The Annals of Mathematical Statistics
1971, Vol. 42, No. 5, 1569–1578

MULTIVARIATE PROCEDURES INVARIANT UNDER LINEAR TRANSFORMATIONS[1]

BY R. L. OBENCHAIN

Bell Telephone Laboratories

Many well-known procedures in multivariate data analysis are invariant under the group, $L(p)$, of translations and nonsingular linear transformations. New maximal $L(p)$ invariant statistics are derived and are shown to have the geometrical interpretation of a scatter of points in Euclidean space. The distribution of maximal $L(p)$ invariants for the case of a single multivariate normal population is shown to follow from a result of James (1954). Finally we consider tests of the null hypothesis that $k > 1$ populations are identical and show that optimal $L(p)$ invariant tests are similar tests of randomness.

1. Introduction and summary. Suppose $p \geqq 1$ variables are measured on individuals in random samples from $k \geqq 1$ populations. We focus interest on procedures invariant under the group, $L(p)$, of translations and nonsingular linear transformations of the variables being observed. Sections 2 through 5 are essentially independent of probabilistic considerations and deal with data reduction to a maximal invariant function, a uniquely determined matrix of squared interpoint distances achieved by a scatter of points in Euclidean space of p or fewer dimensions.

The distribution of maximal $L(p)$ invariants is considered in Section 6. It is shown that $L(p)$ preserves the interchangeability (but not the independence) of the observations in a random sample when the transformation depends (symmetrically) on the data. The distribution of maximal $L(p)$ invariants is displayed for the case of a single multivariate normal population by reinterpreting a result of James (1954).

In Section 7, $L(p)$ invariant tests of randomness are considered for the multivariate, several-sample problem. Optimal $L(p)$ invariant tests are seen to be permutation tests of interchangeability. A test based upon the Pillai (1955) trace criterion is given a geometrical interpretation and is recommended when no specific alternative hypothesis is of interest.

2. The data and the model for invariance. Suppose that a vector, $x = (x_1 \cdots x_p)'$ of $p \geqq 1$ real valued variables is to be observed on individuals from $k \geqq 1$ populations, a random sample of size n_j being taken from the jth population. The

Received August 17, 1970.

[1] Research supported in part by the National Institutes of Health, Institute of General Medical Sciences, Grant No. GM12868–05 at the University of North Carolina, Chapel Hill.

AMS subject classifications. Primary 6205, 6240, 6270; Secondary 1505, 1590.

Key words and phrases. Multivariate analysis, maximal invariant, compatible central measure, projection matrix, squared Euclidean distance, interchangeable random vectors, permutation test.

data to be collected is formed into a $p \times N$ matrix of random variables, X, where the $N = \Sigma n_j$ columns are all stochastically independent, but only certain subsets of the columns are necessarily identically distributed.

A statistic, t, computed from a data matrix, X, will be said to be invariant under the group, $L(p)$, of translations and nonsingular linear transformations if and only if

$$t(X) = t(ae' + BX) \tag{2.1}$$

for every $p \times 1$ vector a and for every $p \times p$ *nonsingular* matrix B, where e' is the $1 \times N$ vector of all unities. The statistic, t, may be real valued or consist of a vector or matrix of real values. Many well-known test statistics in multivariate analysis satisfy (2.1); these include Hotelling's T^2 for detecting difference in location of two populations, the likelihood ratio criterion of Wilks and the trace criteria of Lawley-Hotelling and Pillai for comparing $k \geqq 2$ (normal) populations, and Fisher's sample linear discriminant function.

Let us write $\mathscr{X}(p, N)$ for the set of all $p \times N$ matrices with real elements. $L(p)$ partitions $\mathscr{X}(p, N)$ into equivalence classes or orbits; two data matrices, X_1 and X_2 belonging to $\mathscr{X}(p, N)$, are on the same orbit if and only if there exists $[a, B] \in L(p)$ such that $X_2 = ae' + BX_1$. A statistic satisfying (2.1) is called a *maximal* $L(p)$ *invariant* if it takes on a different value on each orbit of $\mathscr{X}(p, N)$. The maximal $L(p)$ invariants which will be derived here are matrix valued statistics. The value computed at the point X can be interpreted as a reference point on the orbit through X or as a canonical form for the data, X, when considering invariance under $L(p)$.

3. Central measures compatible with $L(p)$. In the next section, a maximal invariant is derived under the subgroup of transformations $[0, B] \in L^\circ(p) \subset L(p)$ in which the translation vector $a = 0$. We now show that it is possible to restrict attention from $L(p)$ to $L^\circ(p)$ by subtracting from each column of the data matrix, X, an appropriately chosen measure of "location," $m = m(X)$.

Let X_1 and $X_2 \in \mathscr{X}(p, N)$ be on the same orbit induced by $L(p)$. Let $m_i = m(X_i)$ for $i = 1, 2$ be the two $p \times 1$ vectors of real valued statistics computed from X_1 and X_2 respectively, and write $Y_i = X_i - m_i e'$. Thus $Y_2 = (a + Bm_1 - m_2)e' + BY_1$, and it follows that Y_1 and Y_2 are on the same orbit induced by $L^\circ(p)$ if and only if m has the following property:

$$m(ae' + BX) = a + Bm(X) \tag{3.1}$$

for every $[a, B] \in L(p)$ and $X \in \mathscr{X}(p, N)$. A statistic, m, which satisfies (3.1) will be called a *central measure compatible with* $L(p)$ because its value on transformed data is equal to its transformed value. Note that the occurrence of the vector a on the right-hand side of (3.1) shows that the statistic m contains no information invariant under $L(p)$.

In the univariate case, $p = 1$, most measures of location (e.g., medians) satisfy (3.1). The multivariate central measures compatible with $L(p)$ are characterized in the following lemma. A statistic, $m = m(X)$, is said to be *continuous* on $\mathscr{X}(p, N)$ if, for every sequence $X_1, X_2, \cdots$ of matrices from $\mathscr{X}(p, N)$ that approaches a limit matrix elementwise, $\lim m(X_i) = m(\lim X_i)$ elementwise.

LEMMA 1. *The only continuous central measures compatible with $L(p)$ when $p > 1$ are weighted means of the form*

$$m(X) = X\gamma \tag{3.2}$$

where γ is a fixed $N\times 1$ vector such that $\gamma' e = 1$. Also, $r = \text{rank } Y \leqq \min(p, N-1)$, *where $Y = X - me'$.*

PROOF. If m is continuous, it follows from (3.1) that $m(BX) = Bm(X)$ even when B is singular. Taking B to be a null matrix except for a unity in the jth position of its first row, it follows that $m_j(X)$, the jth element of $m(X)$, equals $m_1(X_j)$, where X_j is the matrix whose first row is the jth row of X and whose remaining rows are null. But, with B null except for b_{11} and b_{12} arbitrary, $m_1(b_{11}X_1 + b_{12}X_2) = b_{11}m_1(X) + b_{12}m_2(X) = b_{11}m_1(X_1) + b_{12}m_1(X_2)$. Thus m_1 is a linear functional on N-dimensional real space, and $m(X) = X\gamma$ for some fixed $N\times 1$ vector, γ. Taking B null in (3.1) then shows $e'\gamma = 1$. Finally, note that $Y = X(I - \gamma e')$ has rank less than or equal to the minimum of p and $N-1$ because $I - \gamma e'$ is idempotent with trace $N-1$.

4. Maximal $L(p)$ invariants.

Write $Y = X - me' = XV_\gamma$, where $V_\gamma = I - \gamma e'$ as before, and define an $N\times N$ matrix of statistics, $P = P(X)$, by the formula

$$P = Y'(YY')^- Y \tag{4.1}$$

where $(YY')^-$ denotes any generalized inverse for the $p\times p$ matrix YY'. Now P is the *projection matrix* associated with the vector space generated by the rows of Y; P is symmetric, idempotent, and uniquely determined regardless of the choice of the generalized inverse involved in its definition, Rao (1966), Lemma 5. Then we have the following theorem.

THEOREM 1. *If $m(X)$ is compatible with $L(p)$, Equation* (3.1), *then the statistic P defined by* (4.1) *is a maximal $L(p)$ invariant.*

PROOF. The result is now obvious because $[a, B] \in L(p)$ acting upon X transforms Y into BY; the transformations are forming nonsingular linear combinations of the rows of Y. The vector space generated by the rows of Y is thus invariant and maximal. A direct proof of the theorem proceeds as follows: (i) P is invariant because $X_2 = ae' + BX_1$ implies that a possible choice for $[Y_2 Y_2']^-$ is $B'^{-1} \times [Y_1 Y_1']^- B^{-1}$. (ii) P is maximal because $P_1 = P_2$ implies that the rows of Y_1 and Y_2 both form a (possibly singular) basis for the same vector space. The usual argument for change of basis then shows that there exists $[a, B] \in L(p)$ such that $X_2 = ae' + BX_1$; in fact, the transformation which does this is unique only if rank $Y_1 = \text{rank } Y_2 = p$. This concludes the proof.

The statistic P is not the maximal invariant which is the easiest to understand and work with. A simple argument can be used to extend the statement of James (1954), page 45, and to show that the orbits of $\mathscr{X}(p, N)$ induced by $L(p)$ are in one to one correspondence with the elements of the Grassmann manifold, $G_{p,N-p-1}$. See also Fraser (1968), page 228. However, we wish to show that a maximal invariant has the geometrical representation of a scatter of N points in Euclidean space. We note that P is an $N \times N$ *association matrix* in the sense of Gower (1966); p_{ij}, the element in the ith row and jth column of P, compares the ith and jth individuals in the sample taking all p variables measured into consideration. An analysis of the characteristic roots (= c-roots) and vectors of P leads to the representation we desire. If we write r = trace P = rank Y, then the analysis of P produces a $r \times N$ matrix, W, whose jth column, w_j, gives the coordinates of a point, to be plotted using r orthogonal axes, which represents the jth individual.

The calculation of a coordinate representation, W, is complicated by the fact that P is idempotent. W is to be chosen such that

$$W'W = P \tag{4.2}$$

and

$$WW' = I. \tag{4.3}$$

It is clear from (4.1) that γ is a c-vector of P with c-root zero. Since the rows of the W are c-vectors of P with c-root one, it follows that

$$m(W) = W\gamma = 0, \tag{4.4}$$

the weighted mean of the scatter is the origin. If W_1 satisfies (4.2) and (4.3), then so does

$$W_2 = CW_1, \tag{4.5}$$

where C is any $r \times r$ orthogonal matrix which rotates the reference axes. It follows from (4.3) that the sample sum of squares of projections of the scatter onto any direction in r-space is one, and the sample correlation between the projections on any two directions is zero if and only if the directions are orthogonal.

Equation (4.5) suggests that the geometrical interpretation of P is best approached by an analysis of the unique distances among points rather than by references to a nonunique coordinate system. Write $D^{(2)} = ((d_{ij}^2))$ to denote the $N \times N$ matrix where d_{ij}^2 is the squared distance between the representations of the ith and jth individuals in the combined sample. $D^{(2)}$ is symmetric with null diagonal; $D^{(2)}$ cannot be written as a matrix product DD' or $DD = D^2$. Then we have the following corollary to Theorem 1.

COROLLARY 1A. *If $m(X)$ is a central measure compatible with $L(p)$, then the statistic $D^{(2)} = ((d_{ij}^2))$ is a maximal $L(p)$ invariant equivalent to $P = ((p_{ij}))$, where*

$$d_{ij}^2 = (w_i - w_j)'(w_i - w_j) \tag{4.6}$$

$$= p_{ii} + p_{jj} - 2p_{ij} \tag{4.7}$$

$$= (x_i - x_j)'S^-(x_i - x_j), \tag{4.8}$$

and

$$S = YY' = XX' - N(m\bar{x}' + \bar{x}m' - mm') \tag{4.9}$$

is the $p \times p$ pooled within and among samples adjusted sums of squares and products matrix. Also

$$0 \leqq d_{ij}^2 \leqq 2 - (\gamma_i - \gamma_j)^2/\gamma'\gamma. \tag{4.10}$$

PROOF. The expression (4.7) follows immediately from (4.2) and the definition (4.6) as pointed out by Gower (1966). If we rewrite (4.7) in matrix form it becomes

$$D^{(2)} = p_0 e' + e p_0' - 2P, \tag{4.7$'$}$$

where $p_0 = (p_{11}, p_{22}, \cdots, p_{NN})'$. To determine P from $D^{(2)}$, multiply (4.7′) on the right (left) by $V_\gamma(V_\gamma')$ and get

$$P = (I - e\gamma')(-D^{(2)}/2)(I - \gamma e'). \tag{4.11}$$

Now note that $I - P - \gamma\gamma'/\gamma'\gamma$ is symmetric, idempotent, orthogonal to P and γ, and of rank $N - r - 1$. Expression (4.10) follows by noting that the squared distance in this orthogonal space is computed as in (4.7′). This concludes the proof.

The following point is now stressed: when $k > 1$ populations are being sampled, maximal $L(p)$ invariant statistics discard all information about the *average* location and dispersion of the populations. However, information on the *relative* location and dispersion of the populations is preserved. Some further geometrical properties of maximal $L(p)$ invariants are considered in Section 5, the one sample situation ($k = 1$) is examined in Section 6, and we return to the several sample problem in Section 7.

5. Geometrical properties of the representation. One notes immediately that the maximal $L(p)$ invariant, $D^{(2)}$, measures distance in a familiar way. Except for the fact that S contains among sample effects when $k > 1$, formula (4.8) is a Mahalanobis (1936) distance, equivalent to Hotelling's T^2 statistic. In fact, Wald and Wolfowitz (1944), Section 7, show that such statistics are monotonically related to (4.8) when $m = \bar{x}$ ($\gamma = e/N$) and $k = 2$.

If we write

$$S = GL^2G' \tag{5.1}$$

where L^2 is the $r \times r$ diagonal matrix of nonzero c-roots of S, (4.9), and G is the $p \times r$ semi-orthogonal matrix of corresponding c-vectors, then we see that

$$W = L^{-1}G'(X - me') \tag{5.2}$$

satisfies (4.2) and (4.3). That is, the sample principal components, normed to weighted mean zero and sum of squares one, form a possible coordinate representation of P.

Before proceeding further, it should be noted that the specific measure of distance (4.8) was considered by Gower (1966), Section 4.2, for the case $m = \bar{x}$. Gower states that use of this measure of distance "reflects the attitude that the group of individuals is homogeneous." There is, however, reason to doubt this assertion. Suppose, for example, that $p = 2$ characteristics are measured on a population whose bivariate density has unusual equal probability contours. These contours would be ellipses if the distribution were bivariate normal, and representations of such data would tend to convert these contours into circles. Therefore, suppose the contours are not even closed or connected. Then data from such a population and its $L(p)$ invariant representation would both display the same behavior in probability. Empirical evidence could thus be gathered to support the hypothesis that the population consists of a mixture of two or more simple distributions.

The above argument does break down, however, when many variables are measured on only a few subjects. If $r = N-1 \leqq p$, then

$$P = I - \gamma\gamma'/\gamma'\gamma, \tag{5.3}$$

a constant which does not depend upon the data. When $m = \bar{x}$, the representation of (5.3) is that of N equally spaced points in $N-1$ dimensions.

The proof of the following lemma parallels that of Corollary 1A, so it is not given here.

LEMMA 2. *Let d_{0j}^2 denote the squared distance between the representation of the* jth *individual and the origin, the representation of $m(X)$. Then*

$$d_{0j}^2 = p_{jj} = (x_j - m)'S^-(x_j - m), \tag{5.4}$$

where $0 \leqq d_{0j}^2 \leqq 1 - \gamma_j{}^2/\gamma'\gamma$ *and* $\sum_{j=1}^{N} d_{0j}^2 = r \leqq \min(p, N-1)$.

6. The null distribution of maximal $L(p)$ invariants. Suppose that the columns of X are interchangeable random p-vectors. An important special case occurs when, perhaps by hypothesis, the columns of X are independent and identically distributed (i.i.d.). Because of the importance of the following theorem, *in the remainder of this paper* we restrict attention to maximal $L(p)$ invariants constructed using $\gamma = e/N$ in the compatible central measure, m.

THEOREM 2. *If the columns of X are interchangeable random vectors and a maximal $L(p)$ invariant is constructed using $m = \bar{x}$, then the representation of the N individuals is that of interchangeable random quantities.*

PROOF. The columns of $X-\bar{x}e'$ are interchangeable random vectors with a nonrandom sum, Chernoff and Teicher (1958), case (a). Now S defined by (4.9) when $m = \bar{x}$ is a symmetric function of the columns of $X-\bar{x}e'$, so this is true of L and G in (5.1). Thus the coordinator representation, W, defined by (5.2) has interchangeable columns. P and $D^{(2)}$ are *invariant in distribution* under simultaneous permutations of their rows and columns when the columns of X are interchangeable. This concludes the proof.

The distribution of a maximal $L(p)$ invariant is singular because it contains only $r(N-r-1)$ functionally independent entries. This is most easily seen by examining the coordinate representation, W, a matrix with rN elements. The assumption of translation invariance (a in (2.1)) creates r constraints on W summarized by (4.4). The assumption of nonsingular linear transformation invariance (B in (2.1)) adds r^2 constraints. Of these constraints, $r(r+1)/2$ imply (4.3) and $r(r-1)/2$ degrees of freedom are lost because coordinate representations are not unique, Equation (4.5).

The exact distribution of maximal $L(p)$ invariants with $m = \bar{x}$ can be displayed for the case of a multivariate normal population. James (1954) shows how locally defined exterior differential forms can be used to derive sampling distributions in this case; his result on the decomposition of a random sample will be reinterpreted here. Suppose that

$$X =_d \mathcal{N}_{pN}(\mu e'; \Sigma \otimes R) \tag{6.1}$$

where $N > p$, Σ is a $p\times p$ positive definite variance–covariance matrix, and $R = (1-\rho)I+\rho ee'$ is a $N\times N$ correlation matrix with $-1/(N-1) < \rho < 1$. In the following we denote the ith row of W by ω_i' and choose any $(N-p-1)\times N$ matrix U, whose jth row is u_j', such that $[N^{-\frac{1}{2}}e \,\vdots\, W' \,\vdots\, U']$ is an $N\times N$ orthogonal matrix. We can now restate Theorem 8.1 of James (1954), (8.22).

THEOREM 3. *The distribution* (6.1) *of X can be decomposed into four independent distributions*:

$$dF(X) = dF(\bar{x})\cdot dF(S)\cdot dF(D^{(2)})\cdot dF(C), \tag{6.2}$$

where

(i) $\bar{x} =_d \mathcal{N}_p(\mu; (1+(N-1)\rho)\Sigma/N)$,

(ii) $S =_d W_p(N-1, (1-\rho)\Sigma, 0)$,

(iii) $dF(D^{(2)}) = K(p,N)\prod_{j=1}^{N-p-1}\prod_{i=1}^{p} u_j'(-dD^{(2)}/2)\omega_i$,

where the constant is

$$K(p,N) = \pi^{-p(N-p-1)/2}\prod_{i=1}^{p}\Gamma((N-i)/2)/\Gamma(i/2),$$

and

(iv) C has the invariant distribution on the orthogonal group.

$$dF(C) = 2^{-p}\pi^{-p(p+1)/4}\prod_{i=1}^{p}\Gamma(i/2)\prod_{i<j} c_j'\,dc_i.$$

The possible realizations of $D^{(2)}$ consist of the space of $N \times N$ symmetric matrices with nonnegative entries and null diagonal such that the matrix

$$P = (I - ee'/N)(-D^{(2)}/2)(I - ee'/N) \tag{6.3}$$

is idempotent of rank p.

PROOF. We note that $dP = (dW')W + W'(dW)$ is such that

$$U(dP)W' = U(-dD^{(2)}/2)W' = U\,dW' = B'\,dH \tag{6.4}$$

where B' is $(N-p-1) \times (N-1)$, semi-orthogonal, and orthogonal to H of James (1954), (5.17). We arrive at the probability element (6.2 (iii)) by noting that the indicated double product of linear differential forms is the exterior product of the elements of any of the matrices shown in (6.4). It is stressed that the validity of (6.2 (iii)) does not depend upon any uniqueness property of W; the ω_i's and u_j's can be chosen as any orthonormal set of c-vectors of P and $I - ee'/N - P$, respectively, corresponding to c-roots of $+1$.

7. Tests of structure $S(\pi, \alpha)$ invariant under $L(p)$. Suppose we wish to test the null hypothesis, H, that $k > 1$ populations are identical against the alternative hypothesis, K, that at least two of the populations are different. The null hypothesis can be written as

(7.1) H: the columns of X are independent and identically distributed.

In order for a statistic, $t = t(X)$, to be of any use in testing the null hypothesis against the stated alternative, t must *not* be invariant under the group, π, of permutations of the N columns of X. However, t is *invariant in distribution* under this group when H holds.

We now restrict attention to tests invariant under $L(p)$. As shown in Theorem 2, $L(p)$ preserves the interchangeability, but not the independence, of the observations in a random sample. Thus the null hypothesis may as well be weakened to the statement

(7.2) H': the distribution of X is invariant under π.

We are now in a position to apply the results of Lehmann and Stein (1949) who showed that the theory of optimum tests of H' is the same as the theory of optimum *similar* tests of H.

We will represent the transformations in π by $N \times N$ permutation matrices, C, which are null except for one unity in each row and column. We define $t^{(j)}(X)$ as being the jth largest value of $t(XC)$ for $C \in \pi$; actually, we need not consider within sample permutations when $k > 1$, but this complicates the notation. A randomized test function is then defined by

$$\begin{aligned} \phi_\alpha(X) &= 1 && \text{if} \quad t(X) > t^{(j)}(X) \\ &= a(X) && \text{if} \quad t(X) = t^{(j)}(X) \\ &= 0 && \text{if} \quad t(X) < t^{(j)}(X) \end{aligned} \tag{7.3}$$

where, given $0 < \alpha < 1$, the integer j and the function $a(X) \in [0, 1]$ can be so chosen that the test is of size α under H'. The statistic t must be invariant under $L(p)$, (2.1), but the choice of t depends upon the specific alternative of interest.

Most powerful $L(p)$ invariant tests of H against a simple alternative can be constructed in theory. For example, let $g(D^{(2)})$ denote the generalized density of a normal distribution εK with respect to the measure (6.2 (iii)) εH, then one should take $t(X) = g(D^{(2)})$.

When no specific alternative εK is of interest, it is heuristically satisfactory to base the test upon a $L(p)$ invariant statistic which measures the disparity between within-sample and among-sample distances. Thus

(7.4) $$t(X) = \sum{}^* d_{ij}^2 = d_{(a)}^2$$

is suggested, where the symbol Σ^* denotes summation over all subscripts i and j which belongs to different samples. We introduce the following notation:

(7.5) $$\bar{x}^{(m)} = \sum{}^{(m)} x_i / n_m$$

is the sample mean vector for the observations from the mth population, and

(7.6) $$S_w^{(m)} = \sum{}^{(m)} (x_i - \bar{x}^{(m)})(x_i - \bar{x}^{(m)})'$$

is the matrix of adjusted sums of squares and products within the mth sample. We write $\overline{d^2}(m, q)$, to denote the average value of d_{ij}^2 for all $i \neq j$, i from sample m, and j from sample q. It follows that

(7.7) $$\overline{d^2}(m, m) = \frac{2}{n_m - 1} [\sum{}^{(m)} p_{ii} - n_m (\bar{x}^{(m)} - \bar{x})' S^- (\bar{x}^{(m)} - \bar{x})]$$
$$= \frac{2}{n_m - 1} \operatorname{tr} (S^- S_w^{(m)}),$$
and

(7.8) $$\overline{d^2}(m, q) = \operatorname{tr} \left\{ S^- \left[\frac{1}{n_m} S_w^{(m)} + \frac{1}{n_q} S_w^{(q)} + (\bar{x}^{(m)} - \bar{x}^{(q)})(\bar{x}^{(m)} - \bar{x}^{(q)})' \right] \right\}.$$

It is easily shown that the overall average squared distance in the representation is $2r/(N-1)$. We write $S = S_a + S_w$ to divide S into its among-sample and within-sample parts, and we note that $d_{(a)}^2$ of (7.4) takes on the following particularly simple form when S is nonsingular and all samples are of the same size ($n_m = n$ for $1 \leqq m \leqq k$):

(7.9) $$d_{(a)}^2 = n[kp - \operatorname{tr} S^{-1} S_w]$$
$$= n[\operatorname{tr} S^{-1} S_a + p(k-1)].$$

Thus $d_{(a)}^2$ is a monotonically increasing function of the Pillai (1955) trace criterion when all the k samples are of the same size; otherwise $d_{(a)}^2$ is weighted by sample sizes.

8. Conclusions. Geometrical insight may be brought to bear upon multivariate procedures invariant under $L(p)$ by considering the data as a scatter of N points in Euclidean space of p or fewer dimensions. The metric $(x_i - x_j)'S^-(x_i - x_j)$ is used for this purpose, and these squared distances between individuals (observations) can be ordered quite differently from those computed from the scatter of the raw data using the formula $(x_i - x_j)'(x_i - x_j)$.

As an example of the usefulness of geometrical representation, consider the situation described in Mardia (1967). Mardia shows that a concept of "rank on angle" is invariant under $L(2)$ and produces a statistic which is distribution free under the hypothesis (7.1). The results of this paper make it clear that a concept of distance from the origin is also preserved by $L(p)$. Procedures which combine information about radius and angle are not unconditionally distribution free, but clearly are of great interest. The associated probability of an observed test statistic can be computed exactly when N is small or estimated using random permutations when N is large.

The assumption of invariance in multivariate analysis is most commonly made with respect to the group $L(p)$; but other "linear" choices are possible. Maximal invariants are displayed in Obenchain (1969) for the cases where B of (2.1) is restricted to be block diagonal and/or where any information that $k > 1$ populations may differ in location is to be ignored.

Acknowledgment. I am grateful to Professors P. K. Sen, N. L. Johnson, and N. Suguira of the University of North Carolina for their comments on earlier versions of this work. Valuable suggestions were made by the referees.

REFERENCES

Chernoff, H. and Teicher, H. (1958). A central limit theorem for sums of interchangeable random variables. *Ann. Math. Statist.* **29** 118–130.

Fraser, D. A. S. (1968). *The Structure of Inference.* Wiley, New York.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53** 325–338.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

Lehmann, E. L. and Stein, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Statist.* **20** 28–45.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India, Part A.* **12** 49–55.

Mardia, K. V. (1967). A nonparametric test for the bivariate two-sample location problem. *J. Roy. Statist. Soc. Ser. B.* **29** 320–342.

Obenchain, R. L. (1969). Rank tests invariant only under linear transformations. Unpublished dissertation, Univ. of North Carolina.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* **26** 117–121.

Rao, C. R. (1966). Generalized inverse for matrices and its applications in mathematical statistics. *Research Papers in Statistics*, ed. F. N. David. Wiley, New York.

Wald, A. and Wolfowitz, J. (1944). Statistical tests based upon permutations of the observations. *Ann. Math. Statist.* **15** 358–372.