A MULTIVARIATE PLANNING MODEL   -   CITY STRUCTURE

A Thesis presented in partial fulfilment of the requirements for the
degree of Master of Science in Statistics

by

Peter Crawford

1972

Massey University

Preface

The genesis of this study is post graduate research in Urban Geography at Canterbury University in 1966. At that time a crude multivariate Centroid model of 95 New Zealand towns and cities was constructed. Based upon 60 socio-economic variables two factors for each of the years 1951, 1956 and 1961 were extracted and compared. The present study, which is a considerable refinement upon the earlier research, incorporates not only tremendous advancement in multivariate design methodology and application, but also parallel advancements that have been made in computing facilities over the last five years.

The objective of this research is to construct a multivariate statistical planning model that is both statistically precise and meaningful in its application. Particular emphasis is placed upon the need to organise in a systematic and meaningful manner the increasingly greater variety of statistics that portray urban growth. Stress is placed upon the utility of the multivariate technique as a tool in the author's profession of Town Planning.

Acknowledgements

TABLE OF CONTENTS

I.   MULTIVARIATE METHODOLOGY

1.   Introduction:-

Multivariate methods are Statistical techniques concerned with relationships between variables. These relationships attain a particular level of significance in association with the volume of urban area statistics produced in New Zealand and the need to use such statistics in Town Planning. In particular, recent proposed legislation requiring the establishment of planning policy necessitates a more precise understanding of the nature of relationships between statistics used to delineate city development. This legislation is backed by precedence in decisions of the Town and Country Planning Appeals Board which has already stipulated that their determinations will be based upon planning policy where it exists. Few cities in New Zealand have established such policy. The Planner will therefore be required by statute to derive planning policy which will, on the whole, be obtained from a myriad of statistics all of varying degrees of importance. The problem is to develop a statistical technique which will incorporate and account for statistics used in Planning. The multivariate statistical technique of Factor Analysis appears to have the most potential for such an analysis.

2.   Research Objectives:-

Research objectives in this study are two-fold - firstly to investigate the utility of a multivariate statistical technique in the delineation of urban relationships and hence the definition of planning policy, and secondly to assess problems of data distribution and mathematical meaningfulness inherent in multivariate modelling. Both the former and the latter objectives are analysed in terms of an examination of New Zealand's 18 cities over the 1951-71 period. The Multivariate Factor Analysis method is developed as the mathematical planning model.

3.   Statistics:-

Multivariate statistical analysis is associated with a
considerable body of statistical theory and knowledge which has
developed since the 1940s from the work of Lawley. Earlier and simpler
applications focussed upon univariate and bivariate relationships and
the normal distribution. Much of the less systematic statistical
methodology was developed by the early analytic psychologists; Charles
Spearman, Cyril Burt, Karl Pearson, G.H. Thomson, J.C. Maxwell Garnett,
Karl Holzinger, H. Hotelling, L.L. Thurstone, Galton and others. More
recently, and in particular in the last decade, the development of
computer science and more flexible numerical techniques has led to the
relaxation of computational limitations upon applications of multi-
variate statistical theory. Work by Lawley, Howe, Anderson, Rao and
Maxwell, Carroll, Ferguson, Neuhaus and Wrigley, Saunders and Kaiser
on multivariate factor statistical methodology has been of considerable
importance. At the same time refinement of the eigen-value problem by
numerical analysts - Householder, Rutihauser, Francis and others - has
greatly contributed to developments in multivariate analysis. The
breakthrough by Joreskog in the establishment of a numerical method for
the minimisation of a function of many variables in 1966 is of consider-
able importance. Methodological improvement by Joreskog in collabor-
ation with others in the past few years has meant a simplification of
the application of the technique's improved statistical base. Almost
all of the improvements in the technique has meant an increase in
ability to relate many variables in a statistically meaningful manner.

4.   Planning:-

Town Planning involves the establishment of policy for city
development goals, formulated from an interpretation of the patterns of
urban growth. This interpretation involves prior knowledge from a

determination of not only existing relationships and inter-relationships in cities, but also an understanding of the trends in such relationships and their relative degree of importance. Analysis of this kind, while implied under the Third Schedule of the Town and Country Planning Regulations 1960, is not stipulated.

Basic data used in the establishment of urban area inter-relationships are generally available from Census publications or from carefully designed sample surveys. Much of this material requires interpretation, particularly on complex issues where the outcome of a decision or policy implementation, may be consequential upon the complex interaction of a variety of variables. Indicative Planning in New Zealand has, until recently, been involved in the assessment of individual statistics or simple combinations of such statistics. More often than not, only univariate analysis is undertaken and frequently the population statistic was the sole index used in indicative planning.

Recent decisions of the New Zealand Town and Country Planning Appeals Board have emphasised the need for Planners to take cognizance of the more complex issues in establishing Town Planning policy[1]. The definition of the complex issues of planning require a more refined analysis in terms of the available statistics. The problem of the Planner is to arrange these statistics in a meaningful manner so that they may portray the complex issues and clarify the important aspects of city growth and development.

Outside of the classificatory work of the urban Geographers there has been little research undertaken in this area of statistical application. The American Ecological studies by Shevky-Bell, Haynes, Molotch and others have been concerned with spatial inter-relationships

1 An example is the Board's recent decision in the case G.U.S. Properties Ltd. and others v. Timaru City Council 1971 4 N.Z.T.C.P.A. 12.

and classification within urban areas. Most other studies including the above, criticised widely for lack of methodological framework, do not indicate an incorporation of a planning base.

5.   Experimental Multivariate Planning Model:-

        Multivariate statistical models involving determination of the simple and complex inter-relationships between many different statistics appear particularly suitable for an analysis of the characteristics of cities. Moreover, the multivariate Factor Analysis model has considerable potential as a planning model because it incorporates the principle of parsimony, i.e. the ability to precipitate a simple relationship from a complex combination of many variables. This study is an attempt to construct an experimental Factor Analysis planning model and to examine the relationship between the model and observable reality.

        The study format is in three parts. In Chapter Two the mathematical and statistical framework for the model is established. Chapter Three consists of a detailed analysis of the application of the model to the New Zealand situation. In the final section, Chapter Four, the results and meaningfulness of the model are assessed in terms of the statistical accuracy and usefulness as a planning tool.

II.   MATHEMATICAL MODEL - FACTOR ANALYSIS

1.   Theoretical Framework:-

A multivariate mathematical model forms an ideal analytical base
for demonstrating developments in inter-relationships. Initially the
pattern of variable distributions can be portrayed in a univariate
situation. Then may be considered the bivariate distributions which
describe the relationships between pairs of variables. Multivariate
patterns in turn may be portrayed in a Factor Analysis model. The
relationships are essentially linear, but are systematic and the factor
model is developed stage by stage (Figure 1).

Univariate (1)	Bivariate (1,1) (1,2) ...... (1,n)	
Univariate (2)	Bivariate (2,2) (2,3) ..... (2,n)	Multivariate
Univariate (3)	Bivariate (3,3) (3,4) .... (3,n)	Model
Univariate (4)	Bivariate (4,4) (4,5) ... (4,n)	
Univariate (5)	Bivariate (5,5) (5,6) .. (5,n)	
Univariate (6)	Bivariate (n,n)	
Univariate (n)		

Stage I   Stage II   Stage III
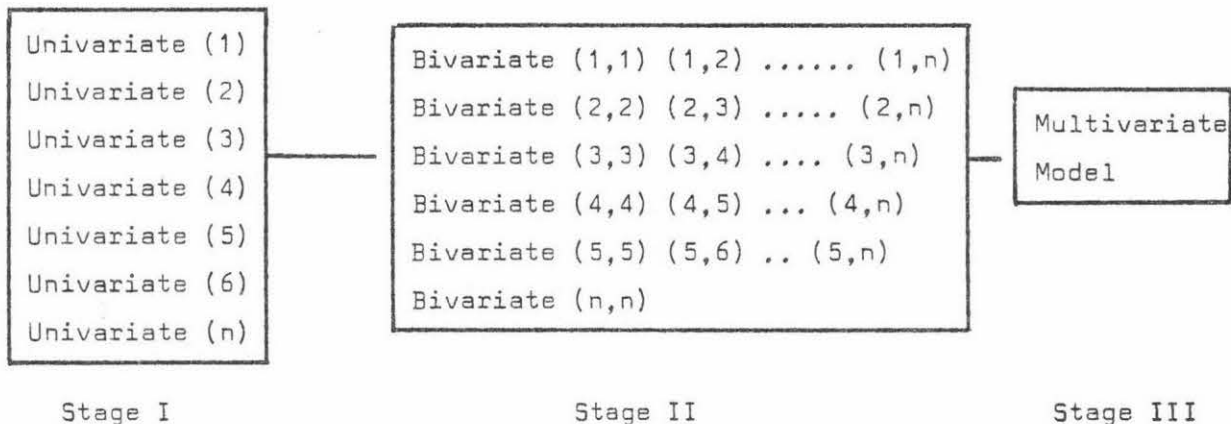
Figure 1. Staged Development of the
Multivariate Factor Model

The advantage of the technique is that in any particular application an
interpretation may be placed upon the various stages of the structuring
of the model.

2.   Data Cube:-

Consider a set of variables or characteristics, X_1, X_2 ....X_n
describing particular entities over a set time period. A standard 'Data
Cube' is formed. Such variables are selected on the basis of a particular
hypothesis or research goal. In this instance the formation of a Data
Cube which describes not only entities and their characteristics also
allows for occasions, provides the basis for an analysis overtime. Such
analysis forms a fundamental structure in the delineation of a planning
model. The data cells of the three dimensional Data Cube form the basis

for the model. In this particular research application occasions are combined in the final model with entities, and the standard data cube becomes two-dimensional. Separate time slices are used to study the structure of this model. The factoring matrix, therefore, conforms to the R-factor analysis. The significance of the datum cell is in the patterns of variation between characteristics over entities. Characteristics are the variables.

occasions

entities

characteristics   datum cell
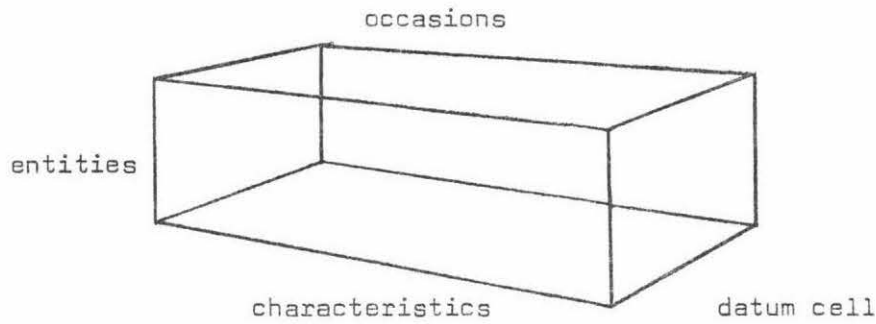
Figure 2. The Data Box

3.   Means and Standard Deviations:-

The mean is a central value of a characteristic calculated as

$$\overline{X}_i = \sum_j X_{ij}/N$$

and indicating the general numerical location of the characteristic. Averaging of one characteristic for different entities at different points in time can reveal a simple pattern of change or a trend.

$$\overline{X}_{i,t_1} \; ; \; \overline{X}_{i,t_2} \; ; \; \overline{X}_{i,t_3} \; \ldots\ldots\ldots$$

$$t = \text{time} \qquad i = 1, \text{———— } n$$

The measure of location, however, may in particular instances, not take cognizance of the arrangement or spread of the individual values of the characteristics. Thus, in some situations the mean value may not provide

enough information about the data distribution. Therefore, a measure of spread may be particularly useful in demonstrating patterns of change. Most commonly used is the variance value for the variable.

$$\sigma_i^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

or its square root, the standard deviation,

$$\sigma_i \quad \sqrt{\frac{\sum (x_i - \bar{x}_1)^2}{N}}$$

which has the advantage of being on the same scale as the variable. For a normal variable, 68.26% of the sample lies within one standard deviation of the mean, etc. Thus, as a consequence a more precise description of data distribution is possible.
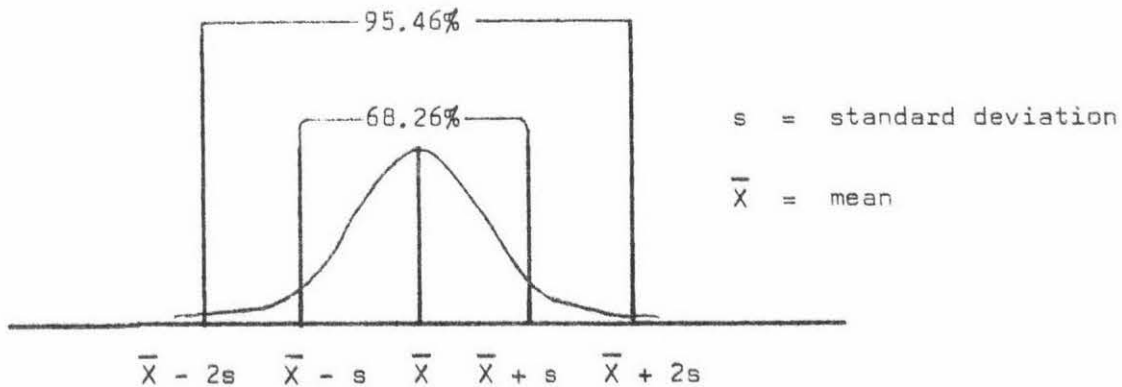
Figure 3. Areas Under a Normal Curve

Hence the standard deviation for a particular variable assessed for different entities at different points in time can reveal a pattern of change or a trend,

$$\sigma_{i,t_1} \quad ; \quad \sigma_{i,t_2} \quad ; \quad \sigma_{i,t_3} \quad \ldots\ldots\ldots$$

$$t = \text{time}, \quad i = 1,2, \quad \ldots\ldots n.$$

More so, if the changes in standard deviations are interpreted with the patterns of change associated with the development of the average values.

4.   Correlation and Covariance:-

Until now the basic descriptive statistics have been associated with univariate situations. Fundamental in data analysis is the bivariate consideration - the pattern of relationships between two variables. The measures of covariance and correlation demonstrate the bivariate relationship. In the latter instance, however, the measure is a scaled quantity while the former retains the numerical data distribution.

$$cov(x_i, x_j) = \sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)$$

$$r_{i,j} = \frac{\sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\left[\sum (x_i - \bar{x}_i)^2\right]\left[\sum (x_j - \bar{x}_j)^2\right]}}$$

Further, $r_{i,j}^2$ represents the amount of variance which the two variables have in common. Either the covariance or the correlation coefficient can be used in an analysis of trends in a particular pair of characteristics. A relationship between two variables may intensify and therefore there will be greater inter-dependence. On the other hand, the converse situation may apply.

$$r_{i,j,t_1} \quad ; \quad r_{i,j,t_2} \quad ; \quad \ldots\ldots\ldots\ldots$$

$$i,j = 1,2,\ldots\ldots n, \quad t_1,t_2\ldots\ldots = time$$

$$cov_{t_1}(x_i,x_j) \quad ; \quad cov_{t_2}(x_i,x_j) \ldots\ldots\ldots$$

Both methods provide systematic measures which may be used to define changes in a bivariate relationship over a particular time period.

5.   Principal Component Multivariate Model:-

Unlike partial, multiple and canonical correlations which are used to analyse the dependence structure of a multinormal population, the primary problem in correlation is the definition of dependent and indep-

endent variables. While the choice of dependent variable may be based upon response patterns and hence the research hypothesis, it is inevitable in a multivariate situation that   the responses are symmetric or there are no a priori patterns of causality available.

Techniques developed to establish a dependence structure of observed responses based upon hypothetical independent variables come within the general category of Factor Analysis. Such statistical techniques attempt to define those hidden factors which have generated the dependence relation between, and the variation in, the responses. Observable variables are represented as functions of a smaller number of latent factor variables. These functions are such that they will generate the covariances or correlations amongst the responses. In this study we are concerned with generating the correlations amongst responses. The objective of the technique is to establish from amongst the responses of many variates a more simple or parsimonious description of dependence structure. It is assumed that the generating model is linear in form.

The principal component model developed by K. Pearson as a method of fitting planes by orthogonal least squares and extended by Hotelling for analysing correlation structures is the simplest of the Factor models and it is usual to use this model as the first step in estimating the structure of a factor model. The technique has widespread use in a variety of fields including human biology, cognitive psychology, mineralogy.

The model which merely partitions the variance amongst the computed components is derived from $X_1, \ldots \ldots \ldots X_p$ random variables with multivariate distribution mean vector $\underset{\sim}{\mu}$ and covariance matrix Σ Both the elements of $\underset{\sim}{\mu}$ and Σ are finite with the rank of Σ being $r \leq p$ and that the q largest characteristic roots

of $\sum$ are distinct. Further, an N x p data matrix is established from a sample of N independent observation vectors.

$$X = \begin{bmatrix} x_{11}\ldots\ldots\ldots\ldots\ldots x_{1p} \\ \\ \\ \\ \\ x_{N1}\ldots\ldots\ldots\ldots\ldots x_{Np} \end{bmatrix}$$

Note that neither $\sum$ nor X need be of full rank p, and further $\sum$ need not contain more than one characteristic root. Full rank, however, ensures simplicity in structure description and is generally assumed in practice.

An estimate of $\sum$ is either the variance-covariance matrix or the correlation matrix R. The latter is preferred instead of the former because of the scaling properties of the correlation coefficient. The first principal component of the observations X is the linear compound

$$Y_1 = a_{11}X_1 + \ldots\ldots + a_{p1}X_p$$

of the responses whose sample variance

$$s_{Y_1}^2 = \sum_{i=1} \sum_{j=1} a_{i1}a_{j1}s_{ij}$$

$$= \lambda_1 \quad \text{(The largest characteristic root)}$$

Continual factoring generates linear compounds of the original variates which account for a progressively smaller amount of the variance. The significant features of the model are that:-

    a) the principal component analysis factorises R

    b) principal component analysis factorization is unique

Because of the model's inherent characteristics it is therefore

possible to construct principal component models portraying the relation-
ships between many variables at different points in time. Moreover, a
comparison between the models can be attempted on the basis of the changes
in relationships and variation.

$$\text{Principal Component} = PC$$

$$PC_{1,t_1} \quad ; \quad PC_{1,t_2} \quad ; \quad PC_{1,t_3} \quad \cdots\cdots\cdots$$

$$PC_{2,t_1} \quad ; \quad PC_{2,t_2} \quad ; \quad PC_{2,t_3} \cdots\cdots\cdots$$

$$\cdot \qquad\qquad \cdot \qquad\qquad \cdot$$
$$\cdot \qquad\qquad \cdot \qquad\qquad \cdot$$
$$\cdot \qquad\qquad \cdot \qquad\qquad \cdot$$

$$PC_{p,t_1} \quad ; \quad PC_{p,t_2} \quad ; \quad PC_{p,t_3} \cdots\cdots\cdots$$

$$PC_{i,t_j} \quad = \quad \text{Principal Component}$$
$$i = 1\ldots\ldots p, \quad t = \text{time} \quad j = 1, \ldots \text{ end of period.}$$

Since the correlation between the original variables and the
individual components can be obtained through the formula $a_{ij}\sqrt{\lambda_j}$
where a_{ij} are the estimated component loadings and λ_j the characteristic
root of the jth component, it is possible to relate components and
variables. Moreover, a simpler or parsimonious description is now possible
in terms of a single linear component if it accounts for the greater part
of the variance of the original variables.

6.   Factor Analysis Multivariate Model:-

Despite its simplicity the Principal Component Multivariate model
has shortcomings. While the model does factorise the covariance matrix
the factorisation is more of a transformation rather than the consequence
of a fundamental model for covariance structure. Further, the forms of
components are not invariant under response scale changes and there is no
strict criteria for deciding when sufficient variance has been accounted

for. It is significant that no provision is made for error variance estimations.

This partition of the variance relates to the factor model in that "each response variate is represented as a linear function of a smaller number of unobservable common factor variates and a single latent specific variate. Common factors generate the covariances among the observable responses while the specific terms contribute only to the variances of their particular responses" (Morrison, 1967). This refinement in description over the Principal Component model is, however, gained at the expense of two assumptions:

a) the observations arose from a multinormal population of full rank.

b) the exact number of common factors can be specified before analysis.

Both these assumptions are an essential part of the Factor philosophy.

The mathematical model is based upon a multivariate system of p responses characterised by observed random variables $x_1 \ldots \ldots \ldots x_p$. x_i having a nonsingular multinormal distribution. The model is of the form:

$$x_1 = a_{11}y_1 + \ldots \ldots + a_{1m}y_m + e_1$$

$$\cdot \qquad \cdot \qquad \qquad \cdot \qquad \cdot$$
$$\cdot \qquad \cdot \qquad \qquad \cdot \qquad \cdot$$
$$\cdot \qquad \cdot \qquad \qquad \cdot \qquad \cdot$$

$$x_p = a_{p1}y_1 + \qquad + a_{pm}y_m + e_p$$

y_j = jth common factor variate, j = 1,2 ..... m

a_{ij} = parameter reflecting importance of jth factor in the composition of the ith response (loading of the ith reponse on the jth common factor)

e_i = ith specific factor variate

In matrix notation the factor model becomes $X = \Lambda Y + \mathcal{E}$

Now let the m common factor variates in y be distributed normally with zero means and unit variances i.e. $y \sim N(0,1)$. Further assume $\mathcal{E} \sim N(0, \Psi_i)$. Ψ_i being the specificity of the ith response.

$$\Psi = \begin{bmatrix} \Psi_1 & \Psi_2 & & \\ & & \ddots & \\ & & & \Psi_p \end{bmatrix}$$

Moreover, it is required that the variates y and $\mathcal{E}$ be independently distributed. Variance on the ith response from the properties of the latent variates are:

$$\sigma^2_i = a_{i1}^2 + \cdots\cdots\cdots + a_{im}^2 + \Psi_i$$

and the covariance of the ith and jth response variate as

$$\sigma_{ij} = a_{i1}a_{j1} + \cdots\cdots\cdots + a_{im}a_{jm}$$

That is $\qquad \sum = \Lambda \Lambda' + \Psi$

Now $\qquad \sigma^2_i - \Psi_i = \sum_{j=1}^{} a_{ij}^2$

are the diagonal elements of $\Lambda\Lambda'$ and are called the communalities of the responses. a_{ij} is the covariance of the ith response with the jth common factor. However, when $\sum$ is the population correlation matrix, R, the a_{ij} as in the case of the principal component model is the correlation of responses and common factors.

The basic problem in factor analysis is the determination of the a_{ij} with the elements of Ψ following as a constraint imposed upon the communalities. The fundamental aspect of the factor model, however, is that linearity becomes part of the research philosophy. Further, the research hypothesis is related directly to the number of factors. If there is not a fit between the hypothesised factors and the observed values then

both the factor hypothesis and the linearity hypothesis may be rejected.
Normally further factors may be hypothesised to test the fit. Some work
by MacDonald has been undertaken on the problem of non-linearity rejection,
but this work is in its early stages of development. Linearity is assumed
throughout this research application.

Similar to the Principal Component technique, but at a more
refined level of statistical analysis it is possible to relate not only
variables to the factors, but also to construct factor models representing
different analyses at different points in time. In addition, it is there-
fore possible to attempt a comparison between models on the basis of the
changes in relationships and variations.

$$F_{1,t_1} \quad ; \quad F_{1,t_2} \quad ; \quad F_{1,t_3} \ldots\ldots\ldots$$

$$F_{2,t_1} \quad ; \quad F_{2,t_2} \quad ; \quad F_{2,t_3} \ldots\ldots\ldots$$

$$F_{m,t_1} \quad ; \quad F_{m,t_2} \quad ; \quad F_{m,t_3} \ldots\ldots\ldots$$

$$F_{i,t_j} = \text{Factor}$$

$i = 1, \ldots\ldots m$ number of factors hypothesised

$t_j = 1, \ldots\ldots$ end of period under study

7.   Varimax Rotation:-

As a corollary to factor production, maximisation of associations
between factors and variables may be obtained by a rotation. The signif-
icant feature of the Principal Component model is not only the unique
factorisation but also the orthogonal relationship between components.
Thus Components are independent and theoretically unrelated. On the other
hand, the Factor model does not have the condition that the sums of the
squares become successively smaller as one passes from the first to the

final factor. As a consequence orthogonal rotation of the loading matrix $\bigwedge$ does not affect the generation of covariances. In fact, it is to be appreciated that in factor analysis an infinity of loading matrices may be obtained from the correlation matrix.

As a result, a more "meaningful" application of the concept of simple structure may be applied to make a clearer definition of loadings. Further, it is reiterated by some that the "particular configuration of numbers obtained in an unrotated factor analysis loading matrix is largely a function of the method used to extract the eigenvalues and eigenvectors and therefore may have no empirical meaning". The concept of simple structure is a non-mathematical technique setting out several criteria for a rotation of factors:

a) existence of a positive manifold (i.e. a minimum number of negative values in the factor loading matrix)

b) a small number of high loadings and a large number of near zero loadings

c) each row of the factor loading matrix to have at least one near zero factor loading and at least one other large positive loading

d) it must account for the relative position of zeros and important high loadings

The principle is one of an application of Occam's Razor to the factor loading matrix and is felt by most to give a better or improved description to the factors.

Most commonly used, and the technique used in this application, is Kaiser's (1958) varimax rotation method which maximises the fourth power of the factor loadings and therefore maximises the scatter amongst the loadings. As the method retains the property of orthogonality which leaves the factors uncorrelated it is as a consequence widely used. In general a transformation matrix 'I' is developed over a cycle of rotations with the

angle of each rotation chosen such that a function 'U' of the factor

matrix is maximised.

$$U = m \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{g_{ij}}{h_i} \right]^4 - \sum_{j=i}^{m} \sum_{i=1}^{n} \left[\frac{g_{ij}^2}{h_j^2} \right]^2$$

m = number of factors

n = number of variables

g_{ij} = element of factors matrix under rotation for ith variable
            jth factor

h_i^2 = communality

The rotated Factors assume particular importance because of this relation-

ship, particularly in respect of the relationship between variables,

factors and factor scores. Factor scores for particular entities are

derived from the factors and demonstrate the relationship between individ-

ual entities in terms of the hypothesised factor constructed from many

related variables. The degree of rotation and hence the dominance of a

particular variable must be assessed in terms of the rotation. Different

rotations applied for separate models representing different points in

time tend to highlight differences between dominant variables.

8.   Multivariate Statistical Factor Model:-

        If occasions are combined with entities and the resultant two-

dimensional data base is factor analysed a more general combined multi-

variate statistical factor model may be constructed. The model is still

the simple format

$$X = \Lambda Y + \mathcal{E}$$

In this instance, however, the entities become entities for different

occasions with the unique and distinct characteristics being associated

with each particular point in time. Further, the descriptive basis allows

the use of the factor hypothesis to delineate aspects of particular

entities, and comparisons can be made between the different factor models

$$F_{1,(t_1, t_2, \ldots t_k)} \quad ; \quad F_{2,(t_1, t_2, \ldots t_k)} \quad ; \quad \ldots\ldots\ldots$$

$$\ldots\ldots\ldots F_{m,(t_1, t_2, \ldots t_k)}.$$

F = factor

F_i $i = 1, 2, \ldots m$ number of hypothesised factors

t_j $j = 1, 2, \ldots k$ time period under consideration

The particular significance of such a combined model is two-fold. Firstly, not only can the model be rotated with some consistency, but also the singular time scale models can be related in terms of specific variable - factor relationships. Secondly, hypothesised factor comparisons can be made. In this latter instance the factor scores, which relate entities and factors defining variation can map a particular trend in patterns of change as shown in the scores over a set time period. The former instance allows a check between the final factor model and observed reality. In fact the staged development from simple arithmetic means, variances, covariances, correlation, single time scale principal component and simple time scale factor models relates the cumulative model to the observable situation. A pattern of growth and inter-relationships may be defined in a complex multivariate situation through such a refinement of the application of the multivariate factor model.

9.   Data Distribution:-

Basically the Factor Analysis model outlined focuses upon a delineation of similarities and differences, relationships and associations. Kendall (1957) stipulates that the application of Factor Analysis is a search for inter-relationships rather than dependency. The preciseness of the definition of such inter-relationships will be dependent upon a

variety of factors. Not the least amongst these factors will be the data distribution.

It is only in recent years that a more sophisticated philosophy of factor analysis has been established from a multivariate normal hypothesis. Lawley's work in the early 1940s demonstrated the need for a sound data base while applications of the maximum likelihood philosophy has given a more formal theoretical statistical framework. Joreskog's breakthrough in 1967 has meant this theoretical framework can be applied to specific research applications. More specifically factors can be tested for significance in terms of a normal sampling situation.

Development of such a body of theory and techniques for application is a breakthrough of considerable importance, but is not undertaken in this study because the need to develop the technique, not yet available in New Zealand, was beyond the scope of the study. There was, however, a need to establish a reasonably consistent framework in which to develop the model. The normality of the data distribution and its effect on the fundamental model is the secondary objective of this piece of experimental research.

Data Distribution and normality considerations assume a particular degree of importance when it is considered that a bivariate normal distribution has the property that the regression relation between two variables is linear (Kendall and Stuart, 1958, vol. 1, p. 387). Further linearity in the bivariate inter-relationship of the data is a basic assumption of the model. Moreover, a sufficient condition for the correlation coefficients to be a true measure of statistical independence between two variables is that the bivariate distribution of the variables be normal. Thus, the importance of the normality of data distribution is

a prime consideration not only in correlation, but in the final factor
model.

In the development of the factor model it is proposed to examine
this relationship between normality of data distribution and the model at
its various stages of construction. Not only will the effect of normal-
ising data be studied in derived correlation coefficients, but also the
implications in terms of the fundamental factor model which is constructed
from the correlation coefficients.

It is proposed to develop the model from basic data and repeat the
application using the same data with a normal transformation. Both models
will be assessed - the crude data model and the statistically exact model.
Final examination will be the **rela**tionship between the model and its
ability to portray the nature of the variation in relationships between
variables.