}\hat{p}_\nu \sum_{m+1}^{k}\hat{p}_\nu}, \tag{5.5}$$

using the relation $\sum_0^{m+1}\hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y}) = -\sum_{m+2}^{k}\hat{p}_\nu(\bar{y}^{(\nu)} - \bar{y})$. In the above

$$nD = \sum_\alpha (y_\alpha - \bar{y})^2 - \sum_0^{m+1} n_\nu(\bar{y}^{(\nu)} - \bar{y})^2 - \left[\sum_0^{m+1} n_\nu(\bar{y}^{(\nu)} - \bar{y})\right]^2 \Big/ \sum_{m+2}^{k} n_\nu$$

$$= \sum_{\nu=0}^{k}\sum_{\lambda=1}^{n_\nu}(y_\lambda^{(\nu)} - \bar{y}^{(\nu)})^2 + \sum_{\nu=0}^{k} n_\nu(\bar{y}^{(\nu)} - \bar{y})^2 - \sum_0^{m+1} n_\nu(\bar{y}^{(\nu)} - \bar{y})^2$$

$$- \left[\sum_0^{m+1} n_\nu(\bar{y}^{(\nu)} - \bar{y})\right]^2 \Big/ \sum_{m+2}^{k} n_\nu$$

Simplifying,

$$nD = \sum_{\nu=0}^{k}\sum_{\lambda=1}^{n_\nu}(y_\lambda^{(\nu)} - \bar{y}^{(\nu)})^2 + \sum_{m+2}^{k} n_\nu(\bar{y}^{(\nu)} - \bar{y}^{(0)})^2. \tag{5.6}$$

Substitution of $\hat{p}_\nu = n_\nu/n$ and (5.6) into (5.5) leads to the result. ||

In particular, if $m + 1 = k - 1$, *then,*

$$(5.7)\qquad \frac{(n-k-1)^{\frac{1}{2}} r_{0(k-1)\cdot}}{(1-r^2_{0(k-1)\cdot})^{\frac{1}{2}}} = \frac{\left(\frac{n_{k-1}\, n_k}{n_{k-1}+n_k}\right)^{\frac{1}{2}} (\bar{y}^{(k-1)} - \bar{y}^{(k)})}{\left(\sum_{\nu=0}^{k}\sum_{\lambda=1}^{n_\nu} (y_\lambda^{(\nu)} - \bar{y}^{(\nu)})^2/(n-k-1)\right)^{\frac{1}{2}}}$$

has a Student's t-distribution with $n - k - 1$ *degrees of freedom, when* $\rho \equiv \rho_{0(k-1)\cdot} = 0$, *and a mixture of non-central* t_{n-k-1}*-distributions with parameter*

$$\tau = \left[\left(\frac{\rho^2}{1-\rho^2}\right)\left(\frac{p_{k-1}+p_k}{p_{k-1}\, p_k}\right)\left(\frac{n_{k-1}\, n_k}{n_{k-1}+n_k}\right)\right]^{\frac{1}{2}}.$$

and mixing coefficients

$$\binom{n}{n_{k-1}, n_k} p_{k-1}^{n_{k-1}} p_k^{n_k} (1 - p_{k-1} - p_k)^{n-n_{k-1}-n_k}$$

Note that $\rho = 0$ if and only if $\mu_{k-1} = \mu_k$.

5.2.1. *Asymptotic distribution.* The asymptotic distribution of $r_{0(k-1)\cdot} \equiv r$ will be obtained by an argument paralleling that of Theorem 4.1. *Let*

$$g(r) = \frac{r^2}{1-r^2} \sim \frac{1}{n-k-1} F_{1,n-k-1},$$

$$h(r) = \{g(r)/[1+g(r)]\}^{\frac{1}{2}} = r,$$

then the limiting distribution of $h(r)$ *is given by* (4.2). As before we find $\lim_{n\to\infty} nV(g(r)) = \lim_{n\to\infty} nd^{-2}V(F_{1,d}(\tau^2))$, using (4.3) and (4.4) with $c = 1$, $d = n - k - 1$. However, now the definition of the non-centrality parameter

$$(5.8)\qquad \tau^2 = g(\rho)\beta n_{k-1} n_k/\alpha(n_{k-1}+n_k), \qquad \alpha = p_{k-1}p_k, \qquad \beta = p_{k-1} + p_k,$$

is different.

LEMMA 5.4.

$$E\tau^2 = g(\rho)[n - 1/\beta].$$

PROOF.

$$E[n_{k-1}n_k/(n_{k-1}+n_k)] = E[n_{k-1}n_k|(n_{k-1}+n_k) = m]$$
$$= Em\beta^{-1}p_{k-1}(1 - \beta^{-1}p_{k-1}) = \alpha\beta^{-2}(Em - 1).$$

Since $m \sim b(n, \beta)$, $Em = n\beta$. ||

LEMMA 5.5.

$$E\tau^4 = \frac{ng^2(\rho)}{\alpha\beta}\left[(n-1)\alpha\beta + \beta^2 - 5\alpha + O\left(\frac{1}{n}\right)\right].$$

PROOF. Let $x = n_1/n$, $y = n_2/n$, $b(x, y) = x^2y^2/(x+y)^2$, then $E\tau^4 = n^2\beta^2\alpha^{-2}g^2(\rho)Eb(x, y)$. Now expand $b(x, y)$ in a Taylor's series about

(p_{k-1}, p_k) to second degree terms. We have $b_x = 2xy^3/(x+y)^3$, $b_{xx} = 2y^3(y-2x)/(x+y)^4$, $b_{xy} = 6x^2y^2/(x+y)^4$. After simplification

$$b(x,y) = \beta^{-4}[x^2p_k^3(p_k - 2p_{k-1}) + y^2p_{k-1}^3(p_{k-1} - 2p_k) + 6\alpha^2x^2y^2] + R.$$

Since $Ex^2 = [(n-1)p_{k-1}^2 + p_{k-1}]/n$, $Exy = (n-1)\alpha/n$, $ER = O(1/n^2)$, we obtain

$$Eb(x,y) = \alpha[\alpha\beta(n-1) + \beta^2 - 5\alpha + O(1/n)]/n\beta^3. \quad \|$$

THEOREM 5.6.

$$r_{0(k-1)} \to \mathfrak{N}\left(\rho, \frac{(1-\rho^2)^2}{n}\left[1 + \rho^2\left(\frac{\beta^2 - 3\alpha\beta - 3\alpha}{4\alpha\beta}\right)\right]\right).$$

PROOF. Using (4.3), (4.4), Lemmas 1, 2,

$$\begin{aligned} V(g(r)) = d^{-2}V(F_{1,d}) = {} & [3(d-2) + E\tau^4 + 6E\tau^2 - (d-4)(1+E\tau^2)^2]/ \\ & (d-2)^2(d-4) = [2n^2g^2(\rho) + n(d-2)g^2(\rho) \\ & (-\alpha\beta + \beta^2 - 5\alpha)/\alpha\beta + 4n(d-1)g(\rho) + 2n(d-4) \\ & g^2(\rho)/\beta + O(n)]/(d-2)^2(d-4) \end{aligned}$$

Recall that $d = n - k - 1$, so that

$$\lim nV(g(r)) = g^2(\rho)(\alpha\beta + \beta^2 - 3\alpha)/\alpha\beta + 4g(\rho).$$

The remaining computations follow directly from (4.2). $\|$

REMARKS. *The parameter ρ may be removed from the variance by the variance stabilizing transformation $\phi(x)$ which satisfies the equation*

$$\phi'(x) = [(1-x^2)(1+cx^2)^{\frac{1}{2}}]^{-1}, \qquad c = (\beta^2 - 3\alpha\beta - 3\alpha)/(4\alpha\beta).$$

The desired solution is

$$\phi(x) = \frac{1}{(c+1)^{\frac{1}{2}}} \tanh^{-1} \frac{x(c+1)^{\frac{1}{2}}}{(1+cx^2)^{\frac{1}{2}}}, \tag{5.9}$$

or, equivalently,

$$\phi(x) = \frac{1}{2(c+1)^{\frac{1}{2}}} \tanh^{-1} \frac{2x[(c+1)(1+cx^2)]^{\frac{1}{2}}}{1 + (2c+1)x^2}. \tag{5.10}$$

If $p_{k-1} = p_k = \frac{1}{2}$, then $c = -\frac{1}{2}$, and $\phi(x) = 2^{\frac{1}{2}} \tanh^{-1} x(2-x^2)^{\frac{1}{2}} = 2^{\frac{1}{2}} \tanh^{-1}[1 - (1-x^2)^2]$, which coincides with the result of [13]. In general, then, we obtain

THEOREM 5.7.

$$\tanh^{-1} \frac{r_{0(k-1)}\,(1+c)^{\frac{1}{2}}}{(1+cr_{0(k-1)}^2)^{\frac{1}{2}}} \to \mathfrak{N}\left(\tanh^{-1} \frac{\rho(1+c)^{\frac{1}{2}}}{(1+c\rho^2)^{\frac{1}{2}}}, \frac{1+c}{n}\right),$$

with $c = (\beta^2 - 3\alpha\beta - 3\alpha)/4\alpha\beta$, $\alpha = p_{k-1}p_k$, $\beta = p_{k-1} + p_k$.

6. Case $k > 1$, $p > 1$.

6.1. *Remarks on sample canonical correlations.* Define $b_{im} = (\bar{y}_i^{(m)} - \bar{y}_i)$, $B = (b_{im}) : p \times k + 1$, $D = \text{diag}\,(n_0/n, \cdots, n_k/n)$,

$$s_{ij} = \sum_{m=0}^{k} \sum_{\lambda=1}^{n_m} (y_{i\lambda}^{(m)} - \bar{y}_i^{(m)})(y_{j\lambda}^{(m)} - \bar{y}_j^{(m)}), \quad S = (s_{ij}) : p \times p,$$

$h_{ij} = \sum_{m=0}^{k} n_m b_{im} b_{jm}/n$, $H = (h_{ij}) : p \times p$. Then $H = BDB'$ is an estimate of UD_pU', and S/n is an estimate of Σ, so that we are interested in the distribution of the roots of $|H - \theta S/n| = 0$ or equivalently the roots of $|S^{-\frac{1}{2}}HS^{-\frac{1}{2}} - \theta/n| = 0$. Even for the simplest model with $\mu^{(0)} = \cdots = \mu^{(k)}$, $\Sigma = I$, and $n_0, \cdots, n_k$ fixed, this problem remains unsolved. The reduction of the following paragraph will serve to focus attention on the difficulties.

Let $B \sim \mathfrak{N}(0, I)$, $S \sim$ Wishart (I, p, n), that is

$$p(B, S) = \text{const}\, |S|^{\frac{1}{2}(n-p-1)} \exp\left[-\tfrac{1}{2} \text{tr}\,(BB' + S)\right]. \tag{6.1}$$

Let $L = S^{-\frac{1}{2}}BD^{\frac{1}{2}}$. The Jacobian is $|S|^{\frac{1}{2}(k+1)}|D|^{-p/2}$, and

$$p(L, S) = \text{const.}\, |S|^{(n-p+k)/2} \exp\left[-\tfrac{1}{2} \text{tr}\, S(LD^{-1}L' + I)\right].$$

Integration over the domain $S > 0$ yields

$$p(L) = \text{const.}\, |D|^{-p/2}|I + LD^{-1}L'|^{-(n+k+1)/2} \tag{6.2}$$
$$= \text{const.}\, |D|^{(n+k-p+1)/2}|D + L'L^{(n+k+1)/2}.$$

Our concern is then to obtain the distribution of the characteristic roots of LL'. Except for the cases $k = 1$ or $p = 1$, this problem is untractable, for it involves the evaluation of integrals of the type $\int g(\Gamma)|\Gamma D_1 \Gamma' - D_2|^{-c} d\Gamma$ over the domain $\Gamma\Gamma' = I$, where D_1, D_2 are diagonal matrices, and $g(\Gamma)$ is a function of Γ alone; it is the determinant in the integral which causes the problem.

6.2. *The sample vector correlation.* The sample counterpart of (3.18), namely

$$r_v^2 = 1 - \frac{|S/n|}{|H + S/n|}, \tag{6.3}$$

using the notation of Section 6.1, is called the sample vector correlation coefficient. For the case $\rho_v^2 > 0$ nothing is known of the distribution. It would of course be some kind of mixture, and would in general contain nuisance parameters.

Under the null hypothesis it is known that $1 - r_v^2 \sim U_{p,k,n-k-1}$, where $U_{p,k,n-k-1}$ is a U-statistic of Wilks (see Anderson [1], ch. 9.7). The exact distribution is available for all p, k values such that at least one of the two quantities is 1, 2, or 3. For the case $p = k = 3$ a table is available for small values of n (Anderson [1], ch. 8.5). For larger sample values the asymptotic distribution of Rao [10] can be adapted to our situation. Denote by $P_x(z)$ the distribution function of the random variable x evaluated at z. Then

$$P_{-\gamma_0 \log(1-r_v^2)}(z) = P_{\chi^2_{pk}}(z) + \frac{\gamma_2}{\gamma_0^2}\left\{P_{\chi^2_{pk+4}}(z) - P_{\chi^2_{pk}}(z)\right\} + O(n^{\frac{1}{2}}), \tag{6.4}$$

where

$$\gamma_0 = n - \tfrac{1}{2}(p + k + 3), \qquad \gamma_2 = (pk/48)(p^2 + k^2 - 5).$$

7. Estimation of parameters. It has been tacitly assumed throughout the paper that the parameters μ_{im} and σ_{ij} may be estimated by their corresponding sample means, and the parameters p_m by their corresponding relative frequencies, regardless of any dependence which might exist between the random vectors y_α and x_α. The purpose of this section is to show that the assumption is valid if the method of maximum likelihood is used.

Let $h(y_\alpha, x_\alpha)$ be the density of the random vector $(y_{1\alpha}, y_{2\alpha}, \cdots, y_{p\alpha}; x_{0\alpha}, x_{1\alpha}, \cdots, x_{k\alpha})$, and $f(x_\alpha)$ the density of $(x_{0\alpha}, x_{1\alpha}, \cdots, x_{k\alpha})$. Also, denote by $\phi_m(y_\alpha)$ the density of a $\mathfrak{N}(\mu^{(m)}, \Sigma)$ random vector; recall that $\mu^{(m)}$ is a column vector of the matrix (μ_{im}). According to our model

$$f(x_\alpha) = p_0^{x_{0\alpha}} \cdots p_k^{x_{k\alpha}}, \qquad h(y_\alpha|x_\alpha) = \sum_0^k x_{m\alpha}\phi_m(y_\alpha). \tag{7.1}$$

Therefore, the density of the whole sample is

$$\prod_\alpha h(y_\alpha, x_\alpha) = \prod_\alpha \Big(\sum_m x_{m\alpha}\phi_m(y_\alpha)\Big) p_0^{x_{0\alpha}} \cdots p_k^{x_{k\alpha}} \tag{7.2}$$

Now change the α-labels in such a way that they start with the α-values for which $x_{0\alpha} = 1$, then those for which $x_{1\alpha} = 1$; more precisely, assume

$$x_{m\alpha} = 1, \qquad \alpha = \lambda + \sum_0^{m-1} n_\nu, \qquad \lambda = 1, 2, \cdots, n_m, \qquad m = 0, 1, \cdots, k.$$

Then, let $y_\lambda^{(m)} = (y_{1\lambda}^{(m)}, \cdots, y_{p\lambda}^{(m)})$ be independent $\mathfrak{N}(\mu^{(m)}, \Sigma)$ random vectors. It is now easy to see that (7.2) becomes

$$\prod_\alpha h(y_\alpha, x_\alpha) = p_0^{n_0} \cdots p_k^{n_k} \prod_{m=0}^{k} \prod_{\lambda=1}^{n_m} \phi_m(y_\lambda^{(m)}). \tag{7.3}$$

Thus, the joint density is factorable; one factor contains the parameters $p_0, p_1, \cdots, p_k$; the other factors contain the parameters μ_{im} and σ_{ij}.

8. Summary of procedures, with examples.

8.1 *Case* $p > 1$, $k = 1$.

(i) $r^2_{x_1(y_1,\cdots,y_p)}$ may be computed from (4.1).

(ii) The hypothesis $H: \rho^2_{x_1(y_1\cdots,y_p)} = 0$ with other parameters arbitrary may be tested by the Q-statistic of Theorem 4.1: Reject if $Q \geqq c$, where c is obtained from a table of the F-distribution.

(iii) Examples of all cases are provided by some of the studies described by Rokeach [11]. In experiments concerning attitudes individuals were classified as Northerners or Southerners; according to membership in the British political parties: Liberals, Conservatives, Laborites, or Communists; according to religious affiliation. The individuals also received scores on Dogmatism and Opinionation scales. Thus, he considers situations in which $p = 2$. For illustrative purposes, we have chosen a restricted set of data from one of his studies.

y_1	y_2	x_0	x_1	y_1	y_2	x_0	x_1
175	148	1	0	133	143	0	1
168	156	1	0	159	135	0	1
158	151	1	0	168	198	0	1
183	136	1	0	192	169	0	1
195	164	1	0	168	117	0	1
106	138	1	0	142	138	0	1

$x_o = 1$ for a Northerner; $x_1 = 1$ for a Southerner.
y_1 = dogmatism score; y_2 = opinionation score.

For this example $n = 12$, $n_0 = 5$, $n_1 = 7$. The formula $\bar{y}_i^{(m)} = \sum_\alpha y_{i\alpha} x_{m\alpha}/n_m$, and the expression for s_{ij}, given at the beginning of Section 4, gives

$$\bar{y}_1^{(0)} = 175.80, \quad \bar{y}_2^{(0)} = 151.00, \quad \bar{y}_1^{(1)} = 152.57, \quad \bar{y}_2^{(1)} = 148.29$$

$$s_{11} = 553.76, \quad s_{12} = 180.87, \quad s_{22} = 471.46.$$

Whence, $s^{11} = .0020645$, $s^{12} = -.0007920$, $s^{22} = .0024249$, and $\bar{y}^{(1)} - \bar{y}^{(0)} = (-23.23, -2.71)$. Then,

$$T^2 = (n_0 n_1/n)(\bar{y}^{(1)} - \bar{y}^{(0)}) S^{-1} (\bar{y}^{(1)} - \bar{y}^{(0)})' = 3.010,$$

and

$$r^2 = T^2/(10 + T^2) = .2314, \quad r = .481.$$

Finally, $Q = 1.355$. Referring to the F-table for 2 and 9 degrees of freedom, we see that a Q-value of 9.38, and hence an r^2 value of .6757, is required for significance. Thus, we accept the hypothesis that there is no particular association between Dogmatism and Opinionation (as defined by Rokeach) on the one hand, and the section of the country in which the individual lives on the other hand.

(iv) In order to obtain the power of the test in part (ii) for a specified alternative (p_0, p_1, ρ) we may consult the tables of Tang [12] or the charts of Pearson and Hartley [9]. The quantity τ^2 of Theorem 4.1 must be computed for p_0, p_1, ρ, and each possible partition $n = n_0 + n_1$. The probability of a Type II error is then obtained as a mixture of various values of P_{II} from the tables or charts. A calculation of this kind is given by Tate ([14], p. 1083).

(v) When p_0, p_1 are known, we can test the hypothesis $H: \rho^2 = \text{constant}$ (not necessarily zero) by using the distribution (4.5) of Theorem 4.2. An alternative to this is to make a variance stabilizing transformation ϕ (this is discussed in Section 5.2.1 in connection with partial correlation), and then test $H: \phi(\rho^2) = \phi(\text{constant})$. For the case at hand we have

$$\phi(r) = (4p_0p_1)^{\frac{1}{2}}(1 - 2p_0p_1)^{-\frac{1}{2}} \tanh^{-1} r(1 - 2p_0p_1)^{\frac{1}{2}}[4p_0p_1 + (1 - 2p_0p_1)r^2]^{-\frac{1}{2}}$$

and, of course, $\phi(r) \to \mathfrak{N}(\phi(\rho), 1/n)$.

(vi) Confidence limits for ρ can be obtained when p_0 and p_1 are known by first finding confidence limits for $\phi(\rho)$ and then obtaining from them the limits for ρ. See Tate [13] for an example with $p = 1$, $k = 1$, $p_0 = p_1 = \frac{1}{2}$.

8.2. *Case* $p = 1$, $k > 1$.

(i) $r_0^2 = r^2_{y(x_1,\cdots x_k)}$ may be computed from (5.1), (5.2), or (5.3).

(ii) The hypothesis $H:\rho_0^2 = 0$ with other parameters arbitrary may be tested by the Z-statistic of Theorem 5.1.

(iii) The asymptotic distribution of r_0 is given by Theorem 5.2. Note that nuisance parameters are present whatever the values of $p_0, \cdots, p_k$ (recall that they must all be positive). To test $H:\rho_0^2 =$ constant (not necessarily zero) we can use the asymptotic distribution, with estimates for the nuisance parameters in any form of the asymptotic variance which is convenient to use. The alternative to this is to use the lower bound $(1 - \rho_0^2)^2(2 - \rho_0^2)/2n$ and reject too often when H is true, or use the upper bound $(1 - \rho_0^2)^2(2 - \rho_0^2)[1 + \frac{1}{2}\rho_0^2(2 - \rho_0^2)^{-1} \sum_0^k (p_m^{-1} - 2)]/2n$ and reject too rarely when H is true.

(iv) The remarks of (iii) apply also to the determination of a confidence interval for ρ_0. The upper bound for the asymptotic variance would lead to a conservative confidence interval worthy of more confidence than we place on it.

8.3 *Case* $p > 1, k > 1$.

(i) r_v^2 is computed from (6.3), using the notation of Section 6.1.

(ii) The hypothesis $H:\rho_v^2 = 0$ with other parameters arbitrary is tested by the statistic $1 - r_v^2$ which has the $U_{p,k,n-k-1}$ distribution under H. See the remarks in Section 6.2 concerning the availability and scope of tables, and Rao's form for the distribution function of the transformed variate $-(n - \frac{1}{2}(p + k + 3)) \log (1 - r_v^2)$.

9. Acknowledgment. The authors are indebted to Professor Milton Rokeach, Michigan State University, for making his data available, and to the referee for the present proofs of Lemma 3.1 and Theorem 3.2.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[2] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey, 1946.

[3] R. A. Fisher, "On a distribution yielding the error functions of several well-known statistics," *Proc. Int. Math. Congress*, Toronto (1924), pp. 805–813.

[4] John W. Hooper, "The sampling variance of correlation coefficients under assumptions of fixed and mixed variates," *Biometrika*, Vol. 45 (1958), pp. 471–477.

[5] Harold Hotelling, "The generalization of Student's ratio," *Ann. Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[6] Harold Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[7] P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci. India*, Vol. 12 (1936), pp. 49–55.

[8] M. D. Moustafa, "Tests of hypotheses on a multivariate population, some of the variables being continuous and the rest catagorical," Institute of Statistics Mimeograph Series No. 179, Chapel Hill, North Carolina, 1957.

[9] E. S. Pearson and H. O. Hartley, "Charts of the power function for analysis of variance tests, derived from the noncentral F-distribution," *Biometrika*, Vol. 38 (1951), pp. 112–130.

[10] C. RADHAKRISHNA RAO, "Tests of significance in multivariate analysis," *Biometrika*, Vol. 35 (1948), pp. 58–79.
[11] MILTON ROKEACH, *The Open and Closed Mind*, Basic Books, New York, 1960.
[12] P. C. TANG, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 126–149 and tables.
[13] R. F. TATE, "Correlation between a discrete and a continuous variable," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 603–607.
[14] R. F. TATE, "Applications of correlation models for biserial data," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1078–1095.
[15] S. S. WILKS, "On the independence of k sets of normally distributed statistical variables," *Econometrica*, Vol. 3 (1935), pp. 309–326.