The Economic Journal of Nepal, Vol.10, No.1, Issue 37, January-March 1987 © DOE-TU

On Some Multivariate Analyses

BAL KUMAR K.C.*

The purpose of this study is to distinguish between four multivariable analyses: factor analysis, multiple regression analysis, canonical correlation analysis, and discriminant analysis.

FACTOR ANALYSIS

Factor analysis is now a common tool in the hand of scientists of several disciplines used extensively to investigate the clustering of attributes in physical, social or decision space. Ever since the pioneering efforts by Charles Spearman,[1] factor analysis has been the domain of psychologists and psychometricians. With the advent of computer systems, this technique has aided in the reduction and organization of vast amount of areal data, and has contributed significantly to regional classification schemes. Whether the technique is to be used for hypothesis testing or for exploratory research is still in debate, although many geographic studies emphasize the inductive use.

Factor analysis takes thousands of measurements and qualitative observations and "resolves them into their distinct patterns of occurrence. It makes explicit and more precise the building of fact linkages continuously going on in human mind."[2] It involves the reduction or simplification of a multitude of variables into a hypothetical set of variates, fewer in number, that explain a large percentage of variance. It attempts to identify the characteristics which the variables have in common and which result in their intercorrelation.

Confusion exists between factor analysis and principal component analysis. Principal component analysis is a generally useful procedure whenever the task is to determine the minimum number of independent dimensions needed to account for most of the variance in the original set of variables. Principal component analysis has a number of components (k) equal to the number of variables (m); it takes m variables and collapses these into another set of k variables. The new variables are called principal components. In factor analysis k is less than m. In principal component analysis percent total r and percent common variance is the same. In factor analysis common variance is only the proportion of total variance explained by all the factors in particular jth component. The communality (h^2) in principal component is 1. In factor analysis it is less than 1.

*Dr. K.C. is Professor of Geography at Tribhuvan University, Kirtipur. This paper was prepared when the author was a Visiting Professor of Public and Environmental Affairs at Indiana University, Bloomington in 1986.

Mathematically, in principal components, dimensions are variables and points are observations.

In factor analysis dimensions are observations and points are variables. In principal component analysis there should be transformation of each variable into a new set of variables (called components), such that the following three criteria are met to make the solution unique. (1) That y_1 accounts for the maximum possible of all the variance in the set of X's and y_2 for maximum possible of remaining y_3 and so on. (2) Total variance of y's = Total variance of X's or (Var y_1 + Var y_2 + ... + Var y_m) = (Var X_1 + Var X_2 + ... + Var X_m) and y's are orthogonal. Factor analysis does not account for the first two criteria due to the absence of unique solution. The third criterion is set or sometimes, dropped. Conceptually, principal component concerns with variance (related to observation). Factor analysis is concerned with covariance and covariance with variables. There is only one principal component matrix which can fit to a regression set of data. On the other hand, there are several or infinity of factor analysis model. The factor analysis solution is very undetermined. It is also a grouping technique if Q factor analysis is performed.[3] Factor analysis is based on a set of well-known equations that imply certain assumptions concerning the variables to be analyzed. Naturally, the application of basic model poses certain technical difficulties, as the assumptions must fit the actual data theoretically and operationally.[4] The basic equations of factor analysis and their corollaries may be compactly stated in terms of matrix algebra for any raw data matrices as follows:

${}_{n}X_{m}$ = raw data matrix: n= # of observations
m= # of variables
In general n should equal 3 (m)

${}_{n}X_{m}$ = matrix of deviation scores for raw data

${}_{n\times m}Z$ = matrix of standardized data scores, i.e., $zij = \frac{xij - xj}{sj}$ where sj = standard deviation of jth column (variable) of X

${}_{m\times m}S$ = variance/covariance matrix of raw data X where $S = \frac{1}{n-1} X^T X$

${}_{m\times m}V$ = variance/covariance matrix of standardized data (Z), where $V = \frac{1}{n-1} Z^T Z$

R = inter-correlation matrix for data, where R = V, because correlation between the kth and ℓth variable is:

$$\gamma k\ell = \text{covariance } \frac{(Xkj\ XL)}{Sk\ SL} = \frac{\Sigma(Xik-\bar{X}k)\ (Xi\ell - \bar{X}\ell)}{SkSL}$$

where Sk, SL = standard deviations SK, SL = 1 if data are standardized.

The basic set of equations may be formulated as:

$$Zj = \underbrace{WjGfG}_{a} + \underbrace{Wj \cdot Sf \cdot Sj}_{b} + \underbrace{Wj \cdot ef \cdot ej}_{c}$$

According to Spearman, (a) there is one general factor which affects all variables, (b) m specific factors, one for each variable, and (c) n error terms, one for each variables.

The Bi-factor or group model postulates three things plus a group model. Such as:

$$Zj = \underbrace{WjGFG}_{a} + \underbrace{WjLFv\ell}_{b} + \underbrace{WjSFSj}_{c} + \underbrace{WjeFej}_{d}$$

where a = general factor for all variables,

b = group factor, c = specific factor, and d = error.

Thurstone's common factor model (1947) postulates that k (< m), common factors associated with two or more variables, m specific factors, and error terms like:

$$Zj = Wj \cdot 1F1 + Wj \cdot 2F2 + Wj \cdot kFk + \underbrace{Wj \cdot Sfsj + WjeFe}_{\text{unique element}}$$

This model is sufficient to incorporate other models. Therefore:

$$Sj = 1 = \underbrace{wj1^2 + wj2^2 + wj3^2 + \ldots + wjk^2}_{h^2} + \underbrace{w^2j \cdot s + w^2j \cdot e}_{1-h^2}$$

In simple form, the total variance can be shown as:

Common Variance						Spec. Var.	Error Var.	
$wj1^2$	$wj2^2$	$wj3^2$	$wj4^2$		wjk^2	$w^2j \cdot s$	$w^2j \cdot e$	= 1

The values of $\sqrt{wj1^2}$... are factor loadings, they represent the correlation of that variable j, with each factor. The common varainces are important as they represent common elements running through the data and resulting in high correlations. The specific variance is only of signifinance in the variable under consideration and is unrelated to the other variables. The value of the square root of the sum of the common variances for each variables is called the factor loading.[5] Also the communality of variable i, is approximated by computing the coefficient of determination resulting from the regression of i on the remaining i variables in the set. In this type of analysis only the common variance is factored, not all of the variance, as in the case of principal component solution.

Berry[6] mentioned that there existed sixteen readily available factorial models. These models may be orthogonal or oblique, treat errors in a constant manner or spread it differently among the variables, treat loadings in a specified or unconstrained sense. Depending upon the method, the output may differ significantly even for the same data set. Factor comparability requires the assumption of homogeneity,[7] by using Kuder-Richarson formula:[8] $\gamma_n = n\bar{\gamma}it^2 - 1/\ (n-1)\bar{\gamma}it^2$. The indexes calculated for two different time periods can be statistically compared using Z score. Regarding the parsimoniousness, principal component solution was the most desirable.[9] There is also a problem of maintaining invariance which requires the closeness of fit of the factors from one study relative to the factors from other studies. Other problems such as relation between factors, selection of variables, common factors, definition of factors, non-uniqueness of solution, level of measurement, linearity, and distributional assumption, have been discussed elsewhere. Factor analysis is no substitute for theoretical and conceptual analysis of a problem. For causal interpretation multiple regression might prove to be better. Often factorial ecologists are worried about how much variance they can explain. It should be checked by means of computing factor scores, as well as by testing for reliability and validity.[10]

MULTIPLE REGRESSION ANALYSIS

Multiple regression is concerned with predicting dependent variable by two or more independent variables. The degree to which the dependent and the independent variables covary is reflected by the proportion of the total variation of the dependent variable that is associated with the variations of the independent variables. Multiple regression may also be viewed as a stepwise procedure in which the dependent variable is regressed on an independent variable holding other independent variables constant statistically. Similar reasoning applies to partial regression coefficients--the rate of change in Y for a unit change in any independent variable is computed, holding constants the effects of the other independent variables statistically. Regression coefficients are beta values which may be compared directly in order to evaluate the relative importance of each independent variable. If stepwise procedure is not used the multiple correlations cause problem of multi-collinearity. The partial correlation coefficients are also employed in stepwise regression procedure in which a sub-set of the total number of independent variables is chosen in order of importance of explaining the variability in the dependent variable. Partial regression coefficients are preferable when attention has been focussed on the prddiction equation itself. Whereas, partial correlation coefficients are preferred when one is interested in knowing the measure of correlation and variability. The linear equation is of the form: $Y = \alpha + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_mX_m + \varepsilon$ or using sample $y = a + b_1x_1 + b_2x_2 + \ldots b_mx_m + e$ where a=regression or y intercept; b = partial regression coefficient. For example, b_1 indicates a change in y per unit change in X_1 holding other variables constant. Multiple regression is based on the following assumptions: for each fixed value of Xs, the ys are:

(1) Normally distributed, (2) have same variance (homo-scedasticity), (3) are independent, and (4) have mean given by $y = \alpha + \beta X$. In other words, for each fixed value of X, the error terms are: (1) normally distributed, (2) equal variance, (3) independent, and (4) have mean of zero.

The criteria used to obtain the loadings in multiple regression is such that the b values related must result in minimizing the unexplained variation of the dependent variable given change in the independent variables. This means that the sum of squared deviation of y actual from $\hat{y}$ estimated by the normal equation must be minimized.

$$(\Sigma X_1 y = b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + bm\Sigma X_1 Xm + a\Sigma X_1) \text{ or}$$

$$(\Sigma y = b_1 \Sigma X_1 + b_2 \Sigma X_2 + \ldots + bm\Sigma Xm + \ldots + Na).$$

This minimization should be through the least square method such that, a and b will be unbiased elements of the population parameter α and β, also the standard error of estimate will be minimized.

The matrices used to obtain b values are:

$y = XB = E$, where y = column vector of estimated y's,

$X = nxm + 1$ = matrix of independent variables plus a column vector of 1's, B = column vector of b's to be obtained, E = column vector of error term. $X^T Y = X^T XB$; where $X^T y$ = vector of error products of dependent variable y and the independent variable X. $X^T X$ = matrix of sum of square and sum of cross products of X's. B = column vector of b's. Therefore $(X^T X)^{-1} X^T y = (X^T y)^{-1} X^T XB$, where: $(X^T X)^{-1}$ is the increase of $X^T X$ and is used to obtain $(X^T X)^{-1} X^T y = IB$ where $(X^T X)^{-1} X^T X = 1$.

The squared multiple correlation coefficient is simply the vector product of b and the predictor-criterion correlations (Cooley and Lohnes, 1971: 63). The overall significance of the multiple regression can be tested by an analysis of variance using F-ratio.11

FACTOR ANALYSIS AND MULTIPLE REGRESSION ANALYSIS

<u>Differences:</u> Although multiple correlation coefficients represent communalities by means of computing common variance, they are not in strict sense of the term the same thing. Communality depends on numbers of factors extracted whereas, multiple correlation coefficients depend on the number of meaningful independent variables explaining the dependent variable. Factor analysis deals with reducing fewer number of dimensions and helps to analyse how these dimensions may have been organized in space. If the investigator is interested in determining how are characteristics of a set of objects can be predicted from other characteristics while all the measures are continuously distributed variables, multiple regression analysis provides an analytical tool. Multiple regression analysis, bases its interpretation on the basis of intercorrelation

matrix and the correlation of each independent variable with the dependent variable. Factor analysis extends further through several matrices. Factor analysis is exploratory and hypothesis testing model. Multiple regression assumes that the errors are normally distributed, however in factor analysis it may not be so, because the first few factors explain much of the variance and the remaining factors can be considered residuals. However, factor analysis employing longitudinal data does pinpoint the pattern for the future if the nature of dimensions are relatively stable. Multiple regression, on the other hand, although, is said to be predictive, does not handle the longer prediction. Spatial auto-correlation is a severe problem in regression analysis, particularly, when the size of areal units are altered from one analysis to another.

Factor analysis assumes that relationships between variables are linear, and thus effects are additive. Many correlation and regression studies have shown that non-linear relationships are more likely. The relative value of using factor analysis over traditional multiple regression is also questionable. Meyer[12] has extensively compared and contrasted the use of multiple regression and factor analysis from empirical data. Examining the variables pertinent to black residential structure, Meyer found that the same variables which are highly intercorrelated in correlation analysis, did not correlate highly in the dimension of factor analysis. Hence, these two methods are conceptually different, their output also should be taken differently.

Similarities: Both analyses assume linearity and require interval scale data. Of course, Berry and Gould[13] have used ordinal scales. Both types of analyses deal with only one population at a time. Intercorrelation matrix of both types of analysis is the same, provided the data are the same. Although factor scores and residuals are not the same, sometimes they seem to serve the same purpose, especially for mapping purposes. From factor analysis, one can postulate causal connections between variables of the same dimension.[14] If these variables are examined by multiple regression they provide the identical results. Both types of analyses assume error terms which reduce the amount of variation explained. Although there is not much functional similarities between the two, multiple regression might be a good starting point for selecting variables for factor analysis. Sometimes, this type of fishing game destroys the necessity of conceptualization. However, for computing factor scores it is desired to find a linear regression of a common factor. This will require the covariance for common factors and the variance-covariance terms for the variables. To estimate the common factor we use the loadings and, to estimate the X's we use Σ. In matrix form it will be $[\hat{F}] = [V]^T \cdot \Sigma^{-1} \cdot [X]$. Where $\hat{F}$ is the estimated common factor and V's are loadings. For the case of 3 variables and 4 observations we use the following:

$$yL = \sum_{j=1}^{m} bjLZj$$

where yL is one particular factor and Zj is the jth variable (Xj) in standarized form (j = 1,2, ... j ... m). In matrix terms, we can specify the whole set of m factors as

$$y = \underset{nxm}{Z}\underset{mxm}{B}, \text{ so that the form of}$$

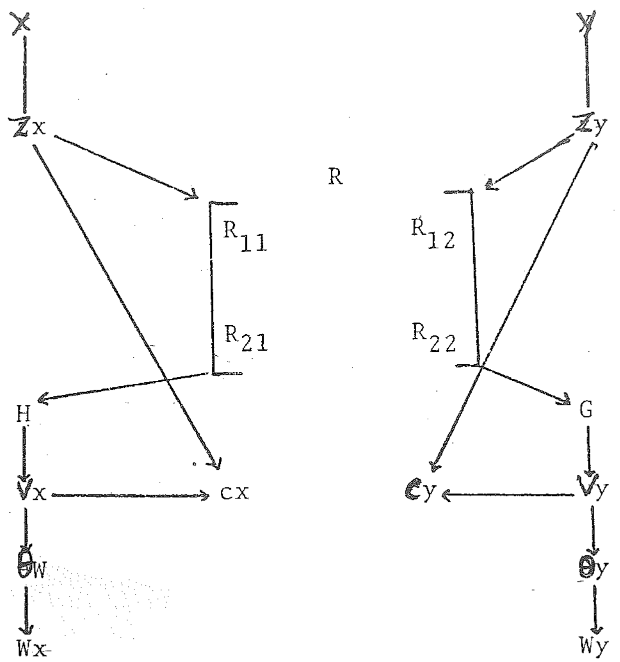
the above equations is the same as for regression, except that y and B are matrices rather than vectors.

CANONICAL CORRELATION ANALYSIS

Canonical correlation analysis is concerned with the determination of a linear combination of each of two sets of variables on the same subject such that the correlation between the two functions is a maximum. "Under certain conditions this analysis is equivalent to discriminant analysis and under certain conditions it is equivalent to multiple regression."[15] It is like multiple regression except that it includes more than one dependent variable. Which of the two sets of variables is considered the dependent does not matter. It is the most general form of correlation analysis. Multiple regression is concerned with the estimation of one predictant by a linear combination of predictors and is a special case of canonical correlation. The regression estimates can be found separately for any number of predictants even though the predictants are linearly related. If dummy variables denoting group membership are formed as predictants, "the resulting regression equations are identical to those derived from the canonical correlation analysis."[16] Furthermore, the interpretation of coefficients has all problems attendant to the beta coefficients of multiple regression.

"A principal distinguishing feature of different applications of canonical analysis is the normalization requirements imposed upon the canonical weights."[17] However, canonical correlation model seems to use the same analytical trick to display the structure of relationships as that of factor analysis except that the canonical model selects linear functions having maximum covariances between domains, and factor analytic model selects the linear function of test having maximum variances. Canonical model is an exploration of the extent to which individuals occupy the relative position in one neasurement as they do in other.[18] We interprete the interrelationships between the variables in each pair of both sets on the basis of how much they are correlated in the respective vector. Each pair of canonical vector is considered independent.

This analytical technique extracts pairs of canonical variates, one half of each pair is associated with the X set, and the other half is associated with the Y set. The matrices involved are:



From the two sets of raw data, a supermatrix R is formed by finding the intercorrelations between all pairs of variables in both data sets. R_{11} contains the intercorrelations between all pairs of variables in the X set, and R_{22} contains the intercorrelations between the variables in the Y set, R_{12} and R_{21} contain the intercorrelations between the variables in the X and Y sets.[19]

The loadings matrix Wx equals $R_{11}Vx\Theta x^{-\frac{1}{2}}$, and the loadings matrix Wy equals $R_{22}\ Vy\Theta y^{-\frac{1}{2}}$. The matrix Vx contains the standardized eigen vectors of a matrix H, where $H = R_{11}^{-1}\ R_{12}\ R_{22}^{-1}\ R_{21}$. The matrix Vy contains the standardized eigen vectors for a matrix G, where $G = R_{22}^{-1}\ R_{21}\ R_{11}^{-1}\ R_{12}$. The matrices Vx and Vy are therefore, the canonical weights which are used to correlate the canonical variates with the original variables in Zx and Zy. The matrices Θx and Θy are diagonal matrices containing the variance for the canonical variates of the x and y data sets, where Θx equals $Vx^{T}R_{11}Vx$ and Θy equals $Xy^{T}\ R_{22}Vy$. The matrices $\Theta x - \frac{1}{2}$ and $\Theta y - \frac{1}{2}$ therefore contain the standard deviations of the canonical variates. The square root of the eigen values λ is the simple correlation coefficient between the canonical variates and original variables. For either of the both x and y, the percent total variance accounted for by a specific canonical variate can be calculated by taking the sum of squared loadings in appropriate one-half of the canonical variate and dividing by the number of variables. The value is labelled Sxi or Syi depending upon the sets it represents.[20]

In addition, the redundancy measure is important, because a very large canonical correlation coefficient could be the result of a very large zero-order correlation of just one variable of one set with just one variable of the other set, and the remainder of the two sets could be essentially uninvolved in the canonical structure.[21] For example, the redundancy measure, $Rdy_1 = Sy_1 \cdot Rcv_1^2$ indicates the amount of variance in the first canonical variate of the Y set that is accounted for by the first canonical variable of the X set.[22] If this redundancy measure is low, then a high canonical correlation coefficient is relatively meaningless. Love and Steward have developed a summary index of redundancy.

$$\bar{R}y = \sum_{i=1}^{p \text{ or } q} Rcv_i^2 \cdot syi .$$

It is this proportion of the total redundancy, $Rdyi/\bar{R}Y$, that determines whether or not a pair of canonical variates is worthy of interpretation. Another problem is to determine the contribution of a variable to the total canonical solution. Love and Steward's another index is equal to multiple coefficient of determination or $R^2 = R_i^2/\Sigma R_i^2$. The means of these indices for X and Y are equal to $\bar{R}x$ and $\bar{R}y$.[23]

As in principal component analysis, the first pair of canonical variates extracts the largest amount of variance from the linear composites of original data. The second pair extracts maximum of the remaining variance while remaining uncorrelated to the first pair. All of the variance also is extracted in canonical correlation analysis. The maximum number of independent dimensions is equal to the number of variables in the smaller set of data. These dimensions are orthogonal to each other.

Since the canonical correlation coefficients are arranged in descending order, significant test has been devised.[24]

The lambda Λ is distributed approximately as chi-square with (p-r) (q-r) degrees of freedom for r canonical dimensions.

$$X = .[n-0.5\ 9q + q + 1)]\ \log e\ \Lambda \quad d.f. = p.q$$

$$\text{where } n = N-1 \text{ and } \Lambda = \Sigma \prod_{l=r+1}^{p \text{ or } q} (1-\lambda i) =$$

$$(1 - \lambda i) \cdot (1 - \lambda_2) \cdot \cdot \cdot (1 - \lambda \ p \text{ or } q)$$

If this proves to be significant, the second can be tested by using

$\Lambda = \prod_{i=2}^{p \text{ or } q} (1 - \lambda i)$. For this case the chi-

square has d.f. = (p - 1) (q - 1).

CANONICAL AND FACTOR ANALYSIS

Koons has suggested[25] some form of factor analysis of the original variables prior to canonical correlation analysis. This suggestion presents difficulty of interpretation.[26] Factor analytic approach may be superfluous. Unlike factor analysis, in which all the variables are treated as a single set, "Canonical analysis identifies the component interrelationships between two sets of data. Factors in factor analysis will fail to appear in the canonical analysis."[27]

However Ray (1971), in number of ways advocated the usefulness of link between factor analysis and canonical analysis. He interpreted the canonical variates as assisted "by their comparison identified in the factor analysis."[28] He further said: "Factor analysis is not as appropriate as canonical analysis for examining interrelationships between the two sets of characteristics ... even though characteristics in both sets suggested their interdependency."[29] Canonical analysis, like factor analysis is a descriptive device that reduces a number of variables to their underlying components and canonical variates, like factors, are made up of linear functions or combination of original variables.[30]

In factor analysis the dimensions may not be orthogonal (whether oblique, or centroid, or orthogonal). But canonical correlation analysis extracts orthogonal dimensions. Path analysis can incorporate more than one dependent variable and it would also be interesting to see the causal connection from the result of canonical analysis. However, the task of going from one to another analysis presents a continuous problem of validity and reliability. The only thing remained to be interpreted is the canonical scores. There seems to be semantic differences in labelling the analysis as "Canonical factor analysis" or "canonical correlation analysis." This creates a considerable confusion between factors as in factor analysis and variates in canonical analysis. In addition, principal component rather than factor analysis seems to have more affinity with canonical correlation analysis, in terms of variance/covariance and number of dimensions extracted. It must be added that canonical correlation analysis is more powerful tool, when, one has to compare between the variables of different time periods. The redundancy measure has no equivalent in factor analysis.

Provided that the measurement scale is the same, while using factor analysis, one has to perform two different sets of factor analysis and then use multiple regression of the factor score. However, a principal axes factor analysis of the attribute matrix produces factor scores.

This is an nxs matrix, where s is the number of factors. From this matrix it is possible to compute a new matrix Δ, with (n^2-n) rows and s columns, which contains the distances between each pair of areas on each factor. A factor analysis of the interaction matrix yields a factor scores matrix B, with (n^2-n) rows and b columns, where b represents the number of dimensions extracted by the factor analysis. Canonical correlation of two matrixes Δ and B allows the statement of relation between the two, "Structural and functional system."[31] Steward and Love (1968) said that "if we have to component analyze two sets of variables independently and then develop weights which would rotate the two component structures to maximize correlation, we would have a canonical solution."[32] Moreover, canonical factor scores are said to be corresponding to the scores computed in principal component analysis.[33]

VI. DISCRIMINANT ANALYSIS

Discriminant analysis is employed for a set of observations which are already classified in some manner (or by cluster analysis). The main purpose is to maximize between group variance and minimize within group variance. Its most common use in geographic research is as aid in classification. "The evaluation of classification is clearly related to so-called "cluster analysis."

Discriminant analysis extracts N observations on m variables. These observations are divided into t groups, with n_2 observations in each group. Thus, $n1 + n2 + \ldots + nt = N$. We can denote each observation by: xhij, where ith observation ($i = 1,2, \ldots n2$) in hth group ($h = 1,2, \ldots t$) on jth variables ($j = 1,2, \ldots m$). Observations may be arranged in the matrix X which is NXm.

		Variables			
		x_{111}	x_{112}	...	x_{11m}
	Group 1	x_{121}	x_{122}	...	x_{12m}
X =		x_{131}	x_{132}	...	x_{13m}
		⋮	⋮		⋮
		x_{1n_11}	x_{1n_12}		x_{1n_2m}
		x_{211}	x_{212}	...	x_{21m}
	Group 2	x_{221}	x_{222}		⋮
		⋮	⋮		
		x_{2n_21}	x_{2n_22}	...	x_{2n_2m}
		x_{t11}	x_{t12}	...	x_{t1m}
	Group 3	x_{t21}	x_{t22}	...	x_{t2m}
		⋮	⋮		⋮
		x_{tn_t1}	x_{tn_t2}		x_{tn_tm}

Three matrices can be defined, which contain sums of squared deviation and sums of cross products.

$$\underset{m\times m}{T} = \underset{m\times n}{DT^T}\ \underset{n\times m}{D^T} \quad ; \quad \underset{m\times m}{G} = \underset{m\times n}{DG^T}\ \underset{n\times m}{DG} \quad ; \text{ and}$$

$$\underset{m\times m}{A} = \underset{m\times n}{DA^T}\ \underset{n\times m}{DA} \quad ; \quad \text{and} \quad T = G + A$$

Equivalent matrices for y's are FT = DTV ; FG = DGV; and FA = DAV. Similarly, we can define sums of square and cross product associated with y's:

$$F^TGFG = (DGV)^T\ DGV = V^TDG^TDGV = V^TGV \text{ and } FA^TFA =$$

$(DAV)^T\ DAV = V^TDA^TDA^V = V^TAV$. So the linear discriminant function can be defined as maximizing the ratio:

$$\lambda = \frac{V^TAV}{V^TGV} = V^TAV(V^TGV)^{-1} = V^TAW^{-1}V$$

A statistical test is available, other than analysis of variance.

$$X^2 = (N = \frac{m + t}{2} - 1) \log \Lambda \ ;$$

$$\text{d.f.} = M(j-1), \text{ where } \Lambda = \prod_{j=1}^{n} \frac{1}{1+\lambda j} \ .$$

The second can be tested by $\Lambda^1 = \prod_{j=2}^{n} \frac{1}{1+\lambda j}$

with d.f. = (m-1) (t-2).

DISCRIMINANT AND OTHER ANALYSIS

Discriminant and Factor Analysis

The discriminant analysis "may be interpreted as a special type of factor analysis that extracts orthogonal factors of the measurement [variables] for the specific task of displaying and capitalizing on differences among criterion groups." The discriminant analysis is similar to factor nalysis and to principle component analysis in identifying major linear dimensions of classifications, much in the same way that factor analysis tends to determine dimensions of variability.[34] It is possible to compute the linear discriminant functions which are linearly related to the factor scores used as input to the algorithm. The coefficients are determined in such a way that discrimination between groups is maximized. "Thus the method has a strong similarity to principal component analysis."[35] In discriminant analysis, the linear discriminant functions are lesser in number by 1 than the number of groups extracted. If these groups are factor analyzed by imposing the number of dimensions as many as the number of linear discriminant functions the results may be similar, more so particularly if we use component analysis.

Discriminant and factor analysis are in some respects as much different as the difference between principal component analysis and factor analysis. Discriminant analysis requires a priori classification, factor analysis does not. The amounts of variance explained are different. However, the Q mode factor analysis does the similar task as discriminant analysis.

Discriminant and Regression Analysis

Similarities: The discriminant function $y = b_1X_1 + b_2X_2 + \ldots + b_pX_p$ is the form of the multiple regression equation except that the constant is omitted. "In classifying observations, this value would be of no importance since it is the same for all items. The values of b coefficients are obtained by application of the least square method."[36] The analysis of variance for discriminant analysis is basically the same as it is for multiple regression analysis.

Discriminant analysis can also be used as a test of whether regression of a criterion on a set of measurements is nearly linear or not. If the regression of the criterion scores on the measurements is truely linear, only one discriminant function should be obtained. If more than one discriminant function is obtained, the regression should be nonlinear.

Differences: The question attacked by discriminant analysis is different from the question answered by the multiple regression technique.[37] No matter how the discriminant analysis turns out, whatever it says is new information not supplied by the multiple regression technique. The discriminant analysis does what the multiple correlation approach does not. It uses group membership as the criterion and tries to maximize between variance and minimize within variance. The multiple correlation technique applied to one group ignores all the data from other groups. The discriminant analysis employs all the data from all the groups. The discriminant technique requires the same kind of measurement on all the members of all the groups. Otherwise, there would be no basis for differentiating the groups. There is no basis for between groups comparison by means of multiple regression technique. In spite of these differences of the two techniques, sometimes they have been taken as the same. Regression equations are not always to be preferred to discriminant function like it is to canonical functions. When the number of groups is large, analysis of the multi-dimensional discriminant space is difficult and involves additional assumptions. Regression analysis is designed to answer the question, "What am I best at," discriminant analysis is designed to answer the question, "What group am I most like?"[38]

Discriminant and Canonical Analysis

Glahn has shown that the discriminant analysis solution is equivalent to the canonical correlation solution, the purposes of both of these are, however, different. Canonical correlation does not classify, discriminant does.

The first canonical variate would give the best quadratic function of X for discriminating between groups only in terms of quadratic transformation but not in terms of linear transformation. However, both types of analysis assume the main diagonal ewual to 1 like in principal component analysis. If two different sets of groups are used, canonical analysis would give interesting results, particularly when we have to compare between two periods.

CONCLUSION

Procedures of multivariate analyses are concerned with the problem of reducing the original set-space to the minimum number of dimensions in order to describe relevant information contained in the original observations. Models which are not criterion oriented and which have logical implications of systematic correlations within self of measurement fall under the rubric of factor analysis. When the researcher is interested in determining how one characteristic of variables can be prdicted from other characteristics, multiple correlation is useful. When one is trying to handle two sets of data of the same subject canonical correlation analysis may be important, which extracts orthogonal dimensions of common variance or redundancy between the two sets. Discriminant analysis is useful in examining the group membership of individuals in such a way that the ratio of the among-groups to the within groups sums of squared deviations from group means on this discriminant function is maximized. The superiority of one type of analysis over another and their complementary uses depend upon a researcher's interest and what he is trying to accomplish.

FOOTNOTES

1. G. Spearman, "General Intelligence, Objective Determined and Measured", American Journal of Psychology, Vol. 15, 1904, pp. 201-293.

2. R.J. Rummel, "Understanding Factor Analysis", Journal of Conflict Resolution, Vol. 11, 1967, p. 445.

3. W.W. Cooley and P.R. Lohnes, Multivariate Procedures for the Behavioural Sciences, (New York: John Wiley and Sons, 1962).

4. C.G. Jansen, "Some Problems of Ecological Factor Analysis", Social Ecology, Edited by M. Dogan and S. Rokkan, (Cambridge: MIT Press, 1969).

5. B. Fruchter, Introduction to Factor Analysis, (New Jersey, Princeton: Van Nostrand, 1954).

6. Brain J.L. Berry, "Introduction: The Logic and Limitations of comparative Factorial Ecology", Economic Geography, Vol. 47, 1971, Supplement, pp. 209-219.

7. R.H. Athanasiou, "The Reliability of Homogeneity", Proceedings of American Psychological Association, Vol. 29, 1969, pp. 119-120.

8. R.H. Gaylord, "Estimating Test Reliability from Item Test Correlations", Educational Psychological Measurement, Vol. 29, 1969, pp. 303-304.

9. R.M. Thorndike and D.J. Weiss, "Parsimony of Factor Analytical Methods for Diverse Scientific Areas", Reprint from the Proceedings of the 78th Annual Convention of the American Psychologist's Association, 1970, pp. 111-112.

10. David R. Heise and George W. Borhnstedt, "Validity, Invalidity and Reliability", Sociological Methodology 1970, edited by Edgar F. Borgatta and George W. Bornstedt, (San Francisco: Jossey-Bass Inc. Publishers 1970), pp. 104-129.

11. Charles T. Clark and Lawrence L. Schkede, Statistical Methods for Business Decisions, (Cincinnati, Ohio: Southwestern Publishing Company, 1969).

12. D.R. Mayer, "Factor Analysis Versus Correlation Analysis: Are Substansive Interpretations Congruent?" Economic Geography, Vol. 47, 1971, Supplement, pp. 336-343.

13. P.R. Gould, "On Mental Maps", Michigan Inter-University Community of Mathematical Geographers, No. 9, 1966.

14. Kevin R. Cox, "Suberbia and Voting Behaviour in a London Metropolitian Area", Annals of the Association of American Geographers, Vol. 58, 1968, pp. 111-127.

15. Harry R. Glahn, "Canonical Correlation and its Relationship to Discriminant Analysis and Multiple Regression", Journal of the Atmospheric Sciences, Vol. 25, 1968, p. 23.

16. Ibid., p. 16.

17. J.J. McKeon, "Canonical Analysis: Some Relations Between Canonical Correlation, Factor Analysis, Discriminant Analysis, and Scaling Theory", Psychometric Mongraphs, No. 3, 1966.

18. H. Hotelling, "Relations Between Sets of Variates", Biometrika, Vol. 28, 1936, pp. 321-377.

19. W.W. Cooley and P.R. Lohnes, Multivariate Data Analysis, (New York: John Wiley and Sons, 1971).

20. William A. Love and Douglas K. Steward, Interpreting Canonical Correlations: Theory and Practice, (Palo Alto, Calif: American Institute of Research, 1968).

21. W.W. Cooley, and P.R. Lohnes, Op. cit., Footnote No. 19, p. 176.

22. Ibid., p. 171.

23. William A. Love and Douglas K K. Steward, Op. cit., Footnote, No. 20.

24. M.S. Barlett, "The Statistical Significance of Canonical Correlations", Biometrika, Vol. 32, No. 1, 1941, pp. 29-38.

25. M.S. Barlett, "Canonical Analysis", in H. Borko (ed.), Computer Applications in the Behaviour Sciences, (Englewood-Cliffs, N.J.: Prentice-Hall, 1962), pp. 266-279.

26. Mark I. Alpert, and Robert A. Peterson, "On the Interpretation of Canonical Analysis", Journal of Marketing Research, Vol. 9, 1972, pp. 187-192.

27. D. Michael Ray, "From Factorial Ecology to Canonical Ecology: The Spatial Interrelationships of Economic and Cultural Differences in Canada", Economic Geography, Vol. 47, 1971, Supplement, p. 346.

28. Ibid., p. 347.

29. Ibid., p. 344.

30. Ibid., p. 346.

31. Leslie J. King, Statistical Analysis in Geography, (Englewood Cliffs, N.J.: Prentice-Hall, 1969).

32. Douglas K. Steward and William A. Love, "A General Canonical Correlation Index", Psychological Bulletin, Vol. 70, 1968.

33. D. Michael Ray and Paul R. Lohnes, "Canonical Correlation in Geographical Analysis", Geographia Polonica, Vol. 25, 1973, p. 49.

34. E. Casetti, Multiple Discriminant Functions, Technical Report No. 11, O.N.R., Task No. 389-135. Contract NONR 1228 (26), (Evanston III: Department of Geography, Northwestern University, 1964), p. 1.

35. Bryn Geer-Weotten, A Bibliography of Statistical Applications in Geography, Association of American Geographers, Commission on College Geography, Technical Paper No. 9, (Washington D.C, 1972).

36. Leslie J. King, Op. cit., Footnote No. 31, 1969.

37. Phillip J. Rulon, "Distinction Between Discriminant and Regression Analysis and Geometric Interpretation of the Discriminant Function", Harvard Education Review, Vol. 21, 1951, p. 86.

38. David V. Tiedman, "The Utility of the Discriminant Function in Psychological and Guidance Investigations", Harvard Education Review, Vol. 21, 1951, p. 79.