A LINEARLY INVARIANT CANONICAL FORM FOR MULTIVARIATE DATA

by

ROBERT L. OBENCHAIN
Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 606

JANUARY 1969

This research was supported in part by the National Institutes of Health of General Medical Sciences, Grant No. GM 12868-05.

TABLE OF CONTENTS

ABSTRACT

Suppose that data on $p \geq 1$ variables in the form of random samples from $c \geq 1$ populations is to be analyzed using statistical procedures invariant with respect to translation and nonsingular linear transformation. Then it is shown that the only information of use from the data is that contained in its canonical form. This representation of the data is constructed using the work of Gower [1] and has many interesting properties, implied by assumptions not usually realized, when viewed as a scatter of points in Enclidean space of p or fewer dimensions. The exact distribution of the representation for a single Multivariate Normal population follows from the work of James [2]. Possible uses of the canonical form are considered.

NOTATION

$\underset{p\,1}{\underline{x}}$ is a $p \times 1$ column vector, and $\underset{1\,p}{\underline{x}'} = (x_1 \ldots x_p)$ is the corresponding row vector.

$\underset{p\,1}{\underline{1}}$ is the $p \times 1$ column vector of all unities, and $\underset{p\,1}{\underline{1}}\,\underset{q}{\underline{1}'}$ is a $p \times q$ matrix of all unities.

$\underset{p\,q}{\underset{\sim}{X}} = ((x_{ij}))$ is a $p \times q$ matrix with x_{ij} as the element in the i-th row and j-th column.

$\underset{\sim}{I}_p$ is the identity matrix of order p.

$\underset{\sim}{A} \otimes \underset{\sim}{B}$ denotes the Kronecker or Right Direct Product of the matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$.

$$\underset{n\;n-1}{\underset{\sim}{Q}^{(n)}} = \begin{bmatrix} \frac{-1}{\sqrt{1.2}} & \frac{-1}{\sqrt{2.3}} & & \frac{-1}{\sqrt{(n-1)n}} \\ \frac{1}{\sqrt{1.2}} & \frac{-1}{\sqrt{2.3}} & & \frac{-1}{\sqrt{(n-1)n}} \\ 0 & \frac{2}{\sqrt{2.3}} & & \frac{-1}{\sqrt{(n-1)n}} \\ 0 & 0 & & \frac{-1}{\sqrt{(n-1)n}} \\ 0 & 0 & 0 & \frac{n-1}{\sqrt{(n-1)n}} \end{bmatrix}$$

is the negative of the semi-orthogonal matrix associated with the Helmert Transformation.

← The columns should be in the reverse order. Eqn. (3.11), page 18, is correct.

Then

$$\underset{1\;n\;n-1}{\underline{1}'\underset{\sim}{Q}^{(n)}} = \underset{1\;n-1}{\underline{0}'}, \quad \underset{\sim}{Q}^{(n)'}\underset{\sim}{Q}^{(n)} = \underset{\sim}{I}_{n-1}, \quad \text{and} \quad \underset{\sim}{Q}^{(n)}\underset{\sim}{Q}^{(n)'} = \underset{\sim}{I}_n - \frac{1}{n}\,\underline{1}\,\underline{1}' .$$

1. INTRODUCTION

Suppose $p \geq 1$ characteristics $(X_1, \ldots, X_p)$ are to be measured on individuals from $c \geq 1$ populations. Suppose $n_i \geq 1$ individuals are selected at random from the i-th population, and let $N = \sum_{i=1}^{c} n_i$. For simplicity of notation, let the individuals be numbered $1, 2, \ldots, N$, let $\underline{X}_j' = (X_{1j}, X_{2j}, \ldots, X_{pj})$ denote the random vector to be observed on the j-th individual, and write

$$\underset{p\sim N}{X} = (\underline{X}_1 \mid \cdots \mid \underline{X}_N) \tag{1.1}$$

so that the first n_1 columns of $\underset{\sim}{X}$ are independent and identically distributed (= *i.i.d.*) random vectors from the first population, columns n_1+1 through n_1+n_2 are *i.i.d.* from the second population, etc. Thus, if $c > 1$, only certain of the columns of $\underset{\sim}{X}$ are interchangeable.

A row of $\underset{\sim}{X}$ contains a particular variable to be measured on all N individuals. We have already imposed a structure on the columns of $\underset{\sim}{X}$, let us now assume a model for the rows of $\underset{\sim}{X}$. In this paper, we will be concerned with situations in which it makes sense to the experimenter to consider linear combinations of the variables. Although this assumption is not always reasonable, this mechanism for combining variables is often assumed in standard techniques of multivariate analysis. In particular, let us suppose that the information we intend to collect by observing $\underset{\sim}{X}$ is only that information which is <u>invariant</u> with respect to <u>all</u> transformations of the form

$$\underset{p\,N}{\underset{\sim}{X}} \rightarrow \underset{p\,N}{\underset{\sim}{X}} + \underset{p\,1}{\underline{a}}\;\underset{1\,N}{\underline{1}'} \tag{1.2}$$

This research was supported in part by the National Institutes of Health of General Medical Sciences, Grant No. GM 12868-05.

and

$$\underset{p\ N}{\underset{\sim}{X}} \rightarrow \underset{p\ p}{\underset{\sim}{B}}\ \underset{p\ N}{\underset{\sim}{X}} \tag{1.3}$$

where $\underline{a}$ is an arbitrary $p\times 1$ vector and $\underset{\sim}{B}$ is an arbitrary, nonsingular $p\times p$ matrix. Note that transformations of this type do not combine data for different individuals (columns) or populations; this is obvious when $\underline{a}$ and $\underset{\sim}{B}$ are not functions of $\underset{\sim}{X}$.

It will be shown that, under the above assumptions, the data $\underset{p\ N}{\underset{\sim}{X}}$ can be replaced by its canonical form

$$\underset{r\ N}{\underset{\sim}{W}} = (\underline{W}_1 \mid \cdots \mid \underline{W}_N) \ . \tag{1.4}$$

$\underset{\sim}{W}$ is a $r\times N$ random matrix with the following properties:

(i) $\underset{\sim}{W}$ is a function of $\underset{\sim}{X}$ only and can be uniquely determined.

(ii) r is an integer valued random variable taking values between zero and $\min(p, N-1)$ inclusive.

(iii) $\underset{r\ N}{\underset{\sim}{W}}$ is <u>maximal invariant</u> under the group of transformations (1.2) and (1.3) on $\underset{p\ N}{\underset{\sim}{X}}$.

(iv) $\underset{\sim}{W}$ is distributed on a subset of the Grassmann manifold $G_{r,N-r}$ homeomorphic to $G_{r,N-r-1}$.

(v) $\underline{W}_j$, the j-th column of $\underset{\sim}{W}$, gives the coordinates of a point, to be plotted using r orthogonal axes in Euclidean space, which represents the j-th individual.

The canonical form will be derived from a somewhat pedagogical point of view so that the properties of $\underset{\sim}{W}$ will be systematically revealed. The experienced user of multivariate techniques may find that the considerations of Lemma 3.3 are a more obvious starting point.

Hotelling's T^2 for detecting difference in location of two populations is one example of a statistical procedure applicable to data which conforms to our model. Indeed, the further assumption that $\underset{\sim}{X}$ has a Multivariate Normal distribution with dispersion $\underset{\sim}{I}_N \otimes \underset{\sim}{\Sigma}_{p\,p}$ could be made so that the significance of T^2 could be judged with respect to the central F distribution. On the other hand, the permutation distribution of T^2 can be constructed and a conditional test can be performed without making further assumptions. In this paper, we wish to study only the implications of our basic assumptions about translation invariance (1.2) and linear transformation invariance (1.3).

2. CONSTRUCTION OF THE REPRESENTATION

2.1. Invariance with respect to Translation.

Write

$$\underset{p\,N}{\underset{\sim}{Y}} = \underset{p\,N}{\underset{\sim}{X}}(\underset{\sim}{I}_N - \frac{1}{N}\underline{1}\,\underline{1}') = \underset{p\,N}{\underset{\sim}{X}} - \underset{p\,1}{\overline{\underline{X}}}\,\underset{1\,N}{\underline{1}'} \tag{2.1}$$

where $\overline{\underline{X}} = \frac{1}{N}\sum_{j=1}^{N}\underline{X}_j$, and note that $\underset{\sim}{Y}$ is invariant under transformations of the form (1.2) on $\underset{\sim}{X}$. Note further that this invariance could be achieved by writing $\underset{p\,N}{\underset{\sim}{Y}} = \underset{p\,N}{\underset{\sim}{X}}\,\underset{N\,N}{\underset{\sim}{T}}$ for any matrix $\underset{\sim}{T}$ such that $\underset{1\,N}{\underline{1}'}\underset{N\,N}{\underset{\sim}{T}} = \underset{1\,N}{\underset{\sim}{0}'}$. The choice for $\underset{\sim}{T}$ specified in (2.1) recognizes the fact that $\underset{\sim}{T}$ combines the columns (individuals) of $\underset{\sim}{X}$ and yet must retain the interpretation that the columns of $\underset{\sim}{Y}$ represent the corresponding columns of $\underset{\sim}{X}$. This choice for $\underset{\sim}{T}$ is symmetric, idempotent, and of rank N-1; it is the projection matrix for the space of dimension N-1 orthogonal the N×1 constant vector, $\underline{1}$.

If $c > 1$, it might be of interest to consider

$$\underset{N\,N}{\underset{\sim}{T}} = \underset{\sim}{I}_N - \begin{bmatrix} \frac{1}{n_1}\underline{1}\,\underline{1}' & \underset{\sim}{0} & & \underset{\sim}{0} \\ \underset{\sim}{0} & \frac{1}{n_2}\underline{1}\,\underline{1}' & & \\ & & & \underset{\sim}{0} \\ \underset{\sim}{0} & & \underset{\sim}{0} & \frac{1}{n_c}\underline{1}\,\underline{1}' \end{bmatrix}, \tag{2.2}$$

a projection matrix of rank N-c. In this way, any possible differences in location among the c populations can be ignored. The choice for $\underset{\sim}{T}$ specified in (2.2) will not, however, be considered in the sequel because it does not lead to a maximal invariant statistic under the group of transformations (1.2).

2.2. Invariance with respect to Nonsingular Linear Transformation.

Write

$$\underset{N\,N}{\underset{\sim}{P}} = \underset{N\,p}{\underset{\sim}{Y}'}(\underset{p\,N}{\underset{\sim}{Y}}\,\underset{N\,p}{\underset{\sim}{Y}'})^{-}\underset{p\,N}{\underset{\sim}{Y}} \tag{2.3}$$

where $(\underset{\sim}{Y}\,\underset{\sim}{Y}')^{-}$ denotes any generalized inverse (= g-inverse) for the $p\times p$ matrix $\underset{\sim}{Y}\,\underset{\sim}{Y}'$. It is well known that $\underset{\sim}{P}$ is symmetric, indempotent, and uniquely determined regardless of the choice of the g-inverse involved in its definition; see, for example, Lemma 5 of Rao [4]. $\underset{\sim}{P}$ is the projection matrix associated with the vector space generated by the rows of $\underset{\sim}{Y}$. It follows that $\underset{\sim}{P}$ is invariant with respect to transformations (1.3) on $\underset{\sim}{Y}$, and, therefore, $\underset{\sim}{P}$ is invariant with respect to transformations (1.2) and (1.3) on $\underset{\sim}{X}$. The proof that $\underset{\sim}{P}$ is maximal invariant must be delayed until the end of the next section.

Note that

$$\underset{p\,N}{\underset{\sim}{Y}}\,\underset{N\,p}{\underset{\sim}{Y}'} = \sum_{j=1}^{N}(\underline{X}_j - \overline{\underline{X}})(\underline{X}_j - \overline{\underline{X}})' = \underset{p\,p}{\underset{\sim}{S}}, \quad \text{say.} \tag{2.4}$$

If we are considering $c > 1$ populations, we have

$$\underset{p\,p}{\underset{\sim}{S}} = \underset{p\,p}{\underset{\sim}{S}^{(w)}} + \underset{p\,p}{\underset{\sim}{S}^{(a)}}$$

where

$$\underset{p\,p}{\underset{\sim}{S}^{(w)}} = \sum_{i=1}^{c}\underset{\sim}{S}_{(i)} = \sum_{i=1}^{c}\sum_{j\in\Pi_i}(\underline{X}_j - \overline{\underline{X}}^{(i)})(\underline{X}_j - \overline{\underline{X}}^{(i)})' \tag{2.5}$$

is the within populations matrix of sample sums of squares and products, Π_i denotes the individuals in the sample from the i-th population,

$$\underline{\bar{X}}^{(i)} = \frac{1}{n_i} \sum_{j \in \pi_i} \underline{X}_j \qquad (2.6)$$

is the sample mean vector from the i-th population, and

$$\underset{p\,p}{\underset{\sim}{S}}^{(a)} = \sum_{i=1}^{c} n_i (\underline{\bar{X}}^{(i)} - \underline{\bar{X}})(\underline{\bar{X}}^{(i)} - \underline{\bar{X}})' \qquad (2.7)$$

is the among populations matrix of sample sums of squares and products.

2.3. Individual Analysis of $\underset{\sim}{P}$

Gower [1] has discussed characteristic root (= c-root) and vector techniques in which a geometrical interpretation is given to the information contained in N×N association matrices. These are techniques in "Individual Analysis" because the element in the ℓ-th row and m-th column of an association matrix compares the ℓ-th and m-th individuals in the sample taking all p variables measured into consideration. Techniques which analyze p×p matrices such as $\underset{\sim}{S}$ are more common in multivariate analysis and proceed from the point of view of "Collective Analysis".

We will apply Gower's technique to the matrix $\underset{\sim}{P}$ given by (2.3). Let $r = \text{rank } \underset{\sim}{P} = \text{rank } \underset{\sim}{Y} \leq \min(p,N-1)$. If $r = 0$, then $\underset{p\,N}{\underset{\sim}{X}} = \underset{p\,1}{\underline{\bar{X}}}\,\underset{1\,N}{\underline{1}'}$ so that there is some indication that the data is non-stochastic. Since $\underset{\sim}{P}$ is idempotent, it has r c-roots of +1 and N-r c-roots of 0. Thus the work of Gower implies that the data should be represented using exactly r orthonormal c-vectors of $\underset{\sim}{P}$ associated with c-roots of +1. If r=1, the c-vector associated with the single c-root of +1 is determined uniquely up to a constant multiple which can change its length and the signs of all its elements. Since the non-zero c-roots of $\underset{\sim}{P}$ are all equal, there is no automatically unique way to pick the $r > 1$ orthogonal c-vectors of interest.

To overcome difficulties in definition, we take $\underset{r\,N}{\underset{\sim}{W}}$ to be the almost everywhere unique matrix with the following four properties:

$$\text{(i)} \quad \underset{N\,r}{\underset{\sim}{W}'}\underset{r\,N}{\underset{\sim}{W}} = \underset{N\,N}{\underset{\sim}{P}} \tag{2.8}$$

$$\text{(ii)} \quad \underset{r\,N}{\underset{\sim}{W}}\underset{N\,r}{\underset{\sim}{W}'} = \underset{\sim}{I}_r \tag{2.9}$$

$$\text{(iii)} \quad \underset{r\,N}{\underset{\sim}{W}}\underset{N\,1}{\underline{1}} = \underset{r\,1}{\underline{0}} \tag{2.10}$$

$$\text{(iv)} \quad \underset{r\,N}{\underset{\sim}{W}} = \left(\underset{r\,N-r}{\underset{\sim}{W}^{(1)}} \,\middle|\, \underset{r\,r}{\underset{\sim}{W}^{(2)}} \right) \tag{2.11}$$

where $\underset{\sim}{W}^{(2)}$ is a $r\times r$ matrix with all elements on the secondary diagonal (lower left to upper right) greater than zero and all elements below the secondary diagonal equal to zero.

Equation (2.8) implies that the rows of $\underset{\sim}{W}$ are c-vectors of $\underset{\sim}{P}$ associated with c-roots of +1. Equation (2.9) implies that the rows of $\underset{r\,N}{\underset{\sim}{W}}$ are orthonormal. Equation (2.10) states that the elements in each row of $\underset{\sim}{W}$ sum to zero. This property of $\underset{\sim}{W}$ follows from the fact that $\underset{p\,N}{\underset{\sim}{Y}}\underset{N\,1}{\underline{1}} = \underset{p\,1}{\underline{0}}$ as is obvious from equation (2.1). Then equation (2.3) implies that $\underset{N\,N}{\underset{\sim}{P}}\underset{N\,1}{\underline{1}} = \underset{N\,1}{\underline{0}}$ so that $\underline{1}$ is a c-vector of $\underset{\sim}{P}$ associated with a c-root of zero. Thus (2.10) follows because c-vectors corresponding to unequal c-roots are orthogonal.

Equation (2.11) orients the representation in a unique way. In the terminology of James [2], (2.9) implies that $\underset{r\,N}{\underset{\sim}{W}}$ is an "r-frame" in the Stiefel manifold, $V_{r,N}$, with the special property (2.10). However, (2.11) makes it possible to interpret $\underset{\sim}{W}$ as an element of the Grassmann manifold,

$G_{r,N-r}$, with the special property (2.10). Actually, James [2] interprets the elements of $G_{r,N-r}$ as the r dimensional planes (passing through the origin) in Euclidean N-space; we are considering the dual interpretation that certain elements of $G_{r,N-r}$ are very special scatters of N points in Euclidean r-space. In geometrical terms, equation (2.11) states that the right-handed system of orthogonal axes should be "rotated" (an orthogonal transformation with determinant +1 or -1) in r-space until

(1) the individual numbered N (the last column of $\underset{\sim}{W}$) lies on the positive half of the axis numbered one,

(2) the individual numbered $N-1$ lies in the plane spanned by the axes numbered one and two and has its coordinate with respect to the axis numbered two greater than zero,

⋮

(k) the individual numbered $(N - k + 1)$ lies in the k-plane spanned by the axes one through k and has its coordinate with respect to the k-th axis greater than zero, and so forth while $k \leq r$.

The set of all possible transformations (1.2) and (1.3) can be termed the general nonsingular linear transformation group, $L(p)$. An element of $L(p)$ will be denoted by $(\underline{a}, \underset{\sim}{B})$ and transforms the matrix $\underset{p\ N}{\underset{\sim}{X}}$ into the matrix $\underset{p}{\underline{a}}\ \underset{1\ N}{\underline{1}'} + \underset{p\ p}{\underset{\sim}{B}}\ \underset{p}{\underset{\sim}{X}}$. The "product" of $(\underline{a}_{(1)}, \underset{\sim}{B}_{(1)})$ and $(\underline{a}_{(2)}, \underset{\sim}{B}_{(2)})$ is defined to be $(\underline{a}_{(1)} + \underline{a}_{(2)}, \underset{\sim}{B}_{(1)}\underset{\sim}{B}_{(2)})$, $(\underline{0}, \underset{\sim}{I}_p)$ is the identity transformation, and $(- \underline{a}, \underset{\sim}{B}^{-1})$ is the transformation inverse to $(\underline{a}, \underset{\sim}{B})$. We are now in a position to prove the following:

THEOREM 2.1. The statistic $\underset{N\ N}{\underset{\sim}{P}}$, defined as a function of $\underset{p\ N}{\underset{\sim}{X}}$ by (2.1) and (2.3), is maximal invariant under $L(p)$. $\underset{r\ N}{\underset{\sim}{W}}$, or any other equivalent of $\underset{N\ N}{\underset{\sim}{P}}$, also has this property.

Proof. Since we have already shown that $\underset{\sim}{P}$ is invariant under $L(p)$, we can state that $\underset{\sim}{P}$ is maximal invariant if $\underset{N\ N}{\underset{\sim}{P}}(\underset{p}{\underset{\sim}{X}}{}_{(1)}{}_N) = \underset{N\ N}{\underset{\sim}{P}}(\underset{p}{\underset{\sim}{X}}{}_{(2)}{}_N)$ implies that there exists $(\underline{a}, \underset{\sim}{B}) \in L(p)$ which transforms $\underset{\sim}{X}_{(1)}$ into $\underset{\sim}{X}_{(2)}$.

If $r = \text{rank}\ \underset{\sim}{P} = p$, $\underset{p}{\underset{\sim}{S}}{}^{-}_{(i)}{}_{p} = \underset{p}{\underset{\sim}{S}}{}^{-1}_{(i)}{}_{p}$ and we have

$$\underset{p}{\underset{\sim}{W}}{}_{(i)}{}_N = \underset{p}{\underset{\sim}{C}}{}_{(i)}{}_p \underset{p}{\underset{\sim}{S}}{}^{-1/2}_{(i)}{}_p (\underset{p}{\underset{\sim}{X}}{}_{(i)}{}_N - \underset{p}{\underline{\overline{X}}}{}_{(i)}{}_1 \underline{1}'_N) \quad \text{for } i = 1, 2 \tag{2.12}$$

where $\underset{\sim}{C}_{(i)}$ is some $p\times p$ orthogonal matrix and $\underset{\sim}{S}^{-1/2}_{(i)}$ is any (necessarily nonsingular) matrix such that $\underset{\sim}{S}^{-1/2'}_{(i)} \underset{\sim}{S}^{-1/2}_{(i)} = \underset{\sim}{S}^{-1}_{(i)}$. Since $\underset{\sim}{P}(\underset{\sim}{X}_{(1)}) = \underset{\sim}{P}(\underset{\sim}{X}_{(2)})$ implies that $\underset{\sim}{W}_{(1)} = \underset{\sim}{W}_{(2)}$, the linear transformation that transforms $\underset{\sim}{X}_{(1)}$ into $\underset{\sim}{X}_{(2)}$ is $(\underline{a}^*, \underset{\sim}{B}^*)$ where

$$\underset{p\ p}{\underset{\sim}{B}^*} = \underset{\sim}{S}^{1/2}_{(2)} \underset{\sim}{C}'_{(2)} \underset{\sim}{C}_{(1)} \underset{\sim}{S}^{-1/2}_{(1)} \quad \text{and} \tag{2.13}$$

$$\underset{p\ 1}{\underline{a}^*} = \underline{\overline{X}}_{(2)} - \underset{\sim}{B}^* \underline{\overline{X}}_{(1)} \tag{2.14}$$

where $\underset{\sim}{S}^{1/2}_{(2)}$ is the inverse of $\underset{\sim}{S}^{-1/2}_{(2)}$.

If $r = \text{rank}\ \underset{\sim}{P} < p$, however, the proof is somewhat complicated by the fact that $(\underline{a}^*, \underset{\sim}{B}^*) \in L(p)$ is not uniquely determined. Since rank $(\underset{\sim}{X}_{(i)} - \underline{\overline{X}}_{(i)}\underline{1}') = r$, there exist nonsingular $p\times p$ matrices $\underset{\sim}{E}_{(i)}$ such that

$$\underset{p}{\underset{\sim}{E}}{}_{(i)}{}_p (\underset{p}{\underset{\sim}{X}}{}_{(i)}{}_N - \underset{p}{\underline{\overline{X}}}{}_{(i)}{}_1 \underline{1}'_N) = \begin{bmatrix} \underset{r}{\underset{\sim}{X}^*}{}_{(i)}{}_N \\ \underset{p-r}{\underset{\sim}{0}}{}_N \end{bmatrix} \tag{2.15}$$

for $i = 1, 2$ where rank $\underset{\sim}{X}^*_{(i)} = r$. Now there exists $(\underline{0}, \underset{\sim}{B}^*) \in L(r)$ such that $\underset{\sim}{X}^*_{(2)} = \underset{\sim}{B}^* \underset{\sim}{X}^*_{(1)}$ as was shown above when $r=p$. It follows that

TABLE 2.1:

Numerical Example for p=r=2 and N=5

$$\underset{2\,5}{\underset{\sim}{X}} = \begin{pmatrix} -2 & -3 & 1 & 2 & 3 \\ 1 & -1 & 1 & -1 & 2 \end{pmatrix}$$

$$\underset{2\,1}{\overline{\underset{\sim}{X}}} = \begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix} \qquad \underset{2\,2}{\underset{\sim}{S}} = \begin{pmatrix} 26.8 & 5.6 \\ 5.6 & 7.2 \end{pmatrix}$$

$$\underset{5\,5}{\underset{\sim}{P}} = \begin{pmatrix} 0.367 & & & & \\ 0.134 & 0.471 & & & \\ 0.010 & -0.148 & 0.055 & & \\ -0.460 & 0.000+ & -0.074 & 0.644 & \\ -0.051 & -0.457 & 0.156 & -0.111 & 0.463 \end{pmatrix}$$

$$\underset{2\,5}{\underset{\sim}{W}^*} = \begin{pmatrix} 0.127 & 0.681 & -0.225 & 0.094 & -0.678 \\ -0.592 & -0.080 & -0.066 & 0.797 & -0.059 \end{pmatrix}$$

$$\underset{2\,5}{\underset{\sim}{W}} = \begin{pmatrix} -0.076 & -0.672 & 0.230 & -0.163 & 0.681 \\ -0.601 & -0.139 & -0.046 & 0.786 & 0 \end{pmatrix}$$

$$\underset{5\,5}{\underset{\sim}{D}^2} = \begin{pmatrix} 0 & & & & \\ 0.569 & 0 & & & \\ 0.401 & 0.822 & 0 & & \\ 1.931 & 1.114 & 0.846 & 0 & \\ 0.933 & 1.849 & 0.205 & 1.329 & 0 \end{pmatrix}$$

FIGURE 2.1. The Raw Data, $\underset{2\sim 5}{X}$

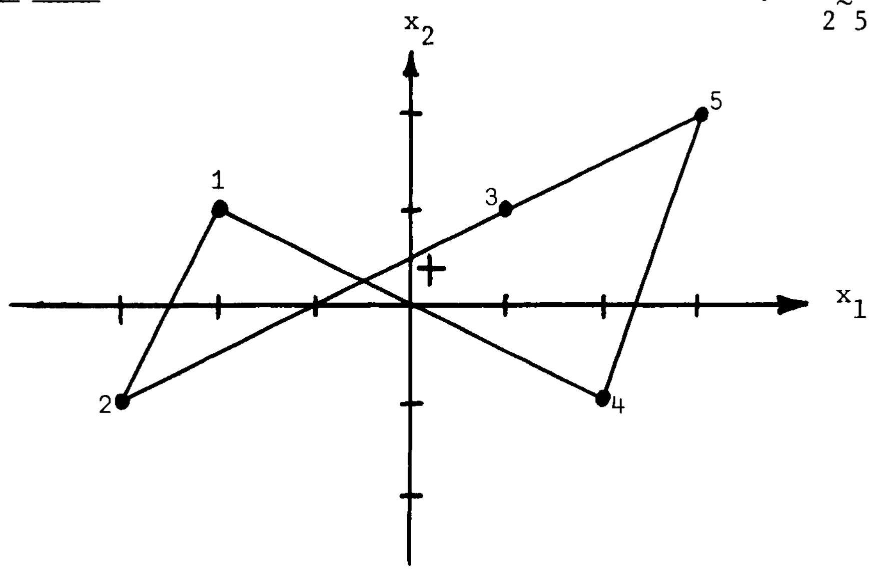
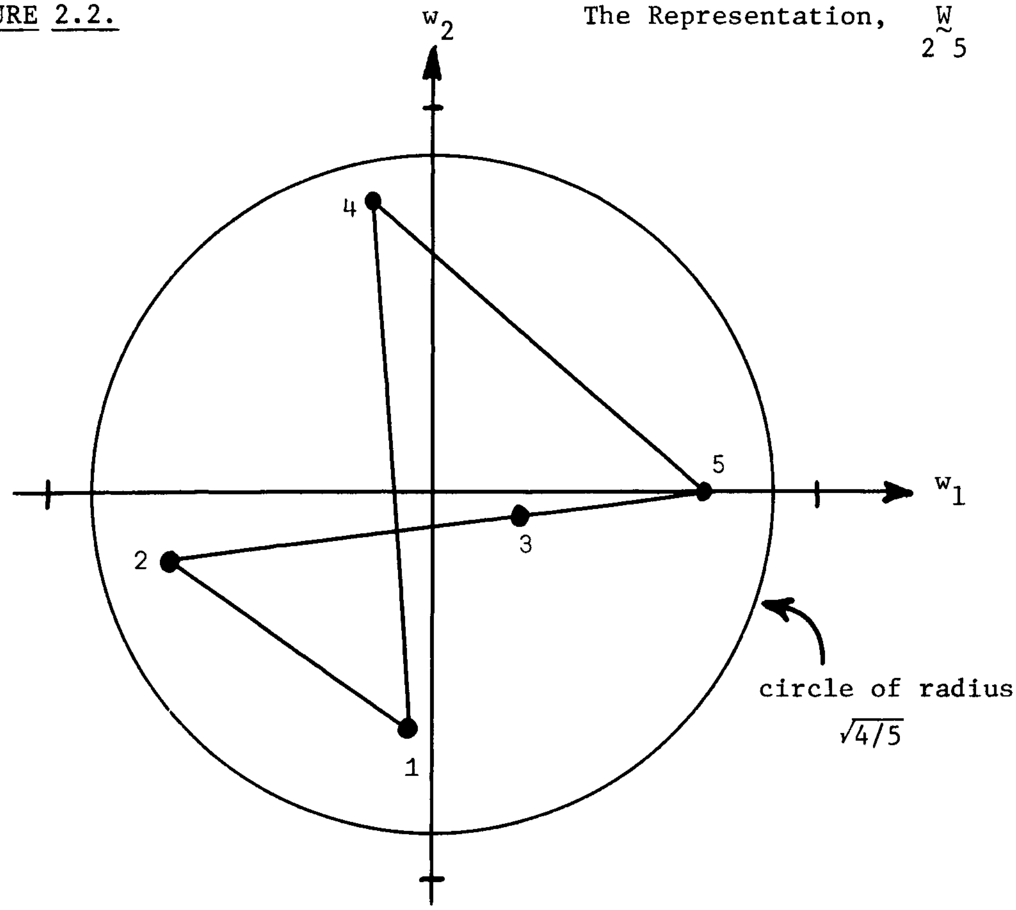


FIGURE 2.2. The Representation, $\underset{2\sim 5}{W}$



$(\underline{a}^{**}, \underset{\sim}{B}^{**}) \in L(p)$ transforms $\underset{\sim}{X}_{(1)}$ into $\underset{\sim}{X}_{(2)}$ where

$$\underset{p\;p}{\underset{\sim}{B}^{**}} = \underset{\sim}{E}_{(2)}^{-1} \begin{bmatrix} \underset{r\;r}{\underset{\sim}{B}^{*}} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{p-r\;p-r}{\underset{\sim}{B}} \end{bmatrix} \underset{\sim}{E}_{(1)} \tag{2.16}$$

and

$$\underset{p\;1}{\underline{a}^{**}} = \underline{\overline{X}}_{(2)} - \underset{\sim}{B}^{**}\underline{\overline{X}}_{(1)} \tag{2.17}$$

where $\underset{\sim}{B}$ is <u>any</u> $(p-r)\times(p-r)$ nonsingular matrix.

The calculation of the linearly invariant canonical form, $\underset{\sim}{W}$, will now be illustrated by a numerical example using artificial data with $p=r=2$ and $N=5$. The results of the calculations are presented in Table 2.1. The matrix $\underset{2\;5}{\underset{\sim}{W}^{*}}$ contains a pair of orthonormal c-vectors of $\underset{5\;5}{\underset{\sim}{P}}$ associated with c-roots of $+1$ calculated using the familiar "power method". Since $\underset{\sim}{W}^{*}$ does not have the property (2.17), "rotation" was required to get $\underset{\sim}{W}$. The raw data $\underset{\sim}{X}$ is plotted in Figure 2.1, and the representation is displayed in Figure 2.2. Note that the linear relationship between individuals 2, 3, and 5 is preserved in the representation. The matrix $\underset{\sim}{D}^2$ displayed in Table 2.1 contains the squared distances between the five individuals in the representation. Note that, as was not obvious in the raw data, our model implies that individual 2 is evidently closer to or more like individual 5 than individual 1 is like individual 4.

3. GENERAL PROPERTIES OF THE REPRESENTATION

The following lemma is useful in the stepwise construction of projection matrices. Let $\underline{v}'_\ell$ denote the ℓ-th row of $\underset{p\,N}{\underset{\sim}{X}}$ and write

$$\underset{k\,N}{\underset{\sim}{X}}^{(\underline{\ell}'_k)} = \begin{bmatrix} \underline{v}'_{\ell_1} \\ \vdots \\ \underline{v}'_{\ell_k} \end{bmatrix} \tag{3.1}$$

where $k \le p$ and $\underline{\ell}'_k = (\ell_1, \ldots, \ell_k)$ is an ordered subset of k integers from 1, 2, ..., p. Thus $\underset{\sim}{X} = \underset{\sim}{X}^{(1,2,\ldots,p)}$. Let $\underline{\pi}'(\underline{\ell}'_k) = (\pi_{\ell_1}, \ldots, \pi_{\ell_k})$ denote a permutation of the elements of $\underline{\ell}'_k$. Write $\underset{N\,N}{\underset{\sim}{P}}^{(\underline{\ell}'_k)}$ for the projection matrix calculated from $\underset{k\,N}{\underset{\sim}{X}}^{(\underline{\ell}'_k)}$, and define $\underset{N\,N}{\underset{\sim}{P}}^{(\phi)} = \underset{N\,N}{\underset{\sim}{0}}$ where ϕ denotes the empty set. Then we have

Lemma 3.1.

$$\underset{N\,N}{\underset{\sim}{P}}^{(\underline{\ell}'_k)} = \underset{\sim}{P}^{(\underline{\pi}'(\underline{\ell}'_k))} \tag{3.2}$$

$$= {}_N(\underset{\sim}{I} - \tfrac{1}{N}\underline{1}\,\underline{1}')\underset{N\,k}{\underset{\sim}{X}}^{(\underline{\ell}'_k)}[\underset{\sim}{X}^{(\underline{\ell}'_k)}(\underset{\sim}{I} - \tfrac{1}{N}\underline{1}\,\underline{1}')\underset{\sim}{X}^{(\underline{\ell}'_k)}]^{-}\underset{k\,N}{\underset{\sim}{X}}^{(\underline{\ell}'_k)}(\underset{\sim}{I} - \tfrac{1}{N}\underline{1}\,\underline{1}')_N \tag{3.3}$$

$$= \underset{\sim}{P}^{(\underline{\ell}'_{k-1})} + \underline{v}^*_{(\underline{\ell}'_k)}\underline{v}^{*\prime}_{(\underline{\ell}'_k)} \tag{3.4}$$

where

$$\underline{v}^{\perp}_{(\underline{\ell}'_k)} = (\underset{\sim}{I} - \underset{\sim}{P}^{(\underline{\ell}'_{k-1})})(\underset{\sim}{I} - \tfrac{1}{N}\underline{1}\,\underline{1}')\underline{v}_{\ell_k} \tag{3.5}$$

is the component of the ℓ_k-th row of $\underset{\sim}{Y} = \underset{\sim}{X}(\underset{\sim}{I} - \frac{1}{N}\underline{1}\,\underline{1}')$ which is orthogonal to the vectors in the vector space generated by the ℓ_1-th, ℓ_2-th, ..., and ℓ_{k-1}-th rows of $\underset{\sim}{Y}$,

and

$$\underset{\sim}{v}^*_{(\ell'_k)} = \begin{cases} \dfrac{1}{\sqrt{\underset{\sim}{v}^{\perp\prime}_{(\ell'_k)}\,\underset{\sim}{v}^{\perp}_{(\ell'_k)}}} \cdot \underset{\sim}{v}^{\perp}_{(\ell'_k)} & \text{if} \quad \underset{\sim}{v}^{\perp}_{(\ell'_k)} \neq \underline{0} \\ 0 & \text{otherwise .} \end{cases} \qquad (3.6)$$

Finally, $\underset{\sim}{P} = \underset{\sim}{P}^{(\pi_1,\pi_2,\ldots,\pi_p)}$, and $\underline{v}^*_{(\ell'_k)}$ is a c-vector of $\underset{\sim}{P}$ and is associated with a c-root of +1 unless $\underline{v}^*_{(\ell'_k)} = 0$.

Proof. Relationship (3.2) follows from the uniqueness of projection matrices and their invariance under transformations (1.3) on $\underset{\sim}{X}^{(\ell'_k)}$. The non-unique generalized inverse, "-", in equation (3.3) can be replaced by the uniquely determined Moore-Penrose inverse, "+", discussed in Rao [4]. The formula (3.4) then follows by noting that

$$\begin{bmatrix} \underset{\sim}{D} & \underline{e} \\ \underline{e}' & f \end{bmatrix}^+ = \left[\begin{array}{c|c} \underset{\sim}{D}^+ + \frac{1}{g}\,\underset{\sim}{D}^+\underline{e}\underline{e}'\underset{\sim}{D}^+ & -\frac{1}{g}\,\underset{\sim}{D}^+\underline{e} \\ \hline -\frac{1}{g}\,\underline{e}'\underset{\sim}{D}^+ & \frac{1}{g} \end{array}\right]$$

if

$$\underset{\sim}{D} = \underset{\sim}{D}', \quad g = f - \underline{e}'\underset{\sim}{D}^+\underline{e} \neq 0, \quad \text{and } \underline{e}'(\underset{\sim}{I} - \underset{\sim}{D}^+\underset{\sim}{D}) = \underline{0}'.$$

In our case $g = \underline{v}^{\perp\prime}_{(\ell'_k)}\underline{v}^{\perp}_{(\ell'_k)}$, so it is clear that $\underset{\sim}{P}^{(\ell'_k)} = \underset{\sim}{P}^{(\ell'_{k-1})}$ when $g=0$. Following (3.5), we have given the geometrical interpretation of the results of the calculation (3.4).

Lemma 3.2. The sample mean of the projections of the representation on any direction in r-space is zero. The sample sums of squares of projections on any direction is one. The sample covariance between the projections on any two orthogonal directions is zero.

Proof. Let $\underline{e}$ be any $r \times 1$ vector such that $\underline{e}'\underline{e} = 1$. Then $\underset{1\ r\ N}{\underline{e}'\underset{\sim}{W}}$ is the row vector of projections of the representation on some direction. Then $\underset{1\ r\ N\ 1}{\underline{e}'\underset{\sim}{W}\,\underline{1}} = \underset{1\ r\ 1}{\underline{e}'\underline{0}} = 0$ by relation (2.10), and $\underset{1\ r\ N\ r\ 1}{\underline{e}'\underset{\sim}{W}\,\underset{\sim}{W}'\underline{e}} = \underset{1\ r\ 1}{\underline{e}'\underline{e}} = 1$ by relation (2.9). Finally, if $\underline{e}^*$ is such that $\underline{e}^{*\prime}\underline{e}^* = 1$, then $\underset{1\ r\ N\ r\ 1}{\underline{e}'\underset{\sim}{W}\,\underset{\sim}{W}'\underline{e}^*} = \underset{1\ r\ 1}{\underline{e}'\underline{e}^*} = 0$ if and only if $\underline{e}$ and $\underline{e}^*$ represent orthogonal directions.

In the following four lemmas, we will study the representation in terms of the distances between its points. Let $d^2_{\ell m}$ denote the squared distance between the representations of the ℓ-th and m-th individuals. Then we have

Lemma 3.3.

$$d^2_{\ell m} = p_{\ell\ell} + p_{mm} - p_{m\ell} - p_{\ell m} \tag{3.7}$$

$$= (\underline{X}_m - \underline{X}_\ell)'\underset{\sim}{S}^-(\underline{X}_m - \underline{X}_\ell) \tag{3.8}$$

where $0 \le d^2_{\ell m} \le 2$ and $\underset{\sim}{S}$ is given by (2.4).

Proof. The expression (3.7) follows immediately from (2.8) as pointed out by Gower [1]. Then (3.8) follows by making use of the fact that

$$p_{\ell m} = p_{m\ell} = (\underline{X}_m - \overline{\underline{X}})'\underset{\sim}{S}^-(\underline{X}_\ell - \overline{\underline{X}}) \tag{3.9}$$

Indeed, a particular choice for $\underset{\sim}{S}^-$ may not be symmetric although $\underset{\sim}{S}$ is symmetric, but $\underset{\sim}{P}$ is symmetric and unique. Thus $d^2_{\ell m} \ge 0$ is uniquely determined. Now $\underset{\sim}{I}_N - \underset{\sim}{P}$ is symmetric, idempotent, of rank N-r, and orthogonal to $\underset{\sim}{P}$. Thus the squared distance in the space orthogonal to the row space of $\underset{\sim}{Y}$ is

$$(d^\perp_{\ell m})^2 = (1 - p_{\ell\ell}) + (1 - p_{mm}) - 2(-p_{\ell m}) = 2 - d^2_{\ell m} \ge 0.$$

Before proceeding to the next lemma, it should be noted that the specific measure of distance (3.8) was considered by Gower [1] in §4.2. Gower states that use of this measure of distance "reflects the attitude that the group of individuals is homogeneous" essentially because of the properties of the representation that we demonstrated in Lemma 3.2. There is, however, reason to question this assertion. First of all, one notes that the distances between points will almost surely be greater than zero and relatively different.

Suppose two characteristics are measured on a population such that the bivariate density of the characteristics has unusual equal probability density contours. These contours would be circles or ellipses for the Bivariate Normal distribution, so suppose the contours are squares or crescents or are not even closed or connected. Then data from this population and its corresponding representation would display the same behavior in probability. For example, the artificial raw data of Figure 3.1 has the canonical form shown in Figure 3.2 in which the original circles become ellipses. If actual data were to split like this into two clusters, there would be empirical evidence for arguing that the population consists of two types of individuals (e.g. men and women) so that the characteristics are distributed as a mixture of two distinct distributions. The linearly invariant canonical form can thus display considerable heterogeneity among the individuals in the sample.

Now let d_{mo}^2 denote the squared distance between the representation of the m-th point and the origin. Then we have

<u>Lemma</u> <u>3.4</u>.

$$d_{mo}^2 = p_{mm} = (\underline{X}_m - \underline{\bar{X}})'\underset{\sim}{S}^-(\underline{X}_m - \underline{\bar{X}}) \tag{3.10}$$

where $0 \le d_{mo}^2 \le \frac{N-1}{N}$ and $\sum_{m=1}^{N} d_{mo}^2 = r \le N-1$.

FIGURE 3.1. The Raw Data, $\underset{\sim}{X}_{2\ 15}$

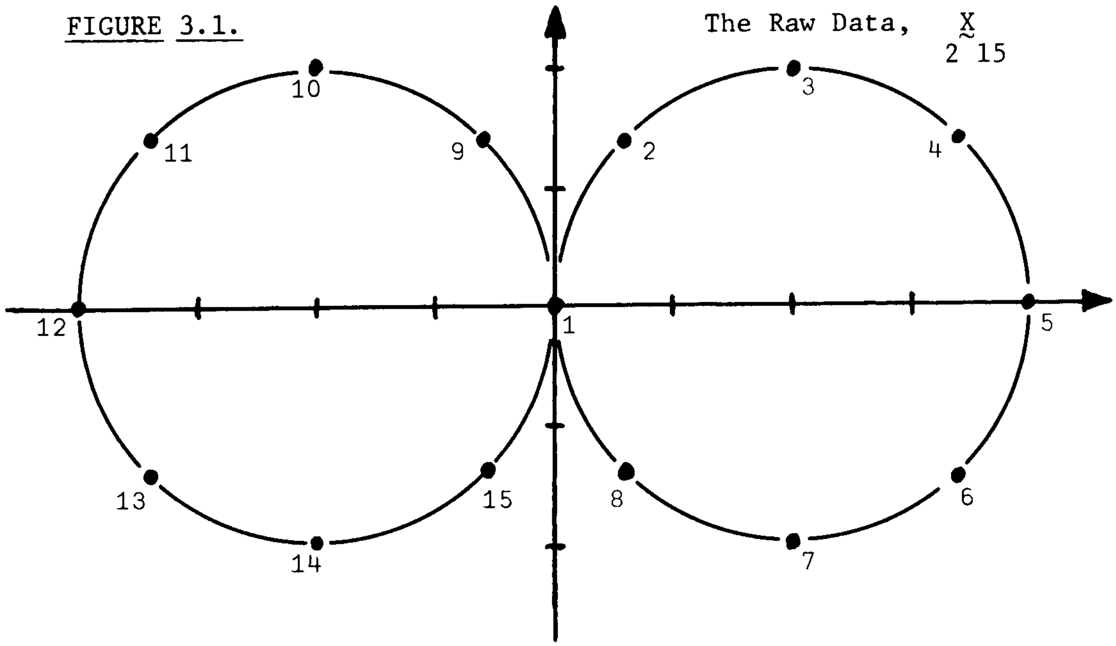
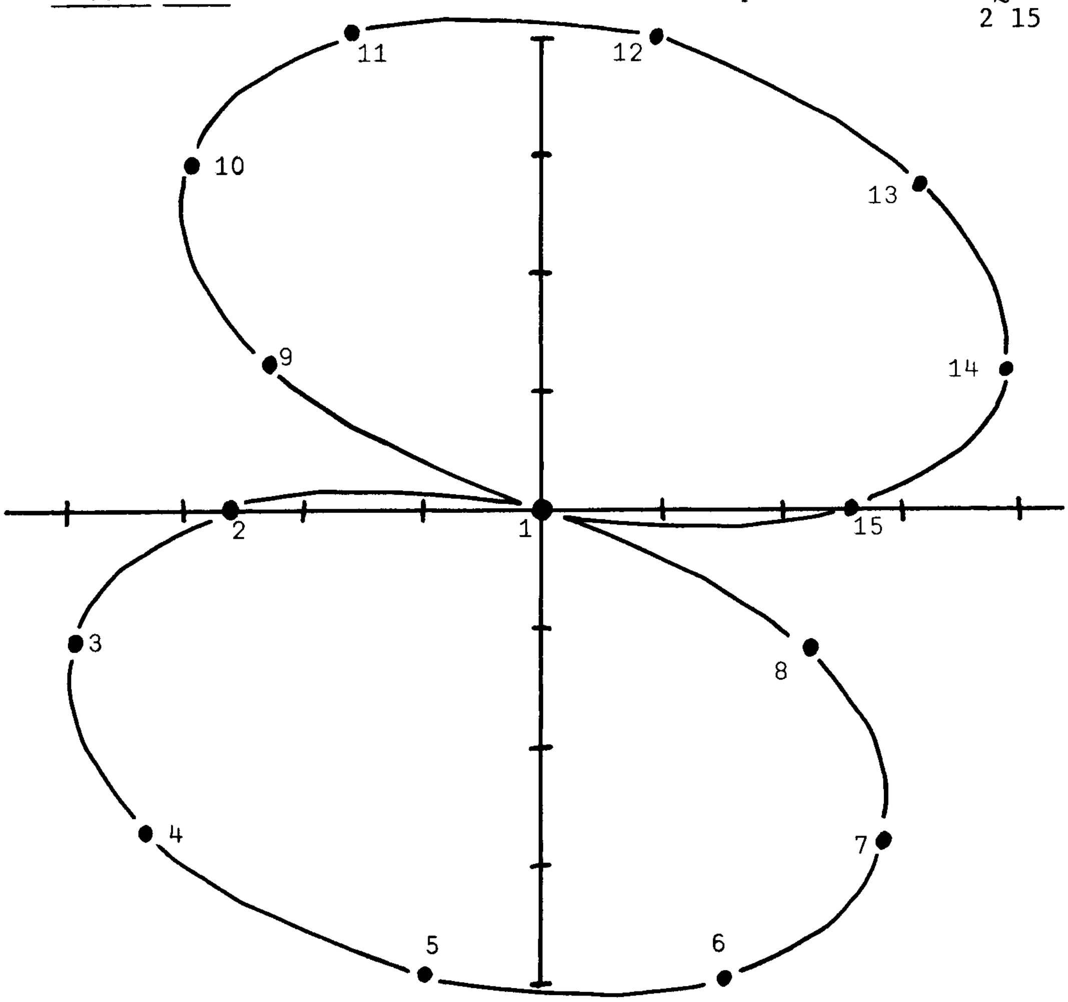


FIGURE 3.2. The Representation, $\underset{\sim}{W}_{2\ 15}$



Proof. The result (3.10) follows immediately from (2.8) as before. To give another interpretation of this result, form the $p \times (N+1)$ matrix $\underset{\sim}{X}^* = (\bar{\underline{X}} \mid \underset{\sim}{X})$ and note that $\underset{\sim}{Y}^* = (\underline{0} \mid \underset{\sim}{Y})$. Thus $\underset{N+1\ N+1}{\underset{\sim}{P}^*} = \begin{pmatrix} 0 & \underline{0}' \\ \underline{0} & \underset{\sim}{P} \end{pmatrix}$ implies that $d^2_{mo} = p_{mm} + 0 - 2(0)$ is the squared distance between the representation of the m-th individual and the representation of $\bar{\underline{X}}$ using Lemma 3.3. Now $\underset{\sim}{I}_N - \frac{1}{N}\underline{1}\,\underline{1}' - \underset{\sim}{P}$ is symmetric, idempotent, of rank $N-r-1$, orthogonal to $\underset{\sim}{P}$, and orthogonal to $\underline{1}$. The squared distance between the m-th point and the origin in this subspace is thus $1 - \frac{1}{N} - d^2_{mo} \geq 0$. Finally

$$\sum_{m=1}^{N} d^2_{mo} = \sum_{m=1}^{N} p_{mm} = \text{tr}\underset{\sim}{P} = \text{rank}\ \underset{\sim}{P} = r.$$

Lemma 3.5. $r = N-1 \leq p$ implies that $\underset{\sim}{P} = \underset{\sim}{I} - \frac{1}{N}\underline{1}\,\underline{1}'$

and

$$\underset{N-1\ N}{\underset{\sim}{W}} = \begin{bmatrix} \frac{-1}{\sqrt{(N-1)N}} & \frac{-1}{\sqrt{(N-1)N}} & & \frac{-1}{\sqrt{(N-1)N}} & \frac{N-1}{\sqrt{(N-1)N}} \\ \frac{-1}{\sqrt{(N-2)(N-1)}} & \frac{-1}{\sqrt{(N-2)(N-1)}} & & \frac{N-2}{\sqrt{(N-2)(N-1)}} & 0 \\ & & & & \\ \frac{-1}{\sqrt{1.2}} & \frac{+1}{\sqrt{1.2}} & 0 & & 0 \end{bmatrix}$$

$$= \underset{\sim}{Q}^{(N)'} \qquad (3.11)$$

The representation is of N equally spaced points in $N-1$ dimensions; each point is of distance $\sqrt{2}$ from every other point and of distance $\sqrt{\frac{N-1}{N}}$ from the origin.

<u>Proof</u>. $\underset{\sim}{P} = \underset{\sim}{P}' = \underset{\sim}{P}^2$, $\underset{\sim}{P}\underline{1} = \underline{0}$, and rank $\underset{\sim}{P} = N-1$ imply that $\underset{\sim}{P}$ is the projection matrix for the N-1 dimensional space of N-vectors orthogonal to the constant vector $\underline{1}$. Thus $\underset{\sim}{P}(\underset{\sim}{I} - \frac{1}{N}\underline{1}\,\underline{1}') = (\underset{\sim}{I} - \frac{1}{N}\underline{1}\,\underline{1}')$. On the other hand, $\underset{\sim}{P}(I - \frac{1}{N}\underline{1}\,\underline{1}') = \underset{\sim}{P}$ is obvious by matrix multiplication. Thus we have $\underset{N-1\ N}{\underset{\sim}{W}} = \underset{\sim}{Q}^{(N)'}$ where $\underset{\sim}{Q}^{(N)}$ is our notation for the semiorthogonal matrix associated with the Helmert transformation. Then $\underline{W}_{\ell}'\underline{W}_m = p_{\ell m} = \delta_{\ell m} - \frac{1}{N}$ where $\delta_{\ell m}$ denotes Kronecker Delta. Therefore $\underline{W}_{\ell}'\underline{W}_{\ell} = \frac{N-1}{N}$ and $(\underline{W}_{\ell} - \underline{W}_m)'(\underline{W}_{\ell} - \underline{W}_m) = 2$.

In the next lemma, we will investigate the properties of average squared distance in the representation with special attention to the case when $c > 1$ populations are under consideration. Let

$$\bar{d}^2(i, j) = \begin{cases} \binom{n_j}{2}^{-1} \sum\limits_{\substack{\ell, m \in \pi_i \\ \ell < m}} d^2_{\ell m} & \text{if } i = j \\ \frac{1}{n_i n_j} \sum\limits_{\substack{\ell \in \pi_i \\ m \in \pi_j}} d^2_{\ell m} & \text{if } i < j \end{cases} \tag{3.12}$$

where $1 \le i \le j \le c$ and Π_i is as used in (2.5). Then we have

<u>Lemma 3.6</u>.

$$\bar{d}^2(i, i) = \mathrm{tr}\{\underset{\sim}{S}^-[\frac{2}{n_i - 1}\underset{\sim}{S}_{(i)}]\} \tag{3.13}$$

$$\bar{d}^2(i, j) = \mathrm{tr}\{\underset{\sim}{S}^-[\frac{1}{n_i}\underset{\sim}{S}_{(i)} + \frac{1}{n_j}\underset{\sim}{S}_{(j)} + (\underline{\bar{X}}^{(i)} - \underline{\bar{X}}^{(j)})(\underline{\bar{X}}^{(i)} - \underline{\bar{X}}^{(j)})']\} \quad i \ne j \tag{3.14}$$

where $\underset{\sim}{S}_{(i)}$ and $\underline{\bar{X}}^{(i)}$ are as in (2.5) and (2.6). Furthermore

$$\sum_{\ell \in \pi_i} d^2_{\ell o} = \mathrm{tr}\{\underset{\sim}{S}^-[\underset{\sim}{S}_{(i)} + n_i(\underline{\bar{X}}^{(i)} - \underline{\bar{X}})(\underline{\bar{X}}^{(i)} - \underline{\bar{X}})']\} \tag{3.15}$$

Finally

$$\bar{d}^2 = \binom{N}{2}^{-1} \sum_{\ell < m} d^2_{\ell m} = \frac{2r}{N-1} \tag{3.16}$$

is non-stochastic except for its dependence upon r .

Proof. Write $\underset{\sim}{d}_o^{2\prime} = (d^2_{1o}, \ldots, d^2_{No}) = \text{diag}(\underset{\sim}{P})$. We have $\underset{\sim}{D}^2 = \underset{\sim}{d}_o^2 \underset{\sim}{1}' + \underset{\sim}{1} \underset{\sim}{d}_o^{2\prime} - 2\underset{\sim}{P}$ from Lemma 3.3. Then $\text{diag}(\underset{\sim}{D}^2) = \underset{\sim}{0}'$ expresses the fact that each individual is of distance zero from itself, and $\underset{\sim}{1}'\underset{\sim}{d}_o^2 = \text{trace } \underset{\sim}{P} = r$. Then $\underset{\sim}{1}'\underset{\sim}{D}^2\underset{\sim}{1} = 2Nr$ shows result (3.16), while the other results of the lemma follow from direct algebraic manipulation.

The many properties of the representation we have considered imply restrictions. Thus we have

Lemma 3.7. $\underset{r\ N}{\underset{\sim}{W}}$ has rN elements but only $r(N-r-1)$ functionally independent variables; the distribution of $\underset{\sim}{W}$ is highly singular.

Proof. If $\underset{p\ N}{\underset{\sim}{X}}$ is considered to contain pN variates which are not functionally dependent, then $(p-r)N$ "independent" variates are lost by restricting attention to linearly combinable variables. This loss will usually be zero because only linear redundancies are detected and these usually occur with probability zero. Next, r constraints are added in achieving translation invariance (1.2) and imply property (2.10), while $\frac{r(r+1)}{2}$ constraints are added in achieving invariance with respect to nonsingular linear transformations (1.3) and imply property (2.9). Finally, $\frac{r(r-1)}{2}$ constraints are added to make the choice of $\underset{\sim}{W}$ unique and imply the property (2.11).

4. SOME EXACT DISTRIBUTIONS FOR NORMAL POPULATIONS

4.1. A Single Multivariate Normal Population.

James [2] shows how locally defined exterior differential forms can be used to derive sampling distributions associated with a Multivariate Normal population. His result on the decomposition of a random sample will be reinterpreted here. Take $c = 1$, $N = n_1$, and suppose that

$$\underset{p\ N}{\underset{\sim}{X}} \overset{d}{=} N_{pN}(\underset{p}{\underline{\mu}}\ \underset{1\ N}{\underline{1}'};\ \underset{\sim}{I}_N \otimes \underset{p\ p}{\underset{\sim}{\Sigma}}) \tag{4.1}$$

where $\underset{\sim}{\Sigma}$ is positive definite. Then $\Pr(r=p) = 1$ if $N > p$ and we have that $\underset{\sim}{S}^- = \underset{\sim}{S}^{-1}$ almost surely. Now (4.1) implies

$$dF(\underset{p\ N}{\underset{\sim}{X}}) = \frac{|\underset{\sim}{\Sigma}|^{-N/2}}{(2\pi)^{\frac{pN}{2}}} \operatorname{etr}\{-\tfrac{1}{2}\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{X} - \underline{\mu}\underline{1}')(\underset{\sim}{X}' - \underline{1}\underline{\mu}')\} \prod_{\substack{1 \le i \le p \\ 1 \le j \le N}} dX_{ij} \ . \tag{4.2}$$

We now consider the Helmert Transformation to get

$$\left[\sqrt{N}\underline{\bar{X}} \;\middle|\; \underset{p\ N-1}{\underset{\sim}{Z}}\right] = \underset{p\ N}{\underset{\sim}{X}}\left[\frac{1}{\sqrt{N}}\underline{1} \;\middle|\; \underset{N\ N-1}{\underset{\sim}{Q}^{(N)}}\right] \overset{d}{=} N_{pN}([\sqrt{N}\underline{\mu} \mid \underset{\sim}{0}];\ \underset{\sim}{I}_N \otimes \underset{p\ p}{\underset{\sim}{\Sigma}}) \ . \tag{4.3}$$

We now transform $\underset{\sim}{Z}$ as in James [2], page 69:

$$\underset{p\ N-1}{\underset{\sim}{Z}} = \underset{p\ p}{\underset{\sim}{G}}\ \underset{p}{\underset{\sim}{L}}\ \underset{p\ N-1}{\underset{\sim}{A}} \tag{4.4}$$

where $\underset{\sim}{L}^2$ is the diagonal matrix of c-roots of $\underset{\sim}{Z}\underset{\sim}{Z}'$, $\underset{\sim}{G}$ is the orthogonal matrix of c-vectors, and $\underset{\sim}{A} = \underset{\sim}{L}^{-1}\underset{\sim}{G}'\underset{\sim}{Z}$. Then

$$\underset{p\ N-1}{\underset{\sim}{Z}}\ \underset{p}{\underset{\sim}{Z}'} = \underset{p\ N}{\underset{\sim}{X}}(\underset{\sim}{I} - \frac{1}{N}\underline{1}\ \underline{1}')\underset{N\ p}{\underset{\sim}{X}'} = \underset{\sim}{S} = \underset{\sim}{G}\underset{\sim}{L}^2\underset{\sim}{G}' \tag{4.5}$$

does not depend upon $\underset{\sim}{A}$ because $\underset{p\ N-1}{\underset{\sim}{A}}\ \underset{p}{\underset{\sim}{A}'} = \underset{\sim}{I}_p$, while

$$\underset{\sim}{Z}'(\underset{\sim}{Z}\underset{\sim}{Z}')^{-1}\underset{\sim}{Z} = \underset{N-1\ p\ N-1}{\underset{\sim}{A}'\underset{\sim}{A}} = \underset{N-1}{\underset{\sim}{Q}^{(N)'}}\ \underset{N}{\underset{\sim}{P}}\ \underset{N\ N-1}{\underset{\sim}{Q}^{(N)}} \tag{4.6}$$

does not depend on $\underset{\sim}{G}$ or $\underset{\sim}{L}$ (and therefore $\underset{\sim}{S}$). Now we can write

$$\underset{p\ N-1}{\underset{\sim}{A}} = \underset{p\ p}{\underset{\sim}{C}}\ \underset{p\ N-1}{\underset{\sim}{H}} \qquad (4.7)$$

where the orthogonal matrix $\underset{\sim}{C}$ is chosen so that $\underset{\sim}{H}$ has property (2.11) with $r \to p$ and $N \to N-1$. Note that $\underset{\sim}{H}$ does not have property (2.10). Finally we have

$$\underset{p\ N}{\underset{\sim}{W}} = \underset{p\ N-1}{\underset{\sim}{H}}\ \underset{N}{\underset{\sim}{Q}}^{(N)'} \ . \qquad (4.8)$$

We can now restate Theorem 8.1 of James [2] as summarized in equation (8.22) of [2].

<u>THEOREM</u> <u>4.1</u>. Let $\underset{p\ N}{\underset{\sim}{X}}$ have the probability element (4.2), and let

$$\underset{p\ N}{\underset{\sim}{W}} = (\underline{W}_1 \,\vdots\, \cdots \,\vdots\, \underline{W}_N) = \begin{bmatrix} \underline{\omega}_1' \\ \vdots \\ \underline{\omega}_p' \end{bmatrix}$$

be the Canonical Form of $\underset{\sim}{X}$. Let

$$\underset{N-p-1\ N}{\underset{\sim}{B}} = \begin{bmatrix} \underline{b}_1' \\ \vdots \\ \underline{b}_{N-p-1}' \end{bmatrix}$$

be chosen so that the $N \times N$ matrix

$$\begin{bmatrix} \frac{1}{\sqrt{N}}\ \underline{1}' \\ \hline \underset{\sim}{W} \\ \hline \underset{\sim}{B} \end{bmatrix}$$

is orthogonal. Then the distribution of $\underset{\sim}{X}$ can be decomposed into four independent distributions:

$$dF(\underset{p\ N}{\underset{\sim}{X}}) = dF(\underline{\overline{X}}) \cdot dF(\underset{\sim}{S}) \cdot dF(\underset{\sim}{W}) \cdot dF(\underset{\sim}{C}) \qquad (4.9)$$

where (1) $\bar{X} \overset{d}{=} N_p(\mu, \frac{1}{N}\Sigma)$ so that

$$dF(\bar{X}) = \frac{N^{\frac{p}{2}}|\Sigma|^{-1/2}}{(2\pi)^{p/2}} \exp\{-\frac{N}{2}(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu)\} \prod_{i=1}^{p} d\bar{X}_i \,, \tag{4.10}$$

(2) $\underset{p\,p}{S} \overset{d}{=} W_p(N-1, \Sigma)$ so that

$$dF(S) = \frac{|\Sigma|^{-\frac{N-1}{2}} |S|^{\frac{N-p-2}{2}}}{2^{\frac{p(N-1)}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma(\frac{N-i}{2})} \operatorname{etr}\{-\frac{1}{2}\Sigma^{-1}S\} \prod_{i \le j} dS_{ij} \tag{4.11}$$

does not depend upon μ

(3) $\underset{p\,N}{W}$ is distributed on the subset of $G_{p,N-p}$ with property (2.10) [homeomorphic to $G_{p,N-p-1}$ because of (4.8)] so that

$$dF(W) = \left[\pi^{-\frac{p(N-p-1)}{2}} \cdot \prod_{i=1}^{p} \frac{\Gamma(\frac{N-i}{2})}{\Gamma(\frac{i}{2})}\right] \cdot \prod_{j=1}^{N-p-1} \prod_{i=1}^{p} b_j' d\omega_i \tag{4.12}$$

does not depend upon μ or Σ, and

(4) C, as chosen in (4.7), has the invariant distribution on the orthogonal group so that

$$dF(C) = 2^{-p}\pi^{-\frac{p(p+1)}{4}} \prod_{i=1}^{p} \Gamma(\frac{i}{2}) \prod_{i<j} c_j' dc_i \tag{4.13}$$

does not depend upon μ or Σ, where $C = (c_1 \mid \cdots \mid c_p)$.

The above results follow because James shows that H of (4.7) has the "invariant distribution" on $G_{p,N-p-1}$. However, this invariance is with respect to the transformations $\underset{p\;N-1}{Z} \to \underset{p\;N-1}{Z}\,\underset{N-1}{U}$ where U is any orthogonal matrix. Thus we have

COROLLARY 4.2. If $\underset{\sim}{X}$ has the distribution (4.1), then $\underset{\sim}{W}$ is invariant in distribution under transformations of the form

$$\underset{p\,N}{\underset{\sim}{X}} \rightarrow \underset{p\,1}{\overline{X}}\;\underset{N}{\underline{1}'} + \underset{p\,N}{\underset{\sim}{X}}\;\underset{N\,N}{\underset{\sim}{V}} \qquad (4.14)$$

where

$$\underset{\sim}{V} = \underset{N\,N-1}{\underset{\sim}{Q}^{(N)}}\;\underset{N-1\,N-1}{\underset{\sim}{U}}\;\underset{N-1\,N}{\underset{\sim}{Q}^{(N)\prime}} \qquad (4.15)$$

and $\underset{\sim}{U}$ is orthogonal, as well as strictly invariant under transformations of the form (1.2) and (1.3).

Note that the transformation matrices (4.15) have the property

$$\underset{\sim}{V}\underset{\sim}{V}' = \underset{\sim}{V}'\underset{\sim}{V} = \underset{\sim}{I} - \frac{1}{N}\underline{1}\,\underline{1}' .$$

Thus $\underset{\sim}{V}$ acts like a projection matrix to remove the row means of $\underset{\sim}{X}$ and then "rotates" these projections. Note also that all reorderings of the columns of $\underset{\sim}{X}$ are included in (4.14). For example,

$$\underset{N-1\,N-1}{\underset{\sim}{U}} = \begin{bmatrix} -1 & \underline{0}' \\ \underline{0} & \underset{\sim}{I}_{N-2} \end{bmatrix} \quad \text{implies} \quad \underset{N\,N}{\underset{\sim}{V}} = \begin{bmatrix} 0 & 1 & \underline{0}' \\ 1 & 0 & \underline{0}' \\ \underline{0} & \underline{0} & \underset{\sim}{I}_{N-2} \end{bmatrix} - \frac{1}{N}\underline{1}\,\underline{1}' \qquad (4.16)$$

so that the data for the first and second individuals are interchanged. In fact, the distribution of $\underset{\sim}{W}$ has to be invariant under reordering of the columns of $\underset{\sim}{X}$ when $c=1$ whether or not $\underset{\sim}{X}$ has the distribution (4.1) because the columns of $\underset{\sim}{X}$ are independent and identically distributed in this case.

COROLLARY 4.3. Let $\underset{p\,N}{\underset{\sim}{X}}$ have the distribution (4.1), let $\underset{\sim}{P}$ be the projection matrix calculated from $\underset{\sim}{X}$, and let $\underset{\sim}{D}^2$ be the matrix of squared distances between points in the representation as in Lemma 3.6. Let the rows of $\underset{p\,N}{\underset{\sim}{W}^*}$ be any orthonormal set of c-vectors of $\underset{\sim}{P} = \frac{1}{2}(\underline{d}_o^2\underline{1}' + \underline{1}\underline{d}_o^{2\prime} - \underset{\sim}{D}^2)$

associated with c-roots of +1, and let the rows of $\underset{N-p-1\ N}{B^*}$ be any orthonormal set of c-vectors of $I - \frac{1}{N}\underline{1}\,\underline{1}' - P$ associated with c-roots of +1. Then $dF(W)$ in decomposition (4.9) can be replaced by

$$dF(P) = K(p, N) \prod_{j=1}^{N-p-1} \prod_{i=1}^{p} \underline{b}_j^{*\prime}(dP)\underline{\omega}_i^* \qquad (4.12')$$

or

$$dF(D^2) = K(p, N) \prod_{j=1}^{N-p-1} \prod_{i=1}^{p} \underline{b}_j^{*\prime}(-\tfrac{1}{2}\, dD^2)\underline{\omega}_i^* \qquad (4.12'')$$

where $K(p, N)$ is the constant term from (4.12).

Proof. $dP = \frac{1}{2}(d\underline{d}_o^2\underline{1}' + \underline{1}d\underline{d}_o^{2\prime} - dD^2) = dW^{*\prime}W^* + W^{*\prime}dW^*$. Thus $B^*(dP)W^{*\prime} = B^*(-\frac{1}{2}\, dD^2)W^{*\prime} = B^*dW^{*\prime}$ because $\underset{1\ N\ p}{\underline{1}'W^{*\prime}} = \underset{1\ p}{\underline{0}'}$, $\underset{N-p-1\ N\ 1}{B^*\underline{1}} = \underset{N-p-1\ 1}{\underline{0}}$, $W^*W^{*\prime} = I_p$, and $\underset{N-p-1\ N\ p}{B^*W^{*\prime}} = \underset{N-p-1\ p}{0}$. So, taking W^* rather than W as our "reference" p-frame in $V_{p,N}$ (James [2], p. 61), we arrive at the probability elements (4.12') and (4.12") by noting that the double product of linear differential forms in (4.12) is simply the exterior product of the elements of $B^*dW^{*\prime}$. Note that P can be constructed from $-\frac{1}{2}D^2$ without knowing $\underline{d}_o^2$, the vector of squared distances from the origin, using the technique of "Principle Co-ordinage Analysis" due to Gower [1].‡ The set of possible realizations of P (the "Projection group") and the set of possible realizations of D^2 are thus both homeomorphic to the Grassmann manifold $G_{p,N-p-1}$.

The probability element (4.12) for W is expressed in terms of locally defined exterior differential forms and cannot be simplified in general. Let us consider the very simple case $p = 1$ and $N = 3$ to get some idea of the sort of distribution implied by (4.12). Then $\underset{1\ 3}{W} = (W_1\ W_2\ W_3)$ contains $p(N-p-1) = 1$ functionally independent variable because (2.9) implies

‡ $p_{ij} = -\frac{1}{2}d_{ij}^2 + \frac{1}{2}\bar{d}_{i.}^2 + \frac{1}{2}\bar{d}_{.j}^2 - \frac{1}{2}\bar{d}_{..}^2$, the symbol "." indicating averaging over the corresponding entire row of column of D^2

$W_1^2 + W_2^2 + W_3^2 = 1$, (2.10) implies $W_1 + W_2 + W_3 = 0$, and (2.11) implies $W_3 > 0$. We have

$$f(W_1, W_2, W_3) = \begin{cases} \frac{1}{\pi}\sqrt{\frac{3}{2-3W_3^2}} & \text{if } 0 < W_3 < \sqrt{\frac{2}{3}} \\ & W_1 = -\frac{W_3}{2} + a\frac{\sqrt{2-3W_3^2}}{2} \\ & \text{and } W_2 = -\frac{W_3}{2} - a\frac{\sqrt{2-3W_3^2}}{2} \\ & \text{where } a = +1 \text{ or } -1 \\ 0 & \text{otherwise} \end{cases} \quad (4.17)$$

Note that the marginal distribution of $\frac{3W_i^2}{2}$ for $i = 1, 2,$ or 3 is Beta with parameters $\frac{1}{2}$ and $\frac{1}{2}$. Figure 4.1 displays the marginal density of W_3 , and Figure 4.2 shows the possible values of W_1 and W_2 given W_3 . Thus we see that, starting with random variables having independent and identical distributions on the entire real line and central tendency displayed by the familiar bell shaped curve, we end up with dependent, bounded random variables that have a tendency to assume their extreme values.

The case $p = 1$ and general N can be handled by making a polar transformation as in James [2], equation (5.7). It is then seen that the distribution of $\underset{1\ N}{\tilde{W}}$ is closely related to the $N-2$ variate Dirichlet distribution with $N-1$ parameters all equal to $\frac{1}{2}$. It follows that the marginal distribution of $\frac{N\,W_i^2}{N-1}$ for $i = 1, 2, \ldots, N$ is Beta with parameters $\frac{1}{2}$ and $\frac{N-2}{2}$; the marginal distribution of W_i for $i = 1, 2, \ldots, N$ is positively skew for large N rather than negatively skew as in Figure 4.1 for $N = 3$.

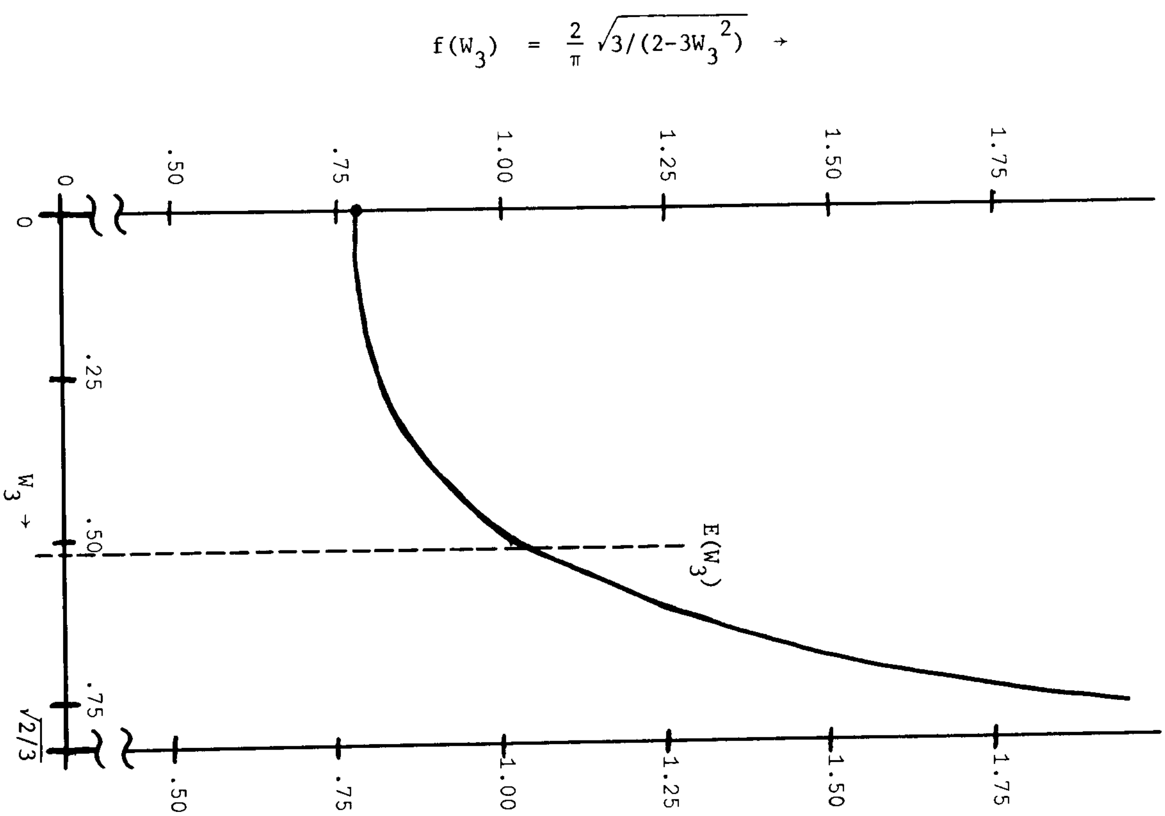


FIGURE 4.1.

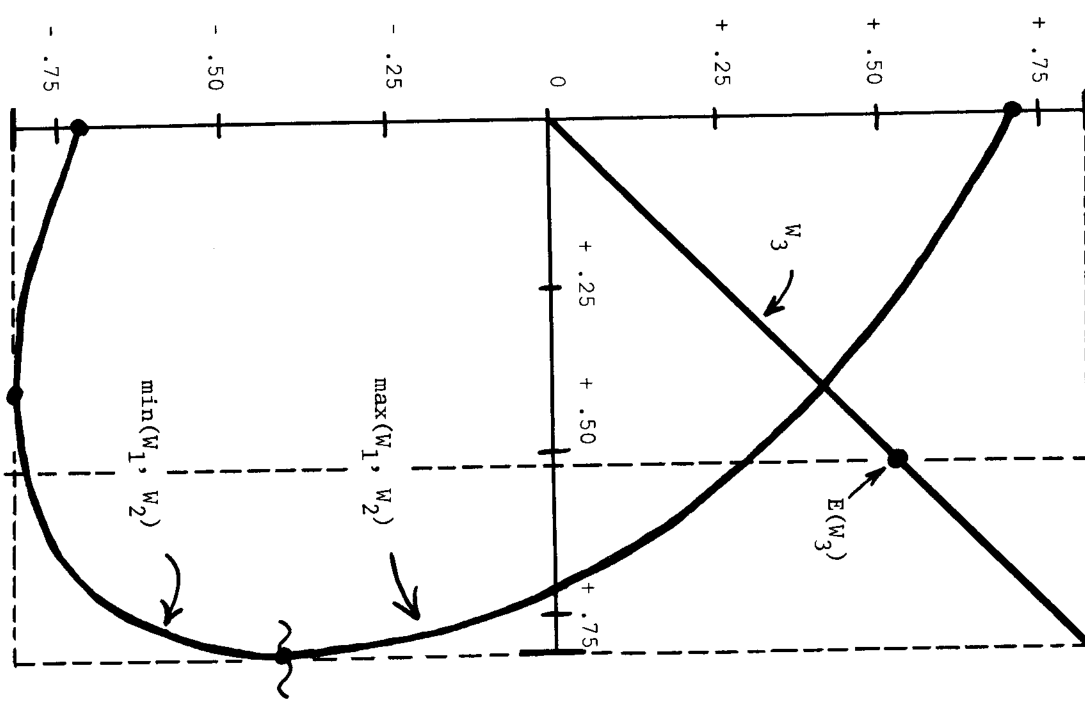


FIGURE 4.2.

4.2. Comments on Several Multivariate Normal Populations.

The distribution of $\underset{\sim}{W}_{p\,N}$ is unknown in this case. If attention is restricted to the special case $\underset{\sim}{\Sigma}_1 = \ldots = \underset{\sim}{\Sigma}_c$, $\underset{\sim}{S}$ will have the noncentral Wishart distribution. The distribution of $\underset{\sim}{W}$ does not follow from the type of analysis performed by James [2] when $c = 1$ because $\underset{\sim}{S}$ and $\underset{\sim}{W}$ are not independent. Thus, when James [3] derived the marginal distribution of $\underset{\sim}{S}$, he did not integrate over the possible realizations of $\underset{\sim}{W}$. Instead, he considered the integral of the density of the Multivariate Normal distribution $N(\underset{\sim}{M}_{p\,N-1}\underset{\sim}{U}_{N-1\,N-1} ; \underset{\sim}{U}'_{N-1\,N-1}\underset{\sim}{U}_{N-1} \otimes \underset{\sim}{\Sigma}_{p\,p})$ with respect to the invariant measure for $\underset{\sim}{U} \in O(N-1)$, the orthogonal group of order $N-1$. He thus replaced the joint density of $\underset{\sim}{S}$ and $\underset{\sim}{W}$ by the density of $\underset{\sim}{S}$ and a statistic independent of $\underset{\sim}{S}$. If one were to integrate the density of $N(\underset{\sim}{C}_{p\,p}\underset{\sim}{M}_{p\,N-1} ; \underset{\sim}{I}_{N-1} \otimes \underset{\sim}{C}_{p\,p}\underset{\sim}{\Sigma}_{p\,p}\underset{\sim}{C}')$ over $\underset{\sim}{C} \in O(p)$, one could replace the joint density of $\underset{\sim}{S} = \underset{\sim}{G}\underset{\sim}{L}^2\underset{\sim}{G}'$ and $\underset{\sim}{W}$ by the joint density of $\underset{\sim}{L}^2$, $\underset{\sim}{W}$, and an independent statistic.

The very simple case $c = 2$, $p = 1$, $n_1 = 2$, and $n_2 = 1$ can, however, be attacked from first principles for comparison with (4.17). Write

$$s^2 = \sum_{j=1}^{3} X_j^2 - 3\overline{X}^2 \geq 0.$$

4.3. Two Normal Populations Differing in Location.

Suppose

$$(X_1\ X_2\ X_3) \overset{d}{=} N_3((0\ 0\ \mu); \underset{\sim}{I}_3) \tag{4.18}$$

where $\mu \geq 0$ without loss of generality. Then

$$f(s^2, W_3) = \frac{1}{\pi}\sqrt{\frac{3}{2-3W_3^2}}\, e^{-s^2/2 - \mu^2/3} \cosh\left(\sqrt{\mu^2 s^2 W_3^2}\right) \tag{4.19}$$

while

$$f(s^2) = e^{-s^2/2 - \mu^2/3} I_o(\sqrt{\tfrac{2}{3}\mu^2 s^2}) \tag{4.20}$$

and

$$f(W_3) = \frac{2}{\pi}\sqrt{\frac{3}{2-3W_3^2}} \cdot \left[e^{-\mu^2/3} + \frac{1}{2}\sqrt{\frac{\mu^2 W_3^2}{2}}\, e^{-\frac{\mu^2}{6}(2-3W_3^2)}\, \gamma\left(\frac{1}{2}, \frac{\mu^2 W_3^2}{3}\right)\right] \tag{4.21}$$

where $I_v(z)$ is the modified Bessel Function and

$$\gamma(a, x) = \int_o^x t^{a-1} e^{-t}\, dt$$

is the incomplete gamma function. Furthermore, $f(W_1, W_2, W_3) = \{\frac{1}{2}f(W_3) \text{ or } 0\}$ as in (4.17).

Note that $\mu = 0$ reduces (4.21) to (4.17), reduces (4.20) to the χ^2 distribution with 2 degrees of freedom, and implies that S^2 and $\underset{\sim}{W}_{1\,3}$ are independent. However, $\mu > 0$ implies that S^2 has non-centrality $\frac{2\mu^2}{3}$, s^2 and $\underset{\sim}{W}_{1\,3}$ are not independent (although both are independent of $\overline{X}$), and (4.21) is even more negatively skew than (4.17) as plotted in Figure 4.1. Note that

$$\gamma\left(\frac{1}{2}, \frac{\mu^2 W_3^2}{3}\right) = \sqrt{\pi}\, \left[2\Phi\left(\sqrt{\frac{2\mu^2 W_3^2}{3}}\right) - 1\right] \tag{4.22}$$

where $\Phi(z)$ is the cumulative distribution of the standard normal distribution. Thus one can claim that (4.21) retains some direct relationship with the normal distribution when $\mu > 0$.

4.4. Two Normal Populations Differing in Dispersion.

Suppose

$$(X_1\ X_2\ X_3) \overset{d}{=} N_3\left((0\ 0\ 0), \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right). \tag{4.23}$$

Then

$$f(s^2, W_3) = \frac{2}{\pi}\sqrt{\frac{3}{2-3W_3^2}} \cdot \exp\{-\frac{s^2}{2}(1 + \frac{3(1-\sigma^2)}{1+2\sigma^2} W_3^2)\} \quad (4.24)$$

while

$$f(s^2) = e^{-s^2/2} {}_1F_1(\tfrac{1}{2}, 1, -\frac{s^2(1-\sigma^2)}{1+2\sigma^2}) \quad (4.25)$$

and

$$f(W_3) = \frac{2}{\pi}\sqrt{\frac{3}{2-3W_3^2}} \cdot \frac{1}{(1 + \frac{3(1-\sigma^2)}{1+2\sigma^2} W_3^2)} \quad (4.26)$$

where ${}_1F_1(a, b, z)$ is Kummer's form of the confluent hypergeometric function. Here $\sigma^2=1$ reduces (4.26) to (4.17) and shows statistical independence. However, if $\sigma^2 \neq 1$, s^2 and ${}_1\underset{\sim}{W}_3$ are not only dependent on each other but also dependent upon $\bar{X}$.

5. SELECTION AND EVALUATION OF VARIABLES

In order for the model for $\underset{p\ N}{\underset{\sim}{X}}$ introduced in Section 2 to be valid in a given experimental situation, the number of variables (p) to be observed, the number of individuals (N) to be examined, and the particular characteristics $(X_1, \ldots, X_p)$ to be measured should be selected in such a way that the properties of $\underset{r\ N}{\underset{\sim}{W}}$ demonstrated in Section 3 have meaningful physical interpretations. Since there will be exactly r equally important orthogonal axes of variation in the invariant representation and $r \le \min(p, N-1)$, we see that we should take $p \le N-1$ and select $(X_1, \ldots, X_p)$ such that $P(r=p) = 1$. A principal components analysis of $\underset{p\ p}{\underset{\sim}{S}}$ or of the corresponding correlation matrix might be used to get some idea of the dimensionality of the data. In view of Lemma 3.5., it is desirable to have N much larger than p. If r is observed to be less than p, there is some linear redundancy between the rows of the data matrix $\underset{p\ N}{\underset{\sim}{X}} - \underset{p\ 1}{\underline{\bar{X}}}\ \underset{1\ N}{\underline{1}'}$, and we should reduce the number of variables under consideration to only r. In particular, there should be no physical reason to believe that the chosen characteristics $(X_1, \ldots, X_p)$ are not equally important in distinguishing the individuals and populations under consideration.

Geometrical intuition can be brought to bear upon the question of how the variables $(X_1, \ldots, X_p)$ interact by considering the stepwise construction of $\underset{r\ N}{\underset{\sim}{W}}$ using Lemma 3.1. as the variables (rows of $\underset{\sim}{X}$) are added in different orders. The elements of $\underline{v}^*_{(\underline{\ell}'_k)}$ in (3.6) are interpreted as the coordinates of N points to be plotted on

$$[-\frac{N-1}{N}, +\frac{N-1}{N}] \subset R^1 .$$

The resulting one dimensional configuration displays the contribution of variable X_{ℓ_k} to the representation $\underset{\sim}{W}$ in the presence of X_{ℓ_1}, X_{ℓ_2}, ..., and $X_{\ell_{k-1}}$ only. Similarly, $\underline{v}^*_{(\underline{\ell}'_k)}$ and $\underline{v}^*_{(\underline{\ell}'_{k+1})}$ can be plotted on orthogonal axes in R^2 to represent the contribution of X_{ℓ_k} and $X_{\ell_{k+1}}$ together after adjusting for the presence of X_{ℓ_1}, ..., $X_{\ell_{k-1}}$. Note that $\underline{v}^*_{(\underline{\ell}'_k)}$ is the projection of $\underset{\sim}{W}$ onto some direction, the orientation of which is determined using only the ℓ_1-th, ℓ_2-th, ..., and ℓ_k-th rows of $\underset{\sim}{X}$. Thus $\underline{v}^*_{(\underline{\ell}'_k)}$ is clearly not invariant under transformations (1.3) although $\underset{\sim}{W}$ is invariant.

Note that if r is less than p, at least two of the vectors $\underline{v}^*_{(\underline{\ell}'_p)}$ will be null. One of the corresponding variables, X_{ℓ_p}, which makes no contribution to the representation in the presence of the $p-1$ other variables should be dropped from consideration. If r is less than $p-1$, the above calculations should be repeated until only r variables are retained. Besides these techniques based upon the theoretical model for $\underset{\sim}{X}$, there may be physical considerations to keep in mind when selecting p and $(X_1, \ldots, X_p)$.

Consider the following artificial experimental situation. A large number of solid metal balls are to be examined. Their respective diameters and weights can easily be measured and would seem to be of interest. Thus data might be collected with $p=2$. However, if the experimenter notices that all the balls are made of the same uniform metal, diameters and weights are clearly related. This redundancy would probably not cause the representation to have rank one because the relationship is nonlinear. However, except for errors in measurement, the points of the two dimensional representation would lie on a smooth curve and thus reveal the redundancy. Since the problem is really one dimensional in this case, only one variable should be observed. Weight might be considered critical if the balls are to be lifted and carried by hand; diameter might be considered critical if the balls are to be manipulated by a mechanical device which opens only to certain widths.

6. SUMMARY AND USES

The linearly invariant canonical form $\underset{r \; N}{\tilde{W}}$ is a representation of the data $\underset{p \; N}{\tilde{X}}$ adjusted for $\underline{\bar{X}}$ and $\tilde{S}$. While $\underline{\bar{X}}$ and $\tilde{S}$ describe individuals in the population(s) in terms of the specific characteristics measured, $\tilde{W}$ describes individuals in the population(s) in terms of all possible linear combinations of the characteristics measured. If it were known that $\tilde{X}$ had the Multivariate Normal distribution (4.1), $\underline{\bar{X}}$ and $\tilde{S}$ would be sufficient for $\underline{\mu}$ and $\tilde{\Sigma}$ and $\tilde{W}$ would contain no information about the population as a whole. However, $\tilde{W}$ would still describe the specific N individuals chosen for analysis. On the other hand, given some experimental data, one might assume that $\tilde{X}$ has the distribution (4.1) but discover that $\tilde{W}$ contains empirical evidence that casts doubt on this assumption. Finally, if $c > 1$ different populations are being sampled, although "average" location and dispersion have been removed, $\tilde{W}$ reveals "relative" location and dispersion of the populations.

The invariant model for $\tilde{X}$ considered here is inadequate and misleading when the variables chosen for analysis are linearly dependent but are measured subject to random experimental error. Let $r^* < p$ be the true rank of the distribution of the p-variables. Because of measurement error, the linear dependence will probabably not be discovered $(r=p)$. Furthermore, the errors in measurement will be given weight equal to $(p-r^*)$ dimensions of variation in $\underset{p \; N}{\tilde{W}}$. Thus the greatest weakness of the representation $\tilde{W}$ is with respect to the type of redundancy in the data it was hoped to automatically detect.

The invariance of $\underset{\sim}{W}$ under (1.3) implies, in particular, that the signs of all the variables can be reversed. Thus procedures appropriate for the analysis of data conforming to the model we have considered can detect differences but do not automatically indicate a direction of difference. For example, in the univariate (p=1) two sample testing problem, we are considering two-sided tests only.

The canonical form $\underset{\sim}{W}$ can be used to bring geometrical intuition to bear on problems in which linear invariance is assumed. The one and two dimensional representations discussed in Section 5 are the easiest to interpret. The results of Lemma 3.6. imply that multivariate test criteria based upon sums of characteristic roots may have an interpretation as average squared distances within and among samples. The representation can be constructed before the use of the T^2 statistic as mentioned in the introduction. This is also the case when Fisher's sample linear discriminant function is to be used if the individual to be identified and the samples known to be from the respective populations are simultaneously represented.

If one is not willing to assume that all the variables can be combined together, but k distinct subsets of the variables are each thought to conform to the invariant model, then the separate representations

$$ {}_{p_1}\underset{\sim}{W}^{[1]}_N, \ldots, {}_{p_k}\underset{\sim}{W}^{[k]}_N $$

can be combined to form a total representation

$$ {}_p\underset{\sim}{W}_N = \begin{bmatrix} \underset{\sim}{W}^{[1]} \\ \vdots \\ \underset{\sim}{W}^{[k]} \end{bmatrix} $$

where $p = \sum_{\ell=1}^{k} p_\ell$. When $k=2$, the above calculations could be performed before use of Hotelling's sample canonical correlation analysis.

The amount of calculation required to find $\underset{\sim}{W}$ increases rapidly with N. A non-singular g-inverse for the $p \times p$ matrix $\underset{\sim}{S}$ can be found by sweep-out to its Hermite row canonical form. When $\underset{\sim}{P}$ is formed using (2.3), the accuracy of the calculations is indicated by the number of digits to which $\underset{\sim}{P}$ is symmetric and idempotent. Since the non-null rows (columns) of $\underset{\sim}{P}$ are all c-vectors of $\underset{\sim}{P}$ associated with c-roots of $+1$, any orthonormalization of these vectors will produce r non-null rows for $\underset{\sim}{W}$. The c-vectors found by this or any other technique will usually not have the uniqueness of orientation property (2.11), but this is not really necessary for geometric interpretation of the representation; only the inter-point distances are of interest. Thus, writing $\underset{\sim}{S} = \underset{\sim}{G}\underset{\sim}{L}^2\underset{\sim}{G}'$ as in (4.5), we see that $\underset{p}{\underset{\sim}{L}}^{-1}\underset{p}{\underset{\sim}{G}}'(\underset{p}{\underset{\sim}{X}} - \bar{\underset{=}{X}}\underset{N}{\underset{=}{1}}')$ has properties (2.8), (2.9), and (2.10); the sample principal components normed to mean zero and variance one constitute an un-oriented decomposition of $\underset{N\ N}{\underset{\sim}{P}}$.

Many univariate nonparametric techniques based upon ranks are invariant with respect to monotonically increasing transformations of the data. These transformations include translation and change of scale. Multivariate, conditionally nonparametric procedures have been considered which rank data on each variable separately. In a future communication, the author will discuss the use of the linearly invariant canonical form in the assignment of "rank scores". New nonparametric and conditionally nonparametric, univariate and multivariate ranking procedures will be introduced which are invariant only with respect to translation and nonsingular linear transformation.

REFERENCES

[1] Gower, J.C. (1966). "Some distance properties of latent root and vector methods used in multivariate analysis". Biometrika 53 325-338.

[2] James, A.T. (1954). "Normal multivariate analysis and the orthogonal group". Ann. Math. Statist. 25 40-75.

[3] James, A.T. (1955). "The noncentral Wishart distribution". Proc. Roy. Soc. London Ser. A. 229 364-366.

[4] Rao, C.R. (1966). "Generalized inverse for matrices and its applications in mathematical statistics." Research Papers in Statistics (Festschrift J. Neyman, Edited by F. N. David), Wiley, New York, 263-279.