BIOMETRICS 28, 3–22
March 1972

RECENT TRENDS OF RESEARCH WORK IN MULTIVARIATE ANALYSIS

C. RADHAKRISHNA RAO

Indian Statistical Institute, Parliament Street, New Delhi-1, India

1. INTRODUCTION

In the issue of *Biometrics* (Vol. 20, part 2) dedicated to the memory of Ronald Aylmer Fisher I reviewed the contributions by Fisher and some of the salient features of research work done in multivariate analysis up to 1964. Fisher's contributions as well as the related methodology developed by Wilks ([1932; 1946]—likelihood ratio criteria for testing multivariate hypotheses), Bartlett ([1947; 1951]—decomposition of Wilks's criteria for testing different aspects of a null hypothesis), Rao ([1946; 1948]—analysis of dispersion[1] as a generalization of the univariate analysis of variance, and tests for additional information supplied by a subset or functions of measurements), and Williams ([1952a; 1959]—residual canonical correlations in testing goodness of fit of specified discriminants) may be described as a study of association between two sets of variables. One set is called predictor and another set, criterion variables. Some of the variables may be hypothetical (unobservable), and others may have values on a dummy or interval scale. I have indicated in Rao [1960] how various multivariate methods such as regression, canonical correlations, analysis of dispersion and canonical analysis, discriminant function, factor and latent structure analyses, treatment of contingency tables etc., can be classified by the nature of predictor and criterion variables. However, the classification provides only a suitable framework for the discussion of different problems but does not imply that the statistical methods appropriate for one problem can be simply deduced from those of another.

Since 1964 considerable progress has been made in several directions of the work initiated by Fisher. Exact distributions have been found of several likelihood ratio criteria and of roots of determinantal equations involving random matrices which arise in multivariate statistical analysis and in some problems of physics. Further progress has been made in testing goodness of fit of assigned discriminant functions. The theory and construction of discriminant functions for deciding between composite hypotheses (involving nuisance parameters) has been developed. A satisfactory

[1] The term Analysis of Dispersion (AD) to denote analysis of variance and covariance was chosen in consultation with Fisher and is the forerunner of MANOVA. The word dispersion implies scatter in all directions.

approach has been made available for statistical analysis of growth data and prediction of growth.

Some new lines of research work, not directly related to Fisherian concepts and contributions and free from multivariate normality assumption, have been established during the last ten years. One is the problem of 'Seriation' in archaeology first considered by Flinders Petrie [1899] seventy years ago, viz., the determination of sequences of prehistoric events by paired comparison of characteristics associated with them—events closer in time would presumably have a larger number of common characteristics. Another is 'Multidimensional Scaling' which seeks to determine the configuration of a set of points using information about interpoint distances—the information may be of a meagre type giving only 'greater or less than relationship' among the distances. A systematic theory of cluster analysis is being developed and new applications made. The last decade also saw a rapid development of nonparametric and graphical methods in the analysis of multivariate data.

A new *Journal of Multivariate Analysis* has been started to meet the increasing demands of research workers and to provide a forum for discussion of current problems.

The object of the present survey is to discuss some of the new multivariate methods which seem to be of immediate value to practical workers in a wide variety of fields, and to review briefly some of the more theoretical developments. A recent paper by Dempster [1971], 'An overview of multivariate analysis,' covers similar ground. There has been considerable work on time series and inference problems on stochastic processes which are somewhat specialized and which are not considered in this article.

Computers have revolutionised research in science (Rao [1970]); more so in statistics, especially in the area of multivariate analysis where most of the new techniques are computer-tied. There seems to be less emphasis on research on choice of variables and refinements of techniques which received considerable attention in the days of desk calculators. What is lost by the use of an inefficient technique is sought to be made up by increasing the number of measurements, often employing the computer itself. A greater danger to scientific research appears to be the much publicised package programs for multivariate methods which have misled and are still likely to mislead applied workers.

2. DISCRIMINANT ANALYSIS

2.1. *Applications*

The discriminant function was suggested by Fisher for identifying an observed specimen as belonging to one of two specified populations. In the previous review (Rao [1964]) I have discussed different aspects of the discriminant function when the measurements are continuous—examining the adequacy of a given function, testing for given ratios of the coefficients of some variables in a discriminant function, providing for the possibility of an observed specimen belonging to a third unknown group (see Rao [1965a] chapter 8). The practical applications of the discriminant function have not been many due to inadequacy of information about the distributions of the measurements and prior probabilities of the alternative populations. However, some important applications have been made in pattern recognition by Gnanadesikan and his collaborators at the Bell Telephone Laboratories and in medical diagnosis (e.g. Truett, Cornfield, and Kannel [1967]).

There are additional difficulties due to some of the measurements being binary in nature. A recent paper by Cox [1971] summarises the work of Bahadur, Lazarsfeld, and others in this direction and recommends the use of a logistic model for the estimation of probabilities.

2.2. *Discrimination between composite hypotheses*

A new line of research in discriminant analysis was started by a problem posed by Burnaby [1966], where a fossil had to be identified as belonging to one of two populations—each population being a composite one consisting of individuals belonging to one species but presumably fossilized at different ages. Given the measurements on a specimen of an unspecified age, how does one identify it as belonging to one of two species? The problem is that of constructing a discriminant function which does not distinguish between individuals of the same species with different ages but is sensitive to differences between species for any given age. The theory of estimation and the construction of a discriminant function in such cases are discussed by Rao [1966].

2.3. *Direction and collinearity tests*

In discriminant analysis, the problems of examining the dimensionality of the configuration of the mean characteristics of the various groups and of testing the adequacy of assigned discriminant functions are of some importance.

Fisher (see also Bartlett [1938]) developed a test for examining whether an assigned discriminant function is adequate for discriminating between two alternative populations. Rao [1946; 1948] extended this test to examine whether a given subset (or functions) of measurements is adequate to explain differences between means of correlated variables in a given population and differences in means of variables between two or more populations. In the case of several populations, if Λ stands for Wilks's criterion for judging differences between k groups in all p variables and Λ_1 for the s assigned discriminant functions, then the appropriate Wilks criterion Λ' for testing their adequacy for discrimination is obtained from the formula due to Bartlett

$$\Lambda'\Lambda_1 = \Lambda. \tag{2.3.1}$$

Illustrations of tests based on Λ' are given in Bartlett [1947; 1951] and Rao [1948].

The decomposition (2.3.1) is the starting point of Williams' work which is admirably reviewed in his 1967 paper. Williams [1952a] observed that Λ' contains two types of deviations, one due to the number ('dimensionality') of specified discriminants being wrong and another due to the specification of the functions ('direction') being wrong although their number is adequate. He developed exact tests for this purpose when $p = 2$ and $s = 1$ for any k using the concept of residual canonical correlations. Bartlett [1951] obtained the corresponding tests for general p and $s = 1$, by considering two kinds of decompositions of Λ', which was later extended by Williams [1961] for $s > 1$. Kshirsagar [1964] gave an analysis of variance type break-up, expressed these factors of Λ' in simpler forms, and gave an analytical derivation of their distributions.

Williams [1959] provided a similar analysis in a more general situation for testing the adequacy of a specified linear structure for mean values of different populations, as in examining whether 'time' alone can be a good discriminator from Barnard's data on Egyptian skulls. The overall criterion in Barnard's problem was given earlier by Rao [1948], which differed somewhat from that given by Bartlett [1947]. Kshirsagar [1962; 1971] provided an appropriate theory for Williams' type of analysis.

Williams [1952b] and Bartlett [1951] also considered similar tests for examining association between attributes in contingency tables.

The goodness-of-fit test of a single discriminant function was carried into the area of principal component analysis by Kshirsagar [1961].

3. EXACT DISTRIBUTION OF MULTIVARIATE TEST CRITERIA

A number of likelihood ratio criteria for testing various multivariate hypotheses have been introduced following the early work of Wilks [1932]. One may refer to books by Anderson [1958], Dempster [1969], and Rao [1952; 1965a]. Till recently only asymptotic expansions of the distributions of these statistics and fairly good approximations in terms of χ^2 and F distributions (Bartlett [1938; 1947], Rao [1948], Radcliffe [1966]) were known, except in some special cases. During the last 10 years considerable progress has been made in obtaining exact distributions in the null and non-null cases and also in obtaining exact percentage points by suitable computer programs. Similar advances took place in the distribution of roots of determinantal equations introduced by Fisher to infer on the dimensionality of the configuration of true mean values of given set of populations.

The principal contributors to exact distributions of test criteria are James [1964] and Constantine [1963] who achieved a breakthrough in deriving exact non-null distributions with the help of zonal polynomials, Mathai [1970], by himself and in collaboration with Rathie and Saxena who used Meijer's G-function and H-function, Krishnaiah and his collaborators, Chang and Waikar (Krishnaiah and Chang [1970; 1971], Krishnaiah and Waikar [1971]) who used inverse Laplace transforms and methods developed by Wigner [1967] and Mehta [1967], Pillai and his collaborators (Khatri and Pillai [1968]; Pillai, Al-Ani Sabri, and Jouris [1969]), Khatri [1967], Khatri and Srivastava [1971] and others. (Mathai [1970] lists a large number of references on the subject.)

Exact distributions in special cases were, however, obtained earlier by a number of authors including Bagai [1965], Consul [1969], Davis [1970], Sugiyama [1970], to give a few references.

The distribution of eigenvalues of a random matrix has also received the attention of physicists (see, e.g., Wigner [1967], Mehta [1967], Porter [1965]). The random matrix they consider is not, however, the Wishart type but has as its (i, j) element an independent normal variable with zero mean and variance equal to 2 when $i = j$ and equal to 1 when $i \neq j$.

4. ANALYSIS OF GROWTH DATA

4.1. *Early work of Wishart*

The forerunner of all research dealing with the statistical analysis of growth data is the classical work of Wishart [1938]. As a first step in the analysis, Wishart fitted orthogonal polynomials to individual growth data and replaced the large number of observations on each individual by a few of the fitted coefficients of the first, second, . . . , degree terms. These are used in subsequent analysis for comparison of different treatments etc. The general approach in recent applications has been essentially the same but several improvements have been made. For our discussion we shall consider different situations.

4.2. *Analysis of comparative experiments*

It has been pointed out that Wishart's approach can be made more efficient by (a) transforming the response variables (such as *log* for weight as in Rao and Rao [1966]) and (b) a suitable choice of time metameter with respect to which the average growth curve assumes a simpler form. Rao [1958] gave a method of constructing such a transformation of the time variable using the data *under analysis* itself and showed that the Wishart type of analysis remains *valid* when polynomials in terms of estimated time metameter are fitted. It was shown in an example that comparison between treatments could be essentially reduced to examining differences in linear growth rate with respect to transformed time, whereas higher order terms were needed otherwise.

4.3. *Estimation of average growth curve over a short time period*

When the average curve is of the polynomial type, Rao discussed methods for: (a) testing the adequacy of a polynomial of a given degree (Rao [1959]), (b) estimating a polynomial of a given degree using suitable functions of observations for covariance adjustment (Rao [1965b; 1967]), and (c) providing a confidence band for the average growth curve (Rao [1959]).

4.4. *A multivariate growth model*

The usual generalization of the Gauss–Markoff model to the multivariate case is of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0} \tag{4.4.1}$$

where $\mathbf{Y}$ is an $n \times p$ matrix of random variables, $\mathbf{X}$ is a given $n \times m$ matrix of coefficients and $\boldsymbol{\beta}$ is an $m \times p$ matrix of unknown parameters. The rows of the random variable $\mathbf{Y}$ are independently distributed while the components in each row may be correlated. If the column vectors of $\boldsymbol{\beta}$ have no restrictions, then we have the usual extension of analysis of variance to analysis of dispersion (Rao [1948], [1965a] p. 459).

As a generalization of the univariate model with concomitant variables, Rao [1965b] considered the multivariate model

$$\underset{n\times p}{E(\mathbf{Y})} = \underset{n\times m}{\mathbf{X}} \; \underset{m\times p}{\boldsymbol{\beta}}, \; \underset{n\times s}{E(\mathbf{Z})} = \mathbf{0} \tag{4.4.2}$$

such that

$$E(\mathbf{Y} \mid \mathbf{Z}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varrho} = (\mathbf{X} \vdots \mathbf{Z})\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\varrho}\end{pmatrix}. \tag{4.4.3}$$

In terms of conditional expectation, the model (4.4.3) is of the same form as (4.4.1) with unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\varrho}$ so that no new problems arise in drawing inferences on $\boldsymbol{\beta}$. However, the appropriate multivariate technique for tests of hypotheses on $\boldsymbol{\beta}$ may be described as analysis of dispersion with adjustment for concomitant variables $\mathbf{Z}$ (see Rao [1965b]).

If the columns of $\boldsymbol{\beta}$ in (4.4.1) are related in such a way that $\boldsymbol{\beta}$ can be written as $\boldsymbol{\beta} = \boldsymbol{\xi}\mathbf{A}$, where $\boldsymbol{\xi}$ is the new $m \times k$ matrix of unknown parameters and $\mathbf{A}$ is a given $k \times p$ matrix, then the model (4.4.1) becomes

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\xi}\mathbf{A} \tag{4.4.4}$$

as considered by Potthoff and Roy [1964]. Let $\mathbf{H} = (\mathbf{H}_1 \vdots \mathbf{H}_2)$ be a $p \times p$ nonsingular matrix such that $\mathbf{A}\mathbf{H}_2 = \mathbf{0}$ and the columns of $\mathbf{H}_1$ form a basis of the vector space generated by the rows of $\mathbf{A}$. Then multiplying both sides of (4.4.4) by $\mathbf{H}$ we have

$$E(\mathbf{Y}_1 = \mathbf{Y}\mathbf{H}_1) = \mathbf{X}\boldsymbol{\xi}\mathbf{A}\mathbf{H}_1 = \mathbf{X}\boldsymbol{\eta}, \quad E(\mathbf{Y}_2 = \mathbf{Y}\mathbf{H}_2) = \mathbf{0}, \tag{4.4.5}$$

introducing the new parameter matrix $\boldsymbol{\eta}$ without any restrictions on its columns. The model (4.4.5) is then of the type (4.4.2) with $\mathbf{Y}_2$ as concomitant variables. We consider the conditional model

$$E(\mathbf{Y}_1 \mid \mathbf{Y}_2) = \mathbf{X}\boldsymbol{\eta} + \mathbf{Y}_2\boldsymbol{\varrho} \tag{4.4.6}$$

which is of the same type as (4.4.3). Thus the model (4.4.4) can be reduced to a model appropriate for analysis of dispersion with adjustment for concomitant variation, and no new problem arises. However, Khatri [1966] tried to derive test criteria, etc., afresh by using the likelihood principle starting from the multivariate normal density function of $\mathbf{Y}$.

When the rank of $\mathbf{A}$ is p we may choose $\mathbf{H}_1 = \mathbf{G}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{G}^{-1}\mathbf{A}')^{-1}$, where $\mathbf{G}$ is any p.d. matrix, in which case

$$E(\mathbf{Y}_1) = \mathbf{X}\boldsymbol{\xi}, \quad E(\mathbf{Y}_2) = \mathbf{0} \tag{4.4.7}$$

Potthoff and Roy [1964] suggested using only the first part of the model $E(\mathbf{Y}_1) = \mathbf{X}\xi$ ignoring $\mathbf{Y}_2$, while Rao [1959] suggested the full use of all the concomitants. Later Rao [1965b; 1967] discussed the possibility of using only some of the concomitants to achieve maximum possible efficiency in estimation. Rao [1967] also showed that the concomitant variable $\mathbf{Y}_2$ does not provide any information on π in (4.4.7) when the dispersion matrix of the variables in any row of $\mathbf{Y}$ is of the form

$$\mathbf{H}_1\mathbf{G}_1\mathbf{H}_1' + \mathbf{H}_2\mathbf{G}_2\mathbf{H}_2' + \sigma^2\mathbf{I}, \tag{4.4.8}$$

where $\mathbf{G}_1$, $\mathbf{G}_2$ and the scalar σ^2 are arbitrary. Some applications of these techniques are given by Grizzle and Allen [1969].

4.5. *Prediction of individual growth*

Knowing the weights of a growing child at some time points in the past, how do we predict the weight at a future time point? This problem has been approached from a Bayesian point of view by Geisser [1971] using the information supplied by complete records of weights on a sample of children observed over the entire growth period. The problem raises some fundamental issues. What aspects of the information provided by complete records of growth observed on some children would be of use in predicting the future growth of a new child from his past observations? A simple-minded regression formula for future weight on previous measurements constructed from complete records does not seem to provide a satisfactory approach. Perhaps a study of an individual's growth process from complete records would provide a better basis for prediction. Further research in this direction is needed.

4.6. *Estimation of age-specific norms from survey data*

Rao and Rao [1966] introduced what is called a Linked Cross Sectional (LCS) approach in collecting data for estimating age-specific norms, growth rates, differential growth rates etc., over a given period of growth. The study consists in taking a sample of individuals at age t_1 and observing them over a specified number s of years, another sample at age $t_2 < s + t_1$ and observing them over s years, etc. The survey which can be completed in s years' time provides estimates of average growth curves in the overlaping time intervals $(t_1, t_1 + s)$, $(t_2, t_2 + s)$, $\cdots$, $(t_k, t_k + s)$. The different segments are then pieced together to obtain the average growth curve over the entire period $(t_1, t_k + s)$. The value of s is chosen to be small compared to $t_k - t_1$. The LCS approach is likely to be of use in many other situations.

5. ARCHAEOLOGICAL SERIATION

This problem was formulated 70 years ago by the archaeologist, Flinders Petrie [1899], and the interest in the problem was revived by D. G. Kendall [1963] who built up an active school of research in this area. There is already considerable literature on the subject due to Kendall, Sibson, Wilkinson, Hole, Shaw, Kaluscha, and others; much of this is contained in and probably all is referred to in the Mamaia Proceedings volume edited by Hodson, Kendall, and Tautu [1971].

Petrie was confronted with some 900 pre-dynastic Egyptian graves containing representatives of about 800 varieties of pottery and the problem was to infer from the 'incidence matrix' of 'graves-versus-varieties of pottery' (the (i, j) element is 1 if i-th grave contained j-th variety of pottery and 0 otherwise) the sequential ordering of the graves in time and to determine the ordinal intervals during which the different varieties of pottery flourished. The basic Petrie principle, without which seriation is hardly possible is as follows: *two graves are the more likely to contain varieties of a similar type the closer together they are in the true time order*. Under this principle it should be possible, by a suitable row-rearrangement, to exhibit the incidence matrix **A** such that all 1's are strung together in each column. Such a matrix is called a *Petrie matrix*. The problem then is one of *petrifying* **A**, i.e. of converting **A** into the Petrie form by a rearrangement of the rows. The sequential ordering of the graves is provided by the order of the corresponding rows in the transformed matrix. In practice such a rearrangement may not be strictly possible. In an early study, Kendall [1963] tried to find a rearrangement of rows by a suitable search method which minimises $\sum n_j \log r_j$, where j runs through various types, n_j is the total number of representatives of the j-th type, and r_j is the 'range' of that type when the graves are given the tested reordering. Later, Kendall [1970] suggested the computation of a similarity matrix $\mathbf{S} = (S_{ij})$ for the graves and an application of multidimensional scaling technique to provide an ordering of the graves over time. The matrix **S** is chosen to be **AA'**, when **A** is the incidence matrix, and $S_{ij} = \sum_k \min (a_{ik}, a_{jk})$ when $\mathbf{A} = (a_{ij})$ is the abundance matrix.

Fulkerson and Gross [1965] examine the conditions under which a matrix **A** is petrifiable. They provide a graph-theoretic algorithm to identify row-permutations of Petrie matrices using the matrix $\mathbf{V} = \mathbf{A'A}$. Kendall [1969] has shown that, when **V** is such that **A** is petrifiable the row-permutations which petrify **A** are exactly those which, when applied to the rows and columns of **S** simultaneously, give to that square matrix the *Robinson form*. Further results in this direction and generalizations are found in Wilkinson [1971].

A square matrix is said to be in Robinson form when its components never decrease as one progresses along a row towards the main diagonal and never increase as one continues to progress along that row beyond the main diagonal. Such matrices play a prominent role in a geological problem of seriation considered by Robinson [1951]. Instead of the incidence matrix he had a matrix **B** giving in its (i, j) component the percentage composition of the i-th deposit which was attributed to the 'object' j. From **B**, a similarity matrix was constructed, and then Robinson's idea was simultaneously to permute the rows and columns of the similarity matrix until it assumed the Robinson form (as nearly as could be managed).

The seriation methods have been used successfully in other areas, classical philology (chronological ordering of written works of Plato), reconstruction of maps by using indices of similarity between places, etc.

6. MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS) is a technique of data analysis in which a configuration of points is determined using information about the interpoint distances. The information about the interpoint distances may take many forms: direct estimates; rank order information; pairwise comparisons; and others. Naturally, random error is assumed present regardless of form.

The term 'multidimensional scaling' was invented by Torgerson in 1952 (see also [1958]). The current surge of interest was largely initiated by Shepard [1962a, b] whose ideas were further refined and developed by Kruskal [1964a, b]. Parallel developments of theory and computer programs are due to Guttman and Lingoes, and to Young. Subsequent work was done by Roskam, McGee, and others. References to papers by these authors and others mentioned in this section can be found in a recent paper by Carrol [1971]. MDS as developed by Shepard and Kruskal has been applied in various fields, study of origin of languages, seriation in archaeology, ordering of an author's works in time, construction of geographical maps from odd bits of information etc. (see the papers in the Mamai proceedings edited by Hodson *et al.* [1971]).

A very striking advance in scaling has occurred recently with the development of 'individual differences scaling' by Carrol and Chang (and independently by Harshman). This method requires several dyadic matrices (pertaining to the same objects) as input. It yields a positive diagonal 'weight' matrix for each input matrix, and a single common configuration. For each dyadic input matrix, the entries correspond to the interpoint distances in a modification of the common configuration. The modification is formed by rescaling the common configuration, using the appropriate weight matrix. This technique of data analysis turns out to be very powerful, as has been demonstrated by many applications.

Meanwhile, ordinary MDS is just starting to undergo development from a data analysis technique into a statistical technique. Several distributional studies have been published (so far, all using Monte Carlo methods) pertaining to significance levels, estimation of dimensionality and standard deviation of error, and the accuracy with which the true underlying configuration has been recovered. The main publications so far are by Klahr, Wagenaar and Padenos, Young, and Stenson and Knoll.

Another advance is the method of scaling invented by Roskam for data of triadic comparisons type. It illustrates concretely an idea that any reasonable form of data can be utilized directly for scaling (if the analysis technique is suitably modified), and that such direct use may well provide advantages over pre-processing the data to fit existing methods of scaling. In this case, one advantage is greater accuracy in estimating the true dimensionality, which results from partitioning the squared stress (just as variance is partitioned in analysis of variance).

An advance in a different direction is a method of looking for clusters (and other structures) in configurations in many dimensions, which was invented by Sammon (and also by Thompson and Woodbury). The general idea is to use the distances in many dimensions as input to scaling in two dimensions. This provides a nonlinear mapping into two dimensions, which can be visually examined, just like the first two principal components.

An imaginative new use of scaling utilizes a pre-existing configuration (due to Henley), based on direct human similarity judgments of 30 animals (tiger, mouse, elephant, etc.). Rumelhart asked subjects to supply the best possible answer to questions of the form: '*Elephant* is to *tiger*, as *goat* is to what?' He found that by forming a parallelogram with three corners at the given animals, and calculating how far each animal is from the fourth corner, he was able to predict the answer frequencies quite well.

Meanwhile, a substantial quantity of more routine applications continue to be made, particularly in psychology and in marketing. A textbook by Green and Carmone [1970] has recently been published, and another by Green and Rao is in press.

While MDS is likely to figure for some time as a useful research tool in diverse fields of application, there is also a potential danger of indiscriminate use leading to wrong and/or over interpretation of data as has happened with the technique of factor analysis when it was first introduced.

7. CHARACTERIZATION OF THE MULTIVARIATE NORMAL DISTRIBUTION

There is a good deal of literature about the characterizations of a univariate normal variable but the corresponding results in the multivariate case have not been fully worked out.

A theorem due to Darmois–Skitovich asserts that if $X_1, \cdots, X_n$ are independent one-dimensional variables, then the independence of two linear functions

$$Y_1 = a_1X_1 + \cdots + a_nX_n, \quad Y_2 = b_1X_1 + \cdots + b_nX_n, \tag{7.1}$$

where $a_ib_i \neq 0$ for any i, implies that each X_i is univariate normal. Ghurye and Olkin [1962] proved the corresponding result when X_i is a p-vector variable.

When all X_i are identically distributed, Linnik [1953] characterized the common law of X_i under the condition that Y_1 and Y_2 in (7.1) have identical distributions, but the corresponding result in the multivariate case is not known. Ramachandran and Rao [1968] have characterized the law of X_i when $E(Y_1 \mid Y_2) = 0$. The extension of this result to the multivariate case has been made only in some special cases.

Recently there has been some interest in multivariate exponential type distribution with the density function

$$f(\mathbf{x}, \boldsymbol{\theta}) = h(\mathbf{x}) \exp [\mathbf{x}'\boldsymbol{\theta} - q(\boldsymbol{\theta})] \tag{7.2}$$

where $\mathbf{x}$ is a p-vector random variable and $\boldsymbol{\theta}$ is a p-vector parameter (Dempster [1971]). Bildikar and Patil [1968] examined the conditions under which (7.2) reduces to a multivariate normal distribution.

A p-vector variable $\mathbf{X}$ is said to have a linear structure if $\mathbf{X} = \mathbf{A}\boldsymbol{\xi}$, where $\mathbf{A}$ is a $p \times m$ matrix of structural coefficients and $\boldsymbol{\xi}$ is an m-vector of independent hypothetical (structural) variables. Two structural representations $\mathbf{X} = \mathbf{A}\boldsymbol{\xi}$ and $\mathbf{X} = \mathbf{B}\boldsymbol{\eta}$ are said to be equivalent if each column of $\mathbf{A}$ is a multiple of some column of $\mathbf{B}$ and vice versa. $\mathbf{X}$ is said to have a unique structure if all alternative representations are equivalent. Rao (see Rao [1969] and other references listed therein) examined the consequences of $\mathbf{X}$ admitting two alternative structures $\mathbf{X} = \mathbf{A}\boldsymbol{\xi}$ and $\mathbf{X} = \mathbf{B}\boldsymbol{\eta}$ and proved the following results:

(i) If the i-th column of $\mathbf{A}$ is not a multiple of any column of $\mathbf{B}$, then ξ_i, the i-th component of $\boldsymbol{\xi}$, is univariate normal.

(ii) If ξ_i is non-normal, then the i-th column of $\mathbf{A}$ must be a multiple of some column of $\mathbf{B}$.

(iii) If no column of $\mathbf{A}$ is a multiple of any column of $\mathbf{B}$, then $\mathbf{X}$ is p-variate normal, i.e., a multivariate normal variable has an arbitrary structure.

(iv) $\mathbf{X}$ has a unique structure if no linear combination of the hypothetical variables $\boldsymbol{\xi}$ has a normal component.

(v) $\mathbf{X}$ can be written as $\mathbf{X}_1 + \mathbf{X}_2$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, $\mathbf{X}_1$ has a unique structure and is therefore non-normal, and $\mathbf{X}_2$ is multivariate normal. However, the decomposition may not be unique.

These results generalize the earlier work of Reiersøl [1950] and are relevant in the discussion of structural models used by psychologists, economists, geneticists, and so on. Related statistical problems on the estimation of structural and functional relationships are discussed in a review paper by Moran [1971].

Rao [1971c] investigated the extent to which the structural variable $\boldsymbol{\xi}$ is identifiable knowing the distribution of $\mathbf{X} = \mathbf{A}\boldsymbol{\xi}$ for given $\mathbf{A}$. A surprising result is that for a suitable choice of $\mathbf{A}$, the joint distribution of p linear functions of $\frac{1}{2}p(p+1)$ independent variables determines the distributions of each of the $\frac{1}{2}p(p+1)$ variables apart from a change of a location.

In a recent paper Khatri and Rao [1971] solved a general functional equation in vector variables and obtained several characterizations of the multivariate normal distribution generalizing all earlier work on the subject.

8. ADVANCES IN OTHER AREAS

8.1. *Nonparametric methods*

A recent book by Puri and Sen [1971] contains most of the generalizations of univariate nonparametric methods to the multivariate case. This book together with *Techniques in Nonparametric Statistical Inference* edited by Puri [1970] cover practically all important work done in this area. There is considerable literature on the subject and all the references can be found in the two volumes mentioned above.

8.2. *Incomplete multiresponse data*

Most of the multivariate techniques are developed for situations in which all the responses or characteristics under study are measured on each sample unit. However, in a large number of cases it is physically impossible, uneconomic, or inadvisable on account of unequal importance, to measure all the responses on each unit. There is very little work done on estimation of parameters and testing of hypotheses when data are incomplete either by design or due to missing observations. Lord [1955], Trawinski and Bargmann [1964], Afifi and Elashoff [1966; 1967; 1969] considered some inference problems in such cases. Rao [1956] discussed an analysis of dispersion test when the observations are incomplete only on one of the characteristics. Srivastava [1966; 1968] examined problems of design for collection of data and their analysis when it is desired to measure only a subset of the characteristics on each unit. An early work on estimation of parameters when data are incomplete is due to Matthai [1951].

The Linked Cross Sectional (LCS) study of Rao and Rao [1966] described in section 5 of this article is an example of a multiresponse design with incomplete observations on units.

Sometimes a unit may be such that it does not naturally admit all measurements. If we are studying a population of ancient skulls which are in a fragmentary condition, different types of fragments admit different sets of measurements; however, two types of fragments may admit some common measurements. In such a case it was pointed out by Rao ([1952] p. 111) that the conditional distribution of a particular measurement given the type of fragment may depend on the latter. Thus the dimensions of a well-preserved skull may be smaller than those that are broken and admit fewer measurements. If so, estimation of mean characteristics of a population from a sample which may consist of some broken and some well-preserved skulls poses a difficult problem.

8.3. *Complex multivariate normal distribution*

A p-vector complex random variable $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ is said to have a complex multivariate normal distribution, $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \boldsymbol{\mu}_1 + i\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ is hermitian positive definite, if the joint density of $\mathbf{x}$ and $\mathbf{y}$ is of the form

$$\pi^{-p} |\boldsymbol{\Sigma}|^{-1} \exp [-\mathrm{tr}\, \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})(\overline{\mathbf{z} - \boldsymbol{\mu}})'], \qquad (8.3.1)$$

which is the same as the density of a $2p$-variate normal variable $(\mathbf{x}' \vdots \mathbf{y}')$ with a covariance structure of the form

$$\begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_2 \\ \boldsymbol{\Sigma}_2' & \boldsymbol{\Sigma}_1 \end{bmatrix}, \qquad (8.3.2)$$

where $\boldsymbol{\Sigma}_1 = D(\mathbf{x}) = D(\mathbf{y})$ and $\mathrm{cov}\,(\mathbf{x}, \mathbf{y}) = \boldsymbol{\Sigma}_2 = [\mathrm{cov}\,(\mathbf{y}, \mathbf{x})]'$ is skew symmetric.

From independent samples drawn from populations with densities of the form (8.3.1) one can construct statistics similar to those in the real case, such as a Wishart matrix, roots of determinantal equations, etc., and derive their sampling distributions by following exactly similar methods. Such programs have been carried out by a number of authors (Wooding [1956], Goodman [1963], Khatri [1970], James [1964], and Pillai and Jouris [1971], to mention a few names).

8.4. *Singular multivariate normal distribution*

When the dispersion matrix $\mathbf{\Sigma}$ of a p-vector normal variable $\mathbf{X}$ is singular, the density function of $\mathbf{X}$ with respect to Lebesgue measure in R^p does not exist. In such a case the density-free approach followed in Rao ([1965a] chapter 8) might be useful. However, when $\mathbf{\Sigma}$ is singular, the vector $\mathbf{X}$ is confined to a hyperplane in R^p and the density of $\mathbf{X}$ on such a hyperplane can be expressed in the form

$$(2\pi)^{-\frac{1}{2}r} (\lambda_1 \cdots \lambda_r)^{-\frac{1}{2}} \exp [-\tfrac{1}{2}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{\Sigma}^-(\mathbf{X} - \boldsymbol{\mu})], \qquad (8.4.1)$$

where $r = R(\mathbf{\Sigma})$, $\lambda_1, \cdots, \lambda_r$ are the non-zero eigenvalues of $\mathbf{\Sigma}$, $\mathbf{\Sigma}^-$ is any g-inverse of $\mathbf{\Sigma}$, and $\boldsymbol{\mu}$ is location parameter (see Rao and Mitra [1971a] for definition of g-inverse). The hyperplane on which the density is defined is $\mathbf{N}'\mathbf{X} = \mathbf{N}'\boldsymbol{\mu}$, where $\mathbf{N}$ is a matrix of rank $p - r$ such that $\mathbf{N}'\mathbf{\Sigma} = \mathbf{0}$. Khatri [1968] obtained maximum likelihood estimates of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, based on the density function (8.4.1) and also studied some distribution problems. Rao and Mitra [1971a, b] used the density function (8.4.1) in constructing a discriminant function between two alternative normal populations with singular dispersion matrices. It would be of interest to explore further uses of the density function (8.4.1).

8.5. *Variance and covariance components*

Consider a linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\boldsymbol{\xi}_1 + \cdots + \mathbf{U}_k\boldsymbol{\xi}_k + \mathbf{W}_1\boldsymbol{\eta}_1 + \cdots + \mathbf{W}_r\boldsymbol{\eta}_r, \qquad (8.5.1)$$

where $\mathbf{X}$, $\mathbf{U}_i$, $\mathbf{W}_j$ are known matrices, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{\xi}_i$, $\boldsymbol{\eta}_j$ are vectors of hypothetical variables such that

$$D(\boldsymbol{\xi}_i) = \mathbf{\Sigma} \qquad i = 1, \cdots, k,$$

$$D(\boldsymbol{\eta}_j) = \sigma_j^2\mathbf{I}_j \qquad j = 1, \cdots, r,$$

$$\operatorname{cov}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j) = \mathbf{0}, \operatorname{cov}(\boldsymbol{\xi}_i, \boldsymbol{\eta}_j) = \mathbf{0}, \operatorname{cov}(\boldsymbol{\eta}_i, \boldsymbol{\eta}_j) = \mathbf{0}, \qquad (8.5.2)$$

where $\mathbf{\Sigma}$ and σ_j^2 are unknown. In such a case

$$D(\mathbf{Y}) = \mathbf{U}_1\mathbf{\Sigma}\mathbf{U}_1' + \cdots + \mathbf{U}_k\mathbf{\Sigma}\mathbf{U}_k' + \sigma_1^2\mathbf{W}_1\mathbf{W}_1' + \cdots + \sigma_r^2\mathbf{W}_r\mathbf{W}_r'. \qquad (8.5.3)$$

There has been no satisfactory method of simultaneously estimating $\boldsymbol{\beta}$, $\mathbf{\Sigma}$, and $\sigma_1^2, \cdots, \sigma_r^2$ in the general case. Rao [1971a, b] has developed a method called MINQUE (Minimum Norm Quadratic Unbiased Estimation) for the estimation of the variance and covariance components (σ_i^2 and $\mathbf{\Sigma}$) and suggested the use of estimated $D(\mathbf{Y})$ to obtain the least squares estimate of $\boldsymbol{\beta}$. J. N. K. Rao and Subrahmaniam [1971] applied such a method with MINQUE estimates of σ_i^2 in the regression problem with heteroscedastic errors and established some properties of regression estimators. Hartley [1971] used a different approach to the problem and the relative merits of J. N. K. Rao's and Hartley's estimators have not yet been assessed.

8.6. *Multivariate discrete models*

A Dictionary and Bibliography of Discrete Distributions by Patil and Joshi (1968) contains a description of a number of multivariate discrete distributions and references to literature on the subject. Most of them are exhibited as sampling distributions of functions of observations from univariate discrete distributions or as formal mathematical extensions of univariate distributions.

8.7. *Graphical techniques in multivariate analysis*

Gnanadesikan and Wilk [1969] mention that 'Man is a geometrical animal and seems to need and want pictures for parsimony and to stimulate insight.' Indeed graphical devices are extremely useful in understanding the nature of data, in detecting unanticipated peculiarities, in the choice of models for statistical analysis, and in the presentation of final results. In a series of papers Wilk and Gnanadesikan (see their 1969 paper for other references) and Gnanadesikan and Lee [1970] have developed systematic graphical aids in the analysis of multivariate data.

8.8. *Least squares theory with a possibly singular dispersion matrix*

Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ be the Gauss–Markoff linear model with $D(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, where $\mathbf{V}$ is possibly singular (p.s.d.). Rao and Mitra [1971b] have given a unified theory of estimation and tests of hypotheses as follows.

(i) Whether $\mathbf{V}$ *is singular or not*, obtain $\boldsymbol{\beta}$ which minimises

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{V} + c\mathbf{X}\mathbf{X}')^{-}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{8.8.1}$$

where c is any constant and $(\mathbf{V} + c\,\mathbf{X}\mathbf{X}')^{-}$ is any generalized inverse of $(\mathbf{V} + c\,\mathbf{X}\mathbf{X}')$.

(ii) The BLUE's of estimable functions $\mathbf{p}'\boldsymbol{\beta}$ and $\mathbf{q}'\boldsymbol{\beta}$ are $\mathbf{p}'\hat{\boldsymbol{\beta}}$ and $\mathbf{q}'\hat{\boldsymbol{\beta}}$, and

$$V(\mathbf{p}'\hat{\boldsymbol{\beta}}) = \sigma^2\{\mathbf{p}'[\mathbf{X}'(\mathbf{V} + c\,\mathbf{X}\mathbf{X}')^{-}\mathbf{X}]^{-}\mathbf{p} - c\,\mathbf{p}'\mathbf{p}\}, \tag{8.8.2}$$

and

$$\text{cov}\,(\mathbf{p}'\hat{\boldsymbol{\beta}}, \mathbf{q}'\hat{\boldsymbol{\beta}}) = \sigma^2\{\mathbf{p}'[\mathbf{X}'(\mathbf{V} + c\,\mathbf{X}\mathbf{X}')^{-}\mathbf{X}]^{-}\mathbf{q} - c\,\mathbf{p}'\mathbf{q}\}. \tag{8.8.3}$$

(iii) An unbiased estimate of σ^2 is

$$f^{-1}R_0^2 = f^{-1}\min_{\boldsymbol{\beta}}\,(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{V} + c\mathbf{X}\mathbf{X}')^{-}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{8.8.4}$$

where $f = R(\mathbf{V} \,\vdots\, \mathbf{X}) - R(\mathbf{X})$, $R(\mathbf{X})$ denoting the rank of $\mathbf{X}$.

(iv) To test a set of k linear hypotheses $\mathbf{p}_i'\boldsymbol{\beta} = d_i$, $i = 1, \cdots, k$, compute $u_i = \mathbf{p}_i'\hat{\boldsymbol{\beta}} - d_i$ and the dispersion matrix $\sigma^2\,\mathbf{D}$ of $\mathbf{u}' = (u_1, \cdots, u_k)$ by using the formulae (8.8.2) and (8.8.3). Compute the two statistics

$$T_1 = \mathbf{u}' \mathbf{D}^- \mathbf{u} \quad \text{and} \quad T_2 = \mathbf{D}\mathbf{D}^-\mathbf{u} - \mathbf{u}, \tag{8.8.5}$$

where $\mathbf{D}^-$ is any g-inverse of $\mathbf{D}$. The hypothesis is rejected if T_2 is nonnull or the statistic

$$F = \frac{\mathbf{u}'\mathbf{D}^-\mathbf{u}}{h} \div \frac{R_0^2}{f}, \quad \text{where} \quad R(\mathbf{D}) = h, \tag{8.8.6}$$

as a variance ratio on h and f D.F. exceeds some chosen critical value.

Rao [1971d] showed that the most general form of the inverse matrix to use in (8.8.1) is $(\mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}')^-$ in place of $(\mathbf{V} + c\mathbf{X}\mathbf{X}')$, where $\mathbf{U}$ is arbitrary, subject to the condition that the space generated by the columns of $\mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}'$ contain the columns of $\mathbf{V}$ and $\mathbf{X}$. Rao [1971d] also gave another unified approach which reduces the problem of linear estimation in the general case to a numerical computation of the inverse of a certain partitioned matrix. A case has also been made for obtaining a BLE (best linear estimator) dropping the condition of unbiasedness.

8.9. *Cluster Analysis*

A recent book by Jardine and Sibson [1971] and the Mamai proceedings volume edited by Hodson *et al.* [1971] contain the recent trends of research in cluster analysis.

The first step in cluster analysis is the construction of a similarity or dissimilarity matrix between units to be classified (Mahalanobis, Majumdar, and Rao [1949] and Majumdar and Rao [1958]). This depends on the characteristics (variables) measured on the units and the measure of similarity chosen. However, there is no adequate discussion in the literature on the choices of variables and measures of similarity *in relation* to objectives of classification.

The second step is to build a hierarchical system which connects units at various levels of similarity. A number of methods have been suggested for this purpose which allow for non-overlapping and overlapping clusters. Attempts are also made to impose a tree structure and estimate the time points at which branching took place (Cavalli-Sforza and Edwards [1964; 1967], Edwards [1970]). The subject is still in its initial stages of development and the relevant problems are not always clearly defined.

8.10. *Whither likelihood principle?*

Godambe (see Godambe and Sprott [1971]) brought to the notice of statisticians through a simple example in sample surveys the need to review the likelihood principle. Let there be n units in a finite population, numbered $1, \cdots, N$ with variate values $X_1, \cdots, X_n$. When a unit is chosen for study we have a bivariate observation (y, x), y standing for the number of the unit and x for the variate value. The object is to estimate the parameter, $T = X_1 + \cdots + X_n$, on the basis of observations $S = (y_1, x_1), \cdots, (y_n, x_n)$ on a sample of n different units chosen according to some design. Let $s =$

$(x_1, \cdots, x_n)$ be the set of observations when $y_1, \cdots, y_n$ are ignored. Now the likelihood of T given S, $L(T \mid S)$ is independent of T while $L(T \mid s)$ depends on T showing that the likelihood does not carry information on T unless the *variable y is ignored* (at least partly). Till now no satisfactory method is known for an effective utilization of the variable y in addition to x (except for dogmatic suggestions to ignore y).

In quite a different context, it was shown that a test based on two variables using Hotelling's T^2 is less efficient than Student's t based on one of the variables, which is a paradoxical situation, since logically speaking two variables contain more information than one of them (see Rao [1952] p. 252).

These situations only serve to prove that the statistical tools which make an effective use of multiple measurements are yet to be forged.

REFERENCES

Afifi, A. A. and Elashoff, R. M. [1966; 1967; 1969]. Missing observations in multivariate statistics. *J. Amer. Statist. Ass. 61*, 595–605; *62*, 10–29; *64*, 337–58, 359–65.

Anderson, T. W. [1958]. *An Introduction to Multivariate Analysis*. Wiley, New York.

Bagai, O. P. [1965]. The distribution of the generalized variance. *Ann. Math. Statist. 36*, 120–30.

Bartlett, M. S. [1938]. Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc. 34*, 33–40.

Bartlett, M. S. [1947]. Multivariate analysis. *J. R. Statist. Soc. Suppl. 9*, 176–97.

Bartlett, M. S. [1951]. The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann Eugen. 16*, 199–214.

Bildikar, S. and Patil, G. P. [1968]. Multivariate exponential-type distributions. *Ann. Math. Statist. 39*, 1316–25.

Burnaby, T. P. [1966]. Growth invariant discriminant functions and generalized distances. *Biometrics 22*, 96–110.

Carrol, D. J. [1971]; Some notes on multidimensional scaling. Bell Telephone Laboratory. Tech. Report.

Cavalli-Sforza, L. L. and Edwards, A. W. F. [1964]. Analysis of human evolution under random genetic drift. *Cold Spring Symp. Quan. Biol. 29*, 9–20.

Cavalli-Sforza, L. L. and Edwards, A. W. F. [1967]. Phylogenetic analysis. *Evolution 21*, 550–70.

Constantine, A. G. [1963]. Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist. 34*, 1270–85.

Consul, P. C. [1969]. On the exact distribution of Votaw's criteria for testing compound symmetry of a covariance matrix. *Ann. Math. Statist. 40*, 836–43.

Cox, D. R. [1971]. The analysis of multivariate binary data. Imperial College, London, Tech. Report.

Davis, A. W. [1970]. On the null distribution of the sum of the roots of a multivariate beta distribution. *Ann. Math. Statist. 41*, 1557–62.

Dempster, A. P. [1969]. *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading, Mass.

Dempster, A. P. [1971]. An overview of multivariate data analysis. *J. Multivariate Analysis 1*, 316–46.

Edwards, A. W. F. [1970]. Estimation of the branching points of a branching diffusion process. *J. R. Statist. Soc. B 32*, 155–74.

Fulkerson, D. R. and Gross, O. A. [1965]. Incidence matrices and interval graphs. *Pacific J. Math. 15*, 835–55.

Geisser, S. [1971]. The inferential use of predictive distributions. *Foundations of Statistical Inference*, 456–69.

Ghurye, S. G. and Olkin, I. [1962]. A characterization of the multivariate normal distribution. *Ann. Math. Statist.* **33**, 533–41.

Gnanadesikan, R. and Wilk, M. B. [1969]. Data analytic methods in multivariate statistical analysis. In: *Multivariate Analysis*, Vol. 2. Krishnaiah, P. R. (Ed.), 593–637.

Gnanadesikan, R. and Lee, T. Elisa [1970]. Graphical techniques for internal comparisons amongst equal degree of freedom groupings in multiresponse experiments. *Biometrika* **57**, 229–37.

Godambe, V. P. and Sprott, D. A. [1971]. *Foundations of Statistical Inference.* Holt, Rinehart and Winston, Toronto.

Goodman, N. R. [1963]. Statistical analysis based on a certain multivariate complex distribution. *Ann. Math. Statist.* **26**, 467–77.

Green, P. E. and Carmone, F. J. [1970]. *Multidimensional Scaling and Related Techniques in Marketing Analysis.* Allyn and Bacon, Boston.

Grizzle, J. E. and Allen, D. M. [1969]. Analysis of growth and dose response curves. *Biometrics* **25**, 357–82.

Hartley, H. O. and Jayatillake, K. S. E. [1971]. Estimation for linear models with unequal variances. Submitted to *J. Amer. Statist. Ass.*

Hodson, F. R., Kendall, D. G., and Tautu, P. [1971]. *Proceedings of the Anglo-Romanian Conference on Mathematics in Archaeological and Historical Sciences.* Edinburgh University Press.

James, A. T. [1964]. Distribution of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35**, 475–501.

Jardine, N. and Sibson, R. [1971]. *Mathematical Taxonomy.* Wiley, New York.

Kendall, D. G. [1963]. A statistical approach to Flinders Petrie's sequence dating. *Bull. Int. Statist. Inst.* **40**, 657–80.

Kendall, D. G. [1969]. Incidence matrices, interval graphs and seriation in archaeology. *Pacific J. Math.* **28**, 565–70.

Kendall, D. G. [1970]. A mathematical approach to seriation. *Phil. Trans. Roy. Soc. London A* **269**, 125–34.

Khatri, C. G. [1966]. A note on Manova model applied to problems in growth curves. *Ann. Inst. Math. Statist.*, Tokyo, **18**, 75–86.

Khatri, C. G. [1967]. Some distribution problems associated with the characteristic roots of $S_1S_2^{-1}$. *Ann. Math. Statist.* **38**, 944–8.

Khatri, C. G. [1968]. Some results for the singular normal multivariate regression models. *Sankhyā A* **30**, 267–80.

Khatri, C. G. [1970]. On the moments of traces of two matrices in three situations for complex multivariate normal populations. *Sankhyā A* **32**, 65–80.

Khatri, C. G. and Pillai, K. S. C. [1968]. On the noncentral distribution of two test criteria in multivariate analysis of variance. *Ann. Math. Statist.* **39**, 215–26.

Khatri, C. G. and Srivastava, M. S. [1971]. Exact distribution of the likelihood ratio test statistics for sphericity for the central and the noncentral case. *Sankhyā A* **33**, 201–6.

Khatri, C. G. and Rao, C. R. [1971]. Functional equations and characterization of probability laws through linear functions of random variables. Discussion paper No. 68, Indian Statistical Institute.

Krishnaiah, P. R. and Chang, T. C. [1970]. On the exact distributions of the traces of $S_1(S_1 + S_2)^{-1}$ and $S_1S_2^{-1}$. Tech. Report. ARL 70-0241, ARL.

Krishnaiah, P. R. [1971]. On the exact distributions of the extreme roots of the Wishart and Manova matrices. *J. Multivariate Analysis* **1**, 108–17.

Krishnaiah, P. R. and Waikar, V. B. [1971]. Exact joint distributions of any few ordered roots of a class of random matrices. *J. Multivariate Analysis* **1**, 308–15.

Kruskal, J. B. [1964a]. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27.

Kruskal, J. B. [1964b]. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**, 115–29.

Kshirsagar, A. M. [1961]. The goodness of fit of a single (non-isotropic) hypothetical principal component. *Biometrika* **48**, 397–407.

Kshirsagar, A. M. [1962]. A note on direction and collinearity factors in canonical analysis. *Biometrika 49*, 255–59.

Kshirsagar, A. M. [1964]. Distribution of the direction and collinearity factors in discriminant analysis. *Proc. Camb. Phil. Soc. 60*, 217–25.

Kshirsagar, A. M. [1971]. Goodness of fit of a discriminant function from the vector space of dummy variables. *J. R. Statist. Soc. B* (in press).

Linnik, Yu. V. [1953]. Linear forms and statistical criteria-I and II. *Ukrain Mat. Zhurnal 5*, 207–43, 247–90.

Lord, F. M. [1955]. Estimation of parameters from incomplete data. *J. Amer. Statist. Ass. 50*, 870–6.

Mahalanobis, P. C., Majumdar, D. N., and Rao, C. R. [1949]. Anthropometric survey of the United Provinces, 1941 (a statistical study). *Sankhyā 9*, 90–324.

Majumdar, D. N. and Rao, C. R. [1958]. Bengal Anthropometric Survey, 1945 (a statistical study). *Sankhyā 19*, 201–408.

Mathai, A. M. [1970]. Application of generalised special functions in statistics. (Mimeographed) McGill University and Indian Statistical Institute.

Matthai, A. [1951]. Estimation of parameters from incomplete data with application to design of sample surveys. *Sankhyā 11*, 145–52.

Mehta, M. L. [1967]. *Random Matrices and the Statistical Theory of Energy Levels*, Academic Press, New York.

Moran, P. A. P. [1971]. Estimating Structural and Functional Relationships. *J. Multivariate Analysis 1*, 232–55.

Patil, G. P. and Joshi, S. W. [1968]. *A Dictionary and Bibliography of Discrete Distributions.* Oliver and Boyd, Edinburgh.

Petrie, Flinders, W. [1899]. Sequences in prehistoric remains. *J. Anthrop. Inst. 29*, 295–301.

Pillai, K. C. S., Ali-Ani, S., and Jouris, M. S. [1969]. On the distributions of the ratios of the roots of a covariance matrix and Wilk's criterion for tests of three hypotheses. *Ann. Math. Statist. 40*, 2033-40.

Pillai, K. C. S. and Jouris, G. M. [1971]. Some distribution problems in the multivariate complex Gaussian case. *Ann. Math. Statist. 42*, 517–25.

Porter, C. S. (Ed.) [1965]. *Statistical Theories of Spectra: Fluctuations.* Academic Press New York.

Pothoff, R. F. and Roy, S. N. [1964]. A generalized multivariate analysis model useful especially for growth curve problems. *Biometrika 51*, 313–26.

Puri, M. L. and Sen, P. K. [1971]. *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

Puri, M. L. (Ed.) [1970]. *Nonparametric Techniques in Statistical Inference.* Cambridge University Press.

Radcliffe, J. [1966]. Factorizations of the residual likelihood criterion in discriminant analysis. *Proc. Camb. Phil. Soc. 62*, 743–52.

Ramachandran, B. and Rao, C. R. [1968]. Some results on characteristic functions and characterizations of the normal and generalized stable laws. *Sankhyā 30*, 125–40.

Rao, C. R. [1946]. Tests with discriminant functions in multivariate analysis. *Sankhyā 7*, 407–14.

Rao, C. R. [1948]. Tests of significance in multivariate analysis. *Biometrika 35*, 58–79.

Rao, C. R. [1952]. *Advanced Statistical Methods in Biometric Research.* Wiley, New York.

Rao, C. R. [1956]. Analysis of dispersion with incomplete observations on one of the characters. *J. R. Statist. Soc. B 18*, 259–64.

Rao, C. R. [1958]. Some statistical methods for comparison of growth curves. *Biometrics 14*, 1–17.

Rao, C. R. [1959]. Some problems involving linear hypotheses in multivariate analysis. *Biometrika 46*, 49–58.

Rao, C. R. [1960]. Multivariate analysis: an indispensable statistical aid in applied research. *Sankhyā 22*, 317–38

Rao, C. R. [1964]. Sir Ronald Aylmer Fisher—the architect of multivariate analysis. *Biometrics 20*, 286–300.

Rao, C. R. [1965a]. *Linear Statistical Inference and its Applications*. Wiley, New York.

Rao, C. R. [1965b]. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika 52*, 447–58.

Rao, C. R. [1966]. Discriminant function between composite hypotheses and related problems, *Biometrika 53*, 339–45.

Rao, C. R. [1967]. Least squares theory using an estimated dispersion matrix and its applications to measurement of signals. *Proc. 5th Berkeley Symp. 1*, 355–72.

Rao, C. R. [1969]. A decomposition theorem for vector variables with a linear structure. *Ann. Math. Statist. 40*, 1845–9.

Rao, C. R. [1970]. Computers: a great revolution in scientific research. *Proc. Indian National Science Academy 36*, 123–39.

Rao, C. R. [1971a]. Estimation of variance and covariance components—MINQUE theory, *J. Multivariate Analysis 1*, 257–75.

Rao, C. R. [1971b]. Minimum variance unbiased estimation of variance components. *J. Multivariate Analysis 1*, (in press).

Rao, C. R. [1971c]. Characterization of probability laws by linear functions. *Sankhyā A 33*, 265–70.

Rao, C. R. [1971d]. Unified theory of linear estimation. *Sankhyā A 33*, 370–96.

Rao, C. R. and Mitra, S. K. [1971a]. *Generalized Inverse of Matrices and its Applications*. Wiley, New York.

Rao, C. R. and Mitra, S. K. [1971b]. Further contributions to the theory of generalized inverse of matrices and its applications. *Sankhyā A 33*, 289–300.

Rao, J. N. K. and Subrahmaniam, K. [1971]. Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics 27*, 971–90.

Rāo, M. N. and Rao, C. R. [1966]. Linked cross sectional study for determining norms and growth rates—a pilot survey of Indian school-going boys. *Sankhyā B 28*, 237–58.

Reiersøl, O. [1950]. Identifiability of a linear relation between variables which are subject to error. *Econometrica 18*, 375–89.

Robinson, W. S. [1951]. A method for chronologically ordering archaeological deposits. *American Antiquity 16*, 293–301.

Roy, S. N. [1957]. *Some Aspects of Multivariate Analysis*. Wiley, New York.

Shepard, R. N. [1962a, b]. The analysis of proximities: multidimensional scaling with an unknown distance function I and II. *Psychometrika 27*, 125–39; 219–46.

Srivastava, J. N. [1966]. Incomplete multiresponse designs. *Sankhyā*.

Srivastava, J. N. [1968]. A general class of designs for multiresponse experiments. *Ann. Math. Statist. 39*, 1825–43.

Sugiyama T. [1970]. Joint distribution of the extreme roots of a covariance matrix. *Ann. Math. Statist. 41*, 655–7.

Torgerson, W. S. [1952]. Multidimensional scaling: I-theory and method. *Psychometrika* 17, 401–19.

Torgerson, W. S. [1958]. *Theory and Method of Scaling*. Wiley, New York.

Trawinski, I. M. and Bargmann, R. E. [1964]. Maximum likelihood estimation with incomplete multivariate data. *Ann. Math. Statist. 35*, 647–57.

Truett, J., Cornfield J., and Kannel, W. [1967]. A multivariate analysis of the risk of coronary heart disease in Framingham. *J. Chron. Dis. 20*, 511–24.

Wigner, E. P. [1967]. Random matrices in physics. *SIAM Rev. 9*, 1–23.

Wilkinson, E. M. [1971]. Seriation Theory (Rayleigh prize essay, University of Cambridge).

Williams, E. J. [1952a]. Some exact tests in multivariate analysis. *Biometrika 39*, 17–31.

Williams, E. J.]1952b]. Use of scores for analysis of association in contingency tables. *Biometrika 39*, 274–89.

Williams, E. J. [1959]. *Regression Analysis*. Wiley, New York.

Williams, E. J. [1961]. Tests for discriminant functions. *J. Aust. Math. Soc. 2*, 243–52.

Williams, E. J. [1967]. The analysis of association among many variates. *J. R. Statist. Soc. B* *29*, 199–228.

Wilks S. S. [1932]. Certain generalizations in the analysis of variance. *Biometrika 24*, 471–94.

Wilks, S. S. [1946]. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math. Statist. 17*, 257–81.

Wishart, J. [1938]. Growth rate determinations in nutrition studies with the bacon pig and their analysis. *Biometrika 30*, 16–28.

Wooding, R. A. [1956]. The multivariate distribution of complex normal variates. *Biometrika 43*, 212–5.

Received October 1971

Key Words: Discriminant analysis; Archaeological seriation; Multidimensional scaling; Growth curves; Characterization of multivariate normal; Multivariate list criteria; Cluster analysis; Variance and covariance components; Least squares; Likelihood.